



EDITORIAL

Proteómica: una aproximación integral

Proteomics: a comprehensive approach

A. Jara^{a,b,c,d} y J.J. Kopchick^{a,b,c,d,e,*}

^a Edison Biotechnology Institute, Konneker Research Laboratories, Ohio University, The Ridges, Athens, Ohio, Estados Unidos

^b Molecular and Cellular Biology Program, Ohio University, Athens, Ohio, Estados Unidos

^c Department of Biological Sciences, College of Arts and Sciences, Ohio University, Athens, Ohio, Estados Unidos

^d Heritage College of Osteopathic Medicine, Ohio University, Athens, Ohio, Estados Unidos

^e Department of Biomedical Science, Ohio University, Athens, Ohio, Estados Unidos

Disponible en Internet el 21 de diciembre de 2012

En las 2 últimas décadas la revolución «-ómica» ha impactado en las ciencias biológicas. Impulsada, en gran medida, por los esfuerzos colectivos resultantes en el esbozo inicial del genoma humano en 2001¹, una cantidad creciente de datos biológicos ha potenciado la evolución del campo de la bioinformática, dedicado al desarrollo de los elementos y algoritmos necesarios para almacenar y analizar grandes cantidades de datos. Una muestra del gran desarrollo de este campo (conocido como «Big Data») es la reciente iniciativa del gobierno federal de los Estados Unidos de Norteamérica de destinar más de 200 millones de dólares para la investigación y el desarrollo de las técnicas (*hardware* y *software*) e infraestructura necesarios para la utilización de estos datos. Parte de esta iniciativa está destinada a la elucidación de todas las proteínas e isoformas de las mismas en células, tejidos y órganos, tanto en estado de salud como en la enfermedad.

Hasta la fecha, se han identificado aproximadamente 22.000 genes (con transcripción por la polimerasa II). En comparación, cientos de miles de proteínas componen las células humanas, demostrando la gran discordancia existente entre el número de genes y el de proteínas. La metodología y los resultados de los estudios proteómicos constituyen un elemento investigador que ayudará a explicar esta discrepancia.

Al hablar de proteómica nos referimos a un conjunto de técnicas empleadas para explorar, simultáneamente, la

presencia y abundancia de miles de proteínas en una muestra biológica determinada. El objetivo fundamental de la proteómica es comprender tanto la composición (incluyendo modificaciones postranscripcionales) como la interacción de proteínas en una muestra; con la ventaja de poder interpretar el flujo de información, no solo en una vía, sino en múltiples redes complejas que interactúan en la célula. A diferencia del «estático» genoma, el proteoma varía tanto en el espacio como en el tiempo. Así, mientras 2 células de un mismo organismo tienen exactamente la misma secuencia de ADN, el conjunto de proteínas expresadas puede ser completamente diferente y depender, no solo del tipo celular, sino también del entorno bioquímico que rodea a la célula.

La proteómica encuentra una aplicación particularmente idónea en el descubrimiento de marcadores útiles para el diagnóstico, tratamiento y seguimiento de diversas entidades clínicas. Por biomarcadores entendemos cualquier cambio cuantificable en un sistema biológico, resultante de un estado de enfermedad (o de su progresión o remisión), tratamiento o intervención². El conocimiento de los cambios en el conjunto de proteínas como consecuencia de la enfermedad proporciona la identificación de aquellas, que podrían permitir la detección precoz o determinar la necesidad de intervención, prácticamente en cualquier proceso patológico. Sin embargo, para disponer de tal utilidad clínica, la muestra debería provenir de un fluido biológico fácilmente extraíble (sangre, orina, saliva) e, idealmente, la modificación proteica debería ser específica de un determinado proceso. En medicina clínica y, específicamente, en pediatría, la posibilidad de predecir la respuesta terapéutica optimizaría el tratamiento del paciente. Recientemente, se

* Autor para correspondencia.

Correo electrónico: kopchick@ohio.edu (J.J. Kopchick).

ha empleado una aproximación proteómica para la identificación de eventuales marcadores pronósticos en niños afectados de leucemia linfoblástica aguda (LLA) en muestras de médula ósea. La abundancia del antígeno de proliferación nuclear celular (PCNA) ha demostrado constituir un marcador específico para la estratificación del riesgo de los pacientes, independientemente del subtipo molecular de leucemia³.

El conjunto de técnicas proteómicas puede ser empleado también para resaltar la existencia de diferencias fisiológicas entre las muestras. En este número de *Anales de Pediatría*, Martos-Moreno et al. emplean una aproximación consistente en electroforesis bidimensional (2DE) seguida de análisis mediante espectrometría de masas secuencial (MS-MS) para examinar las diferencias existentes en el proteoma del tejido adiposo blanco subcutáneo (TAB-Sc) humano de acuerdo con su localización anatómica (6 regiones diferentes) y su disposición en profundidad⁴. Empleando muestras obtenidas mediante liposucción, demuestran cómo el proteoma del TAB-Sc varía, no solo entre las distintas áreas anatómicas, sino también en la misma localización, entre las capas superficial y profunda del TAB-Sc. Es interesante reseñar el hecho de que la abundancia de proteína fijadora de ácidos grasos número 4 fue significativamente inferior en la región dorsal que en el flanco, indicando que el transporte de ácidos grasos puede diferir drásticamente entre las distintas regiones del TAB-Sc, resaltando el hecho de que no todo el TAB-Sc del organismo es igual y postulando la posible existencia de distintas funciones para distintos «subdepósitos» dentro del mismo.

La electroforesis bidireccional se describió a finales de la década de los setenta como una base para el análisis proteómico. Debido al coste relativamente bajo y a la simplicidad del diseño, que obvia la necesidad de equipamiento especializado, esta continúa siendo una de las técnicas proteómicas más comúnmente empleadas. Se basa en un gel de poliacrilamida por medio del cual las proteínas son separadas en 2 dimensiones. En la primera dimensión, las proteínas se separan de acuerdo con su punto isoeléctrico (pI) en tiras finas de gel con gradiente de pH, empleando una técnica denominada iso-electro-enfoque. En ella, bajo la influencia de un campo eléctrico, las proteínas de la muestra vertidas en el gel, tras su tratamiento inicial, migrarán hacia el punto de pH donde su carga resulte neutra. Posteriormente, las tiras de gel son incubadas con un detergente iónico, colocadas perpendicularmente en un gel de poliacrilamida y sometidas a electroforesis, de modo que las proteínas son separadas de acuerdo con su peso molecular aparente (ratio masa-carga). Dicho gel es teñido, empleando distintos tintes, de entre los que destacan por su mayor rendimiento las tinciones fluorescentes (como el SYPRO), que permiten la detección de hasta 0,25 ng de proteína. El resultado de la tinción es una constelación de «puntos» proteicos o potenciales proteínas incluidas en el gel, que representan las proteínas contenidas en dicha muestra y cuya densidad óptica refleja la abundancia de las mismas. Ya mediante el empleo de técnicas bioinformáticas, los «puntos» proteicos de los distintos geles representativos de distintos grupos experimentales se emparejan y comparan. Aquellos «puntos» que resulten de interés pueden ser escindidos del gel manual o automáticamente y preparados para el ulterior

análisis mediante espectrometría de masas. Los beneficios de la 2DE son su bajo coste y accesibilidad y, quizás lo más importante, su capacidad para detectar modificaciones post y cotransduccionales (p. ej., la fosforilación), como diferencias en la localización de un «punto» proteico en el gel.

El paso subsiguiente a la 2DE para la identificación proteica es, habitualmente, la MS. Los «puntos» proteicos que muestran diferencias de intensidad significativas en la 2DE, entre los grupos que se deben comparar (p. ej., entre controles y pacientes), se escinden y someten a digestión con tripsina, concentrados y, posteriormente, reciben el estudio de MS mediante *matrix-assisted laser desorption ionization time of flight*. Este proceso consta de la disolución de la muestra en una matriz cristalina, que se estimula con un láser ultravioleta, lo que conduce a la estimulación de la muestra hacia la fase gaseosa. A continuación, la muestra ionizada se somete a un campo eléctrico en el analizador de masa, que determina la aceleración de la muestra hacia un detector. Cuando las muestras ionizadas alcanzan el detector, los resultados se registran como una ratio de masa/carga (m/z). Se pueden efectuar repeticiones sucesivas de MS en una técnica conocida como MS secuencial (tándem MS o MS-MS). Esta MS secuencial ofrece información detallada sobre la secuencia del péptido, que es compilada, constituyendo la técnica primaria para la génesis de la «huella peptídica» («*peptide fingerprint*»). Esta «huella» se compara con predicciones teóricas de masa basadas en las secuencias conocidas empleando diferentes programas electrónicos, como Swiss-Prot/TrEMBL, Sequest o Mascot.

A pesar de que la combinación de 2DE y MS constituye una técnica «consistente», no adolece de limitaciones que deben ser reconocidas. En primer lugar, las dificultades para la solubilización limitan su empleo en la separación de proteínas de membrana o extremadamente básicas o ácidas. Segundo, la visualización y la cuantificación de proteínas dependen del tinte seleccionado, lo que determina que las proteínas poco abundantes puedan no ser visualizadas aunque se encuentren presentes. Finalmente, no es posible separar proteínas extremadamente grandes o pequeñas. Como consecuencia, el investigador debe decidir los rangos de pI y PM en los que espera encontrar los cambios de interés en abundancia proteica.

Se han empleado con éxito técnicas proteómicas más modernas, como la cromatografía líquida (LC) de alta presión para la preparación de las muestras previa al análisis por MS. Actualmente, hay comercializadas múltiples plataformas automatizadas de LC/MS que permiten realizar la separación de proteínas y el análisis de MS de una forma automática. Esto ha conseguido la separación de miles de proteínas en cada muestra y es, a menudo, la metodología de elección cuando se desea producir un perfil proteómico de alto nivel en una muestra. La principal limitación de la LC/MS es la necesidad de llevar a cabo un fraccionamiento de la muestra previo al análisis mediante MS, ya que la gran cantidad de proteínas producida por la separación mediante LC puede saturar rápidamente la capacidad del detector del espectrómetro.

El análisis mediante MS también puede ser empleado directamente para explorar cuantitativamente el proteoma. Un ejemplo de ello es el empleo de «etiquetas» (*tags*) isotópicas para cuantificación absoluta y relativa (iTRAQ). La

técnica de iTRAQ requiere el «etiquetado» en posición N-terminal de muestras previamente digeridas. Cada grupo experimental es etiquetado con un trazador diferente, las muestras se mezclan, se separan mediante LC y se analizan mediante MS-MS. Los trazadores N-terminales se ionizarán y su abundancia relativa permitirá su cuantificación relativa entre los grupos de estudio. La desventaja fundamental de esta técnica es su alto coste, debido al elevado precio de las «tags» y a los espectrómetros de gran potencia necesarios para su análisis.

Independientemente de la aproximación proteómica empleada, los resultados deben ser validados antes de que cualquier potencial biomarcador se someta a mayor investigación. El modo más común de validar un marcador es el empleo de un ensayo de base inmunológica, como el enzimo-inmuno-ensayo (ELISA) o el *western-blot*. De cualquier modo, la negatividad de un inmunoensayo para una proteína determinada no establece que, necesariamente, los resultados de la MS sean incorrectos. Es posible que los epitopos que reconozca el anticuerpo empleado hayan sufrido una modificación postransduccional, de modo que este no pueda reconocerlos. En esta situación, el ensayo de confirmación puede ser negativo, mientras que los resultados de MS serían positivos. En general, los datos de MS de una proteína son los más sensibles para la asignación de identidad.

Tras la validación de un nuevo biomarcador, el siguiente paso es la génesis de un ensayo clínico que se empleará en distintos estudios retrospectivos desarrollados en muestras con diagnósticos previamente confirmados, para determinar la sensibilidad, la especificidad y el valor predictivo del biomarcador⁵. Posteriormente, este ensayo deberá ser aplicado en grandes series de estudios prospectivos para determinar su capacidad para la detección de la presencia y el estadio de la enfermedad en estudio, así como para compararlo con los ensayos previamente disponibles. Este camino desde la identificación del potencial biomarcador hasta la comercialización de un ensayo clínico es largo y costoso.

Debido a la complejidad del perfil proteómico celular, como consecuencia de las modificaciones postransduccionales, y a la adaptación al micro-medio ambiente circundante (como, por ejemplo, variaciones en el pH), las proteínas presentes en una célula están en un flujo constante. La capacidad para examinar simultáneamente los niveles de cientos de miles de proteínas ofrece una ventana única a través de la que examinar el funcionamiento intrínseco de la célula de una forma integrada y no como una colección

de vías independientes. En lo referente al descubrimiento de biomarcadores, lo ideal es identificar el mayor número de proteínas posible en cada experimento. Uno de los mayores retos en proteómica es sobreponerse a la relativa abundancia de proteínas en una muestra. Entre ellas, el suero, que es una de las muestras de elección en el diagnóstico de enfermedades, muestra un rango dinámico de proteínas con diferencias de hasta 10x en su concentración entre proteínas como la albúmina (mayoritaria) y hormonas circulantes en baja cantidad y de forma intermitente, como la hormona de crecimiento⁶.

Hasta la fecha, incluso los métodos de fraccionamiento más potentes, como la LC, requieren la separación de las proteínas en cientos de fragmentos para su completo análisis. El éxito futuro de la proteómica dependerá, al menos en parte, del desarrollo de procedimientos de fraccionamiento más potentes, fuentes iónicas más potentes para el análisis MS y estrategias bioinformáticas y bioestadísticas más consistentes para almacenar y analizar grandes cantidades de datos. Aun así, es aún más importante que la generación venidera de científicos y estudiantes esté preparada para la aproximación biológica, no solo desde una perspectiva molecular, sino desde una perspectiva integradora que aúne la bioinformática, la biología molecular y la ingeniería, de modo que se puedan desarrollar nuevas estrategias de aproximación que permitan entender mejor las redes metabólicas, así como el descubrimiento y el diseño de nuevas estrategias diagnósticas y terapéuticas.

Bibliografía

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.
2. Strimbu K, Tavel JA. What are biomarkers? *Curr Opin HIV AIDS*. 2010;5:463-6.
3. Jiang N, Kham SK, Koh GS, Suang Lim JY, Ariffin H, Chew FT, et al. Identification of prognostic protein biomarkers in childhood acute lymphoblastic leukemia (ALL). *J Proteomics*. 2011;74:843-57.
4. Martos-Moreno GA, Sackmann-Sala L, Berryman DE, Blome DW, Argente J, Kopchick JJ. El proteoma del tejido adiposo subcutáneo muestra heterogeneidad anatómica. *An Ped*. 2013;78:140-8.
5. Heckman-Stoddard BM. Oncology biomarkers: discovery, validation, and clinical use. *Semin Oncol Nurs*. 2012;28:93-8.
6. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1:845-67.