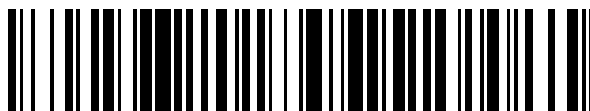


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 539 830**

51 Int. Cl.:

**C12N 15/00** (2006.01)

**C07K 5/00** (2006.01)

**C07H 21/02** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.01.2007 E 07718104 (8)**

97 Fecha y número de publicación de la concesión europea: **25.03.2015 EP 1973946**

54 Título: **ROS quinasa mutante y de translocación en el carcinoma pulmonar no microcítico humano**

30 Prioridad:

**20.01.2006 US 760634 P**

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

**06.07.2015**

73 Titular/es:

**CELL SIGNALING TECHNOLOGY, INC. (100.0%)  
3 TRASK LANE  
DANVERS, MA 01923, US**

72 Inventor/es:

**GUO, AILAN y  
POSSEMATO, ANTHONY**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 539 830 T3**

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

## DESCRIPCIÓN

ROS quinasa mutante y de translocación en el carcinoma pulmonar no microcítico humano

**Solicitudes relacionadas**

5 Esta solicitud reivindica la prioridad y el beneficio del documento USSN 60/760.634, presentado el 20 de enero de 2006.

**Campo de la invención**

La invención se refiere, en general, a proteínas y genes implicado en el cáncer, y a la detección, el diagnóstico y el tratamiento del cáncer.

**Antecedentes de la invención**

10 Muchos cánceres se caracterizan por alteraciones en las vías de señalización celular que conducen a un control aberrante de los procesos celulares, o a un crecimiento y proliferación descontrolados de células. Estas alteraciones a menudo son provocadas por cambios en la actividad de proteínas de señalización concretas, tales como quinasas. Entre estos cánceres se encuentra el carcinoma pulmonar no microcítico (NSCLC). El NSCLC es la principal causa de muerte por cáncer en EEUU, y supone aproximadamente 87% de todos los cánceres de pulmón. Se producen  
15 aproximadamente 151.000 nuevos casos anuales de NSCLC en EEUU, y se calcula que más de 120.000 pacientes mueren al año por esta enfermedad solo en EEUU. Véase "Cancer Facts and Figures 2005", American Cancer Society. El NSCLC, que comprende tres subtipos diferenciados, a menudo solo se detecta después de que haya metastatizado, y así la tasa de mortalidad es del 75% en los dos años siguientes al diagnóstico.

20 Se sabe que las translocaciones de genes que producen proteínas de fusión de quinasas con una actividad de señalización aberrante pueden conducir directamente a ciertos cánceres. Por ejemplo, se ha demostrado directamente que la oncoproteína BCR-ABL, una proteína de fusión de tirosina quinasa, es el agente causativo en la leucemia mielógena crónica humana (CML). La oncoproteína BCR-ABL, que se encuentra en al menos 90-95% de los casos de CML, es generada por la translocación de secuencias génicas desde la proteína tirosina quinasa c-ABL sobre el cromosoma 9 hacia las secuencias BCR sobre el cromosoma 22, produciendo el denominado cromosoma  
25 Filadelfia. Véase, por ejemplo, Kurzock *et al.*, *N. Engl. J. Med.*, 319: 990-998 (1988). También se observa translocación en casos de leucemia linfocítica aguda y AML.

30 Se han descrito translocaciones de genes que conducen a proteínas mutantes o de fusión implicadas en una diversidad de otros cánceres distintos. Por ejemplo, Falini *et al.*, *Blood*, 99(2): 409-426 (2002), analizan las translocaciones que se sabe que aparecen en cánceres hematológicos. Hasta la fecha, sólo se ha descrito un número limitado de translocaciones de genes y proteínas mutantes que aparecen en cánceres de pulmón, incluyendo la translocación t(15;19) que implica a Notch3. Véase Dang *et al.*, *J. Natl. Can. Instit.*, 92(16): 1355-1357 (2000). Se han descubierto defectos en la expresión y/o la actividad de la proteína de unión a ARN-6 (RBM-6) en carcinomas pulmonares microcíticos y no microcíticos. Véase Drabkin *et al.*, *Oncogene*, 8(16): 2589-2597 (1999). Sin embargo, hasta la fecha no se han descrito translocaciones en el cáncer de NSCLC humano que impliquen a  
35 proteína quinasas.

40 Se han encontrado defectos en la expresión y/o la activación de SLC34A2 en el cáncer ovárico humano. Véase Rangel *et al.*, *Oncogene*, 22(46): 7225-7232 (2003). De forma similar, se han descrito defectos en la expresión de ROS quinasa que surgen de la translocación de FIG-ROS del(6)(q21,q21) en el glioblastoma. Véase Charest *et al.*, *Genes Chromos. Canc.*, 37(1): 58-71 (2003). También se ha descrito una forma truncada de ROS quinasa capaz de dirigir el crecimiento tumoral en ratones. Véase Birchmeier *et al.*, *Mol. Cell. Bio.*, 6(9): 3109-3115 (1986). Hasta la fecha, no se conocen mutaciones puntuales activantes que aparezcan en ROS quinasa.

Charest A. *et al.*, *Proceedings of the National Academy of Sciences*, 100(3): 916-921 (2003), describen una proteína oncogénica que comprende el dominio quinasa de c-ROS.

45 Nagarajan L. *et al.*, *Proceedings of the National Academy of Sciences*, 83(17): 6568-6572 (1986), y Xu H. *et al.*, *Genomics*, 62(2): 281-284 (1999), describen sondas adecuadas para detectar polinucleótidos de SLC34A2 y ROS.

Birchmeier C. *et al.*, *BioEssays: News and Reviews in Molecular Cellular and Developmental Biology*, 15(3): 185-190 (1993), describen el control del receptor de c-ROS de la morfogénesis y el crecimiento epitelial.

El documento WO2004/023973 describe las secuencias del polinucleótido y del polipéptido de SLC34A2 y ROS.

50 La identificación de translocaciones y mutaciones en cánceres humanos es muy deseable, porque esto puede conducir al desarrollo de nuevos productos terapéuticos que se dirijan a dichas proteínas de fusión o mutantes, y a nuevos diagnósticos para identificar pacientes que tengan estas translocaciones de genes. Por ejemplo, BCR-ABL se ha convertido en una diana para el desarrollo de productos terapéuticos para tratar la leucemia. En fechas más recientes, Gleevec® (mesilato de imatinib, STI-571), un inhibidor de molécula pequeña de la ABL quinasa, ha sido

aprobado para el tratamiento de la CML. Este fármaco es el primero de una nueva clase de agentes antiproliferativos diseñados para interferir con las vías de señalización que dirigen el crecimiento de las células tumorales. El desarrollo de este fármaco representa un avance significativo frente a las terapias convencionales para la CML y ALL, la quimioterapia y la radiación, que están plagadas de efectos secundarios muy conocidos y a menudo tienen un efecto limitado, puesto que no están dirigidas específicamente a las causas subyacentes de las malignidades. De modo similar, se han descrito reactivos y métodos para detectar específicamente proteínas de fusión de BCR-ABL en pacientes, para identificar a los pacientes que responden con mayor probabilidad a los inhibidores dirigidos, tales como Gleevec®.

Por consiguiente, sigue siendo necesaria la identificación de nuevas translocaciones o mutaciones de genes que produzcan proteínas de fusión o mutantes implicadas en el avance de los cánceres humanos, que incluyen cánceres de pulmón, tales como NSCLC, y el desarrollo de nuevos reactivos y métodos para el estudio y la detección de dichas proteínas de fusión. La identificación de estas proteínas de fusión permitirá desarrollar de modo deseable, entre otras cosas, nuevos métodos para seleccionar pacientes para las terapias dirigidas, así como la selección de nuevos fármacos que inhiban estas proteínas mutantes/de fusión.

### Sumario de la invención

Según la invención, ahora se ha identificado una nueva translocación génica, (4p15, 6q22), en el carcinoma pulmonar no microcítico (NSCLC) que produce proteínas de fusión que combinan parte de la proteína de isoforma NaPi-3b transportadora de fosfato dependiente de sodio (SLC34A2) con la quinasa del precursor de la tirosina proteína quinasa ROS de protooncogén (ROS). Se espera que las dos proteínas de fusión de SLC34A2-ROS conserven la actividad de ROS tirosina quinasa y que dirijan la proliferación y la supervivencia de NSCLC en un subconjunto de aquellos cánceres en los que se expresa la proteína de fusión.

Por tanto, la invención proporciona, en parte, polinucleótidos aislados según la reivindicación 1 y vectores que codifican los polipéptidos de ROS mutantes descritos, es decir, los polipéptidos de fusión de quinasa SLC34A2-ROS, sondas y ensayos para detectarlos, polipéptidos de ROS mutantes aislados, polipéptidos mutantes recombinantes, y reactivos para detectar los polinucleótidos y polipéptidos de ROS mutantes. La identificación descrita de las nuevas proteínas de ROS quinasa mutantes y la translocación de SLC34A2 permite desarrollar nuevos métodos para determinar la presencia de polinucleótidos o polipéptidos de ROS mutantes en una muestra biológica, métodos para seleccionar compuestos que inhiban las proteínas de quinasa mutantes, y métodos para inhibir el avance de un cáncer que se caracteriza por la expresión de polinucleótidos o polipéptidos de ROS mutantes, que también son proporcionados por la invención. Los aspectos y las realizaciones de la invención se describen con más detalle a continuación.

### Breve descripción de los dibujos

Figura 1: Muestra la localización del gen SLC34A2 y el gen ROS sobre los cromosomas 4p y 6q, respectivamente (panel A), y las localizaciones del dominio de las proteínas de SLC34A2 y ROS de longitud completa, así como las de los dos variantes de la proteína de fusión de SLC34A2-ROS (paneles B y C). En el primer variante (largo), la zona de unión de la fusión aparece en el resto 1750 cadena arriba del dominio transmembrana de ROS, mientras que el segundo variante (corto) aparece en el resto 1853.

Figura 2A: Es la secuencia de aminoácidos (código de una letra) del primer variante (largo) de la proteína de fusión de SLC34A2-ROS humana (SEQ ID NO:1) (panel superior), indicándose también la secuencia codificadora de ADN (SEQ ID NO:2) (panel inferior); los restos del resto SCL34A2 se indican en cursiva, mientras que los restos del dominio quinasa de ROS se indican en negrita.

Figura 2B: Es la secuencia de aminoácidos (código de una letra) del segundo variante (corto) de la proteína de fusión de SLC34A2-ROS humana (SEQ ID NO:3) (panel superior), indicándose también la secuencia codificadora de ADN (SEQ ID NO:4) (panel inferior); los restos del resto SCL34A2 se indican en cursiva, mientras que los restos del dominio quinasa de ROS se indican en negrita.

Figura 3: Es la secuencia de aminoácidos (código de una letra) de la proteína SLC34A2 humana (SEQ ID NO:5) (n.º de registro de SwissProt 095436) (panel superior), indicándose también la secuencia codificadora de ADN (SEQ ID NO:6) (n.º de registro de GeneBank NM\_006424) (panel inferior); los restos implicados en la translocación están subrayados.

Figura 4A: Es la secuencia de aminoácidos (código de una letra) de la ROS quinasa humana (SEQ ID NO:7) (n.º de registro de SwissProt P08922); los restos implicados en la translocación del primer variante (largo) están subrayados, mientras que los restos subrayados en negrita son los implicados en la translocación del segundo variante (corto).

Figura 4B: Es la secuencia codificadora de ADN de la ROS quinasa humana (SEQ ID NO:8) (n.º de registro de GeneBank NM\_002944); los restos implicados en la translocación del primer variante (largo) están subrayados, mientras que los restos subrayados en negrita son los implicados en la translocación del segundo variante (corto).

Figura 5: Es un análisis de la transferencia Western de extractos de una línea celular de NSCLC humana (HCC78) que muestra la expresión de la forma de ROS que tiene un peso molecular mucho menor que la ROS de longitud completa/de tipo salvaje.

5 Figura 6: Es un gel que muestra la detección de ROS a través del producto de 5'RACE con cebadores de ROS después de 2 rondas de PCR; se muestran los cebadores empleados (SEQ ID NO:13-15).

Figura 7: Son geles que muestran la detección del gen de fusión formado por la translocación de SLC34A2 y ROS por RT-PCR; se muestra la secuencia de la proteína (y ADN) de la zona de unión de la fusión del exón 4/exón 32 (SEQ ID NO:9 y SEQ ID NO:10) y de la zona de unión de la fusión del exón 4/exón 34 (SEQ ID NO:11 y SEQ ID NO:12) de los dos respectivos variantes (largo y corto).

10 Figura 8: Presenta diagramas (parte superior) que muestran la localización de los exones 1-4 en el gen SLC34A2 y los exones 32-34 en el gen ROS que están implicados en la translocación que produce la proteína de fusión; las flechas indican las localizaciones de los cebadores empleados para la amplificación de PCR de los variantes de la proteína de fusión, mostrándose las secuencias de los cebadores (SEQ ID NO:17-20).

15 Figura 9: Es un gel que muestra la expresión de la proteína de fusión de SLC34A2-ROS (primer variante (largo)) en células 293 transfectadas (riñón embrionario humano), comparado con controles (carriles 1 y 2).

20 Figura 10: Muestra la inhibición de ARNip de la ROS quinasa mutante en líneas celulares de NSCLC humanas: el panel A muestra una gráfica de la inhibición celular después de la transfección de ARNip, el panel B es una inmunotransferencia que muestra la inactivación específica de ROS y la mayor apoptosis (en la línea celular dirigida por ROS mutante), y el panel C es una inmunotransferencia que muestra la menor actividad de las moléculas de señalización corriente abajo de ROS.

Figura 11: Es una imagen que muestra la detección específica de la translocación/fusión de SLC34A2-ROS (en una línea celular de NSCLC humana) mediante FISH empleando una sonda de dos colores de ruptura de dos colores.

#### Descripción detallada de la invención

25 Según la invención, ahora se ha identificado una translocación génica previamente desconocida que produce una proteína de fusión de quinasa mutante, SLC34A2-ROS, en el carcinoma pulmonar no microcítico humano (NSCLC), un subtipo del carcinoma pulmonar. La translocación, que aparece entre el cromosoma (4p15) y el cromosoma (6q22), produce dos variantes de proteína de fusión que combinan el N-terminal de la proteína de isoforma NaPi-3b transportadora de fosfato dependiente de sodio (SLC34A2), una proteína transportadora de fosfato de 690 aminoácidos, con los dominios transmembrana y quinasa de la quinasa del precursor de la tirosina proteína quinasa ROS de protooncogén (ROS), un receptor de tirosina quinasa de 2347 aminoácidos. Se espera que las proteínas de fusión de SLC34A2-ROS resultantes, que tienen una longitud de 724 aminoácidos (variante largo) y de 621 aminoácidos (variante corto), respectivamente, conserven la actividad quinasa y dirijan la proliferación y la supervivencia de un subconjunto de tumores NSCLC humanos en el que se expresa la proteína de fusión.

35 Aunque se han descrito unas pocas translocaciones de genes que producen proteínas de fusión aberrantes que implican a la ROS quinasa, que incluye la translocación FIG-ROS del(6)(q21,q21) en el glioblastoma (véase Charest *et al.*, (2003), *supra*) y una forma truncada activa de ROS (véase Birchmeier *et al.*, *supra*), la translocación de SLC34A2-ROS y las proteínas de fusión dscritas en la presente son nuevas, y esta quinasa de fusión se ha descrito por primera vez en NSCLC humano. La SLC34A2 es una proteína transportadora de fosfato que se expresa en el pulmón y el intestino delgado humanos, y que tiene una actividad dependiente de sodio. Se han descubierto defectos en la expresión y/o la actividad de SLC34A2 en el cáncer de ovario. Véase Rangel *et al.*, *supra*. La ROS es un receptor transmembrana de tirosina quinasa que pertenece a la subfamilia de receptores de la insulina, y está implicada en procesos de diferenciación y proliferación celular. La ROS se expresa, en seres humanos, en células epiteliales de una diversidad de tejidos diferentes. Se han encontrado defectos en la expresión y/o la activación de ROS en el glioblastoma, así como en tumores del sistema nervioso central. Véase, por ejemplo, Charest *et al.* (2003), *supra*.

40 Tal como se describe más a fondo a continuación, el gen de translocación y la proteína de fusión de SLC34A2-ROS se han aislado y secuenciado en la presente, y se han producido ADNc para expresar la proteína de quinasa mutante. Por consiguiente, la invención proporciona, en parte, polinucleótidos aislados según la reivindicación 1, que codifican polipéptidos de fusión de SLC34A2-ROS, sondas de ácidos nucleicos que se hibridan con dichos polinucleótidos, y métodos, vectores y células hospedantes para utilizar dichos polinucleótidos para producir polipéptidos de ROS mutantes recombinantes. La invención también proporciona, en parte, polipéptidos aislados según la reivindicación 7, que comprenden secuencias de aminoácidos que codifican polipéptidos de fusión de SLC34A2-ROS, polipéptidos mutantes recombinantes, y reactivos aislados que se unen específicamente y/o detectan los polipéptidos de fusión de SLC34A2-ROS, pero que no unen ni detectan la SLC34A2 de tipo salvaje ni la ROS de tipo salvaje. Estos aspectos de la invención, que se describirán con más detalle a continuación, serán útiles, entre otras cuestiones, para estudiar más a fondo los mecanismos de los cánceres dirigidos por la expresión/actividad de la ROS quinasa mutante, para identificar carcinomas de pulmón y otros cánceres que se

caracterizan por la translocación y/o las proteínas de fusión de SLC34A2-ROS, y para practicar los métodos de la invención, según se describe con más detalle a continuación.

La identificación de la translocación y los mutantes de ROS quinasa nuevos tiene importantes implicaciones para el diagnóstico y el tratamiento potencial de enfermedades, tales como NSCLC, que se caracterizan por dicha translocación y/o proteína de fusión. El NSCLC es la principal causa de muerte por cáncer en EEUU, y a menudo es difícil de diagnosticar antes de que metastatice, lo cual aumenta la dificultad de curar o tratar con eficacia esta enfermedad. Así, la tasa de mortalidad de NSCLC es del 75% en los dos años siguientes al diagnóstico. Véase American Cancer Society, *supra*. Aunque en la actualidad se han aprobado inhibidores de EGFR dirigidos para el tratamiento del NSCL, se prevé que esta terapia pueda ser parcial o totalmente ineficaz en los pacientes que tengan tumores en los que se expresa la ROS quinasa mutante (en lugar o además de EGFR) y que dirige la enfermedad, en su totalidad o en parte.

Por tanto, el presente descubrimiento de las proteínas de fusión de SLC34A2-ROS que resultan de la translocación génica en NSCLC, que se piensa que dirige la proliferación y la supervivencia de un subconjunto de tumores NSCLC, permite desarrollar importantes métodos nuevos para identificar con precisión los cánceres de pulmón de mamíferos (tales como NSCLC), así como otros cánceres, en los que se expresa la proteína de fusión de SLC34A2-ROS o la ROS quinasa troncada. Es muy probable que estos tumores respondan a inhibidores de la actividad quinasa de las ROS quinasa mutantes. La capacidad para identificar, lo más temprano posible, los cánceres que son dirigidos por una ROS quinasa mutante ayudará en gran medida en la determinación clínica de cuáles serán los productos terapéuticos, o la combinación de productos terapéuticos, que serán más apropiados para un paciente concreto, ayudando así a evitar recetar inhibidores que se dirijan a otras quinasa que no son, de hecho, la principal molécula de señalización que dirige al cáncer.

Por consiguiente, la invención proporciona, en parte, métodos para detectar la presencia de una translocación de SLC34A2-ROS (t(4,6)(p15, q22)) y/o un polipéptido de fusión en un cáncer empleando reactivos específicos de fusión y específicos de mutante de la invención. Estos métodos pueden practicarse, por ejemplo, para identificar un cáncer, tal como un tumor NSCLC, que es probable que responda a un inhibidor de la actividad ROS quinasa de la proteína mutante. La invención también proporciona, en parte, métodos para determinar si un compuesto inhibe el avance de un cáncer caracterizado por un polipéptido de fusión de SLC34A2-ROS. La invención también proporciona un método para inhibir el avance de un cáncer que expresa un polipéptido de fusión de SLC34A2-ROS, mediante la inhibición de la expresión y/o la actividad del polipéptido mutante. Estos métodos se describen con más detalle a continuación.

Otros aspectos, ventajas y realizaciones de la invención se describen con más detalle a continuación.

### Definiciones

Tal como se emplean en la presente, los siguientes términos y expresiones tienen los significados indicados.

Un “anticuerpo” o “anticuerpos” se refiere a todos los tipos de inmunoglobulinas, que incluyen IgG, IgM, IgA, IgD, e IgE, que incluyen sus fragmentos F<sub>ab</sub> o de reconocimiento del antígeno, incluyendo anticuerpos quiméricos, policlonales y monoclonales. La expresión “anticuerpo humanizado”, tal como se emplea en la presente, se refiere a moléculas de anticuerpos en las que se han reemplazado aminoácidos en las regiones que no se unen al antígeno para que se parezcan más a un anticuerpo humano, aunque aún conservan la capacidad de unión original.

La expresión “biológicamente activo” se refiere a una proteína que tiene las funciones estructurales, reguladoras o bioquímicas de una molécula natural. De forma similar, “inmunológicamente activo” se refiere a la capacidad de un polipéptido de fusión ROS troncado o un polipéptido de fusión de SLC34A2-ROS natural, recombinante o sintético, o cualquiera de sus oligopéptidos, para inducir una respuesta inmunológica específica en animales o células apropiados y para unirse con anticuerpos específicos.

La expresión “muestra biológica” se emplea en su sentido más amplio, y significa cualquier muestra biológica sospechosa de contener polinucleótidos o polipéptidos de fusión de SLC34A2-ROS o de ROS troncados, o sus fragmentos, y puede comprender una célula, cromosomas aislados de una célula (por ejemplo, un frotis de cromosomas en metafase), ADN genómico (en disolución o unido a un soporte sólido, tal como para un análisis de la transferencia Southern), ARN (en disolución o unido a un soporte sólido, tal como para un análisis de la transferencia Northern), ADNc (en disolución o unido a un soporte sólido), un extracto procedente de células, sangre, orina, médula ósea o un tejido, y similares.

La expresión “que se caracteriza”, con respecto a un cáncer y a un polinucleótido o polipéptido de ROS mutante, significa un cáncer en el que está presente la translocación del gen SLC34A2-ROS y/o el polipéptido de fusión expresado, comparado con un cáncer en el que dicha translocación y/o polipéptido de fusión no está presente. La presencia de dicho polipéptido de fusión puede dirigir, en su totalidad o en parte, el crecimiento y la supervivencia de dicho cáncer.

“Consenso” se refiere a una secuencia de ácido nucleico que se ha resecuenciado para resolver las bases desconocidas, o que se ha extendido empleando XL-PCR™ (Perkin Elmer, Norwalk, Conn.) en la dirección 5' y/o 3' y

se ha resecuenciado, o que se ha ensamblado a partir de las secuencias solapantes de más de un clon Incyte empleando el sistema de ensamblaje de fragmentos GELVIEW™ (GCG, Madison, Wis.), o que se extendido y también ensamblado.

5 Un “producto terapéutico que inhibe la ROS quinasa” significa cualquier composición que comprende uno o más compuestos, productos químicos o biológicos, que inhibe, directa o indirectamente, la expresión y/o la actividad de la ROS de tipo salvaje o truncada, solo y/o como parte de las proteínas de fusión de SLC34A2-ROS.

10 Un “derivado” se refiere a la modificación química de una secuencia de ácido nucleico que codifica el polipéptido de fusión de SLC34A2-ROS o el propio polipéptido codificado. Los ejemplos de dichas modificaciones serían la sustitución de un hidrógeno por un grupo alquilo, acilo o amino. Un derivado de ácido nucleico codificará un polipéptido que conserva las características biológicas fundamentales de la molécula natural.

Un “marcador detectable”, con respecto a un polipéptido, un polinucleótido o un reactivo descrito en la presente, significa una modificación química, biológica u otra modificación, que incluye modificaciones de fluorescencia, masa, restos, tintes, radioisótopos, marcadores o marcajes, etc., mediante la cual puede detectarse la presencia de la molécula de interés.

15 La “expresión” o “expresado”, con respecto a un polipéptido de fusión de SLC34A2-ROS en una muestra biológica significa significativamente expresado, en comparación con una muestra control en la que este polipéptido de fusión no está significativamente expresado.

20 Un “péptido marcado con un isótopo pesado” (empleado de modo intercambiable con péptido AQUA) significa un péptido que comprende al menos un marcador de un isótopo pesado, que es adecuado para la detección o cuantificación absoluta de una proteína, según se describe en el documento WO/03016861, “Absolute Quantification of Proteins and Modified Forms Thereof by Multistage Mass Spectrometry” (Gygi *et al.*), que se analiza más a fondo a continuación. La expresión “detecta específicamente” con respecto a dicho péptido AQUA significa que el péptido solo detectará y cuantificará polipéptidos y proteínas que contienen la secuencia peptídica AQUA y que no detectará sustancialmente polipéptidos y proteínas que no contienen la secuencia peptídica AQUA.

25 “Aislado” (o “sustancialmente purificado”) se refiere a secuencias de ácidos nucleicos o de aminoácidos que se han retirado de su entorno natural, aislado o separado. Preferiblemente están al menos 60% exentas, más preferiblemente 75% exentas, y lo más preferiblemente 90% o más exentas de otros componentes con los que están asociados en la naturaleza.

30 “Mimético” se refiere a una molécula, cuya estructura se desarrolla a partir del conocimiento de la estructura del polipéptido de fusión de SLC34A2-ROS, o de una de sus porciones, y que, como tal, es capaz de realizar algunas o todas las acciones de moléculas similares a proteínas asociadas con la translocación.

Un polinucleótido o un polipéptido “de ROS mutante” significa un polinucleótido o un polipéptido de fusión de SLC34A2-ROS, según se describe en la presente.

35 Un “polinucleótido” (o “secuencia de nucleótidos”) se refiere a un oligonucleótido, nucleótido o polinucleótido, y a sus fragmentos o porciones, y a ADN o ARN de origen genómico o sintético, que pueden ser monocatenarios o bicatenarios, y que representan la hebra sentido o antisentido.

40 Un “polipéptido” (o “secuencia de aminoácidos”) se refiere a una secuencia de un oligopéptido, péptido, polipéptido o proteína, y a sus fragmentos o porciones, y a moléculas naturales o sintéticas. Cuando en la presente se menciona una “secuencia de aminoácidos” para referirse a una secuencia de aminoácidos de una molécula de proteína natural, la “secuencia de aminoácidos” y los términos similares, tales como “polipéptido” o “proteína”, no pretenden limitar la secuencia de aminoácidos a la secuencia de aminoácidos completa nativa asociada con la molécula de proteína mencionada.

45 Un “polinucleótido de fusión de SLC34A2-ROS” se refiere a la secuencia de ácido nucleico de un polinucleótido de fusión o producto génico de translocación de SLC34A2-ROS sustancialmente purificado, según se describe en la presente memoria, obtenido de cualquier especie, en particular de mamífero, que incluye especies bovinas, ovinas, porcinas, murinas, equinas, y preferiblemente la especie humana, procedente de cualquier fuente, tanto natural como sintética, semisintética o recombinante.

50 Un “polipéptido de fusión de SLC34A2-ROS” se refiere a la secuencia de aminoácidos de un polipéptido de fusión SLC34A2-ROS sustancialmente purificado, descrito en la presente memoria, obtenido de cualquier especie, en particular de mamífero, que incluye especies bovinas, ovinas, porcinas, murinas, equinas, y preferiblemente la especie humana, procedente de cualquier fuente, tanto natural como sintética, semisintética o recombinante.

55 Las expresiones “se une específicamente a” (o “unión específica”) con referencia a la interacción de un anticuerpo y una proteína o un péptido, significan que la interacción depende de la presencia de una estructura concreta (es decir, el determinante antigénico o epitopo) sobre la proteína; en otras palabras, el anticuerpo reconoce y se une a una estructura de proteína específica, en lugar de a proteínas en general. La expresión “no se une”, con respecto a

la unión de un anticuerpo a otras secuencias o determinantes antigénicos distintos a los que le son específicos, significa que no reacciona sustancialmente con estos, comparado con la unión del anticuerpo al determinante antigénico o secuencia para los cuales el anticuerpo es específico.

5 La expresión "condiciones rigurosas", con respecto a las condiciones de hibridación de secuencias o sondas, es la "rigurosidad" que se produce dentro de un intervalo desde aproximadamente  $T_m$  menos 5 °C (5 °C por debajo de la temperatura de fusión ( $T_m$ ) de la sonda o secuencia) hasta aproximadamente 20 °C a 25 °C por debajo de  $T_m$ . Las condiciones rigurosas típicas son una incubación durante la noche a 42 °C en una disolución que comprende formamida al 50%, 5 x SSC (NaCl 750 mM, citrato de trisodio 75 mM), fosfato de sodio 50 mM (pH 7,6), 5 x disolución de Denhardt, sulfato de dextrano al 10%, y 20 microgramos/ml de ADN de esperma de salmón cizallado y desnaturalizado, seguido del lavado de los filtros en 0,1x SSC a aproximadamente 65 °C. Tal como entenderán los expertos en la técnica, la rigurosidad de la hibridación puede alterarse para identificar o detectar secuencias polinucleotídicas idénticas o relacionadas.

15 Un "variante" de un polipéptido de fusión de SLC34A2-ROS se refiere a una secuencia de aminoácidos que está alterada en uno o más aminoácidos. El variante puede tener cambios "conservativos", en los que un aminoácido sustituido tiene propiedades estructurales o químicas similares, por ejemplo, la sustitución de leucina por isoleucina. En casos más raros, un variante puede tener cambios "no conservativos", por ejemplo, la sustitución de una glicina por un triptófano. Otras pequeñas variaciones similares también pueden incluir deleciones o inserciones de aminoácidos, o ambas. Pueden encontrarse directrices para determinar los restos aminoácidos que pueden sustituirse, insertarse o deleccionarse sin abolir la actividad biológica o inmunológica empleando programas informáticos muy conocidos en la técnica, por ejemplo, el software DNASTAR.

#### A. Identificación de ROS quinasa mutante en NSCLC humano

25 La nueva translocación de un gen humano descrita en la presente, que se produce entre el cromosoma (4p15) y el cromosoma (6q22) en NSCLC humano y que resulta en la expresión de dos proteínas de fusión variantes que combinan el N-terminal (exones 1-4) de SLC34A2 con los dominios transmembrana y quinasa (exones 32-43 o exones 34-43, respectivamente) de ROS, se identificó, de forma sorprendente, durante el estudio de los perfiles de péptidos fosforilados globales en extractos de una línea celular (HCC78) de carcinoma pulmonar no microcítico humano (NSCLC), un subtipo de cáncer de pulmón. Los cromosomas, genes y productos de corte y empalme alternativo (largo y corto) implicados en esta translocación se muestran en la figura 1.

30 El perfil de fosforilación de esta línea celular se aclaró empleando una técnica recientemente descrita para el aislamiento y la caracterización espectrométrica de masas de péptidos modificados a partir de mezclas complejas (véase la publicación de patente n.º 20030044848, Rush *et al.*, "Immunoaffinity Isolation of Modified Peptides from Complex Mixtures" (la técnica "IAP"), tal como se describe más a fondo en el ejemplo 1 de la presente. La aplicación de la técnica IAP empleando un anticuerpo específico de fosfotirosina (CELL SIGNALING TECHNOLOGY, INC., Beverly, Mass., 2003/04 n.º de catálogo 9411), indicó que la línea celular HCC78 expresa la ROS quinasa (por contraste con la mayoría de las otras líneas celulares, que no la expresan), pero que la quinasa estaba aparentemente truncada (véase la figura 5). La selección identificó muchas otras quinasa activadas en la línea celular que incluye a ROS. El análisis de la secuencia 5' a ROS mediante 5' RACE indicó que la quinasa estaba condensada al N-terminal de SLC34A2 (véase la figura 6).

40 La expresión del polipéptido de fusión de SLC34A2-ROS en la línea de células HCC78 después fue confirmada mediante análisis de la transferencia Western, para estudiar la expresión de la ROS quinasa (proteína de fusión en células HCC78), y mediante inmunoprecipitación con un anticuerpo específico de p-tirosina para confirmar su fosforilación (véase el ejemplo 2, figura 5). El gen de la fusión de SLC34A2-ROS se amplificó mediante PCR, se aisló y se secuenció (véase el ejemplo 4; figura 7 (panel superior)). Tal como se muestra en el panel B de la figura 1, la translocación de SLC34A2-ROS combina el N-terminal de SLC34A2 (aminoácidos 1-126) con los dominios transmembrana y quinasa de ROS (aminoácidos 1750-2347 o aminoácidos 1853-2347, respectivamente) (véanse también las SEQ ID NO:3 y 5), para producir dos variantes de fusión (largo y corto) (véase el panel C de la figura 1). La translocación conserva el dominio 5'-más transmembrana de SLC34A2. Se espera que las proteínas de fusión de SLC34A2-ROS resultantes, que comprenden 724 aminoácidos y 621 aminoácidos, respectivamente (véase el panel C de la figura 1 y las figuras 2A-B (SEQ ID NO:1 y 3)) conserven la actividad quinasa de ROS. Los exones implicados y las zonas de unión de la fusión aparecen en la figura 8.

50 El ADNc que codifica el variante largo de la proteína de fusión de SLC34A2-ROS después se transfectó en células 293 (células de riñón embrionario humano) para establecer que la proteína de fusión que se expresa con el peso molecular esperado es SLC34A2-ROS, que aparece en células HCC78. Véase la figura 9.

55 La inhibición de la actividad ROS quinasa de la proteína de fusión de SLC34A2-ROS puede demostrarse en la línea de células HCC78 empleando el silenciamiento con ARNi según técnicas muy conocidas, o empleando un inhibidor de quinasa dirigido con actividad contra ROS. Los resultados de este ensayo (véase el ejemplo 3) confirman que la proteína de fusión verdaderamente está dirigiendo la proliferación y la supervivencia de esta línea celular de NSCLC. El perfil de fosfopéptidos global y un análisis FISH de tumores NSCLC humanos indican que un pequeño porcentaje de pacientes de hecho no portan esta mutación (véanse los ejemplos 7 y 9), y que estos pacientes pueden

beneficiarse de la terapia con inhibidores de ROS.

### *B. Polinucleótidos aislados*

5 La presente descripción proporciona, en parte, polinucleótidos aislados que codifican polipéptidos de fusión de SLC34A2-ROS, sondas nucleotídicas que se hibridan con estos polinucleótidos, y métodos, vectores y células hospedantes para utilizar dichos polinucleótidos para producir polipéptidos de fusión recombinantes.

10 A menos que se indique lo contrario, todas las secuencias de nucleótidos determinadas mediante la secuenciación de una molécula de ADN en la presente se determinaron empleando un secuenciador de ADN automático (tal como el modelo 373 de Applied Biosystems, Inc.), y todas las secuencias de aminoácidos de los polipéptidos codificados por las moléculas de ADN determinadas en la presente se determinaron empleando un secuenciador de péptidos automático. Tal como se conoce en la técnica para cualquier secuencia de ADN determinada mediante esta estrategia automática, cualquier secuencia de nucleótidos determinada en la presente puede contener algunos errores. Las secuencias de nucleótidos determinadas de modo automático generalmente son al menos aproximadamente 90% idénticas, más generalmente de al menos aproximadamente 95% hasta al menos aproximadamente 99,9% idénticas a la secuencia de nucleótidos real de la molécula de ADN secuenciada. La secuencia real puede determinarse de modo más preciso mediante otras estrategias, que incluyen métodos de secuenciación de ADN manuales muy conocidos en la técnica. Como también se conoce en la técnica, una única inserción o delección en una secuencia de nucleótidos determinada, comparada con la secuencia real, provocará un desplazamiento de marco en la traducción de la secuencia de nucleótidos, de modo que la secuencia de aminoácidos predicha codificada por una secuencia de nucleótidos determinada será completamente diferente de la secuencia de aminoácidos realmente codificada por la molécula de ADN secuenciada, comenzando en el punto de dicha inserción o delección.

25 A menos que se indique lo contrario, cada secuencia de nucleótidos indicada en la presente se presenta como una secuencia de desoxirribonucleótidos (abreviados como A, G, C y T). Sin embargo, una "secuencia de nucleótidos" de una molécula de ácido nucleico o polinucleótido significa, para una molécula o polinucleótido de ADN, una secuencia de desoxirribonucleótidos, y para una molécula o polinucleótido de ARN, la correspondiente secuencia de ribonucleótidos (A, G, C y U), en la que cada desoxirribonucleótido de timidina (T) en la secuencia de desoxirribonucleótidos especificada está reemplazada por el ribonucleótido uridina (U). Por ejemplo, con referencia a una molécula de ARN que tiene la secuencia de SEQ ID NO:2 o 4, o indicada empleando las abreviaturas de desoxirribonucleótidos, pretende indicar una molécula de ARN que tiene una secuencia en la que cada desoxirribonucleótido A, G o C de SEQ ID NO:2 o 4 ha sido reemplazado por el correspondiente ribonucleótido A, G o C, y cada desoxirribonucleótido T ha sido reemplazado por un ribonucleótido U.

La descripción proporciona un polinucleótido aislado que comprende una secuencia de nucleótidos al menos 95% idéntica a una secuencia seleccionada del grupo que consiste en:

35 (a) una secuencia de nucleótidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos de SEQ ID NO:1 o SEQ ID NO:3;

(b) una secuencia de nucleótidos que codifica un polipéptido de fusión de SLC34A2-ROS, comprendiendo dicha secuencia de nucleótidos la secuencia de nucleótidos de SEQ ID NO:2 o SEQ ID NO:4;

40 (c) una secuencia de nucleótidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos N-terminal de SLC34A2 (restos 1-126 de SEQ ID NO:5) y el dominio quinasa de ROS (restos 1945-2222 de SEQ ID NO:7);

(d) una secuencia de nucleótidos que comprende la secuencia de nucleótidos N-terminal de SLC34A2 (restos 1-378 de SEQ ID NO:6) y la secuencia de nucleótidos del dominio quinasa de ROS (restos 6032-6865 de SEQ ID NO:8);

45 (e) una secuencia de nucleótidos que comprende al menos seis nucleótidos contiguos que incluyen la zona de unión de la fusión (restos 376-381 de SEQ ID NO:2 o restos 376-381 de SEQ ID NO:4) de un polinucleótido de fusión de SLC34A2-ROS;

(f) una secuencia de nucleótidos que codifica un polipéptido que comprende al menos seis nucleótidos contiguos que incluyen la zona de unión de la fusión (restos 126-127 de SEQ ID NO:1 o restos 126-127 de SEQ ID NO:3) de un polipéptido de fusión de SLC34A2-ROS; y

(g) una secuencia de nucleótidos complementaria con cualquiera de las secuencias de nucleótidos de (a)-(f).

50 Empleando la información proporcionada en la presente, tal como las secuencias de nucleótidos en las figuras 2A-B (SEQ ID NO:2 y 4), puede obtenerse una molécula de ácido nucleico de la presente invención que codifica un polipéptido de ROS mutante de la invención, empleando procedimientos de clonación y selección convencionales, tales como los utilizados para clonar ADNc empleando ARNm como material de partida. Como ejemplo de la invención, los polinucleótidos descritos en las figuras 2A-2B (SEQ ID NO:2 y 4) se aislaron a partir de ADN genómico procedente de una línea de células NSCLC humanas (tal como se describe más en fondo en el siguiente



ejemplo 4). El gen de fusión también puede identificarse en bancos de ADN genómico o de ADNc en otros cánceres o carcinomas pulmonares en los que se produce la translocación de SLC34A2-ROS (4p15, 6q22), o en los que una delección u otra translocación diferente resulta en la expresión de una ROS quinasa truncada que carece del dominio extracelular de la quinasa de tipo salvaje.

5 La secuencia de nucleótidos determinada de los productos del gen de translocación de SLC34A2-ROS (SEQ ID NO:2 y SEQ ID NO:4) codifica dos variantes de la proteína de fusión de quinasa (largo y corto) de 724 aminoácidos (véase la figura 2A (SEQ ID NO:1) y la figura 1) y de 621 aminoácidos (véase la figura 2B (SEQ ID NO:3) y la figura 1), respectivamente. Los polinucleótidos de fusión de SLC34A2-ROS comprenden la porción de la secuencia de nucleótidos de SLC34A2 de tipo salvaje (véase la figura 3 (SEQ ID NO:6)) que codifica el N-terminal de esta  
10 proteína (exones 1-4) con la porción de la secuencia de nucleótidos de ROS de tipo salvaje (véase la figura 4 (SEQ ID NO:8) que codifica los dominios transmembrana y quinasa de esta proteína (exones 32-43 o exones 34-43, respectivamente) (véase la figura 1). El dominio quinasa comprende los restos 322-599 en el primer variante de la proteína de fusión (largo) (codificado por los nucleótidos 964-1797 del polinucleótido del primer variante de fusión) y los restos 219-496 en el segundo variante de la proteína de fusión (corto) (codificado por los nucleótidos 655-1488  
15 del polinucleótido del segundo variante de fusión).

Tal como se indicó, la presente invención proporciona, en parte, la forma madura de las proteínas de fusión de SLC34A2-ROS. Según la hipótesis de la señal, las proteínas segregadas por células de mamífero tienen una secuencia conductora secretora o señal que se rompe y escinde de la proteína madura tras haberse iniciado la exportación de la cadena de proteína en crecimiento a través del retículo endoplásmico rugoso. La mayoría de las  
20 células de mamífero, e incluso células de insecto, rompen las proteínas segregadas con la misma especificidad. Sin embargo, en algunos casos, la ruptura de la proteína segregada no es totalmente uniforme, lo cual produce dos o más especies maduras de la proteína. Además, es sabido que la especificidad de ruptura de una proteína segregada viene determinada, en último término, por la estructura primaria de la proteína completa, es decir, es inherente a la secuencia de aminoácidos del polipéptido. Por tanto, la presente invención proporciona, en parte, secuencias de  
25 nucleótidos que codifican un polipéptido de fusión de SLC34A2-ROS maduro cuya secuencia de aminoácidos es codificada por el clon de ADNc identificado como ATCC n.º de depósito PTA-7877, que se depositó en the American Type Culture Collection (Manassas, Virginia, EEUU) el 20 de septiembre, 2006, según las disposiciones del Tratado de Budapest.

Un polipéptido de SLC34A2-ROS maduro cuya secuencia de aminoácidos es codificada por el clon de ADNc deposita significa la forma madura de esta proteína de fusión producida mediante la expresión en una célula de  
30 mamífero (por ejemplo, células COS, tal como se describe a continuación) del marco de lectura abierto completo codificado por la secuencia de ADN humano del clon contenida en el vector en la célula hospedante depositada.

Tal como se ha indicado, los polinucleótidos de la presente invención pueden estar en forma de ARN, tal como ARNm, o en forma de ADN, que incluye, por ejemplo, ADNc y ADN genómico obtenido por clonación o producido de  
35 modo sintético. El ADN puede ser bicatenario o monocatenario. El ARN o ADN monocatenario puede ser la hebra codificadora, también conocida como hebra sentido, o puede ser la hebra no codificadora, también conocida como hebra antisentido.

Los polinucleótidos aislados de la invención son moléculas de ácidos nucleicos, ADN o ARN, que han sido retiradas de su entorno nativo. Por ejemplo, las moléculas de ADN recombinante contenidas en un vector se consideran  
40 aisladas para los fines de la presente invención. Otros ejemplos de moléculas de ADN aisladas incluyen moléculas de ADN recombinante mantenidas en células hospedantes heterólogas o moléculas de ADN purificadas (parcial o sustancialmente) en disolución. Las moléculas de ARN aisladas incluyen transcripciones de ARN *in vivo* o *in vitro* de las moléculas de ADN de la presente invención. Las moléculas de ácidos nucleicos aisladas según la presente invención incluyen también las moléculas producidas de modo sintético.

45 Los polinucleótidos aislados de la descripción incluyen las moléculas de ADN que aparecen en la figura 2A-B (SEQ ID NO:2 y 4), las moléculas de ADN que comprenden la secuencia codificadora de las proteínas de fusión de SLC34A2-ROS maduras que aparecen en la figura 1 (SEQ ID NO:1 y 3), y las moléculas de ADN que comprenden una secuencia sustancialmente diferente de las descritas anteriormente, pero que, debido a la degeneración del código genético, siguen siendo un polipéptido de ROS mutante de la invención. El código genético es muy conocido  
50 en la técnica y, así, los expertos en la técnica están acostumbrados a generar estos variantes degenerados.

En otra realización, la invención proporciona un polinucleótido aislado que codifica el polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de nucleótidos de la translocación SLC34A2-ROS contenida en el clon de ADNc depositado descrito anteriormente. Preferiblemente, esta molécula de ácido nucleico codificará el  
55 polipéptido de fusión maduro codificado por el clon de ADNc depositado. En otra realización, la invención proporciona una secuencia de nucleótidos aislada que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos N-terminal de SLC34A2 (restos 1-126 de SEQ ID NO:5) y el dominio quinasa de ROS (restos 1945-2222 de SEQ ID NO:7). En un ejemplo, el polipéptido que comprende el dominio quinasa de ROS comprende los restos 1750-2347 o 1853-2347 de SEQ ID NO:7 (véase la figura 1, panel B). En otro ejemplo, la secuencia de aminoácidos N-terminal mencionada anteriormente de SLC34A2 y el dominio quinasa de  
60 ROS son codificados por las secuencias de nucleótidos que comprenden los nucleótidos 1-378 de SEQ ID NO:6 y

los nucleótidos 6032-6865 de SEQ ID NO:8, respectivamente.

La invención también proporciona polinucleótidos aislados que comprenden secuencias de nucleótidos que tienen una secuencia complementaria con uno de los polipéptidos de fusión de ROS mutantes de la invención. Estas moléculas aisladas, en particular moléculas de ADN, son útiles como sondas para el cartografiado de genes, para la hibridación *in situ* con cromosomas, y para detectar la expresión de la proteína de fusión de SLC34A2-ROS o el polipéptido de ROS quinasa truncado en tejido humano, por ejemplo, mediante un análisis de la transferencia Northern.

La presente descripción también se dirige a fragmentos de las moléculas de ácidos nucleicos aisladas descritas en la presente. Un fragmento de un polinucleótido de SLC34A2-ROS o de un polinucleótido de ROS truncado aislado de la invención significa fragmentos de al menos aproximadamente 15 nucleótidos, y más preferiblemente de al menos aproximadamente 20 nucleótidos, aún más preferiblemente de al menos aproximadamente 30 nucleótidos y aún más preferiblemente de al menos aproximadamente 40 nucleótidos de longitud, que son útiles como cebadores y sondas de diagnóstico, según se analiza en la presente. Por supuesto, también son útiles en la presente invención fragmentos más largos de aproximadamente 50-1500 nucleótidos de longitud, así como los fragmentos que se corresponden con la mayor parte, sino toda, de la secuencia de nucleótidos de SLC34A2-ROS del ADNc depositado o según se muestra en la figura 2A-B (SEQ ID NO:2 y 4). Un fragmento de al menos 20 nucleótidos de longitud, por ejemplo, son fragmentos que incluyen 20 o más bases contiguas procedentes de las respectivas secuencias de nucleótidos de las que se derivan los fragmentos.

La generación de dichos fragmentos de ADN es un procedimiento habitual para los expertos en la técnica, y puede realizarse, por ejemplo, mediante ruptura por endonucleasas de restricción o cizallamiento por sonicación de ADN que puede obtenerse del clon de ADNc depositado o que puede sintetizarse según la secuencia descrita en la presente. Como alternativa, estos fragmentos pueden generarse directamente de modo sintético.

Los fragmentos de ácidos nucleicos o sondas preferidos de la presente invención incluyen moléculas de ácidos nucleicos que codifican la zona de unión de la fusión de los productos del gen de translocación SLC34A2-ROS (véase la figura 1, paneles B y C, y la figura 7, panel inferior). Por ejemplo, en ciertas realizaciones preferidas, un polinucleótido aislado de la invención comprende una secuencia de nucleótidos/fragmento que comprende al menos seis nucleótidos contiguos que incluyen la zona de unión de la fusión (restos 376-381 de SEQ ID NO:2 o restos 376-381 de SEQ ID NO:4) de un polinucleótido de fusión de SLC34A2-ROS (véase la figura 7, panel inferior). En otra realización preferida, un polinucleótido aislado de la invención comprende una secuencia de nucleótidos/fragmento que codifica un polipéptido que comprende al menos seis aminoácidos contiguos que incluyen la zona de unión de la fusión (restos 126-127 de SEQ ID NO:1 o restos 126-127 de SEQ ID NO:3) de un polipéptido de fusión de SLC34A2-ROS (véase también la figura 7, panel inferior (SEQ ID NO:9 y 11)).

En otro aspecto, la invención proporciona un polinucleótido aislado que se hibrida bajo condiciones de hibridación rigurosas con una porción de un polinucleótido de ROS quinasa mutante de la invención, según se describe en la presente. Unas "condiciones de hibridación rigurosas" significan una incubación durante la noche a 42 °C en una disolución que comprende formamida al 50%, 5 x SSC (NaCl 750 mM, citrato de trisodio 75 mM), fosfato de sodio 50 mM (pH 7,6), 5 x disolución de Denhardt, sulfato de dextrano al 10%, y 20 microgramos/ml de ADN de esperma de salmón cizallado desnaturalizado, seguido del lavado de los filtros en 0,1 x SSC a aproximadamente 65 °C.

Un polinucleótido que se hibrida con una "porción" de un polinucleótido es un polinucleótido (ADN o ARN) que se hibrida con al menos aproximadamente 15 nucleótidos (nt), y más preferiblemente con al menos aproximadamente 20 nt, aún más preferiblemente con al menos aproximadamente 30 nt, y aún más preferiblemente con aproximadamente 30-70 nt del polinucleótido de referencia. Estos son útiles como cebadores y sondas de diagnóstico (por ejemplo, para PCR), según se analizó anteriormente y se analizará con más detalle a continuación.

Por supuesto, los polinucleótidos que se hibridan con una porción mayor del polinucleótido de referencia (por ejemplo, los polinucleótidos de fusión de SLC34A2-ROS maduros descritos en la figura 2A-B (SEQ ID NO:2 y 4)), por ejemplo, con una porción de 50-750 nt de longitud, o incluso con la longitud completa del polinucleótido de referencia, también son útiles como sondas según la presente invención, así como los polinucleótidos que se corresponden con la mayor parte, sino toda, de la secuencia de nucleótidos del ADNc depositado o de las secuencias de nucleótidos mostradas en la figura 2A-2B (SEQ ID NO:2 y 4) o la figura 7 (panel inferior) (SEQ ID NO:10 y 12).

Una porción de un polinucleótido de "al menos 20 nucleótidos de longitud", por ejemplo, significa 20 o más nucleótidos contiguos procedentes de la secuencia de nucleótidos del polinucleótido de referencia. Tal como se indicó, estas porciones son útiles desde el punto de vista del diagnóstico como una sonda para su uso según las técnicas de hibridación de ADN convencionales, o como cebadores para la amplificación de una secuencia diana mediante una reacción en cadena de polimerasa (PCR), según se describe, por ejemplo, en MOLECULAR CLONING, A LABORATORY MANUAL, 2ª edición, Sambrook, J., Fritsch, E. F. y Maniatis, T., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Por supuesto, un polinucleótido que se hibrida solo con una secuencia de poli-A (tal como el tramo de poli(A) 3'-terminal de las secuencias de SLC34A2-ROS mostradas en la figura 2 (SEQ ID NO: 2 o 4)) o con un tramo complementario de restos T (o U), no se incluye en los

polinucleótidos según la invención empleados para hibridarse con una porción de un ácido nucleico de la invención, puesto que dicho polinucleótido se hibridará con cualquier molécula de ácido nucleico que contenga un tramo de poli(A) o con su complemento (por ejemplo, prácticamente cualquier clon de ADNc bicatenario).

5 Tal como se indicó, las moléculas de ácidos nucleicos de la presente descripción, que codifican un polipéptido de ROS quinasa mutante de la invención, pueden incluir las que codifican la secuencia de aminoácidos del polipéptido maduro por sí solas; la secuencia codificadora del polipéptido maduro y secuencias adicionales, tales como las que codifican la secuencia conductora o secretora, tales como una secuencia de pre-, pro- o pre-pro-proteína; la secuencia codificadora del polipéptido maduro, con o sin las secuencias codificadoras adicionales mencionadas anteriormente, junto con secuencias no codificadoras adicionales que incluyen, por ejemplo, intrones y secuencias no codificadoras 5' y 3', tales como las secuencias transcritas no traducidas que desempeñan un papel en la transcripción, el procesamiento de ARNm, que incluyen señales de corte y empalme y de poliadenilación, por ejemplo, la unión a ribosomas y la estabilidad de ARNm; una secuencia codificadora adicional que codifica aminoácidos adicionales, tales como las que proporcionan funcionalidades adicionales.

15 Así, la secuencia que codifica el polipéptido puede estar condensada con una secuencia marcadora, tal como una secuencia que codifica un péptido que facilita la purificación del polipéptido condensado. En ciertas realizaciones preferidas de este aspecto de la invención, la secuencia de aminoácidos marcadora es un péptido de hexa-histidina, tal como el marcador proporcionado en un vector pQE (Qiagen, Inc.), entre otras, muchas de las cuales están disponibles en el mercado. Tal como se describe en Gentz *et al.*, *Proc. Natl. Acad. Sci. USA*, 86: 821-824 (1989), por ejemplo, la hexa-histidina proporciona una purificación conveniente de la proteína de fusión. El marcador "HA" es otro péptido útil para la purificación, que se corresponde con un epitopo derivado de la proteína de la hemaglutinina de la gripe, que se ha descrito en Wilson *et al.*, *Cell*, 37: 767 (1984). Tal como se analiza a continuación, otras de estas proteínas de fusión incluyen el propio polipéptido de fusión de SLC34A2-ROS condensado con Fc en el N- o C-terminal.

25 La presente invención se refiere además a variantes de las moléculas de ácidos nucleicos de la presente invención, que codifican porciones, análogos o derivados de un polipéptido de fusión de SLC34A2-ROS o del polipéptido de ROS quinasa truncado descritos en la presente. Los variantes pueden aparecer en la naturaleza, tales como un variante alélico natural. Un "variante alélico" significa una de varias formas alternativas de un gen que ocupa un locus concreto sobre un cromosoma de un organismo. Véase, por ejemplo, GENES II, Lewin, B., ed., John Wiley & Sons, Nueva York (1985). Pueden producirse variantes que no aparecen en la naturaleza empleando técnicas de mutagénesis conocidas en la técnica.

30 Estos variantes incluyen los producidos mediante sustituciones, deleciones o adiciones de nucleótidos. Las sustituciones, deleciones o adiciones pueden implicar a uno o más nucleótidos. Los variantes pueden estar alterados en regiones codificadoras, en regiones no codificadoras o en ambas. Las alteraciones en las regiones codificadoras pueden producir sustituciones, deleciones o adiciones de aminoácidos conservativas o no conservativas. Entre estas, se prefieren especialmente las sustituciones, adiciones y deleciones silenciosas, que no alteran las propiedades y actividades (por ejemplo, actividad quinasa) de los polipéptidos de ROS quinasa mutantes descritos en la presente. También se prefieren especialmente a este respecto las sustituciones conservativas.

35 La descripción incluye polinucleótidos aislados que comprenden una secuencia de nucleótidos que es al menos 90% idéntica, y más preferiblemente al menos 95%, 96%, 97%, 98% o 99% idéntica a un polinucleótido de ROS mutante de la invención (por ejemplo, una secuencia de nucleótidos que codifica el polipéptido de fusión de RB-ROS que tiene la secuencia de aminoácidos completa que aparece en la figura 2A-B (SEQ ID NO:1 o 3); o una secuencia de nucleótidos que codifica el N-terminal de SLC34A2 y el dominio quinasa de ROS (véase la figura 1, panel B; y las figuras 3 y 4); o un nucleótido complementario a dichos ejemplos de secuencias).

40 Un polinucleótido que tiene una secuencia de nucleótidos que es al menos, por ejemplo, 95% "idéntica" a una secuencia de nucleótidos de referencia que codifica un polipéptido de ROS quinasa mutante significa que la secuencia de nucleótidos del polinucleótido es idéntica a la secuencia de referencia, excepto que la secuencia de polinucleótidos puede incluir hasta cinco mutaciones puntuales por cada 100 nucleótidos de la secuencia de nucleótidos de referencia que codifica el polipéptido ROS mutante. En otras palabras, para obtener un polinucleótido que tenga una secuencia de nucleótidos que es al menos 95% idéntica a una secuencia de nucleótidos de referencia, hasta 5% de los nucleótidos en la secuencia de referencia pueden deleccionarse o sustituirse por otro nucleótido, o una serie de nucleótidos de hasta 5% de los nucleótidos totales en la secuencia de referencia pueden insertarse en la secuencia de referencia. Estas mutaciones de la secuencia de referencia pueden producirse en las posiciones 5'- o 3'-terminales de la secuencia de nucleótidos de referencia o en cualquier punto entre estas posiciones terminales, dispersas de forma individual entre los nucleótidos en la secuencia de referencia, o en uno o más grupo contiguos dentro de la secuencia de referencia.

45 Desde un punto de vista práctico, cualquier molécula de ácido nucleico concreta que sea al menos 90%, 95%, 96%, 97%, 98% o 99% idéntica a, por ejemplo, las secuencias de nucleótidos que aparecen en la figura 2A-B (SEQ ID NO:2 o 4) o a la secuencia de nucleótidos del clon de ADNc depositado descrito anteriormente, puede determinarse de modo convencional empleando programas informáticos conocidos, tales como el programa Bestfit (Wisconsin Sequence Analysis Package, Versión 8 para Unix, Genetics Computer Group, University Research Park, 575

Science Drive, Madison, Wis. 53711). El programa Bestfit emplea el algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics*, 2: 482-489 (1981), para descubrir el mejor segmento de homología entre dos secuencias. Cuando se emplea Bestfit o cualquier otro programa de alineamiento de secuencias para determinar si una secuencia concreta es, por ejemplo, 95% idéntica a una secuencia de un polinucleótido de fusión de SLC34A2-ROS de referencia según la presente invención, los parámetros se ajustan, por supuesto, de modo que el porcentaje de identidad se calcula a lo largo de la longitud completa de la secuencia de nucleótidos de referencia y se permitan huecos en la homología de hasta 5% del número total de nucleótidos en la secuencia de referencia.

La presente descripción incluye, en su alcance, moléculas de ácidos nucleicos que son al menos 90%, 95%, 96%, 97%, 98% o 99% idénticas a las secuencias de ácidos nucleicos que aparecen en la figura 2A-2B (SEQ ID NO:2 o 4), o a los nucleótidos 379-2172 de SEQ ID NO:2 o los nucleótidos 379-1863 de SEQ ID NO:4, o a la secuencia de ácido nucleico del ADNc depositado, independientemente de que codifiquen un polipéptido que tenga actividad ROS quinasa. Esto es porque, aunque una molécula de ácido nucleico concreta no codifique un polipéptido de fusión que tenga actividad ROS quinasa, los expertos en la técnica todavía sabrán cómo emplear la molécula de ácido nucleico, por ejemplo, como una sonda de hibridación o como cebador de la reacción en cadena de la polimerasa (PCR). Los usos de las moléculas de ácido nucleico de la presente invención que no codifican un polipéptido que tenga actividad quinasa incluyen, entre otros, (1) aislar el gen de la translocación de SLC34A2-ROS o sus variantes alélicas en un banco de ADNc; (2) la hibridación *in situ* (por ejemplo, "FISH") con un frotil de cromosomas en metafase para proporcionar la localización cromosómica precisa del gen de la translocación de SLC34A2-ROS, según se describe en Verma *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press, Nueva York (1988); y análisis de la transferencia Northern para detectar la expresión de ARNm de la proteína de fusión de SLC34A2-ROS en tejidos específicos.

Sin embargo, se prefieren las moléculas de ácidos nucleicos que tienen secuencias que son al menos 95% idénticas a una polipéptido de ROS quinasa mutante de la invención, o a la secuencia de ácido nucleico del ADNc depositado, que, de hecho, codifican un polipéptido de fusión que tiene actividad ROS quinasa. Esta actividad puede ser similar, pero no necesariamente idéntica, a la actividad de la proteína de fusión de SLC34A2-ROS descrita en la presente (tanto la proteína de longitud completa como la proteína madura o un fragmento de la proteína que conserve la actividad quinasa), según se mide en un ensayo biológico concreto. Por ejemplo, la actividad quinasa de ROS puede estudiarse determinando su capacidad para fosforilar uno o más sustratos peptídicos que contienen tirosina, por ejemplo, el "péptido relacionado con Src" (RRLIEDAEYAARG), que es un sustrato para muchas tirosina quinasa de receptor y no receptor.

Debido a la degeneración del código genético, los expertos en la técnica reconocerán inmediatamente que un gran número de moléculas de ácidos nucleicos que tienen una secuencia que es al menos 90%, 95%, 96%, 97%, 98% o 99% idéntica a la secuencia de ácido nucleico del ADNc depositado o a la secuencia de ácido nucleico mostrada en la figura 2A-B (SEQ ID NO:2 y 4), codificarán un polipéptido de fusión que tiene actividad ROS quinasa. De hecho, puesto que los variantes degenerados de estas secuencias de nucleótidos codifican todos el mismo polipéptido, esto será evidente para los expertos en la técnica, incluso sin realizar el ensayo de comparación descrito anteriormente. También se reconoce en la técnica que, para las moléculas de ácidos nucleicos que no son variantes degenerados, un número razonable de estas también codificarán un polipéptido que conserve la actividad ROS quinasa. Esto es debido a que los expertos en la técnica conocen perfectamente las sustituciones de aminoácidos que es menos probable que afecten, o que en absoluto afecten significativamente a la función de la proteína (por ejemplo, sustituir un aminoácido alifático por un segundo aminoácido alifático).

Por ejemplo, se proporcionan directrices acerca de cómo conseguir sustituciones de aminoácidos fenotípicamente silenciosas en Bowie *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science*, 247: 1306-1310 (1990), que describe dos estrategias principales para estudiar la tolerancia de una secuencia de aminoácidos al cambio. El primer método se basa en el proceso de la evolución, en el que las mutaciones son aceptadas o rechazadas por la selección natural. La segunda estrategia emplea la ingeniería genética para introducir cambios en aminoácidos en posiciones específicas de un gen clonado, y selecciones para identificar las secuencias que conservan la funcionalidad. Estos estudios han revelado que las proteínas son sorprendentemente tolerantes a las sustituciones de aminoácidos. Los expertos en la técnica que están familiarizados con estas técnicas también conocerán cuáles son los cambios de aminoácidos que es probable que puedan realizarse en una posición concreta de la proteína. Por ejemplo, la mayoría de los aminoácidos enterrados requieren cadenas lateales no polares, mientras que pocas características de las cadenas laterales de la superficie en general se conservan. Otras de estas sustituciones fenotípicamente silenciosas se describen en Bowie *et al.*, *supra*, y en las referencias citadas allí.

Pueden utilizarse los métodos para la secuenciación de ADN que son muy conocidos y que en general están disponibles en la técnica para implementar cualquier realización de polinucleótidos de la invención. Los métodos pueden emplear enzimas, tales como el fragmento Klenow de la ADN polimerasa I, SEQUENASE® (US Biochemical Corp, Cleveland, Ohio), Taq polimerasa (Perkin Elmer), polimerasa T7 termoestable (Amersham, Chicago, Ill.), o combinaciones de polimerasas recombinantes y exonucleasas correctoras, tales como el sistema de amplificación ELONGASE, comercializado por Gibco BRL (Gaithersburg, Md.). Preferiblemente, el proceso se automatiza con máquinas tales como Hamilton Micro Lab 2200 (Hamilton, Reno, Nev.), termociclador Peltier (PTC200; MJ Research,

Watertown, Mass.) y los secuenciadores de ADN ABI 377 (Perkin Elmer).

5 Las secuencias de polinucleótidos que codifican un polipéptido de ROS mutante de la invención pueden extenderse empleando una secuencia de nucleótidos parcial y empleando diversos métodos conocidos en la técnica para detectar secuencias cadena arriba, tales como promotores y elementos reguladores. Por ejemplo, un método que puede emplearse, la PCR "de sitio de restricción", emplea cebadores universales para recuperar una secuencia desconocida adyacente a un locus conocido (Sarkar, G., *PCR Methods Applic.*, 2: 318-322 (1993)). En particular, primero se amplifica el ADN genómico en presencia de un cebador para una secuencia conectora y un cebador específico de la región conocida. Los ejemplos de cebadores son los descritos en el ejemplo 4 de la presente. Las secuencias amplificadas después se someten a una segunda ronda de PCR con el mismo cebador de conector y otro cebador específico interno al primero. Los productos de cada ronda de PCR se transcriben con una ARN polimerasa apropiada y se secuencian empleando una transcriptasa inversa.

10 También puede emplearse una PCR inversa para amplificar o extender secuencias empleando cebadores divergentes basados en una región conocida (Triglia *et al.*, *Nucleic Acids Res.*, 16: 8186 (1988)). Los cebadores pueden diseñarse empleando el software de análisis de cebadores OLIGO 4.06 (National Biosciences Inc., Plymouth, Minn.), u otro programa apropiado, para que tengan una longitud de 22-30 nucleótidos, para que tengan un contenido en GC del 50% o mayor, y para que se reasocian con la secuencia diana a unas temperaturas de aproximadamente 68-72 °C. El método emplea varias enzimas de restricción para generar un fragmento adecuado en la región conocida de un gen. El fragmento después se circulariza mediante acoplamiento intramolecular y se emplea como molde de PCR.

15 Otro método que puede utilizarse es la PCR de captura, que implica una amplificación por PCR de fragmentos de ADN adyacentes a una secuencia conocida en ADN cromosómico artificial de levadura y humano (Lagerstrom *et al.*, *PCR Methods Applic.*, 1: 111-119 (1991)). En este método, también pueden emplearse múltiples digestiones con enzimas de restricción y acoplamientos para colocar una secuencia bicatenaria modificada en una porción desconocida de la molécula de ADN antes de realizar la PCR. Otro método que puede utilizarse para recuperar secuencias desconocidas es el descrito en Parker *et al.*, *Nucleic Acids Res.*, 19: 3055-3060 (1991)). Además, se puede utilizar la PCR, cebadores anidados y bancos PROMOTERFINDER® para realizar un paseo en un ADN genómico (Clontech, Palo Alto, Calif.). Este proceso evita la necesidad de seleccionar bancos y es útil para descubrir zonas de unión de intrón/exón.

20 Cuando se seleccionan ADNc de longitud completa, resulta preferible emplear bancos que han sido seleccionados según el tamaño para que incluyan ADNc más largos. Además, son preferibles bancos cebados aleatoriamente, porque contendrán más secuencias que incluyan las regiones 5' de genes. El empleo de un banco cebado aleatoriamente puede resultar especialmente preferible en situaciones en las que un banco de oligo-d(T) no contenga un ADNc de longitud completa. Los bancos genómicos pueden ser útiles para la extensión de la secuencia hacia las regiones reguladoras no transcritas 5' y 3'.

25 Pueden emplearse sistemas de electroforesis capilar, que están disponibles en el mercado, para analizar el tamaño o confirmar la secuencia de nucleótidos de los productos de la secuenciación o de PCR. En particular, la secuenciación capilar puede emplear polímeros fluidos para la separación electroforética, cuatro tintes fluorescentes diferentes (uno para cada nucleótido) que son activados por láser, y la detección de las longitudes de onda emitidas mediante una cámara con dispositivo de carga acoplada. La intensidad de salida/de la luz puede convertirse en una señal eléctrica empleando un software apropiado (por ejemplo, GENOTYPER® y SEQUENCE NAVIGATOR™, Perkin Elmer), y el proceso completo desde la carga de las muestras al análisis por ordenador y la presentación de los datos electrónicos puede estar controlado por ordenador. La electroforesis capilar es especialmente preferible para la secuenciación de pequeños trozos de ADN que puedan estar presentes en cantidades limitadas en una muestra concreta.

#### 45 C. Vectores y células hospedantes

La presente invención también proporciona vectores recombinantes que comprenden un polinucleótido aislado de la presente invención, células hospedantes que están genéticamente modificadas con los vectores recombinantes, y la producción de polipéptidos de SLC34A2-ROS recombinantes o sus fragmentos por medio de técnicas recombinantes.

50 Las construcciones recombinantes pueden introducirse en las células hospedantes empleando técnicas muy conocidas, tales como infección, transducción, transfección, transvección, electroporación y transformación. El vector puede ser, por ejemplo, un fago, plásmido, vector vírico o retrovírico. Los vectores retrovíricos pueden ser competentes en la replicación o defectuosos en la replicación. En este último caso, la propagación vírica en general se producirá solo en las células hospedantes complementadoras.

55 Los polinucleótidos pueden unirse a un vector que contenga un marcador seleccionable para la propagación en un hospedante. En general, un vector plasmídico se introduce en un precipitado, tal como un precipitado de fosfato de calcio, o en un complejo con un lípido cargado. Si el vector es un virus, este puede encapsularse *in vitro* empleando una línea celular encapsulante apropiada, y después transducirse en las células hospedantes.

- Se prefieren los vectores que comprenden regiones de control de acción en cis para el polinucleótido de interés. Los factores de acción en trans apropiados pueden ser suministrados por el hospedante, ser suministrados por un vector complementador o ser suministrados por el propio vector tras su introducción en el hospedante. En ciertas realizaciones preferidas a este respecto, los vectores proporcionan la expresión específica, que puede ser inducible y/o específica del tipo celular. Entre estos vectores, se prefieren en particular los que son inducibles por factores ambientales que sean fáciles de manipular, tales como la temperatura y aditivos de nutrientes.
- Los vectores de expresión útiles en la presente invención incluyen vectores derivados de cromosomas, episomas y virus, por ejemplo, vectores derivados de plásmidos bacterianos, bacteriófagos, episomas de levaduras, elementos cromosómicos de levaduras, virus, tales como baculovirus, papovavirus, virus de vaccinia, adenovirus, virus de la viruela aviar, virus de la pseudorrabia, y retrovirus, y vectores derivados de sus combinaciones, tales como cósmidos y fágmidos.
- El inserto de ADN que comprende un polinucleótido de SLC34A2-ROS o un polinucleótido ROS truncado de la invención debe estar unido operativamente a un promotor apropiado, tal como el promotor PL del fago lambda, los promotores lac, trp y tac de *E. coli*, los promotores temprano y tardío de SV40, y promotores de LTR retrovíricos, por nombrar unos cuantos. Los expertos en la técnica conocen otros promotores adecuados. Las construcciones de expresión contendrán también sitios para el inicio y terminación de la transcripción y, en la región transcrita, un sitio de unión a ribosomas para la traducción. La porción codificadora de los transcritos maduros expresados por las construcciones incluirá preferiblemente un inicio de la traducción al principio y un codón de terminación (UAA, UGA o UAG) colocado de modo apropiado al final del polipéptido que se va a traducir.
- Tal como se ha indicado, los vectores de expresión incluirán preferiblemente al menos un marcador seleccionable. Estos marcadores incluyen la resistencia a la dihidrofolato reductasa o la neomicina para un cultivo de células eucariotas, y genes de resistencia a la tetraciclina o a la ampicilina para cultivar en *E. coli* y otras bacterias. Los ejemplos representativos de los hospedantes apropiados incluyen células bacterianas, tales como células de *E. coli*, *Streptomyces* y *Salmonella typhimurium*; células fúngicas, tales como células de levadura; células de insecto, tales como células de *Drosophila* S2 y *Spodoptera* Sf9; células animales, tales como células CHO, COS y de melanoma de Bowen; y células vegetales. Los medios de cultivo y las condiciones apropiados para las células hospedantes descritas anteriormente son conocidos en la técnica.
- Entre los vectores preferidos para su uso en las bacterias se incluyen pQE70, pQE60 y pQE-9, disponibles en Qiagen; los vectores pBS, los vectores Phagescript, los vectores Bluescript, pNH8A, pNH16a, pNH18A, pNH46A, disponibles en Stratagene; y ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 disponibles en Pharmacia. Entre los vectores eucariotas preferidos se incluyen pWLNEO, pSV2CAT, pOG44, pXT1 y pSG disponibles en Stratagene; y pSVK3, pBPV, pMSG y pSVL disponibles en Pharmacia. Para los expertos en la técnica serán evidentes otros vectores adecuados.
- Entre los promotores bacterianos conocidos adecuados para su uso en la presente invención se incluyen los promotores lacI y lacZ de *E. coli*, los promotores T3 y T7, el promotor gpt, los promotores lambda PR y PL, y el promotor trp. Los promotores eucariotas adecuados incluyen el promotor temprano inmediato de CMV, el promotor de timidina quinasa de HSV, los promotores temprano y tardío de SV40, los promotores de LTR retrovíricos, tales como los del virus del sarcoma de Rous (RSV), y promotores de metalotioneína, tal como el promotor de metalotioneína-I de ratón.
- En la levadura *Saccharomyces cerevisiae*, pueden emplearse una serie de vectores que contienen promotores constitutivos o inducibles, tales como factor alfa, alcohol oxidasa, y PGH. Para un análisis, véase Ausubel *et al.* (1989), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, Nueva York, N.Y., y Grant *et al.*, *Methods Enzymol.*, 153: 516-544 (1997).
- La introducción de la construcción en la célula hospedante puede realizarse mediante transfección con fosfato de calcio, transfección mediada por DEAE-dextrano, transfección mediada por lípidos catiónicos, electroporación, transducción, infección u otros métodos. Estos métodos se describen en muchos manuales de laboratorio convencionales, tales como Davis *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).
- La transcripción de ADN que codifica un polipéptido de fusión de SLC34A2-ROS de la presente invención por eucariotas superiores puede aumentar insertando una secuencia potenciadora en el vector. Los potenciadores son elementos de acción de cis del ADN, que generalmente tienen de aproximadamente 10 a 300 pb, que actúan para aumentar la actividad transcripcional de un promotor en un tipo de célula hospedante concreta. Los ejemplos de potenciadores incluyen el potenciador de SV40, que está localizado en el lado tardío del origen de la replicación en las pares de bases 100 a 270, el potenciador del promotor temprano de citomegalovirus, el potenciador de polioma en el lado tardío del origen de la replicación, y potenciadores adenovíricos.
- Para la secreción de la proteína traducida hacia el lumen del retículo endoplásmico, hacia el espacio periplásmico o hacia el entorno extracelular, pueden incorporarse señales de secreción apropiadas en el polipéptido expresado. Las señales pueden ser endógenas al polipéptido o pueden ser señales heterólogas.

El polipéptido puede expresarse en una forma modificada, tal como una proteína de fusión (por ejemplo, una fusión-GST), y puede incluir no solo señales de secreción, sino también regiones funcionales heterólogas adicionales. Por ejemplo, una región de aminoácidos adicionales, en particular aminoácidos cargados, puede añadirse al N-terminal del polipéptido para mejorar la estabilidad y la persistencia en la célula hospedante, durante la purificación, o durante la posterior manipulación y conservación. Además, pueden añadirse restos peptídicos al polipéptido para facilitar la purificación. Estas regiones pueden retirarse antes de la preparación final del polipéptido. La adición de restos peptídicos a los polipéptidos para suscitar la secreción o la excreción, para mejorar la estabilidad y para facilitar la purificación, entre otros, son técnicas familiares y rutinarias en la técnica. Una proteína de fusión preferida comprende una región heteróloga procedente de inmunoglobulinas que es útil para solubilizar proteínas.

Por ejemplo, el documento EP-A-O 464 533 (hómologo canadiense 2045869) describe proteínas de fusión que comprenden diversas porciones de la región constante de moléculas de inmunoglobulinas, junto con otra proteína humana o una de sus partes. En muchos casos, la parte Fc en una proteína de fusión es muy ventajosa para su uso en terapia y en el diagnóstico, y esto se traduce, por ejemplo, en mejores propiedades farmacocinéticas (documento EP-A 0 232 262). Por otra parte, para algunos usos sería deseable poder deleccionar la parte Fc después de que la proteína de fusión se haya expresado, detectado y purificado de la manera ventajosa descrita. Este es el caso en el que la porción Fc demuestra ser un impedimento para el uso en terapia y en el diagnóstico, por ejemplo, cuando la proteína de fusión se va a utilizar como antígeno para inmunizaciones. En el descubrimiento de fármacos, por ejemplo, proteínas humanas, hIL-5 se ha condensado con porciones Fc para obtener ensayos de selección de alta capacidad de procesamiento para identificar antagonistas de hIL-5. Véase Bennett *et al.*, *Journal of Molecular Recognition* 8: 52-58 (1995); y Johanson *et al.*, *The Journal of Biological Chemistry*, 270(16): 9459-9471 (1995).

Los polipéptidos de SLC34A2-ROS pueden recuperarse y purificarse a partir de cultivos de células recombinantes mediante métodos muy conocidos, que incluyen precipitación en sulfato de amonio o etanol, extracción con ácidos, cromatografía de intercambio aniónico o catiónico, cromatografía de fosfocelulosa, cromatografía de inerción hidrófoba, cromatografía de afinidad, cromatografía de hidroxilapatito y cromatografía de lectina. Lo más preferiblemente, se emplea una cromatografía líquida de alta resolución ("HPLC") para la purificación. Los polipéptidos de la presente invención incluyen productos purificados de forma natural, productos de procedimientos sintéticos químicos y productos producidos mediante técnicas recombinantes a partir de un hospedante procarionta o eucariota, que incluyen, por ejemplo, células bacterianas, de levadura, de plantas superiores, de insecto y de mamífero. Dependiendo del hospedante empleado en el procedimiento de producción recombinante, los polipéptidos de la presente descripción pueden estar glicosilados o no glicosilados. Además, los polipéptidos de la descripción también pueden incluir un resto metionina modificado inicial, en algunos casos como resultado de procesos mediados por el hospedante.

Por consiguiente, en una realización, la invención proporciona un método para producir un polipéptido de fusión de SLC34A2-ROS recombinante cultivando una célula hospedante recombinante (tal como se describe a continuación) bajo condiciones adecuadas para la expresión del polipéptido de fusión, y recuperando el polipéptido. Las condiciones de cultivo adecuadas para el crecimiento de las células hospedantes y la expresión de los polipéptidos recombinantes a partir de dichas células son muy conocidos por los expertos en la técnica. Véase, por ejemplo, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel F M *et al.*, eds., volumen 2, capítulo 16, Wiley Interscience.

#### 40 D. Polipéptidos aislados

La descripción también proporciona, en parte, polipéptidos de fusión de SLC34A2-ROS aislados y sus fragmentos. En un ejemplo, la invención proporciona un polipéptido aislado que comprende una secuencia de aminoácidos al menos 95% idéntica a una secuencia seleccionada del grupo que consiste en:

45 (a) una secuencia de aminoácidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos de SEQ ID NO:1 o SEQ ID NO:3;

(b) una secuencia de aminoácidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos N-terminal de SLC34A2 (restos 1-126 de SEQ ID NO:5) y el dominio quinasa de ROS (restos 1945-2222 de SEQ ID NO:7); y

50 (c) una secuencia de aminoácidos que codifica un polipéptido que comprende al menos seis aminoácidos contiguos que incluye la zona de unión de la fusión (restos 126-127 de SEQ ID NO:1 o restos 126-127 de SEQ ID NO:3) de un polipéptido de fusión de SLC34A2-ROS.

En una realización preferida, la invención proporciona un polipéptido de fusión de SLC34A2-ROS aislado que tiene la secuencia de aminoácidos codificada por el ADNc depositado descrito anteriormente (ATCC n.º de depósito PTA-7877). En otra realización preferida, se proporcionan polipéptidos mutantes recombinantes de la invención, que pueden producirse empleando un vector recombinante o una célula hospedante recombinante, tal como se describió anteriormente.

En la técnica se reconocerá que algunas secuencias de aminoácidos de un polipéptido de fusión de SLC34A2-ROS

pueden variar sin que esto tenga un efecto significativo en el estructura o la función de la proteína mutante. Si se contemplan dichas diferencias en la secuencia, no debe olvidarse que existen áreas críticas en la proteína que determinan la actividad (por ejemplo, el dominio quinasa de ROS). En general, es posible reemplazar restos que forman la estructura terciaria, con la condición de que se empleen restos que realicen una función similar. En otros casos, el tipo de resto puede no tener ninguna importancia si la alteración se produce en una región no crítica de la proteína.

Así, la descripción incluye además variaciones de un polipéptido de fusión de SLC34A2-ROS que muestran una actividad ROS quinasa sustancial o que incluyen regiones de las proteínas de SLC34A2 y ROS, tales como las porciones de proteínas analizadas a continuación. Estos mutantes incluyen delecciones, inserciones, inversiones, repeticiones y sustituciones de tipos (por ejemplo, se sustituye un resto hidrófilo por otro, pero no uno muy hidrófilo por otro muy hidrófilo, por regla general). En general, unos cambios pequeños o dichas sustituciones de aminoácidos "neutras" tendrán poco efecto sobre la actividad.

Consideradas generalmente como sustituciones conservativas son las sustituciones, de uno por otro, entre los aminoácidos alifáticos Ala, Val, Leu e Ile, el intercambio de los restos hidroxilo Ser y Thr, el intercambio de los restos ácidos Asp y Glu, la sustitución entre los restos amida Asn y Gln, el intercambio de los restos básicos Lys y Arg, y las sustituciones entre los restos aromáticos Phe y Tyr. Los ejemplos de sustituciones conservativas de aminoácidos conocidas en la técnica son: aromáticos: fenilalanina, triptófano, tirosina; hidrófobos: leucina, isoleucina, valina; polares: glutamina, asparagina; básicos: arginina, lisina, histidina; ácidos: ácido aspártico, ácido glutámico; pequeños: alanina, serina, treonina, metionina, glicina. Tal como se indicó en detalle anteriormente, pueden encontrarse más directrices con respecto a los cambios en aminoácidos que es probable que sean fenotípicamente silenciosos (es decir, no es probable que tengan un efecto prejudicial en la función) en Bowie *et al.*, Science, 247, *supra*.

Los polipéptidos de la presente invención se proporcionan preferiblemente en una forma aislada, y preferiblemente están sustancialmente purificados. Una versión producida de modo recombinante de un polipéptido de fusión de SLC34A2-ROS de la invención puede purificarse sustancialmente mediante el método de una etapa descrito en Smith y Johnson, *Gene*, 67: 31-40 (1988).

Los polipéptidos de la presente descripción incluyen los polipéptidos de fusión de SLC34A2-ROS de la figura 2A-B (SEQ ID NO:1 y 3) (que incluyen o no una secuencia conductora), el polipéptido de fusión codificado por el clon de ADNc depositado (ATCC n.º PTA-7877), una secuencia de aminoácidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos N-terminal de SLC34A2 (restos 1-126 de SEQ ID NO:5) y el dominio quinasa de ROS (restos 1945-2222 de SEQ ID NO:7), y una secuencia de aminoácidos que codifica un polipéptido que comprende al menos seis aminoácidos contiguos que incluyen la zona de unión de la fusión (restos 126-127 de SEQ ID NO:1 o restos 126-127 de SEQ ID NO:3) de un polipéptido de fusión de SLC34A2-ROS (véase también la figura 7, panel inferior), así como polipéptidos que tienen al menos 90% de similitud, más preferiblemente al menos 95% de similitud, y aún más preferiblemente al menos 96%, 97%, 98% o 99% de similitud con los descritos anteriormente.

El "porcentaje (%) de similitud" entre dos polipéptidos es una puntuación de similitud producida comparando las secuencias de aminoácidos de los dos polipéptidos empleando el programa Bestfit (Wisconsin Sequence Analysis Package, versión 8 para Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711) y los ajustes por defecto para determinar la similitud. Bestfit emplea el algoritmo de homología local de Smith y Waterman (*Advances in Applied Mathematics*, 2: 482-489 (1981)) para descubrir la mejor similitud entre segmentos de dos secuencias.

Un polipéptido que tiene una secuencia de aminoácidos que es al menos, por ejemplo, 95% "idéntica" a una secuencia de aminoácidos de referencia de un polipéptido de ROS mutante de la invención significa que la secuencia de aminoácidos del polipéptido es idéntica a la secuencia de referencia, excepto que la secuencia del polipéptido puede incluir hasta cinco alteraciones de aminoácidos por cada 100 aminoácidos de la secuencia de aminoácidos de referencia del polipéptido de fusión de SLC34A2-ROS. En otras palabras, para obtener un polipéptido que tiene una secuencia de aminoácidos al menos 95% idéntica a una secuencia de aminoácidos de referencia, hasta 5% de los restos aminoácidos en la secuencia de referencia pueden estar delecionados o sustituidos por otro aminoácido, o una serie de aminoácidos hasta 5% del total de restos aminoácidos en la secuencia de referencia pueden insertarse en la secuencia de referencia. Estas alteraciones en la secuencia de referencia pueden realizarse en las posiciones amino- o carboxi-terminales de la secuencia de aminoácidos de referencia o en cualquier lugar entre estas posiciones terminales, estar dispersas de modo individual entre los restos en la secuencia de referencia, o en uno o más grupos contiguos dentro de la secuencia de referencia.

Cuando se emplea el programa Bestfit o cualquier otro programa de alineamiento de secuencias para determinar si una secuencia concreta es, por ejemplo, 95% idéntica a una secuencia de referencia según la presente invención, los parámetros se ajustan, por supuesto, de modo que el porcentaje de identidad se calcula a lo largo de la longitud completa de la secuencia de aminoácidos de referencia y se permitan huecos en la homología de hasta 5% del número total de nucleótidos en la secuencia de referencia.



Un polipéptido de fusión de SLC34A2-ROS de la presente invención puede emplearse como marcador de peso molecular en geles SDS-PAGE o en columnas de filtración en gel de tamices moleculares, por ejemplo, empleando métodos muy conocidos por los expertos en la técnica.

5 Tal como se describe con más detalle a continuación, los polipéptidos de la presente invención también pueden emplearse para generar reactivos específicos de polipéptidos de fusión, tales como anticuerpos policlonales y monoclonales, que son útiles en ensayos para detectar la expresión del polipéptido de ROS mutante, según se describe a continuación, o como agonistas y antagonistas capaces de potenciar o inhibir la actividad/función de la proteína ROS mutante. Además, estos polipéptidos pueden emplearse en el sistema de dos híbridos de levaduras para "capturar" proteínas de unión al polipéptido de fusión de SLC34A2-ROS, que también son candidatos a  
10 agonistas y antagonistas según la presente invención. El sistema de dos híbridos de levaduras se describe en Fields y Song, *Nature*, 340: 245-246 (1989).

En otro aspecto, la descripción proporciona un péptido o un polipéptido que comprende una porción que porta un epitopo de un polipéptido de la invención, concretamente un epitopo que comprende la zona de unión de la fusión de un variante de polipéptido de fusión de SLC34A2-ROS (véase la figura 7, panel inferior). El epitopo de esta porción del polipéptido es un epitopo inmunogénico o antigénico de un polipéptido de la invención. Un "epitopo inmunogénico" se define como una parte de una proteína que suscita una respuesta de anticuerpos cuando la proteína completa es el inmunógeno. Se cree que estos epitopos inmunogénicos están limitados a unos pocos loci en la molécula. Por otra parte, una región de una molécula de proteína con la cual puede unirse un anticuerpo se define como "epitopo antigénico". El número de epitopos inmunogénicos de una proteína en general es menor que el  
15 número de epitopos antigénicos. Véase, por ejemplo, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:3998-4002 (1983). La producción de anticuerpos específicos del polipéptido de fusión de la invención se describe con más detalle a continuación.

Los anticuerpos generados por polipéptidos o péptidos que portan epitopos antigénicos son útiles para detectar una proteína imitada, y pueden emplearse anticuerpos contra diferentes péptidos para determinar el destino de diversas regiones de un precursor de la proteína que sufre un procesamiento postraducciona. Los péptidos y los anticuerpos antipéptido pueden emplearse en una diversidad de ensayos cualitativos o cuantitativos para la proteína imitada, por ejemplo, en ensayos de competición, puesto que se ha demostrado que incluso péptidos cortos (por ejemplo, de aproximadamente 9 aminoácidos) pueden unirse y desplazar a péptidos más grandes en ensayos de inmunoprecipitación. Véase, por ejemplo, Wilson *et al.*, *Cell*, 37: 767-778 (1984) en 777. Los anticuerpos antipéptido de la invención también son útiles para la purificación de la proteína imitada, por ejemplo, mediante cromatografía de adsorción empleando métodos conocidos en la técnica. Los formatos de ensayo inmunológico se describen con más detalle a continuación.

Los polipéptidos de ROS quinasa mutantes recombinantes también están dentro del alcance de la presente invención, y pueden producirse empleando polinucleótidos de fusión de la invención, según se describe en la anterior sección B. Por ejemplo, la invención proporciona un método para producir un polipéptido de fusión de SLC34A2-ROS recombinante cultivando una célula hospedante recombinante (tal como se describió anteriormente) bajo condiciones adecuadas para la expresión del polipéptido de fusión, y recuperando el polipéptido. Las condiciones de cultivo adecuadas para el crecimiento de las células hospedantes y la expresión de los polipéptidos recombinantes a partir de dichas células son muy conocidas por los expertos en la técnica.

#### 40 *E. Reactivos específicos de mutantes*

Los reactivos específicos del polipéptido de ROS mutante útiles en la práctica de los métodos descritos incluyen, entre otros, los anticuerpos específicos de polipéptidos de fusión y los péptidos AQUA (péptidos marcados con isótopos pesados) que se corresponden con la expresión del polipéptido de fusión de SLC34A2-ROS en una muestra biológica, y son adecuados para la detección y la cuantificación de esta. Un reactivo específico de un polipéptido de fusión es cualquier reactivo, producto biológico o químico, que es capaz de unirse específicamente, detectar y/o cuantificar la presencia/nivel del polipéptido de fusión de SLC34A2-ROS expresado en una muestra biológica. La expresión incluye el anticuerpo preferido y los reactivos de péptidos AQUA analizados a continuación, y otros reactivos equivalentes están dentro del alcance de la presente descripción.

#### Anticuerpos

50 Los reactivos adecuados para su uso en la práctica de los métodos de la descripción incluyen un anticuerpo específico del polipéptido de fusión de SLC34A2-ROS. Un anticuerpo específico de la fusión de la invención es un anticuerpo o anticuerpos aislados que se unen específicamente a un polipéptido de fusión de SLC34A2-ROS de la invención (por ejemplo, SEQ ID NO:1 o 3), pero que no se unen sustancialmente a SLC34A2 de tipo salvaje ni a ROS de tipo salvaje. Otros reactivos adecuados incluyen anticuerpos específicos de epitopos que se unen específicamente a un epitopo en el dominio extracelular de una secuencia de la proteína de ROS de tipo salvaje (cuyo dominio no está presente en la ROS quinasa truncada descrita en la presente) y, por tanto, son capaces de detectar la presencia (o la ausencia) de ROS de tipo salvaje en una muestra.

Los anticuerpos específicos del polipéptido de fusión de SLC34A2-ROS humano también pueden unirse a

secuencias de péptidos epitópicos equivalentes y muy homólogas en otras especies de mamífero, por ejemplo, murinas o de conejo, o viceversa. Los anticuerpos útiles para la práctica de los métodos de la descripción incluyen (a) anticuerpos monoclonales, (b) anticuerpos policlonales purificados que se unen específicamente al polipéptido diana (por ejemplo, la zona de unión de la fusión del polipéptido de fusión de SLC34A2-ROS (véase la figura 7, panel inferior), (c) los anticuerpos descritos en (a)-(b) anteriores que se unen a epitopos equivalentes y muy homólogos o a sitios de fosforilación en otras especies no humanas (por ejemplo, ratón, rata), y (d) fragmentos de (a)-(c) anteriores que se unen al antígeno (o más preferiblemente al epitopo) al cual se unen los ejemplos de anticuerpos descritos en la presente.

El término "anticuerpo" o "anticuerpos", tal como se emplea en la presente, se refiere a todos los tipos de inmunoglobulinas, que incluyen IgG, IgM, IgA, IgD, e IgE. Los anticuerpos pueden ser monoclonales o policlonales, y puede tener su origen en cualquier especie que incluye (por ejemplo) ratón, rata, conejo, caballo, o ser humano, y pueden ser anticuerpos quiméricos. Véase, por ejemplo, M. Walker *et al.*, *Molec. Immunol.*, 26: 403-411 (1989); Morrision *et al.*, *Proc. Nat'l. Acad. Sci.*, 81: 6851 (1984); Neuberger *et al.*, *Nature*, 312: 604 (1984)). Los anticuerpos pueden ser anticuerpos monoclonales recombinantes producidos según los métodos descritos en la patente de EEUU n.º 4.474.893 (Reading) o la patente de EEUU n.º 4.816.567 (Cabilly *et al.*). Los anticuerpos también pueden ser anticuerpos específicos construidos químicamente preparados según el método descrito en la patente de EEUU n.º 4.676.980 (Segel *et al.*).

El sitio epitópico preferido de un anticuerpo específico de un polipéptido de fusión de SLC34A2-ROS de la invención en un fragmento peptídico que consiste fundamentalmente en aproximadamente 11 a 17 aminoácidos de la secuencia de un polipéptido de fusión de SLC34A2-ROS humano (SEQ ID NO:1 y 3), incluyendo dicho fragmento la zona de unión de la fusión (que aparece en el resto 126 en el primer y segundo variante de la proteína de fusión (véase la figura 1 (panel C) y la figura 7 (panel inferior)). Se apreciará que los anticuerpos que se unen específicamente a epitopos/péptidos más cortos o más largos que incluyen la zona de unión de la fusión de un polipéptido de fusión de SLC34A2-ROS están dentro del alcance de la presente descripción.

La descripción incluye el uso de anticuerpos y moléculas equivalentes, tales como dominios de unión de proteínas o aptámeros de ácidos nucleicos, que se unen, de una manera específica de la proteína de fusión o de la proteína truncada, fundamentalmente al mismo epitopo al cual se une un anticuerpo específico de un polipéptido de fusión de SLC34A2-ROS o un anticuerpo específico del epitopo del punto de truncamiento de ROS útiles en los métodos de la invención. Véase, por ejemplo, Neuberger *et al.*, *Nature*, 312: 604 (1984). Estos reactivos equivalentes que no son anticuerpos pueden emplearse de modo adecuado en los métodos de la descripción, tal como se describe más a fondo a continuación.

Los anticuerpos policlonales útiles en la práctica de los métodos de la invención pueden producirse según técnicas convencionales mediante la inmunización de un animal adecuado (por ejemplo, conejo, cabra, etc.) con un antígeno que incluya un epitopo específico de la proteína de fusión deseada (por ejemplo, la zona de unión de la fusión (véase la figura 7, panel inferior), la recolección del suero inmunológico del animal, y la separación de los anticuerpos policlonales del suero inmunológico, y la purificación de los anticuerpos policlonales que tienen la especificidad deseada, según procedimientos conocidos. El antígeno puede ser un antígeno peptídico sintético que comprenda la secuencia epitópica deseada, seleccionado y construido según técnicas muy conocidas. Véase, por ejemplo, ANTIBODIES: A LABORATORY MANUAL, capítulo 5, pp. 75-76, Harlow & Lane eds., Cold Spring Harbor Laboratory (1988); Czernik, *Methods In Enzymology*, 201: 264-283 (1991); Merrifield, *J. Am. Chem. Soc.*, 85: 21-49 (1962)). Los anticuerpos policlonales producidos según se describe en la presente pueden seleccionarse y aislarse como se describe más a fondo a continuación.

Los anticuerpos monoclonales también pueden emplearse de modo beneficioso en los métodos de la invención, y pueden producirse en líneas celulares de hibridoma según la técnica conocida de Kohler y Milstein, *Nature* 265: 495-97 (1975); Kohler y Milstein, *Eur. J. Immunol.*, 6: 511 (1976); véase también CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel *et al.* eds. (1989). Los anticuerpos monoclonales producidos de esta manera son muy específicos, y mejoran la selectividad y la especificidad de los métodos de ensayo proporcionados por la invención. Por ejemplo, una disolución que contiene el antígeno apropiado (por ejemplo, un péptido sintético que comprenda la zona de unión de la fusión de un polipéptido de fusión de SLC34A2-ROS) puede inyectarse en un ratón y, después de un tiempo suficiente (según las técnicas convencionales), el ratón se sacrifica y se obtienen las células esplénicas. Las células esplénicas entonces se immortalizan mediante su fusión con células de mieloma, generalmente en presencia de polietilenglicol, para producir células de hibridoma. Pueden producirse hibridomas de fusión de conejo, por ejemplo, según se describe en la patente de EEUU n.º 5.675.063, C. Knight, otorgada el 7 de octubre de 1997. Las células de hibridoma entonces se cultivan en un medio de selección adecuado, tal como hipoxantina-aminopterina-timidina (HAT), y el sobrenadante se selecciona para detectar los anticuerpos monoclonales que tienen la especificidad deseada, tal como se describe a continuación. El anticuerpo segregado puede recuperarse del sobrenadante del cultivo de tejido mediante métodos convencionales, tales como precipitación, intercambio iónico o cromatografía de afinidad, o similares.

También pueden producirse fragmentos Fab monoclonales en *Escherichia coli* mediante técnicas recombinantes conocidas por los expertos en la técnica. Véase, por ejemplo, W. Huse, *Science*, 246: 1275-1281 (1989); Mullinax *et al.*, *Proc. Nat'l Acad. Sci.*, 87: 8095 (1990). Si se prefieren anticuerpos monoclonales de un isotipo para una

aplicación concreta, los isotipos concretos pueden prepararse directamente, mediante la selección a partir de la fusión inicial, o prepararse de modo secundario, a partir de un hibridoma parental que segregue un anticuerpo monoclonal de un isotipo diferente emplando la técnica de selección sib para aislar variantes de clase de conmutación (Steplewski, *et al.*, *Proc. Nat'l. Acad. Sci.*, 82: 8653 (1985); Spira *et al.*, *J. Immunol. Methods*, 74: 307 (1984)). El sitio de combinación con el antígeno del anticuerpo monoclonal puede clonarse mediante PCR y producirse anticuerpos monocatenarios como anticuerpos recombinantes presentados sobre fagos o anticuerpos solubles en *E. coli* (véase, por ejemplo, ANTIBODY ENGINEERING PROTOCOLS, 1995, Humana Press, Sudhir Paul editor).

Además, la patente de EEUU n.º 5.194.392, Geysen (1990), describe un método general para detectar o determinar la secuencia de monómeros (aminoácidos y otros compuestos) que es un equivalente topológico del epitopo (es decir, un "mimotopo"), que es complementario con un paratopo concreto (sitio de unión al antígeno) de un anticuerpo de interés. Más en general, este método implica detectar o determinar una secuencia de monómeros que es un equivalente topográfico de un ligando que es complementario con el sitio de unión al ligando de un receptor concreto de interés. De forma similar, la patente de EEUU n.º 5.480.971, Houghten *et al.* (1996), describe oligopéptidos peralquilados de alquilo C<sub>1</sub>-C lineal y conjuntos y bancos de dichos péptidos, así como métodos para emplear dichos conjuntos y bancos de oligopéptidos para determinar la secuencia de un oligopéptido peralquilado que se une preferentemente con una molécula aceptora de interés. Así, mediante estos métodos también pueden prepararse, de forma rutinaria, análogos no peptídicos de los péptidos que portan epitopos de la descripción.

Los anticuerpos útiles en los métodos de la invención, tanto policlonales como monoclonales, pueden seleccionarse para la especificidad de la proteína de fusión y el epitopo según técnicas convencionales. Véase, por ejemplo, Czernik *et al.*, *Methods in Enzymology*, 201: 264-283 (1991). Por ejemplo, los anticuerpos pueden seleccionarse frente a un banco de péptidos mediante ELISA para asegurar la especificidad por el antígeno deseado y, si se quiere, para la reactividad solo con un polipéptido de fusión de SLC34A2-ROS de la invención y no con SLC34A2 de tipo salvaje ni con ROS de tipo salvaje. Los anticuerpos también pueden ensayarse mediante un análisis de la transferencia Western frente a preparaciones de células que contienen la proteína diana para confirmar la reactividad solo con la diana deseada y para asegurarse de que no se produce una unión apreciable con otras proteínas de fusión que impliquen a ROS. La producción, la selección y el uso de anticuerpos específicos de proteínas de fusión son conocidos por los expertos en la técnica, y se han descrito. Véase, por ejemplo, la publicación de patente de EEUU 20050214301, Wetzel *et al.*, 29 de septiembre de 2005.

Los anticuerpos específicos de polipéptidos de fusión útiles en los métodos de la invención pueden mostrar cierta reactividad cruzada limitada con epitopos de fusión similares en otras proteínas de fusión o con epitopos en SLC34A2 de tipo salvaje y ROS de tipo salvaje que forman la zona de unión de la fusión. Esto no resulta inesperado, puesto que la mayoría de los anticuerpos muestra algún grado de reactividad cruzada, y los anticuerpos antipéptidos a menudo tienen reactividad cruzada con epitopos que tienen una alta homología o identidad con el péptido inmunizante. Véase, por ejemplo, Czernik, *supra*. La reactividad cruzada con otras proteínas de fusión se caracteriza con facilidad mediante un análisis de la transferencia Western junto con marcadores de peso molecular conocido. Las secuencias de aminoácidos de las proteínas de reacción cruzada pueden estudiarse para identificar los sitios muy homólogos o idénticos a la secuencia del polipéptido de fusión de SLC34A2-ROS a la cual se une el anticuerpo. La reactividad cruzada no deseada puede eliminarse mediante selección negativa empleando la purificación de anticuerpos sobre columnas de péptidos (por ejemplo, eliminando por selección los anticuerpos que se unen a SLC34A2 de tipo salvaje y/o ROS de tipo salvaje).

Los anticuerpos específicos del polipéptido de fusión de SLC34A2-ROS de la descripción que son útiles para practicar los métodos descritos en la presente son específicos, de manera ideal, para el polipéptido de fusión humano, pero no se limitan solo a la unión con la especie humana, *per se*. La descripción incluye la producción y el uso de anticuerpos que también se unen a epitopos conservados y altamente homólogos o idénticos en otras especies de mamífero (por ejemplo, ratón, rata, mono). Las secuencias muy homólogas o idénticas en otras especies pueden identificarse con facilidad mediante comparaciones de secuencias convencionales, tales como el empleo de BLAST, con las secuencias del polipéptido de fusión de SLC34A2-ROS humano descrito en la presente (SEQ ID NO:1 y 3).

Los anticuerpos empleados en los métodos de la invención también pueden caracterizarse y validarse para su uso en un formato de ensayo concreto, por ejemplo, FC, IHC y/o ICC. El uso de anticuerpos específicos del polipéptido de fusión de SLC34A2-ROS en estos métodos se describe más a fondo en la siguiente sección F. Los anticuerpos también pueden conjugarse de modo ventajoso con tintes fluorescentes (por ejemplo, Alexa488, PE), o con marcadores tales como puntos cuánticos, para su uso en análisis multiparamétricos junto con otros anticuerpos de transducción de señales (fosfo-AKT, fosfo-Erk 1/2) y/o de marcador celular (citoqueratina), según se describe más a fondo en la siguiente sección F.

En la práctica de los métodos de la descripción, la expresión y/o la actividad de SLC34A2 de tipo salvaje y/o ROS de tipo salvaje en una muestra biológica concreta también puede estudiarse de modo ventajoso empleando anticuerpos (fosfoespecíficos o totales) para estas proteínas de tipo salvaje. Por ejemplo, están disponibles en el mercado anticuerpos específicos del sitio de fosforilación del receptor CSF (véase CELL SIGNALING TECHNOLOGY, INC., Beverly Mass., 2005/06 n.º de catálogo 3151, 3155, y 3154; y Upstate Biotechnology, 2006, n.º de catálogo 06-457).

Estos anticuerpos también pueden ser producidos según métodos convencionales, tal como se describió anteriormente. Las secuencias de aminoácidos de SLC34A2 y ROS humanas están publicadas (véanse las figuras 3 y 4, y los n.<sup>os</sup> de registro de SwissProt mencionados), así como las secuencias de estas proteínas procedentes de otras especies.

5 La detección de la expresión y/o activación de SLC34A2 de tipo salvaje y de ROS de tipo salvaje, junto con la expresión del polipéptido de fusión de SLC34A2-ROS, en una muestra biológica (por ejemplo, una muestra de tumor), puede proporcionar información sobre si la proteína de fusión está dirigiendo por sí sola al tumor, o si ROS de tipo salvaje también está activada y está dirigiendo al tumor. Esta información es útil desde el punto de vista clínico para evaluar si se deben tomar medidas contra la proteína de fusión o contra la proteína o proteínas de tipo salvaje o todas, o si es probable que sea más beneficioso inhibir el avance del tumor, y para seleccionar un producto terapéutico apropiado o una de sus combinaciones. Los anticuerpos específicos para el dominio extracelular de ROS quinasa de tipo salvaje, que no están presentes en la ROS quinasa truncada descrita en la presente, pueden ser particularmente útiles para determinar la presencia/ausencia de la ROS quinasa mutante.

15 Se entenderá que puede emplearse más de un anticuerpo en la práctica de los métodos descritos anteriormente. Por ejemplo, uno o más anticuerpos específicos del polipéptido de fusión de SLC34A2-ROS, junto con uno o más anticuerpos específicos para otra quinasa, receptor o sustrato de quinasa del cual se sospecha que está, o que potencialmente está, activado en un cáncer en el que el polipéptido de fusión de SLC34A2-ROS se expresa, pueden emplearse simultáneamente para detectar la actividad de estas otras moléculas de señalización en una muestra biológica que comprenda células de dicho cáncer.

20 Los expertos en la técnica apreciarán que los polipéptidos de fusión de SLC34A2-ROS de la presente invención y sus fragmentos que portan el epitopo de la zona de unión de la fusión descritos anteriormente pueden combinarse con partes del dominio constante de inmunoglobulinas (IgG), para producir polipéptidos quiméricos. Estas proteínas de fusión facilitan la purificación y muestran una mayor semivida *in vivo*. Esto se ha demostrado, por ejemplo, para proteínas quiméricas que consisten en los primeros dos dominios del polipéptido de CD4 humano y diversos dominios de las regiones constantes de las cadenas pesada o ligera de inmunoglobulinas humanas (documento 394.827; Trauneker *et al.*, *Nature*, 331: 84-86 (1988)). Las proteínas de fusión que tienen una estructura dimerica unida por disulfuro, debido a la parte IgG, también pueden ser más eficaces para la unión y la neutralización de otras moléculas distintas al polipéptido de fusión de SLC34A2-ROS monomérico por sí solo (Fountoulakis *et al.*, *J. Biochem.*, 270: 3958-3964 (1995)).

### 30 Péptidos marcados con isótopos pesados (péptidos AQUA)

Los reactivos específicos del polipéptido de fusión de SLC34A2-ROS útiles en la práctica de los métodos descritos también pueden comprender péptidos marcados con isótopos pesados adecuados para la cuantificación absoluta del polipéptido de fusión de SLC34A2-ROS expresado en una muestra biológica. Se ha descrito la producción y el uso de péptidos AQUA para la cuantificación absoluta de proteína (AQUA) en mezclas complejas. Véase el documento WO/03016861, "Absolute Quantification of Proteins and Modified Forms Thereof by Multistage Mass Spectrometry," Gygi *et al.*, y también Gerber *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 100: 6940-6945 (2003).

La metodología AQUA emplea la introducción de una cantidad conocida de al menos un patrón de péptido marcado con un isótopo pesado (que tienen una firma exclusiva que puede detectarse mediante cromatografía LC-SRM) en una muestra biológica digerida para determinar, por comparación con el patrón peptídico, la cantidad absoluta de un péptido con la misma secuencia y modificación en la proteína en la muestra biológica. Brevemente, la metodología AQUA tiene dos etapas: la selección y validación de patrones internos peptídicos, y el desarrollo del método; y la puesta en práctica empleando los patrones internos peptídicos validados para detectar y cuantificar una proteína diana en una muestra. El método es una poderosa técnica para detectar y cuantificar una proteína/péptido concreto dentro de una mezcla biológica compleja, tal como un lisado de células, y puede emplearse, por ejemplo, para cuantificar el cambio en la fosforilación de proteínas como resultado de un tratamiento con fármacos, o para cuantificar las diferencias en el nivel de una proteína en diferentes estados biológicos.

En general, para desarrollar un patrón interno adecuado, se elige un péptido (o un péptido modificado) concreto dentro de una secuencia de proteína diana, basándose en su secuencia de aminoácidos y en la proteasa concreta que se va a utilizar para la digestión. Después se genera el péptido mediante una síntesis peptídica en fase sólida, de forma que un resto es reemplazado por el mismo resto que contiene isótopos estables (<sup>13</sup>C, <sup>15</sup>N). El resultado es un péptido que es químicamente idéntico a su homólogo nativo formado mediante proteólisis, pero que puede distinguirse con facilidad mediante MS a través del desplazamiento de masas de 7-Da. El péptido de patrón interno AQUA recién sintetizado después se evalúa mediante LC-MS/MS. Este proceso proporciona información cualitativa acerca de la retención del péptido mediante una cromatografía en fase inversa, la eficacia de ionización, y la fragmentación a través de disociación inducida por colisión. Se eligen los iones de los fragmentos informativos y abundantes para conjuntos de péptidos de patrón interno y nativos, y después se controlan de modo específico en una sucesión rápida como una función de la retención cromatográfica para formar un método de control de la reacción seleccionada (LC-SRM) basado en el perfil exclusivo del patrón peptídico.

La segunda etapa de la estrategia AQUA es su puesta en práctica para medir la cantidad de una proteína o de una

5 proteína modificada en mezclas complejas. Lisados de células completas se fraccionan generalmente mediante una electroforesis en gel de SDS-PAGE, y se cortan y extraen las regiones del gel donde se produce la migración de las proteínas. A este proceso le sigue una proteólisis dentro del gel en presencia de los péptidos AQUA y un análisis de LC-SRM (véase Gerber *et al.*, *supra*). Los péptidos AQUA se siembran en la mezcla compleja de péptidos obtenida mediante la digestión del lisado de células completas con una enzima proteolítica, y se someten a una purificación de inmunoafinidad según se describió anteriormente. El tiempo de retención y el patrón de fragmentación del péptido nativo formado mediante digestión (por ejemplo, tripsinización) son idénticos a los del péptido de patrón interno AQUA determinados previamente; por tanto, un análisis de LC-MS/MS empleando un experimento SRM produce una medición muy específica y sensible del patrón interno y del analito directamente a partir de mezclas extremadamente complejas de péptidos.

15 Puesto que se añade una cantidad absoluta del péptido AQUA (por ejemplo, 250 fmol), puede utilizarse la proporción de las áreas bajo la curva para determinar los niveles de expresión precisos de una proteína o una forma fosforilada de una proteína en el lisado de células original. Además, el patrón interno está presente durante la digestión dentro del gel a medida que se forman los péptidos nativos, de modo que la eficacia de extracción del péptido de los trozos del gel, las pérdidas absolutas durante la manipulación de la muestra (que incluye una centrifugación al vacío), y la variabilidad durante la introducción en el sistema de LC-MS no afectan a la proporción determinada de abundancias de los péptidos nativos y AQUA.

20 Se desarrolla un patrón de péptido AQUA para una secuencia conocida previamente identificada mediante el método IAP-LC-MS/MS dentro de una proteína diana. Si el sitio está modificado, puede desarrollarse un péptido AQUA que incorpore la forma modificada del resto concreto dentro del sitio, y puede desarrollarse un segundo péptido AQUA que incorpore la forma no modificada del resto. De esta forma, los dos patrones pueden emplearse para detectar y cuantificar las formas modificadas y no modificadas del sitio en una muestra biológica.

25 También pueden generarse patrones internos de péptidos estudiando la secuencia de aminoácidos primaria de una proteína y determinando los límites de los péptidos producidos mediante ruptura con proteasa. Como alternativa, una proteína puede ser realmente digerida con una proteasa y después puede secuenciarse un fragmento peptídico concreto producido. Las proteasas adecuadas incluyen serina proteasas (por ejemplo, tripsina, hepsina), metaloproteasas (por ejemplo, PUMP1), quimotripsina, catepsina, pepsina, termolisina, carboxipeptidasas, etc.

30 Se selecciona una secuencia peptídica dentro de una proteína diana según uno o más criterios para optimizar el uso del péptido como patrón interno. Preferiblemente, el tamaño del péptido se selecciona para minimizar las posibilidades de que la secuencia peptídica esté repetida en otro lugar en otras proteínas no diana. Así, un péptido tiene preferiblemente al menos aproximadamente 6 aminoácidos. El tamaño del péptido también se optimiza para maximizar la frecuencia de ionización. Así, no se prefieren péptidos más largos que aproximadamente 20 aminoácidos. El intervalo preferido es de aproximadamente 7 a 15 aminoácidos. También se selecciona una secuencia peptídica que sea poco probable que sea químicamente reactiva durante la espectrometría de masas, y así se evitan las secuencias que comprenden cisteína, triptófano o metionina.

35 Una secuencia peptídica que no incluye una región modificada de la región diana puede seleccionarse de modo que el patrón interno de péptido pueda emplearse para determinar la cantidad de todas las formas de la proteína. Como alternativa, puede resultar deseable un patrón interno de péptido que incluya un aminoácido modificado para detectar y cuantificar solo la forma modificada de la proteína diana. Los patrones de péptidos para ambas regiones modifica y no modificada pueden emplearse juntos, para determinar el grado de una modificación en una muestra concreta (es decir, para determinar cuál es la fracción de la cantidad total de proteína que representa a la forma modificada). Por ejemplo, pueden emplearse patrones de péptidos para la forma fosforilada y no fosforilada de una proteína de la cual se sabe que está fosforilada en un sitio concreto, para cuantificar la cantidad de forma fosforilada en una muestra.

45 El péptido se marca empleando uno o más aminoácidos marcados (es decir, el marcador es una parte real del péptido) o, menos preferiblemente, los marcadores pueden unirse después de la síntesis según métodos convencionales. Preferiblemente, el marcador es un marcador que altera la masa seleccionado que se basa en las siguientes consideraciones: la masa debe ser exclusiva para desplazar las masas de los fragmentos producidos mediante el análisis de MS hasta regiones del espectro con bajo fondo; el componente de firma de masa iónica es la porción del resto marcador que muestra preferiblemente una firma de masa iónica exclusiva en un análisis de MS; la suma de las masas de los átomos constituyentes del marcador es preferible y exclusivamente diferente a la de los fragmentos de todos los aminoácidos posibles. Como resultado, los péptidos y aminoácidos marcados pueden distinguirse con facilidad de los no marcados mediante el patrón de iones/masa en el espectro de masas resultante. Preferiblemente, el componente de firma de masa iónica imparte una masa a un fragmento de la proteína que no se corresponde con la masa del resto para ninguno de los 20 aminoácidos naturales.

60 El marcador debe ser robusto bajo las condiciones de fragmentación de MS y no sufrir una fragmentación desfavorable. La química del marcaje debe ser eficaz bajo una gama de condiciones, en particular condiciones desnaturalizantes, y el marcador marcado preferiblemente permanece soluble en el sistema tampón de MS elegido. El marcador preferiblemente no reprime la eficacia de ionización de la proteína y es químicamente no reactivo. El marcador puede contener una mezcla de dos o más especies isotópicamente diferenciadas para generar un patrón

espectrométrico de masas exclusivo en cada posición del fragmento marcado. Los isótopos estables, tales como  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{17}\text{O}$ ,  $^{18}\text{O}$ , o  $^{34}\text{S}$ , están entre los marcadores preferidos. También pueden prepararse parejas de patrones internos de péptidos que incorporen un marcador de isótopo diferente. Los restos aminoácidos preferidos en los que pueden incorporarse un marcador de isótopo pesado incluyen leucina, prolina, valina y fenilalanina.

- 5 Los patrones internos de péptidos se caracterizan según su proporción de masa a carga ( $m/z$ ), y preferiblemente, también según su tiempo de retención en una columna cromatográfica (por ejemplo, una columna de HPLC). Los patrones internos que coeluyen con los péptidos no marcados de secuencia idéntica se seleccionan como patrones internos óptimos. El patrón interno después se analiza fragmentando el péptido mediante cualquier medio adecuado, por ejemplo, mediante disociación inducida por colisión (CID) empleando, por ejemplo, argón o helio como gas de colisión. Después los fragmentos se analizan, por ejemplo, mediante espectrometría de masas de múltiples etapas ( $\text{MS}^n$ ) para obtener un espectro de iones de fragmentos, para obtener una firma de fragmentación del péptido. Preferiblemente, los fragmentos del péptido presentan diferencias significativas en las proporciones de  $m/z$  para permitir que los picos que se corresponden con cada fragmente estén bien separado, y se obtiene una firma que es exclusiva del péptido diana. Si no se obtiene una firma de los fragmentos adecuada en la primera etapa, se realizan más etapas de MS hasta que se obtiene una firma exclusiva.

Los iones de fragmentos en los espectros de  $\text{MS}/\text{MS}$  y  $\text{MS}^3$  generalmente son muy específicos para el péptido de interés, y, junto con los métodos de LC, permiten un medio muy selectivo para detectar y cuantificar un péptido/proteína diana en una mezcla compleja de proteínas, tal como un lisado celular, que contenga muchos miles o decenas de miles de proteínas. Puede ensayarse cualquier muestra biológica que contenga potencialmente un péptido/proteína diana de interés. Se emplean preferiblemente extractos celulares brutos o parcialmente purificados. En general, la muestra tendrá al menos 0,01 mg de proteína, generalmente una concentración de 0,1-10 mg/ml, y puede ajustarse al pH y la concentración del tampón deseado.

Entonces se añade una cantidad conocida de un patrón interno de péptido marcado, preferiblemente aproximadamente 10 femtomoles, que se corresponde con una proteína diana que se va a detectar/cuantificar, a una muestra biológica, tal como un lisado de células. Después la muestra sembrada se digiere con una o más proteasas durante un periodo de tiempo adecuado para permitir la digestión. Entonces se realiza una separación (por ejemplo, mediante HPLC, HPLC de fase inversa, electroforesis capilar, cromatografía de intercambio iónico, etc.) para aislar el patrón interno marcado y su correspondiente péptido diana de los otros péptidos en la muestra. La LC microcapilar es un método preferido.

Cada péptido aislado después se estudia controlando una reacción seleccionada en el MS. Esto implica emplear el conocimiento previo obtenido a través de la caracterización del patrón interno de péptido y después se requiere del MS para controlar de modo continuo un ion específico en el espectro de  $\text{MS}/\text{MS}$  o  $\text{MS}^n$  para el péptido de interés y el patrón interno. Después de la elución, se calcula el área bajo la curva (AUC) para el patrón de péptido y el péptido diana. La proporción de las dos áreas proporciona la cuantificación absoluta que puede normalizarse para el número de células empleado en el análisis y el peso molecular de la proteína, para proporcionar el número de copias preciso de la proteína por células. Otros detalles de la metodología AQUA se describen en Gygi *et al.*, y Gerber *et al.*, *supra*.

De modo deseable, pueden producirse patrones de péptidos internos AQUA (péptidos marcados con isótopos pesados), según se describió anteriormente, para detectar y cuantificar cualquier sitio exclusivo (por ejemplo, la zona de unión de la fusión dentro de un polipéptido de fusión de SLC34A2-ROS) dentro de un polipéptido de ROS mutante de la invención. Por ejemplo, puede prepararse un fosfopéptido AQUA que se corresponda con la secuencia de la zona de unión de la fusión del polipéptido de fusión de SLC34A2-ROS (véase la figura 7 (panel inferior)). Pueden producirse patrones de péptidos para la zona de unión de la fusión de SLC34A2-ROS, y estos patrones pueden emplearse en la metodología AQUA para detectar y cuantificar la zona de unión de la fusión (es decir, la presencia del polipéptido de fusión de SLC34A2-ROS) en una muestra biológica.

Por ejemplo, un ejemplo de péptido AQUA de la descripción comprende la secuencia de aminoácidos LVGDDF (véase la figura 7, panel inferior), que se corresponde con los tres aminoácidos inmediatamente flanqueantes a cada lado de la zona de unión de la fusión en el segundo variante (corto) del polipéptido de fusión de SLC34A2-ROS (véase SEQ ID NO:11). Se apreciará que los péptidos AQUA más grandes que comprenden la secuencia de la zona de unión de la fusión (y otros restos cadena abajo o cadena arriba de esta) también pueden construirse. De modo similar, un péptido AQUA más pequeño que comprenda menos de todos los restos de dicha secuencia (pero que todavía comprenda el punto de la zona de unión de la fusión en sí mismo) puede construirse de modo alternativo. La selección y la producción de los péptidos AQUA preferidos puede realizarse como se describió anteriormente (véase Gygi *et al.*, y Gerber *et al.*, *supra*).

#### Sondas de ácidos nucleicos

55 Los reactivos específicos de la fusión proporcionados por la invención también incluyen cebadores y sondas de ácidos nucleicos adecuados para la detección de un polinucleótido de SLC34A2-ROS, según se describió en detalle en la anterior sección B. El uso específico de dichas sondas en ensayos, tales como la hibridación in situ de fluorescencia (FISH) o la amplificación por PCR, se describe en la siguiente sección F.

La invención también proporciona un kit para la detección de un polinucleótido y/o polipéptido de fusión de SLC34A2-ROS en una muestra biológica, comprendiendo dicho kit al menos un reactivo específico del polinucleótido o polipéptido de fusión de la invención, y uno o más reactivos secundarios. Los reactivos secundarios adecuados para su empleo en un kit son conocidos por los expertos en la técnica e incluyen, como ejemplo, tampones, sondas o anticuerpos secundarios detectables, quinasas, agentes activantes, sustratos de quinasa y similares.

#### F. Aplicaciones de diagnóstico y formatos de ensayo

Los métodos de la invención pueden realizarse en una diversidad de formatos de ensayo diferentes conocidos por los expertos en la técnica.

#### Inmunoensayos

Los inmunoensayos útiles en la práctica de los métodos de la descripción pueden ser inmunoensayos homogéneos o inmunoensayos heterogéneos. En un ensayo homogéneo, la reacción inmunológica habitualmente implica a un reactivo específico de un polipéptido de ROS mutante (por ejemplo, un anticuerpo específico del polipéptido de fusión de SLC34A2-ROS), un analito marcado, y la muestra biológica de interés. La señal que surge del marcador se modifica, directa o indirectamente, tras la unión del anticuerpo al analito marcado. Tanto la reacción inmunológica como la detección del grado de esta se realizan en una disolución homogénea. Los marcadores inmunoquímicos que pueden emplearse incluyen radicales libres, radioisótopos, tintes fluorescentes, enzimas, bacteriófagos, coenzimas, etc. También pueden emplearse de modo ventajoso marcadores de nanocristales semiconductores, o "puntos cuánticos", y su preparación y uso se ha descrito muy bien. Véase, en general, K. Barovsky, *Nanotech. Law & Bus*, 1(2): artículo 14 (2004) y las patentes citadas allí.

En la estrategia del ensayo heterogéneo, los reactivos son habitualmente la muestra biológica, un reactivo específico del polipéptido de ROS quinasa mutante (por ejemplo, un anticuerpo), y un medio adecuado para producir una señal detectable. Pueden emplearse las muestras biológicas descritas con más detalle a continuación. El anticuerpo en general se inmoviliza sobre un soporte, tal como una esfera, una placa o un portaobjetos, y se pone en contacto con la muestra de la cual se sospecha que contiene el antígeno en una fase líquida. El soporte después se separa de la fase líquida, y la fase de soporte o la fase líquida se estudia para una señal detectable empleando medios para producir dicha señal. La señal se relaciona con la presencia del analito en la muestra biológica. Los medios para producir una señal detectable incluyen el uso de marcadores radiactivos, marcadores fluorescentes, marcadores enzimáticos, puntos cuánticos, etc. Por ejemplo, si el antígeno que se va a detectar contiene un segundo sitio de unión, un anticuerpo que se una a ese sitio puede conjugarse con un grupo detectable y añadirse a la disolución de reacción en fase líquida antes de la etapa de separación. La presencia del grupo detectable sobre el soporte sólido indica la presencia del antígeno en la muestra de ensayo. Los ejemplos de inmunoensayos adecuados son un radioinmunoensayo, métodos de inmunofluorescencia, inmunoensayos ligados a enzimas, y similares.

Los formatos de inmunoensayos y sus variaciones que pueden ser útiles para realizar los métodos descritos en la presente son muy conocidos en la técnica. Véase, en general, E. Maggio, *Enzyme-Immunoassay* (1980) (CRC Press, Inc., Boca Raton, Fla.); véase también, por ejemplo, la patente de EEUU n.º 4.727.022 (Skold *et al.*, "Methods for Modulating Ligand-Receptor Interactions and their Application"); la patente de EEUU n.º 4.659.678 (Forrest *et al.*, "Immunoassay of Antigens"); la patente de EEUU n.º 4.376.110 (David *et al.*, "Immunometric Assays Using Monoclonal Antibodies"). Las condiciones adecuadas para la formación de complejos de reactivo-anticuerpo son muy conocidas por los expertos en la técnica. Véase *id.* Los anticuerpos monoclonales específicos del polipéptido de fusión de SLC34A2-ROS pueden emplearse en un ensayo de "dos sitios" o "sandwich", con una única línea celular de hibridoma como fuente para el anticuerpo monoclonal marcado y el anticuerpo monoclonal unido. Estos ensayos se describen en la patente de EEUU n.º 4.376.110. La concentración del reactivo detectable debe ser suficiente para que la unión del polipéptido de fusión de SLC34A2-ROS sea detectable, comparado con el fondo.

Los anticuerpos útiles en la práctica de los métodos descritos en la presente pueden conjugarse con un soporte sólido adecuado para un ensayo de diagnóstico (por ejemplo, esferas, placas, portaobjetos o pocillos formados por materiales tales como látex o poliestireno) según técnicas conocidas, tales como la precipitación. De forma similar, los anticuerpos u otros reactivos de unión al polipéptido de fusión de SLC34A2-ROS pueden conjugarse con grupos detectables, tales como radiomarcadores (por ejemplo, <sup>35</sup>S, <sup>125</sup>I, <sup>131</sup>I), marcadores enzimáticos (por ejemplo, peroxidasa de rábano, fosfatasa alcalina), y marcadores fluorescentes (por ejemplo, fluoresceína) según técnicas conocidas.

Los ensayos basados en células, tales como la citometría de flujo (FC), la inmunohistoquímica (IHC) o la inmunofluorescencia (IF) son particularmente deseables en la práctica de los métodos de la invención, puesto que dichos formatos de ensayo son adecuados desde el punto de vista clínico, permiten la detección de la expresión del polipéptido de ROS mutante *in vivo*, y evitan el riesgo de cambios accidentales en la actividad que resulten de la manipulación de células obtenidas, por ejemplo, de una muestra de tumor para obtener extractos. Por consiguiente, en alguna realización preferida, los métodos de la invención se ponen en práctica en un formato de ensayo de citometría de flujo (FC), inmunohistoquímica (IHC) o inmunofluorescencia (IF).

La citometría de flujo (FC) puede emplearse para determinar la expresión de un polipéptido de ROS mutante en un

tumor de mamífero antes, durante y después del tratamiento con un fármaco dirigido a la inhibición de la actividad ROS quinasa. Por ejemplo, pueden analizarse células tumorales procedentes de un aspirado con una aguja fina mediante citometría de flujo para la expresión y/o activación del polipéptido de fusión de SLC34A2-ROS, así como para marcadores que identifiquen los tipos de células cancerosas, etc., si así se desea. La citometría de flujo puede realizarse según métodos convencionales. Véase, por ejemplo, Chow *et al.*, *Cytometry (Communications in Clinical Cytometry)*, 46: 72-78 (2001). Brevemente y como ejemplo, puede emplearse el siguiente protocolo para el análisis citométrico: fijación de las células con paraformaldehído al 2% durante 10 minutos a 37 °C, seguido de permeabilización en metanol al 90% durante 30 minutos en hielo. Las células después pueden teñirse con el anticuerpo específico del polipéptido de fusión de SLC34A2-ROS primario, lavarse y marcarse con un anticuerpo secundario marcado con fluorescencia. Las células después se analizarán en un citómetro de flujo (por ejemplo, un Beckman Coulter FC500) según los protocolos específicos del instrumento empleado. Este análisis identificará el nivel del polipéptido de fusión de SLC34A2-ROS expresado en el tumor. Análisis similares después del tratamiento del tumor con un producto terapéutico que inhibe a ROS revelarán el grado de respuesta de un tumor que expresa el polipéptido de fusión de SLC34A2-ROS frente al inhibidor de ROS quinasa dirigido.

La tinción inmunohistoquímica (IHC) también puede emplearse para determinar el estado de expresión y/o activación de un polipéptido de ROS quinasa mutante en un cáncer de mamífero (por ejemplo, NSCLC) antes, durante y después de un tratamiento con un fármaco dirigido a inhibir la actividad ROS quinasa. La tinción IHC puede realizarse según técnicas muy conocidas. Véase, por ejemplo, ANTIBODIES: A LABORATORY MANUAL, capítulo 10, Harlow y Lane eds., Cold Spring Harbor Laboratory (1988). Brevemente y como ejemplo, se prepara un tejido embebido en parafina (por ejemplo, un tejido tumoral procedente de una biopsia) para la tinción inmunohistoquímica desparafinando secciones de tejido con xileno, seguido de etanol; hidratando en agua y después en PBS; exponiendo el antígeno mediante el calentamiento del portaobjetos en tampón citrato de sodio; incubando las secciones en peróxido de hidrógeno; bloqueando en una disolución de bloqueo; incubando el portaobjetos en un anticuerpo primario anti-polipéptido de fusión de SLC34A2-ROS y anticuerpo secundario; y, por último, detectando mediante el empleo de un método de biotina/avidina ABC según las instrucciones del fabricante.

Los ensayos de inmunofluorescencia (IF) también pueden emplearse para determinar el estado de expresión y/o activación de un polipéptido de fusión de SLC34A2-ROS en un cáncer de mamífero antes, durante y después de un tratamiento con un fármaco dirigido a inhibir la actividad ROS quinasa. Los ensayos de IF pueden realizarse según técnicas muy conocidas. Véase, por ejemplo, J. M. Polak y S. Van Noorden (1997), INTRODUCTION TO IMMUNOCYTOCHEMISTRY, 2ª ed., ROYAL MICROSCOPY SOCIETY MICROSCOPY HANDBOOK 37, BioScientific/Springer-Verlag. Brevemente y como ejemplo, las muestras del paciente pueden fijarse en paraformaldehído, seguido de metanol, bloquearse con una disolución de bloqueo, tal como suero de caballo, incubarse con el anticuerpo primario contra el polipéptido de fusión de SLC34A2-ROS, seguido de un anticuerpo secundario marcado con un tinte fluorescente, tal como Alexa 488, y analizarse con un microscopio de epifluorescencia.

Los anticuerpos empleados en los ensayos descritos anteriormente pueden conjugarse de modo ventajoso con tintes fluorescentes (por ejemplo, Alexa 488, PE) u otros marcadores, tales como puntos cuánticos, para su uso en análisis multiparamétricos, junto con otros anticuerpos de transducción de señales (EGFR, fosfo-AKT, fosfo-Erk 1/2) y/o de marcador celular (citoqueratina).

Una diversidad de otros protocolos, que incluyen el ensayo inmunoabsorbente ligado a enzimas (ELISA), radioinmunoensayo (RIA) y clasificación de células activada por fluorescencia (FACS), para medir los polipéptidos de ROS quinasa mutantes, son conocidos en la técnica y proporcionan una base para diagnosticar unos niveles alterados o anómalos de la expresión del polipéptido de fusión de SLC34A2-ROS. Se establecen unos valores normales o patrón para la expresión del polipéptido de fusión de SLC34A2-ROS combinando los fluidos corporales o extractos celulares obtenidos de sujetos mamíferos normales, preferiblemente seres humanos, con un anticuerpo contra el polipéptido de fusión de SLC34A2-ROS bajo condiciones adecuadas para la formación de complejos. La cantidad de formación de complejo de patrón puede cuantificarse mediante diversos métodos, pero preferiblemente por medios fotométricos. Las cantidades del polipéptido de fusión de SLC34A2-ROS expresadas en un sujeto, el control y las muestras de la enfermedad procedentes de los tejidos de biopsia se comparan con los valores patrón. La desviación entre los valores patrón y del sujeto establece los parámetros para diagnosticar la enfermedad.

#### Ensayos de péptidos y nucleótidos

De modo similar, pueden prepararse péptidos AQUA para la detección/cuantificación del polipéptido de ROS mutante expresado en una muestra biológica que comprende células de un tumor y utilizarse en ensayos AQUA convencionales, según se describe en detalle en la anterior sección E. Por consiguiente, en algunas realizaciones preferidas de los métodos de la descripción, el reactivo específico del polipéptido de fusión de SLC34A2-ROS comprende un fosfopéptido marcado con un isótopo pesado (péptido AQUA) que se corresponde con una secuencia peptídica que comprende la zona de unión de la fusión del polipéptido de fusión de SLC34A2-ROS, según se describió anteriormente en la sección E.

Los reactivos específicos del polipéptido de ROS quinasa mutante útiles para practicar los métodos de la invención también pueden ser sondas de ARNm, de oligonucleótidos o de ADN que se pueden hibridar directamente y detectar



los transcritos de expresión del polipéptido de fusión o truncado en una muestra biológica. Estas sondas se analizan en detalle en la anterior sección B. Brevemente y como ejemplo, muestras de un paciente fijadas con formaldehído y embebidas en parafina pueden sondarse con una sonda de ARN marcada con fluoresceína, seguido de unos lavados con formamida, SSC y PBS, y un análisis con un microscopio de fluorescencia.

5 Los polinucleótidos que codifican un polipéptido de ROS quinasa mutante también pueden utilizarse con fines de diagnóstico. Los polinucleótidos que pueden utilizarse incluyen secuencias de oligonucleótidos, moléculas de ADN y ARN antisentido y PNA. Los polinucleótidos pueden emplearse para detectar y cuantificar la expresión génica en tejidos de biopsia en los que la expresión de un polipéptido de fusión de SLC34A2-ROS o un polipéptido de ROS truncado puede correlacionarse con una enfermedad. El ensayo de diagnóstico puede emplearse para distinguir entre la ausencia, la presencia y un exceso de expresión del polipéptido de fusión de SLC34A2-ROS, y para controlar la regulación de los niveles del polipéptido de fusión de SLC34A2-ROS durante una intervención terapéutica.

15 En una realización preferida, la hibridación con sondas de PCR que son capaces de detectar secuencias de polinucleótidos, que incluyen secuencias genómicas, que codifican el polipéptido de fusión de SLC34A2-ROS o el polipéptido de ROS quinasa truncado, o moléculas muy relacionadas, puede utilizarse para identificar secuencias de ácidos nucleicos que codifican un polipéptido de ROS mutante. La construcción y el uso de dichas sondas se describe en la anterior sección B. La especificidad de la sonda, tanto si se produce a partir de una región muy específica, por ejemplo, 10 nucleótidos exclusivos en la zona de unión de la fusión, o de una región menos específica, por ejemplo, la región codificadora 3', y la rigurosidad de la hibridación o la amplificación (máxima, alta, intermedia o baja) determinará si la sonda identifica solo a las secuencias que aparecen en naturaleza que codifican el polipéptido de ROS quinasa mutante, los alelos o secuencias relacionadas.

20 Las sondas también pueden emplearse para la detección de secuencias relacionadas, y deberían contener preferiblemente al menos 50% de los nucleótidos de cualquiera de las secuencias codificadoras del polipéptido de ROS mutante. Las sondas de hibridación de la presente invención pueden ser ADN o ARN y pueden derivarse de las secuencias de nucleótidos de SEQ ID NO:2 o 4, y lo más preferiblemente incluyen la zona de unión de la fusión (véase la figura 7, panel inferior), o derivarse de una secuencia genómica que incluya un promotor, elementos potenciadores e intrones de los polipéptidos SLC34A2 y ROS naturales, como se describe más a fondo en la anterior sección B.

25 El polinucleótido de fusión de SLC34A2-ROS o el polinucleótido de ROS truncado puede utilizarse en análisis de la transferencia Southern o Northern, transferencias por puntos, u otras tecnologías basadas en membranas; en tecnologías de PCR; o en ensayos de bastones de inmersión, espigas, ELISA o de chip empleando fluidos o tejidos procedentes de biopsias de pacientes para detectar una expresión alterada del polipéptido de ROS quinasa mutante. Estos métodos cualitativos o cuantitativos son muy conocidos en la técnica. En un aspecto concreto, las secuencias de nucleótidos que codifican el polipéptido de ROS mutante pueden ser útiles en ensayos que detectan la activación o la inducción de diversos cánceres, incluyendo cánceres de pulmón, que incluye NSCLC. Los polinucleótidos de ROS mutante pueden marcarse por métodos convencionales, y añadirse a una muestra de fluido o tejido procedente de un paciente bajo condiciones adecuadas para la formación de complejos de hibridación. Después de un periodo de incubación adecuado, la muestra se lava y se cuantifica la señal y se compara con un valor patrón. Si la cantidad de señal en la muestra procedente de biopsia o extraída está significativamente alterada con respecto a la de una muestra control comparable, esto indica que las secuencias de nucleótidos se han hibridado con las secuencias de nucleótidos en la muestra, y la presencia de unos niveles alterados de secuencias de nucleótidos que codifican el polipéptido de fusión de SLC34A2-ROS o el polipéptido de ROS quinasa truncado en la muestra indica la presencia de la enfermedad asociada. Estos ensayos también se pueden utilizar para evaluar la eficacia de un régimen de tratamiento terapéutico concreto en estudios animales, en ensayos clínicos, o en el control del tratamiento de un paciente individual.

30 Para proporcionar una base para el diagnóstico de una enfermedad caracterizada por la expresión de un polipéptido de ROS mutante, se establece un perfil normal o patrón para la expresión. Esto puede lograrse combinando fluidos corporales o extractos de células obtenidos de sujetos normales, tanto animales como humanos, con una secuencia, o con uno de sus fragmentos, que codifica el polipéptido de fusión de SLC34A2-ROS o el polipéptido de ROS quinasa truncado, bajo condiciones adecuadas para la hibridación o la amplificación. La hibridación patrón puede cuantificarse comparando los valores obtenidos de sujetos normales con los obtenidos en un experimento en el que se emplea una cantidad conocida de un polinucleótido sustancialmente purificado. Los valores patrón obtenidos a partir de muestras normales pueden compararse con los valores obtenidos de muestras procedentes de pacientes que son sintomáticos para la enfermedad. La desviación entre los valores patrón y del sujeto se emplea para establecer la presencia de la enfermedad.

35 Tras haber establecido la enfermedad e iniciarse un protocolo de tratamiento, los ensayos de hibridación pueden repetirse de manera regular para evaluar si el nivel de expresión en el paciente empieza a aproximarse al observado en un paciente normal. Los resultados obtenidos de los ensayos sucesivos pueden emplearse para demostrar la eficacia del tratamiento a lo largo de un periodo que varía de varios días a meses.

40 Otros usos de diagnóstico de los polinucleótidos de ROS mutante de la invención pueden implicar el uso de la

reacción en cadena de la polimerasa (PCR), otro formato de ensayo preferido que es habitual entre los expertos en la técnica. Véase, por ejemplo, MOLECULAR CLONING, A LABORATORY MANUAL, 2ª edición, Sambrook, J., Fritsch, E. F. y Maniatis, T., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Los oligómeros de la PCR pueden sintetizarse de modo químico, generarse de modo enzimático, o producirse a partir de una fuente recombinante. Los oligómeros preferiblemente consistirán en dos secuencias de nucleótidos, una con la orientación sentido (5' a 3') y otra con la orientación antisentido (3' a 5'), y emplearse bajo condiciones optimizadas para la identificación de un gen o trastorno específico. Pueden emplearse los mismos dos oligómeros, conjuntos anidados de oligómeros, o incluso un agrupamiento degenerado de oligómeros bajo condiciones menos rigurosas para la detección y/o la cuantificación de secuencias de ADN o ARN muy relacionadas.

Los métodos que también pueden emplearse para cuantificar la expresión del polipéptido de fusión de SLC34A2-ROS o el polipéptido de ROS quinasa truncado incluyen el radiomarcaje o la biotilación de nucleótidos, la coamplificación de un ácido nucleico control, y las curvas patrón en las cuales se interpolan los resultados experimentales (Melby *et al.*, *J. Immunol. Methods*, 159: 235-244 (1993); Duplaa *et al.*, *Anal. Biochem.*, 229-236 (1993)). La velocidad de la cuantificación de múltiples muestras puede acelerarse realizando el ensayo en un formato ELISA en el que el oligómero de interés se presenta en diversas diluciones, y una respuesta espectrofotométrica o colorimétrica produce una cuantificación rápida.

En otra realización de la invención, los polinucleótidos de ROS mutantes de la invención pueden emplearse para generar sondas de hibridación que son útiles para el cartografiado de la secuencia genómica natural. Las secuencias pueden cartografiarse en un cromosoma concreto o en una región específica del cromosoma empleando técnicas muy conocidas. Estas técnicas incluyen la hibridación in situ de fluorescencia (FISH), FACS, o construcciones de cromosomas artificiales, tales como cromosomas artificiales de levadura, cromosomas artificiales bacterianos, construcciones P1 bacterianas, o bancos de ADNc de cromosomas individuales, tal como se indica en Price, C. M., *Blood Rev.*, 7: 127-134 (1993), y Trask, B. J., *Trends Genet.*, 7: 149-154 (1991).

En una realización preferida, se emplea FISH (tal como se describe en Verma *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press, Nueva York, N.Y. (1988)) y puede correlacionarse con otras técnicas de cartografiado de cromosomas físicas y datos de mapas genéticos. Pueden encontrarse ejemplos de datos de mapas genéticos en the 1994 Genome Issue of *Science* (265: 1981f). La correlación entre la localización del gen que codifica el polipéptido de fusión de SLC34A2-ROS o el polipéptido de ROS truncado sobre un mapa cromosómico físico y una enfermedad específica, o predisposición a una enfermedad específica, puede ayudar a delimitar la región del ADN asociada con esa enfermedad genética. Las secuencias de nucleótidos de la presente invención pueden emplearse para detectar diferencias en las secuencias génicas entre individuos normales, portadores o afectados.

Puede emplearse la hibridación in situ de preparaciones cromosómicas y técnicas de cartografiado físico, tales como análisis de vinculación que emplean marcadores cromosómicos establecidos, para extender mapas genéticos. A menudo, la colocación de un gen sobre el cromosoma de otra especie de mamífero, tal como un ratón, puede revelar marcadores asociados, incluso si no se conoce el número o brazo de un cromosoma humano concreto. Pueden asignarse nuevas secuencias a los brazos cromosómicos, o a sus partes, mediante cartografiado físico. Esto proporciona información valiosa a los investigadores que buscan genes de enfermedades empleando la clonación posicional u otras técnicas de descubrimiento de genes. Tras haber localizado de forma general la enfermedad o el síndrome mediante vinculación genética con una región genómica concreta, por ejemplo, AT con 11q22-23 (Gatti *et al.*, *Nature*, 336: 577-580 (1988)), cualquier secuencia cartografiada en esa área puede representar genes asociados o reguladores para la posterior investigación. La secuencia de nucleótidos de la presente invención también puede emplearse para detectar diferencias en la localización cromosómica debidas a la translocación, inversión, etc., entre individuos normales, portadores o afectados.

#### 45 Muestras biológicas

Las muestras biológicas útiles en la práctica de los métodos de la descripción pueden obtenerse a partir de cualquier mamífero en el que un cáncer que se caracteriza por la presencia de un polipéptido de fusión de SLC34A2-ROS está o pueda estar presente o en desarrollo. En una realización, el mamífero es un ser humano, y el ser humano puede ser un candidato para emplear un producto terapéutico que inhibe a ROS, para el tratamiento de un cáncer de pulmón, por ejemplo, NSCLC. El candidato humano puede ser un paciente que está siendo tratado en la actualidad, o que se considere para un tratamiento con un inhibidor de ROS quinasa. En otra realización, el mamífero es un animal grande, tal como un caballo o una vaca, mientras que en otras realizaciones el mamífero es un animal pequeño, tal como un perro o un gato, de los cuales se sabe que todos pueden desarrollar cánceres, incluyendo cánceres de pulmón.

Cualquier muestra biológica que comprenda células (o extractos de células) procedentes de un cáncer de mamífero es adecuada para su uso en los métodos de la invención. En una realización, la muestra biológica comprende células obtenidas a partir de una biopsia de tumor. La biopsia puede obtenerse, según las técnicas clínicas convencionales, a partir de tumores primarios que aparezcan en el órgano de un mamífero, o de tumores secundarios que han metastatizado hacia otros tejidos. En otra realización, la muestra biológica comprende células obtenidas a partir de un aspirado con una aguja fina extraídas de un tumor, y las técnicas para obtener estos

aspirados son muy conocidas en la técnica (véase Cristallini *et al.*, *Acta Cytol.*, 36(3): 416-422 (1992)).

La muestra biológica también puede comprender células obtenidas a partir de una efusión, tal como una efusión pleural. Se sabe que, en muchos pacientes con cáncer de pulmón avanzado (que incluye NSCLC), se forman efusiones pleurales (el líquido que se forma fuera del pulmón en la cavidad torácica y que contiene células cancerosas), y la presencia de dicha efusión es predictiva de un resultado malo y un tiempo de supervivencia corto. Se han descrito técnicas convencionales para obtener muestras de efusión pleural y estas son muy conocidas en la técnica (véase Sahn, *Clin Chest Med.*, 3(2): 443-452 (1982)). Las células tumorales en la circulación también pueden obtenerse a partir del suero empleando marcadores tumorales, marcadores de proteína de citoqueratina u otros métodos de selección negativa, según se ha descrito (véase Ma *et al.*, *Anticancer Res.*, 23(1A): 49-62 (2003)). Las muestras de suero y de médula ósea pueden ser particularmente preferidas en pacientes con leucemia. Se ha observado la expresión aberrante de ROS en un glioblastoma. Véase Charest *et al.*, *supra*.

Una muestra biológica puede comprender células (o extractos de células) procedentes de un cáncer en el que un polipéptido de fusión de SLC34A2-ROS o un polipéptido de ROS quinasa truncado se expresa y/o se activa, pero no una ROS quinasa de tipo salvaje. Como alternativa, la muestra puede comprender células procedentes de un cáncer en el que un polipéptido de ROS mutante y una ROS quinasa de tipo salvaje se expresan y/o se activan, o en el que una ROS quinasa de tipo salvaje y/o SLC34A2 se expresan y/o activan, pero no un polipéptido de ROS mutante.

Pueden prepararse extractos celulares de las anteriores muestras biológicas, brutas o parcial (o totalmente) purificadas, según técnicas convencionales, y emplearse en los métodos de la descripción. Como alternativa, pueden utilizarse muestras biológicas que comprenden células completas en formatos de ensayo preferidos, tales como ensayos de inmunohistoquímica (IHC), citometría de flujo (FC), y de inmunofluorescencia (IF), tal como se describió más a fondo anteriormente. Estos ensayos de células enteras son ventajosos porque minimizan la manipulación de la muestra de células tumorales y, así, reducen los riesgos de alterar el estado de señalización/activación *in vivo* de las células y/o introducir señales accidentales. Los ensayos de células completas también son ventajosos porque caracterizan la expresión y la señalización solo en células tumorales, en lugar de una mezcla de células tumorales y normales.

En la práctica del método descrito para determinar si un compuesto inhibe el avance de un tumor caracterizado por un polipéptido de fusión y/o de translocación de SLC34A2-ROS, también pueden emplearse de forma ventajosa muestras biológicas que comprenden células de xenoinjertos de mamífero (o trasplantes de médula ósea). Los xenoinjertos preferidos (o los receptores de trasplantes) son mamíferos pequeños, tales como ratones, que portan tumores humanos (o leucemias) que expresan un polipéptido de ROS quinasa mutante. Los xenoinjertos que portan tumores humanos son muy conocidos en la técnica (véase Kal, *Cancer Treat Res.*, 72: 155-169 (1995)), y la producción de xenoinjertos de mamíferos que portan tumores humanos ha sido bien descrita (véase Winograd *et al.*, *In Vivo*, 1(1): 1-13 (1987)). De forma similar, la generación y el uso de modelos de trasplante de médula ósea ha sido bien descrito (véase, por ejemplo, Schwaller, *et al.*, *EMBO J.*, 17: 5321-333 (1998); Kelly *et al.*, *Blood*, 99: 310-318 (2002)). Un "cáncer caracterizado por" un polipéptido de fusión y/o de translocación de SLC34A2-ROS significa un cáncer en el que dicho polipéptido expresado y/o gen de ROS mutante está presente, comparado con un cáncer en el que dicho polipéptido de fusión y/o de translocación no está presente.

Para evaluar la presencia del polinucleótido de ROS mutante o la expresión del polipéptido en una muestra biológica que comprende células procedentes de un tumor de cáncer de mamífero, puede emplearse de modo deseable una muestra control que representa una célula en la que dicha proteína de fusión y/o translocación no aparece, para fines comparativos. De modo ideal, la muestra control comprende células de un subconjunto del cáncer concreto (por ejemplo, NSCLC) que es representativo del subconjunto en el que no aparece la mutación (por ejemplo, translocación de SLC34A2-ROS) y/o el polipéptido de fusión no se expresa. Así, la comparación del nivel en la muestra control frente a la muestra biológica de ensayo indica si el polipéptido y/o el polinucleótido mutante están presentes. Como alternativa, puesto que el polipéptido y/o el polinucleótido de fusión de SLC34A2-ROS puede no estar presente en la mayoría de los cánceres, cualquier tejido que, de forma similar, no expresa el polipéptido de ROS mutante (o no porta el polinucleótido mutante) puede emplearse como control.

Los métodos descritos a continuación tienen una utilidad diagnóstica valiosa para cánceres que se caracterizan por un polinucleótido y/o un polipéptido de ROS mutante, y para las decisiones de tratamiento que los conciernen. Por ejemplo, pueden obtenerse muestras biológicas de un sujeto que no ha sido previamente diagnosticado con un cáncer que se caracteriza por un polipéptido de fusión y/o de translocación de SLC34A2-ROS, ni que se haya sometido aún a un tratamiento para dicho cáncer, y el método se emplea para determinar diagnósticamente que un tumor en dicho sujeto pertenece a un subconjunto de tumores (por ejemplo, tumores NSCLC) en los que un polinucleótido y/o un polipéptido de ROS mutante está presente/se expresa.

Como alternativa, una muestra biológica puede obtenerse de un sujeto que ha sido diagnosticado con un cáncer dirigido por un tipo de quinasa, tal como EGFR, y que ha estado recibiendo terapia, tal como una terapia con un inhibidor de EGFR (por ejemplo, Tarceva™, Iressa™), para el tratamiento de dicho cáncer, y el método de la descripción se emplea para determinar si el tumor del sujeto también se caracteriza por un polipéptido de fusión y/o de translocación de SLC34A2-ROS y, por tanto, es probable que responda completamente a la terapia existente y/o si resulta deseable o puede garantizarse una terapia de inhibición de la ROS quinasa alternativa o adicional. Los

métodos de la descripción también pueden emplearse para controlar el avance o la inhibición de un cáncer que expresa un polipéptido de ROS mutante después del tratamiento del sujeto con una composición que comprende un producto terapéutico o una combinación de productos terapéuticos que inhiben la ROS quinasa.

5 Este ensayo de diagnóstico puede realizarse después o antes de los procedimientos de la evaluación preliminar o de vigilancia quirúrgica. El método de identificación de la descripción puede emplearse de modo ventajoso como un diagnóstico para identificar pacientes que tienen un cáncer, tal como NSCLC, dirigido por la proteína de fusión de SLC34A2-ROS, para saber cuáles son los pacientes es más probable que respondan a productos terapéuticos dirigidos a la inhibición de la actividad ROS quinasa. La capacidad para seleccionar dichos pacientes también será útil en la evaluación clínica de la eficacia de futuros productos terapéuticos dirigidos a ROS, así como en la prescripción futura de dichos fármacos a pacientes.

#### Diagnósticos

15 La capacidad para identificar selectivamente cánceres en los que un polipéptido de fusión y/o de translocación de SLC34A2-ROS está presente permite desarrollar nuevos métodos importantes para la identificación precisa de dichos tumores para fines de diagnóstico, así como obtener información útil para determinar si es probable que un tumor responda a una composición terapéutica inhibidora de ROS, o si es probable que no responda parcial o totalmente a un inhibidor que se dirija a una quinasa diferentes cuando se administra como agente único para el tratamiento del cáncer.

20 Por consiguiente, en una realización, la descripción proporciona un método para detectar la presencia de un polinucleótido y/o un polipéptido de ROS mutante de la invención en una muestra biológica obtenida de un paciente que tiene cáncer, comprendiendo dicho método la etapa de utilizar al menos un reactivo que detecta un polinucleótido o un polipéptido de ROS mutante de la invención para determinar si un polinucleótido y/o un polipéptido de fusión de SLC34A2-ROS está presente en la muestra biológica.

25 En algunos ejemplos, el cáncer es un cáncer de pulmón, tal como un carcinoma pulmonar no microcítico (NSCLC). En otro ejemplo, la presencia de un polipéptido de ROS quinasa mutante identifica un cáncer que es probable que responda a una composición que comprende al menos un producto terapéutico que inhibe la ROS quinasa.

30 En algunas realizaciones preferidas, los métodos de diagnóstico de la invención se ponen en práctica en un formato de ensayo de citometría de flujo (FC), de inmunohistoquímica (IHC), o de inmunofluorescencia (IF). En otra realización preferida, se detecta la actividad del polipéptido de fusión de SLC34A2-ROS. En otras realizaciones preferidas, los métodos de diagnóstico se ponen en práctica en un formato de ensayo de hibridación in situ de fluorescencia (FISH) o de reacción en cadena de la polimerasa (PCR).

35 La invención proporciona también un método para determinar si un compuesto inhibe el avance de un cáncer que se caracteriza por un polinucleótido o un polipéptido de fusión de SLC34A2-ROS, comprendiendo dicho método la etapa de determinar si dicho compuesto inhibe la expresión y/o la actividad de dicha fusión de SLC34A2-ROS en dicho cáncer. En una realización preferida, la inhibición de la expresión y/o la actividad del polipéptido de fusión de SLC34A2-ROS se determina empleando al menos un reactivo que detecta un polinucleótido o un polipéptido de fusión de SLC34A2-ROS de la invención. Los compuestos adecuados para la inhibición de la actividad ROS quinasa se analizan con más detalle en la siguiente sección G.

40 En las anteriores secciones B y D se describen con más detalles los reactivos específicos de polipéptidos y sondas de polinucleótidos de ROS mutante útiles en la práctica de los métodos de la descripción. En un ejemplo, el reactivo específico del polipéptido de fusión de SLC34A2-ROS comprende un anticuerpo específico del polipéptido de fusión. En otro ejemplo, el reactivo específico del polipéptido de fusión comprende un fosfopéptido marcado con un isótopo pesado (péptido AQUA) que se corresponde con la zona de unión de la fusión del polipéptido de fusión de SLC34A2-ROS (véase la figura 7 (panel inferior)).

45 Los métodos de la invención descritos anteriormente también pueden comprender opcionalmente la etapa de determinar el nivel de expresión o de activación de otras quinasas, tales como EGFR y ROS de tipo salvaje, u otras moléculas de señalización corriente abajo en dicha muestra biológica. La determinación del perfil de expresión/activación del polipéptido de fusión de SLC34A2-ROS y de expresión/activación de otras quinasas y vías en una muestra biológica concreta puede proporcionar información valiosa acerca de cuáles son las quinasas y las vías que dirigen la enfermedad, y de cuál es el régimen terapéutico que, por tanto, es más probable que sea el más beneficioso.

#### Selección de compuestos

55 El descubrimiento de los nuevos polipéptidos de fusión de SLC34A2-ROS descritos en la presente también permite el desarrollo de nuevos compuestos que inhiben la actividad de estas proteínas mutantes, en particular su actividad ROS quinasa. Por consiguiente, la invención también proporciona, en parte, un método para determinar si un compuesto inhibe el avance de un cáncer que se caracteriza por un polipéptido y/o un polinucleótido de fusión de SLC34A2-ROS, comprendiendo dicho método la etapa de determinar si dicho compuesto inhibe la expresión y/o la actividad de dicho polipéptido de fusión de SLC34A2-ROS en dicho cáncer. En una realización preferida, la

inhibición de la expresión y/o la actividad del polipéptido de fusión de SLC34A2-ROS se determina empleando al menos un reactivo que detecte un polinucleótido de ROS mutante y/o un polipéptido de ROS mutante de la invención. Los reactivos de la invención preferidos se han descrito anteriormente. Los compuestos adecuados para la inhibición de la actividad ROS quinasa se describen con más detalle en la siguiente sección G.

5 El compuesto puede ser, por ejemplo, un inhibidor de quinasas, tal como un inhibidor de molécula pequeña o de anticuerpo. Puede ser un inhibidor de pan-quinasa con actividad contra varias quinasas diferentes, o un inhibidor específico de quinasa. Los compuestos que inhiben la ROS quinasa se analizan con más detalle en la siguiente sección G. Las muestras biológicas de los pacientes pueden tomarse antes y después del tratamiento con el inhibidor y después analizarse, empleando los métodos descritos anteriormente, para el efecto biológico del inhibidor sobre la actividad ROS quinasa, que incluyen la fosforilación de la proteína sustrato corriente abajo. Este ensayo farmacodinámico puede ser útil para determinar la dosis biológicamente activa del fármaco que puede ser preferible hasta una dosis máxima tolerable. Esta información también sería útil en propuestas para la aprobación de fármacos mediante la demostración del mecanismo de acción del fármaco. La identificación de compuestos con esas características inhibitorias deseadas se describe con más detalle en la siguiente sección G.

#### 15 *G. Inhibición terapéutica de cánceres*

Según la presente invención, ahora se ha demostrado que el polipéptido de fusión de SLC34A2-ROS aparece en al menos un subgrupo de NSCLC humano. Por consiguiente, el avance de un cáncer de mamífero (por ejemplo, NSCLC), en el que se expresa la proteína de fusión de SLC34A2-ROS, puede ser inhibido *in vivo* mediante la inhibición de la actividad de ROS quinasa en dicho cáncer. La actividad ROS en cánceres que se caracterizan por la expresión de una ROS quinasa mutante puede ser inhibida poniendo en contacto el cáncer (por ejemplo, un tumor) con un producto terapéutico inhibidor de la ROS quinasa. Por consiguiente, la invención proporciona, en parte, un método para inhibir el avance de un cáncer que expresa el polipéptido de fusión de SLC34A2-ROS mediante la inhibición de la expresión y/o la actividad de una ROS quinasa en el cáncer.

Un producto terapéutico inhibidor de la ROS quinasa puede tener cualquier composición que comprenda al menos un compuesto, biológico o químico, que inhiba, directa o indirectamente, la expresión y/o la actividad de la ROS quinasa *in vivo*, que incluye los ejemplos de clases de compuestos descritos a continuación. Estos compuestos incluyen productos terapéuticos que actúan directamente sobre la propia ROS quinasa, o sobre proteínas o moléculas que modifican la actividad de ROS, o que actúan indirectamente inhibiendo la expresión de ROS. Estas composiciones también incluyen composiciones que comprenden solo un único compuesto inhibidor de ROS quinasa, así como composiciones que comprenden múltiples productos terapéuticos (que incluyen productos contra otras RTK), que también pueden incluir un agente terapéutico no específico, tal como un agente quimioterapéutico o un inhibidor de la transcripción en general.

#### Inhibidores de molécula pequeña

En algunas realizaciones preferidas, un producto terapéutico inhibidor de ROS útil en la práctica de los métodos de la descripción es un inhibidor de molécula pequeña dirigido. Los inhibidores de molécula pequeña dirigidos son una clase de moléculas que generalmente inhiben la actividad de su enzima diana mediante la unión específica, y a menudo irreversible, al sitio catalítico de la enzima y/o la unión a un surco de unión a ATP u otro sitio de unión dentro de la enzima que evita que la enzima adopte una conformación necesaria para su actividad. Un ejemplo de inhibidor de quinasa dirigido de molécula pequeña es Gleevec® (Imatinib, STI-571), que inhibe a CSF1R y BCR-ABL, y sus propiedades han sido bien descritas. Véase Dewar *et al.*, *Blood*, 105(8): 3127-3132 (2005).

Los inhibidores de molécula pequeña pueden diseñarse racionalmente empleando la formación de modelos cristalográficos de rayos X o informáticos de la estructura tridimensional de la ROS quinasa, o pueden descubrirse mediante la selección de alta capacidad de procesamiento de bancos de compuestos para la inhibición de ROS. Estos métodos son muy conocidos en el técnica y han sido descritos. La especificidad de la inhibición de ROS puede confirmarse, por ejemplo, estudiando la capacidad de dicho compuesto para inhibir la actividad ROS, pero no otra actividad quinasa, en un panel de quinasas y/o estudiando la inhibición de la actividad ROS en una muestra biológica que comprenda células tumorales de NSCLC, según se describió anteriormente. Estos métodos de selección se describen más a fondo a continuación.

#### Inhibidores de anticuerpos

Los productos terapéuticos que inhiben a ROS quinasa útiles en los métodos de la invención también pueden ser anticuerpos dirigidos que se unen específicamente a sitios de unión o catalíticos críticos o a dominios requeridos para la actividad ROS e inhiben a la quinasa bloqueando el acceso de ligandos, sustratos o moléculas secundarias a  $\alpha$  y/o evitando que la enzima adopte una conformación necesaria para su actividad. La producción, la selección y el uso terapéutico de anticuerpos humanizados específicos de la diana ha sido bien descrito. Véase Merluzzi *et al.*, *Adv Clin Path.*, 4(2): 77-85 (2000). Están disponibles sistemas y tecnologías comerciales, tales como Morphosys, Inc.'s Human Combinatorial Antibody Library (HuCAL®), para la selección y generación de alta capacidad de procesamiento de anticuerpos humanizados inhibidores específicos de la diana.

Se ha descrito la producción de diversos anticuerpos dirigidos anti-receptor de quinasas y su uso para la inhibición de la actividad del receptor diana. Véase, por ejemplo, la publicación de patente de EEUU n.º 20040202655, "Antibodies to IGF-I Receptor for the Treatment of Cancers," 14 de octubre de 2004, Morton *et al.*; la publicación de patente de EEUU n.º 20040086503, "Human anti-Epidermal Growth Factor Receptor Single-Chain Antibodies," Apr. 15, 2004, Raisch *et al.*; la publicación de patente de EEUU n.º 20040033543, "Treatment of Renal Carcinoma Using Antibodies Against the EGFr," 19 de febrero de 2004, Schwab *et al.* En la técnica se conocen métodos estandarizados para la producción y el uso de anticuerpos inhibidores de la actividad del receptor de tirosina quinasa. Véase, por ejemplo, la patente europea n.º EP1423428, "Antibodies that Block Receptor Tyrosine Kinase Activation, Methods of Screening for and Uses Thereof," 2 de junio de 2004, Borges *et al.*

También pueden emplearse estrategias de presentación de fagos para generar inhibidores de anticuerpos específicos de ROS, y se proporcionan protocolos para la construcción de bancos de bacteriófagos y la selección de anticuerpos recombinantes en el texto de referencia muy conocido CURRENT PROTOCOLS IN IMMUNOLOGY, Colligan *et al.* (eds.), John Wiley & Sons, Inc. (1992-2000), capítulo 17, sección 17.1. Véase también la patente de EEUU n.º 6.319.690, 20 de noviembre de 2001, Little *et al.*; la patente de EEUU n.º 6.300.064, 9 de octubre de 2001, Knappik *et al.*; la patente de EEUU n.º 5.840.479, 24 de noviembre de 1998, Little *et al.*; la publicación de patente de EEUU n.º 20030219839, 27 de noviembre de 2003, Bowdish *et al.*

Puede producirse un banco de fragmentos de anticuerpos presentados sobre la superficie de bacteriófagos (véase, por ejemplo, la patente de EEUU n.º 6.300.064, 9 de octubre de 2001, Knappik *et al.*) y seleccionarse para la unión a una forma dimérica soluble de un receptor de proteína tirosina quinasa (como ROS). Se identifica un fragmento de anticuerpo que se une a la forma dimérica soluble de la RTK empleada para la selección, como molécula candidata para bloquear la activación constitutiva de la RTK diana en una célula, Véase la patente europea n.º EP1423428, Borges *et al.*, *supra*.

Los anticuerpos dirigidos a la unión de ROS identificados en la selección de los bancos de anticuerpos según se describió anteriormente después pueden seleccionarse para su capacidad de bloquear la actividad de ROS, en un ensayo de quinasa *in vitro* y en tumores y/o líneas celulares *in vivo*. La inhibición de ROS puede ser confirmada, por ejemplo, estudiando la capacidad de dicho producto terapéutico de anticuerpo para inhibir la actividad ROS quinasa, pero no otra actividad quinasa, en un panel de quinasas y/o estudiando la inhibición de la actividad ROS en una muestra biológica que comprende células cancerosas, según se describió anteriormente. Los métodos para seleccionar estos compuestos para la inhibición de ROS quinasa se describen más a fondo a continuación.

### 30 Inhibidores indirectos

Los compuestos inhibidores de ROS útiles en la práctica de los métodos descritos también pueden ser compuestos que inhiben indirectamente la actividad ROS mediante la inhibición de la actividad de proteínas o moléculas distintas de la propia ROS quinasa. Estos productos terapéuticos inhibidores pueden ser inhibidores dirigidos que modulan la actividad de quinasas reguladoras clave que fosforilan o desfosforilan (y, por tanto, activan o desactivan) a la propia ROS, o que interfieren con la unión de ligandos. Al igual que otros receptores de tirosina quinasa, ROS regula la señalización corriente abajo a través de una red de proteínas adaptadoras y quinasas corriente abajo. Como resultado, la inducción del crecimiento celular y de la supervivencia por la actividad ROS puede ser inhibida afectando a estas proteínas de interacción o corriente abajo.

La actividad ROS quinasa también puede inhibirse indirectamente empleando un compuesto que inhiba la unión de una molécula activante necesaria por ROS para adoptar su conformación activa. Por ejemplo, se ha descrito la producción y el uso de anticuerpos anti-PDGF. Véase la publicación de patente de EEUU n.º 20030219839, "Anti-PDGF Antibodies and Methods for Producing Engineered Antibodies," Bowdish *et al.* La inhibición de la unión del ligando (PDGF) al receptor directamente infrarregula la actividad del receptor.

Los inhibidores indirectos de la actividad ROS pueden diseñarse racionalmente empleando la formación de modelos cristalográficos de rayos X o informáticos de la estructura tridimensional de ROS, o pueden descubrirse mediante la selección de alta capacidad de procesamiento de bancos de compuestos para la inhibición de enzimas reguladoras corriente arriba claves y/o las moléculas de unión necesarias, que producen la inhibición de la actividad ROS quinasa. Estas estrategias son muy conocidas en la técnica y han sido descritos. La inhibición de ROS por dichos productos terapéuticos puede confirmarse, por ejemplo, estudiando la capacidad del compuesto para inhibir la actividad ROS, pero no otra actividad quinasa, en un panel de quinasas y/o estudiando la inhibición de la actividad ROS en una muestra biológica que comprende células cancerosas, por ejemplo, células de NSCLC, según se describió anteriormente. Estos métodos para identificar compuestos que inhiben a un cáncer que se caracteriza por un polipéptido de fusión y/o de translocación de SLC34A2-ROS y/o un polinucleótido y/o un polipéptido de ROS truncado, se describen más a fondo a continuación.

### 55 Inhibidores de la transcripción y/o antisentido

Los productos terapéuticos para la inhibición de ROS también pueden comprender compuestos inhibidores de la transcripción y/o antisentido que inhiben la actividad ROS quinasa bloqueando la transcripción del gen que codifica ROS y/o el gen de fusión de SLC34A2-ROS. La inhibición de diversas quinasas de receptores, que incluyen VEGFR,

EGFR, e IGFR, y FGFR, por productos terapéuticos antisentido para el tratamiento del cáncer se ha descrito. Véanse, por ejemplo, las patentes de EEUU n.ºs 6.734.017; 6.710.174; 6.617.162; 6.340.674; 5.783.683; 5.610.288.

Pueden diseñarse, construirse y emplearse oligonucleótidos antisentido como agentes terapéuticos contra genes diana según técnicas conocidas. Véase, por ejemplo, Cohen, J., *Trends in Pharmacol. Sci.*, 10(11): 435-437 (1989); Marcus-Sekura, *Anal. Biochem.*, 172: 289-295 (1988); Weintraub, H., *Sci. AM.*, pp. 40-46 (1990); Van Der Krol *et al.*, *Bio Techniques*, 6(10): 958-976 (1988); Skorski *et al.*, *Proc. Natl. Acad. Sci. USA* (1994), 91: 4504-4508. Recientemente se ha descrito la inhibición del crecimiento del carcinoma humano *in vivo* empleando un inhibidor de ARN antisentido de EGFR. Véase la publicación de patente de EEUU n.º 20040047847, "Inhibition of Human Squamous Cell Carcinoma Growth *In vivo* by Epidermal Growth Factor Receptor Antisense RNA Transcribed from a Pol III Promoter," 11 de marzo de 2004, He *et al.* De forma similar, un producto terapéutico que inhibe a ROS que comprende al menos un oligonucleótido antisentido contra un gen ROS de mamífero (véase la figura 4 (SEQ ID NO:8) o el polinucleótido de fusión de SLC34A2-ROS o el polinucleótido de ROS truncado (véase la figura 2 (SEQ ID NO:2 o 4) o truncado puede prepararse según los métodos descritos anteriormente. Las composiciones farmacéuticas que comprenden compuestos antisentido inhibidores de ROS pueden prepararse y administrarse como se describe más a fondo a continuación.

#### ARN de interferencia pequeño

También pueden emplearse de forma deseable composiciones de moléculas de ARN de interferencia pequeño (ARNip), que inhiben la traducción y, por tanto, la actividad, de ROS a través del proceso de la interferencia de ARN, en los métodos de la descripción. La interferencia del ARN y el silenciamiento selectivo de la expresión de proteínas diana mediante la introducción de moléculas de ARN bicatenario pequeño exógeno que comprende una secuencia complementaria con el ARNm que codifica la proteína diana, ha sido bien descrito. Véase, por ejemplo, la publicación de patente de EEUU n.º 20040038921, "Composition and Method for Inhibiting Expression of a Target Gene," 26 de febrero de 2004, Kreutzer *et al.*; la publicación de patente de EEUU n.º 20020086356, "RNA Sequence-Specific Mediators of RNA Interference," 12 de junio de 2003, Tuschl *et al.*; la publicación de patente de EEUU n.º 20040229266, "RNA Interference Mediating Small RNA Molecules," 18 de noviembre de 2004, Tuschl *et al.*

Por ejemplo, tal como se describe en el ejemplo 3, el silenciamiento mediado por ARNip de la expresión de la proteína de fusión de SLC34A2-ROS puede realizarse en una línea celular de NSCLC humano que exprese la proteína de fusión.

Las moléculas de ARN bicatenario (ARNbc) ha demostrado bloquear la expresión de genes a través de un mecanismo regulador altamente conservado conocido como ARN de interferencia (ARNi). Brevemente, la ARNasa III Dicer procesa el ARNbc en pequeños ARNp de interferencia (ARNip) de aproximadamente 22 nucleótidos, que actúan como secuencias de guía para inducir la ruptura del ARNm específico de diana mediante un complejo silenciador inducido por ARN RISC (véase Hammond *et al.*, *Nature* (2000), 404: 293-296). El ARNi implica una reacción de tipo catalítico, mediante la cual se generan ARNip a través de la ruptura sucesiva de ARNbc más largos. Así, a diferencia del ARN antisentido, el ARNi degrada el ARN diana de una manera no estequiométrica. Cuando se administra a una célula o a un organismo, el ARNbc exógeno ha demostrado dirigir la degradación específica de secuencia del ARN mensajero (ARNm) endógeno a través del ARNi.

Ahora está disponible en el mercado una amplia diversidad de productos de ARNi específico de diana, que incluyen vectores y sistemas para su expresión y uso en células de mamífero. Véase, por ejemplo, Promega, Inc. ([www.promega.com](http://www.promega.com)); Dharmacon, Inc. ([www.dharmacon.com](http://www.dharmacon.com)). Están disponibles manuales técnicos detallados sobre el diseño, la construcción y el uso de ARNbc para ARNi. Véase, por ejemplo, "RNAi Technical Reference & Application Guide" de Dharmacon; "RNAi: A Guide to Gene Silencing" de Promega. También están disponibles en el mercado productos de ARNip inhibidores de ROS, y pueden ser empleados de modo adecuado en el método de la invención. Véase, por ejemplo, Dharmacon, Inc., Lafayette, Colo. (n.ºs de catálogo M-003162-03, MU-003162-03, D-003162-07 thru -10 (ARNip siGENOME™ SMARTselection y SMARTpool®).

Recientemente se ha establecido que unos ARNbc pequeños menores que 49 nucleótidos de longitud, y preferiblemente de 19-25 nucleótidos, que comprenden al menos una secuencia que es sustancialmente idéntica a una parte de la secuencia de ARNm diana, y cuyo ARNbc óptimamente tiene al menos un grupo pendiente de 1-4 nucleótidos en un extremo, es muy eficaz para mediar en el ARNi en mamíferos. Véase la publicación de patente de EEUU n.º 20040038921, Kreutzer *et al.*, *supra*; la publicación de patente de EEUU n.º 20040229266, Tuschl *et al.*, *supra*. La construcción de estos ARNbc y su uso en preparaciones farmacéuticas para silenciar la expresión de una proteína diana *in vivo* se describen con detalle en dichas publicaciones.

Si se conoce la secuencia del gen diana en un mamífero, pueden producirse ARN de 21-23 nt, por ejemplo, y ensayarse para su capacidad para mediar en el ARNi en una célula de mamífero, tal como una célula humana o de otro primate. Las moléculas de ARN de 21-23 nt que demuestren mediar en el ARNi puede ensayarse, si se desea, en un modelo animal apropiado para evaluar también su eficacia *in vivo*. Los sitios diana que se conozcan, por ejemplo, los sitios diana que se ha determinado que son sitios diana eficaces basándose en estudios con otras moléculas de ácidos nucleicos, por ejemplo, ribozimas o antisentido, o las dianas que se sabe que están asociadas

con una enfermedad o un trastorno, tales como los sitios que contienen mutaciones o deleciones, pueden utilizarse para diseñar moléculas de ARNip que también se dirijan a esos sitios.

Como alternativa, las secuencias de los ARNbc eficaces pueden diseñarse/precedirse racionalmente seleccionando el ARNm diana de interés para sitios diana, por ejemplo, empleando un algoritmo informático de plegamiento. La secuencia diana puede ser analizada mediante simulación por ordenador para producir una lista de todos los fragmentos o subsecuencias de una longitud concreta, por ejemplo, fragmentos de 23 nucleótidos, empleando una secuencia de instrucciones Perl o programas de análisis de secuencias del mercado, tales como Oligo, McVector, o el paquete GCG Wisconsin.

Pueden utilizarse diversos parámetros para determinar los sitios que son los sitios diana más adecuados dentro de la secuencia de ARN diana. Estos parámetros incluyen la estructura de ARN secundaria o terciaria, la composición de bases nucleotídicas de la secuencia diana, el grado de homología entre diversas regiones de la secuencia diana, o la posición relativa de la secuencia diana dentro del transcrito de ARN. Basándose en estas determinaciones, puede elegirse cualquiera de una serie de sitios diana dentro del transcrito de ARN para seleccionar las moléculas de ARNip para su eficacia, por ejemplo, empleando ensayos de ruptura de ARN *in vitro*, cultivos celulares o modelos animales. Véase, por ejemplo, la publicación de patente de EEUU n.º 20030170891, 11 de septiembre de 2003, McSwiggen J. También se ha descrito recientemente un algoritmo para identificar y seleccionar sitios diana de ARNi. Véase la publicación de patente de EEUU n.º 20040236517, "Selection of Target Sites for Antisense Attack of RNA," 25 de noviembre de 2004, Drlica *et al.*

Las técnicas de transferencia de genes empleadas habitualmente incluyen fosfato de calcio, DEAE-dextrano, electroporación y microinyección y métodos víricos (Graham *et al.* (1973), *Virology*, 52: 456; McCutchan *et al.* (1968), *J. Natl. Cancer Inst.*, 41: 351; Chu *et al.* (1987), *Nucl. Acids Res.*, 15: 1311; Fraley *et al.* (1980), *J. Biol. Chem.*, 255: 10431; Capecchi (1980), *Cell*, 22: 479). El ADN también puede introducirse en las células empleando liposomas catiónicos (Feigner *et al.* (1987), *Proc. Natl. Acad. Sci. USA*, 84: 7413). Las formulaciones de lípidos catiónicos disponibles en el mercado incluyen Tfx 50 (Promega) o Lipofectamin 200 (Life Technologies). Como alternativa, pueden emplearse vectores víricos para transportar ARNbc a las células y mediar en el ARNi. Véase la publicación de patente de EEUU n.º 20040023390, "siRNA-mediated Gene Silencing with Viral Vectors," 4 de febrero de 2004, Davidson *et al.*

Los sistemas de transfección y vector/expresión para el ARiNi en células de mamífero están disponibles en el mercado y han sido bien descritos. Véase, por ejemplo, Dharmacon, Inc., sistema DharmaFECT™; Promega, Inc., sistema de horquilla siSTRIKE™ U6 Hairpin; véase también Gou *et al.* (2003), *FEBS*, 548, 113-118; Sui, G. *et al.*, A DNA vector-based RNAi technology to suppress gene expression in mammalian cells (2002), *Proc. Natl. Acad. Sci.*, 99, 5515-5520; Yu *et al.* (2002), *Proc. Natl. Acad. Sci.*, 99, 6047-6052; Paul, C. *et al.* (2002), *Nature Biotechnology*, 19, 505-508; McManus *et al.* (2002), *RNA*, 8, 842-850.

Así, puede realizarse la interferencia de ARNip en un mamífero empleando moléculas de ARNbc preparadas administrando una preparación farmacéutica que comprende el ARNbc al mamífero. La composición farmacéutica se administra en una dosificación suficiente para inhibir la expresión del gen diana. El ARNbc puede administrarse generalmente a una dosificación menor que 5 mg de ARNbc por kilogramo de peso corporal diarios, y es suficiente para inhibir o suprimir completamente la expresión del gen diana. En general, una dosis adecuada de ARNbc estará en el intervalo de 0,01 a 2,5 miligramos por kilogramo de peso corporal del receptor diarios, preferiblemente en el intervalo de 0,1 a 200 microgramos por kilogramo de peso corporal diarios, más preferiblemente en el intervalo de 0,1 a 100 microgramos por kilogramo de peso corporal diarios, aún más preferiblemente en el intervalo de 1,0 a 50 microgramos por kilogramo de peso corporal diarios, y lo más preferiblemente en el intervalo de 1,0 a 25 microgramos por kilogramo de peso corporal diarios. Una composición farmacéutica que comprende el ARNbc se administra una vez diaria, o en múltiples subdosis, por ejemplo, empleando formulaciones de liberación sostenida muy conocidas en la técnica. La preparación y la administración de dichas composiciones farmacéuticas puede realizarse según técnicas convencionales, tal como se describe más a fondo a continuación.

Estos ARNbc después pueden utilizarse para inhibir la expresión y la actividad de ROS en un cáncer, mediante la preparación de una preparación farmacéutica que comprende una cantidad terapéuticamente eficaz de dicho ARNbc, tal como se describió anteriormente, y la administración de la preparación a un sujeto humano que tenga un cáncer que exprese la proteína de fusión de SLC34A2-ROS o el polipéptido de ROS quinasa truncado, por ejemplo, mediante una inyección directa en el tumor. La inhibición similar de otros receptores de tirosina quinasa, tales como VEGFR y EGFR, empleando inhibidores de ARNip se ha descrito recientemente. Véase la publicación de patente de EEUU n.º 20040209832, 21 de octubre de 2004, McSwiggen *et al.*; la publicación de patente de EEUU n.º 20030170891, 11 de septiembre de 2003, McSwiggen; la publicación de patente de EEUU n.º 20040175703, 9 de septiembre de 2004, Kreutzer *et al.*

#### Composición terapéuticas y administración

Las composiciones terapéuticas inhibitorias de ROS quinasa útiles en la práctica de los métodos de la descripción pueden administrarse a un mamífero mediante cualquier medio conocido en la técnica que incluye, pero no se limita a la vía oral o peritoneal, que incluye la administración intravenosa, intramuscular, intraperitoneal, subcutánea,



transdérmica, por las vías respiratorias (aerosol), rectal, vaginal y tópica (que incluye bucal y sublingual).

Para la administración oral, un producto terapéutico inhibidor de ROS en general se proporcionará en forma de comprimidos o cápsulas, como un polvo o gránulos, o como una suspensión o disolución acuosa. Los comprimidos para un uso oral pueden incluir ingredientes activos mezclados con excipientes farmacéuticamente aceptables, tales como diluyentes inertes, agentes disgregantes, agentes ligantes, agentes lubricantes, agentes edulcorantes, agentes aromatizantes, agentes colorantes y conservantes. Los diluyentes inertes adecuados incluyen carbonato de sodio y calcio, fosfato de sodio y calcio, y lactosa, mientras que el almidón de maíz y el ácido algínico son agentes disgregantes adecuados. Los agentes ligantes pueden incluir almidón y gelatina, mientras que el agente lubricante, si está presente, en general será estearato de magnesio, ácido esteárico o talco. Si se desea, los comprimidos pueden estar revestidos con un material, tal como monoestearato de glicerilo o diestearato de glicerilo, para retrasar la absorción en el tracto gastrointestinal.

Las cápsulas para un uso oral incluyen cápsulas de gelatina dura, en las que el ingrediente activo está mezclado con un diluyente sólido, y cápsulas de gelatina blanda, en las que el ingrediente activo se mezcla con agua o un aceite, tal como aceite de cacahuete, parafina líquida o aceite de oliva. Para un uso intramuscular, intraperitoneal, subcutáneo e intravenoso, las composiciones farmacéuticas de la descripción en general se proporcionarán en suspensiones o disoluciones acuosas estériles, tamponadas hasta un pH e isotonicidad adecuados. Los vehículos acuosos adecuados incluyen disolución de Ringer y cloruro de sodio isotónico. El vehículo puede consistir exclusivamente en un tampón acuoso ("exclusivamente" significa que no están presentes agentes auxiliares ni sustancias encapsulantes que puedan afectar o mediar en la captación del producto terapéutico inhibidor de ROS). Estas sustancias incluyen, por ejemplo, estructuras micelares, tales como liposomas o cápsidas, tal como se describe a continuación. Las suspensiones acuosas pueden incluir agentes suspensores, tales como derivados de celulosa, alginato de sodio, polivinilpirrolidona y goma de tragacanto, y un agente humectante, tal como lecitina. Los conservantes adecuados para las suspensiones acuosas incluyen p-hidroxibenzoato de etilo y n-propilo.

Las composiciones terapéuticas inhibidoras de ROS quinasa también pueden incluir formulaciones encapsuladas para proteger al producto terapéutico (por ejemplo, un compuesto de ARNbc) frente a la eliminación rápida del cuerpo, tal como una formulación de liberación controlada, que incluye implantes y sistemas de transporte microencapsulados. Pueden emplearse polímeros biodegradables y biocompatibles, tales como etileno-acetato de vinilo, polianhídridos, poli(ácido glicólico), colágeno, poliortoésteres y poli(ácido láctico). Los métodos para la preparación de dichas formulaciones serán evidentes para los expertos en la técnica. Los materiales también pueden obtenerse en el mercado en Alza Corporation y Nova Pharmaceuticals, Inc. Las suspensiones liposómicas (que incluyen liposomas dirigidos a células infectadas con anticuerpos monoclonales contra antígenos víricos) también pueden emplearse como vehículos farmacéuticamente aceptables. Estos pueden prepararse según métodos conocidos por los expertos en la técnica, por ejemplo, según se describe en la patente de EEUU n.º 4.522.811; la publicación PCT WO 91/06309; y la publicación de patente europea EP-A-43075. Una formulación encapsulada puede comprender una proteína de la envuelta vírica. La proteína de la envuelta vírica puede derivarse o estar asociada con un virus, tal como un virus de polio, o puede ser parcial o totalmente artificial. Por ejemplo, la proteína de la envuelta puede ser una proteína vírica 1 y/o una proteína vírica 2 del virus del polio, o uno de sus derivados.

Las composiciones inhibidoras de ROS pueden comprender también un vehículo de transporte, que incluye liposomas, para la administración a un sujeto, vehículos y diluyentes y sus sales y/o pueden estar presentes en formulaciones farmacéuticamente aceptables. Por ejemplo, los métodos para el transporte de moléculas de ácidos nucleicos se describen en Akhtar *et al.*, 1992, *Trends Cell Bio.*, 2, 139; DELIVERY STRATEGIES FOR ANTISENSE OLIGONUCLEOTIDE THERAPEUTICS, ed. Akhtar, 1995; Maurer *et al.*, 1999, *Mol. Membr. Biol.*, 16, 129-140; Hofland y Huang, 1999, *Handb. Exp. Pharmacol.*, 137, 165-192; y Lee *et al.*, 2000, ACS Symp. Ser., 752, 184-192. En Beigelman *et al.*, patente de EEUU n.º 6.395.713, y Sullivan *et al.*, PCT WO 94/02595, también se describen métodos generales para el transporte de moléculas de ácidos nucleicos. Estos protocolos pueden utilizarse para el transporte de casi cualquier molécula de ácido nucleico.

Los productos terapéuticos inhibidores de ROS pueden administrarse a un tumor de mamífero mediante una diversidad de métodos conocidos por los expertos en la técnica, que incluyen, pero no se limitan a encapsulación en liposomas, mediante iontoforesis o mediante la incorporación en otros vehículos, tales como hidrogeles, ciclodextrinas, nanocápsulas biodegradables, y microesferas bioadhesivas, o mediante vectores proteicos (O'Hare y Normand, publicación PCT internacional n.º WO 00/53722). Como alternativa, la combinación de producto terapéutico/vehículo se administra localmente mediante una inyección directa o mediante el uso de una bomba de infusión. La inyección directa de la composición, por vía subcutánea, intramuscular o intradérmica, puede realizarse empleando metodologías convencionales de aguja y jeringa, o mediante tecnologías sin aguja, tales como las descritas en Conry *et al.*, 1999, *Clin. Cancer Res.*, 5, 2330-2337, y Barry *et al.*, publicación PCT internacional n.º WO 99/31262.

Las formulaciones farmacéuticamente aceptables de productos terapéuticos inhibidores de ROS quinasa incluyen sales de los compuestos descritos anteriormente, por ejemplo, sales de adición de ácidos, por ejemplo, sales del ácido clorhídrico, bromhídrico, acético y bencensulfónico. Una composición o formulación farmacológica se refiere a una composición o formulación en una forma adecuada para la administración, por ejemplo, administración

sistémica, en una célula o paciente, que incluye, por ejemplo, un ser humano. Las formas adecuadas, en parte, depende del uso o de la vía de entrada, por ejemplo, oral, transdérmica o mediante inyección. Estas formas no deben evitar que la composición o la formulación alcance a la célula diana. Por ejemplo, las composiciones farmacológicas inyectadas en la corriente sanguínea deben ser solubles. Otros factores son conocidos en la técnica, e incluyen consideraciones tales como la toxicidad y las formas que evitan que la composición o formulación ejerza su efecto.

Las vías de administración que conducen a una absorción sistémica (es decir, absorción sistémica o acumulación de fármacos en la corriente sanguínea, seguido de su distribución a través de todo el cuerpo) son deseables e incluyen, sin limitación, la vía intravenosa, subcutánea, intraperitoneal, por inhalación, oral, intrapulmonar e intramuscular. Cada una de estas vías de administración expone el producto terapéutico inhibidor de ROS a un tumor o tejido enfermo accesible. La velocidad de entrada de un fármaco en la circulación ha demostrado ser una función del peso molecular o el tamaño.

El uso de un liposoma u otro vehículo de fármaco que comprende los compuestos de la presente invención puede trasladar potencialmente el fármaco, por ejemplo, a ciertos tipos de tejidos, tales como los tejidos del sistema endotelial reticular (RES). Una formulación de liposomas que puede facilitar la asociación del fármaco con la superficie de las células, tales como linfocitos y macrófagos, también es útil. Esta estrategia puede proporcionar una mayor administración del fármaco a las células diana aprovechando la especificidad de los macrófagos y el reconocimiento inmunológico de los linfocitos de las células anómalas, tales como las células cancerosas.

Una "formulación farmacéuticamente aceptable" significa una composición o una formulación que permite la distribución eficaz de las moléculas de ácidos nucleicos de la presente invención en la localización física más adecuada para su actividad deseada. Los ejemplos no limitantes de agentes adecuados para la formulación con las moléculas de ácidos nucleicos de la presente invención incluyen inhibidores de P-glicoproteínas (tales como Pluronic P85), que pueden potenciar la entrada de fármacos en el SNC (Jolliet-Riant y Tillement, 1999, *Fundam. Clin. Pharmacol.*, 13, 16-26); polímeros biodegradables, tales como microesferas de poli(DL-lactida-coglicólido) para la administración de liberación sostenida después de la implatación intracerebral (Emerich *et al.*, 1999, *Cell Transplant*, 8, 47-58) (Alkermes, Inc. Cambridge, Mass.); y nanopartículas cargadas, tales como las fabricadas con polibutilcianoacrilato, que puede trasladar fármacos a través de la barrera hematoencefálica y pueden alterar los mecanismos de captación neuronal (*Prog. Neuro-psychopharmacol. Biol. Psychiatry*, 23, 941-949, 1999). Otros ejemplos no limitantes de estrategias de administración y transporte para los compuestos inhibidores de ROS útiles en el método de la invención incluyen los materiales descritos en Boado *et al.*, 1998, *J. Pharm. Sci.*, 87, 1308-1315; Tyler *et al.*, 1999, *FEBS Lett.*, 421, 280-284; Pardridge *et al.*, 1995, *PNAS USA*, 92, 5592-5596; Boado, 1995, *Adv. Drug Delivery Rev.*, 15, 73-107; Aldrian-Herrada *et al.*, 1998, *Nucleic Acids Res.*, 26, 4910-4916; y Tyler *et al.*, 1999, *PNAS USA*, 96, 7053-7058.

Las composiciones terapéuticas que comprenden liposomas con la superficie modificada que contienen lípidos de polietilenglicol (modificados con PEG, o liposomas de tiempo de circulación largo o liposomas fantasmas) también pueden emplearse de modo adecuado en los métodos de la descripción. Estas formulaciones ofrecen un método para aumentar la acumulación de fármacos en los tejidos diana. Esta clase de vehículos de fármacos resisten la opsonización y la eliminación por el sistema fagocítico mononuclear (MPS o RES), permitiendo con ello unos tiempos de circulación más largos en la sangre y una mayor exposición a los tejidos del fármaco encapsulado (Lasic *et al.*, *Chem. Rev.*, 1995, 95, 2601-2627; Ishiwata *et al.*, *Chem. Pharm. Bull.*, 1995, 43, 1005-1011). Estos liposomas han demostrado acumularse selectivamente en tumores, probablemente debido a la extravasación y la captura en los tejidos diana neovascularizados (Lasic *et al.*, *Science*, 1995, 267, 1275-1276; Oku *et al.*, 1995, *Biochim. Biophys. Acta*, 1238, 86-90). Los liposomas de tiempo de circulación largo potencian la farmacocinética y la farmacodinámica del ADN y ARN, en particular comparado con los liposomas catiónicos convencionales que se sabe que se acumulan en tejidos del MPS (Liu *et al.*, *J. Biol. Chem.*, 1995, 42, 24864-24870; Choi *et al.*, publicación PCT internacional n.º WO 96/10391; Ansell *et al.*, publicación PCT internacional n.º WO 96/10390; Holland *et al.*, publicación PCT internacional n.º WO 96/10392). También es probable que los liposomas de tiempo de circulación largo protejan a los fármacos frente a la degradación por nucleasas en un grado mayor comparado con los liposomas catiónicos, en base a su capacidad para evitar la acumulación en tejidos del MPS metabólicamente agresivos, tales como el hígado y el bazo.

Las composiciones terapéuticas pueden incluir una cantidad farmacéuticamente eficaz de los compuestos deseados en un vehículo o diluyente farmacéuticamente aceptable. Los vehículos o diluyentes aceptables para un uso terapéutico son muy conocidos en la técnica farmacéutica, y se describen, por ejemplo, en REMINGTON'S PHARMACEUTICAL SCIENCES, Mack Publishing Co. (A. R. Gennaro edit., 1985). Por ejemplo, pueden proporcionarse conservantes, estabilizantes, tintes y agentes aromatizantes. Estos incluyen benzoato de sodio, ácido sórbico y ésteres del ácido p-hidroxibenzoico. Además, pueden utilizarse antioxidantes y agentes suspensores.

Una dosis farmacéuticamente eficaz es la dosis requerida para prevenir, inhibir la aparición o tratar (aliviar un síntoma hasta cierto grado, preferiblemente todos los síntomas) de un estado de enfermedad. La dosis farmacéuticamente eficaz depende del tipo de enfermedad, la composición utilizada, la vía de administración, el tipo de mamífero que se está tratando, las características físicas del mamífero específico considerado, la medicación

concurrente, y otros factores que reconocerán los expertos en la técnica médica. En general, se administra una cantidad entre 0,1 mg/kg y 100 mg/kg de peso corporal/día de los ingredientes activos que depende de la potencia del polímero cargado negativamente.

5 Unos niveles de dosificación del orden de aproximadamente 0,1 mg a aproximadamente 140 mg por kilogramo de peso corporal diarios son útiles en el tratamiento de los trastornos indicados anteriormente (de aproximadamente 0,5 mg a aproximadamente 7 g por paciente diarios). La cantidad del ingrediente activo que puede combinarse con los materiales vehículo para producir una forma de dosificación unitaria que varía dependiendo del hospedante tratado y de la vía de administración concreta. Las formas de dosificación unitarias contienen en general entre  
10 aproximadamente 1 mg y aproximadamente 500 mg de un ingrediente activo. Se entiende que el nivel de dosis específico para cualquier paciente concreto depende de una diversidad de factores, que incluyen la actividad del compuesto específico empleado, la edad, el peso corporal, la salud general, el sexo, la dieta, el momento de la administración, la vía de administración y la velocidad de excreción, la combinación de fármacos y la gravedad de la enfermedad concreta que se está sometiendo a terapia.

15 Para la administración a animales no humanos, la composición también puede añadirse al pienso animal o al agua para beber. Puede resultar conveniente formular composiciones para pienso animal y para el agua para beber de modo que el animal ingiera la cantidad terapéuticamente apropiada de la composición junto con su dieta. También puede resultar conveniente presentar la composición como una premezcla para la adición al pienso o al agua para beber.

20 Un producto terapéutico inhibidor de ROS útil en la práctica de la descripción puede comprender un único compuesto, tal como se describió anteriormente, o una combinación de múltiples compuestos, de la misma clase que el inhibidor (es decir, un inhibidor de anticuerpo) o de clases diferentes (es decir, inhibidores de anticuerpo e inhibidores de molécula pequeña). Esta combinación de compuestos puede aumentar el efecto terapéutico global inhibiendo el avance de un cáncer que expresa la proteína de fusión. Por ejemplo, la composición terapéutica puede  
25 incluir un inhibidor de molécula pequeña, tal como STI-571 (Gleevec®) solo, o estar en combinación con otros análogos de Gleevec® dirigidos a afectar la actividad ROS y/o inhibidores de molécula pequeña de EGRF, tales como Tarceva™ o Iressa™. La composición terapéutica también puede comprender uno o más agentes quimioterapéuticos no específicos además de uno o más inhibidores dirigidos. Estas combinaciones han demostrado recientemente que proporcionan un efecto sinérgico destructor del tumor en muchos cánceres. La eficacia de dichas combinaciones para inhibir la actividad de ROS y el crecimiento tumoral *in vivo* puede evaluarse según se describe a  
30 continuación.

#### Identificación de compuestos inhibidores de ROS quinasa mutante

La invención también proporciona, en parte, un método para determinar si un compuesto inhibe el avance de un cáncer que se caracteriza por un polipéptido de fusión y/o de translocación de SLC34A2-ROS, determinando si el  
35 compuesto inhibe la actividad del polipéptido de fusión de SLC34A2-ROS en el cáncer. En algunas realizaciones preferidas, la inhibición de la actividad de ROS se determina estudiando una muestra biológica que comprende células de la médula ósea, sangre o un tumor. En otra realización preferida, la inhibición de la actividad de ROS se determina empleando al menos un reactivo específico del polipéptido o polinucleótido de ROS mutante de la invención.

40 El compuesto ensayado puede ser cualquier tipo de composición o producto terapéutico según se describió anteriormente. Los métodos para evaluar la eficacia de un compuesto, *in vitro* e *in vivo*, están bien establecidos y son conocidos en la técnica. Por ejemplo, una composición puede ensayarse para su capacidad para inhibir ROS *in vitro* empleando una célula o un extracto celular en el que la ROS quinasa está activada. Puede emplearse un panel de compuestos para ensayar la especificidad del compuesto por ROS (en oposición a otras dianas, tales como EGFR o PDGFR).

45 Otra técnica para la selección de fármacos que puede utilizarse proporciona una selección de alta capacidad de procesamiento de compuestos que tienen una afinidad de unión adecuada con una proteína de interés, según se describe en la solicitud PCT publicada WO84/03564. En este método, al igual que se aplica a los polipéptidos de ROS mutante, se sintetiza un gran número de compuestos de ensayo pequeños diferentes sobre un sustrato sólido, tal como varillas de plástico o alguna otra superficie. Los compuestos de ensayo se hacen reaccionar con un  
50 polipéptido de ROS mutante, o con uno de sus fragmentos, y se lavan. El polipéptido mutante unido (por ejemplo, el polipéptido de fusión de SLC34A2-ROS) después se detecta mediante métodos muy conocidos en la técnica. El polipéptido de ROS mutante purificado también puede revestirse directamente sobre placas para su uso en las técnicas de selección de fármacos mencionadas anteriormente. Como alternativa, pueden emplearse anticuerpos no neutralizantes para capturar el péptido e inmovilizarlo sobre un soporte sólido.

55 Un compuesto que se ha descubierto que es un inhibidor eficaz de la actividad ROS *in vitro* después puede estudiarse para su capacidad para inhibir el avance de un cáncer que expresa el polipéptido de fusión de SLC34A2-ROS *in vivo* empleando, por ejemplo, xenoinjertos de mamífero que portan tumores NSCLC humanos que están dirigidos por la proteína de fusión de SLC34A2-ROS. En este procedimiento, líneas celulares que se sabe que están dirigidas por la proteína de fusión de SLC34A2-ROS se colocan de modo subcutáneo en el ratón. Las células

después se hacen crecer hasta una masa tumoral que puede controlarse de forma visual. El ratón después puede ser tratado con el fármaco. El efecto del tratamiento del fármaco sobre el tamaño del tumor puede observarse externamente. El ratón después se sacrifica y el tumor se retira para su análisis mediante IHC y análisis de la transferencia Western. De modo similar, pueden prepararse transplantes de médula ósea de mamífero mediante métodos convencionales para estudiar la respuesta del fármaco en tumores hematológicos que expresen una ROS quinasa mutante. De esta forma pueden observarse los efectos del fármaco en un entorno biológico que es lo más parecido posible a un paciente. Puede determinarse la capacidad del fármaco para alterar la señalización en las células tumorales o las células estromáticas circundantes mediante un análisis con anticuerpos específicos de la fosforilación. También puede observarse la eficacia del fármaco para inducir la muerte celular o la inhibición de la proliferación celular mediante un análisis con marcadores específicos de la apoptosis, tales como caspasa 3 rota y PARP rota.

Puede determinarse la toxicidad y la eficacia terapéutica de estos compuestos mediante procedimientos farmacéuticos convencionales en cultivos celulares o animales experimentales, por ejemplo, para determinar la LD50 (la dosis letal para 50% de la población) y la ED50 (la dosis terapéuticamente eficaz en 50% de la población). La proporción de dosis entre los efectos tóxicos y terapéuticos es el índice terapéutico, y puede expresarse como la proporción de LD50/ED50. Se prefieren los compuestos que muestran unos índices terapéuticos altos.

Los siguientes ejemplos se proporcionan para ilustrar más a fondo la invención.

#### **Ejemplo 1: Identificación de la actividad ROS quinasa en una línea celular NSCLC mediante determinación de los perfiles de fosfopéptidos globales**

Se estudió el perfil de fosforilación global de la activación de quinasas en varias líneas celulares de NSCLC humanas, que incluyen HCC78, empleando una técnica poderosa recientemente descrita para el aislamiento y la caracterización espectrométrica de masas de péptidos modificados a partir de mezclas complejas (la técnica "IAP", véase Rush *et al.*, *supra*). La técnica IAP se realizó empleando un anticuerpo específico de fosfotirosina (CELL SIGNALING TECHNOLOGY, INC., Beverly, Mass., 2003/04, n.º de catálogo 9411) para aislar y posteriormente caracterizar los péptidos que contienen fosfotirosina a partir de extractos de las líneas celulares de NSCLC.

De forma específica, se empleó la estrategia IAP para facilitar la identificación de las tirosina quinasas activadas en las líneas celulares de NSCLC, para identificar nuevos directores de esta enfermedad.

##### Cultivo celular

Las células HCC78 se obtuvieron en DSMZ (German National Resource Centre for Biological Material), se cultivaron en medio RPMI-1640 (Invitrogen) con suero bovino fetal al 10% (FBS) (Sigma).

##### Inmunoprecipitación de fosfopéptidos

Un total de  $2 \times 10^8$  células fueron lisadas en tampón de lisis de urea (HEPES 20 mM, pH 8,0, urea 9 M, vanadato de sodio 1 mM, pirofosfato de sodio 2,5 mM, beta-glicerofosfato 1 mM) a  $1,25 \times 10^8$  células/ml y se sonicaron. Los lisados sonicados se aclararon mediante centrifugación a 20.000 x g, y las proteínas se redujeron y se alquilaron tal como se ha descrito previamente (véase Rush *et al.*, *Nat. Biotechnol.*, 23(1): 94-101 (2005)). Las muestras se diluyeron con HEPES 20 mM, pH 8,0, hasta una concentración final de urea de 2 M. Se añadió tripsina (1 mg/ml en HCl 0,001 M) al lisado aclarado a 1:100 en v/v. Las muestras se digirieron durante la noche a temperatura ambiente.

Después de la digestión, los lisados se acidificaron hasta una concentración final de TFA al 1%. La purificación de los péptidos se realizó empleando columnas Sep-Pak C<sub>18</sub>, tal como se ha descrito previamente (véase Rush *et al.*, *supra*). Después de la purificación, se reunieron todas las eluciones (acetonitrilo al 8%, 12%, 15%, 18%, 22%, 25%, 30%, 35% y 40% en TFA al 0,1%) y se liofilizaron. Los péptidos secados se resuspendieron en tampón MOPS 1,4 ml (MOPS 50 mM/NaOH, pH 7,2, Na<sub>2</sub>HPO<sub>4</sub> 10 mM, NaCl 50 mM) y se retiró el material insoluble mediante centrifugación a 12.000 x g durante 10 minutos.

El anticuerpo monoclonal de fosfotirosina P-Tyr-100 (Cell Signaling Technology) procedente del fluido de ascitis se acopló de modo no covalente a esferas de agarosa-proteína G (Roche) a 4 mg/ml de esferas durante la noche a 4 °C. Después del acoplamiento, el anticuerpo-resina se lavó dos veces con PBS y tres veces con tampón MOPS. El anticuerpo inmovilizado (40 µl, 160 µg) se añadió como una suspensión 1:1 en tampón MOPS IP a la fracción de péptidos solubilizada, y la mezcla se incubó durante la noche a 4 °C. Las esferas de anticuerpos inmovilizados se lavaron tres veces con tampón MOPS y dos veces con ddH<sub>2</sub>O. Los péptidos se eluyeron dos veces de las esferas mediante incubación con 40 µl de TFA al 0,1% durante 20 minutos cada una, y las fracciones se reunieron.

##### Análisis mediante espectrometría de masas LC-MS/MS

Los péptidos en el eluato de IP (40 µl) se concentraron y se separaron del anticuerpo eluido empleando puntas de extracción Stop and Go (StageTips) (véase Rappsilber *et al.*, *Anal. Chem.*, 75(3): 663-670 (2003)). Los péptidos se eluyeron de las microcolumnas con 1 µl de MeCN al 60%, TFA al 0,1% hacia 7,6 µl de ácido acético al 0,4%/ácido heptafluorobutírico al 0,005% (HFBA). La muestra se cargó sobre una columna capilar PicoFrit de 10 cm x 75 µm

(New Objective) cargada con la resina de fase inversa Maic C18 AQ (Michrom Bioresources) empleando un muestreador automático Famos con una válvula de inyección de la muestra inerte (Dionex). La columna se reveló con un gradiente lineal de 45 min de acetonitrilo en ácido acético al 0,4%, HFBA al 0,005% administrado a 280 nl/min (Ultimate, Dionex).

- 5 Los espectros de masas en tándem se adquirieron de una manera dependiente de los datos con un espectrómetro de masas de atrapamiento de iones LCQ Deca XP Plus (ThermoFinnigan) empleando un método de los cuatro superiores, un recuento de repetición de exclusión dinámica de 1, y una duración de repetición de 0,5 min.

#### Análisis de bases de datos y asignaciones

- 10 Los espectros de MS/MS se evaluaron empleando TurboSequest (ThermoFinnigan) (en el paquete Sequest Browser (v. 27, rev. 12) suministrado como parte de BioWorks 3.0). Los espectros de MS/MS individuales se extrajeron del archivo de datos bruto empleando el programa CreateDta de Sequest Browser con los siguientes ajustes: PM inferior, 700; PM superior, 4.500; número mínimo de iones, 20; TIC mínima  $4 \times 10^5$ ; y estado de carga del precursor, no especificado. Los espectros se obtuvieron desde el principio del archivo de datos bruto antes de la inyección de la muestra hasta el final del gradiente de elución. No se emplearon los programas IonQuest ni VuDta para seleccionar los espectros de MS/MS para el análisis Sequest. Los espectros de MS/MS se evaluaron con los siguientes parámetros TurboSequest: tolerancia de masas de péptidos, 2,5; tolerancia de fragmentos de iones, 0,0; número máximo de aminoácidos diferentes por modificación, 4; tipo de masa parental, promedio; tipo de masa de los fragmentos, promedio; número máximo de sitios de ruptura internos, 10; se consideraron las pérdidas neutras de agua y amoniaco de los iones b e y en los análisis de correlación. Se especifica la enzima proteolítica excepto para los espectros recogidos de las digestiones con elastasa.

Se realizaron búsquedas en la base de datos humana NCBI facilitada el 24 de agosto de 2004, que contiene 27.175 proteínas que permiten la metionina oxidada (M+16) y la fosforilación (Y+80) como modificaciones dinámicas.

- 25 En la investigación proteómica, resulta deseable validar las identificaciones de proteínas basándose solamente en la observación de un único péptido en un resultado experimental, para indicar que la proteína está presente de hecho en una muestra. Esto ha conducido al desarrollo de métodos estadísticos para validar asignaciones de péptidos, que aún no están universalmente aceptados, y de líneas directrices para la publicación de resultados de identificación de proteínas y péptidos (véase Carr *et al.*, *Mol. Cell. Proteomics*, 3: 531-533 (2004)), que han sido seguidas en este ejemplo. Sin embargo, debido que la estrategia de inmunoafinidad separa a los péptidos fosforilados de los péptidos no fosforilados, la observación de un solo fosfopéptido procedente de una proteína es un resultado común, puesto que muchas proteínas fosforiladas solo tienen un sitio de tirosina fosforilada.

- 30 Por esta razón, resulta apropiado emplear otros criterios para validar las asignaciones de fosfopéptidos. Es probable que las asignaciones sean correctas si se cumple cualquiera de estos criterios adicionales: (i) la misma secuencia se asigna a iones coeluyentes con diferentes estados de carga, puesto que el espectro de MS/MS cambia notablemente con el estado de carga; (ii) el sitio aparece en más de un contexto de secuencia peptídica debido a solapamientos de secuencia debido a una proteólisis incompleta o al uso de proteasas diferentes de la tripsina; (iii) el sitio aparece en más de un contexto de secuencia peptídica debido a isoformas de proteínas homólogas, pero no idénticas; (iv) el sitio aparece en más de un contexto de secuencia peptídica debido a la existencia de proteínas homólogas, pero no idénticas, entre especies; y (v) sitios validados mediante análisis de MS/MS de fosfopéptidos sintéticos que se corresponden con secuencias asignadas, puesto que el espectrómetro de masas de atrapamiento de iones produce espectros de MS/MS altamente reproducibles. Este último criterio se emplea habitualmente para confirmar nuevas asignaciones de sitios de particular interés.

- 45 Todos los espectros y las asignaciones de secuencias realizados por Sequest se importaron en una base de datos de relaciones. Las secuencias asignadas se aceptaron o se rechazaron después de un proceso de dos etapas conservador. En la primera etapa, un subconjunto de asignaciones de secuencia de puntuación alta se seleccionaron filtrando para unos valores de XCorr de al menos 1,5 para un estado de carga de +1, 2,2 para +2, y 3,3 para +3, permitiendo un valor máximo de RSP de 10. Las evaluaciones en este subconjunto se rechazaron si se cumple cualquiera de los siguientes criterios: (i) el espectro contiene al menos un pico principal (al menos 10% tan intenso como el ion más intenso en el espectro) que no puede cartografiarse en la secuencia asignada como un ion a, b o y, como un ion que surge de la pérdida neutra de agua o amoniaco de un ion b o y, o como un ion protonado múltiple; (ii) el espectro no contiene una serie de iones b o y equivalentes al menos a seis restos ininterrumpidos; o (iii) la secuencia no se observa al menos cinco veces en todos los estudios que los inventores han realizado (excepto para secuencias solapantes debido a una proteólisis incompleta o al uso de proteasas distintas de la tripsina). En la segunda etapa, se aceptaron las asignaciones con puntuaciones por debajo del umbral si el espectro de puntuación baja muestra un alto grado de similitud con el espectro de puntuación alta recogido en otro estudio, que simula una estrategia de búsqueda en bancos de referencia verdadera. Todos los espectros que apoyan la lista final de secuencias asignadas (no se muestran en la presente) fueron revisados por al menos tres científicos para establecer su credibilidad.

Los anteriores análisis de IAP identificaron 454 péptidos que contienen fosfotirosina no redundantes, 395 sitios de fosfotirosina, y 240 proteínas fosforiladas en tirosina, la mayoría de los cuales son nuevos, procedentes de células

HCC78 (los datos no se muestran). Entre las quinasas con tirosina fosforilada, varias de las detectadas no son detectadas normalmente mediante un análisis de MS en otras líneas celulares de NSCLC (datos sin publicar), que incluyen la ROS quinasa.

**Ejemplo 2: Análisis de la transferencia Western de la expresión de ROS quinasa en una línea celular de NSCLC**

La observación de que la línea celular de NSCLC HCC78 (pero no las otras líneas celulares de NSCLC) expresa una ROS quinasa activada se confirmó mediante un análisis de la transferencia Western de extractos celulares empleando anticuerpos específicos para ROS y otros receptores de tirosina quinasa (RTK) y quinasas corriente abajo.

Las células HCC78 se lisaron en 1 x tampón de lisis celular (Cell Signaling Technology) suplementado con Protease Arrest™ (G Biosciencias) y se separaron mediante electroforesis. Todos los anticuerpos y reactivos para la inmunotransferencia proceden de Cell Signaling Technology, Inc. (Beverly, MA). La transferencia Western se realizó como se describe en "Western Immunoblotting Protocol" (Cell Signaling Technology, Inc., catálogo de 2005-2006). El anticuerpo anti-ROS se obtuvo en Santa Cruz Biotechnology, Inc.

La figura 5 muestra los resultados de la transferencia Western. Solo HCC78 expresa la proteína ROS entre muchas líneas celulares de NSCLC diferentes. La proteína de ROS en HCC78 tiene un peso molecular mucho menor que la proteína ROS de tipo salvaje, lo cual indica una proteína de fusión.

La transferencia Western confirma que la proteína de fusión de ROS está fosforilada en la tirosina. El lisado de proteínas procedente de las células HCC78 se inmunoprecipitó con un anticuerpo de fosfotirosina, y se inmunotransfirió con el anticuerpo de ROS total. Fueron detectadas las mismas bandas de pY-iP que en el lisado total por el anticuerpo de ROS, teniendo las bandas IPed una migración un poco más lenta, lo cual también indica la fosforilación de la proteína.

**Ejemplo 3: Inhibición del crecimiento de líneas celulares de NSCLC de mamífero que expresan la ROS quinasa empleando ARNip**

Para confirmar que la forma truncada de ROS está dirigiendo el crecimiento y la supervivencia de las células en la línea celular HCC78, se estudió la capacidad del ARNip silenciador para inhibir el crecimiento de estas células. La expresión de ROS se infrarreguló mediante interferencia de ARN. Se encargó el siguiente ARNip de ROS a Proligo, Inc., indicándose las correspondientes secuencias de ROS entre paréntesis: 5'AAGCCCGGAUGGCAACGUUTT3' (ROS1(6318-6340) (SEQ ID NO:23); 5'AAGCCUGAAGGCCUGAACUTT3' (ROS1(7181-7203) (SEQ ID NO:24).

Se sembraron 2 x 10<sup>5</sup> células en placas de 12 pocillos el día anterior a la transfección. Se transfectaron 100 nM de ARNip de ROS1 empleando el reactivo de transfección Mirus TransIT-TKO. Cuarenta y ocho horas después de la transfección, las células se trasladaron a un medio de inanición manteniéndose en este durante 24 horas más. Las células se recolectaron mediante tripsinización y después se contaron, y el lisado celular se empleó en WB para comprobar el nivel de proteínas de ROS.

Un análisis de inmunotransferencia reveló que la expresión de ROS se redujo específica y significativamente a las 72 horas después de la transfección del ARNip a las células HCC78, y que la línea celular control H2066 no expresa la proteína de ROS (véase la figura 10, panel B). A esto le acompaña una disminución en la fosforilación de los sustratos corriente abajo, tales como p-Erk1/2 y p-Akt, tal como se esperaba (véase la figura 10, panel C). Además, tal como se esperaba, el tratamiento con ARNip de ROS produjo un aumento en la apoptosis de la línea celular HCC78 (pero no en la línea celular control H2066), según se determina mediante la detección de la PARP rota (véase la figura 10, panel B). El 80% de las células murieron 3 días después de la transfección con ARNip de ROS, según se muestra en la figura 10, panel A. Estos resultados indican que la ROS quinasa mutante/truncada en la línea celular HCC78 está dirigiendo la proliferación y el crecimiento de estas células NSCLC, y que este crecimiento y proliferación pueden ser inhibidos empleando ARNip para inhibir la expresión de ROS quinasa.

**Ejemplo 4: Aislamiento y secuenciación del gen de fusión de SLC34A2-ROS**

Dada la presencia de la forma truncada de la ROS quinasas detectada en una línea celular de NSCLC (HCC78), se realizó una amplificación rápida 5' de los extremos del ADNc en la secuencia que codifica el dominio quinasas de ROS, para determinar si está presente un transcrito de ROS quimérico.

Amplificación rápida de los extremos del ADN complementario

Se empleó el kit RNeasy Mini Kit (Qiagen) para extraer el ARN de la línea celular HCC78. El ADN se extrajo empleando el kit DNeasy Tissue Kit (Qiagen). La amplificación rápida de los extremos del ADNc se realizó empleando el sistema 5' RACE (Invitrogen), con los cebadores ROS-GSP1 para la síntesis de ADNc y ROS-GSP2 y ROS-GSP3 para la reacción PCR anidada.

Ensayo de PCR

Para la RT-PCR, se sintetizó ADNc de primera hebra a partir de 2,5 µg de ARN total empleando el sistema de síntesis de primera hebra SuperScript™ III (Invitrogen) con oligo(dT)<sub>20</sub>. Después, el gen de fusión de SLC34A2-ROS se amplificó mediante el uso de las parejas de cebadores SLCROS-F1 y SLCROS-R1, SLCROS-F2 y SLCROS-R2.

#### Construcciones

- 5 El marco de lectura abierto del gen de fusión de SLC34A2-ROS se amplificó mediante PCR a partir del ADNc de células HCC78 empleando la ADN polimerasa de alta fidelidad Platinum Taq (Invitrogen) y las parejas de cebadores SLC-Fb y ROS-Rb (con un sitio de restricción Bgl II). Este producto de la PCR se clonó en el vector retroviral MSCV-Neo. Los cebadores fueron:

ROS-GSP1: ACCCTTCTCGGTTCTTCGTTTCCA (SEQ ID NO:13)

- 10 ROS-GSP2: GCAGCTCAGCCAACTCTTTGTCTT (SEQ ID NO: 14)

ROS-GSP3: TGCCAGACAAAGGTCAGTGGGATT (SEQ ID NO:15)

SLCROS-F1: TCCATCCCAGCACCTGCGGAG (SEQ ID NO:16)

SLCROS-R1: CTCAACTCTCTATTTCCCAAACAACGC (SEQ ID NO:17)

SLCROS-F2: CATGGCTCCCTGGCCTGAATTG (SEQ ID NO:18)

- 15 SLCROS-R2: CAACGCTATTAATCAGACCCATCTCC (SEQ ID NO:19)

SLC-Fb: GAAGATCTCTGACCATGGCTCCCTGGCCTGAA (SEQ ID NO:20)

ROS-Rb: GAAGATCTACGCTATTAATCAGACCCATCTCC (SEQ ID NO:21)

- 20 La figura 7 muestra la detección del producto de la amplificación de PCR después de 2 rondas. El análisis de la secuencia del producto resultante revela que el C-terminal de ROS está condensado con el N-terminal del gen SLC34A2 (véase la figura 1, panel B y C). El gen de fusión de SLC34A2-ROS está dentro del marco y condensado con los primeros 126 aminoácidos de SLC34A2 hasta los últimos 598 o 495 aminoácidos de ROS (véase la figura 1, panel B), respectivamente, lo cual resulta en dos variantes de la proteína de fusión (largo, corto). El SLC34A2 se localizó en el cromosoma 4p15, mientras que ROS está sobre el cromosoma 6q22. Así, se creó el gen de fusión con t(4;6)(p15;q22). Véase la figura 8, panel superior.

- 25 La fusión de SLC34A2 y ROS fue confirmada mediante PCR de transcriptasa inversa del ARN.

#### **Ejemplo 5: La proteína de fusión de SLC34A2-ROS dirige el crecimiento y la supervivencia de células 293 transfectadas**

- 30 Para confirmar que la expresión de la proteína de fusión de SLC34A2-ROS puede transformar células normales en un fenotipo canceroso, células de riñón embrionario humanas (células 293) se transfectaron con la construcción de ADNc descrita anteriormente, que codifica el variante largo de la proteína de fusión de SLC34A2-ROS.

- 35 La construcción del ADNc de SLC34A2-ROS descrita anteriormente (que codifica el variante largo de la proteína de fusión) se insertó en un vector de virus MSCV y se transfectó en células HEK293 empleando el reactivo de transfección SuperFect (Qiagen). Cuarenta y ocho horas después, las células HEK293 transfectadas se recolectaron y se comprobaron mediante transferencia Western para confirmar la expresión de la proteína de fusión de SLC34A2-ROS recombinante (variante largo) con el peso molecular esperado (véase la figura 9).

#### **Ejemplo 6: La proteína de fusión de SLC34A2-ROS dirige el crecimiento y la supervivencia de una línea de células de mamífero transformadas**

- 40 Para confirmar que la expresión de la proteína de fusión de SLC34A2-ROS puede transformar células normales en un fenotipo canceroso, pueden transformarse células 3T3 con una construcción de ADNc descrita anteriormente. Las células se mantienen en medio DMEM (Invitrogen) con suero de ternera fetal al 10% (FCS) (Invitrogen).

- 45 La producción del sobrenadante retroviral y la transducción se realizan como se ha descrito previamente. Véase Schwaller *et al.*, *Embo J.*, 17(18): 5321-5333 (1998). Las células 3T3 se transdujeron con el sobrenadante retroviral que contiene los vectores MSCV-Neo o MSCV-Neo/SLC34A2-ROS (largo) o MSCV-Neo/ROS (corto), respectivamente, y se seleccionaron para G418 (500 µg/ml). Las células transfectadas de forma estable se emplean en un ensayo de agar blando para confirmar que SLC34A2-ROS transformará a las células 3T3.

Este análisis confirmará que la expresión de la proteína de fusión de SLC34A2-ROS transforma a las células 3T3 de modo que el crecimiento celular sea independiente de la unión a un sustrato. Después se realiza un análisis de la transferencia Western para comprobar el estado de fosforilación de ROS, SLC34A2, SHP-1 y otras posibles dianas de ROS corriente abajo.

**Ejemplo 7: Detección de la expresión de la proteína de fusión de SLC34A2-ROS en una muestra de cáncer humano empleando el ensayo FISH**

La presencia de la proteína de fusión de SLC34A2-ROS en muestras de tumor NSCLC humano se detectó empleando un ensayo de hibridación de fluorescencia in situ (FISH), según se ha descrito previamente. Véase, por ejemplo, Verma *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press, Nueva York, N.Y. (1988). Se estudiaron más de 200 muestras de tumores NSCLC humano embebidas en parafina.

Para analizar las redistribuciones que implican a ROS, se diseñó una sonda de ruptura de dos colores. Una sonda proximal (BAC clon RP1-179P9) y dos sondas distales (BAC clon RP11-323017, RP1-94G16) se marcaron con Spectrum Orange dUTP o Spectrum Green dUTP, respectivamente. El marcaje de las sondas mediante FISH de desplazamiento de mella e interfase empleando secciones de tejido FFPE se realizó según las instrucciones del fabricanet (Vysis) con las siguientes modificaciones. Brevemente, las secciones de tejido embebidas en parafina se rehidrataron y se sometieron a la recuperación de los antígenos con microondas en tampón citrato 0,01 M (pH 6,0) durante 11 minutos. Las secciones se digirieron con proteasa (pepsina 4 mg/ml, 2000-3000 U/mg) durante 25 minutos a 37 °C, se deshidrataron y se hibridaron con la sonda FISH a 37 °C durante 18 horas. Después de lavar se aplicó 4',6-diamidino-2-fenilindol (DAPI; mg/ml) en medio de montaje Vectashield (Vector Laboratories, Burlingame, CA) para la contratinción nuclear.

La sonda de redistribución de ROS contiene dos sondas marcada de modo diferente en los lados opuestos del punto de rotura del gen ROS en la secuencia de tipo salvaje (véase la figura 4B y la figura 1). Cuando se hibrida, la región ROS nativa aparecerá como una señal de fusión naranja/verde, mientras que la redistribución en este locus (tal como aparece en la proteína de fusión de SLC34A2-ROS) producirá señales naranjas y verdes distintas. Véase la figura 11.

El análisis FISH reveló una baja incidencia de esta mutación de ROS en la muestra de población estudiada. Dos de 123 tumores, o 1,6% de los tumores, contenían la mutación de fusión. Sin embargo, dada la alta incidencia de NSCLC a nivel mundial (más de 151.000 nuevos casos anuales solo en EEUU), se espera que exista un número significativo de pacientes que porten esta ROS mutante, y estos pacientes pueden beneficiarse de un régimen terapéutico inhibidor de ROS.

**Ejemplo 8: Detección de la expresión de la ROS quinasa mutante en una muestra de cáncer humano empleando el ensayo de PCR**

La presencia de una ROS quinasa truncada y/o una proteína de fusión de SLC34A2-ROS en una muestra de cáncer humano puede detectarse empleando la reacción en cadena de la polimerasa (PCR) con transcriptasa inversa (RT) o genómica, tal como ha sido descrito previamente. Véase, por ejemplo, Cools *et al.*, *N. Engl. J. Med.*, 348: 1201-1214 (2003).

Brevemente y como ejemplo, pueden obtenerse muestras de tumor o de efusión pleural de un paciente que tiene NSCLC empleando técnicas convencionales. Se construyen sondas de PCR contra la ROS quinasa truncada o la proteína de fusión de SLC34A2-ROS. Puede emplearse el kit RNeasy Mini Kit (Qiagen) para extraer ARN de las muestras de tumor o de efusión pleural. El ADN puede extraerse empleando el kit DNeasy Tissue Kit (Qiagen). Para la RT-PCR, el ADNc de primer hebra se sintetiza, por ejemplo, a partir de 2,5 µg de ARN total empleando, por ejemplo, el sistema de síntesis de primera hebra SuperScript™ III (Invitrogen) con oligo(dT)<sub>20</sub>. Después, el gen de fusión de SLC34A2-ROS se amplifica mediante el uso de parejas de cebadores, por ejemplo, SLC34A2-F1 y ROS-P3 (véase el anterior ejemplos 4). Para la PCR genómica, la amplificación del gen de fusión puede realizarse empleando la ADN polimerasa Taq de alta fidelidad Platinum (Invitrogen) con parejas de cebadores, por ejemplo, gSLC34A2-F1 y gROS-R1, o Gslc34A2-F1 y gROS-R2 (véase el anterior ejemplo 4).

Este análisis identificará a un paciente que tenga un cáncer que se caracteriza por la expresión de la ROS quinasa truncada (y/o la proteína de fusión de SLC34A2-ROS), y este paciente es candidato a un tratamiento empleando un producto terapéutico inhibidor de ROS.

**Ejemplo 9: Detección de la expresión de la ROS quinasa mutante en una muestra de cáncer humano empleando la formación de perfiles de fosfopéptidos globales**

Para confirmar la incidencia de la mutación de fusión de ROS en NSCLC humano, se estudió un grupo de 34 tumores NSCLC humanos, empleando la técnica IAP de la formación de perfiles de fosfopéptidos globales descrita anteriormente (véase el ejemplo 1) para identificar los fosfopéptidos ROS en estos tumores. Las muestras de tumores (tumores diseccionados congelados inmediatamente y mantenidos en nitrógeno líquido) se obtuvieron de un colaborador clínico en China (Segundo Hospital Xiangya, Universidad Sur Central Changsha, Hunan).

Se homogeneizaron aproximadamente 300 miligramos de tejido congelado en 3 ml de tampón de lisis de urea empleando un homogeneizador Polytron. El lisado celular se aclaró, se redujo, se alquiló y después se digirió con tripsina durante la noche a temperatura ambiente. Estos 34 tumores se preseleccionaron para la señalización de fosfotirosina mediante inmunohistoquímica empleando protocolos convencionales y se determinó que eran



positivos.

5 La formación de perfiles de fosfotirosina global de estas muestras se realizó como se describió en el anterior ejemplo 1. Los resultados de la formación de los perfiles mostraron que una de estas 34 muestras presentaba fosfopéptidos de ROS y fosfopéptidos de SLC34A2 (véase la siguiente tabla 1 (los otros fosfopéptidos detectados no se muestran) y también moléculas corriente abajo, tales como fosfopéptidos IRS-1 e IRS-2. La firma del perfil de tirosina de este tumor es muy similar a la de la línea celular de NSCLC HCC78 (véase la tabla 1), tal como se esperaba. El análisis FISH también demostró que el tumor tenía una translocación de ROS (véase el ejemplo 7). Una RT-PCR y un análisis de secuenciación de ADN puede utilizarse para confirmar que la activación de ROS en este pacientes (y otros pacientes que portan la translocación de ROS) es debida al transcrito aberrante de SLC34A2/ROS.

10 Tabla 1. Formación de perfiles de fosfopéptidos de tumores NSCLC humanos

Nombre	Registro	Sitio	Péptidos	HCC78 (línea celular)	cs042 (tumor)
ROS	P08922	1923	GLAAGVGLANACyAIHTLPTQEEIENLPAFPR	1	1
ROS	P08922	2110	DIyKNDYyR; DIyKNDYyR; DIyKNDyYR; DIyKNDyyRKRGEGLLPVR	12	4
ROS	P08922	2114	DIYKNDyYR; DIyKNDyYR; DIyKNDyyRKRGEGLLPVR	11	3
ROS	P08922	2115	DIyKNDYyR; DIyKNDyyRKRGEGLLPVR	1	1
ROS	P08922	2274	EGLNyMVLATECGQGEEK; NREGLNyMVLATECGQGEEK; EGLNyMVLATECGQGEEKSEGPLGSQESESCGLR; NREGLNyMVLATECGQGEEKSEGPLGSQESESCGLR	20	
ROS	P08922	2323	QVAyCPSGKPEGLNyACLTHSGyGDGSD; QVAyCPSGKPEGLNyACLTHSGyGDGSD; QVAyCPSGKPEGLNyACLTHSGyGDGSD	4	1
ROS	P08922	2334	QVAYCPSGKPEGLNyACLTHSGyGDGSD; QVAYCPSGKPEGLNyACLTHSGyGDGSD; QVAyCPSGKPEGLNyACLTHSGyGDGSD	7	2
ROS	P08922	2342	QVAYCPSGKPEGLNyACLTHSGyGDGSD; QVAyCPSGKPEGLNyACLTHSGyGDGSD	3	
IRS-1	P35568	612	GGHHRPDSSTLHTDDGyMPMSPGVAPVPSGR		1
IRS-1	P35568	632	KGSGDyMPMSPK VDPNGyMMMSPSGGCSPIGGGPPSSSSSSNAVPSGT	2	1
IRS-1	P35568	662	SYGK	3	
IRS-2	Q9Y4H2	598	QRPVPQPSSASLDEyTLMR		1
IRS-2	Q9Y4H2	653	SSSSNLGADDGyMPMTPGAALAGSGSGSCR	4	5
IRS-2	Q9Y4H2	675	SDDyMPMSPASVSAPK	3	4
IRS-2	Q9Y4H2	742	ASSPAESSPEDSGyMR	3	3
IRS-2	Q9Y4H2	823	APYTCGGSDSDQyVLMSSPVGR; SYKAPYTCGGSDSDQyVLMSSPVGR	2	5
SLC34A2	O95436	54	IELLPSySTATLIDEPTVDDPWNLPTLQDSGIK	1	1

LISTA DE SECUENCIAS

<110> CELL SIGNALING TECHNOLOGY, INC. GUO, Ailan POSSEMATO, Anthony

<120> ROS quinasa mutante y de translocación en el carcinoma pulmonar no microcítico humano

<130> CST-232

<150> 60/760,634

<151> 20-01-2006

<160> 21

<170> PatentIn versión 3.3

<210> 1

<211> 724

<212> PRT

<213> Homo sapiens

<400> 1

Met Ala Pro Trp Pro Glu Leu Gly Asp Ala Gln Pro Asn Pro Asp Lys  
 1 5 10 15

Tyr Leu Glu Gly Ala Ala Gly Gln Gln Pro Thr Ala Pro Asp Lys Ser  
 20 25 30

Lys Glu Thr Asn Lys Thr Asp Asn Thr Glu Ala Pro Val Thr Lys Ile  
 35 40 45

Glu Leu Leu Pro Ser Tyr Ser Thr Ala Thr Leu Ile Asp Glu Pro Thr  
 50 55 60

Glu Val Asp Asp Pro Trp Asn Leu Pro Thr Leu Gln Asp Ser Gly Ile  
 65 70 75 80

Lys Trp Ser Glu Arg Asp Thr Lys Gly Lys Ile Leu Cys Phe Phe Gln  
 85 90 95

Gly Ile Gly Arg Leu Ile Leu Leu Leu Gly Phe Leu Tyr Phe Phe Val  
 100 105 110

Cys Ser Leu Asp Ile Leu Ser Ser Ala Phe Gln Leu Val Gly Ala Gly



Tyr Glu Gly Thr Ala Val Asp Ile Leu Gly Val Gly Ser Gly Glu Ile  
 340 345 350

Lys Val Ala Val Lys Thr Leu Lys Lys Gly Ser Thr Asp Gln Glu Lys  
 355 360 365

Ile Glu Phe Leu Lys Glu Ala His Leu Met Ser Lys Phe Asn His Pro  
 370 375 380

Asn Ile Leu Lys Gln Leu Gly Val Cys Leu Leu Asn Glu Pro Gln Tyr  
 385 390 395 400

Ile Ile Leu Glu Leu Met Glu Gly Gly Asp Leu Leu Thr Tyr Leu Arg  
 405 410 415

Lys Ala Arg Met Ala Thr Phe Tyr Gly Pro Leu Leu Thr Leu Val Asp  
 420 425 430

Leu Val Asp Leu Cys Val Asp Ile Ser Lys Gly Cys Val Tyr Leu Glu  
 435 440 445

Arg Met His Phe Ile His Arg Asp Leu Ala Ala Arg Asn Cys Leu Val  
 450 455 460

Ser Val Lys Asp Tyr Thr Ser Pro Arg Ile Val Lys Ile Gly Asp Phe  
 465 470 475 480

Gly Leu Ala Arg Asp Ile Tyr Lys Asn Asp Tyr Tyr Arg Lys Arg Gly  
 485 490 495

Glu Gly Leu Leu Pro Val Arg Trp Met Ala Pro Glu Ser Leu Met Asp  
 500 505 510

Gly Ile Phe Thr Thr Gln Ser Asp Val Trp Ser Phe Gly Ile Leu Ile  
 515 520 525

Trp Glu Ile Leu Thr Leu Gly His Gln Pro Tyr Pro Ala His Ser Asn  
 530 535 540

Leu Asp Val Leu Asn Tyr Val Gln Thr Gly Gly Arg Leu Glu Pro Pro



ES 2 539 830 T3

gccgcaggtc agcagcccaac tgcccctgat aaaagcaaag agaccaacaa aacagataac 120  
 actgaggcac ctgtaacca gattgaactt ctgccgtcct actccacggc tacactgata 180  
 gatgagccca ctgaggtgga tgaccctgg aacctacca ctcttcagga ctcggggatc 240  
 aagtggtcag agagagacac caaagggag attctctgtt tcttccaagg gattgggaga 300  
 ttgattttac ttctcggatt tctctacttt ttcgtgtgct ccctggatat tcttagtagc 360  
 gccttcagc tggttggagc tggagtcca aataaaccag gcattcccaa attactagaa 420  
 gggagtaaaa attcaatata gtgggagaaa gctgaagata atggatgtag aattacatac 480  
 tatatccttg agataagaaa gagcacttca aataatttac agaaccagaa ttaaggtgg 540  
 aagatgacat ttaatggatc ctgcagtagt gtttgcacat ggaagtccaa aaacctgaaa 600  
 ggaatatttc agttcagagt agtagctgca aataatctag ggtttgggta atatagtgga 660  
 atcagtgaga atattatatt agttggagat gatttttggg taccagaaac aagtttcata 720  
 ctactatta tagttggaat atttctgggt gttacaatcc cactgacctt tgtctggcat 780  
 agaagattaa agaatcaaaa aagtgccaaag gaaggggtga cagtgcctat aaacgaagac 840  
 aaagagttgg ctgagctgag aggtctggca gccggagtag gcctggctaa tgctgctat 900  
 gcaatacata ctcttccaac ccaagaggag attgaaatc ttctgcctt cctcgggaa 960  
 aaactgaetc tgcgtctctt gctgggaagt ggagcctttg gagaagtgta tgaaggaaca 1020  
 gcagtggaca tcttaggagt tggaaagtga gaaatcaaag tagcagtga gactttgaag 1080  
 aagggttcca cagaccagga gaagattgaa ttcctgaagg aggcacatct gatgagcaaa 1140  
 tttaatcatc ccaacattct gaagcagctt ggagtttgtc tgctgaatga accccaatac 1200  
 attatcctgg aactgatgga gggaggagac ctcttactt atttgcgtaa agcccggatg 1260  
 gcaacgtttt atggtccttt actcaccttg gttgaccttg tagacctgtg tgtagatatt 1320  
 tcaaaaggct gtgtctactt ggaacggatg catttcattc acagggatct ggcagctaga 1380  
 aattgccttg tttccgtgaa agactatacc agtccacgga tagtgaagat tggagacttt 1440  
 ggactcgcca gagacatcta taaaantgat tactatagaa agagagggga aggctgctc 1500  
 ccagttcggg gatggctcc agaaagttt atggatgga tcttactac tcaatctgat 1560  
 gtatggtctt ttggaattct gatttgggag attttaactc ttggtcatca gccttatcca 1620  
 gtcattcca accttgatgt gttaaactat gtgcaaacag gaggagact ggagccacca 1680

agaaattgtc ctgatgatct gtggaattta atgacccagt gctgggctca agaaccgcac 1740  
 caaagaccta cttttcatag aattcaggac caacttcagt tattcagaaa ttttttctta 1800  
 aatagcattt ataagtccag agatgaagca aacaacagtg gagtcataaa tgaagcctt 1860  
 gaaggtgaag atggcgatgt gatttgtttg aattcagatg acattatgcc agttgcttta 1920  
 atggaaacga agaaccgaga agggttaaac tatatggtac ttgctacaga atgtggccaa 1980  
 ggtgaagaaa agtctgaggg tcctctaggc tcccaggaat ctgaatcttg tggctgagg 2040  
 aaagaagaga aggaaccaca tgcagacaaa gatttctgcc aagaaaaaca agtggcttac 2100  
 tgccttctg gcaagcctga aggcctgaac tatgcctgtc tcactcacag tggatatgga 2160  
 gatgggtctg attaa 2175

<210> 3  
 <211> 621  
 <212> PRT  
 <213> Homo sapiens

5

<400> 3  
 Met Ala Pro Trp Pro Glu Leu Gly Asp Ala Gln Pro Asn Pro Asp Lys  
 1 5 10 15

Tyr Leu Glu Gly Ala Ala Gly Gln Gln Pro Thr Ala Pro Asp Lys Ser  
 20 25 30

Lys Glu Thr Asn Lys Thr Asp Asn Thr Glu Ala Pro Val Thr Lys Ile  
 35 40 45

Glu Leu Leu Pro Ser Tyr Ser Thr Ala Thr Leu Ile Asp Glu Pro Thr  
 50 55 60

Glu Val Asp Asp Pro Trp Asn Leu Pro Thr Leu Gln Asp Ser Gly Ile  
 65 70 75 80

Lys Trp Ser Glu Arg Asp Thr Lys Gly Lys Ile Leu Cys Phe Phe Gln  
 85 90 95

Gly Ile Gly Arg Leu Ile Leu Leu Leu Gly Phe Leu Tyr Phe Phe Val  
 100 105 110

10

ES 2 539 830 T3

Cys Ser Leu Asp Ile Leu Ser Ser Ala Phe Gln Leu Val Gly Asp Asp  
 115 120 125

Phe Trp Ile Pro Glu Thr Ser Phe Ile Leu Thr Ile Ile Val Gly Ile  
 130 135 140

Phe Leu Val Val Thr Ile Pro Leu Thr Phe Val Trp His Arg Arg Leu  
 145 150 155 160

Lys Asn Gln Lys Ser Ala Lys Glu Gly Val Thr Val Leu Ile Asn Glu  
 165 170 175

Asp Lys Glu Leu Ala Glu Leu Arg Gly Leu Ala Ala Gly Val Gly Leu  
 180 185 190

Ala Asn Ala Cys Tyr Ala Ile His Thr Leu Pro Thr Gln Glu Glu Ile  
 195 200 205

Glu Asn Leu Pro Ala Phe Pro Arg Glu Lys Leu Thr Leu Arg Leu Leu  
 210 215 220

Leu Gly Ser Gly Ala Phe Gly Glu Val Tyr Glu Gly Thr Ala Val Asp  
 225 230 235 240

Ile Leu Gly Val Gly Ser Gly Glu Ile Lys Val Ala Val Lys Thr Leu  
 245 250 255

Lys Lys Gly Ser Thr Asp Gln Glu Lys Ile Glu Phe Leu Lys Glu Ala  
 260 265 270

His Leu Met Ser Lys Phe Asn His Pro Asn Ile Leu Lys Gln Leu Gly  
 275 280 285

Val Cys Leu Leu Asn Glu Pro Gln Tyr Ile Ile Leu Glu Leu Met Glu  
 290 295 300

Gly Gly Asp Leu Leu Thr Tyr Leu Arg Lys Ala Arg Met Ala Thr Phe  
 305 310 315 320

Tyr Gly Pro Leu Leu Thr Leu Val Asp Leu Val Asp Leu Cys Val Asp



ES 2 539 830 T3

				325						330					335
Ile	Ser	Lys	Gly	Cys	Val	Tyr	Leu	Glu	Arg	Met	His	Phe	Ile	His	Arg
			340					345					350		
Asp	Leu	Ala	Ala	Arg	Asn	Cys	Leu	Val	Ser	Val	Lys	Asp	Tyr	Thr	Ser
		355					360					365			
Pro	Arg	Ile	Val	Lys	Ile	Gly	Asp	Phe	Gly	Leu	Ala	Arg	Asp	Ile	Tyr
	370					375					380				
Lys	Asn	Asp	Tyr	Tyr	Arg	Lys	Arg	Gly	Glu	Gly	Leu	Leu	Pro	Val	Arg
385					390					395					400
Trp	Met	Ala	Pro	Glu	Ser	Leu	Met	Asp	Gly	Ile	Phe	Thr	Thr	Gln	Ser
				405					410					415	
Asp	Val	Trp	Ser	Phe	Gly	Ile	Leu	Ile	Trp	Glu	Ile	Leu	Thr	Leu	Gly
			420					425					430		
His	Gln	Pro	Tyr	Pro	Ala	His	Ser	Asn	Leu	Asp	Val	Leu	Asn	Tyr	Val
		435					440					445			
Gln	Thr	Gly	Gly	Arg	Leu	Glu	Pro	Pro	Arg	Asn	Cys	Pro	Asp	Asp	Leu
	450					455					460				
Trp	Asn	Leu	Met	Thr	Gln	Cys	Trp	Ala	Gln	Glu	Pro	Asp	Gln	Arg	Pro
465					470					475					480
Thr	Phe	His	Arg	Ile	Gln	Asp	Gln	Leu	Gln	Leu	Phe	Arg	Asn	Phe	Phe
				485					490					495	
Leu	Asn	Ser	Ile	Tyr	Lys	Ser	Arg	Asp	Glu	Ala	Asn	Asn	Ser	Gly	Val
			500					505					510		
Ile	Asn	Glu	Ser	Phe	Glu	Gly	Glu	Asp	Gly	Asp	Val	Ile	Cys	Leu	Asn
	515						520					525			
Ser	Asp	Asp	Ile	Met	Pro	Val	Ala	Leu	Met	Glu	Thr	Lys	Asn	Arg	Glu
	530					535					540				

ES 2 539 830 T3

Gly Leu Asn Tyr Met Val Leu Ala Thr Glu Cys Gly Gln Gly Glu Glu  
545 550 555 560

Lys Ser Glu Gly Pro Leu Gly Ser Gln Glu Ser Glu Ser Cys Gly Leu  
565 570 575

Arg Lys Glu Glu Lys Glu Pro His Ala Asp Lys Asp Phe Cys Gln Glu  
580 585 590

Lys Gln Val Ala Tyr Cys Pro Ser Gly Lys Pro Glu Gly Leu Asn Tyr  
595 600 605

Ala Cys Leu Thr His Ser Gly Tyr Gly Asp Gly Ser Asp  
610 615 620

<210> 4  
<211> 1866  
5 <212> DNA  
<213> Homo sapiens

<400> 4  
atggctccct ggcctgaatt gggagatgcc cagcccaacc ccgataagta cctcgaaggg 60  
gccgcaggtc agcagcccac tgcccctgat aaaagcaaag agaccaacaa aacagataac 120  
actgaggcac ctgtaaccaa gattgaactt ctgccgtcct actccacggc tacactgata 180  
gatgagccca ctgagggtgga tgaccctctg aacctacca ctcttcagga ctcggggatc 240  
aagtggtcag agagagacac caaaggggaag attctctgtt tcttccaagg gattgggaga 300  
ttgattttac ttctcggatt tctctacttt ttcgtgtgct ccctggatat tcttagtagc 360  
gccttcacgc tggttgagga tgatttttgg ataccagaaa caagtttcat acttactatt 420  
atagttggaa tatttctggt tgttacaatc ccaactgacct ttgtctggca tagaagatta 480  
aagaatcaaa aaagtgccaa ggaaggggtg acagtgtcta taaacgaaga caaagagttg 540  
gctgagctgc gaggtctggc agccggagta ggectggcta atgectgcta tgcaatacat 600  
actcttccaa cccaagagga gattgaaaat cttctgcct tcctcggga aaaactgact 660  
ctcgtctct tgctgggaag tggagccttt ggagaagtgt atgaaggac agcagtggac 720  
atcttaggag ttggaagtgg agaaatcaaa gtagcagtga agactttgaa gaagggttcc 780  
acagaccagg agaagattga attcctgaag gaggcacatc tgatgagcaa atttaacat 840

10

ES 2 539 830 T3

cccaacattc tgaagcagct tggagtttgt ctgctgaatg aacccaata cattatcctg 900  
 gaactgatgg agggaggaga ccttcttact tatttgcgta aagcccggat ggcaacgttt 960  
 tatggtcctt tactcacctt ggttgacctt gtagacctgt gtgtagatat ttcaaaaggc 1020  
 tgtgtctact tggaaaggat gcatttcatt cacagggatc tggcagctag aaattgcctt 1080  
 gtttccgtga aagactatac cagtccacgg atagtgaaga ttggagactt tggactcgcc 1140  
 agagacatct ataaaaatga ttactataga aagagagggg aaggcctgct cccagttcgg 1200  
 tggatggctc cagoaagttt gatggatgga atcttcacta ctcaatctga tgtatggctt 1260  
 tttggaattc tgatttggga gattttaact cttggtcatc agccttatcc agctcattcc 1320  
 aaccttgatg tgttaaacta tgtgcaaaca ggagggagac tggagccacc aagaattgt 1380  
 cctgatgatc tgtggaattt aatgaccagc tgctgggctc aagaaccgga ccaagacct 1440  
 acttttcata gaattcagga ccaacttcag ttattcagaa attttttctt aaatagcatt 1500  
 tataagtcca gagatgaagc aaacaacagt ggagtcataa atgaaagctt tgaagtgaa 1560  
 gatggcgatg tgatttgttt gaattcagat gacattatgc cagttgcttt aatggaaacg 1620  
 aagaaccgag aagggttaaa ctatatggta cttgctacag aatgtggcca aggtgaagaa 1680  
 aagtctgagg gtcctctagg ctcccaggaa tctgaatctt gtggtctgag gaaagaagag 1740  
 aaggaaccac atgcagacaa agatttctgc caagaaaaac aagtggctta ctgcccttct 1800  
 ggcaagcctg aaggcctgaa ctatgcctgt ctcactcaca gtggatatgg agatgggtct 1860  
 gattaa 1866

<210> 5  
 <211> 690  
 <212> PRT  
 <213> Homo sapiens

<400> 5

Met Ala Pro Trp Pro Glu Leu Gly Asp Ala Gln Pro Asn Pro Asp Lys  
 1 5 10 15

Tyr Leu Glu Gly Ala Ala Gly Gln Gln Pro Thr Ala Pro Asp Lys Ser  
 20 25 30

Lys Glu Thr Asn Lys Thr Asp Asn Thr Glu Ala Pro Val Thr Lys Ile

5

10



ES 2 539 830 T3

Gly Glu Asp Ala Pro Asp Leu Leu Lys Val Ile Thr Lys Pro Phe Thr  
 260 265 270

Lys Leu Ile Val Gln Leu Asp Lys Lys Val Ile Ser Gln Ile Ala Met  
 275 280 285

Asn Asp Glu Lys Ala Lys Asn Lys Ser Leu Val Lys Ile Trp Cys Lys  
 290 295 300

Thr Phe Thr Asn Lys Thr Gln Ile Asn Val Thr Val Pro Ser Thr Ala  
 305 310 315 320

Asn Cys Thr Ser Pro Ser Leu Cys Trp Thr Asp Gly Ile Gln Asn Trp  
 325 330 335

Thr Met Lys Asn Val Thr Tyr Lys Glu Asn Ile Ala Lys Cys Gln His  
 340 345 350

Ile Phe Val Asn Phe His Leu Pro Asp Leu Ala Val Gly Thr Ile Leu  
 355 360 365

Leu Ile Leu Ser Leu Leu Val Leu Cys Gly Cys Leu Ile Met Ile Val  
 370 375 380

Lys Ile Leu Gly Ser Val Leu Lys Gly Gln Val Ala Thr Val Ile Lys  
 385 390 395 400

Lys Thr Ile Asn Thr Asp Phe Pro Phe Pro Phe Ala Trp Leu Thr Gly  
 405 410 415

Tyr Leu Ala Ile Leu Val Gly Ala Gly Met Thr Phe Ile Val Gln Ser  
 420 425 430

Ser Ser Val Phe Thr Ser Ala Leu Thr Pro Leu Ile Gly Ile Gly Val  
 435 440 445

Ile Thr Ile Glu Arg Ala Tyr Pro Leu Thr Leu Gly Ser Asn Ile Gly  
 450 455 460

Thr Thr Thr Thr Ala Ile Leu Ala Ala Leu Ala Ser Pro Gly Asn Ala



ES 2 539 830 T3

<213> Homo sapiens

<400> 6

cgggccaggt ttccaggctc ggccgccgcc tccatcccag cacctgcgga gggagcgtg 60  
 accatggctc cctggcctga attgggagat gccagccca accccgataa gtacctgaa 120  
 ggggccgag gtcagcagcc cactgccctt gataaaagca aagagacca caaaacagat 180  
 aacactgagg cacctgtaac caagattgaa cttctgccgt cctactccac ggctacactg 240  
 atagatgagc ccactgaggt ggatgacccc tggaacctac ccactcttca ggactcgggg 300  
 atcaagtggc cagagagaga caccaaaggg aagattctct gtttcttcca agggattggg 360  
 agattgattt tacttctcgg atttctctac ttttctgtgt gctccctgga tattcttagt 420  
 agcgccttcc agctggttgg aggaaaaatg gcaggacagt tcttcagca cagctctatt 480  
 atgtccaacc ctttgttggg gctggtgatc ggggtgctgg tgaccgtctt ggtgcagagc 540  
 tccagcacct caacgtccat cgttgtcagc atggtgtcct cttcattgct cactgttcgg 600  
 gctgccatcc ccattatcat gggggccaac attggaacgt caatcacca cactattgtt 660  
 gcgctcatgc aggtgggaga tcggagtgag ttcagaagag cttttgcagg agccaactgc 720  
 catgacttct tcaactgget gtccgtgttg gtgctcttgc ccgtggagggt ggccacccat 780  
 tacctcgaga tcataacca gcttatagtg gagagcttcc acttcaagaa tggagaagat 840  
 gccccagatc ttctgaaagt catcactaag cccttcacaa agctcattgt ccagctggat 900  
 aaaaaagtta tcagccaat tgcaatgaa gatgaaaag cgaaaaacaa gagtcttgtc 960  
 aagatttggc gcaaaacttt taccaacaag acccagatta acgtcactgt tccctcgact 1020  
 gctaactgca cctcccctc cctctgttgg acggatggca tccaaaactg gaccatgaag 1080  
 aatgtgacct acaaggagaa catcgccaaa tgccagcata tctttgtgaa tttcacctc 1140  
 ccggatcttg ctgtgggac catcttctc atactctccc tgctggtcct ctgtggttgc 1200  
 ctgatcatga ttgtcaagat cctgggctct gtgctcaagg ggcaggctgc cactgtcatc 1260  
 aagaagacca tcaacactga tttcccctt cccttgcac ggttgactgg ctacctggcc 1320

5

atcctcgctg gggcaggcat gaccttcac gtacagagca gctctgtggt cacgtcggcc 1380  
 ttgaccccc tgattggaat cggcgtgata accattgaga gggcttatcc actcacgctg 1440  
 ggctccaaca tcggcaccac caccaccgcc atcctggccg ccttagccag ccctggcaat 1500  
 gcattgagga gttcactcca gatcgcctg tgcactttt tcttcaacat ctccggcatc 1560  
 ttgtgtggt acccgatccc gttcactcgc ctgccatcc gcatggccaa ggggctgggc 1620  
 aacatctctg ccaagtatcg ctggttcgcc gtcttctacc tgatcatctt cttcttctg 1680  
 atcccgctga cgggtgttgg cctctcgctg gccggctggc ggggtgctggt tgggtgcggg 1740  
 gttcccgtcg tcttcatcat cctcctggta ctgtgcctcc gactcctgca gtctcgctgc 1800  
 ccacgcgtcc tgccgaagaa actccagaac tggaaacttc tgccgctgtg gatgcgctcg 1860  
 ctgaagccct gggatgccgt cgtctccaag ttcaccggct gcttccagat gcgctgctgc 1920  
 tactgctgcc gcgtgtgctg ccgcgcgtgc tgcttctgt gtggctgccc caagtgctgc 1980  
 cgctgcagca agtgctgcga ggaactggag gaggcgcagg aggggcagga tgtccctgtc 2040  
 aaggctcctg agaccttga taacataacc attagcagag aggctcaggg tgaggtecc 2100  
 gcctcggact caaagaccga atgcacggcc ttgtagggga cgcccagat tgcagggat 2160  
 ggggggatgg tccttgagtt ttgatgctc tcctccctcc cacttctgca ccctttcacc 2220  
 acctcgagga gatttgcctc ccattagcga atgaattga tgcagtccca aaaaaaaaa 2280

<210> 7  
 <211> 2347  
 <212> PRT  
 <213> Homo sapiens

5

<400> 7  
 Met Lys Asn Ile Tyr Cys Leu Ile Pro Lys Leu Val Asn Phe Ala Thr  
 1 5 10 15

Leu Gly Cys Leu Trp Ile Ser Val Val Gln Cys Thr Val Leu Asn Ser  
 20 25 30

Cys Leu Lys Ser Cys Val Thr Asn Leu Gly Gln Gln Leu Asp Leu Gly  
 35 40 45

Thr Pro His Asn Leu Ser Glu Pro Cys Ile Gln Gly Cys His Phe Trp

10





Pro Glu Ala Glu Ser Ser Ile Thr Thr Ser Ser Ser Ala Val Gln Gln  
 275 280 285

Glu Glu Gln Trp Leu Phe Leu Ser Arg Lys Thr Ser Leu Arg Lys Arg  
 290 295 300

Ser Leu Lys His Leu Val Asp Glu Ala His Cys Leu Arg Leu Asp Ala  
 305 310 315 320

Ile Tyr His Asn Ile Thr Gly Ile Ser Val Asp Val His Gln Gln Ile  
 325 330 335

Val Tyr Phe Ser Glu Gly Thr Leu Ile Trp Ala Lys Lys Ala Ala Asn  
 340 345 350

Met Ser Asp Val Ser Asp Leu Arg Ile Phe Tyr Arg Gly Ser Gly Leu  
 355 360 365

Ile Ser Ser Ile Ser Ile Asp Trp Leu Tyr Gln Arg Met Tyr Phe Ile  
 370 375 380

Met Asp Glu Leu Val Cys Val Cys Asp Leu Glu Asn Cys Ser Asn Ile  
 385 390 395 400

Glu Glu Ile Thr Pro Pro Ser Ile Ser Ala Pro Gln Lys Ile Val Ala  
 405 410 415

Asp Ser Tyr Asn Gly Tyr Val Phe Tyr Leu Leu Arg Asp Gly Ile Tyr  
 420 425 430

Arg Ala Asp Leu Pro Val Pro Ser Gly Arg Cys Ala Glu Ala Val Arg  
 435 440 445

Ile Val Glu Ser Cys Thr Leu Lys Asp Phe Ala Ile Lys Pro Gln Ala  
 450 455 460

Lys Arg Ile Ile Tyr Phe Asn Asp Thr Ala Gln Val Phe Met Ser Thr  
 465 470 475 480

Phe Leu Asp Gly Ser Ala Ser His Leu Ile Leu Pro Arg Ile Pro Phe

				485						490										495	
Ala	Asp	Val	Lys	Ser	Phe	Ala	Cys	Glu	Asn	Asn	Asp	Phe	Leu	Val	Thr						
			500					505					510								
Asp	Gly	Lys	Val	Ile	Phe	Gln	Gln	Asp	Ala	Leu	Ser	Phe	Asn	Glu	Phe						
		515					520					525									
Ile	Val	Gly	Cys	Asp	Leu	Ser	His	Ile	Glu	Glu	Phe	Gly	Phe	Gly	Asn						
	530					535					540										
Leu	Val	Ile	Phe	Gly	Ser	Ser	Ser	Gln	Leu	His	Pro	Leu	Pro	Gly	Arg						
545					550					555					560						
Pro	Gln	Glu	Leu	Ser	Val	Leu	Phe	Gly	Ser	His	Gln	Ala	Leu	Val	Gln						
				565					570					575							
Trp	Lys	Pro	Pro	Ala	Leu	Ala	Ile	Gly	Ala	Asn	Val	Ile	Leu	Ile	Ser						
			580					585					590								
Asp	Ile	Ile	Glu	Leu	Phe	Glu	Leu	Gly	Pro	Ser	Ala	Trp	Gln	Asn	Trp						
		595					600					605									
Thr	Tyr	Glu	Val	Lys	Val	Ser	Thr	Gln	Asp	Pro	Pro	Glu	Val	Thr	His						
	610					615					620										
Ile	Phe	Leu	Asn	Ile	Ser	Gly	Thr	Met	Leu	Asn	Val	Pro	Glu	Leu	Gln						
625					630					635					640						
Ser	Ala	Met	Lys	Tyr	Lys	Val	Ser	Val	Arg	Ala	Ser	Ser	Pro	Lys	Arg						
				645					650					655							
Pro	Gly	Pro	Trp	Ser	Glu	Pro	Ser	Val	Gly	Thr	Thr	Leu	Val	Pro	Ala						
			660					665					670								
Ser	Glu	Pro	Pro	Phe	Ile	Met	Ala	Val	Lys	Glu	Asp	Gly	Leu	Trp	Ser						
		675					680					685									
Lys	Pro	Leu	Asn	Ser	Phe	Gly	Pro	Gly	Glu	Phe	Leu	Ser	Ser	Asp	Ile						
	690					695					700										

Gly Asn Val Ser Asp Met Asp Trp Tyr Asn Asn Ser Leu Tyr Tyr Ser  
 705 710 715 720

Asp Thr Lys Gly Asp Val Phe Val Trp Leu Leu Asn Gly Thr Asp Ile  
 725 730 735

Ser Glu Asn Tyr His Leu Pro Ser Ile Ala Gly Ala Gly Ala Leu Ala  
 740 745 750

Phe Glu Trp Leu Gly His Phe Leu Tyr Trp Ala Gly Lys Thr Tyr Val  
 755 760 765

Ile Gln Arg Gln Ser Val Leu Thr Gly His Thr Asp Ile Val Thr His  
 770 775 780

Val Lys Leu Leu Val Asn Asp Met Val Val Asp Ser Val Gly Gly Tyr  
 785 790 795 800

Leu Tyr Trp Thr Thr Leu Tyr Ser Val Glu Ser Thr Arg Leu Asn Gly  
 805 810 815

Glu Ser Ser Leu Val Leu Gln Thr Gln Pro Trp Phe Ser Gly Lys Lys  
 820 825 830

Val Ile Ala Leu Thr Leu Asp Leu Ser Asp Gly Leu Leu Tyr Trp Leu  
 835 840 845

Val Gln Asp Ser Gln Cys Ile His Leu Tyr Thr Ala Val Leu Arg Gly  
 850 855 860

Gln Ser Thr Gly Asp Thr Thr Ile Thr Glu Phe Ala Ala Trp Ser Thr  
 865 870 875 880

Ser Glu Ile Ser Gln Asn Ala Leu Met Tyr Tyr Ser Gly Arg Leu Phe  
 885 890 895

Trp Ile Asn Gly Phe Arg Ile Ile Thr Thr Gln Glu Ile Gly Gln Lys  
 900 905 910

Thr Ser Val Ser Val Leu Glu Pro Ala Arg Phe Asn Gln Phe Thr Ile

ES 2 539 830 T3

915	920	925																		
Ile	Gln	Thr	Ser	Leu	Lys	Pro	Leu	Pro	Gly	Asn	Phe	Ser	Phe	Thr	Pro					
930						935					940									
Lys	Val	Ile	Pro	Asp	Ser	Val	Gln	Glu	Ser	Ser	Phe	Arg	Ile	Glu	Gly					
945					950					955					960					
Asn	Ala	Ser	Ser	Phe	Gln	Ile	Leu	Trp	Asn	Gly	Pro	Pro	Ala	Val	Asp					
				965					970						975					
Trp	Gly	Val	Val	Phe	Tyr	Ser	Val	Glu	Phe	Ser	Ala	His	Ser	Lys	Phe					
		980						985						990						
Leu	Ala	Ser	Glu	Gln	His	Ser	Leu	Pro	Val	Phe	Thr	Val	Glu	Gly	Leu					
		995					1000					1005								
Glu	Pro	Tyr	Ala	Leu	Phe	Asn	Leu	Ser	Val	Thr	Pro	Tyr	Thr	Tyr						
1010						1015					1020									
Trp	Gly	Lys	Gly	Pro	Lys	Thr	Ser	Leu	Ser	Leu	Arg	Ala	Pro	Glu						
1025						1030					1035									
Thr	Val	Pro	Ser	Ala	Pro	Glu	Asn	Pro	Arg	Ile	Phe	Ile	Leu	Pro						
1040						1045					1050									
Ser	Gly	Lys	Cys	Cys	Asn	Lys	Asn	Glu	Val	Val	Val	Glu	Phe	Arg						
1055						1060					1065									
Trp	Asn	Lys	Pro	Lys	His	Glu	Asn	Gly	Val	Leu	Thr	Lys	Phe	Glu						
1070						1075					1080									
Ile	Phe	Tyr	Asn	Ile	Ser	Asn	Gln	Ser	Ile	Thr	Asn	Lys	Thr	Cys						
1085						1090					1095									
Glu	Asp	Trp	Ile	Ala	Val	Asn	Val	Thr	Pro	Ser	Val	Met	Ser	Phe						
1100						1105					1110									
Gln	Leu	Glu	Gly	Met	Ser	Pro	Arg	Cys	Phe	Ile	Ala	Phe	Gln	Val						
1115						1120					1125									

ES 2 539 830 T3

Arg Ala Phe Thr Ser Lys Gly Pro Gly Pro Tyr Ala Asp Val Val  
 1130 1135 1140

Lys Ser Thr Thr Ser Glu Ile Asn Pro Phe Pro His Leu Ile Thr  
 1145 1150 1155

Leu Leu Gly Asn Lys Ile Val Phe Leu Asp Met Asp Gln Asn Gln  
 1160 1165 1170

Val Val Trp Thr Phe Ser Ala Glu Arg Val Ile Ser Ala Val Cys  
 1175 1180 1185

Tyr Thr Ala Asp Asn Glu Met Gly Tyr Tyr Ala Glu Gly Asp Ser  
 1190 1195 1200

Leu Phe Leu Leu His Leu His Asn Arg Ser Ser Ser Glu Leu Phe  
 1205 1210 1215

Gln Asp Ser Leu Val Phe Asp Ile Thr Val Ile Thr Ile Asp Trp  
 1220 1225 1230

Ile Ser Arg His Leu Tyr Phe Ala Leu Lys Glu Ser Gln Asn Gly  
 1235 1240 1245

Met Gln Val Phe Asp Val Asp Leu Glu His Lys Val Lys Tyr Pro  
 1250 1255 1260

Arg Glu Val Lys Ile His Asn Arg Asn Ser Thr Ile Ile Ser Phe  
 1265 1270 1275

Ser Val Tyr Pro Leu Leu Ser Arg Leu Tyr Trp Thr Glu Val Ser  
 1280 1285 1290

Asn Phe Gly Tyr Gln Met Phe Tyr Tyr Ser Ile Ile Ser His Thr  
 1295 1300 1305

Leu His Arg Ile Leu Gln Pro Thr Ala Thr Asn Gln Gln Asn Lys  
 1310 1315 1320

Arg Asn Gln Cys Ser Cys Asn Val Thr Glu Phe Glu Leu Ser Gly

ES 2 539 830 T3

1325		1330		1335
Ala Met	Ala Ile Asp Thr Ser	Asn Leu Glu Lys Pro	Leu Ile Tyr	
1340		1345	1350	
Phe Ala	Lys Ala Gln Glu Ile	Trp Ala Met Asp	Leu Glu Gly Cys	
1355		1360	1365	
Gln Cys	Trp Arg Val Ile Thr	Val Pro Ala Met	Leu Ala Gly Lys	
1370		1375	1380	
Thr Leu	Val Ser Leu Thr Val	Asp Gly Asp Leu Ile	Tyr Trp Ile	
1385		1390	1395	
Ile Thr	Ala Lys Asp Ser Thr	Gln Ile Tyr Gln Ala	Lys Lys Gly	
1400		1405	1410	
Asn Gly	Ala Ile Val Ser Gln	Val Lys Ala Leu Arg	Ser Arg His	
1415		1420	1425	
Ile Leu	Ala Tyr Ser Ser Val	Met Gln Pro Phe Pro	Asp Lys Ala	
1430		1435	1440	
Phe Leu	Ser Leu Ala Ser Asp	Thr Val Glu Pro Thr	Ile Leu Asn	
1445		1450	1455	
Ala Thr	Asn Thr Ser Leu Thr	Ile Arg Leu Pro Leu	Ala Lys Thr	
1460		1465	1470	
Asn Leu	Thr Trp Tyr Gly Ile	Thr Ser Pro Thr Pro	Thr Tyr Leu	
1475		1480	1485	
Val Tyr	Tyr Ala Glu Val Asn	Asp Arg Lys Asn Ser	Ser Asp Leu	
1490		1495	1500	
Lys Tyr	Arg Ile Leu Glu Phe	Gln Asp Ser Ile Ala	Leu Ile Glu	
1505		1510	1515	
Asp Leu	Gln Pro Phe Ser Thr	Tyr Met Ile Gln Ile	Ala Val Lys	
1520		1525	1530	

ES 2 539 830 T3

Asn Tyr Tyr Ser Asp Pro Leu Glu His Leu Pro Pro Gly Lys Glu  
 1535 1540 1545  
  
 Ile Trp Gly Lys Thr Lys Asn Gly Val Pro Glu Ala Val Gln Leu  
 1550 1555 1560  
  
 Ile Asn Thr Thr Val Arg Ser Asp Thr Ser Leu Ile Ile Ser Trp  
 1565 1570 1575  
  
 Arg Glu Ser His Lys Pro Asn Gly Pro Lys Glu Ser Val Arg Tyr  
 1580 1585 1590  
  
 Gln Leu Ala Ile Ser His Leu Ala Leu Ile Pro Glu Thr Pro Leu  
 1595 1600 1605  
  
 Arg Gln Ser Glu Phe Pro Asn Gly Arg Leu Thr Leu Leu Val Thr  
 1610 1615 1620  
  
 Arg Leu Ser Gly Gly Asn Ile Tyr Val Leu Lys Val Leu Ala Cys  
 1625 1630 1635  
  
 His Ser Glu Glu Met Trp Cys Thr Glu Ser His Pro Val Thr Val  
 1640 1645 1650  
  
 Glu Met Phe Asn Thr Pro Glu Lys Pro Tyr Ser Leu Val Pro Glu  
 1655 1660 1665  
  
 Asn Thr Ser Leu Gln Phe Asn Trp Lys Ala Pro Leu Asn Val Asn  
 1670 1675 1680  
  
 Leu Ile Arg Phe Trp Val Glu Leu Gln Lys Trp Lys Tyr Asn Glu  
 1685 1690 1695  
  
 Phe Tyr His Val Lys Thr Ser Cys Ser Gln Gly Pro Ala Tyr Val  
 1700 1705 1710  
  
 Cys Asn Ile Thr Asn Leu Gln Pro Tyr Thr Ser Tyr Asn Val Arg  
 1715 1720 1725  
  
 Val Val Val Val Tyr Lys Thr Gly Glu Asn Ser Thr Ser Leu Pro



ES 2 539 830 T3

1730		1735		1740
Glu Ser Phe Lys Thr Lys Ala Gly Val Pro Asn Lys Pro Gly Ile				
1745		1750		1755
Pro Lys Leu Leu Glu Gly Ser Lys Asn Ser Ile Gln Trp Glu Lys				
1760		1765		1770
Ala Glu Asp Asn Gly Cys Arg Ile Thr Tyr Tyr Ile Leu Glu Ile				
1775		1780		1785
Arg Lys Ser Thr Ser Asn Asn Leu Gln Asn Gln Asn Leu Arg Trp				
1790		1795		1800
Lys Met Thr Phe Asn Gly Ser Cys Ser Ser Val Cys Thr Trp Lys				
1805		1810		1815
Ser Lys Asn Leu Lys Gly Ile Phe Gln Phe Arg Val Val Ala Ala				
1820		1825		1830
Asn Asn Leu Gly Phe Gly Glu Tyr Ser Gly Ile Ser Glu Asn Ile				
1835		1840		1845
Ile Leu Val Gly Asp Asp Phe Trp Ile Pro Glu Thr Ser Phe Ile				
1850		1855		1860
Leu Thr Ile Ile Val Gly Ile Phe Leu Val Val Thr Ile Pro Leu				
1865		1870		1875
Thr Phe Val Trp His Arg Arg Leu Lys Asn Gln Lys Ser Ala Lys				
1880		1885		1890
Glu Gly Val Thr Val Leu Ile Asn Glu Asp Lys Glu Leu Ala Glu				
1895		1900		1905
Leu Arg Gly Leu Ala Ala Gly Val Gly Leu Ala Asn Ala Cys Tyr				
1910		1915		1920
Ala Ile His Thr Leu Pro Thr Gln Glu Glu Ile Glu Asn Leu Pro				
1925		1930		1935

Ala Phe Pro Arg Glu Lys Leu Thr Leu Arg Leu Leu Leu Gly Ser  
 1940 1945 1950

Gly Ala Phe Gly Glu Val Tyr Glu Gly Thr Ala Val Asp Ile Leu  
 1955 1960 1965

Gly Val Gly Ser Gly Glu Ile Lys Val Ala Val Lys Thr Leu Lys  
 1970 1975 1980

Lys Gly Ser Thr Asp Gln Glu Lys Ile Glu Phe Leu Lys Glu Ala  
 1985 1990 1995

His Leu Met Ser Lys Phe Asn His Pro Asn Ile Leu Lys Gln Leu  
 2000 2005 2010

Gly Val Cys Leu Leu Asn Glu Pro Gln Tyr Ile Ile Leu Glu Leu  
 2015 2020 2025

Met Glu Gly Gly Asp Leu Leu Thr Tyr Leu Arg Lys Ala Arg Met  
 2030 2035 2040

Ala Thr Phe Tyr Gly Pro Leu Leu Thr Leu Val Asp Leu Val Asp  
 2045 2050 2055

Leu Cys Val Asp Ile Ser Lys Gly Cys Val Tyr Leu Glu Arg Met  
 2060 2065 2070

His Phe Ile His Arg Asp Leu Ala Ala Arg Asn Cys Leu Val Ser  
 2075 2080 2085

Val Lys Asp Tyr Thr Ser Pro Arg Ile Val Lys Ile Gly Asp Phe  
 2090 2095 2100

Gly Leu Ala Arg Asp Ile Tyr Lys Asn Asp Tyr Tyr Arg Lys Arg  
 2105 2110 2115

Gly Glu Gly Leu Leu Pro Val Arg Trp Met Ala Pro Glu Ser Leu  
 2120 2125 2130

Met Asp Gly Ile Phe Thr Thr Gln Ser Asp Val Trp Ser Phe Gly

2135		2140		2145
Ile Leu 2150	Ile Trp Glu	Ile Leu 2155	Thr Leu Gly His	Gln Pro Tyr Pro 2160
Ala His 2165	Ser Asn Leu Asp	Val 2170	Leu Asn Tyr Val	Gln Thr Gly Gly 2175
Arg Leu 2180	Glu Pro Pro Arg	Asn 2185	Cys Pro Asp Asp	Leu Trp Asn Leu 2190
Met Thr 2195	Gln Cys Trp Ala	Gln 2200	Glu Pro Asp Gln	Arg Pro Thr Phe 2205
His Arg 2210	Ile Gln Asn Gln	Leu 2215	Gln Leu Phe Arg	Asn Phe Phe Leu 2220
Asn Ser 2225	Ile Tyr Gln Cys	Arg 2230	Asp Glu Ala Asn	Asn Ser Gly Val 2235
Ile Asn 2240	Glu Ser Phe Glu	Gly 2245	Glu Asp Gly Asp	Val Ile Cys Leu 2250
Asn Ser 2255	Asp Asp Ile Met	Pro 2260	Val Val Leu Met	Glu Thr Lys Asn 2265
Arg Glu 2270	Gly Leu Asn Tyr	Met 2275	Val Leu Ala Thr	Glu Cys Gly Gln 2280
Gly Glu 2285	Glu Lys Ser Glu	Gly 2290	Pro Leu Gly Ser	Gln Glu Ser Glu 2295
Ser Cys 2300	Gly Leu Arg Lys	Glu 2305	Glu Lys Glu Pro	His Ala Asp Lys 2310
Asp Phe 2315	Cys Gln Glu Lys	Gln 2320	Val Ala Tyr Cys	Pro Ser Gly Lys 2325
Pro Glu 2330	Gly Leu Asn Tyr	Ala 2335	Cys Leu Thr His	Ser Gly Tyr Gly 2340
Asp Gly 2345	Ser Asp			

5 <210> 8  
 <211> 7368  
 <212> DNA  
 <213> Homo sapiens

ES 2 539 830 T3

<400> 8

caagctttca agcattcaaa ggtctaaatg aaaaaggcta agtattatit caaaaggcoa	60
gtatatccta atatagcaaa acaaacaaag caaaatccat cagctactcc tccaattgaa	120
gtgatgaagc ccaataaatt catatagcaa aatggagaaa attagaccgg ccatctaaaa	180
atctgccatt ggtgaagtga tgaagaacat ttactgtcct attccgaagc ttgtcaatit	240
tgcaactcct ggctgcctat ggatttctgt ggtgcagtgt acagttitaa atagctgcct	300
aaagtcgtgt gtaactaatc tgggccagca gcttgacctt ggcacaccac ataacttgag	360
tgaaccgtgt atccaaggat gtcacttttg gaactctgta gatcagaaaa actgtgcctt	420
aaagtgtcgg gagtcgtgtg aggttgctg tagcagcgcg gaagggtgat atgaagagga	480
agtactggaa aatgcagacc taccaactgc tccctttgct tcttccattg gaagccacaa	540
tatgacatta cgatggaaat ctgcaactt ctctggagta aaatacatca ttcagtggaa	600
atatgcacaa cttctgggaa gctggactta tactaagact gtgtccagac cgtectatgt	660
ggtcaagccc ctgcaccct tcaactgagta cttttccga gtggtttgga tcttcacagc	720
gcagctgcag ctctactccc ctccaagtcc cagttacagg actcatcctc atggagtcc	780
tgaactgca ctttgatta ggaatattga gagctcaagt cccgacactg tggaaagtcag	840
ctgggatcca cctcaattcc caggtggacc tttttgggt tataacttaa ggctgatcag	900
caaaaatcaa aaattagatg cagggacaca gagaaccagt ttccagttit actccactit	960
accaaatact atctacaggt tttctattgc agcagtaaat gaagtgggt agggccaga	1020
agcagaatct agtattacca cttcatctc agcagttcaa caagaggaac agtggctcct	1080
tttatccaga aaaacttctc taagaagag atctttaaa catttagtag atgaagcaca	1140
ttgccttcgg ttgatgcta tataaccataa tattacagga atatctgttg atgtccacca	1200
gcaaattgtt ttttctctg aaggaactct catatgggcg aagaaggctg ccaacatgtc	1260
tgatgtatct gacctgagaa tttttacag aggttcagga ttaatttctt ctatctccat	1320

ES 2 539 830 T3

agattggctt tatcaagaa tgtatttcat catggatgaa ctggatgtg tctgtgattt 1380  
 agagaactgc tcaaacatcg aggaattac tccacctct attagtgcac ctcaaaaaat 1440  
 tgtggctgat tcatacaatg ggtatgtctt ttacctcctg agagatggca tttatagagc 1500  
 agaccttctt gtaccatctg gccgggtgtc agaagctgtg cgtattgtgg agagttgcac 1560  
 gttaaaggac tttgcaatca agccacaagc caagcgaatc atttacttca atgacaactgc 1620  
 ccaagtcttc atgtcaacat ttctggatgg ctctgcttcc catctcatcc tacctcgcac 1680  
 cccctttgct gatgtgaaaa gttttgcttg tgaaaacaat gactttcttg tcacagatgg 1740  
 caaggcatt ttccaacagg atgctttgtc ttttaatgaa ttcatcgtgg gatgtgacct. 1800  
 gagtcacata gaagaatttg ggtttgtaa cttggctcctc tttggctcat cctcccagct 1860  
 gcacctctg ccaggccgcc cgcaggagct ttcgggtctg tttggctctc accaggctct 1920  
 tgttcaatgg aagectcctg cccttgccat aggagccaat gtcacctga tcagtgatat 1980  
 tattgaactc tttgaattag gcccttctgc ctggcagaac tggacctatg aggtgaaagt 2040  
 atccacccaa gacctcctg aagtcactca tttttcttg aacataagtg gaaccatgct 2100  
 gaatgtacct gagctgcaga gtgctatgaa atacaagggt tctgtgagag caagttctc 2160  
 aaagaggcca ggccccgtt cagagccctc agtgggtact accctgggtc cagctagtga 2220  
 accaccattt atcatggctg tgaagaaga tgggcttgg agtaaaccat taaatagctt 2280  
 tggcccagga gagttcttat cctctgatat aggaatgtg tcagacatgg attggtataa 2340  
 caacagctc tactacagtg acacgaaagg cgacgtttt gtgtggctgc tgaatgggac 2400  
 ggatatetca gagaattatc acctaccag cattgcagga gcaggggctt tagcttttga 2460  
 gtggctgggt cactttctc actgggctgg aaagacatat gtgatacaa ggcagtctgt 2520  
 gttgacggga cacacagaca ttgttaccca cgtgaagcta ttggtgaatg acatggtggt 2580  
 ggattcagtt ggtggatate tctactggac cacactctat tcagtggaaa gcaccagact 2640  
 aatggggaa agttccctg tactacagac acagccttg ttttctggga aaaaggtaat 2700  
 tgetctaact ttagacctca gtgatgggct cctgtattgg ttggttcaag acagtcaatg 2760  
 tattcacctg tacacagctg ttcttcgggg acagagcact ggggatacca ccatcacaga 2820  
 atttgcagcc tggagtactt ctgaaattc ccagaatgca ctgatgtact atagtggctg 2880  
 gctgttctgg atcaatggct ttaggattat cacaactcaa gaaataggtc agaaaaccag 2940

ES 2 539 830 T3

tgtctctgtt ttggaaccag ccagatttaa tcagttcaca attattcaga catcccttaa 3000  
 gcccctgccg gggaactttt cctttacccc taaggttatt ccagattctg ttcaagagtc 3060  
 ttcatttagg attgaaggaa atgcttcaag ttttcaaatc ctgtggaatg gtccccctgc 3120  
 ggtagactgg ggtgtagttt tctacagtgt agaatttagt gctcattcta agttcttggc 3180  
 tagtgaacaa cactctttac ctgtatttac tgtggaagga ctggaacctt atgccttatt 3240  
 taatctttct gtcactcctt atacctactg gggaaagggc cccaaaacat ctctgtcact 3300  
 tcgagcacct gaaacagttc catcagcacc agagaacccc agaatattta tattaccaag 3360  
 tggaaaatgc tgcaacaaga atgaagttgt ggtggaatth aggtggaaca aacctaagca 3420  
 tgaaaatggg gtgtaacaa aatttgaaat tttctacaat atatccaatc aaagtattac 3480  
 aaacaaaaca tgtgaagact ggattgctgt caatgtcact ccctcagtga tgtcttttca 3540  
 actgaaggc atgagtccca gatgctttat tgccttcag gttagggcct ttacatctaa 3600  
 ggggccagga ccatatgctg acgttgtaa gtctacaaca tcagaantca acccatttcc 3660  
 tcacctata actcttcttg gtaacaagat agttttttta gatatggatc aaaatcaagt 3720  
 tgtgtggacg ttttcagcag aaagagttat cagtgcggtt tgctacacag ctgataatga 3780  
 gatgggatat tatgctgaag gggactcact ctttctctg cacttgcaaca atcgtctag 3840  
 ctctgagctt ttccaagatt cactggtttt tgatatcaca gttattaca ttgactggat 3900  
 ttcaaggcac ctctactttg cactgaaaga atcacaaaat ggaatgcaag tatttgatgt 3960  
 tgatcttgaa cacaaggatg aatatcccag agaggtgag attcacaata ggaattcaac 4020  
 aataatttct ttttctgtat atcctctttt aagtcgcttg tattggacag aagtttcca 4080  
 ttttggtac cagatgttct actacagtat tatcagtcac accttgacc gaattctgca 4140  
 acccacagct acaaccaac aaacaaaag gaatcaatgt tcttgtaatg tgactgaatt 4200  
 tgagttaagt ggagcaatgg ctattgatac ctctaacctg gagaaccat tgatatactt 4260  
 tgccaaagca caagagatct gggcaatgga tctggaaggc tgtcagtgtt ggagagttat 4320  
 cacagtacct gctatgctg caggaaaaac cettgttagc ttaactgtgg atggagatct 4380  
 tatatactgg atcatcacag caaaggacag cacacagatt tatcaggcaa agaaaggaaa 4440  
 tggggccatc gtttcccagg tgaaggccct aaggagtagg catatcttgg cttacagttc 4500  
 agttatgcag cttttccag ataaagcgtt tctgtctcta gcttcagaca ctgtggaacc 4560

aactatactt aatgccacta acactagect cacaatcaga ttacctctgg ccaagacaaa 4620  
cctcacatgg tatggcatca ccagccctac tccaacatac ctggtttatt atgcagaagt 4680  
taatgacagg aaaaacagct ctgacttga aatatagaatt ctggaatttc aggacagtat 4740  
agctcttatt gaagatttac aaccatttcc aacatacatg atacagatag ctgtaaaaaa 4800  
ttattattca gatcctttgg aacatttacc accaggaaaa gagatttggg gaaaaactaa 4860  
aatggagta ccagaggcag tgcagctcat taatacaact gtgcggtcag acaccagcct 4920  
cattatatct tggagagaat ctcaacagcc aatggacct aaagaatcag tccgttatca 4980  
gttggaatc tcacacctgg ccctaattcc tgaactcct ctaagacaaa gtgaatttcc 5040  
aatggaagg ctcaactctc ttgttactag actgtctggg ggaatatatt atgtgttaa 5100  
ggttcttggc tgccactctg aggaatgtg gtgtacagag agtcactctg tcaactgtga 5160  
aatgtttaac acaccagaga aaccttattc cttggttcca gagaacacta gtttgcaatt 5220  
taattggaag gctccattga atgttaacct catcagattt tgggttgagc tacagaagtg 5280  
gaaatacaat gagttttacc atgttaaac ttcattgcagc caaggctctg ctatgtctg 5340  
taatatacaca aatctacaac cttatacttc atataatgtc agagtagtgg tggttataa 5400  
gacgggagaa aatagcacct cacttcaga aagctttaag acaaaagctg gagtcccaa 5460  
taaaccaggc attcccaaat tactagaagg gagtaaaaat tcaatacagt gggagaaagc 5520  
tgaagataat ggatgtagaa ttacatacta tacccttgag ataagaaaga gcacttcaa 5580  
taatttacag aaccagaatt taaggtgaa gatgacattt aatggatcct gcagtagtgt 5640  
ttgcacatgg aagtccaaa acctgaaagg aatatttcag tcagagtag tagctgcaa 5700  
taatctaggg tttggtgaat atagtggaat cagtgagaat attatattag ttggagatga 5760  
tttttgata ccagaaaca gtttcatact tactattata gttggaatat ttctggtgt 5820  
tacaatccca ctgaccttg tctggcatag aagattaaag aatcaaaaaa gtgccaagga 5880  
aggggtgaca gtgcttataa acgaagaca agagttggct gagctgcgag gtctggcagc 5940  
cggagtaggc ctggctaatt cctgctatgc aatacatact cttccaacc aagaggagat 6000  
tgaaaatctt cctgccttcc ctgggaaaa actgactctg cgtctcttgc tgggaagtgg 6060  
agcctttgga gaagtgtatg aaggaacagc agtggacatc ttaggagttg gaagtggaga 6120  
aatcaagta gcagtgaaga ctttgaagaa gggttcaca gaccaggaga agattgaatt 6180

cctgaaggag gcacatctga tgagcaaatt taatcatccc aacattctga agcagcttgg 6240  
 agtttgtctg ctgaatgaac cccaatacat tatectggaa ctgatggagg gaggagacct 6300  
 tcttacttat ttgcgtaaag cccggatggc aacgttttat ggtcctttac tcaccttgg 6360  
 tgaccttcta gacctgtgtg tagatatttc aaaaggctgt gtctacttgg aacggatgca 6420  
 tttcattcac agggatctgg cagctagaaa ttgccttgtt tccgtgaaag actataccag 6480  
 tccacggata gtgaagattg gagactttgg actcgccaga gacatctata aaaatgatta 6540  
 ctatagaag agaggggaag gcctgtctcc agttcggtgg atggctccag aaagttagat 6600  
 ggatggaatc ttcactactc aatctgatgt atggcttttt ggaattctga tttgggagat 6660  
 tttactctt ggtcatcagc cttatccagc tcattccaac ctgatgtgt taaactatgt 6720  
 gcaaacagga gggagactgg agccaccaag aaattgtcct gatgatctgt ggaatttaat 6780  
 gaccctgtgc tgggctcaag aaccgacca aagacctact tttcatagaa ttcaggacca 6840  
 acttcagtta ttcagaaatt ttttcttaa tagcatttat aagtcagag atgaagcaaa 6900  
 caacagtga gtcataaatg aaagctttga aggtgaagat ggcgatgtga tttgtttgaa 6960  
 ttcagatgac attatgccag ttgctttaat ggaacgaag aaccgagaag ggttaacta 7020  
 tatggtactt gctacagaat gtggccaagg tgaagaaaag tctgagggtc ctctaggctc 7080  
 ccaggaatct gaatcttgtg gtctgaggaa agaagagaag gaaccacatg cagacaaaga 7140  
 tttctgcaa gaaaaacaag tggcttactg ccctctggc aagcctgaag gcctgaacta 7200  
 tgctgtctc actcacagtg gatatggaga tgggtctgat taatagcgtt gtttgggaaa 7260  
 tagagagttg agataaacac tctcattcag tagttactga aagaaaactc tgetagaatg 7320  
 ataaatgtca tgggtgtcta taactccaaa taacaatgc aacgttcc 7368

<210> 9  
 <211> 20  
 <212> PRT  
 <213> Homo sapiens

5

<400> 9  
 Ile Leu Ser Ser Ala Phe Gln Leu Val Gly Ala Gly Val Pro Asn Lys  
 1 5 10 15

Pro Gly Ile Pro

10

20

<210> 10  
 <211> 60  
 <212> DNA  
 <213> Homo sapiens

15

<400> 10  
 attcttagta ggcctcca gctggttga gctggagtcc caaataaacc aggcattccc 60

20



ES 2 539 830 T3

<210> 11  
 <211> 20  
 <212> PRT  
 <213> Homo sapiens  
 5  
 <400> 11  
 Ile Leu Ser Ser Ala Phe Gln Leu Val Gly Asp Asp Phe Trp Ile Pro  
 1 5 10 15  
  
 Glu Thr Ser Phe  
 20  
  
 <210> 12  
 10 <211> 60  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 12  
 15 attcttagta ggcctcca gctggtgga gatgatttt ggataccaga aacaagttc 60  
  
 <210> 13  
 <211> 24  
 <212> DNA  
 20 <213> Artificial  
  
 <220>  
 <223> Secuencia del cebador  
  
 <400> 13  
 25 acccttctcg gttctcgtt tcca 24  
  
 <210> 14  
 <211> 24  
 30 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Secuencia del cebador  
 35  
 <400> 14  
 gcagctcagc caactcttg tctt 24  
  
 <210> 15  
 <211> 24  
 40 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Secuencia del cebador  
 45  
 <400> 15  
 tgccagacaa aggtcagtg gatt 24  
  
 <210> 16  
 <211> 21  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Secuencia del cebador  
 55  
 <400> 16  
 60 tccatcccag cacctgcgga g 21  
 <210> 17

<211> 27  
 <212> DNA  
 <213> Artificial

5 <220>  
 <223> Secuencia del cebador

<400> 17  
 ctcaactctc tatttcccaa acaacgc 27

10 <210> 18  
 <211> 22  
 <212> DNA  
 <213> Artificial

15 <220>  
 <223> Secuencia del cebador

<400> 18  
 20 catggctccc tggcctgaat tg 22

<210> 19  
 <211> 26  
 <212> DNA  
 25 <213> Artificial

<220>  
 <223> Secuencia del cebador

30 <400> 19  
 caacgctatt aatcagaccc atctcc 26

<210> 20  
 <211> 32  
 35 <212> DNA  
 <213> Artificial

<220>  
 <223> Secuencia del cebador

40 <400> 20  
 gaagatctct gaccatggct cctggcctg aa 32

<210> 21  
 <211> 32  
 <212> PRT  
 <213> Artificial

<220>  
 50 <223> Secuencia del cebador

<400> 21  
 Gly Ala Ala Gly Ala Thr Cys Thr Ala Cys Gly Cys Thr Ala Thr Thr  
 1 5 10 15

Ala Ala Thr Cys Ala Gly Ala Cys Cys Cys Ala Thr Cys Thr Cys Cys  
 20 25 30

55

**REIVINDICACIONES**

- 1.- Un polinucleótido aislado que comprende una secuencia de nucleótidos al menos 95% idéntica a una secuencia seleccionada del grupo que consiste en:
- 5 (a) una secuencia de nucleótidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos de SEQ ID NO:1 o SEQ ID NO:3;
- (b) una secuencia de nucleótidos que codifica un polipéptido de fusión de SLC34A2-ROS, comprendiendo dicha secuencia de nucleótidos la secuencia de nucleótidos de SEQ ID NO:2 o SEQ ID NO:4;
- 10 (c) una secuencia de nucleótidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos N-terminal de SLC34A2 que consiste en los restos 1-126 de SEQ ID NO:5, y el dominio quinasa de ROS que consiste en los restos 1945-2222 de SEQ ID NO:7;
- (d) una secuencia de nucleótidos que comprende la secuencia de nucleótidos N-terminal de SLC34A2 que consiste en los nucleótidos 1-378 de SEQ ID NO:6, y la secuencia de nucleótidos del dominio quinasa de ROS que consiste en los nucleótidos 6032-6865 de SEQ ID NO:8; y
- 15 (e) una secuencia de nucleótidos complementaria con cualquiera de las secuencias de nucleótidos de (a)-(d).
- 2.- El polinucleótido aislado de la reivindicación 1, en el que dicha secuencia de nucleótidos de (b) comprende la secuencia de nucleótidos codificadora del clon de ADNc contenido en ATCC n.º de depósito PTA-7877.
- 3.- Un método para producir un vector recombinante que comprende insertar una molécula de ácido nucleico aislada de la reivindicación 1 en un vector.
- 20 4.- Un vector recombinante que comprende una molécula de ácido nucleico aislada según la reivindicación 1.
- 5.- Un método para fabricar una célula hospedante recombinante que comprende introducir el vector recombinante de la reivindicación 4 en una célula hospedante.
- 6.- Una célula hospedante recombinante transformada con el vector recombinante de la reivindicación 4.
- 7.- Un polipéptido aislado que comprende una secuencia de aminoácidos al menos 95% idéntica a una secuencia seleccionada del grupo que consiste en:
- 25 (a) una secuencia de aminoácidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos de SEQ ID NO:1 o SEQ ID NO:3; y
- (b) una secuencia de aminoácidos que codifica un polipéptido de fusión de SLC34A2-ROS que comprende la secuencia de aminoácidos N-terminal de SLC34A2 que consiste en los restos 1-126 de SEQ ID NO:5, y el dominio quinasa de ROS que consiste en los restos 1945-2222 de SEQ ID NO:7.
- 30 8.- Un método para producir el polipéptido de fusión de SLC34A2-ROS de la reivindicación 7, comprendiendo dicho método cultivar una célula hospedante recombinante que comprende un vector recombinante que codifica dicho polipéptido de fusión bajo condiciones adecuadas para la expresión de dicho polipéptido de fusión, y recuperar dicho polipéptido.
- 35 9.- Un reactivo aislado que se une específicamente o detecta un polipéptido de fusión de SLC34A2-ROS de la reivindicación 7, pero que no se une ni detecta la SLC34A2 de tipo salvaje ni la ROS de tipo salvaje, en el que dicho reactivo es un anticuerpo.
- 10.- Un reactivo que detecta un polinucleótido de la reivindicación 1 en un ensayo de PCR, que comprende un primer oligonucleótido en la orientación sentido y un segundo oligonucleótido en la orientación antisentido.
- 40 11.- Un reactivo que detecta un polinucleótido de la reivindicación 1, que comprende una molécula de ácido nucleico aislada que comprende al menos 15 nucleótidos contiguos, o uno de sus complementos, que incluye los restos de la zona de unión de la fusión 376-381 de SEQ ID NO:2, o los restos 376-381 de SEQ ID NO:4 de un polinucleótido de fusión de SLC34A2-ROS.
- 45 12.- Un método para determinar la presencia de un polinucleótido de ROS mutante en una muestra biológica procedente de un cáncer de mamífero, comprendiendo dicho método detectar un polinucleótido según la reivindicación 1 en dicha muestra para determinar si un polinucleótido de ROS mutante está presente en dicha muestra.
- 13.- Un método para determinar la presencia de un polipéptido de ROS mutante en una muestra biológica

procedente de un cáncer de mamífero, comprendiendo dicho método detectar un polipéptido según la reivindicación 7 en dicha muestra para determinar si un polinucleótido de ROS mutante está presente en dicha muestra.

14.- El método de la reivindicación 12 o 13, en el que dicho cáncer es un cáncer de pulmón.

5 15.- El método de la reivindicación 14, en el que dicho cáncer de pulmón es un carcinoma de pulmón no microcítico (NSCLC).

16.- El método de la reivindicación 14, en el que la presencia de un polinucleótido o un polipéptido de ROS mutante identifica a dicho cáncer como un cáncer que es probable que responda a una composición que comprende al menos un producto terapéutico inhibidor de la ROS quinasa.

10 17.- El método de la reivindicación 14, puesto en práctica en un formato de ensayo de citometría de flujo (FC), de inmunohistoquímica (IHC), de inmunofluorescencia (IF), un formato de hibridación in situ de fluorescencia (FISH), o un formato de ensayo de reacción en cadena de polimerasa (PCR).

18.- El método de la reivindicación, 13, en el que se detecta la actividad ROS quinasa de dicho polipéptido en el cáncer.

15 19.- Un método para determinar si un compuesto inhibe el avance de un cáncer, que comprende la etapa de determinar si dicho compuesto inhiben la expresión y/o la actividad ROS quinasa de un polipéptido según la reivindicación 7 en dicho cáncer.

20.- El método de la reivindicación 19, en el que se determina la inhibición de la expresión y/o la actividad ROS quinasa de dicho polipéptido empleando al menos un reactivo que detecta un polinucleótido de la reivindicación 1 y/o al menos un reactivo según una cualquiera de las reivindicaciones 9-11.

20 21.- Un anticuerpo dirigido que se une a ROS e inhibe la actividad ROS quinasa, para su uso en un método para inhibir el avance de un cáncer que expresa un polipéptido de fusión de SLC34A2-ROS según la reivindicación 7, comprendiendo dicho método la etapa de inhibir la expresión y/o la actividad ROS quinasa de dicho polipéptido de fusión de SLC34A2-ROS en dicho cáncer.

25 22.- El anticuerpo de la reivindicación 21 para su uso en un método según la reivindicación 21, en el que dicho cáncer es cáncer de pulmón.

23.- El anticuerpo de la reivindicación 22 para su uso en un método según la reivindicación 21, en el que dicho cáncer de pulmón es un carcinoma pulmonar no microcítico (NSCLC).

30 24.- Un kit para la detección de un polinucleótido y/o un polipéptido de fusión de SLC34A2-ROS en una muestra biológica, comprendiendo dicho kit al menos un polinucleótido de la reivindicación 1 y/o al menos un reactivo según una cualquiera de las reivindicaciones 9-11, y uno o más reactivos secundarios.

Fig. 1

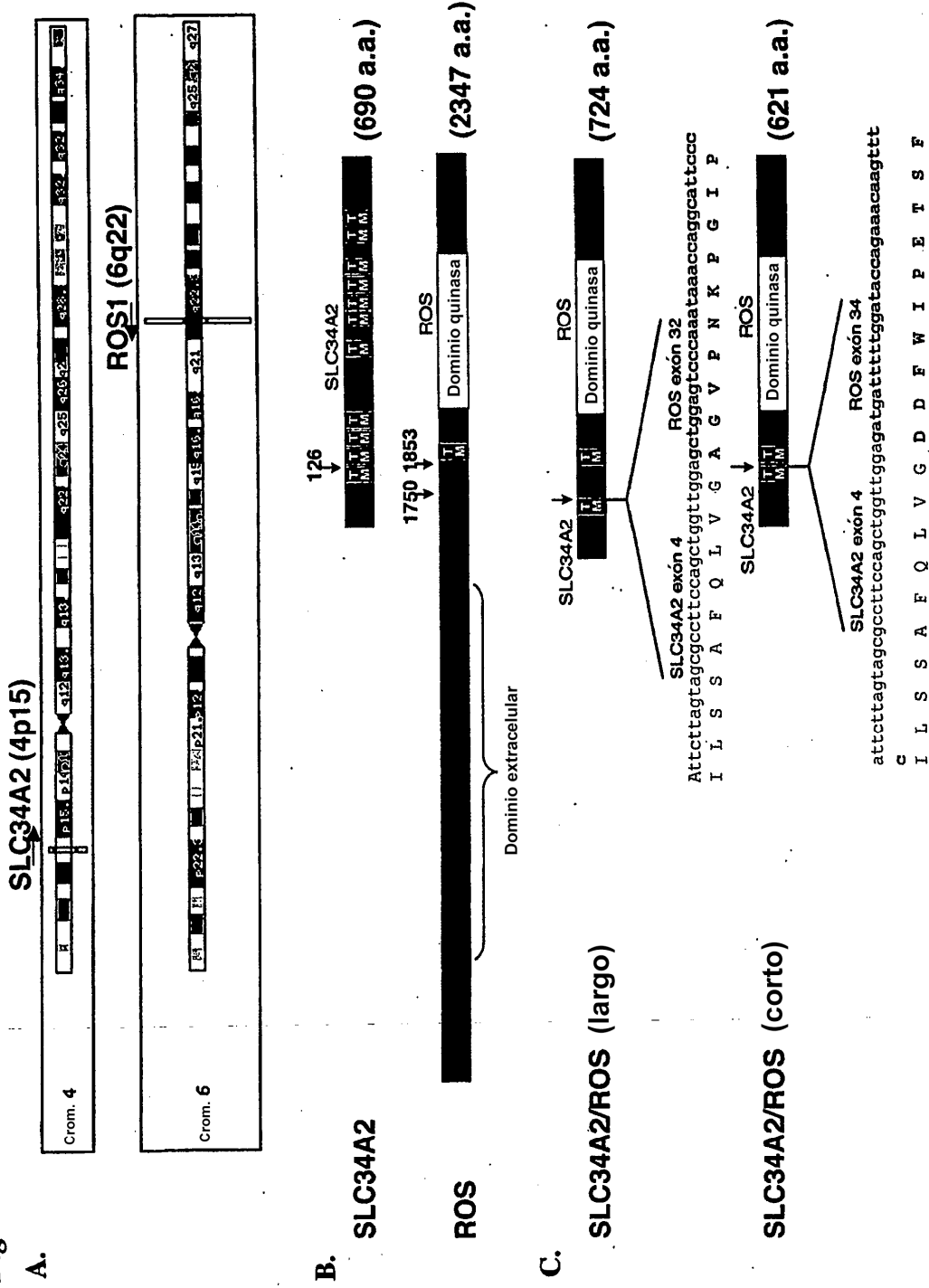


Fig. 2A

MAPWPELGDAQPNPDKYLEGAAGQQPTAPDKSKETNKTDNTEAPVTKIELLPSYSTATLIDEPTEVDDPWNLPQLQDSGI  
 KWSEKRTKGLKLCFFQIGRLILLGFLYFFVCSLDILSSAFQLVAGVVPNKPGIPKLEGGSKNSIQWEKAEDNGCRITY  
 YILEIRKSTSNLQNLNRWKMTFNGSCSSVCTWKSKNLKGIFQFRVVAANNLGFGEYSGISSENIILVGDDFWIPETSFI  
 LTIIVGIFLVVTIPLTFVWHRRLNQKSAKEGVTVLINEDKELAE LRGLAAGVGLANACYAIHTLPTQEEIENLPAFPRE  
 KLTIRLLLLSGAFGEVYEGTAVDILGVSSEIKVAVKTLKKGSTQEKIEFLKEAHLMSKFNHPNILKQLGVCLLNEPQY  
 IILELMEGGDLLTYLRKARMAFYGPLLTLVDLVDLCVDISKGCVYLERMHFIHRDLAARNCLVSVKDYTSPRIVKIGDF  
 GLARDIYKNDYRKRGEGLLPVRWMAPESLMDGIFTTQSDVWSFGILWEILTLGHQPYPAHSNLDVLYVQTGGRELEPP  
 RNCPPDDLWNLMTCWAQEPDQRPTEFHRIQDQLQFRNFFLNSIYKSRDEANNSGVINESFEGEDGDVICLNSDDIMPVAL  
 METKNREGLNYMVLATECGQGEKSEGPLGSQESSESCGLRKEEKEPHADKDFCQEKQVAYCPSGKPEGLNYACLTHSGYG  
 DGSD

a tggctccctggcctgaa tgggagatgccagcccaaccccgataagta cctcgaagggccgcaggtcagcagccca c  
 tggccctgataaaaagcaagagagaccaa caaaa cagataa cactgaggcacctgtaaccaagat tgaactctgccgtcct  
 actccacggctacactgata tagatgagccactgaggtgga tga cccctggaacctaccactcttcaggactcgggga t c  
 aagtggctcagagagagacaccaaagggaagattctctgttctctccaagggattgggagattgattttactctcggatt  
 tctctacttttctgtgtctccctgga tattcttagtagcgccttccagctgggtggagctggagtc ccaataaaccag  
 gcatcccaaatctagaaggagtaaaaaattcaatacagtgaggagaaagctgaagataatggatgtagaattacatac  
 tataccttgagataaagaaagagcacttcaataat tta cagaaccagaat ttaagggtggaagatgacattta atggatc  
 ctgcagtagtgtttgcacatggaagtc caaaaacctgaaaggaatatttcagttcagagtagtagctgcaataatctag  
 ggtttggtaataatagtggaatcagtgagaatattatattagttggagatgattttggataccagaaacaagtttcata  
 ctactattatagttggaatattctgggtgttacaatcccactgacctttgtctggcatagaagattaagaatcaaaa  
 aagtgccaaaggaaggggtgacagtgcttataaa cgaagacaaagagttggctgagctgaggtctggcagccggagtag  
 gcctggcta atgcctgctatgcaatacactcttccaaccaagaggagattgaaaatctcctgccttccctcgggaa  
 aaactgactctgcgtctcttctgctggaagtggaagcctttggagaagtgtatgaaggaaacagcagtggaactcttaggag  
 tggagtggaagaaatcaaaagtagcagtgaaagccttgaagaagggttccacagaccaggagaagattgaaatcctgaagg  
 aggcacatctgatgagcaaa ttaatac tccaacattctgaagcagcttgagtttgtctgctgaatgaaccccaatac  
 attacctggaaactgatggaggaggagaccttctacttatttgctaaagccggatggcaacgttttaggtccttt  
 actcaacttggttgacctgtgtagacctgtgtgtagata tttcaaaaggctgtgtctacttggaaaggatgcaatttcattc  
 acagggatctggcagctagaaattgccttgtttccgtgaaagactataccagtcacaggatagtgaaagattggagacttt  
 ggactcggcagagacatac taaaaatgattactatagaaagagagggaaggcctgctcccaggtcgggtgga tggctcc  
 agaaagtttgatggatggaatcttcaactcaatctgatgtatggctctttggaaatctgat tttgggagattttaactc  
 ttggctcatcagccttaccagctcattccaaccttgatgtgttaactatgtgcaaacaggaggagactgggagccacca  
 agaaattgtcctgatgatctgtggaatttaatagaccagtgctgggctcaagaacccgaccaaagacctacttttcatag  
 aattcaggaccacacttcagttattcagaaat ttttcttaaatagcatttataagtcagagatgaagcaaacacagtg  
 gactcataaatgaaagctttgaaggtgaagatggcgatgtgatttgtttgaattcagatgacattatgccagttgcttta  
 atgaaacgaagaaccgagaagggttaaactataggtacttgctacagatgtggccaaggtgaagaaaagctcagggg  
 tcctctaggtcccaggaatctgaatcttgggtctgaggaaagagaaggaaccaatgcagacaaaagatttctgcc  
 aagaaaaacaagtgcttactgcccttctggcaagcctgaaggcctgaaactatgcctgtctcactcacagtggaatagga  
 gatgggtctgattaa

Fig. 2B

MAPWPELGDAQPNPKYLEGAAGQQPTAPDKSKETNKTDNTEAPVTKIELLPSYSTATLIDEPTVDDPWNLPQLQDSGI  
 KWSERDTKGKILCFQIGIRLILLGFLYFFVCSLDLSSAFQLVGDDEFWI PETSFILTIIVGIFLVVTIPLTFVWHRRL  
 KNQSAKEGVTVLINEDKELAE LRGLAAGVGLANACYAIHTLPTQEEIENLPAPPREKLTLRLLLGSGAFGEVYEGTAVD  
 ILGVSSEIKVAVKTLKKGSTQEKIEFLKEAHLMSKFNHPNLIKQLGVCLLNEPQYIILELMEGGDLLTYLRKARMAF  
 YGPLLTLVDLVDLCVDISKGCVYLERMHFIHRDLAARNCLVSVKDYTSPRIVKIGDFGLARDIYKNDYRKRGEGLLPVR  
 WMAPESLMDGIFTTQSDVWSFGILWEILTLGHQYPYPAHSNLDVLYVQ TGGRLPFRNC PDDLWNLMTCWAQEPDQRP  
 TFHRIQDQLQFRNFFLNSIYKSRDEANNSGVINESFEGEDGDVICLNSDDIMPVALMETKNREGLNYMVLATECGQGE  
 KSEGPLGSQESSEGLRKEEKEPHADKDFCQEKQVAYCPSGKPEGLNYACLTHSGYGDGS D

a tggctccctggcctgaa ttgggaga tgcccagccca accccgata agta cctcga aggggcccaggtcagcagccac  
 tgcccctga taaaagcaa agagaccaa caaaacagata acaactgaggcacctgta ccaaga ttgaa cctctcgccgtcct  
 actccacggctacactga taga tgagccca ctgagggtga tgacccctggaa cctaccca cctctcaggactcggggatc  
 aagtggtcagagagagaccca aagggaagattctctgtttcttcca aggga ttgggaga ttgatttacttctcggatt  
 tctctacttttctgtgtgctccctgga tattcttagtagcgccttccagctggttgagatgatttttgataccagaaa  
 caagtttcatacttactattatagttggaatatttctggttgttacaatcccactgacctttgtctggcatagaagatta  
 aagaatcaaaaaagtgccaaggaaggggtgacagtgtttataaacgaagacaaagagttggctgagctgaggtctggc  
 agccggagtaggcctggcta atgcctgctatgcaatacactcttccaaccaagaggagattgaaaatcttctgcct  
 tccctcgggaaaaactgactctgcgtctcttgctgggaagtgagcctttggagaagtgta tgaaggaa cagcagtgga  
 atcttaggagttggaagtgga gaaatcaa agtagcagtgaa gactttgaagaagggttccacagaccaggagaagattga  
 atcctgaaggaggca catctgatgagcaaat ttaatoatcccaacattctgaagcagcttgagtttgtctgctgaatg  
 aaocccaa tacattatcctggaactga tggagggaggagaccttctacttatttgcgta aagccggatggcaacgttt  
 tatggtcctttactcacttgggtgacctgttagacctgtgtgtaga tatttcaaaaggctgtgtctacttggaa cggat  
 gcaatttcattcacagggatctggcagctagaaattgccttggttccgtgaaagacta taccagctccacggatagtgaga  
 ttggagactttggactcgccagagacatctataaaaaatgattactatagaaagagaggggaaggcctgctccagttcgg  
 tggatggctccagaaagtttgatgga tggaa tcttcaactactcaatctgatgta tggcttttggaa tctctgatttggga  
 gat ttaactcttggtcatcagcctta tccagctcatccaaccttga tgtgttaaactatgtgcaaacaggagggagac  
 tggagccaccaagaaaattgtctgatgatctgtggaattta atgaccagtgctgggctcaagaaccggaccaaaagacct  
 acttttcatagaattcaggaccaacttcagttattcagaaa tttttcttaaatagcatttataagtccagagatgaagc  
 aaacaacagtgagtcataaatgaaagctttgaaagtgaaagatggcgatgtgatttgtttgaa ttcagatgacattatgc  
 cagttgctttaa tggaaacgaagaaccgagaagggtaaaactataatggacttgctacagaatgtggccaaggtgaagaa  
 aagctcagggctcctctaggctcccaggaatctgaa tctgtgtgctgaggaagaaagagaaggaaccaatgcagacaa  
 agatttctgccaagaaaa caagtggttactgccttctggcaagcctgaaggcctgaa cctgacatgcctgtctcactcaca  
 gtggatatggagatgggtctgattaa

Fig. 3

MAPWPELGDAQPNPDKYLEGAAGQOPTAPDKSKETNKTDNTEAPVTKIELLPSYSTATLI  
 DEPTTEVDDPNWNLPTLQDSGIKWSERDTKGKILCFQOGIGRLILLGLFLYFFVCSLDILSS  
 AFQLVGGKMAQOFFSNSSIMSNPLLGLVIGVLVTVLVQSSSTSTSIVVSMVSSLLTVRA  
 APIIIMGANIGTSITNTIVALMQVGRSEFRRAFAGATVHDFNWL SVLVLLPVEVATHY  
 LEIITQLIVESFHFKNGEDAPDLLKVIKPFKTLIVQLDKKVISQIAMNDEKAKNKSLVK  
 IWCKTFTNKTQINVTVPSTANCTSPSLCWT DGIQNWTKMKNV TYKENIAKQHI FVNFHLP  
 DLA VGTILLILSLVLCGCLIMIVKILGSVLKGQVATVIKKTINTDFPFPAWLTGYLAI  
 LVGAGMTFIVQSSSVFTSALTPLIGIGVITIERAYPLTLGNSNIGTTTTAILAALASPGNA  
 LRSSLQIALCHFFFNISGILLWYPIPFTRLP IIRMAKGLGNISAKYRWFVAVFYLIIFFFLI  
 PLTVFGLSLAGWRVLVGVGVVVFII IILVLCRLLLQSRCPRVLPKKLQNNFLPLWMRSL  
 KPWDAVSVSKFTGCFQMRCCYCCRVCCRACCLLCGCPKCCRC SKCCEDLEEAQEGQDVPVK  
 APETF DNITISREAQGEVPASDSKTECTAL

1 cgggccagggt tccaggctc ggccgcccgc tccatcccag cacctgcgga gggagcgctg  
 61 accatggctc cctggcctga attgggagat gccagccca accccgataa gtacctcga  
 121 ggggccgag gtcagcagcc cactgcccct gataaaagca aagagaccaa caaacagat  
 181 aacactgagg cacctgtaac caagattgaa cttctgccgt cctactccac ggctacactg  
 241 atagatgagc ccactgaggt ggatgacccc tggaacctac ccactcttca ggactcgggg  
 301 atcaagtggg cagagagaga caccaaaggg aagattctct gtttcttcca agggattggg  
 361 agattgattt tacttctcgg atttctctac ttttctgtgt gctccctgga tattcttagt  
 421 agcgccttcc agctgggttg aggaaaaatg gcaggacagt tcttcagcaa cagctctatt  
 481 atgtccaacc ctttggtggg gctgggtgac ggggtgctgg tgaccgtctt ggtgcagagc  
 541 tccagcacct caacgtccat cgttgtcagc atgggtgctt cttcattgct cactgttcgg  
 601 gctgccatcc ccattatcat gggggccaac attggaacgt caatcaccaa cactattggt  
 661 cgctcatgc aggtgggaga tcggagtgag ttcagaagag cttttgcagg agccactgtc  
 721 catgacttct tcaactggct gtccgtgttg gtgctcttgc ccgtggaggt ggcacccat  
 781 tacctcgaga tcataacca gcttatagtg gagagcttcc acttcaagaa tggagaagat  
 841 gccccagatc ttctgaaagt catcactaag cccttcacaa agctcattgt ccagctggat  
 901 aaaaaagtta tcagccaaat tgcaatgaac gatgaaaaag cgaaaaaca gagtcttgtc  
 961 aagatttggg gcaaaaacttt taccaacaag acccagatta acgtcactgt tccctcgact  
 1021 gctaactgca cctcccctc cctctgttg acggatggca tccaaaactg gaccatgaag  
 1081 aatgtgacct acaaggagaa catcgccaaa tgccagcata tctttgtgaa tttccacctc  
 1141 ccggatcttg ctgtgggcac catcttgctc atactctccc tgctggctct ctgtgggtgc  
 1201 ctgatcatga ttgtcaagat cctgggctct gtgctcaagg ggcagctgc cactgtcatc  
 1261 aagaagacca tcaacactga tttcccctt cccttgcat ggttgactgg ctacctggcc  
 1321 atcctcgtcg gggcaggcat gaccttcatc gtacagagca gctctgtgtt cagctcggcc  
 1381 ttgaccccc tgattggaat cggcgtgata accattgaga gggcttacc actcacgctg  
 1441 ggctccaaca tcggcaccac caccaccgc atcctggccg ccttagccag ccctggcaat  
 1501 gcattgagga gttcactcca gatcgccctg tgccactttt tcttcaacat ctccggcatc  
 1561 ttgctgtggt acccgatccc gttcactcgc ctgcccacc gcatggcaa ggggctgggc  
 1621 aacatctctg ccaagtatcg ctgggtogcc gtctctacc tgatcatctt cttcttctg  
 1681 atcccgtgta cgggtgtttg cctctcgtg gccggctggc ggggtgctgt tgggtgcggg  
 1741 gttcccgtcg tcttcatcat catcctggta ctgtgcctcc gactcctgca gtctcgtgc  
 1801 ccaagcgtcc tgccgaagaa actccagaac tggaaactcc tgccgctgtg gatgcgctcg  
 1861 ctgaagccct gggatgccgt cgtctccaag ttcaccggt gcttccagat gcgctgctgc  
 1921 tactgctgcc gcgtgtgctg ccgcgcgtgc tgcttgctgt gtggctgcc caagtgtgc  
 1981 cgctgcagca agtgctgca ggacttgag gaggcagag agggcagga tgtccctgtc  
 2041 aaggctcctg agaccttga taacataacc attagcagag aggctcaggg tgaggctcct  
 2101 gcctcggact caaagaccga atgcacggcc ttgtagggga cgcctcagat tgtcagggat  
 2161 ggggggatgg tccttgagtt ttgcatgctc tcctccctcc cacttctgca cccttccacc  
 2221 acctcgagga gatttgctcc ccattagcga atgaaattga tgcagctccta aaaaaaaaa



**Fig. 4A**

MKNIYCLIPKLVNFATLGCLWISVVQCTVLNSCLKSCVTNLGQQDLGTPHNLSEPCIQG  
 CFWNSVDQKNCALKCRESCVEVCSSAEGAYEEVLENADLPTAPFASSIGSHNMTLRWK  
 SANFSGVKYIIQWKYAQLLGSWTYTKTVSRPSYVVKPLHPFTEYIFRVVWIFTAQLQLYS  
 PPSPSYRTHPHGVPETAPLIRNIESSSPDTVEVSWDPPQFPGGPILGYNLRLISKNOKLD  
 AGTQRTSFQFYSTLPNTIYRFSIAAVNEVGEGPEAESSITSSSAVQOEEQWLFLSRKTS  
 LRKRSLKHLVDEAHCLRLDAIYHNITGISVDVHQQIVYFSEGTLIWAKKAANMSDVSDLR  
 IFYRGSGLISSISIDWLYQRMFYIMDELVCVCDLENCNIEEITPPSISAPQKIVADSYN  
 GYVFYLLRDGIYRADLPVPSGRCAEAVRIVESCTLKDFAIKPOAKRIIYFNDAQVFMST  
 FLDGSASHLILPRIPFADVKS FACENNDFLVTDGKVI FQQDALS ENEFIVGCDLSHIEEF  
 GFGNLVIFGSSSQLHPLPGRPQELS VLF GSHQALVQWKPPALAI GANVILISDIIELFEL  
 GPSAWQNWTYEVKVSTQDPPEVTHIFLNISGTMLNVP ELQSAMKYKVS VRASSPKRPGPW  
 SEPSVGTTLVPASEPPFIMAVKEDGLWSKPLNSFGPGEFLSSDIGNVSDMDWYNNSLYS  
 DTKGDV FVWLLNGTDISENYHLPSIAGAGALAFEWLGHFLYWAGKTYVIQRQSVLTGHTD  
 IVTHVKLLVNDMVVDSVGGYLYWTTLYSVESTRLNGESSLVLTQTPWFSGKKVIALTLDL  
 SDGLLYWLVDQSQC IHLTYAVLRGQSTGDTTITEFAAWSTSEISQNALMYYSGRLEFWING  
 FRIITTQEIGQKTSVSVLEPARFNQFTIIQTS LKPLPGNFSFTPKVIPDSVQESSFRIEG  
 NASSFQILWNGPPAVDWGVV FYSVEFSAH SKFLASEQHSLPVFTVEGLEPYALFNLSVTP  
 YTYWGKGPKTSLSLRAPETVPSAPENPRIFILPSGKCCNKNEVVVEFRWNKPKHENGVL  
 KFEIFYNISNQSITNKT CEDWIAVNVTPSVMSFQLEGMSPRCFIAFQVRAFTSKGPGPYA  
 DVVKSTTSEINPFPHLITLLGNKIVFLDMDQNVVWTFSAERVISAVCYTADNEMGYAE  
 GDSLFLHLHNRSSSELFQDSL VFDITVITIDWISRHL YFALKESQNGMQVFDVDLEHKV  
 KYPREVKIHNHNSTIISFSVYPLLSRLYWTEVSNFGYQMFYYSIIISHTLHRILQPTATNQ  
 QNKRNQCS CNVTEFELSGAMAIDT SNLEKPLIYFAKAQEIWAMDLEGCQCWRVITVPAML  
 AGKTLVSLTVDGD LIYWIITAKDSTQIYQAKKNGAIVS QVKALRSRHLAYSSVMQFPF  
 DKAFLSLASDTVEPTILNATNTSLTIRLPLAKTNLTWYGITSPTPTYL VYYAEVNDRKNS  
 SDLKYRILEFQDSIALIEDLQPFSTYMIQIAVKNYSDPLEHLPPGKEIWGKT KNGVPEA  
 VQLINTTVRS DTS LIISWRESHKPNPKESVRYQLAISHLALI PETPLRQSEFPNGRLTL  
 LVTRLSGGNIYVLKVLACHSEEMWCTESHVPTVEMFNTPEKPYSLVPENTSLQFNWKAPL  
 NVNLIRFWVELQWKYNEFYHVKTSCS QGPAYVCNITNLQPYTSYNVRVVVVYKTGENST  
 SLPESFKTKAGVPNKPGI PKLLEGSKNSIQWEKAEDNGCRITYYILEIRKSTSNNLQON  
LRWKMTFNGSCSSVCTWKS KNLKGI FQFRVVAANNLGFGEYSGISENIILVGDDEFWIPET  
SFILTIIVGIFLVVTIPLTFVWHRLKNQKSAKEGVTVLIN EDKELAE LRGLAAGVGLAN  
ACYAIHTLPTQEEIENLPAPPREKLT LRLLLGSGAFGEVYEGTAVDILGVGSGEIKVAVK  
TLKKGSTDOEKIEFLKEAHLMSKFNHPN I LKQLGVCLLNEPQYIILELMEGGDLLTYLRK  
ARMATFYGP LLTLVDLVDLCVDISKGCVYLERMHFIHRDLAARNCLVSVKDYTS PRIVKI  
GDFGLARDIYKNDYYRKRGEGLLPVRWMAPE SLM DGI FT TQSDVWSEGILIWEILTLGHQ  
PYPAHSNLDV LNYVQTGGRLEPPRNC PDDLWNLMTQCWAQEPDQRPTFHRIQNQLQLFRN  
FFLNSIYQCRDEANNSGVINESEFEGEDGDVICLNSDDIMPVVLME TKNREGLNYMVLATE  
CGQGEKSEGP LGSQESE SCGLRKEEKEPHADKDFCQEKQVAYCPSGKPEGLNYACLTHS  
GYGDGSD

(La negrita distingue entre dos proteínas de fusión diferentes)

Fig. 4B

```

1 caagctttca agcattcaaa ggtctaaatg aaaaaggcta agtattatntt caaaaggcaa
61 gtatatccta atatatgcaaa acaaaacaaag caaaatccat cagctactcc tccaattgaa
121 gtgatgaagc ccaaataatt catatagcaa aatggagaaa attagaccgg ccatctaaaa
181 atctgccatt ggtgaagtga tgaagaacat ttactgtctt attccgaagc ttgtcaattt
241 tgcaactcct ggctgcctat ggatttctgt ggtgcagttg acagttttta atagctgcct
301 aaagtcgtgt gtaactaatc tgggcccagca gcttgacctt ggcacaccac ataactctgag
361 tgaaccgtgt atccaaggat gtcacttttg gaactctgta gatcagaaaa actgtgcttt
421 aaagtgctcg gagtcgtgtg aggttggtg tagcagcgcg gaaggtgcat atgaagagga
481 agtactggaa aatgcagacc taccaactgc tcccttgct tcttccattg gaaggccaaa
541 tatgacatta cgatggaaat ctgcaaactt ctctggagta aaatacatca ttcagtggaa
601 atatgcacaa cttctgggaa gctggactta tactaagact gtgtccagac cgtcctatgt
661 ggtcaagccc ctgcacccct tcactgagta cattttccga gtggtttgga tcttcacagc
721 ccgaactgca ctctactccc ctccaagtcc cagttacagg actcatcctc atgggattcc
781 tgaactgca cctttgatta ggaatattga gagctcaagt cccgacactg tggagtcag
841 ctgggatcca cctcaattcc caggtggacc tattttgggt tataacttaa ggtgatcag
901 caaaaatcaa aaattagatg cagggacaca gagaaccagt tccagtttt actccacttt
961 accaaatcact atctacaggt tttctattgc agcagtaaat gaagttggtg agggccaga
1021 agcagaatct agtattacca cttcatcttc agcagttcaa caagaggaac agtggctctt
1081 tttatccaga aaaacttctc taagaaagag atctttaaaa catttagtag atgaagcaca
1141 ttgccttcgg ttggatgcta tataccataa tattacagga atatctgttg atgtccacca
1201 gcaaattggt tatttctctg aaggaactct catatggcgc aagaaggctg ccaacatgac
1261 tgatgtatct gacctgagaa ttttttacag aggttcagga ttaatttctt ctatctccat
1321 agattggcct tatcaaagaa tgtatttcat catggatgaa ctggatgtg tctgtgattt
1381 agagaactgc tcaaacatcg aggaaattac tccaccctct attagtgcac ctcaaaaaat
1441 tgtggctgat tcatacaatg ggtatgtctt ttacctctg agagatggca tttatagagc
1501 agaccctcct gtaccatctg gccgggtgtc agaagctgtg cgtattgtgg agagttgcac
1561 gttaaaggac tttgcaatca agccacaagc caagcgaatc atttacttca atgacactgc
1621 ccaagtcttc atgtcaacat ttctggatgg ctctgcttcc catctcatcc tacctcgcac
1681 cccctttgct gatgtgaaaa gttttgcttg tgaaaaaat gactttcttg tcacagatgg
1741 caaggtcatt ttccaacagg atgctttgtc ttttaatgaa ttcatcgtgg gatgtgacct
1801 gagtcacata gaagaatttg ggtttggtaa cttggtcatc tttggctcat cctcccagct
1861 gcaccctctg ccaggccgcc cgcaggagct ttcgggtgctg tttggctctc accaggctct
1921 tgttcaatgg aagcctcctg ccctggccat aggagccaat gtcacctga atgagatgac
1981 tattgaactt tttgaattag gcccttctgc ctggcagAAC tggacctatg aggtgaaagt
2041 atccacccaa gaccctcctg aagtcactca tattttcttg aacataagtg gaacctatgct
2101 gaatgtacct gagctgcaga gtgctatgaa atacaaggtt tctgtgagag caagttctcc
2161 aaagaggcca ggcccctggg cagagccctc agtgggtact accctggtc cagctagta
2221 accaccattt atcatggctg tgaaagaaga tgggctttgg agtaaaccat taaatagctt
2281 tggcccagga gagttcttat cctctgatat aggaaatgtg tcagacatgg attggtataa
2341 caacagcctc tactacagtg acacgaaagg cgacgttttt gtgtggctgc tgaatgggac
2401 ggatatctca gagaattatc acctaccag cattgcagga gcaggggctt tagcttttga
2461 gtggctgggt cactttctct actgggctgg aaagacat atgtgataaa ggcagctgtt
2521 gttgacggga cacacagaca ttgttaccba cgtgaagcta ttggtgaatg acatgggtgt
2581 ggattcagtt ggtggatata tctactggac cacactctat tcagtggaaa gcaccagact
2641 aaatggggaa agttcccttg tactacagac acagccttgg ttttctggga aaaaggtaat
2701 tgctctaaact ttagacctca gtgatggct cctgtattgg ttggttcaag acagtcaatg
2761 tattcacctg tacacagctg ttcttcgggg acagagcact ggggatacca ccatcacaga
2821 atttgacagc ttgagttact ctgaaatttc ccagaatgca ctgatgtact atagtgtctg
2881 gctgttctgg atcaatggct ttaggattat cacaactcaa gaaataggte agaaaaccag
2941 tgtctctggt ttggaaccag ccagatttaa tcagttcaca attattcaga catcccttaa
3001 gccctgcca gggaaacttt cctttaccba taaggttatt ccagattctg ttcaagagtc
3061 ttcatattag atgaaggaa atgcttcaag ttttcaaatc ctgtggaatg gtcccctgc
3121 ggtagactgg ggtgtagttt tctacagttg agaatttagt gctcattcta agttcttggc
3181 tagtgaacaa cactctttac ctgtatttac tgtggaagga ctggaacctt atgccttatt
3241 taatctttct gtcactcctt atacctactg gggaaagggc cccaaaacat ctctgtcact
3301 tcgagcacc t gaaacagttc catcagcacc agagaacccc agaataatta tattaccaag
3361 tggaaaatgc tgcaacaaga atgaagttgt ggtggaattt aggtggaaca aaacctagca
3421 tgaaaatggg gtgttaacaa aatttgaat tttctacaat atatccaatc aaagtattac
3481 aaacaaaaca tgtgaagact ggattgctgt caatgtcact ccctcagtga tgtcttttca
3541 actgaagyc atgagtcba gatgctttat tgccttccag gttagggcct ttacatctaa
3601 gggcccagga ccatatgtct acgttgtaaa gtctacaaca tcagaaatca acccatttcc
3661 tcacctata actcttcttg gtaacaagat agttttttta gatatggatc aaaaatcaag
3721 tgtgtggacg ttttcagcag aaagagttat cagtgccgtt tgctacacag ctgataatga

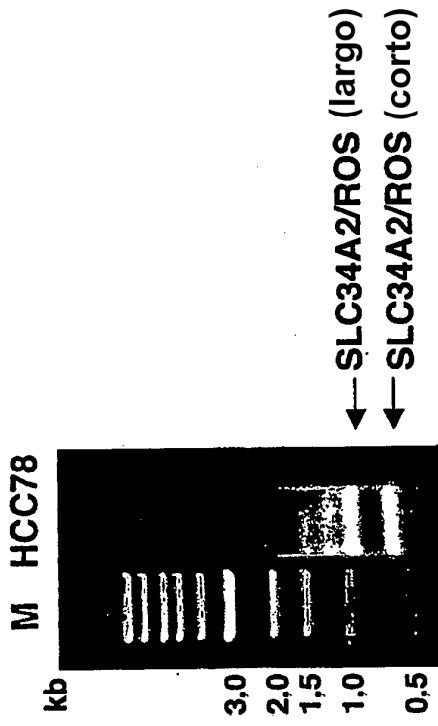
```

Fig. 4B (cont.)

3781 gatgggatat tatgctgaag gggactcact etttcttctg cacttgcaca atcgctctag  
 3841 ctctgagctt ttccaagatt cactggtttt tgatatacaca gttattacaa ttgactggat  
 3901 ttcaaggcac ctctactttg cactgaaaga atcacaaaat ggaatgcaag tatttgatgt  
 3961 tgatcttgaa cacaagtgga aatatcccag agagggtgaag attcacaata ggaattcaac  
 4021 aataatttct ttttctgtat atcctctttt aagtcgcttg tattggacag aagtttccaa  
 4081 ttttggctac cagatgttct actacagtat tadcagtcac accttgcacc gaattctgca  
 4141 acccacagct acaaaccaac aaaacaaaag gaatcaatgt tcttgtaatg tgactgaatt  
 4201 tgagttaagt ggagcaatgg ctattgatac ctctaaccta gagaaacct tgatatactt  
 4261 tgccaaagca caagagatct gggcaatgga tctggaaggc tgtcagtggt ggagagttat  
 4321 cacagtacct gctatgctcg caggaaaaac ccttgttagc ttaactgtgg atggagatct  
 4381 tatatactgg atcatcacag caaaggacag cacacagatt taccaggcaa agaaaggaaa  
 4441 tggggccatc gtttccagc tgaaggccct aaggagtagg catatcttgg cttacagttc  
 4501 agttatgcag ccttttccag ataaagcgtt tctgtctcta gcttcagaca ctgtggaacc  
 4561 aactatactt aatgccacta aactagcct cacaatcaga ttacctctgg ccaagacaaa  
 4621 cctcacatgg tatggcatca ccagccctac tccaacatac ctggtttatt atgcagaagt  
 4681 taatgacagg aaaaacagct ctgacttgaa atatagaatt ctggaatttc aggacagtat  
 4741 agtctttatt gaagatttac aaccttttc aacatacatg atacagatag ctgtaaaaaa  
 4801 ttattattaac gatcctttgg aacctttacc accaggaaaa gagatttggg gaaaaactaa  
 4861 aatggagta ccagaggcag tgcagctcat taatacaact gtgcggtcag acaccagcct  
 4921 cattatatct tggagagaat ctcaaacgac aaatggacct aaagaatcag tccggttatca  
 4981 gttggcaatc tcacacctgg ccctaattcc tgaactcct ctaagacaaa gtgaatttcc  
 5041 aatgggaagg ctactctcc ttgttactag actgtctggt ggaatatttt atgtgttaaa  
 5101 ggttcttgc ttgccactctg aggaaatgtg gtgtacagag agtcatctg tcaactgtgga  
 5161 aatgtttaac acaccagaga aaccttattc cttggttcca gagaacacta gtttgcaatt  
 5221 taattggaag gctccattga atgttaacct catcagattt tgggttgagc tacagaagtg  
 5281 gaaatacaat gagttttacc atgttaaac ttcatgcagc caaggtcctg cttatgtctg  
 5341 taatatacaca aatctacaac cttatacttc atataatgtc agagtagtggt tgggttataa  
 5401 gacgggagaa aatagcacct cacttcacga aagctttaag acaaaaagctg gagtcccaaa  
 5461 taaaccaggc attcccaaat tactagaagg gagtaaaaat tcaatacagt gggagaaggc  
 5521 tgaagataat ggatgtagaa ttacatacta tctccttgag ataagaaaga gcacttcaaa  
 5581 taattttacag aaccagaatt taagtggaag gatgacattt aatggatcct gcagtagtgt  
 5641 ttgcacatgg aagtccaaaa acctgaaagg aatatttcag ttcagagtag tagctgcaaa  
 5701 taatctaggg tttggtgaat atagtggaat cagtgagaat attatattag ttggagatga  
 5761 ttttggata ccagaaacaa gtttcatact tactattata gttggaatat ttctggttgt  
 5821 tacaatccca ctgacctttg tctggcatag aagattaaag aatcaaaaaa gtgccaaagg  
 5881 aggggtgaca gtgcttataa acgaagacaa agagttggct gagctgcagq gtctggcagc  
 5941 cgagtaggc ctggctaagc cctgctatgc aatacatact cttccaaccc aagaggagat  
 6001 tgaaaatctt cctgccttcc ctgggaaaaa actgactctg cgtctcttgc tgggaagtgg  
 6061 agcctttgga gaagtgtatg aaggaacagc agtggacatc ttaggagttg gaagtggaga  
 6121 aatcaaaata gcagtgaaag ctttgaaga ggttccaca gaccaggaca agattgaatt  
 6181 cctgaaggag gcacatctga tgaacaaatt taatcatccc aacattctga agcagcttgg  
 6241 agtttgtctg ctgaatgaac cccaatacat tctcctggaa ctgatggagg gaggagacct  
 6301 tcttacttat ttgcgtaaaag cccggatggc aacgttttat ggtcctttac tcaccttgg  
 6361 tgaccttqta gacctgtgtg tagatatttc aaaaggctgt gtctacttgg aacggatgca  
 6421 tttcattcac agggatctgg cagctagaaa ttgccttgtt tccgtgaaag actataccag  
 6481 tccacggata gtgaagattg gagactttgg actcggccaga gacatctata aaaatgatta  
 6541 ctatagaaag agagggaag gcctgctccc agttcgggtg atggctccag aaagtttgat  
 6601 ggatggaatc ttcactactc aatctgatgt atggtctttt ggaattctga tttgggagat  
 6661 tttaactctt ggtcatcagc cttatccagc tcattccaac cttgatgtgt taaactatgt  
 6721 gcaaacagga gggagactgg agccaccaag aatttgcctt gatgatctgt ggaatttaat  
 6781 gaccagtgq tgggtcaag aaccgacca aagacctact tttcatgaa ttcaggacca  
 6841 acttcagta ttcagaaatt tttcttaaa tagcatttat aagtcagag atgaagcaaa  
 6901 caacagtga gtcataaatg aaagctttga aggtgaagat ggcgatgtga tttgtttgaa  
 6961 ttcagatgac attatgccag ttgctttaat ggaacgaag aaccgagaag ggttaacta  
 7021 tatggtactt gctacagaat gtggccaagg tgaagaaaag totgagggtc ctctagqctc  
 7081 ccaggaatct gaatcttqtg gctctgagaa aqaagagaag gaaccacatg cagacaaaag  
 7141 tttctccaa gaaaaacaag tggctactg ccttctggc aagcctgag gctctgacta  
 7201 tgctgtctc actcacagtg gatatggaga tgggtctgat taatagcgtt gtttggaaa  
 7261 tagagagttg agataaacac tctcattcag tagttactga aagaaaactc tgctagaatg  
 7321 ataaatgtca tgggtgtcta taactccaaa taacaatgc aacgttcc



Fig. 6



Cebadores utilizados para RACE

ROS-GSP1: ACCCTTCTCGGTTCTTCGTTTCCA  
ROS-GSP2: GCAGCTCAGCCAACTCTTTGTCTT  
ROS-GSP3: TGCCAGACAAAGTCAAGTGGGATT

RACE se realizó empleando el sistema 5' RACE de Invitrogen

Fig. 7

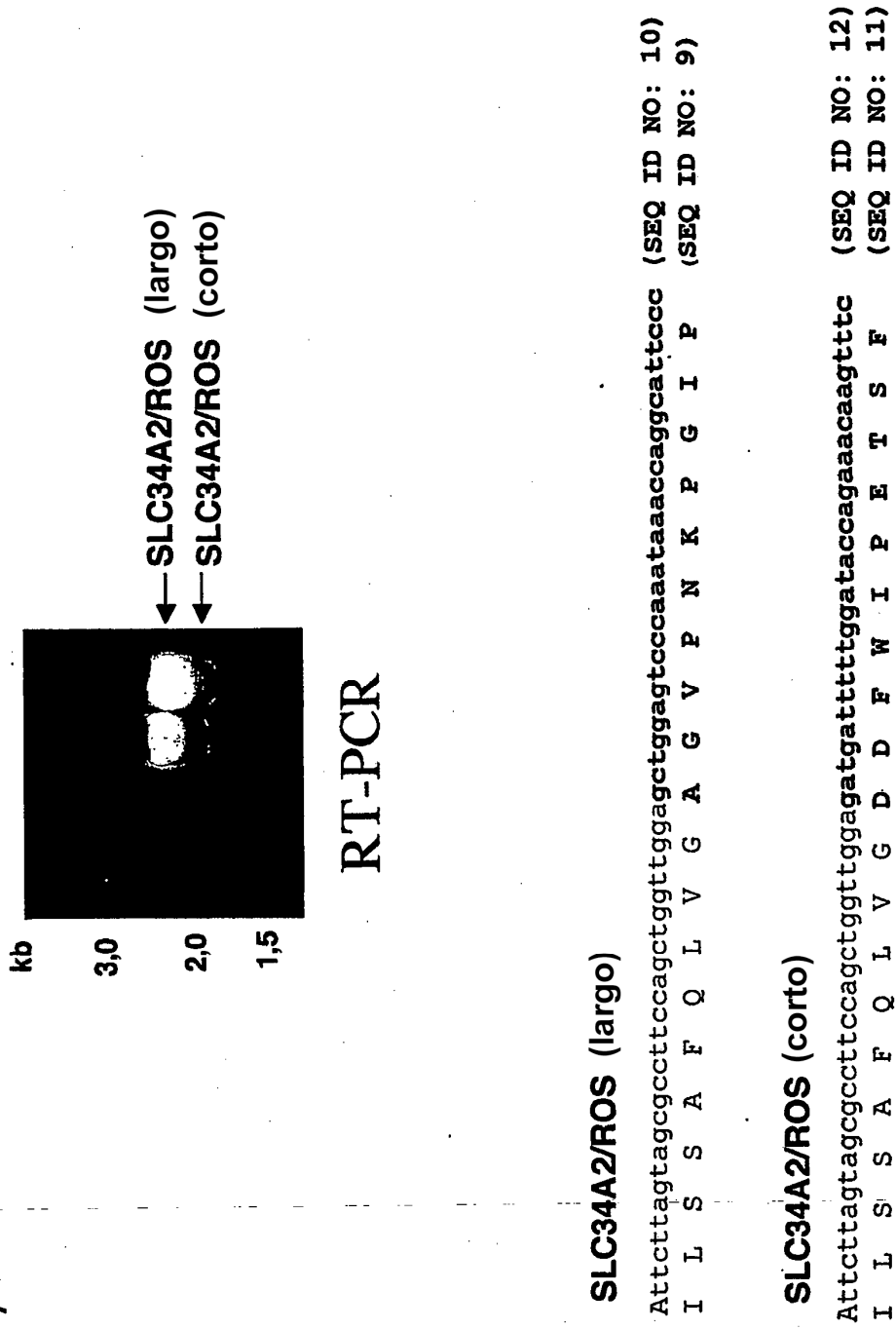


Fig. 8

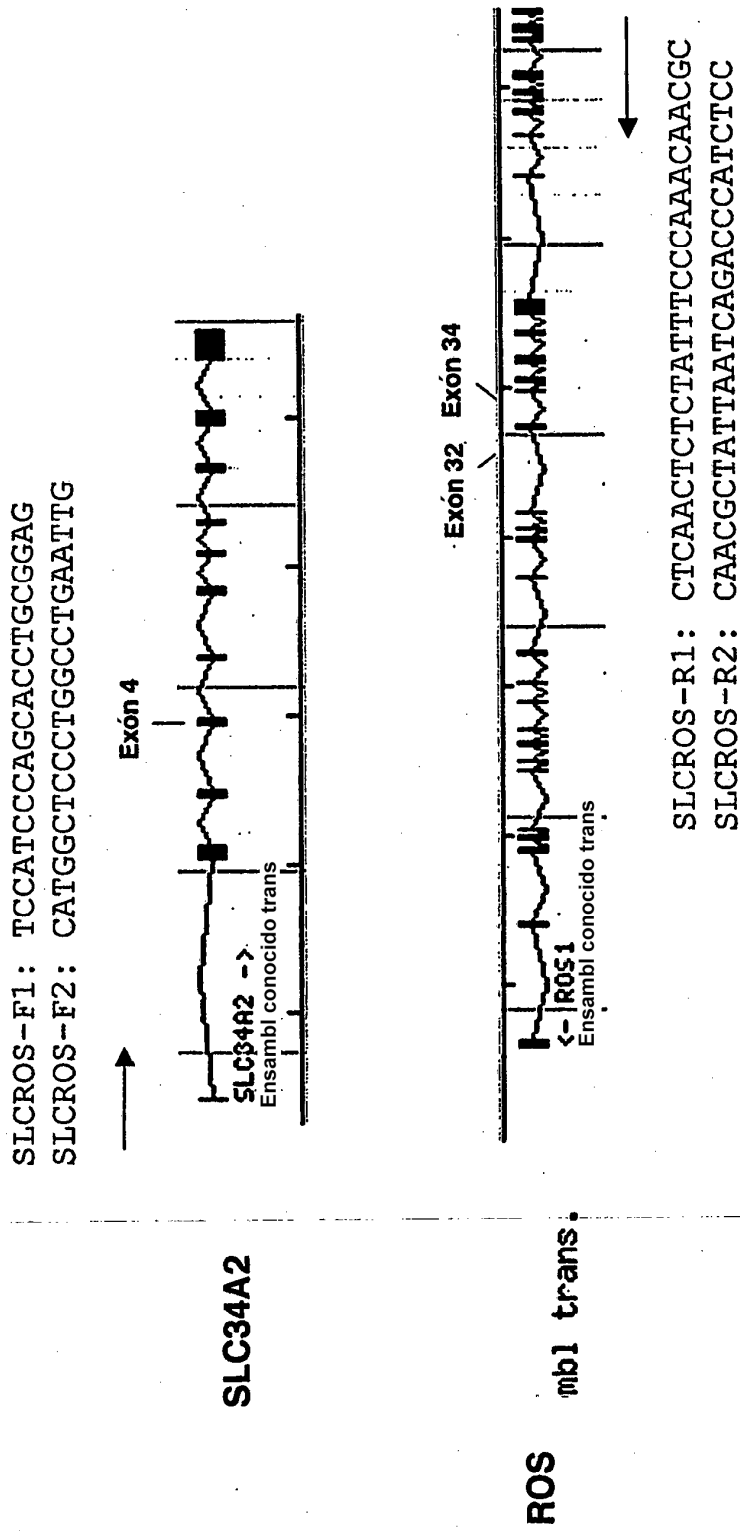
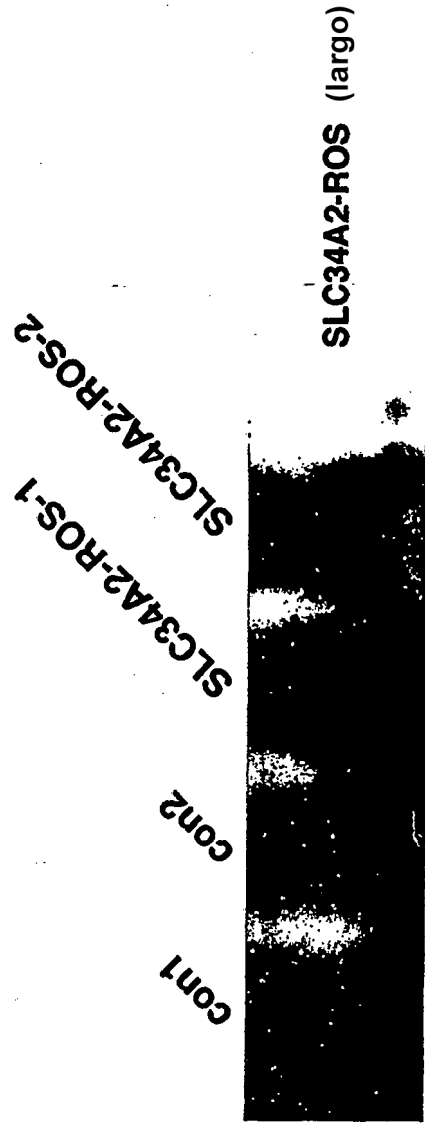


Fig. 9



293 transfectantes



Figura 10

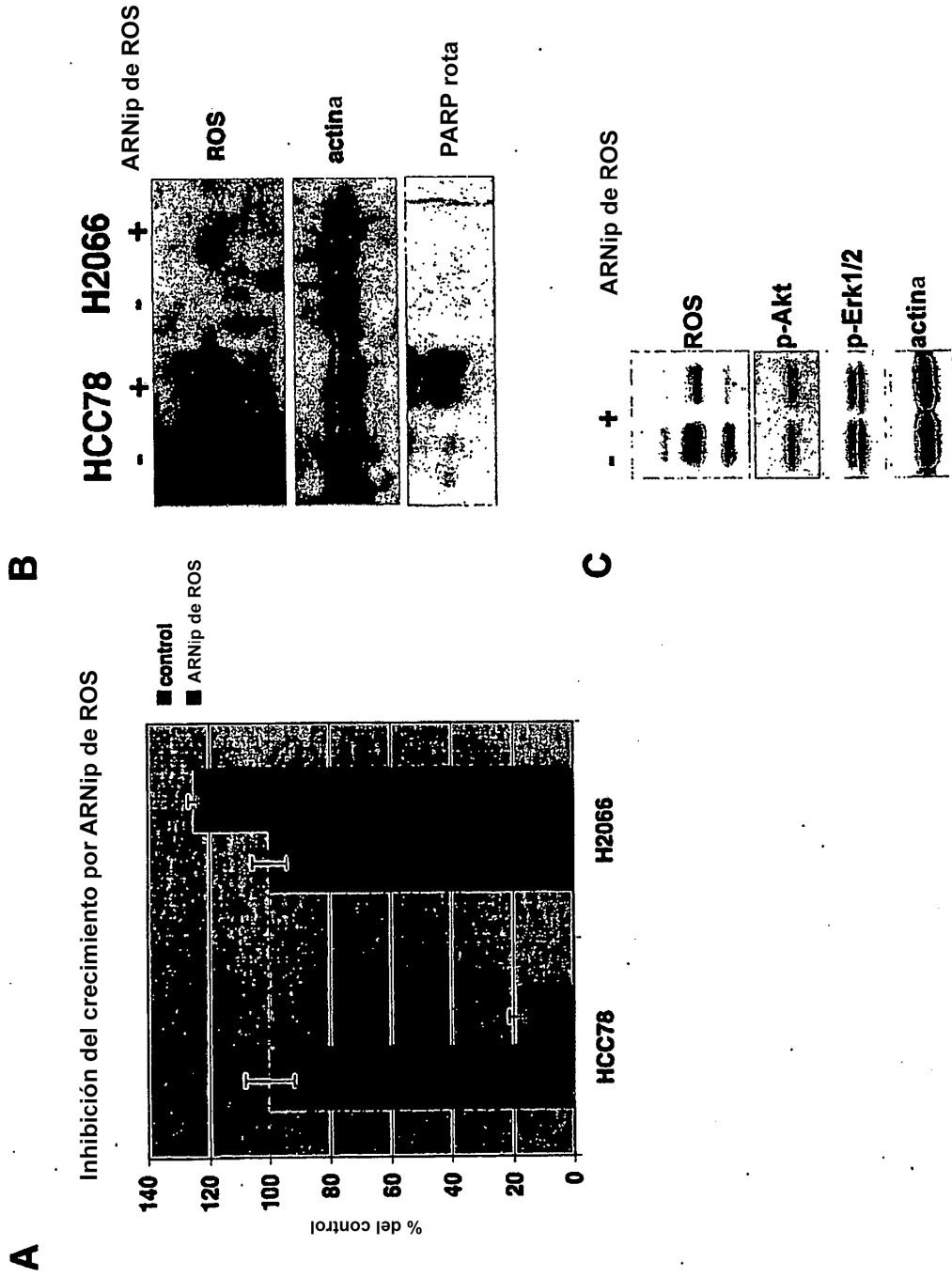


Figura 11

**C.**

