

Guía práctica a los estudios de asociación genética. Consideraciones sobre su utilidad clínica

F. Rodríguez Esparragón¹, J.C. Rodríguez Pérez^{1,3}, M.A. García Bello¹

¹ Unidad de Investigación y ² Servicio de Nefrología. Hospital Universitario de Gran Canaria Dr. Negrín.

³ Universidad de Las Palmas de Gran Canaria

Nefrología 2009;29(6):582-588.

INTRODUCCIÓN

La publicación de una serie de artículos en el número de enero de la revista *JAMA* bajo el título «Cómo usar un artículo de asociación genética»¹⁻³ constituye un punto de partida excelente para intentar un doble propósito: presentar, a similitud de lo publicado, una guía práctica con los requisitos necesarios para enfrentarse a un artículo de asociación genética y, por otro lado, mostrar las herramientas necesarias para efectuar un estudio de estas características. Para nuestro propósito reduciremos el ámbito de actuación a aquellos trabajos de asociación genética de base poblacional que se realizan a través del reclutamiento de casos y de controles considerando que evaluaremos genes candidatos. Los métodos y la interpretación de los resultados de los estudios genéticos con base familiar son distintos y no serán objetivo de esta revisión. Insistimos en nuestra intención de señalar, únicamente, una serie de pautas de aplicación práctica y no la de pretender la realización de una aproximación a la epidemiología genética. Señalaremos algunas consideraciones sobre la relevancia clínica de estos estudios y analizaremos la situación actual y su supervivencia frente a los estudios de asociación a genoma completo.

¿ESTÁN CORRECTAMENTE SELECCIONADOS LOS PACIENTES? ADECUACIÓN DEL FENOTIPO

La caracterización adecuada del fenotipo asociado con una determinada enfermedad ha de realizarse de acuerdo con aquellos criterios clínicos sobre los cuales haya un consenso médico-científico claramente establecido. La mayor parte de las Sociedades científicas establecen estos criterios y, por supuesto, sus correspondientes actua-

lizaciones, a medida que aumenta el conocimiento sobre el desarrollo y la evolución de la enfermedad. En la realidad de la práctica clínica no siempre es posible establecer un fenotipo certero aunque se sigan rigurosamente estos criterios.

¿ES ADECUADO EL TAMAÑO DE LA MUESTRA?

El cálculo del tamaño de la muestra (tabla 1) en un estudio de casos y controles que incluya información genética es objetivo de estudio y modelización constante⁴. Una aproximación común al cálculo del tamaño de la muestra en un estudio de asociación genética no difiere de la que se realiza en un estudio clínico habitual tipo expuestos/no expuestos y se basa en establecer previamente la magnitud de la diferencia que debe detectarse. En nuestro caso en particular, se traduciría en establecer *a priori* la diferencia entre la frecuencia alélica o genotípica entre nuestras poblaciones. Asimismo, deberemos conocer la frecuencia de los alelos (todos los que se quieran considerar) en la población control, el valor del error de tipo I, es decir, el error que se comete cuando se rechaza la hipótesis nula siendo verdadera y el valor del error de tipo II, que es el error que se comete cuando se acepta la hipótesis nula siendo falsa. Habitualmente se trabaja con una seguridad de que no cometamos un error del tipo I del 95% (probabilidad, $\alpha = 0,05$) y del tipo II comprendida entre un 5 y un 20%, si bien, en general, se establece sobre un 20% (probabilidad, $\beta = 0,2$). Así nos aseguramos un poder estadístico ($1 - \beta$) del 80%. Hay otros aspectos que también debemos considerar, como la tasa previsible de errores y el tipo de errores esperables en nuestro procedimiento de genotipificación, que deben compensarse con un mayor tamaño de la muestra a fin de no ver disminuido nuestro poder estadístico. Algunas herramientas «on-line» nos ayudan realizar estos cálculos. Entre otras: <http://linkage.rockefeller.edu/pawe/> y <http://hydra.usc.edu/GxE/>.

Correspondencia: José Carlos Rodríguez Pérez
Unidad de Investigación y Servicio de Nefrología.
Hospital Universitario de Gran Canaria Dr. Negrín.
Universidad de Las Palmas de Gran Canaria.
jrodper@gobiernodecanarias.org

Tabla 1. Variables que influyen en el cálculo del tamaño de la muestra

Error tipo I	Cuando corregimos por múltiples comparaciones, necesitamos calcular el α corregido (α_c) que vamos a utilizar como criterio. Si hay muchas comparaciones (estudios de asociación genómica amplia), el α_c es tan pequeño que necesitamos una muestra muy elevada si queremos tener potencia estadística
Error tipo II	Un valor de 0,20 suele ser el estándar. Un valor de 0,50 supone que los estudios de replicación equivalen a realizar un «lanzamiento de moneda al aire», por lo que debemos incrementar el tamaño de la muestra
Tamaño de efecto esperado	Si se espera una asociación débil, se exige una muestra amplia para localizar la asociación y permitir la replicabilidad del estudio
Errores en genotipado	La tasa de errores y el tipo de error obligan a incrementar la muestra. Si existen errores y no se han tenido en cuenta, realmente no estamos contando con la potencia declarada
Definición de casos excesivamente amplia	Se espera menor fuerza de la asociación, por lo que necesitamos una muestra mayor
Número de variables en nuestro modelo	En acercamientos multivariantes, si no queremos construir un modelo sobreajustado de difícil replicabilidad, se requieren al menos 10 sujetos por variable candidata a entrar en el modelo final
Uso de covariables	Sustraen varianza irrelevante, fortaleciendo la relación entre variables de interés. Equivale a trabajar con mayor tamaño de efecto, por lo que podemos ahorrar en unidades de muestra
El tipo de corrección empleada	Si se usa el False Discovery Rate necesitamos menor tamaño de muestra que si hacemos uso de la corrección de Bonferroni
El test estadístico empleado	Cochran-Armitage es más potente que Chi-cuadrado cuando el modelo está correctamente especificado, por lo que necesita menor tamaño de muestra

CONSIDERACIONES SOBRE LOS CASOS

De forma ideal se han de reclutar casos incidentes. Conociendo previamente la frecuencia alélica de las variantes que van a estudiarse y antes de proceder a la fase de reclutamiento y conociendo la dificultad inherente a un fenotipo certero, podría intentarse aumentar el número de sujetos «casos» que deben estudiarse a través de distintas alternativas: en primer lugar, incrementando su número siendo menos estrictos en la definición de nuestro fenotipo, por ejemplo, «enfermedad renal», con lo que se aumenta la heterogeneidad y disminuye la certeza de causalidad de los alelos que se evalúan; o bien, por otro lado, podríamos ser más rigurosos en los criterios de selección de fenotipos incrementando la homogeneidad pero reduciendo, necesariamente, el número de casos reclutados. De manera alternativa, puede aumentarse la duración de la fase de reclutamiento. La elección entre una u otra alternativa depende de varios factores, de forma principal de la frecuencia alélica de la variación o de las variaciones en el/los gen(es) candidato que debe estudiarse. Si optamos por utilizar un criterio amplio en la selección de casos, resulta esperable que los genes candidatos estén presentes en un subcon-

junto inferior de nuestros casos a estudiar; por tanto, debido al menor tamaño de efecto esperado perderemos al menos parte del poder estadístico que esperábamos alcanzar al hacer una definición más amplia de nuestro fenotipo. Por tanto, a la hora de definir los casos debemos ponderar si el incremento de casos disponibles compensa la pérdida del poder estadístico que se deriva de una menor diferencia de frecuencias esperadas. Si la definición del fenotipo la posponemos hasta el momento en que realizamos los análisis de nuestros resultados disponemos de alternativas. Algunos autores, como Chen y Lee, han creado un método cuantitativo y sencillo que nos permite establecer y sistematizar, en función de las frecuencias alélicas y de la existencia de al menos dos posibles tipos de «casos» cuando podemos o no incrementar el tamaño de la muestra de casos uniendo ambos tipos de casos⁵. No obstante, científicamente lo más adecuado es anticipar estas cuestiones en la fase de diseño del estudio y no cuando procedamos a realizar los análisis estadísticos, ya que podríamos correr el riesgo de sobreajustar los datos y obtener resultados espurios de difícil replicabilidad.

nefrología basada en la evidencia

CONSIDERACIONES SOBRE LOS CONTROLES

Arya Sharma y Xavier Jeunemaitre⁶, reputados autores en el área, señalan una dificultad y un error común en la selección y reclutamiento de la población control. La dificultad deviene de la propia naturaleza de la población de controles. Mientras que para la mayoría de los investigadores hospitalarios la inclusión de pacientes no ha supuesto un problema, sí lo ha sido reclutar controles, ya que requieren de una base poblacional que necesita de recursos específicos. Dado que un control es potencialmente un caso, un sesgo común es incluir a controles de, por ejemplo, bancos de sangre o de trabajadores sanos del medio. La ventaja que teóricamente pudiera suponer la selección de estos candidatos definidos como «hipernormales» ya que, en teoría, redundaría en una mayor diferencia en las frecuencias alélicas entre población afectada y la de controles es, en realidad, mínima y pudiera enmascarar otros fenotipos de selección positiva hacia la supervivencia⁷⁻⁹. De igual forma, la selección de controles que sean casos no diagnosticados supone una reducción del poder estadístico en nuestro estudio¹⁰.

SOBRE EL ANÁLISIS ESTADÍSTICO Y LA ASOCIACIÓN ESTADÍSTICA

Podemos establecer una selección del fenotipo que debe estudiarse caracterizándolo como un rasgo dicotómico: diabéticos frente a no diabéticos, hipertensos frente a normotensos o intentar acotar más la variabilidad reduciendo la incertidumbre ambiental y ganando en influencia del genotipo mediante la evaluación de los denominados fenotipos intermedios, es decir, evaluando uno o varios caracteres mensurables que vinculen a través de rutas biológicas plausibles nuestro(s) gen(es) candidato(s) con la enfermedad. La aproximación estadística más convencional a una u otra alternativa implica la realización de modelos multivariantes.

En estudios con múltiples comparaciones debemos utilizar correcciones que nos permitan no incrementar nuestra probabilidad de cometer error tipo I (tabla 2). El procedimiento de Bonferroni consiste en dividir el alfa por el número de comparaciones a estimar: si existen aproximadamente 10^6 varian-

tes en el genoma, el umbral de valor de p corregida para todas las comparaciones sería $p = 0,05/10^6 = 5 \times 10^{-8}$. Esto nos obliga a necesitar un tamaño de muestra muy elevado si pretendemos detectar diferencias moderadas. Por ello, se utilizan fórmulas de corrección menos conservadoras, como el False Discovery Rate. No obstante, también con este procedimiento se necesita, en caso de hacer numerosas comparaciones, incrementar de manera prohibitiva el tamaño de nuestra muestra.

El análisis estadístico utilizado puede otorgarnos mayor poder estadístico (tabla 3). Así, si conocemos el modelo genético (aditivo-recesivo-dominante), debemos utilizar el test de Cochran-Armitage. No obstante, en general no conocemos el modelo genético de nuestros genes candidatos, y aunque nos ofrezca mayor poder, también es cierto que resulta menos robusto que la tradicional prueba de la Chi-cuadrado de Pearson, por lo que en caso de incumplimiento de los supuestos que establezcamos sobre el modelo genético se invalidarían los resultados.

ANÁLISIS DE MEZCLAS POBLACIONALES

Derivado de la teoría evolutiva actual surge la necesidad en medicina de evaluar la heterogeneidad genética de la población en estudio. Los estudios realizados muestran que la variación en la frecuencia de los alelos asociados con la enfermedad entre poblaciones dispares aunque existe es, en realidad, bastante pequeña⁷. Pudiera ocurrir que la asociación dependiera por completo de la exposición a cierto determinante ambiental cuya frecuencia variase según la localización geográfica y que, por tanto, la frecuencia de estos alelos por selección variase⁷. Cuando se realizan mezclas poblacionales que difieren en frecuencias alélicas por razones genéticas o ambientales la asociación pudiera resultar espuria. Por tanto, se hace necesaria la selección, genotipificación y análisis de marcadores neutrales (alelos nulos, SNP no ligados, inserciones/deleciones) a través de dos estrategias diferentes denominadas «genomic control» y «structured association»: <http://pritch.bsd.uchicago.edu/structure.html> y http://wpicr.wpic.pitt.edu/WPICCompGen/genomic_control/genomic_control.htm

Tabla 2. Alternativas para afrontar la multiplicidad

Stepdown	Supone realizar los análisis en, al menos, dos fases. Cuando tenemos varios grupos, en primer lugar, comparamos si existe alguna diferencia entre ellos y, posteriormente, intentamos identificar entre cuáles de ellos hay diferencias
Bonferroni	En estudios de asociación de casos y controles con muchos genes candidatos se utiliza la corrección de Bonferroni en función del número de comparaciones entre variantes alélicas que vamos a realizar Si tenemos múltiples comparaciones entre variantes alélicas para muchas variables dependientes conviene agruparlas en familias y realizar para cada una de ellas la corrección de Bonferroni
False Discovery Rate	Útil en estudios exploratorios en los que tenemos muchas comparaciones y podemos permitirnos una tasa reducida de falsos positivos (q)

Tabla 3. Tipo de análisis estadísticos empleados

Análisis de contingencia	Variable independiente: estudio de asociación por medio de Cochran-Armitage o Chi-cuadrado de Pearson
Regresión logística	Permiten evaluar la <i>odds ratio</i> del riesgo de enfermedad de una variante alélica frente a otras, ajustando para un conjunto de covariables Con múltiples variables candidatas, se exige explicitar cuál es el mecanismo de selección empleado (algoritmo automático o selección de variables por el investigador según criterios teóricos)
Comparación de medias: técnicas no paramétricas	Comparación sobre fenotipo medido cuantitativamente Otorgan más potencia y el error tipo 1 está mejor ajustado en los casos en que se incumplen los supuestos de las técnicas paramétricas
Comparación de medias: test de la t y ANOVA	Comparación sobre fenotipo medido cuantitativamente. Si hay más de dos grupos, debemos usar ANOVA
Comparación de medias: MANOVA	Comparación de fenotipos medidos cuantitativamente. Se utiliza cuando tenemos múltiples variables dependientes y queremos agruparlas para no acumular error tipo I
Comparación de medias: ANCOVA y MANCOVA	Tratamos de reducir varianza irrelevante para averiguar cuál es la aportación ajustada de nuestra/s variable/s de interés
Interpretación de los coeficientes	
Coefficientes de regresión logística	Cuanto aumenta el riesgo por incrementar en una unidad nuestra puntuación en esa variable, manteniendo el resto de las puntuaciones constantes Si es mayor que 1, implica que se incrementa el riesgo
Diferencia de medias ajustadas	Diferencia en la medición del fenotipo entre dos grupos, una vez se controla el efecto del resto de las variables del modelo
Uso de intervalos de confianza	Indica en qué intervalo estimamos que está el verdadero valor poblacional de los parámetros de interés En estudios de asociación genómica amplia el intervalo de confianza es muy amplio (debido a la corrección de Bonferroni), por lo que se necesita incrementar el tamaño de la muestra Nos permite evaluar la precisión de nuestras estimaciones

ELECCIÓN DE GENES CANDIDATOS

Puesto que en nuestra aproximación hemos decidido evaluar polimorfismos génicos en genes candidatos, nos interesa, en primer lugar, determinar el número y los genes que analizaremos. Tradicionalmente los genes candidatos evaluados se han escogido en función del conocimiento disponible sobre la actividad del producto génico en la enfermedad que debe estudiarse, del conocimiento sobre la función de la proteína codificada por el gen, del derivado de los estudios en modelos animales, del que puede presuponerse a través del fenotipo asociado con las formas monogénicas de la enfermedad, del derivado de estudios de ligamiento génico y de los datos disponibles *a priori*, así como del obtenido a través de metaanálisis⁸. Si además las variantes que deben analizarse se sitúan en regiones de interés en los genes, esto se traduce, por ejemplo, en un cambio de aminoácido (variantes no sinónimas), o bien afectan a la estabilidad o al procesamiento del mensajero, o se encuentra en regiones reguladoras del gen, la variante será probablemente, de mayor utilidad¹¹. Una estrategia alternativa en la selección de genes candidatos es la selección de variantes «tags», es decir, variantes de las que se dispone de información previa, obtenida de estudios de ligamiento, y que a su vez pudieran ligarse a alelos de susceptibilidad¹².

En un área de trabajo enormemente activa como es la bioinformática, sorprende el escaso número de herramientas que ayudan en la labor clave de la selección de genes candida-

tos¹³. Existen, no obstante, una serie de programas *on-line*: <http://omicspace.riken.jp/PosMed/>, http://www.geneniffer.org/index/index_frameset.htm y <http://www.genetics.med.ed.ac.uk/suspects/>. Estos programas integran la información proveniente de estas plataformas de análisis genético de alto rendimiento conjuntamente con la información derivada de los estudios de expresión permitiendo la aparición del término convergencia a la selecciones de regiones y genes candidatos.

FENOTIPO CUANTIFICABLE Y DEMOSTRACIÓN EXPERIMENTAL DEL FENOTIPO CUANTIFICABLE

Identificar y medir el mayor número de parámetros biológicos implicados de manera directa con el gen y su producto o con la ruta biológica en la cual aquel está implicado aumenta sustancialmente la capacidad de información del estudio. Aumenta la rigurosidad si, además, se demuestra de forma experimental que la propia variable se asocia con otras variables en regiones clave del gen o que afecta funcionalmente al gen o a la proteína⁶.

¿TIENE SENTIDO EL ESTUDIO Y ANÁLISIS DE UN NÚMERO RELATIVAMENTE PEQUEÑO DE POLIMORFISMOS?

Un análisis de un solo polimorfismo puede llevar a asociaciones espurias, entre otros motivos, porque la variante pue-

nefrología basada en la evidencia

de encontrarse en desequilibrio de ligamiento con otra u otras variantes, constituyendo un haplotipo característico. Sharma, et al⁶ consideran que para un gen candidato, la selección, genotipificación y análisis de frecuencia de, al menos, tres polimorfismos comunes permite la identificación de variantes en desequilibrio de ligamiento y la identificación de sinergias. En un estudio de casos y controles de base poblacional en el que se hayan genotipificado SNP próximos y dado que por definición se desconoce la fase, pueden inferirse haplotipos mediante el uso de *software* genéticos: (GDA: <http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php> y Arlequín: <http://lgb.unige.ch/arlequin/>), pero no determinarse en su totalidad ni asociarse con un fenotipo mensurable en un sujeto dado (excepto para los genotipificados como homocigóticos para todos los SNP evaluados en un locus). A día de hoy, la información sobre haplotipos comienza a estar disponible en red y por lo tanto resulta de interés elegir y validar la utilidad pronóstica o terapéutica de variantes funcionales.

Las asociaciones positivas encontradas en una determinada población utilizando una o unas pocas variantes suelen no replicarse en otras poblaciones. La ausencia de replicación constituye un argumento fundamental entre los autores más críticos con los estudios de asociación. Al margen de los problemas derivados de errores de genotipificación, mezclas poblacionales, elección de genes candidatos, caracterización inadecuada de los casos y/o de los controles, reclutamiento inadecuado de los casos y/o de los controles, diferencias en la exposición ambiental, la falta de poder estadístico subyace en la mayor parte de los estudios que muestran ausencia de asociación (tabla 4). La solución no es sencilla, puesto que el reclutamiento de un mayor número de casos y de controles tampoco lo es.

Un estudio de asociación genética retrospectivo de casos y controles debe cumplir por tanto una serie de requisitos sucintamente presentados en los párrafos anteriores.

GENOTIPIFICACIÓN A GRAN ESCALA Y ESTUDIOS DE ASOCIACIÓN GENÓMICA AMPLIA O DE GENOMA COMPLETO

El Proyecto Internacional HapMap (<http://www.hapmap.org/index.html.en>) se define como un esfuerzo de

varios países para identificar y catalogar las similitudes y diferencias genéticas de los seres humanos a este nivel¹⁴. Como comentamos al inicio, la metodología y la interpretación de los resultados obtenidos difiere de los análisis de genes candidatos en individuos no relacionados. Las plataformas de análisis genéticos de alto rendimiento ha cambiado el panorama de los estudios de asociación permitiendo genotipificar múltiples polimorfismos aunque necesariamente ampliando el tamaño de la muestra necesaria. La fusión de las tecnologías de genotipificación a gran escala con la información disponible en el proyecto internacional HapMap ha permitido la posibilidad de establecer estudios de asociación genética utilizando la información denominada como de asociación genómica amplia o de genoma completo (WGA). En estos estudios el HapMap provee la información de los denominados «tag SNPs» definidos como el conjunto mínimo de SNP necesarios para detectar un haplotipo. El proyecto ha ido por fases. En la primera se reclutaron tríos (madre, padre e hijos) a partir de los cuales se identificaron SNP localizados a una distancia de 5 kb y con una frecuencia superior al 5%. Caracterizaron entonces su estructura en haplotipos para definir los «tags». En una segunda fase, la identificación de «tags» asociados con una determinada enfermedad permite inferir la estructura del haplotipo disminuyendo la necesidad de genotipificar todas las variantes y permitiendo además localizar genes candidatos próximos. Este tipo de estudios, aunque cada vez más asequibles, requiere a día de hoy de importantes recursos tanto humanos como financieros, pero suponen también un paso importante hacia la caracterización de variables clínicamente relevantes. Sin embargo, la tecnología asociada a la genotipificación no está exenta de problemas¹⁵. Incluso cuando se realiza un esfuerzo de esta magnitud son pocas las variantes estadísticamente asociadas a la enfermedad que sobreviven al proceso de replicado¹⁶. Surge, además, el problema de las correcciones cuando se realizan múltiples comparaciones estadísticas^{9,10}.

UTILIDAD CLÍNICA

El entusiasmo inicialmente suscitado por los estudios de asociación genética se basó en la facilidad con la que permitían ir un paso más allá sobre la aproximación epidemiológica convencional en el conocimiento de la causalidad de la en-

Tabla 4. Poder estadístico y replicabilidad de un estudio

Potencia <i>a priori</i> , ¿para qué tamaño de efecto se ha calculado?	Si se ha elegido potencia de 0,80 para tamaños de efecto grandes, la potencia es menor para efectos medianos y pequeños. Por tanto, cuidado con aceptar la H_0 de ausencia de asociación
Potencia <i>a posteriori</i>	Indica en qué medida esperamos replicabilidad del efecto a igual tamaño de la muestra. Se trata de un índice sesgado positivamente y que no suele aportarse en el informe de resultados
¿Exceso de potencia?	Si tenemos una potencia superior a 0,80, resulta necesario calcular índice de fuerza de asociación (<i>odds ratio</i> , índice de contingencia, etc.)
	La p no es un índice de asociación, ya que depende del tamaño de muestra utilizado en el estudio

fermedad y/o en la de los factores de riesgo asociados a la enfermedad (figura 1). Para uno o varios genes candidatos, la mayoría de los centros hospitalarios podían, por ejemplo, realizar ampliaciones por PCR y digestión enzimática de una serie de polimorfismos en genes de interés una vez reclutados sus casos y controles. Sin embargo, esta información única quedaba sesgada. Esta relativa facilidad permitió un crecimiento exponencial del número de trabajos publicados y, concomitantemente, la aparición de un sector crítico a su utilidad real. Algunos autores señalan además que, puesto que la acción o efecto de una determinada variante dentro de un gen debe interpretarse en el contexto de una red compleja que incluya además de interacciones con otras variantes y con el medio ambiente, la propia complejidad de la ruta biológica en la cual el gen está inmerso, debería plantearse inicialmente la validez de una estrategia de asociación⁷. En muchos de estos estudios inicialmente realizados, la cuestión crítica fundamental la constituyó la ausencia de reproducibilidad en otras series y poblaciones. Sin embargo, cabe señalar que el denominador común que justifica en muchos casos la falta de reproducibilidad no depende tanto de la población analizada sino, como hemos comentado, de la ausencia de poder esta-

dístico, constituyéndose éste en el error principal. Puesto que la tecnología emergente permite además aumentar de forma exponencial el número de variaciones a analizar es reclutar de manera adecuada la población un requisito fundamental y un problema mayor. En nuestro ámbito, la Ley de investigación biomédica 14/2007, de 3 de julio, regula el tipo de estudios genéticos que pueden realizarse, la estructura de los consentimientos informados necesarios, el proceso de anonimización de las muestras y su almacenamiento, utilización y cesión futura. Por tanto, asociado a los problemas sucintamente presentados en los párrafos anteriores: inadecuada caracterización de la población en estudio, falta de una adecuada evaluación de la población, reclutamiento inadecuado de los casos y/o de los controles, tamaño de muestra insuficiente, falta de replicación de las asociaciones analizadas, se genera un cierto escepticismo que puede, sin embargo, rebatirse con, por ejemplo, una caracterización de fenotipos adecuados, análisis de fenotipos intermedios, evaluación de fenotipos mensurables, caracterizando haplotipos, análisis de la estructura poblacional, de forma que los estudios de asociación genética preserven su calidad como una de las herramientas de aproximación práctica más poderosas que existen⁹.

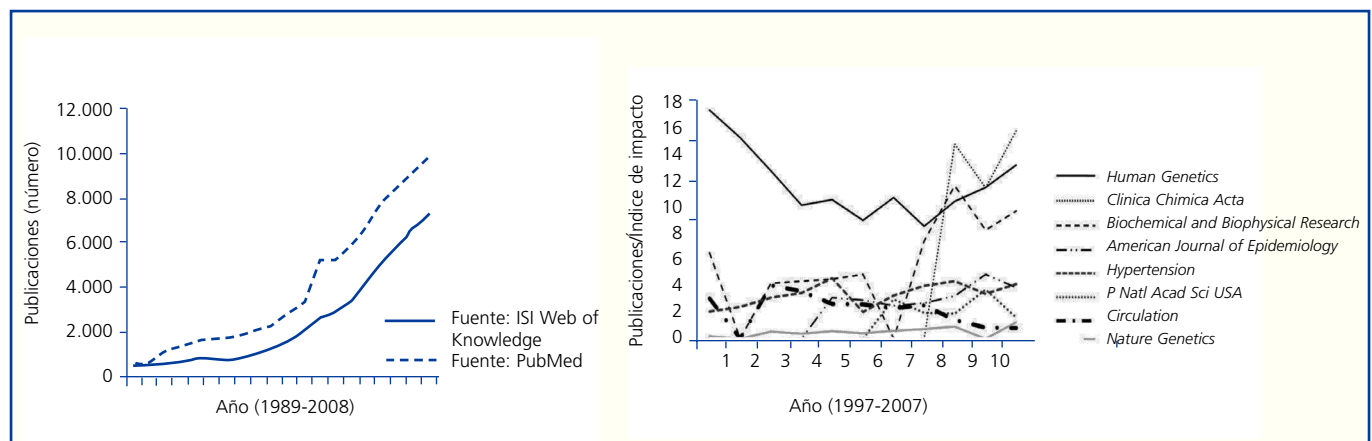


Figura 1. Incremento exponencial en el número de artículos publicados en los años 1989-2008 en las bases de datos PubMed del National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed> e ISI Web of knowledge <http://www.isiknowledge.com/> tras buscar por «polymorphisms».

BIBLIOGRAFÍA

1. Attia J, Ioannidis JP, Thakkinian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: A: Background concepts. *JAMA* 2009;301:74-81.
2. Attia J, Ioannidis JP, Thakkinian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: B: Are the results of the study valid? *JAMA* 2009;301:191-7.
3. Attia J, Ioannidis JP, Thakkinian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: C: What are the results and will they help me in caring for my patients? *JAMA* 2009;301:304-8.
4. Zou G, Zuo Y. On the sample size requirement in genetic association tests when the proportion of false positives is controlled. *Genetics* 2006;172:687-91.
5. Chen CF, Lee WC. Case recruitment in genetic association studies: larger sample size or greater homogeneity? *Int J Epidemiol* 2005;34:711.

nefrología basada en la evidencia

6. Sharma AM, Jeunemaitre X. The future of genetic association studies in hypertension: improving the signal-to-noise ratio. *J Hypertens* 2000;18:811-14.
7. Colhoun HM, McKeigue PM, Davey SG. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361:865-72.
8. Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet* 2005;366:1315-23.
9. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516-7.
10. Silverman EK, Palmer LJ. Case-control association studies for the genetics of complex respiratory diseases. *Am J Respir Cell Mol Biol* 2000; 22:645-8.
11. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000;405:847-56.
12. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.
13. Thornblad TA, Elliott KS, Jowett J, Visscher PM. Prioritization of positional candidate genes using multiple web-based software tools. *Twin Res Hum Genet* 2007;10:861-70.
14. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299-1320.
15. Fu W, Wang Y, Wang Y, Li R, Lin R, Jin L. Missing call bias in high-throughput genotyping. *BMC Genomics* 2009;10:106.
16. Khoury MJ, Little J, Gwinn M, Ioannidis JP. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *Int J Epidemiol* 2007;36:439-45.