

La epidemiología genética: disciplina científica en expansión¹

Diego F. Wyszynski²

RESUMEN

La epidemiología genética es una disciplina relativamente reciente que estudia la interacción entre los factores genéticos y ambientales que dan origen a las enfermedades del ser humano. Valiéndose de marcadores genéticos desarrollados a través de la biología molecular, de complejos algoritmos almacenados en computadoras y de amplias bases de datos, la epidemiología genética se ha desarrollado notablemente durante los últimos 10 años. El presente artículo describe los objetivos de la epidemiología genética y su metodología, empleando ejemplos concretos tomados de la literatura científica reciente.

El paradigma de los “factores ambientales que interactúan con el genoma en el origen de las enfermedades” surgió a mediados del siglo XIX, cuando se observó que ciertos individuos eran más resistentes que otros a las enfermedades infecciosas. Sin embargo, pasaron casi 100 años antes de que los epidemiólogos interesados en la genética y los genetistas interesados en la epidemiología pudieran desarrollar los primeros métodos analíticos para identificar los factores ambientales y genéticos involucrados en los procesos patológicos (1).

Si bien ya se habían acuñado con anterioridad expresiones tales como “genética epidemiológica” (3) y “genética poblacional clínica” (4), Morton y Chung (2) fueron los primeros en

denominar epidemiología genética a la disciplina que se ocupa de controlar y prevenir las enfermedades. Su medio es la identificación de la función que cumplen los factores genéticos, en interacción con factores ambientales, en el origen de las enfermedades en los seres humanos (5).

La prevención puede llevarse a cabo en los niveles primario, secundario y terciario. La prevención primaria se refiere a la prevención de la incidencia de la enfermedad en la población (6). El ejemplo más conocido de prevención primaria es la inmunización destinada a evitar ciertas enfermedades infecciosas. En el ámbito de la epidemiología genética, evitar el factor de riesgo ambiental (por ejemplo, el tabaquismo materno) que interactúa con la susceptibilidad genética (genotipo A2 del marcador *TGF α TaqI* en el feto) conducente a determinado proceso patológico (paladar fisurado) constituye un ejemplo de prevención primaria (7). La prevención secundaria se refiere a la prevención de las manifestaciones clínicas de una enfermedad mediante la detección temprana de la misma y de una intervención eficaz en

la etapa preclínica (6). Ejemplos clásicos de prevención secundaria son la detección e intervención tempranas en los casos de hipotiroidismo congénito y fenilcetonuria. Finalmente, la prevención terciaria consiste en reducir a un mínimo los efectos de una enfermedad al evitar sus complicaciones y el deterioro que causa. Un ejemplo de la prevención terciaria de una enfermedad genética es la aplicación de profilaxis con antibióticos e inmunización en individuos con anemia falciforme para prevenir ciertas infecciones bacterianas que pondrían en grave peligro la vida del paciente.

Las mutaciones genéticas son la base de la variación poblacional (8). Las enfermedades, así como otros rasgos expresados o manifiestos clínicamente (fenotipos), se relacionan con los factores genéticos de tres maneras, que no siempre se excluyen mutuamente:

1. La nueva mutación puede ser directamente perjudicial para el individuo. Esta categoría incluye a los numerosos trastornos transmitidos de manera autosómica dominante a través de un gen único, como la

¹ Próximamente se publicará en esta revista una versión en inglés de este artículo.

² Universidad Johns Hopkins, Facultad de Higiene y Salud Pública, Baltimore, Maryland, EUA. Dirección postal: Department of Epidemiology, School of Hygiene and Public Health, The Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205, EUA. Tel: (410) 955-7961, Fax: (410) 955-0863. E-mail: dfw@welchlink.welch.jhu.edu

acondroplasia y el síndrome de Marfán.

2. La nueva mutación puede ser perjudicial, pero permanece silenciosa a través de las generaciones. Por ejemplo, ciertos errores metabólicos del recién nacido, como la fibrosis quística, solo se hacen evidentes cuando el individuo hereda dos copias (alelos) del gen mutado, es decir, una de cada progenitor.
3. La mutación puede ser perjudicial solo cuando interactúa con otros factores genéticos o ambientales (1). Por ejemplo, los individuos que poseen ambos alelos mutados para la fenilcetonuria o el hipotiroidismo congénito solo padecen estas enfermedades cuando están expuestos a altas concentraciones de fenalanina o a concentraciones reducidas de hormona tiroidea, respectivamente.

Es posible contrastar los objetivos de la epidemiología genética con los de la epidemiología "tradicional" y de la genética poblacional. La epidemiología "tradicional" estudia la relación entre el ambiente y la incidencia de determinada enfermedad, aun reconociendo la importancia del huésped y su constitución genética. La genética poblacional, en cambio, se ocupa de predecir las consecuencias que entrañan la estructura de la población y los fenómenos de selección y mutación para los fenotipos constitucionales y las enfermedades. Finalmente, la epidemiología genética estudia la manera en que los factores de riesgo presentes en el medio ambiente interactúan con la constitución genética de una población determinada.

LOS MÉTODOS DE LA EPIDEMIOLOGÍA GENÉTICA

Las estrategias de investigación en epidemiología genética son de dos tipos: descriptivas y analíticas. Las descriptivas, tanto en el nivel poblacional como familiar, se basan en el estudio del tiempo, el lugar y la persona. Algunas preguntas que ejemplifican esta estrategia son: ¿Cuál es la prevalencia al nacimiento de acondro-

plasia en la población y cuál es la tasa de mutación para esta enfermedad? ¿Cuál es la frecuencia de los grupos sanguíneos y de los antígenos de histocompatibilidad en distintas poblaciones? ¿Existen diferencias geográficas en la prevalencia de un factor genético determinado? Los estudios analíticos, por el contrario, tienen como objetivo identificar la función de factores genéticos en la historia natural de las enfermedades, tanto en poblaciones como en familias. Las preguntas a las que responden los estudios analíticos son ¿por qué . . ? y ¿cómo . . ? .

Estudios de recurrencia familiar

Un aspecto fundamental de la epidemiología genética es el estudio de la agregación (o recurrencia) de ciertas enfermedades en determinadas familias. King, et al. (9) propusieron tres preguntas que permiten identificar los alcances de los estudios sobre recurrencia familiar:

1. ¿Hay enfermedades que afectan a varios miembros de una misma familia?
2. ¿Se relaciona dicha agregación familiar con una exposición ambiental común, con una susceptibilidad heredada, o con una herencia cultural de factores de riesgo?
3. De existir, ¿cómo se hereda la susceptibilidad genética?

La observación de la prevalencia de cierta enfermedad en familiares de los casos índice (caso índice es el individuo afectado por medio del cual se incorpora su familia al estudio) de los controles (individuos no afectados) permite determinar la existencia de una agregación familiar. Dicha agregación existe cuando los familiares de los individuos afectados corren un mayor riesgo de padecer la enfermedad que los familiares de individuos no afectados. Este método es eficiente y poco costoso, pero una de sus limitaciones es que la información sobre las características de los familiares de los casos y controles puede dar lugar a sesgos. Por ejemplo, si el investigador

sabe de la presencia de la enfermedad en la familia del participante, existe la posibilidad de un sobrediagnóstico. La capacidad de recordación de los familiares y su conocimiento de las características del trastorno también pueden ser mayores cuando existe un pariente afectado. El cuadro 1 muestra una forma simple de calcular el riesgo relativo (RR) mediante el uso de una tabla de 2×2 , ejemplificado por el estudio de Mettlin, et al. (10), quienes estudiaron la historia familiar de cáncer de mama en 779 casos y 1 558 controles atendidos en el Roswell Park Memorial Institute, Buffalo, Nueva York. El RR de padecer cáncer de mama asociado a una historia familiar positiva fue de 1,62 [IC95%: 1,28 a 2,06] (cuadro 1). Cuando el análisis fue desglosado según la edad de los casos y controles (< 65 años), los RR fueron de 1,34 (IC95%: 0,94 a 1,92) y 1,88 (IC95%: 1,37 a 2,58), respectivamente (cuadro 2). Esta diferencia pone de manifiesto una limitación de los estudios de familiares de casos y controles, especialmente cuando se estudian trastornos que aparecen en edad avanzada: los familiares de los casos jóvenes tienden a ser más jóvenes que los de los controles.

Otros métodos, como los análisis de cohortes, las regresiones y las ecuaciones de estimación generalizables, permiten ampliar los cálculos para referirlos a situaciones más complejas. Es importante destacar que una elevada agregación familiar no prueba la existencia de un mecanismo genético en el origen de la enfermedad, así como una baja recurrencia tampoco excluye la posibilidad de que dicho mecanismo exista.

En tanto que la comparación de los familiares de los casos y controles puede considerarse una técnica "epidemiológica", es factible también identificar la presencia de agregación familiar a través de métodos de "genética estadística". En este caso, el grado de agregación de una enfermedad en una familia se expresa como λ_R , definido como el cociente entre el riesgo de padecer la enfermedad entre los parientes de los casos y la prevalencia de dicha enfermedad en la población

en general. Esta técnica requiere el cálculo de λ para cada uno de los grados de parentesco. El cuadro 3 muestra los resultados del trabajo de Slater y Cowie (11), quienes recopilaron los datos de los primeros estudios familiares publicados sobre esquizofrenia. Como puede apreciarse, cuanto más alejado es el grado de parentesco, λ se aproxima más a 1. Es importante destacar que dicha relación no es suficiente para atribuir a la esquizofrenia un origen puramente genético.

En el caso de enfermedades de herencia multifactorial, en la covarianza o correlación entre parientes sanguíneos se distinguen dos componentes: el atribuible a diferencias genéticas y el originado por diferencias en exposiciones ambientales. Para fenotipos discretos (afectado frente a no afectado), el modelo estadístico se basa en la premisa de que existe una predisposición continua (*liability*, en inglés) de distribución normal, que determina el riesgo de padecer la enfermedad. Según este modelo, cuando se traspasa determinado umbral (*threshold*, en inglés), se padece la enfermedad. Tanto la susceptibilidad como el umbral pueden ser heredables y las propiedades matemáticas de la distribución normal permiten predecir el parámetro λ . El análisis del modelo multifactorial se diseña con el objeto de estimar la correlación de riesgo entre parientes (12). Este modelo no distingue lo genético de lo ambiental y la estimación de heredabilidad puede ser sobrevaluada, especialmente cuando existen factores ambientales que influyen de manera importante sobre el riesgo entre parientes. El modelo lineal multifactorial también es aplicable a fenotipos que constituyen variables continuas, tales como las concentraciones de lípidos o glucosa en sangre, la tensión arterial y los valores hormonales. El análisis de componentes de la varianza, o, alternatively, el análisis de trayectos (*path analysis*) también son métodos útiles para el estudio de estos fenotipos.

De haber pruebas de agregación familiar y control genético de una enfermedad, la tercera pregunta que surge es cómo identificar el mecanismo

CUADRO 1. Riesgo relativo de padecer cáncer de mama asociado a una historia familiar positiva y negativa, según lo observado en un grupo de 779 casos de cáncer de mama y de 1 558 controles teóricos

		Casos	Controles	Total de casos	Total de controles
Otro familiar afectado ^a	Sí	a	b	144	191
	No	c	d	635	1 367
Riesgo relativo (IC95%)			ad/bc	1,62 (1,28 a 2,06)	1,00

Fuente: Referencia 10.

^a Otro familiar afectado significa un pariente en primer grado (madre, hija o hermana) con cáncer de mama.

CUADRO 2. Riesgo relativo de padecer cáncer de mama asociado a una historia familiar positiva y negativa, según lo observado en 779 casos y 1 558 controles, desglosados por edad

		Edad < 55 años		Edad ≥ 55 años	
		Casos	Controles	Casos	Controles
Otro familiar afectado	Sí	58	90	86	101
	No	300	626	335	741
Riesgo relativo (IC95%)		1,34 (0,94 a 1,92)	1,00	1,88 (1,37 a 2,58)	1,00

Fuente: Referencia 10.

CUADRO 3. Primeros estudios sobre el riesgo familiar de padecer esquizofrenia

Período	Estudios	Parentesco	Incidencia ^c	λ^a
1928–1962	14	Padres	336/7675 = 4,36% (valor corregido ^b = 14,12%)	5,45 17,65 ^b
1928–1962	12	Hermanos	724/8504 = 8,51%	10,6
1921–1962	5	Hijos	151/1226 = 12,31%	15,4
1930–1941	4	Tíos	68/3376 = 2,01%	2,5
1916–1946	3	Medio hermanos	10/311 = 3,22%	4,0
1926–1938	5	Sobrinos	52/2315 = 2,25%	2,8
1928–1938	4	Nietos	20/713 = 2,81%	3,5
1928–1941	4	Primos	71/2438 = 2,91%	3,6

Fuente: Referencia 11.

^a Los valores se calcularon suponiendo una prevalencia poblacional de 0,8%.

^b La corrección se introdujo debido a que, una vez que la esquizofrenia se ha manifestado, es raro que el paciente tenga hijos.

^c Calculada con la siguiente fórmula:

$$\lambda = \frac{\text{Individuos con parentesco X que desarrollaron la enfermedad durante el período}}{\text{Total de individuos con parentesco X durante el período}}$$

genético involucrado. Para ello se han desarrollado varios métodos en los últimos 20 años, gracias a las posibilidades que ofrecen las nuevas técnicas de biología molecular, las computadoras y los complejos algoritmos estadísticos. Seguidamente se describen las metodologías más comúnmente em-

pleadas, con ejemplos tomados de estudios publicados recientemente.

Estudios en gemelos

Clásicamente, el método de estudio en gemelos se ha utilizado para deter-

minar si los factores genéticos desempeñan un papel en el origen de cierta enfermedad. Consiste en comparar la diferencia de concordancia entre gemelos idénticos o monocigóticos (MC) y gemelos fraternales o dicigóticos (DC). Los gemelos MC comparten 100% de su material genético, mientras que los DC comparten, en promedio, 50% de sus genes. Si se estudia una serie de gemelos y se encuentra que los MC son concordantes con mayor frecuencia (por ejemplo, ambos tienen la enfermedad) que los DC, puede concluirse que los factores genéticos están por lo menos parcialmente involucrados en el origen de dicha enfermedad (13). Es importante subrayar, sin embargo, que puede haber diferencias genéticas entre gemelos MC. Estos pueden, por ejemplo, diferir en su repertorio de anticuerpos y receptores de células T, en el número de moléculas de ácido desoxirribonucleico (ADN) mitocondrial, en mutaciones somáticas en general, y en el patrón de inactivación del cromosoma X en el caso de pares femeninos (14). Es bien conocido, además, que los gemelos MC pueden diferenciarse de los DC como consecuencia de factores ambientales.

En estudios de gemelos suele efectuarse uno de dos cálculos de acuerdo al mecanismo de selección de los gemelos: 1) *tasa de concordancia por pares*, que describe la proporción de pares de gemelos en los cuales ambos miembros están afectados, y 2) *tasa de concordancia del caso índice*, determinada por la proporción de individuos afectados entre los cogemelos de los que fueron seleccionados como caso índice. Si bien la tasa de concordancia por pares es el método más sencillo para determinar si los genes ejercen un efecto en determinado fenotipo, ella no mide la magnitud de tal efecto. Para ello es preferible usar la tasa de concordancia del caso índice.

Los estudios de gemelos están limitados por numerosos factores, especialmente por aquellos relacionados con la manera en que los participantes fueron seleccionados para el estudio. Por ejemplo, en estudios basados exclusivamente en voluntarios, se ha observado una mayor proporción de

gemelos MC, pares de sexo femenino y participantes concordantes para el fenotipo en estudio. Dichas diferencias pueden sesgar el cálculo de la tasa de concordancia, motivo por el cual varios países, entre los que se destaca Suecia, han instituido registros de gemelos a nivel poblacional. Otra limitación, especialmente en el estudio de aspectos de la conducta, es que los gemelos MC tienden a compartir factores ambientales con más frecuencia que los DC.

Estudios sobre la interacción entre el gen y el medio ambiente

La existencia de interacciones entre los factores genéticos y los factores ambientales ha sido descrita ampliamente en la última mitad del siglo. La fenilcetonuria es un ejemplo clásico. Este trastorno metabólico recesivo se manifiesta solo en individuos homocigóticos para la mutación que han sido expuestos a fenilalanina (aminoácido presente en la leche y otros alimentos). El xeroderma pigmentosum provee

otro ejemplo: los individuos que lo padecen aumentan su riesgo de padecer cáncer de piel cuando se exponen a rayos ultravioleta. Ottman (15) ha resumido otros ejemplos similares.

Como consecuencia de los avances del Proyecto del Genoma Humano, el método de casos y controles está utilizándose frecuentemente para describir las posibles interacciones entre los factores genéticos y los ambientales. Como puede apreciarse en los cuadros 4 y 5, el tabaquismo materno durante el primer trimestre de la gestación interactúa con el fenotipo fetal (alelo A2 del marcador genético denominado *transforming growth factor-alpha* [TGF α]) en la génesis de fisuras palatinas no sindrómicas [*odds ratio* o razón de posibilidades obtenida en un estudio de casos y controles (RP_{cc}) = 5,5 (IC95%: 2,1 a 14,6)]. Este hallazgo ha sido posteriormente confirmado por Shaw, et al. (16) para fisuras labiopalatinas y palatinas aisladas. Khoury y Flanders (17) recientemente han descrito el estudio de casos únicamente como metodología alternativa a la de casos y controles. En el estudio exclu-

CUADRO 4. Esquema para el análisis de la interacción entre un gen y el medio ambiente en el contexto de un estudio de casos y controles

Exposición ambiental ^a	Susceptibilidad genética	Casos	Controles	Razón de posibilidades
-	-	a	b	1,0
-	+	c	d	RP _g = bc/ad
+	-	e	f	RP _e = be/af
+	+	g	h	RP _{ge} = bg/ah

Fuente: Referencia 17.

^a -: ausente, +: presente; interacción bajo modelo aditivo: RP_{ge} = RP_g + RP_e; interacción bajo modelo multiplicativo: RP_{ge} = RP_g × RP_e.

CUADRO 5. Interacción entre el fenotipo fetal con TGF α y el tabaquismo materno en relación con las fisuras palatinas

Tabaquismo materno	Fenotipo TGF α (alelo A2)	Casos	Controles	Razón de posibilidades (IC95%)
No	No	36	167	1,0
No	Sí	7	34	1,0 (0,3 a 2,4)
Sí	No	13	69	0,9 (0,4 a 1,8)
Sí	Sí	13	11	5,5 (2,1 a 14,6)

Fuente: Referencia 7.

sivo de casos, la tabla de 2×2 se rediseña tal como se ilustra en los cuadros 6 y 7. Como puede apreciarse, la razón de posibilidades obtenida en un estudio de casos únicamente (RP_{cu}) es similar a la RP_{cc} . A pesar de que ambos métodos son estadísticamente potentes y relativamente sencillos de llevar a cabo, la interpretación de sus resultados no es tan fácil. La existencia de una interacción entre el gen y el medio ambiente es, como tal, una asociación estadística, y esta no es necesariamente causal. Sin embargo, es importante recalcar que estos métodos constituyen herramientas útiles en el análisis de interacciones genético-ambientales, ya que identifican factores que podrían resultar importantes para la prevención del trastorno en estudio.

Análisis complejo de segregación

La técnica de análisis complejo de segregación es útil para determinar si cierto fenotipo (representado por una variable continua o discreta) tiene un patrón de transmisión mendeliano en un grupo de genealogías (1). El algoritmo utilizado provee una estimación probabilística de varios factores gené-

CUADRO 6. Esquema para la evaluación de la interacción entre factores genéticos y ambientales en el contexto de un estudio de casos únicamente

Exposición	Genotipo susceptible	
	No	Sí
No	a	b
Sí	c	d

Fuente: Referencia 17.

CUADRO 7. Reanálisis de los datos de Hwang, et al. (7) en el contexto de un estudio de casos únicamente

Tabaquismo materno	Fenotipo con TGF α (alelo A2)	
	No	Sí
No	36	7
Sí	13	13 ^a

Fuente: Referencia 17.

^a RP_{cu} : ad/bc : $(36 \times 13)/(7 \times 13) = 5,14$ (IC95%: 1,68 a 15,71)

cos: probabilidades de transmisión, frecuencias genéticas y parámetros de penetrancia, para los modelos mendelianos; heredabilidad, promedios muestrales y varianzas, para los modelos poligénicos; y ambos tipos de parámetros para lo que se denomina el modelo mixto (18). Por ejemplo, Newman, et al. (19) demostraron que el grado de agregación familiar del cáncer de mama en 1 759 familias era consecutivo con una herencia autosómica dominante, como consecuencia de la acción de un alelo infrecuente (0,06%) al que se atribuyen 4% de los casos afectados (pero 20% de los pares de madres e hijas afectadas), en un contexto mayor de causa multifactorial. Otros ejemplos de fenotipos estudiados usando esta técnica son el asma y la atopia (20), la obesidad (21), las apolipoproteínas plasmáticas (22, 23), la dislexia (24) y las fisuras labiopalatinas (25).³ La principal limitación de este método es su sensibilidad frente al mecanismo de selección de los individuos. Si la selección está sesgada, como suele suceder cuando los casos provienen de una clínica, los resultados tienden a ser espurios. Además, los estudios de segregación son largos y costosos.

Los métodos ya descritos permiten determinar la importancia relativa de los factores genéticos en una enfermedad o fenotipo, pero no identifican el factor causal específico. Para la identificación de genes posiblemente involucrados en el origen de las enfermedades se utilizan las técnicas de "clonación posicional", que incluyen el análisis de asociaciones alélicas y el análisis de ligamiento.

Estudios de asociaciones alélicas

El principal objetivo de los estudios de asociación alélica es comparar la frecuencia de diferentes factores de riesgo en un grupo de individuos afectados por la enfermedad de interés y en un grupo control (27). La evalua-

ción de los factores de riesgo puede incluir exposiciones ambientales o características genéticas. Estas últimas pueden ser tanto productos genéticos, como proteínas o enzimas, o marcadores genéticos basados en secuencias de ADN. Estos últimos, denominados fragmentos de restricción de longitud variable (RFLP, por el inglés *restriction fragment length polymorphisms*), se obtienen usando enzimas de restricción, que cortan el ADN según secuencias específicas. En los últimos años se ha desarrollado otro tipo de marcador genético, los llamados microsatélites, que son, en la mayor parte de los casos, más informativos genéticamente que los RFLP tradicionales (8).

El análisis estadístico en un estudio de asociación es sencillo y puede resumirse en una tabla de 2×2 . El aspecto más difícil, como en la mayor parte de los estudios de casos y controles, consiste en la selección de los controles. Las asociaciones alélicas han ayudado a comprender mejor y a diagnosticar más precozmente ciertas enfermedades autoinmunitarias. El alelo HLA-B27, por ejemplo, ocurre en 90% de los pacientes con espondilitis anquilosante, pero solo en 9% de la población en general (28). Se han descrito otras asociaciones con alelos HLA para diabetes tipo I, artritis reumatoidea, esclerosis múltiple, enfermedad celíaca y lupus eritematoso sistémico (29). Recientemente se describieron asociaciones entre la enzima convertidora de la angiotensina I (ECA) y las enfermedades cardiovasculares (30), entre el angiotensinógeno (AGT) y la hipertensión (31), entre la apolipoproteína E (APOE) y la enfermedad de Alzheimer (32), y entre el gen de la insulina (*INS*) y la diabetes I (33).

La interpretación de una asociación positiva no es trivial. Las asociaciones pueden surgir por tres razones, una de las cuales es completamente artificial (34):

- 1) El alelo en cuestión es realmente la causa del fenotipo.
- 2) El alelo no causa el fenotipo, pero está en *desequilibrio de ligamiento* con otro alelo que sí lo causa. El *desequilibrio de ligamiento* ocurre

³ Varios programas de computadora accesibles en el Internet permiten realizar los cálculos de segregación (véase la referencia 26).

cuando el alelo causante del fenotipo está físicamente cercano (o "ligado") al alelo en estudio. Suele observarse en poblaciones jóvenes, típicamente aisladas (la población finlandesa es un buen ejemplo de un grupo estable en que estudios de asociaciones alélicas suelen arrojar resultados positivos).

- 3) Existe una mezcla poblacional. En una población mezclada, cualquier fenotipo común a un grupo étnico resultará en una asociación positiva con cualquier alelo que también sea más fuerte en dicho grupo étnico. Lander y Schork (34) presentan un pintoresco ejemplo de asociación debida a mezcla poblacional:

'Supongamos que se decide estudiar el fenotipo descrito como habilidad para comer con palitos chinos en la población de San Francisco, a través de un estudio de asociación con el complejo HLA. El alelo HLA-A1 mostraría una asociación positiva con esta habilidad para comer con palitos chinos, no porque ciertos determinantes inmunológicos desempeñen un papel importante en la función manual, sino porque simplemente el alelo HLA-A1 es más común entre asiáticos que entre caucásicos'.

Por ello, el estudio de poblaciones relativamente homogéneas permite evitar este tipo de asociaciones espurias.

En los últimos años se han desarrollado técnicas analíticas que no se ven afectadas por la estructura de la población de interés (35). Una de ellas es la prueba de desequilibrio de transmisión (PDT) (36).⁴ El ejemplo hipotético es un marcador genético con dos alelos M_1 y M_2 , de manera que las combinaciones posibles son M_1M_1 , M_1M_2 (o M_2M_1) y M_2M_2 (cuadro 8). Se selecciona un grupo de casos con el fenotipo de interés y se determina el genotipo de estos casos y de sus progenitores. Posteriormente se evalúa el número de veces que se ha transmitido

el alelo M_1 o el alelo M_2 a cada caso afectado.

Las familias pueden ser tríadas (el individuo afectado y sus progenitores) o ser más complejas (varios miembros afectados por familia más sus progenitores). El método es estadísticamente sólido, aun en casos de mezcla poblacional. La hipótesis que pone a prueba la PDT es que el marcador y el fenotipo no están ligados genéticamente. La teoría empleada deriva del método de Neyman-Pearson (38) y solo utiliza las observaciones b y c (cuadro 8) de padres heterocigóticos (M_1M_2). La fórmula $(b - c)^2 / (b + c)$ permite determinar si existe un número igual de transmisiones de M_1 y M_2 de padres heterocigóticos a sus hijos afectados. De existir ligamiento entre el marcador y el fenotipo, así como asociación alélica, b y c tenderán a ser diferentes. La prueba de significación estadística de la PDT es la X^2 (prueba asintótica de McNemar) o prueba binomial exacta (36); una diferencia significativa confirma que el marcador está ligado al locus del fenotipo. La PDT puede ser utilizada con marcadores genéticos de más de dos alelos e incorporar covariables (39, 40). Es importante señalar que, cuando se aplica la PDT a familias con un fenotipo recurrente (las llamadas "familias múltiples" porque poseen más de un miembro afectado), la evaluación de que existe una asociación alélica no es válida. Esto se debe a que la prueba X^2 supone que las observaciones son independientes, y ello no ocurre cuando los individuos participantes están emparentados. La prueba de ligamiento, sin embargo, es válida aun en estas condiciones (35).

Análisis de ligamiento

El análisis de ligamiento es un recurso valioso para identificar genes que podrían tener asociación causal con el fenotipo de interés, ya que permite evaluar si los loci en un cromosoma son transmitidos juntos más de lo esperado durante la meiosis. Las pruebas estadísticas de ligamiento estiman la fracción de recombinación (ϕ) entre dos loci. Si $\phi = 0$, entonces hay

CUADRO 8. Combinaciones de alelos M_1 y M_2 transmitidos y no transmitidos entre padres ($2n$) de casos afectados (n)

Alelo transmitido	Alelo no transmitido		Total
	M_1	M_2	
M_1	a	b	a + b
M_2	c	d	c + d
Total	a + c	b + d	2n

Fuente: Referencia 36.

ligamiento completo, lo que implica que los alelos en los dos loci siempre se transmiten juntos. El hallazgo de ligamiento positivo entre el locus de un marcador genético (conocido) y el locus del fenotipo de interés (desconocido) permite al investigador determinar la ubicación cromosómica del locus que produce el segundo. De esta manera se han identificado los genes causales de más de 60 trastornos mendelianos y la lista crece diariamente (41).

Si $\phi = 0,50$, entonces no hay ligamiento, es decir, cada uno de los loci se transmite independientemente del otro (como sucede con los loci ubicados en distintos cromosomas). El valor ϕ también sirve como medida de la distancia física entre los dos loci; cuanto mayor es ϕ , más lejos se encuentra un locus del otro. Si se consideran más de dos loci, el análisis de ligamiento puede también utilizarse para establecer el orden de los loci en un cromosoma.

El ligamiento entre dos loci no es un fenómeno en sí, sino una hipótesis que debe demostrarse con una prueba estadística. Para ello se utiliza el método de verosimilitud máxima (*maximum likelihood*) (42). Así, la verosimilitud de una hipótesis, denominada $V(H)$, es proporcional a la probabilidad de la observación experimental bajo dicha hipótesis, $\text{Prob}(O|H)$. En este caso, la hipótesis es ϕ (ligamiento o no ligamiento); por lo tanto, la verosimilitud máxima se expresa como $V(O|\phi)$. Las verosimilitudes relativas de las dos hipótesis (ligamiento ó $\phi < 0,5$ y no ligamiento ó $\phi = 0,5$) se evalúan por medio del cociente de verosimilitud, $CV = V(O|\phi < 0,5) / V(O|\phi = 0,5)$. Para obtener un valor de significación, se transforma logarítmica-

⁴ Se remite a los lectores interesados en otros métodos a leer el trabajo de Thomson sobre el método de riesgo relativo por haplotipo (37).

mente el CV en lo que se denomina un *LOD score* (logaritmo del cociente de verosimilitud del ligamiento, representado por Z). Expresado algebraicamente, $Z = \log_{10}(CV)$.⁵

Es costumbre indicar los *LOD scores* para los diferentes valores de ϕ en una tabla (45). Cuando $\phi = 0,5$, Z es siempre 0, ya que se dividen dos probabilidades idénticas, y el $\log_{10}(1) = 0$. Para fracciones de recombinación menores de 0,5, los *LOD scores* de referencia son 3,00 y -2,00. Un *LOD score* $\geq 3,00$ ($P = 10^{-4}$) es prueba a favor de ligamiento, en tanto que un *LOD score* $< -2,00$ rechaza la hipótesis de ligamiento. Recientemente, Lander y Kruglyak (46) propusieron que se considere un ligamiento significativo a partir de un *LOD score* $\geq 3,3$ ($P = 5 \times 10^{-5}$).

El análisis de ligamiento puede extenderse a sistemas más complejos. Por ejemplo, el análisis de ligamiento multipuntual permite evaluar simultáneamente múltiples marcadores genéticos ubicados en un mismo cromosoma. Como consecuencia de la creciente identificación de los marcadores genéticos disponibles en cada cromosoma, el análisis multipuntual se ha transformado en una técnica de elección para la ubicación precisa de los genes. Dado que esta técnica implica analizar una gran cantidad de marcadores en un mismo cromosoma, los investigadores suelen aplicarla solo después de que cierta región cromosómica ha mostrado indicios de ligamiento.

El análisis de ligamiento también puede llevarse a cabo cuando el fenotipo de interés presenta heterogeneidad genética o cuando es consecuencia de la interacción de dos o más genes. En el primer caso, existe más de un gen que actúa independientemente en las génesis del fenotipo. Por ejemplo, el cáncer de mama hereditario es atribuible en algunas familias a mutaciones del gen *BRCA1*; en otras, se debe a mutaciones del gen *BRCA2*; finalmente, en algunas familias la causa radica en mutaciones de genes aún no

identificados. Fenotipos producidos, al menos parcialmente, por la interacción sinérgica de dos o más genes son el de la esclerosis múltiple (47) y el de los niveles totales de la inmunoglobulina E en suero (48).

Análisis de alelos compartidos

El método de análisis de ligamiento que se ha descrito es sumamente sensible a errores en los modelos de transmisión hereditaria aplicados para explicar el fenotipo en estudio y a variaciones en los valores de frecuencia alélica poblacionales atribuidos a las familias evaluadas. Por ello, se han desarrollado técnicas analíticas que no requieren la imposición de modelos, basadas en la comparación de alelos compartidos entre parientes. Una de ellas, el análisis de pares de hermanos afectados, evalúa cuán frecuentemente una copia determinada de una región cromosómica es compartida idénticamente por descendencia (IPD), es decir, porque proviene de un ancestro común en la familia. Por ejemplo, dos hermanos pueden compartir IPD ninguna, una, o dos copias de cualquier *locus* (con una distribución esperada de 25%, 50% y 25% respectivamente, de existir segregación de alelos al azar). La prueba estadística compara el promedio de alelos compartidos IPD (π) con el promedio esperado (50%). Los resultados se presentan en forma de valores *P*, de *LOD scores*, o de *Z scores* (número de desviaciones estándar en que π excede la expectativa de 50%). Por ejemplo, 100 pares de hermanos que comparten 61% de los alelos en un sector del genoma corresponde a $P = 0,001$, a un *LOD score* de 2,1 y a un *Z score* de 3,1 (46). Según estos autores se obtiene prueba de ligamiento con el método de pares de hermanos cuando el *LOD score* es igual a 3,6 o mayor ($P \geq 2,2 \times 10^{-5}$).

El método de pares de hermanos afectados ha sido empleado con buenos resultados en la ubicación cromosómica de numerosos fenotipos, tales como los de diabetes tipo I, hipertensión esencial, concentraciones de inmunoglobulina E sérica y densidad

ósea en mujeres postmenopáusicas (34). Si bien es más sólido que el análisis de ligamiento, el análisis de pares de hermanos afectados se ve limitado por el elevado número de pares requeridos para llevar a cabo la estadística (en el orden de cientos o miles de pares de hermanos).

COMENTARIOS FINALES

Perspectivas laborales y académicas

La epidemiología genética es una disciplina en plena expansión. Numerosas instituciones académicas y gubernamentales, especialmente en los Estados Unidos de América, Francia e Inglaterra, ofrecen programas educativos y de investigación en epidemiología genética. Las posibilidades laborales para los epidemiólogos genetistas son excelentes, especialmente en los países muy industrializados.

El Proyecto Internacional del Genoma Humano ha estimulado gran interés y controversia. Su principal meta es obtener una descripción completa del genoma humano por medio del análisis secuencial del ADN (49).⁶ El Centro Nacional de Información Biotecnológica de los Estados Unidos comunicó, en octubre de 1996, que estaban ubicados cerca de 16 500 genes, correspondientes a aproximadamente 20% de todos los genes humanos (50). Se estima que el Proyecto terminará en el año 2005, cuando se establezca la secuencia de los 3 000 millones de nucleótidos del ADN humano. En este contexto, una función que compete a los especialistas en epidemiología genética es educar al resto de la comunidad científica, y especialmente a la no científica, acerca de las consecuencias y los alcances del

⁵ Los *LOD scores* pueden calcularse usando programas disponibles para computadoras personales y redes (43) o consultando el Internet (44).

⁶ La página del Centro Nacional de Investigación del Genoma Humano de los Estados Unidos es accesible en el sitio de Internet <http://www.nhgri.nih.gov>; la Base de Datos del Genoma (*Genome Database*), una de las principales bases de datos de genes localizados, puede visitarse en el sitio <http://gdbwww.gdb.org>; el mapa genético integrado puede verse en el sitio <http://www.ncbi.nlm.nih.gov/SCIENCE96/>; y el catálogo de genes y defectos congénitos se encuentra en el sitio <http://www3.ncbi.nlm.nih.gov/omim/>.

Proyecto Internacional del Genoma Humano.

Aspectos éticos, jurídicos y sociales

Desde sus comienzos, los planificadores del Proyecto Internacional del Genoma Humano reconocieron que la identificación de genes tendría profundas consecuencias para el individuo, la familia y la sociedad. Surgieron interrogantes, tales como ¿De qué manera debe interpretarse y utilizarse la información genética?, ¿Quién debe tener acceso a ella? ¿Cómo puede protegerse a los individuos de posibles perjuicios? y ¿Cuál es el beneficio de ofrecer un estudio genético cuando es poco, o nada, lo que puede recomendarse en cuanto a cura o prevención?

Ya se han identificado genes que causan, al menos parcialmente, numerosas enfermedades. A pesar de que la detección y el diagnóstico de las mismas es cada vez más precoz y certero, el objetivo a largo plazo del Proyecto Internacional del Genoma Humano es mejorar su tratamiento, prevenirlas, y a la larga curarlas. El período intermedio, en que la detección temprana es posible, el conocimiento es limitado, y el tratamiento no está disponible, se caracteriza por ser el que conlleva los

conflictos éticos, legales y sociales más importantes.

El Instituto Nacional de Investigación del Genoma Humano de los Estados Unidos es sede, desde 1989, del Grupo de Trabajo sobre las Implicaciones Éticas, Legales y Sociales del Proyecto del Genoma Humano. Multidisciplinario e interinstitucional, este grupo está interesado en las siguientes cuatro esferas (51, 52):

- 1) Tratamiento confidencial y justicia en el uso y la interpretación de la información genética. Su objetivo es evaluar los mecanismos para prevenir la discriminación y estigmatización producidas por el abuso (y las confusiones) en el empleo de esta información.
- 2) Integración de tecnologías genéticas a la actividad clínica. En esta esfera se investigan las consecuencias de la disponibilidad de exámenes genéticos en la práctica médica y los mecanismos para su evaluación.
- 3) Metodología de la investigación en genética. Su principal aspecto consiste en evaluar cómo se informa a los posibles voluntarios sobre los riesgos y beneficios de participar en un estudio de investigación y cómo se obtiene el correspondiente consentimiento.

- 4) Educación para la comunidad y los profesionales médicos acerca de los alcances del Proyecto del Genoma Humano.

La epidemiología genética en el mundo actual

La Sociedad Internacional de Epidemiología Genética,⁷ que posee más de 400 miembros y que se encuentra en plena expansión, edita la revista mensual *Genetic Epidemiology* y organiza congresos internacionales anuales. En los países de América Latina aún no se ha creado un medio común para el desarrollo de la epidemiología genética. Las iniciativas en este sentido suelen ser individuales y esporádicas. Si bien muchos países latinoamericanos experimentan profundas transformaciones sociales y políticas, es necesario que sus instituciones académicas y científicas den cabida y apoyo a nuevas disciplinas, tales como la epidemiología genética, sin descuidar las ya existentes. Solo de esa manera los países de América Latina podrán integrarse al concierto de las naciones científicamente avanzadas.

⁷ Búsquese en el sitio de Internet <http://darwin.mhmc.cwru.edu/IGES/index.html>.

REFERENCIAS

1. Khoury MJ, Beaty TH, Cohen BH. *Fundamentals of genetic epidemiology*. New York: Oxford University Press; 1993.
2. Neel JV, Schull WJ. *Human heredity*. Chicago: The University of Chicago Press; 1954:283-306.
3. Vogel F, Motulsky AG. *Human genetics, problems and approaches*. 2a ed. Berlin: Springer-Verlag; 1986.
4. Introduction. En: Morton NE, Chung CS. *Genetic epidemiology*. New York: Academic Press; 1978:3-11.
5. Cohen BH. Chronic obstructive pulmonary disease: a challenge in genetic epidemiology. *Am J Epidemiol* 1980;112:274-288.
6. Introduction. En: Gordis L. *Epidemiology*. Philadelphia: W.B. Saunders Company; 1996:5-6.
7. Hwang SJ, Beaty TH, Panny Sr, Street NA, Joseph JM, Gordon S, et al. Association study of transforming growth factor alpha (TGF α) TaqI polymorphism and oral clefts: indication of gene-environment interaction in a population-based sample of infants with birth defects. *Am J Epidemiol* 1995;141:629-636.
8. Mutation and instability of human DNA. En: Strachan T, Read AP. *Human molecular genetics*. New York: Wiley-Liss; 1996:259-261.
9. King MC, Lee GM, Spinner NB, Thomson G, Wrensch MR. Genetic epidemiology. *Am Rev Public Health* 1984;5:1-52.
10. Mettlin C, Corghan J, Natarajan N, Lane W. The association of age and familial risk in a case-control study of breast cancer. *Am J Epidemiol* 1990;131:973-986.
11. Slater E, Cowie V. *Genetics of mental disorders*. Oxford: Oxford University Press; 1970.
12. Elston RC. Segregation analysis. *Adv Hum Genet* 1981;11:63-120.
13. Susser M. Separating heredity and environment. *Am J Prev Med* 1985;1:5-23.
14. Ollier WER, MacGregor A. Genetic epidemiology of rheumatoid disease. *Br Med Bull* 1995;51:267-285.
15. Ottman R. An epidemiologic approach to gene-environment interaction. *Genet Epidemiol* 1990;7:177-186.
16. Shaw GM, Wasserman CR, Lammer EJ, O'Malley CD, Murray JC, Basart AM, et al. Orofacial clefts, parental cigarette smoking, and transforming growth factor-alpha gene variants. *Am J Hum Genet* 1996;58:551-561.
17. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! *Am J Epidemiol* 1996; 144:207-213.
18. Morton NE, MacLean CJ. Analysis of family resemblance. III. Complex segregation analysis of quantitative traits. *Am J Hum Genet* 1974; 26:489-503.

19. Newman B, Austin MA, Lee M, King MC. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc Natl Acad Sciences* 1988;85:3044-3048.
20. Panhuysen CIM, Meyers DA, Postma DS, Levitt RC, Bleecker ER. The genetics of asthma and atopy. *Allergy* 1995;50:863-869.
21. Bouchard C. The genetics of obesity: from genetic epidemiology to molecular markers. *Molec Med Today* 1995;45-50.
22. Moll PP, Michels VV, Weidman WH, Kottke BA. Genetic determination of plasma apolipoprotein AI in a population-based sample. *Am J Hum Genet* 1989;44:124-139.
23. Prenger VL, Beaty TH, Kwiterovich PO. Genetic determination of high-density lipoprotein cholesterol and apolipoprotein A-I plasma levels in a family study of cardiac catheterization patients. *Am J Hum Genet* 1992;51:1047-1057.
24. Pennington BF, Gilger JW, Pauls D, Smith SA, Smith SD, DeFries JC. Evidence for major gene transmission of developmental dyslexia. *J Am Med Assoc* 1991;266:1527-1534.
25. Wyszynski DF, Beaty TH, Maestri NE. Genetics of non-syndromic oral clefts revisited. *Cleft Palate Craniof J* 1996;33:406-417.
26. Statistical analysis for genetic epidemiology. Disponible en: <http://darwin.mhmc.cwru.edu/pub/sage.html>. Acceso el 25 de septiembre de 1997.
27. Hwang S-J, Beaty TH, Liang K-Y, Coresh J, Khoury MJ. Minimum sample size estimation to detect gene-environment interaction in case-control designs. *Am J Epidemiol* 1994;140:1029-1037.
28. Ryder LP, Andersen E, Svejgaard A. *HLA and disease registry: third report*. Copenhagen: Munksgaard; 1979.
29. Braun WE. *HLA and disease*. Boca Raton: Chemical Rubber Company Press; 1979.
30. Tiret L, Rigat B, Visvikis S, Breda C, Corvol P, Cambien F, et al. Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. *Am J Hum Genet* 1992;51:197-205.
31. Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RR, Williams CS, Charry A, et al. Molecular basis of human hypertension: role of angiotensinogen. *Cell* 1992;71:169-180.
32. Pericak-Vance MA, Haines JL. Genetic susceptibility to Alzheimer disease. *Trends in genetics* 1995;11:504-508.
33. Bain SC, Prins JB, Hearne CM, Rodriguez NR, Rowe BR, Pritchard LE, et al. Insulin gene region-encoded susceptibility to type 1 diabetes is not restricted to HLA-DR4-positive individuals. *Nature Genet* 1992;2:212-215.
34. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994;265:2037-2048.
35. Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996;59:983-989.
36. Spielman RS, McGinnis RE, Ewens WJ. Transmission disequilibrium test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52:506-516.
37. Thomson G. Mapping disease genes: family-based association studies. *Am J Hum Genet* 1995;57:487-498.
38. Kendall MG, Stuart A. The advanced theory of statistics, 4a ed. *Vol 2: Inference and relationship*. London: Griffin; 1979.
39. Duffy DL. Screening a 2 cM genetic map for allelic association: a simulated oligogenic trait. *Genet Epidemiol* 1995;12:595-600.
40. Bickeböller H, Clerget-Darpoux F. Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers. *Genet Epidemiol* 1995;12:865-870.
41. McKusick VA. History of medical genetics. En: Rimoin DL, Connor JM, Pyeritz RE, eds. *Emery and Rimoin's principles and practice of medical genetics*, 3a ed. New York: Churchill Livingstone; 1996.
42. Edwards AWF. *Likelihood*. Cambridge: Cambridge University Press; 1972.
43. Terwilliger J, Ott J. *Handbook for human genetic linkage*. Baltimore: Johns Hopkins University Press; 1994.
44. An alphabetic list of genetic analysis software. Disponible en: <http://linkage.rockefeller.edu/soft/list.html>. Acceso el 25 de septiembre de 1997.
45. Morton NE. Sequential tests for the detection of linkage. *Am J Hum Genet* 1955;7:277-318.
46. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet* 1995;11:241-247.
47. Tienari PJ, Terwilliger JD, Ott J, Peltonen L. Two-locus linkage analysis in multiple sclerosis (MS). *Genomics* 1994;19:320-325.
48. Xu J, Levitt RC, Panhuysen CIM, Postma DS, Taylor EW, Amelung PJ, et al. Evidence for two unlinked loci regulating total serum IgE levels. *Am J Human Genet* 1995;57:425-430.
49. Engel LW. The human genome project: history, goals, and progress. *Arch Path Lab Med* 1993;117:459-465.
50. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, et al. A gene map of the human genome. *Science* 1996;274:540-546.
51. Pembrey ME, Anionwu EN. Ethical aspects of genetic screening and diagnosis. En: Rimoin DL, Connor JM, Pyeritz RE, eds. *Emery and Rimoin's principles and practice of medical genetics*, 3a. ed. New York: Churchill Livingstone; 1996.
52. Human Genome Project information. Ethical, legal, and social issues (ELSI). Disponible en: http://www.ornl.gov/TechResources/Human_Genome/resource/elsi.html. Acceso el 25 de septiembre de 1997.

Manuscrito recibido el 22 de abril de 1996 y aceptado para publicación en versión revisada el 5 de febrero de 1997.

ABSTRACT

Genetic epidemiology: an expanding scientific discipline

Genetic epidemiology is a relatively new discipline that studies the interaction between genetic and environmental factors in the production of human diseases. Taking advantage of the genetic markers provided by molecular biological research, complex computerized algorithms, and large databases, the field of genetic epidemiology has undergone notable development in the past 10 years. This article describes the objectives and methodology of genetic epidemiology, using concrete examples from the recent scientific literature.