



**UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO**

**DIPARTIMENTO DI  
BIOSCIENZE, BIOTECNOLOGIE  
E BIOFARMACEUTICA**

**Orientamento Consapevole  
Biotecnologie Innovative**

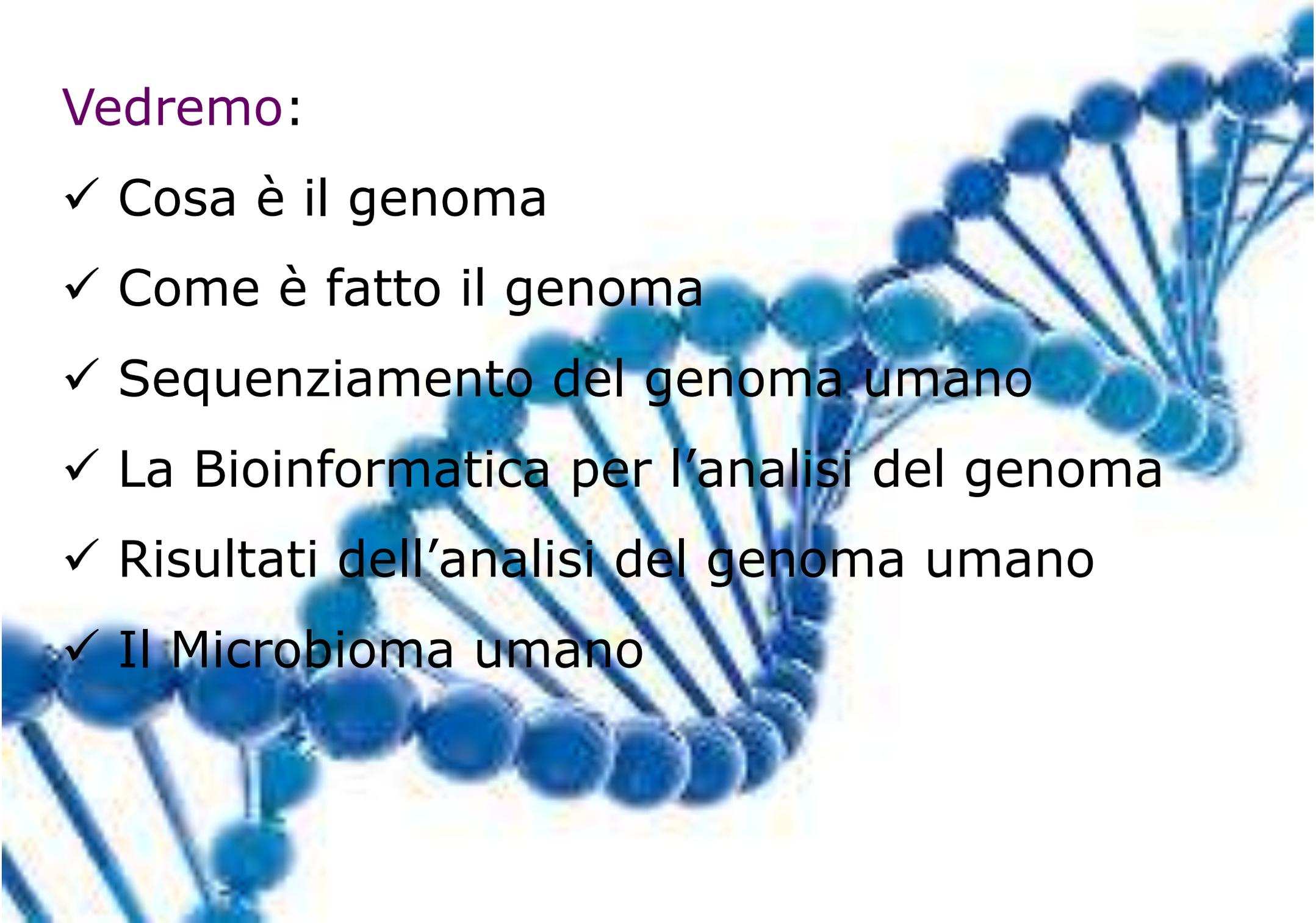
**Bioinformatica e analisi del  
genoma umano**

**Dott.ssa Anna Maria D'Erchia**

**Bari, 1 marzo 2017**

## Vedremo:

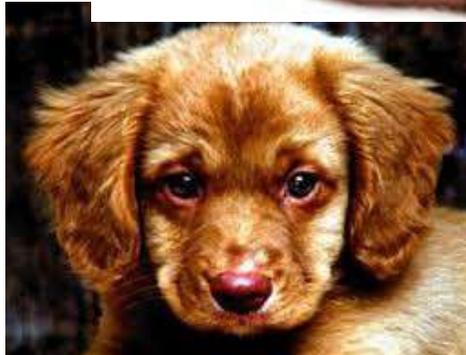
- ✓ Cosa è il genoma
- ✓ Come è fatto il genoma
- ✓ Sequenziamento del genoma umano
- ✓ La Bioinformatica per l'analisi del genoma
- ✓ Risultati dell'analisi del genoma umano
- ✓ Il Microbioma umano



# Che cosa è il genoma?

Il **genoma** è l'insieme di tutte le informazioni biologiche, depositate nella sequenza del DNA, necessarie alla costruzione e al mantenimento di ogni organismo vivente.

Può essere paragonato ad un'enorme manuale in cui sono contenute tutte le istruzioni che regolano lo sviluppo e il funzionamento di ogni organismo.



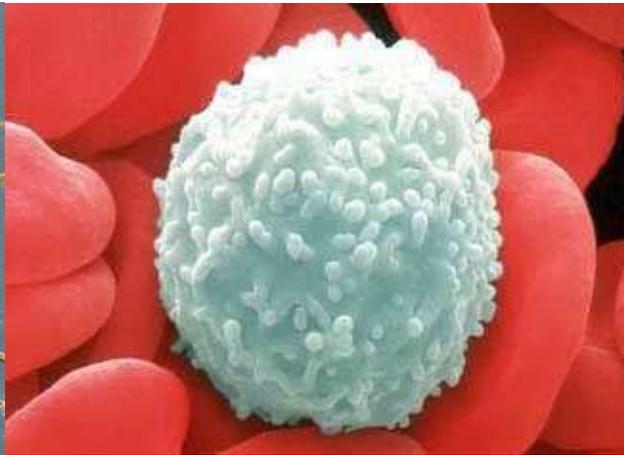
# Come è fatto il genoma?

Il genoma è "scritto" in un composto chimico chiamato **DNA** (**D**eoxyribo**N**ucleic **A**cid, acido desossiribonucleico)

Il **DNA** è identico per tutte le cellule di un individuo, quindi tutte le cellule hanno le stesse informazioni, ma non le utilizzano tutte allo stesso modo.



Neurone



Leucocita

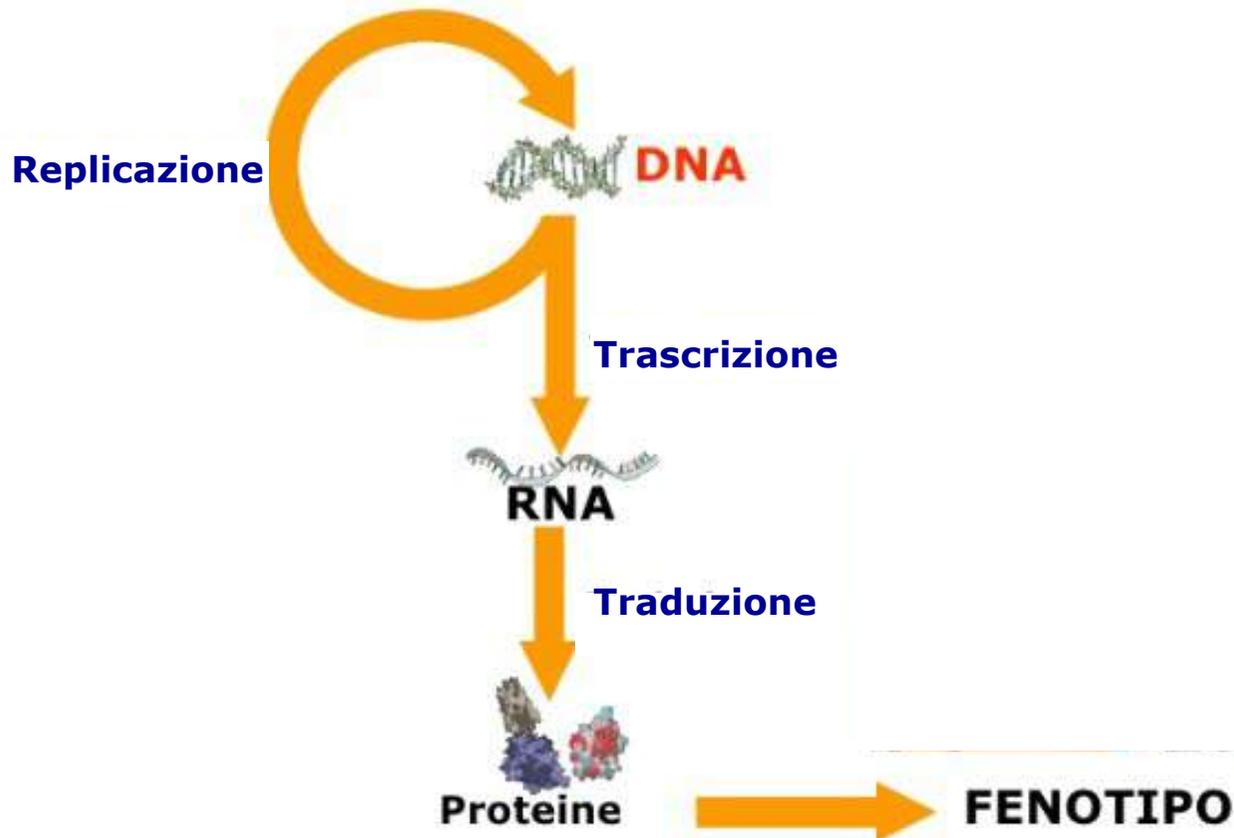


Globuli rossi

# Dal Genoma all'Organismo

L'informazione racchiusa nel DNA è espressa attraverso il processo di **trascrizione** (DNA  $\longrightarrow$  RNA) e **traduzione** (RNA  $\longrightarrow$  proteine) secondo le regole del codice genetico

## Flusso dell'informazione genetica



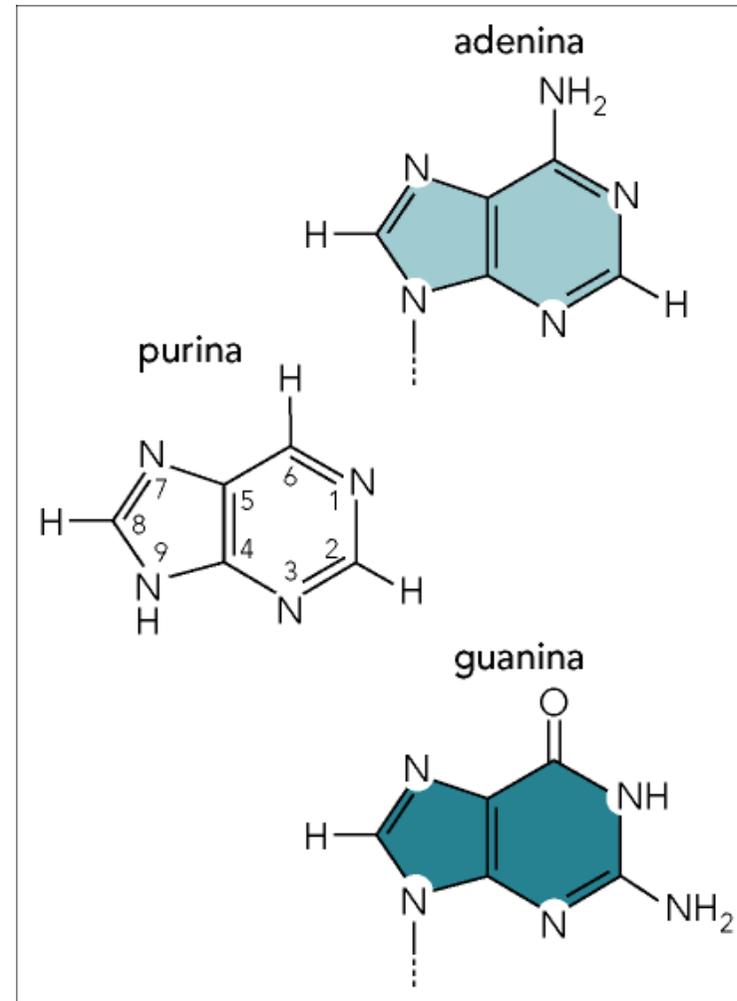
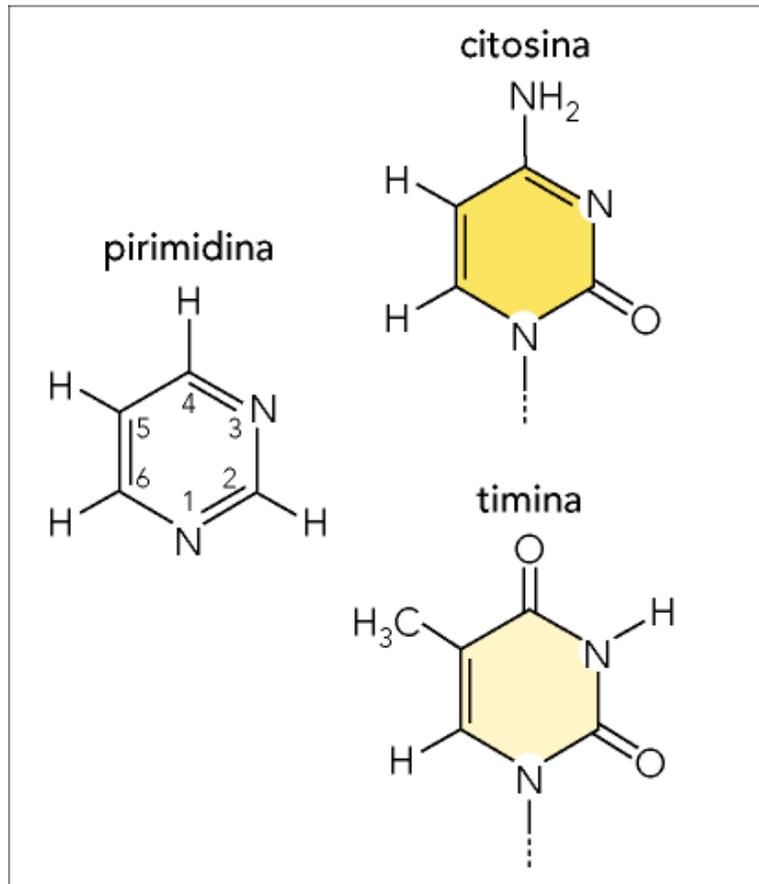
## II DNA

Il **DNA** è un **polimero** costituito da unità chiamate **nucleotidi**, ciascuna composta da:

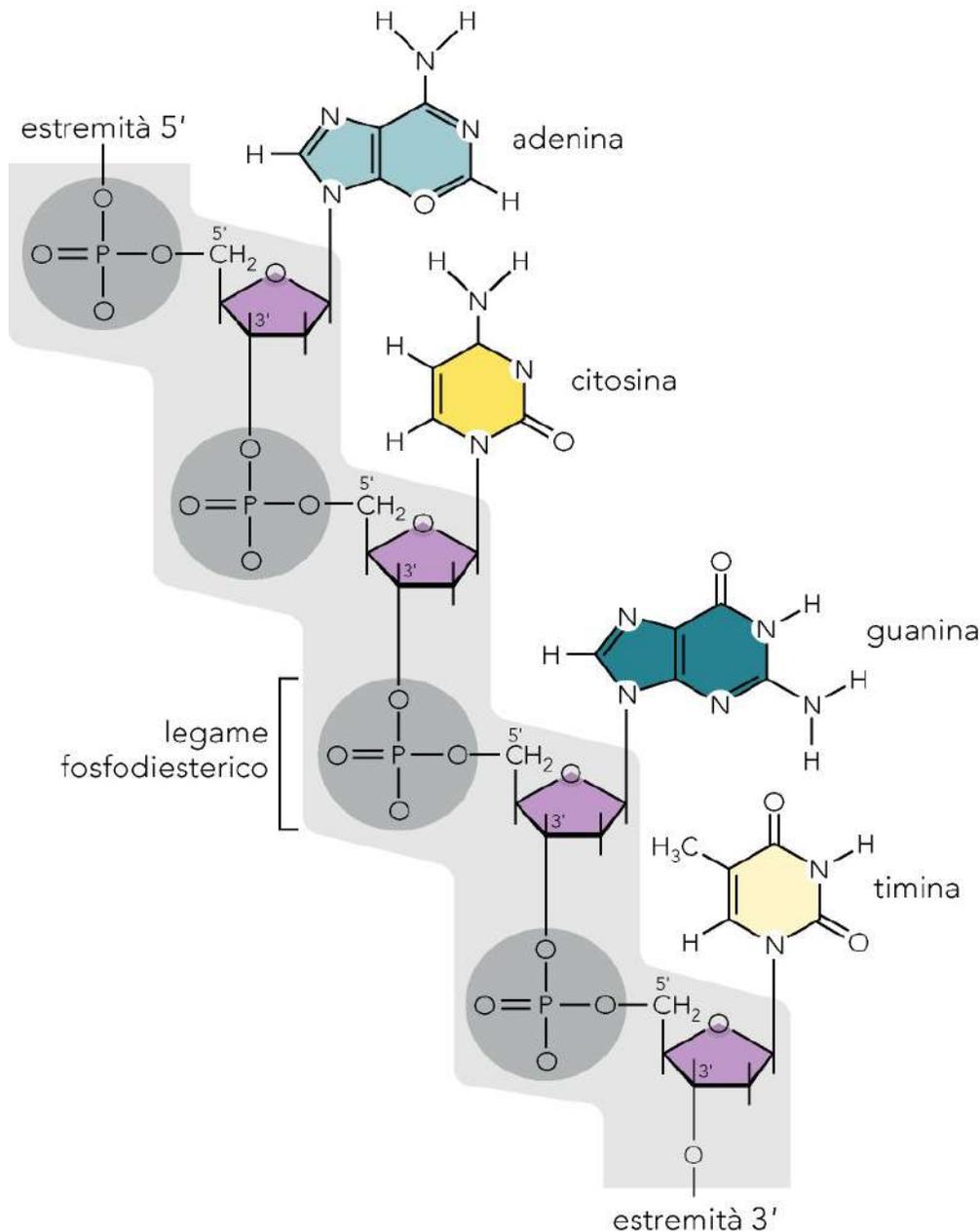
1. uno zucchero a 5 atomi di carbonio (deossiribosio)
2. una base azotata (tra 4 tipi diversi)
3. un gruppo fosfato



# Le basi azotate del DNA: l'alfabeto della vita



# La catena polinucleotidica del DNA



I nucleotidi sono legati tra loro a formare una catena polinucleotidica da **legami fosfodiesterici**.

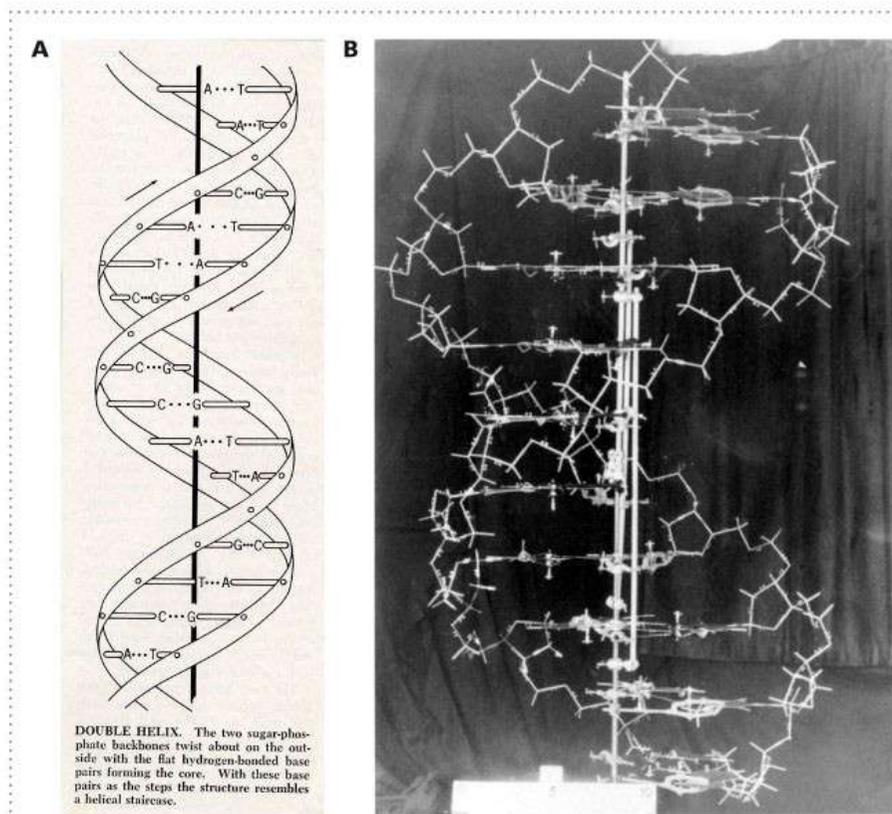
Lo scheletro è costituito dall'alternanza degli zuccheri e dei gruppi fosfato. Le basi azotate sporgono lateralmente dallo scheletro.

I legami fosfodiesterici conferiscono una ben definita **polarità** alla catena del DNA.

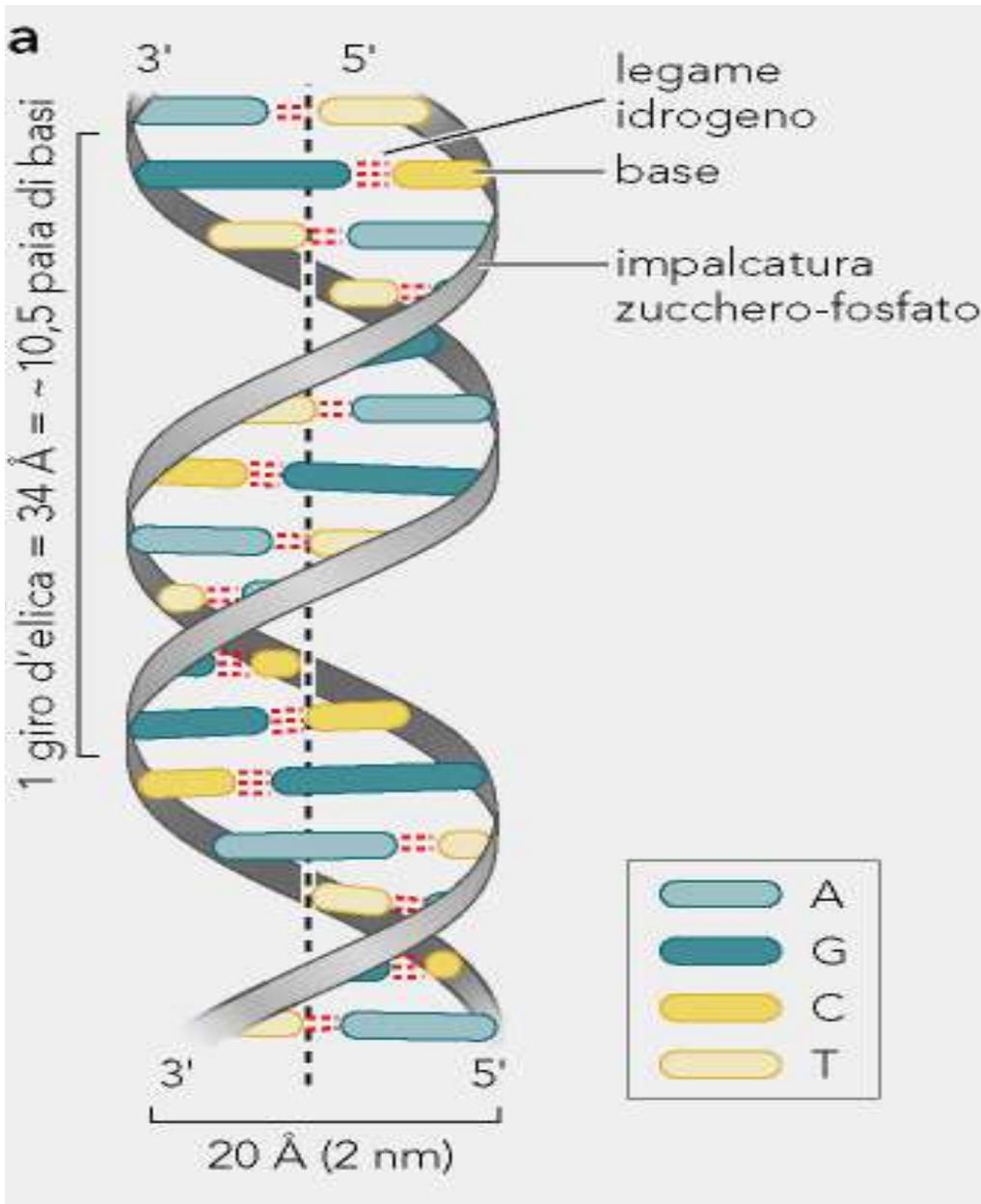
# La struttura del DNA

James Watson e Francis Crick nel 1953 dedussero, anche grazie agli studi condotti da Rosalind Franklin e Maurice Wilkins, la struttura tridimensionale del DNA.

Vincono il premio Nobel nel 1962 (insieme a Wilkins).



# La doppia elica del DNA



Il DNA ha una struttura a **doppia elica**: i due filamenti elicoidali si avvolgono attorno ad un asse centrale e sono tenuti insieme mediante appaiamento tra basi

I due filamenti di DNA sono direzionati (5'→3') e hanno un'**orientamento antiparallelo**:

un filamento ha orientamento **5'-3'**

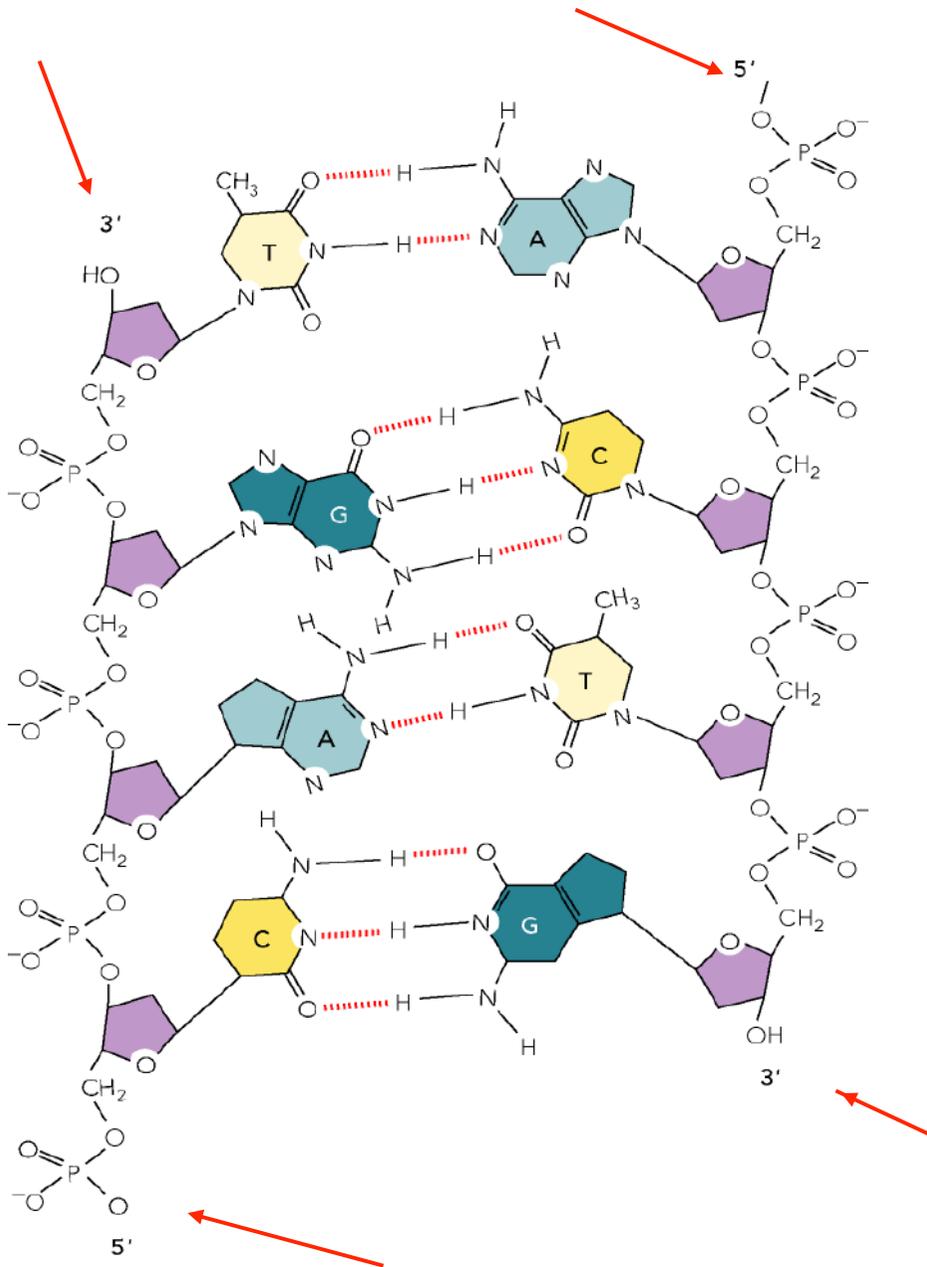
un filamento ha orientamento **3'-5'**

Le impalcature di zucchero-fosfato rimangono all'esterno della struttura

Le basi azotate sono perpendicolari all'asse dell'elica e separate da 3.4 Å

La struttura elicoidale si ripete ogni 34 Å, e quindi ci sono 10 basi per giro dell'elica

# La doppia elica del DNA



I due filamenti sono tenuti insieme mediante formazione di **legami idrogeno tra le basi**, secondo una regola ben precisa:  
ADENINA --- TIMINA  
CITOSINA ---- GUANINA

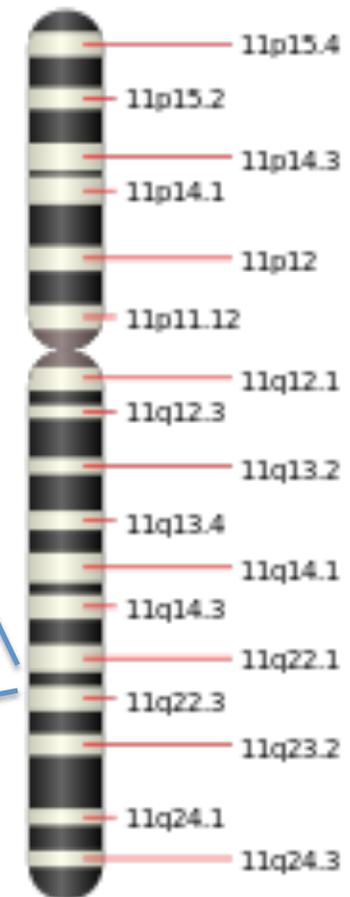
I due filamenti di DNA hanno **sequenze complementari**

# La sequenza del DNA

La **sequenza della molecola di DNA** è definita dalla successione delle quattro basi azotate: adenina, timina, citosina, guanina che vengono indicate rispettivamente con le **lettere A, T, C, G**.

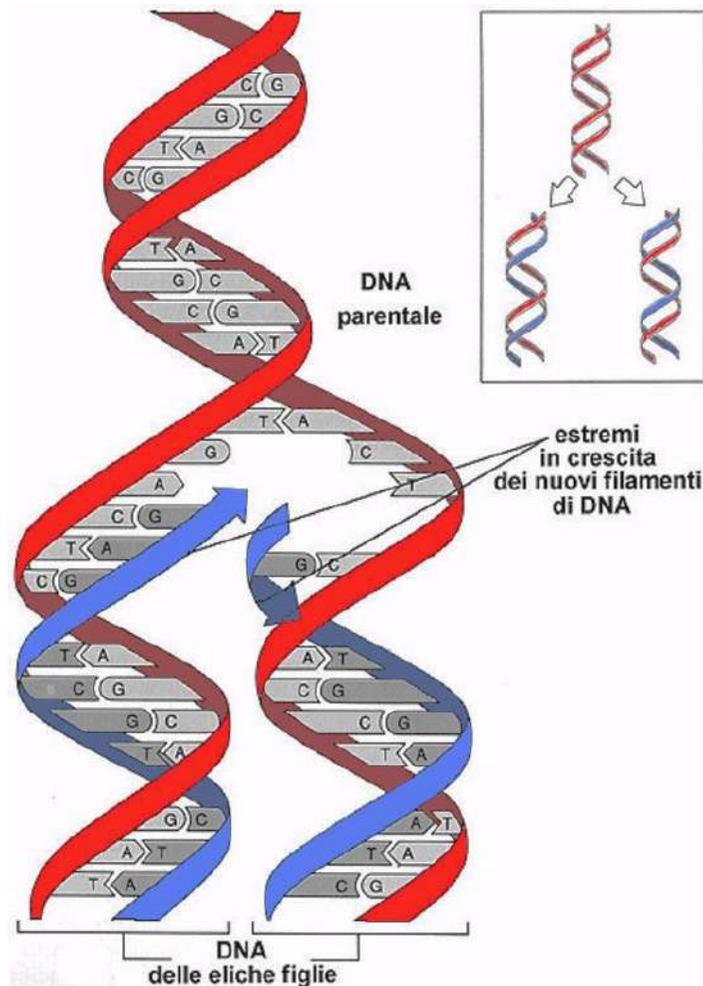
**4 lettere, nessuna punteggiatura**

```
atggcaattaaaattggtatcaatgggttttggtcgtatcggccgtatcgtattccgtgca
gcacaacaccgtgatgacattgaagttgtaggtattaacgacttaatcgacggtgaatac
atggcttataatggtgaaatgatgattcaactcacggtcgtttcgacggcactggtgaagt
aaagatggttaacttagtgggtaatggtaaaactatccgtgtaactgcagaacgtgatcca
gcaacttaaactggggtgcaatcgggtgtgatatcgctggtgaagcgactggtttattc
ttaactgatgaaactgctcgtaaacatatcactgcaggcgcaaaaaaagttgtattaact
ggcccatctaaagatgcaaccctatgttcgttcgtgggtgtaaacttcaacgcatacgc
ggtcaagatatcgtttctaacgcattctgtacaacaaactgtttagctcctttagcacgt
caaaaatagggttaatatgaatctcgatctccatgttgcattcgtattcaacaacaagc
caaaactcgtacaaatgatgaccgcacttcgctataaagaacacggcttcgagatatctct
tgaaaaactttcaagagcaactcaactttctcgagcattgcttgctcacaatatt
gacgtacaagataaaaatcgccatttttgcccataaatggaacggtgggtgttcatgaa
actttcgggatcaaagatgggttaatgaccactgttcacgcaacgactacaatcgttgac
attgcgaccttacaatttcgagcaatcacagtgccattttacgcaaccaatacagcccag
caagcagaatttatcctaaatcacgccgatgtaaaaattctcttcgctcggcgatcaagag
caatacgatcaaacattggaaattgctcatcattgtccaaaattacaaaaaattgtagca
atgaaatccaccattcaattacaacaagatcctctttcttgcaacttg
```



# La replicazione semiconservativa del DNA

Durante la replicazione del DNA, i due filamenti che costituiscono la doppia elica sono srotolati da un enzima chiamato **elicasi** e ciascuno dei due filamenti funge da stampo per la sintesi di un filamento ad esso complementare.



Le due molecole di DNA generate dalla replicazione hanno sequenza identica alla molecola di DNA parentale e contengono ciascuna un filamento originario e un filamento di nuova sintesi.

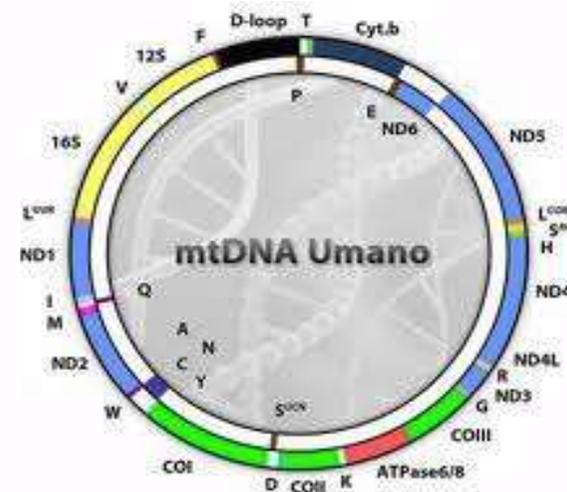
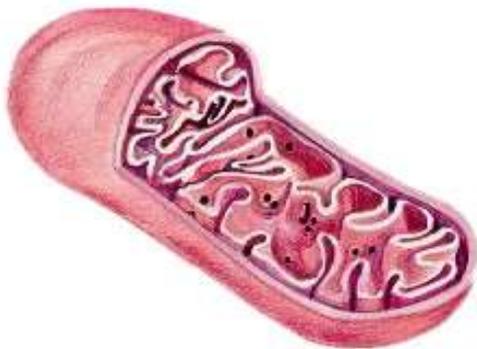
# Il genoma umano

Il genoma umano, similmente ai genomi di tutti gli animali pluricellulari, è costituito da due componenti distinte, il **genoma mitocondriale** e il **genoma nucleare**.

- **Genoma mitocondriale**

- E' una molecola di DNA circolare di circa 16000 nucleotidi, presente in copie numerose nei mitocondri, gli organelli che generano energia.

- Contiene informazioni per la sintesi di alcune proteine, rRNA e tRNA.



# Il genoma umano

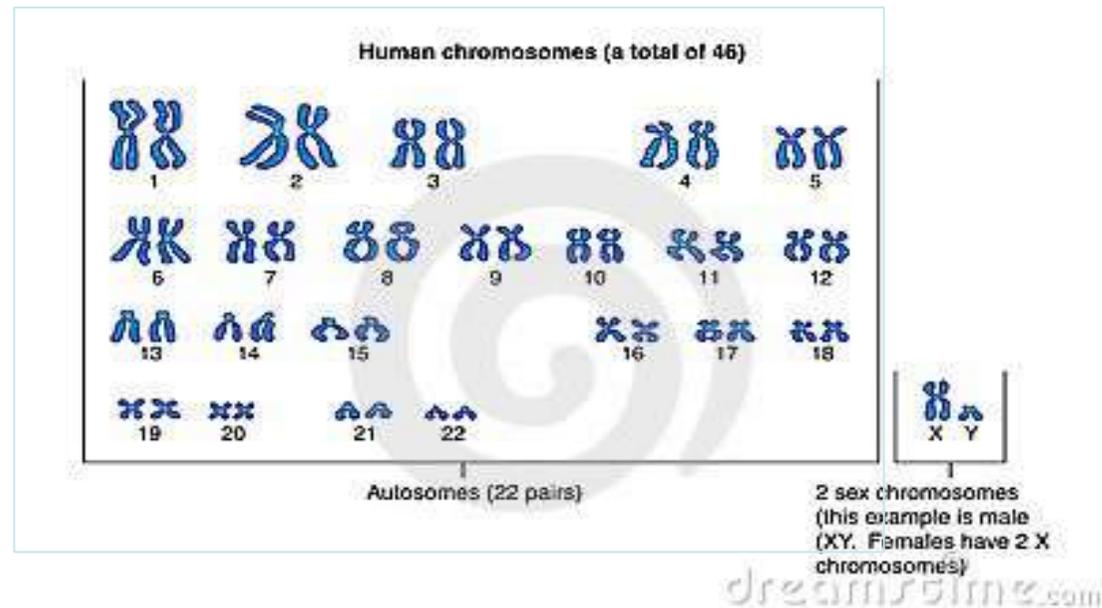
## • Il Genoma nucleare

Contiene circa  $3 \times 10^9$  nucleotidi organizzati in **Cromosomi**

Le cellule somatiche sono **diploidi**, contengono cioè due copie di ciascun cromosoma:

- 22 coppie di autosomi (coppie di cromosomi uguali)
- 1 coppia di cromosomi sessuali, X e Y
  - XX nelle donne, XY negli uomini

Le cellule **sessuali**, o gameti, sono **aploidi**, contengono cioè solo una copia per cromosoma (23 cromosomi in tutto)



# Il genoma è organizzato in Cromosomi

Nelle cellule il DNA è associato a proteine formando un complesso denominato **cromosoma**.

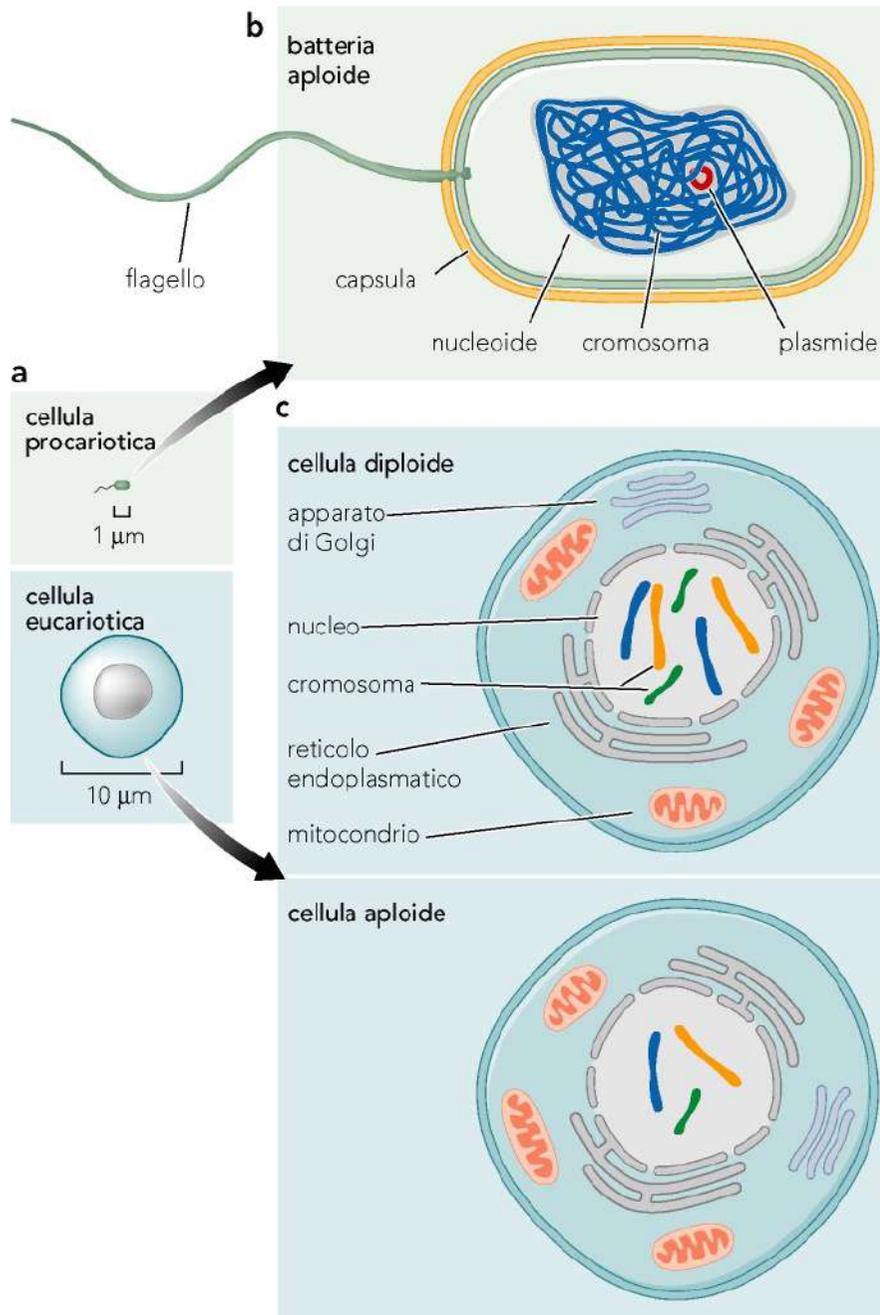
La struttura più compatta del cromosoma oltre a rendere più stabile il DNA ne favorisce la trasmissione più efficiente alle cellule figlie.

**Quoziente di impacchettamento:** rapporto tra lunghezza del DNA e dimensioni del compartimento che lo contiene

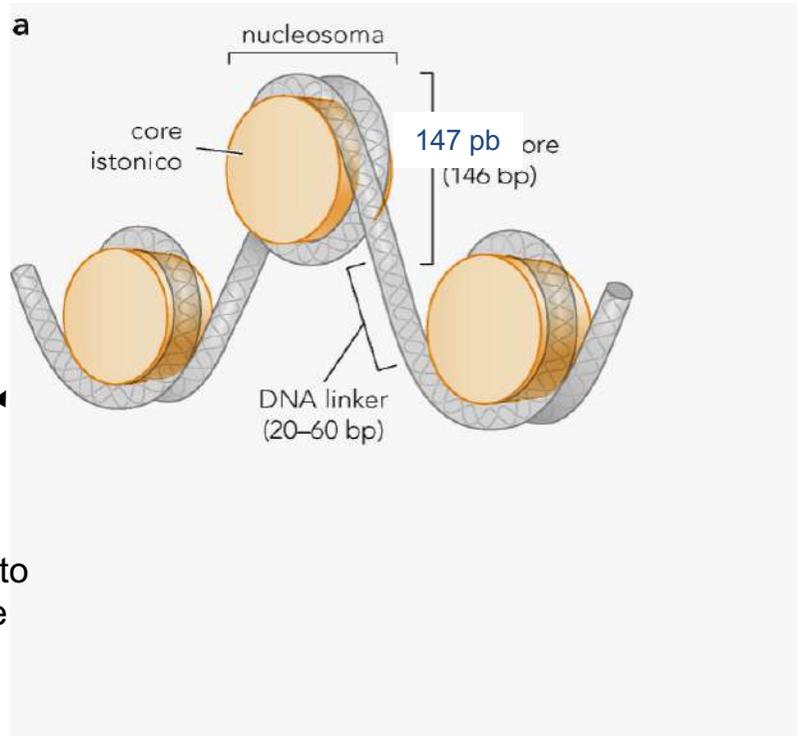
E.coli:  $1\text{mm}/1\mu\text{m}$       **1.000**

Uomo:  $1\text{m}/10\mu\text{m}$       **100.000**

(in realtà è **10.000** per suddivisione del genoma in 46 cromosomi da 4,4 cm )



# La cromatina



Avvolgimento  
di 1.65 volte

La **cromatina** è formata da DNA avvolto su gruppi di proteine dette **istoni**, formando il **nucleosoma**.

Il **nucleosoma** consiste di un nucleo proteico ottomero intorno al quale sono avvolte **147 pb** di DNA.

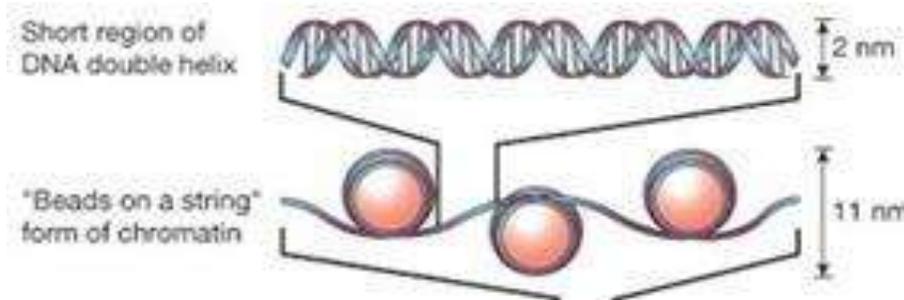
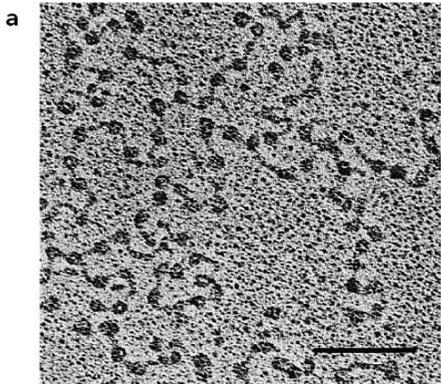
L'ottamero proteico consiste in un tetramero formato da 2 molecole di istone H3 e di istone H4, e da due coppie di dimeri H2A-H2B.

Ciascun nucleosoma è separato da un DNA di connessione detto **DNA linker** (20-60pb).

Ad ogni nucleosoma è associato l'istone **H1** detto "istone linker" che lega il DNA linker.

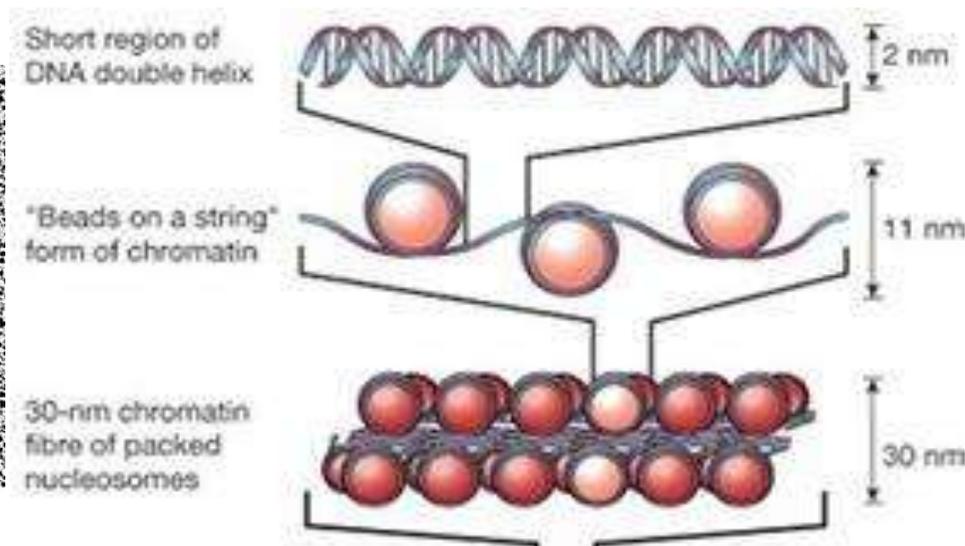
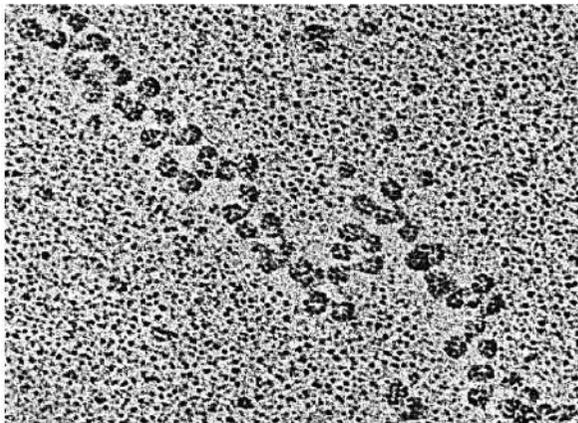
# Fibra da 10 nm e da 30 nm della cromatina

- L'avvolgimento del DNA intorno al nucleo istonico produce la **fibra da 10 nm (filo di perle)**



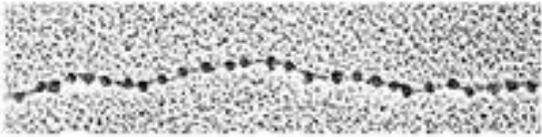
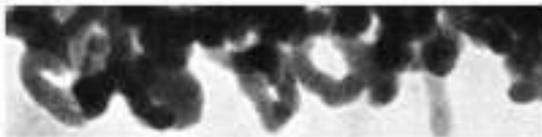
riduzione della lunghezza di 6 volte: 1 m  $\rightarrow$  16 cm

- Ulteriore grado di condensazione: **fibra da 30 nm**



ulteriore riduzione della lunghezza di 40 volte: 1 m  $\rightarrow$  2.5 cm

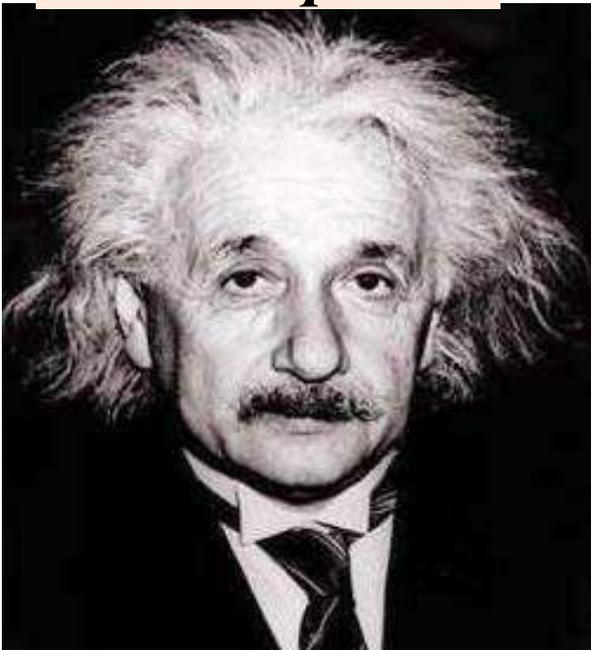
# Strutture di ordine superiore della cromatina

Struttura	Rap. Imp.	
Fibra da 10 nm "collana di perle"	6	
Fibra da 30 nm "solenioide"	40	
Anse cromosomiche da 100-400 nm	1000	
Cromosomi mitotici	10 000	

## Paradosso del valore K

La complessità di un organismo non è correlata al numero di cromosomi

*Homo sapiens*



**46 cromosomi**

*Lysandra atlantica*



**250 cromosomi**

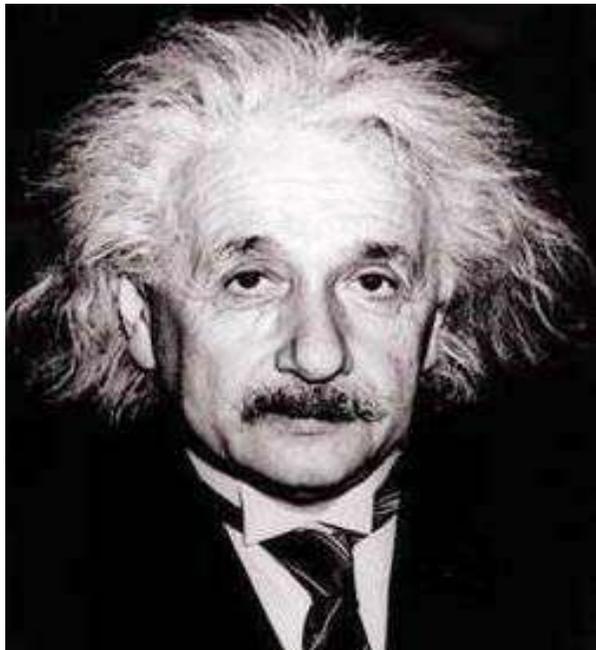
*Ophioglossum reticulatum*



**1260 cromosomi**

## Paradosso del valore C

La complessità di un organismo non è correlata alla grandezza del genoma



*Homo sapiens*  $3 \times 10^9$  pb



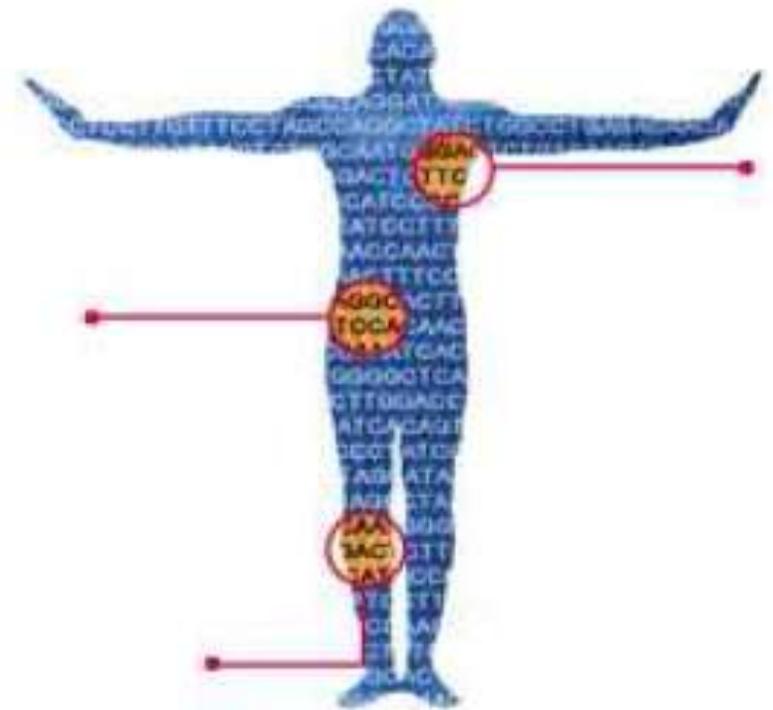
*Amoeba dubia*  $67 \times 10^{10}$  pb

# Il Progetto Genoma Umano

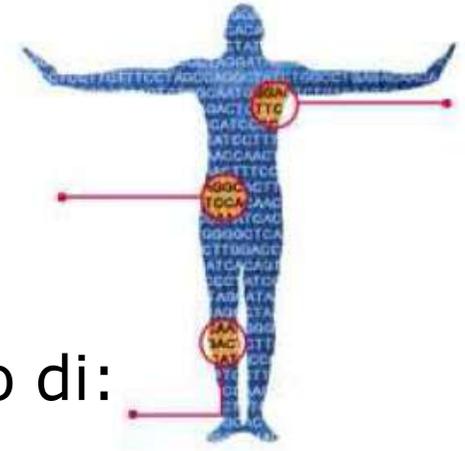
La determinazione e la conoscenza di tutta l'informazione racchiusa nel **genoma** sono la condizione necessaria per comprendere la completa biologia di un determinato organismo, cioè significherebbe comprendere il "segreto della vita".

Conoscere l'intera sequenza del genoma umano è come conoscere tutte le pagine del manuale necessario per costruire il corpo umano.

Nel 1986 il premio Nobel **Renato Dulbecco** e **Leroy Hood** lanciano l'idea di sequenziare l'intero genoma umano



# Il Progetto Genoma Umano



## *Perché?*

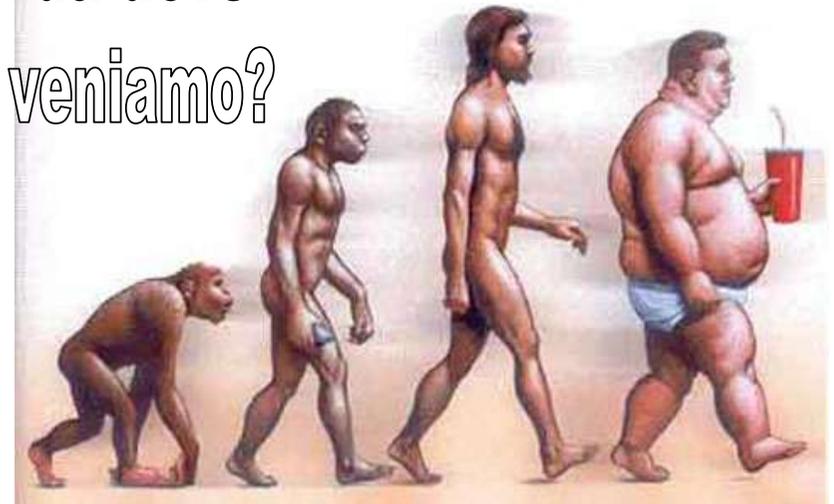
La sequenza del genoma umano avrebbe permesso di:

- identificare **tutti i geni umani**
- determinare **la struttura dei geni**
- identificare le **funzioni dei geni**
- identificare **i geni responsabili delle malattie mendeliane**
- determinare le **regioni non codificanti con funzione regolatoria**
- scoprire **l'inatteso**

# Perché studiamo il genoma umano ?



da dove veniamo?



# Le tappe principali del Progetto Genoma Umano

1990 Negli Stati Uniti nasce ufficialmente lo **Human Genome Project (HGP)**, sotto la guida di **J. Watson** (presso l'U.S. Department of Energy & National Institutes of Health).

Negli anni successivi Regno Unito, Francia, Germania, Giappone, Cina si uniscono al progetto formando un **consorzio pubblico internazionale**.

**Budget iniziale \$ 3.000.000.000; Durata prevista: 15 anni**

1992 **Craig Venter** lascia l'NIH e il progetto pubblico. Fonderà una compagnia privata, la **Celera Genomics**, portando avanti un progetto genoma parallelo

1999 (Dicembre) Pubblicata su **Nature** la sequenza completa del cromosoma 22

2000 (Maggio) Pubblicata su **Nature** la sequenza completa del cromosoma 21

# Le tappe principali del Progetto Genoma Umano

2000 (Giugno) **Francis Collins** (Direttore del National Human Genome Research Institute) e **Craig Venter** annunciano congiuntamente di aver completato la "bozza" (che gli inglesi chiamano **working draft**) del Genoma Umano

2001 (Febbraio) La bozza completa del genoma umano e le prime analisi sono pubblicate su **Nature** (quella del consorzio pubblico) e su **Science** (quella della Celera)



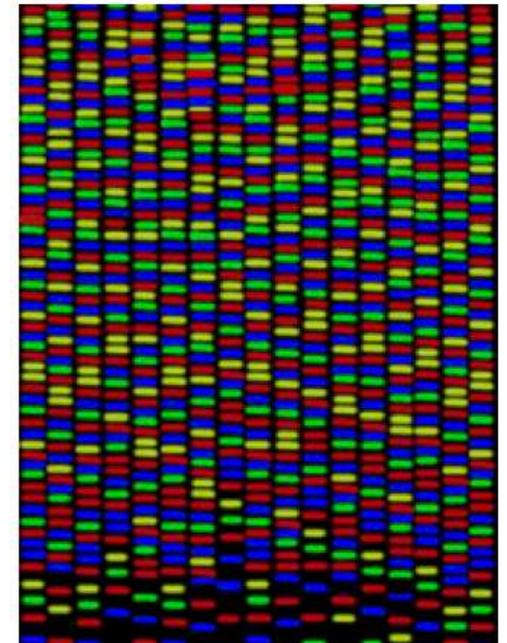
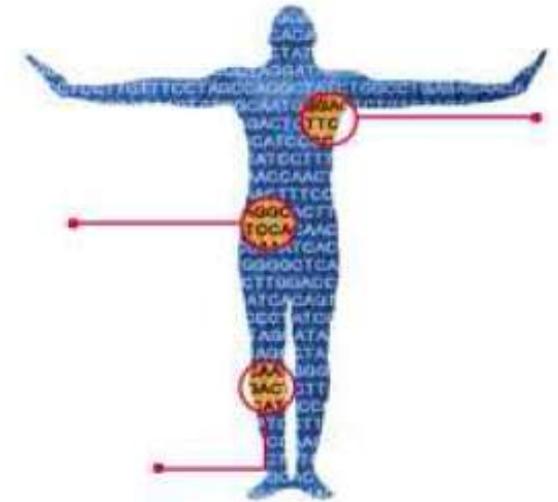
2003 (Aprile) Il progetto è dichiarato finito, con due anni di anticipo rispetto ai tempi stabiliti (sequenziato il 99% del genoma; accuratezza del 99.9%)

# Il Progetto Genoma Umano

*In che modo?*

Sequenziamento del DNA significa determinazione della sequenza lineare delle basi che lo compongono, quindi "leggere" l'ordine in cui A, T, C e G sono disposte lungo il DNA

Per il genoma umano significa determinare la sequenza di **3 miliardi di paia di basi**

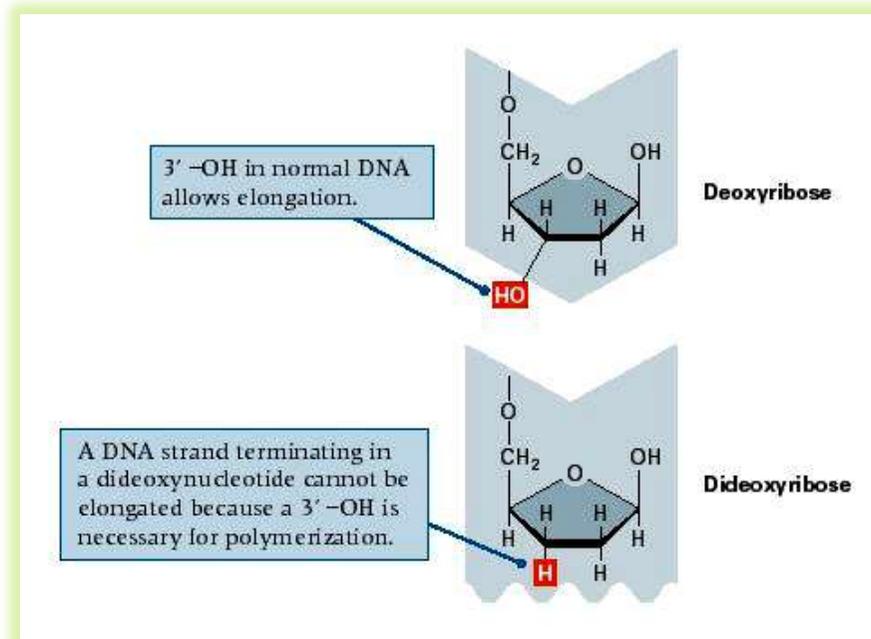


# Metodo di Sanger

Il sequenziamento del genoma umano fu realizzato mediante il **Metodo di Sanger**, che utilizza nucleotidi modificati (**dideossinucleotidi trifosfato**, ddNTPs) per interrompere la reazione di sintesi in posizioni specifiche.

I ddNTPs si differenziano dai dNTP per l'assenza del gruppo idrossilico sul carbonio 3' della molecola.

I ddNTP, a causa della loro struttura, impediscono che un altro nucleotide si leghi ad essi, in quanto non si possono formare legami fosfodiesterici.

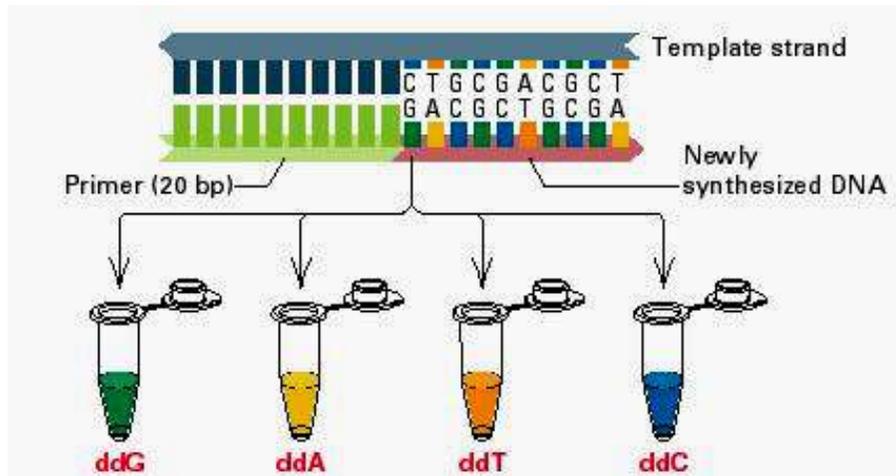


# Metodo di sequenziamento secondo Sanger

Il protocollo classico richiedeva le seguenti componenti:

- Uno stampo di **DNA** a singolo filamento
- un **primer** per iniziare la reazione di polimerizzazione (corto filamento che viene sintetizzato chimicamente e complementare al DNA stampo)
- **enzima** (DNA polimerasi di E.coli per copiare la sequenza di DNA)
- **Deossinucleotidi** (dATP, dCTP, dGTP, dTTP)
- **Dideossinucleotidi** (ddNTPs) marcati radioattivamente per visualizzare i frammenti di DNA sintetizzati

# Metodo di sequenziamento secondo Sanger

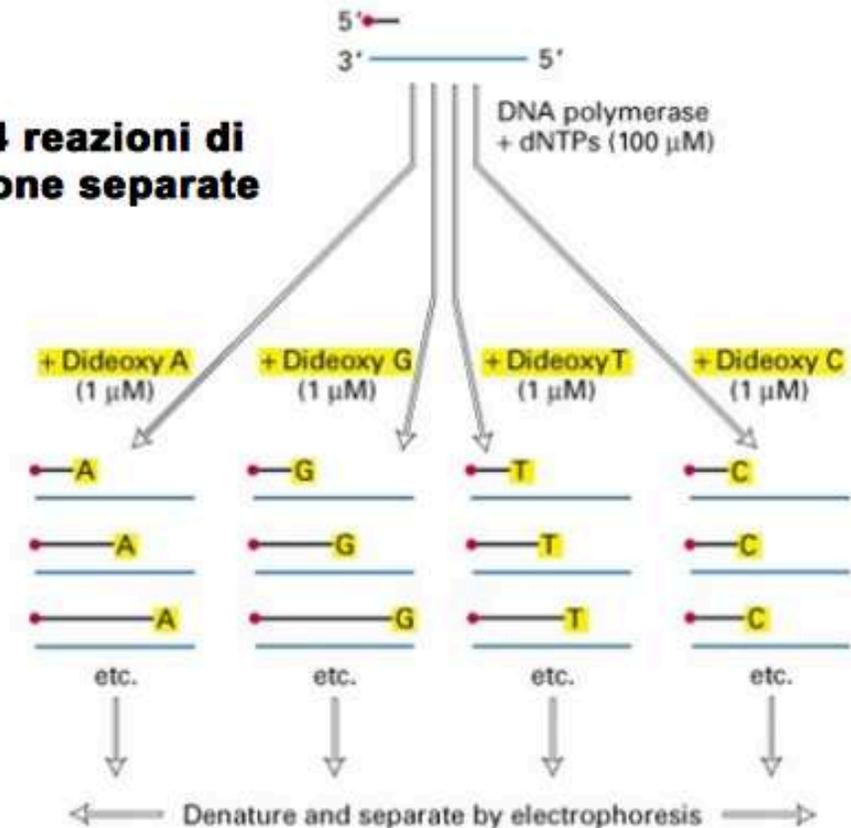


In parallelo 4 reazioni di polimerizzazione ciascuna delle quali contiene il primer, la DNA polimerasi, i 4 dNTP e un singolo ddNTP (100:1)

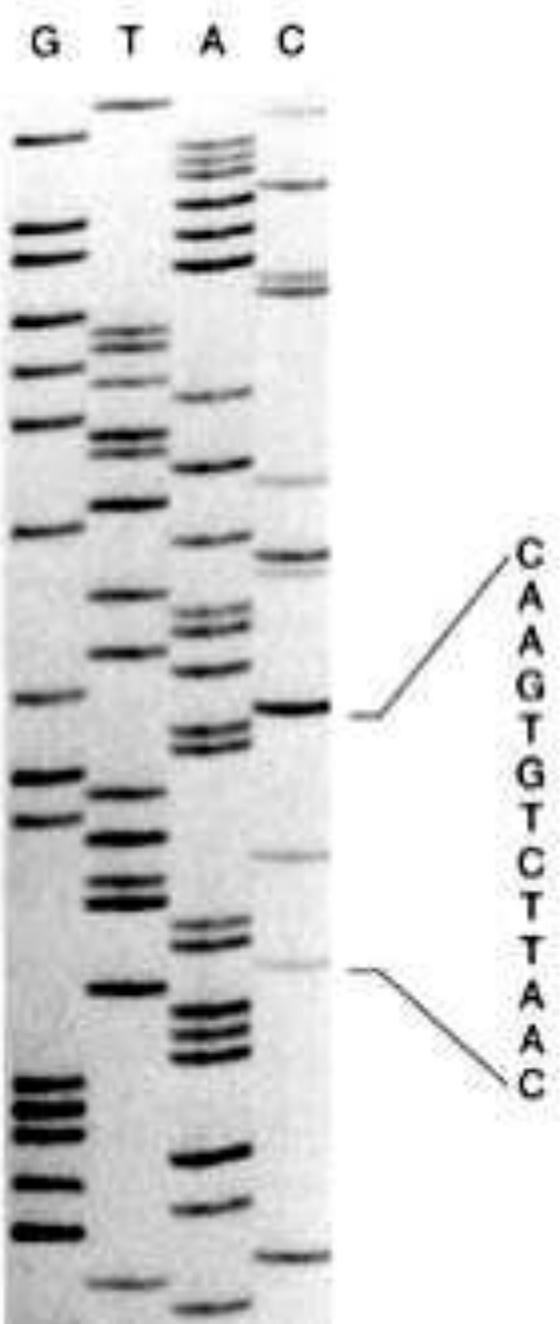
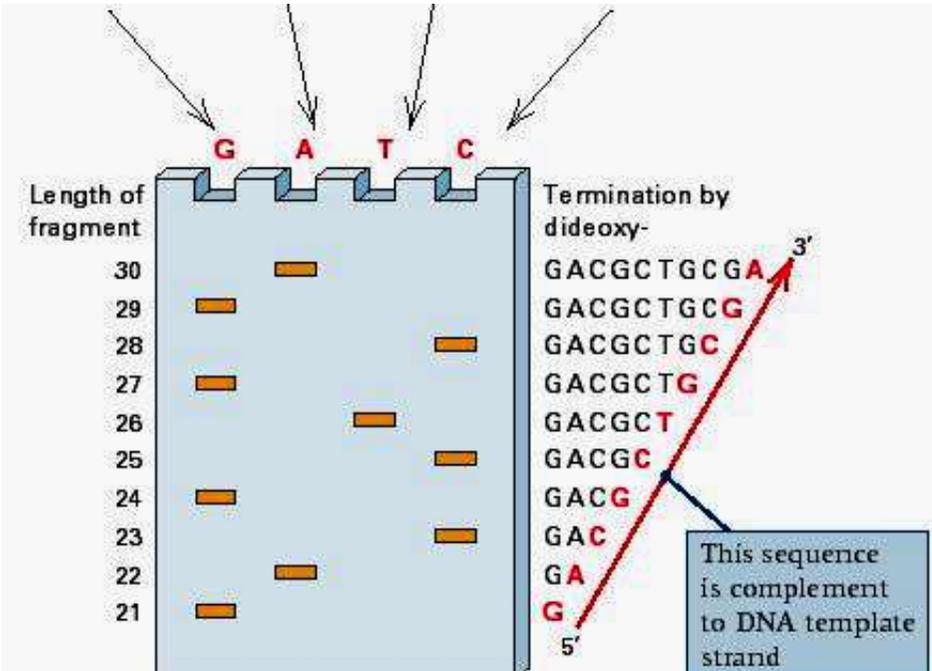
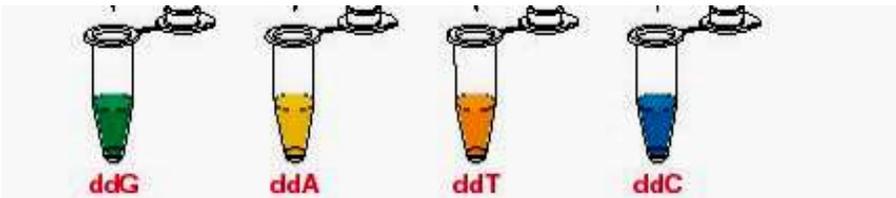
Ogni miscela conterrà una popolazione di filamenti che avranno in comune l'estremità 5' ma con lunghezze diverse perché la loro sintesi è stata bloccata in diverse posizioni al 3'



**Si eseguono 4 reazioni di polimerizzazione separate**



# Metodo di sequenziamento secondo Sanger



# Un passo avanti: l'avvento del sequenziatore automatico



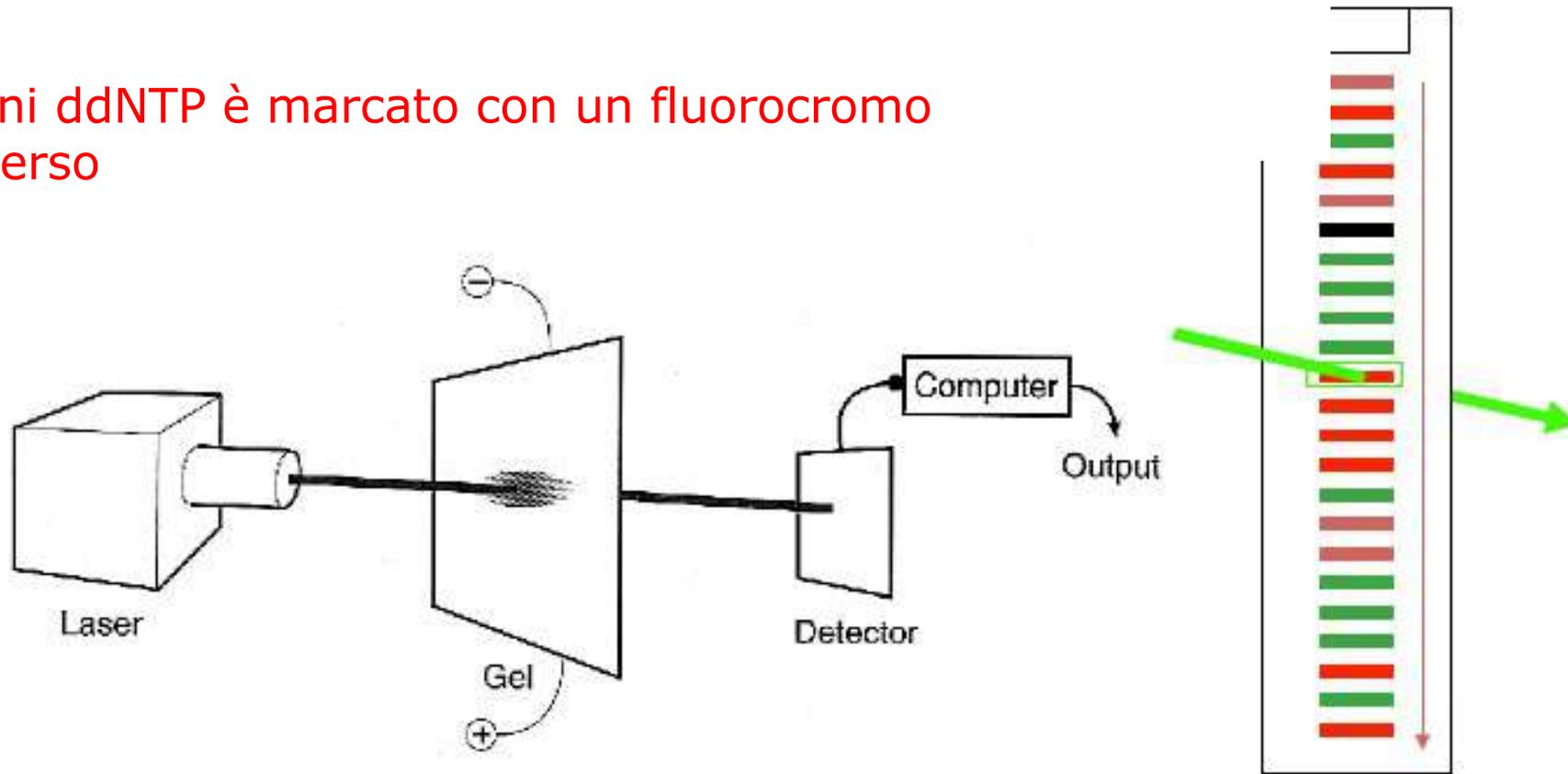
Ogni ddNTP è marcato con un diverso fluorocromo

# Sequenziamento automatico del DNA

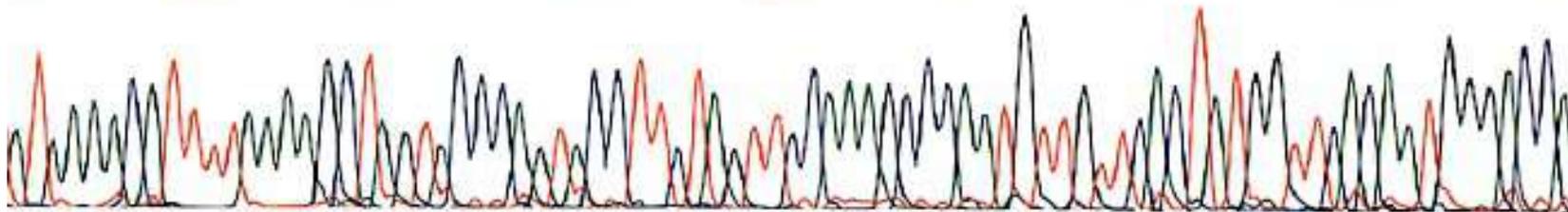
- tecnica non radioattiva
- il DNA stampo da sequenziare a doppio filamento
- la *Taq DNA polimerasi*
- una miscela dei 4 dNTP
- una miscela dei 4 ddNTP marcati
- si genera un'unica miscela di frammenti a singolo filamento che vengono poi caricati in un unico pozzetto e separati per elettroforesi
- durante l'elettroforesi i frammenti vengono eccitati da una luce laser, passano davanti a un rilevatore che capta la lunghezza d'onda e l'intensità delle emissioni fluorescenti e identifica quale nucleotide è presente nella banda
- le informazioni vengono integrate e trasformate in picchi di colore diverso (uno per ogni nucleotide) con aree proporzionali all'intensità di emissione

# Sequenziamento automatico del DNA

Ogni ddNTP è marcato con un fluorocromo diverso

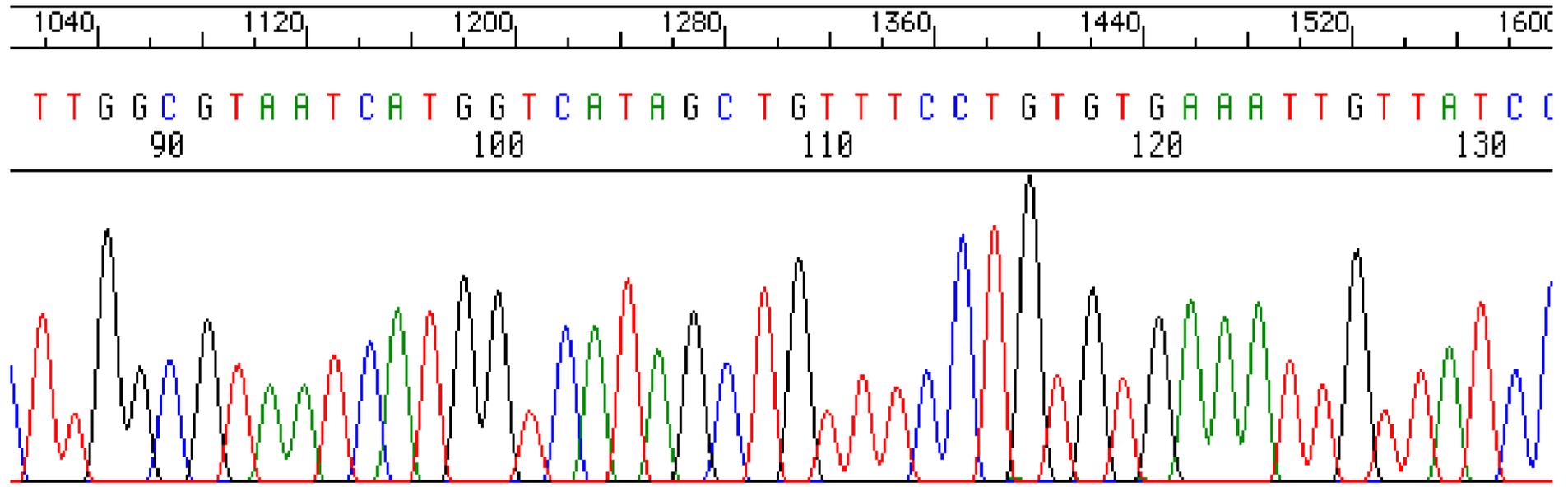


A T A A A A C A T T T T A A A A G C T A G T A C C C A G T A C C T T C T A G T T C C A A A G C C C A A T G T T G T T C A G T A T G G T T C A C A A T G G G A C C A  
0 150 160 170 180 190 200 210 220



# Sequenziamento automatico del DNA

Printout da un sequenziatore automatico

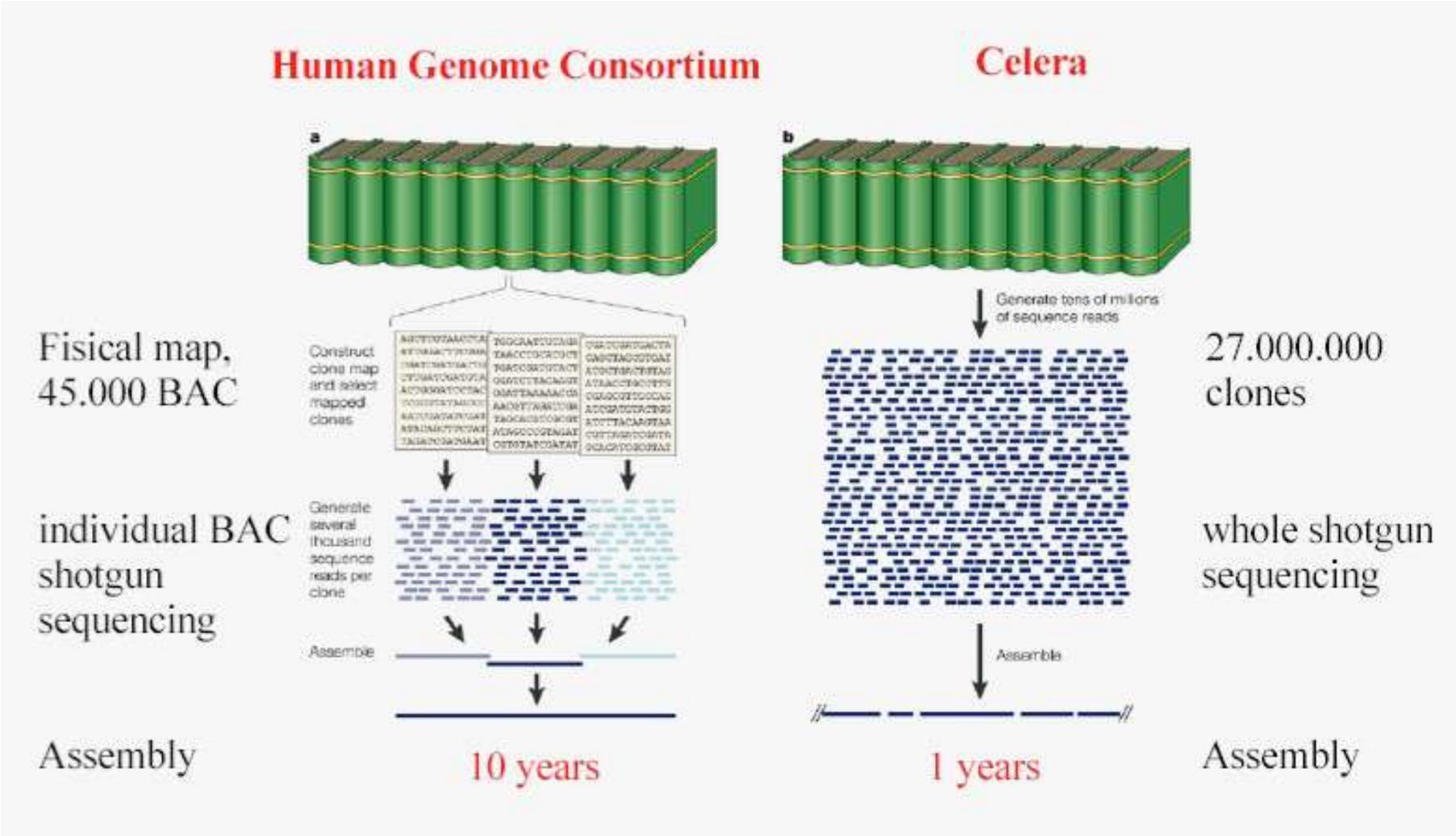


I colori sono generati dalla macchina (che riconosce i 4 fluorocromi associati ai 4 nucleotidi) e indicano le quattro basi: A verde, G nero, C blu, T rosso.

La realizzazione del progetto genoma umano fu possibile grazie ai progressi delle **tecniche del DNA ricombinante**, e allo sviluppo dei **sequenziatori automatici** capaci di sequenziare sino 400.000 basi/giorno.



# Strategie per il sequenziamento del genoma umano



# Approccio Utilizzato dalla Celera Genomics

## *Shotgun Sequencing*

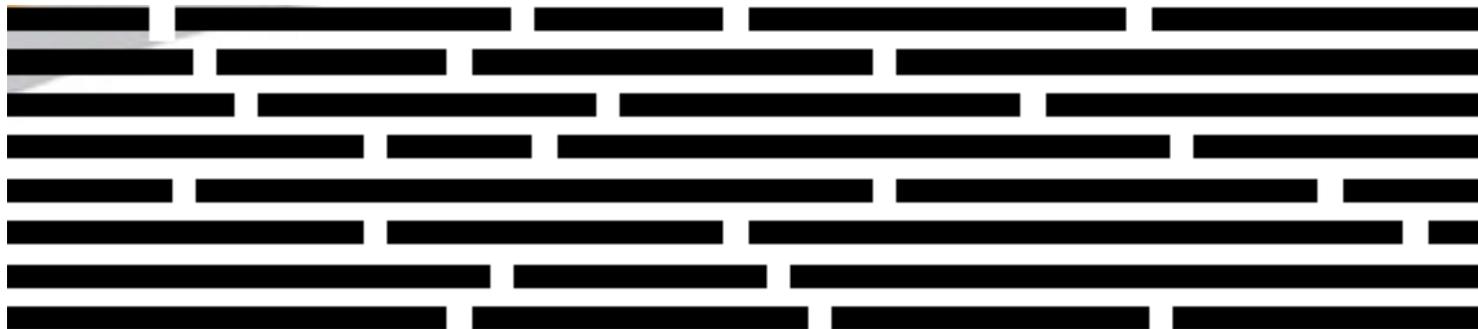
1. Frammentazione casuale del DNA umano in frammenti da 1000-10.000pb
2. Clonaggio dei frammenti in vettore plasmidico
3. Sequenziamento dei frammenti nei sequenziatori automatici
4. Assemblaggio delle sequenze

1. Frammentazione casuale del DNA umano (frammenti da 1000-10.000pb)

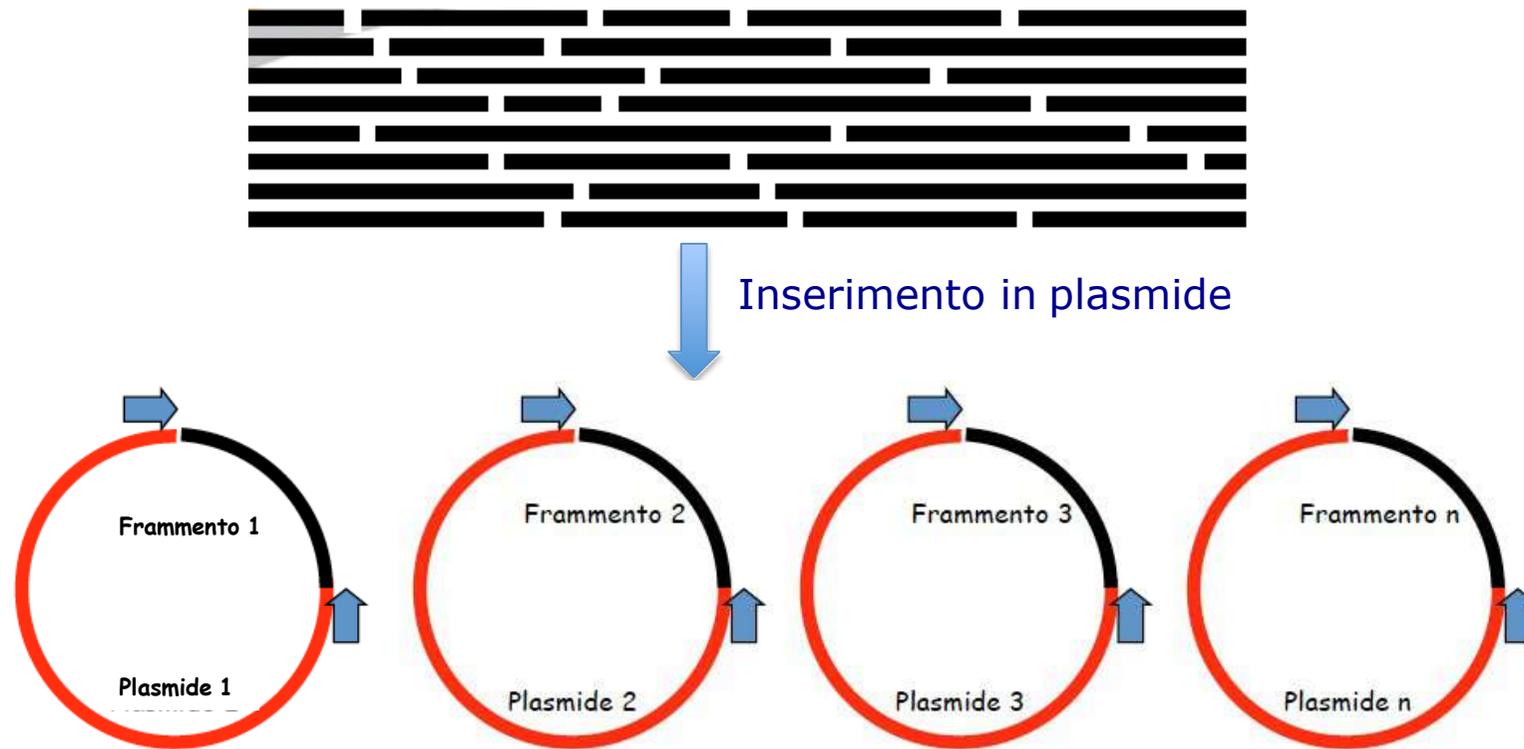
regione di DNA



Frammentazione



## 2. Clonaggio dei frammenti in vettore plasmidico



## 3. Sequenziamento dei frammenti nei sequenziatori automatici



**Milioni di sequenze parzialmente sovrapposte**

## 4. Assemblaggio delle sequenze

Le sequenze ottenute sono confrontate da specifici programmi bioinformatici

**A** TACCTGTGTGGAAGGAAGCAACCA CCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACAT  
AATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGA  
AAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATG

**B** ACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTTATAAACTTGAT  
ATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTG  
TCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGT

**C** GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTG  
GCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGSTTTAAGAATAGTTTTTGTGCTGACTTT  
CTATAGTGAATA

**D** ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT  
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA  
CCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTG

**E** CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT  
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG  
AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC  
TTGGGTTCTTGGGAGCAGCAGGA

**F** GGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA  
AAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGG  
GAGAAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTG

**Frammenti che contengono sequenze sovrapposte sono riuniti in un unico frammento**

**D+ A** { ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT  
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA  
CCTCTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT  
GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGGAAAAATGA  
CATGGTAGAACAGATG

**B** ACTGCTCTTTCAATATCAGCACAAAGCATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGAT  
ATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTG  
TCCAAAGGTATCCTTTGAGCCAATCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGT

**C** GAAAAGAATGAACAAGAAATTATFGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTG  
GCTGTGGTATATAAAATTAATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTTGCTGTACTTT  
CTATAGTGAATA

**E** CATCAAATATTACAGGGCTGCTATTAACAACAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT  
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG  
AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTC  
TTGGGTTCTTGGGAGCAGCAGGA

**F** GGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA  
AAATTAACCCACTCTGTGTTAGTTTAAAGTGCACTGATTTGAAGAATGATACTAATAACCAATAGTAGTAGCGG  
GAGAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAAGCATAAGAGGTAAGGTG

D+ A

ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT  
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA  
CCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT  
GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAAATGA  
CATGGTAGAACAGATG

C

GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGCCAAGTTTGTGGAATTGGTTTAAACATAACAAATTG  
GCTGTGGTATATAAAATTTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTTGTCTGACTTT  
CTATAGTGAATA

E

CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT  
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG  
AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC

F+B

TTGGGTTCTTGGGAGCAGCAGGA  
GGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA  
AAATTAACCCCACTCTGTGTTAGTTTAAAGTGCACCTGATTTGAAGAATGATACTAATAACCAATAGTAGTAGCGG  
GAGAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGC  
AGAAAGAATATGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACA  
AGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGC  
CCCGGCTGGTTTTGCGATTCTAAAATGT

D+A+F+B

ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT  
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA  
CCTACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT  
GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAAACATGTGGAAAAATGA  
CATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCC  
CACTCTGTGTTAGTTTAAAGTGCACTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGAGAATGATA  
ATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATA  
TGCATTTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACA  
CCTCAGTCATTACACAGGCCTGTCCAAGGTATCCTTTGAGCCAATCCCATACATTATTGTGCCCCGGCTGGT  
TTTGCGATTCTAAAATGT

C

GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTG  
GCTGTGGTATATAAAATPATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTTGTCTGTACTTT  
CTATAGTGAATA

E

CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT  
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATGAACCATTAGG  
AGTAGCACCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC  
TTGGGTTCTTGGGAGCAGCAGGA

La sequenza può essere così ricostruita e conoscendo dei marcatori cromosomici è possibile **mapparla sui cromosomi**

# Dopo il sequenziamento del genoma umano

## L'avvento dell'era **Genomica**

Gli enormi sviluppi delle tecniche di indagine bio-molecolare hanno reso possibile acquisire le informazioni relative ai genomi di un numero elevato di organismi e al flusso dell'informazione genetica a ritmi fino a poco fa inimmaginabili.

Questo ha prodotto un drastico cambio di prospettiva e di orizzonti nella ricerca biologica:

Dal **GENE** al **GENOMA**

Dal **TRASCritto** al **TRASCrittOMA**

Dalla **PROTEINA** al **PROTEOMA**

Dal **METABOLITA** al **METABOLOMA**

# La Rivoluzione “OMICA” in Biologia

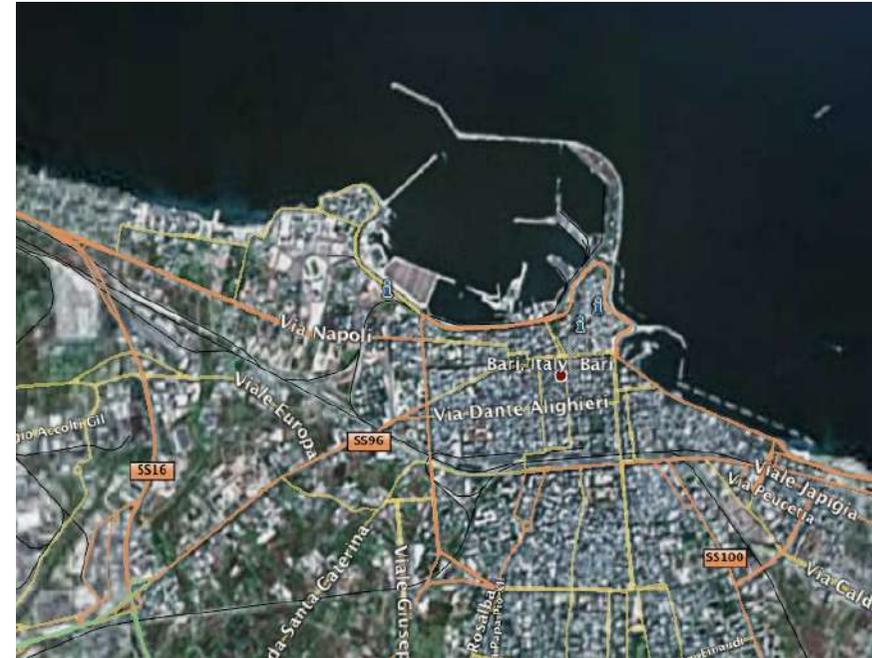
- GENOMICA  
Studio dei genomi
- TRASCRITTOMICA  
analisi del trascrittoma
- PROTEOMICA  
analisi del proteoma
- METABOLOMICA  
analisi dei profili metabolici
- FARMACOGENOMICA  
relazione tra genoma e risposta ai farmaci
- FISIOLOGICA  
Fisiologia dell'intero organismo

Richiede la costituzione di team di ricercatori con competenze multidisciplinari (Biologia Molecolare, Biochimica, Informatica, Matematica, Fisica, Statistica, Chimica, Ingegneria)

# La Genomica

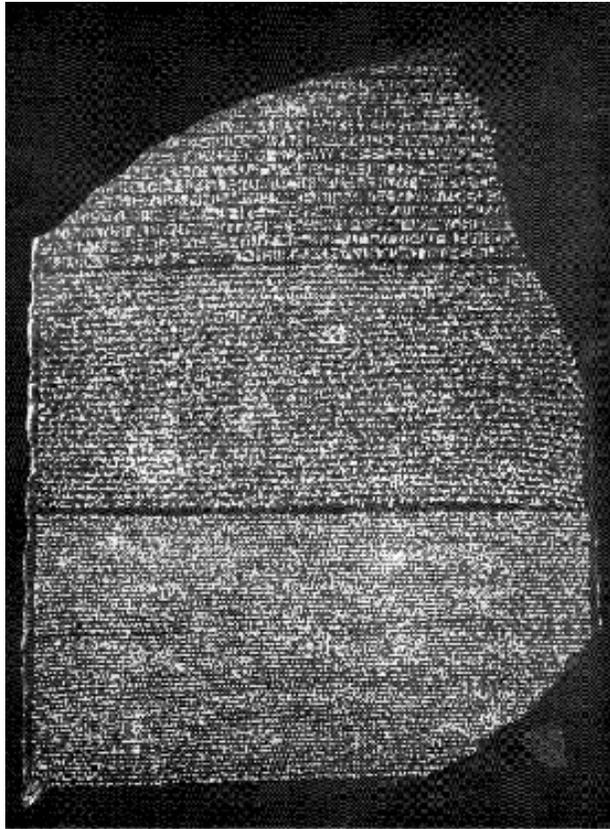


**Genomica Strutturale.** Studio della struttura del genoma, identificazione di geni e dei loro prodotti di espressione, di elementi regolatori ed altre entità informative.



**Genomica Funzionale.** Studio delle funzioni dei geni, delle loro interazioni (pathways metabolici) e dei meccanismi che ne regolano l'espressione.

# La Genomica Comparata



Stele di Rosetta  
(British Museum)

- Non si può capire come è fatto il genoma umano e come funzionano i nostri geni se non si confrontano con quelli dei principali organismi modello
- Il confronto di entità "omologhe" ci aiuta ad interpretare l'informazione genica.

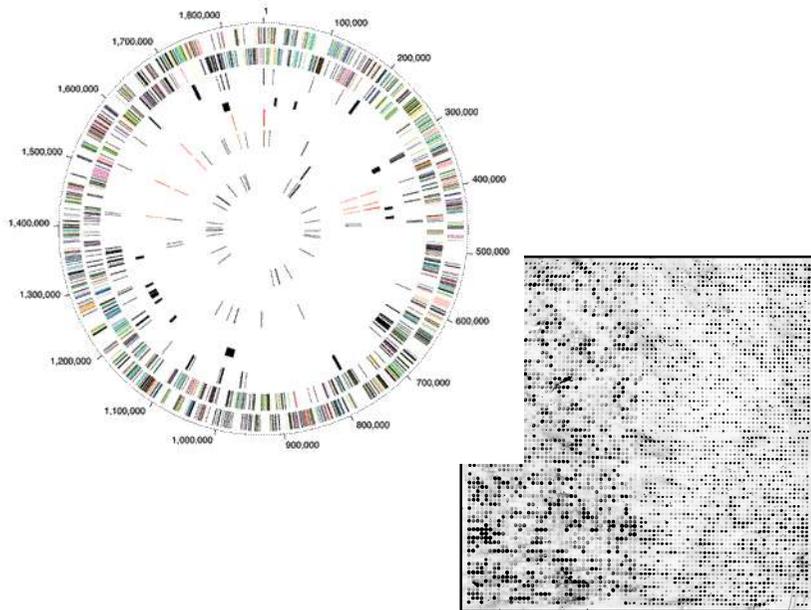
# Genomi completi già sequenziati

	PROCARIOTI	EUCARIOTI
Ottobre 2011	1783	37
Marzo 2013	2600	185
Marzo 2014	2921	221
Marzo 2015	3628	332
Febbraio 2016	5788	349
Febbraio 2017	6968	659

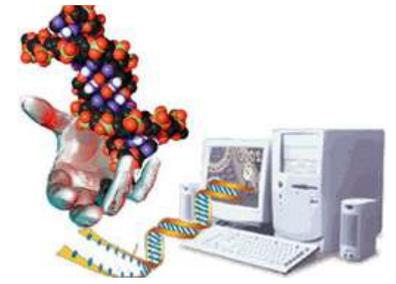
# La Bioinformatica

La necessità di gestire ed interpretare le grandi quantità di informazioni derivanti dal sequenziamento del genoma umano ha richiesto lo sviluppo di adeguati strumenti informatici per la gestione e l'analisi dei dati.

Si è sviluppata così la **Bioinformatica**, una disciplina che si pone l'obiettivo di sviluppare e applicare strumenti adeguati per l'immagazzinamento, l'interrogazione e l'analisi dei dati biologici.



# La Bioinformatica



La ricerca nel settore della **Bioinformatica** si occupa quindi di sviluppare gli approcci più appropriati per:

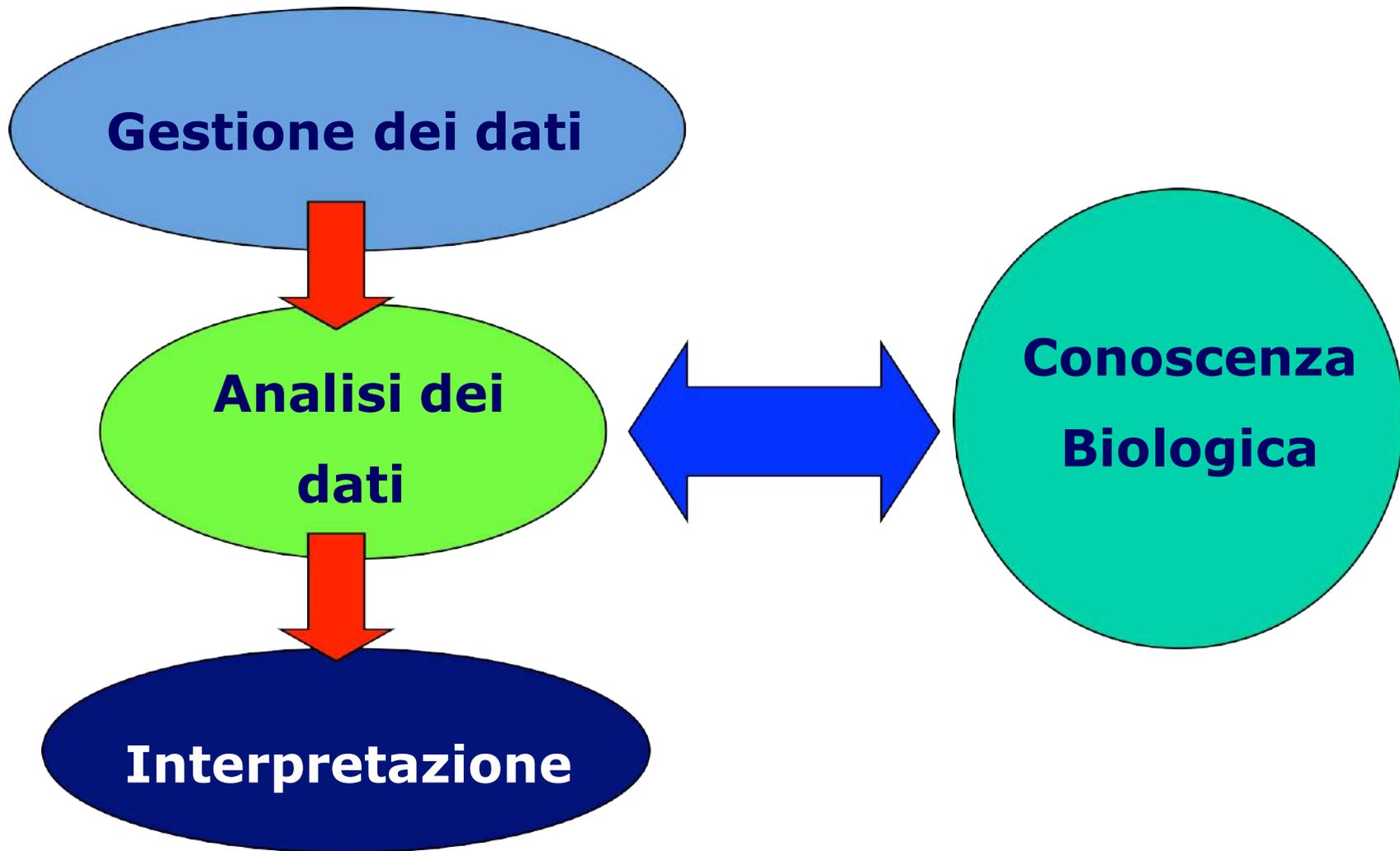
- collezionare i dati biologici in apposite **banche dati**,
- sviluppare **strumenti idonei** per l'interrogazione dei dati biologici;
- sviluppare **metodologie** per l'analisi dei dati biologici.



comprensione dei meccanismi molecolari alla base della vita

La Bioinformatica è per sua natura **multidisciplinare**, richiedendo competenze diverse (Biologia Molecolare, Informatica, Matematica, Statistica, Fisica, Ingegneria, ecc.) ed ha applicazioni pratiche in vasti settori della scienza.

# Il ruolo della Bioinformatica



# Immagazzinamento dei dati di sequenziamento

## Banche Dati Biologiche

In base alla tipo di informazioni in esse contenute, le banche dati biologiche sono distinte in 2 classi principali:

- **Banche dati PRIMARIE**: dette anche collettori primari. Il loro ruolo è quello di raccogliere, giornalmente, tutte le informazioni che riguardano biomolecole prodotte in tutti i laboratori del mondo e renderle disponibili.  
Sono: l'**EMBL / GenBank/ DDBJ**
- **Banche dati SECONDARIE o SPECIALIZZATE**: raccolgono insiemi di dati omogenei dal punto di vista tassonomico e/o funzionale disponibili nelle Banche dati Primarie e/o in Letteratura, rivisti e annotati con informazioni di valore aggiunto (es. **banche dati genomiche**)

# Analisi dei dati di sequenziamento

## Annotazione

Le sequenze genomiche assemblate sono analizzate con opportuni strumenti bioinformatici per arrivare alla identificazione dei geni e degli elementi funzionali, processo definito **Annotazione**.

Data la sequenza completa del genoma, l'**annotazione** permette di rispondere a queste domande :

- Quanti geni contiene?
- Dove sono localizzati i geni?
- A cosa serve ciascun gene (ovvero, qual è la funzione della proteina codificata)?
- Quale è la struttura dei geni?
- quali sono le regioni regolatorie?
- ecc

L'annotazione usa procedure diverse che dipendono dal tipo di elemento funzionale in esame, e integra annotazioni effettuate con diverse metodiche.

# I "browser" genomici

I dati derivanti dal sequenziamento e dall'annotazione dei genomi costituiscono una descrizione completa, strutturale e funzionale, del genoma dei diversi organismi.

Questi dati sono di grande aiuto alla ricerca scientifica, e, per facilitare l'accesso, sono stati riuniti in **collezioni diverse, messe a disposizione della comunità scientifica come risorse accessibili via web.**

Queste collezioni sono chiamate **'browser' genomici**, strumenti che permettono ai ricercatori di "navigare" all'interno dei genomi, visualizzando tutte le annotazioni che sono disponibili.

Sono interfacce web collegate alle banche dati contenenti le sequenze prodotte dai vari progetti di sequenziamento genomico e le relative annotazioni e analisi funzionali.

I browser genomici principali sono:

**Ensembl** (sviluppato da EMBL-EBI e dal Sanger Center) disponibile al sito <http://www.ensembl.org>

**UCSC** (University of California Santa Cruz) - disponibile al sito <http://genome.ucsc.edu>

# Ensembl

si sceglie il genoma d'interesse

Search:  for

e.g. **BRCA2** or rat 5:62797383-63627669 or coronary heart dise

gene o regione cromosomica da visualizzare

### Popular genomes



**Human**  
GRCh38.p2



**Mouse**  
GRCm38.p3



**Zebrafish**  
Zv9

★ [Log in to customize this list](#)

### All genomes

-- Select a species -- ▾

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

ENCODE data in Ensembl

Ve!P

Gene expression in different tissues

Find SNPs and other variants for my gene

Retrieve gene sequence

```
GCTGACTTCGGGTGGT  
GGGCTTGTGGCGGAGC  
GGGCTCTGCTGGCCT  
AGGGACAGATTCTGA  
CACCTCTGGAGCGGT
```

Compare genes across species

Use my own data in Ensembl

Learn about a disease or phenotype

### What's New in Ensembl Release 79

- 015)
  - to Ensembl-Havana GENCODE gene (release 22)
  - q-to-Ensembl model comparison
  - [Scrollable Region in Detail](#)
- [Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

### Latest blog posts

- 30 Apr 2015: [Ensembl website maintenance May 7th](#)
- 16 Apr 2015: [What's coming in Ensembl release 80](#)
- 09 Apr 2015: [dbSNP 142 and 1000 Genomes Phase 3](#)

[Go to Ensembl blog](#) →

### Tweets

[Follow](#)

**e!Ensembl**  
@ensembl

3h

Genomes of different mouse strains annotated with Ensembl genes from #BioMart  
#CitedEnsembl @BMC\_series buff.ly/1H5xzbE

# Ensembl

Current selection:  
< all Species  
Only searching Human

Restrict category to:

- Gene 450
- Transcript 1200
- Somatic Mutation 11029
- GeneTree 19
- GenomicAlignment 1
- ProbeFeature 704
- Clones & Regions 1
- Protein Domain 1
- Protein Family 9
- Variant 36168

Per page:  
10 25 50 100

Layout:  
Standard Table

Tip:  
You can choose which results appear near the top of your search by updating your favourite species.

Only searching Human    
49582 results match TP53 when restricted to species: Human X

**TP53 (Human Gene)**  
ENSG00000141510 17:7661779-7687550:-1  
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]  
LRG\_321 (LRG display in Ensembl record; description: Locus Reference Genomic record for TP53) is an external reference matched to Gene ENSG00000141510  
Variant table • Phenotypes • Location • External Refs. • Regulation • Orthologues • Gene tree

**TP53-020 (Human Transcript)**  
ENST00000604348 17:7675162-7687487:-1  
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].  
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

**TP53-029 (Human Transcript)**  
ENST00000635293 17:7665416-7687491:-1  
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].  
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

**TP53-017 (Human Transcript)**  
ENST00000576024 17:7669569-7673587:-1  
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].  
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

**TP53-024 (Human Transcript)**  
ENST00000618944 17:7668402-7675493:-1  
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]  
LRG\_321t5 (LRG display in Ensembl record; description: Locus Reference Genomic record for TP53) is an external reference matched to Transcript ENST00000618944  
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

**TP53-023 (Human Transcript)**  
ENST00000619186 17:7668402-7675493:-1  
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]  
LRG\_321t5 (LRG display in Ensembl record; description: Locus Reference Genomic record for TP53) is an external reference matched to Transcript ENST00000619186  
Location • External Refs. • cDNA seq. • Exons • Variant table • Protein seq. • Population • Protein summary

**TP53-028 (Human Transcript)**  
ENST00000619485 17:7668421-7687487:-1  
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]  
R-HSA-6811555 (Reactome record; description: PI5P Regulates TP53 Acetylation) is an external reference matched to Translation ENSP00000482537

**Best gene match**

Human Gene

## TP53

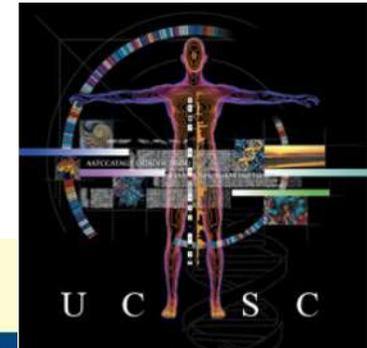
HGNC Symbol; Acc:HGNC:11998



Protein coding gene  
tumor protein p53

**Suggestions**

atp5a atp5b atp5c atp5d atp5e atp5h atp5i atp5k atp5l  
atp5o atp5s tnp03 tee3 tpa tpap tpi tpid tpij tpij tpo  
tpor tpox tpte cfap53 irsp53 tpi1p3 wrap53 dulp5 fatp5 setp5  
tnip3 tnni3 tnp01 tnp2 top2a top2b top3a top3b tor3a torc3  
tpar1 tpam tpbj tpkl tpra1 tpte2 traf3 traj3 tra3 trav3



## UCSC Genome Bioinformatics

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)[Genome Browser](#)[Blat](#)[Table Browser](#)[Gene Sorter](#)[In Silico PCR](#)[Genome Graphs](#)[Galaxy](#)[VizGene](#)[Utilities](#)[Downloads](#)[Release Log](#)[Custom Tracks](#)[Cancer Browser](#)[Microbial](#)

### About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to [ENCODE](#) data at UCSC (2003 to 2012) and to the [Neandertal](#) project. Download or purchase the Genome Browser source code, or the Genome Browser in a Box (GBiB) at our [online store](#).

We encourage you to explore these sequence assemblies and scrolls over chromosomes, showing the work of annotation, gene prediction, homology and other information on groups of genes that can be compared to the genome. The [Table Browser](#) provides convenient access to a large collection of *in situ* mouse and frog images to browse through a large collection of genome-wide data sets. [Genes](#) allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the UCSC Genomics Informatics Group, a cross-departmental team within the [UC Santa Cruz Genomics Institute](#) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our [public mailing list](#).

The Genome Browser project team relies on public funding to support our work. Donations are welcome -- we have many more ideas than our funding supports! If you have ideas, drop a comment in our [suggestion box](#).

[DONATE NOW](#)

Cliccando uno dei due link si accede ai genomi

[News](#)[News Archives](#)

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list. Please see our [blog](#) for posts about Genome Browser tools, features, projects and more.

## Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).  
Software Copyright (c) The Regents of the University of California. All rights reserved.

group	genome	assembly	position	search term
<input type="text" value="Mammal"/>	<input type="text" value="Human"/>	<input type="text" value="Feb. 2009 (GRCh37/hg19)"/>	<input type="text" value="chr17:7,571,720-590,868"/>	<input type="text" value="enter position, gene symbol or search terms"/>

[Click here to reset](#) the browser user interface settings to their defaults.

## Human Genome Browser – hg19 assembly ([sequences](#))

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

### Sample position queries

Scelgo  
il gruppo

Scelgo  
la specie

“versione”

regione  
cromosomica o  
gene da visualizzare

VIA!

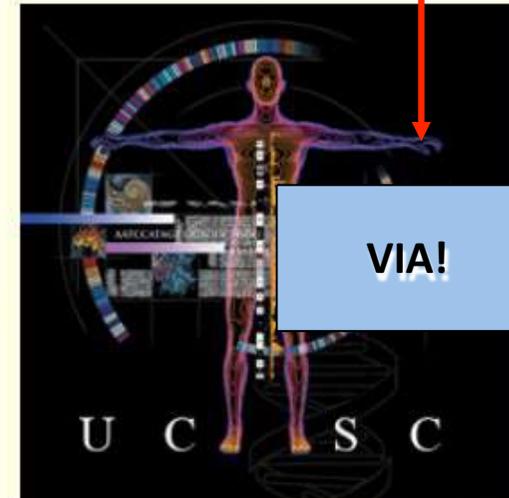
genome. See the [User's Guide](#) for more information.

### Request:

chr7  
chrUn\_gl000212  
20p13

### Genome Browser Response:

Displays all of chromosome 7  
Displays all of the unplaced contig gl000212  
Displays region for band p13 on chr 20



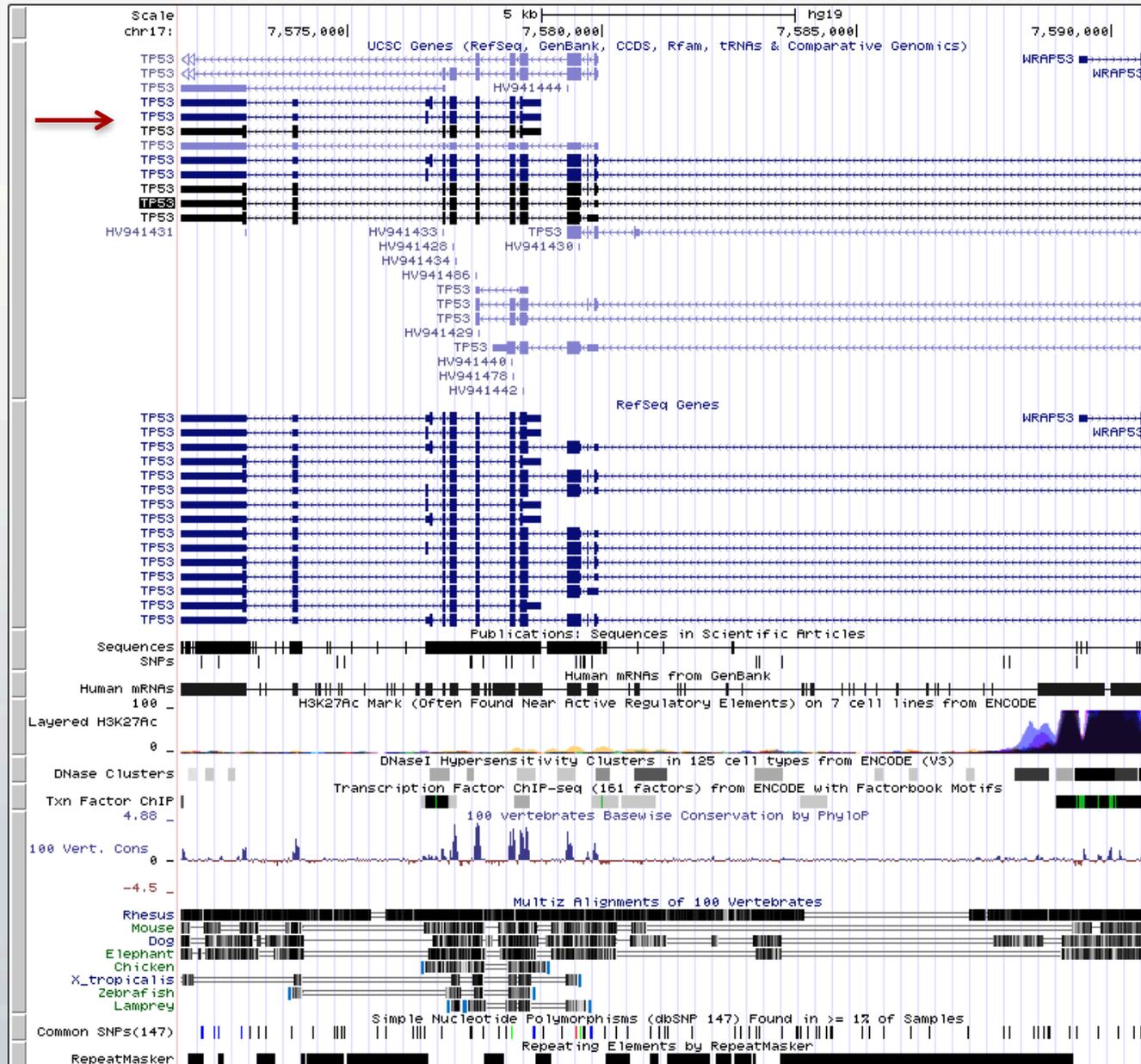
*Homo sapiens*  
(Graphic courtesy of [CBSE](#))

# UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

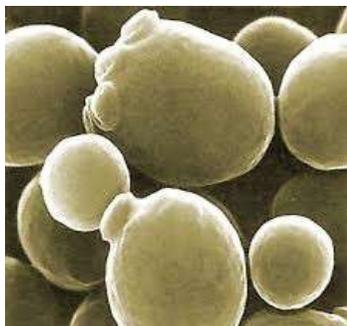
chr17:7,571,720-7,590,868 19,149 bp. TP53 (Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA.) go

chr17 (p13.1) | 3.1 | 17p12 | 17p11.2 | q11.2 | 17q12 | 17q22 | 24.3 | 25.1 | q25.3

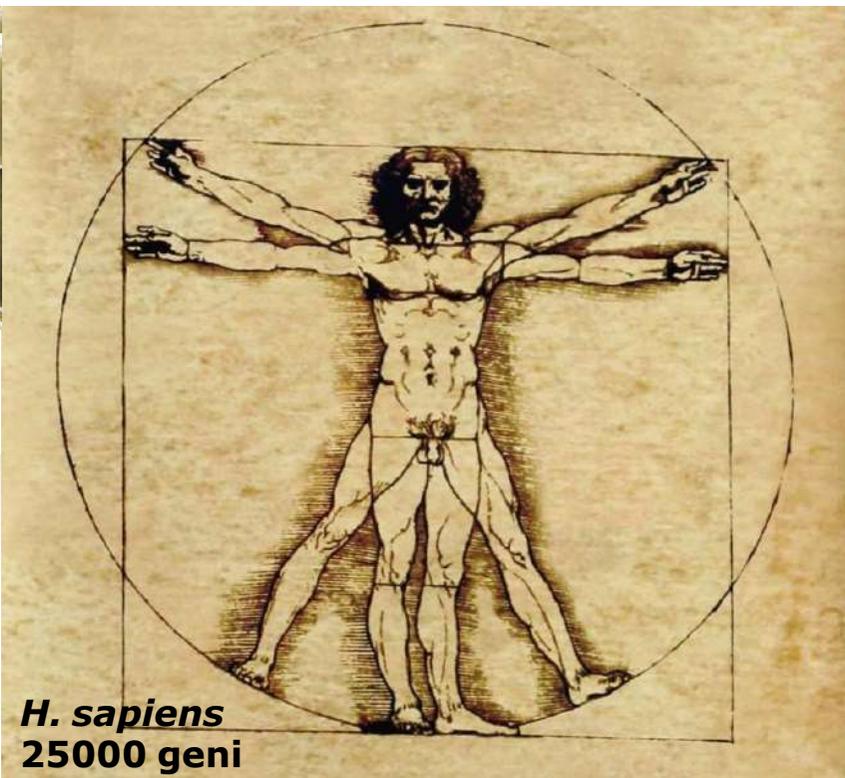


**Quali informazioni ha fornito il sequenziamento del genoma umano?**

# Il numero dei geni umani codificanti per proteine è di gran lunga inferiore rispetto a quanto atteso!!



***S. cerevisiae***  
6000 geni



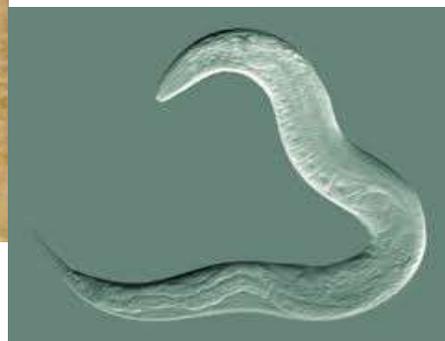
***H. sapiens***  
25000 geni



***M. musculus***  
22000 geni



***D. melanogaster***  
13000 geni



***C. elegans***  
21000 geni



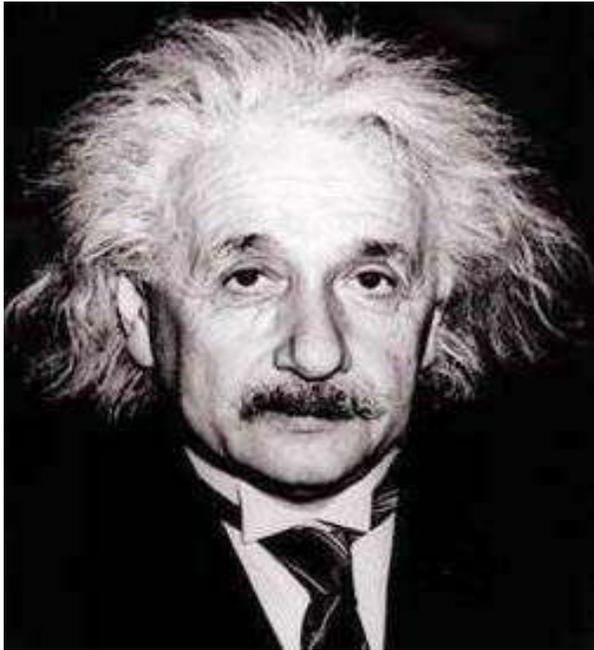
***D. rerio***  
26000 geni



***Z. mays***  
26000 geni

## Quindi:

La complessità dell'organismo umano non è correlata al numero dei geni



~25,000 genes



~22000 geni



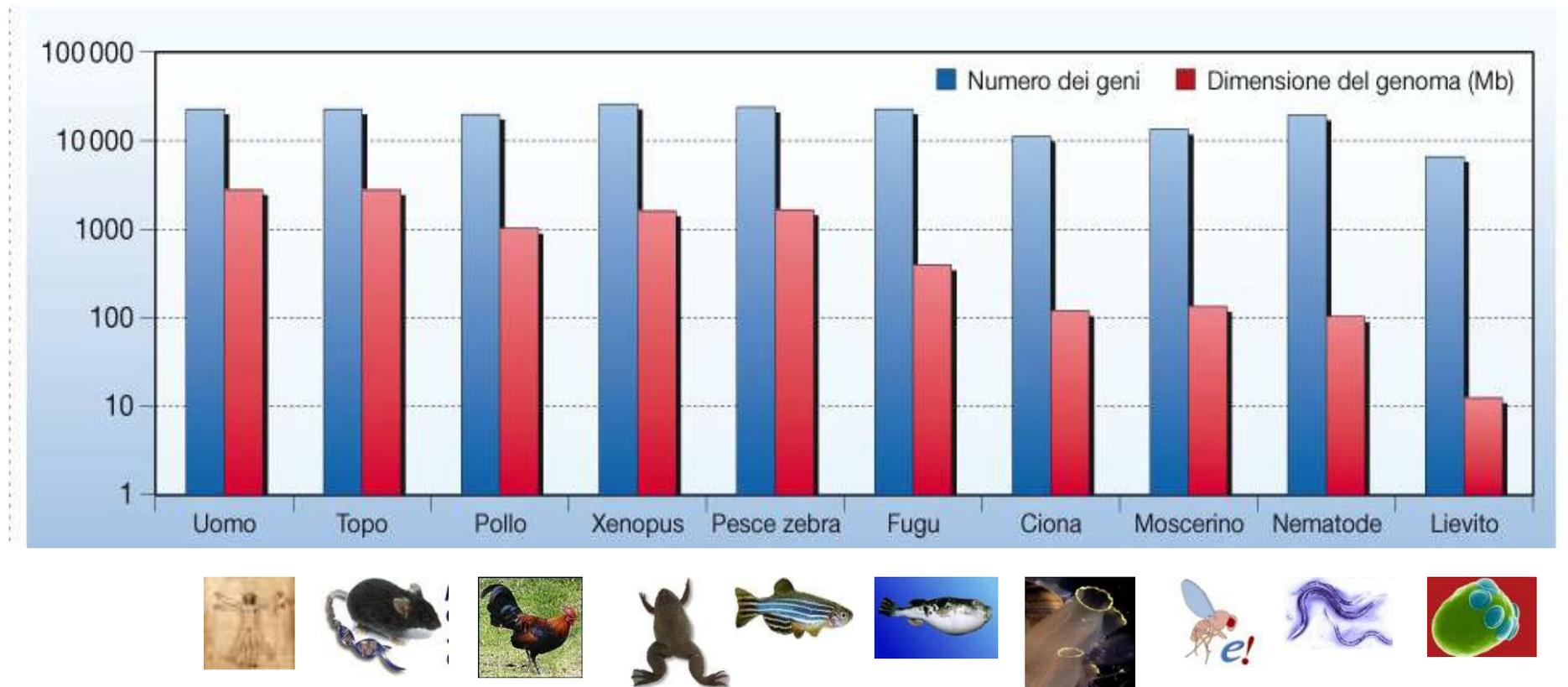
~13000 geni



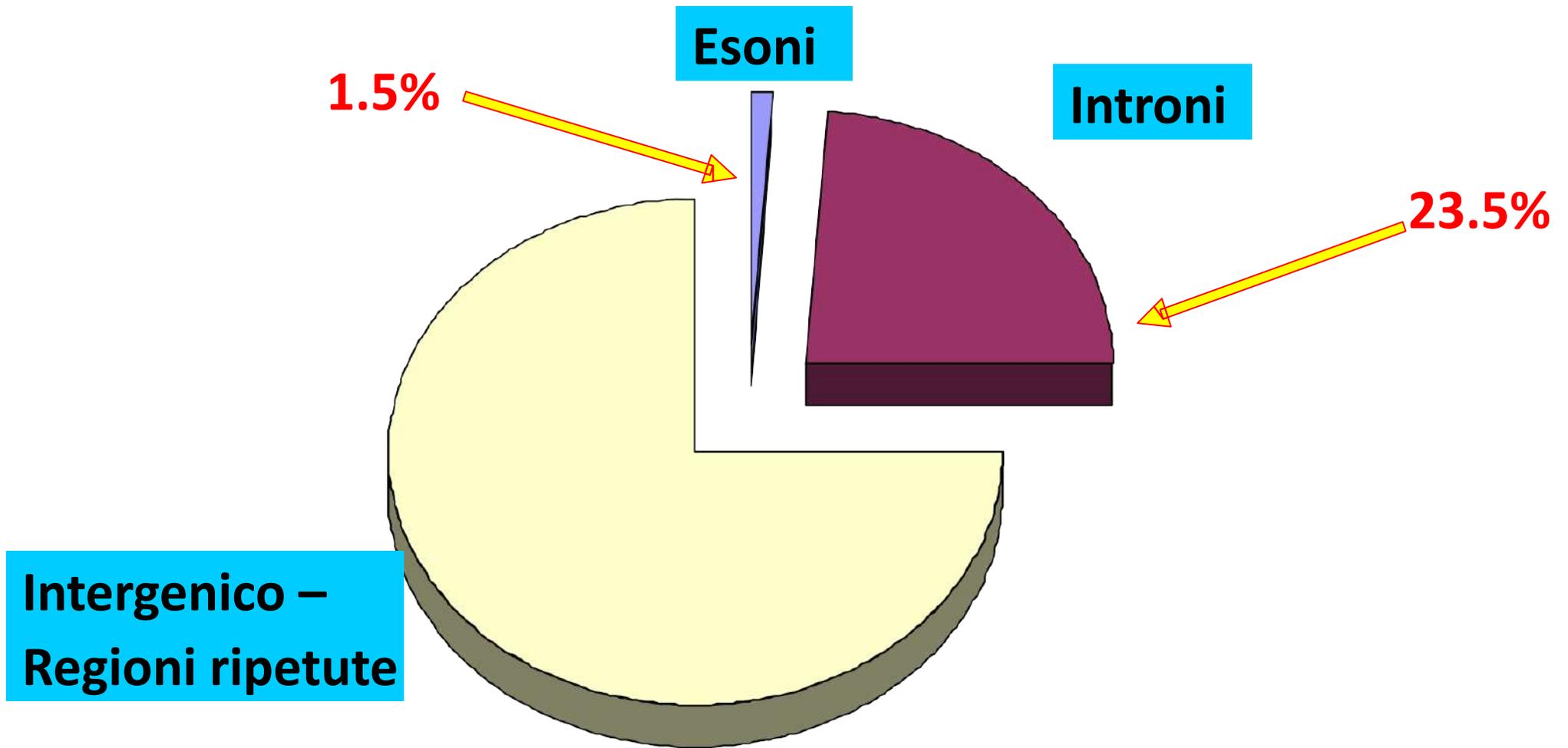
~26000 geni

# Il numero di geni non è correlato alla dimensione del genoma

Il genoma umano è 250 volte più grande del genoma di lievito (13Mbasi), ma il numero di geni codificanti per proteine è dello stesso ordine di grandezza ( 25000 nell'uomo e 6000 nel lievito)



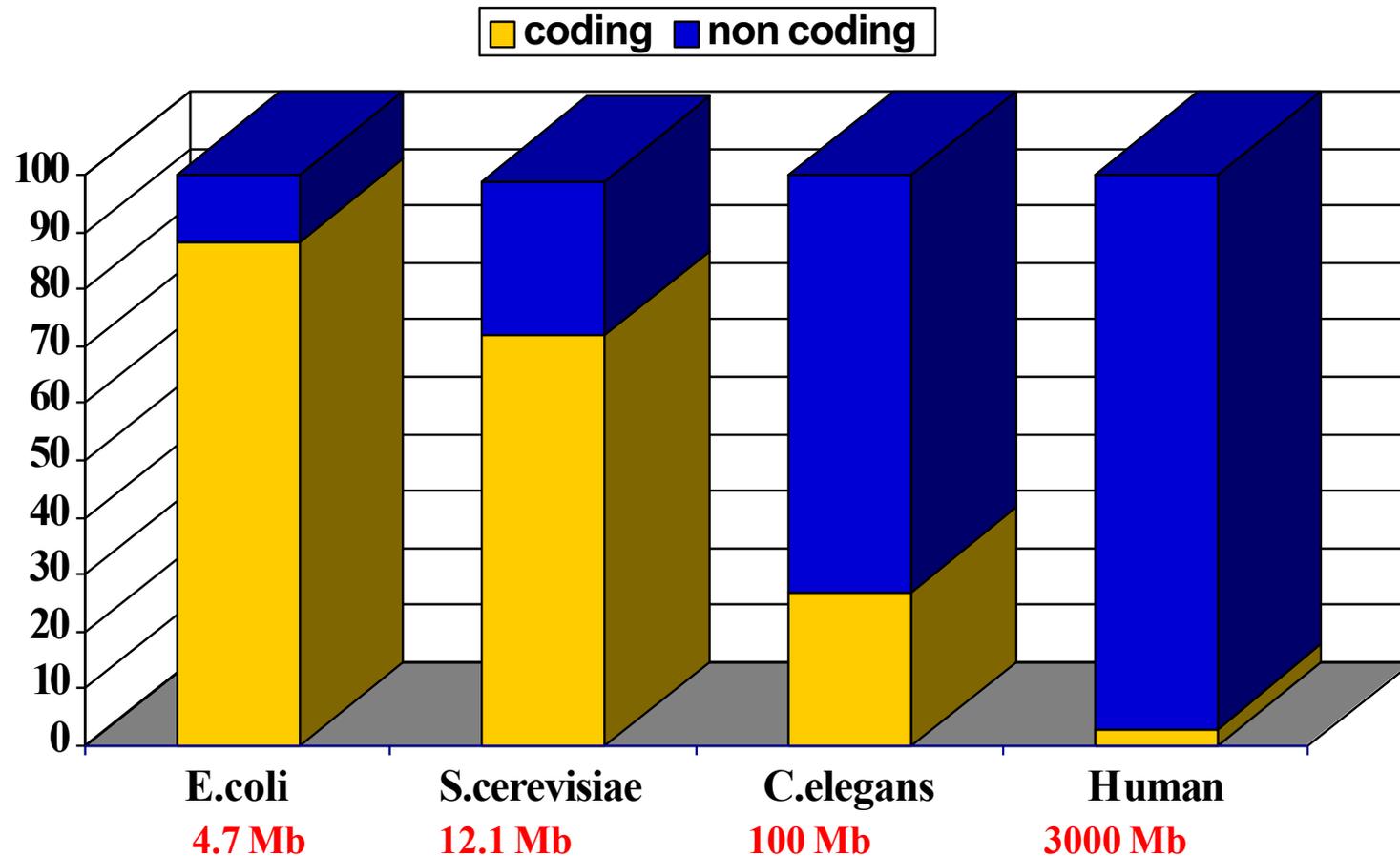
# Il Contenuto del genoma umano



**Il genoma è vuoto??**

# La porzione non codificante del genoma umano è >98%

## La sfida post-genomica



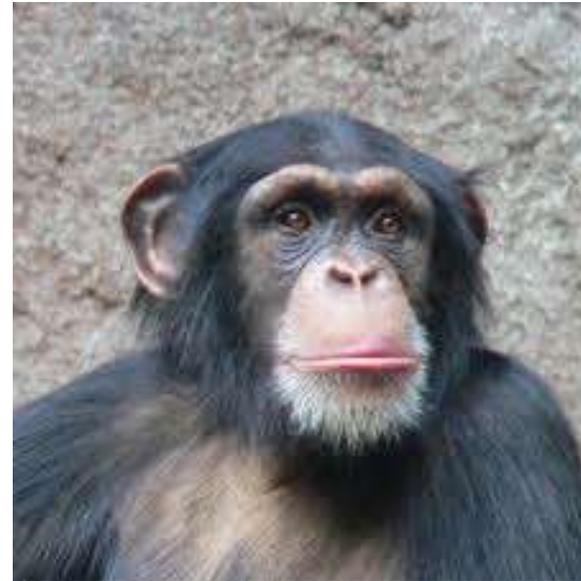
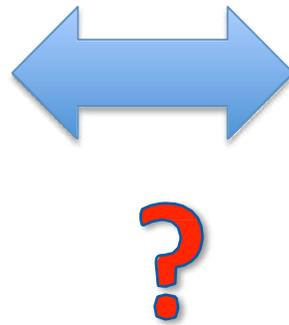
L'annotazione funzionale delle porzioni non-codificanti del genoma è una delle sfide principali dell'era post-genomica.

# I genomi dell'uomo e dello scimpanzé sono uguali all'incirca al 99%



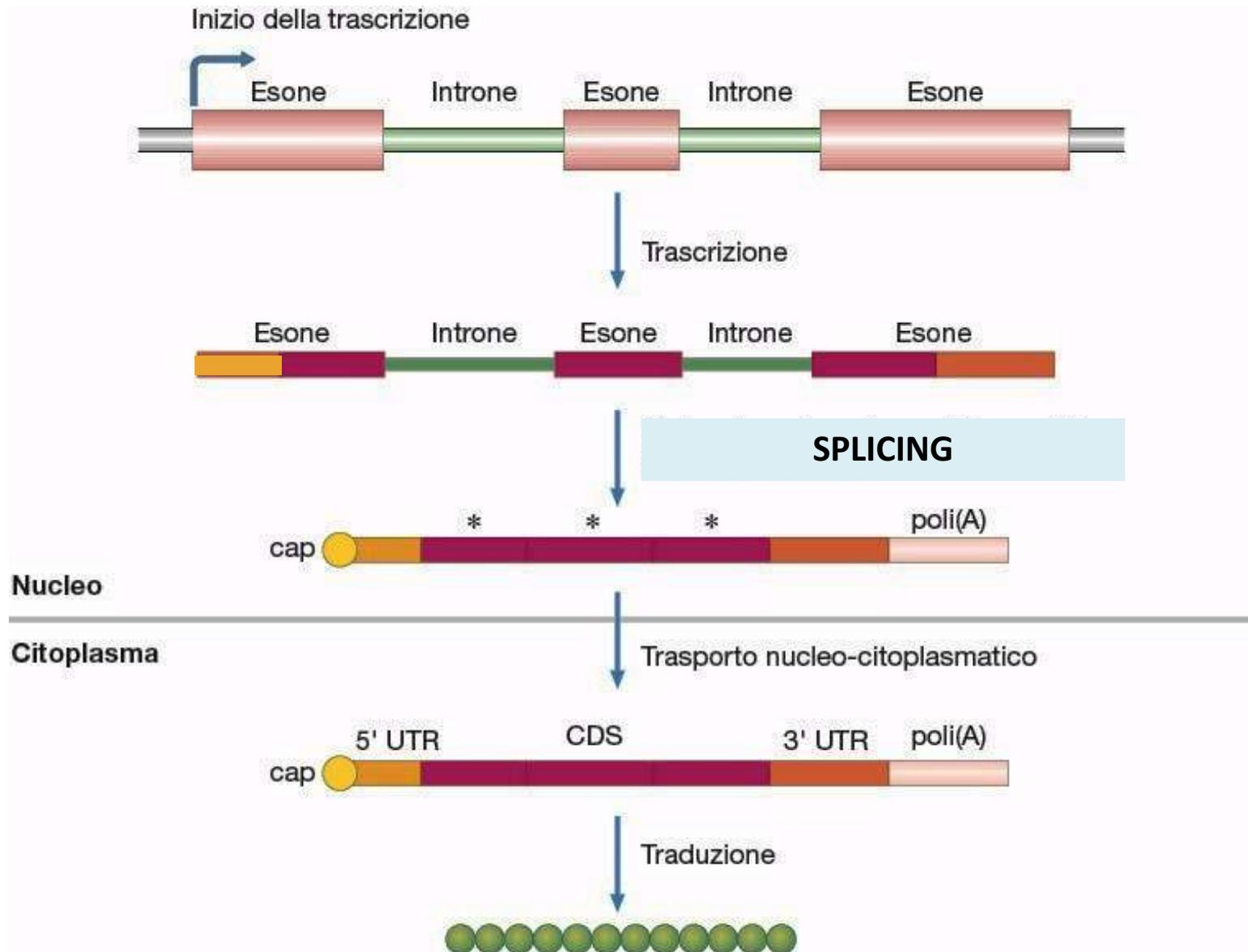
- 87% dei geni umani ha un omologo nello scimpanzé
- 30% delle proteine omologhe sono identiche

## Da cosa dipende la differenza?



E' evidente che geni simili e le interazioni fra gruppi di geni simili possono generare "programmi" molto diversi.

# I geni umani codificanti per proteine sono discontinui



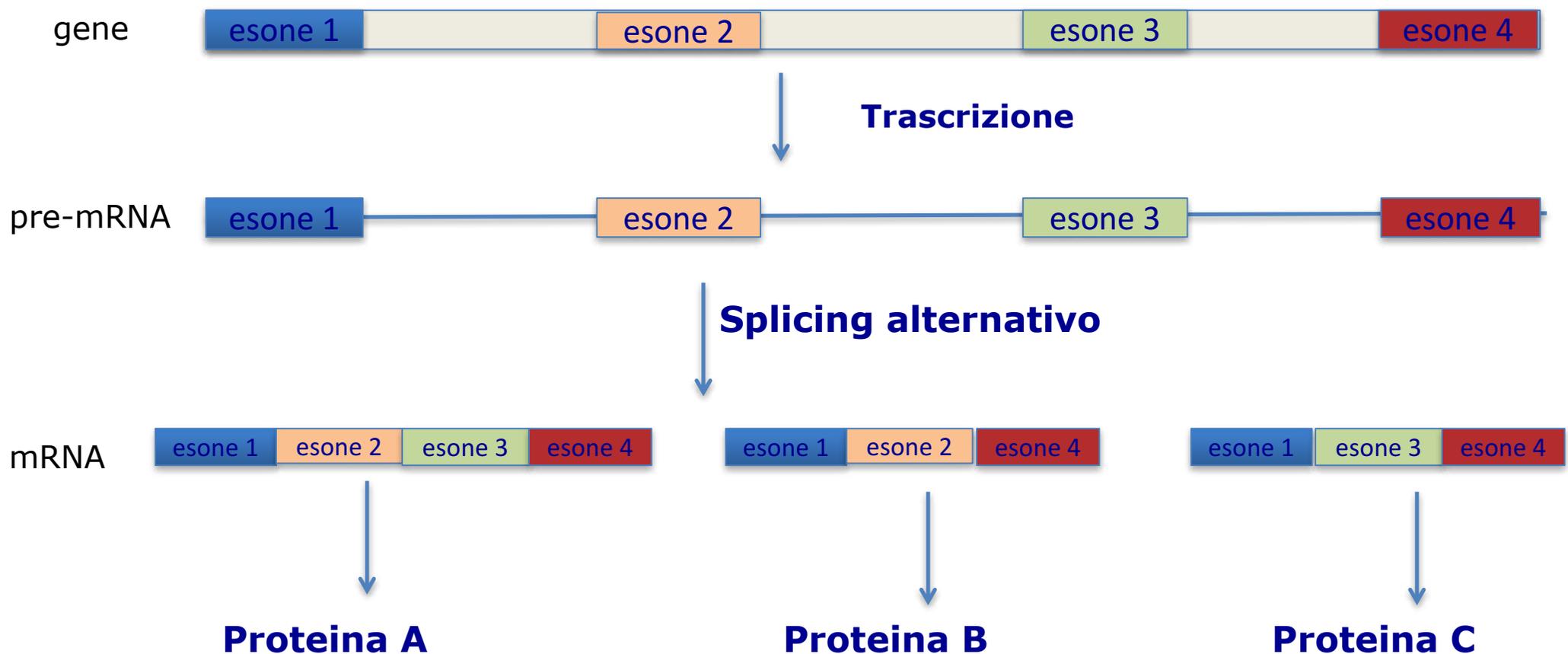
I geni sono espressi sotto forma di lunghi **RNA precursori** che nel nucleo sono sottoposti ad un processo di maturazione che comporta la rimozione di tratti polinucleotidici interni chiamati **introni** e la concatenazione delle restanti regioni chiamate **esoni** (**meccanismo di splicing**).

L'mRNA maturo è traslocato nel citoplasma per essere tradotto in proteina.

# Lo splicing alternativo

E' un meccanismo attraverso il quale uno stesso pre-mRNA può subire eventi di splicing differenti con la produzione di mRNA maturi alternativi che possono dare origine a proteine diverse.

Le proteine **alternative** generate dallo splicing (isoforme) possono avere **funzioni simili, distinte** o perfino **antagoniste**.

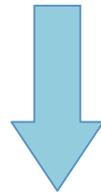


# Lo splicing alternativo contribuisce all'espansione della capacità di espressione del genoma umano

- Oltre il 95% dei geni umani è espresso mediante splicing alternativo
- Il numero medio trascritti prodotti da un gene è di circa 9-10



**Numero di trascritti umani è maggiore di 200.000**



**Aumento della complessità del trascrittoma e del proteoma di almeno un ordine di grandezza**

# Regolazione Epigenetica

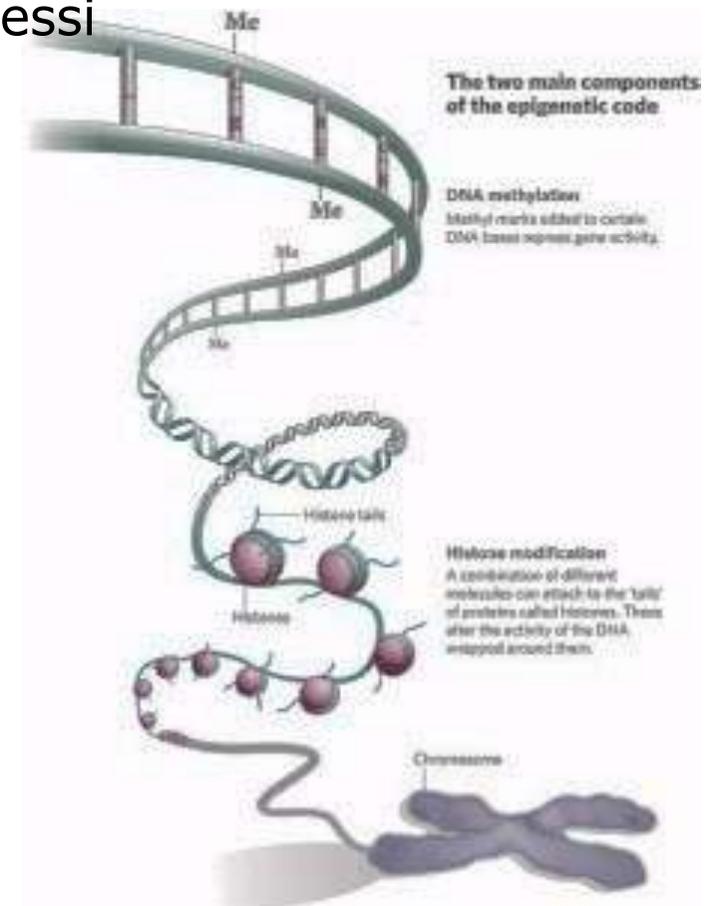
La **regolazione epigenetica** indica un meccanismo di controllo dell'espressione dei geni che non è dipendente dalla sequenza del DNA.

Cio' significa che il fenotipo di una cellula è determinato non solo dal genoma ereditato, quanto anche da una sorta di "impronta" che ne influenza il comportamento funzionale.

Cambiamenti di questi caratteri possono essere trasmessi attraverso le divisioni cellulari.

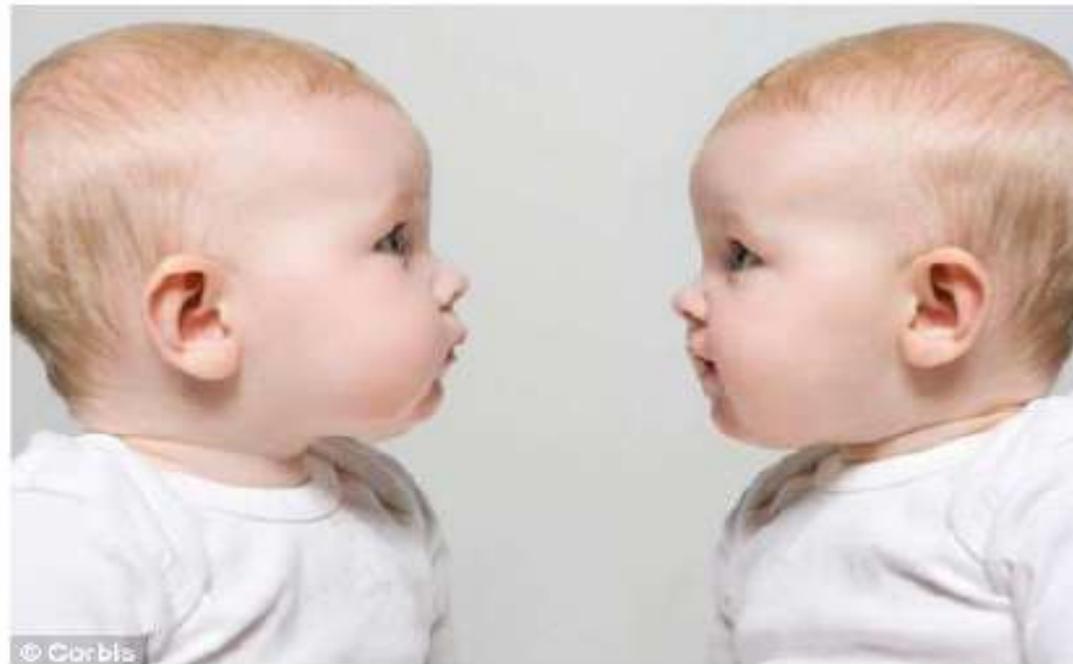
I principali meccanismi alla base dell'ereditarietà epigenetica sono:

- la metilazione del DNA;**
- il **rimodellamento della struttura della cromatina** (modificazione dei nucleosomi)



# Regolazione epigenetica nell'uomo

- I gemelli omozigoti hanno identico corredo genetico.
- Ma due gemelli omozigoti non sono identici dal punto **epigenetico**!
- Due gemelli geneticamente predisposti all'insorgenza di una malattia, potranno o non sviluppare entrambi la malattia.
- In tal caso entrano in gioco i **fattori epigenetici** che determinano l'insorgenza della malattia in uno e non nell'altro.

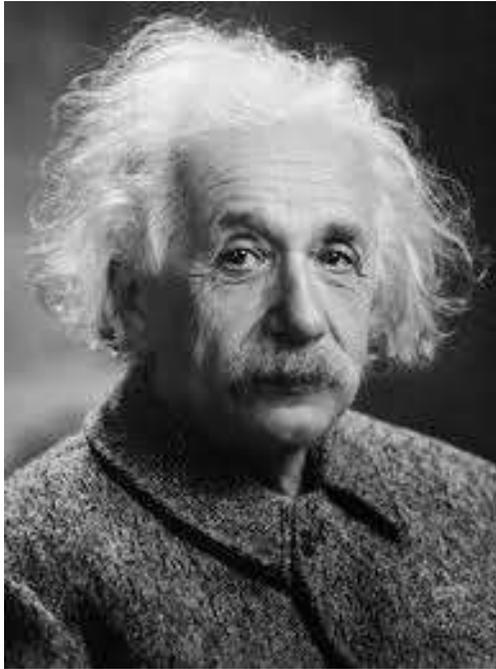


# La variabilità genetica

- Il 99.9% del DNA è identico in tutti gli individui
- Lo 0.1% del DNA mostra variabilità



# La variabilità genetica

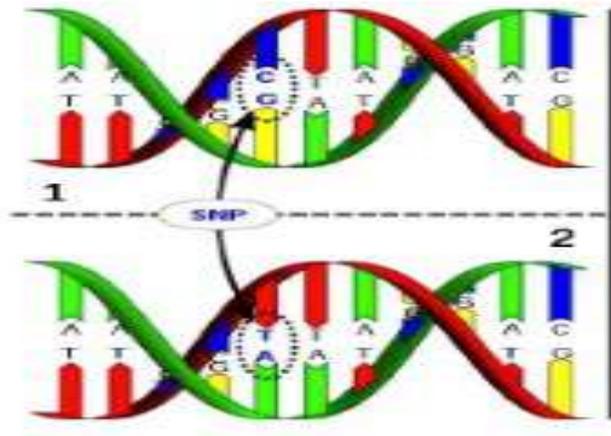


La **componente genetica individuale** è alla base della:

- ✓ **variabilità fenotipica** (diverse altezze, colore degli occhi, dei capelli, ecc);
- ✓ **risposta individuale** all'ambiente
- ✓ diversa **predisposizione** allo sviluppo di malattie complesse;
- ✓ diversa **risposta ai farmaci**

# La variabilità genetica

La maggior parte delle differenze tra individui sono rappresentate da cambiamenti di singole basi nella sequenza del DNA chiamati **singoli polimorfismi nucleotidici (SNPs)**.

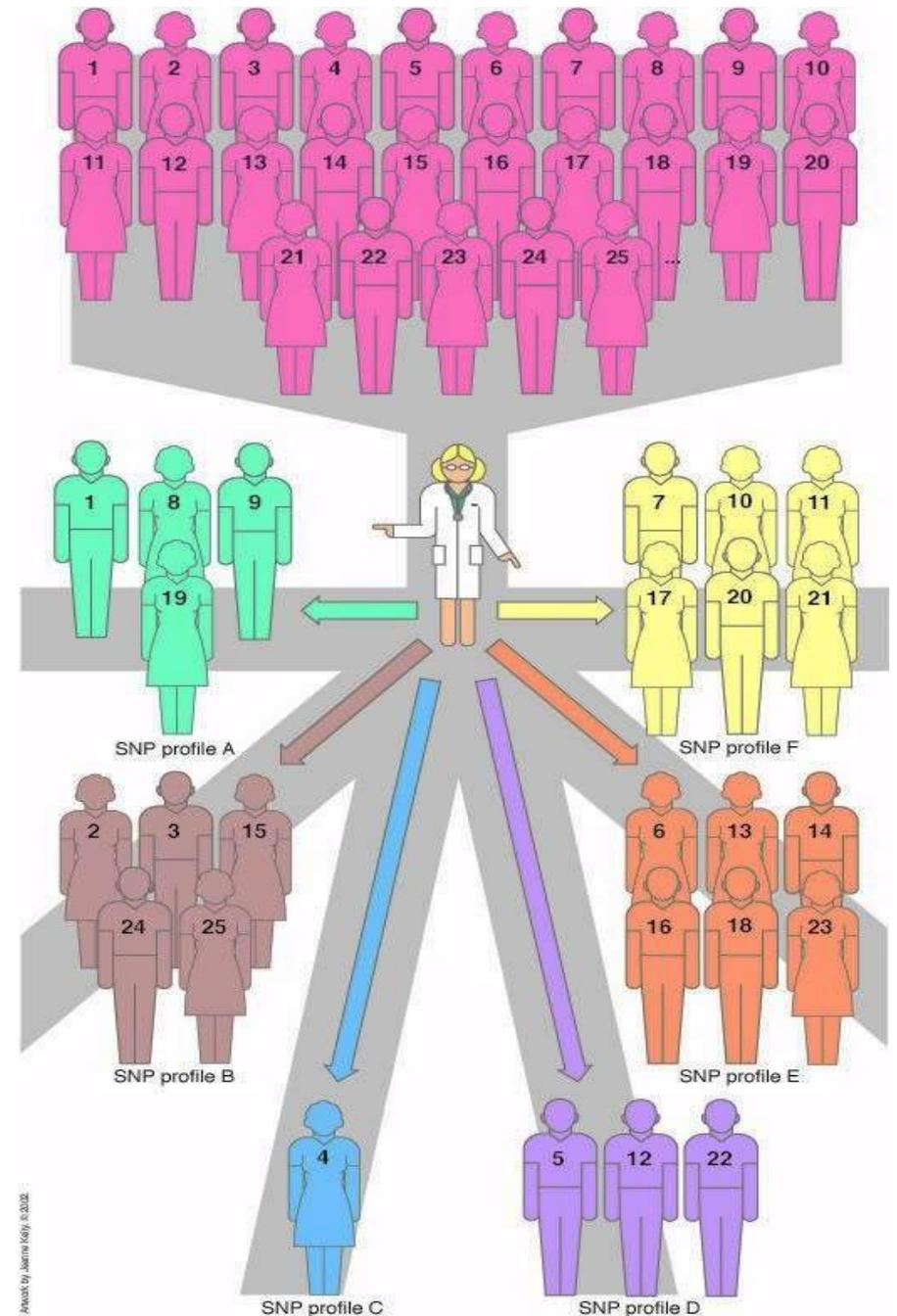


```
...CC A TTGAC...  
...GG T AACTG...  
...CC G TTGAC...  
...GG C AACTG...
```

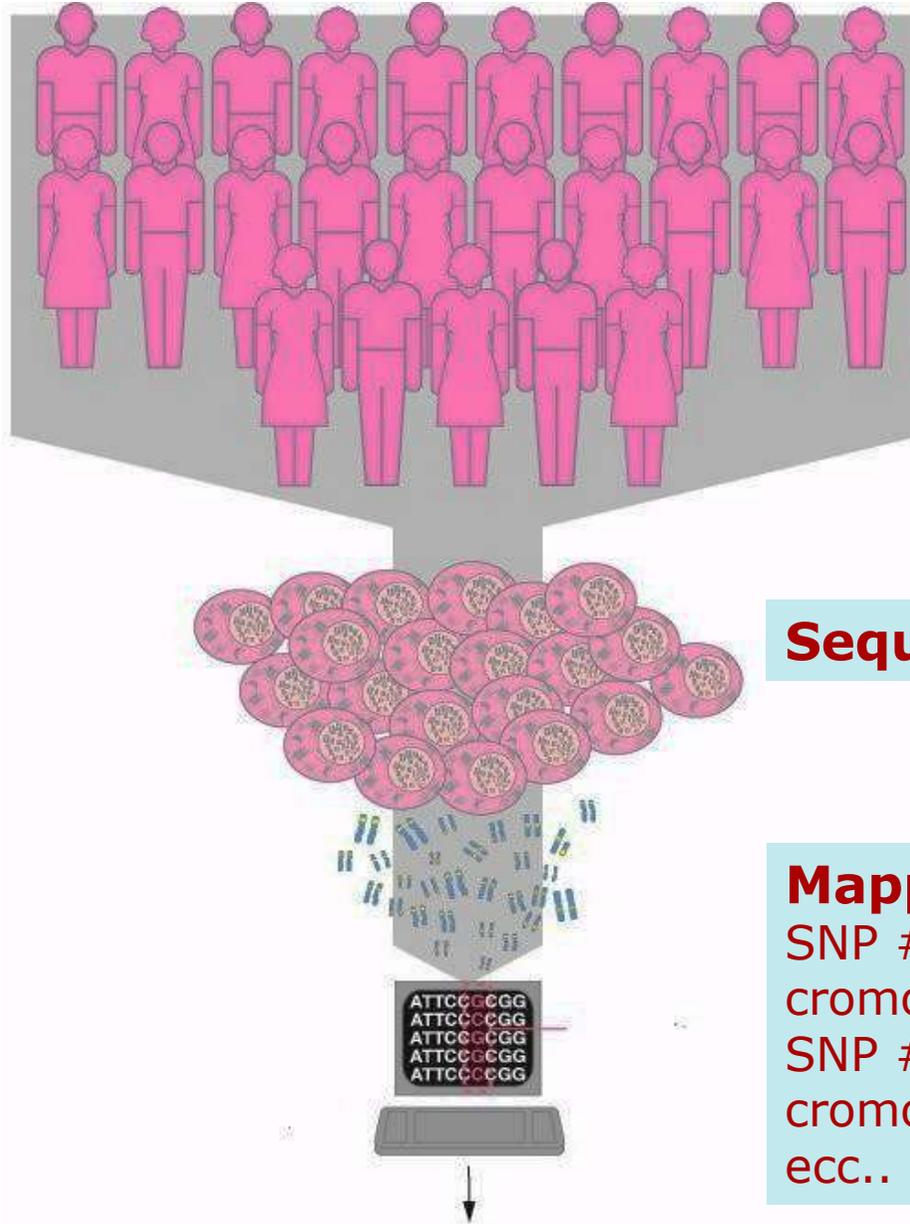


# Profilo individuale degli SNPs

- ▶ Il genoma di ciascun individuo contiene un assortimento diverso di SNPs
- ▶ Gli individui possono essere riuniti in classi in base al loro profilo di SNPs
- ▶ Il profilo di SNPs è importante per identificare la risposta ad un farmaco o la predisposizione ad una malattia
- ▶ Si possono individuare delle correlazioni tra il profilo di SNPs e la risposta ad uno specifico trattamento farmacologico



# Identificazione degli SNPs



**Sequenziamento del genoma**



**Mappatura degli SNPs:**  
SNP #1 nella posizione x sul cromosoma n;  
SNP #2 nella posizione y sul cromosoma m;  
ecc..

**2008**

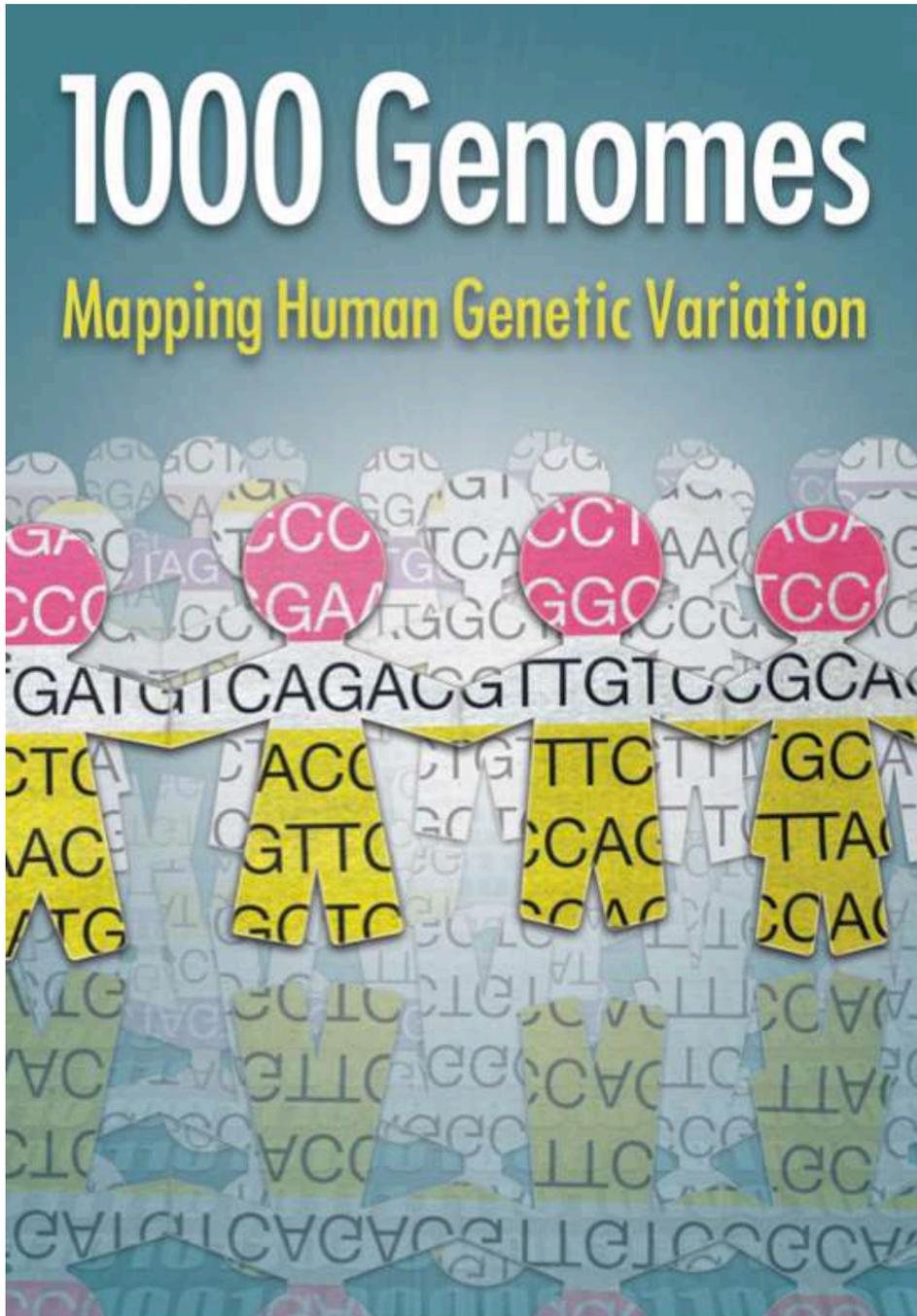
## **1000 Genomes Project**

**Scopo:** identificare le varianti nel genoma umano con una frequenza di almeno 1% nella popolazione

Sequenziati i genomi di 2500 individui appartenenti a 26 popolazioni diverse



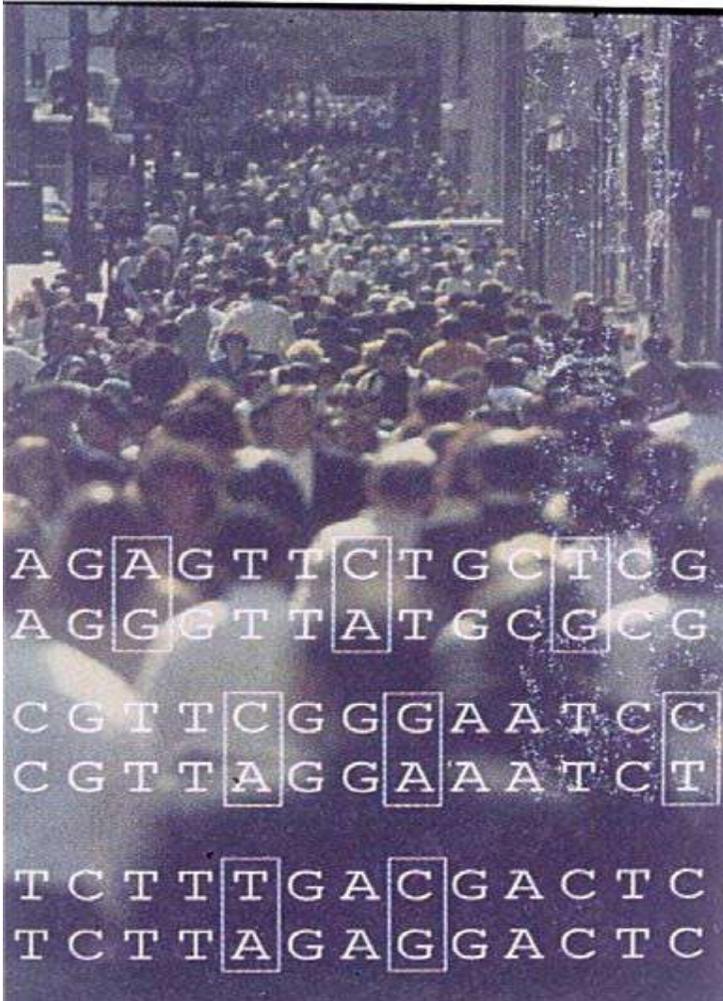
**88 milioni di varianti**



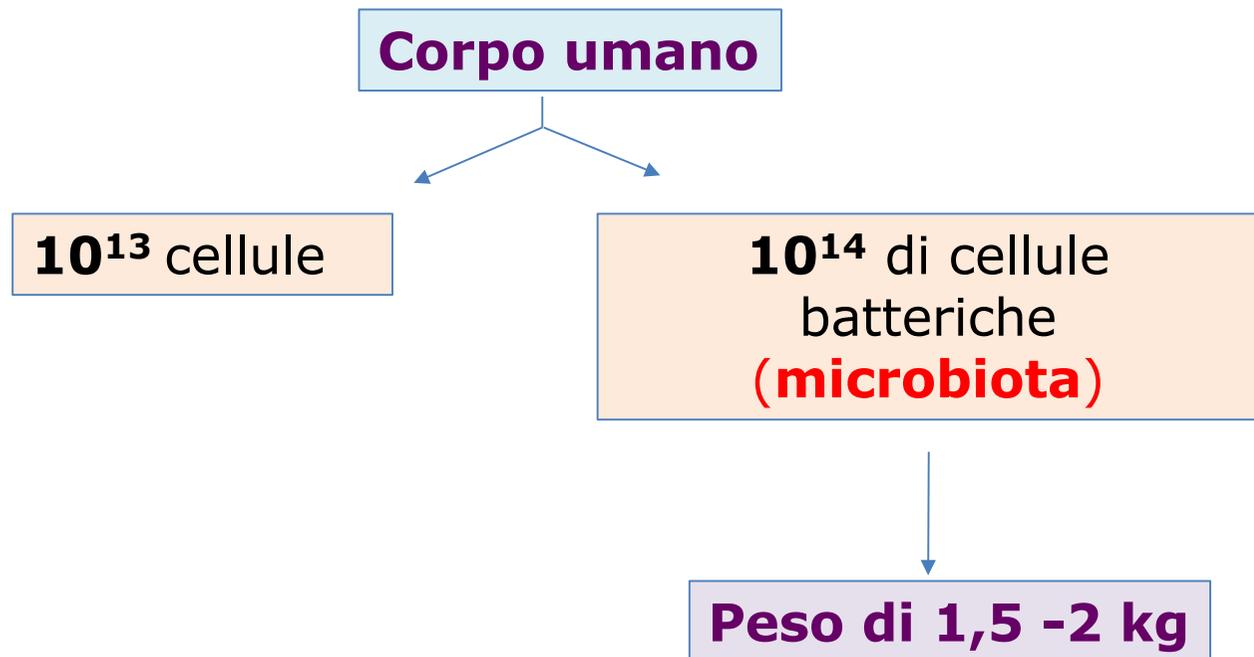
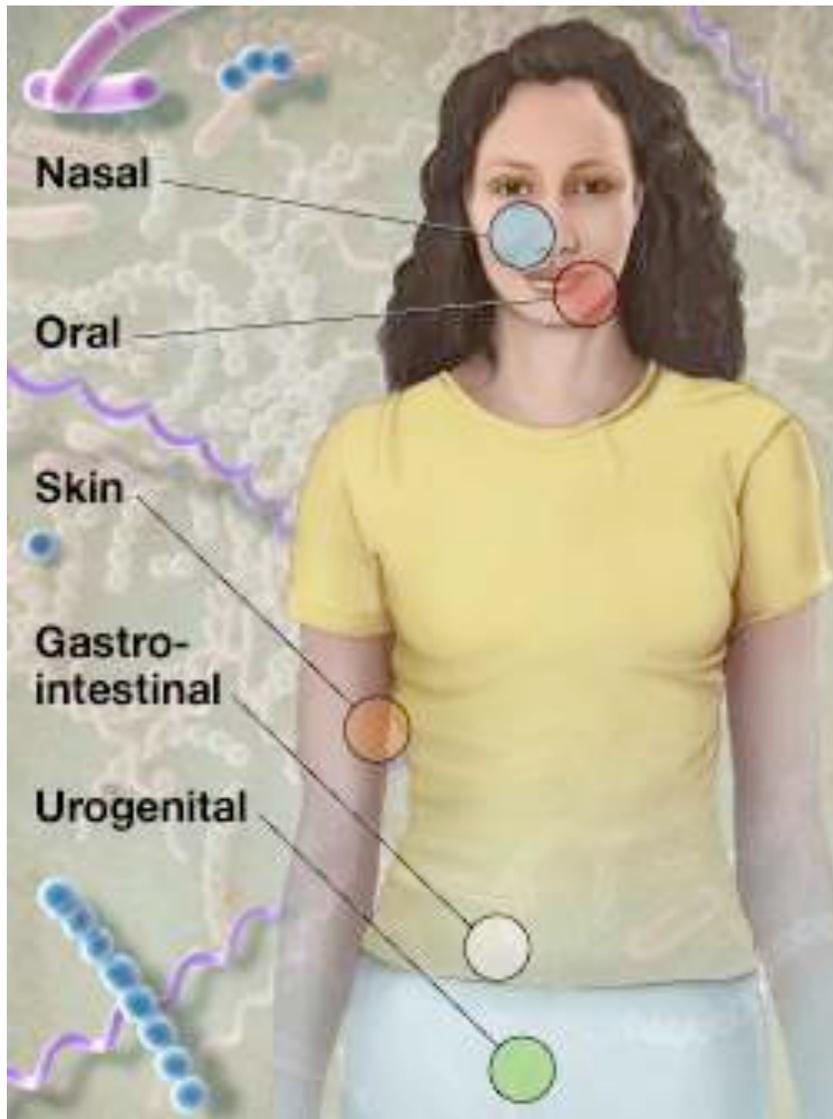
# 1000 Genomes Project

## I numeri

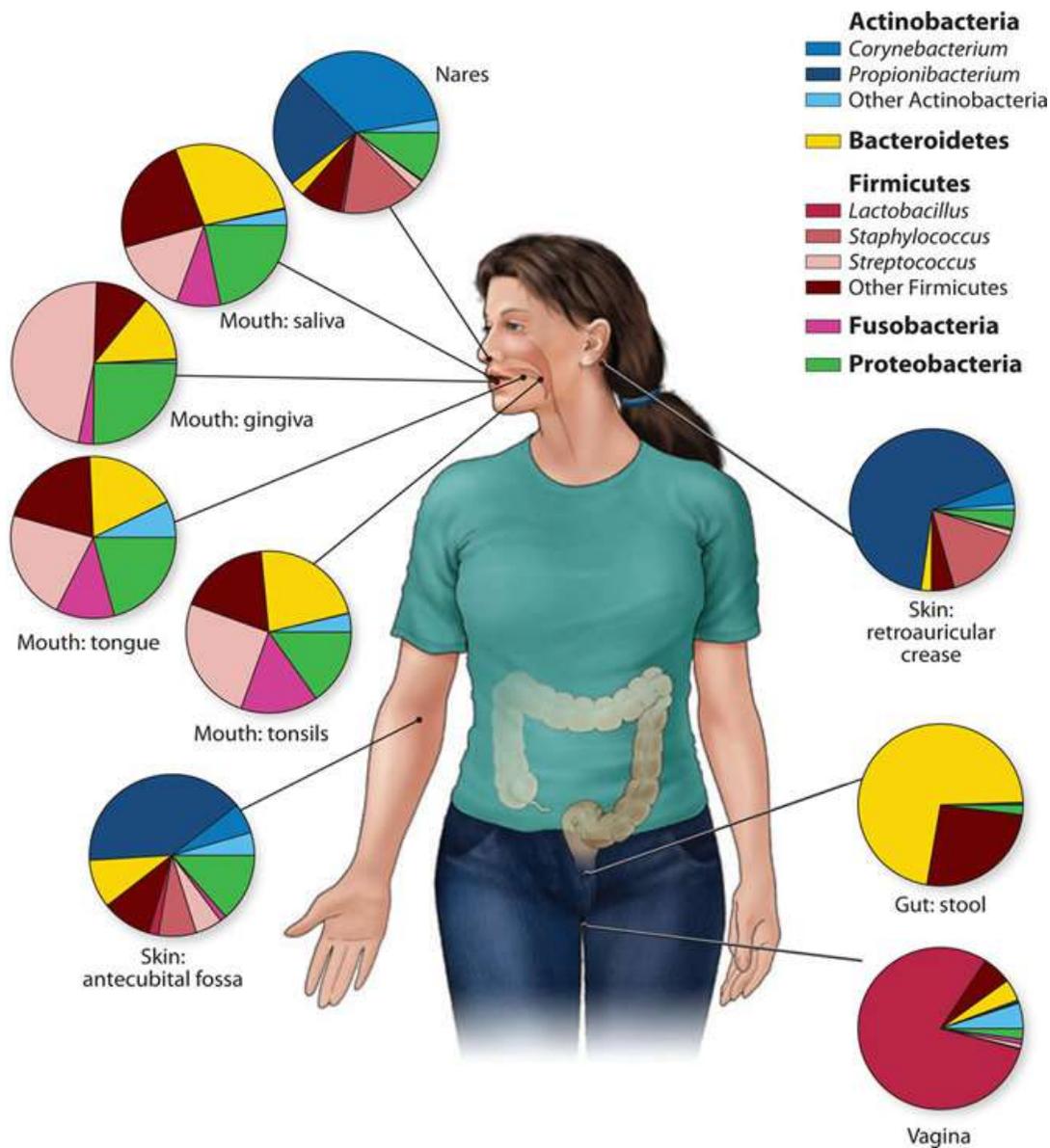
- 88 milioni SNPs
- 4 – 8 SNPs presenti in ogni gene
- 300-1000 bp che separano ogni SNP
- 150.000 SNPs non sinonimi
- 6% SNPs rari (<20% )
- 53% SNPs comuni (>20% )



# Il nostro secondo genoma, il microbioma



# Il nostro secondo genoma, il microbioma



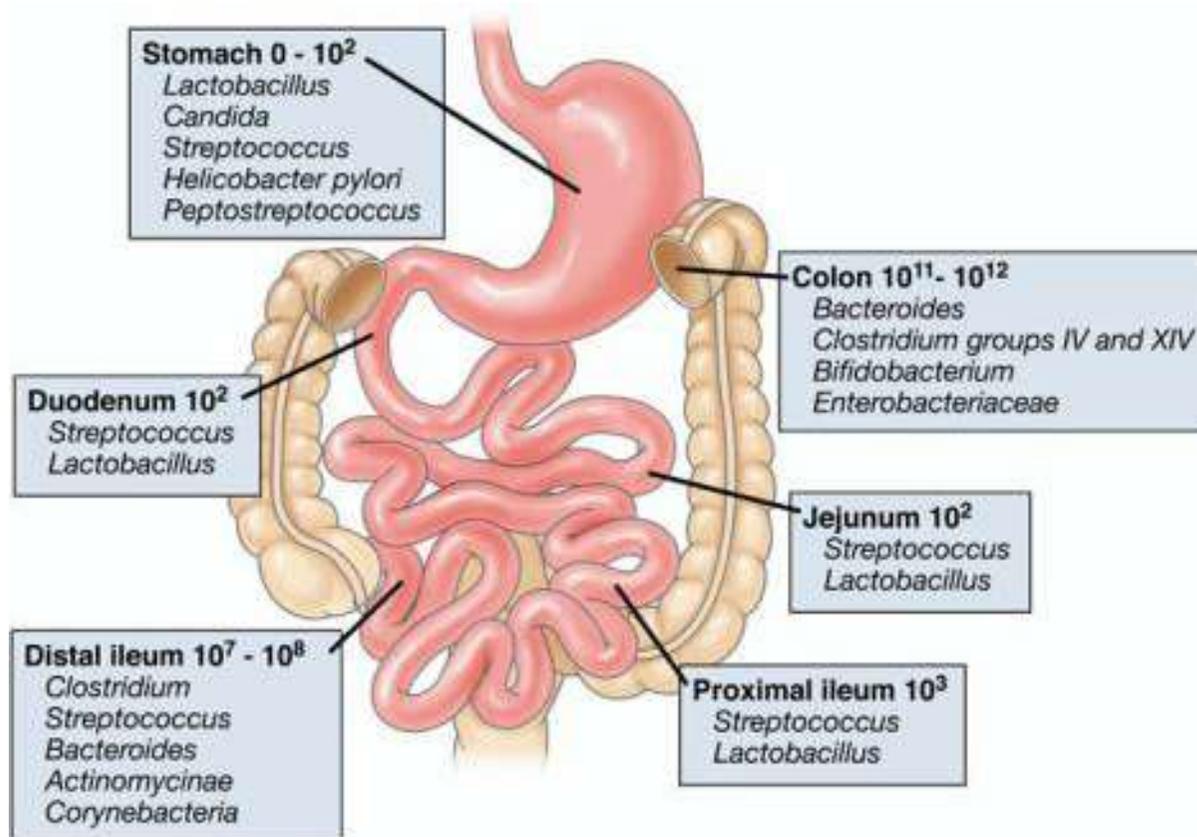
Ad oggi sono state individuate più di **10,000 specie microbiche**

**Genoma umano**  
**25.000 geni**

**Microbioma**  
**8.000.000 geni**

**Metagenoma umano**  
Geni Homo sapiens  
Geni di miliardi di microrganismi

# Il microbioma intestinale



## Microbioma intestinale

–2 kg di microorganismi

–interagiscono su 300mq di intestino (quasi un campo da tennis)

- Inizia a formarsi alla nascita
- Il parto e la nutrizione del neonato influiscono sullo sviluppo iniziale del microbioma
- L'ambiente e la dieta influenzano il microbioma maturo
- La composizione del microbioma cambia con l'età

# Quali sono i benefici del microbioma intestinale?

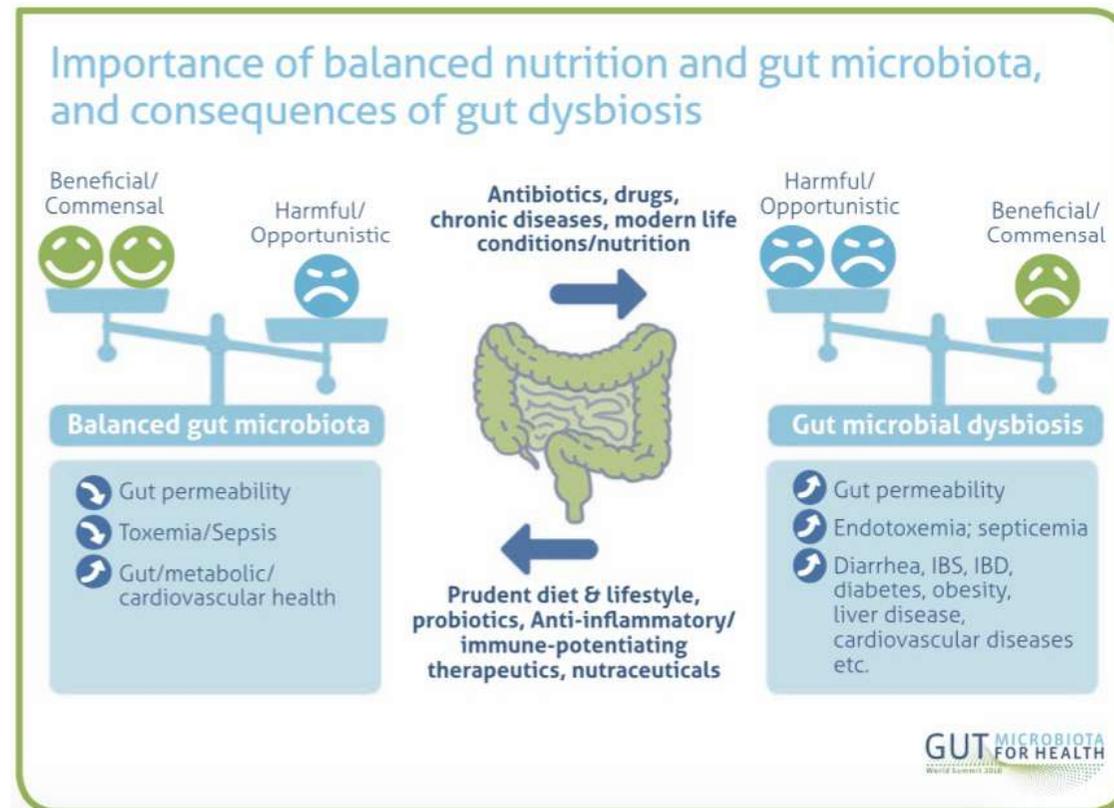
- Influenza lo sviluppo delle cellule della mucosa intestinale
- Serve come barriera contro i microrganismi patogeni (possibili fonti di malattie)
- Stimola le difese immunitarie
- E' fonte di energie e di sostanze nutritive (es. vitamine, prodotti di fermentazione delle fibre)



# Importanza del microbioma intestinale

L'alterazione qualitativa e quantitativa della microflora intestinale (*disbiosi*) è legata allo sviluppo di diverse patologie, quali:

- infiammazioni croniche intestinali
- Obesità
- Diabete di tipo I
- Cancro del colon-retto
- Malattie neurodegenerative (es. Alzheimer, Parkinson, ....)
- .....



# Metagenomica

La **metagenomica** è una nuova disciplina che studia l'insieme dei diversi materiali genetici (detto **metagenoma**) derivanti dalle diverse specie presenti in un determinato ambiente.

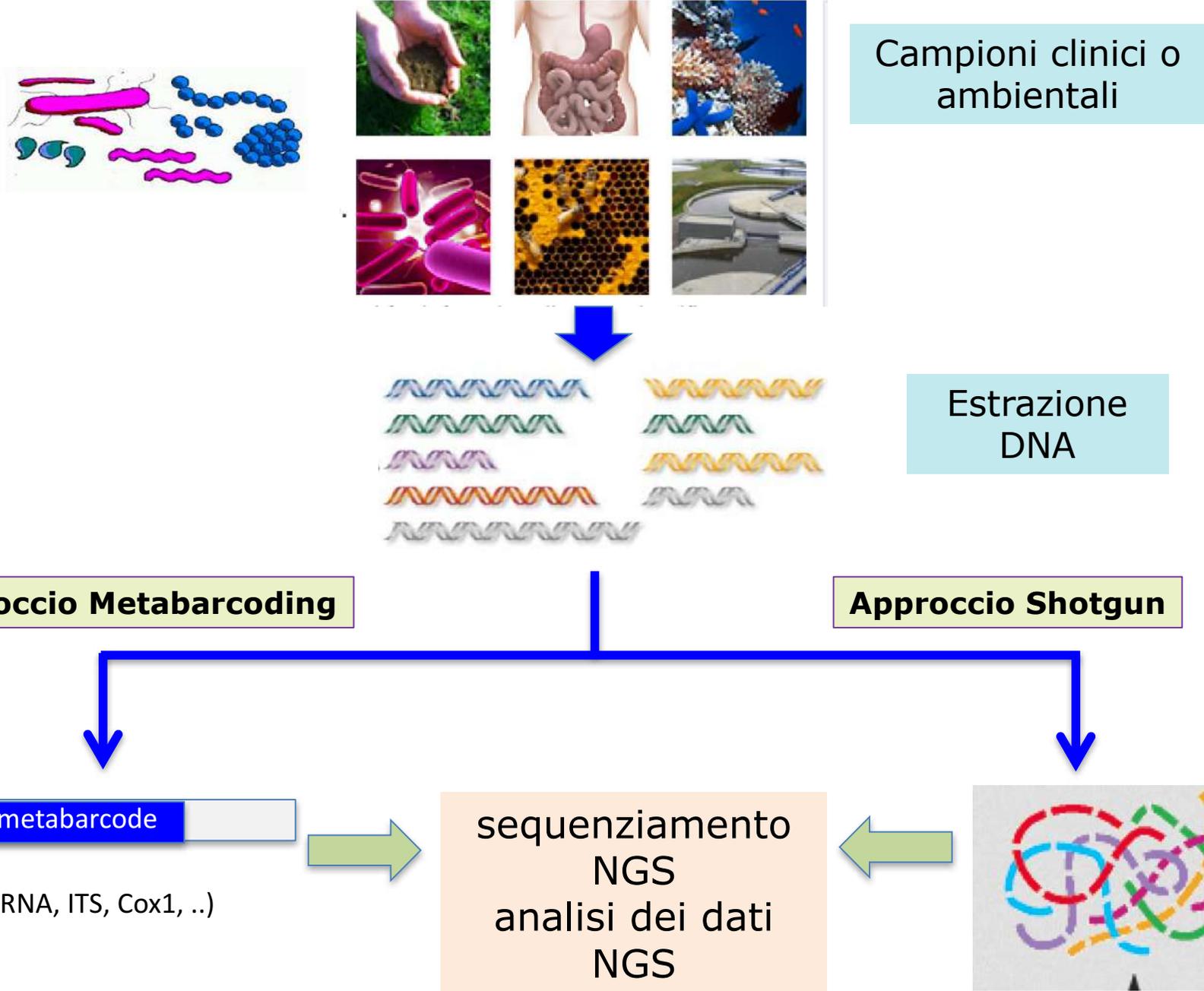
Ad esempio il materiale genetico dell'uomo e quello derivante dal suo microbiota (detto microbioma), rappresenta il **metagenoma umano**.

L'**approccio metagenomico** è applicato con successo per rivelare la diversità genetica e tassonomica delle comunità microbiche.

Non richiede l'isolamento e la messa in coltura di ogni singola specie, e la sua identificazione su basi biochimiche o morfologiche.



# Sequenziamento Metagenomico



# Sequenziamento Metagenomico

## Shotgun NGS

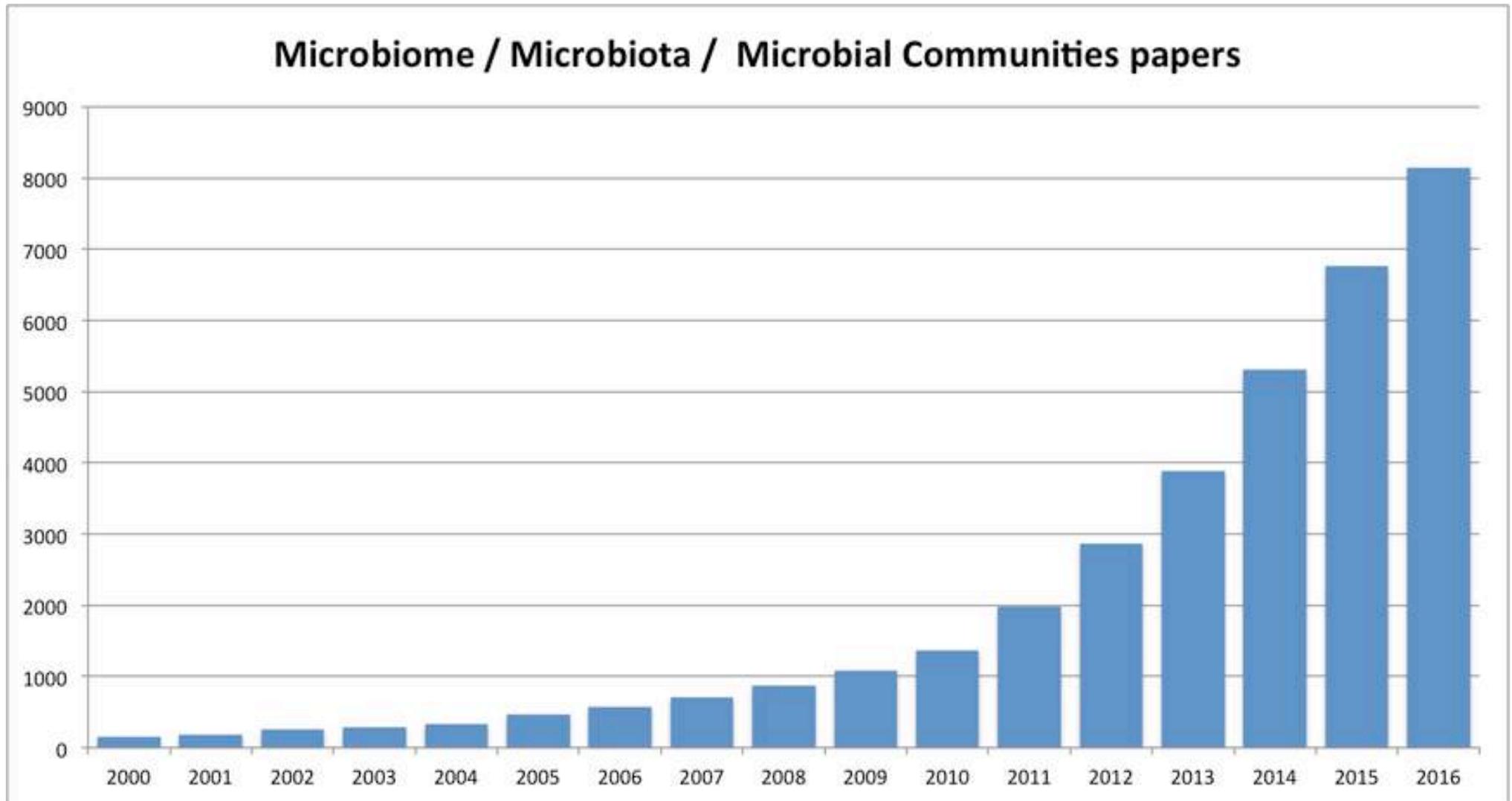
- identifica specie, geni e attività funzionali di comunità microbiche complesse
- coverage limitato per rilevare specie rare
- costoso sia in termini di sequenziamento che di analisi computazione



## Sequenziamento massivamente parallelo di ampliconi

- elevata sensibilità nella identificazione di specie
- poco costoso in termini di sequenziamento che di analisi computazione
- basato su PCR con primer universali
- richiede appropriato database di riferimento (es. RDP per 16S, ITSoneDB per ITS1)
- possibili distorsioni dovute a diversa efficienza di amplificazione in diversi gruppi tassonomici
- nessuna informazione funzionale

# Microbioma: un universo in via di esplorazione



Fonte:PubMed

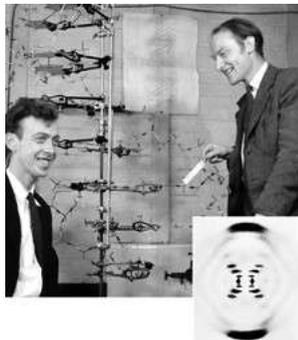
# Il Sequenziamento ....oggi

## Next-Generation Sequencing (NGS)

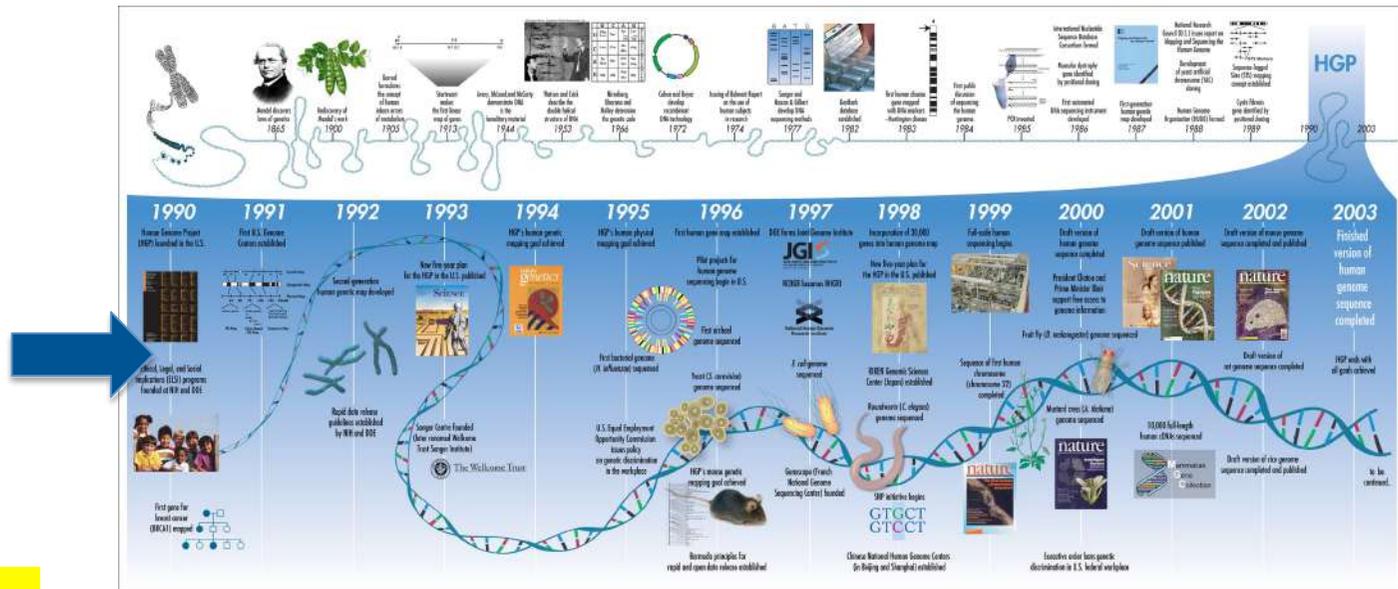
Le piattaforme di Next Generation sequencing consentono di generare centinaia di milioni di sequenze (35bp-300bp) in una sola corsa, in un tempo breve, con un basso prezzo per base sequenziata.



# Il sequenziamento del Genoma Umano oggi



1953 – DNA structure



Human Genome Project – 13 years - > 3 billion dollars

Next generation sequencing (NGS) technologies



Human genome - 2016 – few days - 1000 dollars !!

# Il Progetto Genoma Umano

Dal 2003 ad oggi



## Le sfide future

### Cosa ancora non sappiamo .....

- L' esatto numero di geni presenti nel genoma umano
- La funzione di circa il 50% dei geni
- La regolazione dell'espressione dei geni
- L' informazione contenuta nella porzione non codificante e la sua funzione
- I geni coinvolti nelle malattie complesse
- La correlazione tra variazioni della sequenza del genoma tra singoli individui e la suscettibilità a specifiche malattie
- La relazione tra microbioma e organismo

# Benefici del progetto Genoma Umano

## Medicina

- migliorare la diagnosi delle malattie
- identificazione di predisposizione genetica a specifiche malattie
- creazione di farmaci sulla base di informazioni molecolari
- produzione di " **farmaci personalizzati**" sulla base dei profili genetici individuali

## Forense

- identificazione di responsabili di crimini
- identificazione di vittime di catastrofi ecc.....

# Il Progetto Genoma Umano

oggi

La conoscenza della sequenza del genoma umano ancora **non ci ha fornito** le informazioni necessarie per la comprensione di tutti meccanismi alla base dei processi fisiologici e patologici dell'uomo.

L'analisi integrata di tutti i dati molecolari (genoma, trascrittoma, proteoma, microbioma), metabolici, funzionali, ecc. potrà finalmente far luce sul segreto della vita.....

**La strada è molto lunga!!**