Integración del medio ambiente y la enfermedad en el análisis «ómico»

José M. Ordovás

Laboratorio de Nutrición y Genómica. Laboratorio de Metabolismo Lipídico. Centro de Investigación sobre Nutrición Humana Jean Mayer. Departamento de Agricultura de Estados Unidos. Universidad de Tufts. Boston. Massachusetts. Estados Unidos.

Las nuevas tecnologías englobadas en el concepto general de las «ómicas» suponen una revolución en la investigación biológica, tal como ya se ha hecho patente en el caso de la genómica. Sin embargo, los factores ambientales (tales como los estilos de vida, el ambiente socioeconómico, la dieta, etc.) no se han incorporado en su justa medida entre estas "ómicas", a pesar de la evidencia de que la mayor parte de las enfermedades comunes se deben a la interacción entre factores genéticos y ambientales. Sin embargo, para que su integración en las "ómicas" sea informativa, es crucial que las medidas ambientales tengan mayor precisión y fiabilidad y que estén sujetas a la validación apropiada. Además, dado el ingente volumen de datos generados por las "ómicas" y los factores ambientales, también es esencial el desarrollo de la bioinformática y la bioestadística.

Palabras clave: *Geonómica. Proteómica. Metabolómica. Interacción gen-dieta. Dieta. Fármacos.*

Integrating Environment and Disease Into «Omic» Analysis

New techniques grouped together under the general term of «omic» techniques are bringing about a revolution in biological research, as can be clearly seen in the case of genomics. However, environmental factors (such as lifestyle, socioeconomic context and diet) have not been incorporated to their full extent into these «omic» techniques despite evidence that many common diseases largely result from interactions between genetic and environmental factors. However, if the integration of environmental variables into «omic» analysis is to provide valuable information, it is essential that they are quantified in a highly precise and repeatable manner and that they are adequately validated. In addition, because of the enormous volume of information that will be generated by «omic» techniques and by taking environmental variables into account, it is also essential that bioinformatics and biostatistical methods are further developed.

Key words: *Genomics. Proteomics. Metabolomics. Gene-environment interaction. Diet. Drugs.*

INTRODUCCIÓN

La investigación biológica dedicada a elucidar los factores y mecanismos asociados y que causan enfermedades comunes crónicas-degenerativas se ha visto reforzada enormemente por la aparición de nuevas tecnologías que permiten por primera vez en la historia obtener una visión tanto general como detallada del funcionamiento de los organismos vivos. Estas tecnologías se agrupan dentro de lo que llamamos «ómicas» y se pueden clasificar en una serie de subgrupos principales que incluyen genómica, transcriptómica, proteómica y metabolómica (fig. 1). Con su aparición, también surgió la esperanza de que su uso diera lugar a rápidos descubrimientos que facilitarían de manera casi inmediata nuestra capacidad de prevenir, diagnosticar y desarrollar nuevas terapias contra las enfermedades más comunes, entre las que destacan las cardiovasculares[1,2]. Sin embargo, a pesar de los grandes progresos tecnológicos, la promesa de que las «ómicas» iban a producir rápidos avances en la medicina y la salud pública no se ha cumplido todavía[3], principalmente debido a las limitaciones conceptuales y experimentales que impiden considerar en su totalidad las complejas y dinámicas interacciones entre la plétora de factores involucrados en la iniciación, el progreso y la manifestación de la enfermedad.

El mayor impulso inicial para las «ómicas» vino del Proyecto Genoma Humano. Su valor intrínseco es indiscutible y demostró la posibilidad de lo que, en la mente de tantos, era imposible. Esta lección debe permanecer en nuestra memoria a la hora de confrontar las enormes dificultades que representará la conversión del progreso actual en soluciones prácticas para el futuro. Uno de los puntos más importantes que debemos considerar en nuestro pro-

Correspondencia: Dr. J.M. Ordovas.
Nutrition and Genomics Laboratory. JM-USDA-HNRCA at Tufts University.
711 Washington Street. Boston. MA 02111. Estados Unidos.
Correo electrónico: jose.ordovas@tufts.edu

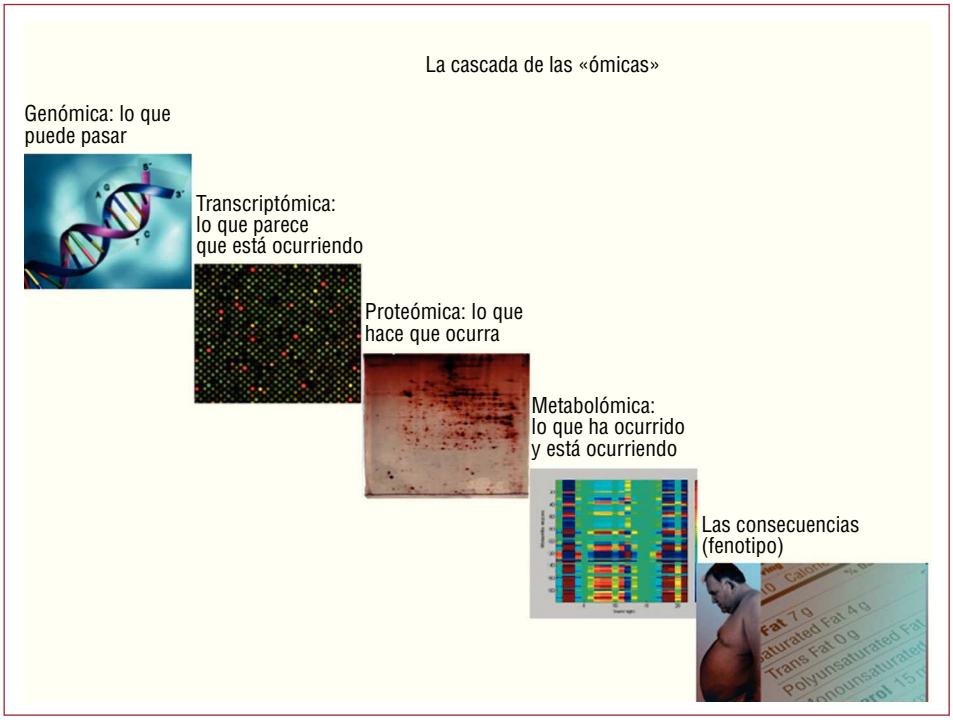

Fig. 1. Grupos de «ómicas» más importantes y descripción de la información biológica proporcionada por cada una de ellas.

greso es que las enfermedades más prevalentes —y por lo tanto nuestros objetivos más importantes y urgentes—, incluyendo en primera línea las enfermedades cardiovasculares, el cáncer y las neurodegenerativas, son extremadamente complejas y que, además de los factores genéticos, los no genéticos (englobados bajo la denominación ambigua de «ambientales») tienen un papel crucial en la etiología de estas enfermedades[4]. Además, es importante considerar que estos factores no son independientes, sino que actúan de una manera interactiva, lo que da lugar al área de investigación conocida como el estudio de las interacciones gen-ambiente. Por lo tanto, debemos reconsiderar la clasificación de las enfermedades como genéticas o ambientales y reconocer la simbiosis íntima de ambos extremos del espectro de causalidad.

Cuando nos referimos a la trama de interacciones entre genes y ambiente, tenemos que tener en cuenta siempre la correcta integración de genotipo, fenotipo y ambiente (fig. 2). A pesar del interés, la expectación y el *glamour* de la tecnología que ha rodeado al genoma, una de las mayores barreras en el progreso de estas investigaciones, y su inmediata aplicación a la mejora de salud, está en la correcta, precisa y estandarizada definición de los fenotipos[5]. Por otra parte, comparadas con la fenotipificación, la obtención de genotipos y la definición de alelos parecen un problema sencillo pues, en la mayor parte de los casos, es binario y sin incertidumbres. Sin embargo, hemos de tener en cuenta que, salvo raras excepciones, los polimorfismos que se han identificado asociados con enfermedad son meros marcadores genéticos y no causales de la enfermedad o de la predisposición a ella. La estrategia utilizada desde la década de los ochenta es la conocida como búsqueda de genes candidatos, que se podría traducir como buscar las llaves bajo la lámpara porque es donde podemos ver. Por muchos años, la limitación tecnológica sólo nos ha permitido buscar variantes genéticas en genes cuyos productos estaban ya involucrados en rutas metabólicas conocidas; por ejemplo, los componentes conocidos implicados en el metabolismo de las lipoproteínas plasmáticas. La incorporación de nuevas técnicas moleculares nos ha permitido con los años ir ampliando el número de marcadores genéticos en estos genes candidatos; de esta manera, hemos podido acceder a decenas, centenares de polimorfismos que podrían ser utilizados como marcadores de riesgo o susceptibilidad a enfermedad. De esta manera, hemos pasado del concepto ingenuo e ineficaz de estudiar polimorfismos uno a uno en nuestra búsqueda de marcadores al más realista, que consiste en el estudio de múltiples polimorfismos simultáneamente[6], incorporando también el concepto de haplotipo (o combinación de alelos polimórficos en una región genómica) al objeto de acercarnos más al descubrimiento de las mutaciones funcionales que finalmente originan las bases mecanísticas del riesgo asociado con una determinada región del genoma[7]. Probablemente la revolución reciente más notable en genotipificación viene en la forma de los chips genómicos, que permite en estos momentos la genotipificación rápida y

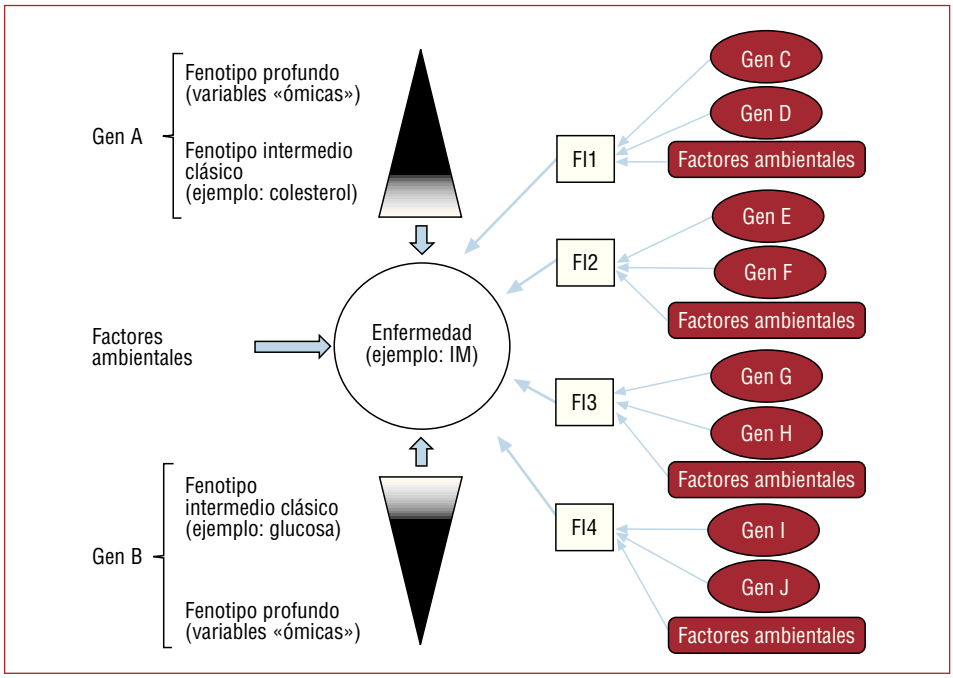

Fig. 2. Esquema de la complejidad implicada en el estudio de la genética de enfermedades comunes como el infarto de miocardio (IM). En la figura se presenta el gradiente de la magnitud de la asociación estadística (alta y probablemente significativa en el caso de los marcadores ómicos —negro en el triángulo de la figura—y baja, y probablemente no significativa —blanco en el triángulo— para el caso de los marcadores intermedios clásicos, colesterol y glucosa en este ejemplo). Esto resulta en una pérdida de predictividad de los marcadores genéticos cuando se consideran individualmente, pues están asociados con fenotipos metabólicamente lejanos del punto final (enfermedad o evento clínico). Esto además se complica por la contribución de factores ambientales y otros fenotipos intermedios (FI), que a su vez se afectan por un número elevado de genes y otros factores ambientales.

relativamente económica de hasta un millón de polimorfismos por individuo. Esta capacidad ha cambiado totalmente los esquemas de la investigación genética de enfermedades complejas y ha promovido la formación de grandes consorcios de investigadores que proporcionan la masa crítica necesaria para abordar la susodicha complejidad.

Con la incorporación de este conocimiento y esa tecnología parece pues al alcance de la mano la consecución de los objetivos indicados; sin embargo, cada vez que estamos a punto de conseguirlo, el listón parece elevarse y se dificulta el reto. Como ilustración de ello está la complejidad creciente que descubrimos en la variación genética. Tras identificar y catalogar millones de polimorfismos consistentes en un cambio único de nucleótido (conocidos como SNP), otro factor inesperado y de considerable importancia ha aparecido en el horizonte de la genética. Nos referimos a la variación en número de copias, que consisten en ganancias o pérdidas de hasta varios miles de kilobases de ADN que pueden incluir en algunos casos uno o varios genes y que, aunque se conocían desde hace años, no se reconoció hasta recientemente lo común de su presencia en el genoma humano[8,9]. Independientemente de los detalles, es evidente que necesitamos establecer claramente las reglas del juego por las que nos tenemos que regir para este tipo de estudios[10]. Estas reglas aplican al diseño de los estudios epidemiológicos, la selección de la población, marcadores genéticos y, muy importante, la generación de fenotipos, y por supuesto el tratamiento estadístico de los datos y las plataformas bioinformáticas necesarias para integrar todos los procesos y los resultados de una manera que, además de ser internamente coherente, pueda integrarse fácilmente a otros estudios externos. Además de genotipificaciones y fenotipificaciones comprehensivas y de buena calidad, no debemos olvidar la tercera columna en estos estudios: la exposición ambiental. Hasta ahora, los factores ambientales han tenido un papel nulo o muy pequeño en los grandes proyectos genómicos conocidos como Genome Wide Association Studies (GWAS), basados en los chips genómicos. Esta omisión actual limita la información resultante de esos esfuerzos, pero es una deficiencia temporal, ya que varios proyectos en marcha tienen como objetivo prioritario la incorporación de factores ambientales, principalmente tabaquismo, consumo de alcohol y dieta habitual, en el análisis de GWAS. Esta evolución es esencial, ya que la evidencia demuestra que los efectos de una gran número de variantes genéticas sólo se ponen de manifiesto cuando se incorpora información ambiental a los análisis.

Sin restar ninguna importancia a la información esencial contenida en el genoma y en su variación, la realidad es que los procesos biológicos no son estáticos, sino altamente dinámicos, y su estudio requiere el uso de las otras «ómicas»[11]. La más avanzada tras la genómica quizá sea la transcriptómica, que estudia la expresión de los genes basados en la cuantificación del ARN mensajero (ARNm). La transcriptómica provee información de qué genes están transcripcionalmente activos en términos cualitativos y cuantitativos. También revela si diferentes variantes genéticas dan lugar a cambios en la

expresión (regulación) génica, y esta información es crítica a la hora de asignar funcionalidad a los polimorfismos genéticos asociados con fenotipos y/o riesgo de enfermedad. El estudio de la transcriptómica se ha visto facilitado por plataformas relativamente sólidas, similares a los chips utilizados en la genómica. La interpretación de los resultados va haciéndose también cada vez más fiable como resultado de la incorporación de nuevas aplicaciones estadísticas. Sin embargo, ninguna «ómica» puede proveer individualmente la información suficiente que nos permita resolver las incógnitas biológicas, y en el caso de la transcriptómica es bien aparente porque comúnmente la cantidad de ARNm no se correlaciona con la de proteínas, que al fin y al cabo son las encargadas de gran parte de los procesos metabólicos. De ahí que necesitemos el complemento de la proteómica. El término proteoma fue usado por vez primera en 1995 para describir el conjunto de proteínas de un genoma, una célula o un tejido. La proteómica es el estudio a gran escala de los productos del genoma, con el fin de obtener una visión integral e integrada de los procesos celulares. El término proteómica se ha asociado tradicionalmente con la separación de un gran número de proteínas de una célula u organismo mediante 2D-PAGE (electroforesis bidimensional en gel de poliacrilamida). Las dificultades técnicas de la 2D-PAGE son tantas que han impedido la aplicación a gran escala de la proteómica. Más recientemente, la espectrometría de masas ha surgido como un método analítico alternativo, que reduce muchas de las limitaciones del análisis mediante 2D-PAGE. Como resultado, la proteómica está evolucionando más rápidamente y ha permitido su utilización para responder a diferentes preguntas solubles sólo a través de esta «ómica», entre otras los estudios de interacciones de proteínas, de modificaciones postraduccionales, el análisis funcional de proteínas y estudios de localización. También podemos distinguir la proteómica de expresión (estudio cuantitativo de la expresión de proteínas entre muestras que difieren en alguna variable), la proteómica estructural (estudio de la localización subcelular de las proteínas y de las interacciones proteína-proteína mediante la purificación de orgánulos o complejos y la posterior identificación de sus componentes mediante espectrometría de masas) y la proteómica funcional, para referirse a un estudio más dinámico. La incorporación de la proteómica a estudios nutricionales todavía es escasa, pero presenta grandes oportunidades[12].

La última de las grandes «ómicas» es la metabolómica, que tiene como objetivo el análisis, en un amplio espectro de muestras biológicas (suero, plasma, orina, tejidos, extractos celulares), de las pequeñas moléculas que son producto del metabolismo, como azúcares, grasas y otras moléculas no proteínicas. Esta información puede resultar fundamental para la comprensión de las rutas metabólicas, ya que habitualmente los compuestos medidos representan los productos finales del metabolismo. Las plataformas tecnológicas utilizadas hasta ahora incluyen la resonancia magnética (RM) y la espectrometría de masas (MS), cada una de ellas con sus ventajas y sus inconvenientes[13,14]. La espectroscopia de RM permite medir el contenido de protones en muestras biológicas complejas de moléculas de pequeño peso molecular. Tiene varias ventajas, entre ellas que las muestras prácticamente no necesitan preparación especial y que el tiempo de medición es muy breve, lo que permite medir cientos o miles de muestras. Además, la tecnología es relativamente fácil de estandarizar, lo que permite la comparación entre laboratorios e instrumentos[15,16]. En el caso de las técnicas basadas en MS, normalmente se utilizan en tándem con técnicas de separación tales como la cromatografía líquida o de gases. La ventaja de estas técnicas está en su sensibilidad; la desventaja, en la necesidad de procesar las muestras antes de su separación. Así, se ha utilizado la metabolómica basada en RM para estudiar los efectos de dietas con diferente contenido de productos de origen animal en las rutas metabólicas[17]. Este estudio demostró que un consumo de dietas ricas en proteína animal resulta en la aparición de altas concentraciones de creatinina, creatina, trimetilamina-N-óxido, taurina y metilhistidina 1 y 3; mientras que una dieta vegetariana se relaciona con elevadas concentraciones de p-hidroxifenilacetato y concentraciones disminuidas de N,N,N-trimetil-lisina. Así, este estudio, junto con otros de más reciente aparición, pone de manifiesto el potencial de la metabolómica para el estudio de cambios metabólicos modulados por modificaciones en la composición de la dieta.

El beneficio de combinar en un mismo análisis el poder analítico de las diferentes «ómicas» lo han puesto de manifiesto recientemente de manera muy precisa Gieger et al[18]. Sus resultados refuerzan el concepto de que polimorfismos genéticos comunes inducen cambios muy importantes en el metabolismo humano. Esto puede conducir a la personalización de las recomendaciones preventivas y terapéuticas basadas en la combinación de genómica y fenotipificación profunda.

Epigenética

Éste es un término propuesto por C.H. Waddington para referirse a los cambios reversibles en el ADN que hacen que unos genes se expresen o no dependiendo de condiciones exteriores. La información epigenética modula la expresión de los genes sin alterar la secuencia de ADN. Existen diversos

tipos de modulación epigenómica, entre ellos, la metilación del ADN es el más conocido y se sabe fuertemente regulado por determinados componentes de los alimentos. La modificación de histonas —como acetilación, metilación y fosforilación— es otro fenómeno epigenético muy estudiado; en cambio se conoce menos los mecanismos implicados en la denominada impronta genética (modificación epigenética del genoma que depende del origen materno o paterno del gameto transmisor). En los últimos años estos conceptos están siendo incorporados a los estudios nutricionales y se ha acuñado el término epigenómica nutricional, que de momento se ha aplicado casi exclusivamente a modelos animales. La compleja interacción entre factores ambientales y nutricionales se manifiesta también en los fenómenos epigenómicos, como se ha demostrado en ratas adultas descendientes de madres que mostraron un comportamiento más maternal durante la primera semana de vida posnatal. Estas ratas mostraron una respuesta más atenuada al estrés debido a la desmetilación de posiciones específicas en el gen del receptor glucocorticoideo[19] que afecta a la unión de factores de trascripción de este gen y, por lo tanto, a su expresión[20]. Por el contrario, factores nutricionales pueden afectar negativamente al comportamiento e interferir con el programa epigenético inducido por el cuidado maternal[21]. En humanos, se ha demostrado que la variación epigenética tiene relación con el envejecimiento, al demostrarse que la metilación en el ADN es casi idéntica en gemelos recién nacidos pero difiere marcadamente en gemelos de edad adulta, lo que indica la importancia del componente ambiental[22].

Genómica nutricional, nutrigenética y nutrigenómica

La genómica nutricional es una disciplina reciente y todavía con cierta crisis de identidad[23]. La genómica nutricional es el estudio conjunto de la nutrición y el genoma incluyendo todas las demás «ómicas» descritas previamente[24,25]. El primer término utilizado fue nutrigenética, acuñado en 1975 por el R.O. Brennan en su libro *Nutrigenetics: New Concepts for Relieving Hypoglycemia*[26]; en 1999 el término nutrigenómica fue utilizado por DellaPenna[27] para describir la disciplina científica dedicada a estudiar el genoma de las plantas con objeto de producir un contenido de micronutrientes optimizados para la protección de la salud humana. Actualmente, se considera que la nutrigenómica es la disciplina que estudia los mecanismos moleculares que explican la distinta respuesta fenotípica a la dieta en función del genotipo. La nutrigenómica, por lo tanto, se centra en estudiar cómo los nutrientes regulan la expresión de los genes, cómo afectan a los polimorfismos en la expresión y la regulación y cómo se interrelacionan estos cambios con aspectos transcriptómicos, proteómicos y metabolómicos[28]. Se han publicado cantidad de ejemplos que demuestran el potencial de esta disciplina para mejorar la prevención nutricional de las enfermedades cardiovasculares. La mayor parte de esos estudios han analizado las interacciones estadísticas entre un número muy reducido de polimorfismos y el consumo de un componente de la dieta como determinantes de fenotipos intermedios de enfermedad. Sin embargo, de momento la reproducibilidad ha sido muy escasa debido a las limitaciones de los diseños experimentales. Por lo tanto, es esencial integrar en los diseños la transcriptómica, la proteómica y la metabolómica, además de un panel más completo de polimorfismos genéticos en rutas metabólicas específicas, así como otros factores ambientales (consumo de tabaco, consumo de alcohol, actividad física y aspectos sociales).

CONCLUSIONES

Las tecnologías basadas en las «ómicas» serán esenciales para nuestro avance en el conocimiento de las bases moleculares de los procesos biológicos. Hasta ahora, la genómica es la que ha tenido un papel más relevante como resultado del Proyecto del Genoma Humano. Sin embargo, es importante resaltar que prácticamente todas las enfermedades comunes son el resultado de complejas interacciones entre genes y factores ambientales, y es esencial que éstos se integren en los estudios genéticos y en los que utilicen el espectro más amplio de las otras «ómicas». Sin embargo, las medidas ambientales requieren mediciones más precisas y estandarizadas para que puedan rendir todo su valor. Asimismo, se va a necesitar el desarrollo de modelos matemáticos más adecuados a la complejidad y la cantidad de medidas resultantes de las ómicas, lo que requerirá avances urgentes en la informática biomédica.

BIBLIOGRAFÍA

1. Hocquette JF. Where are we in genomics? J Physiol Pharmacol. 2005;56 Suppl 3:37-70.
2. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. Nat Genet. 2003;33 Suppl:311-23.
3. Bilello JA. The agony and ecstasy of "OMIC" technologies in drug development. Curr Mol Med. 2005;5:39-52.
4. Willett WC. Related articles, links abstract balancing lifestyle and genomics research for disease prevention. Science. 2002;296:695-8.
5. Schulze TG, McMahon FJ. Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. Hum Hered. 2004;58:131-8.
6. Suh Y, Vijg J. SNP discovery in associating genetic variation with human disease phenotypes. Mutat Res. 2005;573:41-53.

7. Crawford DC, Carlson CS, Rieder MJ, Carrington DP, Yi Q, Smith JD, et al. Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. Am J Hum Genet. 2004;74:610-22.
8. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36:949-51.
9. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science. 2004;305:525-8.
10. Thomas DC, Haile RW, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. Am J Hum Genet. 2005;77:337-45.
11. Kussmann M, Rezzi S, Daniel H. Profiling techniques in nutrition and health research. Curr Opin Biotechnol. 2008;19:83-99.
12. Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J Natl Cancer Inst. 2005;97:307-9.
13. Robertson DG. Metabonomics in toxicology: a review. Toxicol Sci. 2005;85:809-22.
14. Lenz EM, Wilson ID. Analytical strategies in metabonomics. J Proteome Res. 2007;6:443-58.
15. Keun HC, Ebbels TM, Antti H, Bollard ME, Beckonert O, Schlotterbeck G, et al. Analytical reproducibility in (1)H NMR-based metabonomic urinalysis. Chem Res Toxicol. 2002;15:1380-6.
16. Dumas ME, Maibaum EC, Teague C, Ueshima H, Zhou B, Lindon JC, et al. Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. Anal Chem. 2006;78:2199-208.
17. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, et al. Susceptibility of human metabolic phenotypes to dietary modulation. J Proteome Res. 2006;5:2780-8.
18. Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, Meitinger T, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet. 2008;4:e1000282.
19. Liu D, Diorio J, Tannenbaum B, et al. Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. Science. 1997;277:1659-62.
20. Weaver ICG, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, et al. Epigenetic programming by maternal behavior. Nat Neurosci. 2007;7:847-54.
21. Szyf M, Weaver I, Meaney M. Maternal care, the epigenome and phenotypic differences in behavior. Reprod Toxicol. 2007;24:9-19.
22. Kahn A, Fraga MF. Epigenetics and aging: status, challenges, and needs for the future. J Gerontol A Biol Sci Med Sci. 2009;64:195-8.
23. Ordovas JM, Corella D. Nutritional genomics. Annu Rev Genomics Hum Genet. 2004;5:71-118.
24. Corella D, Ordovas JM. Integration of environment and disease into 'omics' analysis. Curr Opin Mol Ther. 2005;7:569-76.
25. Gallou-Kabani C, Vige A, Gross MS, Junien C. Nutri-epigenomics: lifelong remodelling of our epigenomes by nutritional and metabolic factors and beyond. Clin Chem Lab Med. 2007;45:321-7.
26. Brennan RO. Nutrigenetics: new concepts for relieving hypoglycemia. New York: Evans & Co; 1975.
27. DellaPenna D. Nutritional genomics: manipulating plant micronutrients to improve human health. Science. 1999;285:375-9.
28. Muller M, Kersten S. Nutrigenomics: goals and strategies. Nat Rev Genet. 2003;4:315-22.