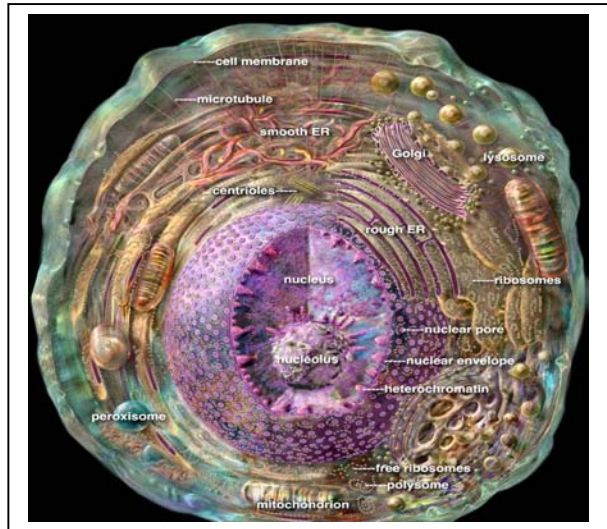


LA COMPRESIONE DEI SISTEMI BIOLOGICI NELL' ERA POST-GENOMICA: L'AVVENTO DELLA PROTEOMICA.

di Gianluca Tell¹

Cosa si intende per Proteoma?

Stiamo soltanto iniziando ad apprezzare la potenza e i limiti della rivoluzione genomica (volta all'identificazione e sequenziamento di geni, che ha portato nel 2001 al completamento del sequenziamento del genoma umano) che un altro innovativo aspetto della ricerca biologica, lo studio del Proteoma, già promette una trasformazione ancor più radicale nell'ambito della ricerca biomedica. Le proteine sono gli effettori finali responsabili dell'estrinsecarsi delle funzioni e dei meccanismi biologici e per comprendere come le cellule funzionino è indispensabile conoscere quali proteine siano presenti in un determinato tipo cellulare, in un determinato momento della vita della cellula, la loro struttura e funzione nonché come queste proteine interagiscano fra di loro a costituire delle reti o 'networks' condizionando la funzione cellulare. Il Termine Proteoma descrive l'intero complemento PROTEICO di un genOMA in una data cellula, tessuto od organismo. In senso lato, la ricerca proteomica comprende lo studio delle attività delle proteine, le loro modificazioni post-traduzionali, la localizzazione e le interazioni (fisiche e/o funzionali) fra proteine a costituire dei complessi sovramolecolari funzionali. E' senza dubbio la più importante sfida tecnologico-culturale che coinvolge biologi, medici, chimici, fisici, bioinformatici e matematici che sia mai stata lanciata al fine di una comprensione più reale e completa dei sistemi biologici e di come essi funzionano. Queste sfide tecnologiche hanno come obiettivo quello di sviluppare metodologie altamente sensibili, fedeli ed efficienti per l'identificazione e la caratterizzazione delle proteine e dei complessi multiproteici in campioni biologici in maniera quantitativa ed integrale.



Attraverso lo studio dei profili proteici globali e delle attività e di come questi cambiano durante il differenziamento cellulare, lo sviluppo embrionale o in condizioni patologiche, la ricerca Proteomica si pone come nuovo motore per la comprensione del funzionamento cellulare su base sistemica. Da questa ricerca di base, la ricerca clinica si aspetta di ottenere larghi benefici per l'identificazione di nuovi bersagli molecolari di farmaci e per lo sviluppo di nuovi marcatori diagnostici.

Come in precedenza per la Genomica, è necessario coordinare gli sforzi dei ricercatori di tutto il mondo al fine di affrontare la sfida titanica del nuovo secolo in maniera sinergica. A tale fine, la comunità scientifica mondiale si è organizzata in una organizzazione detta HUPO (Human Proteome Organization) con lo scopo di coordinare i vari progetti Proteoma in tutto il mondo. La quantità di informazioni che si stanno via via accumulando offre nuove sfide per ambiti scientifico-disciplinari finora poco toccati dai problemi biologici come l'informatica e la matematica. Inoltre, è necessario che vi siano strategie sperimentali comunemente adottate per poter confrontare i dati provenienti da laboratori diversi.

¹ Dipartimento di Scienze e Tecnologie Biomediche, Facoltà di Medicina e Chirurgia, Università degli Studi di Udine

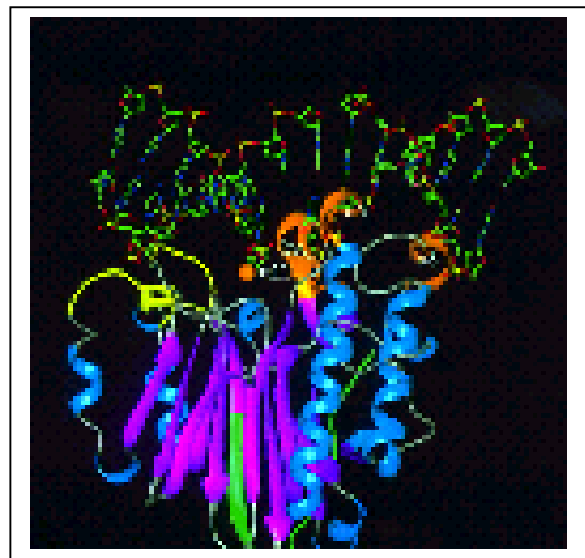
Di qui la necessità di sviluppare metodologie standard di ottenimento ed analisi e confronto dei dati di facile accesso e consultazione, tutti sforzi che si stanno attualmente perseguendo a livello planetario in quello che è un ottimo esempio di 'globalizzazione culturale'.

QUANTO E' COMPLESSO LO STUDIO DEL PROTEOMA?

Il titolo di una conferenza sulla Proteomica del 2001 riassume bene il concetto della difficoltà intrinseca in tali tipi di studi: "Il Progetto Proteoma umano: 'i geni erano semplici'". Le informazioni derivanti dall'ormai completato 'Progetto Genoma Umano' avevano dato l'illusoria presunzione di poter comprendere come l'alfabeto del DNA, costituito da 4 basi chimiche. Adenina (A), Citosina (C), Guanina (G) e Timida (T), potesse essere articolato a dare delle frasi di senso compiuto che fossero significative per il funzionamento del sistema biologico: i geni. In realtà circa 3 miliardi delle succitate basi, che costituiscono il patrimonio genetico di ciascuna nostra cellula, costituiscono l'informazione per 30-40000 geni che, a ben pensare, sono un ristretto numero per distinguerci da cosiddetti 'organismi inferiori' come il comune moscerino della frutta (*Drosophyla melanogaster*) che ne possiede circa 14000 o il lievito della birra (*Saccharomyces cerevisiae*) un organismo unicellulare che ne annovera circa 6000. Le proteine, invece, sono costituite da combinazioni multiple di 20 aminoacidi essenziali la cui successione è controllata dalle informazioni codificate nei geni. Dalla sequenza di un gene è sì possibile determinare la relativa sequenza di una proteina ma non la sua struttura terziaria e quaternaria, né le modificazioni post-traduzionali che essa può subire dopo la propria sintesi (come l'aggiunta di gruppi fosfato, di acidi grassi o di diversi zuccheri) ed, in ultima analisi, la sua funzione.

In realtà si è subito compreso che la filogenesi degli organismi non può essersi semplicemente evoluta in senso quantitativo (aumentando proporzionalmente alla complessità biologica il numero di geni e quindi la dimensione del genoma) ma piuttosto in senso qualitativo attraverso vari tipi di modificazioni che coinvolgano il prodotto dei geni stessi (molecole di mRNA) ed, in ultima analisi, le proteine. Un gene può

quindi dare origine, attraverso meccanismi molecolari ben controllati, a più proteine diverse in maniera molto 'economica' ed estremamente efficiente, in termini cinetici, per il sistema cellulare. Quindi, alla luce di quanto detto, il vecchio Dogma centrale della Biologia: un gene \Rightarrow una proteina non sembra essere integralmente valido ai giorni nostri, nell'Era Post-Genomica. Questo aumento di complessità sarebbe in grado di spiegare sia l'evoluzione della filogenesi degli organismi sia come, sebbene negli organismi pluricellulari tutte le cellule presentino il medesimo genoma, questo produca effetti diversi sul fenotipo cellulare dando origine, a partire da una cellula staminale totipotente, a determinate cellule con funzioni specializzate (cellula nervosa, muscolare, linfocita, ecc.). Quindi, l'alfabeto verrebbe utilizzato con diverse semantiche e facendo uso di grammatiche diverse che dobbiamo cercare di comprendere nel dettaglio per poterle modificare ed adattare 'ad hoc'.



Si stima che, mediamente, ogni tipo cellulare umano sia in grado di esprimere circa 10-15000 geni diversi dei 30-40000 potenziali di cui è fornito il proprio genoma. Da questo numero di geni espressi, ogni cellula sarebbe in grado di dare origine a circa 100000 proteine diverse. L'insieme delle proteine espresse da un organismo umano è stimato essere in circa un milione di specie proteiche diverse. Se consideriamo che queste possano interagire funzionalmente tra loro in maniera combinatoriale e dinamica nel tempo costituendo ogni volta una diversa informazione biologica funzionale, possiamo

comprendere da un lato la complessità dello studio del sistema Proteoma, dall'altro la sua versatilità nel controllare le diverse necessità biologiche di una cellula.

Il Genoma è, pertanto, un'unità statica dei sistemi biologici: non cambia, tipicamente, nel tempo e non si adatta a modifiche ambientali transitorie quali quelle costituite da stimoli extracellulari. Il Proteoma è, invece, un'unità dinamica: può infatti cambiare in risposta agli stimoli cui la cellula è sottoposta e a cui si deve adattare e si può modificare nel corso del tempo (pensiamo all'invecchiamento biologico degli individui, il loro genoma non è cambiato dalla nascita, o all'invecchiamento cellulare, che si chiama senescenza, e durante il quale le cellule perdono via via molte delle loro peculiari caratteristiche biologiche). Tutto questo è reso possibile attraverso il controllo della funzionalità delle proteine e non modificando i loro precursori nucleari, i geni, attraverso meccanismi molecolari quali la degradazione proteica, le modificazioni post-traduzionali, la relocalizzazione subcellulare, le modulazioni delle interazioni con altre proteine. Dato questo dinamismo, di fatto, sarebbe preferibile parlare pertanto di tanti Proteomi quante sono le diverse condizioni cellulari, oltre che le diversità cellulari, piuttosto che di un unico Proteoma.

COME STUDIARE IL PROTEOMA

Due sono principalmente le moderne tecniche che permettono di studiare quali proteine siano presenti in determinate cellule o tessuti: l'elettroforesi bi-dimensionale e la spettrometria di massa.

L'elettroforesi bi-dimensionale consente di separare le proteine di una miscela complessa sulla base delle loro caratteristiche fisico-chimiche: la carica elettrochimica e la massa molecolare facendo uso di opportune matrici polimeriche. Dopo questo tipo di separazione, ogni specie proteica si presenta come una macchiolina visibile ed isolabile dal resto delle proteine della miscela di cui originariamente faceva parte. Questa tecnica consente, attraverso una analisi di tipo differenziale comparativo, di individuare le proteine che siano, ad esempio, caratteristiche di una particolare condizione cellulare o patologica per un tessuto e che consentano quindi, in linea teorica, di individuare dei 'markers' di una

condizione patologica costituendo degli importanti strumenti in ambito diagnostico ed, eventualmente, terapeutico in medicina. Recentemente, ad esempio, alcuni ricercatori statunitensi hanno dimostrato l'efficacia di un approccio proteomico nella diagnosi del tumore ovarico confrontando l'organizzazione delle proteine presenti nel sangue di pazienti affette rispetto ad individui sani e riuscendo a discriminare tutte le pazienti affette dalla patologia.

La fase successiva viene effettuata mediante le moderne tecniche basate sulla spettrometria di massa con cui è possibile individuare precisamente l'identità delle proteine in esame.

Poiché, comunque, queste due metodologie presentano, nonostante l'enorme evoluzione avutasi e i vantaggi che offrono, molte limitazioni dovute alle complessità di indagine, numerose aziende biotecnologiche stanno studiando nuove soluzioni e versioni sempre più sofisticate di analisi per impiegarle in ricerche su base industriale dello stesso ordine di grandezza di quelle che hanno reso possibile il Progetto Genoma Umano. Si prevede che nel 2006 circa 6 miliardi di dollari saranno investiti a livello mondiale in strumentazione e servizi legati alla Proteomica di cui circa un terzo saranno appannaggio dello sviluppo di programmi bioinformatici finalizzati alla diagnostica proteomica e all'ingegnerizzazione del farmaco.

LE PROTEINE COME EFFICACI 'TARGET' FARMACOLOGICI

Da quanto detto finora, dovrebbe essere chiaro che se le proteine sono di fatto gli effettori responsabili della funzionalità biologica cellulare, queste sono anche responsabili dell'instaurarsi di stati patologici cellulari e quindi dell'insorgenza delle malattie. Senza dubbio le nuove frontiere della Proteomica aprono nuovi orizzonti e prospettive di farmacoterapia mirata ove risulta indispensabile conoscere la struttura tridimensionale della proteina stessa per poter disegnare opportunamente dei farmaci specifici in grado, ad esempio, di inibirne o promuoverne la funzione. In questo caso, ulteriori tecniche per la caratterizzazione della struttura tridimensionale delle proteine, non nuove ma estremamente avanzate ed evolute notevolmente negli ultimi anni, quali la cristallografia a raggi X con luce di Sincrotrone e la Risonanza Magnetica

Nucleare (NMR) che consentono una definizione molecolare a livello atomico (dell'ordine dei 2 Angstroms), costituiscono i cardini per i moderni approcci di farmacoterapia mirata.

Si calcola che il 99% dei bersagli farmacologici siano rappresentati da proteine e solo l'1% dal DNA. Di questo 99%, il 52 % sarebbe rappresentato da recettori cellulari di membrana, il 22% da enzimi, il 13% da ormoni e fattori di crescita, il 5% da canali ionici e il 2% dai recettori nucleari. Dato che i recettori di membrana costituiscono una categoria di proteine scarsamente espresse dalle cellule e molto difficili da studiare dal punto di vista strutturale, questo dà l'idea di quanto sia complesso questo ambito di indagine e della necessità di sviluppare nuove adeguate metodologie di caratterizzazione sempre più sensibili.

Dopo il boom della brevettazione di geni in ambito genomico, come effetto di ricaduta industriale del Progetto Genoma Umano, stiamo assistendo oggi ad un drastico mutamento del concetto di proprietà industriale in campo biotecnologico. Dalla brevettazione di sequenze di DNA e della singola sequenza proteica ad esso associata, ora si preferisce brevettare la singola proteina, dato che ciascun gene può dare origine a più proteine diverse. Questo consentirebbe di dare al brevetto molto più valore reale ma soprattutto potrebbe consentire di aggirare i brevetti sui geni già in mano della concorrenza. In questo caso le proteine avranno rubato ai geni anche lo scenario dei tribunali.

BIBLIOGRAFIA CITATA E CONTATTI IN RETE

- Tyers M. and Mann M. 2003 Nature 422, 193-197.
- Hanash S. 2003 Nature 422, 226-232.
- Boguski M.S. and McIntosh M.W. 2003 Nature 422, 233-237.
- Petricoin E.F. et al., 2002 Lancet 359, 572-577.
- www.genomicsglossaries.com/content/omes.asp
- www.hupo.org

