



Editorial

Enfermedades complejas y análisis genéticos por el método GWAS. Ventajas y limitaciones

Genome-wide association studies (GWAS) in complex diseases: Advantages and limitations

José A. Riancho

Departamento de Medicina Interna, Hospital Universitario Marqués de Valdecilla-IFIMAV, Universidad de Cantabria, Santander, España

Las enfermedades hereditarias clásicas se deben a la mutación de un gen y a menudo tienen graves consecuencias para el organismo, pero afortunadamente son poco frecuentes. Frente a ellas se solían situar las enfermedades adquiridas, debidas a factores ambientales. Sin embargo, buena parte de las enfermedades más prevalentes son en realidad el resultado de la combinación de factores hereditarios y ambientales. Muchos trastornos frecuentes, como la osteoporosis, la artrosis, la diabetes o la hipertensión arterial tienden a agruparse en familias, lo que refleja su componente hereditario (aunque pueden existir también factores ambientales compartidos). La importancia de los factores hereditarios en estos trastornos es grande y así, por ejemplo, se ha comprobado que la herencia explica entre el 50 y el 80% de la variabilidad de la masa ósea¹. Por otro lado, los estudios de epidemiología genética mostraron que, a diferencia de las enfermedades hereditarias clásicas, en estos trastornos el riesgo no se explica por la alteración de un solo gen. De ahí la denominación de enfermedades «poligénicas» o «complejas».

La secuenciación del genoma humano permitió comprobar que existen muchas variaciones interindividuales en el ADN. Contrariamente a las mutaciones, esas variaciones son bastante frecuentes y en muchos casos su repercusión funcional es limitada o nula (de hecho, la mayoría están en regiones no codificantes), por lo que se llamaron «polimorfismos». Entre ellos destacan los polimorfismos de un solo nucleótido (*single nucleotide polymorphisms* o SNP), que consisten simplemente en el cambio de una base por otra. Son muy abundantes; hay unos 15 millones en el genoma. Son también frecuentes los polimorfismos de repetición, que consisten en grupos de unos pocos nucleótidos que se repiten un número variable de veces en los individuos. Más recientemente se ha identificado otra forma de variación del ADN, llamada «variación en el número de copias». Consiste en regiones relativamente grandes del genoma, de miles de nucleótidos, que en algunos individuos se encuentran repetidas, en el mismo o en diferentes cromosomas.

El hecho de que las enfermedades complejas no siguieran un patrón clásico de herencia, es decir, que no pudieran explicarse por alteraciones de un solo gen, junto al descubrimiento de

la frecuencia de los polimorfismos, llevó a plantear la hipótesis de las «enfermedades comunes-variantes comunes». Esta asume que las enfermedades complejas comunes son la consecuencia del efecto combinado de muchos polimorfismos frecuentes en la población.

Tras los estudios epidemiológicos que demostraron el importante componente hereditario de estos trastornos, los investigadores se lanzaron a intentar identificar cuáles eran los genes y polimorfismos implicados. Para ello, basándose en los conocimientos de la biología y la etiopatogenia de las enfermedades, identificaron «genes candidatos» y exploraron su asociación con esos trastornos. Así, por ejemplo, dado el importante papel que tienen la vitamina D y las hormonas sexuales en el metabolismo óseo, algunos de los primeros estudios de genes candidatos analizaron la asociación de los polimorfismos de los receptores de vitamina D y de estrógenos con la osteoporosis^{2,3}.

Aunque su diseño admite múltiples variaciones, en esencia en estos estudios se trata de elegir un gen candidato, identificar algunos de sus polimorfismos y analizar si los alelos de esos loci polimórficos se asocian con un rasgo fenotípico determinado o con la frecuencia de una enfermedad. Posteriormente, se han realizado estudios sobre otros muchos genes candidatos en relación con una gran variedad de trastornos, pero globalmente no han respondido a las expectativas generadas por las atractivas hipótesis que los sustentaban. Los resultados obtenidos por algunos investigadores a menudo no se replicaron en otros estudios y, en todo caso, la fuerza de la asociación entre el genotipo y el fenotipo ha sido en general pequeña⁴.

El desarrollo posterior de los *microarrays* hizo posible analizar centenares de miles de SNP de una manera eficiente, con una pequeña muestra de ADN y un coste mucho menor que el que supondría estudiar los SNP de manera individual. Además, al estar esos SNP distribuidos por todos los cromosomas, era posible una aproximación nueva: explorar todo el genoma sin hipótesis previa, es decir, sin necesidad de elegir previamente los genes candidatos. Por otro lado, los SNP incluidos en los *microarrays* se seleccionaron considerando los patrones de ligamiento, de manera que capturaban también otros polimorfismos que no se tipificaban directamente. Estos estudios de asociación genómica (*genome-wide association studies* o GWAS) levantaron

Correo electrónico: rianchoj@Unican.es

grandes expectativas. Se pensó que por fin iban a permitir identificar todos los genes involucrados en la heredabilidad de las enfermedades complejas. De hecho, desde el año 2005 se han publicado GWAS con unas 1.200 asociaciones de SNP con más de 200 enfermedades o rasgos fenotípicos (se puede encontrar un listado en www.genome.gov/GWASStudies).

Sin embargo, los resultados de los GWAS tampoco han respondido a las expectativas iniciales. Por un lado, muchos de los genes que se han encontrado asociados a una determinada enfermedad no tienen un efecto biológico conocido que explique esa relación. No obstante, quizás eso no sea sino reflejo de lo incompleto de nuestros conocimientos. De hecho, esos hallazgos están sirviendo para identificar nuevos mecanismos patogénicos y nuevas dianas terapéuticas⁵. Más sorprendente es que, incluso cuando se combinan todos los GWAS disponibles acerca de un trastorno determinado, los polimorfismos asociados suelen explicar menos del 5-10% del riesgo de enfermedad.

¿Cómo es posible que estudios en los que se analizan 500.000 SNP en varios miles de individuos arrojen esos pobres resultados en cuanto a predicción de riesgo? Uno de los motivos puede ser la potencia de los estudios. Al explorar tantos SNP existe un riesgo elevado de encontrar falsas asociaciones significativas según los criterios convencionales de $p < 0,05$. De hecho, si se analizan 500.000 SNP cabe esperar encontrar unos 25.000 asociados con la enfermedad con un valor $p < 0,05$ simplemente por azar. Para evitar ese error, se hace una corrección de comparaciones múltiples, de manera que sólo se consideran significativas las asociaciones con p menor de 10^{-7} o 10^{-8} . Este criterio reduce los falsos positivos, pero al mismo tiempo disminuye notablemente la potencia para detectar SNP asociados verdaderamente con la enfermedad⁶. Para aumentar la potencia no cabe sino aumentar el tamaño muestral, con estudios más grandes o, lo que es cada vez más frecuente, realizando metaanálisis de varios estudios. De todas maneras, se estima que con los GWAS que se han llevado a cabo presumiblemente se han identificado ya más del 80-90% de los SNP frecuentes que se asocian a trastornos prevalentes como la osteoporosis con una *odds ratio* superior a 1,1-1,2. Por tanto, se piensa que si se amplían los estudios posiblemente se encuentre que otros SNP están también asociados con la enfermedad, pero la influencia individual de cada uno de ellos presumiblemente será muy pequeña^{7,8}.

Entonces, ¿qué determina esa gran parte del riesgo de enfermedad que todavía permanece sin explicar? La respuesta a esta pregunta es incierta. Una posibilidad es que las enfermedades prevalentes no sean el resultado de variantes comunes con efectos individuales relativamente pequeños (como asume la hipótesis de enfermedades comunes-variantes comunes), sino de variantes infrecuentes con efectos relativamente grandes, no identificadas en los GWAS. Hay que tener en cuenta que, en virtud de su diseño, los *microarrays* utilizados en los GWAS no permiten detectar la influencia de SNP con alelos raros, que tienen frecuencias poblacionales inferiores al 1-10%. Es también posible que las diferencias en el riesgo se deban a la interacción entre diferentes SNP, o entre SNP y factores ambientales. Ni los GWAS individuales, ni los metaanálisis efectuados hasta ahora tienen potencia suficiente

para desentrañar esas interacciones. Tampoco puede descartarse que otras formas de variación genética, como los polimorfismos de repetición o las variaciones en el número de copias, que apenas han sido exploradas, desempeñen un papel importante. Posiblemente los mecanismos epigenéticos (potencialmente heredables y capaces de modular la expresión de los genes sin implicar cambios en la secuencia del ADN) influyen también de manera importante en el riesgo de enfermedad. Es también posible que exista una alta heterogeneidad genética en la patogenia de estas enfermedades. Ciertamente, la combinación de varios estudios mediante metaanálisis aumenta la potencia para detectar factores genéticos comunes a todas las poblaciones estudiadas. Pero esa estrategia no es necesariamente eficaz si se combinan estudios con grupos de individuos en los que los genes influyentes son diferentes. Por ejemplo, los polimorfismos de la aromatasas, enzima que convierte los precursores androgénicos en estrógenos en los tejidos periféricos, se asocian a la masa ósea en las mujeres posmenopáusicas, pero no en las jóvenes con una producción activa de estrógenos ováricos⁹. Evidentemente, esa asociación puede quedar enmascarada en un metaanálisis en el que se mezclen mujeres pre y posmenopáusicas.

En definitiva, los GWAS representan un enorme avance tecnológico que ha permitido identificar nuevos genes asociados con distintas enfermedades. Ello ofrece interesantes posibilidades de cara al desarrollo de nuevos tratamientos, pero hasta ahora los resultados han sido decepcionantes en cuanto a predecir el riesgo global de enfermedad. Así pues, los investigadores en este campo tienen por delante el descubrir en qué consiste esa «materia oscura» aun desconocida que explica la heredabilidad de las enfermedades complejas.

Conflicto de intereses

El autor declara no tener ningún conflicto de intereses.

Bibliografía

1. Ralston SH. Osteoporosis as an hereditary disease. *Clin Rev Bone Miner Metab.* 2010;8:68-76.
2. Morrison NA, Qi JC, Tokita A, Kelly PA, Crofts L, Nguyen TV, et al. Prediction of bone density from vitamin D receptor alleles. *Nature.* 1994;367:284-7.
3. Kobayashi N, Fujino T, Shirogane T, Furuta I, Kobamatsu Y, Yaegashi M, et al. Estrogen receptor alpha polymorphism as a genetic marker for bone loss, vertebral fractures and susceptibility to estrogen. *Maturitas.* 2002;41:193-201.
4. Richards JB, Kavvoura FK, Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, et al. Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. *Ann Intern Med.* 2009;151:528-37.
5. Bossini-castillo L, Martin JE, Diaz-Gallo LM, Rueda B, Martin J. Genética de la esclerodermia. *Reumatol Clin.* 2010;6:12-5.
6. Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, et al. Insights into the genetic architecture of osteoarthritis from stage 1 of the arcO-GEN study. *Ann Rheum Dis.* 2011;70:864-7.
7. Duncan EL, Brown MA. Genetic determinants of bone density and fracture risk-state of the art and future directions. *J Clin Endocrinol Metab.* 2010;95:2576-87.
8. Goldstein DB. Common genetic variation and human traits. *N Engl J Med.* 2009;360:1696-8.
9. Zarrabeitia MT, Hernandez JL, Valero C, Zarrabeitia AL, Garcia-Unzueta MT, Amado JA, et al. A common polymorphism in the 5'-untranslated region of the aromatase gene influences bone mass and fracture risk. *Eur J Endocrinol.* 2004;150:699-704.