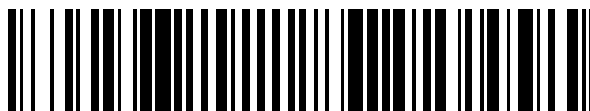


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 402 095**

51 Int. Cl.:

C12Q 1/68 (2006.01)

C12N 15/11 (2006.01)

G01N 33/574 (2006.01)

A61P 35/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.06.1999 E 04015121 (9)**

97 Fecha y número de publicación de la concesión europea: **27.02.2013 EP 1466988**

54 Título: **Genes y productos de expresión génica que son regulados en forma diferencial en el cáncer de próstata**

30 Prioridad:

11.06.1998 US 88877 P
09.06.1999 US 328475

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
26.04.2013

73 Titular/es:

NOVARTIS VACCINES AND DIAGNOSTICS, INC.
(100.0%)
4560 HORTON STREET
EMERYVILLE, CA 94608, US

72 Inventor/es:

ASTEL, JON H.;
CARROLL, EDDIE, III.;
ENDEGE, WILSON O.;
FORD, DONNA M.;
MONAHAN, JOHN E.;
SCHLEGEL, ROBERT;
STEINMANN, KATHLEEN E. y
ZHANG, JIMMY

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 402 095 T3

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

DESCRIPCIÓN

Genes y productos de expresión génica que son regulados en forma diferencial en el cáncer de próstata

Campo de la invención

- 5 Esta invención se refiere al campo del diagnóstico y pronóstico del cáncer, la progresión tumoral, crecimiento de las células hiperproliferativas, y a las manifestaciones físicas y biológicas acompañantes. Más específicamente, la invención incluye polinucleótidos que son regulados diferencialmente en trastornos prostáticos, tales como cáncer de próstata metastático, cáncer de próstata localizado, y la hiperplasia prostática benigna (BPH).

Antecedentes de la invención

- 10 El aumento o la disminución de la expresión de los genes en el cáncer o la progresión del tumor son útiles para fines terapéuticos y de diagnóstico. Por ejemplo, la detección de genes o el aumento de expresión de productos de expresión génica en células hiperproliferativas puede ser un marcador predictivo o de diagnóstico de la aparición o la progresión del cáncer. El diagnóstico precoz puede ser útil si el cáncer, los tumores, o las células hiperproliferativas se puede inhibir, remover o terminar para evitar metástasis o recurrencia del crecimiento canceroso. Tal alerta temprana es de uso particular para pacientes con cáncer de próstata, donde la remoción del crecimiento, el tumor o las células es beneficiosa si la enfermedad se limita a la próstata. Existe una necesidad en el estado del arte por genes relacionados con el cáncer y la progresión del tumor.

Resumen de la invención

- 20 La presente invención proporciona métodos y reactivos para el diagnóstico del cáncer de próstata, la progresión de un tumor de la próstata o el crecimiento de células hiperproliferativas de la próstata. Los reactivos para tales kits de diagnóstico incluyen polinucleótidos que comprenden al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335 o un complemento de los mismos.

La invención proporciona:

- un método para el diagnóstico de cáncer de próstata, el progreso de un tumor de la próstata y del crecimiento de células hiperproliferativas de la próstata que comprende:

- 25 (a) proporcionar una sonda de polinucleótidos que comprenden al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335 o un complemento de los mismos;

(b) poner en contacto una muestra biológica para diagnóstico con dicha sonda bajo condiciones de hibridación lo que permite la formación de un dúplex; y

- 30 (c) determinar la presencia de dicho dúplex, y detectar las diferencias que existen entre los niveles de dúplex en dicha muestra biológica comparado con una muestra biológica normal.

- Un kit de diagnóstico que comprende:

(a) un reactivo de diagnóstico que comprende una sonda de polinucleótidos que contiene al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335 o un complemento de los mismos cuando dicha secuencia está presente en una muestra biológica de ensayo;

- 35 (b) una muestra biológica normal; y

(c) instrucciones para detectar las diferencias que existen entre los niveles de dúplex en dicha muestra biológica de ensayo en comparación con dicha muestra biológica normal.

- Un polinucleótido aislado seleccionado de entre el grupo que consiste de: un polinucleótido que comprende la secuencia de nucleótidos de las SEQ ID NOs: 218 o 219.

- 40 - Un vector comprende el polinucleótido anterior.

- Una célula huésped aislada que comprende al vector anterior.

Los métodos de diagnóstico descritos aquí incluyen tanto ensayos como inmunoensayos de ácido nucleico.

Preferiblemente, el ácido nucleico obtenido a partir de material biológico es ARNm o ADN genómico. El ácido nucleico puede ser también ADNc complementario al ARNm.

También se describe el uso de los polinucleótidos aislados o partes de los mismos como sondas de diagnóstico o como iniciadores.

5 En ciertas realizaciones preferidas, el polinucleótido está operativamente enlazado a una secuencia de control de la expresión. La invención proporciona además una célula huésped, incluidas células bacterianas, de levadura, de mamífero y de insecto, transformadas con la secuencia de polinucleótido. La invención también proporciona el ADNc de longitud completa y el gen humano de longitud completa correspondiente al polinucleótido.

10 También se describe ácido nucleico obtenido por medio de la transformación de una célula huésped con ácido nucleico que comprende al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335, el cultivo de la célula huésped, y la recuperación del ácido nucleico replicado, el ARN expresado, y/o el polipéptido expresado.

Descripción detallada de la invención

15 Los genes que aumentan o disminuyen su expresión en el cáncer o en la progresión del tumor son útiles para fines terapéuticos y de diagnóstico. Por ejemplo, un ensayo de diagnóstico para determinar el estadio de la enfermedad también es útil en la adaptación del tratamiento de un cáncer agresivo frente a un cáncer más suave o progresión tumoral. Las secuencias de polinucleótidos y polipéptidos codificados de la presente invención son útiles para estos propósitos de diagnóstico o de pronóstico.

20 Además, la modulación de genes o de los productos de expresión génica que son mal regulados pueden utilizarse para tratar o mejorar el cáncer, la progresión tumoral, el crecimiento de las células hiperproliferativas, y las manifestaciones físicas y biológicas acompañantes.

En este documento se identifican secuencias de polinucleótidos que aumentan su expresión en una línea celular cancerosa, más específicamente en una línea celular de cáncer de próstata. La presente invención se refiere a métodos y reactivos para diagnóstico.

25 I. Uso de polinucleótidos que tiene una secuencia de una o más de las SEQ ID NOs: 1 - 339 para obtener ADNc de longitud completa y el gen humano de longitud completa y la región promotora.

30 Las moléculas de ADNc de longitud completa que comprenden las secuencias divulgadas se obtienen como sigue. El polinucleótido o una porción del mismo que comprende al menos 12, 15, 18, o 20 nucleótidos se utiliza como sonda de hibridación para detectar los miembros de hibridación de una biblioteca de ADNc usando métodos de diseño de la sonda, métodos de clonación, y técnicas de selección de clones como se describe en la patente de los Estados Unidos No. 5.654.173, "Secreted Proteins and Polynucleotides Encoding Them", incorporada aquí por referencia. Las bibliotecas de ADNc se elaboran a partir de tejidos seleccionados, tales como tejidos normales o tumorales, o de tejidos de un mamífero tratado, por ejemplo, con un agente farmacéutico. Preferiblemente, el tejido es el mismo que el utilizado para generar los polinucleótidos, ya que tanto los polinucleótidos como el ADNc representan genes expresados. Más preferiblemente, la biblioteca de ADNc se elabora a partir del material biológico descrito aquí en los Ejemplos. Alternativamente, muchas bibliotecas de ADNc se encuentran comercialmente disponibles. (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2ª Ed. (Cold Spring Harbor Press, Cold Spring Harbor, NY 1989).

40 Se obtienen los miembros de la biblioteca que son más grandes que el polinucleótido, y preferiblemente que contienen la secuencia completa del mensaje nativo. Con el fin de confirmar que se ha obtenido el ADNc entero, se llevan a cabo experimentos de protección del ARN de la siguiente manera. La hibridación de un ADNc de longitud completa a un ARNm protegerá al ARN de la degradación con ARNasa. Si el ADNc no es de longitud completa, entonces las porciones del ARNm que no se hibridan estarán sometidas a degradación con ARNasa. Este se somete a ensayo, tal como se conoce en la técnica, por medio de los cambios en la movilidad electroforética sobre geles de poliacrilamida, o mediante la detección de monorribonucleótidos liberados. Sambrook et al., Molecular Cloning: A Laboratory Manual, 2ª Ed. (Cold Spring Harbor Press, Cold Spring Harbor, NY 1989). Con el fin de obtener secuencias 5' adicionales al extremo de un ADNc parcial, se lleva a cabo un 5' RACE (PCR Protocols: A Guide to Methods and Applications (Academic Press, Inc. 1990)).

50 El ADN genómico se aísla utilizando polinucleótidos de una manera similar al aislamiento de los ADNc de longitud completa. En resumen, los polinucleótidos, o porciones de los mismos, se utilizan como sondas para las bibliotecas de ADN genómico. Preferiblemente, la biblioteca se obtiene del tipo de célula que se usó para generar los polinucleótidos, pero esto no es esencial. Más preferiblemente, el ADN genómico se obtiene a partir del material biológico descrito en la presente memoria en los Ejemplos. Tales bibliotecas pueden estar en vectores adecuados

para transportar grandes segmentos de un genoma, tales como P1 o YAC, como se describe en detalle en Sambrook et al., 9.4 - 9.30. Además, las secuencias genómicas se pueden aislar a partir de bibliotecas de BAC humanos, que se encuentran comercialmente disponibles por ejemplo de Research Genetics, Inc., Huntsville, Alabama, EE.UU.. Con el fin de obtener secuencias adicionales 5' o 3', se lleva a cabo un desplazamiento sobre el cromosómico, como se describe en Sambrook et al., de tal manera que se aíslan fragmentos adyacentes y superpuestos de ADN genómico. Éstos se mapean y reconstruyen, tal como se conoce en la técnica, usando enzimas de restricción de digestión y la ADN ligasa.

La utilización de las secuencias de polinucleótidos de la invención, correspondientes a genes de longitud completa pueden aislarse utilizando tanto métodos clásicos como PCR para construir y sondear bibliotecas de ADNc. Usando cualquier método, se llevan a cabo preferiblemente transferencias tipo Northern en una cantidad de tipos de células para determinar qué líneas celulares expresan el gen de interés a la tasa más alta.

Los métodos clásicos de la construcción de bibliotecas de ADNc se enseñan en Sambrook et al., véase más arriba. Con estos métodos, se puede producir ADNc a partir de ARNm e insertarlo en vectores virales o de expresión. Típicamente, las bibliotecas de ARNm que comprenden colas de poli (A) se pueden producir con iniciadores de poli (T). Del mismo modo, las bibliotecas de ADNc se pueden producir usando las presentes secuencias como iniciadores.

Se usan métodos de PCR para amplificar los miembros de una biblioteca de ADNc que contienen el inserto deseado. En este caso, el inserto deseado contendrá la secuencia del ADNc de longitud completa que corresponde a las presentes EST. Tales métodos PCR incluyen captura de genes y métodos RACE. Gruber et al., PCT WO 95/04745 y Gruber et al., patente de los Estados Unidos No. 5.500.356. Los kits se encuentran comercialmente disponibles para llevar a cabo experimentos de captura de genes, por ejemplo, Life Technologies, Gaithersburg, Maryland, EE.UU.. PCT., publicación No. WO 97/19110. (Apte y Siebert, Biotechniques 15: 890 - 893, 1993; Edwards et al., Nuc. Acids Res. 19: 5227 - 5232, 1991).

La región promotora de un gen generalmente se encuentra 5' del sitio de iniciación para la ARN polimerasa II, y se puede obtener mediante la realización de 5' RACE utilizando un iniciador de la región codificante del gen. Alternativamente, se puede utilizar el ADNc como una sonda para la secuencia genómica, y la región 5' de la región codificante se identifica por medio de "desplazamiento hacia arriba". Si el gen está altamente expresado o expresado diferencialmente, el promotor del gen puede ser de uso en un constructo regulador para un gen heterólogo.

Una vez se obtiene el gen o el ADNc de longitud completa, se puede preparar ADN que codifica variantes mediante mutagénesis dirigida al sitio, descrita en detalle en Sambrook et al., 15.3 - 15.63. La escogencia del codón o del nucleótido para ser reemplazado se puede basar en la divulgación que se hace aquí sobre cambios opcionales en los aminoácidos para conseguir la estructura y/o función alterada de la proteína.

Como un método alternativo para obtener ADN o ARN a partir de un material biológico, se puede sintetizar un ácido nucleico que contiene los nucleótidos que tienen la secuencia de uno o más polinucleótidos de la invención. Por lo tanto, la invención abarca moléculas de ácido nucleico que varían en longitud desde 20 nucleótidos (que corresponden a al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335) hasta una longitud máxima adecuada para una o más manipulaciones biológicas, incluyendo la replicación y expresión, de la molécula de ácido nucleico. La invención incluye pero no se limita a (a) ácido nucleico que tiene el tamaño de un gen completo, y que comprende al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335; (b) el ácido nucleico de (a) también comprende al menos un gen adicional, operativamente enlazado para permitir la expresión de una proteína de fusión; (c) un vector de expresión que comprende (a) o (b); (d) un plásmido que comprende (a) o (b); y (e) una partícula viral recombinante que comprende (a) o (b).

La secuencia de un ácido nucleico que comprende al menos 20 nucleótidos contiguos de al menos una cualquiera de las SEQ ID NOs: 218, 219 o 335, preferiblemente la secuencia completa de al menos una cualquiera de las SEQ ID NOs: 218, 219 o 335, no está limitada y puede ser cualquier secuencia de A, T, G y / o C (para ADN) y A, U, G y/o C (para ARN) o bases modificadas de la misma, incluyendo la inosina y la pseudouridina. La escogencia de la secuencia dependerá de la función deseada, y puede ser dictada por las regiones codificantes deseadas, las regiones deseadas tipo intrón, y las regiones reguladoras deseadas.

Cuando la secuencia completa de cualquiera de las SEQ ID NO: 218, 219 o 335 está dentro del ácido nucleico, el ácido nucleico obtenido es denominado en este documento como un polinucleótido que comprende la secuencia de una cualquiera de las SEQ ID NOs: 218, 219 o 335.

II. Expresión de un polipéptido codificado por un gen de longitud completa o ADNc de longitud completa

El polinucleótido, el ADNc correspondiente o el gen de longitud completa se utilizan para expresar el producto génico parcial o completo. Las construcciones apropiadas de polinucleótidos se purifican usando técnicas estándar de ADN

recombinante como se describe, por ejemplo, en Sambrook et al., (1989) *Molecular Cloning: A Laboratory Manual*, 2ª ed. (Cold Spring Harbor Press, Cold Spring Harbor, Nueva York). Los polipéptidos codificados por los polinucleótidos se expresan en cualquier sistema de expresión, incluyendo, por ejemplo, sistemas bacterianos, de levadura, de insectos, de anfibios y de mamíferos. Los vectores adecuados y las células huésped se describen en la patente de los Estados Unidos No. 5.654.173.

Bacterias. Los sistemas de expresión en bacterias incluyen a aquellos descritos en Chang et al., *Nature* (1978) 275: 615, Goeddel et al., *Nature* (1979) 281: 544, Goeddel et al., *Nucleic Acids Res.* (1980) 8: 4057; EP 0 036 776, la patente de los Estados Unidos No. 4.551.433, DeBoer et al., *Proc. Natl. Acad. Sci. (EE.UU.)* (1983) 80: 21 - 25, y Siebenlist et al., *Cell* (1980) 20: 269.

Levadura. Los sistemas de expresión en levadura incluyen a aquellos descritos en Hinnen et al., *Proc. Natl. Acad. Sci. (EE.UU.)* (1978) 75: 1929; Ito et al., *J. Bacteriol.* (1983) 153: 163; Kurtz et al., *Mol. Cell. Biol.* (1986) 6: 142; Kunze et al., *J. Basic Microbiol.* (1985) 25: 141; Gleeson et al., *J. Gen. Microbiol.* (1986) 132: 3459, Roggenkamp et al., *Mol. Gen. Genet.* (1986) 202: 302; Das et al., *J. Bacteriol.* (1984) 158: 1165; De Louvencourt et al., *J. Bacteriol.* (1983) 154: 737, Van den Berg et al., *Bio/Technology* (1990) 8: 135; Kunze et al., *J. Basic Microbiol.* (1985) 25: 141; Cregg et al., *Mol. Cell. Biol.* (1985) 5: 3376, las patentes de los Estados Unidos Nos. 4.837.148 y 4.929.555; Beach y Nurse, *Nature* (1981) 300: 706; Davidow et al., *Curr. Genet.* (1985) 10: 380, Gaillardin et al., *Curr. Genet.* (1985) 10: 49, Ballance et al., *Biochem. Biophys. Res. Commun.* (1983) 112: 284 - 289; Tilburn et al., *Gene* (1983) 26: 205 - 221, Yelton et al., *Proc. Natl. Acad. Sci. (EE.UU.)* (1984) 81: 1470 - 1474, Kelly y Hynes, *EMBO J.* (1985) 4: 475 - 479; EP 0 244 234 y WO 91/00357.

Células de insecto. La expresión de genes heterólogos en insectos se logra como se describe en la patente de los Estados Unidos No. 4.745.051, Friesen et al. (1986) "The Regulation of Baculovirus Gen Expression", en: *The Molecular Biology of Baculoviruses* (W. Doerfler, ed), EP 0 127 839, EP 0 155 476, y Vlak et al., *J. Gen. Virol.* (1988) 69: 765 - 776, Miller et al., *Ann. Rev. Microbiol.* (1988) 42: 177, Carbonell et al., *Gene* (1988) 73: 409, Maeda et al., *Nature* (1985) 315: 592 - 594, Lebacqz-Verheyden et al., *Mol. Cell. Biol.* (1988) 8: 3129; Smith et al., *Proc. Natl. Acad. Sci. (EE.UU.)* (1985) 82: 8404, Miyajima et al., *Gene* (1987) 58: 273; y Martin et al., *DNA* (1988) 7: 99. Se describen numerosas cepas y variantes de baculovirus y las correspondientes células huésped permisivas de insectos en Luckow et al., *Bio/Technology* (1988) 6: 47 - 55, Miller et al., *Generic Engineering* (Setlow, J. K. et al. Eds.), Vol. 8 (Plenum Publishing, 1986), páginas 277 - 279, y Maeda et al., *Nature*, (1985) 315: 592 - 594.

Células de mamífero. La expresión en mamíferos se realiza como se describe en Dijkema et al., *EMBO J.* (1985) 4: 761, Gorman et al., *Proc. Natl. Acad. Sci. (EE.UU.)* (1982) 79: 6777, Boshart et al., *Cell* (1985) 41: 521 y la patente de los Estados Unidos No. 4.399.216. Otras características de la expresión en mamíferos son facilitadas como se describe en Ham y Wallace, *Meth. Enz.* (1979) 58: 44, Bames y Sato, *Anal. Biochem.* (1980) 102: 255, las patentes de los Estados Unidos Nos. 4.767.704, 4.657.866, 4.927.762, 4.560.655, WO 90/103430, WO 87/00195, y US. RE 30.985.

Las moléculas de polinucleótido que comprende la secuencia de polinucleótidos se propagan mediante la colocación de la molécula en un vector. Se usan vectores virales y no virales, incluyendo plásmidos. La elección del plásmido dependerá del tipo de célula en la que se desea la propagación y del propósito de la propagación. Ciertos vectores son útiles para amplificar y elaborar grandes cantidades de la secuencia deseada de ADN. Otros vectores son adecuados para expresión en células en cultivo. Incluso otros vectores son adecuados para transferencia y expresión en células en un animal o una persona completa. La elección del vector apropiado está dentro de las habilidades de la técnica. Muchos de tales vectores están disponibles comercialmente. El polinucleótido se inserta en un vector típicamente por medio de unión de ADN ligasa a un sitio de la enzima de restricción escindido en el vector. Alternativamente, la secuencia deseada de nucleótidos puede insertarse mediante recombinación homóloga *in vivo*. Típicamente esto se logra uniendo las regiones de homología al vector sobre los flancos de la secuencia deseada de nucleótidos. Las regiones de homología se añaden por medio de ligación de oligonucleótidos, o mediante reacción en cadena de la polimerasa utilizando iniciadores que comprenden por ejemplo tanto la región de homología como una porción de la secuencia deseada de nucleótidos.

Los polinucleótidos se enlazan a secuencias reguladoras según sea apropiado para obtener las propiedades de expresión deseadas. Éstos pueden incluir promotores (unidos ya sea en el extremo 5' de la cadena codificante o en el extremo 3' de la cadena no codificante), potenciadores, terminadores, operadores, represores, e inductores. Los promotores pueden ser regulados o constitutivos. En algunas situaciones, puede ser deseable utilizar promotores condicionalmente activos, tales como específicos del tejido o específicos de la etapa de desarrollo. Éstos se enlazan a la secuencia deseada de nucleótidos usando las técnicas descritas anteriormente para enlazamiento con vectores. Cualquier técnica conocida en la técnica puede ser utilizada.

Cuando se usa cualquiera de las células huéspedes anteriores, u otras células huésped apropiadas u organismos, para replicar y/o expresar los polinucleótidos o ácidos nucleicos de la invención, el ácido nucleico replicado resultante, ARN, proteína o polipéptido expresado, es un producto de la célula u organismo huésped. El producto se

recupera mediante cualquier medio adecuado conocido en la técnica.

Una vez se identifica el gen correspondiente con el polipéptido, su expresión puede ser regulada en la célula de la que el gen es nativo. Por ejemplo, un gen endógeno de una célula puede ser regulado por una secuencia reguladora exógena, como se divulga en la patente de los Estados Unidos No. 5.641.670, "Protein Production and Protein Delivery".

Identificación de los polipéptidos secretados y unidos a la membrana

Tanto los polipéptidos unidos a la membrana como los secretados son de interés. Por ejemplo, los niveles de polipéptidos secretados pueden ser analizados convenientemente en fluidos corporales, tales como sangre, orina, fluido prostático y semen. Los polipéptidos unidos a la membrana son útiles para construir antígenos de vacunas o inducir una respuesta inmune. Tales antígenos comprenderían toda o parte de la región extracelular de los polipéptidos unidos a la membrana.

Debido a que tanto los polipéptidos unidos a la membrana como los secretados comprenden un fragmento de aminoácidos hidrófobos contiguos, se pueden utilizar algoritmos de predicción de la hidrofobicidad para identificar tales polipéptidos.

Una secuencia de señal es usualmente codificada tanto por los genes para el polipéptido unido a la membrana como el secretado para dirigir un polipéptido a la superficie de la célula. La secuencia de señal comprende habitualmente un tramo de residuos hidrófobos. Dichas secuencias de señal pueden plegarse en estructuras helicoidales.

Los polipéptidos unidos a la membrana comprenden típicamente al menos una región transmembrana que posee un tramo de aminoácidos hidrófobos que pueden atravesar la membrana. Algunas regiones transmembrana también presentan una estructura helicoidal.

Los fragmentos hidrófobos dentro de un polipéptido se pueden identificar mediante el uso de algoritmos informáticos. Tales algoritmos incluyen Hopp y Woods, Proc. Natl. Acad. Sci. EE.UU. 78: 3824 - 3828 (1981); Kyte y Doolittle, J. Mol. Biol. 157: 105 - 132 (1982); y el algoritmo RAOAR, Degli Esposti et al., Eur. J. Biochem. 190: 207 - 219 (1990).

Otro método para identificar polipéptidos unidos a la membrana y secretados consiste en traducir los polinucleótidos descritos aquí, en todos los seis marcos y determinar si al menos están presentes 8 aminoácidos hidrófobos contiguos. Aquellos polipéptidos traducidos con al menos 8, más típicamente, 10; incluso más típicamente, 12 aminoácidos hidrófobos contiguos se considera que son o bien un polipéptido putativo unido a la membrana o secretado. Los aminoácidos hidrófobos incluyen alanina, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, treonina, triptófano, tirosina y valina.

Los polipéptidos putativos unidos a la membrana y/o secretados son codificados por las secuencias de los clones siguientes: SL-5, SL-6, SL-9, SL-11, SL-13, SL-90, SL-100, SL-107, SL-124, SL-135, SL-139, SL-143, SL-152, SL-153, SL-173 y SL-177.

Construcción de polipéptidos y variantes de los mismos

Los polipéptidos descritos aquí incluyen aquellos codificados por los polinucleótidos divulgados. Estos polipéptidos también pueden ser codificados por ácidos nucleicos que, en virtud de la degeneración del código genético, no son idénticos en secuencia a los polinucleótidos divulgados. Por lo tanto, aquí se describen ácidos nucleicos que comprenden polinucleótidos que codifican una proteína o polipéptido expresados por un polinucleótido de la invención. También se describen variantes; las variantes de los polipéptidos incluyen mutantes, fragmentos y fusiones. Los mutantes pueden incluir sustituciones, adiciones o supresiones de aminoácidos. Las sustituciones de aminoácidos pueden ser sustituciones conservadoras de aminoácidos o sustituciones para eliminar aminoácidos no esenciales, tales como para alterar un sitio de glicosilación, un sitio de fosforilación o un sitio de acetilación, o para minimizar el plegamiento incorrecto por sustitución o supresión de uno o más residuos de cisteína que no sean necesarios para la función. Las sustituciones conservadoras de aminoácidos son aquellas que conservan la carga general, la hidrofobicidad / hidrofiliidad, y/o volumen estérico del aminoácido sustituido. Por ejemplo, las sustituciones entre los siguientes grupos son conservadoras: Gly / Ala, Val / Ile / Leu, Asp / Glu, Lys / Arg, Asn / Gln, Ser / Cys, Thr, y Phe / Trp / Tyr.

Las muteínas empobrecidas en cisteína son variantes dentro del alcance de la invención. Estas variantes pueden ser construidas de acuerdo con los métodos divulgados en la patente de los Estados Unidos No. 4.959.314, "Cysteine-Depleted Muteins of Biologically Active Proteins". La patente divulga cómo sustituir otros aminoácidos por cisteínas, y cómo determinar la actividad biológica y el efecto de la sustitución. Tales métodos son adecuados para las proteínas como se describe aquí que tienen residuos de cisteína adecuados para tales sustituciones, por ejemplo

para eliminar la formación de enlaces disulfuro.

Las variantes de proteínas descritas en la presente memoria son codificadas por los polinucleótidos descritos aquí. El código genético se puede utilizar para seleccionar los codones apropiados para construir las variantes correspondientes.

- 5 La invención abarca secuencias de polinucleótidos que tienen al menos 65% de identidad de secuencia y al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335 como se determina por medio del algoritmo de búsqueda de homología de Smith-Waterman tal como el implementado en el programa MSPRCH (Oxford Molecular) usando una búsqueda afín de brechas con los siguientes parámetros de búsqueda: penalización por abertura de brecha de 12 y penalización por extensión de brecha de 1.

- 10 Uso de los polinucleótidos como sondas, en mapeo, y en perfilado de tejidos

Sondas

Las sondas de polinucleótidos que comprenden al menos 12 nucleótidos contiguos seleccionados de la secuencia de nucleótidos de un polinucleótido de la invención se utilizan para una variedad de propósitos, incluyendo la identificación de cromosomas humanos y la determinación de los niveles de transcripción.

- 15 Las sondas de nucleótidos se marcan, por ejemplo, con un marcador radiactivo, fluorescente, biotinilado, o quimioluminiscente, y se detectan por métodos bien conocidos apropiados para el marcador particular seleccionado. Los protocolos para hibridar sondas de nucleótidos para preparaciones de cromosomas en metafase son también bien conocidos en la técnica. Una sonda de nucleótidos hibridará específicamente a secuencias de nucleótidos en las preparaciones de cromosomas que son complementarias a la secuencia de nucleótidos de la sonda. Una sonda
20 que se hibrida específicamente a un polinucleótido deberá proporcionar una señal de detección al menos 5, 10, o 20 veces mayor que la hibridación de fondo proporcionada con otras secuencias no relacionadas.

- En un ejemplo no limitante, se encuentran disponibles programas comerciales para la identificación de regiones de cromosomas comúnmente asociadas con una enfermedad, tal como el cáncer. Los polinucleótidos de la invención se pueden utilizar para sondear estas regiones. Por ejemplo, si a través de perfil de búsqueda se identifica un
25 polinucleótido como correspondiente a un gen que codifica una quinasa, su capacidad para unirse a una región cromosómica relacionada con el cáncer sugerirá su papel como quinasa en una o más etapas de desarrollo / crecimiento de las células del tumor. Aunque se requeriría alguna experimentación para dilucidar el papel, el polinucleótido constituye un nuevo material para aislar una proteína específica que tiene el potencial para desarrollar un diagnóstico o terapéutica para el cáncer.

- 30 Las sondas de nucleótidos se utilizan para detectar la expresión de un gen correspondiente al polinucleótido. Por ejemplo, en transferencias tipo Northern, el ARNm se separa electroforéticamente y se pone en contacto con una sonda. Una sonda se detecta hibridando con una especie de ARNm de un tamaño particular. La cantidad de hibridación se cuantifica para determinar las cantidades relativas de expresión, por ejemplo, bajo una condición particular. Las sondas se utilizan también para detectar productos de amplificación por medio de la reacción en
35 cadena de la polimerasa. Los productos de la reacción se hibridan con la sonda y se detectan los híbridos. Las sondas se usan para la hibridación in situ con células para detectar la expresión. Las sondas también se pueden utilizar in vivo para la detección diagnóstica de secuencias que se hibridan. Las sondas están típicamente marcadas con un isótopo radiactivo. Se pueden utilizar otros tipos de marcadores detectables, tales como cromóforos, compuestos fluorescentes, y enzimas.

- 40 La expresión de ARNm específico puede variar en diferentes tipos de células y puede ser específica del tejido. Esta variación de los niveles de ARNm en diferentes tipos de células se puede explotar con análisis de sondas de ácido nucleico para determinar los tipos de tejidos. Por ejemplo, la PCR, ensayos con sondas de ADN ramificadas, o técnicas de transferencia que utilizan sondas de ácido nucleico sustancialmente idénticas o complementarias a los polinucleótidos enumerados en el Listado de secuencias puede determinar la presencia o ausencia de ADNc o
45 ARNm relacionada con los polinucleótidos de la invención.

Ejemplos de un ensayo de hibridación de nucleótidos se describen en Urdea et al., PCT WO92/02526 y Urdea et al., patente de los Estados Unidos No. 5.124.246, ambas incorporadas aquí por referencia. Las referencias describen un ejemplo de un ensayo de hibridación de nucleótidos tipo sándwich.

- 50 Alternativamente, la Reacción en Cadena de la Polimerasa (PCR) es otro medio para detectar pequeñas cantidades de ácidos nucleicos objetivo, como se describe en Mullis et al., Meth. Enzymol. (1987) 155: 335 - 350; patente de los Estados Unidos No. 4.683.195, y la patente de los Estados Unidos No. 4.683.202. Dos nucleótidos de polinucleótidos iniciadores hibridan con los ácidos nucleicos objetivo y se usan para cebar la reacción. Los

iniciadores pueden estar compuestos de una secuencia dentro o 3' y 5' hacia los polinucleótidos del Listado de secuencias. Alternativamente, si los iniciadores son 3' y 5' hacia esos polinucleótidos, no necesitan hibridarse con ellos o con los complementos. Una polimerasa termoestable crea copias de ácidos nucleicos objetivo a partir de los iniciadores usando los ácidos nucleicos objetivo originales como plantilla. Después de que se generan una gran cantidad de ácidos nucleicos objetivo por la polimerasa, se detectan por métodos tales como transferencias tipo Southern. Cuando se utiliza el método de transferencias tipo Southern, la sonda marcada se hibridará con un polinucleótido del Listado de secuencias o complemento.

Además, se pueden detectar el ARNm o ADNc mediante técnicas de transferencia tradicionales descritas en Sambrook et al., "Molecular Cloning: A Laboratory Manual". (Nueva York, Cold Spring Harbor Laboratory, 1989). El ARNm o el ADNc generados a partir de ARNm usando una enzima polimerasa se pueden purificar y separar usando electroforesis en gel. Los ácidos nucleicos en el gel se transfirieron luego a un soporte sólido, tal como nitrocelulosa. El soporte sólido se expone a una sonda marcada y después se lava para remover cualquier sonda no hibridada. A continuación, se detectan los dúplex que contienen la sonda marcada. Típicamente, se marca la sonda con radiactividad.

15 Mapeo

Los polinucleótidos de la presente invención se utilizan para identificar un cromosoma en el que reside el gen correspondiente. Utilizando hibridación in situ fluorescente (FISH) en propagaciones de metafase normales, la hibridación genómica comparativa permite la evaluación total del genoma de cambios en el número de copias relativo de secuencias de ADN. Véase Schwartz y Samad, *Current Opinions in Biotechnology* (1994) 8: 70 - 74; Kallioniemi et al., *Seminars in Cancer Biology* (1993) 4: 41 - 46; Valdés y Tagle, *Methods in Molecular Biology* (1997) 68:1, Boultonwood, ed., Human Press, Totowa, NJ.

Las preparaciones de cromosomas humanos en metafase se preparan usando técnicas estándar citogenéticas a partir de tejidos humanos primarios o líneas celulares. Se utilizan sondas de nucleótidos que contienen al menos 12 nucleótidos contiguos seleccionados de la secuencia de nucleótidos mostrada en el Listado de Secuencias para identificar el cromosoma correspondiente. Las sondas de nucleótidos se marcan, por ejemplo, con una marca radiactiva, fluorescente, biotinilada, o quimioluminiscente, y se detectan por métodos bien conocidos apropiados para el marcador particular seleccionado. Los protocolos para hibridar sondas de nucleótidos para preparaciones de cromosomas en metafase son también bien conocidos en la técnica. Una sonda de nucleótidos hibridará específicamente con secuencias de nucleótidos en las preparaciones de cromosomas que son complementarias a la secuencia de nucleótidos de la sonda. Una sonda que hibrida específicamente con un gen relacionado con un polinucleótido proporciona una señal de detección al menos 5, 10, o 20 veces mayor que la hibridación de fondo proporcionada con secuencias de codificación que no son EST.

Los polinucleótidos se mapean con cromosomas particulares usando, por ejemplo, híbridos de radiación o paneles híbridos específicos del cromosoma. Véase Leach et al., *Advances in Genetics*, (1995) 33: 63 - 99; Walter et al., *Nature Genetics* (1994) 7: 22 - 28; Walter Goodfellow, *Trends in Genetics* (1992) 9: 352. Tal mapeo puede ser útil en la identificación del gen relacionado con el polinucleótido por su proximidad a otros genes con una función conocida. La función también se puede asignar al gen relacionado cuando síndromes o enfermedades particulares apuntan al mismo cromosoma.

Perfilado del tejido

Los polinucleótidos de la presente invención se puede utilizar para determinar el tipo de tejido del que se deriva una muestra dada. Por ejemplo, una lesión metastática se identifica por su fuente de órgano o de tejido en desarrollo mediante la identificación de la expresión de un marcador particular de ese órgano o tejido. Si un polinucleótido se expresa únicamente en un tipo de tejido específico, y se encuentra una lesión metastática expresa ese polinucleótido, entonces se ha identificado la fuente de desarrollo de la lesión. La expresión de un polinucleótido particular se ensaya mediante la detección ya sea del ARNm correspondiente o el producto proteico. Los métodos inmunológicos, tales como la tinción de anticuerpos, se usan para detectar un producto proteico particular. Se pueden utilizar métodos de hibridación para detectar especies particulares de ARNm, incluyendo pero sin limitarse a hibridación in situ y transferencias tipo Northern.

El uso de Polimorfismos

Un polinucleótido será útil en la medicina forense, análisis genético, mapeo y aplicaciones diagnósticas si la región correspondiente de un gen es polimórfica en la población humana. Una forma polimórfica particular del polinucleótido puede ser utilizada ya sea para identificar una muestra que se derivada de un sospechoso o excluir la posibilidad de que la muestra se derive del sospechoso. Se usa cualquier medio para detectar un polimorfismo en un gen, incluyendo pero sin limitarse a electroforesis de variantes polimórficas de proteína, sensibilidad diferencial a la escisión con enzimas de restricción, e hibridación con una sonda específica del alelo.

Uso de polinucleótidos para generar anticuerpos

5 Los productos de expresión de un polinucleótido, el ARNm o ADNc correspondientes o el gen completo correspondiente se preparan y utilizan para generar anticuerpos para fines experimentales, de diagnóstico y terapéuticos. El polinucleótido o el ADNc relacionado se expresan como se describió anteriormente, y se preparan anticuerpos. Estos anticuerpos son específicos para un epítipo sobre el polipéptido codificado por el polinucleótido y pueden precipitar o unirse a la proteína nativa correspondiente en una preparación celular o tisular o en un extracto libre de células de un sistema de expresión in vitro.

10 Los inmunógenos para generar anticuerpos se preparan mezclando los polipéptidos codificados por el polinucleótido de la presente invención con adyuvantes. Alternativamente, los polipéptidos se elaboran como proteínas de fusión hasta grandes proteínas inmunogénicas. Los polipéptidos están covalentemente enlazados también a otras proteínas inmunogénicas más grandes, tales como hemocianina de lapa. Los inmunógenos se administran típicamente por vía intradérmica, subcutánea, o intramuscular. Los inmunógenos se administra a animales experimentales tales como conejos, ovejas y ratones, para generar anticuerpos. Opcionalmente, se aíslan las células de bazo de animales y se fusionan con células de mieloma para formar hibridomas que secretan anticuerpos monoclonales. Tales métodos son bien conocidos en la técnica. De acuerdo con otro método conocido en la técnica, se administra el polinucleótido directamente, tal como por medio de inyección intramuscular, y se expresa in vivo. La proteína expresada genera una variedad de respuestas inmunes específicas de la proteína, incluyendo la producción de anticuerpos, comparable a la administración de la proteína.

20 Las preparaciones de anticuerpos policlonales y monoclonales específicos para proteínas y polipéptidos codificados por un polinucleótido se elaboran utilizando métodos estándar conocidos en la técnica. Los anticuerpos se unen específicamente a epítopos presentes en los polipéptidos codificados por los polinucleótidos divulgados en el Listado de Secuencias. Típicamente, se requieren al menos 6, 8, 10, o 12 aminoácidos contiguos para formar un epítipo. Sin embargo, los epítopos que implican aminoácidos no contiguos pueden requerir más, por ejemplo al menos 15, 25, o 50 aminoácidos. Una secuencia corta de un polinucleótido puede ser entonces inadecuada para uso como un epítipo para generar anticuerpos para identificar la proteína nueva correspondiente, debido al potencial de reactividad cruzada con una proteína conocida. Sin embargo, los anticuerpos pueden ser útiles para otros propósitos, en particular si identifican características estructurales comunes de una proteína conocida y un nuevo polipéptido codificado por un polinucleótido de la invención.

30 Los anticuerpos que se unen específicamente a polipéptidos codificados por polinucleótidos deben proporcionar una señal de detección al menos 5, 10, o 20 mayor que una señal de detección proporcionada con otras proteínas cuando se usan en transferencias tipo Western u otros ensayos inmunoquímicos. Preferiblemente, los anticuerpos que se unen específicamente a polipéptidos no detectan otras proteínas en ensayos inmunoquímicos y pueden inmunoprecipitar proteínas codificadas por EST a partir de la solución. Para tales inmunoensayos, se pueden utilizar cualquier tipo de muestras, incluyendo tejido, órganos, células, orina, sangre, fluido prostático o semen.

35 De interés son los anticuerpos para los polipéptidos secretados codificados por las secuencias de polinucleótidos de la presente invención. Los anticuerpos para los polipéptidos secretados se puede utilizar para analizar fluidos corporales, tales como sangre, orina, fluido prostático y semen.

40 Para analizar la presencia de anticuerpos en suero con el polipéptido en una población humana, se purifican los anticuerpos humanos por medio de métodos bien conocidos en la técnica. Preferiblemente, los anticuerpos se purifican por afinidad pasando antisuero sobre una columna sobre la que se une una proteína, polipéptido o proteína de fusión. Los anticuerpos unidos se pueden eluir entonces de la columna, por ejemplo usando un amortiguador con una alta concentración de sal.

Además de los anticuerpos discutidos anteriormente, se elaboran derivados de anticuerpos genéticamente modificados por ingeniería genética, tales como anticuerpos de una sola cadena o anticuerpos humanizados.

45 Se prepararon anticuerpos utilizando dos secuencias de clon SL-5: H₂N-CGPRLPSFPCPTHEPSTGQLSK-CONH₂ y H₂N-CKDSQGLSDFKR-NSRTTTRRSYKCCONH₂. El uso de anticuerpos policlonales producidos contra una mezcla de estos polipéptidos, se llevó a cabo inmunohistoquímica en una variedad de tejidos tumorales y el correspondiente tejido normal. Los resultados se muestran en la Figura 3, y se discuten en los Ejemplos. Estos polipéptidos son útiles para la detección de un nivel más alto de expresión de clon SL-5 en tejidos tumorales.

50 Uso de polinucleótidos para construir matrices para diagnósticos

Las presentes secuencias de polinucleótidos y los productos génicos son útiles para determinar la aparición de cáncer, la progresión tumoral, el crecimiento hiperproliferativo, y/o las manifestaciones biológicas o físicas acompañantes. Específicamente, los polinucleótidos de la presente invención pueden ser utilizados para determinar la aparición de trastornos de la próstata, tales como hiperplasia prostática benigna BPH o cáncer localizado de

próstata.

5 Existen una cantidad de trastornos prostáticos, incluyendo adenocarcinoma, BPH, cáncer histológico de próstata, neoplasia intraepitelial prostática, cáncer clínico de próstata, cáncer incidental de próstata, y cáncer localizado de próstata. La BPH es un trastorno común de próstata en los hombres, que se manifiesta clínicamente por lo general después de los cincuenta. En la BPH, el crecimiento hiperplásico de células prostáticas en el tejido glandular periuretral en la zona central de la glándula de la próstata causa un agrandamiento de la próstata que puede comprimir o alargar la uretra y producir síntomas de obstrucción de la uretra que pueden progresar hasta retención urinaria o a una constelación de síntomas conocidos como prostatismo. Una multitud de manifestaciones físicas pueden acompañar a los trastornos de la próstata que incluyen: impotencia, disminución del flujo urinario, vacilación para iniciar la micción, goteo postmiccional, sensación de vaciamiento incompleto de la vejiga, y el desarrollo de infecciones en la vejiga o del tracto urinario superior.

15 Para determinar la aparición de cáncer de próstata, progresión tumoral en la próstata, células prostáticas de crecimiento hiperproliferativo, y/o manifestaciones biológicas o físicas acompañantes, se comparan los niveles de polinucleótidos de la presente invención en una muestra con los niveles en un control normal de los tejidos, células, órganos o fluidos corporales. El control normal puede incluir un conjunto de células de un órgano o tejido particular o tejidos y/o células de todo el cuerpo. Cualquiera de los inmunoensayos descritos anteriormente o los ensayos de ácido nucleico descritos a continuación pueden utilizarse para tales mediciones.

20 Cualquier diferencia observada entre la muestra y el control normal puede indicar la presencia de la enfermedad o el trastorno. Típicamente, si los niveles de los polinucleótidos de la presente invención son superiores a los encontrados en el control normal, los resultados indican la presencia de cáncer de próstata, progresión tumoral en la próstata, crecimiento hiperproliferativo de células de la próstata, y/o manifestaciones biológicas o físicas acompañantes.

25 Además, los polinucleótidos presentes pueden ser útiles para diagnosticar la gravedad, así como la aparición de cáncer de próstata, progresión del tumoral de la próstata, crecimiento de células hiperproliferativas de la próstata, y/o las manifestaciones biológicas o físicas acompañantes, incluyendo trastornos prostáticos. Por ejemplo, cuanto mayor es la diferencia observada en la muestra en comparación con el control normal de los polinucleótidos presentes o los polipéptidos codificados, mayor es la gravedad del trastorno, en particular, cuando se observan niveles más altos en comparación con un control normal.

30 Los polinucleótidos de la presente invención, se expresaron en niveles más altos en una línea de células de cáncer de próstata en comparación con una línea de células normales del epitelio prostático.

Las matrices de polinucleótidos proporcionan una técnica de alto rendimiento que puede analizar un gran número de secuencias de polinucleótidos en una muestra. Esta tecnología se puede utilizar como diagnóstico y como una herramienta para analizar la expresión diferencial para determinar la función de una proteína codificada.

35 Para crear matrices, se colocan sondas de polinucleótidos sobre un sustrato en una matriz o arreglo bidimensional. Las muestras de polinucleótidos se pueden marcar y luego se hibridan con las sondas. Los polinucleótidos bicatenarios que comprenden los polinucleótidos de la muestra marcada unidos a los polinucleótidos de la sonda, se pueden detectar una vez que la porción no unida de la muestra se elimine por lavado.

40 Los polinucleótidos de la sonda se pueden colocar en sustratos incluyendo vidrio, nitrocelulosa, etc. Las sondas se pueden unir al sustrato ya sea mediante enlaces covalentes o por medio de interacciones no específicas, tales como interacciones hidrófobas. Los polinucleótidos de la muestra pueden marcarse usando marcadores radiactivos, fluoróforos, etc.

45 Las técnicas para construir matrices y los métodos de uso de estas matrices se describen en los documentos EP No. 0 799 897; PCT No. WO 97/29212; PCT No. WO 97/27317, EP No. 0 785 280; PCT No. WO 97/02357; la patente de los Estados Unidos No. 5.593.839; la patente de los Estados Unidos No. 5.578.832, EP No. 0 728 520; la patente de los Estados Unidos No. 5.599.695, EP No. 0 721 016; la patente de los Estados Unidos No. 5.556.752; PCT No. WO 95/22058, y la patente de los Estados Unidos No. 5.631.734.

50 Además, las matrices se pueden utilizar para examinar la expresión diferencial de genes y se puede utilizar para determinar la función de genes. Por ejemplo, las matrices de las presentes secuencias de polinucleótidos se puede utilizar para determinar si cualquiera de las secuencias de EST se expresa diferencialmente entre células normales y células cancerosas, por ejemplo. Una expresión alta de un mensaje particular en una célula cancerosa, que no se observa en una célula normal correspondiente, puede indicar una proteína específica de cáncer.

Expresión diferencial

También se describe un método para identificar tejido anormal o enfermo en un humano. Para los polinucleótidos correspondientes a los perfiles de las familias de proteína como se describió anteriormente, la escogencia del tejido puede ser dictada por la función biológica putativa. La expresión de un gen correspondiente a un polinucleótido específico se compara entre un primer tejido que se sospecha que es un tejido enfermo y segundo, tejido normal de humano. El tejido normal es cualquier tejido de humano, especialmente aquellos que expresan el gen relacionado con el polinucleótido, incluyendo, pero sin limitarse a, cerebro, timo, testículo, corazón, próstata, placenta, bazo, intestino delgado, músculo esquelético, páncreas, y el revestimiento mucoso del colon.

Los genes relacionados con el polinucleótido en los dos tejidos se comparan por cualquier medio conocido en la técnica. Por ejemplo, se secuencian los dos genes, y la secuencia del gen en el tejido sospechoso de estar enfermo se compara con la secuencia del gen en el tejido normal. Los genes relacionados con el polinucleótido, o porciones de los mismos, en los dos tejidos se amplifican, por ejemplo, usando cebadores de nucleótidos basados en la secuencia de nucleótidos mostrada en el Listado de secuencias, usando la reacción en cadena de la polimerasa. Los genes amplificados o las porciones de genes se hibridan con sondas de nucleótidos seleccionados de la misma secuencia de nucleótidos mostrada en el Listado de secuencias. Una diferencia en la secuencia de nucleótidos del gen relacionado con polinucleótido en el tejido sospechoso de estar enfermo en comparación con la secuencia normal de nucleótidos sugiere un papel de las proteínas codificadas por polinucleótidos en la enfermedad, y proporciona una ventaja para la preparación de un agente terapéutico. Las sondas nucleotídicas se marcan mediante una variedad de métodos, tales como radiomarcación, biotilación, o marcación con etiquetas fluorescentes o quimioluminiscentes, y se detectan mediante métodos estándar conocidos en la técnica.

Alternativamente, se compara el ARNm relacionado con el polinucleótido en los dos tejidos. El ARN Poli A⁺ se aísla de los dos tejidos como se conoce en la técnica. Por ejemplo, un experto en la técnica puede determinar fácilmente las diferencias en el tamaño o cantidad de transcritos de ARNm relacionados con el polinucleótido entre los dos tejidos utilizando transferencias tipo Northern y sondas de nucleótidos seleccionadas de la secuencia de nucleótidos mostrada en el Listado de secuencias. La mayor o menor expresión de un ARNm relacionado con el polinucleótido en una muestra de tejido sospechoso de estar enfermo, en comparación con la expresión del mismo ARNm relacionado con el polinucleótido en un tejido normal, sugiere que la proteína expresada tiene un papel en la enfermedad, y proporciona también una ventaja para la preparación de un agente terapéutico.

Cualquier método para el análisis de proteínas se utiliza para comparar dos proteínas codificadas por un polinucleótido a partir de muestras emparejadas. Los tamaños de las proteínas en los dos tejidos se comparan, por ejemplo, utilizando anticuerpos de la presente invención para detectar proteínas codificadas por un polinucleótido en transferencias tipo Western de extractos de proteína de los dos tejidos. Otros cambios, tales como los niveles de expresión y la localización subcelular, también se puede detectar inmunológicamente, utilizando anticuerpos para la proteína correspondiente. Un nivel más alto o más bajo de expresión de la proteína codificada por el polinucleótido en un tejido sospechoso de estar enfermo, en comparación con el mismo nivel de expresión de la proteína codificada por el polinucleótido en un tejido normal, es indicativo de que la proteína expresada tiene un papel en la enfermedad, y proporciona otra ventaja para la preparación de un agente terapéutico.

De manera similar, la comparación de secuencias de genes de polinucleótido o de productos de expresión génica de polinucleótido, por ejemplo, ARNm y proteína, entre un tejido humano que es sospechoso de estar enfermo y un tejido normal de un humano, se utilizan para seguir la progresión de la enfermedad o la remisión en el humano. Tales comparaciones de genes relacionados con el polinucleótido, ARNm o proteína se realizan como se describió anteriormente.

Por ejemplo, la expresión aumentada o disminuida del gen relacionado con el polinucleótido en el tejido sospechoso de ser neoplásico puede indicar la presencia de células neoplásicas en el tejido. El grado de aumento de la expresión del gen de polinucleótido en el tejido neoplásico con respecto a la expresión del gen en el tejido normal, o las diferencias en la cantidad de aumento de la expresión del gen de polinucleótido en el tejido neoplásico con el transcurso del tiempo, se utiliza para evaluar la progresión de la neoplasia en que el tejido o para controlar la respuesta del tejido neoplásico con un protocolo terapéutico con el transcurso del tiempo. El patrón de expresión de cualquiera de los dos tipos de células puede ser comparado, tal como las líneas de células tumorales muy y poco metastáticas, o células de tejido que han sido y no han sido expuestas a un agente terapéutico.

Realizaciones relacionadas con el ordenador

En general, una biblioteca de polinucleótidos es una colección de información secuencial, cuya información se proporciona ya sea en forma bioquímica (por ejemplo, como una colección de moléculas de polinucleótido) o en forma electrónica (por ejemplo, como una colección de secuencias de polinucleótidos almacenada en una forma legible por un ordenador, como en un sistema informático y/o como parte de un programa de ordenador). La información de la secuencia de los polinucleótidos se puede usar en una variedad de formas, por ejemplo, como un

recurso para el descubrimiento de genes, como una representación de secuencias expresadas en un tipo seleccionado de célula (por ejemplo, marcadores de tipo celular), y/o como marcadores de una enfermedad dada o estado de enfermedad. En general, un marcador de enfermedad es una representación de un producto génico que está presente en todas las células afectadas por la enfermedad bien sea a un nivel incrementado o disminuido en relación con una célula normal (por ejemplo, una célula del mismo tipo o similar que no esté sustancialmente afectada por enfermedad).

La información de la secuencia de nucleótidos de la biblioteca puede ser incorporada en cualquier forma adecuada, por ejemplo, en formas electrónicas o bioquímicas. Por ejemplo, una biblioteca de información de la secuencia incorporada en forma electrónica comprende un archivo de datos accesible por el ordenador (o, en forma bioquímica, una colección de moléculas de ácido nucleico) que contiene las secuencias representativas de nucleótidos de los genes que son expresados diferencialmente (por ejemplo, sobreexpresados o infraexpresados) como, por ejemplo, entre una célula cancerosa y una célula normal. Las realizaciones bioquímicas de la biblioteca incluyen una colección de ácidos nucleicos que tienen las secuencias de los genes en la biblioteca, donde los ácidos nucleicos pueden corresponder al gen completo en la biblioteca o a un fragmento del mismo, como se describe en mayor detalle a continuación.

Las bibliotecas de polinucleótidos descritas aquí comprenden generalmente información de secuencia de una pluralidad de secuencias de polinucleótidos, en donde al menos uno de los polinucleótidos es un polinucleótido de la presente invención. Por pluralidad se entiende al menos 2, usualmente al menos 3 y puede incluir hasta todos los polinucleótidos de la presente invención. La longitud y el número de polinucleótidos en la biblioteca variará con la naturaleza de la biblioteca, por ejemplo, si la biblioteca es una matriz de oligonucleótidos, una matriz de ADNc, una base de datos informática de la información de la secuencia, etc.

Cuando la biblioteca es una biblioteca electrónica, la información de la secuencia de ácido nucleico puede estar presente en una variedad de medios. "Medios" se refiere a una fabricación, distinta de una molécula de ácido nucleico aislada, que contiene la información de la secuencia de la presente invención. Dicha fabricación proporciona la secuencia del genoma o un subgrupo de la misma en una forma que puede ser examinada por los medios no directamente aplicables a la secuencia tal como existe en un ácido nucleico. Por ejemplo, la secuencia de nucleótidos, por ejemplo, las secuencias de ácido nucleico de cualquiera de los polinucleótidos de la presente invención, se puede grabar en medios legibles por el ordenador, por ejemplo, cualquier medio que pueda ser leído y accedido directamente por un ordenador. Tales medios incluyen, pero no se limitan a: medios de almacenamiento magnéticos, tales como un disquete, un medio de almacenamiento en disco duro y una cinta magnética; medios de almacenamiento óptico tales como CD-ROM; medios de almacenamiento eléctrico tales como RAM y ROM; e híbridos de estas categorías tales como medios de almacenamiento magnético / óptico. Un experto en la técnica puede apreciar fácilmente cómo cualquiera de los medios informáticos conocidos en la actualidad legibles por un ordenador se puede utilizar para crear una fabricación que comprende un registro de la presente información de la secuencia. "Grabado" se refiere a un proceso para almacenar información en un medio legible por ordenador, utilizando cualquiera de tales métodos conocidos en la técnica. Se puede escoger cualquier estructura conveniente de almacenamiento de datos en función de los medios utilizados para acceder a la información almacenada. Una variedad de programas y formatos de procesadores de datos pueden usarse para el almacenamiento, por ejemplo, un procesador de palabra para archivos de texto, formato de bases de datos, etc. Además de la información de la secuencia, se pueden proporcionar versiones electrónicas de las bibliotecas junto o en conexión con otra información legible por ordenador y/o otros tipos de archivos legibles por ordenador (por ejemplo, archivos de búsqueda, archivos ejecutables, etc., incluyendo, pero sin limitarse a, por ejemplo, un software de un programa de búsqueda, etc.)

Al proporcionar la secuencia de nucleótidos en forma legible por ordenador, se puede acceder a la información para una variedad de propósitos. Un software de ordenador para acceder a información de las secuencias se encuentra públicamente disponible. Por ejemplo, se pueden utilizar los algoritmos de búsqueda BLAST (Altschul et al., véase más arriba) y BLAZE (Brutlag et al. Comp. Chem. (1993) 17: 203) sobre un sistema Sybase para identificar los marcos de lectura abiertos (ORF) dentro del genoma que contiene homología con los ORF de otros organismos.

Como se usa en la presente memoria, "un sistema informático" se refiere a los medios de hardware, los medios de software, y los medios de almacenamiento de datos utilizados para analizar la información de la secuencia de nucleótidos de la presente invención. El hardware mínimo de los sistemas basados en ordenador de la presente invención comprende una unidad central de procesamiento (CPU), medios de entrada, medios de salida, y medios de almacenamiento de datos. Un experto en la técnica puede apreciar fácilmente que cualquiera de los sistemas disponibles en la actualidad basados en un ordenador son adecuados para uso en la presente invención. Los medios de almacenamiento de datos pueden comprender cualquier arquitectura que incluya un registro de la presente información de la secuencia tal como se describió anteriormente, o un medio de acceso a la memoria que pueden acceder a dicha arquitectura.

"Medio de búsqueda" se refiere a uno o más programas implementados sobre el sistema basado en ordenador, para

comparar una secuencia objetivo o un motivo estructural objetivo, o niveles de expresión de un polinucleótido en una muestra, con la información almacenada de la secuencia. Se pueden utilizar medios de búsqueda para identificar fragmentos o regiones del genoma que coincidan con una secuencia objetivo particular o un motivo objetivo. Se conocen públicamente una variedad de algoritmos conocidos y están disponibles comercialmente, por ejemplo, MacPattern (EMBL), BLASTN y BLASTX (NCBI). Una "secuencia objetivo" puede ser cualquier secuencia de polinucleótidos o de aminoácidos de seis o más nucleótidos contiguos o de dos o más aminoácidos, preferiblemente aproximadamente de 10 a 100 aminoácidos o aproximadamente de 30 a 300 nt. Se pueden utilizar una variedad de medios de comparación para realizar la comparación de la información secuencial de una muestra (por ejemplo, para analizar secuencias objetivo, motivos objetivo o niveles relativos de expresión) con los medios de almacenamiento de datos. Un experto en la técnica puede reconocer fácilmente que se puede utilizar cualquiera de los programas de búsqueda de homología que se encuentran públicamente disponibles como medio de búsqueda para los sistemas basados en ordenador descritos aquí para llevar a cabo la comparación de secuencias y motivos objetivo. También se conocen en la técnica programas de ordenador para analizar los niveles de expresión en una muestra y en controles.

Un "motivo estructural objetivo" o "motivo objetivo", se refiere a cualquier secuencia seleccionada racionalmente o combinación de secuencias en las que la(s) secuencia (s) se eligen con base en una configuración tridimensional que se forma tras el plegamiento del motivo objetivo, o sobre secuencias consenso de sitios reguladores o activos. Hay una variedad de motivos objetivo conocidos en la técnica. Los motivos objetivo de proteínas incluyen, pero no se limitan a, sitios activos de enzimas y secuencias señal. Los motivos objetivo de ácido nucleico incluyen, pero no se limitan a, estructuras de horquilla, secuencias promotoras y otros elementos de expresión tales como sitios de enlazamiento para factores de transcripción.

Se pueden utilizar una variedad de formatos estructurales para los medios de entrada y de salida para la entrada y salida de la información en los sistemas basados en ordenador de la presente invención. Un formato para un medio de salida clasifica los niveles de expresión relativos de diferentes polinucleótidos. Tal presentación le proporciona a un experto en la técnica una clasificación de los niveles de expresión relativos para determinar un perfil de expresión génica.

Como se discutió anteriormente, la "biblioteca" descrita aquí también abarca bibliotecas bioquímicas de los polinucleótidos de la invención, por ejemplo, colecciones de ácidos nucleicos que representan los polinucleótidos proporcionados. Las bibliotecas bioquímicas pueden tomar una variedad de formas, por ejemplo, una solución de los ADNc, un patrón de ácidos nucleicos sonda asociados en forma estable con una superficie de un soporte sólido (es decir, una matriz) y similares. De particular interés son las matrices de ácido nucleico en las que uno o más de los polinucleótidos de la invención está representado en la matriz. Por matriz se entiende un artículo de fabricación que tiene por lo menos un sustrato con al menos dos objetivos distintos de ácido nucleico en una de sus superficies, donde el número de ácidos nucleicos distintos pueden ser considerablemente más alto, siendo típicamente al menos de 10 nt, usualmente al menos de 20 nt y, a menudo al menos de 25 nt. Se han desarrollado una variedad de formatos de matriz diferentes y son conocidos por los expertos en la técnica. Las matrices de la invención objetivo encuentran uso en una variedad de aplicaciones, incluyendo análisis de expresión génica, cribado de fármacos, análisis de mutaciones y similares, tal como se divulga en los ejemplos de documentos de patente mencionados anteriormente.

Además de las bibliotecas de ácidos nucleicos anteriores, también se proveen bibliotecas de análogos de polipéptidos, donde los polipéptidos de la biblioteca representarán al menos una porción de los polipéptidos codificados por los nucleótidos de la invención.

La presente invención se ilustrará ahora mediante referencia a los siguientes ejemplos que exponen realizaciones particularmente ventajosas. Sin embargo, hay que señalar que estas realizaciones son ilustrativas y no deben interpretarse como limitativas de la invención de ninguna manera.

Ejemplos

Ejemplo 1

Aislamiento de los polinucleótidos

Se prepararon bibliotecas de ADNc a partir de PREC, células epiteliales de próstata humana normal, y LNCaP, una línea de células derivada de cáncer de próstata que hizo metástasis en un ganglio linfático humano. Las células PREC se encuentran disponibles con Clonetics, San Diego, California, EE.UU. Las células LNCaP se encuentran disponibles con la ATCC, Manassas, Virginia, EE.UU.

Utilizando una técnica de PCR y los reactivos disponibles con Clontech, Palo Alto, California, EE.UU. (CLONTECH PCR-Select™), se capturó y amplificó el ARNm que aumentó su expresión en LNCaP. Los insertos de

polinucleótidos capturados se insertaron en el vector pCR2.1, disponible en Invitrogen, Carlsbad, California, EE.UU. Los vectores con los insertos se transformaron en células de *E. coli*.

Ejemplo 2

Confirmación de visualización diferencial

- 5 Diez clones fueron seleccionados al azar, y el aumento de la expresión de las secuencias de estos insertos de clones en células LNCaP frente a células PrEC fue confirmada por medio de transferencias tipo Northern. Se llevaron a cabo transferencias de puntos en 168 clones y se confirmó el aumento de la expresión.

- 10 Además, la secuenciación de los clones mostró que las secuencias del antígeno prostático específico (PSA) y del antígeno prostático específico de membrana (PSMA) fueron aisladas mediante el procedimiento descrito en el Ejemplo 1. Se ha observado una buena correlación entre el aumento de los niveles de PSA en suero y los tumores de próstata. El PSMA, un antígeno de la superficie de la célula, es otro marcador observado para el cáncer de próstata. Véase, Bosland, Encyclopedia of Cancer, Volumen II, páginas 1283 - 1296 (1997), Academic Press. Por lo tanto, los datos confirman que se clonó la característica del ARNm de mayor expresión génica en el cáncer de próstata mediante el método del Ejemplo 1.

15 Ejemplo 3

Secuencias de polinucleótidos

- Algunos de los resultados de las secuencias se muestran en las SEQ ID NOs: 218, 219 o 335. Para los experimentos de secuenciación, cada clon se denominó SL-1 hasta SL-209. Cada secuencia se denominó como una combinación única de dos nombres. Esta combinación única se muestra en la Tabla 1 en las columnas 2 y 3, indicada como "Nombre de secuencia" y "Otro nombre de secuencia".

La Tabla 1 indica todas las secuencias que corresponden al clon SL-68.

A los clones también se les asignan números de agrupación. Véase la columna 4 de la Tabla 1. Los clones con el mismo número de agrupación generalmente comprenden la secuencia derivada de los mismos transcritos de ARNm.

- 25 La última columna de la Tabla 1 indica el vecino más cercano, determinado por medio de un alineamiento de secuencias en una base de datos públicamente disponible.

Se puede construir un consenso para la secuencia de cada clon mediante la alineación de las secuencias correspondientes o complementos inversos de las mismas. La Tabla 1 enlista los nombres de todas las secuencias que corresponden a cada clon, y la Tabla 2 muestra la secuencia específica que corresponde a cada combinación única de Nombre de secuencia y/o "Otro nombre de secuencia".

- 30 El inserto completo de algunos clones no puede ser representado por las secuencias presentadas en la Tabla 2. Por ejemplo, los extremos 5' y 3' de un inserto de clon pueden haber sido secuenciados, pero las secuencias no se superponen. La secuencia adicional correspondiente al inserto del clon se puede aislar y determinar por medio de la construcción de sondas o cebadores a partir de las secuencias presentadas en la Tabla 2 y una biblioteca de ARNm o ADNc de una célula de próstata o línea de células de cáncer de próstata utilizando los métodos descritos anteriormente.

Ejemplo 4

Resultados de la búsqueda en bases de datos públicas

- 40 Tanto la secuencia de nucleótidos como las traducciones de las secuencias enmascaradas mostradas en el Listado de secuencias fueron alineadas con las secuencias individuales que estaban disponibles al público. Se usa la similitud con secuencias individuales para determinar la actividad de los polipéptidos codificados por los genes correspondientes a las secuencias mencionadas en la Tabla 2.

- 45 Las secuencias en las SEQ ID NOs: 218 o 219 primero fueron enmascarados para remover las secuencias del vector pCR2.1 El enmascaramiento se llevó a cabo mediante la alineación de las secuencias de pCR2.1 con cada una de las SEQ ID NOs: 218 o 219 utilizando el programa BLASTN. Cualquier secuencia que produzca una alineación con una puntuación menor a 0,1 fue enmascarada.

Se realizó una búsqueda en BLASTN vs. el GenBank utilizando las secuencias de enmascaradas con parámetros de búsqueda de más de 99% de superposición, 99% de identidad y un valor de p de menos de 1×10^{-40} y esto dio lugar a descartes de secuencias. Secuencias de búsqueda también fueron descartadas si los parámetros de inclusión se cumplieron, pero la secuencia fue ribosomal o derivada del vector.

- 5 Las secuencias resultantes de la búsqueda anterior se clasificaron en tres grupos (1, 2 y 3 a continuación) y se realizaron búsquedas en una búsqueda en las bases de datos BLASTX vs. NRP (proteínas no redundantes): (1) desconocido (sin aciertos en la búsqueda del GenBank), (2) similitud débil (más de 45% de identidad y un valor de p de menos de 1×10^{-5}), y (3) alta similitud (más de 60% de superposición, más de 80% de identidad, y un valor de p inferior a 1×10^{-5}). Esta búsqueda dio lugar a descartes de secuencias por tener más del 99% de superposición, más
10 de 99% de identidad, y un valor de p de menos de 1×10^{-40} .

- Las secuencias restantes se clasificaron como desconocidas (sin aciertos), de similitud débil y de similitud alta (parámetros como los anteriores). Se realizaron dos búsquedas en este conjunto de secuencias. En primer lugar, una búsqueda en la base de datos BLAST vs. EST dio como resultado un descarte de secuencias con más de 99% de superposición, similitud mayor al 99% y un valor de p de menos de 1×10^{-40} ; secuencias con un valor p de menos
15 de 1×10^{-65} cuando se compara con una secuencia de la base de datos de origen humano también fueron excluidas. En segundo lugar, una base de datos BLASTN vs. Patent GeneSeq dio como resultado el descarte de secuencias con identidad mayor al 99%; un valor de p inferior a 1×10^{-40} ; una superposición mayor al 99%.

- Las secuencias enmascaradas fueron traducidas en los seis marcos de lectura para determinar el mejor alineamiento con las secuencias individuales. Estas secuencias de aminoácidos y secuencias de nucleótidos se denominan, generalmente, como secuencias de consulta, que se alinean con las secuencias individuales.
20

Las secuencias individuales y de consulta fueron alineadas utilizando los programas BLAST, disponibles en la red informática mundial.

- La Tabla 2 muestra los resultados de los alineamientos. La Tabla 2 se refiere a cada secuencia por su Nombre de secuencia y/o "Otro nombre de secuencia" e incluye los números de acceso y las descripciones de los vecinos más cercanos del GenBank y búsquedas de proteína no redundantes.
25

La actividad del polipéptido codificado por las secuencias mencionadas en la Tabla 2 se espera que sea la misma o similar a la del vecino más cercano reportado en la Tabla 2. El número de acceso del vecino más próximo se informa, proporcionando una referencia para las actividades exhibidas por el vecino más cercano. El programa de búsqueda y la base de datos usada para la alineación también se indican, así como un cálculo del valor de p.

- 30 Las secuencias de longitud completa o los fragmentos de las secuencias de polinucleótidos de los vecinos más cercanos se pueden utilizar como sondas y cebadores para identificar y aislar la secuencia de longitud completa correspondiente a la secuencia mencionada en la Tabla 2. Aunque se pueden obtener las secuencias de longitud completa a partir de las líneas de células descritas anteriormente, los vecinos más cercanos pueden indicar un tejido o un tipo de célula para ser utilizados para construir una biblioteca para las secuencias de longitud completa de
35 aquellas mencionadas en la Tabla 2.

- Las secuencias mencionadas en la Tabla 2 y las traducciones de las mismas pueden ser homólogos humanos de los genes conocidos de otras especies o nuevas variantes alélicas de genes humanos conocidos. En tales casos, estas nuevas secuencias humanas pueden ser adecuadas como agentes de diagnóstico, pronóstico, o terapéuticos. Como agentes de diagnóstico, las secuencias humanas exhiben una mayor especificidad en la detección y diferenciación
40 de líneas de células humanas y tipos que los homólogos de otras especies. Los polipéptidos humanos son menos propensos a ser inmunogénicos cuando se administran a humanos que los homólogos de otras especies. Además, en la administración a humanos, los polipéptidos codificados pueden mostrar mayor especificidad o pueden ser mejor regulados por otras proteínas humanas que son homólogos de otras especies.

- En las formas de realización preferidas de la invención, las secuencias mostradas en las SEQ ID NOs: 218, 219 o 335 que consisten de las regiones no enmascaradas se deben considerar como la fuente de sondas y cebadores, ya que estas secuencias son más representativas de las porciones distintivas de estos polinucleótidos.
45

Generalmente, el enmascaramiento en sí mismo no influye en los resultados de la búsqueda como se muestra en la Tabla 2, excepto para eliminar múltiples "éxitos" basados en la similitud de las regiones repetitivas comunes a más de un polipéptido.

50 Ejemplo 5

Análisis del clon

Clon SL-68 (SEQ ID NOs: 218 y 219)

Dos transcritos, de 2,6 kb y de 4,3 kb, fueron observados en un bazo y timo normales y en leucocitos de sangre periférica, así como en la leucemia promielocítica, leucemia mieloide crónica y leucemia linfoblástica. El transcrito de 4,3 kb fue observado en testículos y colon normales, células HeLa S3, adenocarcinoma colorrectal y melanoma. La banda de 2,6 kb se encontró en las siguientes líneas de células de próstata: PC-3 (metastático al hueso, insensibles a andrógenos); DU-145 (metastático al cerebro, insensible a andrógenos); FFpz (células primarias derivadas del epitelio de próstata normal); Ffca (células primarias derivadas de epitelio de cáncer de próstata Grado 3 de Gleason), y WO-CA (células primarias derivadas de epitelio de cáncer de próstata Grado 4 de Gleason). Sin embargo, una expresión más alta se observó en LNCaP, MDA PCa 2A, HPV-7 y HPV-10. Se observó también un transcrito de 9,5 kb en MDA PCa 2A y 2B. Una secuencia adicional correspondiente a este clon se divulga en la SEQ ID NO: 335.

Aquellos capacitados en el arte reconocerán, o serán capaces de determinar, utilizando únicamente experimentación de rutina, muchos equivalentes a las realizaciones específicas de la invención descritas aquí. Tales realizaciones específicas y equivalentes pretenden ser abarcadas por las reivindicaciones siguientes.

Tabla 1

Clon #	Nombre de la secuencia	Otro nombre de la secuencia	Clon # Agrupación #	Vecino más cercano si lo hay
SL-068	SL068m13 SL068t7	WE97.B4.M13 WE97.B4.T7	SL-068	

Tabla 2

	BlastN vs. Gb (vecino más cercano)			BlastX vs. NRPdb (vecino más cercano)		
Nombre de la secuencia y/u otro nombre de la secuencia	Acceso	Descripción de éxito	P(V)	Acceso	Descripción de éxito	P(V)
SL68m13	AC004157	***SECUENCIACIÓN EN PROGRESO ***Plasmodium falciparum 3D7 cromosoma 12 Secuencia genómica PFYAC293; HTGS fase 1, 18 piezas no ordenadas.	0,00071	<NINGUNO>	<NINGUNA>	<NING UNO>
SL68t7	AJ226619	Fragmento genómico de Ciona intestinalis, clone 17H6, secuencia genómica en estudio.	0,064	<NINGUNO>	<NINGUNA>	<NING UNO>

Listado de secuencia

<110> Zhang, Jimmy

Aste1, Jon H.

Carroll III, Eddie

Endege, wilson O.

Pord, Donna M.

Monahan, John B.

Schlegel, Robert

5 Steinmann, Kathleen E.

<120> GENES Y PRODUCTOS DE EXPRESIÓN GÉNICA QUE SON DIFERENCIALMENTE REGULADOS EN CÁNCER DE PRÓSTATA

<130> 200130.463

<140> US

10 <141> 1999-06-11

<160> 3

<170> FastSEQ para Windows Versión 3.0

<210> 218

<211> 498

15 <212> ADN

<213> Homo Sapiens

<400> 218

```

cacttgacct gattacgcca agcttggtac cgagctcgga tccactagta acggccgcca      60
gtgtgctgga attcgccgtt tcgagcggcc gcccgggcag gtactttttt tttttttttt      120
taattccac aacaacccat ttcaaaatga gaaaactagg ttgagtgact tgtccacagt      180
tccaaagcta ataaaaatga tgaggcatat ttctcttctg ggccactgt attcagttct      240
ttgttcttta cactgagtgc cgaaaaaaaa aaatcagact attttgattc tagaaagtga      300
gataattgaa aatgttaaca tatttctcca aaactgatca gactgtggag tctgtcactt      360
ttttggtata ataaaggagt ttgaagaaac aaatgacatc attcctgatg atggtagccc      420
actccaacaa aggcgtatat atgtaggcaa gtttgaagat atctataaga gcattaaaag      480
gcaagtgcac cattgtgg

```

<210> 219

20 <211> 818

<212> ADN

<213> Homo Sapiens

<220>

<221> característica nueva

25 <222> (1)...(818)

<223> n = A,T,C o G

<400> 219

ggcctntnga	gctgctcgac	ggccgccatg	tgatggatat	ctgcagaatt	cgcccttagc	60
gtggcgcggc	cgaggtacct	agaaaacaga	aacttgagta	gacatggtaa	tgaccagaaa	120
aggctatctt	tatacatctt	ttttgctacg	cttcaaattc	atgtcaccta	aaagttgtga	180
agtgcacaaa	acaatctac	ttaactgaaa	attatcttca	atgaatggga	tgtttagaac	240
tctgtgaggg	tttttaaggt	cttttcgaat	agcaaattct	aatgaggctt	ttttaagttg	300
gcaatttaaa	ctcatacaag	aaataaaaaac	tcaccagtgt	ggctgggcag	aatatatata	360
ttttctcaaa	tattgtttgt	ttgttttttc	cctgcactgt	atccatgggc	ccatgatgaa	420
actgttatat	tgctgatata	tttattggaa	tatgtgggcc	aacttccttt	ccactcaaca	480
tatggattgg	tagtttaaaa	taattccttt	ctattaagca	aatgtgtggc	taaggcacat	540
ttaaatagcc	cattaaacca	atgagatgac	aatgtgttac	cctcagagaa	agcttaattt	600
ttggagtaat	caattacaca	tatcacagaa	tgtctcatga	gaacattttt	ggctaggtct	660
accaatttat	catgcaaata	attatagatt	ttcatttgag	gcaaagatgc	tgattcatca	720
ttagtaacat	ggtcacaaat	aatcatttat	tttattttgg	taacatctgt	ctttctctgtg	780
gggaaactta	ctatatgtct	tacgttaatt	aaattaaa			818

<210> 335

5 <211> 1312

<212> ADN

<213> Homo sapiens

<400> 335

acctagaaaa	cagaaacttg	agtagacatg	gtaatgacca	gaaaaggcta	tctttatata	60
tttctttttg	tacgcttcaa	attcatgtca	cctaaaagtt	gtgaagtgca	caaaacaaat	120
ctactttaact	gaaaattatt	ttcaatgaat	gggatgttta	gaactctgtg	agggttttta	180
agggtcttttc	gaatagcaaa	ttctaattgag	ggttttttaa	gttggcaatt	taaactcata	240
caagaaataa	aaactcacca	gtgtggctgg	gcagaatata	tatattttct	caaatattgt	300
ttgtttgttt	tttccctgca	ctgtatccat	gggtcccatga	tgaaactgtt	atattgctga	360
tatattttatt	ggaatatgtg	ggccaacttc	ctttccactc	aacatatgga	ttggtagttt	420
aaaataattc	ctttctatta	agcaaatgtg	tggctaaggc	acatttaaat	agcccatata	480
accaatgaga	tgacaatgtg	ttaccctcag	agaaaagctta	atttttggag	taatcaatta	540
cacatatcac	agaatgtctc	atgagaacat	ttttggctag	gtctaccaat	ttatcatgca	600
aataattata	gattttcatt	tgaggcaaaag	atgctgattc	atcattagta	acatggtcac	660
aaataatcat	ttattttatt	ttgtttaaca	tctgtctttc	ctgtggggaa	acttactata	720
tgctctacgt	ttatttaatt	taaaaagtca	attggttatt	ctgaattttt	aaaaataaca	780
taaaactggt	gttctaatac	acagcacctg	cttttctttt	tttagtgaaa	ttatataagc	840
atttagagaa	tgaaagtgtg	agacttgtgg	tttctggtct	ctttttactg	tttgtaagcc	900
tactcgtcat	gatattccac	aatgggtgcac	ttgcctttta	atgctcttat	agatatcttc	960
aaacttgctt	acatatatac	gcctttgttg	gagtgggcta	ccatcatcag	gaatgatgtc	1020
atttggtttc	tcaaaactcc	ttattatacc	aaaaaagtga	cagactccac	agtctgatca	1080
gttttgagaga	aatatgttaa	cattttcaat	tatctcactt	tctagaatca	aaatagtctg	1140
attttttttt	ttcggcactc	agtgtaaaga	acaaagaact	gaatacagtg	ggcccagaag	1200
agaaatatgc	ctcatcattt	ttattagctt	tggaaactgtg	gacaagtcac	tcaacctagt	1260
tttctcattt	tgaaatgggt	tgttgtggga	attaaaaaaa	aaaaaaaaaa	gt	1312

10

REIVINDICACIONES

1. Un método para diagnosticar cáncer de próstata, la progresión de un tumor en la próstata y el crecimiento de células hiperproliferativas de la próstata que comprende:
 - 5 (a) proporcionar una sonda de polinucleótido que comprende al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335 o un complemento de las mismas;
 - (b) poner en contacto una muestra biológica para el diagnóstico con dicha sonda bajo condiciones de hibridación que permitan la formación de un dúplex; y
 - (c) determinar la presencia de dicho dúplex, y detectar las diferencias que existen entre los niveles de dúplex en dicha muestra biológica comparado con una muestra biológica normal.
- 10 2. Un kit de diagnóstico que comprende:
 - (a) un reactivo de diagnóstico que comprende una sonda de polinucleótido que contiene al menos 20 nucleótidos contiguos de las SEQ ID NOs: 218, 219 o 335 o complemento de las mismas, cuando dicha secuencia está presente en una muestra biológica de ensayo;
 - (b) una muestra biológica normal; y
- 15 (c) instrucciones para la detección de las diferencias que existen entre los niveles de dúplex en dicha muestra biológica de ensayo en comparación con dicha muestra biológica normal.
3. Un polinucleótido aislado seleccionado de entre el grupo que consiste de: un polinucleótido que comprende la secuencia de nucleótidos de las SEQ ID NOs: 218 o 219.
4. Un vector que comprende el polinucleótido de la reivindicación 3.
- 20 5. Una célula huésped aislada que comprende el vector de la reivindicación 4.