

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 660 149**

51 Int. Cl.:

**G01N 33/68** (2006.01)

**C12Q 1/68** (2008.01)

**G01N 33/15** (2006.01)

**G01N 33/574** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.02.2011 PCT/IB2011/000382**

87 Fecha y número de publicación internacional: **11.08.2011 WO11095894**

96 Fecha de presentación y número de la solicitud europea: **04.02.2011 E 11739469 (2)**

97 Fecha y número de publicación de la concesión europea: **10.01.2018 EP 2531858**

54 Título: **Identificación, evaluación y terapia de cánceres con resistencia innata o adquirida a inhibidores de ALK**

30 Prioridad:

**04.02.2010 US 337465 P**

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

**21.03.2018**

73 Titular/es:

**JICHI MEDICAL UNIVERSITY (100.0%)  
2-6-3 Hirakawa-cho, Chiyoda-ku  
Tokyo 102-0093 , JP**

72 Inventor/es:

**MANO, HIROYUKI;  
CHOI, YOUNG, L. y  
SODA, MANABU**

74 Agente/Representante:

**ISERN JARA, Jorge**

ES 2 660 149 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**

Identificación, evaluación y terapia de cánceres con resistencia innata o adquirida a inhibidores de ALK

5 Solicitudes relacionadas

Esta solicitud reivindica el derecho de prioridad para la Solicitud de Patente Provisional de Estados Unidos número de serie 61/337.465, presentada el 4 de febrero de 2010.

10 Antecedentes de la invención

Las tirosina quinasa son una clase de enzimas que catalizan la fosforilación de restos de tirosina de sustratos proteicos a través de una transferencia del fosfato terminal de la adenosina trifosfato. En muchos contextos, las tirosina quinasa desempeñan papeles críticos en la transducción de señales para una serie de funciones celulares, incluida la proliferación celular, la carcinogénesis y la diferenciación celular.

15 EML4-ALK es una proteína tirosina quinasa de tipo de fusión que está presente en ~ 5% de los casos de cáncer de pulmón no microcítico (CPNM) y que se genera como resultado de una pequeña inversión dentro del brazo corto del cromosoma 2 humano (Soda, M. *et al.* (2007) *Nature* 448: 561 - 566 ; Mano, H. (2008) *Cancer Sci.* 99: 2349 - 2357).

20 La EML4-ALK experimenta una dimerización constitutiva como resultado de la interacción entre el dominio de superhélice dentro de la región de EML4 de cada monómero y, de ese modo, adquiere una actividad oncogénica pronunciada. Los ratones transgénicos que expresan EML4-ALK, de forma específica en células epiteliales de pulmón, desarrollan cientos de nódulos de adenocarcinoma en ambos pulmones poco después del nacimiento, y la administración oral de un inhibidor específico de la actividad de tirosina quinasa de ALK erradica rápidamente tales nódulos de los pulmones (Soda, M. *et al.* (2008) *Proc. Natl. Acad. Sci. USA* 105: 19893 - 19897). Estas observaciones revelan el papel esencial de EML4-ALK en la carcinogénesis del CPNM que alberga esta quinasa de fusión, y respaldan aún más la factibilidad de la terapia direccionada de forma molecular con inhibidores de ALK para este cáncer. Por ejemplo, están en curso ensayos clínicos de un inhibidor, PF-02341066, de la actividad 25 tirosina quinasa de ALK y MET para el tratamiento del CPNM positivo para EML4-ALK, y sus resultados provisionales son prometedores (Kwak, E.L. *et al.* (2009) *J. Clin. Oncol.* 27 (supl.): 15s (resumen 3509)). Sin embargo, un subconjunto de tumores positivos para EML4-ALK no responde al inhibidor, con una base molecular desconocida para el fracaso del tratamiento.

35 Además de PF-02341066, se ha demostrado que otros inhibidores de tirosina quinasa (los TKI) poseen una actividad terapéutica pronunciada en pacientes con cáncer. El mesilato de imatinib, un TKI para ABL1 y KIT, por ejemplo, mejora notablemente el resultado de individuos con leucemia mieloide crónica positiva para la quinasa de fusión BCR-ABL1 o con un tumor del estroma gastrointestinal positivo para KIT activado (Druker, BJ *et al.* (2001) *N. Engl. J. Med.* 344: 1031-1037; Heinrich, M.C. *et al.* (2008) *J. Clin. Oncol.* 26: 5360 - 5367). Además, gefitinib y 40 erlotinib, que son los TKI para el receptor del factor de crecimiento epidérmico (EGFR), son eficaces en el tratamiento del CPNM asociado con la activación del EGFR (Mok, TS *et al.* (2009) *J. Clin. Oncol.* 27: 5080-2087; Mok, TS *et al.* (2009) *N. Engl. J. Med.* 361: 947 - 957). Desafortunadamente, un subconjunto de tumores diana son insensibles a los TKI correspondientes desde el inicio del tratamiento o se vuelven resistentes después de una respuesta inicial. En algunos casos de fracaso del tratamiento se han detectado mutaciones secundarias en las 45 quinasa diana que afectan directa o alostéricamente la forma del bolsillo de unión a ATP, dando como resultado la obstaculización de la unión al TKI (Deininger, M. *et al.* (2005) *Blood* 105: 2640-2653; Kobayashi, S. *et al.* (2005) *N. Engl. J. Med.* 352: 786-792; Pao, W. *et al.* (2005) *PLoS Med.* 2: e73 ; Shah, NP *et al.* (2002) *Cancer Cell* 2: 117 - 125). El documento WO2010/132888 divulga métodos y composiciones para detectar la presencia de un cáncer en un sujeto y evaluar la eficacia de los tratamientos para el mismo. El documento WO2009/103061 divulga métodos y 50 composiciones para identificar, diagnosticar y tratar el neuroblastoma.

Por consiguiente, existe una necesidad inmediata de identificar mutaciones que confieren resistencia a las tirosina quinasa, tal como EML4-ALK, para desarrollar mejor composiciones, kits y métodos para identificar, evaluar, 55 prevenir y tratar trastornos relacionados con su expresión y/o actividad anómala.

Sumario de la invención

La materia objeto de la protección cuya protección se busca es la definida en las reivindicaciones.

60 La presente divulgación proporciona, al menos, métodos para la identificación y evaluación del cáncer basándose en la identificación de una nueva mutación (o mutaciones) de la quinasa del linfoma anaplásico (ALK) que confiera resistencia a inhibidores de la ALK conocidos. Tales mutaciones de ALK también son importantes desde el punto de vista clínico para la identificación de composiciones farmacéuticas que tienen la capacidad de encajar en el bolsillo de unión a ATP anómalo generado por la mutación (o mutaciones) de la ALK e inhibir la actividad de la ALK.

65

En un aspecto, la presente invención proporciona un método para identificar a un sujeto que tiene cáncer o en riesgo de desarrollar un cáncer como que tiene un riesgo aumentado de falta de respuesta al tratamiento con un inhibidor de la ALK, que comprende:

5 analizar una muestra obtenida del sujeto para detectar la presencia de una o más moléculas de polinucleótido de ALK mutantes que codifican un polipéptido ALK mutante que comprende una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre o para detectar la presencia de uno o más polipéptidos ALK mutantes que comprenden una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre,  
 10 en el que la presencia de una o más moléculas de polinucleótido de ALK mutantes o polipéptidos de ALK mutantes indica que el sujeto tiene un riesgo aumentado de falta de respuesta al tratamiento con el inhibidor de ALK.

En otro aspecto, la presente invención proporciona un método *in vitro* para determinar si un compuesto de prueba modula la actividad de uno o más polipéptidos de ALK mutantes, que comprende:

(a) poner en contacto células de mamífero transfectadas con una construcción que codifica el uno o más polipéptidos ALK mutantes que comprenden una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre con el compuesto de prueba; y

(b) evaluar las células de mamífero para la actividad del uno o más polipéptidos ALK mutantes, en las que la actividad modulada de forma significativa en presencia del compuesto de prueba con respecto a un experimento de control, identifica el compuesto de prueba como un modulador de uno o más polipéptidos ALK mutantes.

En otro aspecto, la presente invención proporciona el uso de una molécula de polinucleótido ALK mutante que codifica un polipéptido ALK mutante que comprende una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre o el uso de un polipéptido ALK mutante que comprende una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre, en un método *in vitro* de diagnóstico del riesgo aumentado de un sujeto de falta de respuesta a un tratamiento con un inhibidor de ALK.

En algunas realizaciones de cualquier aspecto de la presente invención, el sujeto no se ha tratado anteriormente con un inhibidor de ALK, o se ha tratado anteriormente con un inhibidor de ALK y ha desarrollado al menos una resistencia parcial al inhibidor de ALK (por ejemplo, PF-02341066, PDD, 2-metil-11-(2-metilpropil)-4-oxo-4,5,6,11,12,13-hexahidro-2H-indazol[5,4-a]pirrol [3,4-c]carbazol-8-il [4-(dimetilamino)bencil] carbamato, (1S,2S,3R,4R)-3-({5-cloro-2-[(1-etil-2,3,4,5-tetrahidro-6-metoxi-2-oxo-1H-1-benzazepin-7-il)amino]-4-pirimidinil}amino)biciclo[2.2.1]hept-5-en-2-carboxamida y NVP-TAE684). En otras realizaciones, el cáncer se selecciona del grupo que consiste en linfoma anaplásico de células grandes, neuroblastoma, cáncer de mama, cáncer colorrectal, tumores miofibroblásticos inflamatorios y cánceres de pulmón no microcítico. En aún otras realizaciones, la muestra se selecciona del grupo que consiste en esputo, lavado broncoalveolar, derrame pleural, tejido, sangre completa, suero, plasma, raspado bucal, saliva, líquido cefalorraquídeo, orina, heces, células tumorales en circulación, ácidos nucleicos en circulación y médula ósea. En otras realizaciones más, la muestra comprende células o tejido. En algunas realizaciones, el tejido es un tejido tumoral o canceroso. En otras realizaciones, la una o más moléculas de polinucleótidos o polipéptidos de ALK mutantes se seleccionan del grupo que consiste en las moléculas de polinucleótidos o polipéptidos de ALK mutantes enumerados en la Tabla 1. En aún otras realizaciones, la una o más mutaciones de ALK se evalúan mediante un ensayo de hibridación de ácidos nucleicos. En otras realizaciones más, la una o más mutaciones de ALK se evalúan mediante reacción en cadena de la polimerasa. En otras realizaciones, el nivel de expresión del uno o más polipéptidos ALK se detecta usando un reactivo que se une de forma específica a uno o más polipéptidos ALK (por ejemplo, un anticuerpo, un derivado de anticuerpo y un fragmento de anticuerpo). En aún otras realizaciones, la cantidad, estructura y/o actividad del uno o más polipéptidos ALK mutantes se compara con una muestra de control. En aún otras realizaciones, la una o más mutaciones de ALK se evalúan en un primer punto temporal y en al menos un punto temporal posterior. En otras realizaciones, la muestra comprende ADN de estirpe germinal o genómico somático.

La presente divulgación proporciona un método para tratar a un paciente que tiene cáncer, o en riesgo de desarrollar cáncer, que comprende recoger una muestra del paciente, analizar la muestra para detectar la presencia de una o más moléculas de polinucleótido ALK mutante expuestas en la Tabla 1 y administrar a dicho paciente una cantidad terapéuticamente eficaz de un inhibidor de ALK. En algunas realizaciones, el inhibidor de ALK se selecciona del grupo que consiste en PF-02341066, PDD, 2-metil-11-(2-metilpropil)-4-oxo-4,5,6,11,12,13-hexahidro-2H-indazol [5,4-a]pirrol[3,4-c]carbazol-8-il[4-(dimetilamino)bencil]carbamato, (1S,2S,3R,4R)-3-({5-cloro-2-[(1-etil-2,3,4,5-tetrahidro-6-metoxi-2-oxo-1H-1-benzazepin-7-il)amino]-4-pirimidinil}amino)biciclo[2.2.1]hept-5-en-2-carboxamida y NVP-TAE684. En otras realizaciones, el sujeto no se ha tratado anteriormente con un inhibidor de ALK o se ha tratado anteriormente con un inhibidor de ALK y ha desarrollado al menos resistencia parcial al inhibidor de ALK.

La presente divulgación proporciona un kit para determinar la quimiosensibilidad al tratamiento con un inhibidor de ALK de un paciente con cáncer, que comprende: un reactivo que se une de forma específica a una o más moléculas de polinucleótido o polipéptidos de ALK mutante, e instrucciones para su uso. En algunas realizaciones, el kit

comprende adicionalmente un inhibidor de ALK. En otras realizaciones, el reactivo comprende una o más sondas de polinucleótido, cada una de las cuales comprende una secuencia de polinucleótido que es complementaria a una secuencia de nucleótidos enumerada en la Tabla 1 o complementaria a una secuencia de nucleótidos que codifica un polipéptido enumerado en la Tabla 1 (por ejemplo, oligonucleótidos, moléculas de ADNc, moléculas de ARN y sondas de genes sintéticas que comprenden nucleobases). En otras realizaciones más, las sondas comprenden polinucleótidos de aproximadamente 50 a  $10^7$  nucleótidos de longitud. En aún otras realizaciones, el reactivo comprende un anticuerpo y un derivado de anticuerpo, y un fragmento de anticuerpo frente a un polipéptido codificado por una o más secuencias de polinucleótido enumeradas en la Tabla 1.

La presente divulgación proporciona un método para determinar si un compuesto de prueba modula la actividad de uno o más polipéptidos ALK mutantes, que comprende poner en contacto células de mamífero transfectadas con una construcción que codifica el uno o más polipéptidos ALK mutantes con el compuesto de prueba y evaluar las células de mamífero para la actividad del uno o más polipéptidos ALK mutantes, en el que la actividad modulada de forma significativa en presencia del compuesto de prueba con respecto a un experimento de control identifica al compuesto de prueba como un modulador del uno o más polipéptidos ALK mutantes. En algunas realizaciones, la una o más moléculas de polinucleótido o polipéptidos de ALK mutantes se seleccionan del grupo que consiste en las moléculas de polinucleótido o polipéptidos de ALK mutantes enumeradas en la Tabla 1. En otras realizaciones, el control comprende células de mamífero que expresan un polipéptido ALK de tipo silvestre seleccionado del grupo que consiste en los polipéptidos enumerados en la Tabla 1. En aún otras realizaciones, la actividad del uno o más polipéptidos ALK mutantes se selecciona del grupo que consiste en unión a ATP, actividad de tirosina quinasa, proliferación de células cancerosas, crecimiento tumoral, número de tumores, apoptosis y metástasis tumoral. En aún otras realizaciones, el experimento de control comprende células de mamífero que expresan el uno o más polipéptidos ALK mutantes en ausencia del compuesto de prueba, determinado por, por ejemplo, la actividad del uno o más polipéptidos ALK mutantes (por ejemplo, unión a ATP, actividad de tirosina quinasa, proliferación de células cancerosas, crecimiento tumoral, número de tumores, apoptosis y metástasis tumoral).

#### Breve descripción de las figuras

La Figura 1 representa las mutaciones de ALK nuevas de la presente invención asociadas con la resistencia a inhibidores de la tirosina quinasa ALK. La Figura 1A muestra una representación esquemática de la proteína EML4-ALK. Se muestran las posiciones de dos mutaciones *de novo* en el dominio quinasa y las de los cebadores de PCR para la amplificación del dominio quinasa o los ADNc de fusión se indican mediante las flechas negras y blancas superiores, respectivamente. La Figura 1B muestra los resultados de la secuenciación profunda de los ADNc del dominio quinasa de ALK. Los productos de PCR de ~1000 pb de la línea celular de CPNM H2228 o de las muestras de ensayo de las ID J-# 1, J-#12, J-#113, J-#127 o LK-#33 se secuenciaron con el sistema GAIL. Los números para la cobertura de lectura total (Total) y las lecturas no coincidentes (no coincidente) se muestran en cada posición de los ADNc del dominio quinasa con rombos azules y rojos, respectivamente. Los recuadros muestran vistas ampliadas para la región 5' de los ADNc para J-#1 y J-#113 (representados por rectángulos verdes). La Figura 1C muestra electroforetogramas para los clones de ADNc de ALK que rodean las posiciones G4374 y C4493. La PCR se realizó con ADNc preparados a partir de esputo obtenido antes del tratamiento (Inicial) y de las células del derrame pleural obtenido después de la recaída (Recaída). Los nucleótidos A sustituidos se muestran en rojo.

La Figura 2 representa secuencias genómicas que rodean las posiciones que corresponden a G4374 y C4493 del ADNc de ALK. El ADN genómico aislado de células del derrame pleural del paciente se sometió a PCR durante 35 ciclos de 94 °C durante 15 s, 60 °C durante 30 s y 68 °C durante 2 min, con Taq ADN polimerasa Platinum (Invitrogen, Carlsbad, CA) y los siguientes cebadores (5'-GGTAAGAAGTGGCTCACTCTTGAG-3' y 5'-CACAACAAGTGCAGCAAAGACTGG-3'), y los productos se ligaron en el plásmido pT7Blue-2 (Takara Bio). Después, los insertos de los plásmidos se secuenciaron con el Analizador Genético 3130x1, lo que dio como resultado la identificación de clones de PCR que contenían los cambios G4374A (panel izquierdo) o C4493A (panel derecho). Los nucleótidos A sustituidos se muestran en rojo.

La Figura 3 representa los resultados de las células BA/F3 tratadas con PF-02341066. Las células BA/F3 que expresan EML4-ALK (tipo silvestre), EML4-ALK(C1156Y), EML4-ALK(L1196M) o el doble mutante EML4-ALK(C1156Y/L1196M) se incubaron en presencia de las concentraciones indicadas de PF-02341066 durante 48 h, después de lo cual se examinó la morfología celular mediante microscopía de contraste de fase. Barra de escala, 20  $\mu$ m.

La Figura 4 representa propiedades de las mutaciones de ALK nuevas de la presente invención asociadas con la resistencia a los inhibidores de la tirosina quinasa ALK. La Figura 4A muestra el número de células BA/F3 que expresan EML4-ALK (tipo silvestre), EML4-ALK(C1156Y), EML4-ALK(L1196M) o el doble mutante EML4-ALK(C1156Y/L1196M) contadas después de la incubación de  $5 \times 10^5$  células durante 48 h con las concentraciones indicadas de PF-02341066. El porcentaje de células viables se muestra con respecto a las células BA/F3 que expresan la EML4-ALK de tipo silvestre. Los datos son medias  $\pm$  d.t. de tres experimentos distintos. La Figura 4B muestra el efecto de PF-02341066 sobre la fosforilación de tirosina de las formas de tipo silvestre o mutantes de EML4-ALK. Las células BA/F3 que expresan la EML4-ALK de tipo silvestre etiquetada

con FLAG o sus mutantes se expusieron a las concentraciones indicadas de PF-02341066 durante 15 h, tras lo cual se inmunoprecipitó EML4-ALK a partir de lisados celulares y se sometió a análisis por inmunotransferencia con anticuerpos específicos para ALK fosforilada en Tyr<sup>1604</sup> o para el epítipo FLAG (ALK). Se examinaron como control negativo células que expresan un mutante inactivo de EML4-ALK (KM). La Figura 4C muestra un ensayo de quinasa *in vitro* para EML4-ALK de tipo silvestre etiquetada con FLAG o sus mutantes, inmunoprecipitados a partir de células BA/F3 (no expuestas a un inhibidor de ALK). Los inmunoprecipitados se incubaron con  $[\gamma\text{-}^2\text{P}]\text{ATP}$ , un péptido sintético y las concentraciones indicadas de PF-02341066. La fosforilación de los inmunoprecipitados del sustrato peptídico se sometió por separado a análisis por inmunotransferencia con anticuerpos frente a FLAG (panel inferior).

La Figura 5 representa un modelo de estructura tridimensional para el dominio quinasa de ALK. Las posiciones de aminoácido de ALK se superpusieron en la estructura cristalina del receptor de insulina con un análogo de ATP enlazado (ID "1ir3" en el Protein Data Bank of Japan, disponible en internet en [dbj.org/index.html](http://dbj.org/index.html)). El panel derecho muestra la estructura de la proteína observada desde el lado izquierdo del modelo del panel izquierdo. Las hélices  $\alpha$  y las láminas  $\beta$  se muestran en magenta y naranja, respectivamente. También se indican las posiciones Cys<sup>1156</sup> y Leu<sup>1196</sup> de la hélice  $\alpha$ C.

#### Descripción detallada de la invención

La presente invención se basa, al menos en parte, en la identificación de regiones específicas del genoma, que incluyen, por ejemplo, mutaciones de la quinasa del linfoma anaplásico (ALK), asociadas con la predicción de la eficacia de los inhibidores de ALK en el tratamiento del cáncer. En particular, se han identificado en el presente documento mutaciones nuevas del gen de ALK (por ejemplo, mutaciones que codifican el polipéptido EML4-ALK) que pueden conducir a polipéptidos al menos parcialmente resistentes a la terapia con inhibidores de ALK. La presente divulgación proporciona adicionalmente métodos para identificar tales regiones genómicas específicas usando técnicas conocidas en la técnica, que incluyen, pero sin limitación, micromatrices basadas en oligonucleótidos (Brennan, *et al.*, (2004) *Cancer Res.* 64 (14): 4744-8; Lucito, *et al.*, (2003) *Genome Res.* 13: 2291-2305; Bignell *et al.* (2004) *Genome Res.* 14: 287 - 295; Zhao *et al.* (2004) *Cancer Research*, 64 (9): 3060-71) y otros métodos como se describen en el presente documento que incluyen, por ejemplo, métodos basados en la reacción en cadena de la polimerasa (PCR) y en secuenciación directa. La presente invención proporciona adicionalmente kits de diagnóstico para su uso en los métodos.

Varios aspectos de la presente invención se describen con más detalle en las siguientes subsecciones.

#### I. Definiciones

Los artículos "un" y "una" se usan en el presente documento para referirse a uno o más de uno (es decir, a al menos uno) del objeto gramatical del artículo. A modo de ejemplo, "un elemento" significa un elemento o más de un elemento.

La expresión "cantidad modificada" de un marcador o "nivel modificado" de un marcador se refiere a un número de copias aumentado o disminuido de un marcador o región cromosómica, tales como mutaciones y/o productos génicos del gen ALK (por ejemplo, los marcadores expuestos en la Tabla 1), y/o el nivel de expresión aumentado o disminuido de un gen o genes marcadores particulares en una muestra de cáncer, en comparación con el nivel de expresión o el número de copias del marcador en una muestra de control. La expresión "cantidad modificada" de un marcador también incluye un nivel de proteína aumentado o disminuido de un marcador en una muestra, por ejemplo, una muestra de cáncer, en comparación con el nivel de proteína del marcador en una muestra de control normal.

La expresión "nivel alterado de expresión" de mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) se refiere a un nivel de expresión o número de copias de un marcador en una muestra de prueba, tal como una muestra obtenida de un paciente que padece cáncer, que es mayor o menor que el error típico del ensayo empleado para evaluar la expresión o el número de copias, y puede ser de al menos dos veces, al menos dos veces tres, al menos dos veces cuatro, al menos dos veces cinco o al menos dos veces diez, o más veces el nivel de expresión o el número de copias de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) en una muestra de control (por ejemplo, una muestra de un sujeto sano que no tiene la enfermedad asociada), o el nivel de expresión promedio o número de copias de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) en varias muestras de control. El nivel modificado de expresión es mayor o menor que el error típico del ensayo empleado para evaluar la expresión o el número de copias, y es al menos dos veces, al menos tres, al menos cuatro, al menos cinco, al menos diez o más veces el nivel de expresión o número de copias de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) en una muestra de control (por ejemplo, una muestra de un sujeto sano que no tiene la enfermedad asociada), o el nivel de expresión o número de copias promedio de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) en varias muestras de control.

5 El término "actividad modificada" de un marcador se refiere a una actividad de un marcador que está aumentada o disminuida en una enfermedad, por ejemplo, en una muestra de cáncer, en comparación con la actividad del marcador en una muestra de control normal. La actividad modificada de un marcador puede ser el resultado de, por ejemplo, la expresión modificada del marcador, el nivel modificado de proteína del marcador, la estructura modificada del marcador o, por ejemplo, una interacción modificada con otras proteínas implicadas en la misma ruta o en otra distinta a la del marcador.

10 El término "estructura modificada" de un marcador se refiere a la presencia de mutaciones o mutaciones dentro del gen marcador o proteína creadora, por ejemplo, mutaciones que afectan a la expresión o actividad del marcador, en comparación con el gen o proteína normal o de tipo silvestre. Por ejemplo, las mutaciones incluyen, pero sin limitación, un reordenamiento inter e intracromosómico, sustituciones, deleciones y mutaciones de inserción. Las mutaciones pueden estar presentes en la región codificante o no codificante del marcador.

15 "Quinasa del linfoma anaplásico" y "ALK" se usan indistintamente en el presente documento y se refieren a la quinasa del linfoma anaplásico nativa y a determinadas variantes y mutaciones de la misma, obtenidas de cualquier fuente (por ejemplo, roedores, seres humanos y otros mamíferos). En algunas realizaciones, la proteína ALK está representada por el número de identificación Ref. Seq. del NCBI. NP\_004295. A menos que se indique otra cosa, las expresiones se refieren a la proteína humana. El gen que codifica ALK también puede denominarse en el presente documento "ALK". En algunas realizaciones, las secuencias de nucleótidos de ALK están representadas por número de identificación Ref. Seq. del NCBI NM\_004304.3 y número de referencia del GenBank 29029631, secuencias relevantes en ellas (por ejemplo, las secuencias codificantes, la 5'UTR, la 3'UTR, el inicio de la transcripción, el inicio de la traducción, la terminación de la transcripción, la terminación de la traducción, etc.) que puede identificar fácilmente un experto en la materia.

25 Además, "quinasa del linfoma anaplásico" y "ALK" también se usan en el presente documento para incluir las quinazas ALK de fusión y variantes de las mismas, que son bien conocidas por un experto en la materia. Tales quinazas ALK de fusión y variantes de las mismas comprenden la actividad quinasa de la ALK y pueden albergar mutaciones como se describe en el presente documento, haciendo que la actividad quinasa de la ALK sea resistente a los inhibidores de ALK. Los ejemplos representativos incluyen EML4-ALK variante 1 (AB274722.1; BAF73611.1),  
 30 EML4-ALK variante 2 (AB275889.1; BAF73612.1), EML4-ALK variante 3a (AB374361.1; BAG55003.1), EML4-ALK variante 3b (AB374362.1; BAG55004.1), EML4-ALK variante 4 (AB374363.1; BAG75147.1), EML4-ALK variante 5a (AB374364.1; BAG75148.1), EML4-ALK variante 5b (AB374365.1; BAG75149.1), EML4-ALK variante 6 (AB462411.1; BAH57335.1), EML4-ALK variante 7 (AB462412.1; BAH57336.1), KIF5B-ALK (AB462413.1; BAH57337.1), NPM-ALK, TPM3-ALK, TFGXL-ALK, TFGL-ALK, TFGS-ALK, ATIC-ALK, CLTC-ALK, MSN-ALK,  
 35 TPM4-ALK, MYH9-ALK, RANBP2-ALK, ALO17-ALK y CARS-ALK (véase, por ejemplo, Pulford *et al.*, (2004) J. Cell. Physiol. 199: 330 - 358). Además, un experto en la materia entenderá que las variantes de ALK quinasa pueden surgir dependiendo del suceso de fusión particular entre una ALK quinasa y su compañero de fusión (por ejemplo, EML4 puede fusionarse con al menos el exón 2, 6a, 6b, 13, 14 y/o 15, como se describe, por ejemplo, en Horn y Pao, (2009) J. Clin. Oncol. 27: 4247-4253). Por ejemplo, las secuencias de ALK representativas se proporcionan en  
 40 el presente documento como sigue:

Tabla 1

Secuencia de ADNc de ALK de tipo silvestre (NM\_004304.3; GI: 29029631):

ES 2 660 149 T3

1 gggggcggca gcggtgtag cagctggtac ctcccgcgc ctctgttcgg agggctcggg  
 61 ggcaccgagg tgctttccgg ccgccctctg gtcggccacc caaagccgcy ggcgctgatg  
 121 atgggtgagg agggggcggc aagatttcgg ggcgccctgc cctgaacgcc ctgagctgct  
 181 gccgcccggg ccgctccagt gcctgcgaac tctgaggagc cgaggcggcg gtgagagcaa  
 241 ggacgctgca aacttgcgca ggcggggggc tgggattcac gcccagaagt tcagcaggca  
 301 gacagtccga agccttcccg cagcggagag atagcttgag ggtgcgcaag acggcagcct  
 361 ccgccctcgg ttcccgccca gaccgggcag aagagcttgg aggagccaaa aggaacgcaa  
 421 aaggcggcca ggacagcgtg cagcagctgg gagccgccgt tctcagcctt aaaagttgca  
 481 gagattggag gctgccccga gaggggacag accccagctc cgactgcggg gggcaggaga  
 541 ggacggtacc caactgccac ctcccctcaa ccatagtagt tcctctgtac cgagcgcagc  
 601 gagctacaga cgggggcgcy gcaactcggc cggagagcgg gaggctcaag gtcccagcca  
 661 gtgagcccag tgtgcttgag tgtctctgga ctgcgccctg agcttccagg tctgtttcat  
 721 ttagactcct gctcgcctcc gtgcagttgg gggaaagcaa gagacttgcy cgcacgcaca  
 781 gtcctctgga gatcagggtg aaggagccgc tgggtaccaa ggactgttca gagcctcttc  
 841 ccactctcgg gagagcgaag ggtgaggctg ggcccggaga gcagtgtaaa cggcctcctc  
 901 cggcgggatg ggagccatcg ggtcctctgt gctcctgccg ctgctgcttt ccacggcagc  
 961 tgtgggctcc gggatgggga ccggccagcy cggggctcc ccagctgcgg ggcggccgct  
 1021 gcagccccgg gagccactca gctactcgc cctgcagagg aagagtctgg cagttgactt  
 1081 cgtggtgccc tcgctcttcc gtgtctacgc ccgggacctc ctgctgccac catcctcctc  
 1141 ggagctgaag gctggcaggc ccgaggcccy cggctcgtca gctctggact gcgccccgct  
 1201 gctcaggttg tggggccgcy cgcgggggt ctccctggacc gccggttcac cagccccggc  
 1261 agaggcccgg acgctgtcca ggggtgctgaa gggcggctcc gtgcgcaagc tccggcgtgc  
 1321 caagcagttg gtgctggagc tgggcgagga ggcgatcttg gaggttgcy tcgggcccc  
 1381 cggggaggcy gctgtggggc tgtctcagtt caatctcagc gagctgttca gttggtggat  
 1441 tcgccaaggc gaagggcgac tgaggatccg cctgatgccg gagaagaagg cgtcgggaagt  
 1501 gggcagagag ggaaggctgt ccgcggcaat tcgcgcctcc cagccccgcc ttctcttcca  
 1561 gatcttcggg actggtcata gctccttggg atcaccaaca aacatgcctt ctctctctcc  
 1621 tgattathtt acatggaatc tcacctggat aatgaaagac tcttccctt tctgtctca  
 1681 tcgagccga tatggtctgg agtgcagctt tgacttcccc tgtgagctgg agtattcccc  
 1741 tccactgcat gacctcagga accagagctg gtccctggcy cgcaccccc cagaggaggc  
 1801 ctcccagatg gacttgctgg atgggctgg ggcagagcgt tctaaggaga tgcccagagg  
 1861 ctcttttctc cttctcaaca cctcagctga ctccaagcac accatcctga gtccgtggat  
 1921 gaggagcagc agtgagcact gcacactggc cgtctcggtg cacaggcacc tgagccctc  
 1981 tggaaggtag attgcccagc tgtgcccga caacgaggct gcaagagaga tctctctgat  
 2041 gccactcca gggaaagcat gttggacagt gctccaggga agaatcgggc gtccagacaa  
 2101 cccatttctga gtggccctgg aatacatctc cagtggaaac cgcagcttgt ctgcagtgga  
 2161 cttctttgcc ctgaagaact gcagtgaagg aacatcccca ggctccaaga tggccttcca  
 2221 gagctccttc acttgttggg atgggacagt cctccagctt gggcaggcct gtgacttcca  
 2281 ccaggactgt gccaggggag aagatgagag ccagatgtgc cggaaactgc ctgtgggttt  
 2341 ttactgcaac tttgaagatg gcttctgtgg ctggacccaa ggcacactgt caccacacac  
 2401 tcctcaatgg caggtcagga ccctaaagga tgcccggttc caggaccacc aagaccatgc  
 2461 tctattgctc agtaccactg atgtccccgc ttctgaaagt gctacagtga ccagtgtctc  
 2521 gtttctctgca ccgatcaaga gctctccatg tgagctccga atgtcctggc tcattcgtgg  
 2581 agtcttgagg ggaaacgtgt ccttgggtgct agtggagaa aaaaccggga aggagcaagg  
 2641 caggatggtc tggcatgtcg ccgcctatga aggcttgagc ctgtggcagt ggatgggtgt  
 2701 gcctctctc gatgtgtctg acaggttctg gctgcagatg gtcgcatggt ggggacaagg  
 2761 atccagagcc atcgtggctt ttgacaatat ctccatcagc ctggactgct acctaccat  
 2821 tagcggagag gacaagatcc tgcagaatac agcaccaaa tcaagaaacc tgtttgagag  
 2881 aaacccaaac aaggagctga aaccggggga aaattcacca agacagacc ccacttttga  
 2941 ccctacagtt cattggctgt tcaccacatg tggggccagc gggccccatg gccccacca  
 3001 ggcacagtgc aacaacgcct accagaactc caacctgagc gtggagggtg ggagcgggg  
 3061 ccccctgaaa ggcacccaga tctggaagg gcccagccacc gacacctaca gcactcggg  
 3121 ctacggagct gctggcggga aaggcgggaa gaacaccatg atgcggctcc acggcgtgtc  
 3181 tgtgctgggc atcttcaacc tggagaagga tgacatgctg tacatcctgg ttgggcagca  
 3241 gggagaggac gcctgccccg gtacaaacca gttaatccag aaagtctgca ttggagagaa  
 3301 caatgtgata gaagaagaaa tccgtgtgaa cagaagcgtg catgagtggg caggaggcgg  
 3361 aggaggaggg ggtggagcca cctacgtatt taagatgaag gatggagtgc cgggtcccct

ES 2 660 149 T3

```

3421 gatcattgca gccggaggtg gtggcagggc ctacggggcc aagacagaca cgttccaccc
3481 agagagactg gagaataact cctcggttct agggctaaac ggcaattccg gagccgcagg
3541 tgggtggaggt ggctggaatg ataacacttc cttgctctgg gccggaaaat ctttgcagga
3601 ggggtgccacc ggaggacatt cctgccccca ggccatgaag aagtgggggt gggagacaag
3661 agggggtttc ggagggggtg gaggggggtg ctctcaggt ggaggaggcg gaggatata
3721 aggcggcaat gcagcctcaa acaatgacct cgaaatggat ggggaagatg gggtttcctt
3781 catcagtcca ctgggcatcc tgtacacccc agctttaaaa gtgatggaag gccacgggga
3841 agtgaatatt aagcattatc taaactgcag tcaactgtgag gtagacgaat gtcacatgga
3901 ccctgaaagc cacaaggtca tctgcttctg tgaccacggg acggtgctgg ctgaggatgg
3961 cgtctcctgc atttgtgtcac ccaccccggg gccacacctg ccaactctgc tgatcctctc
4021 tgtggtgacc tctgcccctg tggccgccct ggctctggct ttctccggca tcatgattgt
4081 gtaccgccgg aagcaccagg agctgcaagc catgcagatg gagctgcaga gccctgagta
4141 caagctgagc aagctccgca cctcgacct catgaccgac tacaacccca actactgctt
4201 tgctggcaag acctcctcca tcagtgaact gaaggaggtg ccgcgaaaaa acatcacctt
4261 cattgggggt ctggggccatg gcgcctttgg ggaggtgat gaaggccagg tgtccggaat
4321 gcccacgac ccaagcccc tgcaagtggc tgtgaagacg ctgctgaag tgtcctga
4381 acaggacgaa ctggatttcc tcatggaagc cctgatcatc agcaaattca accaccagaa
4441 cattgttcgc tgcattgggg tgagcctgca atccctgccc cggttcatcc tgctggagct
4501 catggcgggg ggagacctca agtccttct cagagagacc cgccctcgcc cgagccagcc
4561 ctctcctctg gccatgctgg accttctgca cgtggctcgg gacattgcct gtggctgtca
4621 gtatttgag gaaaaccact tcatccaccg agacattgct gccagaaact gcctcttgac
4681 ctgtccaggc cctggaagag tggccaagat tggagacttc gggatggccc gagacatcta
4741 cagggcgagc tactatagaa agggaggctg tgccatgctg ccagttaagt ggatgcccc
4801 agaggccttc atggaaggaa tattcacttc taaaacagac acatggtcct ttggagtgt
4861 gctatgggaa atcttttctc ttggatata gccatacccc agcaaaagca accaggaagt
4921 tctggagttt gtcaccagtg gaggccggat ggaccacccc aagaactgcc ctgggcctgt
4981 atacgggata atgactcagt gctggcaaca tcagcctgaa gacaggccca actttgccat
5041 cattttggag aggattgaat actgcaccca ggaccocggat gtaatcaaca ccgctttgcc
5101 gatagaatat ggtccacttg tggaagagga agagaaagtg cctgtgaggc ccaaggaccc
5161 tgagggggtt cctcctctcc tggctctca acaggcaaaa cgggaggagg agcgcagccc
5221 agctgccccca ccacctctgc ctaccacctc ctctggcaag gctgcaaaga aaccacagc
5281 tgcagagatc tctgttcgag tcctagagg gccggccgtg gaagggggac acgtgaatat
5341 ggcattctct cagtccaacc ctctctcgga gttgcacaag gtccacggat ccagaaacaa
5401 gccaccagc ttgtggaacc caacgtacgg ctctgggtt acagagaaa ccacaaaaa
5461 gaataacct atagcaaga agggccaca cgacaggggt aacctggggc tggaggaag
5521 ctgtactgtc ccacctaacg ttgcaactgg gagacttccg gggccctcac tgcctctaga
5581 gccctcttgc ctgactgcca atatgaagga ggtacctctg ttcaggctac gtcacttccc
5641 ttgtgggaat gtcaattacg gctaccagca acagggcttg cccttagaag ccgctactgc
5701 ccctggagct ggtcattacg aggataccat tctgaaaagc aagaatagca tgaaccagcc
5761 tgggccctga gctcggctgc acactcaact ctcttcttg ggatccctaa gaccgtggag
5821 gagagagagg caatggctcc ttcacaaacc agagaccaa tgtcacgttt tgtttgtgc
5881 caacctattt tgaagtacca ccaaaaaagc tgtattttga aaatgcttta gaaaggtttt
5941 gagcatgggt tcactctatt ctttcgaaag aagaaaatat cataaaaatg agtgataaat
6001 acaaggccca gatgtggttg cataagggtt ttatgcatgt ttgtgtata cttccttatg
6061 cttctttcaa atttgtgtgt ctctgcttca atgtagtcag aattagctgc ttctatgttt
6121 catagttggg gtcatagatg tttccttgc ttgttgatgt ggacatgagc catttgaggg
6181 gagagggaa c ggaataaag gagttatttg taatgactaa aa

```

Mutación (o mutaciones) del codón TGC (4373 a 4375) de la secuencia de ADNc de tipo silvestre que codifica un aminoácido distinto de cisteína o una mutación correspondiente en una homóloga de la misma

5 Mutación (o mutaciones) del codón CTG (4493 a 4495) de la secuencia de ADNc de tipo silvestre que codifica un aminoácido distinto de leucina o una mutación correspondiente en una homóloga de la misma

Mutación G4374A de la secuencia de ADNc de tipo silvestre o una mutación correspondiente en una homóloga de la misma

10 Mutación C4493A de la secuencia de ADNc de tipo silvestre o una mutación correspondiente en una homóloga de la misma

Secuencia de la proteína ALK de tipo silvestre (NP\_004295.2; GI: 29029632):

1 mgaigllwll plllstaavg sgmgtgqrag spaagpplqp replsysrlq rkslavdfvv  
 61 pslfrvyard lllppsssel kagrpeargs laldcapllr llgpapgvsw tagspapaea  
 121 rtlsrvlkgg svrklrrakq lvlelgeeeai legcvgppge aavgllqfnl selfswwirq  
 181 gegrlrirlm pekkasevgr egrlsaaira sqprllfqif gtghsslesp tnmppspdy  
 241 ftwnltwimk dsfpflshrs ryglecsfdf pceleysppl hdlrnqswsw rripseeasq  
 301 mdlldgpgae rskemprgsf lllntsadsk htllspwmrs ssehctlavs vhrhlqpsgr  
 361 yiaqllphne aareillmpt pgkhgwtvlq grigrpdnpf rvaleyissg nrslsavdff  
 421 alkncsegts pgskmalqss ftcwngtvlq lggacdfhqd caqgedesqm crklpvgyfc  
 481 nfedgfcgwt qgtlsphtpq wqvrtlkdar fqdhqdhall lsttdvpase satvtsatfp  
 541 apiksspcel rmswlrngvl rgnvslvlve nktgkeqgrm vwhvaayegl slwqwmvlpl  
 601 ldvsdrfwlq mvawwgqgsr aivafdnsi sldcyltisg edkilqntap ksrnlfernp  
 661 nkelkpgens prqtpifdpt vhwlfittcga sgphgptqaq cnnayqnsnl svevgsegpl  
 721 kgiqiwkvpv tdtysisgyg aaggkggknt mmrshgvsvl gifnlekddm lyilvgqqge  
 781 dacpstnqli qkvcigennv ieeeeirvnrs vhwagggggg gggatyvfm kdgvvpplii  
 841 aaggggrayg aktdtfhper lennssvlgl ngnsagaagg ggwndntsll wagkslqega  
 901 tghscpqam kkwgwetrgg fggggggcss ggggggyig naasnndpem dgedvsfis  
 961 plgilytpal kvmeghgevn ikhylnchc evdechmdpe shkvicfcdh gtvlaedgvs  
 1021 civsptpeph lplslilsvv tsalvaalvl afsgimivyr rkhqelqamq melqspeykl  
 1081 sklrtstimt dynpnycfag ktssisdlike vprknlitlr glghgafgev yegqvsmpn  
 1141 dpsplqvavk tlpevcseqd eldfmeali iskfnhqniv rcigvslqsl prfillelma  
 1201 ggdlksflre trprpsqps lamldllhva rdiacgcqyl eenhfihrdi aarnclltcp  
 1261 gpgrvakigd fgwardiyra syyrkggcam lpvkwmppea fmegiftskt dtwsfgvllw  
 1321 eifslgympy psksnqevle fvtsggrmdp pknpcgpvyr imtqcwqhqp edrpnfail  
 1381 erieyctqdp dvintalpie ygplveeek vpvprkdpeg vppllvsqqa kreeerspaa  
 1441 ppplpttssg kaakkptaee isvrvrpgrpa vegghvnmaf sqsnppselh kvhgsrnkpt  
 1501 slwnptygsw ftekptkkn piakkephdr gnlglegsct vppnvatgrl pgasllleps  
 1561 sltanmkevp lfrrlhfcg nvnygyqqqg lpleaatapg aghyedtilk sknsmnqpgp

Mutación Cys1156Xaa de la secuencia de la proteína de tipo silvestre en la que Xaa es un aminoácido distinto de cisteína o una mutación correspondiente en una homóloga de la misma

5 Mutación Leu1196Xaa de la secuencia de la proteína de tipo silvestre en la que Xaa es un aminoácido distinto de leucina o una mutación correspondiente en un homólogo de la misma

Mutación Cys1156Tyr de la secuencia de la proteína de tipo silvestre o una mutación correspondiente en una homóloga de la misma

10 Mutación Leu1196Met de la secuencia de la proteína de tipo silvestre o una mutación correspondiente en una homóloga de la misma

Secuencia del ADNc de EML4-ALK variante 1 (AB274722.1; GI: 152002652)

1 ggcggcgcgcg cgcggcgctc gcggtgctg cctgggaggg aggcggggca ggcggctgag  
 61 cggcgcggtct ctcaacgtga cggggaagtg gttcggggcg cgcgggctta ctaccccagg  
 121 gcgaacggac ggacgacgga ggcgggagcc ggtagccgag ccgggagacc tagagaacga  
 181 gcgggtcagg ctacagctcg gccactctgt cggctcgctg aatgaagtgc ccgcccctct  
 241 gagcccggag cccggcgctt tccccgcaag atggacggtt tcgccggcag tctcgatgat  
 301 agtatttctg ctgcaagtac ttctgatgtt caagatcgcc tgctcagctct tgagtcacga  
 361 gttcagcaac aagaagatga aatcactgtg ctaaaggcgg ctttggtgta tgttttgagg  
 421 cgtcttgcaa tctctgaaga tcatgtggcc tcagtgaaaa aatcagctctc aagtaaaggc  
 481 caaccaagcc ctcgagcagt tattcccatg tcctgtataa ccaatggaag tggtgcaaac  
 541 agaaaaccaa gtacataccag tgctgtctca attgcaggaa aagaaactct ttcatctgct  
 601 gctaaaagtg gtacagaaaa aaagaaagaa aaaccacaag gacagagaga aaaaaagag  
 661 gaatctcatt ctaatgatca aagtccacaa attcgagcat caccttctcc ccagccctct  
 721 tcacaacctc tccaaataca cagacaaaact ccagaaagca agaatgtac tcccaccaa  
 781 agcataaaac gaccatcacc agctgaaaag tcacataatt cttgggaaaa ttcagatgat

ES 2 660 149 T3

841 agccgtaata aattgtcgaa aataccttca acacccaaat taataccaaa agttaccaa  
 901 actgcagaca agcataaaga tgtcatcadc aaccaagaag gagaatata taaaatgttt  
 961 atgcgcggtc ggccaattac catgttcatt ccttccgatg ttgacaacta tgatgacatc  
 1021 agaacggaac tgcctcctga gaagctcaaa ctggagtggg catatggtta tcgaggaaag  
 1081 gactgtagag ctaatgttta ccttcttccg accggggaaa tagtttattt cattgcatca  
 1141 gtagtagtac tatttaatta tgaggagaga actcagcgac actacctggg ccatacagac  
 1201 tgtgtgaaat gccttgctat acatcctgac aaaattagga ttgcaactgg acagatagct  
 1261 ggcgtggata aagatggaag gcctctacaa ccccacgtca gagtgtggga ttctgttact  
 1321 ctatccacac tgcagattat tggacttggc acttttgagc gtggagtagg atgcctggat  
 1381 ttttcaaaag cagattcagg tgttcattta tgtgttattg atgactccaa tgagcatatg  
 1441 cttactgtat gggactggca gaagaaagca aaaggagcag aaataaagac acaaatgaa  
 1501 gttgttttgg ctgtggagtt tcaccaaca gatgcaata ccataattac atgvcggtaaa  
 1561 tctcatatth tcttctggac ctggagcggc aattcactaa caagaaaaca ggaattttt  
 1621 gggaaatag aaaagccaaa atttgtgcag tgtttagcat tcttggggaa tggagatggt  
 1681 cttactggag actcaggtgg agtcatgctt atatggagca aaactactgt agagcccaca  
 1741 cctgggaaag gacctaaagt gtaccgcggc aagcaccagg agctgcaagc catgcagatg  
 1801 gagtgcgaga gccctgagta caagctgagc aagctccgca cctcgacct catgaccgac  
 1861 tacaacccca actactgctt tgctggcaag acctcctcca tcagtgcact gaaggagggtg  
 1921 ccgvcgaaaa acatcaccct cattcggggg ctgggccatg gagcctttgg ggagggtgat  
 1981 gaaggccagg tgtccggaat gcccaacgac ccaagccccc tgcaagtggc tgtgaagacg  
 2041 ctgcctgaag tgtgctctga acaggacgaa ctggatttcc tcatggaagc cctgatcadc  
 2101 agcaaattca accaccagaa cattgttcgc tgcattgggg tgagcctgca atccctgccc  
 2161 cggttcatcc tgctggagct catggcgggg ggagacctca agtcccttct cagagagacc  
 2221 cgcctcgc ccgagccagcc ctctcctctg gccatgctgg accttctgca cgtggctcgg  
 2281 gacattgcct gtggctgtca gtatttggag gaaaaccact tcatccaccg agacattgct  
 2341 gccagaaact gcctcttgac ctgtccaggc cctggaagag tggccaagat tggagacttc  
 2401 gggatggccc gagacatcta cagggcgcgc tactatagaa agggaggctg tgccatgctg  
 2461 ccagttaagt ggatgcccc agaggccttc atggaaggaa tattcacttc taaaacagac  
 2521 acatggctct ttggagtgtc gctatgggaa atcttttctc ttggatataat gccatacccc  
 2581 agcaaaagca accaggaagt tctggagttt gtcaccagtg gaggccggat ggaccacccc  
 2641 aagaactgcc ctgggcctgt ataccggata atgactcagt gctggcaaca tcagcctgaa  
 2701 gacaggccca actttgccat cattttggag aggattgaat actgcacca ggaccvcggt  
 2761 gtaatcaaca ccgctttgccc gatagaatat ggtccacttg tggaaagagga agagaaagtg  
 2821 cctgtgagge ccaaggaccc tgaggggggt cctcctctcc tggctctca acaggcaaaa  
 2881 cgggaggagg agcgcagccc agctgcccce cacctctgc ctaccacct ctctggcaag  
 2941 gctgcaaaga aaccacagc tgcagaggtc tctgttcgag tccctagagg gccggccgtg  
 3001 gaagggggac acgtgaatat ggcattctct cagtccaacc ctcttcgga gttgcacagg  
 3061 gtccacggat ccagaaacaa gccaccagc ttgtggaacc caacgtacgg ctctggttt  
 3121 acagagaaac ccacaaaaa gaataatcct atagcaaaaga aggagccaca cgagaggggt  
 3181 aacctggggc tggagggaaag ctgtaactgtc ccacctaacg ttgcaactgg gagacttccg  
 3241 ggggcctcac tgctcctaga gccctcttcg ctgactgcca atatgaagga ggtacctctg  
 3301 ttcaggctac gtcacttccc ttgtgggaaat gtcaattacg gctaccagca acagggcttg  
 3361 cccttagaag ccgctactgc cctggagct ggtcattacg aggataccat tctgaaaagc  
 3421 aagaatagca tgaaccagcc tgggcctga gctcggtcac aactcactt ctcttcttg  
 3481 ggatccctaa gaccgtggag gagagagagg caatcaatgg ctcttcaca aaccagagac  
 3541 caaatgtcac gttttgttt gtgccaacct attttgaagt accacaaaaa aagctgtatt  
 3601 ttgaaaatgc tttagaaagg ttttgagcat gggttcatcc tattctttcg aaagaagaa  
 3661 atatcataaa aatgagtgat aaatacaagg cccagatgtg gttgcataag gttttatgc  
 3721 atgtttgttg tatacttctt tatgcttctt ttaaattgtg tgtgctctgc ttcaatgtag  
 3781 tcagaattag ctgcttctat gtttcatagt tggggtcata gatgtttcct tgcttgttg  
 3841 atgtggacat gagccatttg aggggagagg gaacggaaat aaaggagtta tttgtaatga  
 3901 aaaaaaaaaa aaaaaaaaaa aaaaaa

Secuencia de la proteína EML4-ALK variante 1 (BAF73611.1; GI: 152002653)

1 mdgfagsldd sisaastsdv qdrilsalesr vqqgedeitv lkaaladvlr rlaisedhva  
 61 svkksvsskg qpspravipm scitngsgan rkpshsavs iagketlssa aksgtekkke  
 121 kpqgqrekke eshsndqspq iraspspqs sqplqihrt pesknaptk sikrpspaek  
 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii nqegeyikmf mrgrpitmfi  
 241 psdvdnyddi rtelppekkl lewaygyrgk dcranvylp tgeivyfias vvvlfnyeer

## ES 2 660 149 T3

301 tqrhylghtd cvkclaihpd kiriatgqia gvdkdgrplq phvrwdsvt lstlqiiglg  
361 tfergvgcld fskadsgvhl cviddsnehm ltvwdwqkka kgaeiktne vvlavefhpt  
421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq claflngdv ltgdsggvml  
481 iwskttept pgkpkvyrr khqelqamqm elqspeyqls klrtstimtd ypnycfagk  
541 tssisdikev prknitlirg lghgafgevy egqvsgmpnd psplqvavkt lpevcseqde  
601 ldflmealii skfnhqnivv cigvslqslp rfillelmag gdlksflret rprpsqpssl  
661 amldllhvar diacgcqyle enhfihrdia arnclltcpg pgrvakigdf gmardiyas  
721 yyrkggcaml pvkwmppeaf megiftsktd twsfgvllwe ifslgympyp sksnqevlef  
781 vtsggrmdpp knpcgpyvri mtqcwqhqpe drpnfaiile riehtqdqpd vintalpiey  
841 gplveeeekv pvrpkdpegv ppllvsqqak reerspaap pplpttssgk aakkptaaev  
901 svrvprgpav egghvnmafs qsnppselhr vhgsrnkpts lwnptygswf tekptkknnp  
961 iakkepherg nlglegsctv pnvatgrlp gaslllepss ltanmkevpl frlrhfpcgn  
1021 vnygyqqqgl pleaatapga ghyedtilks knsmnqpgp

Secuencia del ADNc de EML4-ALK variante 2 (AB275889.1; GI: 152002654)

ES 2 660 149 T3

1 ggcgggcggg cgcggegetc ggggtgctg cctgggaggg aggccgggca ggcggctgag  
 61 cggcgcggtt ctcaacgtga cggggaagtg gttcggggcg ccgcggtta ctaccccagg  
 121 gcgaacggac ggacgacgga ggcgggagcc ggtagccgag ccgggagacc tagagaacga  
 181 gcgggtcagg ctacgctcg gccactctgt cggtcgctg aatgaagtgc ccgcccctct  
 241 gagcccgagg cccggcgctt tccccgcaag atggacgggt tcgcccgcag tctcgatgat  
 301 agtatttctg ctgcaagtac ttctgatggt caagatcgcc tgcagctct tgagtacga  
 361 gttcagcaac aagaagatga aatcactgtg ctaaaggcgg ctttggtgta tgttttgagg  
 421 cgtcttgcaa tctctgaaga tcatgtggcc tcagtgaata aatcagtctc aagtaaaggc  
 481 caaccaagcc ctgagcagt tattcccatg tctgtataa ccaatggaag tggtgcaaac  
 541 agaaaaccaa gtcataccag tgctgtctca attgcaggaa aagaaactct ttcactctgct  
 601 gctaaaagtg gtacagaaaa aaagaaagaa aaaccacaag gacagagaga aaaaaagag  
 661 gaatctcatt ctaatgatca aagtcacaaa attcgagcat caccttctcc ccagccctct  
 721 tcacaacctc tccaaatata cagacaaact ccagaaagca agaatgctac tcccacaaa  
 781 agcataaaa gaccatcacc agctgaaaag tcacataatt cttgggaaaa ttcagatgat  
 841 agccgtaata aattgtcgaa aataccttca acaccaaata taatacaca agttacaaa  
 901 actgcagaca agcataaaga tgcctcctga tgcctcctga tgcctcctga tgcctcctga  
 961 atgcgcggtc ggccaattac catgttcatt ccttcgatg ttgacaacta tgatgacatc  
 1021 agaacggaac tgcctcctga gaagctcaaa ctggagtggg catatgggta tcgaggaaag  
 1081 gactgtagag ctaatgttta ccttcttccg accggggaaa tagtttattt cattgcatca  
 1141 gtagtagtac tatttaatta tgaggagaga actcagcgac actacctggg ccatacagac  
 1201 tgtgtgaaat gccttgctat acatcctgac aaaattagga ttgcaactgg acagatagct  
 1261 ggcgtggata aagatggaag gcctctacaa ccccacgtca gagtgtggga ttctgttact  
 1321 ctatccacac tgcagattat tggacttggc acttttgagc gtggagtagg atgcctggat  
 1381 ttttcaaaag cagattcagg tgttcattta tgtgttattg atgactcaa tgagcatatg  
 1441 cttactgtat gggactggca gaagaaagca aaaggagcag aaataaagac acaaatgaa  
 1501 gttgttttgg ctgtggagtt tcaccaaca gatgcaata ccataattac atgcggtaaa  
 1561 tctcatatth tcttctggac ctggagcggc aattcactaa caagaaaaca gggaaattht  
 1621 gggaaatag aaaagccaaa atttgtgag tgttttagcat tcttggggaa tggagatggt  
 1681 cttactggag actcaggtgg agtcatgctt atatggagca aaactactgt agagcccaca  
 1741 cctgggaaag gacctaaagg tgtatatcaa atcagcaaac aaatcaaagc tcatgatggc  
 1801 agtgtgttca cactttgtca gatgagaaat gggatgttat taactggagg agggaaagac  
 1861 agaaaaataa ttctgtggga tcatgatctg aatcctgaaa gagaaataga ggttcctgat  
 1921 cagtatggca caatcagagc tgtagcagaa ggaaaggcag atcaattht agtaggcaca  
 1981 tcacgaaact ttattttacg aggaacattt aatgatggct tccaaataga agtacagggt  
 2041 catacagatg agctttgggg tcttgccaca catcccttca aagatttgct cttgacatgt  
 2101 gctcaggaca ggcaggtgtg cctgtggaac tcaatggaac acaggctgga atggaccagg  
 2161 ctggtagatg aaccaggaca ctgtgcagat tttcatcaa gtggcacagt ggtggccata  
 2221 ggaacgcact caggcaggtg gtttgttctg gatgcagaaa ccagagatct agtttctatc  
 2281 cacacagacg ggaatgaaca gctctctgtg atgcgctact caatagatgg taccttctg  
 2341 gctgtaggat ctcatgacaa ctttatttac ctctatgtag tctctgaaaa tggaaagaaa  
 2401 tatagcagat atggaaggtg cactggacat tccagctaca tcacacacct tgactgggcc  
 2461 ccagacaaca agtatataat gtctaactcg ggagactatg aaatattgta cttgtaccgc  
 2521 cggaagcacc aggagctgca agccatgcag atggagctgc agagccctga gtacaagctg  
 2581 agcaagctcc gcacctcgac catcatgacc gactacaacc ccaactactg ctttgtctggc

ES 2 660 149 T3

2641 aagacctcct ccatcagtga cctgaaggag gtgccgcgga aaaacatcac cctcattcgg  
 2701 ggtctgggcc atggagcctt tggggaggtg tatgaaggcc aggtgtccgg aatgcccac  
 2761 gacccaagcc ccctgcaagt ggctgtgaag acgctgcctg aagtgtgctc tgaacaggac  
 2821 gaactggatt tcctcatgga agccctgac atcagcaaat tcaaccacca gaacattgtt  
 2881 cgctgcattg ggggtgagcct gcaatccctg ccccggttca tctgtctgga gctcatggcg  
 2941 gggggagacc tcaagtcctt cctccgagag acccgccctc gcccgagcca gccctcctcc  
 3001 ctggccatgc tggaccttct gcacgtggct cgggacattg cctgtggctg tcagtatttg  
 3061 gagaaaacc acttcatcca ccgagacatt gctgccagaa actgcctctt gacctgtcca  
 3121 ggccctggaa gagtggccaa gattggagac ttccggatgg cccgagacat ctacagggcg  
 3181 agtactata gaaagggagg ctgtgccatg ctgccagtta agtggatgcc cccagaggcc  
 3241 ttcatggaag gaatattcac ttctaaaaca gacacatggt cctttggagt gctgctatgg  
 3301 gaaatctttt ctcttgata tatgccatac cccagcaaaa gcaaccagga agttctggag  
 3361 tttgtacca gtggaggccg gatggacca cccaagaact gccctgggcc tgtataccgg  
 3421 ataatgactc agtgtctggca acatcagcct gaagacaggc ccaactttgc catcattttg  
 3481 gagaggattg aatactgcac ccaggaccog gatgtaatca acaccgcttt gccgatagaa  
 3541 tatgtccac ttgtggaaga ggaagagaaa gtgcctgtga ggcccaagga cccgagggg  
 3601 gttcctcctc tctgtgtctc tcaacagga aaacgggagg aggagcgcag cccagctgcc  
 3661 ccaccacctc tgcctaccac ctctctggc aaggctgcaa agaaaccac agctgcagag  
 3721 gtctctgttc gagtccctag agggccggcc gtggaagggg gacacgtgaa tatggcattc  
 3781 tctcagtcca accctccttc ggagttgcac aggggccacg gatccagaaa caagcccacc  
 3841 agcttgtgga acccaacgta cggctcctgg tttacagaga aaccaccaa aaagaataat  
 3901 cctatagcaa agaaggagcc acacgagagg ggtaacctgg ggctggaggg aagctgtact  
 3961 gtcccaccta acgttgcaac tgggagactt cggggggcct cactgctcct agagccctct  
 4021 tcgctgactg ccaatatgaa ggaggtacct ctgttcaggc tacgtcactt cccttgtggg  
 4081 aatgtcaatt acggctacca gcaacagggc ttgcccttag aagccgctac tgcccctgga  
 4141 gctggtcatt acgaggatac cattctgaaa agcaagaata gcatgaacca gcctgggcc  
 4201 tgagctcggg cacacactca ctctcttcc ttgggatccc taagaccgtg gaggagagag  
 4261 aggcaatcaa tggctccttc aaaaaccaga gaccaaatgt cacgttttgt tttgtgcaa  
 4321 cctattttga agtaccacca aaaaagctgt attttgaaaa tgctttagaa aggttttgag  
 4381 catgggttca tcctattctt tcgaaagaag aaaatatcat aaaaatgagt gataaataca  
 4441 aggccagat gtggttgcac aagggtttta tgcattgttg ttgtatactt ccttatgctt  
 4501 cttttaaatt gtgtgtgctc tgcttcaatg tagtcagaat tagctgcttc tatgtttcat  
 4561 agttggggtc atagatggtt ccttgccctg ttgatgtgga catgagccat ttgaggggag  
 4621 agggaacgga aataaaggag ttatttgtaa tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa

Secuencia de la proteína EML4-ALK Variante 2 (BAF73612.1; GI: 152002655)

1 mdgfagsldv sisaastsdv qdralsalesr vqqqedeitv lkaaladvlr rlaisedhva  
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke  
 121 kpggqrekke eshsndqspq iraspspps sqplqihrt pesknaptk sikrpspaek  
 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii ngegeyikmf mrgrpitmfi  
 241 psdvdnyddi rtelppekkl lewaygyrgk dcranvllp tgeivyfias vvvlfnyeer  
 301 tqrhylghtd cvkclaihp kiriatgqia gvdkdgrplq phvrwdsvt lstlqiiglg  
 361 tfergvglcd fskadsgvhl cviddsnehm ltvwdwqkka kgaeikttne vvlavefhpt  
 421 dantiitcgc shiffwtwsg nsltrkqgif gkyekpkfvq clafllngdv ltgdsggvml  
 481 iwsktttvept pgkpgkgyvq iskqikahdg svftlcqmrn gmlttgggkd rkiilwdhdl  
 541 npereievpd qygtiravae gkadqflvgt srnfilrgtf ndgfqievqg htdelwglat  
 601 hpfkdlldtc aqdrqvclwn smehrlewtr lvdephcad fhpsgtvvai gthsgrwfvl  
 661 daetrldvsi htdgneqlsv mrysidgtfl avgshdnfiy lyvvsengr ysrygrctgh  
 721 ssyithldws pdnkyimsns gdeilylyr rkhqelqamq melqspeykl sklrtstimt  
 781 dynpnycfag ktssisdike vprknitlir glghgafgev yegqvsgmpn dpsplqvavk  
 841 tlpevcseqd eldfleali iskfnhqniv rcigvslqsl pfillelma ggdksflre  
 901 trprpsqps lamldllhva rdiacgcqyl eenhfihrdi aarnclltpc gpgrvakigd  
 961 fgmardiyra syyrkggcam lpvkwmppea fmegiftskt dtwsfgvllw eifslgympy  
 1021 psksnqevle fvtsggrmdp pkncpgpvyr imtqcwqhqp edrpnfaiil erieyctqdp  
 1081 dvintalpie ygplveeek vpvprkdpeg vppllvsqqa kreespaa ppplpttssg  
 1141 kaakkptaee vsrvprgpa vegghvnmaf sqsnppselh rvhgrrnkpt slwnptygsw  
 1201 ftekptkkn piakkepher gnlglegsc vppnvatgrl pgasllleps sltanmkevp  
 1261 lfrlrhfpcg nvnygyqqqg lpleaatapg aghyedtilk sknsmnqpgp

Secuencia de ácido nucleico de EML4-ALK Variante 3a (AB374361.1; GI: 194072592)

1 actctgtcgg tccgctgaat gaagtgcccg cccctctaag cccggagccc ggcgctttcc  
 61 ccgcaagatg gacggtttcg ccggcagtct cgatgatagt atttctgctg caagtacttc  
 121 tgatgttcaa gatcgctgtt cagctcttga gtcacgagtt cagcaacaag aagatgaaat  
 181 cactgtgcta aaggcggctt tggctgatgt tttgaggcgt cttgcaatct ctgaagatca  
 241 tgtggcctca gtgaaaaaat cagtctcaag taaaggccaa ccaagccctc gagcagttat  
 301 tcccatgtcc tgtataacca atggaagtgg tgcaaacaga aaaccaagtc ataccagtgc  
 361 tgtctcaatt gcaggaaaag aaactctttc atctgctgct aaaagtggta cagaaaaaaaa  
 421 gaaagaaaaa ccacaaggac agagagaaaa aaaagaggaa tctcattcta atgatcaaa  
 481 tccacaaatt cgagcatcac cttctcccca gccctcttca caacctctcc aaatacacag  
 541 acaaaactca gaaagcaaga atgctactcc caccaaaagc ataaaacgac catcaccagc  
 601 tgaaaagtca cataattctt gggaaaaatc agatgatagc cgtaataaat tctcgaaaaat  
 661 accttcaaca cccaaattaa taccaaaagt taccaaaact gcagacaagc ataaagatgt  
 721 catcatcaac caagtgtacc gccggaagca ccaggagctg caagccatgc agatggagct  
 781 gcagagccct gagtacaagc tgagcaagct ccgcacctcg accatcatga ccgactacia  
 841 ccccaactac tgctttgctg gcaagacctc ctccatcagt gacctgaagg aggtgccgcy  
 901 gaaaaacatc accctcattc ggggtctggg ccatggagcc tttggggagg tgtatgaagg  
 961 ccaggtgtcc ggaatgcccc acgacccaag cccctgcaa gtggctgtga agacgctgcc  
 1021 tgaagtgtgc tctgaacagg acgaactgga tttcctcatg gaagccctga tcatcagcaa  
 1081 attcaaccac cagaacattg ttcgctgcat tggggtgagc ctgcaatccc tgccccggtt  
 1141 catcctgctg gagctcatgg cggggggaga cctcaagtcc ttctccgag agaccgccc  
 1201 tcgcccagac cagccctcct ccctggccat gctggacctt ctgcacgtgg ctcgggacat  
 1261 tgctgtggc tgtcagtatt tggaggaaaa ccacttcac caccgagaca ttgctgccag  
 1321 aaactgcctc ttgacctgtc caggccctgg aagagtggcc aagattggag acttcgggat  
 1381 ggcccagac atctacaggg cgagctacta tagaaaggga ggctgtgcca tgctgccagt  
 1441 taagtggatg cccccagagg ctttcatgga aggaatattc acttctaaaa cagacacatg  
 1501 gtcttttggg gtgctgctat gggaaatctt ttctcttggg tatatgccat accccagcaa  
 1561 aagcaaccag gaagttctgg agtttgtcac cagtggaggc cggatggacc cacccaagaa  
 1621 ctgccctggg cctgtatacc ggataatgac tcagtgtctg caacatcagc ctgaagacag  
 1681 gcccaacttt gccatcattt tggagaggat tgaatactgc accaggacc cggatgtaat  
 1741 caacaccgct ttgccgatag aatatggctc acttgaggaa gaggaagaga aagtgcctgt  
 1801 gaggcccaag gaccctgagg gggttcctcc tctcctggtc tctcaacagg caaaacggga  
 1861 ggaggagcgc agcccagctg cccaccacc tctgcctacc acctcctctg gcaaggctgc  
 1921 aaagaaacc acagctgcag aggtctctgt tcgagtcct agagggcccg ccgtggaagg  
 1981 gggacacgtg aatatggcat tctctcagtc caacctcct tcggagttgc acagggcca  
 2041 cggatccaga aacaagccca ccagcttgctg gaacccaacg tacggctcct ggtttacaga  
 2101 gaaaccacc aaaaagaata atcctatagc aaagaaggag ccacacgaga ggggtaacct  
 2161 ggggctggag ggaagctgta ctgtcccacc taacgttgca actgggagac ttccgggggc  
 2221 ctactgctc cttagaccct cttegtgac tgccaatatg aaggaggtag ctctgttcag  
 2281 gctacgtcac ttcccttggt ggaatgtcaa ttacggctac cagcaacagg gcttgcctt  
 2341 agaagccgct actgcccctg gagctggtea ttacaggat accattctga aaagcaagaa  
 2401 tagcatgaac cagcctgggc cctgagctcg gtcgcacact cacttctctt ccttgggatc  
 2461 ctaagaccg tgg

Secuencia de la proteína EML4-ALK Variante 3a (BAG55003.1; GI: 194072593)

5

1 mdgfagsldd ssaastsdv qdrlesalsr vqqgedeitv lkaaladvlr rlaisedhva  
 61 svkksvsskg qpspravipm scitngsgan rkpshsavs iagketlssa aksgtekkke  
 121 kpqgqrekke eshsndqspq iraspspqs sqplqihrt pesknatptk sikrpspaek  
 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii nqvyrrkhqe lqamqmelqs  
 241 peyklsklrr stimtdynpn ycfagktssi sdlkevprkn itlirglghg afgevyegqv  
 301 sgmpndpspl qvavktlpev cseqdeldfl mealiiskfn hqnivrcigv slqslprfil  
 361 lelmagglk sflretrprp sqpsslamld llhvardiac gcqyleenhf ihrdiaarnc  
 421 llctcpgrv akigdfgmar diyasyyrk ggcamlpvkw mppeafmegi ftsktdtwsf  
 481 gvllweifsl gympypsksn qevlefvtsg grmdppkncp gpvyrimtqc wqhqpdrpn  
 541 failerley ctqdpdvint alpieygpvlv eeeekvpvrp kdpegvppll vsqqakreee  
 601 rspaappplp ttssgkaakk ptaaevsrv prgpaveggh vnmafsqsnp pselhrvhgs  
 661 rnkptslwnp tygswftekp tkknnpiakk ephergnlgl egscvppnv atgrlpgasl  
 721 llepssltan mkevplfrlr hfpcgnvnyg yqqqglplea atapgaghye dtilksknsm

781 nqppp

Secuencia de ácido nucleico de EML4-ALK Variante 3b (AB374362.1; GI: 194072594)

1 actctgtcgg tccgctgaat gaagtgcccg ccctctaaag cccggagccc ggcgctttcc  
61 ccgcaagatg gacggtttcg ccggcagtct cgatgatagt atttctgctg caagtacttc  
121 tgatgttcaa gatcgctgt cagctcttga gtcacgagtt cagcaacaag aagatgaaat  
181 cactgtgcta aaggcggctt tggctgatgt tttgaggcgt cttgcaatct ctgaagatca  
241 tgtggcctca gtgaaaaaat cagtctcaag taaaggccaa ccaagccctc gagcagttat  
301 tcccatgtcc tgtataacca atggaagtgg tgcaaacaga aaaccaagtc ataccagtgc  
361 tgtctcaatt gcaggaaaag aaactctttc atctgctgct aaaagtggta cagaaaaaaa  
421 gaaagaaaaa ccacaaggac agagagaaaa aaaagaggaa tctcattcta atgatcaaa  
481 tccacaaatt cgagcatcac cttctcccca gccctcttca caacctctcc aaatacacag  
541 acaaactcca gaaagcaaga atgctactcc caccaaaagc ataaaacgac catcaccagc  
601 tgaaaagtca cataattctt gggaaaattc agatgatagc cgtataaat tgtcgaanaat  
661 accttcaaca cccaaattaa taccaaaagt taccaaaact gcagacaagc ataaagatgt  
721 catcatcaac caagcaaaaa tgtcaactcg cgaaaaaac agccaagtgt accgcccggaa  
781 gcaccaggag ctgcaagcca tgcagatgga gctgcagagc cctgagtaca agctgagcaa  
841 gctccgcacc tcgaccatca tgaccgacta caaccccaac tactgctttg ctggcaagac  
901 ctctccatc agtgacctga aggaggtgcc gcggaaaaac atcacctca ttcgggggtct  
961 gggccatgga gcctttgggg aggtgtatga aggccaggtg tccggaatgc ccaacgacc  
1021 aagccccctg caagtggctg tgaagacgct gcctgaagtg tgctctgaac aggacgaact  
1081 ggatttctct atggaagccc tgatcatcag caaattcaac caccagaaca ttgttcgctg  
1141 cattgggggtg agcctgcaat ccctgccccg gttcatcctg ctggagctca tggcgggggg  
1201 agacctcaag tccttctctc gagagaccgc ccctcgcccc agccagccct cctccctggc  
1261 catgctggac cttctgcacg tggctcggga cattgcctgt ggctgtcagt atttgaggga  
1321 aaaccacttc atccaccgag acattgctgc cagaaactgc ctcttgacct gtccaggccc  
1381 tggaagagtg gccaaagatt gagacttcgg gatggcccga gacatctaca gggcgagcta  
1441 ctatagaaag ggaggtctg ccatgctgcc agttaaagtg atgccccag aggccttcat  
1501 ggaaggaata ttcacttcta aaacagacac atggtccttt ggagtgtgc tatgggaaat  
1561 cttttctctt ggatatatgc cataccccag caaaagcaac caggaagttc tggagtgtgt  
1621 caccagtgga ggccggatgg acccacccaa gaactgcctt gggcctgtat accggataat  
1681 gactcagtc tggcaacatc agcctgaaga caggccaac tttgccatca ttttgagag  
1741 tactgaatac tgcaccagc acccggatgt aatcaacacc gctttgccga tagaatatgg  
1801 tccacttgtg gaagaggaag agaaagtgcc tgtgaggccc aaggacctg aggggttcc  
1861 tcctctcctg gtctctcaac aggcaaaacg ggaggaggag cgcagcccag ctgccccacc  
1921 acctctgctt accacctctt ctggcaaggc tgcaaaagaa cccacagctg cagaggtctc  
1981 tgttcgagtc cctagagggc cggcctgga agggggacac gtgaatatgg cattctctca  
2041 gtccaacctt ccttcggagt tgcacagggt ccacggatcc agaaacaagc ccaccagctt  
2101 gtggaacca acgtacggct cctggtttac agagaaacc accaaaaaga ataactctat  
2161 agcaaagaag gagccacacg agaggggtaa cctggggctg gaggaagct gtactgtccc  
2221 acctaacgtt gcaactggga gacttcggg ggctcactg ctctagagc cctcttcgct  
2281 gactgccaat atgaaggag tacctctgtt caggctacgt cacttccctt gtgggaatgt  
2341 caattacggc taccagcaac agggcttgcc cttagaagcc gctactgcc ctggagctgg  
2401 tcattacgag gataccattc tgaaaagcaa gaatagcatg aaccagcctg ggcctgagc  
2461 tcggtcgcac actcacttct cttccttggg atccctaaga ccgtgg

5 Secuencia de la proteína EML4-ALK Variante 3b (BAG55004.1; GI: 194072595)

1 mdgfagsldd sīsaastsdv qdrīsalesr vqqqedeitv lkaaladvlr rlaisedhva  
61 svkksvsskg qpspravipm scitngsgan rkpshtsav s iagketlssa aksgtekkke  
121 kpqqqrekke eshsndqspq iraspspqs sqplqihrt pesknaptk sikrpspaek  
181 shnswensdd srnklskips tpklipkvtk tadkhkdvii nqakmstrek nsqvyrkhq  
241 elqamqmelq speyklsklr tstimtdynp nycfagktss isdlkevprk nitlirglgh  
301 gafgevyegq vsqmpndpsp lqvavktlpe vcseqdeldf lmealiiskf nhqnivrcig  
361 vslqslprfi llelmaggdl ksflretrpr psqpsslaml dllhvardia cgcqyleenh  
421 fihrdiaarn clltcppggr vakigdfgma rdiyasyyr kggcamlpvk wmppeafmeg  
481 iftsktdtws fgvllweifs lgympysks nqevlefvt s ggrmdppknc pppvyrimtq  
541 cwqhqpdrp nfaiilerie yctqdpdvin talpieygpl veeekvpvr pkdpegvppl  
601 lvsqqakree erspaapppl pttssgkaak kptaaevsvr vprgpavegg hvnmafsqsn

## ES 2 660 149 T3

661 ppselhrvhg srnkptslwn ptygswftek ptkknnpiak kephergnlg legscvppn  
721 vatgrlpgas lllpsslta nmkevplfrl rhfpcgnvny gyqqqgple aatapgaghy  
781 edtilkskns mnqpgp

Secuencia de ácido nucleico de EML4-ALK Variante 4 (AB374363.1; GI: 209837703)

ES 2 660 149 T3

1 actctgtcgg tccgctgaat gaagtgcccg cccctctaag cccggagccc ggcgctttcc  
 61 ccgcaagatg gacggtttcg ccggcagtct cgatgatagt atttctgctg caagtacttc  
 121 tgatgttcaa gatcgctgt cagctcttga gtcacgagtt cagcaacaag aagatgaaat  
 181 cactgtgcta aaggcggctt tggctgatgt tttgagcgt cttgcaatct ctgaagatca  
 241 tgtggcctca gtgaaaaaat cagtctcaag taaaggccaa ccaagccctc gagcagttat  
 301 tcccatgtcc tgtataacca atggaagtgg tgcaaacaga aaaccaagtc ataccagtgc  
 361 tgtctcaatt gcaggaaaag aaactctttc atctgctgct aaaagtggta cagaaaaaaa  
 421 gaaagaaaaa ccacaaggac agagagaaaa aaaagaggaa tctcattcta atgatcaaag  
 481 tccacaaatt cgagcatcac cttctcccca gccctcttca caacctctcc aaatacacag  
 541 acaaactcca gaaagcaaga atgctactcc caccaaaagc ataaaacgac catcaccagc  
 601 tgaaaagtca cataattctt gggaaaattc agatgatagc cgtaataaat tgtcgaaaat  
 661 accttcaaca cccaaattaa taccaaaagt taccaaaact gcagacaagc ataaagatgt  
 721 catcatcaac caagaaggag aatatattaa aatgtttatg cgcggtcggc caattaccat  
 781 gttcattcct tccgatgttg acaactatga tgacatcaga acggaactgc ctctgagaa  
 841 gctcaaactg gagtgggcat atggttatcg aggaaaggac tgtagagcta atgtttacct  
 901 tcttccgacc ggggaaatag tttatttcat tgcacagta gtagtactat ttaattatga  
 961 ggagagaact cagcgacact acctgggcca tacagactgt gtgaaatgcc ttgctataca  
 1021 tcctgacaaa attaggattg caactggaca gatagctggc gtggataaag atggaaggcc  
 1081 tctacaaccc cacgtcagag tgtgggattc tgttactcta tccacactgc agattattgg  
 1141 acttggcact tttgagcgtg gagtaggatg cctggatttt tcaaaagcag attcagggtgt  
 1201 tcattttatgt gttattgatg actccaatga gcatatgctt actgtatggg actggcagag  
 1261 gaaagcaaaa ggagcagaaa taaagacaac aatgaagtt gttttggctg tggagtttca  
 1321 cccaacagat gcaaatacca taattacatg cggtaaactc catattttct tctggacctg  
 1381 gagcggcaat tctaatacaa gaaaacaggg aatttttggg aaatatgaaa agccaaaatt  
 1441 tgtgcagtgt ttagcattct tggggaatgg agatgttctt actggagact caggtggagt  
 1501 catgcttata tggagcaaaa ctactgtaga gcccacacct gggaaaggac ctaaagggtgt  
 1561 atatcaaatc agcaacaaa tcaagctca tgatggcagt gtgttcacac tttgtcagat  
 1621 gagaaatggg atgttattaa ctggaggagg gaaagacaga aaaataattc tgtgggatca  
 1681 tgatctgaat cctgaaagag aaatagagat atgctggatg agccctgagt acaagctgag  
 1741 caagctccgc acctcgacca tcatgaccga ctacaacccc aactactgct ttgctggcaa  
 1801 gacctcctcc atcagtgacc tgaaggagggt gccgcggaaa aacatcaccc tcattcgggg  
 1861 tctgggcat ggagcctttg gggagggtga tgaaggccag gtgtccggaa tgcccaacga  
 1921 cccaagcccc ctgcaagtgg ctgtgaagac gctgcctgaa gtgtgctctg aacaggacga  
 1981 actggatttc ctcatggaag ccctgatcat cagcaaattc aaccaccaga acattgtctg  
 2041 ctgcatgtgg gtgagcctgc aatccctgcc ccggttcac cctgtggagc tcatggcggg  
 2101 gggagacctc aagtccttc tccgagagac ccgacctcgc ccgagccagc cctcctccct  
 2161 ggccatgctg gaccttctgc acgtggctcg ggacattgcc tgtggctgtc agtatttggg  
 2221 ggaaaaccac ttcattccacc gagacattgc tgccagaaac tgccctctga cctgtccagg  
 2281 ccctggaaga gtggccaaga ttggagactt cgggatggcc cgagacatct acagggcgag  
 2341 ctactataga aaggagggt gtgccatgct gccagttaag tggatgcccc cagaggcctt  
 2401 catggaagga atattcactt ctaaaacaga cacatggtcc tttggagtgc tgctatggga  
 2461 aatcttttct cttggatata tgccataacc cagcaaaagc aaccaagaag ttctggagtt  
 2521 tgtcaccagt ggaggccgga tggaccacc caagaactgc cctgggcctg tataccggat  
 2581 aatgactcag tgctggcaac atcagcctga agacaggccc aactttgcca tcattttgga  
 2641 gaggattgaa tactgcaccc aggaccggga tgtaatcaac accgctttgc cgatagaata  
 2701 tgggccactt gtggaagagg aagagaaagt gcctgtgagg cccaaggacc ctgagggggt  
 2761 tcctcctctc ctggtctctc aacaggcaaa acgggaggag gagcgcagcc cagctgcccc  
 2821 accacctctg cctaccacct cctctggcaa ggctgcaaag aaaccacag ctgcagaggt  
 2881 ctctgttcga gtccctagag ggccggccgt ggaaggggga cacgtgaata tggcattctc  
 2941 tcagtccaac cctccttcgg agttgcacag ggtccacgga tccagaaaca agcccaccag  
 3001 cttgtggaac ccaacgtacg gctcctggtt tacagagaaa cccacaaaa agaataatcc  
 3061 tatagcaaag aaggagccac acgagagggg taacctgggg ctggagggaa gctgtactgt  
 3121 cccacctaac gttgcaactg ggagacttcc gggggcctca ctgctcctag agccctctc  
 3181 cctgactgcc aatatgaagg aggtacctct gttcaggcta cgtcacttcc cttgtgggaa  
 3241 tgtcaattac ggctaccagc aacagggtt gcccttagaa gccgctactg ccctggagc  
 3301 tggtcattac gaggatacca ttctgaaaag caagaatagc atgaaccagc ctgggacctg  
 3361 agctcggctg cacactcact tctcttctt gggatcccta agaccgtgg

Secuencia de la proteína EML4-ALK Variante 4 (BAG75147.1; GI: 209837704)

1 mdgfagslidd sisaastsdv qdrlsalesr vqqqedeitv lkaaladvlr rlaiedhva  
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke  
 121 kpqqgqrekke eshsndqspq iraspspps sqplqihrt pesknaptk sikrpspaek  
 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii ngegeyikmf mrgrpitmf  
 241 psdvdyddi rtelppekl lewaygyrgk dcranvylp tgeivyfias vvvlfnyeer  
 301 tqrhylghtd cvkclaihp kiriatgqia gvdkdgrplq phvrwdsvt lstlqiiglg  
 361 tfergvgcld fskadsgvhl cviddsnehm ltvwdwqrka kgaeiktne vvlavefhpt  
 421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq claflngdv ltgdsggvml  
 481 iwskttevept pgkpkgyvq iskqikahdg svftlcqmrn gmltgggkd rkiilwdhdl  
 541 npereieicw mspeyklsl rtstimtdyn pnycfagkts sisdlkevpr knitlirglg  
 601 hgafgeyevg qvsgmpndps plqvavktrp evcseqdeld flmealiisk fnhqnivrci  
 661 gvslqslprf illelmaggd lksflretrp rpsqpslam ldllhvardi acgcqyleen  
 721 hfihrdiaar nclltcpgpg rvakigdfgm ardiyrasyy rkggcamlpv kwmppeafme  
 781 giftsktdtw sfgvllweif slgympysk snqevlefv sggrmdppkn cpgpvyrimt  
 841 qcwqhpedr pnfaiileri eyctqdpdvi ntalpieygp lveeeekvpv rpkdpegvpp  
 901 llvsqqakre eerspaapp lpttssgkaa kkptaaevsv rvprgpaveg ghvnmfqsqs  
 961 nppselhrvh gsrnkptslw nptygswfte kptkknnpia kkepthergnl glegscvpp  
 1021 nvatgrlpga slleppslt anmkevplfr lrhfpognvn ygyqqqglpl eaatapgagh  
 1081 yedtilskn smnqpgp

Secuencia de ácido nucleico de EML4-ALK Variante 5a (AB374364.1; GI: 209837705)

5

1 actctgtcgg tccgctgaat gaagtgcccg ccctctaaag cccggagccc ggcgctttcc  
 61 ccgcaagatg gacggtttcg ccggcagctc cgatgatagt atttctgctg caagtacttc  
 121 tgatgttcaa gatcgctgt cagctcttga gtcacgagtt cagcaacaag aagatgaaat  
 181 cactgtgcta aaggcggctt tggctgatgt tttgagcgt cttgcaatct ctgaagatca  
 241 tgtggcctca gtgaaaaaat cagtctcaag taaagtgtac cgccggaagc accaggagct  
 301 gcaagccatg cagatggagc tgcagagccc tgagtacaag ctgagcaagc tccgcacctc  
 361 gaccatcatg accgactaca accccaacta ctgctttgct ggcaagacct cctccatcag  
 421 tgacctgaag gaggtgccgc ggaaaaacat caccctcatt cggggctctg gccatggagc  
 481 ctttggggag gtgtatgaag gccaggtgtc cggaatgcc aacgaccaa gccccctgca  
 541 agtggctgtg aagacgctgc ctgaagtgtg ctctgaacag gacgaactgg atttctcat  
 601 ggaagccctg atcatcagca aattcaacca ccagaacatt gttcgtgca ttgggggtgag  
 661 cctgcaatcc ctgccccggt tcatcctgct ggagctcatg gcggggggag acctcaagtc  
 721 cttcctccga gagaccgccc ctgccecgag ccagccctcc tccctggcca tgctggacct  
 781 tctgcacgtg gctcgggaca ttgcctgtgg ctgtcagtat ttggaggaaa accacttcat  
 841 ccaccgagac attgctgcca gaaactgct cttgacctgt ccaggccctg gaagagtggc  
 901 caagattgga gacttcggga tggcccgaga catctacagg gcgagctact atagaaaggg  
 961 aggctgtgcc atgctgccag ttaagtggat gccccagag gccttcatgg aaggaatatt  
 1021 cacttctaaa acagacacat ggtcctttgg agtgtgcta tgggaaatct tttctcttgg  
 1081 atatatgcca taccagca aaagcaacca ggaagttctg gagtttctca ccagtggagg  
 1141 ccgatggac ccaccaaga actgcccctg gctgtatac cggataatga ctcagtgtg  
 1201 gcaacatcag cctgaagaca ggcccactt tgccatcatt ttggagagga tgaatactg  
 1261 caccagac ccgatgtaa tcaacaccgc tttgccgata gaatatggtc cactgtgga  
 1321 agaggaagag aaagtgcctg tgaggccaa ggaccctgag ggggttctc ctctcctggt  
 1381 ctctcaacag gcaaacggg aggaggagc cagcccagct gccccaccac ctctgcctac  
 1441 cacctcctc ggcaaggctg caaagaaacc cacagctgca gaggtctctg ttcgagtccc  
 1501 tagagggccg gccgtggaag ggggacacgt gaatatggca ttctctcagt ccaaccctcc  
 1561 ttcggagtgt cacagggctc acggatccag aaacaagccc accagcttgt ggaaccaac  
 1621 gtacggctcc tggtttacag agaaaccac caaaaagaat aatcctatag caaagaagga  
 1681 gccacacgag aggggtaacc tggggctgga ggggaagctgt actgtcccac ctaacgttg  
 1741 aactgggaga cttccggggg cctcactgct cctagagccc tcttgcctga ctgccaatat  
 1801 gaaggaggta cctctgttca ggctacgtea ctcccttgt gggaaatgca attacggta  
 1861 ccagcaacag ggcttgcct tagaagccgc tactgcccct ggagctggtc attacgagga  
 1921 taccattctg aaaagcaaga atagcatgaa ccagcctggg ccctgagctc ggtcgcacac  
 1981 tcaattctct tccttgggat ccctaagacc gtgg

Secuencia de la proteína EML4-ALK Variante 5a (BAG75148.1; GI: 209837706)

10

ES 2 660 149 T3

1 mdgfagsl<sup>dd</sup> s<sup>isa</sup>aast<sup>sdv</sup> qdr<sup>l</sup>salesr vqq<sup>q</sup>edeitv lkaaladv<sup>lr</sup> rlaisedhva  
 61 svkksvsskv yrrkhqelqa mqmelqspey klsklrtsti mtdynpnycf agkts<sup>sis</sup>dl  
 121 kevprknitl irglghgafg evyegqvsgm pndpsplqva vktlpevcse qdeldflmea  
 181 liiskfnhqn ivrcigvslq slprfillel maggdksfl retrprpsqp sslamldlh  
 241 vardiacgcq yleenhfihr diaarncllt cpgpgrvaki gdfgmardiy rasyyrkggc  
 301 amlpvkwmp eafmegifts ktdtwsfgvl lweifslgym pypsksnqev lefvtsggm  
 361 dppkncpgpv yr<sup>im</sup>tqcwqh qp<sup>edr</sup>pnfai iler<sup>iey</sup>ctq dpdvintalp ieygplveee  
 421 ekvpvrpkdp egvppllvsq qakreeersp aappplptts sgkaakkpta aevsvrvprg  
 481 pavegghvnm afsqsnppse lhrvhgsrnk ptslwnptyg swftekptkk nnp<sup>ia</sup>kkeph  
 541 ergnlglegs ctvppnvatg rlp<sup>gas</sup>llle pssltanmke vplf<sup>rl</sup>rhfp cgnvnygyqq  
 601 qglpleaata pgaghyedti lksk<sup>nm</sup>qp gp

Secuencia de la proteína EML4-ALK Variante 5b (AB374365.1; GI: 209837707)

1 actctgtcgg tccgctgaat gaagtgcccg ccctctaag cccggagccc ggcgctttcc  
 61 ccgcaagatg gacggtttcg ccggcagctc cgatgatagt atttctgctg caagtacttc  
 121 tgatgttcaa gatcgctgt cagctcttga gtcacgagtt cagcaacaag aagatgaaat  
 181 cactgtgcta aaggcggctt tggctgatgt tttgaggcgt cttgcaatct ctgaagatca  
 241 tgtggcctca gtgaaaaaat cagtctcaag taaaggttca gagctcaggg gaggatatgg  
 301 agatccaggg aggcttctg taggaagtgg cctgtgtagt gcttcaaggg ccaggctgcc  
 361 aggccatgtt gcagctgacc acccacctgc agtgtaccgc cggaagcacc aggagctgca  
 421 agccatgcag atggagctgc agagccctga gtacaagctg agcaagctcc gcacctcgac  
 481 catcatgacc gactacaacc ccaactactg ctttctggc aagacctcct cc<sup>at</sup>cagtga  
 541 cctgaaggag gtgccgcgga aaaacatcac cctcattcgg ggtctgggccc atggagcctt  
 601 tggggagggtg tatgaaggcc aggtgtccgg aatgcccac gacccaagcc ccctgcaagt  
 661 ggctgtgaag acgctgcctg aagtgtgctc tgaacaggac gaactggatt tctcatgga  
 721 agccctgatc atcagcaaat tcaaccacca gaacattggt cgctgcattg gggtagcct  
 781 gcaatccctg ccccggttca tctgtctgga gctcatggcg gggggagacc tcaagtcctt  
 841 cctccgagag acccgccctc gcccgagcca gccctcctcc ctggccatgc tggaccttct  
 901 gcagtggtc cgggacattg cctgtggctg tcagtat<sup>ttg</sup> gaggaaaacc acttcatcca  
 961 ccgagacatt gctgccagaa actgctctt gacctgtcca ggccctggaa gagtggccaa  
 1021 gattggagac ttcgggatgg cccgagacat ctacagggcg agctactata gaaagggagg  
 1081 ctgtgccatg ctgccagtta agtggatgcc cccagaggcc ttcattggaag gaatattcac  
 1141 ttctaaaaca gacacatggt cctttggagt gctgctatgg gaaatctttt ctcttgata  
 1201 tatgccatac cccagcaaaa gcaaccagga agtctggag tttgtacca g<sup>tg</sup>gaggccg  
 1261 gatggacca ccaagaact gccctggcc tgtataccg ataatacact agtctggca  
 1321 acatcagcct gaagacaggc ccaactttgc catcattttg gagaggattg aactactgcac  
 1381 ccaggaccgg gatgtaatca acaccgcttt gccgatagaa tatggtccac ttgtggaaga  
 1441 ggaagagaaa gtgctgtga ggcccaagga ccctgagggg gttcctcctc tctggtctc  
 1501 tcaacaggca aaacgggagg aggagcgcag cccagctgcc ccaccacctc tgctaccac  
 1561 ctctctggc aaggctgcaa agaaaccac agctgcagag gtctctgttc gagtccctag  
 1621 agggccggcc gtggaagggg gacacgtgaa tatggcattc tctcagtcca accctcctc  
 1681 ggagttgcac agggctccag gatccagaaa caagcccacc agcttgtgga acccaacgta  
 1741 cggctcctgg tttacagaga aaccaccaa aaagaataat cctatagcaa agaaggagcc  
 1801 acacgagagg ggtaacctgg ggctggaggg aagctgtact gtcccaccta acgttgcaac  
 1861 tgggagactt cgggggcct cactgctcct agagccctct tcgctgactg ccaatagaa  
 1921 ggaggtacct ctgttcaggc tacgtcactt cccttgtggg aatgtcaatt acggctacca  
 1981 gcaacagggc ttgcccttag aagccgctac tgcccctgga gctggtcatt acgaggatac  
 2041 cattctgaaa agcaagaata gcatgaacca gcctgggccc tgagctcggg cgcacactca  
 2101 cttctcttcc ttgggatccc taagaccgtg g

5 Secuencia de la proteína EML4-ALK Variante 5b (BAG75149.1; GI: 209837708)

1 mdgfagsl<sup>dd</sup> s<sup>isa</sup>aast<sup>sdv</sup> qdr<sup>l</sup>salesr vqq<sup>q</sup>edeitv lkaaladv<sup>lr</sup> rlaisedhva  
 61 svkksvsskg selrggygdp grlpvsglc sasrarlpgh vaadhppavy rrk<sup>h</sup>qelqam

## ES 2 660 149 T3

121 qmelqspeyk lsklrtstim tdynpnycfa gktssisdik evprkntli rglghgafge  
181 vyegqvsgmp ndpsplqvav ktlpevcseq deldflmeal iiskfnhqn vrcigvslqs  
241 lprfillelm aggdksflr etrprpsqps slamdlhhv ardiacgcqy leenhfihrd  
301 iaarncllhc ppggrvakig dfgmardiy asyyrkggca mlpvkwmpppe afmegiftsk  
361 tdtwsfgvll weifslgyp ypsksnqevl efvtsqgrmd ppknpcgpvy rimtqcwqhq  
421 pedrpnfaii lerieyctqd pdvintalpi eygplveeee kvpvrpkdpe gvppllvsqq  
481 akreeerspa appplpttss gkaakkptaa evsvrvprgp avegghvma fsqsnppsel  
541 hrvhgsrnkp tslwnptygs wftekptkkn npiakkephe rgnlglegsc tvppnvatgr  
601 lpgasllep ssltanmkev plfrrrhfpc gnvnygyqqq glpleaatap gaghyedtil  
661 ksknsmnqpg p

Secuencia de ácido nucleico de EML4-ALK Variante 6 (AB462411.1; GI: 227452648)

ES 2 660 149 T3

1 tactctgtcg gtccgctgaa tgaagtgcc gccctctaa gcccgagcc cggcgcttc  
 61 cccgcaagat ggacggtttc gcggcgagtc tcgatgatag tatttctgct gcaagtactt  
 121 ctgatgttca agatcgcttg tcagctcttg agtcacgagt tcagcaacaa gaagatgaaa  
 181 tcaactgtgct aaagggcggt ttggctgatg ttttgaggcg tcttgcaatc tctgaagatc  
 241 atgtggcctc agtgaaaaaa tcagctctca gtaaaggcca accaagcctc cgagcagtta  
 301 ttcccatgtc ctgtataacc aatggaagtg gtgcaaacag aaaaccaagt cataccagtg  
 361 ctgtctcaat tgcaggaaaa gaaactcttt catctgctgc taaaagtggc acagaaaaaa  
 421 agaaagaaaa accacaagga cagagagaaa aaaaagagga atctcattct aatgatcaaa  
 481 gtccacaaat tgcagcatca ctttctcccc agccctcttc acaacctctc caaatacaca  
 541 gacaaactcc agaaagcaag aatgctactc ccacaaaagc cataaaacga ccatcaccag  
 601 ctgaaaagtc acataattct tgggaaaatt cagatgatag ccgtaataaa ttgtcgaaaa  
 661 taccttcaac acccaaatta ataccaaaag ttacaaaacg tcagacaagc cataaagatg  
 721 tcatcatcaa ccaagaagga gaatatatta aatgtttat gcgcggtcgg ccaattacca  
 781 tgttcattcc ttccgatggt gacaactatg atgacatcag aacggaactg cctcctgaga  
 841 agtcaaactc ggagtgggca tatggttacc gaggaaagga ctgtagagct aatgtttacc  
 901 ttcttccgac cggggaata gtttatttca ttgcatcagt agtagtacta ttaattatg  
 961 aggagagaac tcagcgacac tacctgggccc atacagactg tgtgaaatgc ctgtctatac  
 1021 atcctgacaa aattaggatt gcaactggac agatagctgg cgtggataaa gatggaaggc  
 1081 ctctacaacc ccacgtcaga gtgtgggatt ctgttactct atccacactg cagattattg  
 1141 gacttggcac ttttgagcgt ggagtaggat gcctggattt ttcaaaagca gattcagggtg  
 1201 ttcatthtat tgttattgat gactccaatg agcatatgct tactgtatgg gactggcaga  
 1261 ggaaagcaaa aggagcagaa ataaagacaa caaatgaagt tgttttggct gtggagtttc  
 1321 acccaacaga tgcaaatacc ataattacat gcggtaaatc tcataatttc ttctggacct  
 1381 ggagcggcaa ttcactaaca agaaaacagg gaatttttgg gaaatatgaa aagccaaat  
 1441 ttgtgcagtg ttttagcattc ttggggaatg gagatgttct tactggagac tcaggtggag  
 1501 tcatgcttat atggagcaaa actactgtag agcccacacc tgggaaagga ctaaaaggaa  
 1561 gtggcctgtg tagtgcttca agggccaggc tgccaggcca tgttgcagct gaccaccac  
 1621 ctgcagtgta ccgccggaag caccaggagc tgcaagccat gcagatggag ctgcagagcc  
 1681 ctgagtacaa gctgagcaag ctccgcacct cgaccatcat gaccgactac aaccccaact  
 1741 actgctttgc tggcaagacc tccctccatca gtgacctgaa ggaggtgccg cggaaaaaca  
 1801 tcaccctcat tcggggtctg ggccatggag cctttgggga ggtgtatgaa ggccagggtg  
 1861 ccggaatgcc caacgacca agccccctgc aagtggctgt gaagacgctg cctgaagtgt  
 1921 gctctgaaca ggacgaactg gatttctca tggaaagcct gatcatcagc aaattcaacc  
 1981 accagaactc tgttcgctgc attggggtga gctgcaatc cctgccccgg ttcatcctgc  
 2041 tggagctcat ggccggggga gacctcaagt ccttctccg agagaccgcg cctgcctcga  
 2101 gccagcctc ctccctggcc atgctggacc ttctgcacgt ggctcgggac attgctgtg  
 2161 gctgtcagta tttggaggaa aaccacttca tccaccgaga cattgctgcc agaaactgcc  
 2221 tcttgacctg tccaggcctt ggaagagtgg ccaagattgg agacttcggg atggcccag  
 2281 acatctacag ggcgagctac tatagaaagg gaggtgtgct catgctgcca gttaagtgga  
 2341 tgcccccaga ggccttcatg gaaggaatat tcaacttctaa aacagacaca tggctctttg  
 2401 gactgtctgct atgggaaatc ttttctcttg gatatatgcc ataccaccagc aaaagcaacc  
 2461 aggaagtctt ggagtttgtc accagtggag gccggatgga cccaccacag aactgcctg  
 2521 ggctgtata ccggataatg actcagtgtc ggcaacatca gcctgaagac aggcccaact  
 2581 ttgccatcat tttggagagg attgaatact gcaccacagga ccggatgta atcaacaccg  
 2641 ctttgccgat agaatatggt ccacttgtgg aagaggaaga gaaagtgcct gtgaggccca  
 2701 aggaccctga gggggttctt cctctctctg tctctcaaca ggcaaaacg gaggaggagc  
 2761 gcagcccagc tgccccacca cctctgccta ccacctctc tggcaaggct gcaagaaac  
 2821 ccacagctgc agaggtctct gttcgagtc cttagaggcc ggccgtggaa gggggacagc  
 2881 tgaatatggc attctctcag tccaacctc cttcggagt gcacagggtc cacggatcca  
 2941 gaaacaagcc caccagcttg tggaaaccaa cgtacggctc ctggtttaca gagaaacca  
 3001 caaaaagaa taatcctata gcaaaagagg agccacacga gaggggtaac ctggggctgg  
 3061 aggaagctg tactgtccca ctaaacgttg caactgggag acttccgggg gcctcactgc  
 3121 tcttagagcc ctcttctctg actgccaata tgaaggagt acctctgttc aggtctactc  
 3181 acttcccttg tgggaatgtc aattacggct accagcaaca gggcttgccc ttagaagccg  
 3241 ctactgccc tggagctggt cattacgagg ataccattct gaaaagcaag aatagcatga  
 3301 accagcctgg gccctgagct cggctgcaca ctcaacttct ttcttgggga tccctaagac  
 3361 cgtgg

Secuencia de la proteína EML4-ALK Variante 6 (BAH57335.1; GI: 227452649)

1 mdgfagslidd sisaastsdv qdrilsalesr vqqqedeitv lkaaladvlr rlaishedva  
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke  
 121 kpqgqrekke eshsndqspq iraspspqs sqplqihrt pesknaptk sikrpspaek  
 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii ngegeyikmf mrgrpitmfi  
 241 psdvndyddi rtelpeklk lewaygyrgk dcranvylp tgeivyfias vvlfnyeer  
 301 tqrhylghtd cvkclaihp kiriatgqia gvdkdgrplq phvrwdsvt lstlqiiglg  
 361 tfergvgcld fskadsgvhl cviddsnehm ltvwdwqrka kgaeiktne vvlavefhpt  
 421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq clafngndv ltgdsggvml  
 481 iwsktttvept pgkpkgsqsl csasarlpq hvaadhppav yrrkhqelqa mqmelqspey  
 541 klsklrtsti mtdynpnycf agkttssidl kevprknitl irglghgafg evyegqvsgm  
 601 pndpsplqva vktlpevcse qdeldflmea liiskfnhqn ivrcigvslq slprfillel  
 661 maggdksfl retrprpsqp sslamlldlh vardiacgcq yleenhfih diaarncllt  
 721 cpqpgrvaki gdfgmardiy rasyyrkqgc amlpvkwmp eafmegifts ktdtwsfgvl  
 781 lweifslgym pypsksnqev lefvtsqgrm dppkncpgpv yrimtqcwqh qpedrpnfai  
 841 ilerieyctq dpdvintalp ieygplveee ekvprpkdp egvppllvsq qakreeersp  
 901 aappplppts sgkaakkpta aevsvrvprg pavegghvnm afsqsnppse lhrvhgsrnk  
 961 ptslwnptyg swftekptkk nnpiakkeph ergnlglegs ctvppnvatg rlpgasllle  
 1021 pssltanmke vplfrlrhfp cgnvnvygyqq qglpleaata pgaghayedti lksknsnmqg  
 1081 gp

Secuencia de ácido nucleico de EML4-ALK variante 7 (AB462412.1; GI: 227452650)

5

1 tactctgtcg gtccgctgaa tgaagtgcc gccctctaa gcccgagacc cggcgctttc  
 61 cccgcaagat ggacggtttc gccggcagtc tcgatgatag tatttctgct gcaagtactt  
 121 ctgatgttca agatcgcttg tcagctcttg agtcacgagt tcagcaaca gaagatgaaa  
 181 tcaactgtcg aaagggcgt ttggctgatg ttttgaggcg tcttgcaatc tctgaagatc  
 241 atgtggcctc agtgaaaaaa tcagttcaa gtaaaggcca accaaggcct cgagcagta  
 301 ttcccatgct ctgtataacc aatggaagtg gtgcaaacag aaaaccaagt caccagctg  
 361 ctgtctcaat tgcaggaaaa gaaactctt catctgctgc taaaagtgg acagaaaaaa  
 421 agaaagaaa accacaagga cagagagaaa aaaaagagga atctcattct aatgatcaaa  
 481 gtccacaaat tgcagcatca cttctctccc agccctctc acaacctctc caaatacaca  
 541 gacaaactcc agaaagcaag aatgctactc ccacaaaag cataaaacga ccatcaccag  
 601 ctgaaaagtc acataattct tgggaaaatt cagatgatag ccgtaataaa ttgtcgaaaa  
 661 taccttcaac acccaaatta ataccaaaag ttaccaaaac tgcagacaag cataaagatg  
 721 tcatcatcaa ccaagaagga gaatatatta aaatgtttat gcgcggtcgg ccaattacca  
 781 tgttcattcc ttccgatgtt gacaactatg atgacatcag aacggaactg cctcctgaga  
 841 agtcaaact ggagtgaggca tatggtatc gaggaaagga ctgtagagct aatgtttacc  
 901 ttcttccgac cggggaata gtttatttca ttgcatcagt agtagtacta ttaattatg  
 961 aggagagaac tcagcgacac tacctgggcc atacagactg tgtgaaatgc cttgctatac  
 1021 atcctgacaa aattaggatt gcaactggac agatagctgg cgtggataaa gatggaaggg  
 1081 ctctacaacc ccacgtcaga gtgtgggatt ctgttactct atccacactg cagattattg  
 1141 gacttggcac ttttgagcgt ggagtaggat gcctggattt ttcaaaagca gattcagggtg  
 1201 ttcatthtat tgttattgat gactccaatg agcatatgct tactgtatgg gactggcaga  
 1261 ggaaagcaaa aggagcagaa ataaagacaa caaatgaagt tgttttggt gtggagtttc  
 1321 acccaacaga tgcaaatacc ataattacat gcggtaaatc tcatattttc ttctggacct  
 1381 ggagcggcaa ttcactaaca agaaaacagg gaatttttgg gaaatatgaa aagcctaaat  
 1441 ttgtgcagt tttagcattc ttggggaatg gagatgttct tactggagac tcaggtggag

ES 2 660 149 T3

1501 tcatgcttat atggagcaaa actactgtag agccccacacc tgggaaagga cctaaagggtg  
 1561 tatatcaaat cagcaaacia atcaaagctc atgatggcag tgtgttcaca ctttgtcaga  
 1621 tgagaaatgg gatgttatta actggaggag ggaaagacag aaaaataatt ctgtgggatc  
 1681 atgatctgaa tcctgaaaga gaaatagagc accaggagct gcaagccatg cagatggagc  
 1741 tgcagagccc tgagtacaag ctgagcaagc tccgcacctc gaccatcatg accgactaca  
 1801 accccaacta ctgctttgct ggcaagacct cctccatcag tgacctgaag gagggtccgc  
 1861 ggaaaaacat caccctcatt cgggggtctgg gccatggagc ctttggggag gtgtatgaag  
 1921 gccaggtgtc cggaatgccc aacgacccaa gccccctgca agtggctgtg aagacgctgc  
 1981 ctgaagtgtg ctctgaacag gacgaactgg atttccatc ggaagccctg atcatcagca  
 2041 aattcaacca ccagaacatt gtctgctgca ttgggggtgag cctgcaatcc ctgccccggt  
 2101 tcatcctgct ggagctcatg gcggggggag acctcaagtc cttcctccga gagaccgcc  
 2161 ctgccccgag ccagccctcc tccctggcca tgctggacct tctgcacgtg gctcgggaca  
 2221 ttgctgtgg ctgtcagtat ttggaggaaa accacttcat ccaccgagac attgctgcca  
 2281 gaaactgcct cttgacctgt ccaggccctg gaagagtggc caagattgga gacttcggga  
 2341 tggcccgaga catctacagg gcgagctact atagaaaggg aggctgtgct atgtgcacat  
 2401 ttaagtggat gccccagag gccttcatgg aaggaatatt cacttctaaa acagacacat  
 2461 ggtcctttgg agtgcgtgta tgggaaatct tttctcttgg atatatcca taccacagca  
 2521 aaagcaacca ggaagtctg gagtttgtca ccagtggagg ccggatggac ccaccaaga  
 2581 actgccctgg gcctgtatac cggataatga ctcagtgtg gcaacatcag cctgaagaca  
 2641 ggcccactt tgccatcatt ttggagagga ttgaatactg caccaggagc ccggatgtaa  
 2701 tcaacaccgc tttgccgata gaatatggct cacttgtgga agaggaagag aaagtgcctg  
 2761 tgaggcccaa ggacctgag ggggttctc ctctcctggt ctctcaacag gcaaacggg  
 2821 aggaggagcg cagcccagct gccccaccac ctctgcctac cacctcctct ggcaaggctg  
 2881 caaagaacc cacagctgca gaggtctctg ttcgagctcc tagagggccg gccgtggaag  
 2941 ggggacacgt gaatatggca ttctctcagt ccaaccctcc ttcggagtgt cacaaggtcc  
 3001 acggatccag aaacaagccc accagcttgt ggaaccaac gtacggctcc tggtttacag  
 3061 agaaaccac caaaaagaat aatcctatag caaagaagga gccacacgac aggggtaacc  
 3121 tggggctgga ggaagctgt actgtcccac ctaacgttgc aactgggaga cttccggggg  
 3181 cctcactgct cctagagccc tcttcgctga ctgccaatat gaaggaggta cctctgttca  
 3241 ggctacgtca cttcccttgt gggaatgtca attacggcta ccagcaacag ggcttgccct  
 3301 tagaagccgc tactgccct ggagctggct attacgagga taccattctg aaaagcaaga  
 3361 atagcatgaa ccagcctggg ccctgagctc ggtcgcacac tcacttctct tccttgggat  
 3421 ccctaagacc gtgga

Secuencia de la proteína EML4-ALK Variante 7 (BAH57336.1; GI: 227452651)

1 mdgfagsldd ssaastsdv qdralsalesr vqqqedeitv lkaaladvlr rlaisedhva  
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke  
 121 kpqgqrekke eshsndqspq iraspspps sqplqihrt pesknaptk sikrpspaek  
 181 shnswnsdd srnklskips tpklipkvtk tadkhkdvii nqegeyikmf mrgrpitmfi  
 241 psdvdnyddi rtelppekkl lewaygyrgk dcranvyllp tgeivyfias vvlfnyeer  
 301 tqrhlyghd cvkclaihp kiriatgqia gvdkdgrplq phvrwdsvt lstlqiiglg  
 361 tfergvglc fskadsgvhl cviddsnehm ltvwdwqrka kgaeiktne vvlavefhpt  
 421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq clafllngdv ltgdsggvm  
 481 iwskttept pgkpkgyvq iskqikahdg svftlcqmrn gmltgggkd rkiilwdhdl  
 541 npereiehqe lqamqmelqs peyklsklrt stimtdynpn ycfagktssi sdlkevprkn  
 601 itlirglghg afgevyegqv sgmpndpspl qvavktlpev cseqdeldfl mealiiskfn  
 661 hqnivrcigv slqslprfil lelmaggdlk sflretrprp sqpsslamlld llhvardiac  
 721 gcqyleenhf ihrdiaarnc lltcpgpgrv akigdfgmar diyrasyyrk ggcamlpvkw  
 781 mppeafmegi ftsktdtwsf gvllweifsl gymypsksn qevleftsv gmdppkncp  
 841 gpvyrimtqc wqhqpdrpn failerley ctqdpdvint alpieyglv eeeekvpvrp  
 901 kdpegvplll vsqqakreee rsapaappplp ttssgkaakk ptaaevsrv prgpaveggh  
 961 vnmafsqsn pselkhvhgs rnkptslwnp tygswftekp tkknnpiakk ephdrgnlgl  
 1021 egstcvppnv atgrlpgasl llepsltan mkevplfrlr hfpcgnvnyg yqqgglplea  
 1081 atapgaghye dtilksknsm nqpgp

5 Secuencia de ácido nucleico de KIF5B-ALK (AB462413.1; GI: 227452652)

1 tgcgagaaag atggcggacc tggccgagtg caacatcaaa gtgatgtgtc gcttcagacc  
 61 tctcaacgag tctgaagtga accgcggcga caagtacatc gccaagtttc agggagaaga  
 121 cacggctgtg atcgcgtcca agccttatgc atttgatcgg gtgttccagt caagcacatc

ES 2 660 149 T3

181 tcaagagcaa gtgtataatg actgtgcaaa gaagattggt aaagatgtac ttgaaggata  
 241 taatggaaca atatattgcat atggacaaac atcctctggg aagacacaca caatggaggg  
 301 taaacttcat gatccagaag gcatgggaat tattccaaga atagtgcaag atatttttaa  
 361 ttatatttac tccatggatg aaaaatttggg atttcatatt aaggtttcat attttgaaat  
 421 atatttggat aagataaggg acctgttaga tgtttcaaag accaaccttt cagttcatga  
 481 agacaaaaac cgagttccct atgtaaaggg gtgcacagag cgttttgtat gtagtccaga  
 541 tgaagttatg gataccatag atgaaggaaa atccaacaga catgtagcag ttacaaatat  
 601 gaatgaacat agctctagga gtcacagtat atttcttatt aatgtcaaac aagagaacac  
 661 acaaacggaa caaaagctga gtggaaaact ttatctgggt gatttagctg gtagtgaaaa  
 721 ggttagtaaa actggagctg aagggtgctg gctggatgaa gctaaaaaca tcaacaagtc  
 781 actttctgct cttggaaatg ttatttctgc tttggctgag ggtagtacat atgttccata  
 841 tcgagatagt aaaatgacaa gaatccttca agattcatta ggtggcaact gtagaaccac  
 901 tattgtaatt tgctgctctc catcatcata caatgagtct gaaacaaaat ctacactctt  
 961 atttggccaa agggccaaaa caattaagaa cacagtttgt gtcaatgtgg agttaactgc  
 1021 agaacagtgg aaaaagaagt atgaaaaaga aaaagaaaa aataagatcc tgcggaacac  
 1081 tattcagtgg cttgaaaatg agctcaacag atggcgtaat ggggagacgg tgctattga  
 1141 tgaacagttt gacaaagaga aagccaactt ggaagctttc acagtggata aagatattac  
 1201 tcttaccat gataaaccag caaccgcaat tggagttata ggaaatttta ctgatgctga  
 1261 aagaagaaag tgtgaagaag aaatgctaa attatacaa cagcttgatg acaaggatga  
 1321 agaaattaac cagcaaagtc aactggtaga gaaactgaag acgcaaagt tggatcagga  
 1381 ggagcttttg gcatctacca gaaggatca agacaatat caagctgagc tgaatcgct  
 1441 tcaagcagaa aatgatgcct ctaaagaaga agtgaagaa gttttacagg ccctagaaga  
 1501 acttgctgtc aattatgatc agaagtctca ggaagttgaa gacaaaacta agaatatga  
 1561 attgcttagt gatgaattga atcagaaatc ggcaacttta gcgagtatag atgctgagct  
 1621 tcagaaactt aaggaaatga ccaaccacca gaaaaaacga gcagctgaga tgatggcatc  
 1681 tttactaaaa gaccttgca aaataggaat tgctgtggga aataatgatg taaagcagcc  
 1741 tgagggaaact ggcattgatg atgaagagtt cactgttgca agactctaca ttagcaaat  
 1801 gaagtcagaa gtaaaaacca tggtagaacg ttgcaagcag ttagaaagca cacaaactga  
 1861 gagcaacaaa aaaatggaag aaaatgaaaa ggagttagca gcatgtcagc ttcgtatctc  
 1921 tcaacatgaa gccaaaatca agtcattgac tgaatacctt caaaatgtgg aacaaaagaa  
 1981 aagacagtgg gaggaatctg tcgatgcct cagtgaagaa ctagtccag ctgagacaca  
 2041 agagaaagtc catgaaatgg aaaagagca cttaataaag gttcagactg caaatgaagt  
 2101 taagcaagct gttgaacagc agatccagag ccatagagaa actcatcaaa aacagatcag  
 2161 tagtttgaga gatgaagtag aagcaaaagc aaaacttatt actgatcttc aagacaaaa  
 2221 ccagaaaatg atgttagagc aggaacgtct aagagtagaa catgagaagt tgaagccac  
 2281 agatcaggaa aagagcagaa aactacatga acttacggtt atgcaagata gacgagaaca  
 2341 agcaagacaa gacttgaagg gtttgaaga gacagtgga aaagaacttc agactttaca  
 2401 caacctgagc aaactctttg ttcaggacct ggctacaaga gttaaaaaga gtgctgagat  
 2461 tgattctgat gacaccggag gcagcgctgc tcagaagcaa aaaatctcct tcttgaaaa  
 2521 taatcttgaa cagctcacta aagtgcacaa acagttggtg cgtgataatg cagatctcgg  
 2581 ctgtgaactt cctaagttgg aaaagcgact tcgagctaca gctgagagag tgaagcttt  
 2641 ggaatcagca ctgaaagaag ctaaagaaaa tgcattctct gatcgcaaac gctatcagca  
 2701 agaagtagat cgcataaagg aagcagtcag gtcaaagaat atggccagaa gaggcattc  
 2761 tgacagatt gtgtaccgcc ggaagcacca ggagctgcaa gccatgcaga tggagctgca  
 2821 gagccctgag tacaagctga gcaagctcgg cacctcgacc atcatgaccg actacaacc  
 2881 caactactgc tttgctggca agacctcctc catcagtgac ctgaaggagg tgccgaggaa  
 2941 aaacatcacc ctcatcggg gctgggcca tggcgcttt ggggaggtgt atgaaggcca  
 3001 ggtgtccgga atgccaacg acccaagccc cctgcaagtg gctgtgaaga cgtgcctga  
 3061 agtgtgctct gaacaggacg aactggattt cctcatggaa gccctgatca tcagcaaat  
 3121 caaccaccag aacattgttc gctgcattgg ggtgagcctg caatccctgc cccggttcat  
 3181 cctgctggag ctcatggcgg ggggagacct caagtcctc ctccgagaga cccgccctcg  
 3241 cccgagccag ccctcctccc tggccatgct ggacctctg cacgtggctc gggacattgc  
 3301 ctgtggctgt cagtatttgg aggaaaacca cttcatccac cgagacattg ctgccagaaa  
 3361 ctgcctcttg acctgtccag gcctggaag agtggccaag attggagact tcgggatggc  
 3421 ccgagacatc tacagggcga gctactatag aaagggaggc tgtgccatgc tccagttaa  
 3481 gtggatgccc ccagaggcct tcatggaagg aatattcact tctaaaacag acagatggtc  
 3541 ctttggagtg ctgctatggg aaatcttttc tcttgatata atgccatacc ccagcaaaag  
 3601 caaccaggaa gttctggagt ttgtcaccag tggaggccgg atggaccac ccaagaactg  
 3661 ccctgggct gtataccgga taatgactca gtgctggcaa catcagcctg aagacaggcc  
 3721 caactttgcc atcattttgg agaggattga atactgcacc caggaccgg atgtaataca  
 3781 caccgctttg ccgatagaat atggtccact tgtggaagag gaagagaaag tgctgtgag

3841 gcccaaggac cctgaggggg ttctctctct cctgggtctct caacaggcaa aacggggagga  
 3901 ggagcgcagc ccagctgccc caccacctct gcctaccacc tcctctggca aggttgcaaa  
 3961 gaaaccacaca gctgcagagg tctctgttcg agtccctaga gggccggccg tgggaaggggg  
 4021 acacgtgaat atggcattct ctcagtccaa ccctccttcg gagttgcaca aggtccacgg  
 4081 atccagaaac aagcccacca gcttgtggaa cccaacgtac ggctcctggt ttacagagaa  
 4141 acccaccaaa aagaataatc ctatagcaaa gaaggagcca cagcacaggg gtaacctggg  
 4201 gctggagggg agctgtactg tcccacctaa cgttgcaact gggagacttc cgggggcctc  
 4261 actgctccta gagccctctt cgctgactgc caatatgaag gaggtacctc tgttcaggct  
 4321 acgtcacttc ccttgtggga atgtcaatta cggctaccag caacagggct tgccttaga  
 4381 agccgctact gccctggag ctggtcatta cgaggatacc attctgaaaa gcaagaatag  
 4441 catgaaccag cctgggacct gagctcggtc gcacactca

Secuencia de la proteína KIF5B-ALK (BAH57337.1; GI: 227452653)

1 madlaecnik vmcfrplne sevrngdkiy akfqqedtvv iaskpyafdr vfqsstsqeq  
 61 vyndcakkiv kvvlegyngt ifaygqtssg kthtmegklh dpegmgiipr ivqdifnyiy  
 121 smdenlefhi kvsyfeiyld kirdlldvsk tnlsvhedkn rvpvykgcte rfvcspdevm  
 181 dtidegksnr hvavtnmneh ssrshsifli nvkqentqte qklsqklylv dlagsekvsk  
 241 tgaegavlde akninkslsa lgnvisalae gstyvpyrds kmtrilqdsi ggnrcrttivi  
 301 ccspssynes etkstllfgq raktikntvc vnveltaeqw kkyekekek nkilrntiqw  
 361 lenelnrwrn getvpideqf dkekanleaf tvdkditltn dkpataigvi gnftdaerrk  
 421 ceeeiaklyk qlddkdeein qqsqvlvklk tqmldqeell astrrdqdnm qaelnrlqae  
 481 ndaskeevke vlqaleelav nydqksqeve dktkeyells delnqksatl asidaelqkl  
 541 kemtnhqkkr aaemmasllk dlaeigiavg nndvkqpegg gmideeftva rlyiskmkse  
 601 vktmvrckrq lestqtesnk kmeenekela acqlrisqhe akikslteyl qnveqkkrql  
 661 eesvdalsee lvqlraqekv hemekehlnk vqtanevkqa veqqiqshre thqkqisslr  
 721 deveakakli tdlqdnqkm mleqerlrve heklkatdqe ksrklheltv mqdrreqarq  
 781 dlkgleetva kelqtlhnlr klfvqdlatr vkksaeidsd dtggsaaqkq kisflennle  
 841 qltkvhkqlv rdnadlrcl pklekrlrat aervkalesa lkeakenasr drkryqqevd  
 901 rikeavrskn marrghsaqi vyrrkhqelq amqmelqspe yklsklrtst imtdynpnyc  
 961 fagktssisd lkevrknit lirlghgaf gevyeqqvsg mpndpsplqv avktlpevcs  
 1021 eqdeldflme aliiskfnhq nivrcigvsl qslprfille lmaggdlksf lretrprpsq  
 1081 psslamldll hvardiacgc qyleenhfih rdiaarncll tcpgpgrvak igdfgmardi  
 1141 yrasyyrkkg camlpvkwp peafmegift sktdtwsfgv llweifslgy mpypsksnqe  
 1201 vlefvtsggr mdppkncpgp vyrimtqcwq hqpdrpnfa iilerieyct qdpdvintal  
 1261 pieygpolve eekvpvrpkd pegvplllvs qqakreers paappplptt ssgkaakkpt  
 1321 aaevsvrvpr gpavegghvn mafsqsnpps elhkvhgsrn kptslwnpty gswftekptk  
 1381 knnpiakkep hdrnglleg sctvppnvat grlpgaslll epssltanmk evplfrlrhf  
 1441 pcgnvnygyq qqglpleaat apgaghyedt ilksknsnmq ppp

5 Secuencia NPM-ALK (t (2; 5) (translocación cromosómica p23; q35))\*

Secuencia TPM3-ALK (translocación cromosómica t (1; 2) (p25; p23))\*

Secuencia de ácido nucleico de TFGXL-ALK (AF390893.1; GI: 20269389)

1 atgaacggac agttggatct aagtgggaag ctaatcatca aagctcaact tggggaggat  
 61 attcggcgaa ttctattca taatgaagat attacttatg atgaattagt gctaattgat  
 121 caacgagttt tcagaggaaa acttctgagt aatgatgaag taacaataaa gtataaagat  
 181 gaagatggag atcttataac aatttttgat agttctgacc tttcctttgc aattcagtgc  
 241 agtaggatac tgaactgac attatttggt aatggccagc caagaccct tgaatcaagt  
 301 caggtgaaat atctccgtcg agaactgata gaacttcgaa ataaagtgaa tcgtttattg  
 361 gatagcttg aaccacctgg agaaccagga ccttcacca atattcctga aatgatact  
 421 gtggatggta ggaagaaaa gtctgcttct gattcttctg gaaaacagtc tactcaggtt  
 481 atggcagcaa gtatgtctgc ttttgatcct ttaaaaaacc aagatgaaat caataaaaat  
 541 gttatgtcag cgtttggtt aacagatgat caggtttcag ggccaccag tgctcctgca  
 601 gaagatcggt caggaacacc cgacagcatt gcttctcct cctcagcagc tcaccacca  
 661 ggcgttcagc cacagcagcc accatataca ggagctcaga ctcaagcagg tcagattgaa  
 721 gtgtaccgcc ggaagcaca ggagctgcaa gccatgcaga tggagctgca gagccctgag  
 781 tacaagctga gcaagctccg cacctcgacc atcatgaccg actacaaccc caactactgc

ES 2 660 149 T3

841 tttgctggca agacctcctc catcagtgc ctgaaggagg tgccgcggaa aaacatcacc  
901 ctcatcggg gtctgggcca tgggccttt ggggagggtg atgaaggcca ggtgtccgga  
961 atgcccacg acccaagccc cctgcaagtg gctgtgaaga cgctgcctga agtgtgctct  
1021 gaacaggacg aactggattt cctcatggaa gccctgatca tcagcaaatt caaccaccag  
1081 aacattgttc gctgcattgg ggtgagcctg caatccctgc cccggttcat cctgctggag  
1141 ctcatggcgg ggggagacct caagtccttc ctccgagaga cccgcctcg cccgagccag  
1201 ccctcctccc tggccatgct ggaccttctg cacgtggctc gggacattgc ctgtggctgt  
1261 cagtatttgg aggaaaacca cttcatccac cgagacattg ctgccagaaa ctgcctcttg  
1321 acctgtccag gccctggaag agtggccaag attggagact tcgggatggc ccgagacatc  
1381 tacagggcga gctactatag aaagggaggc tgtgccatgc tgccagttaa gtggatgccc  
1441 ccagaggcct tcatggaagg aatattcact tctaaaacag acacatggtc ctttggagtg  
1501 ctgctatggg aaatcttttc tcttgatat atgccatacc ccagcaaaaag caaccaggaa  
1561 gttctggagt ttgtcaccag tggaggccgg atggaccac ccaagaactg ccctgggcct  
1621 gtataccgga taatgactca gtgctggcaa catcagcctg aagacaggcc caactttgcc  
1681 atcattttgg agaggattga atactgcacc caggaccgg atgtaatcaa caccgcttg  
1741 ccgatagaat atggccact tgtggaagag gaagagaaag tgctgtgag gcccaaggac  
1801 cctgaggggg ttcctcctct cctggtctct caacaggcaa aacgggagga ggcgcgagc  
1861 ccagctgccc caccacctct gcctaccacc tcctctggca aggctgcaaa gaaaccaca  
1921 gctgcagagg tctctgttcg agtccctaga gggccggccg tggaaagggg acacgtgaat  
1981 atggcattct ctcagtccaa ccctcctcg gagttgcaca aggtccacgg atccagaaac  
2041 aagcccacca gcttgtggaa cccaacgtac ggctcctggg ttacagagaa acccaccaaa  
2101 aagaataatc ctatagcaaa gaaggagcca cacgacaggg gtaacctggg gctggaggga  
2161 agctgtactg tcccacctaa cgttgcaact gggagacttc cgggggcctc actgctccta  
2221 gagcctctt cgctgactgc caatatgaag gaggtacctc tgttcaggct acgtcacttc  
2281 ccttgtggga atgtcaatta cggctaccag caacagggct tgccttaga agccgctact  
2341 gccctggag ctggtcatta cgaggatacc attctgaaaa gcaagaatag catgaaccag  
2401 cctgggcct ga

Secuencia de la proteína TFGXL-ALK (AAM17922.1; GI: 20269390)\*

1 mngqldlsgk liikaqlged irripihned itydelvlmm qrvfrgklls ndevtikykd  
61 edgdilitifd ssdlstfaiqc srilkltlfrv ngqprpless qvkyllrreli elrnkvnrl1  
121 dsleppgepg pstnipendt vdgreeksas dssgkqstqv maasmsafdp lknqdeinkn  
181 vmsafgltd qvsgppsapa edrsgtpdsi assssaahpp gvqpqqppyt gaqtqagqie  
241 vyrrkhqelq amqmelqspe yklsklrtst imtdynpnyc fagktssisd lkevprknit  
301 lirlglhgaf gevyegqvsg mpndpsplqv avktlpevcs eqdeldflme aliiskfnhq  
361 nivrcigvsl qslprfille lmaggdlksf lretrprpsq psslamldll hvardiacgc  
421 qyleenhfih rdiaarncll tcpgpgrvak igdfgmardi yrasyyrkkg camlpvkwmp  
481 peafmegift sktdtwsfgv llweifslgy mpypsksnqe vlefvtsggr mdppkncpgp  
541 vyrimtqcwq hqpedrpnfa iilerieyct qdpdvintal pieyglpvee eekvpvrpkd  
601 pegvppllvs qqakreers paappplptt ssgkaakkpt aaevsrvrpr gpavegghvn  
661 mafsqsnpps elhkvhgsrn kptslwnpty gswftekptk knnpiakkep hdrnlgleg  
721 sctvppnvat grlpgaslll epssltanmk evplfrlrhf pcgnvnygyq qqglpleaat  
781 apgaghyedt ilksknsnmq pgp

5

Secuencia de ácido nucleico de TFGL-ALK (AF143407.1; GI: 6739534)

10

ES 2 660 149 T3

1 cctccgcaag cegtctttct cttagagttgt atatatagaa catcctggag tccaccatga  
 61 acggacagtt ggatctaagt gggagctaa tcatcaaagc tcaacttggg gaggatattc  
 121 ggcgaattcc tattcataat gaagatatta cttatgatga attagtgcta atgatgcaac  
 181 gagttttcag aggaaaactt ctgagtaatg atgaagtaac aataaagtat aaagatgaag  
 241 atggagatct tataacaatt tttgatagtt ctgacctttc ctttgcaatt cagtgcagta  
 301 ggatactgaa actgacatta tttgttaatg gccagccaag accccttgaa tcaagtcagg  
 361 tgaatatct cctgcgagaa ctgatagaac ttcgaaataa agtgaatcgt ttattggata  
 421 gcttggaaac acctggagaa ccaggacctt ccaccaatat tctgaaaat gatactgtgg  
 481 atggtaggga agaaaagtct gcttctgatt cttctggaaa acagtctact caggttatgg  
 541 cagcaagtat gtctgctttt gatcctttaa aaaaccaaga tgaatcaat aaaaatgta  
 601 tgtcagcgtt tggcttaaca gatgatcagg tttcagtgtg ccgccggaag caccaggagc  
 661 tgcaagccat gcagatggag ctgcagagcc ctgagtacaa gctgagcaag ctccgcacct  
 721 cgaccatcat gaccgactac aacccaact actgctttgc tggcaagacc tctccatca  
 781 gtgacctgaa ggaggtgccg cggaaaaaca tcacctcat tccgggtctg ggccatggcg  
 841 cctttgggga ggtgtatgaa ggccaggtgt ccggaatgcc caacgaccca agccccctgc  
 901 aagtggctgt gaagacgctg cctgaagtgt gctctgaaca ggacgaactg gatttctca  
 961 tggaaagcct gatcatcagc aaattcaacc accagaacat tgttcgctgc attgggtga  
 1021 gcctgcaatc cctgccccgg ttcatectgc tggagctcat ggcgggggga gacctcaagt  
 1081 ccttctcccg agagaccgcg cctgcgccga gccagccctc ctccctggcc atgctggacc  
 1141 ttctgcacgt ggctcgggac attgcctgtg gctgtcagta tttggaggaa aaccacttca  
 1201 tccaccgaga cattgctgcc agaaactgcc tcttgacctg tccaggccct ggaagagtgg  
 1261 ccaagattgg agacttcggg atggcccagag acatctacag ggcgagctac tatagaaagg  
 1321 gaggtgtgct catgctgcca gttaagtgga tgccccaga ggccttcatg gaaggaatat  
 1381 tcaattctaa aacagacaca tggctctttg gactgctgct atgggaaatc ttttctcttg  
 1441 gatatatgcc ataccccagc aaaagcaacc aggaagttct ggagtttgtc accagtggag  
 1501 gccgatgga cccaccaag aactgcctg ggcctgtata ccgataatg actcagtgtc  
 1561 ggcaacatca gcctgaagac aggcccaact ttgccatcat tttggagagg attgaatact  
 1621 gcaccagga cccgatgta atcaacaccg ctttgccgat agaataatgg cacttgtgg  
 1681 aagaggaaga gaaagtgcct gtgaggccca aggaccctga gggggttcct cctctcctgg  
 1741 tctctcaaca ggcaaacgg gaggaggagc gcagcccagc tgccccacca cctctgccta  
 1801 ccacctctc tggcaaggct gcaaaagaaac ccacagctgc agaggtctct gttcgagtcc  
 1861 cttagagggc ggccgtggaa gggggacacg tgaatatggc attctctcag tccaaccctc  
 1921 cttagaggtt gcacaaggtc cacggatcca gaaacaagcc caccagcttg tgaacccaa  
 1981 cgtacggctc ctggtttaca gagaaaccca ccaaaaagaa taatcctata gcaagaagg  
 2041 agccacacga caggggtaac ctggggctgg agggaagctg tactgtcca ctaacgttg  
 2101 caactgggag acttccgggg gcctcactgc tcttagagcc ctcttcgctg actgccaata  
 2161 tgaaggaggt acctctgttc aggtacgctc acttccctg tgggaatgct aattacggct  
 2221 accagcaaca gggcttgccc ttagaagccg ctactgcccc tggagctggt cattacgagg  
 2281 ataccattct gaaaagcaag aatagcatga accagcctgg gccctgagct cggctgcaca  
 2341 ctacttctc ttccttggga tccttaagac cgtggaggag agagaggcaa tggctccttc  
 2401 acaaccaga gaccaaatgt cacgttttgt tttgtgcaa cctattttga agtaccacca  
 2461 aaaaagctgt attttgaaaa tgcttttagaa aggttttgag catgggttca tcctattctt  
 2521 tcgaaagaag aaaatatcat aaaaatgagt gataaataca aggccagat gtggttgcat  
 2581 aaggttttta tgcatgtttg ttgtatactt ccttatgctt cttttaaatt gtgtgtgctc  
 2641 tgcttcaatg tagtcagaat tagctgcttc tatgtttcat agttggggctc atagatgttt  
 2701 ccttgccctg ttgatgtgga catgagccat ttgagggggag agggaacgga aataaaggag  
 2761 ttatttghtaa tgactaaaa

Secuencia de la proteína TFGL-ALK (AAF27292.1; GI: 6739535)\*

# ES 2 660 149 T3

1 mngqldlsgk liikaqlged irripihned itydelvlmm qrvfrgklls ndevtikykd  
61 edgdlitifd ssdlsfaiqc srilkltlfv ngqprpless qvkylrreli elrnkvnrl1  
121 dsleppgepg pstnipendt vdgreeksas dssgkqstqv maasmsafdp lknqdeinkn  
181 vmsafgltd d qvsvyrrkhq elqamqmelq speyklsklr tstimtdynp nycfagktss  
241 isdlkevprk nitlirglgh gafgevyegq vsgmpndpsp lqvavktlpe vcseqdelfd  
301 lmealiiskf nhqnivrcig vslqslprfi llelmaggdl ksflretrpr psqpsslaml  
361 dllhvardia cgcqyleenh fihrdiaarn cltccpgpgr vakigdfgma rdiyasyyr  
421 kggcamlpvk wmppeafmeg iftsktdtws fgvllweifs lgympypskns nqevlefvt  
481 ggrmdppknc ppgvyrimtq cwqhqpdrp nfaiilerie yctqdpdvin talpieygp1  
541 veeekvpvr pkdpegvpp1 lvsqqakree erspaapp1 pttssgkaak kptaaevsvr  
601 vprgpavegg hvnmafsqsn ppselkhvhg srnkptslwn ptygswftek ptkknnpiak  
661 kephdrnlg legsvctvppn vatgrlpgas lllpsslta nmkevplf1 rhfpcgnvny  
721 gyqqqglple aatapgaghy edtilkskns mnqpgp

## Secuencia de ácido nucleico de TFGS-ALK (AF125093.1; GI: 7229260)

1 cctccgcaag ccgtttttct ctagagttgt atatatagaa catcctggag tccaccatga  
61 acggacagtt ggatctaagt gggaagctaa tcatcaaagc tcaacttggg gaggatattc  
121 ggcgaattcc tattcataat gaagatatta cttatgatga attagtgcta atgatgcaac  
181 gagttttcag aggaaaactt ctgagtaatg atgaagtaac aataaagtat aaagatgaag  
241 atggagatct tataacaatt tttgatagtt ctgaccttcc ctttgcaatt cagtgcagta  
301 ggatactgaa actgacatta tttgttaatg gccagccaag accccttgaa tcaagtcagg  
361 tgaatatct ccgtcgagaa ctgatagaac ttcgaaataa agtgaatcgt ttattggata  
421 gcttgaacc acctggagaa ccaggacctt ccaccaatat tctgaaaat gtgtaccgcc  
481 ggaagcacca ggagctgcaa gccatgcaga tggagctgca gagccctgag tacaagctga  
541 gcaagctccg cacctcgacc atcatgaccg actacaacc caactactgc tttgctggca  
601 agacctctc catcagtgac ctgaaggagg tgccgcggaa aaacatcacc ctcattcggg

5

661 gtctgggcca tgggcgcttt ggggaggtgt atgaaggcca ggtgtccgga atgcccacg  
 721 acccaagccc cctgcaagtg gctgtgaaga cgctgcctga agtgtgctct gaacaggacg  
 781 aactggattt cctcatggaa gccctgatca tcagcaaatt caaccaccag aacattgttc  
 841 gctgcattgg ggtgagcctg caatccctgc cccggttcat cctgctggag ctcatggcgg  
 901 ggggagacct caagtccttc ctccgagaga cccgccctcg cccgagccag cctcctccc  
 961 tggccatgct ggaccttctg cacgtggctc gggacattgc ctgtggctgt cagtatttgg  
 1021 aggaaaacca cttcatccac cgagacattg ctgccagaaa ctgcctcttg acctgtccag  
 1081 gccctggaag agtggccaag attggagact tgggatggc ccgagacatc tacagggcga  
 1141 gctactatag aaagggagge tgtgccatgc tgccagttaa gtggatgccc ccagaggcct  
 1201 tcatggaagg aatattcact tctaaaacag acacatggtc ctttggagtg ctgctatggg  
 1261 aaatcttttc tcttgatat atgccatacc ccagcaaaag caaccaggaa gttctggagt  
 1321 ttgtcaccag tggaggccgg atggaccac ccaagaactg ccctgggcct gtataccgga  
 1381 taatgactca gtgctggcaa catcagcctg aagacaggcc caactttgcc atcattttgg  
 1441 agaggattga atactgcacc caggaccgg atgtaatcaa caccgctttg ccgatagaat  
 1501 atggtccact tgtggaagag gaagagaaag tgccctgtgag gcccaaggac cctgagggg  
 1561 ttcctcctct cctggtctct caacaggcaa aacgggagga ggagcgcagc ccagctgccc  
 1621 caccacctct gcctaccacc tctctggca aggtgcaaa gaaaccaca gctgcagagg  
 1681 tctctgttcg agtcctaga gggcggccg tggaaagggg acacgtgaat atggcattct  
 1741 ctacgtccaa cctccttcg gagttgcaca aggtccacgg atccagaaac aagcccacca  
 1801 gcttgtggaa cccaacgtac ggctcctggt ttacagagaa acccaccaaa aagaataatc  
 1861 ctatagcaaa gaaggagcca cagcagagg gtaacctggg gctggaggga agctgtactg  
 1921 tcccacctaa cgttgcaact gggagacttc cgggggctc actgctccta gagccctctt  
 1981 cgctgactgc caatatgaag gaggtacctc tgttcaggct acgtcacttc ccttgtggga  
 2041 atgtcaatta cggctaccag caacagggtc tgcccttaga agccgctact gccctggag  
 2101 ctggtcatta cgaggatacc attctgaaaa gcaagaatag catgaaccag cctgggcctt  
 2161 gagctcggtc gcacactcac ttctcttctc tgggatccct aagaccgtgg aggagagaga  
 2221 ggcaatggct ccttcacaaa ccagagacca aatgtcacgt tttgttttgt gccaacctat  
 2281 tttgaagtac caccaaaaaa gctgtatttt gaaaatgctt tagaaagggt ttgagcatgg  
 2341 gttcatccta ttctttcgaa agaagaaaat atcataaaaa tgagtataaa atacaaggcc  
 2401 cagatgtggt tgcataaggt ttttatgcat gtttgttga tacttcctta tgcttctttt  
 2461 aaattgtgtg tgctctgctt caatgtagtc agaattagct gcttctatgt ttcatagttg  
 2521 gggcataga tgtttccttg ccttgttgat gtggacatga gccatttgag gggagaggga  
 2581 acggaataa aggagtatt tgtaatgact aaaa

Secuencia de la proteína TFGS-ALK (AAF42734.1; GI: 7229261)\*

1 mngqldlsgk liikaqlged irripihned itydelvlmm qrvfrgklls ndevtikyk  
 61 edgdlitifd ssdlfaiqc srilkltlfv ngqprpless qvkyllrreli elrnkvnrll  
 121 dsleppgepg pstnipenvy rrxhqlqam qmelqspeyk lsklrtstim tdynpnycfa  
 181 gktssisdsk evprknitli rglghgafge vyegqvsgmp ndpsplqvav ktlpevcseq  
 241 deldflmeal iiskfnhqni vrcigvslqs lprfillelm aggdksflr etrprpsqps  
 301 slamldllhv ardiacgcqy leenhfihrd iaarncllhc ppggrvakig dfgmardiy  
 361 asyyrkggca mlpvkwmppe afmegiftsk tdtwsfgvll weifslgymp ypsksnqevl  
 421 efvtsggrmd ppkncpgpvy rimtqcwqhq pedrpnfaii lerieyctqd pdvintalpi  
 481 eygplveeee kvpvrpkdpe gvppllvsqq akreeerspa appplpttss gkaakkptaa  
 541 evsvrvprgp avegghvnma fsqsnppsel khvhgsrnkp tslwnptygs wftektkkn  
 601 npiakkephd rgnlglegsc tvppnvatgr lpgaslillep ssltanmkev plfrlrhfp  
 5 661 gnvnygyqqq glpleaatap gaghyedtil ksknsmnqpg p

Secuencia de ATIC-ALK (translocación cromosómica inv (2) (p23; q35))\*

Secuencia de CLTC-ALK (translocación cromosómica t (2; 17) (p23; q23))\*

Secuencia de ácido nucleico de MSN-ALK (AF295356.1; GI: 14625823)

10 1 aactccgctg cctttgccgc caccatgccc aaaacgatca gtgtgcgtgt gaccaccatg  
 61 gatgcagagc tggagtttgc catccagccc aacaccaccg ggaagcagct atttgaccag  
 121 gtgggtgaaaa ctattggcct gagggaaagt tggttctttg gtctgcagta ccaggacact  
 181 aaaggtttct ccacctggct gaaactcaat aagaaggtga ctgccagga tgtgcggaag  
 241 gaaagccccc tgctctttaa gttccgtgcc aagttctacc ctgaggatgt gtcccaggaa  
 301 ttgattcagg acatcactca gcgcctgttc tttctgcaag tgaaagaggg cattctcaat

## ES 2 660 149 T3

361 gatgatattt actgcccgc tgagaccgct gtgctgctgg cctcgtatgc tgtccagtct  
421 aagtatggcg acttcaataa ggaagtgc atagctggct acctggccgg agacaagttg  
481 ctcccgcaga gagtctgga acagcaciaa ctcaacaagg accagtggga ggagcggatc  
541 caggtgtggc atgaggaaca cctggtgatg ctcagggagg atgctgtcct ggaatatctg  
601 aagattgctc aagatctgga gatgtatggt gtgaactact tcagcatcaa gaacaagaaa  
661 ggctcagagc tgtggctggg ggtggatgcc ctgggtctca acatctatga gcagaatgac  
721 agactaactc ccaagatagg cttcccctgg agtgaatca ggaacatctc tttcaatgat  
781 aagaaatttg tcatcaagcc cattgacaaa aaagccccgg acttcgtctt ctatgctccc  
841 cggctgcgga ttaacaagcg gatcttggcc ttgtgcatgg ggaacatga actatacatg  
901 cgccgtcgca agcctgatac cattgaggtg cagcagatga aggcacaggc ccgggaggag  
961 aagcaccaga agcagatgga gctgtctatg ctggaaaatg agaagaagaa gcgtgaaatg  
1021 gcagagaagg agaaagagaa gattgaacgg gagaaggagg agctgatgga gaggctgaag  
1081 cagatcgagg aacagactaa gaaggctcag caagaactgg aagaacagac ccgtagggct  
1141 ctggaacttg agcaggaacg gaagcgtgcc cagagcgagg ctgaaaagct ggccaaggag  
1201 cgtcaagaag ctgaagaggc caaggaggcc ttgctgcagg cctcccggga ccagaaaaag  
1261 actcaggaac agctggcctt ggaaatggca gagctgacag ctogaatctc ccagctggag  
1321 atggcccagc agaagaagga gactgaggct gtggagtggc agcagaagca ggagctgcaa  
1381 gccatgcaga tggagctgca gagccctgag tacaagctga gcaagctccg cacctcgacc  
1441 atcatgaccg actacaacc caactactgc tttgctggca agacctcctc catcagtgac  
1501 ctgaaggagg tgcccgga aaacatcacc ctcattcggg gtctgggcca tggcgcctt  
1561 ggggaggtgt atgaaggcca ggtgtccgga atgcccacg acccaag

### Secuencia de la proteína MSN-ALK (AAK71522.1; GI: 14625824)\*

1 mpktisvrvt tmdaelefai qpnttgkqlf dqvvtiglr evwffglqyq dtkgfstwlk  
61 lnkkvtaqdv rkespllfkf rakfypedvs eeliqditqr lfflqvkegi lnddiycppe  
121 tavllasyav qskygdfnke vkhsgylagd kllpqrveq hklnkdwhee riqvwheehr  
181 gmlredavle ylkiaqdlem ygvnyfsikn kkgsewlvq dalglniyeq ndrlltpkigf  
241 pwseirnisf ndkkfvikpi dkkapdfvfy aprlrinkri lalcmgnhel ymrrrkpdti  
301 evqqmkaqar eekhqkqmer amlenekkr emaekekeki erekeelmer lkqieeqtkk  
361 aqqeleeqtr raleleqerk raqseaekla kerqeaeeak eallqasrdq kktqeqlale  
421 maeltarisq lemarqkes eavewqqqe lqamqmelqs peyklsklrt stimtdynpn  
5 481 ycfagktssi sdlkevprkn itlirglghg afgevyegqv sgmpndp

### Secuencia de ácido nucleico de TPM4-ALK variante menor (AF362887.1; GI: 14010353)

1 cgagaagttg agggagaaa gcgggcccgg gaacaggctg aggctgaggt ggcctccttg  
61 aaccgtagga tccagctggt tgaagaagag ctggaccgtg ctcaggagcg tgcggaggtg  
121 tctgaactaa aatgtggtga cctggaagaa gaactcaaga atgttactaa caatctgaaa  
181 tctctggagg ctgcatctga aaagtattct gaaaaggagg acaaatatga agaagaaatt  
241 aaacttctgt ctgacaaact gaaagaggct gagaccctg ctgaatttgc agagagaacg  
301 gttgcaaaa tggaaaagac aattgatgac ctggaagtgt acctccggaa gcaccaagag  
361 ctgcaagcca tgcagatgga gctgcagagc cctgagtaca agctgagcaa gctccgcacc  
421 ctcgac

### Secuencia de la proteína TPM4-ALK variante menor (AAK51964.1; GI: 14010354)

1 revegerrar eqaaeavasl nrriqlveee ldraqeraev selkcgdl ee elknvtnnk  
61 sleaasekys ekedkyeeei kllsdlkea etraefaert vaklektidd levylrkhqe  
121 lqamqmelqs peyklsklrt ld

### Secuencia de ácido nucleico de TPM4-ALK Variante principal (AF362886.1; GI: 14010351)

1 ctggcagagt cccgttgccg agagatggat gagcagatta gactgatgga ccagaacctg  
61 aagtgtctga gtgctgctga agaaaagtac tctcaaaaag aagataaata tgaggaagaa  
121 atcaagattc ttaactgataa actcaaggag gcagagaccg gtgctgaatt tgcagagaga  
181 acggttgcaa aactggaaaa gacaattgat gacctggaag tgtaccgccg gaagcaccag  
241 gagctgcaag ccatgcagat ggagctgcag agccctgagt acaagctgag caagctccgc  
301 acctcgac

### Secuencia de la proteína TPM4-ALK variante principal (AAK51963.1; GI: 14010352)

1 laesrcremd eqirlmdqnl kclsaaeeky sqkedkyeee ikiltdklke aetraefaer  
61 tvaklektid dlevyrrkhq elqamqmelq speyklsklr tst

Secuencia de MYH9-ALK (translocación cromosómica t (2; 22) (p23; q11.2))\*

5 Secuencia de RANBP2-ALK (translocaciones cromosómicas t (2; 2) (p23; q13) o inv (2) (p23; q11-13))\*

Secuencia de ALO17-ALK (translocación cromosómica t (2; 17) (p23; q25))\*

Secuencia de CARS-ALK (translocación cromosómica t (2; 11; 2) (p23; p15; q31))\*

10 \* Con la excepción de MSN-ALK y MYH-9, todas las proteínas de fusión contienen los 563 aminoácidos finales de ALK. MSN-ALK y MYH9 contienen los 567 y 566 aminoácidos finales, respectivamente.

15 “Mutaciones de ALK” se refiere en general a modificaciones en la secuencia del ácido nucleico y/o de aminoácidos con respecto a una secuencia de referencia de la quinasa del linfoma anaplásico. En algunas realizaciones, sin embargo, “mutaciones de ALK” puede referirse a mutaciones de la quinasa de linfomas anaplásicos específicas predictivas de la respuesta al tratamiento con agentes inhibidores de ALK (por ejemplo, PF-02341066 y/o PDD). Por ejemplo, se describen en el presente documento mutaciones del aminoácido cisteína en la posición 1156 (C1156) y/o el aminoácido de leucina en la posición 1196 (L1196) de la proteína ALK de tipo silvestre (NP\_004295) a un aminoácido distinto que confiere resistencia a los agentes inhibidores de la ALK. En una realización, la posición C1156 comprende un aminoácido tirosina y/o la posición L1196 comprende un aminoácido metionina. Un experto en la materia reconocerá que las posiciones de aminoácido que corresponden con las mutaciones “C1156” y “L1196” de la proteína ALK de tipo silvestre tendrán distintos números con respecto a distintas secuencias de referencia (por ejemplo, homólogos de ALK, proteínas ALK de fusión, etc.) sin afectar el valor predictivo de la respuesta al tratamiento con agentes inhibidores de ALK (por ejemplo, PF-02341066 y/o PDD). La presente invención se refiere a mutaciones en la posición 1156 de la ALK humana de tipo silvestre.

25 Un experto en la materia reconocerá además que existe una correspondencia conocida y definida entre la secuencia de aminoácidos de una proteína particular y la secuencia de nucleótidos que puede codificar para la proteína, como se define por el código genético (mostrado a continuación). Asimismo, existe una correspondencia conocida y definida entre la secuencia de nucleótidos de un ácido nucleico particular y la secuencia de aminoácidos que codifica ese ácido nucleico, como se define por el código genético.

CÓDIGO GENÉTICO

Alanina (Ala, A)	GCA, GCC, GCG, GCT
Arginina (Arg, R)	AGA, ACG, CGA, CGC, CGG, CGT
Asparagina (Asn, N)	AAC, AAT
Ácido aspártico (Asp, D)	GAC, GAT
Cisteína (Cys, C)	TGC, TGT
Ácido glutámico (Glu, E)	GAA, GAG
Glutamina (Gln, Q)	CAA, CAG
Glicina (Gly, G)	GGA, GGC, GGG, GGT
Histidina (His, H)	CAC, CAT
Isoleucina (Ile, I)	ATA, ATC, ATT
Leucina (Leu, L)	CTA, CTC, CTG, CTT, TTA, TTG
Lisina (Lys, K)	AAA, AAG
Metionina (Met, M)	ATG
Fenilalanina (Phe, F)	TTC, TTT
Prolina (Pro, P)	CCA, CCC, CCG, CCT
Serina (Ser, S)	AGC, AGT, TCA, TCC, TCG, TCT
Treonina (Thr, T)	ACA, ACC, ACG, ACT
Triptófano (Trp, W)	TGG
Tirosina (Tyr, Y)	TAC, TAT
Valina (Val, V)	GTA, GTC, GTG, GTT
Señal de terminación (final)	TAA, TAG, TGA

35 Una característica importante y bien conocida del código genético es su redundancia, mediante la cual, para la mayoría de los aminoácidos utilizados para hacer proteínas puede emplearse más de un triplete de nucleótidos codificante (por ejemplo, como se ilustra anteriormente). Por lo tanto, varias secuencias de nucleótidos distintas pueden codificar una dada secuencia de aminoácidos. Tales secuencias de nucleótidos se consideran funcionalmente equivalentes dado que dan como resultado la producción de la misma secuencia de aminoácidos en todos los organismos (aunque determinados organismos pueden traducir algunas secuencias de forma más eficaz que otros). Además, en una dada secuencia de nucleótidos se puede encontrar, de forma ocasional, una variante metilada de una purina o una pirimidina. Tales metilaciones no afectan la relación codificante entre el codón de trinucleótidos y el correspondiente aminoácido. Además, un experto en la materia entenderá cómo mutar nucleótidos

de un codón específico de forma que se modifique de forma específica un aminoácido codificado a base de la tabla de codones relevante. Por ejemplo, el codón para la Cys-1156 es "TGC" y el de Tyr puede ser "TAT" o "TAC". Por lo tanto, una única sustitución nucleotídica de G a A en la posición 2 del codón codificará una tirosina en lugar de una cisteína. Un experto en la materia puede realizar manipulaciones similares para diseñar otras mutaciones.

"Compuesto de unión" se referirá a una composición de unión, tal como una molécula pequeña, un anticuerpo, un péptido, un ligando peptídico o no peptídico, una proteína, un oligonucleótido, un análogo de oligonucleótido, tal como un ácido nucleico peptídico, una lectina o cualquier otra entidad molecular que tenga la capacidad de unirse de forma específica a una proteína o molécula diana, o de una formación estable de un complejo con un analito de interés, tal como un complejo de proteínas.

"Fracción de unión" significa cualquier molécula a la que pueden unirse de forma directa o indirecta etiquetas moleculares, que tenga la capacidad de unirse a un analito de forma específica. Las fracciones de unión incluyen, pero sin limitación, anticuerpos, composiciones de unión a anticuerpo, péptidos, proteínas, ácidos nucleicos y moléculas orgánicas que tengan un peso molecular de hasta aproximadamente 1000 daltons, y que contengan átomos seleccionados del grupo que consiste en hidrógeno, carbono, oxígeno, nitrógeno, azufre y fósforo.

Un "biomarcador" o "marcador" es un gen, ARNm o proteína que puede modificarse, en el que dicha modificación está asociada con el cáncer. La modificación puede ser en cantidad, estructura y/o actividad en un tejido canceroso o célula cancerosa, en comparación con su cantidad, estructura y/o actividad en un tejido o célula normal o sano (por ejemplo, un control), y está asociada con una enfermedad, tal como el cáncer. Por ejemplo, un marcador de la presente invención que está asociado con cáncer o es predictivo de la respuesta a una terapia antineoplásica, puede tener una secuencia de nucleótidos, secuencia de aminoácidos, translocación cromosómica, inversión intracromosómica, número de copias, nivel de expresión, nivel de proteína, actividad de proteína o estado de metilación modificado, en un tejido canceroso o célula cancerosa, en comparación con un tejido o célula normal y sano. Adicionalmente, un "marcador" incluye una molécula cuya estructura está modificada, por ejemplo, mutada (contiene una mutación), por ejemplo, difiere de la secuencia de tipo silvestre a nivel de nucleótidos o aminoácidos, por ejemplo, por una sustitución, delección o inserción, cuando está presente en un tejido o célula asociado con una enfermedad, tal como el cáncer.

Los términos "cáncer" o "tumor" se refieren a la presencia de células que poseen características típicas de células que provocan cáncer, tal como proliferación incontrolada, inmortalidad, potencial metastásico, crecimiento rápido y velocidad de crecimiento, y determinados aspectos morfológicos característicos. Las células cancerosas a menudo están en forma de un tumor, pero tales células pueden existir solas, sin un animal, o puede ser una célula cancerosa no tumorigénica, tal como una célula de leucemia. Como se usa en el presente documento, el término "cáncer" incluye cánceres premalignos así como malignos. Los cánceres incluyen, pero sin limitación, cáncer de linfocitos B, por ejemplo, mieloma múltiple, macroglobulinemia de Waldenstrom, enfermedades de la cadena pesada, tales como, por ejemplo, enfermedad de la cadena alfa, enfermedad de la cadena gamma, enfermedad de la cadena mu, gamapatía monoclonal benigna, amiloidosis inmunocítica, melanomas, cáncer de mama, cáncer de pulmón (tal como carcinoma de pulmón no microcítico o CPNM), cáncer de bronquios, cáncer colorrectal, cáncer de próstata, cáncer pancreático, cáncer de estómago, cáncer de ovario, cáncer de vejiga urinaria, cáncer de cerebro o de sistema nervioso central, cáncer del sistema nervioso periférico, cáncer esofágico, cáncer de cuello uterino, cáncer de útero o de endometrio, cáncer de la cavidad oral o de faringe, cáncer de hígado, cáncer de riñón, cáncer testicular, cáncer de las vías biliares, cáncer de intestino delgado o de apéndice, cáncer de glándula salival, cáncer de tiroidea, cáncer de glándula suprarrenal, osteosarcoma, condrosarcoma, cáncer de tejidos hemáticos, adenocarcinomas, tumores miofibroblásticos inflamatorios, tumor de estroma gastrointestinal (TEGI), cáncer de colon, mieloma múltiple (MM), síndrome mielodisplásico (SMD), trastorno mieloproliferativo (TMP), leucemia linfocítica aguda (LLA), leucemia mielocítica aguda (LMA), leucemia mielocítica crónica (LMC), leucemia linfocítica crónica (LLC), policitemia Vera, linfoma de Hodgkin, linfoma no Hodgkin (LNH), sarcoma de tejido blanco, fibrosarcoma, mixosarcoma, liposarcoma, sarcoma osteogénico, cordoma, angiosarcoma, endoteliosarcoma, linfangiosarcoma, linfangioendoteliosarcoma, sinovioma, mesotelioma, tumor de Ewing, liomiosarcoma, rabdomyosarcoma, carcinoma escamoso, carcinoma basocelular, adenocarcinoma, carcinoma de glándula sudorípara, carcinoma de glándula sebácea, carcinoma papilar, adenocarcinomas papilares, estadenocarcinoma, carcinoma medular, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma de vía biliar, coriocarcinoma, seminoma, carcinoma embrionario, tumor de Wilms, carcinoma de vejiga, carcinoma epitelial, glioma, astrocitoma, meduloblastoma, craneofaringioma, epindimoma, pinealoma, hemangioblastoma, neuroma acústico, oligodendroglioma, meningioma, neuroblastoma, retinoblastoma, linfoma folicular, linfoma de linfocitos B grandes difuso, linfoma de células del manto, carcinoma hepatocelular, cáncer de tiroides, cáncer gástrico, cáncer de cabeza y cuello, cánceres microcíticos, trombocitemia esencial, metaplasia mieloide agnógena, síndrome hipereosinofílico, mastocitosis sistémica, hipereosinofilia familiar, leucemia eosinofílica crónica, cánceres neuroendocrinos, tumores carcinoides y similares.

"Agente quimioterapéutico" significa una sustancia química, tal como un agente citotóxico o citostático, que se utiliza para tratar una afección, por ejemplo, un cáncer.

“Complementario” se refiere a un amplio concepto de complementariedad de secuencia entre regiones de dos cadenas de ácido nucleico o entre dos regiones de la misma cadena de ácido nucleico. Se sabe que un resto adenina de una primera región de ácido nucleico tiene la capacidad de formar enlaces de hidrógeno específicos (“emparejamiento de bases”) con un resto de una segunda región de ácido nucleico que es antiparalela con respecto a la primera región si el resto es timina o uracilo. De forma similar, se sabe que un resto citosina de una primera cadena de ácido nucleico tiene la capacidad de emparejamiento de bases con un resto de una segunda cadena de ácido nucleico que es antiparalela con respecto a la primera cadena si el resto es guanina. Una primera región de un ácido nucleico es complementaria a una segunda región del mismo ácido nucleico o uno distinto si, cuando las dos regiones están dispuestas en una forma antiparalela, al menos un resto nucleotídico de la primera región tiene la capacidad de emparejamiento de bases con un resto de la segunda región. En determinadas realizaciones, la primera región comprende una primera porción y la segunda región comprende una segunda porción, por lo cual, cuando la primera y segunda porciones se disponen en una forma antiparalela, al menos aproximadamente el 50 %, al menos aproximadamente el 75 %, al menos aproximadamente el 90 % o al menos aproximadamente el 95 % de los restos nucleotídicos de la primera porción tienen la capacidad de emparejamiento de bases con los restos nucleotídicos en la segunda porción. En otras realizaciones, todos los restos nucleotídicos de la primera porción tienen la capacidad de emparejamiento de bases con los restos nucleotídicos de la segunda porción.

El “número de copias de un gen” o el “número de copias de un marcador” se refiere al número de secuencias de ADN en una célula que codifican un producto génico particular. En general, para un dado gen, un mamífero tiene dos copias de cada gen. El número de copias puede aumentarse, sin embargo, mediante amplificación o duplicación génica, o reducirse mediante delección.

Un marcador está “fijo” en un sustrato si está asociado de forma covalente o no covalente con el sustrato, de forma que el sustrato puede aclararse con un líquido (por ejemplo, solución salina convencional de citrato, pH 7,4) sin que una fracción sustancial del marcador se disocie del sustrato.

“Razón de riesgos instantáneos”, como se usa en el presente documento, se refiere un método estadístico utilizado para generar una estimación del riesgo relativo. “Razón de riesgos instantáneos” es la razón entre el riesgo predicho de un grupo frente a otro grupo. Por ejemplo, las poblaciones de pacientes tratados con un agente inhibidor de ALK frente a sin un agente inhibidor de ALK pueden evaluarse para si el agente inhibidor de ALK es eficaz o no en aumentar el tiempo para una metástasis a distancia de la enfermedad, en particular con respecto al estado de mutación de ALK. Por ejemplo, tratar sujetos que albergan mutaciones de ALK en un tejido canceroso, como se describe en el presente documento, da como resultado un beneficio terapéutico aumentado de los agentes inhibidores de ALK con respecto a sujetos que no tienen dichas mutaciones de ALK en el tejido canceroso.

“Agente inhibidor de ALK” o “inhibidor de ALK”, como se usa en el presente documento, se refiere a un compuesto que puede inhibir la actividad biológica de ALK. Las actividades biológicas también pueden incluir la respuesta del paciente como se expone en la presente solicitud. Los agentes inhibidores de ALK ejemplares incluyen, pero son limitación, PF-02341066, PDD, 2-metil-11-(2-metilpropil)-4-oxo-4,5,6,11,12,13-hexahidro-2H-indazol[5,4-a]pirrol[3,4-c]carbazol-8-il [4-(dimetilamino)bencil]carbamato,(1S,2S,3R,4R)-3-({5-cloro-2-[(1-etil-2,3,4,5-tetrahidro-6-metoxi-2-oxo-1H-1-benzazepin-7-il)amino]-4-pirimidinil}amino)biciclo[2.2.1]hept-5-en-2-carboxamida, y NVP-TAE684 (véase, por ejemplo, PNAS 104: 270-275, 2007; Choi, Y.L. *et al.* (2008) Cancer Res. 68: 4971-2976 y Biochemistry 48: 3600-3609, 2009).

Los términos “homología” o “identidad”, como se usa indistintamente en el presente documento, se refieren a una similitud de secuencia entre dos secuencias de polinucleótido o entre dos secuencias de polipéptido, siendo la identidad una comparación más estricta. Las frases “identidad u homología porcentual” e “identidad u homología %” se refiere al porcentaje de similitud de secuencia encontrado en una comparación de dos o más secuencias de polinucleótido o dos o más secuencias de polipéptido. “Similitud de secuencia” se refiere a la similitud porcentual en la secuencia de pares de bases (determinado mediante cualquier método adecuado) entre dos o más secuencias de polinucleótidos. Dos o más secuencias pueden ser cualquier valor entre 0-100 % similares o cualquier valor entero entre estos. La identidad o similitud puede determinarse comparando una posición en cada secuencia que pueda alinearse para fines de comparación. Cuando una posición en la secuencia comparada está ocupada por la misma base nucleotídica o aminoácido, entonces las moléculas son idénticas en esa posición. Un grado de similitud o identidad entre secuencias de polinucleótido es una función de la cantidad de nucleótidos idénticos o coincidentes en las posiciones compartidas por las secuencias de polinucleótidos. Un grado de identidad de secuencias de polipéptido es una función del número de aminoácidos idénticos en las posiciones compartidas por las secuencias de polipéptido. Un grado de homología o similitud de las secuencias de polipéptido es una función del número de aminoácidos en las posiciones compartidas por las secuencias de polipéptido. La expresión “homología sustancial”, como se usa en el presente documento, se refiere a una homología de al menos el 50 %, al menos el 60 %, al menos el 65 %, al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 % o más.

El cáncer se “inhibe” si al menos un síntoma del cáncer se alivia, acaba, desacelera o previene. Como se usa en el presente documento, el cáncer también se “inhibe” si se reduce, desacelera, retrasa o previene la recaída o metástasis del cáncer.

Un “ácido nucleico marcador” o “ácido nucleico biomarcador” es un ácido nucleico (por ejemplo, ADN, ARNm, ADNc) codificado por o que corresponde a un marcador de la presente invención. Por ejemplo, tales moléculas de ácido nucleico marcadoras incluyen ADN (por ejemplo, ADN genómico o ADNc) que comprenden toda la secuencia o una parcial de cualquiera de las secuencias de ácido nucleico expuestas en la Tabla 1, o el fragmento complementario o que hibrida con tal secuencia. Las moléculas de ácido nucleico marcadoras también incluyen ARN que comprende toda la secuencia o una parcial de cualquiera de las secuencias de ácido nucleico expuestas en la Tabla 1 o la complementaria de tal secuencia, en las que todos los restos timidina se reemplazan por restos uridina. Una “proteína marcadora” es una proteína codificada por, o que corresponde con, un marcador de la presente invención. Una proteína marcadora comprende toda la secuencia o una parcial de una proteína codificada por cualquiera de las secuencias expuestas en la Tabla 1 o un fragmento de las misma. Los términos “proteína” y “polipéptido” se utilizan indistintamente en el presente documento.

El número de copias “normal” de un marcador o el nivel de expresión “normal” de un marcador es el nivel de expresión, el número de copias del marcador, en una muestra biológica, por ejemplo, una muestra que contiene esputo, lavado broncoalveolar, derrame pleural, tejido, sangre entera, suero, plasma, raspado bucal, saliva, líquido cefalorraquídeo, orina, heces y médula ósea, de un sujeto, por ejemplo, un ser humano, no aquejado de cáncer.

Una “sobrexpresión” o “nivel significativamente mayor de expresión, número de copias y/o actividad” de mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) se refiere a un nivel de expresión, número de copias y/o actividad en una muestra de prueba que es mayor que el error típico del ensayo empleado para evaluar la expresión o número de copias y puede ser de al menos dos, al menos tres, al menos cuatro, al menos cinco o al menos diez, o más veces el nivel de expresión o número de copias de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1), en una muestra de control (por ejemplo, una muestra de un sujeto sano no aquejado de cáncer) o el nivel de expresión o número de copias promedio de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la tabla 1) en varias muestras de control.

El término “sonda” se refiere a cualquier molécula que tiene la capacidad de unirse de forma selectiva a una molécula diana diseñada de forma específica, por ejemplo un marcador de la presente invención. Las sondas puede sintetizarlas un experto en la materia u obtenerse a partir de preparaciones biológicas apropiadas. Con fines de detección de la molécula diana, las sondas pueden diseñarse de forma específica para estar marcadas, como se describe en el presente documento. Los ejemplos de moléculas que pueden utilizarse como sondas incluyen, pero sin limitación, ARN, ADN, proteínas, anticuerpos o monómeros orgánicos.

“RECIST” significa un acrónimo que significa “*response evaluation criteria in solid tumours*” (criterios de evaluación de respuesta en tumores sólidos) y es un conjunto de reglas publicadas que define cuándo los pacientes con cáncer mejoran (“responden”), permanece igual (“estables”) o empeora (“progresión”) durante los tratamientos. La respuesta, según se define por los criterios RECIST, se ha publicado, por ejemplo, en el Journal of the National Cancer Institute, Vol. 92, N.º 3, 2 de febrero de 2000, y los criterios RECIST pueden incluir otras definiciones y conjuntos de reglas publicados similares. Un experto en la materia entenderá las definiciones asociadas con los criterios RECIST, como se usan en el presente documento, tales como “PR”, “CR”, “SD” y “PD”.

“Respuesta”, “responder” al tratamiento y otras formas de este verbo, como se usa en el presente documento, se refiere a la reacción de un sujeto al tratamiento con un agente inhibidor de ALK. Como ejemplo, un sujeto responde a un tratamiento con un agente inhibidor de ALK si el crecimiento de un tumor en el sujeto se retarda en aproximadamente el 10 %, aproximadamente el 20 %, aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 70 %, aproximadamente el 80 %, aproximadamente el 90 % o más. En otro ejemplo, un sujeto responde al tratamiento con un agente inhibidor de ALK si un tumor en el sujeto se retrae en aproximadamente el 5 %, aproximadamente el 10 %, aproximadamente el 20 %, aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 % o más, determinado por cualquier medida apropiada, por ejemplo, por la masa o el volumen. En otro ejemplo, un sujeto responde al tratamiento con un agente inhibidor de ALK si el sujeto experimenta una esperanza de vida extendida en aproximadamente el 5 %, aproximadamente el 10 %, aproximadamente el 20 %, aproximadamente el 30 %, aproximadamente el 40 %, aproximadamente el 50 % o más, más allá de la esperanza de vida predicha si no se administrara el tratamiento. En otro ejemplo, un sujeto responde al tratamiento con un agente inhibidor de ALK si el sujeto tiene una supervivencia sin enfermedad, supervivencia global o tiempo transcurrido hasta la progresión aumentados. Pueden utilizarse varios métodos para determinar si un paciente responde a un tratamiento, incluyendo los criterios RECIST, como se expone anteriormente.

“Muestra”, “muestra de tejido”, “muestra de un paciente”, “muestra de células o tejidos de un paciente” o “muestra de ensayo” se refieren cada una recolección de células similares obtenidas a partir de un tejido de un sujeto o paciente. La fuente de la muestra de tejido puede ser de tejido sólido, como de un órgano, muestra de tejido, biopsia o aspirado congelados y/o conservados; sangre o cualquier constituyente de la sangre; fluidos corporales tales como líquido cefalorraquídeo, líquido amniótico, líquido peritoneal o líquido intersticial o células de cualquier momento de la gestación o desarrollo del sujeto. La muestra de tejido puede contener compuestos que en la naturaleza no están

entremezclados de forma natural con el tejido, tales como conservantes, anticoagulantes, tampones, fijadores, nutrientes, antibióticos o similares.

La cantidad de un marcador, por ejemplo, la expresión o número de copias de las mutaciones y/o productos génicos del gen ALK (por ejemplo, los marcadores expuestos en la Tabla 1), en un sujeto, es “significativamente” mayor o menor que la cantidad normal de un marcador, si la cantidad del marcador es mayor o menor, respectivamente, que el nivel normal, en una cantidad mayor que el error típico del ensayo empleado para evaluar la cantidad o al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos diez o más veces esa cantidad. Como alternativa, la cantidad del marcador en el sujeto puede considerarse “significativamente” mayor o menor que la cantidad normal si la cantidad es de al menos aproximadamente dos, al menos aproximadamente tres, al menos aproximadamente cuatro o al menos aproximadamente cinco veces, mayor o menor, respectivamente, que la cantidad normal del marcador.

Como se usa en el presente documento “suceso significativo” se referirá a un suceso en la enfermedad del paciente que es importante según determine un experto en la materia. Los ejemplos de sucesos significativos incluyen, por ejemplo, pero sin limitación, el diagnóstico principal, muerte, recurrencia, la determinación de que la enfermedad del paciente es metastásica, la recaída de la enfermedad del paciente o la progresión de la enfermedad del paciente, desde una cualquiera de las fases indicadas anteriormente a otra. Un suceso significativo puede ser un suceso importante utilizado para evaluar la SG, el TPP y/o utilizando el RECIST o cualquier otro criterio de respuesta, según determine un experto en la materia.

Como se usa en el presente documento, los términos “sujeto” y “paciente” se usan indistintamente. Como se usa en el presente documento, los términos “sujeto” y “sujetos” se refieren a un animal, por ejemplo, un mamífero incluyendo uno no primate (por ejemplo, una vaca, cerdo, caballo, burro, cabra, camello, gato, perro, cobaya, rata, ratón, oveja) y un primate (por ejemplo, un mono, tal como un mono cinomolgo, gorila, chimpancé y un ser humano).

Como se usa en el presente documento, “evolución temporal” se refiere a la cantidad de tiempo entre un suceso inicial y un suceso posterior. Por ejemplo, con respecto al cáncer de un paciente, la evolución temporal puede estar relacionada con la enfermedad del paciente y puede medirse calibrando sucesos significativos en el curso de la enfermedad, en la que el primer suceso puede ser el diagnóstico y el suceso posterior puede ser metástasis, por ejemplo.

“Tiempo transcurrido hasta la progresión” o “TTP” se refiere al tiempo medido desde el inicio del tratamiento hasta la progresión de un cáncer o una censura. La censura puede provenir de la finalización de un estudio o de un cambio en el tratamiento. El tiempo hasta la progresión también puede representarse como una probabilidad como, por ejemplo, en un diagrama de Kaplein-Meier en donde el tiempo hasta la progresión puede representar la probabilidad de estar sin progresión a lo largo de un tiempo particular, siendo ese tiempo el tiempo entre el inicio del tratamiento hasta la progresión o censura.

Un “polinucleótido transcrito” es un polinucleótido (por ejemplo, un ARN, un ADNc o un análogo de uno de ARN o ADNc) que es complementario a u homólogo con todo o una parte de un ARN maduro producido mediante la transcripción de un marcador de la presente invención y el procesamiento po-transcripcional normal (por ejemplo, corte y empalme), si lo hay, del transcrito y la transcripción inversa del transcrito.

“Tratar”, “tratamiento” y otras formas de esta palabra se refiere a la administración de un agente inhibidor de ALK para impedir el crecimiento de un cáncer, para provocar que un cáncer se retraiga en peso o volumen, para extender el tiempo de supervivencia esperado del sujeto y o el tiempo hasta la progresión del tumor, o similar.

Una “subexpresión” o “nivel de expresión, número de copias y/o actividad significativamente menor” de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) se refiere a un nivel de expresión o número de copias en una muestra de prueba que es mayor que el error típico del ensayo empleado para evaluar la expresión o número de copias, por ejemplo, al menos dos, al menos, al menos cuatro, al menos cinco o al menos diez o más veces menos que el nivel de expresión, número de copias y/o la actividad de las mutaciones y/o productos génicos del gen ALK (por ejemplo, los marcadores expuestos en la Tabla 1) en una muestra de control (por ejemplo, una muestra procedente de un sujeto sano no aquejado de cáncer), o el nivel de expresión, número de copias y/o actividad promedio de las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) en varias muestras de control.

## II. Métodos ejemplares de la invención

La presente invención está basada, al menos en parte, en la identificación de regiones específicas del genoma, incluyendo, por ejemplo, mutaciones de ALK asociadas con la predicción de la eficacia de los inhibidores de ALK en el tratamiento del cáncer. El análisis de las secuencias de expresión génica de ALK ha conducido a la identificación de mutaciones nuevas para los polipéptidos ALK (por ejemplo, los biomarcadores enumerados en la Tabla 1, incluyendo los polipéptidos EML4-ALK) que pueden hacer a los polipéptidos al menos parcialmente resistentes a la terapia con los inhibidores de ALK. Por consiguiente, la presente y/o ausencia de uno o más de tales biomarcadores

en diversos métodos descritos en el presente documento está dentro del ámbito de la presente divulgación. Las realizaciones de la invención se refieren a mutaciones en la posición 1156 de la ALK humana de tipo silvestre.

En algunas realizaciones, los métodos de la presente invención pueden utilizarse para controlar la progresión de un cáncer en un sujeto, en los que, si una muestra de un sujeto presenta una o más mutaciones de ALK (por ejemplo, mutaciones de EML4-ALK) identificadas en el presente documento durante la progresión del cáncer, por ejemplo, en un primer punto temporal y en un punto temporal posterior, entonces es menos probable que el cáncer responda al tratamiento mediado por un inhibidor de ALK y viceversa. En aún otra realización, entre el primer punto temporal y un punto temporal posterior el sujeto se ha sometido a un tratamiento, por ejemplo, quimioterapia, radioterapia, cirugía o cualquier otra estrategia terapéutica útil para inhibir el cáncer, ha completado el tratamiento o está en remisión.

Como se describe adicionalmente en el presente documento, uno o más biomarcadores de la presente invención (por ejemplo, mutaciones de ALK, incluyendo mutaciones de EML4-ALK) pueden identificarse de forma específica por la presencia en la secuencia genómica (por ejemplo, de estirpe germinal y/o somática) cuando se compara con una secuencia de referencia, tal como la SEQ ID NO: 1. Por ejemplo, los métodos descritos en el presente documento pueden implicar la detección de biomarcadores de la presente invención llevando a cabo una reacción de amplificación de ácidos nucleicos diana de un tramo de ADN que comprende una o más mutaciones enumeradas en la Tabla 1 y analizando la presencia de una o más mutaciones en el ácido nucleico diana amplificado.

Se conocen en la técnica diversas técnicas para amplificar ácidos nucleicos, tales como: la PCR (reacción en cadena de la polimerasa), descrita en la Patente de Estados Unidos N.º 4.683.195, Patente de Estados Unidos N.º 4.683.202 y Patente de Estados Unidos N.º 4.800.159, y su alternativa RT-PCR (transcripción inversa-PCR), en particular en su formato de una etapa, como se divulga en la Patente EP-B-0.569.272, la LCR (reacción en cadena de la ligasa), como se describe por ejemplo en la solicitud de patente EP-A-0.201.184, la RCR (reacción de reparación en cadena), como se describe por ejemplo en la solicitud internacional WO-A-90/01069, 3SR (replicación de secuencia autosostenida), como se describe por ejemplo en la solicitud de patente WO-A-90/06995, NASBA (Amplificación basada en secuencia de ácidos nucleicos), como se describe por ejemplo en el documento EP-B-0.397.269 y la Patente de Estados Unidos N.º 5.466.586, que utiliza como molde ADN bicatenario, y TMA (amplificación mediada por transcripción), como se describe por ejemplo en la Patente de Estados Unidos N.º 5.399.491.

La detección de la presencia de una o más de las mutaciones en el producto amplificado puede realizarse de diversas maneras que son bien conocidas en la técnica, tales como las metodologías de secuenciación de ADN, como la secuenciación de Sanger y la secuenciación profunda, el uso de enzimas de restricción, amplificación específica de alelos, PCR mediada por ácido peptidónucleico (APN), detección de diferencias conformacionales, como el polimorfismo de conformación de cadena simple (SSCP) y ensayos de electroforesis en gel en gradiente desnaturizante (DGGE) con etapas de detección en membranas (transferencia puntual), utilizando sondas oligonucleotídicas marcadas, ensayos con etapas de detección en placas de microtitulación, como hibridación inversa, ensayo de ligamiento de oligonucleótidos (OLA, MLPA), tecnología del cambio del primer nucleótido (FNC), tecnología de reticulación, PCR de ciclo rápido y análisis simultáneo de fluorescencia (por ejemplo, de nucleasa 5'/Taqman) y PCR seguido de minisequenciación utilizando espectrometría de masas o electroforesis capilar.

### III. Moléculas de ácido nucleico aisladas ejemplares

Un aspecto de la presente divulgación está relacionado con moléculas de ácido nucleico aisladas que corresponden a un biomarcador de la presente invención, incluyendo ácidos nucleicos que codifican un polipéptido que corresponde a un marcador de la presente invención o una porción de tal polipéptido. Las moléculas de ácido nucleico de la presente divulgación incluyen las moléculas de ácido nucleico que residen en las regiones genómicas de ALK o relacionadas con ALK (por ejemplo, de estirpe germinal y/o somática) identificadas en el presente documento, y/o codifican polipéptidos ALK o relacionados con ALK (por ejemplo, EML4-ALK). Las moléculas de ácido nucleico pueden comprender, consistir esencialmente en, o consistir en las secuencias nucleicas, o fragmentos de las mismas, presentadas en la Tabla 1. Las moléculas de ácido nucleico aisladas también incluyen moléculas de ácido nucleico suficientes para su uso como sondas de hibridación para identificar moléculas de ácido nucleico que corresponden a un marcador de la presente invención, incluyendo moléculas de ácido nucleico que codifican un polipéptido que corresponde a un marcador de la presente invención y fragmentos de tales moléculas de ácido nucleico, por ejemplo, las adecuadas para su uso como cebadores de PCR para la amplificación o la mutación de moléculas de ácido nucleico. Como se usa en el presente documento, la expresión "molécula de ácido nucleico" se pretende que incluya moléculas de ADN (por ejemplo, ADNc o ADN genómico) y moléculas de ARN (por ejemplo, ARNm), y análogos de los ADN o ARN generados utilizando análogos de nucleótido. La molécula de ácido nucleico puede ser mono o bicatenaria; en determinadas realizaciones la molécula de ácido nucleico es ADN bicatenario.

Una molécula de ácido nucleico "aislada" es una que está separada de otras moléculas de ácido nucleico que están presentes en la fuente natural de la molécula de ácido nucleico. En determinadas realizaciones, una molécula de ácido nucleico "aislada" está libre de las secuencias (tales como secuencias que codifican proteínas) que flanquean

de forma natural al ácido nucleico (es decir, secuencias emplazadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del que procede el ácido nucleico. Por ejemplo, en diversas realizaciones, la molécula de ácido nucleico aislada puede contener menos de aproximadamente 5 kb, menos de aproximadamente 4 kb, menos de aproximadamente 3 kb, menos de aproximadamente 2 kb, menos de aproximadamente 1 kb, menos de aproximadamente 0,5 kb o menos de aproximadamente 0,1 kb de secuencias de nucleótidos que flanquean de forma natural a la molécula de ácido nucleico en el ADN genómico de la célula de la que procede el ácido nucleico. Además, una molécula de ácido nucleico "aislada", tal como una molécula de ADNc, puede estar sustancialmente libre de otro material celular o medio de cultivo cuando se produce mediante técnicas recombinantes, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza de forma química.

La expresión "sustancialmente libre de otro material celular o medio de cultivo" incluye preparaciones de moléculas de ácido nucleico en las que la molécula está separada de componentes celulares de las células a partir de las que se aísla o se produce de forma recombinante. Por lo tanto, una molécula de ácido nucleico que está sustancialmente libre de material celular incluye preparaciones de la molécula de ácido nucleico que tienen menos de aproximadamente el 30 %, menos de aproximadamente el 20 %, menos de aproximadamente el 10 % o menos de aproximadamente el 5 % (en peso seco) de otro material celular o medio de cultivo.

Una molécula de ácido nucleico de la presente divulgación (por ejemplo las mutaciones del gen de ALK expuestas en la Tabla 1) puede aislarse utilizando técnicas de biología molecular convencionales y la información de secuencias en los registros de las bases de datos descritas en el presente documento. Utilizando todas o una porción de tales secuencias de ácido nucleico, las moléculas de ácido nucleico de la presente divulgación pueden aislarse utilizando técnicas de hibridación y clonación convencionales (por ejemplo, como se describe en Sambrook *et al.*, ed., *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Una molécula de ácido nucleico de la presente divulgación puede amplificarse utilizando como molde ADNc, ARNm o ADN genómico (por ejemplo, ADN genómico de estirpe germinal y/o somático) y cebadores oligonucleotídicos apropiados, de acuerdo con técnicas de amplificación por PCR convencionales. Las moléculas de ácido nucleico así amplificadas pueden clonarse en un vector apropiado y caracterizarse mediante análisis de secuencia de ADN. Además, mediante técnicas de síntesis convencionales pueden prepararse oligonucleótidos que corresponden a toda o una porción de una molécula de ácido nucleico de la presente invención, por ejemplo, utilizando un sintetizador de ADN automatizado.

En otra realización, una molécula de ácido nucleico aislada de la presente divulgación comprende una molécula de ácido nucleico que tiene una secuencia de nucleótidos complementaria con la secuencia de nucleótidos de un ácido nucleico que corresponde a un marcador de la presente invención o a la secuencia de nucleótidos de un ácido nucleico que codifica una proteína que corresponde a un marcador de la presente invención. Una molécula de ácido nucleico que es complementaria a una dada secuencia de nucleótidos es una que es suficientemente complementaria a la dada secuencia de nucleótidos como para hibridar con la secuencia de nucleótidos dada para formar de este modo un dúplex estable.

Además, una molécula de ácido nucleico de la presente divulgación puede comprender solo una porción de una secuencia de ácido nucleico, en la que la secuencia de ácido nucleico de longitud completa comprende un marcador de la presente invención o que codifica un polipéptido que corresponde a un marcador de la presente invención. Tales moléculas de ácido nucleico pueden utilizarse, por ejemplo, como una sonda o cebador. La sonda/cebador normalmente se utiliza como uno o más oligonucleótidos sustancialmente purificados. El oligonucleótido normalmente comprende una región de secuencia de nucleótidos que hibrida en condiciones rigurosas con al menos aproximadamente 7, al menos aproximadamente 8, al menos aproximadamente 9, al menos aproximadamente 10, al menos aproximadamente 11, al menos aproximadamente 12, al menos aproximadamente 13, al menos aproximadamente 14, al menos aproximadamente 15, al menos aproximadamente 16, al menos aproximadamente 17, al menos aproximadamente 18, al menos aproximadamente 19, al menos aproximadamente 20, al menos aproximadamente 21, al menos aproximadamente 22, al menos aproximadamente 23, al menos aproximadamente 24, al menos aproximadamente 25, al menos aproximadamente 26, al menos aproximadamente 27, al menos aproximadamente 28, al menos aproximadamente 29, al menos aproximadamente 30, al menos aproximadamente 31, al menos aproximadamente 32, al menos aproximadamente 33, al menos aproximadamente 34, al menos aproximadamente 35, al menos aproximadamente 36, al menos aproximadamente 37, al menos aproximadamente 38, al menos aproximadamente 39, al menos aproximadamente 40, al menos aproximadamente 45, al menos aproximadamente 50, al menos aproximadamente 55, al menos aproximadamente 60, al menos aproximadamente 65, al menos aproximadamente 70, al menos aproximadamente 75, al menos aproximadamente 80, al menos aproximadamente 85 kb, al menos aproximadamente 90, al menos aproximadamente 95, al menos aproximadamente 100 o más nucleótidos consecutivos de un ácido nucleico de la presente invención.

Las sondas basadas en la secuencia de una molécula de ácido nucleico de la presente divulgación pueden utilizarse para detectar transcritos o secuencias genómicas correspondientes a uno o más marcadores de la presente invención. La sonda comprende un grupo marcador unido a la misma, por ejemplo, un radioisótopo, un compuesto fluorescente, una enzima o un cofactor enzimático. Tales sondas pueden utilizarse como parte de un kit de prueba

diagnóstica para identificar células o tejidos que expresan de forma errónea la proteína, tal como midiendo niveles de una molécula de ácido nucleico que codifica la proteína en una muestra de células de un sujeto, por ejemplo, detectando niveles de ARNm o determinando si un gen que codifica la proteína se ha mutado o delecionado.

5 La divulgación abarca adicionalmente moléculas de ácido nucleico que son sustancialmente homólogas a las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1), de forma que lo son al menos el 60 %, al menos el 65 %, al menos el 65 %, al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 86 %, al menos el 87 %, al menos el 88 %, al menos el 89 %, al menos el 90 %, al menos el 91 %, al menos el 92 %, al menos el 93 %, al menos el 94 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, al menos el 99,5 % o más. En otras realizaciones, la invención abarca adicionalmente moléculas de ácido nucleico que son sustancialmente homólogas a las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1), de forma que difieren solo o al menos en 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 21, al menos 22, al menos 23, al menos 24, al menos 25, al menos 26, al menos 27, al menos 28, al menos 29, al menos 30, al menos 31, al menos 32, al menos 33, al menos 34, al menos 35, al menos 36, al menos 37, al menos 38, al menos 39, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65, al menos 70, al menos 75, al menos 80, al menos 85, al menos 90, al menos 95, al menos 100 nucleótidos o cualquier intervalo entre ellos.

La expresión "polimorfismo de nucleótido único" (SNP, forma siglada de *single nucleotide polymorphism*) se refiere a un sitio polimórfico ocupado por un único nucleótido, que es el sitio de variación entre las secuencias alélicas. El sitio habitualmente está precedido por y seguido de secuencias altamente conservadas del alelo (por ejemplo, secuencias que varían en menos de 1/100 o 1/1000 miembros de una población). Un SNP habitualmente surge debido a la sustitución de un nucleótido por otro en el sitio polimórfico. Los SNP también pueden surgir por una deleción de un nucleótido o una inserción de un nucleótido con respecto a un alelo de referencia. Normalmente, el sitio polimórfico está ocupado por una base distinta de la base de referencia. Por ejemplo, cuando el alelo de referencia contiene la base "T" (timidina) en el sitio polimórfico, el alelo modificado puede contener una "C" (citidina), "G" (guanina) o "A" (adenina) en el sitio polimórfico. Los SNP pueden aparecer en secuencias de ácido nucleico que codifican proteínas, en cuyo caso pueden dar lugar a una proteína defectuosa o variante de otra forma, o a una enfermedad genética. Tal SNP puede modificar la secuencia codificante del gen y, por lo tanto, especificar otro aminoácido (un SNP "de cambio de aminoácido") o un SNP puede introducir un codón de terminación (un SNP "finalizador"). Cuando un SNP no modifica la secuencia de aminoácidos de una proteína, el SNP se denomina "silencioso". Los SNP también pueden aparecer en regiones no codificantes de la secuencia de nucleótidos. Esto puede dar como resultado la expresión defectuosa de proteínas, por ejemplo, como resultado de un corte y empalme alternativo, o puede no tener efecto en la función de la proteína.

En otra realización, una molécula de ácido nucleico aislada de la presente divulgación es de al menos 7, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 55, al menos 60, al menos 65, al menos 70, al menos 75, al menos 80, al menos 85, al menos 90, al menos 95, al menos 100, al menos 125, al menos 150, al menos 175, al menos 200, al menos 250, al menos 300, al menos 350, al menos 400, al menos 450, al menos 550, al menos 650, al menos 700, al menos 800, al menos 900, al menos 1000, al menos 1200, al menos 1400, al menos 1600, al menos 1800, al menos 2000, al menos 2200, al menos 2400, al menos 2600, al menos 2800, al menos 3000, al menos 3500, al menos 4000, al menos 4500 o más nucleótidos de longitud e hibrida en condiciones rigurosas con una molécula de ácido nucleico que corresponde a un marcador de la presente invención o con una molécula de ácido nucleico que codifica una proteína que corresponde a un marcador de la presente invención. Como se usa en el presente documento, la expresión "hibrida en condiciones rigurosas" está destinada a describir condiciones de hibridación y de lavado en las cuales las secuencias de nucleótidos de al menos el 60 %, al menos el 65 %, al menos el 70 %, al menos el 75 %, al menos el 80 % o al menos el 85 % idénticas entre sí normalmente permanecen hibridadas entre sí. Tales condiciones rigurosas son conocidas para los expertos en la materia y pueden encontrarse en las secciones 6.3.1-6.3.6 de *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989). Otro ejemplo no limitativo de condiciones de hibridación rigurosas son la hibridación en cloruro de sodio/citrato de sodio (SSC) 6X a aproximadamente 45 °C, seguido de uno o más lavados en SSC 0,2X, SDS al 0,1 % a 50-65 °C.

La invención también incluye moléculas de ácido nucleico baliza molecular que tienen al menos una región que es complementaria con una molécula de ácido nucleico de la presente divulgación, de forma que la baliza molecular es útil para cuantificar en una muestra la presencia de la molécula de ácido nucleico de la presente invención. Un ácido nucleico "baliza molecular" es una molécula de ácido nucleico que comprende una pareja de regiones complementarias y que tiene un fluoróforo y un inactivador fluorescente asociados a ella. El fluoróforo y el inactivador están asociados con distintas porciones del ácido nucleico en una orientación tal que cuando las regiones complementarias están apareadas entre sí, el inactivador inactiva la fluorescencia del fluoróforo. Cuando las regiones complementarias de las moléculas de ácido nucleico no están apareadas entre sí, la fluorescencia del fluoróforo se inactiva en un menor grado. Las moléculas de ácido nucleico baliza molecular se describen, por ejemplo, en la Patente de Estados Unidos 5.876.930.

## IV. Proteínas y anticuerpos aislados ejemplares

Un aspecto de la presente divulgación se refiere a proteínas aisladas que corresponden a marcadores individuales de la presente invención y a porciones biológicamente activas de las mismas. En una realización, el polipéptido nativo que corresponde a un marcador puede aislarse a partir de fuentes celulares o tisulares mediante un esquema de purificación apropiado, utilizando técnicas convencionales de purificación de proteínas. En otra realización, los polipéptidos que corresponden a un marcador de la presente invención se producen mediante técnicas de ADN recombinante. De forma alternativa a la expresión recombinante, un polipéptido que corresponde a un marcador de la presente invención puede sintetizarse de forma química utilizando técnicas convencionales de síntesis peptídica.

Una proteína "aislada" o "purificada", o una porción biológicamente activa de la misma, está sustancialmente libre de material celular u otras proteínas contaminantes procedentes de la fuente tisular o celular de la que procede la proteína, o sustancialmente libre de precursores químicos u otros productos químicos, cuando se sintetiza de forma química. La expresión "sustancialmente libre de material celular" incluye preparaciones de proteína en que la proteína está separada de los componentes celulares de las células de las que se aisló o se produjo de forma recombinante. Por lo tanto, la proteína que está sustancialmente libre de material celular incluye preparaciones de proteína que tiene menos de aproximadamente el 30 %, menos de aproximadamente el 20 %, menos de aproximadamente el 10 % o menos de aproximadamente el 5 % (en peso seco) de proteína heteróloga (también denominada en el presente documento como una "proteína contaminante"). Cuando la proteína o porción biológicamente activa de la misma se produce de forma recombinante, puede estar sustancialmente libre de medio de cultivo, es decir, el medio de cultivo representa menos de aproximadamente el 20 %, menos de aproximadamente el 10 % o menos de aproximadamente el 5 % del volumen de la preparación de proteína. Cuando la proteína se produce por síntesis química, puede estar sustancialmente libre de precursores químicos u otros productos químicos, es decir, está separada de precursores químicos u otros productos químicos que están implicados en la síntesis de la proteína. Por consiguiente, tales preparaciones de la proteína tienen menos de aproximadamente el 30 %, menos de aproximadamente el 20 %, menos de aproximadamente el 10 %, menos de aproximadamente el 5 % (en peso seco) de precursores químicos o compuestos distintos del polipéptido de interés.

Las porciones biológicamente activas de un polipéptido que corresponde a un marcador de la presente invención incluye polipéptidos que comprenden secuencias de aminoácidos suficientemente idénticas a, o que proceden de, la secuencia de aminoácidos de la proteína que corresponde a las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcados expuestos en la Tabla 1) de la presente invención, las cuales incluyen menos aminoácidos que la proteína de longitud completa y presentan al menos una actividad de la correspondiente proteína de longitud completa. Normalmente, las porciones biológicamente activas comprenden un dominio o motivo con al menos una actividad de la correspondiente proteína. Una porción biológicamente activa de una proteína de la presente invención puede ser un polipéptido que sea de, por ejemplo, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más aminoácidos de longitud. Además, pueden prepararse mediante técnicas recombinantes otras porciones biológicamente activas, en las cuales otras regiones de la proteína se delecionan, y evaluarse para una o más actividades funcionales de la forma nativa de un polipéptido de la presente invención.

En determinadas realizaciones, el polipéptido tiene una secuencia de aminoácidos de una proteína codificada por una molécula de ácido nucleico enumerada en la Tabla 1. Otras proteínas útiles son sustancialmente idénticas (por ejemplo, al menos el 60, al menos el 65, al menos el 70, al menos el 75, al menos el 80, al menos el 85, al menos el 86, al menos el 87, al menos el 88, al menos el 89, al menos el 90, al menos el 91, al menos el 92, al menos el 93, al menos el 94, al menos el 95, al menos el 96, al menos el 97, al menos el 98, al menos el 99, al menos el 99,5 % o más) a una de estas secuencias y conservan la actividad funcional de la proteína (por ejemplo, conferir resistencia o sensibilidad a un inhibidor de ALK) de la correspondiente proteína de longitud completa, aunque difieren en la secuencia de aminoácidos.

Para determinar la identidad porcentual de dos secuencias de aminoácidos o de dos ácidos nucleicos, las secuencias se alinean con el fin de una comparación óptima (por ejemplo, pueden introducirse huecos en la secuencia de una primera secuencia de aminoácidos o de ácido nucleico para el alineamiento óptimo con una segunda secuencia de aminoácidos o de ácido nucleico). Después se comparan los restos de aminoácidos o nucleótidos en las correspondientes posiciones de aminoácidos o posiciones de nucleótidos. Cuando una posición en la primera secuencia está ocupada por el mismo resto de aminoácido o nucleótido que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en esa posición. La identidad porcentual entre dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (por ejemplo, identidad % = n.º de posiciones idénticas/n.º total de posiciones (por ejemplo, posiciones solapantes) x 100). En una realización, las dos secuencias tienen la misma longitud.

La determinación de la identidad porcentual entre dos secuencias puede llevarse a cabo utilizando un algoritmo matemático. Otro ejemplo no limitativo de un algoritmo matemático utilizado para la comparación de dos secuencias es el algoritmo de Karlin y Altschul (1990) Proc. Natl. Acad. Sci. USA 87: 2264-2268, modificado como en Karlin y Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5877. Tal algoritmo se incorpora en los programas NBLAST y XBLAST de Altschul, *et al.* (1990) J. Mol. Biol. 215: 403-410. Las búsquedas de nucleótidos por BLAST pueden

realizarse con el programa NBLAST, puntuación = 100, longitud de palabra = 12, para obtener secuencias de nucleótidos homólogas a moléculas de ácido nucleico de la presente invención. Las búsquedas de proteínas por BLAST pueden realizarse con el programa XBLAST, puntuación = 50, longitud de palabra = 3, para obtener secuencias de aminoácidos homólogas a moléculas proteicas de la presente invención. Para obtener alineamientos con huecos con el fines de comparación, puede utilizarse Gapped BLAST como se describe en Altschul *et al.* (1997) Nucleic Acids Res. 25:3389-3402. Como alternativa, puede utilizarse PSI-Blast para realizar una búsqueda iterativa que detecta relaciones distantes entre moléculas. Cuando se utilizan los programas BLAST, Gapped BLAST y PSI-Blast, pueden utilizarse los parámetros predeterminados de los programas respectivos (por ejemplo, XBLAST y NBLAST) (véase la página del NCBI en la internet en [nc-bi.nlm.nih.gov](http://nc-bi.nlm.nih.gov)). Otro ejemplo no limitativo de un algoritmo matemático utilizado para la comparación de secuencias es el algoritmo de Myers y Miller, (1988) Comput Appl Biosci, 4: 11-7. Tal algoritmo se incorpora en el programa ALIGN (versión 2.0) que es parte del paquete de programas informáticos de alineamiento de secuencias GCG. Cuando se utiliza el programa ALIGN para comparar secuencias de aminoácidos, puede utilizarse una tabla de restos de pesos PAM120, una penalización de longitud de hueco de 12 y una penalización de hueco de 4. Aún otro algoritmo útil para identificar regiones de similitud y alineamiento de secuencias locales es el algoritmo FASTA, como se describe en Pearson y Lipman (1988) Proc. Natl. Acad. Sci. USA 85: 2444-2448. Cuando se utiliza el algoritmo FASTA para comparar secuencias de nucleótidos o aminoácidos, puede utilizarse una tabla de restos de peso PAM120, por ejemplo, con un valor  $k$ -tuple de 2.

La identidad porcentual entre dos secuencias puede determinarse utilizando técnicas similares a las descritas anteriormente, permitiendo o no huecos. En el cálculo de la identidad porcentual solo se cuentan las coincidencias exactas.

Un polipéptido aislado que corresponde a un marcador de la presente invención, o un fragmento del mismo, puede utilizarse como inmunógeno para detectar anticuerpos utilizando técnicas convencionales para la preparación de anticuerpos policlonales y monoclonales. Puede utilizarse el polipéptido o proteína de longitud completa o, como alternativa, la invención proporciona fragmentos peptídicos antigénicos para su uso como inmunógenos. El péptido antigénico de una proteína de la presente invención comprende al menos 8 (o al menos 10, al menos 15, al menos 20 o al menos 30 o más), restos de aminoácido de la secuencia de aminoácidos de uno de los polipéptidos de la presente invención, y abarca un epítipo de la proteína de forma que un anticuerpo generado frente al epítipo forme un complejo inmunitario específico con un marcador de la presente invención al que corresponde la proteína. Los epítopos ejemplares que abarca el péptido antigénico son regiones que se emplazan en la superficie de la proteína, por ejemplo, regiones hidrófilas. Pueden utilizarse un análisis de hidrofobicidad de secuencias, un análisis de hidrofiliidad de secuencias o análisis similares para identificar regiones hidrófilas.

Un inmunógeno normalmente se utiliza para preparar anticuerpos inmunizando un sujeto adecuado (es decir, inmunocompetente) tal como un conejo, cabra, ratón u otro mamífero o vertebrado. Una preparación inmunogénica apropiada puede contener, por ejemplo, un polipéptido expresado de forma recombinante o sintetizado de forma química. La preparación puede incluir adicionalmente un adyuvante, tal como adyuvante de Freund completo o incompleto, o un agente inmunoestimulante similar.

Por consiguiente, otro aspecto de la presente divulgación se refiere a anticuerpos dirigidos contra a un polipéptido de la presente divulgación. Las expresiones "anticuerpo" y "sustancia de anticuerpo" que se usan indistintamente en el presente documento se refieren a moléculas de inmunoglobulina y porciones inmunológicamente activas de moléculas de inmunoglobulina, es decir, moléculas que contienen un sitio de unión a antígeno que se une de forma específica a un antígeno, tal como un polipéptido de la presente invención. Una molécula que se une de forma específica a un dado polipéptido de la presente invención es una molécula que se une al polipéptido, pero no se une sustancialmente a otras moléculas en una muestra, por ejemplo, una muestra biológica, que contiene de forma natural al polipéptido. Los ejemplos de porciones inmunológicamente activas de las moléculas de inmunoglobulina incluyen fragmentos F(ab) y F(ab)<sub>2</sub> que pueden generarse tratando el anticuerpo con una enzima tal como pepsina. La divulgación proporciona anticuerpos policlonales y monoclonales. La expresión "anticuerpo monoclonal" o "composición de anticuerpo monoclonal", como se usa en el presente documento, se refiere a una población de moléculas de anticuerpos que contiene sola una especie de un sitio de unión a antígeno que tiene la capacidad de inmunorreaccionar con un epítipo particular.

Los anticuerpos policlonales pueden prepararse como se describe anteriormente, inmunizando un sujeto adecuado con un polipéptido de la presente invención como inmunógeno. El título de anticuerpos en el sujeto inmunizado puede controlarse a lo largo del tiempo mediante técnicas convencionales que utilizan polipéptido inmovilizado, tal como con un ensayo inmunoabsorbente ligado a enzimas (ELISA). Si se desea, las moléculas de anticuerpo pueden recogerse o aislarse del sujeto (por ejemplo, de la sangre o suero del sujeto) y purificarse adicionalmente mediante técnicas bien conocidas, tal como cromatografía con proteína A para obtener la fracción IgG. En un tiempo apropiado tras la inmunización, por ejemplo, cuando los títulos de anticuerpos específicos es el más alto, pueden obtenerse del sujeto las células productoras de anticuerpos y utilizarse para preparar anticuerpos monoclonales mediante técnicas convencionales, tales como la técnica de hibridoma descrita originalmente por Kohler y Milstein (1975) Nature 256: 495-497, la técnica de hibridoma de linfocitos B humanos (véase Kozbor *et al.*, 1983, Immunol. Today 4: 72), la técnica de hibridoma-EBV (véase Cole *et al.*, pág. 77-96 en Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., 1985) o técnicas de trioma. La tecnología para la producción de hibridomas es bien

conocida (véase, en general, *Current Protocols in Immunology*, Coligan *et al.* ed., John Wiley & Sons, Nueva York, 1994). Las células de hibridoma que producen un anticuerpo monoclonal de la presente invención se detectan explorando sobrenadantes de cultivos de hibridomas para anticuerpos que se unan al polipéptido de interés, por ejemplo, utilizando un ensayo de ELISA convencional.

5 De forma alternativa a la preparación de hibridomas que secretan anticuerpos monoclonales, puede identificarse un anticuerpo monoclonal dirigido contra un polipéptido de la presente divulgación y aislarse mediante exploración de una biblioteca de inmunoglobulinas combinatoria recombinante (por ejemplo, una biblioteca de anticuerpos de presentación en fagos) con el polipéptido de interés. Los kits para la generación y exploración de bibliotecas de presentación en fagos están disponibles en el mercado (por ejemplo, el *Recombinant Phage Antibody System* de Pharmacia, N.º de Catálogo 27-9400-01 y el *SurfZAP Phage Display Kit* de Stratagene, N.º de Catálogo 240612). De forma adicional, los ejemplos de los métodos y reactivos particularmente susceptibles del uso en la generación y exploración de una biblioteca de presentación de anticuerpos pueden encontrar en, por ejemplo, la Patente de Estados Unidos N.º 5.223.409; la Publicación PCT N.º WO 92/18619; la Publicación PCT N.º WO 91/17271; la Publicación PCT N.º WO 92/20791; la Publicación PCT N.º WO 92/15679; la Publicación PCT N.º WO 93/01288; la Publicación PCT N.º WO 92/01047; la Publicación PCT N.º WO 92/09690; la Publicación PCT N.º WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9: 1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3: 81-85; Huse *et al.* (1989) *Science* 246: 1275- 1281; Griffiths *et al.* (1993) *EMBO J.* 12: 725-734.

20 De forma adicional, los anticuerpos recombinantes, tales como anticuerpos monoclonales quiméricos y humanizados, que comprenden tanto porciones humanas como no humanas, que pueden fabricarse utilizando técnicas de ADN recombinante convencionales, están dentro del ámbito de la presente divulgación. Tales anticuerpos monoclonales quiméricos y humanizados pueden producirse mediante técnicas de ADN recombinante conocidas en la técnica, por ejemplo, utilizando los métodos descritos en la Publicación PCT N.º WO 87/02671; Solicitud de Patente Europea 184.187; Solicitud de Patente Europea 171.496; Solicitud de Patente Europea 173.494; Publicación PCT N.º WO 86/01533; Patente de Estados Unidos N.º 4.816.567; Solicitud de Patente Europea 125.023; Better *et al.* (1988) *Science* 240: 1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84: 3439-3443; Liu *et al.* (1987) *J. Immunol.* 139: 3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84: 214-218; Nishimura *et al.* (1987) *Cancer Res.* 47: 999-1005; Wood *et al.* (1985) *Nature* 314: 446-449 y Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80: 1553-1559; Morrison (1985) *Science* 229: 1202-1207; Oi *et al.* (1986) *Bio/Techniques* 4:214; Patente de Estados Unidos 5.225.539; Jones *et al.* (1986) *Nature* 321: 552-525; Verhoeyan *et al.* (1988) *Science* 239: 1534 y Beidler *et al.* (1988) *J. Immunol.* 141: 4053-4060.

35 Los anticuerpos completamente humanos son particularmente convenientes para el tratamiento terapéutico de sujetos humanos. Tales anticuerpos pueden producirse utilizando ratones transgénicos que no tengan la capacidad de expresar los genes endógenos de las cadenas pesada y ligera de inmunoglobulina, pero que puedan expresar los genes de la cadena pesada y ligera de ser humano. Los ratones transgénicos se inmunizan del modo normal con un anticuerpo seleccionado, por ejemplo, todo o una porción de un polipéptido que corresponde a un marcador de la presente divulgación. Los anticuerpos monoclonales dirigidos contra el antígeno pueden obtenerse utilizando la tecnología de hibridomas convencional. Los transgenes de inmunoglobulina de ser humano que portan los ratones transgénicos se reordenan durante la diferenciación de los linfocitos B y posteriormente experimentan el cambio de clase y mutación somática. Por lo tanto, utilizando tal técnica, es posible producir anticuerpos IgG, IgA e IgE terapéuticamente útiles. Para una visión de conjunto de esta tecnología de producción de anticuerpos humanos, véase Lonberg y Huszar (1995) *Int. Rev. Immunol.* 13:65-93). Para una discusión detallada de esta tecnología de producción de anticuerpos humanos y de anticuerpos monoclonales humanos, y de protocolos de producción de tales anticuerpos, véase, por ejemplo, la Patente de Estados Unidos 5.625.126; la Patente de Estados Unidos 5.633.425; la Patente de Estados Unidos 5.569.825; la Patente de Estados Unidos 5.661.016 y la Patente de Estados Unidos 5.545.806. Además, pueden contratarse compañías tales como Abgenix, Inc. (Freemont, CA) para proporcionar anticuerpos humanos dirigidos contra un antígeno seleccionado utilizando una tecnología similar a la descrita anteriormente.

Los anticuerpos completamente humanos que reconocen un epítipo seleccionado pueden generarse utilizando una técnica denominada "selección guiada". En esta estrategia se utiliza un anticuerpo monoclonal no humano seleccionado, por ejemplo, un anticuerpo murino, para guiar la selección de un anticuerpo completamente humano que reconoce el mismo epítipo (Jespers *et al.*, 1994, *Bio/technology* 12: 899-903).

Un anticuerpo dirigido contra un polipéptido que corresponde a un marcador de la presente invención (por ejemplo, un anticuerpo monoclonal) puede utilizarse para aislar el polipéptido mediante técnicas convencionales, tales como cromatografía de afinidad o inmunoprecipitación. Además, tal anticuerpo puede utilizarse para detectar el marcador (por ejemplo, en un lisado celular o sobrenadante celular) para evaluar el nivel y patrón de expresión del marcador. Los anticuerpos pueden utilizarse también para controlar de forma diagnóstica los niveles de proteínas en tejidos y fluidos corporales (por ejemplo, en un fluido corporal que contiene células tumorales) como parte de un procedimiento de análisis clínico, por ejemplo, para, por ejemplo, determinar la eficacia de un dado régimen de tratamiento. La detección puede facilitarse acoplado el anticuerpo a una sustancia detectable. Los ejemplos de sustancias detectables incluyen, pero sin limitación, diversas enzimas, grupos prostéticos, materiales fluorescentes, materiales luminiscentes, materiales bioluminiscentes y materiales radioactivos. Los ejemplos de enzimas

adecuadas incluyen, pero sin limitación, peroxidasa de rábano picante, fosfatasa alcalina,  $\beta$ -galactosidasa o acetilcolinesterasa; los ejemplos de complejos de grupos prostéticos adecuados incluyen, pero sin limitación, estreptavidina/biotina y avidina/biotina; los ejemplos de materiales fluorescentes adecuados incluyen, pero sin limitación, umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinilamina fluoresceína, cloruro de dansilo o ficoeritrina; un ejemplo de un material luminiscente incluye, pero sin limitación, luminol; los ejemplos de materiales bioluminiscentes incluyen, pero sin limitación, luciferasa, luciferina y aequorina, y los ejemplos de materiales radioactivos adecuados incluyen, pero sin limitación 125, 131, 35, 3 limitado a, I, I, S o H.

#### V. Vectores de expresión recombinantes y células hospedadoras ejemplares

Otro aspecto de la presente divulgación se refiere a vectores, tales como vectores de expresión, que contienen un ácido nucleico que codifica un polipéptido que corresponde a un marcador de la presente divulgación (o una porción de tal polipéptido). Como se usa en el presente documento, el término "vector" se refiere a una molécula de ácido nucleico que tiene la capacidad de transportar otro ácido nucleico al que se la ha ligado. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN bicatenario circular en el que se han ligado segmentos de ADN adicionales. Otro tipo de vector es un vector vírico, en el que pueden ligarse en el genoma vírico segmentos de ADN adicionales. Determinados vectores tienen capacidad de replicación autónoma en una célula hospedadora en la que se han introducido (por ejemplo, vectores bacterianos que tienen un origen bacteriano de replicación y vectores episómicos de mamífero). Otros vectores (por ejemplo, vectores no episómicos de mamífero) se integran en el genoma de una célula hospedadora después de la introducción en la célula hospedadora y, de este modo, se replican junto con el genoma del hospedador. Además, determinados vectores, en concreto los vectores de expresión, tienen la capacidad de dirigir la expresión de genes a los que están unidos operativamente. En general, los vectores de expresión de utilidad en las técnicas de ADN recombinante a menudo están en forma de plásmidos (vectores). Sin embargo, se pretende que la divulgación incluya tales otras formas de vectores de expresión, tales como los vectores víricos (por ejemplo, retrovirus, adenovirus y virus adenoasociados con replicación defectuosa), que desempeñan funciones equivalentes.

Los vectores de expresión recombinante de la presente divulgación comprenden un ácido nucleico de la presente invención en una forma adecuada para la expresión del ácido nucleico en una célula hospedadora. Esto significa que los vectores de expresión recombinante incluyen una o más secuencias reguladoras, seleccionadas sobre la base de las células hospedadoras a utilizar para la expresión, que están unidas operativamente a la secuencia de ácido nucleico a expresar. Dentro de un vector de expresión recombinante, "unido operativamente" pretende significar que la secuencia de nucleótidos de interés está ligada a la secuencia (o secuencias) reguladora de una manera que permite la expresión de la secuencia de nucleótidos (por ejemplo, en un sistema de transcripción/traducción *in vitro* o en una célula hospedadora cuando el vector se introduce en la célula hospedadora). La expresión "secuencia reguladora" pretende incluir promotores, potenciadores y otros elementos de control de la expresión (por ejemplo, señales de poliadenilación). Tales secuencias reguladoras se describen, por ejemplo, en Goeddel, *Methods in Enzymology: Gene Expression Technology* vol. 185, Academic Press, San Diego, CA (1991). Las secuencias reguladoras incluyen las que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células hospedadoras y las que dirigen la expresión de la secuencia de nucleótidos solo en determinadas células hospedadoras (por ejemplo, secuencias reguladoras específicas de tejido). Los expertos en la materia apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula hospedadora a transformar, el nivel de expresión de proteína deseado y similares. Los vectores de expresión de la presente divulgación pueden introducirse en las células hospedadoras para, de este modo, producir proteínas o péptidos, incluyendo proteínas o péptidos de fusión, codificados por los ácidos nucleicos como se describe en el presente documento.

Los vectores de expresión recombinante de la presente divulgación pueden diseñarse para la expresión de un polipéptido que corresponde a un marcador de la presente invención en células procariotas (por ejemplo, *E. coli*) o eucariotas (por ejemplo, células de insecto {utilizando vectores de expresión de baculovirus}, células de levadura o células de mamífero). Las células hospedadoras adecuadas se discuten adicionalmente en Goeddel, citado anteriormente. Como alternativa, el vector de expresión recombinante puede transcribirse y traducirse *in vitro*, por ejemplo utilizando secuencias reguladoras del promotor T7 y la polimerasa T7.

La expresión de proteínas en los procariotas se lleva a cabo con mayor frecuencia en *E. coli* con vectores que contienen promotores constitutivos o inducibles, que dirigen la expresión de proteínas de fusión o que no son de fusión. Los vectores de fusión añaden una cantidad de aminoácidos a una proteína codificada en ellos, habitualmente al extremo amino de la proteína recombinante. Tales vectores de fusión normalmente sirven para tres propósitos: 1) aumentar la expresión de la proteína recombinante; 2) aumentar la solubilidad de la proteína recombinante y 3) ayudar en la purificación de la proteína recombinante actuando como un ligando en la purificación por afinidad. A menudo, en los vectores de expresión de fusión se introduce un sitio de escisión proteolítica en la unión de la fracción de fusión y la proteína recombinante para permitir la separación de la proteína recombinante de la fracción de fusión posterior a la purificación de la proteína de fusión. Tales enzimas y sus secuencias de reconocimiento afines incluyen el Factor Xa, la trombina y la enteroquinasa. Los vectores de expresión de fusión típicos incluyen pGEX (Pharmacia Biotech Inc; Smith y Johnson, 1988, *Gene* 67: 31-40), pMAL (New England

Biolabs, Beverly, MA) y pRIT5 (Pharmacia, Piscataway, NJ), los cuales fusionan la glutatión S-transferasa (GST), la proteína de unión a maltosa E o la proteína A, respectivamente, a la proteína recombinante diana.

Los ejemplos de vectores de expresión en *E. coli* no de fusión inducibles adecuados incluyen pTrc (Amann *et al.*, 1988, Gene 69: 301-315) y pET 11d (Studier *et al.*, pág. 60-89, en Gene Expression Technology: Methods in Enzymology vol.185, Academic Press, San Diego, CA, 1991). La expresión del gen diana a partir del vector pTrc se basa en la transcripción con la ARN polimerasa del hospedador a partir de un promotor de fusión híbrido trp-lac. La expresión del gen diana a partir del vector pET 11d se basa en la transcripción a partir de un promotor de fusión T7 gn10-lac mediada por una ARN polimerasa vírica coexpresada (T7 gn1). Esta polimerasa vírica es suministrada por las cepas hospedadoras BL21 (DE3) o HMS174(DE3) a partir de un profago residente que hospeda un gen T7 gn1 bajo el control transcripcional del promotor lacUV 5.

Una estrategia para maximizar la expresión de una proteína recombinante en *E. coli* es expresar la proteína en una bacteria hospedadora con una capacidad deteriorada para escindir de forma proteolítica la proteína recombinante (Gottesman, pág. 119-128, En Gene Expression Technology: Methods in Enzymology vol. 185, Academic Press, San Diego, CA, 1990). Otra estrategia es modificar la secuencia de ácido nucleico del ácido nucleico a insertar en un vector de expresión de forma que los codones individuales para cada aminoácido puedan estar entre los más utilizados en *E. coli* (Wada *et al.*, 1992, Nucleic Acids Res. 20: 2111-2118). Tal modificación de las secuencias de ácido nucleico de la presente invención puede llevarse a cabo mediante técnicas de síntesis de ADN convencionales.

En otra realización, el vector de expresión es un vector de expresión en levaduras. Los ejemplos de vectores para la expresión en la levadura *S. cerevisiae* incluyen pYepSec1 (Baldari *et al.*, 1987, EMBO J. 6: 229-234), pMFa (Kurjan y Herskowitz, 1982, Cell 30: 933-943), pJRY88 (Schultz *et al.*, 1987, Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, CA) y pPicZ (Invitrogen Corp, San Diego, CA).

Como alternativa, el vector de expresión es un vector de expresión de baculovirus. Los vectores de baculovirus disponibles para la expresión de proteínas en células de insecto cultivadas (por ejemplo, células Sf9) incluyen la serie pAc (Smith *et al.*, 1983, Mol. Cell Biol. 3: 2156-2165) y la serie pVL (Lucklow y Summers, 1989, Virology 170: 31-39).

En aún otra realización, un ácido nucleico de la presente divulgación se expresa en células de mamífero utilizando un vector de expresión en mamífero. Los ejemplos de vectores de expresión en mamíferos incluyen pCDM8 (Seed, 1987, Nature 329: 840) y pMT2PC (Kaufman *et al.*, 1987, EMBO J. 6: 187-195). Cuando se utiliza en células de mamífero, las funciones de control del vector de expresión a menudo están proporcionadas por elementos reguladores víricos. Por ejemplo, los promotores habitualmente utilizados se obtienen de polioma, adenovirus 2, citomegalovirus y virus de simio 40. Para otros sistemas de expresión adecuados tanto para células procariontas como eucariotas, véanse los capítulos 16 y 17 de Sambrook *et al.*, citado anteriormente.

En otra realización, el vector de expresión en mamífero recombinante tiene la capacidad de dirigir la expresión del ácido nucleico en un tipo celular particular (por ejemplo, se utilizan elementos reguladores específicos de tejidos para expresar el ácido nucleico). Los elementos reguladores específicos de tejidos son conocidos en la técnica. Los ejemplos no limitativos de promotores específicos de tejido adecuados incluyen el promotor de la albúmina (específico de hígado; Pinkert *et al.*, 1987, Genes Dev. 1: 268-277), promotores específico linfoides (Calame e Eaton, 1988, Adv. Immunol. 43: 235-275), tales como promotores de receptores de linfocitos T (Winoto y Baltimore, 1989, EMBO J. 8: 729-733) e inmunoglobulinas (Banerji *et al.*, 1983, Cell 33: 729-740; Queen y Baltimore, 1983, Cell 33: 741-748), promotores específicos de neurona (por ejemplo, el promotor del neurofilamento; Byrne y Ruddle, 1989, Proc. Natl. Acad. Sci. USA 86: 5473-5477), promotores específicos de páncreas (Edlund *et al.*, 1985, Science 230: 912-916) y promotores específicos de glándula mamaria (por ejemplo, promotor del suero de la leche; Patente de Estados Unidos N.º 4.873.316 y Solicitud Europea N.º de Publicación 264.166). También están abarcados promotores regulados durante desarrollo, por ejemplo los promotores murinos *hox* (Kessel y Gruss, 1990, Science 249: 374-379) y el promotor de la  $\alpha$ -fetoproteína (Camper y Tilghman, 1989, Genes Dev. 3: 537-546).

La divulgación proporciona adicionalmente un vector de expresión recombinante que comprende una molécula de ADN de la presente divulgación clonada en el vector de expresión en una orientación antisentido. Es decir, la molécula de ADN está unida operativamente a una secuencia reguladora de una manera que permite la expresión (mediante la transcripción de la molécula de ADN) de una molécula de ARN que es antisentido para el ARNm que codifica un polipéptido de la presente divulgación. Pueden escogerse secuencias reguladoras unidas operativamente a un ácido nucleico clonado en orientación antisentido, que dirijan la expresión continua de la molécula de ARN antisentido en una diversidad de tipos celulares, por ejemplo promotores y/o potenciadores víricos, o pueden escogerse secuencias reguladoras que dirigen la expresión constitutiva, específica de tejido o específica de tipo celular de un ARN antisentido. El vector de expresión antisentido puede estar en forma de un plásmido recombinante, fagémido o virus atenuado, en el cual se producen ácidos nucleicos antisentido bajo el control de una región reguladora de alta eficacia, cuya actividad puede estar determinada por el tipo celular en que el vector se introduce. Para una discusión de la regulación de la expresión génica utilizando genes antisentido véase Weintraub *et al.*, 1986, Trends in Genetics, Vol. 1(1).

Otro aspecto de la presente divulgación se refiere a células hospedadoras en las que se ha introducido un vector de expresión recombinante de la presente invención. Las expresiones “célula hospedadora” y “célula hospedadora recombinante” se usan indistintamente en el presente documento. Se entiende que tales expresiones se refieren no solo a la célula sujeto particular sino también la progenie o posible progenie de tal célula. Debido a que pueden aparecer determinadas modificaciones en las sucesivas generaciones debido a una mutación o a la influencia del entorno, tal progenie puede, de hecho, no ser idéntica a la célula parental, pero estar aún incluida dentro del ámbito de la expresión como se utiliza en el presente documento.

Una célula hospedadora puede ser cualquier célula procariota (por ejemplo, *E. coli*) o eucariota (por ejemplo, células de insecto, levaduras o células de mamífero).

El ADN de vector puede introducirse en las células procariotas o eucariotas a través de técnicas de transformación o transfección convencionales. Como se usa en el presente documento, los términos “transformación” y “transfección” están destinados a referirse a una diversidad de técnicas reconocidas en la materia para la introducción de un ácido nucleico extraño en una célula hospedadora, incluyendo la coprecipitación con fosfato de calcio o cloruro de calcio, la transfección mediada por DEAE-dextrano, la lipofección o la electroporación. Los métodos adecuados de transformación o transfección de células hospedadoras pueden encontrarse en Sambrook, *et al.* (citado anteriormente) u otros manuales de laboratorio.

Para la transfección estable de células de mamífero se sabe que, dependiendo del vector de expresión y de la técnica de transfección utilizada, solo una pequeña fracción de células puede integrar el ADN ajeno en su genoma. Para identificar y seleccionar estos integrantes, generalmente se introduce en las células hospedadoras un gen que codifica un marcador de selección (por ejemplo, para la resistencia a antibióticos) junto con el gen de interés. Los marcadores de selección ejemplares incluyen los que confieren resistencia a fármacos, tales como G418, higromicina y metotrexato. Las células transfectadas de forma estable con el ácido nucleico introducido pueden identificarse mediante selección con fármacos (por ejemplo, las células que han incorporado el gen marcador de selección sobrevivirán, mientras que las otras células mueren).

Una célula hospedadora de la presente divulgación, tal como una célula hospedadora procariota o eucariota en cultivo, puede utilizarse para producir un polipéptido que corresponde a un marcador de la presente invención. Por consiguiente, la divulgación proporciona adicionalmente métodos de producción de un polipéptido que corresponde a un marcador de la presente invención que utilizan las células hospedadoras de la presente invención. En una realización, el método comprende cultivar la célula hospedadora de la invención (en la cual se ha introducido un vector de expresión recombinante que codifica un polipéptido de la presente invención) en un medio adecuado, de forma que se produzca el marcador. En otra realización, el método comprende adicionalmente aislar el polipéptido marcador del medio o de la célula hospedadora.

Las células hospedadoras de la presente divulgación también pueden utilizarse para producir animales transgénicos no humanos. Por ejemplo, en una realización, una célula hospedadora de la presente divulgación es un ovocito fertilizado o una célula madre embrionaria en la que se han introducido secuencias que codifican un polipéptido que corresponde a un marcador de la presente invención. Tales células hospedadoras pueden utilizarse después para crear animales transgénicos no humanos en los que se han introducido secuencias exógenas que codifican una proteína marcadora de la presente invención en su genoma o animales recombinantes homólogos en los que se ha modificado un gen (o genes) endógeno que codifica un polipéptido que corresponde a un marcador de las secuencias de la presente invención. Tales animales son útiles para estudiar la función y/o actividad del polipéptido que corresponde al marcador, para identificar y/o evaluar moduladores de actividad de polipéptidos, así como en un análisis preclínico de las moléculas terapéuticas o diagnósticas, para el descubrimiento o evaluación de marcadores, por ejemplo, el descubrimiento o evaluación de marcadores terapéuticos y diagnósticos o de criterio indirecto de valoración de la eficacia y especificidad de fármacos.

Como se usa en el presente documento, un “animal transgénico” es un animal no humano, por ejemplo, un mamífero, tal como un roedor, por ejemplo, una rata o ratón, en el que una o más de las células del animal incluyen un transgén. Otros ejemplos de animales transgénicos incluyen primates no humanos, ovejas, perros, vacas, cabras, pollos, anfibios, etc. Un transgén es ADN exógeno que está integrado en el genoma de una célula a partir de la cual se desarrolla el animal transgénico y que permanece en el genoma del animal maduro, dirigiendo de este modo la expresión de un producto génico codificado en uno o más tipos celulares o tejidos del animal transgénico. Como se usa en el presente documento, un “animal recombinante homólogo” es un animal no humano, tal como un mamífero, por ejemplo, un ratón, en el cual se ha modificado un gen endógeno por recombinación homóloga entre el gen endógeno y una molécula de ADN exógena introducida en una célula del animal, por ejemplo, una célula embrionaria del animal, antes del desarrollo del animal. Los animales transgénicos también incluyen animales transgénicos inducibles, tales como los descritos en, por ejemplo, Chan I. T., *et al.* (2004) *J Clin Invest.* 113(4): 528 - 38 y Chin L. *et al.* (1999) *Nature* 400(6743): 468-72.

El animal transgénico de la presente divulgación puede crearse introduciendo un ácido nucleico que codifica un polipéptido que corresponde a un marcador de la presente invención en los pronúcleos masculinos de un ovocito fertilizado, por ejemplo, por microinyección o infección retroviral y permitiendo que el ovocito se desarrolle en un

animal de acogida hembra seudopreñada. También pueden estar incluidas en el transgén secuencias intrónicas y señales de poliadenilación para aumentar la eficacia de la expresión del transgén. Una secuencia (o secuencias) reguladora específica de tejido puede estar unida operativamente al transgén para dirigir la expresión del polipéptido de la presente invención a células particulares. Los métodos para la generación de animales transgénicos a través de la manipulación de embriones y de microinyección, en particular animales tales como ratones, se han vuelto una técnica convencional en la materia y se describen, por ejemplo, en las Patentes de Estados Unidos n.º 4.736.866 y 4.870.009, la Patente de Estados Unidos n.º 4.873.191 y en Hogan, *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1986. Se utilizan métodos similares para la producción de otros animales transgénicos. Un animal fundador transgénico puede identificarse a base de la presencia del transgén en su genoma y/o la expresión de ARNm que codifica el transgén en los tejidos o células de los animales. Después, un animal fundador transgénico puede utilizarse para criar animales adicionales que porten el transgén. Además, los animales transgénicos que portan el transgén pueden adicionalmente criarse para obtener otros animales transgénicos que portan otros transgenes.

Para crear un animal recombinante homólogo, se prepara un vector que contiene al menos una porción de un gen que codifica un polipéptido que corresponde a un marcador de la presente invención en el cual se ha introducido una delección, adición o sustitución para, de este modo, modificar, por ejemplo, alterar de forma funcional el gen. En otra realización, el vector se diseña de forma que, tras la recombinación homóloga, el gen endógeno está funcionalmente alterado (por ejemplo, ya no codifica una proteína funcional; también denominado como vector "genosuprimido"). Como alternativa, el vector puede diseñarse de forma que, tras la recombinación homóloga, el gen endógeno está mutado o modificado de otra forma pero aún codifica la proteína funcional (por ejemplo, la región reguladora cadena arriba puede modificarse para, de este modo, alterar la expresión de la proteína endógena). En el vector de recombinación homóloga, la porción modificada del gen está flanqueada en sus extremos 5' y 3' por un ácido nucleico adicional del gen para permitir que se produzca la recombinación homóloga entre el gen exógeno que porta el vector y un gen endógeno en una célula madre embrionaria. Las secuencias de ácido nucleico flanqueantes adicionales son de suficiente longitud para una recombinación homóloga satisfactoria con el gen endógeno.

Normalmente, están incluidas en el vector varias kilobases de ADN flanqueante (tanto en el extremo 5' como en el 3') (véase, por ejemplo, Thomas y Capecchi, 1987, *Cell* 51: 503 para una descripción de vectores de recombinación homóloga). El vector se introduce en una línea de células madre embrionarias (por ejemplo, mediante electroporación) y se seleccionan las células en las cuales el gen introducido ha recombinado de forma homóloga con el gen endógeno (véase, por ejemplo, Li *et al.*, 1992, *Cell* 69: 915). Las células seleccionadas se inyectan después en un blastocito de un animal (por ejemplo, un ratón) para formar quimeras de agregación (véase, por ejemplo, Bradley, *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, Ed., IRL, Oxford, 1987, pág. 113-152). Después, puede implantarse un embrión quimérico en un animal de acogida hembra seudopreñada y llevar el embrión a término. La progenie que alberga el ADN recombinado de forma homóloga en sus células germinales puede utilizarse para criar animales en los que todas las células del animal contienen el ADN recombinado de forma homóloga por transmisión del transgén en la estirpe germinal. Los métodos para construir vectores de recombinación homóloga y animales recombinantes homólogos se describen adicionalmente en Bradley (1991) *Current Opinion in Bio/Technology* 2: 823-829 y en las Publicaciones PCT n.º WO 90/11354, WO 91/01140, WO 92/0968 y WO 93/04169.

En otra realización, pueden producirse animales no humanos transgénicos que contengan sistemas seleccionados que permitan la expresión regulada del transgén. Un ejemplo de tal sistema es el sistema de recombinasa *cre/loxP* del bacteriófago P1. Para una descripción del sistema de recombinasa *cre/loxP*, véase, por ejemplo, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Otro ejemplo de un sistema de recombinasa es el sistema de recombinasa FLP de *Saccharomyces cerevisiae* (O'Gorman *et al.*, 1991, *Science* 251: 1351-1355). Si se utiliza un sistema de recombinasa *cre/loxP* para regular la expresión del transgén, se requieren animales que contengan transgenes que codifican la recombinasa *Cre* y una proteína seleccionada. Tales animales pueden proporcionarse a través de la construcción de animales "doble" transgénicos, por ejemplo, apareando dos animales transgénicos, uno conteniendo un transgén que codifica una proteína seleccionada y el otro conteniendo un transgén que codifica una recombinasa.

Los clones de los animales transgénicos no humanos descritos en el presente documento también pueden producirse de acuerdo con los métodos descritos en Wilmut *et al.* (1997) *Nature* 385: 810-813 y las Publicaciones PCT n.º WO 97/07668 y WO 97/07669.

#### V. Kits ejemplares

Un kit es cualquier manufactura (por ejemplo, un envase o recipiente) que comprende al menos un reactivo, por ejemplo, una sonda, para detectar de forma específica un marcador de la presente invención, promocionando, distribuyendo o vendiendo la manufactura como una unidad para realizar los métodos de la presente invención. Cuando las composiciones, kits y métodos de la presente invención se utilizan para llevar a cabo los métodos de la presente invención, las mutaciones y/o productos génicos del gen ALK (por ejemplo, los marcadores expuestos de la Tabla 1) de la presente invención pueden seleccionarse de forma que se obtiene un resultado positivo en al menos aproximadamente el 20 %, al menos aproximadamente el 40 %, al menos aproximadamente el 60 %, al menos

aproximadamente el 80 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 99 % o en el 100 % de los sujetos aquejados de cáncer de la correspondiente fase, grado, tipo histológico o naturaleza benigna/premaligna/maligna. En determinadas realizaciones, el marcador o panel de marcadores de la presente invención puede seleccionarse de forma que se obtiene un VPP (valor predictivo positivo) mayor de aproximadamente el 10 % para la población general (por ejemplo, acoplado con una especificidad de ensayo mayor del 99,5 %).

Cuando en las composiciones, kits y métodos de la presente invención se utiliza una pluralidad de mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) de la presente invención, la cantidad, estructura y/o actividad de cada marcador o nivel de expresión, o número de copias puede compararse con la cantidad, estructura y/o actividad normal de cada uno de la pluralidad de marcadores o nivel de expresión, o número de copias, en muestras no cancerosas del mismo tipo, ya sea en una única mezcla de reacción (es decir, utilizando reactivos, tales como distintas sondas fluorescentes, para cada marcador) o en mezclas de reacción individuales que corresponden a una o más de las mutaciones y/o productos genéticos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1). Si se utiliza una pluralidad de mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1), entonces pueden utilizarse o identificarse 2, 3, 4, 5, 6, 7, 8, 9, 10 o más marcadores individuales.

La invención incluye métodos para ensayar las células cancerosas en una muestra (por ejemplo, una muestra de tejido archivada o una muestra obtenida del sujeto). Estos métodos son sustancialmente los mismos que los descritos anteriormente, excepto en que, cuando es necesario, los métodos se adaptan para su uso con determinados tipos de muestras.

La divulgación incluye un kit para evaluar la presencia de células cancerosas que tienen o es probable que tengan una respuesta reducida a los inhibidores de ALK (por ejemplo, en una muestra tal como una muestra de un sujeto). El kit puede comprender uno o más reactivos que tengan la capacidad de identificar mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) de la presente invención, por ejemplo, uniéndose de forma específica a un ácido nucleico o polipéptido que corresponde a las mutaciones y/o producto génico del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) de la presente invención. Los reactivos adecuados para la unión con un polipéptido que corresponde a un marcador de la presente invención incluyen anticuerpos, derivados de anticuerpos, fragmentos de anticuerpos y similares. Los reactivos adecuados para la unión con un ácido nucleico (por ejemplo, un ADN genómico, un ARNm, un ARNm cortado y empalmado, un ADNc o similar) incluyen ácidos nucleicos complementarios. Por ejemplo, los reactivos de ácido nucleico pueden incluir oligonucleótidos (marcados o no marcados) fijados a un sustrato, oligonucleótidos marcados no unidos a un sustrato, parejas de cebadores de PCR, sondas de baliza molecular y similares. En algunas realizaciones, los kits pueden comprender reactivos útiles para realizar los métodos descritos en el presente documento, tales como los que comprenden al menos una pareja de cebadores que reconocen y e hibridan con tramos de ácido nucleico que rodea al menos un tramo de ácido nucleico que comprende al menos una mutación enumerada en la Tabla 1 y medios para detectar la presencia de dicha mutación en el ácido nucleico diana amplificado.

El kit de la presente divulgación puede opcionalmente comprender componentes adicionales útiles para realizar los métodos de la presente invención. A modo de ejemplo, el kit puede comprender líquidos (por ejemplo, tampón SSC) adecuados para el apareamiento de ácidos nucleicos complementarios o para la unión de un anticuerpo con una proteína con la que se unen de forma específica uno o más compartimentos de muestra, un material instructivo que describe la ejecución de un método de la presente invención, una muestra de células normales, una muestra de células cancerosas y similares.

Un kit de la presente divulgación puede comprender un reactivo útil para determinar el nivel de proteína o la actividad de proteína de un marcador. En otra realización, un kit de la presente divulgación puede comprender un reactivo para determinar el estado de metilación de un marcador o puede comprender un reactivo para determinar la modificación de la estructura de un marcador, por ejemplo, la presencia de una mutación.

## VI. Medicina predictiva

La presente invención también se refiere al campo de la medicina predictiva, en la cual se utilizan pruebas de diagnóstico, farmacogenómica y ensayos clínicos de control para fines predictivos para, de este modo, tratar un individuo de forma profiláctica. Por consiguiente, un aspecto de la presente invención se refiere a ensayos para determinar la cantidad, estructura y/o actividad de polipéptidos o ácidos nucleicos que corresponden a uno o más marcadores de la presente invención, para determinar si un individuo con cáncer o en riesgo de desarrollar cáncer tendrá más probabilidades de responder a la terapia mediada por inhibidores de ALK.

Por consiguiente, en un aspecto, la invención se dirige a un método para determinar si un sujeto con cáncer es probable que responda a un tratamiento con un agente inhibidor de la ALK. En otro aspecto, la invención se dirige a un método para predecir un suceso significativo en la evolución temporal de la enfermedad. En aún otro aspecto, el método se dirige a un método para predecir la probabilidad de un suceso significativo en la evolución de la enfermedad. En determinadas realizaciones, el método comprende detectar un biomarcador o combinación de

biomarcadores asociados con la respuesta al tratamiento con un agente inhibidor de ALK (por ejemplo, mutaciones de ALK), como se describe en el presente documento y determinar si es probable que el sujeto responda al tratamiento con el agente inhibidor de ALK.

- 5 En algunas realizaciones, los métodos implican la exploración citogenética de una muestra de tejido biológico de un paciente que se ha diagnosticado con o se sospecha que tiene cáncer (por ejemplo, presenta síntomas de cáncer) para detectar mutaciones de ALK (por ejemplo, las enumeradas en la Tabla 1).

10 Los resultados del método de exploración y la interpretación de los mismos son predictivos de la respuesta del paciente al tratamiento con los agentes inhibidores de ALK (por ejemplo, PF-02341066 y/o PDD). De acuerdo con la presente invención, la presencia de una mutación de ALK es indicativa de que el tratamiento con los agentes inhibidores de ALK (por ejemplo, PF-02341066 y/o PDD) proporcionará un beneficio terapéutico potenciado frente a las células cancerosas con respecto a los pacientes que no tienen una mutación de ALK.

15 En una realización, los métodos de la presente invención comprenden poner en contacto una muestra de ADN, por ejemplo, una muestra que contiene ADN de estirpe germinal y/o somático, tal como una muestra cromosómica, obtenida de células aisladas del paciente con sondas de polinucleótido que son específicas para e hibridan en condiciones rigurosas con el ADN genómico de regiones cromosómicas asociadas con anomalías citogenéticas (por ejemplo, mutaciones de ALK descritas en el presente documento), para determinar la presencia o ausencia de una o más anomalías (por ejemplo, mutaciones) en las células del paciente. Los resultados del análisis son predictivos de la respuesta probable del paciente al tratamiento con agentes terapéuticos, en particular agentes que inhiben la ALK (por ejemplo, PF-02341066 y/o PDD).

20 En otra realización, se mide una evolución temporal determinando el tiempo entre sucesos significativos en el curso de la enfermedad de un paciente, en el que la medición es predictiva de si un paciente tiene un evolución temporal extensa. En otra realización, el suceso significativo es la progresión desde el diagnóstico primario hasta el fallecimiento. En otra realización, el suceso significativo es la progresión desde el diagnóstico primario hasta la enfermedad metastásica. En otra realización, el suceso significativo es la progresión desde el pronóstico primario hasta la recaída. En otra realización, el suceso significativo es la progresión desde la enfermedad metastásica hasta el fallecimiento. En otra realización, el suceso significativo es la progresión desde la recaída hasta el fallecimiento. En determinadas realizaciones, la evolución temporal se mide con respecto a la tasa de supervivencia global, tiempo transcurrido hasta la progresión y/o utilizando el RECIST u otros criterios de respuesta.

25 En determinadas realizaciones, se crea una medida predeterminada dividiendo las muestras de pacientes en al menos dos subgrupos de pacientes. En determinadas realizaciones, la cantidad de subgrupos es de dos, de forma que la muestra del paciente se divide en un subgrupo de pacientes que tienen una mutación (o mutaciones) de ALK y un subgrupo que no tiene una mutación (o mutaciones) de ALK. En determinadas realizaciones, el estado de mutación de ALK en el sujeto se compara con el subgrupo que tiene o el que no tiene una mutación (o mutaciones) de ALK; si el paciente tiene una mutación (o mutaciones) de ALK, entonces es improbable que el paciente responda a un inhibidor de ALK (por ejemplo, PF-02341066 y/o PDD) y/o es probable que el paciente tenga una evolución temporal larga. En determinadas realizaciones, el número de subgrupos es mayor de dos, incluyendo, pero sin limitación, tres subgrupos, cuatro subgrupos, cinco subgrupos y seis subgrupos, dependiendo de la estratificación de la eficacia inhibidora de ALK predicha correlacionada con mutaciones de ALK particulares. En determinadas realizaciones, la probabilidad de responder se mide con respecto a la tasa de supervivencia global, el tiempo transcurrido hasta la progresión y/o utilizando los criterios RECIST. En determinadas realizaciones, el inhibidor de ALK es PF-02341066 y/o PDD.

30 En otro aspecto, la invención se dirige a un método para determinar si un sujeto con un cáncer positivo para una mutación de ALK es probable que responda al tratamiento con un agente inhibidor de ALK (por ejemplo, PF-02341066 y/o PDD) y/o si la evolución temporal de la enfermedad es extensa. En otro aspecto, la invención se dirige a un método para predecir la evolución temporal de la enfermedad en un sujeto con un cáncer positivo para una mutación de ALK. En otro aspecto, la invención se dirige a un método para predecir la probabilidad de un suceso significativo en un sujeto con un cáncer positivo para una mutación de ALK.

### 55 1. Métodos para la detección de mutaciones de ALK

Los métodos para evaluar las mutaciones y/o productos génicos del gen de ALK (por ejemplo, los marcadores expuestos en la Tabla 1) son bien conocidos por los expertos en la materia, incluyendo los ensayos basados en hibridación. Por ejemplo, un método de evaluación del número de copias de un ácido nucleico codificante en una muestra implica una transferencia de Southern. En una transferencia de Southern, el ADN genómico (normalmente fragmentado y separado en un gel electroforético) se hibrida con una sonda específica para la región diana. La comparación de la intensidad de la señal de hibridación de la sonda para la región diana con la señal de la sonda de control a partir del análisis del ADN genómico normal (por ejemplo, una porción no amplificada de la misma célula, tejido, órgano o uno relacionado, etc.) proporciona una estimación de la presencia/ausencia y el número de copias relativo del ácido nucleico diana. Como alternativa, puede utilizarse una transferencia Northern para evaluar el

número de copias de un ácido nucleico codificante en una muestra. En una transferencia de Northern, se hibrida ARNm con una sonda específica para la región diana. La comparación de la intensidad de la señal de hibridación procedente de la sonda para la región diana con la señal de la sonda de control del análisis del ARNm normal (por ejemplo, una porción no amplificada de la misma célula, tejido, órgano o uno relacionado, etc.) proporciona una estimación de la presencia/ausencia y el número de copias relativo del ácido nucleico diana.

Un medio alternativo para determinar el número de copias es la hibridación *in situ* (por ejemplo, Angerer (1987) Meth. Enzymol 152: 649). En general, la hibridación *in situ* comprende las siguientes etapas: (1) la fijación del tejido o estructura biológica a analizar; (2) un tratamiento de prehibridación de la estructura biológica para aumentar la accesibilidad del ADN diana y para reducir la unión no específica; (3) la hibridación de la mezcla de ácidos nucleicos con el ácido nucleico en la estructura biológica o tejido; (4) lavados poshibridación para eliminar los fragmentos de ácido nucleico no unidos en la hibridación y (5) la detección de los fragmentos de ácido nucleico hibridados. El reactivo utilizado en cada una de las etapas y las condiciones de uso varían dependiendo de la aplicación particular.

Los ensayos basados en hibridación ejemplares incluyen, pero sin limitación, métodos de "sonda directa" tradicionales tales como transferencias de Southern o hibridación *in situ* (por ejemplo, FISH y FISH más SKY) y métodos de "sonda comparativa" tales como hibridación genómica comparativa (CGH), por ejemplo, CGH basada en ADNc o basada en oligonucleótidos. Los métodos pueden utilizarse en una amplia diversidad de formatos incluyendo, pero sin limitación, métodos de unión a un sustrato (por ejemplo, membrana o vidrio) o estrategias basadas en matrices.

En un aspecto, se utiliza análisis de FISH. Las muestras de células se obtienen de pacientes de acuerdo con métodos bien conocidos en la técnica para analizarlas mediante un método de análisis citogenético apropiado conocido en la técnica, por ejemplo, el método FISH. En una realización, puede realizarse FISH de acuerdo con el sistema Vysis™ (Abbott Molecular).

Se utilizan sondas que contienen segmentos de ADN que son esencialmente complementarios a secuencias de bases de ADN que existen en distintas porciones de los cromosomas. Los ejemplos de sondas útiles de acuerdo con la invención, y el marcaje y la hibridación de sondas con las muestras se describe en dos patentes de Estados Unidos para Vysis, Inc. las Patentes de Estados Unidos n.º 5.491.224 y 6.277.569 para Bittner, *et al.*

Las sondas cromosómicas son normalmente de aproximadamente 50 a aproximadamente 10<sup>5</sup> nucleótidos de longitud. Las sondas más largas normalmente comprenden fragmentos más pequeños de aproximadamente 100 a aproximadamente 500 nucleótidos de longitud. Las sondas que hibridan con ADN centromérico y ADN específico de locus están disponibles en el mercado, por ejemplo, de Vysis, Inc. (Downers Grove, Ill.), Molecular Probes, Inc. (Eugene, Oreg.) o de Cytocell (Oxfordshire, RU). Como alternativa, las sondas pueden fabricarse de forma no comercial a partir de ADN cromosómico o genómico mediante técnicas convencionales. Por ejemplo, las fuentes de ADN que pueden utilizarse incluyen ADN genómico, secuencias de ADN clonadas, híbridos de células somáticas que contienen un, o una parte de un cromosoma (por ejemplo, un cromosoma humano) junto con el complemento cromosómico normal del huésped, y cromosomas purificados por citometría de flujo o microdissección. La región de interés puede aislarse a través de clonación o por amplificación específica de sitio a través de reacción en cadena de la polimerasa (PCR). Véase, por ejemplo, Nath y Johnson, Biotechnic Histochem., 1998, 73(1): 6-22, Wheelless *et al.*, Cytometry 1994, 17: 319-326 y la Patente de Estados Unidos N.º 5.491.224.

Las sondas a utilizar hibridan con una región específica de un cromosoma para determinar si está presente en esta región una anomalía citogenética. Un tipo de anomalía citogenética es una delección. Aunque las delecciones pueden ser de uno o más cromosomas completos, las delecciones normalmente implican la pérdida de parte de uno o más cromosomas. Si toda la región de un cromosoma que está contenida en una sonda está delecionada de una célula, normalmente no se producirá la hibridación de esa sonda con el ADN de la célula y no estará presente una señal en ese cromosoma. Si la región de un cromosoma que está contenido parcialmente dentro de una sonda está delecionada de una célula, la hibridación de esa sonda con el ADN de la célula aún podría producirse, pero estará presente menos de una señal. Por ejemplo, la pérdida de una señal se compara con la hibridación de la sonda a ADN de las células de control que no contienen las anomalías genéticas que las sondas pretenden detectar. En algunas realizaciones, se cuentan al menos 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200 o más células para la presencia de la anomalía citogenética.

Las anomalías citogenéticas a detectar pueden incluir, pero sin limitación, translocaciones no recíprocas, inversiones intracromosómicas, mutaciones puntuales, delecciones, cambios en el número de copias de genes, cambios del nivel de expresión génica y mutaciones en la estirpe germinal. En particular, un tipo de anomalía citogenética es una duplicación. Las duplicaciones pueden ser de cromosomas completos o de regiones más pequeñas que un cromosoma completo. Si la región de un cromosoma que está contenida en una sonda está duplicada en una célula, la hibridación de la sonda con el ADN de la célula normalmente producirá al menos una señal adicional en comparación con el número de señales presentes en las células de control sin anomalía en la región cromosómica contenida en la sonda. No obstante, puede utilizarse cualquiera de las sondas que detectan el cromosoma humano 2p23 o un ortólogo del mismo, o cualquier región cromosómica que comprende una translocación con el gen de ALK

de 2p23 o un ortólogo del mismo. Las sondas adecuadas son bien conocidas en la técnica (por ejemplo, las disponibles de Vysis, Inc. (Downers Grove, IL).

5 Las sondas cromosómicas están marcadas de forma que puede detectarse la región cromosómica con la que hibridan. Normalmente, las sondas están marcadas directamente con un fluoróforo, una molécula orgánica que fluoresce tras la absorción de luz de una longitud de onda más baja/energía más alta. El fluoróforo permite visualizar la sonda sin una molécula de detección secundaria. Después de unir de forma covalente un fluoróforo a un nucleótido, el nucleótido puede incorporarse directamente en la sonda con técnicas convencionales tales como marcaje por relleno, unión aleatoria y marcaje por PCR. Como alternativa, pueden transaminarse nucleótidos de desoxicitidina con un enlazador dentro de la sonda. Después, el fluoróforo se une de forma covalente a los nucleótidos de desoxicitidina transaminados. Véase la Patente de Estados Unidos N.º 5.491.224 (incorporada como referencia).

15 La Patente de Estados Unidos N.º 5.491.224 describe el marcaje de sondas como una serie de restos citosina que tienen un marcador fluorescente unido de forma covalente a los mismos. La cantidad de bases citosina marcadas de forma fluorescente es suficiente para generar una señal fluorescente detectable aunque los segmentos de ADN individuales así marcados conservan esencialmente sus propiedades de unión complementaria (hibridación) específicas con respecto al cromosoma o región de cromosoma a detectar. Tales sondas se fabrican tomando el segmento de sonda de ADN no marcado, transaminando con un grupo enlazador una serie de nucleótidos desoxicitidina en el segmento, enlazando de forma covalente un marcador fluorescente a al menos una porción de las bases de desoxicitidina transaminadas.

25 Las sondas también pueden marcarse mediante mercado por relleno, marcaje con unión aleatorio o marcaje por PCR. El marcaje se realiza utilizando ya sea nucleótidos fluorescentes (directo) o marcados con hapteno (indirecto). Los ejemplos representativos no limitativos de marcadores incluyen: AMCA-6-dUTP, CascadeBlue-4-dUTP, fluoresceína-12-dUTP, rodamina-6-dUTP, rojo texas-6-dUTP, Cy3-6-dUTP, Cy5-dUTP, biotina(BIO)-11-dUTP, digoxigenina(DIG)-11-dUTP o dinitrofenil (DNP)-11-dUTP.

30 Las sondas también pueden marcarse de forma indirecta con biotina o digoxigenina, o marcarse con isótopos radioactivos tales como <sup>32</sup>P y <sup>3</sup>H, aunque después son necesarias moléculas de detección secundarias o un procesamiento adicional para visualizar las sondas. Por ejemplo, una sonda marcada con biotina puede detectarse con avidina conjugada a un marcador detectable. Por ejemplo, la avidina puede conjugarse a un marcador enzimático tal como fosfatasa alcalina o peroxidasa de rábano picante. Los marcadores enzimáticos pueden detectarse en reacciones colorimétricas convencionales utilizando un sustrato y/o un catalizador para la enzima. El catalizador para la fosfatasa alcalina incluye 5-bromo-4-cloro-3-indolilfosfato y nitroazul de tetrazolio. Puede utilizarse diaminobenzoato como un catalizador de la peroxidasa de rábano picante.

40 Las sondas también pueden prepararse de forma que un marcador fluorescente u otro, no sea parte del ADN antes o durante la hibridación, y se añade tras la hibridación para detectar el cebador hibridado a un cromosoma. Por ejemplo, pueden utilizarse sondas que tengan moléculas antigénicas incorporadas en el ADN. Tras la hibridación, estas moléculas antigénicas se detectan utilizan anticuerpos específicos reactivos con las moléculas antigénicas. Tales anticuerpos pueden ellos mismos incorporar un fluorocromo o pueden detectarse utilizando un segundo anticuerpo con un fluorocromo unido.

45 Sin embargo, tratado o modificado, el ADN de sonda normalmente se purifica para eliminar los productos que no han reaccionado, residuales (por ejemplo, moléculas de fluorocromo incorporadas en el ADN) antes del uso en hibridación.

50 Antes de la hibridación, las sondas cromosómicas se desnaturalizan de acuerdo con métodos bien conocidos en la técnica. En general, las etapas de hibridación comprenden añadir un exceso de ADN bloqueante a la composición de sonda marcada, poner en contacto la composición de sonda bloqueada en condiciones de hibridación con la región cromosómica a detectar, por ejemplo, en un portaobjetos en donde el ADN se ha desnaturalizado, lavar la sonda no hibridada y detectar la unión de la composición de la sonda al cromosoma o región cromosómica.

55 Las sondas se hibridan o aparean al ADN cromosómico en condiciones de hibridación. "Condiciones de hibridación" son las condiciones que facilitan el apareamiento entre una sonda y el ADN cromosómico diana. Dado que el apareamiento de distintas sondas variará dependiendo de la longitud de la sonda, de la concentración de bases y similares, el apareamiento se facilita variando la concentración de la sonda, la temperatura de hibridación, la concentración de sales y otros factores bien conocidos en la técnica.

60 Las condiciones de hibridación se facilitan variando las concentraciones, las composiciones de las bases, la complejidad y la longitud de las sondas, así como las concentraciones de sales, las temperaturas y la duración de la incubación. Por ejemplo, las hibridaciones *in situ* normalmente se realizan en un tampón de hibridación que contiene SSC 1-2x, formamida al 50-65 % y ADN bloqueante para suprimir la hibridación no específica. En general, las condiciones de hibridación, como se describe anteriormente, incluye temperaturas de aproximadamente 25 °C a

aproximadamente 55 °C, y duraciones de incubación de aproximadamente de 0,5 horas a aproximadamente 96 horas.

La unión no específica de las sondas cromosómicas al ADN fuera de la región diana puede eliminarse mediante una serie de lavados. La temperatura y la concentración de sales en cada lavado se varían para controlar la rigurosidad de los lavados. Por ejemplo, para condiciones de alta rigurosidad, los lavados pueden llevarse a cabo de aproximadamente 65 °C a aproximadamente 80 °C, utilizando SSC de 0,2x a aproximadamente 2x y aproximadamente el 0,1 % a aproximadamente el 1 % de un detergente no iónico, tal como Nonidet P-40 (NP40). La rigurosidad puede reducirse disminuyendo la temperatura de los lavados o aumentando las concentraciones de sales en los lavados. En tales aplicaciones es necesario bloquear la capacidad de hibridación de las secuencias repetitivas. Por lo tanto, en algunas realizaciones, se utiliza ARNt, ADN genómico humano o ADN Cot-I para bloquear la hibridación no específica.

Tras el lavado, se permite que el portaobjetos escurra y se seque al aire, después se aplican al portaobjetos un medio de montaje, una contratinción tal como DAPI y un cubreobjetos. Los portaobjetos pueden visualizarse inmediatamente o almacenarse a -20 °C antes del examen.

Para las sondas fluorescentes utilizadas en las técnicas de hibridación de fluorescencia *in situ* (FISH), la fluorescencia puede visualizarse con un microscopio de fluorescencia equipado con un filtro apropiado para cada fluoróforo, o utilizando conjuntos de filtro de paso de banda doble o triple para observar múltiples fluoróforos. Véase, por ejemplo, la Patente de Estados Unidos N.º 5.776.688. Como alternativa, pueden utilizarse técnicas tales como la citometría de flujo para examinar el patrón de hibridación de las sondas cromosómicas. Puede utilizarse FISH para detectar el número de copias de cromosomas o el reordenamiento de regiones de cromosomas. Estas sondas hibridan o se unen al ADN complementario y, debido a que están marcadas con etiquetas fluorescentes, permiten a los investigadores observar el emplazamiento de las secuencias de ADN utilizando un microscopio de fluorescencia. A diferencia de la mayoría de las otras técnicas utilizadas para estudiar cromosomas, que requieren que las células estén dividiéndose de forma activa, la FISH también puede realizarse en células que no están en división, haciéndola un procedimiento altamente versátil. Por lo tanto, la FISH puede realizarse utilizando células en interfase o células en metafase del ciclo de división celular. Muchas de las técnicas implicadas en el análisis por FISH se describen en la Patente de Estados Unidos N.º 5.447.841 por Gray y Pinkel.

Los resultados de FISH pueden interpretarse con referencia a células de control que se sabe que no contienen la anomalía citogenética específica para la que la sonda está diseñada para detectar. El patrón de hibridación por FISH de la sonda con el ADN de las células de control se compara con la hibridación de la misma sonda con el ADN de las células que se están analizando o ensayando para la anomalía citogenética específica. Cuando se diseña una sonda para detectar una delección de un cromosoma o región cromosómica, normalmente hay menos hibridación de la sonda con el ADN de las células que se están analizando que con el de las células de control. Normalmente, en las células analizadas hay una ausencia de una señal de sonda, lo que es indicativo de una pérdida de la región de un cromosoma con la que normalmente hibrida la sonda. Cuando se diseña una sonda para detectar una duplicación o adición cromosómica, normalmente hay más hibridación de la sonda con el ADN de las células que se están analizando que con el de las células de control. Normalmente, hay una adición de la señal de sonda en las células analizadas, lo que es indicativo de la presencia de una región cromosómica con la que la sonda normalmente hibrida adicional.

En los métodos de CGH, se marca una primera colección de ácidos nucleicos (por ejemplo, de una muestra, por ejemplo, de un posible tumor) con un primer marcador, mientras se marca una segunda colección de ácidos nucleicos (por ejemplo, un control, por ejemplo, de una célula/tejido sano) con un segundo marcador. La proporción de hibridación de los ácidos nucleicos se determina por la proporción de los dos marcadores (el primero y el segundo) que se unen a cada fibra en la matriz. Cuando hay delecciones o multiplicaciones cromosómicas, se detectará las diferencias en la proporción de las señales de los dos marcadores y la proporción proporcionará una medida del número de copias. El CGH basado en matrices puede realizarse también con marcaje de un único color (en lugar de marcar el control y la posible muestra tumoral con dos tintes distintos y mezclarlos antes de la hibridación, lo que producirá una proporción debido a la hibridación competitiva de las sondas en las matrices). En el CGH de color único, los controles se marcan e hibridan a una matriz y se leen las señales absolutas, y la posible muestra tumoral se marca e hibrida a una segunda matriz (con un contenido idéntico) y se leen las señales absolutas. Se calcula la diferencia en el número de copias a base de las señales absolutas de las dos matrices. Los protocolos de hibridación adecuados para su uso con los métodos de la presente invención se describen, por ejemplo, en Albertson (1984) EMBO J. 3: 1227-1234; Pinkel (1988) Proc. Natl. Acad. Sci. USA 85: 9138-9142; EPO N.º Pub. 430.402; Methods in Molecular Biology, Vol. 33: *In situ* Hybridization Protocols, Choo, ed., Humana Press, Totowa, N.J. (1994), etc. En una realización, se utiliza el protocolo de hibridación de Pinkel, *et al.* (1998) Nature Genetics 20: 207-211 o de Kallioniemi (1992) Proc. Natl Acad Sci USA 89: 5321-5325 (1992). El CGH basado en matrices se describe en la Patente de Estados Unidos N.º 6.455.258.

En aún otra realización, pueden utilizarse ensayos basados en amplificación para medir la presencia/ausencia y el número de copias. En tales ensayos basados en amplificación, las secuencias de ácido nucleico actúan como un molde en una reacción de amplificación (por ejemplo, reacción en cadena de la polimerasa (PCR)). En una

amplificación cuantitativa, la cantidad de producto de amplificación será proporcional a la cantidad de molde en la muestra original. La comparación con controles apropiados, por ejemplo, tejido sano, proporciona una medida del número de copias.

5 Los métodos de amplificación "cuantitativa" son bien conocidos para los expertos en la materia. Por ejemplo, la PCR cuantitativa implica la coamplificación simultánea de una cantidad conocida de una secuencia de control utilizando los mismos cebadores. Esto proporciona un patrón interno que puede utilizarse para calibrar la reacción de PCR. Los protocolos detallados para la PCR cuantitativa se proporcionan en Innis, *et al.* (1990) *PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc. N.Y.). La medición del número de copias de ADN en locus de microsatélites utilizando el análisis por PCR cuantitativa se describe en Ginzonger, *et al.* (2000) *Cancer Research* 60: 5405-5409. La secuencia de ácido nucleico conocida de los genes es suficiente para permitir que un experto en la técnica seleccione de forma rutinaria cebadores para amplificar cualquier porción del gen. En los métodos de la presente invención también puede utilizarse PCR cuantitativa fluorogénica. En la PCR cuantitativa fluorogénica, la cuantificación es a base de la cantidad de señales de fluorescencia, por ejemplo, TaqMan y verde sybr.

15 Otros métodos de amplificación adecuados incluyen, pero sin limitación, reacción en cadena de la ligasa (LCR) (véase Wu y Wallace (1989) *Genomics* 4: 560, Landegren, *et al.* (1988) *Science* 241: 1077 y Barringer *et al.* (1990) *Gene* 89: 117), amplificación por transcripción (Kwoh, *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173), replicación de secuencia autosostenida (Guatelli, *et al.* (1990) *Proc. Nat. Acad. Sci. USA* 87: 1874), PCR puntual y PCR de adaptador enlazador, etc.

También puede utilizarse para identificar regiones de amplificación o delección el mapeo por pérdida de heterocigosidad (LOH) Wang, Z.C., *et al.* (2004) *Cancer Res* 64(1): 64-71; Seymour, A. B., *et al.* (1994) *Cancer Res* 54, 2761-4; Hahn, S. A., *et al.* (1995) *Cancer Res* 55, 4670-5; Kimura, M., *et al.* (1996) *Genes Chromosomes Cancer* 17, 88-93).

## 2. Métodos para evaluar la expresión génica

30 También puede ensayarse el nivel de expresión de marcadores. La expresión de un marcador de la presente invención puede evaluarse mediante cualquiera de una amplia diversidad de métodos bien conocidos para detectar la expresión de una molécula transcrita o una proteína. Los ejemplos no limitativos de tales métodos incluyen métodos inmunológicos para la detección de proteínas secretadas, de superficie celular, citoplasmáticas o nucleares, métodos de purificación de proteínas, ensayos de función o actividad de proteínas, métodos de hibridación de ácidos nucleicos, métodos de transcripción inversa de ácidos nucleicos y métodos de amplificación de ácidos nucleicos.

40 En determinadas realizaciones, la actividad de un gen particular está caracterizada por una medida de un transcrito génico (por ejemplo, ARNm), mediante la medida de la cantidad de proteína traducida o mediante una medida de la actividad de un producto génico. La expresión de marcadores puede controlarse en una diversidad de formas, incluyendo la detección de los niveles de ARNm, los niveles de proteína o la actividad de proteína, cualquiera de los cuales puede medirse utilizando técnicas convencionales. La detección puede implicar la cuantificación del nivel de expresión génica (por ejemplo, ADN genómico, ADNc, ARNm, proteína o actividad enzimática) o, como alternativa, puede ser una evaluación cualitativa del nivel de expresión génica, en particular en comparación con un nivel de control. El tipo de nivel que se está detectando estará claro a partir del contexto.

45 Los métodos de detección y/o cuantificación del transcrito génico (ARNm o el ADNc hecho a partir de él) utilizando técnicas de hibridación de ácidos nucleicos son conocidos para los expertos en la materia (véase Sambrook *et al.*, citado anteriormente). Por ejemplo, un método para evaluar la presencia, ausencia o cantidad del ADNc implica una transferencia de Southern como se describe anteriormente. Brevemente, el ARNm se aísla (por ejemplo, utilizando un método de extracción por guanidinio-fenol-cloroformo, Sambrook *et al.* citado anteriormente) y transcribiendo de forma inversa para producir ADNc. Después, el ADNc se dirige de forma opcional y se procesa en un gel en un tampón y se transfiere a membranas. Después, se lleva a cabo la hibridación utilizando sondas de ácido nucleico específicas para el ADNc diana.

55 Un principio general de tales ensayos de diagnóstico y pronóstico implica la preparación de una muestra o mezcla de reacción que puede contener un marcador y una sonda, en condiciones apropiadas y durante un tiempo suficiente para permitir interactuar y unirse al marcador y la sonda, formando así un complejo que puede retirarse y/o detectarse en la mezcla de reacción. Estos ensayos pueden realizarse en una diversidad de modos.

60 Por ejemplo, un método para realizar un ensayo implicaría anclar el marcador o sonda en un soporte en fase sólida, también denominado sustrato, y detectar los complejos de marcador/sonda anclados en la fase sólida al final de la reacción. En una realización de tal método, puede anclarse sobre un transportador o soporte en fase sólida una muestra de un sujeto, la cual se va a ensayar para la presencia y/o concentración del marcador. En otra realización es posible la situación inversa, en que la sonda puede anclarse a una fase sólida y puede permitirse reaccionar una muestra de un sujeto como un componente no anclado del ensayo.

65

Existen muchos métodos establecidos de anclaje de componentes de ensayo a una fase sólida. Estos incluyen, pero sin limitación, moléculas de marcador o sonda que se inmovilizan a través de la conjugación de biotina y estreptavidina. Tales componentes de ensayo biotinilados pueden prepararse a partir de biotina-NHS (N-hidroxisuccinimida) utilizando técnicas conocidas en la materia (por ejemplo, kit de biotinilación, Pierce Chemicals, Rockford, IL) e inmovilizarse en pocillos de placas de 96 pocillos recubiertos de estreptavidina (Pierce Chemical). En determinadas realizaciones, las superficies con componentes de ensayo inmovilizados pueden prepararse con antelación y almacenarse.

Otros transportadores o soportes en fase sólida adecuados para tales ensayos incluyen cualquier material que tenga la capacidad de unir la clase de molécula a la cual el marcador o sonda pertenece. Los soportes o transportadores muy conocidos incluyen, pero sin limitación, vidrio, poliestireno, nylon, polipropileno, polietileno, dextrano, amilasas, celulosas naturales o modificadas, poliacrilamidas, gabros y magnetita.

Para realizar los ensayos con las estrategias mencionadas anteriormente, se añade el componente no inmovilizado a la fase sólida, tras lo cual se ancla el segundo componente. Después de acabar la reacción, los componentes que no han formado complejo pueden retirarse (por ejemplo, mediante lavado) en condiciones de forma que cualquiera de los complejos formados permanezca inmovilizado sobre la fase sólida. La detección de los complejos marcador/sonda anclados a la fase sólida puede realizarse en una serie de métodos descritos en el presente documento.

En otra realización, cuando la sonda es el componente no anclado del ensayo puede marcarse con el fin de detección y lectura del ensayo, ya sea de forma directa o indirecta, con marcadores detectables discutidos en el presente documento y que son bien conocidos para un experto en la materia.

También es posible detectar directamente la formación del complejo de marcador/sonda sin manipulación adicional o marcaje de cualquiera de los componentes (marcador o sonda), por ejemplo utilizando la técnica de transferencia de energía de fluorescencia (véase, por ejemplo, Lakowicz *et al.*, Patente de Estados Unidos N.º 5.631.169; Stavrianopoulos, *et al.*, Patente de Estados Unidos N.º 4.868.103). Se selecciona un marcador fluoróforo en la primera molécula "donante" de forma que, tras la excitación con una luz incidente de la longitud de onda apropiada, su energía fluorescente emitida será absorbida por un marcador fluorescente en una segunda molécula "aceptora", que a su vez es capaz de fluorescer debido a la energía absorbida. Como alternativa, la molécula de proteína "donante" puede simplemente utilizar la energía fluorescente natural de los restos triptófano. Se escogen marcadores que emiten distintas longitudes de onda de luz, de forma que el marcador de molécula "aceptora" puede diferenciarse del "donante". Pueden evaluarse las relaciones espaciales entre las moléculas dado que la eficacia de la transferencia de energía entre los marcadores está relacionada con la distancia que separa las moléculas. En una situación en que se produce la unión entre las moléculas, en el ensayo la emisión fluorescente del marcador de molécula "aceptora" debería ser máxima. Un suceso de unión de FET puede medirse de forma conveniente a través de medios convencionales de detección fluorométricos bien conocidos en la técnica (por ejemplo, utilizando un fluorímetro).

En otra realización, la determinación de la capacidad de una sonda para reconocer un marcador puede llevarse a cabo sin marcaje de cualquier componente (sonda o marcador) utilizando una tecnología tal como el análisis de interacción biomolecular (AIB) en tiempo real (véase, por ejemplo, Sjolander, S. y Urbaniczky, C., 1991, *Anal. Chem.* 63: 2338-2345 y Szabo *et al.*, 1995, *Curr. Opin. Struct. Biol.* 5: 699-705). Como se usa en el presente documento, "AIB" o "resonancia de plasmón superficial" es una tecnología para estudiar las interacciones inespecíficas en tiempo real, sin marcar ninguno de los elementos que interactúan (por ejemplo, BIAcore). Los cambios en la masa en la superficie de unión (indicativo de un suceso de unión) da como resultado modificaciones en el índice de refracción de la luz cerca de la superficie (el fenómeno óptico de resonancia de plasmón superficial (RPS)), dando como resultado una señal detectable que puede utilizarse como una indicación de reacciones en tiempo real entre moléculas biológicas.

Como alternativa, en otra realización pueden realizarse ensayos de diagnóstico y pronóstico análogos con un marcador y una sonda como solutos en una fase líquida. En tal ensayo, el marcador y la sonda que están formando un complejo se separan de los componentes que no han formado complejo mediante cualquiera de varias técnicas convencionales incluyendo, pero sin limitación: centrifugación diferencial, cromatografía, electroforesis e inmunoprecipitación. En la centrifugación diferencial, los complejos de marcador/sonda pueden separarse de los componentes de ensayo que no forman complejo a través de una serie de etapas de centrifugación, debido al equilibrio de diferenciación distinto de los complejos a base de sus distintos tamaños y densidades (véase, por ejemplo, Rivas, G. y Minton, A.P., 1993, *Trends Biochem Sci.* 18(8): 284-7). También pueden utilizarse técnicas cromatográficas convencionales para separar las moléculas que forman complejo de las que no forman complejo.

Por ejemplo, la cromatografía de filtración en gel separa moléculas a base de su tamaño y a través de la utilización de una resina de filtración en gel apropiada en un formato de columna, por ejemplo, puede separarse un complejo relativamente más grande de los componentes que no forman complejo relativamente más pequeños. De forma similar, las propiedades de carga relativamente distintas del complejo de marcador/sonda, en comparación con los componentes que no forman complejo, pueden explotarse para diferenciar el complejo de los componentes que no

forman complejo, por ejemplo, a través de la utilización de resinas de cromatografía de intercambio iónico. Tales resinas y técnicas cromatográficas son bien conocidas para un experto en la materia (véase, por ejemplo, Heegaard, N.H., 1998, J. Mol. Recognit. Winter 11(1-6): 141-8; Hage, D.S. y Tweed, S.A. J Chromatogr B Biomed Sci Appl 10 de octubre de 1997; 699(1-2): 499-525). También puede emplearse electroforesis en gel para separar componentes de ensayo que están formando complejo de los componentes no unidos (véase, por ejemplo, Ausubel *et al.*, ed., Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York, 1987-1999). En esta técnica, los complejos de proteína o ácido nucleico se separan a base de su tamaño o carga, por ejemplo. Para mantener la interacción de unión durante el procedimiento electroforético son típicos los materiales de matriz de gel no desnaturizante y condiciones en ausencia de un agente reductor. Las condiciones apropiadas para el ensayo particular y los componentes del mismo serán conocidos por un experto en la materia.

En una realización particular, puede determinarse en una muestra biológica el nivel de ARNm correspondiente al marcador mediante formatos *in situ* e *in vitro*, utilizando métodos conocidos en la técnica. Se pretende que la expresión "muestra biológica" incluya tejidos, células, líquidos corporales y aislados del mismo, aislados de un sujeto, así como tejidos, células y líquidos presentes en un sujeto. Muchos métodos para la detección de la expresión utilizan ARN aislado. Para los métodos *in vitro*, para la purificación de ARN a partir de células puede utilizarse cualquier técnica de aislamiento de ARN que no seleccione por aislamiento de ARNm (véase, por ejemplo, Ausubel *et al.*, ed., Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York 1987-1999). De forma adicional, pueden procesarse fácilmente gran cantidad de muestras de tejido utilizando técnicas muy conocidas para los expertos en la materia, tales como, por ejemplo, el procedimiento de aislamiento de ARN de única etapa de Chomczynski (1989, Patente de Estados Unidos N.º 4.843.155).

El ácido nucleico aislado puede utilizarse en ensayos de hibridación o amplificación que incluyen, pero sin limitación, análisis de Southern o Northern, análisis por reacción en cadena de la polimerasa y matrices de sondas. Un método de diagnóstico para la detección de los niveles de ARNm implica poner en contacto el ARNm aislado con una molécula de ácido nucleico (sonda) que puede hibridar con el ARNm que codifica el gen que se está detectando. La sonda de ácido nucleico puede, por ejemplo, ser un ADNc de longitud completa o una porción del mismo, tal como un oligonucleótido de al menos 7, 15, 30, 50, 100, 250 o 500 nucleótidos de longitud y que es suficiente para hibridar de forma específica en condiciones rigurosas con un ARNm o ADN genómico que codifica un marcador de la presente invención. En el presente documento se describen otras sondas adecuadas para su uso en los ensayos diagnósticos de la presente invención. La hibridación de un ARNm con la sonda indica que el marcador en cuestión se está expresando.

En un formato, el ARNm se inmoviliza en una superficie sólida y se pone en contacto con una sonda, por ejemplo corriendo el ARNm aislado en un gel de agarosa y transfiriendo el ARNm del gel a una membrana, tal como nitrocelulosa. En un formato alternativo, la sonda (o sondas) se inmovilizan en una superficie sólida y el ARNm se pone en contacto con la sonda (o sondas), por ejemplo, en una matriz en chip génica de Affymetrix. Un experto en la materia puede adaptar fácilmente los métodos de detección de ARNm conocidos para su uso en la detección del nivel del ARNm codificado por los marcadores de la presente invención.

Las sondas pueden ser de longitud completa o de menos de la longitud completa de la secuencia de ácido nucleico que codifica la proteína. La especificidad de las sondas más cortas se analiza de forma empírica. Las sondas de ácido nucleico ejemplares son de una longitud de 20 bases o más largas (véase, por ejemplo, Sambrook *et al.* para los métodos de selección de secuencias de sondas de ácido nucleico para su uso en hibridación de ácidos nucleicos). La visualización de las porciones hibridadas permite la determinación cualitativa de la presencia o ausencia de ADNc.

Un método alternativo para determinar el nivel de un transcrito correspondiente a un marcador de la presente invención implica el procedimiento de amplificación de ácidos nucleicos, por ejemplo, mediante rtPCR (la realización experimental expuesta en Mullis, 1987, Patente de Estados Unidos N.º 4.683.202), la reacción en cadena de la ligasa (Barany, 1991, Proc. Natl. Acad. Sci. USA, 88: 189-193), la replicación de secuencia autosostenida (Guatelli *et al.*, 1990, Proc. Natl. Acad. Sci. USA 87: 1874-1878), el sistema de amplificación transcripcional (Kwoh *et al.*, 1989, Proc. Natl. Acad. Sci. USA 86: 1173-1177), replicasa Q-beta (Lizardi *et al.*, 1988, Bio/Technology 6: 1197), replicación en círculo rodante (Lizardi *et al.*, Patente de Estados Unidos N.º 5.854.033) o cualquier otro método de amplificación de ácidos nucleicos, seguido de la detección de las moléculas amplificadas utilizando técnicas muy conocidas para los expertos en la materia. En los métodos de la presente invención también puede utilizarse rtPCR fluorogénica. En la rtPCR fluorogénica la cuantificación está basada en la cantidad de señales fluorescentes, por ejemplo, de TaqMan o verde sybr. Estos esquemas de detección son especialmente útiles para la detección de moléculas de ácido nucleico si tales moléculas están presentes en cantidades muy bajas. Como se usa en el presente documento, los cebadores de amplificación se definen como siendo una pareja de moléculas de ácido nucleico que pueden aparearse en regiones 5' o 3' de un gen (cadenas positiva y negativa, respectivamente, o viceversa) y que contienen una región corta en el medio. En general, los cebadores de amplificación son de aproximadamente 10 a 30 nucleótidos y flanquean una región de aproximadamente 50 a 200 nucleótidos de longitud. En condiciones apropiadas y con los reactivos apropiados, tales cebadores permiten la amplificación de una molécula de ácido nucleico que comprende la secuencia de nucleótidos flanqueada por los cebadores.

Para los métodos *in situ* no se necesita aislar el ARNm de las células antes de la detección. En tales métodos, se prepara o procesa una muestra de células o tejido utilizando métodos histológicos conocidos. Después, se inmoviliza la muestra sobre un soporte, normalmente un portaobjetos de vidrio, y entonces se pone en contacto con una sonda que pueda hibridar con el ARNm que codifica el marcador.

5 Como una alternativa para hacer determinaciones basadas en el nivel de expresión absoluto del marcador, las determinaciones pueden basarse en el nivel de expresión normalizado del marcador. Los niveles de expresión se normalizan corrigiendo el nivel de expresión absoluto de un marcador comparando su expresión con la expresión de un gen que no es un marcador, por ejemplo, un gen constitutivo que se expresa de forma constitutiva. Los genes adecuados para la normalización incluyen genes constitutivos tales como el gen de la actina o genes específicos de células epiteliales. Esta normalización permite la comparación del nivel de expresión en una muestra, por ejemplo, en una muestra del sujeto, con respecto a otra muestra, por ejemplo una muestra no cancerosa, o entre muestras de distintas fuentes.

15 Como alternativa, el nivel de expresión puede proporcionarse como un nivel de expresión relativa. Para determinar un nivel de expresión relativa de un marcador, el nivel de expresión del marcador se determina para 10 o más muestras de aislados de células cancerosas frente a células normales, o incluso 50 o más muestras, antes de la determinación del nivel de expresión para la muestra en cuestión. Se determina el nivel de expresión medio de cada uno de los genes ensayados en el número de muestras más grande y esto se utiliza como un nivel de expresión de referencia para el marcador. El nivel de expresión del marcador determinado para la muestra de prueba (nivel de expresión absoluta) se divide después por el valor medio de expresión obtenido para ese marcador. Esto proporciona un nivel de expresión relativa.

25 En determinadas realizaciones, las muestras utilizadas en la determinación de la medida de referencia serán de células cancerosas o células normales del mismo tipo de tejido. La elección de la fuente de células depende del uso del nivel de expresión relativa. El uso de la expresión hallada en tejidos normales como una puntuación de la expresión media ayuda validar si el marcador ensayado es específico para el tejido del que procede la célula (frente a células normales). Además, a medida que se acumulan más datos, el valor de expresión medio puede revisarse, proporcionando valores de expresión relativa mejorados basados en los datos acumulados. Los datos de expresión de las células normales proporcionan un medio para clasificar la gravedad del estado del cáncer.

35 En otra realización, la expresión de un marcador se evalúa preparando ADN genómico o ARNm/ADNc (es decir, un polinucleótido transcrito) de células en una muestra de un sujeto e hibridando el ADN genómico o el ARNm/ADNc con un polinucleótido de referencia que sea uno complementario de un polinucleótido que comprende el marcador, y fragmentos del mismo. El ADNc puede, de forma opcional, amplificarse utilizando cualquiera de una diversidad de métodos de reacción en cadena de la polimerasa antes de la hibridación con el polinucleótido de referencia. La expresión de uno o más marcadores puede, asimismo, detectarse utilizando PCR cuantitativa (PCRQ) para evaluar el nivel de expresión del marcador (o marcadores). Como alternativa, puede utilizarse cualquiera de los muchos métodos conocidos de detección de mutaciones o variantes (por ejemplo, polimorfismos de nucleótido único, deleciones) de un marcador de la presente invención para detectar la aparición de un marcador mutado en un sujeto.

45 En una realización relacionada, se pone en contacto una mezcla de polinucleótidos transcritos obtenidos de la muestra con un sustrato que tiene fijado a él un polinucleótido complementario u homólogo a al menos una porción (por ejemplo, al menos 7, al menos 10, al menos 15, al menos 20, al menos 25, al menos 30, al menos 40, al menos 50, al menos 100, al menos 500 o más restos de nucleótido) de un marcador de la presente invención. Si los polinucleótidos complementarios u homólogos a un marcador de la presente invención son detectables de forma diferencial sobre el sustrato (por ejemplo, detectables utilizando distintos cromóforos o fluoróforos, o fijados a distintas posiciones seleccionadas), entonces los niveles de expresión de una pluralidad de marcadores puede evaluarse de forma simultánea utilizando un único sustrato (por ejemplo, una micromatriz en "chip génico" de polinucleótidos fijados en posiciones seleccionadas). Cuando se utiliza un método de evaluación de la expresión de marcadores que implica la hibridación de un ácido nucleico con otro, la hibridación puede realizarse en condiciones de hibridación rigurosas.

55 En otra realización, se utiliza una combinación de métodos para evaluar la expresión de un marcador.

Dado que las composiciones, kit y métodos de la presente invención se basan en la detección de una diferencia en el nivel de expresión o número de copias de uno o más marcadores de la presente invención, en determinadas realizaciones el nivel de expresión o número de copias del marcador es significativamente mayor que el límite de detección mínimo del método utilizado para evaluar la expresión o número de copias en al menos uno de células normales y células cancerosas.

### 3. Métodos para evaluar la proteína expresada

65 La actividad o el nivel de una proteína marcadora también pueden detectarse y/o cuantificarse detectando o cuantificando el polipéptido expresado. El polipéptido puede detectarse y cuantificarse mediante cualquiera de una

serie de medios bien conocidos para los expertos en la materia. Estos pueden incluir métodos bioquímicos analíticos tales como electroforesis, electroforesis capilar, cromatografía líquida de alto rendimiento (HPLC), cromatografía en capa fina (CCF), cromatografía de hiperdifusión y similares, o diversos métodos inmunológicos tales como reacciones de precipitina en gel o fluida, inmunodifusión (única o doble), inmunoelectroforesis, radioinmunoensayo (RIA), ensayos inmunoabsorbentes ligados a enzimas (ELISA), ensayos inmunofluorescentes, transferencia de Western, inmunohistoquímica y similares. Un experto en la materia puede adaptar fácilmente métodos de detección de proteínas/anticuerpos conocidos para su uso en la determinación de si las células expresan un marcador de la presente invención.

Otro agente para detectar un polipéptido de la presente invención es un anticuerpo que tiene la capacidad de unirse a un polipéptido que corresponde a un marcador de la presente invención, por ejemplo, un anticuerpo con un marcador detectable. Los anticuerpos pueden ser policlonales o monoclonales. Se puede utilizar un anticuerpo intacto o un fragmento del mismo (por ejemplo, Fab o F(ab')<sub>2</sub>). El término "marcado", con respecto a la sonda o anticuerpo, pretende abarcar el marcaje directo de la sonda o anticuerpo mediante acoplamiento (es decir, conectando físicamente) una sustancia detectable a la sonda o anticuerpo, así como el marcaje indirecto de la sonda o anticuerpo por reactividad con otro reactivo que está marcado directamente. Los ejemplos de marcaje indirecto incluyen la detección de un anticuerpo primario utilizando un anticuerpo secundario marcado de forma fluorescente y el marcado del extremo de una sonda de ADN con biotina, de forma que se pueda detectar con estreptavidina marcada de forma fluorescente.

En otra realización, el anticuerpo está marcado, por ejemplo, un anticuerpo radiomarcado, marcado con cromóforo, marcado con fluoróforo o marcado con enzima. En otra realización, se utiliza un derivado de anticuerpo (por ejemplo, un anticuerpo conjugado con un sustrato, o con la proteína o ligando de una pareja de proteína-ligando {por ejemplo, biotina-estreptavidina}) o un fragmento de anticuerpo (por ejemplo, un anticuerpo monocatenario), un dominio hipervariable de anticuerpo aislado, etc.) que se une de forma específica a una proteína que corresponde al marcador, tal como la proteína codificada por la fase de lectura abierta que corresponde al marcador o la proteína que ha experimentado toda o una parte de su modificación postraduccional normal.

La inmunohistoquímica o IHQ se refiere al procedimiento de localización de antígenos (por ejemplo, proteínas) en células de un corte de tejido que aprovecha el principio de que los anticuerpos se unen de forma específica a los antígenos en tejidos biológicos. La tinción inmunohistoquímica se utiliza ampliamente en el diagnóstico de células anómalas, como las que se encuentran en tumores cancerosos. Marcadores moleculares específicos son característicos de sucesos celulares particulares tales como la proliferación o muerte celular (apoptosis). La IHQ también se utiliza ampliamente en investigación para comprender la distribución y emplazamiento de biomarcadores y proteínas expresadas de forma diferencial en distintas partes de un tejido biológico. La visualización de una interacción anticuerpo-antígeno se puede llevar a cabo de varias maneras. En el caso más común, se conjuga un anticuerpo con a una enzima, tal como la peroxidasa, que puede catalizar una reacción de producción de color. Como alternativa, el anticuerpo también se puede etiquetar con un fluoróforo, tal como fluoresceína, rodamina, DyLight Fluor o Alexa Fluor.

Las proteínas de las células se pueden aislar usando técnicas que son bien conocidas por los expertos en la materia. Los métodos de aislamiento de proteínas empleados pueden ser, por ejemplo, tales como los descritos en Harlow y Lane (Harlow y Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York).

En un formato, se pueden utilizar anticuerpos o fragmentos de anticuerpos en métodos tales como transferencias Western o técnicas de inmunofluorescencia para detectar las proteínas expresadas. En tales usos, se puede inmovilizar el anticuerpo o las proteínas en un soporte sólido. Los soportes o transportadores en fase sólida adecuados incluyen cualquier soporte capaz de unirse a un antígeno o a un anticuerpo. Los soportes o transportadores bien conocidos incluyen vidrio, poliestireno, polipropileno, polietileno, dextrano, nylon, amilasas, celulosas naturales y modificadas, poliacrilamidas, gabros y magnetita.

Un experto en la materia conocerá muchos otros vehículos adecuados para unir un anticuerpo o antígeno, y podrá adaptar tal soporte para su uso con la presente invención. Por ejemplo, la proteína aislada de las células puede correrse en una electroforesis en gel de poliacrilamida e inmovilizarse en un soporte de fase sólida, tal como nitrocelulosa. Después, el soporte puede lavarse con tampones adecuados seguido de tratamiento con el anticuerpo marcado de forma detectable. Después, el soporte de fase sólida se puede lavar con el tampón una segunda vez para eliminar el anticuerpo no unido. La cantidad de marcador unido sobre el soporte sólido puede detectarse entonces por medios convencionales. Los medios para detectar proteínas usando técnicas electroforéticas son bien conocidos para los expertos en la materia (véase en general, R. Scopes (1982) *Protein Purification*, Springer-Verlag, NY.; Deutscher, (1990) *Methods in Enzymology* vol. 182: *Guide to Protein Purification*, Academic Press, Inc., NY).

En otra realización, se utiliza análisis de transferencia Western (inmunotransferencia) para detectar y cuantificar la presencia de un polipéptido en la muestra. En general, esta técnica comprende separar las proteínas de la muestra por electroforesis en gel a base del peso molecular, transferir las proteínas separadas a un soporte sólido adecuado (tal como un filtro de nitrocelulosa, un filtro de nylon o filtro de nylon derivatizado) e incubar la muestra con los

anticuerpos que se unen de forma específica a un polipéptido. Los anticuerpos anti polipéptido se unen de forma específica al polipéptido en el soporte sólido. Estos anticuerpos pueden marcarse directamente o, como alternativa, pueden detectarse posteriormente utilizando anticuerpos marcados (por ejemplo, anticuerpos anti humano de oveja marcados) que se unen de forma específica al anti polipéptido.

En otra realización, el polipéptido se detecta utilizando un inmunoensayo. Como se usa en el presente documento, un inmunoensayo es un ensayo que utiliza un anticuerpo que se une de forma específica al analito. El inmunoensayo se caracteriza así por la detección de la unión específica de un polipéptido a un anti anticuerpo en oposición al uso de otras propiedades físicas o químicas para aislar, dirigirse a y cuantificar el analito.

El polipéptido se detecta y/o cuantifica utilizando cualquiera de una serie de ensayos de unión inmunológica bien reconocidos (véase, por ejemplo, las Patentes de Estados Unidos N.º 4.366.241 ; 4.376.110 ; 4.517.288 y 4.837.168). Para una revisión de los inmunoensayos generales, véase también Asai (1993) *Methods in Cell Biology* Volume 37: Antibodies in Cell Biology, Academic Press, Inc. Nueva York; Stites & Terr (1991) *Basic and Clinical Immunology* 7ª Edición.

Los ensayos de unión inmunológica (o inmunoensayos) normalmente utilizan un "agente de captura" para unirse de forma específica y a menudo para inmovilizar el analito (polipéptido o subsecuencia). El agente de captura es una fracción que se une de forma específica al analito. En otra realización, el agente de captura es un anticuerpo que se une de forma específica a un polipéptido. El anticuerpo (antipéptido) puede producirse mediante cualquiera de una serie de medios bien conocidos por los expertos en la materia.

Los inmunoensayos también utilizan a menudo un agente de marcaje para unirse de forma específica y marcar el complejo de unión formado por el agente de captura y el analito. El propio agente de marcaje puede ser una de las fracciones que comprenden el complejo de anticuerpo/analito. Por lo tanto, el agente de marcaje puede ser un polipéptido marcado o un anti anticuerpo marcado. Como alternativa, el agente de marcaje puede ser una tercera fracción, tal como otro anticuerpo, que se une de forma específica al complejo de anticuerpo/polipéptido.

En una realización, el agente de marcaje es un segundo anticuerpo humano que porta un marcador. Como alternativa, el segundo anticuerpo puede carecer de un marcador, pero a su vez puede estar unido a un tercer anticuerpo marcado específico para anticuerpos de la especie de la que procede el segundo anticuerpo. El segundo puede modificarse con una fracción detectable, por ejemplo, tal como biotina, a la que se puede unir de forma específica una tercera molécula marcada, tal como estreptavidina marcada con una enzima.

También se pueden utilizar como agente marcador otras proteínas que tengan la capacidad de unirse de forma específica a las regiones constantes de las inmunoglobulinas, tales como la proteína A o la proteína G. Estas proteínas son constituyentes normales de las paredes celulares de las bacterias estreptocócicas. Presentan una fuerte reactividad no inmunogénica con las regiones constantes de inmunoglobulinas de una diversidad de especies (véase, en general, Kronval, *et al.*, (1973) *J. Immunol.*, 111: 1401-1406 y Akerstrom (1985) *J. Immunol.*, 135: 2589 - 2542).

Como se indica anteriormente, los inmunoensayos para la detección y/o cuantificación de un polipéptido pueden tomar una amplia diversidad de formatos bien conocidos por los expertos en la materia.

Los inmunoensayos ejemplares para detectar un polipéptido pueden ser competitivos o no competitivos. Los inmunoensayos no competitivos son ensayos en los cuales se mide de forma directa la cantidad de analito capturado. En un ensayo de tipo "sándwich", por ejemplo, el agente de captura (anticuerpos anti péptido) se puede unir directamente a un sustrato sólido donde están inmovilizados. Después, estos anticuerpos inmovilizados capturan el polipéptido presente en la muestra de prueba. El polipéptido así inmovilizado se une entonces por un agente de marcaje, tal como un segundo anticuerpo humano que porta un marcador.

En ensayos competitivos, la cantidad de analito (polipéptido) presente en la muestra se mide de forma indirecta midiendo la cantidad de un analito (polipéptido) añadido (exógeno) desplazado (o competido) de un agente de captura (anticuerpo anti péptido) por el analito presente en la muestra. En un ensayo competitivo, se añade a la muestra una cantidad conocida de, en este caso, un polipéptido y después se pone en contacto la muestra con un agente de captura. La cantidad de polipéptido unido al anticuerpo es inversamente proporcional a la concentración de polipéptido presente en la muestra.

En otra realización, el anticuerpo se inmoviliza en un sustrato sólido. La cantidad de polipéptido unido al anticuerpo puede determinarse midiendo la cantidad de polipéptido presente en un complejo polipéptido/anticuerpo o, como alternativa, midiendo la cantidad de polipéptido que no forma complejo restante. La cantidad de polipéptido puede detectarse proporcionando un polipéptido marcado.

Los ensayos descritos en el presente documento se puntúan (como positivos o negativos o como cantidad de polipéptido) de acuerdo con métodos convencionales bien conocidos para los expertos en la técnica. El método particular de puntuación dependerá del formato del ensayo y la elección del marcador. Por ejemplo, un ensayo de

transferencia de Western puede puntuarse visualizando el producto coloreado producido por el marcador enzimático. Una banda o mancha de color claramente visible con el peso molecular correcto se puntúa como un resultado positivo, mientras que la ausencia de una mancha o banda claramente visible se puntúa como negativo. La intensidad de la banda o mancha puede proporcionar una medida cuantitativa del polipéptido.

Los anticuerpos para su uso en los diversos inmunoensayos descritos en el presente documento pueden producirse como se describe en el presente documento.

En otra realización, el nivel (actividad) se ensaya midiendo la actividad enzimática del producto génico. Los métodos para ensayar la actividad de una enzima son bien conocidos para los expertos en la materia.

Las técnicas *in vivo* para la detección de una proteína marcadora incluyen la introducción en un sujeto de un anticuerpo marcado dirigido frente a la proteína. Por ejemplo, el anticuerpo puede marcarse con un marcador radioactivo cuya presencia y emplazamiento en un sujeto pueden detectarse mediante técnicas de formación de imágenes convencionales.

Determinados marcadores identificados por los métodos de la presente invención pueden ser proteínas secretadas. Es una cuestión simple para el experto en la materia determinar si cualquier proteína marcadora particular es una proteína secretada. Para realizar esta determinación, la proteína marcadora se expresa en, por ejemplo, una célula de mamífero, por ejemplo, una línea celular humana, se recoge líquido extracelular y se evalúa la presencia o ausencia de la proteína en el líquido extracelular (por ejemplo, utilizando un anticuerpo marcado que se une de forma específica a la proteína).

El siguiente es un ejemplo de un método que puede usarse para detectar la secreción de una proteína. Se incuban aproximadamente  $8 \times 10^5$  células 293T a 37 °C en pocillos que contienen medio de cultivo (medio de Eagle modificado por Dulbecco {DMEM} complementado con el 10 % de suero fetal bovino) en CO<sub>2</sub> al 5 % (v/v), el 95 % de atmósfera de aire a aproximadamente el 60-70 % de confluencia. Después, las células se transfectan utilizando una mezcla de transfección convencional que comprende 2 microgramos de ADN que comprende un vector de expresión que codifica la proteína y 10 microlitros de LipofectAMINE™ (GIBCO/BRL n.º de catálogo 18342-012) por pocillo. La mezcla de transfección se mantiene durante aproximadamente 5 horas y después se reemplaza con medio de cultivo recién preparado y se mantiene en una atmósfera de aire. Cada pocillo se enjuaga suavemente dos veces con DMEM que no contiene metionina o cisteína (DMEM-MC; ICN n.º de catálogo 16-424-54). Se añade aproximadamente 1 mililitro de DMEM-MC y aproximadamente 50 microcurios del reactivo Trans-<sup>35</sup>S™ (ICN n.º de catálogo 51006) a cada pocillo. Los pocillos se mantienen en la atmósfera de CO<sub>2</sub> al 5 % descrita anteriormente y se incuban a 37 °C durante un período seleccionado. Después de la incubación, se retiran 150 microlitros de medio acondicionado y se centrifuga para retirar las células flotantes y los residuos. La presencia de la proteína en el sobrenadante es una indicación de que la proteína se secreta.

Se apreciará que las muestras objeto, por ejemplo, una muestra que contiene esputo, lavado broncoalveolar, derrame pleural, tejido, sangre entera, suero, plasma, raspado bucal, saliva, líquido cefalorraquídeo, orina, heces y médula ósea, pueden contener células en ellas, en particular cuando las células son cancerosas y más en particular cuando el cáncer es metastatizante y, por lo tanto, pueden usarse en los métodos de la presente invención. La muestra de células puede, por supuesto, someterse a una diversidad de técnicas conocidas de almacenamiento y preparativas posrecolección (por ejemplo, extracción de ácidos nucleicos y/o proteínas, fijación, almacenamiento, congelación, ultrafiltración, concentración, evaporación, centrifugación, etc.) antes de evaluar el nivel de expresión del marcador en la muestra. Por lo tanto, las composiciones, kits y métodos de la presente invención se pueden usar para detectar la expresión de marcadores correspondientes a proteínas que tengan al menos una porción que está presentada en la superficie de las células que las expresan. Es una cuestión simple para el experto en la materia determinar si la proteína que corresponde a cualquier marcador particular comprende una proteína de superficie celular. Por ejemplo, pueden usarse métodos inmunológicos para detectar tales proteínas en células enteras o pueden usarse métodos bien conocidos de análisis de secuencias basados en ordenadores (por ejemplo, el programa SIGNALP, Nielsen *et al.*, 1997, Protein Engineering 10: 1-6) para predecir la presencia de al menos un dominio extracelular (es decir, que incluya tanto proteínas secretadas como proteínas que tienen al menos un dominio de superficie celular). La expresión de un marcador que corresponde a una proteína que tiene al menos una porción que se presenta en la superficie de una célula que la expresa puede detectarse sin necesariamente lisar la célula (por ejemplo, utilizando un anticuerpo marcado que se une de forma específica a un dominio de superficie celular de la proteína).

La descripción también abarca kits para la detección de la presencia de un polipéptido o ácido nucleico que corresponde a un marcador de la presente invención en una muestra biológica, por ejemplo, una muestra que contiene esputo, lavado broncoalveolar, derrame pleural, tejido, sangre entera, suero, plasma, raspado bucal, saliva, líquido cefalorraquídeo, orina, heces y médula ósea. Dichos kits pueden utilizarse para determinar si un sujeto padece o tiene un riesgo aumentado de desarrollar cáncer. Por ejemplo, el kit puede comprender un compuesto o agente marcado que tenga la capacidad de detectar un polipéptido o un ARNm que codifica un polipéptido que corresponde a un marcador de la presente invención en una muestra biológica y medios para determinar la cantidad del polipéptido o ARNm en la muestra (por ejemplo, un anticuerpo que se une al polipéptido o una sonda de

oligonucleótido que se une al ADN o ARNm que codifica el polipéptido). Los kits también pueden incluir instrucciones para interpretar los resultados obtenidos utilizando el kit.

5 Para kits basados en anticuerpos, el kit puede comprender, por ejemplo: (1) un primer anticuerpo (por ejemplo, unido a un soporte sólido) que se une a un polipéptido que corresponde a un marcador de la presente invención y, de forma opcional, (2) un segundo anticuerpo distinto que se une al polipéptido o al primer anticuerpo, y está conjugado con un marcador detectable.

10 Para kits basados en oligonucleótidos, el kit puede comprender, por ejemplo: (1) un oligonucleótido, por ejemplo, un oligonucleótido marcado de forma detectable, que hibrida con una secuencia de ácido nucleico que codifica un polipéptido que corresponde a un marcador de la presente invención o (2) una pareja de cebadores útiles para amplificar una molécula de ácido nucleico que corresponde a un marcador de la presente invención. El kit también puede comprender, por ejemplo, un agente tamponante, un conservante o un agente estabilizante de proteínas. El kit también puede comprender los componentes necesarios para detectar el marcador detectable (por ejemplo, una enzima o un sustrato). El kit también puede contener una muestra de control o una serie de muestras de control que se pueden ensayar y comparar con la muestra de prueba. Cada componente del kit se puede incluir en un recipiente individual y todos los diversos contenedores pueden estar dentro de un único envase, junto con instrucciones para interpretar los resultados de los ensayos realizados utilizando el kit.

#### 20 4. Método para evaluar modificaciones estructurales

La invención también proporciona un método para evaluar la presencia de una modificación estructural, por ejemplo, una mutación.

25 Otro método de detección es la hibridación específica de alelos que utiliza sondas que solapan con el sitio polimórfico y que tienen aproximadamente 5, aproximadamente 10, aproximadamente 20, aproximadamente 25 o aproximadamente 30 nucleótidos alrededor de la región polimórfica. En otra realización de la presente invención, varias sondas que tienen la capacidad de hibridar de forma específica con las mutaciones están unidas a un soporte en fase sólida, por ejemplo, un "chip". Los oligonucleótidos pueden unirse a un soporte sólido mediante una diversidad de procedimientos, incluyendo la litografía. Por ejemplo, un chip puede contener hasta 250.000 oligonucleótidos (GeneChip, Affymetrix™). El análisis de detección de mutaciones utilizando estos chips que comprenden oligonucleótidos, también denominados "matrices de sondas de ADN" se describe, por ejemplo, en Cronin *et al.* (1996) *Human Mutation* 7: 244. En una realización, un chip comprende todas las mutaciones de al menos una región polimórfica de un gen. El soporte en fase sólida se pone en contacto después con un ácido nucleico de prueba y se detecta la hibridación con las sondas específicas. Por consiguiente, la identidad de numerosas mutaciones de uno o más genes puede identificarse en un experimento de hibridación simple. Por ejemplo, la identidad de la mutación del polimorfismo de nucleótido en elemento regulador 5' cadena arriba puede determinarse en un experimento de hibridación único.

40 En otros métodos de detección, es necesario amplificar en primer lugar al menos una porción de un marcador antes de identificar la mutación. La amplificación puede realizarse, por ejemplo, mediante PCR y/o LCR (véase Wu y Wallace (1989) *Genomics* 4: 560), de acuerdo con métodos conocidos en la técnica. En una realización, el ADN genómico de una célula se expone a dos cebadores de PCR y la amplificación durante un número de ciclos suficiente para producir la cantidad necesaria de ADN amplificado. En determinadas realizaciones, los cebadores se emplazan a entre 150 y 350 pares de bases de distancia.

50 Los métodos de amplificación alternativos incluyen: replicación de secuencia autosostenida (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), sistema de amplificación transcripcional (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173-1177), replicasa Q-beta (Lizardi, P.M. *et al.*, (1988) *Bio/Technology* 6: 1197) y replicación de secuencia autosostenida (Guatelli *et al.*, (1989) *Proc. Nat. Acad. Sci.* 87: 1874), y amplificación de secuencia basada en ácidos nucleicos (NABSA) o cualquier otro método de amplificación de ácidos nucleicos, seguido de la detección de las moléculas amplificadas utilizando técnicas bien conocidas para los expertos en la materia. Estos esquemas de detección son especialmente útiles para la detección de moléculas de ácido nucleico si tales moléculas están presentes en cantidades muy bajas.

55 En una realización, puede utilizarse cualquiera de una diversidad de reacciones de secuenciación conocidas en la técnica para secuenciar de forma directa al menos una porción de un marcador y detectar mutaciones comparando la secuencia de la muestra con la correspondiente secuencia de referencia (control). Las reacciones de secuenciación ejemplares incluyen las basadas en técnicas desarrolladas por Maxam y Gilbert (*Proc. Natl Acad Sci USA* (1977) 74: 560) o Sanger (Sanger *et al.* (1977) *Proc. Nat. Acad. Sci* 74: 5463). También se contempla que, cuando se realizan los ensayos objeto, pueda utilizarse cualquiera de una diversidad de procedimientos de secuenciación automatizados (*Biotechniques* (1995) 19: 448), incluyendo secuenciación por espectrometría de masas (véase, por ejemplo, la Patente de Estados Unidos Número 5.547.835 y la solicitud de Patente Internacional Número de Publicación WO 94/16101, titulada "Secuenciación de ADN por espectrometría de masas" por H. Koster; la Patente de Estados Unidos Número 5.547.835 y la solicitud de Patente Internacional Número de Publicación WO 94/21822 titulada "Secuenciación de ADN por espectrometría de masas a través de degradación con exonucleasa"

por H. Koster) y la Patente de Estados Unidos Número 5.605.798, y la solicitud de Patente Internacional Número PCT/US96/03651 titulada "Diagnóstico de ADN basado en espectrometría de masas" por H. Koster; Cohen *et al.* (1996) *Adv Chromatogr* 36: 127-162 y Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38: 147-159). Será evidente para un experto en la materia que, para determinadas realizaciones, necesita determinarse en la reacción de secuenciación la aparición de solo una, dos o tres bases de ácido nucleico. Por ejemplo, puede llevarse a cabo un tramo de A o similar, por ejemplo, en donde solo se detecta un nucleótido.

Aún otros métodos de secuenciación se divulgan en, por ejemplo, la Patente de Estados Unidos Número 5.580.732 titulada "Método de secuenciación de ADN empleando una sonda mixta de ADN-cadena polimérica" y la Patente de Estados Unidos Número 5.571.676 titulada "Método para la secuenciación de ADN *in vitro* dirigida por desapareamiento".

En algunos casos, la presencia de un alelo específico de un marcador en el ADN de un sujeto puede demostrarse por un análisis de enzima de restricción. Por ejemplo, un polimorfismo de nucleótido específico puede dar como resultado una secuencia de nucleótidos que comprende un sitio de restricción que está ausente en la secuencia de nucleótidos de otra mutación.

En una realización adicional, la protección frente a agentes de escisión (tales como una nucleasa, hidroxilamina o tetróxido de osmio y con piperidina) puede utilizarse para detectar las bases desapareadas en heterodúplex de ARN/ARN ADN/ADN o ARN/ADN (Myers, *et al.* (1985) *Science* 230: 1242). En general, la técnica de "escisión por desapareamiento" se inicia proporcionando heterodúplex formados hibridando un ácido nucleico de control, que opcionalmente está marcado, por ejemplo ARN o ADN, que comprende una secuencia de nucleótidos de una mutación marcadora con un ácido nucleico de muestra, por ejemplo ARN o ADN, obtenido de una muestra de tejido. Los dúplex bicatenarios se tratan con un agente que escinde regiones monocatenarias del dúplex, tal como dúplex formados a base de desapareamientos de pares de bases entre las cadenas de control y de muestra. Por ejemplo, los dúplex de ARN/ADN pueden tratarse con ARNasa y los híbridos de ADN/ADN tratarse con nucleasa S1 para digerir de forma enzimática las regiones desapareadas. En otras realizaciones, los dúplex de ADN/ADN o ARN/ADN pueden tratarse con hidroxilamina o tetróxido de osmio y con piperidina para digerir las regiones desapareadas. Tras la digestión de las regiones desapareadas, el material resultante se separa entonces por tamaño en geles de poliacrilamida desnaturizantes para determinar si los ácidos nucleicos de control y de muestra tienen una secuencia de nucleótidos idéntica o en qué nucleótidos son diferentes. Véase, por ejemplo, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85: 4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217: 286-295. En otra realización, el ácido nucleico de control o de muestra está marcado para la detección.

En otra realización, puede identificarse una mutación por cromatografía líquida de alto rendimiento desnaturizante (DHPLC) (Oefner y Underhill, (1995) *Am. J. Human Gen.* 57: Supl. A266). La DHPLC utiliza cromatografía de par iónico de fase inversa para detectar los heterodúplex que se generan durante la amplificación de fragmentos de PCR a partir de individuos que son heterocigotas en un locus de nucleótidos particular dentro de ese fragmento (Oefner y Underhill (1995) *Am. J. Human Gen.* 57: Supl. A266). En general, los productos de PCR se producen utilizando cebadores de PCR que flanquean el ADN de interés. Se lleva a cabo el análisis por DHPLC y los cromatogramas resultantes se analizan para identificar modificaciones o deleciones de pares de bases a base de perfiles cromatográficos específicos (véase O'Donovan *et al.* (1998) *Genomics* 52: 44-49).

En otras realizaciones, se utilizan las modificaciones en la movilidad electroforética para identificar el tipo de mutación marcadora. Por ejemplo, puede utilizarse polimorfismo de conformación de cadena simple (SSCP) para detectar diferencias en la movilidad electroforética entre los ácidos nucleicos mutante y de tipo silvestre (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA* 86: 2766, véase también Cotton (1993) *Mutat Res* 285: 125-144 y Hayashi (1992) *Genet Anal Tech Appl* 9: 73-79). Los fragmentos de ADN monocatenario de los ácidos nucleicos de muestra y de control se desnaturizan y se permite que renaturalicen. La estructura secundaria de los ácidos nucleicos monocatenarios varía según la secuencia y la modificación resultante de la movilidad electroforética permite la detección incluso de un único cambio base. Los fragmentos de ADN pueden marcarse o detectarse con sondas marcadas. La sensibilidad del ensayo puede potenciarse utilizando ARN (en lugar de ADN), en el que la estructura secundaria es más sensible a un cambio de secuencia. En otra realización, el método objeto utiliza un análisis de heterodúplex para separar moléculas de heterodúplex bicatenarias sobre la base de cambios en la movilidad electroforética (Keen *et al.* (1991) *Trends Genet* 7: 5).

En aún otra realización, la identidad de una mutación de una región polimórfica se obtiene analizando el movimiento de un ácido nucleico que comprende la región polimórfica en geles de poliacrilamida que contienen un gradiente desnaturizante, se ensaya utilizando electroforesis en gel de gradiente desnaturizante (DGGE) (Myers *et al.* (1985) *Nature* 313: 495). Cuando se utiliza DGGE como el método de análisis, el ADN se modificará para asegurar que no se desnaturiza de forma completa, por ejemplo, añadiendo un anclaje de GC de aproximadamente 40 pb de un ADN rico en GC de alta fusión por PCR. En una realización adicional, se utiliza un gradiente de temperatura en lugar de un gradiente de agente desnaturizante para identificar diferencias en la movilidad del ADN de control y de muestra (Rosenbaum y Reissner (1987) *Biophys Chem* 265: 1275).

Los ejemplos de técnicas para detectar diferencias de al menos un nucleótido entre dos ácidos nucleicos incluyen, pero sin limitación, la hibridación selectiva de oligonucleótidos, la amplificación selectiva o la extensión de cebador

selectiva. Por ejemplo, pueden prepararse sondas de oligonucleótido en las cuales el nucleótido polimórfico conocido se coloca de forma central (sondas específicas de alelo) y después se hibridan con el ADN diana en condiciones que permiten la hibridación solo si se encuentra una coincidencia perfecta (Saiki *et al.* (1986) *Nature* 324: 163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86: 6230 y Wallace *et al.* (1979) *Nucl. Acids Res.* 6: 3543). Tales técnicas de hibridación de oligonucleótidos específicos de alelo pueden utilizarse para la detección simultánea de varios cambios de nucleótidos en distintas regiones polimórficas del marcador. Por ejemplo, los oligonucleótidos que tienen secuencias de nucleótidos de mutaciones específicas se unen a una membrana de hibridación y esta membrana se hibrida después con un ácido nucleico marcado de muestra. El análisis de la señal de hibridación revelará después la identidad de los nucleótidos del ácido nucleico de muestra.

Como alternativa, puede utilizarse junto con la presente invención la tecnología de amplificación específica de alelos, que depende de una amplificación por PCR selectiva. Los oligonucleótidos utilizados como cebadores para la amplificación específica pueden portar la mutación de interés en el centro de la molécula (de forma que la amplificación depende de una hibridación diferencial) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17: 2437-2448) o en el extremo 3' extremo de un cebador en donde, en condiciones apropiadas, puede impedirse el desapareamiento o reducirse la extensión de la polimerasa (Prossner (1993) *Tibtech* 11: 238; Newton *et al.* (1989) *Nucl. Acids Res.* 17: 2503). Esta técnica, también denominada "PROBE" por *probe oligo base extension* (extensión de bases de oligonucleótido de sonda). Además, puede ser conveniente introducir un sitio de restricción nuevo en la región de la mutación para crear una detección basada en escisión (Gasparini *et al.* (1992) *Mol. Cell Probes* 6: 1).

En otra realización, la identificación de la mutación se lleva a cabo utilizando un ensayo de ligamiento de oligonucleótido (OLA), como se describe, por ejemplo, en la Patente de Estados Unidos Número 4.998.617 y en Landegren, U. *et al.*, (1988) *Science* 241: 1077-1080. El protocolo del OLA usa dos oligonucleótidos que se diseñan para que tengan la capacidad de hibridar con secuencias colindantes de una única cadena de una diana. Uno de los oligonucleótidos está unido a un marcador de separación, por ejemplo, biotinilado, y el otro está marcado de forma detectable. Si en una molécula diana se encuentra la secuencia complementaria precisa, los oligonucleótidos hibridarán de forma que sus extremos colindan y crean un sustrato de ligamiento. Después, el ligamiento permite recuperar los oligonucleótidos marcados utilizando avidina u otro ligando de la biotina. Nickerson, D. A. *et al.* han descrito un ensayo de detección de ácidos nucleicos que combina atributos de la PCR y el OLA (Nickerson, D. A. *et al.*, (1990) *Proc. Natl. Acad. Sci. (U.S.A.)* 87: 8923-8927. En este método, se utiliza PCR para conseguir la amplificación exponencial del ADN diana, el cual se detecta después utilizando OLA.

La invención proporciona adicionalmente métodos para detectar polimorfismos de nucleótido único en un marcador. Dado que los polimorfismos de nucleótido único constituyen sitios de variación flanqueados por regiones de secuencia invariable, su análisis requiere no más que la determinación de la identidad del nucleótido único presente en el sitio de variación y no es necesario determinar una secuencia génica completa para cada sujeto. Se han desarrollado varios métodos para facilitar el análisis de tales polimorfismos de nucleótido único.

En una realización, el polimorfismo de nucleótido único puede detectarse utilizando un nucleótido resistente a exonucleasa especializado, como se divulga, por ejemplo, en Mundy, C. R. (Patente de Estados Unidos Número 4.656.127). De acuerdo con el método, se permite que un cebador complementario a la secuencia alélica inmediatamente 3' con respecto al sitio polimórfico hibride con una molécula diana obtenida de un animal o ser humano particular. Si el sitio polimórfico en la molécula diana contiene un nucleótido que es complementario al derivado del nucleótido resistente a exonucleasa particular presente, entonces el derivado se incorporará en el extremo del cebador hibridado. Tal incorporación hace resistente al cebador a una exonucleasa y permite de este modo su detección. Dado que la identidad del derivado resistente a exonucleasa de la muestra es conocida, un hallazgo de que el cebador se ha hecho resistente a exonucleasas revela que el nucleótido presente en el sitio polimórfico de la molécula diana era complementario del derivado de nucleótido utilizado en la reacción. Este método tiene la ventaja de que no requiere la determinación de grandes cantidades de datos de secuencia ajenos.

En otra realización de la presente invención, se utiliza un método basado en solución para determinar la identidad del nucleótido de un sitio polimórfico (Cohen, D. *et al.* Patente Francesa 2.650.840; Sol. PCT N.º WO91/02087). Como en el método de Mundy de la Patente de Estados Unidos Número 4.656.127, se emplea un cebador que es complementario con las secuencias alélicas inmediatamente 3' de un sitio polimórfico. El método determina la identidad del nucleótido de ese sitio utilizando derivados de didesoxinucleótido marcados, el cual, si es complementario con el nucleótido del sitio polimórfico se incorporará en el extremo del cebador.

Un método alternativo, conocido como análisis de trozo genético o GBA (forma siglada de *genetic bit analysis* se describe en Goelet, P. *et al.* (Sol. PCT N.º 92/15712). El método de Goelet, P. *et al.* utiliza mezclas de terminadores marcados y un cebador que es complementario a la secuencia 3' con respecto a un sitio polimórfico. El terminador marcado que se incorpora se determina así por el nucleótido presente en el sitio polimórfico, y es complementario a este, de la molécula diana que se esté evaluando. En contraste al método de Cohen *et al.* (Patente Francesa 2.650.840; Sol. PCT N.º WO91/02087) el método de Goelet, P. *et al.* es un ensayo de fase heterogénea, en el cual se inmoviliza en una fase sólida el cebador o la molécula diana.

Se han descrito varios procedimientos de incorporación de nucleótidos guiada por cebador para ensayar sitios polimórficos en el ADN (Komher, J. S. *et al.*, (1989) Nucl. Acids. Res. 17: 7779-7784; Sokolov, B. P., (1990) Nucl. Acids Res. 18: 3671; Syvanen, A. -C., *et al.*, (1990) Genomics 8: 684-692; Kuppaswamy, M. N. *et al.*, (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 1143-1147; Prezant, T. R. *et al.*, (1992) Hum. Mutat. 1: 159-164; Ugozzoli, L. *et al.*, (1992) GATA 9: 107-112; Nyren, P. (1993) *et al.*, Anal. Biochem. 208: 171-175). Estos métodos difieren del GBA en que todos se basan en la incorporación de desoxinucleótidos marcados para discriminar entre bases en un sitio polimórfico. En tal formato, dado que la señal es proporcional a la cantidad de desoxinucleótidos incorporados, los polimorfismos que aparecen en los recorridos del mismo nucleótido pueden dar como resultado señales que son proporcionales a la longitud del recorrido (Syvanen, A.C., *et al.*, (1993) Amer. J. Hum. Genet. 52: 46-59).

Para determinar la identidad de la mutación de una región polimórfica emplazada en la región codificante de un marcador, pueden utilizarse no obstante otros métodos además de los descritos anteriormente. Por ejemplo, puede realizarse la identificación de una mutación que codifica un marcador mutado utilizando un anticuerpo que reconoce de forma específica la proteína mutante en, por ejemplo, una inmunohistoquímica o inmunoprecipitación. Los anticuerpos frente a marcadores de tipo silvestre o formas mutadas de los marcadores pueden prepararse de acuerdo con los métodos conocidos en la técnica.

Como alternativa, también se puede medir la actividad de un marcador, tal como la unión a un ligando del marcador. Los ensayos de unión son conocidos en la técnica e implican, por ejemplo, la obtención de células del sujeto y la realización de experimentos de unión con un ligando marcado para determinar si la unión a la forma mutada de la proteína difiere de la unión al tipo silvestre de la proteína.

#### VI. Métodos de exploración ejemplares basados en la inhibición de ALK

La presente invención proporciona adicionalmente métodos para identificar sustancias que inhiben los polipéptidos ALK (por ejemplo, polipéptidos EML4-ALK), para inhibir de este modo la proliferación, crecimiento, diferenciación, apoptosis y/o metástasis de células cancerosas. Los métodos incluyen poner en contacto un compuesto de prueba con un polipéptido ALK (por ejemplo, los polipéptidos enumerados en la Tabla 1). En algunas realizaciones, el polipéptido ALK comprende una variante (por ejemplo, los polipéptidos enumerados en la Tabla 1) que aumenta el riesgo de una respuesta parcial o de falta de respuesta a la inhibición por uno o más inhibidores de ALK. Un compuesto que es un inhibidor de metástasis tumoral puede identificarse determinando el efecto de un compuesto de prueba sobre la actividad de la variante de polipéptido ALK (incluyendo, por ejemplo, la unión a ligando tal como la unión a ATP y/o la actividad tirosina quinasa). En un ejemplo particular, un compuesto de prueba que inhibe la actividad tirosina quinasa en comparación con la actividad en ausencia del compuesto de prueba identifica el compuesto de prueba como un inhibidor de la metástasis tumoral. Si el compuesto inhibe la actividad de una ALK variante, puede evaluarse adicionalmente su capacidad de inhibir el crecimiento o metástasis tumoral.

En particular, los mutantes activadores de tirosina quinasa, incluyendo los nuevos biomarcadores de la presente invención enumerados en la Tabla 1 (por ejemplo, mutantes de ALK), son útiles para identificar compuestos que pueden utilizarse para tratar, mejorar o prevenir neoplasias, por ejemplo, inhibiendo o previniendo la proliferación, crecimiento, diferenciación, apoptosis y/o metástasis de células cancerosas. La exploración de bibliotecas químicas para moléculas que modulen, por ejemplo, inhiban, antagonicen o agonicen, o imiten, son conocidas en la técnica. Por ejemplo, las bibliotecas químicas pueden ser bibliotecas de péptidos, bibliotecas de peptidomiméticos, bibliotecas sintetizadas de forma química, recombinantes, por ejemplo, bibliotecas de presentación en fagos, y bibliotecas basadas en traducción *in vitro*, otras bibliotecas orgánicas sintéticas no peptídicas.

La exploración o creación, identificación y selección de inhibidores de alta afinidad apropiados de un nuevo biomarcador de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK) puede realizarse mediante una diversidad de métodos. Hablando de forma general, estos puede incluir, pero sin limitación, dos estrategias generales. Una estrategia es utilizar el conocimiento estructural sobre la proteína diana para diseñar una molécula candidata con la que interactuará de forma precisa. Un ejemplo sería el diseño molecular asistido por ordenador, en particular basado en la información de estructura-función nueva divulgada en el presente documento como la Figura 6. Una segunda estrategia es utilizar bibliotecas de moléculas combinatorias u otras, mediante lo cual se explora una gran biblioteca de moléculas por afinidad con respecto a la enzima diana, o la capacidad de inhibir la actividad de la enzima diana. En un ejemplo adicional, puede explorarse la capacidad de un panel de anticuerpos de inhibir la enzima diana.

Algunas realizaciones proporcionadas en el presente documento implican determinar la capacidad de un dado compuesto para inhibir un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK). Los compuestos de prueba pueden evaluarse para su probable capacidad de tratar lesiones neoplásicas de forma directa, o de forma indirecta comparando sus actividades frente a compuestos conocidos como útiles para tratar una neoplasia. Por ejemplo, la capacidad de los compuestos de prueba para inhibir la unión a ligando tal como la unión a ATP y/o la actividad tirosina quinasa frente a los nuevos biomarcadores de la presente invención enumerados en la Tabla 1 (por ejemplo, mutantes de ALK) puede compararse con la de los inhibidores de ALK conocidos, tales como PF-02341066 y/o PDD. En una realización, tales compuestos de prueba tendrían al menos el 100 %, al menos el 99,9 %, al menos el 99,8 %, al menos el 99,7 %, al menos el 99,6 %, al menos el

99,5 %, al menos el 99,4 %, al menos el 99,3 %, al menos el 99,2 %, al menos el 99,1 %, al menos el 99 %, al menos el 98,5 %, al menos el 98 %, al menos el 97,5 %, al menos el 97 %, al menos el 96,5 %, al menos el 96 %, al menos el 95,5 %, al menos el 94 %, al menos el 93,5 %, al menos el 93 %, al menos el 92,5 %, al menos el 92 %, al menos el 91,5 %, al menos el 91 %, al menos el 90,5 %, al menos el 90 %, al menos el 89,5 %, al menos el 89 %, al menos el 88,5 %, al menos el 88 %, al menos el 87,5 %, al menos el 87 %, al menos el 86,5 %, al menos el 86 %, al menos el 85,5 %, al menos el 85 %, al menos el 84,5 %, al menos el 84 %, al menos el 83,5 %, al menos el 83 %, al menos el 82,5 %, al menos el 82 %, al menos el 81,5 %, al menos el 81 %, al menos el 80,5 %, al menos el 80 %, al menos el 79 %, al menos el 78 %, al menos el 77 %, al menos el 76 %, al menos el 75 %, al menos el 74 %, al menos el 73 %, al menos el 72 %, al menos el 71 %, al menos el 70 %, al menos el 69 %, al menos el 68 %, al menos el 67 %, al menos el 66 %, al menos el 65 %, al menos el 64 %, al menos el 63 %, al menos el 62 %, al menos el 61 %, al menos el 60 %, al menos el 59 %, al menos el 58 %, al menos el 57 %, al menos el 56 %, al menos el 55 %, al menos el 54 %, al menos el 53 %, al menos el 52 %, al menos el 51 %, al menos el 50 % o cualquier intervalo intermedio, de inhibición de un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK) con respecto a la de un inhibidor de ALK conocido en las mismas condiciones de ensayo. En determinadas realizaciones, las células pueden transfectarse con una construcción que codifica un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK), ponerse en contacto con un compuesto de prueba que está etiquetado o marcado con un marcador detectable, y analizarse la presencia del compuesto de prueba enlazado. En determinadas realizaciones, se observa que las células transfectadas se unen al compuesto de prueba en comparación con células que no se han transfectado con un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK), lo que es una indicación de que el compuesto de prueba se está uniendo a un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK) expresado por esas células. La unión del compuesto se determina normalmente mediante uno cualquiera de una amplia diversidad de ensayos conocidos en la técnica, tal como ELISA, RIA, y/o ensayos de BIAcore.

Los compuestos pueden explorarse para los efectos inhibidores, u otros, sobre la actividad de un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK) utilizando una versión recombinante de la enzima expresada o un homólogo u ortólogo aislado de otra especie. Como alternativa, las células que expresan uno de estos polipéptidos biomarcadores nuevos pueden tratarse con un compuesto de prueba y puede determinarse el efecto del compuesto de prueba sobre la fosforilación de una diana específica, por ejemplo, utilizando una de las técnicas descritas en el presente documento. En otro ejemplo, se determina la actividad tirosina quinasa. Los métodos para determinar la actividad que afecta la fosforilación de la tirosina quinasa (por ejemplo, inhibición) son bien conocidos para los expertos en la materia. En algunos ejemplos, la actividad tirosina quinasa puede determinarse evaluando la incorporación de un fosfato marcado (tal como un fosfato marcado con <sup>32</sup>P) en un sustrato que tiene la capacidad de ser fosforilado por un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK) (por ejemplo, una proteína o un fragmento peptídico, en especial los de los componentes de señalización aguas abajo). En otras realizaciones, puede medirse la actividad tirosina quinasa utilizando un kit universal de actividad tirosina quinasa (por ejemplo, Kit Universal Tyrosine Kinase Assay (Takara Bio, Inc., Madison, Wis.); Kit Tyrosine Kinase Assay (Millipore, Billerica, Mass.)).

En otra realización, se proporcionan métodos de exploración que implican determinar adicionalmente si el compuesto reduce el crecimiento de célula tumorales, por ejemplo, células tumorales que se sabe que expresan una mutación de tirosina quinasa activada, tal como un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK). Pueden utilizarse diversas líneas celulares, que pueden seleccionarse a base del tejido a analizar, que son conocidas para los expertos en la materia (por ejemplo, células BA/F3). Por ejemplo, muchas líneas celulares están bien caracterizadas y por ejemplo, el United States National Cancer Institute (NCI) las utiliza en su programa de exploración de nuevos fármacos antineoplásicos.

Una inhibición significativa y estadísticamente significativa del crecimiento de células tumorales, tal como se produce para más de aproximadamente el 50 % a una dosis de 100  $\mu$ M, 90  $\mu$ M, 80  $\mu$ M, 70  $\mu$ M, 60  $\mu$ M, 50  $\mu$ M, 40  $\mu$ M, 30  $\mu$ M, 20  $\mu$ M, 10  $\mu$ M, 9  $\mu$ M, 8  $\mu$ M, 7  $\mu$ M, 6  $\mu$ M, 5  $\mu$ M, 4,5  $\mu$ M, 4  $\mu$ M, 3,5  $\mu$ M, 3  $\mu$ M, 2,5  $\mu$ M, 2  $\mu$ M, 1,5  $\mu$ M, 1  $\mu$ M, 900 nM, 850 nM, 800 nM, 750 nM, 700 nM, 650 nM, 600 nM, 550 nM, 500 nM, 450 nM, 400 nM, 350 nM, 300 nM, 250 nM, 200 nM, 150 nM, 100 nM, 95 nM, 90 nM, 85 nM, 80 nM, 75 nM, 70 nM, 65 nM, 60 nM, 55 nM, 50 nM, 45 nM, 40 nM, 35 nM, 30 nM, 25 nM, 20 nM, 15 nM, 10 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM o por debajo, es adicionalmente indicativa de que el compuesto es útil para tratar lesiones neoplásicas. Puede determinarse un valor de CI<sub>50</sub> y utilizarse para fines comparativos. Este valor es la concentración de fármaco necesaria para inhibir el crecimiento de células tumorales en un 50 % con respecto al control.

Estos valores pueden adicionalmente aplicarse a otros criterios. Por ejemplo, en otras realizaciones, los métodos de exploración proporcionados en el presente documento implican adicionalmente determinar si el compuesto de prueba induce apoptosis en cultivos de células tumorales. Pueden describirse por criterios morfológicos y bioquímicos dos formas distintas de muerte celular: necrosis y apoptosis. La necrosis está acompañada de una permeabilidad aumentada de la membrana plasmática, mediante lo cual las células se hinchan y la membrana plasmática se rompe al cabo de minutos. La apoptosis está caracterizada por la formación de protuberancias en la membrana, la condensación de citoplasma y la activación de endonucleasas endógenas.

La apoptosis se produce de forma natural durante la renovación normal de los tejidos y durante el desarrollo embrionario de los órganos y las extremidades. La apoptosis también puede estar inducida por diversos estímulos, incluyendo los linfocitos T citotóxicos y los linfocitos citolíticos naturales, por radiación ionizante y por determinados fármacos quimioterapéuticos. Se cree que la regulación inapropiada de la apoptosis desempeña un papel importante en muchas afecciones patológicas incluyendo el cáncer, el SIDA o la enfermedad de Alzheimer, etc.

Los compuestos de prueba pueden explorarse para la inducción de la apoptosis utilizando cultivos de células tumorales mantenidas en las condiciones descritas anteriormente. En algunas ejemplos de tales métodos de exploración, el tratamiento de las células con compuestos de prueba implica cultivos pre o posconfluentes y el tratamiento durante uno a siete días a diversas concentraciones de los compuestos de prueba. Las células apoptóticas pueden medirse tanto en las porciones adheridas como en las "flotantes" de los cultivos. Ambas se recogen retirando el sobrenadante, tratando con tripsina las células adheridas y combinando ambas preparaciones después de una etapa de lavado por centrifugación (10 minutos, 200 rpm). Después del tratamiento con un compuesto de prueba, los cultivos pueden evaluarse para apoptosis y necrosis, por ejemplo mediante microscopía de fluorescencia seguido de marcaje con naranja de acridina y bromuro de etidio. Los expertos en la materia conocen muchos métodos para medir células apoptóticas, por ejemplo, se ha descrito un método para medir la cantidad de células apoptóticas en Duke y Cohen (Curr. Prot. Immuno., Coligan *et al.*, eds., 3.17.1-3.17.1, 1992). Por ejemplo, las células flotantes y adheridas se recogen mediante tratamiento con tripsina y se lavan tres veces en PBS. Después se centrifugan alcuotas de las células. Los sedimentos se resuspenden en medio y en una mezcla de colorantes que contiene naranja de acridinina y bromuro de etidio preparada en PBS, y se mezcla suavemente. Después, la mezcla puede colocarse en la platina de un microscopio y examinarse las características morfológicas de apoptosis. Además, la apoptosis puede cuantificarse midiendo un aumento de la fragmentación del ADN en las células que se han tratado con los compuestos de prueba. Están disponibles inmunoensayos enzimáticos (EIA) fotométricos comerciales para la determinación cuantitativa *in vitro* de fragmentos citoplásmáticos de ADN asociados con histonas (mono- y oligonucleosomas) (por ejemplo, ELISA de Detección de Muerte Celular, Boehringer Mannheim).

En realizaciones adicionales, los métodos de exploración proporcionados en el presente documento incluyen adicionalmente determinar si el compuesto de prueba disminuye la metástasis tumoral, por ejemplo, en un modelo animal de metástasis. Los métodos de evaluación de metástasis tumoral son bien conocidos en la técnica (véase, por ejemplo, Khanna y Hunter, Carcinogenesis 26: 513-523, 2005). Un modelo de metástasis implica xenoinjertos de humano-ratón, en los cuales se trasplantan líneas celulares o tejidos de cáncer humanos en ratones inmunodeficientes (tales como ratones SCID o ratones desnudos). En métodos similares, puede trasplantarse una línea celular que se ha diseñado técnicamente para expresar un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK) en un ratón inmunodeficiente. En un ejemplo, se inyectan células o líneas celulares tumorales de forma directa en la circulación general. El sitio de inyección define en gran medida el sitio en el que se desarrolla la metástasis en estos sistemas experimentales. El sitio más común de inyección de células tumorales empleado para los modelos de metástasis experimental es la vena lateral de la cola de ratones, lo que principalmente da como resultado metástasis pulmonares. Por el contrario, la inyección de células tumorales intraesplénica o en la vena porta es el sitio más comúnmente empleado para desarrollar metástasis en el hígado y la inyección intracardiaca de células puede dar como resultado metástasis en varios sitios, incluyendo el hueso. Después de la inyección de células tumorales u otras líneas celulares en la circulación, se controla el desarrollo de metástasis en el sitio de interés (tal como pulmón) a lo largo de un periodo de días o semanas.

Otro modelo para evaluar la metástasis tumoral utiliza el trasplante ortotópico, en el que las células cancerosas se trasplantan en el emplazamiento anatómico o tejido del que se obtuvo el tumor (por ejemplo, por inyección directa o implante quirúrgico de fragmentos tumorales). Las metástasis espontáneas que surgen del tumor ortotópico pueden evaluarse a lo largo de un periodo de días o semanas. La capacidad de un compuesto de prueba para disminuir o prevenir la metástasis tumoral puede evaluarse administrando un compuesto de prueba a un animal después de la inyección de células tumorales por vía subcutánea, vía intramuscular o en la circulación, o por trasplante ortotópico. Puede evaluarse el número, tamaño o tiempo de desarrollo de la metástasis. Un compuesto que inhibe la metástasis tumoral puede disminuir el número de metástasis, por ejemplo, en al menos el 10 %, al menos el 20 %, al menos el 30 %, al menos el 40 %, al menos el 50 %, al menos el 55 %, al menos el 60 %, al menos el 65 %, al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 % o incluso el 100 %, en comparación con una muestra de control. Un compuesto que inhibe la metástasis tumoral también puede disminuir el tamaño de la metástasis en comparación con una muestra de control. De forma similar, un compuesto que inhibe la metástasis tumoral puede retrasar el inicio de desarrollo de la metástasis, por ejemplo en al menos una semana, dos semanas, un mes, seis meses, un año o incluso de forma indefinida.

## VII. Inhibidores de ALK ejemplares

Los métodos divulgados en el presente documento incluyen la identificación de un sujeto como un candidato para el tratamiento con un inhibidor de un biomarcador nuevo de la presente invención enumerado en la Tabla 1 (por ejemplo, mutantes de ALK) para inducir la muerte de las células tumorales, reducir el crecimiento tumoral o disminuir el riesgo de metástasis tumoral. Los inhibidores de los polipéptidos ALK son conocidos para un experto en la materia. Por ejemplo, PF-02341066, PDD, 2-metil-11-(2-metilpropil)-4-oxo-4,5,6,11,12,13-hexahidro-2H-indazol[5,4-

a)pirrol[3,4-c]carbazol-8-il [4-(dimetilamino)benzil]carbamato,(1*S*,2*S*,3*R*,4*R*)-3-({5-cloro-2-[(1-etil-2,3,4,5-tetrahidro-6-metoxi-2-oxo-1*H*-1-benzazepin-7-il)amino]-4-pirimidinil}amino)biciclo[2.2.1]hept-5-en-2-carboxamida y NVP-TAE684 9, véase, por ejemplo, PNAS 104:270-275, 2007; Choi, Y.L. *et al.* (2008) Cancer Res. 68: 4971-2976 y Biochemistry 48: 3600-3609, 2009.

## EJEMPLIFICACIÓN

La invención se ilustra adicionalmente mediante los siguientes ejemplos que no deben considerarse como limitativos.

### Ejemplo 1 - Materiales y métodos para los ejemplos 2-4

#### a. Secuenciación de ADN

Se generaron los ADNc cebados con oligo(dT) a partir de los ARN de muestras de ensayo extraídos con el uso del sistema EZ1 (Qiagen, Valencia, CA) y se sometieron a una reacción en cadena de la polimerasa (PCR) de 30 ciclos (que consistían en 90 °C durante 10 s y 68 °C durante 1 min) con ADN polimerasa PrimeSTAR HS (Takara Bio Inc., Shiga, Japón) y los cebadores ALK-TK-F (5'-TACAACCCCAACTACTGCTTTGCT-3') y ALK-TK-R1 (5'-AGGCACTTTCTCTTCTTCCAC-3'). Los productos de PCR correspondientes al dominio quinasa de la ALK se fragmentaron después y se secuenciaron con el Genome Analyzer II (GAII) de Illumina para 76 bases desde ambos extremos mediante el sistema de secuenciación de extremos emparejados (Illumina, San Diego, CA). Se filtró la calidad de los datos de lectura sin procesar a base de la presencia de las secuencias de los cebadores de PCR y un valor Q de  $\geq 20$  para todas las bases. Las lecturas pasadas por filtro se alinearon después con la secuencia del ADNc de ALK con el uso del algoritmo Bowtie (disponible en internet en [bowtie-bio.sourceforge.net/index.shtml](http://bowtie-bio.sourceforge.net/index.shtml)).

Para la secuenciación capilar con 3130x1 Genetic Analyzer (Applied Biosystems, Foster City, CA), se prepararon productos PCR a partir de los ADNc con el mismo conjunto de cebadores o con la combinación de los cebadores EA-F-g-S (5'-CCACACCTGGGAAAGGACCTAAAG-3') y ALK-TK-R2 (5'-CCTCCAATACTGACAGCCACAGG-3').

#### b. EML4-ALK mutante

Se insertó un ADNc que codificaba EML4-ALK variante 1 etiquetado con el epítipo FLAG (Soda, M. *et al.* (2007) Nature 448: 561-566) en el vector retroviral pMX-iresCD8 (Yamashita Y. *et al.* (2001) J. Biol. Chem. 276: 39012-39020) para la expresión simultánea de EML4-ALK etiquetada con FLAG y de CD8 de ratón. Los cambios de nucleótidos correspondientes a las mutaciones C1156Y y L1196M de ALK se introdujeron en el plásmido de forma individual o en combinación para la expresión de EML4-ALK(C1156Y), EML4-ALK(L1196M) o EML4-ALK(C1156Y/L1196M). Los retrovirus recombinantes basados en estos plásmidos se generaron con el uso de la línea celular empaquetadora BOSC23 (Pear, W.S. *et al.* (1993) Proc. Natl. Acad. Sci. USA 90: 8392-8396) y se utilizaron para infectar la línea celular BA/F3, dependiente de interleucina 3 (Palacios, R. *et al.* (1985) Cell 41: 727-734). Las células positivas para CD8 resultantes se purificaron con el uso de una columna de separación de células miniMACS y perlas magnéticas conjugadas con anticuerpos frente a CD8 (ambas de Miltenyi Biotec, Gladbach, Alemania). PF-02341066 se obtuvo de Selleck.

Para el examen de la fosforilación de tirosinas de EML4-ALK, se expusieron células BA/F3 que expresaban la proteína de fusión a inhibidores de ALK durante 15 h, tras lo cual se inmunoprecipitó EML4-ALK de los lisados celulares con anticuerpos frente a FLAG (Sigma-Aldrich, St. Louis, MO) y se sometió a análisis por inmunotransferencia con anticuerpos frente a ALK fosforilada en Tyr<sup>1604</sup> (Cell Signaling Technology, Danvers, MA). Se realizó un ensayo de quinasa *in vitro* a temperatura ambiente durante 30 min como se describe anteriormente (Donella-Deana, A. *et al.* (2005) Biochemistry 44: 8533-8542), con el péptido sintético YFF (Operon Biotechnologies, Huntsville, AL).

### Ejemplo 2 - Mutaciones de ALK nuevas asociadas con la resistencia a los inhibidores de la tirosina quinasa ALK

El paciente era un hombre de 28 años sin antecedentes como fumador y al que se le había diagnosticado un adenocarcinoma de pulmón en una fase clínica T4N3M1 en abril de 2008. Dado que el tumor no portaba ninguna de las mutaciones de EGFR, el paciente se trató con quimioterapia convencional, lo que dio como resultado una progresión de la enfermedad con la formación de múltiples metástasis en el cerebro y hueso. En noviembre de 2008, la presencia de ARNm para EML4-ALK variante 1 en el tumor se confirmó por análisis por transcripción inversa-PCR de esputo así como por análisis de hibridación de fluorescencia *in situ* de una muestra de ensayo de biopsia. El paciente se incorporó así a un ensayo de PF-02341066 y experimentó una notable mejora de su estado general (reducción del nivel 4 a 2). Aunque mostró una "respuesta parcial" al tratamiento, no se había erradicado totalmente su derrame pleural. Tras 5 meses de tratamiento, sin embargo, el tumor comenzó abruptamente a crecer de nuevo, dando como resultado un aumento del derrame pleural y la formación de múltiples nódulos cancerosos en ambos pulmones. El paciente se retiró del ensayo en mayo de 2009 y después se obtuvo derrame pleural para el análisis molecular.

Dado que el tumor reanudó el crecimiento a pesar de la administración sostenida del inhibidor de ALK, se determinó si el tumor había adquirido cambios genéticos secundarios que conferirían resistencia al fármaco. Además, dado que la resistencia a los TKI a menudo es el resultado de mutaciones adquiridas dentro de las quinasas diana, se examinó la posibilidad de que la propia EML4-ALK hubiera experimentado cambios de aminoácidos.

Estaba disponibles muestras de ensayo de esputo (ID J-#1) y derrame pleural (ID J-#113) para el análisis molecular del tumor del paciente antes y después del tratamiento, respectivamente. Dado que la proporción de células tumorales en las dos muestras de ensayo podían haber diferido, se utilizó un secuenciador de última generación para realizar secuenciación profunda de los ADNc de EML4-ALK obtenidos de estas muestras de ensayo. Los ADNc que correspondían al dominio de tirosina quinasa de ALK se amplificaron así a partir de ambas muestras de ensayo (Figura 1A), se fragmentaron y se sometieron a secuenciación de nucleótidos con el sistema GAI. Para la comparación, se analizaron de forma similar una línea celular de CPNM positiva para EML4-ALK, H2228 y otras tres muestras de ensayo clínicas positivas también para la proteína de fusión. Se detectó un polimorfismo de nucleótido único conocido, rs3795850 en los ADNc de cuatro muestras de ensayo (Figura 1B). Además, el cambio T→C en una posición que correspondía al nucleótido 4230 del ADNc de la ALK de tipo silvestre (número de referencia de GenBank, NM\_004304) se detectó con una baja frecuencia (8,9 %) en los ADNc de J-#1. Además, se detectaron dos modificaciones nuevas, los cambios G→A y C→A en las posiciones que correspondían a los nucleótidos 4374 y 4493 del ADNc de ALK de tipo silvestre, con frecuencias del 41,8 y 14,0 %, respectivamente, en los ADNc de J-#113. No hubo otras modificaciones recurrentes (presentes en ≥5 % de las lecturas) en los ADNc del dominio quinasa obtenidos de cualquiera de las muestras de ensayo.

Estos cambios de nucleótidos se confirmaron posteriormente utilizando un secuenciador de Sanger. Para excluir la posibilidad de que las mutaciones hubiesen aparecido en la ALK de tipo silvestre endógena en lugar de en EML4-ALK, también se realizó PCR con un cebador directo dirigido al ADNc de EML4 de forma que solo se amplificara el ADNc de fusión (Figura 1A). El cambio T4230C no se detectó entre los cientos de ADNc de fusión obtenidos de J-#1, lo que indica que era un artefacto que surgió en la PCR inicial o en la etapa de secuenciación por GAI.

Sin embargo, ambos cambios G4374A y C4493A se confirmaron fácilmente mediante secuenciación de Sanger. Entre los 73 clones de ADNc de fusión secuenciados para J-#113, 34 clones (46,6 %) eran positivos para G4374A, 11 (15,1 %) eran positivos para C4493A y el resto (38,4 %) eran del tipo de silvestre (Figura 1C). Mientras que el análisis por PCR cubría ambas posiciones de nucleótidos en los mismos productos, ninguno de los productos contuvo ambas mutaciones, lo que indica que cada mutación apareció de forma independiente. Los fragmentos genómicos que abarcaban las posiciones G4374 o C4493 también se amplificaron por PCR y se sometieron a secuenciación de nucleótidos, dando como resultado la confirmación de cada cambio en el genoma tumoral (Figura 2).

Las sustituciones G4374A y C4493A dan como resultado los cambios Cys→Tyr y Leu→Met en las posiciones que corresponden a los aminoácidos 1156 y 1196, respectivamente, de la ALK humana de tipo silvestre.

Ejemplo 3 - Las mutaciones de ALK nuevas confieren resistencia a inhibidores de la tirosina quinasa ALK

A continuación, se examinó si tales cambios de aminoácidos afectan la sensibilidad de EML4-ALK a los inhibidores de ALK. Se expresaron de forma individual EML4-ALK de tipo silvestre, los mutantes simples EML4-ALK(C1156Y) y EML4-ALK(L1196M) y el doble mutante EML4-ALK(C1156Y/L1196M) en células BA/F3 y las células se expusieron después a inhibidores de ALK. PF-02341066 inhibió de una manera dependiente de la concentración el crecimiento de las células BA/F3 que expresaban EML4-ALK de tipo silvestre (Figura 4A). Por el contrario, las células que expresaban los mutantes C1156Y o L1196M manifestaron una sensibilidad a este fármaco notablemente reducida, mostrando repetidos experimentos que las células BA/F3 que expresaban EML4-ALK(L1196M) eran más resistentes a PF-02341066 que las que expresaban EML4-ALK(C1156Y) (Figura 3). La presencia de ambas mutaciones no dio como resultado un efecto aditivo sobre la resistencia de las células a PF-02341066. Por lo tanto, estos datos demostraron que las mutaciones C1156Y y L1196M confieren cada una resistencia a este fármaco.

Se examinó la fosforilación de tirosinas de EML4-ALK mediante análisis de inmunotransferencia con anticuerpos específicos para ALK fosforilada en Tyr1604. Aunque la exposición de las células BA/F3 a PF-02341066 inhibió notablemente la fosforilación de tirosinas en EML4-ALK de tipo silvestre, no tuvo efecto sustancial sobre la de EML4-ALK(C1156Y) o EML4-ALK(L1196M) (Figura 4B). De forma coherente con esos hallazgos, un ensayo de quinasa *in vitro* reveló que los mutantes C1156Y y L1196M de EML4-ALK eran menos sensibles a la inhibición de la actividad enzimática por PF-02341066 que la proteína de tipo silvestre (Figura 4C). Como fue el caso para la inhibición del crecimiento celular (Figura 4A), el mutante L1196M era más resistente a la inhibición de la actividad quinasa por PF-02341066 que el mutante C1156Y (Figura 4C).

Ejemplo 4 - Relaciones de estructura-función entre las mutaciones de ALK nuevas y la resistencia a los inhibidores de la tirosina quinasa ALK

La Figura 5 muestra las posiciones Cys1156 y Leu1196 en un modelo de estructura tridimensional del dominio quinasa de ALK basado en la estructura cristalina de una quinasa relacionada, el receptor insulina. El anterior resto

5 se posiciona de forma adyacente al extremo amino de la hélice  $\alpha$ C predicha así como cerca de la cubierta superior del bolsillo de unión a ATP. No se han informado mutaciones activadoras en esta posición en otras tirosina quinasas. Leu1196 de ALK corresponde a Thr315 de ALB1 y Thr790 de EGFR, cada una de las cuales es el sitio de mutaciones adquiridas más frecuentes que confieren resistencia a los TKI en estas quinasas (Deininger, M. *et al.* (2005) *Blood* 105: 2640-2653; Linardou, H. *et al.* (2009) *Nat. Rev. Clin. Oncol.* 6: 352-366). Este sitio "guardián" se emplaza en la superficie del fondo del bolsillo de unión a ATP (Figura 5) y se sabe que la presencia en esta posición de un aminoácido con una cadena lateral voluminosa interfiere con la unión de muchos de TKI (Shah, N.P. *et al.* (2002) *Cancer Cell* 2: 117-125; Tsao, M.S. *et al.* (2005) *N. Engl. J. Med.* 353: 133-144).

10 Por lo tanto, se identificaron dos mutaciones *de novo* dentro del dominio quinasa de EML4-ALK que confieren resistencia a múltiples inhibidores de ALK. Dado que no se observaron ADNc de EML4-ALK que portaban ambas mutaciones, se cree que cada mutación se desarrolló de forma independiente en distintos subclones del tumor.

15 Sin ceñirse a teoría alguna, dado que los ADNc preparados de esputo del paciente antes del tratamiento no contenían cambios de nucleótidos que correspondían con las mutaciones C1156Y o L1196M, es probable que los subclones del tumor adquirieran las mutaciones *de novo* durante el tratamiento con PF-02341066. Sin embargo, debido a que el derrame pleural no pudo examinarse antes del tratamiento, no puede excluirse de forma completa la posibilidad de que las células tumorales que portaban los mutantes C1156Y o L1196M estuvieran ya presentes en el derrame pleural en la admisión inicial del paciente. Si este era el caso, el tumor podría haber adquirido otras mutaciones, aún desconocidas, durante el periodo de tratamiento de 5 meses con PF-02341066 que permitieran su posterior crecimiento rápido. Sin embargo, los subclones de células tumorales con las mutaciones C1156Y o L1196M deberían haber sido resistentes al tratamiento inicial y deberían haberse expandido durante el curso del tratamiento. Por el contrario, no hubo signos de expansión tumoral en el paciente durante al menos 5 meses, indicando que las mutaciones C1156Y y L1196M se desarrollaron durante el tratamiento con PF-02341066. Esta idea está avalada adicionalmente por el hecho de que la mutación T790M de EGFR, que confiere resistencia a gefitinib o erlotinib, se detecta frecuentemente en pacientes tratados anteriormente con los TKI pero raramente se encuentra en individuos no tratados (Pao, W. *et al.* (2005) *PLoS Med.* 2:e73).

30 Se han detectado sustituciones de aminoácidos en la posición de guardián de varias tirosina quinasas en tumores tratados con TKI (Kobayashi, S. *et al.* (2005) *N. Engl. J. Med.* 352: 786-792; Pao, W. *et al.* (2005) *PLoS Med.* 2:e73; Shah, N.P. *et al.* (2002) *Cancer Cell* 2: 117-125; Cools, J. *et al.* (2003) *N. Engl. J. Med.* 348: 1201-1214; Tamborini, E. *et al.* (2004) *Gastroenterology* 127: 294-299). Mientras que no se han informado anteriormente mutaciones en este sitio para EML4-ALK o ALK, se han examinado recientemente los efectos de diversas sustituciones artificiales de aminoácidos en la posición de guardián de NPM-ALK, otra oncoquinasa de tipo fusión para ALK (Lu, L. *et al.* (2009) *Biochemistry* 48: 3600-3609). De forma coherente con el presente análisis de células tumorales *in vivo*, se descubrió que la introducción de una Met en esta posición hace que NPM-ALK sea más resistente a múltiples inhibidores de ALK.

40 En contraste con las sustituciones del guardián, no se han informado para otras tirosina quinasas mutaciones activadoras en la posición inmediatamente amino terminal con respecto a la hélice  $\alpha$ C (Cys1156 en ALK). Aunque se describió en un caso de CPNM un cambio Thr→Ile en la correspondiente posición de EGFR, no se examinó su importancia para la sensibilidad a fármacos (Tsao, M.S. *et al.* (2005) *N. Engl. J. Med.* 353: 133-144). La importancia de la hélice  $\alpha$ C para la regulación alostérica de la actividad enzimática se ha demostrado para las serina-treonina quinasas (Hindie, V. *et al.* (2009) *Nat. Chem. Biol.* 5: 758-764). Un cambio en la Cys1156 de ALK podría, por lo tanto, interferir alostéricamente con la unión del TKI o la Cys1156 podría estar directamente implicada en la interacción física entre el dominio quinasa y los TKI.

#### Equivalentes

50 Los expertos en la materia reconocerán o tendrá la capacidad de determinar, utilizando no más que experimentación de rutina, muchos equivalentes de las realizaciones específicas de la presente invención descrita en el presente documento.

## REIVINDICACIONES

1. Un método para identificar a un sujeto que tiene cáncer o en riesgo de desarrollar un cáncer, como que tiene un riesgo aumentado de falta de respuesta al tratamiento con un inhibidor de ALK, que comprende:
- 5 analizar una muestra obtenida del sujeto para detectar la presencia de una o más moléculas de polinucleótido ALK mutantes que codifican un polipéptido ALK mutante, que comprende una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre o para detectar la presencia de uno o más polipéptidos ALK mutantes que comprenden una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre,
- 10 en el que la presencia de las una o más moléculas de polinucleótido de ALK mutantes o polipéptidos ALK mutantes indica que el sujeto tiene un riesgo aumentado de falta de respuesta al tratamiento con el inhibidor de ALK.
- 15 2. El método de la reivindicación 1, en el que el sujeto no ha sido tratado anteriormente con un inhibidor de ALK, o ha sido tratado anteriormente con un inhibidor de ALK y ha desarrollado al menos una resistencia parcial al inhibidor de ALK.
- 20 3. El método de la reivindicación 1, en el que el cáncer se selecciona del grupo que consiste en linfoma anaplásico de células grandes, neuroblastoma, cáncer de mama, cáncer colorrectal, tumores miofibroblásticos inflamatorios y cánceres de pulmón no microcíticos.
- 25 4. El método de la reivindicación 1 o 2, en el que el inhibidor de ALK se selecciona del grupo que consiste en PF-02341066, PDD, 2-metil-11-(2-metilpropil)-4-oxo-4,5,6,11,12,13-hexahidro-2H-indazol[5,4-a]pirrol[3,4-c]carbazol-8-il [4-(dimetilamino)benzil]carbamato, (1S,2S,3R,4R)-3-({5-cloro-2-[(1-etil-2,3,4,5-tetrahidro-6-metoxi-2-oxo-1H-1-benzazepin-7-il)amino]-4-pirimidinil}amino)biciclo[2.2.1]hept-5-en-2-carboxamida y NVP-TAE684.
- 30 5. El método de la reivindicación 1, en el que la muestra se selecciona del grupo que consiste en esputo, lavado broncoalveolar, derrame pleural, tejido, sangre entera, suero, plasma, raspado bucal, saliva, líquido cefalorraquídeo, orina, heces, células tumorales en circulación, ácidos nucleicos en circulación y médula ósea.
- 35 6. El método de la reivindicación 5, en el que la muestra es tejido; y el tejido es un tejido tumoral o canceroso.
7. El método de la reivindicación 1, en el que la muestra comprende células.
- 40 8. El método de la reivindicación 1, en el que las una o más moléculas de polinucleótido o polipéptidos de ALK mutantes se seleccionan del grupo que consiste en moléculas de polinucleótido de ALK mutantes que codifican una proteína ALK mutante que comprende una mutación Cys1156Tyr y polipéptidos ALK mutantes que comprenden una mutación Cys1156Tyr.
- 45 9. El método de la reivindicación 1, en el que las una o más mutaciones de ALK se evalúan por un ensayo de hibridación de ácidos nucleicos o por reacción en cadena de la polimerasa.
- 50 10. El método de la reivindicación 1, en el que la presencia de los uno o más polipéptidos ALK se detecta utilizando un reactivo que se une de forma específica a uno o más polipéptidos ALK, de forma opcional, en el que el reactivo se selecciona del grupo que consiste en un anticuerpo, un derivado de anticuerpo y un fragmento de anticuerpo.
- 55 11. El método de la reivindicación 1, en el que las una o más mutaciones de ALK se evalúan en un primer punto temporal y al menos un punto temporal posterior.
- 60 12. El método de la reivindicación 1, en el que la muestra comprende ADN genómico de estirpe germinal o somático.
13. Un método *in vitro* de determinación de si un compuesto de prueba modula la actividad de uno o más polipéptidos ALK mutantes que comprende:
- (a) poner en contacto células de mamífero transfectadas con una construcción que codifica los uno o más polipéptidos ALK mutantes que comprenden una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre con el compuesto de prueba; y
- (b) evaluar la actividad del uno o más polipéptidos ALK mutantes en las células de mamífero, en las que la actividad modulada de forma significativa en presencia del compuesto de prueba con respecto a un experimento de control identifica al compuesto de prueba como un modulador de los uno o más polipéptidos ALK mutantes.
- 65 14. El método de la reivindicación 13, en el que los uno o más polipéptidos mutantes son polipéptidos ALK mutantes que comprenden una mutación Cys1156Tyr.

15. El método de la reivindicación 13 o 14, en el que el control comprende células de mamífero que expresan un polipéptido ALK de tipo silvestre y, de forma opcional, en el que la actividad del uno o más polipéptidos ALK mutantes se seleccionan del grupo que consiste en: unión a ATP, actividad tirosina quinasa, proliferación de células cancerosas, crecimiento tumoral, número de tumores, apoptosis y metástasis tumoral.

5  
16. El método de la reivindicación 13 o 14, en el que el experimento de control comprende células de mamífero que expresan los uno o más polipéptidos ALK mutantes en ausencia del compuesto de prueba, de forma opcional, en el que la actividad de los uno o más polipéptidos ALK mutantes se selecciona del grupo que consiste en: unión a ATP, actividad tirosina quinasa, proliferación de células cancerosas, crecimiento tumoral, número de tumores, apoptosis y metástasis tumoral.

10  
17. Uso de una molécula de polinucleótido de ALK mutante que codifica un polipéptido ALK mutante que comprende una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre o el uso de un polipéptido ALK mutante que comprende una mutación en una posición que corresponde a la posición 1156 de la ALK humana de tipo silvestre en un método de diagnóstico *in vitro* del riesgo aumentado de un sujeto de falta respuesta a un tratamiento con un inhibidor de ALK.

15  
20  
18. El uso de una molécula de polinucleótido de ALK mutante o el uso de una molécula de polipéptido ALK mutante de acuerdo con la reivindicación 17, en el que la molécula de polinucleótido de ALK mutante o el polipéptido mutante se selecciona del grupo que consiste en una molécula de polinucleótido de ALK mutante que codifica una proteína ALK mutante que comprende una mutación Cys1156Tyr y polipéptidos ALK mutantes que comprenden una mutación Cys1156Tyr.

Figura 1

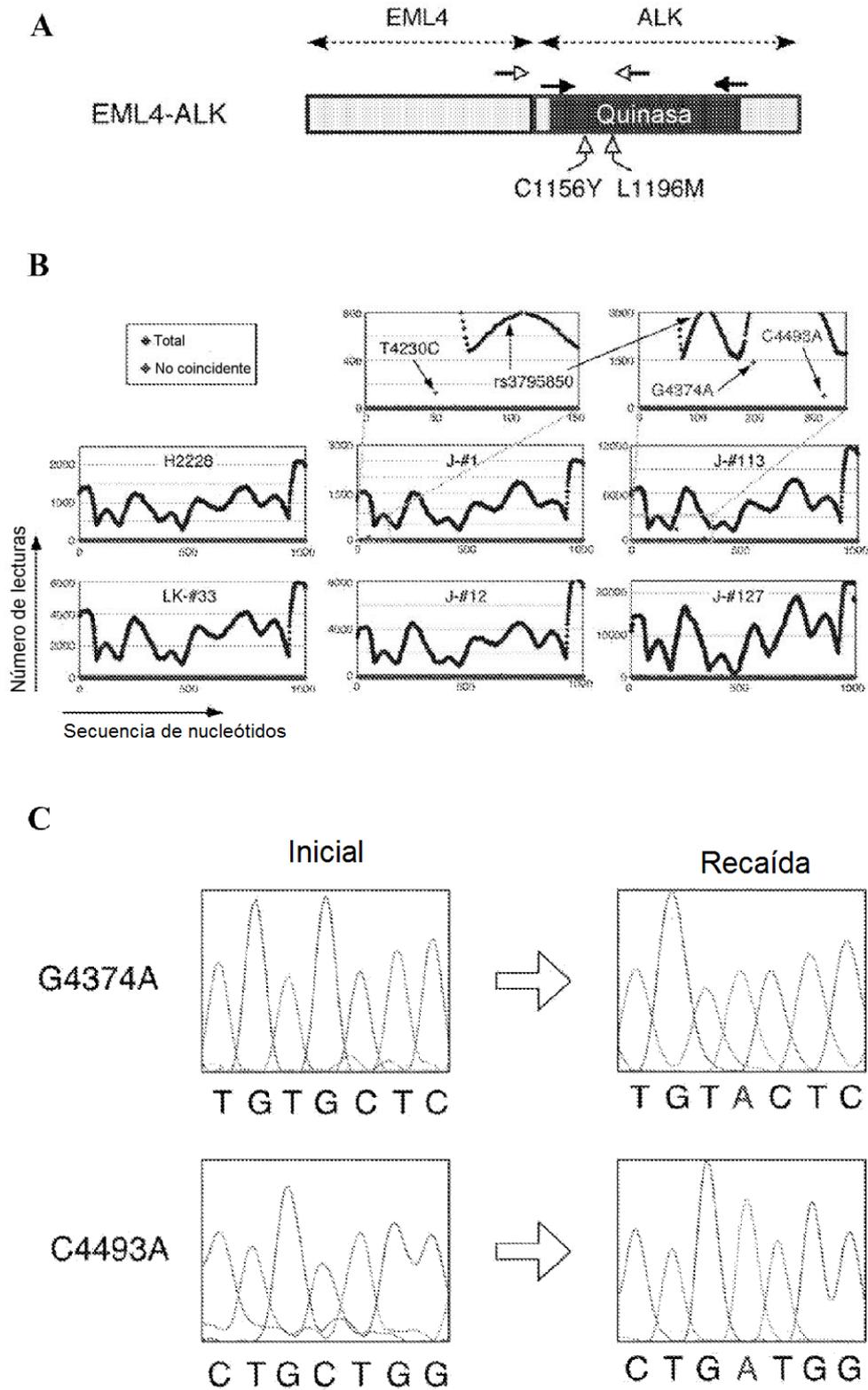


Figura 2

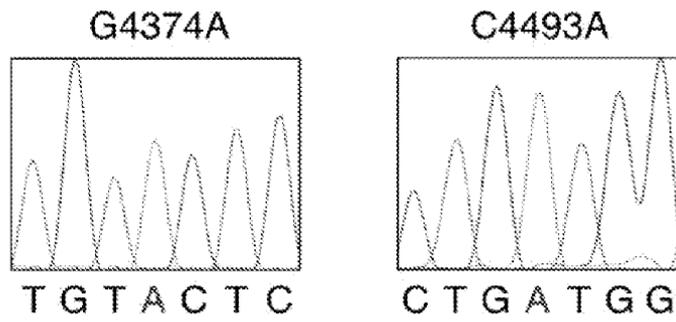
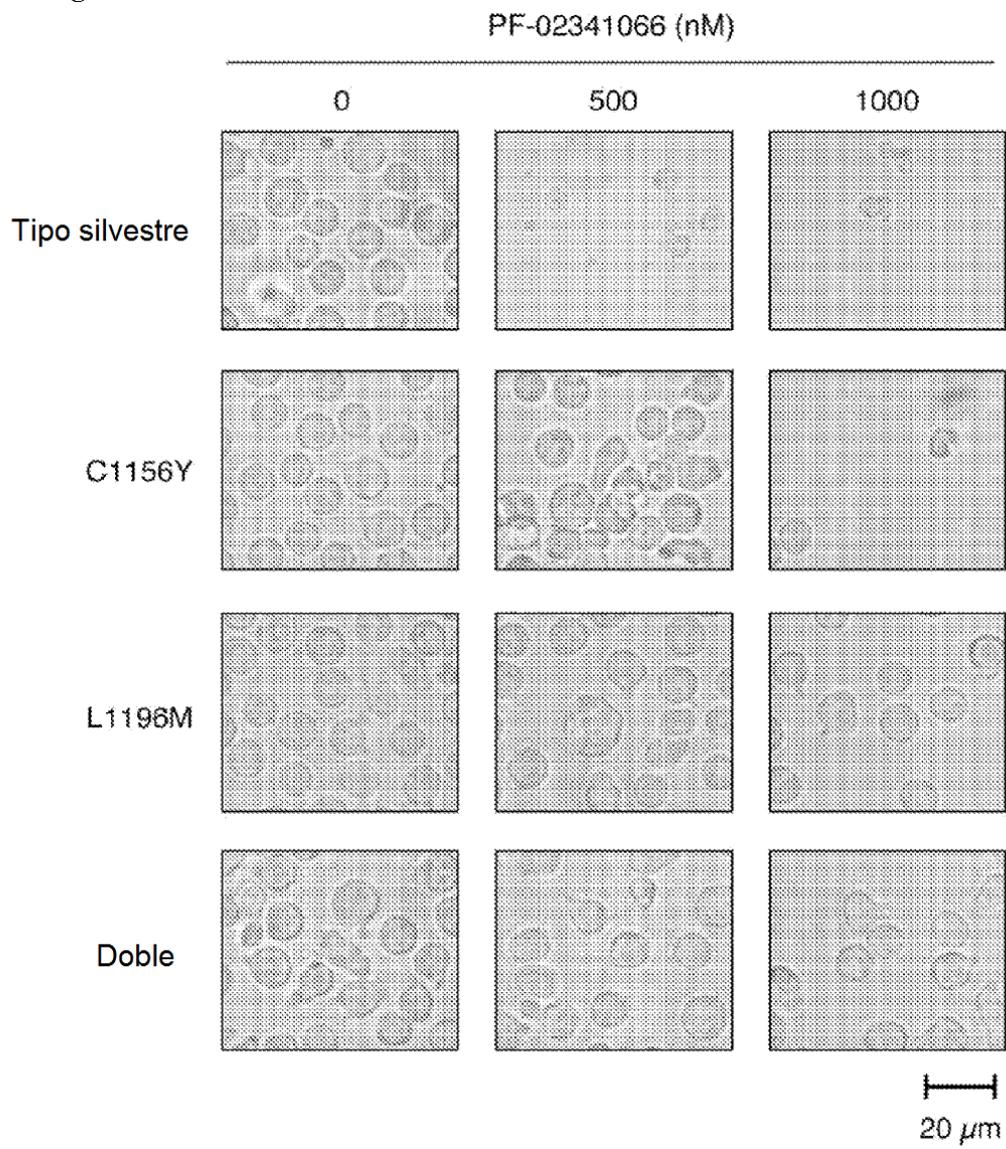


Figura 3



**Figura 4**

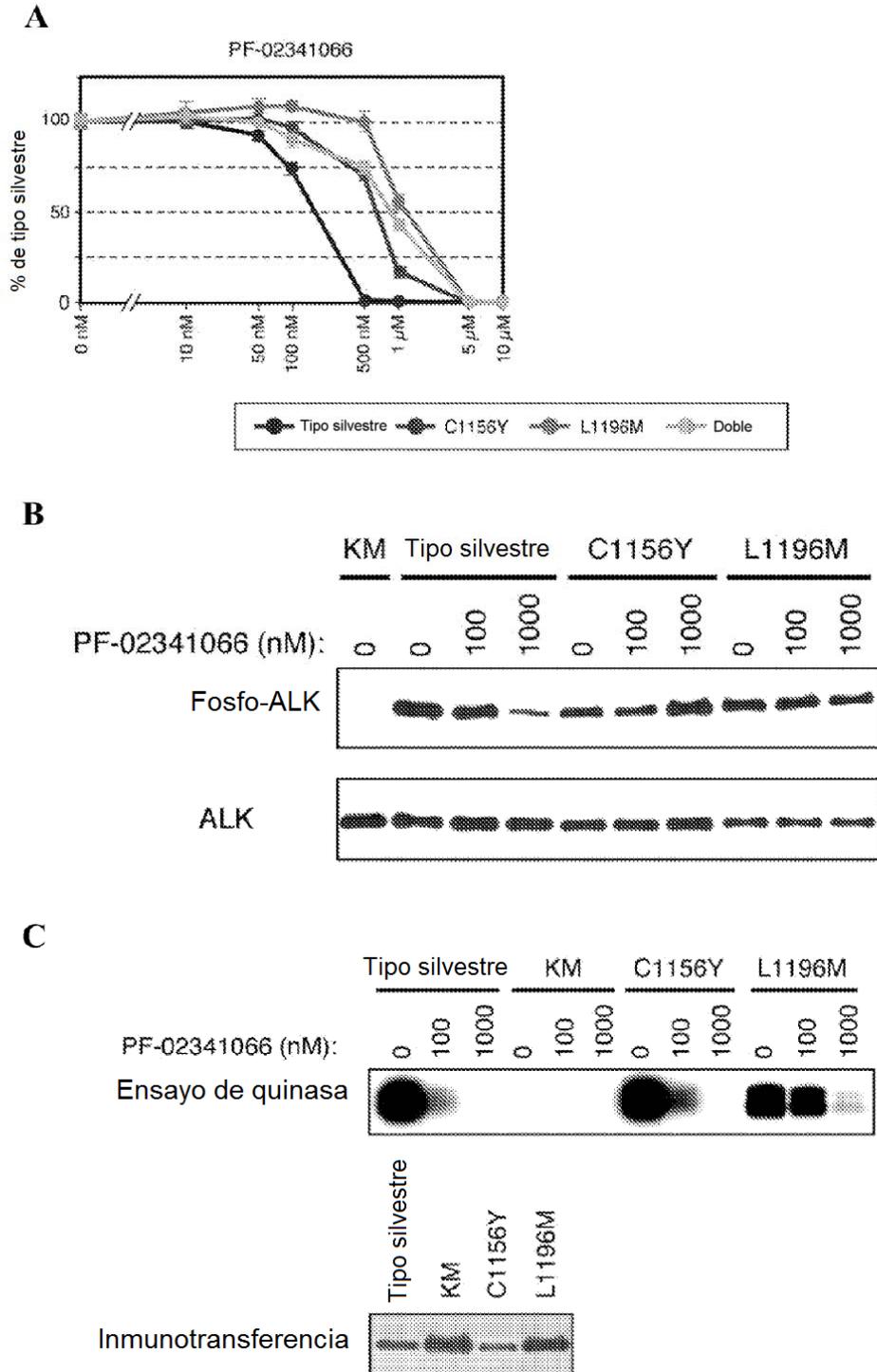


Figura 5

