



①9



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

①1 Número de publicación: **2 330 903**

⑤1 Int. Cl.:

C12N 15/12 (2006.01)

C07K 14/47 (2006.01)

C07K 16/18 (2006.01)

C12Q 1/68 (2006.01)

①2

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑨6 Número de solicitud europea: **98959117 .7**

⑨6 Fecha de presentación : **17.12.1998**

⑨7 Número de publicación de la solicitud: **1037977**

⑨7 Fecha de publicación de la solicitud: **27.09.2000**

⑤4 Título: **ADNcs extendidos para proteínas secretadas.**

③0 Prioridad: **17.12.1997 US 69957 P**
09.02.1998 US 74121 P
13.04.1998 US 81563 P
10.08.1998 US 96116 P

④5 Fecha de publicación de la mención BOPI:
16.12.2009

④5 Fecha de la publicación del folleto de la patente:
16.12.2009

⑦3 Titular/es: **Serono Genetics Institute S.A.**
route Nationale 7
91030 Evry Cédex, FR

⑦2 Inventor/es: **Bougueleret, Lydie;**
Duclert, Aymeric y
Dumas Milne Edwards, Jean-Baptiste

⑦4 Agente: **Elzaburu Márquez, Alberto**

ES 2 330 903 T3

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

ADNcs extendidos para proteínas secretadas.

La presente solicitud se refiere a ADNc extendidos que se han dado a conocer en varias solicitudes de patente provisional de Estados Unidos. La Tabla I enumera los N° de ID de SEC de ADNc extendido en las presentes reivindicaciones, los N° de ID de SEC de ADNc extendido idéntico o casi idéntico en las solicitudes provisionales y la identidad de la solicitud provisional en la que se dio a conocer el ADNc extendido.

Antecedentes de la invención

Los estimadamente 50.000-100.000 genes dispersos por los cromosomas humanos ofrecen una enorme promesa para la comprensión, diagnóstico y tratamiento de enfermedades humanas. Además, las sondas capaces de hibridación específica con loci distribuidos por todo el genoma humano encuentran aplicación en la construcción de mapas cromosómicos de alta resolución y en la identificación de individuos.

En el pasado, la caracterización de incluso un solo gen humano era un proceso cuidadoso que requería años de esfuerzo. Los recientes desarrollos en las áreas de vectores de clonación, secuenciación de ADN y tecnología informática se han combinado para acelerar en gran medida la velocidad a la que pueden aislarse, secuenciarse, cartografiarse y caracterizarse genes humanos. Los vectores de clonación tales como cromosomas artificiales de levadura (YAC) y cromosomas artificiales bacterianos (BAC) son capaces de aceptar insertos de ADN en el intervalo de 300 a 1.000 kilobases (kb) o 100-400 kb de longitud, respectivamente, facilitando así la manipulación y ordenación de las secuencias de ADN distribuidas a lo largo de grandes distancias en los cromosomas humanos. Los dispositivos de secuenciación automatizada de ADN permiten la secuenciación rápida de genes humanos. El software de bioinformática permite la comparación de secuencias de ácido nucleico y proteína, ayudando así a la caracterización de productos génicos humanos.

Actualmente, se están siguiendo dos enfoques diferentes para identificar y caracterizar los genes distribuidos por el genoma humano. En un enfoque, se aíslan, clonan y secuencian fragmentos grandes de ADN genómico. Se identifican los marcos abiertos de lectura potenciales en estas secuencias genómicas usando software de bioinformática. Sin embargo, este enfoque supone secuenciar extensiones grandes de ADN humano que no codifican proteínas para encontrar las secuencias de codificación de proteína dispersadas por todo el genoma. Además de requerir una extensa secuenciación, el software de bioinformática puede caracterizar erróneamente las secuencias genómicas obtenidas. Por tanto, el software puede producir falsos positivos en los que ADN no codificante se caracteriza erróneamente como ADN codificante o falsos negativos en los que ADN codificante se caracteriza erróneamente como ADN no codificante.

Un enfoque alternativo toma una vía más directa para identificar y caracterizar genes humanos. En este enfoque, se sintetizan ADN complementarios (ADNc) a partir de ARN mensajeros aislados (ARNm) que codifican proteínas humanas. Usando este enfoque, se efectúa la secuenciación sólo en ADN que deriva de porciones del genoma codificantes de proteína. A menudo, se secuencian sólo extensiones cortas de ADNc para obtener secuencias denominadas marcadores de secuencia expresada (EST). Los EST pueden usarse entonces para aislar o purificar ADNc extendidos que incluyen secuencias adyacentes a las secuencias de EST. Los ADNc extendidos pueden contener toda la secuencia de EST que se usó para obtenerlos o sólo una porción de la secuencia de EST que se usó para obtenerlos. Además, los ADNc extendidos pueden contener la secuencia de codificación completa del gen del que derivaba el EST o, como alternativa, los ADNc extendidos pueden incluir porciones de la secuencia codificante del gen del que derivaba el EST. Se apreciará que puede haber varios ADNc extendidos que incluyen la secuencia de EST como resultado de un corte y empalme alternativo o de la actividad de promotores alternativos.

En el pasado, las secuencias de EST cortas que podían usarse para aislar o purificar ADNc extendidos se obtenían a menudo a partir de colecciones de ADNc cebadas con oligo-dT. En consecuencia, correspondían principalmente a la región 3' no traducida del ARNm. En el pasado, la prevalencia de las secuencias de EST derivadas del extremo 3' del ARNm es el resultado del hecho de que las técnicas típicas para obtener ADNc no son bien adecuadas para aislar secuencias de ADNc derivadas de los extremos 5' de ARNm (Adams *et al.*, *Nature* 377: 174, 1996, Hillier *et al.*, *Genome Res.* 6: 907-828, 1996).

Además, en aquellos casos reseñados en que se han obtenido secuencias de ADNc más largas, las secuencias reseñadas corresponden típicamente a secuencias de codificación y no incluyen la región 5' no traducida entera del ARNm del que deriva el ADNc. Dichas secuencias incompletas pueden no incluir el primer exón del ARNm, particularmente en situaciones en que el primer exón es corto. Además, pueden no incluir algunos exones, a menudo cortos, que están localizados cadena arriba de sitios de corte y empalme. Por tanto, existe la necesidad de obtener secuencias derivadas de los extremos 5' de ARNm que puedan usarse para obtener ADNc extendidos que puedan incluir las secuencias 5' contenidas en los 5' EST.

Aunque muchas secuencias derivadas de cromosomas humanos tienen aplicaciones prácticas, los enfoques basados en la identificación y caracterización de aquellas secuencias cromosómicas que codifican un producto proteico son particularmente relevantes para usos de diagnóstico y terapéuticos. De los 50.000-100.000 genes codificantes de proteína, aquellos genes que codifican proteínas que se secretan de la célula en la que se sintetizan, así como las proteínas secretadas mismas, son particularmente valiosos como agentes terapéuticos potenciales. Dichas proteínas están a

menudo implicadas en la comunicación de célula a célula y pueden ser responsables de la producción de una respuesta clínicamente relevante en sus células diana.

De hecho, están actualmente en uso clínico varias proteínas secretoras, incluyendo activador de plasminógeno de tejido, G-CSF, GM-CSF, eritropoyetina, hormona de crecimiento humano, insulina, interferón- α , interferón- β , interferón- γ e interleucina-2. Estas proteínas se usan para tratar un amplio intervalo de afecciones, incluyendo infarto de miocardio agudo, apoplejía isquémica aguda, anemia, diabetes, deficiencia de hormona de crecimiento, hepatitis, carcinoma renal, neutropenia inducida por quimioterapia y esclerosis múltiple. Por estas razones, los ADNc extendidos que codifican proteínas secretadas o porciones de las mismas representan una fuente particularmente valiosa de agentes terapéuticos. Por tanto, existe la necesidad de identificación y caracterización de proteínas secretadas y de los ácidos nucleicos que las codifican.

Además de ser terapéuticamente útiles por sí mismas, las proteínas secretadas incluyen péptidos cortos, denominados péptidos señal, en sus extremos amino que dirigen su secreción. Estos péptidos señal están codificados por las secuencias señal localizadas en los extremos 5' de las secuencias de codificación de los genes que codifican proteínas secretadas. Debido a que estos péptidos señal dirigirán la secreción extracelular de cualquier proteína a la que estén ligados operativamente, las secuencias señal pueden explotarse para dirigir la secreción eficaz de cualquier proteína ligando operativamente las secuencias señal a un gen que codifica la proteína para la que se desea secreción. Esto puede demostrarse beneficioso en estrategias de terapia génica en las que se desea suministrar un producto génico particular a células distintas de la célula en que se produce. Las secuencias señal que codifican péptidos señal encuentran también aplicación en la simplificación de técnicas de purificación de proteína. En dichas aplicaciones, la secreción extracelular de la proteína deseada facilita en gran medida la purificación al reducir el número de proteínas indeseadas entre las que debe seleccionarse la proteína deseada. Por tanto, existe la necesidad de identificar y caracterizar las porciones 5' de los genes de proteínas secretadas que codifican péptidos señal.

La información pública sobre el número de genes humanos para los que se han identificado y caracterizado las regiones promotoras y reguladoras cadena arriba es bastante limitada. En parte, esto puede ser debido a la dificultad de aislar dichas secuencias reguladoras. Las secuencias reguladoras cadena arriba tales como sitios de unión a factor de transcripción son típicamente demasiado cortas para utilizarse como sondas para aislar promotores de colecciones genómicas humanas. Recientemente, se han desarrollado algunos enfoques para aislar promotores humanos. Uno de ellos consiste en preparar una colección de islas de CpG (Cross, S.H. *et al.*, "Purification of CpG Islands using a Methylated DNA Binding Column", *Nature Genetics* 6: 236-244 (1994)). El segundo consiste en aislar secuencias de ADN genómico humano que contienen sitios de unión SpeI mediante el uso de la proteína de unión a SpeI (Mortlock *et al.*, *Genome Res.* 6: 327-335, 1996). Ambos enfoques tienen sus límites debido a la falta de especificidad o de integridad.

Los 5' EST y ADNc extendidos obtenibles a partir de ellos pueden usarse para identificar y aislar eficazmente regiones reguladoras cadena arriba que controlan la localización, etapa de desarrollo, velocidad y cantidad de síntesis de proteína, así como la estabilidad del ARNm. (Theil *et al.* *BioFactors* 4: 87-93, (1993). Una vez identificadas y caracterizadas, estas regiones reguladoras pueden utilizarse en terapia génica o esquemas de purificación de proteína para obtener la cantidad y localizaciones deseadas de síntesis de proteína, o para inhibir, reducir o evitar la síntesis de productos génicos indeseables.

Además, los EST que contienen los extremos 5' de genes de proteína secretada o ADNc extendidos que incluyen secuencias adyacentes a las secuencias de EST pueden incluir secuencias útiles como sondas para cartografía cromosómica y la identificación de individuos. Por tanto, existe la necesidad de identificar y caracterizar las secuencias cadena arriba de las secuencias de codificación 5' de genes que codifican proteínas secretadas.

Compendio de la invención

Se describen en la presente memoria ADNc extendidos purificados, aislados o recombinantes que codifican proteínas secretadas o fragmentos de las mismas. Los ADNc purificados, aislados o recombinantes pueden contener el marco abierto de lectura completo o sus correspondientes ARNm, incluyendo un codón de inicio y un codón de terminación. Por ejemplo, los ADNc extendidos pueden incluir ácidos nucleicos que codifican el péptido señal así como la proteína madura. Como alternativa, los ADNc extendidos pueden contener un fragmento del marco abierto de lectura. En algunas realizaciones, el fragmento puede codificar sólo la secuencia de la proteína madura. Como alternativa, el fragmento puede codificar sólo una porción de la proteína madura. Se describe también un ácido nucleico que codifica el péptido señal de una proteína secretada.

Se obtuvieron los presentes ADNc extendidos usando EST que incluyen secuencias derivadas de los extremos 5' auténticos de sus correspondientes ARNm. Como se usa en la presente memoria, los términos "EST" o "5' EST" se refieren a ADNc cortos que se usaron para obtener los ADNc extendidos de la presente invención. Como se usa en la presente memoria, el término "ADNc extendido" se refiere a los ADNc que incluyen secuencias adyacentes al 5' EST usado para obtenerlos. Los ADNc extendidos pueden contener toda o una porción de la secuencia de EST que se usó para obtenerlos. El término "ARNm correspondiente" se refiere al ARNm que era el molde para la síntesis de ADNc que produjo el 5' EST. Como se usa en la presente memoria, el término "purificado" no requiere la pureza absoluta; en lugar de ello, se pretende como una definición relativa. Los clones de ADNc extendido individuales aislados a partir de una colección de ADNc se han purificado convencionalmente hasta homogeneidad electroforética. Las secuencias

obtenidas a partir de estos clones no podían obtenerse directamente a partir de la colección o del ADN humano total. Los clones de ADNc extendido no son de origen natural como tales, sino que en lugar de ello se obtienen mediante manipulación de una sustancia de origen natural parcialmente purificada (ARN mensajero). La conversión de ARNm en una colección de ADNc implica la creación de una sustancia sintética (ADNc) y pueden aislarse clones de ADNc individuales puros a partir de la colección sintética mediante selección clonal. Por tanto, crear una colección de ADNc a partir de ARN mensajero y aislar posteriormente clones individuales a partir de esa colección da como resultado una purificación de aproximadamente 10^4 - 10^6 veces del mensaje nativo. Se contempla expresamente la purificación del material de partida o material natural hasta al menos un orden de magnitud, preferiblemente dos o tres órdenes, y más preferiblemente cuatro o cinco órdenes de magnitud.

Como se usa en la presente memoria, el término “aislado” requiere que el material sea separado de su entorno original (por ejemplo, el entorno natural si es de origen natural). Por ejemplo, un polinucleótido de origen natural presente en un animal vivo no está aislado, pero el mismo polinucleótido, separado de algunos o todos los materiales coexistentes en el sistema natural, está aislado.

Como se usa en la presente memoria, el término “recombinante” significa que el ADNc extendido está adyacente a un ácido nucleico “de cadena principal” al que no está adyacente en su entorno natural. Adicionalmente, para estar “enriquecidos”, los ADNc extendidos representarán 5% o más del número de insertos de ácido nucleico en una población de moléculas de cadena principal de ácido nucleico. Las moléculas de cadena principal según la invención incluyen ácidos nucleicos tales como vectores de expresión, ácidos nucleicos autorreplicativos, virus, ácidos nucleicos integrativos y otros vectores o ácidos nucleicos usados para mantener o manipular un inserto de ácido nucleico de interés. Preferiblemente, los ADNc extendidos enriquecidos representan un 15% o más del número de insertos de ácido nucleico en la población de moléculas de cadena principal recombinantes. Más preferiblemente, los ADNc extendidos enriquecidos representan un 50% o más del número de insertos de ácido nucleico en la población de moléculas de cadena principal recombinantes. En una realización altamente preferida, los ADNc extendidos enriquecidos representan un 90% o más del número de insertos de ácido nucleico en la población de moléculas de cadena principal recombinantes. Las condiciones de hibridación “rigurosa”, “moderada” y “baja” son como se definen en el ejemplo 29.

A menos que se indique otra cosa, una secuencia “complementaria” es completamente complementaria. Por tanto, los ADNc extendidos que codifican polipéptidos secretados o fragmentos de los mismos que están presentes en colecciones de ADNc en las que uno o más ADNc extendidos que codifican polipéptidos secretados o fragmentos de los mismos constituyen un 5% o más del número de insertos de ácido nucleico en las moléculas de cadena principal son “ADNc extendidos recombinantes enriquecidos”, como se definen en la presente memoria. Igualmente, los ADNc extendidos que codifican polipéptidos secretados o fragmentos de los mismos que están en una población de plásmidos en la que se han insertado uno o más ADNc extendidos de la presente invención de tal modo que representen un 5% o más del número de insertos en la cadena principal plasmídica son “ADNc extendidos recombinantes enriquecidos”, como se definen en la presente memoria. Sin embargo, los ADNc extendidos que codifican polipéptidos secretados o fragmentos de los mismos que están en colecciones de ADNc en las que los ADNc extendidos que codifican polipéptidos secretados o fragmentos de los mismos constituyen menos de un 5% del número de insertos de ácido nucleico en la población de moléculas de cadena principal, tales como colecciones en las que las moléculas de cadena principal que tienen un inserto de ADNc que codifica un polipéptido secretado son extremadamente raras, no son “ADNc extendidos recombinantes enriquecidos”.

En particular, se describen en la presente memoria ADNc extendidos que se derivaron de genes que codifican proteínas secretadas. Como se usa en la presente memoria, una proteína “secretada” es aquella que, cuando se expresa en una célula hospedadora adecuada, se transporta por o a través de una membrana, incluyendo transporte como resultado de péptidos señal en su secuencia aminoacídica. Proteínas “secretadas” incluye sin limitación proteínas secretadas enteramente (por ejemplo proteínas solubles) o parcialmente (por ejemplo receptores) a partir de la célula en la que se expresan. Proteínas “secretadas” incluye también sin limitación proteínas que se transportan a través de la membrana del retículo endoplasmático.

Los ADNc extendidos que codifican proteínas secretadas pueden incluir secuencias de ácido nucleico, denominadas secuencias señal, que codifican péptidos señal que dirigen la secreción extracelular de las proteínas codificadas por los ADNc extendidos. Generalmente, los péptidos señal están localizados en los extremos amino de proteínas secretadas.

Las proteínas secretadas se traducen por ribosomas asociados al retículo endoplasmático “rugoso”. Generalmente, las proteínas secretadas se transfieren cotraduccionalmente a la membrana del retículo endoplasmático. La asociación del ribosoma con el retículo endoplasmático durante la traducción de proteínas secretadas está mediada por el péptido señal. El péptido señal se escinde típicamente después de su entrada cotraduccion al retículo endoplasmático. Después del suministro al retículo endoplasmático, las proteínas secretadas pueden proseguir por el aparato de Golgi. En el aparato de Golgi, las proteínas pueden experimentar modificación postraduccion al antes de entrar en vesículas secretoras que las transportan a través de la membrana celular.

Los ADNc extendidos descritos en la presente memoria tienen varias aplicaciones importantes. Por ejemplo, pueden usarse para expresar la proteína secretada entera que codifican. Como alternativa, pueden usarse para expresar porciones de la proteína secretada. Las porciones pueden comprender los péptidos señal codificados por los ADNc ex-

tendidos o las proteínas maduras codificadas por los ADNc extendidos (concretamente, las proteínas generadas cuando se separa por escisión el péptido señal). Las porciones pueden comprender también polipéptidos que tienen al menos 10 aminoácidos consecutivos codificados por los ADNc extendidos. Como alternativa, las porciones pueden comprender al menos 15 aminoácidos consecutivos codificados por los ADNc extendidos. Las porciones pueden comprender al menos 25 aminoácidos consecutivos codificados por los ADNc extendidos. Las porciones pueden comprender al menos 40 aminoácidos codificados por los ADNc extendidos.

Los anticuerpos que reconocen específicamente las proteínas secretadas enteras codificadas por los ADNc extendidos o fragmentos de los mismos que tienen al menos 10 aminoácidos consecutivos, al menos 15 aminoácidos consecutivos, al menos 25 aminoácidos consecutivos o al menos 40 aminoácidos consecutivos pueden obtenerse también como se describe a continuación. Los anticuerpos que reconocen específicamente la proteína madura generada cuando se escinde el péptido señal pueden obtenerse también como se describe a continuación. De forma similar, pueden obtenerse también los anticuerpos que reconocen específicamente los péptidos señal codificados por los ADNc extendidos.

Los ADNc extendidos pueden incluir la secuencia señal. Los ADNc extendidos pueden incluir la secuencia de codificación completa de la proteína madura (concretamente, la proteína generada cuando se separa por escisión el polipéptido señal). Además, los ADNc extendidos pueden incluir regiones reguladoras cadena arriba del sitio de inicio de la traducción o cadena abajo del sitio de terminación que controlan la cantidad, localización o estado de desarrollo de la expresión génica. Como se discute anteriormente, las proteínas secretadas son terapéuticamente importantes, por tanto, las proteínas expresadas a partir de los ADNc pueden ser útiles para tratar o controlar una variedad de afecciones humanas. Los ADNc extendidos pueden usarse también para obtener el ADN genómico correspondiente. El término "ADN genómico correspondiente" se refiere al ADN genómico que codifica ARNm que incluye la secuencia de una de las cadenas del ADNc extendido en la que los residuos de timidina de la secuencia de ADNc extendido se reemplazan por residuos de uracilo en el ARNm.

Los ADNc extendidos o ADNc genómicos obtenidos a partir de los mismos pueden usarse en procedimientos forensicos para identificar individuos o en procedimientos de diagnóstico para identificar individuos que tienen enfermedades genéticas resultantes de una expresión anormal de los genes correspondientes a los ADNc extendidos. Además, los presentes ADN son útiles para construir un mapa de alta resolución de los cromosomas humanos.

Se describen también vectores de secreción capaces de dirigir la secreción de una proteína de interés. Dichos vectores pueden usarse en estrategias de terapia génica en las que se desea producir un producto génico en una célula que ha de suministrarse a otra localización del cuerpo. Los vectores de secreción pueden facilitar también la purificación de proteínas deseadas.

Se describen también vectores de expresión capaces de dirigir la expresión de un gen insertado de la manera deseada desde el punto de vista espacial o temporal o a un nivel deseado. Dichos vectores pueden incluir secuencias cadena arriba de los ADNc extendidos tales como promotores o secuencias reguladoras cadena arriba.

Además, los presentes ADN pueden usarse también para terapia génica para controlar o tratar enfermedades genéticas. Los péptidos señal pueden fusionarse también con proteínas heterólogas para dirigir su secreción extracelular.

Se describe en la presente memoria un ácido nucleico purificado o aislado que comprende la secuencia de SEC ID N° 45 o una secuencia complementaria de la misma. En un aspecto, el ácido nucleico es recombinante.

Se describe también un ácido nucleico purificado o aislado que comprende al menos 10 bases consecutivas de la secuencia de SEC ID N° 45 o una secuencia complementaria de la misma. En un aspecto, el ácido nucleico comprende al menos 15, 25, 30, 40, 50, 75 ó 100 bases consecutivas de la secuencia de SEC ID N° 45 o una secuencia complementaria de la misma. El ácido nucleico puede ser un ácido nucleico recombinante.

Se describe también un ácido nucleico purificado o aislado de al menos 15 bases capaz de hibridar en condiciones rigurosas con la secuencia de SEC ID N° 45 o una secuencia complementaria de la secuencia de SEC ID N° 45. En un aspecto, el ácido nucleico es recombinante.

Se describe también en la presente memoria un ácido nucleico purificado o aislado que comprende la secuencia de codificación completa de SEC ID N° 45, en el que la secuencia de codificación completa comprende opcionalmente la secuencia que codifica el péptido señal así como la secuencia que codifica la proteína madura. En un aspecto, el ácido nucleico es recombinante.

Se describe también un ácido nucleico purificado o aislado que comprende los nucleótidos de SEC ID N° 45 que codifican una proteína madura. En un aspecto, el ácido nucleico es recombinante.

Se describe también un ácido nucleico purificado o aislado que comprende los nucleótidos de SEC ID N° 45 que codifican el péptido señal. En un aspecto, el ácido nucleico es recombinante.

Se describe también un ácido nucleico purificado o aislado que codifica un polipéptido que tiene la secuencia de SEC ID N° 146.

ES 2 330 903 T3

Se describe también un ácido nucleico purificado o aislado que codifica un polipéptido que tiene la secuencia de una proteína madura incluida en la secuencia de SEC ID N° 146.

5 Se describe también un ácido nucleico purificado o aislado que codifica un polipéptido que tiene la secuencia de un péptido señal incluida en la secuencia de SEC ID N° 146.

Se describe también una proteína purificada o aislada que comprende la secuencia de SEC ID N° 146.

10 Se describe también un polipéptido purificado o aislado que comprende al menos 10 aminoácidos consecutivos de la secuencia de SEC ID N° 146. En un aspecto, el polipéptido purificado o aislado comprende al menos 15, 20, 25, 35, 50, 75, 100, 150 ó 200 aminoácidos consecutivos de la secuencia de SEC ID N° 146. En aún otro aspecto, el polipéptido purificado o aislado comprende al menos 25 aminoácidos consecutivos de la secuencia de SEC ID N° 146.

15 Se describe también un polipéptido aislado o purificado que comprende un péptido señal del polipéptido de SEC ID N° 146.

Se describe también un polipéptido aislado o purificado que comprende una proteína madura del polipéptido de SEC ID N° 146.

20 Se describe también un método de preparación de una proteína que comprende la secuencia de SEC ID N° 146, que comprende las etapas de obtener un ADNc que comprende la secuencia de SEC ID N° 45, insertar el ADNc en un vector de expresión de tal modo que el ADNc esté ligado operativamente a un promotor, e introducir el vector de expresión en una célula hospedadora, con lo que la célula hospedadora produce la proteína codificada por dicho ADNc. En un aspecto, el método comprende adicionalmente la etapa de aislar la proteína.

25 Se describe también una proteína obtenible mediante el método descrito en el párrafo precedente.

30 Se describe también un método de preparación de una proteína que comprende la secuencia aminoacídica de la proteína madura contenida en la secuencia de SEC ID N° 146, que comprende las etapas de obtener un ADNc que comprende la secuencia nucleotídica de SEC ID N° 45 que codifica la proteína madura, insertar el ADNc en un vector de expresión de tal modo que el ADNc esté ligado operativamente a un promotor, e introducir el vector de expresión en una célula hospedadora, con lo que la célula hospedadora produce la proteína madura codificada por el ADNc. En un aspecto, el método comprende adicionalmente la etapa de aislar la proteína.

35 Se describe también una proteína madura obtenible mediante el método descrito en el párrafo precedente.

Se describe también una célula hospedadora que contiene los ácidos nucleicos purificados o aislados que comprenden la secuencia de SEC ID N° 45 o una secuencia complementaria de la misma descrita en la presente memoria.

40 Se describe también una célula hospedadora que contiene un ácido nucleico purificado o aislado que comprende la secuencia de codificación completa de SEC ID N° 45, en la que la secuencia de codificación completa comprende la secuencia que codifica el péptido señal y la secuencia que codifica la proteína madura descrita en la presente memoria.

45 Se describe también una célula hospedadora que contiene un ácido nucleico purificado o aislado que comprende los nucleótidos de SEC ID N° 45 que codifican una proteína madura que se describen en la presente memoria.

Se describe también una célula hospedadora que contiene el ácido nucleico purificado o aislado que comprende los nucleótidos de SEC ID N° 45 que codifican el péptido señal que se describen en la presente memoria.

50 Se describe también un anticuerpo purificado o aislado capaz de unirse específicamente a una proteína que tiene la secuencia de SEC ID N° 146. En un aspecto, el anticuerpo es capaz de unirse a un polipéptido que comprende al menos 10 aminoácidos consecutivos de la secuencia de SEC ID N° 146.

55 Se describe también una serie de ADNc o fragmentos de los mismos de al menos 15 nucleótidos de longitud que incluye al menos una de las secuencias de SEC ID N° 40-140 y 242-377, o una de las secuencias complementarias de las secuencias de SEC ID N° 40-140 y 242-377, o un fragmento de las mismas de al menos 15 nucleótidos consecutivos. En un aspecto, la serie incluye al menos dos de las secuencias de SEC ID N° 40-140 y 242-377, las secuencias complementarias de las secuencias de SEC ID N° 40-140 y 242-377, o fragmentos de las mismas de al menos 15 nucleótidos consecutivos. En otro aspecto, la serie incluye al menos cinco de las secuencias de SEC ID N° 40-140 y 242-377, las secuencias complementarias de las secuencias de SEC ID N° 40-140 y 242-377, o fragmentos de las mismas de al menos 15 nucleótidos consecutivos.

60 Se describen también polinucleótidos purificados que comprenden un inserto de un clon depositado en un depósito que tiene un número de acceso enumerado en la Tabla VI o un fragmento del mismo que comprende un tramo continuo de al menos 8, 10, 12, 15, 20, 25, 40, 60, 100 ó 200 nucleótidos de dicho inserto. Se describen también polipéptidos purificados que comprenden, consisten en o consisten esencialmente en una secuencia aminoacídica codificada por el inserto de un clon depositado en un depósito que tiene un número de acceso enumerado en la Tabla VI, así como polipéptidos que comprenden un fragmento de dicha secuencia aminoacídica consistente en un péptido señal, una

proteína madura o un tramo contiguo de al menos 5, 8, 10, 12, 15, 20, 25, 40, 60, 100 ó 200 aminoácidos codificado por dicho inserto.

En consecuencia, en un aspecto la invención proporciona un polipéptido purificado o aislado que comprende una secuencia que tiene al menos un 80% de identidad con el polipéptido de SEC ID N° 146. En un aspecto adicional, la invención proporciona un ácido nucleico purificado o aislado que codifica un polipéptido que comprende o consiste en la SEC ID N° 146. La invención proporciona adicionalmente un vector de expresión que comprende el ácido nucleico y una célula hospedadora que comprende el ácido nucleico recombinante. Se proporciona también un método de preparación del polipéptido de la invención que comprende las etapas de: (a) obtener un ADNc que codifica dicho polipéptido; (b) insertar dicho ADNc en un vector de expresión de tal modo que dicho ADNc esté ligado operativamente a un promotor; y (c) introducir dicho vector de expresión en una célula hospedadora, con lo que dicha célula hospedadora produce dicho polipéptido. La invención proporciona adicionalmente un anticuerpo purificado o aislado capaz de unirse específicamente al polipéptido de la invención.

El n° de acceso Z95114 a la base de datos EMBL proporciona una secuencia de ADN humano del clon 212A2. Duchateau *et al.*, 1997, *Journal of Biological Chemistry*, 272: 25576-25582, se refieren a la apolipoproteína L, una nueva apolipoproteína de lipoproteína humana de alta densidad expresada por el páncreas.

Breve descripción de los dibujos

La Figura 1 es un compendio de un procedimiento para obtener ADNc que se han seleccionado para incluir los extremos 5' de los ARNm de los que derivan.

La Figura 2 es un análisis de los 43 aminoácidos aminotermiales de todas las proteínas SwissProt humanas para determinar la frecuencia de falsos positivos y falsos negativos usando las técnicas para identificación de péptidos señal descritas en la presente memoria.

La Figura 3 muestra la distribución de las puntuaciones de von Heijne para 5' EST en cada una de las categorías descritas en la presente memoria y la probabilidad de que estos 5' EST codifiquen un péptido señal.

La Figura 4 muestra la distribución de 5' EST en cada categoría y el número de 5' EST en cada categoría que tienen una puntuación de von Heijne mínima dada.

La Figura 5 muestra los tejidos de los que se obtuvieron los ARNm correspondientes a los 5' EST en cada una de las categorías descritas en la presente memoria.

La Figura 6 ilustra un método para obtener ADNc extendidos.

La Figura 7 es un mapa de pED6dpc2. pED6dpc2 deriva de pED6dpc1 mediante la inserción de un nuevo poliliga-dor para facilitar la clonación de ADNc. Se clonan los ADNc SSt entre EcoRI y NotI. Se describen los vectores PED en Kaufman *et al.* (1991), *NAR* 19: 4485-4490.

La Figura 8 proporciona una descripción esquemática de los promotores aislados y el modo en que se ensamblan con los correspondientes marcadores 5'.

La Figura 9 describe los sitios de unión a factor de transcripción presentes en cada uno de estos promotores.

Descripción detallada de la realización preferida

I. Obtención de 5' EST

Se obtuvieron los presentes ADNc extendidos usando 5' EST que se aislaron como se describe a continuación.

A. Métodos químicos para obtener ARNm que tengan extremos 5' intactos

Para obtener los 5' EST usados para obtener los ADNc extendidos descritos en la presente memoria, deben obtenerse ARNm que tengan extremos 5' intactos. Actualmente, hay dos enfoques para obtener dichos ARNm. Uno de estos enfoques es un método de modificación química que implica la derivatización de los extremos 5' de los ARNm y la selección de los ARNm derivatizados. Los extremos 5' de ARNm eucarióticos poseen una estructura referida como un "casquete" que comprende una guanosina metilada en la posición 7. El casquete está unido a la primera base transcrita del ARNm por un enlace 5',5'-trifosfato. En algunos casos, la 5' guanosina está metilada en ambas posiciones 2 y 7. Raramente, la 5' guanosina está trimetilada en las posiciones 2,7 y 7. En el método químico para obtener ARNm que tengan extremos 5' intactos, el casquete 5' se derivatiza específicamente y se acopla con un grupo reactivo sobre un sustrato inmovilizado. Esta derivatización específica está basada en el hecho de que sólo la ribosa ligada a la guanosina metilada en el extremo 5' del ARNm y la ribosa ligada a la base en el extremo 3' del ARNm

ES 2 330 903 T3

poseen 2',3'-cis dioles. Opcionalmente, cuando la ribosa 3' terminal tiene un 2',3'-cis diol, el 2',3'-cis diol en el extremo 3' puede estar químicamente modificado, sustituido, convertido o eliminado, dejando sólo la ribosa ligada a la guanosina metilada en el extremo 5' del ARNm con un 2',3'-cis diol. Están disponibles una variedad de técnicas para eliminar el 2',3'-cis diol en la ribosa 3' terminal. Por ejemplo, puede usarse hidrólisis alcalina controlada para generar fragmentos de ARNm en los que la ribosa 3' terminal sea un 3'-fosfato, 2'-fosfato o (2', 3')-ciclofosfato. Después de ello, el fragmento que incluye la 3' ribosa original puede eliminarse de la mezcla mediante cromatografía en una columna de oligo-dT. Como alternativa, puede añadirse una base que carece de 2',3'-cis diol al extremo 3' del ARNm usando una ARN ligasa tal como ARN ligasa T4. El ejemplo 1 siguiente describe un método para el ligamiento de pCp al extremo 3' del ARN mensajero.

Ejemplo 1

Ligamiento del difosfato de nucleósido pCp al extremo 3' de ARN mensajero

Se incubó 1 µg de ARN en un medio de reacción final de 10 µl en presencia de 5 U de ARN ligasa de fago T₄ en el tampón proporcionado por el fabricante (Gibco - BRL), 40 U del inhibidor de ARNasa ARNasina (Promega) y 2 µl de ³²pCp (Amersham PB n° 10208).

Se efectuó la incubación a 37°C durante 2 horas o durante una noche a 7-8°C.

Después de la modificación o eliminación del 2', 3'-cis diol en la ribosa 3', el 2', 3'-cis diol presente en el extremo 5' del ARNm puede oxidarse usando reactivos tales como NaBH₄, NaBH₃CN o peryodato de sodio, convirtiendo así el 2', 3'-cis diol en un dialdehído. El ejemplo 2 describe la oxidación del 2', 3'-cis diol en el extremo 5' del ARNm con peryodato de sodio.

Ejemplo 2

Oxidación de 2',3'-cis diol en el extremo 5' del ARNm

Se trataron del modo siguiente 0,1 unidades de DO de un oligorribonucleótido con casquete de 47 nucleótidos (incluyendo el casquete) o un oligorribonucleótido sin casquete de 46 nucleótidos. Se produjeron los oligorribonucleótidos mediante transcripción *in vitro* usando el kit de transcripción "AmpliScribe T7" (Epicentre Technologies). Como se indica a continuación, el molde de ADN para el transcrito de ARN contenía una sola citosina. Para sintetizar el ARN sin casquete, se incluyeron los cuatro NTP en la reacción de transcripción *in vitro*. Para obtener el ARN con casquete, se reemplazó el GTP por un análogo del casquete, m7G(5')ppp(5')G. Este compuesto, reconocido por la polimerasa, se incorporó al extremo 3' del transcrito naciente durante la etapa de inicio de la transcripción, pero no fue capaz de incorporación durante la etapa de extensión. En consecuencia, el ARN resultante contenía un casquete en su extremo 5'. Las secuencias de los oligorribonucleótidos producidos mediante la reacción de transcripción *in vitro* fueron:

+casquete:

5'-m7GpppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEC ID N° 1)

-casquete:

5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEC ID N° 2)

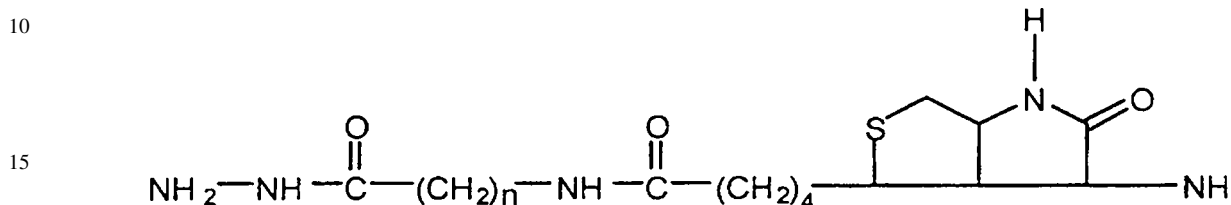
Se disolvieron los oligorribonucleótidos en 9 µl de tampón acetato (acetato de sodio 0,1 M, pH 5,2) y 3 µl de disolución de peryodato de sodio 0,1 M recién preparada. Se incubó la mezcla durante 1 hora en la oscuridad a 4°C o temperatura ambiente. Después de ello, se detuvo la reacción añadiendo 4 µl de etilenglicol al 10%. Se precipitó el producto con etanol, se resuspendió en 10 µl o más de agua o tampón apropiado y se dializó frente a agua.

Los grupos aldehído resultantes pueden acoplarse entonces con moléculas que tienen un grupo amina reactivo, tales como grupos hidrazina, carbazida, tiocarbazida o semicarbazida, para facilitar el enriquecimiento de los extremos 5' de los ARNm. Las moléculas que tienen grupos amina reactivos que son adecuados para uso en la selección de ARNm que tienen extremos 5' intactos incluyen avidina, proteínas, anticuerpos, vitaminas, ligandos capaces de unirse específicamente a moléculas receptoras u oligonucleótidos. El ejemplo 3 describe el acoplamiento del dialdehído resultante con biotina.

Ejemplo 3

Acoplamiento del dialdehído con biotina

- 5 Se disolvió el producto de oxidación obtenido en el ejemplo 2 en 50 μ l de acetato de sodio a un pH entre 5 y 5,2 y 50 μ l de disolución de biotina-hidrazida 0,02 M recién preparada en una mezcla de metoxietanol/agua (1:1) de fórmula:



- 20 En el compuesto usado en estos experimentos, $n = 5$. Sin embargo, se apreciará que pueden usarse también otras hidrazidas comercialmente disponibles, tales como moléculas de la fórmula anterior en la que n varía de 0 a 5.

Se incubó entonces la mezcla durante 2 horas a 37°C. Después de la incubación, se precipitó la mezcla con etanol y se dializó frente a agua destilada.

- 25 El ejemplo 4 demuestra la especificidad de la reacción de biotinilación.

Ejemplo 4

- 30 *Especificidad de biotinilación*

Se evaluó la especificidad de la biotinilación de ARNm con casquete mediante electroforesis en gel de las siguientes muestras:

- 35 Muestra 1. Transcrito *in vitro* de 46 nucleótidos sin casquete preparado como en el ejemplo 2 y marcado con ^{32}pCp como se describe en el ejemplo 1.

- 40 Muestra 2. Transcrito *in vitro* de 46 nucleótidos sin casquete preparado como en el ejemplo 2, marcado con ^{32}pCp como se describe en el ejemplo 1, tratado con la reacción de oxidación del ejemplo 2, y sometido a las condiciones de biotinilación del ejemplo 3.

Muestra 3. Transcrito *in vitro* de 47 nucleótidos con casquete preparado como en el ejemplo 2 y marcado con ^{32}pCp como se describe en el ejemplo 1.

- 45 Muestra 4. Transcrito *in vitro* de 47 nucleótidos sin casquete preparado como en el ejemplo 2, marcado con ^{32}pCp como se describe en el ejemplo 1, tratado con la reacción de oxidación del ejemplo 2, y sometido a las condiciones de biotinilación del ejemplo 3.

- 50 Las muestras 1 y 2 tenían velocidades de migración idénticas, demostrando que los ARN sin casquete no estaban oxidados ni biotinilados. La muestra 3 migraba más lentamente que las muestras 1 y 2, mientras que la muestra 4 exhibía la migración más lenta. La diferencia en migración de los ARN en las muestras 3 y 4 demuestra que los ARN con casquete estaban específicamente biotinilados.

- 55 En algunos casos, los ARNm que tienen extremos 5' intactos pueden enriquecerse uniendo la molécula que contiene un grupo amina reactivo con un sustrato en fase sólida adecuado tal como el interior del recipiente que contiene los ARNm, perlas magnéticas, matrices de cromatografía o membranas de nailon o nitrocelulosa. Por ejemplo, cuando la molécula que tiene un grupo amina reactivo es biotina, el sustrato en fase sólida puede acoplarse con avidina o estreptavidina. Como alternativa, cuando la molécula que tiene el grupo amina reactivo es un anticuerpo o ligando de receptor, el sustrato en fase sólida puede acoplarse con el antígeno o receptor afines. Finalmente, cuando la molécula que tiene un grupo amina reactivo comprende un oligonucleótido, el sustrato en fase sólida puede comprender un oligonucleótido complementario.

- 60 Los ARNm que tienen extremos 5' intactos pueden liberarse de la fase sólida después del procedimiento de enriquecimiento. Por ejemplo, cuando el dialdehído se acopla con biotina-hidrazida y la fase sólida comprende estreptavidina, los ARNm pueden liberarse de la fase sólida simplemente calentando a 95°C en 2% de SDS. En algunos métodos, la molécula que tiene un grupo amina reactivo puede escindirse también de los ARNm que tienen extremos 5' intactos después del enriquecimiento. El ejemplo 5 describe la captura de ARNm biotinilados con perlas recubiertas con estreptavidina y la liberación de los ARNm biotinilados de las perlas después del enriquecimiento.

ES 2 330 903 T3

Ejemplo 5

Captura y liberación de ARNm biotinilados usando perlas recubiertas con estreptavidina

- 5 Se prepararon las perlas magnéticas recubiertas con estreptavidina según las instrucciones del fabricante (CPG Inc., EE.UU.). Se añadieron los ARNm biotinilados a un tampón de hibridación (NaCl 1,5 M, pH 5-6). Después de incubar durante 30 minutos, se retiró el material no unido y no biotinilado. Se lavaron las perlas varias veces con agua con 1% de SDS. Se incubaron las perlas obtenidas durante 15 minutos a 95°C en agua que contenía 2% de SDS.
- 10 El ejemplo 6 demuestra la eficacia con la que se recuperaron los ARNm biotinilados de las perlas recubiertas con estreptavidina.

Ejemplo 6

- 15 *Eficacia de recuperación de ARNm biotinilados*

Se evaluó la eficacia del procedimiento de recuperación del modo siguiente. Se marcaron ARN con ³²pCp, se oxidaron, se biotinilaron y se unieron a perlas recubiertas con estreptavidina como se describe anteriormente. Posteriormente, se incubaron los ARN unidos durante 5, 15 ó 30 min a 95°C en presencia de 2% de SDS.

Se analizaron los productos de reacción mediante electroforesis en geles de poliacrilamida al 12% en condiciones desnaturalizantes (urea 7 M). Se sometieron los geles a autorradiografía. Durante esta manipulación, no se redujeron los enlaces de hidrazona.

25 Se recuperaron cantidades crecientes de ácidos nucleicos a medida que aumentaban los tiempos de incubación en SDS al 2%, demostrando que los ARNm biotinilados se recuperaban eficazmente.

En un método alternativo para obtener ARNm que tienen extremos 5' intactos, se acopla específicamente con ARNm que tienen un casquete intacto un oligonucleótido que se ha derivatizado para contener un grupo amina reactivo. Preferiblemente, el extremo 3' del ARNm se bloquea antes de la etapa en que se unen los grupos aldehído al oligonucleótido derivatizado, como se describe anteriormente, para evitar que el oligonucleótido derivatizado se una al extremo 3' del ARNm. Por ejemplo, puede unirse pCp al extremo 3' del ARNm usando ARN ligasa T4. Sin embargo, como se discute anteriormente, bloquear el extremo 3' del ARNm es una etapa opcional. Pueden prepararse oligonucleótidos derivatizados como se describe a continuación en el ejemplo 7.

Ejemplo 7

- 40 *Derivatización del oligonucleótido*

Se convirtió un oligonucleótido fosforilado en su extremo 3' en una 3' hidrazida en 3' mediante tratamiento con una disolución acuosa de hidrazina o de dihidrazida de fórmula H₂N(R1)NH₂ aproximadamente 1 a 3 M, y a pH 4,5, en presencia de un agente de tipo carbodiimida soluble en agua tal como 1-etil-3-(3-dimetilaminopropil)carbodiimida a una concentración final de 0,3 M a una temperatura de 8°C durante una noche.

Se separó entonces el oligonucleótido derivatizado de los otros agentes y productos usando una técnica estándar para aislar oligonucleótidos.

50 Como se discute anteriormente, los ARNm a enriquecer pueden tratarse para eliminar los grupos 3' OH que pueden estar presentes en los mismos. Esto puede conseguirse mediante ligamiento enzimático de secuencias que carecen de un 3' OH, tales como pCp, como se describe anteriormente en el ejemplo 1. Como alternativa, los grupos 3' OH pueden eliminarse mediante hidrólisis alcalina como se describe en el ejemplo 8 siguiente.

- 55 *Ejemplo 8*

Hidrólisis alcalina de ARNm

60 Los ARNm pueden tratarse con hidrólisis alcalina del modo siguiente. En un volumen total de 100 µl de hidróxido de sodio 0,1 N, se incuban 1,5 µg de ARNm durante 40 a 60 minutos a 4°C. Se neutraliza la disolución con ácido acético y se precipita con etanol.

Después de la eliminación opcional de los grupos 3' OH, se oxidan los grupos diol en los extremos 5' de los ARNm como se describe a continuación en el ejemplo 9.

Ejemplo 9

Oxidación de dioles

5 Se disolvió hasta 1 unidad de DO de ARN en 9 μ l de tampón (acetato de sodio 0,1 M, pH 6-7 o agua) y 3 μ l de disolución de peryodato de sodio 0,1 M recién preparada. Se incubó la reacción durante 1 h en la oscuridad a 4°C o temperatura ambiente. Después de la incubación, se detuvo la reacción añadiendo 4 μ l de etilenglicol al 10%. Después de ello, se incubó la mezcla a temperatura ambiente durante 15 minutos. Después de la precipitación con etanol, se resuspendió el producto en 10 μ l o más de agua o tampón apropiado y se dializó frente a agua.

10 Después de la oxidación de los grupos diol en los extremos 5' de los ARNm, se unió el oligonucleótido derivatizado con los aldehídos resultantes como se describe en el ejemplo 10.

15 Ejemplo 10

Reacción de aldehídos con oligonucleótidos derivatizados

20 Se disolvió el ARNm oxidado en un medio ácido tal como 50 μ l de acetato de sodio a pH 4-6. Se añadieron 50 μ l de disolución de oligonucleótido derivatizado de tal modo que se obtuviera una relación de ARNm:oligonucleótido derivatizado de 1:20, y se redujo la mezcla con un borohidruro. Se dejó incubar la mezcla durante 2 h a 37°C o durante una noche (14 h) a 10°C. Se precipitó la mezcla con etanol, se resuspendió en 10 μ l o más de agua o tampón apropiado y se dializó frente a agua destilada. Si se desea, el producto resultante puede analizarse usando electroforesis en gel de acrilamida, análisis de HPLC u otras técnicas convencionales.

25 Después de la unión del oligonucleótido derivatizado a los ARNm, puede efectuarse una reacción de transcripción inversa como se describe en el ejemplo 11 siguiente.

30 Ejemplo 11

Transcripción inversa de ARNm

35 Se derivatizó un oligodesoxirribonucleótido del modo siguiente. Se disolvieron 3 unidades de DO de un oligodesoxirribonucleótido de secuencia ATCAAGAATTCGACGAGACCATTA (SEC ID N° 3) que tiene extremos 5'-OH y 3'-P en 70 μ l de una disolución de hidroxibenzotriazol 1,5 M, pH 5,3, preparada en dimetilformamida/agua (75:25) que contiene 2 μ g de 1-etil-3-(3-dimetilaminopropil)carbodiimida. Se incubó la mezcla durante 2 h 30 min a 22°C. Se precipitó entonces la mezcla dos veces con LiClO₄/acetona. Se resuspendió el sedimento en 200 μ l de hidrazina 0,25 M y se incubó a 8°C durante 3 a 14 h. Después de la reacción de hidrazina, se precipitó la mezcla dos veces con LiClO₄/acetona.

45 Se extrajeron los ARN mensajeros para transcripción inversa de bloques de placenta de 2 cm de lado que se habían almacenado a -80°C. Se extrajo el ARNm usando técnicas de fenol ácido convencionales. Se usó cromatografía en oligo-DT para purificar los ARNm. Se comprobó la integridad de los ARNm mediante transferencia Northern.

Se oxidaron los grupos diol en 7 μ g de ARNm de placenta como se describe anteriormente en el ejemplo 9. Se unió el oligonucleótido derivatizado con los ARNm como se describe en el ejemplo 10 anterior, excepto porque se reemplazó la etapa de precipitación por una etapa de cromatografía de exclusión para retirar los oligodesoxirribonucleótidos derivatizados que no se habían unido a los ARNm. Se efectuó la cromatografía de exclusión del modo siguiente:

50 Se equilibraron 10 ml de gel AcA34 (BioSeptra n° 230151) en 50 ml de una disolución de Tris 10 mM, pH 8,00, NaCl 300 mM, EDTA 1 mM y 0,05% de SDS. Se dejó sedimentar la mezcla. Se eliminó el sobrenadante y se resuspendió el gel en 50 ml de tampón. Se repitió este procedimiento 2 ó 3 veces.

55 Se introdujo una perla de vidrio (de 3 mm de diámetro) en una pipeta desechable de 2 ml (25 cm de longitud). Se llenó la pipeta con la suspensión de gel hasta la altura del gel estabilizado a 1 cm desde la parte superior de la pipeta. Se equilibró entonces la columna con 20 ml de tampón de equilibrado (Tris-HCl 10 mM, pH 7,4, NaCl 20 mM).

60 Se mezclaron 10 μ l del ARNm que había reaccionado con el oligonucleótido derivatizado en 39 μ l de urea 10 mM y 2 μ l de tampón de glicerol azul, que se había preparado disolviendo 5 mg de azul de bromofenol en glicerol al 60% (v/v), y pasando la mezcla a través de un filtro con un diámetro de filtro de 0,45 μ m.

Se cargó la columna. En cuanto penetró la mezcla, se añadió tampón de equilibrado. Se recogieron fracciones de 100 μ l. El oligonucleótido derivatizado que no se había unido a ARNm apareció en la fracción 16 y fracciones posteriores. Se combinaron las fracciones 3 a 15 y se precipitaron con etanol.

65 Se sembraron los ARNm que habían reaccionado con el oligonucleótido derivatizado sobre una membrana de nailon e hibridaron con una sonda radiactiva usando técnicas convencionales. La sonda radiactiva usada en estas

ES 2 330 903 T3

5 hibridaciones era un oligodesoxirribonucleótido de secuencia TAATGGTCTCGTGCGAATTCTTGAT (SEC ID N° 4) que era anticomplementario del oligonucleótido derivatizado y se marcó en su extremo 5' con ³²P. Se sembró 1/10 del ARNm que había reaccionado con el oligonucleótido derivatizado en dos puntos en la membrana y se visualizó la membrana mediante autorradiografía después de hibridación de la sonda. Se observó una señal, indicando que el oligonucleótido derivatizado se había unido al ARNm.

10 Se transcribieron de modo inverso los 9/10 de los ARNm restantes que habían reaccionado con el oligonucleótido derivatizado del modo siguiente. Se llevó a cabo una reacción de transcripción inversa con transcriptasa inversa siguiendo las instrucciones del fabricante. Para cebar la reacción, se usaron 50 pmol de nonámeros con secuencia aleatoria.

15 Se sembró una porción del ADNc resultante sobre una membrana de nailon cargada positivamente usando métodos convencionales. Se sembraron los ADNc sobre la membrana después de que los heterodúplex de ADNc:ARN se hubieran sometido a una hidrólisis alcalina para eliminar los ARN. Se marcó un oligonucleótido, que tenía una secuencia idéntica a la del oligonucleótido derivatizado, en su extremo 5' con ³²P e hibridó con las transferencias de ADNc usando técnicas convencionales. Se sembraron sobre la membrana ADNc monocatenarios resultantes de la reacción de transcripción inversa. Como controles, la transferencia contenía 1 pmol, 100 fmol, 50 fmol, 10 fmol y 1 fmol respectivamente de un oligodesoxirribonucleótido de control de secuencia idéntica a la del oligonucleótido derivatizado. La señal observada en los puntos que contenían el ADNc indicaba que se habían transcrito de modo inverso 20 aproximadamente 15 fmol del oligonucleótido derivatizado.

Estos resultados demuestran que la transcripción inversa puede efectuarse a través del casquete y, en particular, que la transcriptasa inversa atraviesa el enlace 5'-P-P-5' del casquete de ARN mensajeros eucarióticos.

25 Se usaron los ADNc monocatenarios obtenidos después de la síntesis de primera cadena anterior como molde para reacciones de PCR. Se llevaron a cabo dos tipos de reacciones. En primer lugar, se llevaron a cabo una amplificación específica de los ARNm de alfa-globina, deshidrogenasa, pp15 y factor de elongación E4 usando los siguientes pares de cebadores oligodesoxirribonucleotídicos.

30 alfa-globina

GLO-S: CCG ACA AGA CCA ACG TCA AGG CCG C	(SEC ID N° 5)
35 GLO-As: TCA CCA GCA GGC AGT GGC TTA GGA G 3'	(SEC ID N° 6)

deshidrogenasa

40 3 DH-S: AGT GAT TCC TGC TAC TTT GGA TGG C	(SEC ID N° 7)
3 DH-As: GCT TGG TCT TGT TCT GGA GTT TAG A	(SEC ID N° 8)

45 pp15

PP15-S: TCC AGA ATG GGA GAC AAG CCA ATT T	(SEC ID N° 9)
50 PP15As: AGG GAG GAG GAA ACA GCG TGA GTC C	(SEC ID N° 10)

factor de elongación E4

55 EFA1-S: ATG GGA AAG GAA AAG ACT CAT ATC A	(SEC ID N° 11)
EF1A-As: AGC AGC AAC AAJ CAG GAC AGC ACA G	(SEC ID N° 12)

60 Se llevaron a cabo también amplificaciones no específicas con los oligodesoxirribonucleótidos no codificantes (As) de los pares descritos anteriormente y un cebador elegido de la secuencia del oligodesoxirribonucleótido derivatizado (ATCAAGAATTTCGCACGAGACCATTA) (SEC ID N° 13).

65 Se tiñó con bromuro de etidio un gel de agarosa al 1,5% que contenía las siguientes muestras correspondientes a los productos de PCR de transcripción inversa. (Se usaron 1/20 de los productos de transcripción inversa para cada reacción de PCR).

ES 2 330 903 T3

Muestra 1. Los productos de reacción de una reacción de PCR usando los cebadores de globina de SEC ID N° 5 y 6 en presencia de ADNc.

5 Muestra 2. Los productos de reacción de una reacción de PCR usando los cebadores de globina de SEC ID N° 5 y 6 en ausencia de ADNc añadido.

Muestra 3. Los productos de reacción de una reacción de PCR usando los cebadores de deshidrogenasa de SEC ID N° 7 y 8 en presencia de ADNc.

10 Muestra 4. Los productos de reacción de una reacción de PCR usando los cebadores de deshidrogenasa de SEC ID N° 7 y 8 en ausencia de ADNc añadido.

15 Muestra 5. Los productos de reacción de una reacción de PCR usando los cebadores de pp15 de SEC ID N° 9 y 10 en presencia de ADNc.

Muestra 6. Los productos de reacción de una reacción de PCR usando los cebadores de pp15 de SEC ID N° 9 y 10 en ausencia de ADNc añadido.

20 Muestra 7. Los productos de reacción de una reacción de PCR usando los cebadores de EIE4 de SEC ID N° 11 y 12 en presencia de ADNc añadido.

Muestra 8. Los productos de reacción de una reacción de PCR usando los cebadores de EIE4 de SEC ID N° 11 y 12 en ausencia de ADNc añadido.

25 En las muestras 1, 3, 5 y 7, se observó una banda del tamaño esperado para el producto de PCR, indicando la presencia de la correspondiente secuencia en la población de ADNc.

30 Se llevaron a cabo también reacciones de PCR con los oligonucleótidos no codificantes de los cebadores de globina y deshidrogenasa (SEC ID N° 6 y 8) y un oligonucleótido cuya secuencia corresponde con la del oligonucleótido derivatizado. La presencia de productos de PCR del tamaño esperado en las muestras correspondientes a las muestras 1 y 3 anteriores indicaba que se había incorporado el oligonucleótido derivatizado.

35 Los ejemplos anteriores resumen el procedimiento químico para enriquecer ARNm en aquellos que tienen extremos 5' intactos. Se dan a conocer detalles adicionales respecto a los enfoques químicos para obtener ARNm que tienen extremos 5' intactos en la solicitud internacional n° WO96/34981, publicada el 7 de noviembre de 1996.

40 Pueden utilizarse estrategias basadas en las modificaciones químicas anteriores en la estructura del casquete 5' para generar ADNc que se han seleccionado para incluir los extremos 5' de los ARNm de los que derivan. En una versión de dichos procedimientos, se modifican los extremos 5' de los ARNm como se describe anteriormente. Después de ello, se realiza una reacción de transcripción inversa para extender un cebador complementario del ARNm del extremo 5' del ARNm. Se eliminan los ARN monocatenarios para obtener una población de heterodúplex de ADNc/ARNm en la que el ARNm incluye un extremo 5' intacto. Los heterodúplex resultantes pueden capturarse sobre una fase sólida recubierta con una molécula capaz de interactuar con la molécula usada para derivatizar el extremo 5' del ARNm. Después de ello, se separan las cadenas de los heterodúplex para recuperar las primeras cadenas de ADNc monocatenario que incluyen el extremo 5' del ARNm. Puede procederse entonces a la síntesis de segunda cadena de ADNc usando técnicas convencionales. Por ejemplo, pueden emplearse los procedimientos dados a conocer en el documento WO 96/34981 o en Carninci, P. *et al.* "High-Efficiency Full-Length cDNA Cloning by Biotinylated CAP Trapper". *Genomics* 37: 327-336 (1996) para seleccionar los ADNc que incluyen la secuencia derivada del extremo 5' de la secuencia de codificación del ARNm.

50 Después del ligamiento del marcador oligonucleotídico con el casquete 5' del ARNm, se realiza una reacción de transcripción inversa para extender un cebador complementario del ARNm del extremo 5' del ARNm. Después de la eliminación del componente de ARN del heterodúplex resultante usando técnicas estándar, se realiza una síntesis de segunda cadena de ADNc con un cebador complementario del marcador oligonucleotídico.

55 La Figura 1 resume los procedimientos anteriores para obtener ADNc que se han seleccionado para incluir los extremos 5' de los ARNm de los que derivan.

60 B. Métodos enzimáticos para obtener ARNm que tengan extremos 5' intactos

Otras técnicas para seleccionar ADNc que se extienden hasta el extremo 5' del ARNm del que derivan son totalmente enzimáticas. Se dan a conocer algunas versiones de estas técnicas en Dumas Milne Edwards J.B. (tesis doctoral de la Universidad de París VI, "Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la regulation de l'expression de la tryptophane hydroxylase de rat", 20 de diciembre de 1993), el documento EPO 625572 y Kato *et al.* "Construction of a Human Full-Length cDNA Bank". *Gene* 150: 243-250 (1994).

Brevemente, en dichos enfoques, se trata ARNm aislado con fosfatasa alcalina para retirar los grupos fosfato presentes en los extremos 5' de ARNm incompletos sin casquete. Después de este procedimiento, el casquete presente en ARNm completo se retira enzimáticamente con una enzima eliminadora de casquete tal como polinucleótido cinasa T4 o pirofosfatasa ácida de tabaco. Se liga entonces un oligonucleótido, que puede ser un oligonucleótido de ADN o un oligonucleótido híbrido de ADN-ARN que tiene ARN en su extremo 3', con el fosfato presente en el extremo 5' del ARNm sin casquete usando ARN ligasa T4. El oligonucleótido puede incluir un sitio de restricción para facilitar la clonación de los ADNc después de su síntesis. El ejemplo 12 siguiente describe un método enzimático basado en la tesis doctoral de Dumas.

Ejemplo 12

Enfoque enzimático para obtener 5' EST

Se desfosforilaron 20 µg de ARN de poli A+ usando fosfatasa intestinal de ternero (Biolabs). Después de una extracción con fenol-cloroformo, se hidrolizó la estructura de casquete del ARNm usando la pirofosfatasa ácida de tabaco (purificada como se describe por Shinshi *et al.*, *Biochemistry* 15: 2185-2190, 1976) y se ligó un oligonucleótido hemi 5'-ADN/ARN-3' que tiene un extremo 5' no fosforilado, una extensión de ribofosfato de adenosina en el extremo 3' y un sitio EcoRI cerca del extremo 5' con los extremos 5'P del ARNm usando la ARN ligasa T4 (Biolabs). Los oligonucleótidos adecuados para uso en este procedimiento son preferiblemente de 30-50 bases de longitud. Los oligonucleótidos que tienen un extremo 5' no fosforilado pueden sintetizarse añadiendo un fluorocromo al extremo 5'. La inclusión de una extensión de ribofosfatos de adenosina en el extremo 3' del oligonucleótido aumenta la eficacia de ligamiento. Se apreciará que el oligonucleótido puede contener sitios de clonación distintos de EcoRI.

Después del ligamiento del oligonucleótido al fosfato presente en el extremo 5' del ARNm sin casquete, puede llevarse a cabo la síntesis de primera y segunda cadenas de ADNc usando métodos convencionales o aquellos especificados en el documento EPO 625.572 y Kato *et al.* "Construction of a Human Full-Length cDNA Bank", 150: 243-250 (1994), y Dumas Milne Edwards, *supra*. El ADNc resultante puede ligarse entonces en vectores tales como los dados a conocer en Kato *et al.* "Construction of a Human Full-Length cDNA Bank", *Gene* 150: 243-250 (1994) u otros vectores de ácido nucleico conocidos por los expertos en la técnica usando técnicas tales como las descritas en Sambrook *et al.* "Molecular Cloning: A Laboratory Manual" 2ª Ed., Cold Spring Harbor Laboratory Press, 1989.

II. Caracterización de 5' EST

Se emplearon los enfoques químico y enzimático anteriores para enriquecer ARNm que tienen extremos 5' intactos para obtener 5' EST. En primer lugar, se prepararon ARNm como se describe en el ejemplo 13 siguiente.

Ejemplo 13

Preparación de ARNm

Se adquirieron ARN humanos totales o ARN de poli A+ derivados de 29 tejidos diferentes respectivamente en LABIMO y CLONTECH, y se usaron para generar 44 colecciones de ADNc como se describen a continuación. El ARN adquirido se había aislado de células o tejidos usando extracción con tiocianato de guanidinio ácido-fenol-cloroformo (Chomczynski, P. y Sacchi, N., *Analytical Biochemistry* 162: 156-159, 1987). Se aisló ARN de poli A+ del ARN total (LABIMO) mediante dos pasadas de cromatografía oligo-dT, como se describe por Aviv y Leder (Aviv, H. y Leder, P., *Proc. Natl. Acad. Sci. USA* 69: 1408-1412, 1972) para eliminar el ARN ribosómico.

Se comprobaron la calidad y la integridad del poli A+. Se usaron transferencias Northern hibridadas con una sonda de globina para confirmar que los ARNm no estaban degradados. Se comprobó la contaminación de los ARNm de poli A+ por secuencias ribosómicas usando transferencias de ARN y una sonda derivada de la secuencia de ARN de 28S. Se usaron preparaciones de ARNm con menos de 5% de ARN ribosómico en la construcción de colección. Para evitar construir colecciones con ARN contaminados con secuencias exógenas (procarióticas o fúngicas), se examinó la presencia de secuencias ribosómicas de 16S bacteriano o de dos ARNm altamente expresados usando PCR.

Después de la preparación de los ARNm, se emplearon los procedimientos químicos y/o enzimáticos anteriormente descritos para enriquecer los ARNm que tienen extremos 5' intactos discutidos anteriormente para obtener 5' EST de diversos tejidos. En ambos enfoques, se unió un marcador oligonucleotídico al casquete en los extremos 5' de los ARNm. El marcador oligonucleotídico tenía un sitio EcoRI en el mismo para facilitar procedimientos de clonación posteriores.

Después de la unión del marcador oligonucleotídico al ARNm mediante métodos químicos o enzimáticos, se examinó la integridad del ARNm efectuando una transferencia Northern con 200-500 ng de ARNm usando una sonda complementaria del marcador oligonucleotídico.

Ejemplo 14

Síntesis de ADNc usando moldes de ARNm que tienen extremos 5' intactos

Para los ARNm unidos a marcadores oligonucleotídicos usando métodos tanto químicos como enzimáticos, se efectuó la síntesis de ADNc de primera cadena usando transcriptasa inversa con nonámeros aleatorios como cebadores. Para proteger los sitios EcoRI internos en el ADNc de la digestión en etapas posteriores del procedimiento, se usó dCTP metilado para la síntesis de primera cadena. Después de la retirada del ARN mediante hidrólisis alcalina, se precipitó la primera cadena de ADNc usando isopropanol para eliminar los cebadores residuales.

Tanto para métodos químicos como enzimáticos, se sintetizó la segunda cadena de ADNc con un fragmento Klenow usando un cebador correspondiente al extremo 5' del oligonucleótido ligado descrito en el ejemplo 12. Preferiblemente, el cebador es de 20-25 bases de longitud. Se usó también dCTP metilado para la síntesis de segunda cadena para proteger a sitios EcoRI internos en el ADNc de digestión durante el proceso de clonación.

Después de la síntesis de ADNc, se clonaron los ADNc en pBlueScript como se describe en el ejemplo 15 siguiente.

Ejemplo 15

Inserción de ADNc en BlueScript

Después de la síntesis de segunda cadena, se hicieron romos los extremos del ADNc con ADN polimerasa T4 (Biolabs) y se digirió el ADNc con EcoRI. Puesto que se usó dCTP metilado durante la síntesis de ADNc, el sitio EcoRI presente en el marcador era el único sitio que estaba hemimetilado. En consecuencia, sólo el sitio EcoRI en el marcador oligonucleotídico era susceptible de digestión con EcoRI. Se fraccionó entonces por tamaños el ADNc usando cromatografía de exclusión (AcA, Biosepra). Se combinaron las fracciones correspondientes a ADNc de más de 150 pb y se precipitaron con etanol. Se clonó direccionalmente el ADNc en los extremos SmaI y EcoRI del vector fagémico pBlueScript (Stratagene). Se electroporó la mezcla de ligamiento en bacterias y se propagó mediante selección de antibiótico apropiado.

Se seleccionaron los clones que contenían el marcador oligonucleotídico como se describe en el ejemplo 16 siguiente.

Ejemplo 16

Selección de clones que tienen el marcador oligonucleotídico unido a los mismos

Se purificaron (Qiagen) los ADN de plásmido que contenían colecciones de 5' EST preparadas como se describe anteriormente. Se efectuó una selección positiva de los clones marcados del modo siguiente. Brevemente, en este procedimiento de selección se convirtió el ADN de plásmido en ADN monocatenario usando endonucleasa del gen II del fago F1 en combinación con una exonucleasa (Chang *et al.*, *Gene* 127: 95-8, 1993) tal como exonucleasa III o exonucleasa del gen 6 de T7. Se purificó entonces el ADN monocatenario resultante usando perlas paramagnéticas como se describe por Fry *et al.*, *Biotechniques*, 13: 124-131, 1992. En este procedimiento, se hibridó el ADN monocatenario con un oligonucleótido biotinilado que tiene una secuencia correspondiente al extremo 3' del oligonucleótido descrito en el ejemplo 13. Preferiblemente, el cebador tiene una longitud de 20-25 bases. Se capturaron clones que incluían una secuencia complementaria del oligonucleótido biotinilado mediante incubación con perlas magnéticas recubiertas con estreptavidina seguido de selección magnética. Después de la captura de los clones positivos, se liberó el ADN de plásmido de las perlas magnéticas y se convirtió en ADN bicatenario usando una ADN polimerasa tal como la ThermoSequenase obtenida en Amersham Pharmacia Biotech. Como alternativa, pueden usarse protocolos tales como el kit Gene Trapper (Gibco BRL). Se electroporó entonces el ADN bicatenario en bacterias. Se estimó que el porcentaje de clones positivos que tienen el oligonucleótido marcador 5' está típicamente en el intervalo entre 90 y 98% usando análisis de transferencia puntual.

Después de la electroporación, se ordenaron las colecciones en 384 placas de microvaloración (MTP). Se almacenó una copia de las MTP para necesidades futuras. Se transformaron entonces las colecciones en 96 MTP y se secuenciaron como se describe a continuación.

Ejemplo 17

Secuenciación de insertos en clones seleccionados

Se amplificaron en primer lugar insertos de plásmido por PCR en termocicladores PE 9600 (Perkin-Elmer), usando cebadores SETA-A y SETA-B estándar (Genset SAI, AmpliTaqGold (Perkin-Elmer), dNTP (Boehringer), tampón y condiciones de ciclación como se recomiendan por la compañía Perkin-Elmer.

ES 2 330 903 T3

Se secuenciaron entonces los productos de PCR usando secuenciadores ABI Prism 377 automáticos (Perkin Elmer, Applied Biosystems Division, Foster City, CA). Se efectuaron reacciones de secuenciación usando termocicladores PE 9600 (Perkin Elmer) con química de tinte-cebador estándar y ThermoSequenase (Amersham Life Science). Los cebadores usados eran T7 o 21M13 (disponibles en Genset SA) según fuera apropiado. Se marcaron los cebadores con los tintes JOE, FAM, ROX y TAMRA. Se adquirieron los dNTP y ddNTP usados en las reacciones de secuenciación en Boehringer. El tampón de secuenciación, concentraciones de reactivo y condiciones de ciclación eran como se recomiendan por Amersham.

Después de la reacción de secuenciación, se precipitaron las muestras con EtOH, se resuspendieron en tampón de carga de formamida y se cargaron en un gel de acrilamida al 4% estándar. Se efectuó la electroporesis durante 2,5 horas a 3000 V en un secuenciador ABI 377, se recogieron los datos de secuencia y se analizaron usando el software de análisis de secuencia de ADN ABI Prism, versión 2.1.2.

Se transfirieron los datos de secuencia de las 44 colecciones de ADNc preparadas como se describe anteriormente a una base de datos registrada, donde se efectuaron etapas de control de calidad y validación. Un lector de bases registrado ("Trace"), que trabaja usando un sistema Unix, señaló automáticamente los picos sospechosos, teniendo en cuenta la forma de los picos, la resolución entre picos y el nivel de ruido. El lector de bases registrado efectuó también un ajuste automático. Cualquier extensión de 25 o menos bases que tuviera más de 4 picos sospechosos se consideró no fiable y se desechó. Se retiraron automáticamente las secuencias correspondientes a oligonucleótidos de vector de clonación o ligamiento de las secuencias de EST. Sin embargo, las secuencias de EST resultantes pueden contener 1 a 5 bases pertenecientes a las secuencias anteriormente mencionadas en su extremo 5'. En caso necesario, pueden retirarse éstas fácilmente caso por caso.

Después de ello, se transfirieron las secuencias a la base de datos registrada NETGENE™ para análisis adicional como se describe a continuación.

Después de secuenciar como se describe anteriormente, se introdujeron las secuencias de 5' EST en una base de datos registrada llamada NETGENE™ para almacenamiento y manipulación. Se apreciará por los expertos en la técnica que los datos podrían almacenarse y manipularse en cualquier medio que pueda leer y al que pueda acceder un ordenador. Los medios de lectura informática incluyen medios legibles magnéticamente, medios legibles ópticamente o medios legibles electrónicamente. Por ejemplo, los medios legibles informáticamente pueden ser un disco duro, un disquete, una cinta magnética, CD-ROM, RAM o ROM, así como otros tipos de medios conocidos por los expertos en la técnica.

Además, los datos de secuencia pueden almacenarse y manipularse en una variedad de programas de procesamiento de datos en una variedad de formatos. Por ejemplo, los datos de secuencia pueden almacenarse como texto en un archivo de procesamiento de textos, tal como MicrosoftWORD o WORDPERFECT o como un archivo ASCII en una variedad de programas de bases de datos familiares para los expertos en la técnica, tales como DB2, SYBASE u ORACLE.

Los medios legibles informáticamente en los que se almacena la información de secuencia pueden ser en un ordenador personal, una red, un servidor u otros sistemas informáticos conocidos por los expertos en la técnica. El ordenador u otro sistema incluye preferiblemente los medios de almacenamiento descritos anteriormente, y un procesador para acceder y manipular los datos de secuencia.

Una vez se han almacenado los datos de secuencia, pueden manipularse y buscarse en ellos para localizar aquellas secuencias almacenadas que contienen una secuencia de ácido nucleico deseada o que codifican una proteína que tiene un dominio funcional particular. Por ejemplo, la información de secuencia almacenada puede compararse con otras secuencias conocidas para identificar homologías, motivos implicados en la función biológica o motivos estructurales.

Los programas que pueden usarse para buscar o comparar las secuencias almacenadas incluyen las series de programas MacPattern (EMBL), BLAST y BLAST2 (NCBI), programas de herramienta de búsqueda de alineamiento local básicos para comparaciones de nucleótidos (BLASTN) y péptidos (BLASTX) (Altschul *et al.*, *J. Mol. Biol.* 215: 403 (1990)) y FASTA (Pearson y Lipman, *Proc. Natl. Acad. Sci. USA*, 85: 2444 (1988)). Los programas BLAST extienden entonces los alineamientos basándose en criterios de apareamiento y desapareamiento definidos.

Los motivos que pueden detectarse usando los programas anteriores incluyen secuencias que codifican cremalleras de leucina, motivos de hélice-giro-hélice, sitios de glucosilación, sitios de ubiquitinación, hélices alfa y láminas beta, secuencias señal que codifican péptidos señal que dirigen la secreción de las proteínas codificadas, secuencias implicadas en la regulación de la transcripción tales como homeodominios, extensiones ácidas, sitios activos enzimáticos, sitios de unión a sustrato y sitios de escisión enzimática.

Antes de buscar en los ADNc de la base de datos NETGENE™ los motivos de secuencia de interés, se identificaron los ADNc derivados de ARNm que no eran de interés y se eliminaron de una consideración adicional como se describe en el ejemplo 18 siguiente.

Ejemplo 18

Eliminación de secuencias indeseadas de una consideración adicional

- 5 Se identificaron 5' EST en la base de datos NETGENE™ que derivaban de secuencias indeseadas tales como ARN de transferencia, ARN ribosómicos, ARN mitocondriales, ARN procarióticos, ARN fúngicos, secuencias Alu, secuencias L1 o secuencias repetidas usando los programas FASTA y BLASTN con los parámetros enumerados en la Tabla II.
- 10 Para eliminar los 5' EST que codifican ARNt de una consideración adicional, se compararon las secuencias de 5' EST con las secuencias de 1190 ARNt conocidos obtenidas de EMBL versión 38, de las cuales 100 eran humanas. Se efectuó la comparación usando FASTA en ambas cadenas de los 5' EST. Se identificaron como ARNt las secuencias que tenían más de 80% de homología durante más de 60 nucleótidos. De las 144.341 secuencias examinadas, se identificaron 26 como ARNt y se eliminaron de una consideración adicional.
- 15 Para eliminar los 5' EST que codifican ARNr de una consideración adicional, se compararon las secuencias de 5' EST con las secuencias de 2497 ARNr conocidos obtenidas de EMBL versión 38, de las cuales 73 eran humanas. Se efectuó la comparación usando BLASTN en ambas cadenas de los 5' EST con el parámetro S = 108. Se identificaron como ARNr secuencias que tenían más de 80% de homología durante extensiones de más de 40 nucleótidos. De las 144.341 secuencias examinadas, se identificaron 3.312 como ARNr y se eliminaron de una consideración adicional.
- 20 Para eliminar los 5' EST que codifican ARNmt de una consideración adicional, se compararon las secuencias de 5' EST con las secuencias de los dos genomas mitocondriales conocidos para los que están disponibles las secuencias genómicas enteras y se transcribieron todas las secuencias de estos genomas mitocondriales incluyendo ARNt, ARNr y ARNm para un total de 38 secuencias. Se efectuó la comparación usando BLASTN en ambas cadenas de los 5' EST con el parámetro S = 108. Se identificaron como ARNmt secuencias que tenían más de 80% de homología durante extensiones de más de 40 nucleótidos. De las 144.341 secuencias examinadas, se identificaron 6.110 como ARNmt y se eliminaron de una consideración adicional.
- 30 Se eliminaron de una consideración adicional las secuencias que podrían haber resultado de contaminantes exógenos comparando las secuencias de 5' EST con la versión 46 de las divisiones bacteriana y fúngica de EMBL usando BLASTN con el parámetro S = 144. Se identificaron como contaminantes exógenos todas las secuencias que tenían más de 90% de homología durante al menos 40 nucleótidos. De las 42 colecciones de ADNc examinadas, los porcentajes medios de secuencias procarióticas y fúngicas contenidos en las mismas eran de 0,2% y 0,5%, respectivamente.
- 35 Entre estas secuencias, sólo una no pudo identificarse como una secuencia específica de hongos. Las otras eran secuencias fúngicas o procarióticas que tenían homologías con secuencias de vertebrados o incluían secuencias repetidas que no se habían enmascarado durante la comparación electrónica.
- 40 Además, se compararon los 5' EST con 6093 secuencias Alu y 1115 secuencias L1 para enmascarar de una consideración adicional los 5' EST que contienen dichas secuencias repetidas. Se eliminaron también de una consideración adicional los 5' EST que incluyen las repeticiones THE y MER, secuencias SSTR o repeticiones de satélite, microsatélite o teloméricas. De media, un 11,5% de las secuencias de las colecciones contenían secuencias repetidas. De este 11,5%, un 7% contenía repeticiones Alu, un 3,3% contenía repeticiones L1 y el 1,2% restante derivaba de los otros tipos de secuencias repetitivas que se examinaron. Estos porcentajes son consistentes con aquellos encontrados en colecciones de ADNc preparadas por otros grupos. Por ejemplo, las colecciones de ADN de Adams *et al.* contenían entre 0% y 7,4% de repeticiones Alu dependiendo de la fuente de ARN que se usara para preparar la colección de ADNc (Adams *et al.*, *Nature* 377: 174, 1996).
- 45 Se compararon las secuencias de aquellos 5' EST restantes después de la eliminación de las secuencias indeseables con las secuencias de ARNm humanos conocidos para determinar la exactitud de los procedimientos de secuenciación descritos anteriormente.
- 50

Ejemplo 19

- 55 *Medida de la exactitud de secuenciación por comparación con secuencias conocidas*

- Para determinar adicionalmente la exactitud del procedimiento de secuenciación descrito anteriormente, se identificaron las secuencias de 5' EST derivadas de secuencias conocidas y se compararon con secuencias conocidas. En primer lugar, se realizó un análisis FASTA con salientes menores de 5 pb en ambos extremos en los 5' EST para identificar a aquellos que se aparean con una entrada en la base de datos pública de ARNm humano. Se realinearon entonces los 6655 5' EST que se apareaban con un ARNm humano conocido con su ARNm afín y se usó programación dinámica para incluir sustituciones, inserciones y deleciones en la lista de "errores" que serían reconocidos. Se ignoraron los errores que aparecen en las últimas 10 bases de las secuencias de 5' EST para evitar la inclusión de sitios de clonación espurios en el análisis de la exactitud de secuenciación.
- 60
- 65

Este análisis reveló que las secuencias incorporadas a la base de datos NETGENE™ tenían una exactitud de más de 99,5%.

ES 2 330 903 T3

Para determinar la eficacia con la que los procedimientos de selección anteriores seleccionan ADNc que incluyen los extremos 5' de sus correspondientes ARNm, se efectuó el siguiente análisis.

Ejemplo 20

Determinación de la eficacia de la selección de 5' EST

Para determinar la eficacia a la que los procedimientos de selección anteriores aislaban 5' EST que incluían secuencias cercanas al extremo 5' de los ARNm de los que derivaban, se compararon las secuencias de los extremos de los 5' EST que derivaban de los genes de la subunidad α del factor de elongación 1 y de la cadena pesada de ferritina con las secuencias de ADNc conocidas de estos genes. Puesto que los sitios de inicio de la transcripción para la subunidad α de factor de elongación 1 y cadena pesada de ferritina están bien caracterizados, pueden usarse para determinar el porcentaje de 5' EST derivados de estos genes que incluían los sitios de inicio de transcripción auténticos.

Para ambos genes, más de un 95% de los ADNc incluían secuencias cercanas o cadena arriba del extremo 5' de los ARNm correspondientes.

Para extender el análisis de la fiabilidad de los procedimientos de aislamiento de 5' EST a EST en la base de datos NETGENE™, se realizó un análisis similar usando una base de datos compuesta por secuencias de ARNm humano extraídas de la base de datos GenBank versión 97 para comparación. Para aquellos 5' EST derivados de ARNm incluidos en la base de datos GenBank, más de un 85% tenían sus extremos 5' cercanos a los extremos 5' de la secuencia conocida. Como algunas de las secuencias de ARNm disponibles en la base de datos GenBank se deducen de secuencias genómicas, un apareamiento de extremo 5' con estas secuencias contará como un apareamiento interno. Por tanto, el método usado aquí subestima el rendimiento de EST que incluyen los extremos 5' auténticos de sus ARNm correspondientes.

Las colecciones de EST preparadas anteriormente incluían múltiples 5' EST derivados del mismo ARNm. Se compararon las secuencias de dichos 5' EST entre sí y se identificaron los 5' EST más largos para cada ARNm. Se ensamblaron los ADNc superpuestos en secuencias contiguas (cóntigos). Se compararon entonces las secuencias continuas resultantes con bases de datos públicas para calibrar su similitud con secuencias conocidas, como se describe en el ejemplo 21 siguiente.

Ejemplo 21

Agrupamiento de los 5' EST y cálculo de los índices de novedad para colecciones de ADNc

Para cada colección de EST secuenciada, se agruparon las secuencias por el extremo 5'. Se comparó cada secuencia de la colección con las otras con BLASTN2 (cadena directa, parámetros S = 107). Se agruparon los EST con pares de segmentos de alta puntuación (HSP) de al menos 25 pb de longitud que tienen un 95% de bases idénticas y que empiezan a menos de 10 pb de cada extremo 5' de EST. Se usó la secuencia más larga encontrada en el agrupamiento como representante del agrupamiento. Se efectuó entonces un agrupamiento global entre colecciones, conduciendo a la definición de supercóntigos.

Para evaluar el rendimiento de las nuevas secuencias en las colecciones de EST, se definió un índice de novedad (IN) como: $IN = 100 \times (\text{número de nuevas secuencias únicas encontradas en la colección} / \text{número total de secuencias de la colección})$. Típicamente, se obtuvo un intervalo de índice de novedad entre 10% y 41% dependiendo del tejido del que se obtuviera la colección de EST. Para la mayoría de las colecciones, se prosiguió la secuenciación aleatoria de colecciones de 5' EST hasta que el índice de novedad alcanzara un 20%.

Después de la caracterización como se describe anteriormente, se examinó la colección de 5' EST en NETGENE™ para identificar aquellos 5' EST que portan secuencias señal potenciales como se describe en el ejemplo 22 siguiente.

Ejemplo 22

Identificación de secuencias señal potenciales en 5' EST

Se examinaron los 5' EST en la base de datos NETGENE™ para identificar aquellos que tenían un marco abierto de lectura (ORF) ininterrumpido mayor de 45 nucleótidos, empezando con un codón ATG y extendiéndose hasta el extremo del EST. Aproximadamente la mitad de las secuencias de ADNc en NETGENE™ contenían dicho ORF. Se buscaron los ORF de estos 5' EST para identificar motivos de señal potenciales usando ligeras modificaciones de los procedimientos dados a conocer en von Heijne, G. A "New Method for Predicting Signal Sequence Cleavage Sites". *Nucleic Acids Res.* 14: 4683-4690 (1986). Se consideró que aquellas secuencias de 5' EST que codifican una extensión de 15 aminoácidos de longitud con una puntuación de al menos 3,5 en la matriz de identificación de péptidos de von Heijne poseían una secuencia señal. Aquellos 5' EST que se apareaban con un ARNm humano o secuencia de EST conocida y tenían un extremo 5' más de 20 nucleótidos cadena abajo del extremo 5' conocido, se excluyeron de un análisis adicional. Los ADNc restantes que tenían secuencias señal en los mismos se incluyeron en una base de datos llamada SIGNALTAG™.

ES 2 330 903 T3

Para confirmar la exactitud del método anterior para identificar secuencias señal, se efectuó el análisis del ejemplo 23.

5 Ejemplo 23

Confirmación de la exactitud de la identificación de secuencias señal potenciales en 5' EST

Se evaluó la exactitud del procedimiento anterior para identificar secuencias señal que codifican péptidos señal aplicando el método a los 43 aminoácidos aminoterminales de todas las proteínas SwissProt humanas. Se comparó la puntuación de von Heijne calculada para cada proteína con la caracterización conocida de la proteína por ser una proteína secretada o una proteína no secretada. De esta manera, pudo calcularse el número de proteínas no secretadas que tenían una puntuación mayor de 3,5 (falsos positivos) y el número de proteínas secretadas que tenían una puntuación menor de 3,5 (falsos negativos).

Usando los resultados del análisis anterior, se calculó la probabilidad de que un péptido codificado por la región 5' del ARNm sea de hecho un péptido señal genuino basándose en su puntuación de von Heijne, basándose en la suposición de que un 10% de las proteínas humanas se secretan o en la suposición de que un 20% de las proteínas humanas se secretan. Se muestran los resultados de este análisis en las Figuras 2 y 3.

Usando el método anterior de identificación de proteínas secretadas, se obtuvieron 5' EST de glucagón humano, precursor de monocina inducido por interferón gamma, proteína similar a ciclofilina secretada, pleyotropina humana y precursor de biotinidasa humana, todos los cuales son polipéptidos que se sabe que se secretan. Por tanto, el método anterior identificó exitosamente a aquellos 5' EST que codifican un péptido señal.

Para confirmar que el péptido señal codificado por los 5' EST funciona realmente como péptido señal, pueden clonarse las secuencias señal de los 5' EST en un vector diseñado para la identificación de péptidos señal. Algunos vectores de identificación de péptido señal se diseñan para conferir la capacidad de crecer en medio selectivo en células hospedadoras que tienen una secuencia señal insertada operativamente en el vector. Por ejemplo, para confirmar que un 5' EST codifica un péptido señal genuino, puede insertarse la secuencia señal del 5' EST cadena arriba y en fase con una forma no secretada del gen de invertasa de levadura en vectores de selección de péptido señal tales como los descritos en la patente de EE.UU. n° 5.536.637. El crecimiento de células hospedadoras que contienen vectores de selección de secuencia señal que tienen la secuencia señal de 5' EST insertada en los mismos confirma que el 5' EST codifica un péptido señal genuino.

Como alternativa, puede confirmarse la presencia de un péptido señal mediante la clonación de ADNc extendidos obtenidos usando los EST en vectores de expresión tales como pXT1 (como se describe a continuación), o construyendo vectores de gen informador de secuencia señal promotores que codifican proteínas de fusión entre el péptido señal y una proteína informadora ensayable. Después de la introducción de estos vectores en una célula hospedadora adecuada, tal como células CDS o células NIH 3T3, puede recolectarse el medio de crecimiento y analizarse la presencia de la proteína secretada. Se compara el medio de estas células con el medio de células que contienen vectores que carecen de la secuencia señal o del inserto de ADNc extendido para identificar vectores que codifican un péptido señal funcional o una proteína secretada auténtica.

Aquellos 5' EST que codificaban un péptido señal, como se determina mediante el método del ejemplo 22 anterior, se agruparon adicionalmente en cuatro categorías basándose en su homología con secuencias conocidas. Se describe la categorización de los 5' EST en el ejemplo 24 siguiente.

50 Ejemplo 24

Categorización de 5' EST que codifican un péptido señal

Aquellos 5' EST que tienen una secuencia que no se aparee con ninguna secuencia de vertebrado conocida ni ninguna secuencia de EST conocido públicamente se designaron como "nuevos". De las secuencias en la base de datos SIGNALTAG™, 947 de los 5' EST que tienen una puntuación de von Heijne de al menos 3,5 entraban dentro de esta categoría.

Aquellos 5' EST que tienen una secuencia que no se aparee con ninguna secuencia de vertebrado, pero que se aparee con un EST conocido públicamente, se designaron "EST-ext", a condición de que la secuencia EST conocida se extendiera al menos 40 nucleótidos en la dirección 5'. De las secuencias en la base de datos SIGNALTAG™, 150 de los 5' EST que tienen una puntuación de von Heijne de al menos 3,5 entraban dentro de esta categoría.

Aquellos EST que no se aparean con ninguna secuencia de vertebrado, pero que se aparean con un EST conocido públicamente sin extenderse el EST conocido al menos 40 nucleótidos en la dirección 5', se designaron "EST". De las secuencias en la base de datos SIGNALTAG™, 599 de los 5' EST que tienen una puntuación de von Heijne de al menos 3,5 entraban dentro de esta categoría.

Aquellos 5' EST que se aparean con una secuencia de ARNm humano, pero que extienden la secuencia conocida al menos 40 nucleótidos en la dirección 5', se designaron "VERT-ext". De las secuencias en la base de datos SIGNALTAG™, 23 de los 5' EST que tienen una puntuación de von Heijne de al menos 3,5 entraban dentro de esta categoría. Se incluía en esta categoría un 5' EST que extendía la secuencia conocida del ARNm de translocasa humana más de 200 bases en la dirección 5'. Se identificó también un 5' EST que extendía la secuencia de un gen supresor tumoral humano en la dirección 5'.

La Figura 4 muestra la distribución de 5' EST en cada categoría y el número de 5' EST en cada categoría que tienen una puntuación de von Heijne mínima dada.

Se categorizó cada uno de los 5' EST basándose en el tejido del que se obtuvo su ARNm correspondiente, como se describe a continuación en el ejemplo 25.

Ejemplo 25

Categorización de los patrones de expresión

La Figura 5 muestra los tejidos de los que se obtuvieron los ARNm correspondientes a los 5' EST en cada una de las categorías descritas anteriormente.

Además de categorizar los 5' EST por el tejido del que se obtuvo la colección de ADNc en la que se identificaron por primera vez, pueden determinarse los patrones de expresión espacial y temporal de los ARNm correspondientes a los 5' EST, así como sus niveles de expresión, como se describe en el ejemplo 26 siguiente. La caracterización de los patrones de expresión espacial y temporal de estos ARNm es útil para construir vectores de expresión capaces de producir un nivel deseado de producto génico de la manera deseada desde un punto de vista espacial o temporal, como se discutirá con más detalle a continuación.

Además, pueden identificarse también los 5' EST cuyos ARNm correspondientes están asociados a estados patológicos. Por ejemplo, una enfermedad particular puede ser el resultado de la falta de expresión, sobreexpresión o subexpresión de un ARNm correspondiente a un 5' EST. Al comparar los patrones y cantidades de expresión de ARNm en muestras tomadas de individuos sanos con aquellas de individuos que padecen una enfermedad particular, pueden identificarse 5' EST responsables de la enfermedad.

Se apreciará que los resultados de los procedimientos de caracterización anteriores para 5' EST se aplican también a ADNc extendidos (obtenibles como se describe a continuación) que contienen secuencias adyacentes a los 5' EST. Se apreciará también que si se desea retrasar la caracterización hasta haber obtenido los ADNc extendidos en lugar de caracterizar los EST mismos, pueden aplicarse los procedimientos de caracterización anteriores para caracterizar los ADNc extendidos después de su aislamiento.

Ejemplo 26

Evaluación de los niveles y patrones de expresión de ARNm correspondientes a 5' EST o ADNc extendidos

Pueden analizarse los niveles y patrones de expresión de ARNm correspondientes a 5' EST o ADNc extendidos (obtenibles como se describe a continuación) mediante hibridación en disolución con sondas largas como se describe en la solicitud de patente internacional nº WO 97/05277. Brevemente, se inserta un 5' EST, un ADNc extendido o fragmento del mismo correspondiente al gen que codifica el ARNm a caracterizar en un sitio de clonación inmediatamente cadena abajo de un promotor de ARN polimerasa de bacteriófago (T3, T7 o SP6) para producir ARN no codificante. Preferiblemente, el 5' EST o ADNc extendido tiene 100 o más nucleótidos. Se linealiza el plásmido y se transcribe en presencia de ribonucleótidos que comprenden ribonucleótidos modificados (concretamente, biotina-UTP y DIG-UTP). Se hibrida un exceso de este ARN doblemente marcado en disolución con ARNm aislado a partir de células o tejidos de interés. Se efectúan las hibridaciones en condiciones rigurosas estándar (40-50°C durante 16 horas en un tampón de 80% de formamida, NaCl 0,4 M, pH 7-8). Se retira la sonda no hibridada mediante digestión con ribonucleasas específicas de ARN monocatenario (concretamente, ARNasas CL3, T1, Phy M, U2 o A). La presencia de biotina-UTP posibilita la captura del híbrido en una placa de microvaloración recubierta con estreptavidina. La presencia de la modificación DIG posibilita que el híbrido sea detectado y cuantificado por ELISA usando un anticuerpo anti-DIG acoplado a fosfatasa alcalina.

Los 5' EST, ADNc extendidos o fragmentos de los mismos pueden marcarse también con secuencias nucleotídicas para el análisis en serie de la expresión génica (SAGE) como se da a conocer en la solicitud de patente de R.U. nº 2.305.241 A. En este método, se preparan ADNc de una célula, tejido, organismo u otra fuente de ácido nucleico para la que se desea determinar los patrones de expresión génica. Se separan los ADNc resultantes en dos conjuntos. Se escinden los ADNc en cada conjunto con una primera endonucleasa de restricción llamada "enzima de anclaje", que tiene un sitio de reconocimiento que es probable que esté presente al menos una vez en la mayoría de los ADNc. Se aíslan los fragmentos que contienen la región más 5' ó 3' del ADNc escindido mediante unión a un medio de captura tal como perlas recubiertas con estreptavidina. Se liga un primer ligador oligonucleotídico que tiene una primera

secuencia para la hibridación de un cebador de amplificación y un sitio de restricción interno para una “endonucleasa marcadora” con los ADNc digeridos en el primer conjunto. La digestión con la segunda endonucleasa produce fragmentos “marcadores” cortos de los ADNc.

- 5 Se liga un segundo oligonucleótido que tiene una segunda secuencia para la hibridación de un cebador de amplificación y un sitio de restricción interno con los ADNc digeridos en el segundo conjunto. Se digieren también los fragmentos de ADNc en el segundo conjunto con la “endonucleasa marcadora” para generar fragmentos “marcadores” cortos derivados de los ADNc del segundo conjunto. Los “marcadores” resultantes de la digestión del primer y segundo conjuntos con la enzima de anclaje y la endonucleasa de anclaje se ligan entre sí para producir “dimarcadores”.
- 10 En algunas realizaciones, los dimarcadores se concatamerizan para producir productos de ligamiento que contienen 2 a 200 dimarcadores. Se determinan entonces las secuencias marcadoras y se comparan con las secuencias de los 5' EST o ADNc extendidos para determinar cuál 5' EST o ADNc se expresa en la célula, tejido, organismo u otra fuente de ácidos nucleicos de la que derivaban los marcadores. De este modo, se obtiene el patrón de expresión de 5' EST o ADNc extendidos en la célula, tejido, organismo u otra fuente de ácidos nucleicos.

- 15 Puede efectuarse también el análisis cuantitativo de la expresión génica usando matrices. Como se usa en la presente memoria, el término matriz significa una disposición monodimensional, bidimensional o multidimensional de ADNc completos (concretamente, ADNc extendidos que incluyen la secuencia de codificación del péptido señal, la secuencia de codificación de la proteína madura y un codón de terminación), ADNc extendidos, 5' EST o fragmentos de los ADNc completos, ADNc extendidos o 5' EST de suficiente longitud para permitir una detección específica de la expresión génica. Preferiblemente, los fragmentos son de al menos 15 nucleótidos de longitud. Más preferiblemente, los fragmentos son de al menos 100 nucleótidos de longitud. Más preferiblemente, los fragmentos son de más de 100 nucleótidos de longitud. En algunas realizaciones, los fragmentos pueden ser de más de 500 nucleótidos de longitud.

- 25 Por ejemplo, puede efectuarse el análisis cuantitativo de la expresión génica con ADNc completos, ADNc extendidos, 5' EST o fragmentos de los mismos en una micromatriz de ADN complementario como se describe por Schena *et al.* (*Science* 270: 467-470, 1995; *Proc. Natl. Acad. Sci. U.S.A.* 93: 10614-10619, 1996). Los ADNc completos, ADNc extendidos, 5' EST o fragmentos de los mismos se amplifican mediante PCR y se disponen desde placas de microvaloración de 96 pocillos sobre portaobjetos de microscopio sililados usando robótica de alta velocidad. Se incuban las matrices impresas en una cámara húmeda para permitir la rehidratación de los elementos de la matriz y se aclaran, una vez con SDS al 0,2% durante 1 min, dos veces con agua durante 1 min y una vez durante 5 min con disolución de borohidruro de sodio. Se sumergen las matrices en agua durante 2 min a 95°C, se transfieren a SDS al 0,2% durante 1 min, se aclaran dos veces con agua, se secan al aire y se almacenan en la oscuridad a 25°C.

- 35 Se aísla ARNm de célula o tejido o se obtiene comercialmente y se preparan sondas mediante una sola ronda de transcripción inversa. Se hibridan las sondas con micromatrices de 1 cm² bajo un cubreobjetos de vidrio de 14 x 14 mm durante 6-12 horas a 60°C. Se lavan las matrices durante 5 min a 25°C con tampón de lavado de bajo rigor (1 x SSC/0,2% de SDS), después durante 10 min a temperatura ambiente en tampón de lavado de alto rigor (0,1 x SSC/0,2% de SDS). Se exploran las matrices en 0,1 x SSC usando un dispositivo de exploración por láser de fluorescencia equipado con un grupo de filtros usual. Se obtienen medidas de expresión diferencial exactas tomando la media de las relaciones de dos hibridaciones independientes.

- Puede efectuarse también el análisis cuantitativo de la expresión de genes con ADNc completos, ADNc extendidos, 5' EST o fragmentos de los mismos en una micromatriz de ADN complementario como se describe por Pietu *et al.* (*Genome Research* 6: 492-503, 1996). Se amplifican los ADNc completos, ADNc extendidos, 5' EST o fragmentos de los mismos por PCR y se siembran sobre membranas. Se marcan entonces los ARNm originarios de diversos tejidos o células con nucleótidos radiactivos. Después de la hibridación y lavado en condiciones controladas, se detectan los ARNm hibridados mediante la formación de imágenes por fósforo o autorradiografía. Se efectúan experimentos por duplicado y se efectúa entonces un análisis cuantitativo de ARNm expresados diferencialmente.

- 50 Como alternativa, puede hacerse el análisis de expresión de los 5' EST o ADNc extendidos mediante matrices nucleotídicas de alta densidad como se describen por Lockhart *et al.* (*Nature Biotechnology* 14: 1675-1680, 1996) y Sosnowsky *et al.* (*Proc. Natl. Acad. Sci. USA* 94: 1119-1123, 1997). Se sintetizan directamente sobre el chip oligonucleótidos de 15-50 nucleótidos que corresponden a secuencias de 5' EST o ADNc extendidos (Lockhart *et al.*, *supra*) o se sintetizan y después se dirigen al chip (Sosnowski *et al.*, *supra*). Preferiblemente, los oligonucleótidos son de aproximadamente 20 nucleótidos de longitud.

- Se sintetizan sondas de ADNc marcadas con un compuesto apropiado, tal como biotina, digoxigenina o tinte fluorescente, a partir de la población de ARNm apropiada y se fragmentan entonces aleatoriamente a un tamaño medio de 50 a 100 nucleótidos. Se hibridan entonces dichas sondas con el chip. Después de lavar como se describe en Lockhart *et al.*, *supra* y la aplicación de diferentes campos eléctricos (Sosnowsky *et al.*, *Proc. Natl. Acad. Sci.* 94: 1119-1123), se detectan y cuantifican los tintes o compuestos marcadores. Se efectúan hibridaciones por duplicado. El análisis comparativo de la intensidad de la señal originaria de sondas de ADNc en el mismo oligonucleótido diana en diferentes muestras de ADNc indica una expresión diferencial del ARNm correspondiente al 5' EST o ADNc extendido a partir del que se ha diseñado la secuencia oligonucleotídica.

III. Uso de 5' EST para clonar ADNc extendidos y para clonar los correspondientes ADN genómicos

Una vez se han seleccionado los 5' EST que incluyen el extremo 5' de los correspondientes ARNm usando los procedimientos descritos anteriormente, pueden utilizarse para aislar ADNc extendidos que contienen secuencias adyacentes a los 5' EST. Los ADNc extendidos pueden incluir la secuencia de codificación completa de la proteína codificada por el correspondiente ARNm, incluyendo el sitio de inicio de la traducción auténtico, la secuencia señal y la secuencia que codifica la proteína madura que permanece después de la escisión del péptido señal. Dichos ADNc extendidos se refieren en la presente memoria como "ADNc completos". Como alternativa, los ADNc extendidos pueden incluir sólo la secuencia que codifica la proteína madura que permanece después de la escisión del péptido señal, o sólo la secuencia que codifica el péptido señal.

El ejemplo 27 siguiente describe un método general para obtener ADNc extendidos. El ejemplo 28 siguiente describe la clonación y secuenciación de varios ADNc extendidos, incluyendo ADNc extendidos que incluyen la secuencia de codificación completa y el extremo 5' auténtico del correspondiente ARNm de varias proteínas secretadas.

Pueden usarse también los métodos de los ejemplos 27, 28 y 29 para obtener ADNc extendidos que codifican menos de la secuencia de codificación entera de las proteínas secretadas codificadas por los genes correspondientes a los 5' EST. En algunos aspectos, los ADNc extendidos aislados usando estos métodos codifican al menos 10 aminoácidos de la proteína codificada por la secuencia de SEC ID N° 45. En aspectos adicionales, los ADNc extendidos codifican al menos 20 aminoácidos de las proteínas codificadas por la secuencia de SEC ID N° 45. En aspectos adicionales, los ADNc extendidos codifican al menos 30 aminoácidos de la secuencia de SEC ID N° 45. En una realización preferida, los ADNc extendidos codifican una secuencia de proteína completa que incluye la secuencia de codificación de proteína de SEC ID N° 45.

Ejemplo 27

Método general para usar 5' EST para clonar y secuenciar ADNc extendidos

Se ha usado el siguiente método general para aislar rápida y eficazmente ADNc extendidos, incluyendo la secuencia adyacente a las secuencias de los 5' EST usados para obtenerlos. Este método puede aplicarse para obtener ADNc extendidos para cualquier 5' EST en la base de datos NETGENETM, incluyendo aquellos 5' EST que codifican proteínas secretadas. El método se resume en la Figura 6.

1. Obtención de ADNc extendidos

a) Síntesis de primera cadena

El método aprovecha la secuencia 5' conocida del ARNm. Se realiza una reacción de transcripción inversa en ARNm purificado con un cebador poli 14dT que contiene una secuencia de 49 nucleótidos en su extremo 5' que permite la adición de una secuencia conocida en el extremo del ADNc que corresponde al extremo 3' del ARNm. Por ejemplo, el cebador puede tener la siguiente secuencia: 5'-ATC GTT GAG ACT CGT ACC AGC AGA GTC ACG AGA GAG ACT ACA CGG TAC TGG TTT TTT TTT TTT TTVN -3' (SEC ID N° 14. Los expertos en la técnica apreciarán que pueden añadirse también otras secuencias a la secuencia de poli dT y usarse para cebar la síntesis de primera cadena. Usando este cebador y una transcriptasa inversa tal como la enzima Superscript II (Gibco BRL) o ARNasa H Minus M-MLV (Promega), se genera un transcrito inverso anclado en el sitio 3' poli A de los ARN.

Después de la retirada del ARNm hibridado con la primera cadena de ADNc mediante hidrólisis alcalina, se eliminan los productos de la hidrólisis alcalina y el cebador poli dT residual con una columna de exclusión tal como una matriz AcA34 (Biosepra) como se explica en el ejemplo 11.

b) Síntesis de segunda cadena

Se diseñan un par de cebadores anidados en cada extremo basándose en la secuencia 5' conocida de 5' EST y el extremo 3' conocido añadido por el cebador poli dT usado en la síntesis de primera cadena. Los software usados para diseñar cebadores están basados en el contenido de GC y las temperaturas de fusión de oligonucleótidos, tales como OSP (Illier y Green, *PCR Meth. Appl.* 1: 124-128, 1991), o están basados en el método de disparidad de la frecuencia de octámeros (Griffais *et al. Nucleic Acids Res.* 19: 3887-3891, 1991 tales como PC-Rare (<http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html>).

Preferiblemente, los cebadores anidados en el extremo 5' están separados entre sí por 4 a 9 bases. Las secuencias cebadoras 5' pueden seleccionarse para tener temperaturas de fusión y especificidades adecuadas para uso en PCR.

ES 2 330 903 T3

Preferiblemente, los cebadores anidados en el extremo 3' están separados entre sí por 4 a 9 bases. Por ejemplo, los cebadores 3' anidados pueden tener las siguientes secuencias: (5'- CCA GCA GAG TCA CGA GAG AGA CTA CAC GG -3' (SEC ID N° 15), y 5'- CAC GAG AGA GAC TAC ACG GTA CTG G -3' (SEC ID N° 16). Se seleccionaron estos cebadores porque tienen temperaturas de fusión y especificidades compatibles con su uso en PCR. Sin embargo, los expertos en la técnica apreciarán que pueden usarse también como cebadores otras secuencias.

Se efectúa el primer proceso de PCR de 25 ciclos usando la mezcla de polimerasa Tth Advantage (Clontech) y el cebador externo de cada uno de los pares anidados. Se efectúa entonces una segunda PCR de 20 ciclos usando la misma enzima y el cebador interno de cada uno de los pares anidados en 1/2500 del primer producto de PCR. Después de ello, se retiran cebadores y nucleótidos.

2. Secuenciación de ADNc extendidos completos o fragmentos de los mismos

Debido a la falta de limitaciones de posición en el diseño de cebadores anidados 5' compatibles para uso en PCR usando el software OSP, se obtienen amplicones de dos tipos. Preferiblemente, el segundo cebador 5' está localizado cadena arriba del codón de inicio de la traducción, proporcionando así un producto de PCR anidado que contiene la secuencia de codificación completa. Dicho ADNc extendido completo experimenta un procedimiento de clonación directa como se describe en la sección a siguiente. Sin embargo, en algunos casos, el segundo cebador 5' se localiza cadena abajo del codón de inicio de la traducción, proporcionando así un producto de PCR que contiene sólo parte del ORF. Dichos productos de PCR incompletos se someten a un procedimiento modificado descrito en la sección siguiente.

a) Productos de PCR anidados que contienen ORF completos

Cuando el producto de PCR anidado resultante contiene la secuencia de codificación completa, como se predice por la secuencia 5' EST, se clona en un vector apropiado tal como pED6dpc2, como se describe en la sección 3.

b) Productos de PCR anidados que contienen ORF incompletos

Cuando el amplicón no contiene la secuencia de codificación completa, son necesarias etapas intermedias para obtener tanto la secuencia de codificación completa como un producto de PCR que contenga la secuencia de codificación completa. La secuencia de codificación completa puede ensamblarse a partir de varias secuencias parciales determinadas directamente a partir de diferentes productos de PCR como se describe en la siguiente sección.

Una vez se ha determinado completamente la secuencia de codificación completa, se diseñan nuevos cebadores compatibles para uso en PCR para obtener amplicones que contengan toda la región de codificación. Sin embargo, en dichos casos, se usan cebadores 3' compatibles para uso en PCR dentro del 3' UTR del correspondiente ARNm, proporcionando así amplicones que carecen de parte de esta región, concretamente, el intervalo de poli A y a veces la señal de poliadenilación, como se ilustra en la Figura 6. Se clonan entonces dichos ADNc extendidos completos en un vector apropiado como se describe en la sección 3.

c) Secuenciación de ADNc extendidos

La secuenciación de los ADNc extendidos se efectúa usando un enfoque de Die Terminator con el kit FS de ADN polimerasa AmpliTaq disponible en Perkin Elmer.

Para secuenciar los fragmentos de PCR, se efectúa un desplazamiento por el cebador usando software tal como OSP para elegir cebadores y un software informático automatizado tal como ASMG (Sutton *et al. Genome Science Technol.* 1: 9-19, 1995) para construir cóntigos de secuencias de desplazamiento que incluyen el marcador 5' inicial usando superposiciones mínimas de 32 nucleótidos. Preferiblemente, el desplazamiento por el cebador se efectúa hasta que se obtienen las secuencias de ADNc completas.

La terminación de la secuenciación de un fragmento de ADNc extendido dado se evalúa del modo siguiente. Puesto que las secuencias localizadas después de un intervalo de poli A son difíciles de determinar con precisión en el caso de productos no clonados, los procesos de secuenciación y desplazamiento por el cebador para productos de PCR se interrumpen cuando se identifica un intervalo de poli A en ADNc extendidos obtenidos como se describe en el caso b. Se compara la longitud de secuencia con el tamaño del producto de PCR anidado como se describe anteriormente. Debido a la exactitud limitada de la determinación del tamaño de producto de PCR por electroforesis en gel, se considera completa una secuencia si el tamaño de la secuencia obtenida es al menos un 70% del tamaño del primer producto de PCR anidado. Si la longitud de la secuencia determinada por el análisis informático no es al menos un 70% de la longitud del producto de PCR anidado, se clonan estos productos de PCR y se determina la secuencia de la inserción. Cuando están disponibles los datos de transferencia Northern, se usa el tamaño del ARNm detectado para un producto de PCR dado para evaluar finalmente que la secuencia es completa. Las secuencias que no satisfacen los criterios anteriores se descartan y se someterán a un nuevo procedimiento de aislamiento.

Se transfieren los datos de secuencia de todos los ADNc extendidos a una base de datos registrada, donde se llevan a cabo controles de calidad y etapas de validación como se describe en el ejemplo 15.

5 3. Clonación de ADNc extendidos completos

Se clona entonces el producto de PCR que contiene la secuencia de codificación completa en un vector apropiado. Por ejemplo, los ADNc extendidos pueden clonarse en el vector de expresión pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) del modo siguiente. Se muestra la estructura de pED6dpc2 en la Figura 7. Se prepara ADN del vector pED6dpc2 con extremos romos efectuando una digestión con EcoRI seguida de una reacción de relleno. Se desfosforila el vector de extremos romos. Después de la retirada de los cebadores de PCR y la precipitación con etanol, se fosforila el producto de PCR que contiene la secuencia de codificación completa o el ADNc extendido obtenido como se describe anteriormente con una cinasa posteriormente retirada por extracción con fenol-Sevag y precipitación. Se liga entonces el ADNc extendido bicatenario al vector y se introduce el plásmido de expresión resultante en células hospedadoras apropiadas.

Puesto que los productos de PCR obtenidos como se describe anteriormente son moléculas de extremos romos que pueden clonarse en cualquier dirección, se determina la orientación de varios clones para cada producto de PCR. Se ordenan entonces 4 a 10 clones en placas de microvaloración y se someten a una reacción de PCR usando un primer cebador localizado en el vector cercano al sitio de clonación y un segundo cebador localizado en la porción del ADNc extendido correspondiente al extremo 3' del ARNm. Este segundo cebador puede ser el cebador no codificante usado en PCR anclado en el caso de clonación directa (caso a) o el cebador no codificante localizado dentro del 3' UTR en el caso de clonación indirecta (caso b). Se conservan y secuencian los clones en que el codón de inicio del ADNc extendido está ligado operativamente al promotor en el vector para permitir la expresión de la proteína codificada por el ADNc extendido. Además de los extremos de los insertos de ADNc, se secuencian también aproximadamente 50 pb de ADN de vector en cada lado del inserto de ADNc.

Se secuencian entonces completamente los productos de PCR clonados según el procedimiento anteriormente mencionado. En este caso, se efectúa entonces el ensamblaje de cóntigo de fragmentos largos en secuencias de desplazamiento que se han contigado ya para productos de PCR no clonados durante el desplazamiento por el cebador. La secuenciación de los amplicones clonados es completa cuando los cóntigos resultantes incluyen la región de codificación completa así como las secuencias superpuestas con ADN de vector en ambos extremos.

35 4. Análisis informático de ADNc extendido completo

Se someten entonces secuencias de todos los ADNc extendidos completos a análisis adicional como se describe anteriormente y usando los parámetros encontrados en la Tabla II con las siguientes modificaciones. Para examinar las subdivisiones misceláneas de Genbank, se usó FASTA en lugar de BLASTN y 15 nucleótidos de homología fue el límite en lugar de 17. Para la detección de Alu, se usó BLASTN con los siguientes parámetros: S = 72; identidad = 70%; y longitud = 40 nucleótidos. Se buscaron la señal de poliadenilación y la cola de poli A, que no se buscaron para los 5' EST. Para la detección de la señal de poliadenilación, se buscó la señal (AATAAA) con un desapareamiento permisible en los últimos diez nucleótidos precedentes al extremo 5' de poli A. Para el poli A, se buscó una extensión de 8 aminoácidos en los últimos 20 nucleótidos de la secuencia con BLAST2N en la cadena codificante con los siguientes parámetros (W = 6, S = 10, E = 1000 e identidad = 90%). Finalmente, se buscaron secuencias patentadas y homologías de ORF usando, respectivamente, BLASTN y BLASTP en la base de datos GenSEQ (base de datos de Derwent de secuencias nucleotídicas patentadas) y SWISSPROT para ORF con los siguientes parámetros (W = 8 y B = 10). Antes de examinar en los ADNc completos extendidos las secuencias de interés, se buscan ADNc extendidos que no son de interés del modo siguiente.

a) Eliminación de secuencias indeseadas

Aunque se comprobaron los 5' EST para retirar las secuencias contaminantes como se describe en el ejemplo 18, se llevó a cabo una última verificación para identificar secuencias de ADNc extendido derivadas de secuencias indeseadas tales como ARN de vector, ARN transferentes, ARN ribosómicos, ARN mitocondriales, ARN procarióticos y ARN fúngicos usando los programas FASTA y BLASTN en ambas cadenas de ADNc extendidos como se describe a continuación.

Para identificar los ADNc extendidos que codifican ARN de vector, se comparan los ADNc extendidos con las secuencias conocidas de ARN de vector usando el programa FASTA. Se identifican secuencias de ADNc extendidos con más de 90% de homología durante extensiones de 15 nucleótidos como ARN de vector.

Para identificar los ADNc extendidos que codifican ARNt, se compararon las secuencias de ADNc extendido con las secuencias de 1190 ARNt conocidos obtenidas de EMBL versión 38, de las que 100 eran humanas. Se identificaron como ARNt las secuencias de ADNc extendidos que tenían más de 80% de homología durante 60 nucleótidos usando FASTA.

Para identificar los ADNc extendidos que codifican ARNr, se compararon las secuencias de ADNc extendido con las secuencias de 2497 ARNt conocidos obtenidas de EMBL versión 38, de las que 73 eran humanas. Se identificaron como ARNr las secuencias de ADNc extendidos que tienen más de 80% de homología durante extensiones de más de 40 nucleótidos usando BLASTN.

Para identificar los ADNc extendidos que codifican ARNmt, se compararon las secuencias de ADNc extendidos con las secuencias de los dos genomas mitocondriales conocidos para los que están disponibles las secuencias genómicas enteras y se transcribieron todas las secuencias de estos genomas mitocondriales incluyendo ARNt, ARNr y ARNm para un total de 38 secuencias. Se identificaron como ARNmt las secuencias de ADNc extendidos que tenían más de 80% de homología durante extensiones de más de 40 nucleótidos usando BLASTN.

Se identificaron las secuencias que podrían haber resultado de otros contaminantes exógenos comparando secuencias de ADNc extendido con la versión 105 de las divisiones bacteriana y fúngica de Genbank. Se identificaron como contaminantes procarióticos o fúngicos exógenos secuencias de ADNc extendidos que tienen más de 90% de homología durante 40 nucleótidos usando BLASTN.

Además, se buscaron en ADNc extendidos diferentes secuencias repetidas, incluyendo secuencias Alu, secuencias L1, repeticiones THE y MER, secuencias SSTR o repeticiones de satélite, microsatélite o teloméricas. Se identificaron como secuencias repetidas secuencias de ADNc extendidos con más de 70% de homología durante extensiones de 40 nucleótidos usando BLASTN y se enmascararon en procedimientos de identificación adicionales. Además, se señalaron clones que mostraban una homología extensa con repeticiones, concretamente, apareamientos de más de 50 nucleótidos si la homología era de al menos 75% o de más de 40 nucleótidos si la homología era de al menos 85% o de más de 30 nucleótidos si la homología era de al menos 90%.

b) *Identificación de rasgos estructurales*

Los rasgos estructurales, por ejemplo cola de poli A y señal de poliadenilación, de las secuencias de ADNc extendidos completos se determinan posteriormente del modo siguiente.

Se define una cola de poli A como una extensión homopolimérica de al menos 11 A con como máximo una base alternativa en la misma. La búsqueda de la cola de poli A está restringida a los últimos 20 nt de la secuencia y limitada a extensiones de 11 A consecutivos porque las reacciones de secuenciación a menudo no son legibles después de dicha extensión de poli A. Se identifican como colas de poli A extensiones con 100% de homología durante 6 nucleótidos.

Para buscar una señal de poliadenilación, se fija la cola de poli A de la secuencia completa. Se busca en los 50 pb precedentes de la cola de poli A la señal de poliadenilación canónica AAUAAA, permitiendo un desapareamiento para dar cuenta de posibles errores de secuenciación y la variación conocida de la secuencia canónica de la señal de poliadenilación.

c) *Identificación de rasgos funcionales*

Se determinaron posteriormente los rasgos funcionales, por ejemplo ORF y secuencias señal, de las secuencias de ADNc extendidos completos del modo siguiente.

Se buscan en los 3 marcos de cadena superior de los ADNc extendidos los ORF definidos como los fragmentos de longitud máxima que empiezan con un codón de inicio de la traducción y finalizan con un codón de terminación. Se prefieren ORF que codifican al menos 20 aminoácidos.

Se explora entonces en cada ORF encontrado la presencia de un péptido señal en los primeros 50 aminoácidos o, cuando sea apropiado, en regiones más cortas de hasta 20 aminoácidos o menos en el ORF, usando el método de matriz de von Heijne (*Nuc. Acids Res.* 14: 4683-4690 (1986)) y la modificación descrita en el ejemplo 22.

d) *Homología con secuencias nucleotídicas o proteicas*

Se comparan entonces las secuencias de ADNc extendidos completos con secuencias conocidas basadas en nucleótidos o proteínas.

Se comparan las secuencias de ADNc extendidos completos con las siguientes secuencias de ácido nucleico conocidas: secuencias de vertebrados (Genbank), secuencias de EST (Genbank), secuencias patentadas (Geneseq) y secuencias recientemente identificadas (versiones diarias de Genbank) disponibles en el momento de presentación de los documentos de prioridad. Se comparan entonces también secuencias de ADNc completas con las secuencias de una base de datos registrada (secuencias internas de Genset) para encontrar secuencias que se hayan identificado ya por los solicitantes. Se identifican las secuencias de ADNc extendidos completos con más de 90% de homología durante 30 nucleótidos usando BLASTN o BLAST2N como se indica en la Tabla III como secuencias que ya se han descrito. Se

ES 2 330 903 T3

examinan posteriormente las secuencias de vertebrados apareadas usando FASTA; se identifican los ADNc extendidos completos con más de 70% de homología durante 30 nucleótidos como secuencias que ya se han descrito.

Se comparan posteriormente los ORF codificados por ADNc extendidos completos como se definen en la sección c) con secuencias aminoacídicas conocidas encontradas en las bases de datos públicas Swissprot versión CHP, PIR versión PIR y Genpept versión GPEPT usando BLASTP con el parámetro W=8 y permitiendo un máximo de 10 apareamientos. Se reconocen como proteínas ya identificadas las secuencias de ADNc extendidos completos que muestran una homología extensa con secuencias proteicas conocidas.

Además, se comparan los productos de traducción conceptuales de tres marcos de lectura de la cadena superior de los ADNc extendidos completos con secuencias aminoacídicas conocidas públicamente de Swissprot usando BLASTX con el parámetro E=0,001. Se detectan como proteínas ya identificadas las secuencias de ADNc extendidos completos con más de 70% de homología durante extensiones de 30 aminoácidos.

5. Selección de secuencias completas clonadas de la presente invención

Se someten entonces secuencias de ADNc extendidos completos clonados, que ya se han caracterizado mediante el análisis informático anteriormente mencionado, a un procedimiento automático para preseleccionar ADNc extendidos completos que contienen secuencias de interés.

a) Preselección de secuencia automática

Se consideran todos los ADNc extendidos completos clonados fijados al vector en ambos extremos. En primer lugar, se realiza una selección negativa para eliminar secuencias clonadas indeseadas resultantes de contaminantes o artefactos de PCR del modo siguiente. Se desechan las secuencias que se aparean con secuencias contaminantes tales como secuencias de ARN de vector, ARNt, ARNmt, ARNr, así como aquellas que codifican secuencias de ORF que exhiben una homología extensa con repeticiones como se define en la sección 4a). Se desechan las secuencias obtenidas mediante clonación directa usando cebadores anidados en marcadores 5' y 3' (sección 1, caso a) pero que carecen de cola de poli A. Se mantienen sólo los ORF que contienen un péptido señal y que terminan antes de la cola de poli A (caso a) o antes del final del 3' UTR clonado (caso b). Se eliminan entonces los ORF que contienen proteínas maduras improbables, tales como proteínas maduras cuyo tamaño es menor de 20 aminoácidos o menor de un 25% del tamaño de la proteína inmadura.

En la selección de los ORF, se dio prioridad al ORF y al marco correspondientes a los polipéptidos descritos en SignalTag Patents (solicitudes de patente de Estados Unidos nº de serie: 08/905.223; 08/905.135; 08/905.051; 08/905.144; 08/905.279; 08/904.468; 08/905.134; y 08/905.133). Si el ORF no se encontraba entre los ORF descritos en SignalTag Patents, se eligió el ORF que codificaba el péptido señal con la puntuación mayor según el método de von Heijne como se define en el ejemplo 22. Si las puntuaciones eran idénticas, entonces se eligió el ORF más largo.

Se comparan entonces por pares las secuencias de clones de ADNc extendidos completos con BLAST después de enmascarar las secuencias repetidas. Se agrupan en la misma clase las secuencias que contienen al menos 90% de homología durante 30 nucleótidos. Se somete entonces cada agrupamiento a un análisis de agrupamientos que detecta las secuencias resultantes de cebamiento interno o de corte y empalme alternativo, secuencias idénticas o secuencias con varios desplazamientos de marco. Este análisis automático sirve como base para la selección manual de secuencias.

b) Selección manual de secuencias

Se lleva a cabo la selección manual usando informes generados automáticamente para cada clon de ADNc extendido completo secuenciado. Durante estos procedimientos manuales, se realiza una selección entre clones que pertenecen a la misma clase del modo siguiente. Se alinean y comparan las secuencias de ORF codificadas por clones que pertenecen a la misma clase. Si la homología entre secuencias nucleotídicas de clones que pertenecen a la misma clase es de más de 90% durante extensiones de 30 nucleótidos o si la homología entre secuencias aminoacídicas de clones pertenecientes a la misma clase es de más de 80% durante extensiones de 20 aminoácidos, entonces se considera que los clones son idénticos. El ORF elegido es el mejor según los criterios mencionados a continuación. Si las homologías de nucleótidos y aminoácidos son menores de 90% y 80% respectivamente, se dice que los clones codifican distintas proteínas, pudiendo seleccionarse ambas si contienen secuencias de interés.

Se efectúa la selección de clones de ADNc extendido completo que codifican secuencias de interés usando los siguientes criterios. Se comprueban en primer lugar los parámetros estructurales (marcador inicial, sitio y señal de poliadenilación). Se examinan entonces las homologías con ácidos nucleicos y proteínas conocidos para determinar si la secuencia de clon se aparea con una secuencia de ácido nucleico/proteína conocida y, en este último caso, su índice de cobertura y la fecha en que se hizo pública la secuencia. Si no hay un apareamiento extenso con secuencias distintas de EST o ADN genómico, o si la secuencia del clon porta información sustancialmente nueva, tal como codificar una proteína resultante de corte y empalme alternativo de un ARNm que codifica una proteína ya conocida, se mantiene la secuencia. Se describen en el ejemplo 28 ejemplos de dichos ADNc extendidos completos clonados que contienen secuencias de interés. Se desechan durante este procedimiento las secuencias que resultan de insertos quiméricos o dobles como se evalúa mediante homología con otras secuencias.

Ejemplo 28

Clonación y secuenciación de ADNc extendidos

5 Se usó el procedimiento descrito en el ejemplo 27 anterior para obtener los ADNc extendidos descritos en la presente memoria. Usando este enfoque, se obtuvo el ADNc completo de SEC ID N° 17. Este ADNc entra en la categoría de “EST-ext” descrita anteriormente y codifica el péptido señal MKKVLLLITAILAVAVG (SEC ID N° 18) que tiene una puntuación de von Heijne de 8,2.

10 Se obtuvo también el ADNc completo de SEC ID N° 19 usando este procedimiento. Este ADNc entra en la categoría de “EST-ext” descrita anteriormente y codifica el péptido señal MWWFQQGLSFLPSALVIWTS (SEC ID N° 20) que tiene una puntuación de von Heijne de 5,5.

15 Otro ADNc completo obtenido usando el procedimiento descrito anteriormente tiene la secuencia de SEC ID N° 21. Este ADNc entra dentro de la categoría de “EST-ext” descrita anteriormente y codifica el péptido señal MVLT-TLPSANSANSPVNMPTTGPNSLSYASSALSPCLT (SEC ID N° 22) que tiene una puntuación de von Heijne de 5,9.

20 Se usó también el procedimiento anterior para obtener un ADNc completo que tiene la secuencia de SEC ID N° 23. Este ADNc entra en la categoría de “EST-ext” descrita anteriormente y codifica el péptido señal ILSTVTALTFFAXA (SEC ID N° 24) que tiene una puntuación de von Heijne de 5,5.

25 Se obtuvo también el ADNc completo de SEC ID N° 25 usando este procedimiento. Este ADNc entra en la categoría de “nuevo” descrita anteriormente y codifica un péptido señal LVLTLCTLPLAVA (SEC ID N° 26) que tiene una puntuación de von Heijne de 10,1.

Se obtuvo también el ADNc completo de SEC ID N° 27 usando este procedimiento. Este ADNc entra en la categoría de “nuevo” descrita anteriormente y codifica un péptido señal LWLLFFLVTAIHA (SEC ID N° 28) que tiene una puntuación de von Heijne de 10,7.

30 Se usaron también los procedimientos anteriores para obtener los ADNc extendidos de la presente invención. Se obtuvieron como se describe anteriormente 5’ EST expresados en una variedad de tejidos.

Se usaron los 5’ EST obtenidos como se describe anteriormente para obtener ADNc extendidos que tenían las secuencias de SEC ID N° 40-140 y 242-377. La Tabla IV proporciona los números de identificación de secuencia de los ADNc extendidos de SEC ID N° 45, la localización de la secuencia de codificación completa en la SEC ID N° 45 (concretamente, los nucleótidos que codifican tanto el péptido señal como la proteína madura, enumerados bajo el encabezado localización de FCS en la Tabla IV), la localización de los nucleótidos en la SEC ID N° 45 que codifica los péptidos señal (enumerados bajo el encabezado localización de SigPep en la Tabla IV), la localización de los nucleótidos en la SEC ID N° 45 que codifican las proteínas maduras generadas por escisión del péptido señal (enumerados bajo el encabezado localización de polipéptido maduro en la Tabla IV), la localización en la SEC ID N° 45 de codones de terminación (enumerados bajo el encabezado localización de codón de terminación en la Tabla IV), la localización en la SEC ID N° 45 de señales de poli A (enumeradas bajo el encabezado localización de señal de poli A en la Tabla IV) y la localización de sitios de poli A (enumerados bajo el encabezado localización de sitios de poli A en la Tabla IV).

45 Se examinó en los polipéptidos codificados por los ADNc extendidos la presencia de motivos estructurales o funcionales conocidos o la presencia de firmas, secuencias aminoácídicas pequeñas que están bien conservadas entre los miembros de una familia de proteínas. Las regiones conservadas se han usado para derivar patrones o matrices de consenso incluidos en el banco de datos PROSITE, en particular en el archivo prosite.dat (versión 13.0 de noviembre de 1995, localizado en <http://expasy.hcuge.ch/sprot/prosite.html>). Se usaron los programas Prosit_convert y prosite_scan (http://ulrec3.unil.ch/ftpserveur/prosite_scan) para encontrar firmas en los ADNc extendidos.

50 Para cada patrón obtenido con el programa prosite_convert a partir del archivo prosite.dat, se ha ensayado la exactitud de la detección en una secuencia de una proteína nueva evaluando la frecuencia de aciertos irrelevantes en la población de proteínas secretadas humanas incluida en el banco de datos SWISSPROT. Se usó como índice la relación entre el número de aciertos en proteínas barajadas (con un tamaño de ventana de 20 aminoácidos) y el número de aciertos en proteínas nativas (no barajadas). Se omitió cada patrón para el que la relación fuera mayor de 20% (un acierto en proteínas barajadas por 5 aciertos en proteínas nativas) durante la búsqueda con prosite_scan. El programa usado para barajar secuencias de proteína (db_shuffled) y el programa usado para determinar las estadísticas para cada patrón en los bancos de datos de proteína (prosite statistics) están disponibles en el sitio [ftp http://ulrec3.unil.ch/ftpserveur/prosite_scan](http://ulrec3.unil.ch/ftpserveur/prosite_scan).

55 La Tabla V enumera el número de identificación de secuencia de los polipéptidos de SEC ID N° 146, la localización de los residuos aminoácídicos de SEC ID N° 146 en el polipéptido completo (segunda columna), la localización de los residuos aminoácídicos de la SEC ID N° 146 en los péptidos señal (tercera columna) y la localización de los residuos aminoácídicos de la SEC ID N° 146 en el polipéptido maduro creado escindiendo el péptido señal del polipéptido completo (cuarta columna).

Se proporcionan las secuencias nucleotídicas de la secuencia de SEC ID N° 45 y la secuencia aminoacídica codificada por la SEC ID N° 45 (concretamente la secuencia aminoacídica de SEC ID N° 146) en el listado de secuencias adjunto. En algunos casos, las secuencias son preliminares y pueden incluir algunas secuencias o aminoácidos incorrectos o ambiguos. En las secuencias de SEC ID N° 45 puede examinarse fácilmente cualquier error en las mismas y puede resolverse cualquier ambigüedad de secuencia resecuenciando un fragmento que contiene dichos errores o ambigüedades en ambas cadenas. Los fragmentos de ácido nucleico para resolver errores de secuenciación o ambigüedades pueden obtenerse a partir de los clones depositados o pueden aislarse usando las técnicas descritas en la presente memoria. La resolución de cualquiera de dichas ambigüedades o errores puede facilitarse usando cebadores que hibridan con secuencias localizadas cerca de las secuencias ambiguas o erróneas. Por ejemplo, los cebadores pueden hibridar con secuencias a 50-75 bases de la ambigüedad o error. Tras la resolución de un error o ambigüedad, pueden hacerse las correspondientes correcciones en las secuencias proteicas codificadas por el ADN que contiene el error o ambigüedad. Por ejemplo, en las secuencias descritas en la presente memoria, se resolvieron las ambigüedades en la secuencia de SEC ID N° 131. Puede determinarse también la secuencia aminoacídica de la proteína codificada por un clon particular mediante la expresión del clon en una célula hospedadora adecuada, recogiendo la proteína y determinando su secuencia.

Para cada secuencia aminoacídica, los solicitantes han identificado lo que han determinado que es el marco de lectura mejor identificable con la información de secuencia disponible en el momento de presentación. Algunas de las secuencias aminoacídicas pueden contener indicadores "Xaa". Estos indicadores "Xaa" indican (1) un residuo que no puede identificarse debido a ambigüedad de secuencia nucleotídica o (2) un codón de terminación en la secuencia determinada donde los solicitantes creen que no debería existir (si la secuencia fuera determinada más exactamente).

Se mantienen células que contienen los ADNc extendidos (SEC ID N° 40-140 y 242-377) descritos en la presente memoria en el vector pED6dpc2 en depósito permanente por los inventores en Genset S.A. 24 Rue Royale, 75008 París, Francia.

Se depositaron conjuntos de células que contenían los ADNc extendidos (SEC ID N° 40-140 y 242-377), de los que son obtenibles células que contienen un polinucleótido particular, en la American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209 o la European Collection of Cell Cultures, Vaccine Research and Production Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Oown, Salisbury, Wiltshire SP4 OJG, Reino Unido. Se ha transfectado cada clon de ADNc extendido en células bacterianas separadas (*E. coli*) para este depósito combinado. La Tabla VI enumera el número de depósito del clon que contiene los ADNc extendidos de SEC ID N° 45. La Tabla VII proporciona el número de referencia interna asignado a las SEC ID N° 45 y 146 e indica si la secuencia es una secuencia de ácido nucleico o una secuencia de proteína.

Cada ADNc extendido puede retirarse del vector pED6dpc2 en el que se depositó efectuando una doble digestión con NotI-PstI para producir el fragmento apropiado para cada clon. Las proteínas codificadas por los ADNc extendidos pueden expresarse también a partir del promotor en pED6dpc2.

Pueden obtenerse células bacterianas que contienen un clon particular a partir del depósito combinado del modo siguiente:

Debe diseñarse una sonda o sondas oligonucleotídicas para la secuencia que es conocida para ese clon particular.

Esta secuencia puede derivar de las secuencias proporcionadas en la presente memoria, o de una combinación de esas secuencias. El diseño de la sonda oligonucleotídica debe seguir preferiblemente estos parámetros:

(a) Debe diseñarse en una zona de la secuencia que tenga el menor número de bases ambiguas ("N") o ninguna;

(b) Preferiblemente, la sonda se diseña para tener una T_m de aproximadamente 80°C (suponiendo 2 grados por cada A o T y 4 grados por cada G o C). Sin embargo, pueden usarse también sondas que tengan temperaturas de fusión entre 40°C y 80°C a condición de que no se pierda la especificidad.

El oligonucleótido debe marcarse preferiblemente con (-)[³²P]ATP (actividad específica 6000 Ci/mmol) y polinucleótido cinasa T4 usando técnicas empleadas habitualmente para marcar oligonucleótidos. Pueden usarse también otras técnicas de marcaje. El marcaje no incorporado debe retirarse preferiblemente mediante cromatografía de filtración en gel u otros métodos establecidos. La cantidad de radiactividad incorporada a la sonda debe cuantificarse mediante medida en un contador de centelleo. Preferiblemente, la actividad específica de la sonda resultante debe ser aproximadamente de 4X10⁶ dpm/pmol.

El cultivo bacteriano que contiene el conjunto de clones completos preferiblemente debe descongelarse y usarse 100 µl de la disolución madre para inocular un matraz de cultivo estéril que contiene 25 ml de caldo L estéril que contiene ampicilina a 100 µg/ml. Preferiblemente, el cultivo debe cultivarse hasta saturación a 37°C, y el cultivo saturado debe diluirse preferiblemente en caldo L reciente. Preferiblemente, deben sembrarse alícuotas de estas diluciones para determinar la dilución y el volumen que proporcionarán aproximadamente 5.000 colonias distintas y bien separadas sobre medios bacteriológicos sólidos que contienen caldo L que contiene ampicilina a 100 µg/ml y agar al 1,5% en una placa petri de 150 mm cuando se cultiva durante una noche a 37°C. Pueden emplearse también otros métodos conocidos de obtención de colonias distintas bien separadas.

ES 2 330 903 T3

Deben usarse entonces procedimientos de hibridación de colonia estándar para transferir las colonias a filtros de nitrocelulosa y lisarlos, desnaturalizarlos y hervirlos.

Se incuba entonces preferiblemente el filtro a 65°C durante 1 hora con agitación suave en 6X SSC (la solución madre 20X es 175,3 g de NaCl/l, 88,2 g de citrato de sodio/l, ajustada a pH 7,0 con NaOH) que contiene 0,5% de SDS, 100 pg/ml de ARN de levadura y EDTA 10 mM (aproximadamente 10 ml por filtro de 150 mm). Preferiblemente, se añade entonces la sonda a la mezcla de hibridación a una concentración mayor o igual a 1×10^5 dpm/ml. Se incuba entonces preferiblemente el filtro a 65°C con agitación suave durante una noche. Se lava entonces preferiblemente el filtro con 500 ml de 2X SSC/0,1% de SDS a temperatura ambiente con agitación suave durante 15 minutos. Es opcional un tercer lavado con 0,1X SSC/0,5% de SDS a 65°C durante 30 minutos a 1 hora. Se seca entonces preferiblemente el filtro y se somete a autorradiografía durante un tiempo suficiente para visualizar los positivos en película de rayos X. Pueden emplearse también otros métodos de hibridación conocidos.

Se recogen las colonias positivas, se cultivan y se aísla el ADN de plásmido usando procedimientos estándar. Pueden verificarse entonces los clones mediante análisis de restricción, análisis de hibridación o secuenciación de ADN.

El ADN de plásmido obtenido usando estos procedimientos puede manipularse entonces usando técnicas de clonación estándar familiares para los expertos en la técnica. Como alternativa, puede realizarse un PCR con cebadores diseñados en ambos extremos de la inserción de ADNc extendido. Por ejemplo, puede realizarse una reacción de PCR usando un cebador que tiene la secuencia GGCCATACACTTBAGTGAC (SEC ID N° 38) y un cebador que tiene la secuencia ATATAGACAAACGCACACC (SEC ID N° 39). El producto de PCR que corresponde al ADNc extendido puede manipularse entonces usando técnicas de clonación estándar familiares para los expertos en la técnica.

Además de métodos basados en PCR para obtener ADNc extendidos, pueden emplearse también métodos tradicionales basados en hibridación. Estos métodos pueden usarse también para obtener los ADN genómicos que codifican los ARNm de los que derivaban los 5' EST, los ARNm correspondientes a los ADNc extendidos o ácidos nucleicos que son homólogos de los ADNc extendidos o 5' EST. El ejemplo 29 siguiente proporciona un ejemplo de dichos métodos.

Ejemplo 29

Métodos para obtener ADNc extendidos o ácidos nucleicos homólogos de ADNc extendidos o 5' EST

Puede prepararse una colección de ADNc completos usando las estrategias descritas en los ejemplos 13, 14, 15 y 16 anteriores reemplazando el nonúmero aleatorio usado en el ejemplo 14 por un cebador oligo-dT. Por ejemplo, puede usarse el oligonucleótido de SEC ID N° 14.

Como alternativa, puede obtenerse una colección de ADNc o colección de ADN genómico a partir de una fuente comercial o prepararse usando técnicas familiares para los expertos en la técnica. La colección incluye ADNc que derivan de los ARNm correspondientes a un 5' EST o que tienen homología con un ADNc extendido o 5' EST. Se hibrida la colección de ADNc o colección de ADN genómico con una sonda detectable que comprende al menos 10 nucleótidos consecutivos de 5' EST o ADNc extendido usando técnicas convencionales. Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos de 5' EST o ADNc extendido. Más preferiblemente, la sonda comprende al menos 20-30 nucleótidos consecutivos de 5' EST o ADNc extendido. En algunas realizaciones, la sonda comprende al menos 30 nucleótidos de 5' EST o ADNc extendido. En otras realizaciones, la sonda comprende al menos 40, al menos 50, al menos 75, al menos 100, al menos 150 o al menos 200 nucleótidos consecutivos de 5' EST o ADNc extendido.

Las técnicas para identificar clones de ADNc en una colección de ADNc que hibrida con una secuencia de sonda dada se dan a conocer en Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" 2ª Ed., Cold Spring Harbor Laboratory Press, 1989. Pueden usarse las mismas técnicas para aislar ADN genómicos.

Brevemente, se identifican clones de ADNc o ADN genómico que hibridan con la sonda detectable y se aíslan para manipulación adicional del modo siguiente. Se marca una sonda que comprende al menos 10 nucleótidos consecutivos de 5' EST o ADNc extendido con un marcaje detectable tal como un radioisótopo o una molécula fluorescente. Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos de 5' EST o ADNc extendido. Preferiblemente, la sonda comprende 20-30 nucleótidos consecutivos de 5' EST o ADNc extendido. En algunas realizaciones, la sonda comprende más de 30 nucleótidos de 5' EST o ADNc extendido. En algunas realizaciones, la sonda comprende al menos 40, al menos 50, al menos 75, al menos 100, al menos 150 o al menos 200 nucleótidos consecutivos de 5' EST o ADNc extendido.

Las técnicas para marcar la sonda son bien conocidas e incluyen fosforilación con polinucleótido cinasa, traslado de mella, transcripción *in vitro* y técnicas no radiactivas. Se transfieren los ADNc o ADN genómicos en la colección a un filtro de nitrocelulosa o nailon y se desnaturalizan. Después de la incubación del filtro con una disolución de bloqueo, se pone en contacto con la sonda marcada y se incuba durante una cantidad de tiempo suficiente para que la sonda hibride con ADNc o ADN genómicos que contienen una secuencia capaz de hibridar con la sonda.

ES 2 330 903 T3

Al variar el rigor de las condiciones de hibridación usadas para identificar ADNc extendidos o ADN genómicos que hibridan con la sonda detectable, pueden identificarse y aislarse ADNc extendidos que tienen diferentes niveles de homología con la sonda. Para identificar ADNc extendidos o ADN genómicos que tienen un alto grado de homología con la secuencia de sonda, puede calcularse la temperatura de fusión de la sonda usando las siguientes fórmulas:

Para sondas de entre 14 y 70 nucleótidos de longitud, la temperatura de fusión (T_m) se calcula usando la fórmula: $T_m = 81,5 + 16,6(\log [Na+]) + 0,41 (\text{fracción de G+C}) - (600/N)$, en la que N es la longitud de la sonda.

Si se lleva a cabo la hibridación en una disolución que contiene formamida, la temperatura de fusión puede calcularse usando la ecuación $T_m = 81,5 + 16,6(\log [Na+]) + 0,41 (\text{fracción de G+C}) - (0,63\% \text{ formamida}) - (600/N)$, en la que N es la longitud de la sonda.

Puede llevarse a cabo la prehibridación en 6X SSC, 5X reactivo de Denhardt, 0,5% de SDS, 100 μg de ADN de esperma de salmón fragmentado desnaturalizado o 6X SSC, 5X reactivo de Denhardt, 0,5% de SDS, 100 μg de ADN de esperma de salmón fragmentado desnaturalizado, 50% de formamida. Se enumeran las fórmulas para SSC y disoluciones de Denhardt en Sambrook *et al.*, *supra*.

Se realiza la hibridación añadiendo la sonda detectable a las disoluciones de prehibridación enumeradas anteriormente. Cuando la sonda comprende ADN bicatenario, se desnaturaliza antes de la adición a la disolución de hibridación. Se pone en contacto el filtro con la disolución de hibridación durante un periodo de tiempo suficiente para permitir que la sonda hibride con ADNc extendidos o ADN genómicos que contienen secuencias complementarias de la misma u homólogas de la misma. Para sondas de más de 200 nucleótidos de longitud, la hibridación puede llevarse a cabo a 15-25°C por debajo de la T_m . Para sondas más cortas, tales como sondas oligonucleotídicas, la hibridación puede realizarse a 15-25°C por debajo de la T_m . Preferiblemente, para hibridaciones en 8 X SSC, se realiza la hibridación aproximadamente a 68°C. Preferiblemente, para hibridaciones en disoluciones que contienen 50% de formamida, se realiza la hibridación aproximadamente a 42°C.

Se consideraría que todas las hibridaciones anteriores están en condiciones "rigurosas". Después de la hibridación, se lava el filtro con 2X SSC, 0,1% de SDS a temperatura ambiente durante 15 minutos. Se lava entonces el filtro con 0,1X SSC, 0,5% de SDS a temperatura ambiente durante 30 minutos a 1 hora. Después de ello, se lava la disolución a la temperatura de hibridación en 0,1 X SSC, 0,5% de SDS. Se realiza un lavado final en 0,1 X SSC a temperatura ambiente.

Se identifican los ADNc extendidos, homólogos de ácidos nucleicos de ADNc extendidos o 5' EST o ADN genómicos que han hibridado con la sonda mediante autorradiografía u otras técnicas convencionales.

El procedimiento anterior puede modificarse para identificar ADNc extendidos, ácidos nucleicos homólogos de ADNc extendidos o ADN genómicos que tienen niveles reducidos de homología con la secuencia de sonda. Por ejemplo, para obtener ADNc extendidos, ácidos nucleicos homólogos de ADNc extendidos o ADN genómicos de homología reducida con la sonda detectable, pueden usarse condiciones menos rigurosas. Por ejemplo, la temperatura de hibridación puede reducirse en incrementos de 5°C desde 68°C a 42°C en un tampón de hibridación que tiene una concentración de Na^+ de aproximadamente 1 M. Después de la hibridación, puede lavarse el filtro con 2X SSC, 0,5% de SDS a la temperatura de hibridación. Estas condiciones se consideran condiciones "moderadas" por encima de 50°C y condiciones "bajas" por debajo de 50°C.

Como alternativa, la hibridación puede llevarse a cabo en tampones, tales como 6X SSC, que contienen formamida a una temperatura de 42°C. En este caso, la concentración de formamida en el tampón de hibridación puede reducirse en incrementos de 5% desde 50% a 0% para identificar los clones que tienen niveles reducidos de homología con la sonda. Después de la hibridación, puede lavarse el filtro con 6X SSC, 0,5% de SDS a 50°C. Estas condiciones se consideran condiciones "moderadas" por encima de 25% de formamida y condiciones "bajas" por debajo de 25% de formamida.

Se identifican mediante autorradiografía los ADNc extendidos, homólogos de ácido nucleico de ADNc extendidos o ADN genómicos que han hibridado con la sonda.

Si se desea obtener ácidos nucleicos homólogos de ADNc extendidos, tales como variantes alélicas de los mismos o ácidos nucleicos que codifican proteínas relacionadas con las proteínas codificadas por los ADNc extendidos, puede determinarse fácilmente el nivel de homología entre el ácido nucleico hibridado y el ADNc extendido o 5' EST usado como sonda. Para determinar el nivel de homología entre el ácido nucleico hibridado y el ADNc extendido o 5' EST del que derivaba la sonda, se comparan las secuencias nucleotídicas del ácido nucleico hibridado y los ADNc extendidos o 5' EST de los que derivaba la sonda. Por ejemplo, usando los métodos anteriores, pueden obtenerse e identificarse ácidos nucleicos que tienen al menos un 95% de homología de ácido nucleico con ADNc extendidos o 5' EST de los que derivaba la sonda. De forma similar, pero usando condiciones de hibridación progresivamente menos rigurosas, pueden obtenerse e identificarse ácidos nucleicos que tienen al menos un 90%, al menos un 85%, al menos un 80% o al menos un 75% de homología con los ADNc extendidos o 5' EST de los que derivaba la sonda. El nivel de homología entre el ácido nucleico hibridado y el ADNc extendido o 5' EST usado como sonda puede determinarse adicionalmente usando BLAST2N; los parámetros puede adaptarse dependiendo de la longitud de secuencia y el grado de homología

ES 2 330 903 T3

estudiados. En dichas comparaciones, pueden usarse los parámetros por defecto de los parámetros enumerados en las Tablas II y III.

Para determinar si un clon codifica una proteína que tiene una cantidad dada de homología con la proteína codificada por el ADNc extendido o 5' EST, se compara la secuencia aminoacídica codificada por el ADNc extendido o 5' EST con la secuencia aminoacídica codificada por el ácido nucleico hibridado. Se determina que existe homología cuando una secuencia aminoacídica en el ADNc extendido o 5' EST está estrechamente relacionada con una secuencia aminoacídica del ácido nucleico hibridado. Una secuencia está estrechamente relacionada cuando es idéntica a la del ADNc extendido o 5' EST o cuando contiene una o más sustituciones aminoacídicas en la misma en la que aminoácidos que tienen características similares se han sustituido entre sí. Usando los métodos anteriores, pueden obtenerse ácidos nucleicos que codifican proteínas que tienen al menos un 95%, al menos un 90%, al menos un 85%, al menos un 80% o al menos un 75% de homología con las proteínas codificadas por el ADNc extendido o 5' EST del que derivaba la sonda. Usando los métodos y algoritmos anteriores tales como FASTA, con parámetros dependientes de la longitud de secuencia y el grado de homología estudiados, puede determinarse el nivel de homología. En la determinación del nivel de homología usando FASTA, pueden usarse los parámetros por defecto o los parámetros enumerados en las Tablas II o III.

Como alternativa, pueden prepararse ADNc extendidos obteniendo ARNm del tejido, célula u organismo de interés usando procedimientos de preparación de ARNm que utilizan procedimientos de selección con poli A u otras técnicas conocidas por los expertos en la técnica. Se hibrida con el ARNm un primer cebador capaz de hibridar con la cola de poli A del ARNm y se efectúa una reacción de transcripción inversa para generar una primera cadena de ADNc.

Se hibrida la primera cadena de ADNc con un segundo cebador que contiene al menos 10 nucleótidos consecutivos de las secuencias de 5' EST para las que se desea un ADNc extendido. Preferiblemente, el cebador comprende al menos 12, 15 ó 17 nucleótidos consecutivos de las secuencias de 5' EST. Más preferiblemente, el cebador comprende 20-30 nucleótidos consecutivos de las secuencias de 5' EST. En algunas realizaciones, el cebador comprende más de 30 nucleótidos de las secuencias de 5' EST. Si se desea obtener ADNc que contienen la secuencia de codificación de proteína completa, incluyendo el sitio de inicio de la traducción auténtico, el segundo cebador usado contiene secuencias localizadas cadena arriba del sitio de inicio de la traducción. Se extiende el segundo cebador para generar una segunda cadena de ADNc complementaria de la primera cadena de ADNc. Como alternativa, puede efectuarse una RTPCR como se describe anteriormente usando cebadores de ambos extremos del ADNc a obtener.

Pueden prepararse ADNc extendidos que contienen fragmentos 5' del ARNm poniendo en contacto un ARNm que comprende la secuencia del 5' EST para el que se desea un ADNc extendido con un cebador que comprende al menos 10 nucleótidos consecutivos de las secuencias complementarias del 5' EST, hibridando el cebador con los ARNm y transcribiendo de forma inversa el cebador hibridado para preparar una primera cadena de ADNc a partir de los ARNm. Preferiblemente, el cebador comprende al menos 12, 15 ó 17 nucleótidos consecutivos del 5' EST. Más preferiblemente, el cebador comprende 20-30 nucleótidos consecutivos del 5' EST.

Después de ello, se sintetiza una segunda cadena de ADNc complementaria de la primera cadena de ADNc. La segunda cadena de ADNc puede prepararse hibridando un cebador complementario de secuencias en la primera cadena de ADNc con la primera la primera cadena de ADNc y extendiendo el cebador para generar la segunda cadena de ADNc.

Se aíslan los ADNc extendidos bicatenarios usando los métodos descritos anteriormente y se clonan. Los ADNc extendidos pueden clonarse en vectores tales como plásmidos o vectores víricos capaces de replicar en una célula hospedadora apropiada. Por ejemplo, la célula hospedadora puede ser una célula de bacteria, mamífero, ave o insecto.

Las técnicas para aislar ARNm, transcribir de forma inversa un cebador hibridado con ARNm para generar una primera cadena de ADNc, extender un cebador para preparar una segunda cadena de ADNc complementaria de la primera cadena de ADNc, aislar el ADNc bicatenario y clonar el ADNc bicatenario son bien conocidas en la técnica y se describen en "Current Protocols in Molecular Biology", John Wiley and Sons, Inc. 1997 y Sambrook *et al.* "Molecular Cloning: A Laboratory Manual", 2ª edición, Cold Spring Harbor Laboratory Press, 1989.

Como alternativa, pueden usarse kits para obtener ADNc completos, tales como GeneTrapper (nº de cat. 10356-020, Gibco, BRL), para obtener ADNc completos o ADNc extendidos. En este enfoque, se preparan ADNc completos o extendidos a partir de ARNm y se clonan en fagémidos bicatenarios. Se vuelve entonces monocatenaria la colección de ADNc en los fagémidos bicatenarios mediante tratamiento con una endonucleasa tal como el producto del gen II del fago F1, y exonucleasa III como se describe en el manual que acompaña al kit GeneTrapper. Se hibrida un oligonucleótido biotinilado que comprende la secuencia de un 5' EST, o un fragmento que contiene al menos 10 nucleótidos del mismo, con los fagémidos monocatenarios. Preferiblemente, el fragmento comprende al menos 12, 15 ó 17 nucleótidos consecutivos del 5' EST. Más preferiblemente, el fragmento comprende 20-30 nucleótidos consecutivos del 5' EST. En algunos procedimientos, el fragmento puede comprender más de 30 nucleótidos consecutivos de 5' EST. Por ejemplo, el fragmento puede comprender al menos 40, al menos 50, al menos 75, al menos 100, al menos 150 o al menos 200 nucleótidos consecutivos de 5' EST.

Se aíslan híbridos entre el oligonucleótido biotinilado y fagémidos que tienen insertos que contienen la secuencia 5' EST incubando los híbridos con perlas paramagnéticas recubiertas con estreptavidina y recuperando las perlas con

un imán. Después de ello, se liberan de las perlas los fagémidos resultantes que contienen la secuencia de 5' EST y se convierten en ADN bicatenario usando un cebador específico de la secuencia de 5' EST. Se transforma el ADN bicatenario resultante en bacterias. Se identifican ADNc extendidos que contienen la secuencia de 5' EST mediante PCR de colonia o hibridación de colonia.

Puede proporcionarse una pluralidad de ADNc extendidos que contienen secuencias de codificación de proteína completas o secuencias que codifican sólo la proteína madura restante después de que se ha escindido el péptido señal en forma de colecciones de ADNc para la evaluación posterior de las proteínas codificadas o el uso en ensayos de diagnóstico como se describe a continuación.

IV. Expresión de proteínas codificadas por ADNc extendidos usando 5' EST

Pueden usarse ADNc extendidos que contienen las secuencias de proteína completas de sus correspondientes ARNm o porciones de los mismos, tal como ADNc que codifican la proteína madura, para expresar las proteínas secretadas o porciones de las mismas que se codifican como se describe en el ejemplo 30 siguiente. Si se desea, los ADNc extendidos pueden contener las secuencias que codifican el péptido señal para facilitar la secreción de la proteína expresada. Se apreciará que puede clonarse simultáneamente en vectores de expresión una pluralidad de ADNc extendidos que contienen las secuencias de codificación de proteína completa o porciones de las mismas para crear una colección de expresión para el análisis de las proteínas codificadas como se describe a continuación.

Ejemplo 30

Expresión de las proteínas codificadas por ADNc extendidos o porciones de los mismos

Para expresar las proteínas codificadas por los ADNc extendidos o porciones de los mismos, se obtienen ácidos nucleicos que contienen la secuencia de codificación de las proteínas o porciones de las mismas a expresar como se describe en los ejemplos 27-29, y se clonan en un vector de expresión adecuado. Si se desea, los ácidos nucleicos pueden contener las secuencias que codifican el péptido señal para facilitar la secreción de la proteína expresada. Por ejemplo, el ácido nucleico puede comprender la secuencia de SEC ID N° 45 enumerada en la Tabla IV y en el listado de secuencias adjunto. Como alternativa, el ácido nucleico puede comprender aquellos nucleótidos que constituyen la secuencia de codificación completa de la secuencia de SEC ID N° 45 como se define en la Tabla IV anterior.

Se apreciará que en caso de que la extensión de la secuencia de codificación completa (concretamente, la secuencia que codifica el péptido señal y la proteína madura resultante de la escisión del péptido señal) difiera de la enumerada en la Tabla IV como resultado de un error de secuenciación, error de transcripción inversa o amplificación, corte y empalme de ARNm, modificación postraduccional de la proteína codificada, escisión enzimática de la proteína codificada u otros factores biológicos, un experto en la técnica sería fácilmente capaz de identificar la extensión de la secuencia de codificación completa en la secuencia de SEC ID N° 45. Por ejemplo, la secuencia de SEC ID N° 115 representa un transcrito cortado y empalmado alternativamente de un ARNm identificado anteriormente. En consecuencia, el alcance de cualquier reivindicación de la presente memoria respecto a los ácidos nucleicos que contienen la secuencia de codificación completa de SEC ID N° 45 no ha de considerarse como excluyente de cualquier variación o equivalente fácilmente identificable de las secuencias de codificación completas enumeradas en la Tabla IV. De forma similar, en caso de que la extensión de los polipéptidos completos difiera de los indicados en la Tabla V como resultado de cualquiera de los factores precedentes, el alcance de las reivindicaciones respecto a los polipéptidos que comprenden la secuencia aminoacídica del polipéptido completo no ha de considerarse como excluyente de cualquier variación o equivalente fácilmente identificable de las secuencias enumeradas en la Tabla V.

Como alternativa, el ácido nucleico usado para expresar la proteína o porción de la misma puede comprender aquellos nucleótidos que codifican la proteína madura (concretamente, la proteína creada escindiendo el péptido señal) codificada por la SEC ID N° 45 como se define en la Tabla IV anterior.

Se apreciará que en caso de que la extensión de la secuencia que codifica la proteína madura difiera de la enumerada en la Tabla IV como resultado de un error de secuenciación, error de transcripción inversa o error de amplificación, corte y empalme de ARNm, modificación postraduccional de la proteína codificada, escisión enzimática de la proteína codificada u otros factores biológicos, un experto en la técnica sería fácilmente capaz de identificar la extensión de la secuencia que codifica la proteína madura en la secuencia de SEC ID N° 45. En consecuencia, el alcance de cualquier reivindicación de la presente memoria respecto a ácidos nucleicos que contienen la proteína madura codificada por la SEC ID N° 45 no ha de considerarse como excluyente de cualquier variación o equivalente fácilmente identificable de las secuencias enumeradas en la Tabla IV. Por tanto, las reivindicaciones respecto a ácidos nucleicos que contienen la secuencia que codifica la proteína madura comprenden equivalentes de las secuencias enumeradas en la Tabla IV, tales como secuencias que codifican proteínas biológicamente activas resultantes de modificación postraduccional, escisión enzimática u otras variaciones o equivalentes fácilmente identificables de la proteína secretada además de la escisión del péptido señal. De forma similar, en caso de que la extensión del polipéptido maduro difiera de la indicada en la Tabla V como resultado de cualquiera de los factores precedentes, el alcance de las reivindicaciones respecto a polipéptidos que comprenden la secuencia de una proteína madura incluida en la secuencia de SEC ID N° 146 no ha de considerarse como excluyente de cualquier variación o equivalente fácilmente identificable de las secuencias enumeradas en la Tabla V. Por tanto, las reivindicaciones respecto a polipéptidos que comprenden la secuencia de la proteína madura comprenden equivalentes de las secuencias enumeradas en la Tabla IV, tales como proteínas biológi-

camente activas resultantes de modificación postraducciona, escisión enzimática o cualquier variación o equivalente fácilmente identificable de las proteínas secretadas además de la escisión del péptido señal. Se apreciará también que en caso de que la forma biológicamente activa del polipéptido incluida en la secuencia de SEC ID N° 146 o los ácidos nucleicos que codifican la forma biológicamente activa del polipéptido difiera de los identificados como polipéptido maduro en la Tabla V o los nucleótidos que codifican el polipéptido maduro en la Tabla IV como resultado de un error de secuenciación, error de transcripción inversa o error de amplificación, corte y empalme de ARNm, modificación postraducciona de la proteína codificada, escisión enzimática de la proteína codificada u otros factores biológicos, el experto en la técnica sería fácilmente capaz de identificar los aminoácidos en la forma biológicamente activa de los polipéptidos y ácidos nucleicos que codifican la forma biológicamente activa de los polipéptidos. En dichos casos, las reivindicaciones respecto a polipéptidos que comprenden la proteína madura incluida en la SEC ID N° 146 o ácidos nucleicos que comprenden los nucleótidos de SEC ID N° 45 que codifican la proteína madura, no debe considerarse que excluyan cualquier variación fácilmente identificable de las secuencias enumeradas en la Tabla IV y la Tabla V.

En algunas realizaciones, el ácido nucleico usado para expresar la proteína o porción de la misma puede comprender aquellos nucleótidos que codifican el péptido señal codificado por la SEC ID N° 45 como se define en la Tabla IV anterior.

Se apreciará que en caso de que la extensión de la secuencia que codifica el péptido señal difiera de la enumerada en la Tabla IV como resultado de un error de secuenciación, error de transcripción inversa o error de amplificación, corte y empalme de ARNm, modificación postraducciona de la proteína codificada, escisión enzimática de la proteína codificada u otros factores biológicos, un experto en la técnica sería fácilmente capaz de identificar la extensión de la secuencia que codifica el péptido señal en la secuencia de SEC ID N° 45. En consecuencia, el alcance de cualquier reivindicación de la presente memoria respecto a ácidos nucleicos que contienen el péptido señal codificado por la SEC ID N° 45 no ha de considerarse como excluyente de cualquier variación fácilmente identificable de las secuencias enumeradas en la Tabla IV. De forma similar, en caso de que la extensión del péptido señal difiera de los indicados en la Tabla V como resultado de cualquiera de los factores precedentes, el alcance de las reivindicaciones respecto a polipéptidos que comprenden la secuencia de un péptido señal incluida en la secuencia de SEC ID N° 146 no ha de considerarse como excluyente de cualquier variación fácilmente identificable de las secuencias enumeradas en la Tabla V.

Como alternativa, el ácido nucleico puede codificar un polipéptido que comprende al menos 10 aminoácidos consecutivos de la secuencia de SEC ID N° 146. El ácido nucleico puede codificar un polipéptido que comprende al menos 15 aminoácidos consecutivos de la secuencia de SEC ID N° 146. El ácido nucleico puede codificar un polipéptido que comprende al menos 25 aminoácidos consecutivos de la secuencia de SEC ID N° 146. El ácido nucleico puede codificar un polipéptido que comprende al menos 60, al menos 75, al menos 100 o más de 100 aminoácidos consecutivos de la secuencia de SEC ID N° 146.

Los ácidos nucleicos insertados en los vectores de expresión pueden contener también secuencias cadena arriba de las secuencias que codifican el péptido señal, tales como secuencias que regulan los niveles de expresión o secuencias que confieren una expresión específica de tejido.

El ácido nucleico que codifica la proteína o polipéptido a expresar está operativamente ligado a un promotor en un vector de expresión usando la tecnología convencional de clonación. El vector de expresión puede ser cualquiera de los sistemas de expresión de mamífero, levadura, insecto o bacteria conocidos en la técnica. Los vectores y sistemas de expresión comercialmente disponibles se pueden conseguir de una variedad de proveedores incluyendo Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), e Invitrogen (San Diego, California). Si se desea, para potenciar la expresión y facilitar el plegamiento apropiado de las proteínas, el contexto de los codones y el emparejamiento de los codones de la secuencia puede optimizarse para el organismo particular de expresión en el cual se introduce el vector de expresión, como se explica por Hatfield, *et al.*, patente de EE.UU. n° 5.082.767

A continuación, se proporciona un método ejemplar para expresar las proteínas codificadas por los ADNc extendidos correspondientes a los 5' EST o los ácidos nucleicos descritos anteriormente. En primer lugar, se identifican el codón de iniciación de metionina del gen y la señal de poli A del gen. Si el ácido nucleico que codifica el polipéptido a expresar carece de una metionina que sirva como sitio de iniciación, se puede introducir una metionina de iniciación próxima al primer codón del ácido nucleico usando técnicas convencionales. De forma similar, si los ADNc extendidos carecen de una señal de poli A, puede añadirse esta secuencia al constructo, por ejemplo, mediante corte y empalme de la señal de poli A de pSG5 (Stratagene) usando enzimas endonucleasa de restricción BglII y SailI e incorporándola al vector de expresión de mamífero pXT1 (Stratagene). pXT1 contiene los LTR y una porción del gen *gag* de virus de leucemia de murino Moloney. La posición de los LTR en el constructo permite una transfección estable eficaz. El vector incluye el promotor timidina-cinasa de herpes simplex y el gen de neomicina seleccionable. Se obtiene el ADNc extendido o porción del mismo que codifica el polipéptido a expresar mediante PCR del vector bacteriano usando cebadores oligonucleotídicos complementarios del ADNc extendido o porción del mismo, y que contiene las secuencias de endonucleasa de restricción de PstI incorporadas en el cebador 5' y BglIII en el extremo 5' del correspondiente cebador 3' de ADNc, teniendo cuidado de asegurar que el ADNc extendido se dispone en fase con la señal de poli A. El fragmento purificado obtenido a partir de la reacción de PCR resultante se digiere con PstI, se termina de forma roma con una exonucleasa, se digiere con BglIII, se purifica y se liga a pXT1, que contiene ahora una señal de poli A y se digiere con BglIII.

ES 2 330 903 T3

Se transfecta el producto ligado en células NIH 3T3 de ratón usando lipofectina (Life Technologies, Inc., Grand Island, Nueva York) en las condiciones exhibidas en la especificación de producto. Se seleccionan transfectantes positivos después de cultivar las células transfectadas en G418 600 $\mu\text{g/ml}$ (Sigma, St. Louis, Misuri). Preferiblemente, se libera la proteína expresada en el medio de cultivo, facilitando así la purificación.

Como alternativa, pueden clonarse los ADNc extendidos en pED6dpc2 como se describe anteriormente. Los constructos de pED6dpc2 resultantes pueden transfectarse en una célula hospedadora adecuada, tal como células COS1. Se seleccionan células resistentes a metotrexato y se expanden. Preferiblemente, la proteína expresada a partir del ADNc extendido se libera en el medio de cultivo, facilitando así la purificación.

Se separan las proteínas en el medio de cultivo mediante electroforesis en gel. Si se desea, las proteínas pueden precipitarse con sulfato de amonio o separarse basándose en el tamaño o la carga antes de la electroforesis.

Como control, se introduce el vector de expresión que carece de un inserto de ADNc en células u organismos hospedadores y se recolectan las proteínas en el medio. Se detectan las proteínas secretadas en el medio usando técnicas tales como tinción con Coomassie o plata o usando anticuerpos contra la proteína codificada por el ADNc extendido. Las técnicas de tinción con Coomassie y plata son familiares para los expertos en la técnica.

Pueden generarse anticuerpos capaces de reconocimiento específico de la proteína de interés usando péptidos de 15 unidades sintéticos que tienen una secuencia codificada por el 5' EST, ADNc extendido o porción del mismo apropiado. Se inyectan los péptidos sintéticos en ratones para generar anticuerpo frente al polipéptido codificado por el 5' EST, ADNc extendido o porción del mismo.

Se comparan las proteínas secretadas de las células u organismos hospedadores que contienen un vector de expresión que contiene el ADNc extendido derivado de un 5' EST o porción del mismo con las de células u organismos de control. La presencia de una banda en el medio de células que contienen el vector de expresión, que está ausente en el medio de células de control, indica que el ADNc extendido codifica una proteína secretada. Generalmente, la banda correspondiente a la proteína codificada por el ADNc extendido tendrá una movilidad cercana a la esperada basándose en el número de aminoácidos en el marco abierto de lectura del ADNc extendido. Sin embargo, la banda puede tener una movilidad diferente de la esperada como resultado de modificaciones tales como glucosilación, ubiquinación o escisión enzimática.

Como alternativa, si la proteína expresada a partir de los vectores de expresión anteriores no contiene secuencias que dirijan su secreción, las proteínas expresadas a partir de células hospedadoras que contienen un vector de expresión que contiene un inserto que codifica una proteína secretada o porción de la misma pueden compararse con proteínas expresadas en células hospedadoras que contienen el vector de expresión sin un inserto. La presencia de una banda en muestras de células que contienen el vector de expresión con un inserto que está ausente en muestras de células que contienen el vector de expresión sin inserto indica que se está expresando la proteína o porción de la misma deseada. Generalmente, la banda tendrá la movilidad esperada para la proteína secretada o porción de la misma. Sin embargo, la banda puede tener una movilidad diferente de la esperada como resultado de modificaciones tales como glucosilación, ubiquinación o escisión enzimática.

La proteína codificada por el ADNc extendido puede purificarse usando técnicas inmunocromatográficas estándar. En dichos procedimientos, se aplica una disolución que contiene la proteína secretada, tal como el medio de cultivo o un extracto celular, a una columna que tiene anticuerpos contra la proteína secretada unidos a la matriz de cromatografía. Se permite unir la proteína secretada a la columna de inmunocromatografía. Después de ello, se lava la columna para retirar las proteínas no unidas específicamente. Se libera entonces la proteína secretada unida específicamente de la columna y se recupera usando técnicas estándar.

Si no es posible la producción de anticuerpo, puede incorporarse la secuencia de ADNc extendido o porción del mismo a vectores de expresión diseñados para uso en esquemas de purificación que emplean polipéptidos quiméricos. En dichas estrategias, se inserta la secuencia de codificación del ADNc extendido o porción del mismo en fase con el gen que codifica la otra mitad de la quimera. La otra mitad de la quimera puede ser β -globina o una secuencia que codifica un polipéptido de unión a níquel. Se usa entonces una matriz de cromatografía que tiene un anticuerpo contra β -globina o níquel unido a la misma para purificar la proteína quimérica. Los sitios de escisión de proteasa pueden introducirse por ingeniería genética entre el gen de β -globina o el polipéptido de unión a níquel y el ADNc extendido o porción del mismo. Por tanto, los dos polipéptidos de la quimera pueden separarse entre sí mediante digestión por proteasa.

Es un vector de expresión útil para generar β -globina quimérica el pSG5 (Stratagene), que codifica β -globina de conejo. El intrón II del gen de β -globina de conejo facilita el corte y empalme del transcrito expresado, y la señal de poliadenilación incorporada al constructo aumenta el nivel de expresión. Estas técnicas como se describen son bien conocidas por los expertos en la técnica de la biología molecular. Los métodos estándar están publicados en textos de métodos tales como Davis *et al.*, ("Basic Methods in Molecular Biology", L.G. Davis, M.D. Dibner, y J.F. Battey, ed., Elsevier Press, NY, 1986) y muchos de los métodos están disponibles en Stratagene, Life Technologies, Inc. o Promega. Los polipéptidos pueden producirse adicionalmente a partir del constructo usando sistemas de traducción *in vitro* tales como el kit de traducción *in vitro* ExpressTM (Stratagene).

Después de la expresión y purificación de las proteínas secretadas codificadas por los 5' EST, ADNc extendidos o fragmentos de los mismos, puede ensayarse en las proteínas purificadas la capacidad de unirse a la superficie de diversos tipos celulares como se describe en el ejemplo 31 siguiente. Se apreciará que puede incluirse una pluralidad de proteínas expresadas a partir de estos ADNc en un panel de proteínas para evaluar simultáneamente las actividades específicamente descritas a continuación, así como otros papeles biológicos para los que están disponibles ensayos para determinar la actividad.

Ejemplo 31

Análisis de proteínas secretadas para determinar si se unen a la superficie celular

Las proteínas codificadas por los 5' EST, ADNc extendidos o fragmentos de los mismos se clonan en vectores de expresión tales como los descritos en el ejemplo 30. Se purifican las proteínas por tamaño, carga, inmunocromatografía u otras técnicas familiares para los expertos en la técnica. Después de la purificación, se marcan las proteínas usando técnicas conocidas por los expertos en la técnica. Se incuban las proteínas marcadas con células o estirpes celulares derivadas de una variedad de órganos o tejidos para permitir que las proteínas se unan a cualquier receptor presente en la superficie celular. Después de la incubación, se lavan las células para retirar la proteína no unida específicamente. Se detectan las proteínas marcadas mediante autorradiografía. Como alternativa, pueden incubarse proteínas no marcadas con las células y detectarse con anticuerpos que tienen un marcaje detectable, tal como una molécula fluorescente, unido a los mismos.

Puede analizarse la especificidad de la unión a superficie celular realizando un análisis de competición en el que se incuban diversas cantidades de proteína no marcada junto con la proteína marcada. La cantidad de proteína marcada unida a la superficie celular se reduce a medida que aumenta la cantidad de proteína no marcada competitiva. Como control, se incluyen en algunas reacciones de unión diversas cantidades de proteína no marcada no relacionada con la proteína marcada. La cantidad de proteína marcada unida a la superficie celular no se reduce en reacciones de unión que contienen cantidades crecientes de proteína no marcada no relacionada, indicando que la proteína codificada por los ADNc se une específicamente a la superficie celular.

Como se discutió anteriormente, se ha mostrado que las proteínas secretadas tienen una serie de importantes efectos fisiológicos y, en consecuencia, representan un recurso terapéutico valioso. Las proteínas secretadas codificadas por los ADNc extendidos o porciones de los mismos según los ejemplos 27-29 pueden evaluarse para determinar sus actividades fisiológicas como se describe a continuación.

Ejemplo 32

Ensayo en proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de la actividad citocina, de proliferación celular o diferenciación celular

Como se discute anteriormente, las proteínas secretadas pueden actuar como citocinas o pueden afectar a la proliferación o diferenciación celular. Muchos factores proteicos descubiertos hasta la fecha, incluyendo todas las citocinas conocidas, han exhibido actividad en uno o más ensayos de proliferación celular dependiente de factor, y por tanto los ensayos sirven como una confirmación conveniente de la actividad de citocina. Se evidencia la actividad de una proteína descrita en la presente memoria mediante uno cualquiera de una serie de ensayos de proliferación celular dependientes de factor para estirpes celulares incluyendo, sin limitación, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c y CMK. En las proteínas codificadas por los ADNc extendidos anteriores o porciones de los mismos, puede evaluarse su capacidad de regular la proliferación de linfocitos T o timocitos en ensayos tales como los descritos anteriormente o en las siguientes referencias: "Current Protocols in Immunology", Ed. por J.E. Coligan *et al.*, Greene Publishing Associates y Wiley-Interscience; Takai *et al. J. Immunol.*, 137: 3494-3500, 1986. Bertagnolli *et al. J. Immunol.*, 145: 1706-1712, 1990. Bertagnolli *et al. Cellular Immunology* 133: 327-341, 1991. Bertagnolli *et al. J. Immunol.*, 149: 3778-3783, 1992. Bowman *et al., J. Immunol.*, 152: 1756-1761, 1994.

Además, son conocidos numerosos ensayos de producción y/o proliferación de citocina de células de bazo, células de nódulo linfático y timocitos. Estos incluyen las técnicas dadas a conocer en "Current Protocols in Immunology", J.E. Coligan *et al.* Ed., Vol 1 pág. 3.12.1-3.12.14, John Wiley and Sons, Toronto, 1994; y Schreiber, R.D. "Current Protocols in Immunology", *supra* Vol 1 pág. 6.8.1-6.8.8 John Wiley and Sons, Toronto, 1994.

En las proteínas codificadas por los ADNc, puede ensayarse también la capacidad de regular la proliferación y diferenciación de células hematopoyéticas o linfopoyéticas. Son familiares para los expertos en la técnica muchos ensayos para dicha actividad, incluyendo los ensayos de las siguientes referencias: Bottomly, K., Davis, L.S. y Lipsky, P.E., "Measurement of Human and Murine Interleukin 2 and Interleukin 4", "Current Protocols in immunology", J.E. Coligan *et al.* Ed., Vol 1 pág. 6.3.1-6.3.12, John Wiley and Sons, Toronto, 1991; deVries *et al., J. Exp. Med.* 173: 1205-1211, 1991; Moreau *et al., Nature* 36: 690-692, 1988; Greenberger *et al., Proc. Natl. Acad. Sci. U.S.A.* 80: 2931-2938, 1983; Nordan, R., "Measurement of Mouse and Human Interleukin 6", "Current Protocols in Immunology", J.E. Coligan *et al.* Ed., Vol 1 pág. 6.6.1-6.6.5, John Wiley and Sons, Toronto, 1991; Smith *et al., Proc. Natl. Acad. Sci.*

U.S.A. 83: 1857-1861, 1986; Bennett, F., Giannotti, J., Clark, S.C. y Turner, K.J., "Measurement of Human Interleukin 11", "Current Protocols in Immunology", J.E. Coligan *et al.* Ed., Vol 1 pág. 6.15.1, John Wiley and Sons, Toronto, 1991; Ciarletta, A., Giannotti, J., Clark, S.C. y Turner, K.J., "Measurement of Mouse and Human Interleukin 9", "Current Protocols in Immunology", J.E. Coligan *et al.*, Ed. Vol 1 pág. 6.13.1. John Wiley and Sons, Toronto, 1991.

5

En las proteínas codificadas por los ADNc, puede ensayarse también su capacidad de regular las respuestas de linfocitos T a antígenos. Son familiares para los expertos en la técnica muchos ensayos para dicha actividad, incluyendo los ensayos descritos en las siguientes referencias: capítulo 3 (*In Vitro Assays for Mouse Lymphocyte Function*), capítulo 6 (Cytokines and Their Cellular Receptors) y capítulo 7, (Immunologic Studies in Humans) en "Current Protocols in Immunology", J.E. Coligan *et al.*, Ed. Greene Publishing Associates y Wiley-Interscience; Weinberger *et al.*, *Proc. Natl. Acad. Sci. USA* 77: 6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immun.*, 11: 405-411, 1981. Takai *et al.*, *J. Immunol.*, 137: 3494-3500, 1986. Takai *et al.*, *J. Immunol.*, 140: 508-512, 1988.

10

Aquellas proteínas que exhiben actividad de citocina, proliferación celular o diferenciación celular pueden formularse entonces como productos farmacéuticos y usarse para tratar afecciones en las que es beneficiosa la inducción de proliferación o diferenciación celular. Como alternativa, como se describe con más detalle a continuación, pueden introducirse en células hospedadoras adecuadas genes que codifican estas proteínas o ácidos nucleicos que regulan la expresión de estas proteínas para aumentar o reducir la expresión de las proteínas según se desee.

15

20

Ejemplo 33

Ensayo en las proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de actividad como reguladores del sistema inmunitario

25

Pueden evaluarse también en las proteínas codificadas por los ADNc sus efectos como reguladores inmunitarios. Por ejemplo, puede evaluarse en las proteínas su actividad para influir en la citotoxicidad de timocito o esplenocito. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos descritos en las siguientes referencias: capítulo 3 (*In Vitro Assays for Mouse Lymphocyte Function* 3.1-3.19) y capítulo 7 (Immunologic studies in Humans) en "Current Protocols in Immunology", J.E. Coligan *et al.*, Ed. Greene Publishing Associates y Wiley-Interscience; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA*, 78: 2488-2492, 1981; Herrmann *et al.*, *J. Immunol.*, 128: 1968-1974, 1982; Handa *et al.*, *J. Immunol.*, 135: 1564-1572, 1985; Takai *et al.*, *J. Immunol.*, 137: 3494-3500, 1986; Takai *et al.*, *J. Immunol.*, 140: 508-512, 1988; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA*, 78: 2488-2492, 1981; Herrmann *et al.*, *J. Immunol.*, 128: 1968-1974, 1982; Handa *et al.*, *J. Immunol.*, 135: 1564-1572, 1985; Takai *et al.*, *J. Immunol.*, 137: 3494-3500, 1986; Bowman *et al.*, *J. Virology* 61: 1992-1998; Takai *et al.*, *J. Immunol.*, 140: 508-512, 1988; Bertagnolli *et al.*, *Cellular Immunology* 133: 327-341, 1991; Brown *et al.*, *J. Immunol.*, 153: 3079-3092, 1994.

30

35

Pueden evaluarse también en las proteínas codificadas por los ADNc sus efectos sobre respuestas de inmunoglobulina dependientes de linfocitos T y el cambio de isotipo. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: Maliszewski, *J. Immunol.*, 144: 3028-3033, 1990; Mond, J.J. y Brunswick, M "Assays for B Cell Function": "In vitro Antibody Production", Vol 1 pág. 3.8.1-3.8.16 en "Current Protocols in Immunology", J.E. Coligan *et al.* Ed., John Wiley and Sons, Toronto, 1994.

40

Puede evaluarse también en las proteínas codificadas por los ADNc su efecto sobre células efectoras inmunitarias, incluyendo su efecto sobre células Th1 y linfocitos citotóxicos. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: capítulo 3 (*In Vitro Assays for Mouse Lymphocyte Function* 3.1-3.19) y capítulo 7 (Immunologic studies in Humans) en "Current Protocols in Immunology", J.E. Coligan *et al.* Ed. Greene Publishing Associates y Wiley-Interscience; Takai *et al.*, *J. Immunol.*, 137: 3494-3500, 1986; Takai *et al.*, *J. Immunol.*, 140: 508-512, 1988; Bertagnolli *et al.*, *J. Immunol.*, 149: 3778-3783, 1992.

50

Puede evaluarse también en las proteínas codificadas por los ADNc su efecto sobre la activación mediada por células dendríticas en linfocitos T indiferenciados. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: Guery *et al.*, *J. Immunol.*, 134: 536-544, 1995; Inaba *et al.*, *Journal of Experimental Medicine* 173: 549-559, 1991; Macatonia *et al.*, *Journal of Immunology* 154: 5071-5079, 1995; Porgador *et al.*, *Journal of Experimental Medicine* 182: 255-260, 1995; Nair *et al.*, *Journal of Virology* 67: 4062-4069, 1993; Huang *et al.*, *Science* 264: 961-965, 1994; Macatonia *et al.*, *Journal of Experimental Medicine* 169: 1255-1264, 1989; Bhardwaj *et al.*, *Journal of Clinical Investigation* 94: 797-807, 1994; e Inaba *et al.*, *Journal of Experimental Medicine* 172: 631-640, 1990.

55

60

Puede evaluarse también en las proteínas codificadas por los ADNc su influencia sobre el tiempo de vida de los linfocitos. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: Darzynkiewicz *et al.*, *Cytometry* 13: 795-808, 1992; Gorczyca *et al.*, *Leukemia* 7: 659-670, 1993; Gorczyca *et al.*, *Cancer Research* 53: 1945-1951, 1993; Itoh *et al.*, *Cell* 66: 233, 1991; Zacharchuk, *Journal of Immunology* 145: 4037-4045, 1990; Zamai *et al.*, *Cytometry* 14: 891-897, 1993; Gorczyca *et al.*, *International Journal of Oncology* 1: 639-648, 1992.

65

Los ensayos de proteínas que influyen en las etapas tempranas de caracterización y desarrollo de linfocitos T incluyen, sin limitación, los descritos en: Antica *et al.*, *Blood* 84: 111-117, 1994; Fine *et al.* *Cellular Immunology* 155: 111-122, 1994; Galy *et al.*, *Blood* 85: 2770-2778, 1995; Toki *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 7548-7551, 1991.

Aquellas proteínas que exhiben actividad como reguladores del sistema inmunitario pueden formularse entonces como productos farmacéuticos y usarse para tratar afecciones clínicas en las que la regulación de la actividad inmunitaria es beneficiosa. Por ejemplo, la proteína puede ser útil en el tratamiento de diversas deficiencias y trastornos inmunitarios (incluyendo inmunodeficiencia combinada grave (IDCG)), por ejemplo, en la regulación (positiva o negativa) del crecimiento y proliferación de linfocitos T y/o B, así como afectando la actividad citolítica de células NK y otras poblaciones celulares. Estas deficiencias inmunitarias pueden ser genéticas o causadas por infecciones víricas (por ejemplo VIH), así como bacterianas o fúngicas, o pueden ser el resultado de trastornos autoinmunitarios. Más específicamente, las enfermedades infecciosas causadas por infección vírica, bacteriana, fúngica u otras pueden ser tratables usando una proteína de la presente invención, incluyendo infecciones por VIH, virus de hepatitis, herpesvirus, micobacterias, *Leishmania spp.*, *Malaria spp.* y diversas infecciones fúngicas tales como candidiasis. Por supuesto, a este respecto, una proteína puede ser útil también cuando puede ser deseable en general un refuerzo del sistema inmunitario, concretamente en el tratamiento de cáncer.

Los trastornos autoinmunitarios que pueden tratarse usando una proteína incluyen, por ejemplo, enfermedad de tejido conectivo, esclerosis múltiple, lupus sistémico eritematoso, artritis reumatoide, inflamación pulmonar autoinmunitaria, síndrome de Guillain-Barre, tiroiditis autoinmunitaria, diabetes mellitus insulín dependiente, miastenia grave, enfermedad del injerto contra el hospedador y enfermedad ocular inflamatoria autoinmunitaria. Dicha proteína de la presente invención puede ser también útil en el tratamiento de reacciones y afecciones alérgicas, tales como asma (particularmente asma alérgica) u otros problemas respiratorios. Otras afecciones en las que se desea la supresión inmunitaria (incluyendo por ejemplo trasplante de órganos) pueden ser también tratables usando una proteína.

Usando las proteínas, puede ser posible también regular las respuestas inmunitarias de una serie de modos. La regulación negativa puede ser en forma de inhibición o bloqueo de una respuesta inmunitaria ya en progreso o puede implicar evitar la inducción de una respuesta inmunitaria. Las funciones de linfocitos T activados pueden inhibirse suprimiendo las respuestas de linfocitos T o induciendo tolerancia específica en linfocitos T, o ambas cosas. La inmunosupresión de respuestas de linfocito T es generalmente un proceso activo, no específico de antígeno, que requiere una exposición continua de los linfocitos T al agente supresor. La tolerancia, que implica inducir insensibilidad o anergia en linfocitos T, es distinguible de la inmunosupresión en que generalmente es específica de antígeno y persiste después de cesar la exposición al agente tolerizante. Operativamente, la tolerancia puede demostrarse por la carencia de una respuesta de linfocitos T a la reexposición a antígeno específico en ausencia del agente tolerizante.

La regulación negativa o prevención de una o más funciones antigénicas (incluyendo sin limitación las funciones antigénicas de linfocitos B (tales como, por ejemplo, B7)), por ejemplo evitando la síntesis de linfocina a alto nivel por linfocitos T activados, será útil en situaciones de trasplante de tejido, piel y órganos y en la enfermedad del injerto contra el hospedador (EICH). Por ejemplo, el bloqueo de la función de linfocitos T debe dar como resultado una destrucción de tejido reducida en el trasplante de tejido. Típicamente, en trasplantes de tejido, el rechazo del trasplante se inicia mediante su reconocimiento como extraño por linfocitos T, seguido de una reacción inmunitaria que destruye el trasplante. La administración de una molécula que inhibe o bloquea la interacción de un antígeno de linfocito B7 con su(s) ligando(s) natural(es) en células inmunitarias (tales como una forma soluble monomérica de un péptido que tiene actividad B7-2 solo o junto con una forma monomérica de un péptido que tiene una actividad de otro antígeno de linfocito B (por ejemplo, B7-1, B7-3) o anticuerpo de bloqueo), antes del trasplante puede conducir a la unión de la molécula al(a los) ligando(s) natural(es) en las células inmunitarias sin transmitir la señal coestimulante correspondiente. El bloqueo de función de antígeno de linfocito B evita a este respecto la síntesis de citocina por células inmunitarias, tales como linfocitos T, y actúa por tanto como un inmunosupresor. Además, la falta de coestimulación puede ser suficiente también para anergizar los linfocitos T, induciendo así tolerancia en un sujeto. La inducción de tolerancia a largo plazo por reactivos de bloqueo de antígeno de linfocito B puede evitar la necesidad de una administración repetida de estos reactivos de bloqueo. Para conseguir una inmunosupresión o tolerancia suficiente en un sujeto, puede ser también necesario bloquear la función de una combinación de antígenos de linfocito B.

Puede evaluarse la eficacia de reactivos de bloqueo particulares en la prevención del rechazo de trasplante de órganos o EICH usando modelos animales que son predictivos de eficacia en seres humanos. Los ejemplos de sistemas apropiados que pueden usarse incluyen injertos cardiacos alogénicos en ratas e injertos de célula de islote pancreático xenogénicos en ratones, habiéndose usado ambos para examinar los efectos inmunosupresores de proteínas de fusión CTLA4Ig *in vivo* como se describe en Lenschow *et al.*, *Science* 257: 789-792 (1992) y Turka *et al.*, *Proc. Natl. Acad. Sci. USA* 89: 11102-11105, (1992). Además, los modelos de mûrido de EICH (véase Paul ed., "Fundamental Immunology", Raven Press, Nueva York, 1989, pág. 846-847) pueden usarse para determinar el efecto de bloquear la función antigénica de linfocitos B *in vivo* sobre el desarrollo de esa enfermedad.

El bloqueo de la función antigénica puede ser también terapéuticamente útil para tratar enfermedades autoinmunitarias. Muchas enfermedades autoinmunitarias son el resultado de una activación inapropiada de linfocitos T que son reactivos contra tejido propio y que promueven la producción de citocinas y autoanticuerpos implicados en la patología de las enfermedades. La prevención de la activación de linfocitos T autorreactivos puede reducir o eliminar los síntomas de enfermedad. La administración de reactivos que bloquean la coestimulación de linfocitos T desesta-

bilizando las interacciones de ligando-receptor de antígenos de linfocitos B puede usarse para inhibir la activación de linfocitos T y evitar la producción de autoanticuerpos o citocinas derivadas de linfocitos T que pueden estar implicados en el proceso de la enfermedad. Adicionalmente, los reactivos de bloqueo pueden inducir tolerancia específica de antígeno de linfocitos T autorreactivos, que podría conducir al alivio a largo plazo de la enfermedad. La eficacia de los reactivos de bloqueo en la prevención o alivio de los trastornos autoinmunitarios puede determinarse usando una serie de modelos animales bien caracterizados de enfermedades autoinmunitarias humanas. Los ejemplos incluyen encefalitis autoinmunitaria experimental de mûrido, lupus sistémico eritematoso en ratones MRL/pr/pr o ratones híbridos NZB, artritis colagenosa autoinmunitaria de mûrido, diabetes mellitus en ratones OD y ratas BB, y miastenia grave experimental de mûrido (véase Paul ed., "Fundamental Immunology", Raven Press, Nueva York, 1989, pág. 840-856).

La regulación positiva de una función antigénica (preferiblemente una función antigénica de linfocito B), como medio de regular positivamente respuestas inmunitarias, puede ser también útil en terapia. La regulación positiva de respuestas inmunitarias puede ser en forma de potenciar una respuesta inmunitaria existente o desencadenando una respuesta inmunitaria inicial. Por ejemplo, potenciar una respuesta inmunitaria mediante la estimulación de la función antigénica de linfocitos B puede ser útil en casos de infección vírica. Además, podrían aliviarse enfermedades víricas sistémicas tales como gripe, resfriado común y encefalitis mediante la administración de la forma estimulante de antígenos de linfocito B sistémicamente.

Como alternativa, las respuestas inmunitarias antivíricas pueden potenciarse en un paciente infectado retirando linfocitos T del paciente, coestimulando los linfocitos T *in vitro* con APC con pulsos de antígeno vírico que expresan un péptido de la presente invención o junto con una forma estimulante de un péptido soluble de la presente invención y reintroduciendo los linfocitos T activados *in vitro* en el paciente. Las células infectadas serían capaces entonces de suministrar una señal coestimulante a linfocitos T *in vivo*, activando así los linfocitos T.

En otra aplicación, la regulación positiva o potenciación de la función antigénica (preferiblemente función antigénica de linfocito B) puede ser útil en la inducción de inmunidad tumoral. Pueden administrarse a un sujeto células tumorales (por ejemplo, sarcoma, melanoma, linfoma, leucemia, neuroblastoma, carcinoma) transfectadas con un ácido nucleico que codifica al menos un péptido de la presente invención para superar la tolerancia específica de tumor en el sujeto. Si se desea, el tumor puede transfectarse para expresar una combinación de péptidos. Por ejemplo, las células tumorales obtenidas a partir de un paciente pueden transfectarse *ex vivo* con un vector de expresión que dirige la expresión de un péptido que tiene actividad similar a B7-2 solo o junto con un péptido que tiene actividad similar a B7-1 y/o actividad similar a B7-3. Se devuelven las células tumorales transfectadas al paciente para dar como resultado la expresión de los péptidos en la superficie de la célula transfectada. Como alternativa, pueden usarse técnicas de terapia génica para orientar una célula tumoral para transfección *in vivo*.

La presencia de un péptido que tiene la actividad de antígeno(s) de linfocito B en la superficie de la célula tumoral proporciona la señal de coestimulación necesaria a linfocitos T para inducir una respuesta inmunitaria mediada por linfocitos T contra las células tumorales transfectadas. Además, pueden transfectarse células tumorales que carecen de moléculas MHC de clase I o MHC de clase II, o que no consiguen reexpresar cantidades suficientes de moléculas MHC de clase I o MHC de clase II, con ácidos nucleicos que codifican toda o una porción (por ejemplo una porción truncada de dominio citoplasmático) de una proteína de cadena α de MHC de clase I y proteína macroglobulina β_2 o una proteína de cadena α de MHC de clase II y una proteína de cadena β de MHC de clase II para expresar así proteínas de MHC de clase I o proteínas de MHC de clase II en la superficie celular. La expresión de las MHC de clase I o clase II apropiadas junto con un péptido que tiene actividad de antígeno de linfocito B (por ejemplo, B7-1, B7-2, B7-3) induce una respuesta inmunitaria mediada por linfocito T contra la célula tumoral transfectada. Opcionalmente, puede cotransfectarse también un gen que codifica un constructo no codificante que bloquea la expresión de una proteína asociada a MHC de clase II, tal como la cadena invariante, con un ADN que codifica un péptido que tiene actividad de antígeno de linfocito B para promover la presentación de antígenos asociados a tumor e inducir inmunidad específica de tumor. Por tanto, la inducción de una respuesta inmunitaria mediada por linfocito T en un sujeto humano puede ser suficiente para superar la tolerancia específica de tumor en el sujeto. Como alternativa, como se describe con más detalle a continuación, pueden introducirse en células hospedadoras adecuadas genes que codifican estas proteínas o ácidos nucleicos que regulan la expresión de estas proteínas para aumentar o reducir la expresión de las proteínas según se desee.

Ejemplo 34

Ensayo en las proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de actividad como reguladores de la hematopoyesis

En las proteínas codificadas por los ADNc extendidos o porciones de los mismos, puede evaluarse también su actividad reguladora de la hematopoyesis. Por ejemplo, puede evaluarse el efecto de las proteínas sobre la diferenciación de células madre embrionarias. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: Johansson *et al.* *Cellular Biology* 15: 141-151, 1995; Keller *et al.*, *Molecular and Cellular Biology* 13: 473-486, 1993; McClanahan *et al.*, *Blood* 81: 2903-2915, 1993.

En las proteínas codificadas por los ADNc extendidos o porciones de los mismos, puede evaluarse también su influencia sobre el tiempo de vida de células madre y la diferenciación de células madre. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: Freshney, M.G. "Methylcellulose Colony Forming Assays", en "Culture of Hematopoietic Cells", R.I. Freshney, *et al.* Ed. págs. 265-268, Wiley-Liss, Inc., Nueva York, NY, 1994; Hirayama *et al.*, *Proc. Natl. Acad. Sci. USA*, 89: 5907-5911, 1992; McNiece, I.K. y Briddell, R.A. "Primitive Hematopoietic Colony Forming Cells with High Proliferative Potential", en "Culture of Hematopoietic Cells", R.I. Freshney, *et al.* ed. Vol págs. 23-39, Wiley-liss, Inc., Nueva York, NY, 1994; Neben *et al.*, *Experimental Hematology* 22: 353-359, 1994; Ploemacher, R.E. Cobblestone "Area Forming Cell Assay", en "Culture of Hematopoietic Cells", R.I. Freshney, *et al.* Ed. págs. 1-21, Wiley-liss, Inc., Nueva York, NY, 1994; Spooncer, E., Dexter, M. y Allen, T. "Long Term Bone Marrow Cultures in the Presence of Stromal Cells", en "Culture of Hematopoietic Cells", R.I. Freshney, *et al.* Ed. págs. 163-179, Wiley-liss, Inc., Nueva York, NY, 1994; y Sutherland, H.J. "Long Term Culture Initiating Cell Assay", en "Culture of Hematopoietic Cell", R.I. Freshney, *et al.* Ed. págs. 139-162, Wiley-Liss, Inc., Nueva York, NY, 1994.

Aquellas proteínas que exhiben actividad reguladora de la hematopoyesis pueden formularse entonces como productos farmacéuticos y usarse para tratar afecciones clínicas en las que la regulación de la hematopoyesis es beneficiosa, por ejemplo, una proteína puede ser útil en la regulación de la hematopoyesis y, en consecuencia, en el tratamiento de deficiencias de células mieloides o linfoides. Incluso una actividad biológica marginal en apoyo de células formadoras de colonia o estirpes celulares dependientes de factor, indica implicación en la regulación de la hematopoyesis, por ejemplo, en apoyo del crecimiento y proliferación de células progenitoras eritroides solas o en combinación con otras citocinas, indicando así la utilidad, por ejemplo, en el tratamiento de diversas anemias o para uso junto con irradiación/quimioterapia para estimular la producción de precursores eritroides y/o células eritroides; en el apoyo del crecimiento y proliferación de células mieloides tales como granulocitos y monocitos/macrófagos (concretamente, actividad CSF tradicional) útil, por ejemplo, junto con quimioterapia para evitar y tratar la mielosupresión consiguiente; en apoyo del crecimiento y proliferación de megacariocitos y en consecuencia de plaquetas, permitiendo así la prevención o el tratamiento de diversos trastornos de plaquetas tales como trombocitopenia, y generalmente para uso en lugar de o complementario de transfusiones de plaquetas; y/o en apoyo del crecimiento y proliferación de células madre hematopoyéticas que son capaces de madurar a una cualquiera de las células hematopoyéticas anteriormente mencionadas y por lo tanto encuentran utilidad terapéutica en diversos trastornos de células madre (tales como aquellos tratados habitualmente con trasplantes, incluyendo sin limitación, anemia aplásica y hemoglobinuria nocturna paroxísmica), así como en la despoblación del compartimento de células madre después de irradiación/quimioterapia, *in vivo* o *ex vivo* (concretamente, junto con trasplante de médula ósea o con trasplante de células progenitoras periféricas (homólogas o heterólogas)) como células normales o manipuladas genéticamente para terapia génica. Como alternativa, como se describe con más detalle a continuación, pueden introducirse en células hospedadoras adecuadas genes que codifican estas proteínas o ácidos nucleicos que regulan la expresión de estas proteínas para aumentar o reducir la expresión de las proteínas según se desee.

Ejemplo 35

Ensayo de proteínas expresadas a partir de ADNc extendidos o porciones de los mismos para la regulación del crecimiento de tejido

Puede evaluarse también en las proteínas codificadas por los ADNc extendidos o porciones de los mismos su efecto sobre el crecimiento de tejido. Son familiares para los expertos en la técnica numerosos ensayos de dicha actividad, incluyendo los ensayos dados a conocer en la publicación de patente internacional n° WO95/16035, publicación de patente internacional n° WO95/05846 y publicación de patente internacional n° WO91/07491.

Los ensayos de actividad de curación de heridas incluyen, sin limitación, aquellos descritos en: Winter, "Epidermal Wound Healing", págs. 71-112 (Maibach, H.I. y Rovee, D.T., ed.). Year Book Medical Publishers, Inc., Chicago, modificado por Eaglstein y Mertz, *J. Invest. Dermatol.* 71: 382-84 (1978).

Aquellas proteínas que están implicadas en la regulación del crecimiento de tejido pueden formularse entonces como productos farmacéuticos y usarse para tratar afecciones clínicas en las que la regulación del crecimiento de tejido es beneficiosa. Por ejemplo, una proteína puede tener también utilidad en composiciones usadas para el crecimiento o la regeneración de tejido de hueso, cartílago, tendón, ligamento y/o nervio, así como para la curación de heridas y la reparación y sustitución de tejido, y en el tratamiento de quemaduras, incisiones y úlceras.

Una proteína que induce el crecimiento de cartílago y/o hueso en circunstancias en que normalmente no se forma hueso tiene aplicación en la curación de fracturas óseas y lesión de cartílago o defectos en seres humanos y otros animales. Dicha preparación que emplea una proteína puede tener uso profiláctico en la reducción de fracturas cerradas así como abiertas, y también en la fijación mejorada de articulaciones artificiales. La formación ósea *de novo* inducida por un agente osteogénico contribuye a la reparación de defectos craneofaciales congénitos, inducidos por traumatismo o inducidos por resección oncológica, y también es útil en cirugía plástica cosmética.

Puede usarse también una proteína en el tratamiento de enfermedad periodontal, y en otros procesos de reparación dental. Dichos agentes pueden proporcionar un entorno para atraer células formadoras de hueso, estimular el crecimiento de células formadoras de hueso o inducir la diferenciación de progenitores de células formadoras de hueso.

Una proteína puede ser útil también en el tratamiento de osteoporosis u osteoartritis, tal como mediante la estimulación de la reparación de hueso y/o cartílago o mediante el bloqueo de la inflamación o procesos de destrucción de tejido (actividad colagenasa, actividad de osteoclasto, etc.) mediados por procesos inflamatorios.

5 Otra categoría de actividad de regeneración de tejido que puede ser atribuible a una proteína es la formación de tendón/ligamento. Una proteína que induce la formación de tejido similar a tendón/ligamento u otro tejido en circunstancias en que dicho tejido no se forma normalmente, tiene aplicación en la curación de desgarros de tendón o ligamento, deformidades y otros defectos de tendón o ligamento en seres humanos y otros animales. Dicha preparación que emplea una proteína inductora de tejido similar a tendón/ligamento puede tener uso profiláctico en la
10 prevención de daño en tejido de tendón o ligamento, así como uso en la fijación mejorada de tendón o ligamento a hueso u otros tejidos, y en la reparación de defectos de tejido de tendón o ligamento. La formación de tejido similar a tendón/ligamento *de novo* inducida por una composición contribuye a la reparación de defectos de tendón o ligamento congénitos, inducidos por traumatismo o de otro origen, y es también útil en cirugía plástica cosmética para la unión o reparación de tendones o ligamentos. Las composiciones pueden proporcionar un entorno para atraer células
15 formadoras de tendón o ligamento, estimular el crecimiento de células formadoras de tendón o ligamento, inducir la diferenciación de progenitores de células formadoras de tendón o ligamento, o inducir el crecimiento de células de tendón/ligamento o progenitores *ex vivo* para volver a efectuar la reparación de tejido *in vivo*. Las composiciones pueden ser útiles también en el tratamiento de tendinitis, síndrome del túnel carpiano y otros defectos de tendón o ligamento. Las composiciones pueden incluir también una matriz y/o agente secuestrante apropiado como portador
20 como es bien conocido en la técnica.

La proteína puede ser también útil para la proliferación de células neurales y para la regeneración de tejido de nervios y cerebro, concretamente, para el tratamiento de enfermedades y neuropatías del sistema nervioso central y periférico, así como de trastornos mecánicos y traumáticos que implican la degeneración, muerte o traumatismo de células neurales o tejido nervioso. Más específicamente, puede usarse la proteína en el tratamiento de enfermedades del
25 sistema nervioso periférico, tales como lesiones de nervios periféricos, neuropatía periférica y neuropatías localizadas, y enfermedades del sistema nervioso central tales como enfermedad de Alzheimer, Parkinson, Huntington, esclerosis lateral amiotrófica y síndrome de Shy-Drager. Las afecciones adicionales que pueden tratarse incluyen trastornos mecánicos y traumáticos, tales como trastornos de la médula espinal, traumatismo cerebral y enfermedades cerebrovasculares tales como apoplejía. Pueden ser también tratables usando la proteína neuropatías periféricas resultantes de quimioterapia u otras terapias médicas.
30

Las proteínas pueden ser útiles también para promover un cierre mejor o más rápido de heridas que no curan, incluyendo sin limitación úlceras de presión, úlceras asociadas a insuficiencia vascular, heridas quirúrgicas y traumáticas y similares.
35

Se espera que una proteína pueda exhibir también actividad para la generación o regeneración de otros tejidos, tales como tejido de órganos (incluyendo, por ejemplo, páncreas, hígado, intestino, riñón, piel, endotelio), músculo (liso, esquelético o cardíaco) y vascular (incluyendo endotelio vascular), o para promover el crecimiento de células
40 que comprenden dichos tejidos. Parte de los efectos deseados puede ser mediante la inhibición o modulación de cicatrización fibrótica para permitir que se genere un tejido normal. Una proteína puede exhibir también actividad angiogénica.

Una proteína puede ser también útil para la protección o regeneración intestinal y el tratamiento de fibrosis pulmonar o hepática, lesión por reperfusión en diversos tejidos, y afecciones resultantes de lesión por citocina sistémica.
45

Una proteína puede ser también útil para promover o inhibir la diferenciación de tejidos descritos anteriormente a partir de tejidos o células precursores; o para inhibir el crecimiento de tejidos descritos anteriormente.

50 Como alternativa, como se describe con más detalle a continuación, pueden introducirse en células hospedadoras adecuadas genes que codifican estas proteínas o ácidos nucleicos que regulan la expresión de estas proteínas para aumentar o reducir la expresión de las proteínas según se desee.

55 Ejemplo 36

Ensayo de proteínas expresadas a partir de ADNc extendidos o porciones de los mismos para la regulación de hormonas reproductivas o el movimiento celular

60 Puede evaluarse también en las proteínas codificadas por los ADNc extendidos su capacidad de regular hormonas reproductivas, tales como hormona estimulante de folículo. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: Vale *et al.*, *Endocrinology* 91: 562-572, 1972; Ling *et al.* *Nature* 321: 779-782, 1986; Vale *et al.*, *Nature* 321: 776-779, 1986; Mason *et al.*, *Nature* 318: 659-663, 1985; Forage *et al.*, *Proc. Natl. Acad. Sci. USA*, 83: 3091-3095, 1986; capítulo 6.12 (Measurement of Alpha and Beta Chemokines) "Current Protocols in Immunology", J.E. Coligan *et al.* Ed. Greene Publishing Associates y Wiley-Interscience; Taub *et al.* *J. Clin. Invest.* 95: 1370-1376, 1995; Lind *et al.* *APMIS* 103: 140-146, 1995; Muller *et al.* *Eur. J. Immunol.*, 25: 1744-1748; Gruber *et al.* *J. of Immunol.* 152: 5860-5867, 1994; Johnston *et al.* *J. of Immunol.* 153: 1762-1768, 1994.
65

Pueden formularse entonces aquellas proteínas que exhiben actividad como hormonas reproductivas o reguladores del movimiento celular como productos farmacéuticos y usarse para tratar afecciones clínicas en las que es beneficiosa la regulación de hormonas reproductivas o el movimiento celular. Por ejemplo, una proteína puede exhibir también actividades relacionadas con activina o inhibina. Las inhibinas se caracterizan por su capacidad de inhibir la liberación de hormona estimulante de folículo (FSH), mientras que las activinas se caracterizan por su capacidad de estimular la liberación de hormona estimulante de folículo (FSH). Por tanto, una proteína sola o en heterodímeros con un miembro de la familia de inhibina α puede ser útil como anticonceptivo basándose en la capacidad de las inhibinas de reducir la fertilidad en mamíferos hembra y reducir la espermatogénesis en mamíferos macho. La administración de cantidades suficientes de otras inhibinas puede inducir la infertilidad en estos mamíferos. Como alternativa, la proteína, en forma de homodímero o heterodímero con otras subunidades proteicas del grupo de inhibina B, puede ser útil como producto terapéutico inductor de la fertilidad, basándose en la capacidad de las moléculas de activina de estimular la liberación de FSH de células de la pituitaria anterior. Véase, por ejemplo, la patente de Estados Unidos 4.798.885. Una proteína puede ser también útil para el adelantamiento del inicio de la fertilidad en mamíferos sexualmente inmaduros y para aumentar la actuación reproductiva a lo largo de la vida de animales domésticos tales como vacas, ovejas y cerdos.

Como alternativa, como se describe con más detalle a continuación, pueden introducirse en células hospedadoras adecuadas genes que codifican estas proteínas o ácidos nucleicos que regulan la expresión de estas proteínas para aumentar o reducir la expresión de las proteínas según se desee.

Ejemplo 36A

Ensayo en las proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de actividad quimiotáctica/quimiocinética

En las proteínas codificadas por los ADNc extendidos o porciones de los mismos, puede evaluarse también la actividad quimiotáctica/quimiocinética. Por ejemplo, una proteína puede tener actividad quimiotáctica o quimiocinética (por ejemplo, actuar como una quimiocina) para células de mamífero incluyendo, por ejemplo, monocitos, fibroblastos, neutrófilos, linfocitos T, mastocitos, eosinófilos, células epiteliales y/o endoteliales. Las proteínas quimiotácticas y quimiocinéticas pueden usarse para movilizar o atraer una población celular deseada a un sitio de acción deseado. Las proteínas quimiotácticas o quimiocinéticas proporcionan ventajas particulares en el tratamiento de heridas y otros traumatismos a tejidos, así como en el tratamiento de infecciones localizadas. Por ejemplo, la atracción de linfocitos, monocitos o neutrófilos a tumores o sitios de infección puede dar como resultado respuestas inmunitarias mejoradas contra el tumor o agente infeccioso.

Una proteína o péptido tiene actividad quimiotáctica para una población celular particular si puede estimular, directa o indirectamente, la orientación o el movimiento dirigido de dicha población celular. Preferiblemente, la proteína o péptido tiene la capacidad de estimular directamente el movimiento dirigido de células. Que una proteína particular tenga actividad quimiotáctica para una población de células puede determinarse fácilmente empleando dicha proteína o péptido en cualquier ensayo conocido de quimiotactismo celular.

La actividad de una proteína de la invención puede medirse, entre otros medios, mediante los siguientes métodos:

Los ensayos de actividad quimiotáctica (que identificarán proteínas que inducen o previenen el quimiotactismo) consisten en ensayos que miden la capacidad de una proteína de inducir la migración de células a través de una membrana así como la capacidad de una proteína de inducir la adhesión de una población celular a otra población celular. Los ensayos adecuados para movimiento y adhesión incluyen, sin limitación, aquellos descritos en: "Current Protocols in Immunology", Ed por J.E. Coligan, A.M. Kruisbeek, O.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates y Wiley-Interscience (capítulo 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub *et al. J. Clin. Invest.* 95, 1370-1376, 1995; Lind *et al. APMIS* 103: 140-146, 1995; Mueller *et al., Eur. J. Immunol.*, 25: 1744-1748; Gruber *et al. J. of Immunol.* 152: 5860-5867, 1994; Johnston *et al. J. of Immunol.* 153: 1762-1768, 1994.

Ejemplo 37

Ensayo en proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de la regulación de la coagulación sanguínea

Puede evaluarse también en las proteínas codificadas por los ADNc extendidos o porciones de los mismos sus efectos sobre la coagulación sanguínea. Son familiares para los expertos en la técnica numerosos ensayos para dicha actividad, incluyendo los ensayos dados a conocer en las siguientes referencias: Linet *et al., J. Clin. Pharmacol.* 26: 131-140, 1986; Burdick *et al., Thrombosis Res.* 45: 413-419, 1987; Humphrey *et al, Fibrinolysis* 5: 71-79 (1991); Schaub, *Prostaglandins* 35: 467-474, 1988.

Aquellas proteínas que están implicadas en la regulación de la coagulación sanguínea pueden formularse entonces como productos farmacéuticos y usarse para tratar afecciones clínicas en las que es beneficiosa la regulación de la coagulación sanguínea. Por ejemplo, una proteína puede exhibir actividad hemostática o trombolítica. Como resultado, se espera que dicha proteína sea útil en el tratamiento de diversos trastornos de coagulación (incluyendo trastornos hereditarios tales como hemofilias) o para potenciar la coagulación y otros eventos hemostáticos en el tratamiento de

heridas resultantes de traumatismo, cirugía u otras causas. Una proteína puede ser también útil para disolver o inhibir la formación de trombosis y para el tratamiento y la prevención de afecciones resultantes de la misma (tales como, por ejemplo, infarto de vasos cardíacos y del sistema nervioso central (por ejemplo, apoplejía)). Como alternativa, como se describe con más detalle a continuación, pueden introducirse en células hospedadoras adecuadas genes que codifican estas proteínas o ácidos nucleicos que regulan la expresión de estas proteínas para aumentar o reducir la expresión de las proteínas según se desee.

Ejemplo 38

Ensayo en proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de la implicación en interacciones de receptor/ligando

Puede evaluarse también en las proteínas codificadas por los ADNc extendidos o una porción de los mismos su implicación en interacciones de receptor/ligando. Son familiares para los expertos en la técnica numerosos ensayos para dicha implicación, incluyendo los ensayos dados a conocer en las siguientes referencias: capítulo 7.28 (Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22) en "Current Protocols in Immunology". J.E. Coligan *et al.* Ed. Greene Publishing Associates y Wiley-Interscience; Takai *et al.*, *Proc. Natl. Acad. Sci. USA*, 84: 6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168: 1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169: 149-160, 1989; Stoltzenberg *et al.*, *J. Immunol., Methods* 175: 59-68, 1994; Stitt *et al.*, *Cell* 80: 661-670, 1995; Gyuris *et al.*, *Cell* 75: 791-803, 1993.

Por ejemplo, las proteínas pueden demostrar también actividad como receptores, ligandos de receptor o inhibidores o agonistas de interacciones de receptor/ligando. Los ejemplos de dichos receptores y ligandos incluyen, sin limitación, receptores de citocina y sus ligandos, cinasas receptoras y sus ligandos, fosfatasa receptoras y sus ligandos, receptores implicados en interacciones célula-célula y sus ligandos (incluyendo, sin limitación, moléculas de adhesión celular (tales como selectinas, integrinas y sus ligandos) y pares de receptor/ligando implicados en la presentación de antígeno, el reconocimiento de antígeno y el desarrollo de respuestas inmunitarias celulares y humorales). Los receptores y ligandos son también útiles para examinar inhibidores potenciales peptídicos o de molécula pequeña de la interacción receptor/ligando relevante. Una proteína (incluyendo, sin limitación, fragmentos de receptores y ligandos) puede ser útil por sí misma como inhibidor de interacciones de receptor/ligando.

Ejemplo 38A

Ensayo en las proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de actividad antiinflamatoria

En las proteínas codificadas por los ADNc extendidos o una porción de los mismos puede evaluarse también la actividad antiinflamatoria. La actividad antiinflamatoria puede conseguirse proporcionando un estímulo a células implicadas en la respuesta inflamatoria, inhibiendo o promoviendo interacciones célula-célula (tales como, por ejemplo, adhesión celular), inhibiendo o promoviendo el quimiotactismo de células implicadas en el proceso inflamatorio, inhibiendo o promoviendo la extravasación celular o estimulando o suprimiendo la producción de otros factores que inhiben o promueven más directamente una respuesta inflamatoria. Las proteínas que exhiben dichas actividades pueden usarse para tratar afecciones inflamatorias (incluyendo afecciones crónicas o agudas) incluyendo, sin limitación, inflamación asociada a infección (tal como choque séptico, sepsis o síndrome de respuesta inflamatoria sistémica (SRIS)), lesión por isquemia-reperusión, letalidad por endotoxina, artritis, rechazo hiperagudo mediado por complemento, nefritis, lesión pulmonar inducida por citocina o quimiocina, enfermedad intestinal inflamatoria, enfermedad de Crohn o resultante de la sobreproducción de citocinas tales como TNF o IL-1. Las proteínas pueden ser también útiles para tratar anafilaxis e hipersensibilidad a una sustancia o material antigénico.

Ejemplo 38B

Ensayo en las proteínas expresadas a partir de ADNc extendidos o porciones de los mismos de actividad inhibidora de tumores

En las proteínas codificadas por los ADNc extendidos o una porción de los mismos puede evaluarse también la actividad inhibidora de tumores. Además de las actividades descritas anteriormente para el tratamiento o prevención inmunológico de tumores, una proteína puede exhibir otras actividades antitumorales. Una proteína puede inhibir el crecimiento tumoral directa o indirectamente (tal como, por ejemplo, mediante ADCC). Una proteína puede exhibir su actividad inhibidora de tumores actuando sobre tejido tumoral o tejido precursor tumoral, inhibiendo la formación de tejidos necesarios para apoyar el crecimiento tumoral (tal como, por ejemplo, inhibiendo la angiogénesis), causando la producción de otros factores, agentes o tipos celulares que inhiben el crecimiento celular, o suprimiendo, eliminando o inhibiendo factores, agentes o tipos celulares que promueven el crecimiento tumoral.

Una proteína puede exhibir también una o más de las siguientes actividades o efectos adicionales: inhibir el crecimiento, infección o función, o destruir agentes infecciosos incluyendo, sin limitación, bacterias, virus, hongos y otros parásitos; afectar a (suprimir o potenciar) las características corporales incluyendo, sin limitación, altura, peso, color de cabello, color de ojos, piel, relación de grasa a magro u otra pigmentación de tejido, o tamaño o forma de

órgano o parte del cuerpo (tal como, por ejemplo, aumento o reducción de senos, cambio de la conformación o forma del hueso); afectar a los biorritmos o ciclos o ritmos circadianos; afectar a la fertilidad de sujetos macho o hembra; afectar al metabolismo, catabolismo, anabolismo, procesamiento, utilización, almacenamiento o eliminación de grasa, lípido, proteína, carbohidrato, vitaminas, minerales, cofactores u otros factores o componentes nutricionales de la dieta; afectar a características de comportamiento incluyendo, sin limitación, apetito, libido, estrés, cognición (incluyendo trastornos cognitivos), depresión (incluyendo trastornos depresivos) y comportamientos violentos; proporcionar efectos analgésicos u otros efectos reductores del dolor; promover la diferenciación y el crecimiento de células madre embrionarias en estirpes distintas de estirpes hematopoyéticas; actividad hormonal o endocrina; en el caso de enzimas, corregir las deficiencias de enzima y tratar enfermedades relacionadas con la deficiencia; tratamiento de trastornos hiperproliferativos (tales como, por ejemplo, psoriasis); actividad similar a inmunoglobulina (tal como, por ejemplo, la capacidad de unirse a antígenos o complemento); y la capacidad de actuar como antígeno en una composición de vacuna para crear una respuesta inmunitaria contra dicha proteína u otro material o entidad que es reactivo cruzado con dicha proteína.

15 Ejemplo 39

Identificación de proteínas que interaccionan con polipéptidos codificados por ADNc extendidos

Las proteínas que interaccionan con los polipéptidos codificados por ADNc extendidos o porciones de los mismos, tales como proteínas receptoras, pueden identificarse usando sistemas de dos híbridos tales como Matchmaker Two Hybrid System 2 (nº de catálogo K1604-1, Clontech). Como se describe en el manual que acompaña al Matchmaker Two Hybrid System 2 (nº de catálogo K1604-1, Clontech), se insertan los ADNc extendidos o porciones de los mismos en un vector de expresión de tal modo que estén en fase con ADN que codifica el dominio de unión a ADN del activador transcripcional de levadura GAL4. Se insertan ADNc de una colección de ADNc que codifica proteínas que podrían interaccionar con los polipéptidos codificados por los ADNc extendidos o porciones de los mismos en un segundo vector de expresión de tal modo que estén en fase con ADN que codifica el dominio de activación de GAL4. Se transforman los dos plásmidos de expresión en levadura y se siembra la levadura en medio de selección que selecciona la expresión de marcadores seleccionables en cada uno de los vectores de expresión, así como la expresión dependiente de GAL4 del gen HIS3. Se examina en transformantes capaces de crecer en medio que carece de histidina la expresión de lacZ dependiente de GAL4. Aquellas células que son positivas tanto en selección de histidina como en ensayo de lacZ contienen plásmidos que codifican proteínas que interaccionan con el polipéptido codificado por los ADNc extendidos o porciones de los mismos.

Como alternativa, puede usarse el sistema descrito en Lustig *et al.*, *Methods in Enzymology* 283: 83-99 (1997) para identificar moléculas que interaccionan con los polipéptidos codificados por ADNc extendidos. En dichos sistemas, se efectúan reacciones de transcripción *in vitro* en un conjunto de vectores que contienen insertos de ADNc extendido clonados cadena abajo de un promotor que controla la transcripción *in vitro*. Se introducen los conjuntos resultantes de ARNm en ovocitos de *Xenopus laevis*. Se ensaya entonces en los ovocitos la actividad deseada.

Como alternativa, pueden traducirse *in vitro* los productos de transcripción *in vitro* combinados producidos como se describe anteriormente. Puede ensayarse en los productos de traducción *in vitro* combinados una actividad deseada o la interacción con un polipéptido conocido.

Pueden encontrarse proteínas u otras moléculas que interaccionan con polipéptidos codificados por ADNc extendidos mediante una variedad de técnicas adicionales. En un método, pueden construirse columnas de afinidad que contienen el polipéptido codificado por el ADNc extendido o una porción del mismo. En algunas versiones de este método, la columna de afinidad contiene proteínas químicas en las que la proteína codificada por el ADNc extendido o una porción de la misma se fusiona con glutathione S-transferasa. Se aplica a la columna de afinidad una mezcla de proteínas celulares o un conjunto de proteínas expresadas como se describe anteriormente. Pueden aislarse entonces las proteínas que interaccionan con el polipéptido unido a la columna y analizarse en electroforesis en gel bidimensional como se describe en Ramunsen *et al.* *Electrophoresis*, 18, 588-598 (1997). Como alternativa, las proteínas retenidas en la columna de afinidad pueden purificarse mediante métodos basados en electroforesis y secuenciarse. Puede usarse el mismo método para aislar anticuerpos, para examinar productos de exhibición en fago o para examinar anticuerpos humanos de exhibición en fago.

Pueden examinarse también proteínas que interaccionan con polipéptidos codificados por ADNc extendidos o porciones de los mismos usando un Optical Biosensor como se describe en Edwards & Leatherbarrow, *Analytical Biochemistry*, 246, 1-6 (1997). La ventaja principal del método es que permite la determinación del índice de asociación entre la proteína y otras moléculas de interacción. Por tanto, es posible seleccionar específicamente moléculas de interacción con un alto o bajo índice de asociación. Típicamente, se liga una molécula diana con la superficie del sensor (mediante una matriz de carboximetildextrano) y se pone en contacto una muestra de moléculas de ensayo con las moléculas diana. La unión de una molécula de ensayo con la molécula diana causa un cambio en el índice de refracción y/o el espesor. Este cambio se detecta por el Biosensor a condición de que ocurra en el campo evanescente (que se extiende unos pocos cientos de nanómetros desde la superficie del sensor). En estos ensayos de examen, la molécula diana puede ser uno de los polipéptidos codificados por ADNc extendidos o una porción de los mismos y la muestra de ensayo puede ser una colección de proteínas extraídas de tejidos o células, un conjunto de proteínas expresadas, péptidos combinatorios y/o colecciones químicas o péptidos exhibidos en fago. Los tejidos o células de los que se extraen las proteínas de ensayo pueden proceder de cualquier especie.

En otros métodos, se inmoviliza una proteína diana y la población de ensayo es una colección de polipéptidos únicos codificados por los ADNc extendidos o porciones de los mismos.

Para estudiar la interacción de las proteínas codificadas por los ADNc extendidos o porciones de los mismos con medicamentos, puede usarse el método de microdiálisis acoplado a HPLC descrito por Wang *et al.*, *Chromatographia*, 44, 205-208 (1997) o el método de electroforesis capilar por afinidad descrito por Busch *et al.*, *J. Chromatogr.* 777: 311-328 (1997).

Puede usarse también el sistema descrito en la patente de EE.UU. n°. 5.654.150 para identificar moléculas que interaccionan con los polipéptidos codificados por los ADNc extendidos. En este sistema, se transcriben y traducen *in vitro* conjuntos de ADNc extendidos y se ensaya en los productos de reacción la interacción con un polipéptido o anticuerpo conocido.

Se apreciará por los expertos en la técnica que en las proteínas expresadas a partir de ADNc extendidos o porciones de los mismos pueden ensayarse numerosas actividades además de las enumeradas específicamente anteriormente. Por ejemplo, en las proteínas expresadas pueden evaluarse aplicaciones que implican el control y la regulación de inflamación, proliferación o metástasis tumoral, infección u otras afecciones clínicas. Además, las proteínas expresadas a partir de los ADNc extendidos o porciones de los mismos pueden ser útiles como agentes nutricionales o agentes cosméticos.

Las proteínas expresadas a partir de ADNc extendidos o porciones de los mismos pueden usarse para crear anticuerpos capaces de unirse específicamente a la proteína expresada o fragmentos de la misma como se describe en el ejemplo 40 siguiente. Los anticuerpos pueden ser capaces de unirse a una proteína completa codificada por la secuencia de SEC ID N° 45, una proteína madura codificada por la secuencia de SEC ID N° 45, o un péptido señal codificado por la secuencia de SEC ID N° 45. Como alternativa, los anticuerpos pueden ser capaces de unirse a fragmentos de las proteínas expresadas a partir de los ADNc extendidos que comprenden al menos 10 aminoácidos de las secuencias de SEC ID N° 146. Los anticuerpos pueden ser capaces de unirse a fragmentos de las proteínas expresadas a partir de los ADNc extendidos que comprenden al menos 15 aminoácidos de la secuencia de SEC ID N° 146. Los anticuerpos pueden ser capaces de unirse a fragmentos de las proteínas expresadas a partir de los ADNc extendidos que comprenden al menos 25 aminoácidos de la secuencia de SEC ID N° 146. En realizaciones adicionales, los anticuerpos pueden ser capaces de unirse a fragmentos de las proteínas expresadas a partir de los ADNc extendidos que comprenden al menos 40 aminoácidos de la secuencia de SEC ID N° 146.

Ejemplo 40

Producción de un anticuerpo de una proteína humana

Se aísla proteína o polipéptido sustancialmente puro a partir de células transfectadas o transformadas como se describe en el ejemplo 30. Se ajusta la concentración de proteína en la preparación final, por ejemplo mediante concentración en un dispositivo de filtro Amicon, al nivel de unos pocos microgramos/ml. El anticuerpo monoclonal o policlonal de la proteína puede prepararse entonces del modo siguiente:

A. Producción de anticuerpo monoclonal mediante fusión de hibridoma

Puede prepararse anticuerpo monoclonal de epítomos de cualquiera de los péptidos identificados y aislados como se ha descrito a partir de hibridomas de múrido según el método clásico de Kohfer, G. y Milstein, C, *Nature* 256: 495 (1975) o métodos derivados del mismo. Brevemente, se inocula repetidamente un ratón con unos pocos microgramos de la proteína seleccionada o péptidos derivados de la misma durante un periodo de unas pocas semanas. Se sacrifica entonces el ratón y se aíslan las células productoras de anticuerpo del bazo. Se fusionan las células de bazo mediante polietilenglicol con células de mieloma de ratón, y se destruyen las células no fusionadas en exceso mediante crecimiento del sistema en medios selectivos que comprenden aminopterina (medios HAT). Se diluyen las células fusionadas exitosamente y se disponen alícuotas de la dilución en pocillos de una placa de microvaloración, donde se continúa el crecimiento del cultivo. Se identifican clones productores de anticuerpo mediante la detección de anticuerpo en el fluido sobrenadante de los pocillos mediante procedimientos de inmunoensayo tales como ELISA, como se describe originalmente por Engvall, E., *Meth. Enzymol.* 70: 419 (1980), y métodos derivados del mismo. Los clones positivos seleccionados pueden expandirse y recolectarse para uso su producto de anticuerpo monoclonal. Se describen procedimientos detallados para la producción de anticuerpo monoclonal en Davis, L. *et al.* "Basic Methods in Molecular Biology", Elsevier, Nueva York, sección 21-2.

B. Producción de anticuerpo policlonal mediante inmunización

Puede prepararse antisuero policlonal que contiene anticuerpos de epítomos heterogéneos de una sola proteína inmunizando animales adecuados con la proteína expresada o péptidos derivados de la misma descritos anteriormente, que pueden estar no modificados o modificados para potenciar la inmunogenicidad. La producción de anticuerpo policlonal eficaz está afectada por muchos factores relacionados tanto con el antígeno como con la especie hospedadora. Por ejemplo, las moléculas pequeñas tienden a ser menos inmunogénicas que otras y pueden requerir el uso de portadores y coadyuvante. También los animales hospedadores varían en respuesta al sitio de inoculación y a la dosis, resultando antiseros de bajo título con dosis de antígeno tanto inadecuadas como excesivas. Lo más fiable parece

ser dosis pequeñas (nivel de ng) de antígeno administradas en múltiples sitios intradérmicos. Puede encontrarse un protocolo de inmunización eficaz para conejos en Vaitukaitis, J. *et al. J. Clin. Endocrinol. Metab.* 33: 988-991 (1971).

Pueden administrarse inyecciones de refuerzo a intervalos regulares, y recolectarse el antisuero cuando el título de anticuerpo, determinado semicuantitativamente, por ejemplo mediante doble inmunodifusión en agar frente a concentraciones conocidas del antígeno, empieza a caer. Véase, por ejemplo, Ouchterlony, O. *et al.*, cap. 19 en: "Handbook of Experimental Immunology" D. Wier (ed) Blackwell (1973). La concentración de meseta de anticuerpo está habitualmente en el intervalo de 0,1 a 0,2 mg/ml de suero (aproximadamente 12 μ M). Se determina la afinidad de los antisueros por el antígeno preparando curvas de unión competitiva como se describen, por ejemplo, por Fisher, O., cap. 42 en: "Manual of Clinical Immunology", 2ª ed. (Rose and Friedman, Ed.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Las preparaciones de anticuerpo preparadas de acuerdo con cualquier protocolo son útiles en inmunoensayos cuantitativos que determinan concentraciones de sustancias que portan antígenos en muestras biológicas; también se usan semicuantitativamente o cualitativamente para identificar la presencia de un antígeno en una muestra biológica. Los anticuerpos también pueden usarse en composiciones terapéuticas para destruir células que expresan la proteína o para reducir los niveles de la proteína en el cuerpo.

V. Uso de ADNc extendidos o porciones de los mismos como reactivos

Los ADNc extendidos descritos en la presente memoria pueden usarse como reactivos en procedimientos de aislamiento, ensayos de diagnóstico y procedimientos forensicos. Por ejemplo, pueden marcarse detectablemente secuencias de los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) y usarse como sondas para aislar otras secuencias capaces de hibridar con ellas. Además, pueden usarse secuencias de los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) para diseñar cebadores de PCR para usar en procedimientos de aislamiento, diagnóstico o forense.

Ejemplo 41

Preparación de cebadores de PCR y amplificación de ADN

Los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) pueden usarse para preparar cebadores de PCR para una variedad de aplicaciones, incluyendo procedimientos de aislamiento para clonación de ácidos nucleicos capaces de hibridar con dichas secuencias, técnicas de diagnóstico y técnicas forensicas. Los cebadores de PCR son de al menos 10 bases, y preferiblemente de al menos 12, 15 ó 17 bases de longitud. Más preferiblemente, los cebadores de PCR son de al menos 20-30 bases de longitud. En algunas realizaciones, los cebadores de PCR pueden ser de más de 30 bases de longitud. Se prefiere que los pares cebadores tengan aproximadamente la misma relación de G/C, de modo que las temperaturas de fusión sean aproximadamente la misma. Son familiares para los expertos en la técnica una variedad de técnicas de PCR. Para una revisión de la tecnología de PCR, véase "Molecular Cloning to Genetic Engineering" White, B.A. Ed. en "Methods in Molecular Biology" 67: Humana Press, Totowa 1997. En cada uno de estos procedimientos de PCR, se añaden cebadores de PCR en cualquier lado de las secuencias de ácido nucleico a amplificar a una muestra de ácido nucleico adecuadamente preparada junto con dNTP y una polimerasa termoestable tal como polimerasa Taq, polimerasa Pfu o polimerasa Vent. Se desnaturaliza el ácido nucleico en la muestra y se hibridan específicamente los cebadores de PCR con secuencias de ácido nucleico complementario en la muestra. Se extienden los cebadores hibridados. Después de ello, se inicia otro ciclo de desnaturalización, hibridación y extensión. Se repiten los ciclos múltiples veces para producir un fragmento amplificado que contiene la secuencia de ácido nucleico entre los sitios cebadores.

Ejemplo 42

Uso de ADNc extendidos como sondas

Pueden marcarse sondas derivadas de ADNc extendidos o porciones de los mismos (o ADN genómicos obtenibles a partir de los mismos) con marcajes detectables familiares para los expertos en la técnica, incluyendo radioisótopos y marcajes no radiactivos, proporcionando una sonda detectable. La sonda detectable puede ser monocatenaria o bicatenaria y puede prepararse usando técnicas conocidas en la técnica, incluyendo transcripción *in vitro*, traslado de mella o reacciones de cinasa. Se pone en contacto con la sonda marcada una muestra de ácido nucleico que contiene una secuencia capaz de hibridar con la sonda marcada. Si el ácido nucleico en la muestra es bicatenario, puede desnaturalizarse antes de poner en contacto con la sonda. En algunas aplicaciones, la muestra de ácido nucleico puede inmovilizarse sobre una superficie tal como una membrana de nitrocelulosa o nailon. La muestra de ácido nucleico puede comprender ácidos nucleicos obtenidos a partir de una variedad de fuentes, incluyendo ADN genómico, colecciones de ADNc, ARN o muestras de tejido.

Los procedimientos usados para detectar la presencia de ácidos nucleicos capaces de hibridar con la sonda detectable incluyen técnicas bien conocidas tales como transferencia Southern, transferencia Northern, transferencia puntual, hibridación de colonia e hibridación de placa. En algunas aplicaciones, puede clonarse el ácido nucleico capaz de hibridar con la sonda marcada en vectores tales como vectores de expresión, vectores de secuenciación o vectores de

transcripción *in vitro* para facilitar la caracterización y expresión de los ácidos nucleicos de hibridación en la muestra. Por ejemplo, dichas técnicas pueden usarse para aislar y clonar secuencias en una colección genómica o colección de ADNc que son capaces de hibridar con la sonda detectable como se describe en el ejemplo 30 anterior.

- 5 Pueden usarse cebadores de PCR como se describen en el ejemplo 41 anterior en análisis forensicos, tales como las técnicas de huella genética de ADN descritas en los ejemplos 43-47 siguientes. Dichos análisis pueden utilizar sondas detectables o cebadores basados en secuencias de los ADNc extendidos usando los 5' EST (o ADN genómicos obtenibles a partir de los mismos).

10 Ejemplo 43

Coincidencia forense mediante secuenciación de ADN

- 15 En un método ejemplar, se aíslan muestras de ADN a partir de especímenes forensicos, por ejemplo, de cabello, semen, sangre o células cutáneas mediante métodos convencionales. Se utiliza entonces un panel de cebadores de PCR basado en una serie de ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) según el ejemplo 41 para amplificar ADN de aproximadamente 100-200 bases de longitud del espécimen forense. Se obtienen las secuencias correspondientes a partir de un sujeto de ensayo. Se secuencian entonces cada uno de estos ADN de identificación usando técnicas estándar, y una sencilla comparación de base de datos determina las diferencias, si las hubiera, entre 20 las secuencias del sujeto y las de la muestra. Las diferencias estadísticamente significativas entre las secuencias de ADN del sospechoso y las de la muestra prueban concluyentemente la falta de identidad. Esta falta de identidad puede probarse, por ejemplo, con sólo una secuencia. La identidad, por otro lado, debe demostrarse con un gran número de secuencias, todas apareadas. Preferiblemente, se usan un mínimo de 50 secuencias estadísticamente idénticas de 100 bases de longitud para probar la identidad entre el sospechoso y la muestra.

25 Ejemplo 44

Identificación positiva mediante secuenciación de ADN

- 30 La técnica expuesta en el ejemplo anterior puede usarse también a mayor escala para proporcionar una identificación de huella genética única de cualquier individuo. En esta técnica, se preparan cebadores a partir de un gran número de secuencias identificadas en la presente memoria. Preferiblemente, se usan 20 a 50 cebadores diferentes. Se usan estos cebadores para obtener un número correspondiente de segmentos de ADN generados por PCR a partir del individuo en cuestión según el ejemplo 41. Se secuencian cada uno de estos segmentos de ADN usando los métodos exhibidos en el ejemplo 43. La base de datos generada mediante este procedimiento identifica únicamente al individuo del que se obtuvieron las secuencias. Puede usarse entonces el mismo panel de cebadores en cualquier momento posterior para correlacionar absolutamente tejido u otro espécimen biológico con ese individuo.

40 Ejemplo 45

Identificación forense por transferencia Southern

- Se repite el procedimiento del ejemplo 44 para obtener un panel de al menos 10 secuencias amplificadas a partir de un individuo y un espécimen. Preferiblemente, el panel contiene al menos 50 secuencias amplificadas. Más preferiblemente, el panel contiene 100 secuencias amplificadas. En algunas realizaciones, el panel contiene 200 secuencias amplificadas. Se digiere entonces este ADN generado por PCR con uno de una combinación de, preferiblemente, enzimas de restricción específicas de cuatro bases. Dichas enzimas están comercialmente disponibles y son conocidas por los expertos en la técnica. Después de la digestión, se separan por tamaño los fragmentos génicos resultantes en múltiples pocillos por duplicado en un gel de agarosa y se transfieren a nitrocelulosa usando técnicas de transferencia Southern bien conocidas por los expertos en la técnica. Para una revisión de la transferencia Southern, véase Davis *et al.* ("Basic Methods in Molecular Biology", 1986, Elsevier Press, pág. 62-65).

- Se marcan radiactiva o colorimétricamente un panel de sondas basadas en las secuencias de los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) o fragmentos de los mismos de al menos 10 bases usando métodos conocidos en la técnica, tales como traslación de mella o marcaje de extremo, y se hibridan con la transferencia Southern usando técnicas conocidas en la técnica (Davis *et al.*, *supra*). Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo). Más preferiblemente, la sonda comprende al menos 20-30 nucleótidos consecutivos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo). En algunas realizaciones, la sonda comprende más de 30 nucleótidos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo). En otras realizaciones, la sonda comprende al menos 40, al menos 50, al menos 75, al menos 100, al menos 150 o al menos 200 nucleótidos consecutivos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo).

- Preferiblemente, se usan al menos 5 a 10 de estas sondas marcadas, y más preferiblemente se usan al menos 65 aproximadamente 20 ó 30 para proporcionar un patrón único. Las bandas resultantes que aparecen por la hibridación de una muestra grande de ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) serán un identificador único. Puesto que la escisión por enzima de restricción será diferente para cada individuo, el patrón de banda en la transferencia Southern será también único. Aumentar el número de sondas de ADNc extendido proporcionará un nivel

de confianza estadísticamente mayor en la identificación, puesto que habrá un número aumentado de conjuntos de bandas usadas para identificación.

Ejemplo 46

Procedimiento de identificación por transferencia puntual

Otra técnica para identificar individuos usando las secuencias de ADNc extendido dadas a conocer en la presente memoria utiliza la técnica de hibridación por transferencia puntual.

Se aísla ADN genómico a partir de núcleos del sujeto a identificar. Se sintetizan sondas oligonucleotídicas de aproximadamente 30 pb de longitud que corresponden al menos a 10, preferiblemente a 50 secuencias de los ADNc extendidos o ADN genómicos obtenibles a partir de los mismos. Se usan las sondas para hibridar con el ADN genómico en condiciones conocidas por los expertos en la técnica. Se marcan los extremos de los oligonucleótidos con P^{32} usando polinucleótido cinasa (Pharmacia). Se crean transferencias puntuales sembrando el ADN genómico sobre nitrocelulosa o similar usando un colector de transferencia puntual a vacío (BioRad, Richmond, California). Se hierve el filtro de nitrocelulosa que contiene las secuencias genómicas o se liga por UV al filtro, se prehibrida y se hibrida con sonda marcada usando técnicas conocidas en la técnica (Davis *et al. supra*). Los fragmentos de ADN marcados con ^{32}P se hibridan secuencialmente en condiciones sucesivamente rigurosas para detectar diferencias mínimas entre la secuencia de 30 pb y el ADN. El cloruro de tetrametilamonio es útil para identificar clones que contienen pequeños números de desapareamientos nucleotídicos (Wood *et al.*, *Proc. Natl. Acad. Sci. USA* 82(6): 1585-1588 (1985)). Un único patrón de puntos distingue un individuo de otro individuo.

Pueden usarse ADNc extendidos u oligonucleótidos que contienen al menos 10 bases consecutivas como sondas en la siguiente técnica de huella genética alternativa. Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo). Más preferiblemente, la sonda comprende al menos 20-30 nucleótidos consecutivos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo). En algunas realizaciones, la sonda comprende más de 30 nucleótidos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo). En otras realizaciones, la sonda comprende al menos 40, al menos 50, al menos 75, al menos 100, al menos 150 o al menos 200 nucleótidos consecutivos del ADNc extendido (o ADN genómicos obtenibles a partir del mismo).

Preferiblemente, se usan una pluralidad de sondas que tienen secuencias de genes diferentes en la técnica de huella genética alternativa. El ejemplo 47 siguiente proporciona un procedimiento de huella genética alternativo en el que las sondas derivan de ADNc extendidos.

Ejemplo 47

Técnica de identificación de "huella genética" alternativa

Se preparan oligonucleótidos de 20 unidades a partir de un gran número, por ejemplo 50, 100 ó 200, de secuencias de ADNc extendido (o ADN genómicos obtenibles a partir del mismo) usando servicios de oligonucleótidos comercialmente disponibles tales como Genset, París, Francia. Se procesa el ADN en muestras de células del sujeto de ensayo usando técnicas bien conocidas por los expertos en la técnica. Se digiere el ácido nucleico con enzimas de restricción tales como EcoRI y XbaI. Después de la digestión, se aplican muestras a pocillos para electroforesis. El procedimiento, como es conocido en la técnica, puede modificarse para acomodar electroforesis en poliacrilamida, sin embargo en este ejemplo se cargan muestras que contienen 5 μ g de ADN en pocillos y se separan en geles de agarosa al 0,8%. Se transfieren los geles a nitrocelulosa usando técnicas de transferencia Southern estándar.

Se combinan 10 ng de cada oligonucleótido y se marcan en los extremos con P^{32} . Se prehibrida la nitrocelulosa con disolución de bloqueo y se hibrida con las sondas marcadas. Después de hibridación y lavado, se expone el filtro de nitrocelulosa a película de rayos X X-Omat AR. El patrón de hibridación resultante será único para cada individuo.

Se contempla adicionalmente en este ejemplo que el número de secuencias de sonda usado puede variar para exactitud o claridad adicionales.

Los anticuerpos generados en los ejemplos 30 y 40 anteriores pueden usarse para identificar el tipo de tejido o especie celular de la que deriva una muestra como se describe anteriormente.

Ejemplo 48

Identificación de tipos de tejido o especies celulares mediante anticuerpos específicos de tejido marcados

La identificación de tejidos específicos se consigue mediante la visualización de antígenos específicos de tejido mediante preparaciones de anticuerpo según los ejemplos 30 y 40 que están conjugados, directa o indirectamente, con un marcador detectable. Las especies de anticuerpo marcado seleccionadas se unen a su pareja de unión antigénica específica en secciones de tejido, suspensiones celulares o en extractos de proteínas solubles de una muestra de tejido, proporcionando un patrón para interpretación cualitativa o semicualitativa.

Los antisueros de estos procedimientos deben tener una potencia superior a la de la preparación nativa, y por esa razón, los anticuerpos se concentran a un nivel de mg/ml mediante el aislamiento de la fracción de gammaglobulina, por ejemplo, mediante cromatografía de intercambio iónico o mediante fraccionamiento con sulfato de amonio. Para proporcionar los antisueros más específicos, deben retirarse también los anticuerpos indeseados de la fracción de gammaglobulina, por ejemplo, mediante inmunoabsorbentes insolubles, antes de marcar los anticuerpos con el marcador. Son adecuados antisueros monoclonales o heterólogos para cualquier procedimiento.

A. Técnicas inmunohistoquímicas

Se conjugan anticuerpos purificados de alto título, preparados como se describe anteriormente, con un marcador detectable como se describe, por ejemplo, por Fudenberg, H., cap. 26 en: "Basic 503 Clinical Immunology", 3ª Ed. Lange, Los Altos, California (1980) o Rose, N. *et al.*, cap. 12 en: "Methods in Immunodiagnosis", 2ª Ed. John Wiley 503 Sons, Nueva York (1980).

Se prefiere un marcador fluorescente, fluoresceína o rodamina, pero los anticuerpos pueden marcarse también con una enzima que apoya una reacción productora de color con un sustrato, tal como peroxidasa de rábano picante. Los marcadores pueden añadirse a anticuerpo unido a tejido en una segunda etapa, como se describe a continuación. Como alternativa, los anticuerpos antitejido específicos pueden marcarse con ferritina u otras partículas electrónicamente densas, y conseguirse la localización de los complejos de antígeno-anticuerpo acoplados a ferritina mediante un microscopio electrónico. En aún otro enfoque, se marcan radiactivamente los anticuerpos, por ejemplo con ¹²⁵I, y se detectan cubriendo la preparación tratada con anticuerpo con emulsión fotográfica.

Las preparaciones para llevar a cabo los procedimientos pueden comprender anticuerpos monoclonales o policlonales de una sola proteína o péptido identificado como específico de un tipo de tejido, por ejemplo, tejido cerebral, o pueden usarse preparaciones de anticuerpo de varios antígenos específicos de tejidos antigénicamente distintos en paneles, independientemente o en mezclas, según se requiera.

Se preparan secciones de tejido y suspensiones celulares para examen inmunohistoquímico según técnicas histológicas comunes. Se montan secciones criostáticas múltiples (de aproximadamente 4 µm, no fijadas) de tejido desconocido y control conocido, y se cubre cada portaobjetos con diferentes diluciones de la preparación de anticuerpo. Las secciones de tejidos conocidos y desconocidos deben tratarse también con preparaciones para proporcionar un control positivo, un control negativo, por ejemplo sueros preinmunes, y un control de tinción no específica, por ejemplo tampón.

Se incuban las secciones tratadas en una cámara húmeda durante 30 min a temperatura ambiente, se aclaran y se lavan entonces con tampón durante 30-45 min. Se escurre el fluido en exceso y se revela el marcador.

Si el anticuerpo específico de tejido no estaba marcado en la primera incubación, puede marcarse esta vez en una segunda reacción de anticuerpo-anticuerpo, por ejemplo, añadiendo anticuerpo conjugado con fluoresceína o enzima contra la clase de inmunoglobulina de la especie productora de antisuero, por ejemplo, anticuerpo marcado con fluoresceína de IgG de ratón. Dichos sueros marcados están comercialmente disponibles.

El antígeno encontrado en los tejidos mediante el procedimiento anterior puede cuantificarse midiendo la intensidad de color o fluorescencia en la sección de tejido, y calibrando esa señal usando patrones apropiados.

B. Identificación de proteínas solubles específicas de tejido

Se lleva a cabo la visualización de proteínas específicas de tejido y la identificación de tejidos desconocidos de ese procedimiento usando los reactivos de anticuerpo marcado y la estrategia de detección descritos para la inmunohistoquímica; sin embargo, la muestra se prepara según una técnica electroforética para distribuir las proteínas extraídas del tejido en una disposición ordenada basada en el peso molecular para detección.

Se homogeneiza una muestra de tejido usando un aparato Virtis; se desestabilizan suspensiones celulares mediante homogeneización Dounce o lisis osmótica, usando detergentes en cada caso según sea necesario para desestabilizar las membranas celulares, como es la práctica en la técnica. Se retiran los componentes celulares insolubles, tales como núcleos, microsomas y fragmentos de membrana, mediante ultracentrifugación, y se concentra la fracción que contiene proteína soluble si es necesario y se reservan para análisis.

Se resuelve una muestra de la disolución de proteína soluble en especies de proteína individuales mediante electroforesis en SDS-poliacrilamida como se describe, por ejemplo, por Davis, L. *et al.*, sección 19-2 en: "Basic Methods in Molecular Biology" (P. Leder, ed), Elsevier, Nueva York (1986), usando un intervalo de cantidades de poliácridamida en un conjunto de geles para resolver el intervalo completo de pesos moleculares de proteína a detectar en la muestra. Se procesa un marcador de tamaño en paralelo con el fin de estimar el peso molecular de las proteínas constituyentes. El tamaño de muestra para análisis es un volumen conveniente de 5 a 55 µl, y que contiene de aproximadamente 1 a 100 µg de proteína. Se transfiere una alícuota de cada una de las proteínas resueltas mediante transferencia a un papel de filtro de nitrocelulosa, un proceso que mantiene el patrón de resolución. Se preparan múltiples copias. El

procedimiento, conocido como análisis de transferencia Western, está bien descrito en Davis, L. *et al.*, (anteriormente) sección 19-3. Se tiñe un conjunto de transferencias de nitrocelulosa con tinte azul de Coomassie para visualizar el conjunto completo de proteínas para comparación con las proteínas unidas a anticuerpo. Se incuban entonces los filtros de nitrocelulosa restantes con una disolución de uno o más antisueros específicos de proteínas específicas de tejido preparadas como se describe en los ejemplos 30 y 40. En este procedimiento, como en el procedimiento A anterior, se procesan controles de muestra y reactivo positivos y negativos apropiados.

En cualquier procedimiento A o B, puede unirse un marcador detectable al complejo de antígeno de tejido primario-anticuerpo primario según diversas estrategias y permutaciones de las mismas. En un enfoque directo, puede marcarse el anticuerpo específico primario; como alternativa, puede unirse el complejo no marcado por un anticuerpo anti-IgG secundario marcado. En otros enfoques, se conjuga el anticuerpo primario o secundario con una molécula de biotina que, en una etapa posterior, puede unirse a un marcador conjugado con avidina. Según aún otra estrategia, se une proteína A marcada con enzima o radiactiva, que tiene la propiedad de unirse a cualquier IgG, en una etapa final al anticuerpo primario o secundario.

La visualización de la unión de antígeno específico de tejido a niveles superiores a los observados en tejidos de control a uno o más anticuerpos específicos de tejido, preparados a partir de las secuencias génicas identificadas a partir de secuencias de ADNc extendidos, puede identificar tejidos de origen desconocido, por ejemplo, muestras forensicas o tejido tumoral diferenciado que ha metastatizado a sitios corporales extraños.

Además de sus aplicaciones en ciencia forense e identificación, los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) pueden cartografiarse en sus localizaciones cromosómicas. El ejemplo 49 siguiente describe la cartografía por híbrido de radiación (RH) de regiones cromosómicas humanas usando ADNc extendidos. El ejemplo 50 siguiente describe un procedimiento representativo para cartografiar un ADNc extendido (o un ADN genómico obtenible a partir del mismo) en su localización en un cromosoma humano. El ejemplo 51 siguiente describe la cartografía de ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) en cromosomas en metafase mediante hibridación fluorescente *in situ* (FISH).

Ejemplo 49

Cartografía por híbrido de radiación de ADNc extendidos en el genoma humano

La cartografía por híbrido de radiación (RH) es un enfoque genético de células somáticas que puede usarse para la cartografía de alta resolución del genoma humano. En este enfoque, se irradian letalmente estirpes celulares que contienen uno o más cromosomas humanos, rompiendo cada cromosoma en fragmentos cuyo tamaño depende de la dosis de radiación. Se recuperan estos fragmentos mediante fusión con células de roedor cultivadas, proporcionando subclones que contienen diferentes porciones del genoma humano. Esta técnica se describe por Benham *et al.* (*Genomics* 4: 509-517, 1989) y Cox *et al.*, (*Science* 250: 245-250, 1990). La naturaleza aleatoria e independiente de los subclones permite una cartografía eficaz de cualquier marcador del genoma humano. El ADN humano aislado a partir de un panel de 80-100 estirpes celulares proporciona un reactivo de cartografía para ordenar ADNc extendidos (o ADN genómicos obtenidos a partir de los mismos). En este enfoque, se usa la frecuencia de rotura entre marcadores para medir la distancia, permitiendo la construcción de mapas de resolución fina como se ha hecho usando EST convencionales (Schuler *et al.*, *Science* 274: 540-546, 1996).

La cartografía por RH se ha usado para generar un mapa por híbrido de radiación de alta resolución de todo el genoma del cromosoma humano 17q22-q25.3 mediante los genes de hormona de crecimiento (GH) y timidina cinasa (TK) (Foster *et al.*, *Genomics* 33: 185-192, 1996), la región que rodea al gen del síndrome de Gorlin (Obermayr *et al.*, *Eur. J. Hum. Genet.* 4: 242-245, 1996), 60 loci que cubren todo el brazo corto del cromosoma 12 (Raeymaekers *et al.*, *Genomics* 29: 170-178, 1995), la región del cromosoma humano 22 que contiene el locus de neurofibromatosis de tipo 2 (Frazer *et al.*, *Genomics* 14: 574-584, 1992) y 13 loci en el brazo largo del cromosoma 5 (Warrington *et al.*, *Genomics* 11: 701-708, 1991).

Ejemplo 50

Cartografía de ADNc extendidos en cromosomas humanos usando técnicas de PCR

Los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) pueden asignarse a cromosomas humanos usando metodologías basadas en PCR. En dichos enfoques, se diseñan pares de cebadores oligonucleotídicos a partir de la secuencia de ADNc extendido (o la secuencia de un ADN genómico obtenible a partir del mismo) para minimizar la posibilidad de amplificar por un intrón. Preferiblemente, los cebadores oligonucleotídicos son de 18-23 pb de longitud y se diseñan para amplificación por PCR. La creación de cebadores de PCR a partir de secuencias conocidas es bien conocida por los expertos en la técnica. Para una revisión de la tecnología de PCR, véase Erlich, H.A., "PCR Technology: Principles and Applications for DNA Amplification", 1992, W.H. Freeman y Co., Nueva York.

Se usan los cebadores en reacciones en cadena de la polimerasa (PCR) para amplificar moldes de ADN genómico total. Las condiciones de PCR son las siguientes: Se usan 60 ng de ADN genómico como molde para PCR con 80

ng de cada cebador oligonucleotídico, 0,6 unidades de polimerasa Taq y 1 μ Ci de un trifosfato de desoxicitidina marcado con ^{32} P. Se efectúa la PCR en un termociclador en microplaca (Techne) en las siguientes condiciones: 30 ciclos de 94°C, 1,4 min; 55°C, 2 min; y 72°C, 2 min; con una extensión final a 72°C durante 10 min. Se analizan los productos amplificados en un gel de secuenciación de poliacrilamida al 6% y se visualizan mediante autorradiografía.

Si la longitud del producto de PCR resultante es idéntica a la distancia entre los extremos de las secuencias cebadoras en el ADNc extendido del que derivan los cebadores, entonces se repite la reacción de PCR con moldes de ADN de dos paneles de híbridos de células somáticas de humano-roedor, BIOS PCRable DNA (BIOS Corporation) y NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel n° 1 (NIGMS, Camden, NJ).

Se usa la PCR para examinar en una serie de estirpes celulares de híbridos de células somáticas que contienen conjuntos definidos de cromosomas humanos la presencia de un ADNc extendido dado (o ADN genómico obtenible a partir del mismo). Se aísla ADN de los híbridos somáticos y se usa como moldes de inicio para reacciones de PCR usando los pares cebadores de los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos). Sólo aquellos híbridos de células somáticas que contienen el gen humano correspondiente al ADNc extendido (o ADN genómico obtenible a partir del mismo) proporcionarán un fragmento amplificado. Se asignan los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) a un cromosoma mediante análisis del patrón de segregación de productos de PCR de moldes de ADN híbridos somáticos. El cromosoma humano individual presente en todos los híbridos celulares que da lugar a un fragmento amplificado es el cromosoma que contiene el ADNc extendido (o ADN genómico obtenible a partir del mismo). Para una revisión de las técnicas y análisis de resultados de experimentos de cartografía genética de células somáticas véase Ledbetter *et al.*, *Genomics* 6: 475-181 (1990).

Como alternativa, los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) pueden cartografiarse en cromosomas individuales usando FISH como se describe en el ejemplo 51 siguiente.

Ejemplo 51

Cartografía de 5' EST extendidos en cromosomas usando hibridación fluorescente in situ

La hibridación fluorescente *in situ* permite cartografiar el ADNc extendido (o ADN genómico obtenible a partir del mismo) en una localización particular de un cromosoma dado. Los cromosomas a usar para técnicas de hibridación fluorescente *in situ* pueden obtenerse a partir de una variedad de fuentes, incluyendo cultivos celulares, tejidos o sangre completa.

En una realización preferida, se obtiene la localización cromosómica de un ADNc extendido (o ADN genómico obtenible a partir del mismo) mediante FISH como se describe por Cherif *et al.* (*Proc. Natl. Acad. Sci. U.S.A.* 87: 6639-6643, 1990). Se preparan cromosomas en metafase a partir de donantes de células sanguíneas estimulados con fitohemaglutinina (PHA). Se cultivan los linfocitos estimulados con PHA de hombres sanos durante 72 h en medio RPMI-1640. Para sincronización, se añade metotrexato (10 μ M) durante 17 h, seguido de la adición de 5-bromodesoxiuridina (5-BudR, 0,1 mM) durante 6 h. Se añade colcemida (1 μ g/ml) durante los últimos 15 min antes de recolectar las células. Se recogen las células, se lavan con RPMI, se incuban con una disolución isotónica de KCl (75 mM) a 37°C durante 15 min y se fijan con tres cambios de metanol:ácido acético (3:1). Se deja caer gota a gota la suspensión celular en un portaobjetos de vidrio y se seca al aire. Se marca el ADNc extendido (o ADN genómico obtenible a partir del mismo) con biotina-16 dUTP mediante traslado de mella según las instrucciones del fabricante (Bethesda Research Laboratories, Bethesda, MD), se purifica usando una columna Sephadex G-50 (Pharmacia, Upssala, Suecia) y se precipita. Justo antes de la hibridación, se disuelve el sedimento de ADN en tampón de hibridación (50% de formamida, 2 X SSC, 10% de sulfato de dextrano, 1 mg/ml de ADN de esperma de salmón sonificado, pH 7) y se desnaturaliza la sonda a 70°C durante 5-10 min.

Se tratan los portaobjetos mantenidos a -20°C durante 1 h a 37°C con ARNasa (100 μ g/ml), se aclaran tres veces con 2 X SSC y se deshidratan en una serie de etanol. Se desnaturalizan las preparaciones de cromosoma en 70% de formamida, 2 X SSC durante 2 min a 70°C y se deshidratan entonces a 4°C. Se tratan los portaobjetos con proteinasa K (10 μ g/100 ml en Tris-HCl 20 mM, CaCl₂ 2 mM) a 37°C durante 8 min y se deshidratan. Se dispone la mezcla de hibridación que contiene la sonda sobre el portaobjetos, se cubre con un cubreobjetos, se sella con pegamento de caucho y se incuba durante una noche en una cámara húmeda a 37°C. Después de la hibridación y los lavados posthibridación, se detecta la sonda biotinilada mediante avidina-FITC y se amplifica con capas adicionales de antiavidina de cabra biotinilada y avidina-FITC. Para la localización cromosómica, se obtienen bandas R fluorescentes como se describe anteriormente (Cherif *et al.*, *supra.*). Se observan los portaobjetos bajo un microscopio de fluorescencia LEICA (DMRXA). Se contratiñen los cromosomas con yoduro de propidio y la señal fluorescente de la sonda aparece como dos puntos verdes amarillentos simétricos en ambas cromátidas del cromosoma de banda R fluorescente (rojo). Por tanto, puede localizarse un ADNc extendido particular (o ADN genómico obtenible a partir del mismo) en una banda R citogenética particular de un cromosoma dado.

Una vez se han asignado los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) a cromosomas particulares usando las técnicas descritas en los ejemplos 49-51 anteriores, pueden utilizarse para construir un mapa de alta resolución de los cromosomas en los que están localizados o para identificar los cromosomas en una muestra.

Ejemplo 52

Uso de ADNc extendidos para construir o expandir mapas cromosómicos

La cartografía cromosómica implica asignar una secuencia única dada a un cromosoma particular como se describe anteriormente. Una vez se ha cartografiado la secuencia única en un cromosoma dado, se ordena respecto a otras secuencias únicas localizadas en el mismo cromosoma. Un enfoque de la cartografía cromosómica utiliza una serie de cromosomas artificiales de levadura (YAC) que portan varios miles de insertos largos derivados de los cromosomas del organismo del que se obtienen los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos). Se describe este enfoque en Ramaiah Nagaraja *et al. Genome Research* 7: 210-222, marzo de 1997. Brevemente, en este enfoque se rompe cada cromosoma en trozos superpuestos que se insertan en el vector YAC. Se examinan los insertos YAC usando PCR u otros métodos para determinar si incluyen el ADNc extendido (o ADN genómico obtenible a partir del mismo) cuya posición va a determinarse. Una vez se ha encontrado un inserto que incluye el ADNc extendido (o ADN genómico obtenible a partir del mismo), puede analizarse el inserto mediante PCR u otros métodos para determinar si el inserto contiene también otras secuencias conocidas por estar en el cromosoma o en la región de la que derivaba el ADNc extendido (o ADN genómico obtenible a partir del mismo). Este proceso puede repetirse para cada inserto en la colección de YAC para determinar la localización de cada uno de los ADNc extendidos (o ADN genómico obtenible a partir de los mismos) entre sí y con otros marcadores cromosómicos conocidos. De este modo, puede obtenerse un mapa de alta resolución de la distribución de numerosos marcadores únicos a lo largo de cada uno de los cromosomas del organismo.

Como se describe en el ejemplo 53 siguiente, pueden usarse también ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) para identificar genes asociados a un fenotipo particular, tal como una enfermedad hereditaria o respuesta a medicamento.

Ejemplo 53

Identificación de genes asociados a enfermedades hereditarias o respuesta a medicamento

Este ejemplo ilustra un enfoque útil para la asociación de ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) con características fenotípicas particulares. En este ejemplo, se usa un ADNc extendido particular (o ADN genómico obtenible a partir del mismo) como sonda de ensayo para asociar ese ADNc extendido (o ADN genómico obtenible a partir del mismo) con una característica fenotípica particular.

Se cartografían los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) en una localización particular en un cromosoma humano usando técnicas tales como las descritas en los ejemplos 49 y 50 u otras técnicas conocidas en la técnica. "A search of Mendelian Inheritance in Man" (V. McKusick, "Mendelian Inheritance in Man") (disponible en línea en la Johns Hopkins University Welch Medical Library) revela que la región del cromosoma humano que contiene el ADNc extendido (o ADN genómico obtenible a partir del mismo) es una región muy rica en genes que contiene varios genes conocidos y varias enfermedades o fenotipos para los que no se han identificado genes. El gen correspondiente a este ADNc extendido (o ADN genómico obtenible a partir del mismo) se convierte por tanto en un candidato inmediato para cada una de estas enfermedades genéticas.

Se aíslan células de pacientes con estas enfermedades o fenotipos y se expanden en cultivo. Se usan cebadores de PCR del ADNc extendido (o ADN genómico obtenible a partir del mismo) para examinar ADN genómico, ARNm o ADNc obtenidos a partir de los pacientes. Los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) que no están amplificados en los pacientes pueden asociarse positivamente con una enfermedad particular mediante análisis adicional. Como alternativa, el análisis de PCR puede proporcionar fragmentos de diferentes longitudes cuando las muestras derivan de un individuo que tiene el fenotipo asociado a la enfermedad que cuando la muestra deriva de un individuo sano, indicando que el gen que contiene el ADNc extendido puede ser responsable de la enfermedad genética.

VI. Uso de ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) para construir vectores

Los presentes ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) pueden usarse también para construir vectores de secreción capaces de dirigir la secreción de las proteínas codificadas por genes insertados en los vectores. Dichos vectores de secreción pueden facilitar la purificación o enriquecimiento de las proteínas codificadas por genes insertados en las mismas reduciendo el número de proteínas de fondo a partir de las que la proteína deseada debe purificarse o enriquecerse. Se describen vectores de secreción ejemplares en el ejemplo 54 siguiente.

Ejemplo 54

Construcción de vectores de secreción

Los vectores de secreción descritos en la presente memoria incluyen un promotor capaz de dirigir la expresión génica en la célula, tejido u organismo de interés hospedador. Dichos promotores incluyen el promotor del virus de sarcoma de Rous, el promotor SV40, el promotor de citomegalovirus humano y otros promotores familiares para los expertos en la técnica.

Se liga operativamente una secuencia señal de un ADNc extendido (o ADN genómico obtenible a partir del mismo), tal como las secuencias señal de SEC ID N° 45 como se define en la Tabla IV anterior, de tal modo que el ARNm transcrito a partir del promotor dirigirá la traducción del péptido señal. La célula, tejido u organismo hospedador puede ser cualquier célula, tejido u organismo que reconozca el péptido señal codificado por la secuencia señal en el ADNc extendido (o ADN genómico obtenible a partir del mismo). Los hospedadores adecuados incluyen células, tejidos u organismos de mamíferos, células, tejidos u organismos de aves o células, tejidos u organismos de insectos, o levadura.

Además, el vector de secreción contiene sitios de clonación para insertar genes que codifican las proteínas que se van a secretar. Los sitios de clonación facilitan la clonación del gen de inserto en fase con la secuencia señal de tal modo que se expresa a partir del ARNm transcrito a partir del promotor una proteína de fusión en la que el péptido señal está fusionado con la proteína codificada por el gen insertado. El péptido señal dirige la secreción extracelular de la proteína de fusión.

El vector de secreción puede ser ADN o ARN, y puede integrarse en el cromosoma del hospedador, mantenerse establemente como replicón extracromosómico en el hospedador, ser un cromosoma artificial o estar presente transitoriamente en el hospedador. Son conocidas por los expertos en la técnica muchas cadenas principales de ácido nucleico adecuadas para uso como vectores de secreción, incluyendo vectores retrovíricos, vectores SV40, vectores de papilomavirus bovino, plásmidos de integración en levadura, plásmidos artificiales de levadura, cromosomas artificiales humanos, vectores de elemento P, vectores de baculovirus o plásmidos bacterianos capaces de introducirse transitoriamente en el hospedador.

El vector de secreción puede contener también una señal de poli A de tal modo que la señal de poli A esté localizada cadena abajo del gen insertado en el vector de secreción.

Después de insertar en el vector de secreción el gen que codifica la proteína para la que se desea la secreción, se introduce el vector de secreción en una célula, tejido u organismo hospedador usando precipitación con fosfato de calcio, DEAE-dextrano, electroporación, transfección mediada por liposoma, partículas víricas o como ADN desnudo. Se purifica o enriquece entonces la proteína codificada por el gen insertado a partir del sobrenadante usando técnicas convencionales tales como precipitación con sulfato de amonio, inmunoprecipitación, inmunocromatografía, cromatografía de exclusión por tamaño, cromatografía de intercambio iónico y HPLC. Como alternativa, la proteína secretada puede estar en un estado suficientemente enriquecido o puro en el sobrenadante o medios de crecimiento del hospedador para permitir que se use con su fin pretendido sin enriquecimiento adicional.

Las secuencias señal pueden insertarse también en vectores diseñados para terapia génica. En dichos vectores, se liga operativamente la secuencia señal a un promotor de tal modo que el ARNm transcrito a partir del promotor codifique el péptido señal. Se localiza un sitio de clonación cadena abajo de la secuencia señal de tal modo que pueda insertarse fácilmente en el vector un gen que codifica una proteína cuya secreción es deseada y fusionarse con la secuencia señal. Se introduce el vector en una célula hospedador apropiada. Se secreta extracelularmente la proteína expresada a partir del promotor, produciendo así un efecto terapéutico.

Los ADNc extendidos o 5' EST pueden usarse también para clonar secuencias localizadas cadena arriba de los ADNc extendidos o 5' EST que son capaces de regular la expresión génica, incluyendo secuencias promotoras, secuencias potenciadoras y otras secuencias cadena arriba que influyen en los niveles de transcripción o traducción. Una vez identificadas y clonadas, pueden usarse estas secuencias reguladoras cadena arriba en vectores de expresión diseñados para dirigir la expresión de un gen insertado de la manera deseada desde el punto de vista espacial, temporal, de desarrollo o cuantitativo. El ejemplo 55 describe un método para clonar secuencias cadena arriba de los ADNc extendidos o 5' EST.

Ejemplo 55

Uso de ADNc extendidos o 5' EST para clonar secuencias cadena arriba de ADN genómico

Pueden usarse secuencias de ADNc extendidos o 5' EST para aislar los promotores de los correspondientes genes usando técnicas de desplazamiento por cromosoma. En una técnica de desplazamiento por cromosoma que utiliza el kit GenomeWalker™ disponible en Clontech, se digieren cinco muestras de ADN genómico completo cada una con una enzima de restricción diferente que tiene un sitio de reconocimiento de 6 bases y deja un extremo romo. Después de la digestión, se ligan adaptadores oligonucleotídicos a cada extremo de los fragmentos de ADN genómico resultantes.

Para cada una de las cinco colecciones de ADN genómico, se efectúa una primera reacción de PCR según las instrucciones del fabricante usando un cebador adaptador externo proporcionado en el kit y un cebador específico de gen externo. El cebador específico de gen debe seleccionarse para ser específico del ADNc extendido o 5' EST de interés y debe tener una temperatura de fusión, longitud y localización en el ADNc extendido o EST que sea consistente con su uso en reacciones de PCR. Cada primera reacción de PCR contiene 5 ng de ADN genómico, 5 µl de tampón de reacción 10X Tth, 0,2 mM de cada dNTP, 0,2 µM de cada cebador adaptador externo y cebador específico de gen externo, 1,1 mM de Mg(OAc)₂, y 1 µl de mezcla de polimerasa Tth 50X en un volumen total de 50 µl. El ciclo de reacción para la primera reacción de PCR es el siguiente: 1 min a 94°C/2 s a 94°C, 3 min a 72°C (7 ciclos)/2 s a 94°C, 3 min a 67°C (32 ciclos)/5 min a 67°C.

ES 2 330 903 T3

Se diluye el producto de la primera reacción de PCR y se usa como molde para una segunda reacción de PCR según las instrucciones del fabricante, usando un par de cebadores anidados que se localizan internamente en el amplicón resultante de la primera reacción de PCR. Por ejemplo, pueden diluirse 180 veces 5 µl del producto de reacción de la primera mezcla de reacción de PCR. Se preparan las reacciones en un volumen de 50 µl que tiene una composición idéntica a la de la primera reacción de PCR, excepto porque se usan los cebadores anidados. El primer cebador anidado es específico del adaptador, y se proporciona con el kit GenomeWalker™. El segundo cebador anidado es específico del ADNc extendido o 5' EST particular para el que el promotor se va a clonar y debe tener una temperatura de fusión, longitud y localización en el ADNc extendido o 5' EST que sea consistente con su uso en reacciones de PCR. Los parámetros de reacción de la segunda reacción de PCR son los siguientes: 1 min a 94°C/2 s a 94°C, 3 min a 72°C (6 ciclos)/2 s a 94°C, 3 min a 67°C (25 ciclos)/5 min a 67°C.

Se purifica el producto de la segunda reacción de PCR, se clona y se secuenciar usando técnicas estándar. Como alternativa, pueden construirse dos o más colecciones de ADN genómico usando dos o más enzimas de restricción. Se clona el ADN genómico digerido en vectores que pueden convertirse en ADN monocatenario, circular o lineal. Se hibrida un oligonucleótido biotinilado que comprende al menos 15 nucleótidos de la secuencia de ADNc extendido o 5' EST con el ADN monocatenario. Se aíslan híbridos entre el oligonucleótido biotinilado y el ADN monocatenario que contiene la secuencia de ADNc extendido o EST como se describe en el ejemplo 29 anterior. Después de ello, se libera el ADN monocatenario que contiene la secuencia de ADNc extendido o EST de las perlas y se convierte en ADN bicatenario usando un cebador específico del ADNc extendido o secuencia de 5' EST o un cebador correspondiente a una secuencia incluida en el vector de clonación. Se transforma el ADN bicatenario resultante en bacterias. Se identifican ADN que contienen las secuencias de 5' EST o ADNc extendido mediante PCR de colonia o hibridación de colonia.

Una vez se han clonado y secuenciado las secuencias genómicas cadena arriba como se describe anteriormente, pueden identificarse los promotores y sitios de inicio de la transcripción esperados en las secuencias cadena arriba comparando las secuencias cadena arriba de los ADNc extendidos o 5' EST con bases de datos que contienen sitios de inicio de la transcripción conocidos, sitios de unión a factor de transcripción o secuencias promotoras.

Además, pueden identificarse los promotores en las secuencias cadena arriba usando vectores informadores de promotor como se describen en el ejemplo 56.

Ejemplo 56

Identificación de promotores en secuencias cadena arriba clonadas

Se clonan las secuencias genómicas cadena arriba de los ADNc extendidos o 5' EST en un vector informador de promotor adecuado tal como los vectores informadores de promotor pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, o pEGFP-1, disponibles en Clontech. Brevemente, cada uno de estos vectores informadores de promotor incluye múltiples sitios de clonación situados cadena arriba de un gen informador que codifica una proteína fácilmente ensayable tal como fosfatasa alcalina secretada, β-galactosidasa o proteína fluorescente verde. Se insertan las secuencias cadena arriba de los ADNc extendidos o 5' EST en los sitios de clonación cadena arriba del gen informador en ambas orientaciones y se introducen en una célula hospedadora apropiada. Se ensaya el nivel de proteína informadora 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 al vector de control indica la presencia de un promotor en el inserto. Si es necesario, pueden clonarse las secuencias cadena arriba en vectores que contienen un potenciador para aumentar los niveles de transcripción de secuencias promotoras débiles. Un nivel de expresión significativo superior al observado con el vector que carece de un inserto indica que está presente una secuencia promotora en la secuencia cadena arriba insertada.

Pueden elegirse células hospedadoras apropiadas para los vectores informadores de promotor basándose en los resultados de la determinación anteriormente descrita de patrones de expresión de los ADNc extendidos y 5' EST. Por ejemplo, si el análisis del patrón de expresión indica que el ARNm correspondiente a un ADNc o 5' EST particular se expresa en fibroblastos, puede introducirse el vector informador de promotor en una estirpe celular de fibroblastos humanos.

Las secuencias promotoras en el ADN genómico cadena arriba pueden definirse adicionalmente construyendo delecciones anidadas en el ADN cadena arriba usando técnicas convencionales tales como digestión con exonucleasa III. Los fragmentos de delección resultantes pueden insertarse en el vector informador de promotor para determinar si la delección ha reducido o anulado la actividad promotora. De este modo, pueden definirse los límites de los promotores. Si se desea, pueden identificarse sitios reguladores individuales potenciales en el promotor usando mutagénesis dirigida a sitio o de exploración de ligador para anular los sitios de unión a factor de transcripción potenciales en el promotor individualmente o en combinación. Los efectos de estas mutaciones sobre los niveles de transcripción pueden determinarse insertando las mutaciones en los sitios de clonación en los vectores informadores de promotor.

Ejemplo 57

Clonación e identificación de promotores

5 Usando el método descrito en el ejemplo 55 anterior con 5' EST, se obtuvieron secuencias cadena arriba de varios genes. Usando los pares cebadores GGG AAG ATG GAG ATA GTA TTG CCT G (SEC ID N° 29) y CTG CCA TGT ACA TGA TAG AGA GAT TC (SEC ID N° 30), se obtuvo el promotor que tiene la referencia interna P13H2 (SEC ID N° 31).

10 Usando los pares cebadores GTA CCA GGGG ACT GTG ACC ATT GC (SEC ID N° 32) y CTG TGA CCA TTG CTC CCA AGA GAG (SEC ID N° 33), se obtuvo el promotor que tiene la referencia interna P15B4 (SEC ID N° 34).

15 Usando los pares cebadores CTG GGA TGG AAG GCA CGG TA (SEC ID N° 35) y GAG ACC ACA CAG CTA GAC AA (SEC ID N° 36), se obtuvo el promotor que tiene la referencia interna P29B6 (SEC ID N° 37).

20 La Figura 8 proporciona una descripción esquemática de los promotores aislados y el modo en que se ensamblan con los correspondientes marcadores 5'. Se examinó en las secuencias cadena arriba la presencia de motivos similares a sitios de unión a factor de transcripción o sitios de inicio de la transcripción conocidos usando el programa informático MatInspector versión 2.0, agosto de 1996.

25 La Figura 9 describe los sitios de unión a factor de transcripción presentes en cada uno de estos promotores. La columna marcada como matriz proporciona el nombre de la matriz MatInspector usada. La columna marcada como posición proporciona la posición 5' del sitio promotor. La numeración de la secuencia empieza por el sitio de transcripción como se determina apareando la secuencia genómica con la secuencia de 5' EST. La columna marcada como "orientación" indica la cadena de ADN en la que se encuentra el sitio, siendo la cadena + la cadena de codificación como se determina apareando la secuencia genómica con la secuencia de 5' EST. La columna marcada como "puntuación" proporciona la puntuación MatInspector encontrada para este sitio. La columna marcada como "longitud" proporciona la longitud del sitio en nucleótidos. La columna marcada como "secuencia" proporciona la secuencia del sitio encontrado.

30 Los promotores y otras secuencias reguladoras localizadas cadena arriba de los ADNc extendidos o 5' EST pueden usarse para diseñar vectores de expresión capaces de dirigir la expresión de un gen insertado de la manera deseada desde el punto de vista espacial, temporal, de desarrollo o cuantitativo. Puede seleccionarse un promotor capaz de dirigir los patrones espaciales, temporales, de desarrollo y cuantitativos deseados usando los resultados del análisis de expresión descrito en el ejemplo 26 anterior. Por ejemplo, si se desea un promotor que confiere un alto nivel de expresión a músculo, puede usarse en el vector de expresión la secuencia promotora cadena arriba de un ADNc extendido o 5' EST derivado de un ARNm que se expresa a alto nivel en músculo, como se determina mediante el método del ejemplo 26.

40 Preferiblemente, se dispone el promotor deseado cerca de múltiples sitios de restricción para facilitar la clonación del inserto deseado cadena abajo del promotor, de tal modo que el promotor sea capaz de controlar la expresión del gen insertado. El promotor puede insertarse en cadenas principales de ácido nucleico convencionales diseñadas para replicación extracromosómica, integración en los cromosomas hospedadores o expresión transitoria. Las cadenas principales adecuadas para los presentes vectores de expresión incluyen cadenas principales retrovíricas, cadenas principales de episomas eucarióticos tales como SV40 o papilomavirus bovino, cadenas principales de episomas bacterianos o cromosomas artificiales.

45 Preferiblemente, los vectores de expresión incluyen también una señal de poli A cadena abajo de los múltiples sitios de restricción para dirigir la poliadenilación de ARNm transcrito a partir del gen insertado en el vector de expresión.

50 Después de la identificación de las secuencias promotoras usando los procedimientos de los ejemplos 55-57, pueden identificarse proteínas que interaccionan con el promotor como se describe en el ejemplo 58 siguiente.

55 Ejemplo 58

Identificación de proteínas que interaccionan con secuencias promotoras, secuencias reguladoras cadena arriba o ARNm

60 Las secuencias en la región promotora que es probable que se unan a factores de transcripción pueden identificarse mediante homología con sitios de unión a factor de transcripción conocidos o mediante mutagénesis convencional o análisis de delección de plásmidos informadores que contienen la secuencia promotora. Por ejemplo, pueden prepararse delecciones en un plásmido informador que contiene la secuencia promotora de interés ligada operativamente a un gen informador ensayable. Se transfectan los plásmidos informadores que portan diversas delecciones en la región promotora en una célula hospedadora apropiada y se evalúan los efectos de las delecciones sobre los niveles de expresión. Los sitios de unión a factor de transcripción en las regiones en que las delecciones reducen los niveles de expresión pueden localizarse adicionalmente usando mutagénesis dirigida a sitio, análisis de exploración de ligador u otras técnicas familiares para los expertos en la técnica. Los ácidos nucleicos que codifican proteínas que interaccionan con

secuencias en el promotor pueden identificarse usando sistemas de un híbrido tales como los descritos en el manual que acompaña al kit Matchmaker One-Hybrid System disponible en Clontech (nº de catálogo K1603-1). Brevemente, se usa el sistema Matchmaker One-hybrid del modo siguiente. Se clona la secuencia diana para la que se desea identificar proteínas de unión cadena arriba de un gen marcador seleccionable y se integra en el genoma de levadura. Preferiblemente, se insertan múltiples copias en serie de las secuencias diana en el plásmido informador.

Se transforma una colección que comprende fusiones entre ADNc para evaluar la capacidad de unirse al promotor y se transforma el dominio de activación de un factor de transcripción de levadura, tal como GAL4, en la cepa de levadura que contiene la secuencia informadora integrada. Se siembra la levadura en medios selectivos para seleccionar células que expresan el marcador seleccionable ligado a la secuencia promotora. Las colonias que crecen en los medios selectivos contienen genes que codifican proteínas que se unen a la secuencia diana. Los insertos en los genes que codifican las proteínas de fusión se caracterizan adicionalmente mediante secuenciación. Además, los insertos pueden insertarse en vectores de expresión o vectores de transcripción *in vitro*. Puede confirmarse la unión de los polipéptidos codificados por los insertos al ADN promotor mediante técnicas familiares para los expertos en la técnica, tales como análisis de desplazamiento en gel o análisis de protección de ADNasa.

VII. Uso de ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) en terapia génica

Los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) pueden ser también útiles en estrategias de terapia génica, incluyendo estrategias no codificantes y de triple hélice como se describen en los ejemplos 57 y 58 siguientes. En enfoques no codificantes, se hibridan intracelularmente secuencias de ácido nucleico complementarias de un ARNm con el ARNm, bloqueando así la expresión de la proteína codificada por el ARNm. Las secuencias no codificantes pueden evitar la expresión génica mediante una variedad de mecanismos. Por ejemplo, las secuencias no codificantes pueden inhibir la capacidad de los ribosomas de traducir el ARNm. Como alternativa, las secuencias no codificantes pueden bloquear el transporte del ARNm del núcleo al citoplasma, limitando así la cantidad de ARNm disponible para traducción. Otro mecanismo mediante el que las secuencias no codificantes pueden inhibir la expresión génica es interfiriendo con el corte y empalme de ARNm. En aún otra estrategia, el ácido nucleico no codificante puede incorporarse a una ribozima capaz de escindir específicamente el ARNm diana.

Ejemplo 59

Preparación y uso de oligonucleótidos no codificantes

Las moléculas de ácido nucleico no codificante para usar en terapia génica pueden ser secuencias de ADN o ARN. Pueden comprender una secuencia complementaria de la secuencia del ADNc extendido (o ADN genómico obtenible a partir del mismo). Los ácidos nucleicos no codificantes deben tener una longitud y temperatura de fusión suficientes para permitir la formación de un dúplex intracelular que tenga suficiente estabilidad para inhibir la expresión del ARNm en el dúplex. Se dan a conocer estrategias para diseñar ácidos nucleicos no codificantes adecuados para uso en terapia génica en Green *et al.*, *Ann. Rev. Biochem.* 55: 569-597 (1986) e Izant y Weintraub, *Cell* 36: 1007-1015 (1984).

En algunas estrategias, se obtienen moléculas no codificantes a partir de una secuencia nucleotídica que codifica una proteína invirtiendo la orientación de la región de codificación con respecto a un promotor, para transcribir la cadena opuesta a la que se transcribe normalmente en la célula. Las moléculas no codificantes pueden transcribirse usando sistemas de transcripción *in vitro* tales como aquellos que emplean polimerasa T7 o SP6 para generar el transcrito. Otro enfoque implica la transcripción de los ácidos nucleicos no codificantes *in vivo* mediante ligamiento operativo de ADN que contiene la secuencia no codificante con un promotor en un vector de expresión.

Como alternativa, pueden sintetizarse *in vitro* oligonucleótidos que son complementarios de la cadena transcrita normalmente en la célula. Por tanto, los ácidos nucleicos no codificantes son complementarios del correspondiente ARNm y son capaces de hibridar con el ARNm para crear un dúplex. En las mismas realizaciones, las secuencias no codificantes pueden contener cadenas principales de fosfato de azúcar modificadas para aumentar la estabilidad y hacerlas menos sensibles a la actividad ARNasa. Se describen ejemplos de modificaciones adecuadas para uso en estrategias no codificantes por Rossi *et al.*, *Pharmacol. Ther.* 50(2): 245-254, (1991).

Pueden usarse diversos tipos de oligonucleótidos no codificantes complementarios de la secuencia del ADNc extendido (o ADN genómico obtenible a partir del mismo). En un ejemplo preferido, se usan los oligonucleótidos no codificantes estables y semiestables descritos en la solicitud internacional nº PCT WO94/23026. En estas moléculas, se acoplan el extremo 3' o ambos extremos 3' y 5' por enlace de hidrógeno intramolecular entre pares de bases complementarias. Estas moléculas son capaces de resistir mejor ataques de exonucleasa y exhiben una estabilidad aumentada comparadas con oligonucleótidos no codificantes convencionales.

Es otro ejemplo preferido, los oligodesoxinucleótidos no codificantes contra herpesvirus simplex de tipos 1 y 2 descritos en la solicitud internacional nº WO 95/04141.

En aún otro ejemplo preferido, se usan los oligonucleótidos no codificantes reticulados covalentemente descritos en la solicitud internacional nº WO 96/31523. Estos oligonucleótidos bi- o monocatenarios comprenden uno o más, respectivamente, ligadores cruzados covalentes inter- o intraoligonucleotídicos, en los que el ligador consiste en un

enlace amida entre un grupo amina primario de una cadena y un grupo carboxilo de la otra cadena o de la misma cadena, respectivamente, estando directamente sustituido el grupo amina primario en la posición 2' de la cadena de anillo monosacárido, y estando portado el grupo carboxilo por un grupo espaciador alifático sustituido en un nucleótido o análogo de nucleótido de la otra cadena o la misma cadena, respectivamente.

Pueden usarse también los oligodesoxinucleótidos y oligonucleótidos no codificantes dados a conocer en la solicitud internacional nº WO 92/18522. Estas moléculas son estables a la degradación y contienen al menos una secuencia de reconocimiento de control de la transcripción que se une a proteínas de control y son eficaces como cebos para las mismas. Estas moléculas pueden contener estructuras de "horquilla", estructuras de "mancuerna", estructuras de "mancuerna modificada", estructuras de cebo "reticulado" y estructuras de "bucle".

En otro ejemplo preferido, se usan los oligonucleótidos bicatenarios cíclicos descritos en la solicitud de patente europea nº 0.572.jemplo 62287 A2. Estas "mancuernas" oligonucleotídicas ligadas contienen el sitio de unión de un factor de transcripción e inhiben la expresión del gen bajo el control del factor de transcripción secuestrando el factor.

Se contempla también el uso de oligonucleótidos no codificantes cerrados dados a conocer en la solicitud internacional nº WO 92/19732. Debido a que estas moléculas no tienen extremos libres, son más resistentes a la degradación por exonucleasas que los oligonucleótidos convencionales. Estos oligonucleótidos pueden ser multifuncionales, interaccionando con varias regiones que no están adyacentes al ARNm diana.

Puede determinarse el nivel apropiado de ácidos nucleicos no codificantes requerido para inhibir la expresión génica usando análisis de expresión *in vitro*. La molécula no codificante puede introducirse en las células mediante difusión, inyección o transfección usando procedimientos conocidos en la técnica, por ejemplo, los ácidos nucleicos no codificantes pueden introducirse en el cuerpo como oligonucleótido suelto o desnudo, oligonucleótido encapsulado en lípido, secuencia oligonucleotídica encapsulada en proteína vírica o como oligonucleótido ligado operativamente a un promotor contenido en un vector de expresión. El vector de expresión puede ser cualquiera de una variedad de vectores de expresión conocidos en la técnica, incluyendo vectores retrovíricos o víricos, vectores capaces de replicación extracromosómica o vectores de integración. Los vectores pueden ser de ADN o ARN.

Se introducen moléculas no codificantes en muestras de células a una serie de concentraciones diferentes, preferiblemente entre 1×10^{-10} M y 1×10^{-4} M. Una vez se identifica la concentración mínima que puede controlar adecuadamente la expresión génica, se traduce la dosis optimizada a una dosificación adecuada para uso *in vivo*. Por ejemplo, una concentración inhibidora en cultivo de 1×10^{-7} se traduce en una dosis de aproximadamente 0,6 mg/kg de peso corporal. Pueden ser posibles niveles de oligonucleótido cercanos a 100 mg/kg de peso corporal o mayores después de ensayar la toxicidad del oligonucleótido en animales de laboratorio. Se contempla adicionalmente que se retiren células del vertebrado, se traten con el oligonucleótido no codificante y se reintroduzcan en el vertebrado.

Se contempla adicionalmente que la secuencia oligonucleotídica no codificante se incorpora a una secuencia de ribozima para posibilitar que la no codificante se una específicamente y escinda su ARNm diana. Para aplicaciones técnicas de ribozima y oligonucleótidos no codificantes, véase Rossi *et al*, *supra*.

En una aplicación preferida, se identifica en primer lugar el polipéptido codificado por el gen, de modo que pueda monitorizarse la eficacia de la inhibición no codificante de la traducción usando técnicas que incluyen, pero sin limitación, ensayos mediados por anticuerpo tales como RIA y ELISA, ensayos funcionales o radiomarcaje.

Los ADNc extendidos descritos en la presente memoria (o ADN genómicos obtenibles a partir de los mismos) pueden usarse también en enfoques de terapia génica basados en la formación de triple hélice intracelular. Se usan oligonucleótidos de triple hélice para inhibir la transcripción de un genoma. Son particularmente útiles para estudiar alteraciones de la actividad celular, ya que están asociados a un gen particular. Pueden usarse los ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos), o más preferiblemente una porción de esas secuencias, para inhibir la expresión génica en individuos que tienen enfermedades asociadas a la expresión de un gen particular. De forma similar, puede usarse una porción del ADNc extendido (o ADN genómico obtenible a partir del mismo) para estudiar el efecto de inhibir la transcripción de un gen particular en una célula. Tradicionalmente, las secuencias de homopurina se han considerado las más útiles para estrategias de triple hélice. Sin embargo, las secuencias de homopirimidina pueden inhibir también la expresión génica. Dichos oligonucleótidos de homopirimidina se unen al surco mayor en secuencias de homopurina:homopirimidina. Por tanto, se contemplan ambos tipos de secuencias de ADNc extendido o del gen correspondiente al ADNc extendido.

Ejemplo 60

Preparación y uso de sondas de triple hélice

Se exploran las secuencias de ADNc extendidos (o ADN genómicos obtenibles a partir de los mismos) para identificar tramos de homopirimidina u homopurina de 10 a 20 unidades que podrían usarse en estrategias basadas en triple hélice para inhibir la expresión génica. Después de la identificación de los tramos candidatos de homopirimidina u homopurina, se evalúa su eficacia en la inhibición de la expresión génica introduciendo cantidades variables de oligonucleótidos que contienen las secuencias candidatas en células de cultivo de tejido que expresan normalmente el gen diana. Los oligonucleótidos pueden prepararse en un sintetizador de oligonucleótidos o pueden adquirirse

comercialmente en una compañía especializada en síntesis de oligonucleótidos a medida, tal como GENSET, París, Francia.

Los oligonucleótidos pueden introducirse en las células usando una variedad de métodos conocidos por los expertos en la técnica, incluyendo pero sin limitación precipitación con fosfato de calcio, DEAE-dextrano, electroporación, transfección mediada por liposoma o captación nativa.

Se monitorizan en las células tratadas la función celular alterada o expresión génica reducida usando técnicas tales como transferencia Northern, ensayos de protección de ARNasa o estrategias basadas en PCR para monitorizar los niveles de transcripción del gen diana en células que se han tratado con el oligonucleótido. Se predicen las funciones celulares a monitorizar basándose en las homologías del gen diana correspondiente al ADNc extendido del que derivaba el oligonucleótido con secuencias génicas conocidas que se han asociado a una función particular. Las funciones celulares pueden predecirse también basándose en la presencia de fisiologías anormales en células derivadas de individuos con una enfermedad hereditaria particular, particularmente cuando el ADNc extendido está asociado a la enfermedad, usando técnicas descritas en el ejemplo 53.

Pueden introducirse entonces *in vivo* los oligonucleótidos que son eficaces en la inhibición de la expresión génica en células de cultivo de tejido usando las técnicas descritas anteriormente y en el ejemplo 59 a una dosificación calculada basándose en los resultados *in vitro*, como se describe en el ejemplo 59.

En algunos ejemplos, pueden reemplazarse los anómeros naturales (beta) de los oligonucleótidos por anómeros alfa para volver al oligonucleótido más resistente a nucleasas. Además, puede unirse un agente intercalante tal como bromuro de etidio, o similar, al extremo 3' del oligonucleótido alfa para estabilizar la triple hélice. Para información sobre la generación de oligonucleótidos adecuados para la formación de triple hélice, véase Griffin *et al.* (*Science* 245, 967-971 (1989)).

Ejemplo 61

Uso de ADNc extendidos para expresar una proteína codificada en un organismo hospedador

Los ADNc extendidos descritos en la presente memoria pueden usarse también para expresar una proteína codificada en un organismo hospedador para producir un efecto beneficioso; en dichos procedimientos, la proteína codificada puede expresarse transitoriamente en el organismo hospedador o expresarse establemente en el organismo hospedador. La proteína modificada puede tener cualquiera de las actividades descritas anteriormente. La proteína codificada puede ser una proteína de la que el organismo hospedador carece, o como alternativa, la proteína codificada puede aumentar los niveles existentes de proteína en el organismo hospedador.

Se introduce en el organismo hospedador un ADNc extendido completo que codifica el péptido señal y la proteína madura, o un ADNc extendido que codifica sólo la proteína madura. El ADNc extendido puede introducirse en el organismo hospedador usando una variedad de técnicas conocidas por los expertos en la técnica. Por ejemplo, el ADNc extendido puede inyectarse en el organismo hospedador como ADN desnudo de tal modo que la proteína codificada se exprese en el organismo hospedador, produciendo así un efecto beneficioso.

Como alternativa, el ADNc extendido puede clonarse en un vector de expresión cadena abajo de un promotor que es activo en el organismo hospedador. El vector de expresión puede ser cualquiera de los vectores de expresión diseñados para uso en terapia génica, incluyendo vectores víricos o retrovíricos.

El vector de expresión puede introducirse directamente en el organismo hospedador de tal modo que la proteína codificada se exprese en el organismo hospedador para producir un efecto beneficioso. En otro enfoque, el vector de expresión puede introducirse en células *in vitro*. Se seleccionan después de ello las células que contienen el vector de expresión y se introducen en el organismo hospedador, donde expresan la proteína codificada para producir un efecto beneficioso.

Ejemplo 62

Uso de péptidos señal codificados por 5' EST o secuencias obtenidas a partir de los mismos para importar proteínas en células

Puede usarse también la región hidrófoba nuclear corta (h) de péptidos señal codificados por 5' EST o ADNc extendidos derivados de los 5' EST descritos en la presente memoria como portador para importar un péptido o una proteína de interés, denominada carga, en células de cultivo de tejido (Lin *et al.*, *J. Biol. Chem.*, 270: 14225-14258 (1995); Du *et al.* *J. Peptide Res.* 51: 235-243 (1998); Rojas *et al.*, *Nature Biotech.*, 16: 370-375 (1998)).

Cuando se van a translocar péptidos de tamaño limitado permeables en células (aproximadamente hasta 25 aminoácidos) a través de la membrana celular, puede usarse la síntesis química para añadir la región h al extremo C o al extremo N del péptido de interés de carga. Como alternativa, cuando van a importarse péptidos más largos o proteí-

nas en células, los ácidos nucleicos pueden modificarse por ingeniería genética, usando técnicas familiares para los expertos en la técnica, para ligar la secuencia de ADNc extendido que codifica la región h con el extremo 5' o 3' de una secuencia de ADN que codifica un polipéptido de carga. Dichos ácidos nucleicos modificados por ingeniería genética se traducen entonces *in vitro* o *in vivo* después de transfección en células apropiadas, usando técnicas convencionales para producir el polipéptido permeable en célula resultante. Se incuban entonces simplemente las células hospedadoras adecuadas con el polipéptido permeable en célula, que se transloca a través de la membrana.

Este método puede aplicarse para estudiar diversas funciones intracelulares y procesos celulares. Por ejemplo, se ha usado para sondear dominios funcionalmente relevantes de proteínas intracelulares y para examinar interacciones proteína-proteína implicadas en rutas de transducción de señal (Lin *et al. supra*; Lin *et al. J. Biol. Chem.*, 271: 5305-5308 (1996); Rojas *et al. J. Biol. Chem.*, 271: 27456-27461 (1996); Liu *et al. Proc. Natl. Acad. Sci. USA*, 93: 11819-11824, (1996). Rojas *et al. Bioch. Biophys. Res. Common.*, 234: 675-680 (1997)).

Pueden usarse dichas técnicas en terapia celular para importar proteínas que producen efectos terapéuticos. Por ejemplo, pueden tratarse células aisladas a partir de un paciente con proteínas terapéuticas importadas y reintroducirse entonces en el organismo hospedador.

Como alternativa, podría usarse la región h de señal descrita en la presente memoria en combinación con una señal de localización nuclear para suministrar ácidos nucleicos al núcleo celular. Dichos oligonucleótidos pueden ser oligonucleótidos no codificantes u oligonucleótidos diseñados para formar triples hélices, como se describe en los ejemplos 59 y 60 respectivamente, para inhibir el procesamiento y maduración de un ARN celular diana.

Ejemplo 63

Reensamblado y resecuenciación de clones

Se secuenciaron doblemente clones de ADNc completos obtenidos mediante el procedimiento descrito en el ejemplo 27. Se ensamblaron estas secuencias y se reanalizaron entonces las secuencias consenso resultantes. Se reasignaron marcos abiertos de lectura esencialmente según el mismo proceso que se describe en el ejemplo 27.

Después de este proceso de reanálisis, se revelaron unas pocas anomalías. Las secuencias presentadas en las SEC ID N° 47, 73, 79, 89, 91, 96, 126, 128, 134 y 139 aparentemente es improbable que sean ADNc completos genuinos. Estos clones carecen de un codón de terminación y por tanto son más probablemente secuencias de ADNc truncadas en 3'. De forma similar, las secuencias presentadas en las SEC ID N° 45, 50, 54, 57, 73, 74, 89, 92, 95, 98, 126, 129, 130, 131 y 139 pueden no ser también ADNc completos genuinos basándose en estudios de homología con secuencias proteicas existentes. Aunque ambas de estas secuencias codifican una metionina de inicio potencial, cada una podría representar un ADNc truncado en 5'.

Además, se encontró que la SEC ID N° 115 era un transcrito cortado y empalmado alternativamente y se corrigieron las identidades de algunas de las bases en la SEC ID N° 131.

Finalmente, después de la reasignación de marcos de lectura abiertos para los clones, se eligieron nuevos marcos de lectura abiertos en algunos casos. Por ejemplo, en el caso de las SEC ID N° 41, 47, 50, 52, 54-56, 58, 59, 61, 74, 75, 79, 84, 89, 91, 92, 96, 98, 103, 105, 106, 126, 129, 131 y 133, no se predecía ya que los nuevos marcos de lectura abiertos contuvieran un péptido señal.

Como se discute anteriormente, la Tabla IV proporciona la localización de la secuencia de codificación completa de SEC ID N° 45 (concretamente, los nucleótidos que codifican tanto el péptido señal como la proteína madura, enumerados bajo el encabezado localización de FCS en la Tabla IV), las localizaciones de los nucleótidos en la SEC ID N° 45 que codifican los péptidos señal (enumerados bajo el encabezado localización de SigPep en la Tabla IV), las localizaciones de los nucleótidos en la SEC ID N° 45 que codifican las proteínas maduras generadas por escisión de los péptidos señal (enumerados bajo el encabezado localización de polipéptido maduro en la Tabla IV), las localizaciones en la SEC ID N° 45 de codones de terminación (enumerados bajo el encabezado localización de codón de terminación en la Tabla IV), las localizaciones en la SEC ID N° 45 de señales de poli A (enumeradas bajo el encabezado localización de señal de poli A en la Tabla IV) y las localizaciones de sitios de poli A (enumerados bajo el encabezado localización de sitios de poli A en la Tabla IV).

Como se discute anteriormente, la Tabla V enumera los números de identificación de secuencia del polipéptido de SEC ID N° 146, las localizaciones de los residuos aminoácidos de SEC ID N° 146 en el polipéptido completo (segunda columna), las localizaciones de los residuos aminoácidos de la SEC ID N° 146 en los péptidos señal (tercera columna) y las localizaciones de los residuos aminoácidos de la SEC ID N° 146 en el polipéptido maduro creado escindiendo el péptido señal del polipéptido completo (cuarta columna). En la Tabla V, y en el listado de secuencias adjunto, el primer aminoácido de la proteína madura resultante de la escisión del péptido señal se designa como aminoácido número 1 y el primer aminoácido del péptido señal se designa con el número negativo apropiado, según las regulaciones que gobiernan los listados de secuencias.

Ejemplo 64

Ensayos funcionales de secuencias proteicas predichas

Después de la doble secuenciación, se ensamblaron nuevos cóntigos para cada uno de los ADNc extendidos, y se comparó cada uno con secuencias conocidas disponibles en el momento de la presentación. Estas secuencias se originan en las siguientes bases de datos: Genbank (versión 108 y versiones diarias hasta el 15 de octubre de 1998), Genseq (versión 32), PIR (versión 53) y SwissProt (versión 35). Las proteínas predichas que se aparean con proteínas conocidas se clasificaron adicionalmente en 3 categorías dependiendo del nivel de homología.

La primera categoría contiene proteínas que exhiben más de un 70% de residuos aminoácidos idénticos en toda la longitud de la proteína apareada. Son claramente homólogos cercanos que tienen lo más probablemente la misma función, o una función muy similar, que la proteína apareada.

La segunda categoría contiene proteínas que exhiben homologías más remotas (40 a 70% en toda la proteína), indicando que la proteína puede tener funciones similares a las de la proteína homóloga.

La tercera categoría contiene proteínas que exhiben homología (90 a 100%) con un dominio de una proteína conocida, indicando que la proteína apareada y la proteína de la invención comparten rasgos similares.

Debe observarse que, en la numeración de los aminoácidos en las secuencias proteicas discutidas en las Figuras 1 a 15, la primera metionina encontrada se designa como aminoácido número 1. En el listado de secuencias adjunto, el primer aminoácido de la proteína madura resultante de la escisión del péptido señal se designa como aminoácido número 1 y el primer aminoácido del péptido señal se designa con el número negativo apropiado, según las regulaciones que gobiernan los listados de secuencias.

Además, se exploró en todas las secuencias aminoácidas corregidas (SEC ID N° 141-241 y 378-513) la presencia de distintivos y motivos de proteína conocidos. Se efectuó esta búsqueda en la base de datos Prosite 15.0 usando el software ProScan del paquete GCG.

Proteína de SEC ID N° 146

La proteína de SEC ID N° 146 codificada por el ADNc extendido de SEC ID N° 45 muestra homología con la apolipoproteína L humana (número de acceso a Genbank AF019225). La proteína apareada es una lipoproteína de alta densidad secretada asociada a lipoproteínas que contienen apoA-I que desempeñan un papel clave en el transporte inverso de colesterol.

Tomados en conjuntos, estos datos sugieren que la proteína de SEC ID N° 146 puede desempeñar un papel en el metabolismo lipídico.

Por tanto, esta proteína puede ser útil en el diagnóstico y/o el tratamiento de varios tipos de trastornos incluyendo, pero sin limitación, hiperlipidemia, hipercolesterolemia, aterosclerosis, trastornos cardiovasculares tales como enfermedad cardíaca coronaria y trastornos neurodegenerativos tales como enfermedad de Alzheimer o demencia.

Aunque esta invención se ha descrito en términos de ciertas realizaciones preferidas, están también dentro del alcance de esta invención otras realizaciones que serán evidentes para los expertos en la técnica a la vista de la divulgación de la presente memoria. En consecuencia, se pretende que el alcance de la invención esté definido sólo por referencia a las reivindicaciones adjuntas.

Como se discute anteriormente, los ADNc extendidos descritos en la presente memoria o porciones de los mismos pueden usarse para diversos fines. Los polinucleótidos pueden usarse para expresar proteína recombinante para análisis, caracterización o uso terapéutico; como marcadores para tejidos en los que se expresa preferiblemente la proteína correspondiente (constitutivamente o en una etapa particular de diferenciación o desarrollo de tejido o en estados patológicos); como marcadores de peso molecular en geles Southern; como marcadores cromosómicos o marcadores (cuando están marcados) para identificar cromosomas o para cartografiar posiciones génicas relacionadas; para comparar con secuencias de ADN endógeno en pacientes para identificar trastornos genéticos potenciales; como sondas para hibridar y por tanto descubrir secuencias de ADN relacionadas novedosas; como fuente de información para derivar cebadores de PCR para huella genética; para seleccionar y preparar oligómeros para unión a un "chip genético" u otro soporte, incluyendo para el examen de patrones de expresión; para crear anticuerpos antiproteína usando técnicas de inmunización de ADN; y como antígeno para crear anticuerpos anti-ADN o desencadenar otra respuesta inmunitaria. Cuando el polinucleótido codifica una proteína que se une o se une potencialmente a otra proteína (tal como, por ejemplo, en una interacción de receptor-ligando), el polinucleótido puede usarse también en ensayos de interacción por retención (tales como, por ejemplo, el descrito en Gyuris *et al.*, *Cell* 75: 791-803 (1993)) para identificar polinucleótidos que codifican la otra proteína con la que ocurre la unión o para identificar inhibidores de la interacción de unión.

Las proteínas o polipéptidos proporcionados en la presente memoria pueden usarse de forma similar para determinar la actividad biológica, incluyendo un panel de múltiples proteínas para examen de alto rendimiento; para crear anticuerpos o para desencadenar otra respuesta inmunitaria; como reactivo (incluyendo reactivo marcado) en ensayos diseñados para determinar cuantitativamente los niveles de proteína (o su receptor) en fluidos biológicos; como marcadores para tejidos en los que se expresa preferiblemente la proteína correspondiente (constitutivamente o en una etapa particular de diferenciación o desarrollo de tejido o en un estado patológico); y, por supuesto, para aislar receptores o ligandos correlativos. Cuando la proteína se une o se une potencialmente a otra proteína (tal como, por ejemplo, en una interacción de receptor-ligando), la proteína puede usarse para identificar la otra proteína con la que ocurre la unión o para identificar inhibidores de la interacción de unión). Las proteínas implicadas en estas interacciones de unión pueden usarse también para explorar inhibidores peptídicos o de molécula pequeña o agonistas de la interacción de unión.

Puede desarrollarse cualquiera o todas estas herramientas de investigación hasta pureza de reactivo o formato de kit para comercialización como productos de investigación.

Los métodos para efectuar los usos enumerados anteriormente son bien conocidos por los expertos en la técnica. Las referencias que dan a conocer dichos métodos incluyen, sin limitación, "Molecular Cloning; A Laboratory Manual", 2ª ed., Cole Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch y T. Maniatis ed., 1989 y "Methods in Enzymology; Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. y A. R. Kimmel ed., 1987.

Pueden usarse también polinucleótidos y proteínas como fuentes o suplementos nutricionales. Dichos usos incluyen, sin limitación, el uso de un suplemento de proteína o aminoácido, el uso como fuente de carbono, el uso como fuente de nitrógeno y el uso como fuente de carbohidrato. En dichos casos, la proteína o polinucleótido de la invención puede añadirse a la alimentación de un organismo particular o puede administrarse como una preparación sólida o líquida separada, tal como en forma de polvo, píldoras, disoluciones, suspensiones o cápsulas. En el caso de microorganismos, la proteína o polipéptido puede añadirse al medio en o sobre el que se cultiva el microorganismo.

TABLA I

SEC ID Nº en la presente solicitud	Secuencia de divulgación de solicitud provisional	SEC ID Nº en la solicitud provisional
45	Solicitud de patente de EE.UU. provisional nº de serie 60/074.121, presentada el 9 de febrero de 1998	41

ES 2 330 903 T3

TABLA II

Parámetros usados para cada etapa del análisis de EST

Características de búsqueda				Características de selección	
Etapas	Programa	Cadena	Parámetros	Identidad (%)	Longitud (pb)
Miscelánea	Blastn	ambas	S = 61X=16	90	17
ARNt	Fasta	ambas	-	80	60
ARNr	Blastn	ambas	S = 108;	80	40
ARNmt	Blastn	ambas	S = 108;	80	40
Procariótico	Blastn	ambas	S = 144;	90	40
Fúngico	Blastn	ambas	S = 144;	90	40
Alu	Fasta	ambas	-	70	40
L1	Blastn	ambas	S = 72;	70	40
Repeticiones	Blastn	ambas	S = 72;	70	40
Promotores	Blastn	superior	S = 54X=16	90	15 ⊥
Vertebrado	Fasta*	ambas	S = 108;	90	30
EST	Blastn	ambas	S=108X=16	90	30
Proteínas	Blastx _η	superior	E=0,001	-	-

* usar explorador de la base de datos "Quick Fast"

⊥ alineamiento limitado adicionalmente a empezar más cerca del 10 pb de EST/extremo 5'

η usar la matriz de sustitución BLOSUM62

ES 2 330 903 T3

TABLA III

Parámetros usados para cada etapa del análisis de ADNc extendido

Etapa	Características de búsqueda		Características de selección			
	Programa	Cadena	Parámetros	Identidad (%)	Longitud (pb)	Comentarios
Miscelánea ^{\$}	Fasta	ambas	-	90	IS	
ARNt ^{\$}	Fasta	ambas	-	80	90	
ARNr ^{\$}	Blastn	ambas	S = 108;	80	40	
ARNmt ^{\$}	Blastn	ambas	S = 108;	80	40	
Procariótico ^{\$}	Blastn	ambas	S = 144;	90	40	
Fúngico*	Blastn	ambas	S = 144;	90	40	
Alu*	Blastn	ambas	S = 72;	70	40	máx 5 apareamientos, enmascaramiento
L1 ^{\$}	Blastn	ambas	S = 72;	70	40	máx 5 apareamientos, enmascaramiento
Repeticiones ^{\$}	Blastn	ambas	S = 72;	70	40	Enmascaramiento
Poli A	BLAST2N	superior	W=6, S=10, E=1000	90	8	En los últimos 20 nucleótidos
Señal de poliadenilación		superior	AATAAA que permite 1 desapareamiento			En los 50 nucleótidos precedentes al extremo 5' de poli A
Vertebrado*	BLASTN después FASTA	ambas	-	90 después 70	30	primero BLASTN y después FASTA en secuencias apareadas
EST*	BLAST2N	ambas	-	90	30	
Geneseq	Blastn	ambas	W=8, B=10	90	30	
ORF	BLASTP	superior	W=8, B=10	-	-	En proteínas ORF, máx. 10 apareamientos
Proteínas*	Blastx	superior	E-o.orji	70	30	

^{\$} etapas comunes al análisis de EST y que usan los mismos algoritmos y parámetros

* etapas usadas también en el análisis de EST, pero con diferentes algoritmos y/o parámetros

ES 2 330 903 T3

TABLA IV

ID	Localización de FCS	Localización de SigPep	Localización de polipéptido maduro	Localización de codón de terminación	Localización de señal de poli A	Localización de sitio de poli A
45	276 a 1040	276 a 485	486 a 1040	1041	-	2024 a 2036

TABLA V

ID	Localización de polipéptido completo	Localización de péptido señal	Localización de polipéptido maduro
146	-70 a 185	-70 a -1	1 a 185

TABLA VI

ID	Ref. de colección	Nombre de depósito
45	ATCC nº 98920	SignalTag 67-90

TABLA VII

Número de referencia interna	SEC ID NO:	Tipo de secuencia
26-27-3D7CLCM	45	ADN
26-27-3D7 CL0 1	146	PRT

REIVINDICACIONES

- 5 1. Un polipéptido purificado o aislado que comprende una secuencia que tiene al menos un 80% de identidad con el polipéptido de SEC ID N° 146.
2. El polipéptido de la reivindicación 1, en el que dicho polipéptido comprende la SEC ID N° 146.
3. El polipéptido de la reivindicación 2, en el que dicho polipéptido consiste en la SEC ID N° 146.
- 10 4. Un ácido nucleico purificado o aislado que codifica el polipéptido de la reivindicación 2 ó 3.
5. El ácido nucleico de la reivindicación 4, en el que dicho ácido nucleico comprende la SEC ID N° 45.
- 15 6. Un vector de expresión que comprende el ácido nucleico de la reivindicación 4 ó 5.
7. Una célula hospedadora que contiene un ácido nucleico recombinante de la reivindicación 4 ó 5.
- 20 8. Un método de preparación del polipéptido de cualquiera de las reivindicaciones 1 a 3, que comprende las etapas de:
- a. obtener un ADNc que codifica dicho polipéptido;
- b. insertar dicho ADNc en un vector de expresión de tal modo que dicho ADNc esté ligado operativamente a un promotor; y
- 25 c. introducir dicho vector de expresión en una célula hospedadora, con lo que dicha célula hospedadora produce dicho polipéptido.
- 30 9. El método de la reivindicación 8, que comprende adicionalmente la etapa de aislar dicho polipéptido.
10. El método de la reivindicación 9, que comprende adicionalmente la etapa de purificar dicho polipéptido usando técnicas de inmunocromatografía.
- 35 11. Un anticuerpo purificado o aislado capaz de unirse específicamente al polipéptido de una cualquiera de las reivindicaciones 1 a 3.

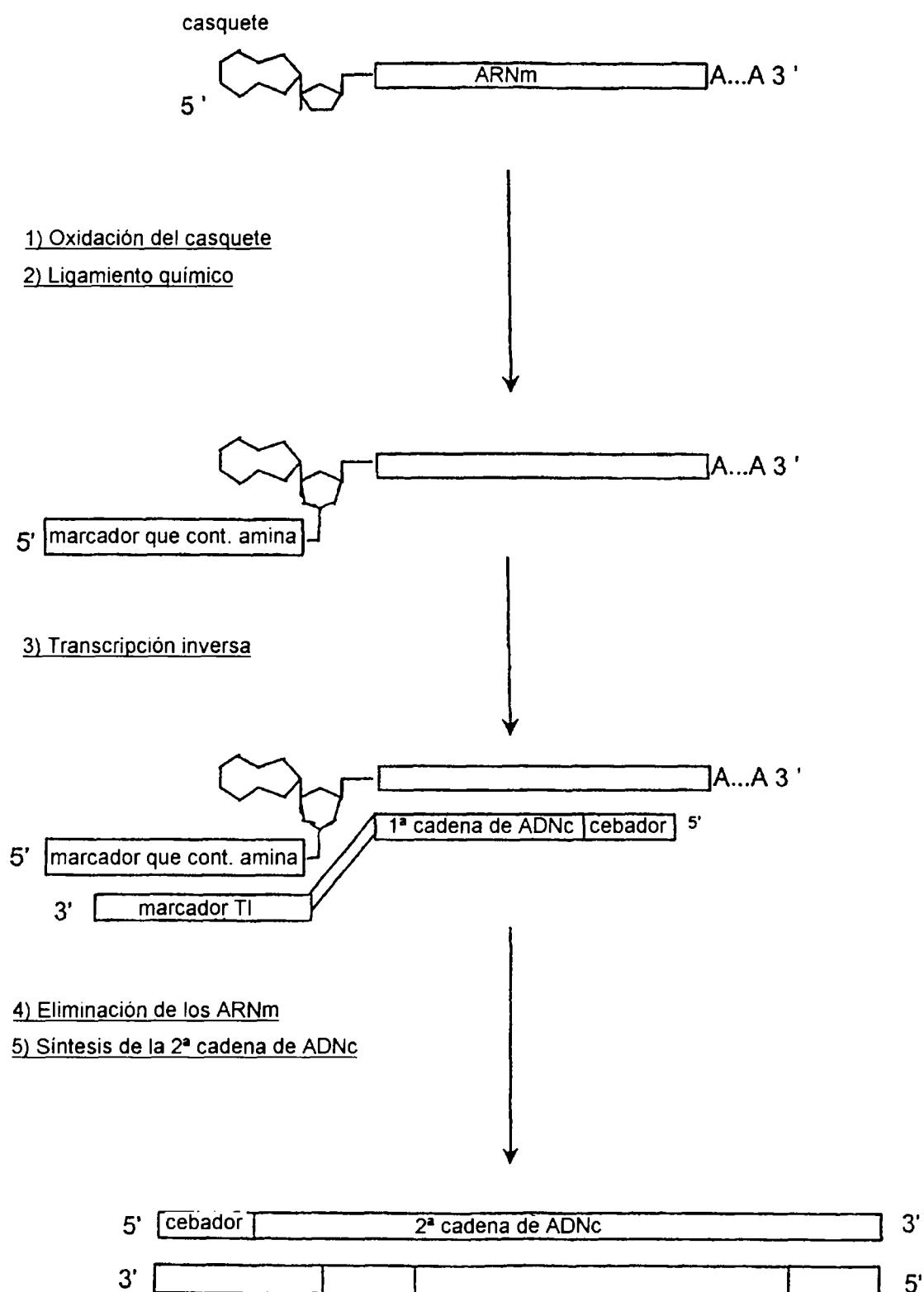


Figura 1

Puntuación de péptido señal mínima	Índice de falsos positivos	Índice de falsos negativos	Proba (0,1)	Proba (0,2)
3,5	0,121	0,036	0,467	0,664
4	0,096	0,06	0,519	0,708
4,5	0,078	0,079	0,565	0,745
5	0,062	0,098	0,615	0,782
5,5	0,05	0,127	0,659	0,813
6	0,04	0,163	0,694	0,836
6,5	0,033	0,202	0,725	0,855
7	0,025	0,248	0,763	0,878
7,5	0,021	0,304	0,78	0,889
8	0,015	0,368	0,816	0,909
8,5	0,012	0,418	0,836	0,92
9	0,009	0,512	0,856	0,93
9,5	0,007	0,581	0,863	0,934
10	0,006	0,679	0,835	0,919

Figura 2

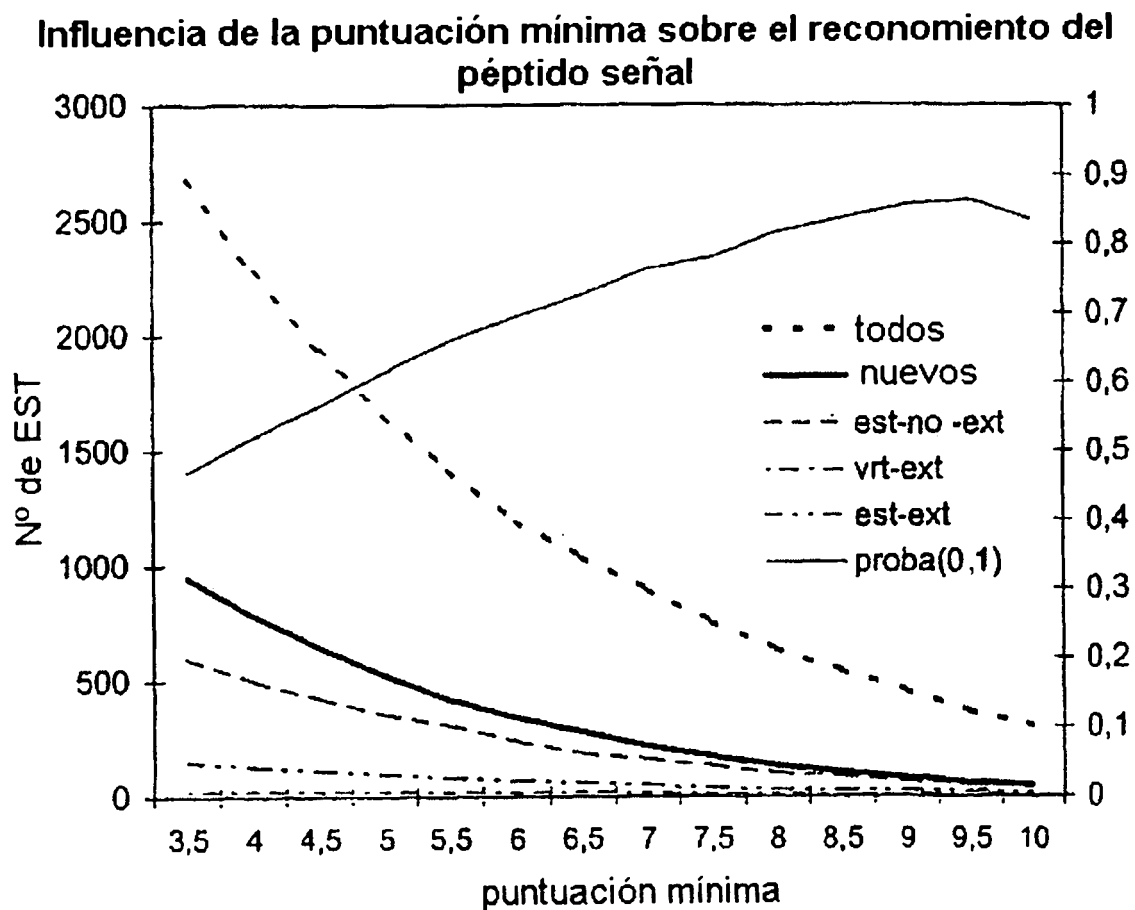


Figura 3

Puntuación de péptido señal mínima	Todos los EST	EST nuevos	EST que se aparean con EST públicos más cerca de 40 pb desde el inicio	EST que extienden ARNm conocidos más de 40 pb	EST que extienden ARNm públicos más de 40 pb
3,5	2674	947	599	23	150
4	2278	784	499	23	126
4,5	1943	647	425	22	112
5	1657	523	353	21	96
5,5	1417	419	307	19	80
6	1190	340	238	18	68
6,5	1035	280	186	18	60
7	893	219	161	15	48
7,5	753	173	132	12	36
8	636	133	101	11	29
8,5	543	104	83	8	26
9	456	81	63	6	24
9,5	364	57	48	6	18
10	303	47	35	6	15

Figura 4

Tejido	Todos los EST	EST nuevos	EST que se aparean con EST públicos más cerca de 40 pb desde el inicio	EST que extienden ARNm conocidos más de 40 pb	EST que extienden ARNm públicos más de 40 pb
Cerebro	329	131	75	3	24
Próstata cancerosa	124	40	37	1	6
Cerebelo	17	9	1	0	6
Colon	21	11	4	0	0
Músculo distrófico	41	18	8	0	1
Cerebro fetal	70	37	16	0	1
Riñón fetal	227	116	46	1	19
Hígado fetal	13	7	2	0	0
Corazón	30	15	7	0	1
Próstata hipertrófica	86	23	22	2	2
Riñón	10	7	3	0	0
Intestino grueso	21	8	4	0	1
Hígado	23	9	6	0	0
Pulmón	24	12	4	0	1
Pulmón (células)	57	38	6	0	4
Ganglios linfáticos	163	60	23	2	12
Linfocitos	23	6	4	0	2
Músculo	33	16	6	0	4
Próstata normal	181	61	45	7	11
Ovario	90	57	12	1	2
Páncreas	48	11	6	0	1
Placenta	24	5	1	0	0
Próstata	34	16	4	0	2
Bazo	56	28	10	0	1
Sustancia negra	108	47	27	1	6
G. suprarrenales	15	3	3	1	0
Testículos	131	68	25	1	8
Tiroides	17	8	2	0	2
Cordón umbilical	55	17	12	1	3
Útero	28	15	3	0	2
Tejido no específico	568	48	177	2	28
Total	2677	947	601	23	150

Figura 5

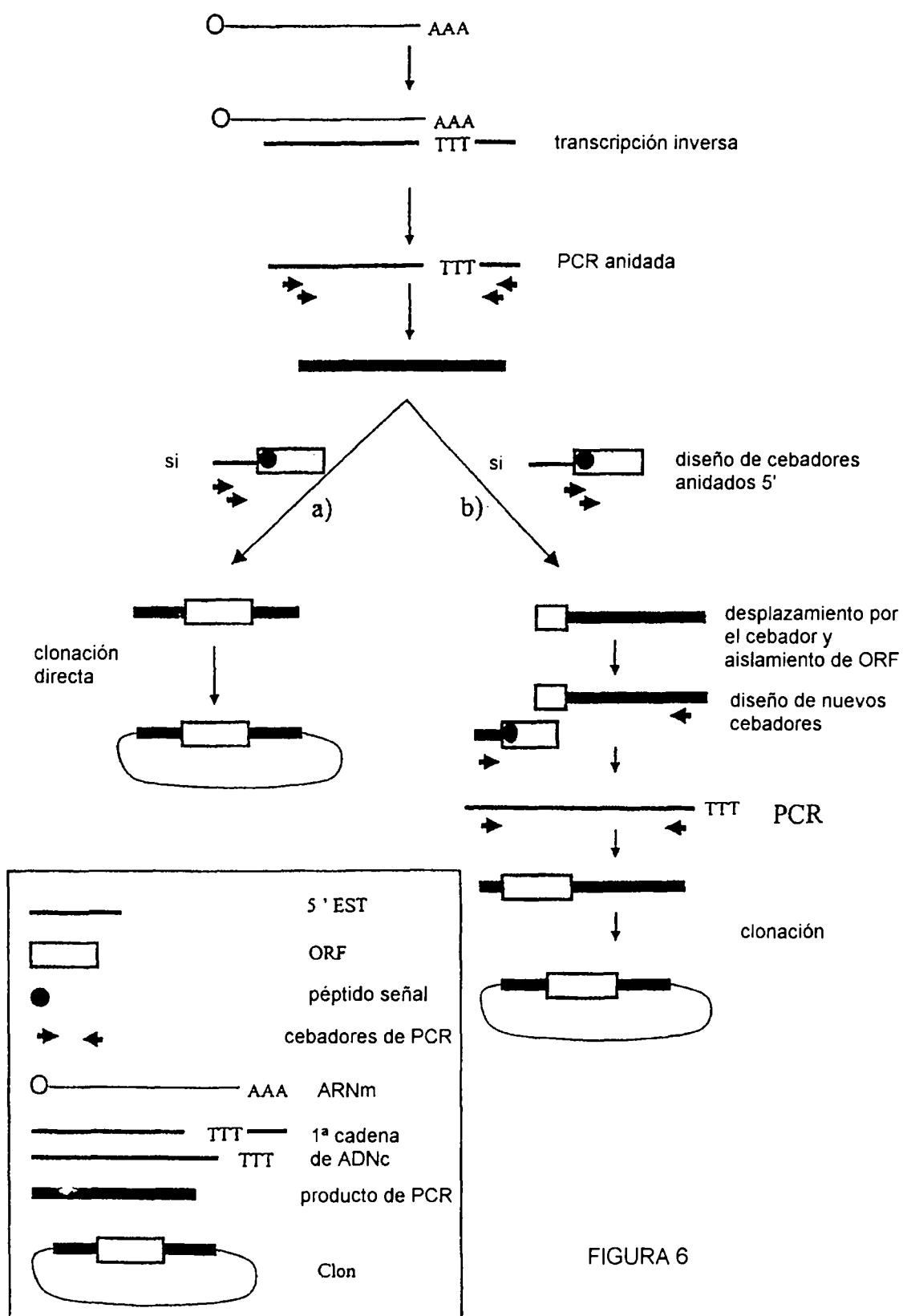
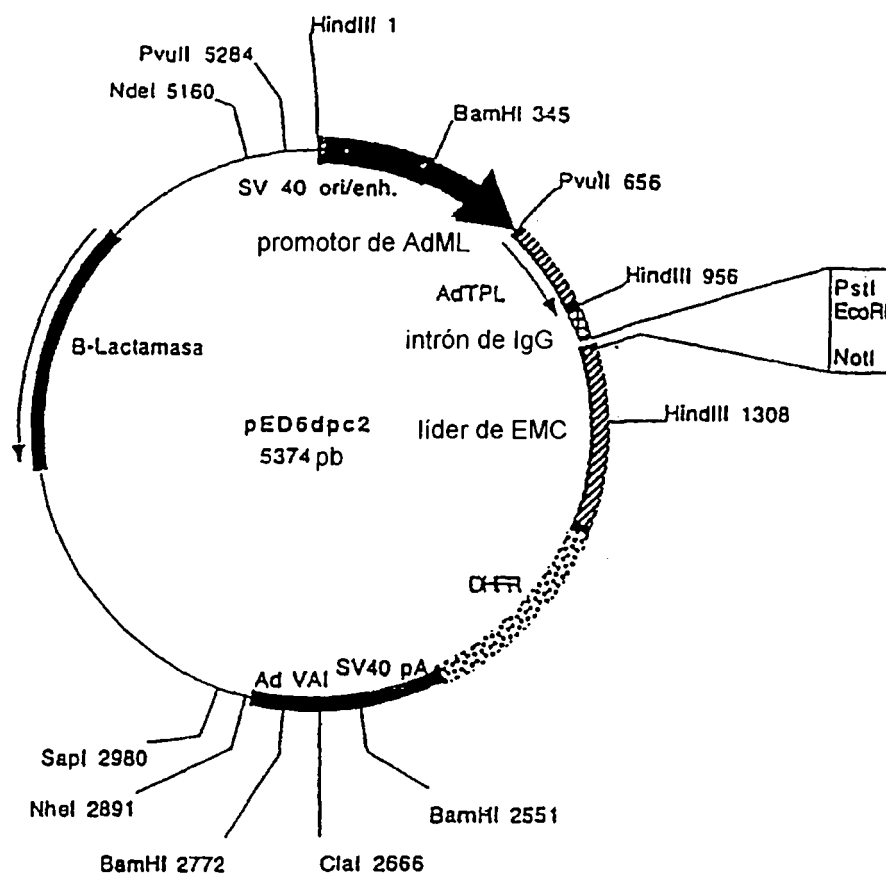


FIGURA 6



nombre del plásmido: **pED6dpc2**

tamaño del plásmido: **5374 pb**

FIGURA 7

Descripción de la estructura de promotores aislados a partir de 5' ESTde SignalTag

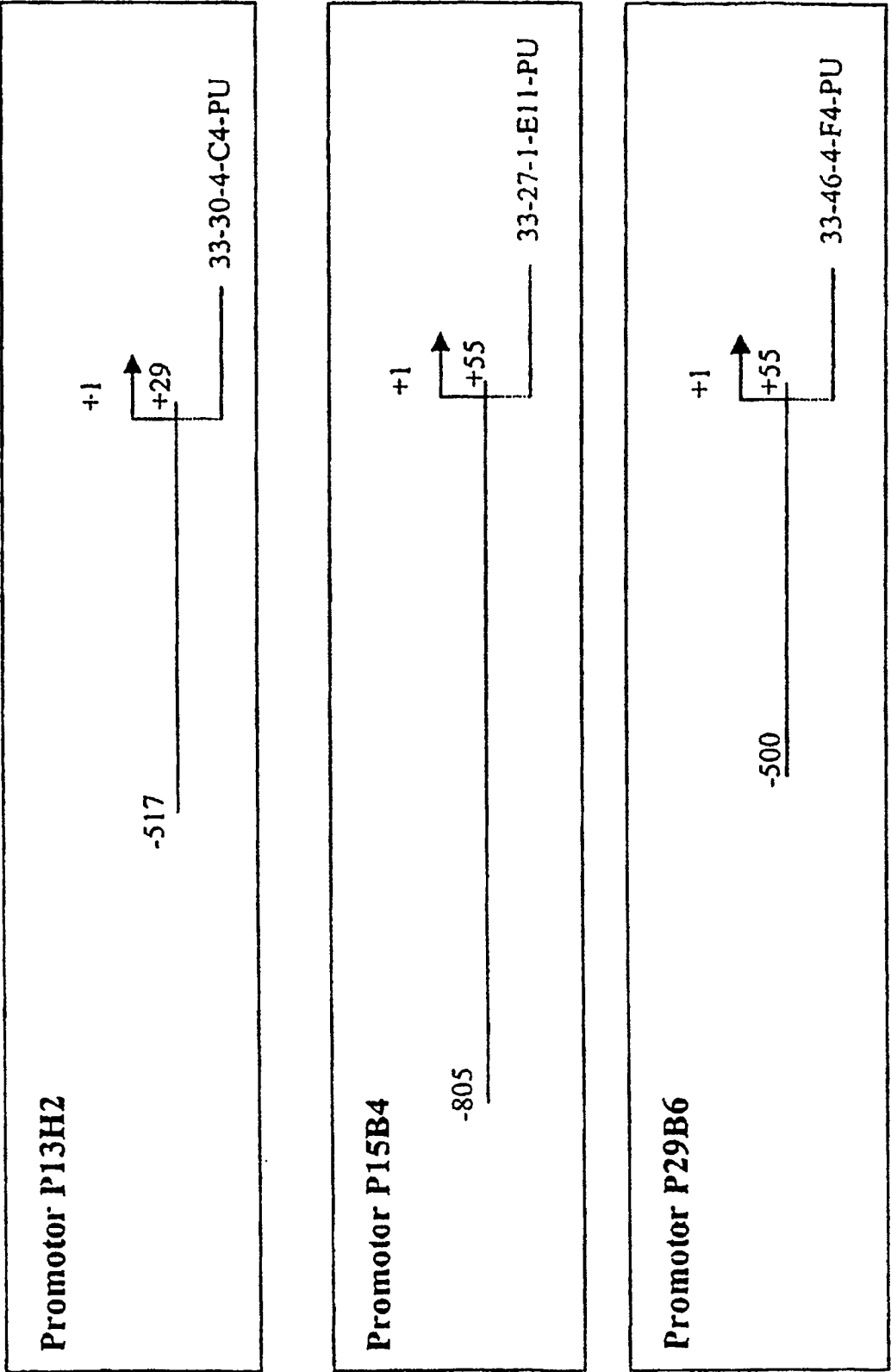


FIGURA 8

Descripción de sitios de unión a factor de transcripción presentes en promotores aislados a partir de secuencias de SignalTag

Secuencia promotora P13H2 (546 pb):

Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
CMYB_01	-502	+	0,983	9	TGTCAGTTG
MYOD_Q6	-501	-	0,961	10	CCCAACTGAC
S8_01	-444	-	0,960	11	AATAGAATTAG
S8_01	-425	+	0,966	11	AACTAAATTAG
DELTAEF1_01	-390	-	0,960	11	GCACACCTCAG
GATA_C	-364	-	0,964	11	AGATAAATCCA
CMYB_01	-349	+	0,958	9	CTTCAGTTG
GATA1_02	-343	+	0,959	14	TTGTAGATAGGACA
GATA_C	-338	+	0,953	11	AGATAGGACAT
TAL1ALPHA47_01	-235	+	0,973	16	CATAACAGATGGTAAG
TAL1BETA47_01	-235	+	0,983	16	CATAACAGATGGTAAG
TAL1BETA1TF2_01	-235	+	0,978	16	CATAACAGATGGTAAG
MYOD_Q6	-232	-	0,954	10	ACCATCTGTT
GATA1_04	-217	-	0,953	13	TCAAGATAAAGTA
IK1_01	-126	+	0,983	13	AGTTGGGAATTCC
IK2_01	-126	+	0,985	12	AGTTGGGAATTC
CREL_01	-123	+	0,962	10	TGGGAATTCC
GATA1_02	-96	+	0,950	14	TCAGTGATATGGCA
SRY_02	-41	-	0,951	12	TAAACAAAACA
E2F_02	-33	+	0,957	8	TTAGCGC
MZF1_01	-5	-	0,975	8	TGAGGGGA

Secuencia promotora P15B4 (861 pb):

Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
NFY_Q6	-748	-	0,956	11	GGACCAATCAT
MZF1_01	-738	+	0,962	8	CCTGGGGA
CMYB_01	-684	+	0,994	9	TGACCGTTG
VMYB_02	-682	-	0,985	9	TCCAACGGT
STAT_01	-673	+	0,968	9	TTCCTGGAA
STAT_01	-673	-	0,951	9	TTCCAGGAA
MZF1_01	-556	-	0,956	8	TTGGGGGA
IK2_01	-451	+	0,985	12	GAATGGGATTTC
MZF1_01	-424	+	0,986	8	AGAGGGGA
SRY_02	-398	-	0,955	12	GAAAACAAAACA
MZF1_01	-216	+	0,960	8	GAAGGGGA
MYOD_Q6	-190	+	0,981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0,958	11	TCCACCTTCC
S8_01	5	-	0,992	11	GAGGCAATTAT
MZF1_01	16	-	0,986	8	AGAGGGGA

Secuencia promotora P29B6 (555 pb):

Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
ARNT_01	-311	+	0,964	16	GGACTCACGTGCTGCT
NMYC_01	-309	+	0,965	12	ACTCACGTGCTG
USF_01	-309	+	0,985	12	ACTCACGTGCTG
USF_01	-309	-	0,985	12	CAGCACGTGAGT
NMYC_01	-309	-	0,956	12	CAGCACGTGAGT
MYC MAX_02	-309	-	0,972	12	CAGCACGTGAGT
USF_C	-307	+	0,997	8	TCACGTGC
USF_C	-307	-	0,991	8	GCACGTGA
MZF1_01	-292	-	0,968	8	CATGGGGA
ELK1_02	-105	+	0,963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0,974	10	TCCGGAAGCC
AP1_Q4	-42	-	0,963	11	AGTGACTGAAC
AP1FJ_Q2	-42	-	0,961	11	AGTGACTGAAC
PADS_C	45	+	1,000	9	TGTGGTCTC

Figura 9

ES 2 330 903 T3

LISTA DE SECUENCIAS

5	<110> Dumas Milne Edwards, Jean-Baptiste Duclert, Aymeric Bougueleret, Lydie	
	<120> ADNc extendidos para proteínas secretadas	
	<130> GENSET.019A	
10	<160> 41	
	<170> Patent.pm	
	<210> 1	
	<211> 47	
15	<212> ARN	
	<213> Secuencia artificial	
	<220>	
	<221> Producto de transcripción <i>in vitro</i>	
20	<221> base_modificada	
	<222> (1)...(1)	
	<223> m7g	
25	<400> 1	
	ngcauccuac ucceauccaa uuccacccua acuccucca ucuccac	47
30	<210> 2	
	<211> 46	
	<212> ARN	
35	<213> Secuencia artificial	
	<220>	
	<223> Producto de transcripción <i>in vitro</i>	
40	<400> 2	
	gcauccuacu cccauccaau uccacccuaa cuccuccau cuccac	46
45	<210> 3	
	<211> 25	
	<212> ADN	
	<213> Secuencia artificial	
50	<220>	
	<223> Producto de transcripción <i>in vitro</i>	
55	<400> 3	
	atcaagaatt cgcacgagac cacta	25
	<210> 4	
60	<211> 25	
	<212> ADN	
	<213> Secuencia artificial	
65	<220>	
	<223> Oligonucleótido	

ES 2 330 903 T3

	<400> 4	
	taatggtctc gtgcgaattc ttgat	25
5	<210> 5	
	<211> 25	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido	
15	<400> 5	
	ccgacaagac caacgtcaag gccgc	25
20	<210> 6	
	<211> 25	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> Oligonucleótido	
30	<400> 6	
	tcaccagcag gcagtggctt aggag	25
35	<210> 7	
	<211> 25	
	<212> ADN	
	<213> Secuencia artificial	
40	<220>	
	<223> Oligonucleótido	
45	<400> 7	
	agtgattcct gctactttgg atggc	25
50	<210> 8	
	<211> 25	
	<212> ADN	
	<213> Secuencia artificial	
55	<220>	
	<223> Oligonucleótido	
	<400> 8	
60	gcttggctctt gttctggagt ttaga	25
65	<210> 9	
	<211> 25	
	<212> ADN	
	<213> Secuencia artificial	

ES 2 330 903 T3

	<220>		
	<223> Oligonucleótido		
5	<400> 9		
	tccagaatgg gagacaagcc aattt		25
10	<210> 10		
	<211> 25		
	<212> ADN		
	<213> Secuencia artificial		
15	<220>		
	<223> Oligonucleótido		
	<400> 10		
20	agggaggagg aaacagcgtg agtcc		25
	<210> 11		
25	<211> 25		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
30	<223> Oligonucleótido		
	<400> 11		
35	atgggaaagg aaaagactca tatca		25
	<210> 12		
40	<211> 25		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
45	<223> Oligonucleótido		
	<400> 12		
50	agcagcaaca atcaggacag cacag		25
	<210> 13		
	<211> 25		
55	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> Oligonucleótido		
60	<400> 13		
	atcaagaatt cgcacgagac catta		25
65	<210> 14		
	<211> 67		

ES 2 330 903 T3

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> Oligonucleótido	
	<400> 14	
10	atcgttgaga ctcgtaccag cagagtcacg agagagacta cacgg tactg gttttttttt ttttttn	60 67
	<210> 15	
	<211> 29	
15	<212> ADN	
	<213> Secuencia artificial	
	<220>	
20	<223> Oligonucleótido	
	<400> 15	
25	ccagcagagt cacgagagag actacacgg	29
	<210> 16	
	<211> 25	
30	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Oligonucleótido	
35	<400> 16	
40	cacgagagag actacacggt actgg	25
	<210> 17	
	<211> 526	
	<212> ADN	
45	<213> <i>Homo sapiens</i>	
	<220>	
	<221> rasgo_misc	
	<222> complemento (261...376)	
50	<223> blastn	
	<221> rasgo_misc	
	<222> complemento (380...486)	
55	<223> blastn	
	<221> rasgo_misc	
	<222> complemento (110...145)	
	<223> blastn	
60	<221> rasgo_misc	
	<222> complemento (196...229)	
	<223> blastn	
65	<221> péptido_señal	
	<222> 90..140	
	<223> matriz de von Heijne	

ES 2 330 903 T3

<400> 17

```

    aatatrarac agctacaata ttccagggcc artcacttgc cattttctcat aacagcgtca      60
    gagagaaaaga actgactgar acgttttgag atg aag aaa gtt ctc ctc ctg atc      113
    Met Lys Lys Val Leu Leu Leu Ile
    -15 -10
    aca gcc atc ttg gca gtg gct gtw ggt ttc cca gtc tct caa gac cag      161
    Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln
    -5 1 5
    gaa cga gaa aaa aga agt atc agt gac agc gat gaa tta gct tca ggr      209
    Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Glv
    10 10 15 20
    wtt ttt gtg ttc cct tac cca tat cca ttt cgc cca ctt cca cca att      257
    Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile
    25 30 35
    cca ttt cca aga ttt cca tgg ttt aga cgt aan ttt cct att cca ata      305
    Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Xaa Phe Pro Ile Pro Ile
    40 45 50 55
    cct gaa tct gcc cct aca act ccc ctt cct agc gaa aag taaacaaraa      354
    Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys
    60 65
    ggaaaagtca crataaacct ggtcacctga aattgaaatt gagccacttc cttgaaraat      414
    caaaattcct gttaataaaa raaaaacaaa tgtaattgaa atagcacaca gcattctcta      474
    gtcaatatct ttagtgatct tctttaataa acatgaaagc aaaaaaaaaa aa      526

```

<210> 18

25 <211> 17

<212> PRT

<213> *Homo sapiens*

<220>

30 <221> SEÑAL

<222> 1..17

<223> matriz de von Heijne

puntuación 8,2

35 sec LLLITAILAVAVAG/FP

<400> 18

```

    Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val
    1 5 10 15
    Gly

```

<210> 19

<211> 822

45 <212> ADN

<213> *Homo sapiens*

<220>

50 <221> rasgo_misc

<222> 260..184

<223> blastn

<221> rasgo_misc

55 <222> 118..184

<223> blastn

<221> rasgo_misc

60 <222> 56..113

<223> blastn

<221> rasgo_misc

<222> 454..485

65 <223> blastn

<221> rasgo_misc

<222> 118..545

ES 2 330 903 T3

```

<223> blastn
<221> rasgo_misc
<222> 65..369
5 <223> blastn
<221> rasgo_misc
<222> 61..399
<223> blastn
10 <221> rasgo_misc
<222> 408..458
<223> blastn
15 <221> rasgo_misc
<222> 69..399
<223> blastn
20 <221> rasgo_misc
<222> 393..432
<223> blastn
<221> péptido_señal
25 <222> 346..408
<223> matriz de von Heijne

<400> 19

30 actccttttta gcatagggggc ttcggcgcca gcggccagcg ctagtcgggc tggtaagtgc 60
ctgatgccga gttccgtctc tcgcgtcttt tcttggtccc aggcaaagcg gasgnagatc 120
ctcaaacggc ctagtgcttc gcgcttcggc agaaaatcag cggctctaatt aattcctctg 180
gtttgttgaa gcagttacca agaattctca accctttccc acaaaagcta attgagtaca 240
cgttcctggt gagtacacgt tctgttgat ttacaaaagg tgcaggtatg agcaggtctg 300
35 aagactaaca ttttgtgaag ttgtaaaaca gaaaacctgt tagaa atg tgg tgg ttt 357
Met Trp Trp Phe
-20
cag caa ggc ctc agt ttc ctt cct tca gcc ctt gta att tgg aca tct 405
Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val Ile Trp Thr Ser
-15 -10 -5
40 gct gct ttc ata ttt tca tac att act gca gta aca ctc cac cat ata 453
Ala Ala Phe Ile Phe Ser Tyr Ile Thr Ala Val Thr Leu His His Ile
1 5 10 15
gac ccg gct tta cct tat atc agt gac act ggt aca gta gct cca raa 501
Asp Pro Ala Leu Pro Tyr Ile Ser Asp Thr Gly Thr Val Ala Pro Xaa
20 25 30
45 aaa tgc tta ttt ggg gca atg cta aat att gcg gca gtt tta tgt caa 549
Lys Cys Leu Phe Gly Ala Met Leu Asn Ile Ala Ala Val Leu Cys Gln
35 40 45
aaa tagaaatcag gaarataatt caacttaaag aakttcattt catgacccaa 602
Lys
50 ctcttcaraa acatgtcttt acaagcatat ctcttgatt gctttctaca ctggtgaatt 662
gtctggcaat atttctgcag tggaaaattt gatctarmta gttcttgact gataaatatg 722
gtaagggtggg cttttccccc tgtgtaattg gctactatgt cttactgagc caagttgtaw 782
tttgaataaa aatgatatga gagtgacaca aaaaaaaaaa 822

55 <210> 20
<211> 21
<212> PRT
60 <213> Homo sapiens
<220>
<221> SEÑAL
<222> 1..21
65 <223> matriz de von Heijne
puntuación 5,5
sec SFLPSALVIWTS/AF

```

ES 2 330 903 T3

<400> 20

Met Trp Trp Phe Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val
1 5 10 15
Ile Trp Thr Ser Ala
20

<210> 21

<211> 405

<212> ADN

<213> *Homo sapiens*

<220>

<221> rasgo_misc

<222> complemento (103..398)

<223> blastn

<221> péptido_señal

<222> 185..295

<223> matriz de von Heijne

<400> 21

atcaccttct tctccatcct tctctggggc agtccccarc ccagtccttc tectgacctg 60
cccagcccaa gtcagccttc agcacgcgct tttctgcaca cagatattcc aggcctacct 120
ggcattccag gacotccgma atgatgctcc agtcccttac aagcgttccc tggatgaggg 180
tggc atg gtg ctg acc acc ctc ccc ttg ccc tct gcc aac agc cct gtg 229
Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val
-35 -30 -25
aac atg ccc acc act ggc ccc aac agc ctg agt tat gct agc tct gcc 277
Asn Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala
-20 -15 -10
ctg tcc ccc tgt ctg acc gct cca aak tcc ccc egg ctt gct atg atg 325
Leu Ser Pro Cys Leu Thr Ala Pro Xaa Ser Pro Arg Leu Ala Met Met
-5 1 5 10
cct gac aac taaatctcct tatccaaatc aataaarwra raatcctccc 374
Pro Asp Asn
tccaraaggg tttctaaaaa caaaaaaaa a 405

<210> 22

<211> 37

<212> PRT

<213> *Homo sapiens*

<220>

<221> SEÑAL

<222> 1..37

<223> matriz de von Heijne

puntuación 5,9

sec LSYASSALSPCLT/AP

<400> 22

Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val Asn
1 5 10 15
Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala Leu
20 25 30
Ser Pro Cys Leu Thr
35

<210> 23

<211> 496

ES 2 330 903 T3

<212> ADN

<213> *Homo sapiens*

<220>

5 <221> rasgo_misc

<222> 149..331

<223> blastn

10 <221> rasgo_misc

<222> 328..485

<223> blastn

<221> rasgo_misc

15 <222> complemento (182..496)

<223> blastn

<221> péptido_señal

<222> 196..240

20 <223> matriz de von Heijne

<400> 23

```

25      aaaaaattgg tcccagtttt caccctgccg cagggctggc tggggagggc agcggtttag      60
      attagccgtg gcctaggccg tttaacgggg tgacacgagc ntgcagggcc gagtccaagg      120
      cccggagata ggaccaaccg tcaggaatgc gaggaatggt tttcttcgga ctctatcgag      180
      gcacacagac agacc atg ggg att ctg tct aca gtg aca gcc tta aca ttt      231
                        Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
30                        -15                        -10                        -5
      gcc ara gcc ctg gac ggc tgc aga aat ggc att gcc cac cct gca agt      279
      Ala Xaa Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser
                        1                        5                        10
      gag aag cac aga ctc gag aaa tgt agg gaa ctc gag asc asc cac tcg      327
      Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Xaa Xaa His Ser
                        15                        20                        25
      gcc cca gga tca acc cas cac cga aga aca acc aga aga aat tat      375
      Ala Pro Gly Ser Thr Xaa His Arg Arg Lys Thr Thr Arg Arg Asn Tyr
      30                        35                        40                        45
40      tct tca gcc tgaaatgaak ccgggatcaa atggttgctg atcaragccc      424
      Ser Ser Ala
      atatttaaat tggaaaagtc aaattgasca ttattaaata aagcttggtt aatatgtctc      484
      aaacaaaaaa aa      496

```

45

<210> 24

<211> 15

<212> PRT

50 <213> *Homo sapiens*

<220>

<221> SEÑAL

55 <222> 1..15

<223> matriz de von Heijne

puntuación 5,5

sec ILSTVTALTFAXA/LD

60

<400> 24

```

      Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe Ala Xaa Ala
      1                        5                        10                        15

```

65

<210> 25

<211> 623

ES 2 330 903 T3

<212> ADN
 <213> *Homo sapiens*
 <220>
 5 <221> péptido_señal
 <222> 49..96
 <223> matriz de von Heijne
 10 <400> 25

```

aaagatccct gcagcccggc aggagagaag gctgagcctt ctggcgctc atg gag agg      57
                                     Met Glu Arg
                                     -15
15   ctc gtc cta acc ctg tgc acc ctc ccg ctg gct gtg gcg tct gct ggc      105
    Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly
                                     -10
                                     -5
                                     1
    tgc gcc acg acg cca gct cgc aac ctg agc tgc tac cag tgc ttc aag      153
    Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys
    5                                     10                                     15
20   gtc agc agc tgg acg gag tgc ccg ccc acc tgg tgc agc ccg ctg gac      201
    Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp
    20                                     25                                     30                                     35
    caa gtc tgc atc tcc aac gag gtg gtc gtc tct ttt aaa tgg agt gta      249
    Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys Trp Ser Val
                                     40                                     45                                     50
25   cgc gtc ctg ctc agc aaa cgc tgt gct ccc aga tgt ccc aac gac aac      297
    Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro Asn Asp Asn
    55                                     60                                     65
    atg aak ttc gaa tgg tgc ccg gcc ccc atg gtg caa ggc gtg atc acc      345
    Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly Val Ile Thr
    70                                     75                                     80
30   agg cgc tgc tgt tcc tgg gct ctc tgc aac agg gca ctg acc cca cag      393
    Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu Thr Pro Gln
    85                                     90                                     95
35   gag ggg cgc tgg gcc ctg cra ggg ggg ctc ctg ctc cag gac cct tcg      441
    Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Leu Gln Asp Pro Ser
    100                                     105                                     110                                     115
    agg ggc ara aaa acc tgg gtg cgg cca cag ctg ggg ctc cca ctc tgc      489
    Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu Pro Leu Cys
    120                                     125                                     130
40   ctt ccc awt tcc aac ccc ctc tgc cca rgg gaa acc cag gaa gga      534
    Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln Glu Gly
    135                                     140                                     145
    taacactgtg ggtgccccca cctgtgcatt gggaccacra cttcaccctc ttggaracaa      594
    taaactctca tgcccccaaa aaaaaaaa
    623
45
    <210> 26
    <211> 15
    <212> PRT
    50 <213> Homo sapiens
    <220>
    <221> SEÑAL
    55 <222> 1..16
    <223> matriz de von Heijne
    puntuación 10,1
    sec LVLTLCPLAVA/SA
    60 <400> 26
  
```

Met Glu Arg Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala
 1 5 10 15

65 <210> 27
 <211> 848

ES 2 330 903 T3

<212> ADN

<213> *Homo sapiens*

<220>

5 <221> péptido_señal

<222> 32..73

<223> matriz de von Heijne

10 <400> 27

```

aactttgcct tgtgtttttcc accctgaaag a atg ttg tgg ctg ctc ttt ttt      52
                               Mat Leu Trp Leu Leu Phe Phe
                               -10
ctg gtg act gcc att cat gct gaa ctc tgt caa cca ggt gca gaa aat      100
Leu Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn
                               -5
gct ttt aaa gtg aga ctt agt atc aga aca gct ctg gga gat aaa gca      148
Ala Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala
                               10
tat gcc tgg gat acc aat gaa gaa tac ctc ttc aaa gcg atg gta gct      196
Tyr Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala
                               15
ttc tcc atg aga aaa gtt ccc aac aga gaa gca aca gaa att tcc cat      244
Phe Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His
                               20
gtc cta ctt tgc aat gta acc cag agg gta tca ttc tgg ttt gtg gtt      292
Val Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val
                               25
aca gac cct tca aaa aat cac acc ctt cct gct gtt gag gtg caa tca      340
Thr Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser
                               30
gcc ata aga atg aac aag aac cgg atc aac aat gcc ttc ttt cta aat      368
Ala Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn
                               35
gac caa act ctg gaa ttt tta aaa atc cct tcc aca ctt gca cca ccc      436
Asp Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro
                               40
atg gac cca tct gtg ccc atc tgg att att ata ttt ggt gtg ata ttt      484
Met Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe
                               45
tgc atc atc ata gtt gca att gca cta ctg att tta tca ggg atc tgg      532
Cys Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp
                               50
caa cgt ada ara aag aac aaa gaa cca tct gaa gtg gat gac gct gaa      580
Gln Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu
                               55
rat aak tgt gaa aac atg atc aca att gaa aat ggc atc ccc tct gat      628
Xaa Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp
                               60
ccc ctg gac atg aag gga ggg cat att aat gat gcc ttc atg aca gag      676
Pro Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu
                               65
gat gag agg ctc acc cct ctc tgaagggtg ttgtctgtct tcttcaaraa      727
Asp Glu Arg Leu Thr Pro Leu
                               70
attaaacatt tgtttctgtg tgactgctga gcaccttgaa ataccaagag cagatcatat      787
wttttgtttc accattcttc ttttgtaata aattttgaat gtgcttgaaa aaaaaaaaaa      847
c                                                                848

```

60 <210> 28

<211> 14

<212> PRT

<213> *Homo sapiens*

65 <220>

<221> SEÑAL

ES 2 330 903 T3

<222> 1..14
 <223> matriz de von Heijne
 puntuación 10,7
 sec LWLLFFLVTAIHA/EL
 5
 <400> 28
Met Leu Trp Leu Leu Phe Phe Leu Val Thr Ala Ile His Ala
1 5 10
 10
 <210> 29
 <211> 25
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido
 20 <400> 29
 gggaagatgg agatagtatt gcctg 25
 25 <210> 30
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Oligonucleótido
 35 <400> 30
 ctgccatgta catgatagag agattc 26
 40 <210> 31
 <211> 546
 <212> ADN
 <213> *Homo sapiens*
 45 <220>
 <221> promotor
 <222> 1..517
 <221> sitio de inicio de la transcripción
 50 <222> 518
 <221> unión_proteína
 <222> 17..25
 55 <223> predicción por matinspector
 nombre CMYB_01
 puntuación 0,983
 secuencia tgcagttg
 60 <221> unión_proteína
 <222> complemento (18..27)
 <223> predicción por matinspector
 nombre MYOD_Q6
 65 puntuación 0,961
 secuencia cccaactgac

ES 2 330 903 T3

<221> unión_proteína
 <222> complemento (75..85)
 <223> predicción por matinspector
 nombre S8_01
 puntuación 0,960
 secuencia aatagaattag

<221> unión_proteína
 <222> 94..104
 <223> predicción por matinspector
 nombre S8_01
 puntuación 0,966
 secuencia aactaaattag

<221> unión_proteína
 <222> complemento (129..139)
 <223> predicción por matinspector
 nombre DELTAEF1_01
 puntuación 0,960
 secuencia gcacacctcag

<221> unión_proteína
 <222> complemento (155..165)
 <223> predicción por matinspector
 nombre GATA_C
 puntuación 0,964
 secuencia agataaatcca

<221> unión_proteína
 <222> 170..178
 <223> predicción por matinspector
 nombre CMYB_01
 puntuación 0,958
 secuencia ctccagttg

<221> unión_proteína
 <222> 176..189
 <223> predicción por matinspector
 nombre GATA1_02
 puntuación 0,959
 secuencia ttgtagataggaca

<221> unión_proteína
 <222> 180..190
 <223> predicción por matinspector
 nombre GATA_C
 puntuación 0,953
 secuencia agataggacat

<221> unión_proteína
 <222> 284..299
 <223> predicción por matinspector
 nombre TAL1ALPHA47_01
 puntuación 0,973
 secuencia cataacagatggaag

ES 2 330 903 T3

<221> unión_proteína
 <222> 284..299
 <223> predicción por matinspector
 nombre TAL1BETAE47_01
 puntuación 0,983
 secuencia cataacagatggtaag

<221> unión_proteína
 <222> 284..299
 <223> predicción por matinspector
 nombre TAL1BETAITF2_01
 puntuación 0,978
 secuencia cataacagatggtaag

<221> unión_proteína
 <222> complemento (287..296)
 <223> predicción por matinspector
 nombre MYOD_Q6
 puntuación 0,954
 secuencia accatctgtt

<221> unión_proteína
 <222> complemento (302..314)
 <223> predicción por matinspector
 nombre GATA1_04
 puntuación 0,953
 secuencia tcaagataaagta

<221> unión_proteína
 <222> 393..405
 <223> predicción por matinspector
 nombre IK1_01
 puntuación 0,963
 secuencia agttgggaattcc

<221> unión_proteína
 <222> 393..404
 <223> predicción por matinspector
 nombre IK2_01
 puntuación 0,985
 secuencia agttgggaattc

<221> unión_proteína
 <222> 396..405
 <223> predicción por matinspector
 nombre CREL_01
 puntuación 0,962
 secuencia tgggaattcc

<221> unión_proteína
 <222> 423..436
 <223> predicción por matinspector
 nombre GATA1_02
 puntuación 0,950
 secuencia tcaagtgatatggca

ES 2 330 903 T3

<221> unión_proteína
 <222> complemento (478..489)
 <223> predicción por matinspector
 nombre SRY_02
 puntuación 0,951
 secuencia taaaacaaaaca

<221> unión_proteína
 <222> 486..493
 <223> predicción por matinspector
 nombre E2F_02
 puntuación 0,957
 secuencia tttagcgc

<221> unión_proteína
 <222> complemento (514..521)
 <223> predicción por matinspector
 nombre MZF1_01
 puntuación 0,975
 secuencia tgagggga

<400> 31

tgagtgcagt	gttacatgtc	agttgggtta	agtttggtta	tgtcattcaa	atcttctatg	60
tcttgatttg	cctgctaatt	ctattatttc	tggaaactaa	ctagtttgat	ggttctatta	120
gttattgact	gaggtgtgct	aatctcccat	tatgtggatt	tatctatttc	ttcagttgta	180
gataggacat	tgatagatac	ataagtacca	ggacaaaagc	agggagatct	tttttccaaa	240
atcaggagaa	aaaaatgaca	tctggaaaac	ctatagggaa	aggcataaca	gatggtaagg	300
atactttatc	ttgagttaga	gagccttcct	gtggcaacgt	ggagaaggga	agaggtcgta	360
gaattgagga	gtcagctcag	ttagaagcag	ggagttggga	attccgttca	tgtgatttag	420
catcagtgat	atggcaaatg	tgggactaag	ggtagtgatc	agaggggtta	aattgtgtgt	480
tttgttttag	cgctgctggg	gcategcctt	gggtcccctc	aaacagattc	ccatgaatct	540
cttcat						546

<210> 32
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido

<400> 32

gtaccaggga ctgtgacat tgc

<210> 33
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido

<400> 33

ctgtgacat tgctcccaag agag

ES 2 330 903 T3

<212> ADN
 <213> *Homo sapiens*
 <220>
 5 <221> promotor
 <222> 1..806
 <221> sitio de inicio de la transcripción
 10 <222> 807

 <221> unión_proteína
 <222> complemento (60...-70)
 15 <223> predicción por matinspector
 nombre NFY_Q6
 puntuación 0,956
 secuencia ggaccaatcat

 20 <221> unión_proteína
 <222> 70..77
 <223> predicción por matinspector
 nombre MZP1_01
 25 puntuación 0,962
 secuencia cctgggga

 <221> unión_proteína
 <222> 124..132
 30 <223> predicción por matinspector
 nombre CMYB_01
 puntuación 0,994
 secuencia tgaccgttg
 35
 <221> unión_proteína
 <222> complemento (126..134)
 <223> predicción por matinspector
 40 nombre VMYB_02
 puntuación 0,985
 secuencia tccaacggt

 45 <221> unión_proteína
 <222> 135..143
 <223> predicción por matinspector
 nombre STAT_01
 50 puntuación 0,968
 secuencia ttctggaa

 <221> unión_proteína
 <222> complemento (135..143)
 55 <223> predicción por matinspector
 nombre STAT_01
 puntuación 0,951
 secuencia ttccaggaa
 60
 <221> unión_proteína
 <222> complemento (252..259)
 <223> predicción por matinspector
 65 nombre MZF1_01
 puntuación 0,956
 secuencia ttggggga

ES 2 330 903 T3

<221> unión_proteína
 <222> 357..368
 <223> predicción por matinspector
 nombre IK2_01
 puntuación 0,965
 secuencia gaatgggatttc

<221> unión_proteína
 <222> 384..391
 <223> predicción por matinspector
 nombre MZF1_01
 puntuación 0,986
 secuencia agagggga

<221> unión_proteína
 <222> complemento (410..421)
 <223> predicción por matinspector
 nombre SRY_02
 puntuación 0,955
 secuencia gaaaacaaaaca

<221> unión_proteína
 <222> 592..599
 <223> predicción por matinspector
 nombre MZF1_01
 puntuación 0,960
 secuencia gaagggga

<221> unión_proteína
 <222> 618..627
 <223> predicción por matinspector
 nombre MYOD_Q6
 puntuación 0,981
 secuencia agcatctgcc

<221> unión_proteína
 <222> 632..642
 <223> predicción por matinspector
 nombre DELTAEF1_01
 puntuación 0,958
 secuencia tccaccttc

<221> unión_proteína
 <222> complemento (813..823)
 <223> predicción por matinspector
 nombre S8_01
 puntuación 0,992
 secuencia gaggcaatcat

<221> unión_proteína
 <222> complemento (824..831)
 <223> predicción por matinspector
 nombre MZF1_01
 puntuación 0,986
 secuencia agagggga

ES 2 330 903 T3

<400> 34		
	tactataggg caccgctggg cgacggccgg gctgttctgg agcagagggc atgtcagtaa 60	
	tgattggtcc ctggggaagg tctggctggc tccagcacag tgaggcattt aggtatctct 120	
5	cggtgaccgt tggattcctg gaagcagtag ctgttctgtt tggatctggg agggacaggg 180	
	ctcagagggc taggcacgag ggaaggctag aggagaaggs aggsarggcc cagtgagarg 240	
	ggagcatgcc tcccccaac cctggcttsc ycttggyam agggcgkty tgggmacttr 300	
	aaytcagggc ccaascagaa scacaggccc aktcntggct smaagcacia tagcctgaat 360	
	gggatttcag gttagncagg gtgagagggg aggetctctg gcttagtctt gttttgtttt 420	
10	ccaaatcaag gtaacttgct cccttctgct acgggccttg gtcttggett gtcctcacc 480	
	agtcggaact ccctaccact ttcaggagag tggttttagg cccgtggggc tgttctgttc 540	
	caagcagtgt gagaacatgg ctggtagagg ctctagctgt gtgcggggcc tgaaggggag 600	
	tgggttctcg cccaaagagc atctgcccac tcccacott cccttctccc accagaagct 660	
	tgcttgagct gtttggacaa aaatccaaac ccacttggc tactctggcc tggcttcagc 720	
15	ttggaaccca atacctaggc ttacaggcca tcctgagcca ggggcctctg gaaattctct 780	
	tcttgatggg cctttaggtt tgggcacaaa atataattgc ctctcccctc tcccattttc 840	
	tctcttggga gcaatgggtc c	861
<210> 35		
20	<211> 20	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> Oligonucleótido	
<400> 35		
30	ctgggatgga aggcacggta	20
<210> 36		
35	<211> 20	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
40	<223> Oligonucleótido	
<400> 36		
45	gagaccacac agctagacaa	20
<210> 37		
	<211> 555	
50	<212> ADN	
	<213> <i>Homo sapiens</i>	
	<220>	
55	<221> promotor	
	<222> 1..500	
	<221> sitio de inicio de la transcripción	
	<222> 501	
60	<221> unión_proteína	
	<222> 191..206	
	<223> predicción por matinspector	
65	nombre ARNT_01	
	puntuación 0,964	
	secuencia ggactcacgtgctgct	

ES 2 330 903 T3

<221> unión_proteína
 <222> 193..204
 <223> predicción por matinspector
 nombre NMYC_01
 puntuación 0,965
 secuencia actcacgtgctg

<221> unión_proteína
 <222> 193..204
 <223> predicción por matinspector
 nombre USP_01
 puntuación 0,985
 secuencia actcacgtgctg

<221> unión_proteína
 <222> complemento (193..204)
 <223> predicción por matinspector
 nombre USP_01
 puntuación 0,985
 secuencia cagcacgtgagt

<221> unión_proteína
 <222> complemento (193..204)
 <223> predicción por matinspector
 nombre NMYC_01
 puntuación 0,956
 secuencia cagcacgtgagt

<221> unión_proteína
 <222> complemento (193..204)
 <223> predicción por matinspector
 nombre MYCMAX_02
 puntuación 0,972
 secuencia cagcacgtgagt

<221> unión_proteína
 <222> 195..202
 <223> predicción por matinspector
 nombre USF_C
 puntuación 0,997
 secuencia tcacgtgc

<221> unión_proteína
 <222> complemento (195..202)
 <223> predicción por matinspector
 nombre USF_C
 puntuación 0,991
 secuencia gcacgtga

<221> unión_proteína
 <222> complemento (210..217)
 <223> predicción por matinspector
 nombre MZF1_01
 puntuación 0,968
 secuencia catgggga

ES 2 330 903 T3

5 <221> unión_proteína
 <222> 397..410
 <223> predicción por matinspector
 nombre ELK1_02
 puntuación 0,963
 secuencia ccctccggaagcct

10 <221> unión_proteína
 <222> 400..409
 <223> predicción por matinspector
 nombre CETS1P54_01
 puntuación 0,974
 15 secuencia tccggaagcc

<221> unión_proteína
 <222> complemento (460..470)
 20 <223> predicción por matinspector
 nombre AP1_Q4
 puntuación 0,963
 secuencia agtgactgaac

25 <221> unión_proteína
 <222> complemento (460..470)
 <223> predicción por matinspector
 nombre AP1FJ_Q2
 30 puntuación 0,961
 secuencia agtgactgaac

<221> unión_proteína
 35 <222> 547..555
 <223> predicción por matinspector
 nombre PADS_C
 puntuación 1,000
 40 secuencia tgtgtgtctc

<400> 37

45	ctatagggca	cgcktggtcg	acggcccggg	ctgggtctggt	ctgtkgtgga	gtcggggtga	60
	aggacagcat	ttgtkacatc	tggtctactg	cacottccct	ctgecgtgca	cttggccttt	120
	kawaagctca	gcacgggtgc	ccatcacagg	gccggcagca	cacacatccc	attactcaga	180
	aggaactgac	ggactcacgt	gctgctccgt	ccccatgagc	tcagtggacc	tgtctatgta	240
	gagcagtcag	acagtgcctg	ggatagagtg	agagttcagc	cagtaaatcc	aagtgattgt	300
50	cattcctgtc	tgcattagta	actcccaacc	tagatgtgaa	aacttagttc	tttctcatag	360
	gttgcctctg	ccatggtccc	actgcagacc	caggcactct	ccggaagcct	ggaaatcacc	420
	cgtgtcttct	gcctgctccc	gtcacatcc	cacacttggt	ttcagtcact	gagttacaga	480
	ttttgcctcc	tcaattctct	ttgtcttagt	cccatcctct	gttcccctgg	ccagtttgtc	540
	tagctgtgtg	gtctc					555

55 <210> 38
 <211> 19
 60 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Oligonucleótido

65

ES 2 330 903 T3

<400> 38

ggccatacac ttgagtgac

19

5

<210> 39

<211> 19

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Oligonucleótido

15 <400> 39

atatagacaa acgcacacc

19

20 <210> 45

<211> 2036

<212> ADN

<213> *Homo sapiens*

25

<220>

<221> CDS

<222> 276..1040

30

<221> péptido_señal

<222> 276..485

<223> matriz de von Heijne
puntuación 3,9

35

sec SVIGVMLAPFTAG/LS

<221> sitio de poli A

<222> 2024..2036

40 <400> 45

gatcctgggt gcagctcatc acaagcgtcg ggggtgcagca aaaccatcca ggctggacag	60
tggttgga gttccaagaa aagaacgct tcaccgaaga agtcattgaa tacttccaga	120
agaaagttag cccagtgcac ctgaaaatcc tgctgactag cgatgaagcc tggaagagat	180
tcgtgcgtgt ggctggattg cccaggggaag aagcagatgc tctctatgaa gctctgaaga	240
atcttacacc atatgtggct attgaggaca aagac atg cag caa aaa gaa cag	293
Met Gln Gln Lys Glu Gln	
-70 -65	
cag ttt agg gag tgg ttt ttg aaa gag ttt cct caa atc taga tgg aag	341
Gln Phe Arg Glu Trp Phe Leu Lys Glu Phe Pro Gln Ile Arg Trp Lys	
-60 -55 -50	
att cag gag tcc ata gaa agg ctt cgt gtc att gca aat gag att gaa	389

55

60

65

ES 2 330 903 T3

	Ile	Gln	Glu	Ser	Ile	Glu	Arg	Leu	Arg	Val	Ile	Ala	Asn	Glu	Ile	Glu	
				-45					-40				-35				
	aag	gtc	cac	aga	ggc	tgc	gtc	atc	gcc	aat	gtg	gtg	tct	ggc	tcc	act	437
5	Lys	Val	His	Arg	Gly	Cys	Val	Ile	Ala	Asn	Val	Val	Ser	Gly	Ser	Thr	
			-30				-25				-20						
	ggc	atc	ctg	tct	gtc	att	ggc	gtt	atg	ttg	gca	cca	ttt	aca	gca	ggg	485
	Gly	Ile	Leu	Ser	Val	Ile	Gly	Val	Met	Leu	Ala	Pro	Phe	Thr	Ala	Gly	
		-15				-10			-5								
10	ctg	agc	ctg	agc	att	act	gca	gct	ggg	gta	ggg	ctg	gga	ata	gca	tct	533
	Leu	Ser	Leu	Ser	Ile	Thr	Ala	Ala	Gly	Val	Gly	Leu	Gly	Ile	Ala	Ser	
	1				5				10					15			
	gcc	acg	gct	ggg	atc	gcc	tcc	agc	atc	gtg	gag	aac	aca	tac	aca	agg	581
	Ala	Thr	Ala	Gly	Ile	Ala	Ser	Ser	Ile	Val	Glu	Asn	Thr	Tyr	Thr	Arg	
15			20				25					30					
	tca	gca	gaa	ctc	aca	gcc	agc	agg	ctg	act	gca	acc	agc	act	gac	caa	629
	Ser	Ala	Glu	Leu	Thr	Ala	Ser	Arg	Leu	Thr	Ala	Thr	Ser	Thr	Asp	Gln	
			35				40					45					
20	ttg	gag	gca	tta	agg	gac	att	ctg	cat	gac	atc	aca	ccc	aat	gtg	ctt	677
	Leu	Glu	Ala	Leu	Arg	Asp	Ile	Leu	His	Asp	Ile	Thr	Pro	Asn	Val	Leu	
		50				55				60							
	tcc	ttt	gca	ctt	gat	ttt	gac	gaa	gcc	aca	aaa	atg	att	gcg	aat	gat	725
	Ser	Phe	Ala	Leu	Asp	Phe	Asp	Glu	Ala	Thr	Lys	Met	Ile	Ala	Asn	Asp	
	65				70		75			80							
25	gtc	cat	aca	ctc	agg	aga	tct	aaa	gcc	act	gtt	gga	cgc	cct	ttg	att	773
	Val	His	Thr	Leu	Arg	Arg	Ser	Lys	Ala	Thr	Val	Gly	Arg	Pro	Leu	Ile	
				85			90			95							
	gct	tgg	cga	tat	gta	cct	ata	aat	gtt	gtt	gag	aca	ctg	aga	aca	cgt	821
	Ala	Trp	Arg	Tyr	Val	Pro	Ile	Asn	Val	Val	Glu	Thr	Leu	Arg	Thr	Arg	
30			100				105					110					
	ggg	gcc	ccc	acc	cgg	ata	gtg	aga	aaa	gta	gcc	cgg	aac	ctg	ggc	aag	869
	Gly	Ala	Pro	Thr	Arg	Ile	Val	Arg	Lys	Val	Ala	Arg	Asn	Leu	Gly	Lys	
			115				120					125					
35	gcc	act	tca	ggt	gtc	ctc	gtt	gtg	ctg	gat	gta	gtc	aac	ctt	gtg	caa	917
	Ala	Thr	Ser	Gly	Val	Leu	Val	Val	Leu	Asp	Val	Val	Asn	Leu	Val	Gln	
			130				135				140						
	gac	tca	ctg	gac	ttg	cac	aag	ggg	gaa	aaa	tcc	gag	tct	gct	gag	ttg	965
	Asp	Ser	Leu	Asp	Leu	His	Lys	Gly	Glu	Lys	Ser	Glu	Ser	Ala	Glu	Leu	
			145			150				155				160			
40	ctg	agg	cag	tgg	gct	cag	gag	ctg	gag	gag	aat	ctc	aat	gag	ctc	acc	1013
	Leu	Arg	Gln	Trp	Ala	Gln	Glu	Leu	Glu	Glu	Asn	Leu	Asn	Glu	Leu	Thr	
				165			170				175						
	cat	atc	cat	cag	agt	cta	aaa	gca	ggc	taggccaat	tggtgcggga						1060
45	His	Ile	His	Gln	Ser	Leu	Lys	Ala	Gly								
			180				185										
	agtcaggac	cccaaacgga	gggactggct	gaagccatgg	cagaagaacg	tggtattgtga											1120
	agatttcacg	gacattttatt	agttccccaa	attaatactt	ttataatttc	ctatgcctgt											1180
	ctttaccgca	atctctaaac	acaaattgtg	aagatttcac	ggacacttat	cacttcccca											1240
	atcaataccc	ttgtgatttc	ttatgcctgt	ctttacttta	atctectaat	cctgtcagct											1300
50	gaggagggtg	tatgtcacct	caggaccatg	tgatmattgc	gttaactgca	caaattgtag											1360
	agcatgtgtg	tttgaacaat	atgaaatctg	ggcaccttga	aaaaagaaca	ggataacagc											1420
	aatcgttcag	gggataagag	agataacctt	aaactctgac	caacagttag	ccgggtggag											1480
	cagagtcata	ttctttttct	ttcaaaagca	aatgggagaa	atatcgctga	attctttttc											1540
	tcagcaagga	acatccctga	gaaagagaat	gcaccctga	gggtgggtct	ataaatggcc											1600
55	tccttggtg	tggtcatctt	ctatggtcga	gactgtagg	atgaaataaa	ccccagtctc											1660
	ccatagtgtc	cccaggctta	ttagggaagag	gaaattcccg	cctaataaat	tttgggtcaga											1720
	ccggttgctc	tcaaaacctt	gtctcctgat	aagatgttat	caatgacaat	ggtgcctgaa											1780
	acctcattag	caattttta	ttctccctcg	tctgtgtgtc	ctgtgatctc	acctgcctc											1840
	cacttgccct	gtgatattct	attaccttgt	gaagtaggtg	atctttgtga	cccacacctt											1900
60	attcatcac	ccccccct	tttgggaagtc	cctaataaaa	acttgctggt	tttgcagctt											1960
	gtgagggcatc	acggaaacct	ctgatgtgtg	atgtctcccc	tggacacct	gcttttaaat											2020
	ttcaaaaaaa	aaaaaa															2036

65 <210> 146
 <211> 255
 <212> PRT

ES 2 330 903 T3

<213> *Homo sapiens*

<220>

<221> SEÑAL

5 <222> -70..-1

<400> 146

```

10      Met Gln Gln Lys Glu Gln Gln Phe Arg Glu Trp Phe Leu Lys Glu Phe
      -70                      -65                      -60                      -55
      Pro Gln Ile Arg Trp Lys Ile Gln Glu Ser Ile Glu Arg Leu Arg Val
      -50                      -45                      -40
      Ile Ala Asn Glu Ile Glu Lys Val His Arg Gly Cys Val Ile Ala Asn
15      -35                      -30                      -25
      Val Val Ser Gly Ser Thr Gly Ile Leu Ser Val Ile Gly Val Met Leu
      -20                      -15                      -10
      Ala Pro Phe Thr Ala Gly Leu Ser Leu Ser Ile Thr Ala Ala Gly Val
      -5                      1                      5                      10
20      Gly Leu Gly Ile Ala Ser Ala Thr Ala Gly Ile Ala Ser Ser Ile Val
      15                      20                      25
      Glu Asn Thr Tyr Thr Arg Ser Ala Glu Leu Thr Ala Ser Arg Leu Thr
      30                      35                      40
      Ala Thr Ser Thr Asp Gln Leu Glu Ala Leu Arg Asp Ile Leu His Asp
25      45                      50                      55
      Ile Thr Pro Asn Val Leu Ser Phe Ala Leu Asp Phe Asp Glu Ala Thr

      60                      65                      70
30      Lys Met Ile Ala Asn Asp Val His Thr Leu Arg Arg Ser Lys Ala Thr
      75                      80                      85                      90
      Val Gly Arg Pro Leu Ile Ala Trp Arg Tyr Val Pro Ile Asn Val Val
      95                      100                      105
      Glu Thr Leu Arg Thr Arg Gly Ala Pro Thr Arg Ile Val Arg Lys Val
35      110                      115                      120
      Ala Arg Asn Leu Gly Lys Ala Thr Ser Gly Val Leu Val Val Leu Asp
      125                      130                      135
      Val Val Asn Leu Val Gln Asp Ser Leu Asp Leu His Lys Gly Glu Lys
40      140                      145                      150
      Ser Glu Ser Ala Glu Leu Leu Arg Gln Trp Ala Gln Glu Leu Glu Glu
      155                      160                      165                      170
      Asn Leu Asn Glu Leu Thr His Ile His Gln Ser Leu Lys Ala Gly
      175                      180                      185

```

45

50

55

60

65