

BIOLOGIA MOLECULAR

Análisis del DNA III: Mapping genético y mapping físico

E. Salido

Departamento de Anatomía Patológica. Universidad de La Laguna. Tenerife.
Department of Pediatrics and Genetics. UCSF (San Francisco, USA).

ANÁLISIS DNA III

El conocimiento de la localización exacta de un gen determinado dentro del genoma es un paso muy importante para llegar a conocer la relación existente entre este gen y un fenotipo concreto. La actividad enfocada a localizar un gen en el genoma, conocida como «*gene mapping*» es una de las áreas de más rápido crecimiento en Medicina en el momento presente. Esto se debe en parte a que se pueden aislar genes asociados a enfermedades en virtud de su mera posición en el genoma, sin necesidad previa de identificar y purificar las proteínas alteradas en dichas enfermedades, sin conocer ni siquiera las rutas metabólicas o defectos bioquímicos implicados. Esta *estrategia posicional*¹ ha sido muy fructífera en el esclarecimiento de las bases moleculares de enfermedades y es conocida también como «Genética Reversa» (fig. 1). Entre las enfermedades renales cuyo gen responsable está siendo definido mediante una estrategia posicional se encuentran la enfermedad poliquística renal autosómica dominante (ADPKD)² (que será tratada en un número futuro de NEFROLOGÍA), la enfermedad de Hippel-Lindau³, la fiebre mediterránea familiar⁴, varias formas de esclerosis tuberosa^{5,6}, y los síndromes de Rubinstein-Taybie⁷ y de Alaguille⁸. Para otras enfermedades con repercusión renal, el gen responsable ya ha sido clonado; haciendo uso de esta estrategia posicional en mayor o menor medida. Este es el caso del tumor de Wilms⁹ (no todos los casos) y una serie de síndromes ligados al cromosoma X, tales como el de Alport¹⁰, el de Lowe¹¹, el de Kallman¹² y la diabetes insípida nefrogénica¹³. La información sobre la posición de genes a lo largo de los cromosomas humanos es relevante no sólo para llegar al aislamiento final de genes de interés sino también para un diagnóstico y consejo genéticos. Existen esfuerzos internacionales dedicados a construir un mapa genético que permita finalmente catalogar todos los genes del hombre y las enfermedades asociadas a los mismos.

Esencialmente disponemos de dos vías para ordenar genes a lo largo de cromosomas mapping genético y mapping físico.

I. MAPPING GENÉTICO

Consiste en la estimación de la tendencia de dos genes a segregarse juntos (ligamento) durante la meiosis en estudios de familias. La meiosis resulta en gametos haploides que tienen un solo ejemplar de cada pareja de cromosomas (fig. 2). Los alelos de loci en cromosomas no-homólogos (cromosomas homólogos son los componentes de una pareja de cromosomas) se segregan, es decir pasan a las células haploides, independientemente durante la meiosis (fig. 3), mientras que los alelos de loci próximos entre sí, en el mismo cromosoma, tienden a heredarse «en bloque» (ligados). Pero no todos los alelos en un mismo cromosoma se heredan ligados debido a la existencia de recombinación (*cross-over*) (fig. 2), un mecanismo base de la variabilidad genética. La distancia entre dos loci se puede estimar analizando el comportamiento de los mismos genes durante un número de meiosis y determinando, en términos estadísticos, qué grado de ligamiento existe entre ellos. Este método no proporciona unidades físicas de distancia (kb: kilobase, mil pares de bases; Mb: megabase, un millón de pares de bases...) sino unidades de recombinación genética: se dice que la distancia entre dos genes es un centimorgan (1 cM) si se detecta recombinación entre ellos en el 1 % de las meiosis estudiadas. Ahora bien, cuanto más próximos estén dos genes a lo largo del cromosoma, menos será el porcentaje de recombinaciones observadas.

Mientras que los procedimientos de mapping físico son cruciales a la hora de clonar genes y determinar sus secuencias, el gran interés suscitado por los estudios de ligamiento genético deriva de que, además de ser el primer paso en la clonación de muchos genes, constituye una herramienta diagnóstica excelente. Mediante análisis de ligamiento genético se pueden diagnosticar los estados de portador y enfermo no sólo de enfermedades cuyo gen responsable ha sido clonado, sino también de aquellas muchas otras en que aún no conocemos la secuencia del gen afecto.

Correspondencia: Dr. Eduardo Salido Ruiz.
Departamento de Anatomía Patológica. Facultad de Medicina
(Universidad de La Laguna).

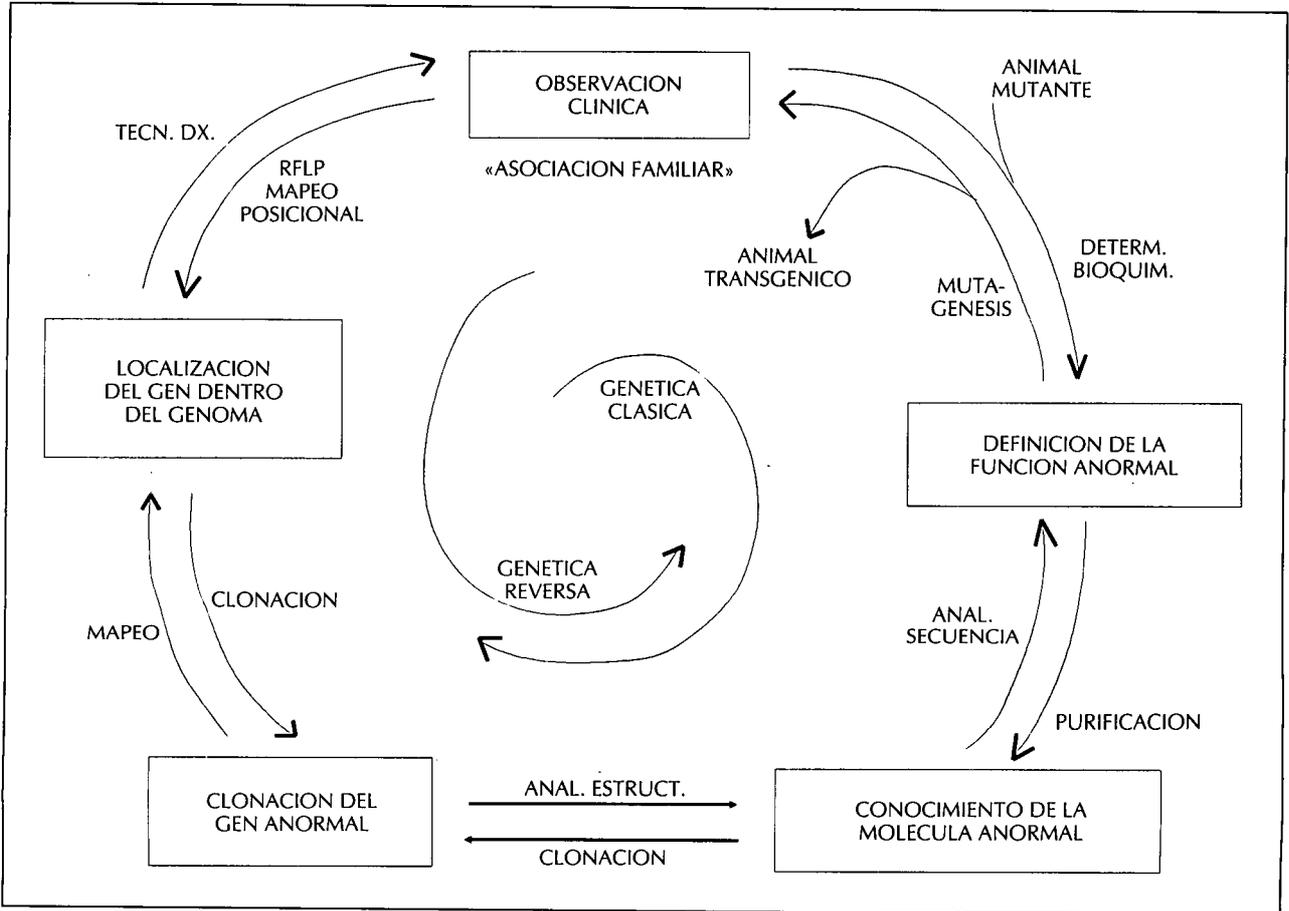


Fig. 1.—Diagrama de las estrategias funcional (clásica) y posicional (reversa) para la clonación de genes responsables de enfermedades.

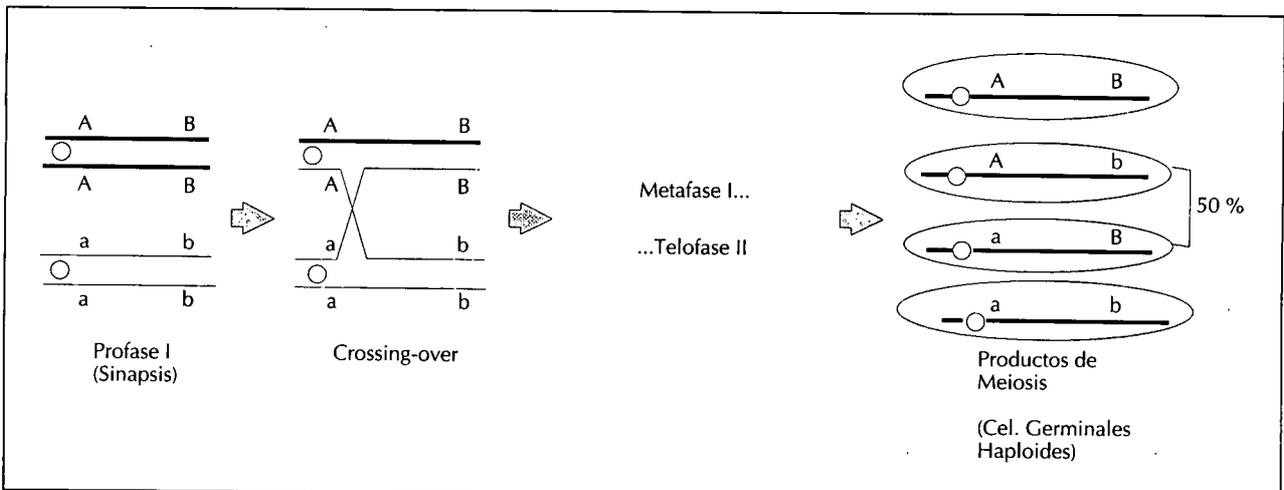


Fig. 2.—Durante la profase I de la meiosis tiene lugar el intercambio recíproco de material genético entre cromátidos conocido como crossing-over. El resultado de un episodio de recombinación entre dos loci (A y B) en una célula germinal heterocigótica (AaBb) son cuatro gametos haploides, dos de los cuales son recombinantes (Ab y aB), mientras que los otros dos reproducen el genotipo paterno (AB y ab).

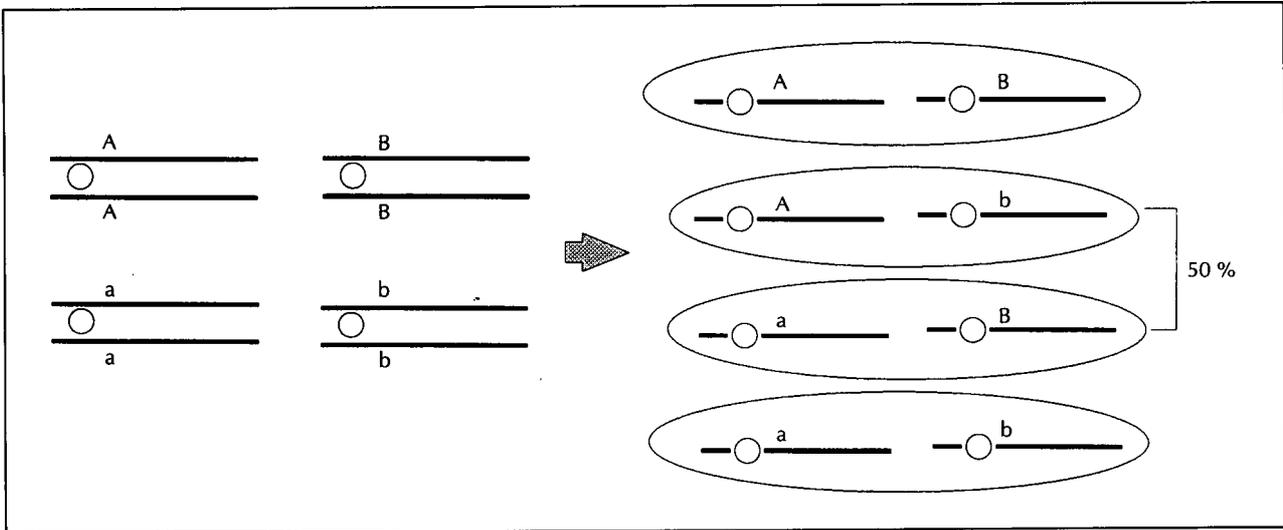


Fig. 3.—Segregación independiente de genes localizados en diferentes cromosomas. La meiosis de una célula germinal heterocigota para ambos genes (AaBb), resulta en cuatro gametos, dos de ellos (50 %) de dotación génica tipo recombinante (Ab y aB).

Análisis del ligamiento genético

Mediante estudios citogenéticos de un gran número de meiosis se llegó a la conclusión de que el genoma humano tiene una longitud de 3.000 cM. Por otro lado, la estimación del tamaño físico de un genoma humano haploide es de 3.000 Mb, por lo que 1 cM corresponde, de media, a 1 Mb de DNA. En otras palabras, dos genes que disten entre sí alrededor de 1 Mb sufren recombinación en el 1 % de las meiosis. También podemos deducir que los cromosomas humanos, tienen una longitud genética de 100-300 cM¹⁴, por lo que en cada meiosis se establecen una-tres recombinaciones por cromosoma. Estas son, no obstante, cifras que resultarían reales sólo si todas las regiones del genoma sufrieran recombinación con igual frecuencia. Esto no es así, sino que diferentes regiones del genoma presentan frecuencias de recombinación distintas, y la frecuencia de recombinación durante la meiosis en células masculinas es diferente de la que ocurre en las femeninas. No obstante, nos da una idea global de la relación entre distancia física y distancia genética en el genoma humano.

Merece la pena revisar el concepto de *recombinación* genética en el que se fundamenta el *ligamiento genético*. Cada par de cromosomas homólogos se aparea durante la meiosis I y se establecen cruces (*crossing over*) entre las cromátidas con intercambio de material homólogo (fig. 2). En cada episodio de recombinación sólo se intercambian material dos de las cuatro cromátidas, por lo que el resultado final de la meiosis es un 50 % de no-recombinantes (del mismo tipo que el progenitor).

Dependiendo de la *posición* de un gen (A) respecto a

otro (B) se puede dar una gama de posibilidades, con dos situaciones extremas:

a) que A y B estén en cromosomas distintos (distancia infinita, al no estar en la misma molécula de DNA). En este caso, los genes se segregan con total independencia y un progenitor que sea AaBb (heterocigoto para ambos genes, cada gen tiene dos alelos: -mayúscula, -minúscula) genera gametos AB, Ab, aB y ab con igual frecuencia (fig. 3). Por tanto, las probabilidades de que un gameto tenga el alelo mayúscula en un gen y el minúscula en el otro son del 50 %;

b) que A y B estén tan próximos entre sí en un determinado cromosoma que las probabilidades de que un episodio de recombinación ocurra entre ellos son prácticamente nulas. En esta situación, los genes no se segregan independientemente sino «ligados» (fig. 4), de modo que todos los gametos son del mismo tipo que el progenitor y no existe ningún recombinante para estos dos loci.

Situaciones intermedias son aquellas en que A y B pertenecen al mismo cromosoma, pero la distancia entre ellos hace que en un porcentaje de meiosis determinado (tanto mayor cuanto más alejado estén) se produzca recombinación genética. Existe, pues, una relación entre el porcentaje de gametos de tipo distinto al progenitor (es decir recombinantes) y la distancia entre los genes. A una distancia tal que el 1 % de los gametos sea recombinante se denomina 1 cM. De esta manera, ligamiento genético se define como la tendencia de dos genes próximos en un cromosoma a transmitirse juntos a la siguiente generación.

El concepto de buscar una cosegregación de caracteres para poder asignar determinados genes a ciertos

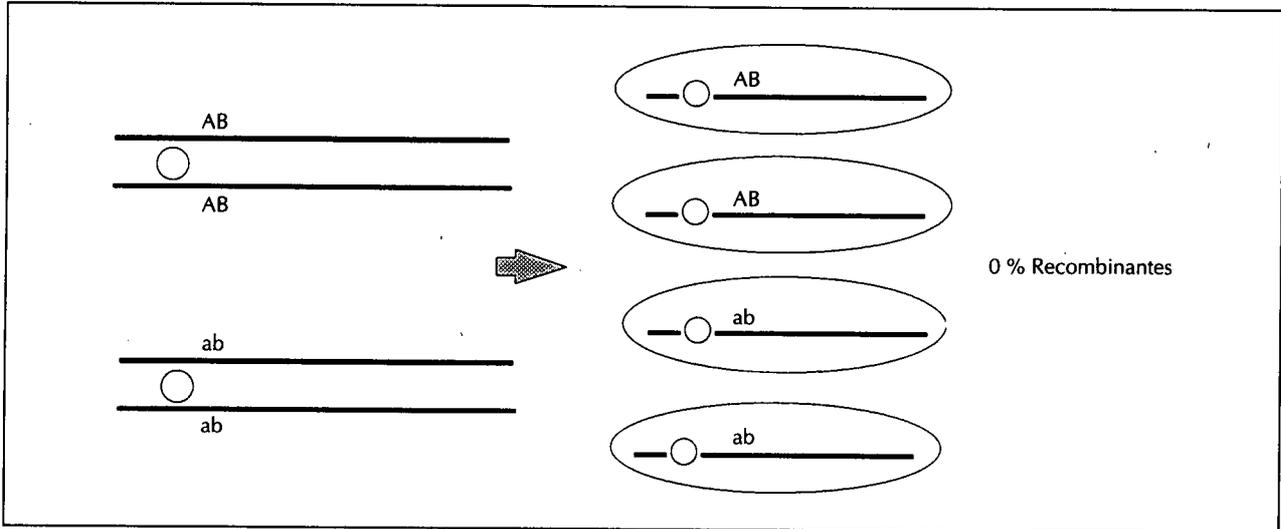


Fig. 4.—Consegregación de genes muy próximos entre sí. La meiosis de una célula germinal heterocigota para ambos genes (AaBb) resulta en un 100 % de gametos de tipo paterno, no observándose recombinaciones.

cromosomas, mapping genético en esencia, fue explotado durante años sin necesidad de nuevas tecnologías. Así, por ejemplo, el gen de la hemofilia fue localizado en el cromosoma X mediante el mero estudio del patrón de herencia. Algunos polimorfismos a nivel de proteínas permitieron también el asignar unos pocos genes a determinados cromosomas antes de que la denominada tecnología del DNA recombinante estuviera al alcance, pero es a partir del uso de sondas moleculares cuando crece exponencialmente el número de loci cuya posición en el genoma es conocida¹⁵. A lo largo de los últimos años se han clonado un gran número de fragmentos de DNA humano. Muchos de ellos no pertenecen a genes concretos, sino que se trata de secuencias «anónimas», pero su valor a la hora de hacer análisis de ligamiento genético es el mismo: una vez conocida su posición en el genoma, cualquier fragmento de DNA de secuencia única es un «marcador» de esa posición del genoma. Si este marcador existe en la población en varias formas alélicas, un análisis molecular en familias afectas permite, con ciertas limitaciones, alcanzar dos objetivos:

- 1) evaluar la distancia genética entre el marcador y el gen responsable de la enfermedad. Si la distancia es corta, el marcador se usa como herramienta para clonar el gen implicado;
- 2) diagnosticar qué miembros de la familia tienen el alelo mutado en uno o ambos cromosomas.

Requerimiento del análisis de ligamiento

No siempre que dispongamos de un marcador más o menos próximo al gen de interés podremos llegar a

conclusiones sobre la distancia entre ambos o usarlo en el diagnóstico de familias. Existen unas condiciones previas para que el análisis de ligamiento sea útil:

1. **Polimorfismo:** Para que podamos determinar el grado de ligamiento entre marcador y gen responsable de enfermedad es preciso que en la familia sujeto de estudio el progenitor portador sea heterocigoto tanto para el marcador como para el gen de la enfermedad. Polimorfismo se define como la presencia en la población de dos o más fenotipos alternativos, definidos genéticamente, con una frecuencia tal que el tipo menos corriente no se puede mantener por simple recurrencia de mutaciones*.

Consideremos un pedigree de la enfermedad de Huntington, autosómica dominante (fig. 5), y supongamos que el gen con la mutación de la enfermedad de Huntington (H) está presente en un alelo paterno. Si, con técnicas moleculares, podemos detectar un marcador (M) próximo al gen H y distinguir los dos alelos de M (1 y 2), la familia será informativa sólo si el padre es heterocigótico (1, 2), pero nunca si es homocigótico (1,1 o 2,2). Por este motivo, los marcadores ideales son aquellos que no sólo están próximos al gen en cuestión sino que son muy polimórficos en la población, de modo que la mayoría de los sujetos sean heterocigotos. En términos teóricos, éste no parece un requisito difícil de cumplir, si consideramos que entre los genomas de dos seres humanos una de cada mil bases es diferente. Un desarrollo técnico, el análisis de DNA tipo Sout-

* En la práctica se considera que un locus es polimórfico si el (los) alelo(s) menos corrientes tienen una frecuencia de 0,01 al menos, que resulta en una frecuencia de heterocigotos para este alelo superior al 2 % (2 pq, donde $p = 1 - q$ y $q = 0,01$).

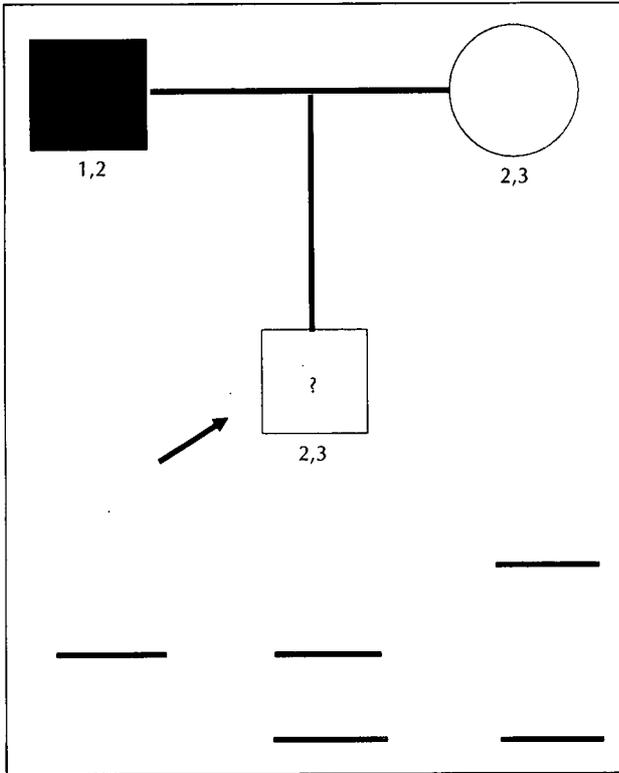


Fig. 5.—Pedigree no informativo debido a homocigosidad del padre afecto de una enfermedad autosómica dominante. A pesar de disponer de un marcador próximo al gen responsable de la enfermedad, resulta imposible predecir si el hijo (que presenta los alelos 1 y 2 del marcador en el análisis tipo Southern) será enfermo, debido a que el padre es homocigoto para el marcador: los dos alelos del marcador (uno en el cromosoma sano y otro en el afecto de la mutación responsable de la enfermedad) dan una misma banda en el Southern y no podemos diferenciar si el alelo 1 heredado por el hijo es el del cromosoma sano o enfermo.

hern, aplicado al estudio de polimorfismos fue el descubrimiento que hizo posible en la práctica, el análisis de ligamiento genético: *RFLP* (*restricción fragment length polymorphism*), *polimorfismos de longitud de fragmentos de restricción*. En un trabajo pionero se demostró que el análisis RFLP es tremendamente informativo en el diagnóstico de una «enfermedad molecular» por excelencia: la anemia falciforme¹⁶. Estos autores encontraron una diana de restricción para *Hpa*I (próxima al gen de la cadena beta de hemoglobina) que es polimórfica, de modo que un análisis de DNA, digerido con *Hpa*I e hibridado con sondas de la cadena beta de hemoglobina, genera bandas de varios tamaños (7 kb, 7,6 kb y 13 kb). Asimismo, demostraron que el alelo de 13 kb se asociaba, en el 60 % de los americanos de orígenes africanos, con la mutación Glu-Val responsable de la anemia falciforme (fig. 6).

Lógicamente, un marcador tiene mayores probabilidades de estar presente en estado heterocigoto (el alelo en un cromosoma es diferente al alelo en el cromoso-

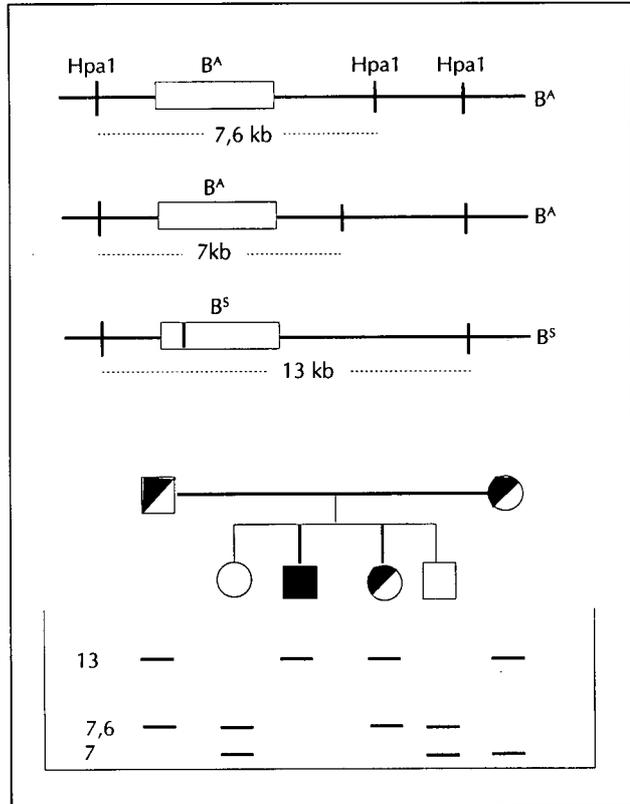


Fig. 6.—Ligamiento entre la mutación de la anemia falciforme (*B^S*) y ausencia de una diana para *Hpa*I próxima al gen de la cadena B de la hemoglobina. El análisis Southern de DNA cortado con *Hpa*I genera una banda de 7 o 7,6 kb en cromosomas normales cuando se hibrida con sondas del gen de la cadena *B^A*. En cambio, el cromosoma con la mutación de la anemia falciforme carece de la diana *Hpa*I, por lo que la banda correspondiente en la autorradiografía es de 13 kb. Esta asociación íntima entre la banda de 13 kb y la mutación Glu-Val en el gen de la cadena B de la hemoglobina se puede usar para determinar que entre los hijos de la familia estudiada hay un varón enfermo y una hembra portadora.

ma homólogo) cuanto más polimórfico sea, cuantas más formas alélicas estén presentes en la población. Es así que el nivel de heterocigosidad de un marcador se conoce como el contenido de información polimórfica (PIC): un valor PIC de 0,4 para un determinado marcador significa que las probabilidades de que un sujeto sea heterocigoto para este marcador son del 40 %. Los polimorfismos basados en la presencia o ausencia de una diana de restricción tienen un valor limitado porque frecuentemente detectan pocos alelos diferentes. Así, aunque se han aislado más de 3.000 marcadores polimórficos humanos¹⁷, el 90 % de ellos tienen PIC de menos de 0,5.

Existe otro tipo de polimorfismo que suele tener valores PIC más altos, y que se basa en la existencia de regiones genómicas con un número variable de repeticiones «en tándem» (VNTR) comprendidas entre dos

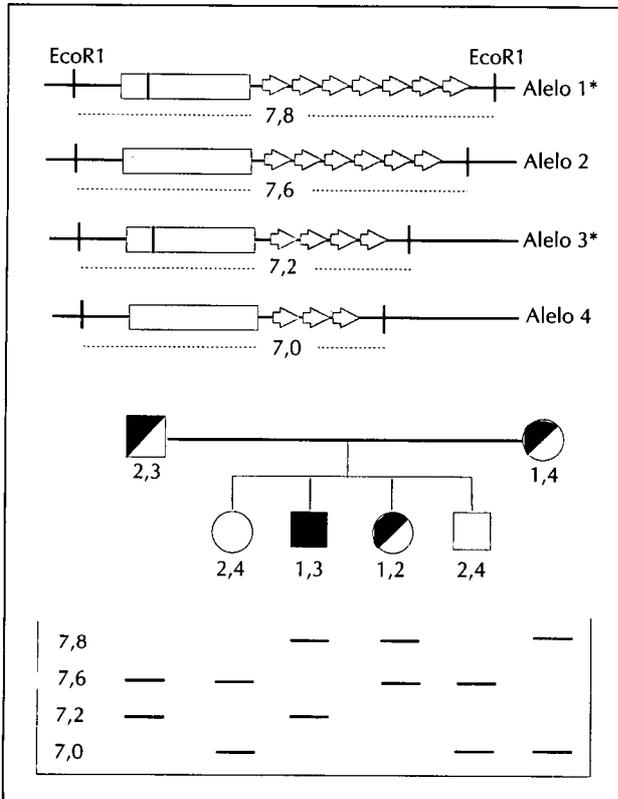


Fig. 7.—RFLP debido a la presencia de un número variable de repeticiones en tándem (VNTR). La digestión del DNA con EcoR1 e hibridación con una sonda próxima al gen de interés distingue cuatro alelos de distinto tamaño (7,0, 7,2, 7,6 y 7,8 kb) según el número de repeticiones en tándem presentes. En un ejemplo de enfermedad autosómica recesiva, resulta probable que los padres sean heterocigotos para el número de repeticiones, por lo que el análisis Southern con una sonda de un marcador próximo al gen responsable permite determinar que una hija es portadora (presenta el alelo de 7,8 kb, de origen materno) y un hijo es enfermo (presenta el alelo de 7,2 kb, paterno, y el de 7,8 kb, materno). La asociación entre los alelos de 7,2 kb y de 7,8 kb con la mutación en el gen responsable de la enfermedad ha sido establecida para esta familia analizando más miembros de la misma (no representados).

dianas de restricción¹⁸ (fig. 7). Estos RFLPs se denominan hipervariables, porque existen muchos alelos, con lo que las probabilidades de que los sujetos a estudio sean heterocigotos son altas.

Con el uso de la reacción en cadena de la polimerasa (PCR) ha surgido un nuevo método de detección de polimorfismos de muy alto PIC: PCR de regiones con repeticiones de dinucleótidos. Se ha comprobado que existen zonas del genoma cuya secuencia consiste en dos nucleótidos (usualmente CA) que se repiten un número n de veces, representado $(CA)_n$. Estas secuencias son frecuentes, se estima que hay una por cada 50 kb de DNA, y el número de repeticiones (n) es muy variable. Por tanto, una PCR que use cebadores (primers) de secuencia única localizados a ambos lados de la región

$(CA)_n$ genera bandas amplificadas de tamaño muy polimórfico. Esta técnica proporciona usualmente un valor PIC mayor a 0,7 y es la base de mapas de ligamiento genético de «segunda generación»¹⁹ (fig. 8).

2. Fase: Un segundo requerimiento para interpretar el ligamiento entre un marcador y un gen determinado consiste en conocer la fase, es decir conocer qué alelo del marcador está en el mismo cromosoma que el gen mutado. Alelos que están en el mismo cromosoma se dice que están en *fase acoplada* o *cis*, mientras que los alelos que están en el cromosoma homólogo al que lleva el gen mutado se dice que están en *fase de repulsión* o *trans*. La necesidad de conocer la fase se comprende fácilmente con un ejemplo de una de las enfermedades en que primero se llevó a cabo con éxito el mapping genético: la enfermedad de Huntington, autosómica dominante²⁰ (fig. 9). El padre es heterocigoto

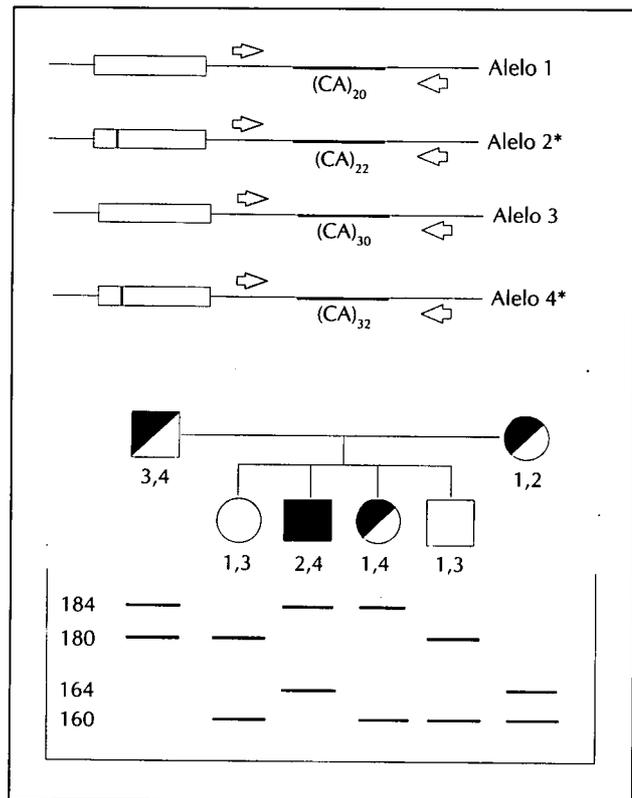


Fig. 8.—PCR de regiones con repeticiones de dinucleótidos CA. Debido a la existencia de un número variable de repeticiones del dinucleótido CA (20, 22, 30 y 32 en este ejemplo), se pueden detectar numerosos alelos. Una vez conocida la fase (CA_{22} y CA_{32} se asocian a la mutación en el gen próximo, responsable de la enfermedad en esta familia), una PCR con primers específicos para la región en torno a la zona con repeticiones permite determinar que uno de los hijos es enfermo por tener los alelos de 184 bp (paterno) y 164 bp (materno), mientras que una hija es portadora de esta enfermedad autosómica recesiva, ya que presenta el alelo de 184 bp.

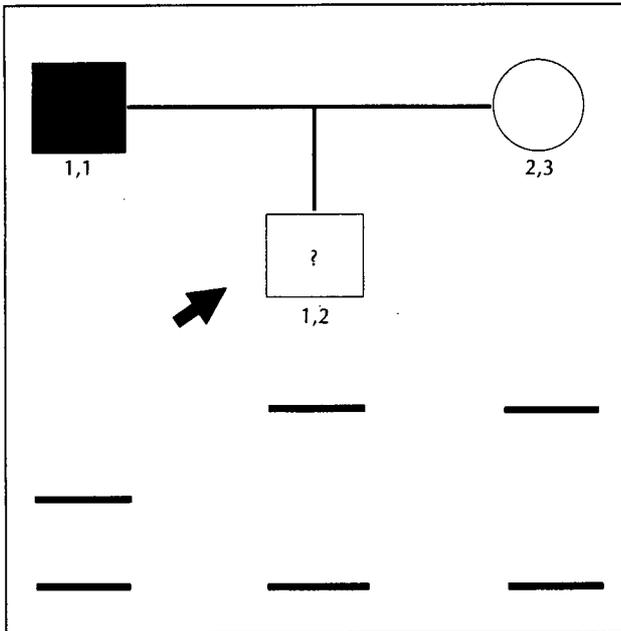


Fig. 9.—Estudio no informativo por desconocerse la fase. Aunque el padre, enfermo de una entidad autosómica dominante, es heterocigoto para el marcador (1,2), no es posible conocer cuál de los dos alelos pertenece al cromosoma con la mutación. Por consiguiente, no se puede predecir si el hijo (2,3) será enfermo o sano.

para el marcador ^(1,2) y del estudio del pedigree se deduce que el hijo heredó el alelo 2 del padre, pero no existe suficiente información para concluir qué alelo paterno (1 o 2) es el que está en el mismo cromosoma que la mutación de corea de Huntington. Para determinar la fase hace falta examinar un número suficiente de miembros de la familia, de ahí que familias numerosas sean más valiosas que familias pequeñas. Por el mismo motivo, familias donde se puedan analizar tres generaciones son más útiles que las que sólo tienen dos generaciones disponibles. Si, por ejemplo, existiera otro hermano ^(1,3) que sabemos es enfermo (fig. 10), podemos concluir que el alelo 1 está en el mismo cromosoma que el gen de la enfermedad de Huntington. Es posible, pues, diagnosticar al caso problema, de alelos 2,3 como normal.

Con excepción de la región pseudoautosómica en los extremos Xp y Yp, el X del varón no sufre recombinación durante la meiosis; por lo que para enfermedades ligadas al X, el genotipo del abuelo materno es particularmente informativo para establecer la fase (figura 11).

Ahora bien, es importante tener presente que la asociación de un cierto alelo del marcador con la mutación en un gen concreto es, en principio, válida sólo para esa familia. Así, es posible que si existen cuatro alelos diferentes (1, 2, 3, 4) para un marcador (G8) próximo al gen de Huntington, la mutación causante de

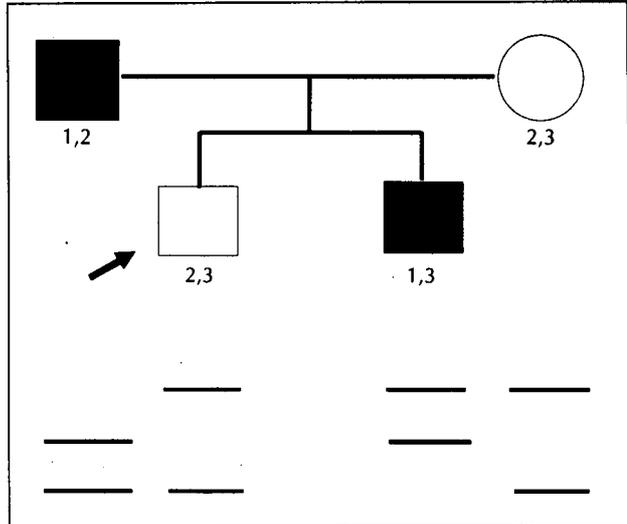


Fig. 10.—Estudio informativo (fase conocida). La información obtenida mediante el análisis de otro hermano (1,3), enfermo, permite establecer que el alelo 1 del marcador está en el mismo cromosoma que la mutación responsable de la enfermedad. Por tanto, podemos concluir que el probando (2,3) será sano.

la enfermedad vaya asociada al alelo 1 en una familia, pero al alelo 2, 3 o 4 en otras. Por tanto, la fase ha de establecerse para cada familia sometida a estudio, y teóricamente, las probabilidades de que se observe una cierta asociación es el producto de las probabilidades de los alelos (equilibrio de ligamiento). En ciertos casos, la mutación causante de la enfermedad se asocia preferentemente a un alelo concreto del marcador, que

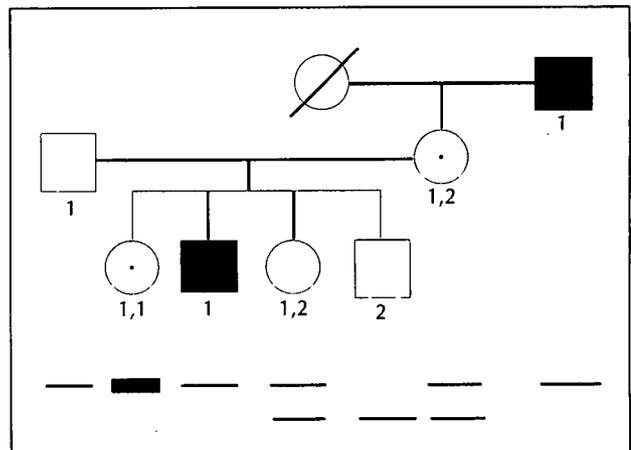


Fig. 11.—Ejemplo de enfermedad recesiva ligada al cromosoma X. El conocimiento del haplotipo del abuelo materno (1) permite establecer la fase: mutación en cis con alelo 1 y, consiguientemente, los estados de portador (), enfermo () o sano () entre los hijos de la portadora (1,2).

no tiene que ser el más frecuente, en cuyo caso se habla de *desequilibrio de ligamiento*. Un ejemplo típico de *desequilibrio de ligamiento* es el observado en la fibrosis quística, donde un haplotipo del marcador D7S23 está presente en el 25 % de la población normal, pero en hasta un 90 % de cromosomas con una mutación de fibrosis quística²¹. Este caso, así como el ya comentado de la asociación del alelo 13 kb (Hpa1) con la mutación de la anemia falciforme en el gen de la cadena β de la hemoglobina¹⁶, suelen ser originales de un sector de la población: población blanca norteeuropea en el ejemplo de fibrosis quística y población negra de África Occidental en el caso de la anemia falciforme. La explicación más aceptada para el fenómeno de *desequilibrio de ligamiento* es la existencia de un origen común para todos esos cromosomas con la mutación, que se asocia a un cierto haplotipo. Este haplotipo sería el que coincidió en tener el sujeto en que se originó la mutación de fibrosis quística o anemia falciforme, y los cromosomas presentes en la actualidad descienden de este ancestro común. En determinados casos, como la anemia falciforme, parece que conocemos el porqué de que esta mutación se extendiera en la población: los sujetos heterocigotos portadores de esta anemia falciforme son resistentes al paludismo, lo que debió proporcionarles una ventaja evolutiva en regiones como África Occidental, endémica de paludismo. En consonancia con esta explicación, marcadores más próximos a la mutación presentan un gran *desequilibrio de ligamiento* debido a que es muy poco probable que exista un fenómeno de recombinación (*crossing-over*) en la distancia entre marcador y gen mutado. Cuanto mayor es la distancia entre marcador y gen mutado, más alta es la probabilidad de recombinación, por lo que a lo largo de generaciones se «diluye» el *desequilibrio de ligamiento* hasta presentarse una situación de equilibrio para ese marcador. Así, los marcadores MET y D7S8, aproximadamente a 1 Mb a un lado y otro del gen de la fibrosis quística, muestran un ligero *desequilibrio de ligamiento*, comparado al más fuerte *desequilibrio de ligamiento* que se observa con los marcadores D7S122 y D7S23, localizados a unas 200 kb y 100 kb del gen de la fibrosis quística, respectivamente²¹.

Un problema que se plantea a veces al intentar establecer la fase en una familia es la *aparición de alelos nuevos*, no presentes en los padres. Supongamos la familia con enfermedad de Huntington ilustrada en la figura 12, cuyo DNA ha sido analizado mediante PCR de una región (CA)_n próxima al gen de Huntington y en el hijo encontramos un alelo 32 (32 repeticiones CA) que no está presente en los padres. Existen dos posibilidades que explican este hallazgo: 1. *Falsa paternidad*: que el padre biológico es distinto del padre social. 2. *Nueva mutación*: que el alelo 32 se ha originado *de novo*. Ahora bien, las mutaciones *de novo* aparecen a una frecuencia de 10⁻⁵-10⁻⁶ para la mayoría de los loci, por lo que ante la aparición de un alelo nuevo

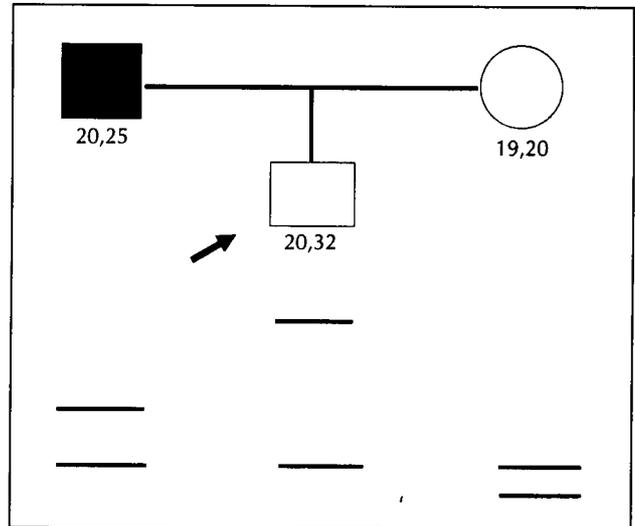


Fig. 12.—Aparición de un alelo (32) no presente en los padres, en una familia con enfermedad autosómica dominante. De las posibles causas: error en la muestra, falsa paternidad, mutación de novo..., esta última suele ser la menos probable.

hay que destacar, en primer lugar, que se trata de un caso de falsa paternidad.

Medición de ligamiento genético

Una vez que el estudio del DNA de una familia con un marcador polimórfico permite establecer la fase, se puede determinar el número de sujetos que han heredado cromosomas con recombinaciones entre el marcador y el gen responsable de la enfermedad. Por ejemplo, consideremos la figura 13, que representa una familia con enfermedad autosómica dominante, analizada con un marcador polimórfico. Resulta claro que el alelo 1 está acoplado, en cis, al gen mutado responsable de la enfermedad. Además, entre los 10 hermanos de la tercera generación, encontramos tres individuos que heredaron un cromosoma que había sufrido recombinación entre el marcador y el gen mutado durante la espermatogénesis en el padre. Por tanto, la frecuencia de recombinación observada en esta familia entre el marcador y el gen responsable es de 3/10. En el caso que ambos loci estuviesen distantes, no sería de esperar que se segregaran juntos en la meiosis y, consecuentemente, en un 50 % de los hijos el alelo 1 estaría asociado al cromosoma con la mutación y en el otro 50 % no. Es importante tener presente que el análisis de ligamiento genético se basa en la clasificación de los miembros de la familia como recombinantes (R) y no-recombinantes (NR). A su vez, cada individuo es clasificado en función de la correlación fenotipo (del gen estudio)-genotipo (del marcador molecular). Por tanto, los parámetros que añaden ambigüedad al esta-

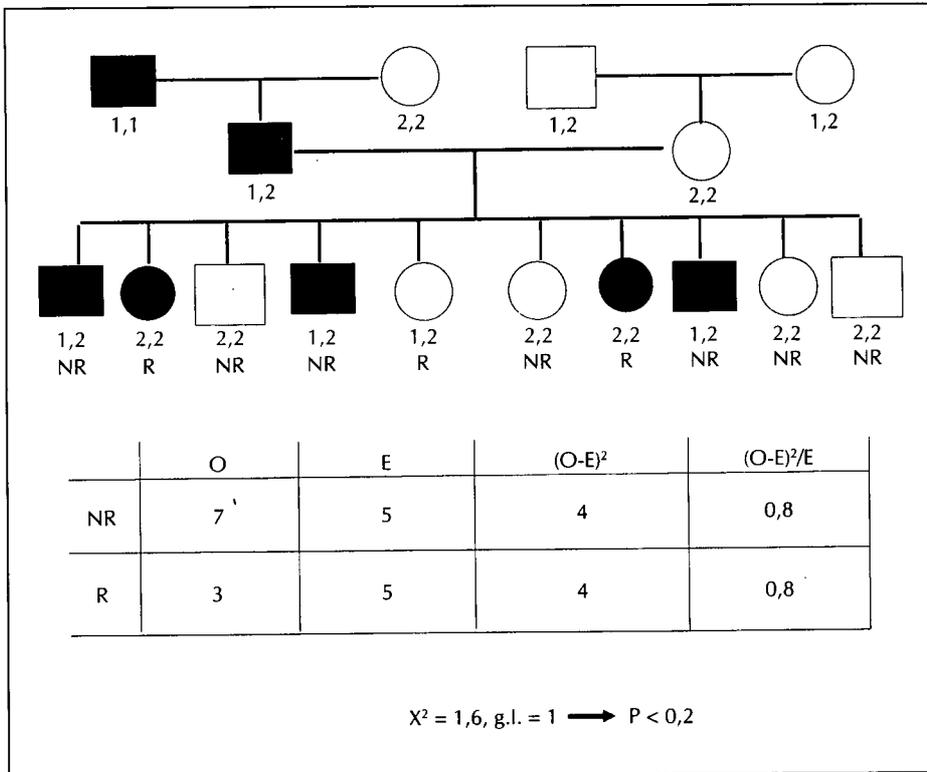


Fig. 13.—Análisis de una gran familia con enfermedad autosómica dominante con un marcador próximo al gen responsable de la enfermedad. La fase, mutación en cis con alelo 1 paterno, queda establecida tras estudio de los abuelos (abuelo paterno 1,1). Por consiguiente, los hijos que tengan el alelo 1 y no sean enfermos, así como los que sean enfermos y no tengan el alelo 1, serán recombinantes. Se puede construir una tabla para análisis chi-cuadrado, donde contrastamos los números observados (siete no recombinantes, tres recombinantes) con los esperados si no existiera ligamiento (segregación independiente entre marcador y gen responsable de la enfermedad: 50 % recombinantes, 50 % no-recombinantes). El valor chi-cuadrado obtenido, 1,6, para un grado de libertad arroja una $p < 0,2$.

blecimiento de esta correlación constituyen las mayores limitaciones del análisis de ligamiento:

1. *Heterogeneidad genética:* Existen situaciones en que la mutación de uno cualquiera de varios genes (usualmente componentes de una misma vía metabólica) resulta en un fenotipo prácticamente idéntico. Así, por ejemplo, existen varios genes cuya mutación resulta en el síndrome de Sanfilippo, una mucopolisacaridosis con acúmulo de heparán sulfato. Evidentemente, si analizamos varias familias con el mismo marcador y resulta que alguna de ellas presenta mutación en un gen completamente distinto, tendrá un efecto negativo sobre el análisis de ligamiento. La ADPKD es un ejemplo típico de heterogeneidad genética: los primeros estudios de familias que demostraron ligamiento significativo localizaron el gen de la ADPKD en el brazo corto del cromosoma 16 (16 p 13.3)²², mientras que trabajos posteriores han identificado familias en que la enfermedad no se segrega con marcadores del cromosoma 16, a pesar de que se trata de casos de ADPKD indistinguibles de los anteriores desde el punto de vista fenotípico²³. Es posible, pues, que varios genes y/o factores no genéticos puedan intervenir en el desarrollo de un fenotipo de ADPKD.

2. *Penetrancia:* La presencia de una mutación en un gen responsable de enfermedad no siempre se traduce en un fenotipo anormal. Es un hecho particular-

mente importante en enfermedades dominantes de presentación clínica en el adulto tales como la referida enfermedad de Huntington o la misma ADPKD, donde con frecuencia encontramos sujetos asintomáticos que presentan la mutación. Supongamos que la enfermedad de interés tiene una penetrancia del 80 %. En este caso, un 20 % de los individuos que en el estudio de ligamiento se interpretan como recombinantes (positivos para el marcador pero sin enfermedad aparente) son realmente no-recombinantes. ¿Cómo podemos analizar el ligamiento cuando no estamos seguros de que hayamos identificado verdaderos recombinantes mediante la observación directa fenotipo-genotipo?

Otra situación que representa ambigüedad es el estudio de familias en que no se puede establecer la fase con total garantía. En el pedigree de la figura 14, la familia afecta de una enfermedad autosómica dominante es investigada con un marcador que presenta dos alelos (1 y 2). La madre enferma es heterocigota (1, 2), pero no poseemos información de suficientes familiares para establecer qué cromosoma lleva la mutación responsable de la enfermedad: si consideramos el alelo 2 en cis con la mutación, ninguno de los hijos es recombinante, mientras que si resulta que el alelo 1 es el que está en el mismo cromosoma que la mutación, todos los hijos son recombinantes. Como en otras ciencias experimentales, estas ambigüedades se afrontan utilizando métodos estadísticos que nos permiten buscar asocia-

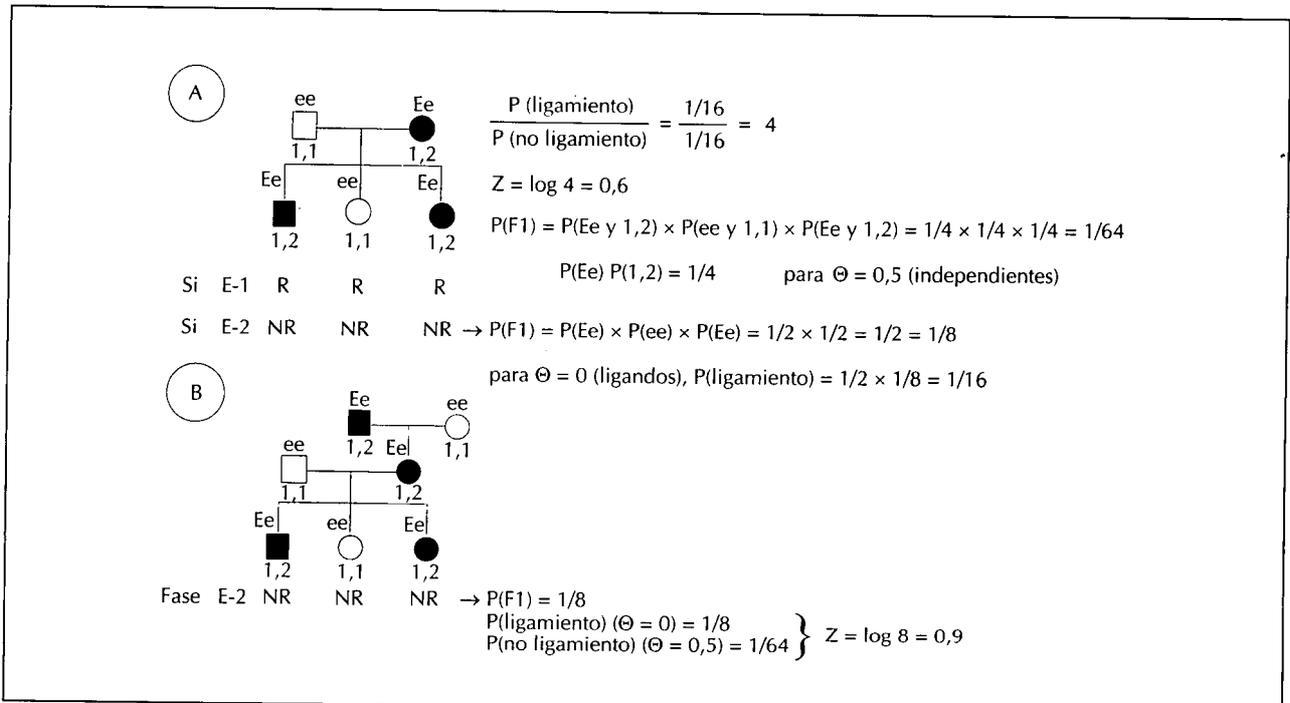


Fig. 14.—Efecto negativo del desconocimiento de la fase en el análisis del ligamiento. El pedigree de una familia con enfermedad autosómica dominante investigada con un marcador muy próximo al gen responsable no permite conocer la fase cuando sólo contamos con DNA de los padres y los hijos (A): tan probable es que el alelo 1 esté en cis con la mutación (E), como que lo esté el alelo 2. Si el primer caso es cierto (E-1), los tres hijos serían recombinantes, mientras que si la fase es E-2, ninguno de los hijos sería recombinante. Si asumimos la fase E-2 (con una probabilidad, a priori, 1/2), la probabilidad de que exista ligamiento (para $\Theta = 0$, no recombinantes) entre marcador y gen responsable es el producto de las probabilidades de cada genotipo encontrado: $P(Ee) \times P(ee) \times P(Ee) = 1/2 \times 1/2 \times 1/2 = 1/8$. Como esta es una de las dos posibles fases, la probabilidad de ligamiento es $1/2 \times 1/8 = 1/16$. La probabilidad de no-ligamiento ($\Theta = 0,5$) es $P(Ee \text{ y } 1,2) \times P(ee \text{ y } 1,1) \times P(Ee \text{ y } 1,2)$. Puesto que en este caso (no-ligamiento) existe segregación independiente, $P(Ee \text{ y } 1,2) = P(Ee) \times P(1,2) = 1/4$. Por tanto, $P(\text{no ligamiento}) = 1/4 \times 1/4 \times 1/4 = 1/64$. Si calculamos el cociente de las probabilidades = $1/16$ entre $1/64 = 4$. Por tanto, el LOD, $Z = \log 4 = 0,6$. Supongamos que tenemos acceso al DNA de los abuelos (B), y podemos establecer que la fase es E-2 con seguridad. Ahora, $P(\text{ligamiento}) = 1/8$ y el cociente de las probabilidades = $1/8$ entre $1/64 = 8$, por lo que $Z = 0,9$.

ciones significativas en los datos obtenidos. El método estadístico diseñado para comprobar lo significativo de aparentes asociaciones fenotipo-genotipo es el análisis de la probabilidad de ligamiento²⁴.

Análisis de la probabilidad de ligamiento

Con un programa de ordenador (LINKAGE, por ejemplo) se calcula cómo de probable es que se den los resultados observados en el estudio fenotipo-genotivo de la familia si el marcador utilizado (M) está a una distancia genética entre 0 y 50 cM del gen responsable de la enfermedad (E), es decir para fracciones de recombinación (Θ) entre 0 y 0,5. Estas probabilidades, calculadas para cada valor de Θ , se comparan con la probabilidad de que se den los resultados observados si no existiera ligamiento entre M y E, es decir que la distancia entre M y E fuese tan grande que se segregarían

independientemente en la meiosis resultando en un 50 % de gametos recombinantes ($\Theta = 0,5$). La división de la probabilidad de ligamiento partido por la probabilidad de no ligamiento (odds ratio) se calcula para cada valor posible de Θ y se presenta como el logaritmo decimal del cociente, lo que se denomina valor LOD (log odds) o Z. Valores positivos de Z sugieren que existe ligamiento entre M y E, con tanta mayor probabilidad cuanto mayor sea Z. En cada familia analizada, los valores Z obtenidos al realizar los cálculos para cada valor de Θ varían en torno a un punto donde Z es máximo: existe una probabilidad máxima de ligamiento si M y E se encuentran a esa distancia genética (valor Θ para el que Z es máximo). No es raro que los valores para los que Z es máximo varíen algo entre distintas familias, pero como Z es un logaritmo se puede calcular la suma de los Z para cada Θ en tantas familias genéticamente homogéneas como sea posible. En la práctica se acepta que un valor $Z > 3$ es evidencia

suficiente de ligamiento entre dos loci²⁵ con una significación estadística de $p < 0,05^*$.

Hay que tener presente que el valor LOD puede tomar valores + o -, de modo que al estudiar numerosas familias y representar la suma de los valores Z para cada Θ , el resultado es una curva que asciende desde valores negativos o próximos a 0 hasta el Z máximo y luego desciende otra vez hasta 0 o valores negativos (fig. 15).

Es importante encontrar marcadores que se localicen a ambos lados del gen de interés, de modo que acotemos la posición del mismo. Supongamos que el marcador A está a 20 cM ($\Theta = 0,2$) del gen causante de enfermedad (E) y que otro locus marcador (B) está a 10 cM ($\Theta = 0,1$) de E. Si calculamos el valor de Θ entre A y B para el que Z es máximo podremos ordenar los tres loci: si resulta que A dista 25 cM de B, el orden es A (20 cM), E (10 cM), B. Con frecuencia, las distancias calculadas por ligamiento genético no son estrictamente aditivas (el cálculo de distancia entre A y B resulta 25 cM en lugar de 30 cM). De hecho, es conveniente estimar las distancias entre varios puntos al mismo tiempo, una estrategia que se conoce como *análisis del ligamiento de múltiples puntos*. Se basa en el principio de que las probabilidades de que un mismo cromosoma sufra dos recombinaciones (*cross-over*) es mucho menos frecuente (la raíz cuadrada) de que sólo sufra un *cross-over*. Entonces, un programa de computador busca el orden de loci que encaja con los pedigreos estudiados con un número mínimo de *cross-over* dobles.

La obtención de mapas con múltiples puntos a ambos lados del gen de interés es un paso muy importante en el logro de los dos objetivos fundamentales del análisis de ligamiento:

1. Clonación posicional de genes: Una vez acotados por marcadores cada vez más próximos a ambos lados es posible pasar a intentar encontrar clonas en genotecas humanas que contengan parte del gen usando el marcador como sonda. La clonación posicional de genes responsables de enfermedades humanas normal-

* Un Z = 3 indica que las probabilidades de que los resultados se presenten si los loci están a una distancia Θ (el valor de Θ al que Z es máximo) son mil veces superiores ($\log 1.000 = 3$) a que se presenten si no existe ligamiento ($\Theta = 0,5$). Aunque, a primera vista este umbral de significación parece muy exigente, en realidad, comparando la fuerza estadística de este valor Z con el parámetro P corrientemente usado en análisis estadísticos, un Z = 3 equivale aproximadamente a un $p = 0,05$. En efecto, Z = 3 indica que la probabilidad de ligamiento calculada es $999/1.000$ y la probabilidad de no ligamiento es $1/1.000$. Pero, *a priori*, cualquier marcador tiene alrededor de $1/50$ probabilidades de localizarse en un cierto cromosoma (46 cromosomas en humanos), mientras que la probabilidad *a priori* de no localizarse en este cromosoma sería $49/50$. Si calculamos la probabilidad conjunta (que esté en un cromosoma concreto y que se segregue ligado al gen analizado): ligamiento = $1/50 \times 999/1.000 = 1/50$ aproximadamente; no ligamiento = $49/50 \times 1/1.000 = 1/1.000$ aproximadamente. ¿Cuál es pues el error que cometemos al aceptar que el marcador está ligado al gen analizado por el hecho de haber obtenido un Z = 3? La probabilidad p de que el marcador no esté ligado a la enfermedad para Z = 3 es $1/1.000$ dividido por $1/50 + 1/1.000$, es decir, $0,05$ ($p = 0,05$).

Θ	0,1	0,2	0,3	0,4	0,5
Familia					
A	0,2	1	3	0,8	0,01
B	1	3	2	0,5	0
C	-0,2	0,8	1,1	1,2	-0,1
	1,0	4,8	6,1	2,5	-0,09

Fig. 15.—Cálculo de la distancia genética entre dos loci (marcador y gen responsable de enfermedad) tras análisis de ligamiento en tres familias afectas. Aunque, individualmente, cada familia arroja un valor Z máximo a fracciones de recombinación diferentes, la suma de los logaritmos proporciona un valor máximo (+6,1) para $\Theta = 0,3$. Podemos concluir que el gen de interés se encuentra con toda probabilidad a unos 30 cM del marcador.

mente requiere la colaboración multicéntrica para compartir el número limitado de familias afectas (con diagnóstico inequívoco) y los marcadores que han sido generados en diferentes laboratorios a lo largo de los años.

2. Diagnóstico indirecto de enfermedades genéticas.

Diagnóstico indirecto de enfermedades genéticas mediante análisis del ligamiento

En el momento presente se conocen abundantes marcadores en torno a la inmensa mayoría de las enfermedades monogenéticas, por lo que es posible el diagnóstico molecular de las mismas incluso cuando el gen responsable no haya sido clonado aún. En el ejemplo de la enfermedad autosómica dominante representado en la figura 16, en que el hijo es sometido a estudio para determinar su riesgo de padecer la enfermedad. Sólo si conocemos el orden de los marcadores en el mapa genético podemos llegar al diagnóstico.

Ahora bien, sólo cuando la estructura y secuencia del gen responsable de la enfermedad son conocidas se puede hacer un diagnóstico directo de las enfermedades. El análisis con marcadores próximos al gen de interés proporciona un *diagnóstico indirecto*, y existe la posibilidad de que se produzca recombinación entre el marcador más próximo y el gen de la enfermedad. Así, por ejemplo, consideremos la familia en la figura 17, en que usamos un marcador (M) de alelos 1 y 2 para el diagnóstico de una enfermedad autosómica dominante (E) y sabemos que la distancia entre (M) y (E) es de 10 cM (estudios de ligamiento genético en familias indican que Z es máximo a un $\Theta = 0,1$). El caso problema es el hijo en la tercera generación, cuyo genotipo

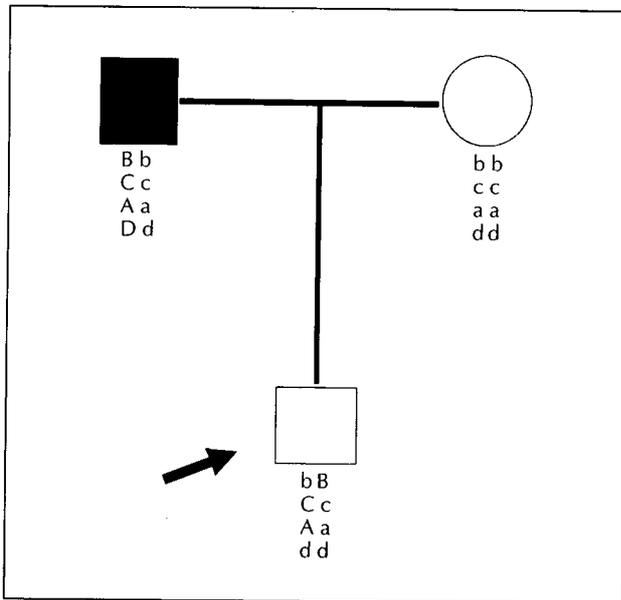


Fig. 16.—Diagnóstico indirecto de una enfermedad autosómica dominante con múltiples marcadores polimórficos. Una vez conocido el orden de los loci entre sí (BCAD) y con respecto al gen responsable de la enfermedad (por análisis de ligamiento) (BCEAD), así como la fase (BCAD-E), podemos concluir que el cromosoma bCAD del hijo ha sufrido dos fenómenos de cross-over (b-C y A-d), pero, puesto que heredó los alelos del cromosoma afecto para los loci que flanquean el gen responsable de la enfermedad (C y A), es de suponer que también heredó la forma mutada de este (E) y será enfermo.

para el marcador es 2,2. Se ha establecido la fase E-2, y puesto que el marcador dista 10 cM del gen involucrado, en cada meiosis hay un 10 % de probabilidades de recombinación entre ambos loci. Teniendo en cuenta estos porcentajes, podemos asegurar que en el 82 % de los casos, un hijo con alelos 2,2 tendrá la mutación (E). El diagnóstico indirecto de enfermedades genéticas mediante análisis de ligamiento se lleva a cabo en la práctica con marcadores que estén muy próximos al gen responsable de la enfermedad (5 cM o menos) para poder emitir diagnósticos correctos con una probabilidad por encima del 90 %.

II. MAPPING FISICO

Consiste en la disección del genoma humano en porciones progresivamente más simples, susceptibles de análisis bioquímico, con el objetivo de conocer la posición relativa de genes (que se expresa en unidades de distancia física: Kb, Mb) y finalmente, clonarlos.

Células híbridas somáticas

Un primer paso, fundamental para el desarrollo de la estrategia de mapping físico fue el descubrimiento de

que el material de una especie de mamífero podía ser transferido a células de otra especie cultivada *in vitro*: transferencia génica en células somáticas. Se consiguen así líneas celulares (híbridos de células somáticas) de ratón o hamster que contienen fracciones del genoma humano: unos pocos cromosomas o un solo cromosoma. Una variante técnica muy útil para acotar la localización de un gen consiste en romper el DNA humano mediante radiación previamente a la fusión celular, con lo que se obtienen híbridos (conocidos como híbridos de radiación) que contienen sólo fracciones de cromosomas humanos. Con el DNA de estos híbridos de células somáticas, correctamente catalogados, se confeccionan paneles de Southern que se someten a hibridación con DNA del gen cuya localización deseamos conocer. También se pueden generar células híbridas que contengan cromosomas humanos con anomalías estructurales, translocaciones por ejemplo, en cuyo caso el material genético humano procede de porciones de dos cromosomas distintos. Los casos con anomalías cromosómicas tipo translocación, deleción, etcétera, han resultado ser extremadamente útiles en la clonación de genes. Una alteración cromosómica que se asocie consistentemente a un cierto fenotipo orienta la búsqueda de genes hacia la región cromosómica donde acontece la anomalía (*breakpoint* de una translocación, por ejemplo).

Casos particularmente informativos son las translocaciones entre el cromosoma X y un autósoma. El estudio molecular de estas translocaciones ha facilitado la clonación de numerosos genes del cromosoma X, tales como el de la distrofia muscular de Duchenne/Becker (DMD-26), coirodermia²⁷, y síndrome de Lowe¹¹. Los cromosomas X en exceso de uno por célula están sometidos a bloqueo transcripcional (ionización o inactivación random del cromosoma X), de modo que los mamíferos hembra son mosaicos que expresan sólo el X paterno en unas células y el X materno en otras. Sin embargo, en una mujer que presente una translocación X-autosómica, sólo se expresa el cromosoma con la translocación en todas las células (inactivación sesgada). Por tanto, si la translocación destruye un gen como el de la DMD, las mujeres portadoras sufren la enfermedad. En definitiva, una estrategia muy fructífera de clonación de genes del X consiste en identificar mujeres que expresen la enfermedad, determinar si presentan translocaciones que afecten al X y, si es así, clonar el *breakpoint* de la translocación porque allí debe estar el gen responsable de la enfermedad.

Dosis génica

Otro procedimiento útil para localizar una secuencia de DNA en el genoma humano consiste en realizar Southern con DNA de pacientes con un número variable de cromosomas. En el caso del cromosoma X,

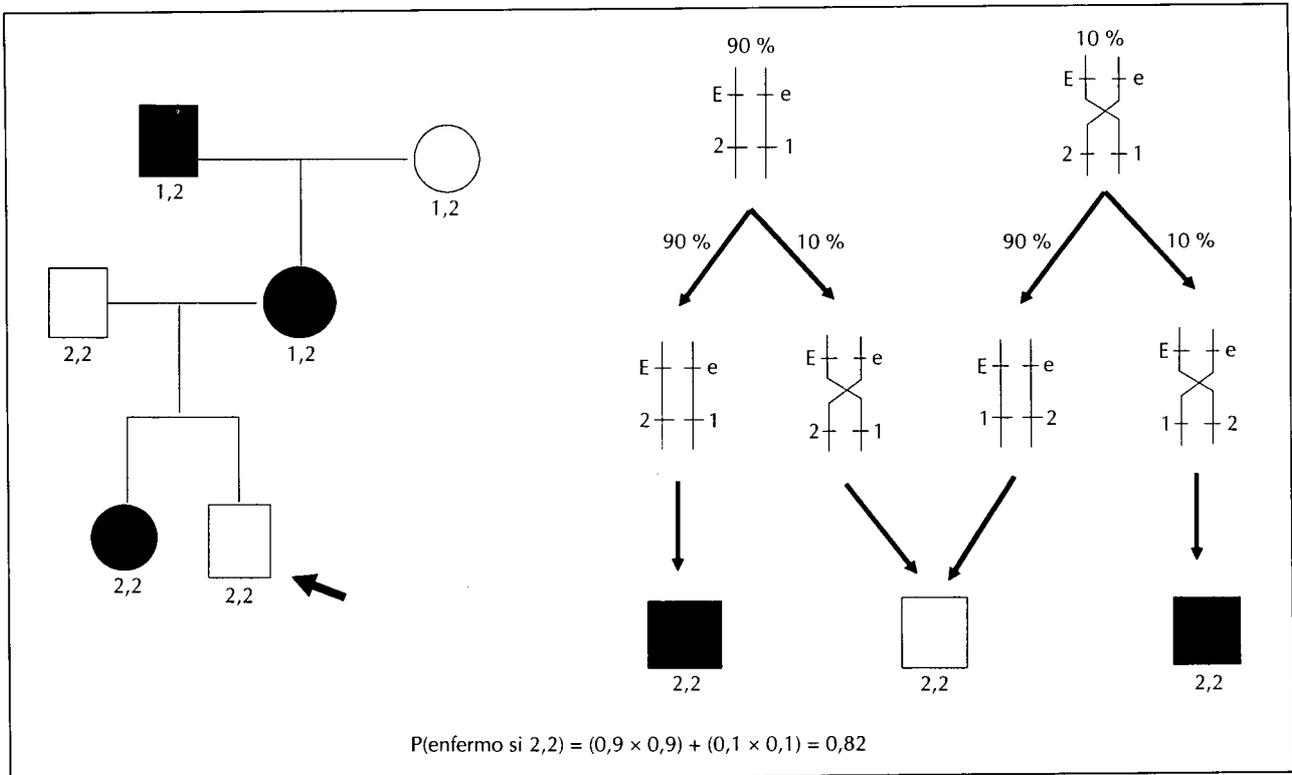


Fig. 17.—Diagnóstico indirecto de una enfermedad autosómica dominante con un marcador polimórfico situado a 10 cM del gen responsable de la enfermedad. Una hermana (2,2) está afectada de la enfermedad, así como la madre (1,2) y el abuelo materno (1,2). Puesto que los abuelos maternos son ambos 1,2, no sabemos qué cromosoma del abuelo está asociado a la enfermedad. No obstante, del estudio de la generación III parece obvio que es el cromosoma del alelo 2 en la madre el que se asocia a la enfermedad. Si el marcador estuviese infinitamente próximo al gen de la enfermedad podríamos asegurar que el probando (2,2) ha de ser enfermo. Pero la distancia entre M y E es 10 cM y, por tanto, hay un 10 % de probabilidades de que el alelo 2 no esté en el mismo cromosoma que la mutación E en la madre (1,2). Es decir el alelo 2 se asocia al gen E mutado una probabilidad del 90 %. Puesto que durante la meiosis materna puede darse un cross-over entre el alelo 2 y el gen mutado E con una probabilidad del 90 % otra vez, al final tenemos que si el hijo (probando) es 2,2, podemos concluir que será afecto de la enfermedad con una probabilidad del 82 % [$0,82 = (0,9 \times 0,9) + (0,1 \times 0,1)$].

por ejemplo, existen líneas celulares de pacientes con aberraciones cromosómicas tales como 47,XXX, 49,XXXXX, etc., cuyo DNA es procesado mediante la técnica de Southern. Si un fragmento de DNA (sonda) se localiza en el X, a igualdad de DNA total por muestra, la señal autorradiográfica será más intensa cuantos más cromosomas X posean las células.

Hibridación *in situ*

Un método relativamente simple para establecer la posición de un fragmento de DNA en el genoma es la hibridación de dicho fragmento, previamente marcado con un fluorocromo, a extensiones de cromosomas en portaobjetos. Aunque se trata de una técnica laboriosa, la hibridación *in situ* es, en muchas ocasiones, una primera aproximación al mapping físico con una resolución de hasta 1-2 Mb, que posteriormente se refina mediante otros procedimientos.

Citometría de flujo

Unos pocos laboratorios se han especializado en separar cromosomas humanos mediante citometría de flujo. Las células se tiñen con colorante intercalares, que se unen al DNA de modo estequiométrico, y los cromosomas en metafase se hacen pasar por delante de un haz láser. La intensidad de la emisión fluorescente generada por el colorante es proporcional al tamaño del cromosoma y, por tanto, se pueden purificar fracciones enriquecidas en un solo cromosoma. De este modo se han preparado genotecas humanas que contienen mayoritariamente fragmentos de DNA de un cromosoma determinado.

Fotólisis y micromanipulación

Recientemente se han propuesto otras tecnologías sofisticadas para aislar una pequeña porción de un cro-

mosoma con el fin de generar marcadores específicos de esa región. En este sentido, algunos laboratorios han diseñado procedimientos que, mediante micromanipulación a través del microscopio, consiguen tomar DNA de una banda concreta de un cromosoma. Estas cantidades mínimas de DNA son amplificadas por PCR para generar marcadores de esta región. Un procedimiento alternativo usa rayos láser para vaporizar la mayoría del cromosoma y dejar intacto sólo determinadas regiones del mismo, definidas bajo el microscopio, de las que se generan marcadores mediante PCR.

Electroforesis en campo pulsátil

Mientras que la electroforesis convencional en geles de agarosa permite separar fragmentos de DNA de hasta 50 kb, el uso de campos eléctricos que varían la dirección periódicamente hace posible resolver fragmentos de DNA de hasta 9 Mb. Estos grandes fragmentos de DNA se generan mediante digestión con enzimas de restricción que reconocen dianas muy poco frecuentes en el genoma. El DNA genómico, tratado especialmente para no cizallarlo, es sometido a digestión con estas enzimas y, tras electroforesis en campo pulsátil (PFGE), a hibridación tipo Southern. Si dos marcadores se hibridan al mismo fragmento de DNA concluimos que están ligados físicamente, en una distancia máxima equivalente al tamaño de la banda autorradiográfica detectada.

Clonación en cromosomas artificiales de levadura (YAC)

Numerosos laboratorios están en la actualidad capacitados para clonar fragmentos de DNA de 0,5-1,0 Mb en cromosomas artificiales de levadura. Estas clonas se pueden utilizar para refinar el mapa físico ya que si dos marcadores hibridan la misma clona han de estar íntimamente ligados (0,5-1,0 Mb). La obtención de clonas de YACs que comparten alguna secuencia (solapadas) y que cubren un intervalo del genoma concreto es uno de los procedimientos de mapping físico más utilizados en el momento presente. Al conjunto de clonas solapadas que cubren una región del genoma se denomina *contig* (de contiguo). A su vez, las clonas YACs se usan para ordenar *contigs* de clonas en vectores tipo fago, de menor capacidad que se utilizan para estudiar la estructura fina y secuencia del gen en cuestión.

Bibliografía

- Collins FS: Positional cloning: let's not call it reverse anymore. *Nature Genet* 1:3-6, 1992.
- Reeders ST: Multilocus polycystic disease. *Nature Genet* 1:235-237, 1992.
- Seizinger BR, Smith DI, Filling-Katz MR: Genetic flanking markers refine diagnostic criteria and provide insights into the genetics of von Hippel Lindan disease. *Proc Natl Acad Sci USA* 88:2864-2868, 1991.
- Pras E, Aksentijerich MD, Gruber L: Mapping of a gene causing familial mediterranean fever to the short arm of chromosome 16. *N Engl J Med* 326:1509-1513, 1992.
- Fahsold R, Rott HD, Loren P: A new gene locus for tuberous sclerosis maps to chromosome 12. *Cytogenet Cell Genet* 58:1976, 1991.
- Kandt RS, Haines JL, Smith M: Linkage of an important gene locus for tuberous sclerosis to a chromosome 16 marker for polycystic kidney disease. *Nature Genet* 2:37-41, 1992.
- Tommerup N, Van der Hagen CB, Heiberg A: Tentative assignment of a locus for Rubinstein-Taybi syndrome to 16 p13.3 by a de novo reciprocal translocation, t(7;16)(q34;p13,3). *Cytogenetic Cell Genet* 58:2002, 1991.
- Anand F, Burn J, Matthews D: Alagille syndrome and deletion of 20p. *J Med Genet*, 27:729-737, 1990.
- Call KM, Glaser T, Ito C: Isolation and characterization of a zinc polypeptide gene at the human chromosome II Wilm's tumor locus. *Cell* 60:509-520, 1990.
- Hostikka SL, Eddy RL, Byers MG: Identification of a distinct type IV collagen chain with restricted kidney distribution and assignment of a gene to the locus of X chromosome-linked. Alport syndrome. *Proc Natl Acad Sci USA* 87:16061610, 1990.
- Attree O, Olivos IM, Okabe I: The Lowe's oculocerebrorenal syndrome gene encodes a protein highly homologous to inositol polyphosphate-5-phosphatase. *Nature* 358:239-242, 1992.
- Legouis R, Hardelin JP, Leveilliers J: The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell* 67:423-435, 1991.
- Van den Ouweland AM, Pang Y: Vasopressin V2 receptor mutations in nephrogenic diabetes insipidus. *Nature Genet* 2:99-106, 1992.
- Morton NE: Parameters of the human genome. *Proc Natl Acad Sci USA* 88:7474-7476, 1991.
- The NIH/CEPH collaborative Mapping Group: A comprehensive Genetic Linkage Map of the human genome. *Science*, 258:67-86, 1992.
- Kan YW, Dozy AO: Polymorphism of DNA sequence adjacent to human β -globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 75:5631-5635, 1978.
- Williamson R, Bowcock A, Kidd KK: Report of the DNA committee and catalogues of cloned and mapped genes and DNA polymorphisms. *Cytogenet Cell Genet* 58:1190, 1991.
- Wyman AR, White R: A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA* 77:6754-6758, 1980.
- Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M: A second-generation linkage map of the human genome. *Nature* 359:794-801, 1992.
- Gusella JF, Wexler NS, Conneally PM: A polymorphic DNA marker genetically linked to Huntington disease. *Nature* 306:234-238, 1983.
- Kerem BS, Rommens JR, Buchanan JA: Identification of the cystic fibrosis gene: Genetic analysis. *Science* 306:234-238, 1983.
- Reeders ST, Breuning MH, Davies KE: A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 317:542-544, 1985.
- Kimberling WJ, Fain PR, Kenyon JB: Linkage heterogeneity of autosomal dominant polycystic kidney disease. *N Engl J Med* 319:913-918, 1988.
- Landers ES: Mapping complex genetic traits in humans. En: *Genome analysis: a practical approach*. Ed. K. Davies, IRL Press, Oxford, 171-189, 1988.
- Risch N: Genetic linkage: Interpreting lod scores. *Science* 255:803-804, 1992.
- Monaco AP, Neve RL; Colletti-Feener C, Bertelson CJ, Kurnit DM, Kundel LM: Isolation of candidate cDNAs for portions of the muscular dystrophy gene. *Nature* 323:646,650, 1986.
- Cremers FP, Van de Pol DJ, Van Kerkhoff LP, Wieringa B, Ropers HH: Cloning of a gene that is rearranged in patients with choroideremia. *Nature* 347:674-677, 1990.