

- Le prossime lezioni saranno incentrate sulla struttura delle principali macromolecole della cellula (DNA, RNA e proteine) e sul processo di traduzione della sequenza di DNA in proteine.
- Lo scopo è quello di dare una rapida visione a questi argomenti (che verranno poi ripresi molto più approfonditamente in corsi specifici) in modo da facilitare la comprensione dei successivi argomenti del corso.

### **Concetti Biologici basilari**

Per approfondire <http://www.ebi.ac.uk/2can/biology/index.html> (in inglese)

## *Le -OMICS*

La bioinformatica, in particolare, si occupa di gestire ed analizzare i dati che vengono prodotti in modo sistematico nelle numerose e più disparate discipline biologiche, quelle a cui spesso ci si riferisce, forse un po' esagerando, come **-OMICS**.

Ad esempio, la Genomica è la disciplina che si occupa di produrre, gestire ed analizzare i dati del genoma.

INSIEME DEI DATI	DISCIPLINA
<b>GENOME</b>	<b>GENOMICS</b>
<b>PROTEOME</b>	<b>PROTEOMICS</b>
<b>TRASCRIPTOME</b>	<b>TRASCRIPTOMICS</b>
<b>METABOLOME</b>	<b>METABOLOMICS</b>
<b>BIBLIOME</b>	<b>BIBLIOMICS</b>

Curiosità. Un sito che elenca tutte le -omics citate in letteratura:  
<http://www.genomicglossaries.com/content/omes.asp>

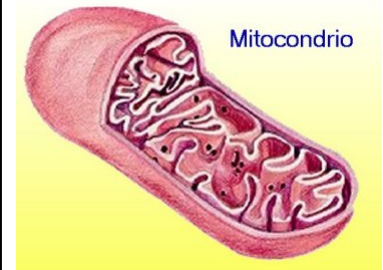
# Procarioti ed Eucarioti

**Procarioti** (conosciuti anche come microbi) sono organismi unicellulari con una organizzazione relativamente semplice. Non contengono particolari organelli caratteristici degli eucarioti. Il materiale genetico (DNA) non è racchiuso in una particolare struttura.

Gli **Eucarioti** hanno un nucleo dove è contenuto il DNA ed hanno dei compartimenti interni, racchiusi da membrane, chiamati organelli, che assolvono a particolari compiti biologici (complesso del Golgi, lisosomi, **mitocondri** ecc.)

**Mitocondri:** Organelli cellulari, racchiusi da due membrane, posseggono un proprio DNA circolare a doppia elica (genoma mitocondriale). Una loro importante funzione è la produzione di energia (sotto forma di ATP) attraverso l'ossidazione di substrati organici.

Si pensa che, in origine, i mitocondri fossero dei batteri, inglobati dalle cellule eucariote con conseguente mutuo beneficio. Successivamente i batteri avrebbero trasferito gran parte del loro materiale genetico a quello cellulare, divenendo così, mitocondri (teoria endosimbiontica).



## Principali molecole biologiche

-**Lipidi**: grazie alle loro caratteristiche chimiche formano membrane che racchiudono le cellule e gli organelli cellulari negli eucarioti.

-**Proteine**: svolgono quasi tutte le funzioni biologiche: formano strutture citoscheletriche; catalizzano alcune reazioni chimiche (enzimi); forniscono attività motorie alle cellule e nelle cellule; dirigono l'esportazione, l'importazione e lo spostamento di varie molecole; trasmettono particolari segnali tra i vari compartimenti cellulari, oppure come ormoni o fattori di crescita trasmettono segnali di controllo a differenti cellule; agiscono come anticorpi, enzimi digestivi; costituiscono tossine e veleni naturali; ecc.

-**Acidi Nucleici** (DNA e RNA) : Il DNA codifica l'informazione per costruire le proteine, l'RNA fornisce lo stampo per la sintesi delle proteine, interviene nella formazione di macromolecole complesse ed in particolari processi biologici.

## Proteine in sintesi:

- Le proteine sono polimeri lineari di **aminoacidi**, uniti chimicamente l'uno all'altro tramite legami **peptidici**.
- Sono costituite essenzialmente da 20 possibili aminoacidi diversi
- La **sequenza** con cui gli aminoacidi si succedono l'uno all'altro determina le proprietà di ogni proteina.
- Le differenti combinazioni dei 20 aminoacidi consentono la formazione di innumerevoli sequenze proteiche (es. proteina di 50 aa può avere  $20^{50}$  sequenze differenti)
- Esistono proteine di lunghezze molto diverse, da pochi aminoacidi (in questo caso sono generalmente chiamate **peptidi**) a diverse migliaia di aminoacidi (Le proteine più comuni sono lunghe 50-1000 aa)

### Dove trovare informazioni 'gratis'

breve corso sulla struttura delle proteine: <http://webhost.bridgew.edu/fgorga/proteins/default.htm>  
A Review of Amino Acids <http://wbiomed.curtin.edu.au/teach/biochem/tutorials/AAs/AA.html>

**Sequenza di una proteina** rappresenta la struttura primaria ed è l'ordine con cui gli aminoacidi si succedono nella molecola

La sequenza viene rappresentata con una 'stringa' di caratteri che rappresentano i simboli degli aminoacidi

es.: P R T W Q E R P R R T W C S S G R

**In una proteina la sequenza di aminoacidi ha una direzione.** Per convenzione la sequenza si scrive a partire dall'estremità **NH<sub>2</sub>-(amino)terminale** all'estremità **COOH (carbossi)terminale** (corrisponde alla direzione di sintesi)

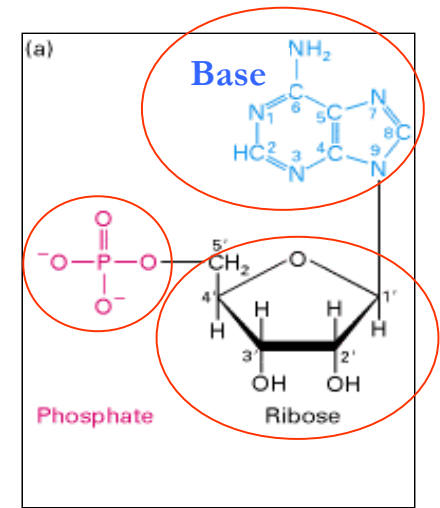
Quindi la sequenza 'ACDE' è diversa da 'EDCA'



## Acidi nucleici (DNA e RNA)

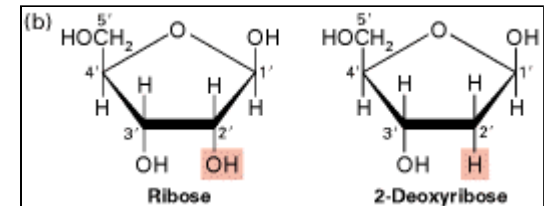
Gli acidi nucleici (DNA: *acido deossiribonucleico* e RNA: *acido ribonucleico*) sono dei polimeri organici costituiti da monomeri chiamati **nucleotidi**.

Tutti i nucleotidi sono costituiti da tre componenti fondamentali: un **gruppo fosfato**, una molecola di **zucchero pentoso** (**deossiribosio nel DNA o ribosio nell'RNA**) e una **base azotata** che si lega al carbonio 1' dello zucchero.

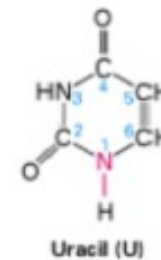
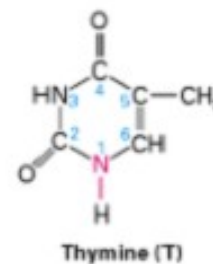
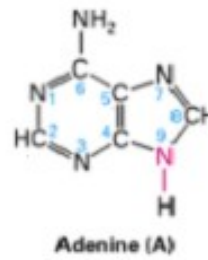
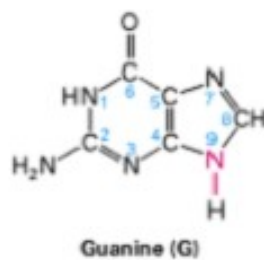
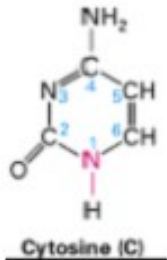


RNA

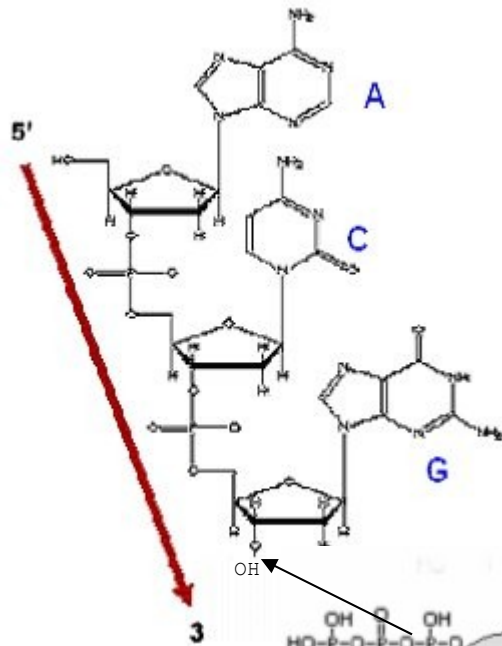
DNA



Gli acidi nucleici possono essere formati da solo quattro possibili basi azotate: **adenina**, **guanina**, **citrosina** (comuni al DNA e all'RNA), la **timina** presente solo DNA mentre l'**uracile** solo nel RNA.



## Direzionalità delle sequenze di acidi nucleici



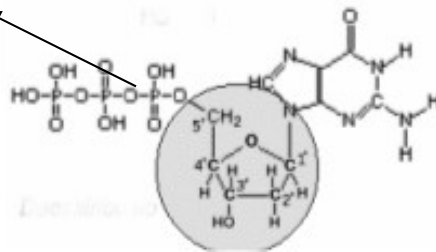
La sequenza di DNA (o RNA) rappresenta l'ordine con cui i differenti nucleotidi si succedono nella catena

Le sequenze hanno una direzione, sono generalmente scritte in direzione 5'-3' (**direzione di sintesi**)

Ogni nuovo nucleotide viene aggiunto al 3' dello zucchero attraverso il gruppo fosfato. Quindi la catena cresce in direzione 5'→3' (facendo riferimento agli atomi di carbonio dello zucchero)

La sequenza in figura viene quindi scritta come segue:

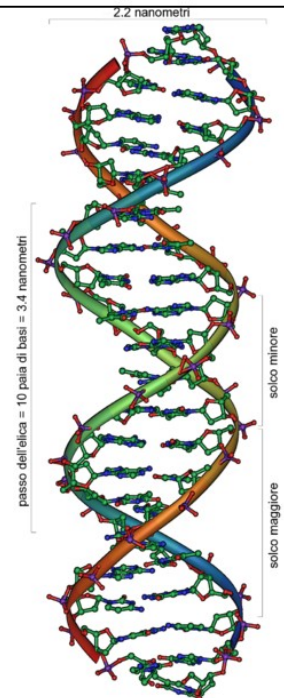
**5'-ACG-3'**



## DOPPIA ELICA

Il DNA esiste prevalentemente in forma di doppia elica (2 molecole di DNA appaiate ed avvolte tra loro).

Questa forma rende il DNA chimicamente più 'stabile' e più facilmente 'compattabile' nelle cellule





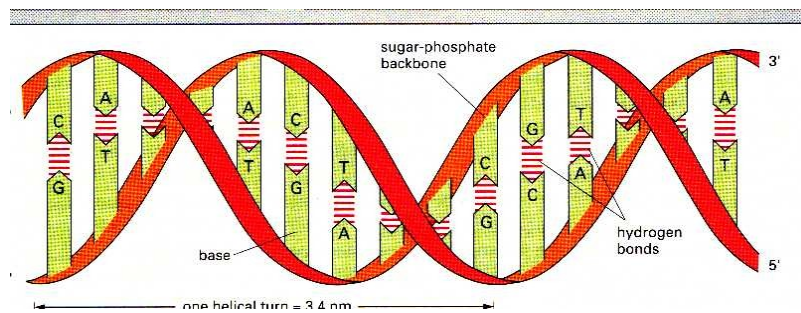
Nel DNA la doppia elica si forma per l'appaiamento delle basi (mediante legami idrogeno) dei due differenti filamenti

In particolare:

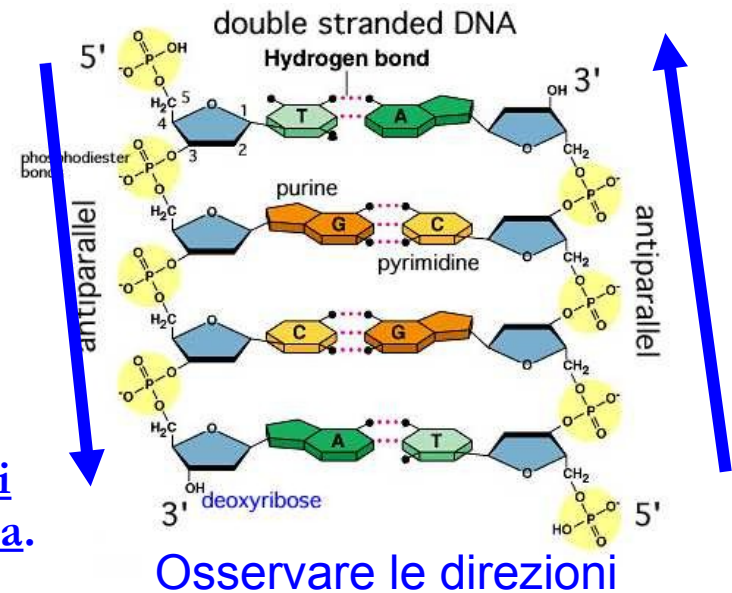
-Una adenina si appaia con una timina (appaiamento A-T o T-A)

-Una citosina si appaia con una guanina (appaiamento C-G o G-C)

Le basi che si appaiano tra loro si dicono **complementari** (C è complementare a G e viceversa, ecc.)



**Nota:** Se si conosce la sequenza di un'elica si può ricavare anche la sequenza dell'altra elica.



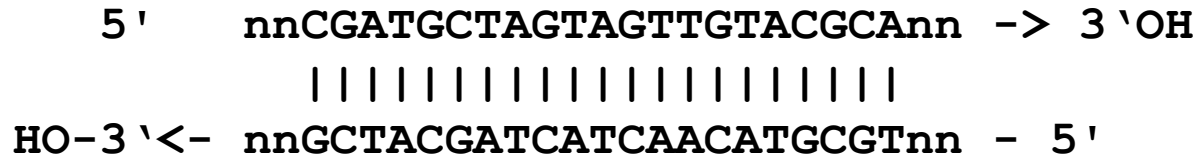
Osservare le direzioni

Negli eucarioti, il DNA si dispone all'interno del nucleo in strutture chiamate **cromosomi**.

Nell'uomo ci sono 23 coppie di cromosomi, di cui 22 sono cromosomi somatici non sessuali (autosomi) mentre una coppia di cromosomi risultano **diversi**: i cromosomi sessuali (eterosomi).

## *Direzionalità e rappresentazione delle molecole di DNA*

Una rappresentazione in formato testo di una doppia elica di DNA



Normalmente è sufficiente rappresentare una sequenza di DNA con una sola stringa (per la complementarità, se si conosce la sequenza di un'elica si può ricavare anche la sequenza dell'altra elica)



### Esempi:

Sequenza d'origine

5' CGATGCTAGTAGTTGTACGCA 3'

Sequenza complementare

3' GCTACGATCATCAACATGCGT 5'

Sequenza inversa-complementare

5' TGCGTACAACACTAGCATCG 3'

[http://promix.cribi.unipd.it/cgi-bin/promix/promix\\_menu.pl](http://promix.cribi.unipd.it/cgi-bin/promix/promix_menu.pl) --> Sequence Utility

Quando la sequenza indica una singola elica lunga pochi nucleotidi, viene detta

### **OLIGONUCLEOTIDE**

Il DNA può essere costituito da catene lunghissime di basi, anche centinaia di milioni!!!  
(nell'uomo circa 3.3 miliardi di basi contenute in 46 cromosomi)

**Genoma:** indica l'intero patrimonio genetico (DNA) di un organismo vivente che si trova in tutte le sue cellule o in particolari organelli (mitocondri e cloroplasti).

**Genomica:** la scienza che studia, definisce e caratterizza il corredo genico nel suo complesso

## DUPLICAZIONE, TRASCRIZIONE E TRADUZIONE

DNA  $\xleftrightarrow{\text{Duplicazione}}$  DNA

↓  
*Trascrizione*

RNA

↓  
*Traduzione*

Proteina

- Il processo di **DUPLICAZIONE** porta alla formazione di copie delle molecole di DNA ed al trasferimento del materiale genetico.

- Il processo di **TRASCRIZIONE** è il trasferimento dell'informazione dal DNA alle molecole di RNA.

- La **TRADUZIONE** è il processo mediante il quale dall'RNA si passa alla sintesi delle proteine

# TRASCRIZIONE (Transcription)

## Processo nel quale l'RNA è sintetizzato a partire dal DNA stampo.

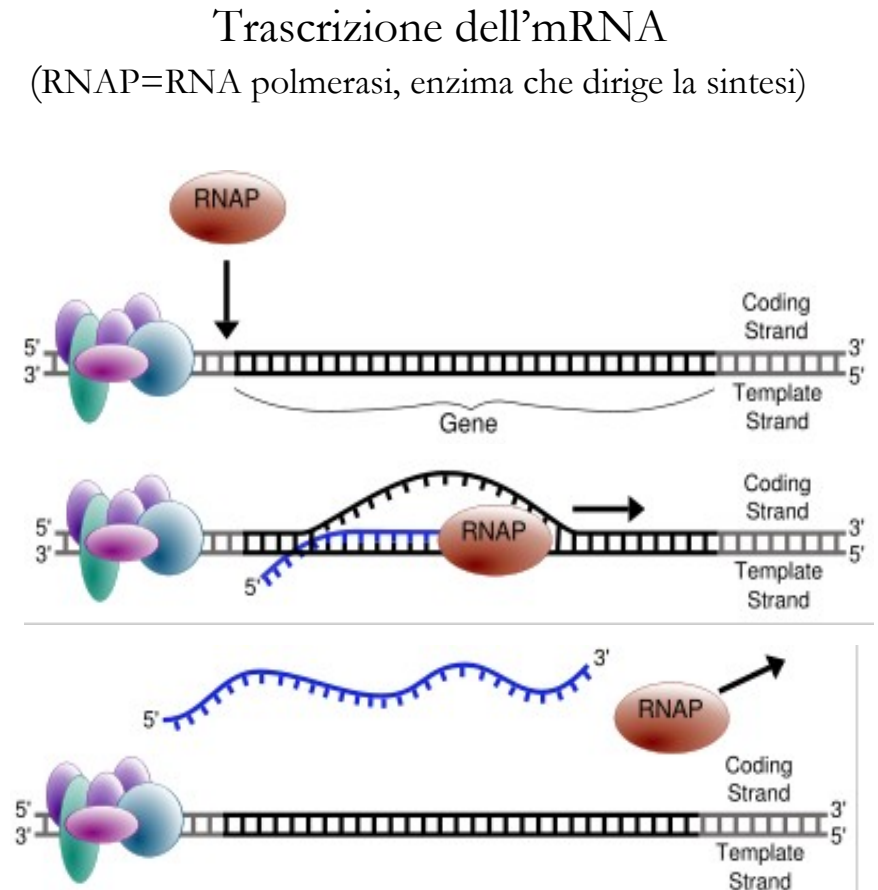
Con la trascrizione si ottengono molecole di RNA con differenti funzioni (es. intervengono nella sintesi proteica, nel silenziamento di geni, nella regolazione della trascrizione stessa):

mRNA (RNA messaggero), tRNA (RNA transfert), rRNA (RNA ribosomale), miRNA (micro RNA), snRNA (small nuclear RNA), ecc.

In particolare **l'mRNA (RNA messaggero)** rappresenta la molecola in cui è copiato il codice del DNA e che serve come stampo per la sintesi delle proteine

**Gene (o cistrone):** sequenza del DNA che determina la sequenza aminoacidica di una proteina.

DNA  
↓  
RNA  
↓  
Protein



## Junk-DNA? (DNA spazzatura?)

Non tutto il DNA viene trascritto in RNA. Alcune parti del DNA forniscono informazioni su:

- inizio (**segnale d'inizio**) e fine (**segnale di stop**) della trascrizione,
- regolazione della trascrizione (nello stesso organismo, non tutti i geni vengono trascritti in tutte le cellule) → **promotori**, **repressori** della trascrizione

Generalmente, particolari sequenze segnale indicano la fine della trascrizione.

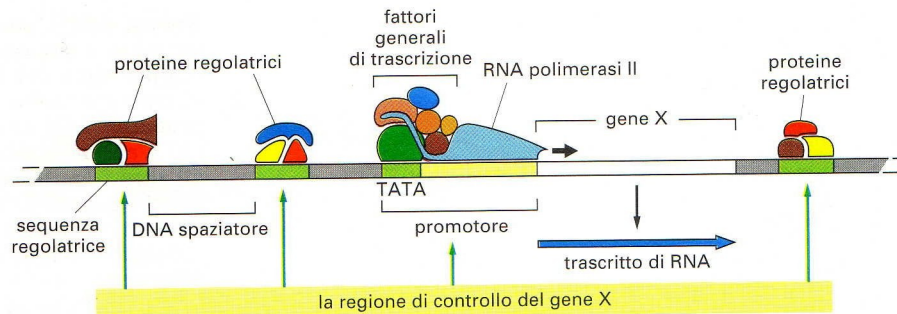
Negli eucarioti, un particolare sito provoca il taglio dell'mRNA nascente e innesca una reazione di poli-adenilazione. Generalmente gli mRNA eucarioti finiscono con una sequenza di 'A' ripetute (**poli A**).

Noterete questo particolare in laboratorio quando effettuerete le ricerche di trascritti nei database.

.....CTGCGCGAACTGCAAAACAAAAAAAAAAAAAAAAAAAAAAAAA (coda di poliA)

## Regolazione della trascrizione (espressione genica)

La trascrizione della maggior parte dei geni è regolata in modo specifico



## *Dall'RNA alle proteine: la traduzione (translation)*

Consideriamo la sequenza lineare di DNA: 5'-ATGATCAGAATCG.....3'

Quante basi servono per poter definire 20 aminoacidi:

- 1 base (A, T, G, A, T, C,.....) : solo 4 aminoacidi
- 2 basi (AT, GA, TC, AG,.....):  $4^2$  combinazioni = 16 aminoacidi, **non basta!**
- 3 basi (ATG, ATC, AGA,.....):  $4^3$  combinazioni = 64 aminoacidi, **anche troppi**, ma è proprio così.

### **Il codice genetico:**

fu decifrato negli anni '60.

Tutti gli organismi hanno essenzialmente lo stesso codice genetico con qualche piccola eccezione in casi molto particolari (ad esempio i mitocondri).

Il codice genetico viene perciò definito universale.

Il codice genetico è letto interpretando tre basi alla volta, **senza sovrapposizioni**: ogni gruppo di tre basi viene chiamato tripletta o più propriamente **codone**.

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC " TAA Ter TAG Ter	TGT Cys (C) TGC " TGA Ter TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " ATG Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

Dei 64 possibili **codoni**, 61 sono detti **codoni senso**, in quanto specificano degli aminoacidi, gli altri 3 codoni ("Ter") indicano la terminazione della sintesi proteica.

Ci sono 61 codoni per 20 aminoacidi; questo comporta che la maggioranza degli aminoacidi è rappresentata da più di un codone:

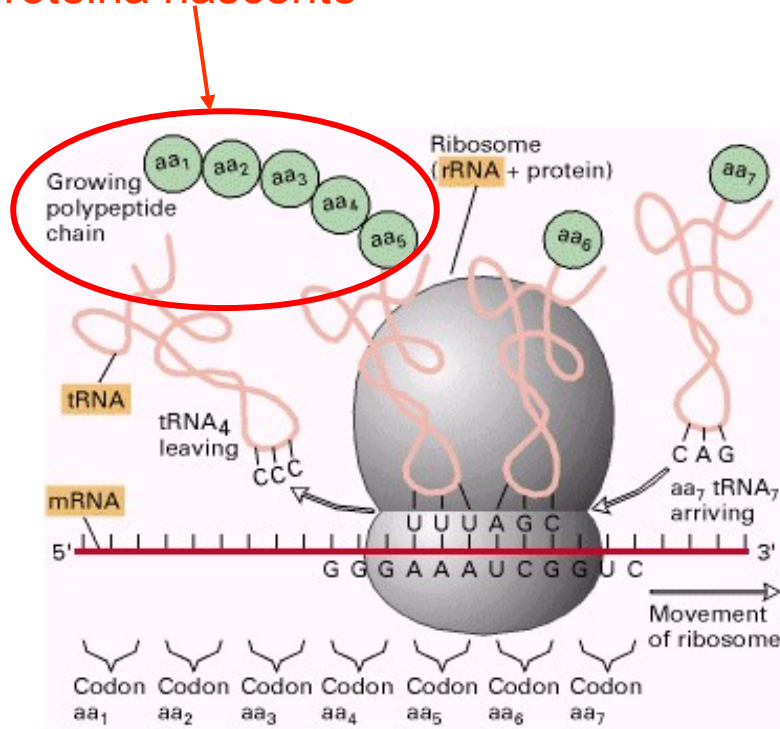
→ il codice genetico è detto **degenere**

Metionina (Met, M) e triptofano (Trp, W) dispongono ciascuno di un solo codone; rappresentano gli aminoacidi meno abbondanti nelle proteine.

**Importante: il codone ATG (AUG nell'RNA), della metionina, è il codone di inizio più comune, specifica l' AA all' N-terminale della catena proteica. Questa conoscenza è importante per individuare sequenze codificanti all'interno di lunghe sequenze di DNA**

# *Sintesi delle proteine*

Proteina nascente





# Sono trascritte anche regioni non tradotte

Sequenza di un trascritto (MYOZ 1) in formato FASTA ottenuta interrogando un database di sequenze nucleotidiche

```
>gi|21359948|ref|NM_021245.2| Homo sapiens myozenin 1 (MYOZ1), mRNA
GTTTCTCCCTAAGTGCTTCTTTGGATCTCAGGCTCTAGGTGCAATGTGAAGGGGAGTCCCTGGGCAGACTGATCCCTGGCTCAGAC
AGTTCAGTGGGAGAATCCCAAAGGCCTTTTCCCTCCTTCCCTGAGCCTCCGGGCAAGGAGGGAGGGATCTTGGTTCCAGGGTCTCAG
TACCCCTGTGCCATTTGAGCTGCTTGCCTCATCATCTCTATTAATAACCAACTTCCCTCCCCACTGCCAGTGTGCCCCACG
CCTGCCAGCTCGTGTCTCCGGTACAGCAGCTCAGTCTCCAAAGCTGCTGGACCCAGGGAGAGCTGACCACTGCCCGAGCAG
CCGGCTGAATCCACCTCCACAATGCCGCTCTCAGGAACCCCGGCCCTAATAAGAAGAGGAAATCCAGCAAGCTGATCATGGAACT
CACTGGAGGTGGACAGGAGAGCTCAGGCTTGAACCTGGGCAAAAAGATCAGTGTCCCAAGGGATGTGATGTTGGAGGAACTGTCGC
TGCTTACCAACCGGGGCTCCAAGATGTTCAAACCTGCGGCAGATGAGGGTGGAGAAGTTTATTTATGAGAACCACCCTGATGTTTTC
TCTGACAGCTCAATGGATCACTTCCAGAAGTTCCTTCCAACAGTGGGGGACAGCTGGGCACAGCTGGTCAGGGATTCTCATACAG
CAAGAGCAACCGGCAGAGGCGGCAGCCAGGCAGGGGGCAGTGGCTCTGCCGGACAGTATGGCTCTGATCAGCAGCACCATCTGGGCTC
CTGGGTCTGGAGCTGGGGGTACAGGTGGTCCCGCGGGCCAGGCTGGCAGAGGAGGAGCTGCTGGCACAGCAGGGGTTGGTGAGACAG
GGATCAGGAGACCAGGCAGGCGGAGAAGGAAAACATATCACTGTGTTCAAGACCTATATTTCCCATGGGAGCGAGCCATGGGGGTT
TGACCCCGAGCAAAAAATGGAACCTTGGCATTGACCTGCTGGCCTATGGGGCCAAAGCTGAACTTCCCAAATATAAGTCTTCAACAA
GGACGGCAATGCCCTATGGTGGATATGAGAAGGCCTCCAAACGCATGACCTTCCAGATGCCCAAGTTTGACCTGGGGCCCTTGCTGT
AGTGAACCCCTGGTCTCTACAACCAAAACCTCTCCAACAGGCCTTCTTTCAATCGAACCCCTATTCCCTGGCTGAGCTCTGGGGAG
GCCTGTAGACTACAACGTGGATATTGGCATCCCCTTGGATGGAGAAACAGAGGAGCTGTGAGGTGTTTTCCTCCTCTGATTTGCATC
ATTTCCCTCTCTGGCTCCAATTTGGAGAGGGAATGCTGAGCAGATAGCCCCATTGTTAATCCAGTATCCTTATGGGAATGGAGG
GAAAAAGGAGAGATCTACCTTTCCATCCTTTACTCCAAGTCCCCACTCCACGCATCCTTCCCTACCAACTCAGAGCTCCCCTTCTA
CTTGCTCCATATGGAACCTGCTCGTTTATGGAATTTGCTCTGCCACCAGTAACAGTCAATAAACTTCAAGGAAAATGAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Solo la parte centrale (in blu e sottolineata) codifica per la relativa proteina (MYOZENIN 1). **La regione a monte si definisce 5'UTR, la regione a valle si definisce 3'UTR (UTR=UnTraslate Region)**

Notare il poliA finale

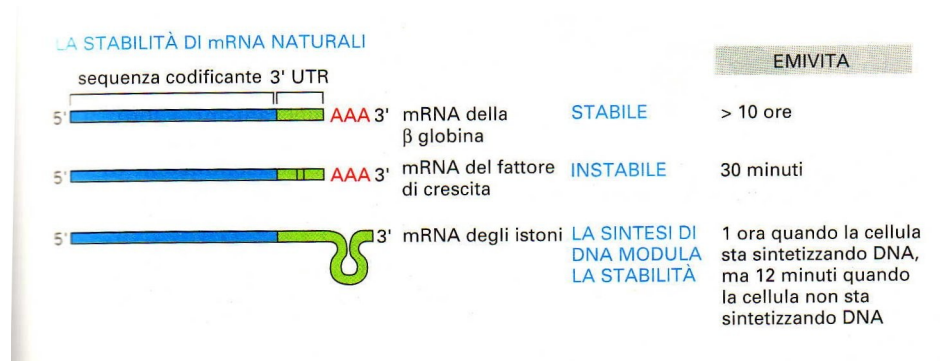
# Sono trascritte anche regioni non tradotte

L'inizio della trascrizione è solitamente 50-200 bp a monte dell'inizio della regione codificante, la fine della trascrizione è invece, mediamente 200 – 1000 bp a valle della regione codificante.

Queste regioni vengono denominate rispettivamente **5'UTR** e **3'UTR**.

(UTR=UnTraslate Region)

Anche se non codificano per le proteine, queste regioni svolgono importanti ruoli di controllo e regolazione (es. stabilità del mRNA, localizzazione cellulare, regolazione della traduzione ecc.ecc.)

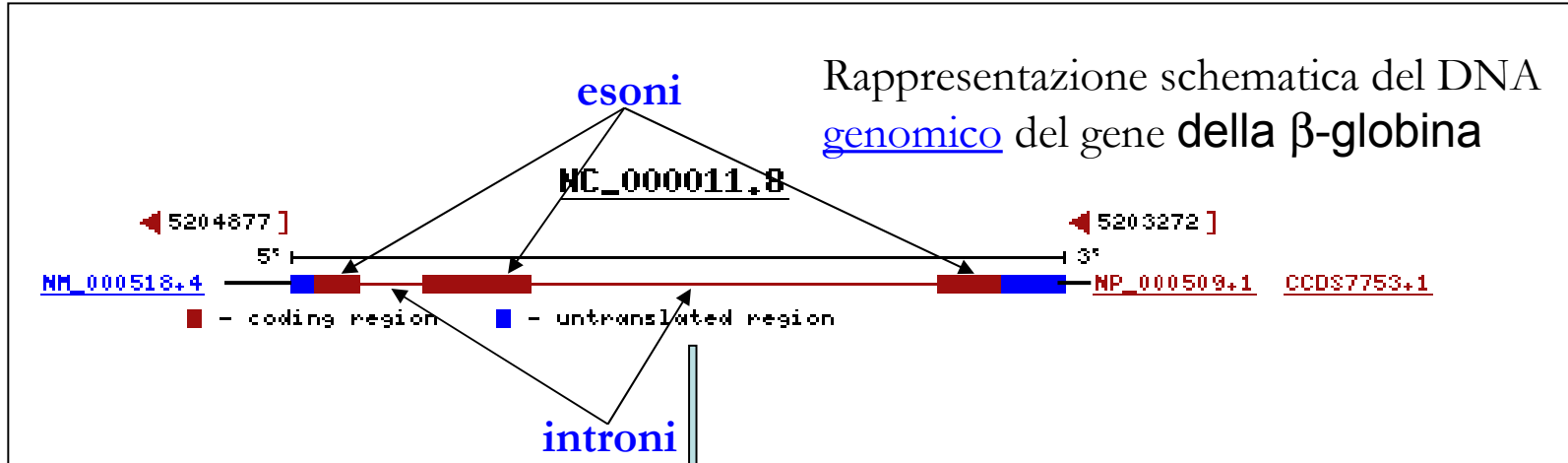


Esempio dell'influenza della regione 3'UTR sulla stabilità dell'mRNA

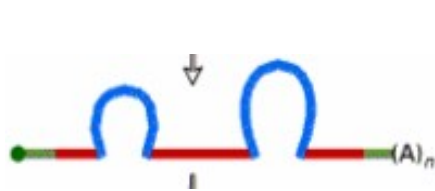
Database regioni UTR: <http://utrdb.ba.itb.cnr.it/>

# ESONI – INTRONI

Generalmente i geni degli eucarioti (in particolare degli eucarioti superiori) sono interrotti da introni.



Trascrizione, poliadenilazione, taglio al 3'  
→ trascritto primario



Messaggero maturo



## Maturazione dell'RNA

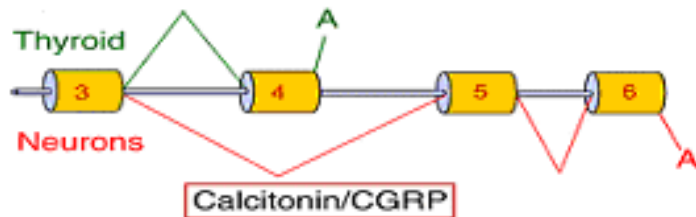
Ad opera di un complesso ribonucleoproteico: *splicing* degli **introni** (la loro eliminazione), e l'unione degli **esoni** a formare il trascritto maturo (RNA messaggero maturo).

## Splicing alternativo

Durante lo splicing, gli esoni degli eucarioti possono essere combinati in modo differente.

Si ottengono così differenti mRNA che codificano differenti proteine a partire dallo stesso gene

Questo meccanismo consente di amplificare la quantità di informazione contenuta nel genoma. (nell'uomo si stimano circa 30-40000 geni che, con questo meccanismo, possono codificare per più di 100.000 differenti proteine.



### Esempio di splicing alternativo

L'uso alternativo dello *splicing* nel gene dei vertebrati per la calcitonina e CGRP (Calcitonin Gene Related Peptide), genera un ormone che regola l'omeostasi del calcio nella ghiandola tiroide (in verde) o un neuropeptide vasodilatatore nel sistema nervoso (in rosso).

Database di splicing alternativi: <http://www.ebi.ac.uk/astd> ; <http://bioinformatics.ucla.edu/ASAP2/>

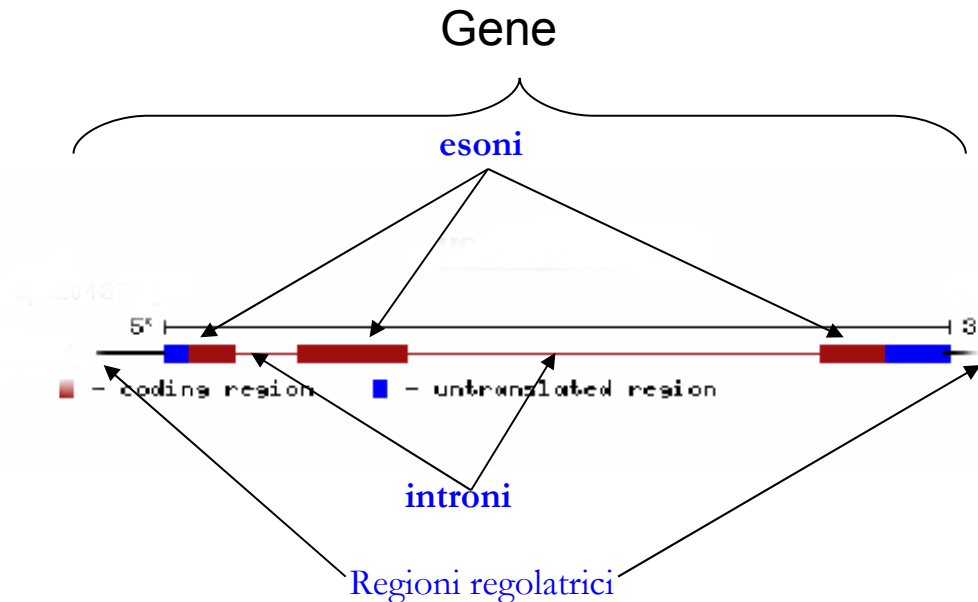
# Definizione di Gene

Una volta, il gene veniva definito come quella parte di DNA dalla quale si originava una proteina (mediante i processi trascrizione/traduzione). Si soleva dire anche: un gene → una proteina.

Oggi il concetto di gene si è evoluto ed è più complesso definirlo (esistono anche regioni di DNA che non esprimono la loro funzione attraverso proteine, ecc).

Per i nostri scopi ci limitiamo a definire gene una unità del DNA che corrisponde ad un'informazione genetica che viene trasferita in modo ereditario.

**Nel nostro corso, semplicemente, consideriamo un gene come la regione del DNA da dove si origina una proteina, comprese le adiacenti regioni regolatrici e le sequenze introniche. Sono comprese nello stesso gene tutte forme derivate dagli splicing alternativi**



## Traduzione in silico

Data una sequenza nucleotidica, è possibile tradurla utilizzando sei differenti frame di lettura, tre per ogni filamento (senso e antisenso), allo scopo di individuare una **Open Reading Frame (ORF)** cioè una sequenza nucleotidica che potenzialmente codifica per una proteina completa (di solito non meno di 70-100 aa).

Con appositi programmi, che scansionano tutto un genoma, è possibile trovare tantissime proteine PUTATIVE, cioè possibili ma non verificate.

Esempio (Traduzione con <http://www.expasy.ch/tools/dna.html>)

```
ACATGATGACCAAGAACCTGGCAGGAAAGACCCCGACGGACCTGGTGCAGCTCTGGCAGGCTGATACCCGGCACGCCCTGGAGCATCCTGA
GCCGGGGGCTGAGCATAACGGGCTGGAGGGGCCTAATGATAGTGGGCGAGAGACCCCTCAGCCTGTGCCAGCCCAGTGAATGCGTCAATAA
AAAAGCTGTTTTTGGCTAAAAAAAAAAAAAAAAAAAAA
```

```
T Stop Stop P R T W Q E R P R R T W C S S G R L I P G T P W S I L S R G L S I T G W R G L Met I V G E R
P L S L C Q P S E C V N K K A V F A K K K K K K
```

```
H D D Q E P G R K D P D G P G A A L A G Stop Y P A R P G A S Stop A G G Stop A Stop R A G G A Stop Stop
Stop W A R D P S A C A S P V N A S I K K L F L L K K K K K K
```

```
Met Met T K N L A G K T P T D L V Q L W Q A D T R H A L E H P E P G A E H N G L E G P N D S G R E T P Q
P V P A Q Stop Met R Q Stop K S C F C Stop K K K K K
```

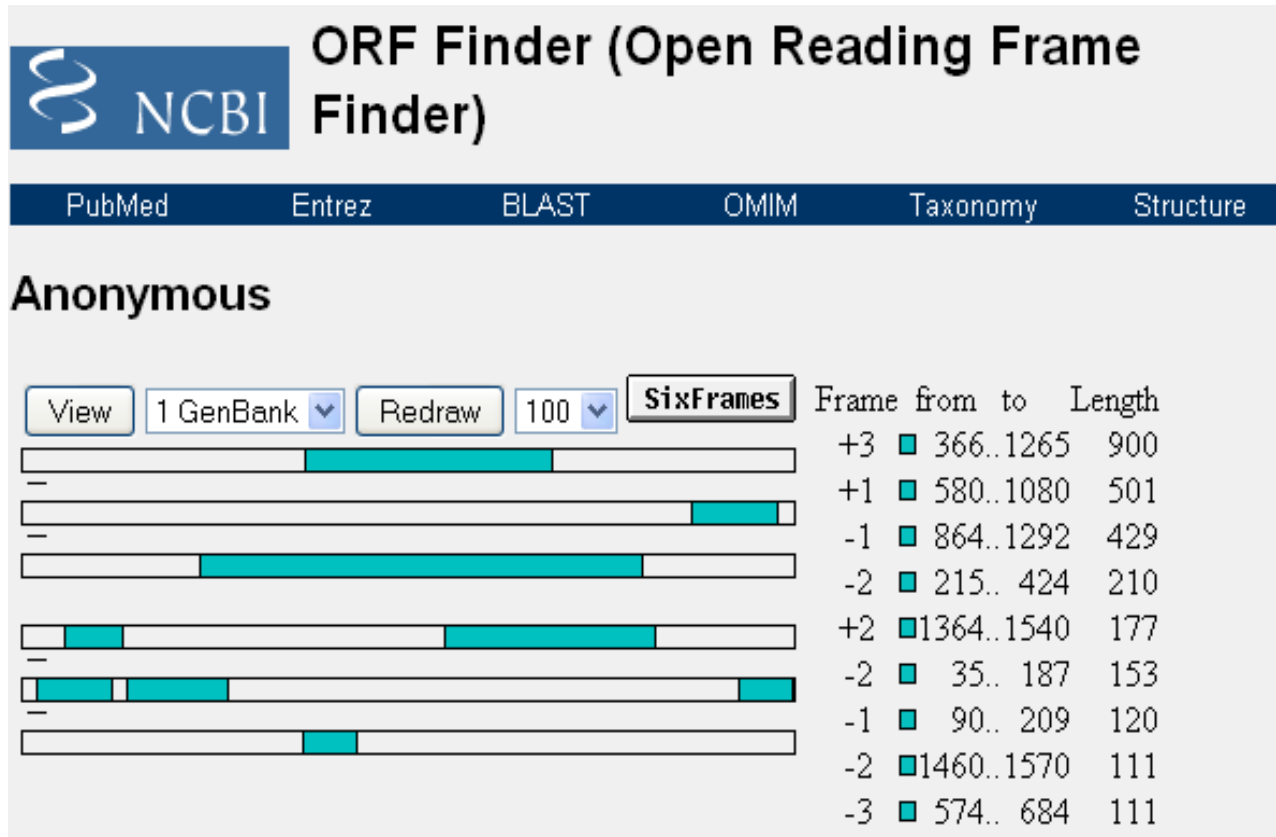
```
F F F F F F Stop Q K Q L F Y Stop R I H W A G T G Stop G V S R P L S L G P S S P L C S A P G S G C S R A
C R V S A C Q S C T R S V G V F P A R F L V I Met
```

```
F F F F F F S K N S F F I D A F T G L A Q A E G S L A H Y H Stop A P P A R Y A Q P P A Q D A P G R A G Y
Q P A R A A P G P S G S F L P G S W S S C
```

```
F F F F F L A K T A F L L T H S L G W H R L R G L S P T I I R P L Q P V Met L S P R L R Met L Q G V P G
I S L P E L H Q V R R G L S C Q V L G H H
```

Vedere esercizio completo di traduzione nel file traduzione.pdf

Esistono anche altri programmi (come ORF Finder all'NCBI) che offrono una veste grafica alla traduzione in silico



Altri Siti interessanti (anche con animazioni)

**Trascrizione e traduzione:**

[www.vcbio.sci.kun.nl/eng/virtuallessons/cellcycle/trans/](http://www.vcbio.sci.kun.nl/eng/virtuallessons/cellcycle/trans/)

**Trascrizione:**

<http://vcell.ndsu.edu/animations/transcription/movie.htm>

**Traduzione:**

<http://vcell.ndsu.edu/animations/translation/movie.htm>

## Nucleotide ambiguity code (codice IUPAC)

Code	Represents
A	Adenine
G	Guanine
C	Cytosine
T	Thymine
Y	Pyrimidine (C or T)
R	Purine (A or G)
W	weak (A or T)
S	strong (G or C)
K	keto (T or G)
M	amino (C or A)
D	A, G, T (not C)
V	A, C, G (not T)
H	A, C, T (not G)
B	C, G, T (not A)
X/N	any base
-	Gap

La sequenza di una regione di DNA (es. di un gene) può essere differente anche tra individui della stessa specie.

Utilizzando semplicemente i 4 simboli delle basi canoniche del DNA, non sarebbe possibile scrivere in un'unica stringa la sequenza di una stessa regione di DNA presente in più individui.

Sono stati introdotti dei simboli con i quali è possibile scrivere una sequenza di DNA in un'unica stringa, con tutte le possibili combinazioni delle 4 basi.

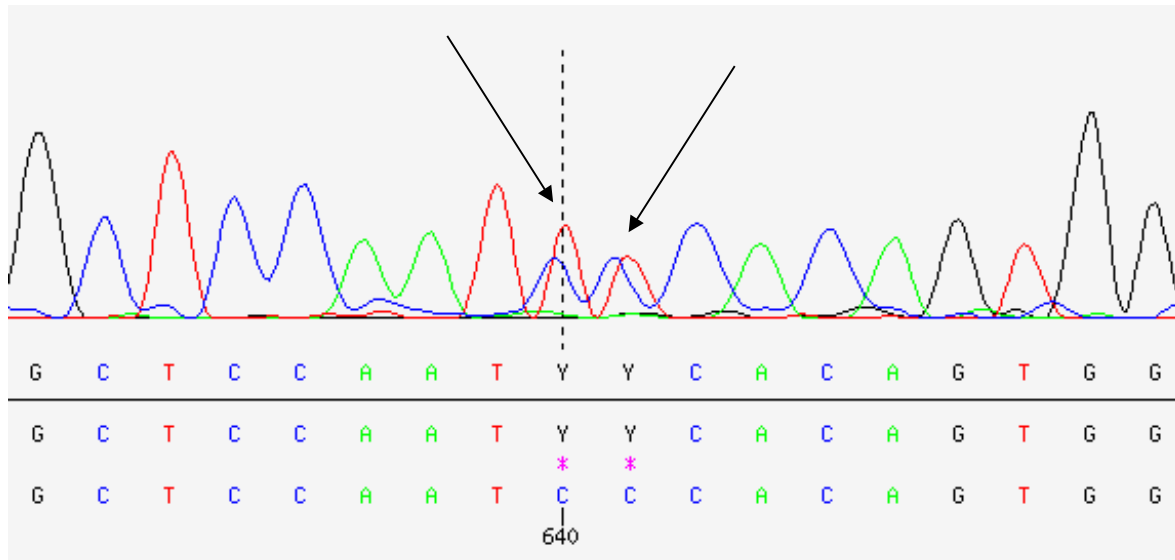
Se nel sequenziamento di una regione di DNA si riscontrano delle ambiguità o incertezze, è possibile annotarle con i simboli del codice IUPAC

### Esempio

1)	CCGTTGAC <b>C</b> GCGCCATT <b>A</b> GCGGCAACGCAGCGTGCC <b>C</b> GAGCCGCGATTAGAGC
2)	CCGTTGAC <b>G</b> GCGCCATT <b>T</b> GCGGCAACGCAGCGTGCC <b>G</b> GAGCCGCGATTAGAGC
3)	CCGTTGAC <b>C</b> GCGCCATT <b>C</b> GCGGCAACGCAGCGTGCA <b>A</b> GAGCCGCGATTAGAGC
Unione	CCGTTGAC <b>S</b> GCGCCATT <b>H</b> GCGGCAACGCAGCGTGCV <b>V</b> GAGCCGCGATTAGAGC



Porzione di un cromatogramma ottenuto sequenziando il DNA col metodo Sanger (vite merlot)



Ad ogni picco colorato corrisponde una base (4 colori, uno per ogni base).

Il DNA sequenziato era eterozigote.

Sono evidenti due doppi picchi (rosso=T, blu=C) A+T corrispondono a Y nel codice IUPA

Il DNA paterno e materno erano differenti in quelle due posizioni (SNPs: Single Nucleotide Polimorphism)