(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 266 693**

(51) Int. Cl.:
C07K 14/47 (2006.01)
C12N 15/12 (2006.01)
C12N 15/62 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **03016497 .4**
(86) Fecha de presentación : **31.07.1998**
(87) Número de publicación de la solicitud: **1375514**
(87) Fecha de publicación de la solicitud: **02.01.2004**

(54) Título: **ESTs 5' para proteínas secretadas expresadas en varios tejidos.**

(30) Prioridad: **01.08.1997 US 905051**

(45) Fecha de publicación de la mención BOPI:
01.03.2007

(45) Fecha de la publicación del folleto de la patente:
01.03.2007

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

(72) Inventor/es:
Dumas Milne Edwards, Jean-Baptiste;
Duclert, Aymeric y
Lacroix, Bruno

(74) Agente: **Elzaburu Márquez, Alberto**

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

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

ESTs 5' para proteínas secretadas expresadas en varios tejidos.

Antecedentes de la invención

Los 50.000-100.000 genes estimados, dispersos a lo largo de los cromosomas humanos suponen una enorme esperanza para la comprensión, el diagnóstico y para el tratamiento de las enfermedades humanas. Además, las sondas capaces de hibridarse específicamente con los loci distribuidos por el genoma humano tienen aplicaciones en la construcción de mapas cromosómicos de alta resolución y en la identificación de individuos.

En el pasado, la caracterización de siquiera un solo gen humano, era un procedimiento meticuloso que requería años de esfuerzo. Recientes progresos en las áreas de clonación de vectores, secuenciación de ADN y tecnología informática se han combinado para acelerar en gran medida la velocidad con la que se pueden aislar, secuenciar, cartografiar y caracterizar los genes humanos. Los vectores de clonación, tales como los cromosomas artificiales de levadura (YACs) y los cromosomas artificiales de bacterias (BACs) son capaces de aceptar insertos de ADN que oscilan desde 300 a 1000 kilobases (kb) o 100-400 kb de longitud, respectivamente, facilitando de este modo la manipulación y la disposición de secuencias de ADN distribuidas a lo largo de grandes distancias sobre los cromosomas humanos. Los aparatos automáticos para la secuenciación del ADN permiten la secuenciación rápida de genes humanos. Los programas bioinformáticos permiten la comparación de secuencias de ácidos nucleicos y proteínas, contribuyendo de este modo a la caracterización de los productos génicos humanos.

Generalmente, se han seguido dos planteamientos distintos para identificar y caracterizar los genes distribuidos a lo largo del genoma humano. En un planteamiento, se aíslan grandes fragmentos de ADN genómico, se clonan y se secuencian. Los marcos de lectura abiertos potenciales en estas secuencias genómicas se identifican empleando programas bioinformáticos. Sin embargo, este planteamiento supone la secuenciación de largos segmentos de ADN humano que no codifican proteínas, para encontrar las secuencias que codifican proteínas dispersas por el genoma. Además de requerir una secuenciación extendida, el programa bioinformático puede caracterizar erróneamente las secuencias genómicas obtenidas. Por tanto, el programa puede producir falsos positivos en los que ADN no codificante se caracteriza erróneamente como codificante o falsos negativos en los que el ADN codificante se marca erróneamente como ADN no codificante.

Un planteamiento alternativo toma una vía más directa para identificar y caracterizar genes humanos. En este planteamiento, se sintetizan ADNs complementarios (ADNc) a partir de ARNs mensajeros aislados (ARNm) que codifican proteínas humanas. Empleando este planteamiento, la secuenciación sólo se realiza sobre el ADN que se obtiene a partir de las partes del genoma que codifican proteínas. Frecuentemente, sólo se secuencian segmentos cortos de los ADNc para obtener secuencias denominadas etiquetas de secuencias expresadas (ESTs). Las ESTs se pueden utilizar entonces para aislar o purificar ADNc extendidos que incluyen secuencias adyacentes a las secuencias EST. Los ADNc extendidos pueden contener toda la secuencia de la EST que se había empleado para obtenerlos o sólo una parte de la secuencia de la EST que se había utilizado para obtenerlos. Además, los ADNc extendidos pueden contener la secuencia codificadora completa del gen a partir de la cual se había obtenido la EST o, alternativamente, los ADNc extendidos pueden incluir partes de la secuencia codificadora del gen a partir de la cual se había obtenido la EST. Se apreciará que puede haber diversos ADNc extendidos que incluyan la secuencia EST, debido al corte y empalme alterno o a la actividad de promotores alternativos.

En el pasado, estas secuencias ESTs cortas se obtenían frecuentemente a partir de genotecas de ADNc cebadas con oligo-dT. Por consiguiente, se correspondían principalmente con la región 3' no traducida del ARNm. En parte, la preponderancia de las secuencias ESTs obtenidas a partir del extremo 3' del ARNm es el resultado del hecho de que las técnicas típicas para obtener ADNc no están bien adaptadas para aislar secuencias de ADNc obtenidas a partir de los extremos 5' de los ARNm (Adams *y col., Nature* **377**:3-174, 1996; Hillier *y col., Genome Res.* **6**:807-828, 1996).

Además, en los ejemplos referidos en los que se han obtenido secuencias de ADNc más largas, las secuencias referidas se corresponden típicamente con secuencias codificadoras y no incluyen la región 5' completa, no traducida del ARNm a partir del cual se ha obtenido el ADNc. Tales secuencias incompletas pueden no incluir el primer exón del ARNm, particularmente en situaciones en las que el primer exón es corto. Adicionalmente, pueden no incluir tampoco algunos exones, frecuentemente los cortos, que están localizados aguas arriba de los sitios de corte y empalme. Por tanto, existe una necesidad de obtener secuencias obtenidas a partir de los extremos 5' de los ARNm.

Aunque muchas secuencias obtenidas a partir de cromosomas humanos tienen aplicaciones prácticas, planteamientos basados en la identificación y la caracterización de estas secuencias cromosómicas que codifican un producto proteico, tienen una importancia particular para los usos de diagnóstico y terapéuticos. Entre los 50.000-100.000 genes que codifican proteínas, los genes que codifican proteínas que se secretan desde la célula en la que se sintetizan, así como las proteínas secretadas mismas, son particularmente valiosos como agentes terapéuticos potenciales. Tales proteínas están implicadas frecuentemente en la comunicación de célula a célula y pueden ser responsables de producir una respuesta clínicamente relevante en sus células diana.

De hecho, algunas proteínas secretoras que incluyen el activador del plasminógeno tisular, G-CSF, GM-CSF, eritropoyetina, hormona de crecimiento humano, insulina, interferón-α, interferón-β, interferón-γ e interleuquina-2, tienen

normalmente uso clínico. Estas proteínas se emplean para tratar una amplia gama de estados que incluyen el infarto agudo de miocardio, el accidente isquémico agudo, la anemia, la diabetes, la deficiencia de hormona de crecimiento, la hepatitis, el carcinoma renal, la neutropenia inducida por quimioterapia y la esclerosis múltiple. Por estas razones, los ADNc extendidos que codifican proteínas secretadas o partes de las mismas, representan una fuente particularmente valiosa de agentes terapéuticos. Por tanto, existe una necesidad de identificar y caracterizar las proteínas secretadas y los ácidos nucleicos que las codifican.

Además de ser terapéuticamente útiles por sí mismas, las proteínas secretoras incluyen péptidos cortos, denominados péptidos señal, en sus extremos amino-terminales que dirigen su secreción. Estos péptidos señal son codificados por las secuencias señal localizadas en los extremos 5' de las secuencias codificadoras de los genes que codifican las proteínas secretadas. Debido a que estos péptidos señal dirigirán la secreción extracelular de cualquier proteína a la que están ligados funcionalmente, las secuencias señal se pueden utilizar para dirigir la secreción eficaz de cualquier proteína ligando funcionalmente las secuencias señal con un gen que codifica la proteína para la que se desea la secreción. Además, también se pueden utilizar partes de las secuencias señal para dirigir la importación intracelular de un péptido o una proteína de interés. Esto se puede mostrar beneficioso para las estrategias de terapia génica en las que se desea suministrar un producto génico en particular a células distintas a la célula en el que es producido. Las secuencias señal que codifican péptidos señal también se aplican para simplificar las técnicas de purificación de proteínas. En tales aplicaciones, la secreción extracelular de la proteína deseada facilita enormemente la purificación, reduciendo el número de proteínas no deseadas entre las que se tiene que seleccionar la proteína deseada. Por tanto, existe una necesidad de identificar y caracterizar las partes 5' de los genes de proteínas secretoras que codifican péptidos señal.

La información pública sobre el número de genes humanos en los que se ha identificado y caracterizado los promotores y las regiones reguladoras aguas arriba, es bastante limitada. En parte, puede ser debido a la dificultad de aislar tales secuencias reguladoras. Las secuencias reguladoras aguas arriba, tales como los sitios de unión a un factor de transcripción son típicamente demasiado cortas para utilizarlas como sondas para aislar promotores a partir de genotecas genómicas humanas. Recientemente, se han desarrollado algunos planteamientos para aislar promotores humanos. Uno de ellos consiste en preparar una genoteca de islas CpG (Cross *y col., Nature Genetics* **6**:236-244, 1994). El segundo consiste en aislar secuencias de ADN genómico humano que contienen sitios de unión a SpeI empleando la proteína de unión de SpeI (Mortlock *y col., Genome Res.* **6**:327-335, 1996). Ambos planteamientos tienen sus limites debido a una falta de especificidad o de detalle.

Las ESTs 5' presentes se pueden utilizar para identificar eficazmente y aislar regiones reguladoras aguas arriba que controlan la posición, el estado de desarrollo, la tasa y la cantidad de síntesis proteica, así como la estabilidad del ARNm. (Theil, *BioFactors* **4**:87-93, 1993). Una vez identificadas y caracterizadas, estas regiones reguladoras se pueden utilizar en la terapia génica o en los esquemas de purificación proteica para obtener la cantidad deseada y las localizaciones de la síntesis de proteínas o para inhibir, reducir o evitar la síntesis de productos génicos no deseables.

Además, las ESTs que contienen los extremos 5' de los genes de las proteínas secretoras pueden incluir secuencias útiles como sondas para cartografiar cromosomas y la identificación de individuos. Por tanto, existe una necesidad de identificar y caracterizar las secuencias aguas arriba de las secuencias 5' codificadoras de los genes que codifican proteínas secretoras.

Sumario de la invención

La presente invención se refiere a ESTs purificadas, aisladas o recombinantes que incluyen secuencias derivadas de los extremos 5' auténticos de sus ARNm correspondientes. La expresión "ARNm correspondiente" se refiere al ARNm que era el molde para la síntesis del ADNc que producía la EST 5'. Estas secuencias se denominarán de aquí en adelante "ESTs 5'". Tal y como se emplea en esta memoria, el término "purificado" no requiere una pureza absoluta sino que pretende ser una definición relativa. Los clones individuales de EST 5', aislados a partir de una genoteca de ADNc se han purificado convencionalmente para homogeneidad electroforética. Las secuencias obtenidas a partir de estos clones no se podrían obtener directamente a partir de la genoteca o del ADN humano total. Los clones de ADNc no están presentes en la naturaleza como tales, pero se obtienen por manipulación de una sustancia presente en la naturaleza, purificada parcialmente (ARN mensajero). La conversión del ARNm en una genoteca de ADNc implica la creación de una sustancia sintética (ADNc) y los clones de ADNc puros individuales se pueden aislar a partir de la genoteca sintética mediante selección de clones. Por tanto, la creación de una genoteca de ADNc a partir de un ARN mensajero y el aislamiento posterior de clones individuales a partir de esta genoteca, da como resultado la purificación de aproximadamente 10^4-10^6 veces del mensaje natural. La purificación del material de partida o el material natural hasta al menos un orden de magnitud, preferentemente dos o tres órdenes y lo más preferido, cuatro o cinco órdenes de magnitud, se contempla expresamente.

Tal y como se emplea en esta memoria, el término "aislado" requiere que el material se retire de su medio original (p. ej., el medio natural si está presente en la naturaleza). Por ejemplo, un polinucleótido presente en la naturaleza, en un animal vivo no se aísla, pero el mismo polinucleótido, separado de alguno o de todos los materiales coexistentes en el sistema natural, se aísla.

Tal y como se emplea en esta memoria, el término "recombinante" significa que la EST 5' es adyacente al ácido nucleico del "esqueleto", junto al que no es adyacente en su medio natural. Adicionalmente, para estar "enriquecidas", las ESTs 5' representarán 5% o más del número de insertos de ácido nucleico en una población de moléculas del

esqueleto de ácido nucleico. Las moléculas del esqueleto de acuerdo con la presente invención, incluyen ácidos nucleicos tales como vectores de expresión, ácidos nucleicos autorreplicativos, virus, ácidos nucleicos integrantes y otros vectores o ácidos nucleicos empleados para mantener o manipular un inserto de ácido nucleico de interés. Preferentemente, las ESTs 5' enriquecidas representan el 15% o más del número de insertos de ácido nucleico en la población de moléculas recombinantes del esqueleto. Más preferentemente, las ESTs 5' enriquecidas representan el 50% o más del número de insertos de ácido nucleico en la población de moléculas recombinantes del esqueleto. En una realización muy preferida, las ESTs 5' enriquecidas representan el 90% o más del número de insertos de ácido nucleico en la población de moléculas recombinantes del esqueleto.

Las condiciones de hibridación "rigurosas", "moderadas" y "suaves", son como se definen en el Ejemplo 29.

A no ser que se indique de otro modo, una secuencia "complementaria" es totalmente complementaria.

Por consiguiente, las ESTs 5' en genotecas de ADNc en las que una o varias ESTs 5' forman hasta el 5% o más del número de insertos de ácido nucleico en las moléculas del esqueleto, son "ESTs 5' recombinantes enriquecidas", tal y como se ha definido en esta memoria. De forma similar, las ESTs 5' en una población de plásmidos en la que una o varias ESTs 5' de la presente invención se han insertado de modo que representen el 5% o más del número de insertos en el esqueleto del plásmido, son "ESTs 5' recombinantes enriquecidas" tal y como se ha definido en esta memoria. Sin embargo, las ESTs 5' en las genotecas de ADNc en las que las ESTs 5' constituyen menos del 5% del número de insertos de ácido nucleico en la población de moléculas del esqueleto, tales como genotecas en las que las moléculas del esqueleto que tienen un inserto de EST 5' son extremadamente raras, no son "ESTs 5' recombinantes enriquecidas".

En particular, la presente invención se refiere a ESTs 5' que se han obtenido a partir de genes que codifican proteínas secretadas. Tal y como se emplea en esta memoria, una proteína "secretada" es una que, cuando se expresa en una célula hospedadora adecuada, se transporta a través o de lado a lado de una membrana, incluyendo el transporte como consecuencia de los péptidos señal en su secuencia de aminoácidos. Las proteínas "secretadas" incluyen sin limitación, las proteínas secretadas de forma completa (p. ej., proteínas solubles) o parcial (p. ej., receptores) a partir de la célula en la que se expresan. Las proteínas "secretadas" también incluyen sin limitación, las proteínas que se transportan a través de la membrana del retículo endoplásmico.

Tales ESTs 5' incluyen secuencias de ácidos nucleicos, denominadas secuencias señal, que codifican los péptidos señal que dirigen la secreción extracelular de las proteínas codificadas por los genes a partir de los cuales se obtienen las ESTs 5'. Generalmente, los péptidos señal se sitúan en el extremo amino terminal de las proteínas secretadas.

Las proteínas secretadas se traducen en ribosomas asociados con el retículo endoplásmico "rugoso". Generalmente, las proteínas secretadas se transfieren de forma simultánea a la traducción a la membrana del retículo endoplásmico. La asociación del ribosoma con el retículo endoplásmico durante la traducción de las proteínas secretadas, está mediada por el péptido señal. El péptido señal se corta típicamente después de su entrada de forma simultánea a la traducción en el retículo endoplásmico. Después de ser entregadas en el retículo endoplásmico, las proteínas secretadas pueden avanzar a través del aparato de Golgi. En el aparato de Golgi, las proteínas pueden sufrir una modificación post-traduccional antes de entrar en las vesículas secretoras que las transportan a través de la membrana celular.

Las ESTs 5' descritas en esta memoria tienen diversas aplicaciones importantes. Por ejemplo, se pueden utilizar para obtener y expresar clones de ADNc que incluyen las secuencias codificadoras de la proteína completa de los productos génicos correspondientes, que incluyen los sitios auténticos de inicio de la traducción obtenidos a partir de los extremos 5' de las secuencias codificadoras de los ARNm a partir de los cuales se obtienen las ESTs 5'. Estos ADNc se denominan de aquí en adelante "ADNc de longitud completa". Estos ADNc también pueden incluir ADN obtenido a partir de secuencias de ARNm aguas arriba del sitio de inicio de la traducción. Las secuencias de ADNc de longitud completa se pueden utilizar para expresar las proteínas que se corresponden con las ESTs 5'. Tal y como se ha expuesto anteriormente, las proteínas secretadas son terapéuticamente importantes. Por ello, las proteínas expresadas procedentes de los ADNc pueden ser útiles para tratar o controlar una variedad de enfermedades humanas. Las ESTs 5' también se pueden utilizar para obtener el ADN genómico correspondiente. La expresión "ADN genómico correspondiente" se refiere al ADN genómico que codifica el ARNm a partir del cual se había obtenido la EST 5'.

Alternativamente, las ESTs 5' se pueden utilizar para obtener y expresar ADNc extendidos que codifican partes de la proteína secretada. Las partes pueden comprender los péptidos señal de las proteínas secretadas o las proteínas maduras generadas cuando el péptido señal se elimina por corte. Las partes pueden comprender también polipéptidos que tienen al menos 10 aminoácidos consecutivos, codificados por los ADNc extendidos o los ADNc de longitud completa. Alternativamente, las partes pueden comprender al menos 15 aminoácidos consecutivos codificados por los ADNc extendidos o los ADNc de longitud completa. En algunas realizaciones, las partes pueden comprender al menos 25 aminoácidos consecutivos codificados por los ADNc extendidos o los ADNc de longitud completa. En otras realizaciones, las partes pueden comprender al menos 40 aminoácidos codificados por los ADNc extendidos o los ADNc de longitud completa.

También se pueden obtener tal y como se describe a continuación, anticuerpos que reconocen específicamente las proteínas secretadas completas, codificadas por los ADNc extendidos, los ADNc de longitud completa 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. Los anticuerpos que reconocen específicamente la proteína madura generadas cuando el péptido señal se separa por corte, también se pueden obtener tal y como se describe a continuación. De forma similar, también se pueden obtener los anticuerpos que reconocen específicamente los péptidos señal codificados por los ADNc extendidos o los ADNc de longitud completa.

En algunas realizaciones, los ADNc extendidos obtenidos empleando los ESTs 5' incluyen la secuencia señal. En otras realizaciones, los ADNc extendidos obtenidos empleando las ESTs 5' pueden incluir la secuencia codificadora completa para la proteína madura (es decir, la proteína generada cuando el polipéptido señal se elimina por corte). Además, los ADNc extendidos obtenidos empleando las ESTs 5' pueden incluir regiones reguladoras aguas arriba del sitio de inicio de la traducción o aguas abajo del codón de detención que controlan la cantidad, la posición o el estado de desarrollo de la expresión génica.

Tal y como se ha expuesto anteriormente, las proteínas secretadas son terapéuticamente importantes. Por ello, las proteínas expresadas a partir de los ADNc extendidos o los ADNc de longitud completa, obtenidos empleando las ESTs 5', pueden ser útiles para tratar o controlar una variedad de enfermedades humanas.

Las ESTs 5' (o los ADNc o los ADNs genómicos obtenidos a partir de las mismas) se pueden utilizar en procedimientos forenses para identificar individuos o en procedimientos de diagnóstico para identificar individuos que tienen enfermedades genéticas como consecuencia de la expresión anormal de los genes correspondientes a las ESTs 5'. Además, la presente invención es útil para construir un mapa de alta resolución de los cromosomas humanos.

La presente invención también se refiere a los vectores de secreción capaces de dirigir la secreción de una proteína de interés. Tales vectores se pueden utilizar en las estrategias de terapia génica en las que se desea producir un producto génico en una célula que se va a suministrar a otra ubicación en el cuerpo. Los vectores de secreción también pueden facilitar la purificación de proteínas deseadas.

La presente invención también se refiere a vectores de expresión capaces de dirigir la expresión de un gen insertado en una forma espacial o temporal deseada o a un nivel deseado. Tales vectores pueden incluir secuencias aguas arriba de las ESTs 5', tales como promotores o secuencias reguladoras aguas arriba.

Finalmente, la presente invención también se puede utilizar en la terapia génica para controlar o tratar enfermedades genéticas. Los péptidos señal también se pueden fusionar con proteínas heterólogas para dirigir su secreción extracelular.

Los clones bacterianos que contienen plásmidos Bluescript que tienen insertos que contienen la EST 5' (SEQ ID NO: 38), están presentes almacenados a -80ºC en 4% (v/v) de glicerol en los laboratorios de los inventores, bajo la denominación enumerada junto a la SEQ ID NO en II. Los insertos se pueden recuperar de los materiales depositados, haciendo crecer los clones adecuados en un medio adecuado. El ADN de Bluescript se puede aislar entonces, empleando procedimientos para aislar plásmidos, familiares para los expertos en la técnica, tales como minipreparados para lisis alcalina o procedimientos para aislar plásmidos por lisis alcalina a gran escala. Si se desea, el ADN del plásmido se puede enriquecer adicionalmente por centrifugación en un gradiente de cloruro de cesio, cromatografía por exclusión por tamaños o cromatografía de intercambio aniónico. El ADN plasmídico obtenido empleando estos procedimientos se puede manipular a continuación empleando técnicas de clonación convencionales, familiares para los expertos en la técnica. Alternativamente, se puede realizar una PCR con los cebadores diseñados en ambos extremos de la inserción de EST. El producto de la PCR que se corresponde con la EST 5', se puede manipular a continuación empleando técnicas de clonación convencionales, familiares para los expertos en la técnica.

Un aspecto de la presente invención es un ácido nucleico purificado o aislado que tiene la secuencia de SEQ ID NO: 38 o que tiene una secuencia complementaria a la misma. En una realización, el ácido nucleico es recombinante.

Otro aspecto de la presente invención es un péptido señal que tiene la secuencia de aminoácidos -26 a -1 de SEQ ID NO: 39.

Aún otro aspecto de la presente invención es una secuencia señal que codifica un péptido señal de la invención. Preferentemente, la secuencia señal tiene la secuencia de los nucleótidos 170 a 247 de SEQ ID NO: 38.

Aún otro aspecto de la presente invención es un método para preparar un ADNc que codifica una proteína secretora humana, comprendiendo las etapas de poner en contacto una colección de moléculas de ARNm procedentes de células humanas, con un cebador que comprende al menos 15 nucleótidos consecutivos de una secuencia complementaria a SEQ ID NO: 38; hibridándose dicho cebador bajo condiciones rigurosas, con un ARNm en dicha colección que codifica dicha proteína; transcribiéndose de forma inversa dicho cebador hibridado para formar una primera hebra de ADNc a partir de dicho ARNm; preparando una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc; y aislando el ADNc resultante que codifica dicha proteína que contiene dicha primera hebra de ADNc y dicha segunda hebra de ADNc.

Otro aspecto de la presente invención es un método para preparar un ADNc que codifica una proteína secretora humana que comprende las etapas de proporcionar un ADNc obtenido a partir de un ARNm de un tejido, de una célula o de un organismo de interés, poner en contacto dicho ADNc con una sonda detectable que comprende al menos 15

nucleótidos consecutivos de dicha secuencia de SEQ ID NO: 38 o una secuencia complementaria a la misma, bajo condiciones que permiten a dicha sonda hibridarse con dicho ADNc; identificar un ADNc que se hibrida con dicha sonda detectable y aislar dicho ADNc que se hibrida con dicha sonda.

Otro aspecto de la presente invención es un método para preparar un ADNc que comprende al menos 15 nucleótidos de la secuencia de SEQ ID NO: 38, que comprende las etapas de poner en contacto una colección de moléculas de ARNm procedentes de células humanas con un primer cebador capaz de hibridarse con la cola poliA de dicho ARNm; hibridar dicho primer cebador con dicha cola poliA; transcribir de forma inversa dicho ARNm para formar una primera hebra de ADNc; preparar una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc empleando al menos un cebador que comprende al menos 15 nucleótidos de la secuencia de SEQ ID NO: 38; y aislar el ADNc resultante que comprende dicha primera hebra de ADNc y dicha segunda hebra de ADNc.

Otro aspecto de la presente invención es un método para preparar una proteína secretada que comprende la etapa de insertar un gen que codifica una proteína no secretada en marco de lectura con una secuencia señal de la invención en un vector, de modo que la proteína codificada por el gen insertado se expresa a partir del ARNm transcrito. Preferentemente, el método comprende adicionalmente la etapa de introducir dicho vector en una célula, un tejido o un organismo hospedador.

Otro aspecto de la presente invención es un vector que comprende una secuencia señal de la invención, ligado funcionalmente a un promotor. Preferentemente, el vector es un vector de secreción o un vector para terapia génica.

Otro aspecto de la presente invención es un ácido nucleico purificado y aislado que codifica un polipéptido que comprende el péptido señal de la presente invención en el extremo 5' de la secuencia codificante y un extremo amino que comprende los aminoácidos 1 a 29 de SEQ ID NO: 39. En una realización, el ácido nucleico comprende una secuencia señal que tiene la secuencia de nucleótidos 170 a 247 de SEQ ID NO: 38. Preferentemente, el ácido nucleico comprende la secuencia señal mencionada anteriormente y el extremo amino del polipéptido es codificado por los nucleótidos 170 a 334 de SEQ ID NO: 38.

Otro aspecto de la invención es un ácido nucleico purificado y aislado que comprende la secuencia señal de la invención, fusionada en marco de lectura con el extremo 5' de una secuencia que codifica una parte de un polipéptido que comprende al menos 25 aminoácidos que es heteróloga de un polipéptido que comprende los aminoácidos 1 a 29 de SEQ ID NO: 39.

Otro aspecto de la presente invención es un vector que comprende los ácidos nucleicos de la invención ligados funcionalmente a un promotor. Otro aspecto de la invención es un polipéptido codificado por el polinucleótido de los dos párrafos anteriores. Preferentemente, el polipéptido es una proteína humana secretada. En una realización adicional, la proteína humana secretada comprende los aminoácidos 1 a 29 de SEQ ID NO: 39.

Otro aspecto de la presente invención es el uso del péptido señal de la invención o el vector de la invención, para dirigir la secreción extracelular de un polipéptido o para simplificar la purificación de proteínas de un polipéptido deseado. Preferentemente, dicho polipéptido es un polipéptido de la presente invención.

Otro aspecto de la invención es una proteína de fusión codificada por un ácido nucleico de la invención.

Otro aspecto de la presente invención es un método para preparar una proteína secretada que comprende las etapas de: introducir un vector de la invención en una célula, un tejido o un organismo hospedador. Preferentemente, los métodos de la invención comprenden adicionalmente la etapa de: aislar la proteína secretada. Preferentemente, la etapa de aislar la proteína secretada comprende purificar la proteína secretada a partir del material sobrenadante, del medio de cultivo o del extracto celular de dicha célula hospedadora.

Otro aspecto de la presente invención es un polipéptido aislado, obtenible por el método descrito anteriormente.

Breve descripción de los dibujos

La Figura 1 es un resumen de un procedimiento para obtener los ADNc que se han seleccionado para incluir los extremos 5' de los ARNm a partir de los cuales se han obtenido.

La Figura 2 muestra la distribución de las puntuaciones de von Heijne para las ESTs 5' en cada una de las categorías descritas en esta memoria y la probabilidad de que estas ESTs 5' codifiquen un péptido señal.

La Figura 3 resume un método general empleado para clonar y secuenciar los ADNc extendidos que contienen secuencias adyacentes a las ESTs 5'.

La Figura 4 (descripción de la estructura de los promotores aislados a partir de las etiquetas señal ESTs 5') proporciona una descripción esquemática de los promotores aislados y del modo en que se ensamblan con las etiquetas 5' correspondientes.

Descripción detallada de la realización preferida

La Tabla IV es un análisis de los 43 aminoácidos localizados en el extremo N-terminal de todas las proteínas SwissProt humanas, para determinar la frecuencia de falsos positivos y de falsos negativos empleando las técnicas para la identificación de péptidos señal, descritas en esta memoria.

La Tabla V muestra la distribución de las ESTs 5' en cada categoría descrita en esta memoria y el número de ESTs 5' en cada categoría que tienen una puntuación mínima dada de von Heijne.

La Tabla VI muestra la distribución de las ESTs 5' en cada categoría descrita en esta memoria en relación con el tejido a partir del cual se habían obtenido las ESTs 5' de los ARNm correspondientes.

La Tabla VII describe los sitios de unión al factor de transcripción presentes en cada uno de estos promotores.

I. *Métodos Generales para obtener ESTs 5' obtenidas a partir de ARNm con extremos 5' intactos*

Para obtener las ESTs 5' descritas en esta memoria, se tienen que obtener ARNm con los extremos 5' intactos. Generalmente, existen dos planteamientos para obtener tales ARNm con los extremos 5' intactos, tal y como se describe a continuación: químico (1) o enzimático (2).

1. *Métodos Químicos para obtener ARNm que tienen los extremos 5' intactos*

Uno de estos planteamientos es un método de modificación química que implica derivatizar los extremos 5' de los ARNm y seleccionar los ARNm derivatizados. Los extremos 5' de los ARNm eucarióticos poseen una estructura denominada "casquete" que comprende una guanosina metilada en la posición 7. El casquete se une a la primera base transcrita del ARNm mediante un enlace 5',5'-trifosfato. En algunos casos, la guanosina 5' está metilada en un ambas posiciones 2 y 7. A veces, la guanosina 5' está trimetilada en las posiciones 2, 7 y 7. En el método químico para obtener ARNm que tienen los extremos 5' intactos, el casquete 5' se derivatiza específicamente y se acopla a un grupo reactivo sobre un sustrato inmovilizante. Esta derivatización específica se basa en el hecho de que sólo la ribosa unida a la guanosina metilada en el extremo 5' del ARNm y la ribosa unida a la base del extremo 3' del ARNm, poseen 2',3'-cis dioles.

Opcionalmente, el 2',3'-cis diol de la ribosa del extremo 3'-terminal se puede modificar, sustituir, convertir o eliminar químicamente, dejando sólo la ribosa unida a la guanosina metilada en el extremo 5' del ARNm con un 2',3'-cis diol. Una variedad de técnicas están a disposición para eliminar el 2',3'-cis diol en la ribosa 3'-terminal. Por ejemplo, la hidrólisis alcalina controlada se puede emplear para generar fragmentos de ARNm en los que la ribosa 3'-terminal es un 3'-fosfato, 2'-fosfato o (2',3')-ciclofosfato. A continuación, el fragmento que incluye la ribosa 3' original se puede eliminar de la mezcla mediante cromatografía sobre una columna de oligodT. Alternativamente, una base que carece del 2',3'-cis diol se puede añadir al extremo 3' del ARNm empleando una ligasa de ARN, tal como la ligasa de ARN de T4. El Ejemplo 1 a continuación describe un método para la ligación de un nucleósido difosfato con el extremo 3' del ARN mensajero.

Ejemplo 1

Ligación del nucleósido difosfato pCp al extremo 3' del ARNm

Un μg de ARN se incubó en un medio de reacción final de 10 μl en presencia de 5 U de la ligasa de ARN del fago T_4, en el tampón proporcionado por el fabricante (Gibco-BRL), 40 U del inhibidor de ARNasa, RNasin (Promega) y 2 μl de ^{32}pCp (Amersham nº PB 10208). La incubación se realizó a 37ºC durante 2 horas o durante una noche a 7-8°C.

Después de la modificación o la eliminación del 2',3'-cis diol en la ribosa 3', el 2',3'-cis diol presente en el extremo 5' del ARNm se puede oxidar empleando reactivos tales como $NaBH_4$, $NaBH_3CN$ o peryodato sódico, convirtiendo de este modo 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 sódico.

Ejemplo 2

Oxidación del 2',3'-cis diol en el extremo 5' del ARNm con peryodato sódico

Se trataron 0,1 unidades de DO de un oligorribonucleótido con casquete de 47 nucleótidos (incluyendo el casquete) o un oligorribonucleótido sin casquete de 46 nucleótidos, del modo siguiente. Los oligorribonucleótidos se produjeron mediante transcripción *in vitro* empleando el equipo de reactivos para la transcripción "AmpliScribe T7" (Epicentre Technologies). Tal y como se indica a continuación, el molde de ADN para el transcrito de ARN contenía una citosina aislada. Para sintetizar el ARN sin casquete, se incluyeron los cuatro NTPs en la reacción de transcripción *in vitro*. Para obtener el ARN con casquete, se sustituyó GTP por un análogo del casquete, m7G(5')ppp(5')G. Este compuesto, reconocido por la polimerasa, se incorporaba en el extremo 5' del transcrito naciente durante la iniciación de la transcripción, pero no se incorporaba durante la etapa de extensión. Por tanto, el ARN resultante contenía un casquete en su extremo 5'. Las secuencias de los oligorribonucleótidos producidos por la reacción de transcripción *in vitro* eran:

+Cap:

5'm7GpppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEQ ID NO: 1)

-Cap:

5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEQ ID NO: 2)

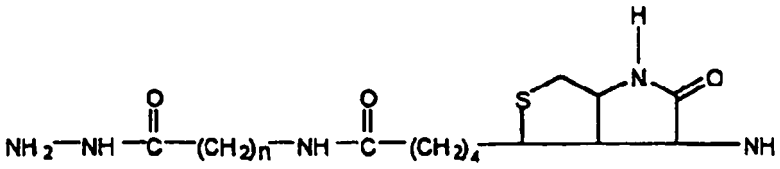
Los oligorribonucleótidos se disolvieron en 9 μl de tampón acetato (acetato sódico 0,1 M, pH 5,2) y 3 μl de solución de peryodato sódico 0,1 M preparada recientemente. La mezcla se incubó durante 1 hora en oscuridad a 4ºC o a temperatura ambiente. A continuación, se detuvo la reacción añadiendo 4 μl de etilenglicol al 10%. El producto se precipitó con etanol, se resuspendió en al menos 10 μl de agua o tampón adecuado y se dializó de nuevo frente a agua.

Los grupos aldehído resultantes se acoplaron después a moléculas que tenían un grupo amino reactivo, tal como hidrazina, carbazida, tiocarbazida o semicarbazida, para facilitar el enriquecimiento de los extremos 5' de los ARNm. Las moléculas que tenían grupos amino reactivos que eran adecuadas para emplear en la selección de ARNm que tenían extremos 5' intactos, incluyen avidina, proteínas, anticuerpos, vitaminas, ligandos capaces de unirse específicamente a moléculas receptoras u oligonucleótidos. El Ejemplo 3 a continuación, describe el acoplamiento del dialdehído resultante con la biotina.

Ejemplo 3

Acoplamiento del dialdehído en el extremo 5' de los transcritos con biotina

El producto de la oxidación obtenido en el Ejemplo 2, se disolvió en 50 μl de acetato sódico con un pH entre 5 y 5,2 y 50 μl de solución de nuevo aporte de biotín-hidrazida 0,02 M en una mezcla de metoxietanol/agua (1:1) de fórmula:



En el compuesto empleado en estos experimentos, n = 5. Sin embargo, se apreciará que otras hidrazidas a disposición comercial, también se pueden utilizar, tales como moléculas de la fórmula anterior en la que n varía entre 0 y 5. La mezcla se incubó a continuación durante 2 horas a 37ºC, se precipitó con etanol y se dializó frente a agua destilada. El Ejemplo 4 muestra la especificidad de la reacción de biotinilación.

Ejemplo 4

Especificidad de la biotinilación de los transcritos con casquete

La especificidad de la biotinilación para los ARNm con casquete se evaluó mediante electroforesis en gel de las siguientes muestras:

Muestra 1. El transcrito *in vitro* de 46 nucleótidos sin casquete, preparado como en el Ejemplo 2 y marcado con ^{32}pCp tal y como se ha descrito en el Ejemplo 1.

Muestra 2. El transcrito *in vitro* de 46 nucleótidos sin casquete, preparado como en el Ejemplo 2 y marcado con ^{32}pCp tal y como se ha descrito 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. El transcrito *in vitro* de 47 nucleótidos con casquete, preparado como en el Ejemplo 2 y marcado con ^{32}pCp tal y como se ha descrito en el Ejemplo 1.

Muestra 4. El transcrito *in vitro* de 47 nucleótidos con casquete, preparado como en el Ejemplo 2, marcado con ^{32}pCp tal y como se ha descrito en el Ejemplo 1, tratado con la reacción de oxidación del Ejemplo 2 y sometido a las condiciones de biotinilación del Ejemplo 3.

Las muestras 1 y 2 tenían las tasas de migración idénticas, mostrando que los ARNs sin casquete no estaban oxidados ni biotinilados. La muestra 3 migraba más lentamente que las muestras 1 y 2, mientras que la muestra 4 mostraba la migración más lenta. Las diferencias en la migración de los ARNs en las muestras 3 y 4 mostraban que los ARNs con casquete estaban específicamente biotinilados.

En algunos casos, los ARNm que tenían los extremos 5' intactos, se podían enriquecer uniendo la molécula que contenía un grupo amino reactivo a un sustrato adecuado en fase sólida, tal como la parte interior del recipiente que contenía los ARNm, perlas magnéticas, matrices de cromatografía o membranas de nailon o de nitrocelulosa. Por ejemplo, cuando la molécula que tiene un grupo amino reactivo es biotina, el sustrato en fase sólida se puede acoplar a avidina o a estreptavidina. Alternativamente, cuando la molécula que tiene el grupo amino reactivo es un anticuerpo o un ligando de receptor, el sustrato en fase sólida se puede acoplar al antígeno o al receptor relacionado-afín. Finalmente, cuando la molécula que tiene un grupo amino reactivo comprende un oligonucleótido, el sustrato en fase sólida puede comprender un oligonucleótido complementario.

Los ARNm que tienen extremos 5' intactos se pueden liberar de la fase sólida siguiendo el procedimiento de enriquecimiento. Por ejemplo, cuando el dialdehído se acopla a biotín-hidrazida y la fase sólida comprende estreptavidina, los ARNm se pueden liberar de la fase sólida, simplemente calentado a 95 grados Celsius en SDS al 2%. En algunos métodos, la molécula que tiene un grupo amino reactivo también se puede escindir de los ARNm que tienen los extremos 5' intactos, después del enriquecimiento. El Ejemplo 5 describe la captura de ARNm biotinilados con perlas revestidas de estreptavidina y la liberación de los ARNm biotinilados de las perlas después del enriquecimiento.

Ejemplo 5

Captura y liberación de ARNm biotinilados empleando perlas revestidas con estreptavidina

Las perlas magnéticas revestidas con estreptavidina se prepararon según las instrucciones del fabricante (CPG Inc., EE.UU.). Los ARNm biotinilados se añadieron 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. Las perlas se lavaron entonces varias veces en agua con SDS al 1%. Las perlas así obtenidas se incubaron durante 15 minutos a 95ºC en agua que contenía SDS al 2%.

El Ejemplo 6 muestra la eficacia con la que los ARNm biotinilados se recuperaban de las perlas revestidas con estreptavidina.

Ejemplo 6

Eficacia de la recuperación de los ARNm biotinilados

La eficacia del procedimiento de recuperación se evaluó del modo siguiente. Los ARNs con casquete se marcaron con ^{32}pCp, se oxidaron, se biotinilaron y se unieron a perlas revestidas con estreptavidina, tal y como se ha descrito anteriormente. Posteriormente, los ARNs unidos se incubaron durante 5, 15 o 30 minutos a 95ºC en presencia de SDS al 2%.

Los productos de la reacción se analizaron por electroforesis sobre geles de poliacrilamida al 12%, bajo condiciones desnaturalizantes (urea 7 M). Los geles se sometieron a autorradiografía. Durante esta manipulación, los enlaces de hidrazona no se reducían.

Se recuperaron cantidades crecientes de ácidos nucleicos cuando se incrementaban los tiempos de incubación en SDS al 2%, demostrando que los ARNm biotinilados se recuperaban con eficacia.

En un método alternativo para obtener los ARNm que tenían los extremos 5' intactos, un oligonucleótido que se había derivatizado para contener un grupo amino reactivo, se acopla específicamente a los ARNm que tienen un casquete intacto. Preferentemente, el extremo 3' del ARNm está bloqueado antes de la etapa en la que los grupos aldehído se unen al oligonucleótido derivatizado, tal y como se ha descrito anteriormente, de modo que se evita que el oligonucleótido derivatizado se una al extremo 3' del ARNm. Por ejemplo, pCp se puede fijar al extremo 3' del ARNm empleando la ligasa de ARN de T4, tal y como se ha descrito en el Ejemplo 1. Sin embargo, tal y como se ha descrito anteriormente, el bloqueo del extremo 3' del ARNm es una etapa opcional. Los oligonucleótidos derivatizados se pueden preparar tal y como se describe en el Ejemplo 7.

Ejemplo 7

Derivatización de oligonucleótidos

Un oligonucleótido fosforilado en su extremo 3' se convirtió en una hidrazida 3' en 3', mediante tratamiento con una solución acuosa de hidrazina o de dihidrazida de fórmula $H_2N(R1)NH_2$ aproximadamente 1 a 3 M y un pH 4,5 a una temperatura de 8ºC, durante una noche. Esta incubación se realizó en presencia de un agente de tipo carbodiimida soluble en agua, tal como 1-etil-3-(3-dimetilaminopropil)carbodiimida, con una concentración final de 0,3 M.

El oligonucleótido derivatizado se separó a continuación de los otros agentes y los productos empleando una técnica convencional para aislar oligonucleótidos.

Tal y como se ha expuesto anteriormente, los ARNm que se van a enriquecer se pueden tratar para eliminar los grupos 3' OH que puedan estar presentes en los mismos. Esto se puede realizar mediante ligación enzimática de las

secuencias que carecen de 3' OH, tal como pCp, tal y como se ha descrito en el Ejemplo 1. Alternativamente, los grupos 3' OH se pueden eliminar por hidrólisis alcalina, tal y como se describe en el Ejemplo 8 a continuación.

Ejemplo 8

Eliminación de los grupos 3' OH del ARNm empleando la hidrólisis alcalina

En un volumen total de 100 μl de hidróxido sódico 0,1 N, se incuban 1,5 μg de ARNm durante 40 a 60 minutos a 4ºC. La solución se neutraliza con ácido acético y se precipita con etanol.

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

Ejemplo 9

Oxidación de los dioles del ARNm

El ARN se disolvió hasta obtener 1 unidad de DO, en 9 μl de tampón (acetato sódico 0,1 M, pH 6-7) o agua y 3 μl de solución de nuevo aporte de peryodato sódico 0,1 M. La reacción se incubó durante 1 h en la oscuridad a 4ºC o a temperatura ambiente. Después de la incubación, la reacción se detuvo añadiendo 4 μl de etilenglicol al 10%. A continuación, la mezcla se incubó a temperatura ambiente durante 15 minutos. Después de la precipitación con etanol, el producto se resuspendió en al menos 10 μl de agua o en tampón adecuado y se dializó frente a agua.

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

Ejemplo 10

Ligación de aldehídos del ARNm a oligonucleótidos derivatizados

El ARNm oxidado se disolvió en un medio ácido, tal como 50 μl de acetato sódico pH 4-6. Cincuenta μl de una solución del oligonucleótido derivatizado se añadieron para obtener una proporción de ARNm:oligonucleótido derivatizado de 1:20. La mezcla se redujo con un borohidruro y se incubó durante 2 h a 37ºC o durante una noche (14 h) a 10ºC. La mezcla se precipitó a continuación con etanol, se resuspendió en 10 μl o más de agua o de tampón adecuado y se dializó frente a agua destilada. Si se desea, el producto resultante se puede analizar empleando electroforesis en gel de acrilamida, análisis con HPLC u otras técnicas convencionales.

Después de la fijación del oligonucleótido derivatizado a los ARNm, se puede realizar una reacción de transcripción inversa, tal y como se describe en el Ejemplo 11 a continuación.

Ejemplo 11

Transcripción inversa de los ARNm ligados a oligonucleótidos derivatizados

Un oligodesoxirribonucleótido se derivatizó del modo siguiente. Tres unidades de DO de un oligodesoxirribonucleótido de secuencia 5'ATCAAGAATTCGCACGAGACCATTA3' (SEQ ID NO:3) que tiene extremos 5'-OH y 3'-P, se disolvieron en 70 μl de una solución de hidroxibenzotriazol 1,5 M, pH 5,3, preparada en dimetilformamida/agua (75:25) que contenía 2 μg de 1-etil-3-(3-dimetilaminopropil)carbodiimida. La mezcla se incubó durante 2 h 30 min. a 22ºC y a continuación se precipitó dos veces en $LiClO_4$/acetona. El sedimento se resuspendió en 200 μl de hidrazina 0,25 M y se incubó a 8ºC de 3 a 14 h. Después de la reacción con hidrazina, la mezcla se precipitó dos veces en $LiClO_4$/acetona.

Los ARNs mensajeros que se iban a transcribir de forma inversa, se extrajeron de los bloques de placenta que tenían lados de 2 cm y que se habían almacenado a -80ºC. El ARN total se extrajo empleando técnicas convencionales con fenol ácido. La cromatografía de los oligo-dT se empleó para purificar los ARNm. La integridad de los ARNm se comprobó con transferencia de tipo Northern.

Los grupos diol en 7 μg de ARNm de placenta se oxidaron tal y como se ha descrito anteriormente en el Ejemplo 9. El oligonucleótido derivatizado se unió a los ARNm tal y como se ha descrito en el Ejemplo 10 anterior, exceptuando que la etapa de precipitación se sustituyó por una etapa de cromatografía por exclusión para eliminar los oligodesoxirribonucleótidos derivatizados que no se habían unido a los ARNm. La cromatografía por exclusión se realizó del modo siguiente:

Diez ml de gel Ultrogel AcA34 (BioSepra nº 230151), una mezcla de agarosa y acrilamida, se equilibraron en 50 ml de una solución de Tris 10 mM, pH 8,0, NaCl 300 mM, EDTA 1 mM y SDS al 0,05%. Se dejó que la mezcla sedimentara. El material sobrenadante se eliminó y el gel se resuspendió en 50 ml de tampón. Este procedimiento se repitió 2 o 3 veces.

Una perla de vidrio (diámetro de 3 mm) se introdujo en una pipeta desechable de 2 ml (longitud de 25 cm). La pipeta se llenó con la suspensión del gel hasta que la altura del gel se estabilizó a 1 cm de la parte superior de la pipeta. La columna se equilibró a continuación con 20 ml de tampón de equilibrio (Tris HCl 10 mM, pH 7,4, NaCl 20 mM).

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

La columna se cargó a continuación con los ARNm acoplados al oligonucleótido. En cuanto la muestra penetró, se añadió el tampón de equilibrio. Fracciones de 100 μl se recogieron a continuación. El oligonucleótido derivatizado que no se había fijado al ARNm aparecía en la fracción 16 y en fracciones posteriores. Por tanto, las fracciones 3 a 15 se combinaron y se precipitaron con etanol.

Para determinar si el oligonucleótido derivatizado estaba realmente ligado al ARNm, una décima parte de las fracciones combinadas se transfirieron dos veces en forma de manchas sobre una membrana de nailon y se hibridaron con una sonda radiactiva, empleando técnicas convencionales. La sonda marcada con ^{32}P se empleada en estas hibridaciones era un oligodesoxirribonucleótido de secuencia 5'TAATGGTCTCGTGCGAATTCTTGAT3' (SEQ ID NO:4) anticomplementaria al oligonucleótido derivatizado. Una señal observada después de la autorradiografía indicaba que el oligonucleótido derivatizado se había ligado verdaderamente al ARNm.

Las nueve décimas partes restantes de los ARNm que habían reaccionado con el oligonucleótido derivatizado, se transcribieron de forma inversa del modo siguiente. Una reacción de transcripción inversa se realizó con transcriptasa inversa, siguiendo las instrucciones del fabricante y 50 pmol de nonámeros con secuencias al azar como cebadores.

Para asegurar que se había realizado la transcripción inversa a través de la estructura del casquete, se realizaron dos tipos de experimentos.

En el primer planteamiento, después de eliminar el ARN de los heterodúplex de ADNc:ARN, obtenidos a partir de la reacción de transcripción inversa mediante una hidrólisis alcalina, una parte de los ADNc monocatenarios resultantes se transfirió en forma de manchas sobre una membrana cargada positivamente y se hibridó empleando métodos convencionales, con una sonda marcada con ^{32}P que tenía una secuencia idéntica a la del oligonucleótido derivatizado. Las manchas testigo que contenían 1 pmol, 100 fmol, 50 fmol, 10 fmol y 1 fmol de un oligodesoxirribonucleótido testigo de secuencia idéntica a la del oligonucleótido derivatizado, se incluyeron. La señal observada en las manchas que contenían el ADNc, indicaba que aproximadamente 15 fmol del oligonucleótido derivatizado se había transcrito de forma inversa. Estos resultados muestran que la transcripción inversa se puede realizar a través del casquete y, en particular, que la transcriptasa inversa cruza el enlace 5'-P-P-P-5' del casquete de los ARNs mensajeros eucariotas.

En el segundo tipo de experimento, los ADNc monocatenarios obtenidos a partir de la síntesis anterior de la primera hebra, se emplearon como molde para reacciones con la PCR. Se realizaron dos tipos de reacciones. Primero, la amplificación específica de los ARNm para alfa globina, deshidrogenasa, pp15 y el factor de elongación E4, se realizó empleando las siguientes parejas de cebadores de oligodesoxirribonucleótidos.

alfa-globina

GLO-S:	5'CCG ACA AGA CCA ACG TCA AGG CCG C3'	(SEQ ID NO:5)
GLO-As:	5'TCA CCA GCA GGC AGT GGC TTA GGA G 3'	(SEQ ID NO:6)

deshidrogenasa

3 DH-S:	5' AGT GAT TCC TGC TAC TTT GGA TGG C3'	(SEQ ID NO:7)
3 DH-As:	5'GCT TGG TCT TGT TCT GGA GTT TAG A3'	(SEQ ID NO:8)

pp15

PP15-S:	5'TCC AGA ATG GGA GAC AAG CCA ATT T3'	(SEQ ID NO:9)
PP15-As:	5'AGG GAG GAG GAA ACA GCG TGA GTC C3'	(SEQ ID NO:10)

factor de elongación E4

EFA1-S:	5' ATG GGA AAG GAA AAG ACT CAT ATC A3'	(SEQ ID NO:11)
EF1A-As:	5' AGC AGC AAC AAT CAG GAC AGC ACA G3'	(SEQ ID NO:12)

Segundo, las amplificaciones no específicas también se realizaron con los oligodesoxirribonucleótidos no codificantes de las parejas descritas anteriormente y con un cebador obtenido a partir de la secuencia del oligodesoxirribonucleótido derivatizado (5'ATCAAGAATTCGCACGAGACCATTA3') (SEQ ID NO:13).

Una veinteava parte de las siguientes muestras del producto de la RT-PCR se aplicaron a un gel de agarosa al 1,5% y se tiñó con bromuro de etidio.

Muestra 1: Los productos de una reacción de la PCR que empleaba cebadores de globina de las SEQ ID NOs 5 y 6 en presencia de ADNc.

Muestra 2: Los productos de una reacción de la PCR que empleaba cebadores de globina de las SEQ ID NOs 5 y 6 en ausencia de ADNc añadido.

Muestra 3: Los productos de una reacción de la PCR que empleaba cebadores de deshidrogenasa de las SEQ ID NOs 7 y 8 en presencia de ADNc.

Muestra 4: Los productos de una reacción de la PCR que empleaba cebadores de deshidrogenasa de las SEQ ID NOs 7 y 8 en ausencia de ADNc añadido.

Muestra 5: Los productos de una reacción de la PCR que empleaba cebadores de pp15 de las SEQ ID NOs 9 y 10 presencia de ADNc.

Muestra 6: Los productos de una reacción de la PCR que empleaba cebadores de pp15 de las SEQ ID NOs 9 y 10 en ausencia de ADNc añadido.

Muestra 7: Los productos de una reacción de la PCR que empleaba cebadores de ElF4 de las SEQ ID NOs 11 y 12 en presencia de ADNc añadido.

Muestra 8: Los productos de una reacción de la PCR que empleaba cebadores de EIF4 de las SEQ ID NOs 11 y 12 en ausencia de ADNc añadido.

Se observó una banda con el tamaño esperado para el producto de la PCR, sólo en las muestras 1, 3, 5 y 7, indicando de este modo la presencia de la secuencia correspondiente en la población de ADNc.

Las reacciones de la PCR se realizaron también con oligonucleótidos no codificantes de los cebadores de globina y de deshidrogenasa (SEQ ID NOs 6 y 8) y un oligonucleótido cuya secuencia se corresponde con la del oligonucleótido derivatizado. La presencia de los productos de la PCR del tamaño esperado en las muestras equivalentes a las muestras anteriores 1 y 3, indicaba que el oligonucleótido derivatizado se había unido al ARNm.

Los ejemplos anteriores resumen el procedimiento químico para enriquecer ARNm que tienen extremos 5' intactos, tal y como se muestra en la Figura 1. Más detalles en relación con los planteamientos químicos para obtener tales ARNm, se describen en el documento de Solicitud de Patente Internacional nº WO 96/34981, publicada el 7 de Noviembre de 1996. Las estrategias basadas en las anteriores modificaciones químicas de la estructura 5' del casquete, se pueden utilizar para generar ADNc seleccionados para incluir los extremos 5' de los ARNm a partir de los cuales se han obtenido. En una versión de tales procedimientos, los extremos 5' de los ARNm se modifican tal y como se ha descrito anteriormente. A continuación, se realiza una reacción de transcripción inversa para extender un cebador complementario al extremo 5' del ARNm. Los ARNs monocatenarios se eliminan para obtener una población de heterodúplex de ADNc/ARNm en los que el ARNm incluye un extremo 5' intacto. Los heterodúplex resultantes se pueden capturar sobre una fase sólida revestida con una molécula capaz de interaccionar con la molécula empleada para derivatizar el extremo 5' del ARNm. A continuación, las hebras de los heterodúplex se separan para recuperar las primeras hebras de ADNc monocatenario aisladas que incluyen el extremo 5' del ARNm. La síntesis de la segunda hebra del ADNc se puede realizar a continuación, empleando técnicas convencionales. Por ejemplo, los procedimientos descritos en el documento WO 96/34981 o en Carninci *y col., Genomics* **37**:327-336, 1996, se pueden emplear para seleccionar los ADNc que incluyen la secuencia obtenida a partir del extremo 5' de la secuencia codificadora de ARNm.

Después de la ligación de la etiqueta oligonucleotídica con el casquete 5' del ARNm, se realiza una reacción de transcripción inversa para extender un cebador complementario al ARNm con el extremo 5' del ARNm. Después de eliminar el componente de ARN del heterodúplex resultante, empleando técnicas convencionales, se realiza la síntesis de la segunda hebra del ADNc, con un cebador complementario a la etiqueta oligonucleotídica.

2. *Métodos enzimáticos para obtener ARNm que tienen extremos 5' intactos*

Otras técnicas para seleccionar los ADNc que se extienden hacia el extremo 5' del ARNm a partir del cual se han obtenido, son completamente enzimáticas. Algunas versiones de estas técnicas las describe Dumas Milne Edwards J.B. (tesis doctoral de la Universidad Paris VI, "Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la régulation de l'expression de la tryptophane hydroxylase de rat", 20 de diciembre de 1993), documento EPO 625572 y Kato *y col., Gene* **150**:243-250, 1994.

Resumiendo, en algunos planteamientos, el ARNm aislado se trata con fosfatasa alcalina para retirar los grupos fosfato presentes en los extremos 5' de los ARNm sin casquete incompletos. Después de este procedimiento, el casquete presente en los ARNm de longitud completa se retira enzimáticamente con una enzima que corta el casquete, tal como la polinucleótido-quinasa de T4 o la pirofosfatasa ácida de tabaco. Un oligonucleótido que puede ser un oligonucleótido de ADN o un oligonucleótido híbrido de ADN-ARN que tiene ARN en su extremo 3', se liga a continuación con el fosfato presente en el extremo 5' del ARNm con el casquete eliminado, empleando la ligasa de ARN de T4. El oligonucleótido puede incluir un sitio de restricción para facilitar la clonación de los ADNc después de su síntesis. El Ejemplo 12 a continuación describe un método enzimático basado en la tesis doctoral de Dumas.

Ejemplo 12

Planteamiento enzimático para obtener ESTs 5'

Se desfosforilaron 20 μg de poliA+ARN empleando fosfatasa de intestino de ternera (Biolabs). Después de extraer con fenol y cloroformo, la estructura del casquete del ARNm se hidrolizó empleando la pirofosfatasa ácida del tabaco (purificada tal y como describen Shinshi *y col., Biochemistry* **15**:2185-2190, 1976) y un oligonucleótido hemi 5'ADN/ARN-3' que tiene un extremo 5' no fosforilado, un segmento de ribofosfato de adenosina y un extremo 3', y un sitio EcoRI cerca del extremo 5', se ligó con los extremos 5'P del ARNm empleando la ligasa de ARN de T4 (Biolabs). Los oligonucleótidos adecuados para utilizar en este procedimiento, tienen preferentemente 30 a 50 bases de longitud. Los oligonucleótidos que tienen un extremo 5' no fosforilado se pueden sintetizar añadiendo un fluorocromo en el extremo 5'. La inclusión de un segmento de ribofosfatos de adenosina en el extremo 3' del oligonucleótido, incrementa la eficacia de la ligación. Se apreciará que el oligonucleótido puede contener sitios de clonación distintos de EcoRI.

Después de la ligación del oligonucleótido con el fosfato presente en el extremo 5' del ARNm con el casquete eliminado, la síntesis de la primera y la segunda hebra del ADNc se realiza empleando métodos convencionales o los especificados en el documento EPO 625.572 y en Kato *y col., véase más arriba*, y Dumas Milne Edwards, *véase más arriba*. El ADNc resultante se puede ligar entonces en vectores, tales como los descritos por Kato *y col., véase más arriba*, u otros vectores de ácido nucleico conocidos por los expertos en la técnica, empleando técnicas tales como las descritas en Sambrook *y col.,* Molecular Cloning: A Laboratory Manual 2ª ed., Cold Spring Harbor Laboratory Press, 1989.

II. *Obtención y caracterización de las ESTs 5' de la presente invención*

Las ESTs 5' descritas en esta memoria, se obtuvieron empleando los planteamientos químicos y enzimáticos mencionados anteriormente, para enriquecer los ARNm que tenían los extremos 5' intactos, tal y como se ha descrito anteriormente.

1. *Obtención de las ESTs 5' empleando ARNm con extremos 5' intactos*

En primer lugar, se prepararon los ARNm tal y como se describe en el Ejemplo 13 siguiente.

Ejemplo 13

Preparación de ARNm con extremos 5' intactos

Los ARNs humanos totales o los ARNs poliA$^+$ obtenidos a partir de 29 tejidos diferentes, se adquirieron respectivamente de LABIMO y CLONTECH y se emplearon para generar 44 genotecas de ADNc del modo siguiente. El ARN comprado se había aislado de las células o tejidos empleando la extracción ácida con tiocianato de guanidinio-fenol-cloroformo (Chomczyniski y Sacchi, *Analytical Biochemistry* **162**:156-159, 1987). El ARN poliA$^+$ se aisló a partir del ARN total (LABIMO) mediante dos pases por cromatografía con oligo dT, tal y como describen Aviv y Leder, *Proc. Natl. Acad. Sci. USA*. **69**:1408-1412, 1972, para eliminar el ARN ribosómico.

La calidad y la estabilidad de los ARNs poliA$^+$ se comprobó. Transferencias de tipo Northern hibridadas con una sonda de globina se emplearon para confirmar que los ARNm no se habían degradado. La contaminación de los ARNm poliA$^+$ por secuencias ribosómicas se comprobó empleando transferencias de tipo Northern y una sonda obtenida a partir de la secuencia del ARNr 28S. Las preparaciones de los ARNm con menos de 5% de ARNr se emplearon en la construcción de las genotecas. Para evitar construir genotecas con ARNs contaminados por secuencias exógenas (procariotas o fúngicas), se examinó la presencia de secuencias ribosómicas 16S bacterianas o de dos ARNm fúngicos altamente expresados, mediante PCR.

Después de la preparación de los ARNm, se emplearon los procedimientos químicos y/o enzimáticos descritos anteriormente para enriquecer los ARNm que tenían extremos 5' intactos, para obtener las ESTs 5' a partir de diversos tejidos. En ambos planteamientos, una etiqueta de oligonucleótido se fijó a los extremos 5' de los ARNm. La etiqueta de oligonucleótido tenía un sitio EcoRI para facilitar los procedimientos posteriores de clonación. Para facilitar el procesamiento del ADNc mono y bicatenario obtenido en la construcción de las genotecas, se empleó la misma secuencia nucleotídica para diseñar el oligonucleótido ligado en ambos planteamientos químico y enzimático. Sin embargo, en el procedimiento químico, la etiqueta empleada era un oligodesoxirribonucleótido que estaba ligado

al casquete del ARNm, mientras que en la ligación enzimática, la etiqueta era un oligonucleótido quimérico hemi 5'ADN/ARN3' que se había ligado al extremo 5' del ARNm sin casquete, tal y como se ha descrito en el Ejemplo 12.

Después de la fijación de la etiqueta de oligonucleótido al ARNm, tanto por el método químico como por el enzimático, se examinó la integridad del ARNm realizando una transferencia tipo Northern con 200 a 500 ng de ARNm, empleando una sonda complementaria a la etiqueta del oligonucleótido, antes de realizar la síntesis de la primera hebra, tal y como se describe en el Ejemplo 14.

Ejemplo 14

Síntesis del ADNc empleando moldes de ARNm que tienen extremos 5' intactos

Para los ARNm unidos a las etiquetas de oligonucleótidos empleando los métodos enzimáticos y químicos, se realizó la síntesis de la primera hebra del ADNc empleando la transcriptasa inversa Superscript II (Gibco BRL) o la ARNasa H Minus de M-MLV (Promega) con nonámeros aleatorios como cebadores. Para proteger los sitios internos de EcoRI en el ADNc, de la digestión en las etapas posteriores del procedimiento, se empleó dCTP metilado para la síntesis de la primera hebra. Después de eliminar el ARN mediante una hidrólisis alcalina, la primera hebra de ADNc se precipitó empleando isopropanol para eliminar los cebadores residuales.

Para los métodos enzimáticos y químicos, la segunda hebra del ADNc se sintetizó con un fragmento Klenow que emplea un cebador que se corresponde con el extremo 5' del oligonucleótido ligado, descrito en el Ejemplo 12. Preferentemente, el cebador tiene 20-25 bases de longitud. El dCTP metilado también se utilizó para la síntesis de la segunda hebra para proteger los sitios internos de EcoRI en el ADNc de la digestión durante el proceso de clonación.

Después de la síntesis del ADNc, los ADNc se clonaron en pBlueScript tal y como se describe en el Ejemplo 15 siguiente.

Ejemplo 15

Clonación de los ADNc obtenidos a partir del ARNm con extremos 5' intactos en BlueScript

Después de la síntesis de la segunda hebra, los extremos del ADNc se volvieron romos con la polimerasa de ADN de T4 (Biolabs) y el ADNc se digirió con EcoRI. Ya que el dCTP metilado se empleó durante la síntesis del ADNc, el sitio EcoRI presente en la etiqueta era el único sitio semi-metilado, por tanto, el único sitio susceptible de digestión con EcoRI. El ADNc se fraccionó por tamaños a continuación, empleando la cromatografía de exclusión (AcA, Biosepra) y las fracciones correspondientes a los ADNc de más de 150 pb se reunieron y se precipitaron con etanol. El ADNc se clonó de forma direccional en los extremos SmaI y EcoRI del vector fagémido pBlueScript (Stratagene). La mezcla de ligación se sometió a electroporación en bacterias y se propagó bajo la selección adecuada con antibióticos.

Los clones que contenían la etiqueta del oligonucleótido fijada, se seleccionaron a continuación tal y como se describe en el Ejemplo 16 siguiente.

Ejemplo 16

Selección de los clones que tienen la etiqueta de oligonucleótido fijada a ellos

Los ADNs plasmídicos que contenían genotecas de EST 5' preparadas tal y como se ha descrito anteriormente, se purificaron (Qiagen). Una selección positiva de los clones etiquetados se realizó del modo siguiente. Brevemente, en este proceso de selección, el ADN plasmídico se convirtió en ADN monocatenario empleando la endonucleasa del gen II del fago F1, junto con una exonucleasa (Chang *y col., Gene* **127**:95-8, 1993) tal como la exonucleasa III o la exonucleasa del gen 6 de T7. El ADN monocatenario resultante se purificó a continuación empleando perlas paramagnéticas tal y como describen Fry *y col., Biotechniques*, **13**:124-131, 1992. En este procedimiento, el ADN monocatenario se hibridó con un oligonucleótido biotinilado que tenía una secuencia correspondiente al extremo 3' del oligonucleótido descrito en el Ejemplo 13. Preferentemente, el cebador tiene una longitud de 20-25 bases. Los clones que incluyen una secuencia complementaria al oligonucleótido biotinilado, se capturaron por incubación con perlas magnéticas revestidas con estreptavidina, seguido de selección magnética. Después de la captura de los clones positivos, el ADN plasmídico se liberó de las perlas magnéticas y se convirtió en ADN bicatenario empleando una polimerasa de ADN, tal como la ThermoSequenase obtenida de Amersham Pharmacia Biotech. Alternativamente, se pueden emplear protocolos tales como el descrito en el equipo de reactivos de Gene Trapper, a disposición en Gibco BRL. El ADN bicatenario se puede electroporar a continuación en bacterias. El porcentaje de clones positivos que tienen el oligonucleótido con la etiqueta 5' se estimó para tener un lugar típicamente entre 90 y 98%, empleando análisis por hibridación puntual.

Después de la electroporación, las genotecas se ordenaron en placas de microtitulación 384 (MTP). Una copia de la MTP se almacenó para necesidades futuras. A continuación, las genotecas se transfirieron a MTP 96 y se secuenciaron tal y como se describe a continuación.

Ejemplo 17

Secuenciación de insertos en clones seleccionados

Los insertos de plásmidos se amplificaron en primer lugar mediante PCR en termocicladores PE 9600 (Perkin-Elmer, Applied Biosystems Division, Foster City, CA) empleando cebadores convencionales SETA-A y SETA-B (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), tampón y condiciones de ciclación como las recomendadas por Perkin-Elmer Corporation.

Los productos de la PCR se secuenciaron a continuación empleando secuenciadores automáticos ABI Prism 377 (Perkin Elmer). Las reacciones de secuenciación se realizaron empleando termocicladores PE 9600 con química convencional de tinción-cebador y ThermoSequenase (Amersham Pharmacia Biotech). Los cebadores empleados eran T7 o 21M13 (a disposición en Genset SA) cuando era adecuado. Los cebadores se marcaron con colorantes JOE, FAM, ROX y TAMRA. Los dNTPs y los ddNTPs empleados en las reacciones de secuenciación se adquirieron en Boehringer. El tampón de secuenciación, las concentraciones de los reactivos y las condiciones de la ciclación eran como las recomendadas por Amersham.

Después de la reacción de secuenciación, las muestras se precipitaron con etanol, se resuspendieron en tampón de carga de formamida y se cargaron en un gel convencional de acrilamida al 4%. La electroforesis se realizó durante 2,5 horas a 3000 V en un secuenciador ABI 377 y los datos de la secuencia se recogieron y se analizaron empleando el programa de análisis de la secuenciación de ADN de ABI Prism, versión 2.1.2.

2. *Análisis por ordenador de las ESTs 5' obtenidas: Construcción de las bases de datos NetGene y SignalTag*

Los datos de las secuencias procedentes de 44 genotecas de ADNc, preparadas tal y como se han descrito anteriormente, se transfirieron a una base de datos privada en donde se realizó el control de la calidad y las etapas de validación. Un asignador de bases privado que funciona empleando un sistema Unix, etiquetaba automáticamente los picos sospechosos, teniendo en cuenta la forma de los picos, la resolución entre los picos y el nivel de ruido de fondo. El asignador de bases privado también realizaba un recorte automático. Cualquier segmento de 25 bases o menos que tuviera más de 4 picos sospechosos, se consideraba no fiable y se eliminaba. Las secuencias correspondientes al vector de clonación o a los oligonucleótidos de ligación se retiraban automáticamente de las secuencias EST. Sin embargo, las secuencias EST resultantes podían contener 1 a 5 bases pertenecientes a las secuencias mencionadas anteriormente en su extremo 5'. Si era necesario, éstas se podían eliminar fácilmente de caso en caso.

Después de la secuenciación tal y como se ha descrito anteriormente, las secuencias de ESTs 5' se introdujeron en una base de datos privada, NetGene®, para el almacenamiento y la manipulación tal y como se describe a continuación. Los expertos en la técnica apreciarán que los datos se podían almacenar y manipular sobre cualquier medio al que se pudiera acceder y leer con un ordenador. Los medios legibles por ordenador incluyen medios legibles de forma magnética, óptica o electrónica. Por ejemplo, los medios legibles por ordenador 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 la secuencia se pueden almacenar y manipular con una variedad de programas de procesadores de datos en una diversidad de formatos. Por ejemplo, los datos de la secuencia se pueden almacenar como un texto en un archivo de procesador de textos, tales como Microsoft Word o WORDPERFECT o como un archivo ASCII en una variedad de programas de bases de datos, conocidos por los expertos en la técnica, tales como DB2, SYBASE u ORACLE.

Los medios legibles por ordenador sobre los que se almacena la información de la secuencia, pueden estar en un ordenador personal, una red, un servidor u otros sistemas de ordenadores conocidos por los expertos en la técnica. El ordenador u otro sistema, incluye preferentemente los medios de almacenamiento descritos anteriormente y un procesador para acceder y manipular los datos de la secuencia. Una vez que se han almacenado los datos de la secuencia, se pueden manipular y escrutar para localizar las secuencias almacenadas que contienen una secuencia deseada de ácido nucleico o que codifica una proteína que tiene un dominio funcional particular. Por ejemplo, la información de la secuencia almacenada se puede comparar con otras secuencias conocidas para identificar homologías, motivos implicados en la función biológica o motivos estructurales.

Los programas que se pueden utilizar para la búsqueda o la comparación de las secuencias almacenadas, incluyen la serie de programas (NCBI) Mac Pattern (EMBL), BLAST y BLAST2, los programas de herramientas de búsqueda de alineación local básica para las comparaciones de nucleótidos (BLASTN) y de péptidos (BLASTX) (Altschul *y col., J. Mol. Biol.* **215**:403, 1990) y FASTA (Pearson y Lipman, *Proc. Natl. Acad. Sci. USA* **85**: 2444, 1988). Los programas BLAST extienden después las alineaciones en base a criterios definidos de emparejamiento o mal emparejamiento.

Los motivos que se pueden detectar empleando los programas anteriores y los descritos en el Ejemplo 28, incluyen secuencias que codifican los cierres en cremallera con leucina, los motivos de hélice-giro-hélice, los sitios de glicosilación, los sitios de la ubiquitina, las hélices alfa y las láminas beta, las secuencias señal que codifican péptidos señal que dirigen la secreción de las proteínas codificadas, secuencias implicadas en la regulación de la transcripción, tales como las cajas homeóticas, segmentos ácidos, sitios activos enzimáticos, sitios de unión a sustrato y sitios de corte enzimático.

Antes de buscar entre los ADNc en la base de datos NetGene®, motivos de secuencias de interés, los ADNc obtenidos a partir de los ARNm que no eran de interés, se identificaron y se eliminaron de otros estudios, tal y como se describe en el Ejemplo 18 siguiente.

Ejemplo 18

Eliminación de secuencias no deseadas de otros estudios

Las ESTs 5' en la base de datos NetGene® que se habían obtenido a partir de secuencias no deseadas tales como ARNs de transferencia, ARNs ribosómicos, ARNs mitocondriales, ARNs procarióticos, ARNs fúngicos, secuencias Alu, secuencias L1 o secuencias repetidas, se identificaron empleando los programas FASTA y BLASTN con los parámetros enumerados en la Tabla I.

Para eliminar de otros estudios las ESTs 5'que codifican los ARNt, se compararon las secuencias de ESTs 5' con las secuencias de 1190 ARNt conocidos, obtenidos a partir de la publicación 38 de EMBL, de las cuales 100 eran humanas. La comparación se realizó empleando FASTA sobre ambas hebras de las ESTs 5'. Las secuencias que tenían más de 80% de homología en más de 60 nucleótidos, se identificaron como ARNt. De las 144.341 secuencias escrutadas, 26 se identificaron como ARNt y se eliminaron de otros estudios.

Para eliminar de otros estudios las ESTs 5'que codifican los ARNr, se compararon las secuencias de ESTs 5' con las secuencias de 2497 ARNr conocidos, obtenidos a partir de la publicación 38 de EMBL, de las cuales 73 eran humanas. La comparación se realizó empleando BLASTN sobre ambas hebras de las ESTs 5' con el parámetro S=108. Las secuencias que tenían más de 80% de homología en más de 40 nucleótidos, se identificaron como ARNr. De las 144.341 secuencias escrutadas, 3.312 se identificaron como ARNr y se eliminaron de otros estudios.

Para eliminar de otros estudios las ESTs 5'que codifican los ARNmt, se compararon las secuencias de ESTs 5' con las secuencias de dos genomas mitocondriales conocidos de los que se disponían las secuencias genómicas completas y todas las secuencias transcritas a partir de estos genomas mitocondriales que incluían ARNt, ARNr y ARNm de un total de 38 secuencias. La comparación se realizó empleando BLASTN sobre ambas hebras de las ESTs 5' con el parámetro S=108. Las secuencias que tenían más de 80% de homología en segmentos de más de 40 nucleótidos, se identificaron como ARNmt. De las 144.341 secuencias escrutadas, 6.110 se identificaron como ARNmt y se eliminaron de otros estudios.

Las secuencias que podían haber sido el resultado de contaminantes exógenos se eliminaron de otros estudios, comparando las secuencias ESTs 5' con la publicación 46 de las divisiones bacterianas y fúngicas de EMBL, empleando BLASTN con el parámetro S=144. Todas las secuencias que tenían más de 90% de homología en al menos 40 nucleótidos, se identificaron como contaminantes exógenos. De las 42 genotecas de ADNc examinadas, los porcentajes promedio de secuencias procariotas y fúngicas contenidas en las mismas, eran 0,2% y 0,5%, respectivamente. Entre estas secuencias, sólo una se pudo identificar como una secuencia específica de hongos. Las otras eran secuencias fúngicas o procariotas que tenían homologías con secuencias de vertebrados o que incluían secuencias repetidas que no se habían ocultado durante la comparación electrónica.

Además, las ESTs 5' se compararon con 6093 secuencias de Alu y 1115 secuencias de L1 para evitar las ESTs 5' que contenían dichas secuencias repetidas. Las ESTs 5' que incluían repeticiones THE y MER, secuencias SSTR o repeticiones satélite, microsatélite o teloméricas también se eliminaron de otros estudios. Como media, 11,5% de las secuencias en las genotecas contenían secuencias repetidas. De estas 11,5%, 7% contenían repeticiones Alu, 3,3% contenían repeticiones L1 y el 1,2% restante procedían de los otros tipos escrutados de secuencias repetitivas. Estos porcentajes son compatibles con los encontrados en las genotecas de ADNc preparadas por otros grupos. Por ejemplo, las genotecas de ADNc de Adams *y col.* contenían entre 0% y 7,4% de repeticiones Alu, dependiendo de la fuente del ARN que se había utilizado para preparar la genoteca de ADNc (Adams *y col., Nature* **377**: 174, 1996).

Las secuencias de estas ESTs 5' que quedaban después de la eliminación de las secuencias no deseadas, cuando se compararon con las secuencias de ARNm humanos conocidos para determinar la exactitud de los procedimientos de secuenciación descritos anteriormente.

Ejemplo 19

Medición de la exactitud de la secuenciación comparando con secuencias conocidas

Para determinar adicionalmente la exactitud del procedimiento de secuenciación descrito anteriormente, se identificaron las secuencias de las ESTs 5' obtenidas a partir de secuencias conocidas y se compararon con las secuencias originales conocidas. En primer lugar, se realizó un análisis FASTA sobre las partes sobresalientes menores de 5 pb en ambos extremos de las ESTs 5', para identificar las que se emparejaban con un registro en la base de datos pública de ARNm humano. Las 6655 ESTs 5' que se emparejaban con un ARNm humano conocido se alinearon de nuevo con su ARNm análogo y se empleó una programación dinámica para incluir sustituciones, inserciones y deleciones en la lista de "errores" que se habían podido reconocer. Los errores presentes en las 10 últimas bases de las secuencias de ESTs 5' se ignoraron para evitar la inclusión de sitios de clonación falsos en el análisis de la exactitud de la secuenciación.

Este análisis reveló que las secuencias incorporadas en la base de datos NetGene® tenían una exactitud superior al 99,5%.

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

Ejemplo 20

Determinación de la eficacia de la selección de EST 5'

Para determinar la eficacia con la que los anteriores procedimientos de selección aislaban las ESTs 5' que incluían secuencias próximas al extremo 5' de los ARNm a partir de los cuales se habían obtenido, las secuencias de los extremos de las ESTs 5' obtenidas a partir de la subunidad α del factor 1 de elongación y los genes de la cadena pesada de ferritina, se compararon con las secuencias de ADNc conocidas de estos genes. Puesto que los sitios de inicio de la transcripción de ambos genes están bien caracterizados, se pueden utilizar para determinar el porcentaje de ESTs 5' derivadas que incluyen los sitios auténticos de inicio de la transcripción.

Para ambos genes, más del 95% de las ESTs 5' obtenidas incluyen realmente las secuencias próximas al extremo 5' o aguas arriba de dicho extremo de los ARNm correspondientes.

Para ampliar el análisis de la fiabilidad de los procedimientos para aislar las ESTs 5' de las ESTs en la base de datos NetGene®, se realizó un análisis similar para comparar empleando una base de datos compuesta por secuencias de ARNm humano extraídas a partir de la publicación 97 de la base de datos de GenBank. Los extremos 5' de más del 85% de las ESTs 5' obtenidas a partir de los ARNm incluidos en la base de datos de GenBank se localizaban cerca de los extremos 5' de la secuencia conocida. Cuando varias de las secuencias de ARNm disponibles en la base de datos de GenBank se deducen de secuencias genómicas, un extremo 5' que se empareja con estas secuencias contará como un emparejamiento interno. Por tanto, el método empleado aquí infraestima el rendimiento de las ESTs que incluyen los extremos 5' auténticos de sus ARNm correspondientes.

Las genotecas de ESTs preparadas anteriormente incluyen ESTs 5' múltiples obtenidas a partir del mismo ARNm. Las secuencias de tales ESTs 5' se compararon con otra y se identificaron las ESTs 5' más largas para cada ARNm. Los ADNc que se solapan se ensamblaron en secuencias continuas (cóntigos). Las secuencias continuas resultantes se compararon a continuación con bases de datos públicas para estimar su similitud con secuencias conocidas, tal y como se describe en el Ejemplo 21 siguiente.

Ejemplo 21

Agrupaciones de las ESTs 5' y cálculo de los índices de novedad para genotecas de ADNc

Para cada genoteca de las ESTs secuenciadas, las secuencias se agruparon por el extremo 5'. Cada secuencia en la genoteca se comparó con las otras con BLASTN2 (hebra directa, parámetros S=107). Las ESTs con pares de segmentos con puntuación elevada (HSPs, del inglés "High Scoring Segment Pairs") de al menos 25 pb de longitud, que tenían 95% de bases idénticas y que comenzaban a una distancia menor de 10 pb de cada extremo 5' de la EST, se agruparon. La secuencia más larga encontrada en la agrupación se empleó como representante del grupo. Una agrupación global entre genotecas se realizó a continuación, conduciendo a la definición de super-cóntigos.

Para determinar el rendimiento de las secuencias nuevas dentro de las genotecas de EST, se definió una tasa novedosa (del inglés, "novelty rate" (NR)) como: NR = 100 X (número de secuencias únicas nuevas encontradas en la genoteca/número total de secuencias de la genoteca). Típicamente, la tasa novedosa oscilaba entre 10% y 41%, dependiendo del tejido a partir del cual se había obtenido la genoteca EST. Para la mayoría de las genotecas, la secuenciación al azar de las genotecas de ESTs 5' se practicó hasta que la tasa novedosa alcanzaba 20%.

Después de la caracterización tal y como se ha descrito anteriormente, se escrutó la colección de ESTs 5' en NetGene®, para identificar las ESTs 5' que eran portadoras de secuencias señal potenciales, tal y como se describe en el Ejemplo 22 siguiente.

Ejemplo 22

Identificación de secuencias señal potenciales en ESTs 5'

Las ESTs 5' en la base de datos NetGene®, se escrutaron para identificar las que tenían un marco de lectura abierto no interrumpido (ORF) superior a 45 nucleótidos, comenzando con un codón ATG y extendiéndose hasta el extremo de la EST. Aproximadamente la mitad de las secuencias de ADNc en NetGene® contenían dicho ORF. En los ORFs de estas ESTs 5' se hizo a continuación una búsqueda para identificar motivos de señales potenciales, empleando ligeras modificaciones de los procedimientos descritos por von Heijne, *Nucleic Acids Res.* **14**:4683-4690, 1986. Las secuencias de ESTs 5'que codifican un segmento de al menos 15 aminoácidos de longitud con una puntuación de al menos 3,5 en la matriz de identificación de péptidos señal de von Heijne, se consideraron que poseían una secuencia señal. Las ESTs 5' que se emparejaban con un ARNm humano conocido o una secuencia EST y que tenían un

extremo 5' más de 20 nucleótidos aguas abajo del extremo 5' conocido, se excluyeron de un análisis adicional. Los ADNc restantes que tenían secuencias señal dentro de ellos mismos, se incluyeron en una base de datos denominada SignalTag®.

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

Ejemplo 23

Confirmación de la exactitud de la identificación de las secuencias señal potenciales en ESTs 5'

La exactitud del procedimiento anterior para identificar secuencias señal que codifican péptidos señal, se evaluó aplicando el método a los 43 aminoácidos localizados en el extremo N-terminal de todas las proteínas SwissProt humanas. La puntuación de von Heijne calculada para cada proteína se comparó con la caracterización conocida de la proteína como si fuera una proteína secretada o una proteína no secretada. De este modo se pudo calcular el número de proteínas no secretadas que tenían una puntuación superior a 3,5 (falsos positivos) y el número de proteínas que tenía una puntuación inferior a 3,5 (falsos negativos).

Empleando los resultados del análisis anterior, 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, se calculó basándose en el supuesto de que se secreta un 10% de las proteínas humanas o el supuesto de que se secreta un 20% de las proteínas humanas. Los resultados de este análisis se muestran en la Figura 2 y en la Tabla IV.

Empleando el método de identificación anterior de proteínas secretoras, se obtuvieron las ESTs 5' de los siguientes polipéptidos, conocidas por ser secretadas: glucagón humano, precursor de la monoquina inducida por el interferón gamma, proteína similar a la ciclofilina secretada, pleyotropina humana y el precursor de la biotinidasa humana. Por tanto, el método anterior identificaba con éxito las ESTs 5' que codifican un péptido señal.

Para confirmar que el péptido señal codificado por las ESTs 5', actúa realmente como un péptido señal, las secuencias señal de las ESTs 5' se pueden clonar en un vector diseñado para identificar péptidos señal. Tales vectores se diseñan para conferir la capacidad de crecimiento en medio selectivo sólo a las células hospedadoras que contienen un vector con una secuencia señal ligada funcionalmente. Por ejemplo, para confirmar que una EST 5' codifica un péptido señal genuino, la secuencia señal de la EST 5' se puede insertar aguas arriba y en marco de lectura con una forma no secretada del gen de la invertasa de levadura, en vectores de selección del péptido señal, tal y como se describen en el documento de Patente de EE.UU. nº 5.536.637. El crecimiento de las células hospedadoras que contienen los vectores de selección de la secuencia señal, con la secuencia señal EST 5' insertada correctamente, confirma que la EST 5' codifica un péptido señal genuino.

Alternativamente, la presencia de un péptido señal se puede confirmar clonando los ADNs extendidos, obtenidos empleando las ESTs en los vectores de expresión, tales como pT1 (tal y como se describe a continuación en el Ejemplo 30) o construyendo vectores con promotor-secuencia señal-gen informador que codifican proteínas de fusión entre el péptido señal y una proteína informadora que se puede someter a ensayo. Después de la introducción de estos vectores en una célula hospedadora adecuada, tal como las células COS o las células NIH 3T3, el medio de crecimiento se puede recolectar y analizar buscando la presencia de la proteína secretada. El medio procedente de estas células se compara con el medio de células testigo que contienen vectores que carecen de la secuencia señal o del inserto de ADNc extendido, para identificar los vectores que codifican un péptido señal funcional o una proteína secretada auténtica.

Estas ESTs 5' que codifican un péptido señal, tal y como se ha determinado por el método del Ejemplo 22, se agruparon adicionalmente en cuatro categorías basadas en su homología con secuencias conocidas, tal y como se describe en el Ejemplo 24 siguiente.

Ejemplo 24

Clasificación de las ESTs 5' que codifican un péptido señal

Las ESTs 5' que tienen una secuencia que no se empareja con ninguna secuencia conocida de vertebrado y con ninguna secuencia de EST disponible publicada, se denominaron "nuevas". Entre las secuencias en la base de datos SignalTag®, 947 de las ESTs 5' que tenían una puntuación de von Heijne de al menos 3,5, entraban dentro de esta categoría.

Las ESTs 5' que tenían una secuencia que no se emparejaba con ninguna secuencia de vertebrado pero que se emparejaba con una EST conocida publicada, se denominaron "EST-ext", con la condición de que la secuencia de EST conocida se extendiera al menos en 40 nucleótidos en la dirección 5'. Entre las secuencias en la base de datos SignalTag®, 150 de las ESTs 5' que tenían una puntuación de von Heijne de al menos 3,5, entraban dentro de esta categoría.

Las ESTs que no se emparejaban con ninguna secuencia de vertebrado pero que se emparejaban con una EST conocida públicamente sin extender la EST conocida en al menos 40 nucleótido en la dirección 5', se denominaron

"EST". Entre las secuencias de la base de datos SignalTag®, 599 de las ESTs 5' que tenían una puntuación de von Heijne de al menos 3,5, entraban en esta categoría.

Las ESTs 5' que se emparejaban con una secuencia de ARNm humano y en las que la secuencia conocida se extendía en al menos 40 nucleótidos en dirección 5', se denominaron "VERT-ext". Entre las secuencias en la base de datos SignalTag®, 23 de las ESTs 5' que tenían una puntuación de von Heijne de al menos 3,5, entraban en esta categoría. También se incluía en esta categoría una EST 5' en la que la secuencia conocida del ARNm de la translocasa humana se extendía más de 200 pares de bases en la dirección 5'. Una EST 5' que extendía la secuencia de un gen supresor tumoral humano en la dirección 5', también se identificó.

La Tabla V muestra la distribución de las ESTs 5' en cada categoría y el número de ESTs 5' en cada categoría que tienen un mínimo dado en la puntuación de von Heijne.

3. *Evaluación de la expresión espacial y temporal de los ARNm correspondientes a las ESTs 5' o los ADNc extendidos*

Cada una de las ESTs 5' también se clasificó basándose en el tejido a partir del cual se había obtenido el ARNm correspondiente, tal y como se describe a continuación en el Ejemplo 25.

Ejemplo 25

Clasificación de los patrones de expresión

La Tabla VI muestra la distribución de las ESTs 5' en cada una de las categorías definidas anteriormente, en relación con el tejido a partir del cual se habían obtenido las ESTs 5' del ARNm correspondiente.

La Tabla II proporciona el número de identificación de secuencia de una secuencia EST 5', obtenida a partir de tejidos diferentes, las categorías en las que entra esta secuencia y la puntuación de von Heijne del péptido señal que codifica. La secuencia EST 5' y la secuencia de los aminoácidos que codifica se proporcionan en las listas de secuencias anejas. La Tabla III proporciona los números ID de la secuencia de la EST 5' y la secuencia del péptido señal que codifica. Las secuencias de la EST 5' y el polipéptido que codifica se proporcionan en la lista de secuencias aneja a esta memoria.

La secuencia de ADN SEQ ID NO: 38, se puede escrutar fácilmente en busca de cualquier error en la misma y cualquier ambigüedad en la secuencia se puede resolver secuenciando de nuevo un fragmento que contiene tales errores o ambigüedades en ambas cadenas. Tales fragmentos se pueden obtener a partir de plásmidos almacenados en el laboratorio del inventor o se pueden aislar empleando las técnicas descritas en esta memoria. La resolución de cualquiera de dichas ambigüedades o errores se puede facilitar empleando cebadores que se hibridan con las secuencias localizadas en la proximidad de las secuencias ambiguas o erróneas. Por ejemplo, los cebadores se pueden hibridar con secuencias a 50-75 bases de la ambigüedad o del error. Después de la resolución de un error o una ambigüedad, se pueden realizar las correcciones correspondientes en las secuencias proteicas codificadas por el ADN que contiene el error o la ambigüedad.

Además de clasificar las ESTs 5' por su tejido de origen, se pueden determinar los patrones de expresión espacial y temporal de los ARNm correspondientes a las ESTs 5', así como sus niveles de expresión, tal y como se describe en el Ejemplo 26 siguiente. La caracterización de los patrones de expresión especial y temporal y los niveles de expresión de estos ARNm, es útil para construir vectores de expresión capaces de producir un nivel deseado de producto génico de una forma espacial o temporal deseada, tal y como se expone con más detalle a continuación.

Además, las ESTs 5' cuyo ARNm correspondiente está asociado con un estado de enfermedad, también se pueden identificar. Por ejemplo, una enfermedad particular puede ser el resultado de la falta de expresión, hiperexpresión o infraexpresión de un ARNm correspondiente a una EST 5'. Comparando los patrones de expresión del ARNm y las cantidades en muestras tomadas de individuos sanos con las de individuos que padecen una enfermedad particular, se pueden identificar las ESTs 5' responsables de la enfermedad.

Se apreciará que los resultados de los anteriores procedimientos de caracterización para las ESTs 5' también se pueden aplicar a ADNc extendidos (obtenibles tal y como se describe a continuación) que contienen secuencias adyacentes a las ESTs 5'. También se apreciará que si se desea, se puede posponer la caracterización hasta que los ADNc extendidos se hayan obtenido, más que hasta la caracterización de las ESTs mismas.

Ejemplo 26

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

Los niveles de expresión y los patrones de los ARNm correspondientes a las ESTs 5' o a los ADNc extendidos (obtenibles tal y como se describe a continuación en el Ejemplo 27), se pueden analizar mediante hibridación en solución con sondas largas, tal y como se describe en el documento de Solicitud de Patente Internacional nº WO 97/05277. Resumiendo, una EST 5', un ADNc extendido o un fragmento del mismo correspondiente al gen que codi-

fica el ARNm que se va a caracterizar, se inserta en un sitio de clonación inmediatamente aguas arriba de un promotor de la polimerasa de ARN de un bacteriófago (T3, T7 o SP6), para producir ARN no codificante. Preferentemente, la EST 5' o el ADNc extendido tienen 100 o más nucleótidos. El plásmido se linealiza y se transcribe en presencia de ribonucleótidos que comprenden ribonucleótidos modificados (es decir, biotina-UTP y DIG-UTP). Un exceso de este ARN marcado por partida doble se hibrida en solución con ARNm aislado de células o de tejidos de interés. Las hibridaciones se realizan bajo condiciones convencionales rigurosas (40-50ºC durante 16 horas en un tampón de formamida al 80%, NaCl 0,4 M, pH 7-8). La sonda no hibridada se retira por digestión con ribonucleasas específicas de ARN monocatenario (es decir, ARNasas CL3, T1, Phy M, U2 o A). La presencia de la modificación biotina-UTP permite la captura del híbrido sobre una placa de microtitulación revestida con estreptavidina. La presencia de una modificación DIG permite que se detecte el híbrido y se cuantifique mediante ELISA, utilizando un anticuerpo anti-DIG acoplado a fosfatasa alcalina.

Las ESTs 5', los ADNc extendidos o fragmentos de los mismos también se pueden etiquetar con secuencias de nucleótidos para realizar un análisis en serie de la expresión génica (SAGE), tal y como se describe en el documento de Solicitud de Patente de GB nº 2305241. En este método, los ADNc se preparan a partir de una célula, un tejido, un organismo u otra fuente de ácidos nucleicos en los que se tiene que determinar los patrones de expresión génica. Los ADNc resultantes se separan en dos grupos. Los ADNc de cada grupo se cortan con una primera endonucleasa de restricción, denominada 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. Los fragmentos que contienen la región más 5' o 3' del ADNc cortado, se aíslan mediante ligación a un medio de captura, tal como perlas revestidas con estreptavidina. Un primer enlazador de oligonucleótido que tiene una primera secuencia para hibridar un cebador de amplificación y un sitio de restricción interno para una endonucleasa denominada de etiquetaje, se liga a los ADNc digeridos en el primer grupo. La digestión con la segunda endonucleasa produce fragmentos cortos de etiquetas procedentes de los ADNc.

Un segundo oligonucleótido que tiene una segunda secuencia para la hibridación de un cebador de la amplificación y un sitio de restricción interno, se liga a los ADNc digeridos en el segundo grupo. Los fragmentos de ADNc en el segundo grupo también se digieren con la endonucleasa de etiquetaje para generar fragmentos cortos de etiquetas obtenidos a partir de los ADNc en el segundo grupo. Las etiquetas resultantes de la digestión del primero y del segundo grupo con la enzima de anclaje y la endonucleasa de etiquetaje, se ligan entre sí para producir las denominadas di-etiquetas. En algunas realizaciones, las dietiquetas se concatemerizan para producir productos de ligación que contienen de 2 a 200 dietiquetas. Las secuencias de las etiquetas se determinan a continuación y se comparan con las secuencias de las ESTs 5' o de los ADNc extendidos para determinar que EST 5' o que ADNc extendido se expresa en la célula, el tejido, el organismo u otra fuente de ácidos nucleicos a partir de la cual se han obtenido las etiquetas. De este modo, se obtiene el patrón de expresión de las ESTs 5' o de los ADNc extendidos en la célula, el tejido, el organismo u otra fuente de ácidos nucleicos.

El análisis cuantitativo de la expresión génica también se puede realizar empleando formaciones ordenadas. Tal y como se emplea en esta memoria, la expresión formación ordenada significa una disposición unidimenensional, bidimensional o multidimensional de ADNc de longitud completa (es decir, los ADNc extendidos que incluyen la secuencia codificante del péptido señal, la secuencia codificante de la proteína madura y un codón de detención), de ADNc extendidos, de ESTs 5' o de fragmentos de los mismos de longitud suficiente para permitir la detección específica de una expresión génica. Preferentemente, los fragmentos tienen al menos 15 nucleótidos de longitud. Más preferentemente, los fragmentos tienen al menos 100 nucleótidos de longitud. Más preferentemente, los fragmentos tienen más de 100 nucleótidos de longitud. En algunas realizaciones, los fragmentos pueden tener más de 500 nucleótidos de longitud.

Por ejemplo, el análisis cuantitativo de la expresión génica se puede realizar con ADNc de longitud completa, tal y como se definen más abajo, con ADNc extendidos, con ESTs 5' o con fragmentos de los mismos en una microformación ordenada de ADN complementario, tal y como describen Schena *y col.* (*Science* **270**:467-470, 1995; *Proc. Natl. Acad. Sci. USA* **93**:10614-10619, 1996). Los ADNc de longitud completa, los ADNc extendidos, las ESTs 5' o fragmentos de los mismos, se amplifican empleando la PCR y se ordenan en conjuntos desde las placas de microtitulación de 96 pocillos sobre portaobjetos sililados para microscopio, empleando tecnología robótica de alta velocidad. Las formaciones ordenadas impresas se incuban en una cámara húmeda para permitir la rehidratación de los elementos de la formación ordenada y se lavan, una vez en SDS al 0,2% durante 1 minuto, dos veces en agua durante 1 minuto y una vez durante 5 minutos en solución de borohidruro sódico. Las formaciones ordenadas se sumergen en agua durante 2 minutos a 95ºC, se transfieren a SDS al 0,2% durante 1 minuto y se lavan dos veces con agua, se secan al aire y se almacenan en la oscuridad a 25ºC.

Se aísla el ARNm celular o tisular o se obtiene comercialmente y se preparan las sondas mediante una única tanda de transcripción inversa. Las sondas se hibridan con microformaciones ordenadas de 1 cm^2 bajo un cubreobjeto de vidrio de 14 x 14 cm, durante 6-12 horas a 60ºC. Las formaciones ordenadas se lavan durante 5 minutos a 25ºC en tampón de lavado con condiciones poco rigurosas (1 x SSC/SDS al 0,2%), a continuación durante 10 minutos a temperatura ambiente en tampón de lavado con condiciones rigurosas (0,1 x SSC/SDS al 0,2%). Las formaciones ordenadas se escanean en 0,1 x SSC empleando un dispositivo de escaneo con láser fluorescente ajustado con un equipo de filtros habitual. Las mediciones de la expresión diferencial exacta se obtienen tomando el promedio de las proporciones de dos hibridaciones independientes.

El análisis cuantitativo de la expresión de genes también se puede realizar con los ADNc de longitud completa, ADNc extendidos, ESTs 5' o fragmentos de los mismos en formaciones ordenadas de ADN complementario, tal y como describen Pietu *y col.* (*Genome Research* **6**:492-503, 1996). Los ADNc de longitud completa, los ADNc extendidos, las ESTs 5' o fragmentos de los mismos se amplifican con PCR y se transfieren a membranas en forma de manchas. A continuación, los ARNm que proceden de diversos tejidos o células, se marcan con nucleótidos radiactivos. Después de hibridar y lavar bajo condiciones controladas, los ARNm hibridados se detectan mediante la fosfoformación de imágenes o la autorradiografía. Se realizan experimentos por duplicado y a continuación se realiza un análisis cuantitativo de los ARNm expresados de forma diferencial.

Alternativamente, el análisis de la expresión de las ESTs 5' o los ADNc extendidos se puede realizar mediante formaciones ordenadas de nucleótidos de alta densidad, tal y como describen Lockhart *y col.* (*Nature Biotechnology* **14**: 1675-1680, 1996) y Sosnowsky *y col.* (*Proc. Natl. Acad. Sci. USA* **94**:1119-1123, 1997). Los oligonucleótidos de 15-50 nucleótidos que se corresponden con las secuencias de las ESTs 5' o con los ADNc extendidos, se sintetizan directamente en el chip (Lockhart *y col., véase más arriba*) o se sintetizan y se dirigen al chip (Sosnowsky *y col., véase más arriba*). Preferentemente, los oligonucleótidos tienen aproximadamente 20 nucleótidos de longitud.

Las sondas de ADNc marcadas con un compuesto adecuado, tal como biotina, digoxigenina o colorante fluorescente, se sintetizan a partir de la población de ARNm adecuada y a continuación se fragmentan al azar hasta un tamaño promedio de 50 a 100 nucleótidos. Dichas sondas se hibridan a continuación con el chip. Después de lavar tal y como describen Lockhart *y col., véase más arriba* y de la aplicación de diferentes campos eléctricos (Sonowsky *y col., véase más arriba*), los colorantes o los compuestos marcadores se detectan y se cuantifican. Se realizan las hibridaciones por duplicado. El análisis comparativo de la intensidad de la señal que originan las sondas de ADNc sobre el mismo oligonucleótido diana en diferentes muestras de ADNc, indica una expresión diferencial del ARNm correspondiente a la EST 5' o al ADNc extendido, a partir del cual se ha diseñado la secuencia de oligonucleótidos.

III. *Uso de las ESTs 5' para clonar los ADNc extendidos y para clonar los ADNs genómicos correspondientes*

Una vez que se han seleccionado las ESTs 5' que incluyen el extremo 5' de los ARNm correspondientes, empleando los procedimientos descritos anteriormente, se pueden utilizar para aislar los ADNc extendidos que contienen secuencias colindantes a las ESTs 5'. Los ADNc extendidos pueden incluir la secuencia codificadora completa de la proteína codificada por el ARNm correspondiente, incluyendo el sitio auténtico de inicio de la traducción, la secuencia señal y la secuencia que codifica la proteína madura que permanece después de escindir el péptido señal. Tales ADNc extendidos se denominan en esta memoria "ADNc de longitud completa". Alternativamente, los ADNc extendidos pueden incluir sólo la secuencia que codifica la proteína madura que permanece después de escindir el péptido señal o sólo la secuencia que codifica el péptido señal.

El Ejemplo 27 siguiente describe un método general para obtener ADNc extendidos que emplea las ESTs 5'. El Ejemplo 28 siguiente proporciona resultados experimentales, empleando el método explicado en el Ejemplo 27, que describe diversos ADNc extendidos que incluyen la secuencia codificadora completa y el extremo 5' auténtico del ARNm correspondiente a diversas proteínas secretadas.

Los métodos de los Ejemplos 27, 28 y 29 también se pueden emplear para obtener los ADNc extendidos que codifican menos que la secuencia codificadora completa de las proteínas secretadas codificadas por los genes correspondientes a las ESTs 5'. En algunas realizaciones, los ADNc extendidos aislados que emplean estos métodos, codifican al menos 10 aminoácidos de una de las proteínas codificadas por la secuencia de SEQ ID NO: 38. En otras realizaciones, los ADNc extendidos codifican al menos 20 aminoácidos de la proteína codificada por la secuencia de SEQ ID NO: 38. En realizaciones adicionales, los ADNc extendidos codifican al menos 30 aminoácidos de la secuencia de SEQ ID NO: 38. En una realización preferida, los ADNc extendidos codifican una secuencia de la proteína de longitud completa, que incluye la secuencia que codifica la proteína de SEQ ID NO. 38.

Ejemplo 27

Método general para emplear las ESTs 5' para clonar y secuenciar los ADNc que incluyen la región codificadora completa y el extremo 5' auténtico del ARNm correspondiente

El siguiente método general se ha empleado para aislar de forma rápida y eficaz los ADNc extendidos que tienen los extremos 5' auténticos de sus ARNm correspondientes, así como la secuencia codificadora completa y que incluyen la secuencia colindante a las secuencias de las ESTs 5' empleadas para obtenerlos. Este método se puede aplicar para obtener los ADNc extendidos para cualquier EST 5' en la base de datos NetGene®, incluyendo las ESTs 5' que codifican polipéptidos que pertenecen a las proteínas secretadas. El método se resume en la Figura 3.

1. *Obtención de ADNc extendidos*

a) *Síntesis de la primera cadena*

El método se beneficia de la secuencia 5' conocida del ARNm. Una reacción de transcripción inversa se realiza sobre ARNm purificado con un cebador poli 14dT que contiene una secuencia de 49 nucleótidos en su extremo 5', permitiendo la adición de una secuencia conocida en el extremo del ADNc que se corresponde con el 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' (SEQ ID NO:14). Los expertos en la técnica apreciarán que también se pueden añadir otras secuencias a la secuencia poli dT y se pueden emplear para cebar la síntesis de la primera cadena. Empleando este cebador y una transcriptasa inversa, tal como la enzima Superscript II (Gibco BRL) o la enzima Rnase H Minus M-MLV (Promega), se genera un transcrito inverso anclado al sitio 3' poliA de los ARNs.

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

b) *Síntesis de la segunda cadena*

Una pareja de cebadores anidados en cada extremo, se diseña basándose en la secuencia 5' conocida procedente de la EST 5' y el extremo 3' conocido, añadido por el cebador poli dT empleado en la síntesis de la primera cadena. Los programas empleados para diseñar cebadores se basan en el contenido en GC y en las temperaturas de fusión de los oligonucleótidos, tales como OSP (Illier y Green, *PCR Meth. Appl.* **1**:124-128, 1991) o se basan en el método de la disparidad de la frecuencia de octámeros (Griffais *y col., Nucleic Acids Res.* **19**:3887-3891) tales como PC-Rare (http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html).

Preferentemente, los cebadores anidados en el extremo 5' están separados entre sí por cuatro a nueve bases. Las secuencias 5' de los cebadores se pueden seleccionar para que tengan temperaturas de fusión y especificidades adecuadas para el uso en PCR. Preferentemente, los cebadores anidados en el extremo 3' están separados entre sí por cuatro a nueve bases. Por ejemplo, los cebadores 3' anidados pueden tener las siguientes secuencias: (5'-CCA GCA GAG TCA CGA GAG AGA CTA CAC GG-3' (SEQ ID NO:15) y 5'-CAC GAG AGA GAC TAC ACG GTA CTG G-3' (SEQ ID NO.16). Estos cebadores se seleccionaron debido a que tenían temperaturas de fusión y especificidades compatibles con su uso en la PCR. Sin embargo, los expertos en la técnica apreciarán que también se pueden utilizar otras secuencias como cebadores.

La primera serie con PCR de 25 ciclos se realiza empleando la mezcla de "Advantage Tth Polymerase Mix" (Clontech) y el cebador externo de cada una de las parejas anidadas. Una segunda PCR con 20 ciclos que emplea la misma enzima y el cebador interno de cada una de las parejas anidadas, se realiza a continuación sobre 1/2500 del producto primero de la PCR. A continuación, se eliminan los cebadores y los nucleótidos.

2. *Secuenciación de los ADNc de longitud completa o de fragmentos de los mismos*

Debido a la falta de limitaciones por la posición en el diseño de los cebadores 5' anidados, compatibles para el uso con la PCR empleando el programa OSP, se obtuvieron amplicones de dos tipos. Preferentemente, el segundo cebador 5' está localizado aguas arriba del codón de iniciación de la traducción, obteniéndose de este modo un producto anidado de la PCR que contiene la secuencia codificadora completa. Un ADNc extendido de longitud completa tal, sufre un procedimiento de clonación directa tal y como se describe en la sección a. Sin embargo, en algunos casos, el segundo cebador 5' está localizado aguas abajo del codón de iniciación de la traducción, proporcionando de este modo un producto de la PCR que contiene sólo parte de la ORF. Dichos productos incompletos de la PCR se someten a un procedimiento modificado descrito en la sección b.

a) *Productos anidados de la PCR que contienen ORFs completos*

Cuando el producto anidado resultante de la PCR contiene la secuencia codificadora completa, tal y como se pronostica de la secuencia EST 5', se clona en un vector adecuado tal como pED6dpc2, tal y como se describe en la sección 3.

b) *Productos anidados de la PCR que contienen ORFs incompletos*

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

Una vez que se ha determinado completamente la secuencia codificadora completa, se diseñan nuevos cebadores compatibles para el uso en la PCR para obtener amplicones que contienen la región codificadora completa. Sin embargo, en tales casos, los cebadores 3' compatibles con el uso en la PCR, están localizados dentro de la UTR 3' del ARNm correspondiente, obteniéndose amplicones que carecen de parte de esta región, es decir la zona poliA y a veces la señal de poliadenilación, tal y como se ilustra en la Figura 3. Tales ADNc extendidos de longitud completa se clonan a continuación en un vector adecuado tal y como se describe en la sección 3.

c) *Secuenciación de ADNc extendidos*

La secuenciación de los ADNc extendidos se realiza empleando un planteamiento "Die Terminator" con el equipo de reactivos "AmpliTaq DNA polymerase FS" disponible en Perkin Elmer.

Para secuenciar los fragmentos de la PCR, se realiza un desplazamiento del cebador empleando un programa como OSP, para escoger los cebadores y un programa automático para ordenador tal como ASMG (Sutton *y col., Genome Science Technol*. **1**: 9-19, 1995) para construir cóntigos de secuencias con desplazamiento que incluyen la etiqueta 5' inicial empleando solapamientos mínimos de 32 nucleótidos. Preferentemente, el desplazamiento de cebadores se realiza hasta que se obtienen las secuencias de ADNc de longitud completa.

La terminación de la secuenciación de un fragmento de ADNc extendido dado, se determina del modo siguiente. Ya que las secuencias localizadas después de una zona poliA son difíciles de determinar, precisamente en caso de productos no clonados, los procedimientos de secuenciación y de desplazamiento del cebador para los productos de la PCR se interrumpen cuando se identifica una zona poliA en ADNc extendidos obtenidos tal y como se describe en el caso b. La longitud de la secuencia se compara con el tamaño del producto de la PCR anidado, obtenido tal y como se ha descrito anteriormente. Debido a la exactitud limitada de la determinación del tamaño del producto de la PCR por electroforesis en gel, se considera que una secuencia está completa si el tamaño de la secuencia obtenida es al menos el 70% del tamaño del primer producto de la PCR anidado, estos productos de la PCR se clonan y se determina la secuencia de la inserción. Cuando están disponibles datos de la transferencia de tipo Northern, el tamaño del ARNm detectado para un producto de la PCR dado, se emplea para determinar finalmente si la secuencia está completa. Las secuencias que no cumplen los criterios anteriores, se descartan y se realiza un nuevo proceso de aislamiento.

Los datos de las secuencias de todos los ADNc extendidos se transfieren a continuación a una base de datos adecuada, en donde se realizan controles de la calidad y etapas de validación, tal y como se ha descrito en el Ejemplo 15.

3. *Clonación de los ADNc extendidos de longitud completa*

El producto de la PCR que contiene la secuencia codificadora completa se clona a continuación en un vector adecuado. Por ejemplo, los ADNc extendidos se pueden clonar en el vector de expresión pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) del modo siguiente. El ADN del vector pED6dpc2 se prepara con extremos romos realizando una digestión con EcoRI seguida por un relleno en reacción. El vector que tiene extremos romos se desfosforila. Después de eliminar los cebadores de la PCR y la precipitación con etanol, el producto de la PCR que contiene la secuencia codificadora completa o el ADNc extendido obtenido tal y como se ha descrito anteriormente, se fosforila con una quinasa que se elimina posteriormente con una extracción con fenol-Sevag y precipitación. El ADNc extendido bicatenario se liga a continuación con el vector y el plásmido de expresión resultante se introduce en células hospedadoras adecuadas.

Puesto que los productos de la PCR obtenidos tal y como se ha descrito anteriormente, son moléculas con extremos romos que se pueden clonar en ambas direcciones, se determina la orientación de diversos clones para cada producto de la PCR. Después, se ordenan 4 a 10 clones en placas de microtitulación y se someten a una reacción de PCR empleando un primer cebador localizado en el vector próximo al sitio de clonación y un segundo cebador localizado en la parte del ADNc extendido que se corresponde con el extremo 3' del ARNm. Este segundo cebador puede ser el cebador no codificante empleado en la PCR anclada, en el caso de clonación directa (caso a) o el cebador no codificante localizado dentro de la UTR 3' en el caso de clonación indirecta (caso b). Los clones en los que el codón de iniciación del ADNc extendido están ligados funcionalmente con el promotor en el vector, de modo que se permite la expresión de la proteína codificada por el ADNc extendido, se conservan y se secuencian. Además de los extremos de los insertos de ADNc, también se secuencian aproximadamente 50 pb del vector de ADN en cada lado del inserto de ADNc.

Los productos de la PCR clonados se secuencian a continuación de forma completa según el procedimiento mencionado anteriormente. En este caso, la contigación (formación de cóntigos) de fragmentos largos se realiza a continuación sobre secuencias de desplazamiento que ya se han contigado con productos de la PCR sin clonar, durante el desplazamiento del cebador. La secuenciación de los amplicones clonados se termina cuando los cóntigos resultantes incluyen la región codificadora completa así como secuencias que se solapan con el ADN del vector en ambos extremos.

4. *Análisis con ordenador del ADNc extendido de longitud completa*

Las secuencias de todos los ADNc extendidos de longitud completa se someten a un análisis adicional tal y como se describe a continuación. Antes de la búsqueda de secuencias de interés en los ADNc extendidos de longitud completa, los ADNc extendidos que no son de interés (ARNs del vector, ARNs de transferencia, ARNs ribosómicos, ARNs mitocondriales, ARNs procarióticos y ARNs fúngicos) se descartan empleando métodos esencialmente similares a los descritos para las ESTs 5' en el Ejemplo 18.

a) *Identificación de características estructurales*

Las características estructurales, p. ej., cola poliA y señal de poliadenilación, de las secuencias de ADNc extendidos de longitud completa, se determinan posteriormente del modo siguiente.

Una cola poliA se define como un segmento homopolímero de al menos 11 A que tiene como máximo una base alternativa dentro del mismo. La búsqueda de colas poliA está restringida a los últimos 100 nt de la secuencia y está limitada a los segmentos de 11 As consecutivas porque las reacciones de secuenciación no se pueden leer frecuentemente después de tales segmentos poliA. Los segmentos que tienen más de 90% de homología en 8 nucleótidos, se identifican como colas poliA empleando BLAST2N.

Para buscar una señal de poliadenilación, la cola poliA se corta en la secuencia de longitud completa. En las 50 pb que preceden a la cola poliA se busca en primer lugar la señal canónica de poliadenilación AAUAAA y, en caso de no detectarse la señal canónica, se busca la señal alternativa AUUAAA (Sheets *y col., Nuc. Acids Res.* **18**:5799-5805, 1990). Si no se encuentra ninguna de las señales de consenso de poliadenilación, se busca de nuevo el motivo canónico, permitiendo que un desemparejamiento se considere en los posibles errores de secuenciación. Más del 85% de las señales de poliadenilación identificadas de ambos tipos, termina realmente entre 10 y 30 pb de la cola poliA. Las señales alternativas AUUAAA representan aproximadamente el 15% del número total de señales de poliadenilación identificadas.

b) *Identificación de las características funcionales*

Las características funcionales, p. ej., ORFs y secuencias señal, de las secuencias de ADNc extendidos de longitud completa, se determinaron posteriormente del modo siguiente.

En los 3 marcos de lectura de las hebras superiores de los ADNc extendidos, se buscaron ORFs definidos como los fragmentos de longitud máxima que comienzan con un codón de inicio de la traducción y terminan con un codón de detención. Se prefieren los ORFs que codifican al menos 20 aminoácidos.

Cada ORF encontrado se escanea a continuación en busca de la presencia de un péptido señal en los 50 primeros aminoácidos o, si es adecuado, en regiones más cortas de menos de 20 aminoácidos o menos en el ORF, empleando el método de la matriz de von Heijne (*Nucl. Acids Res.* **14**: 4683-4690, 1986), tal y como se describe en el Ejemplo 22.

c) *Homología con otras secuencias nucleotídicas o proteicas*

La clasificación de las secuencias de longitud completa se puede conseguir con procedimientos esencialmente similares a los descritos para las ESTs 5' del Ejemplo 24.

Los ADNc extendidos preparados tal y como se ha descrito anteriormente, se pueden someter posteriormente a ingeniería genética, para obtener ácidos nucleicos que incluyen partes deseadas del ADNc extendido, empleando técnicas convencionales tales como la subclonación, la PCR o la síntesis de oligonucleótidos *in vitro*. Por ejemplo, los ácidos nucleicos que incluyen sólo las secuencias codificadoras completas (es decir, las secuencias que codifican el péptido señal y la proteína madura que permanecen después de eliminar por corte el péptido señal) se pueden obtener empleando técnicas conocidas por los expertos en la materia. Alternativamente, se pueden aplicar técnicas convencionales para obtener ácidos nucleicos que contienen sólo las secuencias codificadoras para la proteína madura que permanece después de eliminar por corte el péptido señal o los ácidos nucleicos que contienen sólo las secuencias codificadoras para los péptidos señal. De forma similar, se pueden obtener ácidos nucleicos que contienen cualquier otra parte deseada de las secuencias codificadoras de la proteína secretada. Por ejemplo, el ácido nucleico puede contener al menos 10 bases consecutivas de un ADNc extendido tal como uno de los ADNc extendidos descritos a continuación. En otra realización, el ácido nucleico puede contener al menos 15 bases consecutivas de un ADNc extendido tal como uno de los ADNc extendidos descritos a continuación. Alternativamente, el ácido nucleico puede contener al menos 20 bases consecutivas de un ADNc extendido, tal como uno de los ADNc extendidos descritos a continuación. En otra realización, el ácido nucleico puede contener al menos 25 bases consecutivas de un ADNc extendido tal como uno de los ADNc extendidos descritos a continuación. En aún otra realización, el ácido nucleico puede contener al menos 40 bases consecutivas de un ADNc extendido, tal como uno de los ADNc extendidos descritos a continuación.

Una vez que se ha obtenido un ADNc extendido, se puede secuenciar para determinar la secuencia de aminoácidos que codifica. Una vez que se ha determinado la secuencia de aminoácidos codificada, se puede crear e identificar uno cualquiera entre los muchos ADNc concebibles que codificará esa proteína, empleando simplemente la degeneración del código genético. Por ejemplo, se pueden identificar variantes alélicas u otros ácidos nucleicos homólogos, tal y como se describe a continuación. Alternativamente, los ácidos nucleicos que codifican la secuencia de aminoácidos deseada, se pueden sintetizar *in vitro*.

En una realización preferida, la secuencia codificadora se puede seleccionar empleando las preferencias del codón conocido o de parejas de codones para el organismo hospedador en el que se va a expresar el ADNc.

Los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, se obtuvieron tal y como se describe en el Ejemplo 28 siguiente.

Ejemplo 28

Caracterización de los ADNc extendidos clonados obtenidos empleando ESTs 5'

El procedimiento descrito en el Ejemplo 27 anterior, se empleó para obtener los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, en una variedad de tejidos. La siguiente lista proporciona unos pocos ejemplos de tales ADNc extendidos obtenidos.

Empleando este planteamiento, se obtuvo el ADNc de longitud completa de SEQ ID NO: 17 (número de identificación interno 48-19-3-G1-FL1). Este ADNc entra dentro de la categoría "EST-ext" descrita anteriormente y codifica el péptido señal MKKVLLLITAILAVAVG (SEQ ID NO: 18) que tiene una puntuación de von Heijne de 8,2.

El ADNc de longitud completa de SEQ ID NO: 19 (número de identificación interna 58-34-2-E7-FL2) también se obtuvo empleando este procedimiento. Este ADNc entra dentro de la categoría "EST-ext" descrita anteriormente y codifica el péptido señal MWWFQQGLSFLPSALVIWTSA (SEQ ID NO: 20) que tiene una puntuación de von Heijne de 5,5.

Otro ADNc de longitud completa obtenido empleando el procedimiento descrito anteriormente, tiene la secuencia de SEQ ID NO: 21 (número de identificación interna 51-27-1-E8-FL1). Este ADNc entra dentro de la categoría "EST-ext" descrita anteriormente y codifica el péptido señal MVLTTLPSANSANSPVNMPTTGPNSLSYASSALSPCLT (SEQ ID NO: 22) que tiene una puntuación de von Heijne de 5,9.

El procedimiento anterior también se utilizó para obtener un ADNc de longitud completa que tiene la secuencia de SEQ ID NO: 23 (número de identificación interna 76-4-1-G5-FL1). Este ADNc entra dentro de la categoría "EST-ext" descrita anteriormente y codifica el péptido señal ILSTVTALTFAXA (SEQ ID NO: 24) que tiene una puntuación de von Heijne de 5,5.

El ADNc de longitud completa de SEQ ID NO: 25 (número de identificación interna 51-3-3-B10-FL3) también se obtuvo empleando este procedimiento. Este ADNc entra dentro de la categoría "nueva" descrita anteriormente y codifica un péptido señal LVLTLCTLPLAVA (SEQ ID NO: 26) que tiene una puntuación de von Heijne de 10,1. El ADNc de longitud completa de SEQ ID NO: 27 (número de identificación interna 58-35-2-F10-FL2) también se obtuvo empleando este procedimiento. Este ADNc entra dentro de la categoría "nueva" descrita anteriormente y codifica un péptido señal LWLLFFLVTAIHA (SEQ ID NO: 28) que tiene una puntuación de von Heijne de 10,7.

Plásmidos que contienen clones bacterianos que contienen los ADNc de longitud completa descritos anteriormente, están almacenados en la actualidad en los laboratorios de los inventores, con los números de identificación interna proporcionados anteriormente. Los insertos se pueden recuperar de los materiales almacenados, dejando crecer una parte alícuota del clon bacteriano adecuado en el medio adecuado. El ADN plasmídico se puede aislar a continuación empleando procedimientos de aislamiento de plásmidos conocidos por los expertos en la técnica, tales como los minipreparados para lisis alcalina o los procedimientos de aislamiento de plásmidos a gran escala mediante lisis alcalina. Si se desea, el ADN plasmídico se puede enriquecer adicionalmente por centrifugación en un gradiente de cloruro de cesio, por cromatografía de exclusión por tamaño o por cromatografía de intercambio aniónico. El ADN plasmídico obtenido empleando estos procedimientos se puede manipular a continuación empleando técnicas convencionales de clonación, conocidas por los expertos en la técnica. Alternativamente, se puede realizar una PCR con cebadores diseñados en ambos extremos de la inserción del ADNc. El producto de la PCR que se corresponde con el ADNc, se puede manipular después empleando técnicas de clonación convencionales, conocidas por los expertos en la técnica.

En los polipéptidos codificados por los ADNc extendidos se puede escrutar en busca de la presencia de motivos conocidos estructurales o funcionales o en busca de la presencia de distintivos, secuencias cortas de aminoácidos que están bien conservadas entre los miembros de una familia de proteínas. Las regiones conservadas se han utilizado para obtener patrones o matrices de consenso, incluidos en el banco de datos PROSITE, en particular en el archivo prosite.dat (publicación 13.0 de noviembre de 1995, localizada en http://expasy.hcuge.ch/sprot/prosite.html. Los programas prosite_convert y prosite_scan (http://ulrec3.unil.ch/ftpserveur/prosite_scan) se pueden utilizar para encontrar distintivos en los ADNc extendidos.

Para cada patrón obtenido con el programa prosite_convert procedente del archivo prosite.dat, se puede determinar la exactitud de la detección de una nueva secuencia de proteínas, evaluando la frecuencia de éxitos irrelevantes en la población de proteínas humanas secretadas incluidas en la base de datos SWISSPROT. La proporción entre el número de éxitos con las proteínas mezcladas al azar (con una tamaño de ventana de 20 aminoácidos) y el número de éxitos con las proteínas naturales (sin mezclar), se puede utilizar como índice. Cada patrón en el que la proporción es superior a 20% (un éxito en proteínas mezcladas por 5 éxitos en proteínas naturales) se puede evitar durante la búsqueda con prosite_scan. El programa utilizado para mezclar al azar las secuencias de proteínas (db_shuffled) y el programa empleado para determinar las estadísticas para cada patrón en los bancos de datos de las proteínas (prosite_statistics) están disponibles en el sitio ftp, http://ulrec3.unil.ch/ftpserveur/prosite_scan.

Además de los métodos basados en la PCR para obtener ADNc extendidos, también se pueden emplear métodos tradicionales basados en la hibridación. Estos métodos también se pueden utilizar para obtener los ADNs genómicos

que codifican los ARNm a partir de los cuales se han obtenido las ESTs 5', los ARNm correspondientes a los ADNc extendidos o los ácidos nucleicos que son homólogos a los ADNc extendidos o las ESTs 5'. El Ejemplo 29 siguiente proporciona ejemplos de tales métodos.

Ejemplo 29

Métodos para obtener ADNc que incluyen la región codificadora completa y el extremo 5' auténtico del ARNm correspondiente

Una genoteca de ADNc de longitud completa se puede preparar empleando las estrategias descritas en los Ejemplos 13, 14, 15 y 16 anteriores, sustituyendo el nonámero aleatorio empleado en el Ejemplo 14, por un cebador de oligodT. Por ejemplo, se puede utilizar el oligonucleótido de SEQ ID NO: 14.

Alternativamente, se puede obtener una genoteca de ADNc o una genoteca de ADN genómico a partir de una fuente comercial o prepararla empleando técnicas conocidas por los expertos en la materia. Tales genotecas de ADNc o de ADN genómico se pueden aislar empleando ADNc extendidos aislados, obtenidos a partir de EST 5' o de ácidos nucleicos homólogos a ADNc extendidos o a EST 5', del modo siguiente. La genoteca de ADNc o la genoteca de ADN genómico se hibrida con una sonda detectable que comprende al menos 10 nucleótidos consecutivos de la EST 5' o del ADNc extendido, empleando técnicas convencionales. Preferentemente, la sonda comprende al menos 12, 15 o 17 nucleótidos consecutivos procedentes de la EST 5' o del ADNc extendido. Más preferentemente, la sonda comprende al menos 20 a 30 nucleótidos consecutivos procedentes de la EST 5' o del ADNc extendido. En algunas realizaciones, la sonda comprende más de 30 nucleótidos procedentes de la EST 5' o del ADNc extendido.

Las técnicas para identificar clones de ADNc en una genoteca de ADNc que se hibrida con una secuencia de una sonda dada, se describen en Sambrook *y col., Molecular Cloning: A Laboratory Manual 2ª ed.*, Cold Spring Harbor Laboratory Press, 1989. Se pueden utilizar las mismas técnicas para aislar ADNs genómicos.

Resumiendo, los clones de ADNc o de ADN genómico que se hibridan con la sonda detectable, se identifican y se aíslan para una manipulación posterior del modo siguiente. Una sonda que comprende al menos 10 nucleótidos consecutivos procedentes de la EST 5' o del ADNc extendido, se marca con una etiqueta detectable, tal como un radioisótopo o una molécula fluorescente. Preferentemente, la sonda comprende al menos 12, 15 o 17 nucleótidos consecutivos procedentes de la EST 5' o del ADNc extendido. Más preferentemente, la sonda comprende 20 a 30 nucleótidos consecutivos procedentes de la EST 5' o del ADNc extendido. En algunas realizaciones, la sonda comprende más de 30 nucleótidos procedentes de la EST 5' o del ADNc extendido.

Las técnicas para marcar la sonda son bien conocidas e incluyen la fosforilación con polinucleótido quinasa, el traslado de muescas, la transcripción *in vitro* y técnicas no radiactivas. Los ADNc o los ADNs genómicos en la genoteca se transfieren a un filtro de nitrocelulosa o de nailon y se desnaturalizan. Después de bloquear los sitios no específicos, el filtro se incuba con la sonda marcada durante el tiempo suficiente para permitir la unión de la sonda a los ADNc o ADNs genómicos que contienen una secuencia capaz de hibridarse con la misma.

Variando las condiciones restrictivas de la hibridación empleada para identificar ADNc extendidos o ADNs genómicos que se hibridan con la sonda detectable, se pueden identificar y aislar los ADNc extendidos que tienen niveles diferentes de homología con la sonda, del siguiente modo.

1. *Identificación de las secuencias de ADNc extendido o de ADNc genómico que tienen un alto grado de homología con la sonda marcada*

Para identificar los ADNc extendidos o los ADNs genómicos que tienen un alto grado de homología con la secuencia de la sonda, se puede calcular la temperatura de fusión de la sonda empleando las siguientes fórmulas:

Para sondas que tienen entre 14 y 70 nucleótidos de longitud, la temperatura de fusión (Tf) se calcula empleando la fórmula: Tf=81,5+16,6(log[Na+])+0,41 (fracción G+C)-(600/N) en donde N es la longitud de la sonda.

Si la hibridación se realiza en una solución que contiene formamida, la temperatura de fusión se puede calcular empleando la ecuación Tf=81,5+16,6(log[Na+])+0,41 (fracción G+C)-(0,63% de formamida)-(600/N) en donde N es la longitud de la sonda.

La prehibridación se puede realizar en 6x SSC, 5x reactivo de Denhardt, SDS al 0,5%, 100 μg de ADN de esperma de salmón desnaturalizado y fragmentado o 6x SSC, 5x reactivo de Denhardt, 0,5% de SDS, 100 μg de ADN de esperma de salmón desnaturalizado y fragmentado, 50% de formamida. Las fórmulas para las soluciones de SSC y de Denhardt se exponen en Sambrook *y col., véase más arriba*.

La hibridación se realiza añadiendo la sonda detectable a las soluciones de prehibridación expuestas anteriormente. Cuando la sonda comprende ADN bicatenario, se desnaturaliza antes de añadirlo a la solución de hibridación. El filtro se pone en contacto con la solución de hibridación durante un periodo de tiempo suficiente para permitir que la sonda se hibride con las secuencias que contienen los ADNc extendidos o los ADNs genómicos que son complementarias a la misma o homólogas a la misma. Para sondas de más de 200 nucleótidos de longitud, la hibridación se puede realizar

a 15-25ºC por debajo de la Tf. Para sondas más cortas, tales como sondas de oligonucleótidos, la hibridación se puede realizar a 15-25ºC por debajo de la Tf. Preferentemente, para hibridaciones en 6x SSC, la hibridación se realiza a aproximadamente 68ºC. Preferentemente, para hibridaciones en soluciones que contienen formamida al 50%, la hibridación se realiza a aproximadamente 42ºC.

Todas las hibridaciones anteriores se podrían considerar que están bajo condiciones "restrictivas".

Después de la hibridación, el filtro se lava en 2x SSC, SDS al 0,1% a temperatura ambiente durante 15 minutos. El filtro se lava a continuación con 0,1x SSC, SDS al 0,5% a temperatura ambiente durante 30 minutos a 1 hora. A continuación, la solución se lava a la temperatura de la hibridación en 0,1x SSC, SDS al 0,5%. Un lavado final se realiza en 0,1x SSC a temperatura ambiente.

Los ADNc extendidos, los ácidos nucleicos homólogos a los ADNc extendidos o las ESTs 5' o los ADNs genómicos que se han hibridado con la sonda, se identifican mediante autorradiografía u otras técnicas convencionales.

2. *Obtención de las secuencias de ADNc extendido o de ADNc genómico que tienen grados inferiores de homología con la sonda marcada*

El procedimiento anterior se puede modificar para identificar ADNc extendidos, ácidos nucleicos homólogos a ADNc extendidos o ADNs genómicos que tienen niveles inferiores de homología con la secuencia de la sonda. Por ejemplo, para obtener ADNc extendidos, ácidos nucleicos homólogos a ADNc extendidos o ADNs genómicos con homología inferior con la sonda detectable, se pueden utilizar condiciones menos restrictivas. Por ejemplo, la temperatura de hibridación se puede disminuir en incrementos de 5ºC desde 68ºC hasta 42ºC en un tampón de hibridación que tenga una concentración de sodio de aproximadamente 1 M. Después de la hibridación, el filtro se puede lavar con 2x SSC, 0,5% de SDS a la temperatura de la hibridación. Estas condiciones se consideran que son condiciones "moderadas" por encima de 50ºC y condiciones "reducidas" por debajo de 50ºC.

Alternativamente, la hibridación se puede realizar en tampones, tales como 6x SSC, que contienen formamida a una temperatura de 42ºC. En este caso, la concentración de formamida en el tampón de hibridación se puede reducir en incrementos del 5% desde 50% hasta 0% para identificar clones que tienen niveles decrecientes de homología con la sonda. Después de la hibridación, el filtro se puede lavar con 6x SSC, SDS al 0,5% a 50ºC. Estas condiciones se consideran que son condiciones "moderadas" por encima de 25% de formamida y condiciones "reducidas" por debajo de 25% de formamida.

Los ADNc extendidos, los ácidos nucleicos homólogos a los ADNc extendidos o los ADNs genómicos que se han hibridado con la sonda, se identifican por autorradiografía.

3. *Determinación del grado de homología entre los ADNc extendidos obtenidos y la sonda marcada*

Si se desea obtener ácidos nucleicos homólogos a ADNc extendidos, tales como variantes alélicas de los mismos o ácidos nucleicos que codifican proteínas relacionadas con las proteínas codificadas por los ADNc extendidos, se puede determinar adicionalmente el nivel de homología entre el ácido nucleico hibridado y el ADNc extendido o la EST 5' empleada como sonda, empleando BLAST2N; los parámetros se pueden adaptar dependiendo de la longitud de la secuencia y del grado de homología estudiado. Para determinar el nivel de homología entre el ácido nucleico hibridado y el ADNc extendido o la EST 5' a partir de la cual se ha obtenido la sonda, se comparan las secuencias de nucleótidos del ácido nucleico hibridado y el ADNc extendido o la EST 5' a partir de la cual se ha obtenido la sonda. Por ejemplo, empleando los métodos anteriores, se pueden obtener e identificar los ácidos nucleicos que tienen al menos 95% de homología en los ácidos nucleicos con el ADNc extendido o la EST 5' a partir de la cual se ha obtenido la sonda. De forma similar, empleando condiciones de hibridación que sean progresivamente menos restrictivas, se pueden obtener e identificar ácidos nucleicos que tengan al menos 90%, al menos 85%, al menos 80% o al menos 75% de homología con el ADNc extendido o la EST 5' a partir de la cual se ha obtenido la sonda.

Para determinar si un clon codifica una proteína que tiene una cantidad dada de homología con la proteína codificada por el ADNc extendido o la EST 5', la secuencia de aminoácidos codificada por el ADNc extendido o la EST 5' se compara con la secuencia de aminoácidos codificada por el ácido nucleico que se hibrida. La homología se determina como existente cuando una secuencia de aminoácidos en el ADNc extendido o la EST 5' está estrechamente relacionada con una secuencia de aminoácidos en el ácido nucleico que se hibrida. Una secuencia está estrechamente relacionada cuando es idéntica a la del ADNc extendido o la EST 5' o cuando contiene una o varias sustituciones de aminoácidos en la misma, en las que los aminoácidos que tienen características similares se han sustituido por otro. Empleando los métodos anteriores y algoritmos tales como FASTA con parámetros que dependen de la longitud de la secuencia y del grado de homología estudiado, se pueden obtener ácidos nucleicos que codifican proteínas que tienen al menos 95%, al menos 90%, al menos 85%, al menos 80% o al menos 75% de homología con las proteínas codificadas por el ADNc extendido o la EST 5' a partir de la cual se había obtenido la sonda.

Además de los métodos descritos anteriormente, están disponibles otros protocolos para obtener ADNc extendidos empleando ESTs 5', tal y como se expone en los siguientes párrafos.

Los ADNc extendidos se pueden preparar obteniendo ARNm del tejido, de la célula o del organismo de interés, empleando procedimientos para la preparación de ARNm que emplean procedimientos de selección con poliA u otras técnicas conocidas por los expertos en la materia. Un primer cebador capaz de hibridarse con la cola poliA del ARNm, se hibrida con el ARNm y se realiza una reacción de transcripción inversa para generar una primera hebra de ADNc.

La primera hebra de ADNc se hibrida con un segundo cebador que contiene al menos 10 nucleótidos consecutivos de la secuencia de SEQ ID NO: 38. Preferentemente, el cebador comprende al menos 12, 15 o 17 nucleótidos consecutivos de la secuencia SEQ ID NO: 38. Más preferentemente, el cebador comprende 20 a 30 nucleótidos consecutivos de la secuencia SEQ ID NO: 38. En algunas realizaciones, el cebador comprende más de 30 nucleótidos de la secuencia SEQ ID NO: 38. Si se desea obtener ADNc extendidos que contienen la secuencia que codifica la proteína de longitud completa, que incluye el sitio auténtico de inicio de la traducción, el segundo cebador empleado contiene secuencias localizadas aguas arriba del sitio de inicio de la traducción. El segundo cebador se extiende para generar una segunda hebra de ADNc complementaria a la primera hebra de ADNc. Alternativamente, se puede realizar una RT-PCR tal y como se ha descrito anteriormente empleando cebadores de ambos extremos del ADNc que se va a obtener.

Los ADNc extendidos que contienen fragmentos 5' del ARNm, se pueden preparar hibridando un ARNm que comprende la secuencia de la EST 5' para la que se desea un ADNc extendido, con un cebador que comprende al menos 10 nucleótidos consecutivos de las secuencias complementarias a la EST 5' y transcribir de forma inversa el cebador hibridado para preparar una primera hebra de ADNc a partir de los ARNm. Preferentemente, el cebador comprende al menos 12, 15 o 17 nucleótidos consecutivos procedentes de la EST 5'. Más preferentemente, el cebador comprende 20 a 30 nucleótidos consecutivos de la EST 5'.

A continuación, se sintetiza una segunda hebra de ADNc complementaria a la primera hebra de ADNc. La segunda hebra de ADNc se puede preparar hibridando un cebador complementario a secuencias en la primera hebra de ADNc y extendiendo el cebador para generar la segunda hebra de ADNc.

Los ADNc extendidos bicatenarios preparados empleando los métodos descritos anteriormente, se aíslan y se clonan. Los ADNc extendidos se pueden clonar en vectores tales como plásmidos o vectores víricos capaces de replicarse en una célula hospedadora adecuada. Por ejemplo, la célula hospedadora puede ser una célula bacteriana, de mamífero, de ave o de insecto.

Las técnicas para aislar ARNm, para la transcripción inversa de un cebador que se hibrida con ARNm para generar una primera cadena de ADNc, para extender un cebador para preparar una segunda hebra de ADNc complementaria a la primera hebra de ADNc, para aislar el ADNc bicatenario y clonar el ADNc bicatenario, son bien conocidas por los expertos en la materia y se describen en *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. 1997 y Sambrook *y col., Molecular Cloning: A Laboratory Manual*, 2ª edición, Cold Spring Harbor Laboratory Press, 1989.

Alternativamente, procedimientos tales como los descritos en el Ejemplo 29 se pueden utilizar para obtener ADNc de longitud completa o ADNc extendidos. En este planteamiento, los ADNc de longitud completa o extendidos se preparan a partir de ARNm y se clonan en fagémidos bicatenarios del modo siguiente. La genoteca de ADNc en los fagémidos bicatenarios se vuelve monocatenaria mediante tratamiento con una endonucleasa, tal como el producto del Gen II del fago F1 y una exonucleasa (Chang *y col., Gene* **127**:95-8, 1993). Un oligonucleótido biotinilado que comprende la secuencia de una EST 5' o un fragmento que contiene al menos 10 de sus nucleótidos, se hibrida con los fagémidos monocatenarios. Preferentemente, el fragmento comprende al menos 12, 15 o 17 nucleótidos consecutivos de la EST 5'. Más preferentemente, el fragmento comprende 20-30 nucleótidos consecutivos procedentes de la EST 5'. En algunos procedimientos, el fragmento puede comprender más de 30 nucleótidos consecutivos procedentes de la EST 5'.

Los híbridos entre el oligonucleótido biotinilado y los fagémidos que tienen insertos que contienen la secuencia de EST 5', se aíslan incubando los híbridos con perlas paramagnéticas revestidas con estreptavidina y recuperando las perlas con un imán (Fry *y col., Biotechniques*, **13**:124-131, 1992). A continuación, los fagémidos resultantes que contienen la secuencia de la EST 5' se liberan de las perlas y se convierten en ADN bicatenario empleando un cebador específico para la secuencia EST 5'. Alternativamente, se pueden utilizar protocolos como el del equipo de reactivos de Gene Trapper (Gibco BRL). El ADN bicatenario resultante se transforma en bacterias. Los ADNc extendidos que contienen la secuencia de EST 5' se identifican mediante PCR de colonias o por hibridación de colonias.

Empleando cualquiera de los métodos descritos anteriormente en la sección III, se puede proporcionar una variedad de ADNc extendidos que contienen secuencias que codifican la proteína de longitud completa o secuencias que codifican sólo la proteína madura, cortando la parte que queda detrás del péptido señal, en forma de genotecas de ADNc para la posterior evaluación de las proteínas codificadas o emplear en ensayos de diagnóstico tal y como se describe a continuación.

IV. *Expresión de proteínas codificadas por ADNc extendidos, aislados empleando ESTs 5'*

Los ADNc extendidos que contienen secuencias que codifican proteínas completas de sus ARNm correspondientes o de partes de los mismos, tales como los ADNc que codifican la proteína madura, se pueden utilizar para expresar las proteínas secretadas codificadas o partes de las mismas, tal y 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 una variedad de ADNc extendidos que contienen las secuencias codificadoras de la proteína completa o partes de las mismas, se pueden clonar simultáneamente en vectores de expresión para crear una genoteca de expresión para el análisis de las proteínas codificadas, tal y como se describe más adelante.

Ejemplo 30

Expresión de las proteínas codificadas por los genes correspondientes a las ESTs 5' o a partes de las mismas

Para expresar las proteínas codificadas por los genes correspondientes a las ESTs 5' (o a partes de las mismas), los ADNc de longitud completa que contienen la región que codifica la proteína completa o los ADNc extendidos que contienen secuencias adyacentes a las ESTs 5' (o a partes de las mismas), se obtienen tal y como se ha descrito 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. Los ácidos nucleicos insertados en los vectores de expresión también pueden contener secuencias aguas arriba de las secuencias que codifican el péptido señal, tales como las secuencias que regulan los niveles de expresión o secuencias que confieren una expresión específica de tejido.

El ácido nucleico que codifica la proteína o el polipéptido que se va a expresar, se liga funcionalmente a un promotor en un vector de expresión, empleando tecnología de clonación convencional. El vector de expresión puede ser cualquiera de los sistemas de expresión de mamíferos, de levadura, de insectos o de bacterias, conocidos en la técnica. Están a disposición comercial vectores y sistemas de expresión comercializados por una variedad de proveedores que incluyen Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin) e Invitrogen (San Diego, California). Si se desea, para mejorar la expresión y facilitar un replegado adecuado de la proteína, el contexto de los codones y el emparejamiento de los codones de la secuencia se pueden mejorar para el organismo de expresión en particular, en el que se introduce el vector de expresión, tal y como aclaran Hatfield *y col.,* documento de Patente de EE.UU. nº 5.082.767.

El ADNc clonado en el vector de expresión puede codificar la proteína completa (es decir, el péptido señal y la proteína madura), la proteína madura (es decir, la proteína creada al separar por corte el péptido señal), sólo el péptido señal o cualquier otra parte de la misma.

Lo siguiente se proporciona como un método ejemplar para expresar las proteínas codificadas por los ADNc extendidos correspondientes las ESTs 5' o a los ácidos nucleicos descritos anteriormente. En primer lugar, se identifican el codón de iniciación de metionina para el gen y la señal poliA del gen. Si el ácido nucleico que codifica el polipéptido que se va a expresar carece de una metionina que sirva como sitio de iniciación, se puede introducir una metionina de iniciación junto al primer codón del ácido nucleico, empleando técnicas convencionales. De forma similar, si el ADNc extendido carece de señal poliA, esta secuencia se puede añadir a la estructura artificial, por ejemplo, cortando y empalmando la señal poliA a partir de pSG5 (Stratagene) empleando las enzimas de restricción BglII y SalI e incorporándola en el vector de expresión de mamífero, pXT1 (Stratagene). pXT1 contiene las LTRs y una parte del gen *gag* procedente del virus de la leucemia de múridos de Moloney. La posición de las LTRs en la estructura artificial permite una transfección estable y eficaz. El vector incluye el promotor de la quinasa de timidina de Herpes Simple y el gen seleccionable de la neomicina. El ADNc extendido o una parte del mismo que codifica el polipéptido que se va a expresar, se obtiene mediante PCR a partir del vector bacteriano, empleando cebadores oligonucleotídicos complementarios al ADNc extendido o a partes del mismo y contiene secuencias de las endonucleasas de restricción, PstI incorporada en el cebador 5' y BglII en el extremo 5' del cebador 3' del ADNc correspondiente, teniendo cuidado de asegurar que el ADNc extendido se sitúa con la señal poliA. El fragmento purificado obtenido a partir de la reacción de la PCR resultante, se digiere con PstI, se forman extremos romos con una exonucleasa, se digiere con BglII, se purifica y se liga a pXT1 que contiene una señal poliA y se prepara para esta ligación (romo/BglII).

El producto ligado se transfecta a células de ratón NIH 3T3 empleando Lipofectina (Life Technologies, Inc., Grand Island, Nueva York) bajo las condiciones especificadas en la descripción del producto. Los transfectantes positivos se seleccionan después de crecer las células transfectadas en 600 μg/ml de G418 (Sigma, St. Louis, Missouri). Preferentemente, la proteína expresada se libera al medio de cultivo, facilitando de este modo la purificación.

Alternativamente, los ADNc extendidos se pueden clonar en pED6dpc2 tal y como se ha descrito anteriormente. Las estructuras artificiales de pED6dpc2 se pueden transfectar en una célula hospedadora adecuada, tal como células COS 1. Las células resistentes al metotrexato se seleccionan y se expanden. Preferentemente, la proteína expresada procedente del ADNc extendido, se libera al medio de cultivo facilitando de este modo la purificación.

Las proteínas en el medio se cultivo se separan mediante electroforesis en gel. Si se desea, las proteínas se pueden precipitar con sulfato de amonio o separar basándose en su tamaño o su carga antes de la electroforesis.

Como testigo, el vector de expresión que carece de un inserto de ADNc, se introduce en las células o en los organismos hospedadores y se recolectan las proteínas en el medio. Las proteínas secretadas presentes en el medio se detectan empleando técnicas conocidas por los expertos en la técnica, tales como la tinción con azul de Coomassie o con plata o empleando anticuerpos contra la proteína codificada por el ADNc extendido.

Los anticuerpos capaces de reconocer específicamente la proteína de interés se pueden generar empleando péptidos 15-meros sintéticos que tienen una secuencia codificada por la EST 5' adecuada, el ADNc extendido o partes del mismo. Los péptidos sintéticos se inyectan en ratones para generar el anticuerpo contra el polipéptido codificado por la EST 5', el ADNc extendido o partes de los mismos.

Las proteínas secretadas por las células o los organismos hospedadores que contienen un vector de expresión que contiene el ADNc extendido obtenido a partir de una EST 5' o una parte de la misma, se comparan con las de las células o los organismos testigos. La presencia de una banda en el medio procedente de las células que contienen el vector de expresión que está ausente en el medio procedente de las células testigo, indica que el ADNc extendido codifica una proteína secretada. Generalmente, la banda correspondiente a la proteína codificada por el ADNc extendido, tendrá una movilidad próxima a la esperada, basándose en el número de aminoácidos en el marco de lectura abierto del ADNc extendido. Sin embargo, la banda puede tener una movilidad diferente que la esperada a causa de modificaciones tales como glicosilación, ubiquitinación o escisión enzimática.

Alternativamente, si la proteína expresada por los vectores de expresión anteriores no contiene secuencias que dirigen su secreción, las proteínas expresadas por las células hospedadoras que contienen un vector de expresión con un inserto que codifica una proteína secretada o una parte de la misma, se pueden comparar con las proteínas expresadas en las células hospedadoras testigo que contienen el vector de expresión sin inserto. La presencia de una banda en muestras procedentes de células que contienen el vector de expresión con un inserto que está ausente en muestras procedentes de células que contienen el vector de expresión sin un inserto, indica que la proteína deseada o una parte de la misma se está expresando. Generalmente, la banda tendrá la movilidad esperada para la proteína secretada o para una parte de la misma. Sin embargo, la banda puede tener una movilidad diferente que la esperada como resultado de modificaciones tales como glicosilación, ubiquitinación o escisión enzimática.

La proteína codificada por el ADNc extendido se puede purificar empleando técnicas de inmunocromatografía convencionales. En tales procedimientos, una solución que contiene la proteína secretada, tal como el medio de cultivo o un extracto celular, se aplica a una columna que tenga anticuerpos contra la proteína secretada fijada a la matriz de la cromatografía. Se permite que la proteína secretada se una a la columna de la inmunocromatografía. A continuación, la columna se lava para retirar las proteínas unidas de forma no específica. La proteína secretada unida específicamente se libera a continuación de la columna y se recupera empleando técnicas convencionales.

Si no es posible la producción de anticuerpos, la secuencia de ADNc extendida o una parte de la misma se puede incorporar en vectores de expresión diseñados para emplear en esquemas de purificación que utilizan polipéptidos quiméricos. En tales estrategias, la secuencia codificadora del ADNc extendido o una parte del mismo, se inserta en marco de lectura con el gen que codifica la otra mitad de la quimera. La otra mitad de la quimera puede ser β-globina o un polipéptido que se une a níquel. Una matriz de cromatografía que tiene el anticuerpo de la β-globina o del níquel fijado a la misma, se utiliza a continuación para purificar la proteína quimérica. Los sitios de corte de la proteasa se pueden preparar por ingeniería genética entre el gen de la β-globina o el polipéptido que se une al níquel y el ADNc extendido o una parte del mismo. Por tanto, los dos polipéptidos de la quimera se pueden separar entre sí por digestión con proteasas.

Un vector de expresión útil para generar quimeras con β-globina es pSG5 (Stratagene) que codifica la β-globina de ratón. El intrón II del gen de la β-globina de ratón facilita el corte y empalme del transcrito expresado y la señal de poliadenilación incorporada en la estructura artificial incrementa el nivel de expresión. Estas técnicas descritas son bien conocidas por los expertos en biología molecular. Los métodos convencionales se publican en textos de métodos, tales como Davis y *col., (Basic Methods in Molecular Biology*, Davis, Dibner y Battey, compiladores, Elsevier Press, NY, 1986) y muchos de los métodos los tiene disponibles Stratagene, Life Technologies, Inc., o Promega. El polipéptido se puede producir adicionalmente a partir de la estructura artificial empleando sistemas de traducción *in vitro*, tales como el equipo de reactivos de "*In vitro* Express® Translation Kit" (Stratagene).

Después de la expresión y la purificación de las proteínas secretadas codificadas por las ESTs 5', los ADNc extendidos o fragmentos de los mismos, las proteínas purificadas se pueden someter a ensayo para estudiar la capacidad de unirse a la superficie de diversos tipos celulares, tales como los descritos en el Ejemplo 31 siguiente. Se apreciará que una variedad de proteínas expresadas a partir de estos ADNc, se puede incluir en un panel de proteínas para evaluar simultáneamente las actividades descritas específicamente a continuación, así como otras funciones biológicas para las que están disponibles ensayos para determinar la actividad.

Ejemplo 31

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

Las proteínas codificadas por las ESTs 5', los ADNc extendidos o fragmentos de los mismos se clonan en vectores de expresión tales como los descritos en el Ejemplo 30. Las proteínas se purifican según su tamaño, su carga, su inmunocromatografía u otras técnicas conocidas por los expertos en la técnica. Después de la purificación, las proteínas se marcan empleando técnicas conocidas por los expertos en la técnica. Las proteínas marcadas se incuban con células o con líneas celulares obtenidas a partir de una variedad de órganos o tejidos, para permitir que las proteínas se unan a cualquier receptor presente sobre la superficie celular. Después de la incubación, las células se lavan para eliminar la proteína no unida específicamente. Las proteínas marcadas se detectan por autorradiografía. Alternativamente, las

proteínas no marcadas se pueden incubar con las células y se pueden detectar con anticuerpos que tienen una etiqueta detectable, tal como una molécula fluorescente, fijada a las mismas.

La especificidad de la unión a la superficie celular se puede analizar realizando un análisis de la competición en el que diversas cantidades de proteína sin marcar se incuban junto con la proteína marcada. La cantidad de proteína marcada unida a la superficie celular disminuye cuando aumenta la cantidad de proteína no marcada competitiva. Como testigo, se incluyen diversas cantidades de una proteína no marcada que no está relacionada con la proteína marcada, en algunas reacciones de ligación. La cantidad de proteína marcada unida a la superficie celular no disminuye en las reacciones de ligación que contienen cantidades crecientes de proteína sin marcar no relacionada, indicando que la proteína codificada por el ADNc se une específicamente a la superficie celular.

Tal y como se ha descrito anteriormente, las proteínas secretadas han mostrado tener una cantidad de efectos fisiológicos importantes y, por ello, representan una valiosa fuente terapéutica. Las proteínas secretadas codificadas por los ADNc extendidos o por partes de los mismos, preparados según los Ejemplos 27-29, se pueden evaluar para determinar sus actividades fisiológicas tal y como se describe a continuación.

Ejemplo 32

Ensayos en las proteínas expresadas procedentes de ADNc extendidos o de partes de los mismos para estudiar la actividad de citoquina, de proliferación celular o de diferenciación celular

Tal y como se ha expuesto anteriormente, las proteínas secretadas pueden actuar como citoquinas o pueden afectar a la proliferación o a la diferenciación celular. Muchos factores proteicos descubiertos hasta la fecha que incluyen todas las citoquinas conocidas, han mostrado actividad en ensayos de proliferación celular dependientes de uno o varios factores y, por consiguiente, los ensayos sirven como una confirmación adecuada de la actividad como citoquinas. La actividad de una proteína codificada por los ADNc extendidos se comprueba por uno cualquiera entre una cantidad de ensayos rutinarios de la proliferación celular dependientes de factor, para líneas celulares que incluyen, sin limitación, 32D, DA2, SA1G, 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 por partes de los mismos se puede evaluar su capacidad para regular la proliferación de linfocitos T o timocitos en ensayos tales como los descritos anteriormente o en las referencias siguientes: *Current Protocols in Immunology*, compiladores Coligan *y col.,* Greene Publishing Associates and Wiley-Interscience; Takai *y col., J. Immunol.* **137**:3494-3500, 1986, Bertagnolli *y col., J. Immunol.* **145**:1706-1712, 1990, Bertagnolli *y col., Cell. Immunol.* **133**:327-341, 1991; Bertagnolli *y col., J. Immunol.* **149**:3778-3783, 1992; Bowman *y col., J. Immunol*. **152**:1756-1761, 1994.

Además, se conocen numerosos ensayos para la producción de citoquinas y/o la proliferación de células del bazo, células de nódulos linfáticos y timocitos. Estos incluyen las técnicas descritas en *Current Protocols in Immunology, véase más arriba* **1**:3.12.1-3.12.14; y Schreiber en *Current Protocols in Immunology, véase más arriba* **1**:6.8.1-6.8.8.

En las proteínas codificadas por los ADNc también se puede someter a ensayo la capacidad de regular la proliferación y la diferenciación de células hematopoyéticas o linfopoyéticas. Muchos ensayos de dicha actividad son conocidos por los expertos en la técnica, incluyendo los ensayos de las siguientes referencias: Bottomly *y col., Current Protocols in Immunology, véase más arriba* **1**:6.3.1-6.3.12; de Vries *y col., J. Exp. Med*. **173**:1205-1211, 1991; Moreau *y col., Nature.* **36**:690-692, 1988; Greenberger *y col., Proc. Natl. Acad. Sci. USA*. **80**:2931-2938, 1983; Nordan, R., en *Current Protocols in Immunology, véase más arriba* **1**:6.6.1-6.6.5; Smith *y col., Proc. Natl. Acad. Sci. USA* **83**:1857-1861, 1986; Bennett *y col., Current Protocols in Immunology, véase más arriba* **1**:6.15.1; Ciarletta *y col., Current Protocols in Immunology, véase más arriba*, **1**:6.13.1.

En las proteínas codificadas por los ADNc también se puede someter a ensayo su capacidad para regular las respuestas de linfocitos T frente a antígenos. Muchos ensayos para dicha actividad son conocidos por los expertos en la técnica, 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 el Capítulo 7 (Immunologic Studies in Humans) en *Current Protocols in Immunology, véase más arriba*; Weinberger *y col., Proc. Natl. Acad. Sci. USA*. **77**:6091-6095, 1980; Weinberger *y col., Eur. J. Immun.* **1**:405-411, 1981; Takai *y col., J. Immunol*. **137**:3494-3500, 1986; Takai *y col., J. Immunol.* **140**:508-512, 1988.

Las proteínas que muestran actividad de citoquina, de proliferación celular o de diferenciación celular se pueden formular a continuación como agentes farmacéuticos y emplear para tratar estados clínicos en los que sea beneficioso la inducción de la proliferación o la diferenciación celular. Alternativamente, tal y como se describe con más detalle a continuación, los genes que codifican estas proteínas o los ácidos nucleicos que regulan la expresión de estas proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o disminuir la expresión de las proteínas en la forma deseada.

Ejemplo 33

Ensayos en las proteínas expresadas procedentes de ADNc extendidos o de partes de los mismos para estudiar la actividad como reguladores del sistema inmune

Las proteínas codificadas por los ADNc también se pueden evaluar de cara a sus efectos como inmunorreguladoras. Por ejemplo, en las proteínas se puede evaluar su actividad para influir sobre la citotoxicidad de esplenocitos o timocitos. Numerosos ensayos de dicha actividad son conocidos por los expertos en la técnica, 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*, Coligan *y col.,* compiladores, Greene Publishing Associates and Wiley-Interscience; Herrmann *y col., Proc. Natl. Acad. Sci. USA* **78**:2488-2492, 1981; Herrmann *y col., J. Immunol.* **128**:1968-1974, 1982; Handa *y col., J. Immunol*. **135**:1564-1572, 1985; Takai *y col., J. Immunol* **137**:3494-3500, 1986; Takai *y col., J. Immunol.* **140**:508-512, 1988; Bowman *y col., J. Virology* **61**:1992-1998; Bertagnolli *y col., Cell. Immunol*. **133**:327-341, 1991; Brown *y col., J. Immunol.* **153**:3079-3092, 1994.

En las proteínas codificadas por los ADNc también se puede evaluar su efecto sobre respuestas de inmunoglobulinas dependientes de linfocitos T y sobre el cambio de isotipos. Numerosos ensayos de esta actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Maliszewski, *J. Immunol.* **144**:3028-3033, 1990; Mond *y col., Current Protocols in Immunology* **1**:3.8.1-3.8.16, *véase más arriba*.

En las proteínas codificadas por los ADNc también se puede evaluar su efecto sobre células inmunoefectoras, incluyendo su efecto sobre las células Th1 y los linfocitos citotóxicos. Numerosos ensayos para esta actividad son conocidos por los expertos en la técnica, 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, véase más arriba*; Takai *y col., J. Immunol*. **137**:3494-3500, 1986; Takai *y col., J. Immunol.* **140**:508-512, 1988; Bertagnolli *y col., J. Immunol.* **149**:3778-3783, 1992.

En las proteínas codificadas por los ADNc también se puede evaluar su efecto sobre la activación mediada por células dendríticas de linfocitos T sin tratamiento previo. Numerosos ensayos de dicha actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Guery *y col., J. Immunol.* **134**:536-544, 1995; Inaba *y col., J. Exp. Med.* **173**:549-559, 1991; Macatonia *y col., J. Immunol*. **154**:5071-5079, 1995; Porgador *y col., J. Exp. Med*. **182**:255-260, 1995; Nair *y col., J. Virol.* **67**:4062-4069, 1993; Huang *y col., Science* **264**:961-965, 1994; Macatonia *y col., J. Exp. Med*. **169**:1255-1264, 1989; Bhardwaj *y col., Journal of Clinical Investigation*. **94**:797-807, 1994; e Inaba *y col., J. Exp. Med.* **172**:631-640, 1990.

En las proteínas codificadas por los ADNc también se puede evaluar su influencia sobre el tiempo de vida de los linfocitos. Numerosos ensayos de dicha actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Darzynkiewicz *y col., Cytometry.* **13**:795-808, 1992; Gorczyca *y col., Leukemia* **7**:659-670, 1993; Gorczyca *y col., Cancer Res.* **53**:1954-1951, 1993; Itoh *y col., Cell* **66**:233-243, 1991; Zacharchuk*, J. Immunol.* **145**:4037-4045, 1990; Zamai *y col., Cytometry* **14**:891-897, 1993; Gorczyca *y col., Int. J. Oncol.* **1**:639-648, 1992.

En las proteínas codificadas por los ADNc también se puede evaluar su influencia sobre etapas tempranas de programación y de desarrollo de linfocitos T. Numerosos ensayos de dicha actividad son conocidos por los expertos en la técnica, incluyendo sin limitación los ensayos descritos en las siguientes referencias: Antica *y col., Blood* **84**:111-117, 1994; Fine *y col., Cell. Immunol.* **155**:111-122, 1994; Galy *y col., Blood* **85**:2770-2778, 1995; Toki *y col., Proc. Natl. Acad. Sci. USA* **88**:7548-7551, 1991.

Las proteínas que muestran actividad como reguladoras del sistema inmune se pueden formular a continuación como agentes farmacéuticos y emplear para tratar estados clínicos en los que es beneficiosa la regulación de la actividad inmune. Por ejemplo, la proteína puede ser útil en el tratamiento de diversas deficiencias y enfermedades inmunes (incluyendo la inmunodeficiencia grave combinada), p. ej., regulando (infra o supra) el crecimiento y la proliferación de los linfocitos T y/o B, así como efectuando la actividad citolítica de las células NK y de otras poblaciones celulares. Estas deficiencias inmunitarias pueden ser genéticas o estar causadas por infecciones víricas (p. ej., VIH) así como bacterianas o fúngicas, o pueden ser el resultado de enfermedades autoinmunes. Más específicamente, las enfermedades infecciosas causadas por una infección vírica, bacteriana, fúngica u otra infección, se pueden tratar empleando una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' de la presente invención, que incluyen infecciones con VIH, virus de la hepatitis, herpes virus, micobacterias, Leshmania spp., plasmodio y diversas infecciones fúngicas, tal como candidiasis. Una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' de la presente invención también puede ser útil en este contexto cuando puede ser deseable en general un refuerzo para el sistema inmune, es decir, en el tratamiento del cáncer.

Alternativamente, las proteínas codificadas por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, se pueden utilizar en el tratamiento de enfermedades autoinmunes que incluyen, por ejemplo, la enfermedad del tejido conectivo, la esclerosis múltiple, el lupus eritematoso sistémico, la artritis reumatoide, la inflamación pulmonar autoinmune, el síndrome de Guillain-Barré, la tiroiditis autoinmune, la *diabetes mellitus* dependiente de insulina, la miastenia grave, la enfermedad de injerto contra hospedador y la enfermedad autoinmune inflamatoria del ojo. Una proteína tal, codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria,

también puede ser útil en el tratamiento de reacciones y estados alérgicos, tales como el asma (particularmente asma alérgica) u otros problemas respiratorios. Otros estados en los que se desea la supresión inmune (que incluyen, por ejemplo, el trasplante de órganos), se pueden tratar también con una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria.

Empleando las proteínas de la invención también es posible regular respuestas inmunes tanto favoreciéndolas como deteniéndolas.

La infrarregulación puede implicar inhibir o bloquear una respuesta inmune que ya está en marcha o puede implicar evitar la inducción de una respuesta inmune. Las funciones de los linfocitos T activados se pueden inhibir suprimiendo las respuestas de los linfocitos T o induciendo una tolerancia específica en los linfocitos T, o ambas. La inmunosupresión de las respuestas de los linfocitos T es generalmente un procedimiento activo no específico del antígeno, que requiere una exposición continua de los linfocitos T al agente supresor. La tolerancia que implica inducir la falta de sensibilidad o la anergía en los linfocitos T, se distingue de la inmunosupresión en que generalmente es específica del antígeno y persiste después de terminar la exposición al agente tolerante. Funcionalmente, la tolerancia se puede mostrar por la falta de una respuesta de los linfocitos T después de una nueva exposición al antígeno específico en ausencia del agente tolerante.

La infrarregulación o la prevención de una o varias funciones antigénicas (que incluyen sin limitación las funciones antigénicas de los linfocitos B, tales como por ejemplo, la coestimulación con B7), p. ej., evitar un alto nivel de síntesis de linfoquinas en los linfocitos T activados, será útil en situaciones de trasplante de tejidos, de piel y de órganos y en la enfermedad del injerto contra el hospedador (GVHD). Por ejemplo, el bloqueo de la función de los linfocitos T debe tener como consecuencia una menor destrucción del tejido en los trasplantes de tejido. Típicamente, en los trasplantes de tejido, el rechazo del trasplante se inicia por su reconocimiento como ajeno por los linfocitos T, seguido de una reacción inmune que destruye el trasplante. La administración de una molécula que inhiba o bloquee la interacción de un antígeno de los linfocitos B7 con su ligando natural sobre las células inmunes (tal como una forma soluble, monómera de un péptido que sólo tiene actividad B7-2 o junto con una forma monómera de un péptido que tiene actividad sobre otro antígeno de linfocitos B (p. ej., B7-1, B7-3) o un anticuerpo bloqueante) antes del trasplante, puede conducir a la unión de la molécula con el ligando natural sobre las células inmunes, transmitiendo la correspondiente señal coestimuladora. El bloqueo de esta forma de la función de los antígenos de los linfocitos B, evita la síntesis de citoquinas en las células inmunes, tales como los linfocitos T, y actúa de este modo como un inmunosupresor. Además, la falta de coestimulación puede ser suficiente para producir la anergía de los linfocitos T, induciendo de este modo la tolerancia en un individuo. La inducción de una tolerancia a largo plazo mediante reactivos que bloquean antígenos de los linfocitos B, puede evitar la necesidad de una administración repetida de estos reactivos bloqueantes. Para conseguir una inmunosupresión o una tolerancia suficiente en una persona, puede ser también necesario bloquear la función de una asociación de antígenos de linfocitos B.

La eficacia de los reactivos bloqueantes en particular para evitar el rechazo del trasplante de órganos o la GVHD, se puede determinar empleando modelos animales que sirven para pronosticar la eficacia en humanos. Ejemplos de sistemas adecuados en los que se puede utilizar, incluyen los injertos cardiaco alogénicos en ratas y los injertos xenogénicos en células de los islotes pancreáticos de ratones, habiéndose utilizado los dos para examinar los efectos inmunosupresores de las proteínas de fusión CTLA4Ig *in vivo*, tal y como describen Lenschow *y col., Science* **257**:789-792, 1992 y Turka *y col., Proc. Natl. Acad. Sci. USA* **89**:11102-11105, 1992. Además, los modelos en múridos de GVHD (véase, Paul compilador, *Fundamental Immunology*, Raven Press, Nueva York, 1989, págs. 846-847) se pueden utilizar para determinar el efecto de bloqueo *in vivo* de la función de los antígenos de los linfocitos B, sobre el desarrollo de esta enfermedad.

El bloqueo de la función de los antígenos también puede ser terapéuticamente útil para tratar enfermedades autoinmunes. Muchas enfermedades autoinmunes son el resultado de una activación inadecuada de los linfocitos T que se vuelven reactivos contra su propio tejido y que favorecen la producción de citoquinas y de autoanticuerpos implicados en la patología de las enfermedades. Evitar la activación de los linfocitos T autorreactivos puede reducir o eliminar los síntomas de la enfermedad. La administración de reactivos que bloquean la coestimulación de los linfocitos T, destruyendo las interacciones entre receptor/ligando de los antígenos de los linfocitos B, se puede utilizar para inhibir la activación de los linfocitos T y evitar la producción de autoanticuerpos o de citoquinas derivadas de los linfocitos T que estarían potencialmente implicadas en el proceso de la enfermedad. Adicionalmente, los reactivos bloqueantes pueden inducir la tolerancia específica del antígeno de linfocitos T autorreactivos que podrían conducir a un alivio a largo plazo de la enfermedad. La eficacia de los reactivos bloqueantes para evitar o aliviar las enfermedades autoinmunes, se puede determinar empleando un número de modelos animales bien caracterizados, de enfermedades autoinmunes humanas. Ejemplos incluyen la encefalitis autoinmune experimental de múridos, el lupus eritematoso sistémico en ratones MRL/pr/pr o ratones híbridos NZB, la artritis autoinmune por colágeno en múridos, la *diabetes mellitus* en ratones OD y en ratas BB y la miastenia grave experimental en múridos (*véase más arriba*, Paul compilador, págs. 840-856).

La suprarregulación de una función antigénica (preferentemente una función antigénica de los linfocitos B), como un medio para suprarregular las respuestas inmunes, también puede ser útil en la terapia. La suprarregulación de las respuestas inmunes puede implicar una mejora de la respuesta inmune existente o producir una respuesta inmune inicial, tal y como se muestra en los siguientes ejemplos. Por ejemplo, la mejora de una respuesta inmune mediante la estimulación de la función antigénica de los linfocitos B puede ser útil en casos de infección vírica. Además, las

enfermedades víricas sistémicas, tales como la gripe, el resfriado común y la encefalitis, podrían aliviarse con la administración sistémica de la forma estimuladora de los antígenos de los linfocitos B.

Alternativamente, las respuestas inmunes antivíricas se pueden mejorar en un paciente infectado, eliminando los linfocitos T del paciente, coestimulando los linfocitos T *in vitro* con APCs sometidas a impulsos de antígeno vírico, que expresan un péptido codificado por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria o junto con una forma estimuladora de un péptido soluble codificado por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria e introducir de nuevo los linfocitos T sensibilizados *in vitro* en el paciente. Las células infectadas serán capaces entonces de suministrar una señal coestimuladora a los linfocitos T *in vivo*, activando de este modo los linfocitos T.

En otra aplicación, la suprarregulación o la mejora de la función antigénica (preferentemente la función antigénica de los linfocitos B) puede ser útil para la inducción de inmunidad tumoral. Las células tumorales (p. ej., sarcoma, melanoma, linfoma, leucemia, neuroblastoma, carcinoma) transfectadas con un ácido nucleico que codifica al menos un péptido codificado por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, se pueden administrar a un individuo para superar la tolerancia específica de tumores en el individuo. Si se desea, la célula tumoral se puede transfectar para expresar una combinación de péptidos. Por ejemplo, las células tumorales obtenidas de un paciente se pueden transfectar *ex vivo* con un vector de expresión que dirige la expresión de un péptido que sólo tiene actividad similar a B7-2 o junto con un péptido que tiene actividad similar a B7-1 y/o actividad similar a B7-3. Las células tumorales transfectadas se devuelven al paciente para dar como resultado la expresión de los péptidos sobre la superficie de la célula transfectada. Alternativamente, las técnicas de terapia génica se pueden utilizar para dirigir a la diana, una célula tumoral para la transfección *in vivo*.

La presencia del péptido codificado por los ADNc extendidos derivados de las ESTs 5' descritas en esta memoria que tienen la actividad de un(os) antígeno(s) de linfocitos B en la superficie de la célula tumoral, proporciona la señal necesaria de coestimulación a los linfocitos T para inducir una respuesta inmune mediada por linfocitos T, contra las células tumorales transfectadas. Además, las células tumorales que carecen o que fracasan en la re-expresión de cantidades suficientes de moléculas del CPH de clase I o del CPH de clase II, se pueden transfectar con ácidos nucleicos que codifican todo o una parte de (p. ej., una parte truncada de un dominio citoplásmico) de una cadena α y una microglobulina β_2 de la clase I del CPH o una cadena α de la clase II del CPH y una cadena β de la clase II del CPH, para expresar de este modo proteínas de la clase I del CPH o proteínas de la clase II del CPH sobre la superficie celular, respectivamente. La expresión de moléculas adecuadas de la clase I o de la clase II del CPH junto con un péptido que tiene la actividad de un antígeno de los linfocitos B (p. ej., B7-1, B7-2, B7-3) induce una respuesta inmune mediada por linfocitos T contra la célula tumoral transfectada. Opcionalmente, un gen que codifica una estructura artificial no codificante que bloquea la expresión de una proteína asociada a la clase II del CPH, tal como la cadena invariable, también se puede cotransfectar con un ADN que codifica un péptido que tiene la actividad de un antígeno de los linfocitos B, para favorecer la presentación de los antígenos asociados a tumores e inducir la inmunidad específica tumoral. Por tanto, la inducción de una respuesta inmune mediada por linfocitos T en un ser humano, puede ser suficiente para vencer la tolerancia específica tumoral en el individuo. Alternativamente, tal y como se describe con más detalle a continuación, los genes que codifican estas proteínas reguladoras del sistema inmune o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o disminuir la expresión de las proteínas según se desee.

Ejemplo 34

Ensayo de las proteínas expresadas procedentes de los ADNc extendidos o de partes de los mismos para estudiar la actividad reguladora de la hematopoyesis

Las proteínas codificadas por los ADNc extendidos o por partes de los mismos, también se pueden evaluar en relación con su actividad reguladora de la hematopoyesis. Por ejemplo, se puede evaluar el efecto de las proteínas sobre la diferenciación de células pluripotenciales embrionarias no humanas. Numerosos ensayos de esta actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Johansson *y col., Cell. Biol.* **15**:141-151, 1995; Keller *y col., Mol. Cell. Biol.* **13**:473-486, 1993; McClanahan *y col., Blood* **81**:2903-2915, 1993.

Las proteínas codificadas por los ADNc extendidos o por partes de los mismos, también se pueden evaluar en relación con su influencia sobre el tiempo de vida de las células pluripotenciales y la diferenciación de células pluripotenciales. Numerosos ensayos de esta actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Freshney, Methylcellulose Colony Forming Assays, en *Culture of Hematopoietic Cells*, Freshney *y col.,* compiladores, págs. 265-268, Wiley-Liss Inc., Nueva York, NY, 1994; Hirayama *y col., Proc. Natl. Acad. Sci. USA*. **89**:5907-5911, 1992; McNiece y Briddell, en *Culture of Hematopoietic Cells, véase más arriba*; Neben *y col., Exp. Hematol.* **22**:353-359, 1994; Ploemacher y Cobblestone en *Culture of Hematopoietic Cells, véase más arriba*, Spooncer *y col., Culture of Hematopoietic Cells, véase más arriba*, 163-179 y Sutherland *Culture of Hematopoietic Cells, véase más arriba*, 139-162.

Las proteínas que muestran actividad reguladora de la hematopoyesis se pueden formular a continuación como agentes farmacéuticos y emplear para tratar estados clínicos en los que sea beneficiosa la regulación de la hematopoyesis, tales como el tratamiento de deficiencias en células mieloides o linfoides. La implicación en la regulación de la

hematopoyesis está indicada incluso con actividad biológica marginal para favorecer a células formadoras de colonias o a líneas celulares dependientes de un factor. Por ejemplo, las proteínas que sostienen el crecimiento y la proliferación de células progenitoras eritroides aisladas o junto con otras citoquinas, tienen utilidad, por ejemplo, para tratar diversas anemias o para emplear junto con la radiación/quimioterapia para estimular la producción de precursores eritroides y/o células eritroides. Las proteínas que sostienen el crecimiento y la proliferación de células mieloides, tales como los granulocitos y los monocitos/macrófagos (es decir, actividad CSF tradicional) pueden ser útiles, por ejemplo, junto con la quimioterapia para evitar o tratar una mielosupresión posterior. Las proteínas que mantienen el crecimiento y la proliferación de megacariocitos y, como consecuencia, de las plaquetas, permiten la prevención o el tratamiento de diversas enfermedades plaquetarias tales como la trombocitopenia y generalmente se pueden emplear en lugar de transfusiones de plaquetas o de forma complementaria a la transfusión. Las proteínas que mantienen el crecimiento y la proliferación de células pluripotenciales hematopoyéticas que son capaces de hacer madurar a cualquiera de todas las células hematopoyéticas mencionadas anteriormente, pueden tener un uso terapéutico en diversas enfermedades de las células pluripotenciales (tales como los tratados generalmente con trasplantes, que incluyen sin limitación, la anemia aplásica y la hemoglobinuria nocturna paroxismal), así como para repoblar el compartimiento de las células pluripotenciales después de una radiación/quimioterapia, tanto *in vivo* como *ex vivo* (es decir, junto con un trasplante de médula ósea o con un trasplante (homólogo o heterólogo) de células progenitoras periféricas) como células normales o manipuladas genéticamente para la terapia génica. Alternativamente, tal y como se describe con más detalle a continuación, los genes que codifican proteínas con actividad reguladora de la hematopoyesis o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o disminuir la expresión de las proteínas, según se desee.

Ejemplo 35

Ensayo de las proteínas expresadas a partir de ADNc extendidos o de partes de los mismos para la regulación del crecimiento tisular

Las proteínas codificadas por los ADNc extendidos o por partes de los mismos también se pueden evaluar según su efecto sobre el crecimiento tisular. Numerosos ensayos de dicha actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en los documentos de Publicación de las Patentes Internacionales nº WO 95/16035, WO 95/05846 y WO 91/07491.

Los ensayos para la actividad de curación de heridas incluyen sin limitación, los descritos en: Winter, *Epidermal Wound Healing*, págs. 71-112, Maibach y Rovee compiladores, Year Book Medical Publishers, Inc., Chicago, modificados por Eaglstein y Mertz, *J. Invest. Dermatol.*, **71**:382-384, 1978.

Las proteínas que están implicadas en la regulación del crecimiento tisular se pueden formular a continuación como agentes farmacéuticos y emplear para tratar estados clínicos en los que la regulación del crecimiento tisular sea beneficiosa. Por ejemplo, una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, también pueden tener una utilidad en composiciones empleadas para el crecimiento o la regeneración de tejido óseo, cartílago, tendón, ligamento y/o nervioso, así como para la curación de heridas y la reparación y sustitución tisular, y en el tratamiento de quemaduras, incisiones y úlceras.

Una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria que induce el crecimiento de cartílago y/o hueso en circunstancias en las que el hueso normalmente no se formaría, tiene aplicación en la curación de fracturas óseas y lesiones o defectos del cartílago en seres humanos y en otros animales. Una preparación tal que emplea una proteína de la invención puede tener un uso profiláctico en la reducción de fracturas abiertas o cerradas y también en la fijación mejorada de articulaciones artificiales. La síntesis ósea *de novo* inducida por un agente osteogénico contribuye a la reparación de defectos craneofaciales congénitos, inducidos por un trauma o inducidos por resección oncológica y también es útil en la cirugía plástica.

Una proteína de esta invención también se puede utilizar en el tratamiento de la enfermedad periodontal y en otros procesos de reparación dental. Tales 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 de la invención también puede ser útil en el tratamiento de la osteoporosis o la osteoartritis, mediante la estimulación de la reparación del hueso y/o del cartílago o bloqueando la inflamación o los procesos de destrucción del tejido (actividad colagenasa, actividad de osteoclastos, etc.) mediados por procesos inflamatorios.

Otra categoría de la actividad regeneradora de tejidos que puede ser atribuible a la proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria es la formación de tendón/ligamento. Una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria que induce la formación de tejido similar al tendón/ligamento u otro tejido, en circunstancias en las que dicho tejido no se formaría normalmente, tiene aplicación en la curación de roturas de tendones o ligamentos, deformidades y otros defectos de tendones o ligamentos en humanos y en otros animales. Una preparación tal, que emplea una proteínas que induce el tejido similar al tendón/ligamento puede tener un uso profiláctico para evitar las lesiones del tejido del tendón o del ligamento, así como un uso en la fijación mejorada del tendón o del ligamento al hueso o a otros tejidos y para reparar defectos del tejido del tendón o del ligamento. La formación *de novo* de tejido similar al tendón/ligamento inducida por una composición codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta

memoria, contribuye a la reparación de defectos del tendón o del ligamento de origen congénito, traumático u otro origen y también es útil en la cirugía plástica para fijar o reparar tendones o ligamentos. Las composiciones codificadas por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria pueden proporcionar un entorno para atraer las células formadoras del tendón o del ligamento, estimular el crecimiento de las células formadoras de tendón o de ligamento, inducir la diferenciación de progenitores de las células formadoras de tendón o de ligamento o inducir el crecimiento de células del tendón/ligamento o progenitoras *ex vivo* para restituirlas *in vivo* para efectuar la reparación del tejido. Las composiciones descritas en esta memoria también pueden ser útiles en el tratamiento de la tendinitis, el síndrome de túnel carpiano y otros defectos de los tendones o de los ligamentos. Las composiciones también pueden incluir una matriz adecuada y/o un agente secuestrante como vehículo, tal y como bien se conoce en la técnica.

La proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, también puede ser útil para la proliferación de células neuronales y para la regeneración de nervios y de tejido cerebral, es decir, para el tratamiento de enfermedades del sistema nervioso central y periférico y neuropatías, así como trastornos mecánicos y traumáticos que implican la degeneración, la muerte o el trauma de las células neuronales o del tejido nervioso. Más específicamente, una proteína se puede utilizar en el tratamiento de enfermedades del sistema nervioso periférico, tal como lesiones de los nervios periféricos, neuropatía periférica y neuropatías localizadas y enfermedades del sistema nervioso central, tales como la enfermedad de Alzheimer, de Parkinson, de Huntington, la esclerosis lateral amiotrófica y el síndrome de Shy-Drager. Otros estados que se pueden tratar de acuerdo con la presente invención, incluyen las enfermedades mecánicas y traumáticas, tales como enfermedades de la médula espinal, trauma cerebral y enfermedades cerebrovasculares, tales como el accidente cerebro vascular. Las neuropatías periféricas que son el resultado de la quimioterapia u otras terapias médicas, también se pueden tratar empleando una proteína de la invención.

Las proteínas de la invención también pueden ser útiles para favorecer más o a mayor velocidad el cierre de heridas no restañadas, que incluyen sin limitación, las úlceras por presión, las úlceras asociadas a insuficiencia vascular, las heridas quirúrgicas y traumáticas y similares.

Se espera que una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, pueda mostrar también actividad para la generación o la regeneración de otros tejidos, tales como órganos (incluyendo, por ejemplo, el páncreas, el hígado, el intestino, el riñón, la piel, el endotelio), tejido muscular (liso, esquelético o cardiaco) y tejido vascular (incluyendo el endotelio vascular) o para favorecer el crecimiento de células que comprenden tales tejidos. Parte de los efectos deseados, puede ser la inhibición o la modulación de la cicatrización fibrótica para permitir que se genere tejido normal. Una proteína de la invención también puede mostrar actividad angiogénica.

Una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, también puede ser útil para una buena protección o regeneración y un tratamiento de la fibrosis pulmonar o del hígado, de la lesión por reperfusión en diversos tejidos y estados resultantes de lesiones sistémicas con citoquinas.

Una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria también puede ser útil para favorecer o inhibir la diferenciación de tejidos, descrita anteriormente, a partir de tejidos o células precursoras; o para inhibir el crecimiento de tejidos descritos anteriormente.

Alternativamente, tal y como se describe a continuación con más detalle, los genes que codifican proteínas con actividad reguladora del crecimiento de tejidos o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o disminuir la expresión de las proteínas según se desee.

Ejemplo 36

Ensayo en las proteínas expresadas a partir de ADNc extendidos o de partes de los mismos para estudiar la regulación de hormonas reproductoras

Las proteínas codificadas por los ADNc extendidos o por partes de los mismos también se pueden evaluar por su capacidad para regular hormonas reproductoras, tales como la hormona estimuladora del folículo. Numerosos ensayos para dicha actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Vale *y col., Endocrinol.* **91**:562-572, 1972; Ling *y col., Nature* **321**:779-782, 1986; Vale *y col., Nature* **321**:776-779, 1986; Mason *y col., Nature* **318**:659-663, 1985; Forage *y col., Proc. Natl. Acad. Sci. USA*. **83**:3091-3095, 1986; Capítulo 6.12 en *Currents Protocols in Immunology*, Coligan *y col.* compiladores, Greene Publishing Associates and Wiley-Interscience; Taub *y col., J. Clin. Invest.* **95**:1370-1376, 1995; Lind *y col., APMIS* **103**:140-146, 1995; Muller *y col., Eur. J. Immunol.* **25**:1744-1748; Gruber *y col., J. Immunol*. **152**:5860-5867, 1994; Johnston *y col., J. Immunol.* **153**:1762-1768, 1994.

Las proteínas que muestran actividad como hormonas reproductoras o reguladoras del movimiento celular se pueden formular a continuación como agentes farmacéuticos y emplear para tratar estados clínicos en donde una regulación de las hormonas reproductoras es beneficiosa. Por ejemplo, una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria también puede mostrar actividades relacionadas con la

activina o la inhibina. Las inhibinas se caracterizan por su capacidad para inhibir la liberación de la hormona estimulante del folículo (FSH), mientras que las activinas se caracterizan por su capacidad para estimular la liberación de la FSH. Por tanto, una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, aislada o en heterodímeros con un miembro de la familia de las inhibinas α, puede ser útil como un anticonceptivo basándose en la capacidad de las inhibinas para disminuir la fertilidad en mamíferos hembras y disminuir la espermatogénesis en mamíferos machos. La administración de cantidades suficientes de otras inhibinas puede inducir la infertilidad en estos mamíferos. Alternativamente, la proteína de la invención, como un homodímero o un heterodímero con otras subunidades proteicas del grupo de las inhibinas B, puede ser útil como un agente terapéutico inductor de la fertilidad, basándose en la capacidad de las moléculas de activina para estimular la liberación de FSH a partir de las células de la pituitaria anterior. Véase, por ejemplo, el documento de patente de EE.UU. nº 4.798.885. Una proteína de la invención puede ser también útil para adelantar el inicio de la fertilidad en mamíferos sexualmente inmaduros, de modo que se incrementa la duración de la actividad reproductora de animales domésticos tales como vacas, ovejas y cerdos.

Alternativamente, tal y como se describe con más detalle a continuación, los genes que codifican proteínas con actividad reguladora de las hormonas reproductoras o ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en las células hospedadoras adecuadas para incrementar o disminuir la expresión de las proteínas según se desee.

Ejemplo 37

Ensayo en las proteínas expresadas en los ADNc extendidos o en partes de los mismos para estudiar la actividad quimiotáctica/quimiocinética

Las proteínas codificadas por los ADNc extendidos o por partes de los mismos también se pueden evaluar por su actividad quimiotáctica/quimiocinética. Por ejemplo, una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' descritas en esta memoria, puede tener actividad quimiotáctica o quimiocinética (p. ej., actúa como una quimioquina) sobre células de mamífero, incluyendo, por ejemplo, los monocitos, los fibroblastos, los neutrófilos, los linfocitos T, los mastocitos, los eosinófilos, las células epiteliales y/o endoteliales. Las proteínas quimiotácticas y quimiocinéticas se pueden utilizar para movilizar o atraer una población de células deseadas hacia un sitio de acción deseado. Las proteínas quimiotácticas o quimiocinéticas proporcionan unas ventajas particulares en el tratamiento de heridas y otros traumas en tejidos, así como en el tratamiento de infecciones localizadas. Por ejemplo, la atracción de linfocitos, monocitos o neutrófilos hacia tumores o sitios de infección, puede dar como resultado unas respuestas inmunes mejoradas contra el tumor o el agente infectante.

Una proteína o un péptido tiene actividad quimiotáctica sobre una población celular en particular si puede estimular, directa o indirectamente, la orientación o el movimiento dirigido de dicha población celular. Preferentemente, la proteína o el péptido tiene la capacidad de estimular directamente el movimiento dirigido de las células. Si una proteína en particular tiene actividad quimiotáctica sobre una población de células, se puede determinar fácilmente empleando dicha proteína o péptido en cualquier ensayo conocido de quimiotaxis celular.

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

Los ensayos de la actividad quimiotáctica (que identificarán proteínas que inducen o evitan la quimiotaxis) consisten en ensayos que miden la capacidad de una proteína para inducir la migración de las células a través de una membrana, así como la capacidad de una proteína para inducir la adhesión de una población celular a otra población celular. Los ensayos adecuados del movimiento y la adhesión incluyen, sin limitaciones, los descritos en: *Currents Protocols in Immunology,* compiladores Coligan, Kruisbeek, Margulies, Shevach y Strober, Greene Publishing Associates and Wiley-Interscience, Capítulo **6.12**: 6.12.1-6.12.28; Taub *y col., J. Clin. Invest*. **95**:1370-1376, 1995; Lind *y col., APMIS* **103**:140-146, 1995; Müller *y col., Eur. J. Immunol.* **25**:1744-1748; Gruber *y col., J. Immunol*. **152**:5860-5867, 1994; Johnston *y col., J. Immunol*. **153**:1762-1768, 1994.

Ejemplo 38

Ensayos en las proteínas expresadas a partir de ADNc extendidos o de partes de los mismos de la regulación de la coagulación sanguínea

Las proteínas codificadas por los ADNc extendidos o partes de los mismos también se pueden evaluar por sus efectos sobre la coagulación sanguínea. Numerosos ensayos de dicha actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Linet *y col., J. Clin. Pharmacol.* **26**:131-140, 1986; Burdick *y col., Trombosis Res*. **45**:413-419, 1987; Humphrey *y col., Fibrinolysis* **5**:71-79, 1991; Schaub *y col., Prostaglandins* **35**:467-474, 1988.

Las proteínas que están implicadas en la regulación de la coagulación sanguínea se pueden formular a continuación como agentes farmacéuticos y se pueden emplear para tratar estados clínicos en los que sea beneficiosa la regulación de la coagulación sanguínea. Por ejemplo, una proteína de la invención puede mostrar también actividad hemostática o trombolítica. Como resultado, una proteína tal se espera que sea útil en el tratamiento de diversos trastornos de la coagulación (que incluyen enfermedades hereditarias, tales como la hemofilia) o para mejorar la coagulación y

otros sucesos hemostáticos en el tratamiento de heridas resultado de un trauma, cirugía u otras causas. Una proteína de la invención también puede ser útil para disolver trombos o evitar su formación y para tratar o evitar los estados consecuencia de lo anterior (tales como infarto de los vasos cardiacos y de los vasos del sistema nervioso central (p. ej., accidente cerebro vascular)). Alternativamente, tal y como se describe a continuación con más detalle, los genes que codifican proteínas con actividad sobre la coagulación sanguínea o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o reducir la expresión de las proteínas según se desee.

Ejemplo 39

Ensayos en las proteínas expresadas a partir de ADNc extendidos o de partes de los mismos para estudiar la implicación en las interacciones entre receptor/ligando

Las proteínas codificadas por los ADNc extendidos o por partes de los mismos, también se pueden evaluar por su implicación en las interacciones entre receptor/ligando. Numerosos ensayos de dicha actividad son conocidos por los expertos en la técnica, incluyendo los ensayos descritos en las siguientes referencias: Capítulo 7. 7.28.1-7.28.22 en *Currents Protocols in Immunology,* Coligan *y col.* compiladores, Greene Publishing Associates and Wiley-Interscience; Takai *y col., Proc. Natl. Acad. Sci. USA* **84**:6864-6868, 1987; Bierer *y col., J. Exp. Med*. **168**:1145-1156, 1988; Rosenstein *y col., J. Exp. Med.* **169**:149-160, 1989; Stoltenborg *y col., J. Immunol. Methods* **175**:59-68, 1994; Stitt *y col.,* Cell **80**:661-670, 1995; Gyuris *y col., Cell* **75**:791-803, 1993.

Por ejemplo, las proteínas codificadas por los ADNc extendidos obtenidos a partir de las ESTs 5' de la presente invención también pueden mostrar actividad como receptores, ligandos de receptores o inhibidores o agonistas de las interacciones receptor/ligando. Ejemplos de tales receptores y ligandos incluyen, sin limitación, los receptores de citoquinas y sus ligandos, las quinasas de receptores y sus ligandos, las fosfatasas de receptores y sus ligandos, los receptores implicados en las interacciones célula-célula y sus ligandos (que incluyen sin limitación, las moléculas de adhesión celular (tales como las selectinas, las integrinas y sus ligandos) y parejas de receptor/ligando implicadas en la presentación de antígenos, reconocimiento de antígenos y desarrollo de respuestas inmunes celulares y humorales). Los receptores y los ligandos también son útiles para escrutar inhibidores potenciales peptídicos o de moléculas pequeñas de la interacción relevante entre receptor/ligando. Una proteína codificada por los ADNc extendidos obtenidos a partir de las ESTs 5' de la presente invención (que incluye sin limitación, fragmentos de receptores y ligandos) puede ser ella misma útil como inhibidor de las interacciones entre receptor/ligando. Alternativamente, tal y como se describe a continuación con más detalle, los genes que codifican proteínas implicadas en las interacciones entre receptor/ligando o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o reducir la expresión de las proteínas, según se desee.

Ejemplo 40

Ensayos en las proteínas expresadas a partir de ADNc extendidos o de partes de los mismos para estudiar la actividad anti-inflamatoria

Las proteínas codificadas por los ADNc extendidos o partes de los mismos también se pueden evaluar por la actividad anti-inflamatoria. La actividad anti-inflamatoria se puede conseguir proporcionando un estímulo a células implicadas en la respuesta inflamatoria, inhibiendo o favoreciendo las interacciones entre célula-célula (tales como por ejemplo, la adhesión celular), inhibiendo o favoreciendo la quimiotaxis de las células implicadas en el proceso inflamatorio, inhibiendo o favoreciendo la extravasación celular o estimulando o inhibiendo la producción de otros factores que inhiben o favorecen más directamente la respuesta inflamatoria. Las proteínas que muestran tales actividades se pueden utilizar para tratar estados inflamatorios que incluyen estados crónicos o agudos, incluyendo sin limitación la inflamación asociada con la infección (tal como el choque séptico, la sepsis o el síndrome de respuesta inflamatoria sistémica), la lesión por isquemia-reperfusión, la letalidad por endotoxina, la artritis, el rechazo hiperagudo mediado por el complemento, la nefritis, la lesión pulmonar inducida por citoquinas o quimioquinas, la enfermedad inflamatoria de Bowel, la enfermedad de Crohn, o resultantes de una hiperproducción de citoquinas tales como TNF o IL-1. Las proteínas de la invención también se pueden utilizar para tratar la anafilaxia y la hipersensibilidad a una sustancia o material antigénico. Alternativamente, tal y como se describe a continuación con más detalle, los genes que codifican proteínas con actividad anti-inflamatoria o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o reducir la expresión de las proteínas según se desee.

Ejemplo 41

Ensayos en las proteínas expresadas a partir de ADNc extendidos o de partes de los mismos para estudiar la actividad inhibidora de tumores

Las proteínas codificadas por los ADNc extendidos o por partes de los mismos también se pueden evaluar por la actividad inhibidora tumoral. Además de las actividades descritas anteriormente para el tratamiento inmunológico o la prevención de tumores, una proteína de la invención puede mostrar otras actividades anti-tumorales. Una proteína puede inhibir el crecimiento tumoral directa o indirectamente (como por ejemplo, mediante ADACC). Una proteína puede mostrar su actividad inhibidora tumoral actuando sobre tejido tumoral o sobre tejido precursor tumoral, inhi-

biendo la formación de los tejidos necesarios para sostener el crecimiento tumoral (tal como por ejemplo, inhibiendo la angiogénesis), causando la producción de otros factores, agentes o tipos celulares que inhiben el crecimiento tumoral, o reprimiendo, eliminando o inhibiendo factores, agentes o tipos celulares que favorecen el crecimiento tumoral. Alternativamente, tal y como se describe a continuación con más detalle, los genes que codifican proteínas con actividad inhibidora tumoral o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o reducir la expresión de las proteínas según se desee.

Una proteína de la invención también puede mostrar una o varias de las siguientes actividades o efectos adicionales: inhibir el crecimiento, la infección o el funcionamiento de agentes infecciosos o matándolos, que incluyen sin limitación: bacterias, virus, hongos u otros parásitos; consiguiendo (reprimiendo o mejorando) características corporales, que incluyen sin limitación la altura, el peso, el color del cabello, el color de los ojos, la piel, el porcentaje de grasa, u otra pigmentación del tejido, o el tamaño o la forma de un órgano o una parte del cuerpo (como por ejemplo, aumento o disminución del pecho, un cambio en la forma o el tamaño de un hueso); consiguiendo biorritmos, o ciclos o ritmos circadianos; consiguiendo la fertilidad de individuos machos o hembras; consiguiendo el metabolismo, el catabolismo, el anabolismo, el procesamiento, el uso, el almacenamiento o la eliminación de grasas, lípidos, proteínas, carbohidratos, vitaminas, minerales, cofactores u otros factores nutricionales o componentes de la dieta; consiguiendo característica de la conducta que incluyen sin limitación, el apetito, la líbido, el estrés, la percepción (incluyendo trastornos del conocimiento), la depresión (incluyendo trastornos depresivos) y comportamientos violentos; proporcionando efectos analgésicos u otros efectos que reducen el dolor; favoreciendo la diferenciación y el crecimiento de células pluripotenciales embrionarias en progenies distintas a las progenies hematopoyéticas; actividad hormonal o endocrina; en el caso de enzimas, corrección de deficiencias de la enzima y tratamiento de las enfermedades relacionadas con las deficiencias; tratamiento de trastornos hiperproliferativos (tales como por ejemplo, la psoriasis); actividad similar a la inmunoglobulina (como por ejemplo, la capacidad de unirse a antígenos o al complemento); y la capacidad de actuar como un antígeno en una composición de vacuna para conseguir una respuesta inmune contra dicha proteína u otro material o entidad que tiene una reacción cruzada con dicha proteína. Alternativamente, tal y como se describe a continuación con más detalle, los genes que codifican proteínas implicadas en cualquiera de las actividades mencionadas anteriormente o los ácidos nucleicos que regulan la expresión de tales proteínas, se pueden introducir en células hospedadoras adecuadas para incrementar o reducir la expresión de las proteínas según se desee.

Ejemplo 42

Identificación de proteínas que interaccionan con polipéptidos codificados por los ADNc extendidos

Las proteínas que interaccionan con los polipéptidos codificados por los ADNc obtenidos a partir de las ESTs 5' o de fragmentos de las mismas, tales como proteínas receptoras, se pueden identificar empleando dos sistemas híbridos, tales como el sistema "Matchmaker Two Hybrid System 2" (nº de catálogo K1604-1, Clontech). Tal y como se describe en el manual que acompaña el equipo de reactivos, los ADNc obtenidos a partir de las ESTs 5' o de fragmentos de las mismas, se insertan en un vector de expresión tal como los que están en marco de lectura con el ADN que codifica el dominio de unión del ADN del activador transcripcional de la levadura, GAL4. Los ADNc en una genoteca de ADNc que codifican proteínas que podrían interaccionar con los polipéptidos codificados por los ADNc extendidos o por partes de los mismos, se insertan en un segundo vector de expresión de modo que estén en marco de lectura con el ADN que codifica el dominio de activación de GAL4. Los dos plásmidos de expresión se transforman en levadura y la levadura se extiende en placas sobre medio de selección que selecciona la expresión de los marcadores seleccionables en cada uno de los vectores de expresión, así como la expresión dependiente de GAL4 del gen HIS3. Los transformantes capaces de crecer en un medio que carece de histidina, se escrutan en busca de expresión de lacZ dependiente de GAL4. Las células que son positivas para la selección con histidina y para el ensayo de lacZ, contienen plásmidos que codifican proteínas que interaccionan con el polipéptido codificado por los ADNc extendidos o por partes de los mismos.

Alternativamente, el sistema descrito por Lustig *y col., Methods in Enzymology* **283**:83-99, 1997 y en el documento de patente de EE.UU. nº 5.654.150, se puede utilizar para identificar moléculas que interaccionan con los polipéptidos codificados por los ADNc extendidos. En tales sistemas, las reacciones de transcripción *in vitro* se realizan sobre un conjunto de vectores que contienen insertos de ADNc extendido, clonados aguas arriba de un promotor que dirige la transcripción *in vitro*. Los grupos resultantes de ARNm se introducen en oocitos de *Xenopus laevis*. En los oocitos se somete a ensayo a continuación una actividad deseada.

Alternativamente, los productos reunidos de la transcripción *in vitro* producidos tal y como se ha descrito anteriormente, se traducen *in vitro*. En los productos reunidos de la traducción *in vitro* se puede someter a ensayo una actividad deseada o la interacción con un polipéptido conocido.

Las proteínas u otras moléculas que interaccionan con polipéptidos codificados por ADNc extendidos se pueden encontrar con una variedad de técnicas. En un método, se pueden construir columnas de afinidad que contienen el polipéptido codificado por el ADNc extendido o una parte del mismo. En algunas versiones de este método, la columna de afinidad contiene proteínas quiméricas en las que la proteína codificada por el ADNc extendido o por partes del mismo se fusiona con la S-transferasa de glutatión. Una mezcla de proteínas celulares o el grupo de proteínas expresadas tal y como se ha descrito anteriormente, se aplica a la columna de afinidad. Las proteínas que interaccionan con el polipéptido fijado a la columna, se pueden aislar a continuación y analizar en un gel de electroforesis 2-D, tal y como describen Ramunsen *y col., Electrophoresis* **18**:588-598, 1997. Alternativamente, las proteínas retenidas en la

columna de afinidad se pueden purificar por métodos basados en la electroforesis y secuenciar. El mismo método se puede utilizar para aislar anticuerpos, para escrutar productos que presentan fagos o para escrutar anticuerpos humanos que presentan fagos.

Las proteínas que interaccionan con polipéptidos codificados por ADNc extendidos o por partes de los mismos también se pueden escrutar empleando un biosensor óptico, tal y como se describe en Edwards y Leatherbarrow, *Analytical Biochemistry* **246**:1-6, 1997. La ventaja principal del método es que permite la determinación de la tasa de asociación entre la proteína y otras moléculas interaccionantes. Por tanto, es posible seleccionar específicamente moléculas con una tasa de asociación alta o baja. Típicamente, una molécula diana se une a la superficie de un sensor (a través de una matriz de carboximetil dextrano) y una muestra de las moléculas del ensayo se pone en contacto con las moléculas diana. La unión de una molécula del ensayo con la molécula diana provoca un cambio en el índice de refracción y/o en el espesor. Este cambio se detecta con el biosensor con la condición de que esto ocurra en el campo evanescente (que se prolonga unos pocos cientos de nanómetros desde la superficie del sensor). En estos ensayos de escrutinio, la molécula diana puede ser uno de los polipéptidos codificados por los ADNc extendidos o por una parte de los mismos y la muestra del ensayo puede ser una colección de proteínas extraídas de tejidos o de células, un grupo de proteínas expresadas, bibliotecas combinatorias peptídicas y/o químicas, o péptidos que presentan fagos. Los tejidos o las células a partir de las cuales se extraen las proteínas del ensayo pueden proceder de cualquier especie.

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

Para estudiar la interacción de las proteínas codificadas por los ADNc extendidos o por partes de los mismos, con fármacos, se puede utilizar el método de la microdialisis acoplada a la HPLC, descrito por Wang *y col., Chromatographia* **44**:205-208, 1997 o el método de electroforesis capilar por afinidad, descrito por Busch *y col., J. Chromatogr.* **777**:311-328, 1997.

Los expertos en la técnica apreciarán que las proteínas expresadas a partir de los ADNc extendidos o de partes de los mismos se pueden someter a ensayo para estudiar numerosas actividades, además de las mencionadas específicamente más arriba. Por ejemplo, las proteínas expresadas se pueden evaluar para aplicaciones que implican el control y la regulación de la inflamación, la proliferación tumoral o la metástasis, la infección u otros estados clínicos. Además, las proteínas expresadas a partir de los ADNc extendidos o de partes de los mismos, pueden ser útiles como agentes nutricionales o agentes cosméticos.

Las proteínas expresadas a partir de los ADNc extendidos o de partes de los mismos se pueden utilizar para generar anticuerpos capaces de unirse específicamente a la proteína expresada o a fragmentos de la misma, tal y como se ha descrito en el Ejemplo 40. Los anticuerpos pueden ser capaces de unirse a una proteína de longitud completa codificada por un ADNc obtenido a partir de una EST 5', una proteína madura (es decir, la proteína generada por escisión del péptido señal) codificada por un ADNc obtenido a partir de una EST 5' o un péptido señal codificado por un ADNc obtenido a partir de una EST 5'. Alternativamente, los anticuerpos pueden ser capaces de unirse a fragmentos de al menos 10 aminoácidos de las proteínas codificadas por los ADNc anteriores. En algunas realizaciones, los anticuerpos pueden ser capaces de unirse a fragmentos de al menos 15 aminoácidos de las proteínas codificadas por los ADNc anteriores. En otras realizaciones, los anticuerpos pueden ser capaces de unirse a fragmentos de al menos 25 aminoácidos de las proteínas expresadas a partir de los ADNc extendidos que comprenden al menos 25 aminoácidos de las proteínas codificadas por los ADNc anteriores. En otras realizaciones, los anticuerpos pueden ser capaces de unirse a fragmentos de al menos 40 aminoácidos de las proteínas codificadas por los ADNc anteriores.

Ejemplo 43

Producción de un anticuerpo para una proteína humana

Una proteína o un polipéptido sustancialmente puro se aísla de las células transfectadas o transformadas, tal y como se ha descrito en el Ejemplo 30. La concentración de proteína en la preparación final se ajusta, por ejemplo, mediante concentración con un dispositivo de filtro de Amicon, hasta el nivel de unos pocos μg/ml. El anticuerpo monoclonal o policlonal para la proteína se puede preparar a continuación del modo siguiente:

1. *Producción del anticuerpo monoclonal mediante fusión de hibridomas*

El anticuerpo monoclonal para epítopos de cualquiera de los péptidos identificados y aislados, tal y como se ha descrito, se puede preparar a partir de hibridomas de múridos según el método clásico de Kohler y Milstein, *Nature* **256**:495, 1975 o por métodos derivados del mismo. Resumiendo, un ratón es inoculado repetidamente con unos pocos microgramos de la proteína seleccionada o los péptidos derivados de la misma, durante un periodo de unas pocas semanas. El ratón se sacrifica a continuación y se aíslan del bazo las células que producen el anticuerpo. Las células del bazo se fusionan mediante polietilenglicol con células de mieloma de ratón y el exceso de células no fusionadas se destruye haciendo crecer el sistema sobre medios selectivos que comprenden aminopterina (medios HAT). Las células fusionadas con éxito se diluyen y partes alícuotas de la dilución se colocan en pocillos de una placa de microtitulación, en donde continúa el crecimiento del cultivo. Los clones que producen anticuerpos se identifican por la detección del anticuerpo en el fluido sobrenadante de los pocillos, mediante procedimientos de inmunoensayo, tales como ELISA, tal y como han descrito originalmente Engwall, *Meth. Enzymol*. **70**:419, 1980 y métodos derivados de los mismos.

Los clones positivos seleccionados se pueden expandir y su producto de anticuerpo monoclonal se recolecta para el uso. Procedimientos detallados para la producción de anticuerpos monoclonales se describen en Davis *y col., Basic Methods in Molecular Biology* Elsevier, Nueva York. Sección 21-2.

2. *Producción de anticuerpos policlonales mediante inmunización*

El antisuero policlonal que contiene anticuerpos contra epítopos heterogéneos de una proteína aislada, se puede preparar inmunizando animales adecuados con la proteína o el péptido expresado obtenido a partir de la misma, que puede estar sin modificar o modificado para mejorar la inmunogenicidad. La producción eficaz de anticuerpos monoclonales está afectada por muchos factores relacionados con el antígeno y con la especie hospedadora. Por ejemplo, las moléculas pequeñas tienden a ser menos inmunógenas que otras y pueden requerir el uso de vehículos y coadyuvantes. También, la respuesta de los animales hospedadores varía dependiendo del sitio de las inoculaciones y de las dosis, ya que las dosis inadecuadas o excesivas de antígeno dan como resultado un título menor de antisueros. Las dosis pequeñas (a nivel de ng) de antígeno administrado en sitios intradérmicos múltiples, parecen ser las más adecuadas. Un protocolo de inmunización eficaz para conejos se puede encontrar en Vaitukaitis *y col., J. Clin. Endocrinol. Metab.* **33**:988-991, 1971.

Las inyecciones reforzadas se pueden proporcionar a intervalos regulares y el antisuero se recolecta cuando el título de anticuerpos del mismo comienza a disminuir, tal y como se determina semicuantitativamente, por ejemplo, mediante inmunodifusión doble en agar frente a concentraciones conocidas del antígeno. Véase, por ejemplo, Ouchterlony, *y col.,* Capítulo 19 en: *Handbook of Experimental Immunology* D. Wier (compilador) Blackwell (1973). La concentración de la fase estable del anticuerpo se encuentra generalmente en el intervalo de 0,1 a 0,2 mg/ml de suero (aproximadamente 12 μM). La afinidad de los antisueros hacia el antígeno se determina preparando curvas de unión competitiva, tal y como describen, por ejemplo, Fisher, D., Capítulo 42 en: *Manual of Clinical Immunology*, 2ª ed. (Rose y Friedman, compiladores) Amer. Soc. For Microbiol., Washington, D.C., 1980.

Las preparaciones de anticuerpos preparadas según cualquiera de los dos protocolos, son útiles en inmunoensayos cuantitativos que determinan las concentraciones de sustancias portadoras de antígeno en muestras biológicas; también se emplean de forma semicuantitativa o cualitativa para identificar la presencia de antígeno en una muestra biológica. Los anticuerpos también se pueden emplear en composiciones terapéuticas para destruir células que expresan la proteína o reducir los niveles de proteína en el cuerpo.

V. *Uso de las ESTs 5' o de secuencias obtenibles a partir de las mismas o de partes de las mismas, como reactivos*

Las ESTs 5' descritas en esta memoria (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se pueden utilizar como reactivos en procedimientos de aislamiento, ensayos de diagnóstico y procedimientos forenses. Por ejemplo, las secuencias procedentes de las ESTs 5'(o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se pueden marcar de forma detectable y emplear como sondas para aislar otras secuencias capaces de hibridarse con las mismas. Además, las secuencias de las ESTs 5'(o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se pueden usar para diseñar cebadores de la PCR para uso en procedimientos de aislamiento, diagnósticos o forenses.

1. *Uso de las ESTs 5' o de secuencias obtenibles a partir de las mismas o de partes de las mismas en procedimientos de aislamiento, diagnósticos o forenses*

Ejemplo 44

Preparación de cebadores para la PCR y amplificación del ADN

Las secuencias ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se pueden emplear para preparar cebadores de la PCR para una variedad de aplicaciones que incluyen procedimientos de aislamiento para clonar ácidos nucleicos capaces de hibridarse con tales secuencias, técnicas de diagnóstico y técnicas forenses. Los cebadores de la PCR tienen al menos 10 bases y preferentemente al menos 12, 15 o 17 bases de longitud. Más preferentemente, los cebadores de la PCR tienen al menos 20-30 bases de longitud. En algunas realizaciones, los cebadores de la PCR pueden tener más de 30 bases de longitud. Se prefiere que las parejas de cebadores tengan aproximadamente la misma proporción de G/C, de modo que las temperaturas de fusión sean aproximadamente las mismas. Una variedad de técnicas de la PCR son conocidas por los expertos en la técnica. Para una revisión de la tecnología de la PCR, véase, Molecular Cloning to Genetic Engineering, compilador White, en: *Methods in Molecular Biology* **67**: Humana Press, Totowa 1997. En cada uno de estos procedimientos con la PCR, los cebadores de la PCR en cada uno de los lados de las secuencias de ácido nucleico que se van a amplificar, se añaden a una muestra de ácido nucleico preparada de forma adecuada, junto con los dNTPs y una polimerasa termoestable, tal como la polimerasa Taq, la polimerasa Pfu o la polimerasa Vent. El ácido nucleico en la muestra se desnaturaliza y los cebadores de la PCR se hibridan específicamente con las secuencias de ácido nucleico complementarias en la muestra. Los cebadores hibridados se extienden. A continuación, se inicia otro ciclo de desnaturalización, hibridación y extensión. Los ciclos se repiten varias veces hasta producir un fragmento amplificado que contiene la secuencia de ácido nucleico entre los sitios del cebador.

Ejemplo 45

Uso de las ESTs 5' como sondas

Sondas obtenidas a partir de las ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) que incluyen ADNc de longitud completa o secuencias genómicas, se pueden marcar con etiquetas detectables, conocidas por los expertos en la técnica, que incluyen radioisótopos y etiquetas no radiactivas, para proporcionar una sonda detectable. La sonda detectable puede ser monocatenaria o bicatenaria y se puede preparar empleando técnicas conocidas en la materia, que incluyen la transcripción *in vitro*, el traslado de muescas o las reacciones de quinasas. Una muestra de ácido nucleico que contiene una secuencia capaz de hibridarse con la sonda marcada se pone en contacto con la sonda marcada. Si el ácido nucleico en la muestra es bicatenario, se puede desnaturalizar antes de ponerlo en contacto con la sonda. En algunas aplicaciones la muestra de ácido nucleico se puede inmovilizar sobre una superficie, tal como una membrana de nitrocelulosa o de nailon. La muestra de ácido nucleico puede comprender ácidos nucleicos obtenidos a partir de una variedad de fuentes, que incluyen el ADN genómico, genotecas de ADNc o muestras de tejidos.

Los procedimientos empleados para detectar la presencia de ácidos nucleicos capaces de hibridarse con la sonda detectable, incluyen técnicas bien conocidas tales como la transferencia Southern, la transferencia Northern, la hibridación puntual, la hibridación de colonias y la hibridación de placas. En algunas aplicaciones el ácido nucleico capaz de hibridarse con la sonda marcada se puede clonar en vectores tales como vectores de expresión, vectores de secuenciación o vectores de transcripción *in vitro*, para facilitar la caracterización y la expresión de los ácidos nucleicos que se hibridan en la muestra. Por ejemplo, tales técnicas se pueden emplear para asilar y clonar secuencias en una genoteca genómica o una genoteca de ADNc que son capaces de hibridarse con la sonda detectable, tal y como se ha descrito en el Ejemplo 30 anterior.

Los cebadores de la PCR preparados tal y como se ha descrito en el Ejemplo 44 anterior, se pueden utilizar en análisis forenses, tales como las técnicas de huella genética del ADN, descritas en los Ejemplos 46-50 siguientes. Tales análisis pueden emplear sondas o cebadores detectables, basándose en las secuencias de las ESTs 5' o de los ADNc o ADNs genómicos aislados empleando las ESTs 5'.

Ejemplo 46

Apareamiento forense mediante secuenciación del ADN

En un método ejemplar, se aíslan muestras de ADN a partir de especimenes forenses de, por ejemplo, cabello, semen, sangre o células de la piel según métodos convencionales. Un panel de cebadores de la PCR que se basa en una cantidad de las ESTs 5' del Ejemplo 25, o los ADNc o los ADNs genómicos aislados a partir de las mismas tal y como se ha descrito anteriormente, se utiliza a continuación de acuerdo con el Ejemplo 44 para amplificar el ADN de aproximadamente 100-200 bases de longitud, procedente del espécimen forense. Las secuencias correspondientes se obtienen a partir de un individuo sometido a ensayo. Cada uno de estos ADNs de identificación se secuencia a continuación, empleando técnicas convencionales y una simple comparación de bases de datos determina las diferencias, si es que existe alguna, entre las secuencias del individuo y las de la muestra. Diferencias estadísticamente significativas entre las secuencias del ADN sospechoso y las de la muestra, prueban de forma concluyente una falta de identidad. Esta falta de identidad se puede probar, por ejemplo, sólo con una secuencia. La identidad, por otro lado, se tiene que demostrar con una alto número de secuencias, todas ellas emparejadas. Preferentemente, se emplea un mínimo de 50 secuencias estadísticamente idénticas de 100 bases de longitud, para probar la identidad entre el sospechoso y la muestra.

Ejemplo 47

Identificación positiva mediante secuenciación del ADN

La técnica descrita en el ejemplo anterior también se puede utilizar a mayor escala para proporcionar una identificación única del tipo huella genética de cualquier individuo. En esta técnica, los cebadores se preparan a partir de una alto número de secuencias de ESTs 5' procedentes del Ejemplo 25, o secuencias de ADNc o de ADNs genómicos obtenibles a partir de las mismas. Preferentemente, se emplean de 20 a 50 cebadores diferentes. Estos cebadores se emplean para obtener una cantidad correspondiente de segmentos de ADN generados con PCR, a partir del individuo en cuestión, de acuerdo con el Ejemplo 44. Cada uno de estos segmentos de ADN se secuencia empleando los métodos descritos en el Ejemplo 46. La base de datos de las secuencias generadas mediante este procedimiento, identifica únicamente el individuo a partir del cual se han obtenido las secuencias. El mismo panel de cebadores se puede utilizar a continuación en cualquier momento posterior para poner en correlación de forma absoluta, tejido u otro espécimen biológico con ese individuo.

Ejemplo 48

Identificación forense con transferencia Southern

El procedimiento del Ejemplo 47 se repite para obtener un panel de al menos 10 secuencias amplificadas procedentes de un individuo y un espécimen. Preferentemente, el panel contiene al menos 50 secuencias amplificadas. Más preferentemente, el panel contiene 100 secuencias amplificadas. En algunas realizaciones, el panel contiene 200 secuencias amplificadas. Este ADN generado con la PCR se digiere a continuación con una enzima de restricción o una combinación de, preferentemente, enzimas de restricción específicas para cuatro bases. Tales enzimas están a disposición comercial y son conocidas por los expertos en la técnica. Después de la digestión, los fragmentos génicos resultantes se separan según su tamaño en múltiples pocillos por duplicado, sobre un gel de agarosa y se transfieren a nitrocelulosa empleando técnicas de transferencia Southern, bien conocidas por los expertos en la técnica. Para una revisión de las transferencias Southern, véase Davis *y col*. (Basic Methods in Molecular Biology, 1986, Elsevier Press, págs. 62-65).

Un panel de sondas basadas en las secuencias de las ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas), o fragmentos de las mismas de al menos 10 bases, se marcan radiactiva o colorimétricamente empleando métodos conocidos en la técnica, tales como el traslado de muesca o la marcación del extremo y se hibridan con la transferencia Southern empleando técnicas conocidas por los expertos (Davis *y col., véase más arriba*). Preferentemente, la sonda comprende al menos 12, 15 o 17 nucleótidos consecutivos procedentes de la EST 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de la misma). Más preferentemente, la sonda comprende al menos 20-30 nucleótidos consecutivos procedentes de la EST 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de la misma). En algunas realizaciones, la sonda comprende más de 30 nucleótidos procedentes de la EST 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de la misma).

Preferentemente, al menos 5 a 10 de estas sondas marcadas se emplean para proporcionar un patrón único y aún más preferentemente al menos aproximadamente 20 o 30. Las bandas resultantes que aparecen por la hibridación de una muestra grande de EST 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de la misma) serán un identificador único. Puesto que la escisión con las enzimas de restricción será diferente para cada individuo, el patrón de las bandas en la transferencia Southern, también será único. Un incremento del número de sondas de las ESTs 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de las mismas) proporcionará un nivel de confianza estadísticamente superior en la identificación, ya que habrá un número incrementado de grupos de bandas empleadas para la identificación.

Ejemplo 49

Procedimiento de identificación con hibridación puntual

Otra técnica para identificar individuos empleando las secuencias ESTs 5' descritas en esta memoria, emplea una técnica de hibridación puntual.

El ADN genómico se aísla a partir de los núcleos del individuo que se va a identificar. Las sondas de oligonucleótidos de aproximadamente 30 pb de longitud se sintetizan de forma que se correspondan con al menos 10, preferentemente 50 secuencias procedentes de las ESTs 5' o de los ADNc o ADNs genómicos obtenibles a partir de las mismas. Las sondas se emplean para hibridarse con el ADN genómico mediante condiciones conocidas en la técnica. Los oligonucleótidos se marcan en el extremo con ^{32}P empleando una quinasa de polinucleótidos (Pharmacia). Las hibridaciones puntuales se crean dejando manchas del ADN genómico sobre la nitrocelulosa o similar, empleando dispositivo a vacío para hibridación puntual (BioRad, Richmond California). El filtro de nitrocelulosa que contiene las secuencias genómicas se cuece o se somete a UV, se prehibrida y se hibrida con la sonda marcada empleando técnicas conocidas por los expertos en la materia (Davis *y col., véase más arriba*). Los fragmentos de ADN marcados con ^{32}P se hibridan secuencialmente bajo 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 poca cantidad de desemparejamientos de nucleótidos (Wood *y col., Proc. Natl. Acad. Sci. USA* **82(6)**:1585-1588, 1985). Un patrón único de hibridaciones puntuales distingue un individuo de otro individuo.

Las secuencias ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) o los oligonucleótidos que contienen al menos 10 bases consecutivas procedentes de estas secuencias, se pueden utilizar como sondas en la siguiente técnica alternativa de huella genética. Preferentemente, la sonda comprende al menos 12, 15 o 17 nucleótidos consecutivos procedentes de las secuencias ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas). Más preferentemente, la sonda comprende al menos 20-30 nucleótidos consecutivos procedentes de las secuencias ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas). En algunas realizaciones, la sonda comprende más de 30 nucleótidos procedentes de las secuencias ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas).

Preferentemente, una pluralidad de sondas que tienen secuencias procedentes de diferentes genes se emplea en la técnica alternativa de huella genética. El Ejemplo 50 siguiente proporciona un procedimiento alternativo representativo de la huella genética en el que las sondas se obtienen a partir de las ESTs 5'.

Ejemplo 50

Técnica alternativa para identificación con "huella genética"

Se prepararon oligonucleótidos de 20-meros a partir de un número elevado de ESTs 5', p. ej., 50, 100 o 200, empleando servicios de suministro de oligonucleótidos disponibles comercialmente, tales como Genset, París, Francia. Las muestras celulares procedentes del individuo del ensayo se procesaron para ADN empleando técnicas bien conocidas por los expertos en la materia. El ácido nucleico se digiere con enzimas de restricción tales como EcoRI y XbaI. Después de la digestión, las muestras se aplican a pocillos para electroforesis. El procedimiento, tal y como se conoce en la técnica, se puede modificar para acomodar la electroforesis de poliacrilamida, sin embargo, en este ejemplo las muestras que contienen 5 μg de ADN se cargan en los pocillos y se separan sobre geles de agarosa al 0,8%. Los geles se transfieren a nitrocelulosa empleando técnicas de transferencia Southern convencionales.

Diez ng de cada uno de los oligonucleótidos se reúnen y se marcan en el extremo con ^{32}P. La nitrocelulosa se prehibrida con solución bloqueante y se hibrida con las sondas marcadas. Después de la hibridación y el lavado, el filtro de nitrocelulosa se expone a una película de rayos X, X-Omat AR. El patrón de la hibridación resultante será único para cada individuo.

Se contempla adicionalmente en este ejemplo que el número de secuencias de las sondas empleadas puede variar para una mayor exactitud o claridad.

Las proteínas codificadas por los ADNc extendidos también se pueden utilizar para generar anticuerpos, tal y como se ha explicado en los Ejemplos 30 y 43 para identificar el tipo de tejido o la especie celular a partir de la cual se obtiene una muestra, tal y como se describe en el Ejemplo 51.

Ejemplo 51

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

La identificación de tejidos específicos se realiza mediante la visualización de antígenos específicos del tejido con preparaciones de anticuerpos según los Ejemplos 30 y 43, que se conjugan directa o indirectamente con un marcador detectable. Las especies de anticuerpos marcadas y seleccionadas se unen a sus parejas con unión específica al antígeno en secciones de tejidos, suspensiones celulares o en extractos de proteínas solubles procedentes de una muestra de tejido para proporcionar un patrón de interpretación cualitativa o semi-cualitativa.

Los antisueros para estos procedimientos deben tener una potencia que exceda la de la preparación natural y por esta razón, los anticuerpos se concentran hasta 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. También, para proporcionar los antisueros más específicos, los anticuerpos no deseados, por ejemplo para proteínas comunes, se tienen que retirar de la fracción de gammaglobulina, por ejemplo, mediante inmunoabsorbentes insolubles, antes de que los anticuerpos se marquen con el marcador. Los antisueros monoclonales o heterólogos son adecuados para los dos procedimientos.

A. *Técnicas inmunohistoquímicas*

Los anticuerpos purificados con título alto, preparados tal y como se ha descrito anteriormente, se conjugaron con un marcador detectable, tal y como ha descrito, por ejemplo, Fudenberg, Capítulo 26 en; *Basis and Clinical Immunology*, 3ª ed. Lange, Los Altos, California, 1980 o Rose *y col*., Capítulo 12 in: *Methods in Immunodiagnosis*, 2ª ed. John Wiley and Sons, Nueva York (1980).

Un marcador fluorescente, tanto fluoresceína como rodamina, se prefiere, pero los anticuerpos también se pueden marcar con una enzima que retenga un color que produzca una reacción con un sustrato, tal como la peroxidasa de rábano picante. Los marcadores se pueden añadir al anticuerpo unido al tejido en una segunda etapa, tal y como se describe a continuación. Alternativamente, los anticuerpos específicos anti-tejido se pueden marcar con ferritina o con partículas densas en electrones y la localización de la ferritina acoplada con los complejos antígeno-anticuerpo se consigue mediante un microscopio electrónico. En aún otro planteamiento, los anticuerpos se marcan radiactivamente por ejemplo con ^{125}I y se detectan sobreponiendo la preparación tratada con anticuerpo, con una emulsión fotográfica.

Las preparaciones para llevar a cabo los procedimientos pueden comprender anticuerpos monoclonales o policlonales para una proteína o un péptido aislado, identificado como específico de un tipo de tejido, por ejemplo, tejido cerebral, o las preparaciones de anticuerpos para diversos antígenos específicos de tejidos que sean antigénicamente diferentes se pueden utilizar en paneles, independientemente o en mezclas, según sea necesario.

Las secciones de tejido y las suspensiones celulares se preparan para el examen inmunohistoquímico según técnicas histológicas comunes. Secciones múltiples de criostato (aproximadamente 4 μm, sin fijar) del tejido desconocido y del testigo conocido, se montan y cada portaobjetos se cubre con diluciones diferentes de la preparación de anticuerpo. Las secciones de tejidos conocidos y desconocidos se deben tratar también con preparaciones para proporcionar un

testigo positivo, un testigo negativo, por ejemplo, sueros pre-inmunizados y un testigo para una tinción no específica, por ejemplo, tampón.

Las secciones tratadas se incuban en una cámara húmeda durante 30 minutos a temperatura ambiente, se aclaran, a continuación se lavan en tampón durante 30-45 minutos. El exceso de fluido se retira y el marcador se revela.

Si el anticuerpo específico del tejido no se marcaba en la primera incubación, se puede marcar en ese momento con una segunda reacción de anticuerpo-anticuerpo, por ejemplo, añadiendo fluoresceína o enzima conjugada con el anticuerpo contra la clase de inmunoglobulina de las especies productoras de antisuero, por ejemplo, anticuerpo marcado con fluoresceína contra IgG de ratón. Dichos sueros marcados están a disposición comercial.

El antígeno encontrado en los tejidos mediante el procedimiento anterior, se puede cuantificar midiendo la intensidad del color o la fluorescencia sobre la sección del tejido, y calibrar esta señal empleando patrones adecuados.

B. *Identificación de proteínas solubles específicas del tejido*

La visualización de las proteínas específicas del tejido y la identificación de tejidos desconocidos a partir de este procedimiento, se realiza empleando los reactivos del anticuerpo marcado y la estrategia de detección, tal y como se describe para la inmunohistoquímica, sin embargo, la muestra se prepara según una técnica electroforética para distribuir las proteínas extraídas del tejido en una formación ordenada basada en el peso molecular, para la detección.

Una muestra de tejido se homogeneiza empleando un aparato Virtis; las suspensiones celulares se destruyen por homogeneización Dounce o por lisis osmótica, empleando detergentes, en ambos casos, tal y como es necesario para destruir las membranas celulares, como es habitual en la técnica. Los componentes celulares insolubles, tales como los núcleos, los microsomas y los fragmentos de membranas se retiran por ultracentrifugación y la fracción que contiene la proteína soluble se concentra si es necesario y se reserva para análisis.

Una muestra de la solución de proteína soluble se determina en una especie proteica individual mediante electroforesis convencional con SDS poliacrilamida, tal y como describen por ejemplo, Davis *y col*., Sección **19-2** en: *Basic Methods in Molecular Biology*, Leder compilador, Elsevier, Nueva York, 1986, empleando un intervalo de cantidades de poliacrilamida en un grupo de geles para determinar el intervalo del peso molecular total de las proteínas que se van a detectar en la muestra. Un marcador del tamaño se deja correr paralelamente para estimar los pesos moleculares de las proteínas constituyentes. El tamaño de la muestra a analizar es un volumen conveniente desde 5 hasta 55 μl que contiene desde aproximadamente 1 a 100 μg de proteína. Una parte alícuota de cada una de las proteínas determinadas, se transfiere por transferencia a un papel de filtro de nitrocelulosa, un procedimiento que mantiene el patrón de resolución. Se preparan múltiples copias. El procedimiento conocido como análisis de transferencia Western, está bien descrito por Davis, L. *y col., véase más arriba*, Sección 19-3. Un grupo de transferencias de nitrocelulosa se tiñe con colorante azul de Coomassie para visualizar el grupo completo de proteínas, para comparar con las proteínas unidas al anticuerpo. Los filtros de nitrocelulosa restantes se incuban a continuación con una solución de uno o varios antisueros específicos de las proteínas específicas del tejido, preparados tal y como se ha descrito en los Ejemplos 30 y 43. En este procedimiento, como en el procedimiento A anterior, se dejan correr muestras adecuadas positivas y negativas y testigos reactivos.

Tanto en el procedimiento A como B, se puede fijar una etiqueta detectable al complejo de antígeno del tejido primario-anticuerpo primario, según diversas estrategias y sus permutaciones. En un planteamiento directo, el anticuerpo primario específico se puede marcar; alternativamente, el complejo sin marcar se puede unir a un anticuerpo secundario anti-IgG marcado. En otros planteamientos, tanto el anticuerpo primario como el secundario se conjugan con una molécula de biotina que se puede unir, en una etapa posterior, a un marcador conjugado con avidina. De acuerdo con todavía otra estrategia, la proteína A marcada enzimáticamente o radiactiva que tiene la propiedad de unirse a cualquier IgG, se une en una etapa final al anticuerpo primario o secundario.

Con la visualización de la unión del antígeno específico del tejido con niveles superiores a los observados en los tejidos testigos, con uno o varios anticuerpos específicos del tejido, preparados a partir de las secuencias génicas identificadas a partir de las secuencias de ADNc extendido, se pueden identificar tejidos de origen desconocido, por ejemplo, muestras forenses, o tejido tumoral diferenciado que se ha sufrido metástasis en otros sitios del cuerpo.

Además de sus aplicaciones forenses y en la identificación, las ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se pueden cartografiar para localizar sus posiciones cromosómicas. El Ejemplo 52 siguiente describe el cartografiado de híbridos radiados (RH) de regiones cromosómicas humanas, empleando las ESTs 5'. El Ejemplo 53 siguiente describe un procedimiento representativo para cartografiar una EST 5' en su localización sobre un cromosoma humano. El Ejemplo 54 siguiente describe el cartografiado de las ESTs 5' sobre cromosomas en metafase mediante hibridación *in situ* con fluorescencia (FISH, del inglés, " Fluorescence *In Situ* Hybridization"). Los expertos en la técnica apreciarán que el método de los Ejemplos 52-54 también se puede emplear para cartografiar ADNc o ADNs genómicos obtenibles a partir de las ESTs 5' en sus localizaciones cromosómicas.

2. *Uso de las ESTs 5' o de secuencias obtenibles a partir de las mismas o de partes de las mismas para cartografiar cromosomas*

Ejemplo 52

Cartografiado de híbridos radiados de las ESTs 5' en el genoma humano

El cartografiado de híbridos radiados (RH) es un planteamiento genético en células somáticas que se puede utilizar para el cartografiado de alta resolución del genoma humano. En este planteamiento, líneas celulares que contienen uno o varios cromosomas humanos se irradian letalmente, rompiendo cada cromosoma en fragmentos cuyos tamaños dependen de la dosis de la radiación. Estos fragmentos se rescatan por fusión con células de roedores cultivadas, proporcionando subclones que contienen diferentes partes del genoma humano. Esta técnica la describen Benham *y col., Genomics* **4**:509-517, 1989; Cox *y col., Science* **250**:245-250, 1990. La naturaleza aleatoria e independiente de los subclones permite un cartografiado eficaz de cualquier marcador del genoma humano. El ADN humano aislado a partir de un panel de 80-100 líneas celulares, proporciona un reactivo para el cartografiado para ordenar las ESTs 5'. En este planteamiento, la frecuencia de las roturas entre los marcadores se emplea para medir la distancia, permitiendo la construcción de mapas de alta resolución como se ha ido haciendo empleando ESTs convencionales (Schuler *y col., Science* **274**:540-546, 1996).

El cartografiado de RH se ha empleado para generar un mapa con híbridos radiados del genoma completo, del cromosoma humano 17q22-q25.3, a través de los genes de la hormona de crecimiento (GH) y la quinasa de timidina (TK) (Foster *y col., Genomics* **33**:185-192, 1996), la región que rodea el gen del síndrome de Gorlin (Obermayr *y col., Eur. J. Hum. Genet*. **4**:242-245, 1996), 60 loci que cubren el brazo corto completo del cromosoma 12 (Raeymaekers *y col., Genomics* **29**:170-178, 1995), la región del cromosoma 22 humano que contiene el locus de la neurofibromatosis de tipo 2 (Frazer *y col., Genomics* **14**:574-584, 1992) y 13 loci obre el brazo largo del cromosoma 5 (Warrington *y col., Genomics* **11**:701-708, 1991).

Ejemplo 53

Cartografiado de ESTs 5' en cromosomas humanos empleando técnicas de PCR

Las ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se pueden asignar a cromosomas humanos empleando metodologías basadas en la PCR. En tales planteamientos, las parejas de cebadores de oligonucleótidos se diseñan a partir de las ESTs 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de las mismas) para minimizar la posibilidad de amplificación a través de un intrón. Preferentemente, los cebadores oligonucleótidos tienen 18-23 pb de longitud y se diseñan para la amplificación con PCR. La creación de cebadores para la PCR a partir de secuencias conocidas es bien conocida por los expertos en la técnica. Para una revisión de la tecnología de la PCR, véase Erlich en *PCR Technology, Principles and Applications for DNA Amplification*, Freeman and Co., Nueva York, 1992.

Los cebadores se emplean en reacciones en cadena de la polimerasa (PCR) para amplificar moldes procedentes de ADN genómico humano total. Las condiciones de la PCR son las siguientes: 60 ng de ADN genómico se emplean como molde para la PCR con 80 ng de cada cebador oligonucleótido, 0,6 unidades de polimerasa Taq y 1 μCu de una desoxicitidina trifosfato marcada con ^{32}P. La PCR se realiza en un termociclador de microplacas (Techne) bajo las siguientes condiciones: 30 ciclos a 94ºC, 1,4 minutos; 55ºC, 2 minutos; y 72ºC, 2 minutos; con una extensión final a 72ºC durante 10 minutos. Los productos amplificados se analizan sobre un gel de secuenciación de poliacrilamida al 6% y se visualizan mediante autorradiografía. Si la longitud del producto resultante de la PCR es idéntico a la distancia entre los extremos de las secuencias de cebadores en el ADNc extendido a partir del cual se han obtenido los cebadores, entonces la reacción de la PCR se repite con moldes de ADN procedentes de dos paneles de híbridos de células somáticas humanas-roedores, ADN utilizable en PCR de BIOS (BIOS Corporation) y el panel número 1 del cartografiado de híbridos de células somáticas de ser humano-roedor de NIGMS (NIGMS, Camden, NJ).

La PCR se emplea para escrutar una serie de líneas celulares híbridas de células somáticas que contienen grupos definidos de cromosomas humanos, en busca de la presencia de una EST 5' dada (o del ADNc o del ADN genómico obtenible a partir de la misma). El ADN se aísla a partir de híbridos somáticos y se emplea como molde de partida para reacciones de la PCR que emplean parejas de cebadores procedentes de la EST 5' (o del ADNc o del ADN genómico obtenible a partir de la misma). Sólo los híbridos de células somáticas con cromosomas que contienen el gen humano correspondiente a la EST 5' (o al ADNc o al ADN genómico obtenible a partir de la misma) proporcionarán un fragmento amplificado. La EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma) se asigna a un cromosoma analizando el patrón de segregación de los productos de la PCR procedentes de los moldes de ADN híbrido somático. El cromosoma humano aislado presente en todos los híbridos celulares que proporcionan el incremento de un fragmento amplificado, es el cromosoma que contiene esta EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma). Para una revisión de las técnicas y el análisis de resultados procedentes de experimentos de cartografiado de genes de células somáticas, véase Ledbetter *y col., Genomics* **6**:475-481, 1990.

Ejemplo 54

Cartografiado de las ESTs 5' extendidas para cromosomas empleando la hibridación in situ con fluorescencia

La hibridación *in situ* con fluorescencia permite que la EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma) se cartografíe en una posición en particular sobre un cromosoma dado. Los cromosomas que se van a emplear en las técnicas de hibridación *in situ* con fluorescencia se pueden obtener a partir de una variedad de fuentes que incluyen los cultivos celulares, los tejidos o la sangre completa.

En una realización preferida, la localización cromosómica de una EST 5' (o del ADNc o del ADN genómico obtenible a partir de la misma) se obtiene mediante FISH, tal y como se describe en Cherif *y col., Proc. Natl. Acad. Sci. USA* **87**:6639-6643, 1990). Los cromosomas en metafase se preparan a partir de donantes de células sanguíneas estimuladas con fitohemaglutinina (PHA). Los linfocitos estimulados con PHA, procedentes de machos sanos, se cultivan durante 72 h en medio RPMI-1640. Para la sincronización, se añade metotrexato (10 μM) durante 17 h, seguido de la adición de 5-bromodesoxiuridina (5-BrdU, 0,1 mM) durante 6 h. La colcemida (1 μg/ml) se añade durante al menos 15 min., antes de recolectar las células. Las células se recolectan, se lavan en RPMI, se incuban con una solución hipotónica de KCl (75 mM) a 37ºC durante 15 min y se fijan en tres cambios de metanol:ácido acético (3:1). La suspensión celular se gotea sobre un portaobjetos de vidrio y se seca al aire. La EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma) se marca con dUTP biotina-16 mediante traslado de muesca, de acuerdo con las instrucciones del fabricante (Bethesda Research Laboratories, Bethesda, MD), se purifica empleando una columna G-50 de Sephadex (Pharmacia, Upsala, Suecia) y se precipita. Justo antes de la hibridación, el sedimento de ADN se disuelve en el tampón de hibridación (50% de formamida, 2x SSC, 10% de sulfato de dextrano, 1 mg/ml de ADN de esperma de salmón sometido a ultrasonidos, pH 7) y la sonda se desnaturaliza a 70ºC durante 5-10 minutos.

Los portaobjetos mantenidos a -20ºC y tratados durante 1 hora a 37ºC con ARNasa A (100 μg/ml), se lavan tres veces en 2x SSC y se deshidratan en una serie con etanol. Las preparaciones de cromosomas se desnaturalizan en formamida al 70%, 2x SSC durante 2 min. A 70ºC y a continuación se deshidratan a 4ºC. Los portaobjetos se tratan con proteinasa K (10 μg/100 ml en Tris-HCl 20 mM, $CaCl_2$ 2 mM) a 37ºC durante 8 min. y se deshidratan. La mezcla de hibridación que contenía la sonda se colocó sobre el portaobjetos, se cubre con un cubreobjetos, se sella con un adhesivo de goma y se incuba durante una noche en una cámara húmeda a 37ºC. Después de la hibridación y de los lavados posteriores a la hibridación, la sonda biotinilada se detecta mediante avidina-FITC y se amplifica con capas adicionales de anti-avidina de cabra biotinilada y avidina-FITC. Para la localización en los cromosomas, se obtienen bandas R fluorescentes tal y como se ha descrito previamente (Cherif *y col., véase más arriba*). Los portaobjetos se observan con un microscopio de fluorescencia de LEICA (DMRXA). Los cromosomas se someten a contracoloración con yoduro de propidio y la señal fluorescente de la sonda aparece como dos manchas simétricas amarillo-verdosas en ambas cromátidas del cromosoma con la banda R fluorescente (red). Por tanto, una EST 5' en particular (o el ADNc o el ADN genómico obtenible a partir de la misma) se puede localizar en una banda R citogenética particular sobre un cromosoma dado.

Una vez que se ha asignado a las ESTs 5' (o al ADNc o al ADN genómico obtenible a partir de las mismas) a cromosomas particulares empleando las técnicas descritas en los Ejemplos 52-54 anteriores, se pueden utilizar para construir un mapa de alta resolución de los cromosomas sobre los que están localizadas o para identificar los cromosomas en una muestra.

Ejemplo 55

Uso de ESTs 5' para construir o ampliar mapas cromosómicos

El cartografiado de cromosomas implica asignar a una secuencia dada única, un cromosoma en particular, tal y como se ha descrito anteriormente. Una vez que la secuencia única se ha cartografiado en un cromosoma dado, se ordena en relación con otras secuencias únicas localizadas sobre el mismo cromosoma. Un planteamiento para el cartografiado de cromosomas emplea una serie de cromosomas artificiales de levadura (YACs) que son portadores de algunos insertos de mil de bases, obtenidos a partir de los cromosomas de los organismos a partir de los cuales se han obtenido los ADNc extendidos (o los ADNs genómicos obtenibles a partir de los mismos). Este planteamiento se describe en Nagaraja *y col., Genome Research* **7**:210-222, 1997. Resumiendo, en este planteamiento cada cromosoma se rompe en piezas solapantes que se insertan en el vector YAC. Los insertos del YAC se escrutan empleando PCR o con otros métodos para determinar si incluyen la EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma) cuya posición se va a determinar. Una vez que se ha encontrado un inserto que incluye la EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma), se puede analizar el inserto con la PCR o con otros métodos para determinar si el inserto también contiene otras secuencias que se sabe que están sobre el cromosoma o en la región a partir de la cual se ha obtenido la EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma). Este procedimiento se puede repetir para cada inserto en la genoteca de los YAC para determinar la posición de cada uno de los ADNc extendidos (o de los ADNs genómicos obtenibles a partir de los mismos) en relación con otro y con marcadores cromosómicos conocidos. De este modo, se puede obtener un mapa de alta resolución de la distribución de numerosos marcadores únicos a lo largo de cada uno de los cromosomas de los organismos.

Tal y como se describe en el Ejemplo 56 siguiente, los ADNc extendidos (o los ADNs genómicos obtenibles a partir de los mismos) también se pueden utilizar para identificar genes asociados con un fenotipo particular, tal como una enfermedad hereditaria o una respuesta a fármacos.

3. *Uso de las ESTs 5' o de secuencias obtenidas a partir de las mismas o de fragmentos de las mismas en la identificación génica*

Ejemplo 56

Identificación de genes asociados con enfermedades hereditarias o de respuesta a fármacos

Este Ejemplo ilustra un planteamiento útil para la asociación de las ESTs 5' (o el ADNc o el ADN genómico obtenible a partir de las mismas) con características fenotípicas particulares. En este ejemplo, una EST 5' particular (o el ADNc o el ADN genómico obtenible a partir de la misma) se emplea como una sonda de ensayo para asociar esta EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma) con una característica fenotípica particular.

Las ESTs 5' (o el ADNc o el ADN genómico obtenible a partir de las mismas) se cartografían en una posición particular sobre un cromosoma humano empleando técnicas tales como las descritas en los Ejemplos 52 y 53 u otras técnicas conocidas en la materia. Una búsqueda de la herencia mendeliana en hombres (McKusick en *Mendelian Inheritance in Man* (disponible en línea a través de la biblioteca médica Welch de la universidad John Hopkins) revela que la región del cromosoma humano que contiene la EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma) es una región muy rica en genes que contiene diversos genes conocidos y diversas enfermedades o fenotipos para los que no se han identificado los genes. El gen correspondiente a esta EST 5' (o el ADNc o el ADN genómico obtenible a partir de la misma) se vuelve por consiguiente un candidato inmediato para cada una de estas enfermedades genéticas.

Las células de pacientes con estas enfermedades o fenotipos se aíslan y se amplían en cultivo. Los cebadores de la PCR procedentes de la EST 5' (o del ADNc o del ADN genómico obtenible a partir de la misma) se emplean para escrutar el ADN genómico, el ARNm o el ADNc obtenido a partir de los pacientes. Las ESTs 5' (o el ADNc o el ADN genómico obtenible a partir de las mismas) que no se han amplificado en los pacientes, se pueden asociar positivamente con una enfermedad particular mediante un análisis adicional. Alternativamente, el análisis con la PCR puede proporcionar fragmentos de longitudes diferentes cuando las muestras se obtienen a partir de un individuo que tiene el fenotipo asociado con la enfermedad, más que cuando la muestra se obtiene a partir de un individuo sano, indicando que el gen que contiene la EST 5' puede ser responsable de la enfermedad genética.

VI. *Uso de la EST 5' (o del ADNc o del ADN genómico obtenible a partir de la misma) para construir vectores*

Las presentes ESTs 5' (o el ADNc o el ADN genómico obtenible a partir de las mismas) también se pueden utilizar para construir vectores de secreción capaces de dirigir la secreción de las proteínas codificadas por los genes de los vectores. Tales vectores de secreción pueden facilitar la purificación o el enriquecimiento de las proteínas codificadas por genes insertados en ellos, reduciendo el número de proteínas de señal de fondo a partir de las cuales se tiene que purificar o enriquecer la proteína deseada. Vectores de secreción a modo de ejemplo se describen en el Ejemplo 57 siguiente.

1. *Construcción de vectores de secreción*

Ejemplo 57

Construcción de vectores de secreción

Los vectores de secreción incluyen un promotor capaz de dirigir la expresión génica en la célula, el tejido o el organismo hospedador de interés. Tales promotores incluyen el promotor del virus del sarcoma de Rous, el promotor de SV40 el promotor de citomegalovirus humano y otros promotores conocidos por los expertos en la técnica.

Una secuencia señal procedente de una EST 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de la misma) está ligada funcionalmente con el promotor de modo que el ARNm transcrito desde el promotor dirigirá directamente la traducción del péptido señal. La célula, el tejido o el organismo hospedador puede ser cualquier célula, tejido u organismo que reconozca el péptido señal codificado por la secuencia señal en la EST 5' (o en el ADNc o en el ADN genómico obtenible a partir de la misma). Los hospedadores adecuados incluyen células, tejidos u organismos de mamífero, células, tejidos u organismos aviares, 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 del inserto en marco de lectura con la secuencia señal, de modo que una proteína de fusión tal en la que el péptido señal se fusiona con la proteína codificada por el gen insertado, se expresa a partir del ARNm transcrito a partir del promotor. El péptido señal dirige la secreción extracelular de la proteína de fusión.

El vector de secreción puede ser ADN o ARN y puede estar integrado en el cromosoma del hospedador, estar mantenido de forma estable como un replicón extracromosómico en el hospedador, ser un cromosoma artificial o estar presente de forma transitoria en el hospedador. Muchos esqueletos de ácido nucleico adecuados para emplear como vectores de secreción, son conocidos por los expertos en la técnica, incluyendo los vectores retrovíricos, los vectores de SV40, los vectores del virus de papiloma bovino, los plásmidos que integran levaduras, los plásmidos episomales de

levadura, los cromosomas artificiales de levadura, los cromosomas artificiales humanos, los vectores de elemento P, los vectores de baculovirus o de plásmidos bacterianos capaces de ser introducidos de forma transitoria en el hospedador.

El vector de secreción también puede contener una señal poliA de modo que la señal poliA esté localizada aguas abajo del gen insertado en el vector de secreción.

Después de insertar el gen que codifica la proteína para la que se desea la secreción en el vector de secreción, el vector de secreción se introduce en la célula, el tejido o el organismo hospedador empleando la precipitación con fosfato cálcico, DEAE-dextrano, electroporación, transfección mediada con liposomas, partículas víricas o ADN desprotegido. La proteína codificada por el gen insertado se purifica a continuación o se enriquece a partir del material sobrenadante, empleando técnicas convencionales tales como la precipitación con sulfato de amonio, inmunoprecipitación, inmunocromatografía, cromatografía de exclusión por tamaño, cromatografía de intercambio iónico y HPLC. Alternativamente, la proteína secretada puede estar en estado suficientemente enriquecido o puro en el material sobrenadante o en el medio de crecimiento del hospedador, para permitir que se utilice para los fines pretendidos sin un enriquecimiento adicional.

Las secuencias señal también se pueden insertar en vectores diseñados para la terapia génica. En tales vectores, la secuencia señal está ligada funcionalmente a un promotor de modo que el ARNm transcrito a partir del promotor, codifica el péptido señal. Un sitio de clonación está localizado aguas abajo de la secuencia señal, de modo que un gen que codifica una proteína cuya secreción se desea, se inserta fácilmente en el vector y se fusiona con la secuencia señal. El vector se introduce en una célula hospedadora adecuada. La proteína expresada a partir del promotor, se secreta extracelularmente, produciendo de este modo un efecto terapéutico.

Las ESTs 5' también se pueden utilizar para clonar secuencias localizadas aguas arriba de las ESTs 5' que son capaces de regular la expresión génica, incluyendo las secuencias de promotor, las secuencias potenciadoras y otras secuencias aguas arriba que influyen en los niveles de transcripción o de traducción. Una vez identificadas y clonadas, estas secuencias reguladoras aguas arriba se pueden utilizar en vectores de expresión diseñados para dirigir la expresión de un gen insertado de forma espacial, temporal, de desarrollo o cuantitativa deseada. El Ejemplo 58 describe un método para clonar secuencias aguas arriba de los ADNc extendidos o de las ESTs 5'.

2. *Identificación de las secuencias aguas arriba con actividades promotoras o reguladoras*

Ejemplo 58

Uso de los ADNc extendidos o de las ESTs 5' para clonar secuencias aguas arriba del ADN genómico

Las secuencias obtenidas a partir de los ADNc extendidos o las ESTs 5' se pueden utilizar para aislar los promotores de los genes correspondientes empleando técnicas de desplazamiento cromosómico. En una técnica de desplazamiento cromosómico, que emplea el equipo de reactivos GenomeWalker®, a disposición en Clontech, se digieren cinco muestras de ADN genómico completo cada una con una enzima de restricción diferente que tiene un sito de reconocimiento de 6 bases y deja un extremo romo. Después de la digestión, los adaptadores oligonucleotídicos se ligan con cada extremo de los fragmentos de ADN genómico resultantes.

Para cada una de las cinco genotecas de ADN genómico, se realiza una primera reacción con la PCR según las instrucciones del fabricante, empleando un cebador adaptador externo, proporcionado en el equipo de reactivos y un cebador externo específico del gen. El cebador específico del gen se debe seleccionar para que sea específico del ADNc extendido o la EST 5' de interés y debe tener una temperatura de fusión, una longitud y una posición en el ADNc extendido o en la EST 5' que sea compatible con su uso en las reacciones de PCR. Cada primera reacción de la PCR contiene 5 ng de ADN genómico, 5 μl de tampón de reacción Tth 10x, 0,2 mM de cada dNTP, 0,2 μM de cada cebador adaptador externo y de cada cebador interno específico del gen, 1,1 mM de $Mg(OAc)_2$ y 1 μl de la mezcla de polimerasa Tth 50x en un volumen total de 50 μl. El ciclo de reacción para la primera reacción de la PCR es del modo siguiente: 1 min a 94ºC/2 seg a 94ºC, 3 min a 72ºC (7 ciclos)/2 seg a 94ºC, 3 min a 67ºC (32 ciclos)/5 min a 67ºC.

El producto de la primera reacción de la PCR se diluye y se emplea como un molde para una segunda reacción de la PCR, según las instrucciones del fabricante que emplea una pareja de cebadores anidados que se localizan internamente sobre el amplicón, siendo el resultado de la primera reacción de la PCR. Por ejemplo, 5 μl del producto de la reacción de la primera mezcla de la reacción PCR se puede diluir 180 veces. Las reacciones se realizan en un volumen de 50 μl que tiene una composición idéntica a la de la primera reacción de la PCR, exceptuando que se emplean los cebadores anidados. El primer cebador anidado es específico del adaptador y se proporciona con el equipo de reactivos GenomeWalker®. El segundo cebador anidado es específico del ADNc extendido o de la EST 5' en particular, para el que se va a clonar el promotor y debe tener una temperatura de fusión, una longitud y una posición en el ADNc extendido o en la EST 5' que sea compatible con su uso en las reacciones de la PCR. Los parámetros de la segunda reacción de la PCR son los siguientes: 1 min a 94ºC/2 seg a 94ºC, 3 min a 72ºC (6 ciclos)/2 seg a 94ºC, 3 min a 67ºC (25 ciclos)/5 min a 67ºC. El producto de la segunda reacción de la PCR se purifica, se clona y se secuencia empleando técnicas convencionales.

Alternativamente, dos o más genotecas de ADN genómico humano se pueden construir empleando dos o más enzimas de restricción. El ADN genómico digerido se clona en vectores que se pueden convertir en ADN monocatenario,

circular o lineal. Un oligonucleótido biotinilado que comprende al menos 15 nucleótidos de la secuencia de ADNc extendido o de la EST 5', se hibrida con el ADN monocatenario. Los híbridos entre el oligonucleótido biotinilado y el ADN monocatenario que contiene las secuencias del ADNc extendido o de la EST, se aíslan tal y como se describe en el Ejemplo 29 anterior. A continuación, el ADN monocatenario que contiene la secuencia del ADNc extendido o de la EST, se libera de las perlas y se convierte en ADN bicatenario empleando un cebador específico de la secuencia de ADNc extendido o de EST 5' o un cebador que se corresponde con una secuencia incluida en el vector de clonación. El ADN bicatenario resultante se transforma en bacterias. Los ADNs que contienen las secuencias de la EST 5' o del ADNc extendido, se identifican mediante PCR de colonias o por hibridación de colonias.

Una vez que se han clonado y se han secuenciado las secuencias genómicas aguas arriba, tal y como se ha descrito anteriormente, se pueden identificar los promotores y los sitios de inicio de la transcripción probables, dentro de las secuencias aguas arriba, comparando las secuencias aguas arriba de los ADNc extendidos o las ESTs 5' con las bases de datos que contienen sitios conocidos de inicio de la transcripción, sitios de unión a factores de transcripción o secuencias promotoras.

Además, los promotores en las secuencias aguas arriba se pueden identificar empleando vectores informadores de promotores, tal y como se describe en el Ejemplo 59.

Ejemplo 59

Identificación de promotores en secuencias clonadas aguas arriba

Las secuencias genómicas aguas arriba de los ADNc extendidos o las ESTs 5' se clonan en un vector informador promotor adecuado, tal como los vectores informadores promotores pSEAP-Basis, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer o pEGFP-1, a disposición en Clontech. Resumiendo, cada uno de estos vectores informadores promotores incluye múltiples sitios de clonación situados aguas arriba de un gen informador que codifica una proteína que se somete fácilmente a ensayo, tal como la fosfatasa alcalina secretada, la β-galactosidasa o la proteína verde fluorescente. Las secuencias aguas arriba de los ADNc extendidos o las ESTs 5' se insertan en los sitios de clonación aguas arriba del gen informador con ambas orientaciones y se introducen en una célula hospedadora adecuada. El nivel de proteína informadora se somete a ensayo y se compara con el nivel obtenido a partir de un vector que carece de un inserto en el sito de clonación. La presencia de un elevado nivel de expresión en el vector que contiene el inserto, en relación con el vector testigo, indica la presencia de un promotor en el inserto. Si es necesario, las secuencias aguas arriba se pueden clonar en vectores que contienen un potenciador para aumentar los niveles de transcripción de las secuencias con promotores débiles. Un nivel significativo de expresión superior al observado con el vector que carece de inserto, indica que una secuencia promotora está presente en la secuencia insertada aguas arriba.

Las células hospedadoras adecuadas para los vectores informadores promotores se pueden escoger basándose en los resultados de la determinación descrita anteriormente de los patrones de expresión de los ADNc extendidos y las ESTs. Por ejemplo, si el análisis del patrón de expresión indica que el ARNm correspondiente a un ADNc extendido o una EST 5' en particular, se expresa en fibroblastos, el vector informador promotor se puede introducir en una línea celular de fibroblastos humanos.

Las secuencias promotoras dentro del ADN genómico aguas arriba, se pueden definir adicionalmente construyendo deleciones anidadas en el ADN aguas arriba, empleando técnicas convencionales, tales como la digestión con Exonucleasa III. Los fragmentos resultantes de la deleción se pueden insertar en el vector informador promotor para determinar si la deleción ha reducido o anulado la actividad del promotor. De este modo, se pueden definir los límites de los promotores. Si se desea, se pueden identificar sitios reguladores individuales y potenciales dentro del promotor, empleando la mutagénesis dirigida al sitio o el examen de enlazadores para destruir los sitios potenciales de unión al factor de transcripción dentro del promotor, de forma individual o en combinación. Los efectos de estas mutaciones sobre los niveles de la transcripción se pueden determinar insertando las mutaciones en los sitios de clonación en los vectores informadores del promotor.

Ejemplo 60

Clonación e identificación de promotores

Empleando el método descrito en el Ejemplo 58 anterior con ESTs 5', se obtuvieron secuencias aguas arriba de diversos genes. Empleando los pares de cebadores GGG AAG ATG GAG ATA GTA TTG CCT G (SEQ ID NO: 29) y CTG CCA TGT ACA TGA TAG AGA GAT TC (SEQ ID NO: 30), se obtuvo el promotor que tenía la designación interna P13H2 (SEQ ID NO: 31).

Empleando las parejas de cebadores GTA CCA GGGG ACT GTG ACC ATT GC (SEQ ID NO: 32) y CTG TGA CCA TTG CTC CCA AGA GAG (SEQ ID NO: 33), se obtuvo el promotor que tenía la designación interna P15B4 (SEQ ID NO: 34).

Empleando las parejas de cebadores CTG GGA TGG AAG GCA CGG TA (SEQ ID NO: 35) y GAG ACC ACA CAG CTA GAC AA (SEQ ID NO: 36), se obtuvo el promotor que tenía la designación interna P29B6 (SEQ ID NO: 37).

La Figura 4 proporciona una descripción esquemática de los promotores aislados y del modo en que se ensamblan con las etiquetas 5' correspondientes. Las secuencias aguas arriba se escrutaron en búsqueda de la presencia de motivos que se parecieran a sitios de unión a factores de transcripción o sitios de inicio de la transcripción conocidos, empleando el programa de ordenador MatInspector release 2.0, Agosto 1996.

La Tabla VII describe los sitios de unión del factor de transcripción, presentes en cada uno de estos promotores. La columna con la denominación matriz, proporciona el nombre de la matriz MatInspector utilizada. La columna con la denominación posición, proporciona la posición 5' del sitio del promotor. La numeración de la secuencia comienza en el sitio de la transcripción tal y como se determina emparejando la secuencia genómica con la secuencia EST 5'. La columna denominada "orientación", indica la hebra de ADN sobre la que se encuentra el sitio, siendo la hebra + la hebra codificadora tal y como se determina por el emparejamiento de la secuencia genómica con la secuencia de la EST 5'. La columna denominada "puntuación", proporciona la puntuación de MatInspector encontrada para este sitio. La columna denominada "longitud", proporciona la longitud del sitio en nucleótidos. La columna denominada "secuencia", proporciona la secuencia del sitio encontrado.

Los clones bacterianos que contienen plásmidos que contienen las secuencias del promotor descritas anteriormente, están almacenados realmente en los laboratorios de los inventores con los números de identificación interna proporcionados anteriormente. Los insertos se pueden recuperar de los materiales depositados dejando crecer una parte alícuota del clon bacteriano adecuado en el medio adecuado. El ADN plasmídico se puede aislar a continuación, empleando procedimientos de aislamiento de plásmidos conocidos por los expertos en la técnica, tales como minipreparados para lisis alcalina o los procedimientos de aislamiento de plásmidos a gran escala mediante lisis alcalina. Si se desea, el ADN del plásmido se puede enriquecer adicionalmente por centrifugación sobre un gradiente de cloruro de cesio, cromatografía de exclusión por tamaño o cromatografía de intercambio aniónico. El ADN plasmídico obtenido empleando estos procedimientos se puede manipular a continuación empleando técnicas de clonación conocidas por los expertos en la técnica. Alternativamente, se puede realizar una PCR con los cebadores diseñados en ambos extremos de la inserción de la EST. El producto de la PCR que se corresponde con la EST 5' se puede manipular a continuación empleando técnicas de clonación conocidas por los expertos en la técnica.

Los promotores y otras secuencias reguladoras localizadas aguas arriba de los ADNc extendidos o de las ESTs 5', se pueden utilizar para diseñar vectores de expresión capaces de dirigir la expresión de un gen insertado de forma deseada, espacial, temporal, para desarrollo o cuantitativa. Un promotor capaz de dirigir los patrones deseados espaciales, temporales, de desarrollo y cuantitativos, se puede seleccionar empleando los resultados del análisis de la expresión, descritos en el Ejemplo 26 anterior. Por ejemplo, si se desea un promotor que confiere un alto nivel de expresión en el músculo, la secuencia del promotor aguas arriba de un ADNc extendido o de una EST 5' obtenida a partir de un ARNm que se expresa con una alto nivel en músculo, tal y como se ha determinado por el método del Ejemplo 26, se puede utilizar en el vector de expresión.

Preferentemente, el promotor deseado se coloca cerca de múltiples sitios de restricción para facilitar la clonación del inserto deseado aguas abajo del promotor, de modo que el promotor es capaz de dirigir la expresión del gen insertado. El promotor se puede insertar en esqueletos convencionales de ácidos nucleicos, diseñados para la replicación extracromosómica, para la integración en los cromosomas hospedadores o para la expresión transitoria. Los esqueletos adecuados para los vectores de expresión presentes, incluyen esqueletos retrovíricos, esqueletos de episomas eucarióticos, tales como SV40 o el virus del papiloma bovino, esqueletos de episomas bacterianos o cromosomas artificiales.

Preferentemente, los vectores de expresión también incluyen una señal poliA, aguas abajo de los sitios múltiples de restricción, para dirigir la poliadenilación del ARNm transcrito a partir del gen insertado en el vector de expresión.

Después de identificar las secuencias del promotor empleando los procedimientos de los Ejemplos 58-60, las proteínas que interaccionan con el promotor se pueden identificar tal y como se describe en el Ejemplo 61 siguiente.

Ejemplo 61

Identificación de proteínas que interaccionan con secuencias de promotor, secuencias reguladoras aguas arriba, o ARNm

Las secuencias dentro de la región del promotor que son susceptibles de unirse a factores de transcripción, se pueden identificar mediante homología con los sitios de unión al factor de transcripción conocidos o mediante mutagénesis convencional o análisis de deleciones de plásmidos informadores que contienen la secuencia del promotor. Por ejemplo, se pueden realizar deleciones en un plásmido informador que contiene la secuencia del promotor de interés ligada funcionalmente con un gen informador ensayable. Los plásmidos informadores portadores de diversas deleciones en la región promotora, se transfectan en una célula hospedadora adecuada y se determinan los efectos de las deleciones sobre los niveles de expresión. Los sitios de unión al factor de transcripción dentro de las regiones en las que las deleciones reducen los niveles de expresión, se pueden localizar adicionalmente empleando la mutagénesis dirigida al sitio, el análisis por rastreo de enlazadores u otras técnicas conocidas por los expertos en la técnica.

Los ácidos nucleicos que codifican proteínas que interaccionan con secuencias en el promotor se pueden identificar empleando sistemas de un híbrido, tal como los descritos en el manual que acompaña al equipo de reactivos de Mat-

chmaker One-Hybrid System, de Clontech (nº de catálogo K1603-1). Resumiendo, el sistema de Matchmaker One-Hybrid se emplea del modo siguiente. La secuencia diana para la que se desea identificar las proteínas que se unen, se clona aguas arriba de un gen informador seleccionable y se integra en el genoma de levadura. Preferentemente, se insertan múltiples copias de las secuencias diana en el plásmido informador, en tándem. Una genoteca que comprende fusiones entre los ADNc en los que se va a evaluar su capacidad de unión al promotor, y el dominio de activación de un factor de transcripción de levadura, tal como GAL4, se transforma en la cepa de levadura que contiene la secuencia informadora integrada. La levadura se extiende en placas sobre medios selectivos para seleccionar las células que expresan el marcador seleccionable ligado a la secuencia del promotor. Las colonias que crecen sobre 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 se pueden insertar en vectores de expresión o en vectores de transcripción *in vitro*. La unión de los polipéptidos codificados por los insertos, con el ADN del promotor se puede confirmar por técnicas conocidas por los expertos en la materia, tales como análisis del desvío en gel o el análisis de la protección frente a ADNasas.

VII. *Uso de las ESTs 5' (o de los ADNc o los ADNs genómicos obtenibles a partir de las mismas) en terapia génica*

La presente invención también comprende el uso de las ESTs 5' (o del ADNc o del ADN genómico obtenible a partir de las mismas) en estrategias de terapia génica, que incluyen estrategias con hebras no codificantes y de triple hélice, tal y como se describe en los Ejemplos 62 y 63 a continuación. En los planteamientos de hebra no codificante, las secuencias de ácido nucleico complementarias a un ARNm, se hibridan con el ARNm intracelularmente, bloqueando de este modo la expresión de la proteína codificada por el ARNm. Las secuencias no codificadoras pueden evitar la expresión génica mediante una variedad de mecanismos. Por ejemplo, las secuencias no codificadoras pueden inhibir la capacidad de los ribosomas de traducir el ARNm. Alternativamente, las secuencias no codificadoras pueden bloquear el transporte del ARNm desde el núcleo hasta el citoplasma, limitando de este modo la cantidad de ARNm disponible para la traducción. Otro mecanismo mediante el cual las secuencias no codificantes pueden inhibir la expresión génica, es interfiriendo con el corte y empalme del ARNm. En aún otra estrategia, el ácido nucleico no codificante se puede incorporar en una ribozima capaz de cortar específicamente el ARNm diana.

Ejemplo 62

Preparación y uso de oligonucleótidos no codificantes

Las moléculas de ácido nucleico no codificante que se van a utilizar en la terapia génica pueden ser secuencias de ADN o de ARN. Pueden comprender una secuencia complementaria a la secuencia de la EST 5' (o del ADNc o del ADN genómico obtenible a partir de la misma). Los ácidos nucleicos no codificantes deben tener una longitud y una temperatura de fusión suficientes para permitir la formación de un dúplex intracelular con una estabilidad suficiente para inhibir la expresión del ARNm en el dúplex. Las estrategias para diseñar ácidos nucleicos no codificantes para uso en la terapia génica, se describen en Green *y col., Ann. Rev. Biochem.* **55**:569-597, 1986; e Izant y Weintraub, *Cell* **36**:1007-1015, 1984.

En algunas estrategias, las moléculas no codificantes se obtienen a partir de una secuencia de nucleótidos que codifica una proteína invirtiendo la orientación de la región codificadora en relación con un promotor, de modo que se transcribe la hebra opuesta de la que se transcribe normalmente en la célula. Las moléculas no codificantes se pueden transcribir empleando sistemas de transcripción *in vitro*, tales como los que emplean la polimerasa T7 o SP6, para generar el transcrito. Otro planteamiento implica la transcripción de los ácidos nucleicos no codificantes *in vivo*, ligando funcionalmente el ADN que contiene la secuencia no codificante con un promotor en un vector de expresión.

Alternativamente, los oligonucleótidos que son complementarios a la hebra transcrita normalmente en la célula, se pueden sintetizar *in vitro*. Por tanto, los ácidos nucleicos no codificantes son complementarios a los ARNm correspondientes y son capaces de hibridarse con el ARNm para crear un dúplex. En algunas realizaciones, las secuencias no codificantes pueden contener esqueletos de azúcar-fosfato modificados para incrementar la estabilidad y hacerlas menos sensibles a la actividad ARNasa. Ejemplos de modificaciones adecuadas para emplear en estrategias no codificantes las describen Rossi *y col., Pharmacol. Ther.* **50**(2):245-254, 1991.

Se pueden utilizar diversos tipos de oligonucleótidos no codificantes complementarios a la secuencia de la EST 5' (o del ADNc o del ADN genómico obtenible a partir de la misma). En una realización preferida, se emplean oligonucleótidos no codificantes, estables y semi-estables descritos en el documento de solicitud de patente internacional nº PCT WO 94/23026. En estas moléculas, el extremo 3' o ambos extremos 3' y 5' se conectan con enlaces de puentes de hidrógeno intramoleculares entre pares de bases complementarias. Estas moléculas están mejor dotadas para rechazar los ataques de exonucleasas y muestran una estabilidad incrementada, comparadas con los oligonucleótidos no codificantes convencionales.

En otra realización preferida, se utilizan oligodesoxinucleótidos no codificantes contra el virus herpes simple de tipo 1 y 2, descritos en el documento de Solicitud de Patente Internacional nº WO 95/04141.

En aún otra realización preferida, se emplean los oligonucleótidos no codificantes, reticulados covalentemente, descritos en el documento de la Solicitud Internacional nº WO 96/31523. Estos oligonucleótidos mono o bicatenarios comprenden uno o varios, respectivamente, reticulamientos covalentes entre oligonucleótidos o intraoligonucleótidos,

en donde la reticulación consiste en un enlace amida entre un grupo amino primario de una cadena y un grupo carboxilo de la otra cadena o de la misma cadena, respectivamente, estando el grupo amino primario directamente sustituido en la posición 2' del anillo de monosacáridos de la hebra de nucleótidos y siendo transportado el grupo carboxilo por un grupo espaciador alifático, sustituido sobre un nucleótido o un análogo de nucleótido de la otra hebra o de la misma hebra, respectivamente.

Los oligodesoxinucleótidos no codificantes y los oligonucleótidos descritos en el documento de Solicitud Internacional nº WO 92/18522, también se pueden utilizar. Estas moléculas son estables frente a la degradación y contienen al menos una secuencia de reconocimiento del control de la transcripción que se une a las proteínas de control y son eficaces como señuelos para las mismas. Estas moléculas pueden contener estructuras de "horquilla", estructuras de "barra con pesas", estructuras de "barra con pesas modificada", estructuras de señuelo "reticulado" y estructuras de "lazo".

En otra realización preferida, se emplean los oligonucleótidos bicatenarios cíclicos descritos en el documento de Solicitud de Patente Europea nº 0572287 A2. Estos oligonucleótidos de "barra con pesas" ligados contienen el sitio de unión para un factor de transcripción e inhiben la expresión del gen bajo control del factor de transcripción mediante secuestro del factor.

El uso de los oligonucleótidos no codificantes muy unidos, descritos en el documento de Solicitud Internacional nº WO 92/19732, también se contempla. Debido a que estas moléculas no tienen extremos libres, son más resistentes a la degradación con exonucleasas que los oligonucleótidos convencionales. Estos oligonucleótidos pueden ser multifuncionales, interaccionar con diversas regiones que no son adyacentes al ARNm diana.

El nivel adecuado de ácidos nucleicos no codificantes requerido para inhibir la expresión génica, se puede determinar empleando un análisis de la expresión *in vitro*. La molécula no codificante se puede introducir en las células por difusión, inyección, infección, transfección o por importación mediada por la región h, empleando procedimientos conocidos en la técnica. Por ejemplo, los ácidos nucleicos no codificantes se pueden introducir en el cuerpo como un oligonucleótido desguarnecido o desnudo, un oligonucleótido encapsulado en un lípido, una secuencia de oligonucleótido encapsidada por una proteína vírica o un oligonucleótido ligado funcionalmente a un promotor contenido en un vector de expresión. El vector de expresión puede ser cualquiera entre una variedad de vectores de expresión conocidos en la técnica, que incluyen vectores retrovíricos o víricos, vectores capaces de una replicación extracromosómica o vectores de integración. Los vectores pueden ser ADN o ARN.

Las moléculas no codificantes se introducen en las muestras celulares con una cantidad diferente de concentraciones, preferentemente entre 1 x 10^{-10} M hasta 1 x 10^{-4} M. Una vez que se ha identificado la concentración mínima que puede controlar adecuadamente la expresión génica, la dosis perfeccionada se traduce a una dosificación adecuada para uso *in vivo*. Por ejemplo, una concentración inhibidora en cultivo de 1 x 10^{-7} se traduce a una dosis de aproximadamente 0,6 mg/kg de peso corporal. Los niveles de oligonucleótido que se aproximan a 100 mg/kg de peso corporal o superiores, pueden ser posibles después de someter a ensayo la toxicidad del oligonucleótido en animales de laboratorio. Se contempla adicionalmente que las células procedentes de vertebrados se retiren, se traten con el oligonucleótido no codificante y se introduzcan de nuevo en el vertebrado.

Se contempla adicionalmente qué secuencia de oligonucleótidos no codificantes se va a incorporar en una secuencia de ribozima para permitir que la cadena no codificante se una específicamente y corte su ARNm diana. Para las aplicaciones técnicas de ribozimas y oligonucleótidos no codificantes, véase Rossi *y col., más arriba*.

En una aplicación preferida de esta invención, el polipéptido codificado por el gen se identifica en primer lugar, de modo que se pueda vigilar la eficacia de la inhibición no codificante sobre la traducción, empleando técnicas que incluyen pero que no están limitadas a, los ensayos mediados con anticuerpos, tales como los RIAs y los ELISAs, los ensayos funcionales o la radiomarcación.

Las ESTs 5' descritas en esta memoria (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) también se pueden emplear en planteamientos de terapia génica basados en la formación de una triple hélice intracelular. Los oligonucleótidos de triple hélice se emplean para inhibir la transcripción de un genoma. Son particularmente útiles para estudiar alteraciones en la actividad celular al estar asociados con un gen particular. Las secuencias de las ESTs 5' (o de los ADNc o de los ADNs genómicos obtenibles a partir de las mismas) descritas en esta memoria o, más preferentemente, una parte de estas secuencias, se puede utilizar para inhibir la expresión génica en individuos que tienen enfermedades asociadas con la expresión de un gen particular. De forma similar, una parte de las secuencias ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se pueden utilizar para estudiar el efecto de inhibir la transcripción de un gen particular dentro de una célula. Tradicionalmente, las secuencias de homopurina se consideraban las más útiles para las estrategias de triple hélice. Sin embargo, las secuencias de homopirimidina también pueden inhibir la expresión génica. Tales oligonucleótidos de homopirimidina se unen a la ranura principal en las secuencias de homopurina:homopirimidina. Por tanto, están contemplados ambos tipos de secuencias procedentes de la EST 5' o del gen correspondiente a la EST 5'.

Ejemplo 63

Preparación y uso de sondas de triple hélice

Las secuencias de las ESTs 5' (o los ADNc o los ADNs genómicos obtenibles a partir de las mismas) se escrutan para identificar 10-meros a 20-meros de segmentos de homopirimidina u homopurina que se podrían utilizar en las estrategias basadas en la triple hélice, para inhibir la expresión génica. Después de identificar los segmentos candidatos de homopirimidina o de homopurina, se determina su eficacia para inhibir la expresión génica introduciendo diversas cantidades de oligonucleótidos que contienen las secuencias candidatas en células de cultivo de tejidos que normalmente expresan el gen diana. Los oligonucleótidos se pueden preparar en un sintetizador de oligonucleótidos o se pueden adquirir comercialmente en una compañía especializada en la síntesis habitual de oligonucleótidos, tal como GENSET, París, Francia.

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

Las células tratadas se vigilan en busca de una función celular alterada o una expresión génica reducida, empleando técnicas tales como la transferencia Northern, los ensayos de protección contra ARNasa o las estrategias basadas en la PCR, para vigilar los niveles de transcripción del gen diana en las células que se han tratado con el oligonucleótido. Las funciones celulares que se van a vigilar se pronostican en base a las homologías del gen diana correspondiente con el ADNc extendido a partir del cual se va a obtener el oligonucleótido con secuencias génicas conocidas que se han asociado con una función particular. Las funciones celulares también se pueden pronosticar basándose en la presencia de fisiologías anormales en las células obtenidas a partir de individuos con una enfermedad hereditaria en particular, particularmente cuando el ADNc extendido está asociado con la enfermedad empleando las técnicas descritas en el Ejemplo 56.

Los oligonucleótidos que son eficaces para inhibir la expresión génica en células de cultivo de tejido se pueden introducir a continuación *in vivo*, empleando las técnicas descritas anteriormente y en el Ejemplo 62 con una dosificación calculada en base a los resultados *in vitro*, tal y como se ha descrito en el Ejemplo 62.

En algunas realizaciones, los anómeros naturales (beta) de las unidades de oligonucleótidos se pueden sustituir por anómeros alfa para volver al oligonucleótido más resistente a las nucleasas. Además, un agente intercalante, tal como bromuro de etidio o similar, se puede fijar al extremo 3' del oligonucleótido alfa para estabilizar la triple hélice. Para una información sobre la generación de oligonucleótidos adecuados para la formación de triple hélice, véase Griffin *y col., Science* **245**:967-971, 1989.

Ejemplo 64

Uso de los ADNc obtenidos empleando las ESTs 5' para expresar una proteína codificada en un organismo hospedador

Los ADNc obtenidos tal y como se ha descrito anteriormente empleando las ESTs 5' descritas en esta memoria, también se pueden utilizar para expresar una proteína codificada en un organismo hospedador para producir un efecto ventajoso. En tales procedimientos, la proteína codificada se puede expresar de forma transitoria en el organismo hospedador o se puede expresar de forma estable en el organismo hospedador. La proteína codificada puede tener cualquiera de las actividades descritas anteriormente. La proteína codificada puede ser una proteína que no posee el organismo hospedador o, alternativamente, la proteína codificada puede aumentar los niveles existentes de la proteína en el organismo hospedador.

Un ADNc extendido de longitud completa que codifica el péptido señal y la proteína madura o un ADNc extendido que codifica sólo la proteína madura, se introduce en el organismo hospedador. El ADNc extendido se puede introducir en el organismo hospedador empleando una variedad de técnicas conocidas por los expertos en la materia. Por ejemplo, el ADNc extendido se puede inyectar en el organismo hospedador en forma de ADN desnudo de modo que la proteína codificada se exprese en el organismo hospedador produciendo de este modo un efecto beneficioso.

Alternativamente, el ADNc extendido se puede clonar en un vector de expresión aguas 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 la terapia génica, incluyendo vectores vírico o retrovíricos. El vector de expresión se puede introducir directamente en el organismo hospedador, de modo que la proteína codificada se exprese en el organismo hospedador para producir un efecto ventajoso. En otro planteamiento, el vector de expresión se puede introducir en células *in vitro*. Las células que contienen el vector de expresión se seleccionan después y se introducen en el organismo hospedador, en donde expresan la proteína codificada para producir un efecto beneficioso.

Ejemplo 65

Uso de péptidos señal codificados por las ESTs 5' o por secuencias obtenidas a partir de las mismas para importar las proteína en las células

La región hidrófoba corta del núcleo (h) de los péptidos señal codificados por las ESTs 5' o los ADNc extendidos obtenidos a partir de la SEQ ID NO: 38, también se puede utilizar como vehículo para importar un péptido o una proteína de interés, denominada carga, a las células de cultivo de tejidos (Lin *y col., J. Biol. Chem.*, **270**:14225-14258, 1995; Du *y col., J. Peptide Res.*, **51**:235-243, 1998; Rojas *y col., Nature Biotech.*, **16**:370-375, 1998).

Cuando péptidos de tamaño limitado (aproximadamente hasta 25 aminoácidos) son permeables a las células, se transportan a través de la membrana celular, se puede utilizar la síntesis química para añadir la región h al extremo C-terminal o N-terminal al péptido de carga de interés. Alternativamente, cuando se van a importar en las células péptidos o proteínas más largos, los ácidos nucleicos se pueden modificar por ingeniería genética empleando técnicas conocidas por los expertos en la técnica, para enlazar 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. Tales ácidos nucleicos modificados por ingeniería genética se traducen a continuación después de la transfección, *in vitro* o *in vivo*, en células adecuadas, empleando técnicas convencionales para producir el polipéptido resultante permeable a la célula. Las células hospedadoras adecuadas se incuban a continuación de forma aislada con el polipéptido permeable a la célula que a continuación se transporta a través de la membrana.

Este método se puede aplicar para estudiar diversas funciones intracelulares y procesos celulares. Por ejemplo, se ha empleado para rastrear dominios funcionalmente importantes de las proteínas intracelulares y para examinar las interacciones proteína-proteína implicadas en las rutas de transducción de señales (Lin *y col., supra*; Lin *y col., J. Biol. Chem*., **271**:5305-5308, 1996; Rojas *y col., J. Biol. Chem.*, **271**:27456-27461, 1996; Liu *y col., Proc. Natl. Acad. Sci. USA*, **93**:11819-11824, 1996; Rojas *y col., Bioch. Biophys. Res. Commun.* **234**:675-680, 1997).

Tales técnicas se pueden utilizar en la terapia celular para importar proteínas que producen efectos terapéuticos. Por ejemplo, las células aisladas a partir de un paciente, se pueden tratar con proteínas terapéuticas importadas y a continuación introducirlas de nuevo en el organismo hospedador.

Alternativamente, la región h de los péptidos señal de la presente invención se podría utilizar junto con una señal de localización nuclear para entregar ácidos nucleicos a los núcleos celulares. Tales oligonucleótidos pueden ser oligonucleótidos no codificantes u oligonucleótidos diseñados para formar triple hélices, tal y como se describe en los Ejemplos 62 y 63, respectivamente, para inhibir el procesamiento y/o la maduración de un ARN celular diana.

Tal y como se ha descrito anteriormente, los ADNc o partes de los mismos obtenidos empleando las ESTs 5' descritas en esta memoria, se pueden utilizar para diversos fines. Los polinucleótidos se pueden utilizar para expresar proteínas recombinantes para el análisis, la caracterización o el uso terapéutico; como marcadores para tejidos en los que la proteína correspondiente se expresa preferentemente (tanto de forma constitutiva como en una fase particular de la diferenciación o el desarrollo tisular o en estados de enfermedad); como marcadores del peso molecular sobre geles Southern; como marcadores o etiquetas cromosómicas (cuando se marcan) para identificar cromosomas o para cartografiar las posiciones de genes relacionados; para comparar con secuencias de ADN endógeno en pacientes para identificar enfermedades genéticas potenciales; como sondas para hibridar y, por tanto, descubrir nuevas secuencias de ADN relacionadas; como una fuente de información para obtener cebadores de PCR para la huella genética; para seleccionar y preparar oligómeros para fijarlos a un "circuito integrado génico" u otro soporte, incluyendo el examen de patrones de expresión; para obtener anticuerpos anti-proteína empleando técnicas de inmunización con ADN; y como un antígeno para obtener anticuerpos anti-ADN o para producir otra respuesta inmune. Cuando el polinucleótido codifica una proteína que se une o que se une potencialmente a otra proteína (tal como por ejemplo, en una interacción receptor-ligando), el polinucleótido se puede utilizar también en ensayos de trampa e interacción (tales como por ejemplo, los descritos en Gyuris *y col., Cell* **75**:791-803, 1993); para identificar polinucleótidos que codifican la otra proteína con la que tiene lugar la unión o para identificar inhibidores de la interacción de la unión.

Las proteínas o los polipéptidos proporcionados por la presente invención se pueden utilizar de forma similar en ensayos para determinar la actividad biológica, que incluye un panel de proteínas múltiples para realizar un escrutinio de alto rendimiento, para obtener anticuerpos o para conseguir otra respuesta inmune; como reactivo (incluyendo el reactivo marcado) en ensayos diseñados para determinar cuantitativamente niveles de la proteína (o su receptor) en fluidos biológicos; como marcadores para tejidos en los que la proteína correspondiente se expresa preferentemente (de forma constitutiva o en una fase particular de la diferenciación o el desarrollo tisular o en un estado de enfermedad); 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 receptor-ligando), la proteína se puede utilizar para identificar la otra proteína con la que tiene lugar la unión o para identificar inhibidores de la interacción de la unión. Las proteínas implicadas en estas interacciones de la unión, también se pueden utilizar para escrutar inhibidores peptídicos o de molécula pequeña o agonistas de la interacción de la unión.

Cualquiera de estos usos en investigación se puede desarrollar en grado reactivo o en formato de equipo de reactivos para la comercialización como productos para investigación.

Los métodos para llevar a cabo los usos enumerados arriba, son bien conocidos por los expertos en la técnica. Las referencias que describen tales métodos incluyen sin limitación *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Laboratory Press, Sambrook, Fritsch y Maniatis compiladores, 1989, y *Methods in Enzymology: Guide to Molecular Cloning Techniques*, Academic Press, Berger y Kimmel compiladores, 1987.

Los polinucleótidos y las proteínas de la presente invención también se pueden utilizar como fuentes o suplementos de la nutrición. Tales usos incluyen sin limitaciones el uso como suplemento de proteínas o de aminoácidos, el uso como fuente de carbono, el uso como fuente de nitrógeno y el uso como fuente de carbohidratos. En tales casos, la proteína o el polinucleótido de la invención se puede añadir a la alimentación de un organismo particular o se puede administrar como una preparación sólida o líquida separada, tal y como en forma de polvos, píldoras, soluciones, suspensiones o cápsulas. En el caso de microorganismos, la proteína o el polinucleótido de la invención se puede añadir al medio en el que se cultiva el microorganismo o sobre el que se cultiva.

Aunque esta invención se ha descrito en función de ciertas realizaciones preferidas, otras realizaciones serán evidentes para los expertos en la técnica, de cara a la descripción de esta memoria y también dentro del alcance de esta invención. Por tanto, el alcance de la invención se pretende definir sólo haciendo referencia a las reivindicaciones anejas.

(Tabla pasa a página siguiente)

	Características de la búsqueda		Características de la selección		
Etapa	**Programa**	**Cadena**	**Parámetros**	**Identidad (%)**	**Longitud (pb)**
Variada	blastn	ambas	S = 61 X = 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 = 54 X = 16	90	15†
Vertebrado	fasta*	ambas	S = 108	90	30
ESTs	blastn	ambas	S = 108 X = 16	90	30
Proteínas	blastx□	superior	E = 0,001	-	-

Tabla 1: Parámetros empleados para cada etapa del análisis de las ESTs

* uso del escáner de la base de datos "Quick Fast"
† alineación adicionalmente restringida para que comience a menos de 10 pb del extremo de la EST 5'
□ empleo de la matriz de sustitución BLOSUM62

ES 2 266 693 T3

TABLA II

SEQ ID NO	Categoría	Puntuación de von heijne	Fuente del Tejido	Denominación Internacional
ID38	nueva	6,9	Ganglios linfáticos	48-5-4-B6-PU

TABLA III

SEQ ID NO	Péptido señal
ID38	MTMRHNWTPDLSPLWVLLLCAHVVTL

TABLA IV

Puntuación mínima del péptido señal	Tasa de falsos positivos	Tasa 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	0,853	0,919

TABLA V

Puntuación mínima del péptido señal	Todas las ESTs	ESTs nuevas	ESTs que se emparejan con ESTs públicas a menos de 40 pb del inicio	ESTs que extienden ARNm conocidos más de 40 pb	ESTs que extienden ESTs públicas en más de 40 pb
3,5	2874	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

TABLA VI

Tejido	Todas las ESTs	ESTs nuevas	ESTs que se emparejan con ESTs públicas a menos de 40 pb del inicio	ESTs que extienden ARNm conocidos más de 40 pb	ESTs que extienden públicas más de 40 pb
Cerebro	329	131	75	3	24
Próstata cancerosa	134	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
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

TABLA VII

Descripción de los sitios de unión del factor de transcripción presentes en promotores aislados a partir de secuencias de etiquetas señales

Secuencia del promotor P13H2 (646 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	-339	+	0,953	11	AGATAGGACAT
TAL1ALPHAE47_01	-235	+	0,973	16	CATAACAGATGGTAAG
TAL1BETAE47_01	-235	+	0,983	16	CATAACAGATGGTAAG
TAL1BETAITF2_01	-235	+	0,978	16	CATAACAGATGGTAAG
MYOD_Q6	-232	-	0,954	10	ACCATCTGTT
GATA1_04	-217	-	0,953	13	TCAAGATAAAGTA
IK1_01	-126	+	0,963	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	TAAAACAAAACA
E2F_02	-33	+	0,957	8	TTTAGCGC
MZF1_01	-5	-	0,975	8	TGAGGGGA

Secuencia del promotor 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	+	0,965	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	TCCCACCTTCC
S8_01	5	-	0,992	11	GAGGCAATTAT
MZF1_01	16	-	0,986	8	AGAGGGGA

Secuencia del promotor 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	-	0,958	12	CAGCACGTGAGT
MYCMAX_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

REIVINDICACIONES

1. Un péptido señal que tiene la secuencia de los aminoácidos -26 a -1 de SEQ ID NO: 39.

2. Una secuencia señal aislada que consiste en el ácido nucleico que codifica el péptido señal de la reivindicación 1.

3. La secuencia señal de la reivindicación 2, que consiste en la secuencia de los nucleótidos 170 a 247 de SEQ ID NO: 38.

4. Un ácido nucleico purificado y aislado que codifica un polipéptido que comprende el péptido señal de la reivindicación 1 en el extremo 5' de la secuencia codificadora y un extremo amino que comprende los aminoácidos 1 a 29 de SEQ ID NO: 39.

5. El ácido nucleico de la reivindicación 4, en donde dicho ácido nucleico comprende una secuencia señal que tiene la secuencia de los nucleótidos 170 a 247 de SEQ ID NO: 38.

6. El ácido nucleico de la reivindicación 5, en donde el extremo amino de dicho polipéptido está codificado por los nucleótidos 170 a 334 de SEQ ID NO: 38.

7. Un ácido nucleico purificado y aislado que comprende la secuencia señal de la reivindicación 2, en donde dicha secuencia señal se fusiona en marco de lectura con el extremo 5' de una secuencia que codifica una parte del polipéptido que comprende al menos 25 aminoácidos que es heterólogo de un polipéptido que comprende los aminoácidos 1 a 29 de SEQ ID NO: 39.

8. Un vector de expresión que comprende la secuencia señal de la reivindicación 2 o 3, ligada funcionalmente a un promotor.

9. Un vector de expresión que comprende el ácido nucleico de uno cualquiera de los ácidos nucleicos de las reivindicaciones 4 a 7, ligado funcionalmente a un promotor.

10. El vector de expresión de la reivindicación 8 o 9, en donde dicho vector es un vector de secreción.

11. El vector de expresión de la reivindicación 8 o 9, en donde dicho vector es un vector de terapia génica.

12. Un polipéptido codificado por el polinucleótido de una cualquiera de las reivindicaciones 4 a 7.

13. El polipéptido de la reivindicación 12, en donde dicho polipéptido es una proteína humana secretada.

14. El polipéptido de la reivindicación 13, en donde dicho polipéptido comprende los aminoácidos 1 a 29 de SEQ ID NO: 39.

15. Una proteína de fusión codificada por el ácido nucleico de la reivindicación 7.

16. Uso del péptido señal de la reivindicación 1, para dirigir la secreción extracelular de un polipéptido.

17. Uso del péptido señal de la reivindicación 1, para simplificar la purificación de una proteína de un polipéptido deseado.

18. Uso del vector de una cualquiera de las reivindicaciones 8 a 11, para dirigir la secreción extracelular de un polipéptido.

19. Uso del vector de una cualquiera de las reivindicaciones 8 a 10, para simplificar la purificación de una proteína de un polipéptido deseado.

20. Uso de una cualquiera de las reivindicaciones 16 a 19, en donde dicho polipéptido es el polipéptido de una cualquiera de las reivindicaciones 12 a 15.

21. Un método para preparar una proteína secretada que comprende la etapa de insertar un gen que codifica una proteína no secretada, en marco de lectura con la secuencia señal de la reivindicación 2 o 3, en un vector de modo que la proteína codificada por el gen insertado se expresa a partir del ARNm transcrito.

22. El método de la reivindicación 21, que comprende adicionalmente la etapa de introducir dicho vector en una célula, un tejido o un organismo hospedador.

23. Un método para preparar una proteína secretada que comprende la etapa de introducir el vector de la reivindicación 9, en una célula, un tejido o un organismo hospedador.

24. El método de una cualquiera de las reivindicaciones 21 a 23, que comprende adicionalmente la etapa de aislar la proteína secretada.

25. El método de la reivindicación 24, en el que la etapa de aislar la proteína secretada comprende purificar la proteína secretada a partir del material sobrenadante, del medio de cultivo o del extracto celular de dicha célula hospedadora.

26. Un polipéptido aislado obtenible por uno cualquiera de los métodos de las reivindicaciones 21 a 25.

27. Un método para preparar un ADNc que codifica una proteína humana secretada, en donde dicho método comprende las etapas de:

a) poner en contacto una colección de moléculas de ARNm procedentes de células humanas, con un cebador que comprende al menos 15 nucleótidos consecutivos de una secuencia complementaria a una secuencia de SEQ ID NO: 38;

b) hibridar bajo condiciones rigurosas dicho cebador con una secuencia de ARNm complementario en dicha colección;

c) transcribir de forma inversa dicho cebador hibridado para formar una primera hebra de ADNc a partir de dicho ARNm;

d) preparar una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc; y

e) aislar el ADNc resultante que codifica dicho polipéptido que comprende dicha primera hebra de ADNc y dicha segunda hebra de ADNc.

28. Un método para preparar un ADNc que codifica una proteína humana secretada, en donde dicho método comprende las etapas de:

a) proporcionar un ADNc obtenido a partir de un ARNm de un tejido, una célula o un organismo de interés;

b) poner en contacto dicho ADNc con una sonda detectable que comprende al menos 15 nucleótidos consecutivos de dicha secuencia de SEQ ID NO: 38 o una secuencia complementaria a la misma, bajo condiciones que permitan que dicha sonda se hibride con dicho ADNc;

c) identificar un ADNc que se hibride bajo condiciones rigurosas con dicha sonda detectable; y

d) aislar dicho ADNc que se hibrida con dicha sonda.

29. Un método para preparar un ADNc que comprende al menos 15 nucleótidos consecutivos de una secuencia SEQ ID NO: 38 que codifica una proteína humana secretada, en donde dicho método comprende las etapas de:

a) poner en contacto una colección de moléculas de ARNm procedentes de células humanas con un primer cebador capaz de hibridarse con la cola poliA de dicho ARNm;

b) hibridar dicho primer cebador con dicha cola poliA;

c) transcribir de forma inversa dicho ARNm para formar una primera hebra de ADNc;

d) preparar una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc empleando al menos un cebador que comprende al menos 15 nucleótidos consecutivos de una secuencia SEQ ID NO: 38; y

e) aislar el ADNc resultante que comprende dicha primera hebra de ADNc y dicha segunda hebra de ADNc.

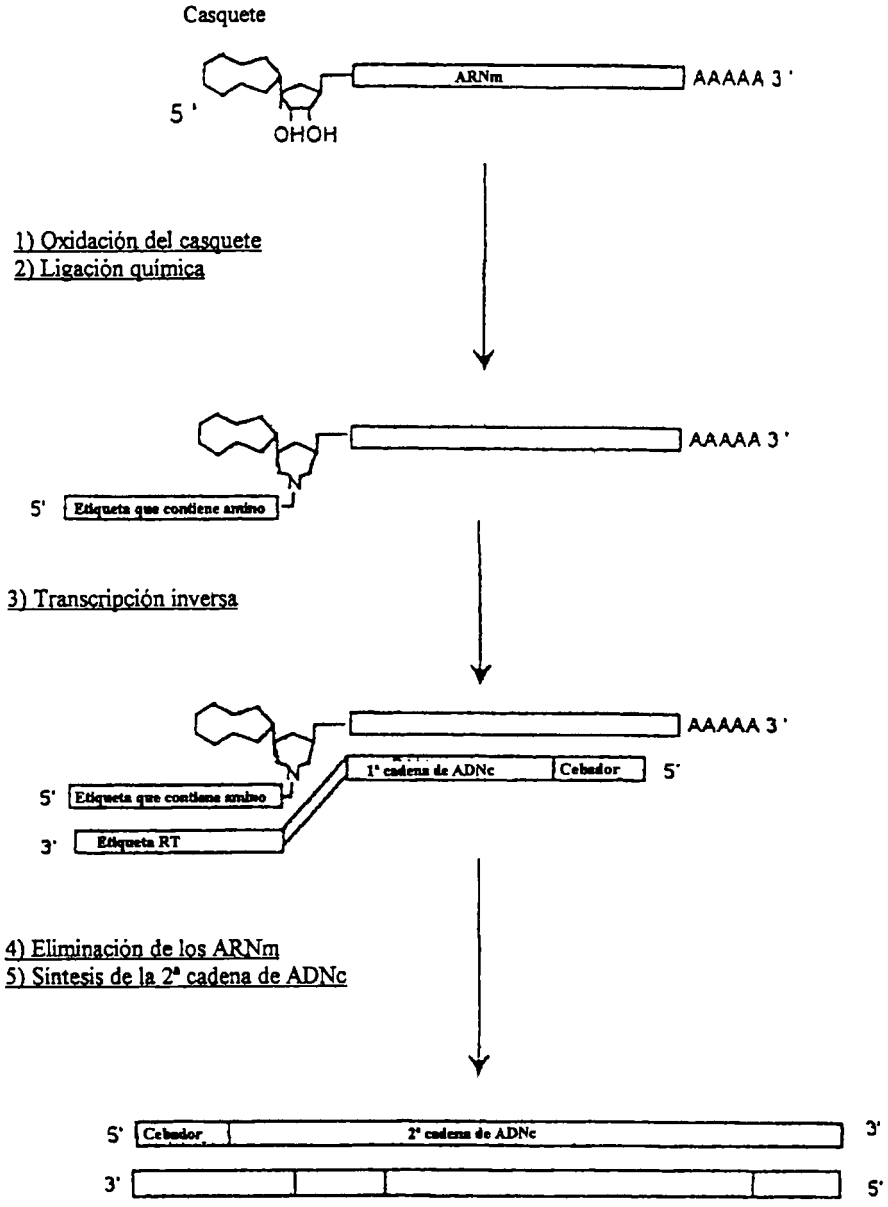

Casquete
ARNm
AAAAA 3'
5'
OHOH
1) Oxidación del casquete
2) Ligación química
AAAAA 3'
5'
Etiqueta que contiene amino
3) Transcripción inversa
AAAAA 3'
1ª cadena de ADNc
Cebador
5'
5'
Etiqueta que contiene amino
3'
Etiqueta RT
4) Eliminación de los ARNm
5) Síntesis de la 2ª cadena de ADNc
5'
Cebador
2ª cadena de ADNc
3'
3'
5'

Figura 1

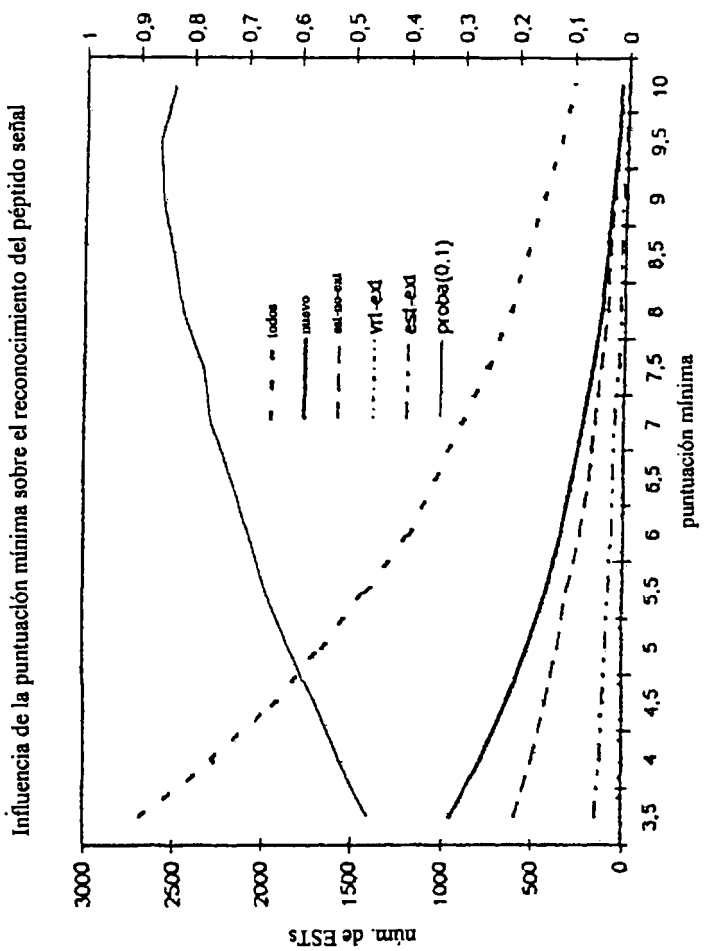

Influencia de la puntuación mínima sobre el reconocimiento del péptido señal
núm. de ESTs
puntuación mínima
todos
nuevo
sal-no-sal
vrl-exl
esl-exl
proba(0.1)
3000
2500
2000
1500
1000
500
0
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
3,5
4
4,5
5
5,5
6
6,5
7
7,5
8
8,5
9
9,5
10

Figura 2

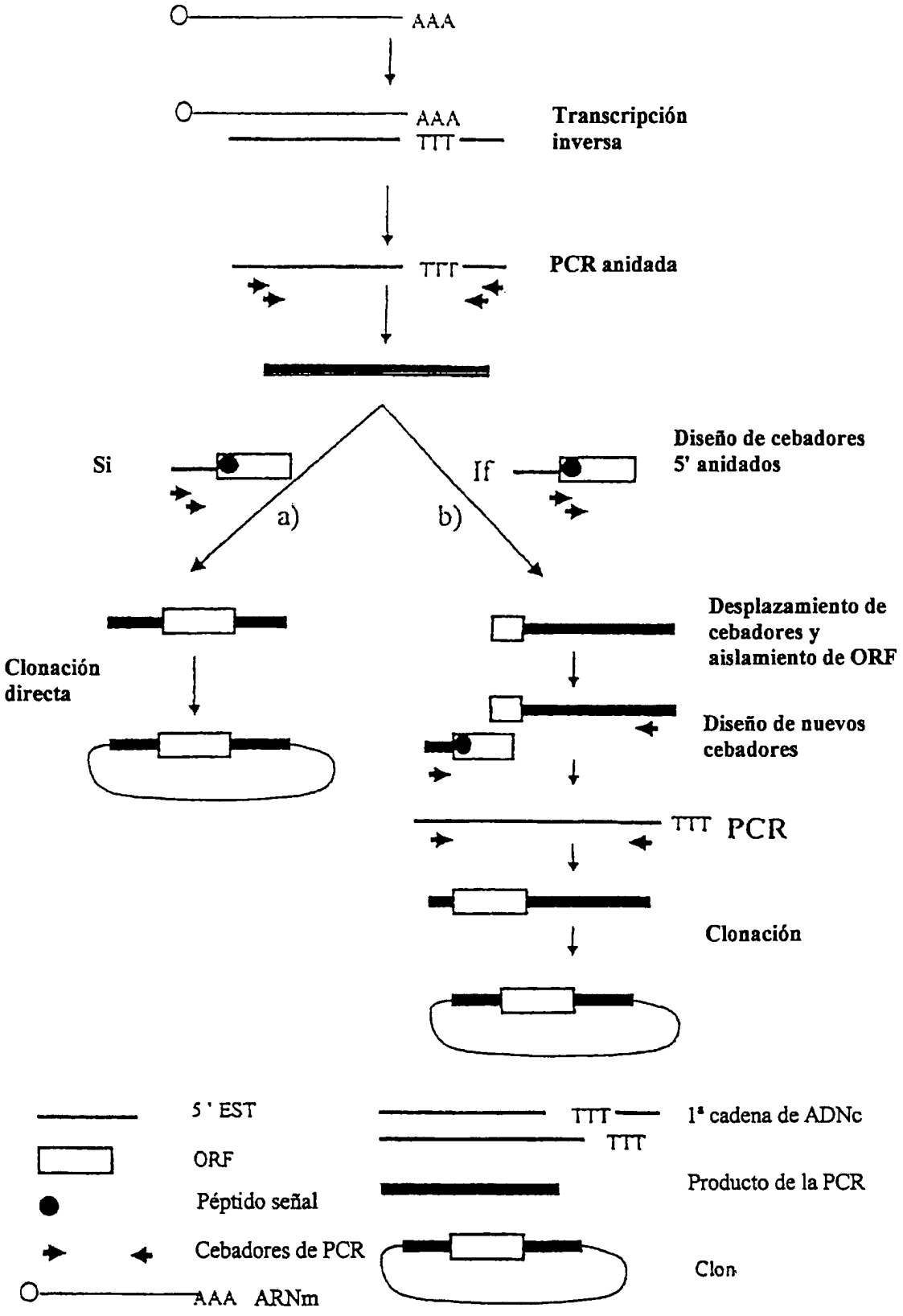

AAA
AAA
TTT
Transcripción inversa
TTT
PCR anidada
Si
If
a)
b)
Diseño de cebadores 5' anidados
Clonación directa
Desplazamiento de cebadores y aislamiento de ORF
Diseño de nuevos cebadores
TTT PCR
Clonación
5 ' EST
ORF
Péptido señal
Cebadores de PCR
AAA ARNm
TTT
TTT
1ª cadena de ADNc
Producto de la PCR
Clon

Figura 3

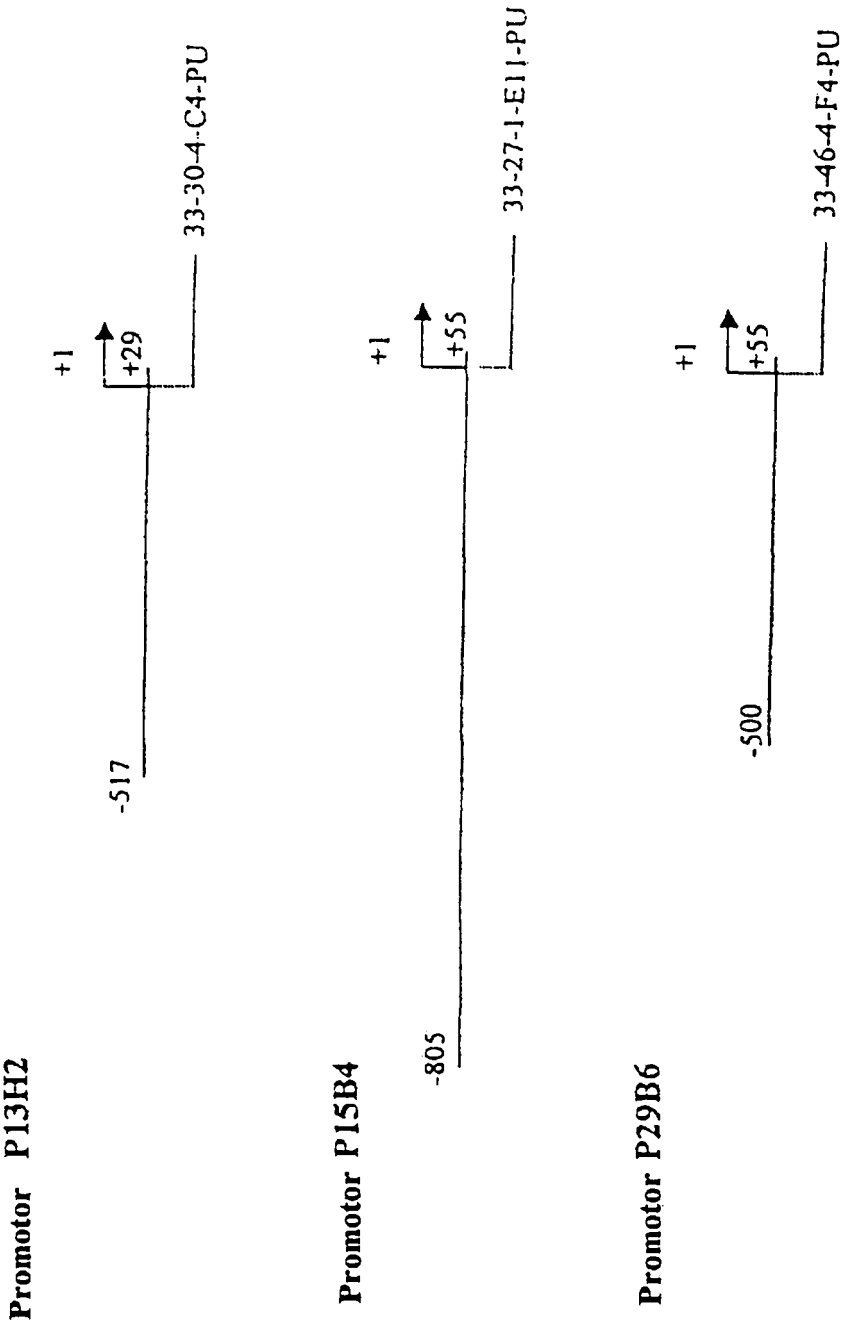

Promotor P13H2
+1
+29
33-30-4-C4-PU
-517
Promotor P15B4
+1
+55
33-27-1-E11-PU
-805
Promotor P29B6
+1
+55
33-46-4-F4-PU
-500

Figura 4

LISTA DE SECUENCIAS

(1) INFORMACIÓN GENERAL:

(A) NOMBRE: GENSET SA

(B) CALLE: 24, RUE ROYALE

(C) CIUDAD: PARÍS

(E) PAÍS: FRANCIA

(F) CÓDIGO POSTAL (ZIP): 75008

(ii) TITULO DE LA INVENCIÓN: ESTs 5' PARA PROTEÍNAS SECRETADAS EXPRESADAS EN DIVERSOS TEJIDOS

(iii) NÚMERO DE SECUENCIAS: 333

(v) FORMA LEGIBLE POR ORDENADOR:

(A) TIPO DE MEDIO: Disquete

(B) ORDENADOR: Compatible con PC IBM

(C) SISTEMA OPERATIVO: Win95

(D) SOPORTE LÓGICO: Word

(2) INFORMACIÓN DE LA SEQ ID NO: 1:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 47 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: Cap

(B) LOCALIZACIÓN: 1

(D) OTRA INFORMACIÓN: m7Gppp añadido a 1

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 1:

GGCAUCCUAC UCCCAUCCAA UUCCACCCUA ACUCCUCCCA UCUCCAC 47

(2) INFORMACIÓN DE LA SEQ ID NO: 2:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 46 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 2:

GCAUCCUACU CCCAUCCAAU UCCACCCUAA CUCCUCCCAU CUCCAC 46

(2) INFORMACIÓN DE LA SEQ ID NO: 3:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 3:

ATCAAGAATT CGCACGAGAC CATTA 25

(2) INFORMACIÓN DE LA SEQ ID NO: 4:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 4:

TAATGGTCTC GTGCGAATTC TTGAT 25

(2) INFORMACIÓN DE LA SEQ ID NO: 5:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 5:

CCGACAAGAC CAACGTCAAG GCCGC 25

(2) INFORMACIÓN DE LA SEQ ID NO: 6:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 6:

TCACCAGCAG GCAGTGGCTT AGGAG 25

(2) INFORMACIÓN DE LA SEQ ID NO: 7:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 7:

AGTGATTCCT GCTACTTTGG ATGGC 25

(2) INFORMACIÓN DE LA SEQ ID NO: 8:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 8:

GCTTGGTCTT GTTCTGGAGT TTAGA 25

(2) INFORMACIÓN DE LA SEQ ID NO: 9:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 9:

TCCAGAATGG GAGACAAGCC AATTT 25

(2) INFORMACIÓN DE LA SEQ ID NO: 10:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 10:

AGGGAGGAGG AAACAGCGTG AGTCC 25

(2) INFORMACIÓN DE LA SEQ ID NO: 11:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 11:

ATGGGAAAGG AAAAGACTCA TATCA 25

(2) INFORMACIÓN DE LA SEQ ID NO: 12:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 12:

AGCAGCAACA ATCAGGACAG CACAG 25

(2) INFORMACIÓN DE LA SEQ ID NO: 13:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 13:

ATCAAGAATT CGCACGAGAC CATTA 25

(2) INFORMACIÓN DE LA SEQ ID NO: 14:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 67 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 14:

ATCGTTGAGA CTCGTACCAG CAGAGTCACG AGAGAGACTA CACGGTACTG GTTTTTTTTT 60
TTTTTVN 67

(2) INFORMACIÓN DE LA SEQ ID NO: 15:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 29 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 15:

CCAGCAGAGT CACGAGAGAG ACTACACGG 29

(2) INFORMACIÓN DE LA SEQ ID NO: 16:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 16:

CACGAGAGAG ACTACACGGT ACTGG 25

(2) INFORMACIÓN DE LA SEQ ID NO: 17:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 526 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(F) TIPO DE TEJIDO: Ganglios linfáticos

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: complemento (261..376)

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 96
región 166..281
id N70479
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: complemento (380..486)

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 97
región 54..160
id N70479
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: complemento (110..145)

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 94
región 403..438
id N70479
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: complemento (196..229)

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 94
región 315..348
id N70479
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 90..140

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 8,2
seq LLLITAILAVAVG/FP

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 17:

```
AATATRARAC AGCTACAATA TTCCAGGGCC ARTCACTTGC CATTTCTCAT AACAGCGTCA      60

GAGAGAAAGA ACTGACTGAR ACGTTTGAG 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 Gly
        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
```

(2) INFORMACIÓN DE LA SEQ ID NO: 18:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 17 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 1..17

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 8,2
seq LLLITAILAVAVG/FP

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 18:

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

Gly

(2) INFORMACIÓN DE LA SEQ ID NO: 19:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 822 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(D) ESTADO DE DESARROLLO: Fetal

(F) TIPO DE TEJIDO: Riñón

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 260..464

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 96
región 153..357
id H57434
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 118..184

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 98
región 98..164
id H57434
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 56..113

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 98
región 35..92
id H57434
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 454..485

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 100
región 348..379
id H57434
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 118..545

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 98
región 1..428
id N27248
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 65..369

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 98
región 41..345
id H94779
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 61..399

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 99
región 6..344
id H09880
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 408..458

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 92
región 355..405
id H09880
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 60..399

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 97
región 56..395
id H29351
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 393..432

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 90
región 391..430
id H29351
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 346..408

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 5,5
seq SFLPSALVIWTSA/AF

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 19:

```
ACTCCTTTTA GCATAGGGGC TTCGGCGCCA GCGGCCAGCG CTAGTCGGTC TGGTAAGTGC      60

CTGATGCCGA GTTCCGTCTC TCGCGTCTTT TCCTGGTCCC AGGCAAAGCG GASGNAGATC     120

CTCAAACGGC CTAGTGCTTC GCGCTTCCGG AGAAAATCAG CGGTCTAATT AATTCCTCTG     180

GTTTGTTGAA GCAGTTACCA AGAATCTTCA ACCCTTTCCC ACAAAAGCTA ATTGAGTACA     240

CGTTCCTGTT GAGTACACGT TCCTGTTGAT TTACAAAAGG TGCAGGTATG AGCAGGTCTG     300

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

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

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 CATGACCAAA            602
Lys

CTCTTCARAA ACATGTCTTT ACAAGCATAT CTCTTGTATT GCTTTCTACA CTGTTGAATT     662

GTCTGGCAAT ATTTCTGCAG TGGAAAATTT GATTTARMTA GTTCTTGACT GATAAATATG     722

GTAAGGTGGG CTTTTCCCCC TGTGTAATTG GCTACTATGT CTTACTGAGC CAAGTTGTAW     782

TTTGAAATAA AATGATATGA GAGTGACACA AAAAAAAAAA                           822
```

(2) INFORMACIÓN DE LA SEQ ID NO: 20:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 21 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 1..21

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 5,5
seq SFLPSALVIWTSA/AF

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 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	

(2) INFORMACIÓN DE LA SEQ ID NO: 21:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 405 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(F) TIPO DE TEJIDO: Testículos

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: complemento (103..398)

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 96
región 1..296
id AA442893
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 185..295

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 5,9
seq LSYASSALSPCLT/AP

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 21:

```
ATCACCTTCT TCTCCATCCT TSTCTGGGCC AGTCCCCARC CCAGTCCCTC TCCTGACCTG      60

CCCAGCCCAA GTCAGCCTTC AGCACGCGCT TTTCTGCACA CAGATATTCC AGGCCTACCT     120

GGCATTCCAG GACCTCCGMA ATGATGCTCC AGTCCCTTAC AAGCGCTTCC 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 CGG 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 TAAATATCCT TATCCAAATC AATAAARWRA RAATCCTCCC TCCARAAGGG   384
Pro Asp Asn

TTTCTAAAAA CAAAAAAAAA A                                             405
```

(2) INFORMACIÓN DE LA SEQ ID NO: 22:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 37 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 1..37

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 5,9
seq LSYASSALSPCLT/AP

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 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
```

(2) INFORMACIÓN DE LA SEQ ID NO: 23:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 496 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(F) TIPO DE TEJIDO: Próstata cancerosa

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 149..331

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 98
región 1..183
id AA397994
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 328..495

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 96
región 179..336
id AA397994
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: complemento (182..496)

(C) MÉTODO DE IDENTIFICACIÓN: blastn

(D) OTRA INFORMACIÓN: identidad 97
región 14..328
id AA399680
est

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 196..240

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 5,5
seq ILSTVTALTFAXA/LD

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 23:

```
AAAAAATTGG TCCCAGTTTT CACCCTGCCG CAGGGCTGGC TGGGGAGGGC AGCGGTTTAG      60

ATTAGCCGTG GCCTAGGCCG TTTAACGGGG TGACACGAGC NTGCAGGGCC GAGTCCAAGG     120

CCCGGAGATA GGACCAACCG TCAGGAATGC GAGGAATGTT 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
                 -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 AAA 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

TCT TCA GCC TGAAATGAAK CCGGGATCAA ATGGTTGCTG ATCARAGCCC ATATTTAAAT    434
Ser Ser Ala

TGGAAAAGTC AAATTGASCA TTATTAAATA AAGCTTGTTT AATATGTCTC AAACAAAAAA     494

AA                                                                    496
```

(2) INFORMACIÓN DE LA SEQ ID NO: 24:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 15 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 1..15

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 5,5
seq ILSTVTALTFAXA/LD

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 24:

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

(2) INFORMACIÓN DE LA SEQ ID NO: 25:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 623 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(F) TIPO DE TEJIDO: Testículo

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 49..96

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 10,1
seq LVLTLCTLPLAVA/SA

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 25:

```
AAAGATCCCT GCAGCCCGGC AGGAGAGAAG GCTGAGCCTT CTGGCGTC ATG GAG AGG      57
                                                      Met Glu Arg
                                                          -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

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

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 TCG 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

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

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

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 AAAAAAAAA                                      623
```

(2) INFORMACIÓN DE LA SEQ ID NO: 26:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 16 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 1..16

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 10,1
seq LVLTLCTLPLAVA/SA

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 26:

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

(2) INFORMACIÓN DE LA SEQ ID NO: 27:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 848 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(D) ESTADO DE DESARROLLO: Fetal

(F) TIPO DE TEJIDO: Riñón

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 32..73

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 10,7
seq LWLLFFLVTAIHA/EL

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 27:

```
AACTTTGCCT TGTGTTTTCC ACCCTGAAAG A ATG TTG TGG CTG CTC TTT TTT CTG     55
                                   Met Leu Trp Leu Leu Phe Phe Leu
                                                   -10

GTG ACT GCC ATT CAT GCT GAA CTC TGT CAA CCA GGT GCA GAA AAT GCT      103
Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn Ala
    -5                  1               5                  10

TTT AAA GTG AGA CTT AGT ATC AGA ACA GCT CTG GGA GAT AAA GCA TAT      151
Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala Tyr
                15                  20                  25

GCC TGG GAT ACC AAT GAA GAA TAC CTC TTC AAA GCG ATG GTA GCT TTC      199
Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala Phe
            30                  35                  40

TCC ATG AGA AAA GTT CCC AAC AGA GAA GCA ACA GAA ATT TCC CAT GTC      247
Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His Val
        45                  50                  55

CTA CTT TGC AAT GTA ACC CAG AGG GTA TCA TTC TGG TTT GTG GTT ACA      295
Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val Thr
    60                  65                  70

GAC CCT TCA AAA AAT CAC ACC CTT CCT GCT GTT GAG GTG CAA TCA GCC      343
Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser Ala
75                  80                  85                  90

ATA AGA ATG AAC AAG AAC CGG ATC AAC AAT GCC TTC TTT CTA AAT GAC      391
Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn Asp
                95                  100                 105

CAA ACT CTG GAA TTT TTA AAA ATC CCT TCC ACA CTT GCA CCA CCC ATG      439
Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro Met
            110                 115                 120

GAC CCA TCT GTG CCC ATC TGG ATT ATT ATA TTT GGT GTG ATA TTT TGC      487
Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe Cys
        125                 130                 135

ATC ATC ATA GTT GCA ATT GCA CTA CTG ATT TTA TCA GGG ATC TGG CAA      535
Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp Gln
    140                 145                 150

CGT ADA ARA AAG AAC AAA GAA CCA TCT GAA GTG GAT GAC GCT GAA RAT      583
Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu Xaa
155                 160                 165                 170

AAK TGT GAA AAC ATG ATC ACA ATT GAA AAT GGC ATC CCC TCT GAT CCC      631
Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp Pro
                175                 180                 185

CTG GAC ATG AAG GGA GGG CAT ATT AAT GAT GCC TTC ATG ACA GAG GAT      679
Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu Asp
            190                 195                 200

GAG AGG CTC ACC CCT CTC TGAAGGGCTG TTGTTCTGCT TCCTCAARAA              727
Glu Arg Leu Thr Pro Leu
        205

ATTAAACATT TGTTTCTGTG TGACTGCTGA GCATCCTGAA ATACCAAGAG CAGATCATAT    787

WTTTTGTTTC ACCATTCTTC TTTTGTAATA AATTTTGAAT GTGCTTGAAA AAAAAAAAAA    847

C                                                                    848
```

(2) INFORMACIÓN DE LA SEQ ID NO: 28:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 14 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 1..14

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 10,7
seq LWLLFFLVTAIHA/EL

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 28:

Met	Leu	Trp	Leu	Leu	Phe	Phe	Leu	Val	Thr	Ala	Ile	His	Ala
1				5					10				

(2) INFORMACIÓN DE LA SEQ ID NO: 29:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 29:

GGGAAGATGG AGATAGTATT GCCTG 25

(2) INFORMACIÓN DE LA SEQ ID NO: 30:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 30:

CTGCCATGTA CATGATAGAG AGATTC 26

(2) INFORMACIÓN DE LA SEQ ID NO: 31:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 546 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN genómico

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: promotor

(B) LOCALIZACIÓN: 1..517

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de inicio de la transcripción

(B) LOCALIZACIÓN: 518

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 17..25

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre CMYB_01
puntuación 0,983
secuencia TGTCAGTTG

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (18..27)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre MYOD_Q6
puntuación 0,961
secuencia CCCAACTGAC

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (75..85)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre S8_01
puntuación 0,960
secuencia AATAGAATTAG

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 94..104

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre S8_01
puntuación 0,966
secuencia AACTAAATTAG

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (129..139)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre DELTAEF1_01
puntuación 0,960
secuencia GCACACCTCAG

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (155..165)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre GATA_C
puntuación 0,964
secuencia AGATAAATCCA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 170..178

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre CMYB_01
puntuación 0,958
secuencia CTTCAGTTG

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 176..189

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre GATA1_02
puntuación 0,959
secuencia TTGTAGATAGGACA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 180..190

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre GATA_C
puntuación 0,953
secuencia AGATAGGACAT

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 284..299

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre TAL1ALPHAE47_01
puntuación 0,973
secuencia CATAACAGATGGTAAG

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 284..299

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre TAL1BETAE47_01
puntuación 0,983
secuencia CATAACAGATGGTAAG

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 284..299

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre TAL1BETAITF2_01
puntuación 0,978
secuencia CATAACAGATGGTAAG

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: complemento (287..296)
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre MYOD_Q6
puntuación 0,954
secuencia ACCATCTGTT

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: complemento (302..314)
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre GATA1_04
puntuación 0,953
secuencia TCAAGATAAAGTA

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 393..405
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre IK1_01
puntuación 0,963
secuencia AGTTGGGAATTCC

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 393..404
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre IK2_01
puntuación 0,985
secuencia AGTTGGGAATTC

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 396..405
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre CREL_01
puntuación 0,962
secuencia TGGGAATTCC

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 423..436
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre GATA1_02
puntuación 0,950
secuencia TCAGTGATATGGCA

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: complemento (478..489)
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre SRY_02
puntuación 0,951
secuencia TAAAACAAAACA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 486..493

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre E2F_02
puntuación 0,957
secuencia TTTAGCGC

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (514..521)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre MZF1_01
puntuación 0,975
secuencia TGAGGGGA

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 31:

```
TGAGTGCAGT GTTACATGTC AGTTGGGTTA AGTTTGTTAA TGTCATTCAA ATCTTCTATG   60
TCTTGATTTG CCTGCTAATT CTATTATTTC TGGAACTAAA TTAGTTTGAT GGTTCTATTA  120
GTTATTGACT GAGGTGTGCT AATCTCCCAT TATGTGGATT TATCTATTTC TTCAGTTGTA  180
GATAGGACAT TGATAGATAC ATAAGTACCA GGACAAAAGC AGGGAGATCT TTTTTCCAAA  240
ATCAGGAGAA AAAAATGACA TCTGGAAAAC CTATAGGGAA AGGCATAACA GATGGTAAGG  300
ATACTTTATC TTGAGTAGGA GAGCCTTCCT GTGGCAACGT GGAGAAGGGA AGAGGTCGTA  360
GAATTGAGGA GTCAGCTCAG TTAGAAGCAG GGAGTTGGGA ATTCCGTTCA TGTGATTTAG  420
CATCAGTGAT ATGGCAAATG TGGGACTAAG GGTAGTGATC AGAGGGTTAA AATTGTGTGT  480
TTTGTTTTAG CGCTGCTGGG GCATCGCCTT GGGTCCCCTC AAACAGATTC CCATGAATCT  540
CTTCAT                                                             546
```

(2) INFORMACIÓN DE LA SEQ ID NO: 32:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 23 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: monocatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 32:

GTACCAGGGA CTGTGACCAT TGC 23

(2) INFORMACIÓN DE LA SEQ ID NO: 33:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 24 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: monocatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 33:

CTGTGACCAT TGCTCCCAAG AGAG 24

(2) INFORMACIÓN DE LA SEQ ID NO: 34:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 861 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: bicatenaria
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN genómico

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: promotor
(B) LOCALIZACIÓN: 1..806

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de inicio de la transcripción
(B) LOCALIZACIÓN: 807

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: complemento (60..70)
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre NFY_Q6
puntuación 0,956
secuencia GGACCAATCAT

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 70..77
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre MZF1_01
puntuación 0,962
secuencia CCTGGGGA

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 124..132
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre CMYB_01
puntuación 0,994
secuencia TGACCGTTG

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (126..134)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre VMYB_02
puntuación 0,985
secuencia TCCAACGGT

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 135..143

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre STAT_01
puntuación 0,968
secuencia TTCCTGGAA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (135..143)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre STAT_01
puntuación 0,951
secuencia TTCCAGGAA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (252..259)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre MZF1_01
puntuación 0,956
secuencia TTGGGGGA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 357..368

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre IK2_01
puntuación 0,965
secuencia GAATGGGATTTC

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 384..391

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre MZF1_01
puntuación 0,986
secuencia AGAGGGGA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (410..421)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre SRY_02
puntuación 0,955
secuencia GAAAACAAAACA

(ix) CARACTERÍSTICA:
- (A) NOMBRE/CLAVE: sitio de unión a TF
- (B) LOCALIZACIÓN: 592..599
- (C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
- (D) OTRA INFORMACIÓN: nombre MZF1_01
  puntuación 0,960
  secuencia GAAGGGGA

(ix) CARACTERÍSTICA:
- (A) NOMBRE/CLAVE: sitio de unión a TF
- (B) LOCALIZACIÓN: 618..627
- (C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
- (D) OTRA INFORMACIÓN: nombre MYOD_Q6
  puntuación 0,981
  secuencia AGCATCTGCC

(ix) CARACTERÍSTICA:
- (A) NOMBRE/CLAVE: sitio de unión a TF
- (B) LOCALIZACIÓN: 632..642
- (C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
- (D) OTRA INFORMACIÓN: nombre DELTAEF1_01
  puntuación 0,958
  secuencia TCCCACCTTCC

(ix) CARACTERÍSTICA:
- (A) NOMBRE/CLAVE: sitio de unión a TF
- (B) LOCALIZACIÓN: complemento (813..823)
- (C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
- (D) OTRA INFORMACIÓN: nombre S8_01
  puntuación 0,992
  secuencia GAGGCAATTAT

(ix) CARACTERÍSTICA:
- (A) NOMBRE/CLAVE: sitio de unión a TF
- (B) LOCALIZACIÓN: complemento (824..831)
- (C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
- (D) OTRA INFORMACIÓN: nombre MZF1_01
  puntuación 0,986
  secuencia AGAGGGGA

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 34:

```
TACTATAGGG CACGCGTGGT CGACGGCCGG GCTGTTCTGG AGCAGAGGGC ATGTCAGTAA   60

TGATTGGTCC CTGGGGAAGG TCTGGCTGGC TCCAGCACAG TGAGGCATTT AGGTATCTCT  120

CGGTGACCGT TGGATTCCTG GAAGCAGTAG CTGTTCTGTT TGGATCTGGT AGGGACAGGG  180

CTCAGAGGGC TAGGCACGAG GGAAGGTCAG AGGAGAAGGS AGGSARGGCC CAGTGAGARG  240
```

GGAGCATGCC TTCCCCCAAC CCTGGCTTSC YCTTGGYMAM AGGGCGKTTY TGGGMACTTR 300

AAYTCAGGGC CCAASCAGAA SCACAGGCCC AKTCNTGGCT SMAAGCACAA TAGCCTGAAT 360

GGGATTTCAG GTTAGNCAGG GTGAGAGGGG AGGCTCTCTG GCTTAGTTTT GTTTTGTTTT 420

CCAAATCAAG GTAACTTGCT CCCTTCTGCT ACGGGCCTTG GTCTTGGCTT GTCCTCACCC 480

AGTCGGAACT CCCTACCACT TTCAGGAGAG TGGTTTTAGG CCCGTGGGGC TGTTCTGTTC 540

CAAGCAGTGT GAGAACATGG CTGGTAGAGG CTCTAGCTGT GTGCGGGGCC TGAAGGGGAG 600

TGGGTTCTCG CCCAAAGAGC ATCTGCCCAT TTCCCACCTT CCCTTCTCCC ACCAGAAGCT 660

TGCCTGAGCT GTTTGGACAA AAATCCAAAC CCCACTTGGC TACTCTGGCC TGGCTTCAGC 720

TTGGAACCCA ATACCTAGGC TTACAGGCCA TCCTGAGCCA GGGGCCTCTG GAAATTCTCT 780

TCCTGATGGT CCTTTAGGTT TGGGCACAAA ATATAATTGC CTCTCCCCTC TCCCATTTTC 840

TCTCTTGGGA GCAATGGTCA C 861

(2) INFORMACIÓN DE LA SEQ ID NO: 35:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: monocatenaria
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 35:

CTGGGATGGA AGGCACGGTA 20

(2) INFORMACIÓN DE LA SEQ ID NO: 36:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: monocatenaria
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 36:

GAGACCACAC AGCTAGACAA 20

(2) INFORMACIÓN DE LA SEQ ID NO: 37:

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 555 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: bicatenaria
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN genómico

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: promotor
(B) LOCALIZACIÓN: 1..500

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de inicio de la transcripción
(B) LOCALIZACIÓN: 501

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 191..206
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre ARNT_01
puntuación 0,964
secuencia GGACTCACGTGCTGCT

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 193..204
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre NMYC_01
puntuación 0,965
secuencia ACTCACGTGCTG

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: 193..204
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre USF_01
puntuación 0,985
secuencia ACTCACGTGCTG

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: complemento (193..204)
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre USF_01
puntuación 0,985
secuencia CAGCACGTGAGT

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: complemento (193..204)
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector
(D) OTRA INFORMACIÓN: nombre NMYC_01
puntuación 0,956
secuencia CAGCACGTGAGT

(ix) CARACTERÍSTICA:
(A) NOMBRE/CLAVE: sitio de unión a TF
(B) LOCALIZACIÓN: complemento (193..204)
(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre MYCMAX_02
puntuación 0,972
secuencia CAGCACGTGAGT

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 195..202

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre USF_C
puntuación 0,997
secuencia TCACGTGC

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (195..202)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre USF_C
puntuación 0,991
secuencia GCACGTGA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (210..217)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre MZF1_01
puntuación 0,968
secuencia CATGGGGA

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 397..410

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre ELK1_02
puntuación 0,963
secuencia CTCTCCGGAAGCCT

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 400..409

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre CETS1P54_01
puntuación 0,974
secuencia TCCGGAAGCC

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (460..470)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre AP1_Q4
puntuación 0,963
secuencia AGTGACTGAAC

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: complemento (460..470)

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre AP1FJ_Q2
puntuación 0,961
secuencia AGTGACTGAAC

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sitio de unión a TF

(B) LOCALIZACIÓN: 547..555

(C) MÉTODO DE IDENTIFICACIÓN: predicción con matInspector

(D) OTRA INFORMACIÓN: nombre PADS_C
puntuación 1,000
secuencia TGTGGTCTC

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 37:

```
CTATAGGGCA CGCKTGGTCG ACGGCCCGGG CTGGTCTGGT CTGTKGTGGA GTCGGGTTGA   60

AGGACAGCAT TTGTKACATC TGGTCTACTG CACCTTCCCT CTGCCGTGCA CTTGGCCTTT  120

KAWAAGCTCA GCACCGGTGC CCATCACAGG GCCGGCAGCA CACACATCCC ATTACTCAGA  180

AGGAACTGAC GGACTCACGT GCTGCTCCGT CCCCATGAGC TCAGTGGACC TGTCTATGTA  240

GAGCAGTCAG ACAGTGCCTG GGATAGAGTG AGAGTTCAGC CAGTAAATCC AAGTGATTGT  300

CATTCCTGTC TGCATTAGTA ACTCCCAACC TAGATGTGAA AACTTAGTTC TTTCTCATAG  360

GTTGCTCTGC CCATGGTCCC ACTGCAGACC CAGGCACTCT CCGGAAGCCT GGAAATCACC  420

CGTGTCTTCT GCCTGCTCCC GCTCACATCC CACACTTGTG TTCAGTCACT GAGTTACAGA  480

TTTTGCCTCC TCAATTTCTC TTGTCTTAGT CCCATCCTCT GTTCCCCTGG CCAGTTTGTC  540

TAGCTGTGTG GTCTC                                                   555
```

(2) INFORMACIÓN DE LA SEQ ID NO: 38:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 334 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: bicatenaria

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(F) TIPO DE TEJIDO: Ganglios linfáticos

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: 170..247

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 6,9
seq LWVLLLCAHVVTL/LV

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 38:

```
ATTTTTCCTG TGGTGGGTTC ACACGCAGCT AGACACAGCT AACTTGAGTC TTGGAGCTCC   60

TAGAGGGAAG CTTCTGGAAA GGAAGGCTCT TCAGGACCTC TTAGGAGCCA GAGAAGAGGA  120

CGTTGTCACA GATAAAGAGC CAGGCTCACC AGCTCCTGAC GCATGCATC ATG ACC ATG  178
                                                      Met Thr Met
                                                         -25

AGA CAC AAC TGG ACA CCA GAC CTC AGC CCT TTG TGG GTC CTG CTC CTG     226
Arg His Asn Trp Thr Pro Asp Leu Ser Pro Leu Trp Val Leu Leu Leu
            -20                 -15                 -10

TGT GCC CAC GTC GTC ACT CTC CTG GTC AGA GCC ACA CCT GTC TCG CAG     274
Cys Ala His Val Val Thr Leu Leu Val Arg Ala Thr Pro Val Ser Gln
         -5                  1                   5

ACC AYC ACA GCT GCS ACT GCC TCA GTT AGA AGC ACA AAG GAC CCC TGC     322
Thr Xaa Thr Ala Ala Thr Ala Ser Val Arg Ser Thr Lys Asp Pro Cys
 10                  15                  20                  25

CCC ACC CAG GGG                                                     334
Pro Thr Gln Gly
```

(2) INFORMACIÓN DE LA SEQ ID NO: 39:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 55 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: *Homo Sapiens*

(F) TIPO DE TEJIDO: Ganglios linfáticos

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: sig_peptide

(B) LOCALIZACIÓN: -26..-1

(C) MÉTODO DE IDENTIFICACIÓN: matriz de von Heijne

(D) OTRA INFORMACIÓN: puntuación 6,9
seq LWVLLLCAHVVTL/LV

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 39:

```
Met Thr Met Arg His Asn Trp Thr Pro Asp Leu Ser Pro Leu Trp Val
-25                 -20                 -15

Leu Leu Leu Cys Ala His Val Val Thr Leu Leu Val Arg Ala Thr Pro
-10                 -5                  1                   5

Val Ser Gln Thr Xaa Thr Ala Ala Thr Ala Ser Val Arg Ser Thr Lys
            10                  15                  20

Asp Pro Cys Pro Thr Gln Gly
        25
```