

# Corso di Biologia Molecolare



Dott.ssa Renata Tisi

Dip. Biotecnologie e Bioscienze

Ed. U4

Tel. 02 6448 3522

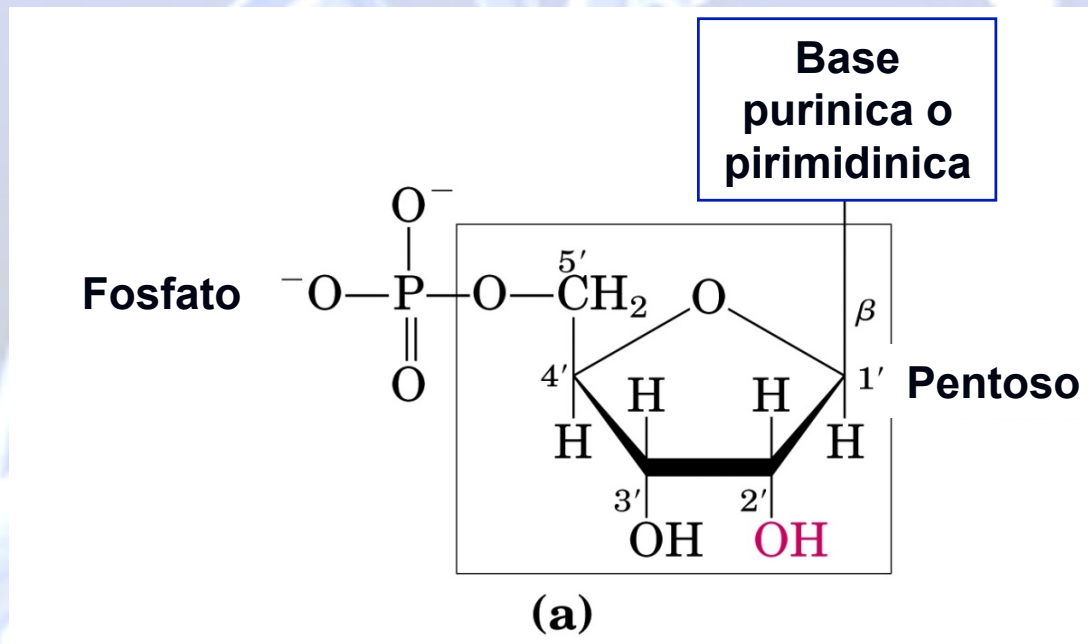
[renata.tisi@unimib.it](mailto:renata.tisi@unimib.it)

# Acidi nucleici

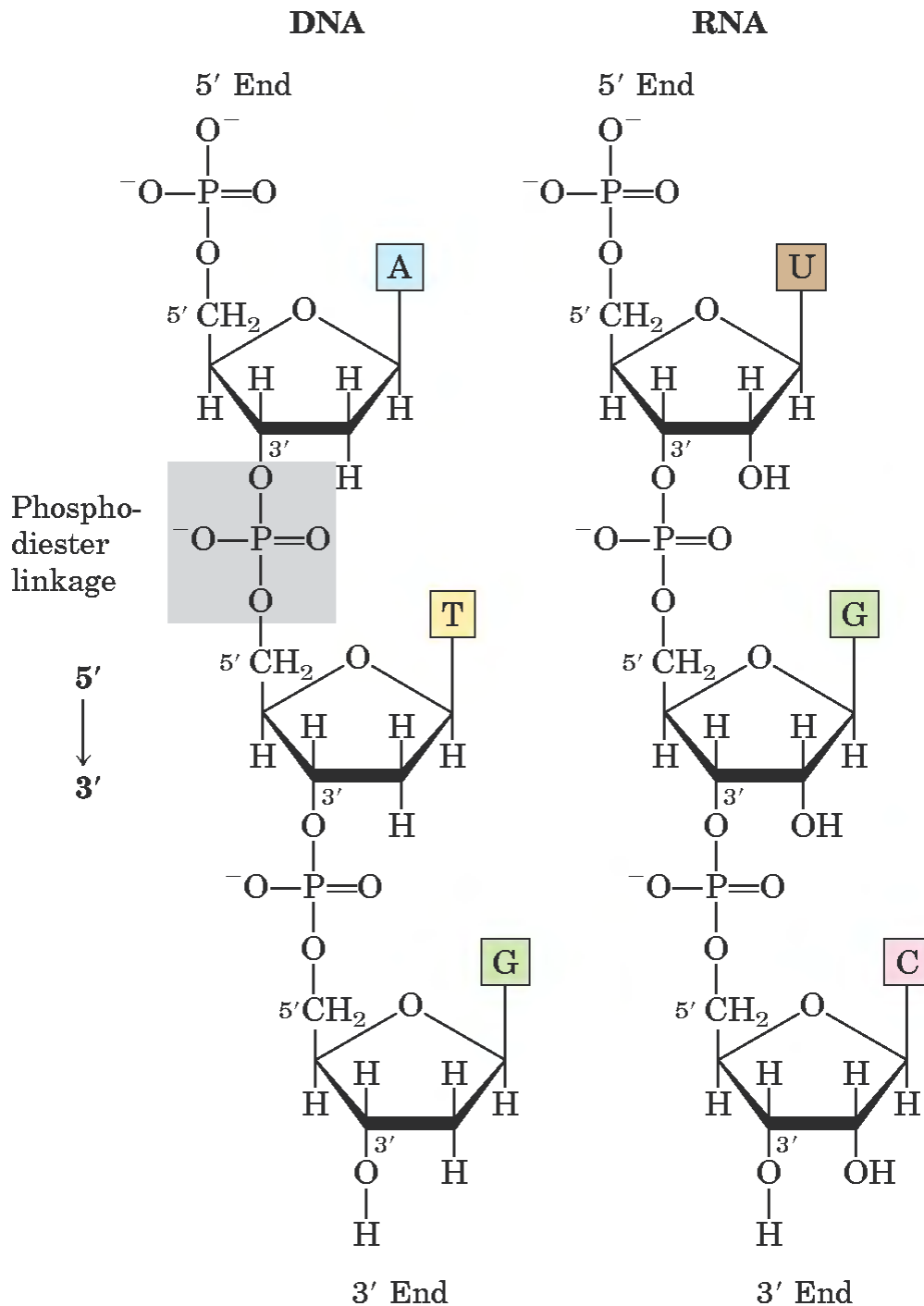
Il ruolo degli acidi nucleici è quello di custodire e trasmettere l'informazione genetica nelle cellule, nei tessuti e negli organismi.

**ACIDO RIBONUCLEICO (RNA) contiene desossiribosio**

**ACIDO DEOSSIRIBONUCLEICO (DNA) contiene ribosio**



La formula base di un **nucleotide monofosfato**



Negli acidi nucleici il collegamento fra unità monomeriche successive è realizzato attraverso un residuo fosfato attaccato all'ossidrile sul carbonio 5' di un'unità e un gruppo 3' ossidrilico della seconda unità.

Questo forma un **legame fosfodiesterico** tra i due residui.

Il legame fosfodiesterico tra le unità monomeriche si trova tra il carbonio 3' di un monomero e il carbonio 5' del successivo. In questo modo le due estremità della catena polinucleotidica risultano distinguibili: un'estremità porterà un gruppo 5' fosfato libero, mentre l'altra porterà un gruppo 3' ossidrilico libero.

**POLARITÀ 5' → 3'**

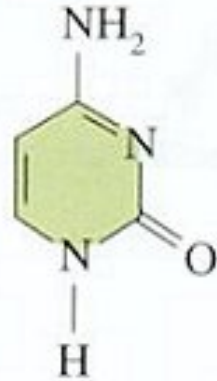


## Struttura Primaria degli Acidi Nucleici

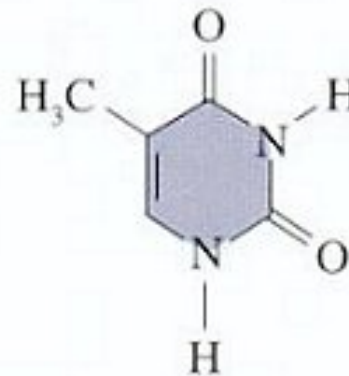
Una catena polinucleotidica ha una propria individualità, determinata dalla sequenza delle sue basi, cioè dalla sua **sequenza nucleotidica**. Questa sequenza costituisce la struttura primaria di quel particolare acido nucleico. L'importanza di tale struttura risiede soprattutto nel fatto che è proprio nella struttura primaria del DNA che viene immagazzinata l'informazione genetica. Un **GENE** non è nient'altro che una particolare sequenza di DNA, che codifica l'informazione in un linguaggio a quattro lettere, nel quale ogni lettera è rappresentata da una base.

# Le basi azotate

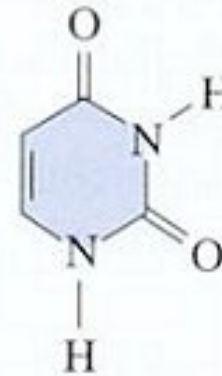
## Pirimidine



Cytosine (C)  
(DNA and RNA)

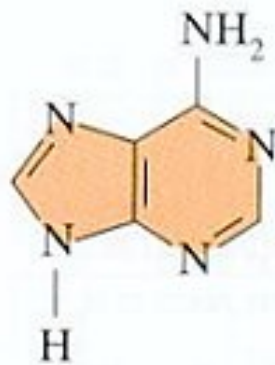


Thymine (T)  
(DNA only)

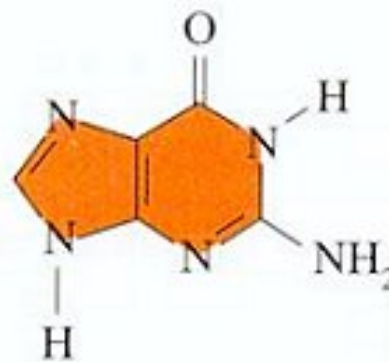


Uracil (U)  
(RNA only)

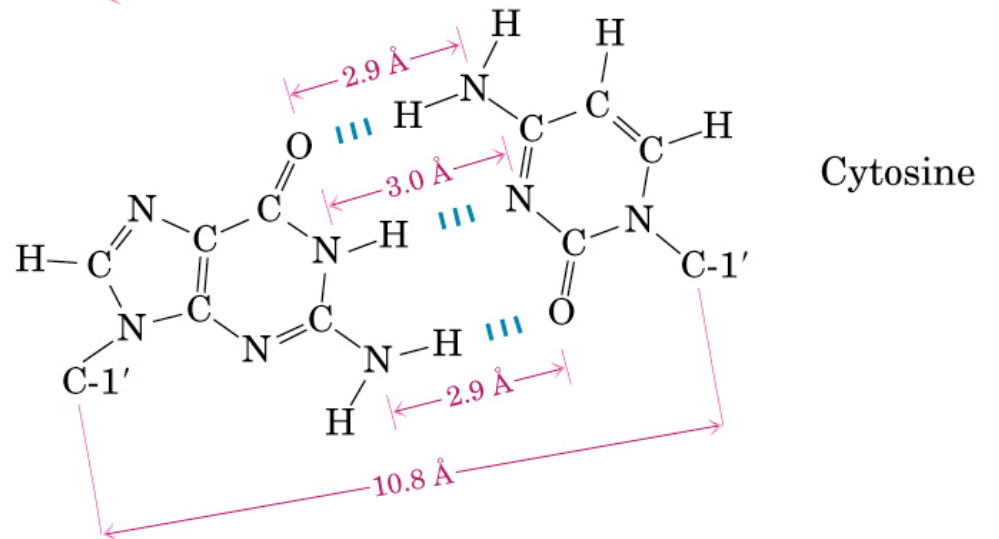
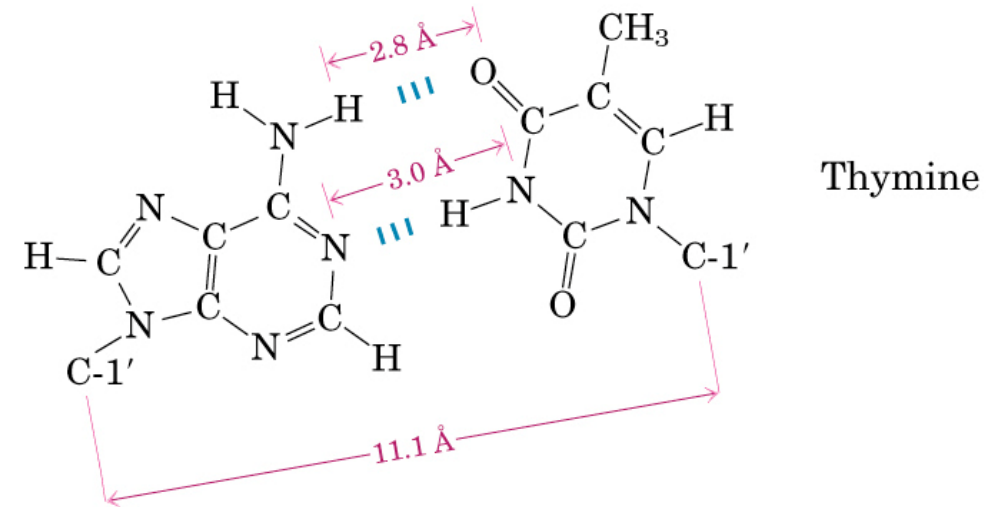
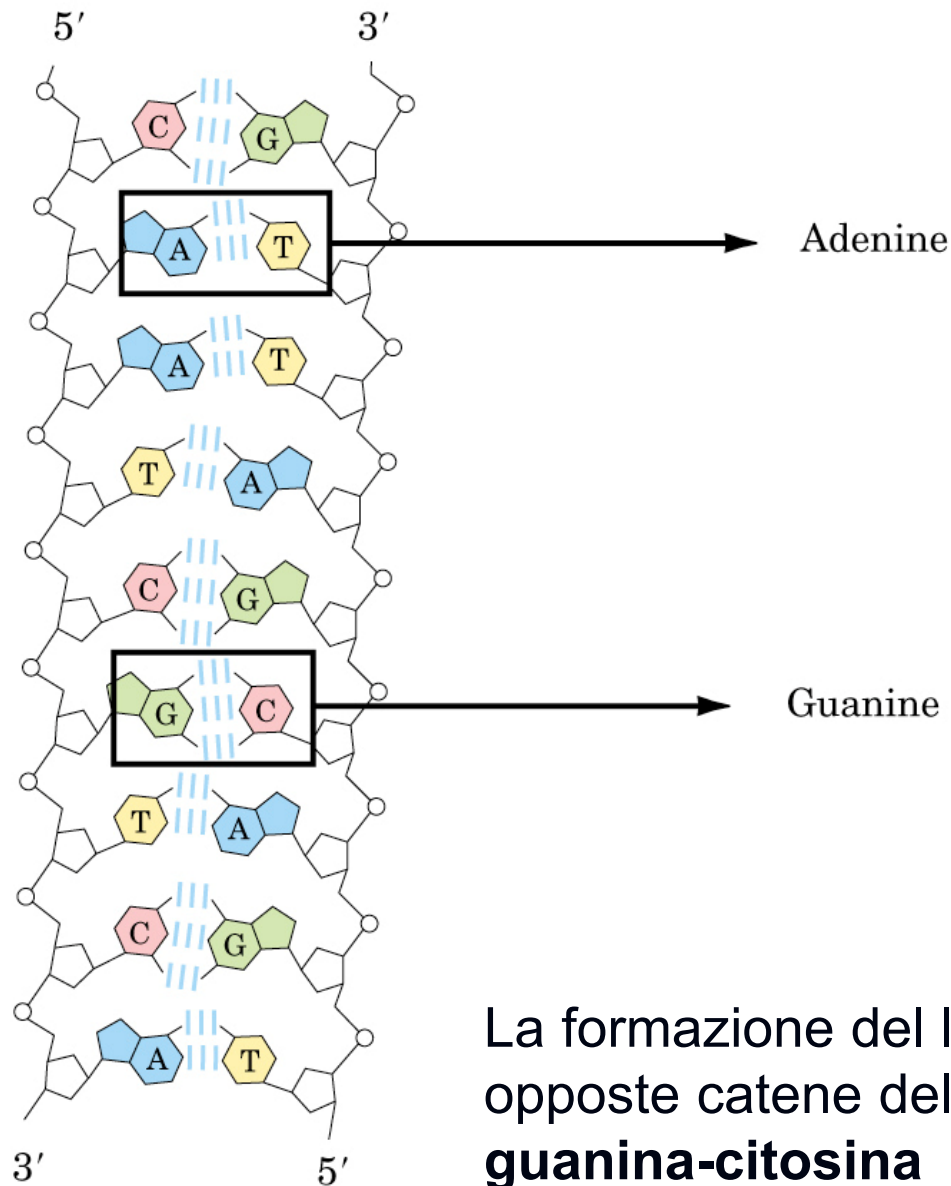
## Purine



Adenine (A)  
(DNA and RNA)

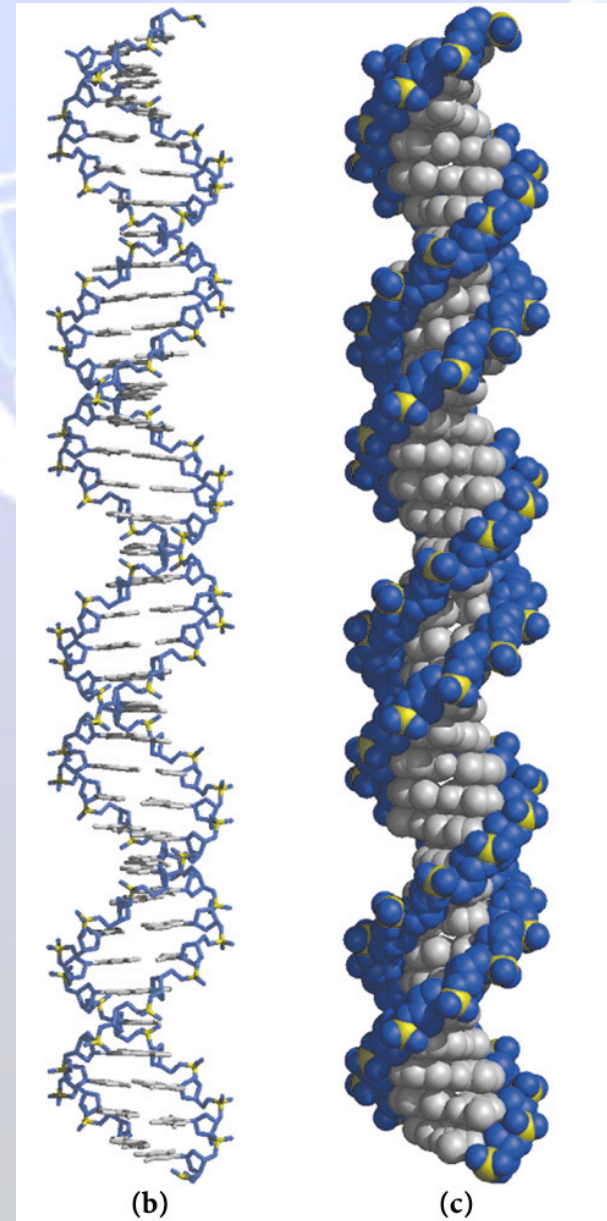
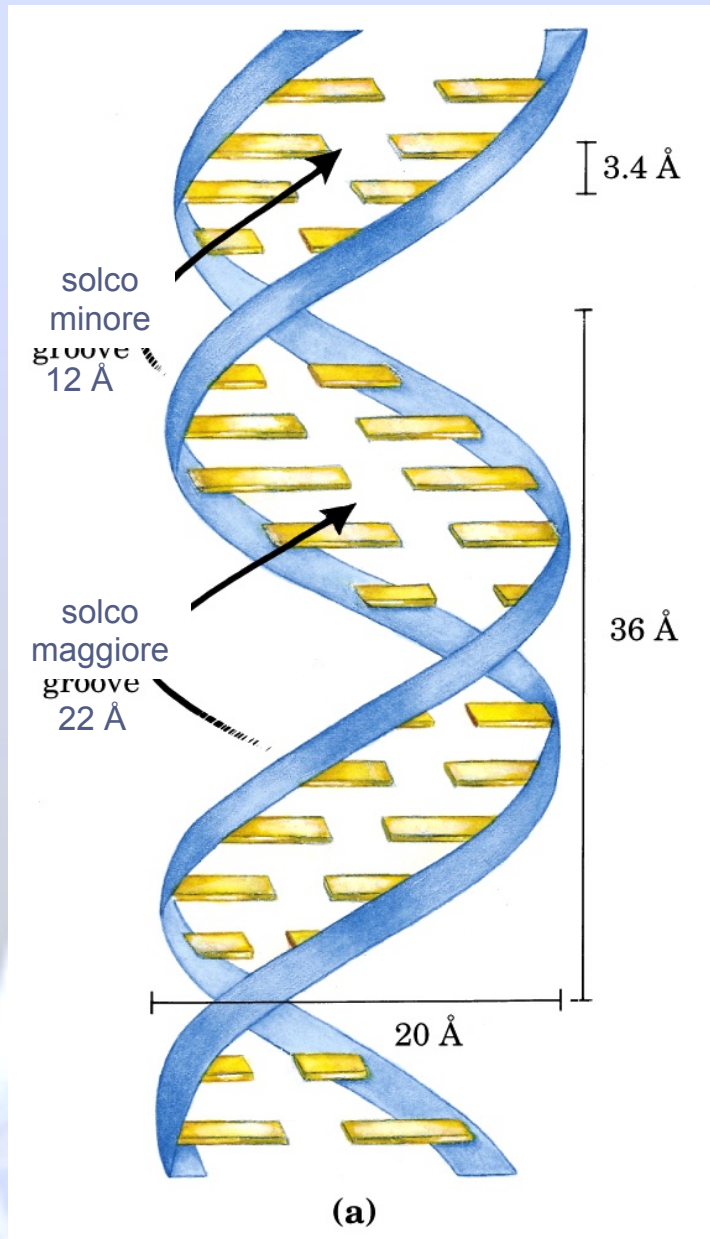


Guanine (G)  
(DNA and RNA)



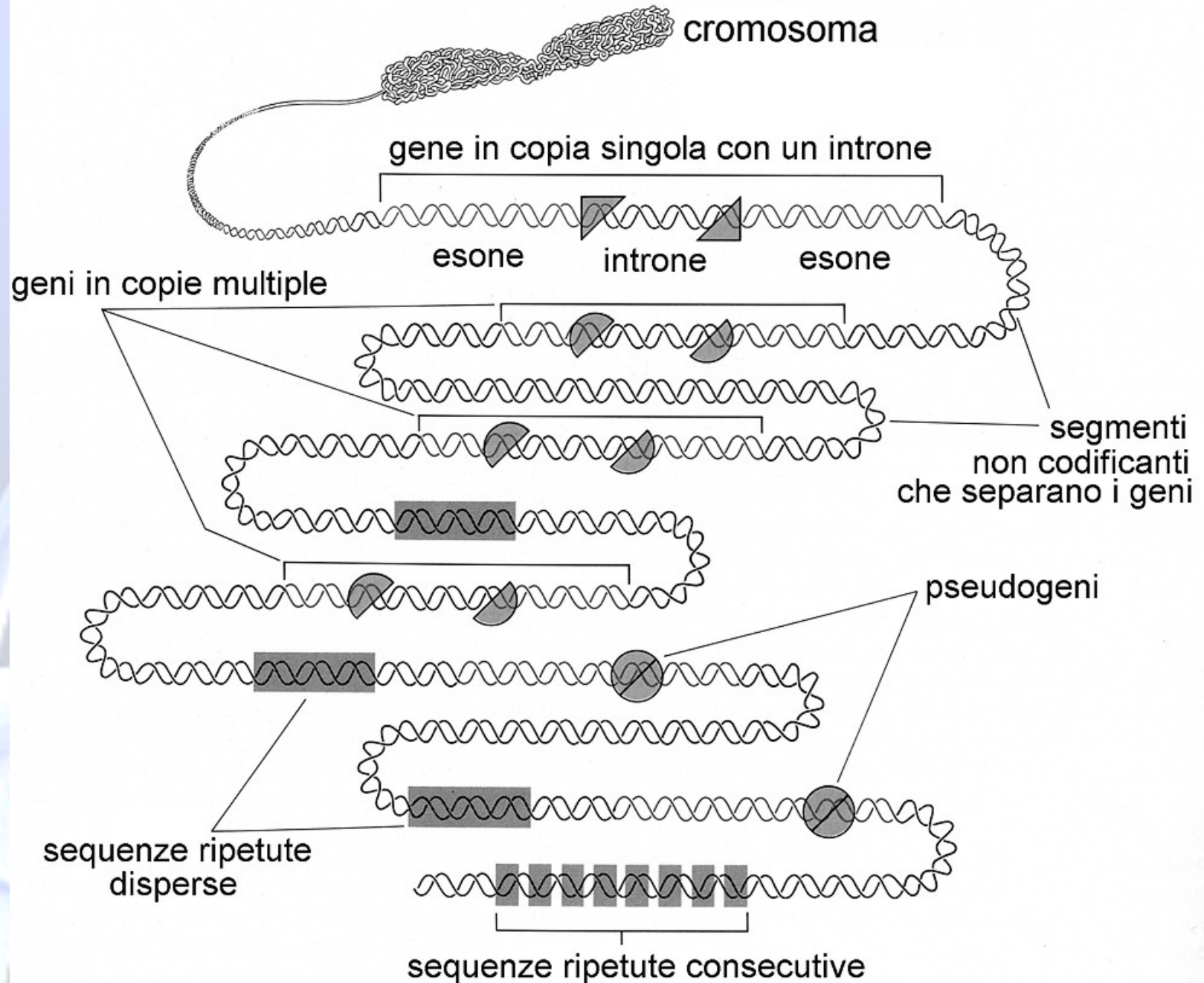
La formazione del legame idrogeno tra le basi presenti sulle opposte catene del DNA porta allo specifico appaiamento di **guanina-citosina** e di **adenina-timina/uracile**.

# Struttura secondaria del DNA



Una rappresentazione schematica (sinistra) e nel dettaglio (destra) dell'andamento nello spazio delle due eliche del DNA nella classica forma B.

# Complessità del DNA





# I geni e l'ereditarietà: genotipo e fenotipo



Il *gene* è l'unità ereditaria fondamentale degli organismi viventi.

Diverse forme alternative di uno stesso gene si chiamano **alleli**; in un individuo diploide i 2 alleli occupano sui cromosomi omologhi lo stesso **locus**.

Un carattere monofattoriale è controllato da un solo gene; un carattere la cui espressione è influenzata da più loci e/o da altri elementi (anche ambientali) si chiama multifattoriale.

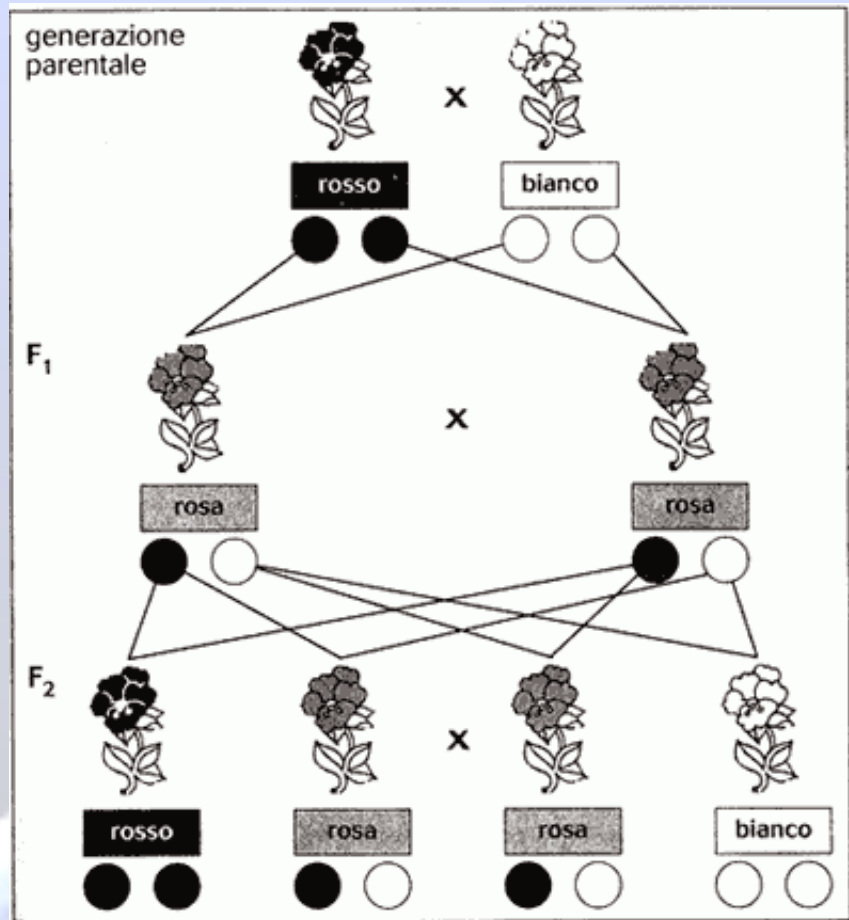
Il **genotipo** di un individuo è dato dal suo corredo genetico.

Il **fenotipo**, invece, è l'insieme dei caratteri che l'individuo manifesta: dipende dal suo genotipo, dalle interazioni fra geni e anche da fattori esterni.

L'**omozigosi** o l'**eterozigosi** è una caratteristica che riguarda il genotipo.

Ad esempio, se per un dato gene esistono 2 alleli che chiamiamo "A" e "a", i genotipi possibili sono 3: AA , aa , Aa.

# I geni e l'ereditarietà: alleli e leggi di Mendel

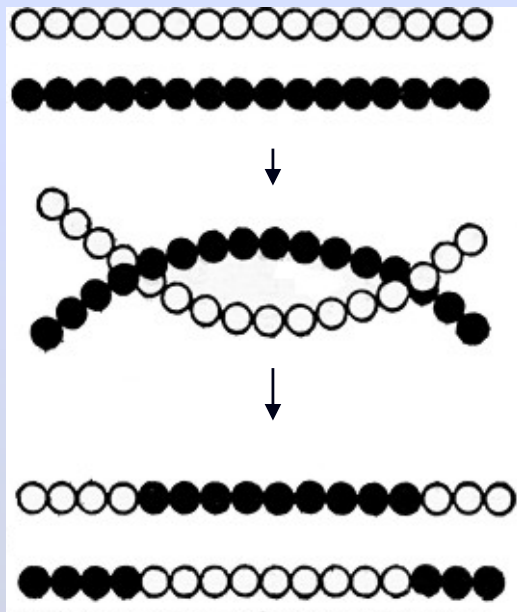


**Legge della dominanza** (o *legge della omogeneità di fenotipo*): Gli individui nati dall'incrocio tra due individui omozigoti che differiscono per una coppia allelica, avranno il fenotipo dato dall'allele dominante. Con significato più ampio rispetto al lavoro di Mendel, può essere enunciata come legge dell'uniformità degli ibridi di prima generazione.

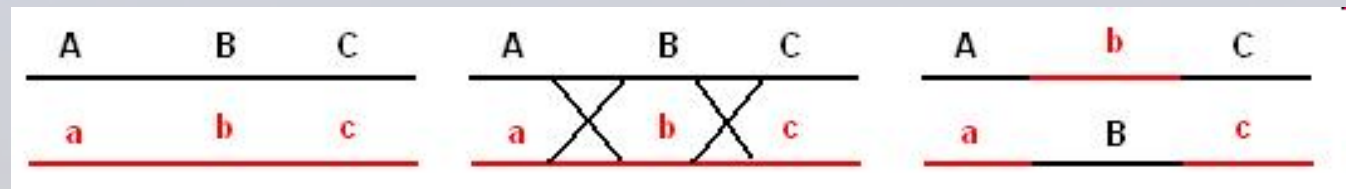
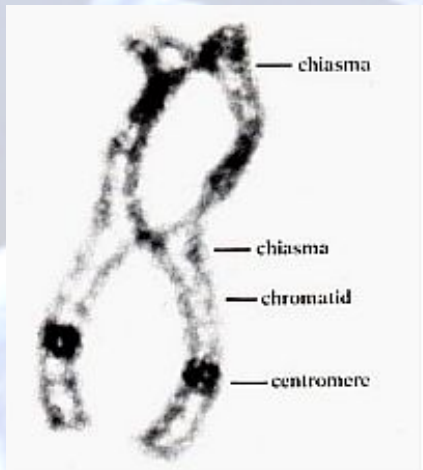
**Legge della segregazione** (o *legge della disgiunzione*): gli alleli di un singolo locus segregano al momento della formazione dei gameti (in seguito fu evidente che ciò era dovuto al fenomeno noto come meiosi). Può essere enunciata come ricomparsa del recessivo nella F<sub>2</sub>.

**Legge dell'assortimento indipendente** (o *legge di indipendenza dei caratteri*): i diversi alleli si trasmettono indipendentemente l'uno dagli altri, secondo precise combinazioni.

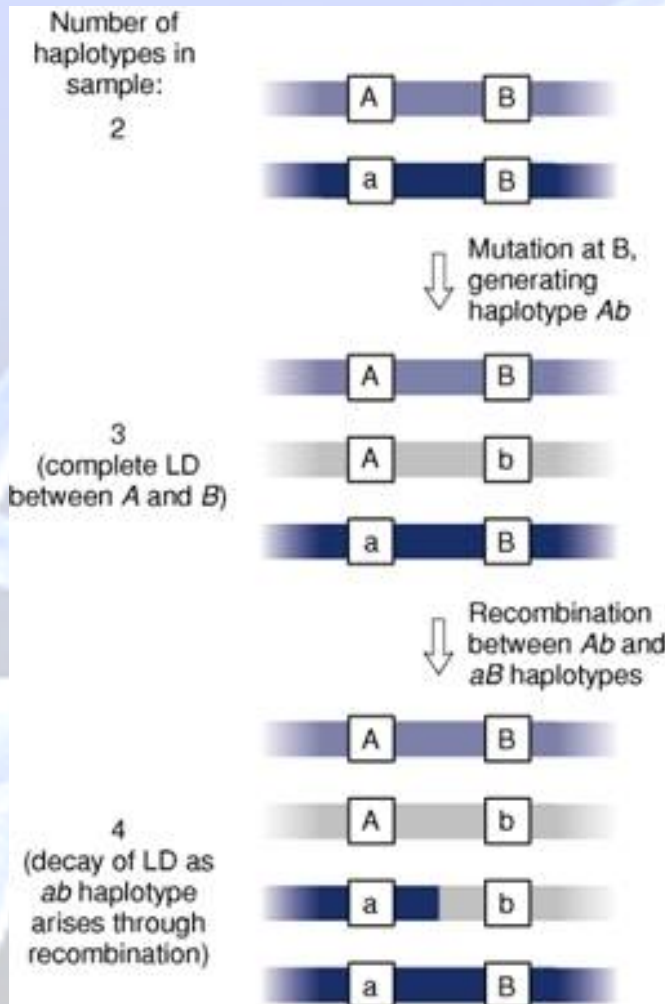
# Indipendenza dei caratteri: la ricombinazione



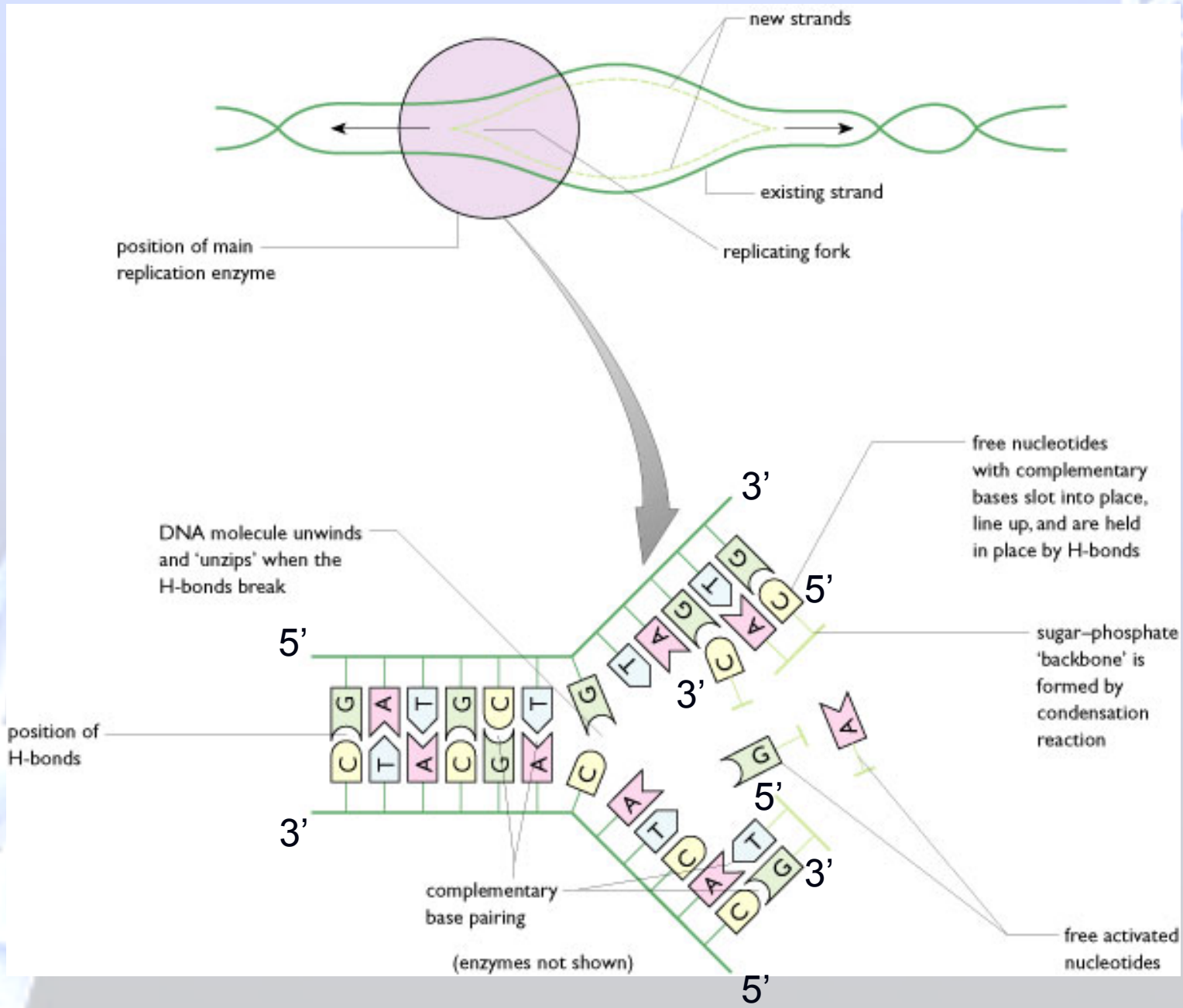
La **frequenza di ricombinazione** misura l'entità del *linkage* o *associazione genica*: più due *loci* sono fisicamente vicini meno frequenti saranno gli eventi di ricombinazione che li separano, quindi più strettamente saranno associati e meno apparirà indipendente la loro segregazione.



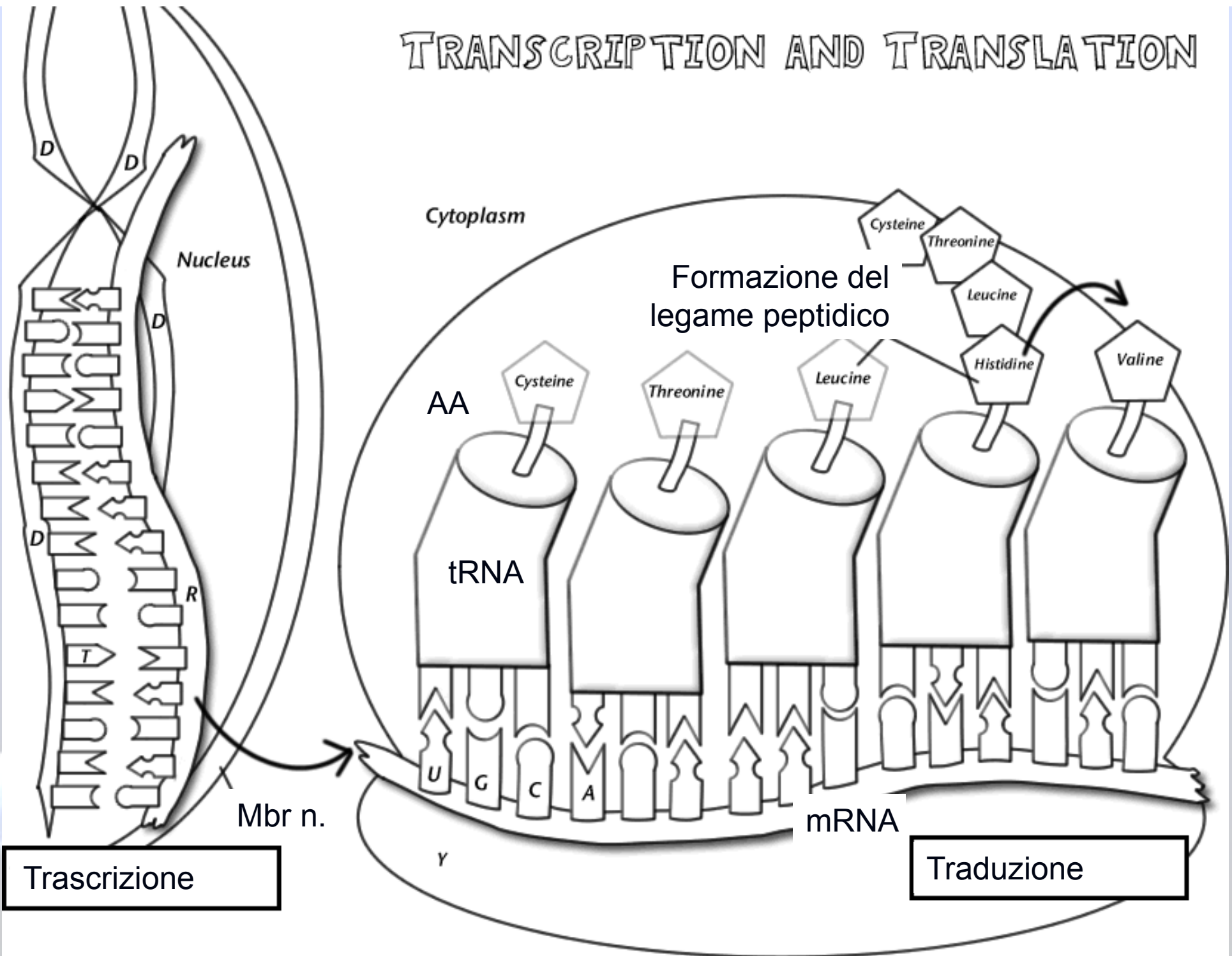
Un aplotipo è una combinazione di alleli a diversi siti polimorfici sullo stesso cromosoma che non segregano in modo indipendente, ma mostrano un certo grado di **linkage**. I siti possono essere SNPs, microsatelliti o qualsiasi tipo di polimorfismo.



LD sta per linkage disequilibrium

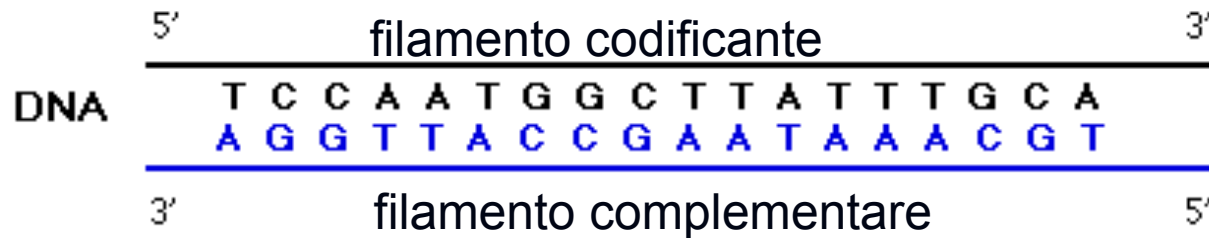


# TRANSCRIPTION AND TRANSLATION

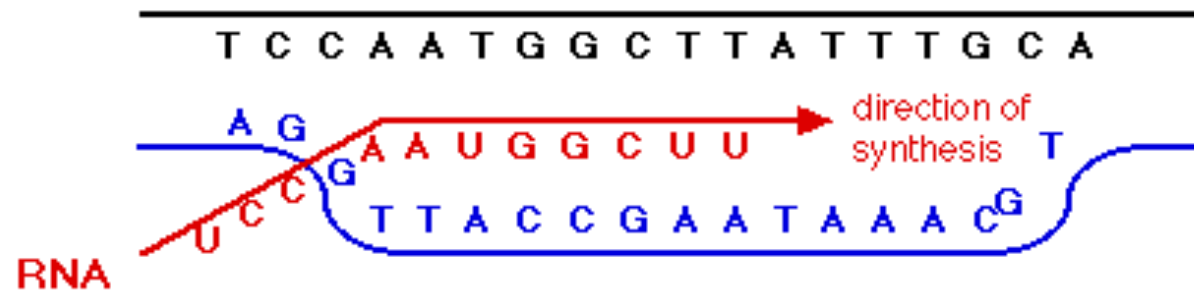


# La trascrizione: da DNA a RNA

## Transcription of RNA from DNA



- The bottom strand of the DNA molecule above is the template for RNA synthesis.
- RNA polymerase makes a copy of the DNA sequence but substitutes uridine (U) in place of thymine (T).



- The bottom strand of the DNA duplex is used as the template to synthesize RNA. However, the sequence of bases in the RNA is the same as in the top strand of the DNA, with U in place of T



Il primo evento dell'espressione genica è la **trascrizione**, che consiste nella produzione di un tratto di RNA corrispondente a uno o più geni.

Il processo è detto trascrizione perché le modalità di immagazzinamento dell'informazione a livello dell'RNA sono concettualmente identiche a quelle del DNA. In altre parole, su uno dei filamenti di DNA viene prodotto un filamento di RNA *complementare*, secondo le regole di appaiamento del DNA, A=T e G=C, salvo il fatto che l'RNA prevede uracile al posto della timina.

# La sintassi della trascrizione

PROMOTORE

terminatore

ORF  
Open Reading Frame

## Typical Bacterial Promoter

82 84 78 65 54 48

TTGACA

-35

← 17 bp →

-10

tss  
↓

TATAAT

80 95 45 50 60 96

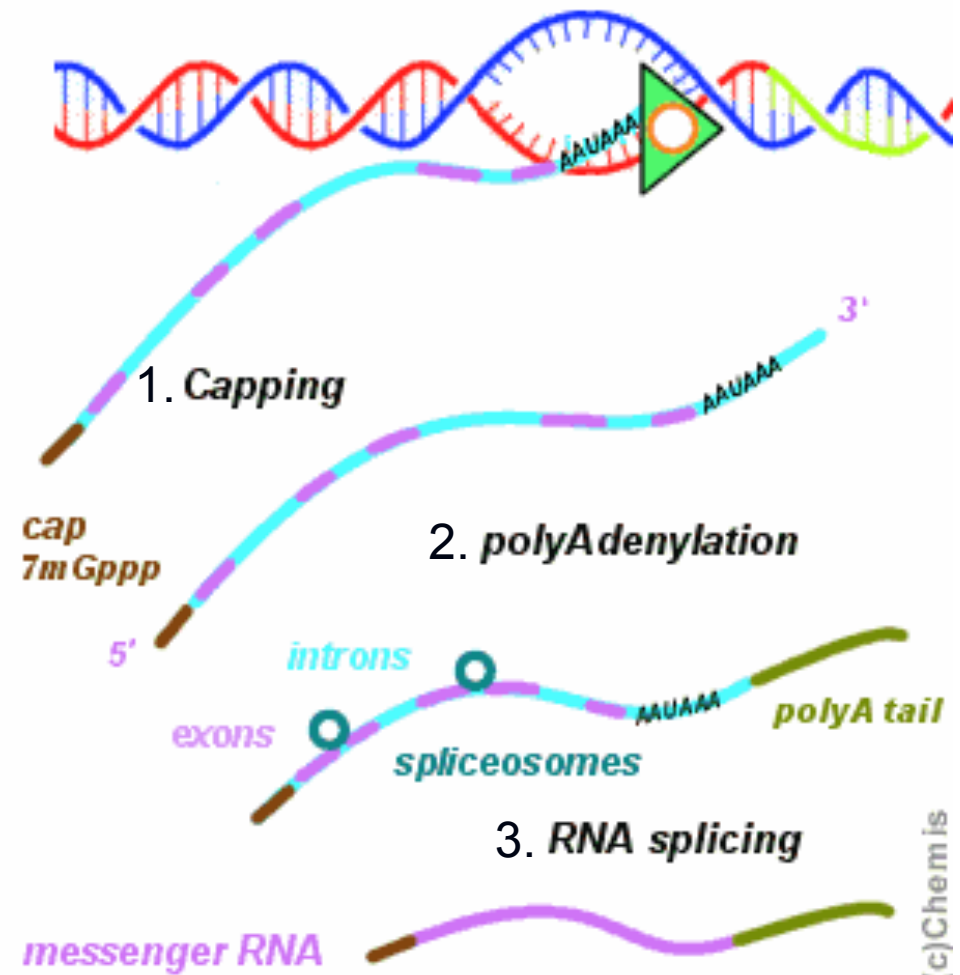
TATA box



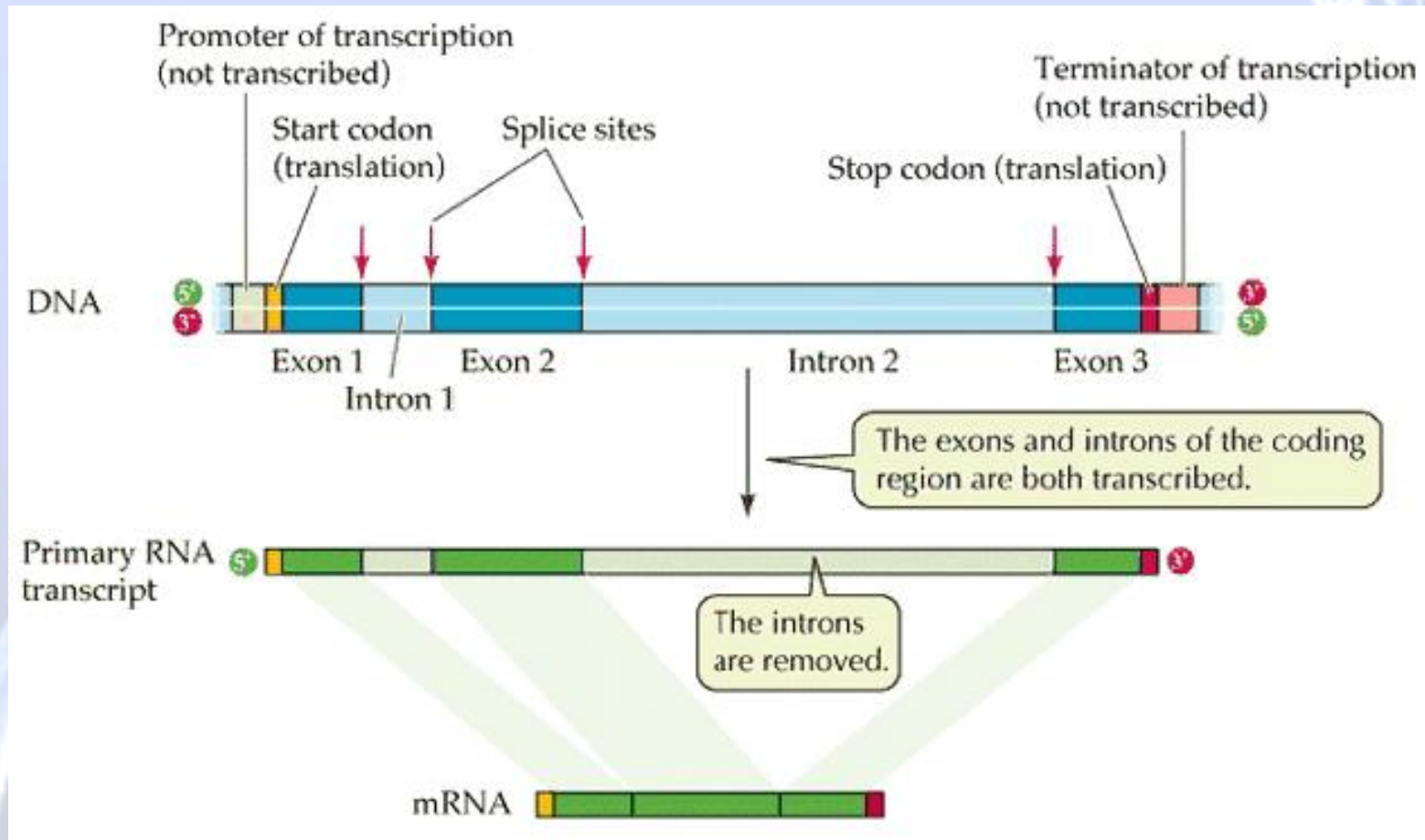
# Modificazioni dell'mRNA negli eucarioti

Subito dopo l'inizio della sua trascrizione, il terminale 5' dell'mRNA viene modificato dalla guanililtransferasi con l'aggiunta di una **7-metilguanossina** (7mGTP), denominata **CAP**, legata attraverso un insolito legame 5'-5' trifosfato. Questa modificazione rende l'mRNA più stabile e resistente all'attacco delle esonucleasi. Il CAP serve anche come punto di attacco per le proteine coinvolte nello splicing e anche per l'inizio della traduzione.

Gli mRNA vengono in seguito modificati attraverso la **poliadenilazione** al terminale 3'. Appena prima della terminazione, una sequenza segnale di poliadenilazione, AAUAAA, viene riconosciuta da una poli-A polimerasi. Il trascritto viene tagliato circa 20 basi a valle di tale sequenza e una coda di poli-adenine (20-250 adenine) viene aggiunta in 3'.



# Esoni e introni



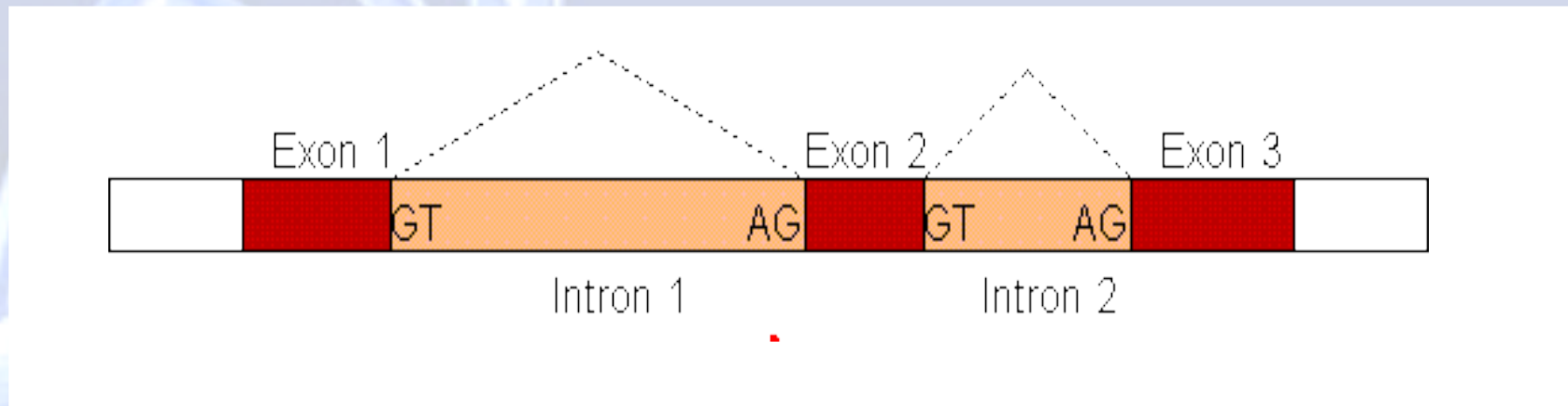
Un **trascritto primario** è una copia fedele del gene da cui deriva, ma comprende anche zone che non verranno tradotte in una proteina. Una breve **sequenza leader** in 5', una generalmente più lunga **sequenza coda** in 3', e sequenze interne al gene che non vengono tradotte: si tratta di **introni**, che devono essere rimossi prima della traduzione in catena polipeptidica.

# Splicing

Una modifica post-trascrizionale molto importante degli RNA messaggeri è la rimozione degli introni o **RNA splicing**.

Alcuni introni sono autocatalitici, cioè sono in grado di rimuoversi da soli. Altri introni sono caratterizzati da alcune basi conservate ai loro estremi, che sono coinvolte nella catalisi del processo di eliminazione dell'introne stesso.

Il meccanismo comprende la formazione di un intermedio a cappio, in un processo guidato da piccole ribonucleoproteine nucleari (snRNPs) che insieme all'mRNA formano lo spliceosoma.



# Spliceosoma

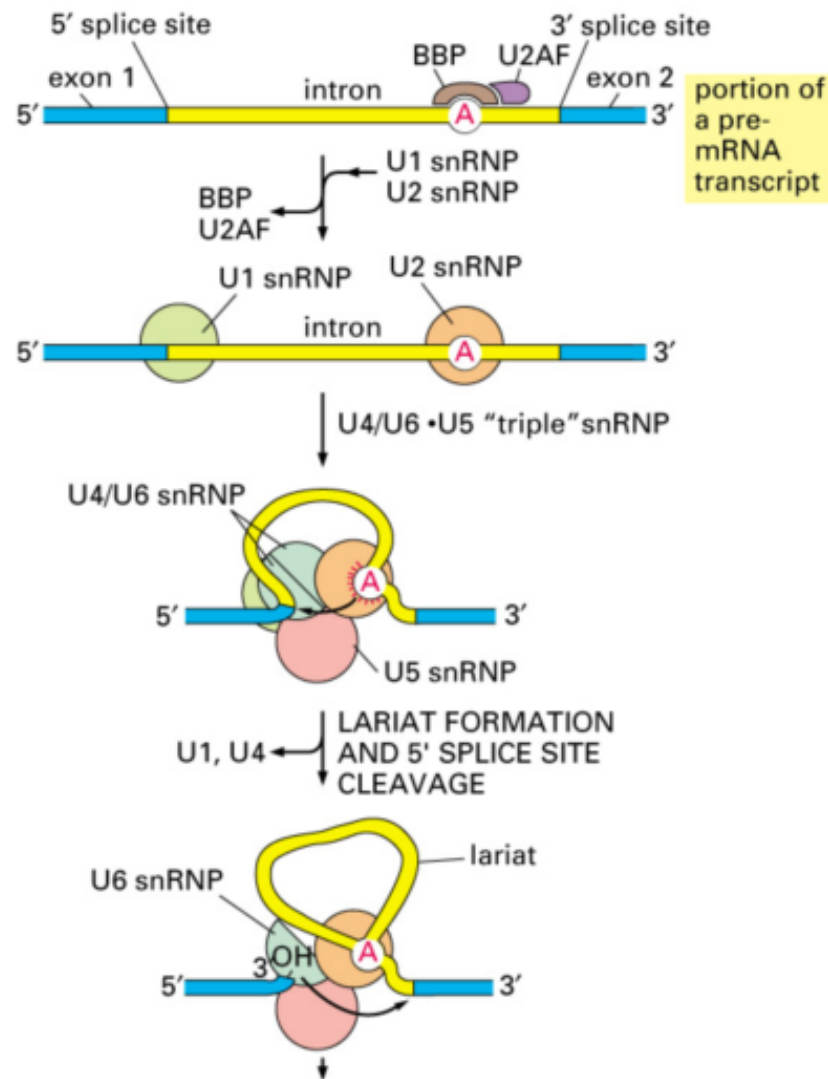


Figure 6-29 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

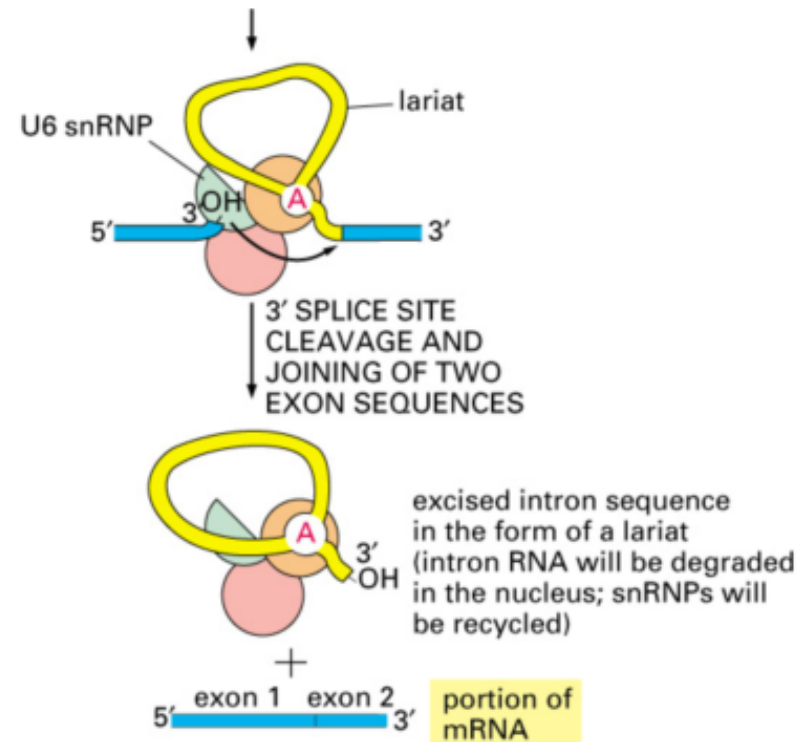
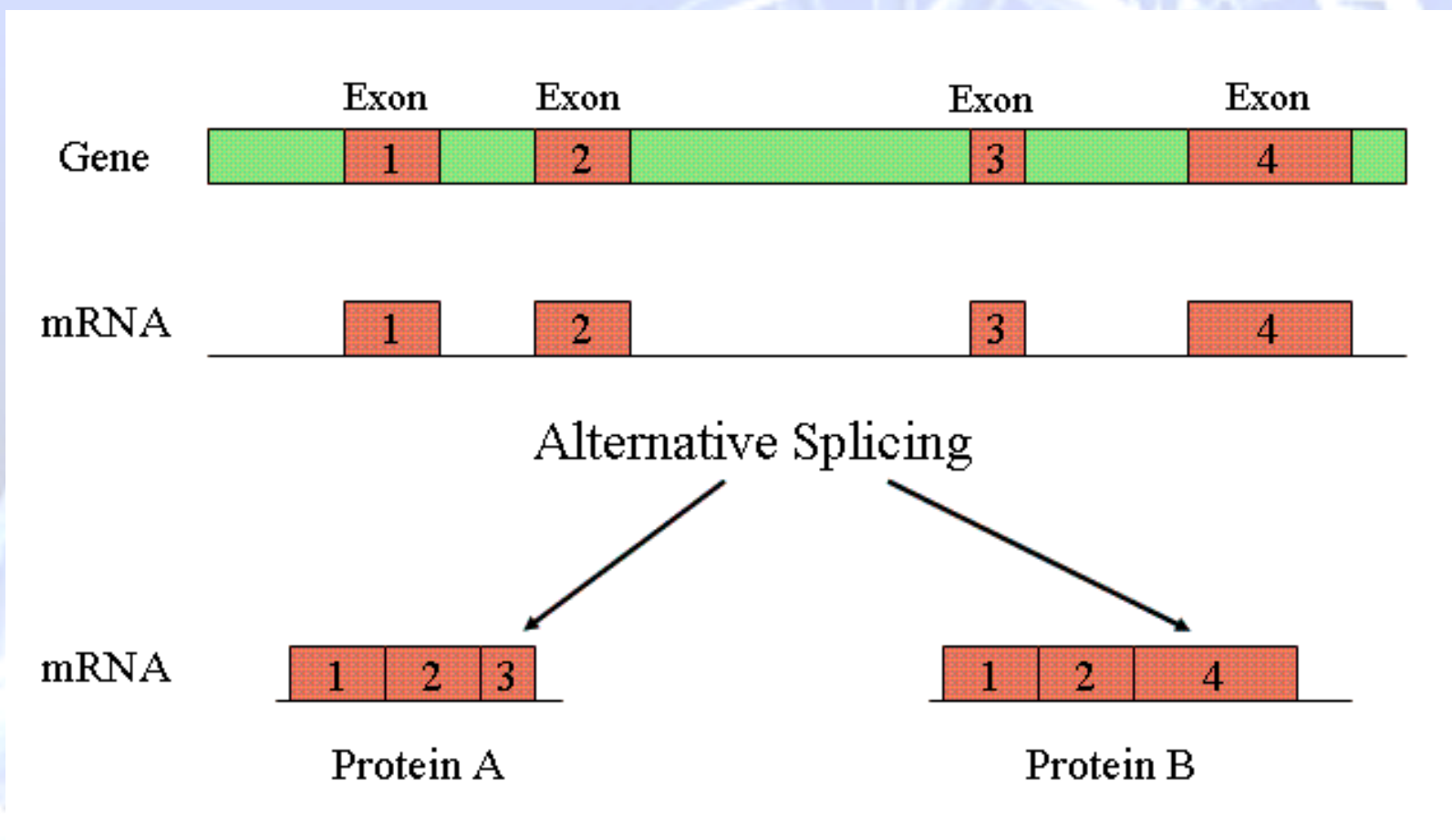


Figure 6-29 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

# Splicing alternativo

Lo stesso gene può produrre diverse proteine grazie all'assemblaggio di una serie differente di esoni.



# Alternative Splicing

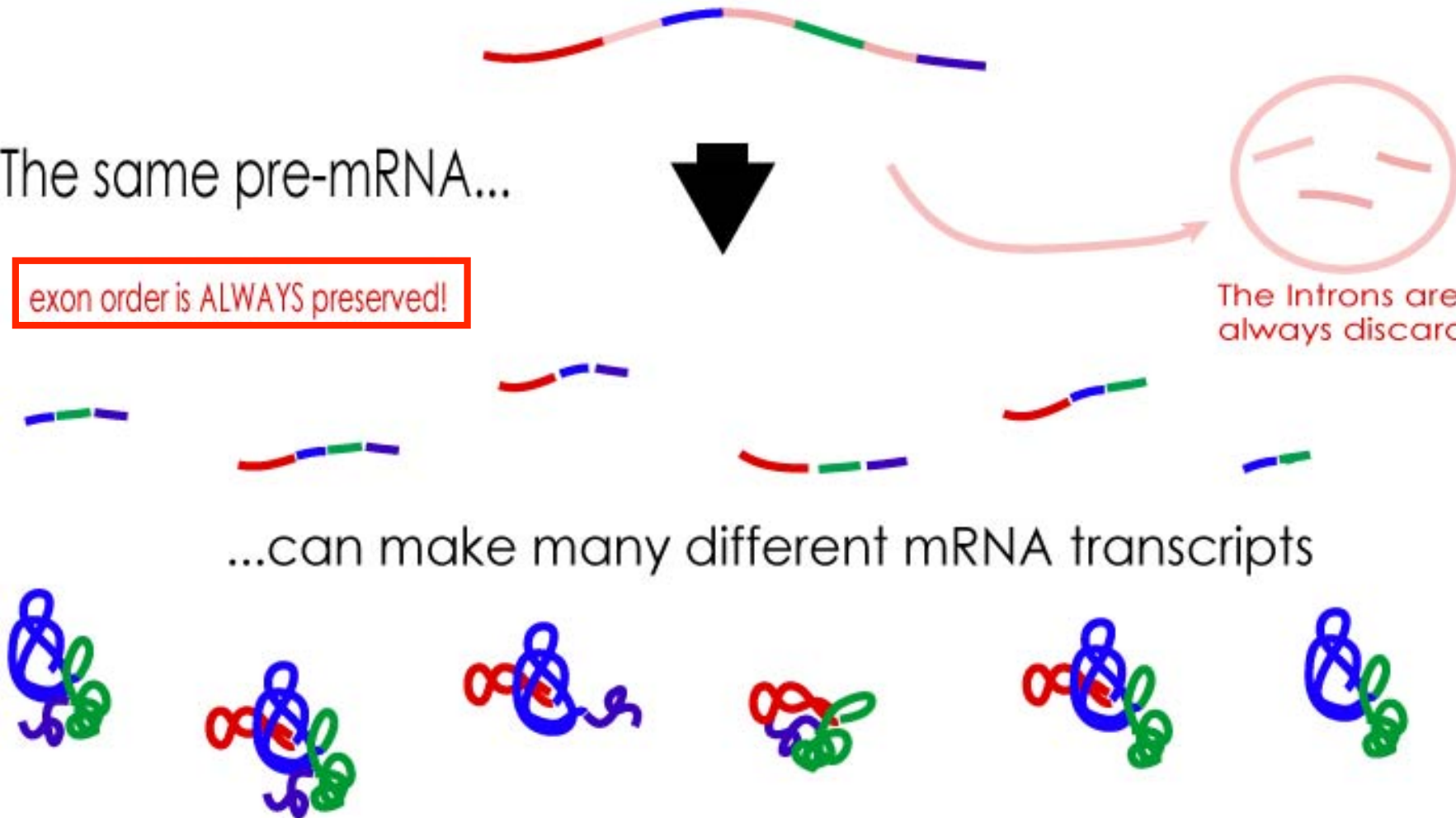
The same pre-mRNA...

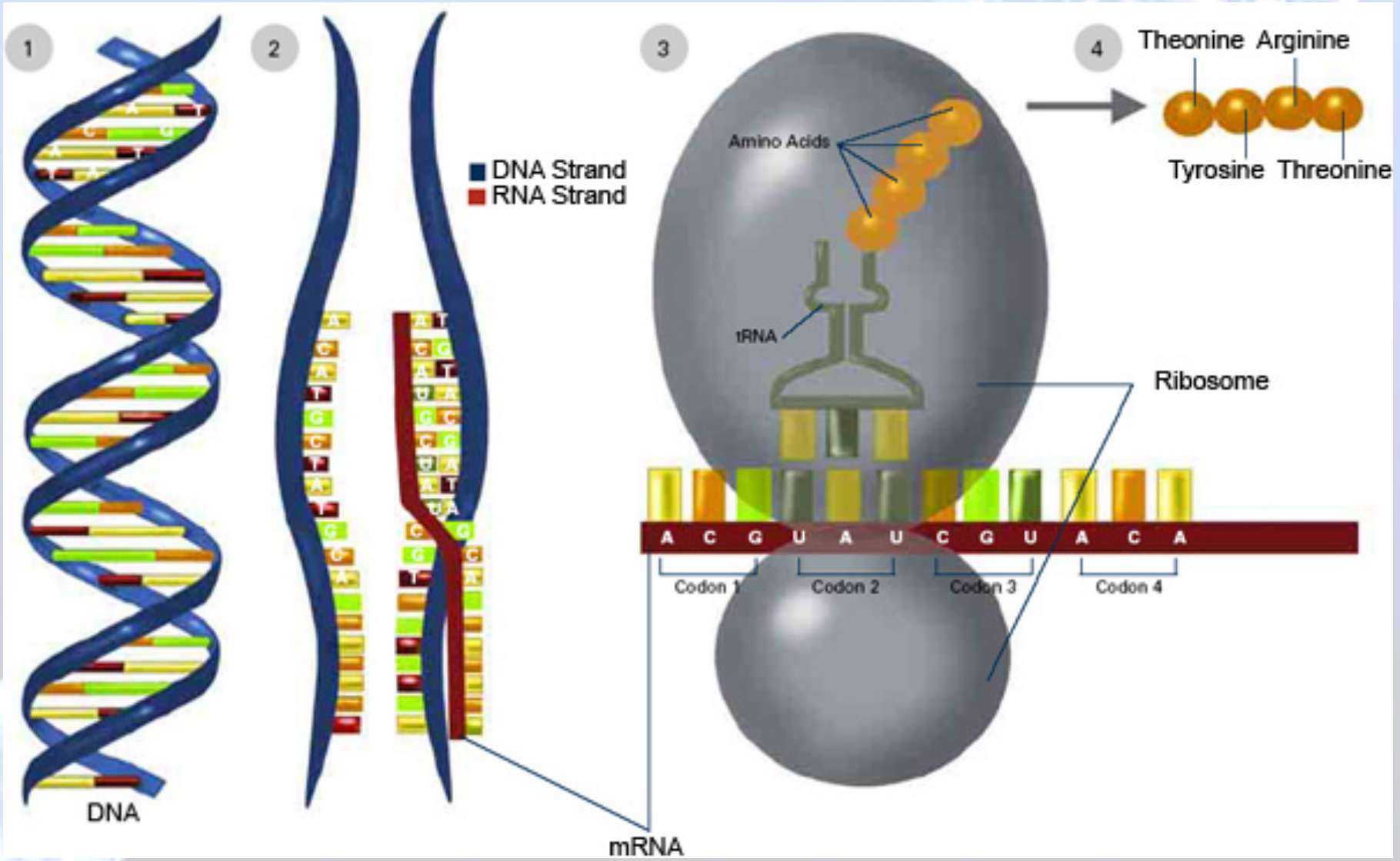
exon order is ALWAYS preserved!

The Introns are always discarded!

...can make many different mRNA transcripts

...and many different proteins





# Il codice genetico

La corrispondenza tra le due sequenze è stata decodificata già negli anni '60 del XX secolo.

In particolare è stato chiarito che ogni amminoacido è codificato da una sequenza di tre basi, detta **tripletta** o **codone**.

Tale corrispondenza tra amminoacidi e codoni è detta **codice genetico**.

## SYNONYMS AND SIMILARITIES

Many of the 64 possible three-letter codons specify the same amino acid, providing alternative ways for genes to spell out most proteins. These synonymous codons tend to differ by just a single letter, usually the last, forming a pattern of blocks. Codons for amino acids with similar affinities for water also tend to differ by their last letter, and codons sharing the same first letter often code for amino acids that are products or precursors of one another. These features, as it turns out, are crucial to the survival of all organisms and may even help speed their evolution. **The genetic code has evolved to minimize the harm caused by errors in the genes or in the translation process.**

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



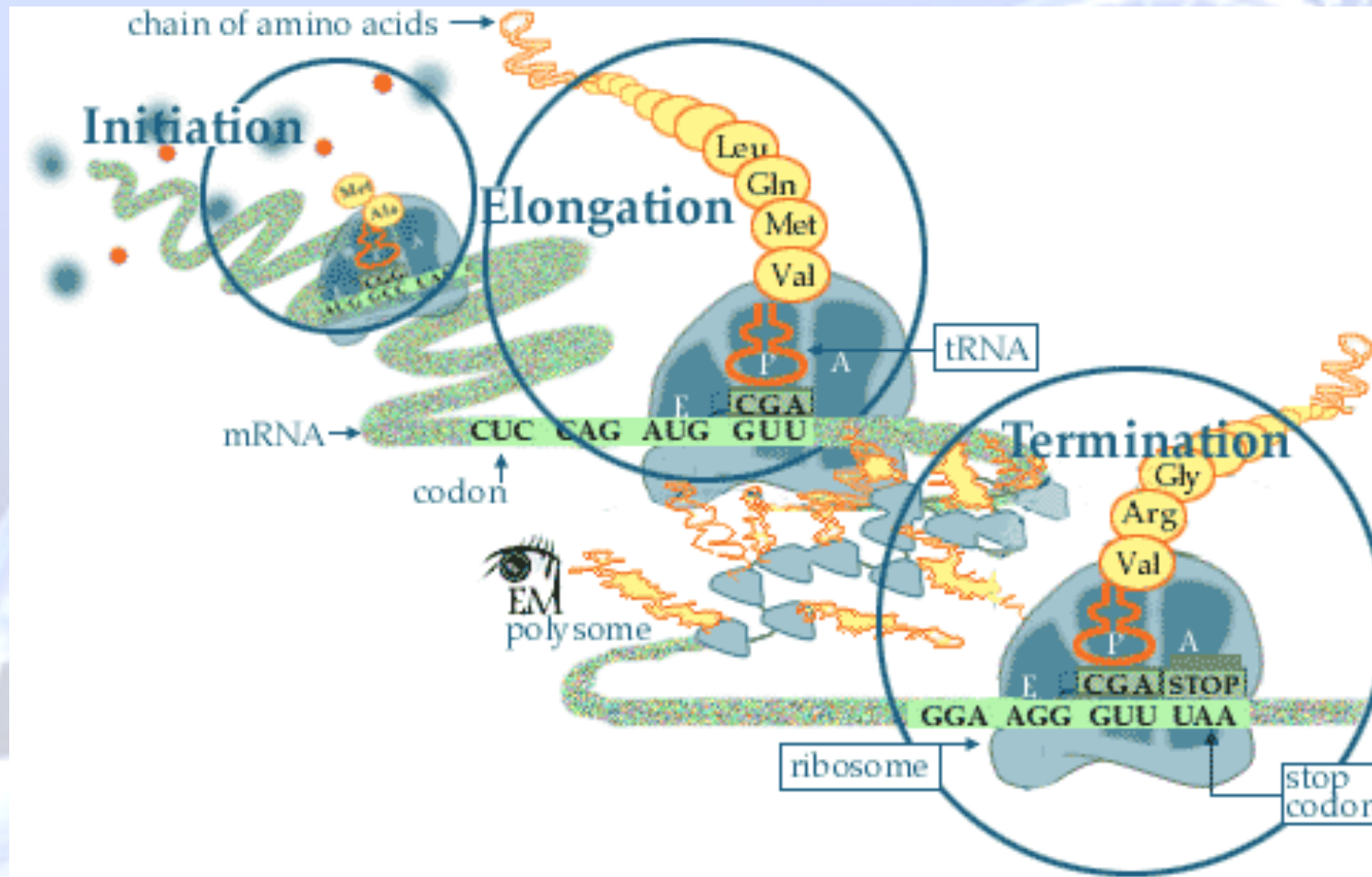
# Frame di lettura

Su un tratto di DNA ci sono sei possibili frame di lettura, tre per filamento.

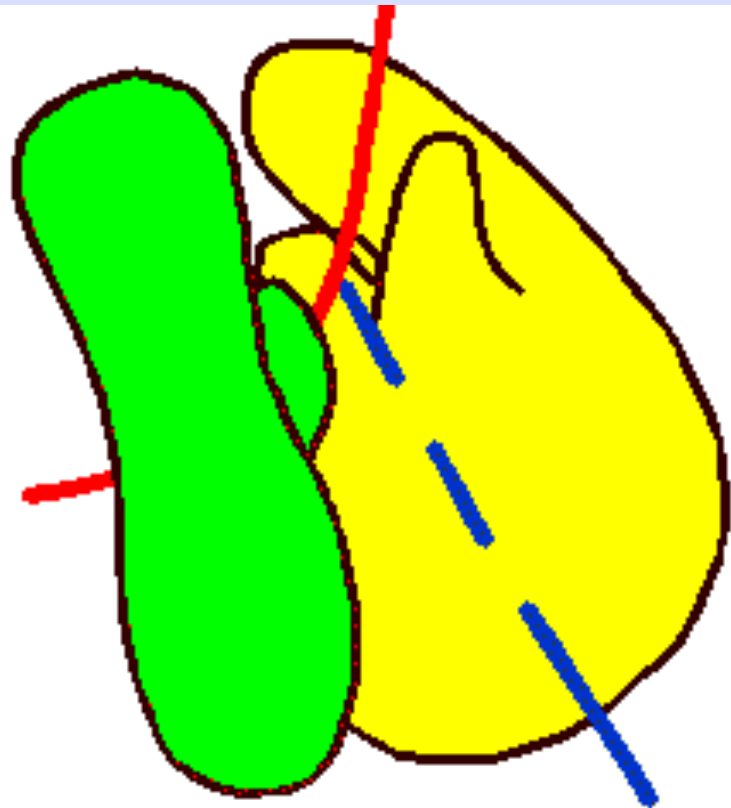






Spesso non più di un frame di lettura risulta aperto su un tratto di DNA. Infatti, nel corso dell'evoluzione solo il frame codificante viene mantenuto invariato dalla pressione selettiva. Negli altri frame si accumulano mutazioni che li chiudono.

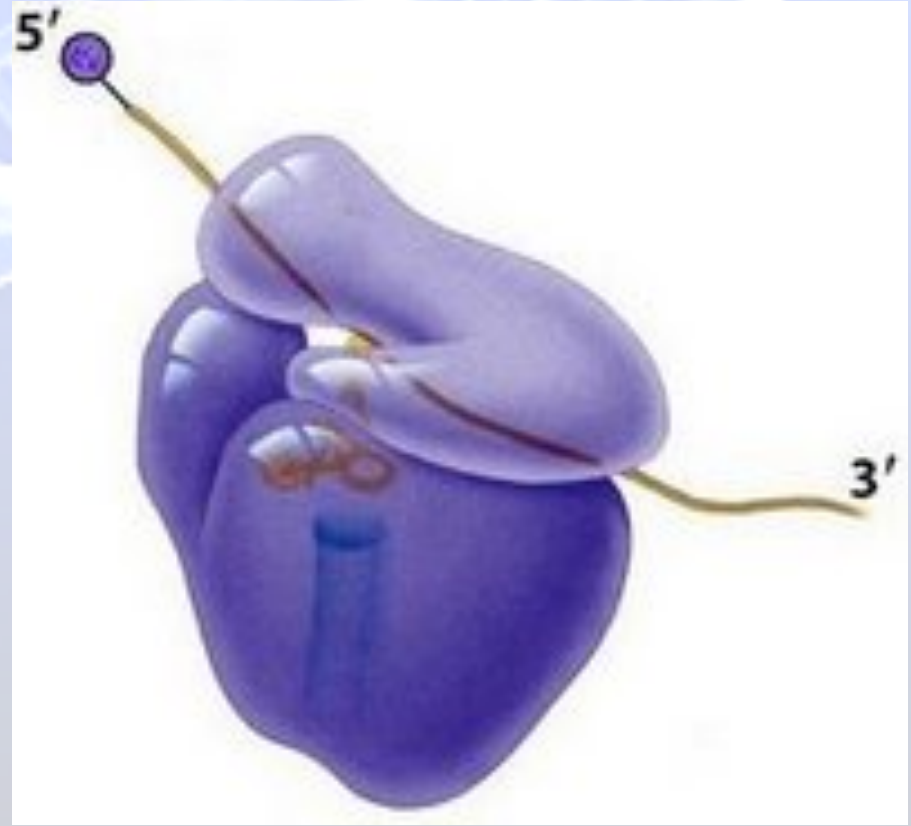
# Sintesi proteica



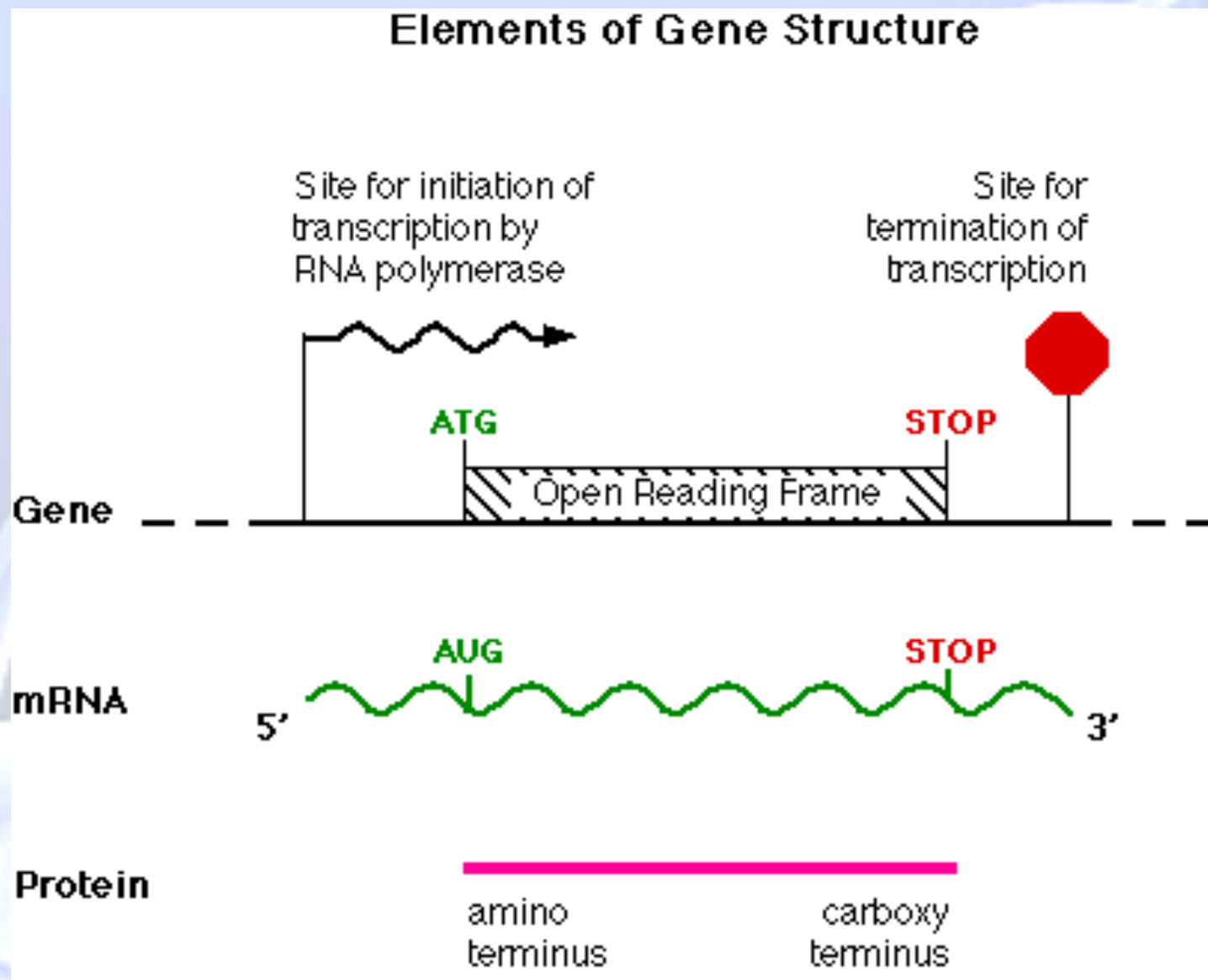
Il processo di sintesi delle proteine coinvolge l'attività catalitica di RNA e proteine.



-  small subunit
-  large subunit
-  amino acid chain
-  messenger RNA



# Sintassi dei geni



# Flusso dell'informazione genetica

**prokaryotes**

DNA



mRNA



protein

**eukaryotes**

DNA

TRANSCRIPTION ↓

pre-mRNA

↓ SPLICING

mRNA

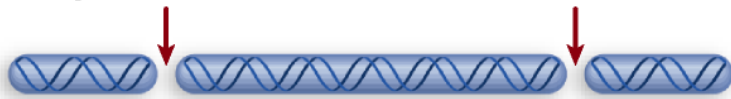


protein

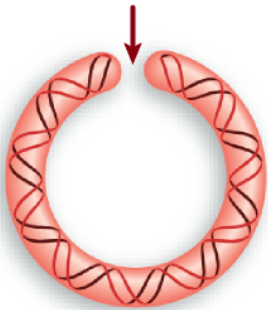
# Le tecniche del DNA ricombinante

## Il DNA viene clonato tramite il suo inserimento in un vettore

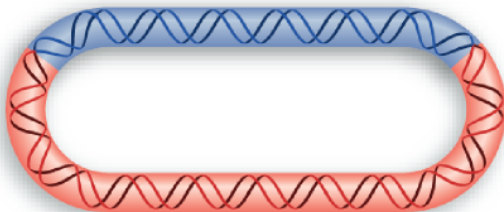
1. Si taglia il DNA donatore



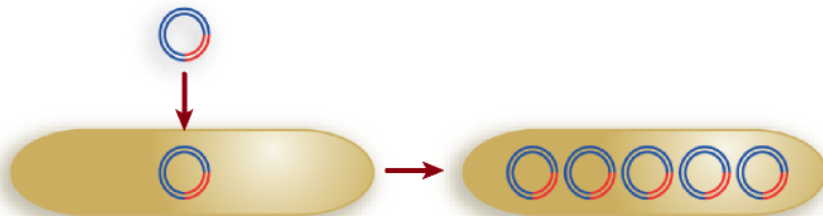
2. Si taglia il DNA vettore



3. Si unisce il DNA bersaglio al vettore



4. Si amplifica il DNA ibrido in cellule di *E. coli* o di lievito



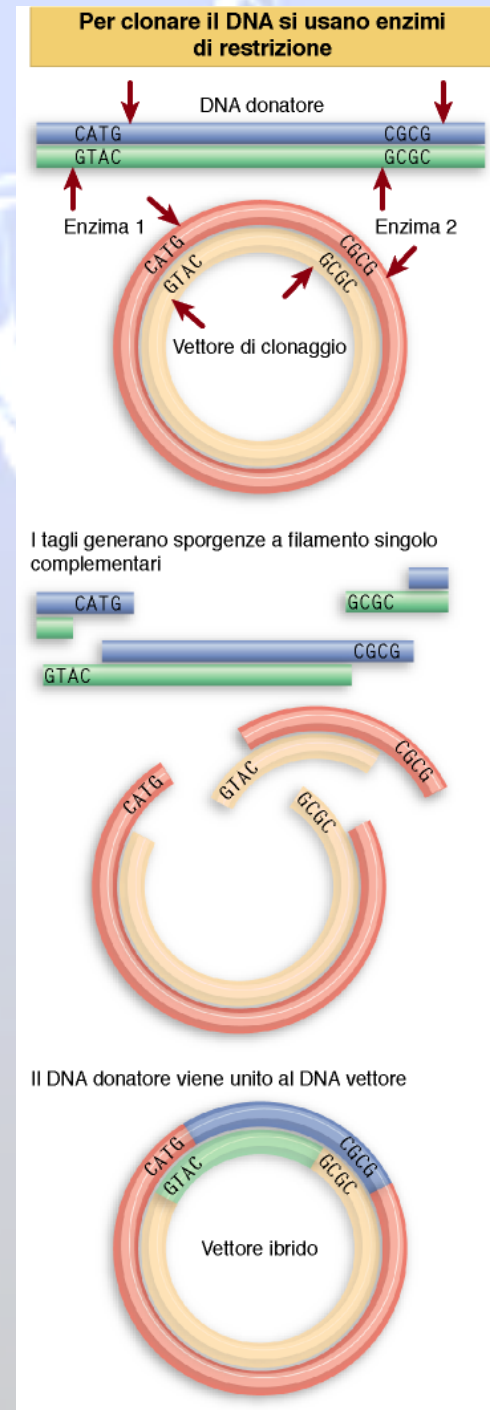
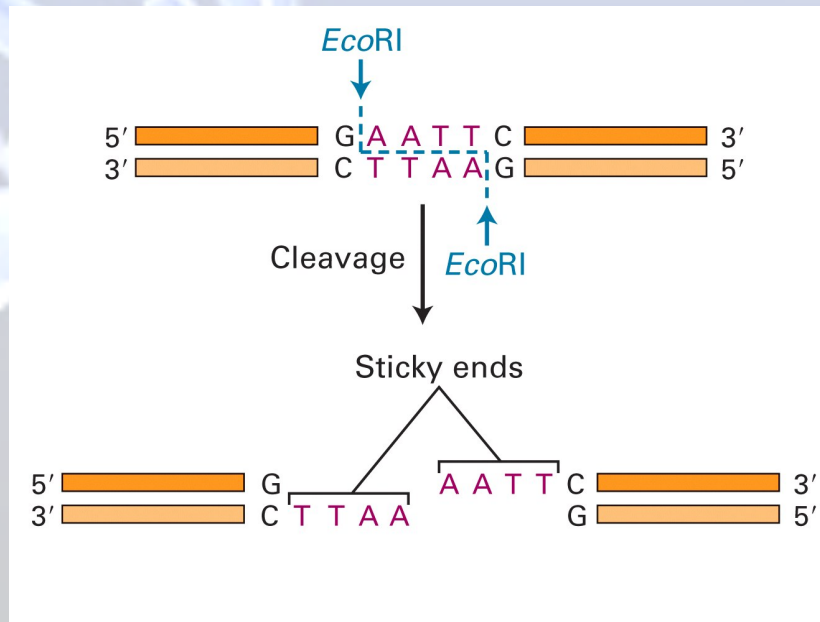
Il **clonaggio** di un frammento di DNA consiste nell'isolarlo dal suo contesto originale per inserirlo in un altro contesto, in genere un vettore, che ne permette la propagazione.

Un **vettore di clonaggio** è una molecola in grado di essere autonomamente replicata e mantenuta all'interno di un ospite, di solito un batterio (*E. coli*). Deve quindi possedere una **origine di replicazione** e un **marcatore selezionabile**, che per i batteri è solitamente un gene che conferisce una resistenza ad un antibiotico.

# Gli enzimi di restrizione

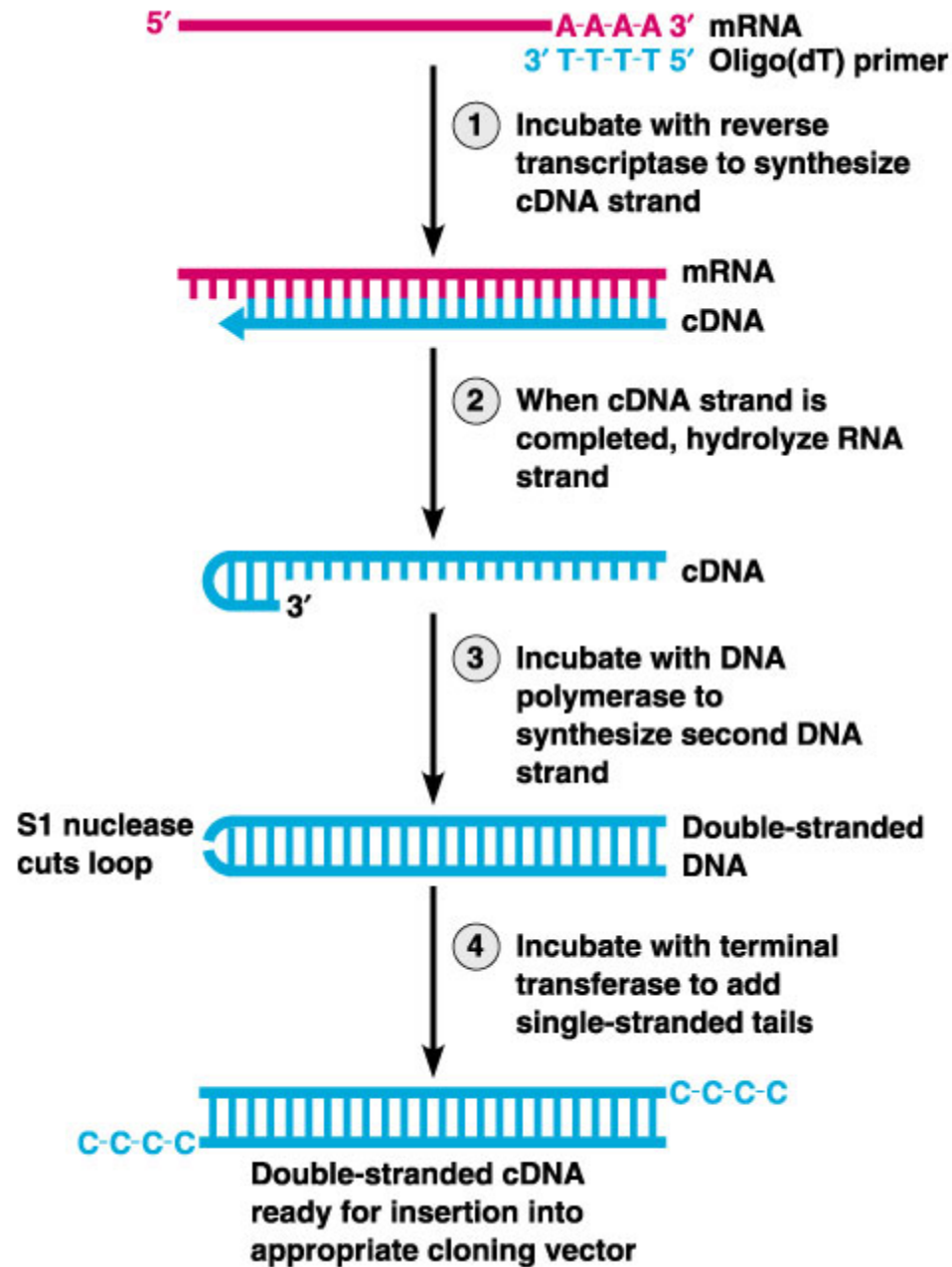
Gli enzimi di restrizione sono delle **endonucleasi** di origine batterica che tagliano entrambi i filamenti del DNA in un punto specifico all'interno di una sequenza palindromica di 4-8 bp. Ogni enzima riconosce una sequenza specifica detta **sito di restrizione** e taglia lasciando estremità tipiche, che quindi permettono a due molecole trattate con lo stesso enzima di essere ricombinate tra loro.

Queste caratteristiche degli enzimi di restrizione hanno permesso lo sviluppo delle tecniche del DNA ricombinante.

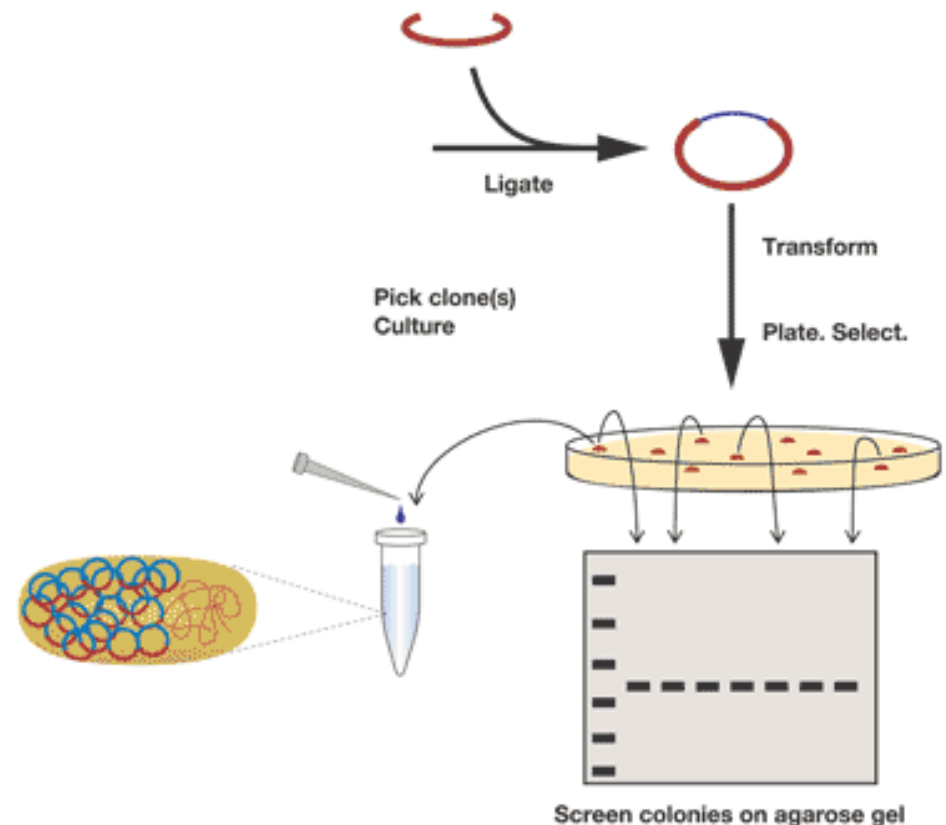


# Le librerie di cDNA

Il cDNA (DNA copia o complementare) è una molecola di DNA retrotrascritta da un mRNA maturo. Contiene perciò le sequenze codificanti, ma anche le sequenze in 3' e in 5' sul messaggero. Non contiene invece gli introni.

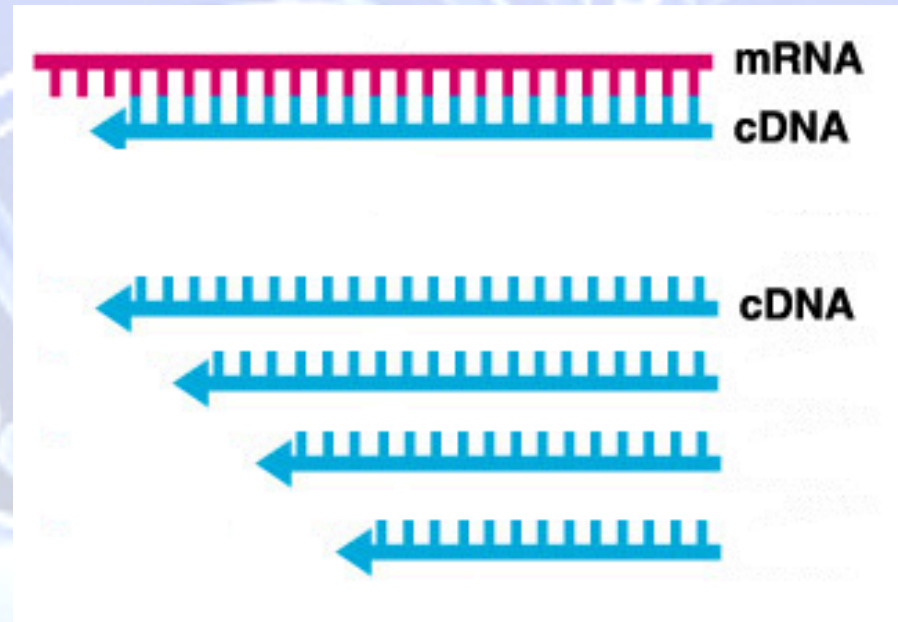


©Addison Wesley Longman, Inc.



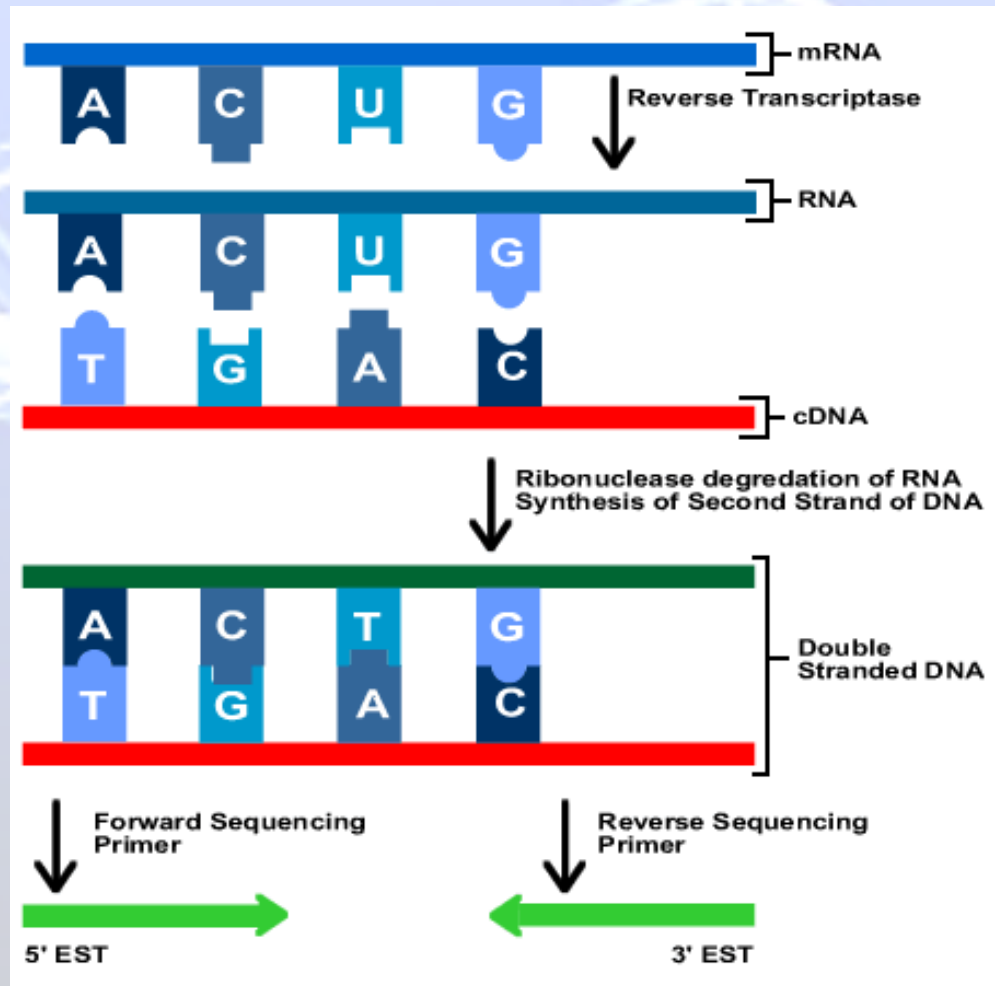


Se la retrotrascrizione termina precocemente si possono ottenere sequenze parziali, che mancano di alcuni tratti iniziali.



# Le EST: Expressed Sequence Tags

Sono brevi (200-500 bp) sequenze di DNA derivate dal sequenziamento rapido di cDNA, non complete ma sufficienti per identificare un trascritto. Possono contenere molti errori e perciò si considerano degenerate. Permettono di ottenere informazioni sull'espressione in determinati organismi o tessuti.



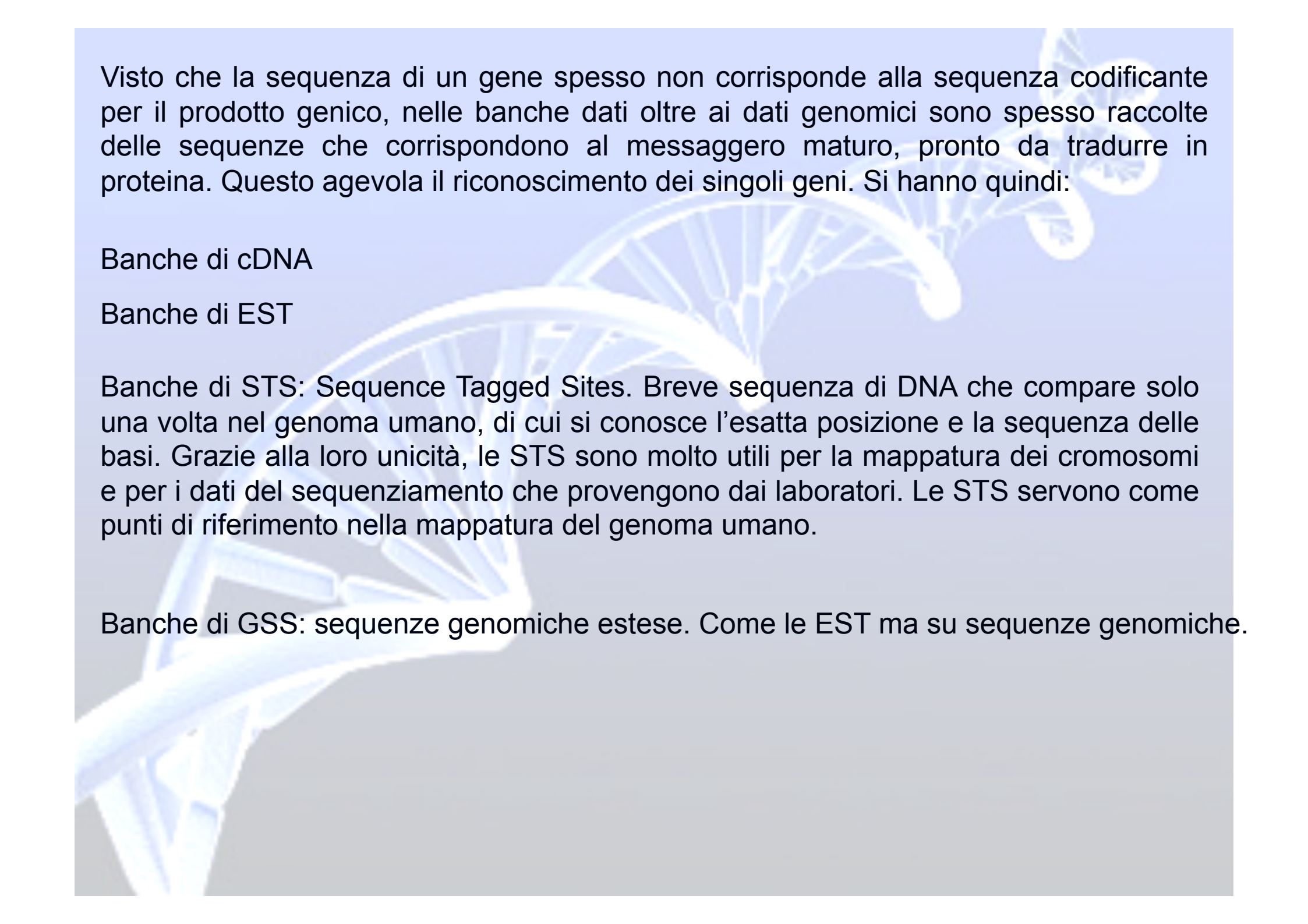
# L'informazione biologica



Un gene non è solo una sequenza di basi: accanto a questa informazione ci sono annotazioni indispensabili (posizione degli introni, splicing alternativo, presenza di siti nel promotore, eventuali enhancers o silencers...) o accessorie (profili di espressione in diversi tessuti/organismi/tempi...). Inoltre, per ogni prodotto genico ci saranno un'altra serie di informazioni che vanno dalla sequenza proteica nuda e cruda ad una serie di altri dati biologici quali le modificazioni cui va incontro, la struttura tridimensionale, la vita media, la localizzazione, le interazioni con altre proteine quali effettori, regolatori, substrati etc.

Tutta questa mole di informazioni tuttavia non dice nulla sul ruolo di un gene o di un prodotto genico all'interno della cellula: la funzione cellulare si può intuire solo dall'integrazione di tutti i dati a disposizione.

Il ruolo della bioinformatica è quello di creare strumenti che permettano di navigare attraverso la complessità biologica in modo da permettere al biologo di integrare tutti i vari dati biologici, tenendo conto della loro diversa natura ed affidabilità.



Visto che la sequenza di un gene spesso non corrisponde alla sequenza codificante per il prodotto genico, nelle banche dati oltre ai dati genomici sono spesso raccolte delle sequenze che corrispondono al messaggero maturo, pronto da tradurre in proteina. Questo agevola il riconoscimento dei singoli geni. Si hanno quindi:

Banche di cDNA

Banche di EST

Banche di STS: Sequence Tagged Sites. Breve sequenza di DNA che compare solo una volta nel genoma umano, di cui si conosce l'esatta posizione e la sequenza delle basi. Grazie alla loro unicità, le STS sono molto utili per la mappatura dei cromosomi e per i dati del sequenziamento che provengono dai laboratori. Le STS servono come punti di riferimento nella mappatura del genoma umano.

Banche di GSS: sequenze genomiche estese. Come le EST ma su sequenze genomiche.