

# Espressione genica: trascrizione (parte 1)

## Gli elementi funzionali di un gene

Il gene è l'unità ereditaria fondamentale degli organismi viventi. I geni corrispondono a porzioni di codice genetico localizzate in precise posizioni all'interno della sequenza (DNA o RNA) e contengono tutte le informazioni necessarie per la produzione di una proteina. Essi sono contenuti ed organizzati all'interno dei cromosomi, presenti in tutte le cellule di un organismo. Le cellule umane contengono tutte 23 coppie di cromosomi, con la sola eccezione dei gameti, che presentano una singola copia di ciascun cromosoma. La definizione di gene presenta diverse accezioni a seconda che si considerino organismi procarioti ed eucarioti: nei primi il gene è costituito esclusivamente da sequenze codificanti, negli altri anche da sequenze non codificanti. Nel gene eucariotico la sequenza codificante si definisce esone e quella non codificante introne. Ogni gene può presentare delle forme alternative, che differiscono leggermente fra loro nella sequenza nucleotidica e che prendono il nome di alleli. Nella cellula eucariotica, un gene consiste concretamente (nella maggior parte dei casi) in una sequenza di DNA. Tale sequenza è caratterizzata dalla presenza di:

- un promotore, che controlla l'espressione genica;
- regioni codificanti, definite esoni;
- sequenze non codificanti, definite introni, che possono avere funzione regolatoria.

Sia gli esoni che gli introni sono trascritti da DNA ad RNA durante un processo chiamato trascrizione in cui si sintetizza un filamento di pre-mRNA, così definito poiché immaturo. Esso viene infatti successivamente processato originando l'RNA messaggero (o mRNA), il quale dirige la sintesi delle proteine:

- ad esso sono infatti sottratti gli introni mediante un processo definito maturazione (in inglese splicing), (in molti casi si ha uno splicing alternativo, che permette alla cellula di sintetizzare più proteine a partire da un unico gene);
- ad esso è aggiunto un cappuccio guanosinico, che ne impedisce la degradazione (in inglese capping);
- ad esso è aggiunta una coda poliadenilica, anch'essa coinvolta nella protezione del trascritto (poliadenilazione).

Alcuni RNA sono utilizzati direttamente in seguito alla trascrizione, ad esempio come parte del ribosoma. In ogni caso, siano essi semplici RNA o proteine, ogni macromolecola direttamente derivante da un gene è definita prodotto genico. Rispetto ai geni eucariotici, quelli di un

organismo procariote si differenziano soprattutto per la rarità degli introni. La maggior parte dei geni procarioti, infatti, è priva di introni e consta di un'unica sequenza ininterrotta di DNA codificante, definita cistron. I geni procariotici sono spesso raggruppati in operoni, regioni in cui diversi geni vicini tra loro sono sotto il controllo di un unico promotore.

Da ogni operone viene trascritto un unico RNA, contenente regioni codificanti differenti, ognuna delle quali preceduta da una sequenza di Shine-Dalgarno (per l'attacco del ribosoma). I geni dirigono lo sviluppo fisico e comportamentale di un essere vivente, in quanto la maggior parte di essi codifica per proteine, le macromolecole maggiormente coinvolte nei processi biochimici e metabolici della cellula. La sintesi proteica è possibile grazie all'esistenza del codice genetico, un linguaggio a tre lettere che associa i codoni (triplette di nucleotidi sull'RNA) agli amminoacidi (i mattoni fondamentali delle proteine). Molti geni non codificano per proteine, ma producono RNA non codificante, che può in ogni caso giocare un ruolo fondamentale nella biosintesi delle proteine e nell'espressione genica. Il fenotipo di un organismo è determinato dall'espressione dei suoi geni e dall'interazione dei prodotti genici con l'ambiente. Tutte le cellule dello stesso organismo sono dotate dello stesso genotipo, ossia dello stesso corredo di geni, ma la loro espressione varia nei diversi tessuti e a seconda dello stadio di sviluppo dell'organismo stesso, nonché dell'ambiente.

Ogni singolo cambiamento nella sequenza del DNA costituisce una mutazione e può causare una conseguente alterazione nella sequenza di amminoacidi di una proteina o nella regolazione della sua espressione (che, in conseguenza, potrebbe anche avere conseguenze patologiche). È stato calcolato che le alterazioni dei nostri geni sono responsabili di circa 5000 malattie ereditarie (per esempio vari tipi di anemia). Altre mutazioni, anziché evidenziarsi in maniera diretta come malattia, possono causare una predisposizione ad esse. In seguito al completamento del Progetto Genoma Umano, i cui risultati furono pubblicati nel 2003, sono stati annoverati circa 20.000-25.000 geni. La densità genica di un genoma è la misura del numero di geni per milione di paia di basi (o megabase, Mb). I genomi procariotici hanno densità geniche più alte di quelli eucariotici. La densità genica del genoma umano è di circa 12-15 geni per paia di megabasi.

## **Direzione di sintesi dell'RNA e definizione del filamento stampo**

L'acido ribonucleico (RNA) è un polimero organico, risultante dalla polimerizzazione di ribonucleotidi. Chimicamente l'RNA è molto simile al DNA, infatti anch'esso è una catena polinucleotidica contenente quattro nucleotidi diversi, ma le molecole di RNA differiscono da quelle di DNA perché:

- contengono lo zucchero ribosio (con un gruppo OH legato al carbonio 2') anziché il deossiribosio (da qui il nome)
- una delle basi, la timina (T), è sostituita dall'uracile (U). In questo caso è l'uracile a legarsi all'adenina, mentre la guanina si lega sempre alla citosina;

- sono di solito a singolo filamento, anziché a filamento doppio (il DNA possiede una struttura più complessa, chiamata a doppia elica, che consiste nel piegamento della struttura a elica singola su se stessa, secondo i canoni dei livelli di organizzazione proteica, mentre l'RNA possiede una struttura semplice a elica singola).

Le molecole di RNA vengono sintetizzate attraverso un processo, conosciuto come trascrizione del DNA, dove un filamento di DNA viene ricopiato nel corrispondente filamento di RNA. Vi sono tre tipi di RNA comuni a tutti gli organismi cellulari:

- mRNA (RNA messaggero) che contiene l'informazione per la sintesi delle proteine;
- rRNA (RNA ribosomiale), che entra nella struttura dei ribosomi;
- tRNA (RNA transfer) necessario per la traduzione nei ribosomi.

Negli eucarioti abbiamo anche:

- hnRNA (RNA eterogeneo nucleare) tipo di molecole di cui fa parte il pre-mRNA;
- snRNA (piccolo RNA nucleare) necessario per la maturazione dell'HnRna.

La sintesi dell'RNA è molto simile a quella del DNA. La RNA polimerasi non richiede però un innesco. La trascrizione può iniziare solo presso una sequenza detta promotore e termina in presenza di altre sequenze particolari. È stata avanzata l'ipotesi che l'RNA abbia assunto un ruolo chiave negli organismi primitivi prima del DNA. A favore di tale ipotesi c'è la capacità catalitica di alcune molecole di RNA (ribozimi). Sull'mRNA viene trascritta l'informazione genetica che poi verrà utilizzata per svariati usi. La sintesi dell'RNA è molto diversa tra gli eucarioti e i procarioti. Negli eucarioti la sintesi avviene nel nucleo attraverso tre diversi enzimi detti RNA polimerasi (nei procarioti ce n'è solo una). Sono parzialmente diverse tra loro, infatti hanno alcune subunità in comune e altre uniche per la loro specie. Ciò è anche dovuto al fatto che esse producono tre tipi diversi di RNA.

- Subunità centrali: rpb1, rpb2, rpb3, sono richiesti per l'attività enzimatica;
- Subunità comuni: rpb5, rpb6, rpb8, rpb10, rpb12 processo trascrizionale;
- Subunità non essenziali: rpb4, rpb9 non sono richieste per attività trascrizionale; rpb7, rpb11, sono richieste per l'attività polimerasica;

La RNA polimerasi I produce gli rRNA 5,8 S, 18 S, 28 S nel nucleolo. La RNA polimerasi II produce mRNA e piccoli RNA stabili che servono a formare gli snRNP nel nucleo. La RNA polimerasi III produce piccoli RNA stabili, tRNA e l'RNA 5 S sempre nel nucleo. La RNA polimerasi più conosciuta è la RNA polimerasi II perché produce l'mRNA, che servirà poi ai ribosomi per sintetizzare le proteine. L'rRNA 18 S più 35 proteine compresse andrà a formare la subunità ribosomiale minore. Gli rRNA 5 S, 5,8 S e 28 S più 50 proteine andranno a formare la subunità ribosomiale maggiore. Durante il corso degli anni i ricercatori si sono accorti che gran

parte dell'RNA sintetizzato dalle polimerasi veniva scartato e solo una piccola parte veniva inviata sotto forma di mRNA per la sintesi proteica. Infatti per una proteina media di circa 400 amminoacidi (quindi 1200 nucleotidi) venivano sintetizzati anche più del doppio dei nucleotidi realmente necessari. Ciò è dovuto al fatto che nel DNA esistono delle sequenze non più codificanti, che servivano alla cellula quando non era ancora specializzata. Queste sequenze vengono comunque trascritte dalla polimerasi e vengono dette introni, quelle che invece vengono copiate e codificanti, esoni. Questo fatto implica che prima della traduzione, esse andranno tagliate; in un processo denominato splicing.

- L'RNA transfer (tRNA) è costituito da un filamento di RNA di circa 80 basi azotate ripiegato a formare tre lobi, che ricordano la caratteristica forma di un trifoglio. Un lobo riconosce tramite un anticodone la tripletta sull'RNA messaggero corrispondente all'amminoacido trasportato dal transfer; un altro riconosce l'enzima che attacca l'amminoacido al transfer; un terzo è il sito di riconoscimento del ribosoma. Trasferisce ai ribosomi i vari amminoacidi, che uniti formano fra loro un legame peptidico per poter formare le proteine.
- L'RNA messaggero (mRNA) si occupa di trasportare le informazioni codificate nel DNA al citoplasma. L'mRNA si presenta sotto forma di filamento sul quale sono presenti triplette di nucleotidi (detti codoni). Il processo di formazione di un nuovo filamento è detto trascrizione. La sequenza base del filamento di mRNA è complementare a quella del filamento del DNA dal quale è stato copiato, e non identica. Quando l'mRNA ha trasmesso l'informazione si scompone nei nucleotidi che lo componevano. La maturazione dell'mRNA differisce molto tra gli eucarioti e i procarioti. L'mRNA procariotico è già maturo dopo la trascrizione e non richiede di essere controllato, tranne che in alcuni casi. Il pre-RNA eucariotico invece richiede diversi passaggi. Le tappe della maturazione sono: splicing, capping, poliadenilazione.
- L'RNA ribosomiale (rRNA) nel citoplasma, forma con le proteine particolari organuli, i ribosomi, sui quali avviene la sintesi proteica.

Il meccanismo utilizzato per la conversione dell'informazione da sequenze di nucleotidi in sequenze di amminoacidi è conosciuto come codice genetico; il processo di trasferimento di informazione dagli mRNA alle proteine costituisce la cosiddetta traduzione dell'informazione genetica. Il codice genetico consiste nell'esistenza di una corrispondenza tra unità di messaggio, costituite a sequenze di tre nucleotidi (triplette o codoni) contenute negli mRNA, e uno dei venti amminoacidi che costituiscono le proteine.

## Caratteristiche del sito promotore

In biologia un promotore è una regione di DNA costituita da specifiche sequenze dette consenso, alla quale si lega la RNA polimerasi per iniziare la trascrizione di un gene, o di più

geni (operone). I promotori, costituiti da DNA a doppia elica, si trovano a monte del gene di cui iniziano la trascrizione e sono lunghi circa 200bp; l'RNA polimerasi riconosce il DNA a doppia elica e si lega ad esso, anche se soltanto un'elica verrà utilizzata come stampo per la trascrizione. I promotori più efficienti sono chiamati promotori forti, dove s'intende il numero di trascritti a cui esso può dare inizio in un dato tempo. I promotori sono costituiti da sequenze nucleotidiche di basi intervallate da corte sequenze che funzionano come moduli di controllo per l'espressione genica.

Sono proprio questi moduli a caratterizzare i promotori: vi sono quelli costitutivi, i cui elementi di controllo possono legarsi ubiquitariamente a fattori di trascrizione differenti (geni housekeeping), e quelli che rispondono a fattori specifici che ne regolano l'espressione genica. I promotori di tutti i geni che codificano per proteine sono stati analizzati confrontando le sequenze di DNA a monte di numerosi geni strutturali e verificando l'eventuale presenza di regioni con sequenze simili. I risultati di questi tipi di esperimenti indicano che i promotori dei geni che codificano per proteine contengono elementi promotori basali ed elementi promotori prossimali. I promotori sono divisibili in costitutivi e inducibili. I promotori costitutivi sono spesso associati a regioni enhancer, che sono in grado di fare aumentare l'attività trascrizionale da 10 a 200 volte, rispetto alla trascrizione in assenza di tale sequenza. L'enhancer è costituito da una serie di moduli per l'associazione con proteine specifiche che la rendono per struttura molto simile ad un promotore. Spesso si trovano in una regione molto lontana rispetto al promotore del gene a cui sono associati. Alcuni enhancer virali sono specifici per un dato ospite, mentre altri (SV40, LTR di RSV e CMV) sono attivi in differenti tipi cellulari di varie specie. Essi sono i determinanti temporali e spaziali dell'espressione di un gene.

Esempi di promotori costitutivi sono: il promotore dei geni precoci di SV40; il promotore maggiore dei geni della regione tardiva dell'adenovirus; il promotore del Raus Sarcoma Virus (RSV); il promotore immediato dei geni precoci del citomegalovirus (CMV). I promotori inducibili nei mammiferi sono sotto il controllo di specifici enhancer che, a seguito di stimoli cioè di fattori trascrizionali specifici, vengono attivati. Questi promotori sono molto sfruttati in ingegneria genetica, quando è necessario controllare il livello di espressione dei geni esogeni. Per utilizzare i promotori inducibili è necessario uno studio approfondito del sistema da usare. Infatti, l'elemento induttore potrebbe agire, oltre che sul promotore del gene da esprimere, anche altrove, generando risposte generali spesso vantaggiose. Un esempio di promotore inducibile è quello del  $\beta$ -interferone, che è indotto nei fibroblasti da un'infezione virale e non risulta funzionale in tutti i tipi cellulari. Per adattarlo a vari tipi cellulari è necessario pretrattare con interferone le cellule, in modo da indurre la sintesi della proteina che riconosce l'enhancer specifico. Un altro esempio di promotore inducibile è quello delle metallotionine, che contiene diversi elementi MRE (metal regulatores elements), che rispondono ai metalli pesanti, come cadmio e zinco.

L'inserzione di più copie di questi elementi a monte del promotore genetico può comportare un'induzione Zn-dipendente del gene clonato. Altri promotori inducibili sono ad esempio:

quello delle Heat-Shock protein, attivato in seguito ad uno stress; quello inducibile dai glucocorticoidi, in cui è il complesso recettore-ormone a traslocare nel nucleo e ad agire direttamente da fattore trascrizionale; quello dipendente dall'espressione dell'antibiotico tetraciclina, che rappresenta un sistema di controllo dell'espressione genica batterica, trasportato anche negli eucarioti. Nel nucleo degli eucarioti sono presenti tre RNA polimerasi: ognuno di questi enzimi riconosce un promotore associato ad una particolare classe di geni.

- RNA polimerasi I: sintetizza tutti gli rRNA escluso il 5S che viene codificato in altre regioni del genoma
- RNA polimerasi II: sintetizza tutti gli mRNA che grazie all'apparato ribosomiale verranno tradotti in proteine nel citoplasma
- RNA polimerasi III: sintetizza l'rRNA 5S, gli snRNA, ossia piccoli RNA che associati a proteine formeranno gli snRNP (small nuclear ribonucleoprotein) coinvolte nel processo di splicing ed infine i tRNA; questi ultimi trasportano l'aminoacido specifico che verrà inserito nella catena polipeptidica mediante il riconoscimento codone-anticodone.

Il motivo di questa specificità risiede nel fatto che ogni tipo di RNA polimerasi riconosce soltanto quei promotori che si rintracciano all'interno di una classe particolare di geni. Il promotore ha un ruolo fondamentale nella trascrizione ma non nella duplicazione del DNA; ciò dipende dal fatto che la trascrizione avviene con coinvolgimento di un promotore specifico riconosciuto dalla RNA polimerasi II. La replicazione, invece, non richiede l'intervento di un promotore, poiché l'enzima DNA polimerasi per poter svolgere la sua funzione necessita della presenza di un primer il cui 3'OH funge da sito di innesco. Negli Eucarioti generalmente un unico promotore controlla l'espressione di un solo gene; si parla infatti di mRNA monocistronici. Al contrario, nei Procarioti uno stesso promotore controlla l'espressione di più geni che codificano per proteine funzionalmente correlate, come ad esempio l'operone lac in cui uno stesso promotore permette la trascrizione dei geni che codificano per enzimi ( $\beta$ -galattosidasi, transacetilasi e lattosio permeasi) che permettono ai batteri di metabolizzare il lattosio. Quindi, nei Procarioti si parla di mRNA policistronici. Inoltre, nei Procarioti il riconoscimento del promotore è mediato dall'oloenzima, un complesso costituito dall'RNA polimerasi (core) e i fattori  $\sigma$  che permettono l'individuazione di promotori specifici. Ad esempio, in E.coli, in condizioni fisiologiche, è maggiormente espresso il fattore  $\sigma 70$  che consente la trascrizione dei geni housekeeping (costitutivi). In condizioni di stress, invece, nella cellula vengono espressi fattori  $\sigma$  alternativi che permettono alla cellula di sopravvivere alle condizioni avverse. Durante lo shock termico (heat shock), ad esempio, aumentano i livelli d'espressione del fattore  $\sigma 32$  che si associa all'RNA polimerasi andando così a legare promotori specifici per l'espressione di proteine necessarie alla sopravvivenza. La sequenza consenso rappresenta la successione migliore di basi, riconosciuta dall'RNA polimerasi per legarsi allo stampo e dare avvio alla trascrizione. Nei procarioti il promotore si compone di tre parti:

- un sito di inizio (quasi sempre una purina)



- DPE (Downstream Promoter Element): situato a +30; presente nei promotori privi della TATA box e funziona solo se è presente anche INR

Elementi regolatori prossimali per determinare quando e quanto frequentemente un gene è espresso. Sono posti tra -50 e -200 rispetto al sito d'inizio della trascrizione:

- GC box: tipica dei geni housekeeping e opera come intensificatore
- CCAAT box: posta a -75 e opera come intensificatore
- CRE elemento di risposta a cAMP

Il promotore della RNA polimerasi III può essere di tre tipi differenti:

- Tipo 1: si trova nei geni per l'RNA ribosomiale 5S e presenta due elementi di controllo a valle del punto d'inizio della trascrizione detti boxA e boxC
- Tipo 2: presente nei geni per il tRNA e contiene due elementi di controllo detti boxA e boxB a valle del punto d'inizio della trascrizione
- Tipo 3: presente nei geni per gli snRNA ed è costituito da tre elementi conservati detti Oct, PSE e TATA a monte del punto d'inizio della trascrizione

Alcuni geni hanno dei promotori alternativi, ovvero siti di legame della RNA polimerasi su uno stesso gene opzionale rispetto a quello principale. Essi danno origine a diverse versioni del trascritto a partire dallo stesso gene e permettono di ottenere una molecola di RNA che differisce all'estremità 5' da quella relativa al promotore principale. I promotori alternativi sono presenti nel 52% geni umani (in media dal 30-50%) e svolgono diverse funzioni:

- possono guidare la trascrizione a partire da un primo esone che può avere versioni alternative;
- se interni, possono portare alla formazione di prodotti tronchi;
- determinano la specificità di tessuto;
- permettono ad una singola cellula di sintetizzare proteine simili con caratteristiche biochimiche leggermente differenti.

Difetti nel funzionamento dei promotori alternativi, possono essere causa di gravi malattie genetiche. Un esempio significativo è la distrofia muscolare di Duchenne, dovuta alla mutazione del gene distrofina, che governa la produzione della omonima proteina, componente fondamentale della struttura delle fibre muscolari scheletriche e cardiache.



