

La proteomica: la storia della cellula raccontata dai *biomarker*

Martina Collotta

Candiolo Cancer Institute - FPO, Torino

Nei numeri precedenti abbiamo percorso i primi due step del dogma centrale della biologia molecolare su scala omica, dal DNA, all'RNA, alle proteine, parlando di genomica e trascrittomica. In questo numero parleremo della proteomica, ovvero dello studio su larga scala delle proteine. Passeremo in rassegna la storia della proteomica e, facendo memoria di quanto avviene nei processi di sintesi e modificazione post-traduzionale delle proteine, analizzeremo i metodi che permettono lo studio del proteoma umano. Vedremo, infine, quali sono le applicazioni utili alla clinica, soffermandoci, in particolare, sui *biomarker* proteici.

Proteomica: definizione e breve storia

La *proteomica* consiste nello studio su larga scala delle proteine (Fig. 1). Così come è possibile studiare a livello "omico", globale, l'insieme dei geni della cellula e dei suoi trascritti, allo stesso modo è possibile studiare la totalità delle proteine da essa sintetizzate, sia per finalità strutturali che funzionali.

Il termine proteomica fu coniato nel 1997 in analogia con il termine genomica, di cui mantiene il suffisso. La storia della proteomica, in realtà, inizia qualche anno prima, nel 1994, quando Marc Wilkins, dottorando della Macquarie University, parlò per la prima volta di *proteoma*, ovvero dell'insieme delle proteine di una data cellula; per la precisione, di una data cellula, di uno *specifico* tessuto e in uno *specifico* organismo.

Questa puntualizzazione potrebbe sembrare superflua, ma è in realtà fondamentale. Facendo memoria di quanto detto parlando sia di epigenomica che di trascrittomica, è

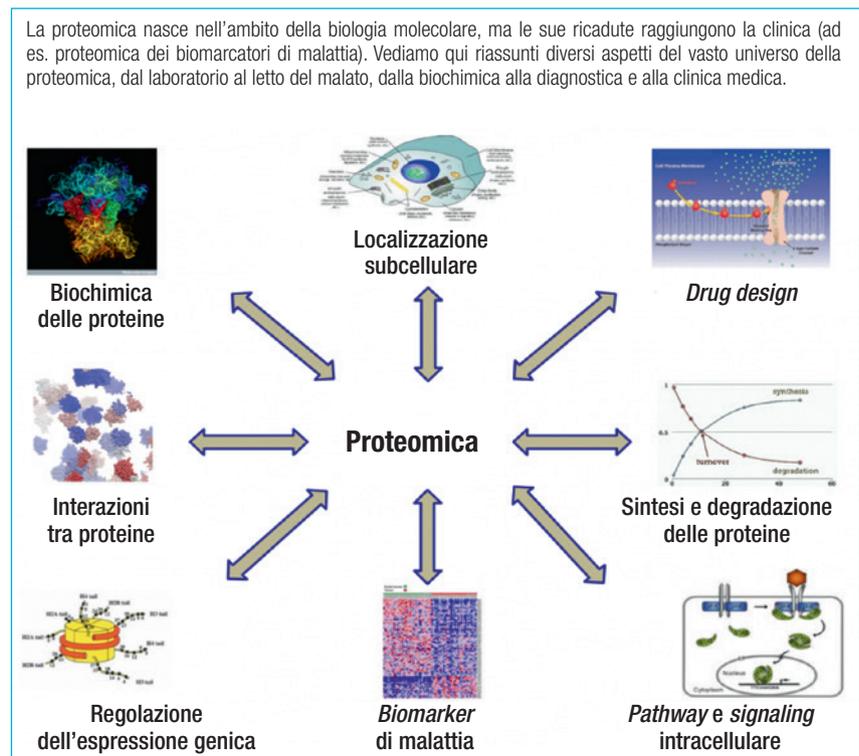
stata sottolineata più volte la differenza a livello di espressione genica tra i diversi tipi cellulari e, addirittura, dello stesso tipo cellulare in differenti momenti della vita dell'organismo. Questo, di conseguenza, significa sintesi di differenti proteine, a seconda delle diverse necessità della cellula, in risposta a stimoli interni ed esterni.

Quando parliamo di proteoma, inoltre, dovremmo precisare che con tale termine si intende l'insieme delle proteine cellulari non solo così come vengono prodotte, ma anche nelle loro *modifiche post-traduzionali*, importanti per il loro significato funzionale. Si stima che il *proteoma umano* contenga

20000-25000 proteine non ridondanti, ma se consideriamo gli eventi di *splicing* alternativo dei trascritti, la proteolisi e le modificazioni post-traduzionali, il numero salirebbe addirittura a qualche milione.

Gli studi di proteomica possono avere diversi obiettivi, al di là della "semplice" caratterizzazione di quali proteine vengono sintetizzate e in quale quantità. La proteomica si occupa anche di analizzare i movimenti di proteine tra i vari compartimenti intracellulari, il loro coinvolgimento nei *pathway* metabolici, le interazioni reciproche. Questo ultimo punto rientra in quella che viene chiamata *interattomica*.

FIGURA 1.
L'universo della proteomica.



Non bastava la trascrittomica?

Abbiamo accennato alla complicazione aggiunta che trascrittomica e proteomica presentano, rispetto alla genomica: il *genoma* di un organismo è lo stesso in tutte le cellule, mentre il *trascrittoma* e il *proteoma*, dipendenti dalla variazione dell'espressione genica, non lo sono.

Quali sono le proteine che sono espresse in tutte le cellule, indipendentemente dal tipo di tessuto e dalla fase del ciclo cellulare o dall'età dell'organismo? Qual è la costante? In passato, si è tentato di dare risposta a questa domanda attraverso l'analisi degli RNA messaggeri (mRNA), ma si è visto che i risultati non correlavano con il contenuto proteico. Sappiamo ora, infatti, che non tutti i mRNA, sebbene trascritti, vengono tradotti in proteine, e che la quantità di proteina sintetizzata a partire dalla traduzione di un mRNA, varia in base al trascritto stesso. Non ultime, la *degradazione proteica* e le *modificazioni post-traduzionali*, hanno un impatto quantitativo sul *pool* proteico intracellulare.

La trascrittomica, dunque, da sola non può confermare la presenza di una certa proteina e, ancor meno, è in grado di predirne la quantità prodotta. Solo la proteomica può rispondere a questa domanda fornendo misure dirette.

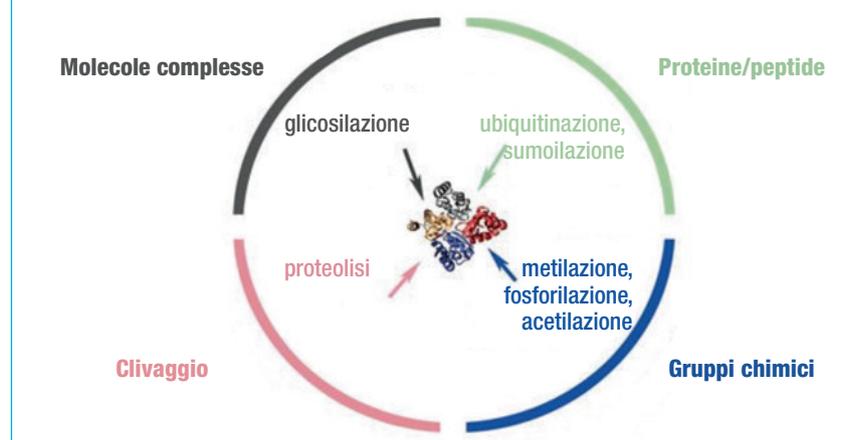
Le modificazioni post-traduzionali

Parlando di proteoma, abbiamo sottolineato come il termine faccia riferimento alle proteine della cellula, dopo i processi di *sintesi* e di *modifica*. Tali modifiche sono di tipo chimico e avvengono dopo il processo di traduzione dell'mRNA; esse vengono pertanto definite *modifiche post-traduzionali* (Fig. 2). La loro importanza risiede nel loro significato funzionale.

Ad esempio, la *fosforilazione*, forse la più nota di tali modifiche, è coinvolta nella funzionalità di molte classi di enzimi e nei processi di trasmissione del segnale (*signaling*), permettendo ad una proteina di diventare il target di legame di altre proteine in grado di riconoscerne il dominio fosforilato. Proprio per la sua importanza funzionale, la fosforilazione è stata oggetto di numerosi studi di proteomica (*fosfoproteomica*) volti ad individuare i *pathway* di *signaling* intracellulare

FIGURA 2.
Modificazioni post-traduzionali.

Vengono qui rappresentate, schematicamente, quattro grandi classi di modificazioni post-traduzionali delle proteine. Come visto, esse non solo hanno un significato funzionale, ma la loro identificazione potrebbe avere interessanti applicazioni nella diagnostica (es. glicosilazione delle proteine in cellule tumorali).



attivi in una data cellula in date condizioni. Altre modificazioni che possiamo ricordare sono l'*ubiquitinazione*, la *metilazione*, la *glicosilazione*, l'*ossidazione*, la *nitrosilazione* e molte altre. Sottolineiamo, inoltre, che una singola proteina può andare incontro a una o più di queste modificazioni, spesso secondo delle precise successioni e combinazioni di modifiche.

È interessante sottolineare come la presenza di una determinata tipologia di modificazioni post-traduzionali, possa rivestire un significato clinico. Nelle cellule tumorali, ad esempio, la velocità di glicosilazione delle proteine è aumentata, risultando in un maggior numero di peptidi che presentano tale modifica post-traduzionale, diventando pertanto possibili *biomarker*, con evidenti ricadute nella diagnostica oncologica.

Metodi per lo studio del proteoma

I due principali metodi utilizzati in proteomica, sono gli *immunoassay* (saggi con utilizzo di anticorpi) e la *spettrometria di massa*. Negli *immunoassay* vengono utilizzati anticorpi in grado di riconoscere specifiche proteine. Gli anticorpi vengono disegnati "su misura" e possono avere specificità per determinate modifiche post-traduzionali (es. anticorpi in grado di riconoscere proteine fosforilate). Sebbene i metodi basati sull'utilizzo di anticorpi siano molto comuni, essi, ovvia-

mente, non sono applicabili quando non esiste un anticorpo specifico, né sono in grado di determinare la sequenza di una proteina o di un peptide. Il *throughput*, che potremmo grossolanamente tradurre come "resa quantitativa", è inoltre inferiore negli *immunoassay* rispetto agli altri metodi. Il più utilizzato dei metodi non basati su anticorpi è la *spettrometria di massa*, utilizzata in combinazione con tecniche di separazione come la *cromatografia*, e che permette di identificare peptidi noti e non noti.

Esistono infine *tecniche ibride* in cui ad una prima fase di purificazione basata sull'utilizzo di anticorpi, segue un'analisi effettuata tramite spettrometria di massa che permette identificazione e quantificazione delle proteine. Attualmente, inoltre, la *proteomica comparativa*, similmente alla genomica comparativa, identifica proteine presenti in un dato organismo, sulla base della loro similarità con quelle presenti in organismi differenti.

Così come in trascrittomica, anche in proteomica si utilizzano dei *microarray* per identificare molte proteine in un solo esperimento. In questo caso, invece del filamento complementare, ancorato all'array troviamo l'anticorpo specifico per una data proteina. Se, invece, lo scopo è studiare l'interazione tra proteina e proteina, o tra proteina e DNA, l'array ospiterà differenti peptidi. Da sottolineare, tuttavia, che da un punto di vista esecutivo, lavorare con le proteine è più complesso rispetto a lavorare con gli oli-

gonucleotidi. Un'applicazione tecnicamente più fattibile dei *microarray* di proteine, sembra piuttosto essere quella di utilizzarli per immobilizzare il set di proteine cellulari in una determinata fase (di malattia), per poi confrontarla con un analogo set raccolto in un momento successivo e, da qui, condurre analisi (tecnica *reverse-phased protein microarrays*).

I dati sul proteoma richiedono, infine, un'elaborazione *bioinformatica* che permette, per esempio, di ipotizzare la struttura 3D della proteina, oppure di identificare il gene da cui è stato trascritto e poi tradotto il mRNA, individuando varianti di *splicing* e le successive modifiche post-traduzionali. L'analisi dell'espressione differenziale delle proteine, infine, richiede un supporto computazionale che solo la bioinformatica è in grado di dare.

Le applicazioni

La proteomica riveste una notevole importanza funzionale nell'identificazione di possibili *biomarker* a partire dall'analisi delle proteine presenti nei fluidi corporei, in alcuni casi attivamente secrete dalle cellule affette da una patologia (*secretoma*). Allo stesso modo permette di identificare antigeni batterici verso cui è rivolta la risposta immune e marker immunostochimici associabili a neoplasie.

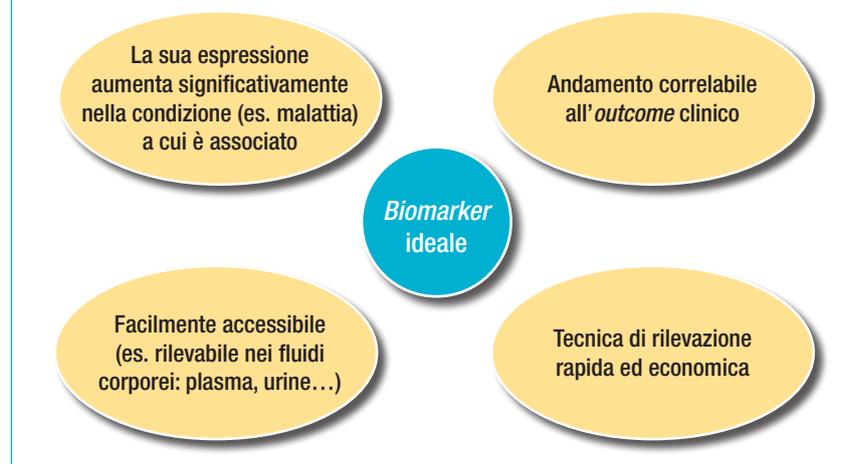
Le cellule tumorali, infatti, producono proteine modificate o proteine che, in condizioni normali, non sarebbero prodotte da quel tipo cellulare, in quel quantitativo e in quella determinata fase dello sviluppo e che, per tanto, possono diventare utili *biomarker* (Fig. 3).

Nell'illustrare i metodi per lo studio del proteoma, abbiamo parlato degli *immunoassay*. L'utilizzo di anticorpi risulta molto efficace quando le tecniche di proteomica vengono sfruttate ai fini della diagnosi precoce. I *biomarker* di malattia, infatti, possono essere presenti in quantità molto limitata, tale che solo la sensibilità dei saggi con anticorpi può essere in grado di rilevarla. Un ulteriore progresso in questa direzione, è avvenuto con l'utilizzo di tecniche digitali di *immunoassay*, accrescendo ulteriormente la sensibilità della tecnica.

L'interesse della ricerca può essere, inoltre, quello di valutare l'*espressione differenzia-*

FIGURA 3.
Il biomarker ideale.

Nello schema vediamo riassunte le caratteristiche di un *biomarker* ideale. È la proteomica, in molti casi, a permettere l'identificazione di possibili *biomarker* e a fornire gli strumenti per la loro analisi quantitativa.



le di determinate proteine in due tipi cellulari diversi o nello stesso tipo cellulare in differenti condizioni, non solo normale vs patologico, ma anche in differenti tappe dello sviluppo o in seguito a stimoli esterni o interni che alterano i *pathway* di *signaling* intracellulare.

Un'altra importante applicazione della proteomica è quella del *drug design*, ovvero della progettazione razionale di farmaci per il trattamento delle più svariate patologie. Lo studio delle proteine associate ad una malattia permette di farne dei target per nuovi farmaci. Questo potrebbe essere ulteriormente implementato da uno studio individuale del proteoma, per giungere allo sviluppo di farmaci personalizzati.

Proteomica del plasma umano

Abbiamo sottolineato l'importanza della proteomica nell'identificazione e nella misura di possibili *biomarker* di malattia. Dove misurare tali *biomarker*? Come trasformare le conquiste del laboratorio in un'applicazione alla portata della clinica?

Il *sangue* contiene informazioni circa lo stato fisiologico di tutti i tessuti e presenta il grande vantaggio, dal punto di vista della pratica clinica, di essere accessibile. Lo studio del proteoma presente nel *plasma*

umano è uno dei più interessanti ambiti di applicazione della proteomica, ma, allo stesso tempo, rappresenta una sfida. Non solo esso contiene ormoni, citochine, immunoglobuline, proteine della cascata coagulatoria o legate ai processi infettivi, ma contiene anche proteine rilasciate nel sangue dai più diversi tessuti dell'organismo attraversati dal torrente circolatorio. Uno degli ostacoli più grandi allo studio del proteoma del plasma umano, sembra essere quello della variabilità nella concentrazione delle diverse proteine (pensiamo all'abbondanza dell'albumina e alla bassissima concentrazione in cui sono invece presenti le citochine). Le proteine non differiscono solo quantitativamente, ma anche per emivita, avendo tempi di turnover differenti.

Bibliografia di riferimento

Liebler DC. *Introduction to proteomics: tools for the new biology*. Totowa, NJ: Humana Press 2002.

Naven T, Westermeier R. *Proteomics in practice: a laboratory manual of proteome analysis*. Weinheim: Wiley-VCH 2002.

Twyman RM. *Principles of proteomics (advanced text series)*. Oxford, UK: BIOS Scientific Publishers 2004.

Wilkins MR, Williams KL. *Proteome research: new frontiers in functional genomics (principles and practice)*. Berlin: Springer 1997.