



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 305 087**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

**C12N 15/11** (2006.01)

**A61K 31/00** (2006.01)

**G01N 33/574** (2006.01)

**G01N 33/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **01948414 .6**

86 Fecha de presentación : **15.06.2001**

87 Número de publicación de la solicitud: **1370684**

87 Fecha de publicación de la solicitud: **17.12.2003**

54 Título: **Polinucleótidos relacionados con el cáncer de colon.**

30 Prioridad: **15.06.2000 US 211835 P**

45 Fecha de publicación de la mención BOPI:  
**01.11.2008**

45 Fecha de la publicación del folleto de la patente:  
**01.11.2008**

73 Titular/es: **Novartis Vaccines and Diagnostics, Inc.**  
**4560 Horton Street**  
**Emeryville, California 94608, US**

72 Inventor/es: **Kennedy, Giulia, C.;**  
**Kang, Sanmao;**  
**Reinhard, Christoph y**  
**Jefferson, Anne, Bennet**

74 Agente: **Carpintero López, Mario**

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

Polinucleótidos relacionados con el cáncer de colon.

5 **Campo de la invención**

La presente invención se refiere a genes expresados diferencialmente en el cáncer de colon y en la displasia. Más concretamente, se refiere a polinucleótidos que se regulan diferencialmente en el cáncer de colon y a los productos génicos codificados.

10 **Antecedentes de la invención**

El cáncer de colon es la segunda causa de mortalidad relacionada con el cáncer en Estados Unidos. La Sociedad Estadounidense contra el Cáncer estimó que habría aproximadamente 94.700 nuevos casos de cáncer de colon en Estados Unidos en 1999 y que el cáncer de colon sería responsable de aproximadamente 47.900 muertes. El colon tiene cuatro partes: el colon ascendente, el colon transversal, el colon descendente y el colon sigmoide, terminando con el recto. Los pólipos adenomatosos o los adenomas, lesiones benignas comunes que se convierten en carcinomas, se pueden desarrollar en cualquiera de las cuatro partes del colon o en el recto. Más del 95% de los cánceres de colon son adenocarcinomas o cánceres de las células que revisten el interior del colon. El cáncer de colon frecuentemente se metastatiza hasta el hígado y el pulmón.

A diferencia del cáncer de pulmón, en el que se ha identificado el consumo del tabaco como el principal factor etiológico responsable de la enfermedad, los mecanismos fundamentales subyacentes del cáncer de colon son complejos y no se comprenden del todo. Se cree que los factores alimenticios favorecen la carcinogénesis, especialmente, un consumo elevado de grasas. A nivel molecular, se sospecha de la existencia de un proceso de múltiples etapas que implica un número de mutaciones en la progresión de los adenomas a tumores de colon (Vogelstein *et al.* (1988) *N. Engl. J. Med.* 319: 525-532). El desarrollo y la progresión del cáncer de colon está conducido por mutaciones secuenciales en tres tipos de genes: oncogenes, genes de supresión tumoral y genes de reparación de apareamientos erróneos, que controlan la velocidad de las mutaciones de otros genes, incluyendo la de los oncogenes y de los genes de supresión tumoral. Estas mutaciones se producen como resultado de una predisposición genética (mutaciones de la línea germinal) o en respuesta a factores ambientales (mutaciones somáticas).

Se han identificado varias mutaciones que están asociadas con el cáncer de colon. Las mutaciones de la línea germinal que se han relacionado con el cáncer de colon hereditario o familiar incluyen al gen de supresión tumoral de la poliposis coli adenomatosa (PCA) (Lengauer *et al.* (1991) *Science* 253: 665-669) y los genes de reparación de apareamientos erróneos MutL y MutS (Modrich (1995) *Phil. Trans. R. Soc. Lond. B* 347: 89-95; Kolodner (1996) *Genes Dev.* 10: 1433-1442). El PCA defectuoso ha sido implicado en la poliposis adenomatosa familiar (PAF) y los genes MutL y MutS en el cáncer colorrectal no polipoide hereditario (CCNPH). Las mutaciones somáticas identificadas en asociación con el cáncer de colon esporádico incluyen los oncogenes K-ras, c-myc y los genes de supresión tumoral p53, PCA, la proteína activadora de GTPasa de la neurofibromatosis de tipo 1 (PAG NF1), eliminadas en el cáncer de colon (ECC) y mutadas en el cáncer de colon (MCC) (Midgley *et al.* (1999) *Lancet* 353: 391-399). Los documentos WO 99/33963 y WO 00/22130 describen genes marcadores expresados diferencialmente que son regulados secuencia arriba en el cáncer de colon. El número de acceso AAZ 52884 describe el aislamiento de una secuencia altamente similar de una genoteca de ADNc del tumor de próstata humano.

El cáncer de colon es muy tratable y a menudo curable cuando se detecta y se trata en fases tempranas. Los procedimientos de diagnóstico convencionales incluyen los procedimientos invasivos, tales como la exploración rectal digital, la sigmoidoscopia, la colonoscopia y la exploración intestinal con bario, así como procedimientos no invasivos, tales como el análisis de sangre oculta fecal y el rastreo genético. El rastreo de marcadores tumorales está particularmente indicado para la identificación de la enfermedad hereditaria, así como para la diagnosis de recurrencia. Por ejemplo, el rastreo del antígeno carcinoembrionario (ACE) se usa para diagnosticar la recurrencia asintomática. Los nuevos procedimientos de diagnóstico incluyen técnicas de formación de imágenes de fluorescencia inducida por láser que pueden detectar células cancerosas sobre la superficie epitelial o en el interior de la pared del colon (véase, p. ej., von Rueden *et al.* (1993) *J. Surg. Oncol.* 53: 43-46).

Los enfoques terapéuticos convencionales para tratar el cáncer de colon incluyen la resección quirúrgica, la radiación y la quimioterapia, incluyendo la terapia con adyuvantes. Los enfoques terapéuticos génicos incluyen la transferencia de genes de citoquina o de inmunoantígenos, la transferencia de sistemas de enzima-profármaco (véase, p. ej., Huber *et al.* (1993) *Cancer Res.* 53: 4619-4626) y el reemplazo de genes de supresión tumoral (véase, p. ej., Venook *et al.* (1998) *Proc. ASCO* 17:431a) usando vectores víricos (Zwacka *et al.* (1998) *Hematol. Oncol. Clin. North Am.* 12: 595-615).

Aunque se han identificado varios genes asociados con el cáncer de colon, la identificación de otros genes asociados con el desarrollo (o la inhibición del desarrollo) del cáncer de colon puede proporcionar otras herramientas de diagnóstico y dianas terapéuticas. La identificación de los genes que se expresan diferencialmente en el cáncer de colon es particularmente importante en el avance del descubrimiento de fármacos, de las tecnologías de diagnóstico y de la comprensión de la progresión y la naturaleza del cáncer de colon. La invención proporciona la identificación de tales genes expresados diferencialmente.

## Resumen de la invención

Esta invención se refiere a polinucleótidos que representan genes expresados diferencialmente en el cáncer de colon, p. ej., el pólipo adenomatoso, el carcinoma colorrectal, el tumor de colon de alto potencial metastásico y el

Por consiguiente, en un aspecto, la invención muestra un procedimiento para identificar una célula de colon cancerosa, implicando el procedimiento la detección de al menos un producto génico expresado diferencialmente, en el que el producto génico está codificado por un gen que comprende una secuencia de SEQ ID NO: 1 en una muestra de análisis, estando la muestra de análisis derivada de una célula de análisis sospechosa de ser una célula de colon cancerosa, y la comparación del nivel de expresión del producto génico expresado diferencialmente detectado con el nivel de expresión del producto génico expresado diferencialmente de una muestra control, estando la muestra control derivada de una célula de colon cancerosa. La detección del nivel de expresión del producto génico expresado diferencialmente de la muestra de análisis que es similar al nivel de expresión del producto génico de la muestra control indica que la célula de análisis es una célula de colon cancerosa. En una realización, la detección se realiza mediante la hibridación de la muestra de análisis con un alineamiento de referencia, comprendiendo el alineamiento de referencia una secuencia identificadora de SEQ ID NO: 1.

La invención también describe un procedimiento para identificar una célula de colon cancerosa, implicando el procedimiento la detección de al menos un producto génico diferencialmente expresado, en el que la detección se realiza detectando la hibridación de un polinucleótido que comprende una secuencia de SEQ ID NO: 1 en una muestra de análisis, estando la muestra de análisis derivada de una célula de análisis sospechosa de ser una célula de colon cancerosa, y la comparación del nivel de hibridación del producto génico expresado diferencialmente detectado con el nivel de hibridación del producto génico expresado diferencialmente de una muestra control, estando la muestra control derivada de una célula de colon cancerosa. La detección del nivel de hibridación del producto génico expresado diferencialmente de la muestra de análisis que es similar al nivel de hibridación del producto génico de la muestra control indica que la célula de análisis es una célula de colon cancerosa. En una realización, la detección se realiza mediante la hibridación de la muestra de análisis con un alineamiento de referencia, comprendiendo el alineamiento de referencia una secuencia identificadora de SEQ ID NO: 1.

La invención también describe un polinucleótido aislado que comprende una secuencia de SEQ ID NO: 1 o variantes degeneradas de la misma. En aspectos relacionados, la invención muestra alineamientos y células huésped recombinantes que comprenden un polinucleótido de la invención. En una realización, el polinucleótido incluye la secuencia de nucleótidos de un inserto contenido en un clon descrito en la presente memoria y depositado en la ATCC.

En otro aspecto, la invención muestra un polipéptido aislado codificado por un gen expresado diferencialmente de la invención, así como anticuerpos que se unen específicamente a tales polipéptidos.

En otro aspecto, la invención muestra composiciones terapéuticas que comprenden un agente activo para la modulación de la expresión de un gen expresado diferencialmente en células de colon canceroso. Por ejemplo, el agente activo de la composición terapéutica puede efectuar una disminución de la actividad biológica de un producto génico codificado por un gen que está sobre-expresado en una célula cancerosa en comparación con una célula normal o puede efectuar un aumento de la actividad biológica de un producto génico codificado por un gen infra-expresado en una célula cancerosa en comparación con una célula normal.

La invención también muestra una genoteca de genes expresados diferencialmente, en la que la genoteca incluye la información secuencial del polinucleótido de SEQ ID NO: 1.

La genoteca puede proporcionarse como un alineamiento de ácidos nucleicos o en un formato electrónico, y puede incluir cantidades relativas del polinucleótido de la SEQ ID NO: 1, siendo las cantidades relativas representativas de las cantidades relativas del polinucleótido encontrado en una célula de colon enfermo.

Un objeto fundamental de la invención consiste en proporcionar polinucleótidos correspondientes a genes expresados diferencialmente, y fragmentos de los mismos, que sean útiles en el diagnóstico del cáncer de colon, así como en la creación de fármacos y terapias.

Tras leer la descripción proporcionada en la presente memoria, habrá diversos aspectos y realizaciones de la invención fácilmente comprensibles por el experto habitual de la técnica.

## Breve descripción de las figuras

La Fig. 1 es una gráfica que muestra los niveles de los mensajes del gen correspondiente a SK2 (c9083, SEQ ID NO: 3) en las líneas celulares indicadas.

La Fig. 2 es una gráfica que muestra el efecto de los oligonucleótidos antisentido SK2 (9083) en los niveles de los mensajes del gen correspondiente a SK2 (SEQ ID NO: 3).

Las gráficas de las figuras 3 y 4 muestran el efecto de los oligonucleótidos antisentido SK2 (9083) en la proliferación de células SW620 (Fig. 3) o de una línea celular no colónica, la HT1080 (Fig. 4).

La Fig. 5 es una gráfica que muestra el efecto de los oligonucleótidos antisentido en el gen correspondiente al grupo 378805 en el crecimiento de células SW620 (31-4as: antisentido; 31-4rc: control inverso; TS: control de tipo silvestre (no oligo)).

Las Fig. 6-8 son gráficas que muestran los resultados del análisis de proliferación con análisis de SW620 para examinar los efectos de la expresión de K-Ras (control, Fig. 6), el gen correspondiente a c3376 (CHIR11-4) y el gen correspondiente a 402380 (CHIR33-4).

La Fig. 9 es una gráfica que muestra los efectos de la expresión de genes correspondientes a K-Ras (control) y a 402380 (CHIR33-4) en la formación en el colon de células SW620 en agar disuelto (valores normalizados con respecto a WST 1).

## Descripción detallada de la invención

Antes de seguir describiendo la presente invención, se entenderá que la invención no se limita a las realizaciones particulares de la invención descritas a continuación, pues se pueden realizar variaciones de las realizaciones particulares sin que dejen de pertenecer al ámbito de las reivindicaciones anexas. También se entenderá que la terminología empleada se hace a efectos de describir las realizaciones particulares, y no pretende ser restrictiva. En cambio, el ámbito de la presente invención estará establecido por las reivindicaciones anexas.

En esta memoria y en las reivindicaciones anexas, las formas singulares de los artículos “un”, “una”, “el” y “la” incluyen sus referencias en plural, a no ser que el contexto establezca claramente algo distinto. A no ser que se definan de otro modo, todos los términos técnicos y científicos usados en la presente memoria tienen el mismo significado entendido comúnmente por cualquier experto habitual en la técnica a la que pertenece esta invención.

La invención se refiere a polinucleótidos que comprenden las secuencias de nucleótidos reveladas, a ADNc de longitud completa, a secuencias genómicas de ARNm y a los genes correspondientes a estas secuencias, así como a las variantes degeneradas de las mismas, y a los polipéptidos codificados por los polinucleótidos de la invención y a las variantes de polipéptido. La siguiente descripción detallada describe las composiciones de los polinucleótidos englobadas por la invención, los procedimientos para obtener ADNc y ADN genómico codificante de un producto génico de longitud completa, la expresión de estos polinucleótidos y genes, la identificación de los motivos estructurales de los polinucleótidos y de los genes, la identificación de la función de un producto génico codificado por un gen correspondiente a un polinucleótido de la invención, el uso de los polinucleótidos proporcionados como sondas y en el mapeo y el perfilado de tejidos, el uso de los correspondientes polipéptidos y de otros productos génicos para generar anticuerpos, y el uso de los polinucleótidos y sus productos génicos codificados a efectos terapéuticos y de diagnóstico.

## Definiciones

Los términos “polinucleótido” y “ácido nucleico”, usados indistintamente en la presente memoria, se refieren a formas poliméricas de nucleótidos de cualquier longitud, bien ribonucleótidos o desoxinucleótidos. De este modo, estos términos incluyen además, pero no se limitan a, ADN o ARN mono-, bi- o multicatenario, ADN genómico, ADNc, híbridos de ADN-ARN, o un polímero que comprende bases de purina o pirimidina u otras bases de nucleótido química o bioquímicamente modificadas, no naturales o derivadas. Estos términos incluyen además, pero no se limitan a, ARNm o ADNc que comprende secuencias intrónicas (véase, p. ej., Niwa *et al.* (1999) *Cell* 99(7): 691-702). El esqueleto del polinucleótido puede comprender azúcares y grupos fosfato (como se pueden encontrar comúnmente en el ARN o el ADN) o azúcares o grupos fosfato sustituidos o modificados. Alternativamente, el esqueleto del polinucleótido puede comprender un polímero de subunidades sintéticas tales como fosforamiditas y, de este modo, puede ser un fosforamidoato de oligodesoxinucleósido o un oligómero mixto de fosforamidoato-fosfodiéster. Peyrottes *et al.* (1996) *Nucl. Acids Res.* 24: 1841-1848; Chaturvedi *et al.* (1996) *Nucl. Acids Res.* 24: 2318-2323. Un polinucleótido puede comprender nucleótidos modificados, tales como nucleótidos metilados o análogos de nucleótidos metilados, uracilo, otros azúcares y grupos de unión, tales como fluororribosa y tioato, así como ramas de nucleótidos. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleotídicos. Se puede modificar más un polinucleótido tras la polimerización, tal como mediante la conjugación con un componente de marcaje. Otros tipos de modificaciones incluidas en esta definición son las caperuzas, la sustitución de uno o más de los nucleótidos naturales por un análogo y la introducción de medios de unión del polinucleótido a proteínas, iones metálicos, componentes de marcaje, otros polinucleótidos o un soporte sólido.

Los términos “polipéptido” y “proteína”, usados indistintamente en la presente memoria, se refieren a una forma polimérica de aminoácidos de cualquier longitud, que puede incluir aminoácidos codificados y no codificados, aminoácidos química o bioquímicamente modificados o derivados, y polipéptidos que tengan esqueletos peptídicos modificados. El término incluye proteínas de fusión, incluyendo, pero no limitándose a, proteínas de fusión con una secuencia de aminoácidos heteróloga, fusiones con secuencias líderes homólogas y heterólogas, con o sin residuos de metionina N-terminales; proteínas marcadas inmunológicamente, y similares.

“Heterólogo” significa que los materiales están derivados de diferentes fuentes (p. ej., de diferentes genes, diferentes especies, etc.).

El término “gen expresado diferencialmente” pretende englobar un polinucleótido que representa o se corresponde con un gen que está expresado diferencialmente en una célula de colon cancerosa en comparación con una célula del mismo tipo celular que no es cancerosa. Tal gen expresado diferencialmente puede incluir un marco de lectura abierto que codifica un producto génico (p. ej., un polipéptido), así como intrones de tales genes y secuencias de nucleótidos no codificantes de 5’ y 3’ adyacentes implicadas en la regulación de la expresión, hasta aproximadamente 20 kb más allá de la región codificante, pero posiblemente además en cualquier dirección. El gen puede introducirse en un vector apropiado para el mantenimiento extracromosómico o para la integración en un genoma huésped. En general, una diferencia en el nivel de expresión asociada con una disminución del nivel de expresión del al menos aproximadamente 25%, habitualmente, del al menos aproximadamente 50% al 75%, más habitualmente del al menos aproximadamente 90% o más es indicativa de un gen expresado diferencialmente de interés, i.e., un gen que está infra-expresado o regulado secuencialmente abajo en una muestra de análisis en comparación con una muestra control. Además, una diferencia en el nivel de expresión asociada con un aumento del nivel de expresión del al menos aproximadamente 25%, habitualmente, del al menos aproximadamente 50% al 75%, más habitualmente del al menos aproximadamente 90%, pudiendo ser un aumento al menos aproximadamente una vez y media más, habitualmente, al menos aproximadamente 2 veces más a aproximadamente 10 veces más, pudiendo ser de aproximadamente 100 veces más a aproximadamente 1.000 veces más, en comparación con una muestra control, es indicativa de un gen expresado diferencialmente de interés, i.e., un gen sobre-expresado o regulado secuencialmente arriba.

“Polinucleótido expresado diferencialmente” como se usa en la presente memoria significa una molécula de ácido nucleico (ARN o ADN) que comprende una secuencia que representa un gen expresado diferencialmente, p. ej., el polinucleótido expresado diferencialmente comprende una secuencia (p. ej., un marco de lectura abierto codificante de un producto génico) que identifica excepcionalmente un gen expresado diferencialmente de manera que la detección de un polinucleótido expresado diferencialmente en una muestra guarda correlación con la presencia de un gen expresado diferencialmente o un producto génico de un gen expresado diferencialmente en una muestra. Por ejemplo, la detección de un polinucleótido en una muestra que se hibrida (p. ej., en condiciones restrictivas) con un polinucleótido expresado diferencialmente es indicativo de la presencia del correspondiente gen expresado diferencialmente en la muestra. “Polinucleótidos expresados diferencialmente” también pretende englobar fragmentos de los polinucleótidos revelados, p. ej., fragmentos que conservan la actividad biológica, así como ácidos nucleicos que son homólogos, sustancialmente similares o sustancialmente idénticos (p. ej., que tienen una identidad secuencial del aproximadamente 90%) con los polinucleótidos revelados.

“Correspondiente a”, “que se corresponde con” o “que representa” cuando se usan en el contexto de, por ejemplo, un polinucleótido o una secuencia “correspondiente a”, “que se corresponde con” o “que representa” un gen significa que una secuencia del polinucleótido está presente en el gen o en el producto génico del ácido nucleico (p. ej., ARNm). El polinucleótido puede estar completamente presente en un exon de una secuencia genómica del gen o diferentes partes de la secuencia del polinucleótido pueden estar presentes en diferentes exones (p. ej., tal que la secuencia de polinucleótido contigua esté presente en un ARNm, bien antes o después del corte y empalme, es decir, un producto de expresión del gen). En algunas realizaciones, el polinucleótido puede representar o corresponderse con un gen que esté modificado en una célula cancerosa en comparación con una célula normal. Por ejemplo, el gen de la célula cancerosa se puede modificar mediante la inserción de un retrovirus endógeno, un elemento transponible u otro ácido nucleico natural o no natural. En tales casos, el polinucleótido puede incluir tanto secuencias del gen nativo (p. ej., el gen sin la secuencia heteróloga) como secuencias no nativas insertadas.

“Gen” se usa generalmente en la presente memoria para englobar un polinucleótido que codifica un producto génico, p. ej., una secuencia de ácido nucleico que define un marco de lectura abierto.

“Producto génico”, como se usa en la presente memoria, pretende englobar todo o una parte de un producto de expresión de un gen correspondiente a un polinucleótido descrito en la presente memoria, incluyendo, pero no necesariamente limitándose a, una molécula de ARN o un polipéptido.

“Diagnosis”, como se usa en la presente memoria, incluye generalmente la determinación de la susceptibilidad de un sujeto a una enfermedad o a un trastorno, la determinación de si un sujeto está afectado en el presente por una enfermedad o un trastorno, así como al pronóstico de un paciente afectado por una enfermedad o un trastorno. La presente invención engloba la diagnosis de sujetos en el contexto del cáncer de colon (p. ej., el pólipo adenomatoso, el carcinoma colorrectal), así como de cualquier fase de tales cánceres (p. ej., fases I a IV de gravedad).

“Cáncer de colon”, como se usa en la presente memoria, pretende englobar las formas benignas o malignas del cáncer de colon y del cáncer rectal; las formas no metastásicas, premetastásicas y metastásicas del cáncer de colon; y cualquier tipo particular de cáncer surgido de células del colon o del recto (p. ej., pólipo adenomatoso, carcinoma colorrectal y similares).

Los términos “individual”, “sujeto”, “huésped” y “paciente” se usan indistintamente en la presente memoria y se refieren a cualquier sujeto mamífero para quien se desea una diagnosis, un tratamiento o una terapia, particularmente a seres humanos. Otros sujetos pueden incluir ganado, perros, gatos, cerdos de guinea, conejos, ratas, ratones, caballos, etc.

El término “muestra” o “muestra biológica” engloba una variedad de tipos de muestras obtenidas de un organismo y que se pueden usar en un análisis de diagnóstico o de control. El término engloba sangre y otras muestras líquidas de origen biológico, muestras de tejidos sólidos, tales como una muestra de biopsia o cultivos tisulares, o células derivadas de los mismos, y la progenie de las mismas. El término engloba muestras que han sido manipuladas de cualquier modo tras su obtención, tal como mediante un tratamiento con reactivos, una disolución o un enriquecimiento en ciertos componentes. El término engloba una muestra clínica y también incluye células de un cultivo celular, sobrenadantes celulares, lisados celulares, suero, plasma, fluidos biológicos y muestras tisulares.

Una “célula huésped”, como se usa en la presente memoria, se refiere a un microorganismo o a una célula eucariota o a una línea celular cultivada como una entidad unicelular que puede ser, o ha sido, usada como receptor de un vector recombinante u otros polinucleótidos de transferencia, e incluye la progenie de la célula original que ha sido transfectada. Se entiende que la progenie de una única célula puede no ser necesariamente completamente idéntica en morfología o en complemento de ADN genómico o total que el progenitor original, debido a una mutación natural, accidental o deliberada.

Los términos “cáncer”, “neoplasma”, “tumor” y “carcinoma” se usan indistintamente en la presente memoria para referirse a células que presentan un crecimiento relativamente autónomo, tal que presentan un fenotipo de crecimiento aberrante caracterizado por una pérdida significativa del control de la proliferación celular. En general, las células de interés para la detección o el tratamiento en la presente solicitud incluyen células precancerosas (p. ej., benignas), malignas, premetastásicas, metastásicas y no metastásicas. La detección de una célula cancerosa es de particular interés.

“Fenotipo canceroso” se refiere generalmente a cualquiera de una variedad de fenómenos biológicos que son característicos de una célula cancerosa, pudiendo los fenómenos variar con el tipo de cáncer. El fenotipo canceroso se identifica generalmente por anomalías en, por ejemplo, el crecimiento o la proliferación celular (p. ej., un crecimiento o una proliferación incontrolada), la regulación del ciclo general, la movilidad celular o la interacción entre células.

“Agente terapéutico” se refiere generalmente a un gen o a un producto génico que, tras la modulación de su actividad (p. ej., mediante la modulación de la expresión, la actividad biológica y similares) puede proporcionar la modulación del fenotipo canceroso.

Como se usa en la presente memoria, “modulación” pretende referirse a un aumento o a un descenso del fenómeno indicado (p. ej., la modulación de una actividad biológica se refiere a un aumento de una actividad biológica o una disminución de una actividad biológica).

### **Perspectiva general de la invención**

En general, la invención se basa en el descubrimiento de polinucleótidos que representan genes que están expresados diferencialmente en células cancerosas del colon. La expresión diferencial de los genes en las células del colon afectadas por un cáncer se determina mediante, por ejemplo, la detección de los genes expresados en una célula cancerosa del colon y mediante la comparación del nivel de la expresión del gen (p. ej., bien cualitativa o cuantitativamente) con la expresión de esos mismos genes en una célula normal del colon (i.e., una célula de colon que no está afectada por un cáncer de colon).

Los polinucleótidos correspondientes a los genes expresados diferencialmente descritos en la presente memoria se identificaron usando visualizaciones diferenciales de muestras de células normales de colon, células colónicas de tumor primario, células colónicas de tumor metastásico y células de pólipo adenomatoso. La secuencia de los polinucleótidos específicos que representan los genes expresados diferencialmente de la presente invención se muestran las SEQ ID NOS: 1, 3, 5, 7, 9, 11-13, 15, 16, 18, 20, 22, 24, 26, 27 y 29.

#### *Composiciones de polinucleótido*

El ámbito de la invención con respecto a las composiciones de polinucleótido incluye, pero no se limita necesariamente a, los polinucleótidos que comprenden una secuencia expuesta en SEQ ID NO: 1, los polinucleótidos obtenidos de los materiales biológicos descritos en la presente memoria o de otras fuentes biológicas (particularmente de fuentes humanas) mediante la hibridación en condiciones restrictivas (particularmente, en condiciones muy restrictivas); los genes correspondientes a los polinucleótidos proporcionados; las variantes de los polinucleótidos proporcionados y sus correspondientes genes, particularmente, aquellas variantes que conservan una actividad biológica del producto génico codificado (p. ej., una actividad biológica atribuida a un producto génico correspondiente a los polinucleótidos proporcionados como resultado de la asignación del producto génico a una/s familia/s de proteínas y/o de la identificación de un dominio funcional presente en el producto génico). Otras composiciones de ácido nucleico contempladas por y en el ámbito de la presente invención resultarán fácilmente comprensibles por cualquier experto habitual en la técnica cuando sea provisto de la presente revelación.

“Polinucleótido” y “ácido nucleico” como se usan en la presente memoria indistintamente con referencia a los ácidos nucleicos de la composición no pretenden estar limitados a la longitud o a una estructura del ácido nucleico a no ser que si indique específicamente lo contrario. Además, los polinucleótidos descritos en la presente memoria pueden constar esencialmente de secuencias de exones, p. ej., secuencias que definen un marco de lectura abierto y codifican

todo o una parte de un producto génico. El término “que consta esencialmente de” en el contexto de un polinucleótido descrito en la presente memoria significa que el polinucleótido está compuesto de una secuencia codificante de un marco de lectura abierto, pudiendo estar la secuencia flanqueada por cualquiera entre una variedad de secuencias que no afectan materialmente a la/s característica/s básica/s del producto génico codificado. Las secuencias flanqueantes adecuadas incluyen, pero no se limitan necesariamente a, secuencias promotoras, secuencias potenciadoras, sitios de inicio y/o terminación de la transcripción, secuencias de constructos o de vectores, (p. ej., secuencias que proporcionan la manipulación del polinucleótido en una molécula lineal o circular, incluyendo, pero no necesariamente limitándose a, secuencias para la replicación y el mantenimiento del constructo o del vector, secuencias codificantes de productos génicos que proporcionan la selección (p. ej., la resistencia o la sensibilidad a antibióticos, los factores que afectan al crecimiento en medios con o sin suplementos y similares)), las secuencias que proporcionan la producción de una proteína de fusión con el polinucleótido y un polipéptido heterólogo (i.e., un polipéptido codificado por un polinucleótido que procede de una fuente distinta de la del polinucleótido al que está ligado operativamente) y similares.

La invención muestra polinucleótidos que están expresados en tejido humano, específicamente, en tejido de colon humano. Las composiciones de ácido nucleico de la invención de particular interés comprenden una secuencia expuesta en SEQ ID NO: 1 o una secuencia de identificación de la misma. Una “secuencia de identificación” es una secuencia contigua de residuos de una longitud de al menos aproximadamente 10 nt a aproximadamente 20 nt, habitualmente, de una longitud de al menos aproximadamente 50 nt a aproximadamente 100 nt, que identifica excepcionalmente una secuencia de polinucleótidos, p. ej., presenta una identidad secuencial de menos del 90%, habitualmente, de menos del aproximadamente 80% al aproximadamente 85% con cualquier secuencia de nucleótidos contigua de más de aproximadamente 20 nt. De este modo, las presentes composiciones de ácido nucleico incluyen ADNc o ARNm de longitud completa que engloban una secuencia de identificación de los nucleótidos contiguos de SEQ ID NO: 1.

Los polinucleótidos de la invención también incluyen los polinucleótidos que tienen una similitud secuencial o una identidad secuencial. Los ácidos nucleicos que tienen una similitud secuencial se detectan por hibridación en condiciones poco restrictivas, por ejemplo, a 50°C y 10 x SSC (solución salina 0,9 M/citrato de sodio 0,09 M) y permanecen unidos cuando se someten a un lavado a 55°C en 1 x SSC. La identidad secuencial se puede determinar por hibridación en condiciones restrictivas, por ejemplo, a 50°C o más y 0,1 x SSC (solución salina 9 mM/citrato de sodio 0,9 mM). Los procedimientos y las condiciones de hibridación son conocidos en la técnica, véase, p. ej., el documento USPN 5.707.829. Los ácidos nucleicos que son sustancialmente idénticos a las secuencias de polinucleótidos proporcionadas, p. ej., las variantes alélicas, las versiones alteradas genéticamente del gen, etc., se unen a las secuencias de polinucleótidos proporcionadas (SEQ ID NO: 1) en condiciones de hibridación restrictivas. Mediante el uso de sondas, particularmente, de sondas marcadas de secuencias de ADN, es posible aislar genes homólogos o relacionados. El origen de los genes homólogos puede ser cualquier especie, p. ej., especie de primates, particularmente, la especie humana; roedores, tales como ratas y ratones; caninas, felinas, bovinas, ovinas, equinas, de levadura, de nematodos, etc.

En general, la hibridación se realiza usando al menos 15 nucleótidos (nt) contiguos de SEQ ID NO: 1. Es decir, cuando se usan al menos 15 nt contiguos de una de las SEQ ID NO reveladas como sonda, la sonda si hibridará preferiblemente con un ácido nucleico que comprende la secuencia complementaria, permitiendo la identificación y recuperación de los ácidos nucleicos que se hibridan excepcionalmente con la sonda seleccionada. Las sondas de más de una SEQ ID NO se pueden hibridar con el mismo ácido nucleico si el ADNc del que fueron derivadas se corresponde con el mismo ARNm de longitud completa. Se pueden usar sondas de más de 15 nt, p. ej., sondas de desde aproximadamente 18 nt hasta aproximadamente 100 nt, pero 15 nt representan una secuencia suficiente para una identificación exclusiva.

Los polinucleótidos de la invención también incluyen las variantes naturales de las secuencias de nucleótidos (p. ej., las variantes degeneradas, las variantes alélicas, etc.). Las variantes de los polinucleótidos de la invención se identifican mediante la hibridación de las variantes putativas con las secuencias de nucleótidos reveladas en la presente memoria, preferiblemente, mediante hibridación en condiciones restrictivas. Por ejemplo, usando las condiciones de lavado apropiadas, se pueden identificar las variantes de los polinucleótidos de la invención, presentando la variante alélica a lo sumo del 25-30% de apareamientos erróneos de los pares de bases (pb) en comparación con la sonda de polinucleótido seleccionada. En general, las variantes alélicas contienen del 15-25% de apareamientos erróneos de los pb, y pueden seleccionar tan poco como incluso del 5-15% o del 2-5%, o del 1-2% de apareamientos erróneos de los pb, así como un único apareamiento erróneo de los pb.

La invención también engloba homólogos correspondientes a los polinucleótidos de la SEQ ID NO: 1, pudiendo ser el origen de los genes homólogos cualquier especie de mamífero, p. ej., especie de primate, particularmente, la especie humana; roedores, tales como ratas; especies caninas, felinas, bovinas, ovinas, equinas, de levadura, de nematodos, etc. Entre las especies de mamíferos, p. ej., los seres humanos y los ratones, los homólogos tienen generalmente una similitud secuencial sustancial, p. ej., una identidad secuencial del 75%, habitualmente del al menos 90%, más habitualmente del al menos 95% entre las secuencias de nucleótidos. La similitud secuencial se calcula en base a una secuencia de referencia, que puede ser un subconjunto de una secuencia más larga, tal como un motivo conservado, una región codificante, una región flanqueante, etc. Una secuencia de referencia tendrá habitualmente una longitud de al menos aproximadamente 18 nt contiguos, más habitualmente, una longitud de aproximadamente 30 nt, y puede extenderse a la secuencia completa que esté siendo comparada. Los algoritmos para los análisis de secuencias son conocidos en la técnica, tales como el BLAST discontinuo, descrito en Altschul, *et al. Nucleic Acids Res.* (1997) 25: 3389-3402.

En general, las variantes de la invención tienen una identidad secuencial de más del al menos aproximadamente 65%, preferiblemente, del al menos aproximadamente 75%, más preferiblemente, del al menos aproximadamente 85%, y pueden ser mayores del al menos aproximadamente 90% o más, según lo determinado por el algoritmo de búsquedas de homología de Smith-Waterman puesto en marcha en el programa MPSRCH (Oxford Molecular). A efectos de esta invención, un procedimiento preferido para calcular el porcentaje de identidad es el algoritmo de Smith-Waterman, usando lo siguiente. La identidad secuencial del ADN global debe ser mayor del 65% según lo determinado por el algoritmo de búsquedas de homología de Smith-Waterman puesto en marcha en el programa MPSRCH (Oxford Molecular) usando una búsqueda de huecos afines con los siguientes parámetros de búsqueda: penalización por apertura de hueco: 12; penalización por extensión de hueco: 1.

Los presentes ácidos nucleicos pueden ser ADNc o ADN genómico, así como fragmentos de los mismos, particularmente, fragmentos que codifican un producto génico biológicamente activo y/o son útiles en los procedimientos revelados en la presente memoria (p. ej., en la diagnosis, como un identificador exclusivo o un gen expresado diferencialmente de interés, etc.): El término “ADNc”, como se usa en la presente memoria, pretende incluir todos los ácidos nucleicos que comparten la distribución de los elementos secuenciales encontrados en las especies de ARNm maduras nativas en las que los elementos secuenciales son exones y regiones no codificantes de 3' y 5'. Normalmente, las especies de ARNm tienen exones contiguos, con intrones intercalados, cuando están presentes, que son eliminados mediante el corte y empalme del ARN nuclear, para crear un marco de lectura abierto codificante de un polipéptido de la invención.

Una secuencia genómica de interés comprende el ácido nucleico presente entre el codón de inicio y el codón de terminación, según lo definido en las secuencias enumeradas, incluyendo todos los intrones que están normalmente presentes en un cromosoma nativo. También puede incluir las regiones no traducidas de 3' y 5' encontradas en el ARNm maduro. También puede incluir secuencias reguladores de la transcripción y de la traducción específicas, tales como promotores, potenciadores, etc., incluyendo aproximadamente 1 kb, pero posiblemente más, del ADN genómico flanqueante en cualquiera de los extremos 5' y 3' de la región transcrita. Es posible aislar el ADN genómico como un fragmento de 100 kpb o menor, y sustancialmente libre de secuencia cromosómica flanqueante. El ADN genómico flanqueante de la región codificante, tanto 3' como 5', o las secuencias reguladoras internas como se encuentran algunas veces en los intrones, contiene las secuencias requeridas para la expresión del tejido apropiado, específica de la etapa o específica del estado patológico.

Las composiciones de ácido nucleico de la invención pueden codificar todos o parte de los presentes polipéptidos. Se pueden obtener fragmentos bi- o monocatenarios de la secuencia de ADN sintetizando químicamente oligonucleótidos conforme a los procedimientos convencionales, mediante la digestión con enzimas de restricción, mediante la amplificación por PCR, etc. Los polinucleótidos y los fragmentos de polinucleótido aislados de la invención comprenden al menos aproximadamente 10, aproximadamente 15, aproximadamente 20, aproximadamente 35, aproximadamente 50, aproximadamente 150 a aproximadamente 200, aproximadamente 250 a aproximadamente 300, o aproximadamente 350 nt contiguos seleccionados de las secuencias de polinucleótidos según lo mostrado en SEQ ID NO: 1. Para la mayor parte, los fragmentos serán de una longitud de al menos 15 nt, habitualmente, de al menos 18 nt o 25 nt, y de hasta al menos aproximadamente 50 nt contiguos o más. En una realización preferida, las moléculas de polinucleótido comprenden una secuencia contigua de al menos 12 nt seleccionada del grupo constituido por el polinucleótido mostrado en SEQ ID NO: 1.

Las sondas específicas de los polinucleótidos de la invención se pueden generar usando las secuencias de polinucleótidos reveladas en SEQ ID NO: 1. Las sondas son preferiblemente al menos aproximadamente un fragmento de 12, 15, 16, 18, 20, 22, 24 o 25 nt de una secuencia contigua correspondiente de SEQ ID NO: 1, y pueden tener una longitud de menos de 2; 1; 0,5; 0,1 ó 0,05 kb. Las sondas se pueden sintetizar químicamente o se pueden generar a partir de polinucleótidos más largos usando enzimas de restricción. Las sondas se pueden marcar, por ejemplo, con una etiqueta radiactiva, biotinilada o fluorescente. Preferiblemente, las sondas se diseñan en base a una secuencia de identificación de un polinucleótido de SEQ ID NO: 1.

Más preferiblemente, las sondas se diseñan en base a una secuencia contigua de uno de los presentes polinucleótidos que permanecen desenmascarados tras la aplicación de un programa de enmascaramiento para enmascarar la baja complejidad (p. ej., XBLAST) con la secuencia, i.e., se seleccionaría una región desenmascarada, según lo indicado por los polinucleótidos exteriores de los tramos de poli-n de la secuencia enmascarada producida por el programa de enmascaramiento.

Los polinucleótidos de la presente invención se aíslan y obtienen con una pureza sustancial, generalmente, distinta a la de un cromosoma intacto. Habitualmente, los polinucleótidos, bien como ADN o como ARN, se obtendrán sustancialmente libres de otras secuencias de ácidos nucleicos naturales, generalmente, siendo al menos aproximadamente un 50%, habitualmente al menos aproximadamente un 90% puros, y comúnmente “recombinantes”, p. ej., flanqueados por uno o más nucleótidos con los que no están normalmente asociados en un cromosoma natural.

Los polinucleótidos de la invención se pueden proporcionar como una molécula lineal o dentro de una molécula circular, y se pueden proporcionar dentro de moléculas de replicación autónoma (vectores) o dentro de moléculas sin secuencias de replicación. La expresión de los polinucleótidos se puede regular mediante sus propias o mediante otras secuencias reguladoras conocidas en la técnica. Los polinucleótidos de la invención se pueden introducir en células huésped adecuadas usando una variedad de técnicas disponibles en la técnica, tales como la transferencia de ADN mediada por policationes de transferrina, la transfección con ácidos nucleicos desnudos o encapsulados, la



transferencia de ADN mediada por liposomas, el transporte intracelular de perlas de látex revestidas de ADN, la fusión de protoplastos, la infección vírica, la electroporación, la pistola de genes, la transfección mediada por fosfato de calcio y similares.

- 5 Las presentes composiciones de ácido nucleico se pueden usar, por ejemplo, para producir polipéptidos, como sondas para la detección de ARNm de la invención en muestras biológicas (p. ej., extractos de células humanas) para generar más copias de los polinucleótidos, para generar ribozimas u oligonucleótidos antisentido, y como sondas de ADN monocatenarias o como oligonucleótidos que forman cadenas triples. Las sondas descritas en la presente memoria se pueden usar, por ejemplo, para determinar la presencia o la ausencia de las secuencias de polinucleótidos según lo mostrado en la SEQ ID NO: 1 o variantes de la misma en una muestra. Éstos y otros usos se describen más detalladamente a continuación.

#### *Uso de polinucleótidos para obtener ADNc de longitud completa, gen y la región promotora*

- 15 Las moléculas de ADNc de longitud completa que comprenden los polinucleótidos revelados se obtienen de la siguiente manera. Se utiliza un polinucleótido que comprende una secuencia de SEQ ID NO: 1 o una parte de la misma que comprende al menos 12, 15, 18 o 20 nt como sonda de hibridación para detectar los miembros hibridantes de una genoteca de ADNc usando procedimientos de diseño de sondas, procedimientos de clonación y técnicas de selección de clones tales como las descritas en el documento USPN 5.654.173. Las genotecas de ADNc se elaboran a partir de tejidos seleccionados, tales como tejido normal o tumoral, o de tejidos de un mamífero tratado con, por ejemplo, un agente farmacéutico. Preferiblemente, el tejido es el mismo que el tejido del que se han aislado los polinucleótidos de la invención, como que tanto los polinucleótidos descritos en la presente memoria como el ADNc representen genes expresados. Lo más preferible es que la genoteca de ADNc se elabore a partir del material biológico descrito en la presente memoria en los ejemplos. La selección del tipo de célula para la construcción de la genoteca se puede hacer tras conocer la identidad de la proteína codificada por el gen correspondiente al polinucleótido de la invención. Esto indicará qué tejido y qué tipos de células tienen tendencia a expresar el gen relacionado y, por consiguiente, representan una fuente adecuada para el ARNm destinado a generar el ADNc. Cuando los polinucleótidos proporcionados se aíslan de genotecas de ADNc, las genotecas se preparan a partir de ARNm de células de colon humano, más preferiblemente, de células de cáncer de colon humano, cuyas células se pueden obtener de tejido del paciente o pueden ser una línea celular de colon, p. ej., la Km12L4-A.

- Las técnicas para producir y sondear las genotecas de secuencias de ácidos nucleicos se describen, por ejemplo en Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", 2ª Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. El ADNc se puede preparar usando cebadores basados en la secuencia de SEQ ID NO: 1. En una realización, la genoteca de ADNc se puede fabricar únicamente a partir de ARNm poli-adenilado. De este modo, se pueden usar cebadores poli-T para preparar ADNc del ARNm.

- Se obtienen miembros de la genoteca que son más largos que los polinucleótidos proporcionados, y preferiblemente, que engloban la secuencia codificante completa del mensaje nativo. Para confirmar que se ha obtenido todo el ADNc, se realizan experimentos de protección del ARN según lo que se explica a continuación. La hibridación de un ADNc de longitud completa con un ARNm protegerá al ARN de la degradación por RNasa. Si el ADNc no es de longitud completa, entonces las partes del ARNm que no se han hibridado serán sometidas a la degradación por RNasa. Esto se analiza, según lo conocido en la técnica, mediante 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., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. Para obtener más secuencias 5' hasta el final de un ADNc parcial, se puede realizar una RACE de 5' ("PCR Protocols: A Guide to Methods and Applications", (1990) Academic Press, Inc).

- El ADN genómico se aísla usando los polinucleótidos proporcionados de una manera similar al aislamiento de los ADNc de longitud completa. En resumen, los polinucleótidos proporcionados, o las porciones de los mismos, se usan como sondas para genotecas de ADN genómico. Preferiblemente, la genoteca se obtiene del tipo de célula que se usó para generar los polinucleótidos de la invención, pero esto no es imprescindible. Lo más preferible es que el ADN genómico se obtenga del material biológico descrito en la presente memoria, en los ejemplos. Tales genotecas pueden estar en vectores adecuados para transportar segmentos largos de un genoma, tales como P1 o YAC, según lo descrito detalladamente en Sambrook *et al.*, 9.4-9.30. Además las secuencias genómicas se pueden aislar de genotecas de BAC humanos, que se encuentran disponibles comercialmente en Research Genetics, Inc., Huntsville, Alabama, EE.UU., por ejemplo. Para obtener más secuencias 5' ó 3', se realiza un paseo cromosómico, según lo descrito en Sambrook *et al.*, tal que se aíslan fragmentos adyacentes y solapantes del ADN genómico. Éstos son mapeados y reconstruidos, como se conoce en la técnica, usando la digestión de enzimas de restricción y ADN ligasa.

- Usando las secuencias de polinucleótidos de la invención, se pueden aislar los genes de longitud completa correspondientes usando procedimientos tanto clásicos como de PCR para construir y sondear genotecas de ADNc. Usando cualquier procedimiento, se realizan preferiblemente transferencias Northern en un número de tipos de células para determinar qué líneas celulares expresan el gen de interés al nivel más elevado. Los procedimientos clásicos para construir genotecas de ADNc se enseñan en Sambrook *et al.*, *supra*. Con estos procedimientos, se puede producir ADNc de ARNm e insertarlo en vectores víricos o de expresión. Comúnmente, las genotecas de ARNm que comprenden colas poli(A) se pueden producir con cebadores poli(T). De manera similar, las genotecas de ADNc se pueden producir usando las presentes secuencias como cebadores.

Los procedimientos de PCR se usan para amplificar los miembros de una genoteca de ADNc que comprende el inserto deseado. En este caso, el inserto deseado contendrá la secuencia del ADNc de longitud completa que se corresponde con los presentes polinucleótidos. Tales procedimientos PCR incluyen el atrapado de genes y los procedimientos de RACE. El atrapado de genes implica insertar un miembro de una genoteca de ADNc en un vector. Luego se desnaturaliza el vector para producir moléculas monocatenarias. A continuación, se usa una sonda unida a un sustrato, tal como un oligo bitinilado, para atrapar a los insertos de ADNc de interés. Las sondas biotiniladas se pueden enlazar a un sustrato sólido unido a avidina. Se pueden usar procedimientos de PCR para amplificar el ADNc atrapado. Para atrapar secuencias correspondientes a los genes de longitud completa, la secuencia de sonda marcada está basada en las secuencias de polinucleótidos de la invención. Los cebadores aleatorios o los cebadores específicos del vector de la genoteca se pueden usar para amplificar el ADNc atrapado. En Gruber *et al.*, documento WO 95/04745 y Gruber *et al.*, documento USPN 5.500.356 se describen tales técnicas de atrapado de genes. Hay equipos comercialmente disponibles para desarrollar los experimentos de atrapado de genes, por ejemplo, en Life Technologies, Gaithersburg, Maryland, EE.UU.

“La amplificación rápida de extremos de ADNc” o la RACE es un procedimiento de PCR de amplificación de ADNc a partir de un número de ARN diferentes. Los ADNc son ligados a un ligador de oligonucleótidos y amplificados por PCR usando dos cebadores. Un cebador se basa en la secuencia de los presentes polinucleótidos, para la que se desea una secuencia de longitud completa, y un segundo cebador comprende la secuencia que se hibrida con el ligador de oligonucleótidos para amplificar el ADNc. En el documento WO 97/19110, se informa de una descripción de estos procedimientos. En realizaciones preferidas de la RACE, se diseña un cebador común para aparearlo a una secuencia adaptadora arbitraria ligada a extremos de ADNc (Apte y Siebert, *Biotechniques* (1993) 15: 890-893; Edwards *et al.*, *Nuc. Acids Res.* (1991) 19: 5227-5232). Cuando se aparea un cebador de RACE específico de un único gen con el cebador común, se produce una amplificación preferencial de las secuencias situadas entre el cebador específico del único gen y el cebador común. Hay disponibles mezclas de ADNc comerciales para su uso en la RACE.

Otro procedimiento basado en la PCR genera una genoteca de ADNc de longitud completa con extremos anclados sin la necesidad de tener un conocimiento específico de la secuencia de ADNc. El procedimiento usa cebadores de anclaje (I-VI), en el que un cebador, poliTV (I-III) se ancla a la cola poli(A) del ARNm eucariota produciendo una síntesis de la primera cadena, y un segundo cebador, el poli(GH) (IV-VI) se ancla a la cola poli(C) añadida por la desoxinucleotidil transferasa terminal (TdT) (Véase, p. ej., el documento WO 96/40998).

La región promotora de un gen se localiza generalmente en 5' con respecto al sitio de inicio para la ARN polimerasa II. Cientos de regiones promotoras contienen la caja “TATA”, una secuencia tal como TATTA o TATAA, que es sensible a las mutaciones. La región promotora se puede obtener realizando una RACE de 5' usando un cebador de la región codificante del gen. Alternativamente, el ADNc se puede usar como una sonda para la secuencia genómica, identificándose la región 5' con respecto a la región codificante mediante un “paseo secuencia arriba”. Si el gen está altamente expresado o expresado diferencialmente, el promotor del gen puede ser útil en un constructo regulador para un gen heterólogo.

Una vez obtenido el ADNc o el gen de longitud completa, se pueden preparar las variantes codificantes del ADN mediante una mutagénesis de sitio dirigida, descrita detalladamente en Sambrook *et al.*, 15.3-15.63. La elección del codón o del nucleótido por reemplazar se puede basar en la revelación de la presente memoria sobre los cambios opcionales en aminoácidos para conseguir una estructura y/o una función proteica alterada.

Como procedimiento alternativo para obtener ADN o ARN de un material biológico, se puede sintetizar ácido nucleico que comprenda nucleótidos que tengan la secuencia de uno o más polinucleótidos de la invención. De este modo, la invención engloba moléculas de ácido nucleico que varían en longitud de 15 nt (correspondiente a al menos 15 nt contiguos de SEQ ID NO: 1) hasta una longitud máxima adecuada para una o más manipulaciones biológicas, incluyendo la replicación y la 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 la SEQ ID NO: 1; (b) el ácido nucleico de (a) que también comprende al menos un gen más, ligado operativamente 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 vírica recombinante que comprende (a) o (b). Una vez proporcionados los polinucleótidos revelados en la presente memoria, la construcción o la preparación de (a)-(e) son conocidas en la técnica.

La secuencia de un ácido nucleico que comprende 15 nt contiguos de la SEQ ID NO: 1, preferiblemente, la secuencia completa de SEQ ID NO: 1, no se limita 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 elección de la secuencia dependerá de la función deseada y puede estar establecida por las regiones codificantes deseadas, las regiones de tipo intrón deseados y las regiones reguladoras deseadas. Cuando la secuencia completa de SEQ ID NO: 1 está dentro del ácido nucleico, el ácido nucleico obtenido es denominado en la presente memoria como un polinucleótido que comprende la secuencia de SEQ ID NO: 1.

#### Expresión de polipéptido codificado por ADNc de longitud completa o gen de longitud completa

Los polinucleótidos proporcionados (p. ej., un polinucleótido que comprende una secuencia de SEQ ID NO: 1), el ADNc correspondiente o el gen de longitud completa se usan para expresar un producto génico parcial o completo. Los constructos de polinucleótidos que comprenden la secuencia SEQ ID NO: 1 también se pueden generar sintéti-

camente. Alternativamente, p. ej., Stemmer *et al.*, Gene (Amsterdam) (1995) 164(1): 49-53 revelan un ensamblaje de una sola etapa de un gen y un plásmido entero procedente de un gran número de oligodesoxirribonucleótidos. En este procedimiento, se describe la PCR de ensamblaje (la síntesis de las secuencias de ADN largas de un gran número de oligodesoxirribonucleótidos (oligos)). El procedimiento deriva de una reorganización del ADN (Stemmer, *Nature* (1994) 370: 389-391), y no se basa en la ADN ligasa, sino que se basa en la ADN polimerasa para construir fragmentos de ADN cada vez más largos durante el procedimiento de ensamblaje.

Los constructos de polinucleótido apropiados se purifican usando técnicas de ADN recombinante estándar según lo descrito en, por ejemplo, Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", 2ª Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY., y en virtud de las regulaciones actuales descritas en las Directrices del Instituto Nacional de Sanidad (NIH) estadounidense para la investigación sobre el ADN recombinante. El producto génico codificado por un polinucleótido de la invención se expresa en cualquier sistema de expresión, incluyendo, por ejemplo, sistemas bacterianos, de levaduras, de insectos, de anfibios y mamíferos. Los vectores, las células huésped y los procedimientos para obtener la expresión en los mismos son conocidos en la técnica. En el documento USPN 5.654.173, se describen vectores y células huésped adecuados.

Las moléculas de polinucleótido que comprenden una secuencia de polinucleótidos proporcionada en la presente memoria se propagan generalmente mediante la colocación de la molécula en un vector. Se usan vectores víricos y no víricos, incluyendo plásmidos. La elección del plásmido dependerá del tipo de célula en la que se desea la propagación y del propósito de la propagación. Determinados vectores son útiles para amplificar y fabricar grandes cantidades de la secuencia de ADN deseada. Otros vectores son adecuados para la expresión en células en cultivo. Otros vectores más son adecuados para la transferencia a y la expresión en células de un animal entero o una persona. La elección del vector apropiado pertenece al alcance de la técnica. Hay muchos vectores tales disponibles comercialmente. Los procedimientos para la preparación de vectores que comprenden una secuencia deseada son conocidos en la técnica.

Los polinucleótidos expuestos en la SEQ ID NO: 1 o sus polinucleótidos correspondientes de longitud completa están ligados a secuencias reguladoras según lo apropiado para obtener las propiedades de expresión deseadas. Éstos pueden incluir promotores (unidos bien en el extremo 5' de la cadena sentido o en el extremo 3' de la cadena antisentido), potenciadores, terminadores, operadores, represores e inductores. Los promotores pueden ser regulados o constitutivos. En algunas situaciones, puede ser deseable usar promotores condicionalmente activos, tales como promotores específicos del tejido o específicos de la etapa de desarrollo. Éstos son ligados a la secuencia de nucleótidos deseada usando las técnicas descritas anteriormente para la unión a vectores. Se puede usar cualquier técnica conocida en la técnica.

Cuando cualquiera de las células huésped anteriores u otras células huésped u organismos se usan para replicar y/o expresar los polinucleótidos o los ácidos nucleicos de la invención, el ácido nucleico replicado resultante, el ARN, la proteína expresada o el polipéptido pertenece al ámbito de la invención como un producto de la célula huésped o el organismo. El producto se recupera mediante cualquier procedimiento apropiado conocido en la técnica.

Una vez identificado el gen correspondiente a un polinucleótido seleccionado, se puede regular su expresión en la célula de la que el gen es nativo. Es posible, por ejemplo, regular un gen endógeno de una célula mediante una secuencia reguladora exógena según lo revelado en el documento USPN 5.641.670.

#### *Identificación de motivos funcionales y estructurales del rastreo de genes frente a las bases de datos a disposición del público*

Es posible alinear las traducciones de la secuencia de nucleótidos de los polinucleótidos proporcionados, de los ADNc o de los genes completos con secuencias conocidas individuales. Se puede usar la similitud con secuencias individuales para determinar la actividad de los polipéptidos codificados por los polinucleótidos de la invención. Además, las secuencias que presentan una similitud con más de una secuencia individual pueden presentar actividades que son características de cualquiera o de ambas secuencias individuales.

Se pueden usar las secuencias de longitud completa y los fragmentos de las secuencias de polinucleótidos de los vecinos más cercanos como sondas y cebadores para identificar y aislar la secuencia de longitud completa correspondiente a los polinucleótidos proporcionados. Los vecinos más cercanos pueden indicar un tejido o un tipo de célula por usar para construir una genoteca para las secuencias de longitud completa correspondientes a los polinucleótidos proporcionados.

Comúnmente, un polinucleótido seleccionado se traduce en el total de los seis marcos para determinar el mejor alineamiento con las secuencias individuales. Las secuencias reveladas en la presente memoria en el listado secuencial están en sentido de 5' a 3', y la traducción en tres marcos puede ser suficiente (con unas cuantas excepciones específicas según lo descrito en los ejemplos). Estas secuencias de aminoácidos son denominadas generalmente secuencias problema, que serán alineadas con las secuencias individuales. En "Computer Methods for Macromolecular Sequence Analysis Methods in Enzymology" (1996) 266, Doolittle, Academic Press, Inc., una filial de Harcourt Brace & Co., San Diego, California, EE.UU., se describen bases de datos con las secuencias individuales. Las bases de datos incluyen la GenBank, la EMBL y la base de datos de ADN de Japón (DDBJ).

Las secuencias problema y las secuencias individuales se pueden alinear usando los procedimientos y los programas informáticos descritos anteriormente, e incluyen el BLAST 2.0., disponible a nivel mundial en <http://www.ncbi.nlm.nih.gov/BLAST/>. Véase también Altschul, *et al. Nucleic Acids Res.* (1997) 25: 3389-3402. Otro algoritmo de alineamiento es el Fasta, disponible en Genetics Computing Group (GCG) package, Madison, Wisconsin, EE.UU., una filial de entera propiedad de Oxford Molecular Group, Inc. En Doolittle, *supra*, se describen otras técnicas para el alineamiento. Preferiblemente, se utiliza un programa de alineamientos que permita huecos en la secuencia para alinear las secuencias. El Smith-Waterman es un tipo de algoritmo que permite huecos en los alineamientos secuenciales. Véase *Meth. Mol. Biol.* (1997) 70:173-187. También, se puede utilizar para alinear secuencias el programa GAP que usa el procedimiento de alineamiento de Needleman y Wunsch.

Una estrategia de búsqueda alternativa usa el programa MPSRCH que funciona en un ordenador MASPARE. El MPSRCH usa un algoritmo de Smith-Waterman para puntuar las secuencias en un ordenador enormemente paralelo. Este enfoque mejora la capacidad para identificar secuencias que son apareamientos relacionados distantemente, y es especialmente tolerante a pequeños huecos y errores de la secuencia de nucleótidos. Las secuencias de ácidos nucleicos codificadas por los polinucleótidos proporcionados se pueden usar para buscar en bases de datos tanto de proteínas como de ADN.

Los resultados de los alineamientos de las secuencias individuales y las secuencias problema se pueden dividir en tres categorías: similitud elevada, similitud débil y ninguna similitud. Los resultados de los alineamientos individuales que varían de similitud elevada a similitud débil proporcionan una base para determinar la actividad y/o la estructura de los polipéptidos. Los parámetros para categorizar los resultados individuales incluyen: el porcentaje de la longitud de la región de alineamiento en la que se encuentra el mayor alineamiento, el porcentaje de identidad secuencial y el valor p. El porcentaje de la longitud de la región de alineamiento se calcula contando el número de residuos de la secuencia individual encontrados en la región de mayor alineamiento, p. ej., la región contigua a la secuencia individual que contiene el mayor número de residuos que son idénticos a los residuos de la región de la secuencia problema alineada correspondiente. Este número se divide entre la longitud de residuos total de la secuencia problema para calcular un porcentaje. Por ejemplo, se podría alinear una secuencia problema de 20 residuos de aminoácido con una región de 20 aminoácidos de una secuencia individual. La secuencia individual podría ser idéntica a los residuos de aminoácido 5, 9-15 y 17-19 de la secuencia problema. La región de mayor alineamiento es, de este modo, la región que va del residuo 9-19, un tramo de 11 aminoácidos. El porcentaje de la longitud de la región de alineamiento es: 11 (longitud de la región de mayor alineamiento) dividido entre (la longitud de la secuencia problema) 20 ó 55%.

El porcentaje de identidad secuencial se calcula contando el número de apareamientos de aminoácidos entre la secuencia problema y la secuencia individual, y dividiendo el número total de apareamientos entre el número de residuos de las secuencias individuales encontradas en la región de mayor alineamiento. De este modo, el porcentaje de identidad del ejemplo anterior sería de 10 apareamientos dividido entre 11 aminoácidos, o aproximadamente, del 90,9%.

El valor p es la probabilidad de que se produzca el alineamiento por casualidad. Se puede calcular el valor p para un único alineamiento según Karlin *et al.*, *Proc. Natl. Acad. Sci.* (1990) 87:2264 y Karlin *et al.*, *Proc. Natl. Acad. Sci.* (1993) 90. El valor p de múltiples alineamientos usando la misma secuencia problema se puede calcular usando un enfoque heurístico descrito en Altschul *et al.*, *Nat. Genet.* (1994) 6:119. Los programas de alineamiento tales como el programa BLAST pueden calcular el valor p. Véase también Altschul, *et al. Nucleic Acids Res.* (1997) 25: 3389-3402.

Otro factor a considerar para determinar la identidad o la similitud es la localización de la similitud o de la identidad. Un alineamiento local elevado puede indicar una similitud incluso si la longitud del alineamiento es corta. La identidad secuencial dispersada por la longitud de la secuencia problema también puede indicar una similitud entre las secuencias problema y perfil. Los límites de la región en la que se alinean las secuencias se pueden determinar según Doolittle, *supra*; BLAST 2.0 (véase, p. ej., Altschul, *et al. Nucleic Acids Res.* (1997) 25: 3389-3402) o programas FAST; o mediante la determinación de la zona en la que la identidad secuencial es más elevada.

**Similitud elevada:** En general, en los resultados de alineamiento considerados de similitud elevada, el porcentaje de la longitud de la región de alineamiento es comúnmente del al menos aproximadamente 55% de la secuencia problema de longitud total; más comúnmente, del al menos aproximadamente 58%; incluso más comúnmente, del al menos aproximadamente 60% de la longitud total de los residuos de la secuencia problema. Habitualmente, el porcentaje de longitud de la región de alineamiento puede ser de tanto como del aproximadamente 62%; más habitualmente, de tanto como del aproximadamente 64%; todavía más habitualmente, de tanto como del aproximadamente 66%. Además, para una similitud elevada, la región de alineamiento presenta comúnmente al menos aproximadamente el 55% de identidad secuencial; más comúnmente, al menos aproximadamente el 78%; incluso más comúnmente, al menos aproximadamente el 80% de identidad secuencial. Habitualmente, el porcentaje de identidad secuencial puede ser de tanto como del aproximadamente 82%; más habitualmente, de tanto como del aproximadamente 84%; todavía más habitualmente, de tanto como del aproximadamente 86%.

El valor p se usa en combinación con estos procedimientos. Si se encuentra una similitud elevada, la secuencia problema se considera que tiene una similitud elevada con una secuencia perfil cuando el valor p es menor de o igual a aproximadamente  $10^{-2}$ ; más habitualmente, menor de o igual a aproximadamente  $10^{-3}$ ; todavía más habitualmente, menor de o igual a aproximadamente  $10^{-4}$ . Más comúnmente, el valor p no es mayor de aproximadamente  $10^{-5}$ ; más comúnmente, es menor de o igual a aproximadamente  $10^{-10}$ ; todavía más comúnmente, menor de o igual a aproximadamente  $10^{-15}$  para considerar a la secuencia problema con una similitud elevada.

*Similitud débil.* En general, cuando los resultados del alineamiento se consideran de una similitud débil, no hay un porcentaje mínimo de longitud de la región de alineamiento ni una longitud de alineamiento mínima. La mejor muestra de una similitud débil se considera cuando la región de alineamiento es, comúnmente, de una longitud de al menos aproximadamente 15 residuos de aminoácido; más comúnmente, de al menos aproximadamente 20; todavía más comúnmente de una longitud de al menos aproximadamente 25 residuos de aminoácido. Habitualmente, la longitud de la región de alineamiento puede ser de tanto como de aproximadamente 30 residuos de aminoácido; más habitualmente, de tanto como de aproximadamente 40; todavía más habitualmente, de tanto como de aproximadamente 60 residuos de aminoácido. Además, para una similitud débil, la región de alineamiento presenta comúnmente al menos aproximadamente el 35% de identidad secuencial; más comúnmente, al menos aproximadamente el 40%; incluso más comúnmente, al menos aproximadamente el 45% de identidad secuencial. Habitualmente, el porcentaje de identidad secuencial puede ser de tanto como del aproximadamente 50%; más habitualmente, de tanto como del aproximadamente 55%; todavía más habitualmente, de tanto como del aproximadamente 60%.

Si se encuentra una similitud baja, la secuencia problema se considera que tiene una similitud baja con una secuencia perfil cuando el valor  $p$  es habitualmente menor de o igual a aproximadamente  $10^{-2}$ ; más habitualmente, menor de o igual a aproximadamente  $10^{-3}$ ; todavía más habitualmente, menor de o igual a aproximadamente  $10^{-4}$ . Más comúnmente, el valor  $p$  no es mayor de aproximadamente  $10^{-5}$ ; más habitualmente, menor de o igual a aproximadamente  $10^{-10}$ ; todavía más habitualmente, menor de o igual a aproximadamente  $10^{-15}$  para considerar a la secuencia problema con una similitud débil.

*Similitud determinada sólo por la identidad secuencial.* La identidad secuencial sola se puede usar para determinar la similitud de una secuencia problema con una secuencia individual, y puede indicar la actividad de la secuencia. Tal alineamiento, preferiblemente, permite huecos para alinear las secuencias. Comúnmente, la secuencia problema está relacionada con la secuencia perfil si la similitud secuencial sobre la secuencia problema completa es del al menos aproximadamente 15%; más comúnmente, del al menos aproximadamente 20%; incluso más comúnmente, del al menos aproximadamente 25%; incluso más comúnmente, del al menos aproximadamente 50%. La identidad secuencial sola como medida de la similitud es más útil cuando la secuencia problema es habitualmente de al menos 80 residuos de longitud; más habitualmente, de 90 residuos; todavía más habitualmente de una longitud de al menos 95 residuos de aminoácidos. Más comúnmente, la similitud se puede concluir sólo en base a la identidad secuencial cuando la secuencia problema es preferiblemente de 100 residuos de longitud; más preferiblemente, de 120 residuos de longitud; todavía más preferiblemente de una longitud de 150 residuos de aminoácidos.

*Alineamientos con secuencias perfil y múltiples secuencias alineadas.* Las traducciones de los polinucleótidos proporcionados se pueden alinear con perfiles de aminoácidos que definen bien familias de proteínas o motivos comunes. Además, las traducciones de los polinucleótidos proporcionados se pueden alinear con alineamientos de múltiples secuencias (AMS) que comprenden las secuencias de polipéptidos de los miembros de las familias de proteínas o de los motivos. La similitud o la identidad con las secuencias perfil o los AMS se puede usar para determinar la actividad de los productos génicos (p. ej., los polipéptidos) codificados por los polinucleótidos proporcionados o el ADNc o los genes correspondientes. Por ejemplo, las secuencias que presentan una identidad o una similitud con un perfil o AMS de quimioquina pueden presentar actividades de la quimioquina.

Los perfiles se pueden diseñar manualmente (1) creando un AMS, que es un alineamiento de la secuencia de aminoácidos de los miembros que pertenecen a la familia y (2) construyendo una representación estadística del alineamiento. Tales procedimientos se describen, por ejemplo, en Birney *et al.*, *Nucl. Acid Res.* (1996) 24 (14): 2730-2739. Los AMS de algunas familias de proteínas y algunos motivos están a disposición del público. Por ejemplo, <http://genome.wustl.edu/Pfam/> incluye AMS de 547 familias y motivos diferentes. Estos AMS se describen también en Sonnhammer *et al.*, *Proteins* (1997) 28: 405-420. Otras fuentes que se encuentran en la red incluyen el sitio <http://www.embl-heidelberg.de/argos/ali/ali.html>; alternativamente, se puede enviar un mensaje a [ALI@EMBL-HEIDELBERG.DE](mailto:ALI@EMBL-HEIDELBERG.DE) para solicitar información. En Pascarella *et al.*, *Prot. Eng* (1996) 9 (3): 249-251, se informa de una breve descripción de estos AMS. En Sonnhammer *et al.*, *supra*; Birney *et al.*, *supra*; y "Computer Methods for Macromolecular Sequence Analysis Methods in Enzymology" (1996) 266, Doolittle, Academic Press, Inc., San Diego, California, EE.UU., se describen técnicas para su construcción a partir de AMS.

La similitud entre una secuencia problema y una familia de proteínas o un motivo se puede determinar (a) comparando la secuencia problema frente al perfil y/o (b) alineando la secuencia problema con los miembros de la familia o el motivo. Comúnmente, se usa un programa tal como el Searchwise para comparar la secuencia problema con la representación estadística del alineamiento múltiple, también conocido como perfil (véase Birney *et al.*, *supra*). En Sonnhammer *et al.* *supra* y Doolittle, *supra*, se describen otras técnicas para comparar la secuencia y el perfil.

A continuación se pueden usar los procedimientos descritos por Feng *et al.*, *J. Mol. Evol.* (1987) 25: 351 y Higgins *et al.*, *CABIOS* (1989) 5: 151 para alinear la secuencia problema con los miembros de una familia o motivo, también conocido como AMS. Los alineamientos de secuencias se pueden generar usando cualquier variedad de herramientas informáticas. Los ejemplos incluyen PileUp, que crea un alineamiento de múltiples secuencias y se describe en Feng *et al.*, *J. Mol. Evol.* (1987) 25: 351. Otro procedimiento, el GAP, usa el procedimiento de alineamiento de Needleman *et al.*, *J. Mol. Biol.* (1970) 48: 443. El GAP es el que está mejor adaptado al alineamiento global de las secuencias. Un tercer procedimiento, el BestFit, funciona insertando huecos para maximizar el número de apareamientos usando el algoritmo de homología local de Smith *et al.*, *Adv. Appl. Math.* (1981) 2: 482. En general, se usan los siguientes

factores para determinar si existe una similitud entre una secuencia problema y un perfil o un AMS: (1) el número de residuos conservados encontrado en la secuencia problema, (2) el porcentaje de residuos conservados encontrado en la secuencia problema, (3) el número de desplazamiento de marcos y (4) el espacio entre los residuos conservados.

Algunos programas de alineamiento que tanto traducen como alinean secuencias pueden hacer cualquier número de desplazamiento de marcos cuando traducen la secuencia de nucleótidos para producir el mejor alineamiento. Cuando menor es el número de desplazamientos de marco necesarios para producir un alineamiento, mayor es la similitud o la identidad entre la secuencia problema y la perfil o el AMS. Por ejemplo, una similitud débil resultante de una carencia de desplazamientos de marcos puede ser una mejor indicación de la actividad o de la estructura de una secuencia problema que una similitud elevada resultado de dos desplazamientos de marco. Es preferible encontrar tres o menos desplazamientos de marco en el alineamiento; más preferiblemente, dos o menos desplazamientos de marco; todavía más preferiblemente, uno o menos desplazamientos de marco; incluso más preferiblemente, ningún desplazamiento de marco en un alineamiento de la secuencia problema y la perfil o los MAS.

Los residuos conservados son aquellos aminoácidos encontrados en una determinada posición en todos o en parte de los miembros de la familia o el motivo. Alternativamente, se considera que una posición está conservada si sólo se encuentra una cierta clase de aminoácidos en una determinada posición en todos o en parte de los miembros de la familia. Por ejemplo, la posición N-terminal puede contener un aminoácido cargado positivamente tal como lisina, arginina o histidina.

Comúnmente, un residuo de un polipéptido está conservado cuando se encuentra una clase de aminoácidos o un único aminoácido en una determinada posición en al menos el aproximadamente 40% de todos los miembros de la clase; más comúnmente, en al menos aproximadamente el 50%; incluso más comúnmente, en al menos aproximadamente el 60% de los miembros. Habitualmente, el residuo está conservado cuando una clase o un único aminoácido se encuentra en al menos el aproximadamente 70% de los miembros de una familia o un motivo; más habitualmente, en al menos aproximadamente el 80%; incluso más habitualmente, en al menos aproximadamente el 90%; incluso más habitualmente, en al menos aproximadamente el 95%.

Un residuo se considera conservado cuando se encuentran tres aminoácidos no relacionados en una determinada posición en todos o en parte de los miembros; más habitualmente, dos aminoácidos no relacionados. Estos residuos están conservados cuando los residuos no relacionados se encuentran en determinadas posiciones en al menos el aproximadamente 40% de todos los miembros de la clase; más comúnmente, en al menos aproximadamente el 50%; incluso más comúnmente, en al menos aproximadamente el 60% de los miembros. Habitualmente, el residuo está conservado cuando una clase o un único aminoácido se encuentra en al menos el aproximadamente 70% de los miembros de una familia o un motivo; más habitualmente, en al menos aproximadamente el 80%; incluso más habitualmente, en al menos aproximadamente el 90%; incluso más habitualmente, en al menos aproximadamente el 95%.

Una secuencia problema tiene similitud con un perfil o un AMS cuando la secuencia problema comprende al menos el aproximadamente 25% de los residuos conservados del perfil o del AMS; más habitualmente, al menos aproximadamente el 30%; incluso más habitualmente, al menos aproximadamente el 40%. Comúnmente, la secuencia problema tiene mayor similitud con una secuencia perfil o un AMS cuando la secuencia problema comprende al menos el aproximadamente 45% de los residuos conservados del perfil o del AMS; más comúnmente, al menos aproximadamente el 50%; incluso más comúnmente, al menos aproximadamente el 55%.

#### *Identificación de polipéptidos secretados y unidos a la membrana*

Tanto los polipéptidos secretados como los unidos a la membrana de la presente invención son de particular interés. Por ejemplo, se pueden analizar los niveles de polipéptidos secretados en fluidos corporales que son convenientes, tales como la sangre, el plasma, el suero y otros fluidos corporales tales como la orina, el fluido prostático y el semen. Los polipéptidos unidos a la membrana son útiles para construir antígenos de vacunas e inducir una respuesta inmune. Tales antígenos comprenderían toda o parte de la región extracelular de los polipéptidos unidos a la membrana. Como tanto los polipéptidos secretados como los unidos a la membrana comprenden un fragmento de aminoácidos hidrófobos contiguos, se pueden usar algoritmos de predicción de la hidrofobicidad para identificar tales polipéptidos.

Las secuencias señal están habitualmente codificadas tanto por genes de polipéptidos secretados como de los unidos a la membrana para dirigir a un polipéptido hacia la superficie de la célula. La secuencia señal comprende habitualmente un tramo de residuos hidrófobos. Tales secuencias señal se pueden plegar en estructuras helicoidales. Los polipéptidos unidos a la membrana comprenden comúnmente 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 de un polipéptido se pueden identificar usando algoritmos informáticos. Tales algoritmos incluyen Hopp y Woods. *Proc. Natl. Acad. Sci. EE.UU.* (1981) 78: 3824-3828; Kyte y Doolittle, *J. Mol. Biol.* (1982) 157: 105-132; y el algoritmo RAOAR, Degli Esposti *et al.*, *Eur. J. Biochem.* (1990) 190: 207-219.

Otro procedimiento para identificar polipéptidos secretados y unidos a la membrana consiste en traducir los polinucleótidos de la invención en el total de los seis marcos y determinar si al menos hay presentes 8 aminoácidos hidrófobos contiguos. Aquellos polipéptidos traducidos con al menos 8; más comúnmente 10; incluso más común-

mente, 12 aminoácidos hidrófobos contiguos se consideran bien un polipéptido secretado putativo o uno enlazado a la membrana. Los aminoácidos hidrófobos incluyen alanina, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, treonina, triptófano, tirosina y valina.

5

#### *Identificación de la función de un producto de expresión de un gen de longitud completa*

Se pueden usar las ribozimas, los constructos antisentido y los mutantes negativos dominantes para determinar la función del producto de expresión de un gen correspondiente a un polinucleótido proporcionado en la presente memoria, y además se pueden usar en la inhibición de la producción de productos génicos funcionales codificados por un gen correspondiente a un polinucleótido descrito en la presente memoria. En el contexto de la caracterización funcional del producto génico codificado, el uso de antisentido, ribozimas y/o mutantes negativos dominantes es particularmente útil cuando el polinucleótido proporcionado no presenta una homología significativa ni sustancial con una secuencia codificante de un gen de función conocida.

15

Las moléculas antisentido y las ribozimas se pueden construir a partir de polinucleótidos sintéticos. Comúnmente, se usa el procedimiento de síntesis de oligonucleótidos con fosoramidita. Véase Beaucage *et al.*, *Tet. Lett.* (1981) 22: 1859 y el documento USPN 4.668.777. Hay dispositivos automatizados para la síntesis destinados a crear oligonucleótidos que usan esta química. Los ejemplos de tales dispositivos incluyen Biosearch 8600, Modelos 392 y 394 por Applied Biosystems, una filial de Perkin-Elmer Corp., Foster City, California, EE.UU.; y Expedite por Perceptive Biosystems, Framingham, Massachusetts, EE.UU. También se pueden producir ARN sintético, oligonucleótidos de análogos de fosfato y oligonucleótidos derivados químicamente, y se pueden unir de forma covalente con otras moléculas. Los oligonucleótidos de ARN se pueden sintetizar, por ejemplo, usando fosoramiditas de ARN. Se puede realizar este procedimiento en un sintetizador automático, tal como el de Applied Biosystems, Modelos 392 y 394, Foster City, California, EE.UU.

Los oligonucleótidos de fosforotioato también se pueden sintetizar para la construcción antisentido. Se puede usar un reactivo sulfurizante, tal como disulfuro de tetraetiltiuramio (TETD) en acetonitrilo para convertir el fosfito de cianoetil internucleótido en triéster de fosforotioato en 15 minutos a temperatura ambiente. El TETD reemplaza al reactivo de yodo, mientras que el resto de los reactivos usados para la química estándar de la fosoramidita permanecen invariables. Tal procedimiento de síntesis se puede automatizar usando los modelos 392 y 394 de Applied Biosystems, por ejemplo.

Se pueden sintetizar oligonucleótidos de hasta 200 nt, más comúnmente, de 100 nt, más comúnmente, de 50 nt; incluso más comúnmente, de 30 a 40 nt. Estos fragmentos sintéticos se pueden aparear y ligar entre sí para construir fragmentos más largos. Véase, por ejemplo, Sambrook *et al.*, *supra*. Los ARN catalíticos trans-escindidores (ribozimas) son moléculas de ARN que poseen actividad endorribonucleasa. Las ribozimas se diseñan específicamente para una diana concreta, y el mensaje de la diana debe contener una secuencia nucleotídica específica. Son diseñadas genéticamente para escindir cualquier especie de ARN con especificidad de sitio en los antecedentes del ARN celular. El evento de escisión convierte al ARNm en inestable y evita la expresión de la proteína. Lo más importante es que las ribozimas se pueden usar para inhibir la expresión de un gen de función desconocida a efectos de determinar su función en un contexto *in vitro* o *in vivo*, mediante la detección del efecto fenotípico. Un motivo de ribozima comúnmente usado es el de cabeza de martillo, para el que los requisitos de la secuencia sustrato son mínimos. En Usman *et al.*, *Current Opin. Struct. Biol.* (1996) 6: 527, se revela el diseño de la ribozima de cabeza de martillo, así como de los usos terapéuticos de las ribozimas. Los procedimientos para la producción de ribozimas, incluyendo los fragmentos de ribozima con estructura de horquilla, los procedimientos para aumentar la especificidad de las ribozimas y similares son conocidos en la técnica.

La región hibridante de la ribozima se puede modificar o se puede preparar como una estructura ramificada según lo descrito en Horn y Urdea, *Nucleic Acids Res.* (1989) 17: 6959. La estructura básica de las ribozimas también se puede alterar químicamente en modos que son familiares para aquellos expertos en la técnica, y se pueden administrar ribozimas sintetizadas químicamente como derivados de oligonucleótido sintéticos modificados por unidades monoméricas. En un contexto terapéutico, la administración mediada por liposomas de las ribozimas mejora la captación celular, según lo descrito en Birikh *et al.*, *Eur. J. Biochem.* (1997) 245: 1.

55

Los ácidos nucleicos antisentido se diseñan para que se unan específicamente a ARN, dando como resultado la formación de híbridos de ARN-ADN o de ARN-ARN, con una detención de la replicación del ADN, la transcripción inversa o la traducción del ARN mensajero. Los polinucleótidos antisentido basados en una secuencia de polinucleótidos seleccionada pueden interferir en la expresión del gen correspondiente. Los polinucleótidos antisentido se generan comúnmente dentro de la célula mediante la expresión desde constructos antisentido que contienen la cadena antisentido como la cadena transcrita. Los polinucleótidos antisentido basados en los polinucleótidos revelados se unirán y/o interferirán en la traducción del ARNm que comprende una secuencia complementaria al polinucleótido antisentido. Los productos de expresión de células control y células tratadas con el constructo antisentido se comparan para detectar el producto de proteína del gen correspondiente al polinucleótido sobre el que está basado el constructo antisentido. La proteína se aísla y se identifica usando procedimientos bioquímicos rutinarios.

Dada la extensa bibliografía de los antecedentes y la experiencia clínica en la terapia antisentido, cualquier experto en la técnica puede usar polinucleótidos seleccionados de la invención como otros posibles compuestos terapéuticos.

La elección del polinucleótido se puede acotar analizándolos primero en cuanto a la unión a regiones de “puntos calientes” del genoma de las células cancerosas de colon. Si se identifica un polinucleótido que se une a un “punto caliente”, el análisis del polinucleótido como un compuesto antisentido en las células de cáncer de colon correspondientes queda garantizado.

Como procedimiento alternativo para identificar la función del gen correspondiente a un polinucleótido revelado en la presente memoria, se generan fácilmente mutaciones negativas dominantes para las proteínas correspondientes que son activas como homomultímeros. Un polipéptido mutante interactuará con polipéptidos de tipo silvestre (elaborados a partir de otro alelo) y formará un multímero no funcional. De este modo, hay una mutación en un dominio de unión al sustrato, un dominio catalítico o un dominio de localización celular. Preferiblemente, el polipéptido mutante será sobre-producido. Se hacen mutaciones puntuales que tengan tal efecto. Además, la fusión de diferentes polipéptidos de diversas longitudes con el terminal de una proteína puede producir mutantes negativos dominantes. Hay disponibles estrategias generales para fabricar mutantes negativos dominantes (véase, p. ej., Herskowitz, *Nature* (1987) 329: 219). Tales técnicas se pueden usar para crear la pérdida de mutaciones de función, que son útiles para determinar la función de las proteínas.

#### *Polipéptidos y variantes de los mismos*

Los polipéptidos de la invención incluyen aquéllos codificados por los polinucleótidos revelados, así como los ácidos nucleicos que, en virtud de la degeneración del código genético, no son idénticos en secuencia con los polinucleótidos revelados. De este modo, la invención incluye dentro de su ámbito un polipéptido codificado por un polinucleótido que comprende la secuencia de SEQ ID NO: 1, o una variante de la misma. Los polipéptidos ejemplares codificados por un marco de lectura abierto de un polinucleótido descrito en la presente memoria incluyen la SEQ ID NO: 2.

En general, el término “polipéptido” como se usa en la presente memoria se refiere tanto al polipéptido de longitud completa codificado por el polinucleótido citado, como al polipéptido codificado por el gen representado por el polinucleótido citado, así como a las porciones o a los fragmentos de los mismos. “Polipéptidos” también incluye las variantes de las proteínas naturales, siendo tales variantes homólogas o sustancialmente similares a la proteína natural, y pudiendo originar de la misma o de distintas especies que la proteína natural (p. ej., la especie humana, murina o alguna otra especie que exprese de forma natural el polipéptido citado, habitualmente, una especie de mamífero). En general, los polipéptidos de las variantes tienen una secuencia que tiene al menos aproximadamente el 80%, habitualmente al menos aproximadamente el 90%, y más habitualmente al menos aproximadamente el 98% de identidad secuencial con un polipéptido expresado diferencialmente de la invención, medido con el BLAST 2.0 usando los parámetros descritos anteriormente. Los polipéptidos de las variantes pueden ser glicosilados de manera natural o no natural, i.e., el polipéptido tiene un patrón de glicosilación que difiere del patrón de glicosilación encontrado en la proteína natural correspondiente.

La invención también engloba homólogos de los polipéptidos revelados (o fragmentos de los mismos), estando los homólogos aislados de otras especies, i.e., de otras especies animales o vegetales, siendo tales homólogos, habitualmente, de una especie de mamífero, p. ej., de roedores, tales como ratones, ratas; animales domésticos, p. ej., caballo, vaca, perro, gato; y seres humanos. El término “homólogo” significa un polipéptido que tiene al menos aproximadamente el 35%, habitualmente al menos aproximadamente el 40%, y más habitualmente, al menos aproximadamente el 60% de identidad secuencial de los aminoácidos con una proteína expresada diferencialmente concreta según lo identificado anteriormente, siendo la identidad secuencial determinada usando el algoritmo del BLAST 2.0, con los parámetros descritos *supra*.

En general, los polipéptidos de la presente invención se proporcionan en un medio no natural, p. ej., se separan de su medio natural. En ciertas realizaciones, la presente proteína está presente en una composición que está enriquecida en la proteína en comparación con un control. Como tal, se proporciona el polipéptido purificado, queriendo significar purificado que la proteína está presente en una composición que está sustancialmente libre de polipéptidos no expresados diferencialmente, queriendo significar sustancialmente libre que menos del 90%, habitualmente, menos del 60% y, más habitualmente, menos del 50% de la composición está compuesta de polipéptidos no expresados diferencialmente.

También están dentro del ámbito de la invención las variantes; las variantes de los polipéptidos incluyen mutantes, fragmentos y fusiones. Los mutantes incluyen sustituciones, adiciones o eliminaciones de aminoácidos. Las sustituciones de aminoácidos pueden ser sustituciones de aminoácidos conservativos o sustituciones para eliminar aminoácidos no esenciales, tales como para modificar un sitio de glicosilación, un sitio de fosforilación o un sitio de acetilación, o para minimizar el plegamiento erróneo por la sustitución o la eliminación de uno o más residuos de cisteína que no son necesarios para la función. Las sustituciones de aminoácidos conservativos son aquéllas que conservan la carga general, la hidrofobicidad/hidrofiliadad y/o la carga estérica del aminoácido sustituido.

Se pueden diseñar las variantes para que conserven o tengan una mayor actividad biológica de una región concreta de la proteína (p. ej., un dominio funcional y/o cuando el polipéptido sea un miembro de una familia de proteínas, una región asociada a una secuencia consenso). La selección de alteraciones de aminoácidos para la producción de variantes se puede basar en la accesibilidad (interior frente a exterior) del aminoácido (véase, p. ej., Go *et al.*, *Int. J. Peptide Protein Res.* (1980) 15:211), la termoestabilidad del polipéptido de la variante (véase, p. ej., Querol *et al.*,



*Prot. Eng.* (1996) 9:265), los sitios de glicosilación deseados (véase, p. ej., Olsen y Thomsen, *J. Gen. Microbiol.* (1991) 137:579), los puentes disulfuro deseados (véase, p. ej., Clarke *et al.*, *Biochemistry* (1993) 32: 4322; y Wakarchuk *et al.*, *Protein Eng.* (1994) 7: 1379), los sitios de unión a metales (véase, p. ej., Toma *et al.*, *Biochemistry* (1991) 30:97, y Haezerebrouck *et al.*, *Protein Eng.* (1993) 6:643) y las sustituciones deseadas con bucles de prolina (véase, p. ej., Masul *et al.*, *Appl. Env. Microbiol.* (1994) 60:3579). Las muteínas reducidas por cisteína se pueden producir según lo revelado en el documento USPN 4.959.314.

Las variantes también incluyen fragmentos de los polipéptidos revelados en la presente memoria, particularmente, de fragmentos biológicamente activos y/o de fragmentos correspondientes a los dominios funcionales. Los fragmentos de interés serán comúnmente de al menos aproximadamente 10 aa a al menos aproximadamente 15 aa de longitud; habitualmente, de al menos aproximadamente 50 aa de longitud, y pueden ser tan largos como de 300 aa de longitud o más largos, pero habitualmente no superarán los aproximadamente 1.000 aa de longitud, teniendo el fragmento un tramo de aminoácidos que es idéntico a un polipéptido codificado por un polinucleótido que comprende una secuencia de SEQ ID NO: 1, o un homólogo de la misma. Las variantes de proteínas descritas en la presente memoria están codificadas por polinucleótidos que pertenecen al ámbito de la invención. Se puede usar el código genético para seleccionar los codones apropiados para construir las correspondientes variantes.

#### *Realizaciones relacionadas con la informática*

En general, una genoteca de polinucleótidos es una colección de información secuencial, cuya información se proporciona en forma bien bioquímica (p. ej., como una colección de moléculas de polinucleótido) o en forma electrónica (p. ej., como una colección de secuencias de polinucleótido almacenadas en una forma de lectura informática, como en un sistema informático y/o como parte de un programa informático). La secuencia de información de los polinucleótidos se puede usar en una variedad de formas, p. ej., como una fuente de descubrimiento de genes, como una representación de secuencias expresadas en un tipo de célula seleccionado (p. ej., marcadores de tipo celular) y/o como marcadores de una enfermedad o de un estado patológico dado. 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 a un nivel mayor o menor en comparación con una célula normal (p. ej., una célula del mismo tipo o de un tipo similar que no esté sustancialmente afectada por la enfermedad). Por ejemplo, una secuencia de polinucleótidos de una genoteca puede ser un polinucleótido que represente un ARNm, un polipéptido u otro producto génico codificado por el polinucleótido, que esté bien sobre-expresado o infra-expresado en una célula de colon afectada por cáncer en comparación con una célula de colon normal (i.e., sustancialmente libre de enfermedad).

La información de las secuencias de nucleótidos de la genoteca se puede realizar en cualquier forma adecuada, p. ej., formas electrónicas o bioquímicas. Por ejemplo, una genoteca de información secuencial realizada en forma electrónica comprende un archivo informático de datos accesible (o, en forma bioquímica, una colección de moléculas de ácido nucleico) que contiene las secuencias de nucleótidos representativas de los genes que están expresados diferencialmente (p. ej., sobre-expresados o infra-expresados) como entre, por ejemplo, i) una célula de colon cancerosa y una célula de colon normal; ii) una célula de colon cancerosa y una célula de colon displásica; iii) una célula de colon cancerosa y una célula de colon afectado por una enfermedad o una condición distinta del cáncer; iv) una célula de colon cancerosa metastásica y una célula de colon normal y/o una célula de colon cancerosa no metastásica; v) una célula de colon cancerosa maligna y una célula de colon cancerosa no maligna (o una célula de colon normal) y/o vi) una célula de colon displásica en relación con una célula de colon normal. Hay otras combinaciones y comparaciones de células de colon afectadas por diversas enfermedades o etapas de una enfermedad que resultarán evidentes para el experto habitual en la técnica. Las realizaciones bioquímicas de la genoteca incluyen una colección de ácidos nucleicos que tienen las secuencias de los genes en la genoteca, pudiendo corresponderse los ácidos nucleicos con todo el gen de la genoteca o con un fragmento del mismo, según lo descrito más detalladamente más adelante.

Las genotecas de polinucleótidos de la presente invención comprenden generalmente la información secuencial de un pluralidad de secuencias de polinucleótidos, comprendiendo al menos uno de los polinucleótidos una secuencia de SEQ ID NO: 1. Pluralidad pretende significar al menos 2, habitualmente al menos 3, y puede incluir hasta todas las SEQ ID NO: 1, 3, 5, 7, 9, 11-13, 15, 16, 18, 20, 22, 24, 26, 27 y 29. La longitud y el número de polinucleótidos de la genoteca variará con la naturaleza de la genoteca, p. ej., si la genoteca es una selección de oligonucleótidos, una selección de ADNc, una base de datos informática de la información secuencial, etc.

Cuando la genoteca es una genoteca electrónica, la información de las secuencias de ácidos nucleicos puede estar presente en una variedad de medios. "Medios" se refiere a una elaboración, distinta de una molécula de ácido nucleico aislada, que contiene la información secuencial de la presente invención. Tal elaboración proporciona la secuencia genómica o un subconjunto de la misma en una forma que pueda ser examinada mediante procedimientos que no son aplicables directamente a la secuencia como existe en un ácido nucleico. Por ejemplo, la secuencia de nucleótidos de la presente invención, p. ej., la secuencia de ácidos nucleicos de los polinucleótidos de SEQ ID NO: 1, se puede registrar en un medio de lectura informática, p. ej., cualquier medio que se pueda leer y al que se pueda acceder directamente por ordenador. Tales medios incluyen, pero no se limitan a: medios de almacenamiento magnético, tales como discos flexibles, un medio de almacenamiento de 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.

Cualquier experto en la técnica puede entender fácilmente cómo cualquiera de los medios de lectura informática conocidos actualmente se puede usar para crear una elaboración que comprenda un registro de la presente información secuencial. “Registrado” se refiere a un procedimiento para almacenar información en un medio de lectura informática, usando cualquiera de tales procedimientos según lo conocido en la técnica. Se puede seleccionar cualquier estructura de almacenamiento de datos conveniente, basada en los medios usados para acceder a la información almacenada. Se puede usar una variedad de programas procesadores de datos y formatos para el almacenamiento, p. ej., un archivo de texto procesador de palabras, formato de base de datos, etc. Además de la información secuencial, las versiones electrónicas de las genotecas de la invención se pueden proporcionar en combinación con o en relación con otra información de lectura informática y/o otros tipos de archivos de lectura informática (p. ej., archivos buscables, archivos ejecutables, etc., incluyendo, pero no limitándose a, por ejemplo, programas informáticos de búsqueda, etc.).

Proporcionando la secuencia de nucleótidos en una forma de lectura informática, se puede acceder a la información para una variedad de propósitos. El programa informático para acceder a la información secuencial se encuentra a disposición del público. Por ejemplo, el BLAST discontinuo (Altschul *et al. Nucleic Acids Res.* (1997) 25: 3389-3402) y el BLAZE (Brutlag *et al. Comp. Chem.* (1993) 17: 203). Los algoritmos de búsqueda basados en un sistema Sybase se pueden usar para identificar marcos de lectura abiertos (ORF) dentro del genoma que contengan homología con ORF de otros organismos.

Como se usa en la presente memoria, “un sistema de base informática” se refiere a los medios de hardware y a los medios de software, a los medios de almacenamiento de datos usados para analizar la información de secuencias de nucleótidos de la presente invención. El hardware mínimo de los sistemas de base informática de la presente invención comprende una unidad de procesamiento central (CPU), medios de entrada, medios de salida y medios de almacenamiento de datos. Un experto en la técnica puede entender fácilmente que cualquiera de los sistemas de base informática disponibles actualmente es adecuado para su uso en la presente invención. El medio de almacenamiento de datos puede comprender cualquier elaboración que comprenda un registro de la presente información secuencial según lo descrito anteriormente, o un medio de acceso a la memoria que pueda acceder a tal elaboración.

“Medio de búsqueda” se refiere a uno o más programas puestos en marcha sobre un sistema de base informática para comparar una secuencia diana o un motivo estructural diana, o niveles de expresión de un polinucleótido de una muestra, con la información secuencial almacenada. Se pueden usar medios de búsqueda para identificar fragmentos o regiones del genoma que se aparen con una secuencia diana o un motivo diana particular. Hay una variedad de algoritmos conocidos por el público y disponibles comercialmente, p. ej., MacPattern (EMBL), BLASTN y BLASTX (NCBI). Una “secuencia diana” puede ser cualquier polinucleótido o secuencia de aminoácidos de seis o más nucleótidos contiguos, o dos o mas aminoácidos, preferiblemente, de aproximadamente 10 a 100 aminoácidos o de aproximadamente 30 a 300 nt. Se puede usar una variedad de medios de comparación para realizar la comparación de la información secuencial de una muestra (p. ej., para analizar secuencias diana, motivos diana o niveles de expresión relativos) con el medio de almacenamiento de datos. Un técnico experto puede reconocer con facilidad que es posible usar uno cualquiera de los programas de búsqueda de homologías que está a disposición del público como medio de búsqueda para los sistemas de base informática de la presente invención con el fin de realizar la comparación de secuencias y motivos diana. También se conocen en la técnica los programas informáticos para analizar los niveles de expresión en una muestra y en controles.

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

Hay una variedad de formatos estructurales para los medios de entrada y de salida que se puede usar para introducir y sacar la información en y de los sistemas de base informática de la presente invención. Un formato para un medio de salida clasifica los niveles de expresión relativos de diferentes polinucleótidos. Tal presentación proporciona al experto en la técnica una clasificación de los niveles de expresión relativa para determinar un perfil de expresión de genes.

Según lo tratado anteriormente, la “genoteca” de la invención también engloba genotecas bioquímicas de los polinucleótidos de la SEQ ID NO: 1, p. ej., colecciones de ácidos nucleicos que representan a los polinucleótidos proporcionados. Las genotecas bioquímicas pueden tener una variedad de formas, p. ej., una solución de ADNc, un patrón de ácidos nucleicos de sonda asociado con una superficie de un soporte sólido (i.e., un alineamiento) y similares. Son de particular interés los alineamientos de ácidos nucleicos en los que está representada la SEQ ID NO: 1 en el alineamiento. Alineamiento pretende significar un artículo de elaboración que tiene al menos un sustrato con al menos dos dianas de ácido nucleico distintas en una de sus superficies, pudiendo ser el número de ácidos nucleicos distintos considerablemente más elevado, siendo comúnmente de al menos 10 nt, habitualmente, de al menos 20 nt y a menudo de al menos 25 nt. Se ha desarrollado una variedad de diferentes formatos de alineamiento y son conocidos para los expertos en la técnica. Los alineamientos de la presente invención encuentran uso en una variedad de aplicaciones, incluyendo el análisis de expresión de genes, el rastreo de fármacos, el análisis de mutaciones y similares, según lo revelado en los ejemplos de documentos de patente enumerados anteriormente. Además de las genotecas de ácidos nucleicos, se

proporcionan las genotecas análogas de polipéptidos, en las que los polipéptidos de la genoteca representarán al menos una parte de los polipéptidos codificados por la SEQ ID NO: 1.

## 5 Utilidades

Las sondas de polinucleótidos, que generalmente comprenden al menos 12 nt contiguos de un polinucleótido según lo mostrado en el listado secuencial, se usan para una variedad de objetivos, tales como para el mapeo cromosómico del polinucleótido y la detección de los niveles de transcripción. En los ejemplos, se encuentra otra revelación sobre las regiones preferidas de las secuencias de polinucleótidos reveladas. Una sonda que se hibrida específicamente con un polinucleótido revelado en la presente memoria debería proporcionar una señal de detección al menos 5, 10 ó 20 veces mayor que la hibridación de fondo proporcionada con otras secuencias no relacionadas.

*Detección de los niveles de expresión.* Las sondas de nucleótido se usan para detectar la expresión de un gen correspondiente al polinucleótido proporcionado. En las transferencias Northern, el ARNm se separa electroforéticamente y se pone en contacto con una sonda. Las sondas se detectan cuando se hibridan con una especie de ARNm de un tamaño particular. La cantidad de hibridación se cuantifica para determinar las cantidades relativas de expresión, por ejemplo, en una determinada condición. Las sondas se usan para la hibridación *in situ* de células para detectar la expresión. Las sondas también se pueden usar *in vivo* para la detección de diagnóstico de secuencias hibridantes. Las sondas se marcan comúnmente con un isótopo radiactivo. Se pueden usar otros tipos de etiquetas detectables tales como cromóforos, fluorescentes y enzimas. En los documentos W092/02526 y USPN 5.124.246, se describen otros ejemplos de análisis de hibridación de nucleótidos.

Alternativamente, la reacción en cadena de la polimerasa (PCR) es otro medio para detectar pequeñas cantidades de ácidos nucleicos diana (véase, p. ej., Mullis *et al.*, *Meth. Enzymol.* (1987) 155: 335; USPN 4.683.195; y USPN 4.683.202). Se usan dos nucleótidos de polinucleótidos cebadores que se hibridan con los ácidos nucleicos diana para cebear la reacción. Los cebadores pueden estar compuestos de una secuencia dentro o 3' ó 5' con respecto a los polinucleótidos del listado secuencial. Alternativamente, si los cebadores son 3' y 5' con respecto a estos polinucleótidos, es necesario que no se hibriden a ellos ni a los complementos. Tras la amplificación de la diana con una polimerasa termoestable, los ácidos nucleicos diana amplificados se pueden detectar mediante procedimientos conocidos en la técnica, p. ej., transferencia Southern. El ARNm o el ADNc también se puede detectar mediante técnicas de transferencia tradicionales (p. ej., transferencia Southern, transferencia Northern, etc.) descritas en Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" (Nueva York, Cold Spring Harbor Laboratory, 1989) (p. ej., sin amplificación PCR). En general, el ARNm o el ADNc generado a partir de ARNm usando una enzima polimerasa se puede purificar y separar usando electroforesis sobre gel y transferir a un soporte sólido, tal como nitrocelulosa. El soporte sólido se expone a una sonda marcada, se lava para eliminar cualquier sonda no hibridada y se detectan los dúplex que contienen la sonda marcada.

*Mapeo.* Los polinucleótidos de la presente invención se pueden usar para identificar un cromosoma sobre el que reside el correspondiente gen. Tal mapeo puede ser útil para identificar la función 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 con el polinucleótido cuando se mapeen síndromes o enfermedades particulares en el mismo cromosoma. Por ejemplo, en el documento USPN 5.783.387, se describe el uso de sondas de polinucleótido para la identificación y la cuantificación de aberraciones de secuencias de ácidos nucleicos. Un ejemplo de procedimiento de mapeo es la hibridación *in situ* con fluorescencia (FISH), que facilita la hibridación genómica comparativa para permitir una evaluación genómica total de los cambios del número relativo de copias de las secuencias de ADN (véase, p. ej., Valdes *et al.*, "Methods in Molecular Biology" (1997) 68:1). Los polinucleótidos también se pueden mapear hasta cromosomas particulares usando, por ejemplo, híbridos de radiación o paneles de híbridos específicos del cromosoma. Véase Leach *et al.*, "Advances in Genetics", (1995) 33: 63-99; Walter *et al.*, "Nature Genetics" (1994) 7:22; Walter y Goodfellow, "Trends in Genetics" (1992) 9: 352. Los paneles para el mapeo de híbridos por radiación están disponibles en Research Genetics, Inc., Huntsville, Alabama, EE.UU. Las bases de datos de marcadores que usan diversos paneles están disponibles en la red en <http://F/shgc-www.stanford.edu>; y <http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl>. Se puede usar el programa estadístico RHMAP para construir un mapa basado en los datos de la hibridación por radiación con una medida de la probabilidad relativa de un orden frente al otro. El RHMAP está disponible en la red en <http://www.sph.umich.edu/group/statgen/software>. Además, hay programas comerciales disponibles para identificar regiones de cromosomas comúnmente asociadas con una enfermedad, tal como el cáncer.

*Clasificación o perfilado de tejidos.* La expresión de ARNm específico correspondiente a los polinucleótidos proporcionados 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, los análisis de sondas de ADN ramificado o las técnicas de transferencia que utilizan sondas de ácido nucleico sustancialmente idénticas o complementarias a los polinucleótidos enumerados en el listado secuencial pueden determinar la presencia o la ausencia del correspondiente ADNc o ARNm.

Se puede usar el tipeado de tejidos para identificar el órgano desarrollacional o la fuente de tejido de la lesión metastásica mediante la identificación de la expresión de un determinado marcador de ese órgano o tejido. Si un polinucleótido se expresa sólo en un tipo de tejido específico, y se encuentra una lesión metastásica para expresar ese polinucleótido, entonces se ha identificado la fuente desarrollacional de la lesión. La expresión de un determinado

polinucleótido se puede analizar mediante la detección de bien el correspondiente ARNm o el producto proteico. Como sería fácilmente comprensible para cualquier científico forense, las secuencias reveladas en la presente memoria son útiles en la diferenciación de tejido humano del tejido no humano. En particular, estas secuencias son útiles para diferenciar el tejido humano del tejido de ave, reptil y anfibio, por ejemplo.

*Uso de polimorfismos.* Se puede usar un polinucleótido de la invención en análisis forenses, genéticos, en mapeos y en aplicaciones de diagnóstico en las que la región correspondiente de un gen es polimórfica en la población humana. Se puede usar cualquier procedimiento para detectar un polimorfismo en un gen, incluyendo, pero no limitándose a electroforesis de variantes polimórficas de proteína, sensibilidad diferencial a una escisión de una enzima de restricción e hibridación con sondas específicas del alelo.

#### *Producción de anticuerpos*

Los productos de expresión de un polinucleótido de la invención, así como el correspondiente ARNm, ADNc o gen completo, se pueden preparar y usar para crear anticuerpos para usos experimentales, de diagnóstico y terapéuticos. Para los polinucleótidos a los que no se les ha asignado un gen correspondiente, se proporciona otro procedimiento para identificar el gen correspondiente. Se expresa el polinucleótido o el ADNc relacionado según lo descrito anteriormente, y se preparan los anticuerpos. Estos anticuerpos son específicos de un epítipo sobre el polipéptido codificado por el polinucleótido y pueden precipitar o unirse a la correspondiente proteína nativa en una preparación celular o tisular, o en un extracto libre de células de un sistema de expresión *in vitro*.

Los procedimientos para la producción de anticuerpos que se unen específicamente a un antígeno seleccionado son conocidos en la técnica. Los inmunógenos para preparar anticuerpos se pueden preparar mezclando un polipéptido codificado por un polinucleótido de la invención con un adyuvante y/o haciendo proteínas de fusión con proteínas inmunogénicas más largas. Los polipéptidos también se pueden unir mediante enlace covalente con otras proteínas inmunogénicas más largas, tales como la hemocianina de lapa californiana. Los inmunógenos se administran comúnmente intradérmica, subcutánea o intramuscularmente en animales experimentales tales como conejos, ovejas y ratones para generar los anticuerpos. Los anticuerpos monoclonales se pueden generar aislando células de bazo y fusionando células de mieloma para formar hibridomas. Alternativamente, el polinucleótido seleccionado se administra directamente, tal como por inyección intramuscular y se expresa *in vivo*. La proteína expresada genera una variedad de respuestas inmunes específicas de la proteína que incluyen la producción de anticuerpos, comparable con la administración de la proteína.

Las preparaciones de anticuerpos policlonales y monoclonales específicos de los polipéptidos codificados por un polinucleótido seleccionado se realizan usando procedimientos estándar conocidos en la técnica. Los anticuerpos se unen específicamente a epítopes presentes en los polipéptidos codificados por los polinucleótidos revelados en el listado secuencial. Comúnmente, se necesitan al menos 6, 8, 10 ó 12 aminoácidos contiguos para formar un epítipo. Los epítopes que implican aminoácidos no contiguos pueden requerir un polipéptido más largo, p. ej., de al menos 15, 25 ó 50 aminoácidos. Los anticuerpos que se unen específicamente a los polipéptidos humanos codificados por los polipéptidos proporcionados deberían proporcionar una señal de detección al menos 5, 10 ó 20 veces mayor que la señal de detección proporcionada con otras proteínas cuando se usan en transferencias Western o en otros análisis inmunoquímicos. Preferiblemente, los anticuerpos que se unen específicamente a los polipéptidos de la invención no se unen a otras proteínas en análisis inmunoquímicos a niveles detectables y pueden inmunoprecipitar el polipéptido específico de la solución.

La invención también contempla los anticuerpos naturales específicos de un polipéptido de la invención. Por ejemplo, se pueden purificar los anticuerpos séricos frente a un polipéptido de la invención en una población humana mediante procedimientos conocidos en la técnica, p. ej., pasando suero sobre una columna a la que se ha unido el correspondiente polipéptido seleccionado o la proteína de fusión seleccionada. Los anticuerpos unidos se pueden eluir entonces de la columna usando, por ejemplo, un tampón con una concentración elevada de sales.

Además de los anticuerpos anteriormente tratados, la invención engloba anticuerpos diseñados genéticamente, derivados de anticuerpo (p. ej., anticuerpos monocatenarios, fragmentos de anticuerpo (p. ej., Fab, etc), según los procedimientos conocidos en la técnica.

#### *Procedimientos de diagnóstico y otros procedimientos que implican la detección de productos génicos expresados diferencialmente*

La presente invención proporciona procedimientos de uso de los polinucleótidos descritos en la presente memoria. En realizaciones no restrictivas específicas, los procedimientos son útiles para detectar las células de cáncer de colon, facilitar el diagnóstico del cáncer y de la gravedad de un cáncer (p. ej., el grado del tumor, la carga tumoral y similares) de un sujeto, facilitar la determinación de un pronóstico de un sujeto y evaluar la respuesta del sujeto a una terapia (p. ej., proporcionando una medida del efecto terapéutico mediante, por ejemplo, la evaluación de la carga tumoral durante o tras el régimen quimioterapéutico). La detección se puede basar en la detección de un polinucleótido que se exprese diferencialmente en una célula de cáncer de colon y/o en la detección de un polipéptido codificado por un polinucleótido que se exprese diferencialmente en una célula de cáncer de colon ("un polipéptido asociado con el

cáncer de colon”). Los procedimientos de detección de la invención se pueden realizar *in vitro* o *in vivo*, sobre células aisladas o en tejidos enteros o en un fluido corporal, p. ej., en sangre, plasma, suero, orina y similares).

En general, los procedimientos de la invención que implican la detección de un producto génico (p. ej., ARNm, ADNc generado por tal ARNm y polipéptidos) implican poner en contacto una muestra con una sonda específica del producto génico de interés. “Sonda” como se usa en la presente memoria en tales procedimientos pretende referirse a una molécula que se une específicamente a un producto génico de interés (p. ej., la sonda se une al producto génico diana con una especificidad suficiente para distinguir la unión a la diana frente a la unión inespecífica a moléculas no dianas (de fondo). Las “sondas” incluyen, pero no se limitan necesariamente a, sondas de ácido nucleico (p. ej., ADN, ARN, ácido nucleico modificado y similares), anticuerpos (p. ej., anticuerpos, fragmentos de anticuerpo que conservan la unión a un epítopo diana, anticuerpos monocatenarios y similares) u otro polipéptido, péptido o molécula (p. ej., ligando receptor) que se una específicamente a un producto génico diana de interés.

La sonda y la muestra sospechosa de tener el producto génico de interés se ponen en contacto en condiciones adecuadas para unir la sonda al producto génico. Por ejemplo, la puesta en contacto se realiza generalmente durante el tiempo suficiente para permitir la unión de la sonda con el producto génico (p. ej., de varios minutos a unas cuantas horas) y a una temperatura y en condiciones de osmolaridad, y similares que proporcionen la unión de la sonda con el producto génico a un nivel que sea suficientemente distinguible de la unión de fondo de la sonda (p. ej., en condiciones que minimicen la unión inespecífica). Las condiciones adecuadas para la unión de la sonda-producto génico diana se pueden determinar fácilmente usando controles y otras técnicas disponibles y conocidos por cualquier experto habitual en la técnica.

En esta realización, la sonda puede ser un anticuerpo u otro polipéptido, péptido o molécula (p. ej., un ligando receptor) que se una específicamente a un polipéptido diana de interés.

Los procedimientos de detección se pueden proporcionar como parte de un equipo. De este modo, la invención proporciona además equipos para detectar la presencia y/o un nivel de un polinucleótido que está expresado diferencialmente en una célula de cáncer de colon (p. ej., mediante la detección de un ARNm codificado por un gen expresado diferencialmente de interés) y/o un polipéptido codificado por el mismo, en una muestra biológica. Los procedimientos que usan estos equipos se pueden realizar en laboratorios clínicos, laboratorios experimentales, por profesionales médicos o por usuarios privados. Los equipos de la invención para detectar un polipéptido codificado por un polinucleótido que está expresado diferencialmente en una célula de cáncer de colon comprenden un resto que se une específicamente al polipéptido, que puede ser un anticuerpo específico. Los equipos de la invención para detectar un polipéptido codificado por un polinucleótido que está expresado diferencialmente en una célula de cáncer de colon comprenden un resto que se hibrida específicamente con tal polinucleótido. El equipo puede proporcionar opcionalmente más componentes que son útiles en el procedimiento, incluyendo, pero no limitándose a, tampones, reactivos de desarrollo, etiquetas, superficies reactivas, medios de detección, muestras control, patrones, instrucciones e información interpretativa.

*Detección de un polipéptido codificado por un polinucleótido que está expresado diferencialmente en una célula de cáncer de colon*

En algunas realizaciones, se proporcionan procedimientos para una célula de cáncer de colon mediante la detección en la célula de un polipéptido codificado por un gen que está expresado diferencialmente en una célula de cáncer de colon. Se puede usar cualquiera de una variedad de procedimientos conocidos para la detección, incluyendo, pero no limitándose a, inmunoanálisis, usando anticuerpo específico del polipéptido codificado, p. ej., mediante análisis de inmunoabsorción ligada a enzimas (ELISA), radioinmunoanálisis (RIA) y similares; y análisis funcionales para el polipéptido codificado, p. ej., actividad de unión o actividad enzimática.

Por ejemplo, se puede realizar fácilmente un análisis de inmunofluorescencia sobre células sin tener que aislar primero el polipéptido codificado. Primero se fijan las células sobre un soporte sólido, tal como un portaobjetos de microscopio o un pozo de microvaloración. La etapa de fijado puede permeabilizar la membrana celular. La permeabilización de la membrana celular permite la unión de la sonda específica y el polipéptido (p. ej., el anticuerpo). Alternativamente, cuando el polipéptido es secretado o está unido a la membrana, o es accesible de otro modo en la superficie celular (p. ej., receptores u otra molécula asociada de forma estable con la membrana celular externa o asociada establemente de otro modo con la membrana celular), puede que tal permeabilización no sea necesaria.

A continuación, se exponen las células fijadas a un anticuerpo específico del polipéptido codificado. Para aumentar la sensibilidad del análisis, se pueden exponer además las células fijadas a un segundo anticuerpo que esté marcado y se una al primer anticuerpo que es específico del polipéptido codificado. Comúnmente, el anticuerpo secundario está marcado detectablemente, p. ej., con un marcador fluorescente. Las células que expresan el polipéptido codificado serán marcadas fluorescentemente y se visualizarán fácilmente en el microscopio. Véase, por ejemplo, Hashido *et al.*, (1992) *Biochem. Biophys. Res. Comm.* 187: 1241-1248.

Como será fácilmente comprensible para un experto habitual en la técnica tras la lectura de la presente memoria, los procedimientos de detección u otros procedimientos descritos en la presente memoria pueden ser fácilmente modificados. Tales variaciones pertenecen al ámbito pretendido de la invención. Por ejemplo, en el esquema de detección

anterior, la sonda para su uso en la detección puede ser inmovilizada sobre un soporte sólido, pudiéndose poner en contacto la muestra de análisis con la sonda inmovilizada. Entonces se puede detectar la unión de la muestra de análisis con la sonda en una variedad de modos, p. ej., detectando una etiqueta detectable unida a la muestra de análisis para facilitar la detección de complejos de muestra de análisis-sonda inmovilizada.

La presente invención proporciona además procedimientos para detectar la presencia de y/o medir un nivel de un polipéptido en una muestra biológica, estando el polipéptido codificado por un polinucleótido que representa un gen expresado diferencialmente en el cáncer, particularmente, en una célula de cáncer de colon, usando una sonda específica para el polipéptido codificado. En esta realización, la sonda puede ser un anticuerpo u otro polipéptido, péptido o molécula (p. ej., un ligando receptor) que se una específicamente a un polipéptido diana de interés.

Los procedimientos comprenden generalmente: a) poner en contacto la muestra con un anticuerpo específico de un polipéptido expresado diferencialmente en una célula de análisis; y b) detectar la unión entre el anticuerpo y las moléculas de la muestra. El nivel de unión al anticuerpo (bien cualitativo o cuantitativo) incide el estado canceroso de la célula. Por ejemplo, cuando se aumenta el gen expresado diferencialmente en células cancerosas, la detección de un aumento del nivel de unión del anticuerpo con la muestra de análisis en comparación con el nivel de unión del anticuerpo asociado con una célula normal indica que la célula de análisis es cancerosa.

Los controles adecuados incluyen una muestra de la que se sabe que no contiene el polipéptido codificado y una muestra puesta en contacto con un anticuerpo inespecífico del polipéptido codificado, p. ej., un anticuerpo anti-idiotípico. Hay una variedad de procedimientos para detectar las interacciones entre anticuerpos específicos y antígenos que es conocida en la técnica, y que se puede usar en el procedimiento, incluyendo, pero no limitándose a, procedimientos inmunohistológicos estándar, de inmunoprecipitación, un inmunoanálisis enzimático y un radioinmunoanálisis.

En general, el anticuerpo específico será marcado detectablemente, bien directa o indirectamente. Las etiquetas directas incluyen radioisótopos; enzimas cuyos productos son detectables (p. ej., luciferasa,  $\beta$ -galactosidasa y similares); etiquetas fluorescentes (p. ej., isotiocianato de fluoresceína, rodamina, ficoeritrina y similares); metales emisores de fluorescencias, p. ej.,  $^{152}\text{Eu}$  u otros de la serie de los lantánidos, unidos al anticuerpo a través de grupos quelantes metálicos tales como EDTA; compuestos quimioluminiscentes, p. ej., luminil, isoluminol, sales de acridinio y similares; compuestos bioluminiscentes, p. ej., luciferina, aecorina (proteína verde fluorescente) y similares.

Se puede unir el anticuerpo (acoplar) a un soporte insoluble, tal como una placa de poliestireno o una perla. Las etiquetas indirectas incluyen segundos anticuerpos específicos de anticuerpos específicos del polipéptido codificado ("primer anticuerpo específico"), estando el segundo anticuerpo marcado según lo descrito anteriormente; y miembros de pares de unión específicos, p. ej., biotina-avidina y similares. Se puede poner en contacto la muestra biológica con e inmovilizarla sobre un soporte sólido o un vehículo, tal como nitrocelulosa, que sea capaz de inmovilizar células, partículas celulares o proteínas solubles. Entonces se puede lavar el soporte con tampones adecuados, siguiendo con la puesta en contacto con un primer anticuerpo específico marcado detectablemente. Los procedimientos de detección son conocidos en la técnica y serán seleccionados según lo apropiado para la señal emitida por la etiqueta detectable. La detección se realiza generalmente en comparación con controles adecuados y con patrones apropiados.

En algunas realizaciones, los procedimientos están adaptados para su uso *in vivo*, p. ej., para localizar o identificar sitios en los que estén presentes células de cáncer de colon. En estas realizaciones, se administra un resto marcado detectablemente, p. ej., un anticuerpo, que sea específico de un polipéptido asociado al cáncer de colon a un individuo (p. ej., mediante inyección) y se localizan las células marcadas usando técnicas estándar de formación de imágenes, incluyendo, pero no limitándose a, formación de imágenes por resonancia magnética, barrido de tomografía por ordenador y similares. De este modo, las células de cáncer de colon quedan marcadas diferencialmente.

#### *Detección de un polinucleótido que representa un gen expresado diferencialmente en una célula de cáncer de colon*

En algunas realizaciones, se proporcionan procedimientos para detectar una célula de cáncer de colon mediante la detección de la expresión en la célula de un transcripto o que está expresada diferencialmente en una célula de cáncer de colon. Se puede usar cualquiera de una variedad de procedimientos conocidos de detección, incluyendo, pero no limitándose a, la detección de un transcripto mediante la hibridación con un polinucleótido que se hibrida con un polinucleótido que está expresado diferencialmente en una célula de cáncer de colon; la detección de un transcripto mediante la reacción en cadena de la polimerasa usando cebadores de oligonucleótido específicos; hibridación *in situ* de una célula usando como sonda un polinucleótido que se hibrida con un gen que está expresado diferencialmente en una célula de cáncer de colon.

Los procedimientos se pueden usar para detectar y/o medir los niveles de ARNm de un gen que está expresado diferencialmente en una célula de cáncer de colon. En algunas realizaciones, los procedimientos comprenden: a) poner en contacto una muestra con un polinucleótido que se corresponde con un gen expresado diferencialmente descrito en la presente memoria en condiciones que permitan la hibridación; y b) detectar la hibridación si la hay. La detección de una hibridación diferencial, en comparación con un control adecuado, es indicativo de la presencia en la muestra de un polinucleótido que se expresa diferencialmente en una célula de cáncer de colon. Los controles apropiados incluyen, por ejemplo, una muestra de la que se sabe que contiene un polinucleótido que está expresado diferencialmente en una célula de cáncer de colon, y el uso de un polinucleótido marcado del mismo "sentido" que el polinucleótido que

está expresado diferencialmente en una célula de cáncer de colon. Las condiciones que permiten la hibridación son conocidas en la técnica y han sido descritas más detalladamente con anterioridad.

También se puede realizar la detección mediante cualquier procedimiento conocido, incluyendo, pero no limitándose a, hibridación *in situ*, PCR (reacción en cadena de la polimerasa), RT-PCR (PCR de transcripción inversa) y transferencia “Northern” o de ARN, o combinaciones de tales técnicas, usando un polinucleótido adecuadamente marcado. Hay una variedad de etiquetas y procedimientos de marcaje de polinucleótidos que es conocida en la técnica y que se puede usar en los procedimientos de análisis de la invención. Es posible determinar la hibridación específica mediante la comparación con controles apropiados.

El polinucleótido que comprende generalmente al menos 12 nt contiguos de un polinucleótido proporcionado en la presente memoria, según lo mostrado en el listado secuencial o en la secuencias de los genes correspondientes a los polinucleótidos del listado secuencial, se usa para una variedad de objetivos, tales como sondas para detección de y/o la medida de los niveles de transcripción de un polinucleótido que esté expresado diferencialmente en una célula de cáncer de colon. En los ejemplos, se encuentra otra revelación sobre las regiones preferidas de las secuencias de polinucleótidos reveladas. Una sonda que se hibrida específicamente con un polinucleótido revelado en la presente memoria debería proporcionar una señal de detección al menos 5, 10 ó 20 veces mayor que la hibridación de fondo proporcionada con otras secuencias no relacionadas. Debería indicarse que “sonda” como se usa en este contexto de detección de ácido nucleico pretende referirse a una secuencia de polinucleótidos usada para detectar un producto génico expresado diferencialmente en una muestra de análisis. Como será fácilmente comprensible para el experto habitual en la técnica, la sonda puede estar marcada detectablemente y puesta en contacto con, por ejemplo, un alineamiento que comprenda polinucleótidos inmovilizados obtenidos de una muestra de análisis (p. ej., ARNm). Alternativamente, se puede inmovilizar la sonda sobre un alineamiento y se puede marcar detectablemente la muestra de análisis. Estas y otras variaciones de los procedimientos de la invención pertenecen al ámbito de la técnica y al ámbito de la invención.

Las sondas de nucleótido se usan para detectar la expresión de un gen correspondiente al polinucleótido proporcionado. En las transferencias Northern, el ARNm se separa electroforéticamente y se pone en contacto con una sonda. Las sondas se detectan cuando se hibridan con una especie de ARNm de un tamaño particular. Se puede cuantificar la cantidad de hibridación para determinar las cantidades relativas de expresión, por ejemplo, en una determinada condición. Las sondas se usan para la hibridación *in situ* de células para detectar la expresión. Las sondas también se pueden usar *in vivo* para la detección de diagnóstico de secuencias hibridantes. Las sondas se marcan comúnmente con un isótopo radiactivo. Se pueden usar otros tipos de etiquetas detectables tales como cromóforos, fluorescentes y enzimas. En los documentos W092/02526 y USPN 5.124.246, se describen otros ejemplos de análisis de hibridación de nucleótidos.

La PCR es otro procedimiento para detectar pequeñas cantidades de ácidos nucleicos diana (véase, p. ej., Mullis *et al.*, *Meth. Enzymol.* (1987) 155: 335; USPN 4.683.195; y USPN 4.683.202). Se usan dos nucleótidos de polinucleótidos cebadores que se hibridan con los ácidos nucleicos diana para cebar la reacción. Los cebadores pueden estar compuestos de una secuencia dentro, o 3' o 5' con respecto de los polinucleótidos del listado secuencial. Alternativamente, si los cebadores están 3' y 5' con respecto a estos polinucleótidos, no es necesario que se hibriden a ellos ni a los complementos. Tras la amplificación de la diana con una polimerasa termoestable, los ácidos nucleicos diana amplificados se pueden detectar mediante procedimientos conocidos en la técnica, p. ej., transferencia Southern. El ARNm o el ADNc también se puede detectar mediante técnicas de transferencia tradicionales (p. ej., transferencia Southern, transferencia Northern, etc.) descritas en Sambrook *et al.*, “Molecular Cloning: A Laboratory Manual” (Nueva York, Cold Spring Harbor Laboratory, 1989) (p. ej., sin amplificación PCR). En general, el ARNm o el ADNc generado a partir de ARNm usando una enzima polimerasa se puede purificar y separar usando electroforesis sobre gel y transferir a un soporte sólido, tal como nitrocelulosa. El soporte sólido se expone a una sonda marcada, se lava para eliminar cualquier sonda no hibridada y se detectan los dúplex que contienen la sonda marcada.

Los procedimientos que usan la amplificación por PCR se pueden realizar sobre el ADN de una sola célula, aunque lo conveniente es usar al menos aproximadamente 10<sup>5</sup> células. En Saiki *et al.* (1985) *Science* 239: 487, se describe el uso de la reacción en cadena de la polimerasa, y en Sambrook, *et al.* “Molecular Cloning: A Laboratory Manual”, CSH Press 1989, pp. 14.2-14.33, se encuentra una revisión de las técnicas actuales. Las etiquetas detectables adecuadas incluyen fluorocromos (p. ej., isotiocianato de fluoresceína (FITC), rodamina, rojo Texas, ficoeritrina, alofocianina, 6-carboxifluoresceína (6-FAM), 2',7'-dimetoxi-4',5'-dicloro-6-carboxifluoresceína, 6-carboxi-X-rodamina (ROX), 6-carboxi-2',4',7',4,7-hexaclorofluoresceína (HEX), 5-carboxifluoresceína (5-FAM) o N,N,N',N'-tetrametil-6-carboxirrodamina (TAMRA)), etiquetas radiactivas, (p. ej., <sup>32</sup>P, <sup>35</sup>S, <sup>3</sup>H, etc.), y similares. La etiqueta puede ser un sistema bifásico, en el que los polinucleótidos sean conjugados con biotina, haptenos, etc., que tengan un patrón de unión de afinidad elevada, p. ej., avidina, anticuerpos específicos, etc., en el que el patrón de unión esté conjugado con una etiqueta detectable. La etiqueta puede estar conjugada con uno o ambos cebadores. Alternativamente, se marca una mezcla de nucleótidos usada en la amplificación, para incorporar la etiqueta al producto de amplificación.

#### Alineamientos

Los alineamientos de polinucleótidos proporcionan una técnica de un rendimiento elevado que puede analizar un gran número de polinucleótidos o polipéptidos de una muestra. Esta tecnología se puede usar como herramienta para analizar la expresión diferencial.

Hay una variedad de procedimientos para producir alineamientos, así como variaciones de estos procedimientos, que se conocen en la técnica y se contemplan para su uso en la invención. Por ejemplo, se pueden crear alineamientos para encontrar sondas de polinucleótido sobre un sustrato (p. ej., vidrio, nitrocelulosa, etc) en una matriz bidimensional o en un alineamiento que tenga sondas unidas. Las sondas se pueden unir al sustrato bien mediante enlaces covalentes o mediante interacciones inespecíficas, tales como mediante interacciones hidrófobas.

Las muestras de polinucleótidos se pueden marcar detectablemente (p. ej., usando etiquetas radiactivas o fluorescentes) y luego se pueden hibridar a las sondas. Los polinucleótidos bicatenarios, que comprenden los polinucleótidos de la muestra marcados unidos a los polinucleótidos de la sonda, se pueden detectar una vez que la parte sin unir de la muestra se haya retirado mediante lavado. Alternativamente, se pueden inmovilizar los polinucleótidos de la muestra de análisis sobre el alineamiento y las sondas se pueden marcar detectablemente. Por ejemplo en Schena *et al.* (1996) *Proc Natl Acad Sci EE.UU.* 93(20): 10614-9; Schena *et al.* (1995) *Science* 270 (5235): 467-70; Shalon *et al.* (1996) *Genome Res.* 6 (7): 639-45, USPN 5.807.522, EP 799.897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; USPN 5.593.839; USPN 5.578.832; EP 728 520; USPN 5.599.695; EP 721 016; USPN 5.556.752; WO 95/22058; y USPN 5.631.734, se describen técnicas para construir alineamientos y los procedimientos para usar estos alineamientos.

Los alineamientos se pueden usar, por ejemplo, para examinar la expresión diferencial de genes y para determinar la función de los genes. Por ejemplo, se pueden usar los alineamientos para detectar la expresión diferencial de un gen correspondiente a un polinucleótido descrito en la presente memoria, comparándose la expresión entre una célula de análisis y una célula control (p. ej., células de cáncer y células normales). Por ejemplo, la expresión alta de un determinado mensaje en una célula cancerosa, que no se observe en una célula normal correspondiente, puede indicar un producto génico específico del cáncer. En, por ejemplo, Pappalarado *et al.*, *Sem. Radiation Oncol.* (1998) 8:217; y Ramsay *Nature Biotechnol.* (1998) 16:40, se describen ejemplos de usos de alineamientos con mayor detalle. Además, hay numerosas variaciones de los procedimientos de detección que usan alineamientos que son conocidos en la técnica y pertenecen al ámbito de la invención. Por ejemplo, en lugar de inmovilizar la sonda en un soporte sólido, se puede inmovilizar la muestra de análisis sobre un soporte sólido que luego se ponga en contacto con la sonda.

### 30 *Diagnosis, pronóstico, evaluación de terapia contra (teramétrica) y gestión del cáncer*

Los polinucleótidos descritos en la presente memoria, así como sus productos génicos y genes correspondientes, son de particular interés como marcadores genéticos o bioquímicos (p. ej., en sangre o tejidos) que detectarán los cambios más tempranos de la ruta de la carcinogénesis y/o controlarán la eficacia de diversas terapias e intervenciones preventivas.

Por ejemplo, el nivel de expresión de determinados polinucleótidos puede ser indicativo de un pronóstico más pobre, y por consiguiente, garantizar una quimioterapia o radioterapia más agresiva para un paciente, o viceversa. La correlación de nuevas características específicas del tumor suplente con respuesta al tratamiento y al resultado en pacientes puede definir los indicadores de pronóstico que permitan el diseño de terapias personalizadas en base al perfil molecular del tumor. Estas terapias incluyen la dirección de anticuerpos, los antagonistas (p. ej., moléculas pequeñas) y la terapia génica.

La determinación de la expresión de determinados polinucleótidos y la comparación de un perfil de pacientes con la expresión conocida en un tejido normal y las variantes de la enfermedad permite la determinación del mejor tratamiento posible para un paciente, tanto en términos de especificidad del tratamiento como en términos del nivel de comodidad para el paciente. Los marcadores tumorales suplentes, tales como la expresión de polinucleótidos, también se pueden usar para clasificar mejor, y de este modo, diagnosticar y tratar diferentes formas y estados patológicos del cáncer. Dos clasificaciones ampliamente usadas en oncología que pueden beneficiarse de la identificación de los niveles de expresión de los genes correspondientes a los polinucleótidos descritos en la presente memoria son el establecimiento de la fase en la que se encuentra el trastorno canceroso y la graduación de la naturaleza del tejido canceroso.

Los polinucleótidos que se corresponden con los genes expresados diferencialmente, así como sus productos génicos codificados, pueden ser útiles para controlar pacientes que tengan o que sean susceptibles al cáncer para detectar eventos potencialmente malignos a un nivel molecular antes de que sean detectables a un nivel morfológico bruto. Además, los polinucleótidos descritos en la presente memoria, así como los genes que se corresponden con tales polinucleótidos, pueden ser útiles como teramétricos, p. ej., para evaluar la eficacia de la terapia usando los polinucleótidos o sus productos génicos codificados, para evaluar, por ejemplo, la carga tumoral en el paciente antes, durante y tras la terapia.

Además, los polinucleótidos identificados como correspondientes a genes que están expresados diferencialmente en, y por consiguiente que son importantes para, un tipo de cáncer también pueden tener implicaciones en el desarrollo o el riesgo de desarrollar otros tipos de cáncer, p. ej., cuando un polinucleótido representa un gen expresado diferencialmente a través de diversos tipos de cáncer. De este modo, por ejemplo, la expresión de un polinucleótido correspondiente a un gen que tiene implicaciones clínicas en el cáncer de colon metastásico también puede tener implicaciones clínicas en el cáncer de mama o el cáncer de ovario.



*Establecimiento de la fase.* El establecimiento de la fase es un procedimiento usado por los médicos para describir cuánto ha avanzado el estado canceroso en un paciente. El establecimiento de la fase ayuda al médico a determinar un pronóstico, a planificar el tratamiento y a evaluar los resultados de tal tratamiento. Los sistemas de establecimiento de la fase varían con los tipos de cáncer, pero generalmente implican el siguiente sistema "TNM": el tipo de tumor, indicado por T; si el cáncer se ha metastasificado hasta cerca de los nódulos linfáticos, se indica con una N; y si el cáncer se ha metastasificado hasta partes más distantes del cuerpo, se indica con una M. Generalmente, si el cáncer sólo se detecta en una zona de la lesión primaria sin que se haya extendido a ningún nódulo linfático, se denomina fase I. Si se ha extendido sólo hasta los nódulos linfáticos más cercanos, se denomina fase II. En la fase II, el cáncer generalmente se ha extendido hasta los nódulos linfáticos en una proximidad cercana al punto de la lesión primaria. Los cánceres que se han extendido hasta una parte distante del cuerpo, tal como el hígado, los huesos, el cerebro u otro sitio, están en fase IV, la fase más avanzada.

Los polinucleótidos y los genes correspondientes, así como los productos génicos descritos en la presente memoria pueden facilitar el ajuste del procedimiento de establecimiento de la fase identificando los marcadores de la agresividad de un cáncer, p. ej., el potencial metastásico, así como la presencia en diferentes zonas del cuerpo. De este modo, un cáncer en fase II con un polinucleótido que expresa un cáncer de elevado potencial metastásico se puede usar para cambiar un tumor en el límite de la fase II en un tumor en fase III, justificando así una terapia más agresiva. Por el contrario, la presencia de un polinucleótido que expresa un potencial metastásico más bajo permite un establecimiento de la fase del tumor más conservador.

*Graduación de los cánceres.* Grado es un término usado para describir cuánto se asemeja un tumor al tejido normal de su mismo tipo. El aspecto microscópico de un tumor se usa para identificar el grado tumoral en base a parámetros tales como la morfología celular, la organización celular y otros marcadores de diferenciación. Como regla general, el grado de un tumor se corresponde con su velocidad de crecimiento o su agresividad, siendo los tumores no diferenciados o de alto grado más agresivos que los tumores diferenciados o de bajo grado. Para graduar los tumores, se usan generalmente las siguientes directrices: 1) el grado GX no se puede evaluar; 2) G1: bien diferenciado; G2: Moderadamente bien diferenciado; 3) G3: Poco diferenciado; 4) G4: No diferenciado. Los polinucleótidos del listado secuencial, así como sus genes y productos génicos correspondientes, pueden ser especialmente valiosos en la determinación del grado de un tumor, pues no sólo pueden ayudar a determinar el estado de diferenciación de las células de un tumor, sino que además pueden identificar factores distintos de la diferenciación que son valiosos para determinar la agresividad de un tumor, tales como el potencial metastásico.

*Detección del cáncer de colon.* Los polinucleótidos correspondientes a los genes que presentan el patrón de expresión apropiado se pueden usar para detectar el cáncer de colon en un sujeto. El cáncer colorrectal es uno de los neoplasmas más comunes en seres humanos y quizás la forma más frecuente de neoplasia hereditaria. La prevención y la detección temprana son factores clave para controlar y curar el cáncer colorrectal. El cáncer colorrectal comienza como pólipos, que son pequeños crecimientos benignos de las células que forman el revestimiento interior del colon. Durante un período de varios años, algunos de estos pólipos van acumulando más mutaciones y se convierten en cancerosos. Se han identificado múltiples trastornos de cáncer colorrectal familiar, que se resumen en los siguientes: 1) Poliposis adenomatosa familiar (PAF); 2) Síndrome de Gardner; 3) Cáncer de colon no polipoide hereditario (CCNPH); y 4) Cáncer colorrectal familiar de judíos Ashkenazi.

La expresión de los polinucleótidos apropiados se puede usar en el diagnóstico, el pronóstico y el tratamiento del cáncer de colon. La detección del cáncer de colon se puede determinar usando los niveles de expresión de cualquiera de estas secuencias solas o en combinación con los niveles de expresión. La determinación de la naturaleza agresiva y/o del potencial metastásico de un cáncer de colon se puede determinar comparando los niveles de uno o más producto génicos de los genes correspondientes a los polinucleótidos descritos en la presente memoria, y comparando los niveles totales de otra secuencia de la que se sabe que varían en tejido canceroso, p. ej., la expresión de p53, DCC, ras, PAF (véase, p. ej., Fearon ER, *et al.*, 759; Hamilton SR *et al.*, Cancer (1993) 72: 957; Bodmer W, *et al.*, Nat Genet. (1994) 4 (3): 217; Fearon ER, Ann. NY Acad. Sci. (1995) 768: 101).

Por ejemplo, es posible detectar el desarrollo del cáncer de colon examinando el nivel de expresión de un gen correspondiente a un polinucleótido descrito en la presente memoria a los niveles de oncogenes (p. ej., ras) o de los genes de supresión tumoral (p. ej., PAF o p53). De este modo, la expresión de los polinucleótidos marcadores específicos se puede usar para discriminar entre el tejido normal y el canceroso, para diferenciar entre los cánceres de colon con diferentes células de origen, para diferenciar entre los cánceres de colon con diferentes velocidades metastásicas potenciales, etc. Para una revisión de los marcadores del cáncer, véase, p. ej., Hanahan *et al.* (2000) Cell 100: 57-70.

#### Tratamiento del cáncer de colon

La invención proporciona además procedimientos para reducir el crecimiento de las células de cáncer de colon. Los procedimientos proporcionan la disminución de la expresión de un gen que está expresado diferencialmente en una célula de cáncer de colon o la disminución del nivel de y/o la disminución de la actividad de un polipéptido asociado con el cáncer de colon. En general, los procedimientos comprenden poner en contacto una célula de cáncer de colon con una sustancia que module (1) la expresión de un gen que se exprese diferencialmente en el cáncer de colon o (2) un nivel de y/o una actividad de un polipéptido asociado con el cáncer de colon. "La reducción del crecimiento de

las células de cáncer de colon” incluye, pero no se limita a, la reducción de la proliferación de células de cáncer de colon y la reducción de la incidencia de una célula que no es de cáncer de colon en convertirse en una célula de colon cancerosa. Es posible determinar fácilmente si se ha logrado la reducción del crecimiento de las células de cáncer de colon usando cualquier análisis conocido, incluyendo, pero no limitándose a, la incorporación de [<sup>3</sup>H]-timidina; el recuento del número de células durante un período de tiempo; la detección y/o la medida de un marcador asociado con el cáncer de colon (p. ej., CEA, CA19-9 y LASA).

La presente invención proporciona procedimientos para tratar el cáncer de colon, comprendiendo generalmente la administración de una sustancia que reduzca el crecimiento de las células de cáncer de colon a un individuo en necesidad de la misma, en una cantidad suficiente para reducir el crecimiento de las células de cáncer de colon y tratar el cáncer de colon. Se puede estimar si una sustancia o una cantidad específica de la sustancia es eficaz para tratar el cáncer de colon usando cualquiera de una variedad de análisis de diagnóstico conocidos para el cáncer de colon, incluyendo, pero no limitándose a, la sigmoidoscopia, la proctoscopia, la exploración rectal, la colonoscopia con biopsia, los estudios radiográficos de contraste, los barridos de CAT, la angiografía y la detección de CAT, la angiografía y la detección de un marcador tumoral asociado con el cáncer de colon en la sangre del individuo. La sustancia se puede administrar sistemática o localmente. De este modo, en algunas realizaciones, la sustancia se administra localmente, y el crecimiento del cáncer de colon disminuye en el lugar de la administración. La administración local puede ser útil para tratar, p. ej., un tumor sólido.

La sustancia que reduce el crecimiento de las células de cáncer de colon se puede dirigir a una célula de cáncer de colon. De este modo, en algunas realizaciones, la invención proporciona un procedimiento para administrar un fármaco a una célula de cáncer de colon, que comprende la administración de un complejo de fármaco-anticuerpo a un sujeto, siendo el anticuerpo específico de un polipéptido asociado con el cáncer de colon, y el fármaco uno que reduzca el crecimiento de las células de cáncer de colon, una variedad de los cuales se conoce en la técnica. La dirección se puede realizar acoplado (p. ej., uniendo, directamente o mediante una molécula ligadora, bien mediante enlace covalente o no covalente, para formar un complejo de fármaco-anticuerpo) un fármaco con un anticuerpo específico de un polipéptido asociado con el cáncer de colon. Los procedimientos para acoplar un fármaco a un anticuerpo son conocidos en la técnica, siendo innecesario describirlos en la presente memoria.

#### Identificación de dianas terapéuticas y agentes terapéuticos contra el cáncer

La presente invención también engloba procedimientos para identificar agentes que tengan la capacidad de modular la actividad de un producto génico expresado diferencialmente, así como los procedimientos para identificar un producto génico expresado diferencialmente como una diana terapéutica para el tratamiento del cáncer, especialmente, del cáncer de colon.

#### Agentes candidatos

La identificación de los compuestos que modulan la actividad de un producto génico expresado diferencialmente se puede realizar usando cualquiera de entre una variedad de técnicas de rastreo de fármacos. Tales agentes son candidatos para desarrollar terapias contra el cáncer. Son de particular interés los análisis de rastreo de agentes que tienen una toxicidad tolerable para las células humanas no cancerosas. Los análisis de rastreo de la invención están generalmente basados en la capacidad del agente para modular una actividad de un producto génico expresado diferencialmente y/o para inhibir o suprimir el fenómeno asociado con el cáncer (p. ej., la proliferación celular, la formación de colonias, la detención del ciclo celular, la metástasis y similares).

El término “agente” como se usa en la presente memoria describe cualquier molécula, p. ej., proteína o compuesto farmacéutico, que tenga la capacidad de modular una actividad biológica de un producto génico de un gen expresado diferencialmente. Generalmente, se ejecuta una pluralidad de mezclas de análisis en paralelo con diferentes concentraciones de agente para obtener una respuesta diferencial a las diversas concentraciones. Comúnmente, una de estas concentraciones sirve como control negativo, i.e., a una concentración cero o por debajo del nivel de detección.

Los agentes candidatos engloban numerosas clases químicas, aunque comúnmente son moléculas orgánicas, preferiblemente, pequeños compuestos orgánicos que tienen un peso molecular de más de 50 y de menos de aproximadamente 2.500 daltons. Los agentes candidatos comprenden grupos funcionales necesarios para la interacción estructural con proteínas, particularmente, enlaces de hidrógeno, y comúnmente incluyen al menos un grupo amino, carbonilo, hidroxilo o carboxilo, preferiblemente, al menos dos de los grupos químicos funcionales. Los agentes candidatos comprenden habitualmente estructuras heterocíclicas o cíclicas de carbonos y/o estructuras aromáticas o poliaromáticas sustituidas por uno o más de los grupos funcionales anteriores. Los agentes candidatos también se encuentran entre las biomoléculas, que incluyen, pero no se limitan a: péptidos, sacáridos, ácidos grasos, esteroides, purinas, pirimidinas, derivados, análogos estructurales o combinaciones de los mismos.

Los agentes candidatos se obtienen de una amplia variedad de fuentes, incluyendo genotecas de compuestos sintéticos o naturales. Por ejemplo, hay numerosos procedimientos disponibles para la síntesis aleatoria y directa de una amplia variedad de compuestos orgánicos y biomoléculas, incluyendo la expresión de oligonucleótidos y oligopéptidos aleatorizados. Alternativamente, hay disponibles o se pueden producir fácilmente genotecas de compuestos naturales

en forma de extractos bacterianos, fúngicos, vegetales y animales (incluyendo los extractos de tejido humano para identificar factores endógenos que afectan a los productos de los genes expresados diferencialmente). Además, las genotecas y los compuestos naturales o producidos sintéticamente se pueden modificar fácilmente a través de procedimientos convencionales químicos, físicos y bioquímicos, y se pueden usar para producir genotecas combinatorias.

5 Los agentes farmacológicos conocidos se pueden someter a modificaciones directas o aleatorias, tales como acilación, alquilación, esterificación, amidificación, etc., para producir análogos estructurales.

Los ejemplos de agentes candidatos de particular interés incluyen, pero no se limitan a, polinucleótidos, y anticuerpos antisentido, receptores solubles y similares. Los anticuerpos y los receptores solubles son de particular interés

10 como agentes candidatos cuando el producto génico expresado diferencialmente diana es secretado o accesible en la superficie celular (p. ej., receptores y otras moléculas asociados establemente con la membrana celular externa).

#### *Rastreo de agentes candidatos*

Los análisis de rastreo pueden estar basados en cualquiera de una variedad de técnicas fácilmente disponibles y conocidas por cualquier experto habitual en la técnica. En general, los análisis de rastreo implican poner en contacto una célula cancerosa (preferiblemente, una célula de colon cancerosa) con un agente candidato, y evaluar el efecto producido sobre la actividad biológica de un producto génico expresado diferencialmente. El efecto generado sobre

20 una actividad biológica se puede detectar mediante, por ejemplo, la detección de la expresión de un producto génico de un gen expresado diferencialmente (p. ej., una disminución del ARNm o de los niveles de polipéptidos, causaría a su vez una disminución de la actividad biológica del producto génico). Alternativamente o además, se puede evaluar el efecto del agente candidato examinando el efecto del agente candidato en un análisis funcional. Por ejemplo, cuando el producto génico expresado diferencialmente es una enzima, entonces se puede evaluar el efecto generado sobre

25 la actividad biológica detectando un nivel de actividad enzimática asociado al producto génico expresado diferencialmente. El análisis funcional será seleccionado según el producto génico expresado diferencialmente. En general, cuando se aumenta la expresión del gen expresado diferencialmente en una célula cancerosa, los agentes de interés son aquellos que disminuyen la actividad del producto génico expresado diferencialmente.

Los análisis descritos *infra* se pueden adaptar en las realizaciones de análisis de rastreo de la invención. Los ejemplos de análisis útiles en la detección de agentes candidatos incluyen, pero no se limitan a, análisis basados en hibridación (p. ej., el uso de sondas de ácido nucleico o de cebadores para evaluar los niveles de expresión), análisis basados en anticuerpos (p. ej., para evaluar los niveles de productos génicos polipeptídicos); los análisis de unión (p. ej., para detectar la interacción de un agente candidato con un polipéptido expresado diferencialmente, pudiendo ser

35 estos análisis competitivos en los que esté disponible un ligando natural o sintético para el polipéptido), y similares. Otros ejemplos de análisis incluyen, pero no se limitan necesariamente a, análisis de proliferación celular, análisis de noqueado antisentido, análisis para detectar la inhibición del ciclo celular, análisis de inducción de la muerte/apoptosis celular, y similares. Generalmente, tales análisis se llevan a cabo *in vitro*, pero muchos análisis se pueden adaptar para que sean *in vivo*, p. ej., en un modelo animal del cáncer.

#### *Identificación de agentes terapéuticos*

En otra realización, la invención contempla la identificación de genes y de productos génicos expresados diferencialmente como dianas terapéuticas. En algunos aspectos, esto es lo opuesto a los análisis descritos anteriormente para la identificación de agentes que tengan actividad en la modulación (p. ej., en la disminución o el aumento) de la actividad de un producto génico expresado diferencialmente.

En esta realización, las dianas terapéuticas se identifican examinando el/los efecto/s de agente del que se puede demostrar o del que se ha demostrado que modula un fenotipo canceroso (p. ej., inhibir o suprimir o prevenir el desarrollo de un fenotipo canceroso). Tales agentes son generalmente denominados en la presente memoria como un “agente anticancerígeno”, englobando éstos los agentes quimioterapéuticos. Por ejemplo, el agente puede ser un oligonucleótido antisentido que sea específico de un transcrito génico seleccionado. Por ejemplo, el oligonucleótido antisentido puede ser una secuencia correspondiente a una secuencia de un gen expresado diferencialmente descrito

55 en la presente memoria, p. ej., una secuencia de SEQ ID NO: 1.

Los análisis de identificación de dianas terapéuticas se pueden realizar en una variedad de modos usando procedimientos que son conocidos por cualquier experto habitual en la técnica. Por ejemplo, se pone en contacto una célula cancerosa de análisis que expresa o sobre-expresa un gen expresado diferencialmente con un agente anticancerígeno,

60 y se evalúa el efecto generado sobre un fenotipo canceroso y una actividad biológica del producto génico candidato. La actividad biológica del producto génico candidato se puede analizar examinando, por ejemplo, la modulación de la expresión de un gen codificante del producto génico candidato (p. ej., según lo detectado mediante, por ejemplo, un aumento o una disminución de los niveles de transcritos o de los niveles de polipéptidos) o la modulación de una actividad enzimática u otra actividad del producto génico. El fenotipo canceroso puede ser, por ejemplo, la proliferación celular, la pérdida de inhibición por contacto del crecimiento (p. ej., la formación de colonias), el crecimiento tumoral (*in vitro* o *in vivo*), y similares. Alternativamente o además, se puede evaluar el efecto de la modulación de una actividad biológica del gen diana candidato sobre la muerte/apoptosis celular o la regulación del ciclo celular.

La inhibición o la supresión de un fenotipo canceroso, o un aumento de la muerte/apoptosis celular como resultado de la modulación de la actividad biológica de un producto génico candidato indica que el producto génico candidato es una diana adecuada para la terapia contra el cáncer. Los análisis descritos *infra* se pueden adaptar fácilmente a los análisis de identificación de diana terapéuticas. Generalmente, tales análisis se llevan a cabo *in vitro*, pero es posible  
 5 adaptar muchos análisis para que sean *in vivo*, p. ej., en un modelo animal del cáncer apropiado aceptado por la técnica.

#### *Identificación de análogos y antagonistas de péptidos*

10 Los polipéptidos codificados por genes expresados diferencialmente identificados en la presente memoria se pueden usar para rastrear genotecas de péptidos para identificar parejas de unión, tales como receptores, de entre los polipéptidos codificados. Las genotecas de péptidos se pueden sintetizar según los procedimientos conocidos en la técnica (véase, p. ej., el documento USPN 5.010.175 y el documento WO 91/17823).

15 Los agonistas o antagonistas de los polipéptidos de la invención se pueden rastrear usando cualquier procedimiento conocido en la técnica, tal como la transducción de señales, la unión de anticuerpos, la unión de receptores, los análisis mitogénicos, los análisis de quimiotaxis, etc. Lo ideal es que las condiciones de análisis se parezcan a las condiciones bajo las que la actividad nativa se muestra *in vivo*, es decir, bajo un pH fisiológico, la temperatura y la fuerza iónica. Los agonistas o antagonistas adecuados mostrarán una fuerte inhibición o aumento de la actividad nativa a concentraciones  
 20 que no causan efectos secundarios tóxicos en el sujeto. Los agonistas o antagonistas que compiten en la unión al polipéptido nativo pueden necesitar concentraciones iguales o mayores que la concentración nativa, mientras que los inhibidores capaces de unirse irreversiblemente al polipéptido se pueden añadir en concentraciones en el orden de la concentración nativa.

25 Tales rastreo y experimentación pueden conducir a la identificación de una pareja de unión polipeptídica, tales como un receptor, codificado por un gen o un ADNc correspondiente a un polinucleótido descrito en la presente memoria, y al menos un agonista o antagonista peptídico de la pareja de unión. Tales agonistas y antagonistas se pueden usar para modular, aumentar o inhibir la función de los receptores en las células de las que el receptor es nativo, o en las células que poseen el receptor como resultado de la ingeniería genética. Además, si el receptor comparte características  
 30 biológicamente importantes con un receptor conocido, la información acerca de la unión de agonistas/antagonistas puede facilitar el desarrollo de mejores agonistas/antagonistas del receptor conocido.

#### *Composiciones farmacéuticas y usos terapéuticos*

35 Las composiciones farmacéuticas de la invención pueden comprender polipéptidos, anticuerpos o polinucleótidos (incluyendo nucleótidos antisentido y ribozimas) de la invención reivindicada en una cantidad terapéuticamente eficaz. El término “cantidad terapéuticamente eficaz” como se usa en la presente memoria se refiere a una cantidad de un agente terapéutico para tratar, mejorar o prevenir un enfermedad o una condición deseada, o para mostrar un efecto  
 40 terapéutico o preventivo detectable. El efecto se puede detectar mediante, por ejemplo, marcadores químicos o niveles de antígenos. Los efectos terapéuticos también incluyen la reducción de síntomas físicos, tales como la disminución de la temperatura corporal. La cantidad eficaz exacta para un sujeto depende del tamaño y de la salud del sujeto, de la naturaleza y del grado de la condición, y de la composición terapéutica o de la combinación de composiciones terapéuticas seleccionada para su administración. De este modo, no es útil especificar una cantidad eficaz exacta por  
 45 anticipado. Sin embargo, la cantidad eficaz para una determinada situación se determina mediante la experimentación rutinaria y depende de la opinión del profesional clínico. A efectos de la presente invención, una dosis eficaz será generalmente de aproximadamente 0,01 mg/kg a 50 mg/kg o de 0,05 mg/kg a aproximadamente 10 mg/kg de los constructos de ADN en el individuo al que se administra.

50 Una composición farmacéutica también puede contener un vehículo farmacéuticamente aceptable. El término “vehículo farmacéuticamente aceptable” se refiere a un vehículo para la administración de un agente terapéutico, tal como anticuerpos o un polipéptido, genes y otros agentes terapéuticos. El término se refiere a cualquier vehículo farmacéutico que no induzca por sí mismo a la producción de anticuerpos dañinos para el individuo que recibe la composición, y que pueda ser administrado sin producir una toxicidad indebida. Los vehículos adecuados pueden ser macromoléculas grandes, lentamente metabolizadas tales como, proteínas, polisacáridos, ácidos polilácticos, ácidos poliglicólicos,  
 55 aminoácidos poliméricos, copolímeros de aminoácidos y partículas víricas inactivas. Tales vehículos son conocidos por aquellos expertos habituales en la técnica. Los vehículos farmacéuticamente aceptables de composiciones terapéuticas pueden incluir líquidos tales como agua, solución salina, glicerol y etanol. Las sustancias auxiliares, tales como agentes humectantes o emulsionantes, sustancias tamponadores del pH, y similares, también pueden estar presentes  
 60 en tales vehículos.

Comúnmente, las composiciones terapéuticas se preparan como inyectables, bien como soluciones líquidas o como suspensiones; también se pueden preparar formas sólidas adecuadas para una solución en, o una suspensión en, vehículos líquidos antes de la inyección. Los liposomas se incluyen en la definición de un vehículo farmacéuticamente  
 65 aceptable. Las sales farmacéuticamente aceptables también pueden estar presentes en la composición farmacéutica, p. ej., sales de ácidos minerales tales como clorhidratos, bromhidratos, fosfatos, sulfatos y similares; y las sales de ácidos orgánicos tales como acetatos, propionatos, malonatos, benzoatos y similares. En “Remington’s Pharmaceutical Sciences” (Mack Pub. Co., N. J. 1991), se tratan en profundidad los excipientes farmacéuticamente aceptables.

*Procedimientos de administración.* Una vez formuladas, las composiciones de la invención se pueden (1) administrar directamente al sujeto (p. ej., como polinucleótido o polipéptidos); (2) administrar *ex vivo* a células derivadas del sujeto (p. ej., como en una terapia génica *ex vivo*). La administración directa de las composiciones se realizará generalmente mediante la inyección parenteral, p. ej., subcutánea, intraperitoneal, intravenosa o intramuscularmente, intratumoral o en el espacio intersticial de un tejido. Otros modos de administración incluyen la administración oral y pulmonar, los supositorios y las aplicaciones transdérmicas, las agujas y las pistolas de genes o los hidropulverizados. El tratamiento de dosificación puede ser un programa de una sola dosis o un programa de múltiples dosis.

Los procedimientos para la administración *ex vivo* y la reimplantación de las células transformadas en un sujeto son conocidos en la técnica y se describen en, p. ej., la publicación internacional n.º: WO 93/14778. Los ejemplos de células útiles en aplicaciones *ex vivo* incluyen, por ejemplo, células madre, particularmente células hematopoéticas, células linfáticas, macrófagos, células dendríticas o células tumorales. Generalmente, la administración de ácidos nucleicos para aplicaciones tanto *ex vivo* como *in vitro* se puede realizar mediante, por ejemplo, una transfección mediada por dextrano, precipitación de fosfato de calcio, transfección mediada por polibreno, fusión protoplástica, electroporación, encapsulación del/de los polinucleótido/s en liposomas y microinyección directa del ADN en núcleos, siendo todas conocidas en la técnica.

Una vez que se descubre que un gen correspondiente a un polinucleótido de la invención está correlacionado con un trastorno proliferativo, tal como neoplasia, displasia e hiperplasia, es posible tratar el trastorno mediante la administración de un agente terapéutico basado en el polinucleótido proporcionado, el polipéptido correspondiente u otra molécula correspondiente (p. ej., antisentido, ribozima, etc.).

La dosis y el medio de administración de las composiciones farmacéuticas de la invención se determinan en base a cualidades específicas de la composición terapéutica, de la condición, la edad, el peso del paciente, la progresión de la enfermedad y otros factores relevantes. Por ejemplo, la administración de las composiciones terapéuticas de los polinucleótidos de la invención incluye la administración local o sistémica, incluyendo la inyección, la administración oral, la pistola de partículas o la administración cateterizada, y la administración tópica. Preferiblemente, la composición de polinucleótido terapéutica contiene un constructo de expresión que comprende un promotor ligado operativamente a un polinucleótido de al menos 12, 22, 25, 30 ó 35 nt contiguos del polinucleótido revelado en la presente memoria.

Se pueden usar diversos procedimientos para administrar la composición terapéutica directamente en un punto específico del cuerpo. Por ejemplo, se localiza una pequeña lesión metastásica y se inyecta varias veces la composición terapéutica en diversos lugares diferentes del cuerpo del tumor. Alternativamente, se identifican las arterias de las que se sirve al tumor, y se inyecta la composición terapéutica en tal arteria, con el fin de administrar la composición directamente en el tumor. Los tumores que tienen un centro necrótico son aspirados, inyectándose la composición directamente en el ahora centro vacío del tumor. La composición antisentido se administra directamente a la superficie del tumor, por ejemplo, mediante la aplicación tópica de la composición. Se utiliza la formación de imágenes por rayos X para ayudar en ciertos procedimientos de administración anteriores.

También se puede usar la administración dirigida mediada por receptores de composiciones terapéuticas que contienen un polinucleótido antisentido, polinucleótidos subgenómicos o anticuerpos frente a tejidos específicos. Las técnicas de administración de ADN mediadas por receptores se describen en, por ejemplo, Findeis *et al.*, *Trends Biotechnol.* (1993) 11: 202; Chiou *et al.*, "Gene Therapeutics: Methods And Applications Of Direct Gene Transfer" (J. A. Wolff, ed.) (1994); Wu *et al.*, *J. Biol. Chem.* (1988) 263: 621; Wu *et al.*, *R Biol. Chem.* (1994) 269: 542; Zenke *et al.*, *Proc. Natl. Acad. Sci.* (EE.UU.) (1990) 87: 3655; Wu *et al.*, *J Biol. Chem.* (1991) 266: 338. Las composiciones terapéuticas que contienen un polinucleótido se administran en una variedad de aproximadamente 100 ng a aproximadamente 200 mg de ADN para una administración local en un protocolo de terapia génica. También se pueden usar intervalos de concentración de aproximadamente 500 ng a aproximadamente 50 mg, de aproximadamente 1  $\mu$ g a aproximadamente 2 mg, de aproximadamente 5  $\mu$ g a aproximadamente 500  $\mu$ g, y de aproximadamente 20  $\mu$ g a aproximadamente 100  $\mu$ g de ADN durante un protocolo de terapia génica. Los factores tales como el procedimiento de acción (p. ej., para aumentar o inhibir los niveles del producto génico codificado) y la eficacia de transformación y expresión son consideraciones que afectarán a la dosis requerida para la eficacia final de los polinucleótidos subgenómicos antisentido. Cuando se desea una mayor expresión sobre una zona más amplia de tejido, se pueden necesitar mayores cantidades de polinucleótidos subgenómicos antisentido o las mismas cantidades readministradas en un protocolo sucesivo de administraciones, o diversas administraciones a diferentes porciones de tejido adyacentes o cercanas de, por ejemplo, un punto tumoral, para efectuar un resultado terapéutico positivo. En todos los casos, la experimentación rutinaria en pruebas clínicas determinará los intervalos específicos para el efecto terapéutico óptimo. Para los genes relacionados con polinucleótidos que codifican polipéptidos o proteínas con actividad antiinflamatoria, en el documento USPN 5.654.173, se describe un uso, dosis y administración adecuados.

Los polinucleótidos terapéuticos y los polipéptidos de la presente invención se pueden administrar usando vehículos de administración de genes. El vehículo de administración de genes puede ser de origen vírico o no vírico (véase, en general, Jolly, "Cancer Gene Therapy" (1994) 1:51; Kimura, "Human Gene Therapy" (1994) 5: 845; Connolly, "Human Gene Therapy" (1995) 1:185; y Kaplitt, "Nature Genetics" (1994) 6:148). Se puede inducir la expresión de tales secuencias codificantes usando promotores de mamífero o heterólogos endógenos. La expresión de la secuencia codificante puede ser bien constitutiva o regulada.

Los vectores basados en virus para su administración a un polinucleótido deseado y la expresión en una célula deseada son conocidos en la técnica. Los ejemplos de vehículos basados en virus incluyen, pero no se limitan a, retrovirus recombinantes (véanse, p. ej., los documentos WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; USPN 5.219.740; WO 93/11230; WO 93/10218; USPN 4.777.127; GB Patente N.º 2.200.651; EP 0 345 242; y WO 91/02805); los vectores basados en alfa-virus (p. ej., vectores del virus Sindbis, del virus Semliki forest (ATCC VR-67; ATCC VR-1247), del virus Ross River (ATCC VR-373; ATCC VR-1246) y del virus de la encefalitis equina venezolana (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), y del virus adeno-asociado (VAA) (véanse, p. ej., los documentos WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 y WO 95/00655). También se puede emplear la administración de ADN ligado a adenovirus muerto según lo descrito en Curiel, *Hum. Gene Ther.* (1992) 3:147.

También se pueden emplear vehículos y procedimientos de administración, incluyendo, pero no limitándose a, ADN condensado policationico ligado o no ligado a adenovirus muerto solo (véase, p. ej., Curiel, *Hum. Gene Ther.* (1992) 3:147); ADN ligado a ligandos (véase, p. ej., Wu, *J. Biol. Chem.* (1989) 264:16985); células de vehículos de administración de células eucariotas (véase, p. ej., los documentos USPN 5.814.482; WO 95/07994; WO 96/17072; WO 95/30763; y WO 97/42338) y la neutralización de la carga nucleica o la fusión con membranas celulares. También se puede emplear ADN desnudo. En los documentos WO 90/11092 y USPN 5.580.859, se describen ejemplos de procedimientos de introducción de ADN desnudo. En los documentos USPN 5.422.120; WO 95/13796; WO 94/23697; WO 91/14445; y EP 0524968, se describen liposomas que pueden actuar como vehículos de administración de genes. En Philip, *Mol. Cell Biol.* (1994) 14:2411, y en Woffendin, *Proc. Natl. Acad. Sci.* (1994) 91:1581, se describen otros enfoques.

Otra administración no vírica adecuada para su uso incluye los sistemas de administración mecánica tales como el enfoque descrito en Woffendin *et al.*, *Proc. Natl. Acad. Sci.* EE.UU. (1994) 91 (24):11581. Además, la secuencia codificante y el producto de expresión de tal se puede administrar a través de la deposición de materiales de hidrogel fotopolimerizados o el uso de radiación ionizante (véase, p. ej., los documentos USPN 5.206.152 y WO 92/11033). Otros procedimientos convencionales para la administración de genes que se pueden usar para administrar la secuencia codificante incluyen, por ejemplo, el uso de un pistola de mano de partículas de transferencia de genes (véase, p. ej., el documento USPN 5.149.655); el uso de radiación ionizante para activar el gen transferido (véanse, p. ej., los documentos USPN 5.206.152 y WO 92/11033).

A continuación se ilustrará la presente invención en referencia a los siguientes ejemplos que exponen las realizaciones particularmente ventajosas. Sin embargo, debería indicarse que estas realizaciones son ilustrativas y no pretenden estar construidas para restringir la invención de ningún modo.

35

## Ejemplos

Los siguientes ejemplos se ofrecen principalmente a efectos ilustrativos. Será fácilmente comprensible para aquellos expertos en la técnica que las formulaciones, las dosis, los procedimientos de administración y otros parámetros de esta invención pueden ser modificados o sustituidos en diversos modos sin alejarse del espíritu ni del ámbito de la invención.

### Ejemplo 1

#### *Fuente de materiales biológicos y perspectiva general de los polinucleótidos expresados por los materiales biológicos*

Para identificar los genes que están expresados diferencialmente en el cáncer de colon, se prepararon genotecas de ADNc a partir de varias líneas celulares y fuentes tisulares diferentes. La tabla 1 proporciona un resumen de estas genotecas, incluyendo el nombre abreviado de la genoteca (usado en lo sucesivo), la fuente del ARNm usada para preparar la genoteca de ADNc, el apodo de la genoteca que se usa en las tablas que figuran a continuación (entre comillas) y el número aproximado de clones de la genoteca. Las genotecas de ADNc se prepararon según los procedimientos conocidos en la técnica, y las secuencias de los insertos de ADNc se determinaron usando procedimientos conocidos.

55

60

65

# ES 2 305 087 T3

TABLA 1

*Descripción de genotecas de ADNc*

Genoteca	Descripción	Número de clones
1	Línea de células de colon humano Km12L4: elevado potencial metastásico (derivada de Km12C)	308731
2	Línea de células de colon humano Km12: bajo potencial metastásico	284771
3	Línea de células de cáncer de mama humana MDA-MB-231: elevado potencial metastásico; micrometástasis en pulmón.	326937
4	Línea de células de cáncer de mama humana MCF7: No metastásica	318979
8	Línea de células de cáncer de pulmón humano MV-522: elevado potencial metastásico	223620
9	Línea de células de cáncer de pulmón humano UCP-3: bajo potencial metastásico	312503
12	Células endoteliales microvasculares humanas (HMEC) – SIN TRATAR (genoteca de ADNc de (OligodT) de PCR)	41938
13	Células endoteliales microvasculares humanas (HMEC) – TRATADAS CON FCFb (genoteca de ADNc de (OligodT) de PCR).	42100
14	Células endoteliales microvasculares humanas (HMEC) – TRATADAS CON FCEV (genoteca de ADNc de (OligodT) de PCR).	42825
15	Colon normal – Paciente UC N.º2 (genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	282718
16	Tumor de colon–Paciente UC N.º2 (genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	298829
17	Metástasis de hígado procedente del tumor de colon del paciente UC N.º2 (genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	303462
18	Colon normal – Paciente UC N.º3 (genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	36216
19	Tumor de colon–Paciente UC N.º3 (genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	41388

20	Metástasis de hígado procedente del tumor de colon	30956
5	del paciente UC N.º3 (genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	
21	Células GRRpz derivadas de epitelio de próstata	164801
10	normal	
22	Células WOca derivadas de epitelio de cáncer de	162088
	próstata de grado 4 de Gleason	
15	Epitelio de pulmón normal del paciente n.º 1006	306198
	(genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	
20	24 Tumor primario, carcinoma de células grandes del	309349
	paciente n.º 1006 (genoteca de ADNc de (OligodT) de PCR MICRODISECCIONADA)	
25	25 Epitelio de próstata normal del paciente IF97-26811	279437
	26 Epitelio de cáncer de próstata Gleason 3+3 del	269366
	paciente IF97-26811	

La línea celular KM12L4 deriva de la línea celular KM12C (Morikawa, *et al.*, *Cancer Research* (1988) 48:6863). La línea celular KM12C, que es pobremente metastásica (poco metastásica) se estableció en cultivo a partir de una muestra quirúrgica en fase B<sub>2</sub> de Dukes (Morikawa *et al.* *Cancer Res.* (1988) 48:6863). La KML4-A es una sublínea altamente metastásica derivada de KM12C (Yeatman *et al.* *Nucl. Acids. Res.* (1995) 23:4007; Bao-Ling *et al.* *Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) 21:3269). Las líneas celulares derivadas de KM12C y de KM12C (p. ej., la KM12L4, la KM12L4-A, etc.) son muy reconocidas en la técnica como una línea celular modelo para el estudio del cáncer de colon (véase, p. ej., Moriakawa *et al.*, *supra*; Radinsky *et al.* *Clin. Cancer Res.* (1995) 1:19; Yeatman *et al.*, (1995) *supra*; Yeatman *et al.* *Clin. Exp. Metastasis* (1996) 14: 246).

La línea celular MDA-MB-231 se aisló originariamente de efusiones pleurales (Cailleau, *J. Natl. Cancer. Inst.* (1974) 53:661), y es de un elevado potencial metastásico, y forma adenocarcinoma pobremente diferenciado de grado II en ratones desnudos que coincide con el carcinoma de mama. La línea celular MCF7 fue derivada de una efusión pleural de un adenocarcinoma de mama y no es metastásica. Estas líneas celulares están muy reconocidas en la técnica como modelos para el estudio del cáncer de mama y de pulmón humano (véase, p. ej., Chandrasekaran *et al.*, *Cancer Res.* (1979) 39:870; Gastpar *et al.*, *J. Med Chem* (1998) 41: 4965; Ranson *et al.*, *Br J Cancer* (1998) 77: 1586; Kuang *et al.*, *Nucleic Acids Res* (1998) 26:1116. Las muestras de las genotecas 15-20 están derivadas de dos pacientes distintos (UC N.º 2 y UC N.º 3). Las líneas celulares GRRpz y WOca fueron proporcionadas por el Dr. Donna M. Peehl, Departamento de Medicina, Escuela Universitaria de Medicina de Stanford. La GRRpz fue derivada del epitelio de próstata normal. La línea celular WOca es una línea celular de grado 4 de Gleason.

Cada genoteca está compuesta por una colección de clones de ADNc que a su vez son representativos de los ARNm expresados en la fuente de ARNm indicada. Para facilitar el análisis de los millones de secuencias de cada genoteca, las secuencias se asignan a grupos. El concepto “grupo de clones” deriva de una clasificación/agrupamiento de clones de ADNc basado en su patrón de hibridación con un panel de sondas de oligonucleótido de aproximadamente 300 7pb (véase Drmanac *et al.*, *Genomics* (1996) 37 (1):29). Los clones de ADNc aleatorios de una genoteca de tejidos se hibridan en condiciones restrictivas moderadas con oligonucleótidos de 300 7pb. Cada oligonucleótido tiene cierta medida de hibridación específica con el clon específico. La combinación de 300 de estas medidas de hibridación para 300 sondas equivale a la “firma de hibridación” para un clon específico. Los clones con una secuencia similar tendrán firmas de hibridación similares. Mediante el desarrollo de un algoritmo de clasificación/agrupamiento para analizar estas firmas, se pueden identificar grupos de clones de una genoteca y reunirles informáticamente. Estos grupos de clones se denominan “grupos”.

Dependiendo de las condiciones restrictivas de la selección en el algoritmo (similares a las condiciones restrictivas de la hibridación en un protocolo clásico de rastreo de ADNc de genotecas), se puede controlar la “pureza” de cada grupo. Por ejemplo, se pueden producir artefactos de agrupamiento en el agrupamiento informático justo como se pueden producir artefactos en el rastreo de “etiquetas húmedas” de una genoteca de ADNc con fragmentos de ADNc



de 400 pb, incluso en las condiciones más restrictivas. Las condiciones restrictivas usadas en la implementación de un grupo en la presente memoria proporcionan grupos de clones que son en general del mismo ADNc o de ADNc muy relacionados. Los clones muy relacionados pueden ser el resultado de clones de diferente longitud del mismo ADNc, clones muy relacionados de familias de genes altamente relacionadas o variantes de empalme del mismo ADNc.

Se evaluó la expresión diferencial para un grupo seleccionado determinando primero el número de clones de ADNc correspondiente al grupo seleccionado en la primera genoteca (clones de la 1ª) y determinando el número de clones de ADNc correspondiente al grupo seleccionado de la segunda genoteca (clones de la 2ª). La expresión diferencial del grupo seleccionado en la primera genoteca en relación con la segunda genoteca se expresa como una “proporción” del porcentaje de expresión entre las dos genotecas. En general, la “proporción” se calcula: 1) calculando el porcentaje de expresión del grupo seleccionado en la primera genoteca dividiendo el número de clones correspondiente a un grupo seleccionado en la primera genoteca entre el número total de clones analizado de la primera genoteca; 2) calculando el porcentaje de expresión del grupo seleccionado en la segunda genoteca dividiendo el número de clones correspondiente al grupo seleccionado en la segunda genoteca entre el número total de clones analizado de la segunda genoteca; 3) dividiendo el porcentaje de expresión calculado de la primera genoteca entre el porcentaje de expresión calculado de la segunda genoteca. Si el “número de clones” correspondiente a un grupo seleccionado en una genoteca es cero, el valor se establece en 1 para ayudar en el cálculo. La fórmula usada para calcular la proporción tiene en cuenta la “profundidad” de cada una de las genotecas que se está comparando, i.e., el número total de clones analizado en cada genoteca.

Como resultado de esta comparación entre genotecas, se identificaron 17 polinucleótidos, enumerados como las SEQ ID NOS: 1, 3, 5, 7, 9, 11-13, 15, 16, 18, 20, 22, 24, 26, 27 y 29 en el listado secuencial anexo y resumidos en la tabla 2, como los correspondientes a los genes expresados diferencialmente en los tejidos de pacientes de cáncer de colon. La tabla 2 proporciona: 1) el número de identificación secuencial (“SEQ ID NO del polinucleótido”) asignado a cada secuencia para su uso en la presente memoria; 2) el número de identificación de grupo (“GRUPO”); 3) el número de identificación de candidato; 4) el número CHIR (que sirve como la referencia cruzada de los oligos antisentido tratados más adelante) teniendo, por ejemplo, CHIR7 los oligos correspondientes CHIR7-2AS (antisentido) y CHIR7-RC (de control inverso); 5) el nombre de secuencia (“NOMBRE DE SEQ”) usado como identificador interno de la secuencia; 6) el nombre asignado al clon del que fue aislada la secuencia (“ID DEL CLON”); 7) el primer nucleótido de los condones de inicio y terminación de los marcos de lectura abiertos (“inicio de ORF” y “terminación de ORF”); y 8) el número de identificación de secuencia (“SEQ ID NO del polipéptido codificado”) asignado al polipéptido codificado, cuando sea apropiado. Como los polinucleótidos proporcionados representan transcritos de ARNm parciales, dos o más polinucleótidos de la invención pueden representar diferentes regiones del mismo transcripto de ARNm y del mismo gen. De este modo, si se identifican dos o más secuencias como pertenecientes al mismo clon, entonces se puede usar cualquier secuencia para obtener el ARNm o el gen de longitud completa.

(Tabla pasa a página siguiente)

# ES 2 305 087 T3

TABLA 2

*Identificación y caracterización de las secuencias de polinucleótidos*

SEQ ID NO	GRUPO	ID de candidato	CHIR	NOMBRE DE SEQ	ORF Inicio	Terminación	SEQ ID NO del polipéptido codificado
1	719	196	CHIR-7	SK1	21	396	2
3	9083	181	CHIR-8	SK2	219	693	4
5	115762	188	CHIR-16	SK5	5	1760	6
7	1665	195	CHIR-9	1665 largo	78	642	8
9	1665	195	CHIR-8	1665 corto	79	232	10
11	2334			SK8 parcial			
12	2334			SK8 de long. comp.			
13	3376	118	CHIR-11	SK19	79	376	14
15	376130			Junc2	181, 363, 731	361, 542, 911	
16	402380	202	CHIR-33	XD4	16	538	17
18	726682	198	CHIR-43	XD1	2	551	19
20	552930	174	CHIR-42	XD7	240	585	21
22	454001	161	CHIR-29	XD10	53	1700	23
24	378805	163	CHIR-31	XD11	10	400	25
26	374641	160	CHIR-32	374641 largo (Junc4)	33, 420	183, 615	
27	374641	160	CHIR-32	374641 corto (XD6)	324	519	28
29	374641	160	CHIR-32	374641 electrónico	40, 388	190, 583	

La tabla 3 resume los polinucleótidos que se corresponden con los genes que se expresan diferencialmente en el tejido de colon de un solo paciente.

TABLA 3

SEQ ID NO	GRUPO	Clones (Lib15) normales	Clones (Lib16) tumores	Clones (Lib17) altam. Met.	(Lib16/Lib15) tumores/ normales	(Lib17/Lib15) altam met./ normales	(Lib17/Lib16) altam. Met./ tumores
1	719	0	20	27	20	27	1
3	9083	0	10	14	10	14	1
5	115762	0	6	7	6	7	1
7	1665	4	14	20	3,5	5	1
12	2334	0	6	1	6	1	0
13	3376	3	20	19	7	6	1
15	376130	0	9	15	9	15	2
16	402380	0	15	2	15	2	0
18	726682	0	52	0	52	0	0
20	552930	1	14	2	14	2	0
22	454001	0	8	13	8	13	2
24	378805	1	12	12	12	12	1
26	374641	9	47	129	5	14	3

### Ejemplo 2

#### *Análisis y caracterización de los polinucleótidos de la invención*

Varios de los polinucleótidos proporcionados contienen uno o más marcos de lectura abiertos putativos (ORF) codificantes de un producto génico. En la tabla 2, se enumeran los sitios de inicio y de terminación de estos ORF.

La SEQ ID NO: 15 contiene tres ORF. El primer ORF se extiende del nucleótido 181 hasta el nucleótido 361. El segundo ORF se extiende del nucleótido 363 hasta el nucleótido 542. El tercer ORF se extiende del nucleótido 731 hasta el nucleótido 911.

La SEQ ID NO: 26 contiene una secuencia de inserción de 39 nucleótidos (del nucleótido 269 al nucleótido 307) y dos ORF. El primer ORF se extiende del nucleótido 33 hasta el nucleótido 183. El segundo ORF se extiende del nucleótido 420 hasta el nucleótido 615.

La SEQ ID NO: 29 es una secuencia electrónica según la RACE de 5' y contiene dos ORF. El primer ORF se extiende del nucleótido 40 hasta el nucleótido 190. El segundo ORF se extiende del nucleótido 388 hasta el nucleótido 583.

### Ejemplo 3

#### *Miembros de familias de proteínas*

Las traducciones de los polinucleótidos proporcionados se alinearon con perfiles de aminoácidos que definen bien familias de proteínas o motivos comunes. Se descubrió que varios de los polinucleótidos de la invención codifican a polipéptidos que tienen características de un polipéptido perteneciente a una familia de proteínas conocida (y que, por consiguiente, representan nuevos miembros de estas familias de proteínas) y/o que comprenden un dominio funcional conocido. La similitud entre una secuencia problema y una familia de proteínas o un motivo se determinó (a) comparando la secuencia problema con el perfil y/o (b) alineando la secuencia problema con los miembros de la familia o el motivo.

Todos los blancos de perfil se describen más detalladamente a continuación. La tabla 4 proporciona la correspondiente SEQ ID NO de los polinucleótidos proporcionados que codifican a los productos génicos con similitud o identidad con las secuencias perfil. En la tabla 4, también se indica la similitud (fuerte o débil). Los acrónimos de los perfiles (proporcionados entre paréntesis) son los usados para identificar el perfil en las bases de datos Pfam y Prosite. La base de datos Pfam tiene acceso a través de cualquiera de las URL: <http://pfam.wustl.edu/index.html>; <http://www.sanger.ac.uk/Software/Pfam/>; y <http://www.cgr.ki.se/Pfam/>. A la base de datos Prosite se puede acceder en <http://www.expasy.ch/prosite/>.

TABLA 4

*Blancos de perfil*

	SEQ ID NO	GRUPO	Perfil	Descripción	Similitud
5	1	719		Glicosil-hidrolasa	débil
	3	9083	ANK	Repeticiones de anquirina	fuerte
10	5	115762	7tm_1	Receptor de 7 dominios transmembrana (familia de las rodopsinas)	débil
	11	2334	EFhand	EF-hand	fuerte
15	12	2334	Efhand	EF-hand	fuerte
	15	376130		Proteasa/integrasa retrógrada endógena	
	16	402380	Rrm	Motivo de reconocimiento de ARN (dominio aka RRM, RBD o RNP)	

20 *Familia 5 de las glicosil-hidrolasas (GLYCOSYL HYDROL F5; N.º de acceso de Pfam*

*PS00659; PDOC00565*). La SEQ ID NO: 1 corresponde a un gen codificante de un polipéptido que tiene homología con los polipéptidos de la familia 5 de las glicosil-hidrolasas (Henrissat *Biochem. J* (1991) 280: 309-316) (también conocida como la familia A de las celulasas (Henrissat *et al. Gene* (1989) 81: 83-95)). Los miembros de esta familia participan en la degradación de la celulosa y los xilanos, y se encuentran por lo general en las bacterias, los hongos y la levadura. El patrón de consenso para los miembros de esta familia es: [LIV]-[LIVMFYWGGA](2)-[DNEQG]-[LIVMGST]-x-N-E-[PV]-[RHDNSTLIVFY] (siendo E un residuo de sitio activo putativo).

30 La SEQ ID NO: 1 corresponde a un gen codificante de un miembro de una de las familias de las glicosil-hidrolasas (Henrissat *et al Biochem. J* (1993) 293:781-788). Estas enzimas contienen al menos un residuo de ácido glutámico conservado (o residuo de ácido aspártico) del que se ha demostrado que está directamente implicado en la escisión de enlaces glicosídicos mediante su actuación como nucleófilo.

35 *Repeticiones Ank (ANK; N.º de acceso de Pfam PF0023)*. La SEQ ID NO: 3 se corresponde con un gen codificante de una proteína que contiene repeticiones Ank. El motivo de anquirina es una secuencia de 33 aminoácidos situada tras la proteína anquirina que tiene 24 motivos de 33 aminoácidos en tándem. Las repeticiones de Ank se identificaron inicialmente en la proteína de control del ciclo celular cdc10 (Breeden *et al., Nature* (1987) 329:651). Las proteínas que contienen repeticiones de anquirina incluyen anquirina, miotropina, proteínas I-kappaB, la proteína del ciclo celular cdc10, el receptor Notch (Matsuno *et al., Development* (1997) 124 (21): 4265); la G9a (o BAT8) de la región de la clase III del complejo de histocompatibilidad principal (*Biochem J.* 290:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14 y SW16. Las funciones de las repeticiones de anquirina son compatibles con un papel en las interacciones entre proteínas (Bork, *Proteins* (1993) 17(4): 363; Lambert y Bennet, *Eur. J. Biochem.* (1993) 211:1; Kerr *et al., Current Op. Cell Biol.* (1992) 4: 496; Bennet *et al., J. Biol. Chem.* (1980) 255: 6424).

45 *Proteínas integrales de la membrana de siete dominios transmembrana: familia de las rodopsinas (7tm 1; N.º de acceso de Pfam PF000011)*. La SEQ ID NO: 3 corresponde a un gen codificante de un polipéptido que es un miembro de la familia de las rodopsinas de receptores de 7 dominios transmembrana (7tm). Los receptores acoplados a proteínas G de la familia de las rodopsinas (7tm) (también denominada R7G) son un extenso grupo de hormonas, neurotransmisores y receptores de la luz que transducen señales extracelulares mediante la interacción con proteínas de unión a nucleótidos de guanina (G) (Strosberg A. D. *Eur. J. Biochem.* (1991) 196: 1; Kerlavage A. R *Curr. Opin. Struct. Biol.* (1991) 1: 394, Probst, *et al., DNA Cell Biol.* (1992) 11: 1, Savarese, *et al., Biochem. J.* (1992) 283:1, <http://www.gcrdb.uthscsa.edu/>, <http://swift.embl-heidelberg.de/7tm/>. El patrón de consenso que contiene el triplete conservado y que también abarca la mayor parte de la tercera hélice transmembrana se usa para detectar esta familia de proteínas ampliamente extendida: [GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM].

60 *EF Hand (EFhand; N.º de acceso de Pfam PF00036)*. Las SEQ ID NO: 11 y 12 se corresponden con los genes codificantes de una proteína de la familia de las proteínas EF-hand. Muchas proteínas de unión a calcio pertenecen a la misma familia evolutiva y comparten un tipo de dominio de unión al calcio conocido como el EF-hand (Kawasaki *et al., Protein. Prof* (1995) 2:305-490). Este tipo de dominio está constituido por doce bucles de residuo flanqueados por ambos lados por doce dominios alfa-helicoidales de residuos. En un bucle EF-hand, el ión de calcio está coordinado en una configuración bipiramidal pentagonal. Los seis residuos implicados en la unión están en las posiciones 1, 3, 5, 7, 9 y 12; estos residuos están denominados por X, Y, Z, -Y, -X y -Z. La invariante Glu o Asp de la posición 12 proporciona dos oxígenos para ligarse al Ca (ligando bidentado). El patrón consenso incluye el bucle EF-hand completo, así como el primer residuo que sigue al bucle y que parece ser siempre hidrófobo: D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW].

*Proteasa/integrasa retroviral endógena.* La SEQ ID NO: 15 corresponde a un gen codificante de un polipéptido que tiene un dominio homólogo a un dominio de proteasa/integrasa de retrovirus endógeno humano de una proteína *pol* retrovítica.

- 5 *Motivo de reconocimiento de ARN* (*rrm*; *N.º de acceso de Pfam*: PF00076). La SEQ ID NO: 16 se corresponde con un gen codificante de un motivo de reconocimiento de ARN también conocido como un dominio RRM, RBD o RNP. Este dominio que es de una longitud aproximada de 90 aminoácidos está contenido en proteínas eucariotas que se unen a ARN monocatenario (Bandziulis *et al. Genes Dev.* (1989) 3:431-437; Dreyfuss *et al. Trends Biochem. Sci.* (1988) 13:86-91). Hay dos regiones en el dominio de unión a ARN que están altamente conservadas: La primera es un segmento hidrófobo de seis residuos (que se denomina el motivo RNP-2), la segunda es un motivo octapeptídico (que se denomina RNP-1 o RNP-CS). El patrón de consenso es: [RK]-G-{EDRKHPCG}-[AGSCI]-[FY]-[LIVA]-x-[FYLM].

#### Ejemplo 4

#### 15 *Detección y cuantificación de los polinucleótidos de la invención*

- Los polinucleótidos de la invención se detectaron y cuantificaron en muestras de tejido de los pacientes mediante una PCR de transcriptasa inversa (RT-PCR). Las amplificaciones del ARN total se realizaron usando el sistema de oscilación térmica LightCycler® (Roche Diagnostics) en una reacción de PCR estándar que contenía los cebadores proporcionados y tinte de unión a ADNbc Verde I SYBR. La amplificación PCR se controló mediante el tinte fluorescente Verde I SYBR, que sólo emite fluorescencia cuando está unido a ADN bicatenario. La especificidad de los productos se verificó mediante un análisis de curvas de fusión.

- 25 *Preparación estándar.* Se transcribieron a la inversa 1 µg de ARN total de placenta humana (Clontech, Palo Alto, CA) a 42°C durante 1 hora, luego se calentó a 94°C durante 5 minutos en un volumen de reacción total de 20 µl (Equipo de síntesis de ADNc 1st-Strand®, Clontech). Se usó la mezcla de reacción como 1 x patrón de molde. Entonces se prepararon diluciones en serie de 1 x el patrón de molde: patrones de molde x 10<sup>-1</sup>, x 10<sup>-2</sup>, x 10<sup>-3</sup>, x 10<sup>-4</sup>, x 10<sup>-5</sup>, x 10<sup>-6</sup>.

- 30 *Preparación de muestras de ARN total.* Las muestras de tejido de los pacientes se enviaron en reactivo de TRIZOL congelado. Se homogeneizaron las muestras en reactivo de TRIZOL. Entonces se añadió cloroformo al ARN aislado, seguido por una precipitación del ARN con isopropanol. Se lavaron los precipitados de ARN con etanol al 75%, se secaron al aire, y luego se disolvieron en agua destilada libre de RNasa. Antes de la transcripción inversa, las muestras de ARN se trataron con DNasa I (libre de RNasa) (2 U/µl, Ambion, Austin, TX) y se aclararon usando el equipo RNeasy Mini Kit (Qiagen, Santa Clarita, CA).

- 35 *RT-PCR.* Las muestras de ARN total fueron transcritas inversamente con cebador de oligo-dT<sub>18</sub> (equipo de síntesis de ADNc 1st-Strand®, Clontech). La PCR se realizó usando los siguientes cebadores específicos del gen:

40	SK1:	Cebador directo	5'-AGGAGTTTCTGAGGACCATGCAC-3'	(SEQ ID NO :
				30)
		Cebador	5'-TCAAGGGTTGGGGATACACACG-3'	(SEQ ID NO :
45		inverso		31)
	SK2:	Cebador directo	5'-CTTGCTTGCTTTCTTCTCTGGC-3'	(SEQ ID NO :
				32)
50		Cebador	5'-AGTCTGGAAATCCACATGACCAAG-3'	(SEQ ID NO :
		inverso		33)
55	SK5:	Cebador directo	5'-CCCAATGAGGAACCTAAAGTTGC-3'	(SEQ ID NO :
				34)
		Cebador	5'-GGTGCCAAATCTGGACTCTTGTC-3'	(SEQ ID NO :
60		inverso		35)
	1665:	Cebador directo	5'-GATCCATTTTCAGCAGTGCTCTG-3'	(SEQ ID NO :
				36)
65		Cebador	5'-CAGTGTTACAGAAGGGGTACTCAC-	(SEQ ID NO :

# ES 2 305 087 T3

		inverso	3'	37)
5	SK8:	Cebador directo	5'-ACGAGAGCGACACGGACAAG-3'	(SEQ ID NO :
				38)
		Cebador	5'-TCTGAGGCTGTGGCAGGTGC-3'	(SEQ ID NO :
		inverso		39)
10	SK19:	Cebador directo	5'-CCAGTCTTTGCCAACTCGTGC-3'	(SEQ ID NO :
				40)
		Cebador	5'-TTCGATCTTCAAACGTGTCCTTG-3'	(SEQ ID NO :
15		inverso		41)
	Junc2:	Cebador directo	5'-TTGGCAACCAGACCAGCATC-3'	(SEQ ID NO :
				42)
20		Cebador	5'-TTTCCCATAGGTGTGAGTGGCG-3'	(SEQ ID NO :
		inverso		43)
25	XD4:	Cebador directo	5'-GACTGGTGTGTTTGTTCGGGGTC-3'	(SEQ ID NO :
				44)
		Cebador	5'-TTTGTCCAAGGCTGCATGGTC-3'	(SEQ ID NO :
30		inverso		45)
	XD1:	Cebador directo	5'-TGCCCTGGTTAAGCCAGAAGTC-3'	(SEQ ID NO :
				46)
35		Cebador	5'-AGCTTCACTTTGGTCTTGACGG-3'	(SEQ ID NO :
		inverso		47)
	XD7:	Cebador directo	5'-GGTCATCTGCATCAAGGTTGGC-3'	(SEQ ID NO :
40				48)
		Cebador	5'-GGTTCGTAACCGTGACTTCAGG-3'	(SEQ ID NO :
		inverso		49)
45	XD10:	Cebador directo	5'-GCATCCTTTTCCAGTCTTCCG-3'	(SEQ ID NO :
				50)
50		Cebador	5'-TGCAGCAAACATGCCTGAGC-3'	(SEQ ID NO :
		inverso		51)
	XD11:	Cebador directo	5'-TGTTCCACGAGCAAAGCATGTG-3'	(SEQ ID NO :
				52)
55		Cebador	5'-ATCCTTCTTCCACTCCCGCTTC-3'	(SEQ ID NO :
		inverso		53)
60	37641:	Cebador directo	5'-TCGGCTTGACTACACTGTGTGG-3'	(SEQ ID NO :
				54)

65

## ES 2 305 087 T3

		Cebador	5'--TACAAAGACCACTGGGAGGCTG--3'	(SEQ ID NO :
		inverso		55)
5	β-	Cebador directo	5'-CGGGAAATCGTGCGTGACATTAAG--3'	(SEQ ID NO :
	actina:			56)
		Cebador	5'-TGATCTCCTTCTGCATCCTGTCTCGG--3'	(SEQ ID NO :
10		inverso		57)
	GAPDH:	Cebador directo	5'-TTTGGCTACAGCAACAGGGTG--3'	(SEQ ID NO :
				58)
15		Cebador	5'-TGTGAGGAGGGGAGATTCAGTG--3'	(SEQ ID NO :
		inverso		59)

20 Como controles positivos, se usaron  $\beta$ -actina y GAPDH. Todos los productos de la PCR son de 150-250 pb. La mezcla de reacción de PCR de 20  $\mu$ l de cada capilar LightCycler® contenía 2  $\mu$ l de 10 x tampón PCR; MgCl<sub>2</sub> 11,3 mM (Perkin-Elmer, Foster City, CA); 140  $\mu$ M de dNTP; 1:50000 de Verde I SYBR; 0,25 mg/ml de BSA; 1 unidad de *Taq* polimerasa (Boehringer Mannheim, Indianápolis, IN); 0,175  $\mu$ M de cada cebador; 2  $\mu$ l de mezcla de reacción a T.A. La amplificación por PCR comenzó con una desnaturalización de 20 segundos a 95°C, seguida por 45 ciclos de desnaturalización a 95°C durante 5 segundos, haciéndose el apareamiento a 60°C durante 1 segundo y la extensión a 72°C durante 30 segundos. Al terminar el ciclo final, se aparearon los productos de la PCR a 60°C durante 5 segundos, luego se calentaron lentamente hasta 95°C a 0,2°C/segundo para medir la curva de fusión de los productos de PCR específicos. Todos los experimentos se realizaron por duplicado.

30 El análisis de datos se realizó usando el programa informático LightCycler® (Roche Diagnostics) con opciones de cuantificación y de curvas de fusión. La fluorescencia está normalizada con respecto a controles positivos y negativos.

35 *Sobre-expresión de genes en tejido entero de pacientes de cáncer de colon.* Los resultados proporcionados en las tablas que figuran a continuación incluyen datos de fluorescencia para los polinucleótidos aislados de muestras de tejido de colon que fueron cosechados directamente, sin microdisseccionar (i.e., tejido entero) y amplificados usando los cebadores indicados. Los tipos de células normales, de tumor primario y metastásicas son denominadas N, TP y Met, respectivamente. La sobre-expresión se determinó comparando bien las células metastásicas o las células de tumor primario, o ambas, con las células normales. Los resultados de cada gen correspondiente a los grupos indicados en cada muestra de paciente se resumen en las siguientes tablas. Todos los valores están ajustados a niveles relativos al control de la beta-actina.

45 Grupo n.º 719 (SK1): sobre-expresión detectada en 4 de 6 pacientes (67%)

Pacientes	N	TP	MET
UC n.º 1	0,022	0,117	0,364
50 UC n.º 2	0,121	0,109	0,142
UC n.º 4	0,083	0,053	0,078
UC n.º 7	0,042	0,199	0,145
55 UC n.º 8	0,215	0,515	0,794
UC n.º 9	0,233	0,585	0,613

60

65

## ES 2 305 087 T3

Grupo n.º 9083 (SK2): sobre-expresión en 3 de 4 pacientes (75%)

	Pacientes	N	TP	MET
5	UC n.º 1	0,0021	0,0013	0,0078
	UC n.º 2	0,008	0,012	0,014
	UC n.º 4	0,0021	0,0022	0,0026
10	UC n.º 7	0,0009	0,0021	0,0039

15 Grupo n.º 115762 (SK5): sobre-expresión en 5 de los 6 pacientes (83%)

	Pacientes	N	TP	MET
	UC n.º 1	0,0053	0,0159	0,044
20	UC n.º 2	0,0195	0,0174	0,0269
	UC n.º 4	0,022	0,033	0,034
	UC n.º 7	0,013	0,028	0,025
25	UC n.º 8	0,0275	0,105	0,143
	UC n.º 9	0,0336	0,0595	0,0541

30

Grupo n.º 1665: sobre-expresión en 4 de 6 pacientes (67%)

	Pacientes	N	TP	MET
35	UC n.º 1	0,00006	0,0003	0,002
	UC n.º 2	0,0015	0,001	0,0012
40	UC n.º 4	0,0016	0,0013	0,0016
	UC n.º 7	0,00003	0,0003	0,0012
	UC n.º 8	0,0016	0,0122	0,0154
45	UC n.º 9	0,006	0,057	0,097

50 Grupo n.º 2334 (SK8): sobre-expresión en 4 de 6 pacientes (67%)

	Pacientes	N	TP	MET
	UC n.º 1	0,011	0,0022	0,017
55	UC n.º 2	0,0266	0,0317	0,026
	UC n.º 4	0,02	0,006	0,01
	UC n.º 7	0,046	0,093	0,042
60	UC n.º 8	0,042	0,168	0,472
	UC n.º 9	0,208	0,322	0,29

65



## ES 2 305 087 T3

Grupo n.º 22376 (SK19): sobre-expresión en 4 de 6 pacientes (67%)

Pacientes	N	TP	MET
UC n.º 1	0,00018	0,00042	0,0012
UC n.º 2	0,002	0,0025	0,0016
UC n.º 4	0,0013	0,0012	0,002
UC n.º 7	0,00024	0,00055	0,00062
UC n.º 8	0,0003	0,00127	0,0023
UC n.º 9	0,001	0,0075	0,009

Grupo n.º 376130 (Junc2): sobre-expresión en 3 de 4 pacientes (75%)

Pacientes	N	TP	MET
UC n.º 1	0,00871	0,0111	0,0142
UC n.º 2	0,000567	0,00663	0,0163
UC n.º 4	0,000107	0,00048	0,000237
UC n.º 7	0,0000401	0,000259	0,00159

Grupo n.º 402380 (XD4): sobre-expresión en 2 de 4 pacientes (50%)

Pacientes	N	TP	MET
UC n.º 1	0,0763	0,123	0,2
UC n.º 2	0,0867	0,0629	0,069
UC n.º 4	0,0735	0,0672	0,0664
UC n.º 7	0,0559	0,112	0,0139

Grupo n.º 726682 (XD1): sobre-expresión en 0 de 4 pacientes

Pacientes	N	TP	MET
UC n.º 1	0,0679	0,0822	0,136
UC n.º 2	0,175	0,124	0,147
UC n.º 4	0,2	0,145	0,145
UC n.º 7	0,108	0,144	0,114

## ES 2 305 087 T3

Grupo n.º 552930 (XD7): sobre-expresión en 1 de 4 pacientes (25%)

Pacientes	N	TP	MET
UC n.º 1	0,018	0,019	0,0902
UC n.º 2	0,204	0,161	0,212
UC n.º 4	0,299	0,25	0,238
UC n.º 7	0,246	0,409	0,248

Grupo n.º 454001 (XD10): sobre-expresión en 2 de 4 pacientes

Pacientes	N	TP	MET
UC n.º 1	0,0197	0,0363	0,0587
UC n.º 2	0,0514	0,0451	0,069
UC n.º 4	0,0587	0,0889	0,096
UC n.º 7	0,0342	0,1	0,0705

Grupo n.º 378805 (XD11): sobre-expresión en 1 de 4 pacientes

Pacientes	N	TP	MET
UC n.º 1	0,00117	0,00269	0,00697
UC n.º 2	0,00864	0,00371	0,00672
UC n.º 4	0,0098	0,00525	0,00497
UC n.º 7	0,00912	0,00989	0,0127

Grupo n.º 374641 (XD1): sobre-expresión en 3 de 4 pacientes (75%)

Pacientes	N	TP	MET
UC n.º 1	0,0124	0,163	0,0947
UC n.º 2	0,28	0,317	0,544
UC n.º 4	0,685	1,809	1,996
UC n.º 7	0,569	1,714	1,073

Sobre-expresión de genes en el epitelio de pacientes de cáncer de colon Los resultados proporcionados en las tablas que se presentan a continuación incluyen datos de fluorescencia para los polinucleótidos aislados de células epiteliales del colon que fueron preparadas mediante el procedimiento de agitación epitelial para obtener epitelio >97% puro sin estroma. Los tipos de células normales, precancerosas (pólipo adenomatoso) y del tumor primario se denominan N, pólipo y TP respectivamente. La sobre-expresión se determinó comparando bien las células del tumor primario o las células precancerosas, o ambas, con las células normales. Todos los valores están ajustados a niveles relativos al control de la beta-actina.

## ES 2 305 087 T3

Grupo n.º 719 (SK1): sobre-expresión en 4 de 4 pacientes (100%)

Pacientes	N	Pólipo	TP
UC n.º 17	0,0924	0,117	N/D
UC n.º 18	0,0864	N/D	0,327
UC n.º 19	0,151	N/D	0,227
UC n.º 20	0,0624	0,162	0,164

Grupo n.º 115762 (SK51): sobre-expresión en 4 de 4 pacientes (100%)

Pacientes	N	Pólipo	TP
UC n.º 17	0,00724	0,0122	N/D
UC n.º 18	0,0156	N/D	0,111
UC n.º 19	0,0158	N/D	0,0461
UC n.º 20	0,00728	0,0187	0,0306

Grupo n.º 1665: sobre-expresión en 5 de 4 pacientes (100%)

Pacientes	N	Pólipo	TP
UC n.º 17	0,0041	0,0306	N/D
UC n.º 18	0,0029	N/D	0,0357
UC n.º 19	0,0045	N/D	0,20357
UC n.º 20	0,0028	0,025	0,047

Grupo n.º 2334 (SK8): sobre-expresión en 1 de 4 pacientes (25%)

Pacientes	N	Pólipo	TP
UC n.º 17	0,1835	0,041	N/D
UC n.º 18	0,0638	N/D	0,0927
UC n.º 19	0,04	N/D	0,04
UC n.º 20	0,2236	0,0576	0,0454

## ES 2 305 087 T3

Grupo n.º 3376 (SK19): sobre-expresión en 4 de 4 pacientes (100%)

	Pacientes	N	Pólipo	TP
5	UC n.º 17	0,0053	0,012	N/D
	UC n.º 18	0,0028	N/D	0,0084
	UC n.º 19	0,003	N/D	0,0135
10	UC n.º 20	10,0023	10,023	10,012

### 15 Ejemplo 5

#### *Análisis de transferencia Northern*

La expresión génica diferencial en las células cancerosas de colon se puede confirmar más mediante otras técnicas, tales como el análisis de transferencia Northern. El análisis Northern se puede realizar mediante procedimientos conocidos en la técnica. En resumen, se precaliente un tampón rapid-Hyb (Amersham Life Science, Little Chalfont, Inglaterra) con 5 mg/ml de ADN de esperma monocatenario desnaturalizado hasta 65°C y se prehibridan las transferencias de ARN total de tumor de colon humano (Invitrogen, Carlsbad, CA) en el tampón con agitación a 65°C durante 30 minutos. Las sondas de ADN específicas del gen (50 ng por reacción) marcadas con [ $\alpha$ -<sup>32</sup>P]dCTP (3.000 Ci/mmol; Amersham Pharmacia Biotech Inc., Piscataway, NJ) (Equipo Prime-It RmT, Stratagene, La Jolla, CA) y purificadas con micro-columnas G-50 ProbeQuant® (Amersham Pharmacia Biotech Inc.) se añaden y se hibridan a las transferencias con agitación a 65°C durante una noche. Se lavan las transferencias en 2 x SSC; SDS al 0,1% (p/v) a temperatura ambiente durante 20 minutos, dos veces en 1 x SSC; SDS al 0,1% (p/v) a 65°C durante 15 minutos, y luego se exponen a hiper-películas (Amersham Life Science).

30

### Ejemplo 6

35 *Análisis de la expresión del gen correspondiente a SK2 (grupo 9083 (c9083)) (SEQ ID NO: 3) en carcinoma colorrectal*

Se examinó la expresión del gen que comprende la secuencia de SK2, que se agrupa en el grupo con ID. N.º: 9083, mediante PCR cuantitativa en varias líneas celulares de cáncer, incluyendo un número de líneas celulares de carcinoma colorrectal. A continuación, se resumen las células en las que se analizó la expresión.

40

	Línea celular	Fuente tisular	Línea celular	Fuente tisular
45	MDA-MB-231	Mama humana; elevado potencial metastásico (micromet. en pulmón; adenocarcinoma; efusión pleural).	Caco-2	Adenocarcinoma colorrectal humano
50	MDA-MB-435	Mama humana; elevado potencial metastásico (micromet. en pulmón)	SW620	Adenocarcinoma colorrectal humano; desde el punto metastásico (nodo linfático)
55	MCF-7	Mama humana; no metastásica	LS 174T	Adenocarcinoma colorrectal humano de elevado potencial
60				

60

65

## ES 2 305 087 T3

				metastásico	
5	MDA-MB-468	Mama humana; adenocarcinoma	LOVO	Adenocarcinoma humano; colon; desde el punto metastásico (colon)	colorrectal
	Alab	Mama humana, metastásica	HT29	Adenocarcinoma humano; colon	colorrectal
10	SKOV3	Adenocarcinoma de ovario humano	SW480	Adenocarcinoma humano; colon	colorrectal
15	OVCAR3	Adenocarcinoma de ovario humano	HCT116	Carcinoma humano; colon	colorrectal
	KM 12C	Colon humano; bajo potencial metastásico	Colo 320DN	Adenocarcinoma humano; colon	colorrectal
20	KM12L4	Colo humano; elevado potencial metastásico (derivado de Km12C)	T84	Carcinoma humano; colon; desde punto metastásico (pulmón)	colorrectal
25	DU 145	Próstata humana; carcinoma; del punto metastásico: cerebro	HCT15	Adenocarcinoma humano; colon	colorrectal
30	HT 1080	Línea celular de sarcoma humano	CCD 112	Adenocarcinoma humano; bajo potencial metastásico	colorrectal
35	HMVEC	Células endoteliales microvasculares humanas primarias	DLD1	Colon humano; adenocarcinoma colorrectal	
40	185B4	Células epiteliales de mama normal; transformadas químicamente	293	Células epiteliales de riñón	
45	LNCAP	Carcinoma de próstata; metástasis hasta el nodo linfático supraclavicular	GRDP	Epitelio de próstata primario	
	U373MG	Célula de glioblastoma	IMR90	Fibroblasto de pulmón primario	
50	WOCA	Epitelio de próstata primario	PC3	Cáncer de próstata; receptor andrógeno negativo	

La PCR en tiempo real cuantitativa se realizó aislando primero ARN de células usando un equipo de aislamiento de ARN de Roche según las instrucciones del fabricante. Se usó un microgramo de ARN para sintetizar un ADNc de la primera cadena usando transcriptasa inversa de MMLV (Ambion), usando el tampón de los fabricantes y las concentraciones recomendadas de oligodT, nucleótidos y ARNsin. El ADNc de la primera cadena sirvió como molde para la PCR a tiempo real cuantitativa usando el LightCycler® de Roche como se recomienda en el manual del instrumento. El gen correspondiente a SK2 (C9083) (SEQ ID NO: 3) se amplificó con el cebador directo: 5'-cgctgacctcaaccag-3' (SEQ ID NO: 60) y el cebador inverso: 5'-ctgtttgcccggtcttattac-3' (SEQ ID NO: 61). El producto se cuantificó en base al ciclo en el que la amplificación entró en la fase lineal de la amplificación en comparación un patrón interno y usando el programa informático suministrado por el fabricante. Las pequeñas diferencias en las cantidades o el molde total en la reacción del ADNc de la primera cadena se eliminaron normalizando con respecto a la cantidad de actina amplificada en una reacción de PCR cuantitativa separada, usando el cebador directo 5'-CGGGAAATCGTGCGT GACATTAAG-3' (SEQ ID NO: 56) y el cebador inverso: 5'-TGATCTCCTTCTGCATCCTGTGCGG-3' (SEQ ID NO: 57). En la Fig. 1, se muestran los resultados.

## Ejemplo 7

*Análisis funcional del gen correspondiente a SK2 (c9083) (SEQ ID NO: 3)*

Para evaluar en mayor profundidad el papel del gen correspondiente a SK2 (c9083) (SEQ ID NO: 3), se obtuvo la información funcional del gen correspondiente a esta secuencia usando una tecnología de noqueado antisentido. En resumen, se emplacó el tipo de célula por analizar, células SW620 o HT1080 que expresan el polipéptido codificado por el gen correspondiente a c9083, hasta una confluencia de aproximadamente 60-80% en placas de 6 pozos, o para los análisis de proliferación, en placas de 96 pozos. Se diluyó el oligonucleótido control antisentido o inverso hasta 2  $\mu$ M en optimem y se añadió a optimem en el que se había diluido el vehículo de administración, lipitoide 116-6 en el caso de células SW620 o lipitoide:colesterole 1:1 en el caso de las células HT1080. Entonces, se siguió diluyendo la mezcla de oligo/vehículo de administración en medio con suero sobre las células. La concentración final del oligonucleótido para todos los experimentos fue de 300 nM, y la proporción final del oligo con respecto al vehículo de administración para todos los experimentos fue de 1,5 nmoles de lipitoide/ $\mu$ g de oligonucleótido. Las células se transfirieron durante una noche a 37°C y se reemplazó la mezcla de transfección con medio fresco a la mañana siguiente.

Se analizaron los siguientes oligonucleótidos antisentido en cuanto a su capacidad para reducir el ARN de c9083 (SEQ ID NO: 3).

Nombre del oligo	Secuencia	Nucleótidos
CHIR-8-4AS	ATTTGGGCATCACTGGCTACAAGCA	(SEQ ID NO: 64)
C9083:P0463		
CHIR-8-4RC	ACGAACATCGGTCACTACGGGTTTA	(SEQ ID NO: 65)
C9083:P0463RC		
CHIR8-5AS C9083:P0157	CAGAGAGGTGAGACACTCGCCGCA	(SEQ ID NO: 66)
CHIR-8-5RC	ACGCCGCTCACAGAGTGGAGAGAC	(SEQ ID NO: 67)
C9083:P0157RC		
RC: oligos de control inverso (oligos de control); AS: oligos antisentido (análisis)		

Se evaluó el efecto del oligonucleótido sobre las células tanto por la cuantificación de los niveles de PCR según lo descrito anteriormente, como en análisis de proliferación usando la cantidad de ADN según lo cuantificado con el equipo Stratagene Quantos® para determinar el número de células.

Los resultados de la cuantificación del nivel de ARNm se muestran en la Fig. 2. En las Fig. 3 y 4, se muestran los efectos de los oligonucleótidos sobre la proliferación durante un período de cuatro días. Las células sin tratamiento con oligonucleótidos (WT) sirvieron como control. El oligo CHIR-8-4AS fue más eficaz en la disminución del ARNm para el gen correspondiente a 9083c. La transfección de estos oligos en células SW620 resultó en una menor velocidad de proliferación en comparación con los oligos de control inverso apareados, siendo CHIR-8-4 algo más eficaz que CHIR-8-5 (Fig. 3). Lo significativo fue que el mismo oligonucleótido antisentido no tuvo efecto sobre el crecimiento de la línea celular de fibrosarcoma, HT1080 (Fig. 4). Esto indica que el papel funcional del gen correspondiente a c9083 es específico del tejido y que además, el gen correspondiente a c9083 tiene un efecto específico sobre el crecimiento.

A continuación, se analizaron los oligos en cuanto a su efecto sobre la formación de colonias en un análisis de agar disuelto. Los análisis de agar disuelto se realizaron estableciendo primero una capa de fondo de 2 ml de agar al 0,6% en medio recién emplacado tras unas cuantas horas de acodadura sobre las células. La capa celular se formó sobre la capa de fondo eliminando las células transfectadas según lo descrito anteriormente (bien un oligo k-Ras antisentido como control positivo, CHIR-8-4, CHIR-8-5, CHIR-8-4RC o CHIR-8-5RC) de las placas usando tripsina al 0,05% y lavando dos veces en el medio. Se contaron las células en un contador Coulter y se volvieron a suspender hasta obtener  $10^6$  por ml en el medio. Se colocan alícuotas de 10  $\mu$ l con medio en placas de 96 pozos (para comprobar el recuento con WST1) o se siguen diluyendo para un análisis en agar disuelto. Se emplacan 2.000 células en 800  $\mu$ l de agar al 0,4% en pozos por duplicado sobre una capa de fondo de agar al 0,6%. Una vez solidificado el agar de la capa celular, se echó un chorro de 2 ml de medio sobre la capa superior y se añadió oligo de control antisentido o inverso sin vehículos de administración. Cada 3-4 días, se añadían medio recién preparado y oligos. Las colonias comenzaron a formarse tras 10 días hasta las 3 semanas. Se contaron a ojo los campos de colonias. Se pueden usar los valores de metabolismo de WST-1 para compensar las pequeñas diferencias en el número inicial de células. Se pueden barrer campos más amplios para el registro visual de las diferencias.

Tanto el oligo antisentido CHIR-8-4 como el CHIR-8-5 condujeron a disminuir el tamaño y el número de las colonias en comparación con los oligos CHIR-8-4RC y CHIR-8-5RC de control. Estos resultados refuerzan la validación del gen correspondiente a c9083 (SEQ ID NO: 3) como una diana para la intervención terapéutica.

## Ejemplo 8

*Efecto de los oligonucleótidos antisentido sobre el nivel de mensajes para los genes diana*

El efecto de los oligonucleótidos antisentido sobre el nivel de los mensajes para los genes correspondientes a las secuencias y a los grupos descritos en la presente memoria se analizó usando una tecnología de noqueado antisentido según lo descrito para c9083 en el ejemplo anterior. Específicamente, se prepararon según lo descrito anteriormente los oligos antisentido para los genes correspondientes a cada uno de c719, c1665, c3376, c115762, c454001, c3788805 y c776682. Una vez sintetizados y cuantificados, se rastrearon los oligómeros en cuanto a la eficiencia de un noqueado de transcritos en un panel de líneas celulares cancerosas. La eficiencia del noqueado se determinó analizando los niveles de ARNm usando una cuantificación con LightCycler®. Los oligómeros que resultaron en el nivel más elevado del noqueado de transcritos, siendo el nivel de al menos aproximadamente el 50%, preferiblemente, del aproximadamente 80-90%, hasta el 95% o más hasta un mensaje indetectable, se seleccionaron para su uso en un análisis de proliferación de base celular, un análisis de crecimiento independiente del anclaje y un análisis de apoptosis.

Se emplacaron células SW620, que expresan el polipéptido codificado por los genes correspondientes por ser analizados, hasta una confluencia del aproximadamente 60-80% en placas de 6 pozos, o para los análisis de proliferación, en placas de 96 pozos. Para cada mezcla de transfección, se preparó una molécula vehículo, preferiblemente, un lipitoide o un colesterioide, hasta una concentración de trabajo de 0,5 mM en agua, se sometió a un tratamiento con ultrasonidos para producir una solución uniforme y se filtró a través de una membrana PVDF de 0,45 µm. Entonces se preparó el oligonucleótido antisentido o control hasta una concentración de trabajo de 100 µM en agua Millipore estéril. Se siguió diluyendo el oligonucleótido en OptiMEM® (Gibco/BRL), en un tubo Microfuge hasta 2 µM, o aproximadamente, 20 µM de oligo/ml de OptiMEM®. En un tubo Microfuge separado, se diluyó lipitoide o colesterioide, comúnmente, en una cantidad de aproximadamente 1,5-2 nmoles de lipitoide/µg de oligonucleótido antisentido en el mismo volumen de OptiMEM® usado para diluir el oligonucleótido. Se añadió inmediatamente el oligonucleótido antisentido diluido al lipitoide diluido y se mezcló pipeteando arriba y abajo. Se añadió el oligonucleótido a las células hasta una concentración final de 30 nM.

Se cuantificó en las líneas celulares de cáncer el nivel de ARNm diana correspondiente a un gen diana de interés en las células transfectadas usando una máquina de PCR en tiempo real LightCycler® de Roche. Los valores para el ARNm diana se normalizaron frente a un control interno (p. ej., beta-actina). Para cada reacción de 20 µl, se colocó el ARN extraído (generalmente, un total de 0,2-1 µg) en un tubo de microcentrifugación de 0,5 ó 1,5 ml estéril, y se añadió agua hasta un volumen total de 12,5 µl. Se añadieron a cada tubo 7,5 µl de mezcla de tampón/enzima, preparada mezclando (por el orden enumerado) 2,5 µl de H<sub>2</sub>O; 2,0 µl de 10 x tampón de reacción, 10 µl de oligo dT (20 pmoles); 1,0 µl de mezcla de dNTP (10 mM cada una); 0,5 µl de RNAsin® (20u) (Ambion, Inc., Hialeah, FL) y 0,5 µl de transcriptasa inversa MMLV (50u) (Ambion, Inc.). Se mezclaron los contenidos pipeteando arriba y abajo, y se incubó la mezcla de reacción a 42°C durante 1 hora. Se centrifugó el contenido de cada tubo antes de la amplificación.

Se preparó una mezcla de amplificación mezclando por el siguiente orden: 1 x tampón PCR II; 3 mM de MgCl<sub>2</sub>; 140 µM de cada dNTP; 0,175 pmoles de cada oligo; dilución 1:50.000 de verde SYBR®; 0,25 mg/ml de BSA; 1 unidad de Taq polimerasa y H<sub>2</sub>O hasta 20 µl. (El tampón PCR II está disponible a una concentración x 10 en Perkin-Elmer, Norwalk, CT). En una concentración x 1, contiene Tris 10 mM a un pH 8,3 y KCl 50 mM. El verde SYBR® (Molecular Probes, Eugene, OR) es un tinte que emite fluorescencias cuando se une a ADN monocatenario. Como se produce producto de PCR bicatenario durante la amplificación, la fluorescencia del verde SYBR® aumenta. A cada alícuota de 20 µl de mezcla de amplificación, se añadieron 2 µl de RT molde, y se llevó a cabo la amplificación según los protocolos estándar.

Se analizaron los siguientes oligonucleótidos antisentido en cuanto a su capacidad para reducir los niveles de mensajes del gen correspondiente al grupo indicado. Gen diana: Localización del oligo proporciona el nombre del grupo al que está asignado el gen diana y el nombre del oligo usado. AS indica antisentido; RC indica control inverso. Los datos de los genes correspondientes a c9083 se proporcionan a modo comparativo.

Gen diana:Loc. del oligo	Secuencia del oligo	SEQ NO:	ID	% KO del mensaje
C719:1-AS	TTGGTGTCATTGGGTCAAGGGTTGG	68		85%
C719:1-RC	GGTTGGGAAGTGGGTTACTGTGGTT	69		
c719:2-AS	ACAGGGCAGATACGGACCTCGGTG	70		93%
c719:2-RC	GTGGCTCCAGGCATAGACGGGACA	71		
c719:3-AS	TTGTGGGTAAGCAGTTTCATGTCGC	72		67%
c719:3-RC	CGCTGTACTTTGACGAATGGGTGTT	73		
c719:4-AS	CCTGGATCAGACGCAAGTTATCGGC	74		85%

# ES 2 305 087 T3

	c719:4-RC	CGGCTATTGAACGCAGACTAGGTCC	75	
	C9083:4-AS	ATTTGGGCATCACTGGCTACAAGCA	64	83,0
5	C9083:4-RC	ACGAACATCGGTCACTACGGGTTTA	65	
	C9083:5-AS	CAGAGAGGTGAGACACTCGCCGCA	66	73,0
	C9083:5-RC	ACGCCGCTCACAGAGTGGAGAGAC	67	
10	C1665:1-AS	CTACTCCCCACACTTCATCGCCAGG	76	73,0
	C1665:1-RC	GGACCGCTACTTCACACCCCTCATC	77	
	C1665:2-AS	CTCTTGATACTCCAGCGGCAAACCA	78	81,0
15	C1665:2-RC	ACCAAACGGCGACCTCATAGTTCTC	79	
	c3376:1-AS	GCGCCCAAGCCGTTTCGTTCTTAAG	80	78,0
	c3376:1-RC	GAATTCTTGCTTGCCGAACCCGCG	81	
20	c3376:2-AS	CCAGGTAGGCACGAGTTGGCAAAGA	82	97,0
	c3376:2-RC	AGAAACGGTTGAGCACGGATGGACC	83	
	c3376:3-AS	GCCATTGAAGATGCCAGATCCCAC	84	56,0
25	c3376:3-RC	CACCCTAGACCCGTAGAAGTTACCG	85	
	c3376:4-AS	CCTGCGTTTGTCCCTCCAGCATCT	86	93,0
30	c3376:4'-4C	TCTACGACCTCCCTGTTTGCGTCC	87	
	c3376:5-AS	AAGTCACAGTCCCCGGATAACCAGTC	88	88,0
	c3376:5-RC	CTGACCATAGGCCCTGACACTGAA	89	
35	c115762:1-AS	TTGTCGCTTTGGCAGGCATAAAACC	90	97,5
	c115762:2-AS	TCTGGTCATCAACTTGCTTTCCGTG	91	99,0
	c115762:3-AS	CAGTGTTTCGTGGTGTGCTCTGTGG	92	98,0
40	c115762:4-AS	GCTCACCATCCGGGCACCAAGCA	93	97,0
	c115762:5-AS	TGAGAGACAGTGTTTCGTGGTGTGC	94	93,0
	454001:1-AS	TGCCTTCACACGCTTGTTATCTTC	95	0
45	454001:2-AS	GACAACATCGGAGGCTTCAATCACC	96	0
	454001:3-AS	GTTGAGGCTCTGAACACCACTGTTG	97	0
	454001:4-AS	GTTTGGCAGCACCTTCAACATTTGG	98	87
50	454001:5-AS	AGCAGTTTGGCAGCACCTTCAACA	99	92
	454001:1-RC	CTTCTATTGGTTCGCACACTTCCGT	100	
	454001:2-RC	CCACTAACTTCGGAGGCTACAACAG	101	
55	454001:3-RC	GTTGTCACCACAAGTCTCGGAGTTG	102	
	454001:4-RC	GGTTTACAACCTCCACGACGGTTTG	103	
	454001:5-RC	ACAACCTCCACGACGGTTTGACGA	104	
60	378805:1-AS	ATCTGGCATGGACGGATGAGCGAA	105	41,0

65



## ES 2 305 087 T3

	378805:2-AS	GCTGGGTGGTTTCCGAACTCAACG	106	97
	378805:3-AS	GTCCCAATCACCTTCCCCACAATCC	107	65,0
5	378805:4-AS	TCAGATCCTTCTTCCACTCCCGCTT	108	100,0
	378805:5-AS	TGCTCGTGGAACAGGTAAAGCTCTG	109	98
	378805:1-RC	AAGCGAGTAGGCAGGTACGGTCTA	110	
10	378805:2-RC	GCAACTCAAGCCTTTGGTGGGTCG	111	
	378805:3-RC	CCTAACACCCCTTCCACTAACCCTG	112	
	378805:4-RC	TTCGCCCTCACCTTCTTCTAGACT	113	
15	378805:5-RC	GTCTCGAAATGGACAAGGTGCTCGT	114	
	776682:1-AS	AGCTTCACTTTGGTCTTGACGGCAT	115	81
	776682:2-AS	CGGAGGGAAGTCAAGTCAGCCACA	116	60
20	776682:3-AS	CGGCATTACCCCTCTCCAGCACCT	117	89
	776682:4-AS	CCTCCACCTGTTTGCGGGCTTCC	118	61
	776682:5-AS	CCACATTGAGGGAGTCCTCTTGCAA	119	80
25	776682:1-RC	TACGGCAGTTCTGGTTTCACTTCGA	120	
	776682:2-RC	ACACCGACTGAACTGAAGGGAGGC	121	
30	776682:3-RC	TCCACGACCTCTCCACTTACGGC	122	
	776682:5-RC	CCTTCGGGCGTTTGTCCACCTCC	123	
	402380:P464:4-AS	CCCCGAACAAAACACCAGTCAACG	124	94
35	402380:P464:4-RC	GCAACTGACCACAAAACAAGCCCC	125	
	402380:P414:5AS	GGCCATTGAGTCCCTCCATAGCAGC	126	92
	402380:P414:5-RC	CGACGATACCTCCCTGAGTTACCGG	127	

El efecto del oligonucleótido sobre las células se evaluó mediante la cuantificación de los niveles de PCR. En la tabla inmediatamente anterior, se resumen los resultados de la cuantificación de los niveles de ARNm.

El efecto de la pérdida de mensaje para cada gen anterior se puede evaluar en análisis basados en células según lo descrito en el ejemplo 7 anterior. Uno de tales usos del oligonucleótido antisentido descrito por la SEQ ID NO: 108 resultó en una inhibición de la proliferación de las células SW620 cuando se usó según lo descrito en los protocolos de transfección y de análisis de proliferación del ejemplo 7 (Fig. 5).

### Ejemplo 9

*El efecto de la expresión de los genes correspondientes a c3376 y 402380 sobre la proliferación*

El efecto de la expresión de los genes correspondientes a c3376 (gen correspondiente a la SEQ ID NO: 13) y el 402380 (gen correspondiente a la SEQ ID NO: 16) sobre la inhibición de la proliferación celular se evaluó en células de carcinoma colorrectal de colon SW620.

Se emplacaron las células hasta una confluencia de aproximadamente 60-80% en placas de 96 pozos. Se diluyó el oligonucleótido control antisentido o inverso hasta 2  $\mu$ M en optiMEM® y se añadió a optiMEM® en el que se había diluido el vehículo de administración, lipitoide 116-6 en el caso de células SW620 o lipitoide:colesterioide 1:1 en el caso de células MDA-MB-231. Entonces siguió diluyendo la mezcla de oligo/vehículo de administración en medio con suero sobre las células. La concentración final del oligonucleótido para todos los experimentos fue de 300 nM, y la proporción final del oligo con respecto al vehículo de administración para todos los experimentos fue de 1,5 nmoles de lipitoide/ $\mu$ g de oligonucleótido.

Se prepararon los oligonucleótidos antisentido según lo descrito anteriormente. Las células se transfectaron durante una noche a 37°C y, a la mañana siguiente, se reemplazó la mezcla de transfección con medio recién preparado. La transfección se llevó a cabo según lo descrito anteriormente en el ejemplo 8. La proliferación se midió usando el

reactivo colorimétrico WST-1 según los procedimientos conocidos en la técnica. En las fig. 6-9, se muestran los resultados de los experimentos antisentido. Los valores del eje Y representan las unidades de fluorescencia relativa. Los oligos control antisentido e inversos para K-Ras sirvieron como control para demostrar que el análisis funcionaba según lo esperado (Fig. 6).

## 5 Ejemplo 10

### *Efecto de la expresión génica sobre la formación de colonias en agar disuelto*

10 El efecto de la expresión del gen correspondiente a 402380 (gen correspondiente a la SEQ ID NO: 16) sobre la formación de colonias de células SW620 se analizó en un análisis en agar disuelto. Los análisis de agar disuelto se realizaron estableciendo primero una capa de fondo de 2 ml de agar al 0,6% en medio recién emplacado con  
15 unas cuantas horas de acodadura sobre las células. La capa celular se formó sobre la capa de fondo eliminando las células transfectadas según lo descrito anteriormente de las placas usando tripsina al 0,05% y lavando dos veces en el medio. Se contaron las células en un contador Coulter y se volvieron a suspender hasta obtener  $10^6$  por ml en el medio. Se colocan alícuotas de 10  $\mu$ l con medio en placas de 96 pozos (para comprobar el recuento con WST-1) o se sigue diluyendo para un análisis en agar disuelto. Se emplacaron 2.000 células en 800  $\mu$ l de agar al 0,4% en pozos por duplicado sobre una capa de fondo de agar al 0,6%. Una vez solidificado el agar de la capa celular, se echó un  
20 chorro de 2 ml de medio sobre la capa superior y se añadió oligo de control antisentido o inverso (producido según lo descrito anteriormente) sin vehículos de administración. Cada 3-4 días, se añadió medio recién preparado y oligos. Las colonias comenzaron a formarse tras 10 días hasta las 3 semanas. Se contaron a ojo los campos de colonias. Se usaron los valores de metabolismo de WST-1 para compensar las pequeñas diferencias en el número inicial de células. Se pueden barrer campos más amplios para el registro visual de las diferencias.

25 En la Fig. 9, se muestran los resultados. 9. El eje Y representa el número de células por un sector definido, usando WST-1 para facilitar el recuento de células y normalizado con respecto a un control. Los oligos control antisentido e inversos para K-Ras ((KRAS 2576-as y KRAS 2576-rc) sirvieron como control para demostrar que el análisis funcionaba según lo esperado.

## 30 Ejemplo 11

### *Efecto de la expresión génica sobre la muerte celular*

35 Se examinó el efecto de la expresión de los genes correspondientes al grupo 719 (gen correspondiente a la SEQ ID NO: 1, CHIR-7); grupo 9083 (gen correspondiente a la SEQ ID NO: 3, CHIR-8); grupo 1665 (gen correspondiente a las SEQ ID NO: 7 y 9, CHIR-9); grupo 3376 (gen correspondiente a la SEQ ID NO: 13, CHIR 11); grupo 115762 (gen correspondiente a la SEQ ID NO: 5, CHIR-16); y grupo 402380 (gen correspondiente a la SEQ ID NO: 16, CHIR-33) sobre la muerte celular en un análisis de citotoxicidad de lactato deshidrogenasa (LDH) en células HT1080 (una línea celular de fibrosarcoma humano), células SW620 y líneas celulares de cáncer de mama metastásicas (MDA-MB-  
40 231 ("231")). El análisis de citotoxicidad de la lactato deshidrogenasa (LDH) fue esencialmente según lo siguiente:

El análisis de citotoxicidad de lactato deshidrogenasa (LDH) se realizó esencialmente según lo siguiente:

45 *Día 1:* Se sembraron células en 4 placas de 96 pozos separadas, comúnmente 5.000 células/pozo, y se incubaron a 37°C y con CO<sub>2</sub> al 5%.

*Día 2:* Se transfectaron las células con controles antisentido así como con los controles complementarios inversos, esencialmente según lo descrito en el ejemplo 4. Se dejó sin transfectar una placa (día 0) como control de sembrado.

50 La transfección se llevó a cabo usando un vehículo de lípidos para la administración según lo descrito en el documento WO 01/16306, incorporado en la presente memoria en su totalidad. En resumen, la transfección usó agentes conocidos como "lipitoides" y "colesteroides", descritos, por ejemplo, en las publicaciones vía PCT WO 01/16306, WO 98/06437 y WO 99/08711.

55 En estas referencias se demuestra que estos conjugados de lípido-peptido catiónico son reactivos eficaces para la administración de ADN de plásmido a células *in vitro*. Cualquiera de los vehículos descritos en las solicitudes a las que se ha hecho referencia anteriormente es adecuado para su uso en la transfección de los oligonucleótidos descritos en la presente memoria.

60 Estos compuestos se pueden preparar mediante la solución convencional o mediante la síntesis en fase sólida. En uno de tales procedimientos, según lo descrito en el documento WO 99/08711, citado anteriormente, se acila el terminal N de un peptido unido a resina con un espaciador tal como ácido Fmoc-aminohexanoico o Fmoc-3-alanina. Tras la eliminación del grupo Fmoc, se hace reaccionar el grupo amino primario con cloroformiato de colesterol para formar un enlace de carbamato. Entonces se escinde el producto de la resina con ácido trifluoroacético y se purifica  
65 mediante CLAR en fase inversa. Se puede usar un resto de lípido derivado de ácido graso, tal como un fosfolípido, en lugar del resto de esteroide. Además se puede unir el esteroide o el otro resto de lípido con el resto de peptido mediante otros enlaces, de cualquier longitud, disponibles fácilmente para el experto.

En función del tipo de célula, se usaron diferentes vehículos de lípido para diferentes duraciones de la transfección. Sin embargo, el tiempo de transfección no superó las 24 h. La transfección se llevó a cabo en un medio completo y la concentración de oligonucleótido antisentido final fue de 300 nM por pozo. En los pozos con fármaco, se añadió el fármaco al cultivo al comienzo de la transfección.

*Comenzando el día 3:* Se recuperaron las células, 1 placa/día, y se midió la liberación de LDH en el sobrenadante así como la LDH de las células intactas usando un equipo de Roche según las instrucciones del fabricante (Roche Diagnostics, Basel, Switzerland) (datos marcados como día 1, 2, 3).

Se analizó cada muestra examinando el nivel relativo de LDH liberada en comparación con la LDH total, significando un aumento como porción de la LDH total un aumento de la muerte celular (debido a una mayor proporción de LDH liberada en el medio). Los datos se evaluaron cualitativamente mediante la comparación con un control sin tratar (no oligo). Este análisis permitió determinar si la pérdida de mensaje inducida por los antisentido para un determinado gen causa la muerte de las células cuando se usa solo, o si esta pérdida de mensaje sintetiza células bajo los efectos de un fármaco.

Los resultados se muestran en la tabla que se presenta inmediatamente a continuación.

	<b>HT1080</b>	<b>SW620</b>	<b>231</b>
Chir7-2	Negativo	Negativo	
Chir8-4	Positivo	Débilmente	
		positivo	
Chir9-5		Positivo	
Chir1-2		Negativo	
Chir16-4		Negativo	
Chir33-4	Muy débilmente positivo	Muy positivo	Muy débilmente positivo

#### Ejemplo 12

##### *Detección de la expresión diferencial usando alineamientos*

Se analizaron ARNm aislado de muestras de tejido de colon canceroso y normal obtenidas de pacientes para identificar genes expresados diferencialmente en células cancerosas y normales. Las células normales y cancerosas recogidas de tejidos criopreservados de pacientes se aislaron usando técnicas de microdissección por captura láser (LCM), siendo estas técnicas conocidas en la técnica (véase, p. ej., Ohyama *et al.* (2000) *Biotechniques* 29:530-6; Curran *et al.* (2000) *Mol. Pathol.* 53:64-8; Suarez-Quian *et al.* (1999) *Biotechniques* 26:328-35; Simone *et al.* (1998) *Trends Genet* 14:272-6; Conia *et al.* (1997) *J. Clin. Lab. Anal.* 11:28-38; Emmert-Buck *et al.* (1996) *Science* 274:998-1001).

La tabla 5 (insertada antes de las reivindicaciones) proporciona la información acerca de cada paciente del que se aislaron las muestras, incluyendo: el "ID del paciente" y "ID del informe pat.", que son números asignados al paciente y a los informes de patología a efectos de identificación; el "Grupo" al que se han asignado los pacientes; la localización anatómica del tumor ("Loc. Anatom."; el "Tamaño del tumor primario"; el "Grado del tumor primario"; la identificación del grado histopatológico ("Grado histopat."); una descripción de los puntos locales invadidos por el tumor ("Invasión local"); la presencia de metástasis en los nódulos linfáticos ("Met. en nódulos linf."); la incidencia de la metástasis en los nódulos linfáticos (proporcionada como un número de nódulos linfáticos con positivo en metástasis frente al número de nódulos linfáticos examinados) ("Incidencia de Met. en N.L."; el "grado de N.L. regionales"; la identificación o la detección de metástasis en puntos distantes al tumor y su localización ("Met. y loc. distante"; una descripción de las metástasis distantes ("descrip. Met. distantes"); el grado de metástasis distante ("Grado de met. distante"); y los comentarios generales sobre el paciente o el tumor ("Comentarios"). No se describieron adenomas en ninguno de los pacientes; la displasia de adenomas (descrita como hiperplasia por el patólogo) se describió en el paciente con n.º ID: 695. Se describieron extensiones extranodales en dos pacientes, n.º ID de paciente: 784 y 791. La invasión linfovascular se describió en siete pacientes, n.º ID de paciente: 128, 278, 517, 534, 784, 786 y 791. Se describieron infiltrados de tipo Crohn en siete pacientes, n.º ID de paciente: 52, 264, 268, 392, 393, 784 y 791.

*Identificación de genes expresados diferencialmente*

Se prepararon sondas de ADNc a partir de ARN total aislado de las células de pacientes descritas anteriormente. Como la LCM proporciona el aislamiento de tipos de células específicos para proporcionar una muestra celular sustancialmente homogénea, se proporciona una muestra de ARN similarmente pura.

Primero se transcribió inversamente el ARN total en ADNc usando un cebador que contenía un promotor de la ARN polimerasa T7, siguiendo con una síntesis de ADN de la cadena secundaria. Entonces se transcribió el ADNc *in vitro* para producir ARN antisentido usando una expresión mediada por el promotor T7 (véase, p. ej., Luo *et al.* (1999) *Nature Med* 5: 117-122), y entonces se convirtió el ARN antisentido en ADNc. Se volvió a transcribir el segundo conjunto de ADNc *in vitro*, usando el promotor T7, para proporcionar ARN antisentido. Opcionalmente, se volvió a convertir el ARN en ADNc, permitiendo hasta una tercera vuelta de amplificación mediada por T7 para producir más ARN antisentido. De este modo, el procedimiento proporcionó dos o tres vueltas de transcripción *in vitro* para producir el ARN final usado para el marcaje fluorescente.

Las sondas fluorescentes se generaron añadiendo primero ARN de control a la mezcla de ARN antisentido, y produciendo ADNc marcado fluorescentemente a partir del material inicial de ARN. Se compararon los ADNc marcados fluorescentemente preparados de la muestra de ARN tumoral con los ADNc marcados fluorescentemente preparados de la muestra de ARN de células normales. Por ejemplo, las sondas de ADNc de las células normales se marcaron con tinte fluorescente Cy3 (verde) y las sondas de ADNc preparadas de las células tumorales se marcaron con tinte fluorescente Cy5 (rojo), y viceversa.

Cada alineamiento usado tenía una distribución espacial de las capas y un establecimiento de las manchas de control idénticos. Cada microalineamiento estaba dividido en dos zonas, teniendo cada zona un alineamiento con, en cada mitad, doce agrupamientos de 32 x 12 manchas, para un total de aproximadamente 9.216 manchas en cada alineamiento. Las dos zonas están manchadas de manera idéntica, lo que proporciona al menos dos duplicados de cada clon por alineamiento.

Los polinucleótidos correspondientes a los genes expresados diferencialmente descritos en la presente memoria para su uso en los alineamientos se obtuvieron tanto de fuentes a disposición del público como de genotecas de ADNc generadas de líneas celulares y de tejidos de pacientes seleccionados. Los productos de PCR de aproximadamente 0,5 kg a 2,0 kb amplificados de estas fuentes fueron manchados sobre el alineamiento usando un marcador de Molecular Dynamics Gen III según las recomendaciones del fabricante. La primera fila de cada una de las 24 regiones del alineamiento tenía aproximadamente 32 manchas de control, incluyendo 4 manchas de control negativas y 8 polinucleótidos de análisis. Se echaron los polinucleótidos de análisis en cada muestra antes de la reacción de marcaje con una variedad de concentraciones de 2-600 pg/muestra y proporciones de 1:1. Para cada diseño de alineamiento, se hibridaron dos muestras con las muestras de análisis marcadas inversamente en la reacción de marcaje. Esto proporcionó aproximadamente cuatro medidas por duplicado de cada clon, dos de un color y dos de otro color, para cada muestra.

El análisis de expresión diferencial se realizó mezclando cantidades iguales de sondas de células tumorales y células normales del mismo paciente. Los alineamientos se pre-hibridaron por incubación durante aproximadamente 2 h a 60°C en 5 x SSC y SDS al 0,2%/EDTA 1 mM, y luego se lavaron tres veces en agua y dos veces en isopropanol. Tras la pre-hibridación del alineamiento, se hibridó la mezcla de sonda con el alineamiento en condiciones muy restrictivas (una noche a 42°C en formamida al 50%; 5 x SSC y SDS al 0,2%). Tras la hibridación, se lavó el alineamiento a 55°C tres veces según lo siguiente: 1) primer lavado en 1 x SSC/SDS al 0,2%; 2) segundo lavado en 0,1 x SSC/SDS al 0,2%; y 3) tercer lavado en 0,1 x SSC.

Entonces se rastrearon los alineamientos en busca de verde y rojo fluorescente usando un detector/escáner láser a color dual de Molecular Dynamics Generation III. Se procesaron las imágenes usando el programa informático Bio-Discovery Autogene, y se normalizaron los datos de cada barrido para obtener una proporción de expresión en relación con la normal. Los datos de los experimentos de microalineamiento se analizaron según los algoritmos descritos en la solicitud estadounidense de n.º de serie 60/252.358, presentada el 20 de noviembre de 2000, por E. J. Moler, M. A. Boyle, y F. M. Randazzo, y titulada "Precision and accuracy in cDNA microarray data".

Se repitió el experimento, esta vez marcando las dos sondas con el color opuesto con el fin de realizar el análisis en ambas "direcciones del color". Se repitió cada experimento algunas veces con dos muestras más (una en cada dirección del color). El nivel de fluorescencia para cada secuencia del alineamiento expresado como una proporción de la media geométrica de 8 manchas/genes por duplicado de cuatro alineamientos o 4 manchas/genes por duplicado de 2 alineamientos, o alguna otra variante. Los datos se normalizaron usando los controles positivos añadidos presentes en cada zona duplicada, y la precisión de esta normalización se incluyó en la determinación final de la significancia de cada diferencial. También se comparó la intensidad fluorescente de cada mancha con los controles negativos de cada zona duplicada para determinar qué manchas han detectado niveles de expresión significativos en cada muestra.

Se aplicó un análisis estadístico de las intensidades fluorescentes a cada conjunto de manchas por duplicado para evaluar la precisión y la significancia de cada medida diferencial, dando como resultado un valor p que prueba la hipótesis nula de que no hay diferencial en el nivel de expresión entre las muestras tumorales y las normales de cada paciente. Durante el análisis inicial de los microalineamientos, se aceptó la hipótesis si  $p > 10^{-3}$ ; y la proporción diferencial se estableció en 1.000 para aquellas manchas. El resto de manchas tienen una diferencia significativa en la

expresión entre la muestra tumoral y la normal. Si la muestra tumoral tiene una expresión detectable y la normal no, la proporción se trunca en 1.000, pues el valor para la expresión en la muestra normal sería cero, y la proporción no sería un valor matemáticamente útil (p. ej., infinito). Si la muestra normal tiene una expresión detectable y la tumoral no, la proporción se trunca en 0,001, pues el valor para la expresión en la muestra tumoral sería cero, y la proporción no sería un valor matemáticamente útil. Estas dos últimas situaciones se denominan en la presente memoria "on/off". Las tablas de la base de datos se poblaron usando un nivel de confianza del 95% ( $p > 0,05$ ).

En la tabla 6 de abajo, se proporcionan los resultados. La tabla incluye: 1) la SEQ ID NO; 2) la identificación de la muestra (ID de muestra); 3) el número de identificación de mancha ("ID de mancha"); y 4) el porcentaje de pacientes analizados en cuyos niveles de expresión del gen resultó ser al menos 2 veces mayores en tejido canceroso que en tejido normal apareado ("val. p corregido  $95 \geq x$  2 de pacientes de colon". Las proporciones de expresión diferencial se expresan como una señal de hibridación normalizada asociada con la sonda tumoral dividida entre la señal de hibridación normalizada con la sonda normal. De este modo, una proporción mayor de 1 indica que la expresión del producto génico aumenta en células cancerosas en comparación con células normales, mientras que una proporción de menos de 1 indica lo opuesto.

<b>Tabla 6</b>	<b>ID de muestra</b>	<b>ID de</b>	<b>val. p corregido <math>95 \geq x</math></b>
<b>SEQ ID</b>		<b>mancha</b>	<b>2 de pacientes de</b>
<b>NO.</b>			<b>colon</b>
1	RG:727787:Order7TM31:E07	29912	82,14
7	M00055209C:B07	24297	30,30
9	M00056908A:H05	21544	42,42
13	M00057000D:E08	21592	30,30
27	RG:1418951:Order7TMII:D12	33623	78,57
29	RG:	33623	78,57
	1418951:Order7TMII:D12		
22	M00001346C:A05	243	55
22	M00054893C:D03	21952	30

Estos datos proporcionan las pruebas de que los genes representados por los polinucleótidos que tienen las secuencias indicadas se expresan diferencialmente en el cáncer de colon.

Los expertos en la técnica reconocerán, o serán capaces de determinar, usando simplemente una experimentación rutinaria, muchos equivalentes a las realizaciones específicas de la invención, descritas en la presente memoria. Tales realizaciones específicas y equivalentes pretenden estar englobados por las siguientes reivindicaciones.

Aunque la invención anterior haya sido descrita en cierto detalle a modo ilustrativo y ejemplificante a efectos de aclarar su comprensión, el experto habitual en la técnica comprenderá fácilmente a la luz de las enseñanzas de esta invención que se pueden hacer ciertos cambios y modificaciones de la misma sin alejarse del espíritu o del alcance de las reivindicaciones anexas.

*Información del depósito.* Se hizo un depósito de cultivos biológicamente puros de los siguientes virus con la colección americana de cultivos tipo, 10801 University Blvd., Manassas, VA 20110-2209, en virtud de las disposiciones del Tratado de Budapest, en o antes de la fecha de presentación de la presente solicitud. El número de acceso indicado fue asignado tras un análisis de viabilidad satisfactorio, habiendo sido pagadas las tarifas exigidas. El acceso a tales cultivos estará disponible durante la vigencia de la solicitud de patente a uno determinado por el comisionado estar titulado para ello bajo 37 C.F.R. § 1.14 y 35 U.S.C. § 122. Todas las restricciones sobre la disponibilidad de dichos cultivos al público serán irrevocablemente eliminadas tras la concesión de una patente basada en la solicitud. Además, los depósitos designados serán mantenidos durante un período de treinta (30) años desde la fecha del depósito, o durante cinco (5) años tras la última petición de depósito; o durante la vida de cumplimiento de la patente estadounidense, independientemente de su duración. En caso de que un cultivo se volviera inviable o fuera inadvertidamente destruido, o, en caso de las cepas que contienen plásmidos, perdiera sus plásmidos, será reemplazado por un/unos cultivo/s viable/s de la misma descripción taxonómica.

Estos depósitos se proporcionan simplemente para la comodidad de los expertos en la técnica, y no reconocen la necesidad de un depósito. Las secuencias de ácido nucleico de estos plásmidos, así como las secuencias de aminoácidos

de los polipéptidos codificados por los mismos, son controladas en el caso de cualquier conflicto con la descripción de la presente memoria. Se puede requerir una licencia para elaborar, usar o vender los materiales depositados, no siendo tal licencia concedida mediante la presente memoria.

- 5 Además, se asignaron a las mezclas de clones seleccionados, así como a las genotecas que contienen clones específicos, un número “ES” (Referencia interna) y se depositaron con la ATCC. La tabla 7 que se presenta a continuación proporciona los n.º de acceso de la ATCC de los clones depositados, la totalidad de los cuales fue depositada en o antes de la fecha de presentación de la solicitud.

TABLA 7

*Mezclas de clones y genotecas depositadas con la ATCC*

Nombre de secuencia	Clones	CMCC	ATCC
SK1	SK-1	5162	PTA-1360
SK2	SK-2	5163	PTA-1361
SK5	SK-5	5164	PTA-1362
1665 corto	1665 corto	5165	PTA-1363
1665 largo	1665 largo	5166	PTA-1363
sk19	SK-19	5167	PTA-1364
Junc2	Junc2-6	5168	PTA-1365
XD4	XD4b	5169	PTA-1366
XD1	XD1b	5170	PTA-1367
XD7	XD7c	5171	PTA-1368
XD10	XD10b	5172	PTA-1369
XD11	XD11b	5173	PTA-1370
Junc4	Junc4-2	5174	PTA-1371

CMCC se refiere al número de referencia interno del solicitante.

*Recuperación de clones individuales del depósito de clones mezclados.* Como el depósito ATCC está compuesto de una mezcla de clones de ADNc o una genoteca de clones de ADNc, se preparó el depósito transfeciendo primero cada uno de los clones en células bacterianas separadas. Los clones de la muestra o de la genoteca se depositaron entonces como una mezcla de mezclas iguales en el depósito compuesto. Se pueden obtener clones concretos del depósito compuesto usando procedimientos conocidos en la técnica. Por ejemplo, se puede identificar una célula bacteriana que contenga un clon particular aislando colonias simples e identificando las colonias que contienen el clon específico a través de técnicas de hibridación de colonias estándar, usando una sonda o sondas de oligonucleótidos diseñadas para que se hibriden específicamente con una secuencia del inserto de clon (p. ej., una sonda basada en una secuencia desenmascarada del polinucleótido codificado que tiene la SEQ ID NO indicada). La sonda debería diseñarse para que tuviera una Tm de aproximadamente 80°C (suponiendo 2°C para cada A o T y 4°C para cada G o C). Entonces se pueden recoger las colonias positivas, dejarlas crecer en cultivos y aislar los clones recombinantes. Alternativamente, se pueden usar las sondas diseñadas de esta manera para una PCR para aislar una molécula de ácido nucleico de los clones mezclados según los procedimientos conocidos en la técnica, p. ej., purificando el ADNc de la mezcla de cultivos depositada, y usando las sondas en reacciones de PCR para producir un producto amplificado que tenga la correspondiente secuencia de polinucleótidos deseada.

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor	Grado del tumor	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Grado de los GL regionales	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios
15	21	III	Colon ascendente	4,0	T3	G2	Extendiéndose en el tejido adiposo subserosal	Positiva	N1	Negativa		MX	Adenocarcinoma invasivo moderadamente diferenciado; se observa invasión perineural focal
52	71	II	Colon ascendente	9,0	T3	G3	Invasión a través de la muscularis propia, implicación subserosal; implicación de la válvula ileocecal	Negativa	N0	Negativa		M0	Pólipo hiperplásico en apéndice.
121	140	II	Sigmoide	6	T4	G2	Invasión de la muscularis propia en la serosa, implicación de la submucosa de la vejiga urinaria.	Negativa	N0	Negativa		M0	Invasión perineural; rodetes de anastomosis negativos. Un adenoma de Tubulovillous y uno

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Grado de los GL regionales	Met y Loc. distante	Descrip. de la Met. distante	Grado de la met. distante	Comentarios
125	144	II	Ciego	6	T3	G2	Invasión a través de la muscularis propia en el tejido adiposo subserosal. Unión ileocecal	Negativa	N0	Negativa		M0	Historial del paciente de melanoma metastásico.
128	147	III	Colon transverso	5,0	T3	G2	Invasión de la muscularis propia.	Positiva	N1	Negativa		M0	
130	149		Ángulo esplénico	5,5	T3		A través de la pared y en el tejido adiposo circundante.	Positiva	N2	Negativa		M1	
133	152	II	Recto	5,0	T3	G2	Invasión a través de la muscularis propia en tejido pericólico no peritonealizado; la configuración gruesa	Negativa	N0	Negativa		M0	Pequeño adenoma tubular separado (0,4 cm)



Tabla 5 Datos de los pacientes

ID del paciente	ID del informe	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL regionales	Grado de los GL	Met y Loc. distante	Descrip. de la Met distante	Grado de la met.	Comentarios
141	160	IV	Ciego	5,5	T3	G2	Invasión de la muscularis propia en el tejido adiposo	Positiva	7/21	N2	Positiva (hígado)	Adenocarcinoma coincidente con el primario	M1	Invasión perineural identificada adyacente al adenocarcinoma metastásico.
156	175	III	Ángulo hepático	3,8	T3	G2	Invasión a través de la muscularis propia en adiposo subserosal/pericólico, sin implicación serosal. Configuración gruesa anular.	Positiva	2/13	N1	Negativa	Adenomas de Tubulovillous y tubular separados.	M0	Adenomas de Tubulovillous y tubular separados.
228	247	III	Recto	5,8	T3	G2 a G3	Invasión a través de la muscularis propia	Positiva	1/8	N1	Negativa		MX	Pólipos hiperplásicos

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe	Grupo	Loc. anatómica	Tamaño del tumor	Grado del tumor	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Grado de Incidencia de la met. de los GL regionales	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios
264	283	II	Colon ascendente	5,5	T3	G2	Invasión a través de la muscularis propia en el tejido adiposo subserosal.	Negativa	0/10	N0	Negativa	M0	Adenoma de Tubulovillous con alto grado de displasia.
266	285	III	Colon transverso	9	T3	G2	Invasión a través de la musc. propia hasta implicar el adiposo pericolónico, se extiende hasta la serosa.	Negativa	0/15	N1	Positiva (depósito mesentérico)	MX	0,4 cm, puede representar GL completamente reemplazada por el tumor

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL	Grado de los GL regionales	Met. y Loc. distante	Descrip. de la Met. distante	Grado de la met. distante	Comentarios
268	287	I	Ciego	6,5	T2	G2	Invade todo el espesor de la muscularis propia, pero el adiposo mesentérico está libre de malignidad.	Negativa	0/12	N0	Negativa		M0	
278	297	III	Recto	4	T3	G2	Invasión en el tejido adiposo perirrectal.	Positiva	7/10	N2	Negativa		M0	Pólipos en el colon descendente, ni HGD ni carcinoma identificados..
295	314	II	Colon ascendente	5,0	T3	G2	Invasión a través de la muscularis propia en el tejido adiposo pericólico.	Negativo	0/12	N0	Negativa		M0	Melanosis coli y enfermedad diverticular.

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL	Grado de los GL regionales	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios
339	358	II	Recto sigmoide	6	T3	G2	Se extiende en la grasa perirrectal pero no alcanza la serosa	Negativa	0/6	N0	Negativa		M0	1 pólipo hiperplásico
341	360	II	Colon ascendente	2 cm de invasión	T3	G2	Invasión a través de la muscularis propia hasta implicar a la grasa pericolónica. Surge de adenoma vellosos.	Negativa	0/4	N0	Negativa		MX	identificado
356	375	II	Sigmoide	6,5	T3	G2	A través de la pared del colon en el tejido adiposo subserosal. No se observa propagación serosal.	Negativa	0/4	N0	Negativa		M0	

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL regionales	Grado de los GL	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios
360	412	III	Colon ascendente	4,3	T3	G2	Invasión a través de la muscularis propia hasta la grasa pericolónica	Positiva	1/5	N1	Negativa			M0 Dos pólipos mucosales
392	444	IV	Colon ascendente	2	T3	G2	Invasión a través de la muscularis propia en el tejido adiposo subserosal, sin serosa.	Positiva	1/6	N1	Positiva (Higado)	Esteatosis macrovesicular y microvesicular	M1	El tumor surge antes de la anastomosis quirúrgica ileocólica.
393	445	II	Ciego	6,0	T3	G2	Ciego, invade a través de la muscularis propia hasta implicar al tejido adiposo subserosal pero no la serosa.	Negativa	0/21	N0	Negativa		M0	

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL regionales	Grado de los GL	Met y Loc. distante	Descrip. de la Met distante	Grado de la met.	Comentarios
413	465	IV	Colon ascendente	4,8	T3	G2	Invasión a través de la muscularis propia hasta implicar a la grasa periserosal; colindando con la unión ileocecal.	Negativa	0/7	N0	Positiva (higado)	Adenocarcinoma en múltiples muestras	M1	Radiagnosis de ruta de ooforectomía hasta el cáncer de colon metastásico.
505	383	IV		Diam max. De 7,5 cm	T3	G2	Invasión a través de la muscularis propia hasta implicar a adiposa pericólica; pero no a superficie serosal	Positiva	2/17	N1	Positiva (Higado)	Adenocarcinoma moderadamente diferenciado, coincidente con primario	M1	Localización anatómica del primario no indicada en el informe. Pruebas de colitis crónica.

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL regionales	Grado de los GL	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios
517	395	IV	Sigmoide	3	T3	G2	Penetra en muscularis propia, implicando a la grasa pericólica.	Positiva	6/6	N2	Negativa		M0	No hay mención de met. distante en el informe
534	553	II	Colon ascendente	12	T3	G3	Invasión a través de la muscularis propia implicando a la grasa pericólica. Serosa libre de tumor.	Negativa	0/8	N0	Negativa		M0	Epilón con fibrosis y necrosis de grasas. Intestino delgado con serositis crónica y aguda; abscesos y adhesiones focales.
546	565	IV	Colon ascendente	5,5	T3	G2	Invasión a través de la muscularis propia extensamente a través de submucosal y extend. a la serosa.	Positiva	6/12	N2	Positiva (Hígado)	Adenocar- cinoma metastásico	M1	

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL	Grado de los GL regionales	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios	
577	596	II	Ciego	11,5	T3	G2	Invasión a través de la pared del intestino a la adiposa subserosal.	Negativa	0/58	N0	Negativa			M0	Apéndice dilatado y fibrótico, pero no afectado por el tumor.
							Superficie serosal libre de tumor								
695	714	II	Ciego	14	T3	G2	Extendiéndose a través de la pared del intestino a la grasa serosal	Negativa	0/22	N0	Negativa		MX	Adenoma tubular y pólipos hiperplásicos presentes, adenoma moderadamente diferenciado con diferenciación mucicida (% no establecido)	
784	803	IV	Colon ascendente	3,5	T3	G3	A través de la muscularis propia en tejidos blandos pericólic.	Positiva	5/17	N2	Positiva (Hígado)		M1	Carcinoma adenoscamoso diferenciado pobremente invasivo.	



Tabla 5 Datos de los pacientes

ID del paciente	ID del informe pat.	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL regionales	Grado de los GL	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios
786	805	IV	Colon descendente	9,5	T3	G2	A través de la muscularis propia a la grasa	Negativa	0/12	N0	Positiva (Hígado)		M1	Adenocarcinoma invasivo moderadamente diferenciado.
791	810	IV	Colon ascendente	5,8	T3	G3	A través de la muscularis propia en la grasa pericólica	Positiva	13/25	N2	Positiva (Hígado)		M1	Adenocarcinoma colónico invasivo pobremente diferenciado
888	908	IV	Colon ascendente	2,0	T2	G1	En la muscularis propia	Positiva	3/21	N0	Positiva (Hígado)		M1	Adenocarcinoma de bien a moderadam. diferenciado; este paciente tiene tumores en colon ascendente y colon sigmoide.

Tabla 5 Datos de los pacientes

ID del paciente	ID del informe	Grupo	Loc. anatómica	Tamaño del tumor principal	Grado del tumor principal	Grado histopat.	Invasión local	Met. de los ganglios linf. (GL)	Incidencia de la met. de los GL	Grado de los GL regionales	Met y Loc. distante	Descrip. de la Met distante	Grado de la met. distante	Comentarios
889	909	IV	Ciego	4,8	T3	G2	A través de la muscularis propia al tejido subserosal	Positiva	1/4	N1	Positiva (Hígado)		M1	Adenocarcinoma moderadamente diferenciado.

## REIVINDICACIONES

1. Un procedimiento para detectar una célula de colon cancerosa que comprende:

poner en contacto una muestra obtenida de una célula de colon de análisis con una sonda para la detección de un producto génico de un gen expresado diferencialmente en el cáncer de colon, comprendiendo el gen una secuencia de SEQ ID NO: 1, siendo dicho contacto durante un tiempo suficiente para unir la sonda con el producto génico; y

comparar un nivel de la unión de la sonda con la muestra con un nivel de la unión de la sonda con una muestra control obtenida de una célula de colon control, siendo la célula de colon control de un estado canceroso conocido;

en el que un nivel de unión de la sonda en la muestra de células de colon de análisis similar al nivel de unión en una muestra control es indicativo del estado canceroso de la célula de colon de análisis.

2. El procedimiento de la reivindicación 1, en el que la sonda es una sonda de polinucleótido y el producto génico es un ácido nucleico.

3. El procedimiento de la reivindicación 1, en el que el producto génico es un polipéptido.

4. El procedimiento de la reivindicación 1, en el que el producto génico está inmovilizado sobre un alineamiento.

5. El procedimiento de la reivindicación 1, en el que la sonda está inmovilizada sobre un alineamiento.

6. Un procedimiento para identificar una célula de colon cancerosa, comprendiendo el procedimiento las etapas de:

detectar al menos un producto génico expresado diferencialmente, estando el producto génico codificado por un gen que comprende la secuencia de SEQ ID NO: 1 en una muestra de análisis, estando la muestra de análisis derivada de una célula de análisis sospechosa de ser una célula de colon cancerosa; y

comparar el nivel de expresión del producto génico expresado diferencialmente detectado con el nivel de expresión del producto génico expresado diferencialmente en una muestra control, estando la muestra control derivada de una célula de colon cancerosa;

en el que la detección del nivel de expresión del producto génico expresado diferencialmente en la muestra de análisis similar al nivel de expresión del producto génico en la muestra control indica que la célula de análisis es una célula de colon cancerosa.

7. El procedimiento de la reivindicación 6, en el que dicha detección se hace por hibridación de la muestra de análisis con un alineamiento de referencia, comprendiendo el alineamiento de referencia un polinucleótido que comprende al menos 12 nucleótidos contiguos de la secuencia de SEQ ID NO: 1.

8. El procedimiento de la reivindicación 6, en el que el producto génico detectado es un polipéptido.

9. Un procedimiento para identificar una célula de colon cancerosa, comprendiendo el procedimiento las etapas de:

detectar al menos un producto génico expresado diferencialmente, estando el producto génico codificado por un gen que comprende la secuencia de SEQ ID NO: 1 en una muestra de análisis, estando la muestra de análisis derivada de una célula de análisis sospechosa de ser una célula de colon cancerosa; y

comparar el nivel de expresión del producto génico expresado diferencialmente detectado con el nivel de expresión del producto génico expresado diferencialmente en una muestra control, estando la muestra control derivada de una célula de colon normal;

en el que la detección de un aumento del nivel de expresión del producto génico expresado diferencialmente en la muestra de análisis con respecto al nivel de expresión del producto génico en la muestra de control indica que la célula de análisis es una célula de colon cancerosa.

10. El procedimiento de la reivindicación 9, en el que la detección del nivel de expresión del producto génico expresado diferencialmente de la muestra de análisis mayor que el nivel de expresión del producto génico de la muestra control indica que la célula de análisis es una célula de tumor de colon.

11. El procedimiento de la reivindicación 9, en el que la detección del nivel de expresión del producto génico expresado diferencialmente en la muestra de análisis mayor al nivel de expresión del producto génico de la muestra control indica que la célula de análisis es una célula de tumor de colon metastásico.

5 12. Un procedimiento para identificar una célula de colon cancerosa, comprendiendo el procedimiento las etapas de:

10 detectar al menos un producto génico expresado diferencialmente, en el que la detección es mediante la detección de la hibridación de un polinucleótido que comprende la secuencia de SEQ ID NO: 1 en una muestra de análisis, estando la muestra de análisis derivada de una célula de análisis sospechosa de ser una célula de colon cancerosa; y

15 comparar el nivel de hibridación del producto génico expresado diferencialmente detectado con el nivel de hibridación del producto génico expresado diferencialmente en una muestra control, estando la muestra control derivada de una célula de colon cancerosa;

20 en el que la detección de un aumento del nivel de hibridación del producto génico expresado diferencialmente en la muestra de análisis con respecto al nivel de hibridación del producto génico de la muestra control indica que la célula de análisis es una célula de colon cancerosa.

25 13. Un procedimiento para identificar una célula de colon cancerosa, comprendiendo el procedimiento las etapas de:

30 detectar al menos un producto génico expresado diferencialmente, en el que la detección es mediante la detección de la hibridación de un polinucleótido que comprende la secuencia de SEQ ID NO: 1 en una muestra de análisis, estando la muestra de análisis derivada de una célula de análisis sospechosa de ser una célula de colon cancerosa; y

35 comparar el nivel de hibridación del producto génico expresado diferencialmente detectado con el nivel de hibridación del producto génico expresado diferencialmente en una muestra control, estando la muestra control derivada de una célula de colon cancerosa;

40 en el que la detección de un aumento del nivel de hibridación del producto génico expresado diferencialmente en la muestra de análisis con respecto al nivel de hibridación del producto génico de la muestra control indica que la célula de análisis es una célula de colon cancerosa.

45 14. Uso de un polinucleótido antisentido para la inhibición de la expresión de un gen que comprende la secuencia de SEQ ID NO: 1 en la fabricación de un medicamento para suprimir o inhibir un fenotipo canceroso de una célula de colon cancerosa.

50 15. El uso de la reivindicación 14, en el que el fenotipo canceroso es metástasis.

55 16. El uso de la reivindicación 14, en el que el fenotipo canceroso es una proliferación celular aberrante en relación con una célula normal.

60 17. El uso de la reivindicación 14, en el que el fenotipo canceroso es la pérdida de la inhibición por contacto del crecimiento celular.

65 18. Uso de un polinucleótido antisentido para la inhibición de la producción de un producto génico codificado por un polinucleótido que comprende la secuencia de SEQ ID NO: 1 o de un anticuerpo que se une específicamente a un polipéptido codificado por un polinucleótido que comprende la secuencia de SEQ ID NO: 1 en la fabricación de un medicamento para inhibir el crecimiento tumoral en el colon de un sujeto que tiene un tumor de colon que expresa un gen que comprende una secuencia de SEQ ID NO: 1.

70 19. Un procedimiento para identificar un producto génico como una diana para un cáncer terapéutico, comprendiendo el procedimiento:

75 poner en contacto una célula de colon cancerosa que expresa un producto génico candidato con un agente anticancerígeno, en el que el producto génico candidato se corresponde con la secuencia de SEQ ID NO: 1; y

80 analizar el efecto del agente anticancerígeno tras la expresión del producto génico candidato y tras un fenotipo canceroso de la célula cancerosa;

85 en el que la modulación de la actividad biológica del producto génico candidato y la modulación del fenotipo canceroso de la célula cancerosa indica que el producto génico candidato es una diana para un cáncer terapéutico.

90 20. El procedimiento de la reivindicación 19, en el que el agente anticancerígeno es un oligonucleótido antisentido.

## ES 2 305 087 T3

21. El procedimiento de la reivindicación 19, en el que el fenotipo canceroso es una proliferación celular aberrante en comparación con una célula normal.

22. El procedimiento de la reivindicación 19, en el que el fenotipo canceroso es una formación de colonias debida a la pérdida de inhibición del crecimiento por contacto.

23. Un procedimiento para identificar agentes que disminuyan la actividad biológica de un producto génico expresado diferencialmente en una célula cancerosa, comprendiendo el procedimiento:

poner en contacto un agente candidato con un producto génico expresado diferencialmente, siendo el producto génico expresado diferencialmente un producto génico de ARNm codificado por la secuencia de SEQ ID NO: 1, un producto génico de ADNc preparado a partir del producto génico de ARNm, o un producto génico polipeptídico expresado por el producto génico de ARNm o ADNc; y

detectar una disminución en una actividad biológica del producto génico en relación con un nivel de actividad biológica del producto génico en ausencia del agente candidato.

24. El procedimiento de la reivindicación 23, en el que dicha detección es mediante la detección de una disminución de la expresión del producto génico expresado diferencialmente.

25. Un polinucleótido aislado que comprende la secuencia de nucleótidos de SEQ ID NO: 1.

26. Un alineamiento que comprende el polinucleótido de la reivindicación 25.

27. Un alineamiento que comprende al menos dos polinucleótidos diferentes, en el que los polinucleótidos comprenden una secuencia que tiene la secuencia de SEQ ID NO: 1.

28. Una célula huésped recombinante que contiene el polinucleótido de la reivindicación 25.

29. Un polipéptido aislado codificado por el polinucleótido de la reivindicación 25.

30. Un anticuerpo que se une específicamente a un polipéptido de la reivindicación 29.

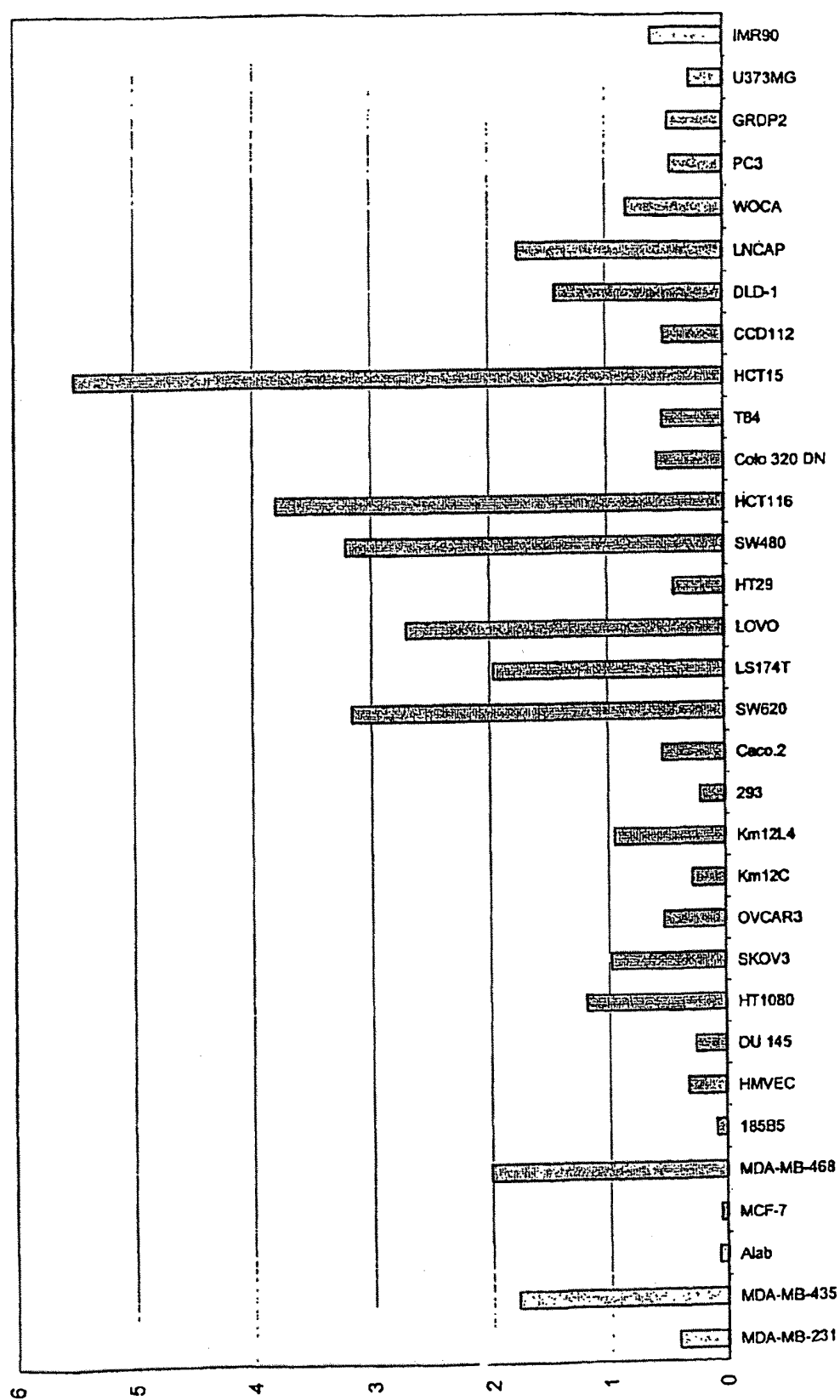
31. Un polinucleótido que comprende la secuencia de nucleótidos de un inserto contenido en el clon SK-1, depositado con el n.º de acceso PTA-1360 en la ATCC.

32. Un polinucleótido aislado que comprende la secuencia codificante del polipéptido de SEQ ID NO: 2.

33. Una composición farmacéutica que comprende un polinucleótido antisentido para la inhibición de la producción de un producto génico codificado por un polinucleótido que tiene una secuencia de SEQ ID NO: 1.

34. La composición farmacéutica de la reivindicación 33, en la que el polinucleótido antisentido comprende una secuencia seleccionada del grupo constituido por SEQ ID NO: 68, 70, 72 y 74.

Fig. 1 Niveles de los mensajes del gen correspondiente a c9083



Evaluación funcional de c9083

Fig. 2 Efecto de los oligonucleótidos AS 9083

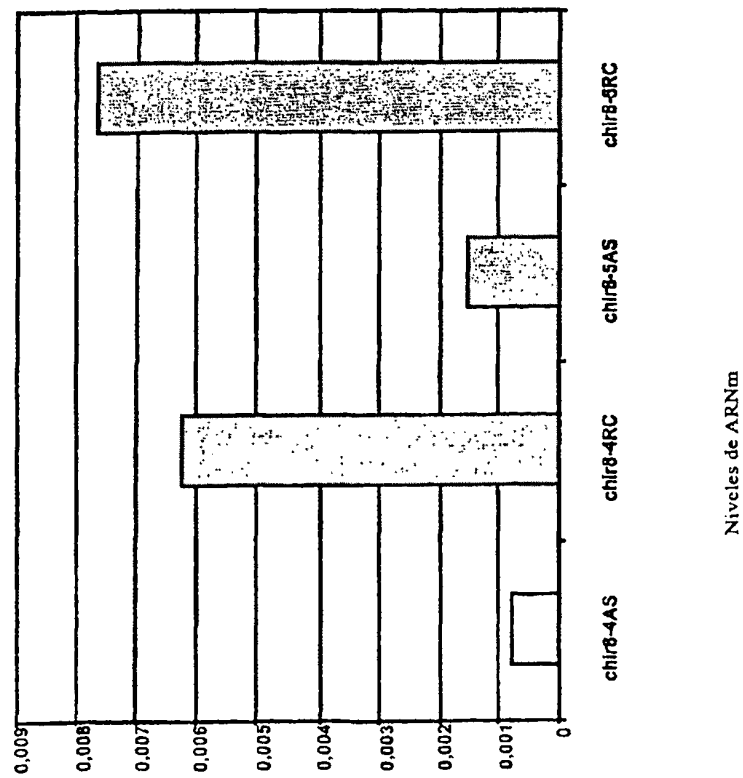
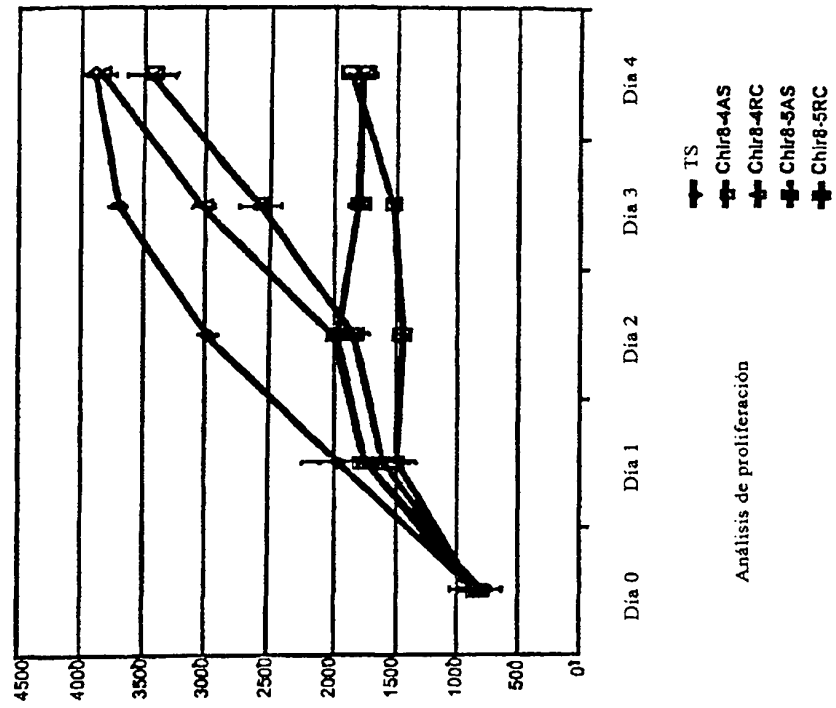
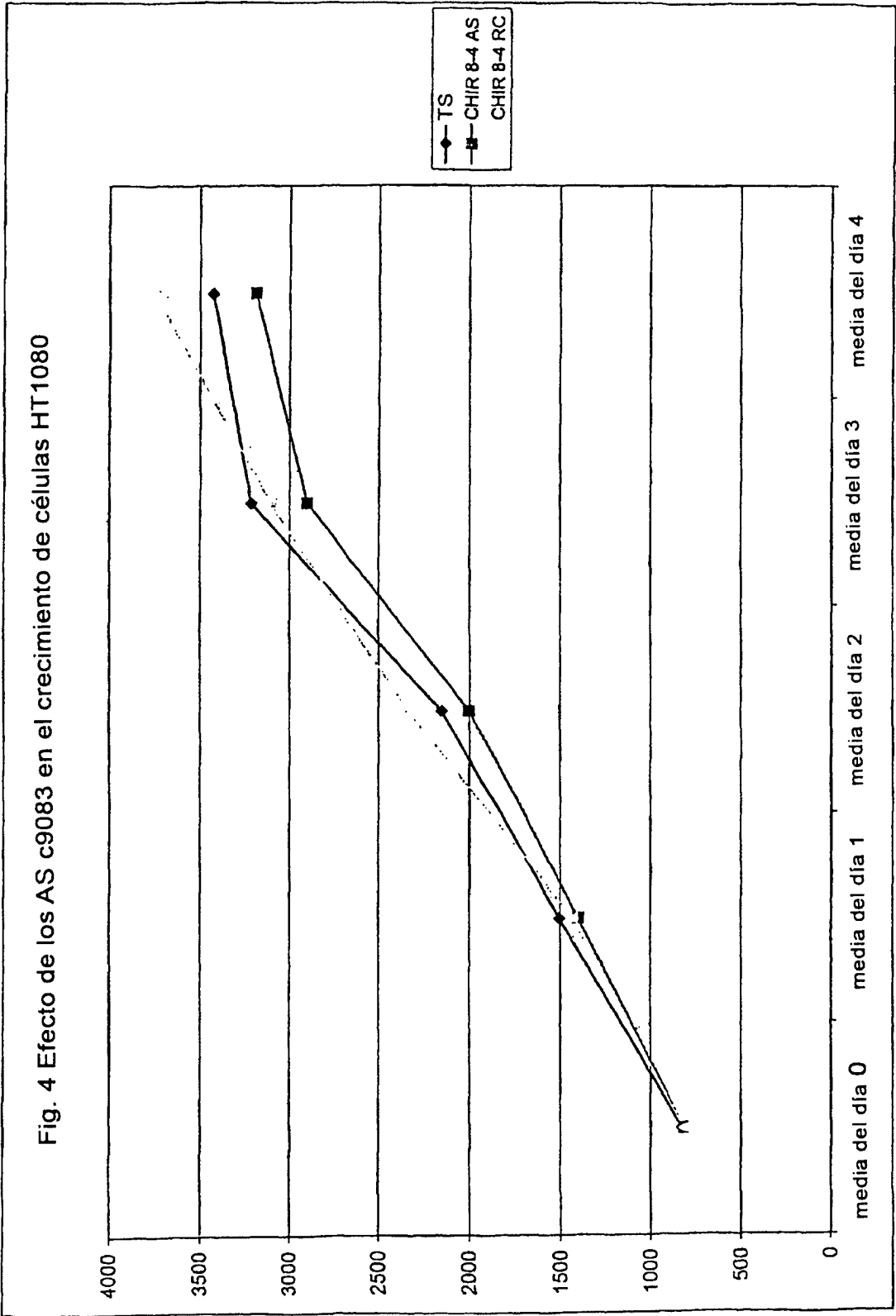


Fig. 3 Efecto de los AS c9083 en el crecimiento de SW620







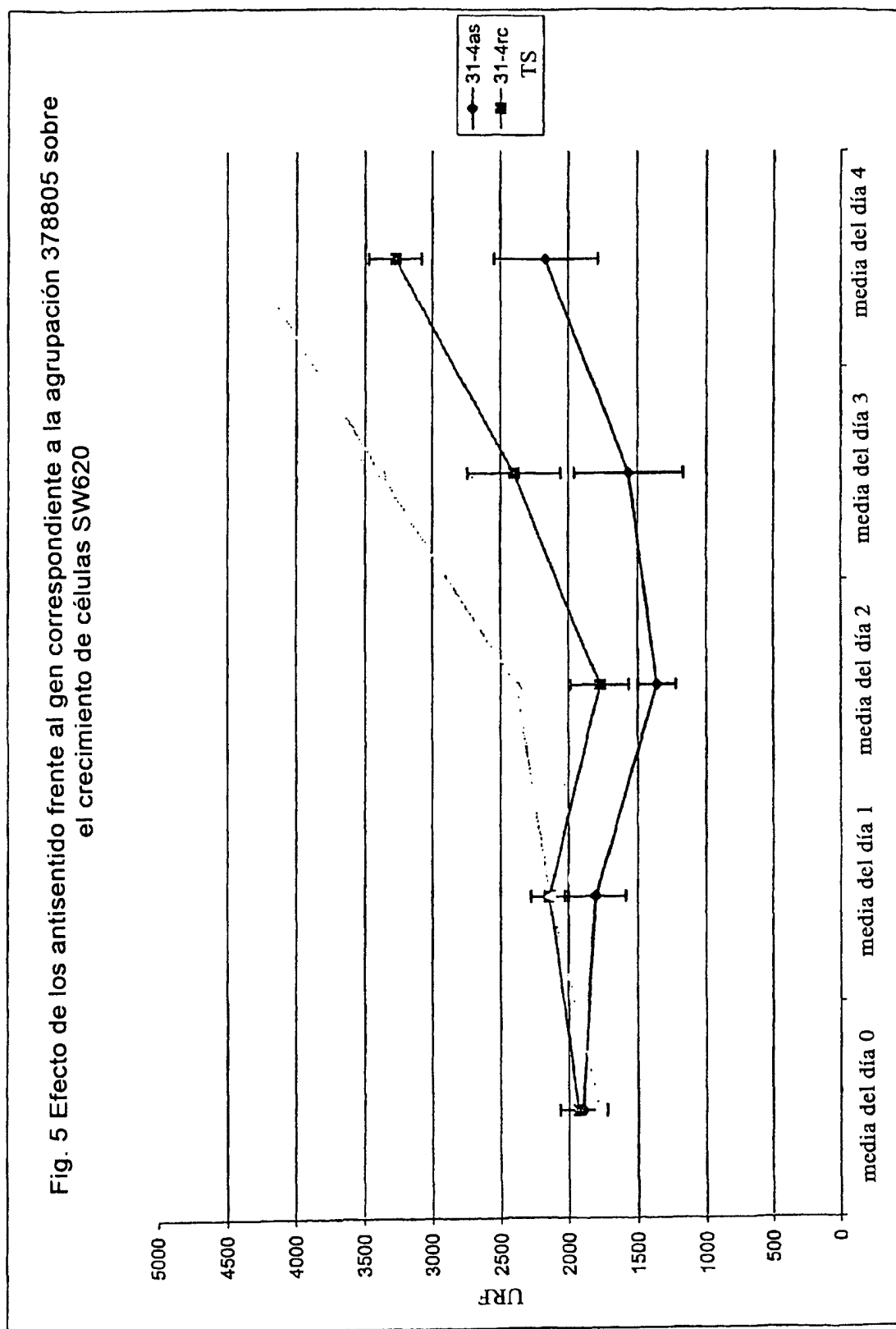


Fig. 6 Prolif. de SW620 con K-Ras oligo 2576 (control)

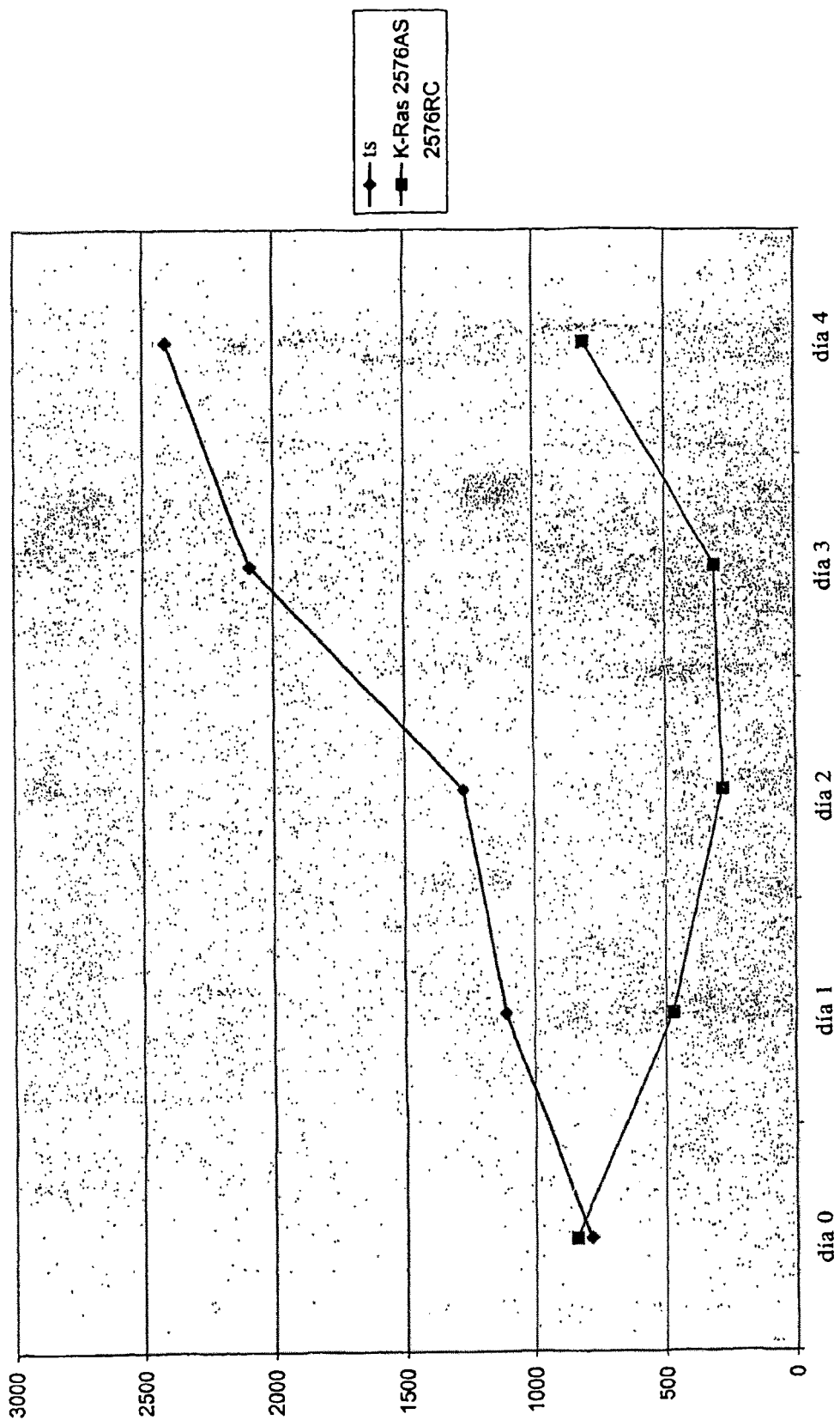


Fig. 7 SW620 transfectadas con chir11-4 AS/RC 300nM

(c3376:4 AS/RC)

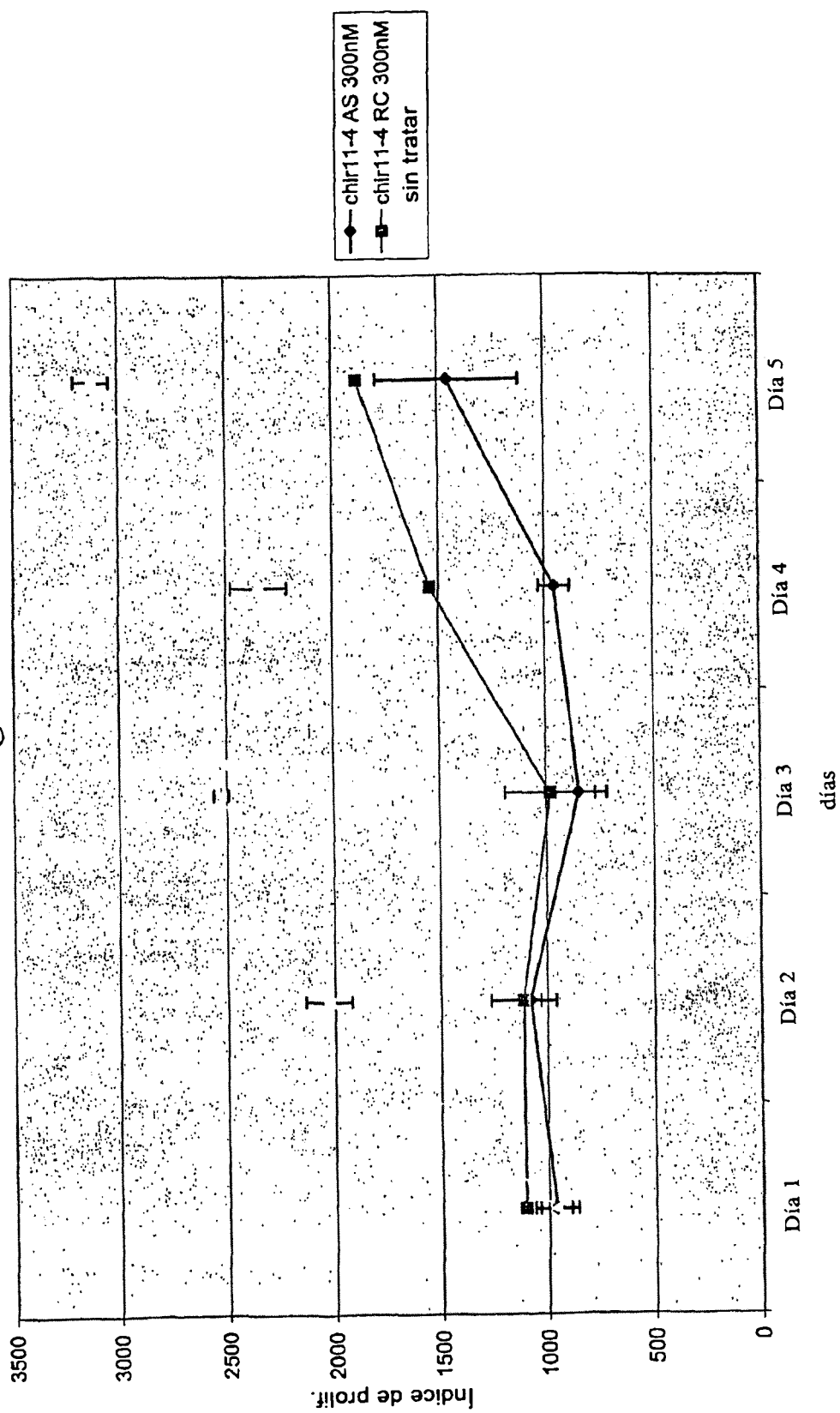


Fig. 8 Prolif. de SW620 con Chir 33-4  
(402380:4-AS(2C))

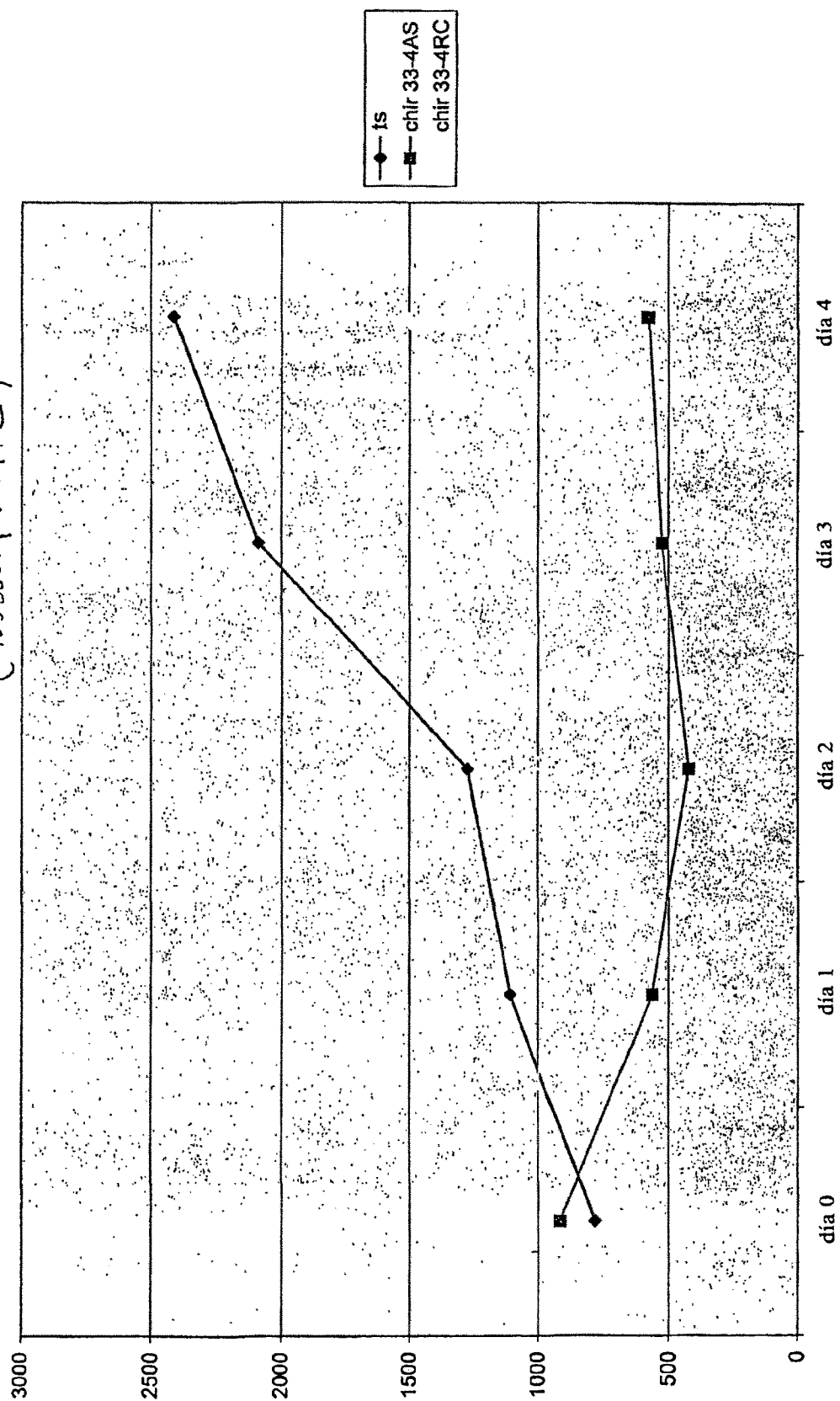
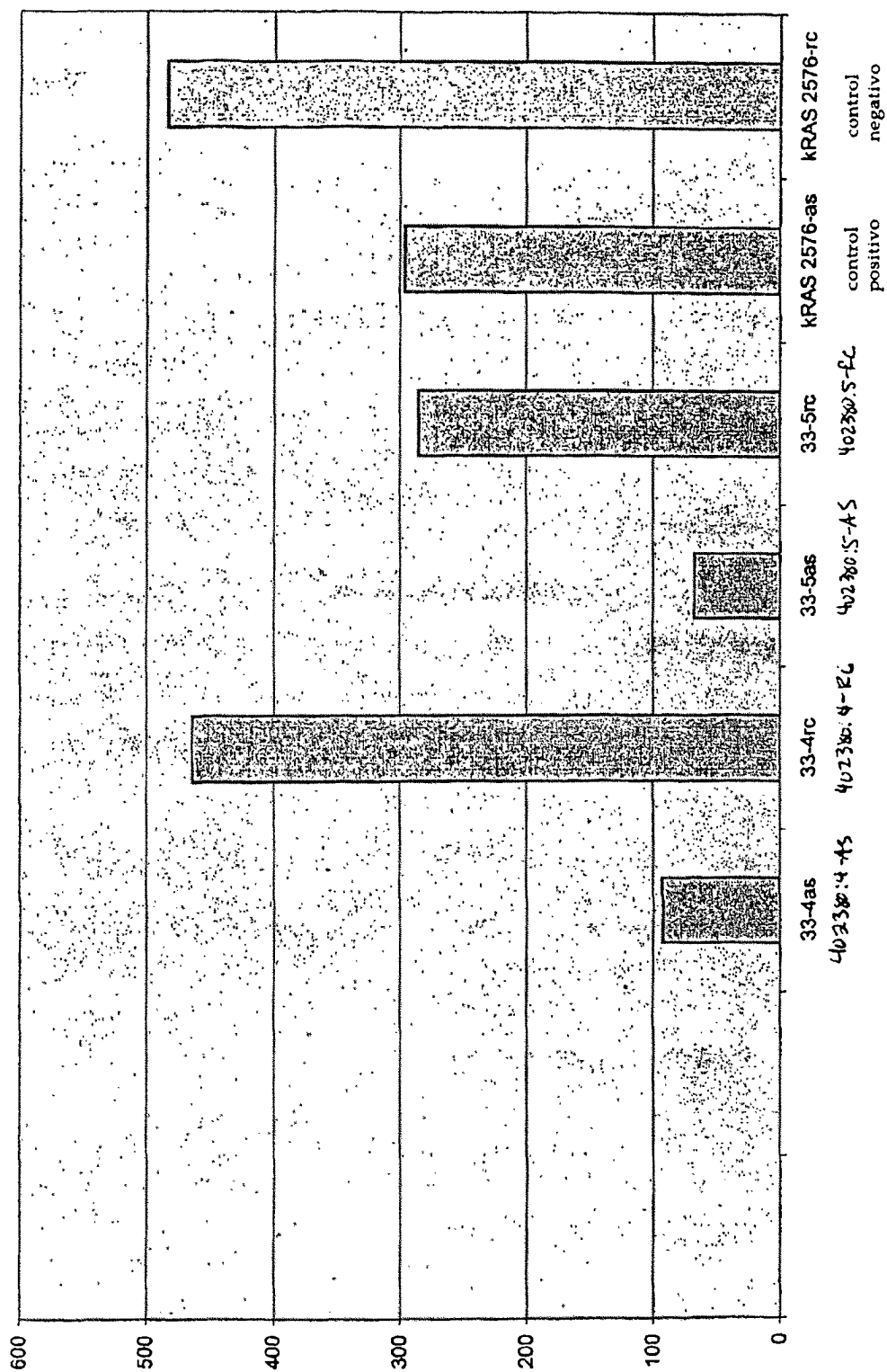


FIG. 9 colonias de SW620 an agar disuelto normalizadas con respecto a WST1



## ES 2 305 087 T3

### LISTA DE SECUENCIAS

5	<110> Kennedy, Giulia C. Kang, Sanmao Reinhard, Christoph Jefferson, Anne Bennett
10	<120> POLINUCLEÓTIDOS RELACIONADOS CON EL CÁNCER DE COLON
	<130> PP-1663.003
15	<140> sin conceder <141> 15-06-2001
20	<150> 60/211.835 <151> 15-06-2000
	<160> 127
25	<170> FastSEQ para Windows versión 4.0
	<210> 1 <211> 564 <212> ADN
30	<213> <i>Homo sapiens</i>
35	<220> <221> CDS <222> (21)...(396)
40	
45	
50	
55	
60	
65	

# ES 2 305 087 T3

<400> 1

```

5      ggcgagattt gtgcggcgac atg aaa ctg ctt acc cac aat ctg ctg agc tcg      53
      Met Lys Leu Leu Thr His Asn Leu Leu Ser Ser
      1 5 10

10     cat gtg cgg ggg gtg ggg tcc cgt ggc ttc ccc ctg cgc ctc cag gcc      101
      His Val Arg Gly Val Gly Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala
      15 20 25

15     acc gag gtc cgt atc tgc cct gtg gaa ttc aac ccc aac ttc gtg gcg      149
      Thr Glu Val Arg Ile Cys Pro Val Glu Phe Asn Pro Asn Phe Val Ala
      30 35 40

20     cgt atg ata cct aaa gtg gag tgg tgc gcg ttc ctg gag gcg gcc gat      197
      Arg Met Ile Pro Lys Val Glu Trp Ser Ala Phe Leu Glu Ala Ala Asp
      45 50 55

25     aac ttg cgt ctg atc cag gtg ccg aaa ggg ccg gtt gag gga tat gag      245
      Asn Leu Arg Leu Ile Gln Val Pro Lys Gly Pro Val Glu Gly Tyr Glu
      60 65 70 75

30     gag aat gag gag ttt ctg agg acc atg cac cac ctg ctg ctg gag gtg      293
      Glu Asn Glu Glu Phe Leu Arg Thr Met His His Leu Leu Leu Glu Val
      80 85 90

35     gaa gtg ata gag ggc acc ctg cag tgc ccg gaa tct gga cgt atg ttc      341
      Glu Val Ile Glu Gly Thr Leu Gln Cys Pro Glu Ser Gly Arg Met Phe
      95 100 105

40     ccc atc agc cgc ggg atc ccc aac atg ctg ctg agt gaa gag gaa act      389
      Pro Ile Ser Arg Gly Ile Pro Asn Met Leu Leu Ser Glu Glu Glu Thr
      110 115 120

45     gag agt t gattgtgcca ggcgccagtt tttcttggtta tgactgtgta tttttgttga      446
      Glu Ser
      125

50     tctataccct gtttccgaat tctgccgtgt gtatcccca cccttgaccc aatgacacca      506
      aacacagtgt ttttgagctc ggtattatat atttttttct cattaaaggt ttaaaacc      564

```

<210> 2

<211> 125

<212> PRT

<213> *Homo sapiens*

<400> 2

```

50     Met Lys Leu Leu Thr His Asn Leu Leu Ser Ser His Val Arg Gly Val
      1 5 10 15
      Gly Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala Thr Glu Val Arg Ile
      20 25 30
      Cys Pro Val Glu Phe Asn Pro Asn Phe Val Ala Arg Met Ile Pro Lys
      35 40 45
55     Val Glu Trp Ser Ala Phe Leu Glu Ala Ala Asp Asn Leu Arg Leu Ile
      50 55 60
      Gln Val Pro Lys Gly Pro Val Glu Gly Tyr Glu Glu Asn Glu Glu Phe
      65 70 75 80
60     Leu Arg Thr Met His His Leu Leu Leu Glu Val Glu Val Ile Glu Gly
      85 90 95
      Thr Leu Gln Cys Pro Glu Ser Gly Arg Met Phe Pro Ile Ser Arg Gly
      100 105 110
65     Ile Pro Asn Met Leu Leu Ser Glu Glu Glu Thr Glu Ser
      115 120 125

```

ES 2 305 087 T3

 $\langle 210 \rangle_3$ 

<211> 919

<212> ADN

5 <213> *Homo sapiens*

 $\langle 220 \rangle$ 

<221> CDS

10  $\langle 222 \rangle$  (219)...(693)

<400> 3

15	tggcacgagg	tggcacgagg	gtccgggtcg	ttgaggatta	ggctgctcgg	gcgtaaccgg	60
	agctggggcg	cggtcgccaa	gggccggccc	ggaagtccca	gcggtcttta	aattctcccg	120
	tgctagggcc	agcctgcgca	ttcttacctg	tcgggggtcg	gcgagtgtct	cacctctctg	180
	cacttccaag	gactcttgtc	atctgcctta	ggcgggaa	atg ctg ttg ctg gat tgc		236
				Met	Leu	Leu Leu Asp Cys	
20				1		5	
	aac ccc gag	gtg gat ggt	ctg aag cat	ttg ctg gag	aca ggg gcc	tcg	284
	Asn Pro Glu	Val Asp Gly	Leu Lys His	Leu Leu Glu	Thr Gly Ala	Ser	
		10		15		20	
25	gtc aac gca	ccc ccg gat	ccc tgc aag	cag tcg cct	gtc cac tta	gcc	332
	Val Asn Ala	Pro Pro Asp	Pro Cys Lys	Gln Ser Pro	Val His Leu	Ala	
		25		30		35	
30	gca gga agc	ggc ctt gct	tgc ttt ctt	ctc tgg cag	ctg caa acg	ggc	380
	Ala Gly Ser	Gly Leu Ala	Cys Phe Leu	Leu Leu Trp	Gln Leu Gln	Thr Gly	
		40		45		50	
35	gct gac ctc	aac cag cag	gat gtt tta	gga gaa gct	cca cta cac	aag	428
	Ala Asp Leu	Asn Gln Gln	Asp Val Leu	Gly Glu Ala	Pro Leu His	Lys	
		55		60		70	
40	gca gca aaa	ggt gga agc	ctg gag tgc	cta agc ctg	ctt gta gcc	agt	476
	Ala Ala Lys	Val Gly Ser	Leu Glu Cys	Leu Ser Leu	Leu Val Ala	Ser	
		75		80		85	
45	gat gcc caa	att gat tta	tgt aat aag	aac ggg caa	aca gct gaa	gat	524
	Asp Ala Gln	Ile Asp Leu	Cys Asn Lys	Asn Gly Gln	Thr Ala Glu	Asp	
		90		95		100	
50	ctc gct tgg	tca tgt gga	ttt cca gac	tgt gcc aag	ttt ctt aca	aca	572
	Leu Ala Trp	Ser Cys Gly	Phe Pro Asp	Cys Ala Lys	Phe Leu Thr	Thr	
		105		110		115	
55	att aaa tgt	atg cag aca	ata aaa gca	agt gaa cac	cct gac agg	aat	620
	Ile Lys Cys	Met Gln Thr	Ile Lys Ala	Ser Glu His	Pro Asp Arg	Asn	
		120		125		130	
60	gat tgt gtt	gcc gtg ctc	aga cag aaa	cgg agt ctc	gga agt gta	gaa	668
	Asp Cys Val	Ala Val Leu	Arg Gln Lys	Arg Ser Leu	Gly Ser Val	Glu	
		135		140		150	
65	aat acc agt	ggg aaa agg	aag tgc t	gatgtcacgt	gggttatgaa		713
	Asn Thr Ser	Gly Lys Arg	Lys Cys				
		155					
70	gaagtctgaa	gaacgccttc	atttcatgca	aatctataag	ctcctgcttt	tggttttacc	773
	atatgttgtg	tctaattctcc	ttctgagaag	gacgaaaaac	tttcttccaa	gtgaagatcc	833
	atttaagaac	acatgtattt	acatgcctat	aatatgctgg	ttgtgtatgc	tttgtctttt	893
	aagttatttaa	aggaacgtct	aaaaaa				919



# ES 2 305 087 T3

<210> 4

<211> 158

<212> PRT

5 <213> *Homo sapiens*

<400> 4

10

```

Met Leu Leu Leu Asp Cys Asn Pro Glu Val Asp Gly Leu Lys His Leu
 1      5      10
Leu Glu Thr Gly Ala Ser Val Asn Ala Pro Pro Asp Pro Cys Lys Gln
      20      25
15 Ser Pro Val His Leu Ala Ala Gly Ser Gly Leu Ala Cys Phe Leu Leu
      35      40      45
Trp Gln Leu Gln Thr Gly Ala Asp Leu Asn Gln Gln Asp Val Leu Gly
 50      55      60
20 Glu Ala Pro Leu His Lys Ala Ala Lys Val Gly Ser Leu Glu Cys Leu
 65      70      75      80
Ser Leu Leu Val Ala Ser Asp Ala Gln Ile Asp Leu Cys Asn Lys Asn
      85      90      95
Gly Gln Thr Ala Glu Asp Leu Ala Trp Ser Cys Gly Phe Pro Asp Cys
      100      105      110
25 Ala Lys Phe Leu Thr Thr Ile Lys Cys Met Gln Thr Ile Lys Ala Ser
      115      120      125
Glu His Pro Asp Arg Asn Asp Cys Val Ala Val Leu Arg Gln Lys Arg
      130      135      140
30 Ser Leu Gly Ser Val Glu Asn Thr Ser Gly Lys Arg Lys Cys
 145      150      155

```

<210> 5

35

<211> 1949

<212> ADN

<213> *Homo sapiens*

40

<220>

<221> CDS

<222> (5)...(1760)

45

50

55

60

65

# ES 2 305 087 T3

<400> 5

5	caac atg gcg ccg tcc acg ccg ctc ttg aca gtc cga gga tca gaa gga	49
	Met Ala Pro Ser Thr Pro Leu Leu Thr Val Arg Gly Ser Glu Gly	
	1 5 10 15	
10	ctg tac atg gtg aat gga cca cca cat ttt aca gaa agc aca gtg ttt	97
	Leu Tyr Met Val Asn Gly Pro Pro His Phe Thr Glu Ser Thr Val Phe	
	20 25 30	
15	cca agg gaa tct ggg aag aat tgc aaa gtc tgt atc ttt agt aag gat	145
	Pro Arg Glu Ser Gly Lys Asn Cys Lys Val Cys Ile Phe Ser Lys Asp	
	35 40 45	
20	ggg acc ttg ttt gcc tgg ggc aat gga gaa aaa gta aat att atc agt	193
	Gly Thr Leu Phe Ala Trp Gly Asn Gly Glu Lys Val Asn Ile Ile Ser	
	50 55 60	
25	gtc act aac aag gga cta ctg cac tcc ttc gac ctc ctg aag gca gtt	241
	Val Thr Asn Lys Gly Leu Leu His Ser Phe Asp Leu Leu Lys Ala Val	
	65 70 75	
30	tgc ctt gaa ttc tca ccc aaa aat act gtc ctg gca acg tgg cag cct	289
	Cys Leu Glu Phe Ser Pro Lys Asn Thr Val Leu Ala Thr Trp Gln Pro	
	80 85 90 95	
35	tac act act tct aaa gat ggc aca gct ggg ata ccc aac cta caa ctt	337
	Tyr Thr Thr Ser Lys Asp Gly Thr Ala Gly Ile Pro Asn Leu Gln Leu	
	100 105 110	
40	tat gat gtg aaa act ggg aca tgt ttg aaa tct ttc atc cag aaa aaa	385
	Tyr Asp Val Lys Thr Gly Thr Cys Leu Lys Ser Phe Ile Gln Lys Lys	
	115 120 125	
45	atg caa aat tgg tgt cca tcc tgg tca gaa gat gaa act ctt tgt gcc	433
	Met Gln Asn Trp Cys Pro Ser Trp Ser Glu Asp Glu Thr Leu Cys Ala	
	130 135 140	
50	cgc aat gtt aac aat gaa gtt cac ttc ttt gaa aac aac aat ttt aac	481
	Arg Asn Val Asn Asn Glu Val His Phe Phe Glu Asn Asn Asn Phe Asn	
	145 150 155	
55	aca att gca aat aaa ttg cat ttg caa aaa att aat gac ttt gta tta	529
	Thr Ile Ala Asn Lys Leu His Leu Gln Lys Ile Asn Asp Phe Val Leu	
	160 165 170 175	
60	tca cct gga ccc caa cca tac aag gtg gct gtc tat gtt cca gga agt	577
	Ser Pro Gly Pro Gln Pro Tyr Lys Val Ala Val Tyr Val Pro Gly Ser	
	180 185 190	
65	aaa ggt gca cct tca ttt gtt aga tta tat cag tac ccc aac ttt gct	625
	Lys Gly Ala Pro Ser Phe Val Arg Leu Tyr Gln Tyr Pro Asn Phe Ala	
	195 200 205	
70	gga cct cat gca gct tta gct aat aaa agt ttc ttt aag gca gat aaa	673
	Gly Pro His Ala Ala Leu Ala Asn Lys Ser Phe Phe Lys Ala Asp Lys	
	210 215 220	
75	gtt aca atg ctg tgg aat aaa aaa gct act gct gtg ttg gta ata gct	721
	Val Thr Met Leu Trp Asn Lys Lys Ala Thr Ala Val Leu Val Ile Ala	
	225 230 235	

# ES 2 305 087 T3

5	agc aca gat gtt gac aag aca gga gct tcc tac tat gga gaa caa act Ser Thr Asp Val Asp Lys Thr Gly Ala Ser Tyr Tyr Gly Glu Gln Thr 240 245 250 255	769
10	cta cac tac att gca aca aat gga gaa agt gct gta gtg caa tta cca Leu His Tyr Ile Ala Thr Asn Gly Glu Ser Ala Val Val Gln Leu Pro 260 265 270	817
15	aaa aat ggc ccc att tat gat gta gtt tgg aat tct agt tct act gag Lys Asn Gly Pro Ile Tyr Asp Val Val Trp Asn Ser Ser Ser Thr Glu 275 280 285	865
20	ttt tgt gct gta tat ggt ttt atg cct gcc aaa gcg aca att ttc aac Phe Cys Ala Val Tyr Gly Phe Met Pro Ala Lys Ala Thr Ile Phe Asn 290 295 300	913
25	ttg aaa tgt gat cct gta ttt gac ttt gga act ggt cct cgt aat gca Leu Lys Cys Asp Pro Val Phe Asp Phe Gly Thr Gly Pro Arg Asn Ala 305 310 315	961
30	gcc tac tat agc cct cat gga cat ata tta gta tta gct gga ttt gga Ala Tyr Tyr Ser Pro His Gly His Ile Leu Val Leu Ala Gly Phe Gly 320 325 330 335	1009
35	aat ctg agg gga caa atg gaa gtg tgg gat gtg aaa aac tac aaa ctt Asn Leu Arg Gly Gln Met Glu Val Trp Asp Val Lys Asn Tyr Lys Leu 340 345 350	1057
40	att tct aaa ccg gtg gct tct gat tct aca tat ttt gct tgg tgc ccg Ile Ser Lys Pro Val Ala Ser Asp Ser Thr Tyr Phe Ala Trp Cys Pro 355 360 365	1105
45	gat ggt gag cat att tta aca gct aca tgt gct ccc agg tta cgg gtt Asp Gly Glu His Ile Leu Thr Ala Thr Cys Ala Pro Arg Leu Arg Val 370 375 380	1153
50	aat aat gga tac aaa att tgg cat tat act ggc tct atc ttg cac aag Asn Asn Gly Tyr Lys Ile Trp His Tyr Thr Gly Ser Ile Leu His Lys 385 390 395	1201
55	tat gat gtg cca tca aat gca gaa tta tgg cag gtt tct tgg cag cca Tyr Asp Val Pro Ser Asn Ala Glu Leu Trp Gln Val Ser Trp Gln Pro 400 405 410 415	1249
60	ttt ttg gat gga ata ttt cca gca aaa aca ata act tac caa gca gtt Phe Leu Asp Gly Ile Phe Pro Ala Lys Thr Ile Thr Tyr Gln Ala Val 420 425 430	1297
65	cca agt gaa gta ccc aat gag gaa cct aaa gtt gca aca gct tat aga Pro Ser Glu Val Pro Asn Glu Glu Pro Lys Val Ala Thr Ala Tyr Arg 435 440 445	1345
70	ccc cca gct tta aga aat aaa cca atc acc aat tcc aaa ttg cat gaa Pro Pro Ala Leu Arg Asn Lys Pro Ile Thr Asn Ser Lys Leu His Glu 450 455 460	1393
75	gag gaa cca cct cag aat atg aaa cca caa tca gga aac gat aag cca Glu Glu Pro Pro Gln Asn Met Lys Pro Gln Ser Gly Asn Asp Lys Pro 465 470 475	1441
80	tta tca aaa aca gct ctt aaa aat caa agg aag cat gaa gct aag aaa Leu Ser Lys Thr Ala Leu Lys Asn Gln Arg Lys His Glu Ala Lys Lys 480 485 490 495	1489

# ES 2 305 087 T3

```

gct gca aag cag gaa gca aga agt gac aag agt cca gat ttg gca cct 1537
Ala Ala Lys Gln Glu Ala Arg Ser Asp Lys Ser Pro Asp Leu Ala Pro
500 505 510

5 act cct gcc cca cag agc aca cca cga aac act gtc tct cag tca att 1585
Thr Pro Ala Pro Gln Ser Thr Pro Arg Asn Thr Val Ser Gln Ser Ile
515 520 525

10 tct ggg gac cct gag ata gac aaa aaa atc aag aac cta aag aag aaa 1633
Ser Gly Asp Pro Glu Ile Asp Lys Lys Ile Lys Asn Leu Lys Lys Lys
530 535 540

ctg aaa gca atc gaa caa ctg aaa gaa caa gca gca act gga aaa cag 1681
Leu Lys Ala Ile Glu Gln Leu Lys Glu Gln Ala Ala Thr Gly Lys Gln
545 550 555

15 cta gaa aaa aat cag ttg gag aaa att cag aaa gaa aca gcc ctt ctc 1729
Leu Glu Lys Asn Gln Leu Glu Lys Ile Gln Lys Glu Thr Ala Leu Leu
560 565 570 575

20 cag gag ctg gaa gat ttg gaa ttg ggt att t aaagattcac ggaaagcaag 1780
Gln Glu Leu Glu Asp Leu Glu Leu Gly Ile
580 585

25 ttgatgacca gaaatcagtg caaacacatc ttctgttaaa cccattggta tacacagaat , 1840
attcctgtgc ccacacttaa tgtcaatcta taattttaac catttatcca agattctact 1900
aagtgtaaaa ttatttaata atgtctatta aattgatatt tatatcttg 1949

```

<210> 6

30 <211> 585

<212> PRT

<213> *Homo sapiens*

35 <400> 6

```

Met Ala Pro Ser Thr Pro Leu Leu Thr Val Arg Gly Ser Glu Gly Leu
1 5 10 15
Tyr Met Val Asn Gly Pro Pro His Phe Thr Glu Ser Thr Val Phe Pro
20 25 30
Arg Glu Ser Gly Lys Asn Cys Lys Val Cys Ile Phe Ser Lys Asp Gly
35 40 45
Thr Leu Phe Ala Trp Gly Asn Gly Glu Lys Val Asn Ile Ile Ser Val
50 55 60
Thr Asn Lys Gly Leu Leu His Ser Phe Asp Leu Leu Lys Ala Val Cys
65 70 75 80
Leu Glu Phe Ser Pro Lys Asn Thr Val Leu Ala Thr Trp Gln Pro Tyr
85 90 95
Thr Thr Ser Lys Asp Gly Thr Ala Gly Ile Pro Asn Leu Gln Leu Tyr
100 105 110
Asp Val Lys Thr Gly Thr Cys Leu Lys Ser Phe Ile Gln Lys Lys Met
115 120 125
Gln Asn Trp Cys Pro Ser Trp Ser Glu Asp Glu Thr Leu Cys Ala Arg
130 135 140
Asn Val Asn Asn Glu Val His Phe Phe Glu Asn Asn Asn Phe Asn Thr
145 150 155 160
Ile Ala Asn Lys Leu His Leu Gln Lys Ile Asn Asp Phe Val Leu Ser
165 170 175
Pro Gly Pro Gln Pro Tyr Lys Val Ala Val Tyr Val Pro Gly Ser Lys
180 185 190
Gly Ala Pro Ser Phe Val Arg Leu Tyr Gln Tyr Pro Asn Phe Ala Gly
195 200 205
Pro His Ala Ala Leu Ala Asn Lys Ser Phe Phe Lys Ala Asp Lys Val
210 215 220

```

# ES 2 305 087 T3

	Thr	Met	Leu	Trp	Asn	Lys	Lys	Ala	Thr	Ala	Val	Leu	Val	Ile	Ala	Ser
	225					230					235					240
5	Thr	Asp	Val	Asp	Lys	Thr	Gly	Ala	Ser	Tyr	Tyr	Gly	Glu	Gln	Thr	Leu
					245					250					255	
	His	Tyr	Ile	Ala	Thr	Asn	Gly	Glu	Ser	Ala	Val	Val	Gln	Leu	Pro	Lys
				260					265					270		
	Asn	Gly	Pro	Ile	Tyr	Asp	Val	Val	Trp	Asn	Ser	Ser	Ser	Thr	Glu	Phe
10			275					280					285			
	Cys	Ala	Val	Tyr	Gly	Phe	Met	Pro	Ala	Lys	Ala	Thr	Ile	Phe	Asn	Leu
		290					295					300				
	Lys	Cys	Asp	Pro	Val	Phe	Asp	Phe	Gly	Thr	Gly	Pro	Arg	Asn	Ala	Ala
	305					310					315					320
15	Tyr	Tyr	Ser	Pro	His	Gly	His	Ile	Leu	Val	Leu	Ala	Gly	Phe	Gly	Asn
					325					330					335	
	Leu	Arg	Gly	Gln	Met	Glu	Val	Trp	Asp	Val	Lys	Asn	Tyr	Lys	Leu	Ile
				340					345					350		
	Ser	Lys	Pro	Val	Ala	Ser	Asp	Ser	Thr	Tyr	Phe	Ala	Trp	Cys	Pro	Asp
20			355					360					365			
	Gly	Glu	His	Ile	Leu	Thr	Ala	Thr	Cys	Ala	Pro	Arg	Leu	Arg	Val	Asn
		370					375					380				
	Asn	Gly	Tyr	Lys	Ile	Trp	His	Tyr	Thr	Gly	Ser	Ile	Leu	His	Lys	Tyr
	385					390					395					400
25	Asp	Val	Pro	Ser	Asn	Ala	Glu	Leu	Trp	Gln	Val	Ser	Trp	Gln	Pro	Phe
					405					410					415	
	Leu	Asp	Gly	Ile	Phe	Pro	Ala	Lys	Thr	Ile	Thr	Tyr	Gln	Ala	Val	Pro
				420					425					430		
30	Ser	Glu	Val	Pro	Asn	Glu	Glu	Pro	Lys	Val	Ala	Thr	Ala	Tyr	Arg	Pro
			435					440					445			
	Pro	Ala	Leu	Arg	Asn	Lys	Pro	Ile	Thr	Asn	Ser	Lys	Leu	His	Glu	Glu
		450				455						460				
	Glu	Pro	Pro	Gln	Asn	Met	Lys	Pro	Gln	Ser	Gly	Asn	Asp	Lys	Pro	Leu
35						470					475					480
	Ser	Lys	Thr	Ala	Leu	Lys	Asn	Gln	Arg	Lys	His	Glu	Ala	Lys	Lys	Ala
				485						490					495	
	Ala	Lys	Gln	Glu	Ala	Arg	Ser	Asp	Lys	Ser	Pro	Asp	Leu	Ala	Pro	Thr
				500					505					510		
40	Pro	Ala	Pro	Gln	Ser	Thr	Pro	Arg	Asn	Thr	Val	Ser	Gln	Ser	Ile	Ser
			515					520					525			
	Gly	Asp	Pro	Glu	Ile	Asp	Lys	Lys	Ile	Lys	Asn	Leu	Lys	Lys	Lys	Leu
		530					535					540				
45	Lys	Ala	Ile	Glu	Gln	Leu	Lys	Glu	Gln	Ala	Ala	Thr	Gly	Lys	Gln	Leu
	545					550					555					560
	Glu	Lys	Asn	Gln	Leu	Glu	Lys	Ile	Gln	Lys	Glu	Thr	Ala	Leu	Leu	Gln
					565					570					575	
	Glu	Leu	Glu	Asp	Leu	Glu	Leu	Gly	Ile							
50				580					585							

<210> 7

<211> 1110

55 <212> ADN

<213> *Homo sapiens*

<220>

60 <221> CDS

<222> (78)...(642)

65

# ES 2 305 087 T3

<400> 7

5	tagtgggccc cagtgttgcg ctctctggcc gttccttaca ctttgcttca ggctccagtg	60
	caggggcgta gtgggat atg gcc aac tcg ggc tgc aag gac gtc acg ggt	110
	Met Ala Asn Ser Gly Cys Lys Asp Val Thr Gly	
	1 5 10	
10	cca gat gag gag agt ttt ctg tac ttt gcc tac ggc agc aac ctg ctg	158
	Pro Asp Glu Glu Ser Phe Leu Tyr Phe Ala Tyr Gly Ser Asn Leu Leu	
	15 20 25	
15	aca gag agg atc cac ctc cga aac ccc tcg gcg gcg ttc ttc tgt gtg	206
	Thr Glu Arg Ile His Leu Arg Asn Pro Ser Ala Ala Phe Phe Cys Val	
	30 35 40	
20	gcc cgc ctg cag gat ttt aag ctt gac ttt ggc aat tcc caa ggc aaa	254
	Ala Arg Leu Gln Asp Phe Lys Leu Asp Phe Gly Asn Ser Gln Gly Lys	
	45 50 55	
25	aca agt caa act tgg cat gga ggg ata gcc acc att ttt cag agt cct	302
	Thr Ser Gln Thr Trp His Gly Gly Ile Ala Thr Ile Phe Gln Ser Pro	
	60 65 70 75	
30	ggc gat gaa gtg tgg gga gta gta tgg aaa atg aac aaa agc aat tta	350
	Gly Asp Glu Val Trp Gly Val Val Trp Lys Met Asn Lys Ser Asn Leu	
	80 85 90	
35	aat tct ctg gat gag caa gaa ggg gtt aaa agt gga atg tat gtt gta	398
	Asn Ser Leu Asp Glu Gln Glu Gly Val Lys Ser Gly Met Tyr Val Val	
	95 100 105	
40	ata gaa gtt aaa gtt gca act caa gaa gga aaa gaa ata acc tgt cga	446
	Ile Glu Val Lys Val Ala Thr Gln Glu Gly Lys Glu Ile Thr Cys Arg	
	110 115 120	
45	agt tat ctg atg aca aat tac gaa agt gct ccc cca tcc cca cag tat	494
	Ser Tyr Leu Met Thr Asn Tyr Glu Ser Ala Pro Pro Ser Pro Gln Tyr	
	125 130 135	
50	aaa aag att att tgc atg ggt gca aaa gaa aat ggt ttg ccg ctg gag	542
	Lys Lys Ile Ile Cys Met Gly Ala Lys Glu Asn Gly Leu Pro Leu Glu	
	140 145 150 155	
55	tat caa gag aag tta aaa gca ata gaa cca aat gac tat aca gga aag	590
	Tyr Gln Glu Lys Leu Lys Ala Ile Glu Pro Asn Asp Tyr Thr Gly Lys	
	160 165 170	
60	gtc tca gaa gaa att gaa gac atc atc aaa aag ggg gaa aca caa act	638
	Val Ser Glu Glu Ile Glu Asp Ile Ile Lys Lys Gly Glu Thr Gln Thr	
	175 180 185	
65	ctt t agaacataac agaatatatc taagggtatt ctatgtgcta atataaaata	692
	Leu	
70	tttttaacac ttgagaacag ggatctgggg gatctccacg tttgatccat tttcagcagt	752
	gctctgaagg agtatcttac ttgggtgatt ccttggtttt agactataaaa aagaaactgg	812
	gataggagtt agacaattta aaaggggtgt atgagggcct gaaatatgtg acaaatgaat	872
	gtgagtaccc cttctgtgaa cactgaaagc tattctcttg aattgatctt aagtgtctcc	932
	ttgctctggt aaaagataga tttgtagctc acttgatgat ggtgctggtg aattgctctg	992
	ctctgtctga gattttttaa aatcagctta atgagagtaa tctgcagaca attgataata	1052
	acattttgaa aattggaaag atgggtatact gtttttagag gaataaacgt atttgtgg	1110

# ES 2 305 087 T3

<210> 8

<211> 188

<212> PRT

5 <213> *Homo sapiens*

<400> 8

10

```

Met Ala Asn Ser Gly Cys Lys Asp Val Thr Gly Pro Asp Glu Glu Ser
 1      5      10      15
Phe Leu Tyr Phe Ala Tyr Gly Ser Asn Leu Leu Thr Glu Arg Ile His
 20      25      30
15 Leu Arg Asn Pro Ser Ala Ala Phe Phe Cys Val Ala Arg Leu Gln Asp
 35      40      45
Phe Lys Leu Asp Phe Gly Asn Ser Gln Gly Lys Thr Ser Gln Thr Trp
 50      55      60
20 His Gly Gly Ile Ala Thr Ile Phe Gln Ser Pro Gly Asp Glu Val Trp
 65      70      75      80
Gly Val Val Trp Lys Met Asn Lys Ser Asn Leu Asn Ser Leu Asp Glu
 85      90      95
Gln Glu Gly Val Lys Ser Gly Met Tyr Val Val Ile Glu Val Lys Val
100      105      110
25 Ala Thr Gln Glu Gly Lys Glu Ile Thr Cys Arg Ser Tyr Leu Met Thr
 115      120      125
Asn Tyr Glu Ser Ala Pro Pro Ser Pro Gln Tyr Lys Lys Ile Ile Cys
130      135      140
30 Met Gly Ala Lys Glu Asn Gly Leu Pro Leu Glu Tyr Gln Glu Lys Leu
145      150      155      160
Lys Ala Ile Glu Pro Asn Asp Tyr Thr Gly Lys Val Ser Glu Glu Ile
165      170      175
35 Glu Asp Ile Ile Lys Lys Gly Glu Thr Gln Thr Leu
180      185

```

<210> 9

40 <211> 965

<212> ADN

<213> *Homo sapiens*

45

<220>

<221> CDS

<222> (79)...(232)

50

55

60

65

# ES 2 305 087 T3

<400> 9

```

5      gtagtggggcc ccagtgttgc gctctctggc cgttccttac actttgcttc aggetccagt      60
      gcagggggcgt agtgggat atg gcc aac tcg ggc tgc aag gac gtc acg ggt      111
              Met Ala Asn Ser Gly Cys Lys Asp Val Thr Gly
                1              5              10

10     cca gat gag gag agt ttt ctg tac ttt gcc tac ggc agc aac ctg ctg      159
      Pro Asp Glu Glu Ser Phe Leu Tyr Phe Ala Tyr Gly Ser Asn Leu Leu
                15              20              25

15     aca gag agg atc cac ctc cga aac ccc tcg gcg gcg ttc ttc tgt gtg      207
      Thr Glu Arg Ile His Leu Arg Asn Pro Ser Ala Ala Phe Phe Cys Val
                30              35              40

20     gcc cgc ctg cag gca aga agg ggt t aaaagtggaa tgtatgttgt      252
      Ala Arg Leu Gln Ala Arg Arg Gly
                45              50

25     aatagaagtt aaagttgcaa ctcaagaagg aaaagaaata acctgtcgaa gttatctgat      312
      gacaaattac gaaagtgttc ccccatcccc acagtataaa aagattattht gcatgggtgc      372
      aaaagaaaat ggtttgccgc tggagtatca agagaagtta aaagcaatag aaccaaatga      432
      ctatacagga aaggtctcag aagaaattga agacatcatc aaaaaggggg aaacacaaac      492
25     tctttagaac ataacagaat atatctaagg gtattctatg tgctaataata aaatattttt      552
      aacacttgag aacagggatc tgggggatct ccacgtttga tccattttca gcagtgtctc      612
      gaaggagtat cttacttggg tgattccttg ttttttagact ataaaaagaa actgggatag      672
      gagttagaca atttaaaagg ggtgtatgag ggcctgaaat atgtgacaaa tgaatgtgag      732
      taccctttct gtgaacactg aaagctattc tcttgaattg atcttaagtg tctccttgct      792
30     ctggtaaaag atagatttgt agctcacttg atgatggtgc tgggtgaattg ctctgctctg      852
      tctgagattt ttaaaaatca gcttaatgag agtaatctgc agacaattga taataacatt      912

35     ttgaaaattg gaaagatggg atactgtttt tagaggaata aacgtatttg tgg      965

```

<210> 10

40 <211> 51

<212> PRT

<213> *Homo sapiens*

45 <400> 10

```

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

```

<210> 11

60 <211> 658

<212> ADN

<213> *Homo sapiens*

65



# ES 2 305 087 T3

<400> 11

	tcgagaggag	ctgacagcct	tcttgcaccc	cgaggagtga	ggtggggccac	tgggtgctgc	60
5	ccccgcccc	ggaccagccc	ctggtggaag	ccaaccacct	gctgcacgag	agcgacacgg	120
	acaaggacgg	gcggctgagc	aaagcgga	tcctgggtaa	ttggaacatg	tttgtgggca	180
	gtcaggccac	caactatggt	gaggacctga	cccggcacca	cgatgagctg	tgagccccgc	240
	gcacctgcca	cagcctcaga	ggcccgacac	atgaccggag	gagggggccg	tgtggtcttg	300
	ccccctccct	gtccaggccc	cgcaggaggc	agatgcagtc	ccaggcatcc	tcctgccccct	360
10	gggctctcag	ggacccccctg	ggtcggcttc	tgtccctgtc	acacccccaa	ccccagggag	420
	gggctgtcat	agtcccagag	gataagcaat	acctatttct	gactgagtct	cccagcccag	480
	acccaggggac	cctggcccca	agctcagctc	taagaaccgc	caccaacccc	tccagctcca	540
	aatctgagcc	tccaccacat	agactgaaac	tcccctggcc	ccagccctct	cctgctggc	600
	ctggcctggg	acacctcctc	tctgccagga	ggcaataaaa	gccagcgccg	ggaccttg	658

<210> 12

<211> 1507

<212> ADN

20 <213> *Homo sapiens*

<220>

<221> misc feature

25 <222> 1047, 1301

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

<400> 12

30	ggaacgcaga	gcggagcgtg	gagagcggag	cgaagctgga	taacagggga	ccgatgatgt	60
	ggcgaccatc	agttctgctg	cttctgttgc	tactgaggca	cggggcccag	gggaagccat	120
	cccagacgc	aggccctcat	ggccagggga	gggtgcacca	ggcgcccccc	ctgagcgacg	180
35	ctcccatga	tgacgcccac	gggaacttcc	agtacgacca	tgaggcttcc	ctgggacggg	240
	aagtggccaa	ggaattcgac	caactcacc	cagaggaaag	ccaggcccgt	ctggggcgga	300
	tcgtggaccg	catggaccgc	gcgggggacg	gcgacggctg	ggtgtcgctg	gccgagcttc	360
	gcgcgtggat	cgcgcacacg	cagcagcggc	acatacggga	ctcgggtgagc	gcggcctggg	420
	acacgtacga	cacggaccgc	gacgggcgtg	tgggttggga	ggagctgcgc	aacgccacct	480
40	atggccacta	cgcgcccgtg	gaagaatttc	atgacgtgga	ggatgcagag	acctacaaaa	540
	agatgctggc	tcgggacgag	cggcgcttcc	gggtggccga	ccaggatggg	gactcgatgg	600
	ccactcgaga	ggagctgaca	gccttcctgc	accccaggga	gttccctcac	atgcgggaca	660
	tcgtgattgc	tgaaccctg	gaggacctgg	acagaaacaa	agatggctat	gtccagggtg	720
	aggagtacat	cgcggatctg	tactcagccg	agcctgggga	ggaggagccg	gcgtgggtgc	780
45	agacggagag	gcagcagttc	cgggacttcc	gggatctgaa	caaggatggg	cacctggatg	840
	ggagtgaggt	gggccactgg	gtgctgcccc	ctgcccagga	ccagcccctg	gtggaagcca	900
	accacctgct	gcacgaragc	gacacggaca	aggaygggcg	gctgagcaaa	gcgsaaatcc	960
	tgggtaattg	gaacatgttt	gtgggcagtc	aggccaccaa	ctatggygag	gacctgacct	1020
50	ggcaccacga	tgagctgtga	gcmccngnca	cctgccacag	cctcagaggc	ccgcacaatg	1080

50

	accggaggag	gggcccgtgt	ggtctggccc	cctccctgtc	caggccccgc	aggaggcaga	1140
	tgacgtccca	ggcatccctc	tkccctggg	ctctcaggga	ccccctgggt	cggcttctgt	1200
55	ccctgtcaca	cccccaaccc	cagggagggg	ctgtcatagt	cccagaggat	aagcaatacc	1260
	tatttctgac	tgagtctccc	agcccagacc	cagggaccct	nggccccaa	ctcagctcta	1320
	agaaccgccc	caacccctcc	agctccaaat	ctgagcctcc	accacataga	ctgaaactcc	1380
	cctggcccca	gcccctctcc	gcctggcctg	gcctgggaca	cctcctctct	gccaggaggc	1440
	aataaaagcc	agcgcgggga	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1500
60	aaaaaan						1507

<210> 13

<211> 661

65 <212> ADN

<213> *Homo sapiens*

# ES 2 305 087 T3

<220>

<221> CDS

<222> (79)...(376)

5

<400> 13

```

ggcgcggtgc agggctctta agaacgaacg gcttgggcgc ggactggtat ccgggggactg      60
tgacttgcag ggtccgcc atg gag cca gag cag atg ctg gag gga caa acg      111
10          Met Glu Pro Glu Gln Met Leu Glu Gly Gln Thr
              1              5              10

cag gtt gca gaa aat cct cac tct gag tac ggt ctc aca gac aac gtt      159
Gln Val Ala Glu Asn Pro His Ser Glu Tyr Gly Leu Thr Asp Asn Val
15          15          20          25

gag aga ata gta gaa aat gag aag att aat gca gaa aag tca tca aag      207
Glu Arg Ile Val Glu Asn Glu Lys Ile Asn Ala Glu Lys Ser Ser Lys
20          30          35          40

cag aag gta gat ctc cag tct ttg cca act cgt gcc tac ctg gat cag      255
Gln Lys Val Asp Leu Gln Ser Leu Pro Thr Arg Ala Tyr Leu Asp Gln
25          45          50          55

aca gtt gtg cct atc tta tta cag gga ctt gct gtg ctt gca aag gaa      303
Thr Val Val Pro Ile Leu Leu Gln Gly Leu Ala Val Leu Ala Lys Glu
30          60          65          70          75

aga cca cca aat ccc att gaa ttt cta gca tct tat ctt tta aaa aac      351
Arg Pro Pro Asn Pro Ile Glu Phe Leu Ala Ser Tyr Leu Leu Lys Asn
35          80          85          90

aag gca cag ttt gaa gat cga aac t gacttaatgg gaagaacaga      396
Lys Ala Gln Phe Glu Asp Arg Asn
35          95

aaaatttagt tgctactgta gatttacatg attaagaggc agctttaatt gccatgatca      456
ttccctcttt ttggatgtat aagaaccttc cggacaacag aacctatttc tggaattgca      516
gaagataaca tatttccctt attttgattt aatcaccata aaccatacct atttaatgag      576
40 tgtattctgt gcaatttttt tctcagattg tctttaactt tgttttttaa atgaccttca      636
aaataaactg tcaaaacacc attat      661

```

<210> 14

45

<211> 99

<212> PRT

<213> *Homo sapiens*

50

<400> 14

```

Met Glu Pro Glu Gln Met Leu Glu Gly Gln Thr Gln Val Ala Glu Asn
1              5              10              15

55 Pro His Ser Glu Tyr Gly Leu Thr Asp Asn Val Glu Arg Ile Val Glu
          20          25          30
Asn Glu Lys Ile Asn Ala Glu Lys Ser Ser Lys Gln Lys Val Asp Leu
          35          40          45
60 Gln Ser Leu Pro Thr Arg Ala Tyr Leu Asp Gln Thr Val Val Pro Ile
          50          55          60
Leu Leu Gln Gly Leu Ala Val Leu Ala Lys Glu Arg Pro Pro Asn Pro
65          65          70          75          80
Ile Glu Phe Leu Ala Ser Tyr Leu Leu Lys Asn Lys Ala Gln Phe Glu
          85          90          95
65 Asp Arg Asn

```

# ES 2 305 087 T3

<210> 15  
<211> 1507  
<212> ADN

5 <213> *Homo sapiens*

<220>  
<221> misc\_feature

10 <222> 1047, 1301  
<223> n = A, T, C o G

15 <400> 15

	<b>ggaacgcaga</b>	<b>gcgagcgctg</b>	<b>gagagcggag</b>	<b>cgaagctgga</b>	<b>taacagggga</b>	<b>ccgatgatgt</b>	<b>60</b>
	<b>ggcgaccatc</b>	<b>agttctgctg</b>	<b>cttctgttgc</b>	<b>tactgaggca</b>	<b>cggggcccag</b>	<b>gggaagccat</b>	<b>120</b>
	<b>ccccagacgc</b>	<b>aggccctcat</b>	<b>ggccagggga</b>	<b>gggtgcacca</b>	<b>ggcgggcccc</b>	<b>ctgagcgacg</b>	<b>180</b>
20	<b>ctccccatga</b>	<b>tgacgcccac</b>	<b>gggaacttcc</b>	<b>agtacgacca</b>	<b>tgaggctttc</b>	<b>ctgggacggg</b>	<b>240</b>
	<b>aaagtggccaa</b>	<b>ggaattcgac</b>	<b>caactcacc</b>	<b>cagaggaaag</b>	<b>ccaggcccgt</b>	<b>ctggggcgga</b>	<b>300</b>
	<b>tcgtggaccg</b>	<b>catggaccgc</b>	<b>gcgggggacg</b>	<b>gcgacggctg</b>	<b>ggtgtcgctg</b>	<b>gccgagcttc</b>	<b>360</b>
	<b>gcgcgtggat</b>	<b>cgcgcacacg</b>	<b>cagcagcggc</b>	<b>acatacggga</b>	<b>ctcggtgagc</b>	<b>gcggcctggg</b>	<b>420</b>
	<b>acacgtacga</b>	<b>cacggaccgc</b>	<b>gacgggcgtg</b>	<b>tgggttggga</b>	<b>ggagctgcgc</b>	<b>aacgccacct</b>	<b>480</b>
25	<b>atggccacta</b>	<b>cgcgcccggg</b>	<b>gaagaatttc</b>	<b>atgacgtgga</b>	<b>ggatgcagag</b>	<b>acctacaaaa</b>	<b>540</b>
	<b>agatgctggc</b>	<b>tcgggacgag</b>	<b>cggcgtttcc</b>	<b>gggtggccga</b>	<b>ccaggatggg</b>	<b>gactcgatgg</b>	<b>600</b>
	<b>ccactcgaga</b>	<b>ggagctgaca</b>	<b>gccttcctgc</b>	<b>accccgagga</b>	<b>gttccctcac</b>	<b>atgcgggaca</b>	<b>660</b>
	<b>tcgtgattgc</b>	<b>tgaaacctcg</b>	<b>gaggacctgg</b>	<b>acagaaacaa</b>	<b>agatggctat</b>	<b>gtccaggtgg</b>	<b>720</b>
	<b>aggagtacat</b>	<b>cgcggatctg</b>	<b>tactcagccg</b>	<b>agcctgggga</b>	<b>ggaggagccg</b>	<b>gcgtgggtgc</b>	<b>780</b>
30	<b>agacggagag</b>	<b>gcagcagttc</b>	<b>cgggacttcc</b>	<b>gggatctgaa</b>	<b>caaggatggg</b>	<b>cacctggatg</b>	<b>840</b>
	<b>ggagtgaggt</b>	<b>gggccactgg</b>	<b>gtgctgcccc</b>	<b>ctgccagga</b>	<b>ccagcccctg</b>	<b>gtggaagcca</b>	<b>900</b>
	<b>accacctgct</b>	<b>gcacgaragc</b>	<b>gacacggaca</b>	<b>aggaygggcg</b>	<b>gctgagcaaa</b>	<b>gcgsaaatcc</b>	<b>960</b>
	<b>tgggtaattg</b>	<b>gaacatgttt</b>	<b>gtgggcagtc</b>	<b>aggccaccaa</b>	<b>ctatggygag</b>	<b>gacctgacct</b>	<b>1020</b>
	<b>ggcaccacga</b>	<b>tgagctgtga</b>	<b>gcmccngca</b>	<b>cctgccacag</b>	<b>cctcagaggc</b>	<b>ccgcacaatg</b>	<b>1080</b>
35	<b>accggaggag</b>	<b>gggcccgtgt</b>	<b>ggtctggccc</b>	<b>cctccctgtc</b>	<b>caggccccgc</b>	<b>aggaggcaga</b>	<b>1140</b>
	<b>tgcagtccca</b>	<b>ggcatcctcc</b>	<b>tkccctggg</b>	<b>ctctcaggga</b>	<b>ccccctgggt</b>	<b>cggcttctgt</b>	<b>1200</b>
	<b>ccctgtcaca</b>	<b>cccccaacct</b>	<b>cagggagggg</b>	<b>ctgtcatagt</b>	<b>cccagaggat</b>	<b>aagcaatacc</b>	<b>1260</b>
	<b>tatttctgac</b>	<b>tgagtctccc</b>	<b>agcccagacc</b>	<b>cagggaccct</b>	<b>nggccccaa</b>	<b>ctcagctcta</b>	<b>1320</b>
	<b>agaaccgccc</b>	<b>caaccctctc</b>	<b>agctccaaat</b>	<b>ctgagcctcc</b>	<b>accacataga</b>	<b>ctgaaactcc</b>	<b>1380</b>
40	<b>cctggcccca</b>	<b>gcctctcct</b>	<b>gcctggcctg</b>	<b>gcctgggaca</b>	<b>cctcctctct</b>	<b>gccaggaggc</b>	<b>1440</b>
	<b>aataaaagcc</b>	<b>agcgccggga</b>	<b>aaaaaaaaaa</b>	<b>aaaaaaaaaa</b>	<b>aaaaaaaaaa</b>	<b>aaaaaaaaaa</b>	<b>1500</b>
	<b>aaaaaan</b>						<b>1507</b>

45 <210> 16  
<211> 716  
<212> ADN  
<213> *Homo sapiens*

50 <220>  
<221> CDS  
55 <222> (16)...(538)

60

65

# ES 2 305 087 T3

<400> 16

5	ctcgcgacgaa ggcgag atg gcg gac gtg cta gat ctt cac gag gct ggg ggc Met Ala Asp Val Leu Asp Leu His Glu Ala Gly Gly	51
	1 5 10	
10	gaa gat ttc gcc atg gat gag gat ggg gac gag agc att cac aaa ctg Glu Asp Phe Ala Met Asp Glu Asp Gly Asp Glu Ser Ile His Lys Leu	99
	15 20 25	
15	aaa gaa aaa gcg aag aaa cgg aag ggt cgc ggc ttt ggc tcc gaa gag Lys Glu Lys Ala Lys Lys Arg Lys Gly Arg Gly Phe Gly Ser Glu Glu	147
	30 35 40	
20	ggg tcc cga gcg cgg atg cgt gag gat tat gac agc gtg gag cag gat Gly Ser Arg Ala Arg Met Arg Glu Asp Tyr Asp Ser Val Glu Gln Asp	195
	45 50 55 60	
25	ggc gat gaa ccc gga cca caa cgc tct gtt gaa ggc tgg att ctc ttt Gly Asp Glu Pro Gly Pro Gln Arg Ser Val Glu Gly Trp Ile Leu Phe	243
	65 70 75	
30	gta act gga gtc cat gag gaa gcc acc gaa gaa gac ata cac gac aaa Val Thr Gly Val His Glu Glu Ala Thr Glu Glu Asp Ile His Asp Lys	291
	80 85 90	
35	ttc gca gaa tat ggg gaa att aaa aac att cat ctc aac ctc gac agg Phe Ala Glu Tyr Gly Glu Ile Lys Asn Ile His Leu Asn Leu Asp Arg	339
	95 100 105	
40	cga aca gga tat ctg aag ggg tat act cta gtt gaa tat gaa aca tac Arg Thr Gly Tyr Leu Lys Gly Tyr Thr Leu Val Glu Tyr Glu Thr Tyr	387
	110 115 120	
45	aag gaa gcc cag gct gct atg gag gga ctc aat ggc cag gat ttg atg Lys Glu Ala Gln Ala Ala Met Glu Gly Leu Asn Gly Gln Asp Leu Met	435
	125 130 135 140	
50	gga cag ccc atc agc gtt gac tgg tgt ttt gtt cgg ggt cca cca aaa Gly Gln Pro Ile Ser Val Asp Trp Cys Phe Val Arg Gly Pro Pro Lys	483
	145 150 155	
55	ggc aag agg aga ggt ggc cga aga cgc agc aga agt cca gac cgg aga Gly Lys Arg Arg Gly Gly Arg Arg Arg Ser Arg Ser Pro Asp Arg Arg	531
	160 165 170	
60	cgt cgc t gacaggctct ctgttggtcca ggtgttctct tcaagattcc atttgacat Arg Arg	588
65	gcagccttgg acaaatagga ctgggggtgga acttgctgtg tttatatatta atctcttacc gtatatgcgt agtattttgag ttgcgaataa atgttccatt ttgttttcta caaaaaaaaaa aaaaaaaaa	648 708 716

<210> 17

60 <211> 174

<212> PRT

<213> *Homo sapiens*

65

# ES 2 305 087 T3

<400> 17

5	Met	Ala	Asp	Val	Leu	Asp	Leu	His	Glu	Ala	Gly	Gly	Glu	Asp	Phe	Ala
	1				5					10					15	
	Met	Asp	Glu	Asp	Gly	Asp	Glu	Ser	Ile	His	Lys	Leu	Lys	Glu	Lys	Ala
				20					25					30		
	Lys	Lys	Arg	Lys	Gly	Arg	Gly	Phe	Gly	Ser	Glu	Glu	Gly	Ser	Arg	Ala
			35					40					45			
10	Arg	Met	Arg	Glu	Asp	Tyr	Asp	Ser	Val	Glu	Gln	Asp	Gly	Asp	Glu	Pro
		50					55					60				
15																
	Gly	Pro	Gln	Arg	Ser	Val	Glu	Gly	Trp	Ile	Leu	Phe	Val	Thr	Gly	Val
	65					70					75					80
	His	Glu	Glu	Ala	Thr	Glu	Glu	Asp	Ile	His	Asp	Lys	Phe	Ala	Glu	Tyr
				85					90						95	
20	Gly	Glu	Ile	Lys	Asn	Ile	His	Leu	Asn	Leu	Asp	Arg	Arg	Thr	Gly	Tyr
				100					105					110		
	Leu	Lys	Gly	Tyr	Thr	Leu	Val	Glu	Tyr	Glu	Thr	Tyr	Lys	Glu	Ala	Gln
			115					120					125			
25	Ala	Ala	Met	Glu	Gly	Leu	Asn	Gly	Gln	Asp	Leu	Met	Gly	Gln	Pro	Ile
		130					135					140				
	Ser	Val	Asp	Trp	Cys	Phe	Val	Arg	Gly	Pro	Pro	Lys	Gly	Lys	Arg	Arg
	145					150					155					160
	Gly	Gly	Arg	Arg	Arg	Ser	Arg	Ser	Pro	Asp	Arg	Arg	Arg	Arg		
30					165					170						

<210> 18

<211> 763

<212> ADN

<213> *Homo sapiens*

<220>

<221> CDS

<222> (2)...(551)

# ES 2 305 087 T3

<400> 18

```

5      c atg gcc aag ccg tgt ggg gtg cgc ctg agc ggg gaa gcc cgc aaa cag      49
      Met Ala Lys Pro Cys Gly Val Arg Leu Ser Gly Glu Ala Arg Lys Gln
      1          5          10          15

10     gtg gag gtc ttc aga cag aat ctt ttc cag gag gct gag gaa ttc ctc      97
      Val Glu Val Phe Arg Gln Asn Leu Phe Gln Glu Ala Glu Glu Phe Leu
      20          25          30

15     tac aga ttc ttg cca cag aaa atc ata tac ctg aat cag ctc ttg caa      145
      Tyr Arg Phe Leu Pro Gln Lys Ile Ile Tyr Leu Asn Gln Leu Leu Gln
      35          40          45

20     gag gac tcc ctc aat gtg gct gac ttg act tcc ctc cgg gcc cca ctg      193
      Glu Asp Ser Leu Asn Val Ala Asp Leu Thr Ser Leu Arg Ala Pro Leu
      50          55          60

25     gac atc ccc atc cca gac cct cca ccc aag gat gat gag atg gaa aca      241
      Asp Ile Pro Ile Pro Asp Pro Pro Pro Lys Asp Asp Glu Met Glu Thr
      65          70          75          80

30     gat aag cag gag aag aaa gaa gtc cct aag tgt gga ttt ctc cct ggg      289
      Asp Lys Gln Glu Lys Lys Glu Val Pro Lys Cys Gly Phe Leu Pro Gly
      85          90          95

35     aat gag aaa gtc ctg tcc ctg ctt gcc ctg gtt aag cca gaa gtc tgg      337
      Asn Glu Lys Val Leu Ser Leu Leu Ala Leu Val Lys Pro Glu Val Trp
      100          105          110

40     act ctc aaa gag aaa tgc att ctg gtg att aca tgg atc caa cac ctg      385
      Thr Leu Lys Glu Lys Cys Ile Leu Val Ile Thr Trp Ile Gln His Leu
      115          120          125

45     atc ccc aag att gaa gat gga aat gat ttt ggg gta gca atc cag gag      433
      Ile Pro Lys Ile Glu Asp Gly Asn Asp Phe Gly Val Ala Ile Gln Glu
      130          135          140

50     aag gtg ctg gag agg gtg aat gcc gtc aag acc aaa gtg aag ctt tcc      481
      Lys Val Leu Glu Arg Val Asn Ala Val Lys Thr Lys Val Lys Leu Ser
      145          150          155          160

55     aga caa cca ttt cca agt act tct cag aac gtg ggg atg ctg tgg cca      529
      Arg Gln Pro Phe Pro Ser Thr Ser Gln Asn Val Gly Met Leu Trp Pro
      165          170          175

60     agg cct cca agg aga ctc atg t aatggattac cgggccttgg tgcattgagcg      581
      Arg Pro Pro Arg Arg Leu Met
      180

65     agatgaggca gcctatgggg agctcagggc catggtgctg gacctgaggg ccttctatgc      641
      tgagctttat catatcatca gcagcaacct ggagaaaatt gtcaacccaa aggggtgaaga      701
      aaagccatct atgtactgaa cccgggacta gaaggaaaat aaatgatcta tatgttgtgt      761
      gg                                                    763

60 <210> 19
    <211> 183
    <212> PRT
    <213> Homo sapiens

```

# ES 2 305 087 T3

<400> 19

5	Met	Ala	Lys	Pro	Cys	Gly	Val	Arg	Leu	Ser	Gly	Glu	Ala	Arg	Lys	Gln
	1				5					10					15	
	Val	Glu	Val	Phe	Arg	Gln	Asn	Leu	Phe	Gln	Glu	Ala	Glu	Glu	Phe	Leu
				20					25					30		
	Tyr	Arg	Phe	Leu	Pro	Gln	Lys	Ile	Ile	Tyr	Leu	Asn	Gln	Leu	Leu	Gln
			35					40					45			
10	Glu	Asp	Ser	Leu	Asn	Val	Ala	Asp	Leu	Thr	Ser	Leu	Arg	Ala	Pro	Leu
		50					55					60				
	Asp	Ile	Pro	Ile	Pro	Asp	Pro	Pro	Pro	Lys	Asp	Asp	Glu	Met	Glu	Thr
	65					70					75					80
	Asp	Lys	Gln	Glu	Lys	Lys	Glu	Val	Pro	Lys	Cys	Gly	Phe	Leu	Pro	Gly
15					85					90					95	
	Asn	Glu	Lys	Val	Leu	Ser	Leu	Leu	Ala	Leu	Val	Lys	Pro	Glu	Val	Trp
				100					105					110		
	Thr	Leu	Lys	Glu	Lys	Cys	Ile	Leu	Val	Ile	Thr	Trp	Ile	Gln	His	Leu
			115					120					125			
20	Ile	Pro	Lys	Ile	Glu	Asp	Gly	Asn	Asp	Phe	Gly	Val	Ala	Ile	Gln	Glu
		130					135					140				
	Lys	Val	Leu	Glu	Arg	Val	Asn	Ala	Val	Lys	Thr	Lys	Val	Lys	Leu	Ser
	145					150					155					160
	Arg	Gln	Pro	Phe	Pro	Ser	Thr	Ser	Gln	Asn	Val	Gly	Met	Leu	Trp	Pro
25					165					170					175	
	Arg	Pro	Pro	Arg	Arg	Leu	Met									
				180												

30 <210> 20  
 <211> 790  
 <212> ADN  
 <213> *Homo sapiens*  
 35 <220>  
 <221> CDS  
 <222> (240)...(585)  
 40 <400> 20

45	acg	tttt	taca	gtct	ttta	aatt	aag	ca	cata	aa	aact	gt	tact	aa	ttta	atat	at	ttct	cc	atga	60
	act	tttt	tgtg	aaat	tc	agat	cgc	agt	gtgt	catt	taca	aaa	tct	ttt	gtct	ttct	tct	gggt	120		
	cat	ctac	acc	tttt	gc	acag	ttct	tga	aga	caac	gt	catc	atcc	ca	cctt	cttt	ta	actt	180		
	tga	agt	tggc	ctga	gg	ctgg	gat	ggg	ccag	tgag	atta	aag	gag	agg	ggtt	ccgt	tc	caga	239		
	atg	ttt	tcc	ata	cga	atc	ctc	tct	tct	tca	gct	ttt	tgt	tct	tgt	tcc			287		
50	Met	Phe	Ser	Ile	Arg	Ile	Leu	Ser	Ser	Ser	Ala	Phe	Cys	Ser	Cys	Ser					
	1				5						10				15						

55

60

65

# ES 2 305 087 T3

ttc ctg gcc tgc tct tca gct ctt tct ttt tta att ttt tcc agt tct 335  
 Phe Leu Ala Cys Ser Ser Ala Leu Ser Phe Leu Ile Phe Ser Ser Ser  
                     20                    25                    30  
 5 gca aga aga gct gca gta tca tca tca tca ctt tct tct tca aaa tct 383  
 Ala Arg Arg Ala Ala Val Ser Ser Ser Ser Leu Ser Ser Ser Lys Ser  
                     35                    40                    45  
 10 tca tct tcc tca tct gtt aga ggg tca tct gca tca agg ttg gcg gca 431  
 Ser Ser Ser Ser Ser Val Arg Gly Ser Ser Ala Ser Arg Leu Ala Ala  
                     50                    55                    60  
 15 gga atc tgg tct aac cgt ggc ttt ttt gac act gaa gag gag gtt gta 479  
 Gly Ile Trp Ser Asn Arg Gly Phe Phe Asp Thr Glu Glu Glu Val Val  
                     65                    70                    75                    80  
 20 tgt tct cgg gtt gga cga tcc cta ttt ttc tct ctt gca gca gct ctc 527  
 Cys Ser Arg Val Gly Arg Ser Leu Phe Phe Ser Leu Ala Ala Ala Leu  
                     85                    90                    95  
 25 tct ctt tct tcc aac tct ctc ctg aag tca cgg tta cga acc tct tca 575  
 Ser Leu Ser Ser Asn Ser Leu Leu Lys Ser Arg Leu Arg Thr Ser Ser  
                     100                    105                    110  
 30 ggg gca tcc t gagtagtctg totgtatttt atctttgtat gagagggtag 625  
 Gly Ala Ser  
                     115  
 35 gtctctgctt gaatactgct ttgaaagttg gctcaaatca ccttctcctt ttccccctcc 685  
 acctctggca ggttcaaagg ttggcctggc tgctgttgtc atcttttatg actggccgag 745  
 gtccgatgca gcaggctccg aagatcatatc agacgccatt accac 790

35 <210> 21  
 <211> 115  
 <212> PRT  
 40 <213> *Homo sapiens*  
 <400> 21

45 Met Phe Ser Ile Arg Ile Leu Ser Ser Ser Ala Phe Cys Ser Cys Ser  
     1                    5                    10                    15  
 Phe Leu Ala Cys Ser Ser Ala Leu Ser Phe Leu Ile Phe Ser Ser Ser  
                     20                    25                    30  
 50 Ala Arg Arg Ala Ala Val Ser Ser Ser Leu Ser Ser Ser Lys Ser  
                     35                    40                    45  
 Ser Ser Ser Ser Ser Val Arg Gly Ser Ser Ala Ser Arg Leu Ala Ala  
                     50                    55                    60  
 Gly Ile Trp Ser Asn Arg Gly Phe Phe Asp Thr Glu Glu Glu Val Val  
                     65                    70                    75                    80  
 55 Cys Ser Arg Val Gly Arg Ser Leu Phe Phe Ser Leu Ala Ala Ala Leu  
                     85                    90                    95  
 Ser Leu Ser Ser Asn Ser Leu Leu Lys Ser Arg Leu Arg Thr Ser Ser  
                     100                    105                    110  
 60 Gly Ala Ser  
                     115

<210> 22  
 <211> 1939  
 65 <212> ADN  
 <213> *Homo sapiens*



# ES 2 305 087 T3

<220>

<221> CDS

<222> (53)...(1700)

5

<400> 22

	gtggcgccag cggaggcagg ttgctgtgtt tgtgcttctt tctacagcca at atg aaa	58
	Met Lys	
10	1	
	agg cct aag tta aag aaa gca agt aaa cgc atg acc tgc cat aag cgg	106
	Arg Pro Lys Leu Lys Lys Ala Ser Lys Arg Met Thr Cys His Lys Arg	
	5 10 15	
15	tat aaa atc caa aaa aag gtt cga gaa cat cat cga aaa tta aga aag	154
	Tyr Lys Ile Gln Lys Lys Val Arg Glu His His Arg Lys Leu Arg Lys	
	20 25 30	
20	gag gct aaa aag cag ggt cac aag aag cct agg aaa gac cca gga gtt	202
	Glu Ala Lys Lys Gln Gly His Lys Lys Pro Arg Lys Asp Pro Gly Val	
	35 40 45 50	
25	cca aac agt gct ccc ttt aag gag gct ctt ctt agg gaa gct gag cta	250
	Pro Asn Ser Ala Pro Phe Lys Glu Ala Leu Leu Arg Glu Ala Glu Leu	
	55 60 65	
30	agg aaa cag agg ctt gaa gaa cta aaa cag cag cag aaa ctt gac agg	298
	Arg Lys Gln Arg Leu Glu Glu Leu Lys Gln Gln Gln Lys Leu Asp Arg	
	70 75 80	
35	cag aag gaa cta gaa aag aaa aga aaa ctt gaa act aat cct gat att	346
	Gln Lys Glu Leu Glu Lys Lys Arg Lys Leu Glu Thr Asn Pro Asp Ile	
	85 90 95	
40	aag cca tca aat gtg gaa cct atg gaa aag gag ttt ggg ctt tgc aaa	394
	Lys Pro Ser Asn Val Glu Pro Met Glu Lys Glu Phe Gly Leu Cys Lys	
	100 105 110	
45	act gag aac aaa gcc aag tcg ggc aaa cag aat tca aag aag ctg tac	442
	Thr Glu Asn Lys Ala Lys Ser Gly Lys Gln Asn Ser Lys Lys Leu Tyr	
	115 120 125 130	
50	tgc caa gaa ctt aaa aag gtg att gaa gcc tcc gat gtt gtc cta gag	490
	Cys Gln Glu Leu Lys Lys Val Ile Glu Ala Ser Asp Val Val Leu Glu	
	135 140 145	
55	gtg ttg gat gcc aga gat cct ctt ggt tgc aga tgt cct cag gta gaa	538
	Val Leu Asp Ala Arg Asp Pro Leu Gly Cys Arg Cys Pro Gln Val Glu	
	150 155 160	
60	gag gcc att gtc cag agt gga cag aaa aag ctg gta ctt ata tta aat	586
	Glu Ala Ile Val Gln Ser Gly Gln Lys Lys Leu Val Leu Ile Leu Asn	
	165 170 175	
65	aaa tca gat ctg gta cca aag gag aat ttg gag agc tgg cta aat tat	634
	Lys Ser Asp Leu Val Pro Lys Glu Asn Leu Glu Ser Trp Leu Asn Tyr	
	180 185 190	
70	ttg aag aaa gaa ttg cca aca gtg gtg ttc aga gcc tca aca aaa cca	682
	Leu Lys Lys Glu Leu Pro Thr Val Val Phe Arg Ala Ser Thr Lys Pro	
	195 200 205 210	
75	aag gat aaa ggg aag ata acc aag cgt gtg aag gca aag aag aat gct	730
	Lys Asp Lys Gly Lys Ile Thr Lys Arg Val Lys Ala Lys Lys Asn Ala	
	215 220 225	

# ES 2 305 087 T3

		gct cca ttc aga agt gaa gtc tgc ttt ggg aaa gag ggc ctt tgg aaa	778
		Ala Pro Phe Arg Ser Glu Val Cys Phe Gly Lys Glu Gly Leu Trp Lys	
		230 235 240	
5		ctt ctt gga ggt ttt cag gaa act tgc agc aaa gcc att cgg gtt gga	826
		Leu Leu Gly Gly Phe Gln Glu Thr Cys Ser Lys Ala Ile Arg Val Gly	
		245 250 255	
10		gta att ggt ttc cca aat gtg ggg aaa agc agc att atc aat agc tta	874
		Val Ile Gly Phe Pro Asn Val Gly Lys Ser Ser Ile Ile Asn Ser Leu	
		260 265 270	
15		aaa caa gaa cag atg tgt aat gtt ggt gta tcc atg ggg ctt aca agg	922
		Lys Gln Glu Gln Met Cys Asn Val Gly Val Ser Met Gly Leu Thr Arg	
		275 280 285 290	
20		agc atg caa gtt gtc ccc ttg gac aaa cag atc aca atc ata gat agt	970
		Ser Met Gln Val Val Pro Leu Asp Lys Gln Ile Thr Ile Ile Asp Ser	
		295 300 305	
25		ccg agc ttc atc gta tct cca ctt aat tcc tcc tct gcg ctt gct ctg	1018
		Pro Ser Phe Ile Val Ser Pro Leu Asn Ser Ser Ser Ala Leu Ala Leu	
		310 315 320	
30		cga agt cca gca agt att gaa gta gta aaa ccg atg gag gct gcc agt	1066
		Arg Ser Pro Ala Ser Ile Glu Val Val Lys Pro Met Glu Ala Ala Ser	
		325 330 335	
35		gcc atc ctt tcc cag gct gat gct cga cag gta gta ctg aaa tat act	1114
		Ala Ile Leu Ser Gln Ala Asp Ala Arg Gln Val Val Leu Lys Tyr Thr	
		340 345 350	
40		gtc cca ggc tac agg aat tct ctg gaa ttt ttt act atg ctt gct cag	1162
		Val Pro Gly Tyr Arg Asn Ser Leu Glu Phe Phe Thr Met Leu Ala Gln	
		355 360 365 370	
45		aga aga ggt atg cac caa aaa ggt gga atc cca aat gtt gaa ggt gct	1210
		Arg Arg Gly Met His Gln Lys Gly Gly Ile Pro Asn Val Glu Gly Ala	
		375 380 385	
50		gcc aaa ctg ctg tgg tct gag tgg aca ggt gcc tca tta gct tac tat	1258
		Ala Lys Leu Leu Trp Ser Glu Trp Thr Gly Ala Ser Leu Ala Tyr Tyr	
		390 395 400	
55		tgc cat ccc cct aca tct tgg act cct cct cca tat ttt aat gag agt	1306
		Cys His Pro Pro Thr Ser Trp Thr Pro Pro Pro Tyr Phe Asn Glu Ser	
		405 410 415	
60		att gtg gta gac atg aaa agc ggc ttc aat ctg gaa gaa ctg gaa aag	1354
		Ile Val Val Asp Met Lys Ser Gly Phe Asn Leu Glu Glu Leu Glu Lys	
		420 425 430	
65		aac aat gca cag agc ata aga gcc atc aag ggc cct cat ttg gcc aat	1402
		Asn Asn Ala Gln Ser Ile Arg Ala Ile Lys Gly Pro His Leu Ala Asn	
		435 440 445 450	
70		agc atc ctt ttc cag tct tcc ggt ctg aca aat gga ata ata gaa gaa	1450
		Ser Ile Leu Phe Gln Ser Ser Gly Leu Thr Asn Gly Ile Ile Glu Glu	
		455 460 465	
75		aag gac ata cat gaa gaa ttg cca aaa cgg aaa gaa agg aag cag gag	1498
		Lys Asp Ile His Glu Glu Leu Pro Lys Arg Lys Glu Arg Lys Gln Glu	
		470 475 480	

# ES 2 305 087 T3

	gag agg gag gat gac aaa gac agt gac cag gaa act gtt gat gaa gaa	1546
	Glu Arg Glu Asp Asp Lys Asp Ser Asp Gln Glu Thr Val Asp Glu Glu	
	485 490 495	
5	ggt gat gaa aac agc tca ggc atg ttt gct gca gaa gag aca ggg gag	1594
	Val Asp Glu Asn Ser Ser Gly Met Phe Ala Ala Glu Glu Thr Gly Glu	
	500 505 510	
10	gca ctg tct gag gag act aca gca ggt gaa cag tct aca agg tct ttt	1642
	Ala Leu Ser Glu Glu Thr Thr Ala Gly Glu Gln Ser Thr Arg Ser Phe	
	515 520 525 530	
15	atc ttg gat aaa atc att gaa gag gat gat gct tat gac ttc agt aca	1690
	Ile Leu Asp Lys Ile Ile Glu Glu Asp Asp Ala Tyr Asp Phe Ser Thr	
	535 540 545	
	gat tat gtg t aacagaacaa tggcttttta tgattttttt ttttaacatt	1740
	Asp Tyr Val	
20		
	ttaagcagac tgctaaactg ttctctgtat aagttatggt atgcatgagc tgtgtaaatt	1800
	ttgtgaatat gtattatatt aaaaccaggc aacttggaat ccctaaattc tgtaaaaaga	1860
	caattcatct cattgtgagt ggaagtagtt atctggaata aaaaaagaag atacctattg	1920
25	aaaaaaaaa aaaaaaaaaa	1939
	<210> 23	
	<211> 549	
30	<212> PRT	
	<213> <i>Homo sapiens</i>	
	<400> 23	
35	Met Lys Arg Pro Lys Leu Lys Lys Ala Ser Lys Arg Met Thr Cys His	
	1 5 10 15	
	Lys Arg Tyr Lys Ile Gln Lys Lys Val Arg Glu His His Arg Lys Leu	
	20 25 30	
40	Arg Lys Glu Ala Lys Lys Gln Gly His Lys Lys Pro Arg Lys Asp Pro	
	35 40 45	
	Gly Val Pro Asn Ser Ala Pro Phe Lys Glu Ala Leu Leu Arg Glu Ala	
	50 55 60	
	Glu Leu Arg Lys Gln Arg Leu Glu Glu Leu Lys Gln Gln Gln Lys Leu	
45	65 70 75 80	
	Asp Arg Gln Lys Glu Leu Glu Lys Lys Arg Lys Leu Glu Thr Asn Pro	
	85 90 95	
	Asp Ile Lys Pro Ser Asn Val Glu Pro Met Glu Lys Glu Phe Gly Leu	
	100 105 110	
50	Cys Lys Thr Glu Asn Lys Ala Lys Ser Gly Lys Gln Asn Ser Lys Lys	
	115 120 125	
	Leu Tyr Cys Gln Glu Leu Lys Lys Val Ile Glu Ala Ser Asp Val Val	
	130 135 140	
	Leu Glu Val Leu Asp Ala Arg Asp Pro Leu Gly Cys Arg Cys Pro Gln	
55	145 150 155 160	
	Val Glu Glu Ala Ile Val Gln Ser Gly Gln Lys Lys Leu Val Leu Ile	
	165 170 175	
	Leu Asn Lys Ser Asp Leu Val Pro Lys Glu Asn Leu Glu Ser Trp Leu	
	180 185 190	
60	Asn Tyr Leu Lys Lys Glu Leu Pro Thr Val Val Phe Arg Ala Ser Thr	
	195 200 205	
	Lys Pro Lys Asp Lys Gly Lys Ile Thr Lys Arg Val Lys Ala Lys Lys	
	210 215 220	
	Asn Ala Ala Pro Phe Arg Ser Glu Val Cys Phe Gly Lys Glu Gly Leu	
65	225 230 235 240	
	Trp Lys Leu Leu Gly Phe Gln Glu Thr Cys Ser Lys Ala Ile Arg	
	245 250 255	

# ES 2 305 087 T3

	Val	Gly	Val	Ile	Gly	Phe	Pro	Asn	Val	Gly	Lys	Ser	Ser	Ile	Ile	Asn
				260					265					270		
	Ser	Leu	Lys	Gln	Glu	Gln	Met	Cys	Asn	Val	Gly	Val	Ser	Met	Gly	Leu
5			275					280					285			
	Thr	Arg	Ser	Met	Gln	Val	Val	Pro	Leu	Asp	Lys	Gln	Ile	Thr	Ile	Ile
		290					295					300				
	Asp	Ser	Pro	Ser	Phe	Ile	Val	Ser	Pro	Leu	Asn	Ser	Ser	Ser	Ala	Leu
	305					310					315					320
10	Ala	Leu	Arg	Ser	Pro	Ala	Ser	Ile	Glu	Val	Val	Lys	Pro	Met	Glu	Ala
					325					330					335	
	Ala	Ser	Ala	Ile	Leu	Ser	Gln	Ala	Asp	Ala	Arg	Gln	Val	Val	Leu	Lys
				340					345					350		
	Tyr	Thr	Val	Pro	Gly	Tyr	Arg	Asn	Ser	Leu	Glu	Phe	Phe	Thr	Met	Leu
15			355					360					365			
	Ala	Gln	Arg	Arg	Gly	Met	His	Gln	Lys	Gly	Gly	Ile	Pro	Asn	Val	Glu
		370					375					380				
	Gly	Ala	Ala	Lys	Leu	Leu	Trp	Ser	Glu	Trp	Thr	Gly	Ala	Ser	Leu	Ala
	385					390					395					400
20	Tyr	Tyr	Cys	His	Pro	Thr	Ser	Trp	Thr	Pro	Pro	Pro	Tyr	Phe	Asn	
				405					410					415		
	Glu	Ser	Ile	Val	Val	Asp	Met	Lys	Ser	Gly	Phe	Asn	Leu	Glu	Glu	Leu
				420				425					430			
	Glu	Lys	Asn	Asn	Ala	Gln	Ser	Ile	Arg	Ala	Ile	Lys	Gly	Pro	His	Leu
25			435					440					445			
	Ala	Asn	Ser	Ile	Leu	Phe	Gln	Ser	Ser	Gly	Leu	Thr	Asn	Gly	Ile	Ile
		450					455				460					
	Glu	Glu	Lys	Asp	Ile	His	Glu	Glu	Leu	Pro	Lys	Arg	Lys	Glu	Arg	Lys
	465					470					475					480
30	Gln	Glu	Glu	Arg	Glu	Asp	Asp	Lys	Asp	Ser	Asp	Gln	Glu	Thr	Val	Asp
					485				490					495		
	Glu	Glu	Val	Asp	Glu	Asn	Ser	Ser	Gly	Met	Phe	Ala	Ala	Glu	Glu	Thr
				500					505					510		
	Gly	Glu	Ala	Leu	Ser	Glu	Glu	Thr	Thr	Ala	Gly	Glu	Gln	Ser	Thr	Arg
35			515					520					525			
	Ser	Phe	Ile	Leu	Asp	Lys	Ile	Ile	Glu	Glu	Asp	Asp	Ala	Tyr	Asp	Phe
		530					535					540				
	Ser	Thr	Asp	Tyr	Val											
40																

<210> 24

<211> 503

45 <212> ADN

<213> *Homo sapiens*

<220>

50 <221> CDS

<222> (10)...(400)

55

60

65

# ES 2 305 087 T3

<400> 24

```

5      ggccacgta atg tcc gta gtt cgc tca tcc gtc cat gcc aga tgg att gtg      51
        Met Ser Val Val Arg Ser Ser Val His Ala Arg Trp Ile Val
        1          5          10

10     ggg aag gtg att ggg aca aaa atg caa aag act gct aaa gtg aga gtg      99
        Gly Lys Val Ile Gly Thr Lys Met Gln Lys Thr Ala Lys Val Arg Val
        15          20          25          30

15     acc agg ctt gtt ctg gat ccc tat tta tta aag tat ttt aat aag cgg      147
        Thr Arg Leu Val Leu Asp Pro Tyr Leu Leu Lys Tyr Phe Asn Lys Arg
        35          40          45

15     aaa acc tac ttt gct cac gat gcc ctt cag cag tgc aca gtt ggg gat      195
        Lys Thr Tyr Phe Ala His Asp Ala Leu Gln Gln Cys Thr Val Gly Asp
        50          55          60

20     att gtg ctt ctc aga gct tta cct gtt cca cga gca aag cat gtg aaa      243
        Ile Val Leu Leu Arg Ala Leu Pro Val Pro Arg Ala Lys His Val Lys
        65          70          75

25     cat gaa ctg gct gag atc gtt ttc aaa gtt gga aaa gtc ata gat cca      291
        His Glu Leu Ala Glu Ile Val Phe Lys Val Gly Lys Val Ile Asp Pro
        80          85          90

30     gtg aca gga aag ccc tgt gct gga act acc tac ctg gag agt ccg ttg      339
        Val Thr Gly Lys Pro Cys Ala Gly Thr Thr Tyr Leu Glu Ser Pro Leu
        95          100          105          110

35     agt tcg gaa acc acc cag cta agc aaa aat ctg gaa gaa ctc aat atc      387
        Ser Ser Glu Thr Thr Gln Leu Ser Lys Asn Leu Glu Glu Leu Asn Ile
        115          120          125

35     tct tca gca cag t gaagcgggag tggaagaagg atctgaaggg aaaaactgac      440
        Ser Ser Ala Gln
        130

40     atgtttatgt tatggaaaaa gaaatttttc taagtttcat cacaaaaaaa aaaaaaaaaa      500
        aaa      503

```

<210> 25

<211> 130

<212> PRT

45 <213> *Homo sapiens*

<400> 25

```

50     Met Ser Val Val Arg Ser Ser Val His Ala Arg Trp Ile Val Gly Lys
        1          5          10          15
        Val Ile Gly Thr Lys Met Gln Lys Thr Ala Lys Val Arg Val Thr Arg
        20          25          30
        Leu Val Leu Asp Pro Tyr Leu Leu Lys Tyr Phe Asn Lys Arg Lys Thr
        35          40          45
55     Tyr Phe Ala His Asp Ala Leu Gln Gln Cys Thr Val Gly Asp Ile Val
        50          55          60
        Leu Leu Arg Ala Leu Pro Val Pro Arg Ala Lys His Val Lys His Glu
        65          70          75          80
        Leu Ala Glu Ile Val Phe Lys Val Gly Lys Val Ile Asp Pro Val Thr
        85          90          95
60     Gly Lys Pro Cys Ala Gly Thr Thr Tyr Leu Glu Ser Pro Leu Ser Ser
        100          105          110
        Glu Thr Thr Gln Leu Ser Lys Asn Leu Glu Glu Leu Asn Ile Ser Ser
        115          120          125
65     Ala Gln
        130

```

# ES 2 305 087 T3

<210> 26  
 <211> 651  
 <212> ADN  
 5 <213> *Homo sapiens*

<400> 26

```

10      ggaattcggc acgaggtcga ctccctgtgag gtatggtgct ggggtgcagat gcagtgtggc      60
      tctggatagc accttatgga cagttgtgtc cccaaggaag gatgagaata gctactgaag      120
      tcctaaagag caagcctaac tcaagccatt ggcacacagg cattagacag aaagctggaa      180
      gttgaaatgg tggagtccaa cttgcctgga ccagcttaat ggttctgctc ctggtaacgt      240
      ttttatccat ggatgacttg cttgggtaag gacatgaaga cagttcctgt catacctttt      300
15      aaaggtatgg agagtcggct tgactacact gtgtggagca agtttttaaag aagcaaagga      360
      ctcagaattc atgattgaag aaatgcaggc agacctgtta tcctaaacta gggtttttaa      420
      tgaccacaac aagcaagcat gcagcttact gcttgaaagg gtcttgctc acccaagcta      480
      gagtgcagtg gcctttgaag cttactacag cctcaaactt ctgggctcaa gtgatcctca      540

20

      gcctcccagt ggtctttgta gactgcctga tggagtctca tggcacaaga agattaaaac      600
      agtgtctcca attttaataa atttttgcaa tccaaaaaaaa aaaaaaaaaa a      651

25

<210> 27
<211> 559
<212> ADN
30 <213> Homo sapiens

<220>  

  <221> CDS  

  35 <222> (324)...(519)



<400> 27



```

40      ggctctggat agcaccttat ggacagttgt gtccccaagg aaggatgaga atagctactg      60
      aagtcctaaa gagcaagcct aactcaagcc attggcacac aggcattaga cagaaagctg      120
      gaagttgaaa tgggtggagtc caacttgccct ggaccagctt aatggttctg ctccctggtaa      180
      cgtttttatc catggatgac ttgcttgggt atggagagtc ggcttgacta cactgtgtgg      240
45      agcaagtttt aaagaagcaa aggactcaga attcatgatt gaagaaatgc aggcagacct      300
      gttatcctaa actaggggtt tta atg acc aca aca agc aag cat gca gct tac      353
                        Met Thr Thr Thr Ser Lys His Ala Ala Tyr
                        1          5          10

50      tgc ttg aaa ggg tct tgc ctc acc caa gct aga gtg cag tgg cct ttg      401
      Cys Leu Lys Gly Ser Cys Leu Thr Gln Ala Arg Val Gln Trp Pro Leu
                        15          20          25

      aag ctt act aca gcc tca aac ttc tgg gct caa gtg atc ctc agc ctc      449
55      Lys Leu Thr Thr Ala Ser Asn Phe Trp Ala Gln Val Ile Leu Ser Leu
                        30          35          40

      cca gtg gtc ttt gta gac tgc ctg atg gag tct cat ggc aca aga aga      497
60      Pro Val Val Phe Val Asp Cys Leu Met Glu Ser His Gly Thr Arg Arg
                        45          50          55

      tta aaa cag tgt ctc caa ttt t aataaatttt tgcaatccaa aaaaaaaaaa      549
      Leu Lys Gln Cys Leu Gln Phe
                        60          65

65      aaaaaaaaaa      559

```


```

# ES 2 305 087 T3

<210> 28

<211> 65

<212> PRT

5 <213> *Homo sapiens*

<400> 28

```

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

```

<210> 29

<211> 623

<212> ADN

25 <213> *Homo sapiens*

<220>

<223> secuencia cebadora

30

<400> 29

```

35      tcgagcggcc gcccgggcag gtgtcgactc ctgtgaggta tgggtgctggg tgcagatgca      60
      gtgtggctct ggatagcacc ttatggacag ttgtgtcccc aaggaaggat gagaatagct      120
      actgaagtcc taaagagcaa gcctaactca agccattggc acacaggcat tagacagaaa      180
      gctggaagtt gaaatgggtg agtccaactt gcctggacca gcttaatggg tctgctcctg      240
      gtaacgtttt tatccatgga tgacttgctt ggggtatggag agtcggcttg actacactgt      300
      gtggagcaag ttttaaagaa gcaaaggact cagaattcat gattgaagaa atgcaggcag      360
40      acctgttatc ctaaactagg gtttttaatg accacaacaa gcaagcatgc agcttactgc      420
      ttgaaagggg cttgcctcac ccaagctaga gtgcagtggc ctttgaagct tactacagcc      480
      tcaaaacttc gggctcaagt gatcctcagc ctcccagtggt tctttgtaga ctgcctgatg      540
      gagtctcatg gcacaagaag attaaaacag tgtctccaat ttttaataaat ttttgcaatc      600
45      caaaaaaaaa aaaaaaaaaa aaa      623

```

<210> 30

<211> 23

<212> ADN

50 <213> Secuencia artificial

<220>

<223> Cebador

55

<400> 30

```

60      aggagttct gaggacatg cac      23

```

<210> 31

<211> 22

65 <212> ADN

<213> Secuencia artificial

## ES 2 305 087 T3

<220>

<223> Cebador

5 <400> 31

tcaaggggtg gggatacaca cg 22

10 <210> 32

<211> 22

<212> ADN

15 <213> Secuencia artificial

<220>

<223> secuencia cebadora

20 <400> 32

ctgtgtgct ttcttctg gc 22

25

<210> 33

<211> 24

<212> ADN

30 <213> Secuencia artificial

<220>

<223> secuencia cebadora

35

<400> 33

agtctggaaa tccacatgac caag 24

40

<210> 34

<211> 23

45 <212> ADN

<213> Secuencia artificial

<220>

50 <223> secuencia cebadora

<400> 34

55 cccaatgagg aacctaaagt tgc 23

<210> 35

<211> 23

60 <212> ADN

<213> Secuencia artificial

<220>

65 <223> secuencia cebadora



## ES 2 305 087 T3

<400> 35

ggtgccaaat ctggactctt gtc 23

5

<210> 36

<211> 23

<212> ADN

10

<213> Secuencia artificial

<220>

15

<223> Cebador

<400> 36

gatccatttt cagcagtgct ctg 23

20

<210> 37

<211> 25

25

<212> ADN

<213> Secuencia artificial

<220>

30

<223> Cebador

<400> 37

cagtgttcac agaaggggta ctcac 25

35

<210> 38

40

<211> 20

<212> ADN

<213> Secuencia artificial

<220>

45

<223> Cebador

<400> 38

acgagagcga cacggacaag 20

50

<210> 39

55

<211> 20

<212> ADN

<213> Secuencia artificial

<220>

60

<223> Cebador

<400> 39

65

tctgaggctg tggcaggtgc 20

## ES 2 305 087 T3

<210> 40  
 <211> 21  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <223> Cebador  
 10  
 <400> 40  
  
 ccagtccttg ccaactcgtg c 21  
 15  
 <210> 41  
 <211> 23  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 25 <223> secuencia cebadora  
  
 <400> 41  
 30 ttcgatcttc aaactgtgcc ttg 23  
  
 <210> 42  
 <211> 20  
 35 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 40 <223> Cebador  
  
 <400> 42  
 45 ttggcaacca gaccagcatc 20  
  
 <210> 43  
 50 <211> 22  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 55 <223> Cebador  
  
 <400> 43  
 60 ttcccatag gtgtgagtg cg 22  
  
 <210> 44  
 65 <211> 22  
 <212> ADN

## ES 2 305 087 T3

<213> Secuencia artificial

<220>

5 <223> Cebador

<400> 44

10 gactggtgtt ttgtcgggg tc 22

<210> 45

15 <211> 21

<212> ADN

<213> Secuencia artificial

20 <220>

<223> Cebador

<400> 45

25 ttgtccaag gctgcatgt c 21

30 <210> 46

<211> 22

<212> ADN

<213> Secuencia artificial

35

<220>

<223> Cebador

40 <400> 46

tgccctggtt aagccagaag tc 22

45 <210> 47

<211> 22

<212> ADN

50 <213> Secuencia artificial

<220>

<223> Cebador

55

<400> 47

agcttcactt tggcttgac gg 22

60

<210> 48

<211> 22

<212> ADN

65 <213> Secuencia artificial

## ES 2 305 087 T3

<220>  
 <223> Cebador

5 <400> 48  
  
 ggatcatctgc atcaagggtg gc 22

10 <210> 49  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

15 <220>  
 <223> Cebador

20 <400> 49  
  
 gggtcgtaac cgtgactca gg 22

25 <210> 50  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> Cebador

35 <400> 50  
  
 gcatcctttt ccagtcttc g 21

40 <210> 51  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <223> Cebador

50 <400> 51  
  
 tgcagcaaac atgcctgagc 20

55 <210> 52  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

60 <220>  
 <223> Cebador

65 <220>  
 <223> Cebador

## ES 2 305 087 T3

<400> 52

tggtccacga gcaaagcatg tg 22

5

<210> 53

<211> 22

<212> ADN

10

<213> Secuencia artificial

<220>

15

<223> Cebador

<400> 53

atctctcttc cactcccgct tc 22

20

<210> 54

<211> 22

25

<212> ADN

<213> Secuencia artificial

<220>

30

<223> Cebador

<400> 54

tcggcttgac tacactgtgt gg 22

35

<210> 55

<211> 22

40

<212> ADN

<213> Secuencia artificial

<220>

45

<223> Cebador

<400> 55

tacaaagacc actgggaggc tg 22

50

<210> 56

55

<211> 24

<212> ADN

<213> Secuencia artificial

<220>

60

<223> Cebador

<400> 56

65

cgggaaatcg tgcgtgacat taag 24

## ES 2 305 087 T3

<210> 57  
 <211> 24  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <223> Cebador  
 10  
 <400> 57  
  
 tgatctcctt ctgcaccttg tcgg 24  
 15  
  
 <210> 58  
 <211> 21  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 25 <223> Cebador  
  
 <400> 58  
  
 ttggctaca gcaacagggt g 21  
 30  
  
 <210> 59  
 <211> 22  
 35 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 40 <223> Cebador  
  
 <400> 59  
 45  
 tgtgaggagg ggagattcag tg 22  
  
 <210> 60  
 50 <211> 16  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 55 <223> Cebador  
  
 <400> 60  
 60  
 cgctgacctc aaccag 16  
  
 <210> 61  
 65 <211> 21  
 <212> ADN

## ES 2 305 087 T3

<213> Secuencia artificial

<220>

5 <223> Cebador

<400> 61

10 ctgtttgcc gttcttatta c 21

<210> 62

<211> 24

15 <212> ADN

<213> Secuencia artificial

<220>

20 <223> Cebador

<400> 62

25 cgggaaatcg tgcgtgacat taag 24

<210> 63

30 <211> 24

<212> ADN

<213> Secuencia artificial

35 <220>

<223> Cebador

<400> 63

40 tgatctcctt ctgcatactg tcgg 24

<210> 64

45 <211> 25

<212> ADN

<213> Secuencia artificial

50 <220>

<223> Oligonucleótido antisentido

55 <400> 64

atttgggcat cactggctac aagca 25

60 <210> 65

<211> 25

<212> ADN

65 <213> Secuencia artificial

## ES 2 305 087 T3

<220>  
 <223> Oligonucleótido control inverso  
 5 <400> 65  
     acgaacatcg gtcactacgg gttta      25  
 10 <210> 66  
     <211> 24  
     <212> ADN  
     <213> Secuencia artificial  
 15 <220>  
     <223> Oligonucleótido antisentido  
 20 <400> 66  
     cagagaggtag agacactcgc cgca      24  
 25 <210> 67  
     <211> 24  
     <212> ADN  
 30 <213> Secuencia artificial  
     <220>  
     <223> Oligonucleótido control inverso  
 35 <400> 67  
     acgccgctca cagagtggag agac      24  
 40 <210> 68  
     <211> 25  
     <212> ADN  
 45 <213> Secuencia artificial  
     <220>  
     <223> Oligonucleótido sintético  
 50 <400> 68  
     ttggtgtcat tgggtcaagg gttgg      25  
 55 <210> 69  
     <211> 25  
 60 <212> ADN  
     <213> Secuencia artificial  
     <220>  
 65 <223> Oligonucleótido sintético



## ES 2 305 087 T3

<400> 69

ggttgggaac tgggtactg tggtt 25

5

<210> 70

<211> 24

<212> ADN

10

<213> Secuencia artificial

<220>

15

<223> Oligonucleótido sintético

<400> 70

20

acagggcaga tacggacctc ggtg 24

<210> 71

<211> 24

25

<212> ADN

<213> Secuencia artificial

<220>

30

<223> Oligonucleótido sintético

<400> 71

35

gtggctccag gcatagacgg gaca 24

<210> 72

40

<211> 25

<212> ADN

<213> Secuencia artificial

45

<220>

<223> Oligonucleótido sintético

<400> 72

50

ttgtgggtaa gcagttcat gtcgc 25

<210> 73

55

<211> 25

<212> ADN

<213> Secuencia artificial

60

<220>

<223> Oligonucleótido sintético

<400> 73

65

cgctgtactt tgacgaatgg gtggt 25

## ES 2 305 087 T3

<210> 74  
 <211> 25  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <223> Oligonucleótido sintético  
 10  
 <400> 74  
  
 cctggatcag acgcaagtta tcggc 25  
 15  
 <210> 75  
 <211> 25  
 <212> ADN  
 20 <213> Secuencia artificial  
  
 <220>  
 25 <223> Oligonucleótido sintético  
  
 <400> 75  
  
 cggtattga acgcagacta ggtcc 25  
 30  
  
 <210> 76  
 <211> 25  
 35 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 40 <223> Oligonucleótido sintético  
  
 <400> 76  
  
 ctactcccca cacttcacag ccagg 25  
 45  
  
 <210> 77  
 50 <211> 25  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 55 <223> Oligonucleótido sintético  
  
 <400> 77  
  
 ggaccgctac ttacaccccc tcatc 25  
 60  
  
 <210> 78  
 65 <211> 25  
 <212> ADN  
 <213> Secuencia artificial

## ES 2 305 087 T3

<220>

<223> Oligonucleótido sintético

5 <400> 78

ctcttgatac tccagcggca aacca 25

10 <210> 79

<211> 25

<212> ADN

15 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

20 <400> 79

accaaacggc gacatcatag ttctc 25

25 <210> 80

<211> 24

<212> ADN

30 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

35 <400> 80

gagcccaagc cgttcgtct taag 24

40

<210> 81

<211> 24

<212> ADN

45 <213> Secuencia artificial

<220>

50 <223> Oligonucleótido sintético

<400> 81

55 gaattctgc ttgccgaacc cgcg 24

<210> 82

<211> 25

60 <212> ADN

<213> Secuencia artificial

<220>

65 <223> Oligonucleótido sintético

## ES 2 305 087 T3

<400> 82  
 ccaggtaggc acgagttggc aaaga 25  
 5  
 <210> 83  
 <211> 25  
 <212> ADN  
 10 <213> Secuencia artificial  
 <220>  
 <223> Oligonucleótido sintético  
 15  
 <400> 83  
 agaaacggtt gagcacggat ggacc 25  
 20  
 <210> 84  
 <211> 25  
 25 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 30 <223> Oligonucleótido sintético  
 <400> 84  
 gccattgaag atgccagat cccac 25  
 35  
 <210> 85  
 40 <211> 25  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 45 <223> Oligonucleótido sintético  
 <400> 85  
 50 caccctagac ccgtagaagt taccg 25  
 <210> 86  
 55 <211> 24  
 <212> ADN  
 <213> Secuencia artificial  
 60 <220>  
 <223> Oligonucleótido sintético  
 <400> 86  
 65 cctgcgttg tccctccagc atct 24

## ES 2 305 087 T3

<210> 87  
 <211> 24  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <223> Oligonucleótido sintético  
 10  
 <400> 87  
  
 tctacgacct ccctgttgc gtcc 24  
 15  
  
 <210> 88  
 <211> 25  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 25 <223> Oligonucleótido sintético  
  
 <400> 88  
  
 aagtcacagt ccccgatac cagtc 25  
 30  
  
 <210> 89  
 <211> 25  
 35 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 40 <223> Oligonucleótido sintético  
  
 <400> 89  
  
 ctgaccatag gccctgaca ctgaa 25  
 45  
  
 <210> 90  
 50 <211> 25  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 55 <223> Oligonucleótido sintético  
  
 <400> 90  
 60  
 ttgtcgctt ggagggcata aaacc 25  
  
 <210> 91  
 65 <211> 25  
 <212> ADN

## ES 2 305 087 T3

	<213> Secuencia artificial	
	<220>	
5	<223> Oligonucleótido sintético	
	<400> 91	
10	tctggtcac aacttgctt ccgtg	25
	<210> 92	
	<211> 25	
15	<212> ADN	
	<213> Secuencia artificial	
	<220>	
20	<223> Oligonucleótido sintético	
	<400> 92	
25	cagtgttcg tgggtgctc tgtgg	25
	<210> 93	
30	<211> 23	
	<212> ADN	
	<213> Secuencia artificial	
35	<220>	
	<223> Oligonucleótido sintético	
	<400> 93	
40	gctcaccac cgggcaccaa gca	23
45	<210> 94	
	<211> 25	
	<212> ADN	
	<213> Secuencia artificial	
50	<220>	
	<223> Oligonucleótido sintético	
55	<400> 94	
	tgagagacag tgttcgtgg tgtgc	25
60	<210> 95	
	<211> 25	
	<212> ADN	
65	<213> Secuencia artificial	

## ES 2 305 087 T3

<220>

<223> Oligonucleótido sintético

5 <400> 95

tgccctcaca cgcttggtta tcttc 25

10 <210> 96

<211> 25

<212> ADN

15 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

20 <400> 96

gacaacatcg gaggtctcaa tcacc 25

25 <210> 97

<211> 25

<212> ADN

30 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

35 <400> 97

gttgaggctc tgaacaccac tgttg 25

40 <210> 98

<211> 25

<212> ADN

45 <213> Secuencia artificial

<220>

50 <223> Oligonucleótido sintético

<400> 98

55 gtttggcagc acctcaaca tttag 25

<210> 99

<211> 24

60 <212> ADN

<213> Secuencia artificial

<220>

65 <223> Oligonucleótido sintético

## ES 2 305 087 T3

<400> 99

agcagtttgg cagcaccttc aaca 24

5 <210> 100  
<211> 25  
<212> ADN  
10 <213> Secuencia artificial

<220>  
15 <223> Oligonucleótido sintético

<400> 100

cttctattgg ttcgcacact tccgt 25

20 <210> 101  
<211> 25  
25 <212> ADN  
<213> Secuencia artificial

<220>  
30 <223> Oligonucleótido sintético

<400> 101

ccactaact cggaggctac aacag 25

35 <210> 102  
<211> 25  
40 <212> ADN  
<213> Secuencia artificial

<220>  
45 <223> Oligonucleótido sintético

<400> 102

gtgtcacca caagtctcgg agttg 25

50 <210> 103  
55 <211> 25  
<212> ADN  
<213> Secuencia artificial

<220>  
60 <223> Oligonucleótido sintético

<400> 103

65 ggttacaac ttccacgacg gtttg 25



## ES 2 305 087 T3

<210> 104  
 <211> 24  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <223> Oligonucleótido sintético  
 10  
 <400> 104  
  
 acaacttcca cgacggttg acga 24  
 15  
 <210> 105  
 <211> 24  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 25 <223> Oligonucleótido sintético  
  
 <400> 105  
  
 atctggcatg gacggatgag cgaa 24  
 30  
 <210> 106  
 <211> 24  
 35 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 40 <223> Oligonucleótido sintético  
  
 <400> 106  
  
 gctgggtggt ttccgaactc aacg 24  
 45  
 <210> 107  
 50 <211> 25  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 55 <223> Oligonucleótido sintético  
  
 <400> 107  
 60  
 gtcccaatca ccttccccac aatcc 25  
  
 <210> 108  
 65 <211> 25  
 <212> ADN

## ES 2 305 087 T3

<213> Secuencia artificial  
 <220>  
 5 <223> Oligonucleótido sintético  
 <400> 108  
 10           tcagatcctt cttccactcc cgctt           25  
 <210> 109  
 <211> 25  
 15 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 20 <223> Oligonucleótido sintético  
 <400> 109  
 25           tgctcgtgga acaggtaaag ctctg           25  
 <210> 110  
 30 <211> 24  
 <212> ADN  
 <213> Secuencia artificial  
 35 <220>  
 <223> Oligonucleótido sintético  
 <400> 110  
 40           aagcgagtag gcaggtacgg tcta           24  
 <210> 111  
 45 <211> 24  
 <212> ADN  
 <213> Secuencia artificial  
 50 <220>  
 <223> Oligonucleótido sintético  
 55 <400> 111  
           gcaactcaag cctttggtgg gtcg           24  
 60 <210> 112  
 <211> 25  
 <212> ADN  
 65 <213> Secuencia artificial

## ES 2 305 087 T3

<220>

<223> Oligonucleótido sintético

5 <400> 112

cctaacaccc ctccactaa ccctg 25

10 <210> 113

<211> 25

<212> ADN

<213> Secuencia artificial

15

<220>

<223> Oligonucleótido sintético

20 <400> 113

ttcgccctca ccttcttct agact 25

25

<210> 114

<211> 25

<212> ADN

30 <213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

35

<400> 114

gtctcgaaat ggacaagggtg ctcgt 25

40

<210> 115

<211> 25

<212> ADN

45

<213> Secuencia artificial

<220>

50 <223> Oligonucleótido sintético

<400> 115

55 agcttcactt tggcttgac ggcatt 25

<210> 116

<211> 24

60

<212> ADN

<213> Secuencia artificial

<220>

65 <223> Oligonucleótido sintético

## ES 2 305 087 T3

<400> 116

5                   cggagggaag tcaagtcagc caca                   24

<210> 117  
<211> 24  
<212> ADN  
10 <213> Secuencia artificial

<220>  
15 <223> Oligonucleótido sintético

<400> 117

20                   cggcattcac cctctccagc acct                   24

<210> 118  
<211> 23  
25 <212> ADN  
<213> Secuencia artificial

<220>  
30 <223> Oligonucleótido sintético

<400> 118

35                   cctccacctg ttgcgggct tcc                   23

<210> 119  
<211> 25  
40 <212> ADN  
<213> Secuencia artificial

<220>  
45 <223> Oligonucleótido sintético

<400> 119

50                   ccacattgag ggagtcctct tgcaa                   25

<210> 120  
55 <211> 25  
<212> ADN  
<213> Secuencia artificial

<220>  
60 <223> Oligonucleótido sintético

<400> 120

65                   tacggcagtt ctggttcac ttcga                   25

## ES 2 305 087 T3

<210> 121  
 <211> 24  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <223> Oligonucleótido sintético  
 10  
 <400> 121  
  
 acaccgactg aactgaaggg aggc 24  
 15  
 <210> 122  
 <211> 24  
 20 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 25 <223> Oligonucleótido sintético  
  
 <400> 122  
  
 30 tccacgacct ctcccactta cggc 24  
  
 <210> 123  
 <211> 23  
 35 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 40 <223> Oligonucleótido sintético  
  
 <400> 123  
  
 45 ccttcggggcg ttgtccacc tcc 23  
  
 <210> 124  
 50 <211> 24  
 <212> ADN  
 <213> Secuencia artificial  
  
 55 <220>  
 <223> Oligonucleótido sintético  
  
 <400> 124  
 60  
 ccccgaaaca aacaccagtc aacg 24  
  
 65 <210> 125  
 <211> 24  
 <212> ADN

## ES 2 305 087 T3

<213> Secuencia artificial

<220>

5 <223> Oligonucleótido sintético

<400> 125

10 gcaactgacc acaaaacaag cccc 24

<210> 126

<211> 25

15 <212> ADN

<213> Secuencia artificial

<220>

20 <223> Oligonucleótido sintético

<400> 126

25 ggccattgag tccctccata gcagc 25

<210> 127

30 <211> 25

<212> ADN

<213> Secuencia artificial

35 <220>

<223> Oligonucleótido sintético

<400> 127

40 cgacgatacc tccctgagtt accgg 25

45

50

55

60

65