

Epidemiología genética: asociaciones con enfermedades complejas e implicancia para la salud pública

Cuando ya no sepas adónde vas, date la vuelta y mira de dónde vienes.

Proverbio senegalés

INTRODUCCIÓN

Se acepta que, en la definición habitual y clásica, la *epidemiología* se refiere al “estudio de la distribución, los determinantes y el control de los estados y de los eventos poblacionales relacionados con la salud”. (1)

Cuando se desarrolló el proyecto del “genoma humano”, a este concepto se le adosó la palabra genética y se comenzó a hablar de “*epidemiología genética*”, la cual pretende estudiar la arquitectura genética de las enfermedades complejas. Con los avances en la tecnología y el conocimiento biológico que subyace en la acción de los genes, podríamos decir que “*epidemiología genética*” es el trabajo que hacen aquellos que investigan las consecuencias sobre la salud de las variantes genéticas que continúan desarrollándose.

Se formó un mito alrededor de la epidemiología genética porque tiene una reputación de misteriosa, difícil y críptica, ya que su terminología y algunos de sus métodos que nacen de la biología genética son tan específicos y con un lenguaje propio que resultan incomprensibles en sus detalles para un clínico no iniciado. Pero lo mismo nos sucedió con la comprensión inicial de la epidemiología tradicional hasta que nos familiarizamos con sus distintos diseños y la convertimos en epidemiología clínica de utilización casi habitual en nuestra práctica.

Vamos a tratar de desmitificar unos pocos puntos cruciales, como un clínico que mira de soslayo estos nuevos campos. En el “glosario” adjunto se puede buscar la explicación de los términos que no nos resultan habituales en nuestra jerga médica. (1)

ASOCIACIÓN DE VARIANTES GENÉTICAS Y SUSCEPTIBILIDAD A ENFERMEDADES COMUNES

Comencemos revisando algunos de los últimos trabajos; en uno de los números de *Circulation* de octubre, Koch y colaboradores (2) nos decepcionan al publicar que la variante común en el gen del receptor α de estrógeno (ESRI -397 T < C) no muestra asociación en hombres ni en mujeres con el infarto de miocardio (IM). Nos decepciona porque en 2003 Shearman y colegas comunicaron que una variante común, el ESRI -397 alelo C, era más frecuente en varones con IM que en los controles sin IM. Al año siguiente, Schuit y colaboradores publican una asociación entre un haplotipo común que contiene el alelo alternante (T) y el IM, pero solamente en mujeres. (3)

GLOSARIO

Alelos

Si la secuencia de ADN en un locus dado (a menudo un gen o un marcador) varía entre diferentes cromosomas en la población, cada diferente versión es un alelo. Si hay dos alelos en un locus dado, el alelo que es menos común en la población es el alelo menor.

Diploide

La mayoría de las células humanas son diploides; todas contienen 46 cromosomas, es decir, una copia de ambos miembros de cada par homólogo (p. ej., dos versiones del cromosoma 18). El número de cromosomas humano diploide completo puede expresarse como 46,XX en una mujer y 46,XY en un hombre.

Haploide

Los gametos (el espermatozoides y los óvulos) son haploides. Contienen solamente un miembro de cada par de cromosomas homólogos (p. ej., sólo una versión del cromosoma 18). Todos los óvulos tienen 23,X cromosomas complementarios y el espermatozoides, 23,X o 23,Y. Cuando el espermatozoides y el óvulo se fusionan para formar un cigoto se restauran los cromosomas complementarios diploides.

Locus

Un locus es una localización cromosómica única que define la posición de un gen individual o una secuencia de ADN. En estudios de encadenamiento genético, el término también puede referirse a la región que implica uno o más genes.

Marcador genético

Un marcador genético es una variable de la secuencia de ADN que tiene un componente no variable que es suficientemente específico para localizarlo en un locus genómico único y un componente variable que es suficientemente heterogéneo para identificar diferencias entre individuos y entre cromosomas homólogos en un individuo. Los marcadores genéticos tienen un papel fundamental en el mapeo genético.

Penetrancia

La penetrancia es la probabilidad de que un fenotipo particular se exprese en una persona con un genotipo particular.

Polimorfismo

Implica que existen dos o más variantes genéticas en un locus designado. Un locus que es polimórfico tiene al menos dos alelos alternativos. Desafortunadamente, la palabra polimorfismo tiene otras definiciones alternativas más específicas (no universalmente aceptadas), por ejemplo alelos, (otras) variantes de secuencias, variantes de estructura cromosómica.

Polimorfismo de nucleótido único (SNP)

Es una variante del ADN que representa una variación en una única base. Un SNP común puede definirse como un locus en el cual dos alelos de SNP están presentes, ambos con una frecuencia 1%. El genoma humano podría albergar más de 10 millones de SNP.

¿Cómo pueden ser posibles resultados tan heterogéneos de asociación genética entre un gen y la posibilidad de aumentar el riesgo de un IM?

¿Acaso los médicos, como el resto de los mortales, no esperábamos que se cumpliera la profecía de Francis

Collins del US National Human Genome Research Institute? En 1999, en la conferencia Shattuck, describió la hipotética consulta en 2010 de un hombre de 23 años con un nivel de colesterol alto al que se le realizan extensas pruebas genéticas. Se le identifican 11 variantes genéticas, 3 desconocidas o de ficción que Collins predijo que se identificarían en 2010; en las 8 restantes, el riesgo elevado de enfermedad coronaria se debía a variantes en sus genes de la apolipoproteína B (APOB) y de la proteína colesterol éster transferasa (CETP). Con estas variantes en la APOB y la CETP se le identificó un riesgo relativo aumentado de 2,5 veces sobre la base de los pequeños estudios que se habían realizado hasta la fecha. (4) La variante APOBThr71Ile se ha investigado extensamente y en un estudio caso-control de 8.145 pacientes –mucho mayor que los estudios previos–, el *odds ratio* (OR) para enfermedad coronaria para Ser/Ser versus Asn/Asn fue de 1,15 (IC 95% 0,91-1,46) y para Ile/Ile versus Thr/Thr 0,95 (0,82-1,11). En el mismo estudio, la CETP variante TaqIb dio también un resultado decepcionante: el OR del genotipo B2/B2 fue 0,94 (0,83-1,06). O sea que en 2005 ni siquiera se cumpliría la profecía que Collins creía que ya era definitiva.

Esto que estamos relatando no es un fenómeno poco común; revisiones recientes indican que el 70% a 95% de las asociaciones que se publican no se confirman en los estudios posteriores. (3)

Este problema de las publicaciones de asociaciones genéticas que son falsamente positivas es tan grave que algunas respetadas revistas científicas declinan publicar trabajos sobre asociaciones genéticas de enfermedades complejas, excepto en circunstancias excepcionales. (5)

La mayoría de estos diseños son del tipo de caso-control, donde se determina si una variante de alelos es más o menos común en los casos que en los controles seleccionados apropiadamente.

Además de los sesgos conocidos de que los casos y los controles no son representativos de “todos” los casos y los controles y de la definición y la medición de los casos y los controles, en las asociaciones genéticas se deben tener en cuenta consideraciones adicionales, por ejemplo, que las variantes genéticas difieren en las poblaciones según la ascendencia (estratificación de la población) y como, además, los alelos marcadores no son uniformes en las poblaciones, puede ser posible que se encuentren asociados con un rasgo en una población y no en otra; en parte se está solucionando con la reciente publicación del mapa de haplotipos, llamado coloquialmente “HapMap” (6) de 269 personas (Ibadan de Nigeria, ancestros de Europa noroccidental de Utah, japoneses de Tokio y han de Pekín).

Pero el problema más importante para estas asociaciones es el número enorme de polimorfismos genéticos (\approx 11 millones con una frecuencia de alelos $>$ 1%) con el desarrollo de una nueva capacidad técnica para detectarlos en una gran cantidad de personas;

esto produjo la irresistible oportunidad de examinar muchísimas variantes (pescar asociaciones) con respecto a numerosísimos fenotipos (enfermedades o factores de riesgo) y por lo tanto, publicar sólo las asociaciones que aparentemente parecen las más promisorias o coherentes. Esto produce dos problemas metodológicos y estadísticos graves.

Uno de ellos es el llamado “sesgo de publicación”, ya que los estudios pequeños negativos no se envían a publicación y, más aún, en el caso de que se envíen, serán rechazados. Por ejemplo, se ha tenido considerable interés entre el polimorfismo ACE I/D y el riesgo de infarto de miocardio; en un metaanálisis de 19 estudios, cuando los *odds ratios* (OR) se asociaban con una medida del tamaño de las diferentes muestras en un *funnel plot* (diagrama del embudo), los estudios caso-control pequeños mostraban un gran efecto, pero sin embargo en el estudio más grande (ISIS de cerca de 5.000 casos y 6.000 controles) el riesgo estimado era pequeño y marginal, OR 1,10 (IC 95% 1,00-1,21). (7) En el gráfico (*funnel plot*) podía observarse que faltaban todos los estudios que contrabalanceaban el gran efecto de los estudios más chicos (sesgo de publicación).

El otro serio problema es que la diferencia de los grupos de casos y controles la expresamos utilizando el error tipo I, con un valor umbral de “*p*”, que indica que si la hipótesis nula es verdadera, la probabilidad de obtener un resultado tan extremo como el umbral será de cerca del 5% (0,05). Pero esto será cierto si planificamos y sólo realizamos una comparación; pero como los investigadores, por lo que comentamos previamente, en realidad realizan muchísimas comparaciones, la mayoría de los resultados que se publican son falsas asociaciones por error de tipo I o las asociaciones son menos notorias que las que se obtienen inicialmente; este fenómeno se ha dado en llamar la “*maldición de los ganadores*”. (3)

Si las pruebas múltiples que se realizan de las distintas asociaciones son independientes entre sí, se debería utilizar el ajuste de Bonferroni, donde el valor umbral de “*p*” se va haciendo más pequeño porque se debe dividir por el número (*n*) de pruebas (asociaciones) que se realizan ($0,05/n$); con este argumento, algunos investigadores estiman que deberían utilizarse valores de “*p*” más pequeños que una millonésima (10^{-8}).

Por lo tanto, sin llegar al extremismo de algunos científicos que dicen que como los estudios de asociación genética expresan tan mal la realidad deberían abandonarse, (3) es necesario no confiar en todas las asociaciones que se publican por primera vez, conocer cómo otros estudios replican los hallazgos y, quizá, desarrollar nuevos métodos *ad hoc* para estos nuevos problemas metodológicos y estadísticos.

Para salvar estas inconsistencias se ha propuesto el desarrollo de una serie de bancos biológicos (*bio-banks*), donde se almacenaría prospectiva y sistemáticamente el material biológico (p. ej., sangre o ADN extraídos) a gran escala, ya sea orientados por enfer-

medades o basados en la población. Los *biobanks* basados en la población almacenarían material e información de personas reclutadas de la población general según su residencia y se utilizarían para la investigación clínica como los estudios tradicionales de cohorte (seguimiento longitudinal). (4) Ejemplo de esto es el "UK Biobank", que de 2006 a 2010 enrollará una muestra de 500.000 personas. Es necesaria esta enorme cantidad de personas porque si se pretende detectar efectos directos pequeños, con un OR de alrededor de 1,15 a 1,30, para un poder del 80% y con una prevalencia en la población del 10% al 25%, al menos se necesitarían 5.000 casos, idealmente 10.000, de la enfermedad de interés. Por ejemplo, aun con este enorme *biobank*, se necesitarán 5 a 8 años para tener el número necesario de infartos de miocardio y muertes por enfermedad coronaria, un tiempo casi igual para el desarrollo de diabetes, pero más de 22 años para la aparición del cáncer de pulmón.

El conocimiento genético será de gran utilidad para la práctica clínica, si puede ponerse en un contexto epidemiológico, médico y de salud pública.

¿SCREENING GENÉTICO PARA PERSONALIZAR LA MEDICINA? ESPERANZA O ILUSIÓN

Persiste el mito de que realizándoles a los individuos un mapeo genético podríamos personalizar la prevención de enfermedades o mejorar el tratamiento de las enfermedades comunes.

Aun cuando aceptemos que el riesgo de cáncer de pulmón de un fumador comparado con otro aumenta 6 veces por la variante NAT2 (N-acetiltransferasa 2) como en el paciente hipotético de Francis Collins [en realidad en un estudio más reciente caso-control de más de 2.000 participantes, que eran o habían sido fumadores, la comparación de la variante genética de los acetiladores lentos *versus* los rápidos no mostraba diferencias, OR 0,96 (0,79-1,16)], aun así los individuos no susceptibles que fuman tienen un riesgo alto y prevenible de cáncer de pulmón y otras enfermedades comunes (EPOC, cáncer de laringe, etc.); de tal manera, lo más efectivo es implementar programas para dejar de fumar en toda la población, a menos que uno sea dueño de una compañía tabacalera. Se podrá argumentar que conocer la susceptibilidad genética puede motivar cambios de conducta en un individuo, pero no hay evidencia de esto y podríamos pensar en cambio que al conocer que es menos susceptible genéticamente que otros fumadores, se muestre menos inclinado a dejar el cigarrillo y de esa manera le estamos dando una falsa tranquilidad.

La otra vertiente es adecuar el tratamiento a las variantes genéticas que influyen en la respuesta (farmacogenómica). En un estudio reciente se encontró una diferencia significativa en la caída del colesterol con el tratamiento con estatina en asociación con dos variantes de la 3-hidroximetil-3-metil-glutaril coenzima A reductasa. En una variante, el LDL-colesterol

cae 34,1 mg/dl y en la otra, 27,7 mg/dl. Es poco probable que la diferencia (6,4 mg/dl), aunque existe, tenga alguna importancia en la práctica clínica, ya que el médico variará la dosis de estatina según la caída que observa y desea lograr en el colesterol más que de acuerdo con el genotipo del paciente.

Como Humphries y otros manifiestan, "*para que una prueba genética sea útil en el manejo de la enfermedad cardiovascular debe tener un poder predictivo más allá y por encima de los factores de riesgo aceptados que se pueden medir fácilmente, habitualmente de bajo costo y con alta reproducibilidad*". (4)

Además, los factores de riesgo de la conducta y las variables fisiológicas muestran cierto grado de agrupamiento, ya que los que tienen colesterol más alto también son hipertensos, con sobrepeso, tienen resistencia a la insulina o fuman. Tales agrupamientos suceden porque lo que subyace son diferentes modos de vida de los procesos sociales o alteraciones como la obesidad y las intervenciones poblacionales, independientemente del genotipo de las personas, producen una reducción sustancial de las enfermedades comunes. Por el otro lado, el tener una variante genética que aumenta el riesgo es independiente de las otras y, por lo tanto, el impacto de producir un riesgo atribuible a nivel poblacional es bajo; por ejemplo, el polimorfismo del gen del MTHFR sólo puede explicar el 2% de la variabilidad de los niveles de homocisteína de la población.

IMPLICACIONES DE LA "ALEATORIZACIÓN MENDELIANA"

Los epidemiólogos buscan las causas modificables de las enfermedades comunes para mejorar la salud de la población. Sin embargo, los estudios epidemiológicos observacionales pueden dar resultados espurios e inducir a asociaciones falsas, como cuando se afirmaba que la terapia de reposición hormonal protegía contra la enfermedad coronaria, o que la vitamina E reducía el riesgo cardiovascular. Cuando se efectuaron los estudios con ensayos clínicos controlados para confirmar estas hipótesis observacionales (donde los confundidores en ambos grupos, de intervención y de control, se igualan por acción de la aleatorización), estas afirmaciones se han refutado contundentemente.

Los hallazgos equívocos de los estudios observacionales epidemiológicos probablemente se deben a la presencia de confundidores en factores como las variables fisiológicas, la conducta o el modo de vida y aun las variables socioeconómicas relacionadas tanto con las "exposiciones" como con los "puntos finales de las enfermedades".

El control de algunos de los confundidores reconocidos por métodos estadísticos multivariados no fue suficiente en muchas de las situaciones como las planteadas y demostró su dificultad cuando la exposición en estudio está relacionada con muchos otros factores conocidos y aun desconocidos, que influyen en el riesgo de padecer la enfermedad.

Hasta hace poco tiempo, la única solución plausible era la realización de experimentos controlados, para que entre el grupo activo y el control lo único distinto fuera la intervención. Esto se logra con los ensayos clínicos aleatorizados que, cuando están bien diseñados, constituyen el *gold standard* de la evidencia clínica.

Claro que resulta muy difícil o diríamos imposible realizar ensayos clínicos aleatorizados de hábitos de conducta, como sería el beber alcohol, fumar, etc.

Otra solución reciente a este viejo problema, clave para la epidemiología clásica y clínica, sería la "aleatorización mendeliana".

La aleatorización mendeliana se basa en la llamada segunda ley de Mendel, en la que se postula que la herencia de un rasgo es independiente de la herencia de otros rasgos.

El principio básico de estos estudios es que las variantes genéticas alteran ya sea el nivel o reflejan los efectos biológicos de una exposición ambiental o en el modo de vida, que en sí misma altera el riesgo de enfermedad; de esa manera, estas variantes genéticas estarían relacionadas con el riesgo de la enfermedad en la extensión predicha por su efecto sobre la exposición al factor de riesgo. Los polimorfismos genéticos comunes que tienen una función biológica bien caracterizada (o son marcadores de tales variantes) pueden, por lo tanto, utilizarse para estudiar el efecto de las exposiciones ambientales o del modo de vida en las que se sospecha un aumento del riesgo de enfermedad. (8)

Dado que, de manera distinta de las exposiciones ambientales habituales, las variantes genéticas no están habitualmente asociadas con los distintos rangos de factores fisiológicos y sociales, la segregación aleatoria de los alelos se comporta como un método de aleatorización clásico y, por lo tanto, la exposición a "un alelo" es similar a la "intervención" en los pacientes expuestos y el "otro alelo" formaría el grupo "control", similar también al grupo de no intervención (véase Figura 1).

En esta situación, la así llamada "aleatorización mendeliana", al igual que la "aleatorización de los ensayos clínicos", permitiría que ambos grupos presentaran confundidores iguales que no influirían en el resultado y, así, comparar los resultados entre los grupos por la intervención o la variante del alelo sin los sesgos habituales en los estudios observacionales.

Este aspecto poderoso de la aleatorización mendeliana se puede ilustrar con el estudio publicado este año por G. Davey Smith y colaboradores, titulado "Asociación de proteína C reactiva con la presión arterial y la hipertensión". (9) Esta asociación se ha publicado reiteradas veces en estudios de corte transversal (encuestas), pero se desconocía si esta asociación era causal. En la muestra representativa de más de 3.500 mujeres británicas de 60 a 79 años se encontró una asociación significativa y creciente entre los cuartiles de niveles de PCR (Tabla 1); pero al mismo tiempo, también existía una fuerte asociación significativa con

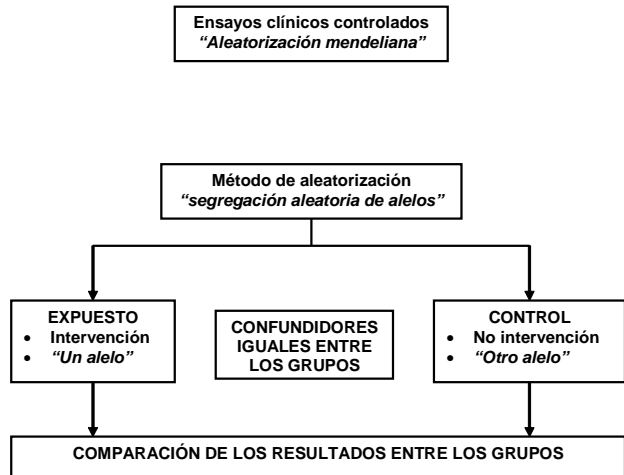


Fig. 1

otras variables, como el aumento del índice de masa corporal (IMC), el descenso del HDL-colesterol, el incremento de la diabetes y del hábito de fumar y la disminución de la actividad física y el consumo de alcohol, que actúan como confundidores de la asociación de la PCR con la hipertensión arterial (Tabla 1).

Para investigar si existe causalidad, utilizaron el método de "aleatorización mendeliana" y examinaron la asociación del polimorfismo 1059G/C del gen de la PCR humana con los niveles de PCR y con la hipertensión. En la Tabla 1 puede observarse que en la aleatorización mendeliana de la PCR con los alelos GG *versus* GC o CC se observa una diferencia pronunciada, de 0,42 mg/dl, y los mismos confundidores que se manifestaban cuando la muestra total se distribuía por cuartiles, en la aleatorización mendeliana desaparecen como confundidores.

El efecto de la asociación de niveles mayores de PCR con hipertensión se atenúa mucho cuando se realiza el ajuste estadístico por los distintos confundidores. El polimorfismo estaba asociado con una diferencia pronunciada en los niveles de PCR (GC o CC 1,39 mg/L *versus* GG 1,81 mg/L) y, por lo tanto, se debería esperar una mayor prevalencia de la hipertensión si ésta era causal porque no existían las variables confundidoras; sin embargo, el resultado mostró que el efecto fue cercano a 0.

Otra diferencia importante con los estudios observacionales clásicos es que en éstos existe la propensión al sesgo de la llamada "causación inversa", o sea que los procesos de enfermedad pueden influir en los niveles de exposición, como disminuir o suspender el alcohol, o la enfermedad puede afectar la medición de fenotipos intermedios como la PCR, el fibrinógeno o el colesterol. Esto no sucede con la aleatorización mendeliana, porque el polimorfismo genético aparece desde la gestación y es previo al desarrollo de cualquier enfermedad.

Otro efecto interesante de las variantes genéticas que influyen en los factores bioquímicos circulantes,

TABLA 1

Proteína C reactiva	Cuartiles				p tend	Promedio (mg/L)		p
	1	2	3	4		GG 1,81	GC o CC 1,39	
Resultado								
Hipertensión (%)	45,8	49,7	57,5	60,7	< 0,001	53,3	53,1	0,95
Confundidores								
IMC (kg/m ²)	25,2	27,0	28,5	29,7	< 0,001	27,5	27,8	0,29
HDL col. (mmol/L)	1,80	1,69	1,63	1,53	< 0,001	1,67	1,65	0,38
Diabetes (%)	3,50	2,80	4,10	8,40	< 0,001	4,70	4,50	0,80
Fumador (%)	1,90	9,60	10,9	15,4	< 0,001	11,2	9,30	0,24
Inactividad física (%)	11,3	14,9	20,1	29,6	< 0,001	18,9	18,9	1,00
Consumo alcohol (%)	22,2	19,6	18,8	14,0	< 0,001	18,6	19,8	0,56

como las concentraciones de colesterol, fibrinógeno, homocisteína, es que en la aleatorización mendeliana los diferentes alelos indican distintos niveles de las variables durante el curso de toda la vida y en consecuencia indicarían el efecto de una variación a muy largo plazo.

Por ejemplo, la hipercolesterolemia familiar es una condición mendeliana dominante, donde la rara mutación de un gen del receptor de lipoproteínas de baja densidad hace que la concentración promedio de colesterol de esta enfermedad se encuentre alrededor de 3,0 mmol/L más alta que en la población general. Con esta variación, la evidencia de los estudios aleatorizados de estatina predeciría un aumento de algo más del doble de enfermedad coronaria para estos pacientes, pero realmente ellos presentan un riesgo cuatro veces mayor. La realidad es que la disminución relativa de la mortalidad por enfermedad coronaria se acrecienta con el paso del tiempo y sería de esperar que a más largo plazo se acerque a la variación genética, que no tiene el problema de la adherencia al tratamiento ni a las fluctuaciones en los niveles de colesterol.

OTRA CIRCUNSTANCIA DE EXPOSICIÓN AL MEDIO INTRAUTERINO Y ALTERACIONES DEL FETO

Las exposiciones a que se halla sometido el feto pueden estar modificadas por el genotipo de los padres. Por ejemplo, actualmente conocemos que el déficit de ácido fólico durante el embarazo causa defectos del tubo neural.

Un polimorfismo en el gen metileno-tetra-hidrofolato reductasa (MTHFR) está asociado con un defecto metabólico equivalente al de los que ingieren poco ácido fólico; por lo tanto, en un metaanálisis de estudios casos-control de defecto del tubo neural, las madres con MTHFR TT (homocigota) comparadas con las MTHFR CC (homocigota) tenían el doble de riesgo. Refuerza la hipótesis que el alelo TT de la madre actúa por una disminución del ácido fólico del medio uterino; si además sabemos que el resultado no está influido por el genotipo TT del padre o del hijo, se

descarta así que una variante genética del descendiente provoque el defecto en el tubo neural.

Estos resultados no deberían llevar a pensar que debe realizarse un *screening* en todas las embarazadas para detectar si son portadoras de los alelos TT de la MTHFR, porque las madres sin el alelo TT pero con baja ingesta de ácido fólico experimentan un incremento del riesgo de tener un bebé con defecto del tubo neural, que es fácilmente prevenible. Por ello, lo único que evidencian estos resultados es que el consumo de ácido fólico materno es el factor etiológico clave para prevenir este defecto inviable.

UN ÚLTIMO ESTUDIO: LOS VEGETALES TIPO COLES REDUCEN EL CÁNCER DE PULMÓN

Si bien algunos estudios observacionales mostraban un papel protector del cáncer de pulmón por el consumo de vegetales, más específicamente vegetales tipo coles como el repollo y el brócoli, y que este efecto se explicaba porque estos vegetales son ricos en isotiocianatos que en animales demostraron una fuerte actividad quimiopreventiva, en una revisión reciente no se pudo identificar un efecto protector definitivo, quizá debido a que eran estudios pequeños con confundidores por otras fuentes de alimentos.

Se conoce que los isotiocianatos son eliminados por la enzima glutatión-S-transferasa (GST), fundamentalmente por dos variantes, la GST M1 y la GST T1, que tienen alelos con genotipo homocigoto de nulidad en los cuales no se produce la enzima y, por lo tanto, las formas inactivas de uno o ambos genes deben tener concentraciones más altas de isotiocianatos debido a una capacidad reducida de eliminación.

Estas condiciones permiten realizar una "aleatorización mendeliana"; por lo tanto, los papeles de los genes de la GST M1 y GST T1 probablemente serán independientes de otros factores dietéticos y de modos de vida, lo cual reducirá la posibilidad de confundidores.

P. Brennan y colaboradores (10) en octubre de este año publicaron un estudio caso-control con 2.141 casos y 2.168 controles en seis países de Europa central y del este.

Como era de esperar, el consumo de vegetales de tipo coles estaba relacionado con otros alimentos, como frutas y otros vegetales, y con la condición de fumador (Tabla 2).

A su vez, en la aleatorización mendeliana por el genotipo positivo o nulo de los genes GST M1 y GST T1, los confundidores se distribuyen igual en ambos grupos (Tabla 2).

Se observó un efecto protector total en los que consumían vegetales tipo coles al menos una vez por semana en comparación con aquellos que lo hacían menos que mensualmente (OR ajustado 0,78; IC 95% 0,64-0,96). Cuando se estratificaron por el estado de la GST, el efecto se restringió a aquellos que presentaban alelos nulos para GST M1 (0,67; 0,49-0,91), GST T1 (0,63; 0,37-1,07) o para ambas GST M1 y GST T1 (0,28; 0,11-0,67). Estos hallazgos nos llevan a considerar el importante efecto quimiopreventivo contra el cáncer de pulmón (reducción del 70%) de los vegetales tipo coles en general y de los isotiocianatos en particular.

LIMITACIONES DE LA ALEATORIZACIÓN MENDELIANA

Hablamos de su potencial, ahora también tenemos que hablar de las limitaciones de este método; la más sencilla de comprender es la ausencia de un polimorfismo adecuado para estudiar exposiciones modificables que nos son de interés (p. ej., no se identificaron variantes genéticas relacionadas con las concentraciones de vitamina C).

La dificultad mayor es común a todos los estudios de asociación genética: el fracaso en obtener una asociación real entre el genotipo y el fenotipo intermedio o la enfermedad debido al tamaño limitado de la muestra de los estudios actuales, como ya discutimos anteriormente. Un ejemplo reciente es la asociación entre el genotipo de la MTHFR y la enfermedad coronaria; un metaanálisis de estudios con el genotipo MTHFR TT, que eleva la homocisteína, mostró un incremento en el riesgo de enfermedad coronaria y se tomó como una fuerte evidencia de la naturaleza causal de la asociación entre homocisteína y enfermedad. Sin embargo, recientemente, un gran estudio demostró que la fuerza de la asociación (y, por lo tanto, el efecto pro-

tektor del ácido fólico) pudo haberse sobrestimado por un sesgo de publicación.

Aunque el control de los confundidores es una gran ventaja de la aleatorización mendeliana, debemos asegurarnos de que es cierto, porque los confundidores pueden aparecer por los efectos múltiples de un genotipo (pleiotropía) o porque una variable genética está físicamente cercana y entonces se transmite con otra variante funcional (desequilibrio de unión), o por canalización (desarrollo compensatorio en la función y la estructura de los tejidos).

Los estudios de variantes genéticas (aleatorización mendeliana) ponen en el tapete la posibilidad de realizar inferencias de cómo los cambios en las exposiciones ambientales en toda la población pueden reducir los riesgos de enfermedades. Lo que es más importante, las inferencias conciernen a atribuciones de causalidad que son relevantes para toda la población y no se refieren a intervenciones puntuales en aquellos con variantes genéticas específicas. Ya referimos que las variantes genéticas de la MTHFR de la madre se relacionan con defecto del tubo neural debido a una disminución del ácido fólico y que este hallazgo es relevante para todas las embarazadas, por lo que deberían recibir suplemento de ácido fólico, y no para realizar *screening* genético en ellas, ya que también se beneficiarían las que no tienen modificación del gen. Aplicando el conocimiento desarrollado por la epidemiología genética, es posible mejorar el conocimiento de cómo las diversas exposiciones ambientales, muchos de ellas modos de vida de grupos poblacionales, modelan la salud de la población y cómo podemos desarrollar inteligentemente futuras medidas de prevención de la salud pública.

CONCLUSIONES

En la última década, la capacidad técnica para producir una cantidad enorme de datos del genoma humano ha dejado muy atrás nuestra habilidad para interpretar esos mismos datos y también estuvo por detrás de nuestra capacidad estadística actual.

¿Qué nos pueden decir el descubrimiento de genes para las enfermedades complejas y comunes? La ma-

TABLA 2

	Consumo de coles			p	Genotipo GST		p
	Bajo	Medio	Alto		TM1+	TM1 nula	
Confundidores							
Hombres (%)	76,0	73,5	71,7	0,344	72,7	72,8	0,93
Alguna vez fumador (%)	73,6	64,2	65,4	0,021	64,6	67,6	0,16
Vegetales (%)							
Bajo	47,6	41,1	25,2		34,5	32,2	
Alto	16,0	27,9	50,5	< 0,001	32,6	34,0	0,53
Frutas (%)							
Bajo	32,8	32,2	30,9		29,3	33,3	
Alto	41,2	35,9	43,3	0,012	41,5	39,9	0,13

yoría de estas enfermedades implican a múltiples genes que predisponen a las distintas enfermedades con un efecto individual modesto, por lo cual, como hemos visto, la detección inicial de los genes que se relacionan con las enfermedades es difícil, al igual que la replicación de esos hallazgos.

Pero deberíamos ser prudentemente optimistas del futuro, ya que si bien la epidemiología tradicional ayudó a dilucidar el rompecabezas de las incoherencias de las asociaciones genéticas y nos permitió abrir nuevas vías de desarrollo de la investigación, como la vuelta a los estudios de cohorte con los *biobanks*, la epidemiología genética en su fase inicial ya contribuye al torrente principal de la epidemiología clásica y clínica y la ha beneficiado con el desarrollo de la aleatorización mendeliana, al permitirle encontrar relaciones causales pequeñas imposibles de probar con los métodos clásicos debido a los sesgos de los confundidores, la causalidad inversa y la heterogeneidad etiológica.

Para seguir avanzando debemos aprender los unos de los otros y trabajar codo a codo el clínico práctico, el epidemiólogo, el epidemiólogo genético, el estadígrafo, el bioinformático y el investigador básico; todos tenemos mucho que aprender del otro.

Hernán C. Doval

BIBLIOGRAFÍA

1. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet* 2005;366:941-51.
2. Koch W, Hoppmann P, Pfeufer A, Mueller JC, Schömig A, Kastrati A. No replication of association between estrogen receptor a gene polymorphisms and susceptibility to myocardial infarction in a large sample of patients of European descent. *Circulation* 2005;112:2138-42.
3. Herrington D. Eliminating the improbable. Sherlock Holmes and standards of evidence in the genomic age. *Circulation* 2005;112:2081-84.
4. Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* 2005;366:1484-98.
5. <http://www.jci.org/misc/jcipoli.pdf>
6. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P; International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299-320.
7. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361:865-72.
8. Davey Smith G, Ebrahim S. What can mendelian randomization tell us about modifiable behavioural and environmental exposures? *BMJ* 2005;330:1076-9.
9. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Rumley A, Lowe GD. Association of C-Reactive Protein with blood pressure and hypertension. *Arterioscler Thromb Vasc Biol* 2005;25:1051.
10. Brennan P, Hsu CC, Moullan N, Szeszenia-Dabrowska N, Lissowska J, Zaridze D, et al. Effect of cruciferous vegetables on lung cancer in patients stratified by genetic status: a mendelian randomisation approach. *Lancet* 2005;366:1558-60.