

Dal genoma alla proteoma passando per la trascrittoma

G. Novelli, E. Giardina

Dipartimento di Biopatologia e Diagnostica per Immagini
Università di Roma Tor Vergata

Il completamento del progetto Genoma Umano, annunciato nella primavera del 2001 ha riservato alcune sorprese, non del tutto preventivate all'epoca delle sue ambiziose premesse (1). Il progetto ha rivelato che ogni persona mostra il 99.9% di identità genetica rispetto ad una qualsiasi altra. Pertanto le caratteristiche proprie di ciascun individuo sono dovute praticamente al restante 0.1% di materiale ereditario che costituisce la variabilità interindividuale. Tale variabilità è sostanzialmente dovuta a piccole variazioni di sequenza (in pratica e molto spesso, sono sostituzioni di singoli nucleotidi che compongono il nostro DNA) chiamate SNPs (Single Nucleotide Polymorphisms). Si calcola che nell'uomo esistano circa 10 milioni di queste variazioni ed è ormai stabilito che gruppi di SNPs sommati insieme determinano la suscettibilità o la resistenza a molte delle malattie umane più comuni (ad es. il diabete, il cancro, l'aterosclerosi, l'Alzheimer). La maggior parte delle malattie che colpiscono l'uomo infatti, dipendono dall'interazione di diversi geni (componente genetica della malattia) con determinati fattori ambientali (componente ambientale della malattia). Dal momento che ognuno di noi, geneticamente, è più o meno di altri suscettibile a sviluppare alcune patologie, attraverso la comparazione delle differenze genetiche interindividuali sarà possibile determinare il contributo offerto dai geni a queste ed altre malattie. La suscettibilità ad una certa patologia naturalmente non coincide con la certezza di ammalarsi, bensì indica un aumentato rischio rispetto alla popolazione generale. I geni, per se, non sono sufficienti a scatenare la malattia: la comparsa dei sintomi è dovuta piuttosto ad un'interazione tra fattori genetici e fattori ambientali (stress, alimentazione, stili di vita). Tale variabilità interindividuale ha inoltre definitivamente screditato qualsiasi ipotesi razziale su base genetica. L'analisi delle variazioni individuali offre inoltre la possibilità di studiare i meccanismi genetici che regolano la risposta ai farmaci (2). Infatti i principi attivi dei farmaci raramente hanno un'efficacia assoluta e spesso sono responsabili di effetti collaterali particolarmente rilevanti (Tabella I). Basti pensare che negli Stati Uniti le reazioni avverse ai farmaci rappresentano tra la quarta e la sesta causa di morte. La presenza di differenziali risposte ai far-

Tabella I. prove di efficacia dei farmaci

Tipologia del farmaco	efficacia
antipertensivi	inferiore al 60%
antidepressivi	inferiore al 50%
antipsicotici	inferiore al 50%

maci è dovuta anche e specialmente alle differenze ereditarie nel metabolismo e nei bersagli dei principi attivi. L'accertamento dell'associazione tra determinati genotipi e la risposta a specifici farmaci consentirebbe la realizzazione di terapie individuali che massimizzerebbero l'effetto terapeutico minimizzando il rischio di reazioni avverse. La sopravvenuta necessità di studiare contemporaneamente un numero sempre maggiore di geni e la conoscenza della maggior parte della sequenza del genoma ha gettato le basi per lo sviluppo della tecnologia del Chip genomico. Con questo termine si intende un supporto solido (in genere silicio, plastica o vetro) sul quale vengono immobilizzate diverse centinaia o migliaia di molecole organiche di varia natura (spesso oligonucleotidi di DNA o RNA complementare). Attraverso la miniaturizzazione e la cibernetica è possibile fissare su un supporto di appena 1 cm per lato, fino a 400.000 molecole diverse corrispondenti a sequenze virali, batteriche, o umane contenenti mutazioni o SNPs. Tali vetrini vengono fatti ibridare con RNA o DNA di un soggetto rendendo possibile l'analisi simultanea di migliaia di segnali di ibridizzazione e permettendo la costruzione del profilo genetico dell'individuo. In tal modo è possibile identificare pazienti portatori di mutazioni a carico di uno o più geni contemporaneamente oppure osservare quali e quanti geni a potenziale attività oncogena siano attivi nelle cellule tumorali di un paziente rispetto ad un altro, o al tessuto normale. Inoltre la diagnosi di una malattia genetica o neoplastica richiede molto spesso la caratterizzazione di mutazioni diverse da individuo a individuo (ad esempio, talassemia, fibrosi cistica, distrofia muscolare di Duchenne, tumore della mammella, leucemie linfoblastiche acute del bambino). Lo studio di queste mutazioni in parallelo evidenzia l'enorme potenzialità di disporre di informazioni su numerose alterazioni o

stato di attività di geni nel medesimo istante. Sarà possibile, ad esempio, caratterizzare tutte le 900 mutazioni responsabili della fibrosi cistica attraverso un unico esame. La tecnologia del chip genomico viene anche utilizzata per studiare l'espressione, quindi il funzionamento dei geni attraverso la loro trascrizione (trascrittomica) e/o produzione di proteine. Tale approccio risulta particolarmente utile nell'indirizzare la ricerca di geni candidati, per patologie mendeliane o complesse la cui base genetica non è nota, esclusivamente verso i geni identificati dal Chip.

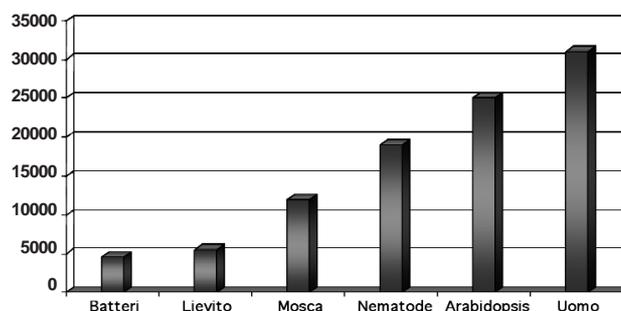
E' stato stimato che sarà possibile disporre di 50 microchip per contenere tutti i geni umani ed effettuare una analisi globale dello stato di attività in una cellula, in un tessuto o in un organo.

Tuttavia, il progetto genoma umano ha riservato alcune sorprese non del tutto preventivate. Infatti, il numero dei geni presenti nel genoma umano è di gran lunga inferiore a quello atteso, essendo circa 30000 (3). Il mondo scientifico si aspettava un numero di geni superiore, dal momento che organismi considerati più semplici rispetto all'uomo possiedono un paragonabile numero di geni (figura 1). Il progetto genoma umano ha quindi posto in risalto che la maggiore complessità della nostra specie non risiede in una maggiore quantità di geni! Tuttavia il nostro livello di complessità è giustificato dall'enorme quantità di prodotti proteici diversi che possono essere codificati. Infatti nell'uomo ogni gene in media codifica per tre proteine diverse a differenza degli altri organismi studiati in cui il numero dei geni e quello delle proteine non differiscono notevolmente. L'attenzione dei ricercatori pertanto si concentra adesso sulla proteomica ovvero l'insieme delle proteine presenti nelle cellule e nei tessuti umani. Infatti la definizione della nostra fisiologia dipende dalle complesse reti di interazione che si instaurano tra le proteine caratteristiche di ciascuna cellula. Alla luce di ciò i ricercatori, tanto nei laboratori privati quanto in quelli accademici, stanno cercando di catalogare tutte le proteine umane e di scoprire come interagiscano tra di loro. Alcuni risultati sono già stati ottenuti: lo scorso gennaio due gruppi di ricerca hanno pubblicato le mappe di interazione di tutte le proteine del lievito della birra (4-5). Nel febbraio scorso un altro gruppo ha annunciato di aver utilizzato tecniche di proteomica per sviluppare un test precoce e accurato per il tumore ovario. La proteomica, insomma, è pronta per diventare un grande business. Secondo un'analisi della Frost&Sullivan, il mercato globale per gli strumenti, i reagenti e i servizi legati alla proteomica raggiungerà circa 5,6 miliardi di dollari entro il 2005, dai 700 milioni di dollari del 1999. E questa stima non comprende i ricavi generati dai farmaci e dai kit diagnostici che saranno sviluppati grazie agli studi di proteomica. Gli studi di proteomica possono distinguersi in due classi differenti: la proteomica basata sulla quantità e la proteomica basata sulla funzione. Gli obiettivi della proteomica basata sulla quantità prevedono la quantificazione di specifici prodotti proteici in determinate cellule e sotto particolari condizioni. Tali studi dovrebbero consentire di individuare i prodotti proteici che differenziano una cellula normale da una anormale (ad esempio tumorale). Gli obiettivi della proteomica basata sulla quantità

prevedono invece la caratterizzazione di tutte le proteine di una cellula e lo studio delle funzioni da queste codificate. Attraverso il confronto dei corredi proteici di cellule con caratteristiche differenti dovrebbe essere possibile l'identificazione di quei particolari prodotti proteici che determinano la transizione tra stato fisiologico e patologico di una cellula.

E' tuttavia doveroso considerare che anche le fasciose promesse della proteomica presentano delle limitazioni notevoli. Innanzitutto alla fissità del genoma (nel tempo) non corrisponde quella del proteoma (una cellula contiene diverse quantità e qualità di prodotti proteici a seconda dell'età e del particolare stato fisiologico). Questa considerazione aumenta il numero, la complessità degli esperimenti necessari e può inficiare la loro imprescindibile riproducibilità. Inoltre ogni proteina può assumere nello spazio differenti strutture (folding) e a tali strutture corrispondono strettamente specifiche funzioni. Alla luce di questa considerazione sarà necessario studiare non solo tutte le interazioni proteiche presenti in una cellula, ma anche tutti i possibili avvolgimenti che tali proteine possono assumere per soddisfare specifiche funzioni. Pertanto anche la proteomica non costituirà un punto di arrivo per i ricercatori o una panacea, bensì un ulteriore e più accurato punto di vista da cui necessariamente ripartire con una diversa e più profonda consapevolezza della complessità della specie umana.

Figura 1. Numero di geni nell'uomo paragonato agli altri genomi noti



Bibliografia

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA et al. The sequence of the human genome. *Science* 2001; 291: 1304-51.
2. Ingelman-Sundberg M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J Intern Med* 2001; 250: 186-200
3. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, et al. Comparative genomics of the eukaryotes. *Science* 2000; 287: 2204-15.
4. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002; 415: 141-7
5. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002; 415: 180-3.