

## LECTURAS RECOMENDADAS

- Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010;464:704-712. Describe el grupo más actualizado de CNV mayores de 443 pares de bases detectados mediante el revestimiento de micromatrices de oligonucleótidos con 42 millones de sondas.
- Dietz HC. New therapeutic approaches to Mendelian disorders. *N Engl J Med*. 2010;363:852-863. Revisión.
- Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. Complete genome sequencing reveals SH3TC2 mutations causing CMT1 neuropathy. *N Engl J Med*. 2010;362:1181-1191. Primer uso de la secuenciación hologenómica personal para identificar las bases genéticas de la enfermedad en un paciente individual.
- Sheehan NA, Meng S, Didelez V. Mendelian randomization: a tool for assessing causality in observational epidemiology. *Methods Mol Biol*. 2011;713:153-166. Métodos para determinar la relevancia independiente de una variante genética basándose en datos epidemiológicos.
- Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437-455. Revisa el papel de los CNV en la enfermedad.

## 41

## BASES HEREDITARIAS DE LAS ENFERMEDADES COMUNES

DAVID ALTSHULER

Una cuestión central en medicina es comprender por qué algunas personas enferman y otras no. La búsqueda de respuestas a este dilema se debe a varios motivos: para ofrecer explicaciones a los pacientes, mejorar la capacidad de predecir el riesgo de sufrir enfermedades y, lo que es más relevante, para comprender la fisiopatología como fundamento para el diseño de estrategias preventivas y terapéuticas racionales. En algunos casos, una única exposición ambiental desempeña un papel principal en la enfermedad (p. ej., el tabaco y el cáncer de pulmón o el virus de la inmunodeficiencia humana [VIH] y el síndrome de inmunodeficiencia adquirida [SIDA]). En otros casos, como el de la enfermedad de Huntington o la fibrosis quística, la mutación de un único gen es necesaria y puede ser suficiente para causar enfermedad. Estas respuestas singulares son la excepción y en la mayoría de los casos la enfermedad no es atribuible ni a un único factor ambiental dominante ni a la mutación de un único gen. Por el contrario, la mayoría de los casos de enfermedad se deben a la acción combinada de alteraciones innatas y adquiridas de la secuencia de los genes, de exposiciones ambientales y conductuales, así como de la mala suerte. Esas enfermedades, que constituyen la mayor parte de la morbimortalidad en la población, se denominan *rasgos complejos*.

La genética humana es una herramienta única para generar nuevas hipótesis sobre las causas básicas de la enfermedad, basada en búsquedas hologenómicas en la población humana que no se ven limitadas por las suposiciones previas sobre los procesos fisiopatológicos subyacentes. Debido a que en la actualidad se conocen la secuencia del genoma humano y gran parte de su variación habitual, y gracias a las herramientas y métodos novedosos para determinar directamente las secuencias genómicas de las personas, estamos entrando en una era en la que la medicina puede obtener información del conocimiento de los genes y variantes específicas que contribuyen al riesgo de las enfermedades humanas comunes.

## HEREDABILIDAD: VARIACIÓN HEREDITARIA DEL RIESGO DE ENFERMEDAD

La susceptibilidad a la enfermedad varía en el seno de una población humana y entre distintas poblaciones. Los estudios de *agregación familiar* pueden determinar el grado en el que la herencia contribuye a estos patrones. Estos estudios son simples en su concepción y se preguntan si los miembros de una misma familia tienen tasas más parecidas de enfermedad que las personas elegidas al azar en la población. El agrupamiento familiar puede reflejar no sólo los genes compartidos sino también el ambiente común. La contribución del genotipo compartido se puede diseccionar aún más comparando las tasas de enfermedad entre familias como una función de la extensión del parentesco genético. El diseño más claro implica comparar la concordancia de enfermedad entre parejas de gemelos dicigóticos y monogigóticos. Para enfermedades comunes, como la diabetes mellitus tipo 1 y 2, la obesidad, la hipertensión, la arteriopatía coronaria, las enfermedades autoinmunitarias, los cánceres habituales, la esquizofrenia y el trastorno bipolar, los estudios de gemelos han demostrado que las tasas de concordancia son significativamente mayores en las parejas de gemelos monogigóticos que en las de dicigóticos. Para muchos otros rasgos de interés clínico (p. ej., la mayoría de las respuestas farmacológicas) todavía no se han realizado análisis formales de heredabilidad, por lo que el papel de la herencia en estas características está peor documentado.

Los datos sobre la agregación familiar hacen posible el cálculo de la *heredabilidad*, o fracción de variabilidad interindividual del riesgo de enfermedad que es atribuible a influencias genéticas aditivas. La variabilidad restante entre individuos se debe a todas

las demás contribuciones: influencia ambiental en la enfermedad, efectos genéticos no aditivos (*epistáticos*), por ejemplo interacciones gen-gen o gen-ambiente, error en la medida de parentesco o enfermedad, así como la casualidad. Para la mayoría de los rasgos clínicos importantes (enfermedades y factores de riesgo), las estimaciones empíricas de heredabilidad varían entre el 20 y el 80% (v. para una información completa, la página de Online Mendelian Inheritance in Man, disponible en <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=OMIM>).

Cuando se interpretan las estimaciones de heredabilidad, se deben tener en cuenta dos factores cruciales: el efecto de los errores de medición y el contexto ambiental. Los errores de medición pueden disminuir la estimación de la heredabilidad de un rasgo. Una única medida de presión arterial es mucho menos heredable que una puntuación compuesta basada en una serie de medidas de presión a lo largo del tiempo. Es decir, la variabilidad día a día y la imprecisión en las medidas clínicas pueden ocultar una susceptibilidad biológica subyacente debida a la herencia. Para el paciente y el médico, esto significa que aunque la presión arterial de un día determinado puede no ser especialmente heredable, la presión arterial a lo largo del tiempo (que es supuestamente el factor de riesgo relevante para la enfermedad vascular) es heredable en un grado mucho mayor.

En segundo lugar, las estimaciones de heredabilidad tienen sentido sólo en el contexto del ambiente en el que se realizó el estudio. En el caso en el que los desencadenantes ambientales de la enfermedad sean relativamente constantes entre la población del estudio, los factores hereditarios pueden explicar gran parte de la variación en las tasas de enfermedad. Por el contrario, si la exposición a causas ambientales de enfermedad varía mucho entre la población del estudio, los factores no genéticos pueden tener mayor peso que la contribución de la susceptibilidad innata. Por ejemplo, la tasa y diversidad de historias de tabaquismo tienen un impacto considerable sobre cuánta variabilidad de las tasas de cáncer de pulmón (en cualquier estudio dado o grupo de la edad del paciente) puede explicarse por la herencia. Si nadie fumase (o si todo el mundo lo hiciera), una pequeña parte de la variación del riesgo de tener cáncer de pulmón se debería al tabaquismo; por el contrario, si la mitad de la población fumase varios paquetes al día y la otra mitad no fumase nada, es indudable que esta conducta predominaría sobre la susceptibilidad innata.

Por estas razones, la heredabilidad no es una característica fija de una enfermedad, sino una evaluación de una población determinada, un grupo de medidas, y el grado en que la variabilidad en la exposición genética y ambiental explica el riesgo de la enfermedad. Esto arroja luz sobre la aparente contradicción entre las tasas de enfermedad que son muy heredables (en una población determinada) y que, sin embargo, varían drásticamente entre poblaciones separadas en el tiempo, por la geografía o por el estatus socioeconómico. En grandes comparaciones entre grupos, la exposición ambiental y los métodos de indagación clínica pueden variar de forma sustancial y contribuir a cambios seculares en los patrones de enfermedad. De forma recíproca, en el seno de un grupo expuesto a un ambiente relativamente uniforme y estudiado de una manera estandarizada, la susceptibilidad genética puede desempeñar un papel principal a la hora de determinar el riesgo individual.

## HETEROCIGOSIS: VARIACIÓN HEREDITARIA EN LA SECUENCIA DEL GENOMA

La *heredabilidad* expresa los patrones de variación hereditaria de las tasas de enfermedad, mientras que la *heterocigosidad* expresa la tasa de variación hereditaria en las secuencias genómicas (tabla 41-1). La heterocigosidad se define como la proporción de lugares del cromosoma en el que dos copias escogidas de forma aleatoria difieren en cuanto a la secuencia de ADN. Debido a que las células son *diploides* (tienen dos copias de la secuencia genómica) y como estas dos copias se seleccionaron de un modo semialeatorio de la población, la heterocigosidad es equivalente a la fracción de pares de bases que varían entre las dos copias que cada persona ha heredado de su padre y de su madre. En otras palabras, la heterocigosidad es la tasa de variación genética en una persona.

Los polimorfismos de nucleótido único (SNP por su acrónimo en inglés) son localizaciones en las que una única letra del código de ADN se ha cambiado por otra alternativa. Estas variantes se observan en alrededor de 1/1.000 posiciones en la secuencia del genoma humano. En las regiones codificantes de proteínas de los genes, las tasas de variación genética son menores (menos de 1/2.000 bases); la tasa de variación que altera de forma sustancial la secuencia de la proteína codificada es aún menor (v. tabla 41-1). Estas tasas pueden comprenderse a la luz de la selección darwiniana frente a los cambios que alteran la secuencia de aminoácidos de las proteínas codificadas.

Los genomas también contienen variación a mayor escala: inserciones y deleciones de nucleótidos, alteración del número de copias de genes y secuencias particulares, así como alteraciones a mayor escala, como inversiones y translocaciones. Tanto los SNP como las alteraciones a mayor escala pueden influir en la función de los genes y contribuir a la aparición de enfermedad.

La variación genética en cada persona se debe en gran medida a variantes comunes. De forma empírica, más del 98% de las localizaciones heterocigóticas en cada persona presentan una frecuencia superior al 1% en la población humana de todo el

**TABLA 41-1 CARACTERÍSTICAS DE LA VARIACIÓN DE LA SECUENCIA DEL GENOMA HUMANO**

Longitud de la secuencia del genoma humano (pares de bases)	3.000.000.000
Número de genes humanos (estimados)	20.000
Fración de pares de bases que difieren entre la secuencia del genoma de un ser humano y el de un chimpancé	1,3% (1/80)
Fración de pares de bases que varían entre las secuencias del genoma de dos seres humanos cualesquiera	0,1% (1/1.000)
Fración de pares de bases de la región codificante que varían de tal manera que alteran la secuencia de la proteína codificada	0,2% (1/5.000)
Número de variantes de secuencia presentes en cada individuo como lugares heterocigotos	3.000.000
Número de variantes que alteran los aminoácidos presentes en cada individuo como lugares heterocigotos	12.000
Número de variantes de secuencia en la población humana con una frecuencia en la población > 1%	10.000.000
Número de polimorfismos de aminoácidos presentes en el genoma humano con una frecuencia en la población > 1%	75.000
Fración de toda la heterocigosis humana atribuible a variantes con una frecuencia > 1%	98%

mundo. Debido a que la mayor parte de la heterocigosidad humana se debe a variantes comunes, se puede elaborar una base de datos que contenga todas las variantes de secuencia comunes (frecuencia superior al 1%) en la población humana mediante la secuenciación de los genomas de tan sólo unos cientos de personas y aún así captar la mayor parte de la variación genética de cualquier individuo.

Como parte de las bases del proyecto genoma humano, se ha creado un catálogo de variantes frecuentes del ADN por una serie de proyectos públicos-privados, como SNP Consortium, International HapMap y 1000 Genomes Projects. En el momento de escribir esta obra, la base de datos pública contiene más de 17 millones de variantes genéticas humanas ([www.ncbi.nlm.nih.gov:80/SNP/index.html](http://www.ncbi.nlm.nih.gov:80/SNP/index.html)). No todas estas entradas representan variantes habituales (algunas son infrecuentes), y algunas constituyen hallazgos técnicos falsos positivos. Sin embargo, la colección existente representa la mayor parte de las variantes comunes en cada persona y ha estimulado los esfuerzos para medir de forma sistemática las variantes genéticas para determinar su contribución a la enfermedad.

El papel destacado de la variación común en la diversidad de la secuencia humana se explica por la historia demográfica única de la población de nuestra especie. A pesar de la distribución global de la población humana actual, ahora está claro que toda procede de una única población que vivió en África hace sólo 10.000-40.000 años. La población ancestral era pequeña (con un tamaño efectivo de unos 10.000 individuos), vivían una existencia de cazadores-recolectores en densidades de población bajas (con relación a otros seres humanos y más tarde a animales domésticos), y habían evolucionado en África durante millones de años. La mayoría de la variación genética surgió en esa fase de la historia humana, antes de las más recientes migraciones, expansiones e invención de tecnologías (p. ej., la agricultura) que llevaron a poblar el globo en toda su extensión. La mayoría de las variaciones genéticas comunes es anterior a la diáspora y es compartida por todas las poblaciones de la Tierra.

Un segundo factor es la lenta tasa de cambio en el ADN humano. Los fenómenos de mutación y recombinación se producen a velocidades muy bajas, del orden de  $10^{-8}$  por par de bases por generación. Aun así, cualquier par de genes humanos permite rastrear un linaje que lleva hasta un antepasado común que vivió hace unas  $10^3$ - $10^4$  generaciones atrás (si una generación corresponde a 20 años,  $10^4$  generaciones equivalen a 200.000 años). Dicho de otro modo, si se considera un nucleótido típico en dos personas no relacionadas, es más probable que se remonte a un antepasado común sin que se haya producido ninguna mutación que el que haya aparecido una mutación en todo ese tiempo. Esto explica por qué el 99,9% de pares de bases son idénticos cuando se comparan dos copias cualesquiera del genoma humano.

Otro aspecto de la variación humana se explica por estas simples relaciones matemáticas y genéticas de la población: el grado de diversidad de la secuencia del ADN humano atribuible a variantes raras y comunes. Cada persona hereda de sus progenitores alguno de los 3 millones de polimorfismos comunes (definidos por lo general como los que tienen una frecuencia superior al 1%). Se heredan variantes comunes que se comparten con personas aparentemente no relacionadas, pero que no alcanzan una frecuencia del 1% o superior. Por último, también se heredan miles de variantes que son específicas de cada persona y de sus familiares más cercanos. La pregunta de cómo estas diferentes clases de variantes influyen en la enfermedad tiene un interés y una relevancia centrales para la genética médica.

La ascendencia compartida de las poblaciones humanas explica otro aspecto de la variación genética en nuestra especie: las correlaciones entre variantes próximas,

conocidas como *desequilibrio de ligamiento* o *haplotipos*. Empíricamente, las personas que portan una variante particular en una posición del genoma tienen más probabilidad de la esperada por el azar de portar un grupo particular de variantes en posiciones próximas en el cromosoma. Es decir, en la población no se observan todas las combinaciones de variantes próximas, sino más bien un pequeño subgrupo de las combinaciones posibles. Estas correlaciones reflejan el hecho antes comentado de que la mayoría de las variantes en nuestros genomas aparecieron una vez en la historia humana (por lo general hace mucho tiempo) y lo hicieron así sobre una copia arbitraria pero única que portaban algunos individuos de la población. La copia ancestral del genoma sobre la que ocurrió la mutación puede reconocerse en la población actual como extensión de alelos particulares (conocida como *haplotipo*) que viajan de la mano en la población. Es decir, aunque la mayoría de las variaciones en nuestro genoma aparecieron antes de la historia escrita de la humanidad, la secuencia de ADN en cada uno de nosotros lleva un registro de la evolución e historia demográfica de la población humana.

Estos haplotipos ancestrales, transmitidos a través de ancestros prehistóricos desde África, se pueden reconocer en la población humana actual. La estructura del haplotipo del genoma humano ofrece una herramienta práctica en los estudios de asociación de enfermedades humanas, porque no es necesario medir directamente cada nucleótido para conseguir gran parte de la información. Estos métodos basados en el haplotipo son la base de los estudios de asociación hologenómicos, que se describen más adelante.

## BÚSQUEDA DE LOS GENES QUE SUBYACEN A ENFERMEDADES MONOGENÉTICAS

El término *arquitectura genética* de una enfermedad se refiere al número y magnitud de los factores de riesgo genéticos que existen en cada paciente y en la población, así como a sus frecuencias e interacciones. Las enfermedades pueden deberse a un solo gen (*monogénicas*) en cada familia o a múltiples genes (*poligénicas*). Lo más fácil es identificar los factores de riesgo genéticos cuando está implicado un solo gen y éste tiene un gran impacto en la enfermedad en una familia dada. En los casos en los que un solo gen es necesario y suficiente para causar la enfermedad, la afección se denomina trastorno *mendeliano* porque la enfermedad se relaciona perfectamente con una mutación (en la familia) que obedece las leyes simples de la herencia de Mendel.

Algunos trastornos monogénicos están causados por el mismo gen en todas las familias afectadas; por ejemplo, la fibrosis quística está causada siempre por mutaciones en *CFTR*. Aunque muchas personas con fibrosis quística portan la misma mutación fundadora (d-508), otras pueden tener cualquier par de una amplia variedad de mutaciones diferentes del gen *CFTR*. La existencia de muchas mutaciones distintas en un gen concreto de enfermedad se denomina *heterogeneidad alélica*.

Un trastorno mendeliano puede deberse a una lesión genética única en cualquier familia concreta, pero en distintas familias puede deberse a mutaciones en diversos genes. Este fenómeno, denominado *heterogeneidad de locus*, queda ilustrado en la retinitis pigmentaria. Aunque la mutación de un único gen es, por lo general, necesaria y suficiente para causar retinitis pigmentaria, hay docenas de genes diferentes en los que se han encontrado mutaciones de la retinitis pigmentaria (Online Mendelian Inheritance in Man #268000). Sin embargo, en cada familia sólo uno de estos genes muta para causar la enfermedad.

La mayoría de los trastornos monogénicos son infrecuentes (presentes en < 1% de la población) y se manifiestan en la infancia. Muchos son graves y producen la muerte antes de la edad de reproducción si no se proporciona una asistencia médica moderna. El hecho de que la mayoría de los trastornos monogénicos sean graves en la infancia y raros en la población no es probablemente una coincidencia, sino más bien el reflejo del impacto de la *selección natural*. El efecto nocivo de estas mutaciones reduce la capacidad reproductiva (en las personas con la poca fortuna de heredarlas), y es improbable que las mutaciones y la enfermedad alcancen una alta frecuencia en la población.

Hay excepciones a esta idea general: casos en los que la mutación que provoca una enfermedad monogénica grave es habitual en la población general (como la *HbS*, que causa la anemia drepanocítica). Estos casos parecen deberse a una forma diferente de selección, denominada *selección estabilizadora*: situaciones en las que la mutación de un gen es beneficiosa en una circunstancia (un genotipo o ambiente) pero nociva en otra. Se cree que los portadores heterocigotos de la *HbS* están relativamente protegidos contra el paludismo y que este beneficio compensa el efecto nocivo de la drepanocitosis en los homocigotos.

Desde la década de 1980, la aparición del análisis de ligamiento hologenómico ha permitido un éxito rápido a la hora de identificar las mutaciones genéticas específicas que causan los trastornos mendelianos, con la identificación de cientos de genes para afecciones clínicas relevantes (para obtener una información completa ver <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=OMIM>). Los avances han sido impulsados por el desarrollo de una serie de técnicas de investigación potentes (*análisis de ligamiento basados en familias* seguido de *clonación posicional*) en las que se realiza una búsqueda hologenómica del gen causal, que se localiza primero en una

región del cromosoma. (La idea inicial del mapeo del ligamiento genético se remonta a Sturtevant en las moscas de la fruta en 1913, pero no se pudo llevar a la práctica en el ser humano hasta la década de 1980.)

Una vez que la búsqueda se ha centrado mediante el descubrimiento de ligamiento entre una región cromosómica y una enfermedad, esa región cromosómica se escudriña para buscar al culpable genético, que se identifica gracias a la observación de mutaciones que alteran la secuencia codificante de la proteína, y que son más llamativas en los casos de enfermedad, en comparación con los familiares no afectados y los controles basados en la población. La potencia de estas estrategias estimuló y fue potenciada a su vez por el Proyecto Genoma Humano, que proporcionó las bases de la información sobre la estructura, secuencia y variación genética del ADN necesarias para llevar a cabo estas búsquedas.

## INVESTIGACIÓN GENÉTICA DE LAS ENFERMEDADES COMUNES

A semejanza de los trastornos mendelianos, en la mayoría de las enfermedades comunes influye la herencia. Sin embargo, a diferencia de los primeros, la contribución genética a las enfermedades comunes parece deberse a la acción de muchos genes más que a un único gen en cada familia. La evidencia empírica a favor de este modelo proviene de los esfuerzos para emplear en los rasgos complejos el mismo enfoque (clonación posicional) que se aplicó con éxito a los trastornos monogénicos.

En la década de 1990, las herramientas de los análisis de ligamiento basados en familias se aplicaron a casi todos los trastornos comunes. Gran parte de este trabajo se realizó en poblaciones fundadoras aisladas (como Finlandia e Islandia), con el objetivo de simplificar la arquitectura genética y acceder a árboles genealógicos ampliados. Salvo por unos pocos éxitos notables, estos estudios revelaron pocos signos sólidos que localizaran los genes responsables de la enfermedad. En la mayoría de los cientos de estos estudios que se han publicado, se observan muchos signos estadísticos débiles (pocos, si es que hay alguno, son estadísticamente significativos dado el gran número de hipótesis evaluadas) y hay poco acuerdo entre los distintos estudios sobre una misma enfermedad.

A la vista de la potencia estadística bien conocida de los métodos de ligamiento basados en familias (lo que se fundamenta en su uso extenso para trastornos monogénicos) y de su éxito relativamente limitado, la mayoría de los investigadores concluyeron que las variantes raras en genes aislados no explican una gran fracción del riesgo de las enfermedades comunes. Si un gen único contuviera mutaciones de efecto amplio que explicaran el 20% o más del riesgo heredado de la diabetes tipo 2, la hipertensión o la esquizofrenia, sería probable que ya se hubiera encontrado hace tiempo su localización mediante análisis de ligamiento.

Un atajo potencial para comprender los determinantes genéticos de enfermedades comunes es identificar y estudiar las formas raras de inicio precoz de enfermedades que demuestran patrones de herencia mendeliana. Dado que estas familias presentan patrones de herencia compatibles con un gen relevante de efecto amplio, las potentes herramientas de clonación posicional se han podido utilizar y se han utilizado con éxito para identificar los genes responsables. Entre los ejemplos destacados se incluyen el papel de los genes *BRCA1* y *BRCA2* en el cáncer de mama de inicio precoz, la diabetes juvenil de inicio en la madurez como forma de diabetes tipo 2, muchos trastornos monogénicos de la presión arterial y la regulación electrolítica, así como la enfermedad de Alzheimer de inicio precoz.

Estos éxitos aportan información diagnóstica para las familias con formas de enfermedad graves y de inicio precoz, y nuevos datos sobre las vías subyacentes responsables de la enfermedad. Por ejemplo, se han identificado más de 20 genes que, una vez mutados, causan trastornos mendelianos raros de la presión arterial y la regulación electrolítica. Hasta ahora, cada uno de estos genes está activo en el riñón y la mayoría están implicados en la vía de la renina-angiotensina-aldosterona. Este resultado es una confirmación convincente de la relevancia central del riñón en la regulación de la presión arterial humana y ha sugerido nuevos objetivos terapéuticos muy prometedores.

Desde hacía tiempo se esperaba que los genes identificados como responsables del inicio precoz de formas monogénicas de enfermedades comunes contribuyeran a las formas más comunes de enfermedad en la población. En este contexto, las mutaciones graves podrían producir formas de inicio precoz, mientras que las alteraciones más prevalentes pero sutiles en los mismos genes podrían contribuir a las formas más comunes de la enfermedad. Para poner a prueba esta hipótesis de forma exhaustiva se requiere el desarrollo de las herramientas derivadas del Proyecto Genoma Humano y mejores métodos de análisis epidemiológico genético.

## ESTUDIOS DE ASOCIACIÓN: DE LOS GENES CANDIDATOS A LOS ESTUDIOS DE ASOCIACIÓN HOLOGENÓMICOS

Los estudios de asociación hologenómicos (EAHG) se basan en un concepto simple. Una vez identificada una variante genética, su frecuencia se mide en aquellas personas que tienen la enfermedad de interés y se compara con controles bien emparejados (extraídos de la población general o de familiares no afectados). Este proceso puede re-

petirse para todas las variantes genéticas que existan, hasta contar con una recopilación del genoma completo. Es preciso realizar análisis apropiados para descartar explicaciones alternativas para una asociación con una enfermedad, como un emparejamiento incorrecto de los casos y los controles, o los artefactos técnicos. Debido a que la distribución nula está bien descrita (bajo la hipótesis de que no existe asociación entre genotipo y fenotipo), es posible calibrar estos análisis e identificar las asociaciones reproducibles, extrayéndolas del amplio mar de los polimorfismos benignos.

Los estudios de asociación genética se iniciaron en el contexto del locus *HLA* en el cromosoma 6. El *HLA* se descubrió gracias a su papel en la tolerancia al trasplante y se caracteriza por diversas variaciones alélicas que pueden medirse a partir de las interacciones entre anticuerpos y antígenos. Cuando se midieron estas lecturas basadas en proteínas (inmunológicas) de la variación genética subyacente, se observó que los alelos *HLA* eran un determinante principal de la susceptibilidad a enfermedades infecciosas y autoinmunitarias. A partir de la década de 1960, se comenzó a recopilar datos empíricos sobre la genética de la población humana y se desarrollaron los estudios de asociación genética en el contexto del *HLA*.

Hacia la década de 1980, las herramientas de biología molecular permitieron medir directamente la variación del ADN (en lugar de utilizar mediciones de proteínas o de fenotipos como sustitutos de la variación genética subyacente), lo que inauguró la era moderna de la investigación genética. En esta era pregenómica, sólo resultaba práctico medir una o unas pocas variaciones genéticas en cada estudio, lo que limitaba los estudios de asociación a evaluaciones incompletas de genes «candidatos» seleccionados según criterios biológicos.

El estudio de genes candidatos llevó a un número modesto de asociaciones sólidas y reproducibles, como la contribución de Apo- $\epsilon$  4 a la enfermedad de Alzheimer, el factor V Leiden a la trombosis venosa profunda, una delección de 32 bases en el receptor de quimiocinas *CCR5* a la infección por el VIH, las variantes comunes del gen de la insulina a la diabetes tipo I, los SNP en el receptor activado por proliferador del peroxisoma (*PPAR- $\gamma$* ) y el canal de potasio Kir6.2 de la célula  $\beta$  al riesgo de diabetes tipo 2.

A principios del siglo XXI, los metaanálisis exhaustivos sobre los estudios de asociación genética publicados demostraron que había pocas asociaciones válidas y estaban muy distanciadas. Además, muchas de las asociaciones que se habían descrito en un principio no pudieron reproducirse, por lo que puede que fuesen falsas. Uno de estos análisis estimó que, en la era previa a los EAHG, sólo se habían documentado 10-20 asociaciones fidedignas de variantes genéticas frecuentes con enfermedades comunes.

Una de las principales razones de la situación en la que se encontraba esta bibliografía era la probabilidad intrínsecamente baja de encontrar un gen y una variante que contribuyesen a cualquier enfermedad concreta. Cada genoma contiene millones de variantes genéticas, y es de suponer que sólo una pequeña fracción de ellas influyen en las enfermedades. Esto suele describirse como un problema de «prueba de hipótesis múltiples», de modo que la comunidad investigadora busca asociaciones entre múltiples genes, múltiples variantes en cada gen y múltiples enfermedades. Un marco estadístico alternativo (bayesiano) aborda este problema basándose en las bajas asociaciones previas de asociación. A pesar de todo, está claro desde un punto de vista conceptual que se requieren umbrales estadísticos mucho más exigentes (que el tradicional  $P < 0,05$ ) para establecer una asociación entre las variantes genéticas y la enfermedad.

Al igual que en el análisis de ligamiento para los rasgos mendelianos, una clave para el éxito en los estudios de asociación fue la aparición de la búsqueda hologenómica, no sesgada por las hipótesis previas sobre los mecanismos biológicos. Gracias a la secuenciación del genoma humano, el desarrollo de bases de datos de SNP a gran escala y las herramientas para la genotipificación de hasta 1 M de SNP por persona, en 2005 se pudo empezar a realizar EAHG para identificar loci genómico que albergasen variación alélica. Una vez que se reconoció que cualquier variante dada tenía una probabilidad muy baja de estar realmente asociada a una enfermedad, se aplicaron umbrales estadísticos mucho más restrictivos (por lo general, con una exigencia de valor  $P$  de  $10^{-7}$  o menos para establecer una «significación hologenómica»).

La degeneración macular relacionada con la edad (DME) fue uno de los primeros éxitos de los EAHG. La DME es una enfermedad poligénica típica y frecuente (cap. 431); los hermanos de pacientes afectados tienen una probabilidad 3-6 veces mayor de padecer la enfermedad que las personas no relacionadas, aunque los análisis de ligamiento basados en familias han mostrado sólo unos resultados modestamente significativos de ligamiento (y con un grado también modesto de reproducibilidad). Los defectos fisiopatológicos que subyacen a la DME se desconocían en gran medida hasta que se descubrió que un polimorfismo codificante común en el gen del factor H del complemento es un factor de riesgo principal para la DME. La variante (*Y402H*) tiene una frecuencia elevada en la población (alrededor del 35% en poblaciones europeas) y aumenta el riesgo 2,5-3 veces en heterocigotos y 5-7 veces en homocigotos. Desde entonces, se ha descubierto que otros muchos factores del complemento albergan una variación genética común que influye en el riesgo de DME de un modo muy reproducible, lo que proporciona una información no ambigua sobre el papel fundamental del complemento en esta enfermedad común.

Desde 2005, los EAHG se han usado para identificar literalmente cientos de nuevas variantes genéticas que muestran asociaciones reproducibles con una gran diversidad de enfermedades humanas comunes. En este contexto, se han desarrollado varios criterios y estándares que han eliminado en gran medida las dificultades previas referentes a las descripciones irreproducibles de asociación, lo que ha convertido a los estudios de asociación en un método fiable para identificar loci genómicos relacionados con enfermedades humanas. El National Human Genome Research Institute de los National Institutes of Health mantiene un catálogo de hallazgos obtenidos en EAHG ([www.genome.gov/26525384](http://www.genome.gov/26525384)) que, en el momento de escribir esta obra, incluía 904 de estas asociaciones para 165 rasgos. Esto supone un progreso rápido en comparación con los alrededor de 24 hallazgos similares que se conocían a principios de la década.

Los resultados de los EAHG respaldan varias conclusiones sobre el papel de las variantes genéticas frecuentes en enfermedades comunes. En primer lugar, la mayoría de las enfermedades estudiadas mediante EAHG han proporcionado nuevos datos, lo que sugiere que este método tiene una utilidad general. En segundo lugar, sólo una pequeña fracción de estos hallazgos se conocían previamente, lo que indica que se pueden obtener nuevas pistas mediante el mapeo genético de las enfermedades comunes. En tercer lugar, la mayoría de las asociaciones demuestran unos cocientes de probabilidades (*odds ratio*, OR) muy modestos (del orden de 1,1-1,5 veces), lo que indica que la selección natural ha eliminado del conjunto de variantes comunes los alelos que tienen un gran efecto. En cuarto lugar, la mayoría de los SNP se sitúan en regiones no codificantes, lo que sugiere que actúan mediante efectos sobre la regulación génica en lugar de alterar directamente las secuencias codificantes de proteínas. En quinto lugar, hasta el momento sólo se ha explicado una fracción modesta de la heredabilidad estimada de cada enfermedad, lo que indica que intervienen otras variantes comunes que tienen un efecto más modesto, variantes raras, interacciones genéticas u otras influencias (aún no previstas).

Debido a que los EAHG son hologenómicos (no se limitan a «genes candidatos»), sirven para poner a prueba la hipótesis de que las investigaciones previas habían identificado grupos de genes relevantes para cada enfermedad (relevantes según la variación genética y la herencia). En el caso de las enfermedades autoinmunitarias, muchos (quizá la mitad) de los 100 o más hallazgos obtenidos de los EAHG están próximos a un gen del que ya se conocía su implicación en el sistema inmunitario. De forma similar, una fracción sustancial de las variantes genéticas en las que se ha observado una influencia sobre los niveles lipídicos se sitúa cerca de los genes de los que ya se sabía que participan en la biología de los lípidos (porque ya se conocían debido a las mutaciones raras que contribuyen a formas mendelianas de hiperlipidemia o gracias a investigaciones biológicas). Estos hallazgos confirman que los mecanismos fisiopatológicos básicos ya conocidos pueden «validarse» desde el prisma de los factores de riesgo heredados y alentar la investigación de los nuevos hallazgos proporcionados por los EAHG.

Por el contrario, en el caso de algunas enfermedades, la mayoría de las variantes genéticas encontradas son nuevas y no se sitúan cerca de genes estudiados previamente. Uno de estos casos corresponde a la diabetes de tipo 2, para la que se han encontrado 35 loci genómicos independientes que influyen en el riesgo de enfermedad, de los que sólo unos pocos se habían implicado previamente por otros métodos. Esto puede indicar y arrojar luz sobre las lagunas existentes en nuestros conocimientos previos de la fisiopatología de la diabetes de tipo 2.

Aunque los resultados de los EAHG son excitantes, han suscitado muchas más preguntas de las que han respondido. Estos descubrimientos implican a regiones genómicas especiales, pero, hasta el momento, los genes causales sólo se han demostrado en pocos casos. Esto supone un gran desafío, debido en gran medida a que muchas de estas variantes comunes son no codificantes y sigue siendo difícil relacionar la variación no codificante con los genes regulados por ese mecanismo. Hasta que se identifiquen genes verdaderamente nuevos, se necesitan muchos más estudios para descubrir sus funciones biológicas y fisiológicas. Por último, los EAHG explican sólo una fracción modesta de la heredabilidad estimada de la mayoría de las enfermedades, dejando abierta la pregunta de qué genes y qué tipos de variantes y efectos genéticos explican el resto.

## DE LAS VARIANTES COMUNES A LOS GENOMAS INDIVIDUALES

Aunque gran parte de la variación genética humana se debe a variantes comunes del ADN (como las analizadas mediante EAHG), cada persona también hereda muchas miles de variantes que surgieron más recientemente y que tienden a tener una frecuencia menor y a ser más específicas de cada población. En la medida en que estas variantes tienen efectos muy amplios sobre el fenotipo, puede que se hayan identificado previamente a partir de los estudios de ligamiento basados en familias de los trastornos mendelianos. Sin embargo, casi con toda certeza existe una amplia gama de variaciones de frecuencia menor que son demasiado infrecuentes para haberse detectado con los EAHG de primera generación y que tienen unos efectos demasiado modestos como para haberse reconocido e identificado en los análisis de ligamiento basados en familias.

El estudio de estas variantes de menor frecuencia y raras se puede llevar a la práctica en la actualidad gracias a los avances en la tecnología para la secuenciación del ADN. Con el drástico abaratamiento y el mayor rendimiento, cada vez es más práctico secuenciar genomas individuales en el contexto de la investigación médica (y, en el futuro, en la práctica clínica). Este método permitirá una evaluación mucho más completa de la variación genética de la que podía lograrse previamente e incorporará variantes comunes y raras.

La primera tarea consistirá en desarrollar métodos para interpretar las millones de variantes de cada genoma. Las variantes de alta frecuencia ya han sido bien estudiadas mediante EAHG, y a cada una se añade cada vez más información sobre su asociación con la enfermedad correspondiente. Para las variantes con una frecuencia menor, pero que aún se observan en un número considerable de personas no relacionadas, se puede aplicar la metodología de asociación básica. Es decir, las frecuencias de cada variante específica pueden medirse en los casos afectados y en los controles no afectados y compararse con una distribución nula elaborada de forma apropiada y con un umbral estadístico adecuado. Por otra parte, la disponibilidad de una base de datos mucho más completa de variación del ADN (como la que está creando el 1000 Genomes Project) dará lugar a una segunda generación de EAHG que será más completa para las variantes de frecuencia más baja.

Sin embargo, muchas variantes de ADN serán específicas de cada persona (y de sus familiares cercanos) y requerirán enfoques diferentes. Si el análisis incluye un árbol genealógico amplio, con muchos familiares afectados y no afectados, puede que sea posible realizar el análisis de asociación en familias únicas. No obstante, lo más habitual es que sea necesario analizar amplios grupos de muestras, medir la tasa de variantes distintas (raras a nivel individual) en cada gen y comparar estas tasas entre casos y controles.

En un futuro lejano, puede que aprendamos a «leer» las secuencias genómicas y predecir el efecto de una variante no observada hasta entonces. Sin embargo, en el futuro inmediato, la interpretación requerirá un análisis estadístico de los genomas, que documente que la variación en genes concretos se asocie de forma sólida y reproducible con cada enfermedad particular.

## IMPLICACIONES Y PERSPECTIVAS FUTURAS

Los factores hereditarios contribuyen sustancialmente a las enfermedades tanto comunes como raras. Los trastornos mendelianos suelen estar causados por mutaciones raras en las regiones de los genes codificantes de proteínas. Las variantes comunes, investigadas mediante EAHG, suelen tener efectos modestos y actúan a menudo a través de efectos no codificantes sobre la regulación de genes. Cada persona porta una gran reserva de variantes menos comunes que pronto se analizarán para determinar su papel en la enfermedad utilizando métodos de secuenciación de ADN de próxima generación. Parece razonable esperar que un análisis integrador de esta información se traducirá en una lista definida de los genes y las variantes en dichos genes (tanto comunes como raros) que contribuyen a cada una de las enfermedades humanas.

El éxito a la hora de identificar los genes y las mutaciones sólo será útil si mejora la predicción, el diagnóstico, la comprensión y el tratamiento. La predicción y la medicina personalizada requieren una base de evidencia que ofrezca beneficios clínicos. Esto implicará la incorporación de la variación del ADN en cohortes válidas desde el punto de vista epidemiológico y poner a prueba métodos concretos de predicción genética en ensayos clínicos. El hecho de que algunas pruebas genéticas puedan ser predictivas no quiere decir que la secuenciación rutinaria del genoma sea útil para los pacientes, y será necesario un gran esfuerzo para ofrecer al público y a los médicos una orientación para interpretar estos datos.

La comprensión biológica requiere combinar la investigación clínica con la investigación básica, de modo que los genes donde se detecten mutaciones en los pacientes se estudien en el laboratorio. Será necesario ubicar estos nuevos genes en vías biológicas conocidas (y en otras aún no conocidas), así como comprender el modo en el que su disfunción y disregulación provocan enfermedad. En algunos casos, como sucede con el papel del complemento en la DME (v. antes), las respuestas iniciales pueden llegar pronto; en otros, en los que la biopatología aún se ignora, la información que se obtendrá a partir de estas pistas es impredecible. Es de suponer que, con el paso del tiempo, los datos genéticos derivados de los pacientes lleven a una nueva generación de terapias que aborden más directamente las causas subyacentes de riesgo en la población. Lo que es más seguro es que se está acumulando información genética y genómica a un ritmo asombroso, lo que supone un gran potencial, así como un gran reto para el futuro de la medicina.

## LECTURAS RECOMENDADAS

- Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322:881-888. *Revisión del mapeo genético de las enfermedades humanas.*
- Lifton RP. Individual genomes on the horizon. *N Engl J Med*. 2010;362:1235-1236. *Comentario sobre la previsión del uso de secuencias genómicas individualizadas en medicina clínica.*
- Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med*. 2010;363:166-176. *Revisión.*
- Fark YJ, Claus R, Weichenhan D, y col. Genome-wide epigenetic modifications in cancer. *Prog Drug Res*. 2011;67:25-49. *Revisión.*