

**Dai geni ai genomi,  
dalla struttura alla funzione**



Michele Morgante

**Dai geni ai genomi,  
dalla struttura alla funzione**

Una rincorsa fra scienza  
e tecnologia alla ricerca  
dell'identità dei viventi

---

Prolusione all'inaugurazione  
dell'anno accademico 2019.2020  
dell'Università degli Studi di Udine

 **FORUM**



Cosa siamo e cosa determina ciò che siamo? Comprendere cosa specifica la nostra identità e quella di tutti i viventi è una questione che ha sempre affascinato l'uomo e forse oggi, mai come prima, siamo vicini a capirlo. Tuttavia, per arrivare dove siamo ora, abbiamo dovuto percorrere un cammino lungo e irto di ostacoli, che abbiamo superato grazie a scienza e tecnologia. In realtà, quando la genetica è nata, nel 1866, l'ambizione era forse minore, in quanto si cercava, partendo dalla semplice osservazione delle somiglianze fra genitori e figli, di spiegare quali meccanismi potevano giustificare il fatto che alcune delle nostre

**Dai geni ai genomi, dalla struttura alla funzione .5**

caratteristiche venissero ereditate di generazione in generazione. C'è voluto parecchio tempo per capire che il segreto dell'identità dei viventi, e quindi anche di quella di noi esseri umani, è scritto in una molecola chiamata DNA e, in particolare, in una più o meno lunga stringa di caratteri con un 'alfabeto' costituito da sole quattro lettere, A, C, G e T. Ci vorrà ancora del tempo per capire esattamente come quella lunga stringa dalla struttura così semplice riesca a determinare ciò che noi siamo.

L'inizio della genetica risale tradizionalmente al 1866 con il lavoro di Gregor Mendel, che derivò le leggi generali di eredità dei caratteri senza sapere quale fosse il materiale genetico che trasmette le informazioni di generazione in generazione. La genetica come disciplina scientifica è nata quindi senza alcuna conoscenza di ciò che il materiale genetico sottostante potesse essere e senza alcuna tecnologia che potesse aiutarne il progresso. Mendel non parlò neppure di 'geni', termine che fu coniato solo nel 1909 dal botanico

## **6. Dai geni ai genomi, dalla struttura alla funzione**

danese Wilhelm Ludvig Johannsen, ma piuttosto di 'elementi cellulari' denotando una visione particellare di tali elementi ereditari, poi dimostrata invece errata una volta che si iniziarono a intuire i legami tra i fattori ereditari e le strutture fisiche presenti nei nuclei delle cellule, ossia i cromosomi. Solo in seguito si giunse alla visione attuale che guarda all'intero 'genoma', parola coniata per la prima volta nel 1920 dal botanico tedesco Hans Winkler, proprio mettendo assieme, pare, i termini 'gene' e 'cromosoma'.

La genetica nacque quindi come una scienza formale che progredì notevolmente, nonostante l'assenza di informazioni sulle basi molecolari di questi fenomeni, grazie a un alto livello di astrazione. Ci volle un bel po' di tempo prima che, nel 1952, Alfred Hershey e Martha Chase conducessero l'esperimento che dimostrò in modo conclusivo che il DNA (acido desossiribonucleico), e non l'RNA o le proteine, è davvero il materiale genetico che trasporta le informazioni da cellula a cellula e di generazione in generazione. Fu solo un

**Dai geni ai genomi, dalla struttura alla funzione .7**



anno più tardi (67 anni fa) che James Watson e Francis Crick descrissero la struttura a doppia elica del DNA, fornendo le evidenze su come l'informazione viene immagazzinata in questa molecola e come tale informazione possa essere fedelmente replicata.

Finalmente la genetica aveva una base molecolare. Tuttavia, si dovette attendere fino agli anni Settanta perché avesse anche una base tecnologica: prima, nel 1972, con lo sviluppo delle procedure di clonazione del DNA, che hanno permesso la moltiplicazione in vivo di molecole di DNA ricombinante, e poi soprattutto nel 1977 quando due gruppi, in modo indipendente, uno guidato da Frederick Sanger, l'altro da Walter Gilbert, svilupparono le prime metodiche di sequenziamento del DNA. La genetica aveva finalmente anche una tecnologia che la poteva supportare e poteva iniziare il lungo lavoro di decodificazione del libretto di istruzioni che è rappresentato dal nostro genoma: un libretto formato da tante parti diverse, capitoli e paragrafi, che nel tempo siamo riusciti a descrivere sempre con maggiore dettaglio.



## **8. Dai geni ai genomi, dalla struttura alla funzione**

Da allora gli sviluppi tecnologici si fecero molto più veloci, cercando di recuperare il ritardo che si era determinato rispetto ai progressi compiuti in genetica riguardanti la comprensione a livello molecolare dei meccanismi di ereditarietà dei caratteri. Nel 2001 è stato possibile ottenere la sequenza completa del genoma umano, contro il parere di molti scienziati che negli anni Novanta pensavano che sarebbe stata necessaria una nuova tecnologia di sequenziamento per raggiungere tale obiettivo, che invece fu ottenuto grazie a una serie incrementale di miglioramenti della tecnica-sviluppata da Frederick Sanger nel 1977. Possiamo considerare questo un momento in cui la scienza della genetica e la tecnologia di sequenziamento del DNA si trovarono in sincronia, allineate l'una rispetto all'altra, con l'identificazione di tutti i 3 miliardi di basi che compongono il nostro DNA e l'individuazione dei blocchi fondamentali, i 21.000 geni, o giù di lì, che definiscono la nostra identità come esseri umani. La genetica poté a questo punto finalmente cercare di affrontare il problema

**Dai geni ai genomi, dalla struttura alla funzione .9**

centrale di come i geni che sono nel nostro DNA interagiscono fra di loro e con l'ambiente per renderci ciò che siamo, avendo ormai a disposizione tutte le informazioni di base di cui aveva bisogno.

Nel frattempo, parallelamente agli sviluppi tecnologici che hanno consentito alla genetica di accedere in maniera diretta al materiale ereditario – essendo ormai in grado di leggere la sequenza dei caratteri che compongono il genoma –, si è andato evolvendo anche il concetto di gene, quella parte del genoma con cui tutti hanno familiarità. A partire dalla visione di Mendel, di cui abbiamo già parlato, di gene come particella fisica, e da quella di Johannsen, il quale come abbiamo detto conìò il termine gene che vedeva come pura astrazione, abbiamo assistito al passaggio da una definizione che dapprima si è andata facendo via via più concreta e fisica, avendo trovato una sua collocazione precisa nella sequenza di DNA, a una definizione più nuovamente più astratta come quella attuale.

## **10. Dai geni ai genomi, dalla struttura alla funzione**

Ma andiamo con ordine: siamo passati dal considerare il gene da un punto di vista puramente formale, come un fattore genetico puntiforme e indivisibile che aveva una certa posizione lungo il cromosoma ed era in grado di influenzare una caratteristica dell'organismo, a pensarlo come unità funzionale e divisibile, riconoscendogli una natura fisica corrispondente a uno spazio lineare che poteva essere ulteriormente suddiviso. Dal momento in cui vi fu lo sviluppo della genetica molecolare e delle tecniche di sequenziamento del DNA la comprensione di come il gene era concretamente fatto progredì notevolmente, riconoscendo che esso è una porzione del DNA che codifica per una proteina attraverso la produzione di un intermedio, l'RNA messaggero (mRNA). Con questo sviluppo il gene divenne qualcosa di ancora più materiale e la definizione fu quindi ulteriormente modificata. Passaggio fondamentale in tutto questo fu la decifrazione del codice genetico, ossia dell'insieme di regole in base alle quali l'informazione contenuta nel DNA viene tradotta in proteine, compren-

dendo prima che si trattava di un codice a triplette (ossia formato da 3 caratteri) e dando poi a ciascuna delle 64 triplette possibili (combinando le 4 lettere) una corrispondenza univoca rispetto ai 21 segnali necessari, quelli per ciascuno dei 20 aminoacidi e quello di fine proteina. Questi esperimenti estremamente eleganti furono compiuti quando ancora il sequenziamento del DNA non era possibile. Ma anche questa definizione venne poi estesa, per ricomprendere nel gene porzioni che vanno a far parte dell'RNA messaggero senza per questo codificare per una proteina (localizzate all'inizio e alla fine dell'mRNA) e poi pure porzioni che inizialmente vengono trascritte in RNA per esserne successivamente eliminate, dette introni. Fino qui eravamo ancora in grado di dire dove fisicamente iniziava e finiva un gene. Anche questa definizione però non durò molto a lungo, perché ben presto ci si rese conto che, oltre a quella porzione del genoma che determina il prodotto di un gene (ad esempio quella che va formare l'mRNA) si doveva ricomprendere altresì tutte quelle re-

## **12. Dai geni ai genomi, dalla struttura alla funzione**

gioni che stabiliscono quanto, quando e dove il prodotto del gene viene formato attraverso il processo di trascrizione, ossia quanto, quando e dove il gene viene espresso, le cosiddette 'sequenze regolatorie', che – oggi sappiamo – possono trovarsi a monte, a valle e all'interno del gene, ma che, in maggioranza, non siamo ancora in grado di identificare e localizzare con precisione lungo la sequenza di DNA. Quindi oggi non siamo in grado di dire dove comincia e dove finisce ciascun gene e, da questo punto di vista, non possiamo neanche dire di avere in mano le sequenze dei geni, poiché conosciamo le sequenze che danno origine alla molecola di RNA, e quindi al trascritto, ma il più delle volte non sappiamo dove siano tutti i pezzettini che servono a determinare quanto e quando quel gene viene trascritto. La complessa organizzazione genomica delle sequenze regolatorie e la nostra incapacità di definire, oltre a un codice genetico, anche un codice regolatorio rendono nuovamente di difficile definizione il concetto stesso di gene. Nonostante i molteplici tentativi, i bio-

logi oggi non hanno raggiunto un consenso su una definizione molecolare del gene ma, soprattutto, le analisi genomiche più recenti spingono a riconsiderare lo stato della sua definizione. Si tratta di un'astrazione (come aveva sostenuto già nel 1909 Johanssen) di cui abbiamo bisogno operativamente e che trova il suo fondamento nelle caratteristiche visibili, ossia il fenotipo, alla cui determinazione il gene contribuisce: non sappiamo però a cosa realmente corrisponda nel genoma e potrebbe anche non trovare una corrispondenza ben definita.

Riprendendo il nostro cammino fra scienza e tecnologia, dopo il 2001 le tecniche di sequenziamento del DNA hanno subito una vera e propria rivoluzione con l'avvento delle cosiddette tecnologie di sequenziamento di nuova generazione (*Next Generation Sequencing*, NGS), che hanno trasformato quello che prima era un processo seriale in uno altamente parallelo, dove centinaia di milioni di molecole di DNA vengono sequenziate tutte assieme, senza bisogno che le singole

#### **14. Dai geni ai genomi, dalla struttura alla funzione**

molecole siano spazialmente separate come nel metodo tradizionale.

In confronto solamente a quindici anni fa, la quantità di sequenza prodotta per giorno è aumentata di 1.640.000 volte e il costo per base è diminuito di 100.000 volte. Grazie a questi enormi progressi sia in termini di processività che di costi, che ricordano quelli previsti dalle leggi di Moore per la capacità di calcolo, oggi non solo abbiamo sequenziato i genomi di centinaia di specie e di decine di migliaia di individui all'interno di molte di queste specie, attraverso un processo di produzione dei dati in continua accelerazione, ma usiamo i metodi NGS anche per analizzare l'espressione genica (avendo sostituito i *microarrays*), per determinare il genotipo di milioni di posizioni del DNA (sostituendo ad esempio i chip di analisi dei *Single Nucleotide Polymorphisms*, SNP, che vengono usati per i test commerciali sul DNA umano reperibili online per i consumatori), per determinare la composizione qualitativa e quantitativa di comunità microbiche complesse (approccio

metagenomico che sostituisce quelli microbiologici tradizionali), per analizzare lo stato epigenetico del DNA (conformazione della cromatina, modificazioni istoniche o delle basi del DNA sostituendo una serie di tecnologie prima usate nel settore dell'epigenetica), per analizzare la struttura tridimensionale dei cromosomi entro i nuclei (sostituendo metodi basati sulla microscopia) e, infine, per analizzare praticamente tutti i fenomeni sopra descritti a livello di singole cellule.

Inoltre, il sequenziamento NGS sta trovando anche ampia applicazione nel campo della diagnostica clinica: senza dimenticare che la rapida e ampia caratterizzazione a livello di sequenza dei ceppi del coronavirus nCoV-2019, che oggi è di drammatica attualità, è stata possibile grazie a queste tecniche, il sequenziamento NGS viene già usato ampiamente nella diagnostica oncologica e in quella prenatale, dove consente di determinare anomalie cromosomiche nel feto con un'analisi non invasiva effettuata con un semplice prelievo del sangue della gestante; oltre a ciò, si iniziano a

## **16. Dai geni ai genomi, dalla struttura alla funzione**

sviluppare metodi diagnostici estremamente innovativi che, sulla base dell'analisi di varie caratteristiche del DNA circolante nel sangue del paziente, permetteranno di anticipare considerevolmente la diagnosi di numerose patologie di grande impatto.

Potremmo quindi pensare che, grazie all'accresciuta capacità di produrre dati di sequenza, che sicuramente non è più un fattore limitante nella ricerca genetica e genomica, siamo destinati a un rapido progresso scientifico verso una comprensione completa della struttura e soprattutto del funzionamento dei genomi, e di conseguenza verso una comprensione completa di come si determina la nostra identità di esseri viventi. In realtà, oggi la nostra capacità di produrre dati di sequenza ha notevolmente superato la nostra capacità di attribuire un significato biologico a tali sequenze. Di nuovo, come agli inizi degli studi genetici, la tecnologia non è più in sincronia con la scienza: questa volta però è la scienza della genetica a essere in ritardo e a dover rincorrere i progressi tecnologici.

**Dai geni ai genomi, dalla struttura alla funzione .17**

Perché è in ritardo la genetica? Soprattutto a causa delle nostre difficoltà nella comprensione del funzionamento del genoma. Quando è stato completato il primo sequenziamento dell'intero genoma umano, sembrava che la chiave per risolvere tanti dei misteri della biologia fosse stata finalmente trovata. È apparso subito evidente, però, che c'è ancora molto da capire e che non è sufficiente conoscere l'intera sequenza del DNA per comprendere i processi che ci mantengono vivi, e nemmeno quelli che ci fanno ammalare. Abbiamo pensato forse troppo ottimisticamente che, una volta determinata la struttura, la funzione ne sarebbe facilmente discesa. Qui possiamo cominciare ad apprezzare una differenza fondamentale tra come procede la biologia rispetto alla fisica.

Diversamente da ciò che succede in fisica, in cui spesso lo sviluppo di una teoria conduce alla definizione degli esperimenti necessari per convalidarla, come è stato recentemente esemplificato dalla vicenda della scoperta del bosone di Higgs, in genetica abbiamo

### **18. Dai geni ai genomi, dalla struttura alla funzione**

effettuato un enorme esperimento, come il sequenziamento del genoma umano, e abbiamo raccolto una immensa quantità di dati sostanzialmente senza avere un quadro teorico robusto che potesse aiutare a capirne il significato, forse pensando che i dati sarebbero stati auto-esplicativi, e siamo ancora in attesa di tale teoria. L'avvento della biologia e della genetica molecolari sembra aver portato a perdere quel notevole livello di astrazione e riflessione teorica che ha caratterizzato non solo il lavoro di Mendel sulle leggi di eredità o quello di Ronald Fisher sulla teoria della genetica quantitativa, ma anche molto più tardi il lavoro pionieristico di alcuni dei padri della genetica molecolare come Francis Crick o Sydney Brenner, solo per citare un paio di premi Nobel.

Cercheremo ora di analizzare quali sono i punti critici da affrontare per tentare di avvicinarci al nostro obiettivo, che deve essere quello di una comprensione completa del funzionamento dei genomi, oltre che della loro struttura, con una prospettiva sia di ca-

rattere generale che di carattere particolare, calata nella realtà italiana.

La genomica oggi, come abbiamo visto, fa uso di una serie di tecnologie che portano alla produzione di grandi set di dati i quali necessitano di nuovi metodi di analisi: per rispondere a questa esigenza, a partire dagli anni Novanta è nata la bioinformatica, settore eminentemente interdisciplinare che, combinando biologia, informatica, ingegneria dell'informazione, matematica e statistica, sviluppa metodi e strumenti software per comprendere e descrivere processi biologici. Abbiamo bisogno di ulteriori e rapidi avanzamenti in questo settore che riguardino sia l'hardware, per migliorare la capacità e la velocità di calcolo e di immagazzinamento dati e l'accesso a database sempre più di grandi dimensioni, sia il software, per lo sviluppo di innovativi algoritmi di analisi in grado di combinare tipologie di dati molto eterogenei, al fine di determinare rapporti di causa-effetto che riescano a spiegare i fenomeni biologici oggetto di studio. Ma abbiamo soprattutto

**20. Dai geni ai genomi, dalla struttura alla funzione**

bisogno di preparare una nuova generazione di biologi e genetisti che, grazie a solide basi di matematica, statistica, fisica e informatica, sappiano utilizzare al meglio i software di analisi e con questi indagare la grande mole di dati disponibili anche pubblicamente per arrivare a una comprensione più completa dei meccanismi di funzionamento dei genomi. Così come a partire dagli anni Settanta la biologia ha iniziato a utilizzare metodi molecolari fino ad arrivare alla situazione attuale in cui tale scienza è pressoché completamente di tipo molecolare, assisteremo presto alla transizione verso una biologia che fa sempre più uso di metodi computazionali, i quali andranno spesso a sostituire o complementare i tradizionali metodi di analisi di laboratorio e permetteranno alla biologia stessa di essere molto più predittiva e non solo descrittiva. Certamente uno dei grandi problemi da affrontare per capire il funzionamento dei genomi è quello di comprendere il significato di quella grande porzione dei genomi che non corrisponde a regioni geniche. Non bisogna

dimenticare che solo meno del 2% del DNA umano è fatto di geni codificanti proteine o geni classicamente definiti. Nella restante porzione ci stanno sicuramente tutti quegli elementi regolatori che sono decisivi nel determinare come ogni gene venga espresso, ossia quando, quanto e dove. Solo una minoranza degli elementi regolatori è oggi conosciuto e il sistema di controllo trascrizionale dell'espressione genica appare sempre più complesso e spesso dipendente da effetti a lunga distanza. Siamo ancora molto lontani dal poter predire *in silico* il pattern di espressione di un gene basandoci solo sulla conoscenza delle sequenze circostanti. Francis Crick, uno dei padri della genetica molecolare e lo scopritore assieme a James Watson della struttura del DNA, era più ottimista circa la nostra capacità di comprendere la complessità dei sistemi di regolazione genica, visto che, quando nel 1969 gli fu chiesto, in occasione del centenario della rivista «Nature», di descrivere la biologia molecolare nel Duemila, dichiarò che questo era uno dei problemi che

## **22. Dai geni ai genomi, dalla struttura alla funzione**

entro l'inizio del nuovo millennio sarebbe stato ovviamente risolto.

Un altro problema che Crick pensava potesse essere appianato entro il Duemila era il significato delle sequenze ripetitive che sono così abbondanti nei genomi eucarioti. Oltre il 53% del nostro genoma è fatto di sequenze ripetute, con il 45-50% corrispondente a elementi trasponibili, ossia sequenze di DNA che hanno la capacità di spostarsi da un punto all'altro all'interno di un genoma e la possibilità di creare a ogni spostamento una nuova copia. Ci sono altri genomi che sono ancora più ripetitivi, come ad esempio il genoma del mais che, pur con una dimensione simile a quella del genoma umano, è ripetitivo per più dell'80% della sua dimensione per la presenza di elementi trasponibili ancora molto attivi. La nostra visione del ruolo di questa grande massa di DNA è cambiata notevolmente nel corso degli ultimi anni e si è progressivamente trasformata dal considerare la frazione ripetitiva come una porzione inerte e dispensabile del genoma di cui potremmo tranquillamente

fare a meno (il cosiddetto 'DNA spazzatura') al vederla come una parte sempre più attiva, che svolge un ruolo importante nel controllo della regolazione genica direttamente o attraverso meccanismi epigenetici. Un esperimento di DNA-suzione, in cui andiamo a rimuovere tutte quelle porzioni del genoma che corrispondono a elementi trasponibili, se fosse tecnicamente possibile, potrebbe darci una risposta definitiva sul ruolo che questi elementi hanno nel determinare il funzionamento del nostro genoma; ma anche in mancanza dei risultati di un simile esperimento, a cui credo nessuno si vorrebbe sottoporre, se pure fosse fattibile, possiamo ragionevolmente pensare che un genoma umano sottoposto a una tale 'cura dimagrante' non darebbe più origine a un individuo che classificheremmo come appartenente alla nostra specie.

Un ulteriore livello di complessità che sta rendendo più difficile il nostro sforzo di comprensione del funzionamento dei genomi deriva dal fatto che le informazioni sono memorizzate non solo nell'alfabeto a quattro

#### **24. Dai geni ai genomi, dalla struttura alla funzione**

lettere (A, C, G, T) della sequenza di DNA, ma anche in modificazioni che possono essere sovrapposte ad esso e che vengono definite 'modificazioni epigenetiche'. Queste consistono in modificazioni chimiche delle basi del DNA e delle proteine che interagiscono direttamente con il DNA e sono trasmissibili attraverso la mitosi (quindi da cellula a cellula dello stesso organismo), ma in parte anche attraverso la meiosi e quindi da generazione a generazione. Sappiamo adesso che esistono meccanismi epigenetici che controllano l'utilizzo dell'informazione contenuta nel nostro DNA e che sono alla base di processi fondamentali per la vita, come sviluppo embrionale, adattamento all'ambiente e metabolismo. 'Epigenetica' vuol dire 'al di sopra della genetica', cioè si tratta di fattori che incidono sul funzionamento del DNA senza cambiare la sua sequenza. Le modificazioni epigenetiche sono in parte, e forse in gran parte, responsabili per la varietà di tipi cellulari che possono essere originati dalla medesima e comune informazione contenuta nel genoma,

perché sono decisive nel differenziare i profili di espressione genica tra cellule, tessuti e organi. Tutte le nostre cellule contengono, in prima approssimazione, la stessa informazione genetica, però sono funzionalmente diverse una dall'altra poiché ciascuna ne usa solo una parte, quella cioè necessaria al suo funzionamento come cellula della pelle, cellula muscolare, neurone ecc. I meccanismi epigenetici introducono 'segnali' sul DNA che indicano a ogni cellula le porzioni di DNA che deve leggere e quelle che deve saltare, quindi sono come dei segnalibri. Diversamente dalla variazione genetica, la variazione epigenetica può essere direttamente influenzata dall'ambiente. In linea con la funzione essenziale che i meccanismi epigenetici svolgono nello sviluppo degli organismi e nell'interazione tra i nostri geni e l'ambiente, non sorprende dunque che alterazioni epigenetiche siano state implicate in alcune malattie, dal cancro all'autismo e alla disabilità intellettiva, dalla distrofia muscolare alle malattie neurodegenerative, cardiovascolari e metaboliche.

## **26. Dai geni ai genomi, dalla struttura alla funzione**

Questo ha delle implicazioni per affinare le nostre capacità di cura (alcune delle quali approdano già verso i primi studi clinici), dal momento che le modificazioni epigenetiche sono reversibili (si possono inserire e rimuovere) e gli interruttori che ne sono responsabili sono bersagli privilegiati di nuovi farmaci che intervengono appunto per rimuovere modifiche inserite nei posti sbagliati, oppure aggiungere quelle che mancano. Più in generale, l'interesse per l'epigenetica deriva dal fatto che l'ambiente, che è una fonte permanente di stimoli, ovvero di fattori di stress che gli organismi devono essere in grado di ricevere, elaborare, imparare e ricordare per sopravvivere, può indurre modificazioni epigenetiche: se queste possono essere stabilmente mantenute nel corso del tempo ed essere trasmesse da generazione a generazione, allora il nostro modo di pensare agli effetti degli stili di vita del genere umano dovrebbe cambiare radicalmente e si dovrebbero recuperare alcune delle idee in merito alla teoria evolutiva di Jean-Baptiste de Lamarck.

La combinazione delle metodiche di analisi della variazione epigenomica con le nuove tecniche di miglioramento genetico di piante e animali di interesse alimentare, basate in particolare sul *genome editing*, promette di poter rivoluzionare il campo della nutrizione, arrivando a una alimentazione che farà sempre più uso di un disegno razionale degli alimenti, per poter apportare i maggiori benefici possibili alla salute umana attraverso una dieta che si basi sulla conoscenza degli effetti dei nutrienti sull'attività dei geni. Si stanno accumulando sempre più evidenze che dimostrano che una quota importante della variazione determinata dall'ambiente, e quindi anche dalla nutrizione, può avere effetti persistenti che vanno oltre l'individuo in cui tali effetti si sono manifestati e possono persistere pure nelle generazioni successive. Da un lato, potremo quindi studiare in maniera molto dettagliata quali sono gli effetti della nutrizione sull'attività genica nelle cellule, sfruttando sia sistemi in vitro che in vivo, e capire se tali effetti determinano modificazio-

**28. Dai geni ai genomi, dalla struttura alla funzione**

ni epigenetiche durature trasmissibili entro una generazione o anche da una generazione all'altra attraverso meccanismi di memoria epigenetica; dall'altro, una volta individuati i singoli nutrienti che sono in grado di influenzare l'attività genica in maniera da conferire un beneficio o un danno alla salute, mediante il miglioramento genetico di precisione potremo andare ad aumentare nel primo caso o diminuire nel secondo il contenuto di tali nutrienti nelle piante e negli animali di cui ci nutriamo.

Un altro aspetto in cui sono richiesti considerevoli progressi è nel chiarire quali sono le relazioni fra la variazione individuale nelle sequenze genomiche, ossia nel genotipo, e la variazione individuale nelle nostre caratteristiche visibili, ossia nel fenotipo. Se 21.000 geni o giù di lì sono ciò che ci rende umani, ciò che rende ciascuno di noi più o meno differente da un altro trova le sue basi, oltre che negli effetti dell'ambiente, nella variazione di sequenza che è presente nei nostri genomi. La nostra conoscenza dei tipi e della frequenza

di queste varianti è enormemente aumentata negli ultimi dieci anni grazie all'avvento delle nuove tecnologie di sequenziamento, che ha permesso il risequenziamento di migliaia di genomi individuali a un costo giunto oggi a essere inferiore ai 1.000 euro. Le differenze a livello di sequenza fra individui della nostra specie si presentano con una frequenza di circa una ogni 700 basi del nostro genoma e sono responsabili, in proporzione più o meno grande a seconda del carattere considerato, della variabilità che osserviamo nelle nostre caratteristiche morfologiche (ad esempio altezza e peso) ma anche in quelle cognitive e comportamentali e nella propensione che abbiamo a sviluppare un'ampia serie di patologie. Comprendere tali relazioni ci potrà aiutare a capire i meccanismi molecolari alla base delle patologie, a individuare metodi di diagnosi estremamente precoce e a sviluppare nuove terapie mirate sui target molecolari. Tuttavia, questo potrebbe anche porci problemi etici perché nel momento in cui capiremo quanto le differenze che osserviamo fra di noi

### **30. Dai geni ai genomi, dalla struttura alla funzione**

sono geneticamente determinate, oltre che influenzate dalle interazioni con l'ambiente che ci circonda, e quindi saremo in grado di definire con precisione i confini di cosa è genetico e cosa non lo è, potremmo scoprire che non ci piace ciò che abbiamo imparato: concetti come libero arbitrio o altruismo, o teorie psicologiche che cercano di spiegare le nostre caratteristiche comportamentali e cognitive, che sono alla base stessa della nostra cultura e della nostra vita sociale, potrebbero infatti essere seriamente messi in discussione.

Se in campo umano il potenziale applicativo della elucidazione delle relazioni fra variazioni a livello di sequenza e variazioni a livello fenotipico è sicuramente notevole, le applicazioni possono essere ancora più immediate in campo agricolo. Comprendere le basi genetiche della variazione di caratteristiche complesse quali la resistenza ai patogeni oppure la capacità di sfruttare le risorse idriche e i fertilizzanti presenti nel suolo potrebbe consentire, soprattutto quando combinata con

le nuove tecniche di modificazione genetica e in particolare con il *genome editing*, di raggiungere molto più facilmente quell'obiettivo di intensificazione sostenibile dell'agricoltura mondiale individuato dalla FAO come fondamentale per coniugare crescita produttiva e aumento della sostenibilità ambientale nella produzione di alimenti.

Fra gli sviluppi tecnologici che ci aspettano nel settore della genomica, per aiutarci ad affrontare i problemi aperti di cui abbiamo appena parlato, appare sicuramente assai promettente quello che ci porterà a effettuare le numerose tipologie di analisi genomiche ed epigenomiche oggi disponibili a livello di singole cellule. Fino ad ora, per limiti di sensibilità delle tecniche utilizzate, ci siamo fermati ad analizzare miscele spesso estremamente eterogenee contenenti tipi cellulari molto diversi, ottenendo delle descrizioni quantitative e qualitative dei fenomeni molecolari che rappresentavano delle medie fra comportamenti a livello di singole cellule, che potevano essere estremamente diversi.

### **32. Dai geni ai genomi, dalla struttura alla funzione**

Da qualche tempo stiamo però osservando uno sviluppo tumultuoso di nuove tecnologie che ci permettono di descrivere il contenuto molecolare delle singole cellule osservandole nel loro contesto spaziale, consentendoci di comprendere meglio le differenze di comportamento fra cellula e cellula e il ruolo degli elementi genetici ed epigenetici nel determinare queste differenze.

Spero di non lasciarvi con l'impressione che nel campo della genetica e della genomica non abbiamo fatto molti progressi: in realtà, ne abbiamo davvero compiuti di enormi, ma c'è ancora molto da imparare e comprendere. Per fare questo abbiamo bisogno del contributo di altre discipline al fine di rendere la genetica e la genomica più predittive e meno descrittive, e soprattutto abbiamo bisogno di giovani di talento che vogliano affrontare grandi e complesse sfide scientifiche.

A livello di sistema paese, inoltre, sarà determinante avere centri di riferimento che mettano tecnologie di sequenziamento, di analisi genomica, epigenomica e bioinforma-

tica a disposizione della ricerca accademica, industriale e delle imprese.

Nel settore della genomica stiamo andando verso un modello di *big science* che ricorda quello da tempo adottato in molti campi della fisica. Tale modello richiede la costruzione di grandi infrastrutture di ricerca che possano centralizzare una serie di procedure analitiche per sfruttare le economie di scala e liberare i singoli gruppi di studio dalla necessità di produrre i dati primari. In tal modo si potrà consentire loro di dedicarsi con maggiore profitto a quella ricerca creativa e di frontiera che ci consenta di avvicinarci sempre più all'obiettivo di comprendere il funzionamento dei genomi e prevederne le manifestazioni fenotipiche.

L'Italia non ha una tradizione, almeno in campo biologico, nell'affidarsi a infrastrutture centralizzate, come invece hanno molti altri paesi avanzati.

Come Università di Udine, grazie alla nostra esperienza e ai numerosi risultati ottenuti nella genomica, nell'epigenomica e nella bio-

#### **34. Dai geni ai genomi, dalla struttura alla funzione**

informatica, stiamo attivamente lavorando, assieme ad altri enti di ricerca e trasferimento tecnologico del territorio attivi in questi settori, alla creazione di una di tali piattaforme che possa mettersi al servizio del sistema della ricerca e delle imprese operando in ambito nazionale e internazionale.



## Le parole per capire

**Acidi nucleici.** Macromolecole organiche polimeriche a struttura lineare formate da unità ripetitive, chiamate nucleotidi, preposte a veicolare l'informazione genetica. Nei viventi sono presenti due tipi di acidi nucleici, il DNA e l'RNA.

**Allele.** Forme alternative dello stesso gene che possono portare a manifestazioni fenotipiche diverse del carattere controllato dal gene.

*Allele dominante.* Una variante genica che, quando presente in eterozigosi, determina la manifestazione di un fenotipo uguale a quello da essa determinato quando è in omozigosi.

*Allele recessivo.* Una variante genica che, quando presente in eterozigosi, non influenza la manifestazione del fenotipo.

**Aploide/diploide.** Le cellule o gli organismi aploidi sono quelli che hanno un solo corredo cromosomico e quindi una sola copia di ogni gene. Le cellule o gli organismi diploidi sono quelli che hanno due corredi cromosomici, uno di origine materna e l'altro di origine paterna, e quindi due copie di ogni gene.

**Bioinformatica.** Settore altamente interdisciplinare, cresciuto in maniera esplosiva a partire dagli anni Novanta, che sviluppa metodi e strumenti software per comprendere e descrivere processi biologici. La bioinformatica combina biologia, informatica, ingegneria dell'informazione, matematica e statistica. Ha una stretta attinenza con la biologia computazionale.

**Cromatina.** Complesso formato da DNA e proteine che si trova nel nucleo delle cellule eucarioti. Le principali proteine presenti nella cromatina insieme al DNA sono gli istoni. La funzione primaria della cromatina è quella di impacchettare le lunghe molecole di DNA rappresentate dai cromosomi in strutture più dense e compatte, determinandone però al tempo stesso la funzionalità. La struttura della cromatina si può modificare dinamicamente nel tempo e nello spazio e ha un ruolo fondamentale nel regolare l'espressione dei geni.

**38. Dai geni ai genomi, dalla struttura alla funzione**

**DNA.** L'acido desossiribonucleico (*DeoxyriboNucleic Acid*) è la molecola principale nella quale è immagazzinata l'informazione che serve a costruire un organismo vivente. È costituita da due filamenti disposti con orientamento antiparallelo con informazione complementare, ossia dato un filamento è determinata anche la struttura dell'altro. Ciascuno dei due filamenti è composto da uno scheletro formato da gruppi fosfato e 2-deossiribosio (uno zucchero pentoso) al quale sono attaccate le basi azotate che possono essere di quattro tipi: adenina (A), citosina (C), guanina (G) e timina (T). I due filamenti si fronteggiano e le basi presenti su ciascuno di essi si appaiano formando legami a idrogeno fra di loro secondo regole definite: A contro T e C contro G.

**Epigenetica.** Branca della genetica che si occupa dello studio dei cambiamenti ereditabili (da cellula a cellula o da individuo a individuo) nell'espressione genica che avvengono senza che vi sia alcuna variazione nella corrispondente sequenza nucleotidica del DNA, ossia cambiamenti fenotipici ereditabili che non presuppongono cambiamenti genotipici. La variazione epigenetica è dovuta a modificazioni che vengono sovrascritte alla sequenza di DNA e consistono di modificazioni chimiche a carico del DNA (in particolare metilazione della citosina, C) e/o di modificazioni

chimiche a carico delle proteine che interagiscono con il DNA a formare la cromatina (istoni).

**Esoni.** Parti del gene che vanno a formare la molecola di RNA matura prodotta dal gene stesso, dopo che sono stati rimossi gli introni attraverso il processo di *splicing* (meccanismo di rimozione) dell'RNA.

**Espressione genica.** Processo attraverso il quale l'informazione presente nel gene sotto forma di sequenza di DNA viene trasformata in prodotti genici funzionali, che possono essere molecole proteiche o di acidi nucleici (in genere RNA). Consiste principalmente di trascrizione e traduzione ed è finemente regolata dalla cellula tramite meccanismi di regolazione genica che operano sia a livello di trascrizione che di traduzione.

**Fenotipo.** Insieme delle caratteristiche osservabili di un individuo che rappresentano la manifestazione visibile del genotipo e risultano dalla interazione fra genotipo e ambiente.

**Genoma.** Insieme del materiale genetico di un organismo, che nel caso degli eucarioti è costituito da DNA. Le piante hanno tre genomi: il principale è quello nucleare, gli altri due quelli del mitocondrio e quello del cloroplasto.

**40. Dai geni ai genomi, dalla struttura alla funzione**

**Genotipo.** Informazione ereditaria di un individuo per un certo carattere. Il genotipo è univocamente definito dalle sequenze di DNA presenti nell'individuo che contribuiscono alla determinazione del carattere in oggetto. Interagendo con l'ambiente, il genotipo determina il fenotipo dell'individuo per quel carattere, ossia la sua manifestazione visibile. Genotipi uguali possono dare fenotipi diversi (in funzione di un differente contributo ambientale) e genotipi diversi possono dare fenotipi uguali (in funzione di un differente contributo ambientale oppure in conseguenza dei rapporti di dominanza e recessività fra gli alleli).

**Introni.** Parti del gene che, dopo essere state inizialmente trascritte, vengono rimosse attraverso il processo di *splicing* dell'RNA e non vanno a formare la molecola di RNA matura. Gli introni in un gene sono sempre fiancheggiati da esoni e quindi ogni gene ha un numero di introni che è pari al numero di esoni meno uno.

**Junk DNA.** Per 'DNA spazzatura' si intende tutta quella parte dei genomi che corrisponde a DNA non codificante per proteine e apparentemente non funzionale. Nel genoma umano il 99% del DNA risulta non codificare per proteine.

**mRNA.** RNA messaggero, molecola di RNA che viene prodotta nel nucleo della cellula attraverso la trascrizione di un gene e che viene poi trasportata nel citoplasma della cellula dove, tramite l'interazione con i ribosomi, porta alla traduzione dell'informazione genetica contenuta nella sequenza del gene in una sequenza di aminoacidi che dà origine alla proteina codificata dal gene stesso.

***Next Generation Sequencing (NGS).*** Tecnologie di sequenziamento del DNA sviluppate a partire dal Duemila che, grazie alla trasformazione di quello che fino ad allora era un processo seriale in uno altamente parallelo, hanno portato a un enorme aumento della produttività del sequenziamento e a una notevole diminuzione dei costi, rendendo il sequenziamento del DNA la tecnologia di elezione per studiare un gran numero di processi biologici.

**Omozigosi/eterozigosi.** L'omozigosi è quella condizione in cui le due copie di un gene (quella di origine materna e quella di origine paterna) in un individuo diploide sono identiche. Al contrario, nell'eterozigosi le due copie sono diverse tra loro.

**Promotore.** Specifica regione di DNA, situata generalmente subito a monte del sito di inizio della

**42. Dai geni ai genomi, dalla struttura alla funzione**

trascrizione di un gene, alla quale si lega la RNA polimerasi per dare inizio alla trascrizione del gene stesso. Il legame della RNA polimerasi al promotore richiede anche il legame di un complesso di fattori di trascrizione.

**RNA.** L'acido ribonucleico (*RiboNucleic Acid*) è una molecola che ha una struttura molto simile a quella del DNA eccetto che per la natura dello zucchero, che è ribosio invece che desossiribosio, e per il fatto che le quattro basi in esso presenti sono ACGU invece di ACGT, con l'uracile che sostituisce la timina presente nel DNA. Nei virus può anche costituire il materiale genetico ereditario, ossia il genoma (vedi ad esempio retrovirus come quello dell'influenza o HIV, e coronavirus come nCoV-2019), mentre negli eucarioti può avere funzioni di trasferimento dell'informazione dal DNA alle proteine (mRNA), strutturali (RNA ribosomale che va a formare i ribosomi) o regolatorie (microRNA).

**Sequenze ripetitive.** Sequenze di DNA che si ripetono in copie multiple in un genoma. Le diverse copie possono essere presenti una a fianco dell'altra oppure essere disperse in posizioni multiple entro il genoma.

**Sequenziamento del DNA.** Processo attraverso il quale si determina la sequenza/ordine dei nucleotidi presenti in una molecola di DNA.

**Traduzione.** Processo biochimico mediante il quale l'informazione contenuta nell'RNA messaggero (mRNA) viene trasformata in una sequenza di aminoacidi che vanno a formare la proteina. L'informazione presente nell'mRNA viene letta a triplette di basi, con una corrispondenza univoca fra tripletta e aminoacido. Il codice di lettura delle triplette è detto codice genetico ed è ridondante, ossia vi sono 64 triplette che codificano per 21 segnali diversi, ciascuno dei 20 aminoacidi e il segnale di fine della proteina (codoni di stop). A ogni tripletta corrisponde un singolo aminoacido o segnale di stop, ma a ogni aminoacido possono corrispondere più triplette (ad esempio, ve ne sono tre per il segnale di stop). La traduzione avviene fuori dal nucleo cellulare, nel citoplasma della cellula, dopo che l'mRNA prodotto nel nucleo è stato portato nel citoplasma a interagire con i ribosomi, organelli nucleoproteici (formati da RNA e proteine) sui quali avviene la sintesi proteica.

**Trascrizione.** Processo attraverso il quale l'informazione contenuta nella sequenza di DNA viene trasferita a una molecola di RNA che potrà poi

#### **44. Dai geni ai genomi, dalla struttura alla funzione**

essere tradotta in proteina o svolgere la sua funzione in quanto tale. Solo uno dei due filamenti di DNA viene in genere trascritto (filamento stampo) per produrre un filamento di RNA che risulta identico all'altro filamento di DNA (filamento senso) ad eccezione della sostituzione della timina con l'uracile. La trascrizione avviene nel nucleo della cellula, l'organello che contiene il genoma.

**Trasposone.** Sequenza di DNA che può spostarsi all'interno di un genoma. Lo spostamento può avvenire secondo due modalità diverse, quella detta 'taglia e incolla' (che si adatta bene alla definizione di *jumping genes*) in cui l'elemento esce dalla posizione originaria per andare in una nuova, e quella detta 'copia e incolla' in cui l'elemento resta nella posizione originaria e crea una sua copia che va a inserirsi in una nuova posizione. I trasposoni o elementi trasponibili sono a tutti gli effetti dei parassiti intragenomici che cercano costantemente di aumentare in numero colonizzando il genoma ospite, il quale da parte sua mette in atto strategie per impedirne il movimento (ad esempio usando l'interferenza a RNA). Da una particolare classe di elementi trasponibili sono derivati alcuni dei più pericolosi parassiti virali dell'uomo, i retrovirus.