



Nuevas herramientas para avanzar en la Genética de enfermedades prevalentes de la infancia: Diabetes tipo 1 y Enfermedad Celiaca

José Ramón Bilbao Catalá

Unidad de Investigación en Endocrinología y Diabetes.

Laboratorio de Inmunogenética

Hospital de Cruces.

Barakaldo-BIZKAIA.

rbilbao@hcr.uosakidetza.net

tel.: 946006376

Introducción

La identificación de los genes causantes de las *enfermedades complejas* (que aparecen como consecuencia de la acción del medio ambiente en sujetos genéticamente predispuestos) es un reto importante para poder seleccionar aquellos individuos con mayor riesgo de desarrollar estas patologías y poder aplicar sobre ellos terapias preventivas con ciertas garantías de éxito. Esta tarea cobra si cabe mayor importancia cuando se trata de patologías crónicas con elevada morbi-mortalidad y alto riesgo de complicaciones a lo largo de la vida. La diabetes mellitus de tipo 1 (DM1) es un buen ejemplo de esta situación: pendientes de que las terapias curativas como el trasplante de islotes pancreáticos o en la ingeniería celular para la obtención de células productoras de insulina a partir de células *madre* se hayan desarrollado lo suficiente

para que puedan ser aplicables en la clínica, la prevención de la enfermedad sigue siendo un objetivo importante. No obstante, los ensayos de prevención secundaria (una vez iniciado el proceso de destrucción autoinmune de la célula β pancreática) en individuos *prediabéticos* que presentan marcadores inmunológicos de la enfermedad (autoanticuerpos anti-insulina y anti-islole circulantes) se han demostrado ineficaces, lo que sugiere la necesidad de contar con herramientas de predicción todavía más precoces, capaces de identificar a la población diana cuando la manifestación fenotípica del riesgo genético (autoinmunidad) todavía no se ha iniciado. La Enfermedad Celíaca (EC) o intolerancia al gluten, es otra enfermedad crónica de aparición en la infancia, que si bien es significativamente más benigna que la DM1, supone un cambio de estilo de vida importante (dieta libre de gluten) y también pueden aparecer complicaciones más graves en el adulto si la EC no se diagnostica (formas silentes y ocultas) o incluso si se realizan transgresiones reiteradas a la dieta sin gluten. Precisamente, se ha sugerido que entre las posibles consecuencias de una EC no diagnosticada (o no tratada) estaría la propia DM1, y en un estudio de nuestro grupo, un 7% de los pacientes diabéticos tipo 1 presentaban una EC no diagnosticada en el momento del debut de DM1.

A pesar de afectar a sistemas y órganos específicos, la DM1 y la EC son trastornos del sistema inmune, en los que el organismo pierde la tolerancia hacia estructuras propias provocando su destrucción. Esta fisiopatología común explicaría la asociación que se observa entre las dos entidades y dado que se trata de enfermedades con un componente hereditario, factores genéticos comunes podrían determinar el riesgo innato a padecerlas. De hecho, se conoce desde hace mucho tiempo que los determinantes más importantes de la susceptibilidad genética a ambas enfermedades se localizan en el llamado *Complejo Mayor de Histocompatibilidad* (MHC), en el cromosoma 6p21, que

sería responsable de aproximadamente el 40% de la heredabilidad en la diabetes autoinmune y de al menos el 30% en la enfermedad celiaca. En población Caucásica, los alelos HLA-DRB1*0301 y HLA-DRB1*04 (DR3 y DR4, respectivamente) confieren el mayor riesgo a padecer DM1. Estos alelos están estrechamente ligados, respectivamente, a los alelos que codifican las moléculas HLA-DQ2 y HLA-DQ8, que se asocian con riesgo a desarrollar EC. No obstante, estos haplotipos también están presentes en la población que no desarrolla estas enfermedades, lo que sugiere que existen otros genes que contribuyen en la predisposición a desarrollar DM1 y EC. En las dos últimas décadas, se han realizado grandes esfuerzos para localizar estos otros genes de susceptibilidad, pero los resultados no han sido del todo satisfactorios y nuestra capacidad de predecir el futuro desarrollo de enfermedades autoinmunes sigue estando muy limitada.

En la actualidad, se están aplicando nuevas alternativas que aprovechan los recientes avances tecnológicos en el *Análisis de la expresión génica del genoma completo en microarrays*, *Genotipado masivo* y en *Bioinformática* para intentar optimizar la identificación de genes candidato.

Estrategias clásicas de búsqueda de genes de susceptibilidad:

Dado que el patrón de herencia que siguen las enfermedades complejas, como DM1 y EC, no se ajusta a los modelos mendelianos clásicos, y el número de casos en cada una de las familias suele ser escaso, es imposible diseñar estudios de ligamiento clásico como los que se han utilizado en la investigación de las enfermedades *monogénicas* o *Mendelianas*. En los últimos años se han empleado fundamentalmente dos tipos de estrategias:

Estrategia de genes candidato: Se trata de una búsqueda dirigida o *educada* de potenciales alteraciones genéticas relacionadas con la enfermedad. En base al conocimiento previo de los mecanismos patogénicos o las rutas metabólicas implicadas en la enfermedad objeto de estudio, se seleccionan los posibles *candidatos* y se realizan *estudios de asociación* de marcadores polimórficos. La pregunta a la que esta estrategia quiere responder podría formularse como sigue: ¿Está alguno de los alelos del gen candidato aumentado en los pacientes con respecto de la población general? En su versión más sencilla, se utiliza un diseño de casos y controles, comparándose la frecuencia de determinada variante en ambos grupos. Debido a la creciente complejidad de las poblaciones actuales, y para evitar posibles asociaciones debidas a estratificaciones genéticas por orígenes étnicos diferentes en los casos y los controles, se recomienda recurrir a los *estudios familiares* de asociación, como el *AFBAC (Affected Family Based Controls)*, donde los cuatro alelos parentales de una familia nuclear (el probando y sus padres), los dos alelos no transmitidos al individuo enfermo son utilizados como genotipo control. Otro ejemplo es el test de desequilibrio de la transmisión o *TDT (Transmisión Disequilibrium Test)* en el que se compara, en un grupo de familias, la frecuencia con la que determinado alelo se transmite a la descendencia enferma, respecto de la proporción teórica del 50%. Cualquier diferencia en la frecuencia de determinado alelo en los grupos enfermo y control o una proporción de transmisión de dicho alelo que esté desviada de la neutralidad se interpreta como evidencia a favor de la implicación de esa variante genética en la susceptibilidad de padecer la enfermedad.

Estudios de mapeo genómico: Esta estrategia pretende determinar la localización genómica (generalmente la banda cromosómica) en la que potencialmente se localiza el gen implicado en la enfermedad. El proceso se basa en el análisis sistemático de la totalidad del genoma en varios miembros enfermos de familias, utilizando segmentos de ADN muy variables (polimórficos) cuya localización exacta es conocida (marcadores). La aproximación más potente, que ha sido empleada con mayor frecuencia, utiliza parejas de hermanos enfermos (*affected sib-pairs*) para determinar si existe un mayor grado de identidad genética (comparten más alelos de ese marcador) del que cabría esperar por azar. Conviene recordar que en el caso de un gen neutro para la enfermedad, el 25% de los hermanos comparten ambos alelos, el 50% comparte uno de los dos y el 25% restante no comparte ninguno de los alelos. Cualquier desviación de estas proporciones indicaría que el marcador polimórfico analizado se encuentra próximo (en ligamiento) con un gen implicado en la enfermedad. La mayor ventaja de este tipo de estudios es que el investigador puede carecer por completo de información acerca de los mecanismos etiológicos y las rutas metabólicas que subyacen la enfermedad, con lo que no es necesario seleccionar candidatos *a priori*.

Estas dos aproximaciones han sido utilizadas en incontables ocasiones a lo largo de estos últimos años, y nos han permitido entender mejor la Genética de enfermedades como la DM1 y la EC. Entre los innumerables trabajos publicados cabe destacar el primer barrido genómico en parejas de hermanos enfermos de DM1 realizado por el grupo del profesor J.A. Todd en 1994, y que localizó 17 regiones del genoma (denominadas IDDM1-IDDM17) en las que presumiblemente se encuentran los genes causantes de DM1. Este trabajo se realizó genotipando *microsatélites* (secuencias de ADN no codificantes, formadas por repeticiones de dos o tres nucleótidos, cuya

longitud (número de repeticiones) es variable, apareciendo diferentes variantes o alelos en la población, lo que las convierte en herramientas útiles para identificar y comparar individuos o grupos (enfermos frente a controles, por ejemplo). Los estudios de mapeo genómico como el descrito utilizan aproximadamente 400 *microsatélites* separados por unos 10cM, lo que supone un trabajo de laboratorio considerable.

Posteriormente, se han replicado algunos de estos *loci* en poblaciones independientes y, mediante estudios de asociación de genes candidato, se han confirmado algunos de los posibles genes en dichas regiones. No obstante, la mayoría de estas regiones son muy extensas y cada una de ellas alberga numerosos genes, lo que dificulta la identificación del verdadero gen implicado y de sus variantes patogénicas. También se han realizado mapeos del genoma completo en la EC, y se conocen también posibles localizaciones de genes de susceptibilidad. No obstante, como se apuntaba al principio, nuestra capacidad de predecir el riesgo genético real de un individuo a padecer cualquiera de estas enfermedades sigue siendo muy limitada.

Nuevas herramientas para los estudios genéticos

Las nuevas herramientas de las que disponemos hoy en día para abordar el análisis genético de las enfermedades humanas se basan tanto en los espectaculares avances de la Biología como en la irrupción de nuevas tecnologías que han hecho posible generar enormes cantidades de información genética, así como en las soluciones informáticas para la gestión y el análisis de estos datos. Es importante mencionar, aunque de manera superficial, algunos de los progresos más significativos de estos últimos tiempos.

El Proyecto Genoma Humano: El que es, sin duda alguna, el logro más espectacular en la Biología moderna ha puesto a nuestra disposición la información de la práctica totalidad de los nucleótidos que componen el material genético humano. Es un proyecto aún en marcha, ya que todavía quedan algunas regiones sin secuenciar, fundamentalmente en los centrómeros y telómeros de los cromosomas.

La secuenciación del genoma ha puesto de manifiesto la existencia de una identidad de más del 99,9% entre todos los seres humanos, con lo que los determinantes genéticos causantes de las enfermedades “comunes”, como la diabetes, se encontrarían en el 0,1% variable. También nos ha permitido conocer que más del 90% de la variabilidad genética entre los individuos corresponde a polimorfismos de un nucleótido (*Single Nucleotide Polymorphism* o SNP) que ocurren cada 300-400 nucleótidos y que pueden aparecer tanto en regiones codificantes de los genes (provocando cambios en la proteína), en regiones genéticas no codificantes (intrones, zonas de *splicing*, promotores, etc., que podrían modular su expresión) o en espacios intergénicos sin aparente expresión fenotípica. Se calcula que existen más de 10.000.000 de SNPs, la mayoría de los cuales son neutros, aunque algunos podrían influir en el fenotipo. Los SNPs constituyen otro conjunto de marcadores genéticos que pueden ser utilizados en estudios de asociación con enfermedades; como veremos más adelante, es posible su genotipado a gran escala con la tecnología actual.

Bioinformática: La enorme cantidad de información “cruda” obtenida de la secuencia humana es compilada junto con los conocimientos ya existentes sobre los genes codificantes: su estructura, localización, etc., creándose enormes bases de datos en las que, además de los genes ya conocidos, se “predice” mediante algoritmos informáticos,

la localización y estructura (e incluso posible función) de genes aún no aislados en el laboratorio, lo que se denomina *anotación genética*. Estas bases de datos ofrecen buscadores y herramientas accesibles por Internet, siendo las dos más importantes la que está gestionada por el Instituto Nacional de Salud (NIH) estadounidense (www.ncbi.nlm.nih.gov/genome/guide/human/) denominada *Entrez*, que alberga también bases de datos más clásicas, como *GenBank*, *OMIM* y *PubMed*, entre otras, así como la plataforma que gestiona el Instituto Bioinformático Europeo (EBI-EMBL) junto con el Instituto Sanger británico (www.ensembl.org). Entre la información disponible, se encuentra la propia secuencia cruda del genoma, la anotación de genes codificantes (promotores, exones, intrones, procesamientos alternativos, etc.), la homología con otras especies que también están siendo secuenciadas, así como toda la variación genética (SNPs), incluyendo las frecuencias alélicas de estos polimorfismos en las diferentes poblaciones humanas que están siendo analizadas.

Microarrays de Expresión Génica: De la mano de la *Nanotecnología*, se ha reducido de manera espectacular el “volumen” físico de las reacciones bioquímicas que se emplean para realizar las determinaciones y en la actualidad, utilizando los denominados *microarrays* o *microchips*, se realizan, de manera simultánea, decenas de miles de reacciones en el espacio de un portaobjetos de microscopía. Una de las variantes más empleadas son los *microarrays* de expresión génica, en los que se analiza el ARN mensajero aislado de una muestra biológica para determinar el nivel de transcripción de la totalidad del genoma (por ejemplo, 30.000 genes humanos) denominado *transcriptoma*. Es posible diseñar un experimento encaminado a identificar genes candidato a una enfermedad, en el cual se compara la expresión de todo el genoma en muestras de tejido enfermo frente a tejido sano, utilizando *microarrays* de

expresión. Aquellos genes que muestren un nivel de expresión diferente en los dos estados comparados estarán potencialmente implicados en el desarrollo de la enfermedad. También se han desarrollado herramientas bioinformáticas para extraer la información relevante (en inglés, *data mining*) y presentarla de manera comprensible y coherente (por ejemplo, genes con *expresión diferencial*). Otro nivel superior de análisis supone ir más allá de los listados de genes individuales sobre- o subexpresados y utilizar los conocimientos existentes sobre interacciones entre genes (o sus productos) para agruparlos en rutas metabólicas o procesos celulares, lo que puede ayudar en la búsqueda de dianas terapéuticas o en la identificación de factores de transcripción o ligandos comunes a los diferentes genes cuya transcripción se presenta alterada en el *microarray*.

También se han desarrollado herramientas para el estudio del conjunto de las proteínas sintetizadas por un tejido (*Proteómica*) y de los metabolitos producidos (*Metabolómica*) pero debido a su mayor complejidad experimental, se encuentran menos avanzados y su uso está menos estandarizado; no obstante, obedecen al mismo planteamiento y los diseños experimentales son análogos a los que emplea la *Transcriptómica*.

Genotipado: También se han producido avances importantes en nuestra capacidad de obtener datos genotípicos mediante el análisis del ADN. En este momento, es posible genotipar hasta 500.000 SNPs de manera simultánea a partir de una muestra de ADN de un individuo, utilizando *microchips* de genotipado masivo. Si se cuenta con los recursos económicos suficientes y se dispone de una amplia colección de ADN de enfermos y de controles, es factible plantearse un estudio de asociación que abarque todo el genoma, en lugar de los estudios de genes candidato del pasado reciente.

Diseño de nuevas estrategias:

Estrategia funcional en enfermedad celiaca.

En nuestro particular intento de identificar nuevos determinantes genéticos para enfermedades autoinmunes, nos hemos aprovechado del hecho de que el diagnóstico de esta enfermedad se basa en el estudio de biopsias intestinales de los pacientes. Hemos realizado estudios de *microarrays* de expresión del genoma completo con el fin de identificar posibles candidatos funcionales. En concreto, hemos analizado la expresión en tejido intestinal de pacientes al debut de la EC (cuando todavía ingieren gluten), comparándola con biopsias obtenidas tras más de dos años con dieta libre de gluten, cuando el paciente es asintomático, no presenta lesión intestinal y no tiene autoanticuerpos circulantes. Así, hemos querido identificar los genes cuya expresión se ve alterada por una exposición *crónica* al gluten y potencialmente intervienen en las fases avanzadas de la enfermedad. Por otro lado, en otro grupo de biopsias de pacientes celíacos tratados, éstas se partieron en dos trozos, incubándose una mitad con gliadina de trigo y la otra mitad sin esta proteína, durante cuatro horas. Se analizó la expresión en estas biopsias para así identificar aquellos genes que se alteraban tras una exposición aguda al gluten y podrían estar involucrados en el inicio de la respuesta autoinmune. Este análisis de *microarrays* ha identificado más de un millar de genes con expresión alterada por el gluten. De todos ellos, analizando las bases de datos de rutas metabólicas y procesos celulares, se han seleccionado aquellos que participan en vías relevantes para la patogenicidad autoinmune (apoptosis, señalización celular, citoquinas, etc.). Posteriormente, se ha cruzado esta información con los resultados de los mapeos genómicos publicados en los últimos años, para obtener la lista de genes candidato funcionales cuya localización coincide con regiones *calientes* del genoma. Finalmente,

se ha realizado un proceso de selección de polimorfismos (SNPs) utilizando herramientas informáticas de predicción de alteración funcional, junto con criterios genéticos (conservación en diferentes especies, patrones de desequilibrio de ligamiento, etc.). Así, hemos obtenido un listado de 384 polimorfismos que serán analizados en un estudio de asociación genética comparando pacientes y controles. El futuro próximo nos aclarará si esta estrategia ha sido acertada.

Estrategia posicional en diabetes tipo 1.

En poblaciones del sur de Europa, como la nuestra, los determinantes genéticos que confieren el máximo riesgo a desarrollar tanto DM1 como EC se encuentran en los haplotipos HLA-DR3-DQ2. La comparación de pacientes de ambas enfermedades homocigotos para este haplotipo ha mostrado que entre los diabéticos, los haplotipos extendidos mayoritarios contienen además los alelos HLA-A*30 y HLA-B*18, lo que en Genética de poblaciones corresponde al haplotipo ancestral AH18.2, mientras que en celíacos el haplotipo AH8.1 es predominante. Esto parece indicar que en el haplotipo 18.2 existiría otro factor diabetogénico independiente de HLA-DR3-DQ2, que lo haría más diabetogénico. Un estudio de genotipado masivo de una región de 5 Mb en la región MHC, analizando 2360 SNPs en estos pacientes demostró que no existen variaciones en el haplotipo 18.2, el cual presenta una conservación extraordinaria y sugiere que el posible alelo diabetogénico del haplotipo estaría fijado en todas las copias del mismo. Por su parte, el resto de haplotipos HLA-DR3-DQ2 (como AH-8.1) presentan un alto grado de variabilidad, lo que significaría que algunos de ellos pueden portar el alelo diabetogénico mientras otros no. Esta situación concuerda con el menor grado de diabetogenicidad de estos haplotipos DR3 y con la menor asociación con diabetes del HLA-DR3 en poblaciones pobres en 18.2 (norte de Europa). La estrategia

que hemos adoptado para localizar este segundo gen diabetogénico en la región MHC consiste en comparar pacientes diabéticos y controles, todos ellos homocigotos para HLA-DR3-DQ2 y portadores de una copia del haplotipo ancestral AH-18.2, que será idéntico en ambos grupos. Analizando los 2360 SNPs del fragmento de 5 Mb, nuestro objetivo es identificar aquellos polimorfismos del *otro haplotipo* que muestren asociación con la enfermedad. Hasta el momento, hemos identificado dos regiones que muestran una fuerte asociación (*Odds Ratio* de 58 y 10, respectivamente) y podrían albergar ese segundo gen de susceptibilidad. Ahora, es necesario confirmar estos resultados en una muestra independiente y localizar con mayor precisión el gen (o genes) implicado(s).

Con estos estudios y muchos otros que están realizando otros Grupos de Investigación, confiamos en poder identificar nuevos marcadores genéticos que nos permitan contar con herramientas eficientes para la predicción de estas enfermedades autoinmunes así como identificar posibles dianas para una actuación preventiva eficaz.