

Renato Dulbecco
La mappa della vita
(2005)

L'interpretazione del codice genetico:
una rivoluzione scientifica
al servizio dell'umanità

NUOVA EDIZIONE AMPLIATA

Il Progetto Genoma

Introduzione

Il Progetto Genoma è stato una grande avventura. È cominciato come il sogno di pochi visionari, è poi stato abbracciato dall'intera comunità scientifica, e ha raggiunto i suoi obiettivi con la cooperazione di istituzioni pubbliche e private. Questo è il vero tragitto di una grande conquista scientifica nel tempo attuale. Il segreto del suo successo comprende molti fattori. Il principale è stata la dedizione assoluta di molti scienziati, che avevano fede di poter raggiungere lo scopo malgrado la scarsità di mezzi tecnici a disposizione. Rapidamente questi mezzi sono stati sviluppati, come tecnologie nuove e tutte automatizzate, per determinare l'organizzazione del DNA, rintracciarvi i geni, leggere i messaggi che essi contengono e i loro significati. Sono stati usati nuovi indirizzi per determinare l'attività dei geni, esplorando in un atto solo tutto il genoma. Straordinario in questo progresso è stato il contributo dell'informatica.

Il primo risultato che ora abbiamo in mano è un abbozzo, un po' approssimativo, di ciò che è scritto nel genoma, cioè la sua sequenza. Questo abbozzo ha bisogno di ulteriori perfezionamenti, che saranno completati in tempo abbastanza breve; ma, anche senza questo passaggio, le conoscenze acquisite costituiscono una vera rivoluzione.

In questo Progetto si sono studiati i geni non soltanto della specie umana, ma anche di molte altre specie: virus, batteri, lieviti, animali dai più semplici ai più complicati e piante. Il risultato stupefacente è che tutte queste specie sono connesse tra di loro perché i loro geni sono molto simili. Perciò è chiaro che tutti gli esseri viventi sono parte di uno stesso mondo, con caratteristiche diverse determinate dall'evoluzione.

Un altro elemento straordinario è che i risultati del Progetto hanno rovesciato il modo di pensare ai geni. Fino a un anno fa li si riteneva elementi indipendenti del genoma, che si dovevano studiare uno per uno per capire cosa facevano e per determinare il loro ruolo nel funzionamento degli organismi viventi. Poi, improvvisamente, studiando i risultati del Progetto, si è visto che ciò non era vero: i geni lavorano insieme in grandi

complessi, ciascuno destinato a una funzione specifica. La visione del gene isolato persiste in alcuni casi, ma è ora inserita nella visione globale dei complessi di geni. Questo cambia moltissimo la nostra visione del ruolo dei geni nel funzionamento normale dell'organismo e nelle malattie.

Come risultato, l'individuo può ora venir considerato connesso ai suoi geni in modo estremamente preciso, completando il percorso di quel processo di caratterizzazione che è cominciato con le impronte digitali e poi si è esteso alla struttura del DNA. Ora culmina nella conoscenza dettagliata dei geni di ciascun essere umano, che va sotto il nome di Profilo Genetico Individuale. Ma tale Profilo non è solo una descrizione fisica dell'individuo: è una descrizione della sua esistenza. Esso determina le condizioni di vita più adatte per lui, i suoi rischi di malattia, quali farmaci possa usare e quali no. Perciò come risultato la sua vita sarà più tranquilla, con meno preoccupazioni per la sua salute; e tutto questo potrà anche contribuire a prolungarla.

Il concetto di malattie e di come combatterle cambia notevolmente grazie al riconoscimento della cooperazione globale dei geni. Gli sforzi diagnostici saranno diretti alle funzioni globali del genoma, per capire la natura delle malattie in cui prendono parte molti geni. Tutti gli sforzi del passato nella cura di patologie dovute a geni alterati sono stati spesso deludenti, ma ora saranno rivalutati sulla base di questo concetto. Procedure come la terapia genica saranno sviluppate, ma limitatamente a situazioni speciali, in cui l'azione di un gene è chiaramente predominante; per gli altri casi si dovranno sviluppare terapie globali, dirette a tutta una cellula o a tutto un organismo. Anche la visione del cancro evolve da quella di una malattia di pochi geni a quella di un disturbo dell'intero genoma, dirigendo la ricerca di nuove terapie in quella direzione. Le nuove terapie saranno basate sulla conoscenza dei geni e del ruolo che essi hanno nel progresso della malattia, disegnando nuovi farmaci sulla base delle conoscenze delle proteine, della loro struttura e delle loro funzioni.

La conoscenza dei nostri geni è la conoscenza di noi stessi, delle nostre origini e del ruolo dell'ambiente nel dirigere la nostra vita e la nostra personalità. Molto rimane da fare per raggiungere la conoscenza completa e dettagliata del genoma; e ancora di più per capire come esso funziona e come determina il destino dell'individuo. Ma questo processo è ormai stato avviato, e andrà avanti senza esitazioni. Il Progetto è come un razzo che è stato lanciato dopo lunghe preparazioni, ma ora è in orbita e alla fine raggiungerà il pianeta a cui è stato diretto, il pianeta uomo.

Che cos'è un gene

Il concetto di gene

È interessante osservare come si arrivò al concetto di gene. Esso fu il culmine delle conoscenze e delle speculazioni sulla struttura ereditaria degli organismi - inizialmente piante - che aumentarono con il tempo, prendendo spunto da semplici osservazioni e poi sviluppando teorie che progressivamente si avvicinarono alla realtà.

L'esistenza di caratteristiche ereditarie, cioè trasmesse da una generazione all'altra, è stata riconosciuta da lungo tempo specialmente dagli allevatori di piante e bestiame, che si sono sempre sforzati di migliorare la qualità dei loro prodotti selezionando i tipi più adatti. Nonostante ciò, lo sviluppo di modelli per spiegare come le caratteristiche vengono trasmesse cominciò solo verso la metà del XIX secolo. Charles Darwin sviluppò l'ipotesi della «pangenesi», secondo cui le cellule di una pianta o di un animale immettono nel sangue piccole particelle, i «pangeni», che poi si uniscono per formare le cellule germinali, quali gli ovociti e gli spermatozoi. In questo modo si cercava di spiegare come le caratteristiche riconoscibili in un individuo possano essere trasmesse attraverso le cellule germinali. Secondo questa teoria, se le cellule vengono alterate da danni ricevuti, anche i pangeni ne risultano alterati, per cui i caratteri acquisiti durante la vita di un individuo sarebbero trasmessi alla progenie.

La teoria continuò a esistere per parecchio tempo, con varie modificazioni. Con il riconoscimento dell'esistenza, nei nuclei delle cellule, di bastoncelli noti come «cromosomi» (dal greco antico *chromo-*, «colore», e *soma*, «corpo», perché si tingono fortemente con certi coloranti) si pensò che i pangeni fossero localizzati entro di essi, e che diventassero attivi in cellule diverse, spiegando così il fenomeno dei cambiamenti cellulari durante lo sviluppo dell'organismo, cioè la «differenziazione». In questa nuova versione della teoria, il materiale ereditario è costituito da fattori presenti in tutte le cellule, che ne

determinano le caratteristiche; le cellule germinali hanno tutti i fattori, che poi vengono distribuiti alle varie cellule dell'embrione. Secondo queste idee, le cellule germinali (gli ovociti, gli spermatozoi) contengono il «germiplasma» immortale, che viene continuamente trasmesso da una generazione all'altra; invece le cellule del corpo «somatiche» contengono il «somatoplasma», che è mortale. Questo concetto portò a rigettare l'ipotesi dell'ereditarietà dei caratteri acquisiti, in quanto risultato di cambiamenti del somatoplasma, che non è trasmesso ereditariamente.

Il primo lavoro sperimentale sull'ereditarietà di caratteristiche corporee fu fatto da Johann Gregor Mendel, e si sviluppò su questo sfondo di idee, studiando i risultati di incroci tra piante di piselli con caratteristiche diverse. La conclusione fu che ogni individuo ha un paio di «fattori» determinanti una certa caratteristica, e ne trasmette uno alla sua progenie; i fattori provenienti dai due genitori si uniscono a caso nella progenie. Le cellule germinali sono pure, hanno cioè solo un fattore, e i vari fattori non si mescolano né si contaminano l'un l'altro. Perciò l'organismo è un mosaico di tali fattori, poi indicati come «geni», un termine coniato dal botanico danese Vilhelm L. Johanssen nel 1909. Questo termine fu subito accettato dalla comunità scientifica, che aborrisce le parole comuni e ama quelle coniate sulla base del greco antico.

Questi risultati sono poi stati convalidati da molte altre osservazioni, concludendo che negli animali e nelle piante ogni cellula del corpo ha due copie di ogni gene in due cromosomi simili che sono ereditati uno dal padre e l'altro dalla madre. Le cellule germinali, sia maschili sia femminili, ricevono una copia ciascuna di ogni gene; quando si forma un nuovo organismo dalla fusione di due cellule germinali, esso riceve una copia di ciascun gene da ciascuna cellula germinale, ricostituendo la composizione delle cellule del corpo.

Un punto centrale per lo sviluppo successivo delle conoscenze genetiche è stato il lavoro di Thomas Hunt Morgan, al principio del XX secolo. All'inizio egli ripudiò le idee di Johann Mendel perché le considerava frutto di speculazioni suggerite da osservazioni che avrebbero potuto avere altre spiegazioni, e non erano assoggettabili a una verifica sperimentale. Il fatto che i cromosomi si comportavano secondo le predizioni non era ritenuto sufficiente perché era un'osservazione, non una prova sperimentale.

Sotto questo aspetto l'attitudine di Morgan rifletteva la nuova filosofia scientifica del suo tempo, che è seguita tuttora. Secondo questa filosofia

uno scienziato, in qualunque campo esso sia attivo, segue nel suo lavoro un percorso ben definito, fatto di stadi successivi. Per ognuno di questi stadi ha un obiettivo, che può essere molto vago oppure preciso, per esempio quello di scoprire il ruolo di un gene. Egli conosce bene ciò che è già noto in quel campo, e basandosi su queste conoscenze sviluppa una nuova idea che suggerisce un'ipotesi precisa, per esempio che quel gene produce una proteina che controlla una certa caratteristica delle cellule. Lo scienziato non accetta che l'ipotesi sia vera senza prendere in considerazione altre eventuali spiegazioni; per cui va a lavorare cercando di ottenere, con esperimenti appropriati, ulteriori dati che possano confermare l'ipotesi. Se egli raggiunge un punto nel suo lavoro in cui tutti i dati sono d'accordo con l'ipotesi, questa rimane valida finché qualche nuovo dato sperimentale non la contraddica. Se questo avviene, sarà necessaria una nuova modificazione dell'ipotesi. In questo modo le conoscenze scientifiche progrediscono a gradi, sulla base di ipotesi che diventano sempre più precise. Ma la verità assoluta non si raggiunge mai. Infatti, per quanto la conoscenza di un fenomeno sia approfondita, c'è sempre la possibilità di qualche aspetto inesplorato che non sia in completo accordo con l'ipotesi su cui si basa. Finché non si scopre una discordanza, la conoscenza è accettata, ma se una discordanza viene scoperta, si deve ricorrere a una nuova modificazione.

Seguendo questo indirizzo filosofico, nel 1908 Morgan intraprese degli studi sperimentali con il moscerino della frutta, la *Drosophila*, che fu selezionato perché si alleva facilmente, ha un tempo di generazione corto (due settimane), per cui si può seguire per molte generazioni, e ha caratteristiche corporee facilmente riconoscibili, come il colore degli occhi o la lunghezza delle sue setole. Studiando l'effetto di cambiamenti dei cromosomi nell'espressione delle caratteristiche genetiche, Morgan confermò le idee di Mendel, aggiungendovi un gran numero di altre caratteristiche, principalmente quella di cambiamenti improvvisi, riconoscibili da modificazioni nel corpo dell'insetto, che vennero chiamati «mutazioni».

Dove sono i geni?

Per capire che cos'è un gene e come funziona, dobbiamo prima vedere come sono fatti i luoghi dove il gene esiste: le «cellule». Queste sono

minuscoli corpicciuoli tutti attaccati l'uno all'altro che costituiscono il corpo di un animale o di altri organismi. Il nome «cellula» fu usato nel 1655 dal fisico inglese Robert Hooke che, esaminando un pezzo di sughero con un microscopio primitivo, osservò che era costituito da piccoli compartimenti, che egli paragonò alle celle di un monastero. Ciò che lui vedeva non erano in realtà le vere cellule, ma le pareti di legno che le circondano nella pianta.

Possiamo perciò cominciare a porre una semplice domanda: i geni si trovano soltanto nelle cellule che mostrano gli effetti delle mutazioni, o in tutte le cellule? Per poter rispondere dobbiamo prima considerare brevemente come è fatta una cellula. È come una goccia di liquido circondata da una sottile membrana, la «membrana cellulare». La cellula può essere paragonata a una fabbrica, un edificio rotondo, coperto da un tetto e circondato da una parete solida. Nel centro dell'edificio c'è una stanza interna, il «nucleo», anch'essa completamente circondata da una membrana, la «membrana nucleare». Il nucleo contiene dei lunghi fili, in forma di coppie, il «DNA» o acido desossiribonucleico; nelle coppie un filo deriva dal padre, l'altro dalla madre. Lo spazio attorno al nucleo, il «citoplasma» è intersecato da una rete di tubi ed è pieno di scatole di forme e dimensioni differenti, che contengono macchinari di vari tipi.

Un organismo composto di molte cellule, come una pianta o un animale, è costituito come una città formata da un gran numero di edifici di questo tipo, separati da spazi ristretti, attraversati da una fitta rete di fili di plastica che li tengono insieme. I vari quartieri della città, che corrispondono ai vari organi del corpo - quali il cervello, il fegato eccetera -, contengono edifici di struttura simile, ma molto diversi per forma e dimensioni nei vari organi. La città è attraversata da canali che portano acqua e sostanze nutritive, e da tubi e cavi per le comunicazioni, che sono l'Internet del corpo.

Il corpo di una pianta o di un animale si sviluppa da una cellula, l'uovo fecondato, fino a formarne molti miliardi, perché le cellule possono dividersi, generandone ciascuna due. La divisione è preceduta da una lunga preparazione, durante la quale la cellula aumenta di volume fino a raddoppiarlo; in questo tempo si raddoppia tutto il macchinario, inclusi i fili di DNA, cosicché, dopo la divisione, le due «cellule figlie» che ne derivano sono uguali alla madre prima che questa crescesse di volume. In questo periodo preparatorio avvengono alcuni cambiamenti straordinari nella stanza centrale, il nucleo. I due fili fratelli (il DNA) cominciano ad

attorcigliarsi in modi complicati, formando delle anse sempre più strette, sino a dar luogo a dei tubi rigidi, che sono dei cromosomi doppi: ogni tubo contiene due fili fratelli completi, ciascuno derivato da uno dei fili esistenti nella cellula madre prima della divisione. Prima che la cellula si divida la membrana attorno al nucleo scompare, i tubi si separano longitudinalmente in due metà, schiudendo i due fili fratelli che vengono spinti ai due estremi della cellula, cosicché ogni cellula figlia ne riceve uno. Appena la divisione è avvenuta, una parete (la membrana nucleare) si forma attorno ai tubi, che immediatamente si dissolvono, rilasciando i fili. Così le cellule figlie sono uguali alla loro madre sotto ogni aspetto.

Localizzare i geni

In una cellula osservata al microscopio prima che sia pronta a dividersi il nucleo sembra vuoto perché i fili sono troppo sottili per essere visibili; essi però lo diventano dopo che si sono condensati nei cromosomi, poco prima della divisione della cellula. È quindi difficile studiare i cromosomi di una cellula perché sono visibili soltanto durante questo brevissimo periodo; però si sviluppò un metodo efficace, usando una sostanza che blocca il progredire della divisione cellulare oltre lo stadio in cui i cromosomi sono visibili. In questo modo si osservò che le cellule di ciascuna specie hanno un numero specifico di cromosomi: nell'uomo sono 46, nel moscerino della frutta 8. In ogni specie ciascun cromosoma ha dimensioni e forme peculiari, e può perciò essere riconosciuto e classificato. Studiandone le caratteristiche, diventò evidente che i cromosomi sono presenti in paia: uno derivato dal padre, l'altro dalla madre. Nell'uomo uno delle 23 paia di cromosomi è formato da cromosomi di forma diversa, noti come X e Y. Essi sono i «cromosomi sessuali», perché differiscono nei due sessi: nei mammiferi, e anche nel moscerino, le femmine hanno due cromosomi X, mentre i maschi ne hanno uno X e uno Y. Come vedremo, questa proprietà determina come le caratteristiche genetiche vengono trasmesse dai genitori ai figli.

Fu possibile localizzare i geni nei cromosomi del moscerino per una circostanza fortunata: in alcune cellule di questi e altri insetti i cromosomi sono molto grandi e chiaramente visibili continuamente al microscopio. Tali «cromosomi giganti» si formano perché in queste cellule i fili del DNA raddoppiano molte volte in assenza di divisione cellulare, formando

dei fasci di un migliaio di copie, tutte perfettamente allineate: è così possibile studiare dettagli della loro costituzione, che non sarebbero altrimenti visibili. Si osservò che quando avviene una mutazione, essa è spesso accompagnata da un cambiamento osservabile in uno di questi cromosomi giganti, il che appoggiò l'idea che i geni sono nei cromosomi.

Nello studio delle mutazioni si fece un'osservazione che aiutò notevolmente a capire cosa fossero i geni. Si osservò che quando si accoppiano due individui con due mutazioni differenti, per esempio un moscerino con occhi rossi (normale) e setole lunghe (anormale) e uno con occhi bianchi (anormale) e setole corte (normale), in una parte della progenie le caratteristiche sono invertite: alcune hanno occhi rossi e setole corte, altri hanno occhi bianchi e setole lunghe. Questo risultato indicò un riarrangiamento dei geni nella progenie, che venne chiamato «ricombinazione». Poi si notò che la ricombinazione è frequente per alcune coppie di geni presenti sullo stesso cromosoma, rara per altre coppie, anche dello stesso cromosoma. Queste osservazioni suggerirono che i geni possono essere a varie distanze sul cromosoma, e che uno scambio tra due geni è tanto più probabile quanto più i geni sono distanti tra di loro.

I risultati vennero confrontati con osservazioni dei cromosomi durante la divisione cellulare, che qualche volta mostrano due cromosomi dello stesso paio in forma di croce, come se si stessero scambiando delle parti. L'osservazione suggerì che la ricombinazione è dovuta allo scambio di parti tra due cromosomi dello stesso paio. Allora emerse il concetto di *linkage*, cioè la connessione tra geni presenti sullo stesso cromosoma, ma a varie distanze. La probabilità di ricombinazione tra due geni venne da allora in poi messa in relazione con la distanza tra di essi; oggi sappiamo che essa non è proprio uguale alla distanza fisica tra i due geni, perché la ricombinazione non avviene con la stessa facilità su cromosomi diversi o nei due sessi; però essa fornì un metodo molto utile, sebbene approssimativo, per determinare la localizzazione dei geni sui cromosomi.

Come è fatto il gene?

Tutte queste osservazioni, pur non chiarificando che cosa sia un gene, portarono a ritenerlo un punto su un cromosoma. Il quadro rimase immutato per parecchio tempo, ma le cose cambiarono quando gli studi di

genetica si estesero a organismi più semplici, quali virus e batteri, che dimostravano caratteristiche genetiche simili a quelle di organismi più complessi. Un risultato importante fu ottenuto con un virus che cresce nei batteri, detto «batteriofago»: osservandolo, si scoprì che due mutazioni di origine indipendente presenti nello stesso gene si possono ricombinare, sebbene raramente. Questo dimostrò che il gene non è un punto sul cromosoma, ma ha una lunghezza: è dunque un segmento del cromosoma.

Particolarmente importanti furono due altri esperimenti che mostrarono la trasmissione di caratteristiche genetiche con il DNA. Nel primo, batteri capaci di infettare e uccidere topi furono lisati (cioè distrutti e frammentati) e disciolti, e il prodotto fu separato in diverse frazioni contenenti vari componenti; ogni frazione fu mescolata con batteri della stessa specie, ma incapaci di infettare e uccidere topi. Si osservò allora che alcuni batteri innocui diventavano capaci di uccidere dopo essere stati mescolati con la frazione che conteneva il DNA estratto dal batterio uccisore. Questo suggerì che il DNA conteneva il gene capace di uccidere i topi.

Questa conclusione non fu accettata con facilità dalla comunità scientifica, per varie ragioni. A quell'epoca si pensava che il DNA fosse solo una sostanza chimica con la composizione di uno zucchero complesso; era impossibile visualizzare come esso potesse mantenere la continuità genetica attraverso varie generazioni. In realtà, è sempre difficile per la gente, scienziati compresi, accettare scoperte rivoluzionarie. Molti suggerirono che l'effetto fosse dovuto a contaminazioni presenti nella frazione che conteneva il DNA, probabilmente di proteine (una componente importante dei corpi di tutti gli esseri viventi), che si sapevano stabili e complesse, e perciò più adatte a mantenere l'informazione genetica. E siccome il DNA non era assolutamente puro, questa possibilità non poteva essere esclusa.

Molti altri esperimenti, condotti usando DNA più purificato, diedero però gli stessi risultati; eppure il dubbio persisteva. Il problema fu finalmente risolto da un secondo esperimento condotto con un batteriofago. Questo virus è formato da due parti: una contiene DNA, l'altra delle proteine. Ciascuna parte fu evidenziata con un marcatore radioattivo diverso, in modo che le due sostanze potessero essere riconosciute separatamente nella progenie del virus. Il risultato fondamentale fu che i batteri infettati dal batteriofago marcato producevano progenie in cui solo il DNA era marcato, non la proteina. Ciò

implicava chiaramente il DNA come il portatore dell'informazione genetica, perché veniva ereditato esattamente come i geni.

Assieme, i due esperimenti suggerivano fortemente che i geni sono fatti di DNA, ma non lo provavano. Questa conclusione fu rinforzata dall'osservazione che i cromosomi di organismi più complessi contengono un'elevata percentuale di DNA. Però era ancora impossibile capire come il DNA, apparentemente una sostanza di scarso interesse, potesse esercitare tale funzione.

Tutti questi studi convinsero un giovane studente all'Università dell'Indiana, James Watson, che per capire i geni, come sono fatti, come funzionano, bisognava elucidare la struttura del DNA. Dopo aver conseguito il dottorato andò in Inghilterra, a Cambridge, dove si unì a Francis Crick, un fisico specializzato nell'identificazione di strutture molecolari attraverso la diffrazione di raggi X, e ad altri ricercatori. Alla fine il gruppo determinò che il DNA è una molecola filiforme costituita da due filamenti attorcigliati l'uno attorno all'altro, che formano una doppia elica. Ciascun filamento contiene quattro «basi», A, T, G, e C (le iniziali dei loro nomi chimici, cioè adenina, timina, guanina e citosina). Un'informazione importante fu che le basi A e T sono sempre in uguali proporzioni, come anche G e C; questa regolarità fu interpretata come indicazione che nella struttura tali basi sono presenti in paia: A-T e G-C. Per questa relazione, tanto le due basi in un paio quanto i due filamenti che costituiscono la doppia elica sono detti «complementari». Fu anche spiegato il meccanismo per cui il DNA si raddoppia nelle cellule in preparazione della divisione cellulare: ciascun filamento produce un nuovo filamento complementare a se stesso, con cui rimane associato. Il complesso dei due filamenti è identico alla molecola di origine.

Il ruolo dei geni venne poi completamente chiarificato solo alcuni anni più tardi, dopo lo sviluppo della «ingegneria genetica» che permise di isolare geni in forma pura, di farli operare in batteri o in cellule coltivate in vitro per ottenere i loro prodotti puri e in buona quantità, e di cambiare le basi per ottenere mutazioni.

L'informazione del DNA e la sua utilizzazione.

RNA e proteine

Oggi sappiamo che il DNA contiene informazioni iscritte nell'ordine

delle sue basi (cioè la «sequenza»), esattamente come il significato di un testo scritto è contenuto nell'ordine delle lettere che lo compongono; cambiando una lettera si altera il significato. L'informazione deve essere mantenuta durante le divisioni cellulari, ed è per questo che i due filamenti del DNA si raddoppiano prima della divisione della cellula, in modo che ciascuna delle cellule figlie abbia una copia di ciascun filamento. Il pericolo di errori, cioè il piazzamento di una base sbagliata, durante le duplicazioni del DNA dovrebbe essere grande: invece gli errori sono rari perché ci sono dei meccanismi che li riconoscono e li correggono. Grazie a questi meccanismi, la probabilità di errore è ridotta a una su 100 milioni di basi aggiunte. Ma qualche errore avviene lo stesso, e dà luogo a una mutazione che viene poi trasmessa alla progenie. In aggiunta ci sono altre cause di errore, per esempio l'azione di agenti chimici o radiazioni che modificano una base in modo che venga riconosciuta durante la duplicazione come una base differente; anche questi errori sono rari, perché ci sono dei meccanismi per ripararli. Però anche questi possono fallire, e le conseguenze di un errore non corretto possono essere gravi: una malattia ereditaria o un cancro.

L'informazione contenuta nei geni è stata oggetto di molti studi, grazie ai quali ora si sa che essa ha un significato preciso: portare alla formazione di un prodotto che, per la maggior parte dei geni, è una proteina. Questa molecola è anch'essa un filamento, ma costituita di elementi di natura diversa, gli «amminoacidi». Un gene può determinare la formazione di più di una proteina, come vedremo presto; perciò, essendoci nel genoma umano circa 30.000 geni, essi tutti insieme possono portare alla formazione di oltre 100.000 proteine. Questo può spiegare il notevole numero di differenti attività che il corpo umano può esercitare, sia fisiche sia mentali.

Per formare la proteina l'informazione contenuta nel gene viene utilizzata in due stadi. Nel primo si crea una copia del gene fatta di RNA (o acido ribonucleico), una sostanza molto simile al DNA. DNA e RNA sono abbreviazioni che descrivono due tipi di molecole molto simili sotto certi aspetti, ma molto diversi sotto altri. In entrambe le molecole, NA significa «acido nucleico», una sostanza chimica abbastanza complessa che include un tipo di zucchero. D e R definiscono due differenti stati dello zucchero. R = riboso, D = deossiriboso, che è un riboso modificato. Perciò DNA e RNA sono chimicamente molto simili; però la piccola differenza ha conseguenze notevoli sul comportamento delle due molecole, che sono

entrambe costituite da lunghi filamenti: nel DNA essi sono quasi sempre attorcigliati l'uno attorno all'altro e sono connessi tra di loro in modo regolare formando una doppia elica, mentre nell'RNA i filamenti sono quasi sempre singoli, e prendono forma aggomitandosi su se stessi, formando connessioni irregolari tra le anse del gomito. Doppie eliche di RNA esistono, ma hanno un uso molto limitato.

Le differenze di forma tra le due molecole determinano le loro proprietà biologiche: il DNA è molto stabile, è molto adatto a conservare l'informazione genetica essenzialmente immutata, mentre l'RNA è molto flessibile, può partecipare a interazioni con altre molecole cambiando forma a seconda delle circostanze. Per queste sue proprietà l'RNA può agire come catalizzatore, aiutando altre molecole a cambiare forma; si comporta perciò come una proteina. Così l'RNA è una molecola con molte funzioni: riceve l'informazione genetica dal DNA, la conserva intatta, la trasferisce ai siti dove si costruiscono le proteine, che poi esprimono la funzione dei geni, indipendentemente dalle proteine.

Il trasferimento di informazione dal DNA all'RNA avviene nel nucleo delle cellule per opera di un sistema di proteine che copia il DNA di un gene in una molecola di RNA; questa viene chiamata il «trascritto» del gene, perché l'RNA usa lo stesso linguaggio di quattro lettere del DNA, con una piccola differenza in una di esse. Il trascritto riproduce uno dei filamenti del DNA, quello che contiene l'informazione, mentre l'altro filamento è lì solo per mantenere la struttura.

Il trascritto viene modificato, anche sostanzialmente, dopo essere stato prodotto. La modificazione più importante deriva da un aspetto speciale della struttura del gene, che è formato da una serie di pezzi «codificanti», cioè capaci di determinare la formazione di una proteina, separati da pezzi che sembrano non avere alcuna funzione; questi ultimi sono chiamati «introni» perché sono dentro i geni. Nel processo noto come *splicing*, tutte le sequenze corrispondenti agli introni vengono eliminate dal trascritto una per una; il risultato è una molecola di RNA, nota come «messaggero», che contiene le sequenze codificanti. Il messaggero è poi responsabile della costruzione della proteina. Nel processo dello *splicing* ci sono spesso delle variazioni; nel saltare un introne, il trascritto può anche saltare un segmento codificante, o anche più di uno, cioè ci può essere uno «*splicing* selettivo». Questo dà luogo alla formazione di più di un messaggero, ciascuno dei quali utilizza alcuni dei pezzi codificanti; il numero di messaggeri corrispondente a un gene può arrivare a una decina. Nella

specie umana lo *splicing* selettivo avviene in almeno un terzo di tutti i geni.

La funzione del messaggero è di dirigere la formazione della proteina che, negli organismi superiori, non è fatta nel nucleo, ma nel citoplasma. Perciò il messaggero deve esser trasportato dal nucleo al citoplasma attraversando la membrana nucleare. Questo avviene attraverso una specie di porte, chiamate «pori», presenti nella membrana; e per fare ciò il messaggero deve ricevere una serie di modificazioni.

La regolazione dei trascritti

In una cellula non tutti i geni vengono trascritti: c'è un complesso sistema di controllo che determina quali geni vengano trascritti in un certo tipo di cellula e quali no. Lo spettro di geni trascritti varia da cellula a cellula, e le differenze sono responsabili delle varie caratteristiche delle cellule. Il controllo principale è basato sulla presenza, all'estremità di ciascun gene, di una sequenza speciale, la «zona di controllo» del gene. In un certo gene, tale zona è riconosciuta da «proteine regolatrici», che si legano a essa attivando il gene, cioè iniziandone la trascrizione. Per fare ciò le proteine si insinuano tra i due filamenti del DNA, permettendo l'entrata del complesso macchinario che opera la trascrizione. Dopo un certo tempo, in qualche modo le proteine regolatrici scompaiono, e la trascrizione del gene si ferma. Ci sono anche degli RNA regolatori, di cui parleremo più avanti.

L'azione delle proteine regolatrici è fortemente influenzata da segnali che provengono dall'esterno della cellula, e che sono necessari per coordinare l'attività dei diversi geni per gli scopi dell'intero organismo. In questo modo l'attività dei geni attivi in una cellula è interconnessa, vale a dire che esiste una rete di interazioni. Questo permette una buona coordinazione delle funzioni dei vari geni; d'altro canto, se un gene viene alterato o cessa di funzionare, l'effetto si riflette su tutti gli altri. La situazione può essere paragonata a quella di una rete di binari in cui i treni viaggiano in varie direzioni; ci sono binari attivi e binari morti. Quello attivo è controllato da un semaforo che risponde sia alle esigenze di quel percorso, sia allo stato degli altri binari. Se un gene non entra in azione quando dovrebbe, è come se si rompesse uno dei semafori, bloccandosi sulla luce rossa. Allora il treno controllato da quel semaforo si fermerebbe,

e progressivamente ciò sconvolgerebbe il funzionamento dell'intera rete. Le interazioni tra i geni sono la chiave delle grandi varietà di tipi cellulari che esistono nell'organismo, nonché dei continui cambiamenti di tipi cellulari che avvengono durante lo sviluppo dall'uovo fecondato.

Le caratteristiche dei geni di cui abbiamo parlato, cioè le varie parti che li costituiscono, sono le chiavi per l'identificazione di geni in una sequenza del genoma attraverso l'uso di computer con programmi sofisticati. Però le caratteristiche variano parecchio da gene a gene, e ciò rende la scoperta dei geni nella sequenza molto difficile. Tuttavia, già esiste una notevole esperienza sulle variazioni esistenti, tanto che si arriva quasi sempre a un'identificazione esatta del gene. Certo si progredirà ancora parecchio in questo campo, come è accaduto nel passato.

Il significato della trascrizione

Il trasferimento dell'informazione da un gene al suo messaggero ha un significato speciale per capire le nostre origini, specialmente per due fatti: l'uso dell'RNA per trasportare l'informazione alla proteina e lo *splicing* degli introni.

Ci si potrebbe chiedere perché sia necessario l'RNA come intermediario, invece di trasferire l'informazione direttamente dal DNA alla proteina. L'uso dell'intermediario potrebbe essere reso necessario dalla presenza della membrana nucleare, che separa il nucleo dal citoplasma, dove si fanno le proteine; in realtà però ciò non è vero, perché anche i batteri, che non hanno membrana nucleare, usano l'RNA come messaggero. L'interpretazione più verosimile è che l'uso dell'RNA riflette l'evoluzione della vita. Nel suo periodo iniziale non c'era il DNA, e i geni erano fatti di RNA (lo sono ancora oggi in alcuni virus). Il DNA venne più tardi, e si stabilì perché è più adatto a conservare l'informazione genetica. Quando avvenne il trasferimento, il meccanismo per fare le proteine già esisteva, ed era basato sull'uso dell'RNA; evidentemente non fu possibile attuare un trasferimento diretto dal DNA, e l'RNA rimase come intermediario. Si può pensare al cambiamento come a un rimaneggiamento in una ditta: i geni esistenti, fatti di RNA, si dimostrarono inefficienti, e vennero ridotti al ruolo di messaggeri, mentre sopra di essi furono piazzati geni di un altro tipo, quelli fatti di DNA.

Un'altra domanda che ci possiamo porre è perché i geni contengono

introni, che sembrano non avere alcuna funzione: infatti quasi tutti i geni di batteri non hanno introni, e i geni prodotti artificialmente, senza introni, funzionano benissimo. La spiegazione più verosimile è che gli introni sono sequenze estranee che hanno invaso il DNA durante l'evoluzione. Infatti i genomi di tutte le specie contengono sequenze autonome che si moltiplicano indipendentemente dal genoma stesso. Esse sono componenti rivoluzionarie del genoma, completamente egoistiche, di solito mantenute sotto controllo; ma di tanto in tanto una nuova copia di una di esse si forma e si insedia in un'altra parte del genoma; così esse possono aumentare di numero. Gli introni degli organismi superiori non hanno caratteristiche invasive, mentre quelli di organismi primitivi, quali le alghe unicellulari, le hanno. Gli introni degli organismi superiori possono dunque avere la stessa origine, ma aver perduto in seguito la capacità di invadere; questo sarebbe un risultato necessario dell'evoluzione per garantire la stabilità degli organismi. Gli introni negli organismi superiori hanno anche acquistato un nuovo e importante ruolo: quello di controllare lo *splicing* dei trascritti dei geni, moltiplicando così il potere dei geni quali controllori dell'informazione.

Il mondo dell'RNA

La conoscenza degli RNA messaggeri potrebbe far pensare che l'RNA ha una posizione secondaria a quella del DNA. Ma le osservazioni più recenti dimostrano che l'RNA costituisce un mondo a sé, indipendente da quello del DNA, ma connesso a esso. Infatti all'origine della vita l'RNA fu la prima molecola depositaria dell'informazione genetica, che trasmetteva da cellula a cellula, e che nello stesso tempo utilizzava per promuovere le funzioni delle cellule. Questo era possibile per le caratteristiche dell'RNA di cui abbiamo già parlato. Tali proprietà erano essenziali per la vita nel mondo molto primitivo (esistente) di circa quattro miliardi di anni fa. Però, con il progredire dell'evoluzione, la flessibilità dell'RNA si dimostrò pericolosa perché non sempre manteneva inalterata l'informazione genetica. Questo portò, circa 3,5 miliardi di anni fa, alla formazione del DNA che ereditò l'informazione dell'RNA, e da allora la mantenne pressoché inalterata. Allora l'RNA diventò il messaggero, capace di trasferire l'informazione del DNA alle proteine, e anche di usarla direttamente per certe funzioni; e le proteine presero il posto dell'RNA

nell'espletare la maggior parte delle funzioni dei geni.

Il mondo dell'RNA, com'è noto oggi, contiene diversi tipi di molecole. Predominanti sono i messaggeri, di cui abbiamo già parlato; numerosi sono gli «RNA ribosomali» che, assieme a proteine, costituiscono i ribosomi, cioè le particelle su cui si ancorano i messaggeri per dar luogo alla sintesi delle proteine; meno numerosi ma sempre importanti i «ribozimi», catalizzatori fatti di RNA che hanno funzioni simili a quelle degli enzimi proteici; e infine gli *small interfering RNA* che controllano la funzione di altri geni. Importanti sono anche gli RNA che causano l'inattivazione di uno dei cromosomi X nelle femmine; essi sono prodotti dal gene Xist, che è specializzato per questa funzione. Il loro modo d'azione non è chiaro. Un fattore essenziale dell'azione di tutti questi RNA è la loro flessibilità strutturale, che favorisce l'interazione con altre molecole e permette l'azione enzimatica; l'altro fattore è la specificità di interazione con altri acidi nucleici, basata sulla sequenza delle basi. Molte delle funzioni dell'RNA consistono nell'interagire con altri RNA in processi di grande significato biologico, come è dimostrato dai seguenti esempi sul ruolo di vari tipi di RNA.

RNA ribosomali. Quando un ribosoma funziona nella sintesi di una proteina, esso contiene tre tipi di RNA: il messaggero, l'RNA che è parte permanente del ribosoma, e un piccolo RNA che porta con sé l'amminoacido che deve essere aggiunto alla proteina in via di formazione. In aggiunta il ribosoma contiene anche altre proteine. Il trasferimento dell'amminoacido al filamento proteico in via di formazione richiede il distacco dell'amminoacido dal piccolo RNA e la sua connessione alla proteina. Entrambe queste funzioni sono il compito dei vari RNA presenti nel ribosoma. Le proteine del ribosoma non partecipano direttamente in questo processo; esse hanno il compito di raggiungere gli RNA ribosomali, e mantenerli nella posizione adatta. Perciò l'RNA è direttamente responsabile per la formazione della nuova proteina, e lo fa con supporto di proteine formate precedentemente.

Ribozimi. Queste sono molecole di RNA che hanno un'attività catalitica paragonabile a quella di proteine specializzate note come «enzimi». All'origine della vita gli RNA erano solo i catalizzatori, con il compito di attuare la funzione dell'informazione genetica che essi portavano. I ribozimi che esistono oggi sono piccoli (40-160 basi) e usano la loro abilità catalitica per spezzare altri RNA o se stessi in due parti, e poi ricongiungere i pezzi in modo da generare la nuova funzione. Per esempio,

i messaggeri di esseri viventi semplici, come batteri, contengono in qualche gene delle sequenze paragonabili agli introni; i ribozimi hanno il compito di tagliare fuori questo segmento, e poi congiungere le estremità libere del messaggero. Anche negli organismi più avanzati, come l'uomo, l'RNA è essenziale per lo *splicing* degli introni, in cui mantiene il suo compito di riconoscere i punti di taglio e di effettuare i tagli.

Small interfering RNA (siRNAs). Fino a pochi anni fa si pensava che in qualsiasi organismo l'attività dei geni fosse controllata da proteine, che si riteneva fossero le sole a determinare se un gene deve o no produrre il messaggero e la proteina corrispondente al gene. Oggi è chiaro che anche gli RNA da soli possono controllare l'attività dei geni. Il primo indizio di tale regolazione emerse nel 1990, quando un ricercatore cercò di cambiare il colore dei fiori in una petunia, introducendo entro le sue cellule dei geni estranei. Il risultato fu sorprendente: non solo i colori attesi non comparvero, ma i fiori prodotti erano o senza colore o con colore a chiazze. Perciò i geni estranei avevano sconvolto l'espressione normale dei geni della pianta.

Questo fenomeno, chiamato «co-soppressione», è dovuto alla forte espressione dei geni introdotti, con una produzione di abbondante RNA messaggero. Un fenomeno simile può anche essere prodotto se una pianta viene infettata da un virus che abbia un genoma fatto di RNA. Fenomeni simili furono successivamente osservati nelle cellule di molti altri organismi, incluse quelle umane.

In tutti i casi l'effetto è prodotto dalla presenza entro le cellule di piccoli RNA a doppia elica, chiamati «siRNA». La loro formazione dipende di solito dalla presenza entro le cellule di RNA a doppia elica, che derivano o da virus o da RNA con sequenze ripetute. In entrambi i casi questi RNA sono tagliati da un enzima apposito, chiamato *dicer*, che genera frammenti da 21 a 25 basi; questi sono i siRNA. SiRNA sono anche prodotti in alcune cellule da geni appositi. I siRNA si associano a proteine, formando complessi che bloccano l'attività di geni con sequenze simili alle loro, o causando la distruzione del messaggero oppure bloccandone la trascrizione. I siRNA possono avere anche altri effetti sul genoma, alterando lo stato della cromatina perciò influenzando l'attività di molti geni. Questo effetto è simile a quello delle proteine regolatrici dei geni.

I siRNA sono anche strumenti eccellenti per studiare la funzione dei geni in una cellula. Ciò è dovuto al fatto che il gene inattivato da un siRNA ne condivide la sequenza. Perciò, conoscendo le sequenze di quasi

tutti i geni di un organismo, si possono sintetizzare siRNA specifici per ogni gene. Quando uno di essi viene introdotto in una cellula, esso inattiva il gene corrispondente. Già esistono librerie di siRNA corrispondenti a 8-10.000 geni umani, che vengono introdotti uno alla volta in cellule umane in coltura per determinare l'effetto dell'inattivatore di quel gene. Questa è una possibilità che parecchi anni fa nessuno avrebbe neppure sognato.

È anche possibile introdurre un siRNA in cellule di animali o umane, per cui si è sviluppato molto interesse nel loro uso per scopi terapeutici. Molti sforzi vengono fatti oggi per cercare di curare malattie come il cancro, malattie della retina dell'occhio o infezioni come l'AIDS. Questi tentativi sono a uno stadio iniziale. I primi risultati preliminari sono incoraggianti, per cui si va avanti, sebbene ci siano molti problemi da superare.