



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 359 401**

51 Int. Cl.:

C12N 15/12 (2006.01)	C07K 14/47 (2006.01)
C12N 15/10 (2006.01)	C12N 15/66 (2006.01)
C12Q 1/68 (2006.01)	G01N 33/50 (2006.01)
C07K 16/18 (2006.01)	G01N 33/53 (2006.01)
A61K 48/00 (2006.01)	A61K 38/17 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **03016783 .7**

96 Fecha de presentación : **31.07.1998**

97 Número de publicación de la solicitud: **1378571**

97 Fecha de publicación de la solicitud: **07.01.2004**

54 Título: **ESTs 5' para proteínas secretadas, expresadas en diversos tejidos.**

30 Prioridad: **01.08.1997 US 905279**

45 Fecha de publicación de la mención BOPI:
23.05.2011

45 Fecha de la publicación del folleto de la patente:
23.05.2011

73 Titular/es: **MERCK SERONO BIODEVELOPMENT**
37, rue Saint Romain
69008 Lyon, 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).

DESCRIPCIÓN

ESTs 5' para proteínas secretadas, expresadas en diversos tejidos

Antecedentes de la Invención

- 5 Los 50.000-100.000 genes estimados diseminados a lo largo de los cromosomas humanos ofrecen una promesa tremenda para la comprensión, diagnóstico y tratamiento de enfermedades humanas. Además, las sondas que son capaces de hibridarse específicamente con loci distribuidos a lo largo del genoma humano encuentran aplicación en la construcción de mapas cromosómicos de alta resolución y en la identificación de individuos.
- 10 En el pasado, la caracterización de incluso un solo gen humano era un proceso laborioso, que requería años de esfuerzo. Recientes progresos en las áreas de vectores de clonación, secuenciación de ADN y tecnología informática se han fundido para acelerar en gran medida la velocidad a la que los genes humanos pueden aislarse, secuenciarse, cartografiarse y caracterizarse. Los vectores de clonación, tales como cromosomas artificiales de levaduras (YAC) y cromosomas artificiales bacterianos (BAC) son capaces de aceptar insertos de ADN que varían de 300 a 1000 kilobases (kb) o
- 15 100-400 kb de longitud, respectivamente, facilitando, con ello, la manipulación y ordenación de secuencias de ADN distribuidas a lo largo de grandes distancias sobre los cromosomas humanos. Las máquinas de secuenciación de ADN automáticas permiten la rápida secuenciación de genes humanos. Los programas bioinformáticos permiten la comparación de secuencias de ácidos nucleicos y proteínas, ayudando, con ello, a la caracterización de productos génicos humanos.
- 20 En la actualidad se siguen dos líneas diferentes para identificar y caracterizar los genes distribuidos a lo largo del genoma humano. En una línea, fragmentos grandes de ADN genómico se aíslan, se clonan y se secuencian. Se identifican los marcos de lectura abiertos potenciales de estas secuencias genómicas utilizando programas bioinformáticos. Sin embargo, esta línea supone secuenciar tramos largos de ADN humano que no codifican proteínas para encontrar las secuencias codificantes de
- 25 proteínas dispersas a lo largo del genoma. Además de requerir una extensa secuenciación, los programas bioinformáticos pueden caracterizar erróneamente las secuencias genómicas obtenidas. Así, el programa puede producir falsos positivos en los que se caracteriza erróneamente un ADN que no codifica como un ADN codificante, o falsos negativos en los que un ADN codificante se designa erróneamente como un ADN no codificante.
- 30 Una línea alternativa toma una ruta más directa para identificar y caracterizar genes humanos. En esta línea, se sintetizan ADN complementarios (ADNc) a partir de ARN mensajeros (ARNm) aislados que codifican proteínas humanas. Utilizando esta línea, sólo se realiza la secuenciación sobre el ADN que se deriva de porciones codificantes de proteínas del genoma. A menudo, sólo se secuencian tramos cortos del ADNc para obtener secuencias denominadas etiquetas de secuencia expresada (EST, "expressed sequence tags"). Las EST a continuación pueden utilizarse para aislar o purificar ADNc
- 35 extendido que incluye secuencias adyacentes a las secuencias de EST. Los ADNc extendidos pueden contener la secuencia completa de la EST que se utilizó para obtenerlos, o sólo una porción de la secuencia de la EST que se utilizó para obtenerlos. Además, los ADNc extendidos pueden contener la secuencia completa del gen a partir del cual se derivó la EST o, como alternativa, los ADNc extendidos pueden incluir porciones de la secuencia codificante del gen a partir del cual se derivó la EST. Se apreciará que pueden existir varios ADNc extendidos que incluyen la secuencia EST como resultado de un corte y empalme alternativo o de la actividad de promotores alternativos.
- 40 En el pasado, estas secuencias EST cortas a menudo se obtenían de bancos de ADNc cebados con oligo-dT. Por consiguiente, corresponden principalmente a la región 3' no traducida del ARNm. En parte, el predominio de secuencias EST derivadas del extremo 3' del ARNm es un resultado del hecho de que las técnicas típicas para obtener ADNc no son adecuadas para aislar secuencias de ADNc derivadas de los extremos 5' de los ARNm (Adams et al., Nature, 377: 3-174, 1996; Hillier et al., Genome Res., 6: 807-828, 1996).
- 45 Además, en los casos indicados en los que se han obtenido secuencias de ADNc más largas, las secuencias indicadas corresponden, de forma típica, a secuencias codificantes y no incluyen la región 5' completa no traducida del ARNm a partir del cual se deriva el ADNc. Estas secuencias incompletas pueden no incluir el primer exón del ARNm, en particular en situaciones en las que el primer exón es corto. Además, pueden no incluir algunos exones, a menudo los que son cortos, que están colocados aguas arriba de los sitios de corte y empalme. Por tanto, existe la necesidad de obtener secuencias derivadas de los extremos 5' de los ARNm.
- 50 Aunque muchas secuencias derivadas de cromosomas humanos tienen aplicaciones prácticas, las líneas basadas en la identificación y caracterización de aquellas secuencias cromosómicas que codifican un producto proteico son particularmente relevantes para usos de diagnóstico y terapéuticos. De los 50.000-100.000 genes codificantes de proteínas, aquellos genes que codifican proteínas que se secretan desde la célula en la que se sintetizan, así como las proteínas secretadas en sí mismas,
- 60

son particularmente valiosos como agentes terapéuticos potenciales. Estas proteínas a menudo están implicadas en la comunicación de célula a célula, y pueden ser responsables de producir una respuesta clínicamente importante en sus células diana.

De hecho, varias proteínas de secreción, incluyendo el activador de plasminógeno tisular, G-CSF, GM-CSF, eritropoyetina, hormona del crecimiento humana, insulina, interferón- α , interferón- β , interferón- γ e interleuquina-2, tienen un uso clínico en la actualidad. Estas proteínas se utilizan para tratar una amplia gama de trastornos, incluyendo infarto agudo de miocardio, ictus isquémico agudo, anemia, diabetes, deficiencia de la hormona del crecimiento, hepatitis, carcinoma renal, neutropenia inducida por quimioterapia, y esclerosis múltiple. Por estas razones, los ADNc extendidos que codifican proteínas secretadas, o porciones de éstas, representan una fuente particularmente valiosa de agentes terapéuticos. Por tanto, existe la necesidad de identificar y caracterizar proteínas secretadas y los ácidos nucleicos que las codifican.

Además de ser terapéuticamente útiles en sí mismas, las proteínas de secreción incluyen péptidos cortos, denominados péptidos señal, en sus extremos amino, que dirigen su secreción. Estos péptidos señal son codificados por las secuencias señal colocadas en los extremos 5' de las secuencias codificantes de genes que codifican proteínas secretadas. Debido a que estos péptidos señal dirigirán la secreción extracelular de cualquier proteína a la que estén operablemente conectados, las secuencias señal pueden aprovecharse para dirigir la secreción eficaz de cualquier proteína mediante la conexión operable de las secuencias señal a un gen que codifica la proteína cuya secreción se desea. Además, también pueden utilizarse porciones de secuencias señal para dirigir la importación intracelular de un péptido o una proteína de interés. Esto puede resultar beneficioso en estrategias de terapia génica, en las que se desea dirigir un producto génico concreto a células distintas a la célula en la que se produce. Las secuencias señal que codifican péptidos señal también tienen aplicación para simplificar las técnicas de purificación de proteínas. En estas aplicaciones, la secreción extracelular de la proteína deseada facilita en gran medida la purificación mediante la reducción del número de proteínas no deseadas, a partir de las cuales debe seleccionarse la proteína deseada. Por tanto, existe la necesidad de identificar y caracterizar las porciones 5' de los genes para proteínas de secreción que codifican péptidos señal.

La información pública sobre el número de genes humanos para los que se han identificado y caracterizado los promotores y las regiones reguladoras aguas arriba es bastante limitada. En parte, esto puede ser debido a la dificultad de aislar estas secuencias reguladoras. Las secuencias reguladoras aguas arriba, tales como los sitios de unión del factor de transcripción, son, de forma típica, demasiado cortas para ser utilizadas como sondas para aislar promotores a partir de bancos genómicos humanos. En fechas recientes se han desarrollado algunas líneas para aislar promotores humanos. Una de ellas consiste en fabricar un banco de isla CpG (Cross et al., Nature Genetics, 6: 236-244, 1994). La segunda consiste en aislar secuencias de ADN genómicas humanas que contienen sitios de unión de Sp1 utilizando la proteína de unión Sp1 (Mortlock et al., Genome Res., 6: 327-335, 1996). Ambas líneas tienen sus límites, debido a la falta de especificidad o de exhaustividad. Las presentes EST 5' pueden utilizarse para identificar y aislar, de forma eficaz, regiones reguladoras aguas arriba que controlan la colocación, etapa del desarrollo, velocidad y cantidad de síntesis proteica, así como la estabilidad del ARNm (Theil, BioFactors, 4: 87-93, 1993). Cuando se han identificado y caracterizado, estas regiones reguladoras pueden utilizarse en terapia génica o en esquemas de purificación de proteínas para obtener la cantidad y colocación deseadas de síntesis proteica, o para inhibir, reducir o evitar la síntesis de productos génicos no deseados.

Además, las EST que contienen los extremos 5' de los genes de proteínas de secreción pueden incluir secuencias útiles como sondas para el cartografiado cromosómico y la identificación de individuos.

Por tanto, existe la necesidad de identificar y caracterizar secuencias aguas arriba de las secuencias codificantes 5' de genes que codifican proteínas de secreción.

Compendio de la invención

La presente invención se refiere a EST purificadas, aisladas o recombinantes que incluyen secuencias derivadas de los extremos 5' auténticos de sus correspondientes ARNm. Más específicamente la invención es un péptido señal que tiene la secuencia de aminoácidos -26 del SEQ ID No 306 y una secuencia señal aislada que codifica dicho péptido. La expresión "ARNm correspondiente" se refiere al ARNm que es el molde para la síntesis del ADNc que produjo la EST 5'. Estas secuencias se indicarán en lo sucesivo como "EST 5'". Tal como se utiliza en la presente memoria, el término "purificado" no requiere una pureza absoluta; en realidad, pretende ser una definición relativa. Los clones de EST 5' individuales aislados a partir de un banco de ADNc se han purificado convencionalmente hasta una homogeneidad electroforética. Las secuencias obtenidas a partir de estos clones no podrían obtenerse directamente del banco ni del ADN humano total. Los clones de ADNc no aparecen en la naturaleza como tales, sino que se obtienen mediante manipulación de una sustancia parcialmente purificada de origen natural (ARN mensajero). La conversión de ARNm en un banco de ADNc implica la creación de

- una sustancia sintética (ADNc), y los clones de ADNc individuales puros pueden aislarse a partir del banco sintético mediante selección clónica. Por tanto, la creación de un banco de ADNc a partir de ARN mensajero y el posterior aislamiento de clones individuales a partir de este banco produce una purificación de aproximadamente 10^4 - 10^6 veces del mensaje nativo. La purificación del material de partida o material natural hasta al menos un orden de magnitud, preferiblemente dos o tres órdenes, y más preferiblemente cuatro o cinco órdenes de magnitud, se contempla expresamente.
- Tal como se utiliza en la presente memoria, el término “aislado” requiere que el material se retire de su medio original (por ejemplo, el medio natural, si aparece en la naturaleza). Por ejemplo, un polinucleótido de origen natural presente en un animal vivo no está aislado, pero el mismo polinucleótido, separado de algunos o de todos los materiales coexistentes en el sistema natural, está aislado.
- Tal como se utiliza en la presente memoria, el término “recombinante” significa que la EST 5' es adyacente al “esqueleto” del ácido nucleico, al cual no está adyacente en su medio natural. Además, para que estén “enriquecidas”, las EST 5' deben representar 5% o más del número de insertos de ácido nucleico en una población de moléculas de esqueleto de ácido nucleico. Las moléculas de esqueleto según la presente invención incluyen ácidos nucleicos, tales como vectores de expresión, ácidos nucleicos autorreplicantes, virus, ácidos nucleicos integrantes, y otros vectores o ácidos nucleicos utilizados para mantener o manipular un inserto de ácido nucleico de interés. Preferiblemente, las EST 5' enriquecidas representan 15% o más del número de insertos de ácido nucleico en la población de moléculas de esqueleto recombinantes. Más preferiblemente, las EST 5' enriquecidas representan 50% o más del número de insertos de ácido nucleico en la población de moléculas de esqueleto recombinantes. En una realización muy preferida, las EST 5' enriquecidas representan 90% o más del número de insertos de ácido nucleico en la población de moléculas de esqueleto recombinantes.
- Las condiciones de hibridación “rigurosas”, “moderadas” y “bajas” se definen como en el Ejemplo 29.
- A menos que se indique lo contrario, una secuencia “complementaria” es totalmente complementaria.
- Por tanto, las EST 5' en bancos de ADNc en los que una o más de las EST 5' constituyen 5% o más del número de insertos de ácido nucleico en las moléculas de esqueleto son “EST 5' recombinantes enriquecidas”, como se define en la presente memoria. De forma similar, las EST 5' en una población de plásmidos en los que una o más de las EST 5' de la presente invención se han insertado de forma que representan 5% o más del número de insertos en el esqueleto del plásmido son “EST 5' recombinantes enriquecidas”, como se define en la presente memoria. Sin embargo, las EST 5' en bancos de ADNc en los que las EST 5' constituyen menos de 5% del número de insertos de ácido nucleico en la población de moléculas de esqueleto, tales como los bancos en los que las moléculas de esqueleto que tienen un inserto EST 5' son extremadamente raras, no son “EST 5' recombinantes enriquecidas”.
- Tal como se utiliza en la presente memoria, una proteína “secretada” es aquella que, cuando se expresa en una célula anfitriona adecuada, es transportada a través de una membrana, incluyendo el transporte como resultado de tener péptidos señal en su secuencia de aminoácidos. Las proteínas “secretadas” incluyen, sin limitación, proteínas secretadas completamente (por ejemplo, proteínas solubles) o parcialmente (por ejemplo, receptores) a partir de la célula en la cual se expresan. Las proteínas “secretadas” también incluyen, sin limitación, proteínas que se transportan a través de la membrana del retículo endoplásmico.
- Estas EST 5' incluyen secuencias de ácidos nucleicos, denominadas secuencias señal, que codifican péptidos señal que dirigen la secreción extracelular de proteínas codificadas por los genes a partir de los cuales se derivan las EST 5'. En general, los péptidos señal se colocan en los extremos amino de las proteínas secretadas.
- Las proteínas secretadas son traducidas por ribosomas asociados con el retículo endoplásmico “rugoso”. En general, las proteínas secretadas se trasladan cotraduccionalmente a la membrana del retículo endoplásmico. La asociación del ribosoma con el retículo endoplásmico durante la traducción de las proteínas secretadas es mediada por el péptido señal. El péptido señal, de forma típica, se escinde después de su entrada cotraduccional en el retículo endoplásmico. Después del transporte al retículo endoplásmico, las proteínas secretadas pueden trasladarse a través del aparato de Golgi. En el aparato de Golgi, las proteínas pueden sufrir una modificación postraduccional antes de entrar en vesículas secretoras que las transportan a través de la membrana celular.
- Las EST 5' descritas en la presente memoria tienen varias aplicaciones importantes. Por ejemplo, pueden utilizarse para obtener y expresar clones de ADNc que incluyen las secuencias codificantes completas de proteínas de los Correspondientes productos génicos, incluyendo los sitios de inicio de la traducción auténticos derivados de los extremos 5' de las secuencias codificantes de los ARNm a partir de los cuales derivan las EST 5'. Estos ADNc se denominarán en lo sucesivo “ADNc completos”.

- Estos ADNc también pueden incluir ADN derivado de secuencias de ARNm aguas arriba del sitio de inicio de la traducción. Las secuencias de ADNc completos pueden utilizarse para expresar las proteínas correspondientes a las EST 5'. Como se analizó anteriormente, las proteínas secretadas son terapéuticamente importantes. Por tanto, las proteínas expresadas a partir de los ADNc pueden ser útiles en el tratamiento o control de una diversidad de trastornos humanos. Las EST 5' también pueden utilizarse para obtener el correspondiente ADN genómico. La expresión "correspondiente ADN genómico" se refiere al ADN genómico que codifica el ARNm a partir del cual deriva la EST 5'.
- Como alternativa, las EST 5' pueden utilizarse para obtener y expresar porciones codificantes del ADNc extendido de la proteína secretada. Las porciones pueden comprender los péptidos señal de las proteínas secretadas o las proteínas maduras generadas cuando el péptido señal se escinde. Las porciones también pueden comprender polipéptidos que tienen al menos 10 aminoácidos consecutivos codificados por los ADNc extendidos o los ADNc completos. Como alternativa, las porciones pueden comprender al menos 15 aminoácidos consecutivos codificados por los ADNc extendidos o los ADNc completos. En algunas realizaciones, las porciones pueden comprender al menos 25 aminoácidos consecutivos codificados por los ADNc extendidos o los ADNc completos. En otras realizaciones, las porciones pueden comprender al menos 40 aminoácidos codificados por los ADNc extendidos o los ADNc completos.
- También pueden obtenerse anticuerpos que reconocen específicamente las proteínas secretadas completas codificadas por los ADNc extendidos, los ADNc completos, o sus fragmentos que tengan al menos 10 aminoácidos consecutivos, al menos 15 aminoácidos consecutivos, al menos 25 aminoácidos consecutivos, o al menos 40 aminoácidos consecutivos, como se describe a continuación. Los anticuerpos que reconocen específicamente la proteína madura generada cuando se escinde el péptido señal también pueden obtenerse como se describe a continuación. De forma similar, también pueden obtenerse anticuerpos que reconocen específicamente los péptidos señal codificados por los ADNc extendidos o los ADNc completos.
- En algunas realizaciones, los ADNc extendidos obtenidos utilizando las EST 5' incluyen la secuencia señal. En otras realizaciones, los ADNc extendidos obtenidos utilizando las EST 5' pueden incluir las secuencias codificantes completas para la proteína madura (es decir, la proteína generada cuando el polipéptido señal se escinde). Además, los ADNc extendidos obtenidos utilizando las EST 5' pueden incluir regiones reguladoras aguas arriba del sitio de inicio de la traducción, o aguas abajo del codón de terminación que controla la cantidad, colocación o etapa de desarrollo de la expresión génica.
- Como se analizó anteriormente, las proteínas secretadas son terapéuticamente importantes. Por tanto, las proteínas expresadas a partir de los ADNc extendidos o los ADNc completos obtenidas utilizando las EST 5' pueden ser útiles para tratar o controlar una diversidad de trastornos humanos.
- Las EST 5' (o los ADNc o ADN genómicos derivados de ellas) pueden utilizarse en procedimientos forenses para identificar individuos, o en procedimientos de diagnóstico para identificar individuos que tienen enfermedades genéticas como resultado de la expresión anómala de los genes correspondientes a las EST 5'. Además, la presente invención es útil para construir un mapa de alta resolución de los cromosomas humanos.
- La presente invención también se refiere al uso de vectores de secreción capaces de dirigir la secreción de una proteína de interés. Estos vectores pueden utilizarse en estrategias de terapia génica en las que se desea producir un producto génico en una célula que se va a dirigir a otro emplazamiento 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 de una manera espacial o temporal deseada, o a un nivel deseado. Estos vectores pueden incluir secuencias aguas arriba de las EST 5', tales como promotores o secuencias reguladoras aguas arriba.
- Finalmente, la presente invención también puede ser utilizada para la terapia génica para controlar o tratar enfermedades genéticas. Los péptidos señal pueden fusionarse con proteínas heterólogas para dirigir su secreción extracelular.
- Los clones bacterianos que contienen plásmidos Bluescript tienen insertos que contienen las EST 5' de la presente invención (SEQ ID NO: 73+75 está conservada en la actualidad a 80°C en glicerol al 4% (v/v) en los laboratorios de los autores de la presente invención con la denominación listada junto a los SEQ ID NO: de II). Los insertos pueden recuperarse a partir de los materiales depositados cultivando los clones apropiados en un medio adecuado. El ADN de Bluescript puede aislarse a continuación utilizando procedimientos de aislamiento de plásmidos familiares para los expertos en la técnica, tal como procedimientos de aislamiento de plásmidos de lisis alcalina a gran escala o de minipreps de lisis alcalina. Si se desea, el ADN del plásmido puede enriquecerse más mediante centrifugación en un gradiente de cloruro de cesio, cromatografía de exclusión molecular, o

- cromatografía de intercambio aniónico. El ADN del plásmido obtenido utilizando estos procedimientos puede entonces manipularse utilizando técnicas de clonación convencionales familiares para los expertos en la técnica. Como alternativa, puede realizarse una PCR con cebadores diseñados en ambos extremos de la inserción de EST. El producto de PCR que corresponde a la EST 5' puede manipularse a continuación utilizando técnicas de clonación convencionales familiares para los expertos en la técnica.
- 5 Se describe en la presente memoria un ácido nucleico purificado o aislado que tiene la secuencia de uno de los SEQ ID NO: 73+75 o que tiene una secuencia complementaria a ésta. En una realización, el ácido nucleico es recombinante.
- 10 Se describe en la presente memoria un ácido nucleico purificado o aislado que comprende al menos 10 bases consecutivas de la secuencia de los SEQ ID NO: 73+75 o una de las secuencias complementarias a ésta.
- 15 Se describe en la presente memoria un ácido nucleico purificado o aislado que comprende al menos 15 bases consecutivas de una de las secuencias de los SEQ ID NO: 73+75 o una de las secuencias complementarias a ésta. En una realización, el ácido nucleico es recombinante. También se describe en la presente memoria un ácido nucleico purificado o aislado de al menos 15 bases capaz de hibridarse bajo condiciones rigurosas con la secuencia de uno de los SEQ ID NO: 73+75 o una de las secuencias complementarias a las secuencias de los SEQ ID NO: 73+75. En una realización, el ácido nucleico es recombinante.
- 20 Se describe en la presente memoria un ácido nucleico purificado o aislado que codifica un producto génico humano, teniendo dicho producto génico humano una secuencia parcialmente codificada por una de las secuencias de los SEQ ID NO: 73+75.
- 25 Se describe en la presente memoria un método para fabricar un ADNc que codifica una proteína de secreción humana, estando dicha proteína de secreción humana parcialmente codificada por uno de los SEQ ID NO:73+75, que comprende 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 uno de los SEQ ID NO:73+75; hibridar dicho cebador con un ARNm en dicha colección que codifica dicha proteína; transcribir de forma inversa dicho cebador hibridado para fabricar una primera hebra de ADNc a partir de dicho ARNm; fabricar una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc; y aislar el ADNc resultante que codifica dicha proteína que comprende dicha primera hebra de ADNc y dicha segunda hebra de ADNc.
- 30 Se describe en la presente memoria un ADNc aislado o purificado que codifica una proteína de secreción humana, comprendiendo dicha proteína de secreción humana la proteína codificada por uno de los SEQ ID NO: 73+75, o un fragmento de ésta de al menos 10 aminoácidos, siendo obtenible dicho ADNc por el método descrito en el anterior párrafo. En una realización, el ADNc comprende la secuencia codificante completa de la proteína de dicha proteína cuya secuencia está parcialmente incluida en una de las secuencias de los SEQ ID NO: 73+75.
- 35 Se describe en la presente memoria un método para fabricar un ADNc que codifica una proteína de secreción humana que está parcialmente codificada por uno de los SEQ ID NO:73+75, que comprende las etapas de obtener un ADNc que comprende una de las secuencias de los SEQ ID NO:73+75; poner en contacto dicho ADNc con una sonda detectable que comprende al menos 15 nucleótidos consecutivos de dicha secuencia del SEQ ID NO:73+75, o una secuencia complementaria a ésta, 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.
- 40 Se describe en la presente memoria un ADNc aislado o purificado que codifica una proteína de secreción humana, comprendiendo dicha proteína de secreción humana la proteína codificada por uno de los SEQ ID NO: 73+75, o un fragmento de ésta de al menos 10 aminoácidos, siendo obtenible dicho ADNc por el método descrito en el anterior párrafo. En una realización, el ADNc comprende la secuencia codificante completa de la proteína parcialmente incluida en una de las secuencias de los SEQ ID NO: 73+75.
- 45 Se describe en la presente memoria un método para fabricar un ADNc que comprende una de las secuencias de los SEQ ID NO:73+75, 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 de poliA de dicho ARNm; hibridar dicho primer cebador con dicha cola de poliA; transcribir de modo inverso dicho ARNm para fabricar una primera hebra de ADNc; fabricar una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc utilizando al menos un cebador que comprende al menos 15 nucleótidos de una de las secuencias de los SEQ ID NO:73+75; y aislar el ADNc resultante que comprende dicha primera hebra de ADNc y dicha segunda hebra de ADNc.
- 50 Se describe en la presente memoria un método para fabricar un ADNc que comprende una de las secuencias de los SEQ ID NO:73+75, 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 de poliA de dicho ARNm; hibridar dicho primer cebador con dicha cola de poliA; transcribir de modo inverso dicho ARNm para fabricar una primera hebra de ADNc; fabricar una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc utilizando al menos un cebador que comprende al menos 15 nucleótidos de una de las secuencias de los SEQ ID NO:73+75; y aislar el ADNc resultante que comprende dicha primera hebra de ADNc y dicha segunda hebra de ADNc.
- 55 Se describe en la presente memoria un método para fabricar un ADNc que comprende una de las secuencias de los SEQ ID NO:73+75, 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 de poliA de dicho ARNm; hibridar dicho primer cebador con dicha cola de poliA; transcribir de modo inverso dicho ARNm para fabricar una primera hebra de ADNc; fabricar una segunda hebra de ADNc complementaria a dicha primera hebra de ADNc utilizando al menos un cebador que comprende al menos 15 nucleótidos de una de las secuencias de los SEQ ID NO:73+75; y aislar el ADNc resultante que comprende dicha primera hebra de ADNc y dicha segunda hebra de ADNc.

- Se describe en la presente memoria un ADNc aislado o purificado que codifica una proteína de secreción humana, comprendiendo dicha proteína de secreción humana la proteína codificada por uno de los SEQ ID NO: 73+75, o un fragmento de ésta de al menos 10 aminoácidos, siendo obtenible dicho ADNc por el método descrito en el anterior párrafo. En una realización, el ADNc comprende la secuencia codificante completa de la proteína parcialmente incluida en una de las secuencias de los SEQ ID NO: 73+75.
- En una realización del método descrito en los dos párrafos anteriores, la segunda hebra de ADNc se fabrica poniendo en contacto dicha primera hebra de ADNc con un primer par de cebadores, comprendiendo dicho primer par de cebadores un segundo cebador que comprende al menos 15 nucleótidos consecutivos de una de las secuencias de los SEQ ID NO:73+75, y un tercer cebador que tiene una secuencia en su interior que está incluida dentro de la secuencia de dicho primer cebador, realizar una primera reacción en cadena de polimerasa con dicho primer par de cebadores anidados para generar un primer producto de PCR; poner en contacto dicho primer producto de PCR con un segundo par de cebadores, comprendiendo dicho segundo par de cebadores un cuarto cebador, comprendiendo dicho cuarto cebador al menos 15 nucleótidos consecutivos de dicha secuencia de uno de los SEQ ID NO:73+75, y un quinto cebador, siendo capaces dichos cuarto y quinto cebadores de hibridarse con secuencias dentro de dicho primer producto de PCR; y realizar una segunda reacción en cadena de la polimerasa, generando, con ello, un segundo producto de PCR.
- Se describe en la presente memoria un ADNc aislado o purificado que codifica una proteína de secreción humana, comprendiendo dicha proteína de secreción humana la proteína codificada por uno de los SEQ ID NO: 73+75, o un fragmento de ésta de al menos 10 aminoácidos, siendo obtenible dicho ADNc por el método descrito en el anterior párrafo. En una realización, el ADNc comprende la secuencia codificante completa de la proteína parcialmente incluida en una de las secuencias de los SEQ ID NO: 73+75. También se describe el método descrito cuatro párrafos antes, en el que la segunda hebra de ADNc se fabrica poniendo en contacto dicha primera hebra de ADNc con un segundo cebador que comprende al menos 15 nucleótidos consecutivos de las secuencias de los SEQ ID NO: 73+75; hibridar dicho segundo cebador con dicha primera hebra de ADNc; y extender dicho segundo cebador hibridado para generar dicha segunda hebra de ADNc.
- Se describe en la presente memoria un ADNc aislado o purificado que codifica una proteína de secreción humana, comprendiendo dicha proteína de secreción humana la proteína parcialmente codificada por uno de los SEQ ID NO: 73+75, o comprendiendo un fragmento de ésta de al menos 10 aminoácidos, siendo obtenible dicho ADNc por el método descrito en el anterior párrafo. En una realización, el ADNc comprende la secuencia codificante completa de la proteína parcialmente incluida en una de las secuencias de los SEQ ID NO: 73+75. También se describe un método para fabricar una proteína que comprende una de las secuencias de los SEQ ID NO:306+308, que comprende las etapas de obtener un ADNc que codifica la secuencia completa de la proteína parcialmente incluida en una de las secuencias de los SEQ ID NO:73+75; insertar dicho ADNc en un vector de expresión, de forma que dicho ADNc está operablemente conectado con un promotor; introducir dicho vector de expresión en una célula anfitriona, mediante lo cual dicha célula anfitriona produce la proteína codificada por dicho ADNc; y aislar dicha proteína.
- Se describe en la presente memoria una proteína aislada obtenible mediante el método descrito en el anterior párrafo.
- Se describe en la presente memoria un método para obtener un ADN promotor que comprende las etapas de obtener ADN colocados aguas arriba de los ácidos nucleicos de los SEQ ID NO: 73-75, o secuencias complementarias a estos; seleccionar dichos ADN aguas arriba para identificar un promotor capaz de dirigir el inicio de la transcripción; y aislar dicho ADN que comprende dicho promotor identificado. En una realización, la etapa de obtención comprende un paseo cromosómico desde dichos ácidos nucleicos de los SEQ ID NO: 73+75, o secuencias complementarias a estos. En otra realización, la etapa de selección comprende insertar dichas secuencias aguas arriba en un vector indicador promotor. En otra realización, la etapa de selección comprende identificar motivos en dichos ADN aguas arriba que son sitios de unión del factor de transcripción o sitios de inicio de la transcripción. También se describe una proteína aislada o purificada que comprende la secuencia de SEQ ID NO: 307.
- Se describe en la presente memoria un promotor aislado obtenible mediante el método descrito antes.
- Se describe en la presente memoria una proteína aislada purificada que comprende una de las secuencias de los SEQ ID NO: 306+308.
- Se describe en la presente memoria la inclusión de al menos una de las secuencias de los SEQ ID NO: 73+75, o una de las secuencias complementarias a las secuencias de los SEQ ID NO: 73+75, o uno de sus fragmento de al menos 15 nucleótidos consecutivos, en una matriz de EST discretas, o fragmentos de las mismas de al menos 15 nucleótidos de longitud. En una realización, la matriz incluye al menos dos de las secuencias de los SEQ ID NO: 73+75, las secuencias complementarias a

las secuencias de los SEQ ID NO: 73+75, o sus fragmentos de al menos 15 nucleótidos consecutivos. En otra realización, la matriz incluye al menos cinco de las secuencias de los SEQ ID NO: 73+75, las secuencias complementarias a las secuencias de los SEQ ID NO: 73+75, o sus fragmentos al menos 15 nucleótidos consecutivos.

- 5 Se describe en la presente memoria un promotor que tiene una secuencia seleccionada del grupo que consiste en los SEQ ID NO: 31, 34 y 37.

La presente invención se refiere a los péptidos, ácidos nucleicos, polipéptidos, vectores, usos y métodos descritos en las reivindicaciones. Cualquier secuencia distinta de la de los SEQ ID NO: 73, 75, 306 o 308 no forma parte de la presente invención.

10 **Breve Descripción de los Dibujos**

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

- 15 La Figura 2 muestra la distribución de puntuaciones de Von Heijne para las EST 5' en cada una de las categorías descritas en la presente memoria, y la probabilidad de que estas EST 5' codifiquen un péptido señal.

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

- 20 La Figura 4 (descripción de estructuras de promotores aislados a partir de EST 5' SignalTag) proporciona una descripción esquemática de promotores aislados y la forma en que se ensamblan con las correspondientes etiquetas 5'.

Descripción Detallada de la Realización Preferida

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

- 25 La Tabla V muestra la distribución de las EST 5' en cada categoría descrita en la presente memoria y el número de EST 5' en cada categoría que tiene una puntuación mínima de Von Heijne dada.

La Tabla VI muestra la distribución de EST 5' en cada categoría descrita en la presente memoria con respecto al tejido a partir del cual se obtuvieron las EST 5' del correspondiente ARNm.

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

I. Métodos Generales para Obtener EST 5' derivados de ARNm con extremos 5' intactos

Para obtener las EST 5' descritas en la presente memoria, deben obtenerse ARNm con extremos 5' intactos. En la actualidad existen dos líneas para obtener dichos ARNm con extremos 5' intactos, como se describe a continuación: (1) métodos químicos, o (2) métodos enzimáticos.

35 **1. Métodos Químicos para Obtener ARNm que Tienen Extremos 5' Intactos**

- Una de estas líneas es un método de modificación química que implica la derivatización de los extremos 5' de los ARNm y la selección de los ARNm derivatizados. Los extremos 5' de los ARNm eucarióticos poseen una estructura denominada "caperuza", que comprende una guanosina metilada en la posición 7. La caperuza se une a la primera base transcrita del ARNm mediante un enlace 5',5'-trifosfato. En algunos casos, la 5'-guanosina se metila en las posiciones 2 y 7. En casos más raros, la 5'-guanosina se trimetila en las posiciones 2, 7 y 7. En el método químico para obtener ARNm que tienen extremos 5' intactos, la caperuza 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 enlazada a la guanosina metilada en el extremo 5' del ARNm y la ribosa enlazada a la base en el extremo 3' del ARNm, poseen 2',3'-cis-dioles.

- Opcionalmente, el 2',3'-cis-diol de la ribosa 3'-terminal puede modificarse, sustituirse, convertirse o eliminarse químicamente, dejando sólo la ribosa unida a la guanosina metilada en el extremo 5' del ARNm con un 2',3'-cis-diol. Están disponibles una diversidad de técnicas para eliminar el 2',3'-cis-diol sobre la ribosa 3'-terminal. Por ejemplo, puede utilizarse una hidrólisis alcalina controlada para generar fragmentos de ARNm en los que la ribosa 3'-terminal es un 3'-fosfato, 2'-fosfato o (2',3')-ciclofosfato. Después, el fragmento que incluye la ribosa 3' original puede eliminarse de la mezcla mediante cromatografía en una columna oligodT. Como alternativa, una base que carece del 2',3'-cisdiol puede añadirse al extremo 3' del ARNm utilizando una ARN ligasa, tal como ARN ligasa de T4.

El ejemplo 1 a continuación describe un método para la ligación de un nucleósido difosfato al extremo 3' del ARN mensajero.

EJEMPLO 1

Ligación del nucleósido difosfato pCp al extremo 3' de un ARNm

- 5 Se incubó 1 µg de ARN en un medio de reacción final de 10 µl en presencia de 5 U de ARN ligasa de fago T4 en el tampón proporcionado por el fabricante (Gibco-BRL), 40 U de inhibidor de RNasa RNasin (Promega), y 2 µl de 32 pCp (Amersham, Núm. PB 10208). La incubación se realizó a 37°C durante 2 horas, o durante la noche a 7-8°C.
- 10 Después de la modificación o eliminación del 2',3'-cis-diol en la ribosa 3', el 2',3'-cis-diol presente en el extremo 5' del ARNm puede oxidarse utilizando reactivos, tales como NaBH, NaBH₃CN, o peryodato de sodio, convirtiendo, con ello, el 2',3'-cis-diol en un dialdehído. El Ejemplo 2 describe la oxidación del 2',3'-cis-diol en el extremo 5' del ARNm con peryodato de sodio.

EJEMPLO 2

Oxidación del 2',3'-cis-diol en el Extremo 5' del ARNm con Peryodato de Sodio

- 15 Se trataron 0,1 unidades OD de un oligorribonucleótido con caperuza de 47 nucleótidos (incluyendo la caperuza), o un oligorribonucleótido sin caperuza de 46 nucleótidos, como sigue. Los oligorribonucleótidos se produjeron mediante transcripción in vitro utilizando el kit de transcripción "AmpliScribe T7" (Epicentre Technologies). Como se indica a continuación, el molde de ADN para el transcrito de ARN contenía una única citosina. Para sintetizar el ARN sin caperuza, se incluyeron los
- 20 cuatro NTP en la reacción de transcripción in vitro. Para obtener el ARN con caperuza, se substituyó el GTP por un análogo de la caperuza, m7G(5')ppp(5')G. Este compuesto, reconocido por la polimerasa, se incorporó en el extremo 5' del transcrito naciente durante el inicio de la transcripción, pero no se incorporó durante la etapa de extensión. Por consiguiente, el ARN resultante contenía una caperuza en su extremo 5'. Las secuencias de los oligorribonucleótidos producidos mediante la reacción de
- 25 transcripción in vitro eran:

+caperuza:

5'm7GpppGCAUCCUACUCCCAUCCAAUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEQ ID NO: 1)

- 30 -caperuza:

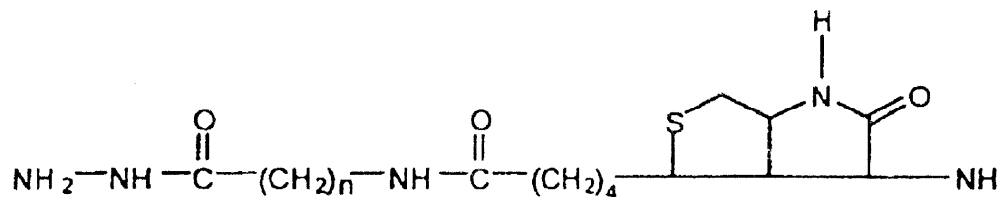
5'-pppGCAUCCUACUCCCAUCCAAUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEQ ID NO: 2)

- 35 Los oligorribonucleótidos se disolvieron en 9 µl de tampón acetato (acetato de sodio 0,1 M, pH 5,2) y 3 µl de disolución de peryodato de sodio 0,1 M recién preparada. La mezcla se incubó durante 1 hora en la oscuridad a 4°C o temperatura ambiente. Después, la reacción se detuvo mediante la adición de 4 µl de etilenglicol al 10%. El producto se precipitó con etanol, se resuspendió en al menos 10 µl de agua o tampón apropiado y se dializó contra agua. Los grupos aldehído resultantes pueden acoplarse a continuación a moléculas que tienen un grupo amina reactivo, tales como grupos hidrazina, carbazida, tiocarbazida o semicarbazida, para facilitar el enriquecimiento de los extremos 5' de los
- 40 ARNm. Las moléculas que tienen grupos amina reactivos que son adecuadas para su uso en la selección de ARNm que tienen extremos 5' intactos incluyen avidina, proteínas, anticuerpos, vitaminas, ligandos capaces de unirse de forma específica a moléculas de receptor, u oligonucleótidos. El Ejemplo 3 a continuación describe el acoplamiento del dialdehído resultante a biotina.

EJEMPLO 3

Acoplamiento del Dialdehído al Extremo 5' de Transcritos con Biotina

El producto de la oxidación obtenido en el Ejemplo 2 se disolvió en 50 µl de acetato de sodio a un pH entre 5 y 5,2, y 50 µl de disolución recién preparada de hidrazida de biotina 0,02 M en una mezcla de metoxietanol/agua (1:1) de fórmula:



En el compuesto utilizado en estos experimentos, $n = 5$. Sin embargo, se apreciará que también pueden utilizarse otras hidrazidas disponibles en el mercado, tales como moléculas de la fórmula anterior en la que n varía de 0 a 5. La mezcla se incubó a continuación durante 2 horas a 37°C , se precipitó con etanol y se dializó contra agua destilada. El Ejemplo 4 demuestra la especificidad de la reacción de biotilación.

EJEMPLO 4

Especificidad de la Biotilación de Transcritos con Caperuza

Se evaluó la especificidad de la biotilación en ARNm con caperuza mediante electroforesis en gel de las siguientes muestras:

Muestra 1. El transcrito de 46 nucleótidos sin caperuza in vitro preparado como en el Ejemplo 2, y marcado con 32 pCp como se describe en el Ejemplo 1.

Muestra 2. El transcrito de 46 nucleótidos sin caperuza in vitro preparado como en el Ejemplo 2, marcado con 32 pCp como se describe en el Ejemplo 1, tratado con la reacción de oxidación del Ejemplo 2, y sometido a las condiciones de biotilación del Ejemplo 3.

Muestra 3. El transcrito de 47 nucleótidos con caperuza in vitro preparado como en el Ejemplo 2, y marcado con 32 pCp como se describe en el Ejemplo 1.

Muestra 4. El transcrito de 47 nucleótidos con caperuza in vitro preparado como en el Ejemplo 2, marcado con 32 pCp como se describe en el Ejemplo 1, tratado con la reacción de oxidación del Ejemplo 2, y sometido a las condiciones de biotilación del Ejemplo 3.

Las muestras 1 y 2 tenían idéntica velocidad de migración, demostrando que los ARN sin caperuza no estaban oxidados y biotinilados. La muestra 3 migra más lentamente que las muestras 1 y 2, mientras que la muestra 4 exhibe la migración más lenta. La diferencia en la migración de los ARN de las muestras 3 y 4 demuestra que los ARN con caperuza estaban específicamente biotinilados.

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

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

EJEMPLO 5

Captura y Liberación de ARNms Biotinilados Utilizando Cuentas Revestidas con Estreptavidina

Las cuentas magnéticas revestidas con estreptavidina se prepararon según las instrucciones del fabricante (CPG Inc., EEUU). 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 cuentas se lavaron a continuación varias veces en agua con SDS al 1%. Las cuentas obtenidas de esta manera se incubaron durante 15 minutos a 95°C en agua que contenía SDS al 2%.

El Ejemplo 6 demuestra la eficacia con la que se recuperaron los ARNm de las cuentas revestidas con estreptavidina.

EJEMPLO 6

Eficacia de Recuperación de ARNms Biotinilados

5 Se evaluó la eficacia del procedimiento de recuperación como sigue. Se marcaron ARN con caperuza con 32 pCp, se oxidaron, se biotinilaron y se unieron a cuentas revestidas con estreptavidina como se describió anteriormente. Posteriormente, los ARN unidos se incubaron durante 5, 15 ó 30 minutos a 95°C en presencia de SDS al 2%.

10 Los productos de la reacción se analizaron mediante electroforesis en 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 redujeron.

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

15 En un método alternativo para obtener ARNm que tienen extremos 5' intactos, un oligonucleótido que se había derivatizado para que contuviera un grupo amina reactivo se acopló de forma específica a los ARNm que tenían una caperuza intacta. Preferiblemente, el extremo 3' del ARNm se bloquea antes de la etapa en la que los grupos aldehído se unen al oligonucleótido derivatizado, como se describió anteriormente, para evitar que el oligonucleótido derivatizado se una al extremo 3' del ARNm. Por ejemplo, puede unirse pCp al extremo 3' del ARNm utilizando ARN ligasa de T4 como se describió en el Ejemplo 1. Sin embargo, como se analizó anteriormente, el bloqueo del extremo 3' del ARNm es una etapa opcional. Los oligonucleótidos derivatizados pueden prepararse como se describe en el Ejemplo 7.

EJEMPLO 7

Derivatización de Oligonucleótidos

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

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

35 Como se analizó anteriormente, los ARNm que van a ser enriquecidos se pueden tratar para eliminar los grupos 3' OH que pueden estar presentes sobre ellos. Esto puede lograrse mediante ligación enzimática de secuencias que carecen de 3' OH, tal como pCp, como se describió en el Ejemplo 1. Como alternativa, los grupos 3' OH pueden eliminarse mediante hidrólisis alcalina como se describe en el Ejemplo 8, a continuación.

EJEMPLO 8

Eliminación de los Grupos 3' OH del ARNm Utilizando Hidrólisis Alcalina

40 En un volumen total de 100 µl de hidróxido de sodio 0,1 N, se incubaron 1,5 µg de ARNm durante 40 a 60 minutos a 4°C. La disolución se neutralizó con ácido acético y se precipitó con etanol.

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

EJEMPLO 9

Oxidación de Dioles de ARNm

45 Se disolvió hasta 1 unidad OD de ARN en 9 µl de tampón (acetato de sodio 0,1 M, pH 6-7) o agua y 3 µl de disolución de peryodato de sodio 0,1 M recién preparada. La reacción se incubó durante 1 h en la oscuridad a 4°C o a temperatura ambiente. Después de la incubación, la reacción se detuvo mediante la adición de 4 µl de etilenglicol al 10%. Después la mezcla se incubó a temperatura ambiente durante 15 minutos. Después de la precipitación en etanol, el producto se resuspendió en al menos 10 µl de agua o del tampón apropiado, y se dializó contra agua.

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

EJEMPLO 10Ligación de Aldehídos de ARNm a Oligonucleótidos Derivatizados

- El ARNm oxidado se disolvió en un medio ácido, tal como 50 µl de acetato de sodio, pH 4-6. Se añadieron 50 µl de una disolución del oligonucleótido derivatizado 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 37°C o durante la noche (14 h) a 10°C. La mezcla se precipitó a continuación con etanol, se resuspendió en 10 µl o más de agua o del tampón apropiado, y se dializó contra agua destilada. Si se desea, el producto resultante puede analizarse utilizando electroforesis en gel de acrilamida, análisis HPLC, u otras técnicas convencionales.
- Después de la unión del oligonucleótido derivatizado a los ARNm, puede realizarse una reacción de transcripción inversa como se describe en el Ejemplo 11, a continuación.

EJEMPLO 11Transcripción Inversa de ARNm Ligados a Oligonucleótidos Derivatizados

- Un oligodesoxirribonucleótido se derivatizó como sigue. Se disolvieron 3 unidades OD de un oligodesoxirribonucleótido de secuencia 5'ATCAAGAATTCGCACGAGACCATTAA3' (SEQ ID NO: 3) que tiene extremos 5'-OH y 3'-P, en 70 µl de una disolución 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 22°C y después se precipitó dos veces en LiClO₄/acetona. El sedimento se resuspendió en 200 µl de hidrazina 0,25 M y se incubó a 8°C de 3 a 14 h. Después de la reacción con hidrazina, la mezcla se precipitó dos veces en LiClO₄/acetona.

- Los ARN mensajeros que se van a someter a una transcripción inversa se extrajeron de bloques de placenta que tenían lados de 2 cm, que se habían conservado a -80°C. El ARN total se extrajo utilizando técnicas de fenol ácidas convencionales. Se utilizó una cromatografía con oligo-dT para purificar los ARNm. La integridad de los ARNm se comprobó mediante análisis de transferencia Northern.

- Los grupos diol en 7 µg de ARNm de placenta se oxidaron como se describió anteriormente en el Ejemplo 9. El oligonucleótido derivatizado se unió a los ARNm como se describió en el Ejemplo 10 anterior, excepto que la etapa de precipitación se sustituyó por una etapa de cromatografía de exclusión para eliminar los oligodesoxirribonucleótidos derivatizados que no se habían unido a los ARNm. La cromatografía de exclusión se realizó como sigue:

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

- Se introdujo una cuenta de vidrio (diámetro de 3 mm) 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 del extremo 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).

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

- La columna se cargó a continuación con los ARNMs acoplados al oligonucleótido. En cuanto la muestra hubo penetrado se añadió tampón de equilibrio. A continuación se recogieron fracciones de 100 µl. El oligonucleótido derivatizado que no se había unido al ARNm apareció en la fracción 16 y las fracciones posteriores. Por tanto, las fracciones 3 a 15 se reunieron y se precipitaron con etanol.

- Para determinar si el oligonucleótido derivatizado estaba realmente unido al ARNm, una décima parte de las fracciones reunidas se rociaron dos veces sobre una membrana de nailon y se hibridaron con una sonda radiactiva utilizando técnicas convencionales. La sonda marcada con ³²P utilizada en estas hibridaciones era un oligodesoxirribonucleótido de secuencia 5' TAATGGTCTCGTGCGAATTCTTGA7 3' (SEQ ID NO:4) anticomplementaria al oligonucleótido derivatizado. Se observó una señal después de la autorradiografía, indicando que el oligonucleótido derivatizado sí estaba realmente unido al ARNm.

- Con el resto de las nueve décimas partes de los ARNm que habían reaccionado con el oligonucleótido derivatizado se realizó una transcripción inversa como sigue. Se llevó a cabo una reacción de

transcripción inversa con transcriptasa inversa siguiendo las instrucciones del fabricante, y se utilizaron 50 pmoles de nonámeros con secuencias aleatorias como cebadores.

Para asegurarse de que la transcripción inversa se había llevado a cabo a través de la estructura de caperuza se realizaron dos tipos de experimentos.

- 5 En el primer tipo, después de la eliminación del ARN de los heterodúplex de ADNc:ARN obtenidos de la reacción de transcripción inversa mediante una hidrólisis alcalina, una porción de los ADNc monocatenarios resultantes se roció sobre una membrana cargada positivamente y se hibridó, utilizando métodos convencionales, con una sonda marcada con ³²P que tenía una secuencia idéntica a la del oligonucleótido derivatizado. Se incluyeron aplicaciones de control que contenían 1 pmol, 100
10 fmol, 50 fmol, 10 fmol y 1 fmol de un oligodesoxirribonucleótido de control con una secuencia idéntica a la del oligonucleótido derivatizado. La señal observada en las aplicaciones que contenían el ADNc indicó que aproximadamente 15 fmol del oligonucleótido derivatizado había sido transcrito de modo inverso. Estos resultados demuestran que la transcripción inversa puede realizarse a través de la caperuza y, en particular, que la transcriptasa inversa atraviesa el enlace 5'-P-P-P-5' de la caperuza de los ARN mensajeros de eucariotas.
15

- En el segundo tipo de experimento, los ADNc monocatenarios obtenidos a partir de la anterior síntesis de la primera hebra se utilizaron como moldes para reacciones PCR. Se realizaron dos tipos de reacciones. En primer lugar, se realizó una amplificación específica de los ARNm para alfa-globina, deshidrogenasa, pp15 y factor de elongación E4 utilizando las siguientes parejas de cebadores de oligodesoxirribonucleótidos.
20

alfa-globina

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

deshidrogenasa

- 25 3 DH-S: 5' AGT GAT TCC TGC TAC TTT GGA TGG C 3' (SEQ ID NO:7) 3 DH-As: 5' GCT TGG TCT TGT TCT GGA GTT TAG A 3' (SEQ ID NO:8)

pp15

PP15-S: 5' TCC AGA ATG GGA GAC AAG CCA ATT T 3' (SEQ ID NO:9) PP15-As: 5' AGG GAG GAG GAA ACA GCG TGA GTC C 3' (SEQ ID NO:10)

- 30 factor de elongación E4

EFA1-S: 5' ATG GGA AAG GAA AAG ACT CAT ATC A 3' (SEQ ID NO:11) EF1A-As: 5' AGC AGC AAC AAT CAG GAC AGC ACA G 3' (SEQ ID NO:12)

- En segundo lugar, también se realizaron amplificaciones no específicas con los oligodesoxirribonucleótidos antisentido de las parejas descritas anteriormente, y con un cebador derivado de la secuencia del oligodesoxirribonucleótido derivatizado
35 (5'ATCAAGAATTCGCACGAGACCATT3') (SEQ ID NO: 13).

Una vigésima parte de las siguientes muestras del producto de RT-PCR se procesaron en un gel de agarosa al 1,5% y se tiñeron con bromuro de etidio.

- 40 Muestra 1: productos de la reacción PCR utilizando los cebadores de globina de los SEQ ID NO 5 y 6 en presencia de ADNc.

Muestra 2: productos de la reacción PCR utilizando los cebadores de globina de los SEQ ID NO 5 y 6 en ausencia de ADNc añadido.

Muestra 3: productos de la reacción PCR utilizando los cebadores de deshidrogenasa de los SEQ ID NO 7 y 8 en presencia de ADNc.

- 45 Muestra 4: productos de la reacción PCR utilizando los cebadores de deshidrogenasa de los SEQ ID NO 7 y 8 en ausencia de ADNc añadido.

Muestra 5: productos de la reacción PCR utilizando los cebadores de pp15 de los SEQ ID NO 9 y 10 en presencia de ADNc.

- 50 Muestra 6: productos de la reacción PCR utilizando los cebadores de pp15 de los SEQ ID NO 9 y 10 en ausencia de ADNc añadido.

Muestra 7: productos de la reacción PCR utilizando los cebadores de EIF4 de los SEQ ID NO 11 y 12 en presencia de ADNc añadido.

Muestra 8: productos de la reacción PCR utilizando los cebadores de EIF4 de los SEQ ID NO 11 y 12 en ausencia de ADNc añadido.

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

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

- 15 Los anteriores Ejemplos resumen el procedimiento químico para enriquecer ARNm para los que tienen extremos 5' intactos, como se ilustra en la Figura 1. Otros detalles con respecto al procedimiento químico para obtener estos ARNm se describen en la Solicitud Internacional nº WO96/34981, publicada el 7 de Noviembre, 1996. Las estrategias basadas en las anteriores modificaciones químicas en la estructura de caperuza 5' pueden utilizarse para generar ADNc seleccionados para que incluyan los extremos 5' de los ARNm de los cuales derivan. En una versión de estos procedimientos, los extremos 5' de los ARNm se modifican como se describió anteriormente. Después se realiza una reacción de transcripción inversa para extender un cebador complementario al extremo 5' del ARNm.
- 20 Los ARN 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 pueden capturarse sobre una fase sólida revestida con una molécula capaz de interaccionar con la molécula utilizada para derivatizar el extremo 5' del ARNm. Después, las hebras de los heterodúplex se separan para recuperar las primeras hebras del ADNc monocatenario que incluyen el extremo 5' del ARNm. A continuación puede realizarse la síntesis de la segunda hebra del ADNc utilizando técnicas convencionales. Por ejemplo, los procedimientos descritos en el documento WO 96/34981, o en Carnici et al., Genomics, 37: 327-336, 1996, pueden emplearse para seleccionar ADNc que incluyen la secuencia derivada del extremo 5' de la secuencia codificante del ARNm.

- 30 Después de la ligación de la etiqueta de oligonucleótido a la caperuza 5' del ARNm se realiza una reacción de transcripción inversa para extender un cebador complementario al ARNm en el extremo 5' del ARNm. Después de la eliminación del componente de ARN del heterodúplex resultante utilizando técnicas convencionales, se realiza la síntesis de la segunda hebra del ADNc con un cebador complementario al marcador de oligonucleótido.

2. Métodos Enzimáticos para Obtener ARNm que Tienen Extremos 5' Intactos

- 35 Otras técnicas para seleccionar ADNc que se extiende hasta el extremo 5' del ARNm del que deriva son totalmente enzimáticas. Algunas versiones de estas técnicas son descritas por Dumas Milne Edwards, J.B. (Tesis doctoral de la Universidad de París VI, Le clonage des ADNc complets: difficultés et perspectives nouvelles. Apports pour l'étude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 de Diciembre, 1993), documento EPO 625572, y Kato et al., Gene, 150: 243-250, 1994.

- 40 Brevemente, en estas técnicas, el ARNm aislado se trata con fosfatasa alcalina para eliminar los grupos fosfato presentes en los extremos 5' de los ARNm incompletos sin caperuza. Después de este procedimiento, la caperuza presente en los ARNm completos se elimina enzimáticamente con una enzima de eliminación de caperuza, tal como polinucleótido quinasa de T4 o 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 acopla a continuación al fosfato presente en el extremo 5' del ARNm sin caperuza utilizando ARN ligasa 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.

50 **EJEMPLO 12**

Método Enzimático para Obtener ESTs 5'

- 55 Se desfosforilaron 20 microgramos de ARN poliA+ utilizando fosfatasa de intestino de ternera (Biolabs). Después de una extracción en fenol y cloroformo, la estructura de caperuza del ARNm se hidrolizó utilizando la pirofosfatasa ácida de tabaco (purificada como se describe en Shinshi et al., Biochemistry, 15: 2185-2190, 1976), y un hemioligonucleótido 5'ADN/ARN-3' que tiene un extremo 5' no fosforilado, un tramo de adenosina ribofosfato en el extremo 3', y un sitio EcoRI cerca del extremo 5', se ligó a los extremos 5'P del ARNm utilizando la ARN ligasa de T4 (Biolabs). Los oligonucleótidos adecuados para su uso en este procedimiento tienen una longitud de preferiblemente 30 a 50 bases.

Los oligonucleótidos que tienen un extremo 5' no fosforilado pueden sintetizarse añadiendo un fluorocromo al extremo 5'. La inclusión de un tramo de adenosina ribofosfatos en el extremo 3' del oligonucleótido aumenta la eficacia de la ligación. Se apreciará que el oligonucleótido puede contener sitios de clonación diferentes de EcoRI.

- 5 Después de la ligación del oligonucleótido al fosfato presente en el extremo 5' del ARNm sin caperuza, se realiza la síntesis de la primera y segunda hebra del ARNc utilizando métodos convencionales, o los indicados en el documento EP 0.625.572, y Kato et al., supra, y Dumas Milne Edwards, supra.

- 10 El ADNc resultante puede ligarse a continuación a vectores como los descritos por Kato et al., supra, u otros vectores de ácido nucleico conocidos por los expertos en la técnica utilizando técnicas como las descritas por Sambrook et al., Molecular Cloning: A Laboratory Manual, 2ª ed., Cold Spring Harbor Laboratory Press, 1989.

II. Obtención y Caracterización de las EST 5' de la Presente Invención

- 15 Las ESTs 5' descritas en la presente memoria se obtuvieron utilizando los métodos químicos y enzimáticos descritos anteriormente para enriquecer ARNm para los que tienen extremos 5' intactos, como se describe a continuación.

1. Obtención de ESTs 5' Utilizando ARNm con Extremos 5' Intactos

En primer lugar, se prepararon los ARNm como se describe en el Ejemplo 13, a continuación.

EJEMPLO 13

Preparación de ARNm con Extremos 5' Intactos

- 20 Se adquirieron ARNs totales humanos o ARNs poliA+ derivados de 29 tejidos diferentes en, respectivamente, LABIMO y CLONTECH, y se utilizaron para generar 44 bancos de ADNc como sigue. El ARN adquirido se había aislado de células o tejidos utilizando extracción ácida de tiocianato de guanidio-fenol-cloroformo (Chomczynski y Sacchi, Analytical Biochemistry, 162: 156-159, 1987). El ARN poliA+ se aisló a partir de ARN total (LABIMO) mediante dos pases de cromatografía de oligo-dT, como describen Aviv y Leder, Proc. Natl. Acad. Sci. USA, 69: 1408- 1412, 1972, para eliminar el ARN ribosómico.

- 30 Se comprobó la calidad e integridad de los ARNs poliA+. Se utilizaron análisis de transferencia Northern hibridados con una sonda de globina para confirmar que los ARNms no estaban degradados. Se comprobó la contaminación de los ARNm poliA+ por secuencias ribosómicas utilizando análisis de transferencia Northern y una sonda derivada de la secuencia del ARNr 28S. Se usaron preparaciones de ARNm con menos de 5% de ARNr en la construcción de los bancos. Para evitar construir bancos con ARNs contaminados por secuencias exógenas (procarióticas o fúngicas), se estudió la presencia de secuencias ribosómicas bacterianas 16S o de dos ARNms fúngicos de alta expresión, utilizando PCR.

- 35 Después de la preparación de los ARNms, se emplearon los procedimientos químicos y/o enzimáticos descritos anteriormente para enriquecer ARNms para los que tienen extremos 5' intactos, para obtener ESTs 5' de diversos tejidos. En ambos procedimientos, se unió una etiqueta de oligonucleótido a los extremos 5' de los ARNms. La etiqueta de oligonucleótido tenía un sitio EcoRI para facilitar los procedimientos de clonación posteriores. Para facilitar el procesamiento del ADNc monocatenario y bicatenario obtenido en la construcción de los bancos, se utilizó la misma secuencia nucleotídica para diseñar el oligonucleótido acoplado en ambos procedimientos químico y enzimático. No obstante, en el procedimiento químico, la etiqueta utilizada fue un oligodesoxirribonucleótido que estaba unido a la caperuza del ARNm, mientras que en la ligación enzimática, la etiqueta fue un hemioligonucleótido 5'ADN/ARN3' quimérico que se ligó al extremo 5' del ARNm sin caperuza como se describe en el Ejemplo 12.

Después de la unión de la etiqueta de oligonucleótido al ARNm mediante el método químico o enzimático, se estudió la integridad del ARNm mediante un análisis de transferencia Northern con 200 a 500 ng de ARNm utilizando una sonda complementaria a la etiqueta de oligonucleótido antes de realizar la síntesis de la primera hebra, como se describe en el Ejemplo 14.

50 **EJEMPLO 14**

Síntesis de ADNc Utilizando Moldes de ARNm que Tienen Extremos 5' Intactos

- 55 Para los ARNms unidos a las etiquetas de oligonucleótidos utilizando los métodos químico y enzimático, se realizó la síntesis de la primera hebra de ADNc utilizando la transcriptasa inversa Superscript II (Gibco BRL) o Rnase H Minus M-MLV (Promega) con nonámeros aleatorios como cebadores. Para proteger los sitios EcoRI internos en el ADNc de la digestión en etapas posteriores

del procedimiento, se utilizó dCTP metilado para la síntesis de la primera hebra. Después de la eliminación del ARN mediante una hidrólisis alcalina, la primera hebra del ADNc se precipitó utilizando isopropanol para eliminar los cebadores residuales.

5 Para los métodos químico y enzimático, la segunda hebra del ADNc se sintetizó con un fragmento de Klenow utilizando un cebador correspondiente al extremo 5' del oligonucleótido ligado descrito en el Ejemplo 12. Preferiblemente, el cebador tiene una longitud de 20-25 bases. También se utilizó dCTP metilado para la síntesis de la segunda hebra para proteger los sitios EcoRI internos en el ADNc de la digestión durante el proceso de clonación.

10 Después de la síntesis del ADNc, los ADNc se clonaron en pBlueScript como se describe en el Ejemplo 15, a continuación.

EJEMPLO 15

Clonación de ADNc derivados de ARNm con extremos 5' intactos en BlueScript

15 Después de la síntesis de la segunda hebra, se crearon extremos romos en el ADNc con ARN polimerasa de T4 (Biolabs), y el ADNc se digirió con EcoRI. Puesto que se utilizó dCTP metilado durante la síntesis de ADNc, el sitio EcoRI presente en la etiqueta era el único sitio hemimetilado y, por tanto, es el único sitio susceptible a la digestión con EcoRI. El ADNc se fraccionó a continuación según su tamaño utilizando cromatografía de exclusión (AcA, Biosepra), y las fracciones que correspondían a ADNc de más de 150 pb se reunieron y se precipitaron con etanol. El ADNc se clonó 20 direccionalmente en los extremos Smal y EcoRI del vector fagémico pBlueScript (Stratagene). La mezcla de ligación se sometió a electroporación en bacterias y se propagó con una selección con antibióticos apropiada.

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

EJEMPLO 16

25 Selección de Clones que Tienen la Etiqueta de Oligonucleótidos Unida

Los bancos de EST 5' que contienen ADN de plásmidos fabricados como se describió anteriormente se purificaron (Qiagen). Se realizó una selección positiva de los clones etiquetados como sigue. Brevemente, en este procedimiento de selección, el ADN del plásmido se convirtió en un ADN 30 monocatenario utilizando endonucleasa del gen II del fago F1, en combinación con una exonucleasa (Chang et al., Gene, 127: 95-98, 1993), tal como exonucleasa III o exonucleasa del gen 6 de T7. El ADN monocatenario resultante se purificó a continuación utilizando cuentas paramagnéticas como describen Fry et al., Biotechniques, 13: 124-131, 1992. En este procedimiento, el ADN monocatenario se hibridó con un oligonucleótido biotinilado que tenía una secuencia que se correspondía con el 35 extremo 3' de los oligonucleótidos descritos en el Ejemplo 13. Preferiblemente, el cebador tiene una longitud de 20-25 bases. Los clones que incluyen una secuencia complementaria al oligonucleótido biotinilado se capturaron mediante incubación con cuentas magnéticas revestidas con estreptavidina, seguido de selección magnética. Después de la captura de los clones positivos, el ADN del plásmido se liberó de las cuentas magnéticas y se convirtió en ADN bicatenario utilizando ADN polimerasa, tal como la ThermoSequenase obtenida de Amersham Pharmacia Biotech. Como alternativa, pueden 40 utilizarse protocolos como el descrito en el kit Gene Trapper, disponible en Gibco BRL. El ADN bicatenario se electroporó a continuación en bacterias. Se estimó que el porcentaje de clones positivos que tenían la etiqueta de oligonucleótido 5' variaba, de forma típica, entre 90% y 98%, utilizando análisis de transferencia puntual.

45 Después de la electroporación, los bancos se ordenaron en placas de microtitulación (MTP) de 384 pocillos. Se conservó una copia de MTP para necesidades futuras. Después los bancos se trasladaron a MTP de 96 pocillos, y se secuenciaron como se describe a continuación.

EJEMPLO 17

Secuenciación de los Insertos en Clones Seleccionados

50 Los insertos de los plásmidos primero se amplificaron mediante PCR en termocicladores PE 9600 (Perkin-Elmer, Applied Biosystems División, Foster City, CA), utilizando cebadores convencionales SETA-A y SETA-B (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTP (Boehringer), tampón y condiciones de ciclación como las recomendadas por Perkin-Elmer Corporation.

Los productos de PCR se secuenciaron después utilizando secuenciadores automáticos ABI Prism 377 (Perkin-Elmer). Las reacciones de secuenciación se realizaron utilizando termocicladores PE 9600 55 con química convencional de colorante-cebador y ThermoSequenase (Amersham Pharmacia Biotech). Los cebadores utilizados fueron T7 o 21M13 (disponible en Genset SA), según fuera apropiado. Los

cebadores se marcaron con los colorantes JOE, FAM, ROX y TAMRA. Los dNTP y ddNTP utilizados en las reacciones de secuenciación se obtuvieron de Boehringer. El tampón de secuenciación, las concentraciones de los reactivos y las condiciones de ciclación fueron los recomendados por Amersham.

- 5 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 de acrilamida al 4% convencional. La electroforesis se realizó durante 2,5 horas a 3000 V en un secuenciador ABI 377, y los datos de las secuencias se recogieron y analizaron utilizando el programa informático de análisis de secuenciación de ADN ABI Prism, versión 2.1.2.

10 2. Análisis por Ordenador de las EST 5' Obtenidas: Construcción de bases de datos NetGene y SignalTag

- 15 Los datos de las secuencias de los 44 bancos de ADNc fabricados como se describió anteriormente se trasladaron a una base de datos patentada, en la que se realizaron las etapas de control de calidad y validación. Un programa de análisis de bases ("base-caller") patentado, que trabaja utilizando el sistema Unix, marca de forma automática los picos sospechosos, tomando en cuenta la forma de los picos, la resolución interpicos, y el nivel de ruido. Este programa de análisis de bases patentado también realizó un recorte automático. Cualquier tramo de 25 o menos bases que tuviera más de 4 picos sospechosos no se consideró fiable y se rechazó. Las secuencias que se corresponden con el vector de clonación o los oligonucleótidos de ligación se retiraron automáticamente de las secuencias EST. Sin embargo, las secuencias EST resultantes pueden contener de 1 a 5 bases que pertenecen a las secuencias mencionadas anteriormente en su extremo 5'. Si fuera necesario, éstas pueden eliminarse con facilidad basándose en el caso concreto.

- 25 Después de realizar la secuenciación como se describió anteriormente, las secuencias de las ESTs 5' se introdujeron en NetGene®, una base de datos patentada utilizada para el almacenamiento y manipulación, como se describe a continuación. Los expertos en la técnica apreciarán que los datos pueden almacenarse y manipularse en cualquier medio que pueda leerse y accederse mediante un ordenador. Los medios de lectura por ordenador incluyen medios de lectura magnéticos, ópticos o electrónicos. Por ejemplo, el medio de lectura por ordenador puede 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.

- 30 Además, los datos de las secuencias pueden almacenarse y manipularse en una diversidad de programas procesadores de datos en una diversidad de formatos. Por ejemplo, los datos de las secuencias pueden almacenarse como un texto en un archivo de procesamiento de textos, tal como Microsoft WORD y WORDPERFECT, o como un archivo ASCII en una diversidad de programas de bases de datos conocidos por los expertos en la técnica, tales como DB2, SYBASE u ORACLE.

- 40 Los medios de lectura por ordenador en los que puede almacenarse la información de las secuencias pueden ser un ordenador personal, una red, un servidor u otros sistemas de ordenador conocidos por los expertos en la técnica. El ordenador u otro sistema incluyen preferiblemente el medio de almacenamiento descrito anteriormente, y un procesador para acceder y manipular los datos de las secuencias. Cuando los datos de la secuencia se han almacenado, pueden manipularse y seleccionarse para localizar las secuencias almacenadas que contienen una secuencia de ácido nucleico deseada o que codifican una proteína que tiene un dominio funcional particular. Por ejemplo, la información almacenada de las secuencias puede compararse con otras secuencias conocidas para identificar homologías, motivos implicados en una función biológica, o motivos estructurales.

- 45 Los programas que pueden utilizarse para seleccionar o comparar las secuencias almacenadas incluyen la serie de programas MacPattern (EMBL), BLAST y BLAST2 (NCBI), programas de herramientas de selección de alineación local básica para comparaciones de nucleótidos (BLASTN) y péptidos (BLASTX) (Altschul et al., J. Mol. Biol., 215: 403, 1990), y FASTA (Pearson y Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444, 1988). Los programas BLAST extienden a continuación los alineamientos en base a criterios definidos de apareamiento y desapareamiento.

- 50 Los motivos que pueden detectarse utilizando los anteriores programas y los descritos en el Ejemplo 28 incluyen secuencias que codifican cremalleras de leucina, motivos de hélice-giro-hélice, sitios de glicosilación, sitios de ubiquitinación, hélices alfa y láminas beta, secuencias señal que codifican péptidos señal que dirigen la secreción de las proteínas codificadas, secuencias implicadas en la regulación de la transcripción, tales como homeodominios, tramos ácidos, sitios enzimáticos activos, sitios de unión al sustrato, y sitios de ruptura enzimática.

55 Antes de seleccionar los ADNc en la base de datos NetGene® para detectar motivos de secuencias de interés, se identificaron los ADNc derivados de los ARNm que no eran de interés y se eliminaron de posteriores consideraciones, como se describe en el Ejemplo 18, a continuación.

EJEMPLO 18Eliminación de las Secuencias no Deseadas de Posterior Consideración

Las ESTs 5' de la base de datos NetGene® que derivaban de secuencias no deseadas, tales como ARN de transferencia, ARN ribosómicos, ARN mitocondriales, ARN procarióticos, ARN fúngicos, secuencias Alu, secuencias L1, o secuencias repetidas, se identificaron utilizando los programas informáticos FASTA y BLASTN con los parámetros listados en la Tabla 1.

Para eliminar las ESTs 5' que codifican ARNt de posterior consideración, las secuencias de EST 5' se compararon con las secuencias de 1190 ARNt conocidos obtenidos de la descarga 38 de EMBL, de las cuales 100 eran humanas. La comparación se realizó utilizando FASTA en ambas hebras de las ESTs 5'. Las secuencias que tienen más de 80% de homología en más de 60 nucleótidos se identificaron como ARNt. De las 144.341 secuencias seleccionadas, 26 se identificaron como ARNt y se eliminaron de posterior consideración.

Para eliminar las ESTs 5' que codifican ARNr de posterior consideración, las secuencias ESTs 5' se compararon con las secuencias de 2497 ARNr conocidos obtenidos de la descarga 38 de EMBL, de las cuales 73 eran humanas. La comparación se realizó utilizando BLASTN en ambas hebras de las ESTs 5' con el parámetro S = 108. Las secuencias que tienen más de 80% de homología en tramos mayores de 40 nucleótidos se identificaron como ARNr. De las 144.341 secuencias seleccionadas, 3.312 se identificaron como ARNr y se eliminaron de posterior consideración.

Para eliminar las ESTs 5' que codifican ARNmt de posterior consideración, las secuencias ESTs 5' se compararon con las secuencias de dos genomas mitocondriales conocidos, para los cuales están disponibles las secuencias genómicas completas, y todas las secuencias transcritas a partir de estos genomas mitocondriales, incluyendo ARNt, ARNr y ARNm para un total de 38 secuencias. La comparación se realizó utilizando BLASTN en ambas hebras de las ESTs 5' con el parámetro S = 108. Las secuencias que tienen más de 80% de homología en tramos mayores de 40 nucleótidos se identificaron como ARNmt. De las 144.341 secuencias seleccionadas, 6.110 se identificaron como ARNmt y se eliminaron de posterior consideración.

Las secuencias que pueden haber surgido de contaminantes exógenos se eliminaron de posterior consideración comparando las secuencias de ESTs 5' con la descarga 46 de las divisiones bacterianas y fúngicas de EMBL, utilizando BLASTN con el parámetro S = 144. Todas las secuencias que tienen más de 90% de homología en al menos 40 nucleótidos se identificaron como contaminantes exógenos. De los 42 bancos de ADNc estudiados, el porcentaje medio de secuencias procarióticas y fúngicas que contienen fue 0,2% y 0,5%, respectivamente. Entre estas secuencias, sólo una pudo identificarse como una secuencia específica de hongos. Las otras eran secuencias fúngicas o procarióticas que tenían homologías con secuencias de vertebrados o que incluían secuencias repetidas que no se habían enmascarado durante la comparación electrónica.

Además, las ESTs 5' se compararon con 6093 secuencias Alu y 1115 secuencias L1 para enmascarar las ESTs 5' que contenían estas secuencias repetidas. Las ESTs 5' que incluían repeticiones THE y MER, secuencias SSTR o satélites, microsatélites o repeticiones teloméricas también se eliminaron de posterior consideración. De media, 11,5% de las secuencias en los bancos contenían secuencias repetidas. De este 11,5%, 7% contenía repeticiones Alu, 3,3% contenía repeticiones L1, y 1,2% restante derivaba de los otros tipos seleccionados de secuencias repetidas. Estos porcentajes son coherentes con los descubiertos en los bancos de ADNc preparados por otros grupos. Por ejemplo, los bancos de ADNc de Adams et al. contienen entre 0% y 7,4% de repeticiones Alu, dependiendo de la fuente de ARN que se utilizó para preparar el banco de ADNc (Adams et al., Nature, 377: 174, 1996).

Las secuencias de las ESTs 5' que permanecieron después de la eliminación de las secuencias no deseables se compararon con las secuencias de ARNm humanos conocidos para determinar la precisión de los procedimientos de secuenciación descritos anteriormente.

EJEMPLO 19Medida de la Precisión de la Secuenciación Mediante Comparación con Secuencias Conocidas

Para determinar aún más la precisión del procedimiento de secuenciación descrito anteriormente, las secuencias de ESTs 5' derivadas de secuencias conocidas se identificaron y se compararon con las secuencias conocidas originales. En primer lugar, se realizó en las ESTs 5' un análisis FASTA con proyecciones más cortas que 5 pb en ambos extremos, para identificar los que se apareaban con una entrada en la base de datos de ARNm humana pública. Las 6655 ESTs 5' que se apareaban con un ARNm humano conocido se realinearon a continuación con su ARNm cognado y se utilizó una programación dinámica para incluir sustituciones, inserciones y deleciones en la lista de "errores" que serían reconocidos. Los errores que se producían en las últimas 10 bases de las secuencias ESTs 5'

se ignoraron para evitar la inclusión de sitios de clonación espúreos en el análisis de la precisión de la secuenciación.

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

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

EJEMPLO 20

Determinación de la Eficacia de la Selección de EST 5'

- 10 Para determinar la eficacia con la que los anteriores procedimientos de selección aíslan ESTs 5' que incluyen los extremos 5' de los ARNm de los que derivan, las secuencias de los extremos de las ESTs 5' derivadas de los genes de la subunidad α del factor de elongación I y 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, pueden utilizarse para determinar el porcentaje de ESTs 5' derivadas que incluye los sitios de inicio de la transcripción auténticos.

15 Para ambos genes, más de 95% de las ESTs 5' obtenidas incluían realmente secuencias cercanas o aguas arriba del extremo 5' de los correspondientes ARNm.

- 20 Para extender el análisis de la fiabilidad de los procedimientos para aislar ESTs 5' a partir de ESTs en la base de datos NetGene®, se realizó un análisis similar utilizando una base de datos compuesta de secuencias de ARNm humano extraídas de la descarga 97 de la base de datos GenBank como comparación. Los extremos 5' de más de 85% de las ESTs 5' derivados de los ARNm incluidos en la base de datos GeneBank estaban localizados cerca de los extremos 5' de la secuencia conocida. Como algunas de las secuencias de ARNm disponibles en la base de datos GenBank se deducen a partir de secuencias genómicas, un apareamiento en el extremo 5' con estas secuencias se contará como un apareamiento interno. Por tanto, el método utilizado aquí subestima el rendimiento de ESTs que incluyen los extremos 5' auténticos de sus correspondientes ARNm.

- 25 Los bancos de EST fabricados anteriormente incluyen múltiples ESTs 5' derivadas del mismo ARNm. Las secuencias de estas ESTs 5' se compararon entre sí, y se identificaron las ESTs 5' más largas para cada ARNm. Los ADNc solapantes se ensamblaron en secuencias continuas (cóntigos). Las secuencias continuas resultantes se compararon después con bases de datos públicas para calibrar su similitud con secuencias conocidas, como se describe en el Ejemplo 21, a continuación.

EJEMPLO 21

Agrupación de las ESTs 5' y Cálculo de los Índices de Novedad en Bancos de ADNc

- 35 Para cada banco de EST secuenciado, las secuencias se agruparon por el extremo 5'. Cada secuencia en el banco se comparó con las otras con BLASTN2 (hebra directa, parámetros S = 107). Las ESTs con los pares de segmentos de alta puntuación (HSP, "High Scoring Segment Pairs") de al menos 25 pb de longitud, que tienen 95% de bases idénticas y que empiezan más cerca que 10 pb de cada extremo 5' de las ESTs se agruparon. Se utilizó la secuencia más larga que se encontró en la agrupación como representativa del grupo. A continuación se realizó una agrupación global entre bancos, lo cual conduce a la definición de supercóntigos.

- 40 Para evaluar el rendimiento de las nuevas secuencias dentro de los bancos de EST, se definió una relación de novedad (NR) como: $NR = 100 \times (\text{número de secuencias exclusivas nuevas descubiertas en el banco} / \text{Número total de secuencias del banco})$. De forma típica, la relación de novedad variaba entre 10% y 41%, dependiendo del tejido a partir del cual se obtuvo el banco de EST. Para la mayoría de los bancos, se buscó la secuenciación aleatoria de los bancos de EST 5' hasta que la relación de novedad alcanzó 20%.

Después de la caracterización como se describió anteriormente, la colección de ESTs 5' en NetGene® se seleccionó para identificar aquellas ESTs 5' que portan secuencias señal potenciales, como se describe en el Ejemplo 22, a continuación.

- 50 **EJEMPLO 22**

Identificación de Secuencias Señal Potenciales en EST 5'

Las ESTs 5' en la base de datos NetGene® se seleccionaron para identificar aquellas que tienen un marco de lectura abierto (ORF, "open reading frame") ininterrumpidos más largos que 45 nucleótidos, que empieza con un codón ATG y se extiende hasta el final de la EST. Aproximadamente la mitad de

las secuencias de ADNc en NetGene® contenían este ORF. Los ORF de estas ESTs 5' se buscaron después para identificar motivos señal potenciales utilizando ligeras modificaciones de los procedimientos descritos por Von Heijne, Nucleic Acids Res., 14: 4683-4690, 1986.

- 5 Se consideró que aquellas secuencias de EST 5' que codificaban un tramo 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 poseían una secuencia señal. Aquellas EST 5' que se aparearon con una secuencia EST o ARNm humana conocida, y tenían un extremo 5' más de 20 nucleótidos aguas abajo del extremo 5' conocido se excluyeron del posterior análisis. El resto de los ADNc que tenían secuencias señal en su interior se incluyeron en una base de datos denominada SignalTag®.
- 10 Para confirmar la precisión del método anterior para identificar secuencias señal se realizó el análisis del Ejemplo 23.

EJEMPLO 23

Confirmación de la Precisión de la Identificación de Secuencias Señal Potenciales en ESTs 5'

- 15 Se evaluó la precisión del procedimiento anterior para identificar secuencias señal que codificaban péptidos señal aplicando el método a los 43 aminoácidos localizados en el extremo N de todas las proteínas SwissProt humanas. La puntuación de Von Heijne informatizada para cada proteína se comparó con la caracterización conocida de la proteína como proteína secretada o proteína no secretada. De esta manera, se pudo calcular el número de proteínas no secretadas que tienen una puntuación mayor que 3,5 (falsos positivos) y el número de proteínas secretadas que tienen una
- 20 puntuación menor que 3,5 (falsos negativos).

- Utilizando los resultados del anterior análisis, 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 la suposición de que 10% de las proteínas humanas son secretadas, o la suposición de que 20% de las proteínas humanas son secretadas. Los resultados de este análisis
- 25 se muestran en la Figura 2 y la Tabla IV.

- Utilizando el método anterior para la identificación de proteínas de secreción, se obtuvieron las ESTs 5' de los siguientes polipéptidos que se sabe que son secretados: glucagón humano, precursor de monoquinas inducido por interferón gamma, proteína de tipo ciclofilina secretada, pleiotropina humana, y precursor de biotinidasa humana. Por tanto, el método anterior identifica con éxito aquellas
- 30 EST 5' que codifican un péptido señal.

- Para confirmar que el péptido señal codificado por las ESTs 5' actúa realmente con un péptido señal, las secuencias señal de las ESTs 5' pueden clonarse en un vector diseñado para la identificación de péptidos señal. Estos vectores se diseñan para conferir la capacidad de crecer en un medio selectivo sólo a las células anfitrionas que contienen un vector con una secuencia señal conectada operablemente. Por ejemplo, para confirmar que una EST 5' codifica un péptido señal genuino, la
- 35 secuencia señal de la EST 5' puede insertarse aguas arriba y dentro del marco con una forma no secretada del gen de invertasa de levadura en los vectores de selección de péptidos señal, como los descritos en la Patente de EE.UU. n° 5.536.637. El crecimiento de las células anfitrionas que contienen los vectores de selección de secuencias señal con la secuencia señal de la EST 5' insertada de forma correcta confirma que la EST 5' codifica un péptido señal genuino.
- 40

- Como alternativa, la presencia de un péptido señal puede confirmarse clonando los ADNc extendidos obtenidos utilizando las ESTs en vectores de expresión, tales como pXT1 (como se describe a continuación en el Ejemplo 30), o construyendo vectores de promotor-secuencia señal-gen indicador que codifican proteínas de fusión entre el péptido señal y una proteína indicadora analizable. Después
- 45 de la introducción de estos vectores en una célula anfitriona adecuada, tal como células COS o células NIH 3T3, el medio de crecimiento puede recolectarse y analizarse en busca de la presencia de la proteína secretada. El medio procedente de estas células se compara con el medio procedente de células control que contienen vectores que carecen de la secuencia señal o el inserto de ADNc extendido, para identificar vectores que codifican un péptido señal funcional o una proteína secretada auténtica.
- 50

Aquellas ESTs 5' que codifican un péptido señal, según se determina mediante el método del Ejemplo 22 anterior, se agruparon después en cuatro categorías basadas en su homología con secuencias conocidas, como se describe en el Ejemplo 24, a continuación.

EJEMPLO 24

- 55 Categorización de las EST 5' que Codifican un Péptido Señal

Aquellas ESTs 5' que tienen una secuencia que no se aparea con ninguna secuencia de vertebrado conocida ni con ninguna secuencia EST disponible públicamente se denominaron "nuevas". De las

secuencias de la base de datos SignalTag®, 947 de las ESTs 5' que tenían una puntuación de Von Heijne de al menos 3,5 se encontraban dentro de esta categoría.

5 Aquellas ESTs 5' que tenían una secuencia que no se apareaba con ninguna secuencia de vertebrado pero sí se apareaba con una EST conocida públicamente se denominaron "EST-ext", con la condición de que la secuencia de EST conocida se extendiera al menos 40 nucleótidos en la dirección 5'. De las secuencias de la base de datos SignalTag®, 150 de las ESTs 5' que tenían una puntuación de Von Heijne de al menos 3,5 se encontraban dentro de esta categoría.

10 Aquellas ESTs 5' que no se apareaban con ninguna secuencia de vertebrado pero sí se apareaban con una EST conocida públicamente sin extender la EST conocida en al menos 40 nucleótidos en la dirección 5' se denominaron "EST". De 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 se encontraban dentro de esta categoría.

15 Aquellas ESTs 5' que se apareaban con una secuencia de ARNm humano pero extendían la secuencia conocida en al menos 40 nucleótidos en la dirección 5' se denominaron "VERT-ext". De las secuencias de la base de datos SignalTag®, 23 de las ESTs 5' que tenían una puntuación de Von Heijne de al menos 3,5 se encontraban dentro de esta categoría. Se incluyó en esta categoría una EST 5' que extendía la secuencia conocida del ARNm de translocasa humana en más de 200 bases en la dirección 5'. También se identificó una EST 5' que extendía la secuencia de un gen supresor de tumor humano en la dirección 5'.

20 La Tabla V muestra la distribución de las ESTs 5' en cada categoría, y el número de ESTs 5' en cada categoría que tienen una puntuación de Von Heijne mínima dada.

3. Evaluación de la Expresión Espacial y Temporal de ARNm que Corresponden a las ESTs 5' o ADNc Extendidos

25 Cada una de las ESTs 5' también se categorizó basándose en el tejido del cual se obtuvo el correspondiente ARNm, como se describe a continuación en el Ejemplo 25.

EJEMPLO 25

Categorización de Patrones de Expresión

La Tabla VI muestra la distribución de las ESTs 5' en cada una de las categorías definidas anteriormente con respecto al tejido del cual se obtienen las ESTs 5' de los correspondientes ARNm.

30 La Tabla II proporciona los números de identificación de secuencia de secuencias EST 5' derivadas de testículo y otros tejidos, las categorías en las que se incluyen estas secuencias, y la puntuación de Von Heijne de los péptidos señal que codifican. Las secuencias EST 5' y las secuencias de aminoácidos que codifican se proporcionan en los listados de secuencia adjuntos. La Tabla III proporciona los números de identificación de secuencia de las ESTs 5' y las secuencias de los péptidos señal que codifican. Las secuencias de las ESTs 5' y los polipéptidos que codifican aparecen en el listado de secuencias adjunto.

40 La secuencia de ADN SEQ ID NO: 73+75 puede seleccionarse con facilidad para detectar cualquier error que presente, y cualquier ambigüedad de la secuencia puede resolverse mediante resecuenciación de un fragmento que contenga tales errores o ambigüedades en ambas cadenas. Estos fragmentos pueden obtenerse a partir de los plásmidos conservados en el laboratorio de los autores de la presente invención, o pueden aislarse utilizando las técnicas descritas en la presente memoria. La resolución de estas ambigüedades o errores puede verse facilitada mediante la utilización de cebadores que se hibridan con secuencias colocadas cerca de las secuencias ambiguas o erróneas. Por ejemplo, los cebadores pueden hibridarse con secuencias dentro de 50-75 bases de la ambigüedad o error. Tras la resolución del error o ambigüedad, pueden realizarse las correspondientes correcciones en las secuencias de proteínas codificadas por el ADN que contiene el error o ambigüedad.

50 Además de categorizar las ESTs 5' con respecto a su tejido de origen, los patrones de expresión espacial y temporal de los ARNm correspondientes a las ESTs 5', así como sus niveles de expresión, pueden determinarse como se describe en el Ejemplo 26, a continuación. La caracterización de los patrones de expresión espacial y temporal y los niveles de expresión de estos ARNm es útil para construir vectores de expresión capaces de producir un nivel deseado de un producto génico de una manera espacial o temporal deseada, como se analizará con más detalle a continuación.

55 Además, las ESTs 5' cuyos correspondientes ARNm se asocian con estados de enfermedad también pueden identificarse. Por ejemplo, una enfermedad particular puede ser el resultado de la falta de expresión, la sobreexpresión, o la infraexpresión de un ARNm que se corresponde con una EST 5'. Comparando los patrones de expresión de ARNm y las cantidades en las muestras tomadas de

individuos sanos, con los de aquellos individuos que padecen una enfermedad particular, pueden identificarse las ESTs 5' responsables de la enfermedad.

- 5 Se apreciará que los resultados del anterior procedimiento de caracterización para ESTs 5' también se aplican a ADNc extendidos (que pueden obtenerse como se describe a continuación) que contienen secuencias adyacentes a las ESTs 5'. También se apreciará que, si se desea, la caracterización puede retrasarse hasta que se hayan obtenido los ADNc extendidos, en lugar de caracterizar las ESTs en sí mismas.

EJEMPLO 26

10 Evaluación de los Niveles de Expresión y Patrones de ARNm que Corresponden a ESTs 5' o ADNc Extendidos

- 15 Los niveles de expresión y patrones de ARNm que corresponden a ESTs 5' o ADNc extendidos (que pueden obtenerse como se describe a continuación en el Ejemplo 27) pueden analizarse mediante hibridación en disolución con sondas largas como se describe en la solicitud de Patente Internacional Núm. WO 97/05277. Brevemente, una EST 5', un ADNc extendido o un fragmento de éstos que se corresponde con el gen que codifica el ARNm que se va a caracterizar se inserta en un sitio de clonación inmediatamente aguas abajo de un promotor de ARN polimerasa de bacteriófago (T3, T7 o SP6) para producir ARN antisentido. Preferiblemente, la EST 5' o ADNc extendido tiene 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 doblemente marcado se hibrida en disolución con ARNm aislado de células o tejidos de interés. Las hibridaciones se realizan bajo condiciones rigurosas convencionales (40-50°C durante 16 horas en formamida al 80%, tampón NaCl 0,4 M, pH 7-8). La sonda no hibridada se retira mediante digestión con ribonucleasas específicas de ARN monocatenario (es decir, ARNasas CL3, T1, Phy M, U2 o A). La presencia de la modificación con biotina-UTP permite la captura del híbrido sobre una placa de microtitulación revestida con estreptavidina. La presencia de la modificación con DIG permite detectar el híbrido y cuantificarlo mediante ELISA utilizando un anticuerpo anti-DIG acoplado a fosfatasa alcalina.

- 30 Las ESTs 5', ADNc extendidos o fragmentos de los mismos también pueden etiquetarse con secuencias de nucleótidos para el análisis en serie de la expresión génica (SAGE), según se describe en la Solicitud de Patente del Reino Unido Núm. 2305241 A. En este método, los ADNcs se preparan a partir de una célula, tejido, organismo u otra fuente de ácido nucleico para la cual deben determinarse los patrones de expresión. Los ADNcs resultantes se separan en dos grupos. Los ADNcs de cada grupo se rompen con una primera endonucleasa de restricción, denominada una enzima de anclaje, que tiene un sitio de reconocimiento que es probable que esté presente al menos una vez en la mayoría de los ADNcs. Los fragmentos que contienen la región más 5' o 3' del ADNc escindido se aíslan mediante la unión a un medio de captura, tal como cuentas revestidas con estreptavidina. Un primer conector de oligonucleótido que tiene una primera secuencia para la hibridación de un cebador de amplificación y un sitio de restricción interno para una endonucleasa llamada marcadora, se liga al ADNc digerido en el primer grupo. La digestión con la segunda endonucleasa produce fragmentos cortos de etiquetas a partir de los ADNcs.

- 45 Un segundo oligonucleótido que tiene una segunda secuencia para la hibridación de un cebador de amplificación y un sitio de restricción interna se liga a los ADNcs digeridos en el segundo grupo. Los fragmentos de ADNc en el segundo grupo también se digieren con la endonucleasa marcadora para generar fragmentos cortos de etiquetas derivados de los ADNc en el segundo grupo. Las etiquetas resultantes de la digestión del primer y segundo grupo con la enzima de anclaje y la endonucleasa marcadora se ligan entre sí para producir las denominadas etiquetas dobles. En algunas realizaciones, las etiquetas dobles se concatamerizan para formar productos de ligación que contienen de 2 a 200 etiquetas dobles. A continuación se determinan las secuencias marcadoras y se comparan con las secuencias de las EST 5' o ADNc extendidos para determinar cuáles son las ESTs 5' o ADNcs extendidos que se expresan en la célula, tejido, organismo u otra fuente de ácidos nucleicos a partir de la cual derivan las etiquetas. De esta manera, se obtiene el patrón de expresión de las ESTs 5' o ADNcs extendidos en la célula, tejido, organismo u otra fuente de ácidos nucleicos.

- 55 También puede realizarse un análisis cuantitativo de la expresión génica utilizando matrices. Tal como se utiliza en la presente memoria, el término matriz significa una ordenación unidimensional, bidimensional o multidimensional de ADNcs completos (es decir, ADNcs extendidos que incluyen la secuencia codificante del péptido señal, la secuencia codificante de la proteína madura, y un codón de terminación), ADNcs extendidos, ESTs 5' o fragmentos de éstos de longitud suficiente para permitir la detección específica de la expresión génica. Preferiblemente, los fragmentos tienen una longitud de al menos 15 nucleótidos. Más preferiblemente, los fragmentos tienen una longitud de al menos 100 nucleótidos. Más preferiblemente, los fragmentos tienen una longitud de más de 100 nucleótidos. En algunas realizaciones, los fragmentos pueden tener una longitud de más de 500 nucleótidos.

Por ejemplo, el análisis cuantitativo de la expresión génica puede realizarse con ADNc completos como se define a continuación, ADNcs extendidos, ESTs 5' o fragmentos de éstos en una micromatriz de ADN complementario, según describen Schena et al. (Science, 270: 467-470, 1995; Proc. Natl. Acad. Sci. USA, 93: 10614-10619, 1996). Los ADNcs completos, ADNcs extendidos, ESTs 5' o fragmentos de éstos se amplifican mediante PCR y se disponen en matrices desde placas de microtitulación de 96 pocillos hacia portaobjetos de microscopio sililados utilizando robótica de alta velocidad. Las matrices impresas se incuban en una cámara húmeda para permitir la rehidratación de los elementos de la matriz, y se enjuagan una vez con SDS al 0,2% durante 1 min, dos veces con agua durante 1 min, y una vez durante 5 min con disolución de borohidruro de sodio. Las matrices se sumergen en agua durante 2 min a 95°C, se trasladan a SDS al 0,2% durante 1 min, se enjuagan dos veces con agua, se secan al aire y se conservan en la oscuridad a 25°C.

El ARNm celular o tisular se aísla o se obtiene en el mercado, y las sondas se preparan mediante un solo ciclo de transcripción inversa. Las sondas se hibridan con micromatrices de 1 cm² bajo un cubreobjetos de vidrio de 14 x 14 mm durante 6-12 horas a 60°C. Las matrices se lavan durante 5 min a 25°C en tampón de lavado poco restrictivo (1 x SSC/SDS al 0,2%), después durante 10 min a temperatura ambiente en tampón de lavado muy restrictivo (0,1 x SSC/SDS al 0,2%). Las matrices se barren en 0,1 x SSC utilizando un dispositivo de barrido con láser de fluorescencia equipado con un ajuste de filtro personalizado. Se obtienen medidas de expresión diferencial precisas tomando la media de las proporciones de dos hibridaciones independientes.

El análisis cuantitativo de la expresión de los genes también puede realizarse con ADNcs completos, ADNcs extendidos, ESTs 5' o fragmentos de éstos en matrices de ADN complementario, según describen Pietu et al. (Genome Research, 6: 492-503, 1996). Los ADNcs completos, ADNcs extendidos, ESTs 5' o fragmentos de éstos se amplifican con PCR y se rocían sobre membranas. Después, los ARNms que se originan de diversos tejidos o células se marcan con nucleótidos radiactivos. Después de una hibridación y un lavado en condiciones controladas, los ARNms hibridados se detectan mediante formación de fosfoimágenes o autorradiografía. Se realizan experimentos por duplicado y después se realiza un análisis cuantitativo de los ARNm expresados diferencialmente.

Como alternativa, el análisis de la expresión de las ESTs 5' o ADNcs extendidos puede realizarse mediante matrices de nucleótidos de alta densidad, según describen Lockhart et al. (Nature Biotechnology, 14: 1675-1680, 1996), y Sosnowsky et al. (Proc. Natl. Acad. Sci., 94: 1119-1123, 1997). Los oligonucleótidos de 15-50 nucleótidos que se corresponden con las secuencias de las ESTs 5' o ADNcs extendidos se sintetizan directamente sobre el chip (Lockhart et al., supra), o se sintetizan y luego se trasladan al chip (Sosnowsky et al., supra). Preferiblemente, los oligonucleótidos tienen una longitud de aproximadamente 20 nucleótidos.

Las sondas de ADNc marcadas con un compuesto apropiado, tal como biotina, digoxigenina o un colorante fluorescente, se sintetizan a partir de la población de ARNm apropiada, y después se fragmentan de forma aleatoria hasta un tamaño medio de 50 a 100 nucleótidos. Dichas sondas se hibridan a continuación con el chip. Después de lavar como describen Lockhart et al., supra, y aplicar diferentes campos eléctricos (Sonowsky et al., supra), los colorantes o los compuestos marcadores se detectan y se cuantifican. Se realizan hibridaciones por duplicado. El análisis comparativo de la intensidad de la señal que se origina de 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 ADNc extendido a partir del cual se diseñó la secuencia del oligonucleótido.

III. Uso de ESTs 5' para Clonar ADNcs Extendidos y para Clonar los Correspondientes ADNcs genómicos

Cuando las ESTs 5' que incluyen el extremo 5' de los correspondientes ARNms se han seleccionado utilizando los procedimientos descritos anteriormente, pueden utilizarse para aislar ADNcs extendidos que contienen secuencias adyacentes a las ESTs 5'. Los ADNcs extendidos pueden incluir la secuencia codificante completa de la proteína codificada por el correspondiente ARNm, incluyendo el sitio de inicio de la traducción auténtico, la secuencia señal, y la secuencia que codifica la proteína madura que permanece después de la ruptura del péptido señal. Estos ADNcs extendidos se denominan en la presente memoria "ADNc completos". Como alternativa, los ADNcs extendidos pueden incluir sólo la secuencia que codifica la proteína madura que permanece después de la escisión del péptido señal, o sólo la secuencia que codifica el péptido señal.

El Ejemplo 27 a continuación describe un método general para obtener ADNcs extendidos utilizando ESTs 5'. El ejemplo 28 a continuación proporciona resultados experimentales, utilizando el método explicado en el Ejemplo 27, que describen varios ADNcs extendidos que incluyen la secuencia codificante completa y el extremo 5' auténtico del correspondiente ARNm para varias proteínas secretadas.

Los métodos de los Ejemplos 27, 28 y 29 también pueden utilizarse para obtener ADNcs extendidos que codifican menos que la secuencia codificante completa de las proteínas secretadas codificadas por los genes correspondientes a las ESTs 5'. En algunas realizaciones, los ADNcs extendidos aislados utilizando estos métodos codifican al menos 10 aminoácidos de una de las proteínas codificada por las secuencias de los SEQ ID NO: 73+75. En otras realizaciones, los ADNcs extendidos codifican al menos 20 aminoácidos de la proteína codificada por las secuencias de los SEQ ID NO: 73+75. En otras realizaciones, los ADNcs extendidos codifican al menos 30 aminoácidos de las secuencias de los SEQ ID NO: 73+75. En una realización preferida, los ADNcs extendidos codifican una secuencia de la proteína completa, que incluye la secuencia codificante de la proteína de los SEQ ID NO: 73+75.

EJEMPLO 27

Método General para Utilizar ESTs 5' para Clonar y Secuenciar ADNcs que Incluyen la Región Codificante Completa y el Extremo 5' Auténtico del Correspondiente ARNm

El siguiente método general se ha utilizado para aislar de forma rápida y eficaz ADNcs extendidos que tienen los extremos 5' auténticos de sus correspondientes ARNm, así como la secuencia codificante de la proteína completa, e incluyen secuencias adyacentes a las secuencias de las ESTs 5' utilizadas para obtenerlos. Este método puede aplicarse para obtener ADNcs 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 ADNcs Extendidos

a) Síntesis de la primera hebra

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

Después de la eliminación del ARNm hibridado con la primera hebra del 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), como se explica en el Ejemplo 11.

b) Síntesis de la segunda hebra

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

Preferiblemente, los cebadores anidados en el extremo 5' están separados entre sí por cuatro a nueve bases. Las secuencias de los cebadores 5' pueden seleccionarse para que tengan unas temperaturas de fusión y especificidades adecuadas para su uso en PCR.

Preferiblemente, los cebadores anidados en el extremo 3' están separados entre sí por cuatro a nueve bases. Por ejemplo, los cebadores anidados 3' 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 porque tienen temperaturas de fusión y especificidades compatibles con su uso en PCR. Sin embargo, los expertos en la técnica apreciarán que otras secuencias también pueden utilizarse como cebadores.

La primera ronda de PCR de 25 ciclos se realiza utilizando Advantage Tth Polymerase Mix (Clontech) y el cebador externo de cada uno de los pares anidados. Después se realiza una segunda PCR de 20 ciclos utilizando la misma enzima y el cebador interno de cada par anidado en 1/2500 del primer producto de PCR. Después se eliminan los cebadores y nucleótidos.

2. Secuenciación de ADNc Extendidos Completos o sus Fragmentos

Debido a la falta de limitaciones de posición en el diseño de cebadores anidados 5' compatibles para PCR utilizando el programa informático OSP, se obtienen amplicones de dos tipos. Preferiblemente, el segundo cebador 5' se localiza aguas arriba del codón de inicio de la traducción produciendo, por tanto, un producto de PCR anidada que contiene la secuencia codificante completa. Este ADNc extendido completo se somete a un procedimiento de clonación directa, según se describe en la sección a). Sin embargo, en algunos casos, el segundo cebador 5' se localiza aguas abajo del codón de inicio de la traducción produciendo, por tanto, un producto de PCR que contiene sólo parte del ORF. Estos productos de PCR incompletos se someten a un procedimiento modificado descrito en la sección b).

a) Productos de PCR anidados que contienen ORFs completos

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

b) Productos de PCR anidados que contienen ORFs incompletos

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

Cuando la secuencia codificante completa se ha determinado completamente, se diseñan nuevos cebadores compatibles para usar en PCR para obtener amplicones que contienen la región codificante completa. Sin embargo, en estos casos, los cebadores 3' compatibles para usar en PCR se localizan dentro del 3' UTR del correspondiente ARNm, produciendo, con ello, amplicones que carecen de parte de esta región, es decir, el tramo de poliA y, a veces, la señal de poliadenilación, como se ilustra en la Figura 3. Estos ADNcs extendidos completos se clonan a continuación en un vector apropiado, como se describe en la sección 3.

c) Secuenciación de ADNcs extendidos

La secuenciación de los ADNcs extendidos se realiza utilizando un enfoque de Die Terminator con el kit de ADN polimerasa FS AmpliTaq disponible de Perkin Elmer.

Para secuenciar fragmentos de PCR, se realiza un paseo de cebadores utilizando un programa informático, tal como OSP, para elegir cebadores, y un programa informático de ordenador automático, tal como ASMG (Sutton et al., Genome Science Technol., 1: 9-19, 1995) para construir cóntigos de secuencias de paseo que incluyen el marcador 5' inicial utilizando solapamientos mínimos de 32 nucleótidos. Preferiblemente, el paseo de cebadores se realiza hasta que se obtienen las secuencias de los ADNcs completos.

Se evalúa si la secuenciación de un fragmento de ADNc extendido concreto se ha completado como sigue. Puesto que las secuencias colocadas después del tramo de poliA son difíciles de determinar con precisión en el caso de productos no clonados, los procesos de secuenciación y paseo de cebadores para productos de PCR se interrumpen cuando un tramo de poliA se identifica en los ADNc extendidos obtenidos como se describe en el caso b). La longitud de la secuencia se compara con el tamaño del producto de PCR anidada obtenido como se describió anteriormente. Debido a la precisión limitada de la determinación del tamaño de los productos de PCR mediante electroforesis en gel, una secuencia se considera completa si el tamaño de la secuencia obtenida es al menos 70% del tamaño del primer producto de PCR anidada. Si la longitud de la secuencia determinada a partir del análisis por ordenador no tiene al menos 70% de la longitud del producto de PCR anidada, estos productos de PCR se clonan y se determina la secuencia de la inserción. Cuando están disponibles datos de un análisis de transferencia Northern, el tamaño del ARNm detectado para un producto de PCR concreto se utiliza para evaluar finalmente que la secuencia está completa. Las secuencias que no cumplen los criterios anteriores se rechazan y se someten a un nuevo procedimiento de aislamiento.

Los datos de las secuencias de todos los ADNcs extendidos son transferidos a continuación a una base de datos patentada, en la que se realizan los controles de calidad y las etapas de validación como se describe en el Ejemplo 15.

3. Clonación de ADNcs Extendidos Completos

El producto de PCR que contiene la secuencia codificante completa se clona después en un vector apropiado. Por ejemplo, los ADNcs extendidos pueden clonarse en el vector de expresión pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) como sigue. Se prepara el ADN del vector

- pED6dpc2 con extremos romos realizando una digestión con EcoRI, seguida de una reacción de relleno. El vector con extremos romos se desfosforila. Después de retirar los cebadores de PCR y de una precipitación en etanol, el producto de PCR que contiene la secuencia codificante completa o el ADNc extendido obtenido como se describió anteriormente se fosforila con una quinasa, que posteriormente se retira mediante extracción con fenol-Sevag y precipitación. El ADNc extendido bicatenario se liga después al vector, y el plásmido de expresión resultante se introduce en células anfitrionas adecuadas.
- Puesto que los productos de PCR obtenidos como se describió anteriormente son moléculas con extremos romos que pueden clonarse en cualquier dirección, se determina la orientación de varios clones para cada producto de PCR. Después se ordenan de 4 a 10 clones en placas de microtitulación y se someten a una reacción de PCR utilizando un primer cebador colocado en el vector cerca del sitio de clonación, y un segundo cebador colocado en la porción del ADNc extendido que se corresponde con el extremo 3' del ARNm. Este segundo cebador puede ser el cebador antisentido utilizado en la PCR anclada en el caso de la clonación directa (caso a), o el cebador antisentido colocado en el interior de la UTR 3' en el caso de la clonación indirecta (caso b). Los clones en los que el codón de inicio del ADNc extendido está operablemente conectado al promotor en el vector para permitir la expresión de la proteína codificada por el ADNc extendido se conservan y secuencian. Además de los extremos de los insertos de ADNc, también se secuencian aproximadamente 50 pb del ADN del vector en cada lado del inserto de ADNc.
- Los productos de PCR clonados se secuencian a continuación completamente según el procedimiento mencionado anteriormente. En este caso, se realiza después la contigación de fragmentos largos en las secuencias de paseo que ya están contigadas para los productos de PCR no clonados durante el paseo de cebadores. La secuenciación de amplicones clonados está completa cuando los cóntigos resultantes incluyen la región codificante completa, así como las secuencias solapantes con el ADN del vector en ambos extremos.

4. Análisis con Ordenador de los ADNcs Extendidos Completos

- Las secuencias de todos los ADNcs extendidos completos se someten después a un posterior análisis como se describe a continuación. Antes de buscar en los ADNcs extendidos completos las secuencias de interés, los ADNcs extendidos que no son de interés (ARNs de vectores, ARNs de transferencia, ARNs ribosómicos, ARNs mitocondriales, ARNs procarióticos y ARNs fúngicos) se rechazan utilizando métodos fundamentalmente similares a los descritos para las ESTs 5' en el Ejemplo 18.

a) Identificación de características estructurales

- Las características estructurales, por ejemplo, cola de poliA y señal de poliadenilación, de las secuencias de ADNcs extendidos completos se determinan posteriormente como sigue.
- Una cola de poliA se define como un tramo homopolimérico de al menos 11 A, con una base alternativa, como mucho, dentro de él. La búsqueda de colas de poliA se restringe a los últimos 100 nucleótidos de la secuencia, y está limitada a tramos de 11 A consecutivas porque las reacciones de secuenciación a menudo no pueden leerse después de este tramo de poliA. Los tramos que tienen más de 90% de homología en 8 nucleótidos se identifican como colas de poliA utilizando BLAST2N.
- Para buscar una señal de poliadenilación, la cola de poliA se corta de la secuencia completa. Las 50 pb que preceden a la cola de poliA se buscan en primer lugar para encontrar la señal de poliadenilación canónica AAUAAA y, si no se detecta la señal canónica, para encontrar la señal alternativa AUUAAA (Sheets et al., Nuc. Acids Res., 18: 5799-5805, 1990). Si no encuentra ninguna de estas señales de poliadenilación consenso, el motivo canónico se busca de nuevo permitiendo un apareamiento para explicar posibles errores de secuenciación. Más de 85% de las señales de poliadenilación identificadas de cualquier tipo terminan realmente de 10 a 30 pb de la cola de poliA. Las señales AUUAAA alternativas representan aproximadamente 15% del número total de señales de poliadenilación identificadas.

b) Identificación de características funcionales

- Las características funcionales, por ejemplo, ORFs y secuencias señal, de las secuencias de ADNcs extendidos completos se determinaron posteriormente como sigue.
- Los marcos de la cadena superior 3 de los ADNcs extendidos se buscan para encontrar ORF definidos como los fragmentos de longitud máxima con un codón de inicio de la traducción y que terminan con un codón de terminación. Se prefieren los ORFs que codifican al menos 20 aminoácidos.
- Cada ORF encontrado se barre a continuación en busca de la presencia de un péptido señal en los primeros 50 aminoácidos o, cuando sea apropiado, dentro de regiones más cortas de hasta 20

aminoácidos o menos en el ORF, utilizando el método de la matriz de Von Heijne (Nuc. Acids Res., 14: 4683-4690, 1986), como se describe en el Ejemplo 22.

c) Homología con secuencias nucleotídicas o proteicas

5 Puede lograrse la categorización de secuencias completas utilizando procedimientos fundamentalmente similares a los descritos para las ESTs 5' en el Ejemplo 24.

10 Los ADNcs extendidos preparados como se describió anteriormente pueden modificarse posteriormente para obtener ácidos nucleicos que incluyen porciones deseadas de ADNc extendido utilizando técnicas convencionales, tales como subclonación, PCR, o síntesis de oligonucleótidos in vitro. Por ejemplo, pueden obtenerse ácidos nucleicos que incluyen sólo las secuencias codificantes completas (es decir, las secuencias que codifican el péptido señal y la proteína madura que permanece después de que se haya escindido el péptido señal) utilizando mecanismos conocidos por los expertos en la técnica. Como alternativa, pueden aplicarse técnicas convencionales para obtener ácidos nucleicos que contienen sólo las secuencias codificantes para la proteína madura que permanece después de que se haya escindido el péptido señal, o ácidos nucleicos que contienen sólo las secuencias codificantes para los péptidos señal.

20 De forma similar, pueden obtenerse ácidos nucleicos que contienen cualquier otra porción deseada de las secuencias codificantes para la proteína secretada. Por ejemplo, el ácido nucleico puede contener al menos 10 bases consecutivas de un ADNc extendido, tal como uno de los ADNcs 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 ADNcs extendidos descritos a continuación. Como alternativa, el ácido nucleico puede contener al menos 20 bases consecutivas de un ADNc extendido, tal como uno de los ADNcs 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 ADNcs extendidos descritos a continuación. En otra realización, el ácido nucleico puede contener al menos 40 bases consecutivas de un ADNc extendido, tal como uno de los ADNcs extendidos descritos a continuación.

30 Cuando se ha obtenido un ADNc extendido, puede secuenciarse para determinar la secuencia de aminoácidos que codifica. Cuando se ha determinado la secuencia de aminoácidos codificada, se puede crear e identificar cualquier de los muchos ADNcs concebibles que codifican esa proteína simplemente utilizando la degeneración del código genético. Por ejemplo, pueden identificarse variantes alélicas u otros ácidos nucleicos homólogos como se describe a continuación. Como alternativa, pueden sintetizarse in vitro ácidos nucleicos que codifican la secuencia de aminoácidos deseada.

35 En una realización preferida, la secuencia codificante puede seleccionarse utilizando el codón conocido o las preferencias de la pareja del codón para el organismo anfitrión en el cual se va a expresar el ADNc.

Los ADNcs extendidos derivados de las ESTs 5' de la presente memoria se obtuvieron como se describe en el Ejemplo 28, a continuación.

EJEMPLO 28

40 Caracterización de ADNcs extendidos clonados obtenidos utilizando EST 5'

Se usó el procedimiento descrito en el Ejemplo 27 anterior para obtener ADNcs extendidos derivados de las ESTs 5' de la presente invención en una diversidad de tejidos. La siguiente lista proporciona unos cuantos ejemplos de los ADNcs extendidos obtenidos de esta manera.

45 Utilizando este enfoque, se obtuvo el ADNc completo del SEQ ID NO: 17 (número de identificación interna 48-19-3-GI-FL1). Este ADNc se incluye en la categoría "EST-ext" descrita anteriormente, y codifica el péptido señal MKKVLLITAILAVAVG (SEQ ID NO: 18), que tiene una puntuación de Von Heijne de 8,2.

50 También se obtuvo el ADNc completo del SEQ ID NO: 19 (número de identificación interna 58-34-2-E7-FL2) utilizando este procedimiento. Este ADNc se incluye en la categoría "EST-ext" descrita anteriormente, y codifica el péptido señal MWWFQQGLSFLPSALVIWTS (SEQ ID NO: 20), que tiene una puntuación de Von Heijne de 5,5.

55 Otro ADNc completo obtenido utilizando el procedimiento descrito anteriormente tiene la secuencia del SEQ ID NO: 21 (número de identificación interna 51-27-1-E8-FL1). Este ADNc se incluye en la categoría "EST-ext" descrita anteriormente, y codifica el péptido señal MVLTTLPANSANSPVNMPTTGPNSLSYASSALSPCLT (SEQ ID NO: 22), que tiene una puntuación de Von Heijne de 5,9.

El anterior procedimiento también se utilizó para obtener un ADNc completo que tiene la secuencia del SEQ ID NO: 23 (número de identificación interna 76-4-1-G5-FL1). Este ADNc se incluye en la categoría “EST-ext” descrita anteriormente, y codifica el péptido señal ILSTVTALTFAA (SEQ ID NO: 24), que tiene una puntuación de Von Heijne de 5,5.

5 También se obtuvo el ADNc completo del SEQ ID NO: 25 (número de identificación interna 51-3-3-B10-FL3) utilizando este procedimiento. Este ADNc se incluye en la categoría “nueva” descrita anteriormente, y codifica el péptido señal LVLTLCTPLAVA (SEQ ID NO: 26), que tiene una puntuación de Von Heijne de 10,1.

10 También se obtuvo el ADNc completo del SEQ ID NO: 27 (número de identificación interna 58-35-2-F10-FL2) utilizando este procedimiento. Este ADNc se incluye en la categoría “nueva” descrita anteriormente, y codifica el péptido señal LWLLFFLVTAIHA (SEQ ID NO: 28), que tiene una puntuación de Von Heijne de 10,7.

15 Los clones bacterianos que contienen plásmidos que contienen los ADNcs completos descritos anteriormente se conservan en la actualidad en los laboratorios de los autores de la presente invención con los números de identificación internos indicados anteriormente. Los insertos pueden recuperarse de los materiales conservados cultivando una parte alícuota del clon bacteriano apropiado en el medio apropiado. El ADN del plásmido puede aislarse después utilizando procedimientos de aislamiento de plásmidos familiares para los expertos en la técnica, tales como procedimientos de aislamiento de plásmidos de minipreps de lisis alcalina o de lisis alcalina a gran escala. Si se desea, el ADN del plásmido puede enriquecerse aún más mediante centrifugación en un gradiente de cloruro de cesio, cromatografía de exclusión molecular, o cromatografía de intercambio aniónico. El ADN del plásmido obtenido utilizando estos procedimientos puede manipularse a continuación utilizando técnicas de clonación convencionales familiares para los expertos en la técnica. Como alternativa, puede realizarse una PCR con cebadores diseñados en ambos extremos de la inserción de ADNc. El producto de PCR que se corresponde con el ADNc puede manipularse después utilizando técnicas de clonación convencionales, familiares para los expertos en la técnica.

20 Los polipéptidos codificados por los ADNcs extendidos pueden seleccionarse para determinar la presencia de motivos estructurales o funcionales conocidos, o la presencia de firmas, secuencias pequeñas de aminoácidos que están bien conservadas entre los miembros de una familia de proteínas. Se han utilizado las regiones conservadas para derivar patrones consenso o matrices incluidas en el banco de datos PROSITE, en particular en el archivo prosite.dat (descarga 13.0 de Noviembre 1995, localizado en <http://expasy.hcuge.ch/sprot/prosite.html>). Pueden utilizarse los programas prosite_convert y prosite_scan (http://ulrec3.unil.ch/ftpserveur/prosite_scan) para encontrar firmas sobre los ADNcs extendidos.

30 Para cada patrón obtenido con el programa prosite_convert del archivo prosite.dat, puede analizarse la precisión de la detección en una nueva secuencia de proteína, evaluando la frecuencia de aciertos irrelevantes en la población de proteínas secretadas humanas incluidas en el banco de datos SWISSPROT. La proporción entre el número de aciertos en proteínas barajadas (con un tamaño de ventana de 20 aminoácidos) y el número de aciertos en proteínas nativas (no barajadas) puede utilizarse como un índice. Los patrones para los cuales la proporción es mayor de 20% (un acierto en proteínas barajadas por 5 aciertos en proteínas nativas) pueden omitirse durante la búsqueda con prosite_scan. El programa utilizado para barajar secuencias de proteínas (db_shuffled) y el programa utilizado para determinar la estadística para cada patrón en los bancos de datos de proteínas (prosite_statistics) están disponibles en el sitio ftp http://ulrec3.unil.ch/ftpserveur/prosite_scan.

45 Además de los métodos basados en PCR para obtener ADNcs extendidos, también pueden emplearse métodos basados en la hibridación tradicionales. Estos métodos también pueden utilizarse para obtener ADNs genómicos que codifican los ARNm a partir de los cuales derivan las ESTs 5', los ARNms que corresponden a los ADNcs extendidos, o los ácidos nucleicos que son homólogos a los ADNcs extendidos o ESTs 5'. El ejemplo 29, a continuación, proporciona ejemplos de estos métodos.

50 EJEMPLO 29

Métodos para Obtener ADNcs que Incluyen la Región Codificante Completa y el Extremo 5' Auténtico del Correspondiente ARNm

55 Puede fabricarse un banco de ADNc completo utilizando las estrategias descritas en los Ejemplos 13, 14, 15 y 16 anteriores, sustituyendo el nonúmero aleatorio utilizado en el Ejemplo 14 por un cebador de oligo-dT. Por ejemplo, puede emplearse el oligonucleótido del SEQ ID NO: 14.

Como alternativa, puede obtenerse un banco de ADNc o un banco de ADN genómico a partir de una fuente comercial, o fabricarse utilizando mecanismos familiares para los expertos en la técnica. Estos bancos de ADNc o ADN genómico pueden usarse para aislar ADNc extendidos obtenidos a partir de EST 5', o ácidos nucleicos homólogos con ADNc extendidos o EST 5' como sigue. El banco de ADNc

- o el banco de ADN genómico se hibrida con una sonda detectable que comprende al menos 10 nucleótidos consecutivos de la EST 5' o del ADNc extendido utilizando técnicas convencionales. Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos de la EST 5' o del ADNc extendido. Más preferiblemente, la sonda comprende al menos de 20 a 30 nucleótidos consecutivos de la EST 5' o del ADNc extendido. En algunas realizaciones, la sonda comprende más de 30 nucleótidos de la EST 5' o del ADNc extendido.
- Las técnicas para identificar los clones de ADNc en un banco de ADNc que se hibridan con una secuencia sonda concreta son descritas por Sambrook et al., Molecular Cloning: A Laboratory Manual, 2ª ed., Cold Spring Harbor Laboratory Press, 1989. Pueden utilizarse las mismas técnicas para aislar ADN genómicos.
- Brevemente, los clones de ADNc o ADN genómico que se hibridan con la sonda detectable se identifican y se aíslan para su posterior manipulación como sigue. Una sonda que comprende al menos 10 nucleótidos consecutivos de la EST 5' o del ADNc extendido se marca con un marcador detectable, tal como un radioisótopo o una molécula fluorescente. Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos de la EST 5' o del ADNc extendido. Más preferiblemente, la sonda comprende 20 a 30 nucleótidos consecutivos de la EST 5' o del ADNc extendido. En algunas realizaciones, la sonda comprende más de 30 nucleótidos de la EST 5' o del ADNc extendido.
- Las técnicas para marcar la sonda son muy conocidas e incluyen la fosforilación con polinucleótido quinasa, la traducción de mellas, la transcripción in vitro, y técnicas no radiactivas. Los ADNcs o ADNs genómicos en el banco se trasladan a un filtro de nitrocelulosa o nailon y se desnaturalizan. Después de bloquear los sitios no específicos, el filtro se incuba con la sonda marcada durante una cantidad de tiempo suficiente para permitir la unión de la sonda a los ADNcs o ADNs genómicos que contienen una secuencia capaz de hibridarse con ella.
- Variando la severidad de las condiciones de hibridación utilizadas para identificar los ADNcs extendidos o ADNs genómicos que se hibridan con la sonda detectable, pueden identificarse los ADNcs extendidos que tienen diferentes niveles de homología con la sonda y aislarse como se describe a continuación.
1. Identificación de Secuencias de ADNc Extendido o ADNc Genómico que Tienen un Alto Grado de Homología con la Sonda Marcada
- Para identificar los ADNcs extendidos o ADNs genómicos que tienen un alto grado de homología con la secuencia de la sonda, puede calcularse la temperatura de fusión de la sonda utilizando las siguientes fórmulas:
- Para sondas con una longitud entre 14 y 70 nucleótidos, la temperatura de fusión (T_f) se calcula utilizando la fórmula: $T_f = 81,5 + 16,6(\log [Na +]) + 0,41(\text{fracción G+C}) - (600/N)$, en la que N es la longitud de la sonda.
- Si la hibridación se realiza en una disolución que contiene formamida, la temperatura de fusión puede calcularse utilizando la ecuación: $T_f = 81,5 + 16,6(\log [Na+]) + 0,41(\text{fracción G+C}) - (\text{formamida al } 0,63\%) - (600/N)$, en la que N es la longitud de la sonda.
- La prehibridación puede realizarse en 6 x SSC, 5 x reactivo de Denhardt, SDS al 0,5%, 100 µg de ADN de esperma de salmón fragmentado desnaturalizado, o 6 x SSC, 5 x reactivo de Denhardt, SDS al 0,5%, 100 µg de ADN de esperma de salmón fragmentado desnaturalizado, formamida al 50%. Las fórmulas para SSC y las disoluciones de Denhardt se enumeran en Sambrook et al., supra.
- La hibridación se realiza añadiendo la sonda detectable a las disoluciones de prehibridación enumeradas anteriormente. Cuando la sonda comprende ADN bicatenario, se desnaturaliza antes de la adición a la disolución de hibridación. El filtro se pone en contacto con la disolución de hibridación durante un periodo de tiempo suficiente para permitir que la sonda se hibride con los ADNcs extendidos o los ADNs genómicos que contienen secuencias complementarias a ella u homólogas a ella. Para sondas con una longitud mayor de 200 nucleótidos, la hibridación puede realizarse a 15-25°C por debajo de la T_f . Para sondas más cortas, tales como sondas de oligonucleótidos, la hibridación puede realizarse a 15-25°C por debajo de la T_f . Preferiblemente, para hibridaciones en 6 x SSC, la hibridación se realiza a aproximadamente 68°C. Preferiblemente, para hibridaciones en disoluciones que contienen formamida al 50%, la hibridación se realiza a aproximadamente 42°C.
- Se consideraría que todas las anteriores hibridaciones estaban bajo condiciones "rigurosas". Después de la hibridación, el filtro se lava en 2 x SSC, SDS al 0,1% a temperatura ambiente durante 15 minutos. El filtro se lava después con 0,1 x SSC, SDS al 0,5% a temperatura ambiente durante 30 minutos a 1 hora. Después la disolución se lava a la temperatura de hibridación en 0,1 x SSC, SDS al 0,5%. Se realiza un lavado final en 0,1 x SSC a temperatura ambiente.

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

2. Obtención de Secuencias de ADNc Extendido o ADNc Genómico que Tienen Grados Menores de Homología con la Sonda Marcada

El anterior procedimiento puede modificarse para identificar ADNcs extendidos, ácidos nucleicos homólogos con los ADNcs extendidos, o ADNcs genómicos que tienen niveles decrecientes de homología con la secuencia de la sonda. Por ejemplo, para obtener ADNcs extendidos, ácidos nucleicos homólogos a los ADNcs extendidos, o ADNcs genómicos que tienen homología decreciente con la sonda detectable, pueden utilizarse condiciones menos rigurosas. Por ejemplo, la temperatura de hibridación puede disminuir en incrementos de 5°C desde 68°C hasta 42°C en un tampón de hibridación que tiene una concentración de sodio de aproximadamente 1 M. Después de la hibridación, el filtro puede lavarse con 2 x SSC, SDS al 0,5% a la temperatura de hibridación. Se considera que estas condiciones son "moderadas" por encima de 50°C, y "bajas" por debajo de 50°C.

Como alternativa, la hibridación puede realizarse en tampones, tales como 6 x SSC, que contienen formamida a una temperatura de 42°C. En este caso, la concentración de formamida en el tampón de hibridación puede reducirse en incrementos de 5% desde 50% hasta 0% para identificar los clones que tienen unos niveles decrecientes de homología con la sonda. Después de la hibridación, el filtro puede lavarse con 6 x SSC, SDS al 0,5% a 50°C. Se considera que estas condiciones son "moderadas" por encima de formamida al 25%, y "bajas" por debajo de formamida al 25%.

Los ADNcs extendidos, ácidos nucleicos homólogos con los ADNcs extendidos, o ADNcs genómicos que se han hibridado con la sonda se identifican mediante autorradiografía.

3. Determinación del Grado de Homología Entre los ADNcs Extendidos Obtenidos y la Sonda Marcada

Si se desea obtener ácidos nucleicos homólogos a los ADNcs extendidos, tales como variantes alélicas de éstos o ácidos nucleicos que codifican proteínas relacionadas con las proteínas codificadas por los ADNcs extendidos, el nivel de homología entre el ácido nucleico hibridado y el ADNc extendido o EST 5' utilizado como sonda puede determinarse con más precisión utilizando BLAST2N; los parámetros pueden adaptarse 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 EST 5' a partir del cual deriva la sonda, se comparan las secuencias de nucleótidos del ácido nucleico hibridado y el ADNc extendido o EST 5' a partir de los cuales deriva la sonda. Por ejemplo, utilizando los métodos anteriores, los ácidos nucleicos que tienen una homología de ácido nucleico de al menos 95% con el ADNc extendido o EST 5' a partir de los cuales deriva la sonda pueden obtenerse e identificarse. De forma similar, usando condiciones de hibridación progresivamente menos rigurosas, se pueden obtener e identificar ácidos nucleicos que tienen una homología de al menos 90%, al menos 85%, al menos 80% o al menos 75% con el ADNc extendido o EST 5' a partir de los cuales deriva la sonda.

Para determinar si un clon codifica una proteína que tiene una cantidad concreta de homología con la proteína codificada por el ADNc extendido o EST 5', la secuencia de aminoácidos codificada por el ADNc extendido o EST 5' se compara con la secuencia de aminoácidos codificada por el ácido nucleico hibridado. Se determina que existe homología cuando una secuencia de aminoácidos en el ADNc extendido o EST 5' está muy relacionada con una secuencia de aminoácidos en el ácido nucleico que se está hibridando. Una secuencia está muy relacionada cuando es idéntica a la del ADNc extendido o EST 5', o cuando contiene una o más sustituciones de aminoácidos en ella, en la que los aminoácidos que tienen características similares se han sustituido por otro. Utilizando los anteriores métodos y algoritmos, tales como FASTA, con parámetros que dependen de la longitud de la secuencia y grado de homología estudiado, se pueden obtener ácidos nucleicos que codifican proteínas que tienen una homología de al menos 95%, al menos 90%, al menos 85%, al menos 80% o al menos 75% con las proteínas codificadas por el ADNc extendido o EST 5' de los cuales deriva la sonda.

Además de los métodos descritos anteriormente, están disponibles otros protocolos para obtener ADNcs extendidos utilizando EST 5' como se indica en los siguientes párrafos.

Pueden prepararse ADNcs extendidos obteniendo ARNm del tejido, célula u organismo de interés utilizando procedimientos de preparación de ARNm que emplean procedimientos de selección de poliA u otros mecanismos conocidos por los expertos en la técnica. Un primer cebador capaz de hibridarse con la cola de 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 las secuencias de los SEQ ID NO: 73+75. Preferiblemente, el cebador comprende al

- menos 12, 15 ó 17 nucleótidos consecutivos de la secuencia de los SEQ ID NO: 73+75. Más preferiblemente, el cebador comprende de 20 a 30 nucleótidos consecutivos de las secuencias de los SEQ ID NO: 73+75. En algunas realizaciones, el cebador comprende más de 30 nucleótidos de las secuencias de los SEQ ID NO: 73+75. Si se desea obtener ADNcs extendidos que contienen la secuencia codificante completa de la proteína, incluyendo el sitio de inicio de la traducción auténtico, el segundo cebador utilizado 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. Como alternativa, puede realizarse una RT-PCR como se describió anteriormente utilizando cebadores de ambos extremos del ADNc que se va a obtener.
- 5 Pueden prepararse ADNcs extendidos que contienen fragmentos 5' del ARNm hibridando un ARNm que comprende la secuencia de la EST 5' de la cual se desea un ADNc extendido, con un cebador que comprende al menos 10 nucleótidos consecutivos de la secuencia complementaria a la EST 5', y realizando una transcripción inversa con el cebador hibridado para fabricar una primera hebra de ADNc a partir de los ARNms. Preferiblemente, el cebador comprende al menos 12, 15 ó 17 nucleótidos consecutivos de la EST 5'. Más preferiblemente, el cebador comprende 20 a 30 nucleótidos consecutivos de la EST 5'.
- Después, se sintetiza una segunda hebra de ADNc complementaria a la primera hebra de ADNc. La segunda hebra de ADNc puede fabricarse hibridando un cebador complementario a secuencias en la primera hebra de ADNc, con la primera hebra de ADNc, y extendiendo el cebador para generar la segunda hebra de ADNc.
- 20 Los ADNcs extendidos bicatenarios fabricados utilizando los métodos descritos anteriormente se aíslan y se clonan. Los ADNcs extendidos pueden clonarse en vectores, tales como plásmidos o vectores víricos, capaces de replicarse en una célula anfitriona adecuada. Por ejemplo, la célula anfitriona puede ser una célula bacteriana, de mamífero, de ave o de insecto.
- 25 Los mecanismos para aislar ARNm, para realizar una transcripción inversa de un cebador hibridado con ARNm para generar una primera hebra de ADNc, para extender un cebador para fabricar una segunda hebra de ADNc complementaria a la primera hebra de ADNc, para aislar el ADNc bicatenario, y para clonar el ADNc bicatenario son muy conocidos por los expertos en la técnica, y se describen en Current Protocols in Molecular Biology, John Wiley and Sons, Inc., 1997, y Sambrook et al., Molecular Cloning: A Laboratory Manual, 2ª edición, Cold Spring Harbor Laboratory Press, 1989. Como alternativa, pueden emplearse procedimientos, tales como el descrito en el ejemplo 29, para obtener ADNcs completos o ADNcs extendidos. En este enfoque, se preparan ADNcs completos o extendidos a partir de ARNm, y se clonan en fagémidos bicatenarios como sigue. El banco de ADNc en los fagémidos bicatenarios se hace después monocatenario mediante un tratamiento con una endonucleasa, tal como el producto del gen II del fago F1, y una exonucleasa (Chang et al., Gene, 127:95-98, 1993). Un oligonucleótido biotinilado que comprende la secuencia de una EST 5', o un fragmento que contiene al menos 10 nucleótidos de ésta, se hibrida con los fagémidos monocatenarios. Preferiblemente, el fragmento comprende al menos 12, 15 ó 17 nucleótidos consecutivos de la EST 5'. Más preferiblemente, el fragmento comprende 20-30 nucleótidos consecutivos de la EST 5'. En algunos procedimientos, el fragmento puede comprender más de 30 nucleótidos consecutivos de la EST 5'.
- Los híbridos entre el oligonucleótido biotinilado y los fagémidos que tienen insertos que contienen la secuencia EST 5' se aíslan incubando los híbridos con cuentas paramagnéticas revestidas con estreptavidina y recuperando las cuentas con un imán (Fry et al., Biotechniques, 13: 124-131, 1992).
- 45 Después, los fagémidos resultantes que contienen la secuencia EST 5' se liberan de las cuentas y se convierten en ADN bicatenario utilizando un cebador específico para la secuencia de EST 5'. Como alternativa, pueden usarse protocolos, tales como el kit Gene Trapper (Gibco BRL). El ADN bicatenario resultante se transforma en bacterias. Los ADNcs extendidos que contienen la secuencia de EST 5' se identifican mediante PCR de colonias o hibridación de colonias.
- 50 Utilizando cualquiera de los métodos descritos anteriormente en la sección III, puede proporcionarse una pluralidad de ADNcs extendidos que contienen secuencias codificantes completas de la proteína o secuencias que codifican sólo la proteína madura que permanece después de escindir el péptido señal, como bancos de ADNc para la posterior evaluación de las proteínas codificadas o para uso en ensayos de diagnóstico como se describe a continuación.
- 55 **IV. Expresión de Proteínas Codificadas por ADNcs Extendidos Aislados Utilizando ESTs 5'**
- Los ADNcs extendidos que contienen las secuencias codificantes completas de la proteína de sus correspondientes ARNms, o porciones de éstos, tales como ADNcs que codifican la proteína madura, pueden usarse para expresar las proteínas secretadas codificadas o porciones de éstas, como se describe en el Ejemplo 30, a continuación. Si se desea, los ADNcs extendidos pueden contener las secuencias que codifican el péptido señal para facilitar la secreción de la proteína expresada. Se apreciará que una pluralidad de ADNcs extendidos que contienen las secuencias codificantes
- 60

completas de la proteína o porciones de éstas pueden clonarse simultáneamente en vectores de expresión para crear un banco de expresión para el análisis de las proteínas codificadas, como se describe a continuación.

EJEMPLO 30

5 Expresión de Proteínas Codificadas por los Genes Correspondientes a las ESTs 5' o Porciones de Éstas

10 Para expresar las proteínas codificadas por los genes correspondientes a las ESTs 5' (o porciones de éstas), se obtienen ADNcs completos que contienen la región codificante completa de la proteína o ADNcs extendidos que contienen secuencias adyacentes a las ESTs 5' (o porciones de éstas) como se describe en los Ejemplos 27-29, y se clonan en un vector de expresión adecuado. Si se desea, los ácidos nucleicos pueden contener la 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 secuencias que regulan los niveles de expresión o secuencias que confieren una expresión específica de tejido.

20 El ácido nucleico que codifica la proteína o polipéptido que se va a expresar se conecta operablemente a un promotor en un vector de expresión utilizando la tecnología de clonación convencional. El vector de expresión puede ser cualquiera de los sistemas de expresión de mamífero, levadura, insecto o bacteriano conocidos en la técnica. Los vectores y sistemas de expresión disponibles en el mercado pueden adquirirse de una diversidad de suministradores, incluyendo Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), e Invitrogen (San Diego, California). Si se desea, para potenciar la expresión y facilitar el correcto plegamiento de la proteína, puede optimizarse el contexto de codones y el apareamiento de codones de la secuencia para el organismo de expresión concreto en el que se introduce el vector de expresión, como explican Hatfield et al., patente de EEUU Núm. 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 por la escisión del péptido señal), sólo el péptido señal o cualquier otra porción de ésta.

30 El siguiente método se proporciona como ejemplo para expresar las proteínas codificadas por los ADNcs extendidos que corresponden a las ESTs 5' o los ácidos nucleicos descritos anteriormente. En primer lugar, se identifica el codón de inicio de metionina para el gen, y la señal de poliA del gen. Si el ácido nucleico que codifica el polipéptido que se va a expresar carece de una metionina para actuar como el sitio de inicio, puede introducirse una metionina de inicio junto al primer codón del ácido nucleico utilizando técnicas convencionales. De forma similar, si el ADNc extendido carece de una señal de poliA, esta secuencia puede añadirse al constructo, por ejemplo, escindiendo la señal de poliA de pSG5 (Stratagene) utilizando las enzimas endonucleasas de restricción BglII y Sall, e incorporándola en el vector de expresión de mamífero pXT1 (Stratagene). El pXT1 contiene las LTRs y una porción del gen gag del virus de la leucemia murina de Moloney. La posición de las LTRs en el constructo permite una transfección estable y eficaz. El vector incluye el promotor de timidina quinasa de Herpes simplex y el gen seleccionable de neomicina. El ADNc extendido o una porción de éste que codifica el polipéptido que se va a expresar se obtiene mediante PCR a partir del vector bacteriano utilizando cebadores de oligonucleótidos complementarios al ADNc extendido o una porción de éste, y contiene secuencias de endonucleasas de restricción para PstI incorporadas en el cebador 5', y BglII en el extremo 5' del correspondiente cebador 3' del ADNc, teniendo cuidado de asegurarse de que el ADNc extendido está colocado con la señal de poliA. El fragmento purificado obtenido de la reacción de PCR resultante se digiere con PstI, se crean extremos romos con una exonucleasa, se digiere con BglII, se purifica y se liga a pXT1 que contiene una señal de poliA, y se prepara para esta ligación (extremo romo/BglII).

50 El producto acoplado se transfecta en células NIH 3T3 utilizando Lipofectina (Life Technologies, Inc., Grand Island, Nueva York) bajo las condiciones indicadas en la especificación del producto. Los transfectantes positivos se seleccionan después de cultivar las células transfectadas en G418 600 µg/ml (Sigma, St. Louis, Missouri). Preferiblemente, la proteína expresada se libera en el medio de cultivo, facilitando, con ello, la purificación.

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

60 Las proteínas en el medio de cultivo se separan mediante electroforesis en gel. Si se desea, las proteínas puede precipitarse en sulfato de amonio o separarse basándose en el tamaño o carga antes

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

Pueden generarse anticuerpos capaces de reconocer específicamente la proteína de interés utilizando péptidos de 15 unidades sintéticos que tienen una secuencia codificada por la EST 5' apropiada, el ADNc extendido apropiado o una porción de éstos. Los péptidos sintéticos se inyectan en ratones para generar anticuerpos contra el polipéptido codificado por la EST 5', el ADNc extendido o una porción de éstos.

Las proteínas secretadas a partir de las células u organismos anfitriones que contienen un vector de expresión que contiene el ADNc extendido derivado de una EST 5' o una porción de ésta se comparan con las de células u organismos control. 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 control, indica que el ADNc extendido codifica una proteína secretada. En general, la banda correspondiente a la proteína codificada por el ADNc extendido tendrá una movilidad cercana a la esperada basándose en el número de aminoácidos en el marco de lectura abierto del ADNc extendido. Sin embargo, la banda puede tener una movilidad diferente de la esperada como resultado de modificaciones, tales como glicosilación, ubiquitinación o escisión enzimática.

Como alternativa, si la proteína expresada a partir de los anteriores vectores de expresión no contiene secuencias que dirigen su secreción, las proteínas expresadas a partir de células anfitrionas que contienen un vector de expresión con un inserto que codifica una proteína secretada o una porción de ésta pueden compararse con proteínas expresadas en células anfitrionas control que contienen el vector de expresión sin inserto. La presencia de una banda en las 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 inserto indica que la proteína deseada, o una porción de ésta, está siendo expresada. En general, la banda tendrá la movilidad esperada para la proteína secretada o una porción de ésta. Sin embargo, la banda puede tener una movilidad diferente de la esperada como resultado de modificaciones, tales como glicosilación, ubiquitinación o escisión enzimática. La proteína codificada por el ADNc extendido puede purificarse utilizando técnicas de inmunocromatografía convencionales. En estos procedimientos, una disolución que contiene la proteína secretada, tal como el medio de cultivo o un extracto celular, se aplica a una columna que tiene anticuerpos contra la proteína secretada unidos a la matriz de cromatografía. Se permite que la proteína secretada se una a la columna de inmunocromatografía. Después, la columna se lava para eliminar las proteínas unidas de forma no específica. La proteína secretada unida de forma específica se libera después de la columna y se recupera utilizando técnicas convencionales.

Si no es posible la producción de anticuerpos, la secuencia del ADNc extendido o una porción de ésta pueden incorporarse en vectores de expresión diseñados para su uso en esquemas de purificación que emplean polipéptidos quiméricos. En estas estrategias, la secuencia codificante del ADNc extendido o una porción de ésta se inserta dentro del marco con un gen que codifica la otra mitad de la quimera. La otra mitad de la quimera puede ser β -globina o un polipéptido de unión a níquel. A continuación se utiliza una matriz de cromatografía que tiene un anticuerpo contra β -globina o níquel unido a ella para purificar la proteína quimérica. Pueden introducirse sitios de escisión por proteasas entre el gen de la β -globina o el polipéptido de unión a níquel y el ADNc extendido o una porción de éste. Por tanto, los dos polipéptidos de la quimera pueden separarse entre sí mediante digestión con proteasas.

Un vector de expresión útil para generar quimeras de β -globina es pSG5 (Stratagene), que codifica la β -globina de conejo. El intrón II del gen de la β -globina de conejo facilita el corte y empalme del transcrito expresado, y la señal de poliadenilación incorporada al constructo aumenta el nivel de expresión. Estos mecanismos, según están descritos, son muy conocidos por los expertos en la técnica de la biología molecular. Los métodos convencionales están publicados en textos metodológicos, tales como Davis et al. (Basic Methods in Molecular Biology, Davis, Dibner y Battey, ed., Elsevier Press, NY, 1986) y muchos de los métodos están disponibles en Stratagene, Life Technologies, Inc., o Promega. El polipéptido también puede producirse a partir del constructo utilizando sistemas de traducción in vitro, tales como el kit In vitro Express® Translation (Stratagene).

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

EJEMPLO 31Análisis de Proteínas Secretadas para Determinar si se Unen a la Superficie Celular

Las proteínas codificadas por las ESTs 5', ADNcs extendidos o fragmentos de éstos se clonan en vectores de expresión, tales como los descritos en el Ejemplo 30. Las proteínas se purifican mediante cromatografía de exclusión, de carga, inmunocromatografía u otros mecanismos familiares para los expertos en la técnica. Después de la purificación, las proteínas se marcan utilizando mecanismos conocidos por los expertos en la técnica. Las proteínas marcadas se incuban con células o líneas celulares derivadas de una diversidad 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 unida de forma no específica. Las proteínas marcadas se detectan mediante autorradiografía. Como alternativa, las proteínas no marcadas pueden incubarse con las células y detectarse con anticuerpos que tienen un marcador detectable, tal como una molécula fluorescente, unido a ellos.

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

Como se indicó anteriormente, se ha demostrado que las proteínas secretadas tienen una serie de importantes efectos fisiológicos y, por consiguiente, representan un valioso recurso terapéutico. Las proteínas secretadas codificadas por los ADNcs extendidos o porciones de éstos fabricadas según los Ejemplos 27-29 pueden evaluarse para determinar sus actividades fisiológicas, como se describe a continuación.

EJEMPLO 32Análisis de las Proteínas Expresadas a Partir de ADNcs Extendidos o Porciones de Éstos para la Actividad de Citoquinas, Proliferación Celular o Diferenciación Celular

Como se analizó anteriormente, las proteínas secretadas pueden actuar como citoquinas, o pueden afectar a la proliferación o diferenciación celular. Muchos factores de proteínas descubiertos hasta la fecha, incluyendo todas las citoquinas conocidas, han mostrado actividad en uno o más análisis de proliferación celular dependientes de factor y, por tanto, los ensayos sirven como una confirmación conveniente de actividad de las citoquinas. La actividad de una proteína codificada por los ADNc extendidos queda demostrada por uno cualquiera de una serie de análisis de proliferación celular dependientes de factor rutinarios para líneas celulares que incluyen, sin limitación: 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c y CMK. Las proteínas codificadas por los anteriores ADNc extendidos o porciones de éstos pueden evaluarse para determinar su capacidad para regular la proliferación de células T o timocitos en análisis tales como los descritos anteriormente, o en las siguientes referencias: Current Protocols in Immunology, Ed. por Coligan et al., Greene Publishing Associates and Wiley-Interscience; Takai et al., J. Immunol., 137: 3494-3500, 1986; Bertagnolli et al., J. Immunol., 125: 1706-1712, 1990; Bertagnolli et al., Cell. Immunol., 133: 327-341, 1991; Bertagnolli et al., J. Immunol., 149: 3778-3783, 1992; Bowman et al., J. Immunol., 152: 1756-1761, 1994.

Además, se conocen numerosos análisis 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, supra, 1: 3.12.1-3.12.14; y Schreiber, en Current Protocols in Immunology, supra, 1: 6.8.1-6.8.8.

Las proteínas codificadas por los ADNcs también pueden analizarse para determinar su capacidad para regular la proliferación y diferenciación de células hematopoyéticas o linfopoyéticas. Muchos análisis para determinar esta actividad son familiares para los expertos en la técnica, incluyendo los ensayos en las siguientes referencias: Bottomly et al., en Current Protocols in Immunology, supra, 1: 6.3.1-6.3.12; deVries et al., J. Exp. Med., 173: 1205-1211, 1991; Moreau et al., Nature, 36: 690-692, 1988; Greenberg et al., Proc. Natl. Acad. Sci. USA, 80: 2931-2938, 1983; Nordan, R., en Current Protocols in Immunology, supra, 1: 6.6.1-6.6.5; Smith et al., Proc. Natl. Acad. Sci. USA, 83: 1857-1861, 1986; Bennett et al., en Current Protocols in Immunology, supra, 1: 6.15.1; Ciarletta et al., en Current Protocols in Immunology, supra, 1: 6.13.1.

Las proteínas codificadas por los ADNcs también pueden analizarse para determinar su capacidad para regular las respuestas de células T a antígenos. Muchos análisis para dicha actividad son

familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias.

- 5 Capítulo 3 (In Vitro Assays for Mouse Lymphocyte Function), Capítulo 6 (Cytokines and Their Cellular Receptors) y Capítulo 7 (Immunologic Studies in Humans) en Current Protocols in Immunology, supra; Weinberg et al., Proc. Natl. Acad. Sci. USA, 77: 6091-6095, 1980; Weinberg et al., Eur. J. Immunol., 11: 405-411, 1981; Takai et al., J. Immunol., 137: 3494-3500, 1986; Takai et al., J. Immunol., 140: 508-512, 1988.

- 10 Las proteínas que muestran actividad de citoquinas, de proliferación celular o de diferenciación celular pueden formularse a continuación como productos farmacéuticos y utilizarse para tratar trastornos clínicos en los que la inducción de la proliferación o diferenciación celular es beneficiosa. Como alternativa, como se describe con más detalle a continuación, los genes que codifican estas proteínas o ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en las células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

EJEMPLO 33

- 15 Análisis de las Proteínas Expresadas a partir de ADNcs Extendidos o Porciones de Éstos para Determinar su Actividad como Reguladores del Sistema Inmunológico

- 20 Las proteínas codificadas por los ADNcs también pueden evaluarse para determinar sus efectos como inmunorreguladores. Por ejemplo, las proteínas pueden evaluarse para determinar su actividad para influir en la citotoxicidad de timocitos o esplenocitos. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis 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 et al., eds., Greene Publishing Associates and Wiley-Interscience; Herrmann et al., Proc. Natl. Acad. Sci. USA, 78: 2488-2492, 1981; Herrmann et al., J. Immunol., 128: 1968-1974, 1982; Handa et al., J. Immunol., 135: 1564-1572, 1985; Takai et al., J. Immunol., 137: 3494-3500, 1986; Takai et al., J. Immunol., 140: 508-512, 1988; Bowman et al., J. Virology, 61: 1992-1998; Bertagnolli et al., Cell Immunol., 133: 327-341, 1991; Brown et al., J. Immunol., 153: 3079-3092, 1994.

- 30 Las proteínas codificadas por los ADNcs también pueden evaluarse para determinar sus efectos sobre respuestas de inmunoglobulinas dependientes de células T y cambio de isotipo. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: Maliszewski, J. Immunol., 144: 3028-3033, 1990; Mond et al., en Current Protocols in Immunology, 1: 3.8.1-3.8.16, supra.

- 35 Las proteínas codificadas por los ADNcs también pueden evaluarse para determinar sus efectos sobre células efectoras inmunológicas, incluyendo su efecto sobre células Th1 y linfocitos citotóxicos. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis 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, supra; Takai et al., J. Immunol., 137: 3494-3500, 1986; Takai et al., J. Immunol., 140: 508-512, 1988; Bertagnolli et al., Cell Immunol., 149: 3778-3783, 1992.

- 40 Las proteínas codificadas por los ADNcs también pueden evaluarse para determinar sus efectos sobre la activación de células T indiferenciadas mediada por células dendríticas. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: Guery et al., J. Immunol., 134: 536-544, 1995; Inaba et al., J. Exp. Med., 173: 549-559, 1991; Macatonia et al., J. Immunol., 154: 5071-5079, 1995; Porgador et al., J. Exp. Med., 182: 255-260, 1995; Nair et al., J. Virol., 67: 4062-4069, 1993; Huang et al., Science, 264: 961-965, 1994; Macatonia et al., J. Exp. Med., 169: 1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation, 94: 797-807, 1994; e Inaba et al., J. Exp. Med., 172: 631-640, 1990.

- 50 Las proteínas codificadas por los ADNcs también pueden evaluarse para determinar su influencia sobre el tiempo de vida de los linfocitos. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: Darzynkiewicz et al., Cytometry, 13: 795-808, 1992; Gorczyca et al., Leukemia, 7: 659-670, 1993; Gorczyca et al., Cancer Res., 53: 1945-1951, 1993; Itoh et al., Cell, 66: 233-243, 1991; Zacharchuk, J. Immunol., 145: 4037-4045, 1990; Zamai et al., Cytometry, 14: 891-897, 1993; Gorczyca et al., Int. J. Oncol., 1: 639-648, 1992.

- 55 Las proteínas codificadas por los ADNcs también pueden evaluarse para determinar su influencia sobre las etapas tempranas del desarrollo y compromiso de células T. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: Antica et al., Blood, 84: 111- 117, 1994; Fine et al., Cell. Immunol., 155: 111-

122, 1994; Galy et al., Blood, 84: 2770-2778, 1995; Toki et al., Proc. Natl. Acad. Sci. USA, 88: 7548-7551, 1991.

5 Aquellas proteínas que muestran actividad como reguladores del sistema inmunológico pueden formularse a continuación como productos farmacéuticos y utilizarse para tratar trastornos clínicos en los que la regulación de la actividad inmunológica es beneficiosa. Por ejemplo, la proteína puede ser útil en el tratamiento de diversas deficiencias y trastornos inmunológicos (incluyendo inmunodeficiencia combinada grave), por ejemplo, para regular (hacia arriba o hacia abajo) el crecimiento y proliferación de linfocitos T y/o B, así como efectuar la actividad citolítica de células NK y otras poblaciones celulares. Estas deficiencias inmunológicas pueden ser genéticas o provocadas por virus (por ejemplo, VIH), así como infecciones bacterianas o fúngicas, o pueden ser el resultado de trastornos autoinmunológicos. Más específicamente, las enfermedades infecciosas provocadas por virus, bacterias, hongos u otras infecciones pueden tratarse utilizando una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención, incluyendo infecciones por VIH, virus de la hepatitis, herpes virus, micobacterias, Leishmania spp., Plasmodium y diversas infecciones fúngicas, tal como candidiasis. Por supuesto, a este respecto, una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también puede ser útil cuando sea deseable, en general, un refuerzo del sistema inmunológico, es decir, en el tratamiento del cáncer.

20 Como alternativa, las proteínas codificadas por los ADNcs extendidos derivados de las ESTs 5' de la presente invención pueden utilizarse en el tratamiento de trastornos autoinmunológicos incluyendo, por ejemplo, enfermedad del tejido conectivo, esclerosis múltiple, lupus eritematoso generalizado, artritis reumatoide, inflamación pulmonar autoinmunológica, síndrome de Guillian-Barre, tiroiditis autoinmunológica, diabetes mellitus dependiente de insulina, miastenia grave, enfermedad del receptor frente al injerto, y enfermedad ocular inflamatoria autoinmunológica. Esta proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también puede ser útil en el tratamiento de reacciones y trastornos alérgicos, tales como asma (en particular, asma alérgica) u otros problemas respiratorios. Otros trastornos, en los que se desea una supresión inmunológica (incluyendo, por ejemplo, el trasplante de órganos) también pueden ser tratables utilizando una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención.

30 Utilizando las proteínas de la invención también es posible regular las respuestas inmunológicas, hacia arriba o hacia abajo.

La infrarregulación puede implicar inhibir o bloquear una respuesta inmunológica que ya está en progreso, o puede implicar prevenir la inducción de una respuesta inmunológica. Las funciones de células T activadas pueden inhibirse mediante la supresión de las respuestas de las células T, o la inducción de una tolerancia específica en células T, o ambas. La inmunosupresión de las respuestas de las células T es, en general, un proceso activo no específico de antígeno que requiere una exposición continuada de las células T al agente supresor. La tolerancia, que implica la inducción de la no respuesta o anergia en células T, puede distinguirse de la inmunosupresión porque, en general, es específica de antígeno y persiste tras el final de la exposición al agente tolerizante. En la práctica, la tolerancia puede demostrarse por la falta de respuesta de células T tras la reexposición a un antígeno específico en ausencia del agente tolerizante.

45 La infrarregulación o prevención de una o más funciones del antígeno (incluyendo, sin limitación, las funciones del antígeno de linfocitos B, tales como, por ejemplo, la coestimulación de B7), por ejemplo, la prevención de unos niveles altos de síntesis de linfoquinas por las células T activadas, será útil en situaciones de trasplante de tejido, piel y órganos, y en la enfermedad del receptor frente al injerto (GVHD). Por ejemplo, el bloqueo de la función de las células T debería producir una menor destrucción del tejido en un trasplante de tejido. De forma típica, en trasplantes de tejidos, el rechazo del trasplante se inicia mediante su reconocimiento como extraño por las células T, seguido de una reacción inmunológica que destruye el trasplante. La administración de una molécula que inhibe o bloquea la interacción de un antígeno de linfocito B7 con su ligando o ligandos naturales en células inmunológicas (tales como una forma soluble y monomérica de un péptido que tiene actividad B7-2 por sí solo o junto con una forma monomérica de un péptido que tiene actividad de otro antígeno de linfocitos B (por ejemplo, B7-1, B7-3) o anticuerpo de bloqueo), antes del trasplante, puede conducir a la unión de la molécula al ligando o los ligandos naturales en las células inmunológicas sin transmitir la correspondiente señal coestimuladora. El bloqueo de la función de antígeno de linfocitos B, en este caso, evita la síntesis de citoquinas por las células inmunológicas, tales como células T, y, por tanto, actúa como un inmunosupresor. Además, la falta de coestimulación también puede ser suficiente para anergizar las células T, induciendo, con ello, la tolerancia en un sujeto. La inducción de tolerancia a largo plazo por reactivos que bloquean antígenos de linfocitos B puede evitar la necesidad de la administración repetida de estos reactivos de bloqueo. Para lograr una tolerancia o inmunosupresión suficiente en un sujeto, también puede ser necesario bloquear la función de una combinación de antígenos de linfocitos B.

La eficacia de los reactivos de bloqueo particulares para prevenir el rechazo de trasplantes de órganos o GVHD puede evaluarse utilizando modelos animales que son predictivos de la eficacia en seres humanos. Los ejemplos de los sistemas apropiados que pueden utilizarse incluyen injertos cardíacos alogénicos en ratas e injertos de células de islotes pancreáticos xenogénicos en ratones, ambos los cuales se han utilizado para estudiar los efectos inmunosupresores de proteínas de fusión CTLA41g in vivo, como describen Lenschow et al., *Science*, 257: 789-792, 1992, y Turka et al., *Proc. Natl. Acad. Sci. USA*, 89: 11102-11105, 1992. Además, pueden usarse modelos murinos de GVHD (véase Paul ed., *Fundamental Immunology*, Raven Press, Nueva York, 1989, págs. 846-847) para determinar el efecto del bloqueo de la función del antígeno de linfocitos B in vivo sobre el desarrollo de esta enfermedad.

El bloqueo de la función del antígeno también puede ser terapéuticamente eficaz para tratar enfermedades autoinmunitarias. Muchas enfermedades autoinmunitarias son el resultado de la activación inapropiada de células T que son reactivas frente al propio tejido y que estimulan la producción de citoquinas y autoanticuerpos implicados en la patología de las enfermedades. La prevención de la activación de las células T autorreactivas puede reducir o eliminar los síntomas de la enfermedad. La administración de reactivos que bloquean la coestimulación de las células T interrumpiendo las interacciones receptor/ligando de antígenos de linfocitos B puede utilizarse para inhibir la activación de las células T y prevenir la producción de autoanticuerpos o citoquinas derivadas de células T que están potencialmente implicados en el proceso de la enfermedad. Además, los reactivos de bloqueo pueden inducir tolerancia específica de antígeno de células T autorreactivas, lo que puede conducir a un alivio a largo plazo de la enfermedad. La eficacia de los reactivos de bloqueo en la prevención o alivio de enfermedades autoinmunitarias puede determinarse utilizando una serie de modelos animales bien caracterizados de enfermedades autoinmunitarias humanas. Los ejemplos incluyen encefalitis autoinmunitaria experimental murina, lupus eritematoso generalizado en ratones MRL/pr/pr o ratones híbridos NZB, artritis por colágeno autoinmunitaria murina, diabetes mellitus en ratones OD y ratas BB, y miastenia grave experimental murina (véase Paul ed., *supra*, págs. 840-856).

La sobreexpresión de una función del antígeno (preferiblemente una función de antígeno de linfocitos B), como un medio para sobreexpresar las respuestas inmunitarias, también puede ser útil en terapia. La sobreexpresión de las respuestas inmunitarias puede implicar potenciar una respuesta inmunitaria existente, o provocar una respuesta inmunitaria inicial como se muestra en los siguientes Ejemplos. Por ejemplo, la potenciación de una respuesta inmunitaria mediante la estimulación de la función del antígeno de linfocitos B puede ser útil en casos de infección vírica. Además, enfermedades víricas sistémicas, tales como la gripe, el resfriado común y la encefalitis, pueden aliviarse mediante la administración de una forma estimuladora de antígenos de linfocitos B por vía sistémica.

Como alternativa, las respuestas inmunitarias antivíricas pueden potenciarse en un paciente infectado retirando células T del paciente, coestimulando las células T in vitro con APC marcadas con antígenos víricos que expresan un péptido codificado por los ADNcs extendidos derivados de las ESTs 5' de la presente invención o junto con una forma estimuladora de un péptido soluble codificado por los ADNcs extendidos derivados de las ESTs 5' de la presente invención, y reintroduciendo las células T cebadas in vitro en el paciente. Las células infectadas serán capaces ahora de dirigir una señal coestimuladora a las células T in vivo, activando, con ello, las células T.

En otra aplicación, la sobreexpresión o potenciación de la función del antígeno (preferiblemente una función de antígeno de linfocitos B) puede ser útil en la inducción de inmunidad tumoral. Células tumorales (por ejemplo, sarcoma, melanoma, linfoma, leucemia, neuroblastoma, carcinoma) transfectadas con un ácido nucleico que codifica al menos un péptido codificado por los ADNcs extendidos derivados de las ESTs 5' de la presente invención pueden administrarse a un sujeto para superar la tolerancia específica de tumor en el sujeto. Si se desea, la célula tumoral puede transfectarse para expresar una combinación de péptidos. Por ejemplo, células tumorales obtenidas de un paciente pueden transfectarse ex vivo con un vector de expresión que dirige la expresión de un péptido que tiene sólo actividad de tipo B7-2, o junto con un péptido que tiene actividad de tipo B7-1 y/o 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. Como alternativa, pueden utilizarse técnicas de terapia génica para seleccionar una célula tumoral para la transfección in vivo.

La presencia del péptido codificado por los ADNc extendidos derivados de las ESTs 5' de la presente invención que tiene la actividad de uno o varios antígenos de linfocitos B sobre la superficie de la célula tumoral proporciona la señal de coestimulación necesaria para que las células T induzcan una respuesta inmunitaria mediada por células T contra las células tumorales transfectadas. Además, las células tumorales que carecen o que no reexpresan cantidades suficientes de moléculas de MHC de clase I o MHC de clase II pueden transfectarse con ácidos nucleicos que codifican toda o una porción (por ejemplo, una porción truncada del dominio citoplásmico) de una cadena α y β 2 de microglobulina MHC de clase I o una cadena α de MHC de clase II y una cadena β de MHC de clase

II, para expresar, con ello, proteínas MHC de clase I o MHC de clase II sobre la superficie celular, respectivamente. La expresión de las moléculas apropiadas de MHC de clase I o clase II junto con un péptido que tiene la actividad de un antígeno de linfocitos B (por ejemplo, B7-1, B7-2, B7-3) induce una respuesta inmunológica mediada por células T contra la célula tumoral transfectada. Opcionalmente, un gen que codifica un constructo antisentido que bloquea la expresión de una proteína asociada con MHC de clase I, tal como la cadena invariable, también puede cotransfectarse con un ADN que codifica un péptido que tiene la actividad de un antígeno de linfocito B para estimular la presentación de antígenos asociados con tumores e inducir una inmunidad específica de tumor. Por tanto, la inducción de una respuesta inmunológica mediada por células T en un sujeto humano puede ser suficiente para superar la tolerancia específica de tumor en el sujeto. Como alternativa, como se describe con más detalle a continuación, los genes que codifican estas proteínas reguladoras del sistema inmunológico o los ácidos nucleicos que regulan la expresión de dichas proteínas pueden introducirse en células anfitrionas adecuadas para aumentar o disminuir la expresión de las proteínas, según se desee.

15 **EJEMPLO 34**

Análisis de las Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para Determinar la Actividad Reguladora de la Hematopoyesis

Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su actividad reguladora de la hematopoyesis. Por ejemplo, puede evaluarse el efecto de las proteínas sobre la diferenciación de células pluripotenciales embrionarias. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: Johansson et al., Cell. Biol., 15: 141-151, 1995; Keller et al., Mol. Cell. Biol., 13: 473-486, 1993; McClanahan et al., Blood, 81: 2903-2915, 1993.

Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su influencia en el tiempo de vida de células pluripotenciales y la diferenciación de células pluripotenciales. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: Freshney, Methylcellulose Colony Forming Assays, en Culture of Hematopoietic Cells, Freshney et al., eds., págs. 265-268, Wiley-Liss, Inc., Nueva York, NY, 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA, 89: 5907-5911, 1992; McNiece y Briddell, en Culture of Hematopoietic Cells, supra; Neben et al., Exp. Hematol., 22: 353-359, 1994; Ploemacher y Cobblestone, en Culture of Hematopoietic Cells, supra, 1-21; Spooncer et al., Culture of Hematopoietic Cells, supra, 163-179; y Sutherland, en Culture of Hematopoietic Cells, supra, 139-162.

Aquellas proteínas que muestran actividad reguladora de la hematopoyesis puede formularse continuación como producto farmacéutico y usarse para tratar trastornos clínicos en los que la regulación de la hematopoyesis es beneficiosa, tales como el tratamiento de deficiencias de células mieloides o linfoides. La implicación en la regulación de la hematopoyesis viene indicada incluso por una actividad biológica marginal en apoyo a células formadoras de colonias o líneas celulares dependientes de factor. Por ejemplo, las proteínas que apoyan el crecimiento y proliferación de células progenitoras eritroides por sí solas, o en combinación con otras citoquinas, indica utilidad, por ejemplo, en el tratamiento de diversas anemias o para su uso junto con terapia de irradiación/quimioterapia para estimular la producción de precursores eritroides y/o células eritroides. Las proteínas que apoyan el crecimiento y proliferación de células mieloides, tales como granulocitos y monocitos/macrófagos (es decir, la actividad CSF tradicional) pueden ser útiles, por ejemplo, junto con quimioterapia, para prevenir o tratar la consiguiente mielosupresión. Las proteínas que apoyan el crecimiento y proliferación de megacariocitos y, por consiguiente, de plaquetas, permiten la prevención o el tratamiento de diversos trastornos plaquetarios, tales como la trombocitopenia y, en general, pueden utilizarse en lugar o como complemento de transfusiones de plaquetas. Las proteínas que apoyan el crecimiento y proliferación de células pluripotenciales hematopoyéticas que son capaces de madurar para formar cualquiera y todas las células hematopoyéticas mencionadas anteriormente pueden encontrar utilidad terapéutica, por tanto, en diversos trastornos de células pluripotenciales (tales como los que se tratan habitualmente con trasplantes, incluyendo, sin limitación, anemia aplásica y hemoglobinuria nocturna paroxística), así como en la repoblación del compartimento de células pluripotenciales después de una terapia de irradiación/quimioterapia, *in vivo* o *ex vivo* (es decir, junto con trasplante de médula ósea o con trasplante (homólogo o heterólogo) de células progenitoras periféricas) como células normales o manipuladas genéticamente para la terapia génica. Como alternativa, como se describe con más detalle a continuación, los genes que codifican proteínas de actividad reguladora de la hematopoyesis o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en las células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

EJEMPLO 35Análisis de Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para la Regulación del Crecimiento de Tejidos

- 5 Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su efecto sobre el crecimiento de tejidos. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en la publicación de Patente Internacional Núm. WO95/16035, publicación de Patente Internacional Núm. WO95/05846, y publicación de Patente Internacional Núm. WO91/07491.
- 10 Los análisis para la actividad de curación de heridas incluyen, sin limitación, los descritos en: Winter, Epidermal Wound Healing, págs. 71-112, Maibach y Rovee, eds., Year Book Medical Publishers, Inc., Chicago, modificado por Eaglstein y Mertz, J. Invest. Dermatol., 71: 382-384, 1978.
- 15 Aquellas proteínas que están implicadas en la regulación del crecimiento de tejidos pueden formularse después como productos farmacéuticos y utilizarse para tratar trastornos clínicos en los que la regulación del crecimiento de tejidos es beneficiosa. Por ejemplo, una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también puede tener utilidad en composiciones utilizadas para crecimiento o regeneración de tejido de hueso, cartílago, tendón, ligamento y/o nervio, así como para la curación de heridas y la reparación y sustitución de tejido, y en el tratamiento de quemaduras, incisiones y úlceras.
- 20 Una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención, que induce el crecimiento de cartílago y/o hueso en circunstancias en las que el hueso no se ha formado con normalidad, tiene aplicación en la curación de fracturas de huesos y daños o defectos en cartílagos en seres humanos y otros animales. Esta preparación que emplea una proteína de la invención puede tener uso profiláctico en la reducción de fracturas abiertas y cerradas, y también en una mejor fijación de articulaciones artificiales. La síntesis de hueso *de novo* inducida por un agente osteogénico contribuye a la reparación de defectos craneofaciales inducidos por resección oncológica, inducidos por traumatismos o congénitos, y también es útil en cirugía plástica cosmética.
- 25 Una proteína de esta invención también puede utilizarse en el tratamiento de la enfermedad periodontal, y en otros procesos de reparación de dientes. Estos 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 u osteoartritis, tal como mediante la estimulación de la reparación de hueso y/o cartílago, o bloqueando la inflamación o los procesos de destrucción de tejidos (actividad colagenasa, actividad de osteoclastos, etc.) mediada por procesos inflamatorios.
- 30 Otra categoría de actividad de regeneración de tejidos que puede atribuirse a la proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención es la formación de tendones/ligamentos. Una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención, que induce tejido de tipo tendón/ligamento u otra formación de tejido en circunstancias en las que dicho tejido no se ha formado con normalidad, tiene aplicación en la curación de desgarros de tendones o ligamentos, deformidades y otros defectos de tendones o ligamentos en seres humanos y otros animales. Esta preparación que emplea proteínas inductoras de tejido de tipo tendón/ligamento puede tener uso profiláctico para prevenir daños en tejido de tendón o ligamento, así como uso en una mejor fijación de tendones o ligamentos al hueso u otros tejidos, y en la reparación de defectos del tejido de tendón o ligamento. La formación *de novo* de tejido de tendón/ligamento inducida por una composición codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención contribuye a la reparación de defectos de tendones o ligamentos de origen congénito, traumático o de otro origen, y también es útil en cirugía plástica cosmética para la unión o reparación de tendones o ligamentos. Las composiciones codificadas por los ADNcs extendidos derivados de las ESTs 5' de la presente invención pueden proporcionar un entorno para atraer células formadoras de tendones o ligamentos, estimular el crecimiento de células formadoras de tendones o ligamentos, inducir la diferenciación de progenitores de células formadoras de tendones o ligamentos, o inducir el crecimiento de progenitores o células de tendones/ligamentos *ex vivo* para devolverlos *in vivo* para realizar la reparación del tejido. Las composiciones de la invención también pueden ser útiles en el tratamiento de la tendinitis, síndrome del túnel carpiano y otros defectos de tendones o ligamentos. Las composiciones también pueden incluir una matriz y/o agente secuestrante apropiados como vehículos, como se conoce en la técnica.
- 35 La proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también puede ser útil para la proliferación de células neurales y para la regeneración de tejido nervioso y cerebral, es decir, para el tratamiento de enfermedades del sistema nervioso central y periférico y neuropatías, así como trastornos mecánicos y traumáticos, que implican degeneración, muerte o traumatismo de células neurales o tejido nervioso. Más específicamente, una proteína puede
- 40
- 45
- 50
- 55
- 60

- usarse en el tratamiento de enfermedades del sistema nervioso periférico, tales como lesiones de nervios periféricos, neuropatía periférica y neuropatías localizadas, y enfermedades del sistema nervioso central, tales como enfermedad de Alzheimer, enfermedad de Parkinson, enfermedad de Huntington, esclerosis lateral amiotrófica y síndrome de Shy-Drager. Otros trastornos que pueden tratarse según la presente invención incluyen trastornos mecánicos y traumáticos, tales como trastornos de la médula espinal, traumatismos cefálicos y enfermedades cerebrovasculares, tales como el ictus. Las neuropatías periféricas que surgen como resultado de quimioterapia u otras terapias médicas también pueden ser tratables utilizando una proteína de la invención.
- Las proteínas de la invención también pueden ser útiles para estimular un cierre mejor o más rápido de heridas que no curan, incluyendo, sin limitación, úlceras de presión, úlceras asociadas con insuficiencia vascular, heridas quirúrgicas y traumáticas, y similares.
- Se espera que una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también pueda mostrar actividad para la generación o regeneración de otros tejidos, tales como tejidos de órganos (incluyendo, por ejemplo, páncreas, hígado, intestino, riñón, piel, endotelio), de músculo (liso, esquelético o cardíaco) y vascular (incluyendo endotelio vascular), o para estimular el crecimiento de células que comprenden dichos tejidos. Parte de los efectos deseados puede producirse mediante la inhibición o modulación de la formación de escaras fibróticas para permitir que el tejido normal se genere. Una proteína de la invención también puede mostrar actividad angiogénica.
- Una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también puede ser útil para la protección o regeneración del intestino y el tratamiento de fibrosis pulmonar o hepática, lesiones por reperfusión en diversos tejidos, y trastornos que se producen como resultado de lesiones sistémicas de citoquinas.
- Una proteína codificada por los ADNc extendidos derivados de las ESTs 5' de la presente invención también puede ser útil para estimular o inhibir la diferenciación de los tejidos descritos anteriormente a partir de tejidos o células precursoras; o para inhibir el crecimiento de los tejidos descritos anteriormente.
- Como alternativa, como se describe con más detalle a continuación, los genes que codifican las proteínas de actividad reguladora del crecimiento de tejidos o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

EJEMPLO 36

Análisis de Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para la Regulación de Hormonas Reproductoras

- Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su capacidad para regular hormonas reproductoras, tales como la hormona estimulante del folículo. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias,
- Vale et al., Endocrinol., 91: 562-572, 1972; Ling et al., Nature, 321: 779-782, 1986; Vale et al., Nature, 321: 776-779, 1986; Mason et al., Nature, 318: 659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA, 83: 3091-3095, 1986, capítulo 6.12 en Current Protocols in Immunology, Coligan et al. eds., Greene Publishing Associates and Wiley-Interscience; Taub et al., J. Clin. Invest., 95: 1370-1376, 1995; Lind et al., APMIS, 103: 140-146, 1995; Muller et al., Eur. J. Immunol., 25: 1744-1748; Gruber et al., J. Immunol., 152: 5860-5867, 1994; Johnston et al., J. Immunol., 153: 1762-1768, 1994.
- Aquellas proteínas que muestran actividad como hormonas reproductoras o reguladores del movimiento celular pueden formularse a continuación como productos farmacéuticos y utilizarse para tratar trastornos clínicos en los que la regulación de las hormonas reproductoras es beneficiosa. Por ejemplo, una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también puede mostrar actividades relacionadas con activina o 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 FSH. Por tanto, una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención, por sí sola o en heterodímeros con un miembro de la familia de inhibinas- α , puede ser útil como anticonceptivo basándose en la capacidad de las inhibinas para disminuir la fertilidad en hembras de mamífero y disminuir la espermatogénesis en machos de mamífero. La administración de cantidades suficientes de otras inhibinas puede inducir la infertilidad en estos mamíferos. Como alternativa, la proteína de la invención, como un homodímero o como un heterodímero con otras subunidades de una proteína del grupo de las inhibinas-B, puede ser útil como producto terapéutico inductor de la fertilidad, basándose en la capacidad de las moléculas de activina para estimular la

liberación de FSH de las células de la pituitaria anterior. Véase, por ejemplo, la Patente de EE.UU. 4.798.885.

5 Una proteína de la invención también puede ser útil para adelantar la aparición de la fertilidad en mamíferos sexualmente inmaduros, para aumentar el tiempo de vida del rendimiento reproductor de animales domésticos, tales como vacas, ovejas y cerdos.

Como alternativa, como se describe con más detalle a continuación, los genes que codifican las proteínas de actividad reguladora de las hormonas reproductoras o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

10 EJEMPLO 37

Análisis de Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para Determinar la Actividad Quimiotáctica/Quimiocinética

15 Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su actividad quimiotáctica/quimiocinética. Por ejemplo, una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención puede tener actividad quimiotáctica o quimiocinética (por ejemplo, puede actuar como una quimioquina) para células de mamífero, incluyendo, por ejemplo, monocitos, fibroblastos, neutrófilos, células T, células cebadas, eosinófilos, células epiteliales y/o endoteliales. Las proteínas quimiotácticas y quimiocinéticas pueden utilizarse para movilizar o atraer una población de células deseada hasta un sitio de acción deseado.

20 Las proteínas quimiotácticas o quimiocinéticas proporcionan ventajas particulares en el tratamiento de heridas y otros traumatismos 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 producir mejores respuestas inmunológicas contra el tumor o el agente infeccioso.

25 Una proteína o péptido tiene actividad quimiotáctica para una población de células particular si puede estimular, directa o indirectamente, la orientación o movimiento dirigido de esta población de células. Preferiblemente, la proteína o péptido tiene la capacidad de estimular directamente el movimiento dirigido de las células. Se puede determinar con facilidad si una proteína concreta tiene actividad quimiotáctica para una población de células empleando dicha proteína o péptido en cualquier análisis conocido para la quimiotaxis de células.

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

35 Los análisis para determinar la actividad quimiotáctica (que identifican proteínas que inducen o previenen la quimiotaxis) consisten en análisis que miden la capacidad de una proteína para inducir la migración de células a través de la membrana, así como la capacidad de una proteína para inducir la adherencia de una población de células a otra población de células. Los análisis adecuados para el movimiento y adherencia incluyen, sin limitación, los descritos en: Current Protocols in Immunology, ed. por Coligan, Kruisbeek, Margulies, Shevach y Strober, pub. Green Publishing Associates and Wiley-Interscience, capítulo 6.12: 6.12.1-6.12.28; Taub et al., J. Clin. Invest., 95: 1370-1376, 1995; Lind et al., APMIS, 103: 140-146, 1995; Mueller et al., Eur. J. Immunol., 25: 1744-1748; Gruber et al., J. Immunol., 152: 5860-5867, 1994; Johnston et al., J. Immunol., 153: 1762-1768, 1994.

40

EJEMPLO 38

Análisis de Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para la Regulación de la Coagulación Sanguínea

45 Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar sus efectos sobre la coagulación sanguínea. Numerosos análisis para dicha actividad son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: Linet et al., J. Clin. Pharmacol., 26: 131- 140, 1986; Burdick et al., Thrombosis Res., 45: 413-419, 1987; Humphrey et al., Fibrinolysis, 5: 71-79, 1991; Schaub, Prostaglandins, 35: 467-474, 1988.

50

55 Aquellas proteínas que están implicadas en la regulación de la coagulación sanguínea pueden formularse a continuación como productos farmacéuticos y utilizarse para tratar trastornos clínicos en los que la regulación de la coagulación sanguínea es beneficiosa. Por ejemplo, una proteína de la invención también puede mostrar actividad hemostática o trombolítica. Como resultado, se espera que esta proteína sea útil en el tratamiento de diversos trastornos de la coagulación (incluyendo trastornos hereditarios, tales como hemofilias), o para potenciar la coagulación y otros acontecimientos hemostáticos en el tratamiento de heridas producidas por traumatismos, cirugía u otras causas. Una

- proteína de la invención también puede ser útil para disolver o inhibir la formación de trombos y para el tratamiento y prevención de trastornos resultantes de éstos (tales como infarto de vasos cardíacos y del sistema nervioso central (por ejemplo, ictus)). Como alternativa, como se describe con más detalle a continuación, los genes que codifican las proteínas con actividad de coagulación sanguínea o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

EJEMPLO 39

Análisis de Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para la Implicación en Interacciones de Receptor/Ligando

- Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su implicación en interacciones de receptor/ligando. Numerosos análisis para determinar esta implicación son familiares para los expertos en la técnica, incluyendo los análisis descritos en las siguientes referencias: capítulo 7, 7.28.1-7.29.22 en Current Protocols in Immunology, Coligan et al. eds., Greene Publishing Associates and Wiley-Interscience; Takai et al., Proc. Natl. Acad. Sci. USA, 84: 6864-6868, 1987; Bierer et al., J. Exp. Med., 168: 1145-1156, 1988; Rosenstein et al., J. Exp. Med., 169: 149-160, 1989; Stoltenberg et al., J. Immunol. Methods, 175: 59-68, 1994; Stitt et al., Cell, 80: 661-670, 1995; Gyuris et al., Cell, 75: 791-803, 1993.

- Por ejemplo, las proteínas codificadas por los ADNcs extendidos derivados de las ESTs 5' de la presente invención también pueden mostrar actividad como receptores, ligandos de receptores, o inhibidores o agonistas de interacciones de receptor/ligando. Los ejemplos de estos receptores y ligandos incluyen, sin limitación, receptores de citoquinas y sus ligandos, receptores de quinasas y sus ligandos, receptores de fosfatasa y sus ligandos, receptores implicados en interacciones de célula-célula y sus ligandos (incluyendo, sin limitación, moléculas de adherencia celular (tales como selectinas, integrinas y sus ligandos), y parejas de receptor/ligando implicadas en la presentación de antígenos, el reconocimiento de antígenos y el desarrollo de respuestas inmunológicas celulares y humorales). Los receptores y ligandos también son útiles para seleccionar péptidos o moléculas pequeñas de inhibidores potenciales de la interacción de receptor/ligando pertinente. Una proteína codificada por los ADNcs extendidos derivados de las ESTs 5' de la presente invención (incluyendo, sin limitación, fragmentos de receptores y ligandos) puede, en sí misma, ser útil como inhibidor de interacciones de receptor/ligando. Como alternativa, como se describe con más detalle a continuación, los genes que codifican las proteínas implicadas en interacciones de receptor/ligando o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

EJEMPLO 40

- Análisis de Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para determinar la Actividad Antiinflamatoria

- Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su actividad antiinflamatoria. La actividad antiinflamatoria puede lograrse proporcionando un estímulo a células implicadas en la respuesta inflamatoria, inhibiendo o estimulando las interacciones célula-célula (tales como, por ejemplo, la adherencia celular), inhibiendo o estimulando la quimiotaxis de células implicadas en el proceso inflamatorio, inhibiendo o estimulando la extravasación celular, o estimulando o suprimiendo la producción de otros factores que inhiben o estimulan, más directamente, una respuesta inflamatoria. Las proteínas que muestran estas actividades pueden utilizarse para tratar trastornos inflamatorios, incluyendo trastornos crónicos o agudos, incluyendo, sin limitación, la inflamación asociada con infección (tal como choque séptico, sepsis o síndrome de respuesta inflamatoria generalizada), lesiones por isquemia-reperusión, letalidad por endotoxinas, artritis, rechazo hiperagudo mediado por el complemento, nefritis, lesiones pulmonares inducidas por citoquinas o quimioquinas, enfermedad inflamatoria intestinal, enfermedad de Crohn, o trastorno resultante de la sobreproducción de citoquinas, tales como TNF o IL-1. Las proteínas de la invención también pueden ser útiles para tratar la anafilaxis e hipersensibilidad frente a una sustancia o material antigénico. Como alternativa, como se describe con más detalle a continuación, los genes que codifican proteínas de actividad antiinflamatoria o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

EJEMPLO 41

Análisis de Proteínas Expresadas a partir de los ADNcs Extendidos o Porciones de Éstos para determinar la Actividad Inhibidora de Tumores

Las proteínas codificadas por los ADNcs extendidos o porciones de éstos también pueden evaluarse para determinar su actividad inhibidora de tumores. Además de las actividades descritas

anteriormente para el tratamiento inmunológico o la prevención de tumores, una proteína de la invención puede mostrar otras actividades antitumorales. Una proteína puede inhibir el crecimiento tumoral directa o indirectamente (tal como, por ejemplo, mediante ADCC). Una proteína puede mostrar su actividad inhibidora de tumores actuando sobre tejido de tumor o tejido precursor de tumor, inhibiendo la formación de tejidos necesarios para soportar el crecimiento del tumor (tal como, por ejemplo, inhibiendo la angiogénesis), provocando la producción de otros factores, agentes o tipos celulares que inhiben el crecimiento del tumor, o suprimiendo, eliminando o inhibiendo factores, agentes o tipos celulares que estimulan el crecimiento del tumor. Como alternativa, como se describe con más detalle a continuación, los genes que codifican proteínas de actividad de inhibición de tumores o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

Una proteína de la invención también puede mostrar una o más de las siguientes actividades o efectos adicionales: inhibiendo el crecimiento, infección o función o eliminando agentes infecciosos, incluyendo, sin limitación, bacterias, virus, hongos u otros parásitos; afectando (suprimiendo o potenciando) a características corporales, incluyendo, sin limitación, la altura, peso, color del pelo, color de los ojos, piel, proporción de grasa a tejido magro u otras pigmentaciones de tejidos, o el tamaño o forma de un órgano o parte del cuerpo (tal como, por ejemplo, aumento o disminución del busto, cambio en la forma del hueso); afectando a los biorritmos o ritmos o ciclos circadianos; afectando a la fertilidad de sujetos macho o hembra; afectando al metabolismo, catabolismo, anabolismo, procesamiento, utilización, almacenamiento o eliminación de grasas, lípidos, proteínas, carbohidratos, vitaminas, minerales, cofactores u otros factores o componentes nutricionales de la dieta; afectando a las características del comportamiento, incluyendo, sin limitación, el apetito, ánimo, estrés, cognición (incluyendo trastornos cognitivos), depresión (incluyendo trastornos depresivos) y comportamientos violentos; proporcionando efectos analgésicos u otros efectos reductores del dolor; estimulando la diferenciación y el crecimiento de células pluripotenciales embrionarias en linajes diferentes de los linajes hematopoyéticos; actividad hormonal o endocrina; en el caso de enzimas, corrigiendo deficiencias de las enzimas y tratando enfermedades relacionadas con las deficiencias; tratamiento de enfermedades hiperproliferativas (tales como, por ejemplo, psoriasis); actividad de tipo inmunoglobulina (tal como, por ejemplo, la capacidad de unir antígenos o complementos); y la capacidad para actuar como antígeno en una composición de vacuna para producir una respuesta inmunológica contra dicha proteína u otro material o entidad que presenta reactividad cruzada con dicha proteína. Como alternativa, como se describe con más detalle a continuación, los genes que codifican proteínas implicadas en cualquiera de las actividades mencionadas anteriormente o los ácidos nucleicos que regulan la expresión de estas proteínas pueden introducirse en células anfitrionas apropiadas para aumentar o disminuir la expresión de las proteínas, según se desee.

EJEMPLO 42

Identificación de Proteínas que Interaccionan con Polipéptidos codificados por los ADNcs Extendidos

Las proteínas que interaccionan con los polipéptidos codificados por los ADNcs derivados de las ESTs 5', o fragmentos de éstos, tales como proteínas de receptores, pueden identificarse utilizando sistemas de dos híbridos, tales como Matchmaker Two Hybrid System 2 (nº de catálogo K1604-1, Clontech). Como se describe en el manual que acompaña al kit, los ADNcs derivados de las ESTs 5', o fragmentos de éstos, se insertan en un vector de expresión, de forma que estén en marco con el ADN que codifica el dominio de unión a ADN del activador transcripcional de levadura GAL4. Los ADNcs en un banco de ADNc que codifica proteínas que pueden interaccionar con los polipéptidos codificados por los ADNcs extendidos, o fragmentos de éstos, se insertan en un segundo vector de expresión, de forma que estén en marco con el ADN que codifica el dominio de activación de GAL4. Los dos plásmidos de expresión se transforman en levaduras y las levaduras se cultivan en placa en medio de selección que selecciona la expresión de marcadores seleccionables sobre cada uno de los vectores de expresión, así como la expresión dependiente de GAL4 del gen HIS3. Los transformantes capaces de crecer sobre un medio que carece de histidina se seleccionan para la expresión de lacZ dependiente de GAL4. Aquellas células que son positivas en la selección de histidina y el ensayo de lacZ contienen plásmidos que codifican proteínas que interaccionan con el polipéptido codificado por los ADNcs extendidos, o porciones de éstos.

Como alternativa, el sistema descrito en Lustig et al., Methods in Enzymology, 283: 83-99, 1997, y en la Patente de EE.UU. nº 5.654.150, puede utilizarse para identificar moléculas que interaccionan con los polipéptidos codificados por los ADNcs extendidos. En estos sistemas, las reacciones de transcripción in vitro se realizan en una agrupación de vectores que contienen insertos de ADNc extendido clonados aguas abajo de un promotor que dirige la transcripción in vitro. Las agrupaciones resultantes de ARNm se introducen en oocitos de *Xenopus laevis*. Los oocitos se analizan a continuación en busca de una actividad deseada.

Como alternativa, los productos de la transcripción *in vitro* agrupados producidos como se describió anteriormente pueden traducirse *in vitro*. Los productos de la traducción *in vitro* agrupados pueden analizarse en busca de una actividad deseada o en busca de la interacción con un polipéptido conocido.

5 Pueden descubrirse proteínas u otras moléculas que interaccionan con polipéptidos codificados por los ADNcs extendidos mediante una diversidad de otras técnicas. En un método, pueden construirse columnas de afinidad que contienen el polipéptido codificado por el ADNc extendido o una porción de éste. En algunas versiones de este método, la columna de afinidad contiene proteínas quiméricas en las que la proteína codificada por el ADNc extendido o una porción de éste se fusiona con glutatión S-transferasa. Una mezcla de proteínas celulares o agrupación de proteínas expresadas como se describió anteriormente se aplica a la columna de afinidad. Las proteínas que interaccionan con el polipéptido unido a la columna pueden aislarse y analizarse a continuación en un gel de electroforesis 2-D, como describen Ramunsen et al., en *Electrophoresis*, 18: 588-598, 1997. Como alternativa, las proteínas retenidas en la columna de afinidad pueden purificarse por métodos basados en electroforesis y secuenciarse. El mismo método puede emplearse para aislar anticuerpos, para seleccionar productos de presentación en fagos, o para seleccionar anticuerpos humanos de presentación en fagos.

Las proteínas que interaccionan con polipéptidos codificados por los ADNcs extendidos o porciones de éstos también pueden seleccionarse utilizando un biosensor óptico, como describen Edwards y Leatherbarrow, en *Analytical Biochemistry*, 246: 1-6, 1997.

La ventaja principal del método es que permite la determinación de la velocidad de asociación entre la proteína y otras moléculas de interacción. Así, es posible seleccionar específicamente moléculas de interacción con una velocidad de asociación alta o baja. De forma típica, una molécula diana se une a la superficie del sensor (mediante una matriz de carboximetildextrano) y una muestra de moléculas de ensayo se pone en contacto con las moléculas diana. La unión de una molécula de ensayo a la molécula diana provoca un cambio en el índice de refracción y/o espesor. Este cambio se detecta mediante el biosensor, con la condición de que se produzca en el campo evanescente (que se extiende a unos pocos cientos de nanómetros de la superficie del sensor). En estos análisis de selección, la molécula diana puede ser uno de los polipéptidos codificados por los ADNcs extendidos o una porción de éstos, y la muestra de ensayo puede ser una colección de proteínas extraída de tejidos o células, una agrupación de proteínas expresadas, bancos de péptidos combinatorios y/o químicos, o péptidos de presentación en fagos. Los tejidos o células de los que se extraen las proteínas de ensayo pueden originarse de cualquier especie.

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

Para estudiar la interacción de las proteínas codificadas por los ADNcs extendidos, o porciones de éstos, con fármacos, puede utilizarse la microdiálisis acoplada a un método de HPLC descrito por Wang et al., *Chromatographia*, 44: 205-208, 1997, o el método de electroforesis capilar por afinidad descrito por Busch et al., *J. Chromatogr.*, 777: 311-328, 1997.

Los expertos en la técnica apreciarán que las proteínas expresadas a partir de los ADNcs extendidos o porciones pueden analizarse para determinar numerosas actividades además de las enumeradas de forma específica anteriormente. Por ejemplo, las proteínas expresadas pueden evaluarse para aplicaciones que implican el control y la regulación de la inflamación, proliferación o metástasis tumoral, infección, u otros trastornos clínicos. Además, las proteínas expresadas a partir de los ADNcs extendidos, o porciones de éstos, pueden ser útiles como agentes nutricionales o agentes cosméticos. Las proteínas expresadas a partir de los ADNcs extendidos, o porciones de éstos, pueden utilizarse para generar anticuerpos capaces de unirse de forma específica a la proteína expresada, o fragmentos de ésta, como se describe en el ejemplo 40 a continuación. Los anticuerpos pueden ser capaces de unirse a una proteína completa codificada por un ADNc derivado de una EST 5', una proteína madura (es decir, la proteína generada por la escisión del péptido señal) codificada por un ADNc derivado de una EST 5', o un péptido señal codificado por un ADNc derivado de una EST 5'. Como alternativa, los anticuerpos pueden ser capaces de unir fragmentos de al menos 10 aminoácidos de las proteínas codificadas por los anteriores ADNcs. En algunas realizaciones, los anticuerpos pueden ser capaces de unir fragmentos de al menos 15 aminoácidos de las proteínas codificadas por los anteriores ADNcs. En otras realizaciones, los anticuerpos pueden ser capaces de unir fragmentos de al menos 25 aminoácidos de las proteínas expresadas a partir de los ADNcs extendidos que comprenden al menos 25 aminoácidos de las proteínas codificadas por los anteriores ADNcs. En otras realizaciones, los anticuerpos pueden ser capaces de unir fragmentos de al menos 40 aminoácidos de las proteínas codificadas por los anteriores ADNcs.

60 EJEMPLO 43

Producción de un Anticuerpo Contra una Proteína Humana

Se aísla una proteína o polipéptido sustancialmente puros a partir de las células transfectadas o transformadas como se describió en el ejemplo 30. La concentración de la proteína en la preparación final se ajusta, por ejemplo, mediante concentración en un dispositivo de filtro Amicon, hasta un nivel de unos pocos $\mu\text{g/ml}$. Después puede prepararse un anticuerpo monoclonal o policlonal contra la proteína como sigue:

1. Producción de Anticuerpos Monoclonales Mediante Fusión de Hibridomas

Pueden prepararse anticuerpos monoclonales contra epítomos de cualquiera de los péptidos identificados y aislados como se describió, a partir de hibridomas murinos según el método clásico de Kohler y Milstein, Nature, 256: 495, 1975, o métodos derivados de éste. Brevemente, un ratón se inocula repetidas veces con unos pocos microgramos de la proteína seleccionada, o péptidos derivados de ésta, a lo largo de un periodo de varias semanas. Después se sacrifica el ratón, y se aíslan las células productoras de anticuerpos del bazo. Las células del bazo se fusionan mediante polietilenglicol con células de mieloma de ratón, y el exceso de células no fusionadas se destruye mediante el crecimiento del sistema en un medio selectivo que comprende aminopterina (medio HAT). Las células que se habían fusionado con éxito se diluyen, y se colocan partes alícuotas de la dilución en pocillos de una placa de microtitulación en los que continúa el crecimiento del cultivo. Los clones productores de anticuerpos se identifican mediante detección del anticuerpo en el fluido sobrenadante de los pocillos mediante procedimientos de inmunoanálisis, tales como ELISA, como describió originariamente Engvall, Meth. Enzymol. 70: 419, 1980, y métodos derivados de éste. Los clones positivos seleccionados se pueden expandir y su producto de anticuerpo monoclonal se recolecta para su uso. Los procedimientos detallados para la producción de anticuerpos monoclonales son descritos por Davis et al., en Basic Methods in Molecular Biology, Elsevier, Nueva York, sección 21-2.

2. Producción de Anticuerpos Policlonales Mediante Inmunización

Puede prepararse antisuero policlonal que contiene anticuerpos contra epítomos heterogéneos de una única proteína, inmunizando animales adecuados con la proteína expresada, o péptidos derivados de ésta, que pueden estar sin modificar o modificados para potenciar la inmunogenicidad. La producción eficaz de anticuerpos policlonales se ve afectada por muchos factores relacionados con el antígeno y la especie anfitriona. Por ejemplo, las moléculas pequeñas tienden a ser menos inmunogénicas que otras, y pueden requerir el uso de vehículos y adyuvantes. Además, la respuesta de los animales anfitriones varía dependiendo del sitio de inoculación y la dosis, produciendo una dosis inadecuada o excesiva de antígeno un antisuero con menor valoración. Las dosis pequeñas (nivel de ng) de antígeno administradas en múltiples sitios intradérmicos parecen ser las más fiables. Un protocolo de inmunización eficaz para conejos puede encontrarse en Vaitukaitis et al., J. Clin. Endocrinol. Metab., 33: 988-991 (1971).

Pueden administrarse inyecciones de refuerzo a intervalos regulares, y el antisuero se recolecta cuando la valoración de anticuerpos, determinada de forma semicuantitativa, por ejemplo, mediante inmunodifusión doble en agar contra concentraciones conocidas del antígeno, empieza a caer. Véase, por ejemplo, Ouchterlony et al., capítulo 19, en: Handbook of Experimental Immunology, D. Wier (ed.), Blackwell (1973). La concentración meseta del anticuerpo normalmente está en el intervalo de 0,1 a 0,2 mg/ml de suero (aproximadamente 12 μM). Se determina la afinidad de los antisueros por el antígeno preparando curvas de unión competitiva, como describen, por ejemplo, Fisher, D., capítulo 42, en: Manual of Clinical Immunology, 2ª ed. (Rose y Friedman, eds.), Amer. Soc. For Microbiol., Washington D.C. (1980).

Las preparaciones de anticuerpos preparadas según cualquiera de los protocolos son útiles en inmunoanálisis cuantitativos que determinan las concentraciones de sustancias que portan antígenos en muestras biológicas; también se emplean de modo semicuantitativo o cualitativo para identificar la presencia de antígenos en una muestra biológica. Los anticuerpos también pueden utilizarse en composiciones terapéuticas para eliminar células que expresan la proteína, o para reducir los niveles de la proteína en el cuerpo.

V. Uso de ESTs 5' o Secuencias Obtenibles a Partir de Éstas o Porciones de Éstas como Reactivos

Las ESTs 5' de la presente invención (o ADNcs o ADNs genómicos obtenibles a partir de éstas) pueden utilizarse como reactivos en procedimientos de aislamiento, análisis de diagnóstico y procedimientos forenses. Por ejemplo, las secuencias de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) pueden marcarse de forma detectable y utilizarse como sondas para aislar otras secuencias capaces de hibridarse con ellas. Además, las secuencias de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) pueden utilizarse para diseñar cebadores de PCR para utilizar en procedimientos de aislamiento, diagnóstico o forenses.

1. Uso de ESTs 5' o Secuencias Obtenibles a Partir de Éstas o Porciones de Éstas en Procedimientos de Aislamiento, Diagnóstico y Forenses

EJEMPLO 44

Preparación de Cebadores de PCR y Amplificación del ADN

Las secuencias de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) pueden utilizarse para preparar cebadores de PCR para una diversidad de aplicaciones, incluyendo procedimientos de aislamiento para clonar ácidos nucleicos capaces de hibridarse con dichas secuencias, técnicas de diagnóstico y técnicas forenses. Los cebadores de PCR tienen una longitud de al menos 10 bases, y preferiblemente al menos 12, 15 ó 17 bases. Más preferiblemente, los cebadores de PCR tienen una longitud de al menos 20-30 bases. En algunas realizaciones, los cebadores de PCR pueden tener una longitud de más de 30 bases. Se prefiere que las parejas de cebadores tengan aproximadamente la misma proporción G/C, de forma que las temperaturas de fusión sean aproximadamente las mismas. Una diversidad de procedimientos de PCR son familiares para los expertos en la técnica. Para un informe acerca de la tecnología PCR, véase Molecular Cloning to Genetic Engineering, White Ed., en Methods in Molecular Biology, 67: Humana Press, Totowa, 1997. En cada uno de estos procedimientos de PCR, se añaden cebadores de PCR a cada lado de las secuencias de ácidos nucleicos que se van a amplificar, a una muestra de ácidos nucleicos preparada de forma adecuada, junto con dNTP y una polimerasa termoestable, tal como polimerasa Taq, polimerasa Pfu o polimerasa Vent. El ácido nucleico en la muestra se desnaturaliza y los cebadores de PCR se hibridan específicamente con las secuencias de ácidos nucleicos complementarias en la muestra. Los cebadores hibridados se extienden. Después se inicia otro ciclo de desnaturalización, hibridación y extensión. Los ciclos se repiten múltiples veces para producir un fragmento amplificado que contiene la secuencia de ácido nucleico entre los sitios de los cebadores.

EJEMPLO 45

Uso de ESTs 5' Como Sondas

Las sondas derivadas de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas), incluyendo los ADNcs completos o las secuencias genómicas, pueden marcarse con marcadores detectables familiares para los expertos en la técnica, incluyendo radioisótopos y marcadores no radiactivos, para proporcionar una sonda detectable. La sonda detectable puede ser monocatenaria o bicatenaria, y puede fabricarse utilizando mecanismos conocidos en la técnica, incluyendo la transcripción *in vitro*, traducción de mellas o 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, puede desnaturalizarse antes de ponerlo en contacto con la sonda. En algunas aplicaciones, la muestra de ácido nucleico puede inmovilizarse sobre una superficie, tal como una membrana de nitrocelulosa o nailon. La muestra de ácido nucleico puede comprender ácidos nucleicos obtenidos a partir de una diversidad de fuentes, incluyendo ADN genómico, bancos de ADNc, ARN o muestras de tejidos.

Los procedimientos utilizados para detectar la presencia de ácidos nucleicos capaces de hibridarse con la sonda detectable incluyen técnicas bien conocidas, tales como análisis de transferencia Southern, análisis de transferencia Northern, análisis de transferencia puntual, hibridación de colonias, e hibridación en placas. En algunas aplicaciones, el ácido nucleico capaz de hibridarse con la sonda marcada puede clonarse en vectores, tales como vectores de expresión, vectores de secuenciación, o vectores de transcripción *in vitro* para facilitar la caracterización y expresión de los ácidos nucleicos que se están hibridando en la muestra. Por ejemplo, estas técnicas pueden utilizarse para aislar y clonar secuencias en un banco genómico o banco de ADNc que son capaces de hibridarse con la sonda detectable como se describe en el Ejemplo 30 anterior.

Los cebadores de PCR fabricados como se describe en el ejemplo 44 anterior puede utilizarse en análisis forenses, tal como las técnicas de huella de ADN descritas en los ejemplos 46-50, a continuación. Estos análisis pueden utilizar cebadores o sondas detectables basadas en las secuencias de las ESTs 5' o de los ADNcs o ADNs genómicos aislados utilizando las EST 5'.

EJEMPLO 46

Correspondencia Forense mediante Secuenciación de ADN

En un método ejemplar, se aíslan muestras de ADN a partir de especímenes forenses, por ejemplo, de pelo, semen, sangre o células de la piel mediante métodos convencionales. A continuación se utiliza un panel de cebadores de PCR basado en una serie de las ESTs 5' del Ejemplo 25, o ADNcs o ADNs genómicos aislados a partir de éstos como se describió anteriormente, según el Ejemplo 44 para amplificar ADN de aproximadamente 100-200 bases de longitud a partir del espécimen forense. Se obtienen las secuencias correspondientes a partir de un sujeto de ensayo. Cada una de estas

- identificaciones de ADN se secuencian después utilizando técnicas convencionales, y una simple comparación en la base de datos determina las diferencias, si hay, entre las secuencias del sujeto y las de la muestra. Las diferencias estadísticamente significativas entre las secuencias de ADN del sospechoso y las de la muestra demuestran, de forma concluyente, la carencia de equivalencia. Esta carencia de equivalencia puede demostrarse, por ejemplo, con sólo una secuencia. La equivalencia, por otra parte, debe demostrarse con una gran cantidad de secuencias, correspondiéndose todas. Preferiblemente, se utiliza un mínimo de 50 secuencias estadísticamente idénticas de 100 bases de longitud para demostrar la equivalencia entre el sospechoso y la muestra.

EJEMPLO 47

10 Identificación Positiva mediante Secuenciación de ADN

- La tecnología esbozada en el ejemplo previo también puede utilizarse a mayor escala para proporcionar una identificación exclusiva de huella de cualquier individuo. En esta técnica, se preparan cebadores a partir de un gran número de secuencias de las ESTs 5' del Ejemplo 25, o secuencias de ADNcs o ADNs genómico que pueden obtenerse de éstos. Preferiblemente, se utilizan de 20 a 50 cebadores diferentes. Estos cebadores se utilizan para obtener un número correspondiente de segmentos de ADN generados por PCR a partir del individuo en cuestión según el Ejemplo 44. Cada uno de estos segmentos de ADN se secuencian, utilizando los métodos indicados en el Ejemplo 46. La base de datos de las secuencias generadas mediante este procedimiento identifica, de forma exclusiva, el individuo del cual se obtuvieron las secuencias. Se puede emplear a continuación el mismo panel de cebadores en cualquier momento posterior para correlacionar absolutamente tejidos u otros especímenes biológicos con ese individuo.

EJEMPLO 48

Identificación Forense con Análisis de Transferencia Southern

- Se repite el procedimiento del Ejemplo 47 para obtener un panel de al menos 10 secuencias amplificadas a partir de un individuo y un espécimen. Preferiblemente, el panel contiene al menos 50 secuencias amplificadas. Más preferiblemente, el panel contiene 100 secuencias amplificadas. En algunas realizaciones, el panel contiene 200 secuencias amplificadas. Este ADN generado por PCR se digiere a continuación con una o con una combinación de, preferiblemente, cuatro enzimas de restricción específicas de base. Estas enzimas están disponibles en el mercado y son conocidas por los expertos en la técnica. Después de la digestión, los fragmentos génicos resultantes se separan por tamaño en múltiples pocillos por duplicado en un gel de agarosa, y se trasladan a nitrocelulosa utilizando mecanismos de análisis de transferencia Southern muy conocidas por los expertos en la técnica. Para un informe acerca del análisis de la transferencia Southern, véase Davis et al. (Basic Methods in Molecular Biology, 1986, Elsevier Press, págs. 62-65).
- Un panel de sondas basadas en las secuencias de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas), o fragmentos de éstas de al menos 10 bases, se marcan de modo radiactivo o colorimétrico utilizando métodos conocidos en la técnica, tales como traducción de mellas o marcaje en el extremo, y se hibridan con la transferencia Southern utilizando mecanismos conocidos en la técnica (Davis et al., supra). Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas). Más preferiblemente, la sonda comprende al menos 20-30 nucleótidos consecutivos de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas). En algunas realizaciones, la sonda comprende más de 30 nucleótidos de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas).
- Preferiblemente, se utilizan al menos 5 a 10 de estas sondas marcadas, y más preferiblemente se utilizan al menos aproximadamente 20 ó 30 para proporcionar un patrón exclusivo. Las bandas resultantes que aparecen tras la hibridación de una muestra grande de EST 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) serán un identificador exclusivo. Puesto que la escisión con enzimas de restricción será diferente en cada individuo, el patrón de bandas del análisis de la transferencia Southern también será exclusivo. El aumento del número de sondas de EST 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) proporciona un nivel estadísticamente superior de confianza en la identificación, puesto que existirá un número mayor de conjuntos de bandas utilizado para la identificación.

EJEMPLO 49

55 Procedimiento de Identificación mediante Transferencia Puntual

Otra técnica para identificar individuos que utiliza las secuencias de las ESTs 5' descritas en la presente memoria utiliza una técnica de hibridación de transferencia puntual.

- Se aísla ADN genómico a partir de núcleos del sujeto que se va a identificar. Se sintetizan sondas de oligonucleótidos con una longitud de aproximadamente 30 pb, que corresponden al menos a 10, preferiblemente a 50 secuencias de las ESTs 5', o los ADNcs o ADNs genómicos obtenibles a partir de éstas. Las sondas se utilizan para hibridarse con el ADN genómico mediante condiciones conocidas en la técnica. Los oligonucleótidos se marcan en el extremo con P^{32} utilizando polinucleótido quinasa (Pharmacia). Las Transferencias Puntuales se crean rociando el ADN genómico sobre nitrocelulosa o similar utilizando un colector de transferencia puntual al vacío (BioRad, Richmond, California). El filtro de nitrocelulosa que contiene las secuencias genómicas se cuece o se une mediante UV al filtro, se prehibrida y se hibrida con sonda marcada utilizando mecanismos conocidos en la técnica (Davis et al., supra). Los fragmentos de ADN marcados con P^{32} se hibridan secuencialmente con condiciones sucesivamente rigurosas para detectar las diferencias mínimas entre la secuencia de 30 pb y el ADN. El cloruro de tetrametilamonio es útil para identificar clones que contienen pocos apareamientos erróneos de nucleótidos (Wood et al., Proc. Natl. Acad. Sci. USA, 82(6): 1585-1588, 1985). Un patrón exclusivo de puntos distingue un individuo de otro individuo.
- Las secuencias de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas), o los oligonucleótidos que contienen al menos 10 bases consecutivas de estas secuencias, pueden utilizarse como sondas en la siguiente técnica alternativa de huella. Preferiblemente, la sonda comprende al menos 12, 15 ó 17 nucleótidos consecutivos de las secuencias de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas). Más preferiblemente, la sonda comprende al menos 20-30 nucleótidos consecutivos de las secuencias de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas). En algunas realizaciones, la sonda comprende más de 30 nucleótidos de las secuencias EST 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas).
- Preferiblemente, se utiliza una pluralidad de sondas que tienen secuencias de genes diferentes en la técnica alternativa de la huella. El Ejemplo 50 a continuación proporciona un procedimiento alternativo de huella representativo, en el que las sondas se derivan de las ESTs 5'.

EJEMPLO 50

Técnica Alternativa de Identificación de la "huella"

- Se preparan oligonucleótidos de 20 unidades a partir de un gran número, por ejemplo, 50, 100 ó 200 de EST 5' utilizando servicios de oligonucleótidos disponibles en el mercado, tales como Genset, París, Francia. Las muestras celulares de los sujetos de ensayo se procesan para obtener ADN utilizando mecanismos muy conocidos por los expertos en la técnica. 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 una electroforesis. El procedimiento, como se conoce en la técnica, puede modificarse para acomodarse a una electroforesis en poliacrilamida, sin embargo, en este ejemplo, las muestras que contienen 5 µg de ADN se cargan en pocillos y se separan en geles de agarosa al 0,8%. Los geles se trasladan a nitrocelulosa utilizando técnicas de análisis de transferencia Southern convencionales.

- Se reúnen 10 ng de cada uno de los oligonucleótidos y se marcan en el extremo con P^{32} . La nitrocelulosa se prehibrida con disolución de bloqueo 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 hibridación resultante será exclusivo para cada individuo.

También se contempla en este Ejemplo que el número de secuencias de sondas utilizado puede variarse para mayor precisión o claridad.

- Las proteínas codificadas por los ADNcs extendidos también pueden utilizarse para generar anticuerpos como se explica en los Ejemplos 30 y 43, para identificar el tipo de tejido o especie celular a partir del cual se deriva una muestra, como se describe en el Ejemplo 51.

EJEMPLO 51

Identificación de los Tipos de Tejidos o Especies Celulares mediante Anticuerpos Específicos de Tejido Marcados

- La identificación de tejidos específicos se logra mediante la visualización de antígenos específicos de tejidos mediante preparaciones de anticuerpos según los Ejemplos 30 y 43, que se conjugan, directa o indirectamente, con un marcador detectable. Las especies de anticuerpos marcados seleccionados se unen a su pareja de unión de antígeno específica en secciones de tejidos, suspensiones celulares, o en extractos de proteínas solubles procedentes de una muestra de tejido para proporcionar un patrón para la interpretación cualitativa o semicualitativa.

Los antisueros para estos procedimientos deben tener una potencia mayor que la de la preparación nativa y, por esta razón, los anticuerpos se concentran hasta un nivel de mg/ml mediante aislamiento

- de la fracción de gammaglobulina, por ejemplo, mediante cromatografía de intercambio iónico, o mediante fraccionamiento en sulfato de amonio. Además, para proporcionar los antisueros más específicos, deben retirarse los anticuerpos no deseados, por ejemplo contra proteínas comunes, de la fracción de gamma-globulina, por ejemplo mediante inmunoabsorbentes insolubles, antes de que los anticuerpos se marquen con el marcador. Los antisueros monoclonales o heterólogos son adecuados para cualquier procedimiento.

A. Técnicas inmunohistoquímicas

- Anticuerpos purificados de alta valoración, preparados como se describió anteriormente, se conjugan con un marcador detectable, como describe, por ejemplo, Fudenberg, capítulo 26, en: Basic and Clinical Immunology, 3ª ed., Lange, Los Altos, California, 1980, o Rose, et al., capítulo 12, en: Methods in Immunodiagnosis, 2ª ed., John Wiley and Sons, Nueva York (1980).

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

- Las preparaciones para realizar estos procedimientos pueden comprender anticuerpos monoclonales o policlonales contra una única proteína o péptido identificado como específico de un tipo de tejido, por ejemplo, tejido cerebral, o pueden emplearse preparaciones de anticuerpos contra varios antígenos específicos de tejidos antigénicamente diferenciados, en paneles, independientemente o en mezclas, según se requiera.

- Las secciones de tejidos y suspensiones celulares se preparan para el examen inmunohistoquímico según técnicas histológicas habituales. Se montan múltiples secciones de criostato (aproximadamente 4 μ m, sin fijar) del tejido desconocido y el control conocido, y cada portaobjetos se cubre con diferentes diluciones de la preparación de anticuerpos. Secciones de tejidos conocidos y desconocidos también deben tratarse con preparaciones para proporcionar un control positivo, un control negativo, por ejemplo, sueros preinmunes, y un control para la tinción no específica, por ejemplo, también.

Las secciones tratadas se incuban en una cámara húmeda durante 30 min a temperatura ambiente, se enjuagan, después se lavan en tampón durante 30-45 min. El exceso de fluido se elimina con papel secante, y el marcador se revela.

- Si el anticuerpo específico de tejido no se marcó en la primera incubación, puede marcarse en este momento en una segunda reacción de anticuerpo-anticuerpo, por ejemplo, añadiendo anticuerpo conjugado con fluoresceína o enzima contra la clase de inmunoglobulinas de la especie productora de antisuero, por ejemplo, anticuerpo marcado con fluoresceína contra IgG de ratón. Estos sueros marcados están disponibles en el mercado.

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

B. Identificación de proteínas solubles específicas de tejido

- La visualización de proteínas específicas de tejido y la identificación de tejidos desconocidos a partir de este procedimiento se realiza utilizando reactivos de anticuerpos marcados y una estrategia de detección como la descrita para la inmunohistoquímica; sin embargo, la muestra se prepara según una técnica electroforética para distribuir las proteínas extraídas a partir del tejido en una matriz ordenada basada en el peso molecular para la detección.

- Una muestra de tejido se homogeneiza utilizando un aparato Virtis; las suspensiones celulares se rompen mediante homogeneización Dounce o lisis osmótica, utilizando detergentes en cada caso según se requiera para romper las membranas celulares, como es la práctica en la técnica. Los componentes celulares insolubles, tales como núcleos, microsomas y fragmentos de membranas se retiran mediante ultracentrifugación, y la fracción que contiene las proteínas solubles se concentra, si es necesario, y se reserva para su análisis.

- Una muestra de la disolución de proteínas solubles se resuelve en especies proteicas individuales mediante electroforesis en SDS-poliacrilamida convencional, como describen, por ejemplo, Davis et al., sección 19-2, en: Basic Methods in Molecular Biology, Leder ed., Elsevier, Nueva York, 1986, utilizando una variedad de cantidades de poliacrilamida en un conjunto de geles para resolver el

intervalo completo de peso molecular de las proteínas que se van a detectar en la muestra. Se ensaya un marcador de tamaño en paralelo para estimar los pesos moleculares de las proteínas constituyentes. Un tamaño de muestra para el análisis es un volumen conveniente de 5 a 55 µl, y contiene de aproximadamente 1 a 100 µg de proteína. Se traslada una parte alícuota de cada una de las proteínas resueltas mediante transferencia a un papel de filtro de nitrocelulosa, un proceso que mantiene el patrón de resolución. Se preparan múltiples copias. El procedimiento, conocido como análisis de transferencia Western, está bien descrito por Davis, L. et al., supra, sección 19-3. Un conjunto de transferencias de nitrocelulosa se tiñe con colorante de azul de Coomassie para visualizar el conjunto completo de proteínas para la comparación con las proteínas unidas a anticuerpos. El resto de los filtros de nitrocelulosa se incuban a continuación con una disolución de uno o más antisueros específicos contra proteínas específicas de tejido preparados como se describe en los ejemplos 30 y 43. En este procedimiento, al igual que en el procedimiento A anterior, se ensayan las muestras positivas y negativas y los controles de reactivo apropiados.

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

La visualización de la unión de antígenos específicos de tejido a niveles superiores a los observados en tejidos control, a uno o más anticuerpos específicos de tejido, preparados a partir de secuencias de genes identificadas a partir de secuencias de ADNc extendidas, puede identificar tejidos de origen desconocido, por ejemplo, muestras forenses, o tejido tumoral diferenciado que se ha metastatizado hacia otros sitios corporales. Además de sus aplicaciones forenses e identificativas, las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) pueden cartografiarse en sus localizaciones cromosómicas. El Ejemplo 52 a continuación describe el cartografiado de híbridos por radiación (RH) de regiones cromosómicas humanas utilizando EST 5'. El Ejemplo 53 a continuación describe un procedimiento representativo para cartografiar una EST 5' en su localización en un cromosoma humano. El Ejemplo 54 a continuación describe el cartografiado de EST 5' en cromosomas en metafase mediante hibridación in situ de fluorescencia (FISH). Los expertos en la técnica apreciarán que el método de los Ejemplos 52-54 también puede utilizarse para cartografiar ADNc o ADN genómicos obtenibles a partir de las EST 5' en sus localizaciones cromosómicas.

2. Uso de ESTs 5' o Secuencias Obtenibles a partir de Éstas o Porciones de Éstas en el Cartografiado Cromosómico

EJEMPLO 52

Cartografiado de Híbridos por Radiación de ESTs 5' en el Genoma Humano

Un cartografiado de híbridos con radiación (RH) es un enfoque genético de células somáticas que puede utilizarse para el cartografiado de alta resolución del genoma humano. En este enfoque, las líneas celulares que contienen uno o más cromosomas humanos se irradian letalmente, rompiendo cada cromosoma en fragmentos cuyo tamaño depende de la dosis de radiación. Estos fragmentos se rescatan mediante fusión con células de roedor cultivadas, produciendo subclones que contienen porciones diferentes del genoma humano. Esta técnica la describen Benham et al., *Genomics*, 4: 509-517, 1989; y Cox et al., *Science*, 250: 245-250, 1990. La naturaleza aleatoria e independiente de los subclones permite 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 de cartografiado para ordenar las ESTs 5'. En este enfoque, la frecuencia de ruptura entre los marcadores se utiliza para medir la distancia, permitiendo la construcción de mapas de resolución finos, como se ha hecho utilizando ESTs convencionales (Schuler et al., *Science*, 274: 540-546, 1996).

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

EJEMPLO 53

Cartografiado de las ESTs 5' en Cromosomas Humanos Utilizando Técnicas de PCR

- Las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) pueden asignarse a cromosomas humanos utilizando metodologías basadas en PCR. En estos enfoques, se diseñan parejas de cebadores de oligonucleótidos a partir de las ESTs 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) para minimizar la probabilidad de amplificar a través de un intrón. Preferiblemente, los cebadores de oligonucleótidos tienen una longitud de 18-23 pb y se diseñan para la amplificación por PCR. La creación de cebadores de PCR a partir de secuencias conocidas es muy conocida por los expertos en la técnica. Para un informe acerca de la tecnología de PCR, véase Erlich en PCR Technology, Principles and Applications for DNA Amplification, Freeman and Co., Nueva York, 1992.
- Los cebadores se utilizan en reacciones en cadena de polimerasa (PCR) para amplificar moldes a partir del ADN genómico humano completo. Las condiciones de PCR son como sigue: se usan 60 ng de ADN genómico como molde para PCR con 80 ng de cada cebador de oligonucleótido, 0,6 unidades de polimerasa Taq, y 1 μ Cu de una desoxicitidina trifosfato marcada con P^{32} . La PCR se realiza en un termociclador de microplacas (Techne) con las siguientes condiciones: 30 ciclos de 94°C, 1,4 min; 55°C, 2 min; y 72°C, 2 min; con una extensión final a 72°C durante 10 min. Los productos amplificados se analizan en un gel de secuenciación de poli(acrilamida) al 6% y se visualizan mediante autorradiografía. Si la longitud del producto PCR resultante es idéntica a la distancia entre los extremos de las secuencias de los cebadores en el ADNc extendido a partir del cual derivan los cebadores, la reacción de PCR se repite a continuación con moldes de ADN procedentes de dos paneles de híbridos de células somáticas de humano-roedor, ADN BIOS PCRable (BIOS Corporation) y panel Núm. 1 de cartografiado de híbridos de células somáticas de humano-roedor NIGMS (NIGMS, Camden, NJ).
- Se utiliza PCR para seleccionar una serie de líneas celulares de híbridos de células somáticas que contienen conjuntos definidos de cromosomas humanos para detectar la presencia de una EST 5' concreta (o ADNc o ADN genómico obtenible a partir de ésta). El ADN se aísla a partir de los híbridos somáticos y se utiliza como molde de partida para reacciones de PCR utilizando las parejas de cebadores de la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta). Sólo aquellos híbridos de células somáticas con cromosomas que contienen el gen humano correspondiente a la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) producirá un fragmento amplificado. La EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) se asigna a un cromosoma mediante análisis del patrón de segregación de los productos de PCR a partir de los moldes de ADN de híbridos somáticos. El único cromosoma humano presente en todos los híbridos celulares que produce un fragmento amplificado es el cromosoma que contiene esa EST 5' (o ADNc o ADN genómico obtenible a partir de ésta). Para un informe acerca de las técnicas y análisis de los resultados de experimentos de cartografiado de genes de células somáticas, véase Ledbetter et al., Genomics, 6: 475-481, 1990.

EJEMPLO 54

Cartografiado de EST 5' Extendidas en Cromosomas Utilizando Hibridación In Situ de Fluorescencia

- La hibridación in situ de fluorescencia permite cartografiar la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) en una localización particular sobre un cromosoma concreto. Los cromosomas que se utilizan para las técnicas de hibridación in situ de fluorescencia pueden obtenerse a partir de una diversidad de fuentes, incluyendo cultivos celulares, tejidos o sangre completa.

- En una realización preferida, la localización cromosómica de una EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) se obtiene mediante FISH, como describen en Cherif et al. (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 de hombres sanos se cultivan durante 72 h en medio RPMI-1640. Para la sincronización se añade metotrexato (10 μ M) durante 17 h, seguido de la adición de 5-bromodesoxiuridina (5-BrdU, 0,1 mM) durante 6 h. Se añade colcemida (1 μ g/ml) durante los últimos 15 min antes de recolectar las células. Las células se recolectan, se lavan en RPMI, se incuban con una disolución hipotónica de KCl (75 mM) a 37°C durante 15 min, y se fijan en tres cambios de metanol:ácido acético (3:1). La suspensión celular se rocía sobre un portaobjetos de vidrio y se seca al aire. La EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) se marca con biotina-16 dUTP mediante traducción de mellas según las instrucciones del fabricante (Bethesda Research Laboratories, Bethesda, MD), se purifica utilizando una columna Sephadex G-50 (Pharmacia, Upsala, Suecia) y se precipita. Justo antes de la hibridación, el sedimento de ADN se disuelve en tampón de hibridación (formamida al 50%, 2 x SSC, sulfato de dextrano al 10%, 1 mg/ml de ADN de esperma de salmón sonificado, pH 7), y la sonda se desnaturaliza a 70°C durante 5-10 min.

- Los portaobjetos mantenidos a -20°C se tratan durante 1 h a 37°C con ARNasa A (100 μ g/ml), se enjuagan tres veces en 2 x SSC, y se deshidratan en una serie de etanol. Las preparaciones de cromosomas se desnaturalizan en formamida al 70%, 2 x SSC durante 2 min a 70°C, después se deshidratan a 4°C. Los portaobjetos se tratan con proteinasa K (10 μ g/100 ml en Tris-HCl 20 mM,

CaCl₂ 2 mM) a 37°C durante 8 min y se deshidratan. La mezcla de hibridación que contiene la sonda se coloca sobre el portaobjetos, se cubre con un cubreobjetos, se sella con un cemento de goma y se incuba durante la noche en una cámara húmeda a 37°C. Después de los lavados de hibridación y poshibridación, la sonda biotinilada se detecta mediante avidina-FITC y se amplifica con más capas de antiavidina de cabra biotinilada y avidina-FITC. Para la localización cromosómica, se obtienen bandas R fluorescentes como se ha descrito previamente (Cherif et al., supra). Los portaobjetos se observan bajo un microscopio de fluorescencia LEICA (DMRXA). Los cromosomas se contratiñen con yoduro de propidio y la señal de fluorescencia de la sonda aparece como dos manchas de color amarillo-verde simétricas en ambas cromátidas del cromosoma de banda R fluorescente (rojo). Por tanto, una EST 5' concreta (o ADNc o ADN genómico obtenible a partir de ésta) puede localizarse en una banda R citogenética concreta sobre un cromosoma dado. Cuando la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) se ha asignado a un cromosoma concreto utilizando las técnicas descritas en los Ejemplos 52-54 anteriores, se puede utilizar para construir un mapa de los cromosomas de alta resolución sobre los cuales se localiza, o para identificar los cromosomas en una muestra.

EJEMPLO 55

Uso de EST 5' para Construir o Expandir Mapas Cromosómicos

El cartografiado de cromosomas implica asignar una secuencia exclusiva concreta a un cromosoma particular, como se describió anteriormente. Cuando la secuencia exclusiva se ha cartografiado en un cromosoma dado, se ordena con relación a otras secuencias exclusivas localizadas en el mismo cromosoma. Un enfoque para el cartografiado de cromosomas utiliza una serie de cromosomas artificiales de levaduras (YAC) que portan varios miles de insertos largos derivados de los cromosomas del organismo a partir del cual se obtienen los ADNcs extendidos (o ADNs genómicos obtenibles a partir de éstos). Este enfoque lo describen Nagaraja et al., *Genome Research*, 7: 210-222, 1997. Brevemente, en este enfoque cada cromosoma se rompe en trozos solapantes que se insertan en el vector YAC. Los insertos de YAC se seleccionan utilizando PCR u otros métodos para determinar si incluyen la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) cuya posición se va a determinar. Cuando se encuentra un inserto que incluye la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta), el inserto puede analizarse mediante PCR u otros métodos para determinar si el inserto contiene también otras secuencias que se sabe que están sobre el cromosoma o en la región de la cual deriva la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta). Este proceso puede repetirse para cada inserto en el banco de YAC para determinar la localización de cada uno de los ADNcs extendidos (o ADNs genómicos obtenibles a partir de éstos) en relación con cada uno y con otros marcadores cromosómicos conocidos. De esta manera, puede obtenerse un mapa de alta resolución de la distribución de numerosos marcadores exclusivos a lo largo de cada uno de los cromosomas del organismo. Como se describe en el Ejemplo 56, a continuación, los ADNcs extendidos (o ADNs genómicos obtenibles a partir de éstos) también pueden utilizarse para identificar genes asociados con un fenotipo particular, tal como una enfermedad hereditaria o la respuesta a un fármaco.

3. Uso de las ESTs 5' o Secuencias Obtenidas a partir de Éstas o Fragmentos de Éstas en la Identificación de Genes

EJEMPLO 56

Identificación de Genes Asociados con Enfermedades Hereditarias o Respuesta a Fármacos

Este Ejemplo ilustra un enfoque útil para la asociación de ESTs 5' (o ADNc o ADN genómico obtenible a partir de éstas) con características fenotípicas particulares. En este Ejemplo, una EST 5' particular (o ADNc o ADN genómico obtenible a partir de ésta) se usa como una sonda de ensayo para asociar esa EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) con una característica fenotípica particular.

Las ESTs 5' (o ADNc o ADN genómico obtenible a partir de éstas) se cartografían en busca de una localización particular sobre un cromosoma humano utilizando mecanismos tales como los descritos en los Ejemplos 52 y 53, u otros mecanismos conocidos en la técnica. Una búsqueda de la Herencia Mendeliana en el Ser Humano (McKusick en *Mendelian Inheritance in Man* (disponible en línea a través de la Johns Hopkins University Welch Medical Library)) revela que la región del cromosoma humano que contiene la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) es una región muy rica en genes que contiene varios genes conocidos y varias enfermedades o fenotipos para los cuales no se han identificado los genes. El gen correspondiente a esta EST 5' (o ADNc o ADN genómico obtenible a partir de ésta), por tanto, se transforma en un candidato inmediato para cada una de estas enfermedades genéticas.

Se aíslan y expanden en cultivo células procedentes de pacientes con estas enfermedades o fenotipos. Se utilizan cebadores de PCR procedentes de la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta) para seleccionar ADN genómico, ARNm o ADNc obtenido de los pacientes. La EST 5'

- (o ADNc o ADN genómico obtenible a partir de ésta) que no se amplifican en los pacientes pueden asociarse de forma positiva con una enfermedad particular mediante posteriores análisis. Como alternativa, el análisis de PCR puede producir fragmentos de longitudes diferentes cuando las muestras se derivan de un individuo que tiene el fenotipo asociado con la enfermedad, que cuando las muestras se derivan de un individuo sano, indicando que el gen que contiene la EST 5' puede ser responsable de la enfermedad genética.

VI. Uso de EST 5' (o ADNc o ADN Genómico Obtenible a partir de Ésta) para Construir Vectores

- Las presentes ESTs 5' (o ADNc o ADN genómico obtenible a partir de éstas) también pueden utilizarse para construir vectores de secreción capaces de dirigir la secreción de las proteínas codificadas por los genes en su interior. Estos vectores de secreción pueden facilitar la purificación o enriquecimiento de las proteínas codificadas por genes insertados en ellos mediante la reducción del número de proteínas de fondo a partir de las cuales la proteína deseada debe purificarse o enriquecerse. En el Ejemplo 57, a continuación, se describen ejemplos de vectores de secreción.

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 de genes en la célula, tejido u organismo anfitrión de interés. Estos promotores incluyen el promotor del virus del sarcoma de Rous, el promotor de SV40, el promotor del citomegalovirus humano, y otros promotores familiares para los expertos en la técnica.

- Una secuencia señal de una EST 5' (o ADNcs o ADNs genómicos obtenibles a partir de ésta) se conecta operablemente con el promotor, de forma que el ARNm transcrito a partir del promotor dirige la traducción del péptido señal. La célula, tejido u organismo anfitrión puede ser cualquier célula, tejido u organismo que reconoce el péptido señal codificado por la secuencia señal en la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta). Los anfitriones adecuados incluyen células, tejidos u organismos de mamíferos, células, tejidos u organismos aviares, células, tejidos u organismos de insectos, o levaduras.

- 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 inserto dentro de marco con la secuencia señal, de forma que una proteína de fusión en la que el péptido señal está fusionado 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 integrarse en el cromosoma del anfitrión para ser mantenido, de manera estable, como un replicón extracromosómico en el anfitrión, para ser un cromosoma artificial, o para estar presente de forma transitoria en el anfitrión. Muchos esqueletos de ácidos nucleicos adecuados para su uso como vectores de secreción son conocidos por los expertos en la técnica, incluyendo vectores retrovíricos, vectores de SV40, vectores del virus del papiloma bovino, plásmidos de integración en levaduras, plásmidos episómicos de levaduras, cromosomas artificiales de levaduras, cromosomas artificiales humanos, vectores de elemento P, vectores de baculovirus, o plásmidos bacterianos capaces de ser introducidos de forma transitoria en el anfitrión.

El vector de secreción también puede contener una señal de poliA, de forma que la señal de poliA se localiza aguas abajo del gen insertado en el vector de secreción.

- Después de que el gen que codifica la proteína de la cual se desea la secreción se inserta en el vector de secreción, el vector de secreción se introduce en la célula, tejido u organismo anfitrión utilizando precipitación con fosfato de calcio, DEAE-dextrano, electroporación, transfección mediada por liposomas, partículas víricas o como ADN desnudo. La proteína codificada por el gen insertado se purifica o enriquece después a partir del sobrenadante utilizando técnicas convencionales, tales como precipitación con sulfato de amonio, inmunoprecipitación, inmunocromatografía, cromatografía de exclusión molecular, cromatografía de intercambio iónico, y HPLC. Como alternativa, la proteína secretada puede estar en un estado suficientemente enriquecido o puro en el sobrenadante o medio de crecimiento del anfitrión para permitir que pueda utilizarse para su fin previsto sin más enriquecimiento.

- Las secuencias señal también pueden insertarse en vectores diseñados para terapia génica. En estos vectores, la secuencia señal está operablemente conectada a un promotor, de forma que el ARNm transcrito a partir del promotor codifica el péptido señal. Un sitio de clonación se localiza aguas abajo de la secuencia señal, de forma que un gen que codifica una proteína cuya secreción se desea puede insertarse con facilidad en el vector y fusionarse con la secuencia señal. El vector se introduce en una

célula anfitriona apropiada. La proteína expresada a partir del promotor se secreta extracelularmente, produciendo, con ello, un efecto terapéutico.

- 5 Las ESTs 5' también pueden utilizarse para clonar secuencias localizadas aguas arriba de las ESTs 5' que son capaces de regular la expresión génica, incluyendo secuencias de promotores, secuencias de potenciadores y otras secuencias aguas arriba que influyen en los niveles de transcripción o traducción. Una vez identificadas y clonadas, estas secuencias reguladoras aguas arriba pueden usarse en vectores de expresión diseñados para dirigir la expresión de un gen insertado de una manera espacial, temporal, de desarrollo o cuantitativa deseada. El ejemplo 58 describe un método para clonar secuencias aguas arriba de los ADNc extendidos o EST 5'.

10 2. Identificación de Secuencias Aguas Arriba con Actividades de Promoción o Regulación

EJEMPLO 58

Uso de ADNcs Extendidos o ESTs 5' para Clonar Secuencias Aguas Arriba del ADN Genómico

- 15 Las secuencias derivadas de ADNcs extendidos o ESTs 5' pueden usarse para aislar los promotores de los correspondientes genes utilizando técnicas de paseo cromosómico. En una técnica de paseo cromosómico, que utiliza el kit GenomeWalker® disponible en Clontech, cinco muestras de ADN genómico completo se digieren cada una con una enzima de restricción diferente que tiene un sitio de reconocimiento de 6 bases y deja un extremo romo. Después de la digestión, los adaptadores de oligonucleótidos se ligan a cada extremo de los fragmentos de ADN genómico resultantes.

- 20 Para cada uno de los cinco bancos de ADN genómico se realiza una primera reacción de PCR según las instrucciones del fabricante (que se incorporan a la presente memoria como referencia), utilizando un cebador adaptador externo proporcionado en el kit y un cebador específico de gen externo. El cebador específico de gen debe seleccionarse para que sea específico para el ADNc extendido o EST 5' de interés, y debe tener una temperatura de fusión, longitud y localización en el ADNc extendido o EST 5' que sea coherente con su uso en las reacciones de PCR. Cada primera reacción de PCR
25 contiene 5 ng de ADN genómico, 5 µl de 10 x tampón de reacción Tth, 0,2 mM de cada dNTP, 0,2 µM de cada cebador adaptador externo y cebador específico de gen externo, 1,1 mM de Mg(OAc)₂, y 1 µl de la mezcla 50 x polimerasa Tth en un volumen total de 50 µl. El ciclo de reacción para la primera reacción de PCR es como sigue: 1 min - 94°C/2 seg - 94°C, 3 min - 72°C (7 ciclos)/2 seg - 94°C, 3 min - 67°C (32 ciclos)/5 min - 67°C.

- 30 El producto de la primera reacción de PCR se diluye y se utiliza como molde para una segunda reacción de PCR según las instrucciones del fabricante utilizando una pareja de cebadores anidados que se localizan internamente en el amplicón resultante de la primera reacción de PCR. Por ejemplo, 5 µl del producto de reacción de la primera mezcla de reacción de PCR pueden diluirse 180 veces. Las reacciones se realizan en un volumen de 50 µl que tiene una composición idéntica a la de la primera
35 reacción de PCR, excepto que se utilizan los cebadores anidados. El primer cebador anidado es específico para el adaptador, y se proporciona en el kit GenomeWalker®. El segundo cebador anidado es específico del ADNc extendido o EST 5' particular para el cual debe clonarse el promotor, y debe tener una temperatura de fusión, longitud y localización en el ADNc extendido o EST 5' que sea coherente con su uso en las reacciones de PCR. Los parámetros de reacción de la segunda reacción de PCR son los siguientes: 1 min - 94°C/2 seg - 94°C, 3 min - 72°C (6 ciclos)/2 seg - 94°C, 3 min - 67°C (25 ciclos)/5 min - 67°C. El producto de la segunda reacción de PCR se purifica, se clona y se secuencia utilizando técnicas convencionales.

- 45 Como alternativa, pueden construirse dos o más bancos de ADN genómico humano utilizando dos o más enzimas de restricción. El ADN genómico digerido se clona en vectores que pueden convertirse en ADN monocatenario, circular o lineal. Un oligonucleótido biotinilado que comprende al menos 15 nucleótidos de la secuencia del ADNc extendido o EST 5' se hibrida con el ADN monocatenario. Los híbridos entre el oligonucleótido biotinilado y el ADN monocatenario que contiene la secuencia del ADNc extendido o EST 5' se aíslan como se describió en el Ejemplo 29 anterior. Después, el ADN monocatenario que contiene la secuencia del ADNc extendido o EST 5' se libera de las cuentas y se
50 convierte en ADN bicatenario utilizando un cebador específico para la secuencia del ADNc extendido o EST 5', o un cebador que corresponde a una secuencia incluida en el vector de clonación. El ADN bicatenario resultante se transforma en bacterias. Los ADN que contienen las secuencias de EST 5' o ADNc extendido se identifican mediante PCR de colonias o hibridación de colonias.

- 55 Cuando las secuencias genómicas aguas arriba se han clonado y secuenciado como se describió anteriormente, pueden identificarse presuntos promotores y sitios de inicio de la transcripción dentro de las secuencias aguas arriba, comparando las secuencias aguas arriba de los ADNcs extendidos o ESTs 5' con bases de datos que contienen sitios de inicio de la transcripción, sitios de unión del factor de transcripción o secuencias promotoras conocidos.

Además, los promotores en las secuencias aguas arriba pueden identificarse utilizando vectores indicadores de promotores como se describe en el Ejemplo.

EJEMPLO 59

Identificación de Promotores en Secuencias Clonadas Aguas Arriba

- 5 Las secuencias genómicas aguas arriba de los ADNcs extendidos o ESTs 5' se clonan en un vector indicador de promotores adecuado, tal como los vectores indicadores de promotores pSEAP-Basic, pSEAP-Enhancer, p β galBasic, p β gal-Enhancer o pEGFP-1, disponibles en Clontech. Brevemente, cada uno de estos vectores indicadores de promotores incluye múltiples sitios de clonación colocados aguas arriba de un gen indicador que codifica una proteína fácilmente analizable, tal como fosfatasa alcalina secretada, β -galactosidasa, o proteína fluorescente verde. Las secuencias aguas arriba de los
- 10 ADNcs extendidos o ESTs 5' se insertan en los sitios de clonación aguas arriba del gen indicador en ambas orientaciones y se introducen en una célula anfitriona adecuada. Se analiza el nivel de proteína indicadora y se compara con el nivel obtenido a partir de un vector que carece de un inserto en el sitio de clonación. La presencia de un nivel de expresión elevado en el vector que contiene el inserto, con respecto al vector de control, indica la presencia de un promotor en el inserto. Si fuera necesario, las
- 15 secuencias aguas arriba pueden clonarse en vectores que contienen un potenciador para aumentar los niveles de transcripción de secuencias promotoras débiles. Un nivel significativo de expresión por encima de la observada cuando el vector carece de un inserto indica que una secuencia promotora está presente en la secuencia insertada aguas arriba.
- 20 Las células anfitrionas apropiadas para los vectores indicadores de promotores pueden elegirse basándose en los resultados de la determinación, descrita anteriormente, de los patrones de expresión de los ADNcs extendidos o ESTs. Por ejemplo, si el análisis del patrón de expresión indica que el ARNm correspondiente a un ADNc extendido o EST 5' particular se expresa en fibroblastos, el vector indicador de promotores puede introducirse en una línea celular de fibroblastos humanos.
- 25 Las secuencias promotoras dentro del ADN genómico aguas arriba pueden definirse aún más construyendo deleciones anidadas en el ADN aguas arriba utilizando técnicas convencionales, tales como una digestión con exonucleasa III. Los fragmentos de deleción resultantes pueden insertarse en el vector indicador de promotores para determinar si la deleción ha reducido o eliminado la actividad promotora. De esta manera pueden definirse los límites de los promotores. Si se desea, pueden
- 30 identificarse sitios reguladores individuales potenciales dentro del promotor utilizando mutagénesis dirigida específica de sitio o barrido de conectores para eliminar los sitios de unión del factor de transcripción potenciales dentro del promotor, de forma individual o en combinación. Los efectos de estas mutaciones sobre los niveles de transcripción pueden determinarse insertando mutaciones en los sitios de clonación en los vectores indicadores de promotores.

EJEMPLO 60

Clonación e Identificación de Promotores

- 35 Utilizando el método descrito en el Ejemplo 58 anterior con EST 5', se obtienen secuencias aguas arriba de varios genes. Utilizando la pareja 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 denominación interna P13H2 (SEQ ID NO: 31).
- 40 Utilizando la pareja de cebadores GTA CCA GGG 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 denominación interna P15B4 (SEQ ID NO: 34).
- 45 Utilizando la pareja 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 denominación interna P29B6 (SEQ ID NO: 37).
- 50 La Figura 4 proporciona una descripción esquemática de los promotores aislados, y de la manera en que son ensamblados con las correspondientes etiquetas 5'. Las secuencias aguas arriba se seleccionaron en busca de la presencia de motivos parecidos a sitios de unión del factor de transcripción, o sitios de inicio de la transcripción conocidos, utilizando el programa de ordenador MatInspector, descarga 2.0, Agosto 1996.
- 55 La Tabla VII describe los sitios de unión del factor de transcripción presentes en cada uno de estos promotores. Las columnas indicadas como "matriz" proporcionan el nombre de la matriz MatInspector utilizada. La columna indicada como "posición" proporciona la posición 5' del sitio del promotor. La numeración de la secuencia comienza desde el sitio de transcripción, según se determina emparejando las secuencias genómicas con la secuencia de la EST 5'. La columna indicada como "orientación" indica la hebra de ADN sobre la cual se encuentra el sitio, siendo la hebra + la secuencia

codificante, según se determina emparejando la secuencia genómica con la secuencia de la EST 5'. La columna indicada como "puntuación" proporciona la puntuación MatInspector encontrada en ese sitio. La columna indicada como "longitud" proporciona la longitud del sitio en nucleótidos. La columna indicada como "secuencia" proporciona la secuencia del sitio encontrado.

- 5 Los clones bacterianos que contienen los plásmidos que contienen las secuencias promotoras descritas anteriormente se conservan, en la actualidad, en los laboratorios de los autores de la presente invención con los números de identificación internos proporcionados anteriormente. Los insertos pueden recuperarse a partir del material depositado cultivando una parte alícuota del clon bacteriano apropiado en el medio apropiado. El ADN del plásmido puede aislarse a continuación utilizando procedimientos de aislamiento de plásmidos familiares para los expertos en la técnica, tales como procedimientos de aislamiento de plásmidos de minipreps de lisis alcalina o de lisis alcalina a gran escala. Si se desea, el ADN del plásmido puede enriquecerse aún más mediante centrifugación sobre un gradiente de cloruro de cesio, cromatografía de exclusión molecular, o cromatografía de intercambio aniónico. El ADN del plásmido obtenido utilizando estos procedimientos puede manipularse a continuación utilizando mecanismos de clonación convencionales familiares para los expertos en la técnica. Como alternativa, puede realizarse una PCR con cebadores diseñados en ambos extremos de la inserción de la EST. El producto de PCR que corresponde a las ESTs 5' puede manipularse a continuación utilizando mecanismos de clonación convencionales familiares para los expertos en la técnica.
- 10
- 15
- 20 Los promotores y otras secuencias reguladoras localizadas aguas arriba de los ADNcs extendidos o ESTs 5' pueden utilizarse para diseñar vectores de expresión capaces de dirigir la expresión de un gen insertado de una manera espacial, temporal, de desarrollo y cuantitativa deseada. Un promotor capaz de dirigir los patrones espaciales, temporales, de desarrollo y cuantitativos deseados puede seleccionarse utilizando los resultados de los análisis de 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 promotora aguas arriba de un ADNc extendido o EST 5' derivado de un ARNm que se expresa con un nivel elevado en el músculo, según se determina mediante el método del Ejemplo 26, puede emplearse en el vector de expresión.
- 25
- 30 Preferiblemente, el promotor deseado se coloca cerca de múltiples sitios de restricción para facilitar la clonación del inserto deseado aguas abajo del promotor, de forma que el promotor sea capaz de dirigir la expresión del gen insertado. El promotor puede insertarse en esqueletos de ácidos nucleicos convencionales diseñados para la replicación extracromosómica, la integración en los cromosomas anfitriones, o la expresión transitoria. Los esqueletos adecuados para los presentes vectores de expresión incluyen esqueletos retrovíricos, esqueletos de episomas eucarióticos, tales como SV40 o virus del papiloma bovino, esqueletos de episomas bacterianos, o cromosomas artificiales.
- 35
- Preferiblemente, los vectores de expresión también incluyen una señal de poliA aguas abajo de los múltiples sitios de restricción para dirigir la poliadenilación del ARNm transcrito a partir del gen insertado en el vector de expresión.
- 40 Después de la identificación de las secuencias promotoras utilizando los procedimientos de los Ejemplos 58-60, pueden identificarse proteínas que interaccionan con el promotor, como se describe en el Ejemplo 61, a continuación.

EJEMPLO 61

Identificación de Proteínas que Interaccionan con Secuencias Promotoras, Secuencias Reguladoras Aguas Arriba, o ARNm

- 45 Las secuencias dentro de la región promotora que es probable que se unan a factores de transcripción pueden identificarse mediante homología con sitios de unión de factores de transcripción conocidos, o mediante mutagénesis convencional o análisis de delección de plásmidos indicadores que contienen la secuencia promotora. Por ejemplo, pueden realizarse delecciones en un plásmido indicador que contiene la secuencia promotora de interés conectada operablemente a un gen indicador analizable.
- 50 Los plásmidos indicadores que portan diversas delecciones dentro de la región promotora se transfectan en una célula anfitriona adecuada, y se evalúan los efectos de las delecciones sobre los niveles de expresión. Los sitios de unión del factor de transcripción dentro de las regiones en las que las delecciones reducen los niveles de expresión pueden localizarse con más precisión utilizando mutagénesis dirigida específica del sitio, análisis de barrido de conectores u otros mecanismos familiares para los expertos en la técnica.
- 55
- Los ácidos nucleicos que codifican proteínas que interaccionan con secuencias en el promotor pueden identificarse utilizando sistemas de un híbrido, tales como los descritos en el manual que acompaña el kit Matchmaker One-Hybrid System, disponible en Clontech (Núm. de catálogo K1603-1). Brevemente, el sistema Matchmaker One-Hybrid se utiliza como sigue. La secuencia diana para la cual se desean
- 60 identificar las proteínas de unión se clona aguas arriba de un gen indicador seleccionable, y se integra

en el genoma de levaduras. Preferiblemente, se insertan múltiples copias de las secuencias diana en el plásmido indicador en tándem. Un banco formado por fusiones entre los ADNcs que se van a evaluar por su capacidad para unirse al promotor y el dominio de activación de un factor de transcripción de levaduras, tal como GAL4, se transforma en la cepa de levaduras que contiene la secuencia indicadora integrada. Las levaduras se cultivan en placa en medios selectivos para seleccionar células que expresan el marcador seleccionable conectado a la secuencia promotora. Las colonias que crecen en los medios selectivos contienen genes que codifican proteínas que se unen a la secuencia diana. Los insertos en los genes que codifican las proteínas de fusión se caracterizan más a fondo mediante secuenciación. Además, los insertos pueden insertarse en vectores de expresión o en vectores de transcripción *in vitro*. La unión de los polipéptidos codificados por los insertos en el ADN del promotor puede confirmarse mediante mecanismos familiares para los expertos en la técnica, tales como análisis de desplazamiento en gel o análisis de protección de ADNasa.

VII. Uso de ESTs 5' (o ADNcs o ADNs Genómicos Obtenibles a partir de Éstas) en Terapia Génica

La presente invención también comprende el uso de las ESTs 5' (o ADNcs o ADNs genómico obtenibles a partir de éstas) en estrategias de terapia génica, incluyendo estrategias antisentido y de triple hélice, como se describe en los Ejemplos 62 y 63, a continuación. En los enfoques antisentido, secuencias de ácidos nucleicos complementarias a un ARNm se hibridan con el ARNm intracelularmente, bloqueando, con ello, la expresión de la proteína codificada por el ARNm. Las secuencias antisentido pueden evitar la expresión génica mediante una diversidad de mecanismos. Por ejemplo, las secuencias antisentido pueden inhibir la capacidad de los ribosomas para traducir el ARNm. Como alternativa, las secuencias antisentido pueden bloquear el transporte del ARNm desde el núcleo al citoplasma, limitando, con ello, la cantidad de ARNm disponible para la traducción. Otro mecanismo mediante el cual dichas secuencias antisentido pueden inhibir la expresión de genes es interfiriendo en el corte y empalme del ARNm. En otra estrategia, el ácido nucleico antisentido puede incorporarse a una ribozima capaz de romper específicamente el ARNm diana.

EJEMPLO 62

Preparación y Uso de Oligonucleótidos Antisentido

Las moléculas de ácidos nucleicos antisentido para ser utilizadas en la terapia génica pueden ser secuencias de ADN o ARN. Pueden comprender una secuencia complementaria a la secuencia de la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta). Los ácidos nucleicos antisentido deben tener una longitud y temperatura de fusión suficientes para permitir la formación de un dúplex intracelular con suficiente estabilidad para inhibir la expresión del ARNm en el dúplex. Las estrategias para diseñar ácidos nucleicos antisentido adecuados para su uso en terapia génica son descritos por Green et al., Ann. Rev. Biochem., 55: 569-597, 1986; e Izant y Weintraub, Cell, 36: 1007-1015, 1984.

En algunas estrategias, se obtienen moléculas antisentido a partir de una secuencia de nucleótidos que codifica una proteína invirtiendo la orientación de la región codificante con respecto a un promotor, de forma que se transcribe la hebra opuesta a la que se transcribe normalmente en la célula. Las moléculas antisentido pueden transcribirse utilizando sistemas de transcripción *in vitro*, tales como aquellos que emplean polimerasa T7 o SP6, para generar el transcrito. Otro enfoque implica la transcripción de los ácidos nucleicos antisentido *in vivo* conectando operablemente ADN que contiene la secuencia antisentido, con un promotor en un vector de expresión.

Como alternativa, pueden sintetizarse *in vitro* oligonucleótidos que son complementarios a la hebra que normalmente se transcribe en la célula. Así, los ácidos nucleicos antisentido son complementarios al correspondiente ARNm, y son capaces de hibridarse con el ARNm para crear un dúplex. En algunas realizaciones, las secuencias antisentido pueden contener esqueletos de azúcar fosfato modificados para aumentar la estabilidad y hacerlas menos sensibles a la actividad ARNasa. Los ejemplos de modificaciones adecuadas para su uso en las estrategias antisentido son descritas por Rossi et al., Pharmacol. Ther., 50(2): 245-254, 1991.

Pueden utilizarse diversos tipos de oligonucleótidos antisentido complementarios a la secuencia de la EST 5' (o ADNc o ADN genómico obtenible a partir de ésta). En una realización preferida, se utilizan los oligonucleótidos antisentido estables y semiestables descritos en la Solicitud de Patente Internacional PCT Núm. WO94/23026. En estas moléculas, el extremo 3', o ambos extremos 3' y 5', están implicados en enlaces de hidrógeno intramoleculares entre pares de bases complementarias. Estas moléculas son más capaces de aguantar los ataques de exonucleasas, y muestran mayor estabilidad comparadas con oligonucleótidos antisentido convencionales.

En otra realización preferida, se utilizan los oligodesoxinucleótidos antisentido contra el virus del Herpes simplex de tipo 1 y 2 descritos en la Solicitud de Patente Internacional Núm. WO 95/04141.

- En otra realización preferida, se utilizan los oligonucleótidos antisentido covalentemente entrelazados descritos en la Solicitud de Patente Internacional Núm. WO 96/31523. Estos oligonucleótidos bicatenarios o monocatenarios comprenden uno o más, respectivamente, enlaces entrelazados covalentes inter- o intraoligonucleótido, en los que el enlace consiste en un enlace amida entre un grupo amina primario de una hebra y un grupo carboxilo de la otra hebra o de la misma hebra, respectivamente, estando el grupo amina primario directamente sustituido en la posición 2' del anillo de monosacárido del nucleótido de la hebra, y estando portado el grupo carboxilo por un grupo espaciador alifático sustituido en un nucleótido o análogo de nucleótido de la otra hebra o de la misma hebra, respectivamente.
- 10 También pueden utilizarse los oligodesoxinucleótidos y oligonucleótidos antisentido descritos en la Solicitud de Patente Internacional Núm. WO 92/18522. Estas moléculas son estables frente a la degradación, y contienen al menos una secuencia de reconocimiento de control de la transcripción que se une a proteínas control, y son eficaces como reclamo para éstas. Estas moléculas pueden
- 15 contener estructuras de "horquilla", estructuras de tipo "dumbbell" (de dos horquillas), estructuras de tipo "dumbbell modificadas", estructuras de reclamo "entrecruzadas", y estructuras de "bucle".
- En otra realización preferida, se utilizan los oligonucleótidos bicatenarios cíclicos descritos en la Solicitud de Patente Europea Núm. 0572287 A2. Estos oligonucleótidos "dumbbell" ligados contienen un sitio de unión para un factor de transcripción e inhiben la expresión del gen bajo el control del factor de transcripción secuestrando el factor.
- 20 También se contempla el uso de los oligonucleótidos antisentido cerrados descrito en la Solicitud de Patente Internacional Núm. WO 92/19732. Debido a que estas moléculas no tienen extremos libres, son más resistentes a la degradación por exonucleasas que los oligonucleótidos convencionales. Estos oligonucleótidos pueden ser multifuncionales, interaccionando con varias regiones que no son adyacentes al ARNm diana.
- 25 El nivel apropiado de ácidos nucleicos antisentido requerido para inhibir la expresión génica puede determinarse utilizando un análisis de expresión *in vitro*. Las moléculas antisentido pueden introducirse en las células mediante difusión, inyección, infección, transfección o importación mediada por la región h, utilizando procedimientos conocidos en la técnica. Por ejemplo, los ácidos nucleicos antisentido pueden introducirse en el cuerpo como un oligonucleótido descubierto o desnudo, un
- 30 oligonucleótido encapsulado en lípidos, una secuencia de oligonucleótido encapsulada por proteínas víricas, o como un oligonucleótido operablemente conectado a un promotor contenido en un vector de expresión. El vector de expresión puede ser cualquiera de una diversidad de vectores de expresión conocidos en la técnica, incluyendo vectores retrovíricos o víricos, vectores capaces de la replicación extracromosómica, o vectores de integración. Los vectores pueden ser de ADN o ARN.
- 35 Las moléculas antisentido se introducen en muestras celulares en una serie de concentraciones diferentes, preferiblemente entre 1×10^{-10} M a 1×10^{-4} M. Cuando se identifica la concentración mínima que puede controlar de forma adecuada la expresión génica, la dosis optimizada se traduce en una dosificación adecuada para uso *in vivo*. Por ejemplo, una concentración inhibidora en cultivo de 1×10^{-7} se traduce a una dosis de aproximadamente 0,6 mg/kg de peso corporal. Pueden ser posibles
- 40 unos niveles de oligonucleótidos cercanos a 100 mg/kg de peso corporal o mayores, después de someter a ensayo la toxicidad del oligonucleótido en animales de laboratorio. También se contempla que las células del vertebrado se retiren, se traten con el oligonucleótido antisentido, y se reintroduzcan en el vertebrado.
- 45 También se contempla que la secuencia del oligonucleótido antisentido se incorpore a una secuencia de ribozima para permitir que el oligonucleótido antisentido se una de forma específica y rompa su ARNm diana. Para las aplicaciones técnicas de ribozimas y oligonucleótidos antisentido, véase Rossi et al., supra.
- En una aplicación preferida de esta invención, en primer lugar se identifica el polipéptido codificado por el gen, de modo que la eficacia de la inhibición del oligonucleótido antisentido sobre la traducción pueda controlarse utilizando técnicas que incluyen, pero no se limitan a ensayos mediados por anticuerpos, tales como RIA y ELISA, análisis funcionales, o radiomarcaje.
- 50 Las ESTs 5' de la presente invención (o ADNcs o ADNs genómicos obtenibles a partir de éstas) también pueden utilizarse en enfoques de terapia génica basados en la formación de triples hélices intracelulares. Los oligonucleótidos de triple hélice se utilizan para inhibir la transcripción de un genoma. Son particularmente útiles para estudiar alteraciones en la actividad celular, asociada con un gen particular. Las secuencias de EST 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) de la presente invención o, más preferiblemente, una porción de esas secuencias, puede utilizarse para inhibir la expresión génica en individuos que tienen enfermedades asociadas con la expresión de un gen particular. De forma similar, una porción de las secuencias de EST 5' (o ADNcs o ADNs genómicos obtenibles a partir de éstas) puede utilizarse para estudiar el efecto de inhibición de la
- 55 transcripción de un gen particular dentro de una célula. De forma tradicional, las secuencias de
- 60

- homopurina se consideraron las más útiles para las estrategias de triple hélice. Sin embargo, las secuencias de homopirimidina también pueden inhibir la expresión génica. Estos oligonucleótidos de homopirimidina se unen al surco mayor en las secuencias de homopurina:homopirimidina. Por tanto, ambos tipos de secuencias procedentes de la EST 5' o procedentes del gen correspondiente a la EST 5' se contemplan como dentro del alcance de esta invención.

EJEMPLO 63

Preparación y Uso de Sondas de Triple Hélice

- Las secuencias de las ESTs 5' (o ADNcs o ADNcs genómicos obtenibles a partir de éstas) se barren para identificar tramos de homopirimidina u homopurina de 10 unidades a 20 unidades que pueden utilizarse en estrategias basadas en triples hélices para inhibir la expresión génica. Después de la identificación de los tramos de homopirimidina u homopurina candidatos, su eficacia para inhibir la expresión génica se evalúa introduciendo cantidades variables de oligonucleótidos que contienen las secuencias candidatas en células de cultivo de tejidos que normalmente expresan el gen diana. Los oligonucleótidos pueden prepararse en un sintetizador de oligonucleótidos, o pueden adquirirse en el mercado en una empresa especializada en síntesis de oligonucleótidos a medida, tal como GENSET, París, Francia.

- Los oligonucleótidos pueden introducirse en las células utilizando una diversidad de métodos conocidos por los expertos en la técnica, incluyendo, pero sin limitarse a precipitación con fosfato de calcio, DEAE-dextrano, electroporación, transfección mediada por liposomas, o captación nativa.
- Las células tratadas se controlan para detectar una función celular alterada o una expresión génica reducida utilizando técnicas, tales como análisis de transferencia Northern, análisis de protección de ARNasa, o estrategias basadas en PCR, para controlar los niveles de transcripción del gen diana en células que se han tratado con el oligonucleótido. Las funciones celulares que se van a controlar se predicen basándose en las homologías del gen diana correspondiente al ADNc extendido a partir del cual derivaba el oligonucleótido, con secuencias génicas conocidas que se han asociado con una función particular. Las funciones celulares también pueden predecirse basándose en la presencia de fisiologías anómalas en células derivadas de individuos con una enfermedad heredada concreta, en particular cuando el ADNc extendido está asociado con la enfermedad utilizando las técnicas descritas en el Ejemplo 56.
- Los oligonucleótidos que son eficaces para inhibir la expresión génica en células de cultivo de tejidos pueden introducirse a continuación *in vivo* utilizando las técnicas descritas anteriormente y en el Ejemplo 62, en una dosificación calculada basada en los resultados *in vitro*, como se describe en el Ejemplo 62.

- En algunas realizaciones, los anómeros naturales (beta) de las unidades de oligonucleótidos pueden sustituirse por anómeros alfa para producir oligonucleótidos más resistentes a nucleasas. Además, puede unirse un agente intercalante, tal como bromuro de etidio o similares, al extremo 3' del alfa-oligonucleótido para estabilizar la triple hélice. Para obtener información sobre la generación de oligonucleótidos estables para la formación de triples hélices, véase Griffin et al., Science, 245: 967-971, 1989.

EJEMPLO 64

Uso de ADNc Obtenidos Utilizando las ESTs 5' para Expresar una Proteína Codificada en un Organismo Anfitrión

- Los ADNc obtenidos como se describió anteriormente utilizando las ESTs 5' también pueden utilizarse para expresar una proteína codificada en un organismo anfitrión para producir un efecto beneficioso. En estos procedimientos, la proteína codificada puede expresarse de modo transitorio en el organismo anfitrión, o puede expresarse de forma estable en el organismo anfitrión. La proteína codificada puede tener cualquiera de las actividades descritas anteriormente. La proteína codificada puede ser una proteína de la que el carece organismo anfitrión o, como alternativa, la proteína codificada puede aumentar los niveles existentes de la proteína en el organismo anfitrión.
- Un ADNc extendido completo que codifica el péptido señal y la proteína madura, o un ADNc extendido que codifica sólo la proteína madura, se introduce en el organismo anfitrión. El ADNc extendido puede introducirse en el organismo anfitrión utilizando una diversidad de mecanismos conocidos por los expertos en la técnica. Por ejemplo, el ADNc extendido puede inyectarse en el organismo anfitrión como ADN desnudo, de forma que la proteína codificada se exprese en el organismo anfitrión, produciendo, con ello, un efecto beneficioso.

Como alternativa, el ADNc extendido puede clonarse en un vector de expresión aguas abajo de un promotor que es activo en el organismo anfitrión. El vector de expresión puede ser cualquier vector de

- expresión diseñado para uso en terapia génica, incluyendo vectores víricos o retrovíricos. El vector de expresión puede introducirse directamente en el organismo anfitrión, de forma que la proteína codificada se exprese en el organismo anfitrión para producir un efecto beneficioso. En otro enfoque, el vector de expresión puede introducirse en células *in vitro*. Las células que contienen el vector de expresión después se seleccionan y se introducen en el organismo anfitrión, donde expresan la proteína codificada para producir un efecto beneficioso.

EJEMPLO 65

Uso de Péptidos Señal Codificados por las ESTs 5' o Secuencias Obtenidas a partir de Éstas para Importar Proteínas hacia Células

- La región hidrófoba (h) corta central de péptidos señal codificados por las ESTs 5' o ADNcs extendidos derivados de los SEQ ID NO: 38-270 también pueden utilizarse como portador para importar un péptido o proteína de interés, denominado cargamento, hacia células de cultivo de tejidos (Lin et al., J. Biol. Chem., 270: 14225-14258, 1995; Du et al., J. Peptide Res., 51: 235-243, 1998; Rojas et al., Nature Biotech., 16: 370-375, 1998).
- Cuando se van a trasladar péptidos permeables a células de tamaño limitado (aproximadamente hasta 25 aminoácidos) a través de la membrana celular, puede utilizarse síntesis química para añadir la región h al extremo C-terminal o al extremo N-terminal del péptido cargamento de interés. Como alternativa, cuando proteínas o péptidos más largos se van a importar hacia células, los ácidos nucleicos pueden modificarse genéticamente, utilizando mecanismos conocidos por los expertos en la técnica, para conectar 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 cargamento. Estos ácidos nucleicos modificados genéticamente se traducen a continuación *in vitro* o *in vivo* después de una transfección en células apropiadas, utilizando técnicas convencionales para producir el polipéptido permeable a células resultante. Las células anfitrionas adecuadas se incuban a continuación sencillamente con el polipéptido permeable a células, que después se traslada a través de la membrana.
- Este método puede aplicarse para estudiar diversas funciones intracelulares y procesos celulares. Por ejemplo, se ha utilizado para sondear dominios funcionalmente relevantes de proteínas intracelulares y para estudiar las interacciones proteína-proteína implicadas en las vías de transducción de señales (Lin et al., supra; Lin et al., J. Biol. Chem., 271: 5305-5308, 1996; Rojas et al., J. Biol. Chem., 271: 27456-27461, 1996; Liu et al., Proc. Natl. Acad. Sci. USA, 93: 11819-11824, 1996; Rojas et al., Bioch. Biophys. Res. Commun., 234: 675-680, 1997).
- Estas técnicas pueden utilizarse en terapia celular para importar proteínas que producen efectos terapéuticos. Por ejemplo, células aisladas de un paciente pueden tratarse con proteínas terapéuticas importadas y después reintroducirse en el organismo anfitrión.
- Como alternativa, la región h de los péptidos señal de la presente invención puede utilizarse en combinación con una señal de localización nuclear para dirigir ácidos nucleicos hacia el núcleo celular. Estos oligonucleótidos pueden ser oligonucleótidos antisentido u oligonucleótidos diseñados para formar triples hélices, como se describió en los ejemplo 62 y 63, respectivamente, para inhibir el procesamiento y/o maduración de un ARN celular diana.
- Como se analizó anteriormente, los ADNcs o porciones de éstos obtenidos utilizando las ESTs 5' de la presente invención pueden utilizarse para diversos fines. Los polinucleótidos pueden utilizarse para expresar proteínas recombinantes para análisis, caracterización o uso terapéutico; como marcadores para tejidos en los que la correspondiente proteína se expresa con preferencia (de manera constitutiva o en una etapa particular de la diferenciación o desarrollo del tejido, o en estados de enfermedad); como marcadores de peso molecular en geles de transferencia Southern; como marcadores o etiquetas de cromosomas (cuando están marcados) para identificar cromosomas o para cartografiar posiciones de genes relacionados; para comparar con secuencias de ADN endógenas en pacientes para identificar potenciales trastornos genéticos; como sondas para hibridarse y, por tanto, descubrir nuevas secuencias de ADN relacionadas; como fuente de información para derivar cebadores de PCR para realizar una huella genética; para seleccionar y fabricar oligómeros para su unión a un "chip genético" u otro soporte, incluyendo para el estudio de patrones de expresión; para producir anticuerpos antiproteína utilizando técnicas de inmunización de ADN; y como antígeno para producir anticuerpos anti-ADN o para provocar otra respuesta inmunológica. Cuando el polinucleótido codifica una proteína que se une o potencialmente se une a otra proteína (tal como, por ejemplo, en una interacción receptor-ligando), el polinucleótido también puede emplearse en ensayos de trampa de interacción (tales como, por ejemplo, los descritos por Gyuris et al., Cell, 75: 791-803, 1993) para identificar polinucleótidos que codifican la otra proteína con la cual se produce la unión, o para identificar inhibidores de la interacción de unión.
- Las proteínas o polipéptidos proporcionados por la presente invención pueden utilizarse, de modo similar, en ensayos para determinar la actividad biológica, incluyendo un panel de múltiples proteínas

5 para un escrutinio de alto rendimiento; para producir anticuerpos o para provocar otra respuesta inmunológica; como reactivo (incluyendo el reactivo marcado) en análisis diseñados para determinar cuantitativamente los niveles de la proteína (o su receptor) en fluidos biológicos; como marcadores para tejidos en los que la correspondiente proteína se expresa de manera preferente (de forma constitutiva o en una etapa concreta de la diferenciación o desarrollo del tejido, o en estados de enfermedad); y, por supuesto, para aislar receptores o ligandos correlativos. Cuando la proteína se une o potencialmente se une a otra proteína (tal como, por ejemplo, en una interacción receptor-ligando), la proteína puede utilizarse para identificar la otra proteína con la que se produce la unión, o para identificar inhibidores de la interacción de unión. Las proteínas implicadas en estas interacciones de unión también pueden utilizarse para seleccionar inhibidores o agonistas peptídicos o de molécula pequeña de la interacción de unión.

Cualquiera o todas estas utilidades de investigación se pueden desarrollar en un formato de calidad de reactivo o kit para la comercialización como productos de investigación.

15 Los métodos para realizar los usos enumerados anteriormente son muy conocidos por los expertos en la técnica. Las referencias que describen dichos métodos incluyen, sin limitación, Molecular Cloning: A Laboratory Manual, 2ª ed., Cold Spring Harbor Laboratory Press, Sambrook, Fritsch y Maniatis eds., 1989, y Methods in Enzymology: Guide to Molecular Cloning Techniques, Academic Press, Berger y Kimmel eds., 1987.

20 Los polinucleótidos y proteínas de la presente invención también pueden utilizarse como suplementos o fuentes nutricionales. Estos usos incluyen, sin limitación, el uso de un suplemento de proteínas o aminoácidos, el uso como una fuente de carbono, el uso como una fuente de nitrógeno, y el uso como una fuente de carbohidratos. En estos casos, la proteína o polinucleótido de la invención puede añadirse al pienso de un organismo particular, o puede administrarse como una preparación sólida o líquida separada, tal como en forma de polvo, píldoras, disoluciones, suspensiones o cápsulas. En el caso de los microorganismos, la proteína o polinucleótido de la invención puede añadirse al medio en el que se cultiva o sobre el que se cultiva el microorganismo.

30 Aunque esta invención se ha descrito en términos de ciertas realizaciones preferidas, también se encuentran dentro del alcance de esta invención otras realizaciones que serán evidentes para los expertos normales en la técnica a la vista de la descripción de la presente memoria. Por consiguiente, se pretende que el alcance de la invención esté definido solamente por la referencia a las reivindicaciones adjuntas.

Tabla 1: Parámetros utilizados para cada etapa de análisis de EST

Etapa	Características de la Búsqueda		Características de la Selección		
	Programa	Hebra	Parámetros	(%) Identidad	Longitud (pb)
miscelánea	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	blasin	ambas	S=72	70	40
Repeticiones	blastn	ambas	S=72	70	40

	Características de la Búsqueda		Características de la Selección		
Etapas	Programa	Hebra	Parámetros	(%) Identidad	Longitud (pb)
Promotores	blastn	superior	S=54X=16	90	15†
Vertebrados	fasta*	ambas	S=108	90	30
EST	blastn	ambas	S=108 X=18	90	30
Proteínas	blastx□	superior	E = 0.001	-	-

* uso de dispositivo de barrido de base de datos "Quick Fast"

† el alineamiento se limitó aún más para empezar antes de 10 pb del extremo 5' de la EST

□ se utiliza la matriz de sustitución BLOSUM62

TABLA II

SEQ. ID NO.	CATEGORIA	PUNTUACIÓN DE VON HEIJNE	FUENTE DE TEJIDO	DESIGNACIÓN INTERNA
ID73	nueva	6,9	Bazo	20-10-1-B12-PU
ID75	nueva	6,9	Ovario	26-23-2-A11-PU

TABLA III

SEQ. ID NO.	PÉPTIDO SEÑAL
ID73	MTMRHNWYPDLSPLWVLLCAHVLT
ID75	MTMPJWWTPDLSPLWVUIXAHVTL

TABLA IV

Puntuación mínima del péptido señal	proporción de falsos positivos	proporción 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

Puntuación mínima del péptido señal	proporción de falsos positivos	proporción de falsos negativos	proba(0,1)	proba(0,2)
5,5	0,05	0,127	0,659	0,813
6	0,04	0,163	0,694	0,838
6,5	0,033	0,202	0,725	0,855
7	0,025	0,248	0,763	0,878
7,5	0,021	0,304	0,78	0,889
8	0,015	0,368	0,816	0,909
8,5	0,012	0,418	0,836	0,92
9	0,009	0,512	0,856	0,93
9,5	0,007	0,581	0,863	0,934
10	0,006	0,679	0,835	0,919

TABLA V

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

Puntuación mínima del péptido señal	Todas las EST	EST nuevas	EST que se corresponden con EST públicas más cerca de 40 pb desde el inicio	EST que extienden más de 40 pb ARNm conocidos	EST que extienden más de 40 pb EST públicas
9,5	364	57	48	6	18
10	303	47	35	6	15

TABLA VI

Tejido	Todas las EST	EST nuevas	EST coincidentes con EST públicas más próximas que 40 pb desde el inicio	EST que extienden ARNm conocidos en más de 40 pb	EST que extienden EST públicas en 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

Tejido	Todas las EST	EST nuevas	EST coincidentes con EST públicas más próximas que 40 pb desde el inicio	EST que extienden ARNm conocidos en más de 40 pb	EST que extienden EST públicas en más de 40 pb
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
Testículo	131	68	25	1	8
Tiroides	17	8	2	0	2
Cordón Umbilical	55	17	12	1	3
Útero	28	15	3	0	2
No específico de tejido	568	48	177	2	28
Total	2677	947	601	23	150

Descripción de los Sitios de Unión del Factor de Transcripción presentes en promotores aislados a partir de Secuencias Señal Etiquetadas

TABLA VII

Secuencia del promotor P13H2 (646 pb):					
Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
CMYB_01	-502	+	0,983	9	TGTCAGTTG
MYOD_Q6	-501	-	0,981	10	CCCAACTGAC
S8_01	-444	-	0,960	11	AATAGAATTAG
S8_01	-425	+	0,968	11	AACTAAATTAG

Secuencia del promotor P13H2 (646 pb):					
Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
DELTAEF1_01	-390	-	0,960	11	GCACACCTCAG
GATA_C	-364	-	0,984	11	AGATAAATCCA
CMYB_01	-349	+	0,958	9	CTTCAGTTG
GATA1_02	-343	+	0,959	14	TTGTAGATAGGACA
GATA_C	-339	+	0,953	11	AGATAGGACAT
TAL1ALPHA47_01	-235	+	0,973	16	CATAACAGATGGTAAG
TAL1BETA47_01	-235	+	0,883	16	CATAACAGATGGTAAG
TAL1BETAITF2_01	-235	+	0,978	16	CATAACAGATGGTAAG
MYOD_Q8	-232	-	0,954	10	ACCATCTGTT
GATA1_04	-217	-	0,953	13	TCAAGATAAAGTA
IK1_01	-126	+	0,983	13	AGTTGGGAATTCC
IK2_01	-128	+	0,985	12	AGTTGGGAATTC
CREL_01	-123	+	0,962	10	TGGGAATTCC
GATA1_02	-98	+	0,950	14	TCAGTGATATGGCA
SRY_02	-41	-	0,951	12	TAAAACAAAACA
E2F_02	-33	+	0,957	6	TTTAGCGC
MZF1_01	-5	-	0,975	8	TGAGGGGA
Secuencia del promotor P15B4 (861 pb) :					
Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
NFY_Q6	-748	-	0,858	11	GGACCAATCAT
MZF1_01	-738	+	0,962	8	CCTGGGGA
CMYB_01	-884	+	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

Secuencia del promotor P13H2 (646 pb):					
Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
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_06	-190	+	0,981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0,958	11	TCCCACCTTCC
S8_01	5	-	0,992	11	GAGGCAATTAT
MZF1_01	16	-	0,986	6	AGAGGGGA
Secuencia del promotor P29B6 (555 pb) :					
Matriz	Posición	Orientación	Puntuación	Longitud	Secuencia
ARMT_01	-311	+	0,964	16	GGA CTCACGTGCTGCT
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	-	0,883	11	AGTGA CTGAAC
AP1FJ_02	-42	-	0,981	11	AGTGA CTGAAC
PADS_C	45	+	1,000	9	TGTGGTCTC

LISTADO DE SECUENCIAS

(1) INFORMACIÓN GENERAL:

(i) SOLICITANTE:

- 5 (A) Nombre : GENSET SA
(B) CALLE :24, RUE ROYALE
(C) CIUDAD: PARIS
(E) PAÍS : FRANCIA
(F) CÓDIGO POSTAL: 75008

10

EN (ii) TÍTULO DE LA INVENCION: ESTs 5' PARA PROTEÍNAS SECRETADAS EXPRESADAS
TESTÍCULO Y OTROS TEJIDOS

15 (iii) NÚMERO DE SECUENCIAS: 503

(v) FORMA LEGIBLE CON ORDENADOR:

- 20 (A) TIPO DE MEDIO: Disquete
(B) ORDENADOR: IBM PC compatible
(C) SISTEMA OPERATIVO: win95
(D) PROGRAMA: Word

(2) INFORMACIÓN PARA EL SEQ ID NO: 1:

25

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 30 (A) LONGITUD: 47 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIDAD: SENCILLA
(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

35

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: Caperuza
(B) LOCALIZACIÓN: 1

(D) OTRA INFORMACIÓN: m7Gppp añadido a 1

5 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 1:
GGCAUCCUAC CCCCAUCCAA UCCACCCUA ACUCCUCCCA UCUCCAC
47

(2) INFORMACIÓN PARA EL SEQ ID NO: 2:

10 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 46 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

(D) TOPOLOGÍA: LINEAL

15

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

20 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 2:
GCAUCCUACU CCCAUCCAAU UCCACCCUAA CUCCUCCCAU CUCCAC
46

(2) INFORMACIÓN PARA EL SEQ ID NO: 3:

25 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

(D) TOPOLOGÍA: LINEAL

30

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

35 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 3:
ATCAAGAATT CGCACGAGAC CATT
25

(2) INFORMACIÓN PARA EL SEQ ID NO: 4:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 25 pares de bases
- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: SENCILLA
- (D) TOPOLOGÍA: LINEAL

5

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 4:
TAATGGTCTC GTGCGAATTC TTGAT
25

10

(2) INFORMACIÓN PARA EL SEQ ID NO: 5:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

15

- (A) LONGITUD: 25 pares de bases
- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: SENCILLA
- (D) TOPOLOGÍA: LINEAL

20

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 5:
CCGACAAGAC CAACGTCAAG GCCGC
25

25

(2) INFORMACIÓN PARA EL SEQ ID NO: 6:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

30

- (A) LONGITUD: 25 pares de bases
- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: SENCILLA
- (D) TOPOLOGÍA: LINEAL

35

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 6:
TCACCAGCAG GCAGTGGCTT AGGAG
25

40

(2) INFORMACIÓN PARA EL SEQ ID NO: 7:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

5

(A) LONGITUD: 25 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

(D) TOPOLOGÍA: LINEAL

10

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 7:
AGTGATTCCT GCTACTTTGG ATGGC
25

15

(2) INFORMACIÓN PARA EL SEQ ID NO: 8:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

20

(A) LONGITUD: 25 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

(D) TOPOLOGÍA: LINEAL

25

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 8:
GCTTGGTCTT GTTCTGGAGT TTAGA
25

30

(2) INFORMACIÓN PARA EL SEQ ID NO: 9:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

35

(A) LONGITUD: 25 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

5 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 9:
TCCAGAATGG GAGACAAGCC AATTT
25

(2) INFORMACIÓN PARA EL SEQ ID NO: 10:

10 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

15 (D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

20 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 10:
AGGGAGGAGG AAACAGCGTG AGTCC
25

(2) INFORMACIÓN PARA EL SEQ ID NO: 11:

25 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

30 (D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

35 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 11:
ATGGGAAAGG AAAAGACTCA TATCA
25

(2) INFORMACIÓN PARA EL SEQ ID NO: 12:

40 (i) CARACTERÍSTICAS DE LA SECUENCIA:

- 5 (A) LONGITUD: 25 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: SENCILLA
(D) TOPOLOGÍA: LINEAL
- (ii) TIPO DE MOLÉCULA: Otro ácido nucleico
- 10 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 12:
AGCAGCAACA ATCAGGACAG CACAG
25
- (2) INFORMACIÓN PARA EL SEQ ID NO: 13:
- 15 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 25 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: SENCILLA
(D) TOPOLOGÍA: LINEAL
- 20 (ii) TIPO DE MOLÉCULA: Otro ácido nucleico
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 13:
ATCAAGAATT CGCACGAGAC CATT
25
- (2) INFORMACIÓN PARA EL SEQ ID NO: 14:
- 30 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 67 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: SENCILLA
(D) TOPOLOGÍA: LINEAL
- 35 (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 PARA EL SEQ ID NO: 15:

5

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 29 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

10

(C) CATENARIEDAD: SENCILLA

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

15

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 15:
CCAGCAGAGT CACGAGAGAG ACTACACGG
29

(2) INFORMATION: FOR SEQ ID NO: 16:

20

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

25

(C) CATENARIEDAD: SENCILLA

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

30

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 16:
CACGAGAGAG ACTACACGGT ACTGG
25

(2) INFORMACIÓN PARA EL SEQ ID NO: 17:

35

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 526 pares de bases

- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: DOBLE
- (D) TOPOLOGÍA: LINEAL

5 (ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

10 (A) ORGANISMO: Homo Sapiens
(F) TIPO DE TEJIDO: Ganglios linfáticos

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: complemento (261..376)
(C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 96 región 166..281 id N70479 est

(ix) CARACTERÍSTICAS:

20 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: complemento (380..486)
(C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 97 región 54..160 id N70479 est

25 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: complemento (110..145)
30 (C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 94 región 403..438 id N70479 est

(ix) CARACTERÍSTICAS:

35 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: complemento (196..229)
(C) MÉTODO DE IDENTIFICACIÓN: blastn

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

(ix) CARACTERÍSTICAS:

- 5 (A) NOMBRE/CLAVE: péptido señal
 (B) LOCALIZACIÓN: 90..140
 (C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne
 (D) OTRA INFORMACIÓN: puntuación 8,2 sec LLLITAILAVAVG/FP

10 (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 PARA EL SEQ ID NO: 18:

- 15 (i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 17 aminoácidos
 (B) TIPO: AMINOÁCIDO
 (D) TOPOLOGÍA: LINEAL

- 20 (ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

5 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: 1..17

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

10 (D) OTRA INFORMACIÓN: puntuación 8,2 sec 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 PARA EL SEQ ID NO: 19:

15

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 822 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

20 (C) CATENARIEDAD: DOBLE

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: ADNc

25 (vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

(D) ETAPA DEL DESARROLLO: Fetal

(F) TIPO DE TEJIDO: riñón

30

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 260..464

35 (C) MÉTODO DE IDENTIFICACIÓN: blastn

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

(ix) CARACTERÍSTICAS:

- 5 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: 118..184
(C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 98 región 98..164 id H57434 est

(ix) CARACTERÍSTICAS:

- 10 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: 56..113
(C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 98 región 35..92 id H57434 est

15

(ix) CARACTERÍSTICAS:

- 20 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: 454..485
(C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 100 región 348..379 id H57434 est

(ix) CARACTERÍSTICAS:

- 25 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: 118..545
(C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 98 región 1..428 id N27248 est

30

(ix) CARACTERÍSTICAS:

- 35 (A) NOMBRE/CLAVE: otro
(B) LOCALIZACIÓN: 65..369
(C) MÉTODO DE IDENTIFICACIÓN: blastn
(D) OTRA INFORMACIÓN: coincidencia 98 región 41..345 id H94779 est

(ix) CARACTERÍSTICAS:

- 5
- (A) NOMBRE/CLAVE: otro
 - (B) LOCALIZACIÓN: 61..399
 - (C) MÉTODO DE IDENTIFICACIÓN: blastn
 - (D) OTRA INFORMACIÓN: coincidencia 99 región 6..344 id H09880 est

(ix) CARACTERÍSTICAS:

- 10
- (A) NOMBRE/CLAVE: otro
 - (B) LOCALIZACIÓN: 408..458
 - (C) MÉTODO DE IDENTIFICACIÓN: blastn
 - (D) OTRA INFORMACIÓN: coincidencia 92 región 355..405 id H09880 est

(ix) CARACTERÍSTICAS:

- 15
- (A) NOMBRE/CLAVE: otro
 - (B) LOCALIZACIÓN: 60..399
 - (C) MÉTODO DE IDENTIFICACIÓN: blastn
 - (D) OTRA INFORMACIÓN: coincidencia 97 región 56..395 id H29351 est

20

(ix) CARACTERÍSTICAS:

- 25
- (A) NOMBRE/CLAVE: otro
 - (B) LOCALIZACIÓN: 393..432
 - (C) MÉTODO DE IDENTIFICACIÓN: blastn
 - (D) OTRA INFORMACIÓN: coincidencia 90 región 391..430 id H29351 - est

(ix) CARACTERÍSTICAS:

- 30
- (A) NOMBRE/CLAVE: péptido señal
 - (B) LOCALIZACIÓN: 346..408
 - (C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne
 - (D) OTRA INFORMACIÓN: puntuación 5,5 sec SFLPSALVIWTS/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
GTTTGTGAA GCAGTTACCA AGAATCTTCA ACCCTTTCCC ACAAAGCTA ATTGAGTACA	240
CGTTCCTGTT GAGTACACGT TCCTGTTGAT TTACAAAAGG TGCAGGTATG AGCAGGTCTG	300
AAGACTAACA TTTTGTGAAG TTGTAAAACA GAAACCTGT 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 AAKTTCATT CATGACCAA	602
Lys	
CTCTTCARAA ACATGTCTTT ACAAGCATAT CTCTTGATT GCTTTCTACA CTGTTGAATT	662
GTCTGGCAAT ATTTCTGCAG TGGAAAATTT GATTTARMTA GTTCTTGACT GATAAATATG	722
GTAAGGTGGG CTTTCCCCC TGTGTAATTG GCTACTATGT CTTACTGAGC CAAGTTGTAW	782
TTTGAAATAA AATGATATGA GAGTGACACA AAAAAAAAAA	822

5

(2) INFORMACIÓN PARA EL SEQ ID NO: 20:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

10

(A) LONGITUD: 21 aminoácidos

(B) TIPO: AMINOÁCIDO

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: PROTEÍNA

15

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: péptido señal
- (B) LOCALIZACIÓN: 1..21
- (C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne
- (D) OTRA INFORMACIÓN: puntuación 5,5 sec SFLPSALVIWTS/AF

5

(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

10 (2) INFORMACIÓN PARA EL SEQ ID NO: 21:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 405 pares de bases
- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: DOBLE
- (D) TOPOLOGÍA: LINEAL

15

(ii) TIPO DE MOLÉCULA: ADNc

20

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: Homo Sapiens
- (F) TIPO DE TEJIDO: Testículo

25

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: otro
- (B) LOCALIZACIÓN: complemento (103..398)
- (C) MÉTODO DE IDENTIFICACIÓN: blastn
- (D) OTRA INFORMACIÓN: coincidencia 96 región 1..296 id AA442893 est

30

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: péptido señal
- (B) LOCALIZACIÓN: 185..295

35

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

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

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

```

ATCACCTTCT TCTCCATCCT TSTCTGGGCC AGTCCCCARC CCAGTCCCTC TCCTGACCTG      60
CCCAGCCCAA GTCAGCCTTC AGCAGCGGCT 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

```

5

(2) INFORMACIÓN PARA EL SEQ ID NO: 22:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

10

(A) LONGITUD: 37 aminoácidos

(B) TIPO: AMINOÁCIDO

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: PROTEÍNA

15

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

20

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: 1..37

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

25

(D) OTRA INFORMACIÓN: puntuación 5,9 sec 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 PARA EL SEQ ID NO: 23:

5

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 496 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

10

(C) CATENARIEDAD: DOBLE

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: ADNc

15

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

(F) TIPO DE TEJIDO: Próstata cancerosa

20

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 149..331

(C) MÉTODO DE IDENTIFICACIÓN: blastn

25

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

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: otro

(B) LOCALIZACIÓN: 328..485

(C) MÉTODO DE IDENTIFICACIÓN: blastn

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

30

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: otro
 (B) LOCALIZACIÓN: complemento (182..496)
 (C) MÉTODO DE IDENTIFICACIÓN: blastn
 (D) OTRA INFORMACIÓN: coincidencia 97 región 14..328 id AA399680 est

5

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: péptido señal
 (B) LOCALIZACIÓN: 196..240
 (C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne
 (D) OTRA INFORMACIÓN: puntuación 5,5 sec ILSTVTALTFAXA/LD

10

(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 TGAATGAAK CCGGGATCAA ATGGTTGCTG ATCARAGCCC ATATTTAAAT	434
Ser Ser Ala	
TGGAAAAGTC AAATTGASCA TTATTAATA AAGCTTGTTT AATATGTCTC AAACAAAAAA	494
AA	496

15 (2) INFORMACIÓN PARA EL SEQ ID NO: 24:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 15 aminoácidos
 (B) TIPO: AMINOÁCIDO
 (D) TOPOLOGÍA: LINEAL

20

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

5 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: 1..15

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

10 (D) OTRA INFORMACIÓN: puntuación 5,5 sec 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

15 (2) INFORMACIÓN PARA EL SEQ ID NO: 25:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 623 pares de bases

20 (B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: DOBLE

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: ADNc

25

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

(F) TIPO DE TEJIDO: Testículo

30

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: 49..96

35 (C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

(D) OTRA INFORMACIÓN: puntuación 10,1 sec 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 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 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 CTTCACCCCTC TTGGARACAA	594
TAAACTCTCA TGCCCCCAA AAAAAAAAAA	623

(2) INFORMACIÓN PARA EL SEQ ID NO: 26:

5

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 16 aminoácidos

(B) TIPO: AMINOÁCIDO

10

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

15

(A) ORGANISMO: Homo Sapiens

(ix) CARACTERÍSTICAS:

- 5 (A) NOMBRE/CLAVE: péptido señal
(B) LOCALIZACIÓN: 1..16
(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne
(D) OTRA INFORMACIÓN: puntuación 10,1 sec 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

10

(2) INFORMACIÓN PARA EL SEQ ID NO: 27:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 15 (A) LONGITUD: 848 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: DOBLE
(D) TOPOLOGÍA: LINEAL (ii) TIPO DE MOLÉCULA: ADNc

20

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: Homo Sapiens
(D) ETAPA DEL DESARROLLO: Fetal
(F) TIPO DE TEJIDO: riñón

25

(ix) CARACTERÍSTICAS:

- 30 (A) NOMBRE/CLAVE: péptido señal
(B) LOCALIZACIÓN: 32..73
(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne
(D) OTRA INFORMACIÓN: puntuación 10,7 sec LWLLFFLVTAIHA/EL

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

AACTTTGCCT TGTGTTTCC 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 TGAAGGCTGA GCATCCTGAA ATACCAAGAG CAGATCATAT	787
WTTTTGTTTC ACCATTCTTC TTTTGAATA AATTTTGAAT GTGCTTGAAA AAAAAAAAAA	847
C	848

5

(2) INFORMACIÓN PARA EL SEQ ID NO: 28:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 14 aminoácidos
- (B) TIPO: AMINOÁCIDO
- (D) TOPOLOGÍA: LINEAL

5 (ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: Homo Sapiens

10

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: péptido señal
- (B) LOCALIZACIÓN: 1..14
- (C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne
- (D) OTRA INFORMACIÓN: puntuación 10,7 sec LWLLFFLVTAIHA/EL

15

(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

20 (2) INFORMACIÓN PARA EL SEQ ID NO: 29:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 25 pares de bases
- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: SENCILLA
- (D) TOPOLOGÍA: LINEAL

25

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

30

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 29:
GGGAAGATGG AGATAGTATT GCCTG
25

35 (2) INFORMACIÓN PARA EL SEQ ID NO: 30:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 5 (A) LONGITUD: 26 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: SENCILLA
(D) TOPOLOGÍA: LINEAL
- (ii) TIPO DE MOLÉCULA: Otro ácido nucleico
- 10 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 30:
CTGCCATGTA CATGATAGAG AGATTC
26
- (2) INFORMACIÓN PARA EL SEQ ID NO: 31:
- 15 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 546 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: DOBLE
20 (D) TOPOLOGÍA: LINEAL
- (ii) TIPO DE MOLÉCULA: ADN genómico
- (ix) CARACTERÍSTICAS:
- 25 (A) NOMBRE/CLAVE: promotor
(B) LOCALIZACIÓN: 1..517
- (ix) CARACTERÍSTICAS:
- 30 (A) NOMBRE/CLAVE: sitio de inicio de la transcripción
(B) LOCALIZACIÓN: 518
- (ix) CARACTERÍSTICAS:
- 35 (A) NOMBRE/CLAVE: Sitio de unión TF
(B) LOCALIZACIÓN: 17..25
(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS::

5

(A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

20 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 94..104

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

30 (A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

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

40 (C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 170..178

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 176..189

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

20 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 180..190

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

30 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 284..299

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre TALIALPHAE47_01 puntuación 0,973 secuencia
CATAACAGATGGTAAG

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 284..299

40 (C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 284..299

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

10

(D) OTRA INFORMACIÓN: nombre TAL1BETAIF2_01 puntuación 0,978 secuencia
CATAACAGATGGTAAG

(ix) CARACTERÍSTICAS:

15

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (287..296)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre MYOD_Q6 puntuación 0,954 secuencia
ACCATCTGTT

20

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (302..314)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

25

(D) OTRA INFORMACIÓN: nombre GATA1_04 puntuación 0,953 secuencia
TCAAGATAAAGTA

(ix) CARACTERÍSTICAS:

30

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 393..405

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre IK1_01 puntuación 0,963 secuencia
AGTTGGGAATTCC

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 393..404

40

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre IK2_01 puntuación 0,985 secuencia
AGTTGGGAATTC

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 396..405

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

10 TGGGAATTCC (D) OTRA INFORMACIÓN: nombre CREL_01 puntuación 0,962 secuencia

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 423..436

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre GATA1_02 puntuación 0,950 secuencia
TCAGTGATATGGCA

20

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (478..489)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

30

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 486..493

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

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

40

(C) MÉTODO DE IDENTIFICACIÓN: predicción 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 TGTCAATCAA ATCTTCTATG   60
TCTTGATTTG CCTGCTAATT CTATTATTTC TGGAACATAA TTAGTTTGAT GGTTCCTATTA  120
GTTATTGACT GAGGTGTGCT AATCTCCCAT TATGTGGATT TATCTATTTC TTCAGTTGTA  180
GATAGGACAT TGATAGATAC ATAAGTACCA GGACAAAAGC AGGGAGATCT TTTTCCAAA  240
ATCAGGAGAA AAAAATGACA TCTGAAAAC CTATAGGGAA AGGCATAACA GATGGTAAGG  300
ATACTTTATC TTGAGTAGGA GAGCCTTCCT GTGGCAACGT GGAGAAGGGA AGAGGTCGTA  360
GAATTGAGGA GTCAGCTCAG TTAGAAGCAG GGAGTTGGGA ATTCCGTTCA TGTGATTAG  420
CATCAGTGAT ATGGCAAATG TGGGACTAAG GGTAGTGATC AGAGGGTTAA AATTGTGTGT  480
TTTGTTTTAG CGCTGCTGGG GCATCGCCTT GGGTCCCCTC AACAGATTC CCATGAATCT  540
CTTCAT                                     546
```

5

(2) INFORMACIÓN PARA EL SEQ ID NO: 32:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

10

- (A) LONGITUD: 23 pares de bases
- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: SENCILLA
- (D) TOPOLOGÍA: LINEAL

15

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

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

23

20

(2) INFORMACIÓN PARA EL SEQ ID NO: 33:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

25

- (A) LONGITUD: 24 pares de bases
- (B) TIPO: ÁCIDO NUCLEICO
- (C) CATENARIEDAD: SENCILLA
- (D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

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

24

5 (2) INFORMACIÓN PARA EL SEQ ID NO: 34:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 10 (A) LONGITUD: 861 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: DOBLE
(D) TOPOLOGÍA: LINEAL

15 (ii) TIPO DE MOLÉCULA: ADN genómico

(ix) CARACTERÍSTICAS:

- 20 (A) NOMBRE/CLAVE: promotor
(B) LOCALIZACIÓN: 1..806

(ix) CARACTERÍSTICAS:

- 25 (A) NOMBRE/CLAVE: sitio de inicio de la transcripción
(B) LOCALIZACIÓN: 807

(ix) CARACTERÍSTICAS:

- 30 (A) NOMBRE/CLAVE: Sitio de unión TF
(B) LOCALIZACIÓN: complemento (60..70)
(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector
(D) OTRA INFORMACIÓN: nombre NFY_Q6 puntuación 0,956 secuencia
GGACCAATCAT

35 (ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: Sitio de unión TF
(B) LOCALIZACIÓN: 70..77
(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre MZF1_01 puntuación 0,962 secuencia
CCTGGGGA

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 124..132

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

10 (D) OTRA INFORMACIÓN: nombre CMYB_01 puntuación 0,994 secuencia
TGACCGTTG

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

20 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 135..143

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

30 (A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

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

40 (C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 357..368

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 384..391

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

20 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

30 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 592..599

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

GAAGGGGA (D) OTRA INFORMACIÓN: nombre MZF1_01 puntuación 0,960 secuencia

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 618..627

40 (C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre MYOD_Q6 puntuación 0,981 secuencia
AGCATCTGCC

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 632..642

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

10 (D) OTRA INFORMACIÓN: nombre DELTAEF1_01 puntuación 0,958 secuencia
TCCCACCTTCC

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (813..823)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre S8_01 puntuación 0,992 secuencia
GAGGCAATTAT

20 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (824..831)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

25 (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 TTCCCCAAC CCTGGCTTSC YCTTGGYMAM AGGGCGKTTY TGGGMACTTR 300
AAYTCAGGGC CCAASCAGAA SCACAGGCCC AKTCNTGGCT SMAAGCACAA TAGCCTGAAT 360
GGGATTTTCAG GTTAGNCAGG GTGAGAGGGG AGGCTCTCTG GCTTAGTTTT GTTTTGTITT 420
CCAAATCAAG GTAACCTGCT CCCTTCTGCT ACGGGCCTTG GTCTTGGCTT GTCCTCACCC 480
AGTCGGAACCT CCCTACCACT TTCAGGAGAG TGTTTTTAGG CCCGTGGGGC TGTCTGTTC 540
CAAGCAGTGT GAGAACATGG CTGGTAGAGG CTCTAGCTGT GTGCGGGGCC TGAAGGGGAG 600
TGGGTTCTCG CCCAAGAGC 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 PARA EL SEQ ID NO: 35:

5 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 20 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

10 (D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

15 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 35:
CTGGGATGGA AGGCACGGTA
20

(2) INFORMACIÓN PARA EL SEQ ID NO: 36:

20 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 20 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: SENCILLA

25 (D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: Otro ácido nucleico

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 36:
CAGACCACAC AGCTAGACAA
20

5 (2) INFORMACIÓN PARA EL SEQ ID NO: 37:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 10 (A) LONGITUD: 555 pares de bases
(B) TIPO: ÁCIDO NUCLEICO
(C) CATENARIEDAD: DOBLE
(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: ADN genómico

15 (ix) CARACTERÍSTICAS:

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

20 ix) CARACTERÍSTICAS:

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

25 (ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: Sitio de unión TF
(B) LOCALIZACIÓN: 191..206
(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector
30 (D) OTRA INFORMACIÓN: nombre ARNT_01 puntuación 0,964 secuencia
GGACTCACGTGCTGCT

(ix) CARACTERÍSTICAS:

- 35 (A) NOMBRE/CLAVE: Sitio de unión TF
(B) LOCALIZACIÓN: 193..204
(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre NMYC_01 puntuación 0,965 secuencia
ACTCACGTGCTG

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 193..204

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

10 (D) OTRA INFORMACIÓN: nombre USF_01 puntuación 0,985 secuencia
ACTCACGTGCTG

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (193..204)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre USF_01 puntuación 0,985 secuencia
CAGCACGTGAGT

20

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (193..204)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

25 (D) OTRA INFORMACIÓN: nombre NMYC_01 puntuación 0,956 secuencia
CAGCACGTGAGT

(ix) CARACTERÍSTICAS:

30

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: complemento (193..204)

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 195..202

40

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

GCACGTGA

10

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

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

15

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

CATGGGGA

(ix) CARACTERÍSTICAS:

20

(A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 397..410

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

25

CTCTCCGGAAGCCT

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

30

(B) LOCALIZACIÓN: 400..409

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

(D) OTRA INFORMACIÓN: nombre CETS1PS4_01 puntuación 0,974 secuencia

TCCGGAAGCC

35

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: Sitio de unión TF

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

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

(ix) CARACTERÍSTICAS:

15 (A) NOMBRE/CLAVE: Sitio de unión TF

(B) LOCALIZACIÓN: 547..555

(C) MÉTODO DE IDENTIFICACIÓN: predicción MatInspector

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

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

```
CTATAGGGCA CGCKTGGTCG ACGGCCCGGG CTGGTCTGGT CTGKGTGGA 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 TTCTCATAG 360
GTTGCTCTGC CCATGGTCCC ACTGCAGACC CAGGCACTCT CCGGAAGCCT GGAAATCACC 420
CGTGTCTTCT GCCTGCTCCC GCTCACATCC CACACTTGTG TTCAGTCACT GAGTTACAGA 480
TTTGGCTCC TCAATTTCTC TTGTCTTAGT CCCATCCTCT GTTCCCTGG CCAGTTTGTG 540
TAGCTGTGTG GTCTC 555
```

(2) INFORMACIÓN PARA EL SEQ ID NO: 73:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

25

(A) LONGITUD: 315 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: DOBLE

(D) TOPOLOGÍA: LINEAL (ii) TIPO DE MOLÉCULA: ADNc

30

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

(F) TIPO DE TEJIDO: Bazo

5

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: 151..228

10

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

(D) OTRA INFORMACIÓN: puntuación 6,9 sec LWVLLLCaHVVT/LV

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 73:

AACACGCAGC TAGACACAGC TAMCTTGAGT CTTGGAGCTC CTAGAGGGAM GCTTCTGGAM	60
AGGAAGGCTC TTCAGGACCT CTTAGGAGCC AGAGMMSMGG ACGTKSACAC AGATAAAGAG	120
CCAGGCTCAC CAGCTCCTGA CGCATGCAKS ATG ACC ATG AGA CAC AAC TGG ACA	174
Met Thr Met Arg His Asn Trp Thr	
CCA GAC CTC AGC CCT TTG TGG GTC CTG CTC CTG TGT GCC CAC GTC GTC	222
Pro Asp Leu Ser Pro Leu Trp Val Leu Leu Leu Cys Ala His Val Val	
-15 -10 -5	
ACT CTC CTG GTC AGA GCC ACA CCT GTC TCG CAG ACC ACC ACA GCT GCC	270
Thr Leu Leu Val Arg Ala Thr Pro Val Ser Gln Thr Thr Thr Ala Ala	
1 5 10	
ACT GCC TCA GTT AGA AGC ACA AAG GAC CCC TGC CCC TCC CAG CGG	315
Thr Ala Ser Val Arg Ser Thr Lys Asp Pro Cys Pro Ser Gln Arg	
15 20 25	

15

(2) INFORMACIÓN PARA EL SEQ ID NO: 75:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

20

(A) LONGITUD: 224 pares de bases

(B) TIPO: ÁCIDO NUCLEICO

(C) CATENARIEDAD: DOBLE

(D) TOPOLOGÍA: LINEAL

25

(ii) TIPO DE MOLÉCULA: ADNc

(vi) FUENTE ORIGINAL:

30

(A) ORGANISMO: Homo Sapiens

(F) TIPO DE TEJIDO: Ovario

(ix) CARACTERÍSTICAS:

5

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: 114..191

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

(D) OTRA INFORMACIÓN: puntuación 6,9 sec LWVLLLCaHVVT/LV

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 75:

```

ACTTCCCAGA AGCAGCTCTG GTGCTGAAGA GAGCACTGCC TCCCTGTGTG ACTGGAGAAG 60
AGGACGTTGT CACAGATAAA GAGCCAGGCT CACCAGCTCC TGACGCATGC ATC ATG 116
Met
ACC ATG AGA CAC AAC TGG ACA CCA GAC CTC AGC CCT TTG TGG GTC CTG 164
Thr Met Arg His Asn Trp Thr Pro Asp Leu Ser Pro Leu Trp Val Leu
-25 -20 -15 -10
CTC CTG TGT GCC CAC GTC GTC ACT CTC CTG GTC AGA GCC ACA CCT GTC 212
Leu Leu Cys Ala His Val Val Thr Leu Leu Val Arg Ala Thr Pro Val
-5 1 5
TCG CAG CCC ACG 224
Ser Gln Pro Thr
10

```

10

(2) INFORMACIÓN PARA EL SEQ ID NO: 306:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

15

(A) LONGITUD: 55 aminoácidos

(B) TIPO: AMINOÁCIDO

(D) TOPOLOGÍA: LINEAL

(ii) TIPO DE MOLÉCULA: PROTEÍNA

20

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

(F) TIPO DE TEJIDO: Bazo

25

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: -26..-1

30

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

(D) OTRA INFORMACIÓN: puntuación 6,9 sec LWVLLLLCAHWTL/LV

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 306:

5
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 Thr Thr Ala Ala Thr Ala Ser Val Arg Ser Thr Lys
10 15 20
Asp Pro Cys Pro Ser Gln Arg
25

(2) INFORMACIÓN PARA EL SEQ ID NO: 308:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

10

(A) LONGITUD: 37 aminoácidos

(B) TIPO: AMINOÁCIDO

(D) TOPOLOGÍA: LINEAL

15

(ii) TIPO DE MOLÉCULA: PROTEÍNA

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: Homo Sapiens

20

(F) TIPO DE TEJIDO: Ovario

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: péptido señal

(B) LOCALIZACIÓN: -26...-1

25

(C) MÉTODO DE IDENTIFICACIÓN: Matriz de Von Heijne

(D) OTRA INFORMACIÓN: puntuación 6,9 sec LWVLLLLCAHVVT/LV

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 308:

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 Pro Thr
10

30

REIVINDICACIONES

1. Un péptido señal que tiene la secuencia de aminoácidos -26 a -1 del SEQ ID NO: 306.
2. Una secuencia señal aislada que codifica el péptido señal de la reivindicación 1.
- 5 3. La secuencia señal de la reivindicación 2, que tiene la secuencia de nucleótidos 151 a 228 del SEQ ID NO: 73.
4. Un ácido nucleico purificado y aislado que codifica un polipéptido que comprende el péptido señal según la reivindicación 1 en el extremo 5' de la secuencia codificante.
5. Un ácido nucleico según la reivindicación 4, en el que el ácido nucleico comprende una secuencia señal que tiene la secuencia de nucleótidos 151 a 228 del SEQ ID NO: 73.
- 10 6. El ácido nucleico según la reivindicación 4, en el que el extremo amino de dicho polipéptido está codificado por los nucleótidos 151 a 315 del SEQ ID NO: 73.
7. El ácido nucleico según la reivindicación 4, en el que el extremo amino de dicho polipéptido está codificado por los nucleótidos 114 a 224 del SEQ ID NO: 75.
- 15 8. El ácido nucleico según la reivindicación 4, en el que dicha secuencia señal está fusionada en marco al extremo 5' de una secuencia que codifica un polipéptido que es heterólogo a un polipéptido que comprende los aminoácidos 1 a 29 del SEQ ID NO: 306.
9. Un vector de expresión que comprende la secuencia señal según la reivindicación 2 ó 3 conectado operablemente a un promotor.
- 20 10. Un vector de expresión que comprende el ácido nucleico de uno cualquiera de los ácidos nucleicos según las reivindicaciones 4 a 8 conectado operablemente a un promotor
11. El vector de expresión según la reivindicación 9 o 10, en el que dicho vector es un vector de secreción.
12. El vector de expresión según la reivindicación 9 o 10, en el que dicho vector es un vector de terapia génica.
- 25 13. Un polipéptido codificado por un polinucleótido de una cualquiera según las reivindicaciones 4 a 8.
14. El polipéptido según la reivindicación 13, en el que dicho polipéptido es una proteína secretada humana.
15. El polipéptido según la reivindicación 13 o la reivindicación 14, en el que dicho polipéptido comprende los aminoácidos 1 a 29 del SEQ ID NO: 306.
- 30 16. El polipéptido según la reivindicación 13 o la reivindicación 14, en el que dicho polipéptido comprende los aminoácidos 1 a 11 del SEQ ID NO: 308.
17. Una proteína de fusión codificada por el ácido nucleico según la reivindicación 8.
18. El uso in vitro del péptido señal según la reivindicación 1, para dirigir la secreción extracelular de un polipéptido.
- 35 19. El uso del péptido señal según la reivindicación 1, para simplificar la purificación de proteínas de un polipéptido deseado.
20. El uso in vitro del vector según una cualquiera de las reivindicaciones 9 a 12, para dirigir la secreción extracelular de un polipéptido.
21. El uso del vector según una cualquiera de las reivindicaciones 9 a 12, para simplificar la purificación de proteínas de un polipéptido deseado.
- 40 22. El uso de un vector según una cualquiera de las reivindicaciones 18 a 21, en el que dicho polipéptido es el polipéptido de una cualquiera de las reivindicaciones 13 a 17.
23. Un método para producir una proteína secretada, que comprende la etapa de insertar un gen que codifica una proteína no secretada en marco con la secuencia señal de la reivindicación 2 ó 3 en un vector de modo que la proteína codificada por el gen insertado es expresada a partir de del ARNm transcrito.
- 45 24. El método de la reivindicación 23, que comprende, además, la etapa de introducir dicho vector en una célula hospedante, tejido no humano u organismo no humano.

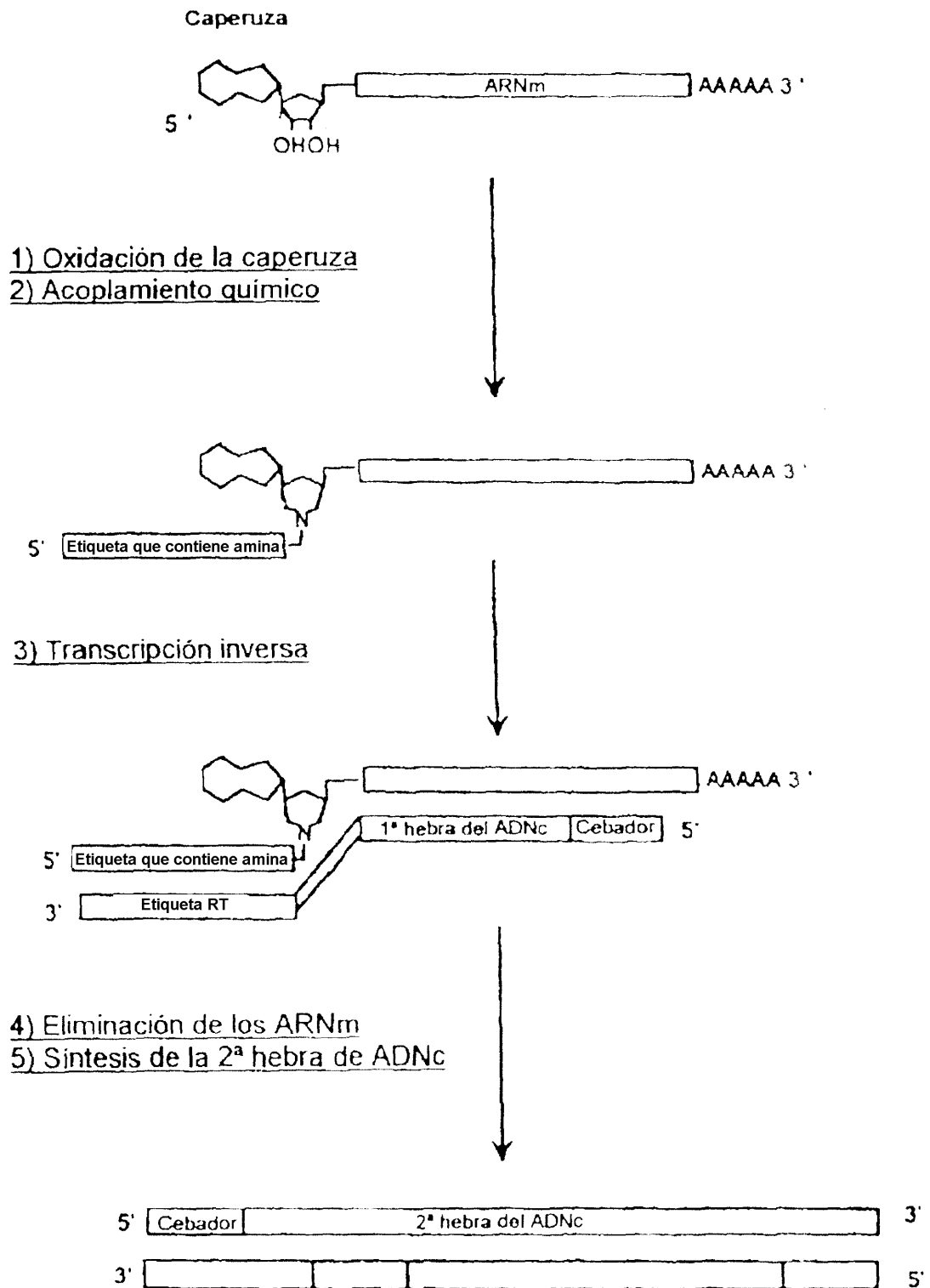


Figura 1

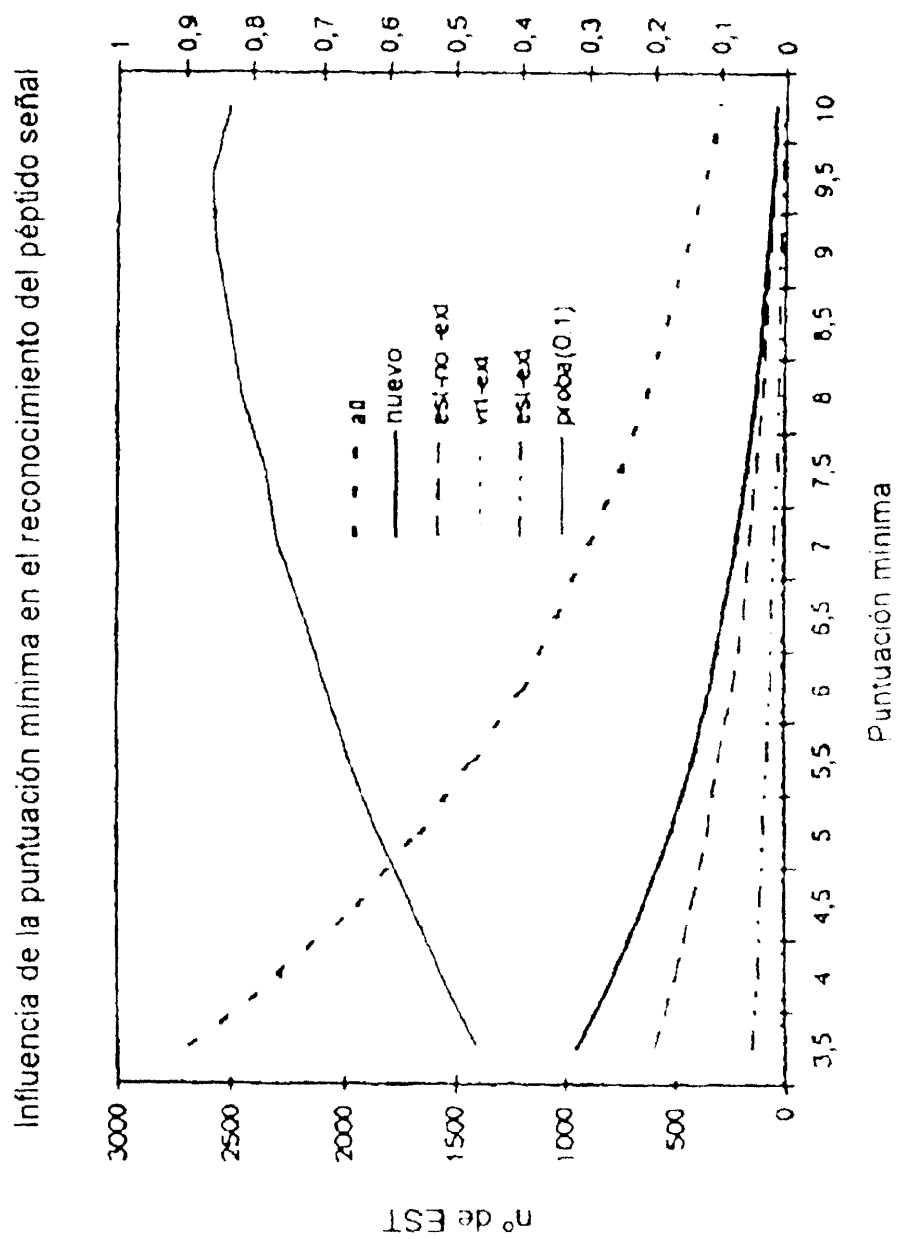


Figura 2

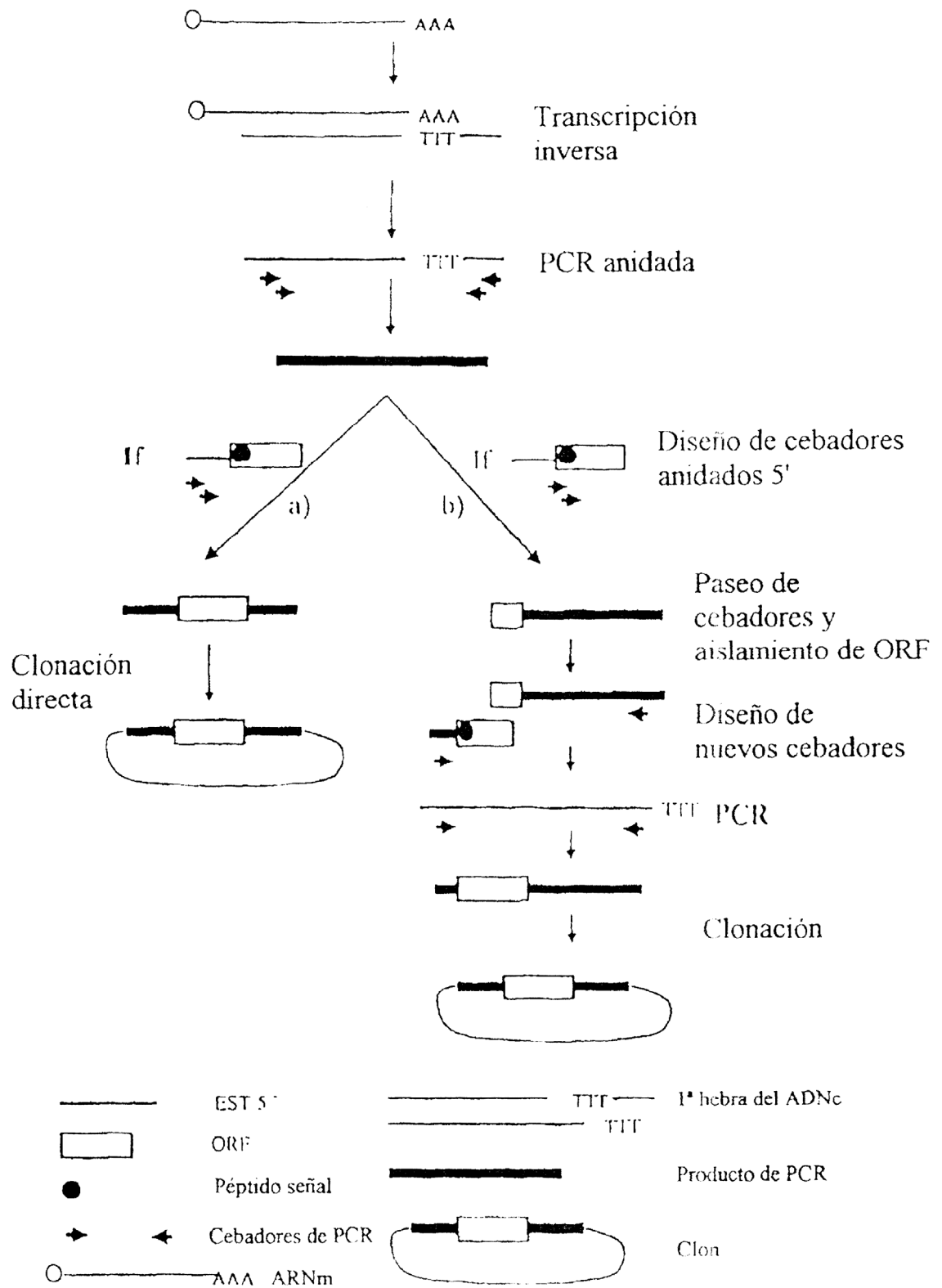
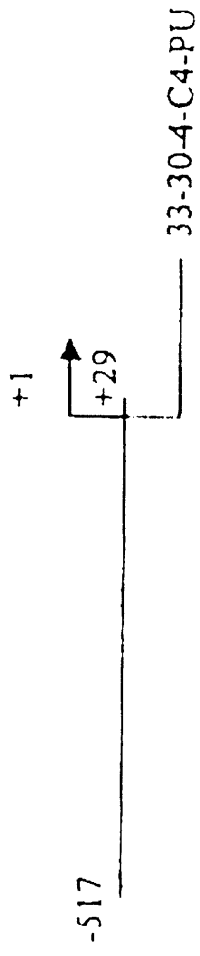
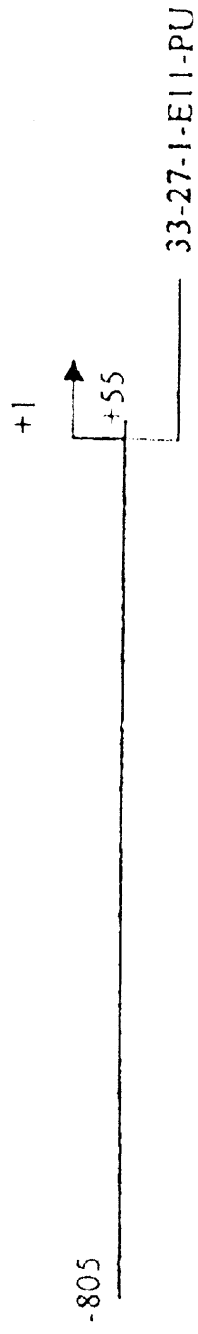


Figura 3

Promotor P13H2



Promotor P15B4



Promotor P29B6

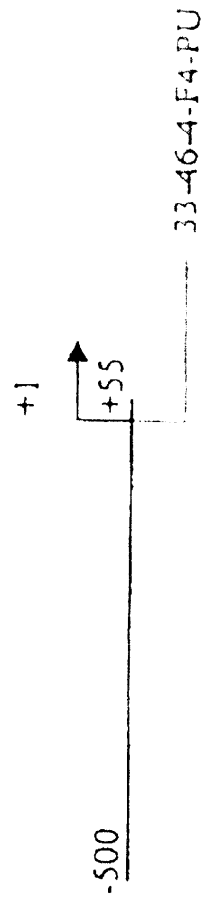


Figura 4