



**UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO**

**DIPARTIMENTO DI
BIOSCIENZE, BIOTECNOLOGIE
E BIOFARMACEUTICA**

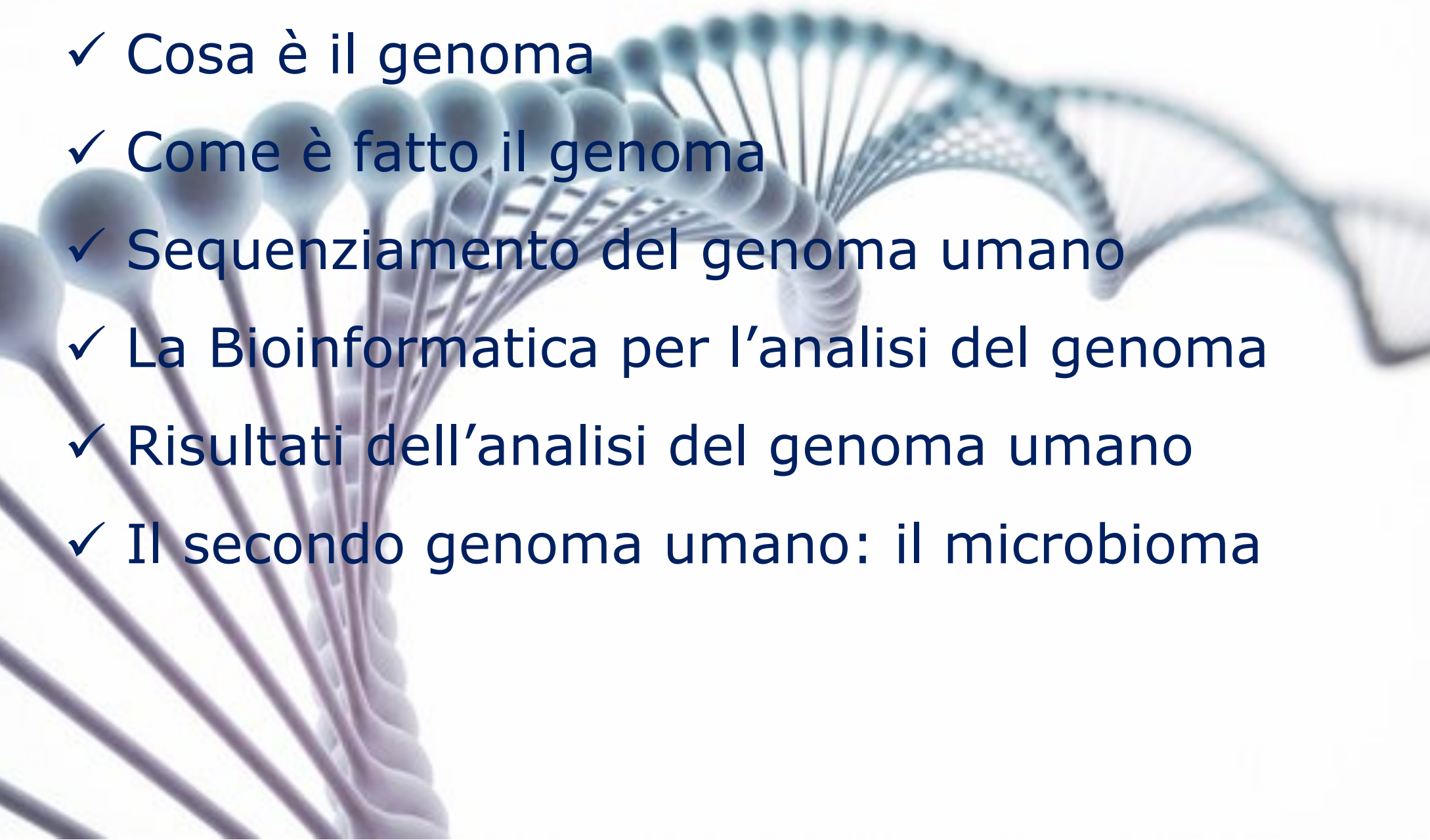
**Orientamento Consapevole
Biotecnologie Innovative**

**Bioinformatica e analisi del
genoma umano**

Anna Maria D'Erchia

Bari, 17 aprile 2020

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 secondo genoma umano: il microbioma
- 

Che cosa è il genoma?

Il **genoma** è l'insieme di tutte le informazioni biologiche necessarie alla costruzione e al mantenimento di ogni organismo vivente.

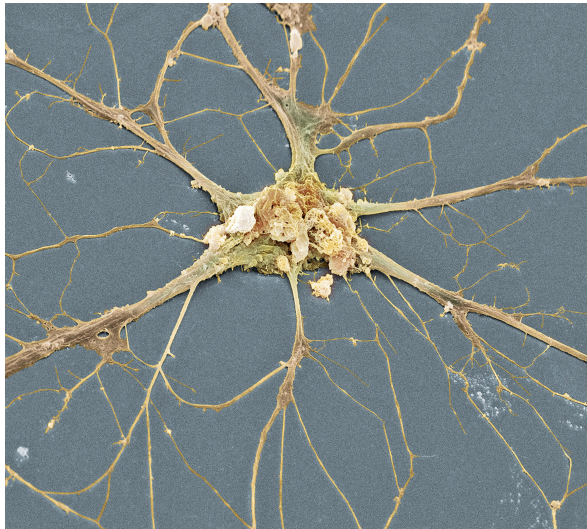
Può essere paragonato ad un'enorme manuale in cui sono contenute tutte le istruzioni che regolano lo sviluppo e il funzionamento di ogni organismo.



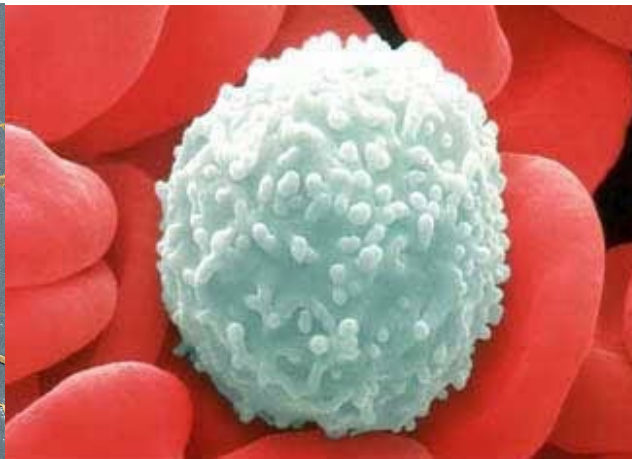
Come è fatto il genoma?

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

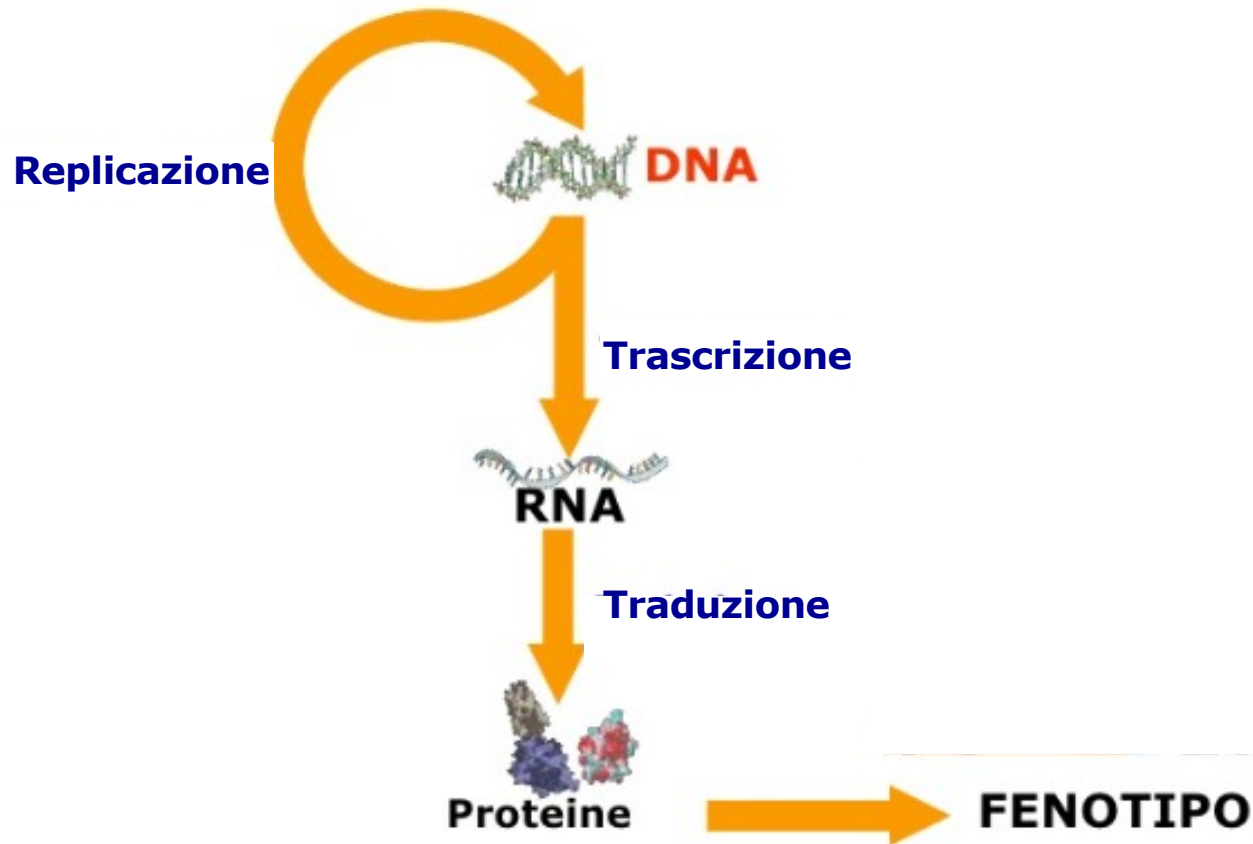


Globuli rossi

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



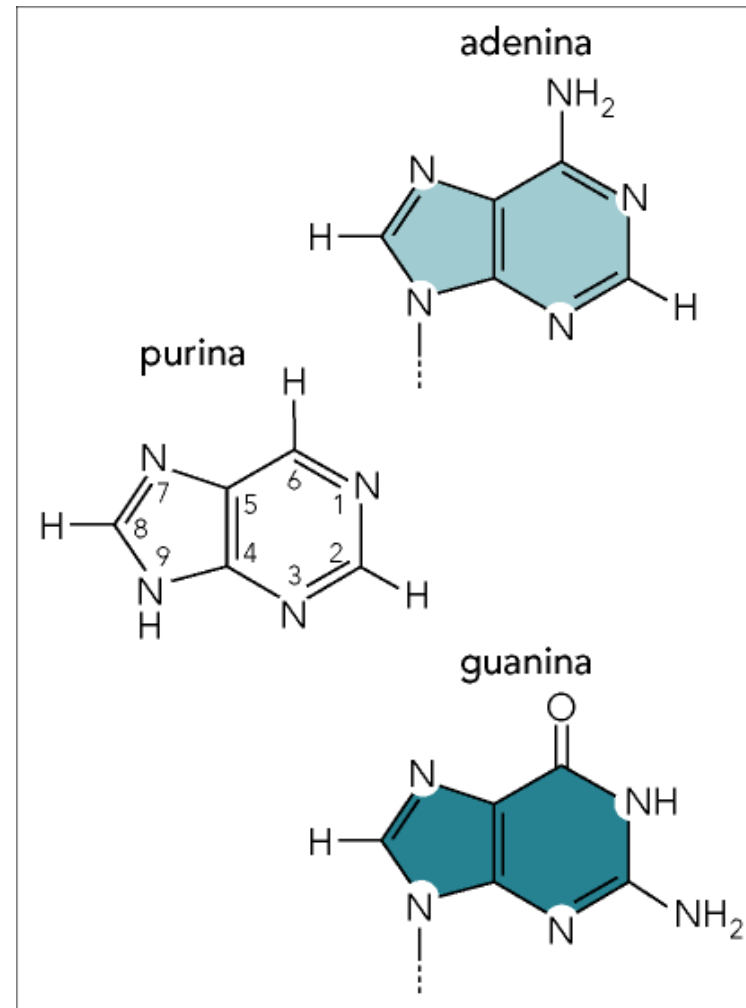
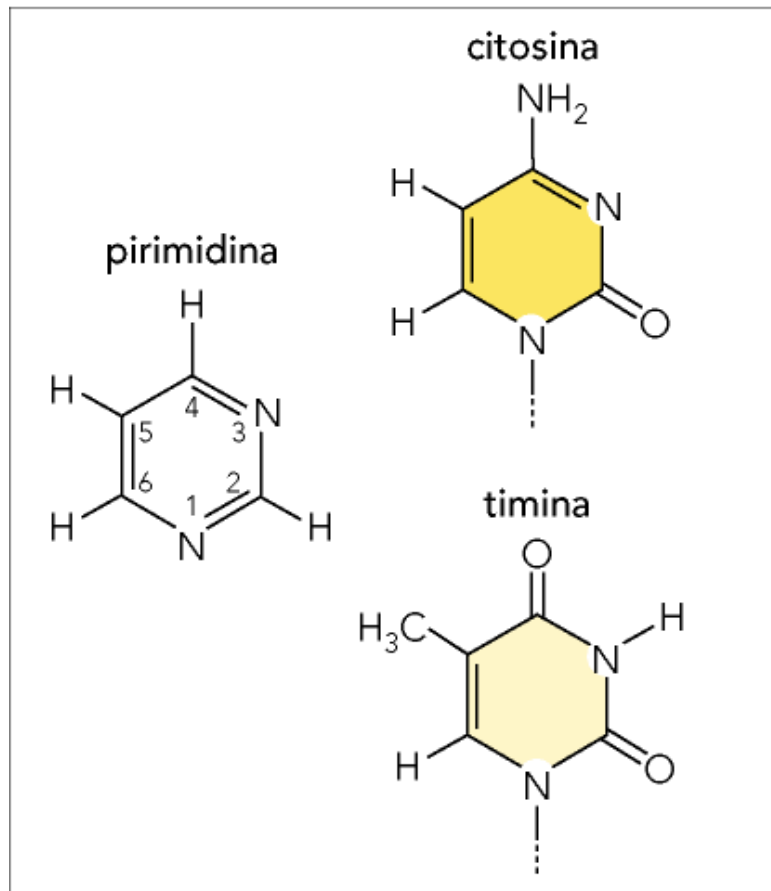
II DNA

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

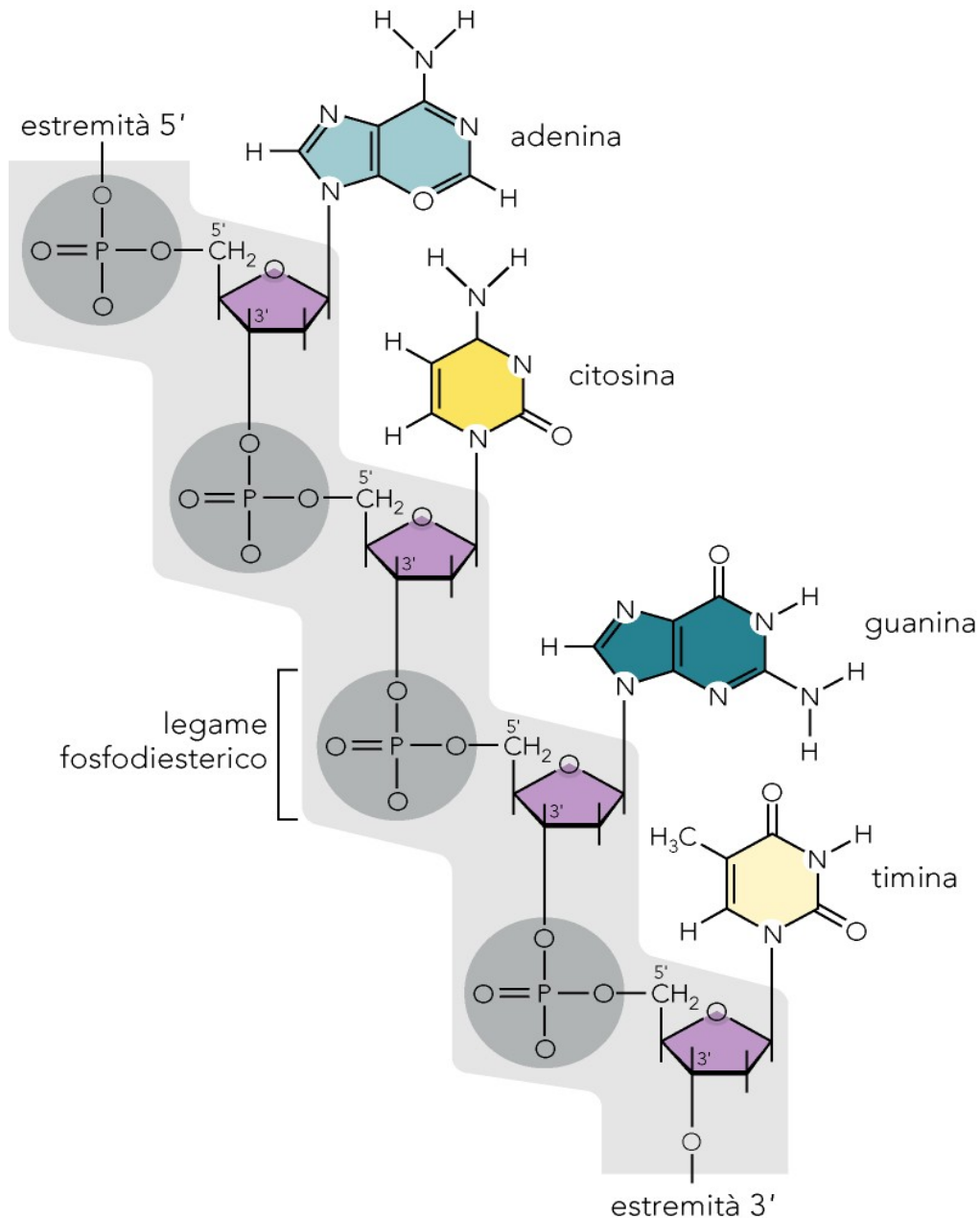
1. uno zucchero a 5 atomi di carbonio (deossiribosio)
2. una base azotata (tra 4 tipi diversi)
3. un gruppo fosfato



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**.

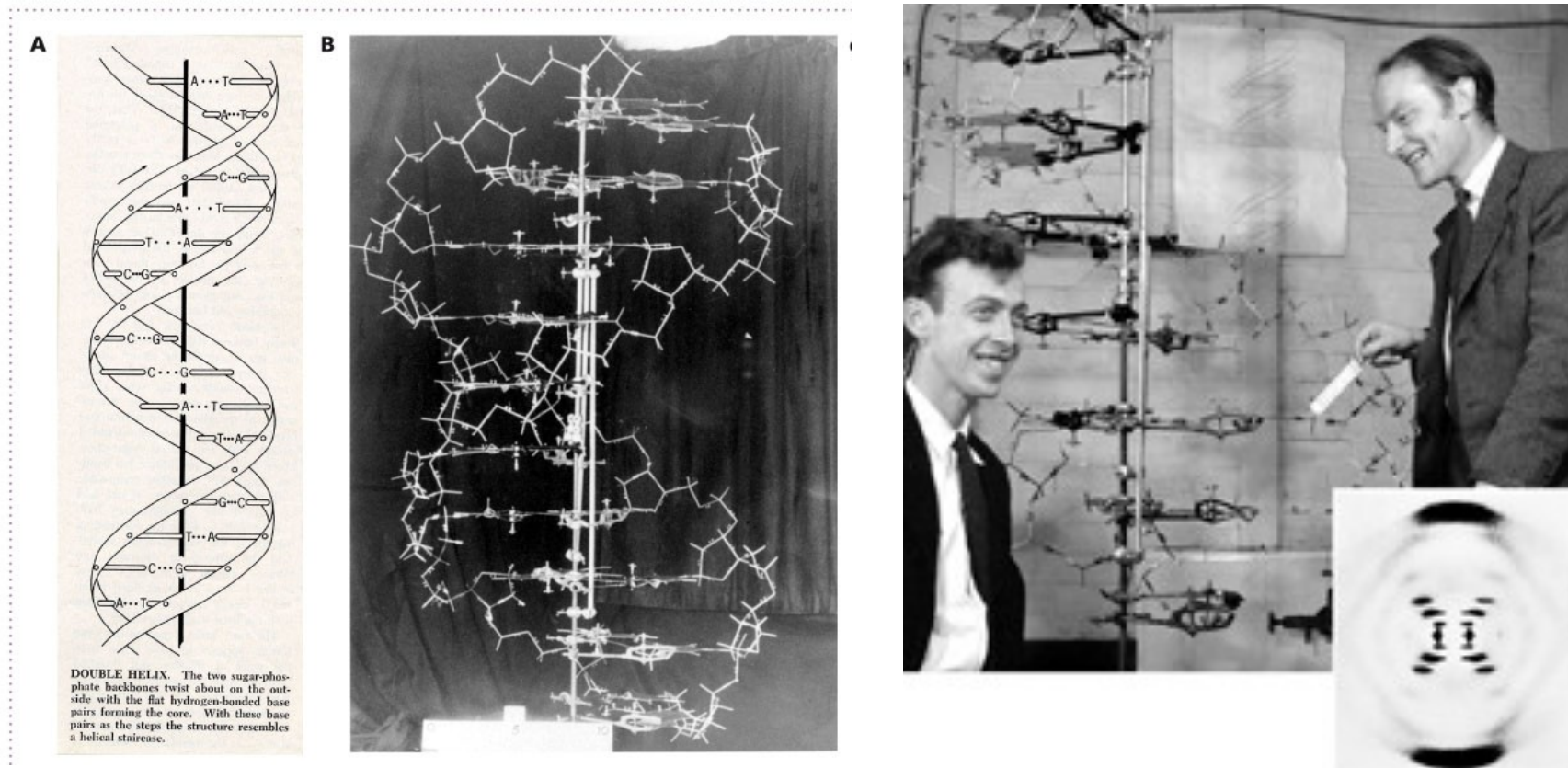
Lo scheletro è costituito dall'alternanza degli zuccheri e dei gruppi fosfato.

Le basi azotate sporgono lateralmente dallo scheletro.

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

La struttura della doppia elica del DNA

James Watson e Francis Crick nel 1953 definirono la struttura tridimensionale del DNA, anche grazie agli esperimenti di diffrazione dei raggi X condotti da Rosalind Franklin e Maurice Wilkins.



La struttura a doppia elica del DNA

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.

¹ Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

² Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

³ Von Arx, W. S., *Woods Hole Papers in Phys. Oceanog. Meteor.*, **11** (3) (1950).

⁴ Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

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.

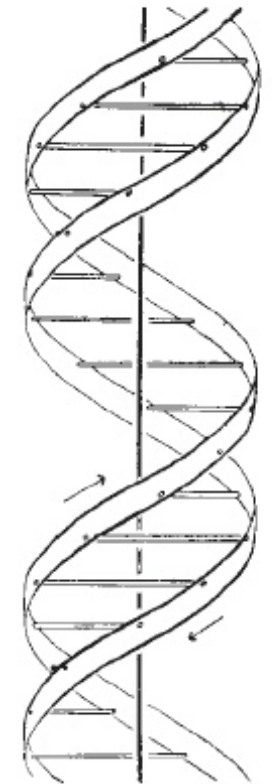
A structure for nucleic acid has already been proposed by Pauling and Corey¹. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure 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

is a residue on each chain every 3.4 Å. in the 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, or after 34 Å. The distance of a phosphate group from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to the fibre axis.

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

The novel feature of the structure is that in which the two chains are held together by hydrogen bonds between purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are held 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 their *z*-coordinates equal. 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.

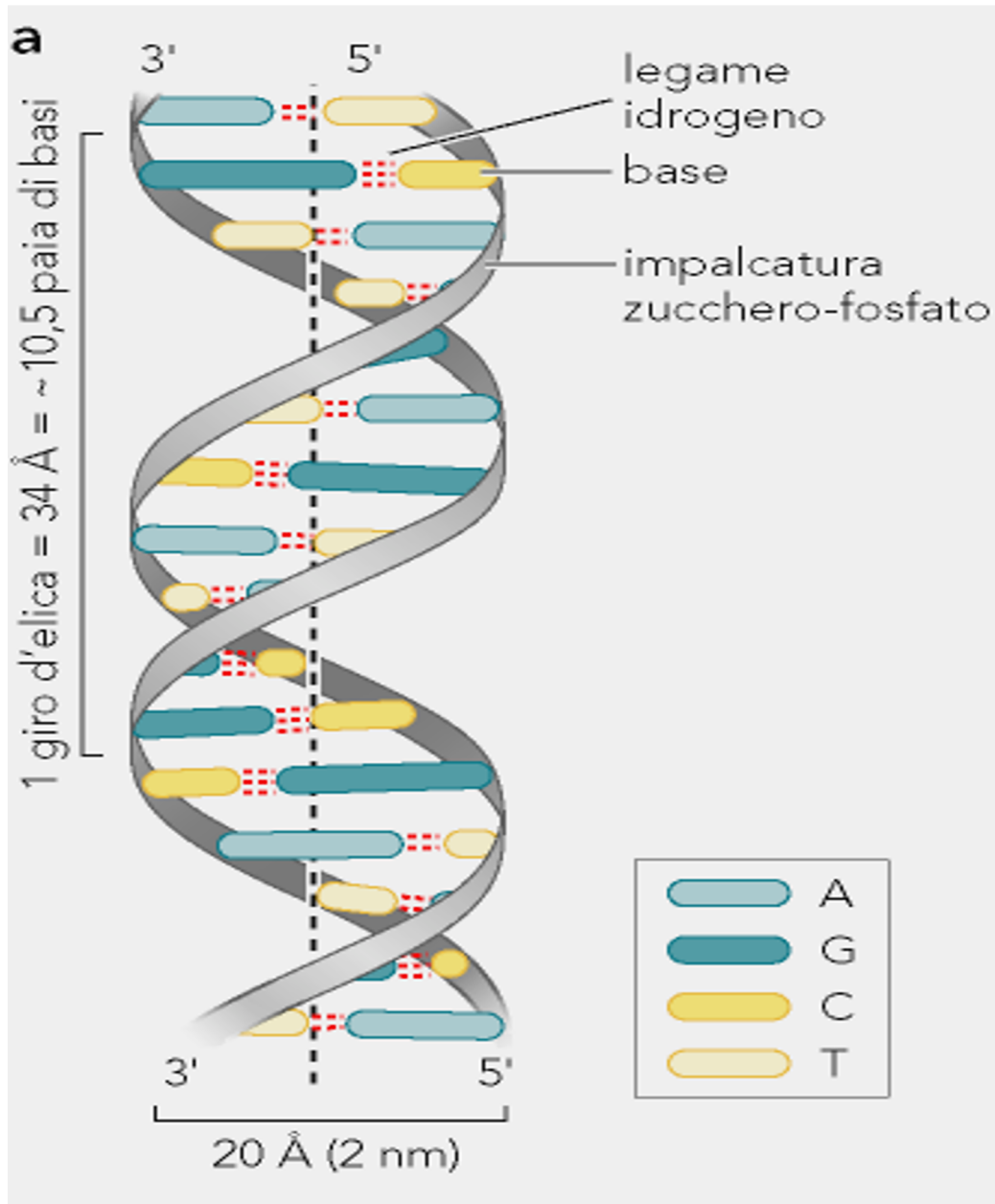
If it is assumed that the bases only occur in the most plausible tautomeric form (that is, with the keto rather than the enol form) it is found that only specific pairs of bases can bond together. These pairs are: (purine) with thymine (pyrimidine), and (purine) with cytosine (pyrimidine).



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.

Watson e Crick 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 ($5' \rightarrow 3'$) e hanno un'**orientamento antiparallelo**:

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

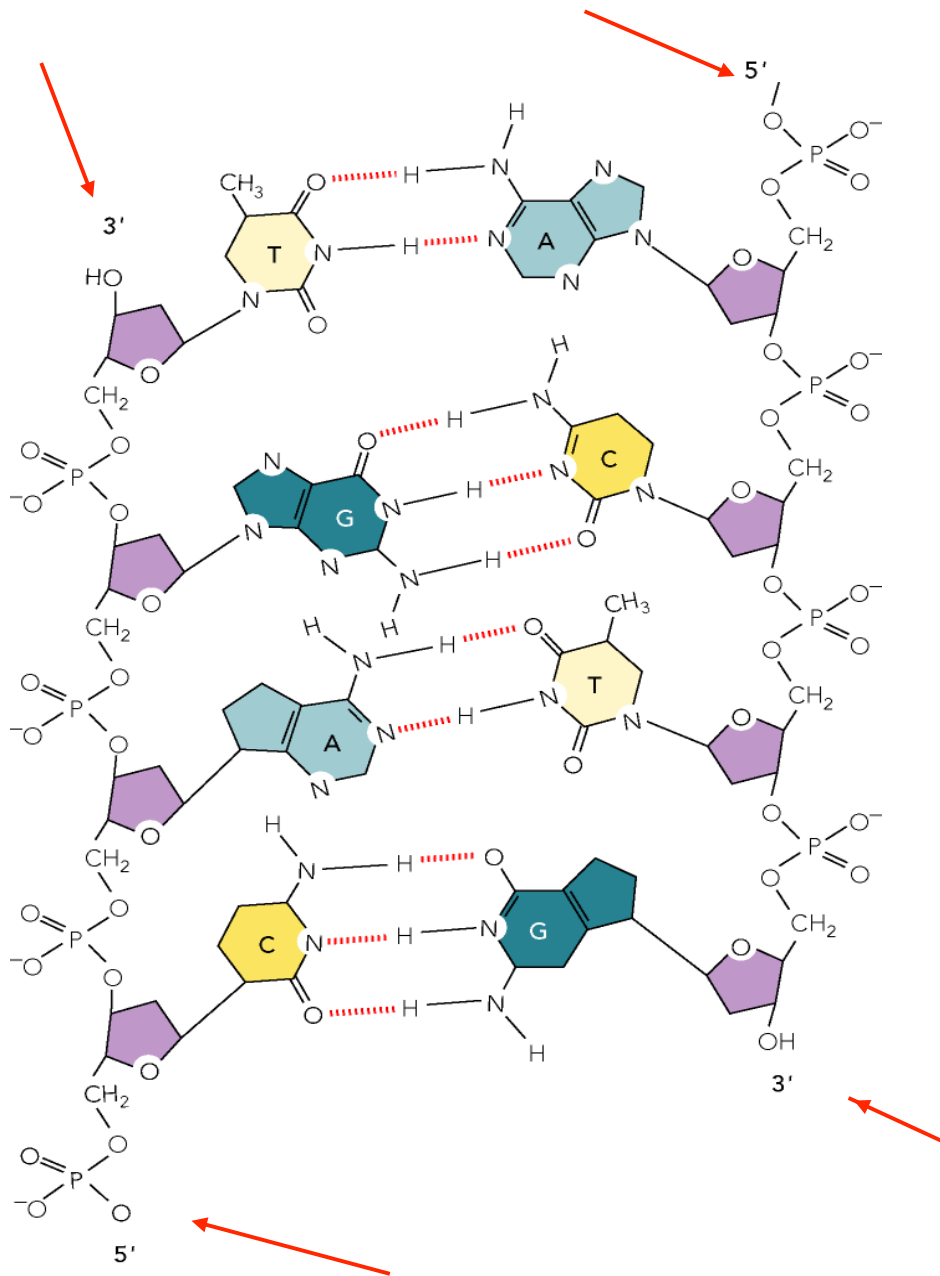
un filamento ha orientamento **3'-5'**

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

Le basi azotate sono perpendicolari all'asse dell'elica e separate da $3,4 \text{ \AA}$

La struttura elicoidale si ripete ogni 34 \AA , 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

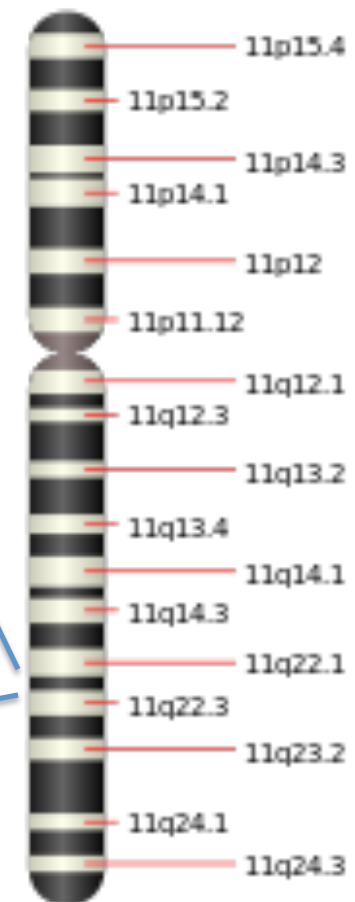
I due filamenti di DNA hanno **sequenze complementari**

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**.

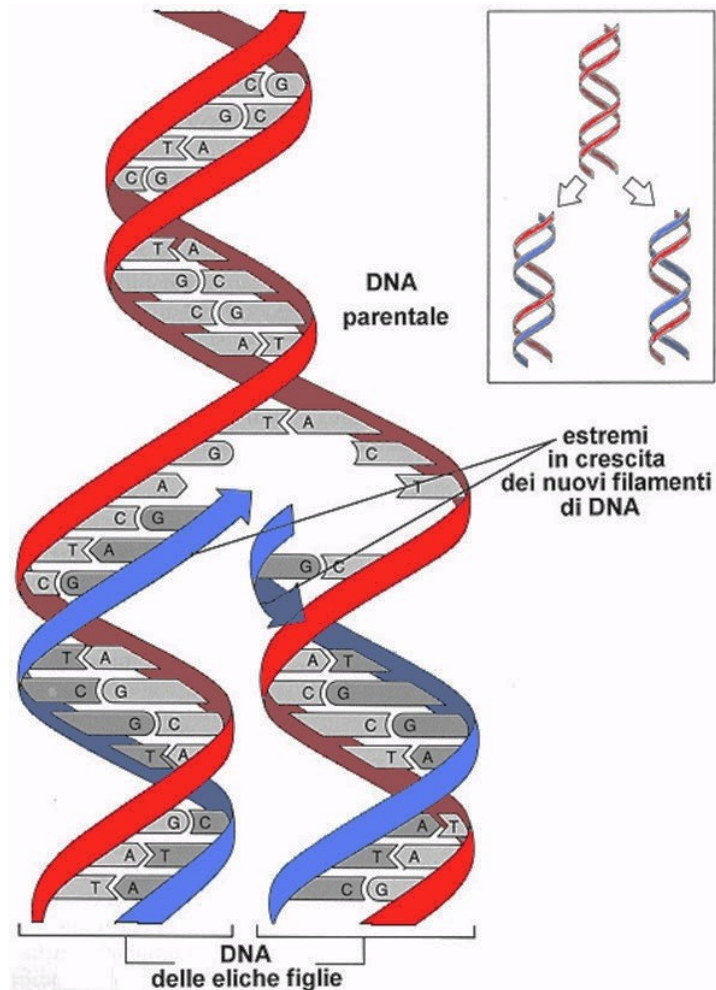
4 lettere, nessuna punteggiatura

```
atggcaattaaaattggtatcaatggttttggtcgtatcggccgtatcgtattccgtgca
gcacaacaccgtgatgacattgaagttgtaggtattaacgacttaatcgacgttgaatac
atggcttatatggtgaaatatgattcaactcacggtcgtttcgacggcactgttgaagtg
aaagatggtaacttagtggttaatggtaaaactatccgtgtaactgcagaacgtgatcca
gcaaacttaaactggggtgcaatcgggtgtgatatcgctggtgaagcgactggtttattc
ttaactgatgaaactgctcgtaaacatatcactgcagggcgcaaaaaaagttgtattaact
ggcccatctaaagatgcaaccctatgttcgttcggtgtaaaacttcaacgcatacgcgca
ggtcaagatatcgtttctaacgcacatcttgtaacaactggttagctccttagcacgt
caaaaatagggttaatatgaatctcgatctccatgttgcattcgtattcaacaacaagc
caaaactcgtaaaaatgaccgcacttcgctataaagaacacggcttcgagatatctct
tggaaaaactttcaagagcaactcaactttctcgagcattgcttgctcacaatatt
gacgtacaagataaaatcgccatgttccataaatggaacggtgggttgttcatgaa
actttcgggtatcaaagatggtttaatgaccactgttcacgcaacgactacaatcgttgac
attgcgaccttacaattcgagcaatcacagtgctatttacgcaaccaatacagcccag
caagcagaatttatcctaaatcacgcogatgtaaaaattctcttcgctcggcgatcaagag
caatacgatcaaacattggaaattgctcatcattgtccaaaattacaataaattgtagca
atgaaatccaccattcaattacaacaagatcctctttcttgacttgg
```



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.

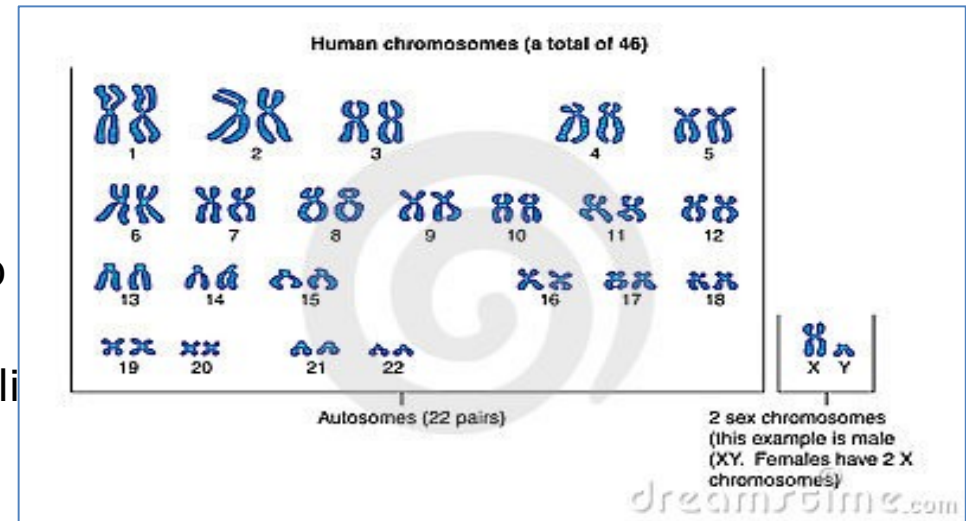
Il genoma umano

• Il Genoma nucleare

Costituito da circa 3×10^9 pb organizzate in 23 **Cromosomi**

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

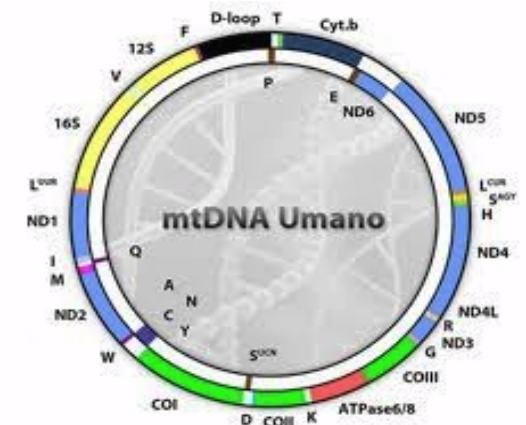
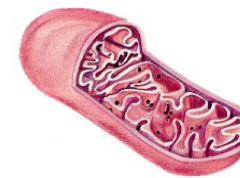
- 22 coppie di autosomi (coppie di cromosomi uguali)
- 1 coppia di cromosomi sessuali, X e Y
 - XX nelle donne, XY negli uomini



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

• Il Genoma mitocondriale

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



Il Progetto Genoma Umano

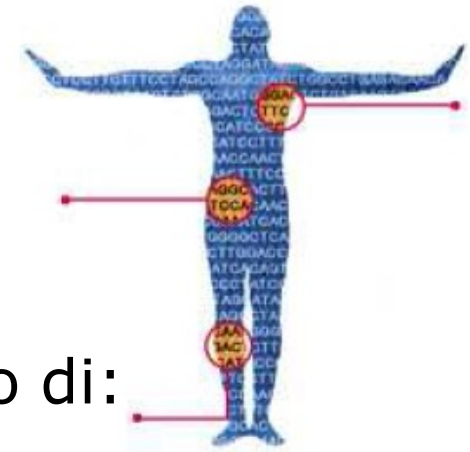
La conoscenza di tutta l'informazione racchiusa nel **genoma** è la condizione necessaria per comprendere la completa biologia di un determinato organismo, cioè significa comprendere il "segreto della vita".

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?

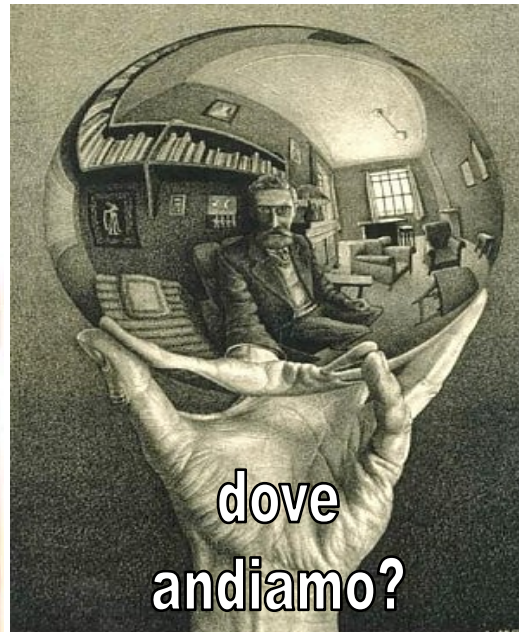
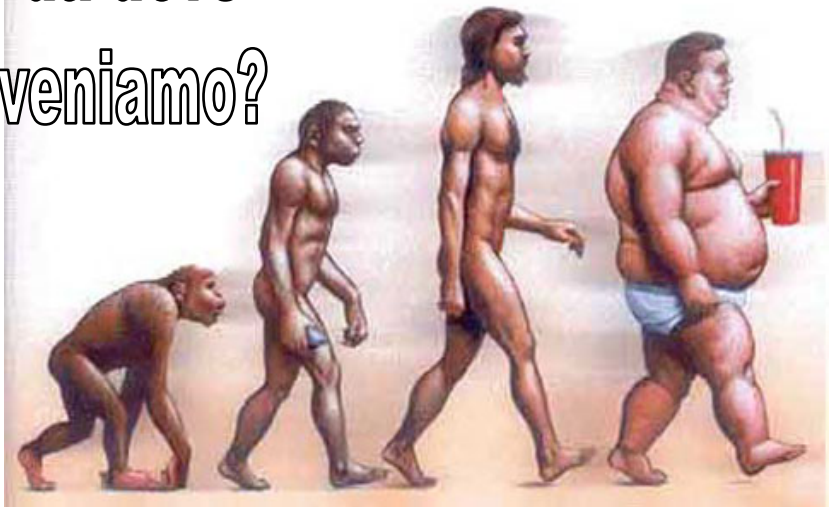
La sequenza del genoma umano avrebbe permesso di:

- identificare **tutti i geni umani**
- determinare **la struttura dei geni**
- definirere 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%)

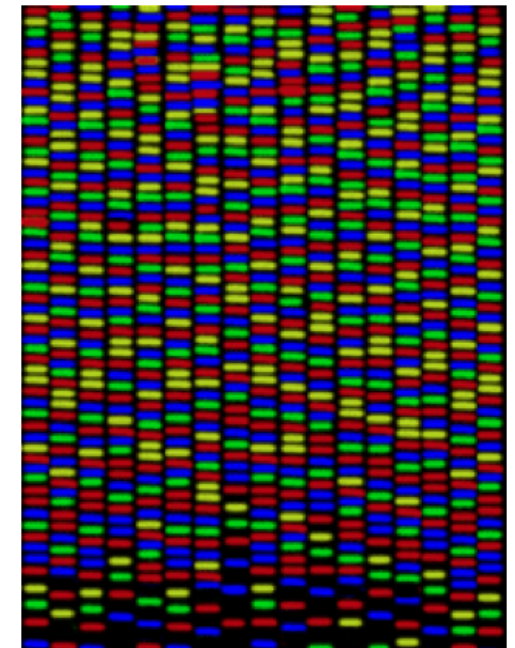
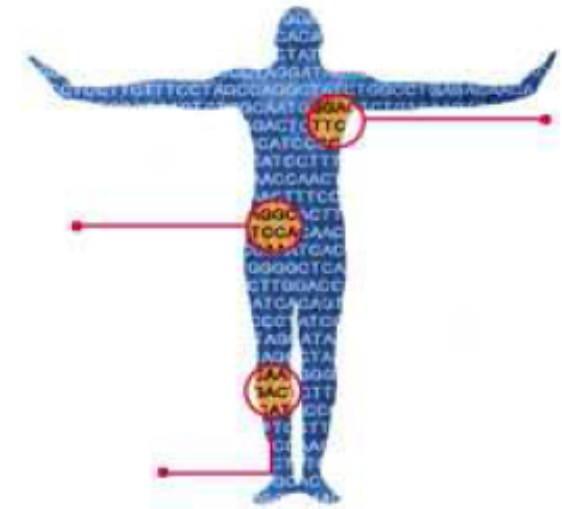
Sequenziamento del genoma

In che modo?

Sequenziamento del DNA significa **determinare la 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**

Il sequenziamento del genoma umano fu realizzato mediante il **Metodo di Sanger**

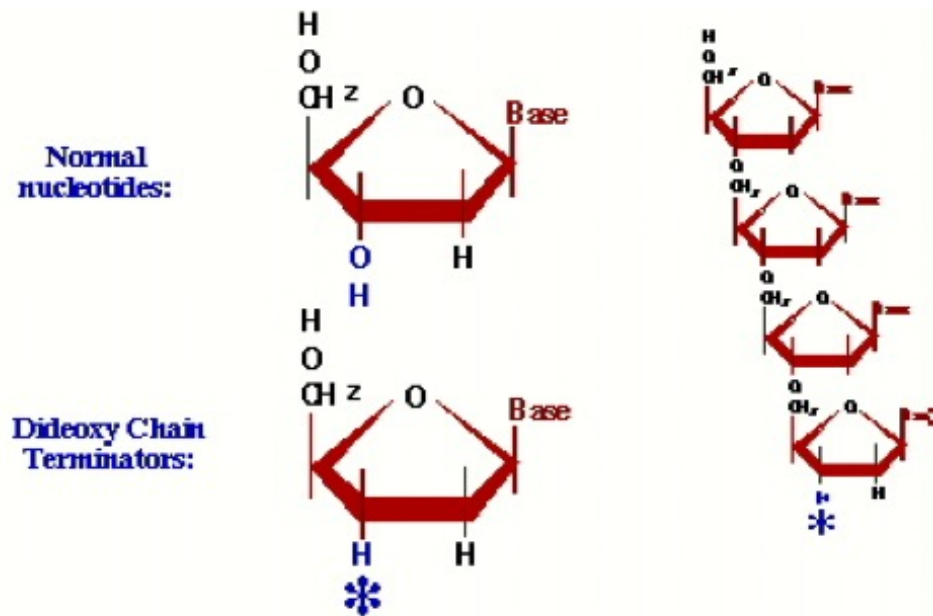


Metodo di Sanger

Il metodo di Sanger utilizza nucleotidi modificati chiamati **dideossinucleotidi trifosfato** (ddNTPs).

I ddNTPs sono deossiribonucleotidi sintetici che presentano lo zucchero deossiribosio privo del gruppo ossidrilico in posizione 3'.

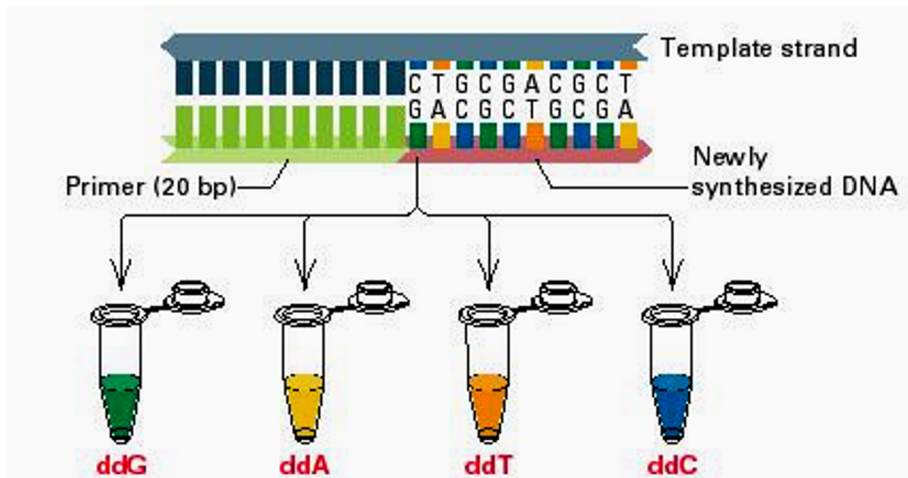
I ddNTP, a causa della loro struttura, impediscono che un altro nucleotide si leghi ad essi, in quanto non permettono la formazione del legame fosfodiesterico.



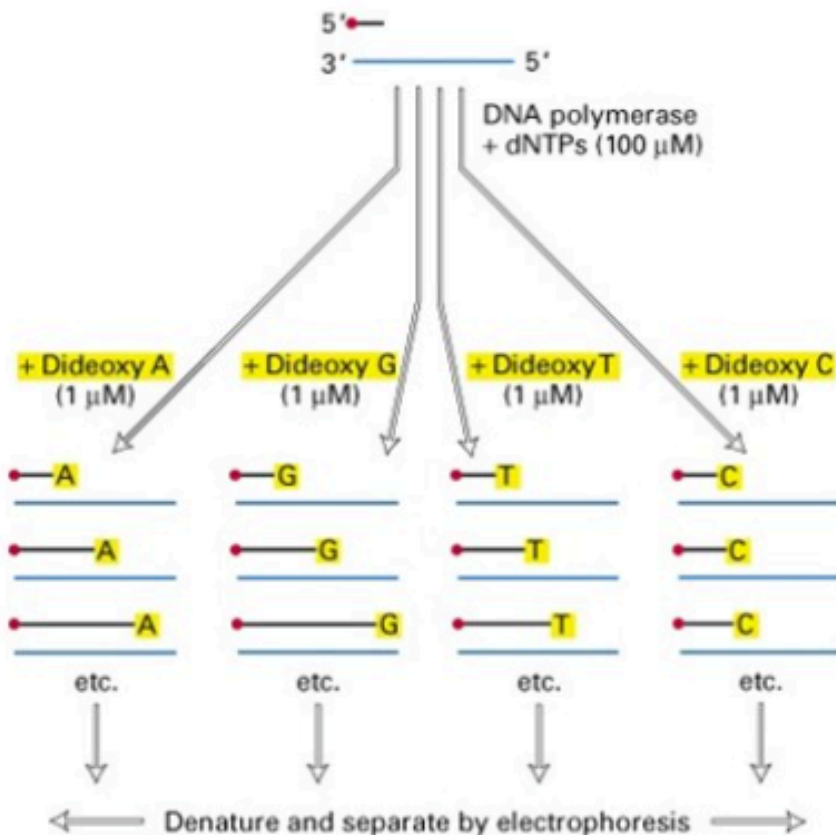
Per sequenziare un frammento di DNA (*filamento stampo*) il metodo di Sanger sfrutta la capacità della DNA polimerasi di copiare la sequenza e sintetizzare un nuovo filamento complementare.

Al procedere della reazione, vengono incorporati nel nuovo filamento sia dNTP normali, che permettono alla reazione di proseguire, sia ddNTP che bloccano la sintesi: al termine della reazione, si avrà una miscela di filamenti con lunghezza diversa, a seconda di dove è avvenuto l'inserimento del ddNTP terminatore di catena.

Metodo di Sanger

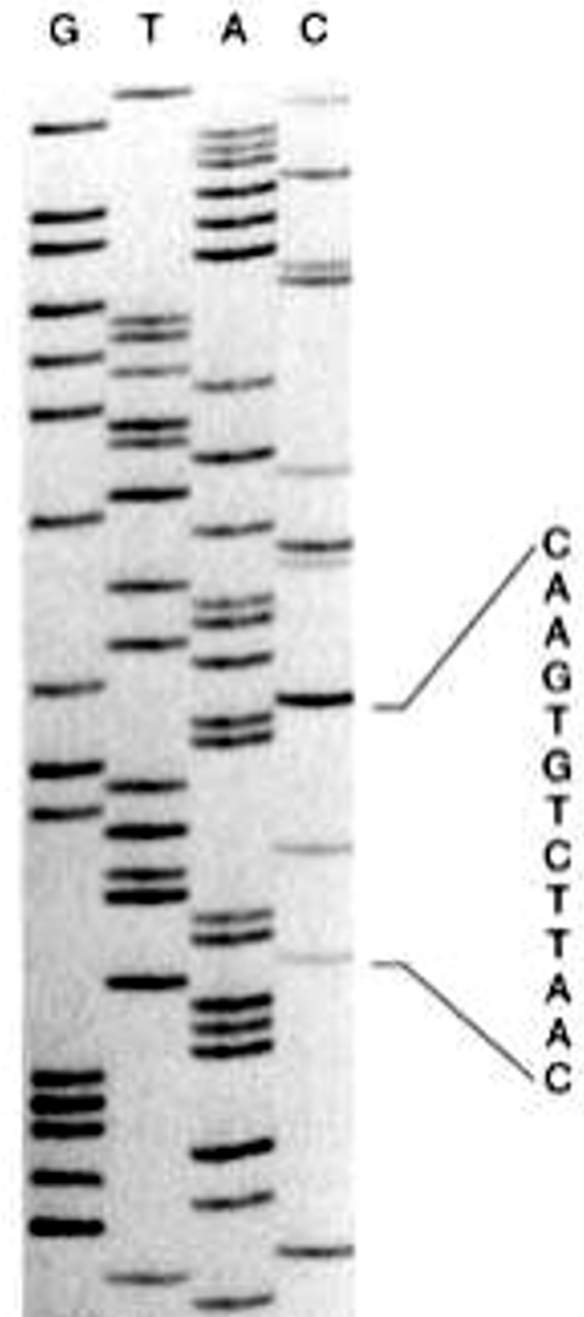
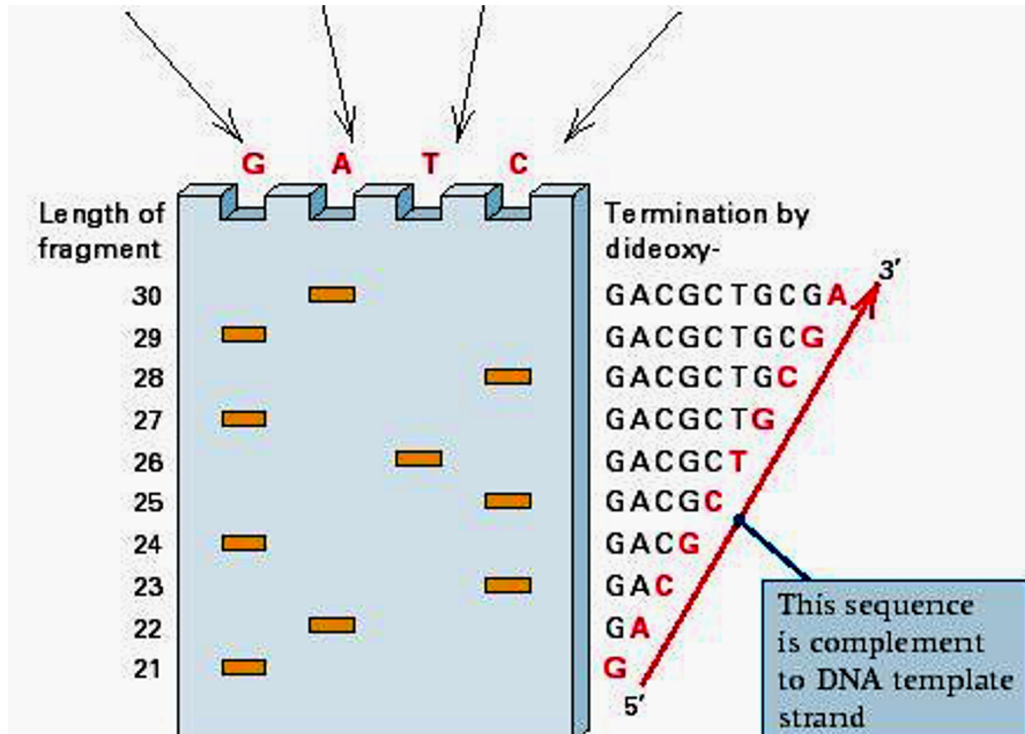
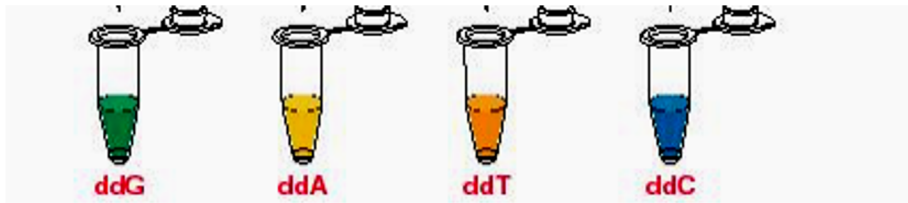


In parallelo 4 reazioni di polimerizzazione ciascuna delle quali contiene: lo stampo di DNA a singolo filamento, il primer, la DNA polimerasi, i 4 dNTP e un singolo ddNTP (100:1)



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'

Metodo di Sanger



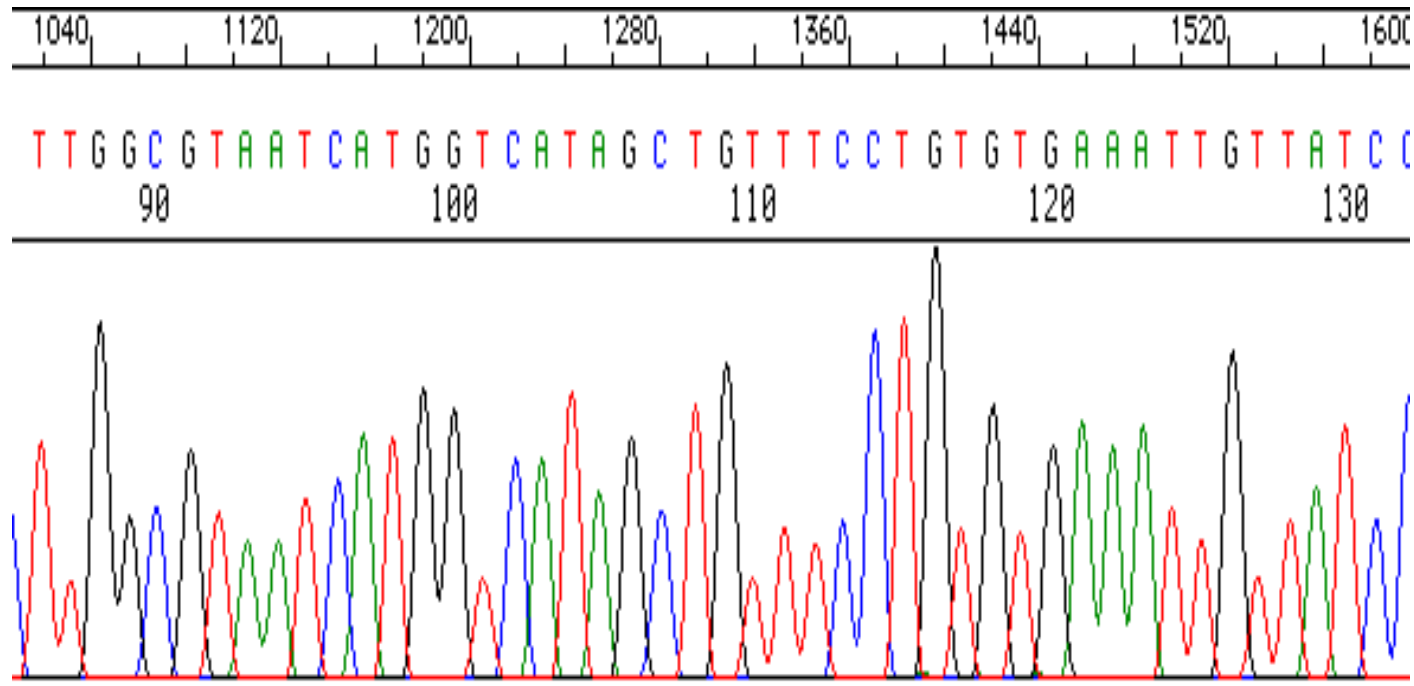
Grazie all'[elettroforesi su gel](#) i diversi frammenti vengono ordinati in base alla loro lunghezza: la disposizione dei frammenti svelerà la sequenza dei nucleotidi e la stringa del DNA stampo che si voleva sequenziare.

Un passo avanti: l'avvento dei sequenziatori automatici

Marcatura a fluorescenza dei frammenti di Sanger

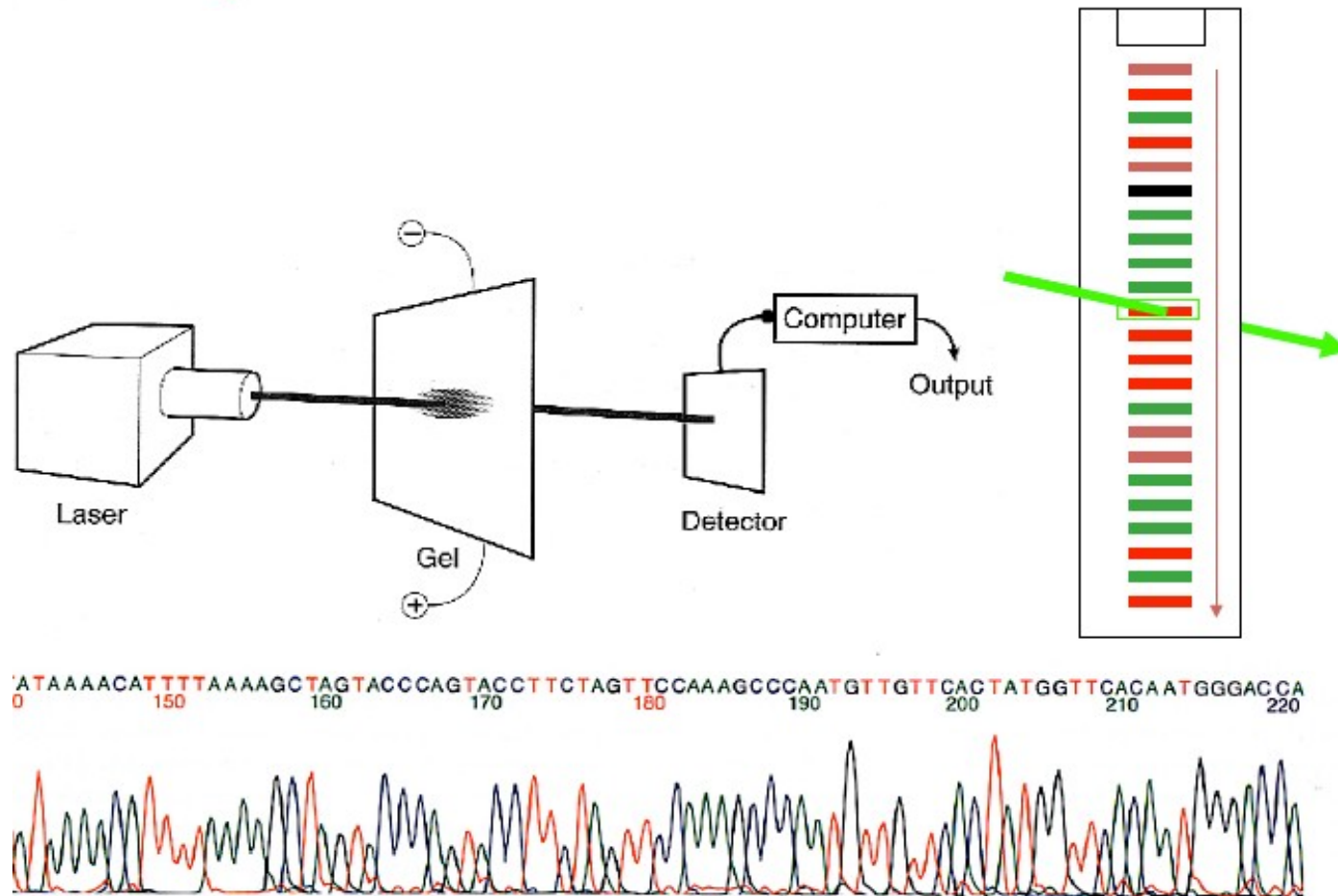


Sequenziamento automatico



Sequenziamento automatico del DNA

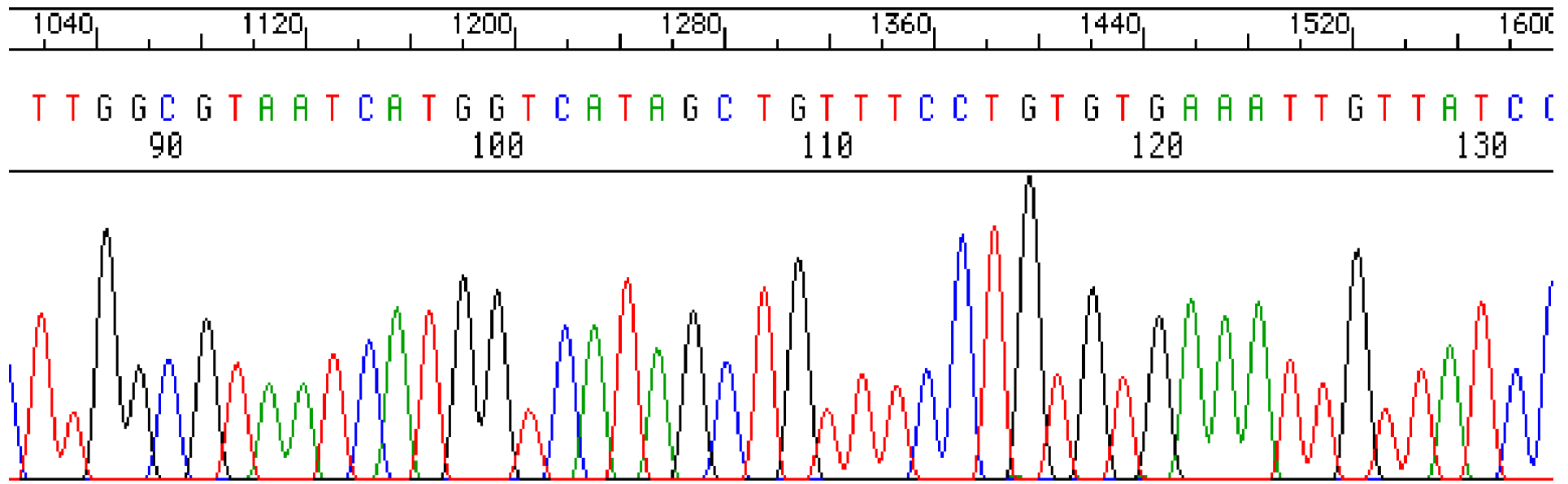
Ogni ddNTP è marcato con un fluorocromo diverso



- durante l'elettroforesi i singoli 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

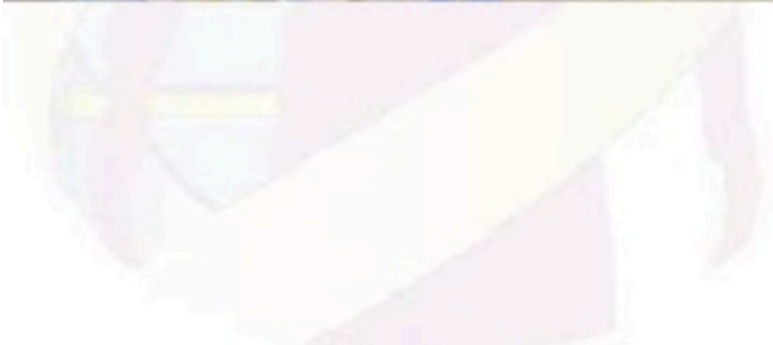
Sequenziamento automatico del DNA

Printout da un sequenziatore automatico

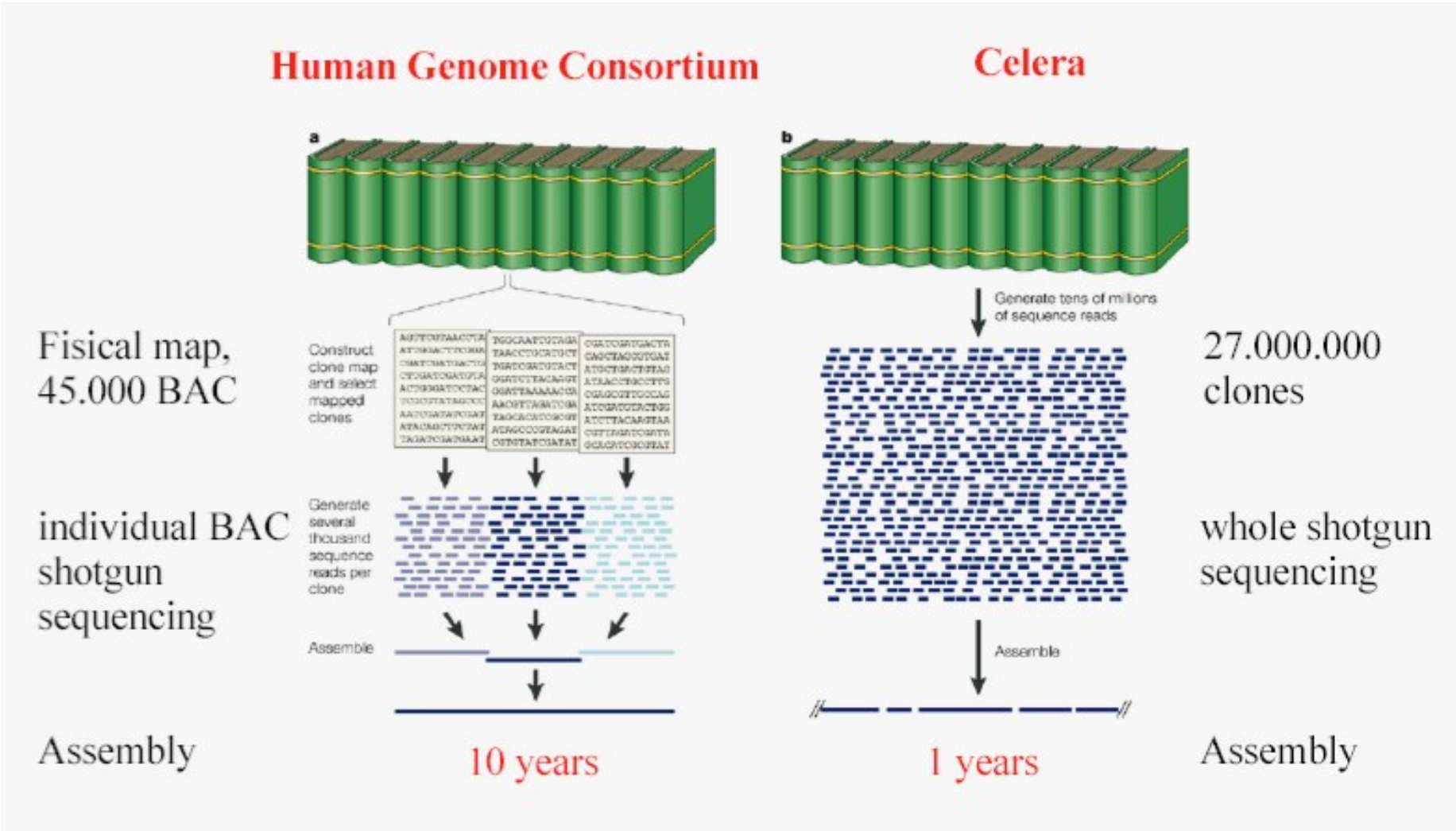


I colori sono generati dalla macchina (che riconosce i 4 fluorocromi associati ai 4 nucleotidi) e indicano le quattro basi: A verde, G nero, C blu, T rosso.

Lo sviluppo dei **sequenziatori automatici** capaci di sequenziare sino a 400.000 basi/giorno e gli enormi progressi delle **tecniche del DNA ricombinante** resero possibile la realizzazione del progetto genoma umano



Strategie per il sequenziamento del genoma umano



“clone by clone”

Whole genome shotgun

Approccio Utilizzato dalla Celera Genomics

Whole genome Shotgun Sequencing

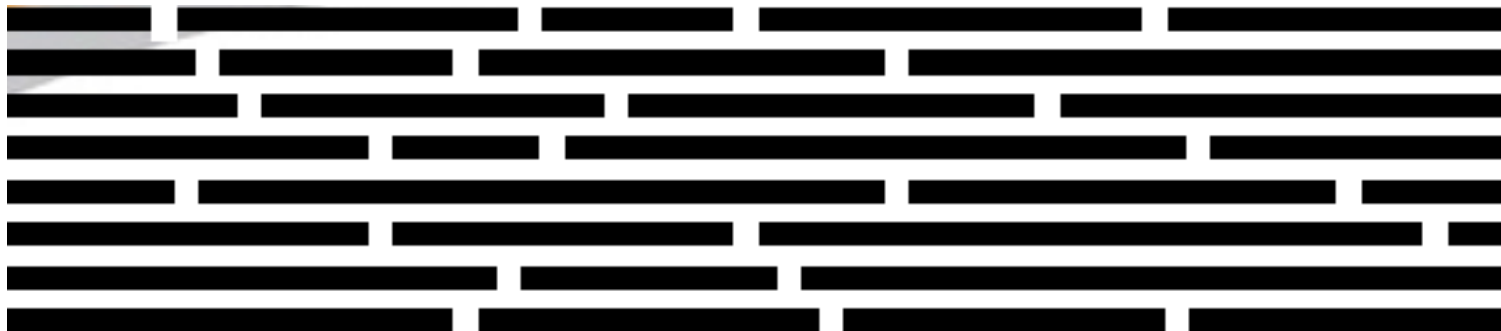
1. Frammentazione casuale del DNA in frammenti da 1.000-10.000pb
2. Clonaggio dei frammenti in vettore plasmidico
3. Sequenziamento dei frammenti nei sequenziatori automatici
4. Assemblaggio delle sequenze
5. Annotazione delle sequenze

1. Frammentazione casuale del DNA umano (frammenti da 1.000-10.000pb)

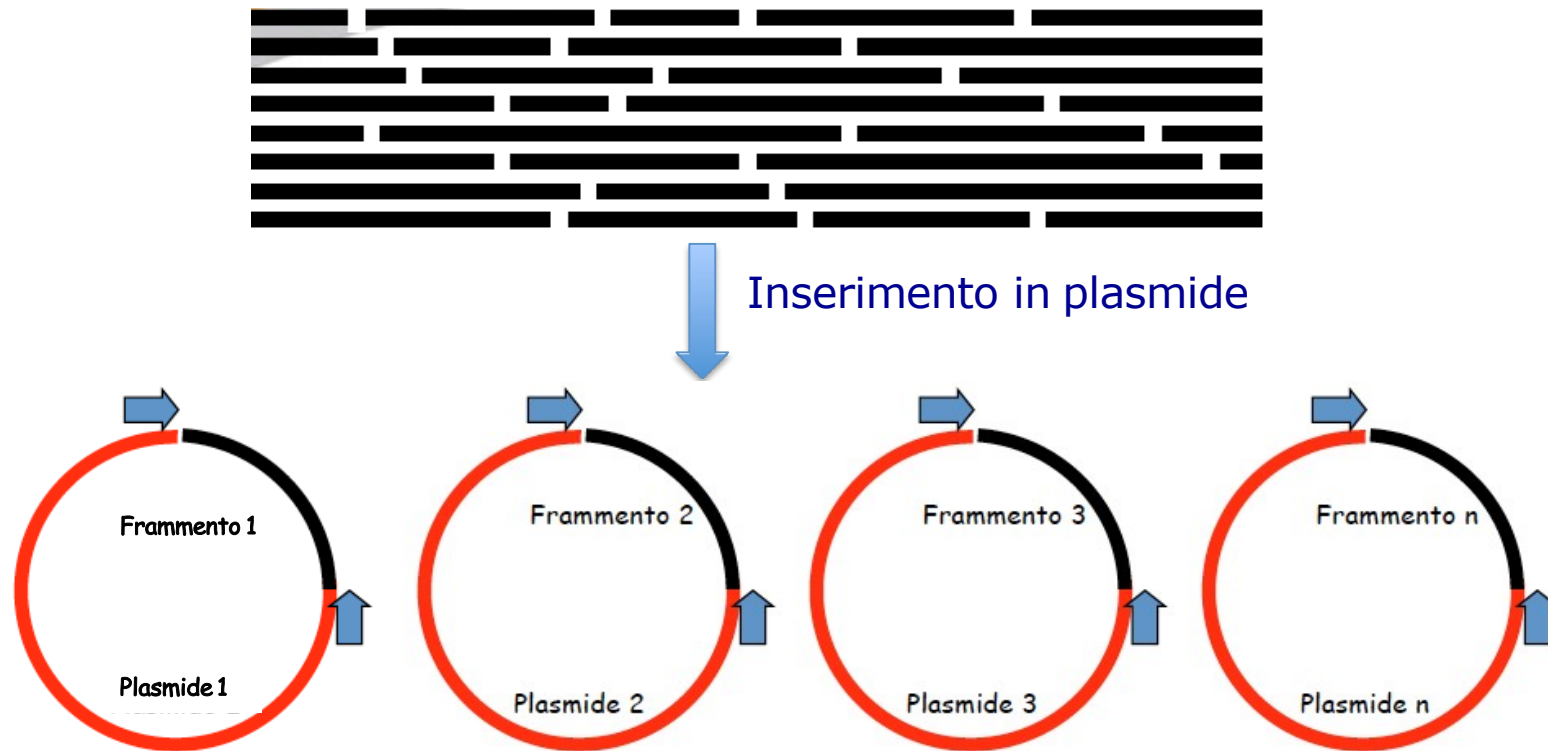
regione di DNA



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 CACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACAT
AATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATGGTAAATGTGACAGA
AAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATG

B ACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTTATAAACTTGAT
ATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTG
TCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTTGCATTCTAAAATGT

C GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTG
GCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGSTTTAAGAATAGTTTTTGTCTGTACTTT
CTATAGTGAATA

D ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT
GATGATCTGTAGTGCTACAGAAAATTTGTGGGTCACAGTCTATTATGGG GTACCTGTGTGGAAGGAAGCAACCA
CCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTG

E CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG
AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC
TTGGGTTCTTGGGAGCAGCAGGA

F GGAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
AAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGAAGAATGATACTAATAACCAATAGTAGTAGCGG
GAGAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTG

Frammenti che contengono sequenze sovrapposte sono riuniti in un unico frammento

D+ A

ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT
 GATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA
 CCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT
 GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAAACATGTGGAAAAATGA
 CATGGTAGAACAGATG

B

ACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTTATAAACTTGAT
 ATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTG
 TCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGT

C

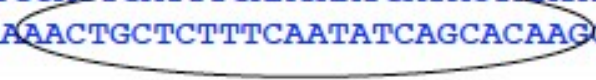
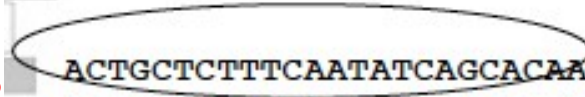
GAAAAGAATGAACAAGAAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTG
 GCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTTGCTGTACTTT
 CTATAGTGAATA

E

CATCAAATATTACAGGGCTGCTATTAACAACAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT
 GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG
 AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTC
 TTGGGTTCTTGGGAGCAGCAGGA

F

GGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
 AAATTAACCCCACTCTGTGTTAGTTTAAAGTGCACTGATTTGAAGAATGATAGTAATAACCAATAGTAGTAGCGG
 GAGAATGATAATGGAGAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTG



D + A

ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA
CCTCTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT
GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAATTTTAACATGTGGAAAAATGA
CATGGTAGAACAGATG

C

GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTG
GCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTTGTCTGACTTT
CTATAGTGAATA

E

CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG
AGTAGCACCCACCAAGGCAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC

F + B

TTGGGTTCTTGGGAGCAGCAGGA
GGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
AAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGAAGAATGATACTAATAACCAATAGTAGTAGCGG
GAGAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGC
AGAAAGAATATGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACA

D+A+F+B

ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA
CCTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT
GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAAATGA
CATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCC
CACTCTGTGTTAGTTTAAAGTGCACTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGAGAATGATA
ATGGAGAAAGGAGAGATAAAAACTGCTCTTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATA
TGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACA
CCTCAGTCATTACACAGGCCGTGCCAAAGGTATCCTTTGAGCCAATCCCATACATTATTGTGCCCCGGCTGGT
TTTGCGATTCTAAAATGT

C

GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTG
GCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTTGTCTGTACTTT
CTATAGTGAATA

E

CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG
AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC
TTGGGTCTTGGGAGCAGCAGGA

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

Analisi dei dati di sequenziamento

4. Annotazione delle sequenze genomiche

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

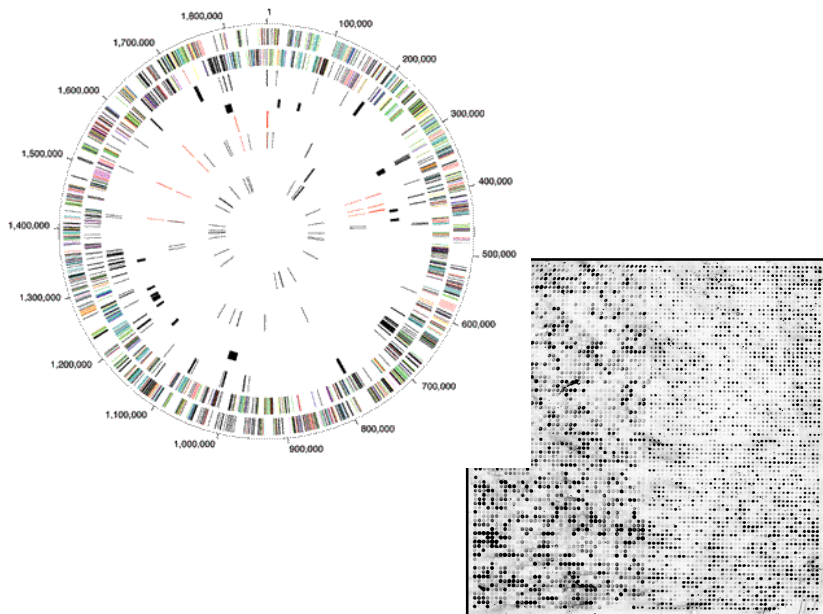
Data la sequenza completa del genoma, l'**annotazione** permette di rispondere a queste domande :

- 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

La Bioinformatica

La necessità di gestire ed interpretare le grandi quantità di informazioni derivanti dal sequenziamento dei genomi 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:

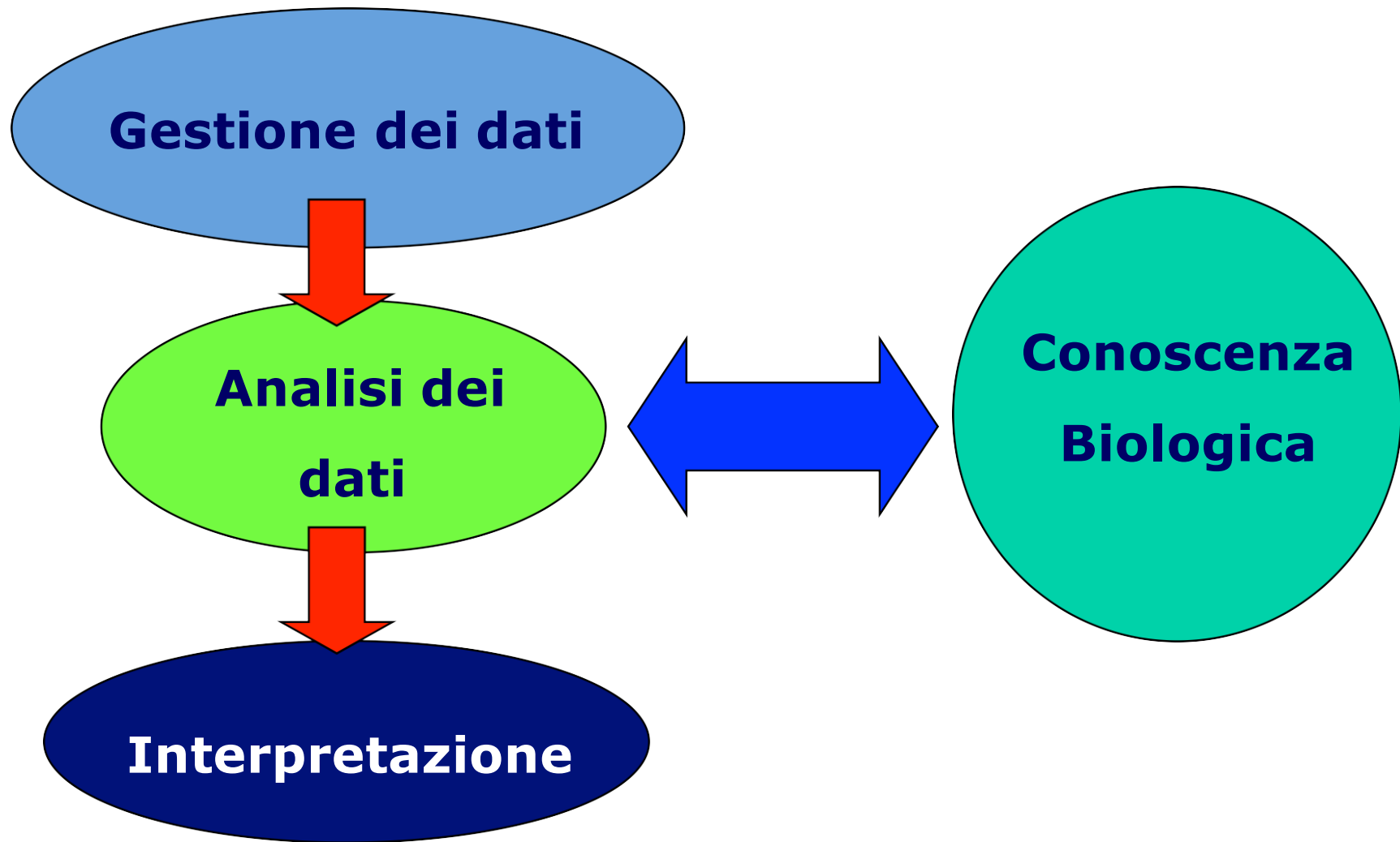
- collezionare i dati biologici in apposite **banche dati**,
- sviluppare **strumenti idonei** per l'interrogazione dei dati biologici;
- sviluppare **metodologie** (algoritmi e software) per l'analisi dei dati biologici.



Comprensione dei meccanismi molecolari alla base della vita

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

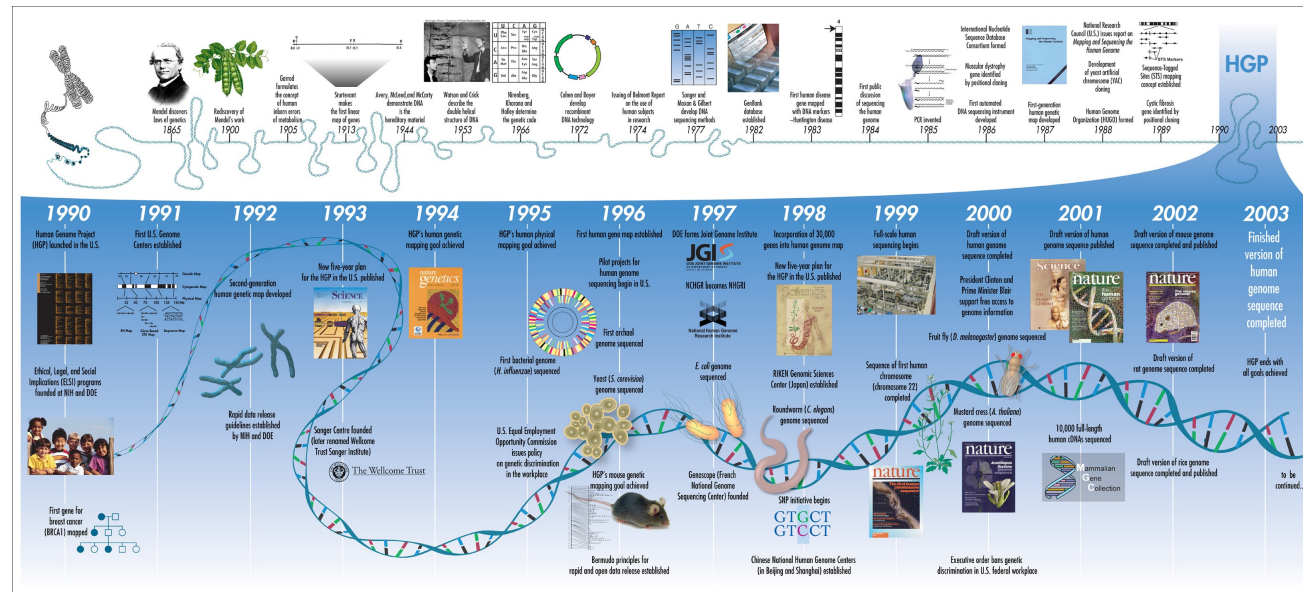
Il ruolo della Bioinformatica



Il sequenziamento del Genoma Umano oggi



1953 – DNA structure



Human Genome Project – 13 years - > 3 billion dollars



Next generation sequencing technologies

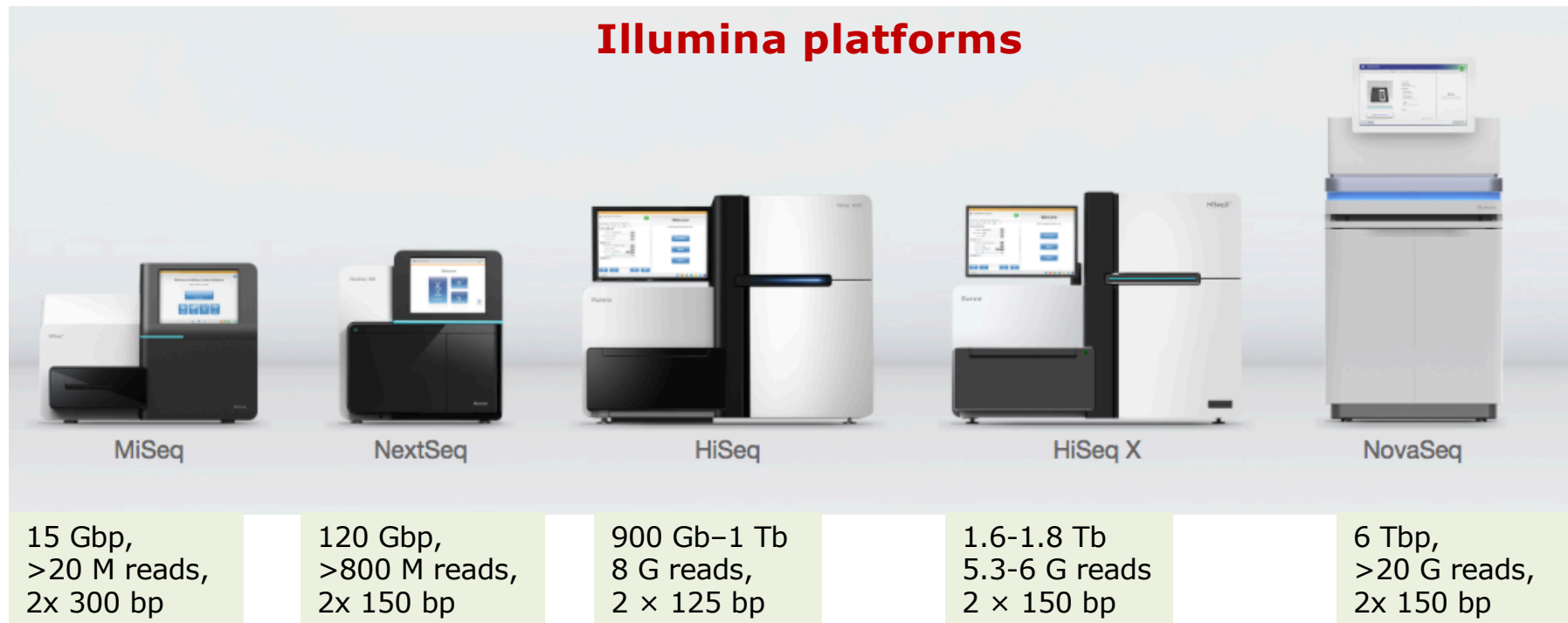


Human genome - 2019 – few days - 1000 dollars !!

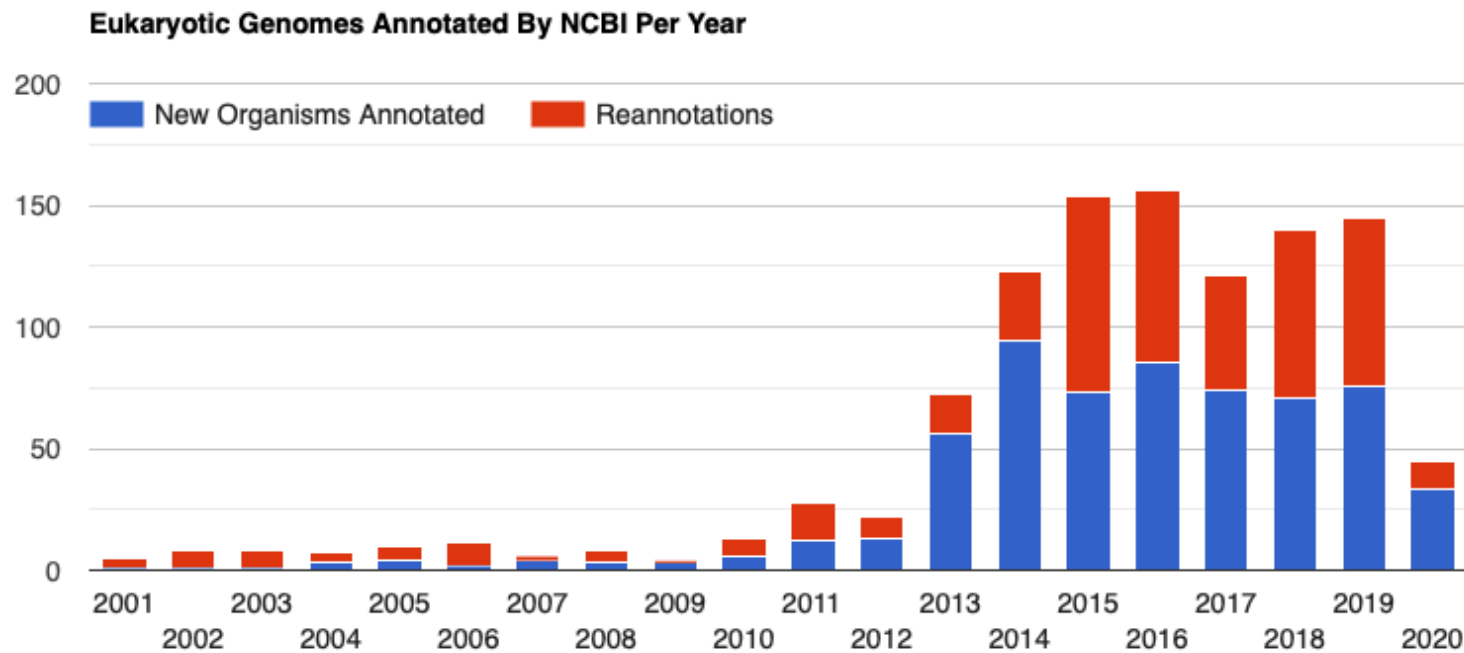
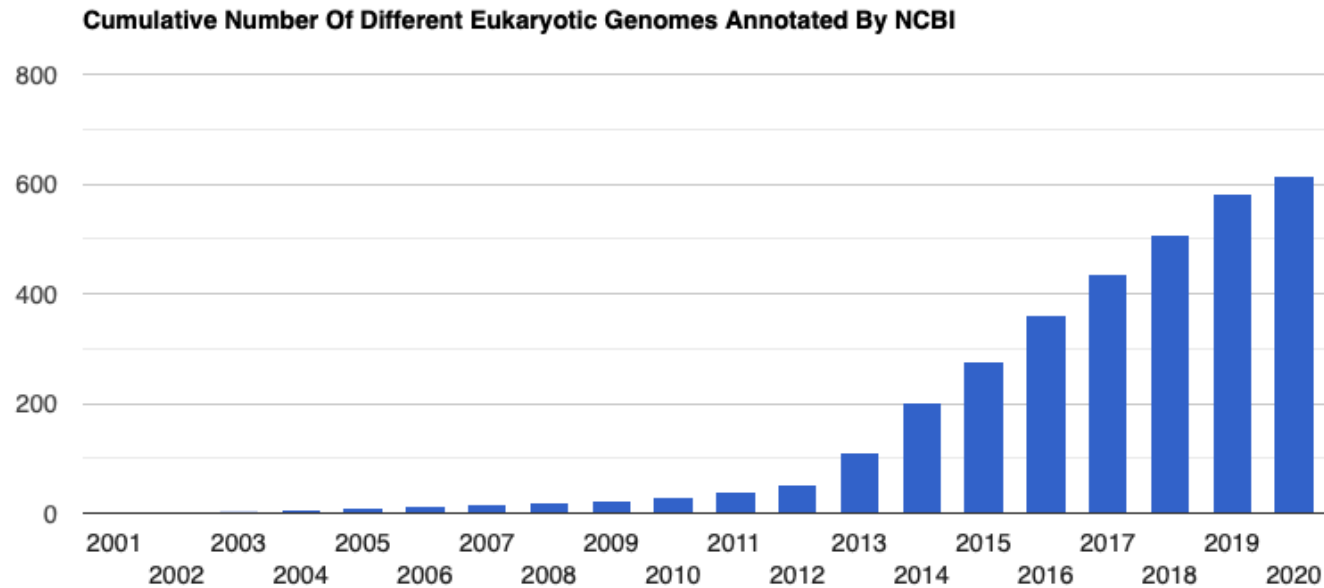
Il Sequenziamentooggi

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.



Genomi eucariotici completi già annotati



E una volta che i genomi sono stati sequenziati?

E' iniziata l'era Post-Genomica

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

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

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

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

La Rivoluzione “OMICA” in Biologia

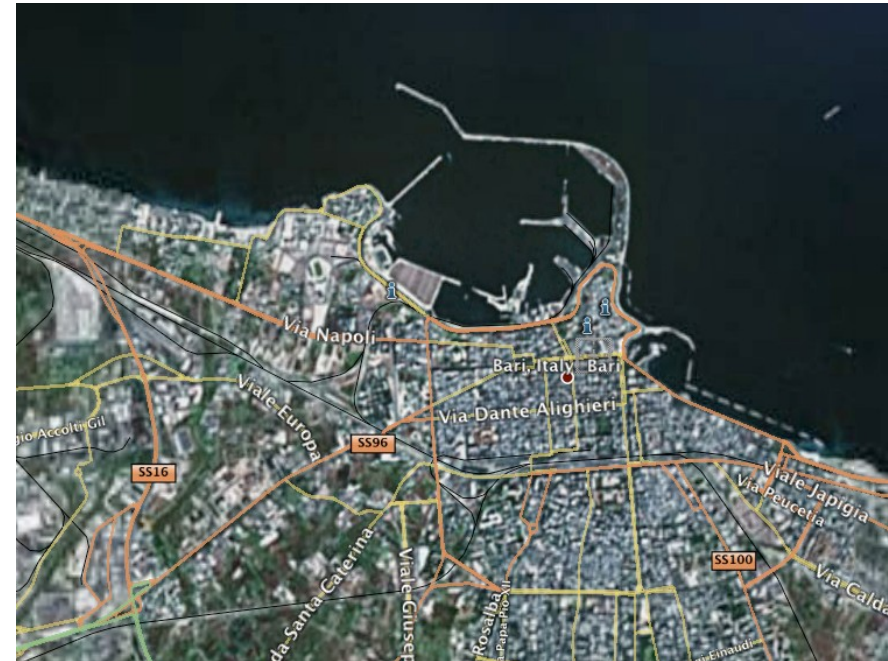
- GENOMICA
Studio dei genomi
- TRASCRITTOMICA
analisi del trascrittoma
- PROTEOMICA
analisi del proteoma
- METABOLOMICA
analisi dei profili metabolici
- FARMACOGENOMICA
relazione tra genoma e risposta ai farmaci
- FISIOMICA
Fisiologia dell'intero organismo

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.

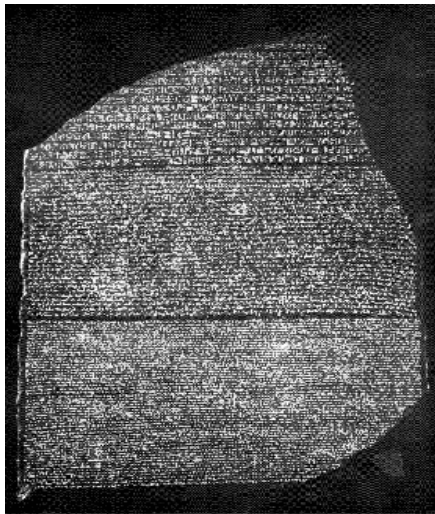


Genomica Funzionale. Studio delle funzioni dei geni, delle loro interazioni (pathways metabolici) e dei meccanismi che ne regolano l'espressione.

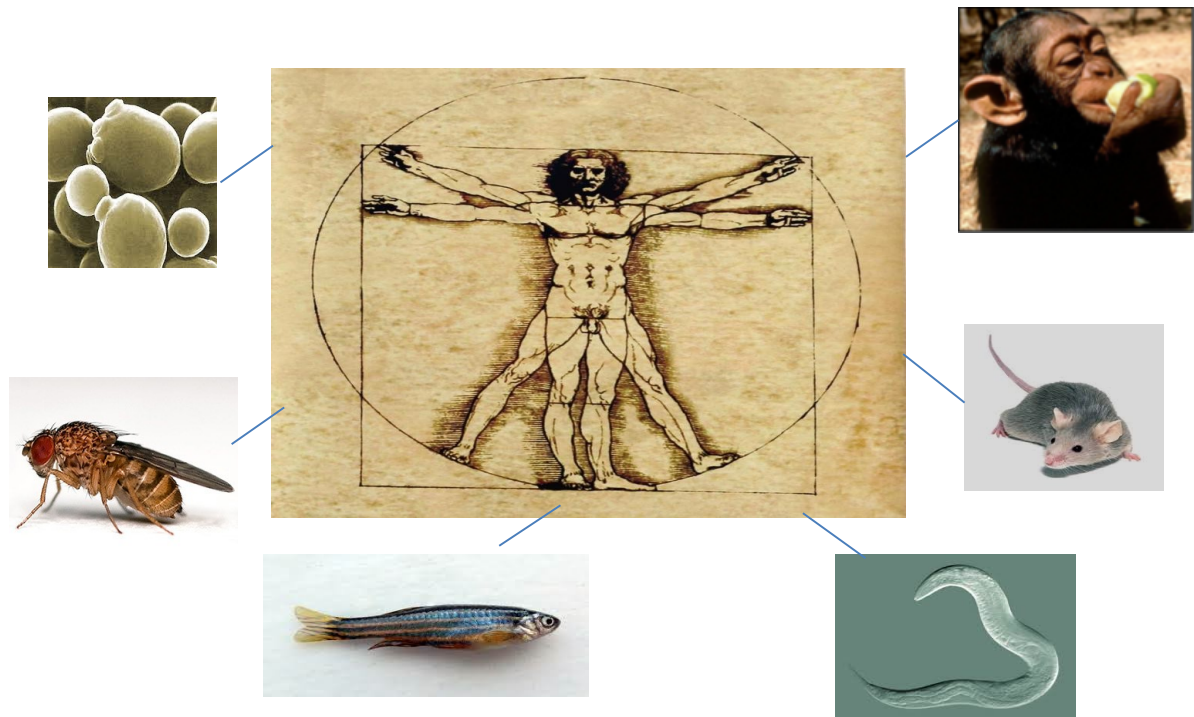
Genomica Comparata

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

BLAST



Stele di Rosetta
(British Museum)



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 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

Tools

[All tools](#)

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

e.g. **BRCA2** or rat 5:62797383-63627669 or rs699 or coronary heart disease

si sceglie il genoma d'interesse

gene o regione cromosomica da visualizzare

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genes



Human
GRCh38

[Still using GRCh37?](#)



Mouse
GRCm38.p6



Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 92 (April 2018)

- [New goat annotation on the ARS1 assembly](#)
- [Update of Marmoset assembly and genebuild](#)
- [Mouse: update to Ensembl-Havana GENCODE gene set](#)
- [Update to Ensembl-Havana human GENCODE gene set \(release 28\)](#)
- [New command line tool for LD](#)

[Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

- 10 Apr 2018: [Do you use transcripts for your work?](#)
- 05 Apr 2018: [Ensembl 92 has been released!](#)
- 23 Mar 2018: [2018 – a year of conferences](#)

[Go to Ensembl blog](#)

Compare genes across species

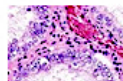


Find SNPs and other variants for my gene

```

GTTTATACATTC
CRTRAAAGTCTT
CTTCTAAATTCT
GTAACATTTTCC
    
```

Gene expression in different tissues



Retrieve gene sequence

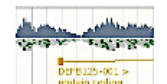
```

GCGTGACTTCGGGTGG;
GGGCTTGTGGGGGAGCC;
GGGCTCTGCTGGGCTT;
AGGGACAGATTTGTGA;
CACCTCTGGAGCGGGTT;
CCGAGTCCAGCGTGGCG;
    
```

Find a Data Display



Use my own data in Ensembl



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

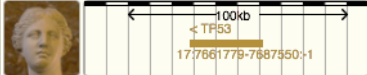
TP53-028 (Human Transcript)
ENST00000619485 17:7668421-7687487:-1
Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]
R-HSA-6811555 (Reactome record; description: PI5P Regulates **TP53** Acetylation) is an external reference matched to Translation ENSP00000482537

Best gene match

Human Gene

TP53

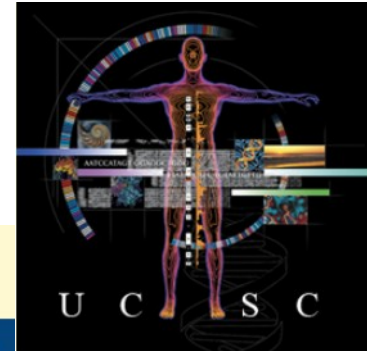
HGNC Symbol; Acc:HGNC:11998



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 tpit tpo
- tpor tpox tpte cfap53 irsp53 tpi1p3 wrap53 dutp5 fatp5 setp5
- tnip3 tnni3 tnp01 tnp02 top2a top2b top3a top3b tor3a torc3
- tpar1 tparam tpbgl tpkil tpra1 tpte2 traf3 traj3 trar3 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](#)[VisiGene](#)[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 [Neanderthal](#) 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 sequences via the [Genome Browser](#) interface. You can zoom in and scroll over chromosomes, showing the work of annotators. The [Table Browser](#) provides convenient access to information on groups of genes that can be related to other genomic features. The [Table Browser](#) also allows you to browse through a large collection of *in situ* mouse and frog images to visualize gene expression patterns. The [Table Browser](#) also allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the [Genome Bioinformatics 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!

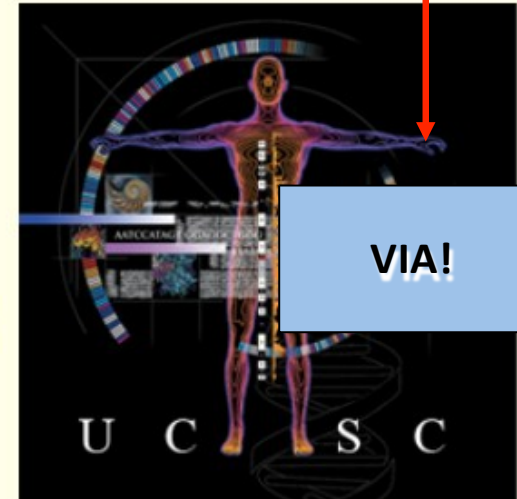
genome. See the [User's Guide](#) for more information.

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

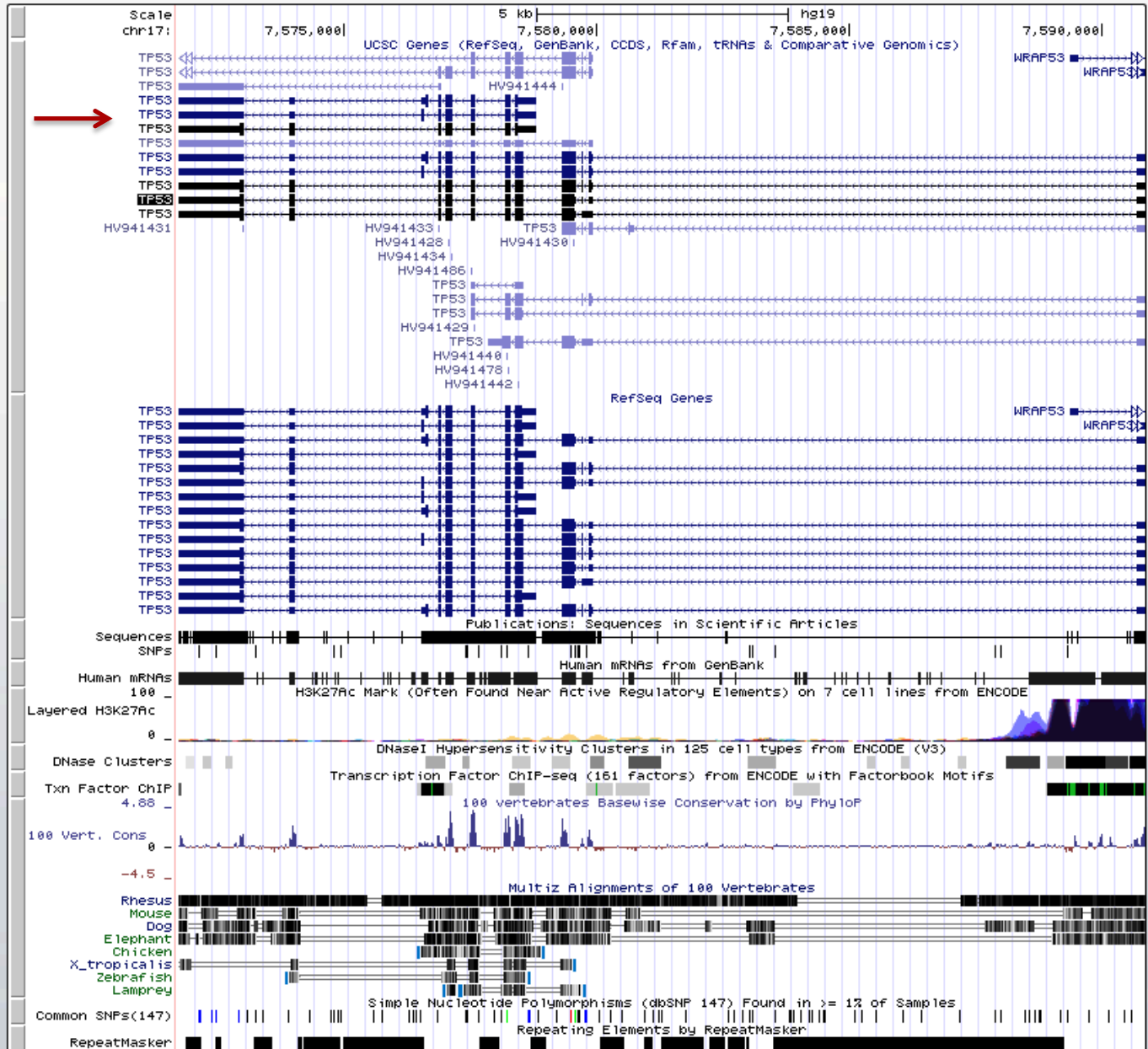
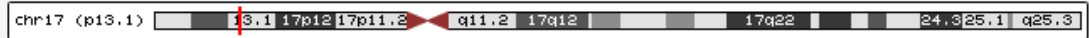


Homo sapiens
(Graphic courtesy of [CBSE](#))

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

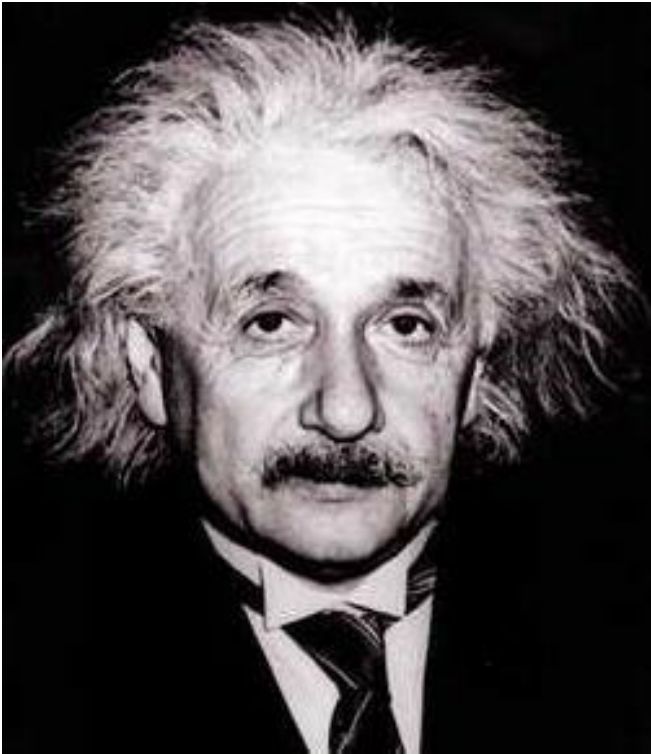
chr17:7,571,720-7,590,868 19,149 bp. TP53 (Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA.) go



Quali informazioni ha fornito il sequenziamento dei genomi?

La grandezza del genoma non è correlata alla complessità dell'organismo

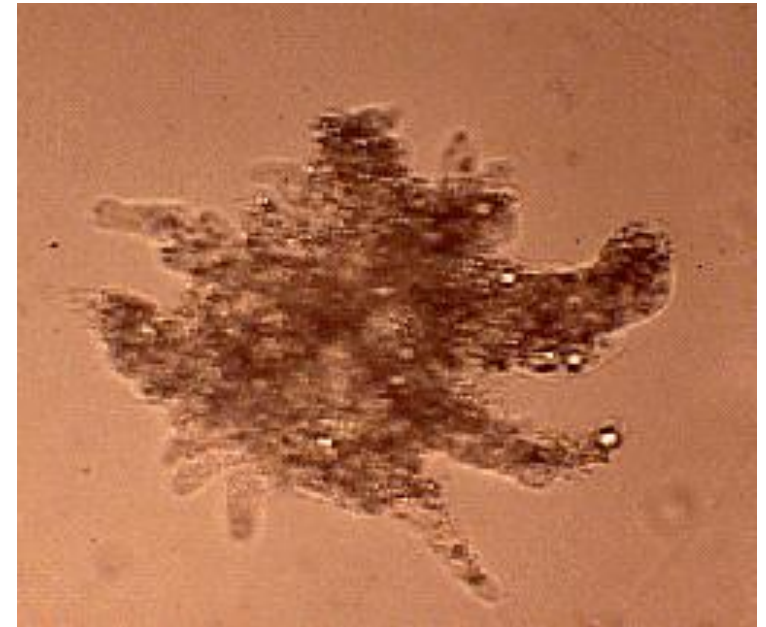
Paradosso del valore C



Homo sapiens 3×10^9 pb

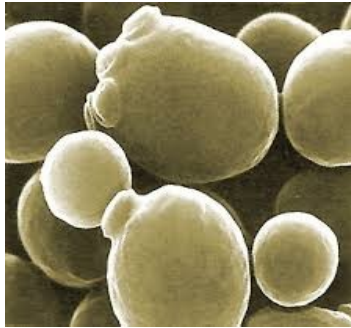


Allium cepa 1.5×10^{10} pb

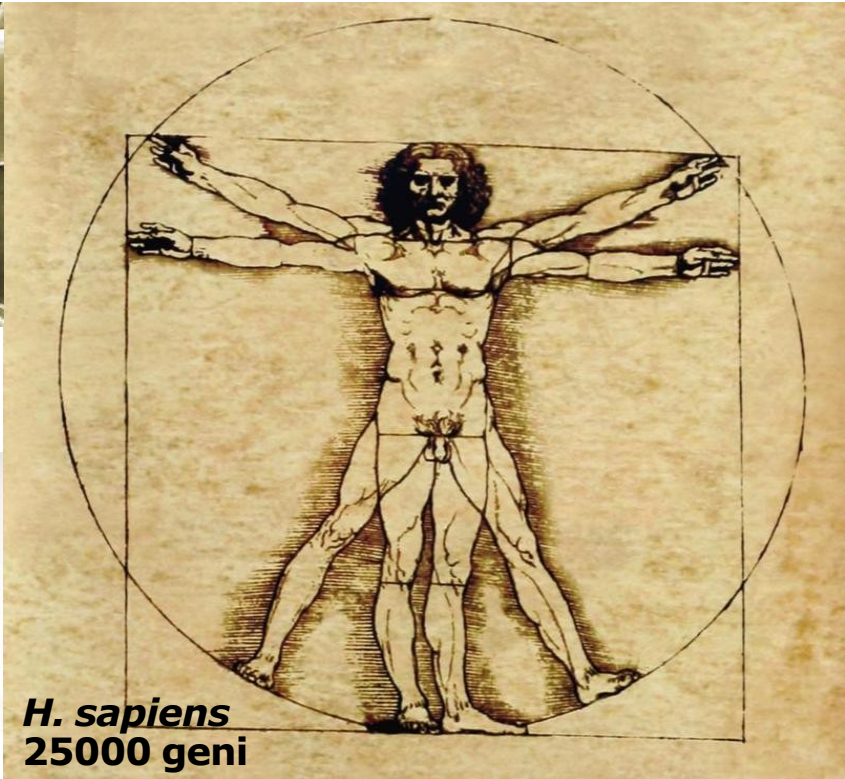


Amoeba dubia 6.7×10^{11} pb

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

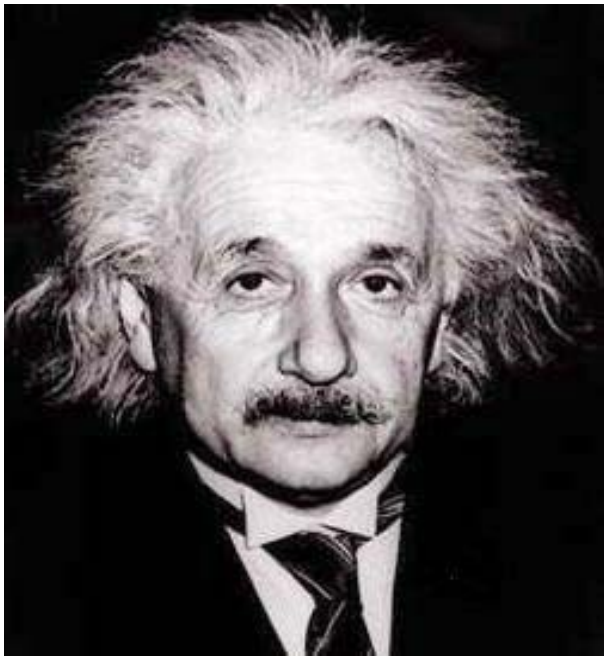


Z. mays
26000 geni

Quindi:

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

Paradosso del valore N



~25.000 geni



~22.000 geni



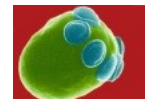
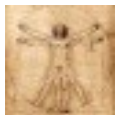
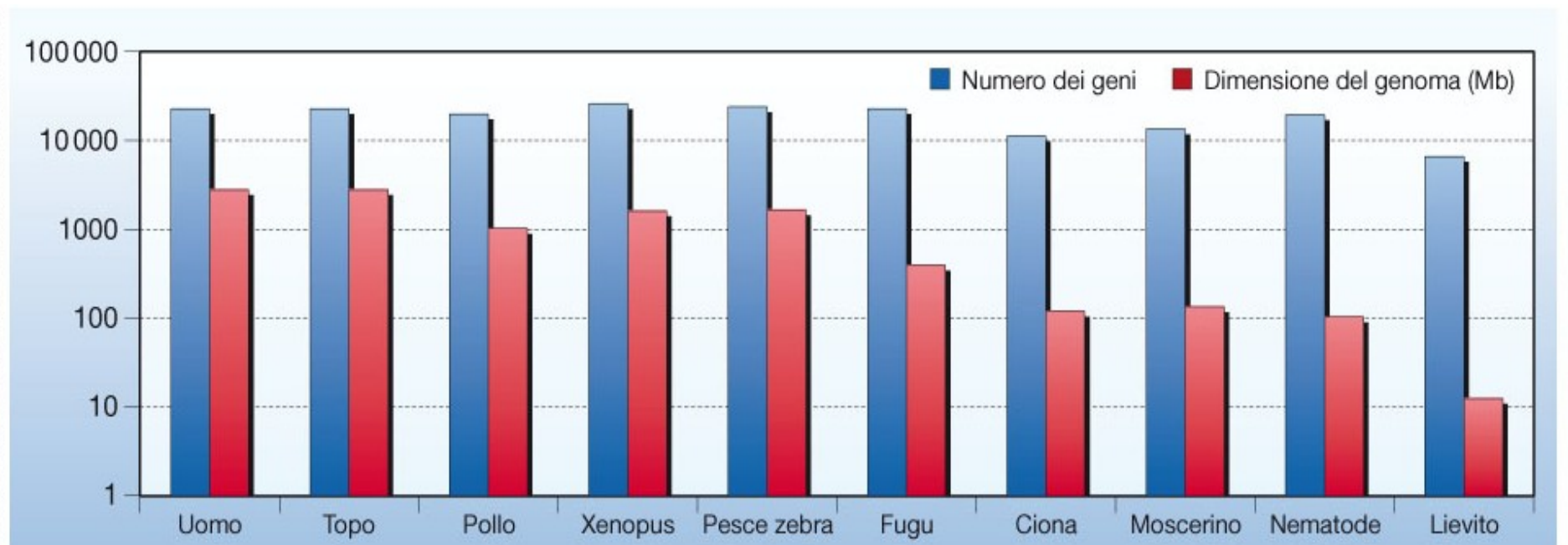
~25.000 geni



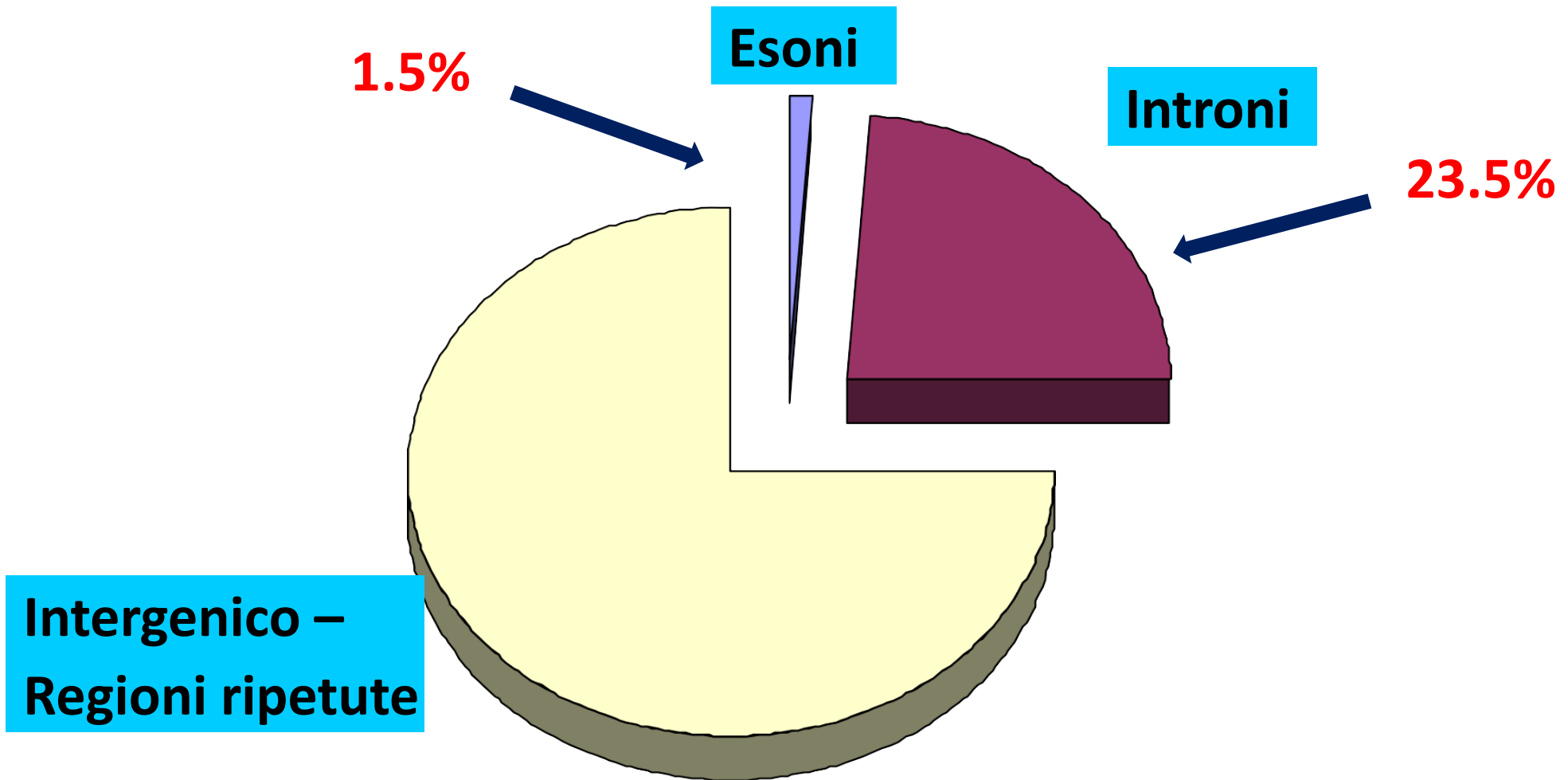
~60.000 geni

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)



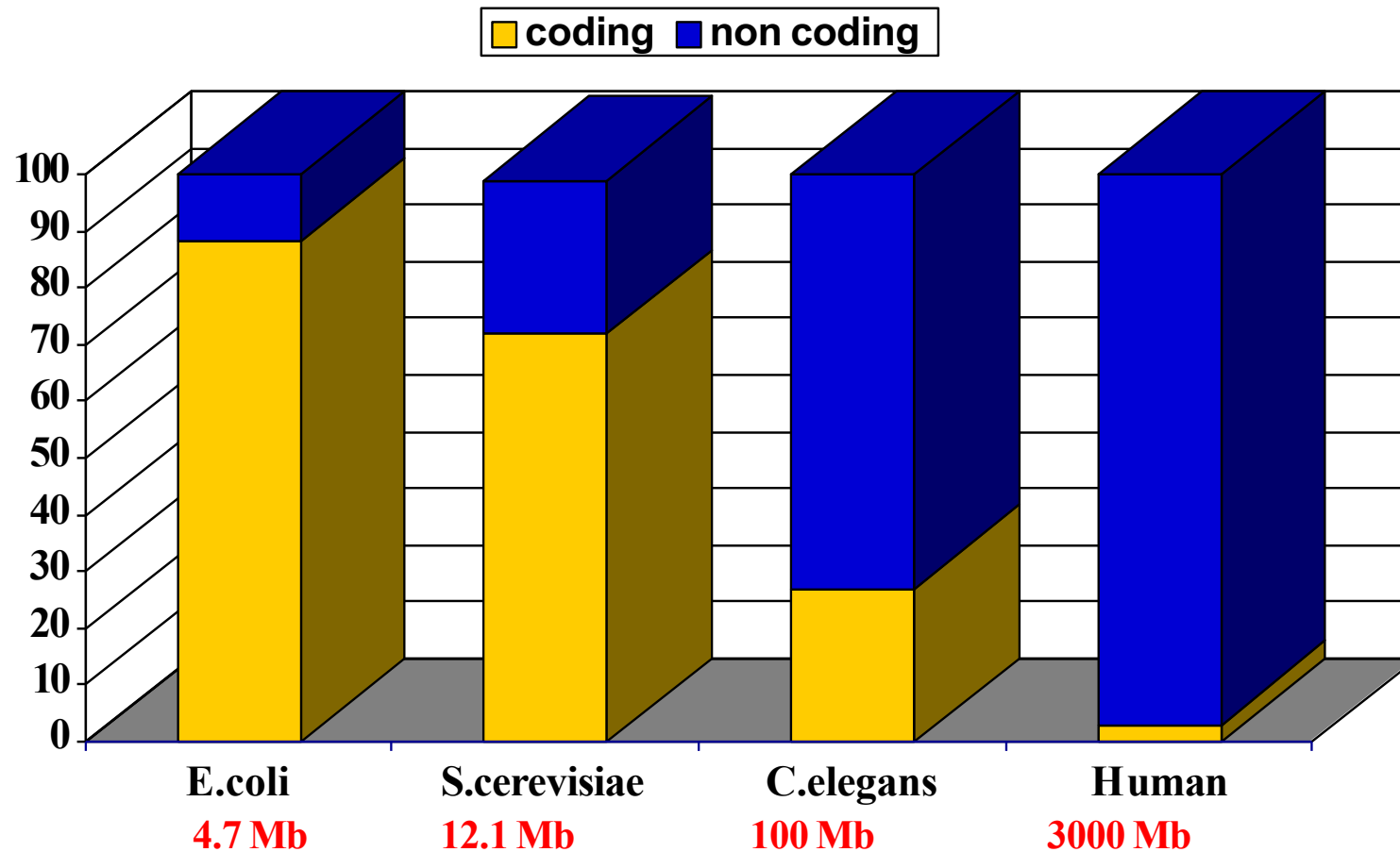
Il Contenuto del genoma umano



Il genoma è vuoto??

La porzione non codificante del genoma umano è >98%

La sfida post-genomica



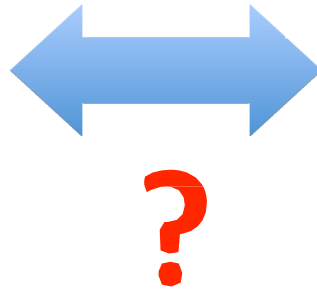
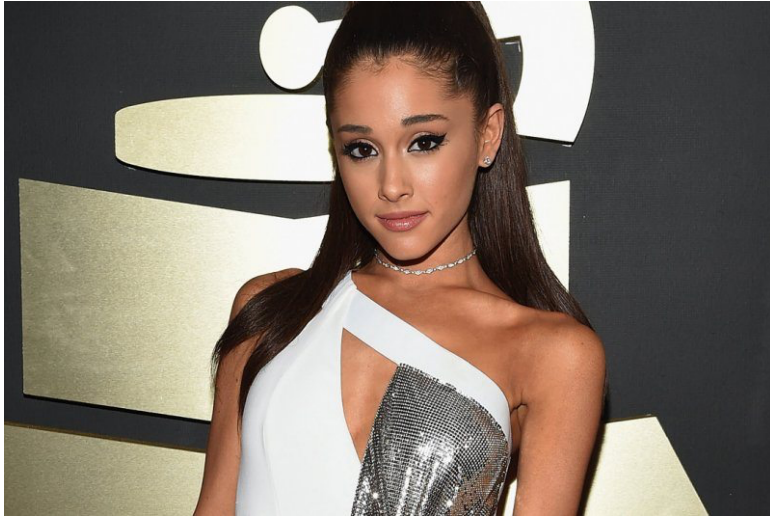
L'annotazione funzionale delle porzioni non-codificanti del genoma è una delle sfide principali dell'era post-genomica.

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

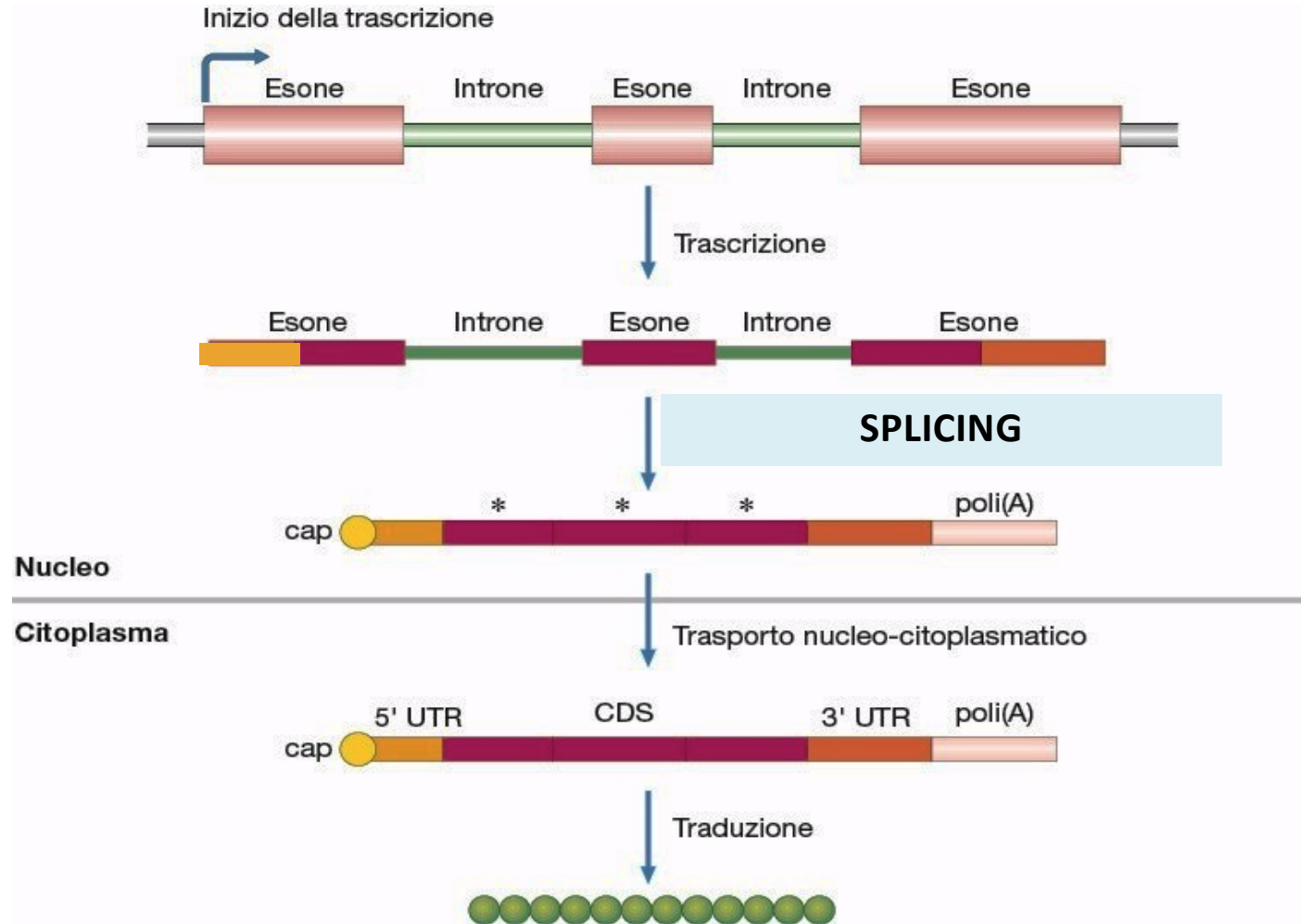
Da cosa dipende la differenza?



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

Da cosa dipendono le differenze?

Lo splicing



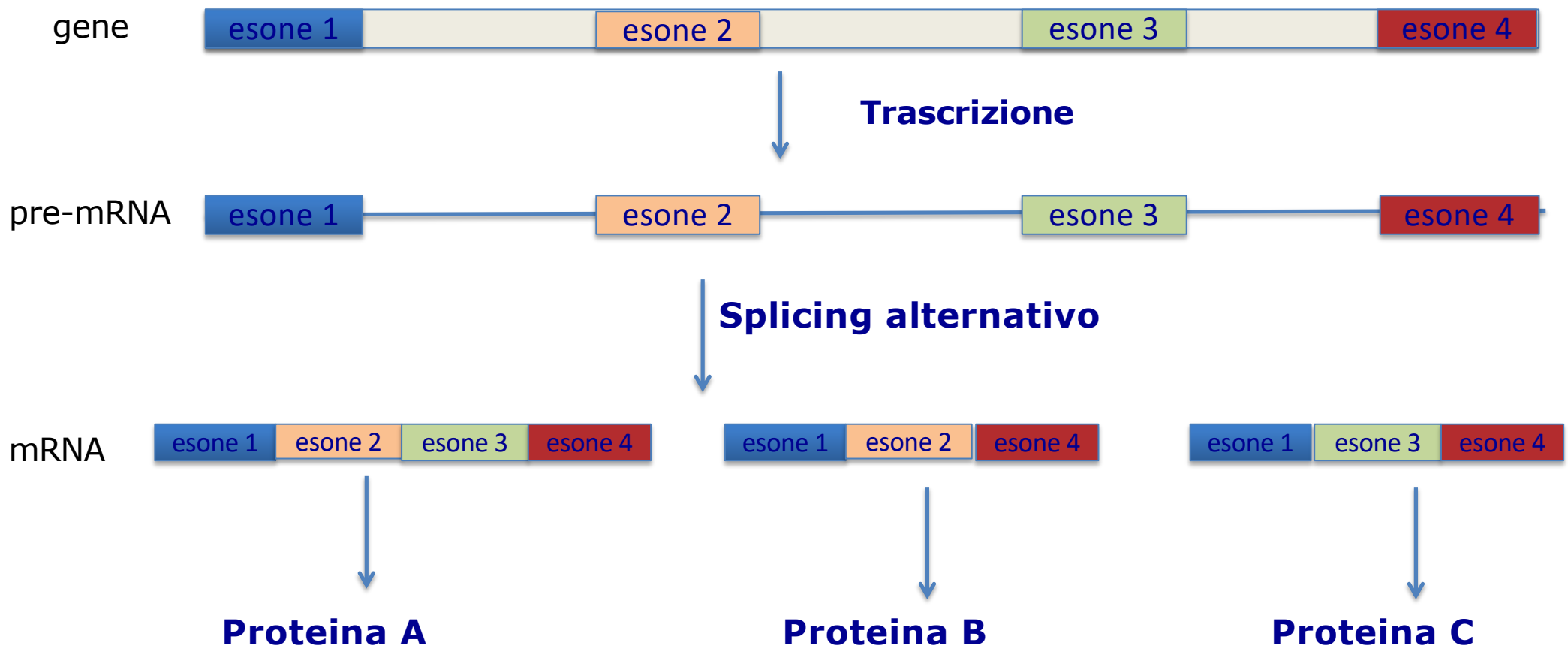
I geni sono espressi sotto forma di lunghi **RNA precursori** che nel nucleo sono sottoposti ad un processo di maturazione che comporta la rimozione di tratti polinucleotidici interni chiamati **introni** e la concatenazione delle restanti regioni chiamate **esoni** (**meccanismo di splicing**).

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

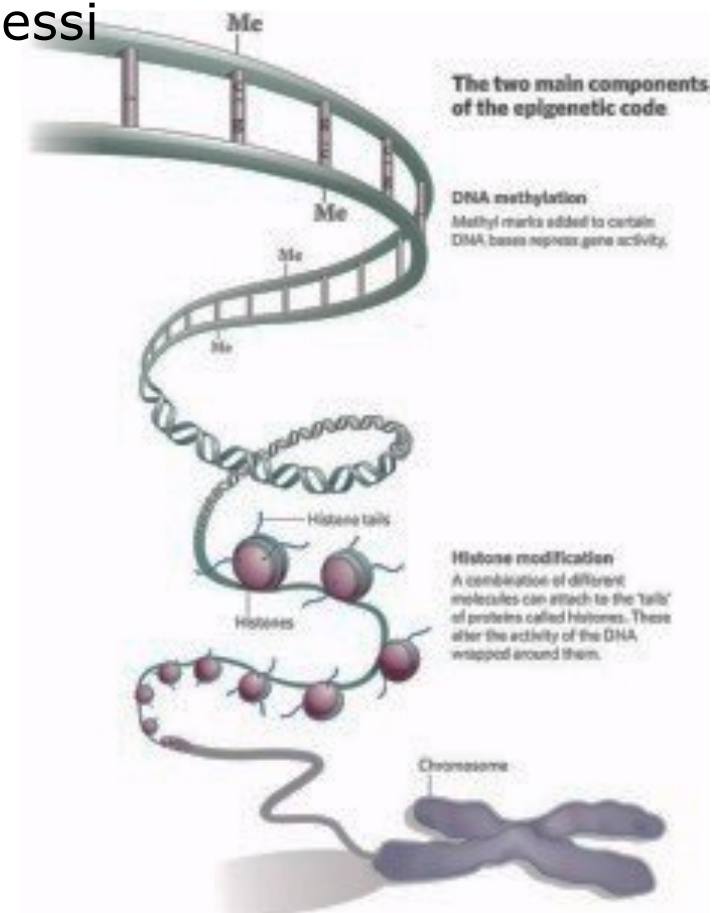
La **regolazione epigenetica** indica un meccanismo di controllo dell'espressione dei geni che non è dipendente dalla sequenza del DNA.

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.

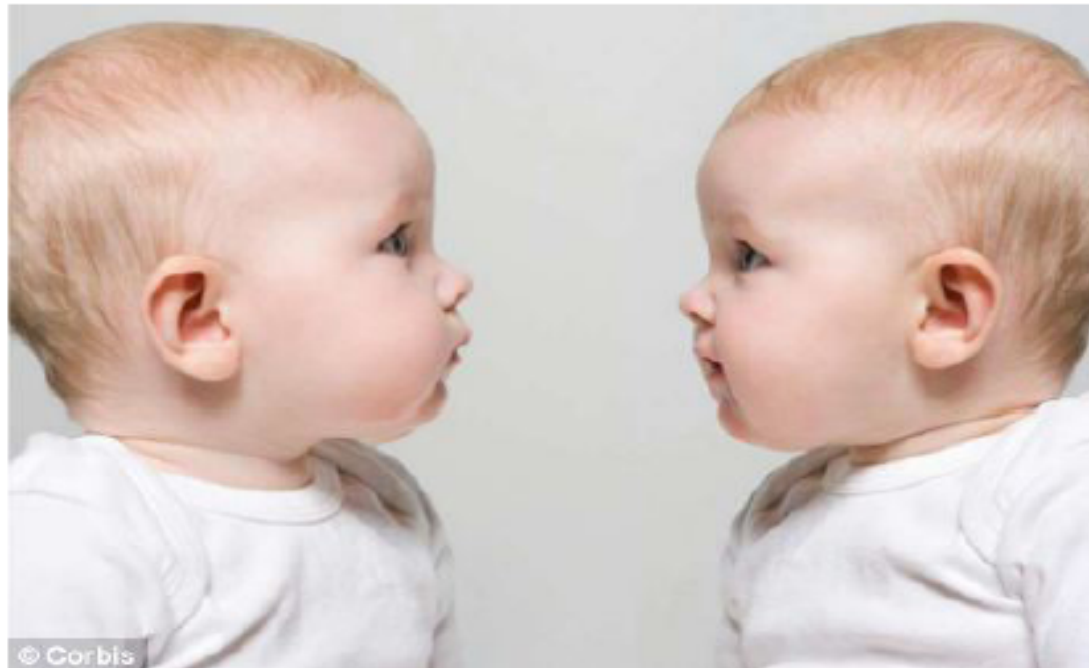
I principali meccanismi alla base dell'ereditarietà epigenetica sono:

- 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.

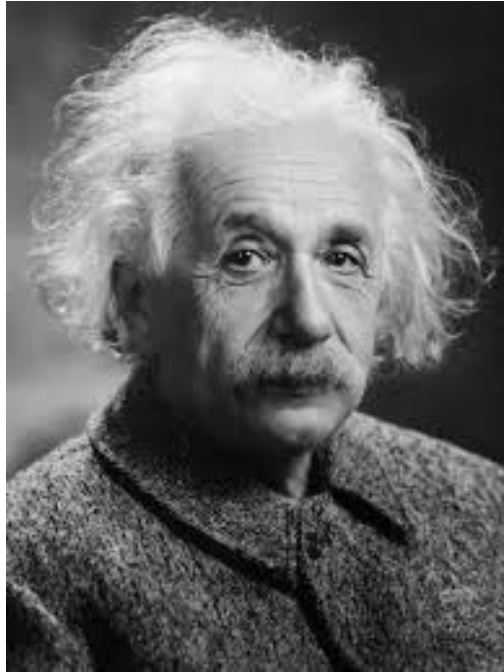


La variabilità genetica

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



La variabilità genetica

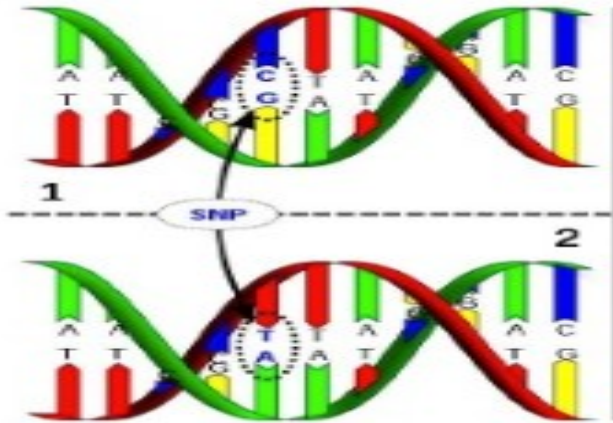


La **componente genetica individuale** è alla base della:

- ✓ **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)**.

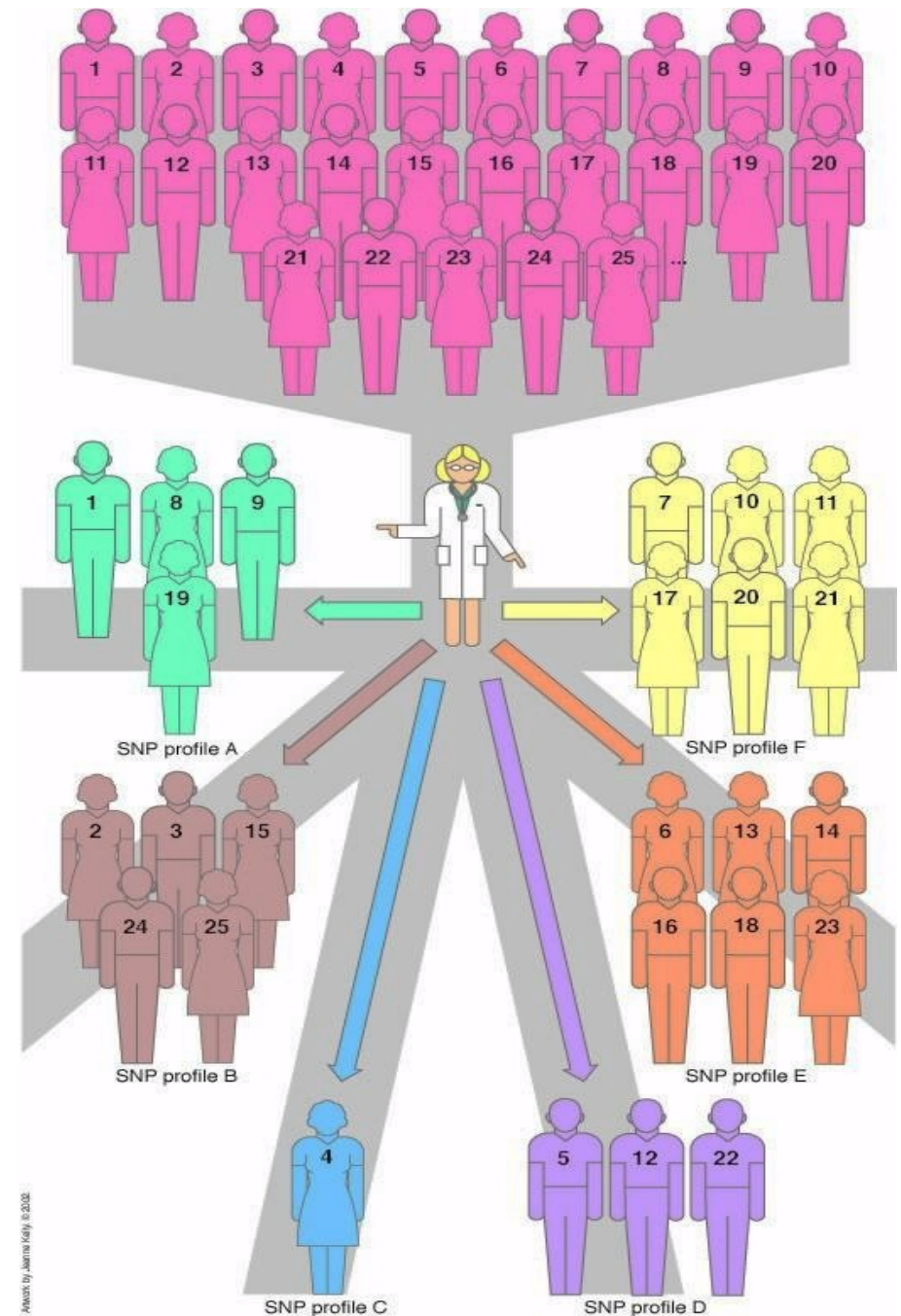


```
...C C A T T G A C...  
...G G T A A C T G...  
  
...C C G T T G A C...  
...G G C A A C T G...
```

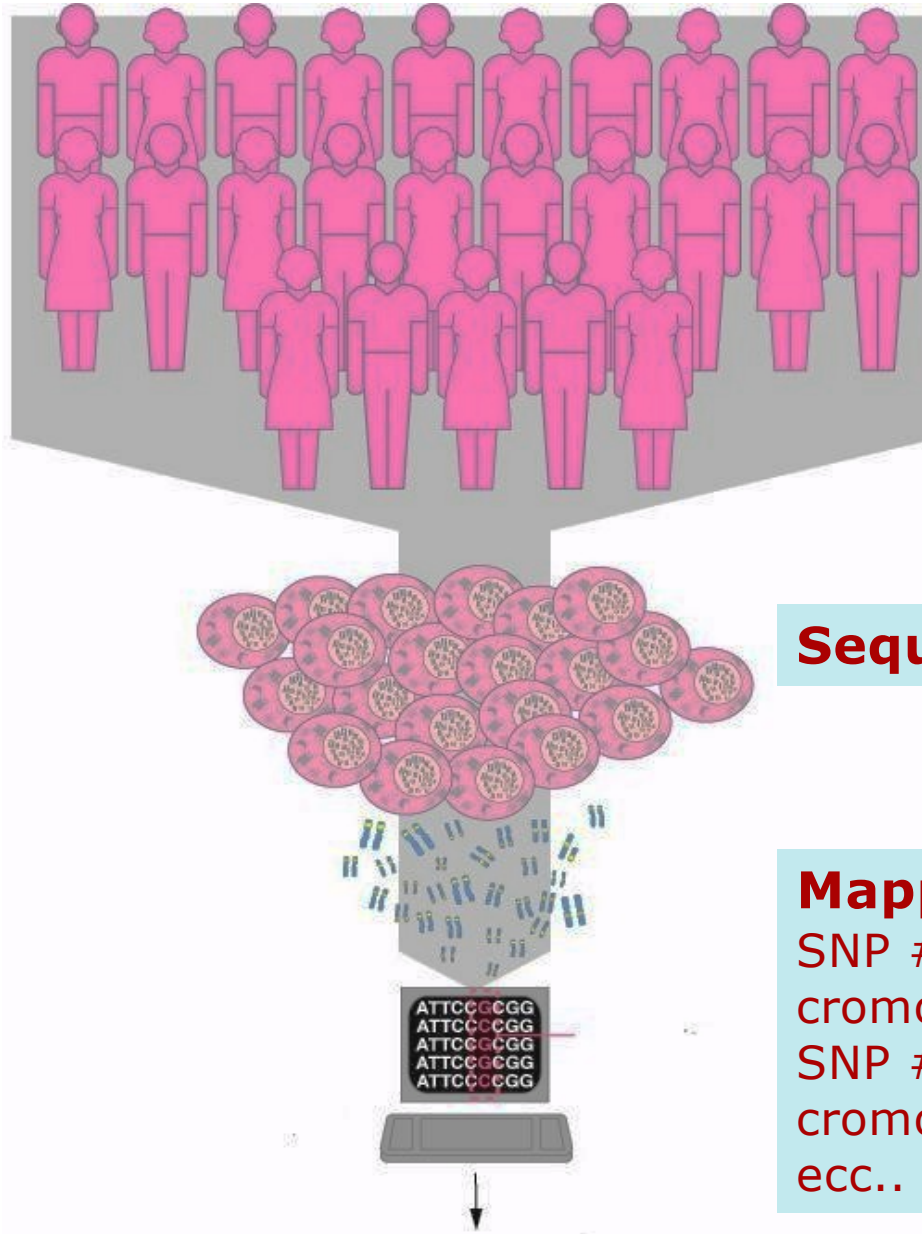


Profilo individuale degli SNPs

- ▶ 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
- ▶ Si possono individuare delle correlazioni tra il profilo di SNPs e la risposta ad uno specifico trattamento farmacologico o la predisposizione ad una malattia



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..

1000 Genomes

Mapping Human Genetic Variation



2008

1000 Genomes Project

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

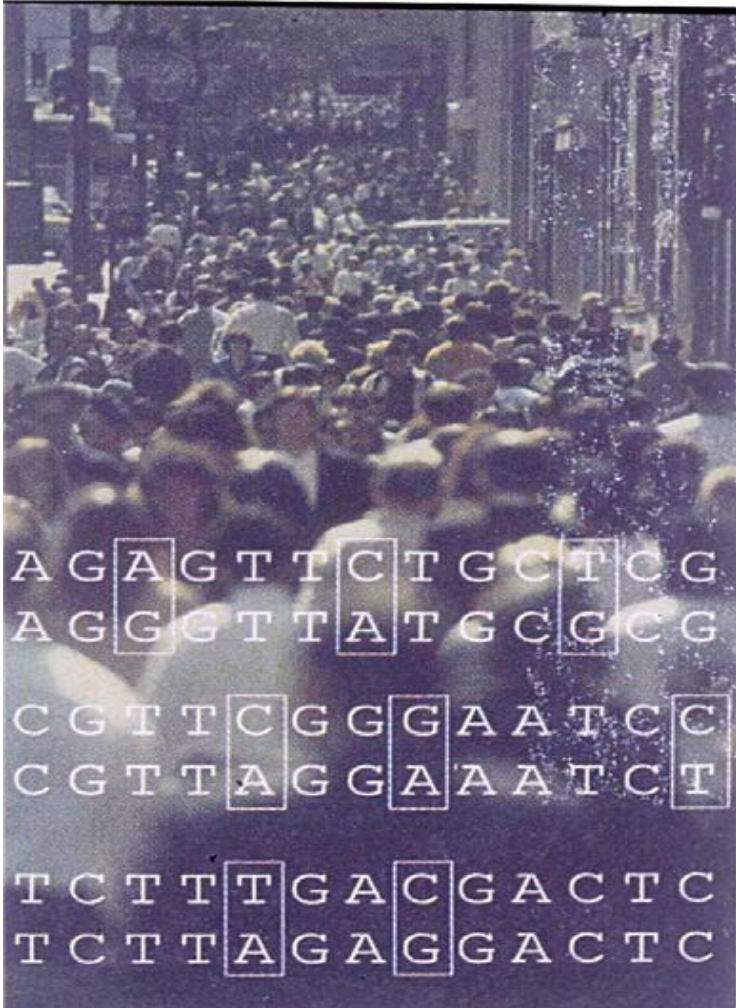
Sequenziati i genomi di 2500 individui appartenenti a 26 popolazioni diverse



88 milioni di varianti

1000 Genomes Project

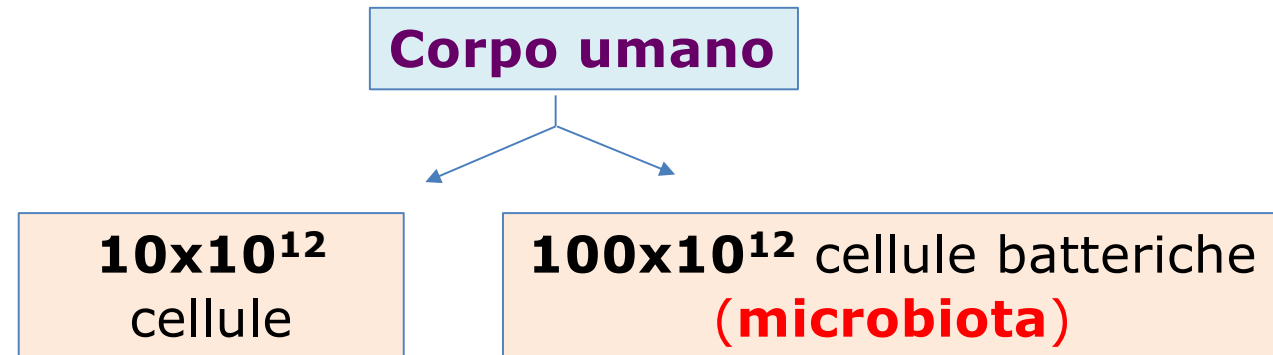
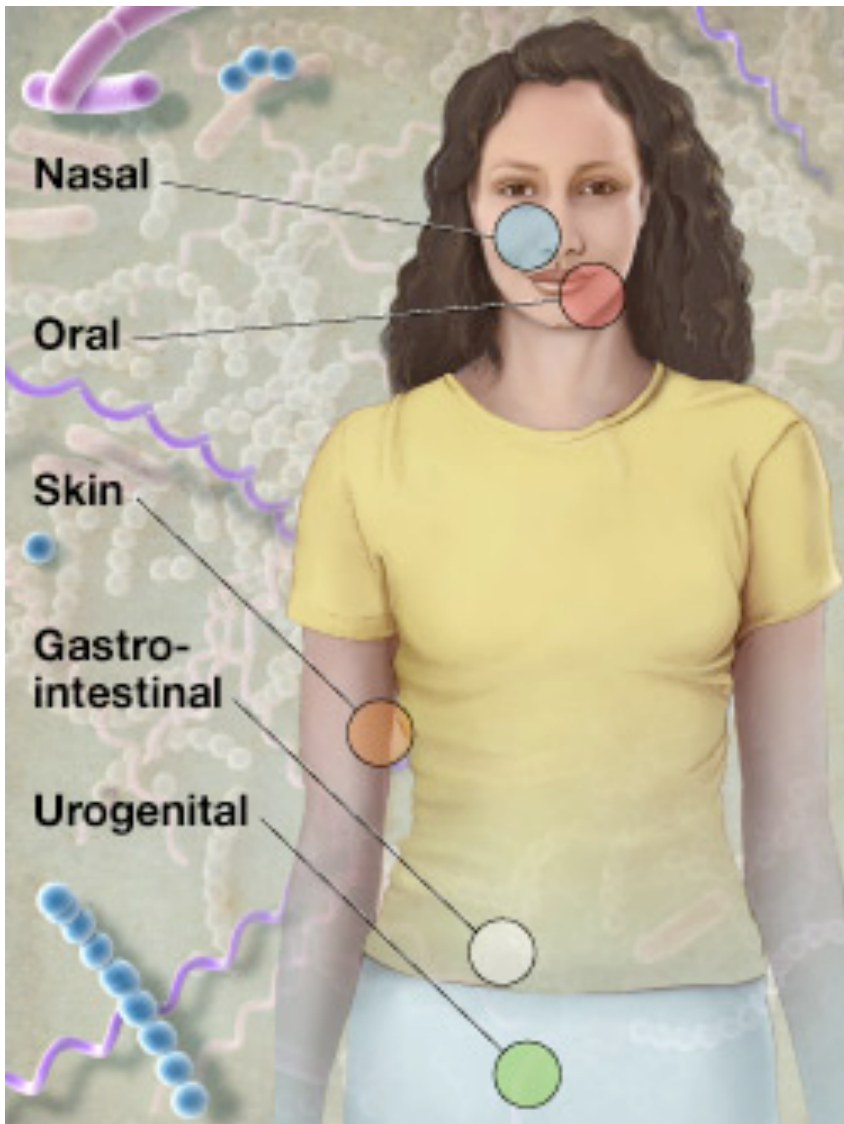
I numeri



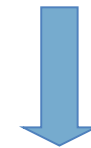
- 88 milioni SNPs
- 4 – 8 SNPs presenti in ogni gene
- 300-1000 bp che separano ogni SNP
- 150.000 SNPs non sinonimi
- 6% SNPs rari (<20%)
- 53% SNPs comuni (>20%)

Il nostro secondo genoma, il microbioma

Microbioma: l'insieme del patrimonio genetico della totalità dei microrganismi di un ambiente definito



Per 1 cellula umana vi sono 10 cellule microbiche



SUPER-ORGANISMO

Il nostro secondo genoma, il microbioma



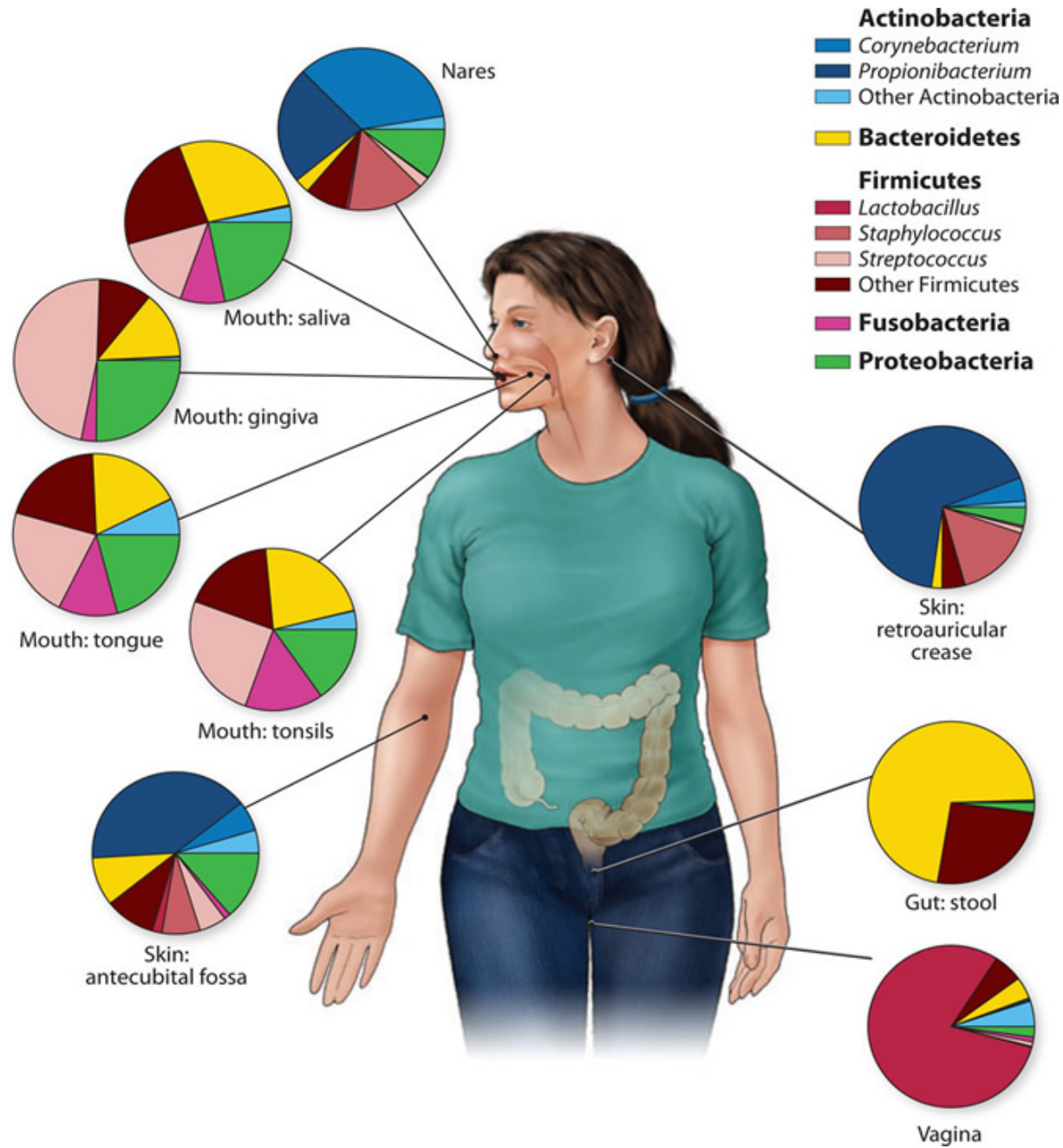
Microbioma
8.000.000 geni



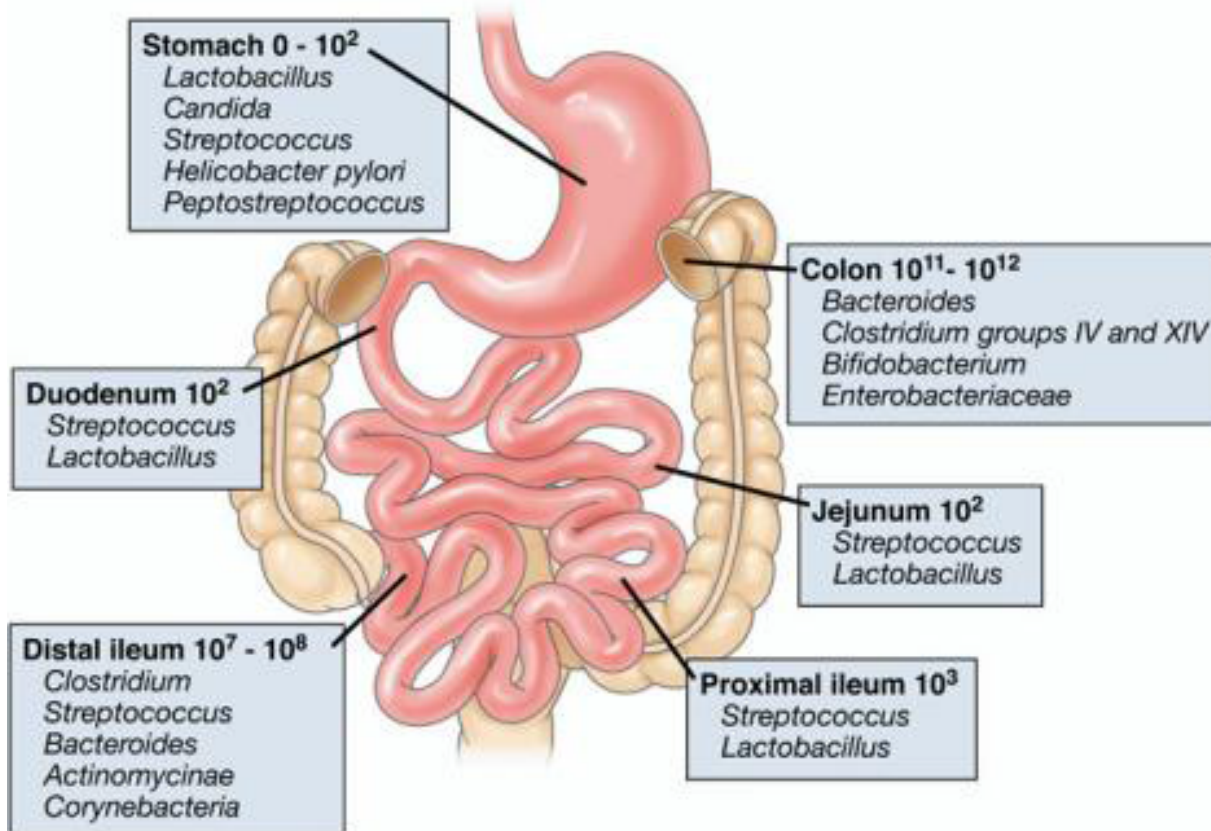
Genoma umano
25.000 geni

Metagenoma umano
Genoma Homo sapiens +
Genoma di miliardi di microrganismi

Il microbiota umano



Il microbiota intestinale



Microbiota 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 microbiota intestinale?

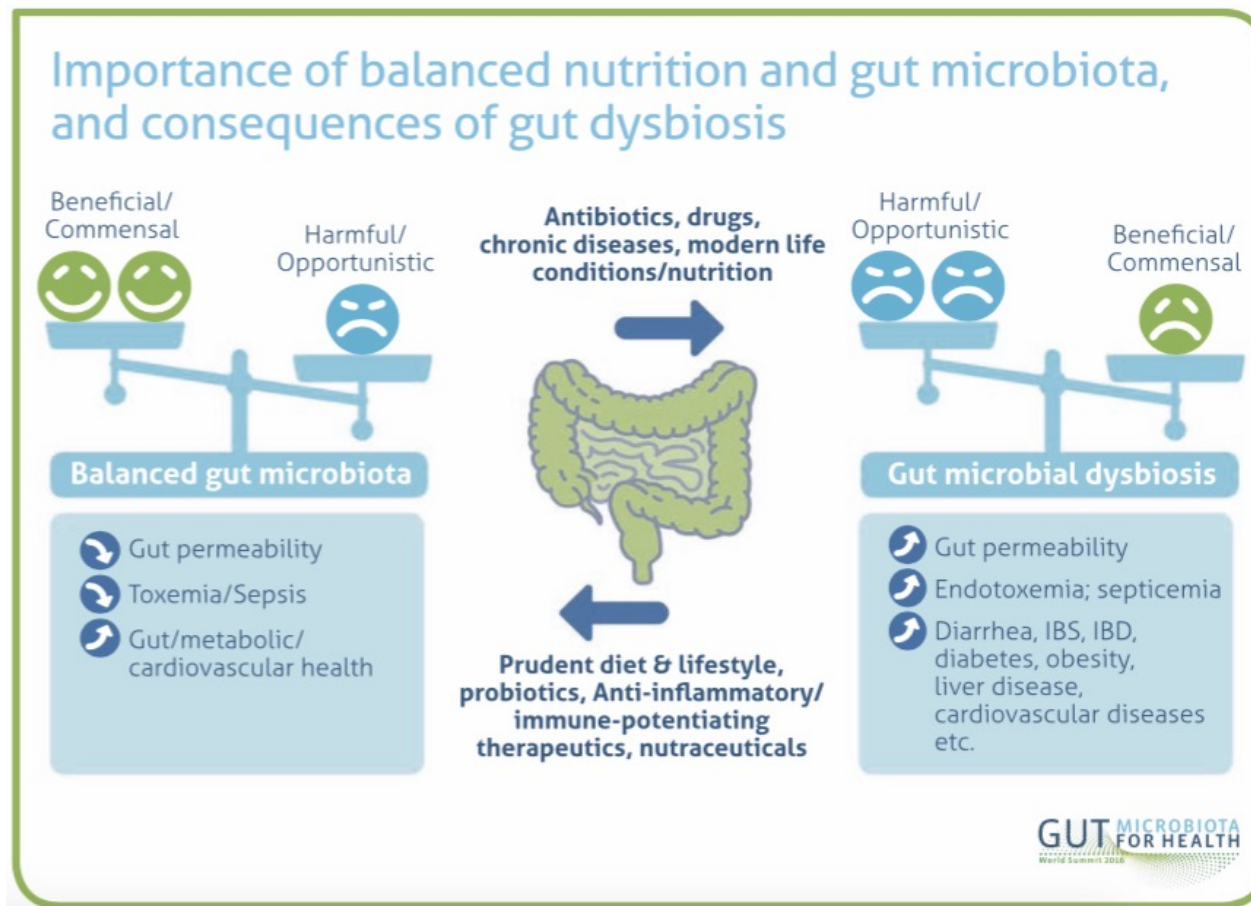
- Influenza lo sviluppo delle cellule della mucosa intestinale
- Serve come barriera contro i microrganismi patogeni (possibili fonti di malattie)
- Stimola le difese immunitarie
- E' fonte di energie e di sostanze nutritive (es. vitamine, prodotti di fermentazione delle fibre)



Importanza del microbiota 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.

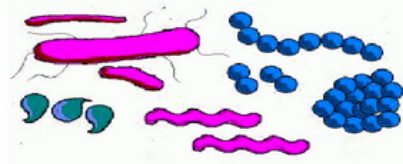
Ad esempio il materiale genetico dell'uomo e quello derivante dal suo microbiota (detto microbioma), rappresenta il **metagenoma umano**.

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

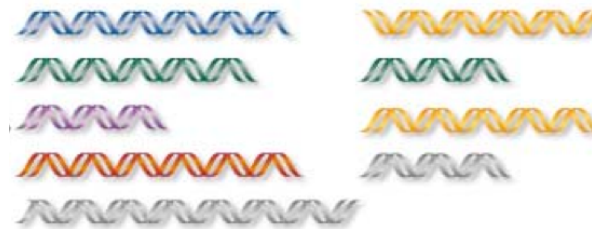
Non richiede l'isolamento e la messa in coltura di ogni singola specie, e la sua identificazione su basi biochimiche o morfologiche.



Sequenziamento Metagenomico



Campioni clinici o ambientali



Estrazione DNA

Approccio Metabarcoding

Approccio Shotgun

metabarcode

(e.g. 16S rRNA, ITS, Cox1, ..)

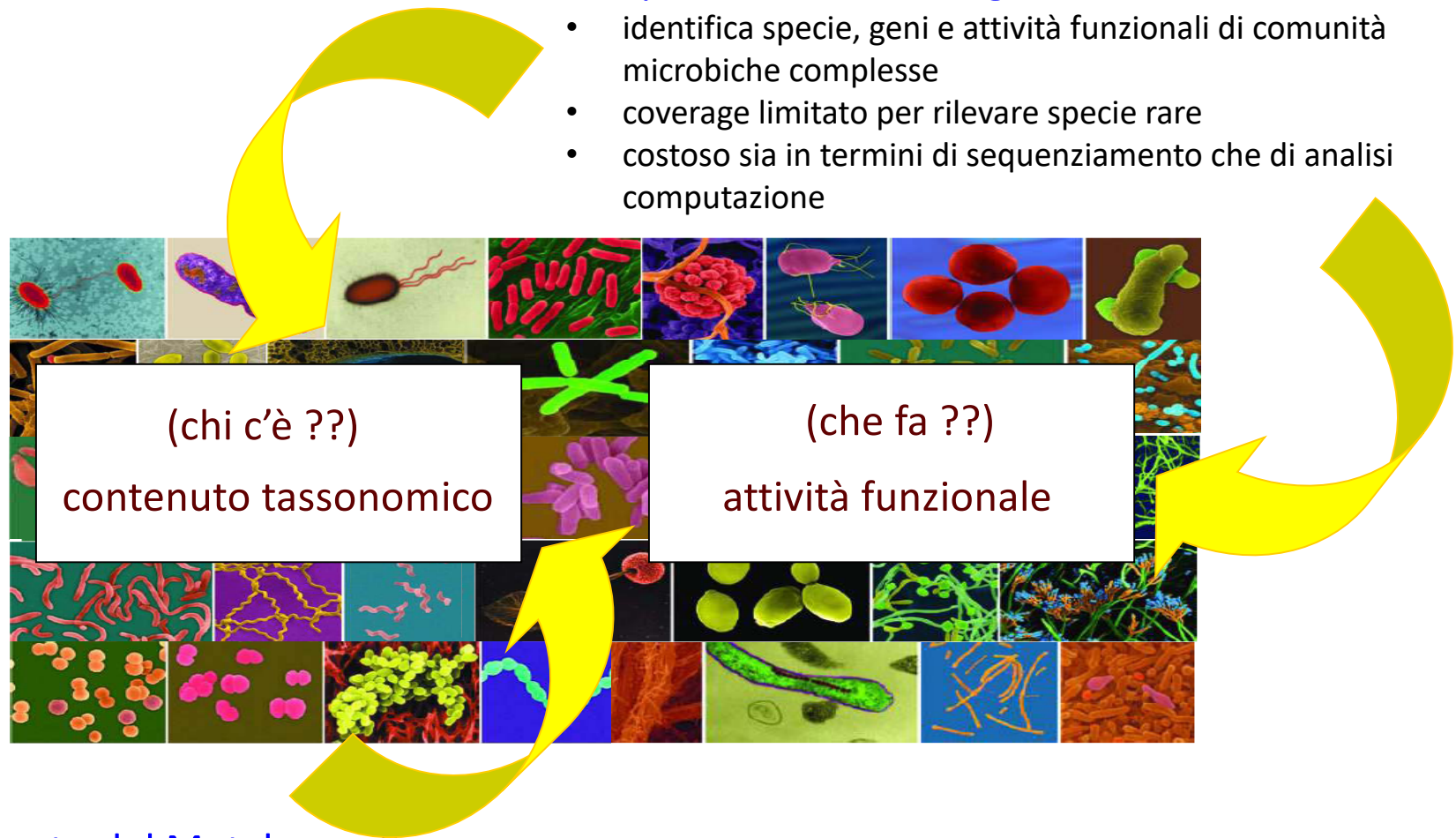
sequenziamento NGS
analisi dei dati NGS



Sequenziamento Metagenomico

Sequenziamento Shotgun

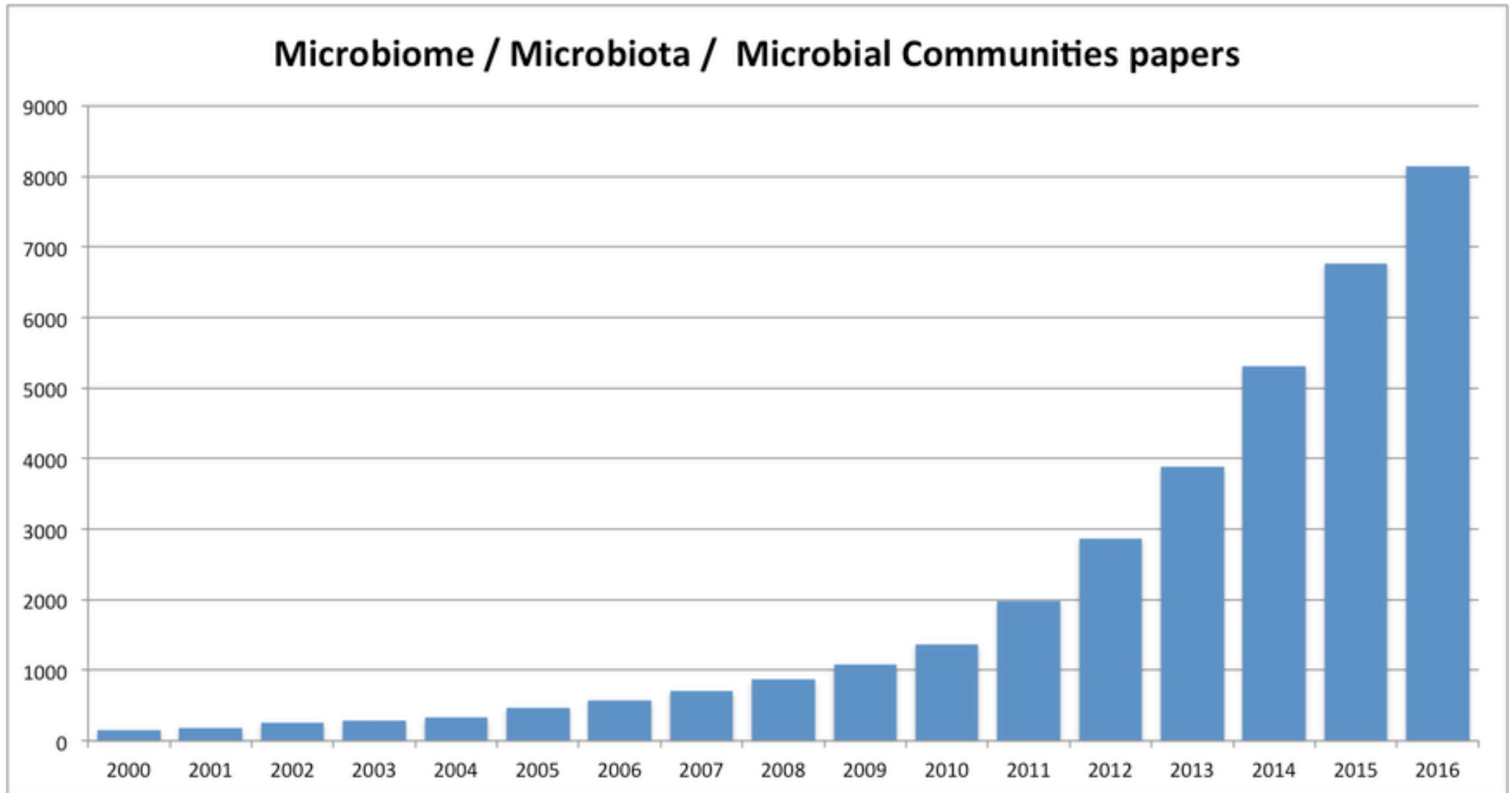
- identifica specie, geni e attività funzionali di comunità microbiche complesse
- coverage limitato per rilevare specie rare
- costoso sia in termini di sequenziamento che di analisi computazione



Sequenziamento del Metabarcoding

- 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 Progetto Genoma Umano..... dal 2003 ad oggi

Benefici del sequenziamento del Genoma Umano

Medicina

- migliorare la diagnosi delle malattie
- identificare la predisposizione genetica a specifiche malattie
- 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..... 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 microbiota e organismo
- ecc

Concludendo....

L'analisi integrata di dati molecolari (genoma, trascrittoma, proteoma, microbioma), metabolici, funzionali, ecc. può essere utile per fornire le informazioni necessarie per la comprensione di tutti i meccanismi alla base dei processi fisiologici e patologici dell'uomo e quindi a far luce sul segreto della vita, ma.....

La strada è ancora lunga!!

