

Proteomica

Lezione 1 – Introduzione

Finalità del corso

Lo studio proteomico consiste nell'identificare proteine espresse in determinate condizioni e nel determinare le interazioni proteina-proteina al fine di definire il ruolo funzionale di una proteina nel contesto biologico di appartenenza.



Il termine proteomica nasce dall'esigenza di coniugare due grossi settori, **PROTEINE** e **genOMA**. Quindi la proteomica è lo studio sistematico di tutte le proteine che in un determinato momento e in determinate condizioni, vengono espresse da un genoma. (In un determinato momento e in determinate condizioni, perché l'istante successivo sarà già cambiato tutto). È un'analisi sistematica di un estratto cellulare, di tessuti, di microrganismi.

È stato possibile approcciarci allo studio proteomico dal momento in cui il sequenziamento del genoma umano è stato completato (nel 2001). La proteomica, abbiamo detto che è l'identificazione, dare un nome e un cognome alle proteine, questo è possibile proprio perché conosco la sequenza dell'intero genoma.

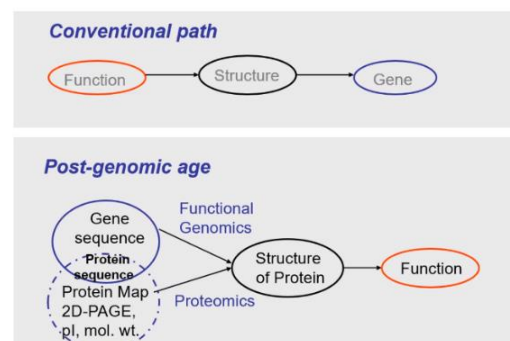
La proteomica è l'**identificazione** (nome e cognome), la **caratterizzazione** (struttura) e la **quantificazione** (quantità) di tutte le proteine coinvolte in un determinato pathway, in una determinata specie a seconda di quelle che sono le condizioni che in quel determinato momento caratterizzano quel sistema, organello, cellula, tessuto, organo.

Ma allo stesso tempo, può anche essere definita come l'analisi della genomica strutturale e funzionale attraverso lo studio sistematico delle funzioni, interazioni, localizzazione cellulare, espressione e modifiche post-traduzionali di proteine su scala high-throughput (ad alto rendimento). Ne va da sé che un approccio di questo tipo, necessita dei tools bioinformatici, cioè di software che mi vengono in soccorso, per darmi il cosiddetto nome e cognome delle proteine di cui ho bisogno.

L'approccio è cambiato nel corso degli anni, dall'approccio convenzionale, si è passati ad un approccio post-genomico.

Prima si partiva dalla funzione di una proteina, si passava alla struttura per arrivare al gene. Si individuava quella che poteva essere l'attività, ad esempio enzimatica, di una determinata proteina, si identificava la modalità con cui si poteva estrarre e

La rivoluzione copernicana dell'era post-genomica



purificare questa proteina, si cercava di arrivare alla sequenza e quindi struttura e da lì si arrivava ad identificarne la sequenza genetica.

Oggi, invece, a partire da quella che è la sequenza del gene, si arriva pian piano con tutti gli esperimenti, a determinarne la funzione. Ho una sequenza genetica, attraverso gli esperimenti arrivo a caratterizzare una proteina o più proteine, a seconda della necessità, e da lì vado a determinarne la funzione.

CONVENZIONALE

- ✓ **Identificazione di attività enzimatica**
- ✓ **Purificazione della singola proteina**
- ✓ **Caratterizzazione funzionale *in vitro***
- ✓ **Caratterizzazione strutturale *in vitro***
- ✓ **Sequenziamento e determinazione struttura primaria mediante EDMAN**
- ✓ **Identificazione del gene**
- ✓ **Predizione dell'intera sequenza amminoacidica a partire da quella genica**
- ✓ **Studio dell'espressione della proteina**

L'approccio convenzionale si basa sul fatto che conoscendo una determinata attività enzimatica (l'attività è sempre la stessa, la proteina avrà sempre la stessa attività nella sua vita, quello che spesso cambia e che ha consentito di creare questa branca è che può avere più funzioni. Questo è possibile perché può interagire con più proteine in diverse condizioni), si purifica la singola proteina, viene caratterizzata sia da un punto di vista strutturale che funzionale. Il sequenziamento veniva fatto mediante EDMAN (un processo che consente aminoacido per aminoacido di accrescere la sequenza peptidica, ovviamente non riesce a farla tutta, però grazie ad un pezzo di sequenza, usando i tools bioinformatici, avrò il nome della proteina). Si arriva all'identificazione del gene e poi è possibile fare delle predizioni.

Si parte dallo studio del proteoma, si identificano le proteine, si determina la struttura primaria mediante de novo sequencing e ricerche in banche dati. Una volta identificato il gene si passa alla caratterizzazione funzionale in vivo (Utilizzando organismi knock out, silenziamento genico, a seconda dei mezzi a disposizione vado a spegnere l'attività di quella proteina, se la spengo, il processo ipotizzato potrebbe venire a mancare). Si passa alla caratterizzazione strutturale in vivo per arrivare poi allo studio dell'espressione della proteina.

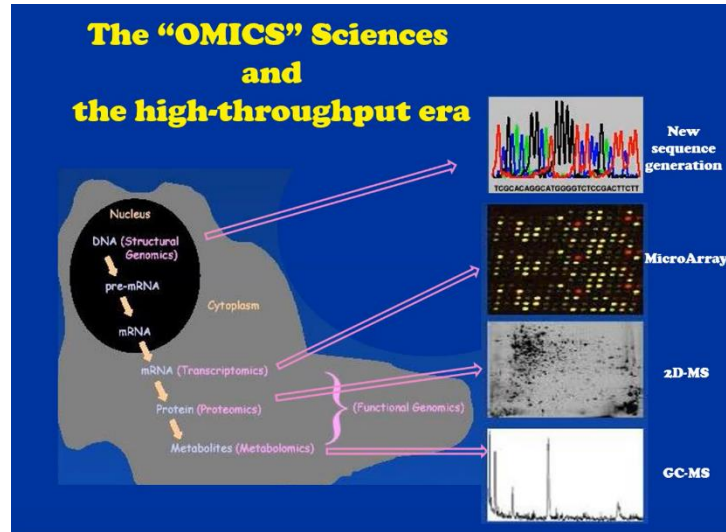
PROTEOMICO

- ✓ **Studio dell'intero proteoma**
- ✓ **Identificazione delle proteine significative**
- ✓ **Determinazione struttura primaria mediante de novo sequencing e ricerche in banche dati**
- ✓ **Identificazione del gene**
- ✓ **Caratterizzazione funzionale *in vivo* (organismi knock out, silenziamento genico, studio dei pathways cellulari ecc)**
- ✓ **Caratterizzazione strutturale *in vivo* (PTMs, NMR, interazioni, ecc)**
- ✓ **Studio dell'espressione della proteina**

L'approccio che si utilizza oggi è un approccio multidisciplinare.

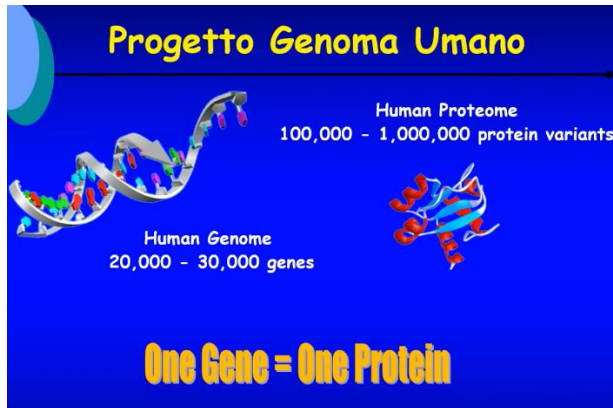
La proteomica è solo una branca, ma ha bisogno della trascrittomica, della metabolomica (voglio conoscere i metaboliti presenti in quel determinato contesto, lo si può fare attraverso un GC-SM, *gas cromatografia per piccole molecole*).

Singolarmente ognuno dà delle informazioni ma messi insieme completano il quadro.



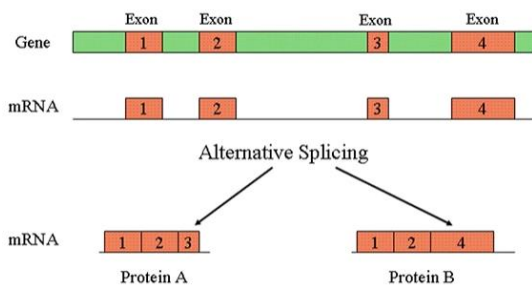
Relazione RNA-proteina

Proteoma e trascrittoma sono due entità temporalmente sfasate e difficilmente correlabili. Non c'è correlazione tra quantità di mRNA e quantità di proteina, questo perché ci sono degli eventi che hanno fatto sì che una certa quantità di mRNA non sia stato tradotto. A monte della traduzione esistono tutta una serie di meccanismi regolatori che fanno in modo che quel determinato mRNA possa o meno essere tradotto e quindi ci dà o no la proteina, possa essere tradotto non nei tempi e nei modi che ci si aspettano, basti pensare allo splicing alternativo.



Consideriamo una sequenza da tradurre, questa sequenza può essere tradotta per intero, non essere tradotta, essere tradotta solo in parte. Questo smentisce la definizione "un gene = una proteina". Il genoma umano contiene 25mila geni, dovrei avere 25mila proteine, invece esistono milioni/miliardi di varianti di proteine.

Splicing Variants



Esempio di splicing alternativo

Le due proteine tradotte, mostrate nello schema, sono proteine multidominio. La proteina A ha i domini 1, 2 e 3 e la proteina B 1, 2 e 4. Possiamo dire che le due proteine hanno in comune due domini che caratterizzano la loro attività, quindi le due proteine hanno la stessa attività biologica.

Ma il dominio 3 e 4 saranno dei domini che potranno consentire alle proteine l'interazione

con proteine diverse, questo farà in modo che queste due proteine con stessa attività biologica possano avere due funzioni completamente differenti, a seconda del contesto in cui si troveranno e a seconda dell'interazione che stabiliranno con altre proteine.

Come si fa a capire che questa proteina ha questa funzione? Da studi di letteratura, si identificano tutte le proteine con cui questa interagisce, quelle proteine verranno in qualche modo ricondotte ad un determinato processo, pathway cellulare, e in base a quella che è la funzione che posseggono quelle proteine partners, ipotizzo anche la funzione che può avere la proteina di interesse, in quel contesto.

Esempio:

l' α -fucosidasi, in base alla sequenza primaria doveva dare come prodotto finale un prodotto tronco al C-terminale. Ma in realtà si aveva contemporaneamente anche la proteina full length funzionante, questo perché era presente una sequenza nucleotidica (AAA AAU) che formava una sorta di ansa, la quale faceva in modo che il ribosoma saltasse la lettura del codone di stop, riuscendo a tradurre fino alla fine della sequenza, dando la proteina full length. Questo è un esempio di

Sequenza genica da *S. solfataricus* codificante per α -Fucosidasi

CAA TAT CAC CTA AAA AAT TCG GCC CAG TAA CCG ATT TCG GAT ATA AGG ATT TCA
 Gln Tyr His Leu Lys Asn Ser Ala Gln OCH
 Ile Ser Pro Lys Lys Phe Gly Pro Val Thr Asp Phe Gly Tyr Lys Asp Phe Ile

TAC CGA TGT TCA CTG GAG AGA ATT GGG ATC -----
 Pro Met Phe Thr Gly Glu Asn Trp Asp -----

Cobucci-Ponzano B. et al. (2003) JBC 278 14622-14631

Prodotto tronco non funzionale

Proteina full length biologicamente attiva

come, anche la sequenza primaria, non è detto che porti ad un'unica proteina.

Un altro motivo per cui ad un gene non corrisponde una sola proteina sono le **modifiche post-traduzionali** (metilazione, fosforilazione, acetilazione). Queste comportano una modifica anche a livello strutturale della proteina e fanno sempre in modo che la proteina possa svolgere funzioni diverse, perché una struttura diversa permette di interagire con proteine diverse. La modifica può avvenire all'interno di un sito specifico, non in maniera random.

La struttura primaria da sola non determina l'identità funzionale di una proteina.

Le caratteristiche principali di un proteoma sono:

- Complessità. Conoscere quello che viene espresso da un genoma, in un determinato contesto cellulare e momento, è diventato necessario per comprendere il perché si verificasse l'alterazione di un determinato processo. Inizialmente si pensava che questa alterazione fosse legata alla presenza o all'assenza di una molecola/proteina, e quindi di individuare in tale molecola/proteina il marcatore, ma poi nella realtà non è così, il proteoma è molto più complesso.
- Variabilità;
- Dinamicità.

