

## Lectura crítica de estudios de asociación genética en enfermedades complejas

Jorge Sapunar Z.<sup>1</sup>

### Critical reading of genetic association studies in complex diseases

#### Introducción

Una enfermedad es compleja cuando su patogenia depende de la interacción de múltiples **factores genéticos** entre sí y con múltiples **factores ambientales**.

En Endocrinología son enfermedades complejas aquellas que constituyen los mayores problemas de Salud Pública como la obesidad común, síndrome metabólico, diabetes mellitus 2, síndrome de ovarios poliquísticos, las dislipidemias y neoplasias endocrinas más frecuentes. Las variantes genéticas más frecuentemente estudiadas en enfermedades complejas son los polimorfismos de un solo nucleótido (*SNPs*).

El desarrollo de tecnologías de genotipificación costo-efectivas ha puesto a disposición del lector miles de artículos de asociación genética en enfermedades complejas. La mayor parte de los estudios comunicados son del tipo *gen candidato*, establecen asociaciones entre *SNPs* y fenotipos relacionados con enfermedades complejas.

El riesgo del desenlace clínico atribuible al haplotipo de susceptibilidad suele ser pequeño ( $OR < 1,2$ ). La frecuencia de los haplotipos de susceptibilidad en las muestras estudiadas es muy variable. La replicación posterior de los resultados, en estudios con poblaciones independientes, fracasa con frecuencia. Todos estos hechos explican las crecientes exigencias metodológicas a los estudios de asociación genética por parte de los Comités Científicos de las principales revistas biomédicas. También obligan al lector, particularmente si es un médico clínico, a manejar herramientas para evaluar la calidad metodológica de los estudios de asociación genética aunque no cuente con formación en ciencias fundamentales.

<sup>1</sup>Centro de Excelencia CIGES. Unidad de Endocrinología, Departamento de Medicina Interna, Facultad de Medicina. Universidad de La Frontera. Temuco.

Correspondencia:  
Jorge Sapunar Zenteno Msc  
Antumalal 01024. Temuco. E-mail: jsapunar@ufro.cl

Recibido: 15 de mayo de 2012  
Aceptado: 24 de mayo de 2012

#### Objetivo de la guía de lectura

- Conocer los factores que explican la baja reproducibilidad de los estudios de asociación genética.
- Entregar algunas recomendaciones para la lectura crítica de artículos de asociación genética.

#### Diseñando un estudio de asociación genética

De un punto de vista epidemiológico los autores de un estudio de asociación genética se formulan la siguiente pregunta de investigación:

**¿Cuál es el efecto de la(s) variante(s) genética(s) X en el riesgo de expresar el (los) fenotipo(s) Y en los sujetos Z considerando variables de control?**

Esta pregunta da origen a 2 tipos de objetivos:

1. Comparar frecuencias alélicas entre sujetos con y sin el desenlace:

**Ha:  $f_{AA}$  (enfermos)  $>$   $f_{AA}$  (no enfermos)**

2. Comparar la frecuencia del desenlace entre sujetos con y sin el haplotipo de susceptibilidad:

**Ha:  $f_E$  (AA)  $>$   $f_E$  (Aa, aa)**

Donde Ha es hipótesis alternativa y  $f$  es la frecuencia.

La Tabla 1 resume los posibles hallazgos de un estudio de asociación genética.

Los estudios de asociación genética se pueden clasificar de acuerdo a su vinculación con una hipótesis:

1. *Estudios de asociación de genes candidatos*: Evalúan la hipótesis alternativa de que alteraciones en uno o más genes de una vía específica se asocian con la enfermedad de interés o desenlaces relacionados.
2. *Estudios de asociación genética amplia (GWA)*: Es un análisis de variación genética conducido sin hipótesis para detectar genes asociados por frecuencia a la enfermedad de interés o desenlaces relacionados.

Independientemente que éste corresponda a un estudio de gen candidato o GWA, puede ser diseñado como:

## Artículo por Invitación

Tabla 1.

	Desenlace (+)	Desenlace (-)	Total
Variante genética (+)	A	B	A+B
Variante genética (-)	C	D	C+D
Total	A + C	B+D	A+B+C+D

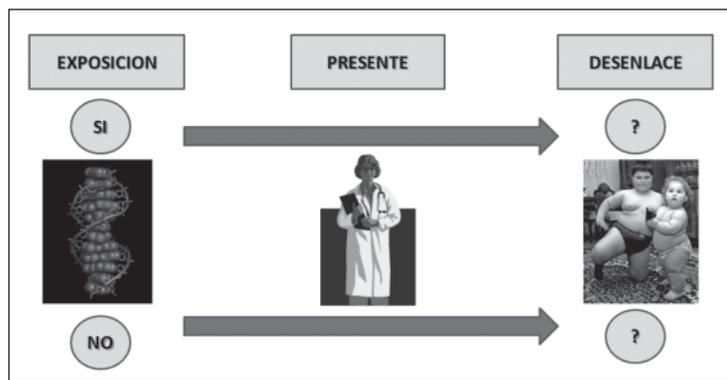


Figura 1. Estudios de cohortes prospectivo secuencia temporal.

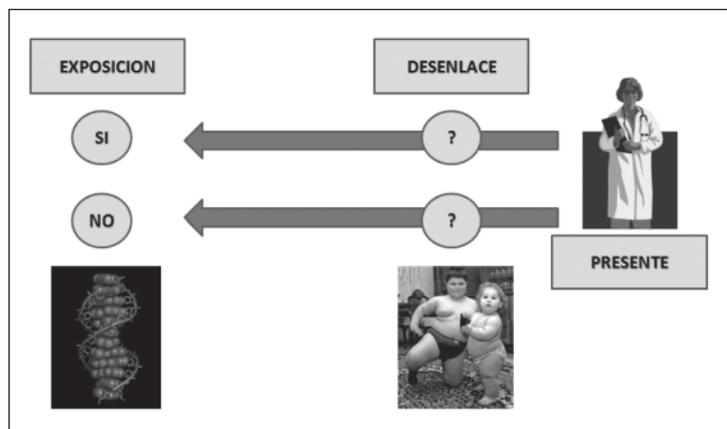


Figura 2. Estudios de cohortes retrospectivo secuencia temporal.

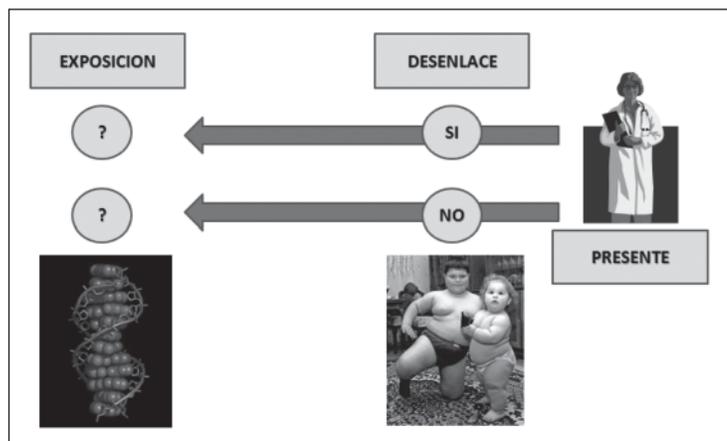


Figura 3. Estudios de casos y controles secuencia temporal.

1. *Estudio de corte transversal:* Este tipo de estudio utiliza los datos extraídos de una muestra representativa de la población de interés. Comunica el estado de un individuo de la muestra en términos de la presencia o ausencia de una variante genética para un determinado desenlace y la presencia o ausencia del mencionado desenlace en un tiempo determinado. Para exposiciones que persisten inalteradas en el tiempo (Género, etnia, genotipo), los estudios de corte transversal pueden establecer asociaciones estadísticas válidas. En la Tabla 1 se puede apreciar que la frecuencia (prevalencia) del desenlace en los expuestos es  $A/A+B$  y la frecuencia del desenlace en los no expuestos es  $C/C+D$ . La **Razón de Prevalencia (RP)** es el cociente entre la frecuencia del desenlace en los expuestos y la frecuencia del desenlace en los no expuestos. RP describe la magnitud de la asociación entre exposición y desenlace siempre y cuando la exposición sea inherente al individuo (Género, etnia, genotipo). Los estudios de corte transversal tienen ventajas como no tener limitaciones de tiempo y poder aportar frecuencia (Prevalencia) total y por subgrupos. Sin embargo tienen limitaciones importantes como el requerir de gran tamaño muestral y que la estimación de frecuencia sólo es válida si se cuenta con una muestra representativa de la población.
2. *Estudios de cohortes:* Este tipo de estudio selecciona a los sujetos a partir de tener o no la variante genética para luego seguirlos en el tiempo registrando la ocurrencia del desenlace (Cohorte Prospectiva) o registrando el antecedente de haberlo tenido (Cohorte Retrospectiva), como lo muestran las Figuras 1 y 2. En la Tabla 1 se puede apreciar como un estudios de cohortes permite estimar la incidencia del desenlace. La magnitud de asociación (**Riesgo**), debido a la secuencia temporal se establece mediante Riesgo Relativo (**RR**) o Riesgo Absoluto (**RA**).

$$RR = A/A + B/C/C + D$$

$$RA = (A/A + B) - (C/C+D)$$

Ej. RR de 2 significa que el riesgo de tener el desenlace en los sujetos con la variante genética es 2 veces mayor que en aquellos que no la tienen.

Los estudios de cohortes son óptimos en exposiciones raras (Variantes infrecuentes), permiten examinar más de un desenlace a la vez y entregan la incidencia del desenlace en expuestos y no expuestos. Sin embargo, son ineficientes con desenlaces raros y de latencia larga, casi siempre requieren de gran tamaño muestral y son demandantes de tiempo.

3. *Estudios de casos y controles:* Este tipo de estudio selecciona a los sujetos a partir del desenlace, es decir los casos serán aquellos con el fenotipo de interés y los controles aquellos sin el desenlace. En ambos grupos se investigará la presencia de la (s) variante (s) genética(s). En la Figura 3 se puede apreciar la secuencia temporal de un estudio de asociación genética de casos y controles. En la Tabla 1 se puede apreciar como la magnitud de asociación (**Riesgo**) se establece mediante Razón de Odds (**OR**).

## Artículo por Invitación

$$OR = A \times D / B \times C$$

Ej. OR de 2 significa que el riesgo de encontrar la variante genética en los sujetos con el desenlace es 2 veces mayor que en aquellos sin el desenlace.

Los estudios de casos y controles requieren de tamaño muestral relativamente pequeño determinado por la frecuencia del haplotipo. Son óptimos para enfermedades poco frecuentes, latencia prolongada y permiten evaluar exposiciones múltiples (GWA). Por estas razones este diseño es el favorito en estudios de asociación genética.

Una vez elegido el diseño del estudio de asociación genética debemos establecer el tamaño de la muestra. Considerando que el *diseño de elección es casos y controles*, el tamaño muestral dependería de:

1. Frecuencia alélica.
2. Nivel de significancia.
3. Poder estadístico.
4. Magnitud del efecto.

El tamaño disminuye en la medida que aumenta la frecuencia del alelo de susceptibilidad y el tamaño del efecto (OR alélico). En la Tabla 2 se puede apreciar el tamaño muestral requerido de acuerdo a la frecuencia alélica.

### Validez interna del estudio

Una vez informados de las características generales de un estudio de asociación genética, podemos iniciar su lectura crítica. La primera etapa es establecer su validez interna a través de comprobar el cumplimiento de los siguientes aspectos:

1. *Definición del fenotipo*: Es la definición estandarizada de la enfermedad y varía de acuerdo al aspecto de la enfermedad considerado:
  - Predisposición (Desenlaces intermediarios).
  - Ocurrencia de la enfermedad.
  - Complicaciones de la enfermedad.
  - Mortalidad, sobrevida.
  - Respuesta a tratamiento.

El fenotipo como variable de respuesta o desenlace puede asumir 2 formas:

- *Variable discreta* (Cualitativas, dicotómicas): Más relevante clínicamente, pero imprecisiones en la definición.
- *Variable continua* (Desenlace intermediario): Fáciles de medir con acuciosidad, le confieren mayor poder estadístico al análisis.

2. *Obtención de los grupos de comparación*: En las enfermedades complejas existen múltiples variables a considerar al momento de evaluar la asociación entre una variante genética y un desenlace. En los estudios de casos y controles se conocen como *variables de control* y en lo posible deben ser emparejadas o modeladas. De acuerdo

al efecto que las variables de control tengan en la asociación se clasifican en *confundentes o modificantes de efecto*.

**Para que un estudio de asociación genética sea adecuado debe existir una completa descripción de las variables de control.**

3. *Estratificación de la Población*: Una de las variables de control de los estudios de asociación genética es la composición étnica de la población. Esta variable confundente e incluso modificadora del efecto de la variante genética sobre el fenotipo se conoce como *estratificación poblacional*. La estratificación poblacional podría generar una asociación espúrea en 2 situaciones:

- Cuando la probabilidad de desarrollar la enfermedad estudiada varía con la ascendencia.
- Cuando los ancestros tienen diferencias en la frecuencia alélica de polimorfismos no relacionados con la enfermedad estudiada.

Para controlar la estratificación de la población existen 3 procedimientos:

- Uso de poblaciones homogéneas en términos de ancestros (autoreporte).
- Uso de controles extraídos de la familia.
- Uso de controles genómicos de marcadores no ligados (DNA mitocondrial, grupos sanguíneos para segregar amerindios).

4. *Error de genotipificación*: El error es inherente a todos los sistemas analíticos y puede depender de:

- Problemas con el material biológico (Comparar muestras frescas con almacenadas).
- Problemas con la técnica utilizada.

La magnitud del error de genotipificación varía ampliamente (1 a 30%) y suele minimizarse en GWA. Una manera de apreciar este error es a través de la tasa de recuperación. Muchos autores recomiendan no analizar *SNPs* con tasa de recuperación inferiores a 90%. Sin embargo,

**Tabla 1. Tamaño muestral necesario para detectar asociación significativa (poder 90%,  $\alpha = 0,001$ )**

OR alélico	Frecuencia del alelo de susceptibilidad en controles					
	1%	5%	10%	20%	30%	40%
1,1	221.997	46.434	24.626	13.987	10.759	9.505
1,2	58.177	12.217	6.509	3.730	2.896	2.581
1,3	27.055	5.702	3.051	1.763	1.380	1.240
1,5	10.604	2.249	1.213	712	566	516
2,0	3.193	687	377	229	188	177
4,0	598	134	78	52	46	47

Adaptado de Hattley A. Lancet; 366: 1315-23.

## Artículo por Invitación

tasas de recuperación altas para los genotipos homocigotos pueden combinarse con tasas de recuperación bajas para heterocigotos.

5. **Equilibrio de Hardy-Weinberg (EHW):** Las frecuencias alélicas en las poblaciones se mantienen inalteradas de una generación a otra cuando se cumplen las siguientes condiciones:
  - Falta de mutaciones que afecten la frecuencia alélica.
  - Gran descendencia poblacional.
  - Apareamiento al azar.
  - Baja migración.
  - Falta de selección natural.

Las proporciones genotípicas de HW para 2 alelos corresponden al cuadrado del binomio de sus frecuencias alélicas (Figura 4). Las proporciones observadas se comparan con las esperadas mediante prueba de hipótesis ( $\chi^2$  o F de Fischer). Un valor de  $p < 0,05$  es el umbral para detectar un desequilibrio.

Las desviaciones de EHW, respetando las 5 condiciones, sugieren error de genotipificación o estratificación poblacional. En GWA, donde se evalúan miles de asociaciones, hasta un 5% de los SNPs podrían violar EHW.

### Un artículo de asociación genética tiene validez interna si:

- Tiene una definición adecuada del fenotipo.
- Tiene una descripción completa de variables de control.
- Tiene algún procedimiento para controlar la estratificación genética.
- Reporta el error de genotipificación.
- Respeta el equilibrio de Hardy-Weinberg.

### Análisis de resultados

Una vez que comprobamos que el estudio de asociación genética cuenta con validez interna podemos analizar sus resultados. Todo estudio de asociación genética debe considerarse en sus resultados 2 aspectos:

1. **Comparaciones múltiples:** En un estudio asociación genética entre 100 SNPs y un desenlace clínico determinado, utilizando un valor  $p < 0,05$ , la probabilidad de asociación espúrea es 99,4%. Este problema afectaría principalmente a los GWA y podría resolverse por el **Método de Bonferroni** que corrige el valor  $p$  de acuerdo al número de comparaciones hechas ( $0,05/n$  comparaciones). En el ejemplo el valor de  $p$  sería  $< 0,0005$ . Para GWA que estudian miles de variantes genéticas los valores de  $p$  por Bonferroni serían extremadamente exigentes, por lo que se ha consensado un valor  $p$  de  $5 \times 10^{-8}$  (**Umbral de significación genómica amplia**).
2. **Probabilidad de falso positivo:** Cuando revisamos un estudio de asociación genética con hallazgos positivos y

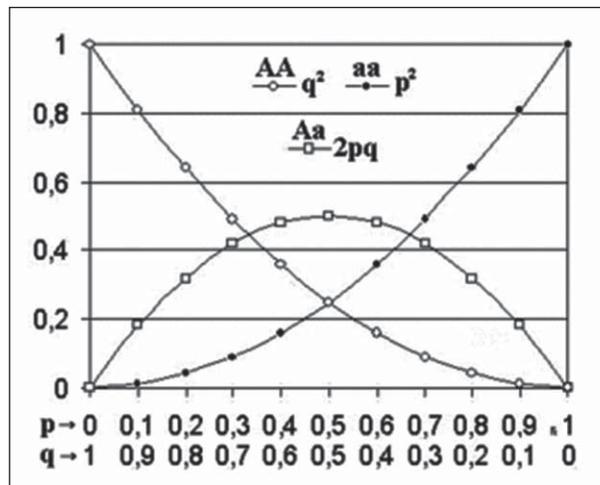


Figura 4. Equilibrio de Hardy-Weinberg.

estamos pensando en replicarlo, existen 2 clases de procedimientos para evaluar la probabilidad de que se trate de un falso positivo:

- **Método de Benjamini:** Utiliza un listado decreciente de valores  $p$  de las asociaciones encontradas en el estudio y ajusta cada valor particular de acuerdo a la posición.
- **Probabilidad de reporte falso (Wacholder):** Establece la probabilidad de una asociación falsa dado el poder del estudio y la OR previa de la asociación.

### Un artículo de asociación genética válido, reporta adecuadamente sus resultados cuando:

- Corrige el valor de  $p$  considerando comparaciones múltiples.
- Reporta la probabilidad de falso positivo.

### Validez externa del estudio

En los estudios de asociación genética la tendencia al falso positivo obliga a no considerar seriamente sus resultados hasta que no hayan sido replicadas por estudios independientes. Muchas asociaciones positivas encontradas por estudios pequeños que fueron publicados en el pasado (**Sesgo de publicación diferencial**), van perdiendo magnitud de efecto con la replicación (**Maldición del ganador**).

**Todo estudio de asociación genética debería establecer si se trata de un estudio inicial o es una replicación en población independiente para el mismo alelo, idéntico fenotipo y dirección de efecto.**

Los GWA evalúan SNPs en muestras adicionales de replicación para incrementar su tamaño muestral y alcanzar el **umbral de significación genómica amplia**. Más allá de los GWA hoy encontramos meta-análisis de estudios de genes candidatos y GWA, que nos permitirían encontrar evidencia

## Artículo por Invitación

de asociaciones genéticas replicadas. El sitio **Human Genome Epidemiology Network (HuGE)** reúne meta-análisis de estudios de asociación genética y dispone de un navegador para la búsqueda de estudios aislados, GWAs y meta-análisis (<http://www.hugenavigator.net/>)

### Referencias bibliográficas

1. Attia J, Ioannidis JPA, Thakkinstian A, McEvoy M, Scott R J, Minelli C, Thompson J, Infante-Rivard C, Guyatt G. 2009. How to use an article about genetic association. Are the results of the study valid? JAMA; 301 (2): 191-197.
2. Simundic AM. 2010. Methodological Issues of Genetic Association Studies. Clin Chem Lab Med; 48 (Suppl 1): S115-S118.
3. Hattley A, McCarthy MI. 2005. What makes a good genetic association study? Lancet; 366: 1315-1323.
4. NCI-NHGRI. 2007. Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. Nature; 477: 655-660.
5. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E, et al. 2009. STrengthening the REporting of Genetic Association studies (STREGA) - an extensión of the STROBE statement. Eur J Clin Invest 39: 247-266.