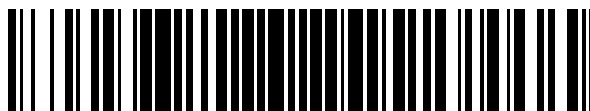


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 390 530**

51 Int. Cl.:
C12N 15/11 (2006.01)
C12Q 1/68 (2006.01)
A61K 48/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09002002 .5**
96 Fecha de presentación: **15.07.1997**
97 Número de publicación de la solicitud: **2113566**
97 Fecha de publicación de la solicitud: **04.11.2009**

54 Título: **Genes procedentes del amplicón 20Q13 y sus usos**

30 Prioridad:
15.07.1996 US 680395
16.10.1996 US 731499
17.01.1997 US 785532

45 Fecha de publicación de la mención BOPI:
13.11.2012

45 Fecha de la publicación del folleto de la patente:
13.11.2012

73 Titular/es:
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (50.0%)
12TH FLOOR 1111 FRANKLIN STREET
OAKLAND, CA 94607-5200, US y
THE HOSPITAL FOR SICK CHILDREN (50.0%)

72 Inventor/es:
GRAY, JOE W.;
COLLINS, COLIN CONRAD;
HWANG, SOO-IN;
GODFREY, TONY;
KOWBEL, DAVID y
ROMMENS, JOHANNA

74 Agente/Representante:
PONTI SALES, Adelaida

ES 2 390 530 T3

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

DESCRIPCIÓN

Genes procedentes del amplión 20Q13 y sus usos

5 REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS

[0001] Esta solicitud es una continuación en parte del documento USSN 08/785.532, presentado el 17 de enero de 1997, del documento USSN 08/731.499 presentado el 16 de octubre de 1996 y del documento USSN 08/680.395 presentado el 15 de julio de 1996, que está relacionada con la Solicitud de patente copendiente de EE.UU., USSN 10 08/546.130, presentada el 20 de octubre de 1995.

ANTECEDENTES DE LA INVENCION

[0002] Esta invención pertenece al campo de la citogenética. Más en particular esta invención está relacionada con la identificación de genes en una región de amplificación en 20q13 aproximadamente en diversos tipos de 15 cáncer. Los genes descritos en el presente documento se pueden utilizar como sondas específicas para el amplión 20q13 así como para el tratamiento de diversos tipos de cáncer.

[0003] Las anomalías cromosómicas a menudo llevan asociados trastornos genéticos, enfermedades 20 degenerativas, y cáncer. En particular, la eliminación o multiplicación de copias de cromosomas completos o segmentos cromosómicos, y niveles superiores de amplificación de regiones específicas del genoma son eventos frecuentes en cáncer. Véase, por ejemplo, Smith, y col., Breast Cancer Res. Treat., 18: Suppl. 1: 5-14 (1991, van de Vijer & Nusse, Biochim. Biophys. Acta. 1072: 33-50 (1991), Sato, y col., Cancer. Res., 50: 7184-7189 (1990). De hecho, la amplificación y eliminación de secuencias de ADN que contienen proto-oncogenes y genes supresores de 25 tumores, respectivamente, habitualmente son características de la tumorigénesis. Dutrillaux, y col., Cancer Genet. Cytogenet., 49: 203-217 (1990). Obviamente, la identificación de regiones amplificadas y eliminadas y la clonación de los genes involucrados es crucial tanto para el estudio de la tumorigénesis como para el desarrollo del diagnóstico del cáncer.

[0004] La detección de regiones cromosómicas amplificadas o eliminadas tradicionalmente se ha realizado 30 mediante citogenética. Debido al complejo empaquetamiento del ADN dentro de los cromosomas, la resolución de las técnicas citogenéticas ha estado limitada a regiones mayores de 10 Mb aproximadamente; aproximadamente la anchura de una banda en cromosomas teñidos con Giemsa. En cariotipos complejos con múltiples traslocaciones y otras variaciones genéticas, el análisis citogenético tradicional es de poca utilidad debido a que se carece de la 35 información del cariotipo o no se puede interpretar. Teyssier, J.R., Cancer Genet. Cytogenet., 37: 103 (1989). Además, el análisis citogenético de bandas convencional consume tiempo, requiere mucho trabajo, y con frecuencia resulta difícil o imposible.

[0005] Más recientemente se han utilizado sondas clonadas para valorar mediante transferencia de Southern la 40 cantidad de una secuencia de ADN dada en un cromosoma. Este procedimiento es eficaz incluso si el genoma se ha reorganizado enormemente para así eliminar la información cariotípica útil. No obstante, la transferencia de Southern solamente proporciona una estimación grosera del número de copias de una secuencia de ADN, pero no proporciona información sobre la localización de dicha secuencia dentro del cromosoma.

[0006] La hibridación genómica comparativa (CGH) es un enfoque más reciente para identificar la presencia y 45 localización de secuencias amplificadas/eliminadas. Véase Kallioniemi, y col., Science, 258: 818 (1992). La CGH, al igual que la transferencia de Southern, revela amplificaciones y eliminaciones independientemente de la reordenación del genoma. Además, la CGH proporciona una estimación del número de copias más cuantitativa que la transferencia de Southern, e incluso también proporciona información de la localización de la secuencia 50 amplificada o eliminada en el cromosoma normal.

[0007] Utilizando la CGH, se ha identificado la región 20q13 del cromosoma como una región que se amplifica 45 frecuentemente en cánceres (véase, por ejemplo, Kallioniemi y col., Genomics, 20: 125-128 (1994)). El análisis inicial de esta región en líneas celulares de cáncer de mama ha identificado una región de 2 Mb aproximadamente 55 en el cromosoma 20 que se amplifica de forma consistente.

[0008] La Base de datos del EMBL, 8 de mayo de 1996 (número de acceso a la base de datos del EBI W05407) describe una EST obtenida a partir de tejido humano de pulmón fetal.

[0009] La Base de datos del EMBL, 16 de marzo de 1996 (número de acceso a la base de datos del EBI N70546) describe una EST adicional obtenida a partir de tejido humano de pulmón fetal.

[0010] El documento WO 95/09929 A describe secuencias de ADNc procedentes de una región de amplificación en el cromosoma 20 relacionadas con la enfermedad.

[0011] Tanner MM y col. 1995 (Clinical Cancer Research, vol. 1, nº 12, páginas 1455-1461) describe la utilización de hibridación *in situ* con fluorescencia con una sonda de un cósmido para la región mínima de amplificación (RMC20C001) con el fin de estudiar la amplificación de 20q13 en 132 carcinomas primarios de mama y 11 metástasis. El tamaño del amplicón se estudió con cuatro sondas flanqueantes.

[0012] Tanner y col. 1994 (Cancer Research, vol. 54, número 16, páginas 4257-4260) describe la utilización de hibridación *in situ* con fluorescencia en interfase con sondas de cósmidos anónimos y clones P1 específicos para los genes con el fin de determinar la región mínima común de mayor número de copias y estudiar la implicación de genes conocidos en 20q13.

[0013] Kallioniemi A y col., 1994 (PNAS, vol. 91, número 6, páginas 2156-2160) describe la aplicación de hibridación genómica comparativa para 5 líneas celulares de cáncer de mama y 33 tumores primarios para descubrir y mapear las regiones del genoma con un mayor número de copias de la secuencia de ADN.

RESUMEN DE LA INVENCION

[0014] La presente invención está relacionada con la identificación de una región estrecha (600 kb aproximadamente) en el interior de un amplicón de 2 Mb localizado aproximadamente en la región cromosómica 20q13 (más exactamente en 20q13.2) que se amplifica de manera consistente en tumores primarios. Además, esta invención proporciona secuencias de ADNc que mapean en esta región. Estas secuencias son útiles como sondas o como dianas de sondas para controlar el número de copias relativas de las secuencias correspondientes procedentes de una muestra biológica, tal como una célula tumoral. También se describe un cóntigo (una serie de clones que extienden de forma contigua este amplicón) que se pueden utilizar para preparar sondas específicas para el amplicón. Las sondas se pueden utilizar para detectar anomalías cromosómicas en 20q13.

[0015] Así, la presente invención proporciona una molécula de ácidos nucleicos aislada que comprende una secuencia de polinucleótidos que contiene una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 o una de sus secuencias complementarias. En algunas formas de realización el ácido nucleico aislado es una molécula de ADNc.

[0016] En algunas formas de realización la secuencia de polinucleótidos comprende las SEQ. ID. Nos. 9 o 10.

[0017] En algunas formas de realización, el ácido nucleico aislado comprende adicionalmente una secuencia promotora unida de manera operable a la secuencia de polinucleótidos.

[0018] La invención también proporciona un procedimiento de selección de células neoplásicas en una muestra, el procedimiento que comprende:

la puesta en contacto de una muestra de ácidos nucleicos procedente de un paciente humano con un ácido nucleico aislado que se hibrida selectivamente a una secuencia de polinucleótidos diana que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 o una de sus secuencias complementarias, o a una subsecuencia de cualquiera de estas susodichas secuencias en donde el ácido nucleico aislado se pone en contacto con la muestra en condiciones en las que el ácido nucleico aislado se hibrida selectivamente con la secuencia de polinucleótidos diana para formar un complejo de hibridación estable; y

la detección de la formación de un complejo de hibridación para controlar el número de copias relativas de secuencias correspondientes a las SEQ. ID. Nos. 9 o 10 en la muestra.

[0019] En algunas formas de realización, la muestra de ácidos nucleicos procede de un paciente con un cáncer seleccionado del grupo constituido por cáncer de ovario, vejiga, cabeza y cuello, colon y cáncer de mama. En algunas formas de realización, el ácido nucleico aislado está marcado.

[0020] En algunas formas de realización, el ácido nucleico aislado está inmovilizado sobre una superficie sólida.

[0021] En algunas formas de realización, la muestra de ácidos nucleicos es un núcleo en propagación durante la metafase o un núcleo en interfase.

5 **[0022]** La invención proporciona adicionalmente un procedimiento para la selección de células neoplásicas en una muestra, el procedimiento que comprende:

la puesta en contacto de una muestra de ácidos nucleicos procedente de un paciente humano con un ácido nucleico aislado que se hibrida selectivamente a una secuencia de polinucleótidos diana que comprende una secuencia
10 seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 o una de sus secuencias complementarias, o a una subsecuencia de cualquiera de estas susodichas secuencias en donde el ácido nucleico aislado se pone en contacto con la muestra en condiciones en las que el ácido nucleico aislado se hibrida selectivamente con la secuencia de polinucleótidos diana para formar un complejo de hibridación estable; y

15 la detección de la formación de un complejo de hibridación para identificar la presencia de una mutación en la secuencia de polinucleótidos diana.

[0023] Incluso se proporciona un procedimiento de selección de células neoplásicas en una muestra biológica, dicho procedimiento que comprende la selección de la muestra para la sobre-expresión del gen ZABC1, en donde la
20 sobre-expresión del gen ZABC1 con respecto al tejido normal es indicativa de células neoplásicas, y en donde el gen ZABC1 corresponde a cualquiera de las secuencias SEQ. ID. Nos. 9 o 10.

[0024] En algunas formas de realización, la muestra para la sobre-expresión del gen ZABC1 comprende la puesta en contacto de la muestra con un anticuerpo que se une específicamente a un antígeno del polipéptido codificado
25 por una secuencia de polinucleótidos que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 y la detección de la formación de un complejo antígeno-anticuerpo.

[0025] En algunas formas de realización, la muestra se selecciona del grupo constituido por tejido de ovario, vejiga, cabeza y cuello, colon y mama.

30 **[0026]** En algunas formas de realización, el procedimiento comprende la selección de la muestra para la sobre-expresión de una proteína codificada por una secuencia de polinucleótidos que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10.

35 **[0027]** La invención proporciona un procedimiento como se ha descrito en este resumen para su utilización en el diagnóstico de la aparición o la progresión de un cáncer en un individuo del cual se ha derivado una muestra biológica, en donde una sobre-expresión de 3 veces o superior del gen ZABC1 en una muestra biológica está correlacionada con una menor supervivencia libre de enfermedad en sujetos con ganglios negativos.

40 **[0028]** Además se proporciona un procedimiento para el diagnóstico del comienzo o la progresión del cáncer en un individuo, cuyo procedimiento comprende las etapas de:

(a) determinar la expresión del gen ZABC1 en una muestra biológica procedente del individuo, y

45 (b) comparar el nivel de expresión del gen ZABC1 observado en la muestra biológica con el nivel de expresión del gen ZABC1 derivado de una segunda muestra de tejido clínicamente normal,

en donde una sobre-expresión del gen ZABC1 de 3 veces o superior en la muestra biológica está correlacionada con una menor supervivencia libre de enfermedad en sujetos con ganglios negativos.

50

Definiciones

[0029] Una "muestra de ácidos nucleicos" como se usa en el presente documento se refiere a una muestra que comprende ADN en una forma adecuada para su hibridación a sondas de la invención. El ácido nucleico puede ser
55 un ADN genómico total, un ARNm total, ADN o ARNm genómicos procedentes de cromosomas particulares, o secuencias seleccionadas (por ejemplo, promotores, genes, fragmentos de amplificación o restricción, ADNc particulares, etc.) dentro de los amplicones particulares descritos en el presente documento. La muestra de ácidos nucleicos se puede extraer a partir de células o tejidos particulares. La muestra de tejido a partir de la cual se prepara la muestra de ácidos nucleicos normalmente se obtiene de un paciente sospechoso de padecer la

enfermedad asociada a la amplificación a detectar. En algunos casos, los ácidos nucleicos se pueden amplificar utilizando técnicas habituales tales como PCR, antes de la hibridación. La muestra puede ser de ácidos nucleicos aislados inmovilizados sobre una superficie sólida (por ejemplo, nitrocelulosa) para su utilización en hibridaciones de Southern o membrana de transferencia de puntos y similares. La muestra también se puede preparar de manera que los ácidos nucleicos individuales permanezcan sustancialmente intactos y comprendan núcleos en interfase preparados de acuerdo con técnicas habituales. Una "muestra de ácidos nucleicos" como se usa en el presente documento también se puede referir a un cromosoma condensado sustancialmente intacto (por ejemplo, un cromosoma en metafase). Dicho cromosoma condensado es adecuado para su utilización como diana de hibridación en técnicas de hibridación *in situ* (por ejemplo, FISH). El uso particular del término "muestra de ácidos nucleicos" (ya sea en forma de ácidos nucleicos extraídos o como cromosoma intacto en metafase) será evidente para la persona experta en la materia según el contexto en el que se utilice el término. Por ejemplo, la muestra de ácidos nucleicos puede ser una muestra de tejidos o de células preparada para procedimientos de hibridación *in situ* habituales descritos a continuación. La muestra se prepara de manera que los cromosomas individuales permanezcan sustancialmente intactos y normalmente comprenden núcleos en propagación durante la metafase o núcleos en interfase de acuerdo con técnicas habituales.

[0030] Una "muestra de cromosomas" como se usa en el presente documento se refiere a una muestra de tejido o de células preparada para los procedimientos de hibridación *in situ* habituales descritos a continuación. La muestra se prepara de manera que los cromosomas individuales permanezcan sustancialmente intactos y normalmente comprenden núcleos en propagación durante la metafase o núcleos en interfase de acuerdo con técnicas habituales.

[0031] Un "ácido nucleico" se refiere a un polímero de desoxirribonucleótidos o ribonucleótidos en cualquiera de las formas de cadena sencilla o de doble cadena, y a menos que esté limitado de otra forma, englobará análogos conocidos de nucleótidos naturales que pueden funcionar de manera similar a los nucleótidos de origen natural.

[0032] Un polinucleótido "aislado" es un polinucleótido que está sustancialmente separado de otros contaminantes que lo acompañan de manera natural, por ejemplo, proteínas, lípidos, y otras secuencias de polinucleótidos. El término abarca secuencias de polinucleótidos que han sido extraídas o purificadas de su entorno natural o de una librería de clones, e incluyen aislados de ADN recombinantes o clonados y análogos sintetizados químicamente o análogos sintetizados biológicamente mediante sistemas heterólogos.

[0033] Una "subsecuencia" se refiere a una secuencia de ácidos nucleicos que comprende una parte de una secuencia de ácidos nucleicos más larga.

[0034] Una "sonda" o una "sonda de ácidos nucleicos", como se usa en el presente documento, se define como una colección de uno o más fragmentos de ácidos nucleicos cuya hibridación a una diana se puede detectar. La sonda se produce a partir de una fuente de ácidos nucleicos procedente de una o más porciones particulares (preseleccionadas) del genoma, por ejemplo, uno o más clones, un cromosoma completo o un fragmento de cromosoma aislados, o una colección de productos de amplificación de la reacción en cadena de la polimerasa (PCR). Las sondas se pueden producir a partir de ácidos nucleicos encontrados en el amplicón 20q13 como se describe en el presente documento. La sonda se puede procesar de alguna manera, por ejemplo, bloqueando o eliminando ácidos nucleicos repetitivos o enriqueciendo con ácidos nucleicos únicos. Así, la palabra "sonda" se puede usar en el presente documento para referirse no sólo a los ácidos nucleicos detectables, sino a los ácidos nucleicos detectables en la forma en que se aplican a la diana, por ejemplo, con los ácidos nucleicos de bloqueo, etc. El ácido nucleico de bloqueo también se puede mencionar por separado. A partir del contexto en el que se utiliza dicha palabra es evidente a qué se refiere específicamente la palabra "sonda".

[0035] La sonda también puede consistir en ácidos nucleicos aislados inmovilizados sobre una superficie sólida (por ejemplo, nitrocelulosa). En algunas formas de realización, la sonda puede ser un miembro de un biochip de ácidos nucleicos como se describe, por ejemplo, en el documento WO 96/17958. También se pueden utilizar para este propósito técnicas capaces de producir biochips de alta densidad (véase, por ejemplo, Fodor y col. Science 261:767-773 (1991) y la patente de EE.UU. Nº 5.143.854).

[0036] "Hibridación" se refiere a la unión de dos ácidos nucleicos de cadena sencilla mediante el apareamiento de bases complementarias.

[0037] "Que se une(n) sustancialmente" o "que se une específicamente" o "que se une selectivamente" o "que se hibrida específicamente" se refieren a la hibridación complementaria entre un oligonucleótido y una secuencia diana y comprende desapareamientos poco importantes que se pueden acomodar reduciendo las condiciones de

rigurosidad del medio de hibridación para conseguir la detección deseada de la secuencia de polinucleótidos diana. Estos términos también se refieren a la unión, formación de dúplex, o hibridación de una molécula sólo a una secuencia de nucleótidos particular en condiciones rigurosas cuando dicha secuencia está presente en una mezcla compleja de ADN o ARN (por ejemplo, celular total). El término "condiciones rigurosas" se refiere a las condiciones 5 bajo las que una sonda se hibridará a su subsecuencia diana, pero no a otras secuencias. Las condiciones rigurosas dependen de la secuencia y serán diferentes en diferentes circunstancias. Una "hibridación en condiciones rigurosas" y unas "condiciones de lavado de la hibridación en condiciones rigurosas" en el contexto de experimentos de hibridación con ácidos nucleicos tales como CGH, FISH, e hibridaciones de Southern y Northern dependen de la secuencia, y son diferentes bajo parámetros medioambientales diferentes. Se puede encontrar una guía exhaustiva 10 para la hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes*, parte I, capítulo 2 "overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, Nueva York. En general, las condiciones de lavado e hibridación altamente rigurosas se seleccionan para que se encuentren 5°C aproximadamente por debajo del punto de fusión térmico (T_m para la secuencia específica a una fuerza iónica y un pH definidos). La T_m es la temperatura (a una 15 fuerza iónica y un pH definidos) a la que el 50% de la secuencia diana se hibrida a una sonda perfectamente complementaria. Las condiciones muy rigurosas se seleccionan para que sean iguales a la T_m para una sonda particular.

[0038] Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que presentan más de 100 restos complementarios sobre un filtro en una transferencia de Southern o Northern es formalina al 50% con 1 mg de heparina a 42°C, y la hibridación se lleva a cabo durante toda la noche. Un ejemplo de condiciones de lavado rigurosas es el lavado en 0,2x SSC a 65°C durante 15 minutos (véase, Sambrook, más arriba para una descripción del tampón SSC). A menudo, el lavado en unas condiciones 25 altamente rigurosas está precedido por un lavado en unas condiciones poco rigurosas para eliminar la señal de fondo de la sonda. Un ejemplo de lavado en condiciones de rigurosidad media para un dúplex de, por ejemplo, 100 nucleótidos o más aproximadamente, es 1x SSC a 45°C durante 15 minutos. Un ejemplo de lavado en condiciones poco rigurosas para un dúplex de, por ejemplo, más de 100 nucleótidos, es 4x SSC a 40°C durante 15 minutos. En general, la observación de una relación de señal/ruido de 2x (o superior) frente a la observada para una sonda no relacionada en el ensayo de hibridación particular indica la detección de una hibridación específica.

[0039] La persona experta reconocerá que la secuencia precisa de las sondas particulares descritas en el presente documento se puede modificar hasta un cierto grado para producir sondas que sean "sustancialmente idénticas" a las sondas descritas, pero que retienen la capacidad para unirse sustancialmente a las secuencias dianas. Dichas modificaciones están cubiertas específicamente por referencia a las sondas individuales del presente 35 documento. El término "identidad sustancial" de las secuencias de polinucleótidos significa que un polinucleótido comprende una secuencia que tiene una identidad de secuencia de al menos el 90%, más preferentemente de al menos el 95%, en comparación con una secuencia de referencia utilizando los procedimientos descritos a continuación usando parámetros habituales.

[0040] Se dice que dos secuencias de ácidos nucleicos son "idénticas" si la secuencia de nucleótidos en las dos secuencias es la misma cuando se alinean para su máxima correspondencia como se describe a continuación. El término "complementario" se utiliza en el presente documento para indicar que la secuencia complementaria es idéntica a toda o una parte de una secuencia de polinucleótidos de referencia. Los ácidos nucleicos que no se hibridan a versiones complementarias entre sí en condiciones rigurosas aún son sustancialmente idénticos si los 45 polipéptidos que codifican son sustancialmente idénticos. Esto se produce, por ejemplo, cuando se genera una copia de un ácido nucleico utilizando la máxima degeneración codónica permitida por el código genético.

[0041] Las comparaciones de secuencia entre dos (o más) polinucleótidos normalmente se llevan a cabo comparando fragmentos de las dos secuencias a lo largo de una "ventana de comparación" para identificar y 50 comparar regiones locales de similitud de secuencia. Una "ventana de comparación", como se usa en el presente documento, se refiere a un segmento de al menos 20 posiciones contiguas aproximadamente, normalmente entre 50 aproximadamente y 200 aproximadamente, y más habitualmente entre 100 aproximadamente y 150 posiciones contiguas aproximadamente, en la que una secuencia se puede comparar con una secuencia de referencia del mismo número de posiciones contiguas después de que las dos secuencias se hayan alineado de forma óptima.

[0042] El alineamiento óptimo de secuencias para su comparación se puede realizar mediante el algoritmo de homología local de Smith y Waterman *Adv. Appl. Math.* 2: 482 (1981), mediante el algoritmo de alineamiento por homología de Needleman y Wunsch *J. Mol. Biol.* 48:443 (1970), mediante la búsqueda con el método de similitud de Pearson y Lipman *Proc. Natl. Acad. Sci. (U.S.A.)* 85: 2444 (1988), o mediante implementaciones computerizadas de 55

estos algoritmos.

[0043] El "porcentaje de identidad de secuencia" se determina comparando dos secuencias alineadas de forma óptima a lo largo de una ventana de comparación, en donde la fracción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o eliminaciones (es decir, huecos) comparada con la secuencia de referencia (que no comprende adiciones o eliminaciones) para el alineamiento óptimo de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que se encuentran bases de ácidos nucleicos o restos de aminoácidos idénticos en ambas secuencias para dar el número de posiciones apareadas, dividiendo el número de posiciones apareadas por el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para dar el porcentaje de identidad de secuencia. Otra indicación de que las secuencias de nucleótidos son sustancialmente idénticas se produce si dos moléculas se hibridan a la misma secuencia de ácidos nucleicos en condiciones rigurosas. Las "variaciones conservativamente modificadas" de una secuencia de ácidos nucleicos particular se refieren a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o en donde el ácido nucleico no codifica una secuencia de aminoácidos, para secuencias esencialmente idénticas. Debido a la degeneración del código genético, un gran número de ácidos nucleicos funcionalmente idénticos codifican cualquier polipéptido dado. Por ejemplo, los codones CGU, CGC, CGA, CCG, AGA, y AGG codifican todos el aminoácido arginina. Así, en cualquier posición en la que una arginina esté especificada por un codón, el codón se puede alterar con cualquiera de los codones descritos correspondientes sin alterar el polipéptido codificado. Dichas variaciones de ácidos nucleicos son "variaciones silenciosas", que son un tipo de "variaciones conservativamente modificadas". Cada secuencia de ácidos nucleicos del presente documento que codifica un polipéptido también describe cada posible variación silenciosa. La persona experta reconocerá que cada codón en un ácido nucleico (excepto AUG, que ordinariamente es el único codón para la metionina) se puede modificar para proporcionar una molécula funcionalmente idéntica mediante técnicas habituales. Por consiguiente, en cada secuencia descrita está implícita cada "variación silenciosa" de un ácido nucleico que codifica un polipéptido. Además, la persona experta reconocerá que sustituciones, eliminaciones o adiciones individuales que alteran, añaden o eliminan un solo aminoácido o un pequeño porcentaje de aminoácidos (normalmente menos del 5%, y más habitualmente menos del 1%) en una secuencia codificada son "variaciones conservativamente modificadas" en donde las alteraciones producen la sustitución de un aminoácido por un aminoácido químicamente similar. Son muy conocidas en la materia las tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares. Cada uno de los siguientes seis grupos contiene aminoácidos que son sustituciones conservativas entre sí:

- 1) Alanina (A), Serina (S), Treonina (T);
- 2) Ácido aspártico (D), ácido glutámico (E);
- 3) Asparagina (N), Glutamina (Q);
- 4) Arginina (R), Lisina (K);
- 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V); y
- 6) Fenilalanina (F), Tirosina (Y), Triptófano (W).

[0044] El término "proteína del amplicón 20q13" se utiliza en el presente documento para referirse a proteínas codificadas por ORFs en el amplicón 20q13 descrito en el presente documento. Los ensayos que detectan las proteínas del amplicón 20q13 están diseñados para detectar el nivel de proteínas del amplicón 20q13 endógenas (nativas) presentes en la muestra biológica del sujeto. No obstante, se pueden añadir proteínas del amplicón 20q13 exógenas (procedentes de una fuente extrínseca a la muestra biológica) a diversos ensayos para proporcionar un marcador o para competir con la proteína del amplicón 20q13 nativa en su unión a un anticuerpo dirigido contra la proteína del amplicón 20q13. La persona experta apreciará que en este contexto se puede utilizar un mimético de la proteína del amplicón 20q13 en lugar de la proteína 20q13 exógena. Una "proteína 20q13", como se usa en el presente documento, se refiere a una molécula que porta uno o más epítomos de la proteína del amplicón 20q13 de manera que se una específicamente a un anticuerpo que se una específicamente a una proteína del amplicón 20q13 nativa.

[0045] Como se usa en el presente documento, un "anticuerpo" se refiere a una proteína que consta de uno o más polipéptidos sustancialmente codificados por genes de inmunoglobulinas o fragmentos de genes de inmunoglobulinas. Los genes de inmunoglobulinas reconocidos incluyen los genes de la región constante kappa, lambda, alfa, gamma, delta, epsilon y mu, así como los miles de genes de la región variable de las inmunoglobulinas. Las cadenas ligeras se clasifican en kappa o lambda. Las cadenas pesadas se clasifican en gamma, mu, alfa, delta o epsilon, que a su vez definen las clases de inmunoglobulinas, IgG, IgM, IgA, IgD e IgE, respectivamente.

[0046] Se sabe que la unidad estructural básica de las inmunoglobulinas (anticuerpo) comprende un tetrámero. Cada tetrámero está compuesto de dos pares idénticos de cadenas polipeptídicas, cada par que presenta una cadena "ligera" (25 kD aproximadamente) y una cadena "pesada" (50-70 kD aproximadamente). El N-término de cada cadena define una región variable de 100 a 110 aminoácidos aproximadamente, o superior, responsables principalmente de reconocimiento de antígenos. Los términos cadena ligera variable (V_L) y cadena pesada variable (V_H) se refieren a estas cadenas ligera y pesada, respectivamente.

[0047] Los anticuerpos pueden existir en forma de inmunoglobulinas intactas o en forma de una serie de fragmentos bien caracterizados producidos por la digestión con diversas peptidasas. Así, por ejemplo, la pepsina digiere un anticuerpo por debajo de los puentes disulfuro en la región bisagra para producir $F(ab)'_2$, un dímero de Fab que es él mismo una cadena ligera unida a V_H-C_{H1} mediante un enlace disulfuro. El $F(ab)'_2$ se puede reducir en condiciones suaves para romper el puente disulfuro en la región bisagra convirtiendo así al dímero $F(ab)'_2$ en un monómero Fab'. El monómero Fab' es esencialmente un Fab con parte de la región bisagra (véase, Fundamental Immunology, W.E. Paul, ed., Raven Press, N.Y. (1993) para una descripción más detallada de otros fragmentos de anticuerpo). Aunque varios fragmentos de anticuerpo se definen en términos de digestión de un anticuerpo intacto, la persona experta apreciará que dichos fragmentos Fab' se pueden sintetizar químicamente *de novo* o utilizando metodologías de ADN recombinante. Así, el término anticuerpo, como se usa en el presente documento también incluye fragmentos de anticuerpos producidos mediante la modificación de anticuerpos completos o sintetizados *de novo* utilizando metodologías de ADN recombinante.

[0048] La frase "que se une específicamente a una proteína" o "específicamente inmunorreactivo con", cuando hace referencia a un anticuerpo, se refiere a una reacción de unión que es determinante para la presencia de la proteína en presencia de una población heterogénea de proteínas y otros componentes biológicos. Así, en condiciones de inmunoensayo designadas, los anticuerpos especificados se unen a una proteína particular y no se unen en una cantidad significativa a otras proteínas presentes en la muestra. La unión específica a una proteína bajo dichas condiciones puede requerir un anticuerpo que se seleccione por su especificidad para una proteína particular. Por ejemplo, se pueden generar anticuerpos contra la proteína del amplicón 20q13 que se unen a la proteína del amplicón 20q13 y no a cualquier otra proteína presente en una muestra biológica. Se pueden utilizar una variedad de formatos de inmunoensayo para seleccionar anticuerpos específicamente inmunorreactivos con una proteína particular. Por ejemplo, los inmunoensayos de ELISA en fase sólida se utilizan de manera rutinaria para seleccionar anticuerpos monoclonales específicamente inmunorreactivos con una proteína. Véase, Harlow y Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, Nueva York, para una descripción de los formatos y las condiciones de los inmunoensayos que se pueden utilizar para determinar la inmunorreactividad específica.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0049] La Figura 1(A) muestra la supervivencia libre de enfermedad de 129 pacientes con cáncer de mama según el nivel de amplificación de 20q13. Las pacientes con tumores que presentan unos niveles elevados de amplificación de 20q13 tienen una supervivencia libre de enfermedad más corta ($p = 0,04$ mediante la prueba de Mantel-Cox) en comparación con aquellas que no presentan amplificación o un bajo nivel de amplificación.

La Figura 1(B) muestra la misma diferencia de supervivencia libre de enfermedad de la Figura 1(A) en el subgrupo de 79 pacientes con ganglios axilares negativos ($p = 0,02$ mediante la prueba de Mantel-Cox).

La Figura 2 muestra una comparación de la amplificación de 20q13 detectada mediante FISH en un carcinoma de mama primario y su metástasis en una paciente de 29 años. Se encontró un bajo nivel de amplificación de 20q13 (20q13 comparado con la sonda de referencia 20p) en el tumor primario. La metástasis, que apareció 8 meses después de la mastectomía, muestra un nivel elevado de amplificación de la región 20q13 del cromosoma. El número total de copias del cromosoma 20 (sonda de referencia 20p) permaneció inalterado. Cada punto de los datos representa los números de copias génicas en células tumorales individuales analizadas.

La Figura 3 muestra una representación gráfica del mapeo citogenético molecular y la posterior clonación del amplicón 20q13.2. La distancia genética se indica en cintimorgans (cM). La línea negra gruesa representa la región con el nivel de amplificación más elevado en la línea celular de cáncer de mama BT474 y cubre una región de 1,5 Mb aproximadamente. Los clones P1 y BAC están representados como líneas horizontales cortas y los clones YAC como líneas horizontales más gruesas. No están representados todos los clones YAC y P1. Los YACs 957f3, 782c9, 931h12, y 902 están truncados. Los sitios de secuencia marcada (STSs) aparecen como líneas verticales finas y los círculos vacíos indican que un YAC dado se ha sometido a ensayo y es positivo para un STS dado. No todos los

STs se han probado sobre todos los YACs. El intervalo a partir del cual se han atrapado más de 100 exones está representado como una caja sólida. El intervalo de 600 kb que abarca la región con el nivel de amplificación más elevado en tumores primarios está representado mediante la caja negra sólida (secuencia marcada). La parte inferior de la figura muestra los niveles de amplificación en dos tumores primarios que acota aún más la región con la amplificación más elevada dentro de los 600 kb aproximadamente.

La Figura 4 proporciona un mapa con una resolución mayor del núcleo del amplicón como se define en tumores primarios.

10 La Figura 5 muestra la localización en el mapa de 15 genes en el amplicón.

La Figura 6 muestra un alineamiento de secuencias entre la ciclofilina de rata y la SEQ. ID. 13.

La Figura 7 es un mapa físico del amplicón 20q13. La columna A de la Figura muestra la posición de STs utilizados para organizar el mapa. La columna B muestra la posición de los YACs. La columna C muestra la posición de los clones P1 y BAC. La columna D muestra la posición de exones atrapados, ADnc directos seleccionados, ESTs, y genes. La columna E muestra los resultados del análisis de hibridación de diversos tumores primarios y de líneas celulares utilizando los clones P1 o BAC como se ha indicado. Un círculo negro sólido indica una amplificación elevada, un círculo sombreado indica una amplificación intermedia y un círculo vacío indica que no hay amplificación tal y como se detecta para cada uno de los clones.

DESCRIPCIÓN DETALLADA

[0050] Esta invención proporciona una serie de secuencias de ADnc que se pueden utilizar como sondas para la detección de anomalías cromosómicas en 20q13. Estudios que utilizan hibridación genómica comparativa (CGH) han demostrado que una región en el cromosoma 20q13 incrementa frecuentemente su número de copias en cánceres de mama (~30%), de ovario (~15%), vejiga (~30%), cabeza y cuello (~75%) y colon (~80%). Esto sugiere la presencia de uno o más genes que contribuyen a la progresión de diversos tumores sólidos que se encuentran en 20q13.

[0051] La amplificación génica es un mecanismo mediante el cual se sobre-expresan oncogenes que actúan de manera dominante, permitiendo que los tumores adquieran nuevas características de crecimiento y/o resistencia a agentes quimioterapéuticos. Los loci implicados en la progresión del cáncer de mama humano y amplificados en el 10-25% de los carcinomas de mama primarios incluyen el locus erbB-2 (Lupu y col., Breast Cancer Res. Treat., 27: 83 (1993), Slamon y col. Science, 235: 177-182 (1987), Heiskanen y col. Biotechniques 17: 928 (1994)) en 17q12, ciclina D (Mahadevan y col., Science, 255: 1253-1255 (1993), Gillett y col., Canc. Res., 54: 1812. (1994)) en 11q13 y MYC (Gaffey y col., Mod. Pathol., 6: 654 (1993)) en 8q34.

[0052] Investigaciones pangenómicas utilizando la hibridación genómica comparativa (CGH) han identificado recientemente unas 20 nuevas regiones con un número de copias incrementado en cáncer de mama (Kallioniemi y col., Genomics, 20: 125-128 (1994)). Uno de estos loci, la banda 20q13, estaba amplificado en el 18% de tumores primarios y en el 40% de las líneas celulares (Kallioniemi y col., Genomics, 20: 125-128 (1994)). Más recientemente, esta misma región se ha encontrado amplificada en el 15% de tumores de ovario, el 80% de tumores de vejiga y el 80% de tumores colorrectales. La resolución de la CGH está limitada a 5-10 Mb. Así, la FISH se llevó a cabo utilizando sondas específicas para el locus para confirmar los datos de la CGH y mapear con exactitud la región de amplificación.

[0053] La región 20q13 ha sido analizada en cáncer de mama a nivel molecular y se ha identificado una región, con una anchura de 600 kb aproximadamente, que se amplificaba de manera consistente, tal y como se ha descrito en el presente documento. Además, como se muestra en el presente documento, la importancia de esta amplificación en el cáncer de mama está indicada por la fuerte asociación entre la amplificación y la reducción en la supervivencia de la paciente y el incremento de la proliferación tumoral (específicamente, fracción incrementada de células en fase S).

[0054] En particular, como se explica con detalle en el Ejemplo 1, un nivel elevado de amplificación de 20q13 estaba asociado ($p = 0,0022$) con una mala supervivencia libre de enfermedad en pacientes con ganglios negativos, en comparación con casos sin amplificación o con un bajo nivel de amplificación (Figura 1). La supervivencia de pacientes con tumores moderadamente amplificados no difería significativamente de la de aquellos sin amplificación. Sin estar limitados por ninguna teoría particular, se sugiere que una explicación para esta observación puede ser

que un bajo nivel de amplificación preceda a un nivel elevado de amplificación. En este aspecto, puede ser significativo que una paciente desarrollase una metástasis local con un nivel elevado de amplificación de 20q13.2 8 meses después de la resección de un tumor primario con un bajo nivel de amplificación (Figura 2).

5 **[0055]** La amplificación de 20q13 estaba asociada a un grado histológico elevado de los tumores. Esta correlación se observó tanto en tumores moderadamente amplificados como en tumores muy amplificados. También había una correlación ($p = 0,0085$) entre el nivel de amplificación elevado de una región complementaria a una sonda particular, RMC20C001 (Tanner y col., Cancer Res., 54: 4257-4260 (1994)), y la proliferación celular, medida mediante la fracción de células en fase S. Este hallazgo es importante debido a que identifica un fenotipo que se
10 puede puntuar en ensayos funcionales, sin saber el mecanismo que subyace en la fracción incrementada en fase S. La amplificación de 20q13 no estaba correlacionada con la edad de la paciente, el tamaño del tumor primario, el estado de los ganglios axilares o del receptor de hormonas esteroideas.

[0056] Este trabajo localizó el amplicón 20q13.2 en un intervalo de 2 Mb aproximadamente. Además, sugiere que
15 el elevado nivel de amplificación, encontrado en el 7% de los tumores, confiere un fenotipo agresivo al tumor, que afecta de manera adversa a los resultados clínicos. El bajo nivel de amplificación (22% de tumores primarios) estaba asociado a características patológicas típicas de tumores agresivos (elevado grado histológico, aneuploidia y proliferación celular) pero no al pronóstico de la paciente.

20 **[0057]** Además, en el presente documento se muestra que el amplicón 20q13 (más precisamente el amplicón 20q13.2) es uno de los tres loci co-amplificados separados del cromosoma 20 humano que se empaquetan juntos en los genomas de algunos tumores primarios y líneas celulares de cáncer de mama. Ningún oncogén conocido mapea en el amplicón 20q13.2.

25 **Identificación de sondas del amplicón 20q13**

[0058] Inicialmente, la escasez de sondas citogenéticas moleculares disponibles dictaba que las sondas de la FISH se generasen mediante la selección aleatoria de cósmidos procedentes de una librería específica del cromosoma 20, LA20NC01, y a continuación se mapeasen en el cromosoma 20 mediante microscopía de imagen
30 digital. Se aislaron 46 cósmidos aproximadamente, que abarcan las 70 Mb del cromosoma, para los que se obtuvieron la medición de la longitud fraccional (FLpter) y las asignaciones de bandas. Se utilizaron 26 de los cósmidos para someter a ensayo el número de copias en la línea celular de cáncer de mama BT474 mediante análisis de FISH en interfase. El número de copias se determinó contando las señales de hibridación en los núcleos en interfase. Este análisis reveló que el cósmido RMC20C001 (FLpter, 0,824; 20q13.2), descrito por Stokke y col.,
35 Genomics, 26: 134-137 (1995), definía el nivel de amplificación más elevado (60 copias/célula) en células BT474 (Tanner y col., Cancer Res., 54: 4257-4260 (1994)).

[0059] Los clones P1 que contienen secuencias genéticamente mapeadas se seleccionaron a partir de 20q13.2 y se utilizaron como sondas de la FISH para confirmar y definir adicionalmente la región de amplificación. Se
40 seleccionaron otros clones P1 para oncogenes candidatos localizados ampliamente en la región 20q13.2 (Flpter, 0,81-0,84). Éstos se seleccionaron a partir de la librería DuPont P1 (Shepherd, y col., Proc. Natl. Acad. Sci. USA, 92: 2629 (1994), disponible comercialmente en Genome Systems), mediante PCR (Saiki y col., Science, 230: 1350 (1985)) utilizando pares de cebadores desarrollados en la región sin traducir en 3' de cada gen candidato. Se obtuvieron clones del gen específico P1 para la proteína tirosina fosfatasa (PTPN1, Flpter 0,78), receptor de melanocortina 3 (MC3R, Flpter 0,81), fosfoenolpiruvato carboxi quinasa (PCK1, Flpter 0,85), dedos de zinc de la proteína 8 (ZNF8, Flpter 0,93), proteína de unión al nucleótido guanina (GNAS 1, Flpter 0,873), oncogén src (SRC, Flpter 0,669), topoisomerasa 1 (TOP1, Flpter 0,675), gen bcl-x, relacionado con bcl-2 (Flpter 0,526) y factor de transcripción E2F-1 (FLpter 0,541). Cada clon se mapeó mediante microscopía de imagen digital y se les asignó los valores Flpter. Cinco de estos genes (SRC, TOPO1, GNAS1, E2F-1 y BC1-x) fueron excluidos como oncogenes
45 candidatos en el amplicón debido a que mapeaban muy fuera de la región crítica en Flpter 0,81-0,84. Tres genes (PTPNR1, PCK-1 y MC3R) estaban localizados suficientemente próximos a la región crítica para justificar una investigación más a fondo.

[0060] La FISH en interfase sobre 14 líneas celulares de cáncer de mama y 36 tumores primarios utilizando 24
55 cósmidos y 3 sondas del gen específico P1 (PTPNR1, PCK-1 y MC3R) tuvieron un elevado nivel de amplificación en el 35% (5/14) de las líneas celulares de cáncer de mama y el 8% (3/36) de los tumores primarios con una o más sondas. La región con el mayor número de copias en 4/5 de las líneas celulares y en 3/3 de los tumores primarios estaba definida por el cósmido RMC20C001. Esto indicaba que PTPNR1, PCK1 y MC3R también se podían excluir como candidatos para oncogenes en el amplicón y, además, se redujo la región crítica de 10 Mb a 1,5-2,0 Mb

(véase, Tanner y col., Cancer Res., 54: 4257-4260 (1994)).

- [0061]** Debido a que la sonda RMC20C001 detectó un nivel elevado (de 3 a 10 veces) de amplificación de 20q13.2 en el 35% de las líneas celulares y el 8% de los tumores primarios, se utilizó para (1) definir la prevalencia de la amplificación en una población tumoral expandida, (2) evaluar la frecuencia y el nivel de amplificación en estos tumores, (3) evaluar la asociación del amplicón 20q13.2 con características patológicas y biológicas, (4) determinar si existe una relación entre la amplificación de 20q13 y los resultados clínicos y (5) evaluar la amplificación de 20q13 en tumores de mama metastásicos.
- 10 **[0062]** Como se detalla en el Ejemplo 1, se utilizó hibridación *in situ* fluorescente (FISH) con RMC20C001 para evaluar la amplificación de 20q13.2 en 132 tumores de mama primarios y 11 tumores de mama recurrentes. Se determinó el número de copias absoluto (media del número de señales de hibridación por célula) y el nivel de amplificación (media del número de señales en relación con la sonda de referencia del brazo p). Se encontraron dos tipos de amplificación: un bajo nivel de amplificación (1.5-3 veces con señales de FISH dispersas por los núcleos del tumor) y un nivel elevado de amplificación (> 3 veces con señales de FISH estrechamente agrupadas). Se encontró un bajo nivel de amplificación de 20q13.2 en 29 de los 132 tumores primarios (22%), mientras que nueve casos (6,8%) presentaban un nivel elevado de amplificación.
- 15 **[0063]** Se utilizaron RMC20C001 y cuatro sondas P1 flanqueantes (MC3R, PCK, RMC20C026, y RMC20C030) para estudiar el grado de amplificación del ADN en tumores altamente amplificados. Sólo RMC20C001 se amplificó de manera consistente en todos los tumores. Este hallazgo confirmó que la región de amplificación común está dentro de un intervalo de 2 Mb flanqueado por, pero que no incluye a, PCK-1 y MC3R.
- 20 **[0064]** Se organizó un mapa físico para acotar la región común mínima de amplificación y aislar el oncogén(es) postulado. La librería DuPont P1 (Shepherd y col. Proc. Natl. Acad. Sci. USA, 91: 2629 (1994) se sometió a selección para STSs que mapearan con probabilidad en la banda 20q13.2. Se aislaron los clones P1 en los loci D20S102, D20S100, D20S120, D20S183, D20S480, D20S211, y la FISH localizó a cada uno en 20q13.2. A continuación se llevó a cabo un análisis de FISH en interfase en la línea celular de cáncer de mama BT474 para evaluar el nivel de amplificación en cada locus. Los loci D20S100, D20S120, D20S183, D20S480, D20S211 estaban muy amplificados en la línea celular BT474, mientras que D20S102 detectó sólo a un bajo nivel de amplificación. Por lo tanto, se localizaron 5 STSs, que abarcan 5 cM, dentro del amplicón 20q13.2 y se utilizaron para someter a selección a la librería CEPH megaYAC.
- 25 **[0065]** La selección de la librería CEPH megaYAC y búsquedas por ordenador de bases de datos públicas revelaron que D20S120-D20S83-D20S480-D20S211 están relacionados en cada uno de los tres clones megaYAC; y820f5, 773h10 y 931h6 (Figura 3). Además, la selección de la librería CEPH megaYAC con los STSs generados a partir de los extremos de los cósmidos RMC20C001, RMC20C30 y RMC20C028 localizó RMC20C001 en cada uno de los mismos tres clones YAC. Se estimó, basándose en el tamaño del más pequeño de estos clones YAC, que D20S120-D20S183-RMC20C001-D20S480-D20S211 mapean en un intervalo de menos de 1,1 Mb. D20S100 estaba localizado a 300 kb distal a D20S120 por FISH en interfase y a YAC901b12 por mapeo de STS. Los datos de los STS combinados hicieron posible la construcción de un cóntigo YAC de 12 miembros que abarca 4 Mb aproximadamente que engloba al amplicón de 1,5 Mb y que contiene los loci RMC20C030-PCK1-RMC20C001-MC3R-RMC20C026. Cada YAC se mapeó por FISH para confirmar la localización en 20q13.2 y para comprobar la existencia de quimerismo. Se clasificaron por tamaños cinco aislados clonales de cada YAC mediante electroforesis en gel de campo pulsado (PFGE). Ninguno de los YACs son quiméricos, sin embargo, varios son muy inestables.
- 30 **[0066]** El cóntigo YAC sirvió como base a partir del cual construir un cóntigo P1 de 2 Mb que abarca el amplicón 20q13. Los clones P1 ofrecían numerosas ventajas sobre los clones YAC, incluyendo (1) estabilidad, (2) una frecuencia quimérica inferior al 1%, (3) aislamiento del ADN mediante procedimientos MiniPrep estándar, (4) sondas de FISH ideales, (5) los extremos se pueden secuenciar directamente, (6) los transposones y/o manipulados genéticamente que portan sitios de unión de cebadores bidireccionales se pueden integrar en cualquier posición del ADN clonado (Strathmann y col., Proc Natl. Acad. Sci. USA, 88: 1247 (1991)) (7) los clones P1 son los moldes para la secuenciación de los genomas del ser humano y de *Drosophila* en el LBNL HGC (Palazzolo y col. DOE Human Genome Program, Contractor-Grantee Workshop IV. Santa Fe, Nuevo México, 13-17 noviembre, 1994).
- 35 **[0067]** Se aislaron aproximadamente 90 clones P1 sometiendo a selección a la librería DuPont P1 por PCR o hibridación en filtro. Para la selección basada en la PCR, se crearon más de 22 nuevos STSs mediante dos procedimientos. En el primer procedimiento, se secuenciaron los extremos de los clones P1 localizados en el amplicón, se crearon los STSs, y la librería P1 se sometió a selección para paseo clónico. En el segundo enfoque,
- 40
45
50
55

se llevó a cabo una PCR inter-Alu (Nelson y col., 86: 6686-6690 (1989)) sobre YACs que abarcan el amplicón y los productos se clonaron y se secuenciaron para la creación de STS. En el esquema de hibridación basada en filtros, se tuvieron clones P1 llevando a cabo PCR inter-Alu sobre YACs que abarcan el amplicón, radiomarcando los productos e hibridando éstos contra filtros que contienen un biochip reticulado de la librería P1. Por último, para 5 tapar los huecos se sometió a selección mediante PCR a una librería de cromosomas bacterianos artificiales (BAC) genómicos humanos (Shizuya y col. Proc. Natl. Acad. Sci. USA, 89: 8794 (1992), disponible comercialmente en Research Genetics, Huntsville, Alabama, EE.UU.). Estos procedimientos combinados para producir más de 100 clones P1 y BAC se localizaron en 20q13.2 mediante FISH. Se utilizó el mapeo para el contenido en STS, las huellas dactilares y la FISH libre de cromatina (Heiskanen y col., BioTechniques, 17: 928 (1994)) para construir el 10 cóntigo de 2 Mb mostrado en la Figura 3 y en la Figura 4. La Figura 7 es un segundo mapa físico de la región que muestra un cóntigo con un camino de recubrimiento mínimo.

Mapeo fino del amplicón 20q13.2 en BT474

15 **[0068]** Se utilizaron clones procedentes del cóntigo P1 de 2 Mb con la FISH para mapear el nivel de amplificación en 20q13.2 en la línea celular de cáncer de mama BT474. Se utilizaron 35 sondas P1 distribuidas a intervalos regulares a lo largo del cóntigo. Los datos resultantes indicaban que la región con el mayor incremento en el número de copias en BT474 se produce entre D20S100 y D20S211, un intervalo de 1,5 Mb aproximadamente. Las sondas de la FISH P1, en este intervalo, detectan una media de 50 señales por núcleos en interfase en BT474, mientras que 20 no se detectó ninguna, o sólo un bajo nivel de amplificación, con los clones P1 fuera de esta región. Así, se clonaron tanto los límites proximales como distales del amplicón.

Mapeo fino del amplicón 20q13.2 en tumores primarios

25 **[0069]** El mapeo fino del amplicón en tumores primarios reveló la región común mínima de elevada amplificación (MCA) que tiene importancia patobiológica. Este proceso es análogo a la selección para la meiosis informativa en el acotamiento de los intervalos genéticos que codifican genes patológicos heredables. El análisis de 132 tumores primarios reveló que en el locus RMC20C001 se amplifican 38 tumores primarios. Nueve de estos tumores presentan un nivel elevado de amplificación en el locus RMC20C001 y se analizaron en profundidad mediante FISH 30 en interfase con 8 P1 que abarcan el cóntigo de \approx 2 Mb. La región común mínima de amplificación (MCA) se mapeó en un intervalo de \approx 600 kb flanqueado por los clones P1 #3 y #12, con el nivel de amplificación más elevado detectado para el clon P1 #38 correspondiente a RMC20C001 (Figuras 3 y 7).

35 **[0070]** Los clones P1 y BAC que abarcan el intervalo de 600 kb del amplicón 20q13 se listan en las Tablas 1 y 2 que proporcionan una referencia cruzada a la librería DuPont P1 descrita por Shepherd, y col., Proc. Natl. Acad. Sci. USA, 92: 2629 (1994). Estas sondas P1 y BAC están disponibles comercialmente en Genetic Systems, y en Research Genetics, respectivamente).

Tabla I. Referencia cruzada a sondas de la librería DuPont P1 (Shepherd, y col., Proc. Natl. Acad. Sci. USA, 92:2629 40 (1994), que está disponible comercialmente en Genomic Systems, St. Louis, Missouri, EE.UU.). Se ilustran los cebadores de la PCR para la amplificación de los sitios de secuencia marcada para cada clon. Además, se proporcionan las condiciones de la PCR (concentración de Mg y temperatura de hibridación), así como el tamaño del producto de la PCR. Tamaño: El tamaño del producto de la PCR; MgCl mM: concentración de Mg; Ann: Temperatura de hibridación; P1: número de identificación de la Sonda P1; PC: Coordenadas de la librería DuPont; SCA: Dirección de un clon único de la librería DuPont; SEQ-directa y SEQ-inversa: cebadores de la PCR directo e inverso, 45 respectivamente; SEQ ID NO:-directa y SEQ ID NO:-inversa: Listado de secuencias SEQ ID NO: para los cebadores directo e inverso, respectivamente.

NOMBRE DEL CEBADOR	TAMAÑO (pb)	MgCl Ann. mM	PI	PC	SCA	SEQ-directa	SEQ-inversa	SEQ	SEQ
								ID	ID
								NO:-	NO:-
								directa	inversa
352.32	136	1,5	52	20	103.5	1228e	TTGGCATTGGTATCAGGTAGCTG	TTGGAGCAGAGAGGGGATTGTGTG	1 2
388.13 F1/B1	201	1,5	52	17	69g6	821	AATCCCCCTCAAACCCTGCTGCTAC	TGGAGCCTGAACTTCTOCAATC	3 4
				19	98f4	1167f	-	-	- -
D20S183	270	3	48	30	124g6	CCC GGATACCGACATTG	TGCACATAAAAACAGCCACC	5 6	
				40	24hl	276h	-	-	- -
D20S211 ½	135	1,5	52	29	119f4	1418f	TTGAAATCAATGGAGCAAAA	AGCITTACCCAAATGTGTGCC	7 8
D205480	300	3	55	68	100d12	1199d2	GTGGTGAACACCCAATAAATGG	AAGCAAATAAAAACCAATAAACTCG	9 10
				41	86c1	1020c	-	-	- -
				2	103d9	1232d	-	-	- -
				67	91b2	1081b9	-	-	- -
9X-SP6 hmF/hmB	165	1,5	55	7	31d11	370d	CAAGATCTGACCCCUCAATC	GACTTCTTCAGGAAAAGATCAGTG	11 12
				9	3519	416f	-	-	- -
11S-17 F2/B4	146	3	58	11	41b1	480b	GCCATGTACCCACCTGAAAAATC	TCAGAACACCCCGTGCAGAATTAAAG	13 14
12T-T7 F2/B1	153	3	58	12	42c2	493c	CCTAAAAACTGGTGCTTAAATCTA	CTCTCACAAGGCAGATGTGG	15 16
28T F1/B1	219	1,5	52	74	888f2	TTTGTGTATGTTGACCCCATC	CTCCAATCTCATTCTATGAGG	17 18	
				2	12c6	137c	-	-	- -

Tabla II. Referencia cruzada a sondas de la librería BAC. Clon # se refiere al número del clon proporcionado, por ejemplo, en las Figuras 3, 4 y 7, aunque las coordenadas de la placa de matriz son las coordenadas de la placa de la librería BAC de Research Genetics. Tamaño: Tamaño del producto de la PCR; MgCl mM: concentración de Mg; Ann: Temperatura de hibridación; BAC #: Número de identificación de la sonda BAC; SEQ-directa y SEQ-inversa: 5 cebadores de la PCR directo e inverso, respectivamente; SEQ ID NO:-directa y SEQ ID NO:-inversa: Listado de secuencias SEQ ID NO: para los cebadores directo e inverso, respectivamente.

NOMBRE DEL CEBADOR	TAMAÑO (pb)	MgCl mM	Ann. BAC.	Coordenadas de la placa del BAC	SEQ-directa	SEQ-inversa	SEQ ID		
							NO:- directa	NO:- inversa	
18T F1/B1	156	3	62	99	L11 I placa 146	AGCAAAGGCAAAATGGCACACAC	TGACATGGGAGAGAAGACACACTTCC	29	30
95 F1/B1	214	1,5	55	97	E8 placa 183	AGGTRTACCAATGTGTTTGG	TCTACATCCCATTCTCTTCTG	31	32
0205480	300	3	55	95	H15 placa 140	GTGGTOAACACCCAATAAAATGG	AAGCAAAATAAAACCAATAAAACTCO	9	10
D20S211 1/2	135	15.5	52	103	A15 placa 188	TRGGAATCAATGGACCAAAA	AGCMACCCAATGTGTCC	7	8
				102	A1 placa 46	-	-	-	-
115-17 F2/84	146	3	58	100	E4 placa 43	GCCATGTACCACCTGAAAAATC	TCAGAACACCCCGTAACAGAAATTAAG	13	14
				101	15 placa 118	-	-	-	-
CYP14				87	J14 placa 96				
ET4211	104			104	C10 placa 754				
ET03.17				121	C10 placa 10				
D205902				166	14 placa 226				
180-R F1/B1				188	021 placa 163				
164-IFf/h2				197	N13 placa 309				

Secuencias de ADNc procedentes del amplicón 20q13

[0071] Se realizó el atrapamiento del exón (véase, por ejemplo, Duyk y col., Proc. Natl. Acad. Sci. USA, 87: 8995-8999 (1990) y Church y col., Nature Genetics, 6: 98-105 (1994)) sobre los clones P1 y BAC que abarcan la región común mínima de amplificación de 600 kb y se han aislado más de 200 exones.

[0072] El análisis de la secuencia de ADN de los exones reveló un número de similitudes de secuencia (85% a 96%) con secuencias de ADNc parcial en la base de datos (dbest) de la secuencia expresada y con un marco de lectura abierto del cromosoma XIV de *S. cerevisiae*. Cada clon P1 sometido al atrapamiento exónico ha producido múltiples exones que constan de al menos una densidad media de genes. Se han atrapado y analizado más de 200 exones, y se han aislado 200 clones mediante selección directa a partir de una librería de ADNc en BT474. Además se está secuenciando un intervalo genómico de 0,6 Mb que abarca el amplicón mínimo descrito a continuación. La predicción exónica y la modelización génica se llevó a cabo con las aplicaciones XGRAIL, SORFIND, y BLAST. Los fragmentos génicos identificados mediante estos enfoques se han analizado mediante RT-PCR, y transferencias de Northern y Southern. De esta forma se identificaron 15 genes únicos (véase, Tabla 3 y Figura 5).

[0073] Además también se demostró que otros dos genes ZABC1 (SEQ. ID. No. 9 y 10) y 1b1 (SEQ. ID. No. 12) se sobre-expresan en una variedad de células cancerosas diferentes.

[0074] A continuación se proporciona la información de secuencia procedente de diversos clones de ADNc. Son los siguientes:

3bf4 (SEQ. ID. No. 1) - transcrito de 3 kb con una identidad de secuencia con el gen de la tirosina quinasa, denominado A6, descrito en Beeler y col., Mol. Cell. Biol. 14:982-988 (1994) y en el documento WO 95/19439. No obstante, estas referencias no describen que el gen esté localizado en el amplicón del cromosoma 20.

1b11 (SEQ. ID. No. 2) - un transcrito de aproximadamente 3,5 kb cuya expresión muestra una gran correlación con el número de copias del amplicón. La secuencia no muestra homología con secuencias de las bases de datos investigadas.

cc49 (SEQ. ID. No. 3) - un transcrito de 6-7 kb que presenta homología con los genes de los dedos de cinc C2H2.

cc43 (SEQ. ID. No. 4) - un transcrito de 4 kb aproximadamente que se expresa en tejidos normales, pero cuya expresión en la línea celular de cáncer de mama no ha sido detectada.

41.1 (SEQ. ID. No. 5) - presenta homología con el gen de la homeocaja T en *Drosophila*.

GCAP (SEQ. ID. No. 6) - codifica una proteína que activa la guanino ciclase que está involucrada en la biosíntesis del AMP cíclico. Como se explica con detalle a continuación, las secuencias procedentes de este gen también se pueden utilizar para el tratamiento de la degeneración de retina.

1b4 (SEQ. ID. No. 7) - una serina-treonina quinasa.

20sa7 (SEQ. ID. No. 8) - un homólogo del gen de rata, BEM-1.

[0075] Además, se proporciona la secuencia completa de nucleótidos para ZABC-1. ZABC-1 significa dedos de cinc amplificados en cáncer de mama. Este gen mapea en el centro del amplicón 20q13.2 y se sobre-expresa en tumores primarios y en líneas celulares de cáncer de mama que presentan amplificación de 20q13.2. La secuencia genómica (SEQ. ID. No. 9) incluye 2 kb aproximadamente de la región promotora. La SEQ ID. No. 10 proporciona el marco de lectura abierto derivado de la secuencia de ADNc y la SEQ ID. No. 11 proporciona la secuencia proteica predicha. Los genes que contienen dedos de cinc a menudo son factores de transcripción que funcionan para modular la expresión de los genes en dirección 3'. Muchos oncogenes conocidos son, de hecho, genes que contienen dedos de cinc.

[0076] En el presente documento se describe la secuencia de ADNc de longitud completa para un ADNc denominado 1b1 (SEQ. ID. No. 12) que también se sobre-expresa en numerosas líneas celulares de cáncer de mama y en algunos tumores primarios.

[0077] La SEQ ID NO: 13 proporciona una secuencia procedente de un clon genómico que es similar a ADNc

conocidos de ciclofilina de rata y ratón. Los ácidos nucleicos de ciclofilina de rata (por ejemplo, ADNc) son conocidos; véase, GenBank™ con el N° de acceso M19533; los ácidos nucleicos de ciclofilina de ratón (por ejemplo, ADNc) son conocidos; véase GenBank™ con el N° de acceso 50620 (véase, Figura 6). Por consiguiente, la SEQ ID NO: 13 es un gen putativo de la ciclofilina humana. La secuencia también está asociada a secuencias amplificadas a partir de 20q13, y se puede utilizar como sonda o diana para la hibridación de sondas para detectar la amplificación de ADN, o la sobre-expresión de ARN del gen correspondiente.

Tabla 3. Fragmentos génicos identificados mediante el atrapamiento de exones y analizados mediante RT-PCR, y transferencias de Northern y Southern.

10

# de Mapa	ID del gen	Tamaño del transcrito (kb)	Identidad de la EST (>95%)	Secuencia clonada (kb)	Homologías de la proteína	Ubicación del mapa
1	20sa7	-	Sí	3	PTP	3, 18, 99
2	1b11	3,5	No	1,5	nueva	18, 3, 99, 69
3	200.1	-	-	-	-	-
4	132.1	-	-	-	nueva	132
5	132.2	-	-	-	-	132
6	3bf4	3	Sí	3	PTK	ambigua
7	7.1	-	-	-	-	7, 11, 97
8	7.2	2,4	-	-	-	7, 11, 97
9	cc49	7,4	Sí	3,6	Proteína	97, 103
10	cc43	1,4	Sí	1,8	hipotética de levadura de Kruppel	97, 7, 11
11	et1807	2,5	Sí	0,7	nueva	97, 9
12	et2106	Ninguno	Sí	-	-	95, 39, 38
13	41.1	detectado	Sí	3	Gen	95, 41, 42
14	67.1	7, 8, 11	Sí	2Kb	homeótico	67
15	67.2		Sí	-	Proteína reg. del GMPc	67
		0				

Proteínas del amplicón 20q13

[0078] Como se ha indicado anteriormente, en el presente documento se describen proteínas codificadas por secuencias de ácidos nucleicos en el amplicón 20q13 (por ejemplo, SEQ. ID. Nos: 1-10 y 12-13) y subsecuencias, más preferentemente subsecuencias con una longitud de al menos 10 aminoácidos, preferentemente de al menos 20 aminoácidos, y lo más preferentemente de al menos 30 aminoácidos. Subsecuencias particularmente preferidas son epítomos específicos para las proteínas 20q13, más preferentemente epítomos específicos para las proteínas ZABC1 y 1b1. Dichas proteínas incluyen, pero no están limitadas a polipéptidos aislados que comprenden al menos 10 aminoácidos contiguos procedentes de un polipéptido codificado por los ácidos nucleicos de las SEQ. ID No. 1-10 y 12-13 o procedente del polipéptido con la SEQ. ID. No. 11 en el que el polipéptido, cuando se presenta en forma de inmunógeno, desencadena la producción de un anticuerpo que se une específicamente a un polipéptido seleccionado del grupo constituido por un polipéptido codificado por los ácidos nucleicos de las SEQ. ID No. 1-10 y

12-13 o del polipéptido con la SEQ. ID. No. 11 y el polipéptido no se une a antisuero generado contra un polipéptido seleccionado del grupo constituido por un polipéptido codificado por los ácidos nucleicos de las SEQ. ID No. 1-10 y 12 o del polipéptido de la SEQ. ID. No. 11 que ha sido completamente inmuoabsorbido con un polipéptido seleccionado del grupo constituido por un polipéptido codificado por los ácidos nucleicos de las SEQ. ID No. 1-10 y 5 12-13 o del polipéptido de la SEQ. ID. No. 11.

[0079] Una proteína que se une específicamente o que es específicamente inmunorreactiva a un anticuerpo generado contra un inmunógeno definido, tal como un inmunógeno constituido por la secuencia de aminoácidos de la SEQ ID NO 11 se determina en un inmunoensayo. El inmunoensayo utiliza un antisuero policlonal que se genera 10 para la proteína de la SEQ ID NO 11 (el polipéptido inmunógeno). Este antisuero se selecciona para que tenga una baja reactividad cruzada contra otros polipéptidos similares conocidos y cualquiera de dicha reactividad cruzada se elimina mediante inmuoabsorción antes de su utilización en el inmunoensayo (por ejemplo, mediante inmuoabsorción del antisuero con el polipéptido relacionado).

15 **[0080]** Con el fin de producir un antisuero para su utilización en un inmunoensayo, el polipéptido, por ejemplo, el polipéptido de la SEQ ID NO 11 se aísla como se ha descrito en el presente documento. Por ejemplo, la proteína recombinante se puede producir en una línea celular de mamífero u otra eucariota. Una cepa endogámica de ratón se inmuniza con la proteína de la SEQ ID NO 11 utilizando un adyuvante habitual, tal como adyuvante de Freund, y un protocolo de inmunización de ratón habitual (véase, Harlow y Lane, más arriba). Alternativamente, se utilizó como 20 inmunógeno un polipéptido sintético derivado de las secuencias descritas en el presente documento y conjugado a una proteína portadora. Los sueros policlonales se recogieron y se titularon frente al polipéptido inmunógeno en un inmunoensayo, por ejemplo, un inmunoensayo en fase sólida con el inmunógeno inmovilizado sobre un soporte sólido. Se seleccionó y se probó antisuero policlonal con un título de 10^4 o superior para su reactividad cruzada frente a polipéptidos conocidos utilizando un inmunoensayo de unión competitiva tal como el descrito en Harlow y 25 Lane, más arriba, en las páginas 570-573. Preferentemente, junto con el polipéptido inmunógeno se utiliza más de un polipéptido conocido en esta determinación.

[0081] Los polipéptidos conocidos se pueden producir en forma de proteínas recombinantes y se pueden aislar utilizando técnicas habituales en biología molecular y química de proteínas, como se describe en el presente 30 documento.

[0082] Se utilizaron inmunoensayos en formato de unión competitiva para la determinación de la reactividad cruzada. Por ejemplo, el polipéptido inmunógeno se inmovilizó en un soporte sólido. Las proteínas añadidas al ensayo compiten con la unión del antisuero al antígeno inmovilizado. La capacidad de las proteínas para competir 35 con la unión del antisuero a la proteína inmovilizada se compara con la del polipéptido inmunógeno. Se calculó el porcentaje de reactividad cruzada para la proteína, utilizando cálculos habituales. Se seleccionan y se reúnen los antisueros con menos del 10% de reactividad cruzada con los polipéptidos conocidos. A continuación los anticuerpos que producen una reacción cruzada se extraen del antisuero reunido mediante inmuoabsorción con polipéptidos conocidos. 40

[0083] A continuación el antisuero inmuoabsorbido y reunido se utiliza en un inmunoensayo de unión competitiva como se describe en el presente documento para comparar un polipéptido "diana" con el polipéptido inmunógeno. Para realizar esta comparación, cada uno de los dos polipéptidos se sometió a ensayo en un amplio espectro de 45 concentraciones y se determinó la cantidad necesaria de cada polipéptido para inhibir el 50% de la unión del antisuero a la proteína inmovilizada utilizando técnicas habituales. Si la cantidad necesaria de polipéptido diana es inferior al doble de la cantidad requerida de polipéptido inmunógeno, entonces se dice que el polipéptido diana se une específicamente a un anticuerpo generado para la proteína inmunógena. Como determinación final de la especificidad, el antisuero reunido se inmuoabsorbe completamente con el polipéptido inmunógeno hasta que en la inmuoabsorción no se puede detectar la unión al polipéptido utilizado. A continuación el antisuero completamente 50 inmuoabsorbido se somete a ensayo para su reactividad con el polipéptido de prueba. Si no se observa reactividad, entonces el polipéptido de prueba se une específicamente al antisuero generado contra la proteína inmunógena.

[0084] De forma similar, en un experimento recíproco, el antisuero reunido se inmuoabsorbió con el polipéptido de prueba. Si el antisuero que permanece después de la inmuoabsorción no se une al polipéptido inmunógeno (es 55 decir, el polipéptido de la SEQ ID NO: 11 utilizado para generar el antisuero), entonces el polipéptido de prueba se une específicamente al antisuero generado contra el péptido inmunógeno.

Detección de anomalías en 20q13

[0085] La persona experta en la materia apreciará que la información de los clones y de las secuencias proporcionadas en el presente documento se puede utilizar para detectar amplificaciones, u otras anomalías cromosómicas, en 20q13 en una muestra biológica. En general, los procedimientos suponen la hibridación de sondas que se unen específicamente a una o más secuencias de ácidos nucleicos del amplicón diana con ácidos nucleicos presentes o procedentes de una muestra biológica.

[0086] Como se usa en el presente documento, una muestra biológica es una muestra de tejido o fluido biológico que contiene células que se desea someter a selección para la determinación de anomalías cromosómicas (por ejemplo, amplificaciones o eliminaciones). En una forma de realización preferida, la muestra biológica es una célula o tejido sospechoso de ser canceroso (transformado). Los procedimientos de aislamiento de muestras biológicas son muy conocidos por las personas expertas en la materia e incluyen, pero no están limitados a, aspiraciones, secciones de tejido, biopsias con aguja, y similares. Frecuentemente la muestra será una "muestra clínica" que es una muestra procedente de un paciente. Se reconocerá que el término "muestra" también incluye el sobrenadante (que contiene células) o las propias células procedentes de cultivos celulares, células procedentes de cultivos de tejidos u otros medios en los que pueda ser deseable detectar anomalías cromosómicas.

[0087] En algunas formas de realización, se prepara una muestra cromosómica depositando células, ya sea en forma de suspensiones de células únicas o en forma de preparación de tejido, sobre soportes sólidos tales como portaobjetos de vidrio y fijadas seleccionando un fijador que proporcione la mejor resolución espacial de las células y la eficacia de hibridación óptima. En otras formas de realización, la muestra se pone en contacto con un biochip de sondas inmovilizadas sobre una superficie sólida.

Preparación de sondas

[0088] Cualquiera de las sondas P1 listadas en la Tabla 1, las sondas BAC listadas en la Tabla 2, y los ADNc descritos en el presente documento son adecuados para su utilización en la detección del amplicón 20q13. Los procedimientos de preparación de sondas son muy conocidos por los expertos en la materia (véase, por ejemplo, Sambrook y col., Molecular Cloning: A Laboratory Manual (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) o Current Protocols in Molecular Biology, F. Ausubel y col., Ed. Greene Publishing y Wiley Interscience, Nueva York (1987)).

[0089] Dada la estrategia para la preparación de los ácidos nucleicos de la presente invención, la persona experta puede construir una variedad de vectores y clones de ácidos nucleicos que contengan ácidos nucleicos funcionalmente equivalentes. La metodología de clonación para conseguir estos fines y los procedimientos de secuenciación para verificar la secuencia de los ácidos nucleicos son muy conocidos en la materia. Ejemplos de técnicas de clonación y secuenciación adecuadas, e instrucciones suficientes para guiar a personas expertas a través de muchos ejercicios de clonación se pueden encontrar en Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volumen 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook y col. (1989) Molecular Cloning - A Laboratory Manual (2ª ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, (Sambrook); y Current Protocols in Molecular Biology, FM Ausubel y col., Eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (Suplemento de 1994) (Ausubel). La información del producto obtenido de los fabricantes de reactivos biológicos y equipos experimentales también resulta útil en procedimientos biológicos conocidos. Dichos fabricantes incluyen la empresa de productos químicos SIGMA (Saint Louis, MO), R & D Systems (Minneapolis, MN), Pharmacia LKB Biotechnology (Piscataway, NJ), CLONTECH Laboratories, Inc. (Palo Alto, CA), Chem Genes Corp., Aldrich Chemical Company (Milwaukee, WI), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersburg, MD), Fluka Chemica-Biochemika Analytika (Fluka Chemie AG, Buchs, Suiza), Invitrogen, San Diego, CA, y Applied Biosystems (Foster City, CA), así como muchas otras fuentes comerciales conocidas por la persona experta.

[0090] Los ácidos nucleicos descritos en el presente documento, ya sea ARN, ADNc, ADN genómico, o un híbrido de las diversas combinaciones, se aíslan a partir de fuentes biológicas o se sintetizan *in vitro*. Los ácidos nucleicos y vectores pueden estar presentes en células enteras transformadas o transfectadas, en lisados celulares transformados o transfectados, o en una forma parcialmente purificada o sustancialmente pura.

[0091] Son conocidas las técnicas de amplificación *in vitro* adecuadas para la amplificación de secuencias para proporcionar un ácido nucleico, o para su posterior análisis, secuenciación o subclonación. Ejemplos de técnicas suficientes para guiar a personas expertas a través de dichos procedimientos de amplificación *in vitro*, que incluyen la reacción en cadena de la polimerasa (PCR), la reacción en cadena de la ligasa (LCR), la amplificación de la replicasa Q β y otras técnicas mediadas por la ARN polimerasa (por ejemplo, NASBA), se encuentran en Berger,

Sambrook, y Ausubel, así como en Mullis y col. (1987), Patente de EE.UU. N° 4.683.202; PCR Protocols A Guide to Methods and Applications (Innis y col. eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (1 de octubre de 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3, 81-94; (Kwoh y col. (1989) Proc. Natl. Acad. Sci. USA 86, 1173; Guatelli y col. (1990) Proc. Natl. Acad. Sci. USA 87, 1874; Lomell y col. (1989) J. Clin. Chem 35, 1826; Landegren y col., (1988) Science 241, 1077-1080; Van Brunt (1990) Biotechnology 8, 291-294; Wu y Wallace, (1989) Gene 4, 560; Banger y col. (1990) Gene 89, 117, y Sooknanan y Malek (1995) Biotechnology 13: 563-564. En Wallace y col., Patente de EE.UU. N° 5.426.039 se describen procedimientos mejorados de clonación de ácidos nucleicos amplificados *in vitro*. Procedimientos mejorados de amplificación de ácidos nucleicos grandes están resumidos en Cheng y col. (1994) Nature 369: 684-685 y las referencias allí citadas.

10

[0092] Los ácidos nucleicos (por ejemplo, oligonucleótidos) para los procedimientos de amplificación *in vitro* o para su utilización como sondas génicas, por ejemplo, normalmente se sintetizan químicamente según el método del triéster de la fosoramidita en fase sólida descrito por Beaucage y Caruthers (1981), Tetrahedron Letts., 22 (20): 1859-1862, por ejemplo, usando un sintetizador automatizado, como se describe en Needham-Vandevanter y col. (1984) Nucleic Acids Res., 12: 6159-6168. La purificación de oligonucleótidos, cuando sea necesaria, normalmente se realiza mediante electroforesis en gel de acrilamida nativa o mediante HPLC de intercambio aniónico como se describe en Pearson y Regnier (1983) J. Chrom. 255:137-149. La secuencia de los oligonucleótidos sintéticos se puede verificar utilizando el método de degradación química de Maxam y Gilbert (1980) en Grossman y Moldave (eds.) Academic Press, Nueva York, Methods in Enzymology 65:499-560.

20

[0093] Las sondas son más fáciles de preparar combinando y marcando una o más de las construcciones listadas en las Tablas 1 y 2. Antes de su uso, las construcciones se fragmentan para proporcionar fragmentos de ácidos nucleicos más pequeños que penetran fácilmente en la célula y se hibridan al ácido nucleico diana. La fragmentación se puede realizar mediante cualquiera de una serie de procedimientos muy conocidos por los expertos en la materia. Los procedimientos preferidos incluyen el tratamiento con una enzima de restricción para escindir selectivamente las moléculas, o alternativamente el tratamiento térmico de los ácidos nucleicos durante un corto periodo de tiempo en presencia de Mg^{2+} . La sonda se fragmenta preferentemente hasta una longitud media del fragmento que abarca entre las 50 pb aproximadamente y las 2000 pb, más preferentemente entre 100 pb aproximadamente y 1000 pb aproximadamente, y lo más preferentemente entre 150 pb aproximadamente y 500 pb aproximadamente.

30

[0094] Alternativamente, la sonda se pueden producir amplificando (por ejemplo, mediante PCR) subsecuencias seleccionadas procedentes del amplicón 20q13 descrito en el presente documento. Las secuencias proporcionadas en el presente documento permiten que la persona experta seleccione cebadores que amplifiquen secuencias a partir de uno o más exones localizados dentro del amplicón 20q13.

35

[0095] Sondas particularmente preferidas incluyen ácidos nucleicos procedentes de las sondas 38, 40, y 79, que corresponden a RMC20C001. Además, los ADNc son particularmente útiles para identificar células que presentan una expresión incrementada de los genes correspondientes, utilizando por ejemplo, análisis de transferencia de Northern.

40

[0096] La persona experta apreciará que con la utilización de la información de secuencia y los clones proporcionados en el presente documento, el experto en la materia puede aislar las mismas sondas o similares a partir de otras librerías genómicas humanas utilizando procedimientos rutinarios (por ejemplo, transferencias de Southern o Northern).

45

[0097] De forma similar, los polipéptidos descritos en el presente documento se pueden preparar sintéticamente mediante una amplia variedad de procedimientos muy conocidos. Por ejemplo, polipéptidos de longitud relativamente corta se pueden sintetizar en disolución o en un soporte sólido de acuerdo con técnicas convencionales. Véase, por ejemplo, Merrifield (1963) J. Am. Chem. 85:2149-2154. Existen diversos sintetizadores automáticos que se encuentran disponibles comercialmente y se pueden utilizar de acuerdo con protocolos conocidos. Véase, por ejemplo, Stewart y Young (1984) Solid Phase Peptide Synthesis, 2ª ed., Pierce Chemical Co. Como se describe con mayor detalle en el presente documento, los polipéptidos aquí descritos se preparan más preferentemente utilizando técnicas recombinantes, por ejemplo, mediante la expresión de los polipéptidos en células hospedadoras y la purificación de las proteínas expresadas.

50

[0098] En una forma de realización preferida, los polipéptidos, o sus subsecuencias, se sintetizan utilizando metodologías de ADN recombinante. En general, esto supone la creación de una secuencia de ADN que codifica la proteína, mediante técnicas recombinantes, sintéticas, o de amplificación *in vitro*, la colocación del ADN en un casete

55

de expresión bajo el control de un promotor particular, la expresión de la proteína en una célula hospedadora, el aislamiento de la proteína expresada y, si fuera necesario, la renaturalización de la proteína.

Marcaje de ácidos nucleicos

5

[0099] Los procedimientos de marcaje de ácidos nucleicos (ya sean sondas o muestras de ácidos nucleicos) son muy conocidos por los expertos en la materia. Los marcadores preferidos son aquellos que son adecuados para su utilización en hibridación *in situ*. Las sondas o muestras de ácidos nucleicos descritas en el presente documento se pueden marcar de manera detectable antes de la reacción de hibridación. Alternativamente, se puede utilizar un

10

marcador detectable que se una al producto de hibridación. Dichos marcadores detectables incluyen cualquier material que presente una propiedad física o química detectable y que esté muy desarrollado en el ámbito de los inmunoensayos.

[0100] Como se usa en el presente documento, un "marcador" es cualquier composición detectable por medios

15

espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos o químicos. Marcadores útiles incluyen marcadores radiactivos (por ejemplo, ^{32}P , ^{125}I , ^{14}C , ^3H , y ^{35}S), colorantes fluorescentes (por ejemplo, fluoresceína, rodamina, Texas Red, etc.), reactivos densos a los electrones (por ejemplo, oro), enzimas (como las utilizadas habitualmente en un ELISA), marcadores colorimétricos (por ejemplo, oro coloidal), marcadores magnéticos (por ejemplo, Dynabeads™), y similares. Ejemplos de marcadores que no se detectan directamente, pero que son detectables

20

mediante la utilización de marcadores directamente detectables incluyen biotina y dioxigenina, así como haptenos y proteínas para los que hay disponibles antisueros o anticuerpos monoclonales marcados.

[0101] El marcador particular utilizado no es crítico, mientras no interfiera con la hibridación *in situ* de la cepa. No obstante, se prefieren cepas marcadas directamente con marcadores fluorescentes (por ejemplo, fluoresceína-12-

25

dUTP, Texas Red-5-dUTP, etc.) para la hibridación de cromosomas.

[0102] Una sonda marcada directa, como se usa en el presente documento, es una sonda a la que se ha unido un marcador detectable. Debido a que el marcador directo ya está unido a la sonda, no son necesarias etapas adicionales para asociar la sonda con el marcador detectable. En contraste, una sonda marcada indirecta es aquella

30

que porta un resto al cual se le une posteriormente un marcador detectable, normalmente después de que la sonda se hibride con el ácido nucleico diana.

[0103] Además, el marcador debe ser detectable en el menor número de copias que sea posible, maximizando así la sensibilidad del ensayo y también detectable por encima de cualquier señal de fondo. Por último, se debe

35

seleccionar un marcador que proporcione una señal altamente localizada, proporcionando así un alto grado de resolución espacial cuando la tinción se mapea físicamente contra el cromosoma. Marcadores fluorescentes particularmente preferidos incluyen la fluoresceína-12-dUTP y Texas Red-5-dUTP.

[0104] Los marcadores se pueden acoplar a las sondas mediante diversos medios conocidos por los expertos en

40

la materia. En una forma de realización preferida, las sondas de ácidos nucleicos se marcarán utilizando la traslación de mellas o la extensión aleatoria del cebador (Rigby, y col. J. Mol. Biol., 113: 237 (1977) o Sambrook, y col., Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1985)).

[0105] La persona experta en la materia apreciará que las sondas descritas en el presente documento no tienen

45

por qué ser absolutamente específicas para la región diana 20q13 del genoma. Más bien, las sondas están destinadas a producir una "contraste de tinción". El "contraste" se cuantifica mediante la relación de la intensidad de la sonda de la región diana del genoma con la de las otras porciones del genoma. Por ejemplo, se puede utilizar una librería de ADN producida por clonación de un cromosoma particular (por ejemplo, el cromosoma 7) como colorante capaz de teñir todo el cromosoma. La librería contiene tanto secuencias encontradas sólo en ese cromosoma, como

50

secuencias compartidas con otros cromosomas. Aproximadamente la mitad del ADN cromosómico entra dentro de cada clase. Si la hibridación de toda la librería fuera capaz de saturar todos los sitios de unión en el cromosoma diana, el cromosoma diana sería dos veces más brillante (relación de contraste de 2) que los otros cromosomas, ya que contendría la señal procedente tanto de las secuencias específicas como de las compartidas en la tinción, mientras que los otros cromosomas sólo estarían teñidos por las secuencias compartidas. Por lo tanto, sólo una

55

modesta disminución en la hibridación de las secuencias compartidas en la tinción mejoraría sustancialmente el contraste. Así, se pueden tolerar en la tinción secuencias contaminantes que sólo se hibridan con secuencias no diana, por ejemplo, las impurezas en una librería, en la medida en que las secuencias no reduzcan el contraste de tinción por debajo de niveles útiles.

Detección del amplicón 20q13.

[0106] Como se ha explicado anteriormente, la detección de la amplificación en el amplicón 20q13 es indicativa de la presencia y/o pronóstico de un gran número de cánceres. Éstos incluyen, pero no están limitados a cáncer de mama, ovario, vejiga, cabeza y cuello, y colon.

[0107] En una forma de realización preferida, se detecta una amplificación de 20q13 mediante la hibridación de una sonda a un ácido nucleico diana (por ejemplo, una muestra cromosómica) en la que se desea seleccionar para la detectar la amplificación. Los formatos de hibridación adecuados son muy conocidos por los expertos en la materia e incluyen, pero no están limitados a, las variaciones de transferencias de Southern, hibridación *in situ* y métodos de amplificación cuantitativos tales como la PCR cuantitativa (véase, por ejemplo, Sambrook, más arriba, Kallioniemi y col., Proc. Natl Acad Sci USA, 89: 5321-5325 (1992), y PCR Protocols, A Guide to Methods and Applications, Innis y col., Academic Press, Inc. N.Y., (1990)).

15 Hibridación *in situ*

[0108] En una forma de realización preferida, el amplicón 20q13 se identifica utilizando hibridación *in situ*. Generalmente, la hibridación *in situ* comprende las siguientes etapas principales: (1) fijación del tejido o de la estructura biológica a analizar, (2) tratamiento de prehibridación de la estructura biológica para aumentar la accesibilidad del ADN diana, y para reducir la unión no específica, (3) hibridación de la mezcla de ácidos nucleicos al ácido nucleico en la estructura biológica o tejido, (4) lavados de posthibridación para retirar los fragmentos de ácidos nucleicos no unidos en la hibridación y (5) detección de los fragmentos de ácidos nucleicos hibridados. El reactivo utilizado en cada una de estas etapas y sus condiciones de uso varían dependiendo de la aplicación particular.

[0109] En algunas aplicaciones es necesario bloquear la capacidad de hibridación de secuencias repetitivas. En este caso, el ADN genómico humano se utiliza como agente para bloquear dicha hibridación. El intervalo de tamaños preferidos abarca entre 200 pb aproximadamente y 1000 pb aproximadamente, más preferiblemente entre 400 pb aproximadamente y 800 pb aproximadamente para ácidos nucleicos con traslación de mella de doble cadena.

[0110] Protocolos de hibridación para las aplicaciones particulares descritas en el presente documento se describen en Pinkel y col. Proc. Natl. Acad. Sci. USA, 85: 9138-9142 (1988) y en el documento Pub EPO. N° 430.402. También se pueden encontrar protocolos de hibridación adecuados en Methods in Molecular Biology Vol. 33: In Situ Hybridization Protocols, K.H.A. Choo, ed., Humana Press, Totowa, Nueva Jersey, (1994). En una forma de realización particularmente preferida, se utiliza el protocolo de hibridación de Kallioniemi y col., Proc. Natl Acad Sci USA, 89: 5321-5325 (1992).

[0111] Normalmente, es deseable el uso de la FISH de dos colores, en donde se utilizan dos sondas, cada una marcada mediante un colorante fluorescente diferente. Una sonda de prueba que se hibrida a la región de interés se marca con un colorante, y una sonda control que se hibrida a una región diferente se marca con un segundo colorante. A menudo es más útil como sonda control un ácido nucleico que se hibride a una porción estable del cromosoma de interés, tal como la región del centrómero. De esta forma, se pueden explicar las diferencias en la eficiencia de hibridación entre muestras.

[0112] Los procedimientos de FISH para detectar anomalías cromosómicas se pueden realizar en el rango de cantidades de nanogramos de los ácidos nucleicos en cuestión. Se pueden utilizar secciones tumorales embebidas en parafina, así como material fresco o congelado. Debido a que la FISH se puede aplicar al material limitado, también se pueden utilizar improntas sobre portas preparados a partir de tumores primarios sin cultivar (véase, por ejemplo, Kallioniemi, A. y col., Cytogenet. Cell Genet. 60: 190-193 (1992)). Por ejemplo, se pueden utilizar pequeñas muestras de tejido de biopsias procedentes de tumores para improntas sobre portas (véase, por ejemplo, Kallioniemi, A. y col., Cytogenet. Cell Genet. 60: 190-193 (1992)). También se puede analizar un pequeño número de células obtenidas por aspiración de la biopsia o de células en fluidos corporales (por ejemplo, sangre, orina, esputo y similares). Para el diagnóstico prenatal, las muestras adecuadas incluirán líquido amniótico y similares.

55 Otros formatos

[0113] Pueden ser útiles una serie de formatos de hibridación. Por ejemplo, se pueden utilizar hibridaciones de Southern. En una transferencia de Southern, un ADN genómico o ADNc (normalmente fragmentado y separado en una electroforesis en gel) se hibrida con una sonda específica para la región diana. La comparación de la intensidad

de la señal de hibridación procedente de la sonda para la región diana (por ejemplo, 20q13) con la señal procedente de una sonda dirigida a un control (no amplificada) tal como ADN centromérico, proporciona una estimación del número relativo de copias del ácido nucleico diana.

5 **[0114]** Otros formatos utilizan biochips de sondas o dianas a las que se hibridan las muestras de ácidos nucleicos, como se describe por ejemplo, en el documento WO 96/17958. Como se usa en el presente documento, un "biochip de ácidos nucleicos" es una pluralidad de elementos diana, cada uno que comprende una o más moléculas de ácidos nucleicos diana inmovilizadas sobre una superficie sólida a la que se hibridan los ácidos nucleicos de la sonda. Los ácidos nucleicos diana de un elemento diana normalmente tendrán su origen en el amplicón 20q13. Los
10 ácidos nucleicos diana de un elemento diana pueden contener, por ejemplo, una secuencia de genes o clones específicos descritos en el presente documento. Se pueden utilizar elementos diana de diferentes dimensiones en los biochips descritos en el presente documento. En general, se prefieren elementos diana más pequeños. Normalmente, un elemento diana será inferior a 1 cm de diámetro aproximadamente. Generalmente el elemento abarca entre 1 μm aproximadamente y 3 mm aproximadamente, preferentemente entre 5 μm y 1 mm
15 aproximadamente.

[0115] Los elementos diana de los biochips se pueden disponer sobre la superficie sólida a diferentes densidades. Las densidades del elemento diana dependerán de una serie de factores, tales como la naturaleza del marcador, el soporte sólido, y similares. La persona experta reconocerá que cada elemento diana puede comprender una mezcla
20 de ácidos nucleicos diana de diferentes longitudes y secuencias. Así, por ejemplo, un elemento diana puede contener más de una copia de un fragmento de ADN clonado, y cada copia se puede romper en fragmentos de diferentes longitudes. La longitud y complejidad de las secuencias diana no es crítica. La persona experta puede ajustar estos factores para proporcionar una hibridación y una producción de señal óptimas para un procedimiento de hibridación dado, y para proporcionar la resolución necesaria entre diferentes genes o localizaciones genómicas.
25 Normalmente, las secuencias diana tendrán una complejidad de entre 1 kb aproximadamente y 1 Mb aproximadamente, en algunas ocasiones entre 10 kb y 500 kb aproximadamente, y normalmente entre 50 kb aproximadamente y 150 kb aproximadamente.

Detección de mutaciones en genes procedentes del amplicón 20q13

30 **[0116]** Las secuencias de ADNc descritas en el presente documento también se pueden utilizar para detectar mutaciones (por ejemplo, sustituciones, inserciones, y eliminaciones) dentro de los genes endógenos correspondientes. La persona experta reconocerá que las técnicas de hibridación de ácidos nucleicos descritas de manera general más arriba se pueden adaptar para detectar muchas de dichas mutaciones. Por ejemplo, en
35 ensayos de hibridación habituales se pueden utilizar sondas de oligonucleótidos que distinguen entre las formas mutante y silvestre del gen diana. En algunas formas de realización se puede utilizar la amplificación (por ejemplo, usando PCR) para incrementar el número de copias de la secuencia diana antes de la hibridación.

Ensayos para detectar las proteínas del amplicón 20q13

40 **[0117]** Como se ha indicado anteriormente, esta invención identifica productos proteicos de genes en el amplicón 20q13 que están asociados a diversos cánceres. En particular, se ha demostrado que las proteínas 20q13 estaban sobre-expresadas en diversos cánceres. La presencia o ausencia y/o nivel de expresión de proteínas 20q13 puede ser indicativo de la presencia, ausencia, o grado de un cáncer. Así, las proteínas 20q13 pueden proporcionar
45 marcadores diagnósticos útiles.

[0118] Las proteínas del amplicón 20q13 (por ejemplo, ZABC1 o 1b1) se pueden detectar y cuantificar mediante cualquiera de una serie de medios muy conocidos por los expertos en la materia. Éstos pueden incluir procedimientos bioquímicos analíticos tales como electroforesis, electroforesis capilar, cromatografía líquida de alto
50 rendimiento (HPLC), cromatografía de capa fina (TLC), cromatografía de hiperdifusión, y similares, o diversos procedimientos inmunológicos tales como reacciones en líquido o en gel con precipitina, inmunodifusión (sencilla o doble), inmunoelectroforesis, radioinmunoensayo (RIA), enzoinmunoanálisis de adsorción (ELISA), ensayos inmunofluorescentes, transferencia de Western, y similares.

55 **[0119]** En una forma de realización preferida, las proteínas del amplicón 20q13 se detectan en una separación electroforética de proteínas tal como electroforesis monodimensional o bidimensional, mientras que en una forma de realización más preferida, las proteínas del amplicón 20q13 se detectan utilizando un inmunoensayo.

[0120] Como se usa en el presente documento, un inmunoensayo es un ensayo que utiliza un anticuerpo que se

une específicamente al analito (por ejemplo, las proteínas ZABC1 o Ib1). Así, el inmunoensayo se caracteriza por la detección de la unión específica de una proteína del amplicón 20q13 a un anticuerpo dirigido contra el amplicón 20q13 (por ejemplo, anti-ZABC1 o anti-1b1) en oposición al uso de otras propiedades físicas o químicas para aislar, identificar, o cuantificar el analito.

5

[0121] A continuación se describe con más detalle una colección de muestras biológicas y su posterior ensayo para proteína(s) del amplicón 20q13.

A) Recogida de muestras y procesamiento

10

[0122] Las proteínas del amplicón 20q13 se cuantifican preferentemente en una muestra biológica procedente de un mamífero, más preferentemente procedente de un paciente humano o de un animal porcino, murino, felino, canino, o bovino. Como se usa en el presente documento, una muestra biológica es una muestra de tejido o fluido biológico que contiene una concentración de la proteína del amplicón 20q13 que se puede correlacionar con una
15 amplificación de 20q13. Las muestras biológicas particularmente preferidas incluyen, pero no están limitadas a fluidos biológicos tales como sangre u orina, o muestras de tejido que incluyen, pero no están limitadas a, muestras de biopsias de tejido (por ejemplo, biopsia con aguja).

[0123] La muestra biológica se puede tratar previamente según sea necesario mediante dilución en una disolución tampón apropiada o se puede concentrar, si se desea. Se puede utilizar cualquiera de una serie de disoluciones tampón acuosas patrón, empleando cualquiera de una serie de tampones, tales como tampón fosfato, Tris, o similares, a pH fisiológico.

20

B) Ensayos electroforéticos

25

[0124] Como se ha indicado anteriormente, la presencia o ausencia de proteínas del amplicón 20q13 en un tejido biológico se puede determinar utilizando procedimientos electroforéticos. Los medios para detectar proteínas utilizando técnicas electroforéticas son muy conocidos por los expertos en la materia (véase, en general, R Scopes (1982) Protein Purification, Springer-Verlag, N.Y; Deutscher, (1990) Methods in Enzymology Vol. 182: Guide to
30 Protein Purification, Academic Press, Inc. N.Y.). En una forma de realización preferida, las proteínas del amplicón 20q13 se detectan utilizando electroforesis monodimensional o bidimensional. Una separación mediante electroforesis bidimensional particularmente preferida se basa en el enfoque isoelectrico (IEF) en gradientes de pH inmovilizados para una dimensión y geles de poliacrilamida para la segunda dimensión. Dichos ensayos se describen en las referencias citadas y por Patton y col. (1990) Biotechniques 8: 518.

35

C) Ensayos inmunológicos de unión

[0125] En una forma de realización preferida, el amplicón 20q13 se detecta y/o cuantifica utilizando cualquiera de una serie de ensayos inmunológicos de unión muy reconocidos (véanse, por ejemplo, patentes de EE.UU. 4.366.241; 4.376.110; 4.517.288; y 4.837.168). Para una revisión de los inmunoensayos generales, véase también
40 Methods in Cell Biology Volume 37: Antibodies in Cell Biology, Asai, ed. Academic Press, Inc. Nueva York (1993); Basic and Clinical Immunology 7th Edition, Stites & Terr, eds. (1991).

[0126] Los ensayos inmunológicos de unión (o inmunoensayos) normalmente utilizan un "agente de captura" para unirse específicamente y, a menudo, inmovilizar al analito (en este caso el amplicón 20q13). El agente de captura es un resto que se une específicamente al analito. En una forma de realización preferida, el agente de captura es un anticuerpo que se une específicamente a las proteína(s) del amplicón 20q13.

45

[0127] El anticuerpo (por ejemplo, anti-ZABC1 o anti-1b1) se puede producir mediante cualquiera de una serie de medios muy conocidos por los expertos en la materia (véase, por ejemplo, Methods in Cell Biology Volume 37: Antibodies in Cell Biology, Asai, ed. Academic Press, Inc. Nueva York (1993); y Basic and Clinical Immunology 7th Edition, Stites & Terr, eds. (1991)). El anticuerpo puede ser un anticuerpo completo o un fragmento de anticuerpo. Puede ser policlonal o monoclonal, y se puede producir exponiendo a un organismo (por ejemplo, ratón, rata, conejo, etc.) a una proteína del amplicón 20q13 o a un epítipo derivado de ella. Alternativamente, el anticuerpo se puede
50 producir *de novo* utilizando metodologías de ADN recombinante. El anticuerpo también se puede seleccionar a partir de una librería de presentación fágica seleccionada contra el amplicón 20q13 (véase, por ejemplo, Vaughan y col. (1996) Nature Biotechnology, 14: 309-314 y las referencias allí citadas).

55

[0128] Los inmunoensayos también utilizan a menudo un agente marcador que se une específicamente y que

marca el complejo de unión formado por el agente de captura y el analito. El agente marcador puede ser él mismo uno de los restos que comprenden el complejo anticuerpo/analito. Así, el agente marcador puede ser una proteína marcada del amplicón 20q13 o un anticuerpo marcado dirigido contra el amplicón 20q13. Alternativamente, el agente marcador puede ser un tercer resto, tal como otro anticuerpo, que se une específicamente al complejo anticuerpo/proteína del amplicón 20q13. En una forma de realización preferida, el agente marcador es un segundo anticuerpo humano que porta un marcador dirigido contra una proteína del amplicón 20q13. Alternativamente, el segundo anticuerpo dirigido contra una proteína del amplicón 20q13 puede carecer de marcador, pero en cambio, puede estar unido por un tercer anticuerpo marcado específico para anticuerpos de las especies de las que procede el segundo anticuerpo. El segundo anticuerpo se puede modificar con un resto detectable, tal como biotina, al que se puede unir específicamente una tercera molécula marcada, tal como estreptavidina marcada enzimáticamente.

[0129] También se puede utilizar como agente marcador otras proteínas capaces de unir específicamente las regiones constantes de inmunoglobulinas, tales como la proteína A o la proteína G. Estas proteínas son constituyentes normales de las paredes celulares de bacterias estreptocócicas. Presentan una fuerte reactividad no inmunógena con las regiones constantes de las inmunoglobulinas procedentes de diversas especies. Véase, en general, Kronval, y col., J. Immunol., 111:1401-1406 (1973), y Akerstrom, y col., J. Immunol., 135:2589-2542 (1985).

[0130] A lo largo de estos ensayos, pueden ser necesarias etapas de incubación y/o lavado después de cada combinación de reactivos. Las etapas de incubación pueden abarcar desde 5 segundos aproximadamente a varias horas, preferentemente desde 5 minutos aproximadamente a 24 horas aproximadamente. No obstante, el tiempo de incubación dependerá del formato del ensayo, del analito, del volumen de disolución, de las concentraciones, y similares. Normalmente, los ensayos se llevarán a cabo a temperatura ambiente, aunque se pueden realizar en un rango de temperaturas, tales como de 10°C a 40°C.

25 **1) Formatos de ensayo no competitivo**

[0131] Los inmunoensayos para detectar las proteínas del amplicón 20q13 pueden ser competitivos o no competitivos. Los inmunoensayos no competitivos son ensayos en los que la cantidad de analito capturado (en este caso, el amplicón 20q13) se mide directamente. En un ensayo en "sándwich" preferido, por ejemplo, el agente de captura (anticuerpos dirigidos contra una proteína del amplicón 20q13) puede estar unido directamente a un sustrato sólido al que están movilizados. A continuación estos anticuerpos inmovilizados capturan la proteína del amplicón 20q13 presente en la muestra de prueba. A la proteína del amplicón 20q13 inmovilizada de esta forma se le une a continuación un agente marcador, tal como un segundo anticuerpo humano que porta un marcador dirigido contra una proteína del amplicón 20q13. Alternativamente, el segundo anticuerpo dirigido contra la proteína del amplicón 20q13 puede carecer de marcador, pero en cambio, puede unirse a un tercer anticuerpo marcado específico para anticuerpos de las especies de las que procede el segundo anticuerpo. El segundo anticuerpo se puede modificar con un resto detectable, tal como biotina, al que se puede unir específicamente una tercera molécula marcada, tal como estreptavidina marcada enzimáticamente.

40 **2) Formatos de ensayo competitivo**

[0132] En ensayos competitivos, la cantidad de analito (proteína del amplicón 20q13) presente en la muestra se mide indirectamente midiendo la cantidad de un analito añadido (exógeno) (proteínas del amplicón 20q13 tales como la proteína ZABC1 o lb1) desplazada (apartada por competición) de un agente de captura (por ejemplo, anticuerpo anti-ZABC1 o anti-lb1) por el analito presente en la muestra. En un ensayo competitivo, se añade una cantidad conocida de, en este caso, proteína del amplicón 20q13 a la muestra y la muestra se pone en contacto con un agente de captura, en este caso, un anticuerpo que se une específicamente a la proteína del amplicón 20q13. La cantidad de proteína del amplicón 20q13 unida al anticuerpo es inversamente proporcional a la concentración de proteína del amplicón 20q13 presente en la muestra.

[0133] En una forma de realización particularmente preferida, el anticuerpo dirigido contra la proteína 20q13 se inmoviliza sobre un sustrato sólido. La cantidad de proteína del amplicón 20q13 unida al anticuerpo se puede determinar midiendo la cantidad de amplicón 20q13 presente en un complejo proteína del amplicón 20q13/anticuerpo, o alternativamente midiendo la cantidad de proteína del amplicón 20q13 que permanece sin complejarse. La cantidad de proteína del amplicón 20q13 se puede detectar proporcionando una proteína marcada del amplicón 20q13.

[0134] Otro ensayo competitivo preferido es el ensayo de inhibición del hapteno. En este ensayo, un analito conocido, en este caso la proteína del amplicón 20q13, se inmoviliza sobre un sustrato sólido. Se añade una

cantidad conocida de anticuerpo dirigido contra la proteína del amplicón 20q13 a la muestra, y la muestra se pone en contacto con la proteína del amplicón 20q13 inmovilizada. En este caso, la cantidad de anticuerpo dirigido contra la proteína del amplicón 20q13 unida a la proteína del amplicón 20q13 inmovilizada es inversamente proporcional a la cantidad de proteína del amplicón 20q13 presente en la muestra. De nuevo, la cantidad de anticuerpo inmovilizado se puede detectar detectando la fracción de anticuerpo inmovilizado o la fracción de anticuerpo que permanece en disolución. La detección puede ser directa, en la que el anticuerpo está marcado, o indirecta, mediante la posterior adición de un resto marcado que se une específicamente al anticuerpo como se ha descrito anteriormente.

3) Otros formatos de ensayo

10

[0135] En una forma de realización particularmente preferida, se utiliza el análisis de transferencia de Western (inmunotransferencia) para detectar y cuantificar la presencia de proteína del amplicón 20q13 en la muestra. En general, la técnica comprende la separación de las proteínas de la muestra mediante electroforesis en gel en base a los pesos moleculares, la transferencia de las proteínas separadas a un soporte sólido adecuado (tal como un filtro de nitrocelulosa, un filtro de nailon, o un filtro de nailon derivado), y la incubación de la muestra con los anticuerpos que se unen específicamente a la proteína del amplicón 20q13. Los anticuerpos dirigidos contra la proteína del amplicón 20q13 se unen específicamente a la proteína del amplicón 20q13 sobre el soporte sólido. Estos anticuerpos se pueden marcar directamente o, alternativamente, se pueden detectar posteriormente utilizando anticuerpos marcados (por ejemplo, anticuerpos de oveja marcados dirigidos contra ratón) que se unen específicamente al anticuerpo dirigido contra la proteína del amplicón 20q13.

[0136] Otros formatos de ensayo incluyen inmunoensayos de liposomas (LIA), que utilizan liposomas diseñados para unir moléculas específicas (por ejemplo, anticuerpos) y liberar reactivos o marcadores encapsulados. A continuación los compuestos químicos liberados se detectan de acuerdo con técnicas habituales (véase, Monroe y col. (1986) Amer. Clin. Prod. Rev. 5:34-41).

25

D) Reducción de unión no específica

[0137] La persona experta apreciará que a menudo es deseable reducir la unión no específica en los inmunoensayos. Particularmente, cuando el ensayo implica a un antígeno o anticuerpo inmovilizado sobre un sustrato sólido, es deseable minimizar la cantidad de unión no específica al sustrato. Los medios para reducir dicha unión no específica son muy conocidos por los expertos en la materia. Normalmente, esto supone el recubrimiento del sustrato con una composición proteácea. En particular, son de uso generalizado composiciones de proteínas tales como seroalbúmina bovina (BSA), leche desnatada en polvo, y gelatina, siendo preferido el uso de la leche en polvo.

35

E) Marcadores

[0138] El marcador o grupo detectable particular utilizado en el ensayo no es crítico, mientras no interfiera significativamente con la unión específica del anticuerpo utilizado en el ensayo. El grupo detectable puede ser cualquier material que presente una propiedad física o química detectable. Dichos marcadores detectables están muy desarrollados en el ámbito de los inmunoensayos y, en general, se pueden aplicar la mayoría de los marcadores útiles en dichos procedimientos. Así, un marcador es cualquier composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunquímicos, eléctricos, ópticos o químicos. Los marcadores útiles incluyen perlas magnéticas (por ejemplo, Dynabeads™), colorantes fluorescentes (por ejemplo, isotiocianato de fluoresceína, Texas Red, rodamina, y similares), radiomarcadores (por ejemplo, ³H, ¹²⁵I, ³⁵S, ¹⁴C, o ³²P), enzimas (por ejemplo, peroxidasa de rábano, fosfatasa alcalina y otros utilizados habitualmente en ELISA), y marcadores colorimétricos tales como oro coloidal o perlas coloreadas de vidrio o plástico (por ejemplo, poliestireno, polipropileno, látex, etc.).

45

[0139] El marcador se puede acoplar directa o indirectamente al componente deseado del ensayo de acuerdo con procedimientos muy conocidos en la materia. Como se ha indicado anteriormente, se pueden utilizar una amplia variedad de marcadores, cuya selección depende de la sensibilidad requerida, la facilidad de conjugación con el compuesto, requerimientos de estabilidad, instrumentación disponible, y disposiciones de eliminación.

50

[0140] Los marcadores no radiactivos normalmente se unen por medios indirectos. En general, una molécula ligando (por ejemplo, biotina) está covalentemente unida a la molécula. A continuación el ligando se une a una molécula anti-ligando (por ejemplo, estreptavidina) que es inherentemente detectable o está covalentemente unida a un sistema de señalización, tal como una enzima detectable, un compuesto fluorescente, o un compuesto

55

quimioluminiscente. Se pueden utilizar una serie de ligandos y anti-ligandos. Cuando un ligando tenga un anti-ligando natural, por ejemplo, biotina, tiroxina, y cortisol, se puede utilizar junto con los anti-ligandos de origen natural marcados. Alternativamente, se puede utilizar cualquier compuesto hapténico o antigénico en combinación con un anticuerpo.

5

[0141] Las moléculas también se pueden conjugar directamente con compuestos que generan señales, por ejemplo, mediante su conjugación con una enzima o fluoróforo. Enzimas de interés como marcadores serán principalmente hidrolasas, particularmente fosfatasas, esterasas y glicosidasas, u óxido-reductasas, particularmente peroxidasas. Los compuestos fluorescentes incluyen fluoresceína y sus derivados, rodamina y sus derivados, dansilo, umbeliferona, etc. Los compuestos quimioluminiscentes incluyen luciferina, y 2,3-dihidroftalazindionas, por ejemplo, luminol. Para una revisión de diversos sistemas marcadores o productores de señales que se pueden utilizar, véase, patente de EE.UU. N° 4.391.904.

[0142] Los medios para detectar marcadores son muy conocidos por los expertos en la materia. Así, por ejemplo, cuando el marcador sea un marcador radiactivo, los medios para su detección incluyen un contador de centelleo o una película fotográfica como en autorradiografía. Cuando el marcador sea un marcador fluorescente, se puede detectar excitando el fluorocromo con la longitud de onda de luz adecuada y detectando la fluorescencia resultante. La fluorescencia se puede detectar visualmente, por medio de una película fotográfica, mediante el uso de detectores electrónicos tales como dispositivos de carga acoplada (CCDs) o fotomultiplicadores y similares. De manera similar, los marcadores enzimáticos se pueden detectar proporcionando los sustratos adecuados para la enzima y detectando el producto de reacción resultante. Por último, los marcadores colorimétricos simples se pueden detectar observando simplemente el color asociado al marcador. Así, en diversos ensayos con tiras reactivas, el oro conjugado a menudo aparece en rosa, mientras que diversas perlas conjugadas aparecen del color de la perla.

[0143] Algunos formatos de ensayo no requieren la utilización de componentes marcados. Por ejemplo, se pueden utilizar ensayos de aglutinación para detectar la presencia de los anticuerpos diana. En este caso, partículas recubiertas de antígeno se aglutinan mediante muestras que comprenden los anticuerpos diana. En este formato, ninguno de los componentes tiene que estar marcado y la presencia del anticuerpo diana se detecta mediante una simple inspección visual.

30

G) Sustratos

[0144] Como se ha mencionado anteriormente, dependiendo del ensayo, diversos componentes, incluyendo el antígeno, el anticuerpo diana, o el anticuerpo dirigido contra seres humanos, pueden estar unidos a una superficie sólida. En la técnica se conocen numerosos procedimientos para inmovilizar biomoléculas a una variedad de superficies sólidas. Por ejemplo, la superficie sólida puede ser una membrana (por ejemplo, nitrocelulosa), o una placa de microtitulación (por ejemplo, PVC, polipropileno, o poliestireno), un tubo de ensayo (de vidrio o plástico), una tira reactiva (por ejemplo, de vidrio, PVC, polipropileno, poliestireno, látex, y similares), un tubo de microcentrífuga, o una perla de vidrio o plástico. El componente deseado puede estar unido covalentemente o fijado no covalentemente mediante uniones no específicas.

[0145] Se pueden utilizar una amplia variedad de polímeros orgánicos e inorgánicos, tanto naturales como sintéticos, como material para la superficie sólida. Polímeros ilustrativos incluyen polietileno, polipropileno, poli(4-metilbuteno), poliestireno, polimetacrilato, poli(tereftalato de etileno), rayón, nailon, poli(butirato de vinilo), difluoruro de polivinilideno (PVDF), siliconas, poliformaldehído, celulosa, acetato de celulosa, nitrocelulosa, y similares. Otros materiales que pueden emplearse, incluyen papel, vidrios, compuestos cerámicos, metales, metaloides, materiales semiconductores, cementos o similares. Además, se incluyen sustancias que forman geles, tales como proteínas (por ejemplo, gelatinas), lipopolisacáridos, silicatos, agarosa y poli(acrilamidas). También son adecuados los polímeros que forman varias fases acuosas, tales como dextranos, glicoles de polialquileño o tensioactivos, tales como fosfolípidos, sales de alquilamonio de cadena larga (12-24 átomos de carbono) y similares. Cuando la superficie sólida sea porosa, se pueden utilizar diversos tamaños de poro dependiendo de la naturaleza del sistema.

[0146] Para la preparación de la superficie, se pueden utilizar una pluralidad de diferentes materiales, particularmente en forma de laminados, para obtener diversas propiedades. Por ejemplo, se pueden utilizar recubrimientos de proteínas, tales como gelatinas, para evitar uniones no específicas, simplificar la conjugación covalente, potenciar la detección de las señales, o similares.

[0147] Si se desea la unión covalente entre un compuesto y la superficie, normalmente la superficie será polifuncional o se podrá polifuncionalizar. Grupos funcionales que pueden estar presentes sobre la superficie y que

se pueden utilizar para la unión pueden incluir ácidos carboxílicos, aldehídos, grupos amino, grupos ciano, grupos etilénicos, grupos hidroxilo, grupos mercapto y similares. La manera de unir una amplia variedad de compuestos a diversas superficies es muy conocida y está ampliamente ilustrada en la bibliografía. Véase, por ejemplo, Immobilized Enzymes, Ichiro Chibata, Halsted Press, Nueva York, 1978, y Cuatrecasas (1970) J. Biol. Chem. 245 5 3059.

[0148] Además de la unión covalente, se pueden utilizar diversos procedimientos para la unión no covalente de un componente del ensayo. La unión no covalente normalmente es una absorción no específica de un compuesto a la superficie. Normalmente, la superficie se bloquea con un segundo compuesto para evitar la unión no específica de componentes marcados del ensayo. Alternativamente, la superficie se diseña de manera que una no específicamente un componente, pero que no una significativamente otro componente. Por ejemplo, una superficie que contiene una lecitina, tal como la concanavalina A unirá un compuesto que contiene carbohidratos, pero no una proteína marcada que carezca de glicosilación. En las patentes de EE.UU. N° 4.447.576 y 4.254.082 hay una revisión de diversas superficies sólidas para su utilización en la unión no covalente de componentes de ensayo.

Kits que contienen sondas del amplicón 20q13

[0149] En el presente documento se describen kits diagnósticos para la detección de anomalías cromosómicas en 20q13. Los kits pueden incluir una o más sondas para el amplicón 20q13 y/o anticuerpos para el amplicón 20q13 (por ejemplo, anti-ZABC1 o anti-1b1) descritos en el presente documento. Adicionalmente los kits pueden incluir sondas de bloqueo, y materiales explicativos que describan cómo usar el contenido del kit para la detección de amplicones 20q13. Los kits también pueden incluir uno o más de lo siguiente: diversos marcadores o agentes marcadores para facilitar la detección de las sondas, reactivos para la hibridación incluyendo tampones, un propagador de la metafase, seroalbúmina bovina (BSA) y otros agentes de bloqueo, dispositivos de toma de muestra incluyendo agujas finas, bastoncillos de algodón, aspiradores y similares, controles de hibridación positivos y negativos, etc.

Expresión de los clones de ADNc

[0150] Los polipéptidos deseados codificados se pueden expresar mediante los clones de ADNc descritos en el presente documento, o mediante la subclonación de partes del ADNc de secuencias genómicas en una célula manipulada genéticamente de forma recombinante tales como bacterias, levaduras, insectos (especialmente empleando vectores de baculovirus), o células de mamífero. Se espera que los expertos en la materia conozcan los numerosos sistemas de expresión disponibles para la expresión de los ADNc. No hay ninguna intención de describir con detalle los diversos procedimientos conocidos para la expresión de proteínas en procariontes o eucariontes.

[0151] En un breve resumen, la expresión de ácidos nucleicos naturales o sintéticos que codifican polipéptidos normalmente se conseguirá mediante la unión de manera operable del ADN o ADNc a un promotor (que sea constitutivo o inducible), seguido por la incorporación a un vector de expresión. Los vectores pueden ser adecuados para la replicación e integración tanto en procariontes como eucariontes. Los vectores de expresión típicos contienen terminadores de la transcripción y la traducción, secuencias de iniciación, y promotores útiles para la regulación de la expresión del ADN que codifica los polipéptidos. Para obtener un nivel de expresión elevado de un gen clonado, es deseable construir plásmidos de expresión que contengan, como mínimo, un promotor fuerte para dirigir la transcripción, un sitio de unión a ribosomas para el inicio de la traducción, y un terminador de la transcripción/traducción.

[0152] Ejemplos de regiones reguladoras adecuadas para este propósito en *E. coli* son la región promotora y operadora de la ruta biosintética del triptófano en *E. coli* como describe Yanofsky, C., 1984, J. Bacteriol., 158:1018-1024 y el promotor hacia la izquierda del fago lambda (P1) como describe Herskowitz, I. y Hagen, D., 1980, Ann. Rev. Genet., 14:399-445. También es útil la inclusión de marcadores de selección en vectores de ADN transformados en *E. coli*. Ejemplos de dichos marcadores incluyen genes específicos de resistencia a ampicilina, tetraciclina, o cloranfenicol. Están disponibles sistemas de expresión que utilizan *E. coli*, especies de *Bacillus* (Palva, I y col., 1983, Gene 22:229-235; Mosbach, K. y col. Nature, 302:543-545 y *Salmonella*. Se prefieren los sistemas de *E. coli*.

[0153] Los polipéptidos producidos por células procariontes pueden no necesariamente plegarse de manera adecuada. Durante la purificación a partir de *E. coli*, los polipéptidos expresados en primer lugar se pueden desnaturar y a continuación renaturalizarse. Esto se puede conseguir solubilizando en un agente caotrópico las proteínas producidas en la bacteria, tal como clorhidrato de guanidina y reduciendo todos los restos cisteína con un

agente reductor tal como beta-mercaptoetanol. A continuación los polipéptidos se renaturalizan, bien por diálisis lenta o mediante filtración en gel. Patente de EE.UU. N° 4.511.503.

5 **[0154]** Los expertos en la materia conocerán una variedad de sistemas de expresión en eucariotas tales como levaduras, líneas celulares de insecto y células de mamífero. Como se explica brevemente a continuación, los polipéptidos también se pueden expresar en estos sistemas eucariotas.

10 **[0155]** La síntesis de proteínas heterólogas en levaduras es muy conocida y está perfectamente descrita. *Methods in Yeast Genetics*, Sherman, F., y col., Cold Spring Harbor Laboratory, (1982) es una obra muy reconocida que describe los diversos procedimientos disponibles para producir los polipéptidos en levaduras. Como vectores se pueden utilizar una serie de plásmidos de expresión en levaduras tales como YEp6, YEp13, YEp4. Un gen de interés se puede fusionar con cualquiera de los promotores en diversos vectores de levaduras. Los plásmidos anteriormente mencionados están ampliamente descritos en la bibliografía (Botstein y col., 1979, *Gene*, 8:17-24; Broach. y col., 1979, *Gene*, 8:121-133).

15 **[0156]** Ejemplos ilustrativos de cultivos celulares útiles para la producción de los polipéptidos son células que tienen su origen en insectos o mamíferos. Los sistemas celulares de mamífero a menudo estarán en forma de monocapas de células, aunque también se pueden utilizar suspensiones de células de mamífero. Ejemplos ilustrativos de líneas celulares de mamífero incluyen las células VERO y HeLa, líneas celulares de ovario de hámster chino (CHO), líneas celulares W138, BHK, COS-7 o MDCK.

20

[0157] Como se ha indicado anteriormente, el vector, por ejemplo un plásmido, que se utiliza para transformar la célula hospedadora, preferentemente contiene secuencias de ADN para iniciar la transcripción y secuencias para controlar la traducción de la secuencia del gen antigénico. Estas secuencias se denominan secuencias para el control de la expresión. Cuando una célula hospedadora tiene su origen en insectos o mamíferos, a menudo se obtienen secuencias ilustrativas para el control de la expresión a partir del promotor de SV-40 (*Science*, 222:524-527, 1983), el promotor IE de CMV (*Proc Natl. Acad. Sci.* 81:659-663, 1984) o el promotor de metalotioneína (*Nature* 296:39-42, 1982). El vector de clonación que contiene las secuencias para el control de la expresión se escinde utilizando enzimas de restricción, y su tamaño se ajusta según sea necesario o deseable, y se liga con el ADN deseado por medios muy conocidos en la materia.

25

30

[0158] Al igual que en el caso de las levaduras, cuando se emplean células hospedadoras de animales superiores, es necesario incorporar al vector secuencias de poliadenilación o secuencias terminadoras de la transcripción procedentes de genes conocidos de mamíferos. Un ejemplo de una secuencia terminadora es la secuencia de poliadenilación procedente del gen bovino de la hormona de crecimiento. También se pueden incluir secuencias para un procesamiento alternativo preciso del transcrito. Un ejemplo de una secuencia de procesamiento alternativo es el intrón VPI de SV40 (Sprague, J. y col., 1983, *J. Virol.* 45: 773-781).

35

[0159] Adicionalmente, se pueden incorporar al vector secuencias génicas para controlar la replicación en la célula hospedadora, tales como las encontradas en los vectores de tipo papilomavirus bovino. Saveria-Campo, M., 1985, "Bovine Papilloma virus DNA a Eukaryotic Cloning Vector" en *DNA Cloning Vol. II a Practical Approach* Ed. DM Glover, IRL Press, Arlington, Virginia pp. 213-238.

Usos terapéuticos y de otro tipo de los ADNc y de sus productos génicos

45 **[0160]** Las secuencias de ADNc y los productos polipeptídicos descritos en el presente documento se pueden utilizar para modular la actividad de los productos génicos de los genes endógenos correspondientes a los ADNc. Modulando la actividad de los productos génicos, se pueden tratar dolencias patológicas asociadas a su expresión o a su falta de expresión. Para este fin se puede utilizar cualquiera de una variedad de técnicas muy conocidas por los expertos en la materia.

50

[0161] Los ADNc descritos en el presente documento se pueden utilizar para el tratamiento de diversos cánceres tales como el cáncer de mama, ovario, vejiga, cabeza y cuello, y colon. Con las secuencias descritas en el presente documento también se pueden tratar otras enfermedades. Por ejemplo, como se ha indicado anteriormente, GCAP (SEQ. ID. No. 6) codifica una proteína que activa la guanino ciclasa que está involucrada en la biosíntesis del AMP cíclico. Se sabe que mutaciones en los genes involucrados en la biosíntesis del AMP cíclico están asociadas a enfermedades hereditarias degenerativas de la retina. Estas enfermedades son un grupo de dolencias hereditarias en las que una degeneración progresiva y bilateral de las estructuras de la retina da lugar a la pérdida de la función de la retina. Estas enfermedades incluyen degeneración macular relacionada con la edad, una causa importante de

55

discapacidad visual en ancianos; amaurosis congénita de Leber, que provoca que sus víctimas nazcan ciegas; y retinitis pigmentosa ("RP"), una de las formas más comunes de ceguera hereditaria. La RP es el nombre dado a las retinopatías hereditarias que se caracterizan por la pérdida de fotorreceptores en la retina (conos y bastones), con una reducción en la amplitud de las respuestas eléctricas de la retina a los destellos de luz (es decir, 5 electrorretinogramas o "ERGs").

[0162] El mecanismo de pérdida de fotorreceptores de la retina o muerte celular en diferentes degeneraciones de la retina no se entiende completamente. Se han identificado mutaciones en una serie de genes diferentes como la lesión genética principal en diferentes formas de RP humana. Los genes afectados incluyen la rodopsina, las 10 subunidades alfa y beta de la GMPc fosfodiesterasa, y la perifera-RDS (Dryja, TP y col., Invest. Ophthalmol. Vis. Sci. 36, 1197-1200 (1995)). En todos los casos las manifestaciones de la enfermedad, independientemente de la mutación genética específica primaria, es similar dando como resultado la degeneración de las células fotorreceptoras y ceguera.

15 **[0163]** Estudios en modelos animales de degeneración de la retina han sido el foco de numerosos laboratorios durante la última década. Se han elucidado los mecanismos que se encuentran alterados en algunas de las mutaciones que producen ceguera. Esto incluiría trastornos hereditarios del ratón *rd*. El gen *rd* codifica la subunidad beta de la GMPc-fosfodiesterasa (PDE) (Bowes, C. y col., Nature 347, 677-680 (1990)), una enzima de importancia 20 fundamental en la función visual normal debido a que es un componente clave en la cascada de acontecimientos que tienen lugar en la fototransducción.

[0164] Los polipéptidos codificados por los ADNc descritos en el presente documento se pueden utilizar como inmunógenos para generar anticuerpos policlonales o monoclonales. Los anticuerpos se pueden utilizar para 25 detectar los polipéptidos con fines diagnósticos, como agentes terapéuticos para inhibir a los polipéptidos, o como restos de direccionamiento en inmunotoxinas. La producción de anticuerpos contra un antígeno deseado es muy conocida por aquellos expertos en la materia y no se revisa con detalle en el presente documento.

[0165] Los expertos en la materia reconocerán que existen numerosos procedimientos para la producción y manipulación de diversas moléculas de inmunoglobulina. Como se usa en el presente documento, los términos 30 "inmunoglobulina" y "anticuerpos" se refieren a una proteína constituida por uno o más polipéptidos codificados sustancialmente por genes de inmunoglobulinas. Las inmunoglobulinas pueden existir en una variedad de formas además de como anticuerpos, incluyendo por ejemplo, Fv, Fab y F(ab)₂, así como en cadenas sencillas. Para generar anticuerpos monoclonales, se immortalizan y se seleccionan células que producen anticuerpos obtenidas de animales inmunizados (por ejemplo, ratones), o en primer lugar se seleccionan para la producción del anticuerpo 35 deseado y a continuación se immortalizan. Para una descripción de los procedimientos generales para la producción de anticuerpos monoclonales véase Harlow y Lane, Antibodies, A Laboratory Manual Cold Spring Harbor Publications, Nueva York (1988).

[0166] Los anticuerpos generados mediante estas técnicas se pueden utilizar en ensayos inmunodiagnósticos 40 para detectar o cuantificar la expresión de los productos génicos procedentes de los ácidos nucleicos descritos en el presente documento. Por ejemplo, se pueden utilizar anticuerpos monoclonales marcados contra polipéptidos de la invención para detectar niveles de expresión en una muestra biológica. Para una revisión de los procedimientos generales en inmunoensayos diagnósticos, véase Basic and Clinical Immunology 7ª Edición, D. Stites y A. Terr ed. (1991).

45 **[0167]** Los polinucleótidos descritos en el presente documento son particularmente útiles para técnicas de terapia génica muy conocidos por los expertos en la materia. Terapia génica, como se usa en el presente documento, se refiere a la multitud de técnicas mediante las que se puede alterar la expresión génica en células. Dichos procedimientos incluyen, por ejemplo, la introducción de ribozimas que codifican ADN o ácidos nucleicos antisentido 50 para inhibir la expresión, así como la introducción de genes silvestres funcionales que sustituyan a genes mutantes (por ejemplo, utilizando genes GCAP silvestres para tratar la degeneración de la retina). Se conocen una serie de vectores víricos adecuados. Dichos vectores incluyen vectores retrovirales (véase Miller, Curr. Top. Microbiol. Immunol. 158: 1-24 (1992); Salmons y Gunzburg, Human Gene Therapy 4: 129-141 (1993); Miller y col., Methods in Enzymology 217: 581-599, (1994)) y vectores asociados a adenovirus (revisado en Carter, Curr. Opinion Biotech. 3: 55 533-539 (1992); Muzyczka, Curr. Top. Microbiol. Immunol. 158: 97-129 (1992)). Otros vectores que se pueden utilizar en los procedimientos incluyen vectores adenovíricos, vectores de herpesvirus y vectores víricos Sindbis, como se describe de forma general en, por ejemplo, Jolly, Cancer Gene Therapy 1:51-64 (1994); Latchman, Molec. Biotechnol. 2:179-195 (1994); y Johanning y col., Nucl. Acids Res. 23:1495-1501 (1995).

[0168] También es conocida la administración de ácidos nucleicos ligados a un elemento promotor-potenciador heterólogo mediante liposomas (véase, por ejemplo, Brigham, y col. (1989) Am. J. Med. Sci., 298:278-281; Nabel, y col. (1990) Science, 249:1285-1288; Hazinski, y col. (1991) Am. J. Resp. Cell Molec. Biol., 4:206-209; y Wang y Huang (1987) Proc. Natl. Acad. Sci. (USA), 84:7851-7855); sistemas de transporte basados en cationes y acoplados a ligandos específicos (Wu y Wu (1988) J. Biol. Chem., 263:14621-14624). También se han descrito vectores de expresión de ADN desnudo (Nabel y col. (1990), más arriba); Wolff y col. (1990) Science, 247:1465-1468).

[0169] Los ácidos nucleicos y polipéptidos codificados descritos en el presente documento se pueden utilizar directamente para inhibir a los genes endógenos o a sus productos génicos. Por ejemplo, se pueden utilizar ácidos nucleicos inhibidores para unirse específicamente a una secuencia de ácidos nucleicos complementaria. Con la unión de la secuencia diana adecuada, se forma un dúplex de ARN-ARN, ADN-ADN, o ARN-ADN. Estos ácidos nucleicos a menudo se denominan "antisentido" debido a que normalmente son complementarios a la hebra codificante o sentido del gen, aunque también se han desarrollado enfoques para la utilización de ácidos nucleicos "sentido". El término "ácidos nucleicos inhibidores" como se usa en el presente documento, se refiere tanto a ácidos nucleicos "sentido" como "antisentido". Los procedimientos con ácidos nucleicos inhibidores engloban una serie de enfoques diferentes para alterar la expresión de genes específicos que operan mediante diferentes mecanismos.

[0170] Resumiendo, los enfoques de terapias con ácidos nucleicos inhibidores se pueden clasificar en aquellos dirigidos a secuencias de ADN, aquellos dirigidos a secuencias de ARN (incluyendo ARNprem y ARNm), aquellos dirigidos a proteínas (enfoques con hebras sentido) y aquellos que provocan la escisión o modificación química de los ácidos nucleicos diana (ribozimas). Estos diferentes tipos de tecnología con ácidos nucleicos inhibidores se describen, por ejemplo, en Helene, C. y Toulme, J. (1990) Biochim. Biophys. Acta., 1049:99-125. Se ha demostrado que ácidos nucleicos inhibidores complementarios a regiones del ARNm de c-myc inhiben la expresión de la proteína c-myc en una línea celular de leucemia promielocítica humana, HL60, que sobre-expresa el proto-oncogén c-myc. Véase, Wickstrom E.L., y col., (1988) PNAS (USA), 85:1028-1032 y Harel-Bellan, A., y col., (1988) Exp. Med., 168:2309-2318.

[0171] Los polipéptidos codificados también se pueden utilizar para diseñar moléculas (peptídicas o no peptídicas) que inhiban a las proteínas endógenas inhibiendo, por ejemplo, la interacción entre la proteína y una segunda molécula reconocida específicamente por la proteína. Los procedimientos para el diseño de dichas moléculas son muy conocidos por los expertos en la materia.

[0172] Por ejemplo, se pueden diseñar polipéptidos que tengan una identidad de secuencia con las proteínas codificadas o que pueden comprender modificaciones (conservativas o no conservativas) de las secuencias. Las modificaciones se pueden seleccionar, por ejemplo, para que alteren su estabilidad *in vivo*. Por ejemplo, la inclusión de uno o más D-aminoácidos en el péptido normalmente incrementa la estabilidad, particularmente si los restos D-aminoácido están sustituidos en uno o ambos términos de la secuencia peptídica.

[0173] El polipéptido también se puede modificar mediante su unión a otras moléculas. Por ejemplo, se pueden introducir diferentes grupos N- o C-terminales para alterar las propiedades físicas y/o químicas de la molécula. Dichas alteraciones se pueden utilizar para modificar, por ejemplo, la adhesión, estabilidad, biodisponibilidad, localización o detección de las moléculas. Para fines diagnósticos, se pueden unir a los extremos una amplia variedad de marcadores, que pueden proporcionar, directa o indirectamente, una señal detectable. Así, los polipéptidos se pueden modificar de diferentes formas para una variedad de propósitos finales mientras aún retienen la actividad biológica.

EJEMPLOS

[0174] Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar, la presente invención.

Ejemplo 1

IMPLICACIONES PRONÓSTICAS DE LA AMPLIFICACIÓN DE LA REGIÓN 20q13 DEL CROMOSOMA EN CÁNCER DE MAMA

Pacientes y material tumoral

[0175] Se obtuvieron muestras de tumores procedentes de 152 mujeres que se habían sometido a cirugía para tratamiento de cáncer de mama entre 1987 y 1992 en la Universidad o en los Hospitales de la ciudad de Tampere.

Ciento cuarenta y dos muestras procedían de carcinomas de mama primarios y 11 de tumores metastásicos. Diez de los tumores primarios que eran carcinomas *in situ* o mucinosos se excluyeron del material, puesto que los especímenes se consideraban inadecuados para los estudios de FISH. De los restantes 132 tumores primarios, 128 eran carcinomas ductales invasivos y 4 carcinomas lobulares. Las edades de las pacientes están comprendidas entre los 29 y los 92 años (61 años de media). Hubo seguimiento clínico disponible para 129 pacientes. El periodo de seguimiento medio fue de 45 meses (intervalo de 1,4-1,77 meses). Se administró terapia de radiación a 77 de las 129 pacientes (51 pacientes con ganglios linfáticos positivos y 26 con ganglios linfáticos negativos), y terapia sistémica adyuvante a 36 pacientes (33 con quimioterapia endocrina y 3 con quimioterapia citotóxica). El tamaño de los tumores primarios y la implicación de los ganglios axilares se determinaron de acuerdo con la clasificación de metástasis del tumor-nódulo (TNM). El diagnóstico histopatológico se evaluó de acuerdo con la Organización Mundial de la Salud (11). Los carcinomas se clasificaron en base a la disposición tubular de las células cancerosas, la atipia nuclear, y la frecuencia de figuras nucleares mitóticas o hiper cromáticas de acuerdo con Bloom y Richardson, Br. J. Cancer, 11: 359-377 (1957).

15 **[0176]** Los especímenes procedentes de la biopsia quirúrgica se congelaron a -70°C un máximo de 15 minutos después de la extracción. Se prepararon secciones criostáticas (5-6 µm) para el diagnóstico histopatológico intraoperatorio, y se cortaron secciones finas adicionales para los estudios inmunohistoquímicos. Se cortó una sección adyacente de un grosor de 200 µm para la citometría de flujo de ADN y los estudios de FISH.

20 **Preparación de las células para la FISH**

[0177] Después de la verificación histológica de que los especímenes de la biopsia contenían una elevada proporción de células tumorales, los núcleos se aislaron a partir de las secciones congeladas de 200 µm de acuerdo con una modificación del procedimiento Vindelov para la citometría de flujo de ADN, se fijaron y se pusieron en portaobjetos para el análisis de FISH como describe Hyytinen y col., Cytometry 16: 93-99 (1994). Como controles negativos en los estudios de amplificación se utilizaron fibroblastos de prepucio y se prepararon recogiendo células en confluencia para obtener núcleos en interfase enriquecidos en la fase G1. Todas las muestras se fijaron en metanol-ácido acético (3:1).

30 **Sondas**

[0178] Se utilizaron cinco sondas que mapean en la región 20q13 (véase Stokke, y col., Genomics, 26: 134-137 (1995)). Las sondas incluían clones-PI para el receptor de melanocortina-3 (Sonda MC3R, longitud fraccional a partir del brazo p del telómero (Flpter 0,81) y fosfoenolpiruvato carboxi quinasa (PCK, Flpter 0,84), así como clones del cósmido anónimo RMC20C026 (Flpter 0,79). Además, se utilizaron RMC20C001 (Flpter 0,825) y RMC20C030 (Flpter 0,85). Previamente se había demostrado que la sonda RMC20C001 define la región de máxima amplificación (Tanner y col., Cancer Res, 54: 4257-4260 (1994)). Se utilizó una sonda del cósmido que mapea en el brazo p proximal, RMC20C038 (Flpter 0.237) como sonda de referencia específica para el cromosoma. Las sondas de pruebas se marcaron con biotina-14-dATP y la sonda de referencia con digoxigenina-11-dUTP utilizando la traslación de mellas (Kallioniemi y col., Proc. Natl Acad Sci USA, 89: 5321-5325 (1992)).

Hibridación *in situ* con fluorescencia

45 **[0179]** Se llevó a cabo una FISH con dos colores utilizando sondas específicas de 20q13 marcadas con biotina y la sonda de referencia de 20p marcada con digoxigenina esencialmente como se ha descrito (Id.). Las muestras tumorales se postfijaron en paraformaldehído al 4%/tampón fosfato salino durante 5 minutos a 4°C antes de su hibridación, se deshidrató en etanol al 70%, 85% y 100%, se secó al aire, y se incubó durante 30 minutos a 80°C. Los portas se desnaturalizaron en formamida al 70%/disolución patrón de citrato salino 2x a 72-74°C durante 3 minutos, seguido por una digestión con proteinasa K (0,5 mg/ml). La mezcla de hibridación contenía 18 ng de cada una de las sondas marcadas y 10 µg de ADN de placenta humana. Después de la hibridación, las sondas se detectaron inmunológicamente con avidina-FITC y anti-digoxigenina rodamina. Los portas se contratiñeron con 4,6-diamino-2-fenilindol (DAPI) 0,2 µM en una disolución anti-desteñimiento.

Microscopía de fluorescencia y puntuación de las señales en los núcleos en interfase

55 **[0180]** Se utilizó un microscopio de fluorescencia Nikon equipado con filtros paso bajo de doble banda (Chromatechnology, Brattleboro, Vermont, EE.UU.) y un objetivo 63x (NA 1,3) para la visualización simultánea de las señales de FITC y rodamina. Se puntuaron al menos 50 núcleos no solapados con una morfología intacta en base a la contratiñencia DAPI para determinar el número de pruebas y las señales de hibridación de la sonda de referencia.

Los leucocitos que se infiltran en el tumor se excluyeron del análisis. Se realizaron hibridaciones control a núcleos de fibroblastos normales en interfase para cerciorarse de que las sondas reconocían una única copia diana y de que las eficacias de hibridación de las sondas de prueba y de referencia eran similares.

- 5 **[0181]** Los resultados de la puntuación se expresan tanto en forma de número medio de señales de hibridación por célula como nivel medio de amplificación (= media del número de señales en relación al número de señales de la sonda de referencia).

Citometría de flujo del ADN y análisis del receptor de esteroides

10

[0182] Se llevó a cabo la citometría de flujo del ADN a partir de secciones congeladas de 200 µm como describe Kallioniemi, Cytometry 9: 164-169 (1988). Se llevó a cabo un análisis utilizando el citómetro de flujo EPICS C (Coulter Electronics Inc., Hialeah, Florida, EE.UU.) y el programa MultiCycle (Phoenix Flow Systems, San Diego, California, EE.UU.). Se utilizó un índice de ADN por encima de 1,07 (en más del 20% de las células) como criterio para la aneuploidía del ADN. En histogramas aneuploides del ADN, se analizó la fase S procedente únicamente del clon aneuploide. La evaluación del ciclo celular tuvo éxito en el 86% de los tumores (108/126).

15

[0183] Se detectaron inmunohistoquímicamente los receptores de estrógeno (ER) y progesterona (PR) a partir de secciones criostáticas como se ha descrito previamente (17). Los resultados de la tinción se evaluaron semicuantitativamente y una puntuación histológica superior o igual a 100 se consideran positivas tanto para ER como para PR (17).

20

Procedimientos estadísticos

25 **[0184]** Las tablas de contingencia se analizaron con pruebas de Chi cuadrado para la tendencia. La asociación entre la fracción en fase S (variable continua) y la amplificación de 20q13 se analizó con la prueba de Kruskal-Wallis. El análisis de supervivencia libre de enfermedad se realizó mediante el programa BMDPIL y en el análisis de regresión multivariante se utilizaron la prueba Mautel-Cox y el modelo de riesgos proporcionales de Cox (programa BMDP2L) (Dixon BMDP Statistical Software. Londres, Berkeley, Los Angeles: University of California Press, (1981)).

30

Amplificación de 20q13 en carcinomas de mama primarios mediante hibridación *in situ* con fluorescencia

[0185] En el análisis de FISH se utilizó la región mínima de la sonda RMC20C001 para evaluar la amplificación de 20q13. Se utilizó la FISH para analizar tanto el número total de señales en células tumorales individuales como para determinar el nivel medio de amplificación (número medio de copias con la sonda RMC20C001 en relación a una sonda de referencia 20p). Además, también se valoró la distribución del número de señales en los núcleos tumorales. Los tumores se clasificaron en tres categorías: sin amplificación, bajo nivel de amplificación y elevado nivel de amplificación. Los tumores clasificados como no amplificados mostraban menos de 1,5 veces el número de copias de RMC20C001 en comparación con el control de brazo p. Los clasificados como con un bajo nivel de amplificación presentaban un nivel medio de amplificación de 1,5-3 veces. Los tumores que presentaban un nivel de amplificación superior a 3 veces el nivel de amplificación medio se clasificaron como muy amplificados.

35

40

[0186] Los tumores muy amplificados a menudo presentaban una amplia heterogeneidad intratumoral con hasta 40 señales en células tumorales individuales. En tumores altamente amplificados, las señales de la sonda RMC20C001 siempre estaban dispuestas en grupos para la FISH, lo que indica una localización de las secuencias de ADN amplificadas muy próximas entre sí, por ejemplo, en una conformación en tándem. Se encontró un bajo nivel de amplificación de 20q13 en 29 de los 132 tumores primarios (22%), mientras que nueve casos (6,8%) mostraban un elevado nivel de amplificación. La prevalencia global de un mayor número de copias en 20q13 fue de esta forma del 29% (38/132).

45

Definición de la región mínima de amplificación.

[0187] Se determinó el número medio de copias de cuatro sondas flanqueantes RMC20C001 en los nueve tumores altamente amplificados. Las sondas flanqueantes probadas fueron el receptor de melanocortina-3 (MC3R, FLpter 0,81), la fosfoenolpiruvato carboxiquinasa (PCK, 0,84), RMC20C026 (0,79) y RMC20C030 (0,85). El tamaño y la ubicación del amplificación varió ligeramente de un tumor a otro pero RMC20C001 fue la única sonda altamente amplificada de manera consistente en los nueve casos.

50

55

Asociación de la amplificación de 20q13 con características patológicas y biológicas.

[0188] La amplificación de 20q13 estaba asociada de manera significativa con el alto grado histológico de los tumores ($p = 0,01$). Esta correlación se observó tanto en tumores moderada como altamente amplificados (Tabla 4). La amplificación de 20q13 también estaba asociada de manera significativa con aneuploidía determinada por citometría de flujo de ADN ($p = 0,01$, Tabla 4) La actividad de proliferación celular media, medida como el porcentaje de células en la fracción en fase S, se incrementó ($p = 0,0085$ mediante la prueba de Kruskal-Wallis) con el nivel de amplificación en los tumores sin amplificación, con bajos niveles de amplificación y elevados niveles de amplificación (Tabla 4). No se encontró asociación con la edad de la paciente, el tamaño del tumor primario, el estado de los ganglios axilares o del receptor-hormona esteroide (Tabla 4).

10

Tabla 4. Correlaciones clínico-patológicas de la amplificación en la región cromosómica 20q13 en 132 cánceres de mama primarios

ES 2 390 530 T3

Característica patobiológica	Estado de amplificación de 20q13			Valor de p ¹
	Nº de pacientes SIN AMPLIFICACIÓN (%)	Nº de pacientes con un BAJO NIVEL DE AMPLIFICACIÓN (%)	Nº de pacientes con un ELEVADO NIVEL DE AMPLIFICACIÓN (%)	
Todos tumores primarios	94 (71%)	29 (22%)	9 (6,8%)	
Edad de las pacientes				
< 50 años	17 (65%)	6 (23%)	3 (12%)	
≥ 50 años	77 (73%)	23 (22%)	6 (5,7*)	,39
Tamaño del tumor				
< 2 cm	33 (79%)	7 (17%)	2 (4,8%)	
≥ 2 cm	58 (67%)	22 (25%)	7 (8,0%)	,16
Estado de los ganglios				
Negativo	49 (67%)	19 (26%)	5 (6,8%)	
Positivo	41 (75%)	10 (18%)	4 (7,3%)	,41
Grado histológico				
I - II	72 (76%)	18 (19%)	5 (5,3%)	
III	16 (52%)	11 (35%)	4 (13%)	,01
Estado del receptor de estrógeno				
Negativo	30 (67%)	10 (22%)	5 (11%)	
Positivo	59 (72%)	19 (23%)	4 (4,9%)	,42
Estado del receptor de progesterona				
Negativo	57 (69%)	20 (24%)	6 (7,2%)	
Positivo	32 (74,4%)	8 (19%)	3 (7,0%)	,53
Ploidía del ADN				
Diploide	45 (82%)	8 (14,5%)	2 (3,6%)	,01
Aneuploide	44 (62%)	20 (28%)	7 (10%)	
Fracción en fase S (%)	media_+ DE 9,9 ± 7,2	media + DE 12,6 ± 6,7	media +_DE 19,0 ± 10,5	,0085 ¹

Prueba de Kruskal-Wallis

Relación entre la amplificación de 20q13 y la supervivencia libre de enfermedad.

[0189] La supervivencia libre de enfermedad de pacientes con elevado nivel de amplificación de 20q13 fue 5 significativamente más corta que en pacientes con nada o sólo un bajo nivel de amplificación ($p = 0,04$). La supervivencia libre de enfermedad de pacientes con tumores moderadamente amplificados no difirió significativamente de la de pacientes sin amplificación. Entre las pacientes con ganglios negativos ($n = 79$), un elevado nivel de amplificación de 20q13 fue un factor pronóstico muy importante para una supervivencia libre de enfermedad más corta ($p = 0,002$), incluso en el análisis de regresión multivariado de Cox ($p = 0,026$) tras ajustar el 10 tamaño del tumor, el grado de ER y PR, la ploidía y la fracción en fase S.

Amplificación de 20q13 en tumores de mama metastásicos.

[0190] Dos de los 11 tumores de mama metastásicos presentaban un nivel bajo y un elevado nivel de 15 amplificación de 20q13. Por lo tanto, la prevalencia global (27%) del aumento en el número de copias de 20q13 en tumores metastásicos era similar a la observada en tumores primarios. Ambos especímenes tumorales primarios y metastásicos estaban disponibles procedentes de una de las pacientes. Esta paciente de 29 años de edad desarrolló una metástasis infiltrada en el músculo pectoral ocho meses después de mastectomía total. La paciente no recibió terapia adyuvante o radioterapia después de la mastectomía. La mayoría de las células tumorales en el 20 tumor primario mostraban un bajo nivel de amplificación, a pesar de que las células tumorales individuales (menos de 5% del total) contenían 8-20 copias por célula por FISH. En contraste, todas las células tumorales procedentes de la metástasis mostraron un elevado nivel de amplificación de 20q13 (12-50 copias por célula). El número de copias absoluto de la sonda de referencia siguió siendo el mismo, lo que sugiere que el elevado nivel de amplificación no produjo un mayor grado de aneuploidía.

25

Valor diagnóstico y pronóstico de la amplificación de 20q13.

[0191] Estos hallazgos sugieren que la amplificación de 20q13 recién descubierta puede ser un componente importante de la vía de progresión genética de ciertos carcinomas de mama. Específicamente, los experimentos 30 anteriores establecen que: 1) un elevado nivel de amplificación de 20q13, detectado en el 7% de los tumores, estaba asociado de manera significativa con la disminución de la supervivencia libre de enfermedad en pacientes con ganglios negativos en cáncer de mama, así como con indicadores indirectos de elevado potencial maligno, tal como un grado elevado y la fracción en fase S. 2) Un bajo nivel de amplificación, que era mucho más habitual, también llevaba asociado las características clínico-patológicas de tumores agresivos, pero no era significativo a nivel de 35 pronóstico. 3) El nivel de amplificación de RMC20C001 sigue siendo superior a la amplificación de genes y loci candidatos cercanos, lo que indica que hay un nuevo oncogén en las proximidades de RMC20C001.

[0192] Un elevado nivel de amplificación de 20q13 estaba definido por la presencia de un número de copias de 40 amplificación de 20q13 más de 3 veces superior, que es algo inferior a las frecuencias de amplificación reportadas para algunos de los otros oncogenes de cáncer de mama, tales como ErbB2 (17q12) y ciclina D (11q13) (Borg y col., Oncogene. 6: 137-143 (1991), Van de Vijver y col. Adv. Canc. Res., 61: 25-56 (1993)). Sin embargo, similar a lo que se ha encontrado previamente con estos otros oncogenes (Swab, y col., Genes Chrom. Canc., 1: 181-193 (1990), Borg y col., más arriba) un elevado nivel de amplificación de 20q13 era más habitual en los tumores con un alto grado o una fracción elevada en fase S y en los casos de mal pronóstico. Aunque sólo se analizó un pequeño 45 número de pacientes con ganglios negativos, nuestros resultados sugieren que la amplificación de 20q13 podría tener un papel independiente como indicador pronóstico. Los estudios para abordar esta cuestión están garantizados con una gran cantidad de material de pacientes. Además, basándose en estas correlaciones de supervivencia, el oncogén putativo desconocido en la actualidad, amplificado en este locus puede conferir un fenotipo agresivo. Por lo tanto, la clonación de este gen es un objetivo importante. En base a la asociación de la 50 amplificación con tumores muy proliferativos se podría proponer la hipótesis del papel de este gen en la regulación del crecimiento de la célula.

[0193] El papel del bajo nivel de amplificación de 20q13 como acontecimiento significativo en la progresión tumoral parece menos claro. El bajo nivel de amplificación se definió como un incremento de 1,5 a 3 veces el número medio 55 de copias de la sonda 20q13 con respecto al control brazo p. Además, estos tumores carecían de forma característica de células tumorales individuales con un número de copias muy elevado, y mostraban un aspecto disperso y no agrupado de las señales. Una distinción precisa entre un elevado y un bajo nivel de amplificación de

20q13 sólo se puede realizar de manera fiable mediante FISH, mientras que los análisis de transferencia de Southern y de transferencia de ranuras es probable que sean capaces de detectar sólo un elevado nivel de amplificación, en el que tiene lugar la elevación sustancial del número medio de copias de genes. Esta distinción es importante, ya que sólo los tumores muy amplificados llevaban asociados resultados clínicos adversos. Los tumores con un bajo nivel de amplificación de 20q13 parecían tener muchas características clínico-patológicas que estaban a medio camino de las encontradas en tumores sin amplificación y aquellos con un elevado nivel de amplificación. Por ejemplo, la fracción promedio del tumor en fase S fue la más baja en tumores no amplificados y la más elevada en tumores muy amplificados. Una posibilidad es que el bajo nivel de amplificación preceda al desarrollo de un elevado nivel de amplificación. Este ha demostrado ser el caso, por ejemplo, en el desarrollo de amplificación *in vitro* de genes con resistencia a fármacos (Stark, Adv. Canc. Res., 61: 87-113 (1993)). Se encontraron evidencias que apoyan esta hipótesis en uno de nuestras pacientes, cuya metástasis local contenía un nivel mucho más alto de amplificación de 20q13 que el tumor primario operado 8 meses antes.

[0194] Por último, nuestro trabajo anterior informó de una región crítica de 1,5 Mb definida por la sonda RMC20C001 y la exclusión de genes candidatos en líneas celulares de cáncer de mama y en un número limitado de tumores primarios de mama. Los resultados del presente estudio confirman estos hallazgos al demostrar de manera concluyente en un conjunto más amplio de tumores primarios que la región crítica de amplificación está, de hecho, definida por esta sonda.

[0195] Los datos actuales sugieren que el elevado nivel de amplificación de 20q13 puede ser un paso importante en la progresión de ciertos tumores de mama a un fenotipo más maligno. Las implicaciones clínicas y pronósticas de la amplificación de 20q13 son sorprendentes y ahora se ha definido la ubicación de la región mínima de amplificación en 20q13.

[0196] Se entiende que los ejemplos y las formas de realización descritas en el presente documento sólo tienen fines ilustrativos y que en vista de ella los expertos en la materia podrán sugerir diversas modificaciones o cambios que se deben incluir dentro del espíritu y alcance de esta solicitud y el alcance de las reivindicaciones adjuntas. Todas las publicaciones, patentes y solicitudes de patente citadas en el presente documento se incorporan por referencia para todos los propósitos.

30

Discusión del listado de secuencias acompañantes

[0197] Las SEQ ID NO: 1-10 y 12-13 proporcionan secuencias de ácidos nucleicos. En cada caso, la información se presenta como una secuencia de ADN. Un experto comprenderá fácilmente que la secuencia también describe el ARN correspondiente (es decir, por sustitución de los restos T con restos U) y sus diversas variaciones conservativamente modificadas. La secuencia complementaria se describe completamente por comparación con la secuencia existente, es decir, la secuencia complementaria se obtiene utilizando las reglas convencionales de emparejamiento de bases de ADN (por ejemplo, A a T, C a G). Además, la secuencia de ácidos nucleicos proporciona la secuencia de aminoácidos correspondiente mediante la traducción de la secuencia de ADN dada utilizando el código genético.

[0198] Para la SEQ ID NO 11, la información se presenta como una secuencia polipeptídica. Un experto comprenderá fácilmente que la secuencia también describe todas las secuencias de ARN y ADN correspondientes que codifican el polipéptido, por conversión de la secuencia de aminoácidos en la secuencia de nucleótidos correspondiente utilizando el código genético, asignando alternativamente cada posible codón a cada posible posición del codón. De forma similar, cada secuencia de ácidos nucleicos proporcionada también proporciona inherentemente todos los ácidos nucleicos que codifican la misma proteína, puesto que la persona experta puede traducir simplemente un ácido nucleico seleccionado en una proteína y, a continuación utilizar el código genético para hacer la traducción inversa a todos los posibles ácidos nucleicos a partir de la secuencia de aminoácidos.

50

[0199] Las secuencias también proporcionan diversas variaciones modificadas conservativamente mediante la sustitución de restos apropiados con las sustituciones conservativas de aminoácidos proporcionadas como ejemplo, por ejemplo, en la sección de Definiciones anterior.

55 [0200]

LISTADO DE SECUENCIAS

SEQ. ID. No. 1

3bf4 3000 pb

CCGCCGGCCGGGGGGCCCTGGCTGCACTCAGCGCCGGAGCCGGGAGCTAGCGGCCGCCCATGTCCCACCAGACCGGCAT
CCAAGCAAGTGAAGATGTTAAAGAGATCTTTGCCAGAGCCAGAAAATGGAAAGTACAGACTTCTGAAAAATCTATTGAAA
ATGAGCAACTTGTGATGGATCATA TAGTCAGCCCTTCAGATTCCTGGGATAGGATTTATGATTCCTTTGTTTTACCCTGT
TTGGAGGACAAAACCAATGCTATA TATTAATCAGGTTAGATTCAGAAATGCCAGGGATATGAATGGATATTCATGGC
ATGGTC TCCAGATCATTTCATGTTCTGTCAAAAAATGTTGTATGCAGCAACAAGAGCAACTCTGAAGAAGGAATTTGGAG
GTGGCCACATTAAGATGAAGTATTGGAAACAGTAAAGGAAGA GTATCA TTACATGGATATAAAAAATACTTGTCTCA
CAATCTTCCCTCCCTCAGCTGACTGCACTGAGGAAGAACTACGACAGATTA AAAATCAA TGAGGTACAGACTGACGTGGG
TGTGGACACTAAGCATCAAACACTACAAGGAGTAGCATTCCCATTTCTCGAGAAGCCCTTTCAGGCTTTGGAAAAATGA
ATAATAGACAGCTCAACTATGTGCAGTTGGAAA TAGATA TAAAAATGAAATTAATA TTTGGCCAACACAAACAA TACA
GAACTGAAGATTTGCCAAGAGGATTC CCAAGGATTCAGCTCGTTACCA TTTCTTCTGTATAAA CATTCCCA TGAAGG
AGACTATTTAGAGTCCATAGTTTTTATTTA TTTCAATGCC TGGATACACATGCAGTAEAGAGAGCCGGATGCTGTATTTCTA
GCTGAAGAGCCCGTCTGTAGAAA TGTAGAAAAGACAAC TACAAA TGGATGTAATTAGAAAAGATGAGATAGACAA TGGG
GATGAGTTGACTGCACTTCTTTA TGAAGAAGTACA TCCCAAGCAGCA TGCACACAAGCAAAGTTTTCGAAAACCAA
AGGTCTCGCAGGAAAAGAGGAA TTCGAAGACTAATTAGGGGCCAGCGGAAACTGAAAGCTACTACTGATTAAGTCA TC
ACATTAACATTTGTAATAC TAGTTTTTAAAGTCCAGCTTTTAGTACAGGAGAAC TGAATCA TTTCCATGTTGATAEAA
AGTAGGGA AAAAATTTGACTTTTTGGAAA TAGCACTTTCCATCTGTGTTTTTAAATTAATGTATAGAA GACT
CATGATTTCTATTTTTGAGTAAAGCTAGAAAAGGGTCAACATAATGTTTAAATTTGTCACACTGTTTTCATAGCGTTG
ATTCACACTTCAAATACTTCTTAAAATTTTATACAGTTGGGCCAGTCTTAGAAAAGTCTGATGCTCAAAGGGTAAACTT
ACTACTTTCTTGTGGGACAGAAAGACCTTAAAATATTCATA TTTACTTAA TGAATATG TTAAGGACCAGGCTAGAGTATTT
TCTAAGCTGGAAACTTAGTGTGCC TTGGAAAAGCCGCAAGTTGCTTACTCCGAGTAGCTGTGCTAGCTCTGTCAGACTGT
AGGATCATGCTGCAACTTTTAGAAA TAGTGTCTTATA TGTGCAGCAGTCTTTTATATTTGACTTTTTTTTAA TAGCATT
AAATTCAGACTCAGCTCAGCTCTGAAACTTTAAGGGTACCAGATA TTTCTATACTGCAGGATTTCTGATGACATTGAAAG
ACTTTAAACAGCC TTAGTAAATATCTTTCTAA TGTCTGTGAGGCCAAACATTTATGTTTCAGATGAAATTTAAATTA
TATCATTCAAAGGAAACAAAAAATGTTGAGTTTAAAAATCAGGATGACTTTTTCTCAAACCATACATTTAAGGG
CAAATGTTGTTCTTATCAGCTCCGAGCAAATAC TCAAGTTTAAATTTACTTTAAAGTCTGGTACTTAAAGGCTAAG
TAGATAAACACCTTAATAATCTCAGTTAATACTGTATTTCAAACACATTTAACTGTTTTCTAA TGTCTTGCATTATCAG
TTACAACCTAGAGAGATTTGAGCCTCATATTTCTTTGATACTTGAATAGAGGGAGCTAGAACACTAA TGTTTAACTCT
GTAAACCTGCTGCAAGAGCCATAACTTTGAGGCATTTTCTAAATGAACTGTGGGGATCCAGGATTTGTAATTTCTTGAT
CTAAACTTTATGCTGCATAAATCAGTTATCGGAAATGCACATTTCA TAGTGTGAAGCACTCATTCTAAACCTEATATC
TAAGGTAA ZATA TGCACCTTTCAGAAAATTTGTTGTGAGTAAAGTAAAGCATA TTAGAA TAATTTGTTGGTTGACAGATTTT
TAAAATAGAAATTAGAGTATTTGGGGTTTTGTTGTTTACAAAATAATCAGACTATAATATTAAACATGCAAAAATACTG
ACAATAATGTTGCAC TTTGTTACTAAAGATATAAGTTGTTCCATGGGTGTACACGTAGACAGACACACATACACC CAAAT
TATTTGCA TTAAGAA TCTGGAGCAGACCATAGCTGAAGCTGTTA TTTTCAGTCAAGGAGACTACCTGTCA TGAAGGTATA
AAAATAATTTAGAAGTGAATGTTTTTCTGTACCATCTATGTGCAATTA TACTCTAAATTCACCTACACTACATTAAGTAA
ATGGACATTCAGAA ZATA GATGATATAGCTTAAACTAATTA TTTAATTAACCAA TGAATGCTGAAAATCAGTGAATG
CAFTTGTATA TAGAGTATAACTCATCGTTTACAGTATGTTTTAGTTGCCAGTATCATACCTAGATGGTGAATAAACATA TTC
CCAGTAAATTTATA TAGCAGTGAAGAA TACA TGCCTTCTGGTGGACATTTTATAAGTGCATTTTATATCAAA TAAAAA
TTTTTTCTCTTTAAAAA AAAAAAACAAGAAAAA AAAAAA

5 SEQ. ID. No. 2
1b11 723 pb

TGGAAGCTGTCATGTTTACCGTCTCTAACGTTGGACTCTTAAGAAAATGATTA TTTCTGGTTCTAGACAGGCCAAATGT
AATTCACCTACGTGGCAGATTAAGAGGTGGGCTTACTAGATTTGATTTGGGTA TTGAGCATGCTCTGAATGACAGTCCCC
AAAAAGGACCTCTTATCCGTCTTCCCTTTGGGGAAGGGCTTTTGGCCACTTCCATGTCAA TGTGGCAGTTGAGCTTGGAA
ATTGGTGGCTTGTACAACA TAAGCA TTACTTCTCAAGATGTGCCCTGTGTAGAAATGGTCA TAGATTCAAAAC TGTAGCT
ACTATGTGGACAGGGGGCAGCAAGGAC CCCCATTGTAAAACATGTTTTGGGGGAATGTTTTGTTTTTCA TTTTCTTAT
TACC TGGCAAAAATA TCCAGGTGGTGTGTGAGTCACCAAGTAGAGATTA TAAAGTCCAAGGAAGTGAATCAGCCTTACAA
ACAGTGGACCTCAACGAAGGAGATGCTGCACC TGAACCCACGGAAGGAAACTCAAAGAGAAGAAAGCAAACCAAGAAC
CTCTCTGATGRC TTTCTCAGACAAA TGGTAAGCCCTTACTTCCAGTATAGGAAAACCTAAGATACCTAGAGCGGCTTTT
GGGAACAATGGGCTCATGCCACAGGTAG TAGGAGACATAA TGTAGCTGGTGTGTATGGAATGTGAATGGAA TATGGATT
GCG

10 SEQ. ID. No. 3
cc49 1507 pb

GCAGGTTGCTGGGATTGACTTCTTGCTCAATTGAAACACTCAATCAATGGAGACAAAGAGCACAAATGCTTTGTGCTGAT
TCATATTTGAAATCGAGGCATTGGGAACCCGTATGCCCTTGTTTGGAAAGAACCAGTGACACCATCACTGAGCTTCCTA
AAAGTTCGAAGAAGTTAGAGGACTATACACTTCTTTTGAACCTTTATAATAAATATTTGCTCTGGTTTGGAAACCCAGG
ACTGTTAGAGGGTGAAGTGACAGGCTTACAGTGGCCCTAATCCAACTCCAGAAATGGCCCAACGGAACTTTGAGATTATA
TGCAATCGAAAGTGACAGGAACA TGCCAACTCAATCCCTCTTAAATGTACATGGATGGC CAAGAGTGATGGCAGCTCTC
TTGCCAGTCCGATGGAGATGGAGATGCCTTGTCAA TGAAAGGGCCNCTGTTGTCAATTCGGACTACACAAAGAAAAA
ATGTCAATCCGAAATCGAGGGGAAATGCCCC TTGGATTGCAATGTTCTGCAGCCAGACCTTCACACATTCAGAAGACC TAA
TAAACATGCTTAAATGCAACACCCGGCTACCCCTCTGTGAACCAGCAGTTCTTCGGGTTGAAGCAGAGTATCTCAGTCCGC

TTGATAAAAGTCAAGTGC GAACAGAACCTCCCAAGGAAAAGAAATGCAAGGAAAA TGAATTTAGCTGTGAGGTATGTGGG
CAGACATTTAGAGTCCGCTTTTGAATGTTGAGATCCACATGAGAACACACAAAGATTCCTTCACCTACGGGTGTAACA TGTG
CGGAAGAAGATTC AAGGAGCCTTGGTTTCTTAAAAATCACA TGCGGACRCATAATGGCAAAATCGGGGGCCAGAAAGCAAA
TGCAGCAAGGCCTGGAGAGTAGTCCAGCAACGATCAACGAGGTCGTCAGGTGCAGCGGGCCGAGAGCATCTCCTCTCCT
TGCAAAATCTGCATGGTTTGTGGCTTCTTATTTCCAAAATAAAGAAAGTCTAATTTGAGCACCGCAAGGTGCACACCAAAA
AACTGCTTTCGGTACCAGCAGCCGCGCAGACAGACTCTCCACAAGGAGGAAATGCCGCTCTCGAGGGAGGACTTCTCTCAGT
TGTTCAACTTGAGACCAAAAATCTCACCTGAAACGGGGAAGAAGCCTGTGAGATGCA TCCCTCAGCTCGATCCGTTCCACC
ACCTTCAGGC TTGGCAKCTGGCTACCAAAGGAAGATTGCTCAATTTGCCAAGAAGTGAAGGAAATGGGGCAAGAAGGGAG
CACCGACAACGACGATTCGAGTTCGGAGAAGGAGCTTGGAGAACAATAAGAACAATTGTGTCAGGCCTCTCGCAAGAGA
AAGAGAAGTGCAAAACACTCCACCGCGGAAGCGCCCTCCGTGGACCGGATCCCAAGTACCCAGTAGCAAGGAGAAGCC
ACTCACTGCTCCGAGTCCGGCAAAGCTTTCAGAACCTACCACCAGCTGGTCTTGCACCTCCAGGGTCC

SEQ. ID. No. 4
cc43 2605 pb

5

CAAGCTCGAAATTAACCC TCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGAT
CCCCGGGCTGCAGGAATTCGGCACGAGCTGGGCTACTACGATGGCGATGAGTTTCGAGTGGCCGTGGCAGTATCGCTTC
CCACCCCTCTCTTACGTTACAAACCGAATGTGGACACTCGGCAGAAAGACTGGCCGCTGGTGTCTCGCTGGTCTCTCT
CTGGCCCTCTCAAAACAGTCCAGCA TGACGGTGAATGGAAGCTCAGGAGAGCCGCTTCAAAACGTTCAAGCTACAGC
GAAAGCTTCCCTGTGGAGTCCATCCAGATTGATAGAGGAACTGAGGAAGAAAGGGAACCTCGAGTGGTTGGATAAGAGC
AAGTCCAGCTTCTGATCAATGTGGCCGAGGCCAGAAGAA TGGGGGAAACTCATCTATCAGTGGGTTTCCAGGAGTGGCCA
GAACAATCCGCTCTTACCTGTATGAACTGACTAATGGGGAAGACACAGAGGATGAGGAGTCCACGGGCTGGATGAAG
CCACTCTACTCGGGCTCTGCAGGCCCTACAGCAGGAGCACAAAGGACTCCAGTCTGAAAGGCTGGGACTGGGGATGGGTT
AAGTTCTCTAGCAGGGACCTGTCTCCCTTACTTCTTACCTCCACCTTTCAGGGCTTCAAAAGGAGACAGACCCAG
TGTCCCCAAAGACTGGATCTGTGACTCCACAGACTCAAAGGACTCCAGTCTGAAAGGCTGGGACTGGGGATGGGTT
TCTCACACCCCAATATGCTGTCTCTTGGATAGGGTGAAGCTGAAAGCACAGGGAGAAAATATGTGCTTCTTCTCGCCCTA
CCTCCTTCCCACTCTAGACTGTCTCTGAGCCAGGGTCTGTAAACCTGACACTTATATGTGTTTACACATCTTAAGTACA
TACACACA TGCGCTGCAGCACATGCTTCTGTCTCTCC TCCCTCCACCCCTTAGCTGTGTTGCCCTCCCTCTCAGGC
TGGTGTGGATCTTCTTAGGGGATGGGGGAAGCCCTGGCTGCAGGCAGCCTTCCAGGCAATATGAAGATAGGAGGCCCA
CGGGCTGGCAGTGAAGGTTGTGGCCACACCGATTTA TGATATTAATAATCAACTCCCAAAAAAAAAAAAAAAAAA
CTGAGACTAGTTCCTCTCTCTCGAGAACTAGTCTCGAGTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
GCTTAAAGGATTTATTAATGTTTCTCTTACAGTGTCCACTTTTCTACTTAAACTACTTTCCAGTCTCAGAAGCC
CAGAGGCAAAAAAAAAAAGACCATGAATCTTCTCTCCAGATTAAGTACACACTTTGGA AAAACAGATTGGAAAACCTTT
CTGAAAAAAGTTEACTGAACTCCAAACCAACATGCCATAATGTTGATGTTGCTCATGAAAAATGTTAAAAACCTGTTCT
AGATAAAGAACAGTCTCAAGTTTTTGTACAGCTTACACATAGTACAAAGGGTCCCTATGATGATTTCTGTAGGACGAA
ATAATGTAATTTTTTTCAGTTTCTGGTTTATAACTCTCTCGATCTCAGAGTTGACTGATTA AAAACACTTACTCA TGCAACA
GAGAATAAAGCACTCATTTTTATAAATATA TGGACCAAACATTTTGGAAAATCTTATCTATTTGGAGACAAAATATGC
TGGACTAAAGCAA TAATTA TTTTATCTCAAATGCTGTGCTTAACCTCAAATGACTTAGAATGCTTTGCTATATTTTGCCCTC
TATGCCCTAACACACTGGCTTCTTTTAGCTCTTGAACAAGCCA
AACTGCTTCTCTGCTCAGGACCAGATAATTTGGGACTTCTCTEAAGAA TCTATTTCTTDAATCTTATCTGGGTAAC T
TAGTTTTATCCAACACTTCAGATCTCTGCCGTA AAAACTCTCTCTTATAGAAGCCTGTCA TGACACTGTCTCTCTCTCTCCAA
CATACTCACAGCACACATGTAGACTAGATTAGAACCCTCTGTTTTTCTTTTTTCA TACTTTTCTCTATCA TGGCTTCCCTC
CATTATAATTTTTTATAATGTTGTTGAAATGCTGCCCCAAAGTCAAGTTTCTCACTAAACTATAAATCCGTA AAGCTGG
GATCTTCCAATTTTGAATCACCACTTAGTACAGTAGGAACACAGTAAAGATTCAATTTGGTATTTGTGGAAATGAA TGAATG
AATTTGTTTGTCTAGTAAAGTCTGGGGGAACCCAGGTTGAGAAGAGCCTAGAAAGCAGGTCGAAATCCAAGGCTAGATAGACT
TAGTGTACTCAAGAAAGGGTAGCCTGAAAAATAAAGGTTCAAATATAGTCAAGAA TAGTCAAGACA TGGGCAAGACAAG
AGTGTCTCTCGTCCGAAATCGATATCAAGCTTATCGATACCTCGACTCGAGGGGGGGCCCGGTACCCAAATTCGCCCT
ATAGTGAAGTCTATACAAATCACTGGCCGCTGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACTTAAT

SEQ. ID. No. 5
41.1 1288 pb

10

GAGGGCAGCGAGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGCGCCCTCGCCAACCCAGC
CCGGCCCTGCCATGCATCAACCCACTCAGCGCCCTGCACTCCGTCCGTAACAACTCACCTGGGCAAAGCCAGGAGCCCT
TGCGCTCACCTTCC TGCTCCAGCCAAAGTTCAAGCACAATTTCCATGTTCCACAAGTCGAATTCAA TGTCATGGACAAG
CCGGCTTGAGTCC TGCCCTCCACAAGGTCAGCCAGCGTGTCCAGGGCGCTACCTGTTTGTGAAACAGCGATCAGCCCATTTGT
CTGACATCGCCGACATGGTCAAGTCTCTCCCAAAGCCACCTCCCAAAGCCAGCCCTCCCTCCAGGGTCCCCCTCATG
AAGCTGGAAA TGGATGTCAGGCGCTTTGAGGATGTC TCCAGTGAAGTCTCAACTTTGCA TAAAAGAAAAGGCCGGCAGTC
CAACTGGAA TCC T CAGCA TCTTCTGATCTTACAAAGCCAGTTTGCTCGAGCCCTTCCAGACATCAGAGGGCAAATACC
TGCTGTCTGATCTGGGCCACAAAGAGCGTATGCAAA TCTCTAAGTTTACGGGACTCTCAA TGACCACATATCAGTCACTGG
CTGGCCAACTCAAGTACCAGCTTAGGAAAACGGCGGGGACAAAATTTCTGAAAACA TGGACAAAAGGCCACCCCATCTT
TTATTGCACTGAGTGTGCC TCCAGTTCAGAACCCCTTCTACCTACATCAGTCACTTAGAA TCTCACCTGGGTTTCCAAA
TGAAGGACATGACCCGCTTGTCTAGTGGACCAGCAAGCAAGGTGGAGCAAGAGATCTCCCGGGTATCTCGGCTCAGAGG
TCTCCAGAAAACA TAGCTGCCAAGAGGACACAGACTCTAAATTCAGTGTAACTGTGCTGTCTGGACATTTGTGAGCAA
ACATCCGGTAAAAC TCCACCTTAAGCAAACGCCACAGCAAGTCAACCCGAACACCAATTCACAGTTTGTAAACAGGCTGGAT
AAGAATAGCTCTGCAGGACGAA TGCCTTAGT TTTCCACTTTCCAGCCTGGATCCCTTCACACTGAACCTTCTCTGTTGCA
CCATCTGCTTCTGCATTTGAAC TCAATGAAC TCTCTCTGACACCTTGCTCTGAGAAGACTGCCAAAAA AAAAAAAAAA
AAAAATTC

SEQ. ID. No. 6
GCAP 2820 pb

5

ATCCTAAGACGCACAGCC TGGGAAGCCAGCACTGGGGAAAGTGGTGTGAGGGATGTGGGTCACTGGGGTGAAGGTGGAGC
TTTCAGGGTCTCCCGTCAA TGCAGCTGAGT TTTCTTTGGCAGGGAA TTTACCAGCTGAAGAAAGCC TGCCGGCGAGAGCT
ACAAACTGAGCAAGGCCAGCTGCTCACACCCGAGGAGGTCGTGGACAGGATCTTCTCTCTGGTGGATGAGAA TGGAGATG
GTAAGAGGGCCAGAGATGGGAGAGTGTCTCCACTCTGCATCATGCCACTTCTGGCCGACGCTCTTTGGGCAAGGCC
CTCCACTTTCAAACCTGGGGTCTCTACTGTGAGAAGGCTGTGGAGAAGA TGTCAATGAAC TAAAGGGACTCAAGAG
CACGTGT TTTGAGGAGTGC TAAAAGTCTTACAGGAGTGTCTGATGGAGGCCAGGCAC GCAGAA TAGAAAGAA TAGGAAC
TTTGGAGTCAGGCAGGGAGTGA TATA TTAGCTTCTCTGCTCTAGTCTCAA TTTCTCTCA TCTGGAAAATGGGGATAATAAT
AGTGGTTGAGAGGAA TGAATAGGATAA TGTGTTAAGAGCAGGCA TAGGGTAGACTTCAATTCAGGCTGCTTGGGCTTTC
TCCCTGTAGCCCAAAGCCAGCCTCAGGGCTATGTGGGGAGAGAGCTGGCTTGGAA TACACACTTGAGCCCTCCAGCTC
TCTCAGCTCCACCAGCATTTCCGTTGTTACCTGCGCAAAGTAAAAC TTTCAA TTTCA T CAGCAAAGAAAGCCCTTAAAG
GTGGCAGGAGACTCTGGAGATTCAGACACTGACAAGCCGAAGCTTGAGGCTCTGAGACTGCAGGATAGTTGGCATAAG
ACGTTAGGGCCATCTCTGGGAGCGAGTCTCTCTCTGCCCCAGACCCAGTCTCCCTTCTCTACATGACCACCTC
TCTCCCTCTGCTCAGGCAGCTGTCTCTGAAAGAGTTTGTGAAAGGTGCCGTCCGGGCAAGTGGGTGA TAGAATGC
TGCAGATGGACATGAATCC CAGCAGCTGGCTCGCTCAGCAGAGACGGAAAAGTGC CATGTTCTGAGGAGTCTGGGGCCC
TCCACGACTCCAGGCTCACCCAGGTTTCAGGGTAGTAGAGGGTCCCTGCTCAGCCTGCTCA TGCCACACTTCTCCCT
TGGCTTGACTTCTTGGCACCCTCTGTGCAGGGCTGAGTGGGGATGGGAAAGGGCTGCTGGGTTTGAAGTGGCCAAACAGG
GCATAGTCCATTTTGGAGGAGTCCCTGGGATGGTGAAGGGAATCAGTTACTTTTCTGTT CAGCCGCTCTGGGAGGAC
TGTGCTTGGCTGGGTGGTTGTGGGGCTCCACAGT TTTCTGGGTGTCTCAGTTGGAAGCAAGAGCCAACTGAGGGGTGA
GGGTCC CACAGACAAA TCAAAA TGAAGAACAAAAGACTGGTAGGAGCCAGGGGTGGAGGGTGTGAGACTGAAGAAA
AGGCAGGAGTTGCCGGCACGGTGGCTCAGGCC TGTAA TCCAGCACTTTGGGAGGCC GAGCGGGCAGATCACGAGGTC
AGGAGATCGAGACCATCTGGCTAACACGGGGTGA AACCCCGTCTCTACTAAAAATACAAAAA T CAGCCGGGTGAGGTG
GCGGGCGCTGTAGTCC CAGCTACTCAGGAGGCTGAGGCCAAGAGAA TGGCGTGAACCCAGGGGGCCGAGCTACAGTGA
GCCGAGATTTGGCCACTGC ACTCCAGCCTGGACGACAGTGAGACTCCGTCTCAAAAAA AAAAAAAGAAAAGAAAAG
GCAGGAGTTTGGGGGGCAGGGGGCAGCAA TAA TTTCTA TAACTTCCGGGATGCTGAGGGGCGTTCA TGGGGAGGACC TG
GCCTCTCTCTCCCAAGGCA TCTCTCAGTGGTGTCAACAGGAAAAATGGCAGCAAATACGCTGCAGGCTGTGGTCTTT
CTGCTTTGAAAGGGT CAGCTGTACTTAAAGGGACTGTTTCAGCTCTGCC TGGGTGCTGCTCTGGGACCCCTGCTGCCA
ACCCACACTCCCCCAACAA TCTCTCTTTCCATCCA TATCCCCAGTATGGACCTTCCACAAC TCCAGCCATAAGCTG
AATGTTCTCTTTAAAGGATGGAGAAAAC TCTGCTGTCTCTGGCAAGAA TGGGGGACTGTTGACTGGGA TTTGTGGC
TGGGCTTGGCTTCTAACTGCTGTGTGACCCAAGACAGCCACTTCTCTCTCTAACTTGGTGTATGCTCTGGCAGCACAGT
GAGCAGGTCGGACTAGGCCAA CAGTTTGGATTA TTTGTGTTT TTAGATGTGGAATTA TTTTGTGTTATAAATCTTAT
GTGTAACCCCAA TA TAGAACTAGATAAAAGGGAGTCTCTCTGGTTGAAAGGGGAGCTGAGTACCTCTGGAAC TGGAG
GCACCTCTGAAAAAAGCAAAC TGA AAAACAGTCCCTGGGTCACTGTACTCTATAAGACAGTTTAAAGTGAACCTGG
AAAAACA TTTGCTTTACCTTGAATAGATAGGTTTTTATGTTGGTATA TAAGAAA TAAAAC TAACCTATTAACCTGAGAC
TTTACAGGTGTGTTATTTCA TATGATAGTCA TATAAAA TTTCTTTAGACATCAATTTTAGGTAAAAA TAAATGATTAG
AAAA TA TTTGGCCAGGTGCAGCAGCTCACACCTGCAATCC CAGGACTTTGGGAGGCCGAGCCGGGTGGATCACTGAGGT
CAGGGGTTCAAGACCAGCCCTG

SEQ. ID. No. 7
1b4 1205 pb

10

GGCGCGTGAGTCGGCCCCCCAGTCACGTGACCGCTGACTCGGGGGGTTCTCCACATCGCTTACCTACCTCCCTCTGC
AGGAACCOGGOGATATGGCTGCCGCTGTGCCCCCGCCCGCATTTCTCTCCCCGCTGCTTCCCTTCTCCTGGGCTTCTCTGC
TCCTCTCGCTCCGCATGGCGGCAGCGGCCTGCACACCAAGCGCCCTTCCCTGGATACGGTCACTTCTACAAGGTCA
TTCCAAAAGCAAGTTCTGCTGGTGAAGTTCGACACCCAGTACCCCTACGGTGAGAAGCAGGATGAGTTCAAGCGTCTTC
TGAAAACTCGGCTTCCAGCGATGATCTCTTGGTGGCAGAGGTGGGGATCTCAGATTATGTGACAAGCTGAACATGGAGCT
GAGTGAGAAAATACAAGCTGGACAAAGAGAGCTACCCATCTTCTACCTCTTCCGGGATGGGGACTTTGAGAACCAGTCCC
ATACACTGGGGCAGTTAGGTGGAGCCATCCAGCGCTGGCTGAAGGGGCAAGGGGTCTACCTAGGTATGCCTGGTGCCTG
CCTGTATACGACGCCCTGGCCGGGGAGTTCATCAGGGCCTCTGGTGTGGAGCCC GCCAGGCCCTCTTGAAGCAGGGGCAA
GATAACCCTCAAGTGTGAAGGAGACTCAGAAGAGTGGGCCGAGCAA TACTGAAGATCATGGGGGAAGATCTTAGACCAA
GGGGAGCACTTCCAGCATCAGAGATGACACGGATCGCCAGGCTGATTGAGAAGAACAAGA TGAAGTGACGGCCAGAAGGAGG
AGCTCCAGAAGAGCTTAAACATCTGACTGCCCTTCCAGAAGAAGGGGGCCGAGAAAGAGGAGCTGTAAAAGGCTGTCTG
TGATTTTCCAGGCTTTGGTGGGGTAGGGAGGGGANAGTTAACCTGCTGGCTGTGANTCCCTTGTGGAAATATAAGGGGGY
MSKGGGAAAAGTGTACTAACCACGATTTCTGACCCTGAGTATGCC TGGACATGATGCTAACATGACCATGCTTGGGA
TGTCTCTAGCTGGCTGGGGATAGCTGGAGCAC TACTCAGGTGGCTGGTGAATGACACCTCAGAAGGAATGAGTCCTA
TAGAGAGGAGAGAGGAGTGTACTGCCAGGCTTTGACAGATGTAATCTCATTC AATTAAAGTTTCAGTGTTTTGCTTA
AGTGG

SEQ. ID. No. 8
20sa7 456 pb

5

GAAATCAGAAGTTTAAATATGACACAA TAAAATATA TTTGTATA TCTCACACCGGAGNTTCTCTTCAAACATAAGGAGTTA
GAAATEACAAGTAGGCATA TGCTTCCATA TATTCAGATAAAATTCATTTGATTAAATTAATTC CAGATAGAGAGAAGTAAT
TTTCGGAAAAGAAA TGA TAGCTATAT TAAAGCAGATATTCATTACAATACCA TG TAGAGACATAAGCAATATTTGGCAT

CATCTGTCCGCTCAGTAGGCCGTGTTCCCTCTGGTAGGGCCTTTGGAGAGTACCATCTATCTAAGATGGAGGAATGCTG
TGGGAAGGGCGGGA TGGAGGTGCGTTTCTACGCTGAACCCACACAGGAAATCTGCAGCCACACAGCTGCCCTCTGCC
CGCCTTCCATGTGATCATCTGGTCAATGAAGTGAATGTCTTATTTGNGGGGT

SEQ. ID. No. 9
Secuencia genómica que codifica ZABC 1

10

CCATCATATTTCTTATTTTTTTGGGCGGAGAGGGGAGACTTGCTCTGTTGCCAGGCTGGACCAGTGGTGGAT
 CT
 TGGCTCACTGCAACCTCCACCTCCTGGGTTCAAGTGATTCCCAAATAGCTGGGATTACAGGTGTGTATTACCAT
 GC
 CCAGCTAATTTTTGTATTTTTAGCAGATAAGGGGTTTCACCATGTTGGCCAGGCTGGTCTCCAACCTCTGGCCT
 CA
 TGTGATCCACCCACTTCGGCTTCCCAAAGCATTGGGAGTATAGGTGTGAGCCACTATACCCGTCCTCACATCAT
 AT
 TTCTAATCCCGAGACTGTAGAGCTGGTGTCTTTTTCTAAAGGATGTCAGTAGAGAAGTGGAGTCCCCAAAA
 TT
 ACAGTTTCACGTATTAGTCAAGTTTCTAAAAACAGTAATAATGTTGAGAGCTGACATAGGGACTAACTTGGTT
 TT
 TTTTTTTTTTTTTTTTTTCAAATTCCTACTGAACTTTGATTTTGCTAAATAAGGACATTAAAAAAAAAAACAAAA
 AACTCCACTATTGCCTATTGCCACTATTTGATTTTTAAAAATAAGCGTATTTTAGCATCTAAAAGTAGGAAGG
 A
 CCTCAAATAAATGAGTCTTTGTTCTTGGCCAGGAAAAACAGCGTTGTCAGAATTTGATAAAGTGTCTTTCTAGGG
 TA
 TGTGCTGTTATTTCAGTTAAACCTTGCCCTGGGACGCTAGCATTGAGTAAATACTTGTGAATAAGCAAATGAAA
 CT
 TAAGCTTCTATGTATAGAAACCTAAGTCACTTCACATTCTGATTAGCAGAGTAATTGAATATTCTTTTCAATGTG
 T
 AGCTCTATCCCCAGAACACAGAATATTGAACTGTAAAGGCCATCTATAGTTAAACCAACTGCGTTAAATAG
 AT
 AATAGAAAAGATGTGGTATGTGGCAGTGACAACCTGAAGGTTGTGACTAGAACTCGGGTCTCTGGAGTGTCTA
 TTA
 TATCACACCAAGCTGGTCACCAGCCCATGTGTTGATCCTCCATTGTGATAGCAACAAAGAAAAGACTTCAGGAC
 AT
 TCTTTCCTTACCCTAATCCTTGATCTGCAGTCTTATTTAGAAAAGCTTAATGTTAAAGATCTAGTTTATTCAA
 A
 CTAAGATAACAAGGAGTATGAGAATTTCTATTTCCGGAGTGTAAGGAGGAGATGTTTCTTGGCTTCTCTGAG
 CC
 TGCAGGCCCTTCTTGCTCTTTAAGGAAGTAGAGAGAGGGAGGAAAGTAAAGTATGCTTTTGTTTTTAAGGTTA
 CT
 TTGCTGGGAGTAGTTTGCATGCCTTTTGGTTTTCTTGGGTGGAATTAAGTACTTAAGTTTAAAGTAGTTGGGA
 CT
 ATTTAAAAACAATGCCTATCCAATGTTTGCATAAAGGCAGAGGGTATTGGCTTTAGAAGTTAATCTTCTCCA
 GG
 AGTGAAAATTAGCTTCTAAACCAGAAGCAGCAGAGCTAAATAAAGTAATTTCCACCTGGCCAGTGCATGATGT
 GA
 AAGGTAGATTAAAAAATGAGAGGGCCATTTTCTGATGAAAGACTAAGCCATGTTGAAACAGCCCTGTTGAG
 GAT
 TTTATTTTAAATCTATACATTCACAAAGGAGCTTTGTGTATGTCTTCCCTATTTGTTGTTGGACTAGGAAGCC
 C
 CACCCAGTGCTTGTGAAGGCAGAAAAGTCGTTGAAAAGCAAGCTGGGATTGAAACAGTGGATTGAGGTTTCGAA
 TAT
 CCAGTGAACCAAAATATATCAGGGTTCCCTGGCCAAGATGAGTGACCATTCTGAGGTGTTAAGTATTTCTTGA
 AT
 GGGGATTTAGGAAAAGTTTCTGTATTTCTGTGCTCATTTTGTGACCTCTGTATGTGCAAAATCTCTAAGGGG
 GT
 GTTTGGGCACTTAGATTTCTTGGATGCAGATTTGTTTGTATATGAAACAAATTTTAAATGTTTTGTATACACTG
 G
 ATTTAAATAGTTTACTAAAGTGTTTTAATTTTTTTCATCTTAATTTTACAGTTCTTATAGTCTTTAGATTTAGGG
 AGGCTGTTGATGGCATCCACATGTGCATTTTAGTGGCATTAAAAATGTATTCAGCTGAATTTAAACAATTTCTGA
 CC
 TAAAACCTGACATTTTAGATTTAAGTCGGTAAAGCACTGATTTAAACTGGATTTTAACTGGATGAAATTCGATT
 T
 AATAAGTGTACTGACTGGATAAAATGCCAATGATTTAATTAACAAGCACGTTTAAACAGGATGCCCTATATATTA
 GT
 TAAAAGTGAAGCAATTGAATTAGGTACCTTCTCTGCTGCGTGGAAAAGACCGTATGACTCACCCACACCAGCCT
 TC
 TCTTCGCTCTGAGTGTAGCTAACCGTTTCTGTTTTTTTTCTCTAGGGTTTGGAAATCCCTTGTCTCCAGGTTGC
 T

GGGATTGACTTCTTGCTCAATTGAAACACTCATTCAATGGAGACAAAGAGAACTAATGCTTTGTGCTGATTCAT
AT
TTGAATCGAGGCATTGGGAACCCCTGTATGCCTTGTGTTGTGGAAAGAACCAAGTGACACCATCACTGAGCTTCCTA
AA
AGTTCGAAGAAGTTAGAGGACTATACACTTTCTTTTGAACTTTTATAATAAATATTTGCTCTGGTTTTTGGAAAC
C
AGGGCTGTTAGAGGGGTGAGTGACAAGTCTTACAAGTGGCCTTATTCCAATCCAGAAATTGCCCAAAGGAAC
TTT
GAGATTATATGCAATCGAAAGTGACAGGAAACATGCCAACTCAATCCCTCTAATGTACATGGATGGGCCAGA
AGT
GATTGGCAGCTCTCTTGGCAGTCCGATGGAGATGGAGGATGCCTTGTCAATGAAAGGGACCGCTGTTGTTCCA
TTC
CGAGCTACACAAGAAAAAATGTCATCCAAATCGAGGGGTATATGCCCTTGGATTGCATGTTCTGCAGCCAGA
CCT
TCACACATTGAGAAGACCTTAATAAACATGCTTAATGCAACACCGGCCTACCCTCTGTGAACCAGCAGTTCCT
CG
GGTTGAAGCAGAGTATCTCAGTCCGCTTGATAAAAGTCAAGTGCGAACAGAACCTCCAAGGAAAAGAAATTGC
AAG
GAAAATGAATTTAGCTGTGAGGTATGTGGGCAGACATTTAGAGTCGCTTTTGATGTTGAGATCCACATGAGAAC
AC
ACAAAGATTCTTCACTTACGGGTGTAACATGTGCGGAAGAAGATTCAAGGAGCCTTGGTTTTCTAAAAATCAC
AT
GCGGACACATAATGGCAAATCGGGGGCCAGAAGCAAACCTGCAGCAAGGCTTGGAGAGTAGTCCAGCAACGAT
CAAC
GAGGTCTCCAGGTGCACGCGGCCGAGAGCATCTCCTCTCCTTACAAAACTGCATGGTTTTGTGGCTTCTATT
TC
CAAAATAAGAAAAGTCTAATTGAGCACCGCAAGGTGCACACCAAAAAAAGTCTTTCGGTACCAGCAGCGCGCA
GAC
AGACTCTCCACAAGGAGGAATGCCGCTCTCGAGGGAGGACTTCTGTCAGTTGTTCAACTTGAGACCAAAATCT
CAC
CCTGAAACGGGGAAGAAGCCTGTGAGATGCATCCCTCAGCTCGATCCGTTACCACCTTCCAGGCTTGGCAGC
TGG
CTACCAAAAGGAAAAGTTGCCATTTGCCAAGAAGTGAAGGAATCGGGGCAAGAAGGGAGCACCGACAACGACG
ATTC
GAGTTCGAGAAGGAGCTTGGAGAAACAAATAAGGGCAGTTGTGCAGGCCTCTCGCAAGAGAAAAGAGAAGTG
CAAA
CACTCCACGGCGAAGCGCCCTCCGTGGACGCGGATCCCAAGTTACCCAGTAGCAAGGAGAAGCCCACTCACT
GCT
CCGAGTGGCGCAAAGCTTTCAGAACCTACCACCAGCTGGTCTTGCATCCAGGTTCCACAAGAAGGACCGGAG
GGC
CGGCGCGGAGTCGCCCACCATGTCTGTGGACGGGAGGCAGCCGGGGACGTGTTCTCCTGACCTCGCCGCCCC
TCTG
GATGAAAATGGAGCCGTGGATCGAGGGGAAGGTGGTTCTGAAGACGGATCTGAGGATGGGCTTCCCGAAGGA
ATCC
ATCTGGGTAAGCTGCCCTGTCTCCGTCCCGTCTGTTCGGCTGTGTCTGTCTGTCTCCCCGTCTCCCCCTCT
TA
TTCCCATCTCCAGACAACGCTGGCCAGGAATGGGGTTGGAGAGCCAGAGTCAAGTCCAGGCTCTTTTTGGTA
TCA
CTCTGTGTAAGTCATTTAACCTCTCAGGGCCTAATTTTCTCATTCTGTAATAACAGGGTTGAGTTAAGAGGTC
T
CCTTGTCTGAAAATATATATATATTTTTTAAACGTGTATCGTTTTTGTCTCACAAAAACACTTTAAAAAATAA
A
ACTTGTGCATCCAGCCCAAATGCACTGCTTCTTAACTGGGGCGATTTTGTCCCAATCAGTATCTGGCAATGTC
TG
GAGGCATTTTGGTTGTCATACTGTGTGTGTGGGTGTGCTGCTGGCATCCAGTGGGCAGAGGCCAGGGACACT
GCT
CAGCATGGTACAGTGCACAGGACAGCCCCATCATCAAAGAATTATCTGGTCCCAAATGCAATAGTTTGAAGCAT
TG
AGAGACCCTAGCCTTCACTTAAGTTTTTCTGGCGTTCCTGATCTTTTTCTGTAGTGAATTTCTAGTGGCCATAAA
A
GGTACTGGGAGTGATCAACTAGAGCCAGGAATATTATTTGGGCAGCCGTTTGGTGCTGTCCAAAACCTTGTCC
TTT
CTGTCTGGCAAGCTAGTATCCATTTATAGGTACCTCAGGAACCCAAATGATTTGTCAATAAATAACAAGGAATGT
GA
GCACACTGAAGACATTTTAAAGAAGGCTCATTTGCTCAGCAGAATTTTCAAGTGTACTAGTGGCATTATAGAAA
GA
GAAGGTGATCACTGAAGGCATGCTCACATAATATTCTGAGCCCTGGTGGCGTTATCTAGGGCAAAGGATTC
CAC
CTGTGTTTGGAGTTGCGCCCATCTCACTGTAGCCAGAGCTTCTCCTATCAGAGTTTAGTATTTGTTGAATA
GA

GGATCTTGCTGCTTAAAAACAGTTGAAAAGACCCTGATGGCCAGGCCGTAATTGACAAGCGAATGATGGGAACA
TGA
ATCGGTCTTAGGGAAGCATCTGTCAAAGTGGTCTTGGTTAAAAACAAGTGCCTCCTCCTCAGTGTCACCTGA
TT
GTGTGCTTGAATCTTTCGGAAAACTGGGTGTATGAGACCCACGATGAATTTGCCACACGATTGATTGGACTCT
TC
CTTCACCTGCTCTTCAGCCAGTGCCAGTTCCTTTTCTGATCATGTGATTGACGTGAGAAGTGTAGTCTGTATAT
CA
AATCTTTAGAAATGTTTTTGTAGTTTCTGGGACACAGGAAACCCAGCACTTAGCATACTACAAATCTAATGTCTT
AA
TGGCATCATAAAAAAGAGGCTTTAAACACAGACTCCAGTTAGCTAAGTGGTTTCTGCTAGTGCCGGTACTGTTC
AG
GGGCCCTGTGAGATGCCCCAGTTCCTGAAAAGAAATGAAAAGGCCAGTTACCGGTAGGTGGTGTGGAAAAACAT
CGG
CTAGATCATCAGGCAGGACAGAATGCCTGGCTGTGGGTGGGAGCACCCAGCTTGGCGTTGAGTTCGGTTCT
ACC
ACTGCGTTGTTTTGTGACCAATTATGAGTTGCTTAACCTTTCTTTGCTACTATTTCCCTGTTTGCAAAAATGGTTC
A
TFGACCCTGTCTTCCACCTCCCAAGGACAATTTCAACAGCCTATTTGTAAAAAGATCACAGTCTTTAAAAAAT
A
TAACTGTAAGTCAAGGTTGATGCTTGAAGAGCAGGAACCCAGGTAGATGTGGAATGTGATGCTCTTTGTTTC
TAA
AGAAAAGGCATTTTATAGCTTTTTGGATATGACGCAACATACCATAAATCCTGACACATAGTTGGGAGTCGGAA
AT
TGCAACAACGCCAGTTATAAACCCAGCTAGTTGGGTATGATTGTAAGAAAAAAGCTGGCCATTCTGTATT
TG
GGGAATTGATTTTCTAACTTATATTATCTTAGTAGTCTAGATTTATCATATTGTACTATCATCTGGCTTTTTT
AAGACTTAAGAAAGATCAAGTAAATTTTTTTTTCTTTCTTTAGACACTATATAGATCATCAAGGGTGTCTGTCTTA
C
AGGTGGATAGTGATATGATCTACAGTGAGGGGACATTTATTTAAAACTTAAACATTCATGTGTTTTGGGGGTGG
TA
TTTTAACGGCAGCACCTCTGATTGTCTTTTGGAGGGCTGGTGTGTGTTTGAAGTTCTGTCTCCTTCCAGTGGA
CT
CTAACTTCTCTGATGCACGTGAGACACATTGTCTATTGTCTGACAGAACTAAAGCCAAACACTGTGATCTG
GG
GACAGGTTTTTCATTTGTGAGATCTCTTTCGCCACATGAGTGTGTTGTTGGACAATACAGCCTGCTTTCCAAAACT
TT
GCTAAATTTTACAGACTTTCTAGGTGCTTGCCCAATGCCAGACTTTCTTTTCTGTTGAAGATTAAGTTGTGCT
T
GCTGCCCTCTAGTGGTCAGTTGTTAATCCTAACCTTAAACGGCTTATTTTTCCCTGGTGGTGGGAAGTTGA
CG
GTTTGTAAATGGCTCATTTTTCTAAATTAATCTGAAGAAGATAATTTTTCCCGCCAGTATGTATGTCCACCTTCA
G
TTTCCAGATCCTGCCTGCTCAGAGACACTGAGAACCAGGAAAGCTGCCCGGCAATTCAGTCTATGAAATGATC
TTT
CTTGTGATTAAGGCAAAACGAAGAACTGAATGTTAATAGTGTACTCTGCTGTACCCAGAAAAAACAACAAA
AT
CATGTTATAACACTCTAAAACTTCAACAACCTCCAACAGCATTGGTGTGTGTCTAGCCGTTTTGTTCTAACCC
G
ATGTTATATAAAGAATTTTTTCATGCTTTCCAAAAATGTTTATGTCAAGAATATTTAAGTCAGCATGCCTTATT
C
AGGTACTTCAGCTACCTTCTTATATAAATATTTTTGTTTTTCTTTAAGATAAAAAATGATGATGGAGGAAAAATA
A
AACATCTTACATCTTCAAGAGAGTGTAGTTATTGTGGAAAGTTTTCCGTTCAAATTATTACCTCAATATTCATC
T
CAGAACGCATACAGGTAAGAACTTTTATTTTTTAACCATGCATTAGTTAAATTATGTAGTTATCTAATTTTTT
T
GTTGTTGTTGTTTTCAGATACTCTGCCAGATCCTTGGACTAGCTTAAGGATAAATATGTAGCATGTTGATTGCAGT
GG
TTATTTTTATTCTTTTAGTGCCATTGTAACCTGAGCCATTGFTCTTATTTGCAGTTCATTTCTTTTCTTTT
TGTTTTTTGAGACGGAGTCTTGCTCTGTACCTCGGCTGGAGTGCAGTGGTGAATTTTCGGCTCACTGCAGCCT
CC
ACCTCCCTGGTTCAAGCAATACTCCTGCCTCAGCCTCCCCAGTAGTTGGGATTACAGGTACCTGCCACCACACC
CG
GCTAATTTCTGTATTTTTAGTAGAGATGGGGTTTACCATGCTGGCCAGGCTGGTTTGAACCTCTGACCTCAA
GT
GATCCGCTCACCTTGGCCTCCCATAGTGTGGCCTCCCATAGTGTGGGATTACAGGCGTGAGCCACCGCGCC
CGG
ACAAAAGTTCATTTGTTAGTTTATGACTGCTATGTCTGACTCTTATCTTATTTAAAAAGCTACAGTATTTTTAAAT
G

CTGCATCTTATGTCTTTATGATTGAGAATGAAATGAGAATCTATTTAGTAGTCTTGAGATTGTGAAAGGAGCTA
 TG
 ACATCATGATGTAGGAGGCTGCGTAGATTTGAAATTCATCTCTCCACTTACTATCTGTGCACCCCTGGGCAA
 GT
 TATTTAACCTTTTTGTGCTTTTGTGCTTTTCTTTGCTGTAAAAGTAGAATAATACATATTTCCCTAGGGCTGTTAGG
 A
 AGATTAATAAGTTAGAAGTGTGCTGTTAATTTTTCTATTGAAGATAGGCATTCATAATTTCAAATATTCATTA
 C
 AGTAAGGATGATAAAGAAGTGTGAGAAATCCTATGTGATAGTAGATCGAGAAAGCAAAGGAGGAAAAGAACG
 CTG
 TTTTCTTAATAAATAGATATTTGATCTATTTTCAGTGTCTTTTCATACACTTCTATAATAAAGTGCCATTTCTTGCT
 TAGGTGAAAAACCATACAAATGTGAATTTTGTGAATATGCTGCAGCCAGAAAGACATCTCTGAGGTATCACTTG
 GA
 GAGACATCACAAGGAAAAACAACCGATGTTGCTGTGCTGAAGTCAAGAACGATGGTAAAAATCAGGACACTGAA
 GAT
 GCACATTAACCGCTGACAGTGCAGCAAAACAAAAATTTGAAAAGATTTTTGATGGTGCCAAAGATGTTACAGG
 CA
 GTCCACCTGCAAAGCAGCTTAAGGAGATGCCTTCTGTTTTTCAGAATGTTCTGGGCAGCGCTGTCTCTCACCA
 GC
 ACACAAAGATACTCAGGATTTCCATAAAAAATGCAGCTGATGACAGTGTGATAAAGTGAATAAAAAACCTACCC
 CT
 GCTTACCTGGACCTGTTAAAAAAGAGATCAGCAGTTGAAACTCAGGCAAATAACCTCATCTGTAGAACCAAGG
 CGG
 ATGTTACTCCTCCTCCGGATGGCAGTACCACCCATAACCTTGAAGTTAGCCCCAAAGAGAAGCAAACGGGAGAC
 CGC
 AGCTGACTGCAGATACAGGCCAAGTGTGGATTGTCACGAAAAACCTTTAAATTTATCCGTGGGGGCTCTTCACA
 AT
 TGCCCCGCAATTTCTTTGAGTAAAAAGTTTGATTCCAAGTATCACCTGTCCATTTTGTACCTTCAAGACATTTTAT
 C
 CAGAAGTTTTAATGATGCACCAGAGACTGGAGCATAAATACAATCCTGACGTTTAAAAAAGTGTGAAACAAG
 TC
 CTTGCTTAGAAGTGCAGCTACCGGATGCCCGCCAGCGTTGCTGGGAAAAGATGTGCTCCCTCCCTCCTAGTTTC
 TGT
 AAACCCAAGCCCAAGTCTGCTTTCCCGGCGCAGTCCAAATCCCTGCCATCTGCGAAGGGGAAGCAGAGCCCTC
 CTG
 GGCCAGGCAAGGCCCTCTGACTTCAGGGATAGACTCTAGCACTTTAGCCCAAGTAACCTGAAGTCCACAG
 ACC
 ACAGCAGAATGTGGGGTCCAAGGGGCGCCACCAGGCAACAGCAATCTGAGATGTTTCTAAAACAGTGT
 TCC
 CCTGCACCGGATAAGACAAAAAGACCCGAGACAAAATTGAAACCTTTCAGTAGCTCCTTCTCAGCCACCCT
 CG
 GCAGCAGTAACATCAATGGTTCCATCGACTACCCCGCCAAGAACGACAGCCCGTGGGCACCTCCGGGAAGAGA
 CTA
 TTTCTGTAATCGGAGTGCCAGCAATACTGCAGCAGAATTTGGTGAGCCCTTCCAAAAAGACTGAAGTCCAGC
 GTG
 GTTGCCCTTGACGTTGACCAGCCCGGGGCCAATTACAGAAGAGGCTATGACCTTCCCAAGTACCATATGGTCA
 GAG
 GCATCACATCACTGTTACCGCAGGACTGTGTGTATCCGTGCAGGCGCTGCCCTCCAAACCAAGGTTCTGAG
 CTC
 CAGCGAGGTCGATTCTCCAAATGTGCTGACTGTTTACAAGCCCTATGGTGGCTCCGGGCCACTTTACACTGT
 GTG
 CCTGCTGGTAGTCCAGCATCCAGCTCGACGTTAGAAGGATTGTCATGAGGGGCGTCTGTTTAAATGGCTGCC
 TAC
 AGTGATTAATAGCTAATCCAGGCATTCTCAGTGGAGATGTTACCCTCCCAAGGGTGGGGGGTAGGCAGCCAG
 AAG
 TTCTGGGGGTACAGAGAGAAGCATTCTTAGATACGGCAGTGGTTTGTGGTCTCCAAAGGCTTACTTAACTCT
 GT
 GGGTTAACTCTTAACCTGTGTATTTTATCTTTTGTATTGTTTAGTCTTACTTTATTTTAGAGAAAGGGTCTT
 GCTCCGTCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGA
 TC
 CTTCTGCCCTCAGCTTCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTT
 G
 TAGAGATGGAGTTGCCAGGCTGGTCTTGAACCTCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTA
 TCT
 TAGACTACAGATGCACTCCACCAGCTTG

**ATGCAATCGAAAGTGACAĠGAAACATGCCAACTCAATCCCTCTTAATGTACATGGATGGGCCAGAAGTGATTG
GCA**

GCTCTCTTGGCAGTCCGATGGAGATGGAGGATGCCTTGTCATGAAAGGGACCGCTGTTGTTCCATTCCGAGC
 TAC
 ACAAGAAAAAATGTCATCCAAATCGAGGGGTATATGCCCTTGGATTGCATGTTCTGCAGCCAGACCTTCACAC
 AT
 TCAGAAGACCTTAATAAACATGTCTTAATGCAACACCGGCCTACCCTCTGTGAACCAGCAGTTCTTCGGGTTGA
 AG
 CAGAGTATCTCAGTCCGCTTGATAAAAAGTCAAGTGCGAACAGAACCTCCCAAGGAAAAAAGATTGCAAGGAAAA
 TGA
 ATTTAGCTGTGAGGTATGTGGGCAGACATTTAGAGTCGCTTTTGATGTTGAGATCCACATGAGAACACACAAAAG
 AT
 TCTTTCACCTACGGGTGTAACATGTGCGGAAGAAGMITSRRSSAGCCTTGGTTTCTTAAAAATCACATGCGGAC
 AC
 ATAATGGCAAATCGGGGGCCAGAAGCAAAGTGCAGCAAGGCTTGGAGAGTAGTCCAGCAACGATCAACGAGG
 TCGT
 CCAGGTGCACGCGGCCGAGAGCATCTCCTCTCCTTACAAAATCTGCATGGTTTGTGGCTTCTATTTCAAAATA
 AA
 GAAAGTCTAATTGAGCACCAGGAGTGCACACCAAAAAAAGTCTTTCGGTACCAGCAGCGCGCAGACAGACT
 CTC
 CACAAGGAGGAATGCCGTCCTCGAGGGAGGACTTCTGCAGTTGTTCAACTGAGACCAAAATCTCACCTGA
 AAC
 GGGGAAGAAGCCTGTGAGATGCATCCCTCAGCTCGATCCGTTACCACCTTCCAGGCTTGGCAGCTGGCTACC
 AAA
 GGAAAAGTTGCCATTTGCCAAGAAGTGAAGGAATCGGGGCAAGAAGGGAGCACCAGACAACGACGATTTCGAGT
 TCCG
 AGAAGGAGCTTGGAGAAACAAATAAGGGCAGTTGTGCAGGCCCTCTCGCAAGAGAAAAGAGAAGTGCAAAACT
 CCA
 CGGCGAAGCGCCCTCCGTGGACGGGATCCCAAGTTACCCAGTAGCAAGGAGAAGCCCACTCACTGCTCCGA
 GTGC
 GGCAAAGCTTTCAGAACCTACCACCAGCTGGTCTTGCACCTCCAGGGTCCACAAGAAGGACCGGAGGGCCGGC
 GCGG
 AGTEGCCACCATGTCTGTGGACGGGAGGACGCCGGGGACGTTCTCTGACCTCGCCGCCCTCTGGATGA
 AAA
 TGGAGCCGTGGATCGAGGGGAGGTGGTTCTGAAGACGGATCTGAGGATGGGCTTCCCGAAGGAATCCATCT
 GGAT
 AAAAAATGATGATGGAGGAAAAATAAAACATCTTACATCTTCAAGAGAGTGTAGTTATTGTGGAAAAGTTTTCCG
 TT
 CAAATTATTACCTCAATATTACATCTCAGAACGCATACAGGTGAAAAACCATACAAATGTGAATTTTGTGAATAT
 GC
 TGCAGCCCAGAAGACATCTCTGAGGTATCACTTGGAGAGACATCACAAAGAAAAACAAACCGATGTTGCTGCT
 GAA
 GTCAGAACGATGGTAAAAATCAGGACACTGAAGATGCACTATTAACCGCTGACAGTGCAGCAAAACAAAAATT
 TGA
 AAAGATTTTTGATGGTGCCAAAGATGTTACAGGCAGTCCACCTGCAAAGCAGCTTAAGGAGATGCCTTCTGTT
 TT
 TCAGAATGTTCTGGGCAGCGCTGTCTCTCACCAGCACACAAAGATACTCAGGATTTCCATAAAAAATGCAGCTG
 AT
 GACAGTGCTGATAAAGTGAATAAAAAACCTACCCCTGCTTACCTGGACCTGTTAAAAAAGAGATCAGCAGTTGA
 AA
 CTCAGGCAAATAACCTCATCTGTAGAACCAAGGCGGATGTTACTCCTCCTCCGGATGGCAGTACCACCATAAC
 CT
 TGAAGTTAGCCCCAAAGAGAAGCAAACGGAGACCGCAGCTGACTGCAGATACAGGCCAAGTGTGGATTGTCAC
 GAA
 AAACCTTTAAATTTATCCGTGGGGGCTCTTACAATTGCCCGCAATTTCTTTGAGTAAAAGTTTGATTCCAAG
 TA
 TCACCTGTCCATTTTGTACCTTCAAGACATTTTATCCAGAAGTTTTAATGATGCACCAGAGACTGGAGCATAAA
 TA
 CAATCTGACGTTCCATAAAAACTGTGGAACAAGTCCTTGCTTAGAAGTCGACGTACCGGATGCCCGCCAGCG
 TTG
 CTGGGAAAAAGATGTGCCCTCCCTCTCTAGTTTCTGTAACCCAAAGCCCAAGTCTGCTTTCCCGGCGCAGTCCAA
 AT
 CCCTGCCATCTGCGAAGGGGAAGCAGAGCCCTCCTGGGCCAGGCAAGGCCCTCTGACTTCAGGGATAGACTC
 TAG
 CACTTTAGCCCCAAGTAACTGAAGTCCACAGACCACAGCAGAATGTGGGGGTCCAAGGGGCCGCCACCAGG
 CAA
 CAGCAATCTGAGATGTTTCTAAAACAGTGTTCCTCCTGCACCGGATAAGACAAAAAGACCCGAGACAAAATT
 GA
 AACCTCTCCAGTAGCTCCTTCTCAGCCCACCCTCGGCAGCAGTAACATCAATGTTCCATCGACTACCCCGCC
 AA
 GAACGACAGCCCGTGGGCACCTCCGGGAAGAGACTATTTCTGTAATCGGAGTGCCAGCAATACTGCAGCAGAA
 TTT

GGTGAGCCCTTCCAAAAAGACTGAAGTCCAGCGTGGTTGCCCTTGACGTTGACCAGCCCGGGGCCAATTACA
GAA
GAGGCTATGACCTTCCCAAGTACCATATGGTCAGAGGCATCACATCACTGTTACCGCAGGACTGTGTGTATCC
GTC
GCAGGCGCTGCCTCCCAAACCAAGGTTCTGAGCTCCAGCGAGGTCGATTCTCCAAATGTGCTGACTGTTGAG
AAG
CCCTATGGTGGCTCCGGGCCACTTTACTTGTGTGCTGCTGGTAGTCCAGCATCCAGCTCGACGTTAGAAG
GTC
TTGGTGGATGTCAGTGCTTACTCCCATGAAATTAATTTTACTTCATCCTTTGAGAAGCGAATGGTGAAGCT
AC
TGAAATAAGCTGTGATTGTACTGTACATAAAACATATGAGGAATCTGCAAGGAACACTACAGTTGTGTAA

SEQ. ID. No. 11
Proteína ZABC

5

**MQSKVTGNMPTQSLLMYMDGPEVIGSSLSGSPMEMEDALSMKGTAVVFFRATQEKVNIQIEGYMPLDCMFCSTFT
H**
SEDLNKBVLMQHRPTLCEPAVLRVEAEVLSPLDKSQVRTEPPKEKNCKENEFSCVCGQTFRVAFDVEIHMRTKID
SFTYGCNMCGRJODDPWFLKNHMRTHNGKSGARSKLQQGLESPATNEVVQVHAAESSIPYKICMVCGLFPNK
ESLIEHRKVBTKKTAFGTSSAQTDSPQGGMPSSREDFLQLFNLRPKSHPETGKKPVRCIPQLDPFTTFQAWQLATE
GKVAICQEVKESGQEGSTDNDSSSEKELGETNKGSCAGLSQEKEKCKHSHGEAPSVADPKLPSSKEKPTHCSCEC
GKAFRTYBQLVLSRVBKKDRRAGAESPTMSVDGRQPGTCSFDLAAPLDENGAVDRGEGGSEDGSEDGLPEGIHLD
KNDDGGKIKHLTSSRECSYCGKFFRSNYLNIHLRTHTEKPYKCEFCYAAAQKTSLRVHLERHHKEKQTDVAEE
VKNDGKNQDTEALLTADSAQTKNLKRFFDGAKDVTGSPPAKQLKEMPSVFQNVLGSAVLSPAHKDTQDFBKNA
D
DSADKVNKNPTPAYLDLLKKRSVETQANNLICRTKADVTPPPDGSTTHNLEVSPKEKQTETAADCRYRPSVDCHE
KPLNLSVGALENCPAISLSKSLPSITCPFCFKTFYPEVLMNHQRLEHKYNDVHKNCRNKSLLSRRTGCPPAL
LGKDVPLSSFCPKPKSAFFAQSKSLPSAKGKQSPPGPKAPLTSGIDSSTLAPSNLKSERPOQNVGVQGAATRQ
QQSEMFPKTSVSPAPDKTKRPETKLIKPLPVAPSQPTLGSNNINGSIDYPAKNDSPWAPPGRDYFCNRSASNTAAEF
GEPLPKRLKSSVVALDQPGANYRRGYDLPKYHMVRGITSLLPQDCVYPSQALPPKPRFLSSSEVDSPNVLTVQK
PYGGSGPLYTCVPAGSPASSSTLEGLGQCCLLPMKLNPTSSFEKRMVKATEISCDCCTVRIKTYEESARNTTVV

SEQ. ID. No. 12
1b1

10

GGAAACAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCC
ACC
GCGGTGGCGGCCGCTCTAGAAGTAGTGGATCCCCGGGCTGCAGGAATTCGGCACGAGGCTCCACCGACAGC
CAGG
CACTGGGCAGCACGCACTGGAGACCCAGGACCCTGTGCAGGAGCAGCTCCGGGTGACACGAGGGGACTGAAG
ATAC
TCCCACAGGGGCTCAGCAGGAGCAATGGGTAACCAAATGAGTGTCCCCAAAGAGTTGAAGACCAAGAGAATG
AAC
CAGAAGCAGAGACTTACCAGGACAACCGCTCTGCTCTG AACGGGGTTCCAGTGGTGGTGTGACCCACACAGT
TCA
GCACCTAGAGGAAGTCGACTTGGGAATAAGTGTCAAGACGGATAATGTGGCCACTTCTTCCCCCGAGACAACG
GAG
ATAAGTGTCTTGGCGGATGCCAACGGAAAGAATCTTGGGAAAAGAGGCCAAACCCGAGGCACCAGCTGCTAAAT
CTC
GTTTTTCTTGATGCTCTCTCGGCCTGTACCAGGACGTACCCGGAGACCAAGCCGCAGATTTCATCCCTTGGATCA
GT
GAAGCTTGATGTCAGCTCCAATAAAGCTCCAGCGAACAAAGACCCAAGTGAGAGCTGGACACTTCCCGTGGCA
GCT
GGACCGGGGCAGGACACAGATAAAACCCAGGGCACGCCCCGGCCCAAGACAAGGTCTCTCTGCCGCCAGG
GATC
CCACGCTTCTCCACCTGAGACAGGGGGAGCAGGAGGAGAAGCTCCCTCCAAGCCCAAGGACTCCAGCTTTTT
TGA
CAAATTTCTTAAGCTGGACAAGGGACAGGAAAAGGTGCCAGGTGACAGCCAACAGGAAGCCAAGAGGGCAGA
GCAT
CAAGACAAGGTGGATGAGGTTCTTGGCTTATCAGGGCAGTCCGATGATGTCCCTGCAGGGAAGGACATAGTTG
ACG
GCAAGGAAAAAGAAGGACAAGAAGCTTGGAACTGCGGATTGCTCTGTCCCTGGGGACCCAGAAGGACTGGAGA
CTGC
AAAGGACGATCCCAGGCAGCAGCTATAGCAGAGAATAATAATTCATCATGAGTTTCTTTAAAAGCTGTTT
CA
CCTAACAAAGCTGAAACAAAAAAGGACCCAGAAGACACGGGTGCTGAAAAGTCACCCACCACTTCAGCTGACC
TTA
AGTCAGACAAAGCCAACCTTTACATCCCAGGAGACCCAAGGGGCTGGCAAGAATTCCAAAGGATGCAACCCATC
GGG
GCACACACAGTCCGTGACAACCCCTGAACCTGCGAAGGAAGGCACCAAGGAGAAATCAGGACCCACCTCTCTG
CCT

CTGGGCAAAC TGTTTTGGAAAAAGTCAGTTAAAGAGGACTCAGTCCCCACAGGTGCGGAGGAGAATGTGGTGT
 GTG
 AGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAGGAGA
 TGC
 TGCACCTGAACCCACAGAAGCGAAACTCAAAAGAGAAGAAAGCAAACCAAGAACCTCTCTGATGGCGTTTCTC
 AGA
 CAAATGTCAGTGAAGGGGATGGAGGGATCACCCACTCAGAAGAAATAAATGGGAAAGACTCCAGCTGCCAAA
 CAT
 CAGACTCCACAGAAAAGACTATCACACCGCCAGAGCCTGAACCAACAGGAGCACCACAGAAGGGTAAAGAGG
 GCTC
 CTCGAAGGACAAGAAAGTCAGCAGCCGAGATGAACAAGCAGAAGAGCAACAAGCAGGAAGCCAAAGAACCCAGC
 CCAG
 TGCACAGAGCAGGCCACGGTGGACACGAACCTCACTGCAGAATGGGGACAAGCTCCAAAAGAGACCTGAGAAG
 CGGC
 AGCAGTCCCTTGGGGGCTTCTTTAAAGGCCTGGGACCAAAGCGGATGTTGGATGCTCAAGTGCAAACAGACCC
 AGT
 ATCCATCGGACCAGTTGGCAAACCCAAAGTAAACAAATCAGCACGGTTCCCACCAGGTTCTCTGCCACCAAGA
 TGT
 GTTCTCCTTACTCCATCTCCTCCCCAAACAGCTCCATGTATATATTCTTCTGATGGCCAGCAAATGAAATCTG
 C
 CTAGAAATTAAGCCCGAGCTGTTGTATATTGAGGTGTATTATTTACGTCTCTGGTCCAGTCTTTTCTGGCAAAT
 AA
 CAGTAAAGATGGTTTAGCAGGTCACCTAGTTGGGTCAGAAGAGTCGATGATCACCAAGCAGGAAAGGGAGGG
 AATA
 GAGGAATGTGTTCCGGGTTAAGTGATGAAAATGGCAGTGGTGGCCGGCCGTGGTGGCTCTCGCCTGTAATCTCA
 GCA
 CTTTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACTAGCCTGGCCAACATCATGAAACC
 CCG
 TCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCACACACCTGTAGTCCCAGCTACTCGGGAGCCCAA
 CGC
 ACGAGAACCCTGTGATCCAGGAGGTGGAGGTTGCAGTGAGCCGAAGTTGCACCATTGCACTCCACCCTGGGC
 GAC
 AGAGCAAGATTCTATCAAAAAAAAAAGGCAGTGGCAAGTAAAGTTATAGAAGAGAAATGCTGCTAGAAGGAATT
 AAG
 CGTTGTAGTAAACCGCTGCTCATCCTCTAAGCTTGAAGAAGGGAGACGAAAATCCATTTGTTTAAATTCACATC
 TC
 AAGGAGGGAGAACCCGGGCTGTGTTGGGTGGTTGCCAATTTCTAGAACGGAATGTGTGGGGTATAGAAAA
 GGAA
 TGAATAAGCGTTGTTTTCAAATAGGGTCTTGTAAAGTTATTGATGAGAGGGAAAAGATTGACTGGGGAGGGC
 TTA
 AAATGATTGGGAAAAAATGCTTTTGGGGCTCAGTGACAACGGCAAAGATTACAACCTAAAAAAAAAAAAAAAA
 AA
 AAACCGAGACTAGTTCTCTCTCTCTCTCGTGCCGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGG
 GG
 GGGCCCGGTACCCAATTCGCCCTATA

SEQ. ID. No. 13

Secuencia genómica procedente del clon 97 de BAC

5 Secuencia filtrada en la búsqueda:

> query_seq

12. El ácido nucleico aislado del párrafo 1, en el que la subsecuencia se hibrida específicamente en condiciones rigurosas a la SEQ. ID. No. 7.
- 5 13. El ácido nucleico aislado del párrafo 12, en el que la subsecuencia es la SEQ. ID. No. 7.
14. El ácido nucleico aislado del párrafo 1, en el que la subsecuencia se hibrida específicamente en condiciones rigurosas a la SEQ. ID. No. 8.
- 10 15. El ácido nucleico aislado de los párrafos 14, 16, 18, 20, en el que la subsecuencia es la SEQ. ID. No. 8.
16. El ácido nucleico aislado del párrafo 1, en el que la subsecuencia se hibrida específicamente en condiciones rigurosas a la SEQ. ID. No. 9.
- 15 17. El ácido nucleico aislado del párrafo 16, en el que la subsecuencia es la SEQ. ID. No. 9.
18. El ácido nucleico aislado del párrafo 1, en el que la subsecuencia se hibrida específicamente en condiciones rigurosas a la SEQ. ID. No. 10.
- 20 19. El ácido nucleico aislado del párrafo 1, en el que la subsecuencia es la SEQ. ID. No. 10.
20. El ácido nucleico aislado del párrafo 1, en el que la subsecuencia se hibrida específicamente en condiciones rigurosas a la SEQ. ID. No. 12.
- 25 21. El ácido nucleico aislado del párrafo 20, en el que la subsecuencia es la SEQ. ID. No. 12.
22. El ácido nucleico aislado del párrafo 1, en el que la subsecuencia se hibrida específicamente en condiciones rigurosas a la SEQ. ID. No. 13.
- 30 23. El ácido nucleico aislado del párrafo 22, en el que la subsecuencia es la SEQ. ID. No. 12.
24. El ácido nucleico aislado del párrafo 1, que comprende adicionalmente una secuencia promotora unida de manera operable a la secuencia de polinucleótidos.
- 35 25. El ácido nucleico aislado del párrafo 1, cuyo ácido nucleico es una molécula de ADNc.
26. Un procedimiento de selección para células neoplásicas en una muestra, el procedimiento que comprende:
 la puesta en contacto de una muestra de ácidos nucleicos procedente de un paciente humano con una sonda que se hibrida selectivamente a una secuencia de polinucleótidos diana que comprende una secuencia seleccionada del
 40 grupo constituido por la SEQ. ID. No. 1, SEQ. ID. No. 2, SEQ. ID. No. 3, SEQ. ID. No. 4, SEQ. ID. No. 5, SEQ. ID. No. 6, SEQ. ID. No. 7, SEQ. ID. No. 8, SEQ. ID. No. 9, SEQ. ID. No. 10, SEQ. ID. No. 11, SEQ. ID. No. 12, y, SEQ. ID. No. 13 en el que la sonda se pone en contacto con la muestra en condiciones en las que la sonda se hibrida selectivamente con la secuencia del polinucleótido diana para formar un complejo de hibridación estable; y
 45 detectar la formación de un complejo de hibridación.
27. El procedimiento del párrafo 26, en el que la muestra de ácidos nucleicos procede de una paciente con cáncer de mama.
- 50 28. El procedimiento del párrafo 26, en el que la muestra de ácidos nucleicos procede de un núcleo en propagación durante la metafase o de un núcleo en interfase.
29. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 1.
- 55 30. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 2.
31. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se

expone en la SEQ. ID. No. 3.

32. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 4.

5

33. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 5.

34. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 6.

10

35. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 7.

36. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 8.

15

37. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 9.

20

38. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 10.

39. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 12.

25

40. El procedimiento del párrafo 26, en el que la sonda comprende una secuencia de polinucleótidos como se expone en la SEQ. ID. No. 13.

41. El procedimiento del párrafo 26, en el que la sonda se utiliza para identificar la presencia de una mutación en la secuencia de polinucleótidos diana.

30

42. Un procedimiento para detectar una célula neoplásica en una muestra biológica, el procedimiento que comprende:

35

la puesta en contacto de la muestra con un anticuerpo que se une específicamente a un antígeno polipeptídico codificado por una secuencia de polinucleótidos que comprende una secuencia seleccionada del grupo constituido por la SEQ. ID. No. 1, SEQ. ID. No. 2, SEQ. ID. No. 3, SEQ. ID. No. 4, SEQ. ID. No. 5, SEQ. ID. No. 6, SEQ. ID. No. 7, SEQ. ID. No. 8, SEQ. ID. No. 9, SEQ. ID. No. 10, SEQ. ID. No. 12, y SEQ. ID. No. 13; y

40

detectar la formación de un complejo antígeno-anticuerpo.

43. El procedimiento del párrafo 42, en el que la muestra procede de tejido de mama.

44. Un procedimiento para inhibir la proliferación patológica de células cancerosas, el procedimiento que comprende la inhibición de la actividad de un producto génico de un gen endógeno que presenta una subsecuencia que se hibrida en condiciones rigurosas con una secuencia seleccionada del grupo constituido por la SEQ. ID. 1, SEQ. ID. No. 2, SEQ. ID. No. 3, SEQ. ID. No. 4, SEQ. ID. No. 5, SEQ. ID. No. 6, SEQ. ID. No. 7, SEQ. ID. No. 8, SEQ. ID. No. 9, SEQ. ID. No. 10, SEQ. ID. No. 12, y SEQ. ID. No. 13.

50

45. Un procedimiento para detectar un cáncer, dicho procedimiento que comprende la detección de la sobre-expresión de una proteína codificada en un amplicón 20q13.

46. El procedimiento del párrafo 45, en el que dicha proteína codificada en un amplicón 20q13 es ZABC1.

55

47. El procedimiento del párrafo 45, en el que dicha proteína codificada en un amplicón 20q13 es 1b1.

REIVINDICACIONES

1. Una molécula aislada de ácidos nucleicos que comprende una secuencia de polinucleótidos que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 o una de sus secuencias complementarias.
5
2. El ácido nucleico aislado de la reivindicación 1, en el que la secuencia de polinucleótidos comprende las SEQ. ID. Nos. 9 o 10.
- 10 3. El ácido nucleico aislado de la reivindicación 1, que comprende adicionalmente una secuencia promotora unida de manera operable a la secuencia de polinucleótidos.
4. El ácido nucleico aislado de la reivindicación 1, que es una molécula de ADNc.
- 15 5. Un procedimiento de selección para células neoplásicas en una muestra, el procedimiento que comprende:

la puesta en contacto de una muestra de ácidos nucleicos procedente de un paciente humano con un ácido nucleico aislado que se hibrida selectivamente a una secuencia de polinucleótidos diana que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 o una de sus secuencias complementarias, o a una subsecuencia de cualquiera de estas susodichas secuencias en las que el ácido nucleico aislado se pone en contacto con la muestra en condiciones en las que el ácido nucleico aislado se hibrida selectivamente con la secuencia de polinucleótidos diana para formar un complejo de hibridación estable, y
20
25 detectar la formación de un complejo de hibridación para controlar el número de copias relativo de secuencias correspondientes a las SEQ. ID. Nos. 9 o 10 en la muestra.
6. El procedimiento de la reivindicación 5, en el que la muestra de ácidos nucleicos procede de un paciente con un cáncer seleccionado del grupo constituido por cáncer de ovario, vejiga, cabeza y cuello, colon y mama.
30
7. El procedimiento de una cualquiera de la reivindicación 5 o la reivindicación 6 en el que el ácido nucleico aislado está marcado.
- 35 8. El procedimiento de una cualquiera de la reivindicación 5 o la reivindicación 6 en el que el ácido nucleico aislado está inmovilizado sobre una superficie sólida.
9. El procedimiento de la reivindicación 5, en el que la muestra de ácidos nucleicos es un núcleo en propagación durante la metafase o un núcleo en interfase.
40
10. Un procedimiento de selección para células neoplásicas en una muestra, el procedimiento que comprende:

la puesta en contacto de una muestra de ácidos nucleicos procedente de un paciente humano con un ácido nucleico aislado que se hibrida selectivamente a una secuencia de polinucleótidos diana que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 o una de sus secuencias complementarias, o a una subsecuencia de cualquiera de estas susodichas secuencias en las que el ácido nucleico aislado se pone en contacto con la muestra en condiciones en las que el ácido nucleico aislado se hibrida selectivamente con la secuencia de polinucleótidos diana para formar un complejo de hibridación estable, y
45
50 detectar la formación de un complejo de hibridación para identificar la presencia de una mutación en la secuencia de polinucleótidos diana.
11. Un procedimiento para la detección de células neoplásicas en una muestra biológica, dicho procedimiento que comprende la selección de la muestra para la sobre-expresión del gen ZABC1, en donde la sobre-expresión del gen ZABC1 con respecto al tejido normal, es indicativa de una célula neoplásica,
55
en donde el gen ZABC1 corresponde a cualquiera de las SEQ. ID. Nos. 9 o 10.

12. Un procedimiento de acuerdo con la reivindicación 11, en el que la selección de la muestra para la sobre-expresión del gen ZABC1 comprende la puesta en contacto de la muestra con un anticuerpo que se une específicamente a un antígeno polipeptídico codificado por una secuencia de polinucleótidos que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10 y la detección de la formación de un complejo antígeno-anticuerpo.
13. El procedimiento de la reivindicación 11 o 12, en el que la muestra se selecciona del grupo constituido por tejido de ovario, vejiga, cabeza y cuello, colon y mama.
- 10 14. Un procedimiento de acuerdo con la reivindicación 11, dicho procedimiento que comprende la selección de la muestra para la sobre-expresión de una proteína codificada por una secuencia de polinucleótidos que comprende una secuencia seleccionada del grupo constituido por las SEQ. ID. Nos. 9 o 10.
15. Un procedimiento según una cualquiera de las reivindicaciones 5 a 9 u 11 a 14 para su utilización en el diagnóstico de la aparición o progresión de un cáncer en un individuo del que se ha derivado la muestra biológica, en el que una sobre-expresión de 3 veces o superior del gen ZABC1 en la muestra biológica está correlacionada con una disminución de la supervivencia libre de enfermedad en sujetos con ganglios negativos.
16. Un procedimiento para diagnosticar la aparición o progresión del cáncer en un individuo, 20 procedimiento que comprende las etapas de:
- (a) determinación de la expresión del gen ZABC1 en una muestra biológica procedente del individuo, y
- (b) comparar el nivel de expresión del gen ZABC1 observado en la muestra biológica con el nivel de expresión del 25 gen ZABC1 derivado de una segunda muestra de tejido clínicamente normal,
- en el que una sobre-expresión de 3 veces o superior del gen ZABC1 en la muestra biológica está correlacionada con una disminución de la supervivencia libre de enfermedad en sujetos con ganglios negativos.

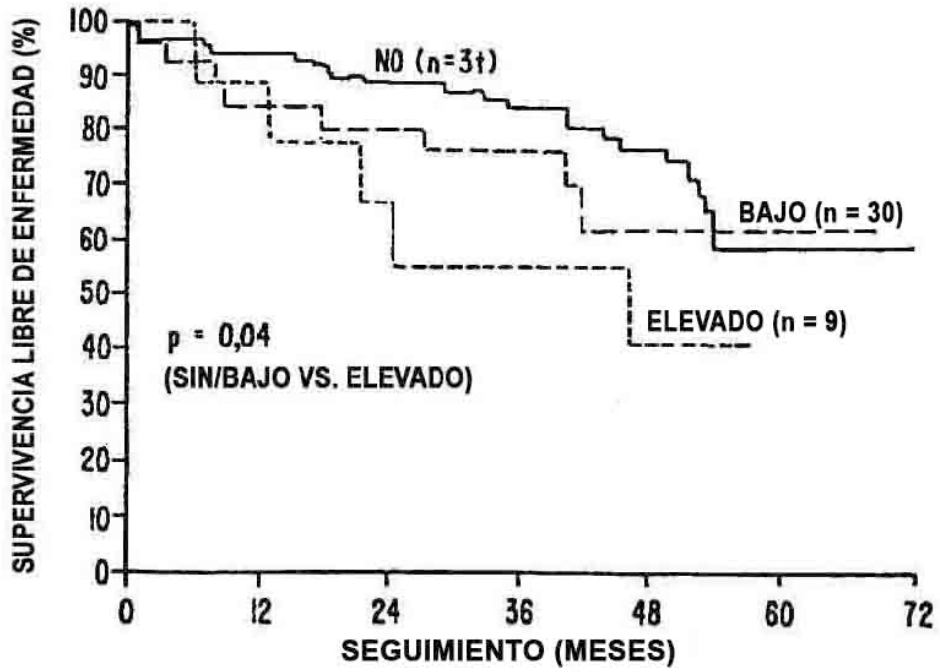


FIG. IA.

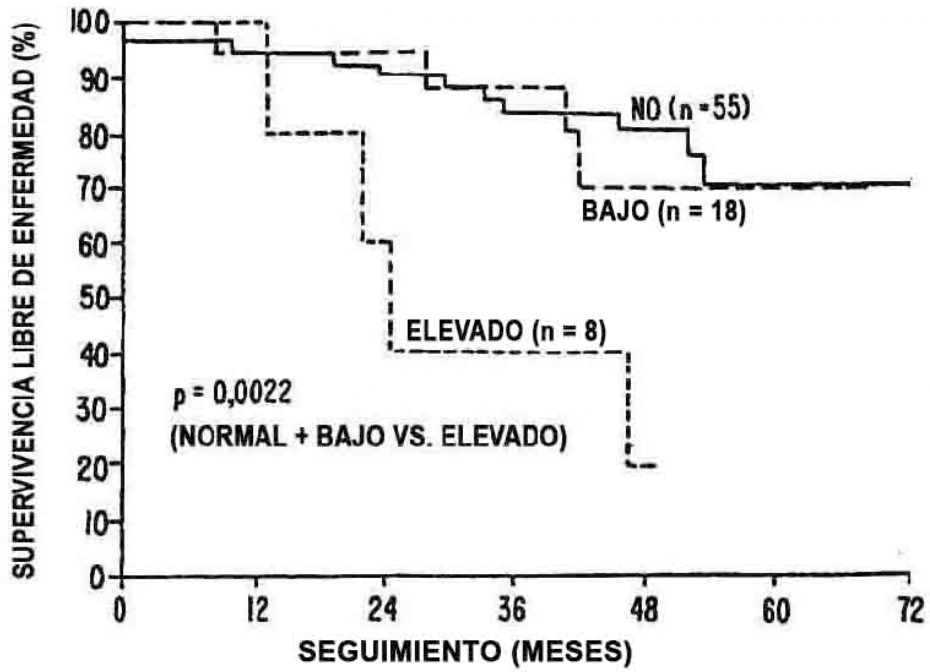


FIG. IB.

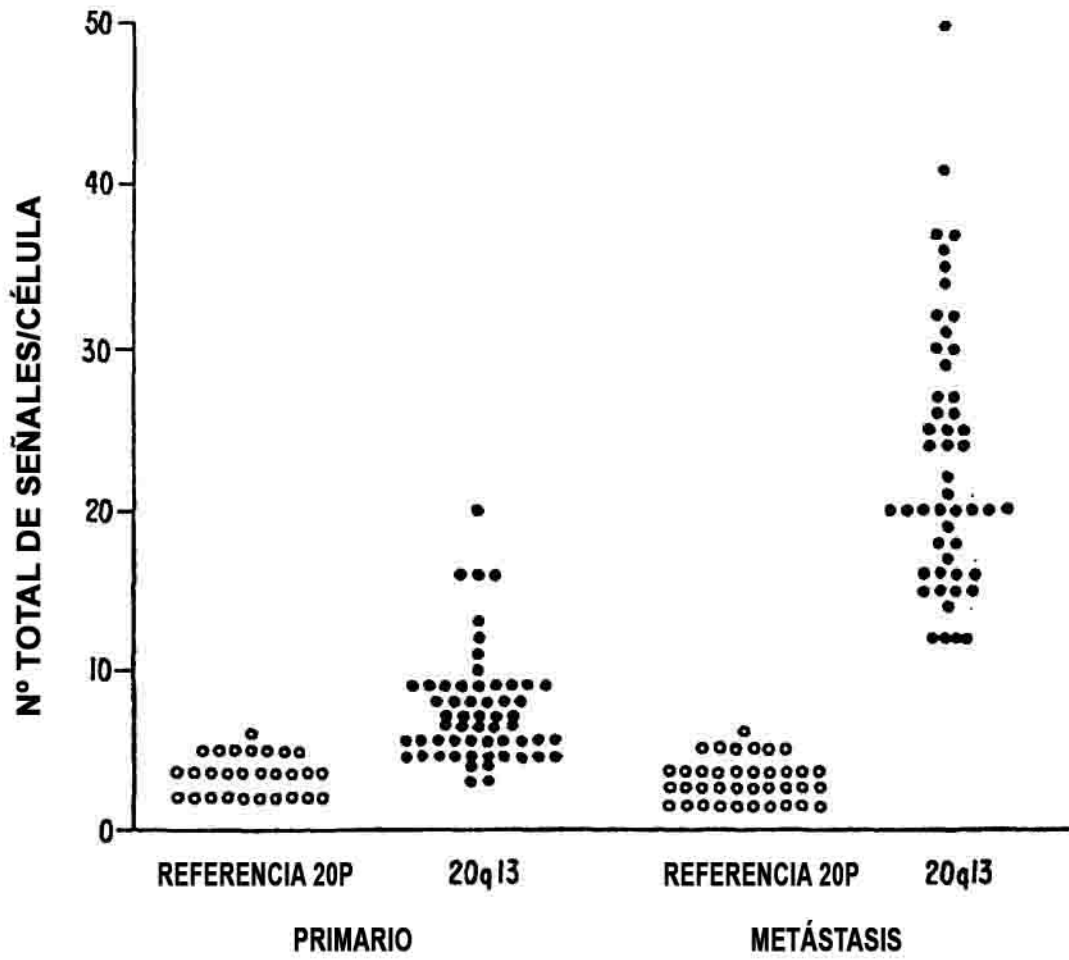


FIG. 2.

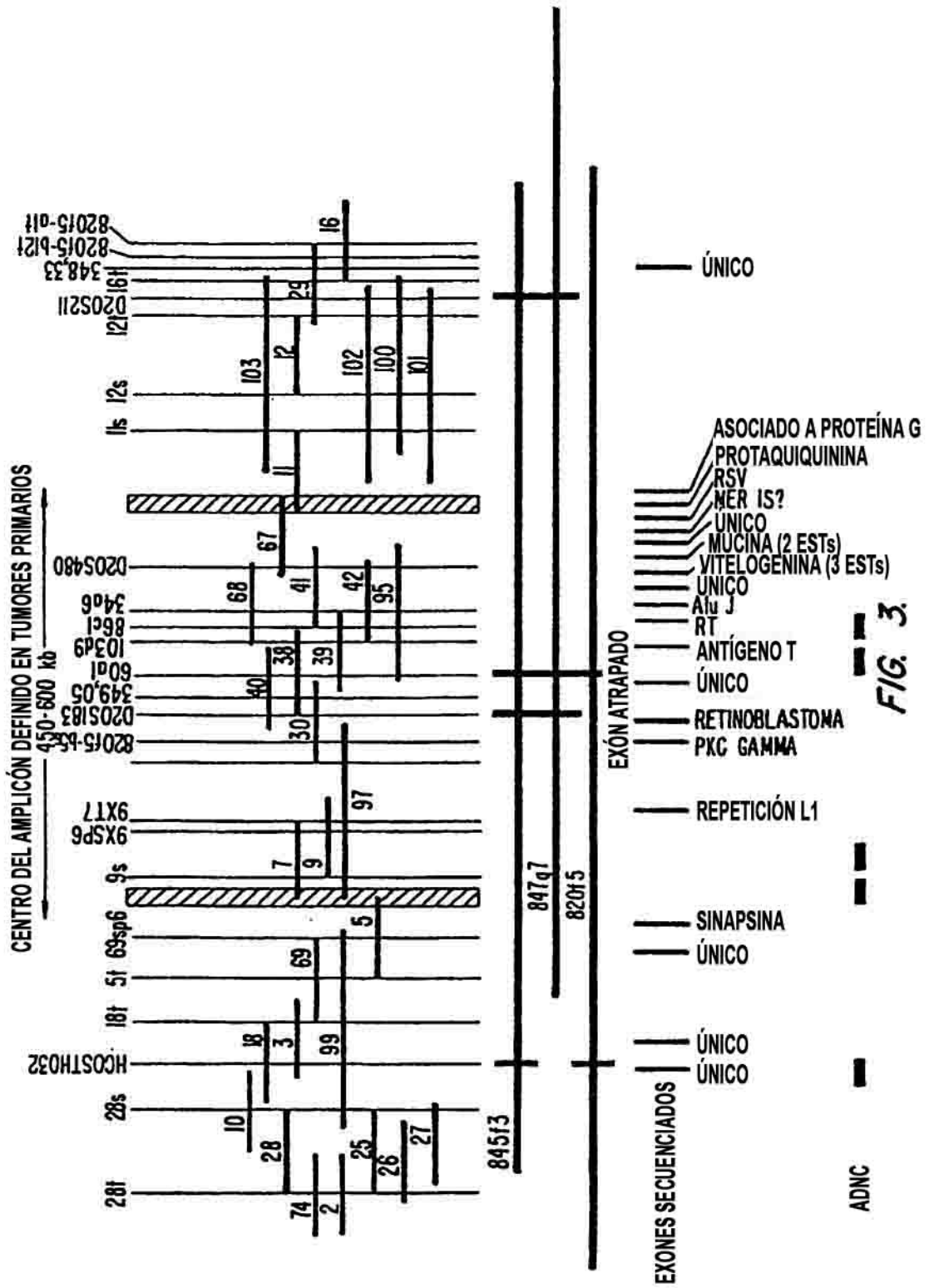


FIG. 3

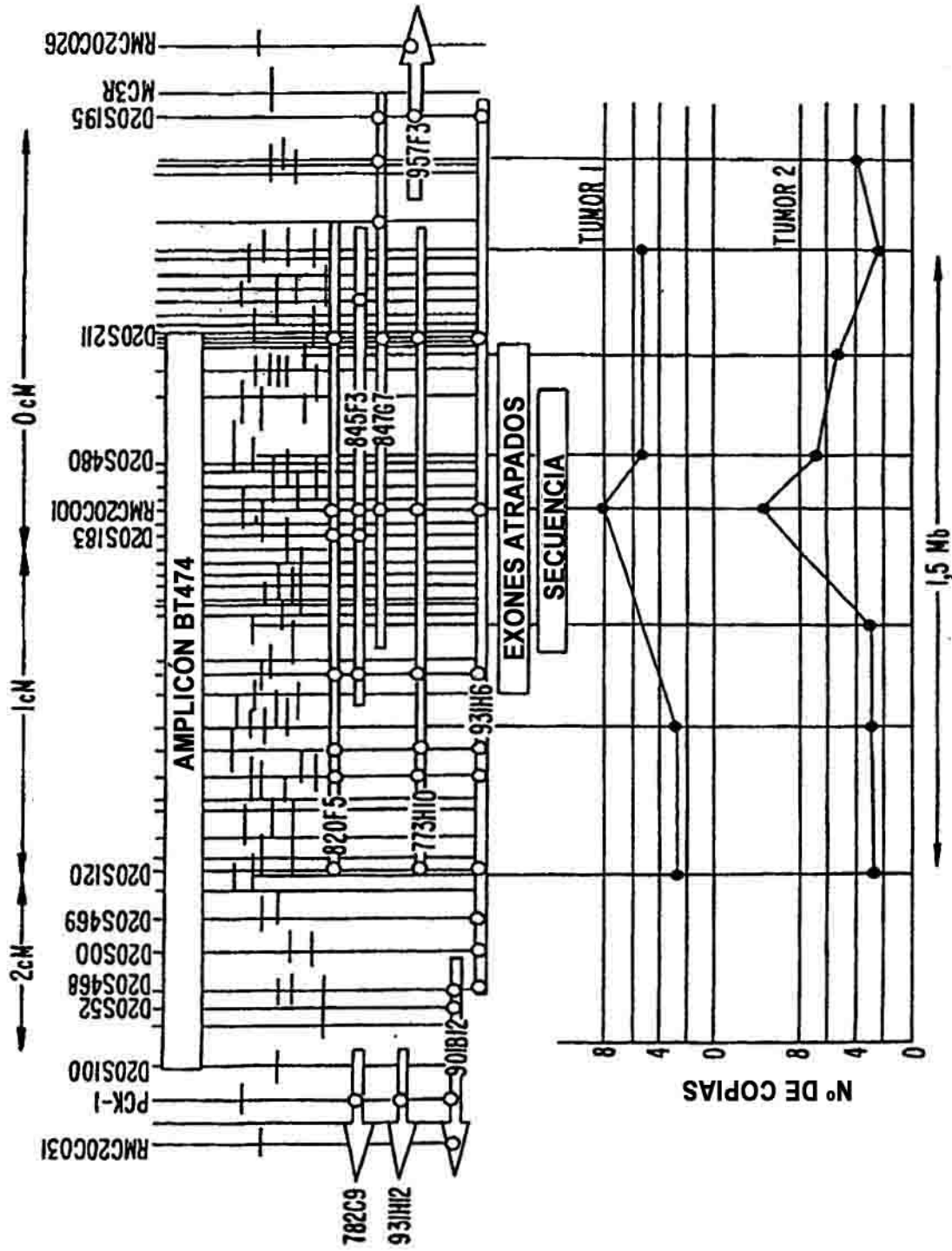


FIG 4

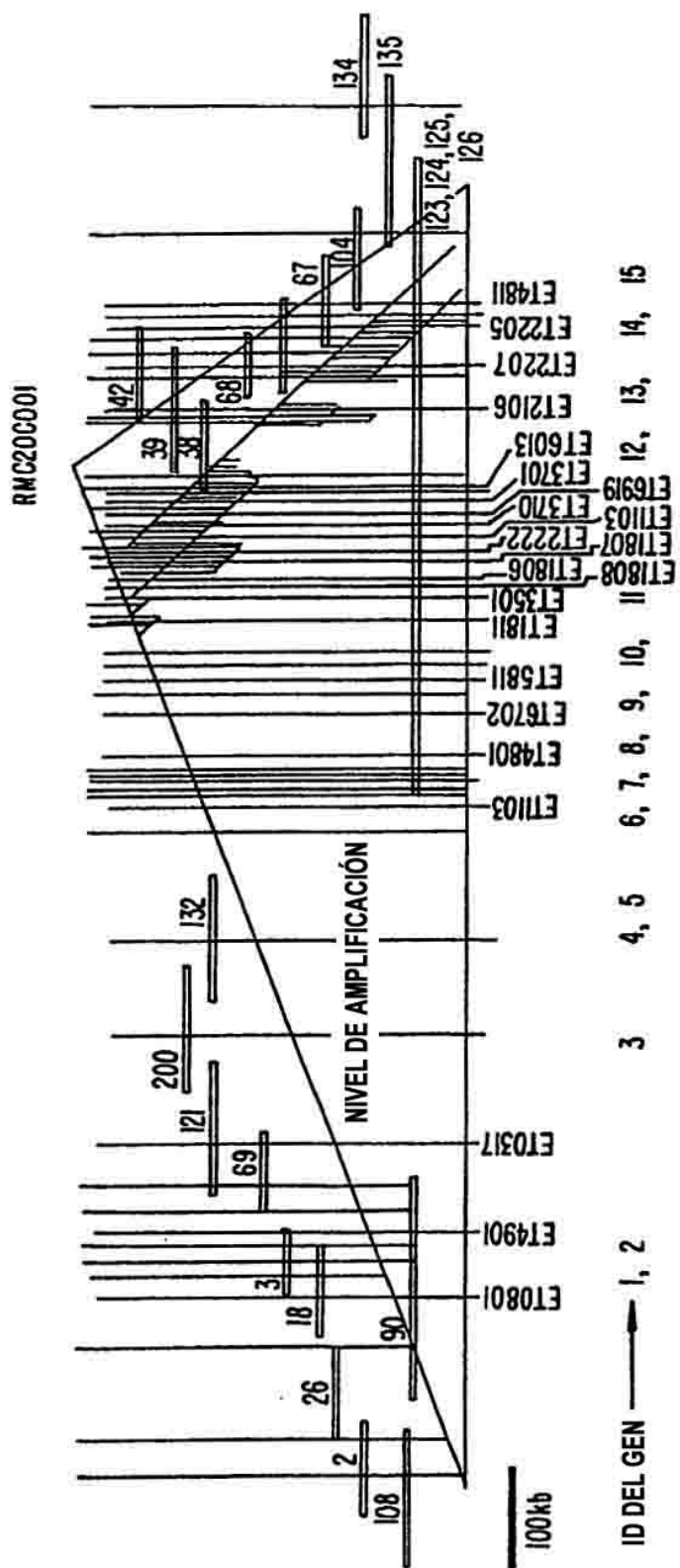


FIG. 5.

Secuencia genómica procedente del clon 97 de BAC
Secuencia filtrada en la búsqueda:

```
> query seq
TGATATGGATTCATGCCCTCTTGCACCTTGCCAAACATCACACGCTG
CCATCCAGTCCACTCGATTTTGGCAGTGCAGATGAAAACGCGAACCAT
TTGTGTTGAGTCCAGCAAGATGCCAGGACCTGCATGTTTCAGAACGAAGT
TCTTCATCATCCAATTTCTCCCTGTATATGGGCTTACCACNACTGCCGTT
AAGTCGTGTNAAGTCACCCTCAGGTACATAATGGAATAATTCTGCAAAG
GCAGGAGNCACTTTCTCTCCAGTGTCTCAGACCATGAAAAGTTTCTGAATGT
CTTTGGAACCTTTGTCTGCAAATAGCTCGAAGGAGACATGGCCTAAAGGCT
CGCCATCTGCCGTGATATTGNAACAATGGTAGGCCTGACCCTGGCTGTGGC
CATGACTTTTTACANTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNCCAAATGCCGGACAGAGAATCNAAGAAACTGTATTA
GGGAAAGGGCTCCTGAGTTTATGCCAAAGTTTCCAGATTGCTTTCCATTG
AAACGTAGCTCTGTGAGATAACCATCAGGTGTATGTTGAAGAAATGCTGTT
GTAGTCAAATATGTTTGTAGTGAGTGAGCCTGAGCTGAGCAAGACTTTACT
GCAAGACTTCAGACTCTTCGTCCCTTTTATGCTAATGGGTAACACAAAC
TCCAAAGTGGGGTGTACAGCATGAGGCATTAACAAAATTTATTGGACC
CCACACACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCTC
```

gb | M19533 | RATCYCA ARNm de ciclofilina de rata, cds completo.
Longitud = 743

HSPs de la hebra menos:

Puntuación = 418 (115,5 bits), Esperado = 1,5e-58, Suma P(5) = 1,5e-58
Identidades = 96/112 (85%), Positivos = 96/112 (85%), Hebra = Menos /
Más

Búsqueda: 372 TNCAA?ATCACCGCAGATGGCGAGCCCTTAGGCCATG?CTCCCTCGAGCTATTTGCAGAC
313
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Objeto: 64 TTCGACATCACGGCTGATGGCGAGCCCTTGGGTCGCGTCTGCTTCGAGCTGTTGCAGAC
123

Búsqueda: 312 AAAGTTCCAAAGACATCAGAA?ACTTT?CATGGTCTGAGCACTGGAGAGAAAG 261
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Objeto: 124 AAAGTTCCAAAGACAGCAGAAA?ACTTTCTGTGCTCTGAGCACTGGGGAGAAAG 175

Puntuación = 236 (65,2 bits), Esperado = 1,5e-58, Suma P(5) = 1,5e-58
Identidades = 52/58 (89%), Positivos = 52/58 (89%), Hebra = Menos / Más

Búsqueda: 117 TGCTGGACTCAAACACAAATGGTTCACAGTTTTCATCTGCACTGCCAAAATCGAGTGG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Objeto: 348 TGCTGGACCAAACACAAATGGTTCACAGTTTTCATCTGCACTGCCAAGACTGAGTGG
405

Puntuación = 177 (48,9 bits), Esperado = 1,5e-58, Suma P(5) = 1,5e-58
Identidades = 41/48 (85%), Positivos = 41/48 (85%), Hebra = Menos / Más

Búsqueda: 60 GACTGGATGGCAAGCGTGTGATGTTTGGCAAGGTGCAAGAGGGGCATGA 13
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Objeto: 404 GGCTGGATGGCAAGCATGTGGTCTTTGGGAAGGTGAAA?GAAGGCATGA 451

Puntuación = 154 (42,5 bits), Esperado = 1,5e-58, Suma P(5) = 1,5e-58
Identidades = 34/38 (89%), Positivos = 34/30 (89%), Hebra = Menos / Más

Búsqueda: 153 AGAACTTCG?TCTGA?AACATGCAGGTTCCTGGCATCTTG 115
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Objeto: 299 AGAACTTCATCCTGAAGCATA?AGGTCCTGGCATCTTG 336

Puntuación = 86 (23,8 bits), Esperado = 1,5e-58, Suma P(5) = 1,5e-58
Identidades = 22/28 (78%), Positivos = 22/28 (78%), Hebra = Menos / Más

Búsqueda: 256 TCCTGCCTTTGCAGAA?TATTCCATTA? 229
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Objeto: 193 TCCTCCTTTCACAGA?TATTCCAGGA? 220

Figura 6

Mapa físico del amplicón 20q13.2 de cáncer de mama

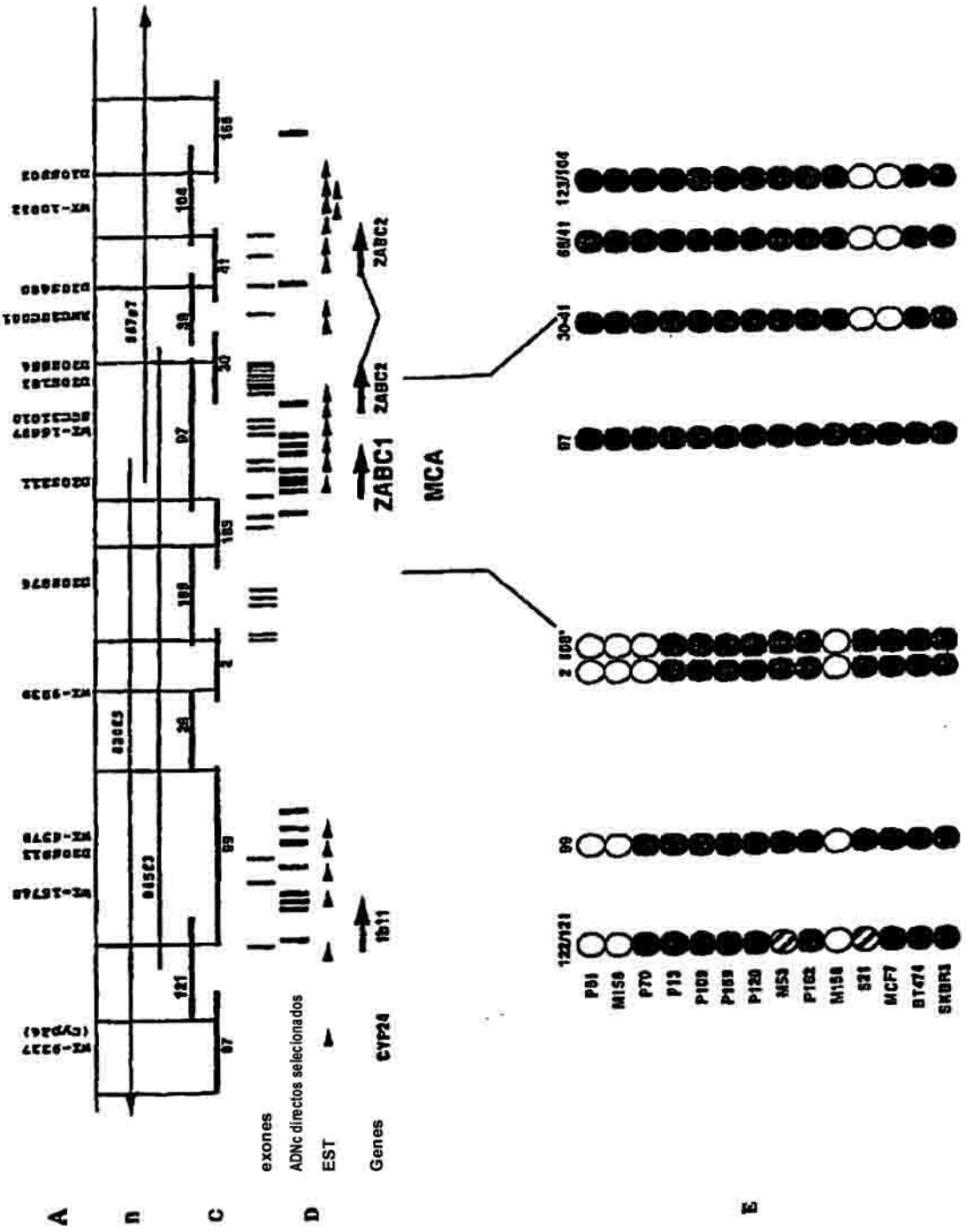


Figura 7