

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 811 927**

51 Int. Cl.:

**C12N 9/12** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.04.2015 PCT/US2015/025289**

87 Fecha y número de publicación internacional: **15.10.2015 WO15157624**

96 Fecha de presentación y número de la solicitud europea: **10.04.2015 E 15777435 (7)**

97 Fecha y número de publicación de la concesión europea: **20.05.2020 EP 3129477**

54 Título: **Isoforma novedosa de cinasa de linfoma anaplásico y sus usos**

30 Prioridad:

**10.04.2014 US 201461978106 P**

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

**15.03.2021**

73 Titular/es:

**MEMORIAL SLOAN-KETTERING CANCER  
CENTER (100.0%)  
1275 York Avenue, Box 524  
New York, NY 10065, US**

72 Inventor/es:

**CHI, PING;  
WIESNER, THOMAS y  
CHEN, YU**

74 Agente/Representante:

**ARIAS SANZ, Juan**

ES 2 811 927 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

Isoforma novedosa de cinasa de linfoma anaplásico y sus usos

5 **1. Introducción**

La presente invención se refiere a una isoforma truncada de cinasa de linfoma anaplásico (TALK). La expresión de esta isoforma se asocia con cáncer y con la capacidad de respuesta a inhibidores de ALK. La detección de la isoforma puede usarse en métodos diagnósticos y terapéuticos. Debido a que surge como resultado de la transcripción de variantes en lugar de un reordenamiento genético, su presencia no se detectaría por pruebas genómicas.

15 **2. Antecedentes de la invención**

La caracterización integral de aberraciones genéticas subyacentes al cáncer humano es esencial para mejorar el diagnóstico tumoral, identificar dianas terapéuticas, desarrollar terapias combinadas racionales y optimizar los diseños de ensayos clínicos<sup>3</sup>. Los estudios de secuenciación a gran escala, como el proyecto Atlas del Genoma del Cáncer (TCGA), continúan revelando una imagen cada vez más detallada de las aberraciones genéticas en muchos tipos de cáncer, pero se centran principalmente en caracterizar aberraciones genéticas en las regiones codificantes del ADN<sup>2-5</sup>. Sin embargo, en una proporción significativa de tumores, el número de aberraciones genéticas detectables en oncogenes indicadores es demasiado bajo para explicar la transformación maligna<sup>6</sup>. Esto se ejemplifica en un estudio reciente, en el que el ~15% del adenocarcinoma pulmonar carecía de mutaciones génicas que afectaban a cualquiera de las características distintivas de cáncer<sup>7,8</sup>. Estas observaciones sugieren que mecanismos distintos de aberraciones genéticas pueden estar involucrados en la transformación maligna.

Isabelle Janoueix-Lerosey y otros "Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma", *Nature*, vol. 455, n.º 7215, 16 de octubre de 2018, páginas 967-970, el documento XP055004092 da a conocer que la línea celular de neuroblastoma humano CLB-BAR expresó altos niveles de una proteína de ALK de masa molecular menor, y que esta proteína de ALK truncada de CLB-BR es reactiva con el anticuerpo de ALK específico de Y1586, mientras que ALK de tipo silvestre no reacciona con este anticuerpo.

L Lobo De Figueiredo-Pontes *et al.* "Identification and characterization of ALK kinase splicing isoforms in nonsmall cell lung cancer", *Journal of Thoracic Oncology*, vol. 9, 1 de febrero de 2014, páginas 248-253, el documento XP055405883 da a conocer la identificación de dos isoformas de corte y empalme del dominio de cinasa de ALK, es decir, ALKde123 que omite el exón 23 y ALKde127 que omite el exón 27, en cáncer de pulmón de células no pequeñas.

Hiroyuki Mano *et al.* "EML4-ALK Fusion in Lung", *American Journal of Pathology*, vol. 176, n.º 3, 1 de marzo de 2010, páginas 1552-1554, el documento XP055406811 da a conocer que la longitud del intrón 19 del gen ALK es 1932 pb de largo.

El documento WO 2013/059740 describe una molécula de fusión DCTN1-ALK que comprende los exones 1-26 de DCTN1 y exones 20-29 de ALK, en la que la unión de fusión incluye una fusión en el marco que tiene un punto de interrupción dentro del intrón 26 de DCTN1 o fragmento del mismo y un punto de interrupción dentro del intrón 19 de ALK o un fragmento del mismo; se detectó la expresión de DCTN1-ALK en tumores de Spitz que van desde los nevos benignos de Spitz hasta melanomas spitzoides.

45 **3. Sumario de la invención**

La presente divulgación se refiere a una isoforma truncada de cinasa de linfoma anaplásico, también conocida como "TALK" en el presente documento, y su uso en el diagnóstico y tratamiento de pacientes con cáncer. Se basa, al menos en parte, en el descubrimiento de una TALK oncogénica novedosa, denominada "ALK<sup>ATl</sup>", que surgió como resultado del establecimiento de un sitio de iniciación transcripcional alternativo (ATI) *de novo* en el intrón 19 de ALK independiente de aberraciones genéticas en el locus de ALK. El transcrito de ALK<sup>ATl</sup> codifica tres proteínas con un peso molecular de 61,1, 60,8 y 58,7 kDa que consisten principalmente en el dominio de tirosina cinasa intracelular. Se encontró que ALK<sup>ATl</sup> estimula múltiples vías de señalización oncogénica, impulsa la proliferación celular independiente del factores de crecimiento *in vitro* y promueve la tumorigénesis *in vivo*. Se encontró que los inhibidores de ALK suprimen la actividad de cinasa de ALK<sup>ATl</sup>, lo que sugiere que los pacientes con tumores que expresan ALK<sup>ATl</sup> pueden beneficiarse de inhibidores de ALK. La expresión de ALK<sup>ATl</sup> se encontró en más del 10% (~11%) de los melanomas y esporádicamente en otros tipos de cáncer, pero no en tejidos normales. La detección de TALK, por ejemplo, ALK<sup>ATl</sup> en una célula de un sujeto puede usarse para diagnosticar un cáncer en el sujeto y puede usarse para determinar si el tratamiento del sujeto con un inhibidor de ALK sería más probable que confiriera un beneficio terapéutico. Es una ventaja de la divulgación que un sujeto que expresa TALK que parecería normal por pruebas genómicas pueda identificarse detectando transcripción de TALK, ARNm o proteína.

La presente invención proporciona una molécula de ácido nucleico aislado que comprende los exones 20-29 y una parte pero no la totalidad del intrón 19 de un gen de cinasa de linfoma anaplásico (ALK), donde la molécula de ácido nucleico no comprende los exones 1-19 del gen ALK, o bien individualmente o bien en cualquier combinación, en el

que la molécula de ácido nucleico es de aproximadamente 2300-2600 pb de longitud. En determinadas realizaciones, el gen ALK es un gen ALK humano. En una realización no limitante, la parte del intrón 19 del gen ALK es de aproximadamente 400 pb de longitud ubicada en el sentido de 5' del exón 20 del gen ALK. En determinadas realizaciones, la molécula de ácido nucleico es de aproximadamente 2300-2600 pb de longitud, como

5 aproximadamente de 2500 pb de longitud, por ejemplo, aproximadamente de 2513 pb de longitud. En una realización no limitante, la molécula de ácido nucleico es un transcrito de ARN mensajero (ARNm). En determinadas realizaciones, la molécula de ácido nucleico aislado comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma.

10 La presente invención también proporciona una molécula de ácido nucleico aislado que consiste esencialmente en los exones 20-29 y una parte, pero no todo, del intrón 19 de un gen de cinasa de linfoma anaplásico (ALK), en el que la molécula de ácido nucleico no comprende los exones 1-19 del gen ALK, o bien individualmente o bien en cualquier combinación, en el que la molécula de ácido nucleico es de aproximadamente 2300-2600 pb de longitud. En determinadas realizaciones, la molécula de ácido nucleico aislado consiste esencialmente en la secuencia de

15 nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma.

La presente invención proporciona además una molécula de ADN complementario (ADNc) aislado que comprende los exones 20-29 de un gen de cinasa de linfoma anaplásico (ALK), en el que la molécula de ADNc no comprende los

20 exones 1-19 del gen ALK, o bien individualmente o bien en cualquier combinación, en la que la molécula de ácido nucleico es de aproximadamente 2300-2600 pb de longitud. En determinadas realizaciones, el gen ALK es un gen ALK humano. En determinadas realizaciones, la molécula de ADNc aislado comprende ácidos nucleicos 405-2063 de la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma. En determinadas realizaciones, la molécula de ADNc aislado comprende ácidos

25 nucleicos 411-2063 de la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma. En determinadas realizaciones, la molécula de ADNc aislado comprende ácidos nucleicos 465-2063 de la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma.

30 Adicionalmente, la presente invención proporciona una molécula de ADN complementario (ADNc) aislado que consiste esencialmente en los exones 20-29 de un gen de cinasa de linfoma anaplásico (ALK), en el que la molécula de ADNc no comprende los exones 1-19 del gen ALK, o bien individualmente o bien en cualquier combinación, en el que la molécula de ácido nucleico es de aproximadamente 2300-2600 pb de longitud. También se proporciona una molécula de ADNc aislado que consiste esencialmente en los ácidos nucleicos 405-2063 de la secuencia de nucleótidos

35 expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma, en el que la molécula de ADNc no comprende los exones 1-19 del gen ALK, o bien individualmente o bien en cualquier combinación. La presente invención proporciona además una molécula de ADNc aislado que consiste esencialmente en los ácidos nucleicos 411-2063 de la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma, en el que la molécula de

40 ADNc no comprende los exones 1-19 del gen ALK, o bien individualmente o bien en cualquier combinación. La presente invención proporciona además una molécula de ADNc aislado que consiste esencialmente en los ácidos nucleicos 465-2063 de la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma, en la que la molécula de ADNc; 4. La presente invención proporciona además un polipéptido aislado que comprende los aminoácidos 21 a 552 de la secuencia de aminoácidos

45 expuesta en SEQ ID NO: 12 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma, en la que el polipéptido no comprende la secuencia de aminoácidos expuesta en SEQ ID NO; 4.

La invención también proporciona un anticuerpo que se une a cualquiera de los polipéptidos descritos anteriormente, en el que el anticuerpo no se une a un ALK de tipo silvestre. En determinadas realizaciones, el ALK de tipo silvestre es un ALK humano de tipo silvestre, por ejemplo, uno que comprende la secuencia de aminoácidos expuesta en SEQ

50 ID NO: 1.

La presente invención también proporciona un método para diagnosticar una célula como una célula cancerosa que comprende detectar la presencia de una isoforma truncada de cinasa de linfoma anaplásico (TALK) en la célula, donde

55 la presencia de la TALK indica que la célula es una célula cancerosa. En determinadas realizaciones, el método comprende determinar la presencia de un transcrito de ARNm de TALK detectable, una molécula de ADNc de TALK detectable correspondiente a la misma, un polipéptido de TALK detectable codificado mediante el mismo y/o un sitio de iniciación transcripcional alternativa detectable (ATI) en el intrón 19 de un gen ALK en la célula.

La presente invención proporciona un método para determinar si un efecto anticancerígeno es probable que se produzca en un cáncer por un inhibidor de ALK, que comprende determinar si una o más células del cáncer contiene una TALK detectable, en el que la presencia de una TALK detectable en la célula indica que un inhibidor de ALK tendría un efecto anticancerígeno sobre el cáncer. En determinadas realizaciones, el método comprende determinar la presencia de un transcrito de ARNm de TALK detectable, una molécula de ADNc de TALK detectable correspondiente a la misma, un polipéptido de TALK detectable codificado por el mismo y/o un sitio de ATI detectable

60 en el intrón 19 de un gen ALK en la célula.

65

La presente invención proporciona además un método *in vitro* para diagnosticar una célula en una muestra recogida de un sujeto como célula cancerosa que comprende detectar la presencia de una isoforma truncada de cinasa de linfoma anaplásico (TALK) en la célula, en el que la presencia de la TALK indica que el célula es una célula cancerosa, en el que detectar comprende determinar la presencia de la molécula de ácido nucleico según lo descrito anteriormente o la molécula de ADNc según lo descrito anteriormente y/o el polipéptido de TALK según lo descrito anteriormente. Se da a conocer además un método para tratar un sujeto que tiene un cáncer que comprende: (a) determinar si un sujeto es posible que obtenga beneficio terapéutico de un inhibidor de ALK, que comprende determinar si una o más células cancerosas del sujeto contienen una TALK detectable, en el que la presencia de una TALK detectable en la célula cancerosa del sujeto indica que es probable que el sujeto se beneficie de un inhibidor de ALK; y (b) tratar al sujeto que es posible que se beneficie de un inhibidor de ALK con una cantidad terapéutica de un inhibidor de ALK. En determinadas realizaciones, el método comprende tratar al sujeto con una terapia alternativa distinta de un inhibidor de ALK si no está presente TALK detectable en la célula cancerosa del sujeto. Dado a conocer en el presente documento, el método comprende además obtener una muestra de una o más células cancerosas del sujeto antes del tratamiento con un inhibidor de ALK. Dado a conocer en el presente documento, el sujeto que es posible que se beneficie de un inhibidor de ALK recibe un inhibidor de ALK y uno o más tratamientos de cáncer adicionales seleccionados del grupo que consiste en uno o más inhibidores de BRAF, uno o más inhibidores de MEK, uno o más inhibidores inmunológicos, uno o más inhibidores de CDK4, uno o más inhibidores de CDK6, uno o más agentes alquilantes, uno o más inhibidores de topoisomerasa, uno o más antimetabolitos, uno o más antimicrotúbulos, uno o más antibióticos citotóxicos, radioterapia, quimioterapia y combinaciones de los mismos. Dado a conocer en el presente documento, el inhibidor inmunológico se selecciona del grupo que consiste en un anticuerpo anti-PD-1, un anticuerpo anti-CTLA-4 y un anticuerpo anti-PDL1. En un ejemplo no limitativo, el sujeto que es posible que se beneficie de un inhibidor de ALK recibe un inhibidor de ALK, un anticuerpo anti-CTLA-4 y un anticuerpo anti-PD-1. En un ejemplo no limitativo, el sujeto recibe además radiación y quimioterapia. Dado a conocer en el presente documento, el método comprende determinar la presencia de un transcrito de ARNm de TALK detectable, una molécula de ADNc de TALK detectable, un polipéptido de TALK detectable codificado por el mismo y/o un sitio de ATI detectable en el intrón 19 de un gen ALK en la célula.

La presente divulgación también proporciona un kit para determinar si un efecto anticancerígeno es posible que se produzca en un cáncer por un inhibidor de ALK, que comprende un medio para determinar la presencia de una TALK detectable en una o más células del cáncer. En determinados ejemplos, el kit comprende medios para determinar el nivel de un transcrito de ARNm de TALK, una molécula de ADNc de TALK correspondiente a la misma y/o un polipéptido de TALK codificado por el mismo. En determinados ejemplos, el medio para determinar el nivel de un transcrito de ARNm de TALK se selecciona del grupo que consiste en hibridación de sonda, reacción en cadena de la polimerasa (PCR), transferencia de tipo Northern, secuenciación, microalineamiento y combinaciones de los mismos. En determinados ejemplos, el kit comprende una sonda de ácido nucleico que hibrida con el transcrito de ARNm de TALK para determinar el nivel del transcrito de ARNm de TALK. En determinados ejemplos, la hibridación de sonda es un ensayo nCounter de NanoString basado en sonda. En determinados ejemplos, el kit comprende un par de sondas codificadas por colores que hibrida con el transcrito de ARNm de TALK para determinar el nivel del transcrito de ARNm de TALK. En determinados ejemplos, la PCR se selecciona del grupo que consiste en reacción en cadena de la polimerasa de transcripción inversa (RT-PCR), PCR de transcriptasa inversa cuantitativa, PCR en tiempo real, PCR en tiempo real cuantitativa y combinaciones de las mismas. En determinados ejemplos, el kit comprende un par de cebadores de ácido nucleico que hibrida con el ARNm de TALK para determinar el nivel del transcrito de ARNm de TALK. En determinados ejemplos, el medio para determinar el nivel de una molécula de ADNc de TALK es PCR. Dado a conocer en el presente documento, el kit comprende un par de cebadores de ácido nucleico que hibridan con la molécula de ADNc de TALK para determinar el nivel de la molécula de ADNc de TALK. En determinados ejemplos, el medio para determinar el nivel de un polipéptido de TALK se selecciona del grupo que consiste en unión a anticuerpos, inmunohistoquímica, inmunotransferencia de tipo Western, un ensayo funcional, enzimoimmunoanálisis de adsorción (ELISA), radioinmunoensayos (RIA), inmunoensayos enzimáticos (EIA), espectrometría de masas, un sistema de análisis basado en gel 1-D o 2-D, inmunoprecipitación y combinaciones de los mismos. En determinados ejemplos, el kit comprende un anticuerpo o un fragmento del mismo que se une específicamente al polipéptido de TALK. En un ejemplo no limitativo, los ensayos funcionales son un ensayo de cinasa.

Dado a conocer en el presente documento, el kit comprende medios para determinar la presencia de un sitio de ATI en el intrón 19 de un gen ALK en una o más células del cáncer. En determinados ejemplos, los medios para determinar la presencia de un sitio de ATI en intrón 19 de un gen ALK se seleccionan del grupo que consiste en transferencia de tipo Northern, secuenciación de inmunoprecipitación de cromatina (ChIP), ChIP-qPCR, PCR de amplificación rápida de extremos de ADNc (RACE) y combinaciones de los mismos.

La presente divulgación proporciona un medio para determinar la presencia de una TALK detectable para su uso en un método de determinación de si un efecto anticancerígeno es probable que se produzca en un cáncer por un inhibidor de ALK, el método caracterizado por determinar si una o más células del cáncer contienen una TALK detectable, donde la presencia de una TALK detectable en la célula indica que un inhibidor de ALK tendría un efecto anticancerígeno sobre el cáncer. En determinados ejemplos, el medio es para determinar el nivel de un transcrito de ARNm de TALK en una o más células del cáncer. En determinados ejemplos, los medios se seleccionan del grupo que consiste en hibridación de sonda, reacción en cadena de la polimerasa (PCR), transferencia de tipo Northern,

secuenciación, microalineamiento y combinaciones de las mismas. En un ejemplo no limitativo, la hibridación de sonda es un ensayo NCounter de NanoString basado en sonda. En determinados ejemplos, la PCR se selecciona del grupo que consiste en reacción en cadena de la polimerasa de transcripción inversa (RT-PCR), PCR de transcriptasa inversa cuantitativa, PCR en tiempo real, PCR en tiempo real cuantitativa y combinaciones de las mismas. En determinados ejemplos, los medios son para determinar el nivel de una molécula de ADNc de TALK en una o más células del cáncer. En un ejemplo no limitativo, el medio es PCR. En determinados ejemplos, los medios son para determinar el nivel de un polipéptido de TALK en una o más células del cáncer. En determinados ejemplos, los medios se seleccionan del grupo que consiste en unión de anticuerpos, inmunohistoquímica, inmunotransferencia de tipo Western, un ensayo funcional, enzimoimmunoanálisis de adsorción (ELISA), radioinmunoensayos (RIA), inmunoensayos enzimáticos (EIA), espectrometría de masas, un sistema de análisis basado en gel 1-D o 2-D, inmunoprecipitación y combinaciones de los mismos. En un ejemplo no limitativo, los ensayos funcionales son un ensayo de cinasa. En determinados ejemplos, los medios son para determinar la presencia de un sitio de ATI en intrón 19 de un gen ALK en una o más células del cáncer. En determinados ejemplos, los medios se seleccionan del grupo que consiste en transferencia de tipo Northern, secuenciación de inmunoprecipitación de cromatina (ChIP), ChIP-qPCR, PCR de amplificación rápida de extremos de ADNc (RACE) y combinaciones de los mismos.

La presente divulgación proporciona además un inhibidor de ALK para su uso en el tratamiento de un sujeto que tiene cáncer, en el que el inhibidor de ALK es más probable que produzca un efecto anticancerígeno si una o más células cancerosas del sujeto contienen una TALK detectable.

En determinadas realizaciones, el cáncer se selecciona del grupo que consiste en melanoma, carcinoma de tiroides, adenocarcinoma de pulmón, carcinoma de células escamosas pulmonares, carcinoma de células claras renales y cáncer de mama. En una realización no limitante, el cáncer es melanoma. En una realización no limitante, el cáncer es carcinoma de tiroides anaplásico.

En determinados ejemplos, el inhibidor de ALK se selecciona del grupo que consiste en crizotinib, ceritinib, NVP-TAE684, alectinib, AP26113, ASP-3026, CEP-37440, NMS-E628, PF-06463922, TSR-011, RXDX-101 y X-396. En un ejemplo no limitativo, el inhibidor de ALK es crizotinib.

En determinados ejemplos, el transcrito de ARNm de TALK es la molécula de ácido nucleico descrita anteriormente. En determinados ejemplos, la molécula de ADNc de TALK es la molécula de ADNc descrita anteriormente. En determinados ejemplos, el polipéptido de TALK es el polipéptido descrito anteriormente.

#### 4. Breve descripción de las figuras

FIGURAS 1A-1H. La iniciación transcripcional alternativa (ATI) da como resultado un transcrito de ALK novedoso. 1A, Distribución de lecturas de secuenciación de ARN de transcritos de variantes de ALK: lecturas de secuenciación de ARN de  $ALK^{ATI}$  alineadas tanto con el intrón 19 como con los exones 20-29 de ALK; lecturas de secuenciación de ARN de  $ALK^{WT}$  de longitud completa alineadas con todos los exones de ALK, pero no con los intrones; lecturas de secuenciación de ARN de ALK translocada alineadas solo con los exones 20-29 de ALK. 1B, Mapeo de los sitios de ATI de  $ALK^{ATI}$  de RACE en 5' y secuenciación de última generación. Más del 95% de los transcritos de  $ALK^{ATI}$  comienzan dentro de una región de 25 pares de base en intrón 19 de ALK (hg19 ch2:29,446,768-29,446,744; área sombreada azul). 1C, perfil de secuenciación de ChIP de H3K4me3 en el sitio de ATI de un tumor que expresa  $ALK^{ATI}$  (MM-15), una línea celular de melanoma (SKMEL-524) sin expresión de  $ALK^{ATI}$ , y una línea celular de cáncer de pulmón (H3122) con una fusión de EML4-ALK. La forma del pico muestra un estrechamiento lento hacia el cuerpo genético característico de los promotores activos. 1D, validación de ChIP-qPCR de la unión de H3K4me3 en el sitio de ATI. Media  $\pm$  SEM, n=3. 1E, ChIP-qPCR de unión de ARN polimerasa II en el sitio de ATI. Media  $\pm$  SEM, n=3. 1F, Análisis de perfil de ARNm cuantitativo de diferentes variantes de ALK usando nCounter de Nanostring. Dos líneas celulares de neuroblastoma que expresan  $ALK^{WT}$  (SK-N-BE2, SK-N-DZ), dos líneas celulares de cáncer de pulmón (H3122, H2228) con translocaciones de EML4-ALK, nueve tumores que expresan  $ALK^{ATI}$ : 8 melanomas (MM) y 1 carcinoma de tiroides anaplásico (ATC). 1G, frecuencias de SNV similares en los datos de secuenciación de ADN, secuenciación de ARN y secuenciación de ChIP (H3K4me3) indican que  $ALK^{ATI}$  se expresa de manera bi-alélica y que ambos alelos de ALK están decorados con H3K4me3. 1H, Tabla que muestra el conjunto de sondas usado para el ensayo nCounter de Nanostring para cuantificar los niveles de ARNm de  $ALK^{ATI}$ ,  $ALK^{WT}$  y ALK translocada.

FIGURAS 2A-2F. El transcrito de  $ALK^{ATI}$  codifica tres proteínas de ALK acortadas que contienen principalmente el dominio de cinasa ALK. 2A, Ilustración de isoformas de proteína ALK. ALK, de tipo silvestre ( $ALK^{WT}$ ) consiste en una posición 1-18 de péptido de señalización de aminoácido (aa), un segmento extracelular con dos dominios de MAM (meprina, proteína A-5 y proteína-tirosina fosfatasa mu receptora) y una región rica en glicina (Gly), un segmento transmembrana, un segmento de yuxtamembrana y un segmento intracelular con el dominio de tirosina cinasa (cinasa, aa 1116-1392). La activación de las mutaciones de ALK ocurren normalmente en el dominio de cinasa, como el  $ALK^{F1174L}$ . La proteína de fusión EML4-ALK y  $ALK^{ATI}$  contienen todo el dominio de cinasa y partes del segmento de yuxtamembrana. La traducción de  $ALK^{ATI}$  se inicia en uno de los tres codones de inicio en marco (ATG) como se indica. 2B, Inmunotransferencias de ALK totales (t-) y fosforilados (p-) en dos líneas celulares de neuroblastoma que expresan  $ALK^{WT}$  (SK-N-DZ, SK-N-BE2), dos líneas celulares de cáncer de pulmón (H3122, H2228) que expresan diferentes variantes de fusión EML4-ALK, tres muestras de tumor que expresan  $ALK^{ATI}$  (MM-15, MM-74, ATC- 28) y

- una línea celular de melanoma de control negativo (SKMEL-28). 2C, Inmunotransferencias de células 293T transducidas con  $ALK^{AT7}$ , en las que los tres codones de inicio previstos se mutaron de ATG a AAG, individualmente o en combinación como se indica. 2D, Co-inmunoprecipitación (IP) e inmunotransferencias (IB) de los epítomos indicados en células 293T con expresión exógena de  $ALK^{AT7}$  etiquetados con V5 ( $V5-ALK^{AT7}$ ) o  $ALK^{AT7}$  etiquetado HA ( $HA-ALK^{AT7}$ ), o ambos, demostrando que las proteínas  $ALK^{AT7}$  interactúan entre sí. 2E, Inmunofluorescencia de ALK en células NIH-3T3 que expresan las isoformas de ALK indicadas. Barra de escala, 25  $\mu\text{m}$ . 2F, Tinción de HE (hematoxilina-eosina) e inmunohistoquímica de ALK en muestras de tumor humano que expresan  $ALK^{AT7}$ . Barra de escala, 50  $\mu\text{m}$ .
- 5
- 10 FIGURAS 3A-3C.  $ALK^{AT7}$  promueve la proliferación *in vitro* independiente del factor de crecimiento y la tumorigénesis *in vivo*. 3A, Curvas de crecimiento de células Ba/F3 que expresan de manera estable las isoformas de ALK indicadas en ausencia de interleucina 3 (IL-3). Media  $\pm$  SEM, n=4. 3B, Inmunotransferencias para determinar los niveles de ALK en células Ba/F3. Las células se transformaron previamente por la expresión de diferentes isoformas de ALK y se seleccionaron para el crecimiento en ausencia de IL-3. Todas las variantes de ALK se fosforilaron cuando se expresaron a niveles requeridos para el crecimiento independiente de IL-3. 3C, Curvas de crecimiento de injertos tumorales de células NIH-3T3 que expresan de manera estable las isoformas de ALK indicadas.  $ALK^{F1174L}$ , EML4-ALK, y sobreexpresión y amplificación de  $ALK^{WT}$  son iniciadores oncogénicos bien establecidos en varios tumores. Media  $\pm$  SEM, n=8.
- 15
- 20 FIGURAS 4A-4H. La expresión de  $ALK^{AT7}$  confiere sensibilidad a inhibidores de ALK *in vitro* e *in vivo*. 4A, Curvas de dosis-respuesta para el crizotinib inhibidor de ALK en células Ba/F3 que expresan las isoformas de ALK indicadas en presencia o ausencia de IL-3. Media  $\pm$  SEM, n=3. 4B, Inmunotransferencias representativas de células Ba/F3 que expresan  $ALK^{AT7}$  y tratadas con concentraciones crecientes de crizotinib durante 2 horas. 4C, Volumen tumoral normalizado a lo largo del tiempo en ratones con SCID implantados con células NIH-3T3 que expresan las isoformas de ALK indicadas y tratados con o bien vehículo o bien con crizotinib (100 mg/kg/día). Media  $\pm$  SEM, n=8. 4D, Tinción de Hematoxilina-eosina (HE) e inmunohistoquímica (IHC) de tumores que expresan  $ALK^{AT7}$  explantados 48 horas después del primer tratamiento con crizotinib. Barra de escala, 50  $\mu\text{m}$ . 4E, Bioluminiscencia normalizada de tumores injertados NIH-3T3 marcados con luciferasa que expresan  $ALK^{AT7}$  a lo largo del tiempo en ratones con SCID tratados con vehículo o crizotinib (100 mg/kg/día). Media  $\pm$  SEM, n=4. 4F, Tinción HE y ALK-IHC (inserto) de la metástasis de melanoma del paciente 1 (MM-382). 4G, Análisis de perfil de ARNm cuantitativo de  $ALK^{AT7}$  usando nCounter de Nanostring. Controles que expresan  $ALK^{WT}$ , EML4-ALK o  $ALK^{AT7}$  en comparación con la metástasis de melanoma del paciente MM-382. Barra de escala, 50  $\mu\text{m}$ . 4H, Imágenes de tomografía computarizada (TC) de una metástasis de melanoma subcutánea representativa en la cadera izquierda del paciente 1 (MM-382) antes y después del tratamiento con crizotinib, respectivamente.
- 25
- 30
- 35 FIGURAS 5A-5C. Comparación de los perfiles de secuenciación de ARN de varias transcripciones de ALK. Los datos de secuenciación de ARN se muestran en la herramienta Integrative Genomics Viewer (IGV). Las barras/flechas grises indican las lecturas de secuenciación. Las líneas azules conectan lecturas de secuenciación que están alineadas sobre el sitio de corte y empalme de unión de exones. 5A, el transcrito de  $ALK^{AT7}$  muestra la expresión de exones 20-29 de ALK y de aproximadamente 400 pb en intrón 19 (área sombreada azul). No se observa ninguna expresión de exón 1-19 o áreas intrónicas, excepto en intrón 19. La vista detallada ilustra que las lecturas de secuenciación se alinean de manera continua entre el exón 20 y el intrón 19, lo que indica una transcripción ininterrumpida. El UTR en 5' de  $ALK^{AT7}$  (intrón 19) y el exón 20-29 se expresan a niveles comparables. 5B, el transcrito de  $ALK^{WT}$  de longitud completa muestra la expresión de todos los exones de ALK y solo muy poca expresión de los intrones. La vista detallada muestra que las lecturas de secuenciación se alinean considerablemente con los exones, pero no con la región de intrón 19, que está presente en  $ALK^{AT7}$  (área sombreada azul). 5C, el transcrito de fusión de ALK de un cáncer de pulmón de células no pequeñas con una translocación EML4-ALK muestra expresión de exones 20-29, y poca expresión de exones 1-19 y todos los intrones de ALK. La vista detallada ilustra que la transcripción comienza principalmente en el exón 20 debido al sitio de corte y empalme conservado. Pocas lecturas están alineadas con la región de intrón 19 (área sombreada azul). Las lecturas marcadas en verde resaltan pares de lectura quimérica que indican la translocación EML4-ALK.
- 40
- 45
- 50
- 55 FIGURAS 6A-6G. Identificación del transcrito de  $ALK^{AT7}$  novedoso. 6A, Vista de IGV de los fragmentos de ADNc de RACE en 5' obtenidos por secuenciación masivamente paralela. La gran mayoría de las lecturas de secuenciación (flechas grises) comienzan dentro del sitio de ATI principal de 25 pares de bases (hg19 chr2:29,446,744-768). 6B, Porcentaje de lecturas que comienzan en el sitio de ATI en muestras de tumor que expresa  $ALK^{AT7}$ . 6C, Secuenciación de Sanger de los fragmentos de ADNc de RACE en 5' clonados confirman la transcripción continua comenzando en el intrón 19 de ALK y extendiéndose a los exones 20-21. 6D, el transcrito de  $ALK^{AT7}$  consiste en aproximadamente 400 pb en el sentido de 5' del exón 20 y de los exones 20-29 de ALK. El sitio de iniciación transcripcional se definió como el primer par de bases en el que se iniciaron más del 5% de los transcritos (chr2:29,446,766). Otros sitios principales de iniciación transcripcional están marcados en rojo, los UTR en 5' y 3' en azul oscuro, la secuencia de codificación de ADN (CDS) en negro y encuadrada, y la primera y última base de cada exón en azul claro. La traducción se inicia a partir de 3 codones de inicio (ATG; en negrita y subrayados): 1<sup>er</sup> ATG: hg19 chr2:29,446,360-2, 2<sup>o</sup> ATG (+7-9), and 3<sup>er</sup> ATG (+61-3). 6E, La secuencia de aminoácidos de  $ALK^{AT7}$ . La traducción se inicia en uno de los 3 codones de inicio. Las 3 metioninas correspondientes (en negrita y subrayadas) dan como resultado 3 proteínas diferentes, 61,08
- 60
- 65

kDa (552 aminoácidos), 60,82 kDa (550 aminoácidos) y 58,71 kDa (532 aminoácidos). El dominio de cinasa se resalta en letras rojas. La lisina en el dominio de unión de ATP en *ALK<sup>WT</sup>* está marcada en negrita y subrayada, y se mutó a metionina (p.K1150M) en el *ALK<sup>AT1</sup>* sin actividad de cinasa. 6F, Transferencia de tipo Northern de *ALK<sup>WT</sup>* de longitud completa (líneas celulares de neuroblastoma), EML4-ALK (líneas celulares de cáncer de pulmón; variante 1 y 3), *ALK<sup>AT1</sup>* de melanomas humanos y controles negativos (líneas celulares de melanoma). A excepción de los controles negativos, cada carril muestra dos bandas; la banda V2 inferior coincide con el transcrito de ALK canónico más corto (RefSeq) que termina en ~chr2:29,415,640; la banda V1 superior corresponde a un transcrito con un 1,8 kb más largo que UTR en 3' que termina en el ~chr2:29,413,840. Dos tumores que expresan *ALK<sup>AT1</sup>*, MM-284 y MM-74, muestran solo señales débiles porque se dispuso de menos de 1 µg de ARN; para todas las demás muestras se usaron 5-10 µg de ARN. 6G, Gráfico de tipo Sashimi de secuenciación de ARN que ilustra los transcritos de ALK V2 más cortos y V1 más largas por la brusca caída de lecturas de secuenciación en el UTR en 3' en chr2:29,415,640 para V2 y en chr2:29,413,840 para V1.

FIGURAS 7A-7D. ARN polimerasa II y H3K4me3 se enriquecen en el sitio de ATI de muestras de tumor que expresa *ALK<sup>AT1</sup>*. 7A, Secuencias de cebador usadas para ChIP-qPCR. 7B, Esquemas de los sitios de unión de cebador de ChIP-qPCR en el locus de *ALK<sup>AT1</sup>*. 7C y 7D, ChIP-qPCR de (C) H3K4me3 y (D) ARN polimerasa II en el sitio de ATI que demuestra el enriquecimiento de ambas marcas en las muestras de tumor humano que expresan *ALK<sup>AT1</sup>*, pero no en los controles negativos, incluyendo una línea celular de cáncer de pulmón con translocación EML4-ALK (H3122), y una línea celular de melanoma (SKMEL-524). Se muestran los resultados del par de cebador 1 (P1) en las figuras 1D-1E. Media ± SEM, n=3.

FIGURAS 8A-8E. Tumores que expresan *ALK<sup>AT1</sup>* en el conjunto de datos de TCGA. 8A, La frecuencia de tumores que expresan *ALK<sup>AT1</sup>* en más de 5.000 muestras tumorales de 15 tipos diferentes de cáncer en el conjunto de datos de secuenciación de ARN de TCGA. 8B-8E, Vistas de IGV del locus de ALK de los tipos representativos de tumor que expresa *ALK<sup>AT1</sup>* en el conjunto de datos de TCGA incluyendo B, melanoma; C, adenocarcinoma de pulmón; D, carcinoma invasivo de mama; E, carcinoma de células renales claras. La ausencia de pares de lectura de secuenciación quimérica indica que no hay translocaciones en estos casos.

Figuras 9A-9D *ALK<sup>AT1</sup>* se transcribe desde un locus de ALK genómicamente intacto. 9A, FISH en interfase con sondas de flanqueamiento de ALK demuestra señales verdes y naranjas yuxtapuestas que indican que no hay reordenación de ALK en MM-15. Barra de escala, 10µm. 9B, FISH en interfase no muestra reordenación de ALK, pero 3 señales de fusión verde/naranja en la mayoría de los núcleos celulares que indican una trisomía 2 en MM-74. Barra de escala, 10µm. 9C, el panel superior muestra el perfil de alineamiento CGH de genoma completo de MM-15 con numerosas ganancias y pérdidas cromosómicas en todo el genoma, que son características del melanoma metastásico. Los cromosomas están alineados a lo largo del eje X. La línea azul ilustra el número de copia relativo (razón de log<sub>2</sub>) y las barras azules resaltan las ganancias y pérdidas del número de copia. El panel central ilustra el número de copia relativo (línea azul) a través del cromosoma 2. Distal al locus de ALK, se indica una pérdida en el brazo p del cromosoma 2. El panel inferior ilustra el número de copia relativo a través del gen ALK. Los cuadrados rojos y verdes representan la razón de log<sub>2</sub> de sondas individuales aCGH (verde: razón de log<sub>2</sub> positiva; rojo: razón de log<sub>2</sub> negativa). No se encuentran interrupciones ni ganancias o pérdidas selectivas en el locus de ALK. 9D, El perfil de alineamiento CGH de genoma completo de MM-74 muestra numerosas ganancias y pérdidas cromosómicas a través de todo el genoma en el panel superior. El panel central muestra una ganancia de número de copia relativa de todo el cromosoma 2, que está en línea con la trisomía del cromosoma 2 según lo indicado por FISH. El panel inferior también muestra la trisomía del cromosoma 2 y no muestra ganancias ni pérdidas focales en el locus de ALK.

FIGURAS 10A-10D. La secuenciación dirigida y la secuenciación del genoma completo no revela aberraciones genómicas recurrentes en el locus de ALK. 10A, Datos de la secuenciación ultra profunda de todo el locus de ALK se muestran en IGV. La región genómica alrededor del intrón 19 revela varias variaciones de nucleótidos únicos (SNV). Sin embargo, la gran mayoría de los SNV en el locus de ALK también se encuentran en la población general, ya que se detectan en el grupo de ADN normal, que se usó como el control (panel inferior normal común). Numerosos SNV también se documentan en la base de datos The Single Nucleotide Polymorphism (dbSNP - [www.ncbi.nlm.nih.gov/SNP/](http://www.ncbi.nlm.nih.gov/SNP/)). No se encontraron aberraciones genómicas en el sitio de iniciación transcripcional de *ALK<sup>AT1</sup>*. 10B y 10C, Gráficos Circos de los datos de secuenciación del genoma completo de (B) MM-15 y (C) ATC-28 que ilustran numerosas aberraciones estructurales y SNV. 10D, polimorfismos de nucleótido único y pequeñas inserciones o deleciones en el locus de ALK detectadas por secuenciación ultra profunda.

FIGURAS 11A-11H. Contexto local de cromatina en el sitio de iniciación de transcripción alternativa (ATI). 11A, Vista del navegador genómico (Genome Browser) de UCSC en el sitio de ATI. La pista RepeatMasker muestra elementos transponibles en la región de ATI, incluyendo una repetición de terminal largo (LTR) en el intrón 19 (LTR16B2) y un elemento de largo interespaciado (LINE) en el intrón 18. Las pistas de ENCODE revelan una agrupación de hipersensibilidad de ADNasa I y enriquecimiento de H3K4me1, pero no enriquecimiento de H3K27ac. 11B, El estado de metilación del locus de ALK se evaluó mediante la captura personalizada de toda la huella de ALK, seguida por el tratamiento de bisulfito y la secuenciación de última generación. Los resultados de la secuenciación por bisulfito de H3122 y MM-15 se muestran en el modo CG-bisulfito de IGV. El color rojo indica "C" (citosina) correspondiente a citosina metilada, que se conserva durante la reacción de bisulfito. El color azul indica "T" (timina) correspondiente a

5 citosina no metilada, que se convierte en uracilo en la reacción de bisulfito, y posteriormente se amplifica a timina durante la PCR. 11C, Nivel de metilación en CpG en muestras de tumor que expresa  $ALK^{AT1}$  (MM-15 y ATC-28) y células de control que no expresan  $ALK^{AT1}$  (H3122, una línea celular de cáncer de pulmón con expresión de EML4-ALK y SKMEL-28, una línea celular de melanoma sin expresión de  $ALK^{AT1}$ ) en los loci genómicos indicados. Las líneas negras marcan el LTR16B2 (LTR), las líneas rojas marcan la LINE. 11D, Comparación del estado de metilación de CpG adyacentes al sitio de ATI en muestras de tumor que expresa  $ALK^{AT1}$  (MM-15 y ATC-28) y células control que no expresan  $ALK^{AT1}$  (H3122 y SKMEL-28). Las regiones que flanquean LTR16B2 tienen niveles de metilación de CpG significativamente más bajos en muestras que expresan  $ALK^{AT1}$  que los controles; los puntos rojos indican una diferencia estadísticamente significativa ( $p < 0,05$ ) entre muestras que expresan  $ALK^{AT1}$  y muestras que no expresan ALKATI. Los puntos negros indican diferencias estadísticamente no significativas. 11E, Perfil de secuenciación de ChIP de H3K27ac en el locus  $ALK^{AT1}$ . Los 17 perfiles azules se recuperaron de ENCODE, los 5 perfiles rojos son datos originales del laboratorio de los inventores. Solo las tres muestras de melanoma (MM-15, SKMEL-28, SKMEL-524; parte inferior), pero no las 19 líneas celulares de no-melanoma, muestran enriquecimiento de H3K27ac en el sitio de ATI. 11F, Validación de ChIP-qPCR para el enriquecimiento de H3K27ac en el sitio de ATI en 6 líneas celulares de melanoma. Media  $\pm$  SEM,  $n=3$ . 11G, Ensayo de indicador de luciferasa de LTR16B2 en líneas celulares de melanoma (rojo) y líneas celulares de cáncer de pulmón que expresan EML4-ALK (verde). A diferencia de las líneas celulares de cáncer de pulmón sin actividad de luciferasa, las líneas celulares de melanoma muestran actividad de luciferasa de baja a moderada. 11H, Análisis de motivo de factor de transcripción de la región cis-reguladora proximal (hg19 chr2: 29,445,000 a 29,447,100).

20 FIGURAS 12A-12I.  $ALK^{AT1}$  es activo *in vitro*, muestra localización nuclear y citoplasmática por inmunohistoquímica e induce tumorigénesis. 12A, Ensayo de cinasa *in vitro*. Las variantes de ALK indicadas se expresaron de manera estable en células NIH-3T3, se inmunoprecipitaron y se ensayaron para actividad de tirosina cinasa. Después de la reacción enzimática, se usó el material inmunoprecipitado para inmunotransferencias para evaluar la cantidad de proteína ALK usada en el ensayo de cinasa. Media  $\pm$  SEM,  $n=4$ . 12B, Tumor humano (MM-15) que expresa  $ALK^{AT1}$  muestra la localización citoplasmática y nuclear de ALK por inmunohistoquímica. Tumor melanocítico que expresa una translocación de TPM3-ALK mostró la localización citoplasmática de la proteína de fusión ALK sin ninguna tinción nuclear por inmunohistoquímica. Fibroblastos, células epiteliales y linfocitos reactivos sirven como controles negativos internos. Barras de escala, 100  $\mu$ m. 12C, Análisis de citometría de flujo para proteína fluorescente verde (GFP) coexpresada con las isoformas de ALK indicadas. Se cultivaron células en medio suplementado IL-3 hasta el día 0 (curva azul) y se evaluó el número de positivas para GFP. 14 días después de la retirada de IL-3, se evaluó nuevamente el número de células que expresan ALK positivas para GFP (curva roja). 12D, Inmunotransferencias de injertos de tumor explantado NIH-3T3 que expresan diversas isoformas de ALK.  $ALK^{AT1}$  se expresó a niveles de proteína similares a los de dos muestras de tumor humano que expresan  $ALK^{AT1}$ . 12E, Curvas de crecimiento de injertos de tumor de células melan-a que expresan de manera estable las isoformas de ALK indicadas. Media  $\pm$  SEM,  $n=8$ . 12F, Inmunotransferencias de injertos de tumor melan-a explantados que expresan diversas isoformas de ALK en comparación con muestras de tumores humanos que expresan  $ALK^{AT1}$ . 12G, Análisis de citometría de flujo de la señal de GFP en células NIH-3T3 que expresan de manera estable niveles bajos (bajo  $ALK^{AT1}$ ) o altos (alto  $ALK^{AT1}$ ) de  $ALK^{AT1}$  antes de injertar en ratones con SCID. 12H, Inmunotransferencia de t-ALK en células con alto  $ALK^{AT1}$  y bajo  $ALK^{AT1}$ , confirmando la expresión diferencial de ATI. 12I Curvas de crecimiento de injertos de tumor de células con bajo  $ALK^{AT1}$  y alto  $ALK^{AT1}$ .

45 FIGURAS 13A-13E. Inhibición de ALK dependiente de concentración en células Ba/F3 que expresan  $ALK^{AT1}$ ,  $ALK^{WT}$ ,  $ALK^{F1174L}$  y EML4-ALK. 13A y 13B, Ensayo de viabilidad celular de células Ba/F3, o bien en presencia o bien en ausencia de IL-3 (1 ng/ml), que expresan las isoformas de ALK indicadas y tratadas con las dosis indicadas de inhibidores de ALK (A) ceritinib y (B) TAE-684. La viabilidad celular se midió después de 72 horas de tratamiento con fármaco. Media  $\pm$  SEM,  $n=4$ . 13C-13E, Inmunotransferencias representativas de células Ba/F3 que expresan de manera estable (C)  $ALK^{WT}$ , (D)  $ALK^{F1174L}$  o (E) EML4-ALK y tratadas con concentraciones crecientes de crizotinib durante 2 horas. Se muestran inmunotransferencias para  $ALK^{AT1}$  en la figura 4B.

50 FIGURAS 14A-14F. La expresión de  $ALK^{WT}$ ,  $ALK^{F1174L}$  y EML4-ALK confiere sensibilidad al inhibidor de ALK, crizotinib, *in vivo*. 14A-14C, Bioluminiscencia de tumores injertados NIH-3T3 marcados con luciferasa que expresan (A)  $ALK^{WT}$ , (B)  $ALK^{F1174L}$  o (C) EML4-ALK a lo largo del tiempo en ratones con SCID tratados con o bien vehículo o bien crizotinib (100 mg/kg/día). Los datos para tumores que expresan  $ALK^{AT1}$  se muestran en la figura 4E. 14D-14F, Tinción de hematoxilina-eosina (HE) e inmunohistoquímica de tumores explantados que expresan (D)  $ALK^{WT}$ , (E)  $ALK^{F1174L}$  o (F) EML4-ALK 48 horas después del primer tratamiento con crizotinib. Los datos de tumores que expresan  $ALK^{AT1}$  se muestran en la figura 4D. Barra de escala, 50  $\mu$ m.

### 60 Descripción detallada de la invención

Para mayor claridad y no a modo de limitación, la descripción detallada de la invención se divide en las siguientes subsecciones:

#### 65 (i) Definiciones

(ii) TALK;

(iii) Detección de TALK;

5 (iv) Métodos de diagnóstico;

(v) Métodos de tratamiento; y

(vi) Kits.

10

### 5. 1 Definiciones

15 Tal como se utiliza en el presente documento, el término “alrededor de” o “aproximadamente” significa dentro de un intervalo de error aceptable para el valor particular como se determina por un experto en la técnica, que dependerá en parte de cómo se mide o determina el valor, es decir, las limitaciones del sistema de medición. Por ejemplo, “alrededor de” puede significar dentro de 3 o más de 3 desviaciones estándar, según la práctica de la técnica. Alternativamente, “alrededor de” puede significar un intervalo de hasta el 20%, preferiblemente hasta el 10%, más preferiblemente hasta el 5%, y más preferiblemente aún hasta el 1% de un valor dado. Alternativamente, particularmente con respecto a sistemas o procesos biológicos, el término puede significar dentro de un orden de magnitud, preferiblemente dentro de cinco veces, y más preferiblemente dentro de dos veces, de un valor.

20 Como se usa en el presente documento, el término “anticuerpo” significa no solo moléculas de anticuerpo intactas, sino también fragmentos de moléculas de anticuerpo que conservan la capacidad de unión de inmunógenos. Estos fragmentos también son bien conocidos en la técnica y se emplean regularmente tanto *in vitro* como *in vivo*. Por consiguiente, como se utiliza en el presente documento, el término “anticuerpo” significa no solo moléculas de inmunoglobulina intactas, sino también los fragmentos activos bien conocidos F(ab')<sub>2</sub> y Fab. Los fragmentos F(ab')<sub>2</sub> y Fab que carecen del fragmento Fc de anticuerpo intacto, se limpian más rápidamente de la circulación, y pueden tener menos unión tisular no específica de un anticuerpo intacto (Wahl *et al.*, J. Nucl. Med. 24:316-325 (1983)). Los anticuerpos de la presente divulgación comprenden anticuerpos nativos completos, anticuerpos biespecíficos; anticuerpos quiméricos; fragmentos de región V de cadena sencilla Fab, Fab'(scFv), polipéptidos de fusión, anticuerpos humanizados derivados de un anticuerpo no humano y anticuerpos no convencionales.

25 Como se usa en el presente documento, el término “vector” se refiere a cualquier elemento genético, como un plásmido, bacteriófago, transposón, cósmido, cromosoma, virus, virión, etc., que es capaz de replicarse cuando se asocia con los elementos de control adecuados y que puede transferir secuencias genéticas a células. Por tanto, el término incluye los vehículos de clonación y expresión, así como vectores virales y vectores de plásmido.

30 Como se usa en el presente documento, el término “vector de expresión” se refiere a una secuencia de ácido nucleico recombinante, por ejemplo, una molécula de ADN recombinante, que contiene una secuencia de codificación deseada y secuencias de ácido nucleico apropiadas necesarias para la expresión de la secuencia de codificación ligada operativamente en un organismo huésped particular. Secuencias de ácido nucleico necesarias para su expresión en procariontes normalmente incluyen un promotor, un operador (opcional) y un sitio de unión de ribosoma, a menudo junto con otras secuencias. Se sabe que las células eucarióticas utilizan promotores, potenciadores y señales de terminación y poliadenilación.

35 Como se usa en el presente documento, el término “tratar” o “tratamiento” se refiere a la intervención clínica en un intento de alterar el curso de la enfermedad del individuo o célula a tratar, y puede realizarse o bien para profilaxis o bien durante el curso de la patología clínica. Los efectos terapéuticos del tratamiento incluyen, sin limitación, prevenir la aparición o recurrencia de la enfermedad, alivio de síntomas, disminución de las consecuencias patológicas directas o indirectas de la enfermedad, prevenir metástasis, disminuir la tasa de progresión de la enfermedad, mejora o atenuación del estado de la enfermedad, y remisión o mejor pronóstico. Al prevenir la progresión de una enfermedad o trastorno, un tratamiento puede prevenir el deterioro debido a un trastorno en un sujeto afectado o diagnosticado o un sujeto sospechoso de tener el trastorno, pero también un tratamiento puede prevenir la aparición del trastorno o un síntoma del trastorno en un sujeto en riesgo para el trastorno o sospechoso de tener el trastorno.

40 Como se usa en el presente documento, el término “cantidad efectiva” (o “cantidad terapéuticamente efectiva”) es una cantidad suficiente para afectar un resultado clínico beneficioso o deseado durante el tratamiento. Puede administrarse una cantidad efectiva a un sujeto en una o más dosis. En términos de tratamiento, una cantidad efectiva es una cantidad suficiente para paliar, mejorar, estabilizar, revertir o ralentizar la progresión de la enfermedad (por ejemplo, un cáncer), o reducir de otra manera las consecuencias patológicas de la enfermedad (por ejemplo, un cáncer). La cantidad efectiva se determina generalmente por el médico caso por caso y está dentro de la habilidad de un experto en la técnica. Varios factores se tienen en cuenta normalmente al determinar una dosificación adecuada para lograr una cantidad efectiva. Estos factores incluyen la edad, el sexo y el peso del sujeto, la afección que se está tratando, la gravedad de la afección y la forma y concentración efectiva de las células inmunosensibles administradas.

45 Como se usa en el presente documento, el término “sujeto” se refiere a cualquier animal (por ejemplo, un mamífero),

incluyendo, pero no limitado a, humanos, primates no humanos, roedores y similares (por ejemplo, que va a ser el receptor de un tratamiento particular, o de quien se cosechan células).

5 Como se usa en el presente documento, el término “un efecto anticancerígeno” significa uno o más de una reducción de masa de células cancerosas agregadas, una reducción de la tasa de crecimiento de células cancerosas, una reducción de la proliferación de células cancerosas, una reducción de la masa tumoral, una reducción del volumen tumoral, una reducción de la proliferación de células tumorales, una reducción en la tasa de crecimiento tumoral, una reducción en la metástasis tumoral y/o un aumento en la proporción de células cancerosas senescentes.

## 10 5.2 TALK

15 El gen de cinasa de linfoma anaplásico humano (ALK) está ubicado en 2p23, contiene 29 exones y codifica una tirosina cinasa de la superfamilia de insulina clásica de 220 kDa, de 1620 aminoácidos. La proteína ALK humana madura experimenta una glicosilación ligada a N post-traslacional y consiste en un dominio de unión a ligandos extracelular, un dominio transmembrana y un único dominio de tirosina cinasa intracelular.

El término “TALK” puede usarse para referirse a las isoformas truncadas de ALK a las que se hace referencia en el presente documento y al ARNm y la correspondiente molécula de ADNc que codifica las mismas.

20 En algunos ejemplos no limitativos, la presente divulgación proporciona una isoforma de ALK que comprende los exones 20-29 pero que carece del dominio transmembrana y/o el dominio extracelular. La TALK surge como resultado del establecimiento de un sitio de iniciación transcripcional alternativo *de novo* (“ATI”) en lugar de una reordenación genómica o una aberración genómica en el locus de ALK. Por tanto, la TALK se denomina como *ALK<sup>ATI</sup>*.

25 En determinados ejemplos no limitativos, la isoforma de ALK no está compuesta por fusión con una parte de otra proteína nativa, por ejemplo, que surge de un evento de translocación, por ejemplo, donde la parte 3’ de ALK se fusiona con una parte de la proteína ATIC, C2orf44, CARS, CLTC, EML4, FN1, KIF5B, KLC1, MSN, NPM1, PPFIBP1, RANBP2, SEC31A, SQSTM1, STRN, TFG, TPM3, TPM4, o VCL (véase la base de datos COMIC, <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>)

30 En determinadas realizaciones no limitativas, la presente invención proporciona una isoforma de ALK codificada por un transcrito de ARNm que comprende los exones 20-29 y una parte, pero no todo, del intrón 19, pero no exones comprendidos en el dominio transmembrana y/o dominio extracelular de ALK. En determinadas realizaciones no limitativas, la presente invención proporciona una isoforma de ALK codificada por un transcrito de ARNm que comprende los exones 20-29 y una parte, pero no todo, del intrón 19, pero no los exones 1-19, del gen ALK, o bien individualmente o bien en cualquier combinación. En determinadas realizaciones no limitativas, la presente invención proporciona una molécula de ácido nucleico que comprende el transcrito. En determinadas realizaciones no limitativas, el tamaño del transcrito es de aproximadamente 2300 - 2600 bases, o aproximadamente 2500 bases, o aproximadamente 2513 bases. Determinadas realizaciones no limitativas del transcrito comprenden aproximadamente 400 bases de intrón 19 en el sentido de 5’ del exón 20. Determinadas realizaciones no limitativas de la invención proporcionan una molécula de ADNc correspondiente a dicho transcrito de ARNm. Dicha molécula de ADNc puede estar opcionalmente vinculada a un promotor y/o incorporada en un vector de ácido nucleico. Dicho promotor puede estar situado directamente antes de la secuencia de codificación. En determinados ejemplos no limitativos, el promotor es un promotor heterólogo (es decir, no el promotor de ALK). En determinados ejemplos no limitativos, dicho transcrito de ARNm o molécula de ADNc carece de secuencia de ácido nucleico que codifica al menos un fragmento largo de 10 o 20 aminoácidos de una proteína distinta de ALK.

En determinadas realizaciones no limitativas, el ALK es ALK humano.

50 En determinados ejemplos no limitativos, el gen ALK humano correspondiente tiene una secuencia según lo dispuesto por la secuencia de referencia de NCBI: NG\_009445.1 n.º de acceso NG\_009445 o una variante alélica del mismo.

55 En determinados ejemplos no limitativos, la proteína de ALK humana tiene la secuencia de aminoácidos proporcionada por la secuencia de referencia de NCBI: NP\_004295.2 n.º de acceso NP\_004295 (proporcionado más adelante) o una variante alélica de la misma.

## ES 2 811 927 T3

1 mgaigllwll plllstaavg sgmgtgqrag spaagpplqp replsysrlq rkslavdfvv  
61 pslfrvyard lllppsssel kagrpeargs laldcapllr llgpapgvsw tagspapaea  
121 rtilsrvlkkg svrklrrakq lvlelgeeai legcvppge aavglqlqfnl selfswwirq  
181 gegrlrirlm pekkasevgr egrlsaaira sqprllfqif gtghsslesp tnmppspdy  
241 ftwnltwimk dsfpflshrs ryglecsfdf pceleysppl hdlrnqswsw rripseeasq  
301 mdlldqpgae rskemprgsf lllntsadsk htllspwmrs ssehctlavs vhrhlqpsgr  
361 yiaqllphne aareillmpt pgkhgwtvlq grigrpdnpf rvaleyissg nrslsavdff  
421 alknscsegs pgskmalqss ftcwngtvqlq lqqacdfhqd caqgedesqm crklpvgfyc  
481 nfedgfcgwt qgtlsphtpq wqvrtlkdar fqdhqdhall lsttdvpase satvtsatfp  
541 apiksspcel rmswlrngvl rgnvslvlve nktgkeggrm vwhvaayegl slwqwmvpl  
601 ldvsdrfwlq mvawwqgsr aivafdnisi sldcyltisg edkilqntap ksrnlfernp  
661 nkelkpgens prqtpifdpt vhwlfctoga sqphgptqaa cnnayqnsnl svevgsegpl  
721 kgiqiwwkpa tdtysisgyg aaggkqgknt mmrshgvsvl gifnlekddm lyilvgqqge  
781 dacpstnqli qkvcigennv ieeeirvnrs vhwagggggg gggatyvfkmd kgdvpvplii  
841 aaggggrayg aktdtfhper lennssvlgl ngnsagaagg ggwndntsll wagkslqega  
901 tggscpqam kkwgwetrgg fggggggcss ggggggyigg naasnndpem dgedgvsfis  
961 plgilytpal kvmeghgevn ikhylncshc evdechmdpe shkvicfcdh gtvlaedgvs  
1021 civsptpeph lplslilsvv tsalvaalvl afsgimivyr rkhqelqamq melqspeykl  
1081 sklrtstimt dynpnycfag ktssisdike vprknitlir glghgafgev yeqvsgmpn  
1141 dpsplqvavk tlpevcseqd eldfmeali iskfnhqniv rcigvslqsl prfillelma  
1201 ggdlksflre trprpsqps lamldllhva rdiacgcqyl eenhfihrdi aarnclltcp  
1261 gpgrvakigd fgwardiyra syyrkgcam lpvkmppea fmegiftskt dtwsfgvllw  
1321 eifslgymphy psksnqevle fvtsggrmdp pkncpgpvyr imtqcwqhqp edrpnfaiil  
1381 erieyctqdp dvintalpie ygplveeeek vpvprkdpeg vppllvsqqa kreespaa  
1441 ppplpttssg kaakkptaee isvrprgpa vegghvnmf sqsnppselh kvhgsrnkpt  
1501 slwnptygsw ftekptkkn piakkephdr gnlglegsct vppnvatgrl pgasllleps [SEQ  
ID NO: 1]

5 En determinados ejemplos no limitativos, el ARNm de ALK humano de tipo silvestre o no truncado tiene la secuencia de nucleótidos proporcionada por la secuencia de referencia de NCBI: NM\_004304.4 n.º de acceso NM\_004304 (proporcionado a continuación) o una variante alélica del mismo.

1 agctgcaagt ggcggcgcc caggcagatg cgatccagcg gctctggggg cggcagcgt  
61 ggtagcaget ggtacctccc gccgcctctg ttcggagggt cgcggggcac cgaggtgctt  
121 tccggccgcc ctctggtcgg ccacccaaag ccgcggcgcc tgatgatggg tgaggagggg  
181 gcggcaagat ttcggcgcc cctgcctga acgcctcag ctgctgccgc cggggccgct  
241 ccagtgcctg cgaactctga ggagccgagg cgccggtgag agcaaggacg ctgcaaactt

# ES 2 811 927 T3

301 gcgcagcgcg ggggctggga ttcacgcca gaagttcagc aggcagacag tccgaagcct  
 361 tcccgcagcg gagagatagc ttgagggtgc gcaagacggc agcctccgcc ctcggttccc  
 421 gccagaccg ggcagaagag cttggaggag ccaaaggaa cgaaaaggc ggccaggaca  
 481 gcgtgcagca gctgggagcc gccgttctca gccttaaaag ttgcagagat tggaggatgc  
 541 cccgagaggg gacagacccc agctccgact gcggggggca ggagaggacg gtacccaact  
 601 gccacctccc ttcaaccata gtagttctc tgtaccgagc gcagcgagct acagacgggg  
 661 gcgcggcact cggcgcgag agcgggaggc tcaaggtecc agccagtgag cccagtgtgc  
 721 ttgagtgtct ctggactcgc ccctgagctt ccaggctctgt ttcatttaga ctctgtctg  
 781 cctccgtgca gttgggggaa agcaagagac ttgcgcgcac gcacagtctct ctggagatca  
 841 ggtggaagga gccgctgggt accaaggact gttcagagcc tcttcccatc tcggggagag  
 901 cgaaggggta ggctgggccc ggagagcagt gtaaacggcc tctccggcg ggatgggagc  
 961 catcgggctc ctgtggctcc tgccgtgct gctttccacg gcagctgtgg gctccgggat  
 1021 ggggaccggc cagcgcgcgg gctcccagc tgcggggccg ccgctgcagc cccgggagcc  
 1081 actcagctac tcgcgctgc agaggaagag tctggcagtt gacttcgtgg tgccctcgct  
 1141 cttccgtgtc tacgcccggg acctactgct gccaccatcc tctcggagc tgaaggctgg  
 1201 caggcccag gcccgcggct cgctagctct ggactgcgcc ccgctgctca ggttgctggg  
 1261 gccggcgcg ggggtctcct ggaccgccg ttcaccagcc ccggcagag cccggacgct  
 1321 gtccagggtg ctgaagggcg gctccgtgcg caagctccgg cgtgccaaagc agttggtgct  
 1381 ggagctgggc gaggagcga tcttgagggt ttgcgtcggg cccccgggg aggcggctgt  
 1441 ggggctgctc cagttcaatc tcagcgagct gttcagttgg tggattcgcc aaggcgaagg  
 1501 gcgactgag atccgcctga tgcccagaaa gaagcgctcg gaagtgggca gagagggaag  
 1561 gctgtccgcg gcaattcgcg cctcccagcc ccgcttctc ttccagatct tcgggactgg  
 1621 tcatagctcc ttggaatcac caacaaacat gccttctcct tctcctgatt atttacatg  
 1681 gaatctcacc tggataatga aagactcctt ccctttctg tctcatcga gccgatatgg  
 1741 tctggagtgc agctttgact tcccctgtga gctggagtat tcccctccac tgcatgacct  
 1801 caggaaccag agctggtcct ggcgcgcgat cccctccgag gaggcctccc agatggactt  
 1861 gctggatggg cctggggcag agcgttctaa ggagatgcc agaggctcct ttctccttct  
 1921 caacacctca gctgactcca agcacaccat cctgagtccg tggatgagga gcagcagtga  
 1981 gcaactgcaca ctggccgtct cggtgacag gcacctgcag ccctctggaa ggtacattgc  
 2041 ccagctgctg cccacaacg aggctgcaag agagatcctc ctgatgcca ctccagggaa  
 2101 gcatggttgg acagtgtccc agggaagaat cgggcgtcca gacaacccat ttcgagtggc  
 2161 cctggaatac atctccagtg gaaaccgcag ctgtctgca gtggacttct ttgccctgaa  
 2221 gaactgcagt gaaggaacat ccccaggctc caagatggcc ctgcagagct ctttacttg  
 2281 ttggaatggg acagtctcc agcttgggca ggctgtgac ttccaccag actgtgcca  
 2341 gggagaagat gagagccaga tgtgccgga actgcctgtg ggttttact gcaacttga  
 2401 agatggcttc tgtggctgga cccaaggcac actgtcacc cactctcctc aatggcaggt  
 2461 caggacccta aaggatgcc ggttccagga ccaccaagac catgctctat tgctcagtac  
 2521 cactgatgtc cccgcttctg aaagtgtac agtgaccagt gctacgtttc ctgcaccgat  
 2581 caagagctct ccatgtgagc tccgaatgtc ctggctcatt cgtggagtct tgaggggaaa  
 2641 cgtgtccttg gtgctagtgg agaacaaac cgggaaggag caaggcagga tggctggca  
 2701 tgtcgcgcc tatgaaggct tgagcctgtg gcagtggatg gtgttgctc tctcagatgt  
 2761 gtctgacagg ttctggctgc agatggctgc atgggggga caaggatcca gagccatcgt  
 2821 ggcttttgac aatatctcca tcagcctgga ctgctacctc accattagcg gagaggaaa  
 2881 gatcctgcag aatacagcac ccaaatcaag aaacctgttt gagagaaacc caaacaagga  
 2941 gctgaaacc ggggaaaatt caccaagaca gaccccatc tttgacccta cagttcattg

# ES 2 811 927 T3

3001 gctgttcacc acatgtgggg ccagcgggcc ccatggcccc acccaggcac agtgcaacaa  
 3061 cgctaccag aactccaacc tgagcgtgga ggtggggagc gagggcccc tgaaaggcat  
 3121 ccagatctgg aagggtgccag ccaccgacac ctacagcatc tcgggctacg gagctgctgg  
 3181 cgggaaaggc gggagaacaa ccatgatgcg gtcccacggc gtgtctgtgc tgggcatctt  
 3241 caacctggag aaggatgaca tgcgtacat cctggttggg cagcagggag aggacgcctg  
 3301 ccccagtaca aaccagttaa tccagaaagt ctgcattgga gagaacaatg tgatagaaga  
 3361 agaaatccgt gtgaacagaa gcgtgcatga gtgggcagga ggccggaggag gaggggggtg  
 3421 agccacctac gtatttaaga tgaaggatgg agtgccgggtg cccctgatca ttgcagccgg  
 3481 aggtggtggc agggcctacg gggccaagac agacacgttc caccagaga gactggagaa  
 3541 taactcctcg gttctagggc taaacggcaa ttccggagcc gcaggtggtg gaggtggctg  
 3601 gaatgataac acttccttgc tctgggccgg aaaatctttg caggaggggtg ccaccggagg  
 3661 acattcctgc ccccaggcca tgaagaagtg ggggtgggag acaagagggg gtttcggagg  
 3721 ggggtggagg ggggtgctct caggtggagg aggcggagga tatataggcg gcaatgcagc  
 3781 ctcaacaat gaccccgaaa tggatgggga agatggggtt tccttcatca gtccactggg  
 3841 catcctgtac accccagctt taaaagtgat ggaaggccac ggggaagtga atattaagca  
 3901 ttatctaaac tgcagtcact gtgaggtaga cgaatgtcac atggaccctg aaagccacaa  
 3961 ggtcatctgc ttctgtgacc acgggacggg gctggctgag gatggcgtct cctgcattgt  
 4021 gtcaccacc cggagccac acctgccact ctgctgac ctcctgtg tgcctctgc  
 4081 cctcgtggc gccctgttcc tggctttctc cggcatcatg attgtgtacc gccggaagca  
 4141 ccaggagctg caagccatgc agatggagct gcagagccct gagtacaagc tgagcaagct  
 4201 ccgcacctcg accatcatga ccgactacaa cccaactac tgctttgctg gcaagacctc  
 4261 ctccatcagt gacctgaagg aggtgccggc gaaaaacatc accctcattc ggggtctggg  
 4321 ccatggcgcc tttggggagg tgtatgaagg ccagggtgcc ggaatgcca acgaccgaag  
 4381 ccccctgcaa gtggctgtga agacgctgcc tgaagtgtgc tctgaacagg acgaactgga  
 4441 tttcctcatg gaagccctga tcatcagcaa attcaaccac cagaacattg ttcgctgcat  
 4501 tgggggtgagc ctgcaatccc tgccccggtt catcctgctg gagctcatgg cggggggaga  
 4561 cctcaagtcc ttcctccgag agaccgcc tcgcccgagc cagccctcct ccctggccat  
 4621 gctggacctt ctgcacgtgg ctccggacat tgctgtggc tgtcagtatt tggagaaaa  
 4681 ccaactcatc caccgagaca ttgctgccag aaactgcctc ttgacctgtc caggccctgg  
 4741 aagagtggcc aagattggag acttcgggat ggcccagac atctacaggc cgagctacta  
 4801 tagaaaggga ggctgtgcca tgcctccagt taagtggatg ccccagagg ccttcatgga  
 4861 aggaatattc acttctaaaa cagacacatg gtcccttggg gtgctgctat gggaaatctt  
 4921 ttctcttggg tatatgccat accccagcaa aagcaaccag gaagtctctg agtttgtcac  
 4981 cagtggaggc cggatggacc caccgaaga ctgccctggg cctgtatacc ggataatgac  
 5041 tcagtgtctg caacatcagc ctgaagacag gcccaacttt gccatcattt tggagaggat  
 5101 tgaatactgc acccaggacc cggatgtaat caacaccgct ttgccgatag aatatggtcc  
 5161 acttgtggaa gaggaagaga aagtgcctgt gaggcccaag gaccctgagg gggttcctcc  
 5221 tctcctggtc tctcaacagg caaacggga ggaggagcgc agcccagctg ccccaccacc  
 5281 tctgcctacc acctcctctg gcaaggctgc aaagaaacc acagctgcag agatctctgt  
 5341 tcgagtcctt agaggccggc cgtggaagg gggacacgtg aatatggcat tctctcagtc  
 5401 caaccctcct tcggagtgc acaaggcca cggatccaga aacaagcca ccagcttgtg  
 5461 gaaccacacg tacggctcct ggtttacaga gaaaccacc aaaaagaata atcctatagc  
 5521 aaagaaggag ccacacgaca ggggtaacct ggggctggag ggaagctgta ctgtcccacc  
 5581 taacggttga actgggagac ttccgggggc ctcaactgct ctagagccct cttegtgac  
 5641 tgccaatatg aaggaggtac ctctgttcag gctacgtcac ttcccttgtg ggaatgtcaa

## ES 2 811 927 T3

5701 ttacggctac cagcaacagg gcttgccctt agaagccgct actgccctg gagctggctca  
5761 ttacgaggat accattctga aaagcaagaa tagcatgaac cagcctgggc cctgagctcg  
5821 gtcgcacact cacttctctt ccttgggatc cctaagaccg tggaggagag agaggcaatg  
5881 gtccttcac aaaccagaga ccaaatgtca cgttttgttt tgtgccaacc tattttgaag  
5941 taccacaaa aaagctgtat ttgaaaatg ctttagaaag gttttgagca tgggttcac  
6001 ctattctttc gaaagaagaa aatatcataa aaatgagtga taaatacaag gccagatgt  
6061 ggttgcataa ggtttttatg catgtttgtt gtatacttcc ttatgcttct ttcaaattgt  
6121 gtgtgctctg cttcaatgta gtcagaatta gctgcttcta tgtttcatag ttggggtcac  
6181 agatgtttcc ttgccttggt gatgtggaca tgagccattt gaggggagag ggaacggaaa  
6241 taaaggagtt atttgtaatg actaaaa [SEQ ID NO: 2]

5 En determinadas realizaciones no limitativas, la presente invención proporciona una molécula de ácido nucleico aislado que comprende la secuencia de ácido nucleico expuesta en SEQ ID NO: 11 como se muestra en la figura 6D, o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a SEQ ID NO:11, pero que carecen de secuencias que codifican los exones 1-19 de ALK, o bien individualmente o bien en cualquier combinación.

10 En determinadas realizaciones no limitativas, la molécula de ácido nucleico comprende una secuencia de nucleótidos que codifica la secuencia de aminoácidos expuesta en SEQ ID NO: 12 tal como se muestra en la figura 6E, o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a la SEQ ID NO: 12. En determinadas realizaciones no limitativas, la molécula de ácido nucleico comprende una secuencia de nucleótidos que codifica una secuencia de aminoácidos que comprende al menos una (por ejemplo, no más de 2, no más de 3, no más de 4, no más de 5, no más de 6, no más de 7, no más de 8, no más de 9, no más de 10, no más de 11, no más de 12, no más de 13, no más de 14, no más de 15, no más de 16, no más de 17, no más de 18, no más de 19 o no más de 20) variación de aminoácidos con respecto a SEQ ID NO: 12, donde la secuencia resultante no se encuentra en una proteína que se produce de manera natural, por ejemplo, una proteína de tipo silvestre.

20 En determinadas realizaciones no limitativas, la molécula de ácido nucleico comprende una secuencia de nucleótidos que codifica aminoácidos 3-552 de SEQ ID NO: 12 o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a los aminoácidos 3-552 de SEQ ID NO:12. En determinadas realizaciones no limitativas, la molécula de ácido nucleico comprende una secuencia de nucleótidos que codifica una secuencia de aminoácidos que comprende al menos una (por ejemplo, no más de 2, no más de 3, no más de 4, no más de 5, no más de 6, no más de 7, no más de 8, no más de 9, no más de 10, no más de 11, no más de 12, no más de 13, no más de 14, no más de 16, no más de 17, no más de 18, no más de 19 o no más de 20) variación de aminoácidos con respecto a aminoácidos 3-552 de SEQ ID NO: 12, donde la secuencia resultante no se encuentra en una proteína que se produce de manera natural, por ejemplo, una proteína de tipo silvestre.

30 En determinadas realizaciones no limitativas, la molécula de ácido nucleico comprende una secuencia de nucleótidos que codifica aminoácidos 21-552 de SEQ ID NO: 12 o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a los aminoácidos 21-552 de SEQ ID NO:12. En determinadas realizaciones no limitativas, la molécula de ácido nucleico comprende una secuencia de nucleótidos que codifica una secuencia de aminoácidos que comprende al menos una (por ejemplo, no más de 2, no más de 3, no más de 4, no más de 5, no más de 6, no más de 7, no más de 8, no más de 9, no más de 10, no más de 11, no más de 12, no más de 13, no más de 14, no más de 15, no más de 16, no más de 17, no más de 18, no más de 19, o no más de 20) variación de aminoácidos con respecto a aminoácidos 21-552 del SEQ ID NO: 12, donde la secuencia resultante no se encuentra en una proteína que se produce de manera natural, por ejemplo, una proteína de tipo silvestre.

40 En determinadas realizaciones no limitativas, la secuencia de aminoácidos tiene patrones de glucosilación alterados en comparación con una proteína que se produce de manera natural, incluyendo, pero no limitado a, una secuencia de aminoácidos producida en células no humanas, por ejemplo, células de levadura, células de insectos, células CHO, etc.

50 En determinadas realizaciones no limitativas, la presente invención proporciona una molécula de ácido nucleico aislado, que es una molécula de ADNc, comprende ácidos nucleicos 405-2063 de la secuencia de ácidos nucleicos expuesta en SEQ ID NO: 11 tal como se muestra en la figura 6D o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a ácidos nucleicos 405-2063 de SEQ ID NO:11.

5 En determinadas realizaciones no limitativas, la presente invención proporciona una molécula de ácido nucleico aislado, que es una molécula de ADNc, comprende ácidos nucleicos 411-2063 de la secuencia de ácido nucleico expuesta en SEQ ID NO: 11 tal como se muestra en la figura 6D o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a los ácidos nucleicos 411-2063 de SEQ ID NO:11.

10 En determinadas realizaciones no limitativas, la presente invención proporciona una molécula de ácido nucleico aislado, que es una molécula de ADNc, comprende ácidos nucleicos 465-2063 de la secuencia de ácido nucleico expuesta en SEQ ID NO: 11 tal como se muestra en la FIGURA 6D o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a los ácidos nucleicos 465-2063 de SEQ ID NO:11.

15 En determinados ejemplos no limitativos, dicha molécula de ácido nucleico está vinculada operativamente a un promotor, que puede ser un promotor heterólogo. En determinados ejemplos no limitativos, la molécula de ácido nucleico comprende una molécula de ADNc. En determinados ejemplos no limitativos, la molécula de ácido nucleico aislado se compone de un vector que puede, por ejemplo, ser un vector de expresión o un virus.

20 En determinadas realizaciones no limitativas, la presente invención proporciona una célula huésped que comprende un vector que comprende las moléculas de ácido nucleico descritas en el presente documento.

25 La homología de secuencia o identidad de secuencia puede medirse usando software de análisis de secuencia (por ejemplo, paquete de software de análisis de secuencia de Genetics Computer Group, Centro de Biotecnología de la Universidad de Wisconsin, 1710 University Avenue, Madison, Wis. 53705, programas BLAST, BESTFIT, GAP o PILEUP/PRETTYBOX). Este software apareja secuencias idénticas o similares asignando grados de homología a diversas sustituciones, deleciones y/u otras modificaciones. En un enfoque a modo de ejemplo para determinar el grado de identidad, puede usarse un programa BLAST, con una puntuación de probabilidad entre e-3 y e-100 que indica una secuencia estrechamente relacionada.

30 En determinadas realizaciones no limitativas, la presente invención proporciona un polipéptido aislado codificado por cualquiera de las moléculas de ácido nucleico descritas en el presente documento, incluyendo moléculas de ADNc.

35 En determinadas realizaciones no limitativas, la presente invención proporciona un polipéptido aislado que comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 12 como se muestra en la figura 6E, pero que carece de la secuencia de aminoácidos expuesta en SEQ ID NO: 3 (proporcionada a continuación) o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos aproximadamente un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a SEQ ID NO:12.

MGAIGLLWLLPLLLSTAAVGSGMGTGQRAGSPAAGPPLQPREPL  
 SYSRLQRKSLAVDFVVPSTLFRVYARDLLLPPSSSELKAGRPEARGLALDCAPLLRLL  
 GPAPGVSWTAGSPAPAEARTLSRVLKGGSVRKLRRKQLVLELGEEAILEGCVGPPGE  
 AAVGLLQFNLSSELSFWWIRQEGRLRIRLMPEKKASEVREGRLSAAIRASQPRLLFQ  
 IFGTGHSSLESPTNMPSPSPDYFTWNLTWIMKDSFPFLSHRSRYGLECSFDFPCELEY  
 SPPLHDLRNQSWRRI PSEEASQMDLLDGPGAERSKEMPRGSFLLLNTSADSKHTIL

SPWMRSSEHCTLAVSVHRHLQPSGRYIAQLLPHNEAAREILLMPTPGKHGWTVLQGR  
IGRPDNPFRVALEYISSGNRSLSAVDFFALKNCSEGTSPGSKMALQSSFTCWNGTVLQ  
LGQACDFHQDCAQGEDESQMCRKLPVGFYCNFEDGFCGWTQGTLSPHTPQWQVRTLKD  
ARFQDHDHALLLSTTDVPASESATVTSATFPAPIKSSPCELMSWLIRGVLRGVNSL  
VLVENKTGKEQGRMVHVAAYEGLSLWQWMVLPDLLVSDRFWLQMVAVWGWGGSRAIVA  
FDNISISLDCYLTISGEDKILQNTAPKSRNLFERNPNKELKPGENSPRQTPIFDPTVH  
WLFITTCGASGPHGPTQAQCNNAYQNSNLSVEVGSEGPLKGIQIWKVPATDTYSISGYG  
AAGGKGGKNTMMRSHGVSVLGFIFNLEKDDMLYILVGQQGEDACPSTNQLIQKVCIGEN  
NVIEEEEIRVNRSVHEWAGGGGGGGGATYVFKMKDGVVPLIIAAGGGGRAYGAKTDTF  
HPERLENNSSVLGLNGNSGAAGGGGGWINDNTSLLWAGKSLQEGATGGHSCPQAMKKWG  
WETRGGFGGGGGGCSSGGGGGGYIGGNAASNNDPEMDGEDGVSFISPLGILYTPALKV  
MEGHGEVNIKHLYLNCSEVDECHMDPESHKVICFCDHGTVLAEDGVSCIVSPTPEPH  
LPLSLILSVVTSALVAALVLAFFSGIMIVYRRKHQELQA [SEQ ID NO: 3]

5 En determinadas realizaciones no limitativas, la presente invención proporciona un polipéptido aislado que comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 12 como se muestra en la figura 6E, pero que carece de la secuencia de aminoácidos expuesta en SEQ ID NO: 4 (proporcionada a continuación) o una secuencia que es al menos un 80%, al menos un 85%, al menos un 90% o al menos aproximadamente un 95% homóloga con respecto a la misma, incluyendo secuencias que no son idénticas a SEQ ID NO:12.

MGAIGLLWLLPLLLSTAAGVSGMGTGQRAGSPAAGPPLQPREPL  
SYSRLQRKSLAVDFVVPVSLFRVYARDLLLPPSSSELKAGRPEARGLALDCAPLLRLL  
GPAPGVSWTAGSPAPAEARTLSRVLKGGSVRKLRRAKQLVLELGEAAILEGCVGPPGE  
AAVGLLQFNLSSELSFWWIRQEGRLRIRLMPEKKASEVREGRLSAAIRASQPRLLFQ  
IFGTGHSSLESPTNMPSPSPDYFTWNLTWIMKDSFPFLSHRSRYGLECSFDFPCELEY  
SPPLHDLRNQSWSWRRIPSEEASQMDLLDGPGAERSKEMPRGSFLLLNTSADSKHTIL  
SPWMRSSEHCTLAVSVHRHLQPSGRYIAQLLPHNEAAREILLMPTPGKHGWTVLQGR  
IGRPDNPFRVALEYISSGNRSLSAVDFFALKNCSEGTSPGSKMALQSSFTCWNGTVLQ  
LGQACDFHQDCAQGEDESQMCRKLPVGFYCNFEDGFCGWTQGTLSPHTPQWQVRTLKD  
ARFQDHDHALLLSTTDVPASESATVTSATFPAPIKSSPCELMSWLIRGVLRGVNSL  
VLVENKTGKEQGRMVHVAAYEGLSLWQWMVLPDLLVSDRFWLQMVAVWGWGGSRAIVA  
FDNISISLDCYLTISGEDKILQNTAPKSRNLFERNPNKELKPGENSPRQTPIFDPTVH  
WLFITTCGASGPHGPTQAQCNNAYQNSNLSVEVGSEGPLKGIQIWKVPATDTYSISGYG  
AAGGKGGKNTMMRSHGVSVLGFIFNLEKDDMLYILVGQQGEDACPSTNQLIQKVCIGEN  
NVIEEEEIRVNRSVHEWAGGGGGGGGATYVFKMKDGVVPLIIAAGGGGRAYGAKTDTF  
HPERLENNSSVLGLNGNSGAAGGGGGWINDNTSLLWAGKSLQEGATGGHSCPQAMKKWG  
WETRGGFGGGGGGCSSGGGGGGYIGGNAASNNDPEMDGEDGVSFISPLGILYTPALKV  
MEGHGEVNIKHLYLNCSEVDECHMDPESHKVICFCDHGTVLAEDGVSCIVSPTPEPH  
LPLSLILSVVISA [SEQ ID NO: 4]

10 En determinadas realizaciones no limitativas, el polipéptido comprende los aminoácidos 3 a 552 de SEQ ID NO: 12.  
En determinadas realizaciones no limitativas, el polipéptido comprende los aminoácidos 21 a 552 de SEQ ID NO: 12.

15 En determinadas realizaciones no limitativas, el polipéptido aislado comprende una secuencia de aminoácidos que comprende al menos una (por ejemplo, no más de 2, no más de 3, no más de 4, no más de 5, no más de 6, no más de 7, no más de 8, no más de 9, no más de 10, no más de 11, no más de más de 12, no más de 13, no más de 14, no

- más de 15, no más de 16, no más de 17, no más de 18, no más de 19, o no más de 20) variación de aminoácidos con respecto a SEQ ID NO: 12, donde la secuencia resultante no se encuentra en una proteína que se produce de manera natural, por ejemplo, una proteína de tipo silvestre. En determinadas realizaciones, la secuencia de aminoácidos comprende al menos una variación de aminoácidos con respecto a SEQ ID NO: 12 pero carece de la secuencia de aminoácidos expuesta en SEQ ID NO: 3. En determinadas realizaciones, la secuencia de aminoácidos comprende al menos una variación de aminoácidos con respecto a SEQ ID NO: 12 pero carece de la secuencia de aminoácidos expuesta en SEQ ID NO: 4.
- En determinados ejemplos no limitativos, la secuencia de aminoácidos tiene patrones de glucosilación alterados comparada con una proteína que se produce de manera natural, incluyendo, pero no limitada a, una secuencia de aminoácidos producida en células no humanas, por ejemplo, células de levadura, células de insecto, células CHO, etc.
- Las variaciones o modificaciones de aminoácidos pueden incluir sustituciones de aminoácidos, adiciones y/o deleciones. Pueden introducirse variaciones o modificaciones mediante técnicas estándar conocidas en la técnica, como la mutagénesis dirigida al sitio y la mutagénesis mediada por PCR. En determinados ejemplos, la al menos una variación de aminoácidos comprende al menos una modificación conservadora de aminoácidos. Los aminoácidos pueden clasificarse en grupos según sus propiedades fisicoquímicas como carga y polaridad. Las sustituciones conservadoras de aminoácidos son aquellas en las que el residuo de aminoácidos se reemplaza por un aminoácido dentro del mismo grupo. Por ejemplo, los aminoácidos pueden clasificarse por carga: aminoácidos cargados positivamente incluyen lisina, arginina, histidina, aminoácidos cargados negativamente incluyen ácido aspártico, ácido glutámico, aminoácidos de carga neutra incluyen alanina, asparagina, cisteína, glutamina, glicina, isoleucina, leucina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina y valina. Además, los aminoácidos pueden clasificarse por polaridad: aminoácidos polares incluyen arginina (polar básico), asparagina, ácido aspártico (polar ácido), ácido glutámico (polar ácido), glutamina, histidina (polar básico), lisina (polar básico), serina, treonina y tirosina; aminoácidos no polares incluyen alanina, cisteína, glicina, isoleucina, leucina, metionina, fenilalanina, prolina, triptófano y valina.
- En determinadas realizaciones no limitativas, el polipéptido comprende los aminoácidos 3 a 552 de SEQ ID NO: 12.
- En determinadas realizaciones no limitativas, el polipéptido comprende los aminoácidos 21 a 552 de SEQ ID NO: 12.
- En determinadas realizaciones no limitativas, la presente invención proporciona un polipéptido aislado que comprende la secuencia de aminoácidos expuesta en los residuos 1-552 de SEQ ID NO: 12 como se muestra en la figura 6E. El polipéptido tiene un peso molecular estimado de aproximadamente 61 kDa (por ejemplo, 61,08 kDa).
- En determinadas realizaciones no limitativas, la presente invención proporciona un polipéptido aislado que comprende la secuencia de aminoácidos expuesta en los residuos 3-552 de SEQ ID NO: 12 como se muestra en la figura 6E. El polipéptido tiene un peso molecular estimado de aproximadamente 61 kDa (por ejemplo, 60,82 kDa).
- En determinadas realizaciones no limitativas, la presente invención proporciona un polipéptido aislado que comprende la secuencia de aminoácidos expuesta en los residuos 21-552 de SEQ ID NO: 12 como se muestra en la figura 6E. El polipéptido tiene un peso molecular estimado de aproximadamente 59 kDa (por ejemplo, 58,71 kDa).
- En determinados ejemplos no limitativos, el polipéptido tiene un peso molecular estimado de entre 58-64, por ejemplo, aproximadamente 61 kDa o aproximadamente 59 kDa.
- En determinados ejemplos no limitativos, el polipéptido está vinculado a una etiqueta de péptido heterólogo de entre aproximadamente 2-20 aminoácidos, o entre aproximadamente 3-15 aminoácidos.
- En determinados ejemplos no limitativos, el polipéptido está etiquetado de manera detectable.
- En determinados ejemplos no limitativos, el polipéptido contiene al menos una sustitución de aminoácidos conservadora con respecto a la secuencia de tipo silvestre.
- En determinadas realizaciones no limitativas, la presente invención proporciona un anticuerpo que se une a uno de los polipéptidos descritos anteriormente, pero que no se une a un polipéptido de ALK de tipo silvestre, por ejemplo, uno que comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 1.
- 5.3 Detección de TALK
- La detección de TALK puede realizarse por cualquier método conocido en la técnica. Una TALK puede detectarse en forma de un transcrito de ARNm, una molécula de ADNc correspondiente, su polipéptido codificado, o un sitio de ATI en el intrón 19 de ALK. En una realización no limitante, la TALK es  $ALK^{ATI}$ .
- En determinados ejemplos no limitativos, un transcrito de ARNm que codifica una TALK puede detectarse mediante un método que comprende hibridación de sonda (por ejemplo, ensayo basado en sonda nCounter, de NanoString),

amplificación, transcripción inversa, reacción en cadena de la polimerasa (PCR, por ejemplo, reacción en cadena de la polimerasa de transcripción inversa (RT-PCR), PCR de transcriptasa inversa cuantitativa, PCR en tiempo real, PCR en tiempo real cuantitativa), transferencia de tipo Northern, secuenciación, microalineamiento o una combinación de los mismos.

5 En determinados ejemplos no limitativos, un polipéptido de TALK puede detectarse mediante un método que comprende unión a anticuerpos, inmunohistoquímica, inmunotransferencia de tipo Western, ensayo funcional (por ejemplo, cinasa), ensayos inmunosorbentes ligados a enzimas (ELISA), radioinmunoensayos (RIA), inmunoensayos enzimáticos (EIA), espectrometría de masas, sistemas de análisis basados en gel 1-D o 2-D (por ejemplo, electroforesis en gel de poliacrilamida (PAGE)), inmunoprecipitación o una combinación de los mismos.

En determinados ejemplos no limitativos, una molécula de ADNc de TALK puede detectarse por PCR.

15 En determinados ejemplos no limitativos, un sitio de ATI en el intrón 19 de un gen ALK puede detectarse por un método que comprende transferencia de tipo Northern, secuenciación de inmunoprecipitación de cromatina (ChIP), ChIP-qPCR, amplificación rápida de extremos de ADNc (RACE)-PCR, o una combinación de los mismos.

Puede ponerse en práctica un método de detección en una célula o tejido o lisado celular recolectado de un sujeto. La célula puede recogerse como muestra, por ejemplo, una muestra de un tumor o neoplasia u otra lesión. La muestra puede ser una muestra de sangre periférica u otra muestra de fluido (por ejemplo, derrame pleural).

25 En un ejemplo particular no limitativo, el sitio de inicio de transcripción y de extremo 5' de un transcrito de TALK, por ejemplo el transcrito de *ALK<sup>ATL</sup>*, pueden determinarse usando la técnica RACE en 5'. Por ejemplo, una técnica tabaco-ácido-pirofosfatasa RACE en 5' puede realizarse según el protocolo de fabricación (kit de RACE de ARNm eucariótico ExactStart, n.º ES80910, Epicentre) usando los siguientes cebadores: 5'-TCATACACATACGATTTAGGTGACACTATAGAGCGGCCGCCTGCAGGAAA -3' [SEQ ID NO: 5]; inverso 5'-CAGGTCCTGATGGAGGAGGTCTTGCCAGCAAAGCA-3' [SEQ ID NO: 6]. Los productos RACE en 5' pueden entonces secuenciarse, por ejemplo, usando un sistema Illumina MiSeq con un protocolo de extremos apareados de 150 pb según el protocolo de fabricación. Alternativamente, pueden utilizarse los siguientes cebadores inversos: 30 cebador directo 5'-CTAATACGACTCACTATAGGGC-3' [ID SEQ NO: 7], cebador inverso 5'-ACACCTGGCCTTCATACACCTCC-3' [SEQ ID NO: 8].

35 En otro ejemplo particular no limitativo, una molécula de ADNc de un transcrito de TALK (por ejemplo, la *ALK<sup>ATL</sup>*) puede amplificarse usando los cebadores específicos de ALK 5'-CACCATCCCATCTCCAGTCTGCTTC-3' [SEQ ID NO: 9] y 5'-AGAGAAGTGAGTGTGCGACC-3' [SEQ ID NO: 10].

40 En otro ejemplo particular no limitativo, la presencia de una TALK puede detectarse usando un anticuerpo específico de ALK, así como una técnica que demuestra un peso molecular inferior a la proteína de tipo silvestre, tal como, pero no limitado a electroforesis en gel de poliacrilamida e inmunotransferencia de tipo Western. Ejemplos no limitativos de anticuerpos que pueden usarse incluyen anticuerpo anti- $\alpha$ -tubulina (#T9026-0.5ML), anti-V5 (#MA1-81617) y anti-HA3F10 (#12158167001), AcM de conejo frente a fosfo-ALK (Tyr1278/1282/1283) (#3983), AcM de conejo frente a fosfo-ALK (Tyr1604) (#3341), AcM de conejo frente a ALK (D5F3) (#3633), fosfo-Akt (Ser473) (#4060), Akt (#4685), fosfo-Stat3 (Tyr705) (#9145), Stat3 (#4904), fosfo-S6 (Ser235/236) (#4858), S6 (#2217), fosfo-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (#4370), p44/42 MAPK (Erk1/2) (#4695), fosfo-MEK1/2 (#2338), MEK1/2 (#9122).

45 En determinados ejemplos no limitativos, puede detectarse una TALK a través de la presencia de un sitio de ATI en el intrón 19 de ALK, por ejemplo, usando secuenciación de ChIP y ChIP-qPCR. Las marcas de cromatina "histona H3 lisina 4 trimetilada (H3k4me3)" y "ARN polimerasa II" están bien caracterizadas para enriquecer en sitio de ATI equilibrado y activamente transcrito.<sup>17-19</sup>

50 En un determinado ejemplo no limitativo, un ensayo nCounter de NanoString con conjuntos de sondas en los exones 1-19, exones 20-29 e intrón 19 de ALK puede usarse (véase la figura 1H y su leyenda) para distinguir ALK de tipo silvestre (*ALK<sup>WT</sup>*), ALK translocada y transcritos de *ALK<sup>ATL</sup>*. El ensayo nCounter de NanoString se basa en la detección digital y la codificación en barras molecular directa de moléculas diana mediante el uso de un par de sondas codificadas por colores. El par de sondas consiste en una sonda indicadora, que lleva la señal en su extremo 5', y una sonda de 55 captura que lleva una biotina en el extremo 3'.<sup>51</sup>

60 En determinados ejemplos no limitativos, una TALK puede detectarse a través de su localización celular, en la que TALK tiene una mayor presencia en el núcleo con respecto a ALK de tipo silvestre. Tales estudios pueden realizarse, por ejemplo, usando inmunohistoquímica y sondas de anticuerpos fluorescentes primarios o secundarios.

#### 5.4 Métodos de diagnóstico

65 Se da a conocer en el presente documento un método de diagnóstico de una célula como célula cancerosa que comprende detectar la presencia de una TALK en una célula, donde la presencia de una TALK indica que la célula es una célula cancerosa. La célula puede estar en una muestra recolectada de un sujeto. En determinados ejemplos no

limitativos, una muestra incluye, pero no se limita a, una muestra clínica, células en cultivo, sobrenadantes celulares, lisados celulares, suero, plasma sanguíneo, fluido biológico (por ejemplo, fluido linfático) y muestras de tejido. La fuente de la muestra puede ser tejido sólido (por ejemplo, a partir de un órgano, muestra de tejido, biopsia o aspirado en fresco, congelado y/o preservado), sangre o cualquier componente sanguíneo, fluidos corporales (como, por ejemplo, orina, linfa, fluido espinal cerebral, líquido amniótico, fluido peritoneal o fluido intersticial), o células del individuo, incluyendo células tumorales circulantes.

En determinados ejemplos no limitativos, cuando una TALK se encuentra en una célula, la célula u otra célula del mismo sujeto (por ejemplo, una o más células adicionales de la misma muestra) puede someterse a prueba para otros marcadores indicativos de un diagnóstico de cáncer, como se conoce en la técnica.

Ejemplos no limitativos de cánceres que pueden indicarse por la presencia de una TALK incluyen melanoma, carcinoma de tiroides (por ejemplo, carcinoma de tiroides anaplásico), adenocarcinoma de pulmón, carcinoma de células escamosas pulmonares, carcinoma de células claras renales y cáncer de mama. En una realización particular no limitativa, el cáncer es melanoma.

Se da a conocer en el presente documento un método para determinar si un efecto anticancerígeno es probable que se produzca en un cáncer por un inhibidor de ALK, que comprende determinar si una o más células del cáncer contiene una TALK detectable, en el que la presencia de una TALK detectable en la célula indica que un inhibidor de ALK tendría un efecto anticancerígeno sobre el cáncer.

Se da a conocer en el presente documento un método para determinar la probabilidad de que un sujeto que tiene un cáncer pueda obtener un beneficio terapéutico de la terapia con un inhibidor de ALK, que comprende determinar si una célula cancerosa del sujeto contiene una TALK detectable, donde la presencia de una TALK detectable indica que es más probable que el sujeto se beneficie de la terapia inhibidora de ALK que un sujeto que carece de TALK o una mutación genética activadora de ALK.

En determinados ejemplos, la presencia de una TALK detectable comprende la presencia de un transcrito de ARNm de TALK detectable, una molécula de ADNc de TALK detectable correspondiente a la misma, un polipéptido de TALK detectable codificado mediante el mismo, y/o un sitio de ATI detectable en el intrón19 de un gen ALK en la célula.

En determinados ejemplos no limitativos, una célula cancerosa de un sujeto puede tener un locus de ALK de tipo salvaje (o normal) y sin embargo expresar una TALK, y por lo tanto no sería identificado como que tiene un ALK activado por pruebas genómicas.

### 5.5 Métodos de tratamiento

Se da a conocer en el presente documento un método para tratar a un sujeto que padece cáncer, que comprende (i) determinar la probabilidad de que el sujeto obtenga un beneficio terapéutico de la terapia con un inhibidor de ALK, que comprende determinar si una célula cancerosa del sujeto contiene una TALK detectable, cuando la presencia de una TALK detectable indica que es más probable que el sujeto se beneficie de la terapia inhibidora de ALK que un sujeto que carece de TALK o una mutación genética activadora de ALK; (ii) cuando se determina que el sujeto tiene más probabilidades de beneficiarse de la terapia con inhibidores de ALK, tratar al sujeto con un inhibidor de ALK, y (iii) cuando se determina que el sujeto no tiene más probabilidades de beneficiarse de la terapia con inhibidores de ALK, tratar al sujeto con una terapia alternativa que no sea un inhibidor de ALK pero que pueden ser tratamiento quirúrgico, radioterapia o quimioterapia.

Ejemplos no limitativos de cánceres que pueden tratarse con el método anterior incluyen melanoma, carcinoma de tiroides (por ejemplo, carcinoma de tiroides anaplásico), adenocarcinoma de pulmón, carcinoma de células escamosas pulmonares, carcinoma de células claras renales, cáncer de tiroides y cáncer de mama. En una realización particular no limitativa, el cáncer es melanoma.

Ejemplos no limitativos de inhibidores de ALK que pueden utilizarse en el método anterior incluyen crizotinib ("Xalkori®"; (Pfizer); ceritinib (Zykadia®; también conocido como "LDK-378; Novartis), NVP-TAE684 (Novartis), alectinib (Chugai); AP26113 (Ariad); ASP-3026 (Astellas); CEP-37440 (Teva); NMS-E628 (Nerviano); PF-06463922 (Pfizer); TSR-011 (Tesoro); RXDX-101 (Ignyta Inc.), y X-396 (Xcovery). En una realización particular no limitativa, el inhibidor de ALK es crizotinib.

Sujetos tratados con uno o más inhibidores de ALK pueden además tratarse con uno o más tratamientos de cáncer adicionales, incluyendo pero no limitado a, uno o más inhibidor de BRAF, uno o más inhibidores de MEK, uno o más inhibidores inmunológicos (por ejemplo, un anticuerpo anti-CTLA4, un anticuerpo anti-PD1, un anticuerpo anti-PD-L1), uno o más inhibidores de CDK4, uno o más inhibidores de CDK6, uno o más agentes alquilantes, uno o más inhibidores de topoisomerasa, uno o más antimetabolitos, uno o más agentes antimicrotúbulos, uno o más antibióticos citotóxicos, radioterapia, quimioterapia o una combinación de los mismos. En un ejemplo particular no limitante, el sujeto recibe una combinación de un anticuerpo anti-CTLA-4 y un anticuerpo PD-1 además de un inhibidor de ALK. En un ejemplo no limitativo, el sujeto ha sido tratado con una combinación de un anticuerpo anti-CTLA-4 y un anticuerpo anti-PD-1

antes del tratamiento con un inhibidor de ALK. En determinados ejemplos, el sujeto recibe además radiación y/o quimioterapia.

#### 5.6 Kits

5 La presente divulgación comprende un kit que comprende un medio para detectar una TALK como se ha descrito anteriormente, opcionalmente junto con la divulgación escrita de uno o más métodos anteriormente expuestos, y opcionalmente junto con una molécula de control positiva como un ácido nucleico y/o proteína de TALK.

10 En determinados ejemplos, el kit comprende medios para determinar el nivel de un transcrito de ARNm de TALK, una molécula de ADNc de TALK correspondiente a la misma y/o un polipéptido de TALK codificado por el mismo. Los medios para determinar el nivel de un transcrito de ARNm de TALK pueden ser la hibridación de sonda (por ejemplo, ensayo nCounter de NanoString basado en sonda), reacción en cadena de la polimerasa (PCR, por ejemplo, RT-PCR, PCR cuantitativa de transcriptasa inversa, PCR en tiempo real o PCR cuantitativa en tiempo real), transferencia de tipo Northern, secuenciación, microalineamiento o una combinación de los mismos.

15 En determinados ejemplos, el medio para determinar el nivel de una molécula de ADNc de TALK puede ser PCR.

20 En determinados ejemplos, los medios para determinar el nivel de un polipéptido de TALK pueden ser unión a anticuerpos, inmunohistoquímica, inmunotransferencia de tipo Western, un ensayo funcional (por ejemplo, cinasa), ensayos inmunoabsorbentes ligados a enzimas (ELISA), radioinmunoensayos (RIA), inmunoensayos enzimáticos (EIA), espectrometría de masas, sistemas de análisis basados en gel 1-D o 2-D (por ejemplo, electroforesis en gel de poliacrilamida (PAGE)), inmunoprecipitación o una combinación de los mismos.

25 El kit puede comprender medios para determinar la presencia de un sitio de ATI en el intrón 19 de un gen ALK en una o más células del cáncer, que puede ser transferencia de tipo Northern, secuenciación de inmunoprecipitación de cromatina (ChIP), ChIP-qPCR, PCR-amplificación rápida de extremos de ADNc (RACE) o una combinación de los mismos.

30 La presente divulgación proporciona además un medio para determinar la presencia de una TALK detectable para su uso en un método de determinación de si un efecto anticancerígeno es probable que se produzca en un cáncer por un inhibidor de ALK, el método caracterizado por determinar si una o más células del cáncer contiene una TALK detectable, donde la presencia de una TALK detectable en la célula indica que un inhibidor de ALK tendría un efecto anticancerígeno sobre el cáncer.

35 Este kit puede, por tanto, comprender una o más sondas de ácido nucleico (por ejemplo, que hibrida con un transcrito de ARNm de TALK (incluyendo un par de sondas codificadas por colores usadas en un ensayo nCounter de NanoString), o que hibrida con una molécula de ADNc de TALK), cebador, y/o un par de cebadores (por ejemplo, con un transcrito de ARNm de TALK o una molécula de ADNc de TALK), y/o uno o más anticuerpos, fragmentos de anticuerpos o anticuerpos de cadena única (por ejemplo, que se une específicamente a un polipéptido de TALK), cualquiera de los cuales está opcionalmente etiquetado directamente con un marcador químico, fluorescente, enzimático o radiactivo. Alternativamente, la detección puede realizarse a través de una sonda secundaria, enzima, sustrato, ligando, anticuerpo, fragmento de anticuerpo, anticuerpo de cadena única, etc., usando técnicas conocidas en la técnica.

45 En un ejemplo específico no limitativo, el kit comprende el par de cebadores: 5'-TCATACACATACGATTTAGGTGACACTATAGAGCGGCCGCCTGCAGGAAA -3' [SEQ ID NO: 5]; inverso 5'-CAGGTCCTGATGGAGGAGGTCTTGCCAGCAAAGCA- 3' [SEQ ID NO: 6], por ejemplo, pero no a modo de limitación, para su uso en una técnica RACE en 5'.

50 En un ejemplo específico no limitativo, el kit comprende el par de cebadores: cebador directo 5'-CTAATACGACTCACTATAGGGC- 3' [SEQ ID NO: 7], cebador inverso 5'-ACACCTGGCCTTCATACACCTCC-3' [SEQ ID NO: 8], por ejemplo, pero no a modo de limitación, para su uso en una técnica RACE en 5'.

55 En un ejemplo específico no limitativo, el kit comprende el par de cebadores: 5'-CACCATCCCATCTCCAGTCTGCTTC- 3' [SEQ ID NO: 9] y 5'-AGAGAAGTGAGTGTGCGACC-3' [SEQ ID NO: 10].

60 En un ejemplo específico, no limitativo, el kit comprende un anticuerpo o un fragmento del mismo que se une a ALK en una región comprendida en los exones 20-29 y opcionalmente un anticuerpo o un fragmento del mismo que se une a ALK en una región comprendida en los exones 1-19.

En un ejemplo específico, no limitativo, el kit comprende medios para identificar un sitio de ATI en el intrón 19 de ALK.

65 En un ejemplo específico no limitativo, el kit comprende medios para detectar la localización nuclear de TALK, por ejemplo, una sonda de ácido nucleico o anticuerpo y una segunda sonda de ácido nucleico o anticuerpo que se une específicamente a un compañero de unión en el núcleo o un medio químico para identificar o teñir el núcleo.

## 6. Ejemplo: La iniciación transcripcional alternativa conduce a la expresión de una isoforma novedosa de *ALK* en cáncer

### 5 6.1 Materiales y métodos

Muestras de tumor humano. El estudio se aprobó por los Comités Institucionales de Revisión/Comités de Ética del Centro del cáncer Memorial Sloan-Kettering, Nueva York, y se llevó a cabo según la Declaración de Helsinki. Las partes representativas de los tumores extirpados se ultracongelaron en nitrógeno líquido o se fijaron en formol tamponado neutro al 4%, se incrustaron en parafina, se procesaron usando métodos histológicos rutinarios y se tiñeron con hematoxilina-eosina. Se excluyeron especímenes con cantidad insuficiente de tejido o ácidos nucleicos gravemente degradados.

15 Secuenciación de ARN. Se extrajo ARN total de secciones de tejido congeladas en fresco (17 de melanoma metastásico y 6 de carcinoma de tiroides) usando un minikit RNEasy de Qiagen (#74104, Qiagen). El ARN aislado se procesó usando el kit de preparación de muestras de ARN TruSeq (#15026495, Illumina) según el protocolo de fabricación. Brevemente, el ARN se seleccionó por poli-a, se sometió a transcripción inversa y el ADNc obtenido se sometió a un proceso de reparación de extremos, preparación de colas-A, ligadura de los índices y adaptadores y enriquecimiento por PCR. Las bibliotecas creadas se secuenciaron en una plataforma HiSeq-2500 de Illumina con lecturas de extremos apareados de 50, 75 o 100 pb para obtener en promedio de 40 a 100 millones de lecturas por muestra. Los datos de secuenciación se mapearon con respecto al genoma de referencia humano (hg19) usando Bowtie o BWA y se analizaron usando paquetes de software disponibles públicamente: SAMtools<sup>26</sup>, Tophat<sup>27</sup>, FusionSeq<sup>28</sup>, GATK<sup>29</sup>, Picard (<http://picard.sourceforge.net>) e IGV<sup>30</sup>.

25 Cribado de cinasas expresadas de manera aberrante. Para el cribado inicial de datos de secuenciación de ARN, se definieron los genes de tirosina cinasa receptora candidatos (RTK) mediante la anotación de Gene Ontology GO:0004714 tal como se encuentra en AmiGO<sup>31</sup>. Se usó DEXSeq<sup>9</sup> para calcular los recuentos de niveles de exón usando identificadores de gen de RTK Ensembl. Para cada gen en cada muestra, se calculó la razón de lecturas en la primera mitad del gen con respecto a la segunda mitad. Se identificó *ALK* como el mayor acierto.

30 Análisis de conjuntos de datos públicos. Los datos de secuenciación de ARN se descargaron desde el portal de expresión de genotipo tisular GTEx del Broad Institute (<http://www.broadinstitute.org/gtex/>), ejecución del 23 de 09 de 2013, usando datos de cuantificación de exones (del inglés, exon\_quantification) de illuminaiseq\_rnaseqv2\_unc\_uvu. Se identificaron candidatos de *ALK*<sup>AT7</sup> como muestras con un nivel de expresión de *ALK* de RSEM  $\geq 100$ ,  $\geq 500$  lecturas totales en todos los exones de *ALK* y  $\geq 10$  veces mayor expresión promedio (por RPKM de nivel de exón) en los exones 20-29 en comparación con los exones 1-19. Para confirmar la expresión de *ALK*<sup>AT7</sup>, se examinaron manualmente los candidatos en IGV<sup>30</sup>. Los datos de secuenciación de ChIP de ENCODE para H3K27ac, mapeados a hg19 y convertidos al formato de pista bigwig, se descargaron de <http://genome.ucsc.edu/ENCODE/dataMatrix/encodeChipMatrixHuman.html>.

40 Análisis de promotor/motivo. La región reguladora en cis proximal, chr2:29,445,000-29,447,100, se exploró por motivos de factor de transcripción usando FIMO<sup>48</sup> con parámetros por defecto frente a los motivos conocidos de factor de transcripción de vertebrados en la base de datos JASPAR<sup>52</sup>.

45 Amplificación rápida de extremos de ADNc en 5' (RACE en 5'). Se usaron dos técnicas independientes de RACE en 5' para mapear el sitio de ATI y el extremo 5' del transcrito de *ALK*<sup>AT7</sup>. Una técnica de RACE en 5' de tabaco-ácido-pirofosfatasa según el protocolo de fabricación (kit RACE de ARNm eucariota ExactSTART, #ES80910, Epicentre) usando los siguientes cebadores: 5'-TCATACACATACGATTTAGGTGACACTATAGAGCGGCCGCCTGCAGGAAA-3' [SEQ ID NO: 5]; inverso 5'-CAGGTCAGTGGAGGAGGTCTTGCCAGCAAAGCA-3' [SEQ ID NO: 6]. Los productos de RACE en 5' se secuenciaron en un sistema MiSeq de Illumina con un protocolo de extremos apareados de 150 pb según el protocolo de fabricación. Las lecturas de secuenciación se mapearon al genoma de referencia humano (hg19) usando BWA, se analizaron usando Tophat<sup>27</sup> y se visualizaron usando IGV<sup>30</sup>. La transcripción continua empezando en el intrón 19 de *ALK* se confirmó con un kit de RACE en 5' basado en oligonucleótidos independientes (kit de amplificación de ADNc RACE SMARTer™, #634923, Clontech) según el protocolo de fabricación usando los siguientes cebadores inversos: cebador inverso 5'-CTAATACGACTCACTATAGGGC-3' [SEQ ID NO: 7], cebador inverso 5'-ACACCTGGCCTTCATACACCTCC-3' [SEQ ID NO: 8]. Los productos de ADNc de RACE se clonaron para dar plásmidos (kit de clonación por PCR Zero Blunt® TOPO®, #K2800-20, Invitrogen) y se analizaron con secuenciación de Sanger usando procedimientos estándar. Se usaron como controles dos líneas celulares de cáncer de pulmón (H3122, H2228) con translocaciones de *EML4-ALK*, y como se esperaba, ambos controles mostraron el gen *EML4* junto al exón 20 de *ALK*.

65 Secuenciación por inmunoprecipitación de cromatina (ChIP) y ChIP-qPCR. Se aisló cromatina a partir de líneas celulares y tejido tumoral humano. El tejido tumoral humano congelado en fresco (MM-15, MM-74, ATC-28) se seccionó con un microtomo y se reticuló en paraformaldehído al 1% durante 15 min. Las muestras de tejido reticulado se extinguieron en glicina 125 mM durante 10 min, se lavaron en PBS, se resuspendieron en tampón de lisis y se homogeneizaron en una trituradora de tejido de estilo Tenbroeck y se sonicaron.<sup>32</sup> El aislamiento de cromatina a partir

de las líneas celulares H3122 y células SKMEL-524 y la inmunoprecipitación se realizó como se describió anteriormente<sup>32</sup>. La cromatina solubilizada a partir de tumores humanos y líneas celulares se sometió a inmunoprecipitación con anticuerpos contra H3K4me3 (#39159, Active Motif), H3K27ac (#Ab4729, Abcam) y ARN polimerasa II (#39097, Active Motif).

Se realizó secuenciación de ChIP en un HiSeq2500 de Illumina con lecturas simples de 51 pb. Las lecturas se alinearon con el genoma humano hg19 usando el software de alineación Bowtie dentro del canal de análisis de Illumina. Las lecturas duplicadas se eliminaron para el análisis posterior. La designación de picos se realizó usando MACS 1.4 comparando la cromatina sometida a inmunoprecipitación con cromatina de entrada.<sup>33</sup> Se realizó ChIP-qPCR en un sistema de PCR en tiempo real ViiA™ 7 (Life Technologies) usando mezcla maestra Power SYBR (#4367659, Life Technology). Los cebadores de ChIP-qPCR se diseñaron usando la herramienta primer blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast>) y se describen en las figuras 7A y 7B.

Secuenciación dirigida ultraprofunda del locus de *ALK* completo. La secuenciación dirigida del locus de *ALK* completo se realizó usando sondas de captura de hibridación personalizadas con microalineamientos de hg19 chr2:29400000-30300000 (Seq-Cap EZ de Roche/NimbleGen). Esta región abarcaba toda la huella genómica completa de *ALK* así como ~150 kb de secuencia en el sentido de 5'. Después de que el ADN genómico se fragmentara (E220, Covaris), se prepararon bibliotecas de secuencias con código de barras (New England Biolabs, Kapa Biosystems), y se realizó la captura de hibridación en grupos con código de barras usando sondas personalizadas (Nimblegen SeqCap). Se usaron 250 ng de ADN genómico para la construcción de bibliotecas a partir de 7 muestras separadas: 2 tumores de melanoma (MM-15, MM-74), 1 carcinoma de tiroides anaplásico (ATC-28), 2 líneas celulares de cáncer de pulmón (H3122, H2228) con translocaciones de *EML4-ALK*, 1 melanoma (SKMEL-28) y 1 grupo de control de 10 muestras de sangre "normales". Las bibliotecas se agruparon a concentraciones equimolares (100 ng por biblioteca) y se usaron en la reacción de captura como se describió anteriormente.<sup>34</sup> Para evitar la hibridación fuera de diana, se realizaron adiciones conocidas en un conjunto de oligonucleótidos bloqueantes complementarios a las secuencias completas de todos los adaptadores con código de barras. Las bibliotecas capturadas se secuenciaron en un HiSeq 2500 de Illumina para generar lecturas de extremos apareados de 75 pb. Los datos de secuencia se desmultiplexaron usando CASAVA, y se alinearon con el genoma humano de referencia (hg19) usando BWA.<sup>35</sup>

El realineamiento local y la recalibración de puntuación de calidad se realizaron usando el kit de herramientas de análisis de genoma (GATK) según las mejores prácticas de GATK.<sup>36</sup> Se logró una cobertura de secuencia diana única media de 1778 veces por muestra (intervalo: 1293 veces – 2188 veces). Los datos de secuencia se analizaron para identificar variantes de un solo nucleótido, inserciones/delecciones (*indels*) pequeñas y reordenamientos estructurales. Las variantes de un solo nucleótido se designaron usando muTect<sup>37</sup> y se compararon con el grupo de control negativo (muestras de sangre "normal" agrupadas). Se conservaron variantes si la frecuencia de alelos variantes en el tumor era >5 veces que en el control negativo y la frecuencia en el control negativo era <0,02. Se filtraron SNP validados en la base de datos de dbSNP. Las *indels* se designaron usando la herramienta SomaticIndelDetector en GATK<sup>36</sup> y se conservaron si el tumor albergaba >3 lecturas de soporte y la frecuencia en el control negativo era <0,02. Se usó DELLY para buscar reordenamientos estructurales.<sup>38</sup>

Secuenciación por bisulfito del locus de *ALK* completo. La captura personalizada del locus de *ALK* completo se realizó usando sondas de captura de hibridación personalizadas con microalineamientos de la huella genómica completa de *ALK* (900 kb, chr2:29400000-30300000) seguido de secuenciación por bisulfito. Después de la fragmentación (E220, Covaris) de 3 µg de ADN genómico de cada muestra (MM-15, ATC-28, H3122 y SKMEL-28), se prepararon bibliotecas con el kit KAPA HyperPrep (#KR0961, Kapa Biosystems) sin amplificación por PCR para preservar el estado de metilación. 1 µg de cada biblioteca con código de barras se agrupó a concentraciones equimolares y se capturó según el protocolo de fabricación (SeqCap EZ de Roche/NimbleGen). Después de lavar las Dynabeads M-270 (#65306, Life Technologies), el ADN de línea celular/tumor no biotinilado se disoció de las perlas de captura biotiniladas con NaOH 0,5 M. Para la conversión con bisulfito se usó ADN eluido monocatenario usando el kit EZ DNA Methylation-Gold™ (#D5005, Zymo Research) según el protocolo de fabricación, excepto para la etapa de desnaturalización a 98°C. Después de la conversión con bisulfito, se usó la polimerasa KAPA HiFi Uracil PCR (#KK280, Kapa Biosystems) para amplificar la biblioteca, se purificó la reacción con perlas Agencourt AMPure XP (A63881, Beckman Coulter) y se secuenció la biblioteca en un MiSeq de Illumina con un protocolo de extremos apareados de 150 pb según las instrucciones del fabricante. Los datos de secuencia se alinearon con hg19 y se analizaron usando BISMARKE<sup>53</sup>. Se comparó el nivel de metilación en sitios de CpG en todas las muestras; no se detectó metilación en los contextos de CHG y CHH. La metilación se calculó por primera vez como el número de lecturas de CpG metiladas frente al número de lecturas totales que cubren cada sitio de CpG (se excluyeron los sitios con <10 lecturas). Se usó una ventana deslizante para determinar el nivel de metilación medio para cada región de 250 pb (con al menos tres CpG) cerca de la región de promotor de *ALK* (chr2:29,444,000-29,452,000). La metilación diferencial se evaluó usando una prueba de Mann-Whitney.

Secuenciación de genoma completo. La secuenciación del genoma completo se realizó en el New York Genome Center (Nueva York). Brevemente, se prepararon bibliotecas de ADN genómico a partir de MM-15 y ATC-28 (no estaba disponible ADN normal coincidente) usando el kit sin PCR de Illumina. Se secuenciaron las bibliotecas en un HiSeq 2500 usando el protocolo de secuenciación de genoma completo de extremos apareados de 100 pb de Illumina. Se mapearon lecturas de secuencia usando BWA<sup>35</sup> y se procesaron usando GATK<sup>36</sup>. Se realizaron análisis de genoma

completo de mutaciones (HaplotypeCaller<sup>36</sup>), alteraciones del número de copias (FREEC)<sup>54</sup> y variaciones estructurales (CREST)<sup>55</sup>. Las mutaciones se anotaron con el Ensembl Variant Effect Predictor<sup>56</sup> y se filtraron hasta eliminar 13 polimorfismos comunes. Se visualizaron mutaciones no sinónimas junto con alteraciones del número de copias y variaciones estructurales usando Circos<sup>57</sup>.

5 Alineamiento de CGH. Las muestras de ADN genómico se etiquetaron usando un kit de marcaje genómico de alineamientos de CGH de Bioprime (#18095-011, Life Technologies) según las instrucciones del fabricante. Brevemente, 1 µg de ADN tumoral y ADN de referencia (#G1471, Promega) se etiquetaron diferencialmente con dCTP-Cy5 (#45-001-291, GE Healthcare) y dCTP-Cy3 (#45-001-290, GE Healthcare). El análisis de genoma completo de cambios del número de copias de ADN se llevó a cabo usando un microalineamiento de CGH humano SurePrint G3 de oligonucleótidos (#G4447A, Agilent) que contiene 1 millón de sondas según el protocolo de fabricación. Los portaobjetos se exploraron usando el escáner de microalineamientos G2505B de Agilent y se analizaron usando el banco de trabajo genómico de Agilent.

15 Hibridación *in situ* con fluorescencia en interfase (FISH). Se usó una sonda de separación de *ALK* disponible comercialmente (#06N38-020, Abbott) según el protocolo de fabricación. Las sondas se hibridaron en secciones de tejido de 5 µm de grosor. Se evaluó el número y la localización de las señales de hibridación en un mínimo de 100 núcleos en interfase con contornos bien delineados. Al menos el 10% de las células neoplásicas tenían que mostrar una señal de división para notificar un reordenamiento de *ALK*.

20 Transferencia de tipo Northern. Se extrajo el ARN total de tejidos o líneas celulares congeladas en fresco usando el mini kit RNeasy de Qiagen (#74104, Qiagen). Hasta 10 µg de ARN se usaron para realizar un análisis de tipo Northern basado en formaldehído según el protocolo de fabricación usando el kit RNA Ambion NorthernMax® (#AM1940, Ambion). Después de la hibridación con una sonda etiquetada con 32P, que consiste en el exón 20-29 de *ALK*, la membrana se lavó y se visualizó.

25 NanoString. Se notificaron previamente detalles del sistema de análisis nCounter (NanoString Technologies). En resumen, se construyeron dos sondas específicas de secuencia para los exones 1-19, intrón 19 y exones 20-29 de *ALK*, respectivamente. Para la normalización, se usaron cuatro genes de control (*RPS13*, *RPL27*, *RPS20*, *ACTB*). Las sondas eran complementarias a una región de 100 pares de bases del ARNm diana y se enumeran en la figura 1H. Se hibridaron 100 ng de ARN total de cada muestra, los datos sin procesar se normalizaron a la curva estándar generada a través del sistema nCounter y el valor promedio de las dos sondas en cada región diana (exones 1-19, intrón 19, exones 20-29) se imprimieron en diagramas de barras usando el software GraphPad Prism 6.0.

30 Líneas celulares. Se obtuvieron células de fibroblastos embrionarios de ratón NIH-3T3 de la 'American Type Culture Collection' (#CRL-1658, ATCC) y se mantuvieron en DMEM. La línea celular pro B murina dependiente de interleucina 3 (IL-3), Ba/F3, se obtuvo de la "Deutsche Sammlung von Mikroorganismen und Zellkulturen" (#ACC-300, DSMZ) y se cultivó en RPMI suplementado con interleucina 3 (1 ng/ml; #403-ML-010, R&D Systems). Se proporcionaron células melan-A por el Dr. Dorothy Bennett (Hospital de St. George, Universidad de Londres, Londres, Reino Unido)<sup>40</sup> y se mantuvieron en RPMI suplementado con 200 nM de 12-O-tetradecanoilforbol-13-acetato (TPA; #4174, Cell Signaling). Para la producción de retrovirus, se adquirieron y cultivaron células 293T (#631507, Clontech) en DMEM. Todos los medios de cultivo celular contenían FBS al 10%, L-glutamina (2 mM), penicilina (100 U/ml) y estreptomina (100 µg/ml). Todas las células se cultivaron a 37°C en el 5% de CO<sub>2</sub>.

35 Plásmidos. Para investigar los papeles funcionales y la activación de las rutas de señalización oncogénica, se generaron constructos de expresión de *ALK<sup>AT1</sup>*, *EML4-ALK* y *ALK<sup>F1174L</sup>*. Para el vector de *ALK<sup>AT1</sup>*, el ARN a partir de MM-15 se transcribió de manera inversa con cebadores de oligo(dT) anclados para dar ADNc (#04379012001, Roche), se amplificó por PCR con cebadores específicos de *ALK* 5'-CACCATCCCATCTCCAGTCTGCTTC-3' [SEQ ID NO: 9] y 5'-AGAGAAGTGAGTGTGCGACC-3' [SEQ ID NO: 10] y el producto de PCR se clonó para dar un vector pENTR (#K2400-20, Life Technologies). El plásmido de *ALK* de longitud completa (HsCD00079531) se adquirió del DF/HCC DNA Resource Core (<http://plasmid.med.harvard.edu>), y se sintetizó *EML4-ALKv1* en GeneArt (Life Technologies). Se realizó mutagénesis dirigida al sitio usando QuikChange (#200523, Agilent): Para *ALK<sup>AT1</sup>* sin actividad cinasa (*ALK<sup>AT1-KD</sup>*), la lisina en el sitio de unión de ATP del dominio de cinasa se mutó a metionina (p.K1150M) en *ALK<sup>AT1</sup>*, y para *ALK<sup>F1174L</sup>*, se introdujo una mutación p.F1174L en *ALK<sup>WT</sup>*. Los plásmidos se subclonaron en un vector pMIG-w<sup>41</sup> (#12282, [www.addgene.org](http://www.addgene.org)), dando como resultado constructos *MSCV-ALK<sup>variante</sup>-IRES-GFP*, los cuales se confirmaron por digestión y secuenciación. En todos los experimentos de expresión de *ALK* se usó como control un vector pMIG-w vacío con proteína verde fluorescente (GFP). Para confirmar los codones de inicio, los tres codones de inicio se mutaron de ATG a AAG. Para la co-inmunoprecipitación, se clonó *ALK<sup>AT1</sup>* en pcDNA3.1/nV5-Dest (#12290-010, Life Technologies) y MSCV N-HA FLAG-Dest (#41033, Addgene). Para la obtención de imágenes de bioluminiscencia, se usó un plásmido indicador retroviral de triple modalidad (proteína fluorescente roja (RFP) - timidina cinasa - luciferasa)<sup>42</sup>.

65 Expresión génica estable. Se produjeron retrovirus en células 293T por métodos estándar usando vector de empaquetamiento ecotrópico o amfotrópico y XtremeGene 9 (#06365809001, Roche). Se recolectó el sobrenadante que contenía virus durante 48, 64, 64 h y 72 horas después de la transfección. El sobrenadante se agrupó, se filtró a

través de una membrana de PVDF de 0,45  $\mu\text{m}$  y se usó para la transducción en presencia de polibreno (8  $\mu\text{g}$  ml<sup>-1</sup>). Las células que expresan de manera estable EGFP+ o RFP+ transducidas se clasificaron con un FACS Aria II (BD Biosciences).

5 Co-inmunoprecipitación. V5-ALK<sup>AT1</sup> y HA-ALK<sup>AT1</sup> se transfectaron de manera transitoria en células 293T y, después de 24 horas, se lisaron las células en Tris HCl 10 mM pH 7,5, Triton X-100 al 1%, NaCl 150 mM, EDTA 1 mM, DTT 1 mM, PMSF 1 mM, inhibidores de proteinasa/fosfatasa. Después de la incubación y centrifugación, se usaron 100  $\mu\text{l}$  de sobrenadante como entrada, y 300  $\mu\text{l}$  para inmunoprecipitación usando los siguientes anticuerpos: 2  $\mu\text{g}$  de anticuerpo anti-V5 (#MA1-81617, Thermo Scientific), 10  $\mu\text{l}$  de gel de afinidad anti-HA rojo EZview (#E6779-1ML, Sigma), 2  $\mu\text{g}$  de anticuerpo anti-IgG de ratón (#sc-2025, Santa Cruz). Se usaron 20  $\mu\text{l}$  de resina UltraLink de proteína A/G (#53133, Thermo Scientific) para la inmunoprecipitación. El material inmunoprecipitado se eluyó en tampón de carga SDS 4 veces para la inmunotransferencia.

15 Ensayo de cinasa *in vitro*. Se cultivaron células NIH-3T3 transducidas de manera estable en una placa de 15 cm, se lavaron en PBS y se lisaron en Tris 20 mM pH 8,0, NP-40 al 1%, NaCl 125 mM, MgCl<sub>2</sub> 2,5 mM, EDTA 1 mM con inhibidor de proteinasa/fosfatasa. Se incubaron los lisados en hielo, se centrifugaron, se aclararon previamente con 25  $\mu\text{l}$  de resina UltraLink de proteína A/G (#53133, Thermo Scientific) durante 30 min a 4°C bajo rotación y se inmunoprecipitaron con 10  $\mu\text{l}$  de AcM de conejo ALK (D5F3) XP® (#3633) y 25  $\mu\text{l}$  de resina UltraLink de proteína A/G. Después de la rotación durante 120 min a 4°C, se lavó el material inmunoprecipitado y se usó según las instrucciones del kit de ensayo de tirosina cinasa universal (#MK410, Clontech). Después de la reacción enzimática, el material inmunoprecipitado se mezcló con tampón de carga SDS 4 veces para la inmunotransferencia.

25 Inmunohistoquímica. Se realizó inmunohistoquímica en muestras de tumores FFPE de archivo usando un protocolo estándar de detección de multímero/DAB en un sistema Discovery Ultra (Roche/Ventana Medical Systems) con controles positivos y negativos apropiados. Los siguientes anticuerpos se adquirieron de Cell Signaling y se diluyeron en diluyente de anticuerpos SignalStain (#8112, Cell Signaling) como se indica: AcM de conejo frente ALK (D5F3) XP® (#3633) 1:250, AcM de conejo frente a fosfo-Akt (Ser473) (D9E) (#4060) 1:50, AcM de conejo frente a fosfo-Stat3 (Tyr705) (D3A7) (#9145) 1:400, AcM de conejo frente a proteína ribosomal fosfo-S6 (Ser235/236) (D57.2.2E) (#4858) 1:400, AcM de conejo frente a fosfo-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (D13.14.4E) (#4370) 1:400, anticuerpos frente a caspasa-3 escindida (Asp175) (#9661) 1:400. El anticuerpo anti-Ki67 se adquirió a partir de Abcam (#ab15580) y se diluyó 1:600.

35 Inmunofluorescencia. Se cultivaron células NIH-3T3 transducidas de manera estable sobre cubreobjetos, se fijaron en formaldehído al 4%, se lavaron en PBS durante 15 min a temperatura ambiente, se lavaron en PBS durante 10 min y se incubaron en disolución de bloqueo (suero de cabra al 5%, Triton X-100 al 0,1% en PBS) durante 1 hora a temperatura ambiente. Después de aspirar la disolución de bloqueo, se incubaron las células con un AcM de conejo frente a ALK (D5F3) (#3633, Cell Signaling Technology) diluido 1:1000 en tampón de bloqueo durante la noche a 4°C. Después de lavar las células 3 veces con Tween 20 al 0,05% y PBS, se incubaron las células con un anticuerpo secundario (#A-11012, Life Technologies) diluido 1:500 en tampón de bloqueo durante 2 horas a temperatura ambiente. Después de lavar en PBS, se montaron los portaobjetos con reactivo antidesvanecimiento Prolong® Gold con DAPI (#8961, Cell Signaling Technology) y se obtuvieron imágenes con un microscopio confocal Leica TCS SP5 II.

45 Inmunotransferencia. Se prepararon lisados celulares en tampón RIPA (#9806, Cell Signaling) suplementado con cóctel inhibidor de fosfatasa y proteasas Halt (#78440, Thermo Scientific). Se redisolieron cantidades iguales de proteína, tal como se midió por el ensayo de proteína BCA (#23225, Thermo Scientific) en geles de proteína Bis-Tris NuPAGE® Novex® al 4-12% (#NP0321BOX, Life Technologies) y se transfirieron electroforéticamente sobre una membrana de nitrocelulosa de 0,45  $\mu\text{m}$  (#162-0115, BioRad). Las membranas se bloquearon durante 1 hora a temperatura ambiente en tampón de bloqueo Odyssey al 50% en PBS (#927-40000, LICOR) y se incubaron durante la noche a 4°C con los anticuerpos primarios diluidos a 1:1000 en tampón de bloqueo Odyssey al 50% en PBS más Tween 20 al 0,1%. Se usaron los siguientes anticuerpos primarios: Anticuerpo anti- $\alpha$ -tubulina (#T9026-0.5ML, Sigma-Aldrich), anti-V5 (#MA1-81617, Thermo Scientific) y anti-HA3F10 (#12158167001, Roche); Cell Signaling Technology: AcM de conejo frente a fosfo-ALK (Tyr1604) (#3341), AcM de conejo frente a ALK (D5F3) (#3633), AcM de conejo frente a fosfo-Akt (Ser473) (D9E) (#4060), AcM de conejo frente a Akt (pan) (11E7) (#4685), AcM de conejo frente a fosfo-Stat3 (Tyr705) (D3A7) (#9145), AcM de conejo frente a Stat3 (79D7) (#4904), AcM de conejo frente a proteína ribosomal fosfo-S6 (Ser235/236) (D57.2.2E) (#4858), AcM de conejo frente a proteína ribosomal S6 (5G10) (#2217), AcM de conejo frente a fosfo-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (D13.14.4E) (#4370), AcM de conejo frente a p44/42 MAPK (Erk1/2) (137F5) (#4695), AcM de conejo frente a fosfo-MEK1/2 (Ser221) (166F8) (#2338), anticuerpo frente a MEK1/2 (#9122). Después de 4 lavados de 5 minutos en PBST, se incubaron membranas con anticuerpos secundarios (IRDye 800CW de cabra anti-conejo #926-32211, 1:20.000, LI-COR; IRDye 680RD de cabra anti-ratón #926-68070, 1:20.000, LI-COR) en tampón de bloqueo Odyssey al 50% en PBS más Tween 20 al 0,1% durante 45 minutos a temperatura ambiente. Después de otros 4 lavados en PBS-T y un lavado final con PBS, las membranas se exploraron con un escáner LI-COR Odyssey CLx y se ajustaron usando LI-COR Image Studio.

65 Ensayo de indicador de luciferasa. La repetición terminal larga en el intrón 19 de ALK en el sitio de ATI (LTR16B2,

- chr2:29,446,649-29,447,062; 414 pb) se amplificó usando el ADN genómico del paciente MM-15 y los cebadores 5'-GTCCTCATGGCTCAGCTTGT-3' y 5'-AGCACTACACAGGCCACTTC-3'. El producto de PCR (chr2:29,446,444-29,447,174; 731 pb) se clonó en el vector de luciferasa de luciérnaga pGL4.14 (#E6691, Promega). Para determinar la actividad promotora de LTR16B2, se transfectaron  $10^5$  células con 500  $\mu$ g de pGL4.14-LTR16B2 o vector solo; como control interno, se cotransfectaron 200  $\mu$ g de vector indicador de luciferasa pRL-TKRenilla (#E2241, Promega). La actividad luciferasa se midió usando un sistema de ensayo de luciferasa Dual-Glo (#E2920, Promega) 48 horas después de la transfección. La actividad promotora se calculó mediante la normalización de la actividad luciferasa de luciérnaga a la actividad luciferasa de renilla de control y se comparó entre pGL4.14-LTR16B2 y el vector solo.
- 5 Citometría de flujo y clasificación celular activada por fluorescencia (FACS). El análisis de citometría de flujo para ensayos de transformación *in vitro* con células Ba/F3 se realizó en un LSRFortessa (BD Biosciences). Se clasificaron células positivas para GFP o RFP usando un canal de FITC (láser azul) o PE (láser amarillo), respectivamente, en un FACS Aria II configurado con 5 láseres (BD Biosciences).
- 10 Ensayos de transformación *in vitro* y de tratamiento con fármacos. Se transdujeron células Ba/F3 de manera estable con constructos de *MSCV-ALK<sup>variante</sup>-IRES-GFP* con una multiplicidad de infección (MOI) de  $\sim 0,26$ . Basándose en los cálculos de MOI, el  $\sim 78\%$  de las células no estaban infectadas, el  $\sim 20\%$  de las células estaban infectadas con una partícula de virus y  $\sim 2\%$  de las células estaban infectadas por más de una partícula viral. Se cultivaron células Ba/F3 transducidas de manera estable en medio RPMI suplementado con IL-3 (1 ng ml<sup>-1</sup>) y se validó la tasa de transducción de  $\sim 20\%$  usando citometría de flujo para GFP, que se expresó de manera conjunta con las variantes de *ALK*. Para el ensayo de proliferación celular, las células Ba/F3 transducidas se transfirieron a medio RPMI agotado en IL-3 y se cuantificó el crecimiento celular cada 2-4 días con un ensayo de luminiscencia (#G7571, Promega). Para la curva de respuesta a la dosis de ensayos de viabilidad celular (inhibidor de *ALK*), se sembraron en placa 2000 células Ba/F3 por cuadruplicado en pocillos de placas de 96 pocillos con concentraciones crecientes de los inhibidores de *ALK* crizotinib (#C-7900, LC Laboratories), TAE684 (#CT-TAE684, ChemieTek) o ceritinib (#CT-LDK378, ChemieTek) tal como se indica. Todos los fármacos se suspendieron en DMSO. La viabilidad celular se evaluó después de 72 horas mediante un ensayo de luminiscencia CellTiter-Glo (#G7571, Promega). Los resultados se normalizaron al crecimiento de células en un medio que contiene un volumen equivalente de DMSO. La curva de inhibición se determinó con el software GraphPad Prism 6.0 usando el modelo de regresión no lineal de "log(inhibidor) frente a respuesta-pendiente variable". Para el análisis de inmunotransferencia de tipo Western, se recogieron 10 millones de células Ba/F3 después de 2 horas de tratamiento con crizotinib, se lavaron en PBS enfriado con hielo y se lisaron en tampón RIPA (#9806, Cell Signaling).
- 15 Ensayos de tratamiento con fármacos y tumorigenicidad *in vivo*. Todos los experimentos con animales se realizaron de acuerdo con un protocolo aprobado por el Comité Institucional de Cuidado y Uso de Animales de MSKCC. Se resuspendieron  $10^6$  células transfectadas de manera estable NIH-3T3 o melan-a en 50  $\mu$ l de mezcla 1:1 de PBS y Matrigel (#356237, BD Biosciences) y las células se inyectaron por vía subcutánea en los costados de ratones CB17-SCID hembra de 6-8 semanas de edad (#CB17SC-F, Taconic). Para los ensayos de crecimiento tumoral, se les inyectó a 4 ratones cada línea celular y se evaluaron 8 tumores. Los tamaños tumorales se midieron con calibres cada 2 a 7 días durante un período de hasta 100 días. Para los estudios de sensibilidad a fármacos *in vivo*, se les inyectó a 8 ratones las células NIH-3T3 transducidas de manera estable que expresan un constructo indicador de luciferasa y los plásmidos indicados. Cuando los tumores alcanzaron un tamaño promedio de 200-250 mm<sup>3</sup>, los ratones se aleatorizaron en un grupo de tratamiento y o vehículo. Los ratones se alimentaron con sonda por vía oral una vez al día con crizotinib (100 mg/kg/d) o vehículo. Se midieron 8 tumores con calibres cada de 2 a 3 días y se visualizaron las curvas de crecimiento con Prism GraphPad 6.0. Paralelamente, se monitorizó el crecimiento tumoral mediante obtención de imágenes de bioluminiscencia de ratones anestesiados mediante inyección por vía retroorbital de d-luciferina (150 mg por kg de peso corporal) y obtención de imágenes con la máquina IVIS Spectrum Xenogen (Caliper Life Sciences). Después de sacrificar los ratones, se explantaron los tumores o bien lisados en tampón RIPA (#9806, Cell Signaling Technology) o bien fijados durante la noche en paraformaldehído al 4%, se lavaron, se incrustaron en parafina y se cortaron para la tinción con hematoxilina y eosina (HE) o inmunohistoquímica.
- 20 Estadística. Todas las comparaciones estadísticas entre dos grupos se realizaron mediante el software GraphPad Prism 6.0 usando una prueba de la t desapareada de dos colas.

## 55 6.2 RESULTADOS Y DISCUSIÓN

- Para identificar mecanismos novedosos activantes oncogénicos, se realizaron análisis de transcriptoma (secuenciación de ARN) de melanoma metastásico y carcinoma de tiroides. Se usó un algoritmo<sup>9</sup> para investigar la expresión diferencial de exones en tirosina cinasas receptoras (RTK). El análisis se centró en transcritos con una alta expresión del dominio de cinasa. Se identificó un transcrito de *ALK* novedoso en dos melanomas (MM-15, MM-74) y en un carcinoma de tiroides anaplásico (ATC-28). El transcrito novedoso contenía los exones 20-29 de *ALK* precedidos por  $\sim 400$  pares de bases de intrón 19, pero no los exones 1-19 (figuras 1A y 5A). Este patrón de expresión era distinto de translocaciones de *ALK* y *ALK* de tipo silvestre (ALT<sup>WT</sup>). El ALT<sup>WT</sup> muestra expresión de todos los exones, pero no expresión de intrones (figuras 1A, 5B), y se encuentra comúnmente en neuroblastoma en asociación con mutaciones activantes.<sup>10-13</sup> Se observan translocaciones de *ALK* en diversos tipos de cáncer, incluyendo linfoma, sarcoma y cáncer

de pulmón y normalmente ocurren en el intrón 19.<sup>14-16</sup> Debido a los sitios de corte y empalme preservados, las translocaciones de *ALK* normalmente abarcan los exones 20-29 de *ALK* con poca expresión intrónica (figuras 1A y 5C). La sobreexpresión<sup>10, 58-60</sup> y amplificación<sup>10,60</sup> de *ALKWT*, mutaciones de *ALK* activantes<sup>10,11-13</sup> y translocaciones de *ALK*<sup>15,16</sup> son iniciadores oncogénicos bien establecidos en diversos tipos de cáncer, incluyendo tumores melanocíticos<sup>62</sup>.

Para evaluar si el transcrito novedoso de *ALK* surge de la iniciación transcripcional alternativa (ATI), se realizó una amplificación rápida en 5' de extremos de ADNc (RACE en 5'). El sitio de ATI se mapeó a una región de 25 pb en el intrón 19 y la presencia del transcrito novedoso se confirmó y se denominó *ALK<sup>ATI</sup>* por transferencia de tipo Northern (figuras 1B y 6A-6G). La secuenciación ChIP y ChIPqPCR mostraron que solo tumores que expresan *ALK<sup>ATI</sup>*, pero no los controles, tuvieron un enriquecimiento significativo de las marcas de cromatina histona H3 lisina 4 trimetilada (H3K4me3) y ARN polimerasa II en el sitio de ATI, que son características de los promotores activos<sup>17,18</sup> (figuras 1C-1E y 7A-7D). En conjunto, estos datos sugieren que *ALK<sup>ATI</sup>* se origina a partir de un sitio de ATI genuino recientemente establecido asociado con alteraciones de cromatina características.

Para determinar la prevalencia de expresión de *ALK<sup>ATI</sup>*, se examinaron más de 5000 muestras de 15 tipos diferentes de cáncer en el conjunto de datos de secuenciación de ARN de TCGA. Se expresó *ALK<sup>ATI</sup>* en ~11% de melanoma (38 de 334 tumores) y esporádicamente en otros tipos de cáncer, incluyendo adenocarcinoma de pulmón, carcinoma de células escamosas pulmonares, carcinoma de células renales claras y carcinoma de mama (figuras 8A-8E). No se encontró expresión de *ALK<sup>ATI</sup>* en más de 1600 muestras a partir de 43 tejidos normales diferentes en el conjunto de datos de secuenciación de ARN de expresión tisular de genotipo (GTEx)<sup>1</sup>, lo que indica que *ALK<sup>ATI</sup>* se expresa principalmente en cánceres, particularmente melanoma.

Para cuantificar de manera precisa la expresión de *ALK<sup>ATI</sup>* en especímenes clínicos, se desarrolló un ensayo nCounter de NanoString con conjuntos de sondas en los exones 1-19, intrón 19 y exones 20-29 de *ALK* (figura 1H). Este ensayo era capaz de distinguir *ALK<sup>ATI</sup>*, *ALK<sup>WT</sup>* y *ALK* translocada e identificó tumores que expresan *ALK<sup>ATI</sup>* adicionales derivados tanto de especímenes clínicos congelados en fresco como fijados en formalina, incrustados en parafina (FFPE) (figura IF).

Para determinar si las aberraciones genómicas somáticas en el locus de *ALK* contribuyen al establecimiento del sitio de ATI *de novo*, se realizaron análisis genéticos integrales incluyendo hibridación *in situ* con fluorescencia en interfase (FISH), alineamiento de genoma completo CGH, secuenciación del genoma completo y secuenciación ultraprofunda del locus de *ALK* completo, cuyos resultados se muestran en las figuras 9A-9D y 10A-10F. Por ejemplo, dado que la activación transcripcional puede surgir debido a reordenamientos genómicos,<sup>20</sup> se realizó FISH en interfase usando sondas que reconocen los extremos 5' y 3' del locus de *ALK*, pero no se encontraron reordenamientos de *ALK* (figuras 9A y 9B). Para examinar el locus de *ALK* en busca de pequeñas deleciones descritas anteriormente<sup>21,22</sup> y duplicaciones en tándem,<sup>16</sup> se usó alineamiento de genoma completo CGH, pero no se descubrieron alteraciones genéticas (figuras 9C y 9D). Finalmente, para investigar si *ALK<sup>ATI</sup>* podría surgir a través de alteraciones genéticas que crean un sitio de ATI *de novo* similar a las mutaciones promotoras de TERT recientemente descritas en melanoma<sup>23,24</sup>, se realizó secuenciación dirigida ultraprofunda del locus de *ALK* completo. No se identificaron variaciones de un solo nucleótido (SNV), inserciones o deleciones recurrentes en los exones e intrones de *ALK* (figuras 10A-10D). En resumen, no se encontraron aberraciones genómicas que pudieran explicar la expresión *de novo* de *ALK<sup>ATI</sup>*. Razonando que las aberraciones genómicas locales normalmente actúan en cis y solo afectan a la expresión del alelo afectado,<sup>25,43</sup> se analizaron los SNV en los datos de secuenciación de ADN, ARN y ChIP. En comparación con el ADN genómico, se encontraron frecuencias de SNV alélicas similares en los datos de secuenciación de ARN y ChIP, lo que indica que ambos alelos de *ALK* se transcriben y decoran activamente con H3K4me3 (figura 1G). La expresión de *ALK<sup>ATI</sup>* bi-alélica indica que la activación transcripcional de *ALK<sup>ATI</sup>* es independiente de las aberraciones genéticas. Sin embargo, aberraciones que afectan a elementos de actuación en trans, tales como factores de transcripción o modificadores de cromatina, pueden contribuir a la expresión de *ALK<sup>ATI</sup>*.

La región de ATI de *ALK* contiene elementos transponibles, incluyendo una repetición terminal larga (LTR) en el intrón 19 y un elemento largo intercalado (LINE) en el intrón 18, ambos de los cuales pueden regular la transcripción<sup>44, 45</sup> (figura 11A). Para evaluar si la metilación de CpG de estos elementos podría contribuir a la expresión de *ALK<sup>ATI</sup>*, se realizó secuenciación por bisulfito del locus de *ALK* completo. En comparación con los controles, las muestras que expresan *ALK<sup>ATI</sup>* mostraron menor metilación de CpG en regiones que flanquean el sitio de ATI, incluyendo el LINE (figuras 11B-D). El LTR contenía solo algunos CpG con bajos niveles de metilación en todas las muestras. Los datos de ENCODE<sup>46</sup> revelaron una agrupación de hipersensibilidad de ADNasa I y enriquecimiento de H3K4me1 en la región de ATI. Independientemente de la expresión de *ALK<sup>ATI</sup>*, se encontró enriquecimiento de H3K27ac (un marca de histona característica de promotores y potenciadores activos<sup>47</sup>) en todas las muestras de melanoma analizadas, pero no en las líneas celulares de cáncer de pulmón o en las 17 líneas celulares que no son de melanoma de ENCODE (figuras 11E y 11F). Mediante la integración de los datos de ChIP, hipersensibilidad de ADNasa I y RACE en 5', se definió la región regulatoria en cis proximal como chr2:29,445,000-29,447,100 y se determinaron bioinformáticamente los posibles motivos de unión a factor de transcripción.<sup>48</sup> (Figura 11H). Para someter a prueba si el LTR podía funcionar como promotor, se usó un ensayo de indicador de luciferasa, y se encontró que, a diferencia de las líneas celulares de cáncer de pulmón, las líneas celulares de melanoma mostraron actividad de luciferasa baja pero constante (figura 11G). Tomados en conjunto, estos datos sugieren que la marca H3K27ac en el sitio de ATI podría sensibilizar a

melanomas para la expresión de  $ALK^{AT1}$ , lo que es consecuente con la mayor frecuencia de  $ALK^{AT1}$  en melanomas en comparación con otros tipos de cáncer. Sin embargo, estos datos también indican que la activación completa de la expresión de  $ALK^{AT1}$  requiere factores activantes en trans adicionales.

5 Se usó RACE en 5' (amplificación rápida en 5' de extremos de ADNc) acoplado con secuenciación de última generación para determinar el transcrito de  $ALK^{AT1}$ . El sitio de AT1 se mapeó a una región de 25 pares de bases (pb) en el intrón 19, aproximadamente 400 pb en el sentido de 5' del exón 20 de  $ALK$  (figuras 1B y 6A-C). El transcrito novedoso de  $ALK$  tiene un tamaño de aproximadamente 2500 pb (figura 7A). El transcrito de  $ALK^{AT1}$  tiene tres codones de inicio en marco (ATG) predichos dando como resultado proteínas con pesos moleculares (PM) de 61,1 kDa (552 aminoácidos), 60,8 kDa (550 aminoácidos) y 58,7 kDa (532 aminoácidos) (figuras 6D y 6E). Las tres proteínas mantienen el dominio de tirosina cinasa intracelular, pero carecen de los dominios extracelular y transmembrana de  $ALK^{WT}$  (figura 2A).

15 Las inmunotransferencias de dos líneas celulares de neuroblastoma que expresan  $ALK^{WT}$  y dos líneas celulares de cáncer de pulmón que expresa la variante EML4- $ALK$  mostraron proteínas con el PM esperado de ~220 kDa para  $ALK^{WT}$  (y un producto de escisión más pequeño que carece de parte de la región extracelular<sup>49</sup>) y de ~120 kDa y -90 kDa para dos variantes EML4- $ALK$ . Tumores que expresan  $ALK^{AT1}$  revelaron una banda doble a -60 kDa, sugiriendo que  $ALK^{AT1}$  se traduce a partir de más de un codón de inicio (figura 2B). Para confirmar experimentalmente los codones de inicio predichos, los tres codones de inicio se mutaron individualmente o en combinación y se expresaron en células 293T. Las inmunotransferencias revelaron que cada uno de los  $ALK^{AT1}$  mutantes perdieron la banda de proteína correspondiente, lo que indica que los tres codones de inicio en  $ALK^{AT1}$  son funcionales y dan lugar a tres proteínas distintas (figura 2C).

25 Las proteínas de  $ALK^{AT1}$  se fosforilaron en tumores con expresión de  $ALK^{AT1}$  endógena y en células con expresión de  $ALK^{AT1}$  exógena (figuras 2B y 2C) indicando que  $ALK^{AT1}$  está activo. Esto se confirmó por un ensayo de cinasa *in vitro* (figura 12A). Un  $ALK^{AT1}$  sin actividad cinasa ( $ALK^{AT1-KD}$ ), en el que una lisina en el sitio de unión a ATP del dominio de cinasa se reemplazó por una metionina<sup>15</sup>, no estaba fosforilada ni activa. Razonando que  $ALK^{AT1}$  puede autoactivarse formando homodímeros como otras RTK<sup>50</sup>, la capacidad de autointeracción se sometió a prueba usando co-inmunoprecipitación con proteínas  $ALK^{AT1}$  etiquetadas con V5 o HA. El V5- $ALK^{AT1}$  se co-inmunoprecipitó fácilmente con el HA- $ALK^{AT1}$  y viceversa, lo que indica que  $ALK^{AT1}$  interactúa consigo mismo dando como resultado autofosforilación y activación de cinasa (figura 2D). Usando inmunofluorescencia, se detectó  $ALK^{AT1}$  tanto en el núcleo como en el citoplasma, mientras que se encontraron  $ALK^{F1174L}$  y EML4- $ALK$  en el citoplasma o la membrana celular (figura 2E). La inmunohistoquímica de  $ALK$  en muestras clínicas confirmó la localización nuclear y citoplasmática de  $ALK^{AT1}$ , sugiriendo que la tinción de  $ALK$  nuclear en inmunohistoquímica podría servir como biomarcador clínico para identificar pacientes con tumores que expresan  $ALK^{AT1}$  (figuras 2F y 12B).

35 Basándose en el análisis del conjunto de datos secuenciación de ARN de GTEx,  $ALK^{AT1}$  no se expresa en tejidos normales. Para establecer las consecuencias funcionales de la expresión de  $ALK^{AT1}$ , se transdujeron de manera estable células Ba/F3, NIH-3T3 y melan-a con  $ALK^{AT1}$ , controles negativos ( $ALK^{AT1}$ , vector vacío) y controles positivos (las variantes de  $ALK$  oncogénicas  $ALK^{F1174L}$ , EML4- $ALK$  y  $ALK^{WT}$ , que anteriormente se mostró que son suficientes para impulsar la oncogénesis a altos niveles de expresión endógena<sup>6,12,58-61</sup>). En las células Ba/F3, la expresión de  $ALK^{AT1}$  condujo a un crecimiento celular independiente de IL-3, al igual que los controles positivos, pero no los controles negativos (figura 3A). En células Ba/F3 transformadas con  $ALK$  que crecen en medios sin IL-3, se confirmó que  $ALK^{AT1}$  se expresaba a un nivel similar en comparación con los tumores humanos y que todas las isoformas de  $ALK$  estaban fosforiladas y, por lo tanto, activas (figura 3B). La dependencia de  $ALK$  del crecimiento independiente de IL-3 se reflejó en la selección de células Ba/F3 que expresan proteína verde fluorescente (GFP), la cual se expresó de manera conjunta a partir de los vectores de expresión de  $ALK$  (figura 12C). Consecuente con los datos *in vitro*, las células melan-a y NIH-3T3 que expresan  $ALK^{AT1}$  indujeron eficientemente el crecimiento tumoral en ratones SCID (figuras 3C y 12D-12F).

50 En resumen, todas las células que expresan variante de  $ALK$  ( $ALK^{AT1}$ ,  $ALK^{F1174L}$ , EML4- $ALK$ ,  $ALK^{WT}$ ) eran capaces de establecer proliferación y tumorigénesis independientes de factores de crecimiento, con tasas de crecimiento similares una vez se establecieron los tumores. De manera importante, la capacidad oncogénica observada de  $ALK^{WT}$  es consecuente con informes previos de que la alta expresión endógena o amplificación de  $ALK^{WT}$  impulsa la oncogénesis y confiere sensibilidad a los inhibidores de  $ALK$  en neuroblastomas<sup>10,12, 58-61</sup>. Para explorar adicionalmente el papel patógeno de niveles de expresión de  $ALK^{AT1}$ , se transdujeron células NIH-3T3 de manera estable con un título o bien bajo o bien alto de  $ALK^{AT1}$  dando como resultado células que expresan  $ALK^{AT1}$  a niveles o bien bajos o bien altos. Se encontró que un aumento adicional en los niveles de expresión de  $ALK^{AT1}$  no aceleró la formación de injertos tumorales y el crecimiento tumoral, indicando que  $ALK^{AT1}$  puede impulsar la tumorigénesis una vez se alcanza un umbral de expresión (figuras 12G-12I).

65 Para examinar las respuestas terapéuticas a la inhibición de  $ALK$  farmacológica, se trataron células Ba/F3 que expresan de manera estable diversas isoformas de  $ALK$  con tres inhibidores de  $ALK$  diferentes (crizotinib, ceritinib, TAE-684). Todos los inhibidores de  $ALK$  inhibieron eficazmente el crecimiento independiente de IL-3 de células Ba/F3 transformadas con  $ALK$ , mientras que no tuvieron efecto sobre el crecimiento en presencia de IL-3 (figuras 4A y 13A-13B). Crizotinib inhibió la fosforilación de  $ALK^{AT1}$  y la señalización en el sentido de 3' de manera dependiente de la

concentración, corroborando además que  $ALK^{AT1}$  se activa mediante autofosforilación (figuras 4B y 13E-13E). El tratamiento con crizotinib también indujo regresión de tumores de NIH-3T3 impulsados por  $ALK^{AT1}$  *in vivo*, y la inmunohistoquímica de tumores explantados confirmó proliferación celular reducida, apoptosis aumentada e inhibición de varias rutas de señalización oncogénicas (figuras 4C-4E y 14A-14F).

Basándose en estos datos preclínicos alentadores, se identificó un paciente con melanoma metastásico con expresión de  $ALK^{AT1}$  (figuras 4F y 4G). El paciente había progresado previamente en la combinación de inmunoterapia con ipilimumab e nivolumab en un ensayo clínico, seguido de radiación paliativa y luego quimioterapia con dacarbazina. El uso compasivo de crizotinib dio como resultado una marcada mejoría sintomática y una contracción tumoral en el plazo de 6 semanas de terapia (figura 4H). Este caso de paciente proporciona evidencia adicional de que  $ALK^{AT1}$  confiere ganancia de función en los pacientes y es susceptible de selección como diana farmacológica, lo que justifica una investigación clínica adicional.

En conjunto, se identificó un transcrito de  $ALK$  novedoso,  $ALK^{AT1}$ , que surge independientemente de aberraciones genómicas en el locus de  $ALK$  mediante iniciación transcripcional alternativa.  $ALK^{AT1}$  codifica proteínas de  $ALK$  acortadas que son capaces de impulsar la oncogénesis *in vitro* e *in vivo*. Los tumores impulsados por  $ALK^{AT1}$  son sensibles a inhibidores de  $ALK$ , lo que sugiere que los pacientes que albergan tales tumores podrían beneficiarse potencialmente de la terapia con inhibidores de  $ALK$ , pero no pueden identificarse usando ensayos genómicos clínicos actuales, en particular los basados en secuenciación de ADN. De manera importante, se descubrió iniciación transcripcional alternativa como mecanismo novedoso para la activación oncogénica, además de mecanismos genéticos bien establecidos tales como mutaciones, translocaciones o amplificaciones. Otros oncogenes pueden activarse a través de mecanismos similares en otros tumores malignos humanos y su identificación puede proporcionar nuevos conocimientos sobre la oncogénesis y oportunidades de intervención terapéutica.

## 7. Bibliografía

1. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45, 580-5 (2013).
2. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214-8 (2013).
3. Lawrence, M.S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumor types. *Nature* 505, 495-501(2014).
4. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330-7 (2012).
5. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489, 519-25 (2012).
6. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* 339, 1546-58 (2013).
7. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150, 1107-20 (2012).
8. Weinstein, J.N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45, 1113-20 (2013).
9. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* 22, 2008-17 (2012).
10. Mosse, Y.P. *et al.* Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* 455, 930-5 (2008).
11. Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* 455, 971-4 (2008).
12. Janoueix-Lerosey, I. *et al.* Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* 455, 967-70 (2008).
13. George, R.E. *et al.* Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature* 455, 975-8 (2008).
14. Morris, S.W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 263, 1281-4 (1994).
15. Soda, M. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448, 561-6 (2007).
16. Lipson, D. *et al.* Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat Med*

- 18, 382-4 (2012).
17. Kouzarides, T. Chromatin modifications and their function. *Cell* 128, 693-705 (2007).
- 5 18. Ram, O. *et al.* Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* 147, 1628-39 (2011).
19. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40, 897-903 (2008).
- 10 20. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nature reviews. Cancer* 7, 233-45 (2007).
- 15 21. Cazes, A. *et al.* Characterization of rearrangements involving the ALK gene reveals a novel truncated form associated with tumor aggressiveness in neuroblastoma. *Cancer Res* 73, 195-204 (2013).
22. Okubo, J. *et al.* Aberrant activation of ALK kinase by a novel truncated form ALK protein in neuroblastoma. *Oncogene* 31, 4667-76 (2012).
- 20 23. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* 339, 959-61 (2013).
24. Huang, F.W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* 339, 957-9 (2013).
- 25 25. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506-11 (2013).
26. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-9 (2009).
27. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-11 (2009).
- 30 28. Sboner, A. *et al.* FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNAsequencing data. *Genome Biol* 11, R104 (2010).
- 35 29. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297-303 (2010).
30. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* 29, 24-6 (2011).
- 40 31. Carbon, S. *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* 25, 288-9 (2009).
32. Chi, P. *et al.* ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumors. *Nature* 467, 849-53 (2010).
- 45 33. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137 (2008).
34. Won, H.H., Scott, S.N., Brannon, A.R., Shah, R.H. & Berger, M.F. Detecting somatic genetic alterations in tumor specimens by exon capture and massively parallel sequencing. *J Vis Exp*, e50710 (2013).
- 50 35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-60 (2009).
36. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43, 491-8 (2011).
- 55 37. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31, 213-9 (2013).
38. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333-i339 (2012).
- 60 39. Geiss, G.K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 26, 317-25 (2008).
- 65 40. Bennett, D.C., Cooper, P.J. & Hart, I.R. A line of non-tumorigenic mouse melanocytes, syngeneic with the B16 melanoma and requiring a tumor promoter for growth. *Int J Cancer* 39, 414-8 (1987).

41. Refaeli, Y., Van Parijs, L., Alexander, S.I. & Abbas, A.K. Interferon gamma is required for activation-induced death of T lymphocytes. *J Exp Med* 196, 999-1005 (2002).
- 5 42. Ponomarev, V. *et al.* A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur J Nucl Med Mol Imaging* 31, 740-51 (2004).
43. Northcott, P.A. *et al.* Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature* 511, 428-434 (2014).
- 10 44. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32, D493-496 (2004).
45. Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* 45, 836-841 (2013).
- 15 46. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74 (2012).
47. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49, 825-837 (2013).
- 20 48. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017-1018 (2011).
49. Moog-Lutz, C. *et al.* Activation and inhibition of anaplastic lymphoma kinase receptor tyrosine kinase by monoclonal antibodies and absence of agonist activity of pleiotrophin. *J Biol Chem* 280, 26039-26048 (2005).
- 25 50. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* 141, 1117-1134 (2010).
51. Reis, P. P. *et al.* mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol* 11, 46 (2011).
- 30 52. Mathelier, A. *et al.* JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 42, D142-147 (2014).
53. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571-1572 (2011).
- 35 54. Boeva, V. *et al.* Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics* 27, 268-269 (2011).
- 40 55. Wang, J. *et al.* CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat Methods* 8, 652-654 (2011).
56. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26, 2069-2070 (2010).
- 45 57. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* 19, 1639-1645 (2009).
58. Schulte, J. H. *et al.* High ALK receptor tyrosine kinase expression supersedes ALK mutation as a determining factor of an unfavorable phenotype in primary neuroblastoma. *Clin Cancer Res* 17, 5082-5092 (2011).
- 50 59. Passoni, L. *et al.* Mutation-independent anaplastic lymphoma kinase overexpression in poor prognosis neuroblastoma patients. *Cancer Res* 69, 7338-7346 (2009).
- 55 60. Bresler, S. C. *et al.* Differential inhibitor sensitivity of anaplastic lymphoma kinase variants found in neuroblastoma. *Sci Transl Med* 3, 108-114 (2011).
61. Montavon, G. *et al.* Wild-type ALK and activating ALK-R1275Q and *ALK<sup>F1174L</sup>* mutations upregulate Myc and initiate tumor formation in murine neural crest progenitor cells. *Oncotarget* 5, 4452-4466 (2014).
- 60 62. Wiesner, T. *et al.* Kinase fusions are frequent in Spitz tumors and spitzoid melanomas. *Nat Commun* 5, 3116 (2014).

## REIVINDICACIONES

1. Molécula de ácido nucleico aislado que comprende o que consiste esencialmente en los exones 20-29 y una parte, pero no todo, del intrón 19 de un gen de cinasa de linfoma anaplásico (*ALK*), en la que la molécula de ácido nucleico no comprende los exones 1-19 del gen *ALK*, o bien individualmente o bien en cualquier combinación, en la que la molécula de ácido nucleico es de aproximadamente 2300-2600 pb de longitud.
2. Molécula de ácido nucleico aislado según la reivindicación 1, en la que el gen *ALK* es un gen *ALK* humano.
3. Molécula de ácido nucleico aislado según la reivindicación 1 o 2, en la que la parte del intrón 19 del gen *ALK* es de aproximadamente 400 pb de longitud ubicada en el sentido de 5' del exón 20 del gen *ALK*.
4. Molécula de ácido nucleico aislado según una cualquiera de las reivindicaciones 1-3, en la que la molécula de ácido nucleico es de aproximadamente 2500 pb de longitud, o aproximadamente 2513 pb de longitud.
5. Molécula de ácido nucleico aislado según una cualquiera de las reivindicaciones 1-4, en la que la molécula de ácido nucleico es un transcrito de ARN mensajero (ARNm).
6. Molécula de ácido nucleico aislado según una cualquiera de las reivindicaciones 1-5, que comprende o que consiste esencialmente en la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma.
7. Molécula de ADN complementario (ADNc) aislado que corresponde al ácido nucleico aislado según una cualquiera de las reivindicaciones 1-6, en la que la molécula de ADNc comprende o consiste esencialmente en los exones 20-29 de un gen de cinasa de linfoma anaplásico (*ALK*), en la que la molécula de ADNc no comprende los exones 1-19 del gen *ALK*, o bien individualmente o bien en cualquier combinación, en la que la molécula de ácido nucleico es de aproximadamente 2300-2600 pb de longitud.
8. Molécula de ADNc aislado según la reivindicación 7, que comprende los ácidos nucleicos 405-2063, 411-2063 o 465-2063 de la secuencia de nucleótidos expuesta en SEQ ID NO: 11 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma.
9. Molécula de ácido nucleico aislado según una cualquiera de las reivindicaciones 1-8, que comprende una secuencia de nucleótidos que codifica la secuencia de aminoácidos expuesta en SEQ ID NO: 12 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma.
10. Molécula de ácido nucleico aislado según una cualquiera de las reivindicaciones 1-8, que comprende una secuencia de nucleótidos que codifica los aminoácidos 3 a 552, o aminoácidos 21 a 552 de la secuencia de aminoácidos expuesta en SEQ ID NO: 12 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma.
11. Vector que comprende (a) la molécula de ácido nucleico aislado según una cualquiera de las reivindicaciones 1-6, 9 y 10, o (b) la molécula de ADNc aislado según la reivindicación 7 o la reivindicación 8.
12. Célula huésped que comprende el vector según la reivindicación 11.
13. Polipéptido aislado codificado por (a) la molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-6, 9 y 10 o (b) la molécula de ADNc según la reivindicación 7 o la reivindicación 8.
14. Polipéptido aislado según la reivindicación 13, que comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 12 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma, en el que el polipéptido no comprende (a) la secuencia de aminoácidos expuesta en SEQ ID NO: 3 o (b) la secuencia de aminoácidos expuesta en SEQ ID NO: 4.
15. Polipéptido aislado según la reivindicación 14, que comprende los aminoácidos 3 a 552 o los aminoácidos 21 a 552 de la secuencia de aminoácidos expuesta en SEQ ID NO: 12 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma, en el que el polipéptido no comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 3.
16. Polipéptido aislado según la reivindicación 14, que comprende los aminoácidos 3 a 552 o los aminoácidos 21 a 552 de la secuencia de aminoácidos expuesta en SEQ ID NO: 12 o una secuencia que es al menos aproximadamente un 95% homóloga con respecto a la misma, en el que el polipéptido no comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 4.
17. Anticuerpo que se une al polipéptido según una cualquiera de las reivindicaciones 14-16, en el que el anticuerpo no se une a un *ALK* de tipo silvestre.

18. Anticuerpo según la reivindicación 17, en el que el ALK de tipo silvestre es un ALK humano de tipo silvestre.
- 5 19. Anticuerpo según la reivindicación 18, en el que el ALK humano de tipo silvestre comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 1.
- 10 20. Método *in vitro* de diagnóstico de una célula en una muestra recogida de un sujeto como una célula cancerosa que comprende detectar la presencia de una isoforma truncada de cinasa de linfoma anaplásico (TALK) en la célula, en el que la presencia de la TALK indica que la célula es una célula cancerosa, en el que detectar comprende determinar la presencia de la molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-6, 9 y 10 o la molécula de ADNc según la reivindicación 7 o la reivindicación 8, y/o el polipéptido de TALK según una cualquiera de las reivindicaciones 14-16.
- 15 21. Método según la reivindicación 20, en el que el cáncer se selecciona del grupo que consiste en melanoma, carcinoma de tiroides, adenocarcinoma de pulmón, carcinoma de células escamosas pulmonares, carcinoma de células claras renales y cáncer de mama.
- 20 22. Método según la reivindicación 20 o la reivindicación 21, en el que el cáncer es melanoma o carcinoma de tiroides anaplásico.

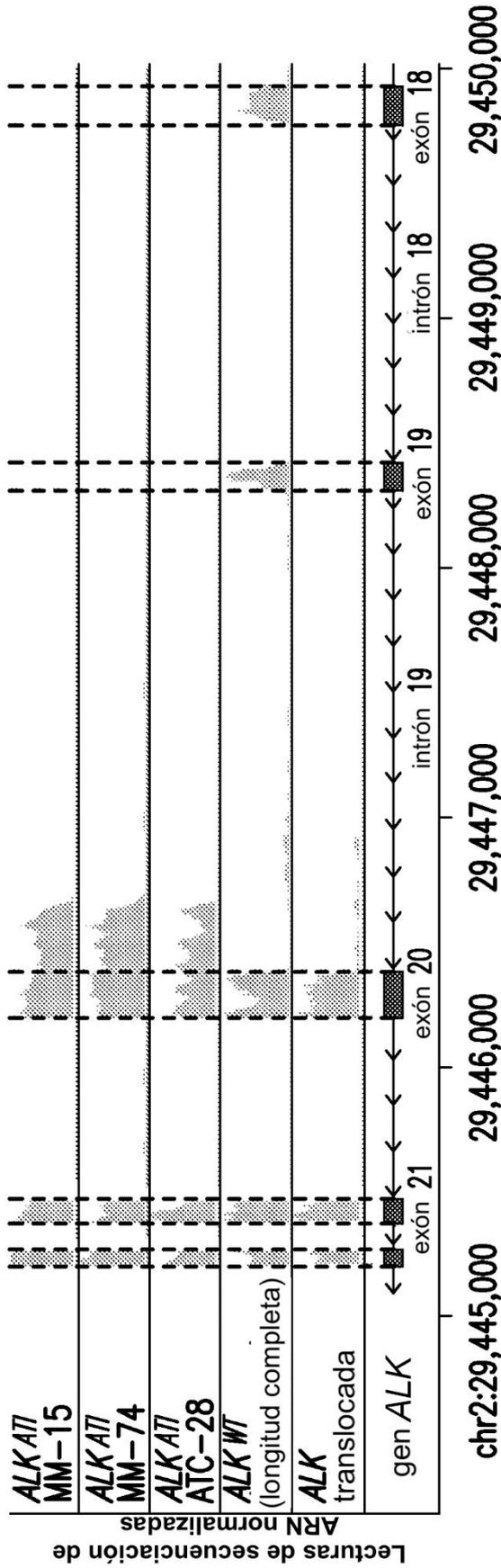


FIG. 1A

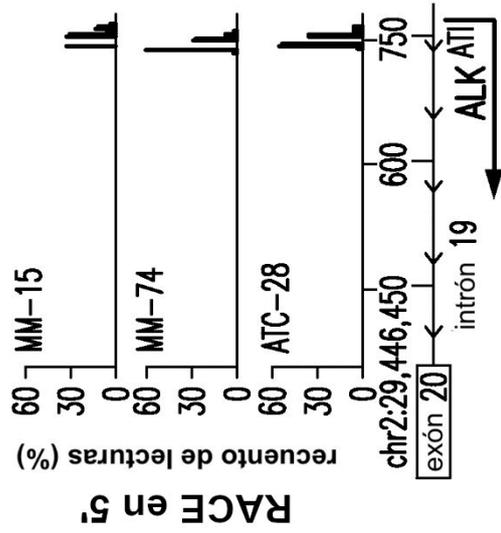


FIG. 1B

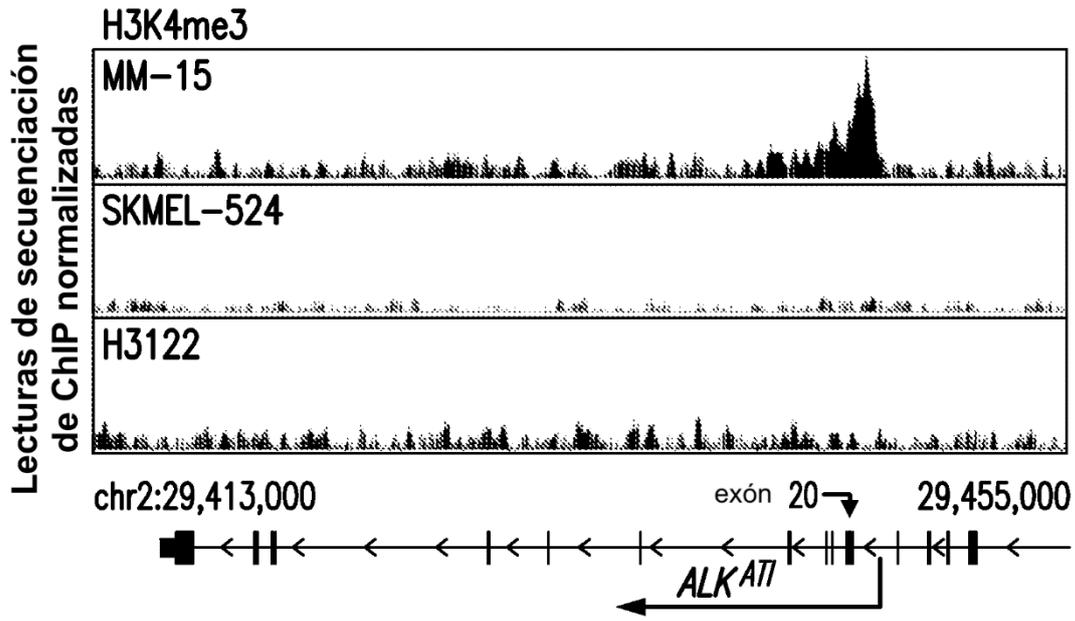


FIG. 1C

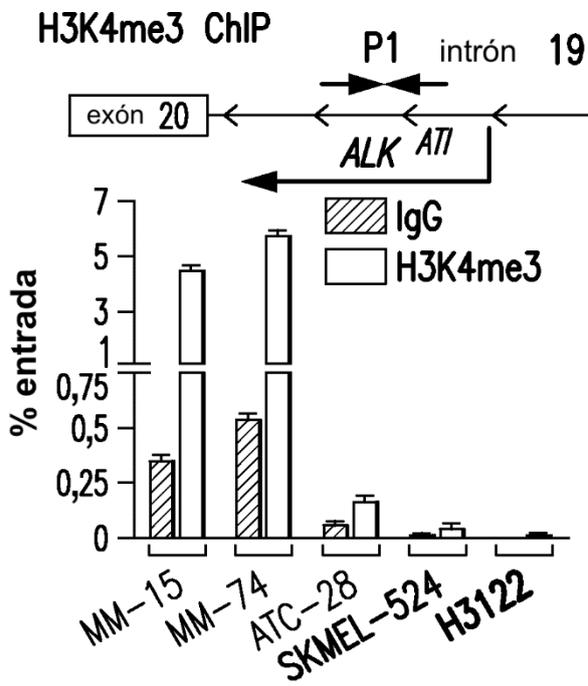


FIG. 1D

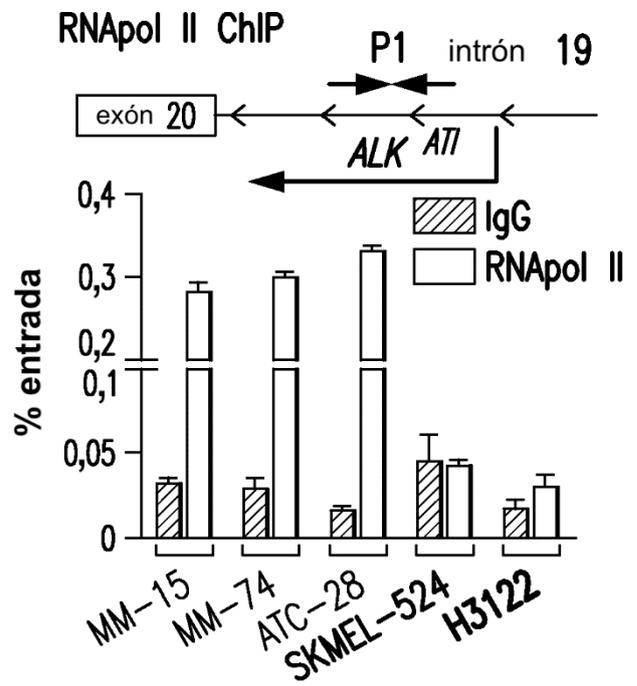


FIG. 1E

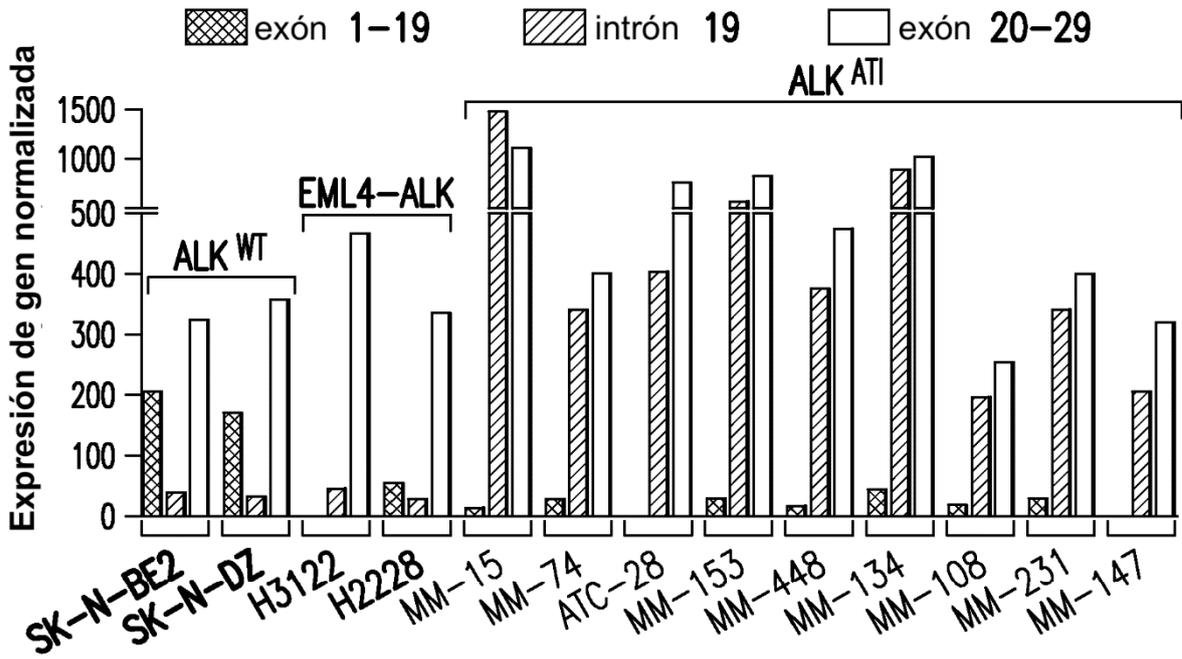


FIG. 1F

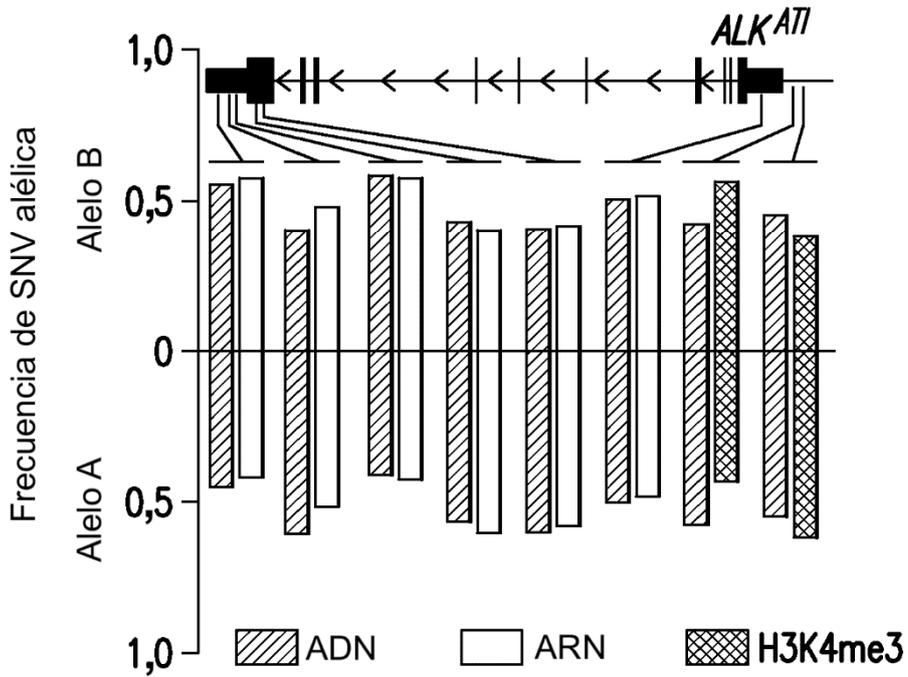


FIG. 1G

ID	Nombre	Acceso	Región diana	Secuencia diana
1	ALK_ex10-19_v1	NM_004304.4	2844-2943	GCCTGGACTGCTACCTCACCATTAGCGGAGAGGACAAAGATCCTGCAGAATACAGC ACCCAAATCAAGAAACCTGTTTGAGAGAAACCCAAACAAAGGAGCT
2	ALK_ex10-19_v2	NM_004304.3	3496-3595	TAACTCCTCGTTCAGGGCTAAACGGCAATCCGGAGCCCGCAGGTGGTGGAGG TGGCTGGAATGATAACACTTCCTTGCTCTGGCCGGAAAAATCTTTG
3	ALK_in19-UTR_v1	ALK_001.1	261-360	TTTATTAGTATTTCTAAGTATGATGGAAGGTTCAAGCTCAGGGGAGGATATGGA GATCCAGGGAGGCTTCCTGTAGGAAGTGGCCTGTGTAGTGTCTC
4	ALK_in19-UTR_v2	ALK_002.1	10-109	CTTCAAGGGCCAGGCTGCCAGGCCATGTTGCAGCTGACCCACCCCTGCAGTGT ACGGCCGGAAGCACCAGGAGCTGCAAGCCATGCAGATGGAGCTGCA
5	ALK_ex20-29_v1	NM_004304.4	4846-4945	AGAGGCCTTCATGGAAGGAATATTCACCTTCTAAACAGACACATGGTCTTTGGAG TGCTGCTATGGGAAATCTTTTCTCTTGGATATATGCCATACCC
6	ALK_ex20-29_v2	NM_004304.4	5825-5924	CACACTCACTTCTCTTCCCTTGGATCCCTAAGACCCGTGGAGGAGAGAGGCAAT GGCTCCTTCACAAACCAGAGACCAAAATGTCACGTTTTGTTTTGTG
7	ACTB	NM_001101.2	1011-1110	TGCAGAAGGAGATCACTGCCCTGGCACCAGCCACAAATGAAGATCAAGATCATTTGC TCCTCCTGAGCGCAAGTACTCCGTGTGGATCGCCGGCTCCATCCT
8	RPL27	NM_000988.3	24-123	GGGCGGGTGGTTCGTCGCCGAAATGGGCAAGTTCATGAAACCTGGGAAGGTGGT GCTTGTCTGGCTGGAGCGCTACTCCGGACGCAAGCTGTCATCGTG
9	RPS13	NM_001017.2	332-431	GCATCTTGAGAGGAACAGAAAGGATAAGGATGCTAAATTCCTGCTGATTTCTAATAG AGAGCCGGATTCACCGTTTTGGCTCGATATTATAAGACCAAGCGA
10	RPS20	NM_001023.2	187-286	TCGAATCACCCCTAACAAAGCCGCAACGTAATACTTTGGAAAAGGTGTGTGCTGAC TTGATAAGAGGGCGAAAAGAAAGAAATCTCAAGTGAAGGACCA

FIG. 1H

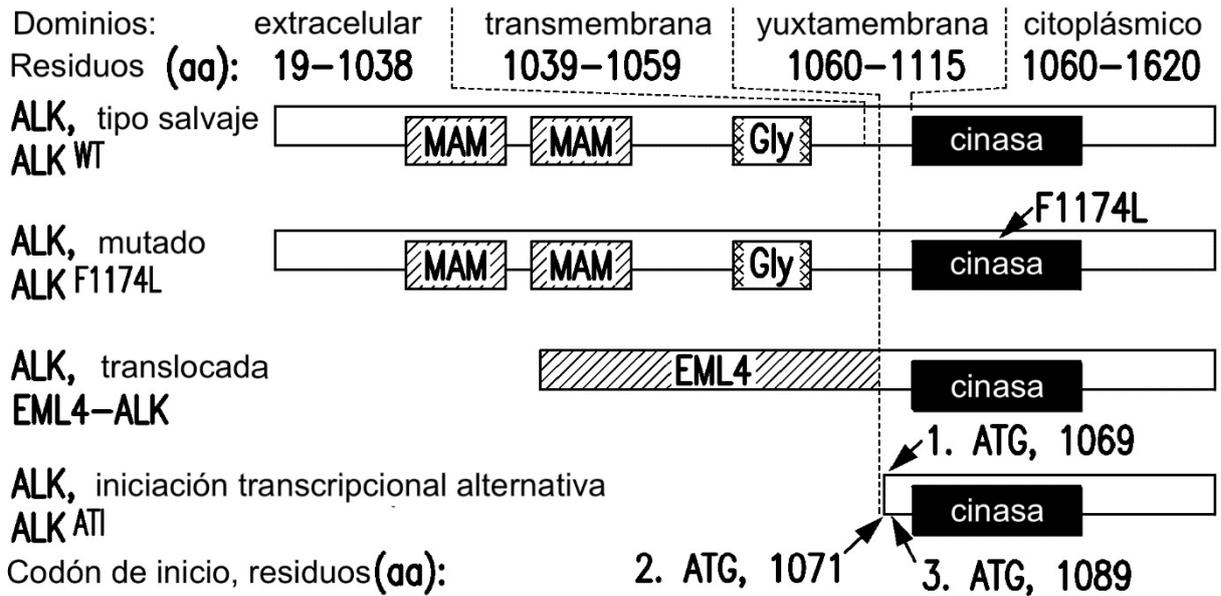


FIG. 2A

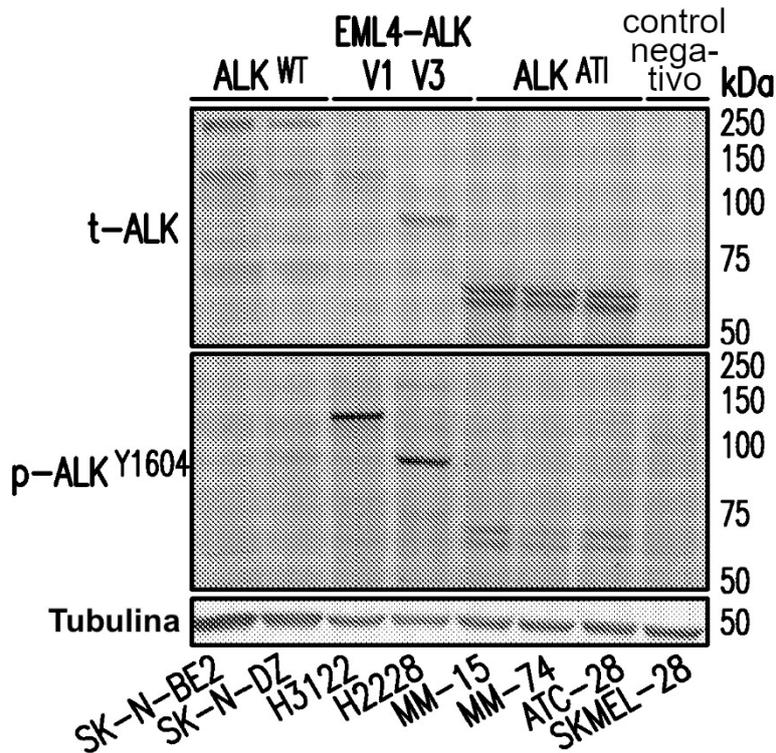


FIG. 2B

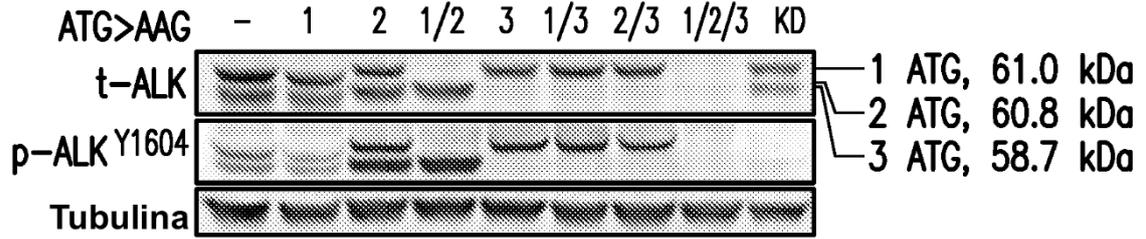


FIG. 2C

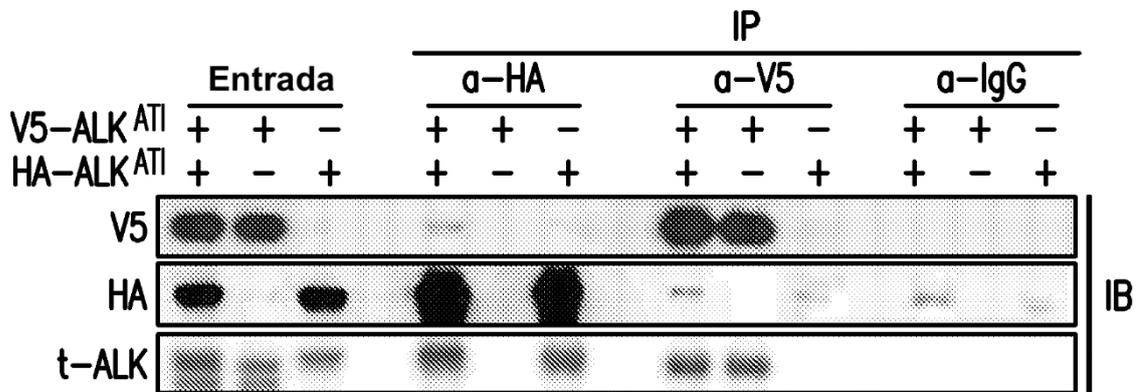


FIG. 2D

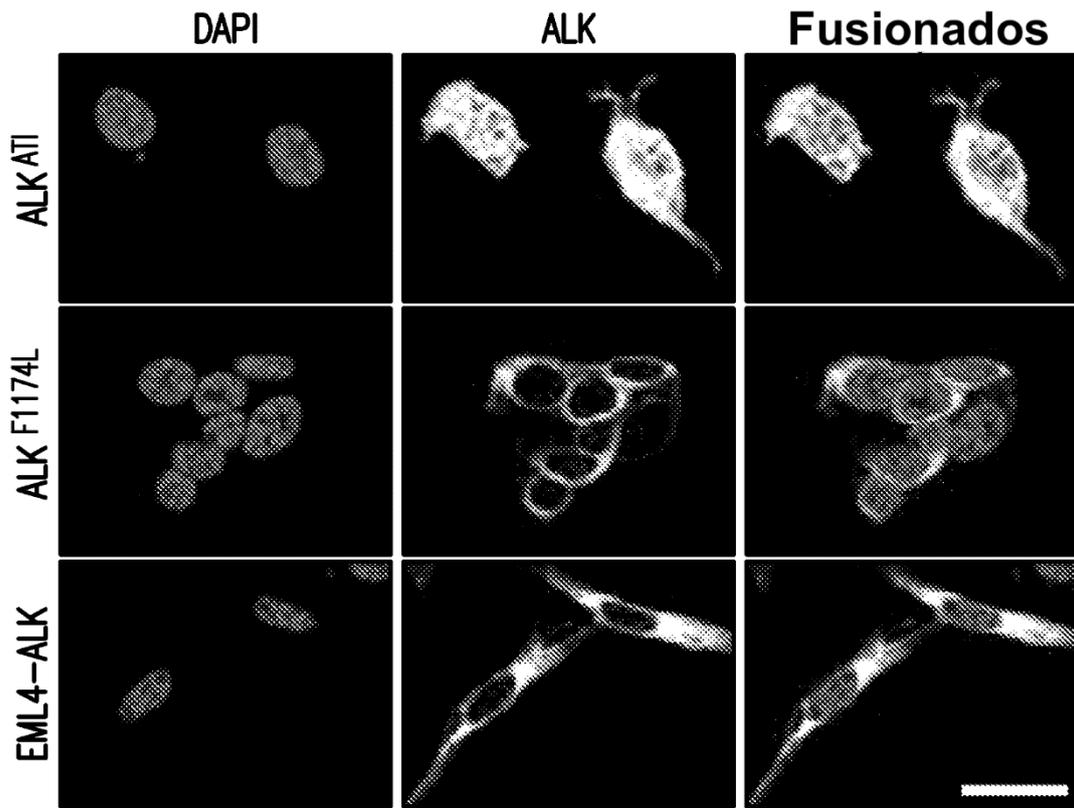


FIG. 2E

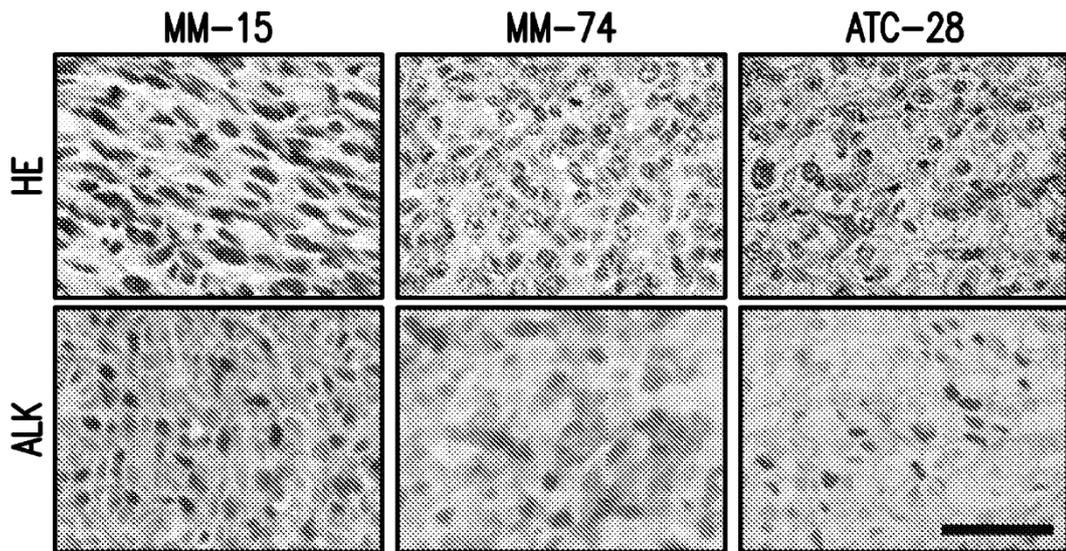


FIG. 2F

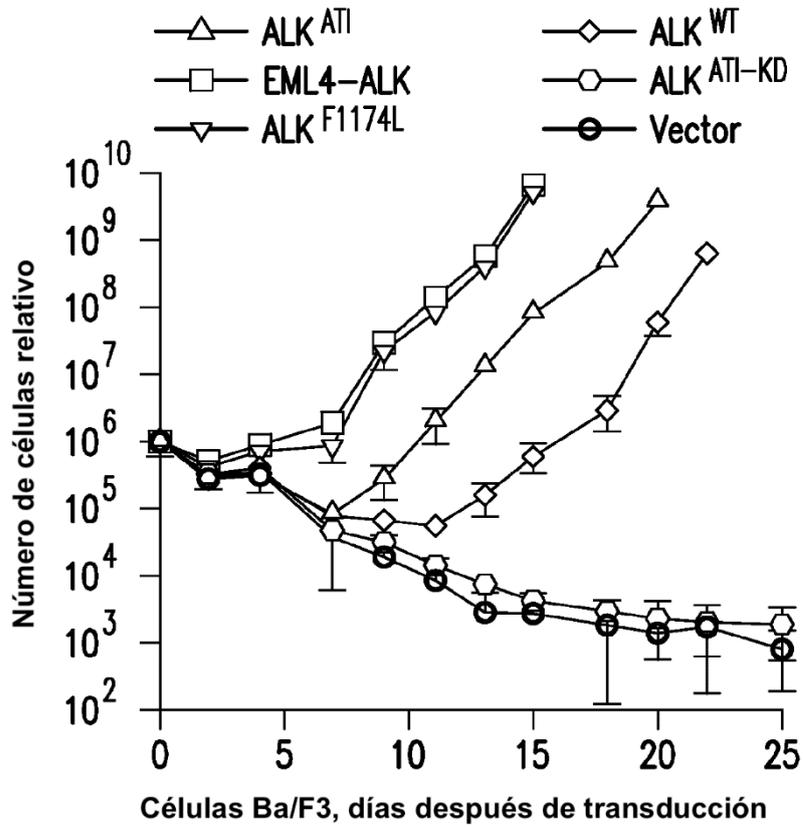


FIG. 3A

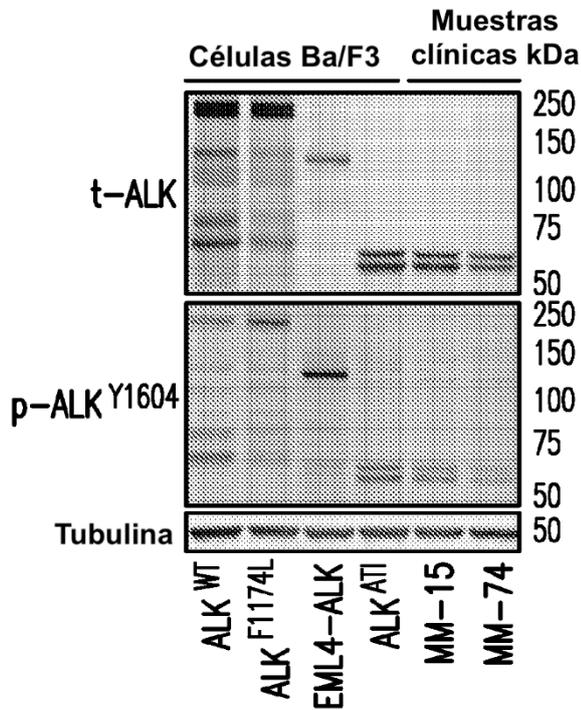


FIG. 3B

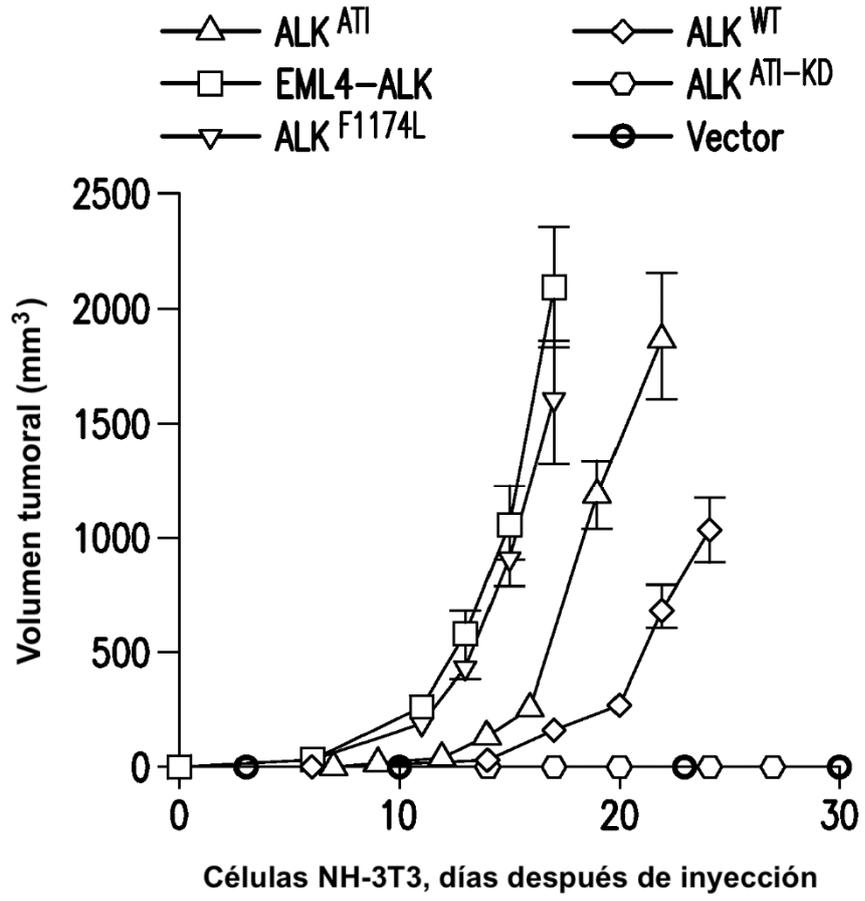


FIG. 3C

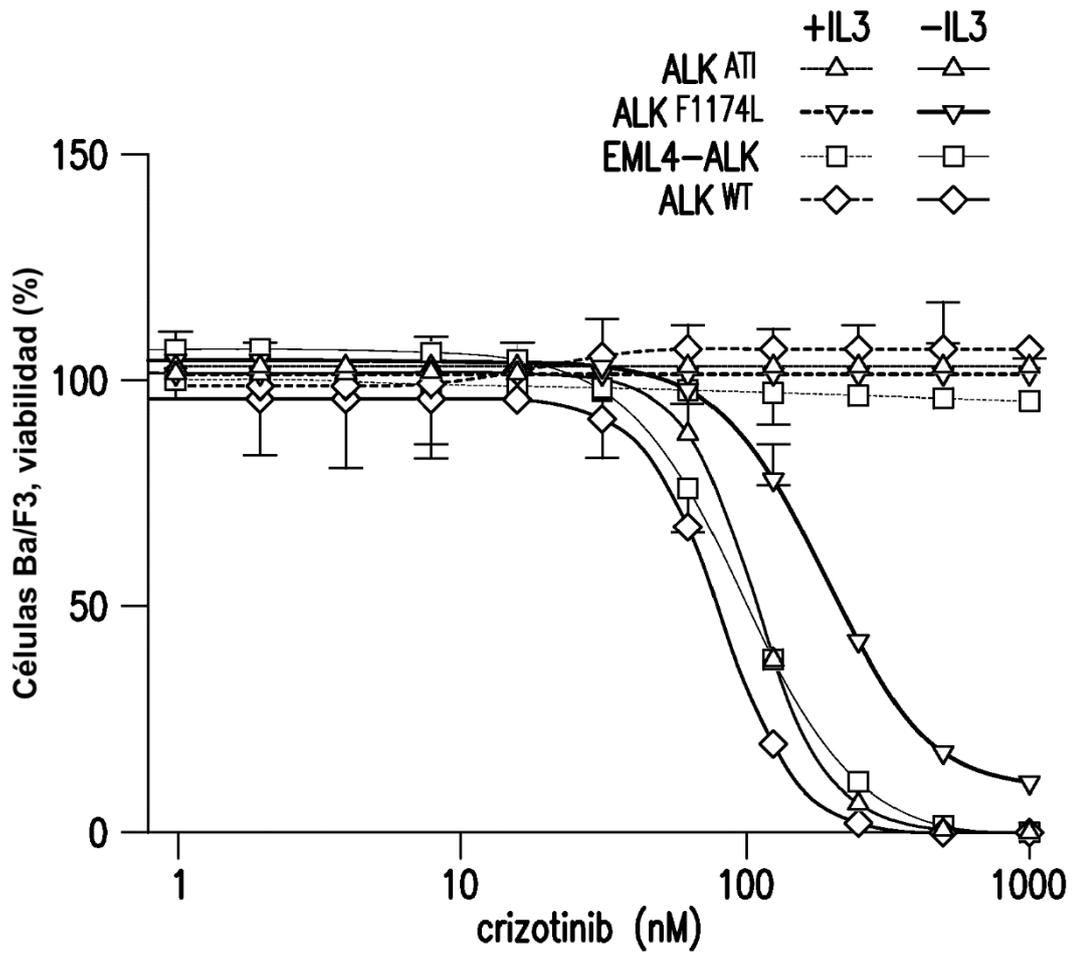
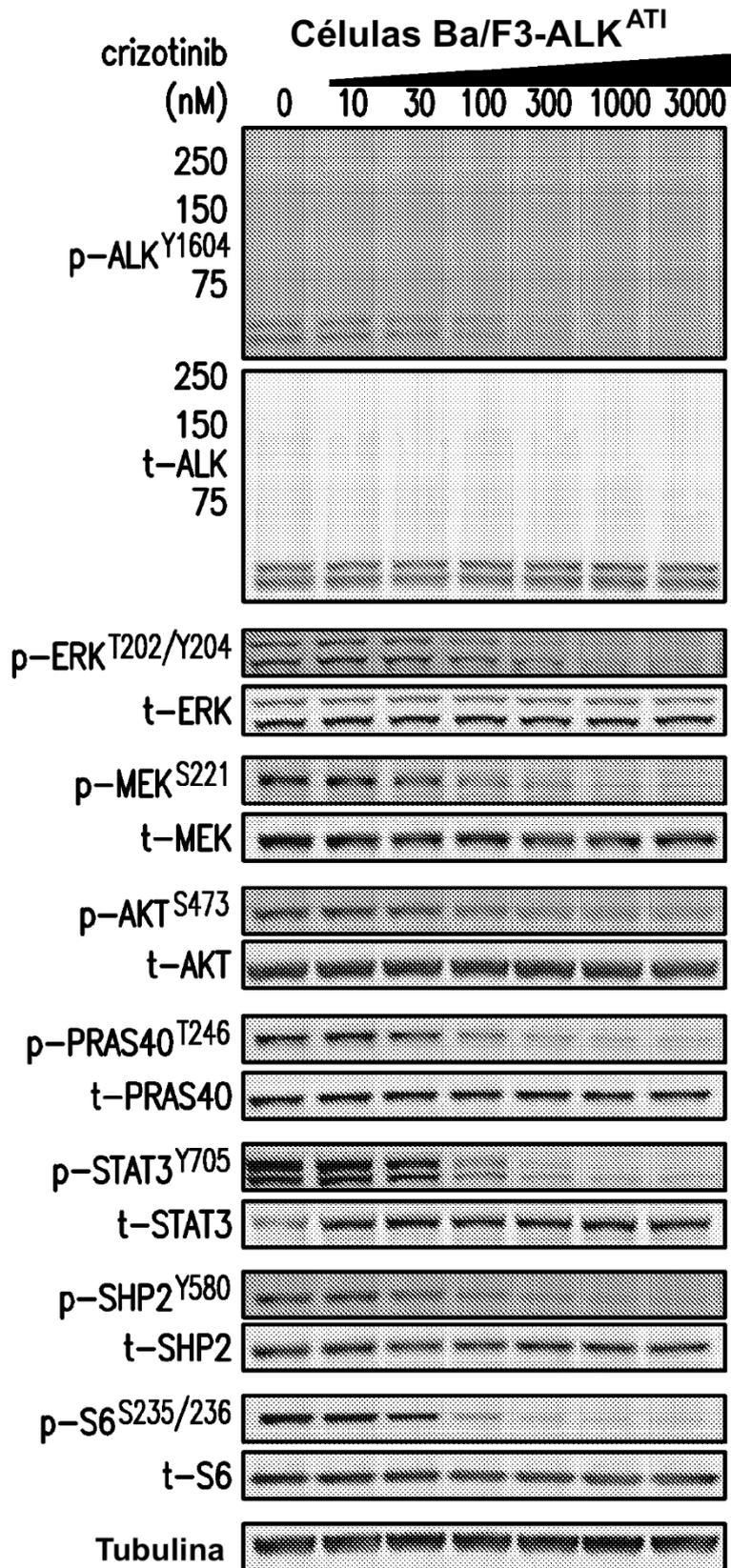


FIG. 4A



**FIG. 4B**

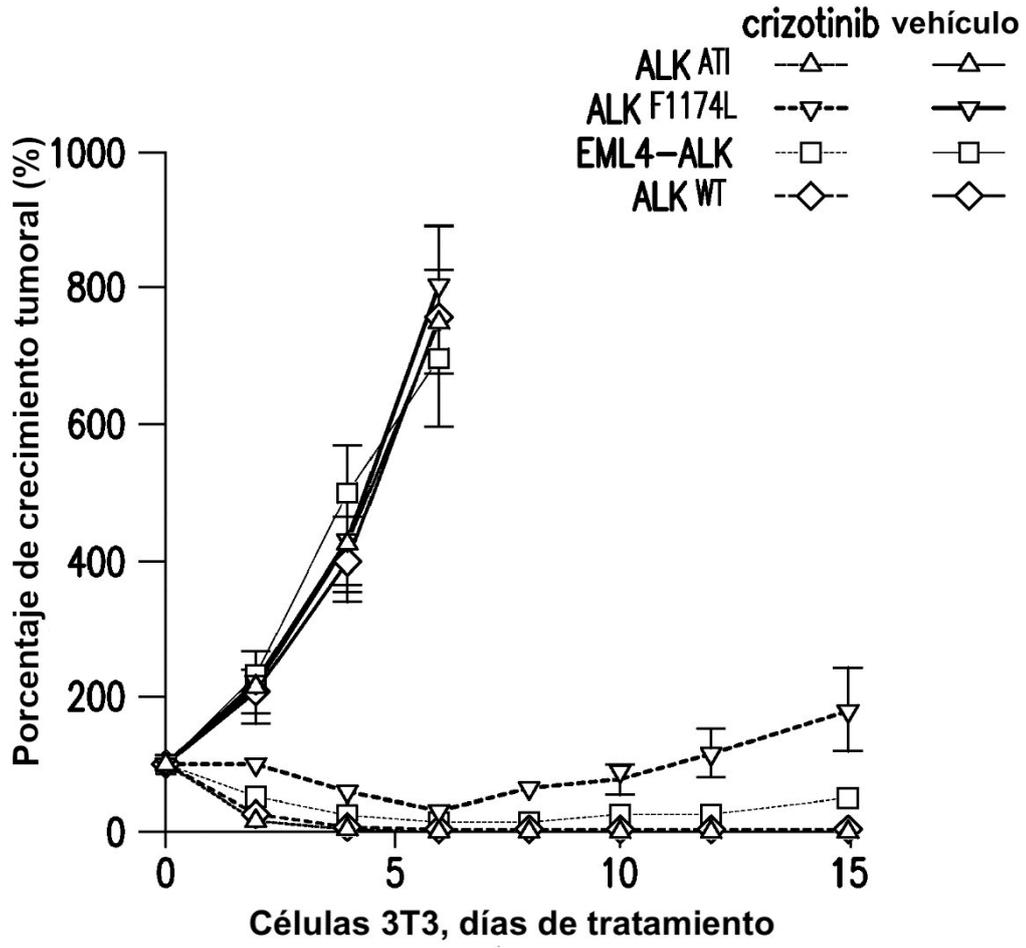
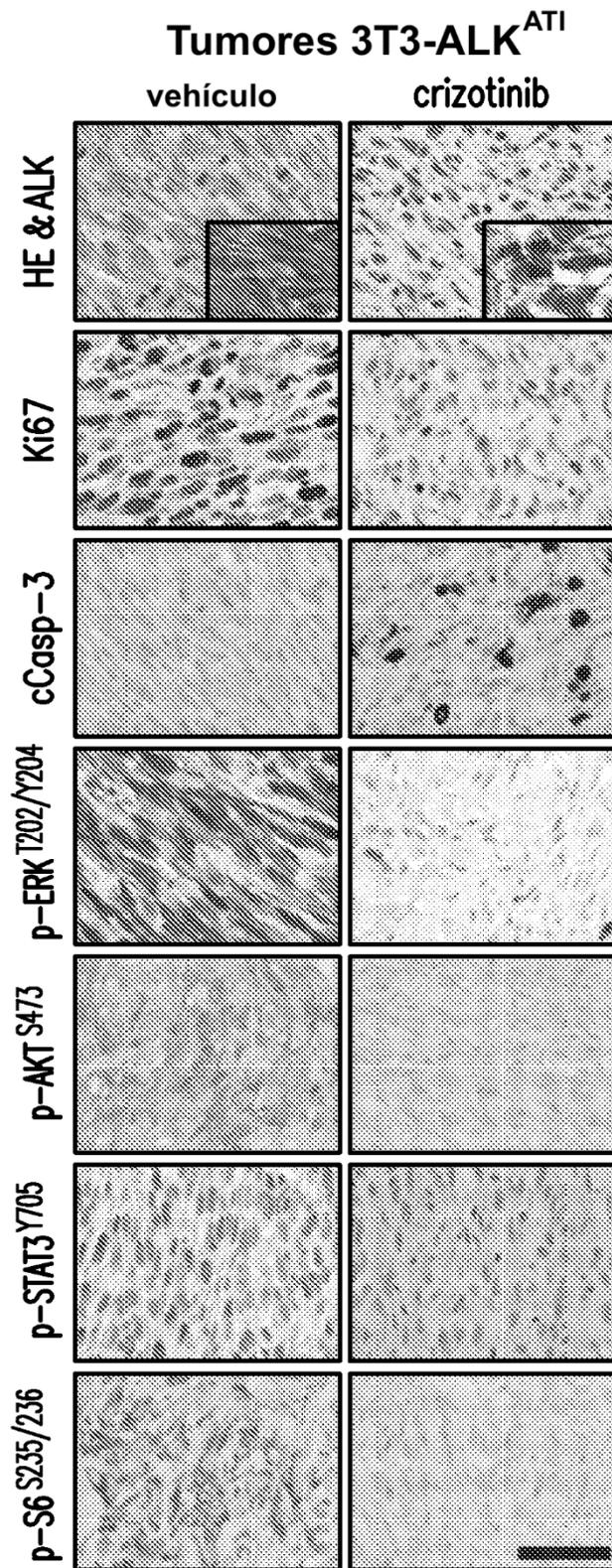
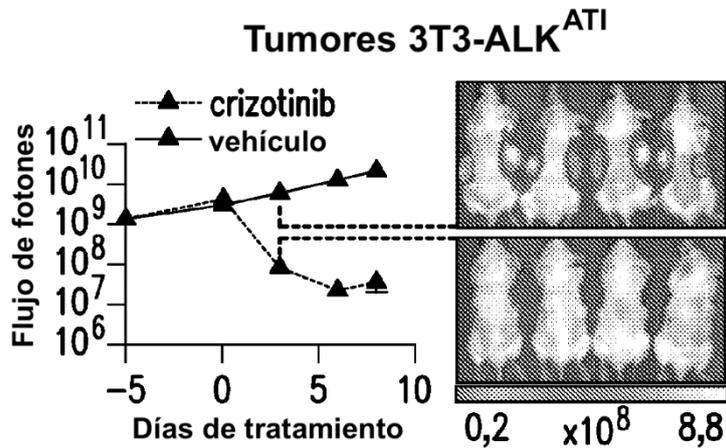


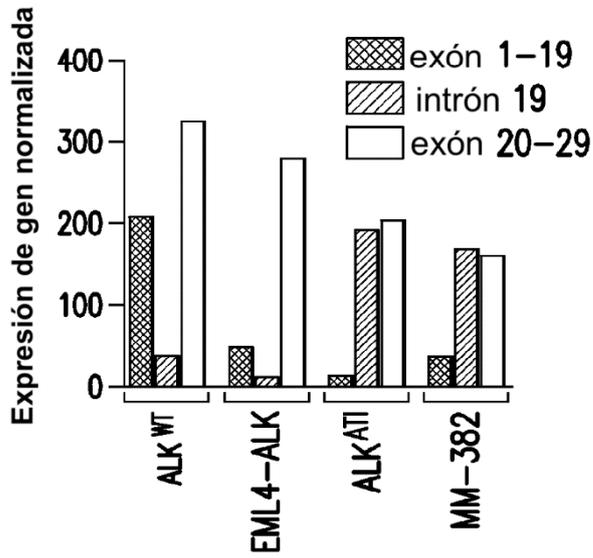
FIG. 4C



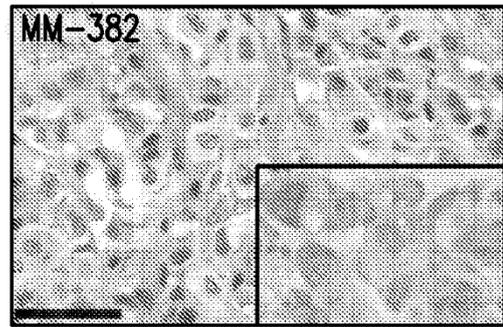
**FIG. 4D**



**FIG. 4E**



**FIG. 4F**



**FIG. 4G**



**FIG. 4H**

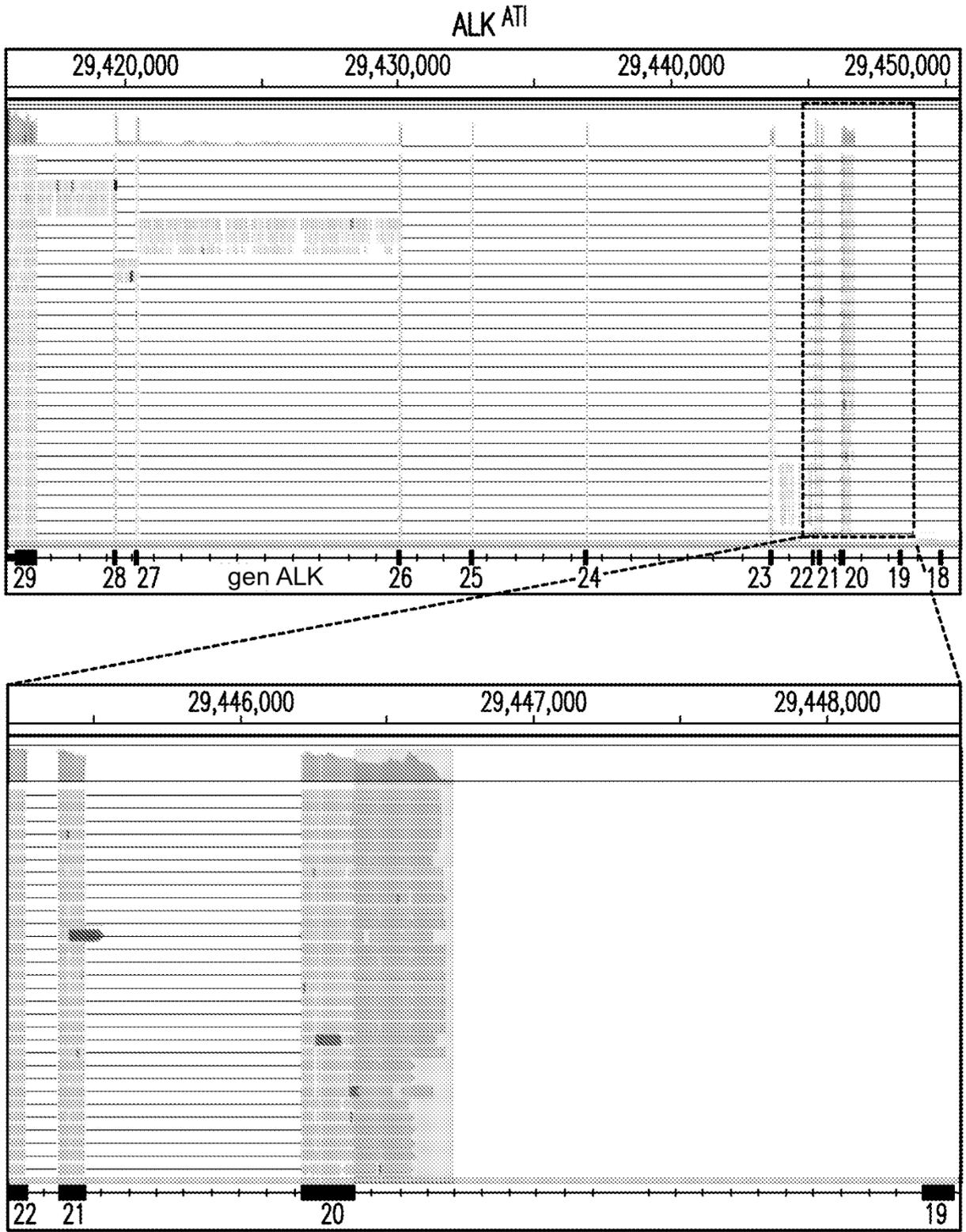


FIG. 5A

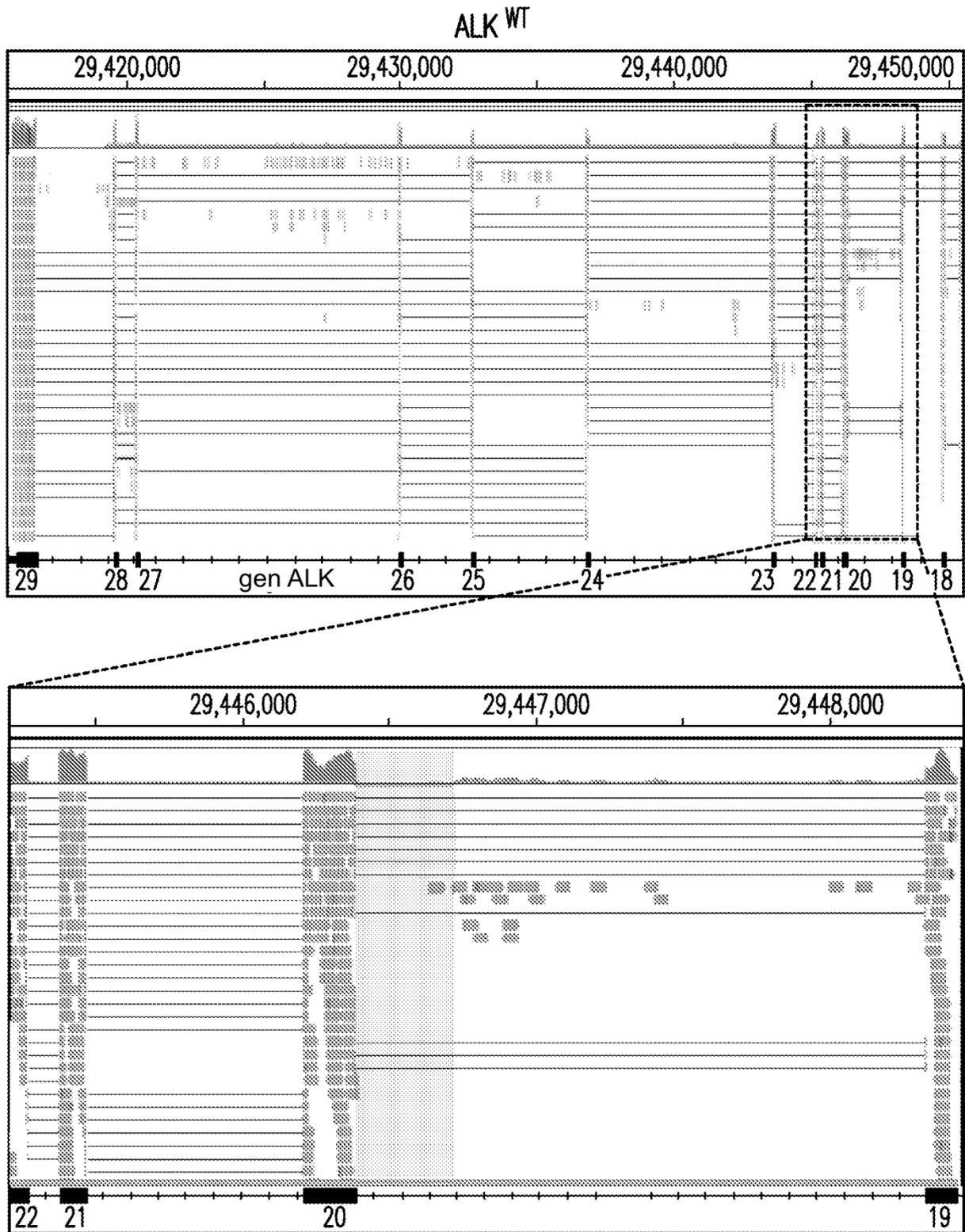
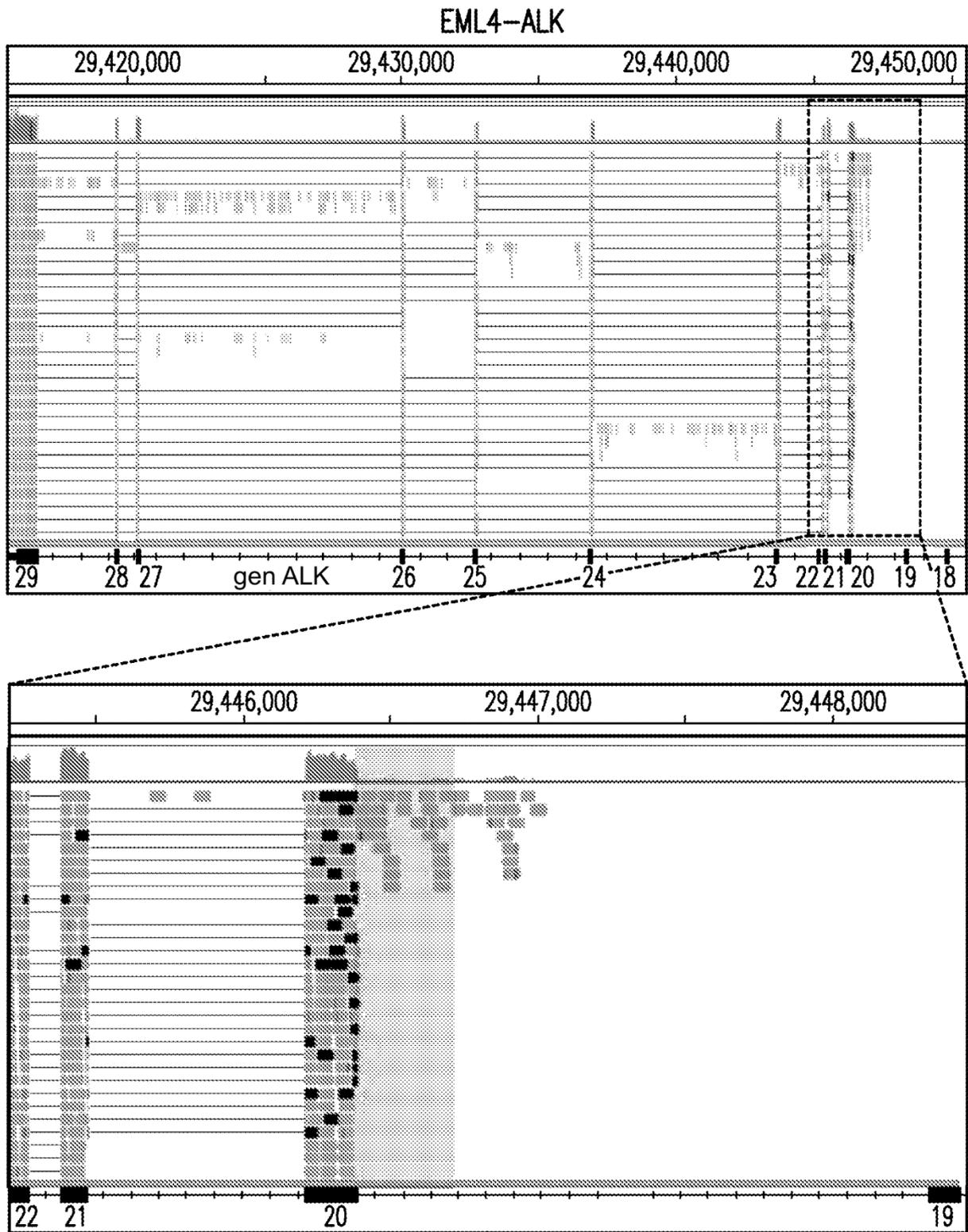


FIG. 5B



**FIG. 5C**

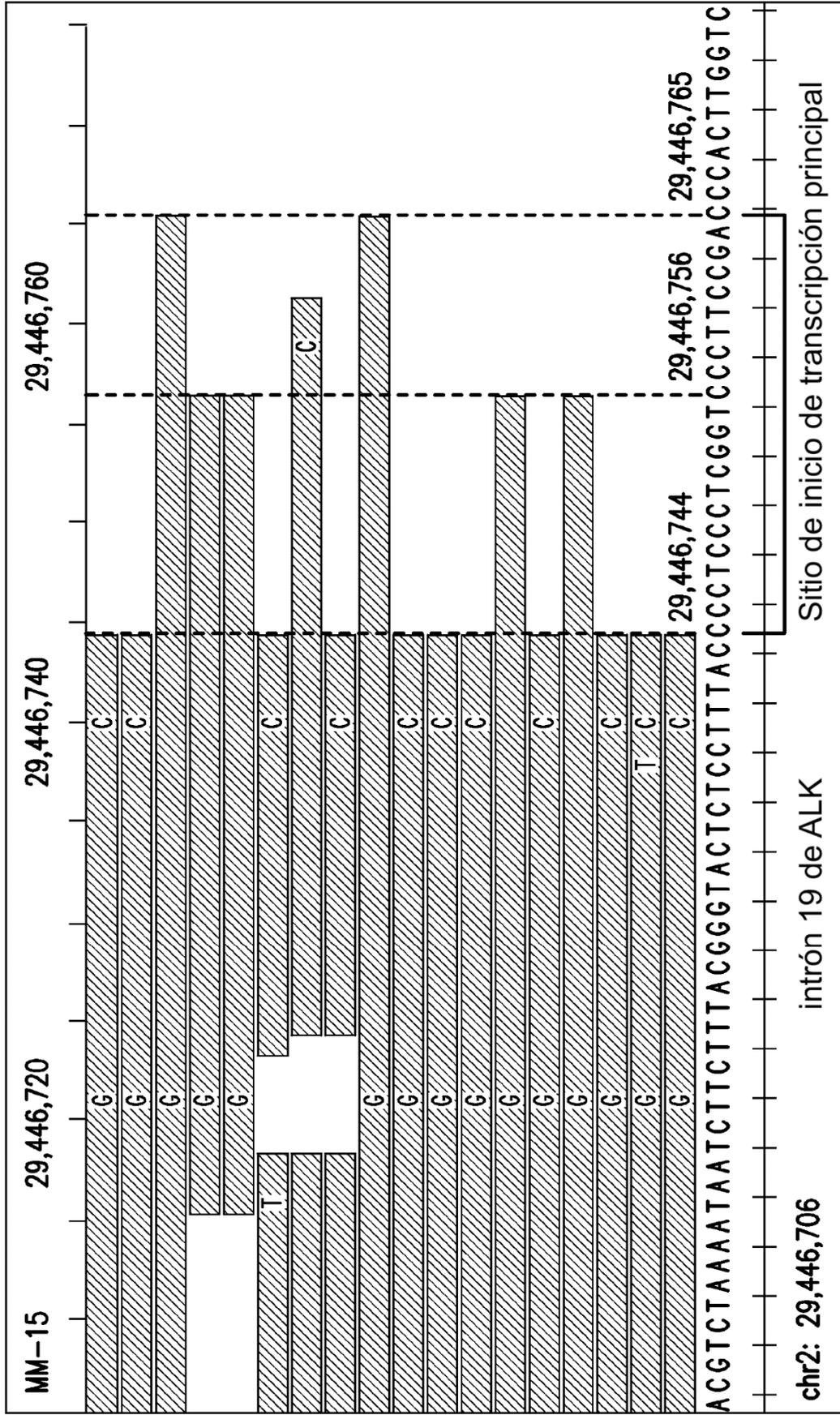


FIG. 6A

Sitio de iniciación de transcripción	MM-15 en %	MM-74 en %	ATC-28 en %	Sitio de iniciación de transcripción	MM-15 en %	MM-74 en %	ATC-28 en %
29,446,768	0,05	0,01	0,01	29,446,755	2,39	0,63	0,83
29,446,767	1,98	0,47	0,23	29,446,754	0,38	0,21	0,24
29,446,766	7,05	1,05	1,02	29,446,753	0,24	0,15	0,12
29,446,765	11,99	1,35	3,27	29,446,752	0,13	0,03	0,04
29,446,764	0,33	0,05	0,08	29,446,751	0,13	0,09	0,09
29,446,763	0,43	0,11	0,19	29,446,750	0,04	0,03	0,02
29,446,762	0,39	0,25	0,32	29,446,749	0,03	0,04	0,02
29,446,761	4,58	6,25	3,87	29,446,748	0,03	0,02	0,02
29,446,760	0,07	0,03	0,03	29,446,747	0,04	0,01	0,04
29,446,759	0,07	0,04	0,04	29,446,746	0,04	0,01	0,01
29,446,758	0,12	0,11	0,07	29,446,745	0,12	0,11	0,09
29,446,757	1,56	0,57	0,15	29,446,744	32,07	59,48	53,12
29,446,756	31,22	27,12	34,01	29,446,743	0,33	0,28	0,27

FIG. 6B

chr2:29,446,765 UTR en 5' de ALK<sup>ATI</sup> (intrón 19)

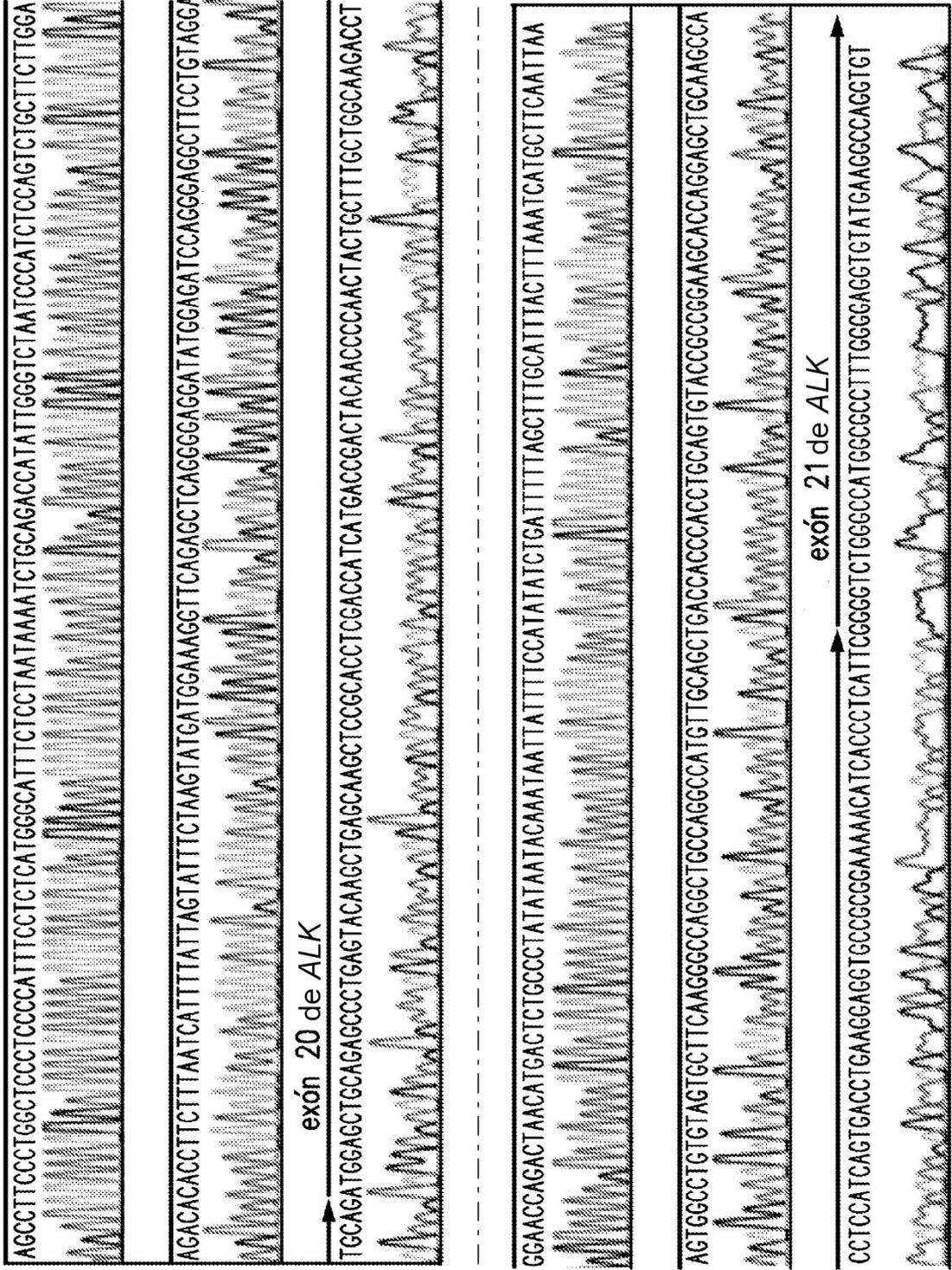


FIG. 6C

**CA**GGCCTTCCC **T**GGCTCCCTC **CC**GAATTTCCCT CTCATGGGCA TTTCTTCTAA TAAATCTGC AGACCATATT GGGCTAATC CCATCTCCAG TCTGCTTCTT 100  
 GGAGGAACCA GACTAACATG ACTCTGCCCT ATATAATACA AATAATTATT TTCCATATAT CTGATTTTAA GCTTTGCATT TACTTTAAT CATGCTTCAA 200  
 TTAAGACAC ACCTTCTTTA ATGATTTTAT TAGTATTTCT AAGTATGATG GAAAGTTCA GAGCTCAGGG GAGGATATGG AGATCCAGGG AGGCTTCCCTG 300  
 TAGGAAGTGG CCTGTGTAGT GCTTCAAGGG CCAGGCTGCC AGGCCATGTT GCAGCTGACC ACCCACCTGC AGTGTACCGC CGGAAGCACCC AGGAGCTGCA 400  
 AGCCATGCAG ATGGAGCTGC AGAGCCCTGA GTACAAGCTG AGCAAGCTCC GCACCTCGAC CATCATGACC GACTACAACC CCAACTACTG CTTTGTCTGGC 500  
 AAGACCTCCT CCATCAGTGA CCTGAAGGAG GTGCCGGGA AANAATCAC CCTCATTGGG GGTCTGGCC ATGGCGCCTT TGGGGAGGTG TATGAAGGCC 600  
 AGGTGTCCGG AATGCCCAAC GACCCAAGCC CCTGTCAAGT GCGTGTGAAG **AC**GGTGGCTG AAGTGTGCTC TGAACAGGAC GAACCTGGATT TCCTCATGGA 700  
 AGCCCTGATC ATCAG**CC**AAAT TCAACCACGA GAACATTGTT CCGTGCATTG GGGTGAGCCT GCAATCCCTG CCCCAGTTGA TCCTGTCTGA GCATCATGGCG 800  
 GGGGAGACC TCAAGTCCCT CCTCCGAGAG ACCCGCCCTC GCCCCAGCCA GCCCCCTCC CTGGCCATGC TGGACCTTCT GCACGTGGCT CGGGACATTG 900  
 CCTGTGGCTG TCAGTATTG GAGGAAACC ACTTCATCCA CCGAGACATT GCTGCCAGAA CTGCCCTCTT GACCTGTCCA GGGCTTGGAA GAGTGGCCAA 1000  
 GATTGGAGAC TTCGGGATGG CCCCAGACAT CTACAGGGCG AGCTACTATA GAAAGGGAGG CTGTGCCATG CTGCCAGTTA AGTGGATGCC CCCAGAGGCC 1100  
 TTCATGGAAG GAATTTCCAC TTCTAAACA GACACAT**GG**T CCTTTGGAGT GCTGCTATGG GAAATCTTTT CICTTGGATA TATGCCAIAC CCCAGCAAAA 1200  
 GCRAACCAGGA AGTCTGGAG TTGTACCCA GTGGAGGCCG GATGGACCCA CCCAAGAACT GCCCCGGCC TGTATACCGG ATAATGACTC ACACCCCTT GCCGATAGAA 1300  
 ACATCAGCCT GAAGACAGGC CCAACTTIGC CATCATTTIG GAGAGGATIG AATACIGCAC CCA**GG**ACCCCG GATGTAATCA ACACCCCTT GCCGATAGAA 1400  
 TATGGTCCAC TTGTGGAAGA GGAAGAGAAA GTGCCGTGTA GCCCCAAGGA CCGTGAAGGG GTTCTCTCTC TCCTGGTCTC TCAACAGGCA AAACGGGAGG 1500  
 AGGAGCCAG CCCAGCTGCC CCACCACCTC TCCCTACCAC TCCCTCTGGC AAGGCTGCAA AGAAACCCAC AGCTGCAGAG ATCTCTGTTC GAGTCCCTAG 1600  
 AAGGCGGCCG GTGGAAGGGG GACACGTGAA TATGGCATTG TCTCAGTCCA ACCCTCCTTC GGAGTGGCAC AAGGTCACG GATCCAGAAA CAAGCCACCC 1700  
 AGCTTGTGGA ACCCAACGTA CGGCTCCTGG TTIACAGAGA AACCCACCAA AAAGAATAAT CCTATAGCAA AGAAGGAGCC ACAGGACAGG GGTAACTGG 1800  
 GCGTGGAGGG AAGCTGTACT GTCCCACCTA ACGTTGCAAC TGGGAGACTT CCGGGGGCCCT CACTGTCTCT AGAGCCCTCT TCGCTGACTG CCAATATGAA 1900  
 GGAGTACCT CTGTTACGGC TAGTCACTT CCCTTGTGGG AATGTCAATT ACGGCTACCA GCAACAGGGC TTGCCCTTAG AAGCCGCTAC TGCCCTTGG 2000  
 GCTGGTCATT ACGAGGATAC CATTCTGAAA AGCAAGAATA GCATGAACCA GCGTGGGCCG TGAGCTCGGT CGCACACTCA CTTCTCTTCC TTGGGATCCC 2100  
 TAAGACCGTG GAGGAGAG AGCCAATGGC TCCTTCACAA ACCAGAGACC AAATGTACG TTTTGTGTTG TGCCAACCTA TTTTGAAGTA CCACCAAAA 2200  
 AGCTGTATT TGAAATGCT TTAGAAGGT TTTGAGCATG GGTTCATCCT ATTCTTTCGA AAGAAGAAA TATCATAAA ATGAGTGATA AATACAAGGC 2300  
 CCAGATGGG TTGCATAAGG TTTTATGCA TGTGTTGT ATACTTCTT ATGCTTCTT CAATTTGT GTGCTCTGCT TCAATGTAGT CAGAATTAGC 2400  
 TGCTTCTATG TTTCATAGT GGGTTCATAG ATGTTTCCCT GCCTTGTGTA TGTGGACATG AGCCATTTGA GGGGAGAGG AACGGAAATA AAGGAGTTAT 2500  
 TTGTAATGAC TAA [ SEQ ID NO: 11]

2513

**FIG. 6D**

MQMELQSP EY KLSKLR TSTI MTDYNP NYCF AGKTSSISDL KEVPRKNITL IRGLGHGAFG EVYEQVSGM PNDPSPLOVA VKTLPEVCSE QDELDFLMEA 100  
 LIISKFNHQN IVRCIGVSIQ SLPRFILLLEL MAGGDLKSEL RETRPRPSQP SSLAMLDLLH VARDIACCCQ YLEENHFIHR DIAARNCLLT CPGCRVAKTI 200  
 GDEGMARDIY RASYRKGCC AMILPVKIMPP EAFMEGIFTS KTDIWSFGVL LWEIFSLGYM PYPKSNQEV LEFVTSGRM DPPKNCQGPV YRINTQCWQH 300  
 QPEDRPNFAT ILERIEYCTQ DPDVINTALP IEYGPLVEE EKVPVRPKDP EGVPLLVSQ QAKREERSP AAPPLPTTS SGKAAKKPTA AEISVRVPRG 400  
 PAVEGGHVM AFSQSNPPSE LHKVHGSRNK PTLWNPTYG SWFTEKTKK NNPIAKKEPH DRGNLGLEGS CTYPPNWTG RLPGASLLE PSSLTANMKE 500  
 VPLFRLRHFP CGNVNYGYQQ QGLPLEAATA PGAGHYEDTI LKSKNSMNQP GP\* [SEQ ID NO: 12] 552

FIG. 6E

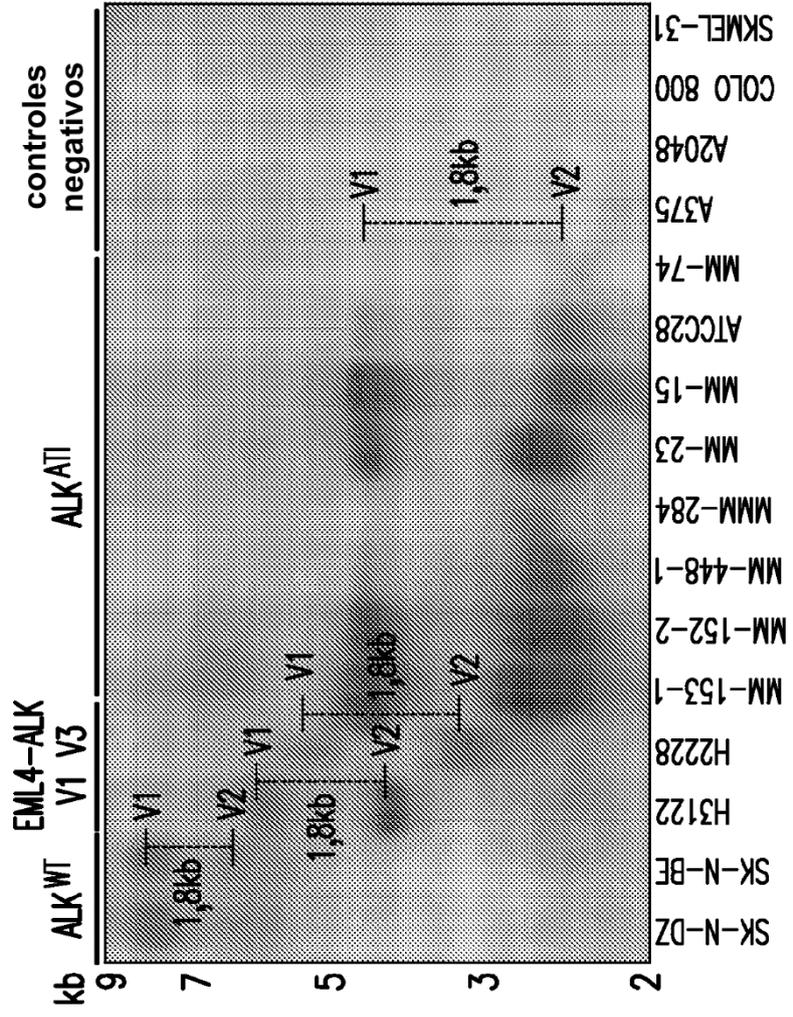


FIG. 6F

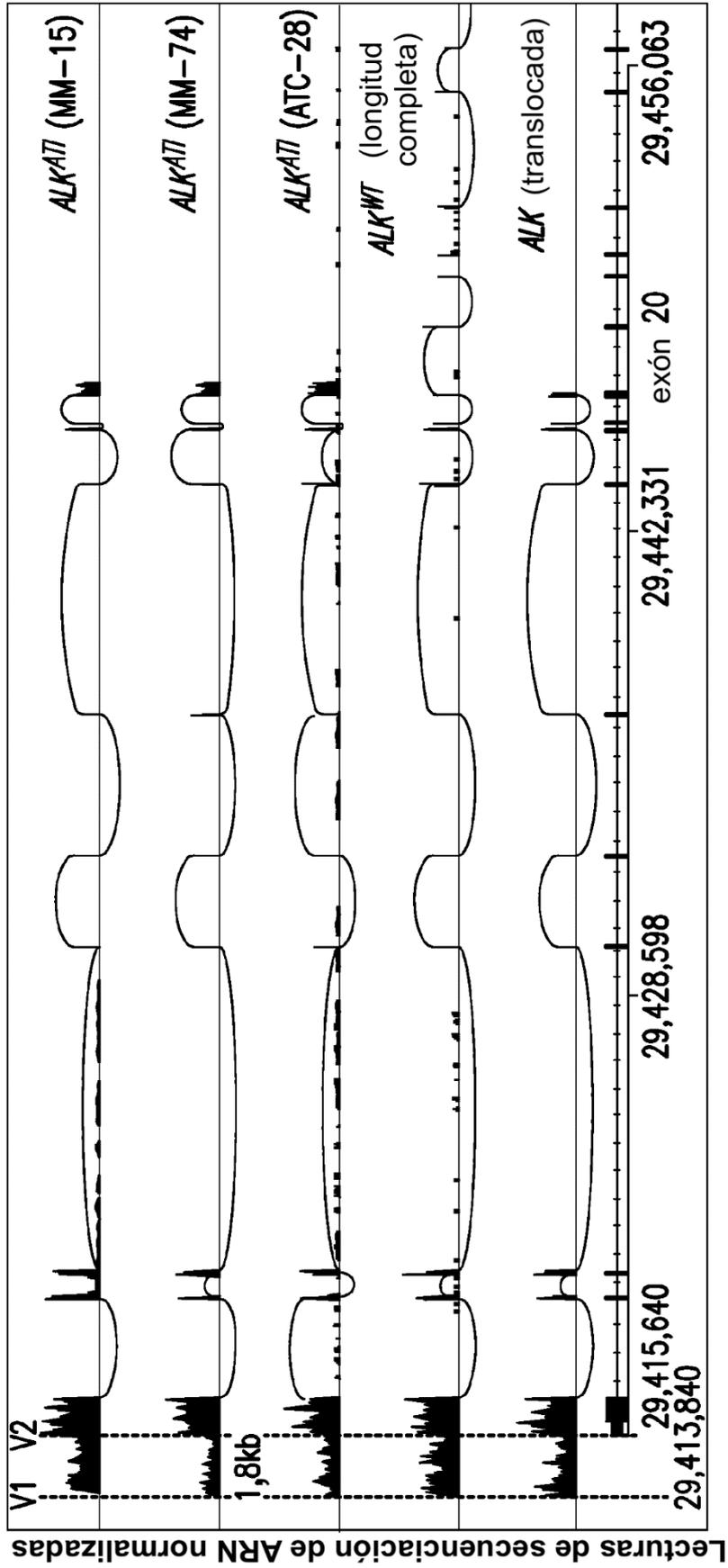


FIG. 6G

n.º	Cebador directo	Cebador inverso	Alineamiento de amplicón
P1	ACAAGCTGAGCCATGAGGAC	AGGAGTTACCATCCCTGCCT	chr2:29,447,155–29,447,235
P2	AAGGGGAGCCAGGTAGACTT	ATGAGAGGAGAACCAGGGCT	chr2:29,445,931–29,446,002
P3	ATTCAGCCCCTACACTGCAC	CGCGGAAAAACATCACCTC	chr2:29,446,154–29,446,232
P4	GAGGCGGGGTAACATACAC	GTGAAGAACTGGAAGCCCGA	chr2:29,445,577–29,445,657
P5	TCATGCTCCTTGGGGAGAGA	TCTAGAGCAACTGCCCTTC	chr2:29,445,799–29,445,887

FIG. 7A

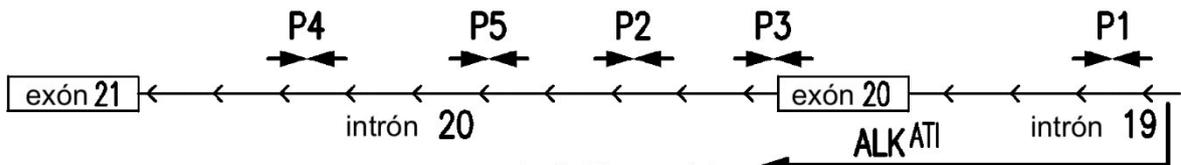


FIG. 7B

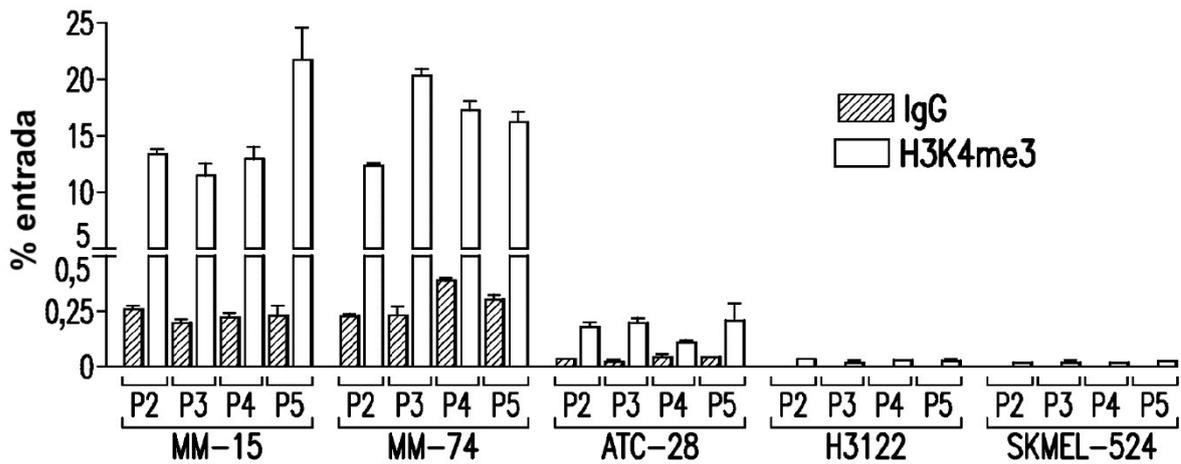


FIG. 7C

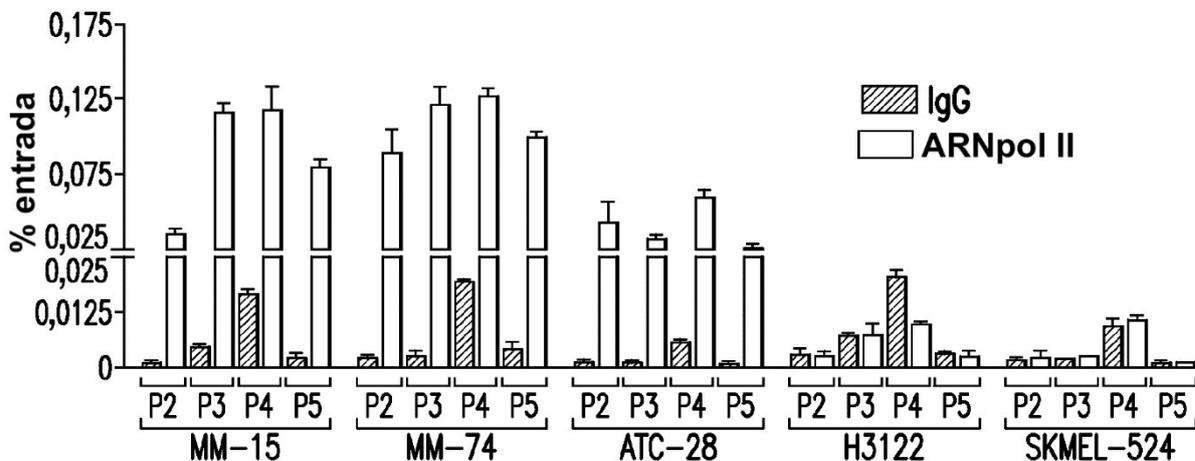


FIG. 7D

<b>Tipo</b>	<b>ALK<sup>AT</sup></b>	<b>Número total de casos</b>	<b>%</b>
Melanoma cutáneo de piel (SKCM)	38	334	11,34
Adenocarcinoma de pulmón (LUAD)	3	470	0,64
Carcinoma de células escamosas de pulmón (LUSC)	1	482	0,20
Carcinoma de células claras renales de riñón (KIRC)	2	480	0,42
Carcinoma invasivo de mama (BRCA)	1	988	0,10
Carcinoma de tiroides (THCA)	0	482	0,00
Glioblastoma multiforme (GBM)	0	153	0,00
Glioma de grado inferior de cerebro (LGG)	0	271	0,00
Carcinoma urotelial de vejiga (BLCA)	0	182	0,00
Adenocarcinoma de próstata (PRAD)	0	195	0,00
Carcinoma endometrial de cuerpo uterino (UCEC)	0	118	0,00
Renal cromóforo (KICH)	0	66	0,00
Adenocarcinoma colorrectal (COADREAD)	0	316	0,00
Carcinoma ovárico (OV)	0	261	0,00
Carcinoma de células escamosas de cuello y cabeza (HNSC)	0	303	0,00

**FIG. 8A**

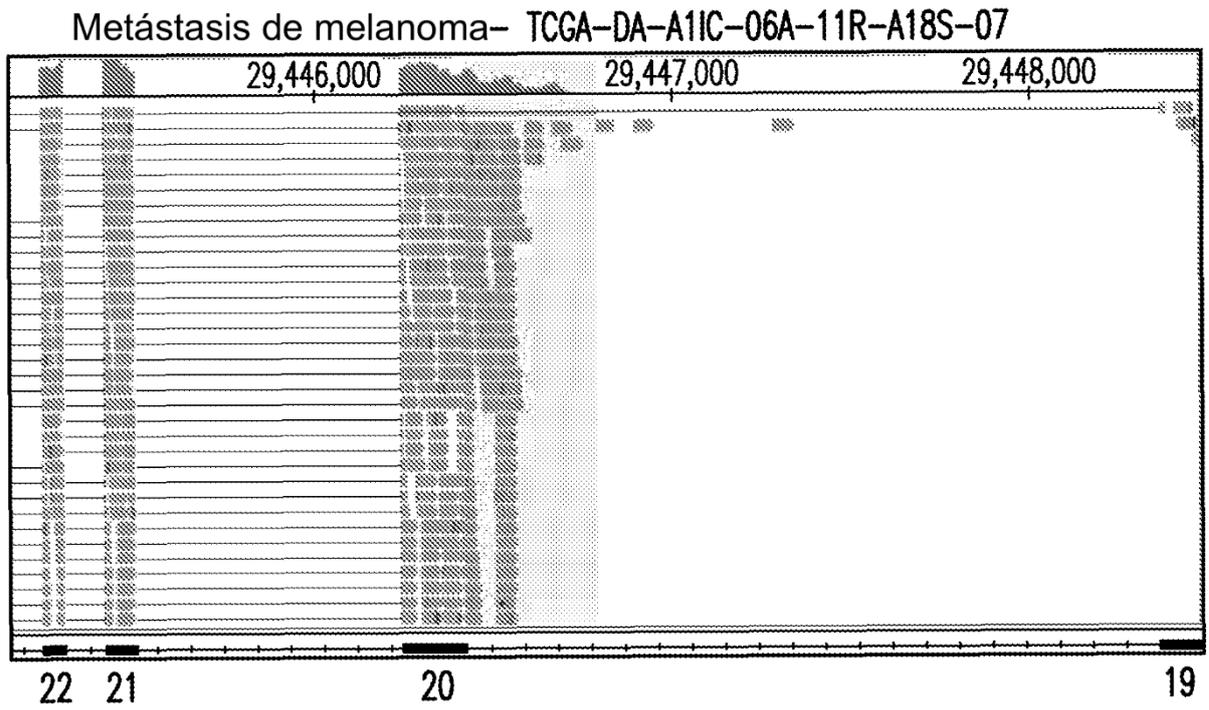


FIG. 8B

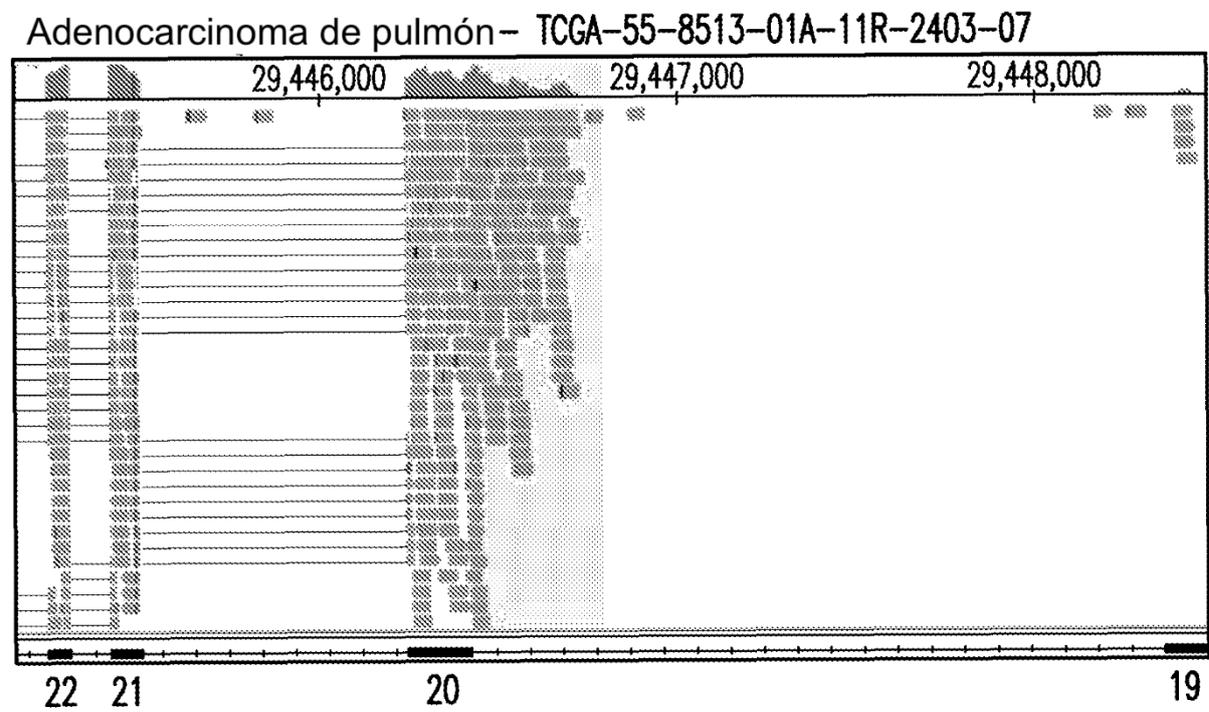


FIG. 8C

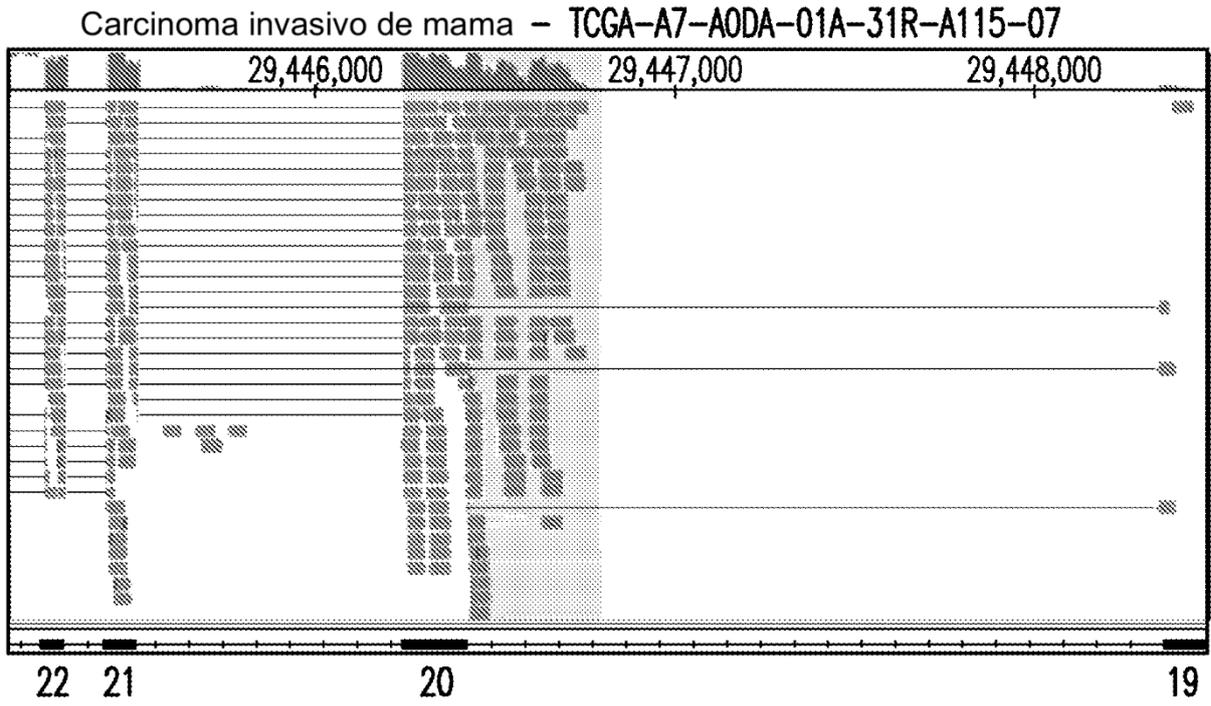


FIG. 8D

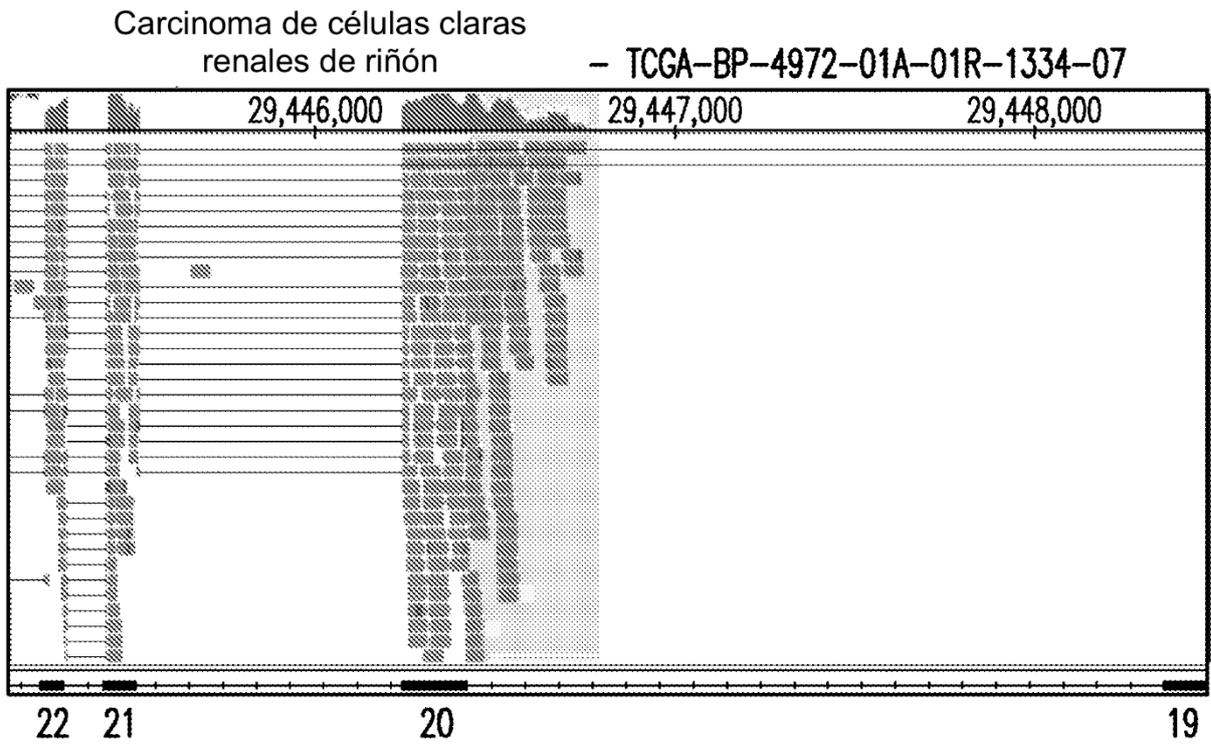


FIG. 8E

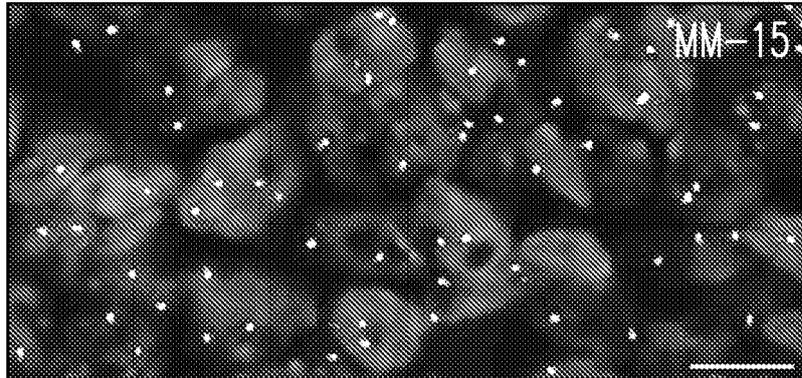


FIG. 9A

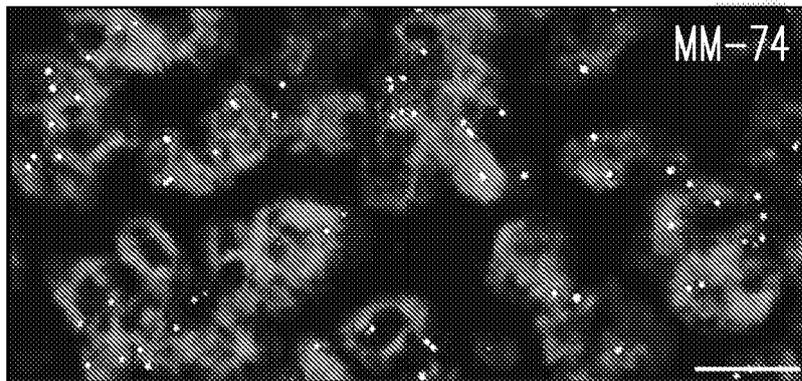


FIG. 9B

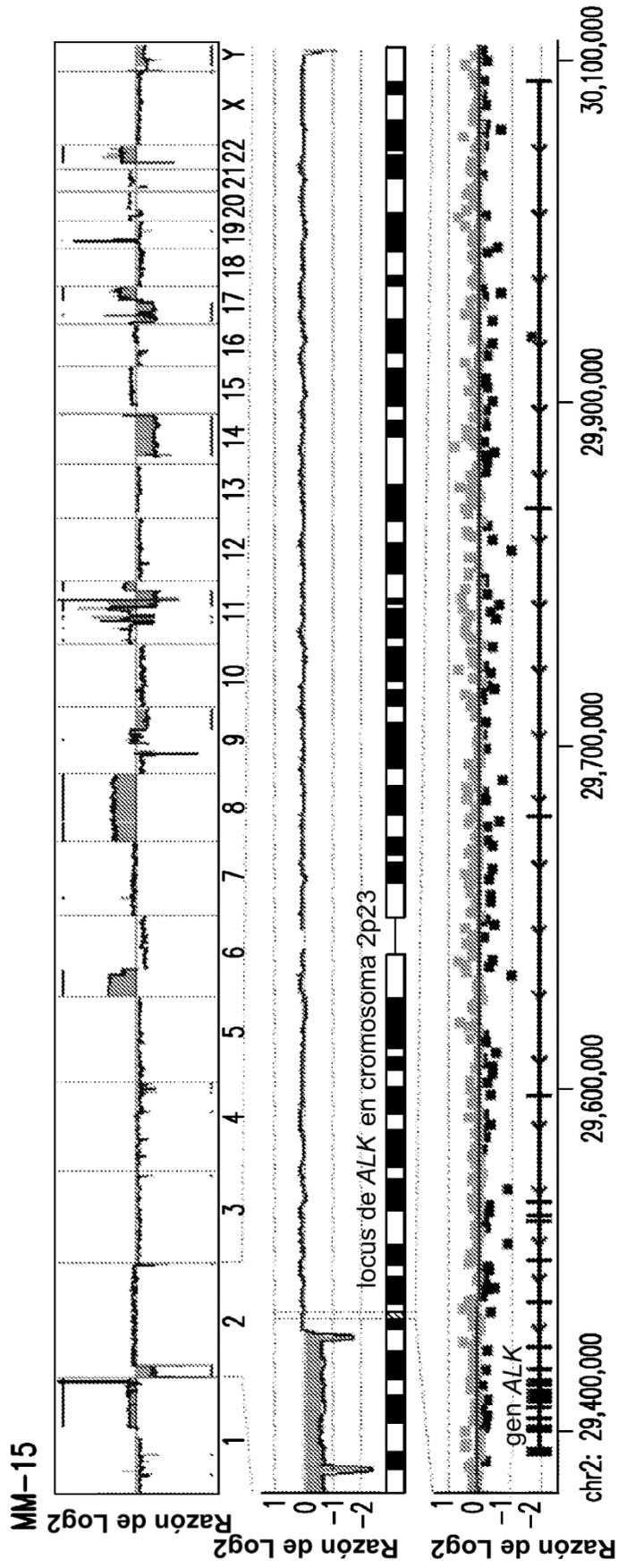


FIG. 9C

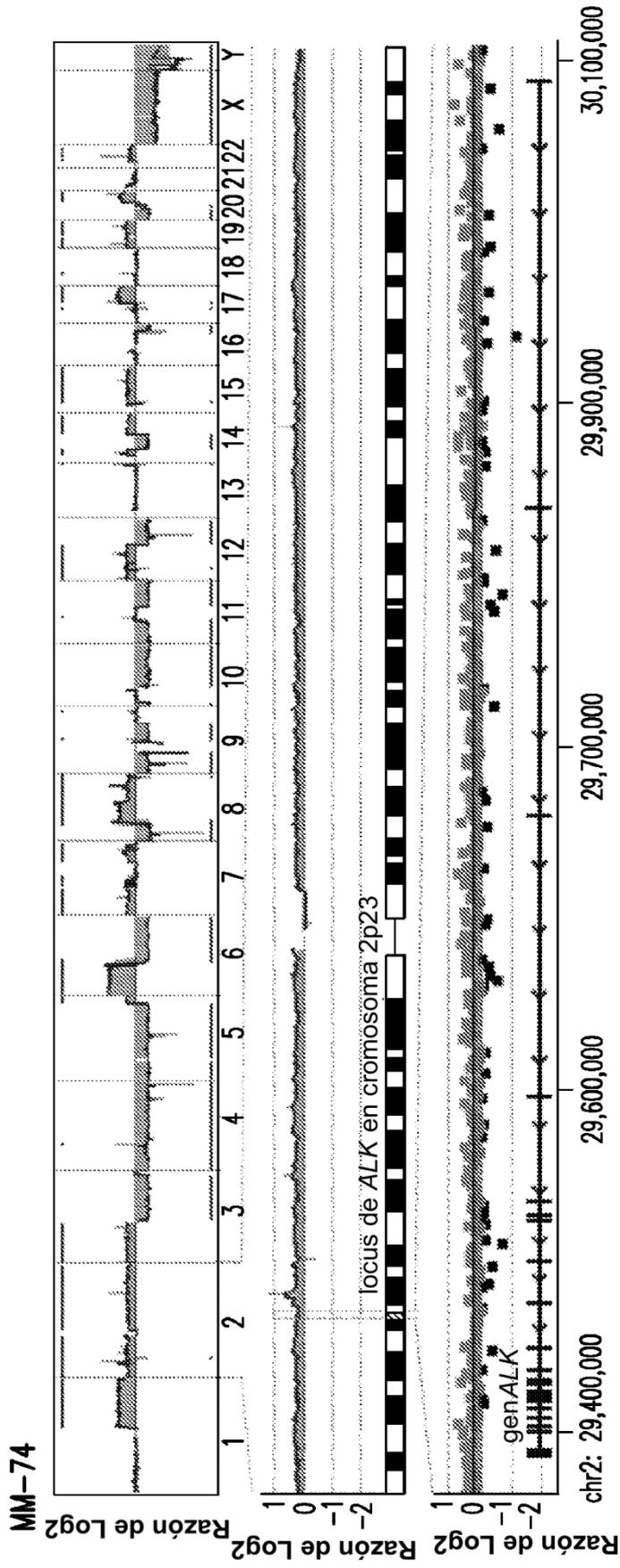


FIG. 9D

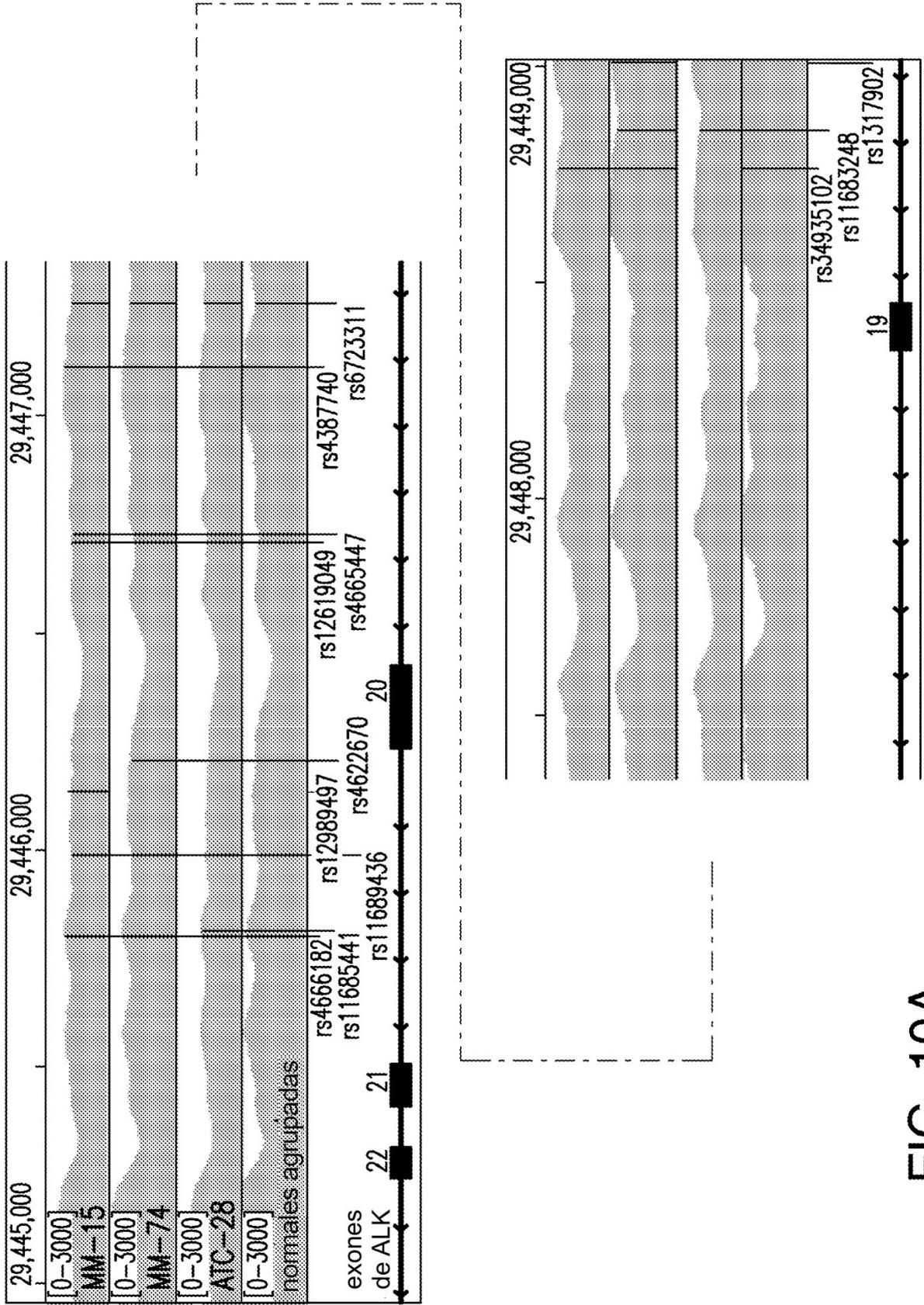


FIG. 10A

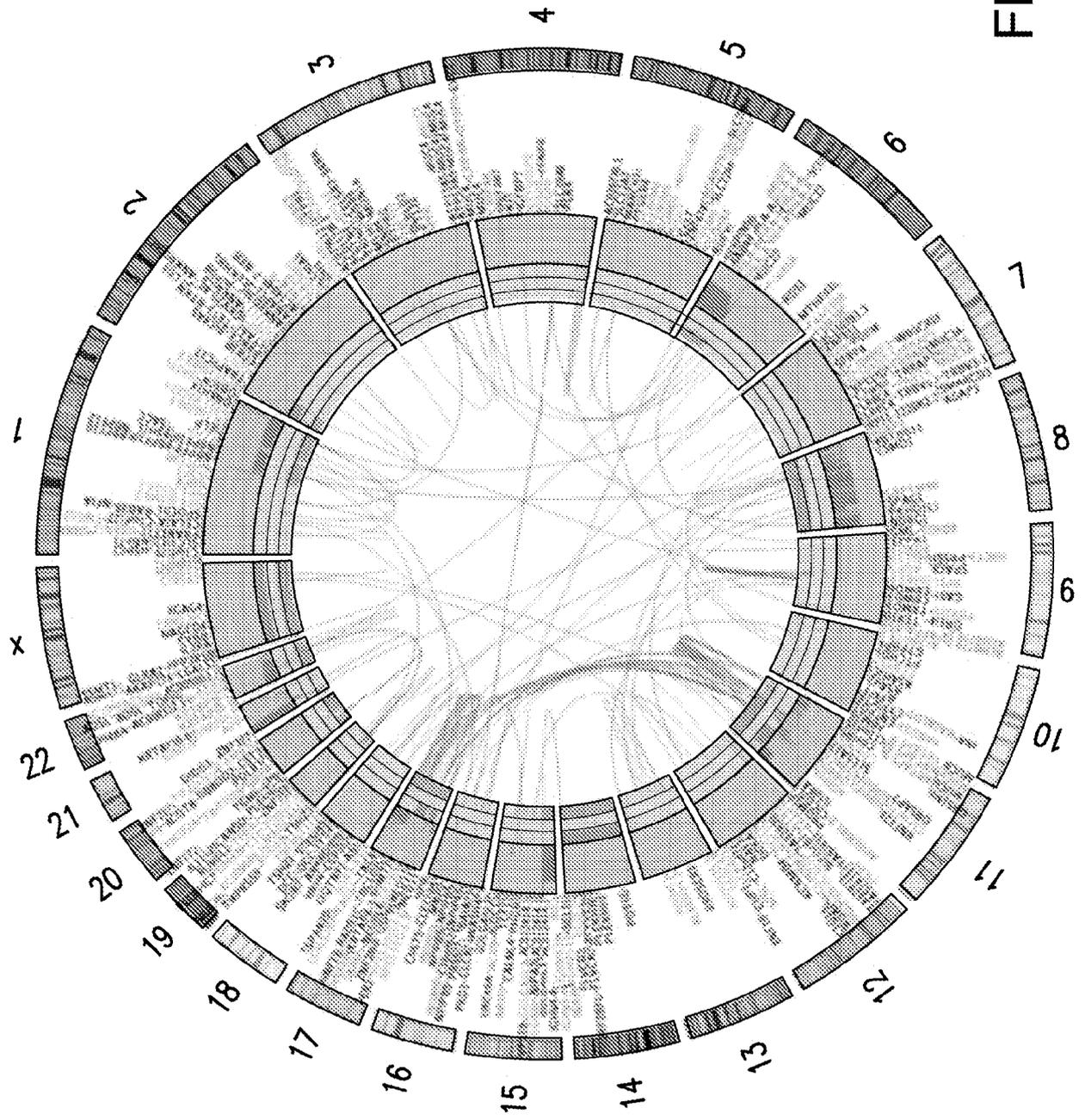


FIG. 10B

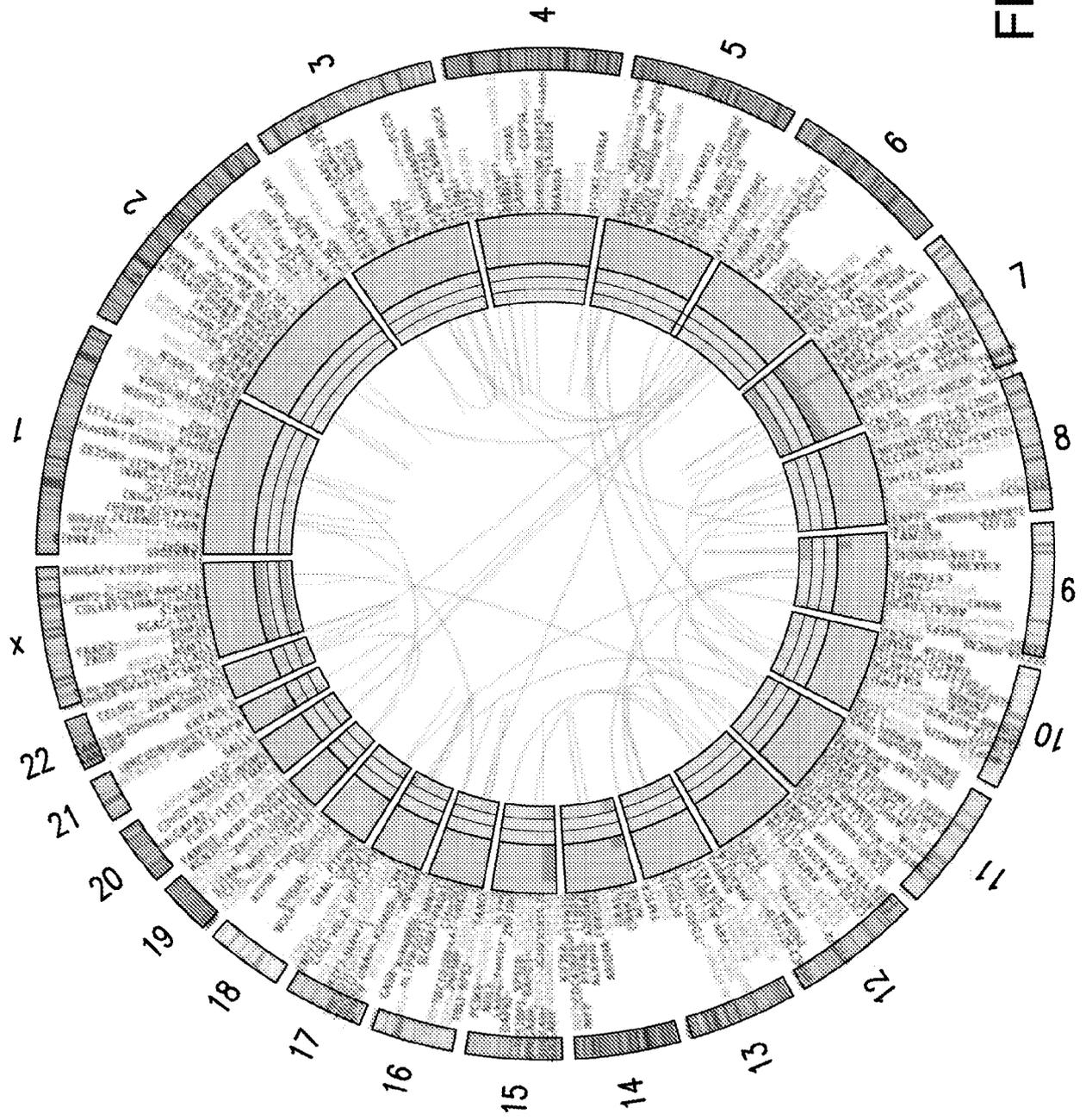


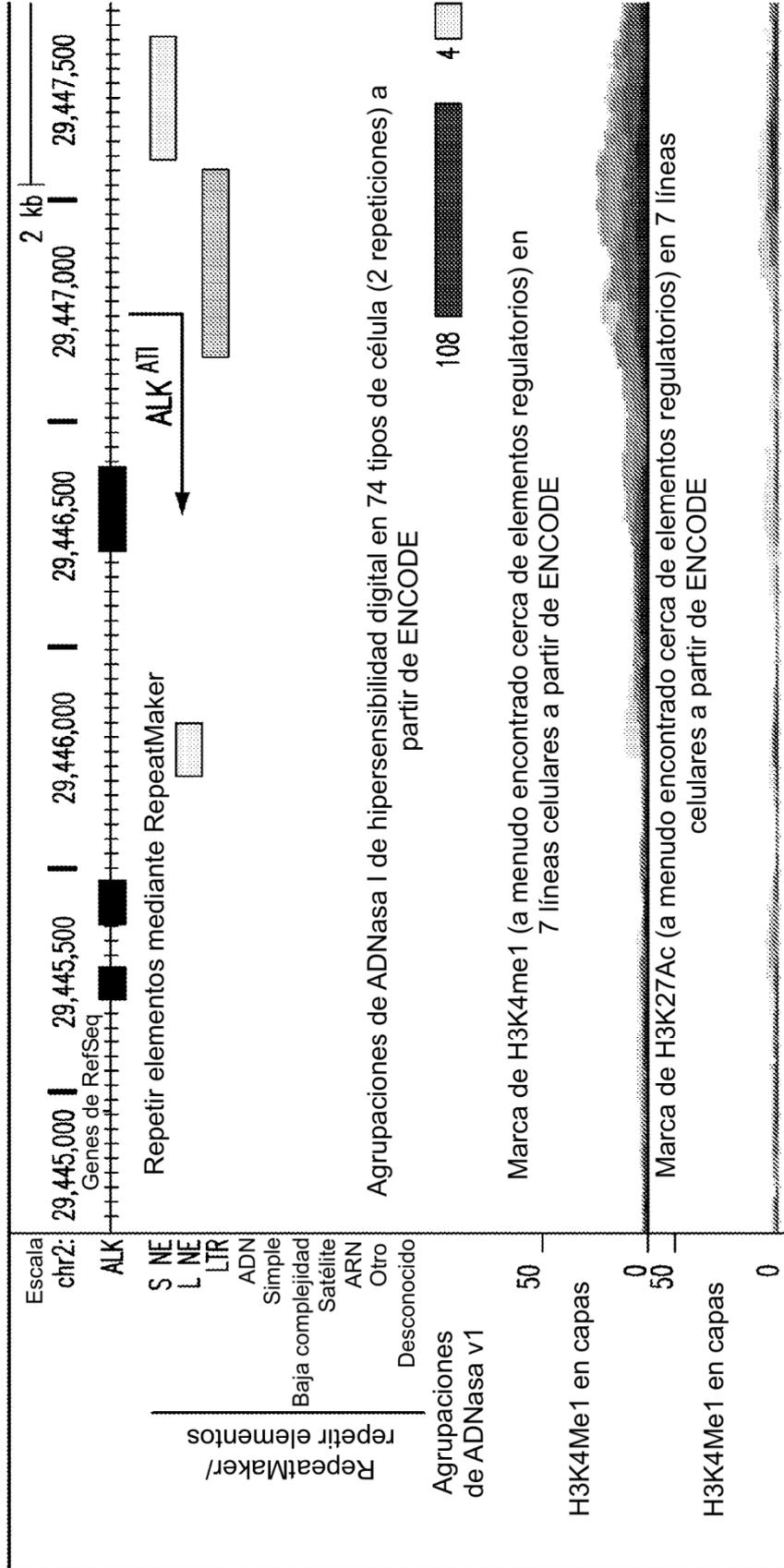
FIG. 10C

n.º	Posición inicial	Posición final	Alelo de referencia	Alelo de tumor	Frecuencia	Localización
1	29,420,634	29,420,634	T	-	0,37	Intrón
2	29,460,522	29,460,522	G	A	0,43	Intrón
3	29,470,032	29,470,032	T	-	0,39	Intrón
4	29,492,095	29,492,095	G	A	0,26	Intrón
5	29,499,955	29,499,955	C	T	0,45	Intrón
6	29,519,637	29,519,637	G	A	0,44	Intrón
7	29,537,765	29,537,765	C	A	0,74	Intrón
8	29,566,924	29,566,924	C	T	0,57	Intrón
9	29,567,833	29,567,833	G	A	0,46	Intrón
10	29,587,837	29,587,837	G	C	0,47	Intrón
11	29,604,104	29,604,104	G	A	0,45	Intrón
12	29,624,468	29,624,468	G	A	0,56	Intrón
13	29,642,007	29,642,007	G	A	0,69	Intrón
14	29,661,584	29,661,584	C	T	0,12	Intrón
15	29,677,601	29,677,601	T	G	0,55	Intrón
16	29,679,303	29,679,303	T	C	0,58	Intrón
17	29,684,682	29,684,682	G	A	0,41	Intrón
18	29,690,777	29,690,778	AT	-	0,33	Intrón
19	29,709,870	29,709,870	T	C	0,41	Intrón
20	29,726,177	29,726,177	T	G	0,57	Intrón
21	29,750,006	29,750,006	T	C	0,38	Intrón
22	29,758,564	29,758,564	C	T	0,44	Intrón
23	29,761,096	29,761,096	G	T	0,42	Intrón
24	29,764,306	29,764,306	A	T	0,28	Intrón
25	29,770,527	29,770,527	C	T	0,26	Intrón
26	29,804,173	29,804,173	G	A	0,74	Intrón

FIG. 10D

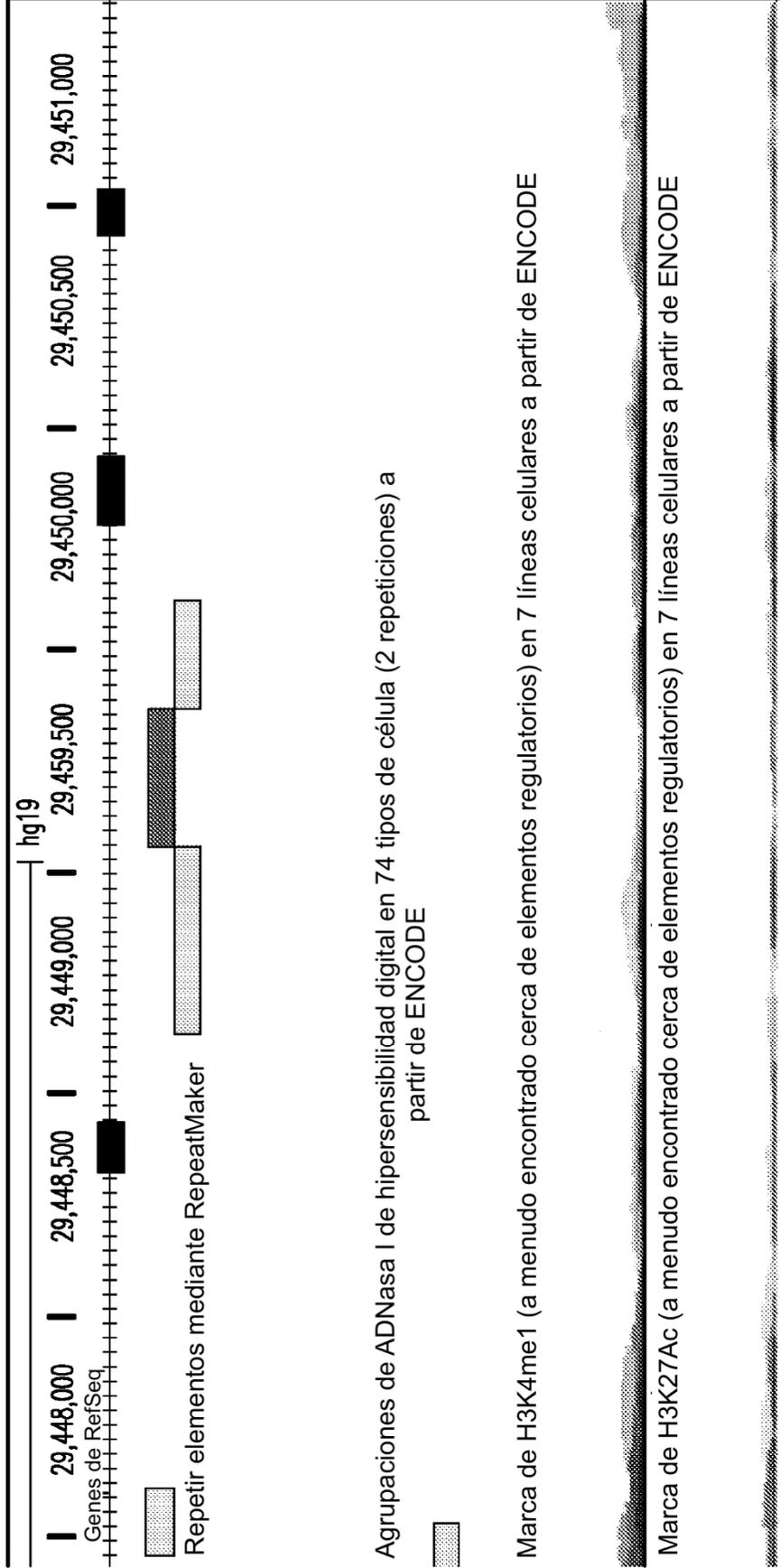
27	29,819,803	29,819,803	X	G	A	0,39	Intrón
28	29,857,339	29,857,340	X	-	A	0,45	Intrón
29	29,863,910	29,863,910	X	A	C	0,49	Intrón
30	29,867,114	29,867,114		C	T	0,1	Intrón
31	29,874,965	29,874,965		C	G	0,42	Intrón
32	29,878,627	29,878,627	X	T	C	0,41	Intrón
33	29,879,612	29,879,612	X	C	-	0,42	Intrón
34	29,879,617	29,879,617	X	T	G	0,46	Intrón
35	29,885,468	29,885,468		G	A	0,27	Intrón
36	29,887,819	29,887,819	X	T	C	0,41	Intrón
37	29,891,658	29,891,658	X	T	C	0,22	Intrón
38	29,897,910	29,897,910	X	G	A	0,59	Intrón
39	29,915,442	29,915,442	X	T	C	0,51	Intrón
40	29,953,528	29,953,528	X	T	C	0,71	Intrón
41	29,961,164	29,961,164	X	T	C	0,27	Intrón
42	29,972,014	29,972,014	X	C	T	0,46	Intrón
43	29,978,898	29,978,898	X	G	A	0,42	Intrón
44	29,985,171	29,985,171	X	C	T	0,41	Intrón
45	29,994,457	29,994,457	X	G	A	0,5	Intrón
46	30,004,526	30,004,527		TG	-	0,61	Intrón
47	30,033,145	30,033,145	X	G	A	0,48	Intrón
48	30,044,269	30,044,269	X	C	T	0,42	Intrón
49	30,044,579	30,044,579	X	T	G	0,47	Intrón
50	30,070,597	30,070,597	X	T	C	0,61	Intrón
51	30,075,004	30,075,004	X	G	A	0,4	Intrón
52	30,096,355	30,096,355		C	T	0,46	Intrón
53	30,100,743	30,100,743	X	T	G	0,54	Intrón
54	30,102,577	30,102,577	X	G	A	0,47	Intrón
55	30,112,153	30,112,153	X	C	G	0,41	Intrón
56	30,146,896	30,146,896	X	T	C	0,41	Fianqueo en <sup>5</sup>

FIG. 10D (continuación)



A: Fig. 11A (continuación)

FIG. 11A



Agrupaciones de ADNasa I de hipersensibilidad digital en 74 tipos de célula (2 repeticiones) a partir de ENCODE

Marca de H3K4me1 (a menudo encontrado cerca de elementos regulatorios) en 7 líneas celulares a partir de ENCODE

Marca de H3K27Ac (a menudo encontrado cerca de elementos regulatorios) en 7 líneas celulares a partir de ENCODE

De: FIG. 11A

**FIG. 11A (Continuación)**

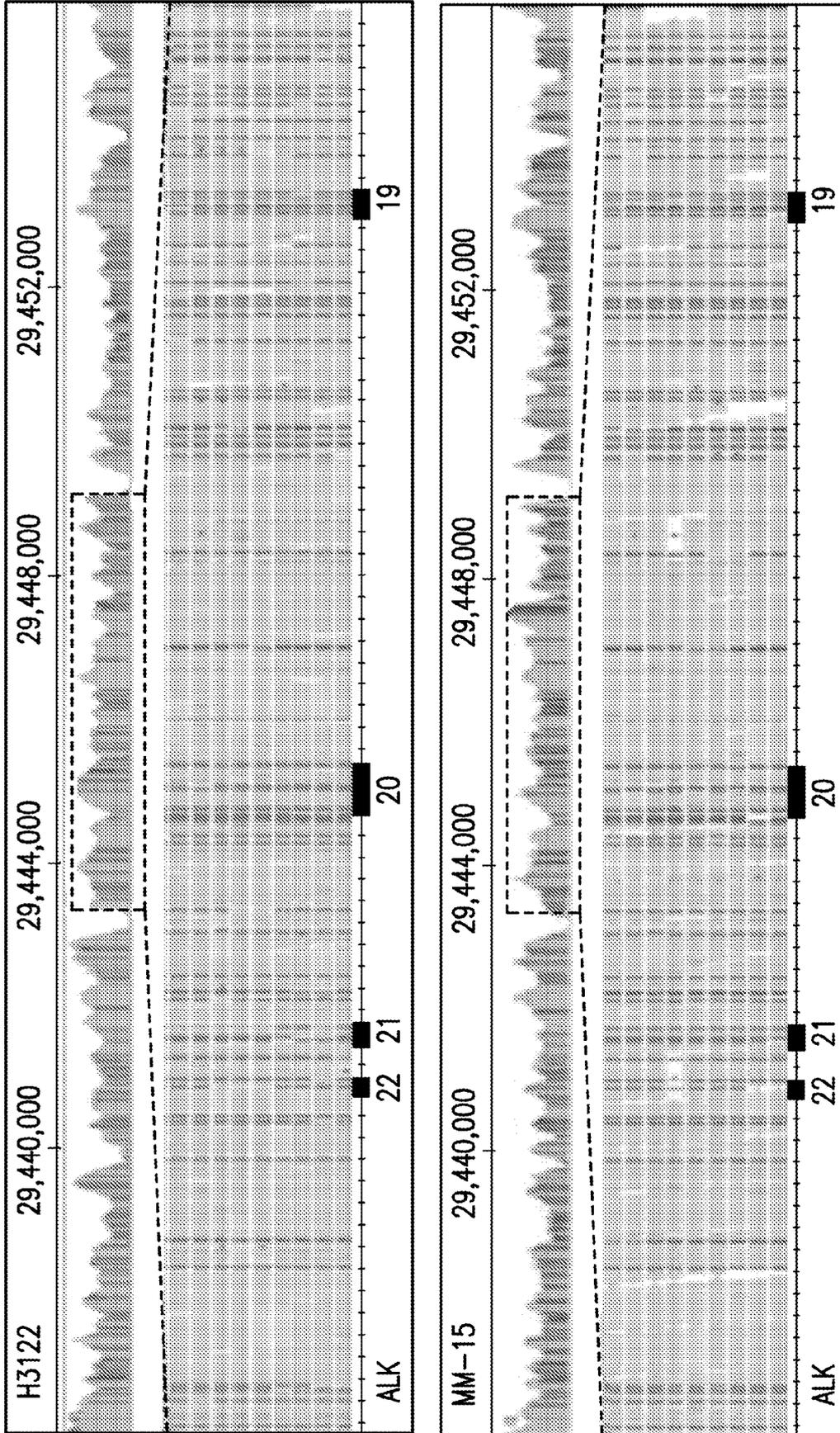


FIG. 11B

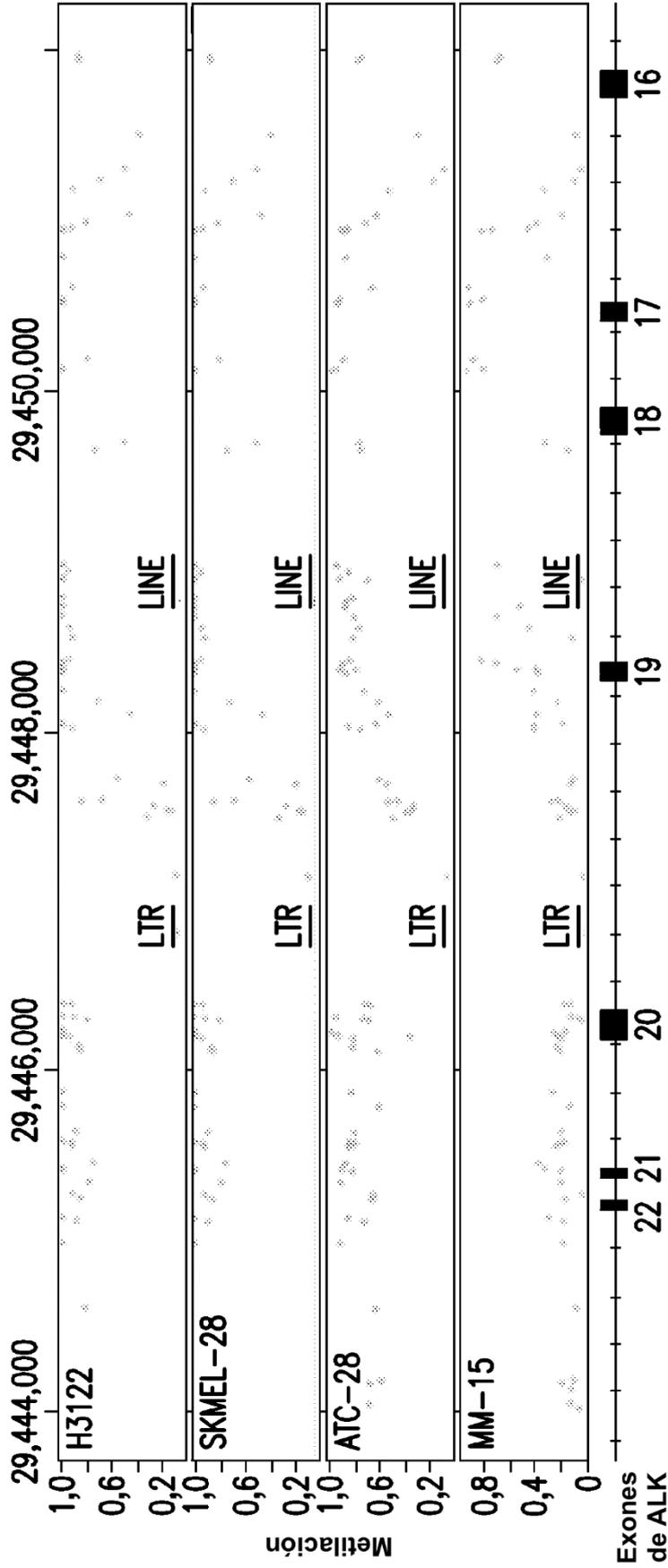


FIG. 11C

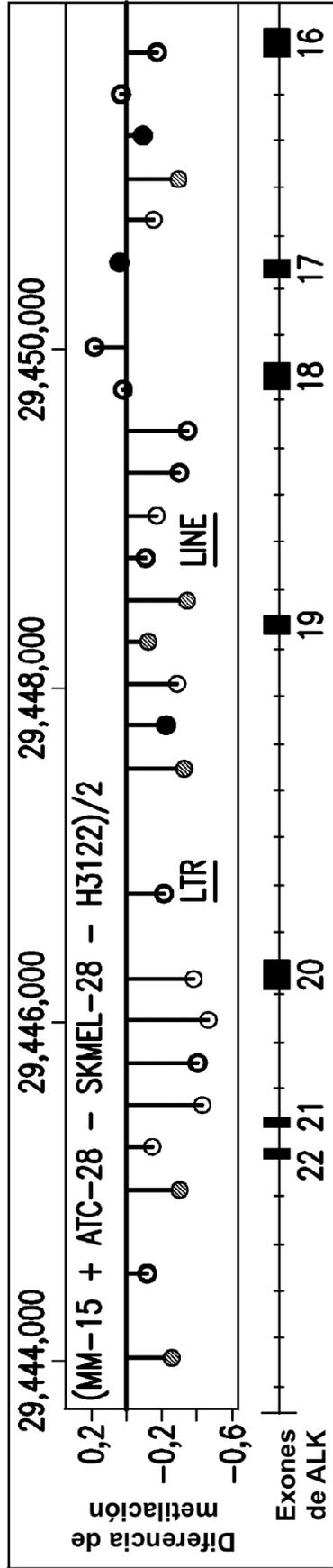


FIG. 11D

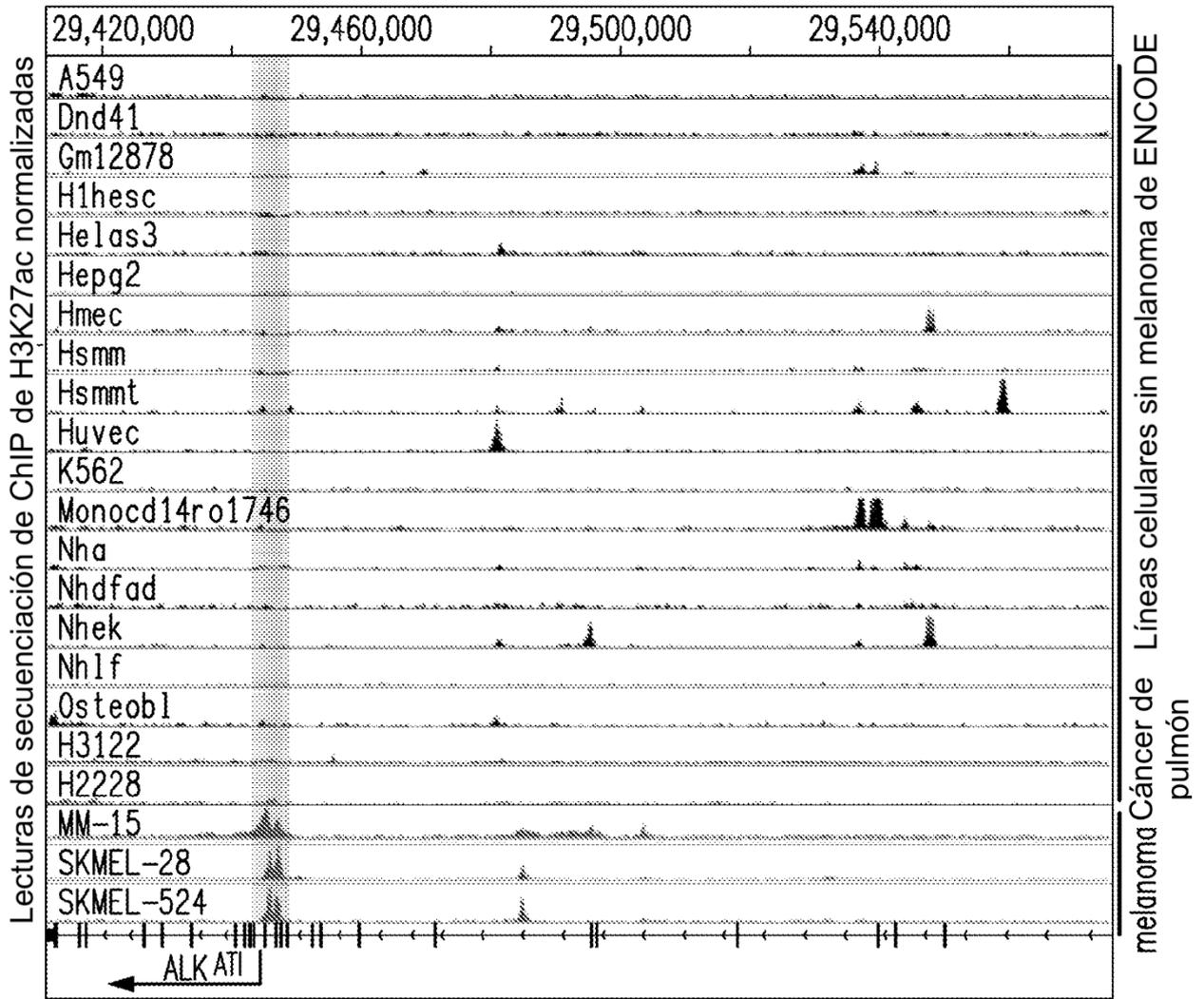


FIG. 11E



Factor de transcripción	Posición de inicio de motivo a partir de Chr2:29,445,000	Posición de detención de motivo a partir de Chr2:29,445,000	Puntuación	Valor de p	Valor de q	Secuencia coincidente
AR	1339	1325	11.948	4,47E-05	0,18	GAGTACAAGCTGAGC
Bach1::Mafk	1653	1639	10.8694	1,86E-05	0,073	AACATGACTCTGCC
Bach1::Mafk	1653	1639	10.8694	1,86E-05	0,073	AACATGACTCTGCC
Bach1::Mafk	521	535	8.6477	4,32E-05	0,085	ATAATGAGGCAGCTG
Bach1::Mafk	521	535	8.6477	4,32E-05	0,085	ATAATGAGGCAGCTG
BHLHE40	1800	1810	11.3822	7,93E-05	0,325	CCCACGTGAAC
BRCA1	110	116	9.01131	6,51E-05	0,258	ACAACAC
CTCF	1250	1232	13.6407	7,73E-06	0,031	TGACCTGAAGGAGGTGCCG
CTCF	91	73	9.20027	6,33E-05	0,127	TCTCCAGTAGGAGACTCCC
Ddit3::Cebpa	1988	1999	12.5036	2,87E-05	0,12	CAGTGCAATCAC
Ddit3::Cebpa	1988	1999	12.5036	2,87E-05	0,12	CAGTGCAATCAC
Ddit3::Cebpa	1988	1999	12.5036	2,87E-05	0,12	CAGTGCAATCAC
Ddit3::Cebpa	629	640	11.3202	6,55E-05	0,137	CTATGCAATCTC
Ddit3::Cebpa	629	640	11.3202	6,55E-05	0,137	CTATGCAATCTC
Ddit3::Cebpa	629	640	11.3202	6,55E-05	0,137	CTATGCAATCTC
E2F3	1825	1811	15.6755	3,89E-06	0,0142	TTCCCTCCCTCCCTC
E2F3	1821	1807	14.0027	1,14E-05	0,0208	CTCCCTCCCTCGTTC
E2F3	933	919	12.1626	2,84E-05	0,0344	TTCCCTCCAGGGGAG
E2F3	1817	1803	10.9916	4,48E-05	0,0408	CTCCCTCGTTCACGT
E2F3	1754	1740	10.3225	5,62E-05	0,0409	CTCCCTCCCCATTTT
E2F6	1817	1827	15.5096	3,49E-06	0,0105	GGGAGGGAAGG
E2F6	1813	1823	15.1708	5,28E-06	0,0105	GGGAGGGAAGG
E2F6	925	935	13.6464	1,82E-05	0,0208	TGGAGGGAAGG
E2F6	1746	1756	13.477	2,08E-05	0,0208	GGGAGGGAAGG

FIG. 11H

E2F6	1809	1819	11.9525	5,00E-05	0,0399	ACGAGGGAGGG
EBF1	931	921	13.3621	2,20E-05	0,081	CCCTCCAGGGG
EBF1	1495	1505	11.7176	5,46E-05	0,081	CTCCCCTGAGC
EBF1	816	806	11.512	6,07E-05	0,081	CTCCCCAAGGA
EBF1	1739	1729	10.4842	9,12E-05	0,0913	CTCTCATGGC
EGR1	312	299	15.5788	4,20E-06	0,0169	CCTCCTCCTCCTCC
EGR1	315	302	13.097	1,97E-05	0,0396	CAGCCTCCTCCTCC
EGR1	801	788	11.1879	5,35E-05	0,0717	GAACCTCCCCCGGG
EGR1	297	284	10.2334	8,31E-05	0,0728	CACCTCCCCCTTCT
EGR1	411	398	10.0425	9,05E-05	0,0728	CCCAAGCCCCCTGC
EGR2	312	298	10.2839	7,97E-05	0,315	CCTCCTCCTCCTCCC
EHF	873	866	14.3145	2,41E-05	0,0972	CCTTCCTC
ELK1	1391	1382	12.6123	1,09E-05	0,0454	CCGCCGGAAG
ELK4	1379	1389	14.1514	1,08E-05	0,0433	GTGCTTCCGGC
Erg	1736	1746	13.8707	1,48E-05	0,0606	AGAGGAAATGG
Erg	1736	1746	13.8707	1,48E-05	0,0606	AGAGGAAATGG
Erg	1736	1746	13.8707	1,48E-05	0,0606	AGAGGAAATGG
Erg	865	875	11.9167	5,42E-05	0,1	AGAGGAAGGCG
Erg	1467	1477	11.3306	7,32E-05	0,1	ACAGGAAGCCT
Erg	865	875	11.9167	5,42E-05	0,1	AGAGGAAGGCG
Erg	1467	1477	11.3306	7,32E-05	0,1	ACAGGAAGCCT
Erg	865	875	11.9167	5,42E-05	0,1	AGAGGAAGGCG
Erg	1467	1477	11.3306	7,32E-05	0,1	ACAGGAAGCCT
Erg	865	875	11.9167	5,42E-05	0,1	AGAGGAAGGCG
Erg	1467	1477	11.3306	7,32E-05	0,1	ACAGGAAGCCT
ESR1	1421	1440	11.205	4,17E-05	0,119	ATGGCCTGGCAGCCTGGCCC

**FIG. 11H (Continuación)**

ESR1	449	468	10.5073	6,31E-05	0,119	CCCCAAGGCGCCATGGCCC
ESR1	1780	1761	9.80968	9,43E-05	0,119	CTGCTGTTTACCCAGCCTT
ESR2	183	197	10.5306	6,49E-05	0,265	GGTGAGGGTGTCTC
ESRRA	350	360	13.5237	1,38E-05	0,0552	CCAAGGGCAGG
Esrfb	347	358	12.2027	3,32E-05	0,132	AAGCCAAGGGCA
Ets1	1747	1733	14.2216	8,55E-06	0,0347	CCCATTTCCCTCAT
Ets1	876	862	11.7651	4,44E-05	0,0861	GCGCCTTCCCTGTGTC
Ets1	1487	1501	11.11	6,37E-05	0,0861	TCCATATCCTCCCT
FLI1	1736	1746	12.5362	3,03E-05	0,104	AGAGGAAATGG
FLI1	1389	1379	11.6687	5,18E-05	0,104	GCCGGAAGCAC
FLI1	1467	1477	10.9747	7,76E-05	0,104	ACAGGAAGCCT
Foxd3	707	718	13.6322	1,31E-05	0,0525	AAATGTTTATTC
Foxd3	1048	1037	12.8787	2,51E-05	0,0525	GTATGTTGGCTT
FOXH1	2013	2003	11.2367	7,28E-05	0,304	TCAAATCCACT
Foxq1	707	717	14.3755	8,64E-06	0,0361	AAATGTTTATT
Foxq1	1025	1015	12.1488	4,63E-05	0,0967	CATAGTTTATG
GABPA	1388	1378	14.2072	1,10E-05	0,0431	CCGGAAGCAC
GATA2	1159	1146	12.6573	3,33E-05	0,139	TGAATGTTATCACA
Gfi1	1992	2001	11.6665	3,64E-05	0,152	GCAATCACAG
HNF4G	916	902	12.5633	2,18E-05	0,0881	GGGAGCAAAGGGCC
HNF4G	1515	1501	11.0704	5,34E-05	0,108	AAGGTTCAGAGCTCA
HSF1	68	54	13.1771	1,48E-05	0,0577	ATTCTAGACACTTCT
INSM1	396	407	12.5085	2,34E-05	0,0938	TTGCAGGGGGCT
Klf4	292	301	12.5704	2,59E-05	0,1	AGGGTGGGGA
KLF5	301	292	14.1793	1,36E-05	0,0522	TCCCCACCCCT
KLF5	1754	1745	11.1006	7,53E-05	0,0522	CTCCCTCCCC
KLF5	306	297	10.9081	8,33E-05	0,0522	CCTCCTCCCC

FIG. 11H (Continuación)

KLF5	904	895	10.9081	8,33E-05	0,0522	GCCCTCCTC
KLF5	191	182	10.5233	9,29E-05	0,0522	CCCTCACCCC
KLF5	1821	1812	10.5233	9,29E-05	0,0522	CTCCCTCCCT
KLF5	1825	1816	10.5233	9,29E-05	0,0522	TTCCCTCCCT
Meis1	868	854	12.0071	4,41E-05	0,131	CTCTGTCACTCACTG
Meis1	824	810	11.3328	6,42E-05	0,131	ATCTGTCTCTCCCCA
Myb	536	527	10.9878	7,98E-05	0,33	CCAGCTGCCT
MYCN	1800	1807	12.8822	3,18E-05	0,124	CCCACGTG
NFATC2	1619	1613	12.1547	7,93E-05	0,33	TTTTCCA
NFE2::MAF	1650	1636	11.9662	1,81E-05	0,0754	ATGACTCTGCCCTAT
Nfe2l2	1654	1640	12.005	2,53E-05	0,105	TAACATGACTCTGCC
Nfe2l2	1765	1779	9.60768	7,83E-05	0,163	CTGGGTGAACCAGCA
NFKB1	779	789	11.7116	4,50E-05	0,182	TGGCTTTTCCC
NHLH1	527	538	11.731	3,33E-05	0,0565	AGGCAGCTGGGG
NHLH1	884	873	11.3939	3,88E-05	0,0565	GAGCAACTGCCG
NHLH1	2027	2016	11.1691	4,31E-05	0,0565	TCCCAGCTGCAC
Nkx2-5	147	157	12.7707	3,59E-05	0,146	AACCTCTCCAG
Nkx2-5	1564	1570	9.93746	9,66E-05	0,405	TTAATTG
NKX3-1	1522	1528	12.2428	9,66E-05	0,405	ATACTTA
NR1H2::RXRA	1515	1499	0.20754	1,66E-05	0,0692	AAGGTTCAAGAGCTCAGG
NR2C2	164	178	15.4648	3,28E-06	0,013	TGGGGCAGAGGGGA
NR2F1	1501	1514	10.4372	5,44E-05	0,224	TGAGCTCTGAACCT
Nr5a2	346	360	11.4456	4,03E-05	0,163	CAAGCCAAGGGCAGG
Nr5a2	1058	1072	9.49644	9,70E-05	0,196	AATGTCAAGGCTTGT
Pax6	1579	1566	12.0533	2,99E-05	0,0995	ATCATGCTTCAATT
Pax6	1568	1581	11.2937	4,77E-05	0,0995	TTGAAGCATGATTT
PLAG1	1081	1094	13.5435	3,77E-06	0,0143	GAGGGCTAGGGGTG

FIG. 11H (Continuación)

PLAG1	165	178	10.155	1,37E-05	0,0239	GGGGGCAGAGGGGA
PLAG1	1875	1888	9.2837	1,89E-05	0,0239	AAGGCCCTAAGGGG
PLAG1	164	177	4.83031	8,41E-05	0,0796	TGGGGCAGAGGGG
PPARG::RXRA	164	178	14.69	3,84E-06	0,0152	TGGGGCAGAGGGGA
PPARG::RXRA	164	178	14.69	3,84E-06	0,0152	TGGGGCAGAGGGGA
PPARG::RXRA	916	902	13.1479	1,33E-05	0,0227	GGGGAGCAAAGGGCC
PPARG::RXRA	85	99	12.8052	1,72E-05	0,0227	CTGGAGAAAAAGGGGA
PPARG::RXRA	916	902	13.1479	1,33E-05	0,0227	GGGGAGCAAAGGGCC
PPARG::RXRA	85	99	12.8052	1,72E-05	0,0227	CTGGAGAAAAAGGGGA
REL	416	425	12.7526	2,27E-05	0,095	TGGGCATTCC
RELA	242	233	11.0708	5,10E-05	0,125	CTGGATTCC
RELA	416	425	10.2963	6,12E-05	0,125	TGGGCATTCC
RFX5	1857	1843	12.8927	2,23E-05	0,089	CTCCATGGCACCCAG
RREB1	190	171	4.46988	5,78E-05	0,152	CCTCACCCCAACTCCCCTCT
RREB1	193	174	3.48298	7,81E-05	0,152	CACCTCACCCCAACTCCCC
RUNX1	197	207	13.6666	8,73E-06	0,0176	CTCTGTGGCTT
RUNX1	2076	2066	13.6666	8,73E-06	0,0176	CTCTGTGGCTT
RUNX1	1119	1129	11.3637	6,81E-05	0,0916	TTTTGTGGCTA
RUNX2	194	208	13.094	2,25E-05	0,0918	TCTCTGTGGCTTT
RUNX2	2079	2065	11.2266	6,84E-05	0,101	CAACTCTGTGGCTTG
RUNX2	1116	1130	11.0709	7,41E-05	0,101	GCCTTTTGTGGCTAG
Rxra	350	360	11.1142	7,78E-05	0,19	CCAAGGGCAGG
Rxra	1443	1433	10.7563	9,48E-05	0,19	CAAGGGCCAGG
RXRA::VDR	1022	1008	11.7512	8,50E-06	0,0355	AGTTTATGGAGTTTA
SMAD2::SMAD3 ::SMAD4	822	810	11.2364	5,20E-05	0,108	CTGTCTCTCCCCA

FIG. 11H (Continuación)

SMAD2::SMAD3 ::SMAD4	1758	1746	11.0832	5,66E-05	0,108	CTGGCTCCCTCCC
SMAD2::SMAD3 ::SMAD4	822	810	11.2364	5,20E-05	0,108	CTGTCTCTCCCCA
SMAD2::SMAD3 ::SMAD4	1758	1746	11.0832	5,66E-05	0,108	CTGGCTCCCTCCC
SMAD2::SMAD3 ::SMAD4	822	810	11.2364	5,20E-05	0,108	CTGTCTCTCCCCA
SMAD2::SMAD3 ::SMAD4	1758	1746	11.0832	5,66E-05	0,108	CTGGCTCCCTCCC
Sox17	705	697	11.9846	4,55E-05	0,191	ACCATTGTC
Sox2	704	697	13.0228	6,51E-05	0,272	CCATTGTC
SP2	587	573	14.5465	6,55E-06	0,0145	CCCCGGCCTCTCGTG
SP2	312	298	13.735	1,12E-05	0,0145	CCTCCTCCTCCTCCC
SP2	904	890	13.735	1,12E-05	0,0145	GCCCCCTCCTGAGC
SP2	301	287	12.9236	1,84E-05	0,0179	TCCCCACCCCTCCCCT
SP2	1825	1811	12.4367	2,42E-05	0,0189	TTCCTCCCTCCCTC
SP2	309	295	11.7876	3,43E-05	0,0223	CCTCCTCCTCCCCAC
SP2	1821	1807	10.6515	5,99E-05	0,0333	CTCCCTCCCTCGTTC
SP2	1754	1740	10.327	6,94E-05	0,0334	CTCCCTCCCCATTTC
SP2	306	292	10.0024	8,01E-05	0,0334	CCTCCTCCCCACCCT
SP2	297	283	9.84008	8,59E-05	0,0334	CACCCCTCCCTTCTC
Spi1	1732	1746	12.0562	3,40E-05	0,135	CATGAGAGGAAATGG
Spi1	861	875	10.7681	6,97E-05	0,138	TGACAGAGGAAAGGCG
SPIB	865	871	11.288	6,51E-05	0,136	AGAGGAA
SPIB	1736	1742	11.288	6,51E-05	0,136	AGAGGAA
SREBF1	324	315	11.2408	7,58E-05	0,3	CTCAGCTCAC
SRF	1690	1707	9.62959	8,57E-05	0,357	TTAGACCCAATATGGTCT

**FIG. 11H (Continuación)**

Stat5a.:Stat5b	1926	1936	11.5378	7,54E-05	0,204	ACTCCCAGGAA
Stat5a.:Stat5b	1533	1523	10.9522	9,89E-05	0,204	ATTTCTAAGTA
Stat5a.:Stat5b	1926	1936	11.5378	7,54E-05	0,204	ACTCCCAGGAA
Stat5a.:Stat5b	1533	1523	10.9522	9,89E-05	0,204	ATTTCTAAGTA
Stat5a.:Stat5b	1926	1936	11.5378	7,54E-05	0,204	ACTCCCAGGAA
Stat5a.:Stat5b	1533	1523	10.9522	9,89E-05	0,204	ATTTCTAAGTA
TAL1::GATA1	1154	1137	15.2089	4,32E-06	0,018	GTTATCACAGCACCCGAG
TAL1::GATA1	633	650	9.8736	9,93E-05	0,207	GCAATCTCGGGCTTCCAG
Tcf12	528	538	14.8249	1,10E-05	0,0434	GGCAGCTGGGG
Tcf3	2026	2016	16.394	1,17E-06	0,00455	CCCAGCTGCAC
Tcf3	537	527	15.8065	3,64E-06	0,006	CCCAGCTGCCT
Tcf3	1406	1396	15.4148	4,63E-06	0,006	CCCACCTGCAG
Tcf3	427	437	11.302	8,89E-05	0,0691	GACACCTGGCC
Tcf3	1409	1419	11.302	8,89E-05	0,0691	GTCAGCTGCCA
TEAD1	875	864	8.21141	9,73E-05	0,402	CGCCTTCCTCTG
TFAP2A	799	785	11.5594	4,54E-05	0,17	ACCTCCCCCGGGGAA
TFAP2A	1107	1093	10.3116	8,93E-05	0,17	CCCTCCCTATGGGCA
TFAP2C	799	785	12.3611	3,43E-05	0,0663	ACCTCCCCCGGGGAA
TFAP2C	934	920	12.3611	3,43E-05	0,0663	CTTCCCTCCAGGGGA
THAP1	283	275	10.2961	5,25E-05	0,162	CTGCCCAGA
THAP1	172	164	9.90304	7,89E-05	0,162	CTGCCCCCA
TLX1::NFIC	2066	2053	8.9703	2,09E-05	0,0825	TGGCAGTTCTCCAA
TLX1::NFIC	2090	2077	5.9266	6,13E-05	0,121	TGGCTCTTCTCCAA
TP53	1072	1058	8.41515	8,75E-05	0,365	ACAAGCCTTGACATT
USF1	436	426	10.3469	9,56E-05	0,399	GCCAGGTGTCC
YY1	616	627	11.4769	4,22E-05	0,125	CAACATGGCTTT
YY1	1417	1428	10.7443	5,97E-05	0,125	CAACATGGCCTG

FIG. 11H (Continuación)

ZEB1	653	661	12.939	3,41E-05	0,0974	CTTCACCTG
ZEB1	1407	1399	12.4199	4,84E-05	0,0974	ACCCACCTG
znf143	1774	1755	14.312	6,38E-06	0,0248	GTTACCCAGCCTTCCCTGG
znf143	304	285	10.8665	3,08E-05	0,0566	TCCTCCACCCCTCCCTTC
znf143	1852	1833	9.98301	4,38E-05	0,0566	TGCACCCAGGGTGCTTCCA
znf143	908	927	8.74616	7,03E-05	0,0623	TTGCTCCCTGCTCCCTGG
znf143	1765	1746	8.39278	8,03E-05	0,0623	GCCTTCCCTGGCTCCCTCCC
ZNF263	287	307	22.9123	6,75E-09	2,70E-05	AGGGAGGGTGGGAGGAGGA
ZNF263	299	319	19.6374	1,25E-07	0,00025	GGAGGAGGAGGAGGCTGTGAG
ZNF263	290	310	18.0963	4,13E-07	0,000551	GGAGGGTGGGAGGAGGAGGA
ZNF263	282	302	17.3258	7,24E-07	0,00058	AGAGAAAGGGAGGGTGGGGAG
ZNF263	1811	1831	17.3258	7,24E-07	0,00058	GAGGGAGGAGGGAAAGTTGG
ZNF263	291	311	14.0509	6,31E-06	0,00421	GAGGGTGGGAGGAGGAGGAG
ZNF263	1744	1764	13.473	8,94E-06	0,00511	TGGGAGGGAGCCAGGGAAGG
ZNF263	293	313	13.0877	1,12E-05	0,00558	GGGTGGGAGGAGGAGGAGGGC
ZNF263	1815	1835	12.8951	1,25E-05	0,00558	GAGGAGGGAAAGTTGGGTGG
ZNF263	166	186	12.1245	1,94E-05	0,00707	GGGCAGAGGGAGTTGGGGT
ZNF263	1185	1165	12.1245	1,94E-05	0,00707	GGAGGAGAGGGGTGCAGTGTA
ZNF263	1819	1839	11.7392	2,40E-05	0,00801	GAGGAAAGTTGGGTGGAAGC
ZNF263	165	185	11.5466	2,67E-05	0,00821	GGGGCAGAGGGGAGTTGGGG
ZNF263	922	902	11.1613	3,28E-05	0,00909	GGAGCAGGGGAGCAAAGGGCC
ZNF263	284	304	10.9687	3,63E-05	0,00909	AGAAGGGGAGGGTGGGGAGGA
ZNF263	481	501	10.9687	3,63E-05	0,00909	AAAGGAGAAGACAAGAGGAGA
ZNF263	927	947	10.7761	4,02E-05	0,00947	GAGGAAAGGGGAGCCAGGTAG
ZNF263	1244	1224	10.1981	5,41E-05	0,012	GAAGGAGGTGCCCGGGAAAA
ZNF263	1804	1824	10.0055	5,97E-05	0,0126	CGTGAACGAGGGAGGAGGAGGA
ZNF263	277	297	9.81286	6,58E-05	0,0132	TGGGCAGAGAAGGGGAGGGTGG

FIG. 11H (Continuación)

ZNF263	285	305	9.62022	7,24E-05	0,0138	GAAGGGGAGGGTGGGGAGGAG
ZNF263	281	301	9.23494	8,74E-05	0,0152	CAGAGAAAGGGGAGGGTGGGA
ZNF263	302	322	9.23494	8,74E-05	0,0152	GGAGGAGGAGGCTGTGAGCTG
ZNF263	928	948	9.0423	9,60E-05	0,0154	AGGGAAGGGGAGGCCAGGTAGA
ZNF263	1736	1756	9.0423	9,60E-05	0,0154	AGAGGAAATGGGGAGGGAGCC

**FIG. 11H (Continuación)**

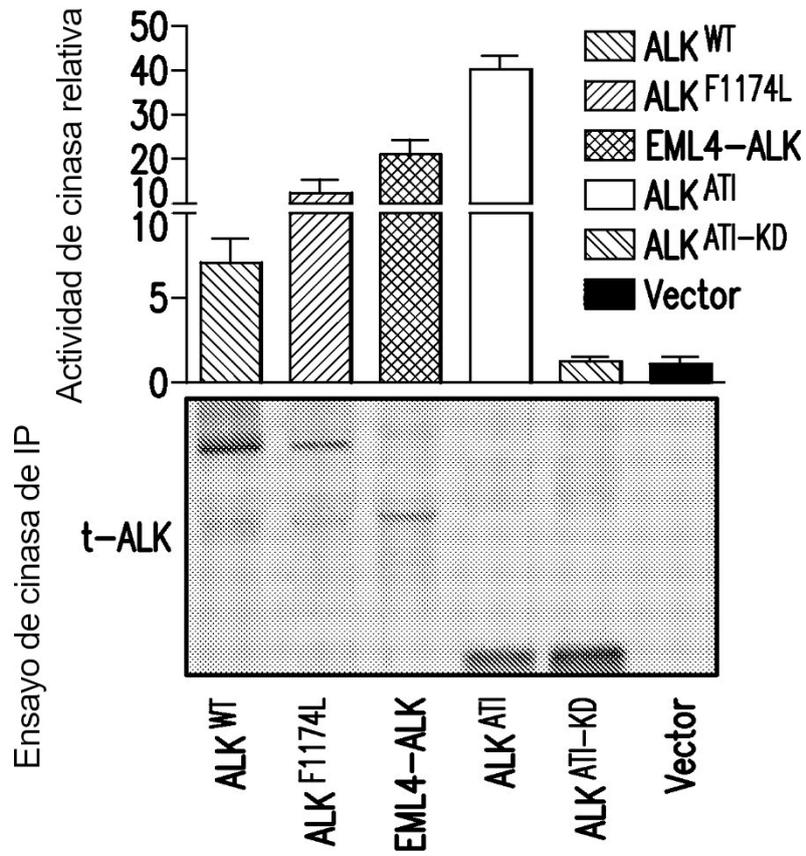


FIG. 12A

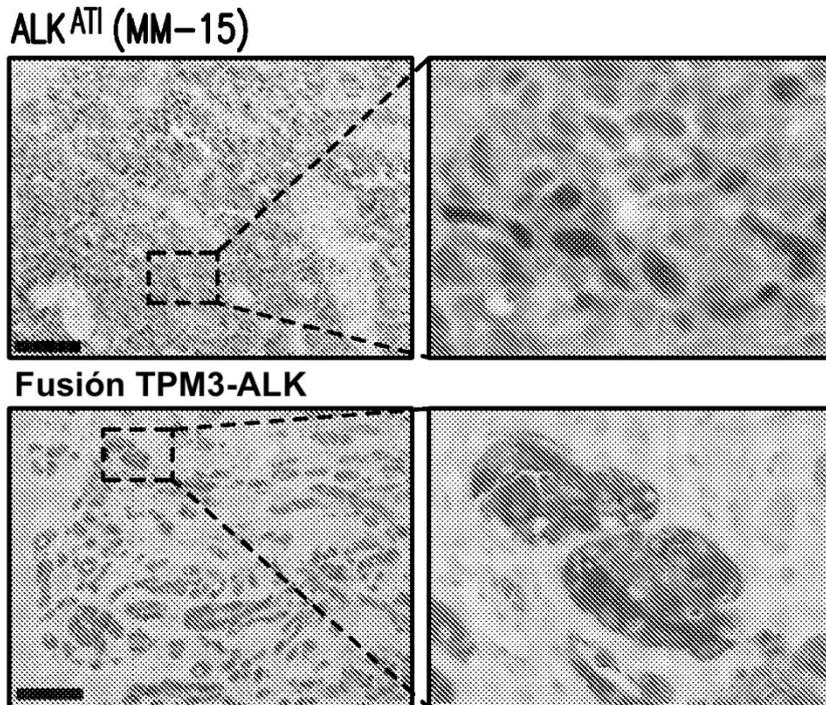


FIG. 12B

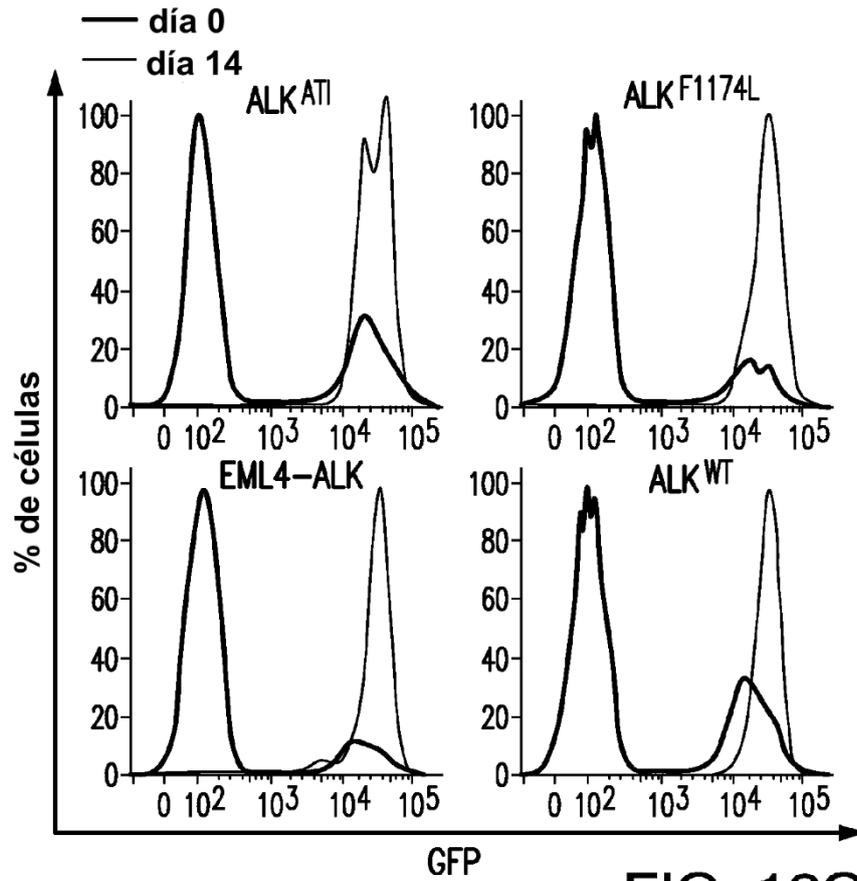


FIG. 12C

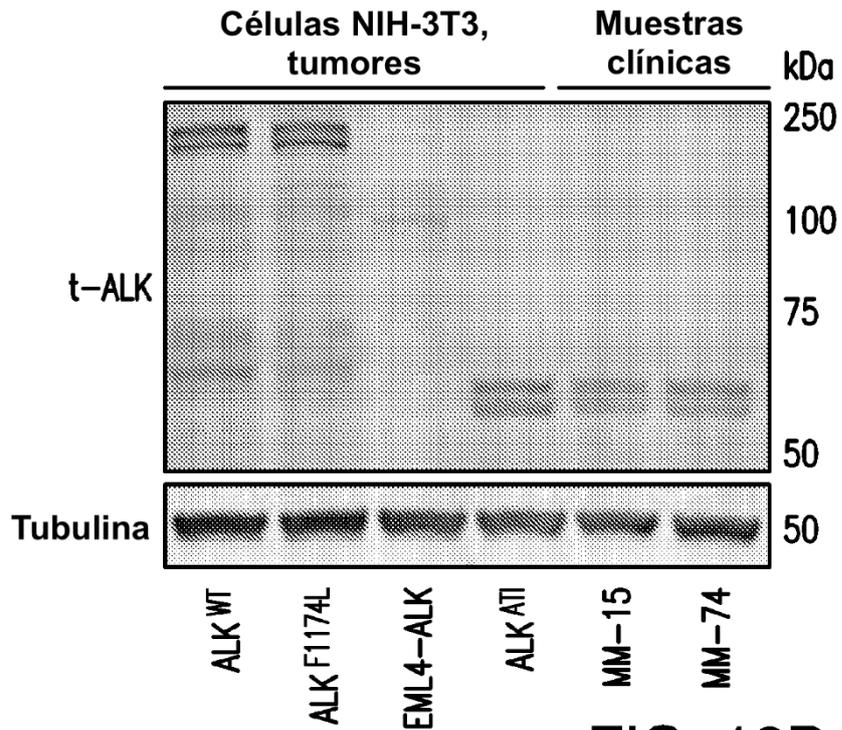


FIG. 12D

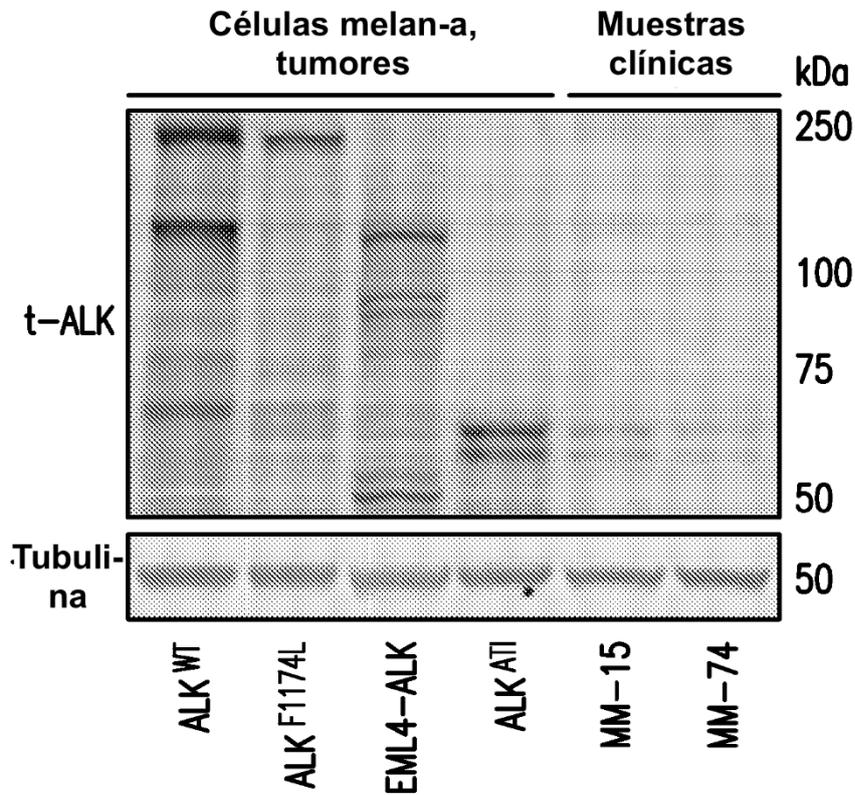
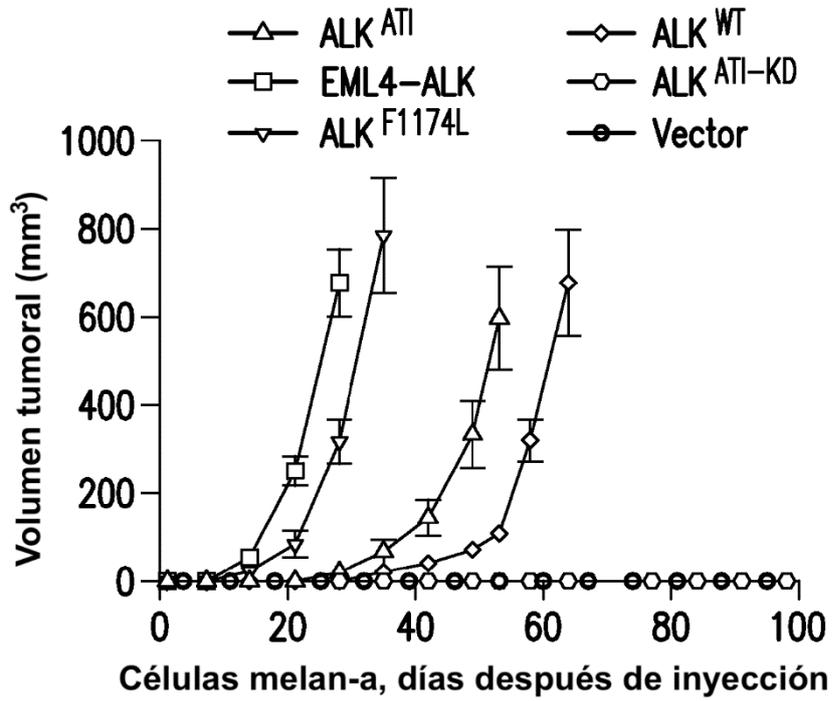


FIG. 12F

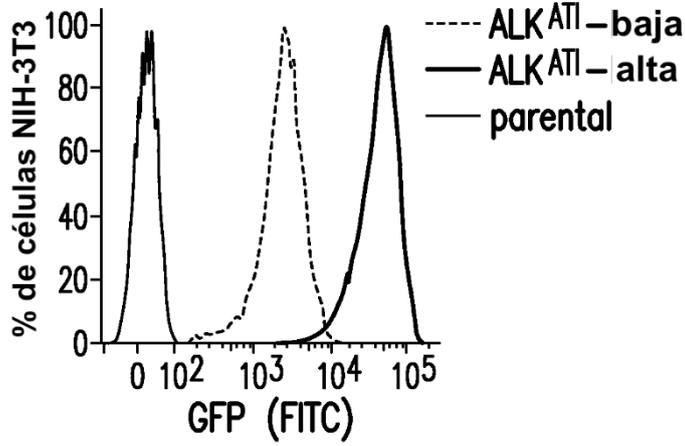


FIG. 12G

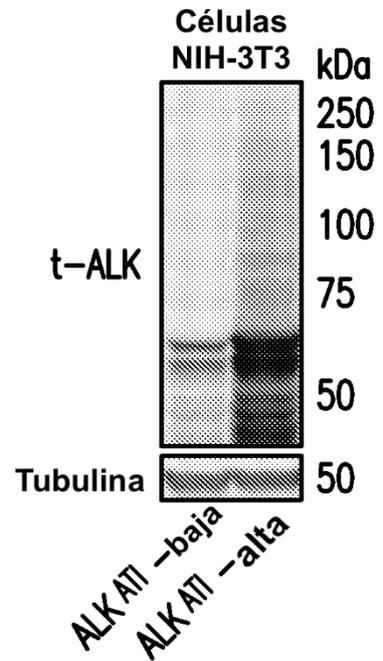


FIG. 12H

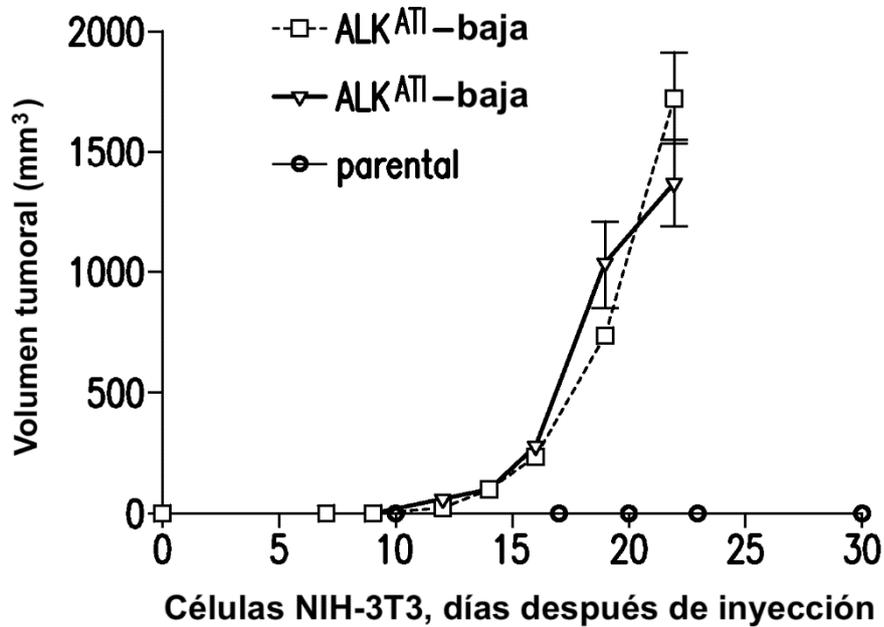


FIG. 12I

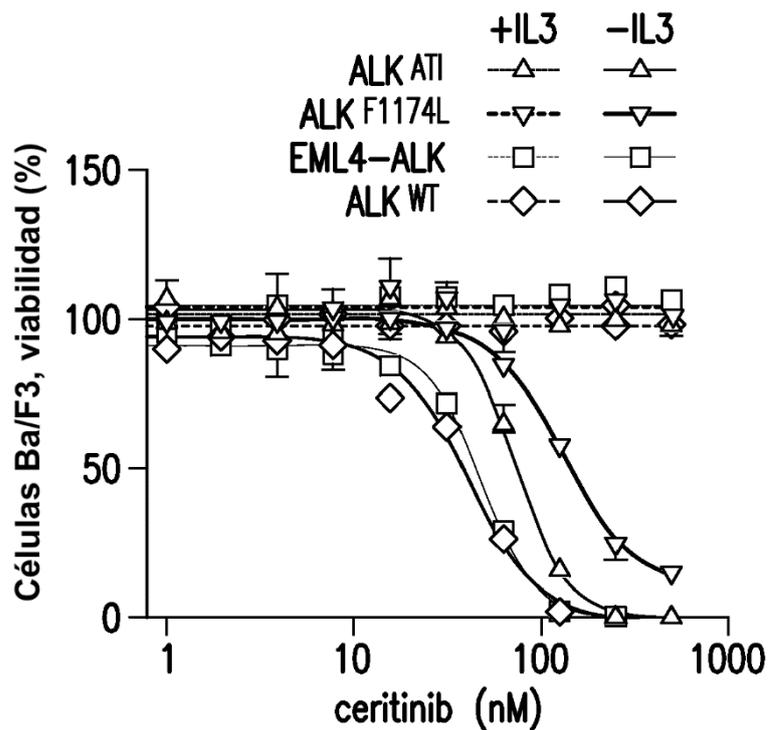


FIG. 13A

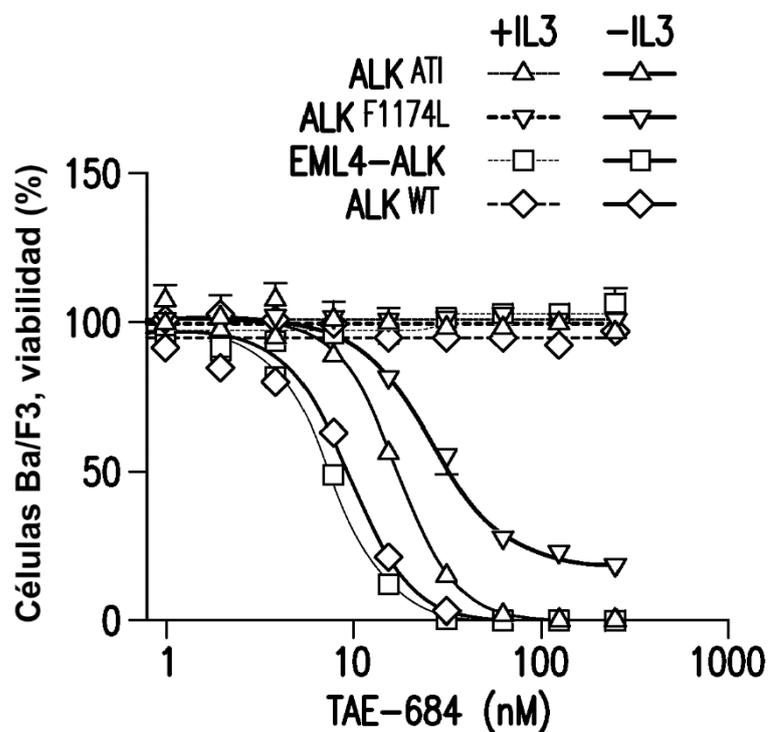
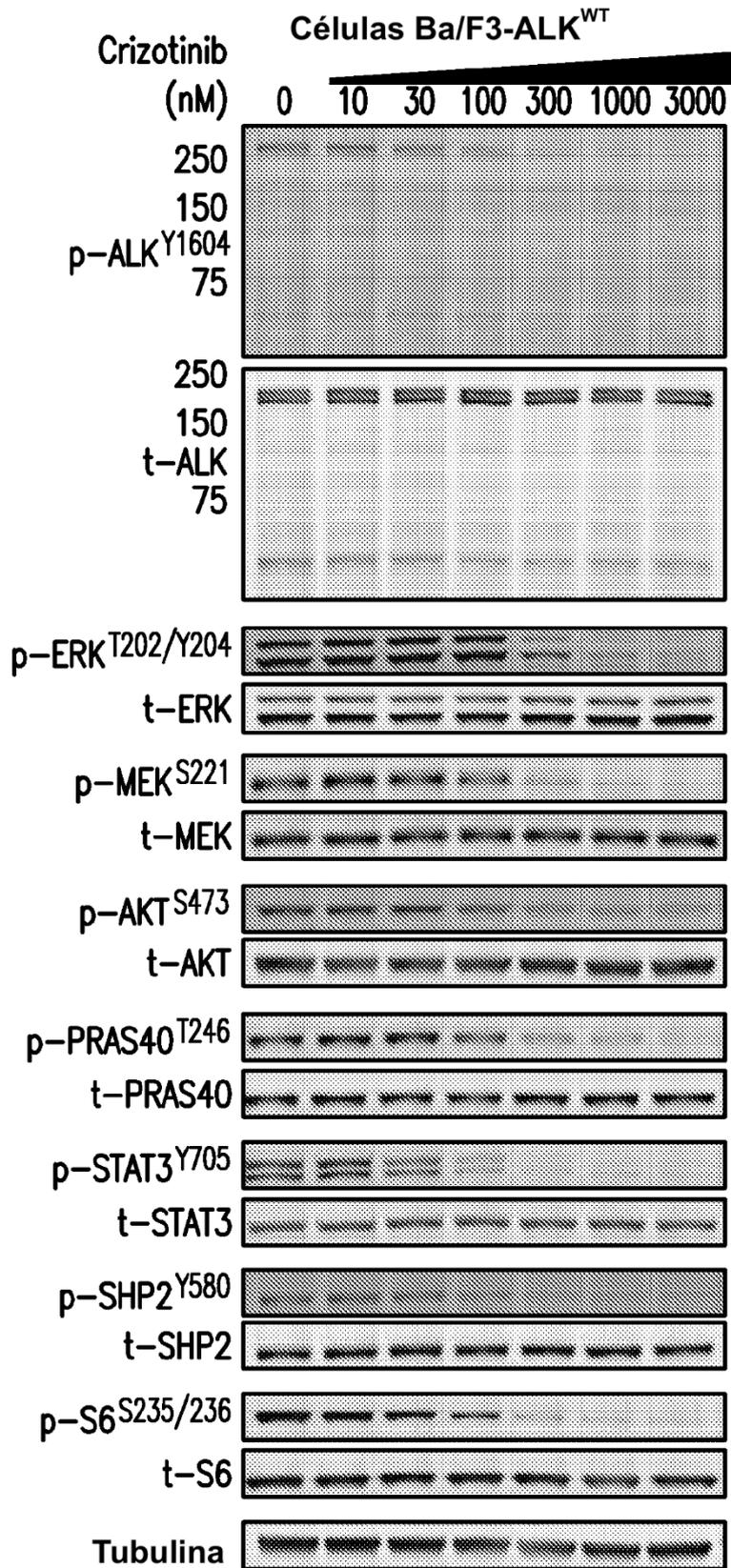
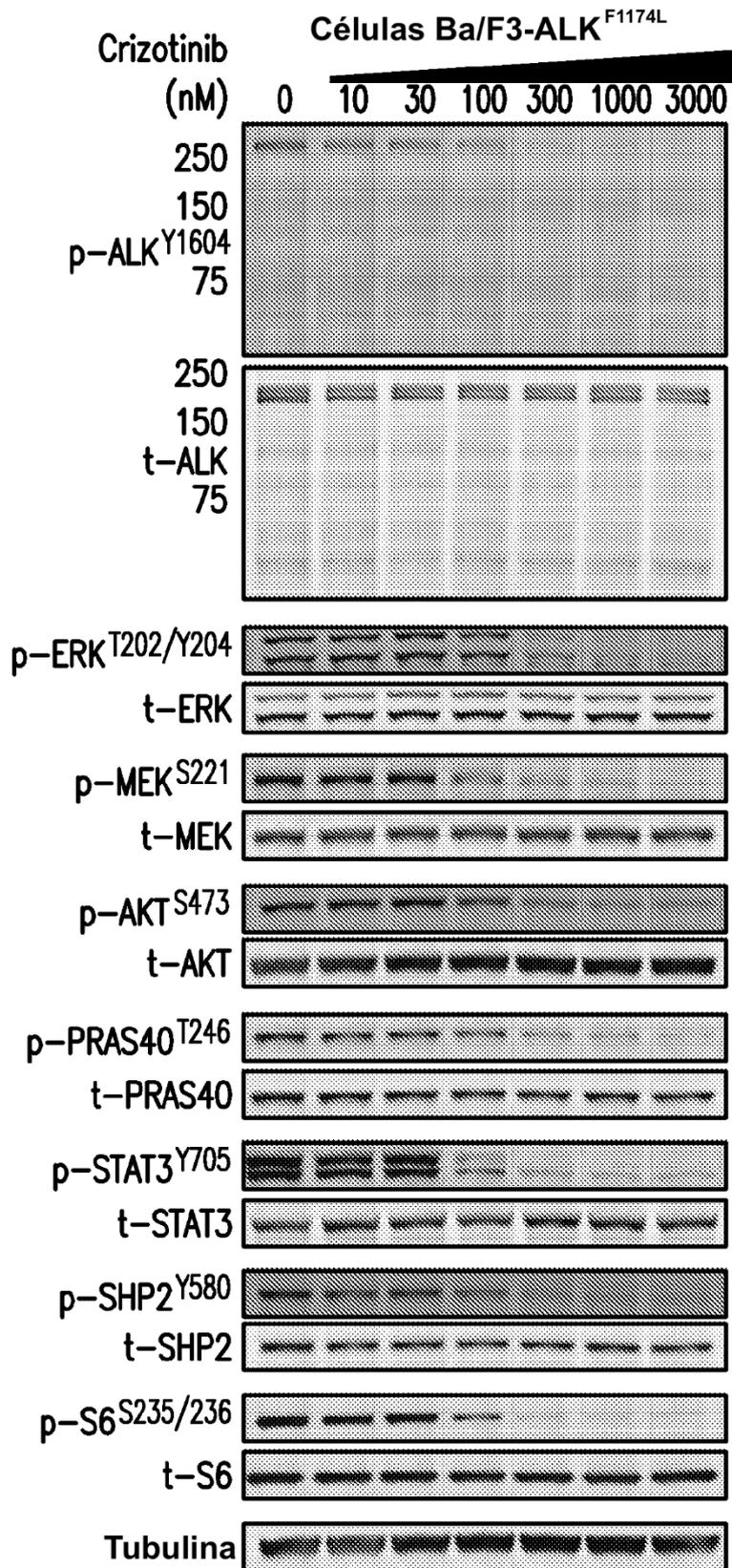


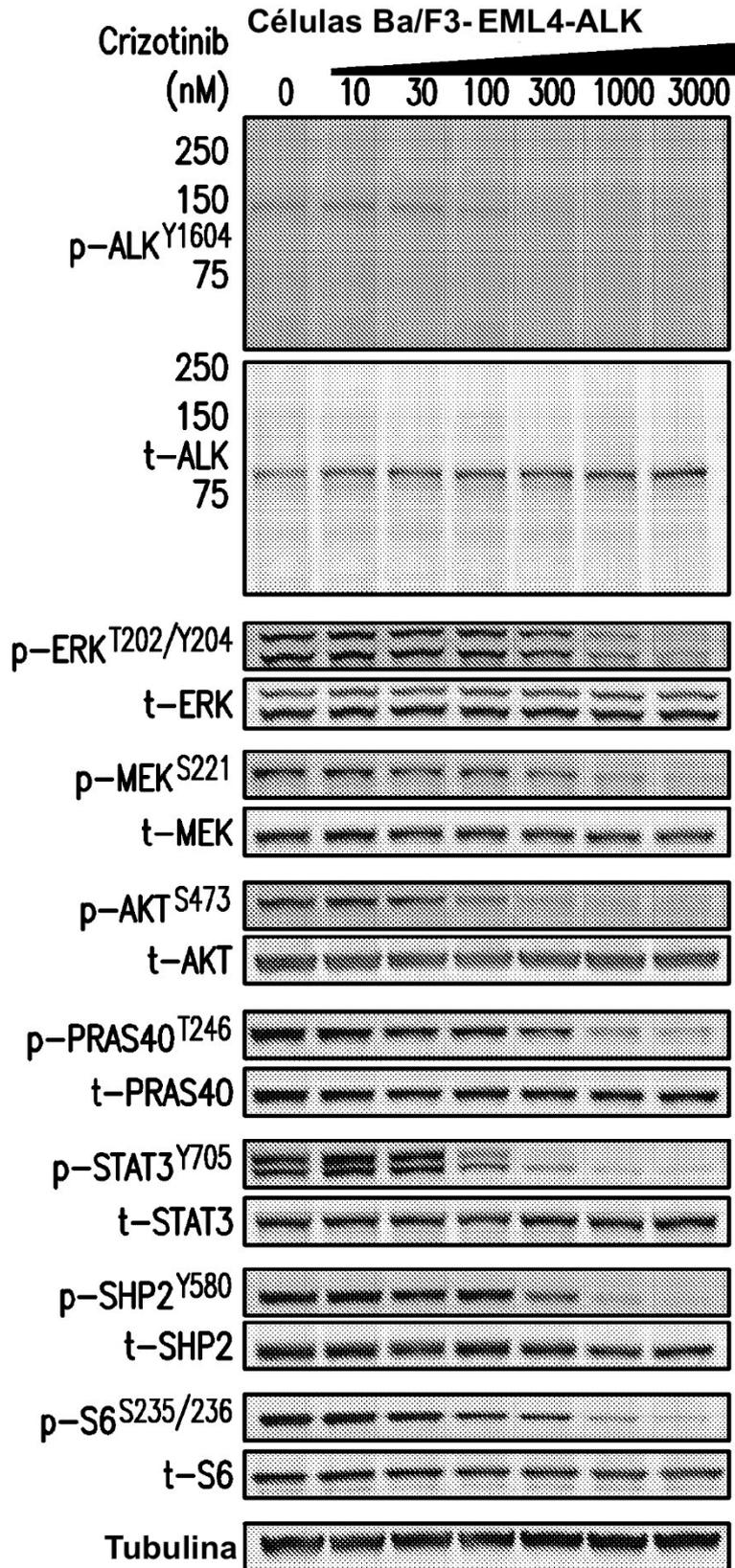
FIG. 13B



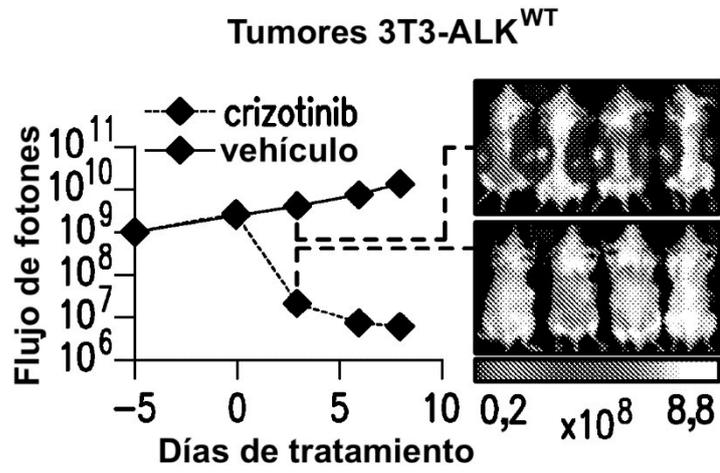
**FIG. 13C**



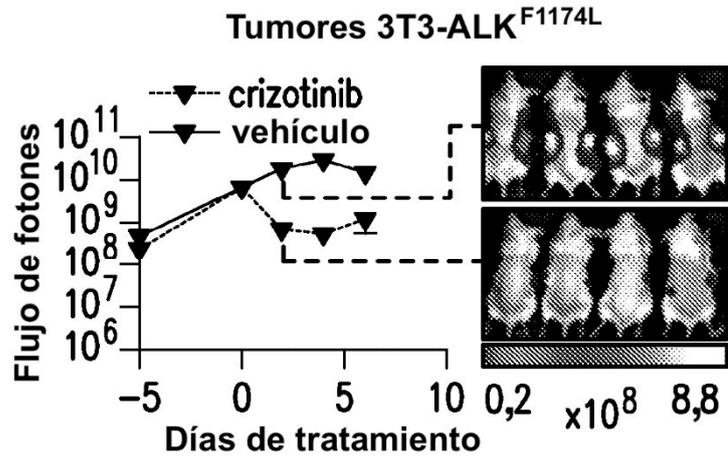
**FIG. 13D**



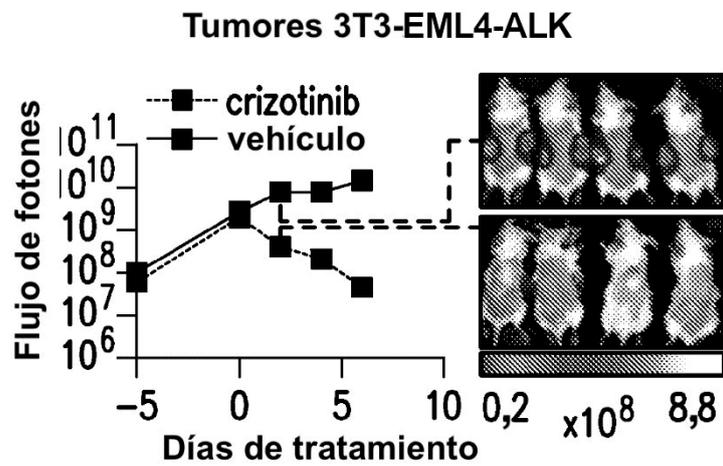
**FIG. 13E**



**FIG. 14A**



**FIG. 14B**



**FIG. 14C**

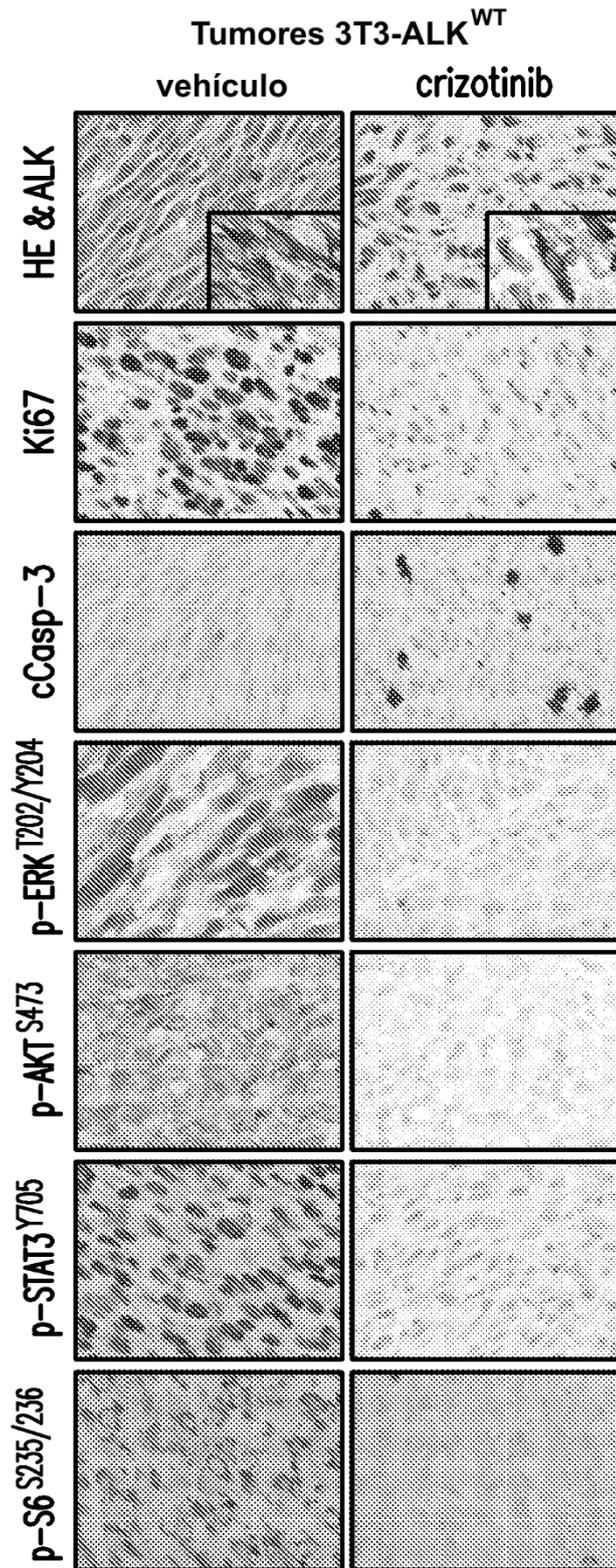
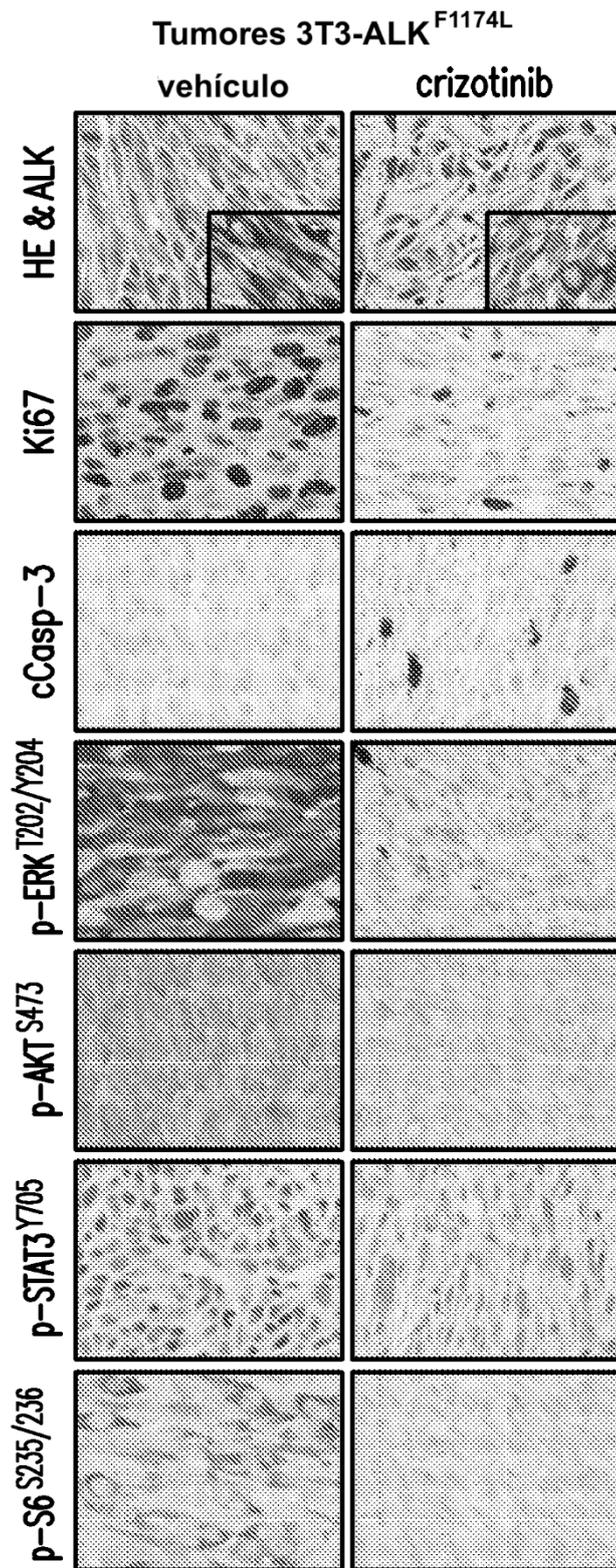


FIG. 14D



**FIG. 14E**

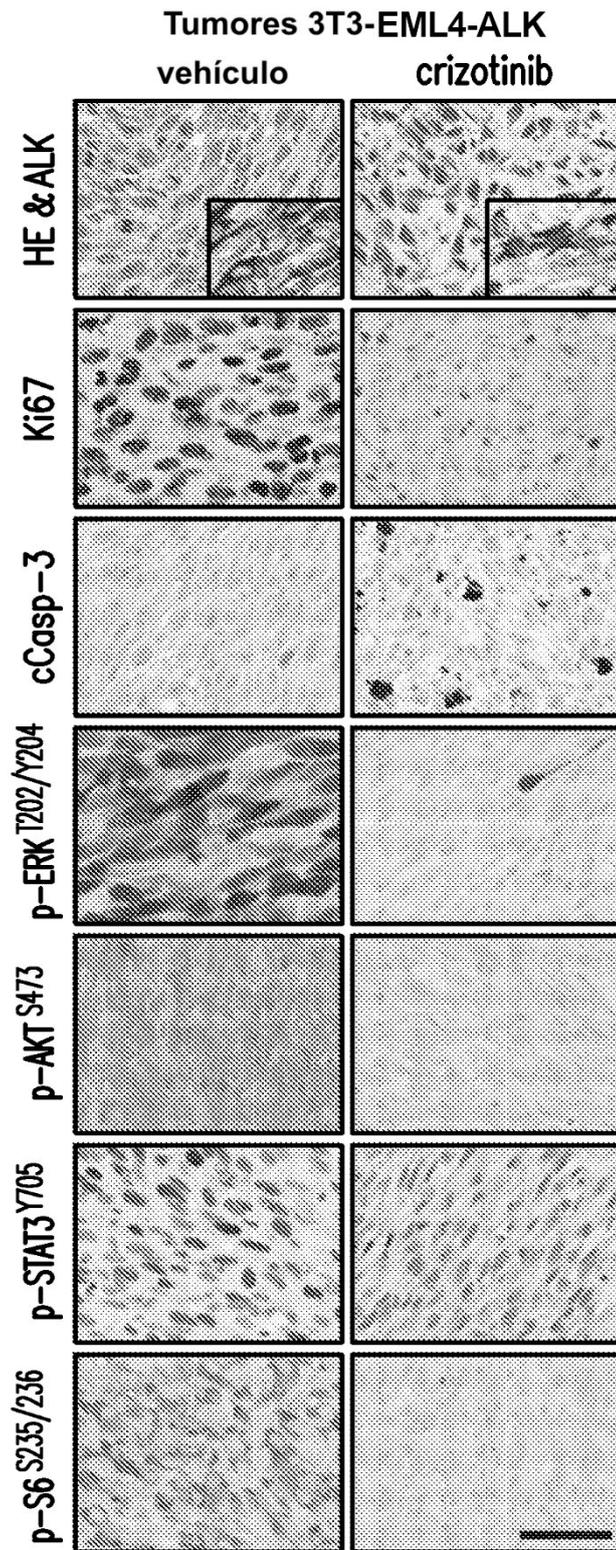


FIG. 14F