



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 308 852**

51 Int. Cl.:

C12N 15/12 (2006.01)

C07K 14/47 (2006.01)

C12Q 1/68 (2006.01)

C12N 15/11 (2006.01)

A01K 67/027 (2006.01)

C07K 16/18 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **99954295 .4**

96 Fecha de presentación : **04.11.1999**

97 Número de publicación de la solicitud: **1127120**

97 Fecha de publicación de la solicitud: **29.08.2001**

54

Título: **Secuencias genómicas y completas de cDNA de APM1 humano específico para adipocitos y marcadores bialélicos de los mismos.**

30

Prioridad: **04.11.1998 US 107113 P**
10.02.1999 US 119593 P

45

Fecha de publicación de la mención BOPI:
01.12.2008

45

Fecha de la publicación del folleto de la patente:
01.12.2008

73

Titular/es: **Serono Genetics Institute S.A.**
route Nationale 7
91000 Evry Cédex, FR

72

Inventor/es: **Dufaure-Gare, Isabelle;**
Bougueleret, Lydie;
Blumenfeld, Marta y
Bihain, Bernard

74

Agente: **Elzaburu Márquez, Alberto**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Secuencias genómicas y completas de cDNA de APM1 humano específico para adipocitos y marcadores bialélicos de los mismos.

Esta solicitud reivindica prioridad con respecto a la solicitud de patente provisional de EE.UU. número de serie 60/107.113, presentada el 4 de noviembre de 1.988, de Bougeleret, titulada "Secuencia genómica y secuencia completa de cDNA de APM1", y con respecto a la solicitud de patente provisional de EE.UU. número de serie 60/119.593, presentada el 10 de febrero de 1.999, de Blumenfeld y col. y titulada "Secuencia genómica y secuencia completa de cDNA de APM1 y sus marcadores bialélicos", por lo que ambas se incorporan aquí en su totalidad, incluyendo todas las figuras, tablas o dibujos.

Campo de la invención

Esta invención concierne a las secuencias genómicas y de cDNA del gen APM1, así como a métodos y kits para detectar estos polinucleótidos. La invención también concierne a las regiones reguladoras, en particular a la región del promotor del gen APM1. La invención comprende marcadores bialélicos del gen APM1 que pueden ser útiles en el diagnóstico de la obesidad o de trastornos relacionados con la obesidad.

Antecedentes

La obesidad es un problema de salud pública que es importante y está muy generalizado. La tercera parte de la población de los países industrializados presenta un exceso de peso de al menos el 20% con respecto al peso ideal. Este fenómeno continúa empeorando, en particular en regiones del globo en las que los sistemas económicos se están modernizando. En Estados Unidos, el número de personas obesas se ha incrementado desde el 25% al final de la década de los 70 hasta el 33% al comienzo de la década de los 90.

La obesidad aumenta considerablemente el riesgo de contraer enfermedades cardiovasculares y metabólicas. Se estima que si toda la población tuviera un peso ideal, el riesgo de la insuficiencia coronaria decrecería en el 25% y el de la insuficiencia cardíaca y los accidentes cerebrovasculares decrecería en el 35%. La insuficiencia coronaria, la enfermedad ateromatosa y la insuficiencia cardíaca son el primer frente de las complicaciones cardiovasculares inducidas por la obesidad. En el caso de un exceso de peso mayor que el 30%, la incidencia de las enfermedades coronarias se dobla en sujetos de edad inferior a 50 años. Estudios llevados a cabo con respecto a otras enfermedades son igualmente significativos. En el caso de un exceso de peso del 20%, se dobla el riesgo de padecer una presión sanguínea alta. En el caso de un exceso de peso del 30%, el riesgo de contraer una diabetes insulino-dependiente se triplica y el riesgo de las hiperlipidemias se multiplica seis veces.

La lista de enfermedades cuya aparición está provocada por la obesidad es larga: hiperuricemia (11,4% en sujetos obesos, en comparación con el 3,4% en la población general), patologías digestivas, anormalidades en las funciones hepáticas e incluso determinados cánceres.

Tanto si los cambios fisiológicos en la obesidad se caractericen por un aumento en el número de células adiposas o por un aumento en la cantidad de triglicéridos almacenados en cada célula adiposa, o por ambos aumentos, este exceso de peso es principalmente el resultado de un desequilibrio entre la cantidad de calorías consumida y la cantidad de calorías utilizadas por el organismo. Algunos estudios realizados sobre las causas de este desequilibrio se han centrado en estudiar el mecanismo de la absorción de alimentos y, por consiguiente, de las moléculas que controlan la ingestión de alimentos y la sensación de saciedad. Otros estudios han caracterizado las rutas a través de las cuales el organismo usa sus calorías.

Los tratamientos contra la obesidad que se han propuesto son de cuatro tipos. 1) La restricción de alimento es el usado más frecuentemente usado. Se aconseja a los individuos obesos que cambien sus hábitos alimenticios de manera que consuman menos calorías. Este tipo de tratamiento es eficaz a corto plazo. Sin embargo la tasa de recidiva es muy alta. 2) También se propone un mayor gasto de las calorías a través del ejercicio físico. Este tratamiento no es eficaz cuando se aplica solo, pero sin embargo mejora la pérdida de peso en los sujetos que siguen una dieta baja en calorías. 3) La cirugía gastrointestinal, que disminuye la absorción de las calorías ingeridas, es eficaz pero ha sido prácticamente abandonada debido a los efectos secundarios que produce. 4) La estrategia medicamentosa usa o bien la acción anorexigénica de moléculas que participan a nivel del sistema nervioso central, o bien el efecto de moléculas que aumentan la utilización de la energía aumentando la producción de calor. Los prototipos de esta clase de molécula son las hormonas tiroideas que desacoplan la fosforilaciones oxidativas de la cadena respiratoria mitocondrial. Los efectos secundarios y la toxicidad de este tipo de tratamiento hacen que su uso sea peligroso. También se está utilizando una estrategia que tiene como objetivo disminuir la absorción de los lípidos de la dieta secuestrándolos en el lumen del tubo digestivo. Sin embargo, esta estrategia induce la aparición de desequilibrios fisiológicos que son difíciles de tolerar: deficiencia en la absorción de vitaminas liposolubles, flatulencia y esteatorrea. Cualquiera que sea la estrategia terapéutica contemplada, los tratamientos de la obesidad se caracterizan todos ellos por una tasa de recidiva extremadamente elevada.

Los mecanismos moleculares responsables de la obesidad en el hombre son complejos y en ellos participan factores genéticos y medioambientales. Debido a la baja eficiencia de los tratamientos conocidos hasta ahora, es urgente definir

los mecanismos genéticos que determinan la obesidad para ser capaces de desarrollar medicamentos que se dirijan mejor a ella.

Se han estudiado más de 20 genes como posibles candidatos, ya sea por que han sido implicados en enfermedades en las que la obesidad es una de las manifestaciones clínicas, o porque son homólogos de genes implicados en la obesidad en modelos animales. Situado en la región cromosómica 7q31, el gen OB es uno de los genes más ampliamente estudiados. Su producto, la leptina, participa en los mecanismos de la saciedad. La leptina es una proteína plasmática de 16 kDa, producida por los adipocitos bajo la acción de diversos estímulos. Ratones obesos de tipo *ob/ob* exhiben una deficiencia en el gen de la leptina; esta proteína es indetectable en el plasma de estos animales. La administración de leptina obtenida mediante ingeniería genética a ratones *ob/ob* corrige su hiperfagia relativa y permite la normalización de sus pesos. Este efecto anorexigénico de la leptina pone en juego a un receptor del sistema nervioso central: el receptor ob, que pertenece a la familia de la clase I de los receptores de citoquinas. El receptor ob es deficiente en ratones obesos de la cepa db/db. La administración de leptina a estos ratones no tiene efecto sobre su ingestión de alimento y no permite obtener una disminución sustancial en sus pesos. Los mecanismos mediante los que los receptores ob transmiten la señal de saciedad no se conocen con precisión. Es posible que el neuropéptido Y esté implicado en esta vía de señalización. Es importante especificar en este momento que los receptores ob no son los únicos reguladores del apetito. El receptor de la Melanocortina 4 está también implicado ya que ratones a los que se les ha hecho deficientes en este receptor son obesos (Gura, (1997)).

El descubrimiento de la leptina y la caracterización del receptor de la leptina en el sistema nervioso central abrió una nueva vía para la búsqueda de medicamentos contra la obesidad. Este modelo, sin embargo, pronto resultó decepcionante. De hecho, con sólo una excepción (Montague y col., (1997)), los genes que codifican la leptina o su receptor OB han resultado ser normales en sujetos humanos obesos. Además y paradójicamente, las concentraciones plasmáticas de leptina, la hormona de la saciedad, son anormalmente altas en la mayoría de los sujetos humanos obesos.

Es evidente que se necesitan medicamentos nuevos que sean útiles para reducir el peso corporal en seres humanos. Esas composiciones farmacéuticas, ayudarían favorablemente a controlar la obesidad y con ello a mitigar muchas de las consecuencias cardiovasculares asociadas con esta afección.

El gen APM1 específico de adipocitos humanos codifica una proteína secretora del tejido adiposo y es probable que desempeñe un papel en la patogénesis de la obesidad. El conocimiento de la secuencia genómica de APM1 y en particular de las secuencias tanto del promotor como de las uniones de corte y empalme, permite el diseño de nuevas herramientas diagnósticas y terapéuticas que actúen sobre el metabolismo de los lípidos y que sean útiles para diagnosticar y tratar trastornos relacionados con la obesidad.

Sumario de la invención

La presente invención tiene como punto de partida el aislamiento y caracterización de la secuencia genómica del gen *APM1*, incluyendo sus regiones reguladoras y la secuencia completa del cDNA que codifica la proteína APM1. Las sondas y cebadores oligonucleotídicos que se hibridan específicamente con una secuencia genómica de *APM1* forman también parte de la invención. Un objetivo más de la invención consiste en vectores recombinantes que comprenden cualquiera de las secuencias de ácido nucleico descritas en la presente invención y en particular en vectores recombinantes que comprenden la región del promotor de *APM1* o una secuencia que codifica la proteína APM1, así como células hospedadoras que comprenden dichas secuencias de ácido nucleico o dichos vectores recombinantes. La invención se dirige también a marcadores bialélicos que se localizan en la secuencia genómica de *APM1*, marcadores bialélicos que representan herramientas útiles para identificar una asociación estadísticamente significativa entre alelos específicos del gen *APM1* y uno o varios trastornos relacionados con la obesidad.

Breve descripción de las figuras

Figura 1: Mapa de la secuencia genómica de *APM1* humano (Tránsito 1 del gen más abundante en el tejido adiposo) (del inglés, "*Adipose Most Abundant Gene Transcript 1*") con los marcadores bialélicos de la invención.

Breve descripción de las secuencias proporcionadas en el listado de secuencias

El SEQ ID NO: 1 contiene una secuencia genómica de *APM1* que comprende la región reguladora en 5' (región aguas arriba que no se transcribe), los exones e intrones y la región reguladora en 3' (región agua abajo que no se transcribe).

El SEQ ID NO: 2 contiene una región reguladora en 5' (región aguas arriba que no se transcribe) del gen *APM1*.

El SEQ ID NO: 3 contiene una región reguladora en 3' (región agua abajo que no se transcribe) del gen *APM1*.

El SEQ ID NO: 4 contiene un cDNA parcial en 5' de *APM1*.

El SEQ ID NO: 5 contiene un cDNA completo del *APM1* humano.

ES 2 308 852 T3

El SEQ ID NO: 6 contiene la proteína APM1 codificada por el cDNA del SEQ ID NO: 5.

El SEQ ID NO: 7 contiene un cebador que contiene la secuencia adicional de PU en 5' descrita más adelante en el Ejemplo 2.

El SEQ ID NO: 8 contiene un cebador que contiene la secuencia adicional de RP en 5' descrita más adelante en el Ejemplo 2.

De acuerdo con las regulaciones relativas a los Listados de secuencias, los siguientes códigos se han usado en el Listado de secuencias para indicar las posiciones de los marcadores bialélicos dentro de las secuencias y para identificar cada uno de los alelos presentes en la base polimórfica. El código "r" en las secuencias indica que uno de los alelos de la base polimórfica es guanina mientras que el otro alelo es adenina. El código "y" en las secuencias indica que uno de los alelos de la base polimórfica es timina mientras que el otro alelo es citosina. El código "m" en las secuencias indica que uno de los alelos de la base polimórfica es adenina mientras que el otro alelo es citosina. El código "k" en las secuencias indica que uno de los alelos de la base polimórfica es guanina mientras que el otro alelo es timina. El código "s" en las secuencias indica que uno de los alelos de la base polimórfica es guanina mientras que el otro alelo es citosina. El código "w" en las secuencias indica que uno de los alelos de la base polimórfica es adenina mientras que el otro alelo es timina. El código del nucleótido del alelo original de cada uno de los marcadores bialélicos es el siguiente:

<i>Marcador bialélico</i>	<i>Alelo original</i>
9-27-261	G
99-14387-129	A
9-12-48	T
9-12-124	T
9-12-355	G
9-12-428	A
99-14405-105	G

En algunos casos, las bases polimórficas de los marcadores bialélicos cambian la identidad de los aminoácidos en el polipéptido codificado. Esto se indica en el Listado de secuencias que se adjunta mediante el uso de la característica VARIANTE, la colocación de un Xaa en la posición del aminoácido polimórfico y la definición de Xaa como los dos aminoácidos alternativos. Por ejemplo, si uno de los alelos de un marcador bialélico es el codón CAC, que codifica histidina, mientras que el otro alelo del marcador bialélico es CAA, que codifica glutamina, el Listado de secuencias correspondiente al polipéptido codificado contendrá un Xaa en la posición del aminoácido polimórfico. En este caso, Xaa se definiría como histidina o glutamina.

En otros casos, Xaa puede indicar un aminoácido cuya identidad se desconoce debido a la ambigüedad de la secuencia de nucleótidos. En este caso, se usa incierto, Xaa se coloca en la posición del aminoácido desconocido y Xaa se define como cualquiera de los 20 aminoácidos o cualquiera de un número limitado de los aminoácidos sugeridos por el código genético.

Descripción detallada

El objetivo de la presente invención es proporcionar polinucleótidos derivados del gen *APM1* que son particularmente útiles para diseñar medios adecuados para detectar la presencia de este gen en una muestra a ensayar o, como alternativa, para detectar la presencia de moléculas del mRNA de *APM1* que están presentes en una muestra a ensayar. También se describen polinucleótidos implicados en la expresión del gen *APM1* y que se pueden usar para diseñar medios capaces de modular la expresión de *APM1*. Otros polinucleótidos aquí descritos son útiles para diseñar medios adecuados para expresar un polinucleótido de interés deseado. La presente invención también abarca marcadores bialélicos del gen *APM1*.

Definiciones

Antes de describir la invención con mayor detalle, se exponen las siguientes definiciones para ilustrar y definir el significado y el alcance de los términos y expresiones usados para describir la presente invención.

La expresión "gen *APM1*", cuando se usa en el presente documento, incluye secuencias genómicas, de mRNA y de cDNA que codifican la proteína APM1, incluyendo la regiones del DNA genómico que no se traducen.

La expresión “proteína heteróloga”, cuando se usa en el presente documento, quiere designar cualquier proteína o polipéptido diferente de la proteína APM1. Más en particular, esta proteína heteróloga es un compuesto que se puede usar como marcador en experimentos adicionales, con una región reguladora de APM1.

El término “aislado” requiere que el material sea extraído de su entorno original (p. ej., del entorno natural si el material está presente en la naturaleza). Por ejemplo, un polinucleótido presente en la naturaleza o un polipéptido presente en un animal vivo no está aislado, pero el mismo polinucleótido o DNA o polipéptido, separado de parte o de la totalidad de los materiales coexistentes en ese sistema natural, está aislado. Ese polinucleótido podría ser parte de un vector y/o ese polinucleótido o polipéptido podría ser parte de una composición, y aún así podría estar aislado porque ni el vector ni la composición forman parte de su entorno natural.

El término “purificado” no requiere de una pureza absoluta; más bien, se concibe como una definición relativa. La purificación del material de partida o de un material natural en al menos un orden de magnitud, preferiblemente en dos o tres órdenes de magnitud, y más preferiblemente en cuatro o cinco órdenes de magnitud, se contempla expresamente. Como ejemplo, la purificación desde una concentración al 0,1% hasta una concentración al 10% es de dos órdenes de magnitud.

La expresión “polinucleótido purificado” o “vector polinucleotídico purificado” se usa aquí para describir un polinucleótido o un vector polinucleotídico de la invención que ha sido separado de otros compuestos que incluyen, pero que no se limitan a, otros ácidos nucleicos, carbohidratos, lípidos y proteínas (tales como las enzimas usadas en la síntesis del polinucleótido), o la separación de polinucleótidos cerrados covalentemente de entre polinucleótidos lineales. Un polinucleótido es sustancialmente puro cuando al menos aproximadamente el 50%, preferiblemente del 60% al 75% de una muestra exhibe una única secuencia polinucleotídica y conformación (lineal frente a cerrada covalentemente). Un polinucleótido sustancialmente puro típicamente comprende aproximadamente el 50%, preferiblemente del 60% al 90% peso/peso de una muestra de ácido nucleico, más usualmente, aproximadamente el 95%, y preferiblemente es puro en más de aproximadamente el 99%. La pureza u homogeneidad del polinucleótido se indica mediante varios medios muy conocidos en la técnica, tal como la electroforesis de una muestra en gel de agarosa o de poliacrilamida, seguida de la visualización de una sola banda del polinucleótido tras la tinción del gel. Para determinados fines, se puede proporcionar una resolución más alta usando HPLC u otros medios muy conocidos en la técnica.

El término “polipéptido” se refiere a un polímero de aminoácidos sin tener en consideración la longitud del polímero; por lo tanto, péptidos, oligopéptidos y proteínas están incluidos en la definición de polipéptido. Este término además no especifica ni excluye modificaciones postraduccionales de polipéptidos. Por ejemplo, polipéptidos que incluyen la unión covalente de grupos glicosilo, grupos acetilo, grupos fosfato, grupos lipídicos y grupos similares están expresamente incluidos en el término polipéptido. También incluidos en la definición están polipéptidos que contienen uno o más análogos de un aminoácido (que incluyen, por ejemplo, aminoácidos no presentes en la naturaleza, aminoácidos que solamente están presentes en la naturaleza en un sistema biológico no emparentado, aminoácidos modificados procedentes de sistemas de mamíferos, etc.), polipéptidos con uniones sustituidas, así como otras modificaciones conocidas en la técnica, tanto presentes en la naturaleza como no presentes en la naturaleza.

La expresión “polipéptido recombinante” se usa aquí para hacer referencia a polipéptidos que se han diseñado artificialmente y que comprenden al menos dos secuencias polipeptídicas que no se encuentran en forma de secuencias polipeptídicas contiguas en su entorno natural inicial, o para hacer referencia a polipéptidos que han sido expresados por un polinucleótido recombinante.

La expresión “polipéptido purificado” se usa aquí para describir un polipéptido de la invención que ha sido separado de otros compuestos que incluyen, pero que no se limitan a, ácidos nucleicos, lípidos, carbohidratos y otras proteínas. Un polipéptido es sustancialmente puro cuando una muestra contiene al menos el 50%, preferiblemente del 60% al 75%, de una única secuencia polipeptídica. Un polipéptido sustancialmente puro típicamente comprende aproximadamente el 50%, preferiblemente del 60% al 90%, más preferiblemente del 95% al 99% peso/peso de una muestra de proteína. La pureza u homogeneidad del polipéptido se indica mediante varios medios muy conocidos en la técnica, tales como la electroforesis de una muestra en gel de agarosa o de poliacrilamida, seguida de la visualización de las bandas del polipéptido tras la tinción del gel. Para determinados fines, se puede proporcionar una resolución más alta usando HPLC u otros medios muy conocidos en la técnica.

Según aquí se usa, la expresión “animal no humano” se refiere a cualquier vertebrado no humano, aves y más usualmente mamíferos, preferiblemente primates, animales de granja tales como cerdos, cabras, ovejas, burros y caballos, conejos o roedores, más preferiblemente ratas o ratones. Según aquí se usa, el término “animal” se usa para hacer referencia a cualquier vertebrado, preferiblemente un mamífero. Ambos términos “animal” y “mamífero” abarcan expresamente sujetos humanos salvo que vayan precedidos por la expresión “no humanos”.

Según aquí se usa, el término “anticuerpo” se refiere a un polipéptido o a un grupo de polipéptidos que comprenden al menos un dominio de unión, cuando un dominio de unión del anticuerpo se forma por el plegamiento de dominios variables de una molécula de anticuerpo para formar espacios de unión tridimensionales con una forma de la superficie interna y una distribución de cargas que son complementarios a las características de un determinante antigénico de un antígeno, lo que permite que tenga lugar una reacción inmunológica con ese antígeno. Los anticuerpos incluyen proteínas recombinantes que comprenden los dominios de unión así como fragmentos, incluyendo los fragmentos Fab, Fab', F(ab)₂ y F(ab')₂.

Según aquí se usa, un “determinante antigénico” es la porción de una molécula de antígeno, es este caso un polipéptido APM1, que determina la especificidad de la reacción antígeno-anticuerpo. Un “epítipo” se refiere a un determinante antigénico de un polipéptido. Un epítipo puede comprender tan sólo tres aminoácidos en una conformación espacial que es exclusiva de ese epítipo. Generalmente un epítipo consiste en al menos 6 de esos aminoácidos, y más usualmente en al menos 8-10 de esos aminoácidos. Métodos para determinar los aminoácidos que componen un epítipo incluyen cristalografía por rayos X, resonancia magnética nuclear bidimensional y mapeo epitópico, p. ej., el método de Pepscan, descrito por H. Mario Geysen y col. 1984. *Proc. Natl. Acad. Sci. U.S.A.* 81: 3998-4002; Publicación de PCT documento WO 84/03564; y Publicación de PCT documento WO 84/03506.

A lo largo de la presente memoria descriptiva, la expresión “secuencia de nucleótidos” se puede utilizar para designar indistintamente un polinucleótido o un ácido nucleico. De manera más precisa, la expresión “secuencia de nucleótidos” incluye el material nucleico propiamente dicho y por ello no se limita a la información sobre la secuencia (es decir, a la sucesión de letras elegidas entre las letras de las bases) que caracteriza bioquímicamente a una molécula específica de DNA o RNA.

Según aquí se usa indistintamente, la expresión “ácidos nucleicos” y los términos “oligonucleótidos” y “polinucleótidos” incluyen secuencias de RNA, DNA o secuencias híbridas de RNA/DNA de más de un nucleótido ya sea en forma monocatenaria o en forma de dúplex. El término “nucleótido” se usa aquí como adjetivo para describir moléculas que comprenden RNA, DNA o secuencias híbridas de RNA/DNA de una longitud cualquiera en forma monocatenaria o en forma de dúplex. El término “nucleótido” también se usa aquí como nombre para hacer referencia a nucleótidos individuales o a variedades de nucleótidos, dando a entender una molécula, o una unidad individual de una molécula de ácido nucleico más grande, que comprenden una purina o una pirimidina, un resto de azúcar ribosa o desoxirribosa y un grupo fosfato, o una unión fosfodiéster en el caso de nucleótidos contenidos en un oligonucleótido o polinucleótido. El término “nucleótido” también se usa aquí para incluir “nucleótidos modificados” que comprenden al menos una de las siguientes modificaciones: (a) un grupo de unión alternativo, (b) una forma análoga de purina, (c) una forma análoga de pirimidina o (d) un azúcar análogo. Con respecto a ejemplos de grupos enlazadores, purinas, pirimidinas y azúcares análogos, véase por ejemplo la Publicación de PCT documento WO 95/04064. Las secuencias polinucleótídicas de la invención se pueden preparar mediante alguno de los métodos conocidos, que incluyen métodos de síntesis, métodos recombinantes, generación *ex vivo* o una de sus combinaciones, así como utilizando los métodos de purificación conocidos en la técnica.

Un “promotor” se refiere a una secuencia de DNA reconocida por la maquinaria de síntesis de la célula, que es requerida para iniciar la transcripción específica de un gen.

Que una secuencia está “operativamente unida” a una secuencia reguladora, tal como un promotor, significa que dicho elemento regulador está en la posición y orientación correctas en relación con el ácido nucleico, para controlar la iniciación de la RNA-polimerasa y la expresión del ácido nucleico de interés. Según aquí se usa, la expresión “operativamente unida”, se refiere a una unión de elementos polinucleotídicos en una relación funcional. Por ejemplo, un promotor o un elemento estimulador están operativamente unidos a una secuencia codificadora si afectan a la transcripción de esa secuencia codificadora. De manera más precisa, dos moléculas de DNA (tales como un polinucleótido que contiene una región de promotor y un polinucleótido que codifica un polipéptido o polinucleótido deseado) se dice que están “operativamente unidas” si la naturaleza de la unión entre los dos polinucleótidos (1) no da como resultado la introducción de una mutación en el marco de lectura, o (2) no interfiere con la capacidad del polinucleótido que contiene el promotor para dirigir la transcripción del polinucleótido codificador.

El término “cebador” indica una secuencia oligonucleotídica específica que es complementaria a una secuencia de nucleótidos diana y se usa para que se hibride con la secuencia de nucleótidos diana. Un cebador sirve como punto de iniciación de una polimerización de nucleótidos catalizada por una DNA-polimerasa o por una RNA-polimerasa o por una transcriptasa inversa.

El término “sonda” indica un segmento de ácido nucleico definido (o un segmento de análogos de nucleótidos, p. ej., un polinucleótido según se define más adelante) que se puede usar para identificar una secuencia polinucleotídica específica presente en muestras, comprendiendo dicho segmento de ácido nucleico una secuencia de nucleótidos que es complementaria a la secuencia polinucleotídica específica que ha de ser identificada.

Los términos “rasgo” y “fenotipo” se usan aquí indistintamente y se refieren a cualquier propiedad visible, detectable o de alguna otra manera medible, de un organismo, tal como los síntomas de, o la susceptibilidad a padecer, una enfermedad por ejemplo. Típicamente, los términos “rasgo” y “fenotipo” se usan aquí para hacer referencia a los síntomas de, o la susceptibilidad a padecer, ya sea obesidad o trastornos relacionados con la obesidad, más en particular aterosclerosis, resistencia a la insulina, hipertensión, microangiopatía en individuos obesos con diabetes de tipo II, lesiones oculares asociadas con una microangiopatía en individuos obesos con diabetes de tipo II, o lesiones renales asociadas con una microangiopatía en individuos obesos con diabetes de tipo II.

El término “alelo” se usa aquí para hacer referencia a variantes de una secuencia de nucleótidos. Un polimorfismo bialélico presenta dos formas. Típicamente, el primer alelo identificado se designa como alelo original, mientras que los otros alelos se designan como alelos alternativos. Los organismos diploides pueden ser homocigotos o heterocigotos con respecto a una forma alélica.

La expresión “tasa de heterocigosidad” se usa aquí para hacer referencia a la incidencia de los individuos de una población que son heterocigotos en un alelo particular. En un sistema bialélico, la tasa de heterocigosidad es en promedio igual a $2P_a(1-P_a)$, cuando P_a es la frecuencia del alelo menos común. Con el fin de ser útil en estudios genéticos, un marcador genético debe tener un nivel adecuado de heterocigosidad para permitir una probabilidad razonable de que una persona seleccionada al azar sea heterocigota.

El término “genotipo”, según aquí se usa, se refiere a la identidad de los alelos presentes en un individuo o en una muestra. En el contexto de la presente invención, un genotipo se refiere preferiblemente a la descripción de los alelos del marcador bialélico presentes en un individuo o en una muestra. El término “genotipificar” una muestra o un individuo con respecto a un marcador bialélico consiste en determinar el alelo específico o el nucleótido específico portado por un individuo en el sitio del marcador bialélico.

El término “mutación” según aquí se usa, se refiere a una diferencia en la secuencia de DNA entre genomas o individuos diferentes, que tiene una frecuencia inferior al 1%.

El término “haplotipo” se refiere a una combinación de alelos presentes en un individuo o en una muestra. En el contexto de la presente invención, un haplotipo preferiblemente se refiere a una combinación de alelos de marcadores bialélicos que se encuentra en un individuo determinado y que puede estar asociada con un fenotipo.

El término “polimorfismo” según aquí se usa, se refiere a la existencia de dos o más secuencia genómicas alternativas o alelos alternativos entre genomas o individuos diferentes. “Polimórfico” se refiere al estado en el que dos o más variantes de una secuencia genómica específica se pueden encontrar en una población. Un “sitio polimórfico” es el locus en el que la variación tiene lugar. Un polimorfismo de un solo nucleótido es la sustitución de un único nucleótido por otro nucleótido en ese sitio polimórfico. La delección o la inserción de un solo nucleótido también da lugar a polimorfismos de un solo nucleótido. En el contexto de la presente invención, “polimorfismo de un solo nucleótido” preferiblemente se refiere a una sustitución de un solo nucleótido. Típicamente, entre individuos diferentes, el sitio polimórfico puede estar ocupado por dos nucleótidos diferentes.

La expresión “polimorfismo bialélico” y “marcador bialélico” se usan aquí indistintamente para hacer referencia a un polimorfismo de un solo nucleótido que tiene dos alelos presentes en la población con una frecuencia bastante alta. Un “alelo de marcador bialélico” se refiere a las variantes de nucleótidos presentes en el sitio de un marcador bialélico. Típicamente, la frecuencia del alelo menos común de los marcadores bialélicos de la presente invención se ha validado para que sea mayor que el 1%, preferiblemente la frecuencia es mayor que el 10%, más preferiblemente la frecuencia es al menos del 20% (es decir, una tasa de heterocigosidad de al menos 0,32), aún más preferiblemente la frecuencia es al menos del 30% (es decir, una tasa de heterocigosidad de al menos 0,42). Un marcador bialélico en el que la frecuencia del alelo menos común es del 30% o mayor se denomina “marcador bialélico de alta calidad”.

La posición de los nucleótidos en un polinucleótido con respecto al centro del polinucleótido se describe aquí de la siguiente manera. Cuando un polinucleótido tiene un número impar de nucleótidos, el nucleótido situado a la misma distancia de los extremos 5' y 3' del polinucleótido se considera que está “en el centro” del polinucleótido, y cualquier nucleótido situado en posición inmediatamente adyacente al nucleótido del centro, o el propio nucleótido situado en el centro, se considera que está a “una distancia de 1 nucleótido del centro”. Con un número impar de nucleótidos en un polinucleótido, cualquiera de las cinco posiciones de los nucleótidos situados en el centro del polinucleótido, se consideraría que están a una distancia de 2 nucleótidos del centro, y así sucesivamente. Cuando un polinucleótido tiene un número par de nucleótidos, habría un enlace y no un nucleótido situado en el centro del polinucleótido. Por lo tanto, cualquiera de los dos nucleótidos centrales se consideraría que está a “una distancia de 1 nucleótido del centro” y cualquiera de los cuatro nucleótidos situados en el medio del polinucleótido se consideraría que está a “una distancia de 2 nucleótidos del centro”, y así sucesivamente. En el caso de polimorfismos que implican la sustitución, inserción o delección, de 1 o más nucleótidos, el polimorfismo, el alelo o el marcador bialélico está “en el centro” de un polinucleótido si la diferencia entre la distancia desde los polinucleótidos sustituidos, insertados o delecionados del polimorfismo y el extremo 3' del polinucleótido, y la distancia desde los polinucleótidos sustituidos, insertados o delecionados del polimorfismo y el extremo 5' del polinucleótido, es cero o un solo nucleótido. Si esta diferencia es de 0 a 3, entonces se considera que el polimorfismo está a “una distancia de 1 nucleótido del centro”. Si la diferencia es de 0 a 5, se considera que el polimorfismo está a “una distancia de 2 nucleótidos del centro”. Si la diferencia es de 0 a 7, se considera que el polimorfismo está a “una distancia de 3 nucleótidos del centro”, y así sucesivamente.

El término “agua arriba” se usa aquí para hacer referencia a una posición que está situada hacia el extremo 5' del polinucleótido desde un punto de referencia específico.

Las expresiones “bases emparejadas” y “bases emparejadas según Watson y Crick” se usan aquí indistintamente para hacer referencia a nucleótidos que pueden estar unidos por enlaces de hidrógeno con otro nucleótido en virtud de las identidades de sus secuencias y de una manera como la encontrada en el DNA de doble hélice, con los residuos de timina o uracilo unidos a residuos de adenina por dos enlaces de hidrógeno, y con los residuos de citosina y guanina unidos por tres enlaces de hidrógeno. (Véase Stryer, L., *Biochemistry*, 4ª edición, 1.995).

El término “complementaria” o la expresión “su complemento” se usan aquí para hacer referencia a las secuencias de polinucleótidos que son capaces de formar parejas de bases según Watson y Crick con otro polinucleótido especificado a lo largo de la totalidad de la región complementaria. Para los fines de la presente invención, se considera

que un primer polinucleótido es complementario a un segundo polinucleótido cuando cada una de las bases del primer polinucleótido se empareja con su base complementaria. Bases complementarias son, en términos generales, A y T (o A y U), o C y G. “Complemento” se usa aquí como sinónimo de “polinucleótido complementario”, “ácido nucleico complementario” y “secuencia de nucleótidos complementaria”. Estos términos y expresiones se aplican a las parejas
5 de polinucleótidos sobre la base exclusivamente de sus secuencias y no de cualquier conjunto particular de situaciones en las que los dos polinucleótidos se unirían realmente.

La expresión “no génicos” se usa aquí para describir marcadores bialélicos relacionados con *APM1*, así como polinucleótidos y cebadores que se encuentran fuera de las posiciones nucleotídicas mostradas en la secuencia
10 genómica de *APM1* humano del SEQ ID NO: 1. El término “génico” se usa aquí para describir marcadores bialélicos relacionados con *APM1* así como polinucleótidos y cebadores que sí se encuentran en las posiciones nucleotídicas mostradas en la secuencia genómica de *APM1* humano del SEQ ID NO: 1.

Variantes y fragmentos

La invención también se refiere a variantes y fragmentos de los polinucleótidos aquí descritos, en particular, a un gen *APM1* que contiene uno o más marcadores bialélicos según la invención.

Variantes de polinucleótidos, según aquí se usa esta expresión, son polinucleótidos que difieren de un polinucleótido de referencia. Una variante de un polinucleótido puede ser una variante presente en la naturaleza, tal como una variante alélica presente en la naturaleza, o puede ser una variante que no se conoce que esté presente en la naturaleza. Esas variantes del polinucleótido no presentes en la naturaleza se pueden preparar mediante técnicas de mutagénesis, que incluyen las que se aplican a polinucleótidos, células u organismos. En general, las diferencias son limitadas de manera que las secuencias de nucleótidos de la referencia y de la variante son muy similares en conjunto y, en muchas
25 regiones, son idénticas.

Las variantes de polinucleótidos según la invención incluyen, sin que se esté limitado a ellas, secuencias de nucleótidos que son idénticas en al menos el 95% a un polinucleótido seleccionado entre el grupo que consiste en los SEQ ID NOS: 1-4, o a cualquier fragmento polinucleotídico de al menos 8 nucleótidos consecutivos de un polinucleótido seleccionado entre el grupo que consiste en los SEQ ID NOS: 1-3, y que preferiblemente son idénticas en al menos el 99%, más preferiblemente son idénticas en al menos el 99,5% y muy preferiblemente son idénticas en al menos el 99,8%, a un polinucleótido seleccionado entre el grupo que consiste en los SEQ ID NOS: 1-5, o a cualquier fragmento polinucleotídico de al menos 8 nucleótidos consecutivos de un polinucleótido seleccionado entre el grupo que consiste en los SEQ ID NOS: 1-3.

Los cambios de nucleótido que se encuentran en un polinucleótido variante pueden ser silenciosos, lo que significa que no cambian los aminoácidos codificados por el polinucleótido. Sin embargo, los cambios de nucleótido puedan también dar como resultado sustituciones, adiciones, delecciones, fusiones y truncamientos de aminoácidos en el polipéptido codificado por la secuencia de referencia. Las sustituciones, delecciones o adiciones pueden involucrar a uno o más nucleótidos. Las variantes pueden presentar cambios en regiones codificadoras o en regiones no codificadoras o en ambas. Los cambios en las regiones codificadoras pueden producir sustituciones, delecciones o adiciones de aminoácidos conservativas o no conservativas.

En el contexto de la presente invención, las realizaciones particularmente preferidas son aquellas en las que los polinucleótidos codifican polipéptidos que aumentan o retienen sustancialmente la misma función biológica o actividad que la proteína *APM1* madura, o aquellas en las que los polinucleótidos codifican polipéptidos que mantienen o aumentan una actividad biológica particular, al mismo tiempo que disminuyen o mantienen una segunda actividad biológica.

Un fragmento polinucleotídico es un polinucleótido que tiene una secuencia que es en su totalidad igual que una parte, pero no que la totalidad, de una secuencia de nucleótidos determinada, preferiblemente de la secuencia de nucleótidos de un gen *APM1*, y sus variantes. El fragmento puede ser una porción de un intrón de un gen *APM1*. También puede ser una porción de las regiones reguladoras de *APM1*, preferiblemente de la secuencia del promotor del gen *APM1*. Preferiblemente, esos fragmentos comprenden al menos uno de los marcadores bialélicos del A1 al A8, o sus complementos, o un marcador bialélico en desequilibrio de ligamiento con uno o más de los marcadores bialélicos del A1 al A8.

Esos fragmentos pueden ser “autónomos”, es decir, no son parte de, ni están fusionados a, otros polinucleótidos, o pueden estar comprendidos dentro de un polinucleótido individual más grande del que forman una parte o región. De hecho, varios de estos fragmentos pueden estar presentes dentro de un polinucleótido individual más grande.

Opcionalmente, esos fragmentos pueden consistir en, o consistir esencialmente en, un tramo contiguo de al menos 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 ó 1.000 nucleótidos de longitud. Un conjunto de fragmentos preferidos contiene al menos uno de los marcadores bialélicos del A1 al A8 del gen *APM1*, que se describen en el presente documento, o sus complementos.

Identidad entre ácidos nucleicos o polipéptidos

Las expresiones “porcentaje de identidad de secuencias” y “porcentaje de homología” se usan aquí indistintamente para hacer referencia a comparaciones entre polinucleótidos y polipéptidos, y se determinan comparando dos secuencias óptimamente alineadas a lo largo de una ventana de comparación, en las que la porción de la secuencia polinucleotídica o polipeptídica en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones ni deleciones, aunque se pueden producir huecos en los sitios en los que la otra secuencia contiene adiciones) para el alineamiento óptimo de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que la base de ácido nucleico idéntica o el residuo de aminoácido idéntico se encuentran en ambas secuencias para obtener el número de posiciones emparejadas, dividiendo el número de posiciones emparejadas entre el número total de posiciones de la ventana de comparación y multiplicando el resultado por 100 para obtener el porcentaje de identidad de las secuencias. La homología se evalúa usando cualquiera de los diversos algoritmos y programas de comparación de secuencias conocidos en la técnica. Esos algoritmos y programas incluye, pero no se limitan a, TBLASTN, BLASTP, FASTA, TFASTA y CLUSTALW (Pearson y Lipman, 1988, *Proc. Natl. Acad. Sci. USA* 85(8): 2444-2448; Altschul y col., 1990, *J. Mol. Biol.* 215(3): 403-410; Thompson y col., 1994, *Nucleic Acids Res.* 22(2): 4673-4680; Higgins y col., 1996, *Methods Enzymol.* 266: 383-402; Altschul y col., 1990, *J. Mol. Biol.* 215(3): 403-410; Altschul y col., 1993, *Nature Genetics* 3: 266-272). En una realización particularmente preferida, las homologías entre secuencias de proteínas y ácidos nucleicos se evalúan usando la herramienta básica de búsqueda de alineamientos locales (“BLAST”) (del inglés, “*Basic Local Alignment Search Tool*”) que es muy conocida en la técnica (véase, p. ej., Karlin y Altschul, 1990, *Proc. Natl. Acad. Sci. USA* 87: 2267-2268; Altschul y col., 1990, *J. Mol. Biol.* 215: 403-410; Altschul y col., 1993, *Nature Genetics* 3: 266-272; Altschul y col., 1997, *Nuc. Acids Res.* 25: 3389-3402). En particular, se usan cinco programas BLAST específicos para realizar la siguiente tarea:

(1) BLASTP y BLAST3 comparan una secuencia de aminoácidos a consultar (del inglés, “*query*”) frente a una base de datos de secuencias de proteínas;

(2) BLASTN compara una secuencia de nucleótidos a consultar frente a una base de datos de secuencias de nucleótidos;

(3) BLASTX compara los productos resultantes de la traducción conceptual de seis marcos de lectura de una secuencia de nucleótidos a consultar (las dos cadenas) frente a una base de datos de secuencias de proteínas;

(4) TBLASTN compara una secuencia de proteína a consultar frente a una base de datos de secuencias de nucleótidos traducidos en los seis marcos de lectura (las dos cadenas); y

(5) TBLASTX compara las traducciones de seis marcos de lectura de una secuencia de nucleótidos a consultar frente a las traducciones de los seis marcos de lectura de una base de datos de secuencias de nucleótidos.

Los programas BLAST identifican secuencias homólogas identificando segmentos similares, que aquí se denominan “parejas de segmentos de puntuación alta”, entre una secuencia de aminoácidos a consultar o una secuencia de ácido nucleico a consultar y una secuencia de prueba que se obtiene preferiblemente procedente de una base de datos de secuencias de proteínas o de ácidos nucleicos. Las parejas de segmentos de puntuación alta preferiblemente se identifican (es decir, se alinean) por medio de una matriz de puntuación, muchas de las cuales se conocen en la técnica. Preferiblemente, la matriz de puntuación utilizada es la matriz BLOSUM62 (Gonnet y col., 1992, *Science* 256: 1443-1445; Henikoff y Henikoff, 1993, *Proteins* 17: 49-61). Menos preferiblemente, también se pueden utilizar las matrices PAM o PAM250 (véase, p. ej., Schwartz y Dayhoff, redactores, 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). Los programas BLAST evalúan la significación estadística de todas las parejas de segmentos de puntuación alta identificadas, y preferiblemente seleccionan los segmentos que satisfacen un umbral de satisfacción especificado por el usuario, tal como un porcentaje de homología especificado por el usuario. Preferiblemente, la significación estadística de una pareja de segmentos de puntuación alta se evalúa usando la fórmula de significación estadística de Karlin (véase, p. ej., Karlin y Altschul, 1990, *Proc. Natl. Acad. Sci. USA* 87: 2267-2268).

Condiciones rigurosas de hibridación

A modo de ejemplo y no de limitación, los siguientes son procedimientos que usan condiciones de rigurosidad alta: La prehibridación de filtros que contienen DNA se lleva a cabo durante desde 8 h hasta durante toda la noche, a 65°C, en un tampón compuesto por SSC 6X, Tris-HCl 50 mM (pH 7,5), EDTA 1 mM, PVP al 0,02%, Ficoll al 0,02%, BSA al 0,02% y DNA desnaturalizado de esperma de salmón en concentración de 500 µg/ml. Los filtros se hibridan durante 48 h a 65°C, que es la temperatura de hibridación preferida, en una mezcla de hibridación que contiene DNA desnaturalizado de esperma de salmón en concentración de 100 µg/ml y 5-20 x 10⁶ cpm de una sonda marcada con ³²P. Como alternativa, la etapa de hibridación de puede llevar a cabo a 65°C en presencia de tampón de SSC, SSC 1X que corresponde a NaCl 0,15 M y citrato de Na 0,05. Posteriormente, los lavados de los filtros se pueden realizar a 37°C durante 1 h en una disolución que contiene SSC 2X, PVP al 0,01%, Ficoll al 0,01% y BSA al 0,01%, seguidos de un lavado en SSC 0,1X a 50°C durante 45 min. Como alternativa, los lavados de los filtros se pueden llevar a cabo en una disolución que contiene SSC 2X y SDS al 0,1%, o SSC 0,5X y SDS al 0,1%, o SSC 0,1X y SDS al 0,1% SDS a 68°C durante intervalos de 15 minutos. Después de la etapa de lavados, las sondas hibridadas son detectables mediante

autorradiografía. Otras condiciones de rigurosidad alta que pueden ser usadas son muy conocidas en la técnica y, como las citadas en Sambrook y col., 1989; y en Ausubel y col., 1989, se incorporan aquí en su totalidad. Estas condiciones de hibridación son adecuadas para una molécula de ácido nucleico de aproximadamente 20 nucleótidos de longitud. Las condiciones de hibridación anteriormente descritas se adaptan según la longitud del ácido nucleico deseado, siguiendo métodos muy conocidos por el experto en la técnica. Las condiciones de hibridación se pueden adaptar, por ejemplo, según las enseñanzas descritas en el libro de Hames y Higgins (1985) o en Sambrook y col. (1989).

Secuencias genómicas de APM1

La presente invención concierne a la secuencia genómica de *APM1*. La presente invención incluye polinucleótidos, genes *APM1* o secuencias genómicas de *APM1* que consisten en, o que comprenden, la secuencia del SEQ ID NO: 1, una secuencia complementaria de ella, así como sus fragmentos. Estos polinucleótidos pueden estar purificados, aislados o pueden ser recombinantes. Esta secuencia genómica de *APM1* se ha localizado en el locus 3p27 mediante FISH.

También se describen polinucleótidos purificados, aislados o recombinantes que comprenden una secuencia de nucleótidos que tiene una identidad de nucleótidos al menos del 70%, 75%, 80%, 85%, 90% o 95% con una secuencia de nucleótidos del SEQ ID NO: 1, o una secuencia complementaria a ella, o uno de sus fragmentos. Las diferencias de nucleótidos respecto a la secuencia de nucleótidos del SEQ ID NO: 1, pueden por lo general estar distribuidas al azar por todo el ácido nucleico. Sin embargo, los ácidos nucleicos preferidos son aquellos en los que las diferencias de nucleótidos respecto a la secuencia de nucleótidos del SEQ ID NO: 1 se localizan en su mayoría fuera de las secuencias codificadoras contenidas en los exones. Estos ácidos nucleicos, así como sus fragmentos y variantes, se pueden usar como cebadores oligonucleotídicos o sondas oligonucleotídicas con el fin de detectar la presencia de una copia del gen *APM1* en una muestra a ensayar, o como alternativa, con el fin de amplificar una secuencia nucleotídica diana contenida en las secuencias de *APM1*.

También se describen ácidos nucleicos purificados, aislados o recombinantes que se hibridan con la secuencia de nucleótidos del SEQ ID NO: 1, o con una secuencia complementaria a ella, o con una de sus variantes, en condiciones rigurosas de hibridación según se han definido anteriormente.

Los ácidos nucleicos de la invención particularmente preferidos incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5 ó 10 de las siguientes posiciones nucleotídicas del SEQ ID NO: 1: de la posición 1 a 3.528, de 4.852 a 15.143, de 15.366 a 16.276 y de 20.560 a 20.966. Otros ácidos nucleicos preferidos de la invención incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende las posiciones de 4.150 a 4.154, o de 17.169 a 17.170 del SEQ ID NO: 1. Ácidos nucleicos preferidos de la invención adicionales incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende una G en la posición 3.787, una G en la posición 3.809, una T en la posición 4.311, una A en la posición 4.328, una A en la posición 4.683 o una A en la posición 15.319 del SEQ ID NO: 1. Debe destacarse que los fragmentos de ácido nucleico de cualquier tamaño y secuencia pueden comprender también los polinucleótidos descritos en esta sección. Otros ácidos nucleicos particularmente preferidos de la invención incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5 ó 10 de las posiciones nucleotídicas de 1 a 4.833 del SEQ ID NO: 1.

El ácido nucleico genómico de *APM1* comprende 3 exones. El exón 1 comienza en el nucleótido de la posición 4.812 y termina en el nucleótido de la posición 4.851 de la secuencia de nucleótidos del SEQ ID NO: 1; El exón 2 comienza en el nucleótido de la posición 15.144 y termina en el nucleótido de la posición 15.365 de la secuencia de nucleótidos del SEQ ID NO: 1; El exón 3 comienza en el nucleótido de la posición 16.277 y termina en el nucleótido de la posición 20.559 de la secuencia de nucleótidos del SEQ ID NO: 1; Así, la invención abarca polinucleótidos purificados, aislados o recombinantes que comprenden una secuencia de nucleótidos seleccionada entre el grupo que consiste en los tres exones del gen *APM1*, o una secuencia complementaria a ella. La invención también trata sobre ácidos nucleicos purificados, aislados o recombinantes que comprenden una combinación de al menos dos exones del gen *APM1*, en los que los polinucleótidos están ordenados dentro del ácido nucleico desde el extremo 5' hasta el extremo 3' de dicho ácido nucleico, en el mismo orden que en el SEQ ID NO: 1.

El intrón 1 (secuencia de nucleótidos localizada entre el exón 1 y el exón 2) comienza en el nucleótido de la posición 4.852 de la secuencia de nucleótidos del SEQ ID NO: 1 y termina en el nucleótido de la posición 15.143 de la secuencia de nucleótidos del SEQ ID NO: 1; El intrón 2 (secuencia de nucleótidos localizada entre el exón 2 y el exón 3) comienza en el nucleótido de la posición 15.366 y termina en el nucleótido de la posición 16.276 de la secuencia de nucleótidos del SEQ ID NO: 1. Así, la invención abarca polinucleótidos purificados, aislados o recombinantes que comprenden una secuencia de nucleótidos seleccionada entre el grupo que consiste en el intrón 1 y en el intrón 2 del gen *APM1*, y secuencias complementarias a ella.

ES 2 308 852 T3

Aunque esta sección se titula "Secuencias genómicas de *APM1*", se debe señalar que fragmentos de ácido nucleico de cualquier tamaño y secuencia pueden también estar comprendidos en los polinucleótidos descritos en esta sección, incluyendo los que flanquean las secuencias genómicas de *APM1* por cada lado y/o los situados entre dos o más de estas secuencias genómicas.

Secuencias de cDNA de APM1

La expresión del gen *APM1* se ha demostrado que da lugar a la producción de al menos una especie de mRNA que tiene la secuencia de ácido nucleico expuesta en el SEQ ID NO: 5.

Otro objetivo de la invención es un ácido nucleico purificado, aislado o recombinante que comprende la secuencia de nucleótidos del SEQ ID NO: 5 y secuencias complementarias a ella. También se describen sus variantes alélicas y sus fragmentos. Además, los polinucleótidos preferidos de la invención incluyen los cDNA de *APM1* purificados, aislados o recombinantes, que consisten en, o que comprenden, la secuencia del SEQ ID NO: 5. Las realizaciones de la invención particularmente preferidas incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 5, o sus complementos, en los que dicho comprende posiciones seleccionadas entre el grupo que consiste en un nucleótido T en la posición 93, posiciones 1.154-1.157, un nucleótido G en la posición 1.997, posiciones 2083-2086, un nucleótido C en la posición 2.367, 2.456, 2.467, 2.475 ó 2.631, un nucleótido A en la posición 2.778, posiciones 2.785-2788, posiciones 2.797-2.801, un nucleótido T en la posición 3.594, un nucleótido G en la posición 3.684, posiciones 3.697-3.701, posiciones 4.026-4.027, un polinucleótido T en la posición 4.053, 4.078, 4.533 ó 4.536 del SEQ ID NO: 5. Realizaciones adicionales particularmente preferidas incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos, del SEQ ID NO: 5, o de sus complementos, en los que dicho comprende al menos 1, 2, 3, 5 ó 10 de las posiciones nucleotídicas de 1 a 22 del SEQ ID NO: 5.

El cDNA del SEQ ID NO: 5 incluye una región 5'-UTR que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 48 del SEQ ID No 5. El cDNA del SEQ ID NO: 5 incluye una región 3'-UTR que comienza en el nucleótido de la posición 785 y termina en el nucleótido de la posición 4.545 del SEQ ID NO: 5. Al menos dos sitios de poliadenilación están presentes en la posición de 2.937 a 2.942 y en la posición de 4.525 a 4.530 del SEQ ID NO: 5.

Por consiguiente, también se describen ácidos nucleicos purificados, aislados o recombinantes que comprenden una secuencia de nucleótidos del 5'-UTR del cDNA de *APM1*, una secuencia complementaria a ella, o una de sus variantes alélicas.

La secuencia situada en el extremo 5' de este cDNA, más en particular la secuencia de nucleótidos que comprende del 1 al 367 del SEQ ID NO: 5, corresponde a la secuencia de nucleótidos de un 5'-EST que se había obtenido procedente de una genoteca de cDNA de músculo distrófico humano y que se había caracterizado siguiendo las enseñanzas de la solicitud de PCT documento WO 96/34981 y de la solicitud de Patente de EE. UU. número de serie 08/905.134, presentada el 1 de agosto de 1.997. También se describen polinucleótidos que comprenden esta 5'-EST. Esta 5'-EST se expone en el SEQ ID NO: 4.

Aunque esta sección se titula "Secuencias de cDNA de *APM1*", se debe señalar que fragmentos de ácido nucleico de cualquier tamaño y secuencia pueden también estar comprendidos en los polinucleótidos descritos en esta sección, incluyendo los que flanquean las secuencias genómicas de *APM1* y/o los situados entre dos o más de estas secuencias genómicas.

Secuencias reguladoras de APM1

La secuencia genómica del gen *APM1* contiene secuencias reguladoras tanto en la región no codificadora flanqueante en 5' como en la región no codificadora flanqueante en 3' que bordean la región codificadora de *APM1* que contiene los tres exones de este gen, así como también en los intrones.

La secuencia reguladora en 5' del gen *APM1* comprende la secuencia de nucleótidos del SEQ ID NO: 2, y desde el 1 al 4.811 del SEQ ID NO: 1. Este polinucleótido contiene el sitio del promotor.

La secuencia reguladora en 3' del gen *APM1* comprende la secuencia de nucleótidos del SEQ ID NO: 3, y desde el 20.560 al 20.966 del SEQ ID NO: 1.

Los polinucleótidos derivados de los SEQ ID NOS: 2 ó 3 son útiles con el fin de detectar la presencia de al menos una copia de una secuencia de nucleótidos del SEQ ID NO: 1 o de uno de sus fragmentos en una muestra a ensayar.

La actividad del promotor de las regiones reguladoras de *APM1* se puede evaluar según se describe más adelante.

Los métodos para identificar fragmentos polinucleotídicos relevantes y biológicamente activos o variantes de los SEQ ID NOS: 2 y 3, son conocidos por un experto en la técnica, y se describen métodos ilustrativos en Sambrook y

col. (Sambrook, 1989). Por ejemplo, la presencia de un promotor (o de otras secuencias reguladoras) en las secuencias a ensayar se puede determinar sometiendo las secuencia a ensayar (fragmentos o variantes de los SEQ ID NOS: 2 y 3, por ejemplo) a un procesamiento por corte y empalme en un vector recombinante que porta un gen marcador (es decir, beta-galactosidasa, cloranfenicol-acetil-transferasa, etc.) que se expresa sólo si un promotor u otras secuencias reguladoras están presentes en las secuencias a ensayar. Las secuencias genómicas situadas aguas arriba del primer exón del gen *APM1* se pueden clonar en un vector adecuado de reporte de un promotor, tal como los vectores de reporte de promotores pSEAP-*Basic*, pSEAP-*Enhancer*, ppgal-*Basic*, pβgal-*Enhancer*, or pEGFP-1, disponibles procedentes de Clontech, o un vector que contiene un gen de reporte de luciferasa y que carece de promotor, pGL2-*basic* o pGL3-*basic*, procedentes de Promega. Brevemente expuesto, cada uno de estos vectores de reporte de un promotor incluyen sitios de clonación múltiple situados aguas arriba de un gen de reporte que codifica una proteína fácilmente ensayable, tal como fosfatasa alcalina secretada, luciferasa, β-galactosidasa o la proteína fluorescente verde. Las secuencias aguas arriba desde la región codificadora de *APM1* se insertan en los sitios de clonación situados aguas arriba del gen de reporte, en ambas orientaciones, y se introducen en una célula hospedadora adecuada. El nivel de la proteína de reporte se ensaya y se compara con el nivel obtenido con un vector que carece de un inserto en el sitio de clonación. La presencia de un nivel de expresión elevado en el vector que contiene el inserto, con respecto a la del vector de control, indica la presencia de un promotor en el inserto. En caso necesario, las secuencias aguas arriba se pueden clonar dentro de vectores que contienen un elemento estimulador para aumentar los niveles de transcripción desde las secuencias de un promotor débil. Un nivel significativo de expresión por encima del observado con el vector que carece de inserto, indica que una secuencia de promotor se encuentra presente en la secuencia insertada aguas arriba.

Una secuencia de promotor contenida en el DNA genómico aguas arriba se puede definir también construyendo delecciones anidadas en 5' y/o 3' en el DNA aguas arriba, usando técnicas convencionales tales como una digestión con Exonucleasa III o con endonucleasas de restricción adecuadas. Los fragmentos deleccionados resultantes, se pueden insertar en el vector de reporte de un promotor para determinar si la delección ha disminuido o ha destruido la actividad del promotor, según se describe, por ejemplo, en Coles y col. (1998). De esta manera, se pueden definir los límites de los promotores. Si se desea, se pueden identificar sitios potenciales e individuales de regulación, contenidos en el promotor, usando una mutagénesis dirigida de sitio específico o una exploración de fragmentos conectores para destruir sitios potenciales de unión a factores de transcripción, contenidos en el promotor, de manera individual o en combinación. Los efectos de estas mutaciones sobre los niveles de la transcripción se pueden determinar insertando las mutaciones en los sitios de clonación de los vectores de reporte de un promotor. Este tipo de ensayo es muy conocido por los expertos en la técnica y se encuentra descrito en los documentos WO 97/17359, US 5.374.544, EP 582 796, US 5.698.389, US 5.643.746, US 5.502.176 y US 5.266.488, incorporados aquí como referencia en su totalidad, incluyendo todos los dibujos, figuras o tablas.

La fuerza y la especificidad del promotor del gen *APM1* se puede determinar a través de los niveles de expresión de un polinucleótido detectable, operativamente unido al promotor de *APM1* en diferentes tipos de células y tejidos. El polinucleótido detectable puede ser o un polinucleótido que se hibrida específicamente con una sonda oligonucleotídica predefinida, o un polinucleótido que codifica una proteína detectable, incluyendo un polipéptido de *APM1* o uno de sus fragmentos o una de sus variantes. Este tipo de ensayo es muy conocido por los expertos en la técnica y se encuentra descrito en los documentos US 5.502.176 y US 5.266.488. Además, algunos de los métodos se comentan con más detalle más adelante.

Se pueden usar ventajosamente polinucleótidos que portan los elementos reguladores situados en el extremo 5' y en el extremo 3' de la región codificadora de *APM1* para controlar la actividad transcripcional y traduccional de un polinucleótido heterólogo de interés.

Por lo tanto, también se describe un ácido nucleico purificado o aislado que comprende un polinucleótido que se selecciona entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NOS: 2 y 3, o una secuencia complementaria a ella, o uno de sus fragmentos biológicamente activos, o una de sus variantes.

También se describe un ácido nucleico purificado o aislado que comprende un polinucleótido que tiene una identidad de nucleótidos de al menos el 95% con respecto a un polinucleótido seleccionado entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NOS: 2 y 3, que tiene ventajosamente una identidad de nucleótidos del 99%, preferiblemente una identidad de nucleótidos del 99,5% y muy preferiblemente una identidad de nucleótidos del 99,8%, con respecto a un polinucleótido seleccionado entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NOS: 2 y 3, o una secuencia complementaria a ella, o una de sus variantes, o uno de sus fragmentos biológicamente activos.

También se describen ácidos nucleicos purificados, aislados o recombinantes que comprenden un polinucleótido que se hibrida, en condiciones rigurosas de hibridación aquí definidas, con un polinucleótido seleccionado entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NO: 2 y 3, o una secuencia complementaria a ella, o una de sus variantes, o uno de sus fragmentos biológicamente activos.

Los fragmentos preferidos del ácido nucleico del SEQ ID NO: 2 tiene una longitud de aproximadamente 1.500 ó 1.000 nucleótidos, preferiblemente de aproximadamente 500 nucleótidos, más preferiblemente de aproximadamente 400 nucleótidos, aún más preferiblemente de 300 nucleótidos y muy preferiblemente de aproximadamente 200 nucleótidos. Preferiblemente, los fragmentos del SEQ ID NO: 2 están dentro de las posiciones de 1 a 3.528.

ES 2 308 852 T3

Los fragmentos preferidos del ácido nucleico del SEQ ID NO: 3 tienen una longitud de al menos 50, 100, 150, 200, 300 ó 400 bases.

Por derivados de polinucleótidos “biológicamente activos” de los SEQ ID NOS: 1, 2 y 3 se entiende polinucleótidos que comprenden, o que como alternativa consisten en, un fragmento de dicho polinucleótido que es funcional como región reguladora de la expresión de un polipéptido recombinante o de un polinucleótido recombinante en una hospedador celular recombinante. Dicho fragmento podría actuar o como secuencia estimuladora o como represor.

Para los fines de esta descripción, un ácido nucleico o un polinucleótido es “funcional” como región reguladora de la expresión de un polipéptido recombinante o de un polinucleótido recombinante si dicho polinucleótido regulador contiene secuencias de nucleótidos que contienen información transcripcional y traduccional relacionada con la regulación, y esas secuencias se encuentran “operativamente unidas” a secuencias de nucleótidos que codifica el polipéptido deseado o el polinucleótido deseado.

Los polinucleótidos reguladores se pueden preparar a partir de cualquiera de las secuencias de nucleótidos de los SEQ ID NOS: 1, 2 y 3 mediante escisión usando enzimas de restricción adecuadas, como se describe por ejemplo en Sambrook y col. (1989).

Los polinucleótidos reguladores se pueden también preparar mediante digestión de cualquiera de los SEQ ID NOS: 1, 2 y 3 con una enzima exonucleasa, tal como Bal31 (Wabiko y col., 1986).

Estos polinucleótidos reguladores también se pueden preparar mediante síntesis química de ácidos nucleicos, según se describe en otra parte de esta memoria descriptiva.

Los polinucleótidos reguladores pueden formar parte de un vector de expresión recombinante que se puede usar para expresar una secuencia codificadora en una célula hospedadora deseada o en un organismo hospedador deseado. Los vectores de expresión recombinantes según la invención se describen en otra parte de esta memoria descriptiva.

Un polinucleótido regulador en 5' preferido de la invención incluye la región en 5' que no se traduce (5'-UTR) (del inglés, “5'-*Untranslated Region*”) del cDNA de APM1, o uno de sus fragmentos biológicamente activos, o una de sus variantes.

Un polinucleótido regulador en 3' preferido de la invención incluye la región en 3' que no se traduce (3'-UTR) del cDNA de APM1, o uno de sus fragmentos biológicamente activos, o una de sus variantes.

También se describe un ácido nucleico purificado o aislado que comprende:

a) un ácido nucleico que comprende una secuencia de nucleótidos reguladora, seleccionada entre el grupo que consiste en:

- (i) una secuencia de nucleótidos que comprende un polinucleótido del SEQ ID NO: 2, o una secuencia complementaria a ella;
- (ii) una secuencia de nucleótidos que comprende un polinucleótido que tiene una identidad de nucleótidos de al menos el 95% con respecto a la secuencia de nucleótidos del SEQ ID NO: 2, o una secuencia complementaria a ella;
- (iii) una secuencia de nucleótidos que comprende un polinucleótido que se hibrida en condiciones de hibridación rigurosas con la secuencia de nucleótidos del SEQ ID NO: 2, o una secuencia complementaria a ella; y
- (iv) un fragmento biológicamente activo, o una variante, de los polinucleótidos de (i), (ii) y (iii);

b) un polinucleótido que codifica un polipéptido deseado o un ácido nucleico de interés, operativamente unido al ácido nucleico definido en el anterior apartado (a);

c) Opcionalmente, un ácido nucleico que comprende un polinucleótido regulador en 3', preferiblemente un polinucleótido regulador en 3' del gen *APM1*.

En una realización específica del ácido nucleico anteriormente definido, dicho ácido nucleico incluye la región en 5' que no se traduce (5'-UTR) del cDNA de APM1, o uno de sus fragmentos o variantes biológicamente activos.

En una segunda realización específica del ácido nucleico anteriormente definido, dicho ácido nucleico incluye la región en 3' que no se traduce (3'-UTR) del cDNA de APM1, o uno de sus fragmentos o variantes biológicamente activos.

El polinucleótido regulador del SEQ ID NO: 2, y sus fragmentos o variantes biológicamente activos, está operativamente unido al extremo 5' del polinucleótido que codifica el polipéptido o polinucleótido deseado.

El polinucleótido regulador del SEQ ID NO: 3, o sus fragmentos o variantes biológicamente activos, está de manera ventajosa unido operativamente al extremo 3' del polinucleótido que codifica el polipéptido o polinucleótido deseado.

El polipéptido deseado codificado por el ácido nucleico anteriormente descrito, puede ser de distinta naturaleza u origen, e incluye proteínas de origen procariótico o eucariótico. Entre los polipéptidos expresados bajo el control de una región reguladora de *APM1* se incluyen antígenos bacterianos, fúngicos o víricos. También están abarcadas proteínas eucariotas que incluyen, pero que no se limitan a, proteínas intracelulares tales como proteínas "domésticas", proteínas unidas a membranas tales como receptores, y proteínas secretadas tales como mediadores endógenos, por ejemplo citoquinas. El polipéptido deseado puede ser la proteína APM1, especialmente la proteína de la secuencia de aminoácidos del SEQ ID NO: 6, o uno de sus fragmentos, o una de sus variantes.

Los ácidos nucleicos codificados por el polinucleótido anteriormente descrito, normalmente una molécula de RNA, pueden ser complementarios a un polinucleótido codificador, por ejemplo, a la secuencia codificadora de *APM1*, y por lo tanto pueden ser útiles como polinucleótidos antisentido.

Ese polinucleótido se puede incluir en un vector de expresión recombinante con el fin de que exprese el polipéptido deseado o el ácido nucleico deseado en una célula hospedadora o en un organismo hospedador. Los vectores recombinantes adecuados que contiene un polinucleótido tal como se ha descrito aquí en lo que antecede, se describen en otra parte de esta memoria descriptiva.

Regiones codificadoras

El marco de lectura abierto de *APM1* está contenido en el RNA correspondiente del SEQ ID NO: 5. De manera más precisa, la secuencia codificadora (CDS) efectiva de *APM1* incluye la región situada entre la posición nucleotídica 49 (primer nucleótido del codón ATG codon) y la posición nucleotídica 783 (último nucleótido del codón TGA) del SEQ ID NO: 5. También se describen polinucleótidos aislados, purificados y recombinantes, que codifican polipéptidos que comprenden un tramo contiguo de al menos 6 aminoácidos, preferiblemente de al menos 8 ó 10 aminoácidos, más preferiblemente de al menos 12, 20, 25, 30, 40, 50 ó 100 aminoácidos del SEQ ID NO: 6, en los que dicho tramo contiguo incluye un residuo de ácido glutámico en la posición de aminoácido 56 en el SEQ ID NO: 6.

La secuencia codificadora de *APM1* puede ser expresada en una célula hospedadora deseada o en un organismo hospedador deseado, cuando este polinucleótido se coloca bajo el control de señales de expresión adecuadas. Las señales de expresión pueden ser o señales de expresión contenidas en las regiones reguladoras del gen *APM1* de la invención o pueden ser secuencias nucleicas reguladoras exógenas. Ese polinucleótido, cuando se coloca bajo el control de señales de expresión adecuadas, se puede también insertar en un vector para su expresión y/o amplificación.

Construcciones polinucleotídicas

Las expresiones "construcción polinucleotídica" y "polinucleótido recombinante" se usan aquí indistintamente para hacer referencia a polinucleótidos lineales o circulares, purificados o aislados, que han sido diseñados artificialmente y que comprenden al menos dos secuencias de nucleótidos que no se encuentran en forma de secuencias de nucleótidos contiguas en sus entornos naturales iniciales.

Construcción de DNA que permite dirigir la expresión temporal y espacial del gen APM1 en hospedadores celulares recombinantes y en animales transgénicos

Con el fin de estudiar las consecuencias fisiológicas y fenotípicas de una falta de síntesis de la proteína APM1, tanto a nivel de células como a nivel de organismos multicelulares, la invención también incluye construcciones de DNA y vectores recombinantes que permiten una expresión condicional de un alelo específico de la secuencia genómica o del cDNA de *APM1* y también de una copia de esta secuencia genómica o de este cDNA, que alberga sustituciones, deleciones o adiciones de una o más bases con respecto a la secuencia de nucleótidos de *APM1* de los SEQ ID NOS: 1 y 5, o de uno de sus fragmentos, estando localizadas estas sustituciones, deleciones o adiciones de bases o en un exón o en un intrón o en una secuencia reguladora, pero preferiblemente en la secuencia reguladora en 5' o en un exón de la secuencia genómica de *APM1* o dentro del cDNA de *APM1* del SEQ ID NO: 5. En una realización preferida, la secuencia de APM1 comprende un marcador bialélico de la presente invención.

La presente invención también abarca vectores recombinantes que comprenden polinucleótidos aislados, purificados o recombinantes, que codifican polipéptidos de comprenden un tramo contiguo de al menos 6 aminoácidos, preferiblemente de al menos 8 ó 10 aminoácidos, más preferiblemente de al menos 12, 15, 20, 25, 30, 40, 50 o 100 aminoácidos del SEQ ID NO: 6, en los que dicho tramo contiguo incluye un residuo de ácido glutámico en la posición de aminoácido 56 del SEQ ID NO: 6. Las realizaciones de la invención particularmente preferidas incluyen vectores recombinantes que comprenden polinucleótidos aislados, purificados o recombinante, que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID No: 5 o sus complementos, en los que dicho tramo contiguo comprende posiciones seleccionadas entre el grupo que consiste en un nucleótido T en la posición 93, posiciones 1.154-1.157, un nucleótido G en la posición 1.997, posiciones 2.083-2.086, un nucleótido C en la posición 2.367, 2.456, 2.467, 2.475 ó 2.631, un nucleótido A en la posición 2.778, posiciones 2.785-2.788, posiciones 2.797-2.801, un nucleótido T en la posición 3.594, un nucleótido G en la posición 3.684, posiciones 3.697-3.701, posiciones 4.026-4.027, un nucleótido T en la posición 4.053, 4.078, 4.533 ó 4.536 del

SEQ ID NO: 5. Otras realizaciones particularmente preferidas de la invención incluyen vectores recombinantes que comprenden polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 5 o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5 ó 10 de las posiciones nucleotídicas de 1 a 22 del SEQ ID NO: 5. Esas realizaciones son particularmente útiles en vectores de expresión y cuando se usan para transfectar de manera estable células hospedadoras y animales. Los vectores recombinantes particularmente preferidos de la invención comprenden polinucleótidos aislados, purificados o recombinantes, que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5 ó 10 de las siguientes posiciones nucleotídicas del SEQ ID NO: 1: de 1 a 3.528, de 4.852 a 15.143, de 15.366 a 16.276 y de 20.560 a 20.966. Otros vectores recombinantes preferidos de la invención comprenden polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1 o sus complementos en los que dicho tramo contiguo comprende las posiciones de 4.150 a 4.154, o de 17.169 a 17.170 del SEQ ID NO: 1. Vectores recombinantes adicionales y preferidos de la invención comprenden polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1 o sus complementos, en los que dicho tramo contiguo comprende una G en la posición 3.787, una G en la posición 3.809, una T en la posición 4.311, una A en la posición 4.328, una A en la posición 4.683 o una A en la posición 15.319 del SEQ ID NO: 1. Otros vectores recombinantes particularmente preferidos de la invención incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1 o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5 ó 10 de las posiciones nucleotídicas de 1 a 4.833 del SEQ ID NO: 1.

Una primera construcción de DNA preferida está basada en el operón *tet* de resistencia a tetraciclina procedente del trasposón Tn110 de *E. coli*, encargado de controlar la expresión del gen *APM1*, tal como se describe en Gossen y col. (1992, 1995) y en Furth y col. (1994). Esa construcción de DNA contiene siete secuencias del operador *tet* procedentes de Tn10 (*tetop*) que están fusionadas o a un promotor mínimo o a una secuencia reguladora en 5' del gen *APM1*, estando dicho promotor mínimo o dicha secuencia reguladora de *APM1* operativamente unidos a un polinucleótido de interés que codifica o un oligonucleótido con sentido o antisentido, o un polipéptido, incluyendo un polipéptido *APM1* o uno de sus fragmentos peptídicos. Esta construcción de DNA es funcional como sistema de expresión condicional de la secuencia de nucleótidos de interés cuando la misma célula comprende también una secuencia de nucleótidos que codifica o (*tTA*) de tipo salvaje o el represor (*rTA*) mutante, fusionada con el dominio de activación de la proteína vírica VP16 del virus del herpes simple, situado bajo el control de un promotor, tal como el promotor/estimulador HCMVIE1 o la MMTV-LTR. De hecho, una construcción de DNA preferida de la invención comprende tanto el polinucleótido que contiene las secuencias del operador *tet* como el polinucleótido que contiene una secuencia que codifica el *tTA* o el represor *rTA*.

En una realización específica, la construcción de DNA de expresión condicional contiene la secuencia que codifica el represor mutante de tetraciclina *rTA*, cuando la expresión del polinucleótido de interés es silenciosa en ausencia de tetraciclina y es inducida en su presencia.

Construcciones de DNA que permiten una recombinación homóloga: Vectores de sustitución

Una segunda construcción de DNA preferida comprenderá desde el extremo 5' hasta el extremo 3': (a) una primera secuencia de nucleótidos que está comprendida en la secuencia genómica *APM1*; (b) una secuencia de nucleótidos que comprende un marcador de selección positiva, tal como el marcador de resistencia a neomicina (*neo*); y (c) una segunda secuencia de nucleótidos que está comprendida en la secuencia genómica de *APM1*, y que está situada aguas abajo de la primera secuencia de nucleótidos (a) de *APM1*.

En una realización preferida, esta construcción de DNA comprende también un marcador de selección negativa situado aguas arriba de la secuencia de nucleótidos (a) o situado aguas abajo de la secuencia de nucleótidos (c). Preferiblemente, el marcador de selección negativa consiste en el gen de la timidina-quinasa (*tk*) (Thomas y col., 1986), el gen de la higromicina beta (*Te Riele y col.*, 1990), el gen *hprt* (Van der Lugt y col., 1991; Reid y col., 1990) o el gen del fragmento A de la toxina diftérica (*Dt-A*) (Nada y col., 1993; Yagi y col. 1990). Preferiblemente, el marcador de selección positiva se sitúa dentro de una secuencia exónica de *APM1* a fin de interrumpir la secuencia que codifica la proteína *APM1*. Estos vectores de sustitución se encuentran descritos, por ejemplo, en Thomas y col. (1986; 1987), Mansour y col. (1988) y Koller y col. (1992).

Las secuencias de nucleótidos primera y segunda (a) y (c) pueden estar situadas dentro de una secuencia reguladora de *APM1*, de una secuencia intrónica, de una secuencia exónica o de una secuencia que contiene una secuencia reguladora y/o una secuencia intrónica y/o una secuencia exónica. El tamaño de las secuencias de nucleótidos (a) y (c) varía de 1 a 50 kb, preferiblemente de 1 a 10 kb, más preferiblemente de 2 a 6 kb y muy preferiblemente de 2 a 4 kb.

Construcciones de DNA que permiten una recombinación homóloga: Sistema Cre-LoxP

Estas nuevas construcciones de DNA hacen uso del sistema de recombinación específica de sitio del fago P1. El fago P1 posee una recombinasa denominada Cre que interacciona específicamente con un sitio *loxP* de 34 pares de bases. El sitio *loxP* se compone de dos secuencias palindrómicas de 13 pb separadas por una secuencia conservada

de 8 bp (Hoess y col., 1986). La recombinación por acción de la enzima Cre entre dos sitios *loxP* que tienen una orientación idéntica da lugar a la delección de ese fragmento de DNA.

El sistema Cre-*loxP* usado en combinación con una técnica de recombinación homóloga ha sido descrito por Gu y col. (1993, 1994). Brevemente expuesto, una secuencia de nucleótidos de interés que ha de ser insertada en una posición diana del genoma, alberga al menos dos sitios *loxP* en la misma orientación y se ha de colocar en los respectivos extremos de una secuencia de nucleótidos que ha de ser escindida por el genoma recombinante. El evento de escisión requiere la presencia de la enzima recombinasa (Cre) dentro del núcleo del hospedador celular recombinante. La enzima recombinasa se puede proporcionar (a) incubando los hospedadores celulares recombinantes en un medio de cultivo que contiene esta enzima, inyectando la enzima Cre directamente en la células deseada, tal como se describe en Araki y col. (1995), o mediante lipofección de la enzima en las células, tal como se describe en Baubonis y col. (1993); (b) transfectando el hospedador celular con un vector que comprende la secuencia codificadora de Cre operativamente unida a un promotor funcional presente en el hospedador celular recombinante, promotor que es opcionalmente inducible, y siendo dicho vector introducido en el hospedador celular recombinante, tal como se describe en Gu y col. (1993) y en Sauer y col., (1988); (c) introduciendo en el genoma del hospedador celular un polinucleótido que comprende la secuencia codificadora de Cre, operativamente unida a un promotor funcional en el hospedador celular recombinante, promotor que es opcionalmente inducible e insertándose dicho polinucleótido en el genoma de la célula del hospedador celular, ya sea por un evento de inserción al azar o por un evento de recombinación homóloga, tal como se describe en Gu y col. (1994).

En una realización específica, el vector que contiene la secuencia que ha de ser insertada en el gen *APM1* mediante recombinación homóloga se construye de manera que los marcadores de selección estén flanqueados por sitios *loxP* en la misma orientación. Los marcadores de selección se pueden retirar mientras que dejen las secuencias de *APM1* de interés que han sido insertadas mediante un evento de recombinación homóloga usando la enzima Cre. De nuevo, se necesitan dos marcadores de selección: un marcador de selección positiva para realizar la selección con respecto al evento recombinante, y un marcador de selección negativa para realizar la selección con respecto al evento de recombinación homóloga. Los vectores y métodos que usan el sistema Cre-*loxP* se encuentran descritos en Zou y col. (1994).

Así, una tercera construcción de DNA preferida comprende desde el extremo 5' hasta el extremo 3': (a) una primera secuencia de nucleótidos que está comprendida en la secuencia genómica *APM1*; (b) una secuencia de nucleótidos que comprende un polinucleótido que codifica un marcador de selección positiva, comprendiendo adicionalmente dicha secuencia de nucleótidos dos secuencias que definen un sitio reconocido por una recombinasa, tal como un sitio *loxP*, estando los dos sitios situados en la misma orientación; y (c) una segunda secuencia de nucleótidos que está comprendida en la secuencia genómica de *APM1*, y que se localiza en el genoma aguas abajo de la primera secuencia de nucleótidos (a) de *APM1*.

Las secuencias que definen un sitio reconocido por una recombinasa, tal como un sitio *loxP*, se sitúan preferiblemente dentro de la secuencia de nucleótidos (b) en posiciones convenientes que limitan la secuencia de nucleótidos para la que se busca la escisión condicional. En una de las realizaciones específicas, dos sitios *loxP* se sitúan a cada uno de los lados de la secuencia del marcador de selección positiva con el fin de permitir su escisión en el momento que desee, una vez que haya tenido lugar el evento de recombinación homóloga.

En una realización preferida del método que usa la tercera construcción de DNA anteriormente descrita, la escisión del fragmento polinucleotídico limitado por los dos sitios reconocidos por una recombinasa, preferiblemente dos sitios *loxP*, se realiza en el momento deseado debido a la presencia dentro del genoma de la célula hospedadora recombinante de una secuencia que codifica la enzima Cre, operativamente unida a una secuencia de promotor, preferiblemente de un promotor inducible, más preferiblemente una secuencia de un promotor específico de tejidos, y muy preferiblemente una secuencia de un promotor que es tanto inducible como específico de tejidos, tal como se describe en Gu y col. (1994).

La presencia de la enzima Cre dentro del genoma de la célula hospedadora recombinante puede ser resultado del cruce de dos animales transgénicos, portando el primer animal transgénico la secuencia de interés derivada de *APM1*, que contiene los sitios *loxP* según se ha descrito anteriormente, y portando el segundo animal transgénico la secuencia codificadora de Cre, operativamente unida a una secuencia de un promotor adecuado, tal como se describe en Gu y col. (1994).

El control espacio-temporal de la expresión de la enzima Cre se puede obtener también con un vector basado en adenovirus que contiene el gen Cre, permitiendo así la infección de células, o la infección *in vivo* de órganos, para suministrar la enzima Cre, tal como se describe en Anton y Graham (1995) y en Kanegae y col. (1995).

Las construcciones de DNA anteriormente descritas se pueden usar para introducir una secuencia de nucleótidos deseada de la invención, preferiblemente una secuencia genómica de *APM1* o una secuencia de cDNA de *APM1*, y muy preferiblemente una copia modificada de una secuencia genómica o de cDNA de *APM1*, dentro de una posición predeterminada del genoma diana, produciéndose o bien la generación de una copia modificada de un gen diana (recombinación homóloga *knock-out*) (en la que se elimina del genoma un gen endógeno), o bien la sustitución de una copia del gen diana por otra copia lo suficientemente homóloga para permitir que tenga lugar un evento de recombinación homóloga (recombinación homóloga *knock-in*) (en la que se inserta en el genoma un gen exógeno). En una

realización específica, las construcciones de DNA anteriormente descritas se pueden usar para introducir una secuencia genómica de *APM1* o una secuencia de cDNA de *APM1* que comprende al menos uno de los marcadores bialélicos de la presente invención, preferiblemente al menos uno de los marcadores bialélicos seleccionados entre el grupo que consiste en de A1 a A8.

Construcciones de DNA nuclear antisentido

Otras composiciones contienen un vector de la invención que comprende un fragmento oligonucleotídico de la secuencia de nucleótidos del SEQ ID NO: 5, preferiblemente un fragmento que incluye el codón de iniciación del gen *APM1*, como herramienta antisentido que inhibe la expresión del correspondiente gen *APM1*. Métodos preferidos que usan un polinucleótido antisentido según la presente invención se describen en Sczakiel y col. (1995) o en la solicitud de PCT documento WO 95/24223, por lo que se incorporan aquí como referencia en su totalidad, incluyendo todas las dibujos, figuras o tablas.

Preferiblemente, las herramientas antisentido se eligen entre los polinucleótidos (de 15 pb-200 pb de longitud) que son complementarios al extremo 5' del RNA de *APM1*. En una de las realizaciones, se usa una combinación de diferentes polinucleótidos antisentido, complementarios a diferentes partes del gen diana deseado.

Los polinucleótidos antisentido preferidos de la invención son complementarios a secuencias de mRNA de *APM1* que contienen o el codón ATG de iniciación a la traducción o un sitio de corte y empalme. Otros polinucleótidos antisentido preferidos son complementarios a un sitio de corte y empalme del mRNA de *APM1*.

Preferiblemente, los polinucleótidos antisentido de la invención tienen una señal de poliadenilación en 3' que ha sido reemplazada por una secuencia de auto-escisión por acción de ribozima, de manera que se producen transcritos de la RNA-polimerasa II sin poli(A) en sus extremos 3'. Estos polinucleótidos antisentido no tienen la capacidad de ser exportados desde el núcleo (Liu y col. (1994)). En una realización preferida, estos polinucleótidos antisentido de *APM1* también comprenden, incluida en la casete de la ribozima, una estructura tallo-bucle de las histonas para estabilizar los transcritos escindidos frente a una degradación 3'-5' exonucleolítica, tal como la estructura descrita por Eckner y col. (1991).

Sondas y cebadores oligonucleotídicos

Los polinucleótidos derivados del gen *APM1* son útiles para detectar la presencia de al menos una copia de una secuencia de nucleótidos del SEQ ID NO: 1, o de uno de sus fragmentos, complementos o variantes, en una muestra a ensayar.

Los cebadores y las sondas según la invención consisten en ácidos nucleicos que comprenden al menos 12, 15, 18, 20, 25, 30, 40, 50 ó 100 nucleótidos consecutivos de un ácido nucleico seleccionado entre el grupo que consiste en:

a) la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 4.811 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella; más en particular, la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 3.528 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

b) la secuencia de nucleótidos que comienza en el nucleótido de la posición 4.853 y que termina en el nucleótido de la posición 15.143 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

c) la secuencia de nucleótidos que comienza en el nucleótido de la posición 15.366 y que termina en el nucleótido de la posición 16.276 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

d) la secuencia de nucleótidos que comienza en el nucleótido de la posición 20.560 y que termina en el nucleótido de la posición 20.966 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

e) la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 22 de la secuencia de nucleótidos del SEQ ID NO: 5, o una de sus variantes o una secuencia complementaria a ella.

Por lo tanto, la invención también se refiere a sondas de ácidos nucleicos caracterizadas porque se hibridan específicamente en condiciones rigurosas de hibridación anteriormente definidas, con un ácido nucleico seleccionado entre el grupo que consiste en:

a) la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 4.811 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes o una secuencia complementaria a ella; más en particular, la secuencia de nucleótidos que comienza en el nucleótido de la posición 1

y que termina en el nucleótido de la posición 3.528 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes o una secuencia complementaria a ella;

b) la secuencia de nucleótidos que comienza en el nucleótido de la posición 4.853 y que termina en el nucleótido de la posición 15.143 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes o una secuencia complementaria a ella;

c) la secuencia de nucleótidos que comienza en el nucleótido de la posición 15.366 y que termina en el nucleótido de la posición 16.276 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes o una secuencia complementaria a ella;

d) la secuencia de nucleótidos que comienza en el nucleótido de la posición 20.560 y que termina en el nucleótido de la posición 20.966 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes o una secuencia complementaria a ella;

e) la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 22 de la secuencia de nucleótidos del SEQ ID NO: 5, o una de sus variantes o una secuencia complementaria a ella.

La formación de híbridos estables depende de la temperatura de fusión (T_m) (del inglés, “*melting Temperature*”) del DNA. La T_m depende de la longitud del cebador o de la sonda, de la fuerza iónica de la disolución y del contenido en G + C. Cuanto mayor es el contenido en G + C del cebador o de la sonda, mayor es la temperatura de fusión debido a que los pares G:C se mantienen unidos por tres enlaces de H, mientras que los pares A:T tienen solamente dos. El contenido en GC en las sondas de la invención normalmente varía entre el 10% y el 75%, preferiblemente entre el 35% y el 60%, y más preferiblemente entre el 40% y el 55%.

Una sonda o un cebador según aquí se describen, tiene entre 8 y 1.000 nucleótidos de longitud y varía preferiblemente desde al menos 8, 10, 12, 15, 18 ó 20 hasta 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 ó 1.000 nucleótidos contiguos de la secuencia de nucleótidos de los SEQ ID NOS: 1-3, o una de sus variantes o una secuencia complementaria a ella, o se especifica que tiene al menos 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 ó 1.000 nucleótidos contiguos de la secuencia de nucleótidos del los SEQ ID NOS: 1-3, o una de sus variantes o una secuencia complementaria a ella. Más en particular, la longitud de estas sondas puede variar desde 8, 10, 15, 20, o 30 nucleótidos hasta 100 nucleótidos, preferiblemente de 10 a 50 nucleótidos, más preferiblemente de 15 a 30 nucleótidos. Las sondas más cortas tienden a presentar una falta de especificidad por una secuencia de ácido nucleico diana y en general requieren temperaturas más frías para formar complejos híbridos relativamente estables con el molde. Las sondas más largas son costosas de producir y a veces pueden auto-hibridarse para formar estructuras de tipo horquilla. La longitud adecuada para los cebadores y las sondas. La longitud adecuada para los cebadores y las sondas bajo un conjunto particular de las condiciones del ensayo pueden ser determinadas empíricamente por un experto en la técnica. Una sonda preferida o un cebador preferido consiste en un ácido nucleico que comprende un polinucleótido seleccionado entre el grupo de secuencias de nucleótidos que consisten en de B1 a B15, de C1 a C15, de D1 a D8 y de E1 a E8.

De manera adicional, otra realización preferida de una sonda según la invención consiste en un ácido nucleico que comprende un marcador bialélico seleccionado entre el grupo que consiste en de A1 a A8, o sus complementos. En la Tabla 4 de los Ejemplos se proporcionan sondas ilustrativas.

Sondas y cebadores particularmente preferidos de la invención incluyen polinucleótidos aislados, purificados o recombinantes, que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1 o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5, ó 10 de las siguientes posiciones nucleotídicas del SEQ ID NO: 1: de 1 a 3.528, de 4.852 a 15.143, de 15.366 a 16.276 y de 20.560 a 20.966. Otros cebadores y sondas preferidos de la invención incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende las posiciones de 41.150 a 4.154, o de 17.169 a 17.170 del SEQ ID NO: 1. Cebadores y sondas adicionales preferidos de la invención incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende una G en la posición 3.787, una G en la posición 3.809, una T en la posición 4.311, una A en la posición 4.328, una A en la posición 4.683 o una A en la posición 15.319 del SEQ ID NO: 1. Otros cebadores y sondas de la invención particularmente preferidos incluyen polinucleótidos aislados, purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5 ó 10 de las posiciones nucleotídicas de 1 a 4.833 del SEQ ID NO: 1.

Otro objetivo de la invención es el constituido por cebadores y sondas purificados, aislados o recombinantes que comprenden la secuencia de nucleótidos del SEQ ID NO: 5, secuencias complementarias a ella, así como sus variantes alélicas y sus fragmentos. Además, los cebadores y sondas preferidos de la invención incluyen los cDNA purificados, aislados o recombinantes de *APM1* que consisten en, que consisten esencialmente en, o que comprenden, la secuencia del SEQ ID NO: 5. Las realizaciones particularmente preferidas de la invención incluyen polinucleótidos aislados,

ES 2 308 852 T3

purificados o recombinantes que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 5, o sus complementos, en los que dicho tramo contiguo comprende posiciones seleccionadas entre el grupo que consiste en un nucleótido T en la posición 93, posiciones 1.154-1.157, un nucleótido G en la posición 1.997, posiciones 2083-2086, un nucleótido C en la posición 2.367, 2.456, 2.467, 2.475 ó 2.631, un nucleótido A en la posición 2.778, posiciones 2.785-2.788, posiciones 2.797-2.801, un nucleótido T en la posición 3.594, un nucleótido G en la posición 3.684, posiciones 3.697-3.701, posiciones 4.026-4.027, un nucleótido T en la posición 4.053, 4.078, 4.533 ó 4.536 del SEQ ID NO: 5. Otras realizaciones particularmente preferidas de la invención incluyen polinucleótidos aislados, purificados o recombinantes, que comprenden un tramo contiguo de al menos 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 ó 1.000 nucleótidos del SEQ ID NO: 5, o sus complementos, en los que dicho tramo contiguo comprende al menos 1, 2, 3, 5 ó 10 de las posiciones nucleotídicas de 1 a 22 del SEQ ID NO: 5.

En una de las realizaciones, la invención abarca polinucleótidos aislados, purificados o recombinantes que consisten en, o que consisten esencialmente en, un tramo contiguo de 8 a 50 nucleótidos del SEQ ID NO: 1, y sus complementos, en los que dicho tramo incluye un marcador bialélico relacionado con APM1 en dicha secuencia; opcionalmente, en los que dicho marcador bialélico relacionado con APM1 se selecciona entre el grupo que consiste en A1, A2, A3, A5, A6 y A7; opcionalmente, en los que dicho marcador bialélico relacionado con APM1 es A4 o A8; opcionalmente, en los que dicho tramo contiguo tiene una longitud de 18 a 35 nucleótidos y dicho marcador bialélico está a una distancia de 4 nucleótidos del centro de dicho polinucleótido; opcionalmente, en los que dicho polinucleótido consiste en dicho tramo contiguo y dicho tramo contiguo tiene una longitud de 25 nucleótidos y dicho marcador bialélico está en el centro de dicho polinucleótido; opcionalmente, en los que el extremo 3' de dicho tramo contiguo está presente en el extremo 3' de dicho polinucleótido; y opcionalmente, en los que el extremo 3' de dicho tramo contiguo está situado en el extremo 3' de dicho polinucleótido y dicho marcador bialélico se encuentra en el extremo 3' de dicho polinucleótido.

En otra realización la invención incluye polinucleótidos aislados, purificados y recombinantes que consisten en, o que consisten esencialmente en, un tramo contiguo de 8 a 50 nucleótidos del SEQ ID NO: 1, o sus complementos, en los que el extremo 3' de dicho tramo contiguo está situado en el extremo 3' de dicho polinucleótido, y en los que el extremo 3' de dicho polinucleótido está situado a una distancia de 20 nucleótidos aguas arriba de un marcador bialélico relacionado con APM1 en dicha secuencia; opcionalmente, en los que dicho marcador bialélico relacionado con APM1 se selecciona entre el grupo que consiste en A1, A2 y A7, o en los que dicho marcador bialélico relacionado con APM1 es A4 o A8; opcionalmente, en los que el extremo 3' de dicho polinucleótido está situado a 1 nucleótido aguas arriba de dicho marcador bialélico relacionado con APM1 en dicha secuencia; y opcionalmente, en los que dicho polinucleótido consiste esencialmente en una secuencia seleccionada entre las siguientes secuencias: D1, D2, D3, D4, D5, D6, D7, D8, E1, E2, E3, E4, E5, E6, E7 y E8.

En otra realización más, la invención abarca polinucleótidos aislados, purificados o recombinantes que consisten en, o que consisten esencialmente en, una secuencia seleccionada entre las siguientes secuencias: B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13, B14, B15, C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14 y C15.

En una realización adicional, la invención incluye polinucleótidos para usar en ensayos de hibridación, ensayos de secuenciación y ensayos de amplificación específica de alelo, para determinar la identidad del nucleótido situado en un marcador bialélico relacionado con APM1 en el SEQ ID NO: 1, o sus complementos, así como también polinucleótidos para usar en la amplificación de segmentos de nucleótidos que comprenden un marcador bialélico relacionado con APM1 en el SEQ ID NO: 1, o sus complementos; opcionalmente, en los que dicho marcador bialélico relacionado con APM1 se selecciona entre el grupo que consiste en A1, A2 y A7, o en los que dicho marcador bialélico relacionado con APM1 es A4 o A8.

Los cebadores y las sondas se pueden preparar mediante cualquier método adecuado, que incluye, por ejemplo, la clonación y restricción de secuencias adecuadas y la síntesis química dirigida mediante un método tal como el método del fosfodiéster de Narang y col. (1979), el método del fosfodiéster de Brown y col. (1979), el método de la dietilfosforamidita de Beaucage y col. (1981) y el método que utiliza un soporte sólido descrito en el documento EP 0 707 592. Las descripciones de todos estos documentos se incorporan aquí como referencia.

Las sondas de detección por lo general son secuencias de ácidos nucleicos o análogos de ácidos nucleicos sin carga tales como, por ejemplo, ácidos nucleicos peptídicos, que se describen en la solicitud de Patente internacional, documento WO 92/20702, y análogos morfolínicos que se describen en las Patentes de EE. UU. Núms. 5.185.444, 5.034.506 y 5.142.047. La sonda puede haberse hecho "no extensible" en el sentido de que dNTP adicionales no se puedan añadir a la sonda. Los análogos por sí mismos normalmente son no extensibles y las sondas de ácido nucleico se pueden hacer no extensibles modificando el extremo 3' de la sonda de manera que el grupo hidroxilo no sea ya capaz de participar en el alargamiento. Por ejemplo, el extremo 3' de la sonda se puede hacer funcional con el marcador de captura o el marcador de detección para así consumir, o de alguna otra manera bloquear, el grupo hidroxilo. Como alternativa, el grupo hidroxilo en 3' puede ser escindido, reemplazado o modificado. La solicitud de Patente de EE. UU. Núm. de serie 07/049.061, presentada el 19 de abril de 1.993, describe modificaciones que se pueden usar para hacer que una sonda sea no extensible.

Cualquiera de los polinucleótidos de la presente invención se pueden marcar, si se desea, incorporando un marcador detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunohistoquímicos o químicos. Por ejemplo, los marcadores útiles incluyen sustancias radiactivas (^{32}P , ^{35}S , ^3H , ^{125}I), colorantes fluorescentes (5-bromodesoxiuridina, fluoresceína, acetilaminofluoreno, digoxigenina) y biotina. Preferiblemente, los polinucleótidos se marcan en sus extremos 3' y/o 5'. Ejemplos de marcaje no radiactivo de fragmentos de ácidos nucleicos se describen en la Patente de Francia documento FR-7810975 o en Urdea y col. (1988) o en Sanchez-Pescador y col. (1988). Además, las sondas según la presente invención pueden tener características estructurales tales que permiten la amplificación de la señal, siendo esas características estructurales, por ejemplo, sondas de DNA ramificado como las descritas por Urdea y col. en 1.991 o en la Patente europea documento EP 0 225 807 (Chiron).

Un marcador también se puede usar para capturar al cebador, con el fin de facilitar la inmovilización o del cebador o de un producto de extensión del cebador, tal como un DNA amplificado, sobre un soporte sólido. Un marcador de captura se une a los cebadores o a las sondas y puede ser un miembro de una unión específica que forma una pareja de unión con el miembro de unión específica del reactivo de fase sólida (p. ej., biotina y estreptavidina). Por lo tanto, dependiendo del tipo de marcador portado por un polinucleótido o por una sonda, éste o ésta se puede utilizar para capturar o para detectar el DNA diana. Además, se entenderá que los polinucleótidos, cebadores o sondas aquí proporcionados, pueden ellos mismos servir como marcador de captura. Por ejemplo, en el caso en el que el miembro de la pareja de unión presente en el reactivo de fase sólida sea una secuencia de ácido nucleico, ésta se puede seleccionar de manera que se una a una porción complementaria de un cebador o de una sonda para de esta manera inmovilizar al cebador o a la sonda en la fase sólida. En los casos en los que la propia sonda polinucleotídica hace de miembro de la pareja de unión, los expertos en la técnica reconocerán que la sonda contendrá una secuencia o una "cola" que no es complementaria a la diana. En el caso en el que el propio cebador polinucleotídico hace de marcador de captura, al menos una porción del cebador estará libre para hibridarse con un ácido nucleico presente sobre una fase sólida. Las técnicas de marcaje de DNA son muy conocidas para el especialista experto.

Las sondas de la presente invención son útiles para varios fines. Se pueden usar en particular en una hibridación Southern a un DNA genómico. Las sondas se pueden también usar para detectar productos de amplificación por PCR. También se pueden usar para detectar desapareamientos en el gen o en el mRNA de *APMI* usando otros métodos muy conocidos en la técnica.

Cualquiera de los polinucleótidos, cebadores y sondas de la presente invención, se pueden inmovilizar convenientemente sobre un soporte sólido. Los soportes sólidos son conocidos por los expertos en la técnica e incluyen las paredes de pocillos de una bandeja de reacción, tubos de ensayo, bolitas de poliestireno, bolitas magnéticas, tiras de nitrocelulosa, membranas, micropartículas tales como partículas de látex, glóbulos rojos de oveja (o de otros animales), *Duracytes* y otros soportes sólidos. El soporte sólido no es crítico y puede ser seleccionado por un experto en la técnica. Así, las partículas de látex, micropartículas, bolitas magnéticas o no magnéticas, membranas, tubos de plástico, paredes de pocillos de microtitulación, chips de vidrio o de silicio, glóbulos rojos de oveja (u otros animales adecuados) y *Duracytes*, constituyen todos ellos ejemplos adecuados. Los métodos adecuados para inmovilizar ácidos nucleicos sobre fases sólidas incluyen interacciones iónicas, hidrófobas, covalentes y métodos similares. Un soporte sólido, según aquí se usa, se refiere a cualquier material que es insoluble o que se puede hacer insoluble en una reacción posterior. El soporte sólido se puede elegir por su capacidad intrínseca para atraer e inmovilizar al reactivo de captura. Como alternativa, la fase sólida puede retener un receptor adicional que posee la capacidad de atraer e inmovilizar al reactivo de captura. El receptor adicional puede incluir una sustancia con carga que tiene una carga opuesta a la del propio reactivo de captura o a la de una sustancia cargada que está conjugada con el reactivo de captura. Como otra alternativa más, la molécula del receptor puede ser cualquier miembro de una unión específica que está inmovilizado en, (unido a), el soporte sólido y que posee la capacidad de inmovilizar al reactivo de captura a través de una reacción de unión específica. La molécula del receptor permite que tenga lugar la unión indirecta del reactivo de captura a un material de soporte sólido antes de la ejecución del ensayo o durante la ejecución del ensayo. La fase sólida, por lo tanto, puede ser un plástico, un derivado de plástico, un metal magnético o no magnético, una superficie de vidrio o de silicio de un tubo de ensayo, un pocillo de microtitulación, láminas, bolitas, micropartículas, glóbulos rojos de oveja (o de otros animales adecuados), *Duracytes*[®] y otras configuraciones conocidas por los que poseen una experiencia normal en la técnica. Los polinucleótidos de la invención se pueden unir a, o inmovilizar sobre, un soporte sólido, individualmente o en grupos de al menos 2, 5, 8, 10, 12, 15, 20 ó 25 polinucleótidos distintos de la invención, en un único soporte sólido. Además, otros polinucleótidos distintos a los de la invención se pueden unir al soporte sólido al que se han unido uno o más de los polinucleótidos de la invención.

Por consiguiente, la invención trata también sobre un método para detectar la presencia de un ácido nucleico que comprende una secuencia de nucleótidos seleccionada entre un grupo que consiste en los SEQ ID NOS: 1-5, uno de sus fragmentos, o una de sus variantes, o una secuencia complementaria a ella, en una muestra, comprendiendo dicho método las siguientes etapas:

a) poner en contacto una sonda de ácido nucleico o una pluralidad de sondas de ácido nucleico que son capaces de hibridarse con una secuencia de nucleótidos incluida en un ácido nucleico seleccionado entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NOS: 1-5, uno de sus fragmentos, o una de sus variantes, o una secuencia complementaria a ella, y la muestra que ha de ser sometida al ensayo.

b) detectar el complejo híbrido formado entre la sonda y un ácido nucleico presente en la muestra.

En una primera realización preferida de este método de detección, dicha sonda de ácido nucleico o la pluralidad de sondas de ácido nucleico están marcadas con una molécula detectable. En una segunda realización preferida de dicho método, dicha sonda de ácido nucleico, o la pluralidad de sondas de ácido nucleico, se ha inmovilizado sobre un sustrato. En una tercera realización preferida, la sonda de ácido nucleico, o la pluralidad de sondas de ácido nucleico, comprenden, o una secuencia que se selecciona entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NOS: de B1 a 15, de C1 a C15, de D1 a D8 y de E1 a E8, o un marcador bialélico seleccionado entre el grupo que consiste en de A1 a A8, o sus complementos.

La invención también concierne a un kit para detectar la presencia de un ácido nucleico que comprende una secuencia de nucleótidos seleccionada entre un grupo que consiste en los SEQ ID NOS: 1-5, uno de sus fragmentos, o una de sus variantes, o una secuencia complementaria a ella, en una muestra, comprendiendo dicho kit:

a) una sonda de ácido nucleico o una pluralidad de sondas de ácido nucleico que se pueden hibridar con una secuencia de nucleótidos incluida en un ácido nucleico seleccionado entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NOS: 1-5, uno de sus fragmentos, o una de sus variantes, o una secuencia complementaria a ella;

b) opcionalmente, los reactivos necesarios para llevar a cabo la reacción de hibridación.

En una primera realización preferida del kit de detección, la sonda de ácido nucleico o la pluralidad de sondas de ácido nucleico están marcadas con una molécula detectable. En una segunda realización preferida del kit de detección, la sonda de ácido nucleico, o la pluralidad de sondas de ácido nucleico, se ha inmovilizado sobre un sustrato. En una tercera realización preferida del kit de detección, la sonda de ácido nucleico o la pluralidad de sondas de ácido nucleico comprenden, o una secuencia que se selecciona entre el grupo que consiste en las secuencias de nucleótidos de los SEQ ID NOS: de B1 a 15, de C1 a C15, de D1 a D8 y de E1 a E8, o un marcador bialélico seleccionado entre el grupo que consiste en de A1 a A8 o sus complementos.

Matrices de oligonucleótidos

Un sustrato que comprende una pluralidad de cebadores o sondas oligonucleotídicas de la invención se puede usar para detectar o para amplificar secuencias diana del gen *APM1* y se puede usar también para detectar mutaciones en las secuencias codificadoras o en las secuencias no codificadoras del gen *APM1*.

Todos los polinucleótidos que aquí se proporcionan se pueden unir a áreas superpuestas o a posiciones elegidas al azar sobre el soporte sólido. Como alternativa, los polinucleótidos de la invención se pueden unir a una matriz ordenada en la que cada polinucleótido se une en una región distinta del soporte sólido que no se superpone con el sitio al que se une cualquier otro polinucleótido. Preferiblemente, esa matriz ordenada de polinucleótidos se diseña para que sea “accesible” cuando las distintas posiciones están registradas y se puede acceder a ellas como parte del procedimiento de un ensayo. Las matrices accesibles de polinucleótidos típicamente comprenden una pluralidad de sondas oligonucleotídicas diferentes que están acopladas a la superficie de un sustrato en diferentes posiciones conocidas. El conocimiento de la localización precisa de cada uno de los polinucleótidos, hace que estas matrices “accesibles” sean particularmente útiles en ensayos de hibridación. Cualquiera de las tecnologías de matrices accesibles conocidas en la técnica se puede utilizar con los polinucleótidos de la invención. Una de las realizaciones particulares de estas matrices de polinucleótidos, conocida como *Genechips*TM, se encuentra descrita en líneas generales en la Patente de EE. UU. 5.143.854 y en las publicaciones de PCT documento WO 90/15070 y documento WO 92/10092. Estas matrices generalmente se pueden producir usando métodos de síntesis mecánica o métodos de síntesis dirigida por la luz que incorporan una combinación de métodos fotolitográficos y una síntesis de oligonucleótidos en fase sólida (Fodor y col., 1991). La inmovilización de matrices de oligonucleótidos sobre soportes sólidos se ha hecho posible gracias al desarrollo de una tecnología identificada en términos generales como “Síntesis a muy gran escala de polímeros inmovilizados”) (VLSIPSTM) (del inglés, “*Very Large Scale Immobilized Polymer Synthesis*”TM) en la que, típicamente, las sondas se inmovilizan en una matriz de alta densidad sobre la superficie sólida de un chip. Ejemplos de tecnologías basadas en VLSIPSTM se proporcionan en las Patentes de EE. UU. 5.143.854 y 5.412.087 y en las publicaciones de PCT documento WO 90/15070, documento WO 92/10092 y documento WO 95/11995, que describen métodos para formar matrices de oligonucleótidos mediante técnicas tales como las técnicas de síntesis dirigida por la luz. Durante el diseño de estrategias que tienen como objetivo proporcionar matrices de oligonucleótidos inmovilizados sobre soportes sólidos, se desarrollaron otras estrategias de presentación para ordenar y presentar las matrices de oligonucleótidos sobre los chips en un intento por maximizar los patrones de hibridación y la información sobre secuencia. En las publicaciones de PCT documento WO 94/12305, documento WO 94/11530, documento WO 97/29212 y documento WO 97/31256 se describen ejemplos de estas estrategias de presentación.

En otra realización de las matrices de oligonucleótidos de la invención, una matriz de sondas oligonucleotídicas se puede usar ventajosamente para detectar mutaciones que tienen lugar en el gen *APM1* y preferiblemente en su región reguladora. Con este fin particular, se diseñan específicamente sondas que contienen una secuencia de nucleótidos que permite su hibridación a los genes que portan mutaciones conocidas (ya sea por delección, inserción o sustitución de uno o varios nucleótidos). Por mutaciones conocidas se entiende mutaciones en el gen *APM1* que se han identificado, por ejemplo, conforme a la técnica usada por Huang y col. (1996) o por Samson y col. (1996).

Otra técnica que se usa para detectar mutaciones en el gen *APM1* es una matriz de DNA de alta densidad. Cada sonda oligonucleotídica que constituye un elemento unitario de la matriz de DNA de alta densidad se diseña para que se empareje con una secuencia específica del DNA genómico o del cDNA de *APM1*. Por lo tanto, una matriz de oligonucleótidos *Apml* de tipo salvaje, complementarios a subsecuencias de la secuencia del gen diana, se usa para determinar la identidad de la secuencia diana, para medir su cantidad y para detectar diferencias entre la secuencia diana y la secuencia de referencia. En uno de estos diseños (una matriz de baldosas 4L) se usa un conjunto de cuatro sondas (A, C, G, T), preferiblemente oligómeros de 15 nucleótidos. En cada conjunto de cuatro sondas, el complemento perfecto se hibridará con más fuerza que las sondas que contienen desapareamientos. Por consiguiente, un ácido nucleico diana de longitud L se explora con respecto a la presencia de mutaciones con una matriz de baldosas que contiene sondas 4L, conteniendo el conjunto total de sondas todas las mutaciones posibles de la secuencia de tipo salvaje conocida de referencia. Las señales de hibridación de la matriz de baldosas que lleva un conjunto de sondas de 15-mer se perturban por el cambio de una sola base en la secuencia diana. Como consecuencia de ello, se produce una pérdida de señal característica o "huella" correspondiente a las sondas que flanquean la posición de una mutación. Esta técnica fue descrita por Chee y col. en 1996, y se incorpora aquí como referencia.

Por consiguiente, la invención concierne a una matriz de moléculas de ácido nucleico que comprende al menos uno de los polinucleótidos anteriormente descritos como sodas y cebadores. Preferiblemente, la invención concierne a una matriz de secuencias de ácido nucleico que comprende al menos dos de los polinucleótidos anteriormente descritos como sondas y cebadores.

Otro objetivo más de la invención consiste en una matriz de secuencias de ácido nucleico que comprende, o al menos una de las secuencias seleccionadas entre el grupo que consiste en los SEQ ID NOS: de B1 a B15, de C1 a C15, de D1 a D8 y de E1 a E8, o las secuencias complementarias a ellas, o uno de sus fragmentos, o al menos sus 8, 10, 12, 15, 18, 20, 25, 30 ó 40 nucleótidos consecutivos, o al menos una secuencia que comprende un marcador bialélico seleccionado entre el grupo que consiste en de A1 a A8, o sus complementos.

La invención también concierne a una matriz de secuencias de ácido nucleico que comprende al menos dos de las secuencias seleccionadas entre el grupo que consiste en los SEQ ID NOS: de B1 a B15, de C1 a C15, de D1 a D8 y de E1 a E8, o las secuencias complementarias a ellas, o uno de sus fragmentos, o al menos sus 8 nucleótidos consecutivos, o al menos dos secuencias que comprenden una marcador bialélico seleccionado entre el grupo que consiste en de A1 a A8, o sus complementos.

Proteínas APM1 y fragmentos polipeptídicos de APM1

La expresión "polipéptidos APM1" se usa aquí para abarcar las proteínas y los polipéptidos de la presente invención. También forman parte de la invención los polipéptidos codificados por los polinucleótidos de la invención, así como los polipéptidos de fusión que comprenden esos polipéptidos. La invención describe proteínas APM1 procedentes de seres humanos, que incluyen proteínas APM1 aisladas o purificadas, que consisten en, que consisten esencialmente en, o que comprenden, la secuencia del SEQ ID NO: 6. Se debe destacar que las proteínas APM1 de la invención están basadas en la variante presente en la naturaleza de la secuencia de aminoácidos de la APM1 humana, en la que el residuo de ácido aspártico de la posición del aminoácido 56 se ha reemplazado con un residuo de ácido glutámico. Esta proteína variante y sus fragmentos que contienen la posición del aminoácido 56 del SEQ ID NO: 6, se denominan aquí colectivamente "variantes 56-Glu".

También se describen polipéptidos aislados, purificados y recombinantes, que codifican polipéptidos que comprenden un tramo contiguo de al menos 6 aminoácidos, preferiblemente de al menos de 8 a 10 aminoácidos, más preferiblemente de al menos 12, 20, 25, 30, 40, 50 ó 100 aminoácidos del SEQ ID NO: 6, en los que dicho tramo contiguo incluye un residuo de ácido glutámico en la posición del aminoácido 56 en el SEQ ID NO: 6. En otras realizaciones preferidas, la extensión contigua de aminoácidos comprende el sitio de una mutación o de una mutación funcional, que incluye una delección adición, intercambio o truncamiento de los aminoácidos de la secuencia de proteínas de APM1.

Las proteínas APM1 se aíslan preferiblemente procedentes de muestras de tejidos humanos o de mamíferos, o son expresadas por genes humanos o de mamíferos. Los APM1 se pueden fabricar usando métodos de expresión de rutina, conocidos en la técnica. El polinucleótido que codifica el polipéptido deseado, se liga dentro de un vector de expresión adecuado para cualquier hospedador conveniente. Se usan sistemas hospedadores tanto de eucariotas como de procariotas para formar polipéptidos recombinantes, y en el presente documento se proporciona un sumario de alguno de los sistemas más comunes. El polipéptido después se aísla a partir de células lisadas o a partir de medio de cultivo y se purifica hasta obtener el grado necesario para el uso que de él se requiere. La purificación se realiza mediante cualquiera de los métodos conocidos en la técnica, por ejemplo, extracción diferencial, fraccionamiento con sales, cromatografía, centrifugación y métodos similares. Véase, por ejemplo, *Methods in Enzymology* en relación con diversos métodos para purificar proteínas.

Además, se pueden producir fragmentos de proteína más cortos mediante síntesis química. Como alternativa, las proteínas se extraen de las células o de los tejidos de seres humanos o de animales no humanos. Los métodos para purificar proteínas son conocidos en la técnica e incluyen el uso de detergentes o de agentes caotrópicos para disgregar las partículas, seguido de una extracción diferencial y de la separación de los polipéptidos mediante cromatografía de intercambio iónico, cromatografía de afinidad, sedimentación conforme a la densidad y electroforesis en gel, por ejemplo.

Cualquier cDNA de APM1, incluyendo el SEQ ID NO: 5, se puede usar para expresar proteínas y polipéptidos APM1. El ácido nucleico que codifica la proteína o el polipéptido que ha de ser expresado, se une operativamente a un promotor presente en un vector de expresión usando la tecnología convencional de clonación. El inserto que contiene APM1 presente en el vector de expresión, puede comprender la secuencia codificadora completa de la proteína APM1 o una de sus porciones. Por ejemplo, el inserto derivado de APM1 puede codificar un polipéptido que comprende al menos 10 aminoácidos consecutivos de la proteína APM1 del SEQ ID NO: 6, en el que dichos aminoácidos consecutivos comprenden un residuo de ácido glutámico en la posición del aminoácido 56.

El vector de expresión es cualquiera de los sistemas de expresión de mamíferos, levaduras, insectos o bacterias conocidos en la técnica. Los vectores y los sistemas de expresión comercialmente disponibles, se encuentran disponibles procedentes de diversos proveedores que incluyen el *Genetics Institute* (Cambridge, MA), *Stratagene* (La Jolla, California), *Promega* (Madison, Wisconsin) e *Invitrogen* (San Diego, California). Si se desea, para aumentar la expresión y facilitar el plegamiento correcto de la proteína, el contexto de los codones y el emparejamiento de los codones de la secuencia se pueden optimizar para su expresión en el organismo en el que el vector de expresión se introduce, según se explica en Hatfield y col., Patente de EE. UU. Núm. 5.082.767.

En una de las realizaciones, la secuencia codificadora completa del cDNA de APM1 hasta la señal de poliA del cDNA, está operativamente unida a un promotor en el vector de expresión. Como alternativa, si al ácido nucleico que codifica una porción de la proteína APM1 le falta la metionina que hace de sitio de iniciación, se puede introducir una metionina de iniciación al lado del primer codón del ácido nucleico usando técnicas convencionales. De manera similar, si el inserto procedente del cDNA de APM1 carece de señal de poliA, esta secuencia se puede añadir a la construcción, por ejemplo, extrayendo mediante corte y empalme la señal de poliA procedente del plásmido pSG5 (Stratagene) usando las enzimas endonucleasas de restricción BglII y SalI, e incorporándola en el vector pXTI (Stratagene) de expresión en mamíferos. pXTI contiene las LTR y una porción del gen *gag* procedente del virus de la leucemia murina de Moloney. La posición de las LTR en la construcción permite que tenga lugar una transfección eficiente y estable. El vector incluye el promotor de la timidina-quinasa del virus del herpes simple y el gen de selección de resistencia a neomicina. El ácido nucleico que codifica la proteína APM1 o una de sus porciones, se obtiene mediante PCR a partir de un vector bacteriano que contiene el cDNA de APM1 del SEQ ID NO: 5. Los cebadores oligonucleotídicos usados son complementarios al cDNA de APM1, o a una de sus porciones, y contienen secuencias para endonucleasas de restricción, para Pst I incorporada en el cebador 5', y para BglIII incorporada en el extremo 5' del correspondiente cebador 3' del cDNA, teniendo cuidado de comprobar que la secuencia que codifica la proteína APM1, o una de sus porciones, está situada correctamente con respecto a la señal de poliA. El fragmento purificado obtenido de la reacción de PCR resultante, se digiere con PstI, se hace romo en sus extremos con una exonucleasa, se digiere con BglIII, se purifica y se liga a pXTI.

Con el producto ligado se transfectan células NIH 3T3 de ratón usando Lipofectina (*Life Technologies, Inc., Grand Island*, Nueva York) en las condiciones descritas en las especificaciones del producto. Los transfectantes positivos se seleccionan después de hacer crecer las células transfectadas en una concentración de 600 µg/ml de G418 (Sigma, St. Louis, Missouri).

Como alternativa, los ácidos nucleicos que codifican la proteína APM1 o una de sus porciones se clonan en pED6dpc2 (*Genetics Institute*, Cambridge, MA). Con la construcción/construcciones pED6dpc2 resultantes se transfectan células hospedadoras adecuadas, tales como las células COS 1. Las células resistentes a metotrexato se seleccionan y se expanden.

Los procedimientos anteriormente mencionados se pueden también usar para expresar una proteína APM1 mutante responsable de un fenotipo detectable, o una de sus porciones.

Las proteínas expresadas se purifican usando técnicas de purificación convencionales, tales como la precipitación con sulfato amónico o la separación cromatográfica basada en el tamaño o la carga. La proteína codificada por el inserto del ácido nucleico, se puede también purificar usando técnicas estándar de inmunocromatografía. En estos procedimientos, una disolución que contiene la proteína APM1 expresada o una de sus porciones, tal como un extracto celular, se aplica a una columna que contiene anticuerpos contra la proteína APM1, o contra una de sus porciones, unidos a la matriz cromatográfica. Se deja que la proteína expresada se una a la columna de inmunocromatografía. Seguidamente, la columna se lava para retirar las proteínas unidas inespecíficamente. La proteína expresada y unida de manera específica, se separa luego de la columna y se recupera usando técnicas estándar.

Para confirmar la expresión de la proteína APM1 o de una de sus porciones, las proteínas expresadas en las células hospedadoras que contienen un vector de expresión que contiene un inserto que codifica la proteína APM1 o una de sus porciones, se pueden comparar con las proteínas expresadas en células hospedadoras que contienen el vector de expresión que no lleva un inserto. La presencia de una banda en las muestras procedentes de las células que contienen el vector de expresión que lleva un inserto, que está ausente en las muestras procedentes de células que contienen el vector de expresión sin un inserto, indica que se está expresando la proteína APM1 o una de sus porciones. Generalmente, la banda tendrá la movilidad que se espera que tenga la proteína APM1 o una de sus porciones. Sin embargo, la banda puede tener una movilidad diferente de la esperada como resultado de modificaciones tales como glicosilación, ubiquitinación o escisión enzimática.

A continuación se describen anticuerpos capaces de reconocer específicamente a la proteína APM1 expresada o una de sus porciones.

Si la producción de anticuerpos no fuera posible, los ácidos nucleicos que codifican la proteína APM1 o una de sus porciones, se pueden incorporar en vectores de expresión diseñados para usar en esquemas de purificación utilizando polipéptidos quiméricos. En estas estrategias, el ácido nucleico que codifica la proteína APM1 o una de sus porciones, se inserta en el marco de lectura con un gen que codifica la otra mitad de la quimera. La otra mitad de la quimera puede ser una secuencia que codifica un polipéptido de unión a β -globina o de unión a níquel, por ejemplo. Una matriz cromatográfica que lleva unido a ella níquel o un anticuerpo contra β -globina se puede usar entonces para purificar la proteína quimera. Por ingeniería genética se construyen sitios de escisión por proteasas entre el gen de la β -globina o el polipéptido de unión a níquel y la proteína APM1 o una de sus porciones. De esta manera, los dos polipéptidos de la quimera se separan uno de otro mediante digestión con proteasas.

Uno de los vectores de expresión útiles para generar quimeras con β -globina es pSG5 (*Stratagene*), que codifica la β -globina de conejo. El intrón II del gen de la β -globina de conejo facilita el corte y empalme del transcrito expresado, y la señal de poliadenilación incorporada en la construcción aumenta el nivel de expresión. Estos métodos son muy conocidos por los expertos en la técnica de biología molecular. Existen publicados métodos estándar en textos sobre métodos tales como Davis y col., (*Basic Methods in Molecular Biology*, L. G. Davis, M. D. Digner y J. F. Battey, redactores, *Elsevier Press*, NY, 1986) y muchos de los métodos se encuentran disponibles procedentes de *Stratagene*, *Life Technologies, Inc.*, o procedentes de *Promega*. Se pueden producir además polipéptidos desde la construcción que usa sistemas de traducción *in vitro* tales como el *In vitro ExpressTM Translation Kit* (*Stratagene*).

Anticuerpos que se unen a polipéptidos APM1 de la invención

Se puede usar cualquier polipéptido APM1 o la proteína APM1 completa para generar anticuerpos capaces de unirse específicamente a una proteína APM1 expresada o a sus fragmentos según se han descrito. Las composiciones de anticuerpos son capaces de realizar una unión específica con, o de unirse específicamente a, la variante 56-Glu de la proteína APM1. En el caso de una composición de anticuerpos que se une específicamente a la variante 56-Glu de APM1, esa composición debe demostrar que presenta una afinidad de unión mayor que al menos el 5%, 10%, 15%, 20%, 25%, 50% o 100% por la variante 56-Glu de APM1 que por la variante 56-Asp de APM1, en un ensayo ELISA, RIA u otro ensayo basado en anticuerpos.

En una realización preferida, las composiciones de anticuerpos son capaces de realizar una unión selectiva con, o de unirse selectivamente a, un fragmento de un polipéptido que contiene epítipo, que comprende un tramo contiguo de al menos 6 aminoácidos, preferiblemente de al menos de 8 a 10 aminoácidos, más preferiblemente de al menos 12, 15, 20, 25, 30, 40, 50 ó 100 aminoácidos del SEQ ID NO: 6, en el que dicho epítipo comprende un residuo de ácido glutámico en la posición del aminoácido 56 del SEQ ID NO: 6, en el que dicha composición de anticuerpos opcionalmente es policlonal o monoclonal.

También se describe el uso de polipéptidos que comprenden un tramo contiguo de al menos 6 aminoácidos, preferiblemente de al menos de 8 a 10 aminoácidos, más preferiblemente de al menos 12, 20, 25, 30, 40, 50 ó 100 aminoácidos de un polipéptido APM1, en la fabricación de anticuerpos, en los que dicho tramo contiguo comprende un residuo de ácido glutámico en la posición del aminoácido 56 en el SEQ ID NO: 6. En una realización preferida, esos polipéptidos son útiles en la fabricación de anticuerpos para detectar la presencia y ausencia de la variante 56-Glu.

Los animales y mamíferos no humanos, ya sean de tipo salvaje o transgénicos, que expresan una especie de APM1 diferente de la especie con la que se desea que tenga lugar la unión al anticuerpo, y los animales que no expresan APM1 (es decir, un animal *knock out* en el que se ha eliminado la expresión de APM1, según aquí se ha descrito) son particularmente útiles para preparar anticuerpos. Los animales *knock out* en los que se ha eliminado la expresión de APM1, reconocerán la totalidad o la mayor parte de las regiones expuestas de APM1 como antígenos extraños, y por lo tanto producirán anticuerpos con una matriz más amplia de epítopos de APM1. Por otra parte, polipéptidos más pequeños, de solamente de 10 a 30 aminoácidos, pueden ser útiles para obtener una unión específica a la variante 56-Glu. Además, el sistema inmunitario humoral de los animales que producen una especie de APM1 que se asemeja a la secuencia antigénica, reconocerá preferentemente las diferencias entre la especie APM1 natural del animal y la secuencia del antígeno. Esta técnica será particularmente útil para obtener anticuerpos que se unen específicamente a la variante 56-Glu.

Amplificación del gen de APM1

1. Extracción de DNA

En lo que se refiere a la fuente del DNA genómico que ha de ser sometida a análisis, se puede usar casi cualquier muestra a ensayar sin una limitación particular. Estas muestras a ensayar incluyen muestras biológicas que se pueden ensayar con los métodos de la presente invención aquí descritos, e incluyen fluidos biológicos humanos y de animales tales como sangre completa, suero, plasma, líquido cerebroespinal, orina, líquidos linfáticos y diversas secreciones externas de los aparatos respiratorio, intestinal y genitourinario, lágrimas, saliva, leche, glóbulos blancos, mielomas y muestras similares; fluidos biológicos tales como sobrenadantes de cultivos celulares; especímenes de tejidos fijados

que incluyen tejido tumoral y no tumoral, y tejidos de los ganglios linfáticos; aspirados de médula ósea y especímenes de células fijadas. La fuente de DNA genómico preferida, usada en el contexto de la presente invención, es sangre venosa periférica de cada donante.

- 5 Las técnicas de extracción de DNA son muy conocidas por el especialista con una experiencia normal en la técnica. Esas técnicas se encuentran descritas en particular en Lin y col. (1998) y en Mackey y col. (1998).

2. Amplificación de DNA

- 10 Las técnicas de amplificación de DNA son muy conocidas por los expertos en la técnica. Las técnicas de amplificación que se pueden usar en el contexto de la presente invención incluyen, pero no se limitan a, la reacción en cadena de la ligasa (LCR) descrita en los documentos EP-A-320 308, WO 9320227 y EP-A-439 182, cuyas descripciones se incorporan aquí como referencia, la reacción en cadena de la polimerasa (PCR, RT-PCR) y técnicas tales como la amplificación basada en una secuencia de ácido nucleico (NASBA) (del inglés, "*Nucleic Acid Sequence Based*
15 *Amplification*") descrita en Guatelli J. C. y col. (1990) y en Compton J. (1991), amplificación Q-beta según se describe en la solicitud de Patente europea Núm. 4544610, amplificación por desplazamiento de la cadena según se describe en Walker y col. (1996) y en el documento EP A 684 315, y amplificación mediada por una diana según se describe en la publicación de PCT documento WO 9322461, cuyas descripciones se incorporan aquí como referencia. En el caso de la amplificación de los mRNA, dentro del alcance de la presente invención está la transcripción inversa de mRNA
20 en cDNA seguida de la reacción en cadena de la polimerasa (RT-PCR) (del inglés, "*Reverse Transcription followed by Polymerase Chain Reaction*"); o el uso de una única enzima en las dos etapas según se describe en la Patente de EE. UU. Núm. 5.322.770, o el uso de una LCR basada en huecos asimétricos (RT-AGLCR) (del inglés, "*Reverse Transcription-Asymmetric Gap LCR*") según se describe en Marshall y col. (1994). La AGLCR es una modificación de la GLCR que permite la amplificación de RNA.

- 25 La tecnología de PCR es la técnica de amplificación preferida usada en la presente invención. Diversas técnicas de PCR son familiares a los expertos en la técnica. En cuanto a una revisión de la tecnología de PCR, véase White (1997) y la publicación titulada "*PCR Methods and Applications*" (1991, Cold Spring Harbor Laboratory Press). En cada uno de estos procedimientos de PCR, los cebadores de PCR de cada lado de las secuencias de ácido nucleico que se han de amplificar, se añaden a una muestra de ácido nucleico convenientemente preparada, junto con los
30 dNTP y con una polimerasa termoestable tal como la Taq polimerasa, Pfu polimerasa o Vent polimerasa. El ácido nucleico de la muestra se desnaturaliza y los cebadores de PCR se hibridan de manera específica con las secuencias complementarias del ácido nucleico de la muestra. Los cebadores hibridados se extienden. Después se inicia otro ciclo de desnaturalización, hibridación y extensión. Los ciclos se repiten muchas veces para producir un fragmento amplificado que contiene la secuencia de ácido nucleico entre los sitios de los cebadores. La PCR se encuentra descrita más extensamente en varias patentes que incluyen la Patentes de EE. UU. Núms. 4.683.195, 4.683.202 y 4.965.188. Cada una de estas publicaciones se incorpora aquí como referencia.

- Uno de los aspectos de la presente invención es un método para la amplificación del gen *APM1* humano, en particular de la secuencia genómica del SEQ ID NO: 1 o de la secuencia de cDNA del SEQ ID NO: 5, o de uno de sus fragmentos o de una de sus variantes, en una muestra a ensayar, preferiblemente usando la tecnología de PCR. Este método comprende las etapas de poner en contacto una muestra a ensayar de la que se sospecha que contiene la secuencia diana que codifica *APM1*, o una de sus porciones, con reactivos de la reacción de amplificación que comprenden una pareja de cebadores de amplificación, y finalmente en algunos casos, con una sonda de detección que
45 se puede hibridar con una región interna de las secuencias del amplicón para confirmar que la reacción de amplificación deseada ha tenido lugar.

- Por lo tanto, la presente invención también se refiere a un método para la amplificación de una secuencia del gen *APM1* humano, en particular de una porción de las secuencias genómicas del SEQ ID NO: 1 o de la secuencia de cDNA del SEQ ID NO: 5, o de una de sus variantes, en una muestra a ensayar, comprendiendo dicho método las etapas de:

- a) poner en contacto una muestra a ensayar de la que se sospecha que contiene la secuencia del gen *APM1* diana comprendida en una secuencia de nucleótidos seleccionada entre un grupo que consiste en los SEQ ID NOS: 1 y 5, o
55 sus fragmentos, o sus variantes, con reactivos de la reacción de amplificación que comprenden una pareja de cebadores de amplificación según se han descrito en lo que antecede y que se sitúan a ambos lados de la región polinucleotídica que ha de ser amplificada, y

- b) opcionalmente, detectar los productos resultantes de la amplificación.

- 60 En una primera realización preferida del anterior método de amplificación, el producto de la amplificación se detecta mediante hibridación con una sonda marcada que contiene una secuencia que es complementaria a la región amplificada. En una segunda realización preferida, los cebadores de ácido nucleico comprenden una secuencia que se selecciona entre el grupo que consiste en de B1 a B15, de C1 a C15, de D1 a D8 y de E1 a E8. Los cebadores se caracterizan más concretamente porque tienen suficiente complementariedad con cualquier secuencia de una cadena de
65 la secuencia genómica próxima a la región que ha de ser amplificada, por ejemplo, con una secuencia no codificadora adyacente a los exones que han de ser amplificados.

ES 2 308 852 T3

La invención también concierne a un kit para la amplificación de una secuencia del gen *APM1* humano, en particular de una porción de la secuencia genómica del SEQ ID NO: 1 o de la secuencia de cDNA del SEQ ID NO: 5, o una de sus variantes, en una muestra a ensayar, en el que dicho kit comprende:

- 5 a) una pareja de cebadores oligonucleotídicos situados a ambos lados de la región de *APM1* que ha de ser amplificada;
 - b) opcionalmente, los reactivos necesarios para llevar a cabo la reacción de amplificación.
- 10 En una de las realizaciones del kit de amplificación anteriormente mencionado, el producto de la amplificación se detecta mediante hibridación con una sonda marcada que tiene una secuencia que es complementaria a la de la región amplificada.
- 15 En otra de las realizaciones del kit de amplificación anteriormente mencionado, los cebadores comprenden una secuencia que se selecciona entre el grupo que consiste en de B1 a B15, de C1 a C15, de D1 a D8 y de E1 a E8.

Marcadores bialélicos relacionados con APM1

- 20 Los marcadores bialélicos generalmente consisten en un polimorfismo en la posición de una sola base. Cada uno de los marcadores bialélicos corresponde por lo tanto a las dos formas de una secuencia de polinucleótidos que, cuando se comparan una con la otra, presentan la modificación de un nucleótido en una sola posición. Normalmente, la modificación de un nucleótido implica la sustitución de uno nucleótido por otro (por ejemplo C en lugar de T).

Ventajas de los marcadores bialélicos de la presente invención

- 25 Los marcadores bialélicos relacionados con *APM1* de la presente invención ofrecen varias ventajas importantes sobre otros marcadores genéticos tales como los marcadores RFLP (polimorfismo de longitud de fragmentos de restricción) y los marcadores VNTR (número variable de repeticiones en tándem).

- 30 La primera generación de marcadores fueron los RFLP, que son variaciones que modifican la longitud de un fragmento de restricción. Pero los métodos usados para identificar y tipificar los RFLP representan un desperdicio relativo de materiales, esfuerzo y tiempo. La segunda generación de marcadores genéticos fueron los VNTR, que se pueden clasificar en minisatélites o microsatélites. Los minisatélites son secuencias de DNA repetidas en tándem, que se presentan en unidades de 5-50 repeticiones que se distribuyen a lo largo de regiones de los cromosomas humanos que
- 35 tienen una longitud que varía de 0,1 kilobases a 20 kilobases. Debido a que presentan muchos alelos posibles, su contenido informativo es muy elevado. Los minisatélites se puntúan realizando transferencias Southern para identificar el número de repeticiones en tándem presentes en una muestra de ácido nucleico procedentes del individuo sometido al ensayo. Sin embargo, existen solamente 10^4 VNTR potenciales que se puedan tipificar mediante transferencias Southern. Además, la generación y el ensayo en grandes cantidades tanto de los marcadores RFLP como de los marcadores
- 40 VNTR son costosos y requieren mucho tiempo.

- Los marcadores de polimorfismos de un solo nucleótido o marcadores bialélicos se pueden usar de la misma manera que los RFLP y los VNTR pero ofrecen varias ventajas. Los SNP están densamente distribuidos en el genoma humano y representan el tipo de variación más frecuente. Un número estimado de más de 10^7 sitios están diseminados
- 45 por los 3×10^9 pares de bases del genoma humano. Por lo tanto, los SNP ocurren con una frecuencia mayor y con una uniformidad mayor que los marcadores RFLP o VNTR, lo que significa que existe una mayor probabilidad de que un marcador se encuentre muy próximo a un locus de interés. Los marcadores SNP son menos variables que los marcadores VNTR pero son más estables desde el punto de vista mutacional.

- 50 Asimismo, las diferentes formas de un polimorfismo de un solo nucleótido caracterizado, tal como los marcadores bialélicos de la presente invención, suelen ser más fáciles de distinguir y pueden por lo tanto ser tipificados fácilmente de manera rutinaria. Los marcadores bialélicos contienen alelos basados en un solo nucleótido y tienen sólo dos alelos comunes, lo que permite una detección muy paralela y una puntuación automática. Los marcadores bialélicos de la presente invención ofrecen la posibilidad de realizar una genotipificación rápida y de alta capacidad de procesamiento de un gran número de individuos.
- 55

- Los marcadores bialélicos se encuentran densamente distribuidos en el genoma, son suficientemente informativos y pueden ser analizados en grandes cantidades. Los efectos combinados de estas ventajas hacen que los marcadores bialélicos sean extremadamente valiosos para estudios genéticos. Los marcadores bialélicos se pueden usar en estudios de ligamiento en familias, en métodos de reparto de alelos, en estudios sobre el desequilibrio de ligamiento en poblaciones, en estudios de asociación de poblaciones caso-control o de poblaciones positivas con respecto a un rasgo o negativas con respecto a un rasgo. Un aspecto importante de la presente invención es que los marcadores bialélicos permiten que se puedan llevar a cabo estudios de asociación para identificar genes implicados en rasgos complejos. Los estudios de asociación investigan la frecuencia de los alelos de los marcadores en poblaciones caso y en poblaciones control no emparentadas y generalmente se utilizan para la detección de rasgos poligénicos o esporádicos. Los estudios de asociación se pueden llevar a cabo en la población general y no se limitan a estudios realizados con individuos emparentados de las familias afectadas (estudios de ligamiento). Los marcadores bialélicos de genes diferentes pueden ser escrutados en paralelo con respecto a una asociación directa con una enfermedad o con la respuesta a un
- 60
- 65

tratamiento. Esta estrategia génica múltiple es una potente herramienta para realizar diversos estudios genéticos en seres humanos ya que proporciona el poder estadístico necesario para investigar el efecto sinérgico de múltiples factores genéticos sobre un fenotipo particular, la respuesta a un fármaco, un rasgo esporádico o un estado patológico particular que tiene una etiología genética compleja.

Gen candidato de la presente invención

Se pueden utilizar diferentes estrategias para llevar a cabo estudios de asociación: estudios de asociación en todo el genoma, estudios de asociación en regiones candidatas y estudios de asociación en genes candidatos. Los estudios de asociación en todo el genoma se basan en el escrutinio de marcadores genéticos distribuidos uniformemente y que cubren el genoma completo. La estrategia que usa el gen candidato se basa en el estudio de marcadores genéticos específicamente situados en genes que están potencialmente implicados en una ruta biológica relacionada con el rasgo de interés. En la presente invención, el gen candidato es *APM1*. De hecho el gen *APM1* parece estar implicado en la obesidad y en otros trastornos relacionados con la obesidad. El análisis del gen candidato proporciona claramente una estrategia de tipo atajo para la identificación de genes y de polimorfismos génicos relacionados con un rasgo particular cuando se dispone de cierta información concerniente a la biología de ese rasgo. Sin embargo, se debe señalar que todos los marcadores bialélicos descritos en la presente solicitud se puede emplear como parte de estudios de asociación en todo el genoma o como parte de estudios de asociación en regiones candidatas, y esos usos se contemplan de manera específica en la presente invención y en las reivindicaciones.

Marcadores bialélicos relacionados con APM1 y polinucleótidos relacionados con ellos

La invención también concierne a marcadores bialélicos relacionados con *APM1*. Según aquí se usa, la expresión “marcador bialélico relacionado con *APM1*” se refiere a un conjunto de marcadores bialélicos en desequilibrio de ligamiento con el gen *APM1*. La expresión marcador bialélico relacionado con *APM1* incluye los marcadores bialélicos designados de A1 a A8.

Una porción de los marcadores bialélicos de la presente invención se describe en las Tablas A y B. Su posición en el gen *APM1* se indica en las Tablas A y B y también se indica como polimorfismo de una sola base en las características del SEQ ID NO: 1. Las parejas de cebadores que permiten la amplificación de un ácido nucleico que contiene la base polimórfica de uno de los marcadores bialélicos de *APM1* están listadas en la Tabla 1 del Ejemplo 2.

TABLA A

Lista de marcadores bialélicos rodeados por una secuencia que nunca antes ha sido propuesta en la técnica

Marcador bialélico	Nombre del marcador	Situación en el gen <i>APM1</i>	Polimorfismo	Frecuencia del alelo 2	Posición del marcador en SEQ ID NO: 1
A1	9-27/261	Región reguladora en 5'	Alelo 1: G Alelo 2: C		3787
A2	99-14387/129	Intrón 1	Alelo 1: A Alelo 2: C		11118
A3	9-12/48	Intrón 1	Alelo 1: T Alelo 2: C	1,5%	15120
A5	9-12/355 ó 9-13/297	Intrón 2	Alelo 1: G Alelo 2: T	26%	15427
A6	9-12/355 ó 9-13/297	Intrón 2	Alelo 1: A Alelo 2: G	11%	15500
A7	99-14405/105	Intrón 2	Alelo 1: G Alelo 2: A	37%	15863

TABLA B

Lista de marcadores bialélicos rodeados por una secuencia previamente propuesta, en los que uno de los alelos, el alelo 2, nunca antes ha sido propuesto en la técnica. La identidad del nucleótido del alelo original ha sido propuesta previamente en la técnica

Marcador bialélico	Nombre del marcador	Situación en el gen <i>APM1</i>	Polimorfismo	Frecuencia del alelo 2	Posición del marcador en SEQ ID NO: 1
A4	9-12/124 o 9-13/66	Exón 2	Alelo 1: T Alelo 2: G	11,5%	15196
A8	9-16/189	Exón 3	Alelo 1: A Alelo 2: Del	40%	17170

La invención también se refiere a una secuencia de nucleótidos purificada y/o aislada que comprende una base polimórfica de un marcador bialélico situado en la secuencia del gen *APM1*, preferiblemente de un marcador bialélico seleccionado entre el grupo que consiste en A1, A2, A3, A4, A5, A6, A7 y A8, y sus complementos. La secuencia tiene entre 8 y 1.000 nucleótidos de longitud, y preferiblemente comprende al menos 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 ó 1.000 nucleótidos contiguos de una secuencia de nucleótidos seleccionada entre el grupo que consiste en los SEQ ID NOS: 1 y 5, o una de sus variantes o una de sus secuencias complementarias. Estas secuencias de nucleótidos comprenden la base polimórfica del alelo 1 o del alelo 2 del marcador bialélico considerado. Opcionalmente, dicho marcador bialélico puede estar a una distancia de 6, 5, 4, 3, 2 ó 1 nucleótidos del centro de dicho polinucleótido, o puede estar en el centro de dicho polinucleótido. Opcionalmente, el extremo 3' de dicho tramo contiguo puede encontrarse en el extremo 3' de dicho polinucleótido. Opcionalmente, el marcador bialélico puede encontrarse en el extremo 3' de dicho polinucleótido. Opcionalmente, el extremo 3' de dicho polinucleótido puede estar situado a una distancia de, o al menos a, 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500 ó 1.000 nucleótidos aguas arriba de un marcador bialélico del gen *APM1* en dicha secuencia. Opcionalmente, el extremo 3' de dicho polinucleótido puede estar situado 1 nucleótido aguas arriba de un marcador bialélico del gen *APM1* en dicha secuencia. Opcionalmente, dicho polinucleótido puede comprender también un marcador. Opcionalmente, dicho polinucleótido puede estar unido a un soporte sólido. En otra realización, los polinucleótidos anteriormente definidos se pueden usar solos o en cualquier combinación.

En una realización preferida, las secuencias que comprenden una base polimórfica de uno de los marcadores bialélicos listados en las Tablas A y B se seleccionan entre el grupo que consiste en las secuencias de nucleótidos que contiene un tramo contiguo de, que consisten en, que están comprendidas en, o que comprenden, un polinucleótido seleccionado entre el grupo que consiste en las secuencias de ácido nucleico expuestas como los Núms. 9-27, 99-14.387, 9-12, 9-13, 99-14.405 y 9-16 (listados en la Tabla 1), o una de sus variantes, o una de secuencias complementarias.

La invención también concierne a un ácido nucleico que codifica la proteína *APM1*, en el que dicho ácido nucleico comprende una base polimórfica de un marcador bialélico seleccionado entre el grupo que consiste en A1, A2, A4, A7 y A8, y sus complementos.

La invención también incluye el uso de cualquier polinucleótido para, o de cualquier polinucleótido para usar en, determinar la identidad de uno o más nucleótidos situados en un marcador bialélico relacionado con *APM1*. Además, los polinucleótidos de la invención adecuados para determinar la identidad de uno o más de los nucleótidos situados en el marcador bialélico relacionado con *APM1*, incluyen polinucleótidos que tienen alguna otra limitación descrita en esta divulgación, o los que vienen a continuación, especificados individualmente o en cualquiera de sus combinaciones. Opcionalmente, dicho marcador bialélico relacionado con *APM1* se puede seleccionar entre el grupo que consiste en A1, A2, A4, A7 y A8, y sus complementos. Opcionalmente, dicho polinucleótido puede comprender una secuencia descrita en la presente memoria descriptiva. Opcionalmente, dicho polinucleótido puede consistir en, o consistir esencialmente en, cualquier polinucleótido descrito en la presente memoria descriptiva. Opcionalmente, dicha determinación se puede llevar a cabo en un ensayo de hibridación, ensayo de microsecuenciación o ensayo de amplificación específica de alelo. Opcionalmente, dicho polinucleótido puede estar unido a un soporte sólido, a una matriz o a una matriz accesible. Opcionalmente, dicho polinucleótido puede estar marcado. Un polinucleótido preferido se puede usar en un ensayo de hibridación para determinar la identidad del nucleótido situado en un marcador bialélico del gen *APM1*. Otro polinucleótido preferido se puede usar en un ensayo de secuenciación o de microsecuenciación para determinar la identidad del nucleótido situado en un marcador bialélico del gen *APM1*. Un tercer polinucleótido preferido se puede usar en un ensayo de amplificación específica de alelo para determinar la identidad del nucleótido situado en un marcador bialélico del gen *APM1*. Un cuarto polinucleótido preferido se puede usar para amplificar un segmento de polinucleótidos que comprenden un marcador bialélico del gen *APM1*. Opcionalmente, cualquiera de los polinucleótidos anteriormente descritos puede estar unido a un soporte sólido, matriz o matriz accesible. Opcionalmente, dicho polinucleótido puede estar marcado.

De manera adicional, la invención incluye el uso de cualquier polinucleótido para, o de cualquier polinucleótido para usar en, la amplificación de un segmento de nucleótidos que comprenden un marcador bialélico relacionado con *APM1*. Además, los polinucleótidos de la invención adecuados para usar en la amplificación de un segmento de nucleótidos que comprende un marcador bialélico relacionado con *APM1*, incluyen polinucleótidos con alguna otra limitación descrita en esta memoria descriptiva, o los que vienen a continuación, especificados individualmente o en cualquiera de sus combinaciones. Opcionalmente dicho marcador bialélico relacionado con *APM1*, se puede seleccionar entre el grupo que consiste en A1, A2, A4, A7 y A8, y sus complementos. Opcionalmente, dicho polinucleótido puede comprender una secuencia descrita en la presente memoria descriptiva. Opcionalmente, dicho polinucleótido puede consistir en, o consistir esencialmente en, cualquiera de los polinucleótidos descritos en la presente memoria descriptiva. Opcionalmente, dicha amplificación se puede realizar mediante PCR o LCR. Opcionalmente, dicho polinucleótido puede estar unido a un soporte sólido, a una matriz o a una matriz accesible. Opcionalmente dicho polinucleótido puede estar marcado.

Los cebadores de una reacción de amplificación o secuenciación de un polinucleótido que comprende un marcador bialélico de la invención, se pueden diseñar a partir de las secuencias descritas, mediante cualquiera de los métodos conocidos en la técnica. Un conjunto de cebadores preferido se diseña de manera que el extremo 3' del tramo contiguo de identidad con una secuencia seleccionada entre el grupo que consiste en los SEQ ID NOS: 1 ó 5, o una de sus secuencias complementarias, o una de sus variantes, se encuentra en el extremo 3' del cebador. Esta configuración permite que el extremo 3' del cebador se hibride con una secuencia de ácido nucleico seleccionada y aumente espectacularmente la eficiencia del cebador de las reacciones de amplificación o secuenciación. Se pueden diseñar cebadores específicos de alelo tales que una base polimórfica de un marcador bialélico está en el extremo 3' del tramo contiguo, y el tramo contiguo está en el extremo 3' del cebador. Esos cebadores específicos de alelo tienden a cebar de manera selectiva una reacción de amplificación o secuenciación siempre que se usen con una muestra de ácido nucleico que contenga uno de los dos alelos presentes en un marcador bialélico. El extremo 3' del cebador de la invención se puede situar a una distancia de, o al menos a, 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500 ó 1.000 nucleótidos aguas arriba de un marcador bialélico de *APM1* en dicha secuencia o en cualquier otra posición que sea adecuada para el uso que de él se requiere para la secuenciación, amplificación o localización de nuevas secuencias o marcadores. Por lo tanto, otro conjunto de cebadores de amplificación preferidos comprende un polinucleótido aislado que consiste esencialmente en un tramo contiguo de 8 a 50 nucleótidos en una secuencia seleccionada entre el grupo que consiste en los SEQ ID NOS: 1 y 5, o una secuencia complementaria a ella, o una de sus variantes, en el que el extremo 3' de dicho tramo contiguo está situado en el extremo 3' de dicho polinucleótido, y en el que el extremo 3' de dicho polinucleótido está situado a aguas arriba de un marcador bialélico del gen *APM1* en dicha secuencia. Preferiblemente, esos cebadores de amplificación comprenden una secuencia seleccionada entre el grupo que consiste en las secuencias de B1 a 15 y de C1 a C15. Los cebadores que tienen sus extremos 3' situados 1 nucleótido aguas arriba de un marcador bialélico de *APM1*, tienen una en ensayos de microsecuenciación. Los cebadores de microsecuenciación preferidos se encuentran descritos en la Tabla 3. Opcionalmente, el marcador bialélico del gen *APM1* se selecciona entre el grupo que consiste en A1, A2 y A7, y sus complementos. Opcionalmente, el marcador bialélico del gen *APM1*, se selecciona entre el grupo que consiste en A4 y A8 y sus complementos. Opcionalmente, los cebadores de microsecuenciación se seleccionan entre el grupo que consiste en las secuencias de nucleótidos de D1 a D8 y de E1 a E8, y sus complementos. Los cebadores de microsecuenciación más preferidos se seleccionan entre el grupo que consiste en las secuencias de nucleótidos D3, E4, E5, E6, D7 y D8.

Las sondas de la presente invención se pueden diseñar a partir de las secuencias descritas mediante cualquiera de los métodos conocidos en la técnica, en particular, mediante métodos que permiten realizar ensayos si está presente un marcador aquí descrito. Se puede diseñar un conjunto de sondas preferidas para usar en ensayos de hibridación del invento de cualquiera de las maneras conocidas en la técnica, de manera que se unen selectivamente a uno de los alelos de un marcador bialélico, pero no al otro alelo, en cualquier conjunto particular de condiciones de ensayo. Las sondas de hibridación preferidas comprenden la base polimórfica o del alelo 1 o del alelo 2 del marcador bialélico considerado. Opcionalmente, dicho marcador bialélico puede estar a una distancia de 6, 5, 4, 3, 2 o 1 nucleótidos del centro de la sonda de hibridación o está situado en el centro de dicha sonda. Se proporcionan sondas ilustrativas en la Tabla 4 de los Ejemplos.

Los polinucleótidos de la presente invención no están limitados a tener las secuencias flanqueantes exactas que rodean rodeando las bases polimórficas que se enumeran en el Listado de secuencias. Las secuencias flanqueantes que rodean los marcadores bialélicos se pueden alargar o acortar hasta una extensión compatible con el uso que se requiere de ellas, y la presente invención contempla específicamente esas secuencias. Las regiones flanqueantes situadas fuera del tramo contiguo no necesitan ser homólogas a las secuencias flanqueantes naturales que se sabe que se encuentran en sujetos humanos. La adición de cualquier secuencia de nucleótidos que sea compatible con el uso que se quiere obtener de los nucleótidos se contempla de manera específica.

Los cebadores y las sondas pueden estar marcados o inmovilizados sobre un soporte sólido según se describe en "Sondas y cebadores oligonucleotídicos".

Los polinucleótidos de la invención que están unidos a un soporte sólido incluyen polinucleótidos con alguna limitación más descrita en esta divulgación, o los que vienen a continuación especificados individualmente o en cualquiera de sus combinaciones. Opcionalmente, dichos polinucleótidos pueden estar especificados como unidos individualmente o unidos en grupos de al menos 2, 5, 8, 10, 12, 15, 20 ó 25 polinucleótidos de la invención diferentes, a un único soporte sólido. Opcionalmente, polinucleótidos distintos a los de la invención se pueden unir al mismo soporte sólido

al que están unidos los polinucleótidos de la invención. Opcionalmente, cuando muchos polinucleótidos se unen a un soporte sólido, se pueden unir a posiciones al azar o en una matriz ordenada. Opcionalmente, dicha matriz ordenada puede ser accesible.

5 La invención también concierne a un método de genotipificación que comprende determinar la identidad de un nucleótido situado en un marcador bialélico del gen *APM1*, seleccionado entre el grupo que consiste en A1, A2 y A7, y sus complementos, en una muestra biológica.

10 La invención también trata sobre un método de genotipificación que comprende determinar la identidad de un nucleótido situado en un marcador bialélico relacionado con *APM1*, preferiblemente un marcador bialélico del gen *APM1*, seleccionado entre el grupo que consiste en A3 y A8, y sus complementos, en una muestra biológica.

15 Opcionalmente, la muestra biológica deriva de un solo sujeto. Opcionalmente, la identidad de los nucleótidos de dicho marcador bialélico se determina en ambas copias de dicho marcador bialélico presente en el genoma de dicho individuo. Opcionalmente, la muestra biológica se obtiene procedente de muchos sujetos. Opcionalmente, el método de genotipificación anteriormente descrito comprende también amplificar una porción de dicha secuencia que comprende el marcador bialélico antes de dicha etapa de determinación, por ejemplo mediante una amplificación por PCR.

20 La etapa de determinación del método de genotipificación anteriormente mencionado se puede llevar a cabo usando un ensayo de hibridación, un ensayo de secuenciación, un ensayo de amplificación específica de alelo o un ensayo de microsecuenciación. Por lo tanto, la invención también incluye métodos para genotipificar una muestra biológica que comprenden determinar la identidad de un nucleótido situado en un marcador bialélico relacionado con *APM1*. Además, los métodos de genotipificación de la invención incluyen métodos con alguna otra limitación descrita en esta divulgación, o los siguientes métodos especificados individualmente o en cualquiera de sus combinaciones. Opcionalmente, dicho marcador bialélico relacionado con *APM1* se puede seleccionar entre el grupo que consiste en A1, A2, A4, A7 y A8, y sus complementos. Opcionalmente, dicha muestra biológica deriva de un solo individuo o sujeto. Opcionalmente, dicho método se realiza *in vitro*. Opcionalmente, dicho marcador bialélico se determina para ambas copias de dicho marcador bialélico presentes en el genoma de dicho individuo. Opcionalmente, dicha muestra biológica deriva de muchos sujetos o individuos. Opcionalmente dicho método comprende también amplificar una porción de dicha secuencia que comprende el marcador bialélico antes de dicha etapa de determinación. Opcionalmente, en el que dicha amplificación se realiza mediante PCR, LCR o replicación de un vector recombinante que comprende un origen de replicación y dicha porción, en una célula hospedadora. Opcionalmente, en el que dicha determinación se realiza mediante un ensayo de hibridación, un ensayo de secuenciación, un ensayo de microsecuenciación o un ensayo de amplificación específica de alelo.

Se describen también kits de diagnóstico que comprenden uno o más de los polinucleótidos de la invención con una porción o con todos los reactivos necesarios y las instrucciones para genotipificar un sujeto en estudio determinando la identidad de un nucleótido situado en un marcador bialélico de *APM1*. Los polinucleótidos de un kit pueden opcionalmente estar unidos a un soporte sólido, o formar parte de una matriz, o de una matriz accesible, de polinucleótidos. El kit puede facilitar la determinación de la identidad del nucleótido situado en una posición del marcador, mediante cualquiera de los métodos conocidos en la técnica que incluyen, pero que no se limitan a, un método de ensayo de secuenciación, un método de ensayo de microsecuenciación, un método de ensayo de hibridación o un método de amplificación específica de alelo. Opcionalmente, un kit de ese tipo puede incluir instrucciones para puntuar los resultados de la determinación con respecto al riesgo que presenta el sujeto en estudio de padecer obesidad o trastornos relacionados con la obesidad.

Métodos para la identificación de novo de marcadores bialélicos

50 Se puede usar cualquiera de los diversos métodos que existen para escrutar un fragmento genómico con respecto a la presencia de polimorfismos de un solo nucleótido, tales como la hibridación diferencial con sondas oligonucleotídicas, la detección de cambios en la movilidad medida mediante electroforesis en gel, o la secuenciación directa del ácido nucleico amplificado. Un método preferido para identificar marcadores bialélicos implica la secuenciación comparativa de fragmentos de DNA genómico procedentes de un número adecuado de individuos no emparentados.

55 En una primera realización, se reúnen muestras de DNA procedentes de individuos no emparentados, después de lo cual el DNA genómico de interés se amplifica y se secuencia. Las secuencias de nucleótidos así obtenidas se analizan después para identificar polimorfismos importantes. Una de las principales ventajas de este método reside en el hecho de que la reunión de las muestras de DNA disminuye sustancialmente el número de reacciones de amplificación de DNA y el de reacciones de secuenciación que se deben llevar a cabo. Además, este método es bastante sensible de manera que un marcador bialélico obtenido con él normalmente demuestra una frecuencia suficiente de su alelo menos común como para ser útil para llevar a cabo estudios de asociación.

65 En una segunda realización las muestras de DNA no se reúnen y por lo tanto se amplifican y se secuencian individualmente. Este método es el normalmente preferido cuando los marcadores bialélicos necesitan ser identificados dentro de los genes candidatos con el fin de llevar a cabo estudios de asociación. Preferiblemente, las regiones de genes altamente relevantes, tales como las regiones de promotor o las regiones exónicas, se pueden escrutar con respecto a la presencia de marcadores bialélicos. Un marcador bialélico obtenido usando este método puede mostrar un

menor grado de aporte de información para llevar a cabo estudios de asociación, p. ej., si la frecuencia de su alelo menos frecuente fuera menor que aproximadamente el 10%. Sin embargo, ese marcador bialélico será suficientemente informativo para llevar a cabo estudios de asociación y se valorará además que la inclusión de marcadores bialélicos menos informativos en los estudios de análisis genéticos de la presente invención, puede permitir en muchos casos la identificación directa de mutaciones causales, que pueden ser, dependiendo de su penetrancia, mutaciones raras.

A continuación viene una descripción de los diferentes parámetros de un método preferido usado por los autores de la invención para la identificación de los marcadores bialélicos de la presente invención.

10 *Muestras de DNA genómico*

Las muestras de DNA genómico a partir de las que se generan los marcadores bialélicos de la presente invención se obtiene preferiblemente a partir de individuos no emparentados que corresponden a una población heterogénea de antecedentes étnicos conocidos. El número de individuos de los que se obtienen muestras de DNA puede variar sustancialmente, preferiblemente desde aproximadamente 10 hasta aproximadamente 1.000, preferiblemente desde aproximadamente 50 hasta aproximadamente 200 individuos. Normalmente se prefiere recoger las muestras de DNA procedentes de al menos aproximadamente 100 individuos con el fin de tener una diversidad polimórfica suficiente en una población determinada para identificar el mayor número de marcadores posibles y para generar resultados estadísticamente significativos.

En lo que se refiere a la fuente del DNA genómico que ha de ser sometido a análisis, cualquier muestra a ensayar se puede prever sin ninguna limitación particular. Estas muestras a ensayar incluyen muestras biológicas, que se pueden ensayar con los métodos de la presente invención aquí descritos, e incluyen fluidos biológicos humanos y de animales tales como sangre completa, suero, plasma, líquido cerebroespinal, orina, fluidos linfáticos y diversas secreciones externas de los aparatos respiratorio, intestinal y genitourinario, lágrimas, saliva, leche, glóbulos blancos, mielomas y muestras similares; fluidos biológicos tales como sobrenadantes de cultivos celulares; especímenes de tejidos fijados que incluyen tejido tumoral y no tumoral y tejidos de los ganglios linfáticos; aspirados de médula ósea y especímenes de células fijadas. La fuente preferida de DNA genómico usada en la presente invención es sangre venosa periférica de cada donante. Las técnicas para preparar DNA genómico a partir de muestras biológicas son muy conocidas por el especialista experto en la técnica. Los detalles de una realización preferida se proporcionan en el Ejemplo 1. El experto en la técnica puede elegir muestras de DNA reunido o no reunido.

Amplificación de DNA

La identificación de marcadores bialélicos en una muestra de DNA genómico se puede facilitar a través del uso de métodos de amplificación de DNA. Las muestras de DNA se pueden reunir o no reunir para la etapa de amplificación. Las técnicas de amplificación son muy conocidas por los expertos en la técnica. Diferentes métodos para amplificar fragmentos de DNA que portan marcadores bialélicos se han descrito con más detalle en lo que antecede en "Amplificación del gen *APM1*". La tecnología de PCR es la técnica de amplificación preferida usada para identificar nuevos marcadores bialélicos. Un ejemplo típico de una reacción de PCR adecuada para los fines de la presente invención se proporciona en el Ejemplo 2.

En una primera realización de la presente invención, los marcadores bialélicos se identifican usando la información de la secuencia genómica generada por los autores de la invención. Se usan fragmentos de DNA genómico secuenciados para diseñar cebadores para la amplificación de fragmentos de 500 pb. Estos fragmentos de 500 pb se amplifican a partir del DNA genómico y son escrutados con respecto a la presencia de marcadores bialélicos. Los cebadores se pueden diseñar usando los programas informáticos OSP (Hillier L. y Green P., 1991). Todos los cebadores pueden contener, situados aguas arriba de las bases diana específicas, una cola oligonucleotídica común que sirve de cebador de secuenciación. Los expertos en la técnica están familiarizados con las extensiones de los cebadores, que se pueden usar para estos fines.

Los cebadores preferidos, útiles para la amplificación de secuencias genómicas que codifican los genes candidatos, se centran en los promotores, exones y sitios de corte y empalme de los genes. Un marcador bialélico presenta una probabilidad más alta de ser una mutación causal si está situado en estas regiones funcionales del gen. Los cebadores de amplificación preferidos de la invención incluyen las secuencias de nucleótidos Núms. de B1 a B15 y las secuencias de nucleótidos Núms. de C1 a C15 descritas en el Ejemplo 2.

Secuenciación del DNA genómico amplificado e identificación de polimorfismos de un solo nucleótido

Los productos resultantes de la amplificación generados como se ha descrito anteriormente, se secuencian después usando alguno de los métodos conocidos y disponibles para el especialista experto. Los métodos de secuenciación de DNA que usan el método mediado por didesoxinucleótidos (método de Sanger) o el método de Maxam-Gilbert son muy conocidos por los expertos con una capacitación normal en la técnica. Esos métodos se encuentran descritos, por ejemplo, Sambrook y col. (1989). Estrategias alternativas incluyen la hibridación a matrices de alta densidad de sondas de DNA según se describe en Chee y col. (1996).

Preferiblemente, el DNA se somete a reacciones de secuenciación automática con terminadores didesoxinucleotídicos usando un protocolo de secuenciación por ciclos con cebador unido a un colorante. Los productos de las reac-

ciones de secuenciación se procesan sobre geles para secuenciación y las secuencias se determinan usando un análisis de imágenes en gel. La búsqueda de polimorfismos se basa en la presencia de picos superpuestos en el patrón de electroforesis, que son el resultado de bases diferentes que se encuentran en la misma posición. Debido a que cada uno de los terminadores didesoxinucleotídicos está marcado con una molécula fluorescente diferente, los dos picos correspondientes a un sitio dialélico presentan distintos colores que corresponden a los dos nucleótidos diferentes situados en la misma posición en la secuencia. Sin embargo, la presencia de dos picos puede ser un artefacto debido a ruido de fondo. Para excluir este artefacto, las dos cadenas de DNA se secuencian y se lleva a cabo una comparación entre los picos. Con el fin de ser registrada como secuencia polimórfica, el polimorfismo se ha de detectar en las dos cadenas.

El procedimiento anterior permite identificar los productos resultantes de la amplificación que contienen marcadores bialélicos. El límite de detección correspondiente a la frecuencia de los polimorfismos bialélicos detectados secuenciando materiales reunidos de 100 individuos, es aproximadamente de 0,1 para el alelo minoritario, según se comprueba secuenciando materiales reunidos de secuencias alélicas conocidas. Sin embargo, más del 90% de los polimorfismos bialélicos detectados con el método que usa materiales reunidos, tienen una frecuencia para el alelo minoritario mayor que 0,25. Por lo tanto, los marcadores bialélicos seleccionados con este método tienen y una frecuencia de al menos 0,1 para el alelo minoritario y menor que 0,9 para el alelo mayoritario. Preferiblemente, de al menos 0,2 para el alelo minoritario y menor que 0,8 para el alelo mayoritario, más preferiblemente, de al menos 0,3 para el alelo minoritario y menor que 0,7 para el alelo mayoritario, y por lo tanto una tasa de heterocigosidad mayor que 0,18, preferiblemente mayor que 0,32 y más preferiblemente mayor que 0,42.

En otra realización, los marcadores bialélicos se detectan secuenciando muestras individuales de DNA, y la frecuencia del alelo minoritario de ese marcador bialélico puede ser menor que 0,1.

Validación de los marcadores bialélicos de la presente invención

Los polimorfismos se evalúan con respecto a su utilidad como marcadores genéticos validando el que los dos alelos estén presentes en la población. La validación de los marcadores bialélicos se lleva a cabo genotipificando un grupo de individuos mediante un método de la invención y demostrando que los dos alelos están presentes. La microsecuenciación es un método preferido para fenotipificar alelos. La validación mediante una etapa de genotipificación se puede llevar a cabo en muestras individuales obtenidas de cada uno de los individuos del grupo, o genotipificando una muestra reunida que se obtiene de más de un solo individuo. El grupo puede ser tan pequeño como de un solo individuo si ese individuo es heterocigoto con respecto al alelo en cuestión. Preferiblemente, el grupo contiene al menos tres individuos, más preferiblemente el grupo contiene cinco o seis individuos, de manera que un único ensayo de validación dará probablemente como resultado la validación de más marcadores bialélicos de los que se están estudiando. Sin embargo, se debe tener en cuenta que cuando el ensayo de validación se lleva a cabo en un grupo pequeño, se puede producir un resultado falso negativo si como resultado del error de muestreo ninguno de los individuos estudiados porta uno de los dos alelos. Por lo tanto, el procedimiento de validación es menos útil para demostrar que un resultado particular inicial es un artefacto, que para demostrar que existe un marcador bialélico *bona fide* (genuino) en una posición particular de la secuencia. Todos los métodos de genotipificación, haplotipificación y de estudios de interacción de la invención se pueden llevar opcionalmente a cabo exclusivamente con marcadores bialélicos validados.

Evaluación de la frecuencia de los marcadores bialélicos de la presente invención

Los marcadores bialélicos validados se evalúan también con respecto a su utilidad como marcadores genéticos determinando la frecuencia del alelo menos común situado en el sitio del marcador bialélico. Cuanto más alta es la frecuencia del alelo menos común, mayor es la utilidad del marcador bialélico en estudios de asociación y en estudios de interacción. La determinación del alelo menos común se lleva a cabo genotipificando un grupo de individuos con un método de la invención y demostrando que los dos alelos están presentes. Esta determinación de la frecuencia mediante una etapa de genotipificación se puede llevar a cabo en muestras individuales procedentes de cada uno de los individuos del grupo, o genotipificando una muestra reunida obtenida de más de un solo individuo. El grupo debe ser lo suficientemente grande para ser representativo de la población en su totalidad. Preferiblemente, el grupo contiene al menos 20 individuos, más preferiblemente el grupo contiene al menos 50 individuos, muy preferiblemente el grupo contiene al menos 100 individuos. Por supuesto que cuanto más grande es el grupo, mayor es la exactitud de la determinación de la frecuencia debido a un menor error de muestreo. Para tener una indicación de la frecuencia del alelo menos común del marcador bialélico particular de la invención, véanse la Tabla A y la Tabla B. Un marcador bialélico en el que la frecuencia del alelo menos común es del 30% o mayor se denomina "marcador bialélico de alta calidad". Todos los métodos de genotipificación, haplotipificación, estudios de asociación y de interacción de la invención opcionalmente pueden realizarse sólo con marcadores bialélicos de alta calidad.

La invención también se refiere a métodos para estimar la frecuencia de un alelo en una población, que comprenden determinar la representación proporcional de un nucleótido situado en un marcador bialélico relacionado con *APM1*, en dicha población. Además, los métodos de la invención para estimar la frecuencia de un alelo en una población incluyen métodos con alguna otra limitación descritos en esta divulgación, o los siguientes métodos, especificados individualmente o en cualquiera de sus combinaciones. Opcionalmente, dicho marcador bialélico relacionado con *APM1* se puede seleccionar entre el grupo que consiste en A1, A2, A3, A4, A5, A6, A7 y A8, y sus complementos. Opcionalmente, la determinación proporcional de un nucleótido en un marcador bialélico relacionado con *APM1* se puede llevar a cabo determinando la identidad de los nucleótidos en las dos copias de dicho marcador bialélico presente en el genoma de cada uno de los individuos de dicha población, y calculando la representación proporcional

de dicho nucleótido en dicho marcador bialélico relacionado con *APM1* con respecto a la población. Opcionalmente, la determinación de la representación proporcional se puede llevar a cabo realizando un método de genotipificación de la invención en una muestra biológica reunida, obtenida de un número representativo de individuos, o en cada uno de los individuos, en dicha población, y calculando la cantidad proporcional de dicho nucleótido en comparación con la cantidad total.

Métodos para genotipificar un individuo con respecto a la presencia de marcadores bialélicos

Se proporcionan métodos para genotipificar una muestra biológica con respecto a uno o más marcadores bialélicos de la presente invención, todos los cuales se pueden realizar *in vitro*. Esos métodos de genotipificación comprenden determinar la identidad de un nucleótido situado en el sitio de un marcador bialélico relacionado con *APM1* mediante cualquiera de los métodos conocidos en la técnica. Estos métodos encuentran aplicación en la genotipificación de poblaciones caso-control en estudios de asociación, así como en la genotipificación de individuos en el contexto de la detección de alelos de marcadores bialélicos que se sabe que están asociados con un rasgo determinado, en cuyo caso se determinan las dos copias del marcador bialélico presente en el genoma del individuo de manera que un individuo se puede clasificar como homocigoto o heterocigoto con respecto a un alelo particular.

Estos métodos de genotipificación se pueden llevar a cabo en muestra de ácido nucleico obtenidas de un solo individuo o procedente de muestras de DNA reunidas.

La genotipificación se puede realizar usando métodos similares a los anteriormente descritos para la identificación de los marcadores bialélicos, o usando otros métodos de genotipificación como los que se describen con más detalle a continuación. En las realizaciones preferidas, la comparación de las secuencias de fragmentos genómicos amplificados procedentes de individuos diferentes, se usa para identificar marcadores bialélicos nuevos mientras que la secuenciación se usa para genotipificar marcadores bialélicos conocidos en aplicaciones para estudios diagnósticos y estudios de asociación.

En una de las realizaciones, la invención incluye métodos de genotipificación que comprenden determinar la identidad de un nucleótido situado en un marcador bialélico relacionado con *APM1*, del SEQ ID NO: 1, o su complemento, en una muestra biológica. Opcionalmente la muestra biológica procede de un solo sujeto. Opcionalmente, la identidad de los nucleótidos situados en dicho marcador bialélico se determina en las dos copias de dicho marcador bialélico presente en el genoma de dicho individuo. Opcionalmente, la muestra biológica procede de muchos sujetos. Opcionalmente, el método comprende además amplificar una porción de dicha secuencia que comprende el marcador bialélico, antes de dicha etapa de determinación. Opcionalmente, la etapa de amplificación se realiza mediante PCR. Opcionalmente, la etapa de determinación se realiza mediante un ensayo de hibridación, un ensayo de secuenciación, un ensayo de microsecuenciación o un ensayo de amplificación específica de alelo.

Fuente de DNA para la genotipificación

Cualquier de ácidos nucleicos, en forma purificada o no purificada, se puede utilizar como ácido nucleico de partida con la condición de que contenga, o se sospeche que contenga, la secuencia específica de ácido nucleico deseada. El DNA o el RNA se puede extraer de células, tejidos, fluidos corporales y fuentes similares como las descritas en lo que antecede. Aunque los ácidos nucleicos adecuados para usar en los métodos de genotipificación de la invención pueden tener su origen en cualquier fuente originaria de mamíferos, los sujetos e individuos del estudio de los que se recogen las muestras de ácidos nucleicos, se entiende en general que son seres humanos.

Amplificación de fragmentos de DNA que comprenden marcadores bialélicos

Se proporcionan métodos y polinucleótidos para amplificar un segmento de nucleótidos que comprende uno o más de los marcadores bialélicos de la presente invención. Se comprenderá que la amplificación de fragmentos de DNA que comprenden marcadores bialélicos se puede usar en distintos métodos y para distintos fines, y no está restringida a la genotipificación. No obstante, muchos de los métodos de genotipificación, aunque no todos, requieren la amplificación previa de la región de DNA que porta el marcador bialélico de interés. Esos métodos aumentan específicamente la concentración del número total de secuencias que abarcan el marcador bialélico o que incluyen el sitio y las secuencias situados en posición distal o proximal respecto a él. Los ensayos diagnósticos se pueden basar también en la amplificación de segmentos de DNA que portan un marcador bialélico de la presente invención. La amplificación del DNA se puede realizar mediante cualquiera de los métodos conocidos. Técnicas de amplificación se han descrito en lo que antecede en la sección titulada, amplificación del gen *APM1*.

Algunos de estos métodos de amplificación son particularmente adecuados para la detección de polimorfismos de un solo nucleótido y permiten la amplificación simultánea de una secuencia diana y la identificación del nucleótido polimórfico como se describirá con detalle más adelante.

La identificación de marcadores bialélicos según se ha descrito anteriormente, permite el diseño de oligonucleótidos adecuados, que se pueden usar como cebadores para amplificar fragmentos de DNA que comprenden los marcadores bialélicos de la presente invención. La amplificación se puede realizar usando los cebadores inicialmente usados para descubrir los nuevos marcadores bialélicos que aquí se describen, o cualquier conjunto de cebadores que permiten la amplificación de un DNA que comprende un marcador bialélico de la presente invención.

En algunas realizaciones la presente invención proporciona cebadores para amplificar un fragmento de DNA que contiene uno o más de los marcadores bialélicos de la presente invención. Los cebadores de amplificación preferidos se listan en el Ejemplo 2. Se valorará que tendrá en cuenta que los cebadores listados son meramente ilustrativos y que se puede usar cualquier otro conjunto de cebadores que produce productos de amplificación que contienen uno o más de los marcadores bialélicos de la presente invención.

El espaciado entre los cebadores determina la longitud del segmento que se ha de amplificar. En el contexto de la presente invención, los segmentos amplificados que portan marcadores bialélicos, pueden variar de tamaño desde al menos aproximadamente 25 pb hasta 35 kpb. Los fragmentos resultantes de la amplificación de 25 pb-3.000 pb son típicos. Se prefieren los fragmentos de 50 pb a 1.000 pb, y los fragmentos de 100 pb-600 pb son muy preferidos. Se tendrá en cuenta que los cebadores de amplificación correspondientes a los marcadores bialélicos pueden ser cualquier secuencia que permite la amplificación específica de un fragmento de DNA que porta los marcadores. Los cebadores de amplificación pueden estar marcados o pueden estar inmovilizados en un soporte sólido, según se describe en "Sondas y cebadores oligonucleotídicos".

15 *Métodos para genotipificar muestras de DNA con respecto a la presencia de marcadores bialélicos*

Se puede usar cualquiera de los métodos conocidos en la técnica para identificar el nucleótido presente en el sitio de un marcador bialélico. Debido a que el alelo del marcador bialélico que ha de ser detectado ha sido identificado y especificado en la presente invención, la detección se hará fácil para una persona con una experiencia normal en la técnica, utilizando cualquiera de las diversas técnicas. Muchos métodos de genotipificación requieren la amplificación previa de la región de DNA que porta el marcador bialélico de interés. Aunque la amplificación de la diana o de la señal se suele preferir en la actualidad, métodos de detección ultrasensibles que no requieren amplificación están también incluidos en los presentes métodos de genotipificación. También se pueden usar métodos muy conocidos por los expertos en la técnica para detectar polimorfismos bialélicos que incluyen métodos tales como, análisis convencionales de transferencia de puntos, análisis de polimorfismos conformacionales de una sola cadena (SSCP) (del inglés, "*Single Strand conformational polymorphism analysis*") (SSCP) descrito por Orita y col. (1989), electroforesis en gel en gradiente desnaturalizante (DGGE), análisis de los heterodúplex, detección de desapareamientos mediante escisión, y otras técnicas convencionales según se describen en Sheffield y col. (1991), White y col. (1992), Grompe y col. (1989 y 1993). Otro método para determinar la identidad del nucleótido presente en un sitio polimórfico particular utiliza un derivado de nucleótido especializado, resistente a exonucleasa, según se describe en la Patente de EE. UU. 4.656.127.

Los métodos preferidos implican determinar directamente la identidad del nucleótido presente en el sitio de un marcador bialélico mediante un ensayo de secuenciación, un ensayo de amplificación específica de alelo, o un ensayo de hibridación. A continuación sigue una descripción de algunos métodos preferidos. Un método muy preferido es la técnica de microsecuenciación. El término "secuenciación" se usa aquí para hacer referencia a la extensión de los complejos dúplex cebador/molde por acción de la polimerasa e incluye tanto la secuenciación tradicional como la microsecuenciación.

40 1) *Ensayos de secuenciación*

El nucleótido presente en un sitio polimórfico se puede determinar mediante métodos de secuenciación. En una realización preferida, las muestras de DNA se someten a una amplificación por PCR antes de la secuenciación según se ha descrito anteriormente. Los métodos de secuenciación de DNA se describen en el apartado "Secuenciación del DNA genómico amplificado e identificación de polimorfismos de un solo nucleótido".

Preferiblemente, el DNA amplificado se somete a reacciones de secuenciación automática con terminadores dideoxinucleotídicos, usando un protocolo de secuenciación por ciclos con un cebador unido a un colorante. El análisis de la secuencia permite la identificación de la base presente en el sitio del marcador bialélico.

50 2) *Ensayos de microsecuenciación*

En los métodos de microsecuenciación, el nucleótido situado en un sitio polimórfico de un DNA diana se detecta mediante una reacción de extensión de un solo nucleótido con cebador. Este método implica el uso de cebadores de microsecuenciación adecuados que se hibridan justamente aguas arriba de la base polimórfica de interés en el ácido nucleico diana. Se usa una polimerasa para extender de manera específica el extremo 3' del cebador con un único ddNTP (terminador de cadena) complementario al nucleótido situado en el sitio polimórfico. Después se determina la identidad del nucleótido incorporado de alguna manera que convenga.

Típicamente, las reacciones de microsecuenciación se llevan a cabo usando ddNTP fluorescentes y los cebadores de microsecuenciación extendidos se analizan mediante electroforesis en máquinas de secuenciación ABI 377 para determinar la identidad del nucleótido incorporado según se describe en el documento EP 412 883. Como alternativa, se puede usar la electroforesis capilar a fin de procesar simultáneamente un número más alto de ensayos. Un ejemplo de un procedimiento de microsecuenciación típico que se puede usar en el contexto de la presente invención, se proporciona en el Ejemplo 4.

Se pueden usar estrategias diferentes para el marcaje y detección de los ddNTP. Un método de detección en fase homogénea, basado en la transferencia de energía por resonancia de fluorescencia ha sido descrito por Chen y Kwok (1997) y por Chen y col. (1997). En este método, fragmentos de DNA genómico amplificados, que contienen sitios polimórficos, se incuban con un cebador marcado con fluoresceína en 5' en presencia de los trifosfatos de didesosxirribonucleósidos marcados con un colorante alélico y con una Taq polimerasa. El cebador marcado con colorante se extiende sólo una base con el terminador unido a colorante específico para el alelo presente en el molde. Al final de la reacción de genotipificación, las intensidades de fluorescencia de los dos colorantes presentes en la mezcla de reacción se analizan directamente sin separar ni purificar. Todas estas etapas se pueden realizar en el mismo tubo y los cambios en la fluorescencia se pueden detectar en tiempo real. Como alternativa, el cebador extendido se puede analizar mediante espectrometría de masas MALDI-TOF. La base situada en el sitio polimórfico se identifica por la masa añadida en el cebador de la microsecuenciación (véase Haffand Smimov, 1997).

La microsecuenciación se puede llevar a cabo mediante el método de microsecuenciación establecido o mediante evoluciones o métodos derivados del mismo. Métodos alternativos incluyen diversas técnicas de microsecuenciación en fase sólida. El protocolo básico de microsecuenciación es el mismo que el anteriormente descrito excepto que el método se lleva a cabo en forma de ensayo en fase heterogénea, en el que o el cebador o la molécula diana está inmovilizado o capturado sobre un soporte sólido. Para simplificar la separación del cebador y el análisis de adición del nucleótido terminal, los oligonucleótidos se unen a los soportes sólidos o se modifican de maneras que permitan una separación por afinidad así como la extensión por acción de la polimerasa. Los extremos 5' y los nucleótidos internos de los oligonucleótidos sintéticos se pueden modificar de varios modos diferentes para permitir estrategias diferentes de separación por afinidad, p. ej., biotinilación. Si se usa un único grupo de afinidad con los oligonucleótidos, los oligonucleótidos se pueden separar del reactivo terminador incorporado. Esto elimina la necesidad de una separación física o por tamaños. Se puede separar más de un solo oligonucleótido del reactivo terminador y se puede analizar simultáneamente si se usa más de un único grupo de afinidad. Esto permite el análisis de varias especies de ácidos nucleicos o de más información sobre secuencias de ácido nucleico por cada una de las reacciones de extensión. El grupo de afinidad necesita estar en el oligonucleótido cebador pero como alternativa podría estar en el molde. Por ejemplo, la inmovilización se puede llevar a cabo a través de una interacción entre DNA biotinilado y pocillos de microtitulación revestidos con estreptavidina, o partículas de poliestireno revestidas con avidina. De la misma manera, los oligonucleótidos o moldes se pueden unir a un soporte sólido en un formato de alta densidad. En esas reacciones de microsecuenciación en fase sólida, los ddNTP incorporados pueden estar marcados radiactivamente (Syvänen, 1994) o pueden estar unidos a fluoresceína (Livak y Hainer, 1994). La detección de los ddNTP marcados radiactivamente se puede realizar a través de técnicas basadas en centelleo. La detección de los ddNTP unidos a fluoresceína se puede basar en la unión de un anticuerpo anti-fluoresceína conjugado con fosfatasa alcalina, seguida de una incubación con un sustrato cromogénico (tal como el fosfato de p-nitrofenilo). Otras parejas posibles de molécula de reporte-molécula de detección incluyen: ddNTP unidos a dinitrofenilo (DNP) (del inglés, "*DiNitroPhenyle*") y el conjugado de anticuerpo anti-DNP y fosfatasa alcalina (Harju y col., 1993) o ddNTP biotinilado y estreptavidina conjugada con peroxidasa de rábano rusticano, con o-fenilendiamina como sustrato (documento WO 92/15712). Como otro procedimiento alternativo más de microsecuenciación en fase sólida, Nyren y col. (1993) han descrito un método que se basa en la detección de la DNA-polimerasa mediante un ensayo enzimático y luminométrico de detección de pirofosfato inorgánico (ELIDA) (del inglés, "*Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay*").

Pastinen y col. (1997) describen de detección multiplex de polimorfismos de un solo nucleótido en el que el principio de la microsecuenciación en fase sólida se aplica a un formato de matrices de oligonucleótidos. Matrices de alta densidad de sondas de DNA unidas a un soporte sólido (chips de DNA) se describen con detalle más adelante.

En uno de los aspectos la presente invención proporciona polinucleótidos y métodos para genotipificar uno o más de los marcadores bialélicos de la presente invención realizando un ensayo de microsecuenciación. Los cebadores de microsecuenciación preferidos incluyen las secuencias de nucleótidos Núms. de D1 a D8 y de E1 a E8. Los cebadores de microsecuenciación más preferidos se seleccionan entre el grupo que consiste en las secuencias de nucleótidos Núms. D3, E4, E5, E6, D7 y D8. Se percibirá que los cebadores de microsecuenciación listados en el Ejemplo 4 son meramente ilustrativos y que se puede usar cualquier cebador que tenga un extremo 3' inmediatamente adyacente al nucleótido polimórfico. De manera similar, se percibirá que el análisis de microsecuenciación se puede realizar para cualquier marcador bialélico o para cualquier combinación de los marcadores bialélicos de la presente invención. Uno de los aspectos de la presente invención es un soporte sólido que incluye uno o más de los cebadores de microsecuenciación listados en el Ejemplo 4, o sus fragmentos que comprenden al menos 8, 12, 15, 20, 25, 30, 40 ó 50 nucleótidos consecutivos y que tienen una terminación 3' inmediatamente aguas arriba del correspondiente marcador bialélico, para determinar la identidad de un nucleótido en el sitio de un marcador bialélico.

3) Métodos de ensayos de amplificación específica de alelo

En uno de los aspectos la presente invención proporciona polinucleótidos y métodos para determinar el alelo de uno o más de los marcadores bialélicos de la presente invención, en una muestra biológica, mediante ensayos de amplificación específica de alelo. Los métodos, los cebadores y los diferentes parámetros para amplificar fragmentos de DNA que comprenden marcadores bialélicos de la presente invención, se han descrito con detalle en lo que antecede en el apartado "Amplificación de fragmentos de DNA que comprenden marcadores bialélicos".

Cebadores de la amplificación específica de alelo

La discriminación entre los dos alelos de un marcador bialélico se puede realizar también mediante amplificación específica de alelo, estrategia selectiva en la que uno de los alelos se amplifica sin que se amplifique el otro alelo. Esto se lleva a cabo colocando la base polimórfica en el extremo 3' de uno de los cebadores de la amplificación. Debido a que la extensión se crea desde el extremo 3' del cebador, un desapareamiento en, o cerca de, esta posición tiene un efecto inhibitorio sobre la amplificación. Por lo tanto, en las condiciones de amplificación adecuadas, estos cebadores sólo dirigen la amplificación en el alelo complementario a ellos. La determinación de la situación precisa del desapareamiento y de las correspondientes condiciones del ensayo está dentro de la experiencia normal en la técnica.

Métodos basados en ligación/amplificación

El ensayo de ligación de oligonucleótidos (OLA) (del inglés, "*Oligonucleotide Ligation Assay*") usa dos oligonucleótidos que se diseñan para que sean capaces de hibridarse con secuencias contiguas de una sola de las cadenas de una molécula diana. Uno de los oligonucleótidos está biotinilado y el otro está marcado de manera detectable. Si la secuencia complementaria exacta se encuentra en una molécula diana, los oligonucleótidos se hibridarán de manera que sus terminaciones se empalman y crean un sustrato de ligación que se puede capturar y detectar. El ensayo OLA es capaz de detectar polimorfismos de un solo nucleótido y se puede combinar de manera provechosa con la PCR según se describe en Nickerson y col. (1990). En este método, la PCR se usa para realizar la amplificación exponencial del DNA diana, que es después detectado usando el ensayo OLA.

Otros métodos de amplificación que son particularmente adecuados para la detección de polimorfismos de un solo nucleótido incluyen la LCR (reacción en cadena de la ligasa) y la LCR basada en la existencia de huecos (GLCR) (del inglés, "*Gap LCR*"), que se han descrito en lo que antecede en el apartado "Amplificación del gen *APMI*". La LCR usa dos pares de sondas para amplificar de manera exponencial una secuencia diana. Las secuencias de cada pareja de oligonucleótidos se seleccionan para permitir que la pareja se hibride con secuencias contiguas de la misma cadena de la diana. Esa hibridación forma un sustrato para la ligasa dependiente de molde. De acuerdo con la presente invención, la LCR se puede realizar con oligonucleótidos que contienen las secuencias proximal y distal de la misma cadena de un sitio de marcador bialélico. En una de las realizaciones, ambos oligonucleótidos se diseñarán de manera que incluyan el sitio del marcador bialélico. En esa realización, las condiciones de reacción se seleccionan de manera que los oligonucleótidos se pueden ligar entre sí solamente si la molécula diana contiene, o carece de, el nucleótido específico que es complementario al marcador bialélico en el oligonucleótido. En una realización alternativa, los oligonucleótidos no incluirán el marcador bialélico, de manera que cuando se hibridan a la molécula diana, se crea un "hueco" según se describe en el documento WO 90/01069. Este hueco se "rellena" luego con los dNTP complementarios (de manera mediada por la DNA-polimerasa) o por medio de un par adicional de oligonucleótidos. De esta manera al final de cada ciclo, cada una de las cadenas sencillas tiene un complemento capaz de servir de diana durante el ciclo siguiente y se obtiene una amplificación exponencial específica de alelo, de la secuencia deseada.

El análisis *Ligase/Polymerase-mediated Genetic Bit Analysis*TM es otro de los métodos para determinar la identidad de un nucleótido situado en un sitio preseleccionado en una molécula de ácido nucleico (documento WO 95/21271). Este método implica la incorporación de un nucleósido trifosfato que es complementario al nucleótido presente en el sitio preseleccionado, en la terminación de una molécula de cebador, y su posterior ligación a un segundo oligonucleótido. La reacción se controla detectando un marcador específico que está unido a la fase sólida de la reacción, o mediante una detección en la disolución.

4) Métodos de ensayos de hibridación

Un método preferido para determinar la identidad del nucleótido presente en un sitio de marcador bialélico implica una hibridación de ácidos nucleicos. Las sondas de hibridación, que se pueden usar convenientemente en esas reacciones, incluyen preferiblemente las sondas aquí definidas. Se puede usar cualquiera de los ensayos de hibridación que incluyen la hibridación Southern, hibridación Northern, hibridación puntual e hibridación en fase sólida (véase Sambrook y col., 1989).

Hibridación se refiere a la formación de una estructura dúplex producida por dos ácidos nucleicos monocatenarios, debido al apareamiento de bases complementarias. La hibridación puede tener lugar entre cadenas de ácidos nucleicos exactamente complementarias o entre cadenas de ácido nucleico que contienen regiones minoritarias de desapareamiento. Se pueden diseñar sondas específicas que se hibridan a una de las formas de un marcador bialélico y no a la otra y que por lo tanto son capaces de discriminar formas alélicas diferentes. Las sondas específicas de alelo se suelen utilizar por parejas, mostrando uno de los miembros de la pareja un apareamiento perfecto a una secuencia diana que contiene el alelo original, y mostrando el otro miembro de la pareja un apareamiento perfecto con la secuencia diana que contiene el alelo alternativo. Las condiciones de la hibridación deben ser lo suficientemente rigurosas para que exista una diferencia significativa en la intensidad de hibridación entre los alelos, y preferiblemente una respuesta esencialmente binaria mediante la cual una sonda se hibrida con sólo uno de los alelos. Condiciones de hibridación rigurosas y específicas de secuencia, bajo las cuales una sonda se hibridará solamente con la secuencia diana exactamente complementaria son muy conocidas en la técnica (Sambrook y col., 1989). Las condiciones rigurosas son dependientes de la secuencia y serán diferentes en circunstancias diferentes. Generalmente, las condiciones rigurosas se seleccionan para que sean de aproximadamente 5°C por debajo del punto de fusión térmica (T_m) (del

inglés, “*Thermal melting point*”) de la secuencia específica, a una fuerza iónica y a un pH definidos. Aunque esas hibridaciones se pueden llevar a cabo en una disolución, se prefiere utilizar un ensayo de hibridación en fase sólida. El DNA diana que comprende un marcador bialélico de la presente invención se puede amplificar antes de la reacción de hibridación. La presencia de un alelo específico en la muestra se determina detectando la presencia o la ausencia de

5 dúplex híbridos estables, formados entre la sonda y el DNA diana. La detección de los dúplex híbridos se puede llevar a cabo mediante diversos métodos. Los diferentes formatos de ensayos de detección son muy conocidos y utilizan marcadores detectables unidos o a la diana o a la sonda para permitir la detección de los dúplex híbridos. Típicamente, los dúplex resultantes de la hibridación se separan de los ácidos nucleicos no hibridados, y los marcadores unidos a los dúplex se detectan después. Los expertos en la técnica reconocerán que se pueden usar etapas de lavado para

10 arrastrar el exceso del DNA diana o de sonda así como el conjugado no unido. Además, los formatos estándar de ensayos heterogéneos son adecuados para detectar los híbridos usando los marcadores presentes en los cebadores y sondas.

Dos ensayos recientemente desarrollados permiten una discriminación de alelos basada en hibridación, sin necesidad de separaciones o de lavados (véase Landegren U. y col., 1998). El ensayo TaqMan aprovecha la actividad de 5'-nucleasa de la Taq DNA-polimerasa para digerir una sonda de DNA renaturalizada específicamente con producto de amplificación que se acumula. Las sondas TaqMan están marcadas con una pareja de colorantes dador y aceptor que interacciona a través de la transferencia de energía de fluorescencia. La escisión de la sonda TaqMan por el avance de la polimerasa durante la amplificación disocia el colorante dador del colorante aceptor de extinción, aumentando

20 enormemente la fluorescencia del dador. Todos los reactivos necesarios para detectar dos variantes alélicas se pueden reunir al principio de la reacción y los resultados se detectan en tiempo real (véase Livak y col., 1995). En un procedimiento alternativo basado en una hibridación homogénea, se usan balizas moleculares para realizar las discriminaciones entre alelos. Las balizas moleculares son sondas oligonucleotídicas con forma de horquilla que informan de la presencia de ácidos nucleicos específicos en disoluciones homogéneas. Cuando se unen a sus dianas experimentan una reorganización conformacional que restituye la fluorescencia de un fluoróforo de extinción interna (Tyagi y col., 1998).

Los polinucleótidos que aquí se proporcionan se pueden usar para producir sondas que se pueden usar en ensayos de hibridación para la detección de alelos de marcadores bialélicos en muestras biológicas. Estas sondas se caracterizan porque comprenden preferiblemente entre 8 y 50 nucleótidos y porque son lo suficientemente complementarias a una secuencia que comprende un marcador bialélico de la presente invención como para hibridarse con ella, y preferiblemente, lo suficientemente específicas como para ser capaces de discriminar la secuencia diana por solamente la variación de un nucleótido. Una sonda particularmente preferida tiene una longitud de 25 nucleótidos. Preferiblemente el marcador bialélico está a una distancia de 4 nucleótidos del centro de la sonda polinucleotídica.

35 En las sondas particularmente preferidas, el marcador bialélico está en el centro de dicho polinucleótido. Las sondas preferidas comprenden una secuencia de nucleótidos seleccionada entre el grupo que consiste en los Núms. 9-27, 99-14.387, 9-12, 9-13, 99-14.405 y 9-16, y las secuencias complementarias a ella, o uno de sus fragmentos, comprendiendo dicho fragmento al menos 8 nucleótidos consecutivos, preferiblemente 10, 15, 20 nucleótidos consecutivos, más preferiblemente 25, 30, 40, 47 ó 50 nucleótidos consecutivos, y conteniendo una base polimórfica. En las realizaciones

40 preferidas la base polimórfica está a una distancia de 5, 4, 3, 2, 1 nucleótidos del centro de dicho polinucleótido, más preferiblemente está en el centro de dicho polinucleótido.

Preferiblemente, las sondas de la presente invención están marcadas o inmovilizadas sobre un soporte sólido. Los marcadores y los soportes sólidos se describen con más detalle en el apartado “Sondas y cebadores oligonucleotídicos”.

45 Las sondas pueden ser no extensibles según se describe en el apartado “Sondas y cebadores oligonucleotídicos”.

Realizando ensayos con respecto a la hibridación a una sonda específica de alelo, se puede detectar la presencia o ausencia de un alelo del marcador bialélico presente en una muestra determinada. Las hibridaciones paralelas de gran capacidad de procesamiento, en formato de matriz, están específicamente incluidas en el apartado “Ensayos de hibridación” y se describen más adelante.

50

5) Hibridación a matrices accesibles de oligonucleótidos

Los ensayos de hibridación basados en matrices de oligonucleótidos se fundamentan en las diferencias en la estabilidad de la hibridación de oligonucleótidos cortos a variantes de la secuencia diana perfectamente emparejadas y a variantes de la secuencia diana desapareadas. Un acceso eficiente a la información sobre polimorfismos se obtiene a través de una estructura básica que comprende matrices de alta densidad de sondas de oligonucleotídicas, unidas a un soporte sólido (p. ej., el chip) en posiciones seleccionadas. Cada chip de DNA puede contener de miles a millones de sondas individuales de DNA sintético, ordenadas en un patrón parecido a una rejilla y miniaturizadas hasta el tamaño de una moneda de 10 centavos.

60

La tecnología basada en los chips ya se ha aplicado con éxito en numerosos casos. Por ejemplo, se ha llevado a cabo el escrutinio de mutaciones en el gen BRCA1, en cepas mutantes de *S. cerevisiae*, y en el gen de la proteasa del virus HIV-1 (Hacia y col., 1996; Shoemaker y col., 1996; Kozal y col., 1996). Affymetrix (*GeneChip™*), Hyseq (*HyChip* y *HyGnostics*) y Laboratorios *Protogene* producen de manera personalizada a las necesidades del cliente chips de diferentes formatos para usar en la detección de polimorfismos bialélicos.

65

En general estos métodos utilizan matrices de sondas oligonucleotídicas que son complementarias a segmentos de secuencias de ácido nucleico diana procedentes de un individuo, secuencias diana que incluyen un marcador polimórfico. El documento EP 785280 describe una estrategia basada en el embaldosado para la detección de polimorfismos de un solo nucleótido. Brevemente expuesto, las matrices generalmente pueden contener “baldosas” correspondientes a un gran número de polimorfismos específicos. Por “embaldosado” generalmente se entiende la síntesis de un conjunto definido de sondas oligonucleotídicas, que se compone de una secuencia complementaria a la secuencia diana de interés, así como también de variaciones preseleccionadas de esta secuencia, p. ej., sustitución de una o más posiciones determinadas con uno o más miembros del conjunto de bases de los monómeros, es decir, de los nucleótidos. Las estrategias de embaldosado se encuentran descritas con más detalle en la solicitud de PCT, documento WO 95/11995.

En un aspecto particular, las matrices se embaldosan con respecto a un número de secuencias de marcadores bialélicos específicos e identificados. En particular, la matriz se embaldosa para que incluya varios bloques de detección, siendo cada bloque de detección específico para un marcador bialélico específico, o para un conjunto de marcadores bialélicos. Por ejemplo, un bloque de detección se puede embaldosar de manera que incluya varias sondas, que abarcan el segmento de secuencia que incluye un polimorfismo específico. Para tener la seguridad de que las sondas son complementarias a cada uno de los alelos, las sondas se sintetizan por parejas que difieren en el marcador bialélico. Además de las sondas que difieren en la base polimórfica, generalmente también se embaldosan sondas monosustituidas dentro del bloque de detección. Estas sondas monosustituidas contienen bases en, y hasta, un determinado número de bases en cada uno de los sentidos desde la posición del polimorfismo, sustituidas con los nucleótidos restantes (seleccionados entre A, T, G, C y U). Típicamente, las sondas de un bloque de detección embaldosado incluirán sustituciones de las posiciones de la secuencia hasta, e incluyendo, las que están a una distancia de 5 bases del marcador bialélico. Las sondas monosustituidas proporcionan controles internos para la matriz embaldosada, para distinguir una hibridación real de hibridaciones cruzadas de artefacto. Una vez completada la hibridación con la secuencia diana y los lavados de la matriz, la matriz se explora para determinar la posición sobre la matriz a la que la secuencia diana se hibrida. Los datos de las hibridaciones obtenidos de la matriz explorada se analizan luego para identificar qué alelo o alelos del marcador bialélico están presentes en la muestra. La hibridación y la exploración se pueden llevar a cabo según se describe en la solicitud de PCT documento WO 92/10092 y documento WO 95/11995 y en la Patente de EE. UU. Núm. 5.424.186.

Así, en algunas realizaciones, los chips pueden comprender una matriz de secuencias de ácido nucleico, de fragmentos de aproximadamente 15 nucleótidos de longitud. En otras realizaciones, el chip puede comprender una matriz que incluye al menos una de las secuencias seleccionadas entre el grupo que consiste en 9-27, 99-14.387, 9-12, 9-13, 99-14.405 y 9-16, y las secuencias complementarias ella, o uno de sus fragmentos, comprendiendo dicho fragmento al menos 8 nucleótidos consecutivos, preferiblemente 10, 15, 20 nucleótidos consecutivos, más preferiblemente 25, 30, 40, 47 ó 50 nucleótidos consecutivos, y conteniendo dicho fragmento una base polimórfica. En las realizaciones preferidas, la base polimórfica está a una distancia de 5, 4, 3, 2, 1 nucleótidos del centro de dicho polinucleótido, más preferiblemente está situada en el centro de dicho polinucleótido. En algunas realizaciones, el chip puede comprender una matriz de al menos 2, 3, 4, 5, 6, 7, 8 o más de estos polinucleótidos de la invención. Los soportes sólidos y polinucleótidos de la presente invención unidos a soportes sólidos, se describen con más detalle en el apartado “Sondas y cebadores oligonucleotídicos”.

6) Sistemas integrados

Otra técnica, que se puede usar para analizar polimorfismos, incluye sistemas integrados multicomponente que miniaturizan y compartimentan procedimientos tales como las reacciones de PCR y la electroforesis capilar en un único dispositivo funcional. Un ejemplo de esta técnica se describe en la Patente de EE. UU. 5.589.136, que describe la integración de una amplificación por PCR y una electroforesis capilar en chips.

Los sistemas integrados se pueden contemplar principalmente cuando se usan sistemas microfluídicos. Estos sistemas comprenden un patrón de microcanales diseñados sobre una oblea de vidrio, silicio, cuarzo o plástico, incluida en un microchip. Los movimientos de las muestras se controlan mediante fuerzas eléctricas, electroosmóticas o hidrostáticas aplicadas a través de las diferentes áreas del microchip para crear válvulas y bombas microscópicas funcionales con partes no móviles. Variando el voltaje se controla el flujo de líquido en las intersecciones entre los canales micro-mecanizados y se cambia el caudal del líquido para bombearlo a través de las diferentes secciones del microchip.

Para genotipificar marcadores bialélicos, el sistema microfluídico puede integrar la amplificación, microsecuenciación y electroforesis capilar de ácidos nucleicos, y un método de detección tal como la detección por fluorescencia inducida con láser. En una primera etapa, las muestras de DNA se amplifican, preferiblemente mediante PCR. Seguidamente, los productos de la amplificación se someten a reacciones automáticas de microsecuenciación usando los ddNTP (con una fluorescencia específica para cada uno de los ddNTP) y los cebadores oligonucleotídicos de microsecuenciación adecuados, que se hibridan justo aguas arriba de la base polimórfica diana. Una vez completada la extensión del extremo 3', los cebadores se separan de los ddNTP fluorescentes no incorporados mediante electroforesis capilar. El medio de separación usado en la electroforesis capilar puede ser, por ejemplo, poliacrilamida, polietilenglicol o dextrano. Los ddNTP incorporados en los productos resultantes de la extensión del cebador en un solo nucleótido, se identifican mediante una detección por fluorescencia. Este microchip se puede usar para procesar desde al menos 96 hasta 384 muestras en paralelo. Puede usar la detección usual de los ddNTP mediante fluorescencia inducida con un láser de cuatro colores.

Métodos de análisis genético que usan los marcadores bialélicos de la presente invención

Existen disponibles diferentes métodos de análisis genético de rasgos complejos (véase Lander y Schork, 1994). La búsqueda de genes de susceptibilidad a padecer enfermedades se lleva a cabo usando dos métodos principales: la estrategia basada en el ligamiento, en la que se busca evidencia de una cosegregación entre un locus y un locus de un rasgo putativo usando estudios genealógicos, y la estrategia basada en la asociación en la que se busca evidencia de una asociación estadísticamente significativa entre un alelo y un rasgo o un alelo causante de un rasgo (Khoury y col., 1993). En general, los marcadores bialélicos de la presente invención tienen aplicación en cualquier método conocido en la técnica que demuestra una correlación estadísticamente significativa entre un genotipo y un fenotipo. Los marcadores bialélicos se pueden usar en métodos paramétricos y no paramétricos de análisis de ligamiento. Preferiblemente, los marcadores bialélicos de la presente invención se usan para identificar genes asociados con rasgos detectables, usando estudios de asociación, estrategia que no requiere el uso de las familias afectadas y que permite la identificación de genes asociados con rasgos complejos y esporádicos.

Los análisis genéticos que usan los marcadores bialélicos de la presente invención se pueden llevar a cabo a cualquier escala. Se puede usar el conjunto completo de marcadores bialélicos de la presente invención o cualquier subconjunto de los marcadores bialélicos de la presente invención que se correspondan con el gen candidato. Además, se puede usar cualquier conjunto de marcadores genéticos que incluya un marcador bialélico de la presente invención. Un conjunto de polimorfismos bialélicos que se podrían usar como marcadores genéticos junto con los marcadores bialélicos de la presente invención se ha descrito en el documento WO 98/20165. Según se ha mencionado en lo que antecede, se debe destacar que los marcadores bialélicos de la presente invención pueden estar incluidos en cualquier mapa genético completo o parcial del genoma humano. Estos diferentes usos se contemplan específicamente en la presente invención y en las reivindicaciones.

La invención también comprende métodos para detectar una asociación entre un genotipo y un fenotipo, que comprende las etapas de: a) genotipificar al menos un marcador bialélico relacionado con *APM1* en una población positiva con respecto a un rasgo, conforme a un método de genotipificación de la invención; b) genotipificar dicho marcador bialélico relacionado con *APM1* en una población de control, conforme a un método de genotipificación de la invención; y c) determinar si existe una asociación estadísticamente significativa entre dicho genotipo y dicho fenotipo. Además, los métodos para detectar una asociación entre un genotipo y un fenotipo de la invención incluyen métodos con alguna otra limitación descrita en esta divulgación, o los métodos que siguen a continuación, especificados individualmente o en cualquiera de sus combinaciones. Opcionalmente, dicho marcador bialélico relacionado con *APM1* se puede seleccionar entre el grupo que consiste en A1, A2, A3, A4, A5, A6, A7 y A8, y sus complementos. Opcionalmente, dicha población de control puede ser una población negativa con respecto a un rasgo, o una población al azar. Opcionalmente, cada una de dichas etapas a) y b) de genotipificación se pueden llevar a cabo sobre una muestra biológica reunida obtenida de cada una de dichas poblaciones. Opcionalmente, dicho genotipificación de las etapas a) y b) se realiza por separado sobre muestras biológicas obtenidas de cada uno de los individuos de dicha población o de una submuestra de la misma. Opcionalmente, dicho fenotipo es la obesidad o los trastornos relacionados con la obesidad. Opcionalmente, en los que dicho trastorno relacionado con la obesidad se selecciona entre el grupo que consiste en aterosclerosis, resistencia a la insulina, hipertensión, microangiopatía en individuos obesos con diabetes de tipo II, lesiones oculares asociadas con una microangiopatía en individuos obesos con diabetes de tipo II, y lesiones renales asociadas con una microangiopatía en individuos obesos con diabetes de tipo II.

La invención también incluye métodos para estimar la frecuencia de un haplotipo de un conjunto de marcadores bialélicos en una población, que comprende las etapas de: a) genotipificar al menos dos marcadores bialélicos relacionados con *APM1* en cada uno de los individuos de dicha población o de una submuestra de la misma, conforme a un método de genotipificación de la invención; y b) aplicar un método de determinación de haplotipos a las identidades de los nucleótidos determinados en las etapas a) para obtener una estimación de dicha frecuencia. Además, los métodos de estimación de la frecuencia de un haplotipo de la invención, incluyen métodos con alguna otra limitación descrita en esta divulgación, o los métodos que siguen a continuación, especificados individualmente o en cualquiera de sus combinaciones. Opcionalmente, dicho marcador bialélico se puede seleccionar entre el grupo que consiste en A1, A2, A3, A4, A5, A6, A7 y A8, y sus complementos. Opcionalmente, dicho método de determinación de haplotipos se realiza mediante una amplificación por PCR asimétrica, mediante doble amplificación por PCR de alelos específicos, mediante el algoritmo de Clark o mediante un algoritmo de tipo esperanza-maximización.

Una realización adicional de la presente invención incluye métodos para detectar una asociación entre un haplotipo y un fenotipo, que comprenden las etapas de: a) estimar la frecuencia de al menos un haplotipo en una población positiva con respecto a un rasgo, conforme a un método de la invención para estimar la frecuencia de un haplotipo; b) estimar la frecuencia de dicho haplotipo en una población de control, conforme a un método de la invención para estimar la frecuencia de un haplotipo; y c) determinar si existe una asociación estadísticamente significativa entre dicho haplotipo y dicho fenotipo. Además, los métodos para detectar una asociación entre un haplotipo y un fenotipo de la invención, incluyen métodos con alguna otra limitación descrita en esta divulgación, o los métodos que siguen a continuación. Opcionalmente, dicho marcador bialélico se puede seleccionar entre el grupo que consiste en A1, A2, A3, A4, A5, A6, A7 y A8, y sus complementos. Opcionalmente, dicha población de control es una población negativa con respecto a un rasgo, o una población al azar. Opcionalmente, dicho fenotipo es la obesidad o un trastorno relacionado con la obesidad. Opcionalmente, dicho método comprende las etapas adicionales de determinar el fenotipo en dicha población positiva con respecto a un rasgo y en dicha población de control, antes de la etapa c). Opcionalmente, en los que dicho trastorno relacionado con la obesidad se selecciona entre el grupo que consiste en

ateroesclerosis, resistencia a la insulina, hipertensión, microangiopatía en individuos obesos con diabetes de tipo II, lesiones oculares asociadas con una microangiopatía en individuos obesos con diabetes de tipo II, o lesiones renales asociadas con una microangiopatía en individuos obesos con diabetes de tipo II.

5 *Análisis de ligamiento*

El análisis de ligamiento se basa en establecer una correlación entre la transmisión de marcadores genéticos y la transmisión de un rasgo específico a lo largo de las generaciones de una familia. Por ello, el objetivo del análisis de ligamiento es detectar loci de marcadores que muestran una cosegregación con un rasgo de interés presente en las genealogías.

Métodos paramétricos

Cuando se puede disponer de datos provenientes de generaciones sucesivas, existe la oportunidad de estudiar el grado de ligamiento entre pares de loci. Las estimaciones de la fracción de recombinación permiten ordenar los loci y situarlos sobre un mapa genético. Con loci que son marcadores genéticos, se puede establecer un mapa genético, y luego se puede calcular la fuerza del ligamiento entre los marcadores y los rasgos, y usar para indicar las posiciones relativas de los marcadores y de los genes que afectan a esos rasgos (Weir, 1996). El método clásico de los análisis de ligamiento es el método de puntuaciones basado en el logaritmo de *odds* (ventajas) (lod) (del inglés, “*logarithm of odds*”) (véanse Morton, 1955; Ott, 1991). El cálculo de las puntuaciones lod requiere la especificación del modo de herencia de la enfermedad (método paramétrico). Generalmente, la longitud de la región candidata identificada usando el análisis de ligamiento está entre 2 Mb y 20 Mb. Una vez identificada una región candidata, según se ha descrito anteriormente, un análisis de los individuos recombinantes usando marcadores adicionales permite una mejor delineación de la región candidata. Los estudios de análisis de ligamiento se han basado por lo general en el uso de un máximo 5.000 marcadores microsatélite, limitando así la resolución máxima teóricamente alcanzable del análisis de ligamiento a aproximadamente 600 kb como promedio.

El análisis de ligamiento se ha aplicado con éxito a la elaboración de mapas de rasgos genéticos simples que muestran unos patrones claros de herencia mendeliana y que tienen una penetrancia alta (es decir, el cociente entre el número de portadores del alelo *a* positivos con respecto a un rasgo y el número total de portadores del alelo *a* en la población). Sin embargo, el análisis paramétrico del ligamiento adolece de varios inconvenientes. Primero, está limitado por su fiabilidad en la elección de un modelo genético adecuado para cada uno de los rasgos estudiados. Además, como se ha mencionado anteriormente, la resolución alcanzable cuando se usa usando el análisis de ligamiento es limitada, y se requieren estudios complementarios para perfeccionar el análisis de las regiones típicas de 2 Mb a 20 Mb, inicialmente identificadas por el análisis de ligamiento. Además, las estrategias basadas en análisis paramétricos de ligamiento han demostrado ser difíciles cuando se aplican a rasgos genéticos complejos, tales como los debidos a la acción combinada de múltiples genes y factores medioambientales. Es muy difícil modelar estos factores de manera adecuada en un análisis basado en puntuaciones lod. En estos casos, son necesarios un esfuerzo y un coste demasiado grandes para reclutar el número adecuado de familias afectadas, requerido para aplicar un análisis de ligamiento a estas situaciones, según ha sido recientemente tratado por Risch, N. y Merikangas, K. (1996).

Métodos no paramétricos

La ventaja de los denominados métodos no paramétricos de análisis de ligamiento es que no requieren especificación del modo de herencia de la enfermedad, y tienden a ser más útiles para el análisis de rasgos complejos. En los métodos no paramétricos, se trata de probar que el patrón de herencia de una región cromosómica no es consistente con la segregación mendeliana al azar, demostrando que los familiares afectados heredan copias idénticas de la región más frecuentemente de lo esperado por casualidad. Los familiares afectados deberían mostrar “un reparto de alelos” en exceso incluso en presencia de una penetrancia incompleta y una herencia poligénica. En análisis no paramétricos de ligamiento, el grado de concordancia en el locus de un marcador en dos individuos se puede medir o bien por el número de alelos idénticos por función (IBS) (del inglés, “*identical by state*”) o bien por el número de alelos idénticos por ascendencia (IBD) (del inglés, “*identical by descent*”). El análisis de parejas de familiares afectados es un caso especial muy conocido y es la forma más sencilla de estos métodos.

Los marcadores bialélicos de la presente invención se pueden usar en análisis de ligamiento tanto paramétricos como no paramétricos. Preferiblemente, los marcadores bialélicos se pueden usar en métodos no paramétricos que permiten el mapeo de genes implicados en rasgos complejos. Los marcadores bialélicos de la presente invención se pueden usar en métodos tanto IBD como IBS para mapear genes que afectan a un rasgo complejo. En esos estudios, aprovechando la elevada densidad de los marcadores bialélicos, se pueden reunir varios loci de marcadores bialélicos adyacentes para obtener la eficiencia alcanzada por los marcadores multialélicos (Zhao y col., 1998).

Estudios de asociación en poblaciones

La presente invención comprende métodos para identificar si el gen *APM1* está asociado con un rasgo detectable usando los marcadores bialélicos de la presente invención. En una de las realizaciones, la presente invención comprende métodos para detectar una asociación entre un alelo de un marcador bialélico, o un haplotipo de un marcador bialélico, y un rasgo. Además, la invención comprende métodos para identificar un alelo causante de un rasgo, en desequilibrio de ligamiento con un alelo de un marcador bialélico de la presente invención.

Según se ha descrito anteriormente, se pueden utilizar estrategias alternativas para realizar estudios de asociación: estudios de asociación en la totalidad del genoma, estudios de asociación en regiones candidatas y estudios de asociación en genes candidatos. En una realización preferida, los marcadores bialélicos de la presente invención se usan para realizar estudios de asociación en genes candidatos. El análisis de genes candidatos proporciona claramente una estrategia de tipo atajo para la identificación de genes y de polimorfismos génicos relacionados con un rasgo particular, cuando se dispone de cierta información concerniente con la biología del rasgo. Además, los marcadores bialélicos de la presente invención se pueden incorporar a un mapa de marcadores genéticos del genoma humano con el fin de realizar estudios de asociación en la totalidad del genoma. Se han descrito métodos para generar un mapa de alta densidad de marcadores bialélicos en la solicitud de Patente provisional de EE. UU. número 60/082.614. Los marcadores bialélicos de la presente invención se pueden también incorporar a un mapa de una región candidata y específica del genoma (un cromosoma específico o un segmento cromosómico específico, por ejemplo).

Según se ha mencionado anteriormente, los estudios de asociación se pueden llevar a cabo en la población general y no se limitan a estudios realizados en individuos emparentados de las familias afectadas. Los estudios de asociación son extremadamente valiosos porque permiten el análisis de rasgos esporádicos o multifactoriales. Además, los estudios de asociación representan un método poderoso para realizar mapas a escala fina, lo que hace posible mapeos mucho más precisos de alelos causantes de rasgos que los de los estudios de ligamiento. Los estudios basados en genealogías muchas veces solamente estrechan la localización del alelo causante del rasgo. Los estudios de asociación que usan los marcadores bialélicos de la presente invención se pueden usar por lo tanto para afinar la localización de un alelo causante de un rasgo en una región candidata identificada mediante métodos de análisis de ligamiento. Además, una vez identificado un segmento cromosómico de interés, la presencia de un gen candidato, tal como un gen candidato de la presente invención, en la región de interés, puede proporcionar un atajo para la identificación del alelo causante del rasgo. Los marcadores bialélicos de la presente invención se pueden usar para demostrar que un gen candidato está asociado con un rasgo. Estos usos se contemplan específicamente en la presente invención.

Determinación de la frecuencia de un alelo de un marcador bialélico o de un haplotipo de un marcador bialélico en una población

Los estudios de asociación exploran las relaciones entre las frecuencias correspondientes a los conjuntos de alelos entre loci.

Determinación de la frecuencia de un alelo en una población

Las frecuencias alélicas de los marcadores bialélicos en las poblaciones, se pueden determinar usando uno de los métodos anteriormente descritos en el epígrafe "Métodos para genotipificar a un individuo con respecto a la presencia de marcadores bialélicos", o usando cualquier procedimiento de genotipificación adecuado para el fin propuesto. La genotipificación de una mezcla de muestras o de muestras individuales puede determinar la frecuencia de un alelo de un marcador bialélico en una población. Una de las maneras de disminuir el número de genotipificaciones requeridas es usar una mezcla de muestras. Uno de los principales obstáculos del uso de una mezcla de muestras está en relación con la exactitud y la reproducibilidad para determinar concentraciones exactas de DNA al constituir las mezclas. La genotipificación de muestras individuales proporciona una mayor sensibilidad, reproducibilidad y exactitud, y es el método preferido usado en la presente invención. Preferiblemente, cada individuo se genotipifica por separado y se aplica un simple recuento de genes para determinar la frecuencia de un alelo de un marcador bialélico o de un genotipo en una población determinada.

Determinación de la frecuencia de un haplotipo en una población

La fase gamética de los haplotipos se desconoce cuando individuos diploides son heterocigotos en más de un locus. Usando la información genealógica de las familias, la fase gamética se puede inferir a veces (Perlin y col., 1994). Cuando no se puede disponer de información genealógica, se pueden usar diferentes estrategias. Una de las posibilidades es que los diploides heterocigotos en múltiples sitios se puedan eliminar del análisis, dejando sólo a los homocigotos y a los individuos heterocigotos en un único sitio, pero esta estrategia podría dar lugar a un posible sesgo en la composición de la muestra y a la infraestimación de los haplotipos de baja frecuencia. Otra posibilidad es que los cromosomas individuales se puedan estudiar independientemente, por ejemplo, mediante amplificación por una PCR asimétrica (véanse Newton y col., 1989; Wu y col., 1989) o mediante el aislamiento de un cromosoma individual por dilución límite, seguida de una amplificación por PCR (véase Ruano y col., 1990). Además, se puede haplotipificar una muestra con respecto a la presencia de marcadores bialélicos suficientemente próximos mediante amplificación por doble PCR de alelos específicos (Sarkar, G. y Sommer S. S., 1991). Estas estrategias no son del todo satisfactorias, o bien debido a su complejidad técnica, o al coste adicional que conllevan, o bien debido a su falta de generalización a gran escala, o a los posibles sesgos que introducen. Para solucionar estas dificultades, se puede usar un algoritmo que infiere la fase de los genotipos de DNA amplificados por PCR, introducido por Clark, A. G. (1990). Brevemente expuesto, el principio es empezar rellenando una lista preliminar de haplotipos presentes en la muestra, estudiando los individuos no ambiguos, es decir, los homocigotos completos y los heterocigotos en un único sitio. Después otros individuos de la misma muestra se someten a escrutinio con respecto a la posible ocurrencia de haplotipos previamente reconocidos. Por cada identificación positiva, el haplotipo complementario se añade a la lista de haplotipos reconocidos, hasta que la información de la fase para todos los individuos o bien se resuelva o bien se identifique como no resuelta. Este método asigna un solo haplotipo a cada individuo multiheterocigoto, aunque algunos haplotipos son posibles cuando hay más de un solo sitio heterocigoto. Como alternativa, se pueden usar métodos que

estiman las frecuencias de los haplotipos en una población sin asignar haplotipos a cada individuo. Preferiblemente se usa un método basado en el algoritmo de esperanza-maximización (EM) (Dempster y col., 1977) que da lugar a estimaciones de máxima verosimilitud de las frecuencias de haplotipos en el supuesto de las proporciones de Hardy-Weinberg (de emparejamiento al azar) (véase Excoffier L. y Slatkin M., 1995). El algoritmo EM es una estrategia generalizada e iterativa basada en la máxima verosimilitud, para obtener una estimación, que es útil cuando los datos son ambiguos y/o incompletos. El algoritmo EM se usa para resolver individuos heterocigotos en haplotipos. Las estimaciones de los haplotipos se describen con detalle más adelante en el epígrafe “Métodos estadísticos”. Se puede usar cualquier otro método conocido en la técnica para determinar o para estimar la frecuencia de un haplotipo en una población.

Análisis de desequilibrio de ligamiento

El desequilibrio de ligamiento es una asociación no al azar de alelos situados en dos o más loci y representa una poderosa herramienta para mapear genes implicados en rasgos asociados con enfermedades (véase Ajioka R. S. y col., 1997). Los marcadores bialélicos, debido a que están densamente repartidos en el genoma humano y a que se pueden genotipificar en mayores cantidades que otros tipos de marcadores genéticos (tales como los marcadores RFLP o VNTR), son particularmente útiles en análisis genéticos basados en el desequilibrio de ligamiento.

Cuando una mutación patológica se introduce por primera vez en una población (por medio de una mutación nueva o de la inmigración de un vehículo de la mutación), esa mutación necesariamente reside en un solo cromosoma y por ello en un único haplotipo “antecesor” o “ancestral” de marcadores ligados. Por consiguiente, existe un desequilibrio completo entre estos marcadores y la mutación causante de la enfermedad: la mutación causante de la enfermedad se descubre solamente en presencia de un conjunto específico de alelos de marcadores. A través de las generaciones sucesivas, se producen eventos de recombinación entre la mutación causante de la enfermedad y estos polimorfismos marcadores, y el desequilibrio se disipa gradualmente. El ritmo de esta disipación es una función de la frecuencia de recombinación, de manera que los marcadores más cercanos al gen causante de la enfermedad manifestaran niveles más altos de desequilibrio que los que están más alejados. Cuando no se rompen a causa de una recombinación, los haplotipos “ancestrales” y el desequilibrio de ligamiento entre alelos de marcadores situados en diferentes loci, pueden ser rastreados no solamente a través de las genealogías sino también a través de las poblaciones. El desequilibrio de ligamiento se suele observar en forma de una asociación entre un alelo específico situado en uno de los locus, y el otro alelo específico situado en un segundo locus.

El patrón o curva de desequilibrio entre una enfermedad y los loci de marcadores se espera que exhiba un máximo que se produce en el locus relacionado con la enfermedad. Por consiguiente, la cantidad de desequilibrio de ligamiento entre un alelo relacionado con una enfermedad y marcadores genéticos estrechamente ligados, puede dar lugar a una información valiosa concerniente a la localización del gen causante de la enfermedad. Para el mapeo a escala fina de un locus relacionado con una enfermedad, es útil tener cierto conocimiento de los patrones de desequilibrio de ligamiento que existen entre los marcadores de la región estudiada. Según se ha mencionado anteriormente, la resolución de los mapeos conseguida a través del análisis de desequilibrio de ligamiento, es mucho más alta que la obtenida de otros estudios de ligamiento. La alta densidad de los marcadores bialélicos junto con el análisis de desequilibrio de ligamiento, proporciona poderosas herramientas para la elaboración de mapas a escala fina. Diferentes métodos para calcular el desequilibrio de ligamiento se describen más adelante en el epígrafe “Métodos estadísticos”.

Estudios caso-control en poblaciones, de asociaciones de rasgo-marcador

Según se ha mencionado anteriormente, la ocurrencia de parejas de alelos específicos en loci diferentes situados en el mismo cromosoma no es al azar y la desviación respecto al suceso al azar se denomina desequilibrio de ligamiento. Los estudios de asociación se centran en las frecuencias halladas en la población y se basan en el fenómeno del desequilibrio de ligamiento. Si un alelo específico presente en un gen determinado está directamente implicado en la causa de un rasgo particular, su frecuencia se incrementará estadísticamente en una población afectada (población positiva con respecto a ese rasgo), cuando se compara con la frecuencia hallada en una población negativa con respecto a ese rasgo, o en una población de control elegida al azar. Como consecuencia de la existencia del desequilibrio de ligamiento, la frecuencia de los demás alelos presentes en el haplotipo que porta el alelo causante del rasgo se incrementará también en individuos positivos con respecto a ese rasgo, en comparación con la frecuencia encontrada en individuos negativos con respecto a ese rasgo, o en controles elegidos al azar. Por lo tanto, la asociación entre el rasgo y un alelo (específicamente un alelo de un marcador bialélico) en desequilibrio de ligamiento con el alelo causante del rasgo, bastará para sugerir la presencia de un gen relacionado con el rasgo en esa región concreta. Las poblaciones caso-control se pueden genotipificar con respecto a la presencia de marcadores bialélicos para identificar asociaciones que localizan muy estrechamente un alelo causante de un rasgo. Al igual que cualquier marcador en desequilibrio de ligamiento con un marcador determinado asociado con un rasgo, el alelo estará asociado con ese rasgo. El desequilibrio de ligamiento permite que las frecuencias relativas halladas en poblaciones caso-control, de un número limitado de polimorfismos genéticos (específicamente de marcadores bialélicos), sean analizadas como una alternativa al escrutinio de todos los polimorfismos funcionales posibles con el fin de encontrar alelos causantes de rasgos. Los estudios de asociación comparan la frecuencia de alelos marcadores en poblaciones caso-control no emparentadas y representan herramientas poderosas para la disección de rasgos complejos.

Poblaciones caso-control (Criterios de inclusión)

Los estudios de asociación basados en poblaciones no conciernen a la herencia familiar pero comparan la prevalencia de un marcador genético particular, o de un conjunto de marcadores, en poblaciones caso-control. Son estudios caso-control basados en la comparación de individuos caso no emparentados (afectados o positivos con respecto a un rasgo) e individuos de control no emparentados (no afectados, negativos con respecto a un rasgo o elegidos al azar). Preferiblemente, el grupo de control se compone de individuos no afectados o de individuos negativos con respecto a un rasgo. Además, el grupo de control está étnicamente emparejado con la población caso. Además, el grupo de control preferiblemente se empareja con la población caso con respecto al principal factor de confusión conocido del rasgo en estudio (por ejemplo, una correspondencia por edad en el caso de un rasgo dependiente de la edad). En el mejor de los casos, los individuos de las dos muestras se emparejan de manera que se espera que difieran únicamente en el estado de la enfermedad. Las expresiones “población positiva con respecto a un rasgo”, “población caso” y “población afectada” se usan aquí indistintamente.

Una etapa importante en la disección de rasgos complejos usando estudios de asociación es la elección de las poblaciones caso-control o (véase Lander y Schork, 1994). Una etapa muy importante en la elección de las poblaciones caso-control es la definición clínica de un rasgo o fenotipo determinado. Cualquier rasgo genético se puede analizar mediante el método de asociación aquí propuesto, seleccionando cuidadosamente los individuos que han de ser incluidos en los grupos fenotípicos positivos con respecto al rasgo y negativos con respecto al rasgo. Suelen ser útiles cuatro criterios: el fenotipo clínico, la edad de aparición, la genealogía y la gravedad. El procedimiento de selección para rasgos continuos o cuantitativos (tales como la presión sanguínea por ejemplo) implica seleccionar individuos situados en extremos opuestos de la distribución fenotípica del rasgo en estudio, para incluir en estas poblaciones positivas con respecto al rasgo y negativas con respecto al rasgo a individuos con fenotipos que no se superponen. Preferiblemente, las poblaciones caso-control consisten en poblaciones fenotípicamente homogéneas. Las poblaciones positivas con respecto a un rasgo y negativas con respecto a un rasgo consisten en poblaciones fenotípicamente uniformes de individuos que representan en cada caso entre el 1% y el 98%, preferiblemente entre el 1% y el 80%, más preferiblemente entre el 1% y el 50%, y más preferiblemente entre el 1% y el 30%, y lo más preferiblemente entre el 1% y el 20% de la población total en estudio, y que preferiblemente se seleccionan entre individuos que exhiben fenotipos que no se superponen. Cuanto más clara es la diferencia entre los dos fenotipos del rasgo, mayor es la probabilidad de detectar una asociación con marcadores bialélicos. La selección de estos fenotipos radicalmente diferentes pero relativamente uniformes, permite hacer comparaciones eficientes en los estudios de asociación y la posible detección de diferencias marcadas a nivel genético, a condición de que los tamaños de las muestras de las poblaciones en estudio sean suficientemente importantes.

En las realizaciones preferidas, un primer grupo de entre 50 y 300 individuos positivos con respecto a un rasgo, preferiblemente un grupo de aproximadamente 100 individuos, se recluta de acuerdo con sus fenotipos. Un número similar de individuos negativos con respecto a un rasgo se incluye en estos estudios.

En la presente invención, los ejemplos típicos de criterios de inclusión incluyen la obesidad y los trastornos relacionados con la obesidad.

Análisis de asociación

La estrategia general para llevar a cabo estudios de asociación usando marcadores bialélicos obtenidos de una región que porta un gen candidato, consiste en explorar a dos grupos de individuos (poblaciones caso-control) con el fin de medir y de comparar estadísticamente las frecuencias alélicas de los marcadores bialélicos de la presente invención en ambos grupos.

Si una asociación estadísticamente significativa con un rasgo se identifica para al menos uno o más de los marcadores bialélicos analizados, se puede suponer que: o el alelo asociado es directamente responsable de causar el rasgo (es decir, el alelo asociado es el alelo causante del rasgo), o más probablemente, el alelo asociado está en desequilibrio de ligamiento con el alelo causante del rasgo. Las características específicas del alelo asociado con respecto a la función del gen candidato, normalmente proporcionan una nueva visión de la relación entre el alelo asociado y el rasgo (causal o en desequilibrio de ligamiento). Si la evidencia indica que el alelo asociado presente en el gen candidato es muy probable que no sea el alelo causante del rasgo sino que está en desequilibrio de ligamiento con el verdadero alelo causante del rasgo, en ese caso el alelo causante del rasgo se puede hallar secuenciando las inmediateces del marcador asociado y llevando a cabo nuevos estudios de asociación con los polimorfismos que se han revelado de manera iterativa.

Los estudios de asociación se suelen ejecutar en dos etapas sucesivas. En una primera fase, las frecuencias de un número reducido de marcadores bialélicos procedentes del gen candidato, se determinan en la población positiva con respecto al rasgo y en la población negativa con respecto al rasgo. En una segunda fase del análisis, la posición de los loci genéticos responsables del rasgo concreto se afina aún más usando una densidad mayor de marcadores procedentes de la región relevante. Sin embargo, si el gen candidato en estudio tiene una longitud relativamente pequeña, como es el caso de APM1, una sola fase puede ser suficiente para establecer asociaciones significativas.

Análisis de haplotipos

Según se ha descrito anteriormente, cuando un cromosoma que porta un alelo relacionado con una enfermedad aparece por primera vez en una población como resultado de una mutación o de una migración, el alelo mutante necesariamente reside en un cromosoma que tiene un conjunto de marcadores ligados: el haplotipo ancestral. Este haplotipo se puede rastrear a través de las poblaciones y se puede analizar su asociación estadística con un rasgo concreto. Si se complementan los estudios de asociación monopuntuales (alélicos) con estudios de asociación multipuntuales, también denominados estudios de haplotipos, se aumenta el poder estadístico de los estudios de asociación. De este modo, un estudio de asociación en haplotipos permite que se defina la frecuencia y el tipo del haplotipo portador ancestral. Un análisis de haplotipos es importante porque aumenta el poder estadístico de un análisis que involucra a marcadores individuales.

En una primera fase de un análisis de la frecuencia de haplotipos, se determina la frecuencia de los posibles haplotipos basados en las diferentes combinaciones de los marcadores bialélicos identificados de la invención. La frecuencia de los haplotipos se compara luego para las distintas poblaciones de individuos positivos con respecto a un rasgo y de individuos de control. El número de individuos positivos con respecto a un rasgo, que se deberían someter a este análisis para obtener resultados estadísticamente significativos, normalmente varía entre 30 y 300 individuos, variando un número preferido de individuos entre 50 y 150. Las mismas consideraciones se aplican al número de individuos no afectados (o de individuos de control elegidos al azar) usados en este estudio. Los resultados de este primer análisis proporcionan frecuencias de haplotipos en poblaciones caso-control y para cada frecuencia de haplotipo evaluada se calcula un valor de p y un cociente de *odds* (ventajas). Si se encuentra una asociación estadísticamente significativa, el riesgo relativo de que un individuo que porta el haplotipo concreto esté afectado por el rasgo en estudio, se puede obtener por aproximación.

Análisis de interacciones

Los marcadores bialélicos de la presente invención se pueden usar también para identificar patrones de marcadores bialélicos asociados con rasgos detectables resultantes de interacciones poligénicas. El análisis de una interacción genética entre alelos situados en loci no ligados, requiere la genotipificación de los individuos usando las técnicas aquí descritas. El análisis de interacciones alélicas entre un conjunto seleccionado de marcadores bialélicos con un nivel adecuado de significación estadística, se puede considerar como un análisis de haplotipos. El análisis de interacciones consiste en estratificar las poblaciones caso-control con respecto a un haplotipo concreto correspondiente a los primeros loci y realizar un análisis de haplotipos con los segundos loci en cada subpoblación.

*Métodos estadísticos usados en los estudios de asociación**Ensayos de ligamiento en presencia de una asociación*

Los marcadores bialélicos de la presente invención se pueden también usar en la prueba TDT (del inglés, "*Transmission/Disequilibrium Test*") (Prueba de transmisión del desequilibrio). La TDT determina tanto el ligamiento como la asociación y no se afecta por la estratificación de la población. La TDT requiere datos de los individuos afectados y de sus padres, o datos procedentes de familiares no afectados en lugar de los de sus padres (véanse Spielmann S. y col., 1993; Schaid D. J. y col., 1996; Spielmann S. y Ewens W. J., 1998). Estos ensayos combinados generalmente disminuyen los errores falsos positivos producidos en análisis independientes.

Métodos estadísticos

En general, se puede usar cualquiera de los métodos conocidos en la técnica que determinan si un rasgo y un genotipo muestran una correlación estadísticamente significativa.

1) *Métodos de análisis de ligamiento*

Los métodos estadísticos y programas informáticos útiles para análisis de ligamiento son muy conocidos por los expertos en la técnica (véase Terwilliger J. D. y Ott J., 1994; Ott J., 1991).

2) *Métodos para estimar frecuencias de haplotipos en una población*

Según se ha descrito anteriormente, cuando los genotipos son puntuados, muchas veces no es posible distinguir a los heterocigotos por lo que las frecuencias de los haplotipos no se pueden inferir fácilmente. Cuando no se conoce la fase gamética, las frecuencias de los haplotipos se pueden estimar a partir de los datos genotípicos multiloci. Cualquiera de los métodos conocidos por una persona experta en la técnica se puede usar para estimar frecuencias de haplotipos (véanse Lange K., 1997; Weir, B. S., 1996). Preferiblemente, las frecuencias de haplotipos de máxima verosimilitud se calculan usando un algoritmo Esperanza-Maximización (EM) (véanse Dempster y col., 1977; Ex-

coffier L. y Slatkin M., 1995). Este procedimiento es un proceso iterativo que tiene como meta obtener estimaciones de máxima verosimilitud de frecuencias de haplotipos a partir de datos genotípicos multiloci cuando no se conoce la fase gamética. Las estimaciones de haplotipos se suelen realizar aplicando el algoritmo EM usando por ejemplo el programa EM-HAPLO (Hawley M. E. y col., 1994) o el programa Arlequin (Schneider y col., 1997). El algoritmo EM es una estrategia generalizada e iterativa, basada en la máxima verosimilitud, para realizar una estimación, y se describe brevemente más adelante.

Obsérvese que en la presente sección “Métodos para estimar frecuencias de haplotipos en una población” de este texto, los fenotipos harán referencia a genotipos multiloci con fase desconocida. Los genotipos harán referencia a genotipos multiloci con fase conocida.

Una muestra de N individuos no emparentados se tipifica con respecto a K marcadores. Los datos observados son los fenotipos de K loci con fase desconocida que se pueden clasificar en F fenotipos diferentes. Supóngase que se tienen H posibles haplotipos subyacentes (en el caso de K marcadores bialélicos, $H = 2^K$).

En el caso del fenotipo j, supóngase que son posibles c_j genotipos. Por lo tanto se tiene la siguiente ecuación:

$$P_j = \sum_{i=1}^{c_j} pr(genotipo_i) = \sum_{i=1}^{c_j} pr(h_k, h_l)$$

Ecuación 1

en la que P_j es la probabilidad del fenotipo j, h_k y h_l son los dos haplotipos que componen el genotipo i. En las condiciones de equilibrio de Hardy-Weinberg, $pr(h_k, h_l)$ se hace:

$$pr(h_k, h_l) = pr(h_k)^2 \quad \text{si} \quad h_k = h_l, \quad pr(h_k, h_l) = 2pr(h_k).pr(h_l) \quad \text{si} \quad h_k \neq h_l$$

Ecuación 2

Las sucesivas etapas del algoritmo EM se pueden describir de la siguiente manera:

Partiendo de los valores iniciales de las frecuencias de los haplotipos, indicados $p_1^{(0)}, p_1^{(0)}, \dots, p_H^{(0)}$, estos valores iniciales sirven para estimar las frecuencias de los genotipos (etapa de Esperanza) y después para estimar otro conjunto de frecuencias de haplotipos (etapa de Maximización), indicados $p_1^{(1)}, p_1^{(1)}, \dots, p_H^{(1)}$, y estas dos etapas se repiten hasta que los cambios en los conjuntos de frecuencias de los haplotipos sean muy pequeños.

Un criterio de terminación puede ser que la diferencia máxima entre las frecuencias de los haplotipos entre dos iteraciones sea menor que 10^{-7} . Estos valores se pueden ajustar de acuerdo con la precisión deseada de las estimaciones.

A una iteración determinada, la etapa de Esperanza consiste en calcular las frecuencias de los genotipos por medio de la siguiente ecuación:

$$pr(genotipo_i)^{(s)} = pr(fenotipo_j).pr(genotipo_i | fenotipo_j)^{(s)}$$

$$= \frac{n_j}{N} \cdot \frac{pr(h_k, h_l)^{(s)}}{P_j^{(s)}}$$

Ecuación 3

en la que el genotipo i ocurre en el fenotipo j, y en la que h_k y h_l constituyen el genotipo i. Cada probabilidad se obtiene conforme a la Ecuación 1 y la Ecuación 2, anteriormente descritas.

Luego la etapa de Maximización simplemente estima otro conjunto de frecuencias de haplotipos dadas las frecuencias de los genotipos. Esta estrategia también se conoce como método de recuento de genes (Smith, 1957).

$$p_i^{(s+1)} = \frac{1}{2} \sum_{j=1}^F \sum_{i=1}^{c_j} \delta_{it} \cdot pr(genotipo_i)^{(s)}$$

Ecuación 4

En la que δ_{it} es un indicador variable que cuenta el número de veces del haplotipo t en el genotipo i . Toma los valores de 0, 1 ó 2.

Para comprobar que la estimación finalmente obtenida es la estimación de máxima verosimilitud, se requieren varios valores de desviación. Las estimaciones obtenidas se comparan y si son diferentes, se mantienen las estimaciones que dan lugar a la mejor verosimilitud.

3) Métodos para calcular el desequilibrio de ligamiento entre marcadores

Se pueden usar varios métodos para calcular el desequilibrio de ligamiento entre dos posiciones genéticas cualesquiera y, en la práctica, el desequilibrio de ligamiento se mide aplicando una prueba estadística de asociación a los datos sobre haplotipos recogidos en una población.

El desequilibrio de ligamiento entre cualquier pareja de marcadores bialélicos que comprenden al menos uno de los marcadores bialélicos de la presente invención (M_i , M_j) que contienen los alelos (a_i/b_i) en el marcador M_i y los alelos (a_j/b_j) en el marcador M_j , se puede calcular para todas las combinaciones de alelos (a_i, a_j ; a_i, b_j ; b_i, a_j y b_i, b_j) conforme a la fórmula de Piazza:

$$\Delta_{aiaj} = \sqrt{\theta 4} - \sqrt{(\theta 4 + \theta 3)(\theta 4 + \theta 2)},$$

en la que:

$\theta 4 = --$ = frecuencia de los genotipos que no contienen el alelo a_i en M_i y que no contienen el alelo a_j en M_j .

$\theta 3 = -+$ = frecuencia de los genotipos que no contienen el alelo a_i en M_i y que contienen el alelo a_j en M_j .

$\theta 2 = +- =$ frecuencia de los genotipos que contienen el alelo a_i en M_i y que no contienen el alelo a_j en M_j .

El desequilibrio de ligamiento (LD) entre parejas de marcadores bialélicos (M_i , M_j) se puede también calcular para todas las combinaciones de alelos (a_i, a_j ; a_i, b_j ; b_i, a_j y b_i, b_j) conforme a la estimación de máxima verosimilitud (MLE) (del inglés, “*Maximum-Likelihood Estimate*”) para delta (coeficiente de desequilibrio genotípico compuesto), de la manera descrita por Weir (Weir B.S., 1996). La MLE para el desequilibrio de ligamiento compuesto es:

$$D_{aiaj} = (2n_1 + n_2 + n_3 + n_4 / 2) / N - 2(pr(a_i) \cdot pr(a_j))$$

en la que

$n_1 = \Sigma$ fenotipo (a_i/a_i , a_j/a_j),

$n_2 = \Sigma$ fenotipo (a_i/a_i , a_j/b_j),

$n_3 = \Sigma$ fenotipo (a_i/b_i , a_j/a_j),

$n_4 = \Sigma$ fenotipo (a_i/b_i , a_j/b_j) y N es el número de individuos de la muestra.

Esta fórmula permite que el desequilibrio de ligamiento entre alelos se estime cuando únicamente existen disponibles datos sobre el genotipo y no sobre el haplotipo.

Otra manera de calcular el desequilibrio de ligamiento entre marcadores es la siguiente. En el caso de una pareja de marcadores bialélicos M_i (a_i/b_i) y M_j (a_j/b_j), que se ajustan al equilibrio de Hardy-Weinberg, se puede hacer una estimación de las cuatro posibles frecuencias de haplotipos en una población determinada, conforme a la estrategia anteriormente descrita.

La estimación del desequilibrio gamético entre a_i y a_j es simplemente:

$$D_{aiaj} = pr(haplotipo(a_i, a_j)) - pr(a_i) \cdot pr(a_j)$$

en donde $pr(a_i)$ es la probabilidad del alelo a_i ; y $pr(a_j)$ es la probabilidad del alelo a_j , y en donde $pr(haplotipo(a_i, a_j))$ se estima como en la anterior Ecuación 3.

En el caso de una pareja de marcadores bialélicos en la que sólo uno de ellos mide el desequilibrio, es necesario describir la asociación entre M_i y M_j .

Luego se calcula un valor normalizado de la fórmula anterior de la siguiente manera:

$$D'_{aiaj} = D_{aiaj} / \max(-pr(a_i) \cdot pr(a_j), -pr(b_i) \cdot pr(b_j)) \quad \text{con} \quad D_{aiaj} < 0$$

$$D'_{aiaj} = D_{aiaj} / \max(pr(b_i) \cdot pr(a_j), pr(a_i) \cdot pr(b_j)) \quad \text{con} \quad D_{aiaj} > 0$$

El experto pronto se percatará de que se pueden usar otros métodos para realizar el cálculo del LD.

El desequilibrio de ligamiento entre un conjunto de marcadores bialélicos que tienen una tasa de heterozigosidad adecuada, se puede determinar genotipificando entre 50 y 1.000 individuos no emparentados, preferiblemente entre 75 y 200 individuos no emparentados, y más preferiblemente alrededor de 100 individuos no emparentados.

4) Ensayos de asociación

Métodos para determinar la significación estadística de una correlación entre un fenotipo y un genotipo, en este caso un alelo situado en un marcado dialélico o un haplotipo constituido por esos alelos, se pueden determinar por medio de cualquiera de las pruebas estadísticas conocidas en la técnica y requiriéndose un umbral aceptado de significación estadística. La aplicación de los métodos particulares y de los umbrales de significación están dentro de la experiencia normal del profesional de la técnica.

Un ensayo de asociación se realiza determinando la frecuencia de un alelo de un marcador bialélico en poblaciones caso y en poblaciones control y comparando estas frecuencias con una prueba estadística para determinar si existe una diferencia estadísticamente significativa en la frecuencia, lo que indicaría la existencia de una correlación entre el rasgo y el alelo del marcador bialélico en estudio. De manera similar, un análisis de haplotipos se realiza estimando las frecuencias de todos los haplotipos posibles de un conjunto determinado de marcadores bialélicos en poblaciones caso y en poblaciones control, y comparando estas frecuencias con una prueba estadística para determinar si existe una correlación estadísticamente significativa entre el haplotipo y el fenotipo (rasgo) en estudio. Se puede usar cualquier herramienta estadística que sirva para probar la existencia de una asociación estadísticamente significativa entre un genotipo y un fenotipo. Preferiblemente, la prueba estadística utilizada es una prueba de Chi cuadrado con un grado de libertad. Se calcula un valor de P (el valor P es la probabilidad de que un suceso estadístico de la misma extensión o de una extensión mayor que el observado, pudiera ocurrir por casualidad).

Significación estadística

En las realizaciones preferidas, la significación para fines diagnósticos, ya sea como base positiva para nuevos ensayos diagnósticos, o como un punto de partida preliminar para una terapia temprana preventiva, el valor de p en relación con una asociación a marcadores bialélicos, preferiblemente es de aproximadamente 1×10^{-2} o inferior, más preferiblemente es de aproximadamente 1×10^{-4} o inferior, en el caso del análisis de un único marcador bialélico, y es de aproximadamente 1×10^{-3} o menor, aún más preferiblemente es de 1×10^{-6} o menor, y lo más preferiblemente es de aproximadamente 1×10^{-8} o menor, en el caso de un análisis de haplotipos en el que están implicados dos o más marcadores. Se piensa que estos valores son aplicables a todos los estudios de asociación en los que están implicados marcadores individuales o combinaciones de marcadores múltiples.

La persona experta puede usar el intervalo de valores anteriormente presentado como punto de partida, con el fin de llevar a cabo estudios de asociación con marcadores bialélicos de la presente invención. Haciendo esto, se pueden poner de manifiesto asociaciones significativas entre los marcadores bialélicos de la presente invención y la obesidad o los trastornos relacionados con la obesidad, y se pueden usar con fines diagnósticos y para el escrutinio de fármacos.

Permutación fenotípica

Con el fin de confirmar la significación estadística de los análisis de haplotipos de la primera fase, anteriormente descritos, podría ser conveniente realizar otros análisis en los que los datos de genotipificación procedentes de individuos caso-control se reúnen y se aleatorizan con respecto al fenotipo de un rasgo. Cada uno de los datos individuales de genotipificación se distribuye aleatoriamente en dos grupos, que contienen el mismo número de individuos que las poblaciones caso-control usadas para compilar los datos obtenidos en la primera fase. Un análisis de haplotipos de la segunda fase se ejecuta preferiblemente sobre estos grupos artificiales, preferiblemente con respecto a los marcadores incluidos en el haplotipo del análisis de la primera fase que muestra el coeficiente de riesgo relativo más alto. Este experimento se repite preferiblemente al menos entre 100 y 10.000 veces. Las iteraciones repetidas permiten la determinación del porcentaje de haplotipos obtenidos con un valor de p significativo, inferior a aproximadamente 1×10^{-3} .

Evaluación de una asociación estadística

Para abordar el problema de los falsos positivos, se puede realizar un análisis similar con las mismas poblaciones caso-control en regiones genómicas aleatorias. Los resultados obtenidos de regiones al azar y en la región candidata se comparan según se describe en la solicitud de Patente provisional titulada "Métodos, programan y aparatos para identificar regiones genómicas que albergan un gen asociado con un rasgo detectable", número de serie 60/107.986, presentada el 10 de Noviembre de 1998, cuyos contenidos se incorporan aquí como referencia.

5) Evaluación de factores de riesgo

La asociación entre un factor de riesgo (en epidemiología genética el factor de riesgo es la presencia o ausencia de un determinado alelo o haplotipo en los loci de un marcador) y una enfermedad se mide por medio del cociente de ventajas (OR) (del inglés, "*Odds Ratio*") y por medio del riesgo relativo (RR). Cuando $P(R^+)$ es la probabilidad de contraer la enfermedad que tienen los individuos con R, y $P(R^-)$ es la probabilidad que tienen los individuos sin el factor de riesgo, el riesgo relativo es simplemente el cociente entre estas dos probabilidades, es decir:

$$RR = P(R^+)/P(R^-)$$

En los estudios caso-control, no se pueden obtener medidas directas del riesgo relativo debido al diseño del muestreo. Sin embargo, el cociente de ventajas (*odds*) permite una buena aproximación del riesgo relativo en el caso de enfermedades de baja incidencia, y se puede calcular de la siguiente manera:

$$OR = (F^+/(1-F^+))/(F^-/(1-F^-))$$

F^+ es la frecuencia de la exposición al factor de riesgo en los casos y F^- es la frecuencia de la exposición al factor de riesgo en los controles. F^+ y F^- se calculan usando las frecuencias alélicas o las frecuencias de los haplotipos del estudio y además dependen del modelo genético subyacente (dominante, recesivo, aditivo, etc.).

Se puede también hacer una estimación del riesgo atribuible (AR), que describe la proporción de individuos de una población que exhiben un rasgo debido a un factor de riesgo concreto. Esta medida es importante para la cuantificación del papel de un factor específico en la etiología de una enfermedad, y en lo que se refiere al impacto de un factor de riesgo en la salud pública. La relevancia en la salud pública de esta medida se basa en la estimación de la proporción de casos de la enfermedad que ocurren en la población que podrían prevenirse si la exposición de interés estuviera ausente. El AR se determina de la siguiente manera:

$$AR = P_E(RR-1)/(P_E(RR-1)+1)$$

AR es el riesgo atribuible a un alelo de un marcador bialélico o a un haplotipo de marcadores bialélicos. P_E es la frecuencia de exposición a un alelo o a un haplotipo dentro de la población en general; y RR es el riesgo relativo que se calcula de manera aproximada con el cociente de ventajas (*odds*) cuando el rasgo en estudio presenta una incidencia relativamente baja en la población general.

Identificación de marcadores bialélicos en desequilibrio de ligamiento con los marcadores bialélicos de la invención

Una vez se ha identificado un primer marcador bialélico en una región genómica de interés, un profesional con una experiencia normal en la técnica, usando las enseñanzas de la presente invención, es capaz de identificar con facilidad marcadores bialélicos adicionales que están en desequilibrio de ligamiento con este primer marcador. Según se ha mencionado anteriormente, cualquier marcador en desequilibrio de ligamiento con un primer marcador asociado con un rasgo, estará asociado con ese rasgo. Por lo tanto, una vez que se ha demostrado la existencia de una asociación entre un marcador bialélico concreto y un rasgo, el descubrimiento de marcadores bialélicos adicionales asociados con

ese rasgo tiene un gran interés con el fin de aumentar la densidad de marcadores bialélicos en esa región concreta. El gen causal o la mutación causal se encontrarán en las proximidades del marcador o del conjunto de marcadores que muestran la correlación más alta con ese rasgo.

- 5 La identificación de marcadores adicionales en desequilibrio de ligamiento con un marcador concreto comprende: (a) amplificar un fragmento genómico que comprende un primer marcador bialélico procedente de una pluralidad de individuos; (b) identificar los segundos marcadores bialélicos presentes en la región genómica que alberga dicho primer marcador bialélico; (c) llevar a cabo un análisis de desequilibrio de ligamiento entre dicho primer marcador bialélico y dichos segundos marcadores bialélicos; y (d) seleccionar dichos segundos marcadores bialélicos cuando están en desequilibrio de ligamiento con dicho primer marcador. También se contemplan subcombinaciones que comprenden las etapas (b) y (c).

- 15 Los métodos para identificar marcadores bialélicos y para llevar a cabo análisis de desequilibrio de ligamiento se describen aquí y pueden ser llevados a cabo por una persona experta sin una experimentación excesiva. La presente invención también concierne por lo tanto a marcadores bialélicos que están en desequilibrio de ligamiento con los marcadores bialélicos específicos A1, A2, A3, A4, A5, A6, A7 y A8, y que se espera que presenten características similares en relación con su asociación respectiva con un rasgo concreto.

Estudios de mapeo: Identificación de mutaciones funcionales

- 20 Una vez se ha confirmado la existencia de una asociación positiva con un marcador bialélico de la presente invención, un gen situado en la región candidata asociada (en el desequilibrio de ligamiento del gen *APM1*) puede ser explorada con respecto a la búsqueda de mutaciones, comparando las secuencias de un número seleccionado de individuos positivos con respecto al rasgo y de individuos negativos con respecto al rasgo. En una realización preferida, las regiones funcionales tales como los exones y sitios de corte y empalme, promotores y otras regiones reguladoras del gen *APM1*, son exploradas con respecto a la búsqueda de mutaciones. Preferiblemente, los individuos positivos con respecto al rasgo portan el haplotipo que se ha demostrado que está asociado con el rasgo, y los individuos negativos con respecto al rasgo no portan el haplotipo o el alelo asociado con el rasgo. El procedimiento de detección de mutaciones es esencialmente similar al procedimiento usado para identificar de sitios bialélicos.

- 30 El método usado para detectar esas mutaciones generalmente comprende las siguientes etapas de: (a) amplificar una región del gen candidato que comprende un marcador bialélico o un grupo de marcadores bialélicos asociados con ese rasgo, en muestras de DNA de pacientes positivos con respecto al rasgo y de controles negativos con respecto al rasgo; (b) secuenciar la región amplificada; (c) comparar secuencias de DNA procedentes de pacientes positivos con respecto al rasgo y de controles negativos con respecto al rasgo; y (d) determinar mutaciones específicas de los pacientes positivos con respecto al rasgo. Se contemplan de manera específica subcombinaciones que comprenden las etapas (b) y (c).

- 40 Se prefiere que los polimorfismos candidatos sean luego verificados sometiendo a escrutinio una población más grande de casos y de controles por medio de cualquiera de los procedimientos de genotipificación como los aquí descritos, preferiblemente usando una técnica de microsecuenciación en un formato de ensayo individual.

- Los polimorfismos se consideran como mutaciones candidatas cuando están presentes en los casos y en los controles en frecuencias compatibles con los resultados de asociación esperados.

Marcadores bialélicos de la invención en métodos de diagnóstico genético

- 50 Los marcadores bialélicos de la presente invención también se pueden usar para desarrollar ensayos diagnósticos capaces de identificar individuos que expresan un rasgo detectable como resultado de un genotipo específico, o individuos cuyo genotipo los pone en riesgo de manifestar un rasgo detectable en un momento posterior.

- 55 Los facultativos expertos en el tratamiento o diagnóstico de la obesidad y de trastornos relacionados con la obesidad entenderán sin duda que la presente invención no pretende proporcionar una identificación absoluta de individuos que podrían estar en riesgo de contraer una enfermedad particular en la que están implicados la obesidad y los trastornos relacionados con la obesidad, sino más bien indicar un cierto grado o probabilidad de contraer una enfermedad. Sin embargo, esta información es extremadamente valiosa ya que, en determinadas circunstancias, se puede usar para iniciar tratamientos preventivos o para permitir que un individuo que porta un haplotipo significativo vislumbre signos de alarma tales como síntomas menores. En enfermedades en las que los ataques pueden ser extremadamente violentos y a veces mortales si no se tratan a tiempo, el conocimiento de una predisposición potencial, incluso si esta predisposición no fuera absoluta, podría contribuir de manera importante a la eficacia del tratamiento.

- 65 Las técnicas diagnósticas de la presente invención pueden emplear diversas metodologías para determinar si un sujeto en estudio presenta un patrón de marcadores bialélicos asociado con un mayor riesgo de manifestar un rasgo detectable, o si ese individuo padece un rasgo detectable como resultado de una mutación particular, que incluyen métodos que permiten el análisis de los cromosomas del individuo con respecto a la determinación de haplotipos, tales como estudios genealógicos, análisis individuales del DNA de esperma o la obtención de híbridos somáticos. El rasgo analizado usando los presentes métodos diagnósticos puede ser cualquier rasgo detectable, incluyendo la obesidad y los trastornos relacionados con la obesidad.

Otro aspecto se refiere a un método para determinar si un individuo está en riesgo de manifestar un rasgo o si un individuo expresa un rasgo como consecuencia de poseer un alelo concreto causante del rasgo. También se describe un método para determinar si un individuo está en riesgo de manifestar una pluralidad de rasgos, o si un individuo expresa una pluralidad de rasgos como resultado de poseer un alelo concreto causante de esos rasgos. Estos métodos implican obtener una muestra de ácido nucleico del individuo y determinar si la muestra de ácido nucleico contiene uno o más de los alelos de uno o más de los marcadores bialélicos indicativos de un riesgo de manifestar el rasgo, o indicativos de que el individuo expresa el rasgo como resultado de poseer un alelo concreto causante del rasgo.

Según aquí se describe, los ensayos diagnósticos pueden estar basados en un único marcador bialélico o en un grupo de marcadores bialélicos.

Vectores recombinantes

El término “vector” se usa aquí para designar una molécula de DNA o RNA, circular o lineal, que es bicatenaria o monocatenaria, y que comprende al menos uno de los polinucleótidos de interés que se buscan para transferirlos a un hospedador celular o a un organismo hospedador unicelular o multicelular.

La presente invención incluye una familia de vectores recombinantes que comprenden un polinucleótido regulador obtenido de la secuencia genómica de *APM1*, o un polinucleótido codificador procedente de la secuencia genómica de *APM1*. Por consiguiente, la presente invención también trata sobre un vector recombinante que comprende, o un polinucleótido regulador comprendido en los ácidos nucleicos de los SEQ ID NOS: 2 y 3, o un polinucleótido que comprende la secuencia codificadora de *APM1*, o ambos.

En una primera realización preferida, un vector recombinante de la invención se usa para amplificar el polinucleótido insertado, obtenido de una secuencia genómica de *APM1*, seleccionada entre el grupo que consiste en los ácidos nucleicos de los SEQ ID NOS: 1, 2 y 3, o un cDNA de *APM1*, por ejemplo el cDNA del SEQ ID NO: 5, en un hospedador celular adecuado, amplificándose este polinucleótido cada vez que el vector recombinante se replica.

Generalmente, un vector recombinante de la invención puede comprender cualquiera de los polinucleótidos aquí descritos, que incluyen secuencias reguladoras y secuencias codificadoras, así como cualquiera de los cebadores o sondas de *APM1* según se han definido anteriormente.

Una segunda realización preferida de los vectores recombinante según la invención, consiste en vectores de expresión que comprenden, o un polinucleótido regulador, o un ácido nucleico codificador de la invención, o ambos. En algunas realizaciones, los vectores de expresión se utilizan para expresar el polipéptido APM1 que se puede luego purificar y, por ejemplo, usar en ensayos de escrutinio de ligandos, o usar como inmunógeno para generar anticuerpos específicos dirigidos contra la proteína APM1. En otras realizaciones, los vectores de expresión se usan para construir animales transgénicos y también para terapia génica. La expresión requiere que los vectores estén provistos de las señales adecuadas, incluyendo dichas señales diversos elementos reguladores, tales como elementos estimuladores/promotores originarios tanto de virus como de mamíferos, que dirigen la expresión de los genes de interés en las células hospedadoras. Marcadores dominantes de selección por fármacos, usados para establecer clones celulares permanentes y estables que expresan los productos, generalmente se incluyen en los vectores de expresión de la invención, cuando son elementos que conectan la expresión de los marcadores de selección de resistencia a un fármaco, con la expresión del polipéptido.

Más en particular, la presente invención se refiere a vectores de expresión que incluyen ácidos nucleicos que codifican una proteína APM1, preferiblemente la proteína APM1 de la secuencia de aminoácidos del SEQ ID NO: 6, o sus variantes, o sus fragmentos, bajo el control de una secuencia reguladora seleccionada entre los polinucleótidos reguladores de *APM1*, o como alternativa, bajo el control de una secuencia reguladora exógena.

Por consiguiente, los vectores de expresión preferidos de la invención se seleccionan entre el grupo que consisten en: (a) la secuencia reguladora de *APM1* comprendida en dicho gen, dirige la expresión de un polinucleótido codificador operativamente ligado a ella; (b) la secuencia codificadora de *APM1* está operativamente ligada a secuencias de regulación que permiten su expresión en un hospedador celular adecuado y/o en un organismo hospedador.

Un vector de expresión recombinante que comprende un ácido nucleico seleccionado entre el grupo que consiste en el SEQ ID NO: 2, o sus fragmentos o variantes biológicamente activos, forma también parte de la presente invención.

En una realización preferida, un vector de expresión recombinante de la invención comprende una secuencia de nucleótidos reguladora, seleccionada entre el grupo que consiste en:

(i) una secuencia de nucleótidos que comprende un polinucleótido del SEQ ID NO: 2, o una secuencia complementaria a ella;

(ii) una secuencia de nucleótidos que comprende un polinucleótido que tiene una identidad de nucleótidos de al menos el 95% con respecto a la secuencia de nucleótidos del SEQ ID NO: 2, o una secuencia complementaria a ella;

(iii) una secuencia de nucleótidos que comprende un polinucleótido que se hibrida en condiciones rigurosas de hibridación con la secuencia de nucleótidos del SEQ ID NO: 2, o una secuencia complementaria a ella; y

(iv) un fragmento o una variante biológicamente activos, de los polinucleótidos de (i), (ii) y (iii).

La invención también incluye un vector de expresión recombinante que comprende:

a) un ácido nucleico que comprende una secuencia de nucleótidos reguladora seleccionada entre el grupo que consiste en:

(i) una secuencia de nucleótidos que comprende un polinucleótido del SEQ ID NO: 2, o una secuencia complementaria a ella;

(ii) una secuencia de nucleótidos que comprende un polinucleótido que tiene una identidad de nucleótidos de al menos el 95% con respecto a la secuencia de nucleótidos del SEQ ID NO: 2, o una secuencia complementaria a ella;

(iii) una secuencia de nucleótidos que comprende un polinucleótido que se hibrida en condiciones rigurosas de hibridación con la secuencia de nucleótidos del SEQ ID NO: 2, o una secuencia complementaria a ella; y

(iv) un fragmento biológicamente activo, o una variante, de los polinucleótidos de (i), (ii) y (iii); y

b) un polinucleótido que codifica un polipéptido deseado o un ácido nucleico de interés, operativamente unido al ácido nucleico definido en el anterior apartado (a).

Adicionalmente, el vector de expresión recombinante anteriormente descrito puede también comprender un ácido nucleico que comprende un polinucleótido regulador en 3', preferiblemente un polinucleótido regulador en 3' del gen *APM1*. El polinucleótido regulador en 3' de *APM1* también puede comprender la secuencia 3'-UTR contenida en la secuencia de nucleótidos del SEQ ID NO: 5.

El polinucleótido regulador en 5' puede también incluir la secuencia 5'-UTR del cDNA de *APM1*, o uno de sus fragmentos o variantes biológicamente activos.

La invención también concierne a un vector de expresión recombinante útil para la expresión de la secuencia codificadora de *APM1*, en el que dicho vector comprende un ácido nucleico del SEQ ID NO: 5.

Otro vector de expresión recombinante preferido consiste en un vector de expresión de una secuencia codificadora de *APM1*, en el que dicho vector comprende un ácido nucleico del SEQ ID NO: 1, o uno de sus fragmentos, o un ácido nucleico que tiene una identidad de nucleótidos de al menos el 95% con un polinucleótido del SEQ ID NO: 1, o uno de sus fragmentos.

También forman parte de la invención vectores recombinantes que comprenden un ácido nucleico que contiene un marcador bialélico relacionado con *APM1*. En una realización preferida, dicho marcador bialélico se selecciona entre el grupo que consiste en A1, A2, A4, A7 y A8, y sus complementos.

Algunos de los elementos que se pueden encontrar en los vectores de la presente invención, se describen con más detalles en las secciones que vienen a continuación.

1. Características generales de los vectores de expresión de la invención

Un vector recombinante según la invención comprende, pero no se limita a, un YAC (cromosoma artificial de levaduras) (del inglés, "*Yeast Artificial Chromosome*"), un BAC (cromosoma artificial de bacterias), un fago, fagómido, cósmido, plásmido, o incluso una molécula de DNA lineal que puede consistir en un DNA cromosómico, no cromosómico, semisintético y sintético. Ese vector recombinante puede comprender una unidad transcripcional que comprende un ensamblaje de:

(1) un elemento o elementos genéticos que tienen un papel regulador en la expresión génica, por ejemplo, promotores o elementos estimuladores. Los elementos estimuladores son elementos de DNA que actúan en cis, normalmente de una longitud de aproximadamente 10 pb a 300 pb, que actúan sobre el promotor para aumentar la transcripción.

(2) una secuencia estructural o secuencia codificadora, que se transcribe en mRNA y que finalmente se traduce en un polipéptido, estando dicha secuencia estructural o codificadora operativamente unida a los elementos reguladores descritos en el apartado (1); y

(3) secuencias adecuadas de iniciación y terminación de la transcripción. Las unidades estructurales pensadas para usar en sistemas de expresión eucarióticos o de levaduras, preferiblemente incluyen una secuencia conductora que facilita la secreción extracelular de la proteína traducida por parte de una célula hospedadora. Como alternativa, cuando una proteína recombinante se expresa sin una secuencia conductora o de transporte, puede incluir un residuo N-terminal. Este residuo puede ser o no posteriormente escindido de la proteína recombinante expresada para proporcionar un producto final.

En general, los vectores de expresión recombinantes incluirán orígenes de replicación, marcadores de selección que permiten la transformación de la célula hospedadora y un promotor obtenido de un gen de alta capacidad de expresión, que dirigir la transcripción de una secuencia estructural situada aguas abajo. La secuencia estructural heteróloga se ensambla en la fase adecuada con secuencias de iniciación y terminación de la traducción, y preferiblemente con una secuencia conductora capaz de dirigir la secreción de la proteína traducida hacia el espacio periplásmico o el medio extracelular. En una realización específica en la que el vector se adapta para transfectar células hospedadoras de mamífero y expresar las secuencias deseadas en ellas, los vectores preferidos comprenderán un origen de replicación en el hospedador conveniente, un promotor adecuado y un elemento estimulador, y también todos los sitios necesarios de unión a ribosoma, sitio de poliadenilación, sitios dadores y aceptores de corte y empalme, secuencias de terminación de la transcripción y secuencias flanqueantes en 5' que no se transcriben. También se pueden usar secuencias de DNA obtenidas del genoma vírico de SV40, por ejemplo, el origen, promotor temprano, elemento estimulador, sitios de corte y empalme y sitios de poliadenilación de SV40, para proporcionar los elementos genéticos requeridos que no se transcriben.

La expresión *in vivo* de un polipéptido APM1 del SEQ ID NO: 6, o de sus fragmentos o variantes, puede ser útil con el fin de corregir un defecto genético relacionado con la expresión del gen natural en un organismo hospedador, o para la producción de una proteína APM1 biológicamente activa.

Por consiguiente, también se describen vectores de expresión recombinantes diseñados principalmente para la producción *in vivo* del polipéptido APM1 del SEQ ID NO: 6, o sus fragmentos o variantes, mediante la introducción del material genético adecuado en el organismo del paciente que ha de ser tratado. Este material genético se puede introducir *in vitro* en una célula que ha sido previamente extraída del organismo, siendo esa célula modificada reintroducida posteriormente en dicho organismo, directamente *in vivo* en el tejido adecuado.

2. Elementos reguladores

Promotores

Las regiones de promotor adecuadas, usadas en los vectores de expresión según la presente invención se eligen teniendo en cuenta la célula hospedadora en la que el gen heterólogo ha de ser expresado. El promotor particular utilizado para controlar la expresión de una secuencia de ácido nucleico de interés no se considera que sea importante siempre y cuando sea capaz de dirigir la expresión de ese ácido nucleico en la célula diana. Por lo tanto, cuando una célula humana es la célula diana, es preferible situar la región codificadora del ácido nucleico adyacente a, y bajo el control de, un promotor que es capaz de expresarse en una célula humana, como por ejemplo, un promotor humano o un promotor vírico.

Un promotor adecuado puede ser un promotor heterólogo con respecto al ácido nucleico cuya expresión controla o, como alternativa, puede ser un promotor endógeno en relación con el polinucleótido natural que contiene la secuencia codificadora que ha de ser expresada. De manera adicional, el promotor generalmente es heterólogo con respecto a las secuencias del vector recombinante dentro de las cuales se ha de insertar la construcción promotor/secuencia codificadora.

Las regiones de promotor se pueden seleccionar entre las de cualquiera de los genes convenientes usando, por ejemplo, vectores CAT (cloranfenicol-transferasa) y más preferiblemente los vectores pKK232-8 y pCM7.

Los promotores bacterianos preferidos son los promotores LacI, LacZ, los promotores de la RNA-polimerasa del bacteriófago T3 o T7, los promotores gpt, lambda PR, PL trp (documento EP 0036776), el promotor de la polihedrina o el promotor de la proteína p10 procedentes de baculovirus (Kit Novagen) (Smith y col., 1983; O'Reilly y col., 1992), el promotor PR del fago lambda o también el promotor trc.

Los promotores de células eucariotas incluyen el promotor temprano inmediato de CMV, el promotor de la timidina-quinasa de HSV, los promotores temprano y tardío de SV40, las LTR procedentes de retrovirus y el promotor de la metalotioneína L de ratón. La selección de un vector y un promotor convenientes está dentro del nivel normal de experiencia en la técnica.

La elección de un promotor está dentro de la capacidad de una persona experta en el campo de la ingeniería genética (Sambrook y col. (1989) y Fuller y col., (1996)).

Otros elementos reguladores

En los casos en los que se utiliza un inserto de cDNA, típicamente será conveniente incluir una señal de poliadenilación para efectuar una correcta poliadenilación del transcrito génico. La naturaleza de la señal de poliadenilación no se considera que sea crucial para la correcta práctica de la invención, y se puede utilizar alguna de esas señales tales como las señales de poliadenilación de la hormona de crecimiento humana y del SV40. También se contempla a un terminador como elemento de la casete de expresión. Estos elementos pueden servir para aumentar los niveles del mensaje y para minimizar la lectura del principio a fin de la casete que da lugar a otras secuencias.

El vector que contiene la secuencia de DNA correcta según se ha descrito anteriormente, más preferiblemente un polinucleótido regulador del gen *APM1*, un polinucleótido que codifica el polipéptido APM1 seleccionado entre el grupo que consiste en el SEQ ID NO: 1, o uno de sus fragmentos o variantes, y el SEQ ID NO: 5, o ambos de ellos, se pueden utilizar para transformar un hospedador adecuado que permita la expresión del polipéptido o polinucleótido deseado.

3. Marcadores de selección

Los marcadores de selección confieren a la célula un cambio identificable que permite una fácil identificación de las células que contienen la construcción de expresión. Los genes de marcadores de selección usados para la selección de las células hospedadoras transformadas, son preferiblemente el gen de la dihidrofolato-reductasa o el gen de resistencia a neomicina para cultivos de células eucariotas, el gen *TRP1* para *S. cerevisiae*, o el gen de resistencia a tetraciclina, rifampicina o ampicilina en *E. coli*, o el gen de la leva-sacarasa para micobacterias, siendo este último marcador un marcador de selección negativa.

4. Vectores preferidos

Vectores bacterianos

Como ejemplo representativo pero no limitativo, los vectores de expresión útiles para usar en bacterias pueden comprender un marcador de selección y un origen de replicación en bacterias obtenidos de plásmidos comercialmente disponibles que comprenden elementos genéticos del pBR322 (ATCC 37017). Esos vectores comerciales incluyen, por ejemplo, pKK223-3 (*Pharmacia*, Uppsala, Sweden) y GEM 1 (*Promega Biotec.*, Madison, WI, USA).

Los expertos en la técnica conocen una gran cantidad de otros vectores adecuados y comercialmente disponibles, tales como los siguientes vectores bacterianos: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, *phagescript*, psiX174, *pbluescript* SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (*Stratagene*); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (*Pharmacia*); pWLNEO, pSV2CAT, pOG44, pXTI, pSG (*Stratagene*); pSVK3, pBPV, pMSG, pSVL (*Pharmacia*); pQE-30 (*QIAexpress*).

Vectores bacteriofágicos

El vector bacteriofágico P1 puede contener insertos grandes que varían desde aproximadamente 80 kb hasta aproximadamente 100 kb.

La construcción de vectores bacteriofágicos P1, tales como p158 o p158/neo8, se describen de manera particular en Sternberg (1992, 1994). Los clones recombinantes de P1 que comprenden secuencias de nucleótidos de *APM1*, se pueden diseñar para insertar en ellos polinucleótido grandes de más de 40 kb (Linton y col., 1993). Para generar DNA de P1 para experimentos transgénicos, un protocolo preferido es el protocolo descrito por McCormick y col., (1994). Brevemente expuesto, bacterias *E. coli* (preferiblemente la cepa NS3529) que albergan el plásmido P1, se hacen crecer durante la noche en un caldo de cultivo adecuado que contiene una concentración de kanamicina de 25 µg/ml. El DNA de P1 se prepara a partir de esa *E. coli* mediante lisis alcalina usando el kit *Qiagen Plasmid Maxi kit* (Qiagen, Chatsworth, CA, USA) conforme a las instrucciones del fabricante. El DNA de P1 se purifica a partir del lisado bacteriano en dos columnas Qiagen-tip 500, usando los tampones de lavado y de elución contenidos en el kit. A continuación se lleva a cabo una extracción con fenol/cloroformo antes de precipitar el DNA con etanol al 70%. Después de disolver el DNA en TE (Tris-HCl 10 mM, pH 7,4, EDTA 1 mM), la concentración del DNA se determina por espectrofotometría.

Para expresar un clon P1 que comprende secuencias de nucleótidos de *APM1*, en un animal transgénico, típicamente, en ratones transgénicos, es conveniente retirar secuencias del vector procedentes del fragmento de DNA de P1, por ejemplo, escindiendo el DNA de P1 en sitios de corte poco frecuentes incluidos en el fragmento adaptador de policlonación (*SfiI*, *NotI* o *Sall*) de P1. El inserto de P1 se purifica después de entre las secuencias del vector sobre un gel de agarosa en campo pulsante, usando métodos similares a los originariamente descritos para el aislamiento del DNA procedente de los YAC (Schedl y col., 1993a; Peterson y col., 1993). En esta fase, el DNA del inserto purificado resultante se puede concentrar, en caso necesario, en una unidad de filtración *Millipore Ultrafree-MC Filter Unit* (*Millipore*, Bedford, MA, USA, límite de peso molecular en 30.000) y después se puede dializar frente al tampón de

microinyección (Tris-HCl 10 mM, pH 7,4; EDTA 250 μ M) que contiene NaCl 100 mM, espermina 30 μ M, espermina 70 μ M en una membrana de microdiálisis (de tipo VS, 0,025 μ M, procedente de *Millipore*). La integridad del inserto de DNA de P1 purificado se determina mediante electroforesis sobre un gel de agarosa al 1% (Sea Kem GTG; FMC Bio-products) en campo pulsante y tinción con bromuro de etidio.

Vectores baculovíricos

Un vector adecuado para la expresión del polipéptido APM1 del SEQ ID NO: 6, o de sus fragmentos o variantes, es un vector baculovírico que se puede propagar en células de insectos y en líneas celulares de insectos. Un sistema de vector, específico para un hospedador adecuado, es el vector pVL1392/1393 de transferencia de baculovirus (*Pharming*) que se usa para transfectar la línea celular SF9 (ATCC Núm. CRL 1711) que deriva de *Spodoptera frugiperda*.

Otros vectores adecuados para la expresión del polipéptido APM1 del SEQ ID NO: 6, o de sus fragmentos o variantes, en un sistema de expresión en baculovirus, incluyen los vectores descritos por Chai y col., (1993), Vlasak y col. (1983) y Lenhard y col. (1996).

Vectores víricos

En una de las realizaciones específicas, el vector deriva de un adenovirus. Vectores adenovíricos preferidos según la invención son los descritos por Feldman y Steg (1996) o por Ohno y col. (1994). Otro adenovirus recombinante preferido según esta realización específica de la presente invención es el adenovirus de tipo 2 o de tipo 5 (Ad 2 o Ad 5) de seres humanos o un adenovirus de origen animal (solicitud de Patente francesa Núm. FR-93.05954).

Los vectores retrovíricos y los vectores del tipo de los virus adenoasociados generalmente se entiende que son los sistemas de suministro de genes recombinantes de elección para la transferencia de polinucleótidos exógenos *in vivo*, en particular a mamíferos, incluyendo seres humanos. Estos vectores proporcionan un suministro eficiente de genes a las células, y los ácidos nucleicos transferidos se integran de manera estable en el DNA cromosómico del hospedador.

Los retrovirus particularmente preferidos para la preparación o construcción de vehículos retrovíricos de suministro de genes *in vivo* o *in vitro*, de la presente invención, incluyen retrovirus seleccionados entre el grupo que consiste en el virus inductor de focos en células de visión, virus del sarcoma murino, virus de la reticuloendoteliosis y el virus del sarcoma de Rous. Los virus de la leucemia murina particularmente preferidos incluyen los virus 4070A y los virus 1504A, Abelson (ATCC Núm. VR-999), Friend (ATCC Núm. VR-245), Gross (ATCC Núm. VR-590), Rauscher (ATCC Núm. VR-998) y el virus de la leucemia murina de Moloney (ATCC Núm. VR-190; solicitud de PCT documento WO 94/24298). Los virus de sarcoma de Rous particularmente preferidos incluyen la cepa Bryan de título alto (ATCC Núms. VR-334, VR-657, VR-726, VR-659 y VR-728). Otros vectores retrovíricos preferidos son los descritos en Roth y col. (1996), solicitud de PCT documento WO 93/25234, solicitud de PCT documento WO 94/06920, Roux y col., 1989, Julian y col., 1992 y Neda y col., 1991.

Otro sistema más de vectores retrovíricos que es contemplado por la invención, consiste en el virus adenoasociado (AAV). El virus adenoasociado es un virus defectivo presente en la naturaleza, que requiere otro virus, tal como un adenovirus o un herpesvirus, como virus auxiliar para una replicación eficiente y un ciclo de vida productivo (Muzyczka y col., 1992). Es también uno de los pocos virus que puede integrar su DNA en células que no están en división, y exhibe una alta frecuencia de integración estable (Flotte y col., 1992; Samulski *et al.*, 1989; McLaughlin *et al.*, 1989). Una de las características ventajosas de los AAV deriva de su menor eficacia para transducir células primarias en comparación con su eficacia para transducir células transformadas.

Vectores BAC

Los sistemas de clonación basados en cromosomas artificiales de bacterias (BAC) (Shizuya y col., 1992) se ha desarrollado para mantener de manera estable fragmentos grandes de DNA genómico (de 100 kb-300 kb) en *E. coli*. Un vector BAC preferido consiste en el vector pBeloBAC11 que ha sido descrito por Kim y col. (1996). Las genotecas de BAC se preparan con este vector usando DNA genómico seleccionado por tamaños, que ha sido parcialmente digerido usando enzimas que permiten la ligación en los sitios *Bam*HI o en los sitios *Hind*III del vector. Flanqueando estos sitios de clonación hay sitios de iniciación a la transcripción de la RNA-polimerasa de T7 y SP6, que se pueden usar para generar sondas de los extremos mediante métodos de transcripción en RNA o métodos de PCR. Después de la construcción de una genoteca de BAC en *E. coli*, el DNA de los BAC se purifica procedente de la célula hospedadora en forma de un círculo superenrollado. La conversión de estas moléculas circulares en una forma lineal precede tanto a la determinación de los tamaños como a la introducción de los BAC en las células receptoras. El sitio de clonación está flanqueado por dos sitios *Not*I, que permiten que los segmentos clonados sean escindidos del vector mediante digestión con *Not*I. Como alternativa, el inserto de DNA contenido en el vector pBeloBAC11 se puede linearizar mediante tratamiento del vector BAC con la enzima terminasa del fago lambda, comercialmente disponible, que conduce a la escisión en el sitio único *cos*N, pero este método de escisión da como resultado un clon de BAC de longitud completa que contiene tanto el DNA del inserto como las secuencias del BAC.

5. Suministro de los vectores recombinantes

Con el fin de efectuar la expresión de los polinucleótidos y construcciones polinucleotídicas de la invención, estas construcciones se deben suministrar a las células. Este suministro se puede llevar a cabo *in vitro*, en forma de procedimientos de laboratorio para transformar líneas celulares, o se puede llevar a cabo *in vivo* o *ex vivo*, como sucede en el tratamiento de determinados estados patológicos.

Uno de los mecanismos es la infección vírica en los casos en los que la construcción de expresión se encuentra encapsulada en una partícula vírica infecciosa.

La presente invención también contempla varios métodos no víricos para la transferencia de polinucleótidos a células de mamífero en cultivo, e incluyen, sin que se esté limitado a ellos, precipitación con fosfato cálcico (Graham y col., 1973; Chen y col., 1987;), DEAE-dextrano (Gopal, 1985), electroporación (Tur-Kaspa y col., 1986; Potter y col., 1984), microinyección directa (Harland y col., 1985), liposomas cargados con DNA (Nicolau y col., 1982; Fraley y col., 1979) y transfección mediada por receptores (Wu y Wu, 1987; 1988). Algunas de estas técnicas se pueden adaptar satisfactoriamente para su uso *in vivo* o *ex vivo*.

Una vez que el polinucleótido de expresión ha sido suministrado a la célula, se puede integrar de manera estable en el genoma de la célula receptora. Esta integración puede tener lugar en la posición y orientación cognadas a través de una recombinación homóloga (reemplazamiento de genes) o se puede integrar en una posición no específica al azar (incremento génico). En otras realizaciones más, el ácido nucleico se puede mantener de manera estable en forma de un segmento episómico independiente de DNA. Esos segmentos de ácido nucleico o “episomas” codifican las secuencias suficientes para permitir el mantenimiento y replicación independientes de, o en sincronización con, el ciclo de la célula hospedadora.

Una de las realizaciones específicas de un método para suministrar *in vivo* una proteína o un péptido al interior de una célula de un vertebrado, comprende la etapa de introducir una preparación que comprende un vehículo fisiológicamente aceptable y un polinucleótido desnudo que codifica operativamente el polipéptido de interés, en el espacio intersticial de un tejido que comprende esa célula, por medio de lo cual el polinucleótido desnudo es absorbido hacia el interior de la célula y ejerce un efecto fisiológico. Esto es aplicable de manera particular en el caso de una transferencia *in vitro* pero se puede aplicar también a una transferencia *in vivo*.

Las composiciones para usar *in vitro* e *in vivo* que comprenden un polinucleótido “desnudo” se describen en la solicitud de PCT documento WO 90/11092 (*Vical Inc.*) y también en la solicitud de PCT documento WO 95/11307 (Instituto Pasteur, INSERM, Universidad de Ottawa) así como en los artículos Tacson y col. (1996) y de Huygen y col., (1996).

En otra realización más de la invención, la transferencia de un polinucleótido desnudo de la invención, que incluye una construcción polinucleotídica de la invención, al interior de las células, puede ir precedida por un bombardeo de partículas (bombardeo biolístico), siendo dichas partículas microproyectiles revestidos con DNA y acelerados a una velocidad elevada que les permite atravesar membranas celulares y penetrar en las células sin destruirlas, tal como se describe en Klein y col., 1987).

En una realización más, el polinucleótido de la invención puede encontrarse encerrado en un liposoma (Ghosh y Bacchawat, 1991; Wong y col., 1980; Nicolau y col., 1987).

En una realización específica, la invención proporciona una composición para la producción *in vivo* de la proteína o polipéptido APM1 aquí descrito. La composición comprende un polinucleótido desnudo que codifica operativamente ese polipéptido, disuelto en un vehículo fisiológicamente aceptable y adecuado para su introducción en un tejido, para hacer que las células de ese tejido expresen dicha proteína o polipéptido.

La cantidad de vector que se ha de inyectar al organismo hospedador deseado varía según el sitio de la inyección. Como dosis indicativa, se inyectarán entre 0,1 μg y 100 μg del vector en un organismo animal, preferiblemente en el organismo de un mamífero, por ejemplo en el organismo de un ratón.

En otra realización del vector según la invención, éste se puede introducir *in vitro* en una célula hospedadora, preferiblemente en una célula hospedadora previamente recogida del animal que ha de ser tratado, y más preferiblemente en una célula somática tal como una célula muscular. En una etapa posterior, la célula que ha sido transformada con el vector que codifica el polipéptido APM1 deseado, o el fragmento deseado del mismo, se reintroduce en el organismo del animal con el fin de suministrar la proteína recombinante al interior del organismo, ya sea por vía local o sistémica.

Células hospedadoras

Otro objetivo de la invención consiste en una célula hospedadora que ha sido transformada o transfectada con uno de los polinucleótidos descritos en la invención, y de manera más precisa, con un polinucleótido que comprende o un polinucleótido regulador de *APM1* o la secuencia codificadora del polipéptido APM1 seleccionada entre el grupo que consiste en el SEQ ID NO: 1, o uno de sus fragmentos, o una de sus variantes, y el SEQ ID NO: 5. Están

ES 2 308 852 T3

incluidas células hospedadoras transformadas (células procariotas) o transfectadas (células eucariotas) con un vector recombinante tal como uno de los anteriormente descritos.

Generalmente, una célula hospedadora recombinante de la invención comprende uno cualquiera de los polinucleótidos o de los vectores recombinantes descritos en la invención.

Una célula hospedadora recombinante preferida según la invención, comprende un polinucleótido seleccionado entre el siguiente grupo de polinucleótidos:

a) un ácido nucleico purificado o aislado que codifica un polipéptido APM1, o uno de sus fragmentos polipeptídicos o una de sus variantes;

b) un ácido nucleico purificado o aislado que comprende al menos 8, preferiblemente al menos 15, más preferiblemente al menos 25 nucleótidos consecutivos de una secuencia de nucleótidos seleccionada entre el grupo que consiste en:

1) la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 4.811 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella; más en particular, la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 3.529 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

2) la secuencia de nucleótidos que comienza en el nucleótido de la posición 4.852 y que termina en el nucleótido de la posición 15.142 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

3) la secuencia de nucleótidos que comienza en el nucleótido de la posición 15.366 y que termina en el nucleótido de la posición 16.276 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella; y

4) la secuencia de nucleótidos que comienza en el nucleótido de la posición 20.560 y que termina en el nucleótido de la posición 20.966 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

c) un ácido nucleico purificado o aislado que comprende al menos 8 nucleótidos consecutivos, preferiblemente al menos 15 nucleótidos consecutivos de la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 22 de la secuencia de nucleótidos del SEQ ID NO: 5, o una de sus variantes, o una secuencia complementaria a ella;

d) un ácido nucleico purificado o aislado que comprende un exón del gen *APM1*, una secuencia complementaria a él, o una de sus variantes;

e) un ácido nucleico purificado o aislado que comprende combinación de al menos dos exones del gen *APM1*, o las secuencias complementarias a ellos, en el que los polinucleótidos están ordenados dentro del ácido nucleico desde el extremo 5' hasta el extremo 3' de dicho ácido nucleico, en el mismo orden que en el SEQ ID NO: 1;

f) un ácido nucleico purificado o aislado que comprende la secuencia de nucleótidos del SEQ ID NO: 2, o las secuencias complementarias a ella, o uno de sus fragmentos o variantes biológicamente activos;

g) un ácido nucleico purificado o aislado que comprende la secuencia de nucleótidos del SEQ ID NO: 3, o la secuencia complementaria a ella, o uno de sus fragmentos o variantes biológicamente activos;

h) un polinucleótido que consiste en:

(1) un ácido nucleico que comprende un polinucleótido regulador del SEQ ID NO: 2, o las secuencias complementarias a él, o uno de sus fragmentos o variantes biológicamente activos;

(2) un polinucleótido que codifica un polipéptido o un ácido nucleico deseados; o

(3) opcionalmente, un ácido nucleico que comprende un polinucleótido regulador del SEQ ID NO: 3, o la secuencia complementaria a él, o uno de sus fragmentos o variantes biológicamente activos; y

i) una construcción de DNA como la previamente descrita en la presente memoria descriptiva.

Otro hospedador celular recombinante preferido, según la presente invención, se caracteriza porque su genoma o su transgenoma (que incluye cromosomas, plásmidos) es modificado por el ácido nucleico que codifica el polipéptido APM1 del SEQ ID NO: 5, o por sus fragmentos o variantes.

ES 2 308 852 T3

Otro hospedador celular recombinante según la invención, comprende un polinucleótido que contiene un marcador bialélico seleccionado entre el grupo que consiste en A1, A2, A3, A4, A5, A6, A7 y A8, y sus complementos.

Las células hospedadoras preferidas usadas como receptoras de los vectores de expresión de la invención son las siguientes:

- a) Células hospedadoras procariotas: Cepas de *Escherichia coli* (cepa I.E.DH5- α), *Bacillus subtilis*, *Salmonella typhimurium* y cepas procedentes de especies de géneros similares a *Pseudomonas*, *Streptomyces* y *Staphylococcus*;
- b) Células hospedadoras eucariotas: Células HeLa (ATCC Núm. CCL2; Núm. CCL2.1; Núm. CCL2.2), células Cv 1 (ATCC Núm. CCL70), células COS (ATCC Núm. CRL1650; Núm. CRL1651), células Sf-9 (ATCC Núm. CRL1711), células C127 (ATCC Núm. CRL-1804), células 3T3 (ATCC Núm. CRL-6361), células CHO (ATCC Núm. CCL-61), células 293 de riñón humano (ATCC Núm. 45504; Núm. CRL-1573) y células BHK (ECACC Núm. 84100501; Núm. 84111301); y
- c) otras células hospedadoras procedentes de mamíferos.

La expresión del gen *APM1* en mamíferos, y típicamente en células humanas, puede hacerse defectuosa o como alternativa, puede ir precedida de la inserción de una secuencia genómica o de cDNA de *APM1*, con el reemplazamiento de la parte equivalente del gen *APM1* en el genoma de una célula animal, con un polinucleótido *APM1* según la invención. Estas modificaciones genéticas se pueden generar por medio de eventos de recombinación homóloga usando las construcciones de DNA específicas que han sido previamente descritas.

Las células hospedadoras que se pueden usar incluyen cigotos de mamíferos, tales como cigotos murinos. Por ejemplo, los cigotos murinos pueden recibir una microinyección con una molécula de DNA purificada de interés, por ejemplo, una molécula de DNA purificada que se ha ajustado previamente a un intervalo de concentraciones desde 1 ng/ml, en el caso de los insertos de BAC, hasta 3 ng/ μ l, en el caso de los insertos del bacteriófago P1, en Tris-HCl 10 mM, pH 7,4, EDTA 250 μ M, que contiene NaCl 100 mM, espermina 30 μ M y espermidina 70 μ M. Cuando el DNA que ha de ser microinyectado tiene un tamaño grande, se pueden usar poliaminas y concentraciones elevadas de sales, con el fin de evitar la rotura mecánica de este DNA, según se describe en Schedl y col. (1993b).

Cualquiera de los polinucleótidos de la invención, incluyendo las construcciones de DNA aquí descritas, se puede introducir en una línea de células troncales embrionarias (ES) (del inglés, "*Embryonic Stem cell line*"), preferiblemente una línea de células ES de ratón. Las líneas de células ES derivan de células pluripotenciales no comprometidas de la masa celular interna de blastocistos en fase de preimplantación. Las líneas de células ES preferidas incluyen las siguientes líneas: ES-E14TG2a (ATCC Núm. CRL-1821), ES-D3 (ATCC Núm. CRL1934 y Núm. CRL-11632), YS001 (ATCC Núm. CRL-11776), 36,5 (ATCC Núm. CRL-11116). Para mantener las células ES en un estado no comprometido, se cultivan en presencia de células alimentadoras inhibidas en su crecimiento, que proporcionan las señales adecuadas para preservar este fenotipo embrionario y sirven como matriz para la adherencia de las células ES. Las células alimentadoras preferidas consisten en fibroblastos embrionarios primarios que se establecen a partir de tejido de embriones de 13 días-14 días de desarrollo, de prácticamente cualquiera de las cepas de ratones, que se mantienen en cultivo, tal como se describe en Abbondanzo y col. (1993), y cuyo crecimiento se ha inhibido mediante irradiación, tal como se describe en Robertson (1987), o mediante la presencia de una concentración inhibidora de LIF, tal como se describe en Pease y Williams (1990).

Las construcciones en las células hospedadoras se pueden usar de una manera convencional para producir el producto génico codificado por la secuencia recombinante.

Después de la transformación de un hospedador conveniente y del crecimiento de ese hospedador hasta alcanzar una densidad celular adecuada, el promotor seleccionado es inducido con medios adecuados, tales como un cambio de temperatura o una inducción química, y las células se cultivan durante un periodo de tiempo adicional.

Las células típicamente se recogen mediante centrifugación, se rompen por medios físicos o químicos, y el extracto en crudo resultante se conserva para su posterior purificación.

Las células microbianas utilizadas para la expresión de las proteínas se pueden romper por medio de cualquier método conveniente, que incluye ciclos de congelación-descongelación, aplicación de ultrasonidos, rotura mecánica o el uso de agentes que lisan las células. Esos métodos son muy conocidos por el experto en la técnica.

Animales transgénicos

Las expresiones "animales transgénicos" o "animales hospedadores" se usan aquí para designar a animales que tienen su genoma genéticamente y artificialmente manipulado a fin de incluir uno de los ácidos nucleicos según la invención. Los animales preferidos son mamíferos no humanos e incluyen los pertenecientes a un género seleccionado entre *Mus* (p. ej. ratones), *Rattus* (p. ej. ratas) y *Oryctogalus* (p. ej. conejos) que tienen su genoma genéticamente y artificialmente modificado por la inserción de un ácido nucleico según la invención. En una de las realizaciones, la invención incluye mamíferos hospedadores no humanos y animales que comprenden un vector recombinante de

ES 2 308 852 T3

la invención o un gen *APM1* alterado mediante recombinación homóloga con un vector *knock out* (en el que se ha eliminado o inactivado un gen endógeno).

Los animales transgénicos de la invención incluyen todos ellos, en una pluralidad de sus células, una secuencia clonada de DNA recombinante o sintético, más específicamente, uno de los ácidos nucleicos purificados o aislados que comprenden una secuencia codificadora de *APM1*, un polinucleótido regulador de *APM1* o una secuencia de DNA que codifica un polinucleótido antisentido, tal como se describen en la presente memoria descriptiva.

Los animales transgénicos preferidos según la invención contienen en sus células somáticas y/o en sus células germinales, un polinucleótido seleccionado entre el siguiente grupo de polinucleótidos:

a) un ácido nucleico purificado o aislado que codifica un polipéptido APM1, o uno de sus fragmentos polipeptídicos, o una de sus variantes;

b) un ácido nucleico purificado o aislado que comprende al menos 8, preferiblemente al menos 15, más preferiblemente al menos 25, nucleótidos consecutivos de una secuencia de nucleótidos seleccionada entre el grupo que consiste en:

1) la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 4.811 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella; más en particular, la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 3.529 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

2) la secuencia de nucleótidos que comienza en el nucleótido de la posición 4.852 y que termina en el nucleótido de la posición 15.142 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

3) la secuencia de nucleótidos que comienza en el nucleótido de la posición 15.366 y que termina en el nucleótido de la posición 16.276 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella; y

4) la secuencia de nucleótidos que comienza en el nucleótido de la posición 20.560 y que termina en el nucleótido de la posición 20.966 de la secuencia de nucleótidos del SEQ ID NO: 1, o una de sus variantes, o una secuencia complementaria a ella;

c) un ácido nucleico purificado o aislado que comprende al menos 8 nucleótidos consecutivos, preferiblemente al menos 15, de la secuencia de nucleótidos que comienza en el nucleótido de la posición 1 y que termina en el nucleótido de la posición 22 de la secuencia de nucleótidos del SEQ ID NO: 5, o una de sus variantes, o una secuencia complementaria a ella;

d) un ácido nucleico purificado o aislado que comprende un exón del gen *APM1*, una secuencia complementaria a él, o una de sus variantes;

e) un ácido nucleico purificado o aislado que comprende combinación de al menos dos exones del gen *APM1*, o las secuencias complementarias a ellos, en el que los polinucleótidos están ordenados dentro del ácido nucleico desde el extremo 5' hasta el extremo 3' de dicho ácido nucleico, en el mismo orden que en el SEQ ID NO: 1;

f) un ácido nucleico purificado o aislado que comprende la secuencia de nucleótidos del SEQ ID NO: 2, o las secuencias complementarias a ella, o uno de sus fragmentos o variantes biológicamente activos;

g) un ácido nucleico purificado o aislado que comprende la secuencia de nucleótidos del SEQ ID NO: 3, o la secuencia complementaria a ella, o uno de sus fragmentos o variantes biológicamente activos;

h) un polinucleótido que consiste en:

(1) un ácido nucleico que comprende un polinucleótido regulador del SEQ ID NO: 2, o las secuencias complementarias a él, o uno de sus fragmentos o variantes biológicamente activos;

(2) un polinucleótido que codifica un polipéptido o un ácido nucleico deseados; o

(3) opcionalmente, un ácido nucleico que comprende un polinucleótido regulador del SEQ ID NO: 3, o la secuencia complementaria a él, o uno de sus fragmentos y variantes biológicamente activos; y

i) una construcción de DNA como la previamente descrita en la presente memoria descriptiva.

Los animales transgénicos de la invención contienen por lo tanto secuencias específicas de material genético exógeno, tales como las secuencias de nucleótidos anteriormente descritas con detalle.

Otros animales transgénicos según la invención contienen en sus células somáticas y/o en sus células germinales, un polinucleótido que comprende un marcador bialélico seleccionado entre el grupo que consiste en A1, A2, A4, A7 y A8, y sus complementos.

En una primera realización preferida, estos animales transgénicos pueden ser buenos modelos experimentales con el fin de estudiar las diversas patologías relacionadas con la diferenciación celular, en particular las que conciernen a los animales transgénicos dentro de cuyo genoma se ha insertado una o varias copias de un polinucleótido que codifica un proteína APM1 natural, o como alternativa, una proteína APM1 mutante.

En una segunda realización preferida, estos animales transgénicos pueden expresar un polipéptido de interés deseado, bajo el control de los polinucleótidos reguladores del gen *APM1*, dando lugar a rendimientos buenos en la síntesis de esta proteína de interés y, finalmente, a una expresión específica de tejido de esta proteína de interés.

El diseño de los animales transgénicos de la invención puede hacerse conforme a las técnicas convencionales, muy conocidas por un experto en la técnica. Para más detalles en relación con la producción de animales transgénicos, y específicamente de ratones transgénicos, se puede hacer referencia a las Patentes de EE. UU. Núm. 4.873.191, expedida el 10 de octubre de 1.989, Núm. 5.464.764 expedida el 7 de noviembre de 1.995 y Núm. 5.789.215 expedida el 4 de agosto de 1.998, incorporándose aquí estos documentos como referencia en relación con la descripción de métodos que producen animales transgénicos.

Los animales transgénicos de la presente invención se producen mediante la aplicación de procedimientos que dan como resultado un animal con un genoma que lleva incorporado material genético exógeno. El procedimiento implica obtener el material genético, o una porción del mismo, que codifica o una secuencia codificadora de *APM1*, o un polinucleótido regulador de *APM1*, o una secuencia de DNA que codifica un polinucleótido *APM1* antisentido, tal como se describen en la presente memoria descriptiva.

Un polinucleótido recombinante de la invención se inserta en una línea de células embrionarias o en una línea de células troncales ES. La inserción se realiza preferiblemente usando electroporación, tal como se describe en Thomas y col. (1987). Las células sometidas a electroporación se someten a escrutinio (p. ej. mediante una selección a través de marcadores seleccionables, mediante PCR o mediante análisis de transferencia Southern) para encontrar las células positivas que han integrado en su genoma el polinucleótido recombinante exógeno, preferiblemente a través de un evento de recombinación homóloga. Un procedimiento ilustrativo de selección positiva-negativa que se puede usar conforme a la invención, se describe en Mansour y col. (1988).

A continuación, las células positivas se aíslan, se clonan y se inyectan en blastocistos de 3,5 días, procedentes de ratones, tal como se describe en Bradley (1987). Los blastocistos se insertan luego en un animal hospedador hembra y se dejan crecer hasta llegar a término.

Como alternativa, las células ES positivas se ponen en contacto con embriones en la fase de 8-16 células, 2,5 días (mórulas), tal como se describe en Wood y col. (1993) o en Nagy y col. (1993), siendo las células ES internalizadas para colonizar el blastocisto en toda su extensión, incluyendo las células que darán lugar a la línea germinal.

La descendencia de la hembra hospedadora se analiza para determinar que animales son transgénicos, p. ej. qué animales incluyen la secuencia de DNA exógeno insertada y qué animales son de tipo salvaje.

Por lo tanto, la presente invención también concierne a un animal transgénico que contiene un ácido nucleico, un vector de expresión recombinante o una célula hospedadora recombinante según la invención.

Líneas de células recombinantes derivadas de los animales transgénicos de la invención

Un objetivo más de la invención consiste en células hospedadoras recombinantes obtenidas de un animal transgénico aquí descrito. En una de las realizaciones, la invención incluye células derivadas de mamíferos y de animales hospedadores no humanos que comprenden un vector recombinante de la invención o un gen *APM1* alterado por una recombinación homóloga con un vector *knock out* (en el que se ha eliminado o inactivado un gen endógeno).

Las líneas de células recombinantes se pueden establecer *in vitro* a partir de células obtenidas de cualquier tejido de un animal transgénico según la invención, por ejemplo, mediante transfección de cultivos de células primarias con vectores que expresan genes *onc*, tales como el antígeno T largo de SV40, según se describe en Chou (1989) y Shay y col. (1991).

Método para producir un polipéptido APM1

En la actualidad es fácil producir proteínas en grandes cantidades mediante técnicas de ingeniería genética por medio de vectores de expresión tales como plásmidos, fagos o fagómidos. El polinucleótido que codifica la proteína APM1 se inserta en un vector de expresión adecuado con el fin de producir *in vitro* el polipéptido de interés.

ES 2 308 852 T3

Por lo tanto, también se describe un método para producir una proteína APM1, y en especial un polipéptido del SEQ ID NO: 6, en el que dicho método comprende:

- 5 a) cultivar en un medio de cultivo adecuado, un hospedador celular previamente transformado o transfectado con el vector recombinante que comprende un ácido nucleico que codifica la proteína APM1;
- b) recoger el medio de cultivo así acondicionado o lisar el hospedador celular, por ejemplo mediante aplicación de ultrasonidos o mediante choque osmótico;
- 10 c) separar o purificar, a partir de dicho medio de cultivo, o a partir del sedimento del lisado de las células hospedadoras resultante, el polipéptido de interés de esta manera producido; y
- d) opcionalmente caracterizar el polipéptido de interés producido.

15 En una realización específica del método anterior, el ácido nucleico que codifica la proteína APM1 se inserta en un vector adecuado, opcionalmente después de una escisión adecuada de este ácido nucleico amplificado con una o varias endonucleasas de restricción. En una realización preferida, el ácido nucleico que codifica la proteína APM1 se selecciona entre el grupo que consiste en el SEQ ID NO: 1, o uno de sus fragmentos, y el SEQ ID NO: 5. En otra
20 realización, el ácido nucleico que codifica la proteína APM1, comprende un alelo de al menos uno de los marcadores bialélicos de A1 a A8. El ácido nucleico que codifica la proteína APM1 puede ser el producto resultante de una reacción de amplificación en la que se usa una pareja de cebadores según la invención (mediante SDA, TAS, 3SR NASBA, TMA, etc.).

25 Los polipéptidos se pueden caracterizar por su unión a una columna de cromatografía de inmunoafinidad en la que se han inmovilizado previamente anticuerpos policlonales o monoclonales dirigidos contra un polipéptido del SEQ ID NO: 6.

30 Los polipéptido o péptidos así obtenidos se pueden purificar, por ejemplo, mediante cromatografía líquida de alta resolución, tal como una HPLC de fase inversa y/o una HPLC de intercambio catiónico, según se describe en Rougeot y col. (1994). La razón para preferir esta clase de purificación de péptidos o proteínas es la ausencia de productos secundarios encontrados en las muestras eluidas, que hace que la proteína o péptido purificado resultante sea más adecuado para un uso terapéutico.

35 *Método para realizar el escrutinio de sustancias que interaccionan con las secuencias reguladoras del gen APM1*

También se describe un método para realizar el escrutinio de sustancias o moléculas que son capaces de interaccionar con las secuencias reguladoras del gen *APM1*, tales como por ejemplo secuencias de promotor o de elementos reguladores.

40 Los ácidos nucleicos que codifican proteínas que son capaces de interaccionar con las secuencias reguladoras del gen *APM1*, más en particular, una secuencia de nucleótidos seleccionada entre el grupo que consiste en los polinucleótidos de los SEQ ID NOS: 2 y 3, o uno de sus fragmentos o variantes, y preferiblemente una variante que comprende uno de los marcadores bialélicos de la invención, se pueden identificar usando un sistema basado en la estrategia de un solo híbrido, tal como el descrito en el folleto incluido en el kit *Matchmaker One-Hybrid System kit* procedente de
45 *Clontech* (Ref. del catálogo Núm. K1603-1). Brevemente expuesto, la secuencia de nucleótidos diana se clona aguas arriba de una secuencia de reporte seleccionable y la construcción de DNA resultante se integra en el genoma de una levadura (*Saccharomyces cerevisiae*). Las células de levadura que contienen la secuencia de reporte en su genoma, se transforman después con una genoteca que consiste en moléculas de fusión entre los cDNA que codifican proteínas candidatas a unirse a las secuencias reguladoras del gen *APM1*, y secuencias que codifican el dominio activador de un
50 factor de transcripción de levaduras tal como GAL4. Las células de levadura recombinantes se siembran en placa en un caldo de cultivo para seleccionar las células que expresan la secuencia de reporte. Las células de levadura recombinantes así seleccionadas contienen una proteína de fusión que es capaz de unirse a la secuencia reguladora diana del gen *APM1*. A continuación los cDNA que codifican las proteínas de fusión se secuencian y se pueden clonar *in vitro* en vectores de expresión o de transcripción. La unión de los polipéptidos codificados a las secuencias reguladoras diana del gen *APM1* se puede confirmar mediante métodos familiares para el experto en la técnica, tales como los ensayos
55 de retardo en gel o los ensayos de protección contra DNasas.

Los ensayos de retardo en gel también se pueden realizar independientemente con el fin de escrutar moléculas candidatas que son capaces de interaccionar con las secuencias reguladoras del gen *APM1*, tal como se describe en
60 Fried y Crothers (1981), Garner y Revzin (1981) y Dent y Latchman (1993). Estas técnicas se basan en el principio según el cual un fragmento de DNA que está unido a una proteína, migra más despacio que ese mismo fragmento de DNA que no está unido. Brevemente expuesto, la secuencia de nucleótidos diana está marcada. Entonces la secuencia de nucleótidos diana marcada se pone en contacto o con un extracto nuclear completo procedente de células que contienen factores de transcripción, o con las diferentes moléculas candidatas que han de ser ensayadas. La interacción
65 entre la secuencia reguladora diana del gen *APM1* y la molécula candidata o el factor de transcripción se detecta después de una electroforesis en gel o de una electroforesis capilar por medio de un retardo en la migración.

Métodos para realizar el escrutinio de ligandos que modulan la expresión del gen APM1

También se describe un método para realizar el escrutinio de moléculas que regulan la expresión de la proteína APM1. Ese método de escrutinio comprende:

a) cultivar una célula procariota o una célula eucariota que ha sido transfectada con una secuencia de ácidos nucleicos que codifica la proteína APM1, o una de sus variantes, o uno de sus fragmentos, situada bajo el control de su propio promotor;

b) poner en contacto la célula cultivada con una molécula a ensayar; y

c) cuantificar la expresión de la proteína APM1, o de una de sus variantes, o de uno de sus fragmentos.

En una realización, la secuencia de nucleótidos que codifica la proteína APM1, o una de sus variantes, o uno de sus fragmentos, comprende un alelo de al menos uno de los marcadores bialélicos A1, A2, A3, A4, A5, A6, A7 y A8, y sus complementos.

Usando técnicas de recombinación de DNA muy conocidas por el experto en la técnica, la secuencia de DNA que codifica la proteína APM1 se inserta en un vector de expresión, aguas abajo desde la secuencia de su promotor. Como ejemplo ilustrativo, la secuencia del promotor del gen *APM1* está contenida en el ácido nucleico del SEQ ID NO: 2.

La cuantificación de la expresión de la proteína APM1 se puede realizar a nivel de mRNA o a nivel de proteína. En este último caso, se pueden usar anticuerpos policlonales o monoclonales para cuantificar las cantidades de la proteína APM1 que se han producido, por ejemplo, en un ensayo ELISA o en un ensayo RIA.

En una realización preferida, la cuantificación del mRNA de *APM1*, se realiza mediante una amplificación por PCR cuantitativa del cDNA obtenido por transcripción inversa del mRNA total de la célula hospedadora transfectada con *APM1* y cultivada, usando una pareja de cebadores específicos para *APM1*.

También se describe un método para realizar el escrutinio de sustancias o moléculas que son capaces de aumentar, o por el contrario, de disminuir, el nivel de expresión del gen *APM1*. Ese método puede permitir al experto en la técnica seleccionar sustancias que ejercen un efecto regulador sobre el nivel de expresión del gen *APM1* y que pueden ser útiles como componentes activos incluidos en composiciones farmacéuticas para tratar pacientes que padecen deficiencias en la regulación de la expresión del gen *APM1*, en particular de pacientes que padecen obesidad.

También se describe un método para realizar el escrutinio de una sustancia o una molécula candidata para modular la expresión del gen *APM1*, que comprende:

a) proporcionar una célula hospedadora recombinante que contiene un ácido nucleico, en el que dicho ácido nucleico comprende una secuencia de nucleótidos del SEQ ID NO: 2, o uno de sus fragmentos o variantes biológicamente activos, situada aguas arriba de un polinucleótido que codifica una proteína detectable.

b) obtener una sustancia candidata; y

c) determinar la capacidad de la sustancia candidata para modular los niveles de expresión del polinucleótido que codifica la proteína detectable.

En una realización específica, el ácido nucleico que comprende una secuencia de nucleótidos del SEQ ID NO: 2, o uno de sus fragmentos o variantes biológicamente activos, incluye un marcador bialélico seleccionado entre el grupo que consiste en A1, A2, A3, A4, A5, A6, A7 y A8, o sus complementos.

En una realización más, el ácido nucleico que comprende la secuencia de nucleótidos del SEQ ID NO: 2, o uno de sus fragmentos o variantes biológicamente activos, también incluye una región 5'UTR del cDNA de *APM1* del SEQ ID NO: 5, o uno de sus fragmentos o variantes biológicamente activos.

Entre los polinucleótidos preferidos que codifican una proteína detectable, se pueden citar polinucleótidos que codifican beta-galactosidasa, proteína fluorescente verde (GFP) (del inglés, "*Green Fluorescent Protein*") y cloranfenicol-acetil-transferasa (CAT).

La invención también concierne a kits útiles para llevar a cabo los métodos de escrutinio descritos en lo que antecede. Preferiblemente, esos kits comprenden un vector recombinante que permite la expresión de una secuencia de nucleótidos del SEQ ID NO: 2, o uno de sus fragmentos o variantes biológicamente activos, situada aguas arriba y operativamente ligada a un polinucleótido que codifica una proteína detectable o la proteína APM1 o uno de sus fragmentos o una de sus variantes.

En otra realización, un método para realizar el escrutinio de una sustancia o una molécula candidata para modular la expresión del gen *APM1*, comprende:

- 5 a) proporcionar una célula hospedadora recombinante que contiene un ácido nucleico, en el que dicho ácido nucleico comprende una secuencia 5'UTR del cDNA de *APM1* del SEQ ID NO: 5, o uno de sus fragmentos o variantes biológicamente activos, estando la secuencia 5'UTR o sus fragmentos o variantes biológicamente activos, operativamente ligados a un polinucleótido que codifica una proteína detectable;
- 10 b) obtener una sustancia candidata; y
- c) determinar la capacidad de la sustancia candidata para modular los niveles de expresión del polinucleótido que codifica la proteína detectable.

15 En una realización específica del anterior método de escrutinio, el ácido nucleico que comprende una secuencia de nucleótidos seleccionada entre el grupo que consiste en la secuencia 5'UTR del cDNA de *APM1* del SEQ ID NO: 5, o uno de sus fragmentos o variantes biológicamente activos, incluye una secuencia de promotor que es endógena con respecto a la secuencia 5'UTR de *APM1*.

20 En otra realización específica del anterior método de escrutinio, el ácido nucleico que comprende una secuencia de nucleótidos seleccionada entre el grupo que consiste en la secuencia 5'UTR del cDNA de *APM1* del SEQ ID NO: 5, o uno de sus fragmentos o variantes biológicamente activos, incluye una secuencia de promotor que es exógena con respecto a la secuencia 5'UTR de *APM1* allí definida.

25 En una realización preferida más, el ácido nucleico que comprende la secuencia 5'UTR del cDNA de *APM1* del SEQ ID NO: 5, o sus fragmentos biológicamente activos, incluye un marcador bialélico seleccionado entre el grupo que consiste en A1, A2, A4, A7 y A8, o sus complementos.

30 También se describe un kit para el escrutinio de una sustancia candidata que modula la expresión del gen *APM1*, en el que dicho kit comprende un vector recombinante que comprende un ácido nucleico que incluye una secuencia 5'UTR del cDNA de *APM1* del SEQ ID NO: 5, o uno de sus fragmentos o variantes biológicamente activos, estando esa secuencia 5'UTR o su fragmento o variante biológicamente activo, operativamente ligado a un polinucleótido que codifica una proteína detectable.

35 Para el diseño de vectores recombinantes adecuados, útiles para llevar a cabo los métodos de escrutinio anteriormente detallados, se hará referencia a la sección de la presente memoria descriptiva en la que se detallan los vectores recombinantes preferidos de la invención.

40 Los niveles de expresión y los patrones de *APM1* se pueden analizar mediante una hibridación en disolución con sondas largas según se describe en la solicitud de Patente internacional documento WO 97/05277, que por la presente se incorpora aquí como referencia en su totalidad, incluyendo todas las figuras, tablas o referencias. Brevemente expuesto, el cDNA de *APM1* o el DNA genómico de *APM1* anteriormente descritos, o sus fragmentos, se inserta en un sitio de clonación situado justo aguas abajo de un promotor de RNA-polimerasa de un bacteriófago (T3, T7 o SP6) para producir un RNA antisentido. Preferiblemente, el inserto de *APM1* comprende al menos 100 o más nucleótidos consecutivos de la secuencia de DNA genómico o de las secuencias de cDNA. El plásmido se lineriza y se transcribe en presencia de ribonucleótidos que comprenden ribonucleótidos modificados (es decir, biotina-UTP y DIG-UTP). Un exceso de este RNA doblemente marcado se hibrida en disolución con un mRNA aislado a partir de las células o los tejidos de interés. Las hibridaciones se realizan en condiciones estándar de rigurosidad (40°C -50°C, durante 16 horas, en un tampón de formamida al 80%, NaCl 0,4 M, pH 7-8). La sonda no hibridada se retira mediante digestión con ribonucleasas específicas de RNA monocatenario (es decir, RNases CL3, T1, Phy M, U2 o A). La presencia de la modificación con biotina-UTP permite la captura del híbrido sobre una placa de microtitulación revestida con estreptavidina. La presencia de la modificación con DIG permite detectar y cuantificar el híbrido mediante ELISA usando un anticuerpo anti-DIG acoplado a fosfatasa alcalina.

55 El análisis cuantitativo de la expresión del gen *APM1* se puede también realizar usando matrices. Según aquí se usa, el término "matriz" significa una ordenación unidimensional, bidimensional o multidimensional de una pluralidad de ácidos nucleicos de una longitud suficiente para permitir la detección específica de la expresión de los mRNA capaces de hibridarse a ellos. Por ejemplo, las matrices pueden contener una pluralidad de ácidos nucleicos obtenidos de genes cuyos niveles de expresión han de ser determinados. Las matrices pueden incluir el DNA genómico de *APM1*, las secuencias de cDNA de *APM1*, o las secuencias complementarias a ellas, o sus fragmentos, en particular las secuencias que comprenden al menos uno de los marcadores bialélicos según la presente invención, preferiblemente uno de los marcadores bialélicos del A1 al A8. Preferiblemente, los fragmentos tienen una longitud de al menos 15 nucleótidos. En otras realizaciones, los fragmentos tienen una longitud de al menos 25 nucleótidos. En algunas realizaciones, los fragmentos tienen una longitud de al menos 50 nucleótidos. Más preferiblemente, los fragmentos tienen una longitud de al menos 100 nucleótidos. En otra realización preferida, los fragmentos tienen una longitud de más de 100 nucleótidos. En algunas realizaciones los fragmentos pueden tener una longitud de más de 500 nucleótidos.

Por ejemplo, el análisis cuantitativo de la expresión del gen *APM1* se puede llevar a cabo con una micromatriz de DNA complementario de la manera descrita por Schena y col. (1.995 y 1.996). Los cDNA de longitud completa de *APM1*, o sus fragmentos, se amplifican mediante PCR y se ordenan tomándolos de una placa de microtitulación de 96 pocillos y disponiéndolos sobre portaobjetos de microscopio sililados, usando un robot de alta velocidad. Las matrices impresas se incuban en una cámara húmeda para permitir la rehidratación de los elementos de la matriz y se enjuagan, una vez en SDS al 0,2% durante 1 min, dos veces en agua durante 1 min y una vez durante 5 min en una disolución de borohidruro sódico. Las matrices se sumergen en agua durante 2 min a 95°C, se transfieren a SDS al 0,2% durante 1 min, se enjuagan dos veces con agua, se secan al aire y se almacenan en la oscuridad a 25°C.

El mRNA de células o de tejidos se aísla o se obtiene comercialmente y se preparan sondas mediante un ronda única de transcripción inversa. Las sondas se hibridan con micromatrices de 1 cm² bajo un cubreobjetos de vidrio de 14 mm x 14 mm, durante 6-12 horas, a 60°C. Las matrices se lavan durante 5 min a 25°C en un tampón de lavado de rigurosidad baja (SSC 1X/SDS al 2%), a continuación durante 10 min a temperatura ambiente en un tampón de lavado de rigurosidad alta (SSC 0,1X/SDS al 2%). Las matrices se exploran en SSC 0,1X usando un aparato de barrido láser de fluorescencia equipado con un conjunto de filtros hechos de encargo. Se obtienen medidas exactas de la expresión diferencial tomando el promedio de los cocientes de dos hibridaciones independientes.

El análisis cuantitativo de la expresión del gen *APM1* también se puede realizar con los cDNA de longitud completa de *APM1*, o con sus fragmentos, en matrices de DNA complementario, según se describe en Pietu y col. (1996). El cDNA de longitud completa de *APM1*, o sus fragmentos, se amplifica mediante PCR y se puntea sobre membranas. A continuación, mRNA originarios de diferentes tejidos o células se marcan con nucleótidos radiactivos. Después de la hibridación y los lavados en condiciones controladas, los mRNA hibridados se detectan mediante formación de imágenes por fosforescencia o autorradiografía. Se realizan experimentos por duplicado y se realiza después un análisis cuantitativo de los mRNA expresados de manera diferencial.

Como alternativa, el análisis de la expresión usando en DNA genómico de *APM1*, el cDNA de *APM1*, o sus fragmentos, se puede realizar por medio de matrices de nucleótidos de alta densidad, según se describe en Lockhart y col. (1996) y Sosnowsky y col. (1997). Oligonucleótidos de 15-50 nucleótidos procedentes de las secuencias de DNA genómico de *APM1*, de las secuencia de cDNA de *APM1*, en particular las que comprenden al menos uno de los marcadores bialélicos según la presente invención, preferiblemente al menos uno de los marcadores bialélicos seleccionados entre el grupo que consiste en de A1 a A8, o las secuencias complementarias a ellos, se sintetizan directamente sobre el chip (Lockhart y col., *supra*) o se sintetizan y después se remiten al chip (Sosnowski y col., *supra*). Preferiblemente, los oligonucleótidos tienen una longitud de aproximadamente 20 nucleótidos.

Se sintetizan sondas de cDNA de *APM1*, marcadas con un compuesto adecuado, tal como biotina, digoxigenina o un colorante fluorescente, a partir de una población de los mRNA adecuados, y después se fragmentan aleatoriamente hasta un tamaño promedio de 50 a 100 nucleótidos. Dichas sondas se hibridan después con el chip. Después de los lavados según se describe en Lockhart y col., *supra*, y de la aplicación de diferentes campos eléctricos (Sosnowsky y col., 1997), los colorantes o los compuestos marcadores se detectan y se cuantifican. Se realizan hibridaciones por duplicado. Un análisis comparativo de la intensidad de la señal que se origina procedente de las sondas de cDNA que están sobre el mismo oligonucleótido diana en muestras de cDNA diferentes indica una expresión diferencial del mRNA de *APM1*.

Métodos para inhibir la expresión de un gen APM1

Otras composiciones terapéuticas comprenden ventajosamente un fragmento oligonucleotídico de la secuencia de ácido nucleico de *APM1*, como herramienta antisentido o como una herramienta formación de triple hélice que inhibe la expresión del correspondiente gen *APM1*. Un fragmento preferido de la secuencia de ácido nucleico de *APM1* comprende un alelo de al menos uno de los marcadores bialélicos de A1 a A8.

Estrategia antisentido

Los métodos preferidos que usan un polinucleótido antisentido son los procedimientos descritos por Sczakiel y col., (1995).

Preferiblemente, las herramientas antisentido se eligen entre los polinucleótidos (15 pb-200 pb de longitud) que son complementarios al extremo 5' del mRNA de *APM1*. En otra realización, se usa una combinación de diferentes polinucleótidos antisentido, que son complementarios a las diferentes partes del gen diana deseado.

Los polinucleótidos antisentido preferidos según la presente invención son complementarios a una secuencia de los mRNA de *APM1* que contiene el codón ATG de iniciación a la traducción o un sitio dador o un sitio aceptor de corte y empalme.

Los ácidos nucleicos antisentido deben tener una longitud y una temperatura de fusión suficientes para permitir la formación de un dúplex intracelular que tiene la estabilidad suficiente para inhibir la expresión del mRNA de *APM1* en el dúplex. Estrategias para diseñar ácidos nucleicos antisentido adecuados para usar en terapia génica se describen en Green y col., (1986) y en Izant y Weintraub, (1984), cuyas descripciones se incorporan aquí como referencia.

En algunas estrategias, se obtienen moléculas antisentido invirtiendo la dirección de la región que codifica *APM1*, con respecto a un promotor, con el fin de transcribir la cadena opuesta de la que normalmente se transcribe en la célula. Las moléculas antisentido se pueden transcribir usando sistemas de transcripción *in vitro*, tales como los que utilizan la polimerasa de T7 o de SP6 para generar el transcrito. Otra estrategia implica la transcripción *in vivo* de ácidos nucleicos antisentido de *APM1*, uniendo operativamente el DNA que contiene la secuencia antisentido a un promotor presente en un vector de expresión adecuado.

Como alternativa, estrategias antisentido adecuadas son las descritas por Rossi y col. (1991), en las solicitudes internacionales, documentos WO 94/23026, WO 95/04141, WO 92/18522 y en la solicitud de Patente europea Núm. EP 0 572 287 A2.

Una alternativa a la tecnología antisentido consiste en usar ribozimas que se unirán a una secuencia diana a través de su cola polinucleotídica complementaria, y que escindirán el correspondiente RNA hidrolizando su sitio diana (en concreto, las "ribozimas cabeza de martillo"). Brevemente expuesto, el ciclo simplificado de una ribozima cabeza de martillo consiste en: (1) la unión de una secuencia específica al RNA diana a través de secuencias antisentido complementarias; (2) la hidrólisis específica de sitio del motivo escindible de la cadena diana; y (3) la liberación de los productos de escisión, que da lugar a otro ciclo catalítico. De hecho, es ventajoso el uso de un polinucleótido antisentido de cadena larga (de al menos 30 bases de longitud) o de ribozimas con brazos antisentido largos. Un sistema de suministro preferido de ribozimas antisentido se consigue uniendo covalentemente estas ribozimas antisentido a grupos lipófilos, o mediante el uso de liposomas como vector conveniente. Las ribozimas antisentido preferidas se preparan según se describe en Sczakiel y col., (1995), incorporándose aquí como referencia los procedimientos de preparación específicos que se mencionan en dicho artículo.

Estrategia de la triple hélice

El DNA genómico de *APM1* se puede usar también para inhibir la expresión del gen *APM1* basándose en la formación de una triple hélice intracelular.

Los oligonucleótidos de una triple hélice se usan para inhibir la transcripción de un genoma. Son particularmente útiles para estudiar cambios en la actividad celular cuando ese cambio está asociado con un gen particular.

De manera similar, una porción del DNA genómico de *APM1* se puede usar para estudiar el efecto de la inhibición de la transcripción de *APM1* dentro de una célula. Tradicionalmente, las secuencias de homopurinas se consideraron las más útiles para las estrategias basadas en una triple hélice. Sin embargo, las secuencias de homopirimidinas también pueden inhibir la expresión génica. Esos oligonucleótidos con homopirimidinas se unen al surco mayor en secuencias de homopurinas:homopirimidinas. Por lo tanto, ambos tipos de secuencias del DNA genómico de *APM1* se contemplan dentro del alcance de esta invención.

Para llevar a cabo estrategias de terapia génica usando el enfoque basado en la doble hélice, las secuencias del DNA genómico de *APM1* primero se exploran para identificar extensiones de homopirimidinas u homopurinas de 10-mer a 20-mer que se podrían usar en las estrategias basadas en la triple hélice para inhibir la expresión de *APM1*. Después de la identificación de las extensiones de homopirimidinas u homopurinas candidatas, se evalúa su eficiencia en cuanto a inhibir la expresión de *APM1*, introduciendo cantidades variables de oligonucleótidos que contienen las secuencias candidatas en células de cultivos de tejidos que expresan el gen *APM1*.

Los oligonucleótidos se pueden introducir en las células usando diversos métodos conocidos por los expertos en la técnica, que incluyen pero que no se limitan a, precipitación con fosfato cálcico, DEAE-dextrano, electroporación, transfección mediada por liposomas o absorción natural.

Se realiza un seguimiento de las células tratadas para detectar una función celular alterada o a una expresión disminuida de *APM1*, usando técnicas tales como la transferencia Northern, ensayos de protección contra RNasa o estrategias basadas en la PCR, para hacer el seguimiento de los niveles de transcripción del gen *APM1* en células que han sido tratadas con el oligonucleótido.

Los oligonucleótidos que son efectivos en cuanto a inhibir la expresión génica en células en cultivo de tejidos, se pueden después introducir *in vivo* usando las técnicas anteriormente descritas en la estrategia antisentido, en una dosificación calculada sobre la base de los resultados *in vitro*, según se ha descrito en la estrategia antisentido.

En algunas realizaciones, los anómeros naturales (beta) de las unidades oligonucleotídicas se pueden reemplazar por anómeros alfa para hacer al oligonucleótido más resistente a las nucleasas. Además, un agente intercalante tal como bromuro de etidio, o un agente similar, se puede unir al extremo 3' del oligonucleótido alfa para estabilizar la triple hélice. Para obtener información sobre la generación de oligonucleótidos adecuados para la formación de una triple hélice, véase Griffin *et al.* (1989).

Ejemplos**Ejemplo 1**5 *Identificación de marcadores bialélicos-Extracción de DNA*

Los individuos dadores fueron individuos no emparentados y sanos. Presentaban una diversidad suficiente para ser representativos de una población francesa heterogénea. Se extrajo el DNA de 100 individuos y se sometió a un ensayo de detección de marcadores bialélicos.

10

Se extrajeron 30 ml de sangre venosa periférica de cada uno de los individuos dadores, en presencia de EDTA. Se reunieron las células (sedimento) después de una centrifugación durante 10 minutos a 2.000 rpm. Los glóbulos rojos se lisaron con una disolución de lisis (volumen final 50 ml: Tris 10 mM pH 7,6; MgCl₂ 5 mM; NaCl 10 mM). La disolución se centrifugó (10 minutos, 2.000 rpm) tantas veces como fueron necesarias para eliminar los glóbulos rojos residuales presentes en el sobrenadante, después de la resuspensión del sedimento en la disolución de lisis.

15

El sedimento de glóbulos blancos se lisó durante toda la noche a 42°C con 3,7 ml de una disolución de lisis compuesta por:

20

- 3 ml de TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM)/NaCl 0,4 M

- 200 µl de SDS al 10%

- 500 µl de proteinasa K (2 mg proteinasa K en TE 10-2/NaCl 0,4 M).

25

Para la extracción de las proteínas, se añadió 1 ml de una disolución saturada de NaCl (6 M) (1/3,5 v/v). Después de una agitación vigorosa, la disolución se centrifugó durante 20 minutos a 10.000 rpm.

30

Para la precipitación del DNA, al sobrenadante anterior se añadieron de 2 a 3 volúmenes de etanol al 100% y la disolución se centrifugó durante 30 minutos a 2.000 rpm. La disolución de DNA se enjuagó tres veces con etanol al 70% para eliminar las sales y se centrifugó durante 20 minutos a 2.000 rpm. El sedimento se secó a 37°C y se resuspendió en 1 ml de TE 10-1 o en 1 ml de agua. La concentración de DNA se evaluó midiendo la OD a 260 nm (1 unidad de OD = una concentración de DNA de 50 µg/ml).

35

Para determinar la presencia de proteínas en la disolución de DNA, se determinó el cociente OD 260/OD 280. Solamente las preparaciones de DNA que tenían un cociente OD 260 / OD 280 entre 1,8 y 2 se usaron en los posteriores ejemplos que se describen a continuación.

40

El material reunido se constituyó mezclando cantidades equivalentes de DNA procedente de cada uno de los individuos.

Ejemplo 245 *Identificación de marcadores bialélicos: Amplificación de DNA genómico mediante PCR*

La amplificación de secuencias genómicas específicas de las muestras de DNA del ejemplo 1 se llevó a cabo en el material reunido de los DNA previamente obtenidos. Además, muestras de 50 individuos se amplificaron de manera similar.

50

Los ensayos de PCR se realizaron usando el siguiente protocolo:

55

Volumen final	25 µl
DNA	2 ng/µl
MgCl ₂	2 mM
dNTP (cada uno)	200 µM
Cebador (cada uno)	2,9 ng/µl
Ampli Taq Gold DNA-polimerasa	0,05 unidades/µl
Tampón para PCR (10X = TrisHCl 0,1 M pH 8,3 KCl 0,5 M)	1X

65

ES 2 308 852 T3

Cada pareja de primeros cebadores se diseñó usando la información sobre las secuencias del gen *APM1* aquí descrita y el programa informático OSP (Hillier y Green, 1991). Esta primera pareja de cebadores tenía una longitud de aproximadamente 20 nucleótidos y tenía la secuencia correspondiente a las posiciones del SEQ ID descritas en la Tabla 1 en las columnas designadas PU y RP.

TABLA 1

Amplificón	Posición del amplificón en el SEQ ID 1	PU	Posición de PU en el SEQ ID 1	RP	Posición de RP en el SEQ ID 1
9-27	3528-3946	B1	3528-3946	C1	3928-3946
9-28	3892-4321	B2	3892-3911	C2	4303-4321
99-14402	4155-4602	B3	4155-4175	C2	4584-4602
9-29	4223-4642	B4	4223-4242	C4	4623-4642
9-30	4599-5027	B5	4599-4618	C5	5008-5027
99-14387	10990-11442	B6	10990-11008	C6	11423-11442
99-14389	12472-12966	B7	12472-12491	C7	12946-12966
9-12	15073-15520	B8	15073-15092	C8	15503-15520
9-13	15131-15551	B9	15131-15150	C9	15532-15551
99-14405	15759-16211	B10	15759-15776	C10	16191-16211
9-14	16233-16652	B11	16233-16251	C11	16633-16652
9-15	16604-17025	B12	16604-16621	C12	17006-17025
9-16	16982-17402	B13	16982-17001	C13	17384-17402
9-17	17216-17517	B14	17216-17233	C14	17498-17517
9-18	17300-17503	B15	17300-17317	C15	17486-17503

Preferiblemente, los cebadores contenían una cola oligonucleotídica común aguas arriba de las bases específicas que son diana de la amplificación, y que fue útil en la secuenciación.

Los cebadores PU contienen la siguiente secuencia adicional PU 5':

TGTAAAACGACGGCCAGT; los cebadores RP contienen la siguiente secuencia RP 5':

CAGGAAACAGCTATGACC. El cebador que contiene la secuencia PU 5' adicional está listado en el SEQ ID NO: 7. El cebador que contiene la secuencia RP 5' adicional está listado en el SEQ ID NO: 8.

La síntesis de estos cebadores se llevó a cabo siguiendo el método de la fosforamida, en un sintetizador GENSET UFPS 24.1.

La amplificación de DNA se llevó a cabo en un termociclador Genius II. Tras calentar a 95°C durante 10 min, se realizaron 40 ciclos. Cada ciclo comprendió: 30 s a 95°C, 54°C durante 1 min y 30 s a 72°C. Para el alargamiento final, 10 min a 72°C pusieron fin a la amplificación. Las cantidades de los productos obtenidos en la amplificación se determinaron en placas de microtitulación de 96 pocillos, usando un fluorómetro y *Picogreen* como agente intercalante (*Molecular Probes*).

Ejemplo 3

Identificación de marcadores bialélicos. Secuenciación del DNA genómico amplificado e identificación de polimorfismos

La secuenciación del DNA amplificado, obtenido en el ejemplo 2, se llevó a cabo en secuenciadores ABI 377. Las secuencias de los productos resultantes de la amplificación se determinaron usando reacciones de secuenciación automática con terminadores dideoxi, con un protocolo de secuenciación por ciclos y terminadores unidos a un colorante. Los productos de las reacciones de secuenciación se hicieron correr sobre geles de secuenciación y las secuencias se determinaron usando un análisis de imágenes en gel (programa informático *ABI Prism DNA Sequencing Analysis* (versión 2.1.2) y el lector de nucleótidos patentado "Trace" anteriormente mencionado.

Los datos de las secuencias se evaluaron de nuevo usando el programa de análisis de polimorfismos anteriormente mencionado, diseñado para detectar la presencia de marcadores bialélicos entre la mezcla de fragmentos amplificados. La búsqueda de polimorfismos se basó en la presencia de picos superpuestos en el patrón electroforético, resultante de la existencia de diferentes bases en la misma posición, según se ha descrito previamente.

Se analizaron 15 fragmentos de la amplificación. En 5 de estos segmentos, se detectaron 8 marcadores bialélicos. La localización de estos marcadores bialélicos se muestra en la Tabla 2.

ES 2 308 852 T3

TABLA 2

Amplicón	Marcador bialélico	Nombre del marcador	Posición en el gen <i>APM1</i>	Polimorfismo	Posición del marcador en SEQ ID NO:1
9-27	A1	9-27/261	región reguladora en 5'	Alelo 1: G Alelo 2: C	3787
99-14387	A2	99-4387/129	Intrón 1	Alelo 1: A Alelo 2: C	11118
9-12	A3	9-12/48	Intrón 1	Alelo 1: T Alelo 2: C	15120
9-12 y 9-13	A4	9-12/124 ó 9-13/66	Exón 2	Alelo 1: T Alelo 2: G	15196
9-12 y 9-13	A5	9-12/355 ó 9-13/297	Intrón 2	Alelo 1: G Alelo 2: T	15427
9-12 y 9-13	A6	9-12/428 ó 9-13/370	Intrón 2	Alelo 1: A Alelo 2: G	15500
99-14405	A7	99-14405/105	Intrón 2	Alelo 1: G Alelo 2: A	15863
9-16	A8	9-16/189	Exón 3	Alelo 1: A Alelo 2: Del	17170

Ejemplo 4

Validación de los polimorfismos por medio de microsecuenciación

Los marcadores bialélicos identificados en el ejemplo 3 fueron de nuevo confirmados y sus respectivas frecuencias se determinaron por medio de microsecuenciación. La microsecuenciación se llevó a cabo en cada una de las muestras de DNA de los individuos, descritas en el Ejemplo 1.

La amplificación del DNA genómico de los individuos se realizó mediante PCR según se ha descrito anteriormente para la detección de los marcadores bialélicos, con el mismo conjunto de cebadores de PCR (Tabla 1).

Los cebadores preferidos usados en la microsecuenciación tenían una longitud de aproximadamente 19 nucleótidos y se hibridaron justo aguas arriba de la base polimórfica considerada. Según la invención, los cebadores usados en la microsecuenciación se detallan en la Tabla 3.

TABLA 3

Nombre del marcador	Marcador	Mis 1	Posición de mis 1 en SEQ ID NO:1	Mis 2	Posición de mis 2 en SEQ ID NO:1
9-27/261	A1	D1	3768-3786	E1	3788-3806
99-14387/129	A2	D2	11099-11117	E2	11119-11137
9-12/48	A3	D3	15101-15119	E3	15121-15139
9-12/124 ó 9-13/66	A4	D4	15177-15195	E4	15197-15215
9-12/355 ó 9-13/297	A5	D5	5408-15426	E5	15428-15446
9-12/428 ó 9-13/370	A6	D6	15481-15499	E6	15501-15519
99-14405/105	A7	D7	15844-15862	E7	15864-15882
9-16/189	A8	D8	17151-17169	E8	17171-17189

ES 2 308 852 T3

Mis 1 y Mis 2 se refieren a cebadores para microsecuenciación que se hibridan con la cadena no codificadora del gen *APM1* y con la cadena codificadora del gen *APM1* respectivamente.

La reacción de microsecuenciación se realizó de la siguiente manera:

Después de la purificación de los productos resultantes de la amplificación, la mezcla de reacción de microsecuenciación se preparó añadiendo, en un volumen final de 20 μ l, 10 pmol del oligonucleótido objeto de la microsecuenciación, 1 U de *Thermosequenase* (Amersham E79000G), 1,25 μ l de tampón para *Thermosequenase* (Tris-HCl 260 mM pH 9,5, MgCl₂ 65 mM) y los dos ddNTP fluorescentes adecuados (*Perkin Elmer*; *Dye Terminator Set* 401095) complementarios a los nucleótidos situados en el sitio polimórfico de cada uno de los marcadores bialélicos sometidos al ensayo, siguiendo las recomendaciones del fabricante. Después de 4 minutos a 94°C, se llevaron a cabo 20 ciclos de PCR, de 15 s a 55°C, 5 s a 72°C y 10 s a 94°, en un termociclador Tetrad PTC-225 (*MJ Research*). Los terminadores unidos al colorante no incorporados se retiraron mediante precipitación en etanol. Por último, las muestras se resuspendieron en un tampón de carga de formamida-EDTA y se calentaron durante 2 min a 95°C antes de ser cargadas en un gel de poliacrilamida para secuenciación. Los datos se recogieron en un secuenciador de DNA ABI PRISM 377 y se procesaron usando el programa informático GENESCAN (*Perkin Elmer*).

Después del análisis en gel, los datos se procesaron automáticamente con un programa informático que permite la determinación de los alelos de los marcadores bialélicos presentes en cada uno de los fragmentos amplificados.

El programa informático evalúa factores como el si las intensidades de las señales resultantes de los procedimientos de microsecuenciación anteriores son débiles, normales o saturadas, o si las señales son ambiguas. Además, el programa informático identifica picos significativos (conforme a criterios de la forma y altura). Entre los picos significativos, los picos que corresponden al sitio diana se identifican por su posición. Cuando dos picos significativos se detectan en la misma posición, cada una de las muestras se clasifica como de tipo homocigoto o de tipo heterocigoto basándose en el cociente de las alturas.

Se pueden usar sondas oligonucleotídicas en la genotipificación de marcadores bialélicos mediante ensayos de hibridación. La muestra de ácido nucleico se pone en contacto con una o más sondas oligonucleotídicas específicas de alelo, que se hibridan específicamente con uno o más alelos asociados con un fenotipo detectable. Las sondas son de 25-mer, y llevan un marcador bialélico relacionado con *APM1* en la posición central. Las sondas usadas en el ensayo de hibridación pueden incluir las sondas listadas en la Tabla 4.

TABLA 4

BM	Nombre del marcador	Intervalo de posiciones de las sondas en un SEQ ID no genómico		Sondas
A1	9-27/261	3775	3799	P1
A2	99-14387/129	11106	11130	P2
A3	9-12/48	15108	15132	P3
A4	9-12/124	15184	15208	P4
A5	9-12/355	15415	15439	P5
A6	9-12/428	15488	15512	P6
A7	99-14405/105	15851	15875	P7
A8	9-16/189	17158	17182	P8

Ejemplo 5

Preparación de composiciones de anticuerpos contra la variante 56-Glu de *APM1*

Se aísla la proteína o polipéptido sustancialmente puro a partir de células transfectadas o transformadas que contienen un vector de expresión que codifica la proteína *APM1* o una de sus porciones. La concentración de proteína en la preparación final se ajusta, por ejemplo, mediante concentración en un aparato de filtración de Amicon, hasta alcanzar el nivel de unos cuantos microgramos/ml. Un anticuerpo monoclonal o policlonal dirigido contra la proteína se puede preparar entonces de la siguiente manera:

A. Producción de un anticuerpo monoclonal mediante fusión en hibridomas

Se puede preparar un anticuerpo monoclonal dirigido contra epítomos de la proteína APM1, o contra una porción de la misma, a partir de hibridomas murinos conforme al método clásico de Kohler, G. y Milstein, C., *Nature* 256: 495 (1975), o conforme a métodos derivados del mismo. Véase también Harlow, E. y D. Lane. 1988. *Antibodies. A Laboratory Manual. Cold Spring Harbor Laboratory*. págs. 53-242.

Brevemente expuesto, un ratón se inocula repetidamente con unos pocos microgramos de la proteína APM1 o de una de sus porciones, a lo largo de un período de unas cuantas semanas. El ratón es después sacrificado y se aíslan las células del bazo. Las células de bazo se fusionan por medio de polietilenglicol con células de mieloma de ratón, y el exceso de células no fusionadas se destruye mediante cultivo del sistema en un medio selectivo que comprende aminopterina (medio HAT). Las células correctamente fusionadas se diluyen y partes alícuotas de la dilución se ponen en los pocillos de una placa de microtitulación en donde se continúa el crecimiento del cultivo. Los clones productores de anticuerpo se identifican mediante detección del anticuerpo en el fluido sobrenadante de los pocillos mediante procedimientos de inmunoensayo, tales como ELISA, de la manera descrita en un principio por Engvall, E., *Meth. Enzymol.* 70: 419 (1980), y mediante métodos derivados del mismo. Los clones positivos seleccionados se pueden expandir y el anticuerpo monoclonal que producen se puede recoger para su utilización. Procedimientos detallados para la producción de anticuerpos monoclonales se encuentran descritos en Davis, L. y col. *Basic Methods in Molecular Biology*, Elsevier, Nueva York. Sección 21-2.

B. Producción de anticuerpos policlonales mediante inmunización

Un antisuero policlonal que contiene anticuerpos dirigidos contra epítomos heterogéneos de la proteína APM1 o de una de sus porciones, se puede preparar inmunizando animales no humanos adecuados con la proteína APM1 o con una de sus porciones, que puede estar no modificada o modificada para aumentar su inmunogenicidad. Un animal no humano adecuado es preferiblemente un mamífero no humano seleccionado, normalmente un ratón, rata, conejo, cabra o caballo. Como alternativa, para generar anticuerpos se puede usar una preparación en crudo que ha sido enriquecida en concentración de APM1. Esas proteínas, fragmentos o preparaciones se introducen en el mamífero no humano en presencia de un adyuvante adecuado (p. ej. hidróxido de aluminio, RIBI, etc.) conocido en la técnica. Además, la proteína, el fragmento o la preparación se puede pretratar con un agente que incrementa la antigenicidad, agentes que son conocidos en la técnica y que incluyen, por ejemplo, albúmina de suero bovino metilada (mBSA), albúmina de suero bovino (BSA), antígeno de superficie de la hepatitis B y hemocianina de lapa californiana (KLH) (del inglés, "*Keyhole Limpet Hemocyanin*"). El suero del animal inmunizado se recoge, se trata y se ensaya conforme a procedimientos conocidos. Si el suero contiene anticuerpos policlonales contra epítomos no deseados, los anticuerpos policlonales se pueden purificar mediante cromatografía de inmunoafinidad.

La producción eficaz de anticuerpos policlonales se ve afectada por muchos factores relacionados tanto con el antígeno como con la especie hospedadora. Asimismo, la respuesta de los animales hospedadores varía dependiendo del sitio de las inoculaciones y de la dosis, produciendo unas dosis de antígeno tanto inadecuadas como excesivas un antisuero de título bajo. Dosis bajas (del orden de ng) de antígeno, administradas en múltiples sitios por vía intradérmica parecen ser más fiables. En la técnica se conocen métodos para producir y procesar antisueros policlonales, véase por ejemplo, Mayer y Walker (1987). Un protocolo de inmunización eficaz para aplicar en conejos se puede encontrar en Vaitukaitis, J. y col. *J. Clin. Endocrinol. Metab.* 33: 988-991 (1971).

Se pueden administrar inyecciones de refuerzo a intervalos regulares y el antisuero se puede recoger cuando su título de anticuerpo, determinado de manera semicuantitativa, por ejemplo, mediante inmunodifusión doble en agar frente a concentraciones conocidas del antígeno, comienza a decaer. Véase, por ejemplo, Ouchterlony, O. y col., Capítulo 19, en: *Handbook of Experimental Immunology*, D. Wier (redactor) Blackwell (1973). La concentración de estabilización del anticuerpo en la meseta está normalmente en el intervalo de 0,1 mg/ml a 0,2 mg/ml de suero (aproximadamente 12 μ M). La afinidad del antisuero por el antígeno se determina preparando curvas de unión competitiva, según se describe, por ejemplo, en Fisher, D., Capítulo 42, en: *Manual of Clinical Immunology*, 2ª Ed. (Rose y Friedman, redactores), *Amer. Soc. For Microbiol.*, Washington, D. C. (1980).

Las preparaciones de anticuerpo realizadas conforme al protocolo monoclonal o al protocolo policlonal son útiles en inmunoensayos cuantitativos que determinan concentraciones de sustancias que portan antígenos en muestras biológicas; también se usan de manera semicuantitativa o cualitativa para identificar la presencia de un antígeno en una muestra biológica. Los anticuerpos también se pueden usar en composiciones terapéuticas para destruir las células que expresan la proteína o para disminuir los niveles de la proteína en el organismo.

Aunque se han ilustrado y descrito las realizaciones preferidas de la invención, se tendrá en cuenta que un experto en la técnica puede realizar diferentes cambios en ella sin apartarse del espíritu y alcance de la invención.

Referencias

Abbondanzo S.J. et al. (1993) *Methods in Enzymology*, Academic Press, New York pp. 803-823.

Ajioka R.S. et al. (1997) *Am. J. Hum. Genet.* 60:1439-1447.

Anton M. (1995) *et al.*, *J. Virol.* 69: 4600-4606.

Araki K et al. (1995) *Proc. Natl. Acad. Sci. USA.* 92(1):1604.

Ausubel et al. (1989) *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y.

Bates G.P. et al. (1997a) *Hum. Mol. Genet.* 6(10):633-1637.

Bates GP et al. (1997b) *Molecular Medicine today*, 508:515.

Baubonis W. (1993) *Nucleic Acids Res.* 21(9):2025-9.

Beaucage et al. (1981) *Tetrahedron Lett.* 22:1859-1862.

Bradley A., (1987) Production and analysis of chimeric mice. In: E.J. Robertson (Ed.), *Teratocarcinomas and embryonic stem cells: A practical approach*. IRL Press, Oxford, pp. 113.

Brown E.L., Belagaje R., Ryan M.J., Khorana H.G. (1979) *Methods Enzymol.* 68:109-151.

Burright et al. (1997) *Brain Pathology.* 7:965-977.

Chai H. et al. (1993) *Biotechnol Appl. Biochem.* 18:259-273.

Chee et al. (1996) *Science.* 274:610-614.

Chen et al. (1987) *Mol. Cell. Biol.* 7:2745-2752.

Chen and Kwok (1997) *Nucleic Acids Research.* 25:347-353.

Chen et al. (1997) *Proc. Natl. Acad. Sci. USA.* 94(20):10756-10761.

Chou J.Y. (1989) *Mol. Endocrinol.* 3:1511-1514.

Clark A.G. (1990) *Mol. Biol. Evol.* 7:111-122.

Coles R., Caswell R., and Rubinsztein D.C. (1998) *Hum. Mol. Genet.* 7(5):791-800.

Compton J. (1991) *Nature.* 350(6313):91-92.

Davies S.W., Turmaine M., Cozens B.A., DiFiglia M., Sharp A.H., Ross C.A., Scherzinger E., Feldman and Steg. (1996) *Medecine/Sciences.* 12:47-55.

Dempster et al., (1977) *J. R. Stat. Soc.,* 39B:1-38.

Dent D.S. and Latchman D.S. (1993) The DNA mobility shift assay. In: *Transcription Factors: A Practical Approach* (Latchman DS, ed.) Oxford: IRL Press. ppl-26.

Eckner R. et al. (1991) *EMBO J.* 10:3513-3522.

Excoffier L. and Slatkin M. (1995) *Mol. Biol. Evol.,* 12(5): 921-927.

Flotte et al. (1992) *Am. J. Respir. Cell Mol. Biol.* 7:349-356.

Fodor et al. (1991) *Science* 251:767-777.

Fraley et al. (1979) *Proc. Natl. Acad Sci. USA.* 76:3348-3352.

Fried M. and Crothers D.M. (1981) *Nucleic Acids Res.* 9:6505-6525.

Fuller S. A. et al. (1996) *Immunology in Current Protocols in Molecular Biology*, Ausubel et al. Eds, John Wiley & Sons, Inc., USA.

Furth P.A. et al. (1994) *Proc. Natl. Acad Sci USA.* 91:9302-9306.

Garner M.M. and Revzin A. (1981) *Nucleic Acids Res* 9:3047-3060.

Ghosh and Bacchawat (1991) *Targeting of liposomes to hepatocytes, IN: Liver Diseases, Targeted diagnosis and therapy using specific receptors and ligands.* Wu *et al.*Eds., Marcel Dekeker, New York, pp. 87-104.

Gopal (1985) *Mol. Cell. Biol.*, 5:1188-1190.

Gossen M. et al. (1992) *Proc. Natl. Acad. Sci. USA*, 89:5547-5551.

Gossen M. et al. (1995) *Science*. 268:1766-1769.

Graham et al. (1973) *Virology* 52:456-457.

Green et al. (1986) *Ann. Rev. Biochem.* 55:569-597.

Griffin et al. (1989) *Science*. 245:967-971.

Grompe, M. et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:5855-5892.

Grompe, M. (1993) *Nature Genetics*. 5:111-117.

Gu H. et al. (1993) *Cell* 73:1155-1164.

Gu H. et al. (1994) *Science* 265:103-106.

Guatelli JC et al. Proc. Natl. Acad. Sci. USA. 35:273-286.

Gura. (1997) *Science* 275:751.

Hacia J.G., et al. (1996) *Nat. Genet.* 14(4):441-447.

Haff L. A. and Smirnov I. P. (1997) *Genome Research*, 7:378-388.

Haines B.D. and Higgins S.J. (1985) *Nucleic Acid Hybridization: A Practical Approach.* Hames and Higgins Ed., IRL Press, Oxford.

Harju L. et al. (1993) *Clin Chem.*, 39(11 Pt 1) 2282-2287.

Harland et al. (1985) *J. Cell. Biol.* 101:1094-1095.

Hawley M.E. et al. (1994) *Am. J. Ayr. Anthropol.* 18:104.

Hillier L. and Green P. (1991) *Methods Appl.* 1:124-8.

Hoess et al. (1986) *Nucleic Acids Res.* 14:2287-2300.

Hu E., Liang P., and Spiegelman B.M. (1996) *J. Biol. Chem.* 271:10697-10703.

Huang L. et al. (1996) *Cancer Res* 56(5): 1137-1141.

Huygen et al. (1996) *Nature Medicine*. 2(8):893-898.

Izant J.G. and Weintraub H. (1984) *Cell* 36(4):1007-1015.

Julan et al. (1992) *J. Gen. Tirol.* 73:3251-3255.

Kanegae Y. et al., Nucl. Acids Res. 23:3816-3821.

Khoury J. et al. (1993) *Fundamentals of Genetic Epidemiology*, Oxford University Press, NY.

Kim U -J. et al. (1996) *Genomics* 34:213-218.

Klein et al. (1987) *Nature*. 327:70-73.

Koller et al. (1992) *Annu. Rev. Immunol.* 10:705-730.

Kopp M.U., Mello A.J., Manz A., (1998) *Science*. 280(5366):1046-1048.

Kozal M.J. et al. (1996) *Nat. Med.* 2(7):753-759.

Landegren U. et al. (1998) *Genome Research*, 8:769-776.

Lander and Schork (1994) *Science*. 265:2037-2048.

Lange K. (1997) *Mathematical and Statistical Methods for Genetic Analysis*. Springer, New York.

Lenhard T. et al. (1996) *Gene*. 169:187-190.

Lin M.W. et al. (1997) *Hum. Genet.* 99(3): 417-420.

Linton M.F. et al. (1993) *J. Clin. Invest.* 92:3029-3037.

Liu Z. et al. (1994) *Proc. Natl. Acad. Sci. USA*. 91: 4528-4262.

Livak K.J. and Hainer J.W. (1994) *Hum. Mutat.* 3(4):379-385.

Lockhart et al. (1996) *Nature Biotechnology* 14:1675-1680.

Mackey K., Steinkarnp A., and Chomczynski P. (1998) *Mol Biotechnol.* 9(1):1-5.

Maeda et al. (1996) *Biochem. Biophys. Res. Comm.* 221:286-289.

Mangiarini L., Sathasivam K., Mahal A., Mott R., Seller M., and Bates G.P. (1997) *Nat. Genet.* 15(2):197-200.

Mansour S.L. et al. (1988) *Nature*. 336:348-352.

Manz et al. (1993) *Adv. in Chromatogr.* 33:1-66.

Marshall R. L. et al. (1994) *PCR Methods and Applications*. 4:80-84.

McCormick et al. (1994) *Genet. Anal. Tech. Appl.* 11:158-164.

McLaughlin BA. et al. (1996) *Am. J. Hunt Genet.* 59:561-569.

Montague et al. (1997) *Nature*. 387:903.

Morton N.E. (1955) *Am. J. Hum. Genet.* 7:277-318.

Muzyczka et al. (1992) *Curr. Topics in Micro. and Immunol.* 158:97-129.

Nada S. et al. (1993) *Cell* 73:1125-1135.

Nagy A. et al. (1993) *Proc. Natl. Acad. Sci. USA*. 90: 8424-8428.

Narang S.A., Hsiung H.M. (1979) *Brousseau R., Methods Enzymol.* 68:90-98.

Neda et al. (1991) *J. Biol. Chem.* 266:14143-14146.

Newton et al. (1989) *Nucleic Acids Res.* 17:2503-2516.

Nickerson DA. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927.

Nicolau et al. (1982) *Biochem. Biophys. Acta.* 721:185-190.

Nyren P. et al. (1993) *Anal. Biochem.* 208(1):171-175.

O'Reilly et al. (1992) *Baculovirus Expression Vectors: A Laboratory Manual*. W. H. Freeman and Co., New York.

Olmo et al. (1994) *Science*. 265:781-784.

Orita et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86: 2776-2770.

Ott J. (1991) *Analysis of Human Genetic Linkage*. John Hopkins University Press, Baltimore.

Pastinen et al. (1997) *Genome Research*. 7:606-614.

Pease S. and William RS. (1990) *Exp. Cell Res.* 190:09-211.

Perlin et al. (1994) *Am. J. Hum. Genet.* 55:777-787.

Peterson et al. (1993) *Proc. Natl. Acad. Sci. USA.* 90: 7593-7597.

Pietu et al. (1996) *Genome Research.* 6:492-503.

Potter et al. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81(22):7161-7165.

Reid L.H. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:4299-4303.

Risch, N. and Merilcangas, K. (1996) *Science.* 273:1516-1517.

Robertson E. (1987) "Embryo-Derived Stem Cell Lines." In: E.J. Robertson Ed. *Teratocarcinomas And Embryonic Stem Cells: A Practical Approach.* IRL Press, Oxford, pp. 71.

Rossi et al. (1991) *Pharmacol. Ther.* 50:245-254.

Roth J.A. et al. (1996) *Nature Medicine.* 2(9):985-991.

Rougeot, C. et al. *Eur. J. Biochem.* 219(3):765-773.

Roux et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:9079-9083.

Ruano et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:6296-6300.

Sambrook, J., Fritsch, E.F., and T. Maniatis. (1989) *Molecular Cloning: A Laboratory Manual.* 2ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Samson M, et al. (1996) *Nature*, 382(6593):722-725.

Samulski et al. (1989) *J. Tirol.* 63:3822-3828.

Sanchez-Pescador R. (1988) *J. Clin. Microbiol.* 26(10):1934-1938.

Sandou et al. (1994) *Science.* 265:1875-1878.

Sarkar, G. and Sommer S.S. (1991) *Biotechniques.*

Sauer B. et al. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:5166-5170.

Schaid D.J. et al. (1996) *Genet. Epidemiol.* 13:423-450.

Schedl A. et al. (1993a) *Nature.* 362:258-261.

Schedl et al. (1993b) *Nucleic Acids Res.* 21:4783-4787.

Schena et al. (1995) *Science.* 270:467-470.

Schena et al. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93(20):10614-10619.

Schneider et al. (1997) *Arlequin: A Software For Population Genetics Data Analysis.* University of Geneva.

Sczakiel G. et al. (1995) *Trends Microbiol.* 3(6):213-217.

Shay J.W. et al. (1991) *Biochem. Biophys. Acta.* 1072:1-7.

Sheffield, V.C. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 49:699-706.

Shizuya et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:8794-8797.

Shoemaker D.D. et al. (1996) *Nat. Genet.* 14:450-456.

Smith (1957) *Ann. Hum. Genet.* 21:254-276.

Smith et al. (1983) *Mol. Cell. Biol.* 3:2156-2165.

Sosnowski ILG. et al. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1119-1123.

ES 2 308 852 T3

Spielmann S. et al. (1993) *Am. J. Hum. Genet.* 52:506-516.

Spielmann S. and Ewens W. J. (1998) *Am. J. Hum. Genet.* 62:450-458.

Sternberg N.L. (1992) *Trends Genet.* 8:1-16.

Sternberg N.L. (1994) *Mamm. Genome.* 5:397-404.

Syvanen A.C. et al. (1994) *Clin. Chem. Acta.* 226(2):225-236.

Tacson et al. (1996) *Nature Medicine.* 2(8):888-892.

Te Riele et al. (1990) *Nature.* 348:649-651.

Terwilliger J.D. and Ott J. (1994) *Handbook of Human Genetic Linkage.* John Hopkins University Press, London.

Thomas K.R. et al. (1986) *Cell.* 44:419-428.

Thomas K.R. et al. (1987) *Cell.* 51:503-512.

Tur-Kaspa et al. (1986) *Mol. Cell. Biol.* 6:716-718.

Tyagi et al. (1998) *Nature Biotechnology.* 16:49-53.

Urdea M.S. (1988) *Nucleic Acids Research.* 11:4937-4957.

Urdea M.S. et al. (1991) *Nucleic Acids Symp. Ser.* 24:197-200.

Van der **Lugt et al.** (1991) *Gene.* 105:263-267.

Vlasak R. et al. (1983) *Eur. J. Biochem.* 135:123-126.

Wabiko et al. (1986) *DNA.* 5(4):305-314.

Walker et al. (1996) *Clin. Chem.* 42:9-13.

Wanker E.E., Mangiarini L., and Bates G.P. (1997) *Cell.* 90(3):537-48 .

Weir, B.S. (1996) *Genetic data Analysis IL' Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, U.S.A.*

White, M.B. et al. (1992) *Genomics.* 12:301-306.

White, M.B. et al. (1997) *Genomics.* 12:301-306.

Wong et al. (1980) *Gene.* 10:87-94.

Wood S.A. et al. (1993) *Proc. Natl. Acad. Sci. USA.* 90:4582-4585.

Wu and Wu (1987) *J. Biol. Chem.* 262:4429-4432.

Wu and Wu (1988) *Biochemistry.* 27:887-892.

Wu et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:2757.

Yagi T. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:9918-9922.

Zhao et al. (1998) *Am. J. Hum. Genet.* 63:225-240.

Zou Y. R. et al. (1994) *Curr. Biol.* 4:1099-1103.

ES 2 308 852 T3

Texto libre del listado de secuencias

El siguiente texto libre aparece en el Listado de secuencias que se adjunta:

5 región reguladora en 5'
 región reguladora en 3'
 base polimórfica
10 o
 complemento
15 sonda
 homología con una 5'-EST de un banco privado
 potencial
20 cebador oligonucleotídico PU de secuenciación
 cebador oligonucleotídico RP de secuenciación
25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Un polinucleótido aislado, purificado o recombinante que comprende la secuencia de nucleótidos del SEQ ID NO: 1 o su complemento.
2. Un polinucleótido aislado, purificado o recombinante que comprende la secuencia de nucleótidos del SEQ ID NO: 5 o su complemento.
3. Un polinucleótido aislado, purificado o recombinante que consiste en un tramo contiguo de 8 a 50 nucleótidos del SEQ ID NO: 1 o su complemento, en el que dicho tramo incluye un marcador bialélico relacionado con APM1 en dicha secuencia,
en el que dicho marcador bialélico relacionado con APM1 se selecciona entre el grupo que consiste en A1 (SEQ ID NO: 1, posición 3.787), A2 (SEQ ID NO: 1, posición 11.118), A4 (SEQ ID NO: 1, posición 15.196), A7 (SEQ ID NO: 1, posición 15.863) y A8 (SEQ ID NO: 1, posición 17.170).
4. Un polinucleótido según la reivindicación 3, en el que dicho tramo contiguo tiene una longitud de 18 a 35 nucleótidos y dicho marcador bialélico está a una distancia de 4 nucleótidos del centro de dicho polinucleótido.
5. Un polinucleótido según la reivindicación 4, en el que dicho polinucleótido consiste en dicho tramo contiguo y dicho tramo contiguo tiene una longitud de 25 nucleótidos y dicho marcador bialélico está en el centro de dicho polinucleótido.
6. Un polinucleótido según la reivindicación 3, en el que el extremo 3' de dicho tramo contiguo está situado en el extremo 3' de dicho polinucleótido y dicho marcador bialélico se encuentra en el extremo 3' de dicho polinucleótido.
7. Un polinucleótido según la reivindicación 3, en el que el extremo 3' de dicho tramo contiguo está situado en el extremo 3' de dicho polinucleótido y en el que dicho marcador bialélico presente en dicha secuencia está situado a una distancia de 20 nucleótidos aguas arriba del extremo 3' de dicho polinucleótido.
8. Un polinucleótido según la reivindicación 7, en el que dicho marcador bialélico presente en dicha secuencia está situado 1 nucleótido aguas arriba del extremo 3' de dicho polinucleótido.
9. El uso de un polinucleótido de cualquiera de las reivindicaciones de 3 a 8,
 - (i) en un ensayo de hibridación para determinar la identidad del nucleótido situado en un marcador bialélico relacionado con APM1 en el SEQ ID NO: 1 o su complemento;
 - (ii) en un ensayo de secuenciación para determinar la identidad del nucleótido situado en un marcador bialélico relacionado con APM1 en el SEQ ID NO: 1 o su complemento;
 - (iii) en un ensayo de amplificación específica de alelo para determinar la identidad del nucleótido situado en un marcador bialélico relacionado con APM1 en el SEQ ID NO: 1 o su complemento; o
 - (iv) en la amplificación de un segmento de nucleótidos que comprende un marcador bialélico relacionado con APM1 en el SEQ ID NO: 1 o su complemento,
en el que dicho marcador bialélico relacionado con APM1 se selecciona entre el grupo que consiste en A1 (SEQ ID NO: 1, posición 3.787), A2 (SEQ ID NO: 1, posición 11.118), A4 (SEQ ID NO: 1, posición 15.196), A7 (SEQ ID NO: 1, posición 15.863) y A8 (SEQ ID NO: 1, posición 17.170).
10. Un vector recombinante que comprende un polinucleótido según las reivindicaciones 1 ó 2.
11. Una célula hospedadora, o un animal o mamífero hospedador no humano, que comprende el vector recombinante según la reivindicación 10.
12. Un método de genotipificación que comprende determinar la identidad de un nucleótido situado en un marcador bialélico relacionado con APM1 del SEQ ID NO: 1, o su complemento, en una muestra biológica,
en el que dicho marcador bialélico relacionado con APM1 se selecciona entre el grupo que consiste en A1 (SEQ ID NO: 1, posición 3.787), A2 (SEQ ID NO: 1, posición 11.118), A4 (SEQ ID NO: 1, posición 15.196), A7 (SEQ ID NO: 1, posición 15.863) y A8 (SEQ ID NO: 1, posición 17.170).
13. Un método según la reivindicación 12, en el que dicha muestra biológica se obtiene procedente de un único sujeto.

ES 2 308 852 T3

14. Un método según la reivindicación 13, en el que la identidad de los nucleótidos situados en dicho marcador bialélico se determina en ambas copias de dicho marcador bialélico presente en el genoma de dicho individuo.

5 15. Un método según la reivindicación 12, en el que dicha muestra biológica se obtiene procedente de muchos sujetos.

16. Un método según la reivindicación 12, que además comprende amplificar una porción de dicha secuencia que comprende el marcador bialélico, antes de dicha etapa de determinación.

10 17. Un método según la reivindicación 12, en el que dicha determinación se realiza mediante:

(i) un ensayo de hibridación;

15 (ii) un ensayo de secuenciación;

(iii) un ensayo de microsecuenciación; o

(iv) un ensayo de amplificación específica de alelo.

20 18. Un método para detectar una asociación entre un genotipo y un fenotipo, que comprende las etapas de:

a) genotipificar al menos uno de los marcadores bialélicos relacionados con APM1 en una población positiva con respecto a un rasgo, según el método de la reivindicación 12;

25 b) genotipificar dicho marcador bialélico relacionado con APM1 en una población de control, según el método de la reivindicación 12; y

c) determinar si existe una asociación estadísticamente significativa entre dicho genotipo y dicho fenotipo.

30 19. Un método para estimar la frecuencia de un haplotipo correspondiente a un conjunto de marcadores bialélicos en una población, que comprende:

a) genotipificar al menos uno de los marcadores bialélicos relacionados con APM1 según la reivindicación 13, en cada uno de los individuos de dicha población;

35 b) genotipificar un segundo marcador bialélico para determinar la identidad de los nucleótidos situados en dicho segundo marcador bialélico, en ambas copias de dicho segundo marcador bialélico presente en el genoma de cada uno de los individuos de dicha población; y

40 c) aplicar un método de determinación de haplotipos a las identidades de los nucleótidos determinados en las etapas a) y b) para obtener una estimación de dicha frecuencia.

20. Un método para detectar una asociación entre un haplotipo y un fenotipo, que comprende las etapas de:

45 a) estimar la frecuencia de al menos uno de los haplotipos en una población positiva con respecto a un rasgo, según el método de la reivindicación 19;

b) estimar la frecuencia de dicho haplotipo en una población de control, según el método de la reivindicación 19; y

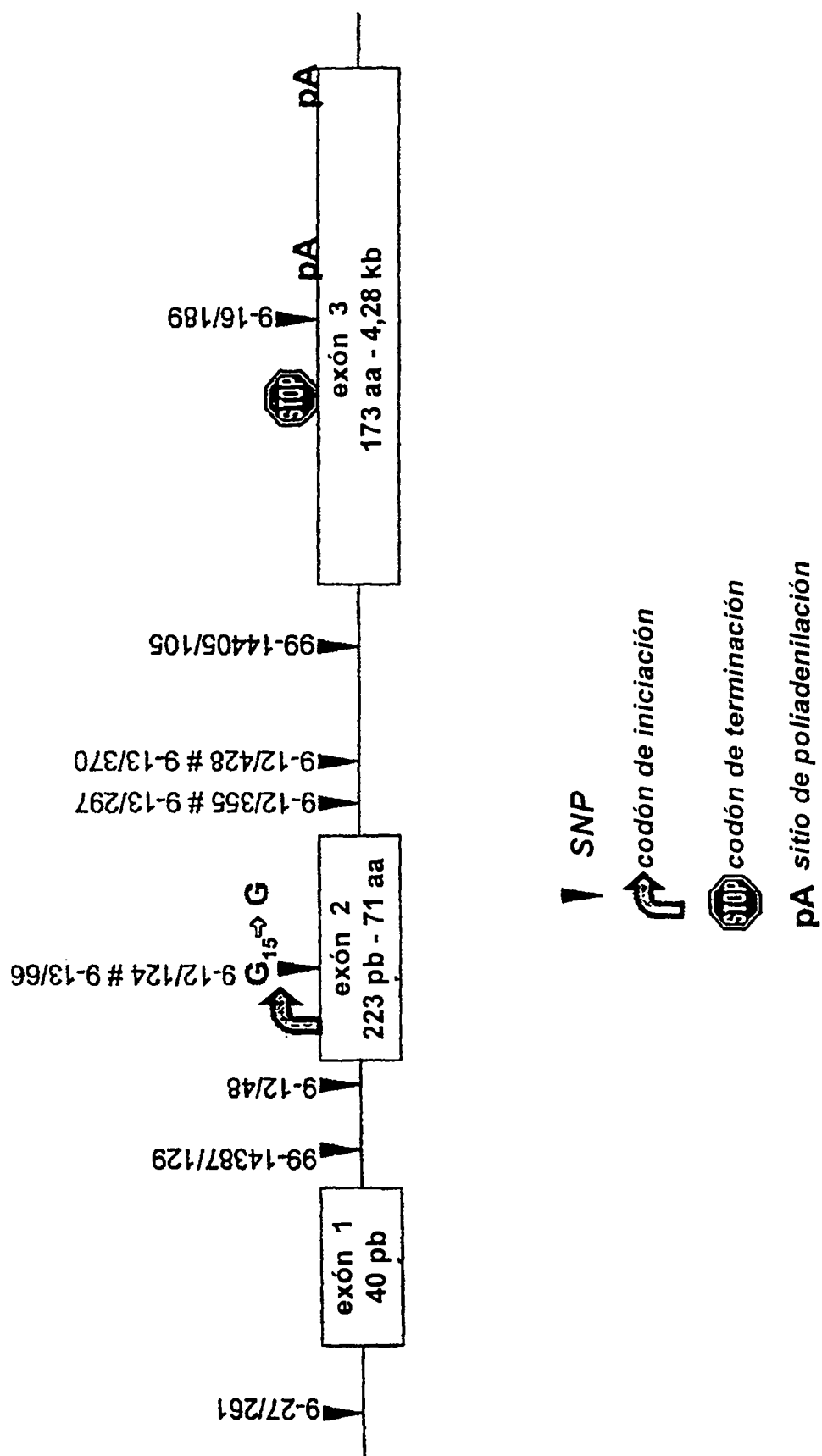
50 c) determinar si existe una asociación estadísticamente significativa entre dicho haplotipo y dicho fenotipo.

21. Un método o según la reivindicación 18 o según la reivindicación 20, en el que dicho fenotipo es una enfermedad en la que está implicada la obesidad o los trastornos relacionados con la obesidad, en el que esa enfermedad o trastorno es aterosclerosis, resistencia a la insulina, hipertensión, microangiopatía en individuos obesos con diabetes de tipo II, lesiones oculares asociadas con una microangiopatía en individuos obesos con diabetes de tipo II, o lesiones renales asociadas con una microangiopatía en individuos obesos con diabetes de tipo II.

60

65

Fig. 1



ES 2 308 852 T3

LISTA DE SECUENCIAS

<110> *Genset*

5 <120> Secuencia genómica y secuencia completa de cDNA de Apm1 y de sus marcadores bialélicos

<150> US 60/107.113

<151> 04-11-1.998

10 <150> US 60/119.593

<151> 10-02-1.999

15 <160> 8

<170> Patent.pm

20 <210> 1

<211> 20.966

<212> DNA

<213> *Homo sapiens*

25 <220>

<221> característica misc.

<222> 1...4.811

30 <223> región reguladora en 5'

<220>

<221> exón

35 <222> 4.812..4.851

<223> exón 1

<220>

40 <221> exón

<222> 15.144..15.365

<223> exón 2

45 <220>

<221> exón

<222> 16.277..20.559

50 <223> exón 3

<220>

<221> característica misc.

55 <222> 20.560..20.966

<223> región reguladora en 3'

<220>

60 <221> alelo

<222> 3.787

<223> 9-27-261: base polimórfica G o C

65 <220>

<221> alelo

ES 2 308 852 T3

<222> 11.118
<223> 99-14.387-129: base polimórfica A o C

5 <220>
<221> alelo
<222> 15.120
<223> 9-12-48: base polimórfica C o T

10 <220>
<221> alelo
<222> 15.196
15 <223> 9-12-124: base polimórfica G o T

<220>
20 <221> alelo
<222> 15.427
<223> 9-12-355: base polimórfica G o T

25 <220>
<221> alelo
<222> 15.500
<223> 9-12-428: base polimórfica A o G

30 <220>
<221> alelo
<222> 15.863
35 <223> 9-14.405-105: base polimórfica A o G

<220>
40 <221> alelo
<222> 17.170
<223> 9-16-189: delección de la base polimórfica A

45 <220>
<221> sitio de unión de cebador
<222> 3.528..3.545
<223> 9-27.pu

50 <220>
<221> sitio de unión de cebador
<222> 3.928..3.946
55 <223> 9-27.rp complemento

<220>
60 <221> sitio de unión de cebador
<222> 10.990..11.008
<223> 99-14.387.pu

65 <220>
<221> sitio de unión de cebador
<222> 11.423..11.442

ES 2 308 852 T3

<223> 99-14.387.rp complemento

<220>

5 <221> sitio de unión de cebador

<222> 15.073..15.092

<223> 9-12.pu

10 <220>

<221> sitio de unión de cebador

<222> 15.503..15.520

15 <223> 9-12.rp complemento

<220>

<221> sitio de unión de cebador

20 <222> 15.759..15.776

<223> 99-14.405.pu

<220>

25 <221> sitio de unión de cebador

<222> 16.191..16.211

<223> 99-14.405.rp complemento

30 <220>

<221> sitio de unión de cebador

<222> 16.982..17.001

35 <223> 9-16.pu

<220>

<221> sitio de unión de cebador

40 <222> 17.384..17.402

<223> 9-16.rp complemento

<220>

45 <221> sitio de unión misc.

<222> 3.775..3.799

<223> 9-27-261.sonda

50 <220>

<221> sitio de unión misc.

<222> 11.106-11.130

<223> 99-14.387-129.sonda

55

<220>

<221> sitio de unión misc.

<222> 15.108..15.132

60 <223> 9-12-48.sonda

<220>

<221> sitio de unión misc.

65 <222> 15.184..15.208

<223> 9-12-124.sonda

ES 2 308 852 T3

<220>
<221> sitio de unión misc.
<222> 15.415..15.439
5 <223> 9-12-355.sonda

<220>
<221> sitio de unión misc.
10 <222> 15.488..15.512
<223> 9-12-428.sonda

<220>
15 <221> sitio de unión misc.
<222> 15.851..15.875
<223> 99-14.405-105.sonda

20 <220>
<221> sitio de unión misc.
<222> 17.158..17.182
25 <223> 9-16-189.sonda

<220>
<221> sitio de unión de cebador
30 <222> 3.768..3.786
<223> 9-27-261.mis

<220>
35 <221> sitio de unión de cebador
<222> 3.788..3.806
<223> 9-27-261.mis complemento

40 <220>
<221> sitio de unión de cebador
<222> 11.099..11.117
45 <223> 99-14.387-129.mis.

<220>
<221> sitio de unión de cebador
50 <222> 11.119..11.137
<223> 99-14.387-129. mis complemento

<220>
55 <221> sitio de unión de cebador
<222> 15.101..15.119
<223> 9-12-48.mis

60 <220>
<221> sitio de unión de cebador
<222> 15.121..15.139
65 <223> 9-12-48.mis complemento

ES 2 308 852 T3

<220>
<221> sitio de unión de cebador
<222> 15.177..15.195
5 <223> 9-12-124.mis

<220>
<221> sitio de unión de cebador
10 <222> 15.197..15.215
<223> 9-12-124.mis complemento

<220>
15 <221> sitio de unión de cebador
<222> 15.408..15.426
<223> 9-12-355.mis

20 <220>
<221> sitio de unión de cebador
<222> 15.428..15.446
25 <223> 9-12-355.mis complemento

<220>
<221> sitio de unión de cebador
30 <222> 15.481..15.499
<223> 9-12-428.mis

<220>
35 <221> sitio de unión de cebador
<222> 15.501..15.519
<223> 9-12-428.mis complemento

40 <220>
<221> sitio de unión de cebador
<222> 15.844..15.862
45 <223> 99-14.405-105.mis

<220>
<221> sitio de unión de cebador
50 <222> 15.864..15.882
<223> 99-14.405-105.mis complemento

<220>
55 <221> sitio de unión de cebador
<222> 17.151..17.169
<223> 9-16-189.mis

60 <220>
<221> sitio de unión de cebador
<222> 17.171..17.189
65 <223> 9-16-189.mis complemento

ES 2 308 852 T3

<400> 1

5	gctgatctgc tgcctcagcc ttcccaaagt gctgtaattt attaggcata agccactgtg	60
	cctgcctagt gttgtacatt ctgtggggtt tgacaattgt atgcatctac atgtatgtac	120
	catttatagt attcctgttt ttaatttttag ccattctagt aggcattgtg tgatatctca	180
10	tgggtgatttt aattttgcgtt tccgtaattgg ttaataatgc tgaacatctt tgcattgtgt	240
	tgtttgtcat ttgtgtttcc tacttgggtg aataattgtt catgtccttt gtccattttc	300
	taattgaatt tttttttacc atttagtttt gagatttctt tatacaatct agatccaaat	360
15	ctcttgtctc aaatatgggt tgcaaataca ttctctaat tcatatattg ccttttcttc	420
	ctcttaacag gatgtttcac agagcaaaag ttttagtttt gttgaaatct cacttttcat	480
	ttttttcttt agtggattgt gcttttgtt tcatatgtaa gaactcttca ctggccctag	540
20	atccttgtat tgggttctca agattgccat agcaaatcac catgaactta gtgacaaaaa	600
	gacagaaatt tatttttact tctactgtg ggcagactag acgttaatta ttttcatgta	660
	tgtcatttcc tatgacatct ttctgatata ataattatag ttattcttaa gcttcacct	720
25	tttttctatt agctttgtta ccttgggtgt cactttttct tttttgacat tgtgacctat	780
	gccagatcat gtctgttagt acttagccct ccattcacct ctccataatc ccttttgtat	840
	tcctggagct tgatgcctga aatgacacat cctacattcc tttgccagat gggtagcagt	900
30	tagcttgtgc acatgggaga caaccgtgaa aagactgaag tggggaagaa gggaggagct	960
	gttgtgtttc agtgagcgcc cttggcagtg gcggtgacag tggctcctgt tcagtggcaa	1020
	tggtaggagca gctagcaaga catgcagtaa ggcaggtct ataggctatg gtccaggagc	1080
	agtcaccgat tcctgggtct taggcaatat catctcctt tgccttctca gcctttctaa	1140
35	aattattgta ccttgactag tacaattttt tagtattggg ggtagtccaa ggacacaggc	1200
	tttaaaaagt atgaattcag ggttgcctac ctgcattgac tgcgcttgaa tcatgatggc	1260
	cttctggtcg gtggcaggag gtgacagtc aaatcatgca gtagcaaacc agatacttaa	1320
40	attatcatct gagatacttc agaagtacag ccgtagccat accttcagaa gagataaaga	1380
	aatgttctcc tggccaggcg cggtaggtca cgcctgtcat tccagcactt tgggaggccg	1440
	agggggtgga tcacctgagg tcgggagttc gagaccagcc tgaccaacat ggggaaaccc	1500
45	tgtctctact aaaaatacaa aatttagcggg gcggtgtggc acatgcccat aatcccagct	1560
	actcgggagg ctaaggcagg ataatcgctt gaacctgaga ggcagagggt gcggtgaact	1620
	gagatcatgc catagtactc cagcctgggc aacaagagtg aaactccatc tcaaaaaaaa	1680
	aaaaaagaaa aaaagataaa gaaatgttct cctttcttgc catttctagg ggtttgggga	1740
50	tggcgtacat tgctgcaggc cgtgtcact ctaccatctt gctccaatct ttatttttca	1800
	aaatacagtg cttatgcttg gttacttcag ttaagattat ttttaaaaat cataattaag	1860
	caaaaatata tggccatgct taaacatatt taagataaat taagtattt ggctgtttc	1920
55	agtatcccaa ctacatgct aacaggggct tgacctgtag ctacggtagc ctggaggaaa	1980
	tgatcgcat tatttgggtta tttcgggtcta agtagtaata gttctgtcct gggaaaaaga	2040
	ctagcctcaa ggcatttctg attgaatgtt tttcaattac agtcttttaa ccagtatgcc	2100
60	acagaactgg ctctttccac atgacggcct ttgtgggtggg tggcagattg ccctgaggcc	2160
	tcgcaaaatg ctaggctttc acaatgtcac tgactgacag ccaggcccag cacagtcttg	2220
	gtgtgattgt ggggctaaag ttattccacc ttgtgcaata gctacagcct tctctaacca	2280

65

ES 2 308 852 T3

gctgcattct tataaagtta gaagaaaata cttttttttt tttgagatgg attctcgcctc 2340
 tgttgcccag gctggagtgc aatggtgcga tctcggtcgc ctgcaacctc cgctcctggt 2400
 5 gttcaaacga ttctcctccc tcagaccccc gagtagctgg gattgcaggt gcctgccacc 2460
 acgcccggct aacttttttg tatttttagt ggagacgggg tttcaccatc ttcgtcaggc 2520
 tggctctcaga ctctgacct caagtgatct gcccgctca gcctcccaa atgctgggat 2580
 10 tacaggcatg agctactgtg cccggccaaa gaaaatactt tttatgccag ccctgaaact 2640
 accctgaagc acatacatca accctgaggg ctacactcc atcaagaggg gtgaagggca 2700
 tgaggaatta gaaagcatag ggatttttag ttagacagat ctggttcaaa tcctagactt 2760
 gtgccttgaa caaattatct accctcattg aactctagat tcattatttg taaaatgaaa 2820
 15 gacaataata gttatctcca aaggaaagt gaatatgatc attcatttat tcattaattc 2880
 aacatttatt attgcctact ttgtgccagg ttctatteta ggaactaagg gatacaactt 2940
 tgaataggca aaatctctgc tctcctgaag tttacttttt tttttttttt ttgagacaga 3000
 20 gtttcaactc tgtcaccag gctggagcgc aatggtgctc ttggctcact gcaacctcca 3060
 cctcctgggt tcaagtgat ctctgtctc agcctcccaa gtagctggga ctacaggat 3120
 gtgccaccac gcccggtat ttctgcatct ttagtagaga tggggtttca ccatgttggc 3180
 cagactggtc tcaaactcct gatctcaggt gatatgcctg tcttggcctt ccaaagtact 3240
 25 gggattacag gcctgagcca ctgcacctga cctgaagttt atgttctatt aaatagcaac 3300
 agacagtaac ataaaccaa aataaatagg aaacacccat aacaaaaatc aaacagtgat 3360
 ataattgaga gttgcttcta tttctttttt ttgtcttctt ggttcaatca gcctgctaaa 3420
 30 ctatatggaa cctcatcttc atgggccact tatttaagcc gggggacctt ggaaagtctc 3480
 tcatgtctct catctcaacg gcctaagtgt acttctcttg aaatatttgg acattagcag 3540
 gaagctgagg ctttacatca gatctttact ttaatgggtg acttgacttt actggttagat 3600
 35 ttttaggctc tgtgtggact gtggagatga tatctggggg gcaggcagac acttgccctg 3660
 cctctgtctg agaaaattct gttttggatg tcttgttgaa gttggtgctg gcatcctaag 3720
 ccttgcctgg ggtcgtatt taattcatca gaatgtgtgg cttgcaagaa ccggctcaga 3780
 40 tcctgcctct caaaaaaaa acatgagcgt gccaagaaag tccaaggtgt tgaatgttgc 3840
 cacttcaagc ctaactttc taggaacacc taagtgggtg gcagcttcca gttctccagg 3900
 ctgcttctag gccagagctg ggttccacaa gagacagaat aggcataat atgcttaagg 3960
 aactggaaaa acaggctctc tctctctcac aaacacacac acacacatac caaggtagct 4020
 45 gtcaaaatgt tatccgaaat ttggaacca aaaaatcttg aaagatggta ttccaatatc 4080
 acattttatg taagttttct attatattag attcaaatta cgattcgagg ccacaagctt 4140
 taagaattca gggccttttt aacttgccaa gcccacacc actccaggaa cttcccaca 4200
 50 cccagttct cagaattcat gtgcaaggct tttcctaaat ccagggtcca ggtcagagag 4260
 tggaggatgt gctctatttc ttacctgatt gcagaccct ctgacagtgc tcccttctga 4320
 agcactcact gtctgaacgt acacagtctc agacttaatc atgcacagtg agcaagactg 4380
 55 tgggtgtgata attggcgtcc ctgacttatt agggcaaatc tatgggaggg ggagacctcc 4440
 tggaccactg agcaattaat tcatttacat taggaagttt ctccgtcaga tgcaggaaaa 4500
 aaatcttgtt ttctgtgtg ggttttgact tttgccccat cttctgttgc tgtttagga 4560
 ggcaaaataa ggtcaaggc ctggaacac aagtgtttg actgaagctc cacttggtt 4620
 60 ccgaagccca agctgggttg taccagggtc cctaggggtg aggctgtggg caactgccag 4680
 ggacatgtgc ctgcccaccg gcctctggcc ctactgagt tggccaatgg gaaatgacaa 4740
 ttgtgagggt gggactgcct gcccctgta gtaccaggct gttgaggctg ggcatctcc 4800

ES 2 308 852 T3

	tcctcacttc cattctgact gcagtcctgtg gttctgattc cataccagag ggtaagagca	4860
	attctgtgaa gttccaggct ggggtgggga tgcattgcata gcctctggct gggatcaccc	4920
5	aggctctccc gtccgtagta gtgtgggagt ggatacaggt ggatactctg gtcagagcag	4980
	cactgggtga ggcagatatg cactgggctt ctctctccgt tctcccacag cccaagaga	5040
	gaaagggtta tttcagacat tccttctaag atgcatggaa ccattctgaa ttttaccag	5100
10	ttcgctctgt agcaggatac ctattgagaa aaagttaggg tcagtaagggt ggaagggtct	5160
	gtccacagat gaagtccaat tcgattaagg gggataaggg aatacattgt ctcttagctt	5220
	gaccaggtag ggcaaaggaa gaagcatata tgaaggcagc ttcagaaaag tcaagctgag	5280
15	cactgacttc agactggaat taggaatcca gctctgccac tttattctac tcagcaaata	5340
	tttactgagc aaattctatg ggctagacag tggattgggt tcacaagata caatgagtgt	5400
	gacatggttg ttgtctatgg atttggggat atatgtagggt atagggatat cttacaagggt	5460
	aatcaagagg ttctaattgag gccagccatg gtggctcaca cctgtaatcc cagcaatttg	5520
20	ggagaccgag gcgggtggat cacctgagggt caggagtcc agactagcct gaccaacatg	5580
	gtgaaacccc gcctctacca aaaatacaaa aattagttgg gcgtgatggc aggtgcctgt	5640
	aatcccagct tctcgggagg ctgaggcagg agaattgtct gaacctggga ggcagagggt	5700
25	gcagtgagcc gagattgttg ccactgcatt ccagcctggg tgacagagcg agactttgtg	5760
	tcaaaaaaaa aaaaaaaaaaag aaagaaaaga aaaagaggct ctaatgagat aaaatgagaa	5820
	aagcctggca tgtagtggca acttatgaaa aattgtaatt aaaaaaaaaa attttctgac	5880
30	agaagaaact ggatctacct ggtttttctg aagcctaate ctgctcgccc cagtgagtgc	5940
	tgtttctgag gcatcctgggt tgttttgagc tgtggatgct gaaggtaga gtgggagggg	6000
	tttttagagg taggtctgcc cctcttgtgt tagaggacat ggatccctgg tctggagagg	6060
35	ttctggtttt tggatcaagc ctcacaagggt gtggcaccaa ctcaactcta ggaactccgc	6120
	tagaagggaag gccagctctg cctaattcgg ttggggagat ggggggccct ttatgctagc	6180
	agaatatgtc cgaaggagca tgatggtgtc agctttgttc atgaaggcca gtggtacaca	6240
	gggagcccg cagcttcctc agcagtcctt gctgccactc ttccttaagt cttgaggagt	6300
40	cttttttttg cacaatctca gctcactgca acctccgctt cccagggttca agcgattctc	6360
	ctgcctcagt ctcccaagta gctgagacta caggcatgcg ccaccacgcc cagctaattt	6420
	ttatatTTTT agtagagatg gggttcacca tattggccag gatggtctcg atctcttgac	6480
45	ctcatattcc acctgcctcg gcctcccaa gtgctggtat tacagggtgtg agccactgcg	6540
	cctggccgag gagtcttaag ctgagatcac agcattgcac tccagcctgg gcaaaaagag	6600
	caaaactcca tctcaaaaaa aaaaaaaaaa tagacacaag actggctcct tgtctttttt	6660
50	ggggacaggg tctcactcta tcaccagggc tggagtgcag tggtgcaatc acagctcact	6720
	gcagcctcga tttcccaggc tcaagtgacc ctcccatctt agcctcctga gtagctggga	6780
	ctacagggtgt gtgcaaccat gcctggctaa tttttaaaaa tttttttag agatgagggtc	6840
55	tcactatatt ggcctggggg cctcaaactc ctgggctcag cagtcctccc acctcagcct	6900
	cccaaaaggc tgggattata tgcttgctct ttttaagggt gctgtaggga caaactttcc	6960
	acctactcct tgtcaagcca gtggaccgggt ggtcccagac atacggctaa agtcaagagg	7020
60	tgatgtcttt tggagagata ctttcaatca ggaatttcaa tcagaaattc aatcatgtgg	7080
	agagagactt atcctaaaaa tgtggtggtg cgtgggatgc tctgttttat tagttccttg	7140
	acagtatgta tgtgtgtgag tgtgtgtgtg tgcgcgcgca cactcatttg gatgggtgtg	7200
	tatgtgtgtg ggggggtgggt gcgtacgtat gtggatgtgt ggatgtgggt tgtgggtgtg	7260
65	cgcgtgcata ggtggagggtg tgtgtatggg tgcgggtatg tgtgtgtgtt gggcatggag	7320

ES 2 308 852 T3

	atattgacag	ctctcccagg	gctgagtga	ggctttcggg	caaagctcct	gggagctagg	7380
	caaagctgag	ttgattcctg	gttatgccat	ttattattgg	gttgaccctg	gtgaaactgc	7440
5	caatattcta	cactttgact	tttatttatt	tttattttta	ttttttttga	gacagagttt	7500
	cacacttgtc	acctaggctg	gagtgcagtg	gcgcgatctc	agctcactgc	aacctctgcc	7560
	tcatggattc	aagtgattct	cctgcctcag	cctcccaagt	agctggaatt	acaggtgccc	7620
10	gctaccacgc	ctgactaatt	tttgtatttt	tggtagagac	gggatttcac	catgttgctc	7680
	aggctgggtc	gaaactcctg	acatcaggta	atccaccac	ctcagcctcc	caaagtgcgt	7740
	ggattacagg	catgagccac	tgcccccggc	ccattttgac	ttttaaaaa	gggagtttga	7800
15	tataattcaa	tccagtgggt	gaattagcta	gcacgcgttc	ctctccaagt	ctcaggttct	7860
	cctacacggt	agagtcaaaa	gcagggctat	gggaagatta	agtaaaataa	attttgaaaa	7920
	tgccctatga	aaattacact	ccaaagaact	cgccgcagtg	tcagtgttct	catgttctct	7980
20	atctcacatg	atcacatttc	gcggattagg	aagctgagtc	tgagaagctc	cgtgtagtgc	8040
	tttttcggag	gcaccgtgat	gtgatggaag	gctcactcgt	taggaagtca	gaacagagtc	8100
	tctgagggat	catttcctta	atctgtcagt	ttcctcatct	ctgaagttgg	gctcatttcc	8160
	ttccttcatg	gagttattgt	aaagatgaag	ataaataacg	tgtaaaatct	agcatgggaa	8220
25	ctggcttcta	taaggttcta	ataagtgcac	tcctactcct	ccccctcagc	cttcccattt	8280
	gtaaaagcaa	ggcaggggtg	aggtgatttc	tggggtcctc	tttggtctct	acatttgagg	8340
	attttgtatc	cttttttttt	tcagagtctt	gctctgtcac	ccaggttgga	gtgcagctca	8400
30	atgcaaattc	cgcctcccag	gctcaagcaa	ttcttatgtc	tcagcctcct	gagtacctgg	8460
	gattacaggg	aggcaccacc	acccccagct	aattttttgt	attttcagta	gagacggggg	8520
	tttgccatat	tgccagggct	ggtcttgaac	tcctgacttc	atgtgaccca	cccatctcag	8580
35	cctcccaaag	tgtctgagtg	acaggtgtga	gctaccgtgc	ctggccaatt	ttgtgtgctt	8640
	taatgccctt	ttctgctgga	agagttggca	ccaggtttgg	tgatctcttt	cccccacacg	8700
	gctctgcctc	ctgccagtc	cagaggggac	cctgtccttg	catttcacag	gattctgctg	8760
40	ttgcaactga	aattccagta	ggtcaaagtg	aaattttctc	tacacttta	catgaagata	8820
	aatgatcaca	gtatggccct	ttaggatcct	gagaacatca	cggctatccc	ctggtataat	8880
	tttaaaagca	gatgaatcca	tgccgtgtcg	aggtttgcc	ggaaagccag	tgtctgggatt	8940
45	acagtggga	tctttttatg	ctactttttt	cttgtatccc	tcaccccatg	gggtggcata	9000
	ttgaaaggca	ggatgtgtga	ccacgatact	tttctcctcc	tggaactatg	ctaagagtct	9060
	gttattgggt	tctgaagatc	agagttta	ttccgactcc	tctctgtgta	gctctgggat	9120
	cttggaagc	cacttaacct	ttctgaagtc	ccctttcctc	atctctaaaa	tgcatacact	9180
50	catcactaac	atctactgag	cactgacatg	tgccagacac	cattctaagc	attttacaca	9240
	gactacacca	tttgatcttc	caacaaacag	aacactgaaa	cgcattacag	gtcagaacaa	9300
	atgatttgtg	cctaagcacc	aagaccgtag	agcccgtgct	ccctattcta	ccctatcctg	9360
55	tctctcaaaa	tgattgtgag	aatcgaaatg	gacactaggt	gagaaaaggg	ttttataaat	9420
	agcattttta	aaatttttta	aagtccacaa	aaattttta	tttaatacag	ataaaataga	9480
	tccctttgtt	ttataaaaag	taacaaaatt	tggtatacaa	caactatggt	atctattaat	9540
60	tttgcccttt	tgtatgctgc	caggaaagaa	acattaagaa	atcttaaat	gattatgggtg	9600
	aatcagaagg	tctgcctgga	ctttttattg	ctctaactgt	acagctgac	atactacctc	9660
	attttttttt	atgacacttc	aaggggtgcgc	ttagcttcat	cactccttcg	ttgccaaaag	9720
	ctttgtgacc	aaaaacaatt	aagcagattc	ctgagtcact	aaatgacaca	taaccagagt	9780
65	tgagacttag	gaacttttag	tgccatgcta	agcccacagg	gacacaacaa	atagcatttt	9840

ES 2 308 852 T3

acaaaggcaa agaattgtga cacttgagat ttagcttggt gatccttgta aaagttttct 9900
 ttttaggcat aattgagttt tagatcatag tactcactat tacttagtaa taattttttt 9960
 5 ctgatagaaa tacagtgtaa caggccgggc gcagtggctc atgcctgtaa tcccagcact 10020
 ttgggaggcc gaggcgggcg gatcacttga ggtcaggagt ttgagaccag cccggccaac 10080
 atggtgaaat cccatctcta ctaaaaatac aaaaaattag ccagggtgtgg tcgtggattc 10140
 10 ctgtgatccc agctacttgg gaggggtgagg caggagcatc agttgaaccc aggaggcgga 10200
 ggttgacgtg agccaagatg gtgccattgc actccagcct gggccacaaa gcgagactcc 10260
 acttcagaaa caaaaaaaaa aagagagaga gagaaaagaa ggaaggaagg aaggaaggaa 10320
 ggaagaagg aaggaaggaa ggaagaagg aaggaaggaa ggaaggaag aaggaaggaa 10380
 15 ggaagaagg aaggaaggaa agaaggaagg aaggaaggaa ggaagggtaa caagcaaagt 10440
 gtaacaatgg caatatctaa aaaaataggt atttttatat gtttgcgtt ttatatatat 10500
 gacccccact ttagagatga ggaaactgag agattaagga aacgatccct gagagactct 10560
 20 gttctgactt ccaaatcggg gagctttcca tcgcatcacg gtgcctccga aagcatgaca 10620
 cggagcttct cagacttagc ttctaatacc gctaaacggg attatgtgag atgaggaaca 10680
 tgagaacgct gacatgggtg agggttcctt ggagtatcat ttcatgtgg cattttcaaa 10740
 25 acttatttta cctaattctt ccaaagccct gcttttgact ctaatgtgtc tcctgagact 10800
 tggagagcgc aagatgctag cgacagagca agactccatc tccagataaa taaataagta 10860
 aaataaaaaa gaacacaaat aattttgaaa atttttttga aaattaggca cgtttgcact 10920
 30 gaccttcaat tgttattaat tgctggtttc ccaccagaa ttaagttgga atgcaacttt 10980
 cttttacaat cagagtcctt tcttggctct ggaaacttct gaggctcctg tgctaatecc 11040
 actcttgat ttttggcacc tctacccgt gccactgtca tggaaaccag gctgatcgca 11100
 35 cctattagtg gagaaatmtg tccataatac tgaagtttg ggacaaacag tgttccctta 11160
 gggtaggaga aagagatctt tatttttaac aaagggggag gagccagaaa actccagaga 11220
 cccctgagtt tgccctctct ccaaggtttg gggtaagccc cccgtcacc tttatctctg 11280
 gggctttcac atattctgga ttctctctc ctgtttccca gcagaaaagg atggagcctc 11340
 40 acagattctt cccatttctg gagaaaaaca tgcattggag tcaaagtct tctcaggagt 11400
 tttattgcca aagccataat aagaaaggt ggaggtgaca agcagtgagg aagtttaag 11460
 atgcatgaaa tctgtaaagt ctcagaacaa gaattctcct aaaatgcaaa aggggctttg 11520
 45 ctggtctccc cttggcttct catgtagctc acctcttttt tcttatcttg agactagtca 11580
 aacctaaagt gtttctcatt ttatttccag aagctattga gaacactctc ctgaattctt 11640
 caaattcagt agagggcgac aaatgtacat ataaatgatg gtagtgggtc ttaaataaag 11700
 50 actcatgaca cctaaagggg cagcacctga gtctgattgc acctgtttct gttgctgttt 11760
 ctgtctctct tctctctgtc tgccatttca ttatcaatgg ttactttact tataagatca 11820
 tattagaacc tgatatttga taaatgatgc atcagatcta tagtgagaga aaaaattaat 11880
 55 gcaattaaag gtgttgtaac agctagtctt caagtgggga gaaatcattt gagtacctta 11940
 ggtcacagct tacatcaaaa caaaaaatca gagctacatt aaaaagtga attttaacta 12000
 tatcaacaa tagaaaaaaa cagaagaaaa ttgaatactt actaaatctt agcatgaata 12060
 60 agaactgttt aacacttaga ggcaaggact gggcgtgggt gctcatgctt ttaatcccag 12120
 gactttggga gcccaaggcg ggcggatcac ctgaggtcag gagtttgaga ctagcctggc 12180
 caacatggtg aaaccccgct tctactaaaa aatgcaaaaa ttagctgcgt gtggtggtgc 12240
 atgcctgtaa tctcagctac ttgggaggct aaggcatgag aatcgcttga acctgggagg 12300
 65 tggaggctgt agtgagccga gattgtgcca ctgcactaca gcctgggtga cagtgtgaaa 12360

ES 2 308 852 T3

	tcctctctct	caaaaaaaaa	aaaaaaaaaa	agcaaaactag	agcagtgagg	taccattatt	12420
	tcctttgtct	actaaactga	caacacacaa	atgtttttta	taatacccaa	agctgatgag	12480
5	ggtagttaag	gtatgccctt	ttatacacac	actaatgatg	tactactggt	tggcagtata	12540
	acatatgctg	ccatgtgggg	atatgtatca	ggagacttaa	aaatgtgcat	accttttggc	12600
	ccagtaattt	acttctggga	atctgtcata	acagaataat	aatcttgggg	aaagctacat	12660
10	gcctaaggat	attttaaata	ttatttaaaa	atcaaagtat	aatttcttac	agaatataaa	12720
	ataatatatt	aaaatgaaaa	tatgctaaaa	gtttgatgaa	atataaatgg	tcaaataatat	12780
	attgattata	tccacttact	agactagcac	tcactctgag	acgttaaaaa	tagtcattat	12840
15	aaaaactaga	aaatgccaaa	gacaaaataa	aggaataaaag	ttttacataa	agtatgatcc	12900
	cactatgttt	aaaaataaac	agagacattc	ttggagttga	gtattgtttt	cttttctgtc	12960
	atgtccaaag	aactatataa	ctattatttt	taatgaacta	tatatgtaat	atatacatat	13020
20	agtttatatg	tatatacaaa	atttatctca	tatatatgat	aaagatgaaa	gatgagttgg	13080
	atgtgccacg	tgaagtgggt	agtatagaaa	cccaggtaat	ggggcatagg	agtgggatcc	13140
	cagataccag	gcccattgtt	ttggggtgag	attgccaatc	acggtctttc	ttccatccct	13200
	cacagaggag	taggtttgtc	ttcaacaaac	cttcagttgt	cctgaagaca	aacctaatcc	13260
25	tggagacttc	atataatcta	gaagagacaa	gcaaactgat	gaaaaatagt	gaatttttaa	13320
	ggtaaaataa	agtacatgga	ctacactttg	tttagaatca	gattcttggg	attaaccaca	13380
	ttaaccacac	gagggtctta	gtgatgcctc	taatccagga	tcctaggacc	tatttctctc	13440
30	tgtgagatgc	tttctcccaa	ctccttggtg	agagtgggaa	gactaagacc	tcagcaatct	13500
	gagggtggag	cctaagatcc	ccctaagatc	ggaggcagaa	tctgagaggg	gataaaaagtc	13560
	cctatacctg	tattgggccc	tttctgsga	gggggatatc	aaagaatgat	tttgagacag	13620
35	ggaggctttt	gactacctgt	gccacttgag	ctctttgcta	gggctccaga	atacatattt	13680
	caaatacatt	ccccctccct	cttctcttcc	ctcttccact	cttctctttt	atcttctctt	13740
	cttcttttcc	ttctctcttc	cttctcttcc	tctggctctc	tcattgatttc	tttctctcat	13800
40	tataaaagtg	cttattttagt	ccctactctg	ctattagtgt	gttagtcttt	gtcccttggt	13860
	acttgctgtt	taatggagaa	atgggtgagc	aaaacagaaa	ttacagcaga	gtgcaataat	13920
	agagctaagc	caggtgtata	aatccattct	cacactgctg	taaaaaacta	ctgggtaatt	13980
	tataaagaaa	agaggtttta	ttgactcaca	gttccacagg	ctgtacagga	agcatggctg	14040
45	gggaggcctc	agaaaactta	caatcatggt	ggaagaaaga	gcgaagggga	agcaagcaca	14100
	tcacacagca	gcaggagaga	gagagagaaa	gagagagaga	gagaatatag	gggaagtgtc	14160
	acacactttc	aaccagatct	tgtgagaatt	cacctactat	catgagaaca	gcaagggata	14220
50	agtctgcctc	catgattcag	tcacctccta	ccaggccctc	tctccaacac	atgtcgacgt	14280
	gctattttgg	tggggacaca	gacccaaacc	atattaccag	ggcactggag	aaacacagag	14340
	gggaaagaac	cagccaagga	gtgagatgga	gaacaaggag	gacttcttga	aacagatgac	14400
55	atccaaactg	ggctcctgaaa	gctgaataga	gattagacag	gggaggaggg	gcagctaaaag	14460
	atggctcagg	caaacaaagg	gccaggggat	atgttcattg	gatgatgtgt	ctctcgttgt	14520
	ctgcttaaca	caagggtgagt	ctctccctcc	ctctctctct	ctttttctct	gtgtgtgttt	14580
60	gtgtgtgtgc	atgtgtgcaa	atgtaataata	cccaatagtc	aaacatgtgc	cccaggagag	14640
	gggtagagga	agaaagagaa	tgagagagta	agaaggagga	atagacacag	aaaatgagag	14700
	agaagggggg	aaagaaaaag	aagaaaggag	ccagaggaga	gaagctggtt	agcattgaat	14760
	ggagcaatct	gtgtcatcgt	acttgggaaa	ccaagggatg	gattcttggc	aagtcgactc	14820
65	ttggagcttt	ccctgtgctt	ggtcctgtgc	tcagacatgg	gaaaattaga	ggagtgtcat	14880

ES 2 308 852 T3

ctgtgcaatc actgaattca taatcttggg gaggaagga gactacacac agggaaataat 14940
 gctaagtatt acagatttca gggcagaaag agatcaagggt gggctgcaat attcagaaaa 15000
 5 gtcttctctg aaaagttgaa tacttagaaa gcagctccta gaagtagact ctgctgagat 15060
 ggacggagtc ctttgtaggt cccaactggg tgtgtgtgtg gggctctgtct ctccatggcy 15120
 gacagtgcac atgtggattc cagggtcag gatgctgttg ctgggagctg ttctactgct 15180
 10 attagctctg cccggkcatg accaggaaac cagcactcaa gggcccgag tctgcttcc 15240
 cctgcccag ggggcctgca cagggtggat ggcgggcac ccagggcac cgggcatata 15300
 tggggcccca ggcgtgatg gcagagatgg caccctggg gagaagggtg agaaaggaga 15360
 tccaggttaag aatgtttctg gcctctttca tcacagacct cctacactga tataaactat 15420
 15 atgaagkcat tcattattaa ctaaggccta gacacagga gaaagcaaag cttttttatg 15480
 ttaaccataa gcaacctgar gtgatttggg gttggctctc caaggatgag tgtagatgg 15540
 gcctctataa ccaagacttt ggctttgctg catctgcagc tcttttcca tcccccttcc 15600
 20 catcttcacc ctcaccccta tcccagtac attcatattc tgattcctct ttctgtctgc 15660
 ttaacttcca tttcaccag tggcattcaa ccacatttac tgcacacccc ctgaaaggct 15720
 cagtccctgcc tttggggaac tcttgatcta ggtaagatgt ctaatgtgca aggctctgtt 15780
 25 ggtggttacc acaagaaagt ctactctaaa aatgtcaaac tgaatgtgaa caagtattca 15840
 aagtatggag catagagaaa atrtactcac cgtggacctg atgaagaatg aaggcttcaa 15900
 ggaggaggca gagcttcagc taggccttga atgatgggtg ggcagaatag aggaggagag 15960
 30 acatccctaga tggagggggg agaattgcaa aaccagggtt gatggtgcca gcacataaag 16020
 ggctggcagg gtggagggtc tatgatagag acctatagga gataaagata gagttgaaat 16080
 tatgggagcc tccatgtctg tgggagatat agaaggagga ggtaacacct ctctcctttt 16140
 35 gggagctctt attggtttct tgatctataa gtcaagaagg ttgtgagtgg gagccacagg 16200
 gatggtaatt taggctgtaa ccaacctagg caggagtctt gttctttgta gtcactgagg 16260
 tcttctcatt ccttaggtct tattggtcct aaggagaca tcggtgaaac cggagtacc 16320
 ggggctgaag gtccccgagg ctttcgggga atccaaggca ggaaggaga acctggagaa 16380
 40 ggtgcctatg tataccgctc agcattcagt gtgggattgg agacttacgt tactatcccc 16440
 aacatgcca ttcgctttac caagatcttc tacaatcagc aaaaccacta tgatggctcc 16500
 actggtaaat tccactgcaa cattcctggg ctgtactact ttgcctacca catcacagtc 16560
 45 tatatgaagg atgtgaagg cagcctcttc aagaaggaca aggtatgct cttcacctat 16620
 gatcagtacc aggaaaataa tgtggaccag gcctccggct ctgtgctcct gcatctggag 16680
 gtgggagacc aagtctggct ccaggtgtat ggggaaggag agcgtaatgg actctatgct 16740
 50 gataatgaca atgactccac cttcacaggc tttcttctct accatgacac caactgatca 16800
 ccactaactc agagcctcct ccaggccaaa cagcccaaaa gtcaattaaa ggctttcagt 16860
 acggttagga agttgattat tatttagttg gaggccttta gatattattc attcatttac 16920
 55 tcattcattt attcattcat tcatcaagta actttaaaaa aatcatatgc tatgttccca 16980
 gtcttgggga gcttcacaaa catgaccaga taactgacta gaaagaagta gttgacagt 17040
 ctattttgtg cccactgtct ctctgatgc tcatatcaat cctataaggc acagggaaca 17100
 agcattctcc tgtttttaca gattgtatcc tgaggctgag agagttaagt gaatgtctaa 17160
 60 ggtcacacaa gtattaagtg acagtgttag aaatcaaacc cagagctgtg gactttgttc 17220
 actagactgt gcccttttat agaggtacat gttctctttg gagtgttggg aggtgtctgt 17280
 tccccacctc acctgagagc cattgaattt gccttctca tgaattaaaa cctcccccaa 17340
 65 gcagagcttc ctcagagaaa gtggttctat gatgaagtcc tgtcttgga ggactactac 17400

ES 2 308 852 T3

	tcaatggccc	ctgcaactact	ctacttcctc	ttacctatgt	cccttctcat	gcctttccct	17460
	ccaacgggga	aagccaactc	catctctaag	tgctgaactc	atccctgttc	ctcaaggcca	17520
5	cctggccagg	agcttctctg	atgtgatata	cacttttttt	tttttttgag	atggagtctc	17580
	actctgtcac	ccaggctgga	gtacagtgac	acgacctcgg	ctcactgcag	cctccttctc	17640
	ctgggtccaa	gcaattattg	tgcctcagcc	tcccagtag	ctgagacttc	aggtgcattc	17700
10	caccacacat	ggctaatttt	tgtattttta	gtagaaatgg	ggtttcgtca	tgttggccag	17760
	gctggtctcg	aactcctggc	ctaggtgatc	caccgcctc	gacctcccaa	agtgtctggga	17820
	ttacaggcat	gagccaccat	gcccagtcga	tatctcactt	tttattttgc	catggatgag	17880
	agtcctgggt	gtgaggaaca	cctcccacca	ggctagaggc	aactgcccag	gaaggactgt	17940
15	gcttccgtca	cctctaaatc	ccttgccagat	ccttgataaa	tgccctcatga	agaccaatct	18000
	cttgaatccc	gtatctaccc	agaattaact	ccattccagt	ctctgcattg	aatcagtttt	18060
	atccacagaa	acattttcat	tttaggaaat	ccctggtttt	aagtatcaat	ccttgttcag	18120
20	ctggacaata	tgaatctttt	ccactgaagt	tagggatgac	tgtgattttc	agaacacgtc	18180
	cagaattttt	catcaagaag	gtagcttgag	cctgaaatgc	aaaacccatg	gaggaattct	18240
	gaagccattg	tctccttgag	taccaacagg	gtcaggggaag	actgggcctc	ctgaatttat	18300
25	tattgttctt	taagaattac	aggttgaggt	agttgatggg	ggtaaacatt	ctctcaggag	18360
	acaataactc	cagtgtatgt	cttcaaagat	tttagcaaaa	acagagtaaa	tagcattctc	18420
	tatcaatata	taaattttaa	aaactatctt	tttgcttaca	gtttttaaat	ctgaacaatt	18480
30	ctctcttaca	tgtgtattgc	taatcattaa	ggtattattt	tttccacata	taaagctttg	18540
	tctttttgtt	gttggtgttg	tttttaagat	ggagtttccc	tctgttgcca	ggctagagtg	18600
	cagtggcatg	atctcggtt	actgcaacct	ttgcctccca	ggttcaagcg	attcttctgc	18660
	ctcagcctcc	cgagttagctg	ggaccacagg	tgccctaccac	catgccaggc	taatttttgt	18720
35	attttttagta	aagacagggt	ttcaccatat	tggccagggt	ggtctcgaac	tcctgacctt	18780
	gtgatctgcc	cacctccatt	tttggtgtta	ttttttgaga	aagatagata	tgagggttag	18840
	agagggatga	agaggtgaga	gtaagccttg	tgttagtcag	aactctgtgt	tgtgaatgtc	18900
40	attcacaaca	gaaaacccaa	aatattatgc	aaactactgt	aagcaagaaa	aataaaggaa	18960
	aaatggaaac	atttattcct	ttgcataata	gaaattacca	gagttgttct	gtcttttagat	19020
	aagggttgaa	ccaaagctca	aaacaatcaa	gacccttttc	tgtatgtcct	tctgttctgc	19080
45	cttccgcagt	gtaggcttta	ccctcagggt	ctacacagta	tagttctagg	gtttccctcc	19140
	cgatatcaaa	aagactgtgg	cctgcccagc	tctcgtatcc	ccaagccaca	ccatctgggt	19200
	aaatggacat	catgttttct	ggatgtgccc	aaagaggaga	gaggaagctc	tctttccag	19260
50	atgccccagc	aagtgttaacc	ttgcatctca	ttgctctggc	tgagttgtgt	gcctgtttct	19320
	gaccaatcac	tgagtcagga	ggatgaaata	ttcatattga	cttaattgca	gcttaagtta	19380
	ggggtatgta	gaggtatttt	ccctaagca	aaattgggac	actgttatca	gaaataggag	19440
55	agtggtatgat	agatgcaaaa	taatacctgt	ccacaacaaa	ctcttaatgc	tgtgtttgag	19500
	ctttcatgag	tttccagag	agacatagct	ggaaaattcc	tattgatttt	ctctaaaatt	19560
	tcaacaagta	gctaaagtct	ggctatgctc	acagtctcac	atctgggttg	ggtgggctcc	19620
60	ttacagaaca	cgctttcaca	gttaccctaa	actctctggg	gcaggggttat	tcctttgttg	19680
	aaccagaggc	acagagagag	tcaactgagg	ccaaaagagg	cctgagagaa	actgagggtca	19740
	agatttcagg	attaatggtc	ctgtgatgct	ttgaagtaca	attgtggatt	tgtccaattc	19800
	tcttttagttc	tgtcagcttt	tgttccatat	attttagcgc	tctattatta	gatatatata	19860
65	tgtttagtat	tatgtcttat	tgggtgcattt	actctcttat	cattatgtaa	tgtccttctt	19920

ES 2 308 852 T3

	tatctgtgat aattttctgt gttctgaagt ctactttgtc taaaaataac atacgcactc	19980
	aaacttccctt tctttctctcc ttcctttctt tcttccctcc tttctttctc tctctctctc	20040
5	tttccctctc tecttccctcc ttttctttct ctctctctct ctctctctct ttttgacaga	20100
	ctctcgttct gtggccctgg ctggagttca gtggtgtgat ctggctcac tgetacctct	20160
	accatgagca atttccctgc ctccgcctcc caagtagctg gaactacagg ctcatgccac	20220
10	tgcgcccagc taatttttgt atttttctga gagacggggg ttcaccacat tgcgcagggt	20280
	ggtttc aaac tccctgacttt gtgateccacc cgcctcggcc tcccaaagtg ctgggattac	20340
	aggcatgagc catcacacct ggtcaacttt cttttgatta gtgtttttgt ggtatatctt	20400
15	tttccatcat gttactttta atatatctat attattgtat ttaaaatgtg tttcttacag	20460
	actgcatgta gttgggtata atttttatcc agtctaaaaa tatctgtctt ttaattgggtg	20520
	tttagacaat ttatatttta taaaattggt gaatttaaga tggatgactg ttttatttgt	20580
20	ttgctgttca ccacttctgt tttattctct ttcagaatt cttttggatt gtttaaata	20640
	ttcataatat tttatcttaa tttatttatt gggattttgc ctatatctct ttgtggatt	20700
	ttttagtggg tgcttgaggg attacaatgt acttaacttt tcacagtgtg cataaagtta	20760
	atattttgcc acttgagta aaccgtagaa ggcttataat catattagta cctctatcca	20820
25	ctttctttta tgttgtagtt gtcatatata ttacatctat atacactgaa acattatagg	20880
	caatgttatg atttttgcat tgcgtcagtc tatatatatt ttaagaatt taagaggaga	20940
	aaaatacata ttcagatatt catcat	20966
30	<210> 2	
	<211> 4.811	
	<212> DNA	
35	<213> <i>Homo sapiens</i>	
	<400> 2	
40	gctgatctgc tgcctcagcc tcccaaagt gctgtaattt attaggcata agccactgtg	60
	cctgcctagt gttgtacatt ctgtgggttt tgacaattgt atgcatctac atgtatgtac	120
	catttatagt attcctgttt ttaatttttag ccattctagt aggcattgtg tgatatctca	180
45	tggtgatttt aatttgctt tccgtaatgg ttaataatgc tgaacatctt tgcattgtct	240
	tggttgcatt ttgtgtttcc tacttggtga aataattgtt catgtccttt gtccattttc	300
	taattgaatt tttttttacc atttagtttt gagatttctt tatacaatct agatccaaat	360
50	ctcttgtctc aaatatgggt tgcaaataca ttcctctaatt tcatatattg ctttttctc	420
	ctcttaacag gatgtttcac agagcaaaag ttttagtttt gttgaaatct cacttttcat	480
	ttttttcttt agtggattgt gcttttgttg tcatatgtaa gaactcttca ctggccctag	540
55	atccttgtat tgggtttccta agattgccat agcaaatac catgaactta gtgacaaaaa	600
	gacagaaatt tattttcact tccactgtg ggcagactag acgttaatta ttttcatgta	660
	tgctcattcc tatgacatct ttctgatata ataattatag ttattcttaa gcttcacct	720
60	tttttctatt agctttgtta ccttgggtgt cactttttct tttttgacat tgtgacctat	780
	gccagatcat gctctgttagt acttagccct ccattcacct ctccataatc ctttttgtat	840
	tcctggagct tgatgcctga aatgacacat cctacattcc tttgccagat gggtagcagt	900
	tagcttgtgc acatgggaga caaccgtgaa aagactgaag tggggaagaa gggaggagct	960
65	gttgtgtttc agtgagcgcc cttggcagtg gcggtgacag tggctcctgt tcagtggcaa	1020

ES 2 308 852 T3

	tggtggagca gctagcaaga catgcagtaa ggcagggctc ataggctatg gtccaggagc	1080
	agtcacccgat tcctggctct taggcaatat catctccctt tgcttctcca gcctttctaa	1140
5	aattattgta ccttgactag tacaattttt tagtattggg ggtagtcaa ggacacaggc	1200
	tttaaaaagt atgaattcag ggttgccctac ctgcattgac tgcgcttgaa tcatgatggc	1260
	cttctggctg gtggcaggag gtgacagtcc aaatcatgca gtagcaaacc agatacttaa	1320
10	attatcatct gagatacttc agaagtacag ccgtagccat accttcagaa gagataaaga	1380
	aatgtttctcc tggccaggcg cggtaggtca cgcctgtcat tccagcactt tgggaggccg	1440
	agggggtgga tcacctgagg tcgggagttc gagaccagcc tgaccaacat ggggaaaccc	1500
15	tgtctctact aaaaatacaa aattagcggg gcgtgggtgg acatgcccat aatcccagct	1560
	actcgggagg ctaaggcagg ataatcgctt gaacctgaga ggcagagggt gcggtgaact	1620
	gagatcatgc catagtactc cagcctgggc aacaagagtg aaactccatc tcaaaaaaa	1680
	aaaaaagaaa aaaagataaa gaaatgttct cctttcttgc catttctagg ggtttgggga	1740
20	tggcgctacat tgctgcaggg cgtgctcact ctaccatctt gctccaatct ttatttttca	1800
	aaatacagtg cttatgcttg gttacttcag ttaagattat ttttaaaaat cataattaag	1860
	caaaaatata tggccatgct taacatatt taagataaat taagtattt ggctgtttc	1920
25	agtatcccaa ctcatatgct aacaggggct tgacctgtag ctacgggtacc ctggaggaaa	1980
	tgatcgcaat tatttgggta tttcggctca agtagtaata gttctgtcct gggaaaaaga	2040
	ctagcctcaa ggcatttctg attgaatgtt tttcaattac agtcttttaa ccagtatgcc	2100
30	acagaactgg ctctttccac atgacggcct ttgtgggtgg tggcagattg ccctgaggcc	2160
	tcgcaaaatg ctaggctttc acaatgtcac tgactgacag ccaggcccag cacagtcttg	2220
	gtgtgattgt ggggctaaag ttattccacc ttgtgcaata gctacagcct tctctaacca	2280
35	gctgcattct tataaagtta gaagaaaata cttttttttt tttgagatgg attctcgctc	2340
	tggtgcccag gctggagtgc aatgggtgca tctcggtctg ctgcaacctc cgcctcctgg	2400
	gttcaaacga ttctcctccc tcagaccccc gagtagctgg gattgcagggt gcctgccacc	2460
	acgcccggct aacttttttg tatttttagt ggagacgggg tttcaccatc ttcgtcaggc	2520
40	tggtctcaga ctctgacct caagtgatct gccgcctca gcctcccaa atgctgggat	2580
	tacaggcatg agctactgtg cccggccaaa gaaaatactt tttatgccag ccctgaaact	2640
	accctgaagc acatacatca accttgaggc ctcacactcc atcaagaggg gtgaagggca	2700
45	tgaggaatta gaaagcatag ggatttttag ttagacagat ctggttcaaa tcctagactt	2760
	gtgccttgaa caaattattt accctcattg aactctagat tcattatttg taaaatgaaa	2820
	gacaataata gttatctcca aaggaaagt gaatatgatc attcatttat tcattaattc	2880
50	aacattttatt attgcctact ttgtgccagg ttctattcta ggaactaagg gatacaactt	2940
	tgaataggca aaatctctgc tctcctgaag tttacttttt tttttttttt ttgagacaga	3000
	gtttcactct tgtcaccag gctggagcgc aatgggtgctc ttggctcact gcaacctcca	3060
55	cctcctgggt tcaagtgatt ctctgtctc agcctcccaa gtagctggga ctacaggat	3120
	gtgccaccac gcccggctat ttctgcattt ttagtagaga tggggtttca ccattgtggc	3180
	cagactggtc tcaaactcct gatctcagggt gatatgcctg tcttggcctt ccaaagtact	3240
	gggattacag gcctgagcca ctgcacctga cctgaagttt atgttctatt aaatagcaac	3300
60	agacagtaac ataaaccaaa aataaatagg aaacacccat aacaaaaatc aaacagtgat	3360
	ataattgaga gttgcttcta tttctttttg ttgtcttctt ggttcaatca gcctgctaaa	3420
	ctatatggaa cctcattttc atgggccact tatttaagcc gggggacctt ggaaagtctc	3480
65	tcatgtctct catctcaacg gcctaagtgt acctctcttg aaatatttgg acattagcag	3540

ES 2 308 852 T3

gaagctgagg ctttacatca gatctttact ttaatggtgg acttgacttt actggttagat 3600
 ttttaggctc tgtgtggact gtggagatga tatctggggg gcaggcagac acttgccctg 3660
 5 cctctgtctg agaaaattct gttttggatg tcttgttgaa gttggtgctg gcatcctaag 3720
 cccttgctgg ggtcgtaatt taattcatca gaatgtgtgg cttgcaagaa ccggctcaga 3780
 tcctgcgctt caaaaacaaa acatgagcgt gccaaagaaag tccaaggtgt tgaatgttgc 3840
 10 cacttcaagc cttaaactttc taggaacacc taagtgggtg gcagcttcca gttctccagg 3900
 ctgcttctag gccagagctg gggtccacaa gagacagaat aggcataatat atgcttaagg 3960
 aactggaaaa acaggctctc tctctctcac aaacacacac acacacatac caaggtagct 4020
 15 gtcaaaatgt tatccgaaat tttggaacca aaaaatcttg aaagatggta ttccaatatc 4080
 acattttatg taagttttct attatattag attcaaatta cgattcgagg ccacaagctt 4140
 taagaattca gggccttttt aacttgccaa gccccacacc actccaggaa cttccccaca 4200
 20 cccagttct cagaattcat gtgcaaggct tttcctaata ccagggtcca ggtcagagag 4260
 tggaggatgt gctctatttc ttacctgatt gcagaccct ctgacagtgc tcccttctga 4320
 agcactcact gtctgaacgt acacagtctc agacttaatc atgcacagtg agcaagactg 4380
 25 tgggtgtgata attggcgctc ctgacttatt agggcaaatac tatgggaggg ggagacctcc 4440
 tggaccactg agcaattaat tcatttacat taggaagttt ctccgtcaga tgcaggaaaa 4500
 aaatcttggt ttctgtgtgt ggttttgact tttgccccat cttctgttgc tgttgtagga 4560
 30 ggcaaaataa ggggtcaaggc ctggaaacac aagtgtcttg actgaagctc cacttggtt 4620
 ccgaagccca agctgggttg taccaggctc cctaggggtg aggtgtgtgg caactgccag 4680
 35 ggacatgtgc ctgcccaccg gcctctggcc ctactgagt tggccaatgg gaaatgacaa 4740
 ttgtgagggt gggactgcct gccccgtga gtaccaggct gttgaggctg ggccatctcc 4800
 tcttcacttc c 4811

40 <210> 3
 <211> 407
 <212> DNA
 45 <213> *Homo sapiens*
 <400> 3

atggatgact gttttatttg tttgctgttc accacttctg ttttattctc tttccagaat 60
 50 tcttttggat tgtttaaata tttcataata ttttatctta atttatttat tgggtatttg 120
 cctatatctc tttgtggtat tttttagtgg ttgcttgagg gattacaatg tacttaactt 180
 55 ttcacagtgt gcataaagtt aatattttgc cacttgagcgt aaaccgtaga aggcttataa 240
 tcatattagt acctctatcc actttctttt atgttgtagt tgtcatatat attacatcta 300
 tataactga aacattatag gcaatgttat gatttttgca ttcgtcagtc atatatatat 360
 60 tttaaagaat ttaagaggag aaaaatacat attcagatat tcatcat 407

<210> 4
 <211> 366
 <212> DNA
 65 <213> *Homo sapiens*

ES 2 308 852 T3

<220>

<221> característica misc.

<222> 267

5 <223> n = a, g, c o t.

<400> 4

```

10      attctgactg cagtctgtgg ttctgattcc ataccagagg ggctcaggat gctggtgctg      60
      ggagctgttc tactgctatt agctctgccc gggcattgacc aggaaaccac gactcaaggg      120
      cccggagtcc tgcttccccct gcccaagggg gcctgcacag gttggatggc gggcattcca      180
15      gggcattccg gccataatgg ggccccaggc cgtgatggca gagatggcac ccctggtgag      240
      aagggtgaga aaggagatcc aggtctnatt ggtcctaagg gagacatcgg tgaaacggag      300
20      taccgggggc tgaagggtccc cgaggctttc cgggaatcca aggcaggaaa ggagaaccgg      360
      agaagg
  
```

<210> 5

25 <211> 4.545

<212> DNA

<213> *Homo sapiens*

30 <220>

<221> CDS

<222> 49..783

35 <220>

<221> característica misc.

<222> 1..367

40 <223> homología con la 5'-EST A254990 de un banco privado: GENSET

<220>

<221> señal de poliA

45 <222> 2.937..2.942

<223> AATAAA potencial

<220>

50 <221> señal de poliA

<222> 4.525..4.530

<223> AATAAA

55 <220>

<221> alelo

<222> 93

60 <223> 9-12-124: base polimórfica G o T

65

ES 2 308 852 T3

<400> 5

	attctgactg cagtctgtgg ttctgattcc ataccagagg ggctcagg atg ctg ttg	57
5	Met Leu Leu	
	1	
	ctg gga gct gtt cta ctg cta tta gct ctg ccc ggk cat gac cag gaa	105
10	Leu Gly Ala Val Leu Leu Leu Leu Ala Leu Pro Gly His Asp Gln Glu	
	5 10 15	
	acc acg act caa ggg ccc gga gtc ctg ctt ccc ctg ccc aag ggg gcc	153
15	Thr Thr Thr Gln Gly Pro Gly Val Leu Leu Pro Leu Pro Lys Gly Ala	
	20 25 30 35	
	tgc aca ggt tgg atg gcg ggc atc cca ggg cat ccg ggc cat aat ggg	201
20	Cys Thr Gly Trp Met Ala Gly Ile Pro Gly His Pro Gly His Asn Gly	
	40 45 50	
	gcc cca ggc cgt gat ggc aga gat ggc acc cct ggt gag aag ggt gag	249
25	Ala Pro Gly Arg Asp Gly Arg Asp Gly Thr Pro Gly Glu Lys Gly Glu	
	55 60 65	
	aaa gga gat cca ggt ctt att ggt cct aag gga gac atc ggt gaa acc	297
30	Lys Gly Asp Pro Gly Leu Ile Gly Pro Lys Gly Asp Ile Gly Glu Thr	
	70 75 80	
	gga gta ccc ggg gct gaa ggt ccc cga ggc ttt ccg gga atc caa ggc	345
35	Gly Val Pro Gly Ala Glu Gly Pro Arg Gly Phe Pro Gly Ile Gln Gly	
	85 90 95	
	agg aaa gga gaa cct gga gaa ggt gcc tat gta tac cgc tca gca ttc	393
40	Arg Lys Gly Glu Pro Gly Glu Gly Ala Tyr Val Tyr Arg Ser Ala Phe	
	100 105 110 115	
	agt gtg gga ttg gag act tac gtt act atc ccc aac atg ccc att cgc	441
45	Ser Val Gly Leu Glu Thr Tyr Val Thr Ile Pro Asn Met Pro Ile Arg	
	120 125 130	
	ttt acc aag atc ttc tac aat cag caa aac cac tat gat ggc tcc act	489
50	Phe Thr Lys Ile Phe Tyr Asn Gln Gln Asn His Tyr Asp Gly Ser Thr	
	135 140 145	
	ggg aaa ttc cac tgc aac att cct ggg ctg tac tac ttt gcc tac cac	537
55	Gly Lys Phe His Cys Asn Ile Pro Gly Leu Tyr Tyr Phe Ala Tyr His	
	150 155 160	
	atc aca gtc tat atg aag gat gtg aag gtc agc ctc ttc aag aag gac	585
60	Ile Thr Val Tyr Met Lys Asp Val Lys Val Ser Leu Phe Lys Lys Asp	
	165 170 175	
	aag gct atg ctc ttc acc tat gat cag tac cag gaa aat aat gtg gac	633
65	Lys Ala Met Leu Phe Thr Tyr Asp Gln Tyr Gln Glu Asn Asn Val Asp	
	180 185 190 195	
	cag gcc tcc ggc tct gtg ctc ctg cat ctg gag gtg ggc gac caa gtc	681

ES 2 308 852 T3

	Gln	Ala	Ser	Gly	Ser	Val	Leu	Leu	His	Leu	Glu	Val	Gly	Asp	Gln	Val	
	200					205					210						
5	tgg	ctc	cag	gtg	tat	ggg	gaa	gga	gag	cgt	aat	gga	ctc	tat	gct	gat	729
	Trp	Leu	Gln	Val	Tyr	Gly	Glu	Gly	Glu	Arg	Asn	Gly	Leu	Tyr	Ala	Asp	
	215					220					225						
10	aat	gac	aat	gac	tcc	acc	ttc	aca	ggc	ttt	ctt	ctc	tac	cat	gac	acc	777
	Asn	Asp	Asn	Asp	Ser	Thr	Phe	Thr	Gly	Phe	Leu	Leu	Tyr	His	Asp	Thr	
	230					235					240						
15	aac	tga	tcaccactaa	ctcagagcct	cctccaggcc	aaacagcccc	aaagtcaatt										833
	Asn	*															
	245																
	aaaggctttc	agtacgggta	ggaagttgat	tattatttag	ttggaggcct	ttagatatta											893
20	ttcattcatt	tactcattca	tttattcatt	cattcatcaa	gtaacttta	aaaaatcata											953
	tgctatgttc	ccagtcctgg	ggagcttcac	aaacatgacc	agataactga	ctagaaagaa											1013
	gtagttgaca	gtgctatttt	gtgcccactg	tctctcctga	tgctcatatc	aatcctataa											1073
25	ggcacaggga	acaagcattc	tctgtttttt	acagattgta	tcctgaggct	gagagagtta											1133
	agtgaatgtc	taaggtcaca	caagtattaa	gtgacagtgc	tagaaatcaa	accagagct											1193
	gtggactttg	ttcactagac	tgtgcccttt	tatagaggta	catgttctct	ttggagtgtt											1253
30	ggtaggtgtc	tgtttcccac	ctcacctgag	agccattgaa	tttgccctcc	tcatgaatta											1313
	aaacctcccc	caagcagagc	ttcctcagag	aaagtgggtc	tatgatgaag	tcctgtcctg											1373
	gaaggactac	tactcaatgg	cccctgcact	actctacttc	ctcttaccta	tgteccctct											1433
35	catgcctttc	cctccaacgg	ggaaagccaa	ctccatctct	aagtgtgtaa	ctcatccctg											1493
	ttcctcaagg	ccacctggcc	aggagcttct	ctgatgtgat	atccactttt	tttttttttt											1553
	gagatggagt	ctcactctgt	cacctcaggct	ggagtacagt	gacacgacct	cggctcactg											1613
	cagcctcctt	ctcctgggtc	caagcaatta	ttgtgcctca	gcctccagag	tagctgagac											1673
40	ttcaggtgca	ttccaccaca	catggcta	ttttgtattt	ttagtagaaa	tggggtttcg											1733
	tcattgttggc	caggctgggtc	tcgaactcct	ggcctagggtg	atccaccctg	ctcgacctcc											1793
	caaagtgtctg	ggattacag	catgagccac	catgcccagt	cgatatctca	ctttttattt											1853
45	tgccatggat	gagagtcctg	ggtgtgagga	acacctccca	ccaggctaga	ggcaactgcc											1913
	caggaaggac	tgtgtcttcg	tcacctctaa	atcccttgca	gatccttgat	aaatgcctca											1973
	tgaagaccaa	tctcttgaat	cccgatctca	cccagaatta	actccattcc	agtctctgca											2033
50	tgtaatcagt	tttatccaca	gaaacatttt	catttttagga	aatccctggg	tttaagtatc											2093
	aatccttggt	cagctggaca	atatgaatct	tttccactga	agttagggat	gactgtgatt											2153
	ttcagaacac	gtccagaatt	tttcatcaag	aaggtagctt	gagcctgaaa	tgcaaaaccc											2213
55	atggaggaat	tctgaagcca	ttgtctcctt	gagtaccaac	agggtcaggg	aagactgggc											2273
	ctcctgaatt	tattattggt	ctttaagaat	tacagggtga	ggtagttgat	ggtggtaaac											2333
	attctctcag	gagacaataa	ctccagtgat	gttcttcaaa	gatttttagca	aaaacagagt											2393
60	aaatagcatt	ctctatcaat	atataaattt	aaaaaactat	ctttttgctt	acagttttaa											2453
	atcctgaaca	attctctctt	acatgtgtat	tgctaatacat	taagggtatta	ttttttccac											2513
	atataaagct	ttgtcttttt	gttgttgttg	ttgtttttta	gatggagttt	ccctctgttg											2573
	ccaggctaga	gtgcagtggc	atgatctcgg	cttactgcaa	cctttgcctc	ccagggtcaa											2633
65	gcgattcttc	tgccctcagc	ccccagtag	ctgggaccac	aggtgcctac	caccatgcca											2693

ES 2 308 852 T3

ggctaattttt tgtatttttta gtaaagacag ggtttcacca tattggccag gctgggtctcg 2753
 aactcctgac ctttgtgatct gccacacctc atttttgttg ttattttttg agaaagatag 2813
 5 atatgaggtt tagagagggg tgaagaggtg agagtaagcc ttgtgttagt cagaactctg 2873
 tgttgtgaat gtcattcaca acagaaaacc caaaatatta tgcaaactac tgtaagcaag 2933
 aaaaataaag gaaaaatgga aacattttatt cctttgcata atagaaatta ccagagttgt 2993
 10 tctgtcttta gataagggtt gaaccaaage tcaaaacaat caagaccctt ttctgtatgt 3053
 ccttctgttc tgccttccgc agtgtaggct ttaccctcag gtgctacaca gtatagttct 3113
 agggtttccc tcccgatata aaaaagactg tggcctgccc agctctcgta tccccaagcc 3173
 15 acaccatctg gctaaatgga catcatgttt tctgggtgat cccaaagagg agagaggaag 3233
 ctctctttcc cagatgcccc agcaagtgtt accttgcata tcatgtctct ggctgagttg 3293
 tgtgcctgtt tctgaccaat cactgagtca ggaggatgaa atattcatat tgacttaatt 3353
 20 gcagcttaag ttaggggtat gtagaggtat tttccctaaa gcaaaattgg gacactgtta 3413
 tcagaaatag gagagtggat gatagatgca aaataatacc tgtccacaac aaactcttaa 3473
 tgctgtgttt gagctttcat gagtttccca gagagacata gctggaaaat tcctattgat 3533
 25 tttctctaaa atttcaacaa gtagctaaag tctggctatg ctcacagtct cacatctggg 3593
 tgggggtgggc tccttacaga acacgctttc acagttaccc taaactctct ggggcagggt 3653
 tattcctttg tgggaaccaga ggcacagaga gagtcaactg aggccaaaag aggcctgaga 3713
 30 gaaactgagg tcaagatttc aggattaatg gtccgtgat gcttgaagt acaattgtgg 3773
 atttgtccaa ttctcttttag ttctgtcagc ttttgcctca tatatttttag cgctctatta 3833
 35 ttagatatat acatgttttag tattatgtct tattgggtgca tttactctct tatcattatg 3893
 taatgtcctt ctttatctgt gataattttc tgtgttctga agtctacttt gtctaaaaat 3953
 aacatacgca ctcaacttcc ttttctttct tccttccctt ctttcttctt tcctttcttt 4013
 40 ctctctctct ctcttctctt ccttccctcc tccttttctt tctctctctc tctctctctc 4073
 tttttttgac agactctcgt tctgtggccc tggctggagt tcagtgggtg gatcttggct 4133
 cactgctacc tctaccatga gcaattctcc tgcctcagcc tcccaagtag ctggaactac 4193
 45 aggctcatgc cactgcgccc agctaatttt tgtatttttc gtagagacgg ggtttcacca 4253
 cattcgtcag gttgggttca aactcctgac tttgtgatcc accgcctcg gcctccaaa 4313
 gtgctgggat tacaggcatg agccatcaca cctgggtcaac tttcttttga ttagtgtttt 4373
 50 tgtgggtatat ctttttccat catgttactt taaatatatc tatattattg tatttaaaat 4433
 gtgtttctta cagactgcat gtagttgggt ataattttta tccagtctaa aaatatctgt 4493
 55 cttttaattg gtgttttagac aatttatatt taataaaatt gttgaattta ag 4545

<210> 6

<211> 244

60 <212> PRT

<213> *Homo sapiens*

65

ES 2 308 852 T3

<400> 6

5 Met Leu Leu Leu Gly Ala Val Leu Leu Leu Leu Ala Leu Pro Gly His
1 5 10 15
10 Asp Gln Glu Thr Thr Thr Gln Gly Pro Gly Val Leu Leu Pro Leu Pro
20 25 30
Lys Gly Ala Cys Thr Gly Trp Met Ala Gly Ile Pro Gly His Pro Gly
35 40 45
15 His Asn Gly Ala Pro Gly Arg Asp Gly Arg Asp Gly Thr Pro Gly Glu
50 55 60
20 Lys Gly Glu Lys Gly Asp Pro Gly Leu Ile Gly Pro Lys Gly Asp Ile
65 70 75 80
25 Gly Glu Thr Gly Val Pro Gly Ala Glu Gly Pro Arg Gly Phe Pro Gly
85 90 95
Ile Gln Gly Arg Lys Gly Glu Pro Gly Glu Gly Ala Tyr Val Tyr Arg
100 105 110
30 Ser Ala Phe Ser Val Gly Leu Glu Thr Tyr Val Thr Ile Pro Asn Met
115 120 125
35 Pro Ile Arg Phe Thr Lys Ile Phe Tyr Asn Gln Gln Asn His Tyr Asp
130 135 140
40 Gly Ser Thr Gly Lys Phe His Cys Asn Ile Pro Gly Leu Tyr Tyr Phe
145 150 155 160
Ala Tyr His Ile Thr Val Tyr Met Lys Asp Val Lys Val Ser Leu Phe
45 165 170 175
Lys Lys Asp Lys Ala Met Leu Phe Thr Tyr Asp Gln Tyr Gln Glu Asn
180 185 190
50 Asn Val Asp Gln Ala Ser Gly Ser Val Leu Leu His Leu Glu Val Gly
195 200 205
55 Asp Gln Val Trp Leu Gln Val Tyr Gly Glu Gly Glu Arg Asn Gly Leu
210 215 220
Tyr Ala Asp Asn Asp Asn Asp Ser Thr Phe Thr Gly Phe Leu Leu Tyr
60 225 230 235 240
His Asp Thr Asn

<210> 7

<211> 18

ES 2 308 852 T3

	<212> DNA	
	<213> Secuencia artificial	
5	<220>	
	<223> Cebador oligonucleotídico PU de secuenciación	
	<400> 7	
10	tgtaaaacga cggccagt	18
15	<210> 8	
	<211> 18	
	<212> DNA	
	<213> Secuencia artificial	
20	<220>	
	<223> Cebador oligonucleotídico RP de secuenciación	
25	<400> 8	
	caggaaacag ctatgacc	18
30		
35		
40		
45		
50		
55		
60		
65		