

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 547 053**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.07.2008 E 08794916 (0)**

97 Fecha y número de publicación de la concesión europea: **15.04.2015 EP 2183379**

54 Título: **Enriquecimiento de una secuencia diana**

30 Prioridad:

**01.08.2007 US 962838 P**

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

**01.10.2015**

73 Titular/es:

**DANA FARBER CANCER INSTITUTE (100.0%)  
44 BINNEY STREET  
BOSTON, MA 02155, US**

72 Inventor/es:

**MAKRIGIORGOS, GERASSIMOS**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

**ES 2 547 053 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

Enriquecimiento de una secuencia diana

## ANTECEDENTES

5 Una situación habitualmente encontrada en análisis genético implica la necesidad de identificar un porcentaje bajo de secuencias de ADN variantes ("secuencias diana") en presencia de un gran exceso de secuencias no variantes ("secuencias de referencia"). Los ejemplos de tales situaciones incluyen: (a) la identificación y secuenciación de unos pocos alelos mutados en presencia de un gran exceso de alelos normales; (b) la identificación de unos pocos alelos metilados en presencia de un gran exceso de alelos no metilados (o viceversa) en análisis epigenético; (c) la identificación y genotipado de unas pocas secuencias de ADN fetales que circulan en la sangre de la madre, en la que también está presente un gran exceso de secuencias de ADN de la madre; y (d) la identificación de ADN circulante de tumores en sangre de pacientes con cáncer (o personas que se sospecha que tienen cáncer) en presencia de un gran exceso de alelos de tipo salvaje.

15 Aunque se han descrito recientemente métodos de identificación de alto rendimiento para mutaciones de línea germinal o somáticas de alto predominio (Thomas, R.K., et al. (2007) *Nat Genet*, 39, 347-351; Chou, L.S., et al. (2005) *Am J Clin Pathol*, 124; 330-338; Thomas, R.K., et al. (2006) *Nat Med*, 12; 852-855), todavía sigue siendo problemática la detección de mutaciones somáticas de bajo predominio en tumores con heterogeneidad, contaminación estrómicica o en fluidos corporales. Y además, la significancia clínica de la identificación de estas mutaciones es importante en varias situaciones. Por ejemplo: (a) en adenocarcinoma pulmonar, las mutaciones de EGFR de bajo nivel que no se pueden identificar mediante secuenciación normal pueden conferir respuesta positiva a inhibidores de tirosina cinasas (Paez, J.G., et al. (2004) *Science*, 304; 1497-1500.) o resistencia a fármacos (Janne, P.A., et al. (2006) *Clin Cancer Res*, 12; 751-758), (b) las mutaciones en plasma útiles como biomarcadores para la detección temprana (Diehl, F., et al. (2005) *Proc Natl Acad Sci U S A*, 102; 16368-16373) o respuesta tumoral al tratamiento (Kimura, T., et al. (2004) *Ann N Y Acad Sci*, 1022; 55-60) no se pueden secuenciar usando métodos convencionales; y (c) las mutaciones en tumores con contaminación estrómicica frecuente, tales como pancreáticos o de próstata, se pueden "enmascarar" por la presencia de alelos de tipo salvaje, requiriendo así microsección laboriosa o perdiendo mutaciones en general.

20 El documento WO 03/072809 (véanse los ejemplos 1 y 2) describe protocolos de PCR que usan una temperatura de desnaturalización de 95°C que provocará la desnaturalización tanto de la secuencia diana como de la de referencia. Después, la temperatura se disminuye hasta primeramente 58°C, seguido de 72°C. Subsiguientemente, la temperatura se incrementa hasta una temperatura entre la  $T_m$  de la secuencia diana y de la secuencia de referencia. En las siguientes etapas, se deja que los cebadores se unan a la secuencia diana y se extiendan. Las secuencias metiladas de forma diferencial se pueden enriquecer durante la amplificación.

## SUMARIO DE LA INVENCION

35 La presente invención se refiere a métodos para enriquecer alelos poco abundantes procedentes de una muestra, como se define en las reivindicaciones.

El método se basa en parte en un protocolo de amplificación de ácidos nucleicos modificado que incluye incubar la mezcla de reacción a una temperatura de desnaturalización crítica o "Tc". Empleando la presente invención, se mejora enormemente los límites de detección actuales de todas las tecnologías a base de PCR.

40 La "temperatura crítica" o "Tc" se refiere a una temperatura por debajo de la temperatura de fusión " $T_m$ " de la secuencia de referencia. En algunas realizaciones, la Tc está por debajo de la  $T_m$  tanto de la secuencia de referencia como de la secuencia diana. La temperatura crítica aprovecha la menor  $T_m$  de la secuencia diana bicatenaria o del dúplex de ADN bicatenario hibridado de forma cruzada entre la diana y la referencia para desnaturalizar preferentemente estos dúplex con respecto a los homodúplex de referencia/referencia. Cuando la secuencia diana y la secuencia de referencia se hibridan de forma cruzada, las diferencias pequeñas de secuencia de uno o más mal apareamientos de un solo nucleótido en cualquier parte a lo largo de una secuencia de ADN bicatenario corta (por ejemplo, <200 pb) generarán un cambio pequeño pero predecible en la temperatura de fusión ( $T_m$ ) para esa secuencia (Lipsky, R.H., et al. (2001) *Clin Chem*, 47, 635-644; Liew, M., et al. (2004) *Clin Chem*, 50, 1156-1164). Se contemplan cambios de la temperatura de fusión de 0,1-20°C, dependiendo del contexto y posición exactos del mal apareamiento de la secuencia.

50 La temperatura de desnaturalización crítica (Tc) es la temperatura por debajo de la cual la eficiencia de la PCR cae abruptamente para una secuencia de ácidos nucleicos de referencia. Por ejemplo, una secuencia de p53 de 167 pb se amplifica bien si la temperatura de desnaturalización de la PCR se fija a 87°C, se amplifica de forma modesta a 86,5°C, y no produce producto detectable si la desnaturalización de la PCR se fija a 86°C o menos. Por lo tanto, en este ejemplo, Tc~ 86,5°C.

55 En un primer aspecto, la invención se refiere a un método para enriquecer una secuencia diana en una muestra de ácido nucleico que se sospecha que tiene secuencias diana y de referencia. El método incluye someter la mezcla de reacción de amplificación a una primera temperatura de desnaturalización que está por encima de la temperatura de

fusión "T<sub>m</sub>" de una secuencia de referencia. A continuación, la temperatura de la mezcla de reacción de amplificación se disminuye, permitiendo que las secuencias diana y secuencias de referencia monocatenarias se hibriden para formar moléculas bicatenarias. De este modo, la reacción incluye homodúplex de hibridación de diana-diana, referencia-referencia, y heterodúplex de hebras de diana-referencia. Por definición, un heterodúplex es un dúplex emparejado de forma imperfecta que no obstante contiene suficiente homología entre las hebras para mantener una forma de dúplex en la mezcla de reacción. Por definición, un homodúplex es un dúplex perfectamente apareado. La temperatura de la mezcla de reacción se incrementa entonces hasta la T<sub>c</sub>, dando como resultado la desnaturalización preferente de los dúplex de hibridación de secuencias diana-referencia. La T<sub>c</sub> o temperatura crítica está por debajo de la T<sub>m</sub> de la secuencia de referencia, y se puede determinar mediante los métodos descritos aquí. A la T<sub>c</sub>, los dúplex de secuencia diana-referencia (y los dúplex de secuencia diana-diana solamente si tienen una T<sub>m</sub> menor que la secuencia de referencia) están sustancialmente desnaturalizados, mientras que los dúplex de diana-diana (si tienen una T<sub>m</sub> igual o mayor que la T<sub>m</sub> de la secuencia de referencia) y los dúplex de secuencia de referencia-referencia están sustancialmente no desnaturalizados. "Sustancialmente" significa al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 80%, incluso más preferiblemente al menos 90%, y lo más preferible al menos 98% en una forma desnaturalizada o no desnaturalizada dada. Tras la desnaturalización preferente de los dúplex de diana-referencia y dúplex de diana-diana (aquellos que tienen una T<sub>m</sub> menor que la secuencia de referencia), se aplica una temperatura reducida a la mezcla de reacción para permitir que un par de cebadores se hibride a la secuencia diana. Los cebadores hibridados se alargan entonces, enriqueciendo así la secuencia diana con respecto a la secuencia de referencia en la muestra.

Se describe otro método para enriquecer una secuencia diana. En este método, una muestra de ácido nucleico sospechosa de contener cada una de una secuencia diana y una secuencia de referencia se desnaturaliza aplicando una primera temperatura de desnaturalización que está por encima de la T<sub>m</sub> de la secuencia de referencia. A continuación, las hebras diana y de referencia se hibridan entre sí para formar dúplex de secuencia diana-referencia bicatenarios. El dúplex de secuencia diana-referencia se forma y está presente en la mezcla de reacción junto con dúplex de secuencia diana-diana y referencia-referencia bicatenarios. Los dúplex de secuencia diana-referencia y diana-referencia bicatenarios se desnaturalizan preferentemente aplicando la T<sub>c</sub> a la muestra. A la T<sub>c</sub>, los dúplex de secuencia diana-referencia (y los dúplex de secuencia diana-diana solamente si tienen una T<sub>m</sub> menor que la secuencia de referencia) están sustancialmente desnaturalizados, mientras que los dúplex de diana-diana (si tienen una T<sub>m</sub> igual o mayor que la T<sub>m</sub> de la secuencia de referencia) y los dúplex de secuencia de referencia-referencia están sustancialmente no desnaturalizados. "Sustancialmente" significa al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 80%, incluso más preferiblemente al menos 90% y lo más preferible al menos 98%, en una forma desnaturalizada o no desnaturalizada dada. Entonces, un par de cebadores se hibrida a la secuencia diana y se alargan, incrementando así la concentración de la secuencia diana con respecto a la secuencia de referencia en la muestra.

Se describe además un método para enriquecer una secuencia diana llevando a cabo un protocolo de reacción de amplificación de ácidos nucleicos. El protocolo de reacción de amplificación incluye una primera temperatura de desnaturalización y una segunda temperatura de desnaturalización. La primera temperatura de desnaturalización está por encima de la T<sub>m</sub> de la secuencia de referencia, y la segunda temperatura de desnaturalización está por debajo de la T<sub>m</sub> de la secuencia de referencia.

Todavía se describe además un método para enriquecer una secuencia diana, que tiene una T<sub>m</sub> menor que la secuencia de referencia correspondiente sometiendo una mezcla de reacción de amplificación a una T<sub>c</sub>, reduciendo la temperatura de la mezcla de reacción y alargando un par de cebadores. La mezcla de reacción de amplificación se sospecha que contiene cada una de una secuencia diana y una secuencia de referencia. La T<sub>c</sub> está por debajo de la T<sub>m</sub> de la secuencia de referencia, permitiendo así la desnaturalización preferente de la secuencia diana con la menor T<sub>m</sub>. A la T<sub>c</sub>, los dúplex de secuencia diana-referencia y los dúplex de secuencia diana-diana están sustancialmente desnaturalizados, mientras que los dúplex de secuencia de referencia-referencia están sustancialmente no desnaturalizados. "Sustancialmente" significa al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 80%, incluso más preferiblemente al menos 90% y lo más preferible al menos 98%, en una forma desnaturalizada o no desnaturalizada dada. La etapa de reducir la temperatura de la mezcla de reacción permite que el par de cebadores se hibride a la secuencia diana. Estos cebadores hibridados se alargan entonces mediante una polimerasa, incrementando la cantidad de la secuencia diana en la muestra con respecto a la secuencia de referencia.

También se describe un método para enriquecer una secuencia diana sometiendo una mezcla de reacción de amplificación a múltiples ciclos de condiciones de hibridación y a la T<sub>c</sub>. La mezcla de reacción de amplificación, sospechosa de tener cada una de una secuencia diana y una secuencia de referencia, se somete primeramente a una primera temperatura de desnaturalización que está por encima de la T<sub>m</sub> de la secuencia de referencia. A continuación, la muestra se cicla entre dos etapas de incubación a temperatura diferente. En la primera etapa de incubación, la temperatura se disminuye para permitir la hibridación de la secuencia diana con la secuencia de referencia para formar un dúplex. En la segunda etapa de incubación, la temperatura se incrementa hasta la T<sub>c</sub>, que está por debajo de la T<sub>m</sub> de la secuencia de referencia. A la T<sub>c</sub>, los dúplex de secuencia diana-referencia (y los dúplex de secuencia diana-diana solamente si tienen una T<sub>m</sub> menor que la secuencia de referencia) están sustancialmente desnaturalizados, mientras que los dúplex de diana-diana (si tienen una T<sub>m</sub> igual o mayor que la T<sub>m</sub> de la secuencia de referencia) y los dúplex de secuencia de referencia-referencia están sustancialmente no

desnaturalizados. "Sustancialmente" significa al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 80%, incluso más preferiblemente al menos 90% y lo más preferible al menos 98%, en una forma desnaturalizada o no desnaturalizada dada. Estas etapas primera y segunda se repiten entonces una o más veces. Una vez que la etapa de incubación cíclica está terminada, la temperatura de la mezcla de reacción se disminuye para permitir que un par de cebadores se hibride a la secuencia diana. Estos cebadores se alargan entonces mediante una polimerasa, enriqueciendo así la secuencia diana con respecto a la secuencia de referencia en la muestra.

#### BREVE DESCRIPCIÓN DE LOS DIBUJOS

10 La Figura 1 ilustra una realización del procedimiento de enriquecimiento de la secuencia diana de la invención utilizando una primera temperatura de desnaturalización y la Tc.

La Figura 2 ilustra una realización del procedimiento de enriquecimiento de la secuencia diana de la invención utilizando una etapa oscilante de hibridación/Tc.

La Figura 3 ilustra las mutaciones de p53 y Kras ensayadas para enriquecimiento vía un protocolo de enriquecimiento de secuencias diana.

15 La Figura 4 ilustra el enriquecimiento de alelos mutantes de p53 con un protocolo de enriquecimiento de secuencias diana.

La Figura 5 ilustra el enriquecimiento de alelos mutantes de Kras con un protocolo de enriquecimiento de secuencias diana.

20 La Figura 6 ilustra el enriquecimiento de alelos mutantes en muestras clínicas procedentes de tumores de pulmón y de colon.

La Figura 7 ilustra el enriquecimiento de alelos mutantes mediante una realización del protocolo de enriquecimiento que carece de la etapa de formación de mal apareamientos.

25 La Figura 8A-8D ilustra las gráficas de amplificación para el exón 8 de p53 de tipo salvaje y mutante en PCR en tiempo real normal o vía un método de enriquecimiento de la invención aplicado en un formato en tiempo real. (a) muestra gráficas de amplificación en PCR en tiempo real normal para estirpes celulares diluidas en serie que contienen un mutante en la mutación del exón 8 de p53. (b) muestra gráficas de amplificación en un procedimiento de enriquecimiento de la invención según se aplica en un formato de PCR en tiempo real para estirpes celulares diluidas en serie que contienen un mutante en la mutación del exón 8 de p53. (c) muestra gráficas de amplificación en PCR en tiempo real normal de cuatro muestras tumorales clínicas, una de las cuales se sabe que contiene una mutación del exón 8 de p53 (CT20). (d) muestra gráficas de amplificación para un procedimiento de enriquecimiento de la invención según se aplica en un formato de PCR en tiempo real para las cuatro muestras tumorales clínicas.

30 La Figura 9A-D muestra gráficas de amplificación de PCR en tiempo real normal y enriquecida para el exón 8 de p53 de tipo salvaje y mutante utilizando un colorante de detección de ADN (LCGreen). (a) muestra gráficas de amplificación para PCR en tiempo real normal de muestras que contienen el exón 8 de p53 mutante (SW480, TL6, CT20) y de tipo salvaje (R27, TL8, TL18, TL81, TL82). (b) muestra gráficas de amplificación de un procedimiento de enriquecimiento de la invención según se aplica en un formato en tiempo real para muestras que contienen el exón 8 de p53 mutante (SW480, TL6, CT20) y de tipo salvaje (R27, TL8, TL18, TL81, TL82). (c) muestra gráficas de amplificación para PCR en tiempo real normal para muestras que contienen el exón 8 de p53 mutante (SW480, CT7, HCC, CT20) y de tipo salvaje. (d) muestra gráficas de amplificación para un procedimiento de enriquecimiento de la invención según se aplica en un formato en tiempo real para muestras que contienen el exón 8 de p53 mutante (SW480, CT7, HCC, CT20) y de tipo salvaje.

35 La Figura 10A-B ilustra el impacto de disolventes orgánicos (por ejemplo, 3% de DMSO) sobre el enriquecimiento de la diana. (a) protocolos de amplificación en PCR en tiempo real para el exón 8 de p53 de tipo salvaje y mutante. (b) gráficas de amplificación de un procedimiento de enriquecimiento de la invención según se aplica en un formato en tiempo real para el exón 8 de p53 mutante y de tipo salvaje.

40 La Figura 11 ilustra la detección mejorada de Polimorfismo de Longitud de Fragmento de Restricción (RFLP) usando un procedimiento de enriquecimiento de la invención tras la digestión con TaqI de la mutación de EGFR AA761. (a) ilustra la detección de EGFR de tipo salvaje y mutante (AA761) cuando se diluye 1:10.000 (mutante:genoma) tras PCR normal o tras un procedimiento de enriquecimiento. (b) ilustra la detección de EGFR de tipo salvaje según se evalúa mediante PCR normal.

#### DESCRIPCIÓN DETALLADA

La presente invención se refiere a métodos para enriquecer alelos poco abundantes (por ejemplo, secuencias diana) procedentes de una muestra. La presente invención se refiere, en parte, al enriquecimiento selectivo de secuencias diana aplicando una temperatura de desnaturalización crítica a una mezcla de reacción. La temperatura de desnaturalización crítica o  $T_c$  es una temperatura por debajo de la temperatura de fusión  $T_m$  de la secuencia de referencia. La temperatura crítica aprovecha la menor  $T_m$  del dúplex de ADN bicatenario hibridado de forma cruzada entre la diana y la referencia para desnaturalizar preferentemente estos dúplex con respecto a los homodúplex de referencia-referencia.

Muchos métodos de detección de mutaciones conocidos se postulan para beneficiarse de la presente invención que emplea un protocolo de amplificación que enriquece selectivamente la secuencia diana. El uso de PCR como la etapa inicial para el ensayo genético se usa en la mayoría de reacciones de mutaciones y de secuenciación. Generalmente, una secuencia de ácidos nucleicos (por ejemplo, ADN genómico/ADNc) se amplifica como una primera etapa, seguido de un método de didesoxisequenciación o de identificación de mutaciones (por ejemplo, SSCP, dHPLC, MALDI-TOF, pirosecuenciación, fusión de alta resolución. En consecuencia, los límites de esencialmente todos los métodos de detección de mutaciones se beneficiarían de forma uniforme si el enriquecimiento de secuencias que contienen mutaciones (por ejemplo, secuencias diana) se realizase durante la etapa de PCR que antecede a la identificación.

#### Definiciones

Como se usa aquí, la expresión "enriquecer una secuencia diana" se refiere a incrementar la cantidad de una secuencia diana e incrementar la relación de secuencia diana con respecto a la secuencia de referencia correspondiente en una muestra. Por ejemplo, cuando la relación de secuencia diana a secuencia de referencia es inicialmente 5% a 95% en una muestra, la secuencia diana se puede amplificar preferentemente en una reacción de amplificación para producir una relación de 70% de secuencia diana a 30% de secuencia de referencia. De este modo, se produce un enriquecimiento de 14 veces de la secuencia diana con respecto a la secuencia de referencia. El enriquecimiento de una secuencia diana da como resultado un incremento de 2X a 200X en la secuencia diana con respecto a la secuencia de referencia antes del enriquecimiento. El enriquecimiento de la diana es al menos un enriquecimiento de 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X, 15X, 20X, 25X, 30X, 35X, 40X, 45X, 50X, 60X, 70X, 80X, 90X 100X, 150X, 200X o más veces con respecto a la secuencia de referencia. El enriquecimiento de una secuencia diana da como resultado que una muestra tenga 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 80%, 90%, 95% o más de secuencia diana en comparación con la secuencia de referencia (por ejemplo, 10% de secuencia diana:90% de secuencia de referencia a 95% de secuencia diana:5% de secuencia de referencia).

Como se usa aquí, la expresión "secuencia diana" se refiere a un ácido nucleico que es menos predominante en una muestra de ácido nucleico que una secuencia de referencia correspondiente. La secuencia diana constituye menos del 50% de la cantidad total de secuencia de referencia + secuencia diana en una muestra. Preferiblemente, la secuencia diana se expresa a nivel del ARN y/o ADN 1:10, 1:15, 1:20, 1:25X, 1:30, 1:35, 1:40, 1:45, 1:50, 1:60, 1:70, 1:80, 1:90, 1:100, 1:150, 1:200X o menos que la secuencia de referencia. En una realización, la secuencia diana es un alelo mutante. Por ejemplo, una muestra (por ejemplo, muestra de sangre) puede contener numerosas células normales y unas pocas células cancerosas. Las células normales contienen alelos no mutantes o de tipo salvaje, mientras que el número pequeño de células cancerosas contiene mutaciones somáticas. En este caso, el mutante es la secuencia diana, mientras que la secuencia de tipo salvaje es la secuencia de referencia. En otra realización, la invención se refiere a la detección de ADN fetal en una muestra de ácido nucleico obtenida de una madre. En esta realización, la secuencia diana está presente en el ADN fetal mientras que el ADN de la madre más predominante contiene la secuencia de referencia. Como se usa aquí, una secuencia diana incluye ADN fetal obtenido de una madre gestante. Como se usa aquí, una "hebra diana" se refiere a una hebra de ácido nucleico individual de una secuencia diana.

La secuencia diana tiene una longitud de alrededor de 17-2000 nucleótidos. En una realización, la secuencia diana tiene una longitud de 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 o más nucleótidos. Las secuencias diana comparten al menos 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de homología con la secuencia de referencia correspondiente, pero difieren en al menos un nucleótido de la secuencia de referencia. Las secuencias diana se pueden amplificar vía PCR con el mismo par de cebadores que los usados para la secuencia de referencia.

Como se usa aquí, la expresión "secuencia de referencia" se refiere a un ácido nucleico que es más predominante en una muestra de ácido nucleico que una secuencia diana correspondiente (por ejemplo, mismo gen pero diferente secuencia de ácidos nucleicos). La secuencia de referencia constituye alrededor del 50% del total de secuencia de referencia + secuencia diana en una muestra. Preferiblemente, la secuencia de referencia se expresa a nivel del ARN y/o ADN 10X, 15X, 20X, 25X, 30X, 35X, 40X, 45X, 50X, 60X, 70X, 80X, 90X 100X, 150X, 200X o más que la secuencia diana. En una realización, la secuencia de referencia es un alelo de tipo salvaje. Por ejemplo, una muestra (por ejemplo, muestra de sangre) puede contener numerosas células normales y unas pocas células cancerosas. Las células normales contienen alelos no mutantes o de tipo salvaje, mientras que el número pequeño de células cancerosas contiene mutaciones somáticas. En este caso, la secuencia de tipo salvaje es la secuencia de

referencia, mientras que la secuencia mutante es la secuencia diana (por ejemplo, alelo mutante). Como se usa aquí, una "hebra de referencia" se refiere a una hebra de ácido nucleico individual de una secuencia de referencia.

La secuencia de referencia tiene una longitud de alrededor de 17-2000 nucleótidos. En una realización, la secuencia de referencia tiene una longitud de 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 o más nucleótidos. Las secuencias de referencia compartirán al menos 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de homología con la secuencia diana correspondiente, pero diferirán en al menos un nucleótido de la secuencia diana. Las secuencias de referencia se pueden amplificar mediante PCR con el mismo par de cebadores que el usado para la secuencia de referencia.

El término "alelo" se refiere a formas alternativas de un gen, porción del mismo o región no codificante de ADN que ocupa el mismo locus o posición en cromosomas homólogos que tienen al menos una diferencia en la secuencia nucleotídica. El término alelo se puede usar para describir ADN de cualquier organismo, incluyendo, pero sin limitarse a, bacterias, virus, hongos, protozoos, mohos, levaduras, plantas, seres humanos, seres no humanos, animales, y arqueobacterias. Los alelos se pueden encontrar en una célula individual (por ejemplo, dos alelos, uno heredado del padre y uno de la madre) o en una población de células (por ejemplo, un alelo de tipo salvaje procedente de tejido normal y un alelo mutante somático procedente de tejido enfermo).

Un alelo puede tener una longitud de 17-2000 nucleótidos. En una realización, el alelo tiene una longitud de 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 o más nucleótidos. Los alelos compartirán generalmente 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de homología entre sí. Los alelos se pueden amplificar mediante PCR con el mismo par de cebadores.

En una realización, la presente invención se usa para enriquecer un polimorfismo. Cualquier gen dado puede no tener ninguna, puede tener una, o puede tener muchas formas alélicas (polimorfismo). Los cambios mutacionales habituales que dan lugar a alelos pueden ser el resultado de supresiones, adiciones o sustituciones naturales o artificiales (por ejemplo, carcinógenos químicos) de nucleótidos. Cada uno de estos tipos de cambios puede ocurrir solo, o en combinación con los otros, una o más veces en una secuencia diana.

Se han dado a conocer varios tipos diferentes de polimorfismos. Un polimorfismo de longitud de fragmento de restricción (RFLP) es una variación en la secuencia de ADN que altera la longitud de un fragmento de restricción (Botstein et al., *Am. J. Hum. Genet.* 32:314-331 (1980)). El polimorfismo de longitud de fragmento de restricción puede crear o suprimir un sitio de restricción, cambiando así la longitud del fragmento de restricción. Los RFLPs se han usado ampliamente en análisis genéticos de seres humanos y de animales (véanse el documento WO 90/13668; documento WO 90/11369; Donis-Keller, *Cell* 51:319-337 (1987); Lander et al., *Genetics* 121:85-99 (1989)). Cuando un rasgo heredable puede estar relacionado con un RFLP particular, la presencia del RFLP en un individuo se puede usar para predecir la probabilidad de que el individuo también mostrará el rasgo. El RFLP se puede usar en combinación con los métodos de enriquecimiento descritos aquí para potenciar la detección de polimorfismos. Tales métodos se describen aquí en el Ejemplo 9.

Otros polimorfismos toman la forma de repeticiones en tándem cortas (STRs) que incluyen motivos repetidos di-, tri- y tetranucleotídicos en tándem. Estas repeticiones en tándem también se denominan como polimorfismos de repeticiones en tándem de número variable (VNTR). Las VNTRs se han usado para en análisis de identidad y paternidad (patente U.S. nº 5.075.217; Armour et al., *FEBS Lett.* 307:113-115 (1992); Horn et al., documento WO 91/14003; Jeffreys, documento EP 370.719) y en un gran número de estudios de cartografiado genético.

Otros polimorfismos tienen la forma de variaciones de un solo nucleótido entre individuos de la misma especie. Tales polimorfismos son mucho más frecuentes que RFLPs, STRs (repeticiones en tándem cortas) y VNTRs (repeticiones en tándem de número variable). Algunos polimorfismos de un solo nucleótido se producen en secuencias que codifican proteínas, en cuyo caso, una de las formas polimórficas puede dar lugar a la expresión de una proteína defectuosa u otra proteína variante y, potencialmente a una enfermedad genética. Otros polimorfismos de un solo nucleótido se producen en regiones no codificantes. Algunos de estos polimorfismos también pueden dar como resultado la expresión de proteínas defectuosas (por ejemplo, como resultado del ajuste defectuoso). Otros polimorfismos de un solo nucleótido no tienen efectos fenotípicos.

Aún otras mutaciones incluyen mutaciones somáticas. Las mutaciones somáticas son alteraciones en el ADN que se producen después de la concepción. Las mutaciones somáticas se pueden producir en cualquiera de las células del cuerpo, excepto en las células germinales (esperma y óvulo), y por lo tanto no se transmiten al niño. Estas alteraciones pueden conducir (pero no siempre) a cáncer u otras enfermedades.

Como se usa aquí, la expresión "tipo salvaje" se refiere a la secuencia polinucleotídica o alelo más habitual para un cierto gen en una población. Generalmente, el alelo de tipo salvaje se obtendrá de células normales.

Como se usa aquí, el término "mutante" se refiere a un cambio nucleotídico (es decir, una sustitución, supresión, o inserción de un solo nucleótido o de múltiples nucleótidos) en una secuencia de ácidos nucleicos. Un ácido nucleico que posee mutación tiene una secuencia de ácido nucleico (alelo mutante) que es diferente en secuencia de la de la

secuencia polinucleotídica de tipo salvaje correspondiente. Los métodos de la invención son especialmente útiles enriqueciendo selectivamente un alelo mutante que contiene entre 1 y 500 cambios de secuencia nucleotídica. Un alelo mutante puede tener 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400 o 500 cambios de secuencia nucleotídica en comparación con un alelo de tipo salvaje correspondiente. Preferiblemente, una mutación en un alelo mutante contendrá entre 1 y 10 cambios de secuencia nucleotídica, y más preferiblemente entre 1 y 5 cambios de secuencia nucleotídica. El alelo mutante tendrá 50%, 60%, 70%, 80%, 85%, 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más de homología con el alelo de tipo salvaje. Generalmente, el alelo mutante se obtendrá a partir de tejidos o células enfermos, y está asociado con un estado mórbido.

Como se usa aquí, la expresión "temperatura de fusión" o " $T_m$ " se refiere a la temperatura a la que un polinucleótido se disocia de su secuencia complementaria. Generalmente, la  $T_m$  se puede definir como la temperatura a la que la mitad de los pares de bases de Watson-Crick en una molécula de ácido nucleico dúplex se rompe o disocia (es decir, se "funden"), mientras que la otra mitad de los pares de bases de Watson-Crick permanecen intactos en una conformación bicatenaria. En otras palabras, la  $T_m$  se define como la temperatura a la que el 50% de los nucleótidos de dos secuencias complementarias están hibridados (hebras dobles) y el 50% de los nucleótidos están desnaturalizados (hebras sencillas). Por lo tanto, la  $T_m$  define un punto medio en la transición desde moléculas de ácido nucleico bicatenarias a monocatenarias (o, al contrario, en la transición de moléculas de ácido nucleico monocatenarias a bicatenarias).

La  $T_m$  se puede estimar mediante un número de métodos, por ejemplo mediante un cálculo del vecino más cercano de acuerdo con Wetmur 1991 (Wetmur, J. G. 1991. DNA probes: applications of the principles of nucleic acid hybridization. Crit Rev Biochem Mol Biol 26: 227-259, incorporada aquí como referencia) y mediante programas comerciales que incluyen Oligo™ Primer Design y programas disponibles en internet. Como alternativa, la  $T_m$  se puede determinar mediante experimentación real. Por ejemplo, los colorantes de unión a ADN bicatenario o intercalantes, tales como bromuro de etidio o SYBR-green (Molecular Probes) se pueden usar en un ensayo de curva de fusión para determinar la  $T_m$  real del ácido nucleico. Métodos adicionales para determinar la  $T_m$  de un ácido nucleico son bien conocidos en la técnica y se describen aquí.

Como se usa aquí, la expresión "temperatura crítica" o " $T_c$ " se refiere a una temperatura por debajo de la  $T_m$  de la secuencia de referencia. La  $T_c$  se aplica a dúplex bicatenario de secuencia diana/secuencia de referencia preferiblemente desnaturalizado, para permitir el enriquecimiento selectivo de la secuencia diana durante una reacción de amplificación. La temperatura de desnaturalización crítica ( $T_c$ ) es la temperatura por debajo de la cual la eficiencia de la PCR cae abruptamente para una secuencia de ácido nucleico dada. Por ejemplo, una secuencia de p53 de 167 pb se amplifica bien si la temperatura de desnaturalización de la PCR se fija a 87°C, se amplifica de forma modesta a 86,5°C, y no produce producto detectable si la desnaturalización de la PCR se fija a 86°C o menos. Por lo tanto, en este ejemplo,  $T_c \sim 86,5^\circ\text{C}$ . La  $T_c$  está alrededor de 0,1-20°C por debajo de la  $T_m$  de la secuencia de referencia. Más preferiblemente, la  $T_c$  está alrededor de 0,1-10°C, 0,1-9°C, 0,1-8°C, 0,1-7°C, 0,1-6°C, 0,2°C-5°C, 0,3°C-4,5°C, 0,4-4°C, 0,5-3,5°C, 0,5-3°C, 0,5-3°C, 0,5-2,5°C, 0,5-2°C, 0,5-1,5°C, 0,5-1°C por debajo de la  $T_m$  de la secuencia de referencia. En algunas realizaciones, la  $T_c$  está por debajo de la  $T_m$  tanto de la secuencia de referencia como de la secuencia diana. Por ejemplo, la  $T_c$  puede estar alrededor de 0,1-10°C, 0,1-9°C, 0,1-8°C, 0,1-7°C, 0,1-6°C, 0,2°C-5°C, 0,3°C-4,5°C, 0,4-4°C, 0,5-3,5°C, 0,5-3°C, 0,5-3°C, 0,5-2,5°C, 0,5-2°C, 0,5-1,5°C, 0,5-1°C por debajo de la  $T_m$  de la secuencia diana.

Como se usa aquí, la expresión "desnaturalización selectiva" o "desnaturalización preferente" se refiere a la ruptura preferente de enlaces de hidrógeno entre pares de bases en una molécula de ácido nucleico bicatenaria de una secuencia diana o de un dúplex de secuencia diana/de referencia, para producir una secuencia diana monocatenaria. La desnaturalización selectiva de la secuencia diana se logra aplicando la temperatura crítica a la muestra que contiene las secuencias diana y de referencia.

Como se usa aquí, "par de cebadores" se refiere a dos cebadores que se hibridan a hebras opuestas de una secuencia diana y de referencia para formar un producto de amplificación durante una reacción de PCR. El par de cebadores se diseña para que tenga una  $T_m$  menor que la  $T_c$  de la reacción.

Como se usa aquí, la expresión "hibridación cruzada" se refiere a un dúplex bicatenario formado entre dos secuencias de ácidos nucleicos, que difieren en uno o más nucleótidos, en virtud de la formación de uniones de hidrógeno entre bases G y C complementarias y entre bases A y T o A y U complementarias. Las dos secuencias de ácidos nucleicos complementarias se enlazan mediante hidrógeno en una configuración antiparalela. El ADN hibridado de forma cruzada contendrá mal apareamientos en las posiciones de diferencia entre las secuencias de referencia y diana. Los mal apareamientos pueden incluir polimorfismos, mutaciones, inserciones, supresiones y otros cambios que dan como resultado tales diferencias. Por ejemplo, metilación. Por ejemplo, en los sitios de supresiones/inserciones se producen bucles y/o regiones monocatenarias de uno o más nucleótidos.

La hibridación cruzada implica típicamente desnaturalizar las secuencias diana y de referencia, por ejemplo mediante calentamiento, seguido de la renaturalización en condiciones (tal como temperatura) que permitan que se produzca la hibridación y la formación del dúplex.

Como se usa aquí, una “mezcla de reacción” es una mezcla que se sospecha que contiene un dúplex de secuencia diana que comprende un tampón adecuado para permitir la desnaturalización de una secuencia diana.

Como se usa aquí, “identidad” u “homología” se refiere a la similitud de secuencia subunidad entre dos moléculas poliméricas, por ejemplo dos polinucleótidos o dos polipéptidos. Cuando una posición subunitaria en ambas moléculas está ocupada por la misma subunidad monomérica, por ejemplo si una posición en cada uno de dos péptidos está ocupada por serina, entonces son idénticas en esa posición. La identidad entre dos secuencias es una función directa del número de posiciones que se aparean o idénticas, por ejemplo, si la mitad (por ejemplo, 5 posiciones en un polímero de 10 subunidades de longitud) de las posiciones en dos secuencias peptídicas o compuestas son idénticas, entonces las dos secuencias son idénticas al 50%; si el 90% de las posiciones, por ejemplo 9 de 10, están apareadas, las dos secuencias comparten una identidad de secuencia del 90%.

El porcentaje de identidad nucleotídica se puede determinar mediante los parámetros por defecto de BLAST. Para la comparación de secuencias, típicamente una secuencia actúa como una secuencia de referencia, con la que se comparan secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de ensayo y de referencia se introducen en un ordenador, se designan coordenadas de subsecuencia, si es necesario, y se designan parámetros del programa de algoritmo de secuencias. Se pueden usar parámetros del programa por defecto, o se pueden designar parámetros alternativos. El algoritmo de comparación de secuencias calcula entonces el porcentaje de identidades de secuencia para las secuencias de ensayo con respecto a la secuencia de referencia, basado en los parámetros del programa.

La ventana de comparación incluye la referencia a un segmento de una cualquiera del número de posiciones contiguas seleccionadas del grupo que consiste en 20 a 600, habitualmente alrededor de 50 a alrededor de 200, más habitualmente alrededor de 100 a alrededor de 150 en el que una secuencia se puede comparar con una secuencia de referencia del mismo número de posiciones contiguas después de que las dos secuencias se alinean de forma óptima. Los métodos de alineamiento de secuencias para comparación son bien conocidos en la técnica. El alineamiento óptimo de secuencias para comparación se puede realizar, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, *Adv. Appl. Math.* 2:482 (1981), mediante el algoritmo de alineamiento de homología de Needleman y Wunsch, *J. Mol. Biol.* 48:443 (1970), mediante el método de búsqueda de similitud de Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), mediante implementaciones computerizadas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA en el paquete de Software de Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, Wis.), o mediante un alineamiento manual e inspección visual (véase, por ejemplo, *Current Protocols in Molecular Biology* (Ausubel et al., eds. 1995 suplemento)).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul et al., *Nuc. Acids Res.* 25:3389-3402 (1977) y Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), respectivamente. El software para llevar a cabo los análisis mediante BLAST está disponible al público a través del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica identificar primeramente pares de secuencias de alta puntuación (HSPs) identificando palabras cortas de longitud  $W$  en la secuencia de búsqueda, ya sea que se apareen o satisfagan cierta puntuación umbral  $T$  de valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de una base de datos.  $T$  se denomina como el umbral de puntuación de palabra vecina (Altschul et al., más arriba). Estos resultados de palabra vecina iniciales actúan como semillas para iniciar las búsquedas para encontrar HSPs más largos que los contienen. Los resultados de la palabra se extienden en ambas direcciones a lo largo de cada secuencia en la medida en que se puede incrementar la puntuación de alineamiento acumulativa. Las puntuaciones acumulativas se calculan usando, para secuencias nucleotídicas, los parámetros  $M$  (puntuación de recompensa para un par de restos que se aparean; siempre  $>0$ ) y  $N$  (puntuación de penalización para restos mal apareados; siempre  $<0$ ). Para secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los resultados de la palabra en cada dirección se detiene cuando: la puntuación de alineamiento acumulativa cae en la cantidad de  $X$  desde su valor alcanzado máximo; la puntuación acumulativa llega a cero o por debajo, debido a la acumulación de uno o más alineamientos de restos que puntúan negativamente; o se alcanza el extremo de cualquiera de las secuencias. Los parámetros  $W$ ,  $T$  y  $X$  del algoritmo BLAST determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencia nucleotídica) usa como defectos una longitud de palabra ( $W$ ) de 11, un valor esperado ( $E$ ) de 10,  $M = 5$ ,  $N = -4$ , y una comparación de ambas hebras. Para secuencias de aminoácidos, el programa BLASTP usa como defectos una longitud de palabra de 3, un valor esperado ( $E$ ) de 10, y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)), alineamientos ( $B$ ) de 50, expectativa ( $E$ ) de 10,  $M = 5$ ,  $N = -4$ , y una comparación de ambas hebras.

El algoritmo BLAST también lleva a cabo un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, *Proc. Nat'l Acad. Sci. USA* 90:58735787 (1993)). Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña ( $P(N)$ ), que proporciona una indicación de la probabilidad mediante la cual se producirá por casualidad un apareamiento entre dos secuencias nucleotídicas o de aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es menor que alrededor de 0,2, más preferiblemente menor que alrededor de 0,01, y lo más preferible menor que alrededor de 0,001.

En un primer aspecto, la invención se refiere a un método para enriquecer una secuencia diana en una muestra de ácido nucleico sospechosa de tener secuencias diana y de referencia. Las secuencias de referencia y diana se pueden amplificar antes de usarlas en el presente método. Esto es, las secuencias de referencia y diana de interés se pueden amplificar a partir de un molde genómico en una reacción de PCR antes de usarlas en el presente método. Una alícuota procedente de esta reacción de PCR se transfiere entonces para uso en el método de enriquecimiento selectivo. Como alternativa, las secuencias de referencia y diana no necesitan ser sometidas a una primera reacción de PCR sino que se pueden usar en su forma nativa (por ejemplo, ADN genómico) en el método de enriquecimiento selectivo. Las secuencias diana y de referencia se pueden obtener a partir de cualquier secuencia de ácido nucleico, incluyendo ADN genómico, ADNc, ADN vírico, ADN de mamífero, ADN fetal, o ADN bacteriano. Mientras que la secuencia de referencia es generalmente el alelo de tipo salvaje y la secuencia diana es el alelo mutante, lo contrario también puede ser cierto. El alelo mutante puede incluir una cualquiera o más supresiones, inserciones o alteraciones nucleotídicas. En algunas realizaciones, el alelo mutante es una mutación somática. En otras realizaciones, la secuencia diana es ADN metilado, mientras que la secuencia de referencia es ADN no metilado. Como alternativa, la secuencia diana es ADN no metilado, mientras que la secuencia de referencia es ADN metilado. Los cebadores usados en el presente método se diseñan generalmente para producir productos de amplificación de las secuencias de referencia y diana de alrededor de 17 a 1000 bases, más preferiblemente alrededor de 25 a 500 bases, y lo más preferible alrededor de 50 a 100 bases de tamaño.

El método incluye someter la mezcla de reacción de amplificación a una primera temperatura de desnaturalización que está por encima de la temperatura de fusión "T<sub>m</sub>" de una secuencia de referencia. La T<sub>m</sub> de un ácido nucleico se puede determinar mediante experimentación, o se puede estimar por cálculo. El experto estará bien al tanto de numerosos métodos bien conocidos para determinar la T<sub>m</sub> de un ácido nucleico, algunos de los cuales se describen aquí. La primera temperatura de desnaturalización se fija según procedimientos estándar usados en PCR. De este modo, la primera temperatura de desnaturalización debería de ser suficientemente alta para permitir la desnaturalización completa de las secuencias diana y de referencia (por ejemplo, 96°C). En una realización, la primera temperatura de desnaturalización está alrededor de 1°C a 30°C por encima de la T<sub>m</sub> de la secuencia de referencia, más preferiblemente la T<sub>m</sub> de la secuencia de referencia está alrededor de 5°C a 20°C por encima de la T<sub>m</sub> de la secuencia de referencia.

A continuación, la temperatura de la mezcla de reacción de amplificación se disminuye, permitiendo que las secuencias diana y secuencias de referencia se hibriden.

Esta temperatura de hibridación o temperatura intermedia (estando la temperatura por debajo de la primera temperatura de desnaturalización y T<sub>c</sub> pero por encima de la temperatura de hibridación/alargamiento del cebador, por ejemplo alrededor de 60°C a 80°C) está por encima de la T<sub>m</sub> del par de cebadores, y de este modo permite que las secuencias diana y de referencia se hibriden mientras evita la unión del par de cebadores a las secuencias diana y/o de referencia. Esta etapa de hibridación da como resultado la formación de dúplex de hibridación de secuencias diana-diana, referencia-referencia y diana-referencia bicatenarias. Los dúplex de hibridación de diana-referencia se desnaturalizan entonces preferentemente incrementando la temperatura de la mezcla de reacción hasta la T<sub>c</sub>. La T<sub>c</sub> o temperatura crítica está por debajo de la T<sub>m</sub> de la secuencia de referencia, y se puede determinar mediante los métodos descritos aquí. En una realización, la T<sub>c</sub> está por debajo de alrededor de 0,3°C-5°C, y más preferiblemente por debajo de alrededor de 0,5°C a 1,5°C de la T<sub>m</sub> de la secuencia de referencia. Generalmente, la T<sub>c</sub> será alrededor de 70-90°C. Los dúplex de hibridación de diana-diana también se pueden desnaturalizar preferentemente si la secuencia diana tiene una secuencia nucleotídica que da como resultado una menor T<sub>m</sub> en comparación con la secuencia de referencia. A la T<sub>c</sub>, los dúplex de secuencia diana-referencia (y los dúplex de secuencia diana-diana solamente si tienen una menor T<sub>m</sub> que la secuencia de referencia) están sustancialmente desnaturalizados, mientras que los dúplex de diana-diana (si tienen una T<sub>m</sub> igual o mayor que la T<sub>m</sub> de la secuencia de referencia) y los dúplex de secuencia de referencia-referencia están sustancialmente no desnaturalizados. "Sustancialmente" significa al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 80%, incluso más preferiblemente al menos 90%, y lo más preferible al menos 98% en una forma desnaturalizada o no desnaturalizada. La T<sub>c</sub> se aplica generalmente de alrededor de 1 segundo a 5 minutos, más preferiblemente 2 segundos a 1 minuto, y lo más preferible 5 segundos a 30 segundos.

Tras la desnaturalización preferente de los dúplex de hibridación de secuencias diana-referencia y/o diana-diana, la temperatura de la mezcla de reacción se reduce para permitir que un par de cebadores se hibride a la secuencia diana. Los cebadores hibridados se alargan entonces mediante una polimerasa de ácido nucleico, enriqueciendo así la secuencia diana con respecto a la secuencia de referencia en la muestra.

Las etapas del método se repiten generalmente durante múltiples ciclos a fin de obtener una amplificación suficiente de las secuencias diana y de referencia. En una realización, las etapas del método se repiten durante 5-40 ciclos, y más preferiblemente 10-30 ciclos. El número óptimo de ciclos se puede determinar por una persona experta normal en la técnica. Preferiblemente, los presentes métodos se llevan a cabo en un dispositivo de PCR, más preferiblemente en condiciones de reacción en tiempo real en un dispositivo de PCR de detección en tiempo real, tal como el dispositivo de PCR en tiempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) y el dispositivo de PCR en tiempo real Mx3005P (Stratagene, La Jolla, CA). En esta realización, la mezcla de reacción puede incluir un agente de detección de ácido nucleico (por ejemplo, un colorante de detección de ácido nucleico tal como el colorante SYBR Green o el colorante LC-Green, o una sonda acoplada de forma operativa a un colorante fluorescente) para

cuantificar y/o monitorizar los productos de amplificación de la reacción. Una vez que el enriquecimiento de la secuencia diana está terminado, la muestra se puede procesar adicionalmente (por ejemplo, para identificación de cualesquiera alteraciones genéticas enriquecidas por el método, por ejemplo sometidas a una reacción de secuenciación). Las secuencias de referencia enriquecidas se pueden procesar adicionalmente mediante una variedad de procedimientos, incluyendo: MALDI-TOF, fusión de HR, secuenciación didesoxi, secuenciación de una sola molécula, pirosecuenciación, RFLP, PCR digital y PCR cuantitativa.

Se describe un método para enriquecer una secuencia diana llevando a cabo un protocolo de reacción de amplificación de ácido nucleico. El protocolo de reacción de amplificación incluye una primera temperatura de desnaturalización y una segunda temperatura de desnaturalización. La primera temperatura de desnaturalización está por encima de la  $T_m$  de la secuencia de referencia, y la segunda temperatura de desnaturalización está por debajo de la  $T_m$  de la secuencia de referencia. El método incluye someter la mezcla de reacción de amplificación a una primera temperatura de desnaturalización que está por encima de la temperatura de fusión " $T_m$ " de una secuencia de referencia. La  $T_m$  de un ácido nucleico se puede determinar mediante experimentación, o se puede estimar mediante cálculo. El experto estará bien al tanto de numerosos métodos bien conocidos para determinar la  $T_m$  de un ácido nucleico. La primera temperatura de desnaturalización se selecciona generalmente como aquella que se seleccionaría generalmente como la temperatura de desnaturalización de una reacción de PCR, y debería de ser suficientemente elevada para permitir la desnaturalización de las secuencias diana y de referencia. En una realización, la primera temperatura de desnaturalización está  $1^\circ\text{C}$  a  $30^\circ\text{C}$  por encima de la  $T_m$  de la secuencia de referencia, más preferiblemente la  $T_m$  de la secuencia de referencia está alrededor de  $5^\circ\text{C}$  a  $20^\circ\text{C}$  por encima de la  $T_m$  de la secuencia de referencia.

La segunda temperatura de desnaturalización está por debajo de la  $T_m$  de la secuencia de referencia, y se puede determinar mediante los métodos descritos aquí. En una realización, la  $T_c$  está alrededor de  $0,3^\circ\text{C}$ - $5^\circ\text{C}$  por debajo de la  $T_m$  de la secuencia de referencia, y más preferiblemente alrededor de  $0,5^\circ\text{C}$  a  $1,5^\circ\text{C}$  por debajo de la  $T_m$  de la secuencia de referencia. Generalmente, la  $T_c$  será alrededor de  $70$ - $90^\circ\text{C}$ . La segunda temperatura de desnaturalización se aplica generalmente de alrededor de 1 segundo a 5 minutos, más preferiblemente 2 segundos a 1 minuto, y lo más preferible 5 segundos a 30 segundos.

Se describe además un método para enriquecer una secuencia diana sometiendo una mezcla de reacción de amplificación a una  $T_c$ , reduciendo la temperatura de la mezcla de reacción y alargando un par de cebadores. La mezcla de reacción de amplificación es sospechosa de contener una secuencia diana y una secuencia de referencia. En este aspecto, la secuencia diana tiene una  $T_m$  por debajo de la  $T_m$  de la secuencia de referencia. La  $T_c$  está por debajo de la  $T_m$  de la secuencia de referencia, permitiendo así la desnaturalización preferente de la secuencia diana que tiene una menor  $T_m$  que la secuencia de referencia como resultado de su composición nucleotídica (por ejemplo, supresión). Con respecto a los otros aspectos de la invención, las secuencias de referencia y diana se pueden amplificar antes de usarlas en el presente método. Esto es, las secuencias de referencia y diana de interés se pueden amplificar a partir de un molde genómico en una reacción de PCR antes de usarlas en el presente método. Entonces se transfiere una alícuota de esta reacción de PCR para uso en el método de enriquecimiento selectivo. Como alternativa, no es necesario someter a las secuencias de referencia y diana a una primera reacción de PCR, sino que se pueden usar en su forma nativa en el método de enriquecimiento selectivo, por ejemplo, ADN genómico. Las secuencias diana y de referencia se pueden obtener a partir de cualquier secuencia de ácido nucleico, incluyendo ADN genómico, ADNc, ADN vírico, ADN de mamífero, ADN fetal, o ADN bacteriano. Mientras que la secuencia de referencia es generalmente el alelo de tipo salvaje y la secuencia diana es el alelo mutante, lo contrario también puede ser cierto. El alelo mutante puede incluir una cualquiera o más supresiones, inserciones o alteraciones nucleotídicas. En algunas realizaciones, el alelo mutante es una mutación somática. Los cebadores usados en el presente método se diseñan generalmente para producir productos de amplificación de secuencias de referencia y diana de alrededor de 15 a 1000 bases, más preferiblemente alrededor de 25 a 500 bases, y lo más preferible alrededor de 50 a 100 bases de tamaño.

Los dúplex de hibridación de diana-diana se desnaturalizan preferentemente incrementando la temperatura de la mezcla de reacción hasta la  $T_c$ . La  $T_c$  o temperatura crítica está por debajo de la  $T_m$  de la secuencia de referencia, y se puede determinar mediante los métodos descritos aquí. En una realización, la  $T_c$  está alrededor de  $0,3^\circ\text{C}$ - $5^\circ\text{C}$  por debajo de la  $T_m$  de la secuencia de referencia, y más preferiblemente alrededor de  $0,5^\circ\text{C}$  a  $1,5^\circ\text{C}$  por debajo de de la  $T_m$  de la secuencia de referencia. Generalmente, la  $T_c$  será alrededor de  $70$ - $90$ . La  $T_c$  se aplica generalmente de alrededor de 1 segundo a 5 minutos, más preferiblemente 2 segundos a 1 minuto, y lo más preferible 5 segundos a 30 segundos. A la  $T_c$ , los dúplex de secuencia diana-referencia y los dúplex de secuencia diana-diana están sustancialmente desnaturalizados, mientras que los dúplex secuencia de referencia-referencia están sustancialmente no desnaturalizados. "Sustancialmente" significa al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 80%, incluso más preferiblemente al menos 90%, y lo más preferible al menos 98% en una forma desnaturalizada o no desnaturalizada.

La etapa de reducir la temperatura de la mezcla de reacción permite que el par de cebadores se hibride a la secuencia diana. Estos cebadores hibridados se alargan entonces mediante una polimerasa, incrementando la cantidad de la secuencia diana en la muestra. Las etapas del método se repiten generalmente durante múltiples ciclos a fin de obtener una amplificación suficiente de las secuencias diana y de referencia. En una realización, las etapas del método se repiten durante 5-40 ciclos, y más preferiblemente 10-30 ciclos. El número óptimo de ciclos se

puede determinar por un experto normal en la técnica. Preferiblemente, los presentes métodos se llevan a cabo en un dispositivo de PCR, más preferiblemente en condiciones de reacción en tiempo real en un dispositivo de PCR de detección en tiempo real, tal como el dispositivo de PCR en tiempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) y el dispositivo de PCR en tiempo real Mx3005P (Stratagene, La Jolla, CA). En esta realización, la mezcla de reacción puede incluir un agente de detección de ácido nucleico (por ejemplo, colorante de detección de ácido nucleico tal como el colorante SYBR Green o el colorante LC-Green, o una sonda acoplada de forma operativa a un colorante fluorescente) para cuantificar y/o monitorizar los productos de amplificación de la reacción. Una vez que el enriquecimiento de la secuencia diana está terminado, la muestra se puede procesar adicionalmente, por ejemplo se puede someter a una reacción de secuenciación. Los alelos enriquecidos se pueden procesar adicionalmente mediante una variedad de procedimientos, incluyendo: MALDI-TOF, fusión de HR, secuenciación didesoxi, secuenciación de una sola molécula, pirosecuenciación, RFLP, PCR digital y PCR cuantitativa.

Todavía se describe además un método para enriquecer una secuencia diana sometiendo una mezcla de reacción de amplificación a etapas alternas de condiciones de hibridación y condiciones de desnaturalización aplicando la  $T_c$ . La mezcla de reacción de amplificación, que tiene una secuencia diana y una secuencia de referencia, se somete en primer lugar a una primera temperatura de desnaturalización que está por encima de la  $T_m$  de la secuencia de referencia. Como en los otros aspectos, las secuencia de referencia y diana se pueden amplificar antes de usarlas en el presente método. Esto es, las secuencias de referencia y diana de interés se pueden amplificar a partir de un molde genómico en una reacción de PCR antes de usarlas en el presente método. Entonces, una alícuota de esta reacción de PCR se transfiere para uso en el método de enriquecimiento selectivo. Como alternativa, no es necesario someter a las secuencias de referencia y diana a una primera reacción de PCR, sino que se pueden usar en su forma nativa en el método de enriquecimiento selectivo, por ejemplo ADN genómico. Las secuencias diana y de referencia se pueden obtener a partir de cualquier secuencia de ácido nucleico, incluyendo ADN genómico, ADNc, ADN vírico, ADN de mamífero, ADN fetal, o ADN bacteriano. Mientras que la secuencia de referencia es generalmente el alelo y la secuencia diana es el alelo mutante, lo contrario también puede ser cierto. El alelo mutante puede incluir una cualquiera o más supresiones, inserciones o alteraciones nucleotídicas. En algunas realizaciones, el alelo mutante es una mutación somática. En otras realizaciones, la secuencia diana es ADN metilado, mientras que la secuencia de referencia es ADN no metilado. Como alternativa, la secuencia diana es ADN no metilado, mientras que la secuencia de referencia es ADN metilado. Los cebadores usados en la presente invención se diseñan generalmente para producir productos de amplificación de secuencia de referencia y diana de alrededor de 15 a 1000 bases, más preferiblemente alrededor de 25 a 500 bases, y lo más preferible alrededor de 50 a 100 bases de tamaño.

La  $T_m$  de un ácido nucleico se puede determinar mediante experimentación, o se puede estimar mediante cálculo. El experto estará bien al tanto de numerosos métodos bien conocidos para determinar la  $T_m$  de un ácido nucleico. La primera temperatura de desnaturalización se selecciona generalmente como aquella que se seleccionaría generalmente como la temperatura de desnaturalización de una reacción de PCR, y debería de ser suficientemente elevada para permitir la desnaturalización de las secuencias diana y de referencia. En una realización, la primera temperatura de desnaturalización está alrededor de 1°C a 30°C por encima de la  $T_m$  de la secuencia de referencia, más preferiblemente la  $T_m$  de la secuencia de referencia está alrededor de 5°C a 20°C por encima de la  $T_m$  de la secuencia de referencia.

A continuación, la muestra se cicla entre dos etapas de incubación de temperaturas diferentes. En la primera etapa de incubación, la temperatura se disminuye para permitir la hibridación de la secuencia diana con la secuencia de referencia. En la segunda etapa de incubación, la temperatura se incrementa hasta la  $T_c$ , que está por debajo de la  $T_m$  de la secuencia de referencia. Estas etapas primera y segunda se repiten entonces una o más veces, más preferiblemente 3-20 veces, y lo más preferible 5-10 veces.

La primera etapa de incubación da como resultado la formación de dúplex de hibridación de secuencias diana-diana, referencia-referencia y diana-referencia. En una realización preferida, esta temperatura de hibridación o temperatura intermedia (estando la temperatura por debajo de la primera temperatura de desnaturalización y  $T_c$  pero por encima de la temperatura de hibridación/alargamiento de cebadores, por ejemplo alrededor de 60°C a 80°C) está por encima de la  $T_m$  del par de cebadores, y de este modo permite que las secuencias diana y de referencia se hibriden mientras evita la unión del par de cebadores a las secuencias diana y/o de referencia. Los dúplex de hibridación de diana-referencia y de diana-diana (en tanto que la diana tenga una menor  $T_m$  que la secuencia de referencia) se desnaturalizan entonces preferentemente incrementando la temperatura de la mezcla de reacción hasta la  $T_c$  en la segunda etapa de incubación. La  $T_c$  o temperatura crítica está por debajo de la  $T_m$  de la secuencia de referencia, y se puede determinar mediante los métodos descritos aquí. En una realización, la  $T_c$  está alrededor de 0,3°C-5°C por debajo de la  $T_m$  de la secuencia de referencia, y más preferiblemente alrededor de 0,5°C a 1,5°C por debajo de la  $T_m$  de la secuencia de referencia. Generalmente, la  $T_c$  será alrededor de 70-90°C. Los dúplex de hibridación de diana-diana también están preferentemente desnaturalizados si la secuencia diana tiene una secuencia nucleotídica que da como resultado una menor  $T_m$  en comparación con la secuencia de referencia. La  $T_c$  se aplica generalmente de alrededor de 1 segundo a 5 minutos, más preferiblemente 2 segundos a 1 minuto, y lo más preferible 5 segundos a 30 segundos. Una vez que la etapa de incubación cíclica está terminada, la temperatura de la mezcla de reacción se disminuye para permitir que el uno o más cebadores se hibride a la secuencia diana. Estos cebadores se alargan entonces mediante una polimerasa, enriqueciendo así la secuencia diana.

Una vez que cada etapa está terminada, la reacción se puede repetir durante múltiples ciclos a fin de obtener suficiente amplificación de las secuencias diana y de referencia. En una realización, las etapas del método se repiten durante 5-40 ciclos, y más preferiblemente 10-30 ciclos. El número óptimo de ciclos se puede determinar por una persona de pericia normal en la técnica. Preferiblemente, los presentes métodos se llevan a cabo en un dispositivo de PCR, más preferiblemente en condiciones de reacción en tiempo real en un dispositivo de PCR de detección en tiempo real, tal como el dispositivo de PCR en tiempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) y el dispositivo de PCR en tiempo real Mx3005P (Stratagene, La Jolla, CA). En esta realización, la mezcla de reacción puede incluir un agente de detección de ácido nucleico (por ejemplo, colorante de detección de ácido nucleico tal como colorante SYBR Green o colorante LC-Green, o una sonda acoplada de forma operativa a un colorante fluorescente) para cuantificar y/o monitorizar los productos de amplificación de la reacción. Una vez que el enriquecimiento de la secuencia diana está terminado, la muestra se somete a procesamiento posterior. El procedimiento posterior incluye MALDI-TOF, fusión de HR, secuenciación didesoxi, secuenciación de una sola molécula, pirosecuenciación, RFLP, PCR digital y PCR cuantitativa.

En otra realización, el método de la presente invención se puede usar para detectar si la metilación se ha producido en una secuencia diana o en una secuencia de referencia. En una realización adicional, el método utiliza ADN genómico para ensayar la metilación.

El método de detección de la metilación comprende un enfoque químico o enzimático para el tratamiento de ADN sensible a la metilación. Los tratamientos químicos incluyen la incubación de ADN con bisulfito de sodio, que convierte selectivamente citocinas no metiladas en uracilos. El ADN se desnaturaliza primeramente por calor, y después se puede tratar con bisulfito 5M, pH 5-7. El pretratamiento de ADN genómico para eliminar uracilos preexistentes se usa antes del tratamiento con bisulfito. Este pretratamiento consiste en el tratamiento con uracilo glucosilasa en presencia de 5 mM de hidroxilamina, pH 7. El ADN modificado se puede usar ahora en los métodos de la invención.

Debido a que las citocinas metiladas de la secuencia de referencia o de la secuencia diana se convierten en uracilos, ahora formarán malos apareamientos cuando forman un dúplex con la hebra opuesta (diana o referencia), que no estaba metilada, durante su etapa de hibridación cruzada de la reacción.

En todavía otro aspecto, cualquiera de los métodos de la invención se usa para enriquecer múltiples secuencias diana diferentes en una reacción de múltiplex. En esta realización, el método incluye conjuntos adicionales de pares de cebadores para las secuencias diana adicionales.

En otro aspecto, la invención se refiere a un medio legible por ordenador que tiene instrucciones de programa para llevar a cabo cualquiera de los métodos de la invención. Se describe además un sistema de PCR para enriquecer la secuencia diana. El sistema incluye memoria para implementar las instrucciones del programa del medio legible por ordenador.

Las Figuras 1 y 2 ilustran dos aspectos diferentes de la invención. La Figura 1 ilustra un aspecto de la invención en el que el método utiliza una reacción de amplificación que tiene una primera temperatura de desnaturalización y una temperatura de desnaturalización crítica o  $T_c$ . La Figura 2 también ilustra una reacción de amplificación que tiene una primera temperatura de desnaturalización y una  $T_c$  pero que incluye además cambiar de forma oscilante, o repetir, la etapa de temperatura de hibridación y de temperatura de desnaturalización crítica múltiples veces antes de la fase de hibridación y alargamiento de los cebadores de la reacción.

La Figura 1 muestra el procedimiento para enriquecer una secuencia diana en una muestra de ácido nucleico que tiene secuencias diana y de referencia. Las secuencias diana y de referencia se pueden obtener a partir de cualquier secuencia de ácido nucleico, incluyendo ADN genómico, ADNc, ADN vírico, ADN de mamífero, ADN fetal, o ADN bacteriano. Mientras que la secuencia de referencia es generalmente el alelo y la secuencia diana es el alelo mutante, lo contrario también puede ser cierto. El alelo mutante puede incluir una cualquiera o más supresiones, inserciones o alteraciones nucleotídicas. En algunas realizaciones, el alelo mutante es una mutación somática. En otras realizaciones, la secuencia diana es ADN metilado, mientras que la secuencia de referencia es ADN no metilado. Como alternativa, la secuencia diana es ADN no metilado, mientras que la secuencia de referencia es ADN metilado.

El método incluye someter la mezcla de reacción de amplificación a una primera temperatura de desnaturalización (Fig. 1A) que está por encima de la temperatura de fusión " $T_m$ " de una secuencia de referencia. La  $T_m$  de un ácido nucleico se puede determinar mediante experimentación, o se puede estimar mediante cálculo. El experto está bien al tanto de numerosos métodos bien conocidos para determinar la  $T_m$  de un ácido nucleico, algunos de los cuales se describen aquí. La primera temperatura de desnaturalización se selecciona generalmente como aquella que se seleccionaría generalmente como la temperatura de desnaturalización de una reacción de PCR, y debería de ser suficientemente elevada para permitir la desnaturalización total de las secuencias diana y de referencia (por ejemplo, 94°C). En una realización, la primera temperatura de desnaturalización está alrededor de 1°C a 30°C por encima de la  $T_m$  de la secuencia de referencia, más preferiblemente la  $T_m$  de la secuencia de referencia está alrededor de 5°C a 20°C por encima de la  $T_m$  de la secuencia de referencia.

5 A continuación, la temperatura de la mezcla de reacción de amplificación se disminuye, permitiendo que las  
 10 secuencias diana y las secuencias de referencia se hibriden (Fig. 1B). Esta etapa de hibridación da como resultado  
 la formación de dúplex de hibridación de secuencias diana-diana, referencia-referencia y diana-referencia. La  
 determinación de una temperatura de hibridación es bien conocida por el experto. Los cebadores de la PCR usados  
 en el método se diseñan para que tengan una  $T_m$  que evita que se unan a las secuencias diana y de referencia a  
 esta temperatura intermedia, de manera que no interfieran con la hibridación cruzada de las secuencias mutante  
 (diana) y de tipo salvaje (de referencia). Debido a las mutaciones en la secuencia diana, la mayoría de las  
 secuencias diana terminan en una estructura mal apareada con la secuencia de referencia, y de este modo tienen  
 una menor temperatura de fusión, cuando forman un heterodúplex con la secuencia de referencia, que los  
 homodúplex de referencia/referencia completamente apareados.

15 Los dúplex de hibridación de diana-referencia se desnaturalizan entonces preferentemente incrementando la  
 temperatura de la mezcla de reacción hasta la  $T_c$  (Fig. 1C). La  $T_c$  o temperatura crítica está por debajo de la  $T_m$  de  
 la secuencia de referencia, y se puede determinar mediante los métodos descritos aquí. En una realización, la  $T_c$   
 está alrededor de 0,3°C-5°C por debajo de la  $T_m$  de la secuencia de referencia, y más preferiblemente alrededor de  
 0,5°C a 1,5°C por debajo de de la  $T_m$  de la secuencia de referencia. Generalmente, la  $T_c$  será alrededor de 70-90°C.  
 Los dúplex de hibridación de diana-diana también se pueden desnaturalizar preferentemente si la secuencia diana  
 tiene una secuencia nucleotídica que da como resultado una menor  $T_m$  en comparación con la secuencia de  
 referencia. La  $T_c$  se aplica generalmente de alrededor de 1 segundo a 5 minutos, más preferiblemente 2 segundos a  
 1 minuto, y lo más preferible 5 segundos a 30 segundos.

20 Después de la desnaturalización preferente de los dúplex de hibridación de las secuencias diana-referencia y/o  
 diana-diana, la temperatura de la mezcla de reacción se reduce para permitir que uno o más cebadores se hibriden  
 a la secuencia diana (Fig. 1D). Los cebadores hibridados se alargan entonces mediante una polimerasa de ácido  
 nucleico, enriqueciendo así la secuencia diana en la población de ácidos nucleicos contenida en la muestra.

25 Las etapas del método se repiten generalmente durante múltiples ciclos a fin de obtener la suficiente amplificación  
 de las secuencias diana y de referencia. En una realización, las etapas del método se repiten durante 5-40 ciclos, y  
 más preferiblemente 10-30 ciclos. El número óptimo de ciclos se puede determinar por una persona de pericia  
 normal en la técnica. Preferiblemente, los presentes métodos se llevan a cabo en un dispositivo de PCR, más  
 preferiblemente en condiciones de reacción en tiempo real en un dispositivo de PCR de detección en tiempo real, tal  
 como el dispositivo de PCR en tiempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) y el dispositivo de PCR en  
 tiempo real Mx3005P (Stratagene, La Jolla, CA). En esta realización, la mezcla de reacción puede incluir un agente  
 de detección de ácido nucleico (por ejemplo, un colorante de detección de ácido nucleico tal como el colorante  
 SYBR Green o el colorante LC-Green, o una sonda acoplada de forma operativa a un colorante fluorescente) para  
 cuantificar y/o monitorizar los productos de amplificación de la reacción. Una vez que el enriquecimiento de la  
 secuencia diana está terminado, la muestra se puede procesar adicionalmente, por ejemplo se puede someter a una  
 35 reacción de secuenciación. Los alelos enriquecidos se pueden procesar adicionalmente mediante una variedad de  
 procedimientos que incluyen: MALDI-TOF, fusión de HR, secuenciación didesoxi, secuenciación de una sola  
 molécula, pirosecuenciación, RFLP, PCR digital y PCR cuantitativa.

40 Llevando a cabo el método de enriquecimiento en cada ciclo de PCR, la cantidad de secuencias mutantes  
 (secuencia diana) se enriquece gradualmente con respecto a las secuencias (secuencia de referencia). Mediante el  
 método se enriquecen tanto mutaciones homocigotas como heterocigotas. El enriquecimiento en unas 10-60 veces  
 de secuencias que contienen mutaciones con respecto a la realización de la PCR normal a una temperatura de  
 desnaturalización de 94°C es habitual. A una temperatura de desnaturalización crítica ( $T_c$ ) dada, el enriquecimiento  
 de las mutaciones se produce simultáneamente en todas las posiciones de la secuencia, no obstante con una  
 eficiencia diferente dependiendo del contexto de la secuencia y del tamaño global del amplicón de PCR. Tanto la  
 45 temperatura de desnaturalización crítica  $T_c$  como el enriquecimiento anticipado en cualquier posición son  
 predecibles usando el software de fusión de ADN apropiado, y se pueden verificar experimentalmente. De este  
 modo, dependiendo de dónde está la mutación, se puede anticipar un enriquecimiento en cierto modo diferente; sin  
 embargo, en todos los casos se logra un enriquecimiento sustancial y por lo tanto mejoran los límites de detección  
 de ensayos aguas abajo, por ejemplo las reacciones de secuenciación.

50 La Figura 2 ilustra una realización del método para enriquecer una secuencia diana sometiendo una mezcla de  
 reacción de amplificación a múltiples etapas alternas de hibridación y temperatura de desnaturalización crítica. Esta  
 realización aprovecha tanto la desnaturalización preferente a la temperatura crítica como la hibridación cruzada  
 preferente a la temperatura de hibridación.

55 La mezcla de reacción de amplificación, que tiene una secuencia diana y de referencia, se somete en primer lugar a  
 una primera temperatura de desnaturalización, que está por encima de la  $T_m$  de la secuencia de referencia (Fig. 2A).

60 A continuación, la muestra se cicla entre dos etapas de incubación de temperatura diferente. En la primera etapa de  
 incubación, la temperatura se disminuye para permitir la hibridación preferente de la secuencia diana con la  
 secuencia de referencia (Fig. 2B). En la segunda etapa de incubación, la temperatura se incrementa hasta la  $T_c$  (Fig.  
 2C). Estas etapas primera y segunda se repiten entonces una o más veces, más preferiblemente 3-20 veces, y lo  
 más preferible 5-10 veces.

La temperatura de hibridación preferente para una secuencia dada es una temperatura a la que los alelos de tipo salvaje se hibridan a sí mismos a una velocidad más rápida que lo que lo hacen los alelos que contienen mutaciones. Debido a que los alelos mutados predominan mucho menos que los alelos de tipo salvaje, la hibridación cruzada de los alelos de tipo salvaje con sí mismos transcurre de forma más rápida que la hibridación cruzada de los alelos con alelos mutantes, o de los alelos mutantes con alelos (formando estos últimos un mal apareamiento). Como resultado, cuando la temperatura de la PCR se reduce hasta una temperatura de hibridación cruzada, los alelos mutantes no se hibridan de forma cruzada en el mismo grado que lo que lo hacen los alelos de tipo salvaje. Esta temperatura de hibridación o temperatura intermedia (estando la temperatura por debajo de la primera temperatura de desnaturalización y de  $T_c$  pero por encima de la temperatura de hibridación/alargamiento de los cebadores, por ejemplo alrededor de 60°C a 80°C) está por encima de la  $T_m$  del par de cebadores, y de este modo permite que las secuencias diana y de referencia se hibriden mientras evita la unión del par de cebadores a las secuencias diana y/o de referencia.

A continuación, la temperatura de la reacción se incrementa hasta la  $T_c$  dando como resultado que los dúplex de hibridación de diana-referencia y diana-diana se desnaturalicen preferentemente (Fig. 2C). La  $T_c$  o temperatura crítica está por debajo de la  $T_m$  de la secuencia de referencia, y se puede determinar mediante los métodos descritos aquí. En una realización, la  $T_c$  está alrededor de 0,3°C-5°C por debajo de la  $T_m$  de la secuencia de referencia, y más preferiblemente alrededor de 0,5°C a 1,5°C por debajo de la  $T_m$  de la secuencia de referencia. Generalmente, la  $T_c$  será alrededor de 70-90°C. Los dúplex de hibridación de diana-diana también se pueden desnaturalizar preferentemente si la secuencia diana tiene una secuencia nucleotídica que da como resultado una menor  $T_m$  en comparación con la secuencia de referencia. La  $T_c$  se aplica generalmente de alrededor de 1 segundo a 5 minutos, más preferiblemente 2 segundos a 1 minuto, y lo más preferible 5 segundos a 30 segundos. Este procedimiento se repite varias veces, oscilando entre la temperatura de hibridación (Fig. 2B) y la temperatura crítica (Fig. 2C), generando selectivamente cada vez más secuencias mutantes en la forma monocatenaria que secuencias en la forma monocatenaria.

Una vez que la etapa de incubación cíclica está terminada, la temperatura de la mezcla de reacción se disminuye para permitir que el uno o más cebadores se hibriden a la secuencia diana (Fig. 2D). Estos cebadores se alargan entonces mediante una polimerasa, enriqueciendo de este modo la secuencia diana. Las etapas del método se repiten generalmente durante múltiples ciclos a fin de obtener una amplificación suficiente de las secuencias diana y de referencia. En una realización, las etapas del método se repiten durante 5-40 ciclos, y más preferiblemente 10-30 ciclos. El número óptimo de ciclos se puede determinar por una persona de pericia normal en la técnica. Preferiblemente, los presentes métodos se llevan a cabo en un dispositivo de PCR, más preferiblemente en condiciones de reacción en tiempo real en un dispositivo de PCR de detección en tiempo real.

Una vez que el enriquecimiento de la secuencia diana está terminado, la muestra se puede someter a procesamiento posterior. El procesamiento posterior incluye MALDI-TOF, fusión de HR, secuenciación dideoxi, secuenciación de una sola molécula, pirosecuenciación, RFLP, PCR digital y PCR cuantitativa.

La PCR digital se puede usar en la detección de mutaciones de niveles ultrabajos en combinación con un procedimiento de enriquecimiento. En la PCR digital, la muestra de ADN se diluye hasta moléculas individuales, de manera que en cada reacción de PCR el material de partida es tipo salvaje o mutante. Tras un gran número de reacciones de PCR del mismo material de partida, las moléculas mutantes se aíslan y se detectan. Fluidigm (South San Francisco, CA) comercializa una variedad de sistemas a base de PCR digital que se pueden usar con la presente invención. En consecuencia, el enriquecimiento en tiempo real se puede llevar a cabo simultáneamente a partir de moléculas individuales en miles de reacciones de enriquecimiento paralelas, permitiendo la identificación de la reacción de PCR que originó el ADN mutante. Tal sistema es particularmente útil para la detección de mutaciones ultrabajos en genomas de cáncer. La combinación de PCR digital con el procedimiento de enriquecimiento puede beneficiar de forma similar a aplicaciones de secuenciación de una sola molécula.

#### Reacción de amplificación de ácido nucleico

En una realización, una muestra de ácido nucleico utilizada en el método de la invención comprende ADN genómico que tiene una secuencia diana y una secuencia de referencia. En otra realización, la muestra de ácido nucleico del método de la invención comprende una secuencia diana y una secuencia de referencia que se amplificaron previamente en una reacción de amplificación de ácido nucleico. El experto apreciará que hay muchos métodos disponibles para amplificar un ácido nucleico. Quizás el método más popular es la reacción en cadena de la polimerasa (PCR; por ejemplo, véanse, patentes U.S. n<sup>os</sup> 4.683.195 y 4.683.202, así como Saiiki et al., Science 230:1350-1354 (1985) y Gyllensten et al., PNAS (USA) 85:7652-7656 (1985)). Una variación preferida del método de la PCR es la PCR asimétrica (por ejemplo, véanse Mao et al., Biotechniques 27(4):674-678 (1999); Lehbein et al., Electrophoresis 19(8-9):1381-1384 (1998); Lazaro et al., Molec. Cell. Probes 6(5):357-359 (1992); y patente U.S. n<sup>o</sup> 6.197.499). Otros métodos de amplificación incluyen, pero no se limitan a, amplificación de desplazamiento de hebra (SDA) (véanse, Walker et al., Nuc. Acids Res. 20(7):1691-1696 (1992), así como las patentes U.S. n<sup>os</sup> 5.744.311, 5.648.211 y 5.631.147), amplificación por círculo rodante (RCA) (véase la publicación PCT WO 97/19193), la amplificación a base de secuencias de ácidos nucleicos (NASBA) (véanse Compton, Nature 350:91-92 (1991); así como las patentes U.S. n<sup>os</sup> 5.409.818 y 5.554.527), amplificación mediada por transcritos (TMA) (véanse Kwoh et al., PNAS (USA) 86:1173-1177 (1989), así como la patente U.S. n<sup>o</sup> 5.399.491), replicación de secuencia autosostenida

(3SR) (véase Guatelli et al., PNAS (USA) 87:1874-1879 (1990) y la reacción en cadena de ligasa (LCA) (véanse las patentes U.S. n<sup>os</sup> 5.427.930 y 5.792.607).

El presente método utiliza una PCR modificada. La PCR se lleva a cabo como se describe en Mullis y Faloona, 1987, Methods Enzymol., 155: 335. La técnica de la reacción en cadena de la polimerasa (PCR) se describe en las patentes U.S. n<sup>os</sup> 4.683.202, 4.683.195 y 4.800.159. En su forma más simple, la PCR es un método *in vitro* para la síntesis enzimática de secuencias de ADN específicas, usando dos cebadores oligonucleótidos que se hibridan a hebras opuestas y flanquean la región de interés en el ADN diana. Una serie repetitiva de etapas de reacción que implican la desnaturalización del molde, la hibridación del cebador y el alargamiento de los cebadores hibridados mediante ADN polimerasa da como resultado la acumulación exponencial de un fragmento específico cuyos términos se definen mediante los extremos 5' de los cebadores. Se da a conocer que la PCR es capaz de producir un enriquecimiento selectivo de una secuencia de ADN específica en un factor de 10<sup>9</sup>. El método de la PCR también se describe en Saiki et al., 1985, Science 230:1350.

La PCR se lleva a cabo usando ADN molde (secuencias diana y de referencia) (al menos 1 fg; de forma más útil, 1-1000 ng) y al menos 25 pmoles de cebadores oligonucleotídicos. Una mezcla de reacción típica incluye: 2 µl de ADN, 25 pmoles de cebador oligonucleotídico, 2,5 µl de un tampón adecuado, 0,4 µl de dNTP 1,25 µM, 2,5 unidades de Taq DNA polymerase (Stratagene), y agua desionizada hasta un volumen total de 25 µl. La PCR se lleva a cabo usando un ciclador térmico programable.

La duración y temperatura de cada etapa de un ciclo de PCR, así como el número de ciclos, se ajustan según los requisitos de restricción en efecto. La temperatura y tiempo de hibridación se determinan tanto mediante la eficiencia con la que se espera que un cebador se hibride a un molde como por el grado de mal apareamiento que se va a tolerar. La capacidad para optimizar la restricción de las condiciones de hibridación de cebadores está dentro del conocimiento de alguien de pericia moderada en la técnica. Se usa una temperatura de hibridación de entre 30°C y 72°C. La desnaturalización inicial de las moléculas del molde se produce normalmente entre 92°C y 99°C durante 4 minutos, seguido de 20-40 ciclos que consisten en desnaturalización (94-99°C durante 15 segundos a 1 minuto), hibridación (temperatura determinada como se explica anteriormente; 1-2 minutos), y alargamiento (72°C durante 1 minuto). La etapa de alargamiento final se lleva a cabo generalmente durante 4 minutos a 72°C, y puede ser seguida de una etapa indefinida (0-24 horas) a 4°C.

La PCR utiliza una polimerasa de ácido nucleico, o enzima que cataliza la polimerización de trifosfatos de nucleósidos. Generalmente, la enzima iniciará la síntesis en el extremo 3' del cebador hibridado a la secuencia diana, y transcurrirá en la dirección 5' a lo largo del molde. Las ADN polimerasas conocidas incluyen, por ejemplo, ADN polimerasa I de *E. coli*, T7 ADN polimerasa, ADN polimerasa de *Thermus thermophilus* (Tth), ADN polimerasa de *Bacillus stearothermophilus*, ADN polimerasa de *Thermococcus litoralis*, ADN polimerasa de *Thermus aquaticus* (Taq) y ADN polimerasa de *Pyrococcus furiosus* (Pfu). La expresión "polimerasa de ácido nucleico" también engloba ARN polimerasas. Si el molde de ácido nucleico es ARN, entonces "polimerasa de ácido nucleico" se refiere a una actividad de polimerización dependiente de ARN, tal como una transcriptasa inversa.

En los métodos de la invención, el protocolo de PCR incluye además una etapa de desnaturalización crítica. Estos protocolos se describen aquí completamente y se ilustran en las Figuras 1 y 2.

Preferiblemente, los procedimientos de enriquecimiento se llevan a cabo en un dispositivo de PCR, más preferiblemente en condiciones de reacción en tiempo real en un dispositivo de PCR en tiempo real. Las condiciones de reacciones en tiempo real utilizan además un agente de detección de ácido nucleico (por ejemplo, colorante o sonda) a fin de medir/detectar el producto de la PCR según se produce.

En una realización, los métodos de enriquecimiento se practican en un formato de múltiplex. La PCR múltiplex es aquella cuando se usa más de un par de cebadores en una reacción de PCR para detectar más de una secuencia diana. El objetivo de la PCR múltiplex es amplificar más de una secuencia diana simultáneamente y de ese modo ahorrar tiempo y minimizar el gasto. Permite la amplificación de múltiples secuencias diana en un solo experimento. Generalmente, el uso de PCR múltiplex junto con la presente invención con el fin de enriquecer simultáneamente los alelos minoritarios en todas las secuencias amplificadas, los cebadores se diseñan de manera que los amplicones de la PCR resultantes comparten todos ellos aproximadamente la misma T<sub>c</sub> (temperatura de desnaturalización crítica).

Determinación de la T<sub>m</sub> y de la T<sub>c</sub>

La T<sub>m</sub> se puede definir como la temperatura a la que la mitad de los pares de bases de Watson-Crick en una molécula de ácido nucleico dúplex se rompen o se disocian (es decir, se "funden") mientras que la otra mitad de los pares de bases de Watson-Crick permanece intacta en una conformación bicatenaria, mientras que la "temperatura crítica", "temperatura de desnaturalización crítica" o "T<sub>c</sub>" se refiere a una temperatura por debajo de la T<sub>m</sub> de la secuencia de referencia. La T<sub>c</sub> se aplica para desnaturalizar selectivamente la secuencia diana bicatenaria o el dúplex bicatenario de secuencia diana/secuencia de referencia en una muestra de ácido nucleico para permitir el enriquecimiento selectivo de la secuencia diana durante una reacción de amplificación.

La  $T_m$  de un par dado de hebras de ácido nucleico es indicativa de la estabilidad de la unión de hebra a hebra, y depende de la complementariedad de las hebras, de la longitud de la secuencia, del contenido de GC, de la presencia o ausencia de mal apareamientos en la región bicatenaria, y de otros factores de menor importancia, por ejemplo la concentración salina de la muestra (Lewin, Genes V, Capítulo 5, Oxford University Press and Cell Press: Nueva York, (1994) p. 109-126; SantaLucia, 1998).

La temperatura del punto de fusión se determina habitualmente de forma experimental sometiendo la muestra a un incremento constitutivo de temperatura y midiendo continuamente la disociación del dúplex de hibridación en hebras individuales. La disociación se puede detectar mediante una variedad de diferentes métodos, por ejemplo mediante un desplazamiento en la absorbancia de UV, mediante la fluorescencia de colorantes que se unen al ADN bicatenario, mediante resonancia de plasmones de superficie, o preferiblemente por medio de fluorescencia. En este último caso, la sonda de hibridación se marca habitualmente con una entidad fluorescente, y la generación de una señal fluorescente depende en cierto modo de la formación del dúplex de hibridación.

La  $T_m$  se puede determinar experimentalmente, o se puede estimar basándose en métodos bien definidos conocidos por aquellos de pericia normal en la técnica. Los métodos para observar y analizar las transiciones de la desnaturalización de ácidos nucleicos incluyen: medir el cambio de entalpía en una muestra a medida que se desnaturaliza mediante calorimetría de barrido diferencial (DSC) (Kulinski et al., Nucleic Acids Res. 19(9):2449-2455 (1991); Paner et al., Biopolymers 29:1715-1734 (1990); Volker et al., Biopolymers 50:303-318 (1999)), midiendo la fluorescencia de pares de fluoróforos unidos covalentemente (Vamosi y Clegg, Biochemistry 37:14300-14316 (1998)), y monitorizando el cambio en la hipercromicidad de los ácidos nucleicos (Haugland, "In Vitro Applications for Nucleic Acid Stains and Probes", en Handbook of Fluorescent Probes and Research Chemicals, 6ª ed., Molecular Probes Inc, Eugene OR (1996) p. 161-174).

Los valores de  $T_m$  de los ácidos nucleicos bicatenarios también se pueden observar monitorizando la fluorescencia de colorantes específicos del ADN bicatenario combinados con los ácidos nucleicos (Wittwer et al., 1996). Los colorantes específicos bicatenarios son fluoróforos que se unen a ácidos nucleicos. Típicamente, la fluorescencia de estos colorantes aumenta cuando se unen a ácidos nucleicos en forma de dúplex (Wittwer et al., BioTechniques 22:176-181 (1997)). Ririe et al. (1997) demostraron que los productos post-PCR se pueden diferenciar mediante análisis de la curva de fusión usando el colorante de unión a ácidos nucleicos bicatenarios SYBR® Green I. SYBR® Green I se une preferentemente a ácido nucleico bicatenario (Haugland, 1996). Otros colorantes adecuados para determinar la  $T_m$  de un ácido nucleico bicatenario incluyen SYBR® Gold, bromuro de etidio, naranja de acridina, bromuro de propidio, PicoGreen®, Hoechst 33258, Hoechst 33342, Hoechst 34580, YO-PRO®-1 y YOYO®-1. Cada uno de estos colorantes está comercialmente disponible. Por ejemplo, el Capítulo 8 del catálogo de Molecular Probes (Eugene, Oreg.) Handbook of Fluorescent Probes and Research Products, Decimotava Edición (en CD-ROM, mayo de 2001) enumera una multitud de colorantes que se pueden usar en la presente invención.

Como apreciarán los expertos, la fusión de cualquier estructura de ácido nucleico bicatenario se produce generalmente en una proporción sustancial de una población de ácidos nucleicos similares a lo largo de un intervalo limitado de temperatura, y tendrá típicamente un pico de fusión (transición más rápida) a aproximadamente la  $T_m$  para ese ácido nucleico. Por lo tanto, tales picos de cambio en la emisión de fluorescencia se pueden usar para calcular la  $T_m$  para ácidos nucleicos bicatenarios.

Un perfil de temperatura de fusión se puede representar gráficamente dibujando  $-dF/dT$  frente a  $T$ , en el que  $dF$  es el cambio en la emisión de fluorescencia medida,  $dT$  es el cambio en la temperatura del ácido nucleico, y  $T$  es la temperatura del ácido nucleico. Tal representación gráfica mostrará picos a temperaturas a las que se producen los cambios más rápidos en la fluorescencia, indicando temperaturas de fusión.

En la técnica se conocen otros métodos para determinar la  $T_m$  de un ácido nucleico bicatenario, incluyendo aquellos descritos en las patentes U.S. n<sup>os</sup> 7.226.736 y 6.030.115.

La  $T_m$  se puede predecir a partir de ecuaciones empíricas. Se sabe que la  $T_m$  depende de la longitud de la secuencia de un par de bases complementarias formadoras de ácido nucleico ( $n$ ), del contenido de G y C en la secuencia, de las concentraciones de sal ( $\mu$ ) y del agente desnaturalizante (% FA) en la disolución de la muestra, y, en general, sigue una ecuación empírica  $T_m = 81,5 + 16,6 \log(\mu) + 0,41(\% GC) - 500/n - 0,61(\% FA)$ . La  $T_m$  se puede estimar mediante un número de métodos, por ejemplo mediante un cálculo de vecino más próximo de acuerdo con Wetmur 1991 (Wetmur, J. G. 1991. DNA probes: applications of the principles of nucleic acid hybridization. Crit Rev Biochem Mol Biol 26: 227-259) y mediante programas comerciales, incluyendo Oligo™ Primer Design y programas disponibles en internet.

La "temperatura crítica" o " $T_c$ " se refiere a una temperatura por debajo de la  $T_m$  de la secuencia de referencia. La  $T_c$  se aplica para desnaturalizar preferentemente el dúplex de secuencia diana bicatenario o el dúplex bicatenario de secuencia diana/secuencia de referencia con respecto al dúplex de secuencia de referencia/referencia, para permitir el enriquecimiento selectivo de la secuencia diana durante una reacción de amplificación. La temperatura crítica aprovecha la menor  $T_m$  de la secuencia diana bicatenaria o del dúplex de ADN bicatenario hibridado de forma cruzada de diana-referencia. Cuando la secuencia diana o la secuencia de referencia se hibridan de forma cruzada, diferencias de secuencia menores de uno o más mal apareamientos de un solo nucleótido en cualquier parte a lo

largo de una secuencia de ADN bicatenaria corta (por ejemplo, < 200 pb) generarán un cambio pequeño pero predecible en la temperatura de fusión ( $T_m$ ) para esa secuencia (Lipsky, R.H., et al. (2001) Clin Chem, 47, 635-644; Liew, M., et al. (2004) Clin Chem, 50, 1156-1164). Dependiendo del contexto exacto de la secuencia y de la posición del mal apareamiento, son habituales cambios de temperatura de fusión de 0,5-1,5°C para secuencias de hasta 200 pb. De este modo, la hibridación de la secuencia diana-referencia, debido a los mal apareamientos, tendrá esencialmente una menor  $T_m$  que el alelo conocido (por ejemplo, secuencia de referencia). Al menos en parte, la presente invención aprovecha la pequeña diferencia de  $T_m$  entre secuencias completamente apareadas y secuencias mal apareadas. Debido a que la desnaturalización crítica se lleva a cabo en cada ciclo de PCR, el enriquecimiento diferencial de alelos que contienen mutaciones se realiza exponencialmente, y resulta en una gran diferencia en la eficiencia de la amplificación global entre alelos mutantes y de tipo salvaje al final del ciclo. La  $T_c$  está alrededor de 0,1-20°C por debajo de la  $T_m$  de la secuencia de referencia. Más preferiblemente, la  $T_c$  está alrededor de 0,1-15°C, 0,1-10°C, 0,1-9°C, 0,1-8°C, 0,1-7°C, 0,1-6°C, 0,2-5°C, 0,3-4,5°C, 0,4-4°C, 5-3,5°C, 0,5-3°C, 0,5-3°C, 0,5-2,5°C, 0,5-2°C, 0,5-1,5°C, 0,5-1°C por debajo de la  $T_m$  de la secuencia de referencia.

En algunas realizaciones, la  $T_c$  puede estar por debajo de la  $T_m$  tanto de la secuencia de referencia como de la secuencia diana. Por ejemplo, en un caso, la  $T_m$  de una secuencia de tipo salvaje fue 84°C y la  $T_m$  de la secuencia mutada fue 83,8°C. Cuando se usó una versión rápida de un procedimiento de enriquecimiento, la  $T_c$  óptima fue 83,5°C.

En algunas realizaciones preferidas, la  $T_c$  se escoge de manera que está por debajo de la  $T_m$  tanto de la secuencia de referencia como de todas las posibles secuencias diana. Por ejemplo, en un caso, la  $T_m$  de una secuencia de tipo salvaje fue 84°C, y la  $T_m$  de secuencias mutadas en diferentes posiciones (mutaciones de un solo punto) fue 83,8°C, 83,7°C, 83,9°C, 83,6°C, 83,75°C. Cuando se usó una versión rápida de un procedimiento de enriquecimiento, la  $T_c$  óptima fue 83,5°C.

La temperatura de desnaturalización crítica ( $T_c$ ) es la temperatura por debajo de la cual la eficiencia de la PCR cae abruptamente para una secuencia de ácido nucleico dada. Por ejemplo, una secuencia de p53 de 167 pb se amplifica bien si la temperatura de desnaturalización de la PCR se fija a 87°C, se amplifica de forma modesta a 86,5°C, y no produce ningún producto detectable si la desnaturalización de la PCR se fija a 86°C o menos.

Al igual que la  $T_m$ , la  $T_c$  para una secuencia dada se puede identificar experimentalmente o mediante cálculo. Para identificar experimentalmente la  $T_c$  para un producto de PCR dado, se lleva a cabo una curva de fusión en tiempo real en presencia de un colorante intercalante (LC-GREEN o SYBR-Green), para obtener la temperatura de fusión media  $T_m$  de la secuencia (véase anteriormente con respecto a la determinación de  $T_m$ ). Las temperaturas de alrededor de 0,5-1,5°C menores que  $T_m$  son habitualmente temperaturas de desnaturalización críticas  $T_c$  apropiadas que dan como resultado el enriquecimiento de secuencias diana.

La  $T_c$  también se puede estimar determinando el  $\Delta T_m$  para una secuencia de ADN debido a su mal apareamiento de pares de bases, como estaría presente en el dúplex de diana-referencia hibridado de forma cruzada. Esta diferencia es de alrededor de 0,1°C a alrededor de 12,5°C en comparación con un apareamiento perfecto de secuencia de referencia/referencia. De este modo, en una realización,  $T_c$  se puede representar mediante la ecuación  $T_c = T_m - \Delta T_m$ , en la que  $T_m$  es la temperatura de fusión del dúplex de secuencia de referencia/referencia, y  $\Delta T_m$  es el cambio en la  $T_m$  de la secuencia de referencia como resultado de uno o más mal apareamientos de pares de bases que se pueden formar durante la hibridación cruzada de las secuencias diana/referencia. Se ha encontrado que el  $\Delta T_m$  depende de la longitud del dúplex de secuencia diana/referencia, del porcentaje de contenido de guanina-citosina (%GC), y también de la localización de la mutación de punto o del mal apareamiento de los pares de bases en el dúplex. Por ejemplo, si el mal apareamiento se encuentra hacia el extremo del dúplex,  $\Delta T_m$  será generalmente menor, típicamente en el intervalo de alrededor de 0,1°C a alrededor de 8°C. Si el mal apareamiento se encuentra hacia el centro del dúplex,  $\Delta T_m$  será relativamente mayor, típicamente en el intervalo de alrededor de 0,2°C a alrededor de 11°C. El  $\Delta T_m$  es equivalente en general a alrededor de 0,5°C a alrededor de 1,5°C por porcentaje de mal apareamiento de bases en el dúplex hibridado de forma cruzada de diana-referencia. Se ha de apreciar que  $\Delta T_m$  variará dependiendo no solo de la longitud del dúplex y de la localización de la mutación, sino que también variará dependiendo del orden específico de la secuencia. Consiguientemente, todas las citas diferencias están dentro del alcance de la descripción.

50 Secuencias de ácidos nucleicos útiles en la invención

La invención proporciona métodos para enriquecer una secuencia diana en una muestra de ácido nucleico, y también utiliza cebadores para amplificar una secuencia de ácido nucleico molde.

Como se usa aquí, las expresiones “ácido nucleico”, “polinucleótido” y “oligonucleótido” se refieren a cebadores, sondas, y fragmentos oligoméricos a detectar, y deben ser genéricas a polidesoxirribonucleótidos (que contienen 2-desoxi-D-ribosa), a polirribonucleótidos (que contienen D-ribosa), y a cualquier otro tipo de polinucleótido que sea un N-glucósido de una base purínica o pirimidínica, o bases purínicas o pirimidínicas modificadas (incluyendo sitios abásicos). No hay ninguna distinción pretendida en longitud entre la expresión “ácido nucleico”, “polinucleótido” y “oligonucleótido”, y estas expresiones se usarán de forma intercambiable. Estas expresiones se refieren solamente a

la estructura primaria de la molécula. De este modo, estos términos incluyen ADN mono y bicatenario, así como ARN mono y bicatenario.

El oligonucleótido no deriva necesariamente de forma física de ninguna secuencia existente o natural, sino que se puede generar de cualquier manera, incluyendo síntesis química, replicación de ADN, transcripción inversa, o una combinación de las mismas. Las expresiones "oligonucleótido" o "ácido nucleico" significan un polinucleótido de ADN o ARN genómico, ADNc, origen semisintético o sintético que, en virtud de su origen sintético o manipulación: (1) no está asociado con todo o con una porción del polinucleótido con el que está asociado en la naturaleza; y/o (2) está ligado a un polinucleótido distinto de aquel al que está ligado en la naturaleza.

En una realización, los ácidos nucleicos usados en el método contienen nucleótidos modificados para incrementar las diferencias de temperatura de desnaturalización entre el homodúplex de secuencia de referencia/referencia y el heterodúplex de secuencia diana/referencia. Tales modificaciones potenciarían el enriquecimiento de la secuencia diana. Los nucleótidos modificados o no naturales se pueden incorporar antes o durante el procedimiento de enriquecimiento. Los nucleótidos modificados contemplados para uso en los métodos de la invención incluyen: análogos de diamino-purina (por ejemplo 2'-O-metil-2,6-diaminopurina), uracilo, análogos de ácidos nucleicos peptídicos, análogos de los anteriores modificados con biotina, análogos de los anteriores modificados con fluoróforos, inosina, 7-desazaguanina, nucleótidos de ácido 2'-desoxi-2'-fluoro-β-D-arabinonucleico (2'F-ANA), ácidos nucleicos bloqueados (LNAs), ENAs: ácidos nucleicos con puentes de 2'-O,4'-C-etileno y otros. Los nucleótidos modificados se pueden incorporar en cualquiera de los ácidos nucleicos de la invención, incluyendo ácidos nucleicos molde, cebadores y sondas.

Estas modificaciones pueden incrementar la diferencia de  $T_m$  entre bases apareadas y mal apareadas, y de ese modo pueden incrementar el enriquecimiento obtenido con los presentes métodos. Por ejemplo, los ácidos nucleicos bloqueados representan una clase de análogos nucleotídicos conformacionalmente restringidos descritos, por ejemplo, en el documento WO 99/14226, que incrementan la temperatura de fusión de un ácido nucleico. Los oligonucleótidos que contienen el nucleótido bloqueado se describen en Koshkin, A.A., et al., *Tetrahedron* (1998), 54: 3607-3630) y Obika, S. et al., *Tetrahedron Lett.* (1998), 39: 5401-5404), las cuales se incorporan aquí como referencia. La introducción de un nucleótido bloqueado en un oligonucleótido mejora la afinidad por secuencias complementarias e incrementa la temperatura de fusión en varios grados (Braasch, D.A. y D.R. Corey, *Chem. Biol.* (2001), 8:1-7). La invención se puede llevar a cabo con cualquiera de los LNAs conocidos en la técnica, por ejemplo los descritos en el documento WO 99/14226 y en Latorra D, et al., 2003. *Hum. Mutat.* 22: 79-85.

La complementariedad no necesita ser perfecta; los dúplex estables pueden contener pares de bases mal apareados o bases no apareadas. Los expertos en la técnica de tecnología de ácidos nucleicos pueden determinar empíricamente la estabilidad del dúplex considerando un número de variables que incluyen, por ejemplo, la longitud del oligonucleótido, la composición y secuencia de las bases del oligonucleótido, la fuerza iónica, y la incidencia de pares de bases mal apareados. La estabilidad de un dúplex de ácido nucleico se mide mediante la temperatura de fusión, o " $T_m$ ".

La invención usa cebadores oligonucleotídicos para amplificar una secuencia de ácido nucleico molde.

El término "cebador" se puede referir a más de un cebador, y se refiere a un oligonucleótido, ya sea de origen natural, como en una digestión de restricción purificada, o producido sintéticamente, que es capaz de actuar como un punto de inicio de la síntesis a lo largo de una hebra complementaria cuando se coloca en condiciones en las que está catalizada la síntesis de un producto de alargamiento del cebador que es complementario a una hebra de ácido nucleico. Tales condiciones incluyen la presencia de cuatro trifosfatos de desoxirribonucleósidos diferentes y un agente inductor de la polimerización, tal como ADN polimerasa o transcriptasa inversa, en un tampón adecuado ("tampón" incluye sustituyentes que son cofactores, o que afectan al pH, fuerza iónica, etc.), y a una temperatura adecuada. El cebador es preferiblemente monocatenario para una eficiencia máxima en la amplificación.

Los cebadores oligonucleotídicos útiles según la invención son moléculas de ADN o ARN monocatenarias que son hibridables a una secuencia de ácido nucleico molde y ceban la síntesis enzimática de una hebra de ácido nucleico. El cebador es complementario a una porción de una molécula diana presente en un conjunto de moléculas de ácido nucleico. Se contempla que los cebadores oligonucleotídicos se preparen mediante métodos sintéticos, ya sea químicos o enzimáticos. Como alternativa, tal molécula o un fragmento de la misma es de origen natural, y se aísla de su fuente natural o se adquiere de un proveedor comercial. Los cebadores oligonucleotídicos tienen una longitud de 5 a 100 nucleótidos, idealmente de 17 a 40 nucleótidos, aunque son de uso cebadores de diferente longitud. Los cebadores para la amplificación tienen preferiblemente alrededor de 17-25 nucleótidos. Los cebadores útiles según la invención también se pueden diseñar para que tengan una temperatura de fusión ( $T_m$ ) particular mediante el método de estimación de la temperatura de fusión. Los programas comerciales, incluyendo Oligo<sup>2</sup>, Primer Design y programas disponibles en internet, incluyendo Primer3 y Oligo Calculator, se pueden usar para calcular una  $T_m$  de una secuencia de ácido nucleico útil según la invención. Preferiblemente, la  $T_m$  de un cebador de amplificación útil según la invención, según se calcula por ejemplo mediante Oligo Calculator, está preferiblemente entre alrededor de 45 y 65°C, y más preferiblemente entre alrededor de 50 y 60°C.

Típicamente, la hibridación selectiva se produce cuando dos secuencias de ácidos nucleicos son sustancialmente complementarias (al menos alrededor de 65% complementarias a lo largo de un tramo de al menos 14 a 25 nucleótidos, preferiblemente al menos alrededor de 75%, más preferiblemente al menos alrededor de 90% complementarias). Véase Kanehisa, M., 1984, *Nucleic Acids Res.* 12: 203.

- 5 Como resultado, se espera que se tolere cierto grado de mal apareamiento en el sitio de cebado. Tal mal apareamiento puede ser pequeño, tal como un mono-, di- o trinucleótido. Como alternativa, una región de mal apareamiento puede englobar bucles, que se definen como regiones en las que existe un mal apareamiento en una serie ininterrumpida de cuatro o más nucleótidos.

- 10 Se describe que los procedimientos de enriquecimiento se pueden usar en combinación con cebadores de ácidos nucleicos peptídicos (PNA) para incrementar la sensibilidad del enriquecimiento de las mutaciones. Los PNAs se usan para suprimir la amplificación de las secuencias de tipo salvaje (de referencia) solamente. Los cebadores se pueden sintetizar para distinguir entre una secuencia diana (mutante) y una secuencia de referencia. Los cebadores a base de PNA reconocen y se unen a la secuencia de tipo salvaje complementaria con una mayor estabilidad térmica y especificidad que los cebadores que se unen a la secuencia mutada. Esto no solamente incrementa la diferencia de  $T_m$  entre el cebador de PNA-ácido nucleico de referencia y el cebador normal-ácido nucleico diana, sino que también evita que el cebador a base de PNA se alargue mediante una ADN polimerasa, dando así como resultado un enriquecimiento adicional de la secuencia diana. Tales ensayos son conocidos en la técnica y se describen en Orum et al. *Nucleic Acids Research*, 21(23): 5332-5336 (1993).

- 20 Los cebadores oligonucleotídicos se pueden diseñar con estas consideraciones en mente y se pueden sintetizar según los siguientes métodos.

#### Estrategia de diseño de cebadores oligonucleotídicos

- 25 El diseño de un cebador oligonucleotídico particular con el fin de secuenciar o de realizar la PCR implica seleccionar una secuencia que es capaz de reconocer la secuencia diana, pero tiene una estructura secundaria predicha mínima. La secuencia oligonucleotídica puede unirse o no solamente a un único sitio en el ácido nucleico diana. Además, la  $T_m$  del oligonucleótido se optimiza mediante análisis de la longitud y contenido de GC del oligonucleótido. Además, cuando se diseña un cebador de PCR útil para la amplificación de ADN genómico, la secuencia del cebador seleccionada no demuestra apareamientos significativos a secuencias en la base de datos de GenBank (u otras bases de datos disponibles).

- 30 El diseño de un cebador útil según la invención se facilita mediante el uso de programas de ordenador fácilmente disponibles, desarrollados para ayudar en la evaluación de los varios parámetros descritos anteriormente y la optimización de secuencias de cebadores. Los ejemplos de tales programas son "PrimerSelect" del paquete de software DNASTar™ (DNASTar, Inc.; Madison, WI), OLIGO 4.0 (National Biosciences, Inc.), PRIMER, Oligonucleotide Selection Program, PGEN and Amplify (descrito en Ausubel et al., 1995, *Short Protocols in Molecular Biology*, 3ª Edición, John Wiley & Sons).

- 35 Los cebadores se diseñan para que tengan una  $T_m$  por debajo de la temperatura aplicada durante la etapa de hibridación cruzada de secuencia diana/referencia. De este modo, los cebadores no se hibridan a las secuencias diana o de referencia durante esta etapa de hibridación (véase la Fig. 1). En una realización, la  $T_m$  de los cebadores está 5-10°C por debajo de la temperatura de la etapa de hibridación cruzada.

#### Síntesis

- 40 Los propios cebadores se sintetizan usando técnicas que también son bien conocidas en la técnica. Los métodos para preparar oligonucleótidos de secuencia específica son conocidos en la técnica, e incluyen, por ejemplo, clonación y análisis de digestión de restricción de secuencias apropiadas, y síntesis química directa. Una vez diseñados, los oligonucleótidos se preparan mediante un método de síntesis química adecuado, incluyendo, por ejemplo, el método de fosfotriéster descrito por Narang et al., 1979, *Methods in Enzymology*, 68:90, el método de fosfodiéster descrito por Brown et al., 1979, *Methods in Enzymology*, 68:109, el método de fosforamidato de dietilo descrito en Beaucage et al., 1981, *Tetrahedron Letters*, 22:1859, y el método de soporte sólido descrito en la patente U.S. nº 4.458.066, o mediante otros métodos químicos usando un sintetizador oligonucleotídico automatizado comercial (que está comercialmente disponible) o tecnología VLSIPS™. Los cebadores también se pueden sintetizar con ácidos nucleicos modificados mediante métodos bien conocidos en la técnica.

- 50 Muestras

- Como se usa aquí, "muestra" se refiere a cualquier sustancia que contiene o que se presume que contiene un ácido nucleico de interés (secuencias diana y de referencia), o que es en sí mismo un ácido nucleico que contiene o que se presume que contiene un ácido nucleico diana de interés. El término "muestra" incluye así una muestra de ácido nucleico (ADN genómico, ADNc, ARN), célula, organismo, tejido, fluido, o sustancia, incluyendo, pero sin limitarse a, por ejemplo, plasma, suero, fluido espinal, fluido linfático, fluido sinovial, orina, lágrimas, heces, secreciones externas de la piel, aparato respiratorio, tubo digestivo y aparato genitourinario, saliva, células de la sangre, tumores, órganos, tejido, muestras de constituyentes de cultivos celulares *in vitro*, aislados naturales (tales como agua de

beber, agua marina, materiales sólidos), muestras microbianas, y objetos o muestras que se han “marcado” con moléculas trazadoras de ácidos nucleicos.

5 Las secuencias de ácidos nucleicos se pueden amplificar a partir de ADN genómico. El ADN genómico se puede aislar de tejidos o células según el siguiente método. Como alternativa, las secuencias de ácidos nucleicos se pueden aislar de la sangre mediante métodos bien conocidos en la técnica.

10 Para facilitar la detección de una forma variante de un gen a partir de un tejido particular, el tejido se aísla. Para aislar ADN genómico de tejido de mamífero, el tejido se trocea y se congela en nitrógeno líquido. El tejido congelado se muele a un polvo fino con un mortero y mano de almirez previamente enfriados, y se suspende en tampón de digestión (100 mM de NaCl, 10 mM de Tris-HCl, pH 8,0, 25 mM de EDTA, pH 8,0, 0,5% (p/v) de SDS, 0,1 mg/ml de

15 proteína K) a 1,2 ml de tampón de digestión por 100 mg de tejido. Para aislar ADN genómico de células de cultivo tisular de mamífero, las células se peletizan mediante centrifugación durante 5 min. a 500 x g, se resuspenden en 1-10 ml de PBS enfriado con hielo, se vuelven a peletizar durante 5 min. a 500 x g y se resuspenden en 1 volumen de tampón de digestión.

20 Las muestras en el tampón de digestión se incuban (con agitación) durante 12-18 horas a 50°C, y después se extraen con un volumen igual de fenol/cloroformo/alcohol isoamílico. Si las fases no están resueltas tras una etapa de centrifugación (10 min. a 1700 x g), se añade otro volumen de tampón de digestión (sin proteína K), y se repite la etapa de centrifugación. Si es evidente un material blanco espeso en la interfaz de las dos fases, se repite la etapa de extracción orgánica. Tras la extracción, la capa acuosa superior se transfiere a un nuevo tubo, al que se añadirá 1/2 volumen de acetato de amonio 7,5 M y 2 volúmenes de etanol al 100%. El ácido nucleico se peletiza

25 mediante centrifugación durante 2 min. a 1700 x g, se lava con etanol al 70%, se seca al aire y se resuspende en tampón de TE (10 mM de Tris-HCl, pH 8,0, 1 mM de EDTA, pH 8,0) a 1 mg/ml. El ARN residual se elimina incubando la muestra durante 1 hora a 37°C en presencia de 0,1% de SDS y 1 µg/ml de RNasa libre de DNasa, y repitiendo las etapas de extracción y de precipitación con etanol. Se espera que el rendimiento de ADN genómico según este método sea aproximadamente 2 mg de ADN/1 mg de células o tejido (Ausubel et al., más arriba). El ADN genómico aislado según este método se puede usar según la invención.

30 El ADN diana también se puede extraer de sangre completa. Por ejemplo, la sangre se puede extraer por métodos estándar en un tubo de recogida, que comprende preferiblemente vidrio siliconizado, ya sea sin anticoagulante para preparación de suero, o con EDTA, citrato de sodio, heparina, o anticoagulantes similares, muy preferiblemente EDTA, para la preparación de plasma. El método preferido, aunque no se requiere absolutamente, es que el plasma o suero se fraccione a partir de sangre completa. El plasma o suero se puede fraccionar a partir de sangre completa mediante centrifugación, preferiblemente centrifugación suave a 300 a 800 X g durante 5-10 minutos, o se puede fraccionar mediante otros métodos estándar. Puesto que la heparina puede interferir con la PCR, el uso de sangre

35 heparinizada puede requerir un pretratamiento con heparinasa. De este modo, EDTA es el anticoagulante preferido para muestras de sangre. El plasma sanguíneo o suero recientemente recogido, el plasma o suero congelado (almacenado) y subsiguientemente descongelado, se puede usar en los métodos de la invención. El plasma o suero almacenado se debería de mantener a -20°C a -70°C, y el plasma o suero recientemente recogido se debería de mantener refrigerado o mantener en hielo hasta el uso. El ADN se puede extraer entonces por métodos bien conocidos en la técnica así como aquellos descritos aquí.

#### Ensayos de diagnóstico

40 La invención proporciona un método para enriquecer secuencias diana de una muestra de paciente, para diagnóstico, detección, monitorización, evaluación o tratamiento de un trastorno, en particular una enfermedad neoplásica o proliferativa en un animal o un ser humano. En una realización preferida, el ácido nucleico deriva de un ácido nucleico que codifica un oncogén u otro ADN asociado a tumor.

45 El enriquecimiento preferente de la secuencia diana permite el análisis o manipulación posterior del ADN. Por ejemplo, los alelos enriquecidos se pueden analizar para definir las características de la célula a partir de la que se origina el ADN. Se puede usar cualquiera de varios métodos, dependiendo de la información deseada, incluyendo secuenciación de ácidos nucleicos, RFLP, PCR digital, espectroscopía, incluyendo espectroscopía de RMN de protón, análisis bioquímico, y análisis inmunológico. En una realización, el ADN amplificado se aísla cortando bandas de ADN mutante a partir de un gel de agarosa, se reamplifica, se clona en un vector plasmídico, por ejemplo el vector plasmídico pGEM-T (Promega), y se secuencia usando un kit comercial tal como Sequenase 2.0 (USB). El análisis para definir las características del ADN diana, y de este modo por ejemplo un tumor, produce un conjunto

50 amplio de utilidad clínica, incluyendo la descripción, caracterización, o clasificación de la célula (por ejemplo, tumor), si se conoce o está oculta, tal como mediante tejido de origen, por tipo (tal como premaligno o maligno), fenotipo, y genotipo, y mediante descripción o caracterización del comportamiento tumoral, fisiología y bioquímica, para obtener una comprensión de la invasividad tumoral, de la propensión a metastatizarse, y de la sensibilidad o resistencia a diversas terapias, permitiendo de ese modo la predicción de la respuesta a una terapia continuada o planeada y, además, permitiendo la evaluación del pronóstico. La comparación de las características del ADN diana con muestras biológicas o quirúrgicas previas permite la evaluación adicional de la heterogeneidad o similitud tumoral en comparación con esa muestra, y de este modo la evaluación de la recurrencia tumoral.

También tras el enriquecimiento selectivo de la secuencia diana, el ácido ribonucleico (ARN) complementario se puede transcribir o fabricar a partir del ADN. En una realización preferida, la transfección del ARN se lleva a cabo empleando un cebador con una región promotora de la ARN polimerasa unida a la secuencia del cebador estándar para el ADN de interés en la reacción de amplificación (etapa tres). El ARN complementario al ADN se transcribe entonces a partir de la región promotora adjunta. En un método alternativo, el ADN del alelo amplificado se clona en un vector de expresión, y se transcribe el ARN complementario al ADN. Además, como una realización preferida opcional, el ARN complementario se usa en una reacción de traducción *in vitro* para fabricar proteína asociada a tumor o específica del tumor.

La caracterización del alelo, la amplificación del ADN derivado de tumor o asociado a tumor, y la caracterización, transcripción del ARN complementario, y traducción a proteína asociada a tumor o específica de tumor proporciona utilidad significativa, tanto la asignación de la terapia como en el desarrollo de terapias específicas de tumor. La secuenciación de ADN extracelular o la transcripción de ARN complementario permite la asignación o desarrollo de compuestos antisentido, incluyendo oligonucleótidos sintéticos y otros constructos antisentido apropiadamente específicos para el ADN extracelular, tal como mediante la construcción de un plásmido de expresión tal como adaptando el método de Aoki et al. (1995, *Cancer Res.* 55: 3810-3816). De forma similar, la definición de las características tumorales permite la asignación de terapias de vacunas o de anticuerpos monoclonales específicas apropiadamente específicas para el ADN amplificado. La producción de proteína inmunológica correspondiente se puede usar en el desarrollo de anticuerpos monoclonales específicos de tumor. De forma similar, la proteína traducida se puede usar en el desarrollo de una vacuna específica de tumor.

Especialmente importante, la invención permite el desarrollo y aplicación de estas terapias o diagnósticos específicos de tumor incluso cuando están presentes solamente tumores premalignos, cánceres tempranos, o cánceres ocultos. De este modo, la invención permite la intervención terapéutica cuando la carga tumoral es baja, la función inmunológica está relativamente intacta, y el paciente no está comprometido, incrementando todo ello el potencial para la cura.

Procesamiento posterior de secuencias enriquecidas

La combinación de los presentes métodos con MALDI-TOF, fusión de alta resolución o secuenciación de una sola molécula abordaría 3 necesidades distintas en la detección de mutaciones: la detección rápida de mutaciones somáticas conocidas o sospechosas de estar correlacionadas con el resultado clínico (MALDI-TOF); el escaneado rápido de muestras de pacientes individuales en busca de mutaciones somáticas desconocidas (fusión de HR), seguido de la secuenciación selectiva de unos pocos exones; y la secuenciación masivamente paralela (secuenciación de una sola molécula, SMS) de múltiples genes en "muestras difíciles", es decir, en muestras procedentes de tumores con heterogeneidad, contaminación estrómicca o fluidos corporales en los que las mutaciones clínicamente relevantes pueden estar al nivel de 0,5-5%.

Espectrometría de masas

En una realización, una secuencia diana enriquecida se somete a secuenciación mediante MALDI-TOF. La espectrometría de masas (MS) ha surgido como una herramienta poderosa en la secuenciación del ADN. Los espectrómetros de masas producen una medida directa de la masa, que se puede adquirir en segundo o minutos en el intervalo de femtomolar a picomolar. La MS con analizador de tiempo de vuelo (TOF) y desorción/ionización mediante láser asistida por matriz (MALDI) se ha usado con éxito para la secuenciación rápida del ADN y la determinación eficiente de los tamaños de moléculas de ADN. El advenimiento de la MS de MALDI-TOF ha hecho más fácil ionizar moléculas de ADN grandes intactas y medir sus relaciones masa a carga. Productos monocatenarios y bicatenarios de la reacción en cadena de polimerasa (PCR®) de 500 nucleótidos (nt) de longitud se han detectado mediante MS de MALDI-TOF. Usando combinaciones optimizadas de matriz-láser que reducen la fragmentación del ADN, se han dado a conocer espectros de masas mediante MALDI con infrarrojos de ADN sintético, fragmentos de enzimas de restricción de ADN plasmídico, y transcritos de ARN de hasta un tamaño de 2180 nt con una exactitud de  $\pm 0,5-1\%$ . Aunque se han detectado oligómeros grandes mediante MS de MALDI-TOF, generalmente se acepta que actualmente es habitual hasta un 100-mero.

Para varios genes, las mutaciones clínicamente importantes no se producen aleatoriamente en el genoma (por ejemplo, mutaciones puntuales de ganancia de función que se producen en el oncogén más conocido). En su lugar, los cambios que afectan a un número relativamente pequeño de codones dan cuenta a menudo de la mayoría de las mutaciones somáticas. En principio, entonces, un número limitado de ensayos genéticos juiciosamente diseñados debería de interrogar eficazmente una gran proporción de mutaciones clínicamente relevantes. Por ejemplo, Garraway y colegas (Thomas, R.K., et al. (2007) *Nat Genet.* 39, 347-351) mostraron que 16-44 ensayos de MALDI-TOF por gen en *RAS*, *EGFR* y *BRAF* capturaron el 90%-99% del predominio mutacional observado hasta ahora para estos genes en neoplasias humanas. Por lo tanto, se propone que el genotipado de alto rendimiento puede proporcionar un medio eficaz para detectar mutaciones cancerosas críticas y/o "dianizables" a gran escala en muestras clínicas. MALDI-TOF es idealmente adecuado para la detección de mutaciones previamente identificadas en muestras tumorales no heterogéneas. Sin embargo, aunque no hay dudas sobre la fiabilidad de MALDI-TOF para la identificación de mutaciones de línea germinal o de SNP, la experiencia a la hora de detectar mutaciones somáticas es relativamente reciente. De este modo, la fiabilidad de MALDI-TOF disminuye sustancialmente cuando

se usan muestras heterogéneas con <10% de células mutadas (por ejemplo, cáncer pancreático, de pulmón o de próstata), o cuando se va a identificar ADN procedente de fluidos corporales. Al mejorar la sensibilidad, el presente método de enriquecimiento permite que MALDI-TOF detecte mutaciones somáticas de bajo nivel, y también proporcionará la fiabilidad requerida que es necesaria para la identificación de muestras tumorales quirúrgicas convencionales.

#### Fusión de alta resolución

En otra realización, una secuencia diana enriquecida se somete a fusión de alta resolución. Los genes que contienen mutaciones clínicamente relevantes en numerosas posiciones a lo largo de los exones, tales como p53, se identifican de forma más fácil vía el escaneado mutacional que mediante el genotipado de mutaciones individuales. La fusión de HR es una tecnología de escaneado mutacional de alto rendimiento introducida en los últimos años recientes, con excelentes capacidades para descubrir SNPs o mutaciones de línea germinal (Chou, L.S., et al. (2005); Am J Clin Pathol, 124, 330-338; Wittwer, C.T., et al. (2003); Clin Chem, 49, 853-860; Reed, G.H. y Wittwer, C.T. (2004); Clin Chem, 50, 1748-1754).

Inmediatamente tras la amplificación mediante PCR de una región genómica de interés en presencia de un colorante fluorescente intercalante (LC Green o Sybr-Green), la presencia de una mutación se identifica en tiempo real mediante el análisis cuidadoso de la curva de fusión y mediante comparación con el tipo salvaje sin ningún tratamiento post-amplificación. Además, el análisis de fusión de alta resolución logra el escaneado génico y el genotipado mutacional (es decir, identificación de SNPs) simultáneos en una fracción del tiempo requerido cuando se usan métodos tradicionales, a la vez que mantiene un entorno de tubo cerrado. La PCR necesita <30 min. (capilares) o 1,5 h (placas de 96 o de 384 pocillos), y la adquisición por fusión tarda 1-2 min. por capilar o 5 min. por placa.

Sin embargo, al igual que con MALDI-TOF, las ventajas de usar fusión de HR no se pueden aprovechar para mutaciones somáticas por debajo de aproximadamente ~20% de relaciones de mutante a tipo salvaje; de ese modo, varias clases de muestras clínicas no se pueden identificar de forma fiable vía la fusión de HR. Al incrementar los límites de detección, la presente invención permite que la conveniencia y el rendimiento de la fusión de HR se apliquen a la identificación de muestras tumorales quirúrgicas convencionales y también para detectar mutaciones somáticas de bajo nivel en muestras clínicas "difíciles" con contaminación estrómicica o ADN procedente de fluidos corporales.

#### Secuenciación de una sola molécula

La secuencia diana enriquecida se puede someter a secuenciación de una sola molécula (Thomas, R.K., et al. (2006); Nat Med, 12, 852-855). Las capacidades de la secuenciación de una sola molécula también se beneficiarían de la incorporación de la presente invención. Por ejemplo, para la identificación de mutaciones en muestras de pacientes, todavía es necesaria la PCR de los exones seleccionados en una máquina de PCR normal a partir de ADN genómico antes de iniciar la secuenciación de la segunda generación. Además, la detección de mutaciones a nivel de una relación de mutante a tipo salvaje de 1-5% en muestras clínicas requiere la secuenciación repetida de numerosos "sucesos individuales" a fin de lograr estadísticas aceptables. Esto reduce básicamente las capacidades de rendimiento, y, para mutaciones al nivel de 1%, solo se pueden identificar simultáneamente 10-20 secuencias, contrariamente a las ~4.000 secuencias si las mutaciones fuesen predominantes (por 454 Life Sciences, Technical Service). Al llevar a cabo la presente invención antes de la secuenciación de una sola molécula, aumentará el predominio de mutaciones en 1-2 órdenes de magnitud como una fracción del número global de alelos, incrementando de ese modo el rendimiento de la secuenciación de una sola molécula hasta un grado equivalente.

El método de enriquecimiento selectivo de la invención se puede aplicar durante la etapa en emulsión de una reacción de secuenciación de una sola molécula. Las secuencias diana enriquecidas se someten entonces a pirosecuenciación.

#### Alargamiento del cebador

La secuencia diana enriquecida se puede someter a una reacción de secuenciación de alargamiento del cebador. En el alargamiento del cebador, se usan oligonucleótidos para evaluar la variación en secuencia en una posición predeterminada de la misma con respecto a un ácido nucleico, cuya secuencia es conocida. Se proporciona un oligonucleótido de muestra como una molécula monocatenaria, la molécula monocatenaria se mezcla con un agente inductor, un nucleótido marcado, y un cebador que tiene una secuencia idéntica a una región que flanquea la posición predeterminada para formar una mezcla, teniendo la mezcla una ausencia esencial de nucleótidos constituidos por bases distintas de la base de la que está constituido el nucleótido marcado. La mezcla se somete entonces a condiciones que conduzcan a la hibridación del cebador a la molécula monocatenaria y la formación de un producto de alargamiento del cebador que incorpora el nucleótido marcado, y la mezcla se analiza en busca de la presencia de producto de alargamiento del cebador que contiene nucleótido marcado (patente U.S. nº 5.846.710).

#### Tampones

Se contempla en el método de amplificación de la actual invención la inclusión de disolventes orgánicos para incrementar las diferencias de  $T_m$  entre dúplex mal apareados y apareados. La inclusión de ciertos disolventes orgánicos puede mejorar el enriquecimiento de la secuencia diana. Por ejemplo, la inclusión de disolventes orgánicos puede incrementar la diferencia de temperatura de desnaturalización entre las secuencias de ADN de referencia y diana, y de este modo puede ayudar a la amplificación preferente de la secuencia diana. Los disolventes orgánicos tales como DMSO, formamida, betaína o glicerol (Pomp, D. y Medrano, J.F., *Biotechniques*, 10, 58-59 (1991)) pueden incrementar la diferencia de  $T_m$  entre secuencias apareadas (referencia/referencia) y mal apareadas (diana/referencia). En consecuencia, puesto que la etapa de hibridación intermedia (hibridación cruzada) en el método de la presente invención forma secuencias diana-referencia que contienen mal apareamientos, es ventajosa la inclusión de disolventes orgánicos hasta el grado de que no inhiban la acción de la polimerasa. Por lo tanto, en algunas realizaciones, la mezcla de reacción se suplementa con DMSO, formamida, betaína, glicerol, o una combinación de los mismos, a niveles de 1-10% volumen a volumen, o preferiblemente 3-8% volumen a volumen, o lo más preferible 5-6% volumen a volumen. El Ejemplo 8 ilustra el uso de DMSO en un método de enriquecimiento.

Una ventaja práctica de usar disolventes orgánicos es que la  $T_c$  que es apropiada para una secuencia dada cambia al usar un disolvente orgánico en la reacción. De este modo, mientras que en ausencia de DMSO la  $T_c$  para una secuencia es 83,5°C, en presencia de DMSO la  $T_c$  es 80,5°C. Como resultado, al añadir una cantidad diferente de DMSO u otro disolvente a diferentes secuencias, se puede asegurar que la  $T_c$  es la misma para todas las secuencias. Esto es útil para llevar a cabo numerosas reacciones de enriquecimiento para una variedad de secuencias en un solo experimento con una máquina de PCR, puesto que la temperatura de desnaturalización es entonces la misma para todas.

## EJEMPLOS

### Ejemplo 1. Materiales y métodos para enriquecer una secuencia diana

Secuencias usadas para la validación de PCR COLD: Para validar la presente invención, se usó una serie de ADN genómico y estirpes celulares que contienen mutaciones en diferentes posiciones del exón 8 de p53 y del exón 2 de Kras (codones 12-13) (Figura 3). Las mutaciones del exón 8 de p53 se correlacionan con un mal pronóstico en cáncer de pulmón, y son mutaciones de bajo predominio en el plasma de pacientes con cáncer. De forma similar, las mutaciones de Kras tienen significancia de pronóstico en adenocarcinoma pulmonar.

Protocolo de enriquecimiento y cebadores: La PCR se llevó a cabo en presencia de 0,1 x de agente intercalante LC-Green y fue seguida en tiempo real en una máquina Cepheid. El seguimiento en tiempo real de la PCR no es necesario pero es conveniente, de manera que se adoptó para todos los experimentos.

Para la secuencia de p53 de 167 pb, en primer lugar se llevó a cabo la PCR normal durante 10 ciclos a fin de generar suficiente producto para uso en el protocolo de enriquecimiento. Una máquina Cepheid se programó con los siguientes parámetros de ciclación:

95°C, 120 s; (95°C, 15 s/55°C lectura de fluorescencia ON, 30 s/72°C, 1 min. de alargamiento) x 10 ciclos.

El producto de la PCR resultante se diluyó entonces 1:1000 y se sometió al protocolo de enriquecimiento a continuación, que también se ilustra en la Figura 1:

95°C, 15 s; 70°C durante 120 s; desnaturalización a una  $T_c = 86,5^\circ\text{C}$  durante 3 s; 55°C lectura de fluorescencia ON durante 30 s; después 72°C, 1 min. de alargamiento durante 30 ciclos.

A fin de preparar el producto de PCR para la didesoxisequenciación de Sanger, el producto se trató con exonucleasa I y fosfatasa alcalina de gamba. Se usaron los siguientes cebadores en el método de secuenciación:

fragmento de 167 pb: 5' - GCT TCT CTT TTC CTA TCC TG - 3' directo;

5' - CTT ACC TCG CTT AGT GCT - 3' inverso

Para los fragmentos de p53 de 87 pb y 210 pb y los fragmentos de Kras de 135 pb, el protocolo de enriquecimiento fue como se describe anteriormente, pero las temperaturas de desnaturalización críticas se fijaron a  $T_c = 83,5, 87,5,$  y  $80^\circ\text{C}$ , respectivamente. Los cebadores usados en la reacción de secuenciación fueron:

5' - TGG TAA TCT ACT GGG ACG-3' directo;

5' CGG AGA TTC TCT TCC TCT - 3' inverso (fragmento del exón 8 de p53 de 87 pb)

5' - GCT TCT CTT TTC CTA TCC TG - 3' directo;

5' - TAA CTG CAC CCT TGG TC - 3' inverso (fragmento del exón 8 de p53 de 210 pb)

5'-AACTTGTTGGTAGTTGGACCT-3' directo;

5'-CTCTATTGTTGGATCATATT-3' inverso (fragmento del exón 2 de Kras).

La reproducibilidad de todos los protocolos de enriquecimiento se evaluó en 3-6 experimentos independientes.

#### Resultados

5 Mutaciones del exón 8 de p53: Cuando el protocolo de enriquecimiento que utiliza una temperatura de desnaturalización crítica  $T_c = 86,5^\circ\text{C}$  se aplicó para el fragmento del exón 8 de 167 pb, el enriquecimiento fue evidente para todas las mutaciones ensayadas. La Figura 4 representa los resultados representativos. Por ejemplo, el ADN procedente de células HCC diluido inicialmente en células de tipo salvaje hasta una relación de mutante a tipo salvaje de 5% se convierte en una relación de mutante a tipo salvaje de ~70% tras el protocolo de enriquecimiento, según se estima observando el cromatograma de secuenciación (es decir, enriquecimiento por un factor de ~14). De forma similar, el ADN procedente de células SW480 (mutación G>A homocigota en el codón 273) diluido por un factor de 10 en tipo salvaje se enriquece por un factor de ~7 tras el procedimiento de enriquecimiento. También se observó el enriquecimiento por un factor de ~12 para la muestra de CT7 (mutación C>A heterocigota) y por un factor de 6 para la muestra de MDA-MB231 (mutación C>T heterocigota). La muestra de p53 de tipo salvaje, amplificada vía el método de enriquecimiento, no indicó ninguna mutación (Figura 4). Para todas las mutaciones de p53 que se estudiaron para el fragmento de 167 pb, como se da a conocer en la Figura 2, el enriquecimiento varió en 5-14 veces. De este modo, el procedimiento de enriquecimiento incrementó el predominio de todas las secuencias que contienen mutaciones, independientemente de dónde se encuentra la mutación.

20 Mutaciones del codón 12/13 de Kras: La Figura 5 ilustra los resultados del fragmento de 135 pb procedente de Kras. Los resultados se compararon con una PCR anidada normal realizada a una temperatura de desnaturalización de  $94^\circ\text{C}$  y seguida de una secuenciación de Sanger. La Figura 5 ilustra que las mutaciones hasta 33% de mutante a tipo salvaje se pueden detectar claramente usando secuenciación de Sanger.

#### Ejemplo 2: Secuenciación de Sanger de muestras clínicas

25 Para aplicar la presente invención para el análisis de muestras clínicas, 20 muestras clínicas de tumor de colon y de adenocarcinoma de pulmón que se secuenciaron previamente tras la PCR normal se sometieron al protocolo de enriquecimiento y a secuenciación de Sanger como se describe en el Ejemplo 1. Los resultados indicaron que todas las mutaciones identificadas vía PCR normal-secuenciación de Sanger también se identificaron vía el procedimiento de enriquecimiento seguido de secuenciación. Sin embargo, el procedimiento de enriquecimiento también identificó mutaciones que se perdieron por la secuenciación normal. La Figura 6 representa 2 muestras clínicas, TL64 y CT20, en las que se detectaron mutaciones G>A de bajo predominio vía el protocolo de enriquecimiento-secuenciación de Sanger del exón 8 de p53, codón 273, pero que no se detectaron mediante secuenciación tras PCR normal. Se llevó a cabo una verificación independiente de la presencia de mutaciones usando secuenciación a base de RFLP.

30 Además, la mutación del exón 8 de p53 (G>A) se detectó en ADN circulante en plasma procedente de 5 pacientes con cáncer de colon vía el procedimiento de enriquecimiento-secuenciación de Sanger, pero no vía PCR normal-secuenciación de Sanger. A continuación, las mutaciones de p53 (C>T) que se perdieron mediante secuenciación normal también se revelaron usando ADN obtenido de una muestra fijada en formalina (FFPE) obtenida de un paciente con cáncer de pulmón no microcítico (NSCLC) (Figura 6). El cromatograma inferior en la Figura 6 demuestra la detección de mutaciones del codón 12 de Kras en otra muestra de FFPE obtenida de un paciente con NSCLC. Las mutaciones identificadas vía el procedimiento de enriquecimiento se verificaron subsiguientemente de forma independiente a partir de ADN genómico vía métodos de RFLP. De este modo, las mutaciones relevantes perdidas por PCR normal-secuenciación son fácilmente detectables usando PCR COLD-secuenciación.

#### Ejemplo 3: Mutaciones que disminuyen la $T_m$ se pueden enriquecer sin la etapa de hibridación de mal apareamiento

45 La dependencia de la PCR sobre la secuencia nucleotídica es tan pronunciada cuando la temperatura de desnaturalización se fija a la temperatura crítica ( $T_c$ ) que incluso sin formar un mal apareamiento durante la PCR existe enriquecimiento de aquellas mutaciones que disminuyen la  $T_m$ . De este modo, cuando está presente un alelo G y un alelo A, el alelo A se enriquecerá durante la PCR COLD a medida que ésta disminuye la  $T_m$  del alelo. Para demostrar este hecho, y también para examinar la dependencia del enriquecimiento con el tamaño de la secuencia examinada, se examinó un fragmento de 87 pb y un fragmento de 210 pb que contiene las mismas mutaciones que el fragmento de 167 pb del exón 8 de p53 (Figura 3). Al igual que con el fragmento de 167 pb, estos dos fragmentos se amplificaron a partir del amplicón del exón 8 de p53 inicial vía PCR anidada seguida del protocolo de enriquecimiento. Sin embargo, en este caso se usó una versión truncada del protocolo de amplificación del Ejemplo 1, pero la etapa de formación de mal apareamiento a  $70^\circ\text{C}$  así como la etapa de  $94^\circ\text{C}$  se omitieron ambas (temperaturas de desnaturalización críticas  $T_c = 83,5^\circ\text{C}$  para el fragmento de 87 pb y  $T_c = 87,5^\circ\text{C}$  para el fragmento de 210 pb). Por lo tanto, en esta versión del protocolo de enriquecimiento, la PCR cicla solamente entre la temperatura de desnaturalización crítica ( $T_c$ ), la etapa de unión del cebador (por ejemplo  $55^\circ\text{C}$ ) y la etapa de síntesis del cebador (por ejemplo  $72^\circ\text{C}$ ).

La Figura 7A demuestra cromatogramas de secuenciación para el fragmento de 87 pb (la secuenciación directa o la inversa se llevaron a cabo dependiendo de si la mutación se encuentra en la secuencia de 87 pb). Los datos indican que, usando esta versión del protocolo de enriquecimiento modificado, el fragmento de 87 pb se enriquece 20-50 veces. Por ejemplo, una dilución inicial del 1% de ADN mutante a tipo salvaje de ADN de SW480 dio como resultado 50% de mutante a tipo salvaje tras el enriquecimiento, es decir, un enriquecimiento de ~50 veces. La secuenciación subsiguiente de Sanger reveló una secuencia "heterocigota". El efecto del tamaño sobre el enriquecimiento vía el protocolo de enriquecimiento abreviado se representa en la Figura 7B. Este dato ilustra que el enriquecimiento es más elevado para fragmentos <100 pb, pero todavía claramente evidente (~8-10 veces) para fragmentos de hasta 210 pb.

#### 10 **Ejemplo 4: Mutaciones que incrementan o disminuyen la $T_m$ se pueden enriquecer vía el método de enriquecimiento completo**

Aunque la mayoría (~70%) de las mutaciones encontradas en diversas muestras de cáncer disminuyen la  $T_m$ , ~15% de las mutaciones incrementan la  $T_m$  (por ejemplo A>G), mientras que ~15% mantienen la  $T_m$  (por ejemplo G>C). A fin de ser capaces de enriquecer todas las mutaciones posibles, incluyendo tanto mutaciones G>A y A>G como supresiones, se prefiere el programa de enriquecimiento completo (Figura 1). Para demostrar la capacidad de enriquecer secuencias diana que tienen mutaciones que incrementan o disminuyen la  $T_m$ , se amplificaron fragmentos del exón 8 de p53 de 167 pb con el nucleótido C o T (estirpe celular salvaje frente a estirpe celular HCC) vía el protocolo de enriquecimiento de la Fig. 1.

Se formaron dos mezclas, una con el alelo C en minoría (C:T 1:10) y otra con el alelo T en minoría (T:C 1:10). Tras el procedimiento de enriquecimiento, o alternativamente la PCR normal, los productos se secuenciaron. En ambos casos, se enriqueció el alelo minoritario, es decir, el C o el T dependiendo de cuál estaba más diluido antes de la amplificación. Presumiblemente, la secuencia mal apareada tiene una temperatura de fusión menor que el alelo C o el alelo T; por lo tanto, al llevar a cabo el protocolo de la Figura 1, las secuencias mal apareadas siempre se desnaturalizan preferentemente. Por lo tanto, al formar un mal apareamiento a una temperatura intermedia (~70°C) durante el protocolo de enriquecimiento, siempre hay enriquecimiento del alelo minoritario incluso si el cambio nucleotídico específico tiende a incrementar la  $T_m$  local.

#### **Ejemplo 5: Secuenciación de MALDI-TOF**

También se espera que la presente invención mejore la mayoría de las otras tecnologías a base de PCR, incluyendo MALDI-TOF para la detección de mutaciones somáticas. Para demostrar este punto, se usó el mismo modelo aplicado en el Ejemplo 1, para la secuenciación de Sanger para identificar mutaciones del exón 8 de p53 específicas usando diluciones en serie de estirpes celulares que contienen mutaciones en muestras de tipo salvaje, para comparar el protocolo de enriquecimiento frente a la PCR normal seguida de MALDI-TOF.

Tras el protocolo de enriquecimiento o la PCR normal, los dNTPs en exceso se eliminaron de la reacción mediante incubación con 0,3 U de fosfatasa alcalina de gamba (USB) a 37°C durante 20 minutos, seguido de una incubación durante 5 minutos a 85°C para desactivar la enzima. El alargamiento del cebador individual con respecto al SNP o la inserción/supresión se llevaron a cabo a una concentración final de: 600 nM de cada cebador de alargamiento, 50 uM de d/ddNTP y 0,126 U de Thermosequenase (Solis Biodyne), y se incubaron a 94°C durante 2 minutos seguido de 45 ciclos de 94°C durante 5 segundos, 52°C durante 5 segundos, y 72°C durante 5 segundos. Los cebadores de alargamiento usados fueron diseñados por la MALDI-TOF Harvard Core Facility para cada mutación de p53 estudiada, usando el software MassArray Assay Design versión 3.1.2.2. Los cebadores usados para cada mutación fueron:

p53\_sw480: CAGGACAGGCACAAACA;

p53\_CT7: AGGACAGGCACAAACAC;

p53\_DU145: ACAGCTTTGAGGTGCGT;

45 KRAS\_SW480: TGTGGTAGTTGGACCTG;

KRAS\_A549: ACTCTTGCTACGCCAC.

La reacción se desaló entonces mediante adición de una resina de intercambio catiónico, seguido del mezclado y centrifugación para sedimentar los contenidos del tubo. El producto del alargamiento se colocó sobre un spectroCHIP de 384 pocillos antes de llevarlo al espectrómetro de masas de MALDI-TOF (Sequenom).

50 Los resultados se ilustran en la Tabla I. El factor de enriquecimiento de la mutación se da en la tercera columna de la Tabla I. El enriquecimiento se calcula comparando los valores obtenidos cuando se aplica PCR normal con MALDI-TOF. Para la mayoría de las mutaciones estudiadas, se obtienen factores de enriquecimiento de 10-60. El

## ES 2 547 053 T3

enriquecimiento aumenta a medida que disminuye la relación de mutante a tipo salvaje, indicando una dependencia no lineal del factor de enriquecimiento con la concentración inicial de mutaciones.

Tabla I

Mutaciones de p53. Se indica el % de alelo mutante o de tipo salvaje para diversas mutaciones del exón 8 de p53			
	RESULTADO DE MALDI-TOF		
mutación del exón 8 de p53=14484G>A			PCR DE ENRIQUECIMIENTO
DILUCIONES DE ESTIRPES CELULARES IDENTIFICADAS	%A mutante	%G tipo salv.	ENRIQUECIMIENTO
Relación SW480-tipo salv. 1:5 PCR normal	15	85	
Relación SW480-tipo salv. 1:10 PCR normal	5 (límite del espec. masas)	95	
Relación SW480-tipo salv. 1:33 PCR normal	0	100	
Relación SW480-tipo salv. 1:10 PCR CO.L.D	45	55	~9
Relación SW480-tipo salv. 1:100 PCR CO.L.D	33	67	>10
Relación SW480-tipo salv. 1:300 PCR CO.L.D	31	69	>30
TIPO SALVAJE SOLAMENTE, PCR CO.L.D	0	100	Ninguno
	RESULTADO DE MALDI-TOF		
mutación del exón 8 de p53=14483C>T			PCR CO.L.D
DILUCIONES DE ESTIRPES CELULARES IDENTIFICADAS	%T mutante	%C tipo salv.	ENRIQUECIMIENTO
Relación CT7-tipo salv. 1:5 PCR normal	28	72	N/A
Relación CT7-tipo salv. 1:10 PCR normal	18	82	N/A
Relación CT7-tipo salv. 1:33 PCR normal	5 (límite del espec. masas)	95	N/A
Relación CT7-tipo salv. 1:100 PCR CO.L.D	45	54	~18
Relación CT7-tipo salv. 1:200 PCR CO.L.D	28	72	~40
Relación CT7-tipo salv. 1:300 PCR CO.L.D	27	73	~60
TIPO SALVAJE SOLAMENTE, PCR CO.L.D	0	100	Ninguno
	RESULTADO DE MALDI-TOF		
mutación del exón 8 de p53=14486G>T			PCR CO.L.D
DILUCIONES DE ESTIRPES CELULARES IDENTIFICADAS	%G mutante	%T tipo salv.	ENRIQUECIMIENTO
Relación DU145-tipo salv. 1:5 PCR normal	28	72	N/A
Relación DU145-tipo salv. 1:10 PCR normal	18	82	N/A
Relación DU145-tipo salv. 1:33 PCR normal	0	100	N/A
Relación DU145-tipo salv. 1:33 PCR CO.L.D	27	73	~7

ES 2 547 053 T3

Relación DU145-tipo salv. 1:100 PCR CO.L.D	12	88	>10
Relación DU145-tipo salv. 1:300 PCR CO.L.D	0	100	indetectable
TIPO SALVAJE SOLAMENTE, PCR CO.L.D	0	100	Ninguno
Mutaciones Kras			
Se indica el % de alelo mutante o de tipo salvaje para mutaciones del codón 12 de Kras			
	RESULTADO DE MALDI-TOF		
mutación del codón 12 de Kras=GGT>AGT			PCR CO.L.D
DILUCIONES DE ESTIRPES CELULARES IDENTIFICADAS	%A mutante	%G tipo salv.	ENRIQUECIMIENTO
Relación A549-tipo salv. 1:10 PCR normal	25	75	N/A
Relación A549-tipo salv. 1:33 PCR CO.L.D	40	60	~5,5
Relación A549-tipo salv. 1:100 PCR CO.L.D	25	75	~10
Relación A549-tipo salv. 1:200 PCR CO.L.D	12	88	~10
mutación del codón 12 de Kras=GGT>GTT			
DILUCIONES DE ESTIRPES CELULARES IDENTIFICADAS	%A mutante	%G tipo salv.	ENRIQUECIMIENTO
Relación SW480-tipo salv. 1:10 PCR normal	15	85	N/A
Relación SW480-tipo salv. 1:33 PCR CO.L.D	15	85	~3
Relación SW480-tipo salv. 1:100 PCR CO.L.D	7	93	~5
Relación SW480-tipo salv. 1:200 PCR CO.L.D	5	95	~7
TIPO SALVAJE SOLAMENTE, PCR CO.L.D	0	100	Ninguno
Comparación del PCR de enriquecimiento-MALDI-TOF con PCR normal-MALDI-TOF.			
El enriquecimiento de mutaciones ganado se da en la última columna.			

**Ejemplo 6: Comparación de PCR en tiempo real normal basada en TAQMAN y de PCR en tiempo real basada en el enriquecimiento**

- 5 Se llevaron a cabo reacciones de amplificación de ácido nucleico para comparar ensayos de sondas TAQMAN en PCR en tiempo real normal y una PCR en tiempo real basada en el enriquecimiento. Para comparar los dos métodos de detección en tiempo real, tanto el ADN genómico como las muestras tumorales clínicas que tienen una mutación G>A en el exón 8 de p53 se ensayaron a diversas diluciones en la secuencia de tipo salvaje. Se realizaron diluciones en serie (1:3, 1:10, 1:30, 1:100 y 1:300) de ADN genómico procedente de SW480 en ADN de tipo salvaje.
- 10 Específicamente, las reacciones de PCR en tiempo real se llevaron a cabo directamente a partir de 20 ng de ADN genómico en presencia de 0,2 µM de sonda Taqman 5'-6-Fam-TTT GAG GTG CAT GTT TGT GCC-BHQ\_1-3' que tiene un apareamiento total con la secuencia que contiene la mutación p53 en ADN procedente de células SW480. Las concentraciones finales de otros reactivos fueron: 1X de tampón GoTaq Flexi (Promega), 1X GoTaq polymerase (Promega) 0,2 mM de cada dNTP, 0,2 µM de cebador directo, 5'-TGG TAA TCT ACT GGG ACG-3', 0,2 µM de cebador inverso, 5'-CGG AGA TTC TCT TCC TCT-3', MgCl<sub>2</sub> 3 mM, más ADN. El tamaño del amplicón de la PCR fue 87 pb y T<sub>c</sub> = 83,5°C. El ciclo de la PCR COLD rápido fue: 95°C, 120 s; (95°C, 15 s; 58°C lectura de fluorescencia ON, 60 s) x 25 ciclos; (83,5°C 15 s; 58°C lectura de fluorescencia ON, 60 s) x 25 ciclos. Para el ciclo de la PCR normal, se empleó el mismo programa, pero la temperatura de desnaturalización a lo largo de la PCR fue 95°C. Los experimentos se repitieron al menos 5 veces en experimentos independientes.
- 15

En la FIG. 8A y 8B se representan gráficas de amplificación que ilustran la sensibilidad de la PCR en tiempo real normal y enriquecida según se aplica a ADN genómico procedente de la estirpe celular de cáncer de colon SW480. La FIG. 8B demuestra que la PCR en tiempo real enriquecida puede detectar la presencia de la mutación a una relación de alelo mutante a alelo de tipo salvaje de 1:300. Por el contrario, la PCR en tiempo real normal realizada en condiciones idénticas, con la excepción de la etapa de T<sub>c</sub>, solamente puede detectar el mutante a dilución máxima de 1:10 (FIG. 8A). Por lo tanto, la sensibilidad del ensayo es 30 veces mejor usando el procedimiento de enriquecimiento.

En la FIG. 8C y 8D se ilustran gráficas de amplificación que comparan la sensibilidad de la PCR en tiempo real normal y enriquecida en muestras tumorales clínicas que tienen la mutación del exón 8 de p53 (una de las cuales se sabe que contiene una mutación de bajo nivel (5% de mutante a tipo salvaje) en el exón 8 de p53, CT20). La PCR en tiempo real enriquecida fue capaz fácilmente de detectar la mutación (FIG. 8D), mientras que la PCR normal no lo hizo (FIG. 8C). La muestra restante (TL6, TL8 y TL18), que se sabía que eran muestras de tipo salvaje, no se amplificaron (FIG. 8C) en las mismas condiciones.

#### **Ejemplo 7: Escaneado de mutaciones vía PCR en tiempo real con enriquecimiento**

Se crearon mezclas de reacción de amplificación de ácidos nucleicos para comparar la capacidad de la PCR en tiempo real normal y enriquecida que utiliza colorantes de detección de ADN para detectar muestras que contienen mutaciones en cualquier parte a lo largo del exón 8 de p53. El método proporciona un método rápido y conveniente para identificar mutaciones desconocidas o SNPs heterocigotos. De este modo, el presente método se puede adaptar por alguien de pericia normal en la técnica para escanear grandes números de genes en busca de mutaciones de línea germinal o somáticas para una variedad de diferentes aplicaciones (por ejemplo, escaneado en busca de mutaciones BRCA1/2 en poblaciones con riesgo elevado de desarrollar CA de mama/ovárico, escaneado de rutas genéticas completas para identificar mutaciones, etc).

La FIG. 9A y 9B ilustran gráficas de amplificación que comparan PCR en tiempo real tanto normal como enriquecida que usa el colorante LC-Green en una variedad de estirpes celulares y muestras clínicas que se sabe que contienen mutaciones en el exón 8 de p53. Los datos muestran que la PCR normal en tiempo real (FIG. 9A) es incapaz de distinguir entre las muestras mutantes (SW480, TL6 y CT20) y las de tipo salvaje (R27, TL8, TL18, TL81 y TL82), mientras que la PCR en tiempo real enriquecida sí lo hace (FIG. 9B). El método de enriquecimiento proporciona una detección umbral más temprana para las muestras que contienen mutaciones que para las muestras de tipo salvaje.

La FIG. 9C y 9D ilustran los resultados de reacciones de amplificación preparadas en condiciones de reacción idénticas a las ilustradas en la FIG. 9A y 9B, pero aquí las muestras son todas tumores pulmonares. Los datos muestran que la PCR normal en tiempo real (FIG. 9C) no puede diferenciar entre las muestras mutantes y las de tipo salvaje, mientras que la PCR enriquecida en tiempo real (FIG. 9D) sí lo hace. El método de enriquecimiento proporciona una detección umbral más temprana para las muestras que contienen mutaciones que para las muestras de tipo salvaje.

Además, el método de detección enriquecido fue capaz de identificar una mutación C>T previamente desconocida en la muestra TL6.

#### **Ejemplo 8: Los disolventes orgánicos incrementan el enriquecimiento de mutaciones durante PCR en tiempo real normal y potenciada**

Se prepararon mezclas de reacción de amplificación de ácidos nucleicos para evaluar el impacto de disolventes orgánicos sobre la PCR en tiempo real normal y enriquecida. Las reacciones se llevaron a cabo en presencia o en ausencia de un disolvente orgánico (3% de DMSO). El procedimiento usado fue el mismo que el descrito en el Ejemplo 7 y representado en la FIG. 9C y 9D, excepto por la adición de 3% de DMSO.

La presencia del disolvente orgánico potenció la discriminación entre las gráficas de amplificación de muestras que contienen las secuencias mutadas con aquellas que contienen secuencias de tipo salvaje (FIG. 10). Por ejemplo, la diferencia umbral entre muestras de tipo salvaje y muestras mutantes aumentó desde ~5 ciclos (enriquecimiento en tiempo real sin DMSO; véase la FIG. 9A) hasta más de 10 ciclos (enriquecimiento en tiempo real con DMSO; véase la FIG. 10A).

#### **Ejemplo 9: Detección de mutaciones de nivel ultrabajo usando RFLP en combinación con PCR de enriquecimiento**

La PCR de enriquecimiento combinada con RFLP-PCR se puede usar para la identificación mejorada de mutaciones de nivel ultrabajo, por ejemplo para identificar mutaciones aleatorias en un genoma canceroso, o mutaciones de resistencia en muestras de cáncer en una etapa muy temprana, es decir, antes del tratamiento.

Por ejemplo, muestras que contienen el exón 19 de EGFR de tipo salvaje se digirieron selectivamente con la enzima TaqI. Entonces, diluciones de hasta 1:10.000 de ADN mutante a genómico se sometieron a PCR en un formato de enriquecimiento (T<sub>c</sub> = 81,5 o 81°C) o en un formato de PCR normal (95°C). El enriquecimiento de la mutación se cuantificó mediante la digestión con TaqI seguido de dHPLC (FIG. 11). La cantidad de mutación presente se

cuantificó por la presencia de un pico de mutación separado a un tiempo de retención de alrededor de 7 min. Tras el procedimiento de enriquecimiento, el pico de mutación es mucho más evidente que tras la PCR normal. En conclusión, el procedimiento de enriquecimiento mejoró sustancialmente la detección de mutaciones de nivel muy bajo identificadas vía RFLP-PCR.

- 5 Aunque esta invención se ha mostrado y descrito particularmente con referencias a realizaciones preferidas de la misma, se entenderá por los expertos en la técnica que se pueden realizar en ella diversos cambios en forma y detalles sin separarse del alcance de la invención englobado por las reivindicaciones anejas.

## REIVINDICACIONES

1. Un método para enriquecer una secuencia diana en una mezcla de reacción, comprendiendo dicho método:
- 5 a. someter una mezcla de reacción sospechosa de tener un dúplex de secuencia diana y un dúplex de secuencia de referencia a una primera temperatura de desnaturalización que está por encima de la temperatura de fusión ( $T_m$ ) del dúplex de la secuencia de referencia y del dúplex de la secuencia diana, para permitir la desnaturalización de dicho dúplex de secuencia diana y de dicho dúplex de secuencia de referencia, en el que dicho dúplex de secuencia diana comprende una o más inserciones, supresiones o alteraciones y difiere en al menos un nucleótido de dicho dúplex de secuencia de referencia, y en el que dicha secuencia diana es al menos 50% homóloga a dicho dúplex de secuencia de referencia y es amplificable por el mismo par de cebadores que dicho dúplex de secuencia de referencia, y en el que la secuencia diana es menos predominante en la mezcla de reacción que la secuencia de referencia;
- 10 b. reducir la temperatura de la mezcla de reacción hasta una temperatura de hibridación para permitir la formación de dúplex de hebra diana/hebra de referencia y evitar la unión de dicho par de cebadores a las hebras diana y de referencia desnaturalizadas en la mezcla de reacción, en el que la temperatura de hibridación está por debajo de la temperatura crítica ( $T_c$ ) y por encima de la temperatura de hibridación del cebador;
- 15 c. incrementar dicha mezcla de reacción de amplificación hasta una temperatura crítica ( $T_c$ ) que está por debajo de la  $T_m$  de dicho dúplex de secuencia de referencia, para permitir la desnaturalización preferente de dichos dúplex de la etapa (b) para formar hebras diana y de referencia desnaturalizadas;
- 20 d. reducir la temperatura de la mezcla de reacción hasta una temperatura de hibridación del cebador para permitir que dicho par de cebadores se hibride a dichas hebras diana y de referencia; y
- e. alargar dicho par de cebadores para enriquecer dicha secuencia diana con respecto a dicha secuencia de referencia;
- en el que el método se repite durante dos o más ciclos.
- 25 2. El método de la reivindicación 1, en el que dichas secuencia diana y de referencia se amplifican en primer lugar sometiendo la mezcla de reacción a PCR.
3. El método de la reivindicación 1, en el que dicha secuencia diana es un alelo mutante que difiere de la secuencia de referencia en entre 1 y 10 nucleótidos.
4. El método de una cualquiera de las reivindicaciones 1 a 3, en el que:
- 30 (i) dichas secuencias diana y de referencia comprenden 25 a 500 bases, y/o
- (ii) dicha  $T_c$  está 0,3°C a 5°C por debajo de la  $T_m$  de dicha secuencia de referencia, y/o
- (iii) dicha  $T_c$  está por debajo de la  $T_m$  de la secuencia diana, y/o
- (iv) dicho método se repite entre 5 y 40 ciclos, y/o
- (v) dicho método se repite entre 10 y 30 ciclos.
- 35 5. El método de una cualquiera de las reivindicaciones 1 a 4, que comprende además la etapa de analizar dicha mezcla de reacción con secuencia diana enriquecida
- en el que dicho análisis usa opcionalmente uno o más de los métodos seleccionados del grupo que consiste en: MALDI-TOF, fusión de HR, secuenciación didesoxi, secuenciación de una sola molécula, pirosecuenciación, SSCP, RFLP, dHPLC, PCR digital y PCR cuantitativa.
- 40 6. El método de la reivindicación 1, en el que dicha  $T_c$  se aplica entre 1 segundo y 5 minutos.
7. El método de una cualquiera de las reivindicaciones anteriores, en el que dicha secuencia diana está metilada diferentemente de la secuencia de referencia, y, antes de implementar el método de la reivindicación 1 en la mezcla de reacción, la mezcla de reacción se trata opcionalmente con bisulfito de sodio.
8. El método de una cualquiera de las reivindicaciones anteriores, en el que dicha mezcla de reacción contiene un colorante de detección de ácido nucleico, y se lleva a cabo opcionalmente en un dispositivo de PCR en tiempo real.
- 45 9. El método de la reivindicación 1, llevado a cabo en condiciones de reacción en tiempo real utilizando un agente de detección de ácido nucleico.

10. El método de la reivindicación 1, en el que dicho método se usa para enriquecer dos o más secuencias diana diferentes, y dicho método comprende además uno o más pares adicionales de cebadores específicos para dichas secuencias diana.
- 5 11. El método de la reivindicación 1, en el que dicho par de cebadores tiene una temperatura de fusión que está por debajo de la temperatura aplicada en la etapa (b), opcionalmente al menos 5°C por debajo de la temperatura en la etapa (b).
12. El método de la reivindicación 1, en el que la mezcla de reacción incluye un ácido nucleico modificado.
13. Un método para enriquecer una secuencia diana como se cita en la reivindicación 1, que comprende además: repetir las etapas b) y c) más de una vez por ciclo antes de ejecutar la etapa d).
- 10 14. El método de la reivindicación 13, en el que la secuencia diana es al menos 70% homóloga a la secuencia de referencia, y preferiblemente al menos 80% homóloga a la secuencia de referencia.
15. Un medio legible por ordenador que comprende instrucciones de programa adaptadas para llevar a cabo el método de la reivindicación 1.

FIGURA 1

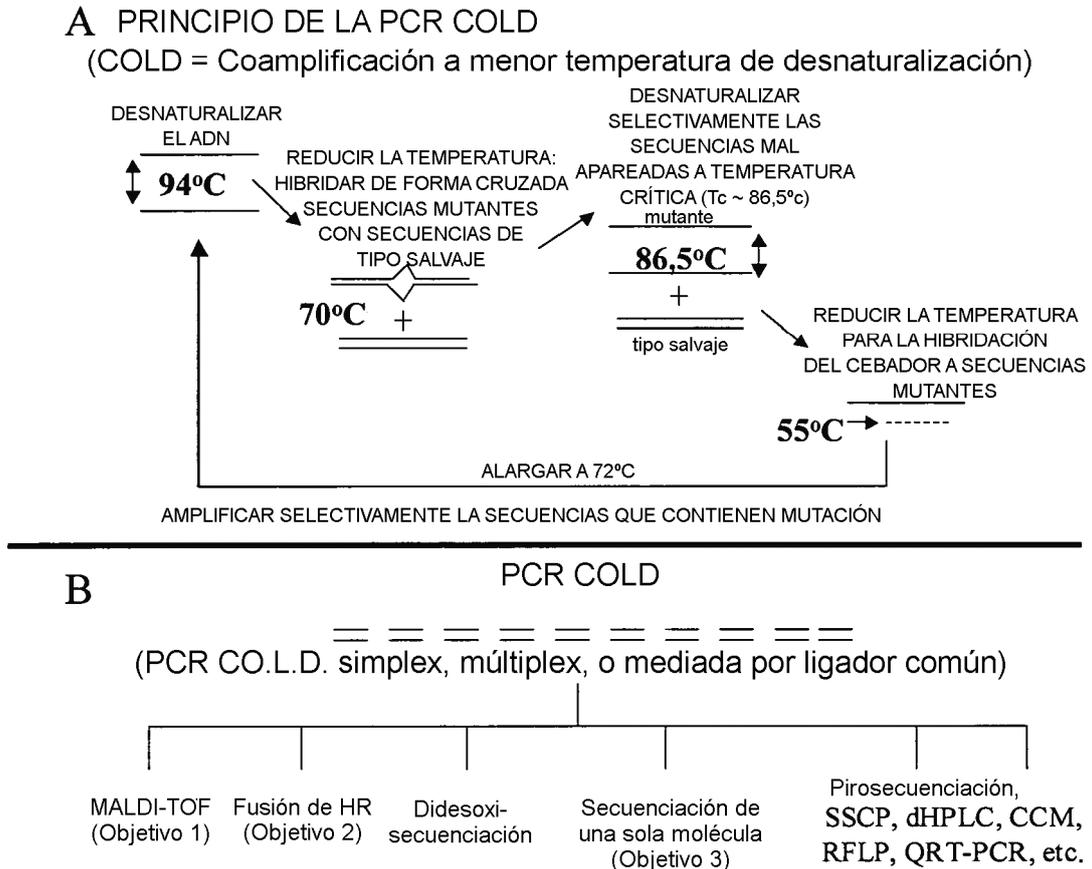


Figura 1. A. Ilustración del protocolo de enriquecimiento. Se representa un ejemplo para una secuencia del exón 8 de p53 de 167 pb. La formación de mal apareamientos en cualquier parte a lo largo de la secuencia durante la PCR permite la desnaturalización y amplificación preferentes de alelos (mutantes) minoritarios en cada ciclo de PCR. B. Sustitución de PCR por el método de enriquecimiento: Todos los ensayos de pruebas genéticas a base de PCR se beneficiarán mediante el enriquecimiento de mutaciones durante la etapa de PCR que los precede.

FIGURA 2

PRINCIPIO DE PCR COLD OSCILANTE

(se muestra un ejemplo de una secuencia del exón 8 de p53 de 87 pb con Tc=83,5)

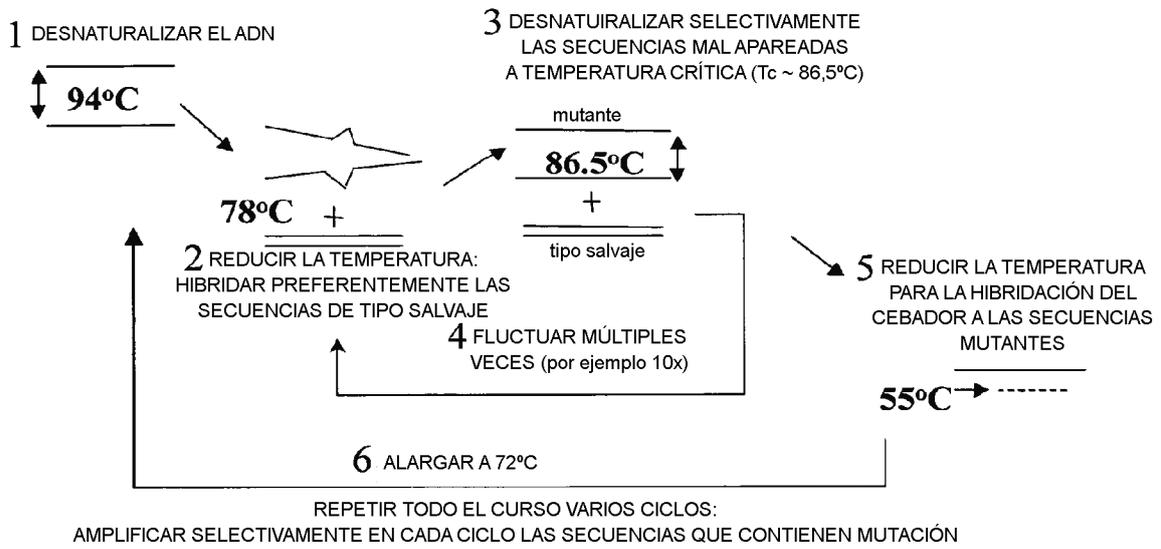
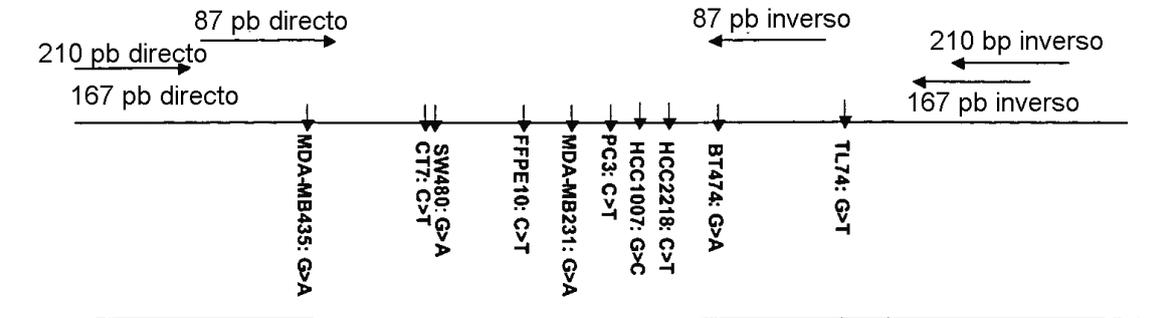


Figura 2: PRINCIPIO DE PROTOCOLO DE ENRIQUECIMIENTO OSCILANTE

FIGURA 3

mutaciones del exón 8 de p53 ensayadas para enriquecimiento vía PCR CO.L.D.



mutaciones de los codones 12/13 de Kras ensayadas para enriquecimiento vía PCR CO.L.D.

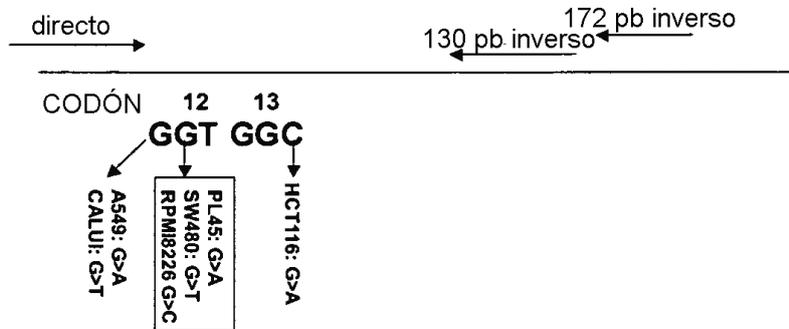


FIGURA 4

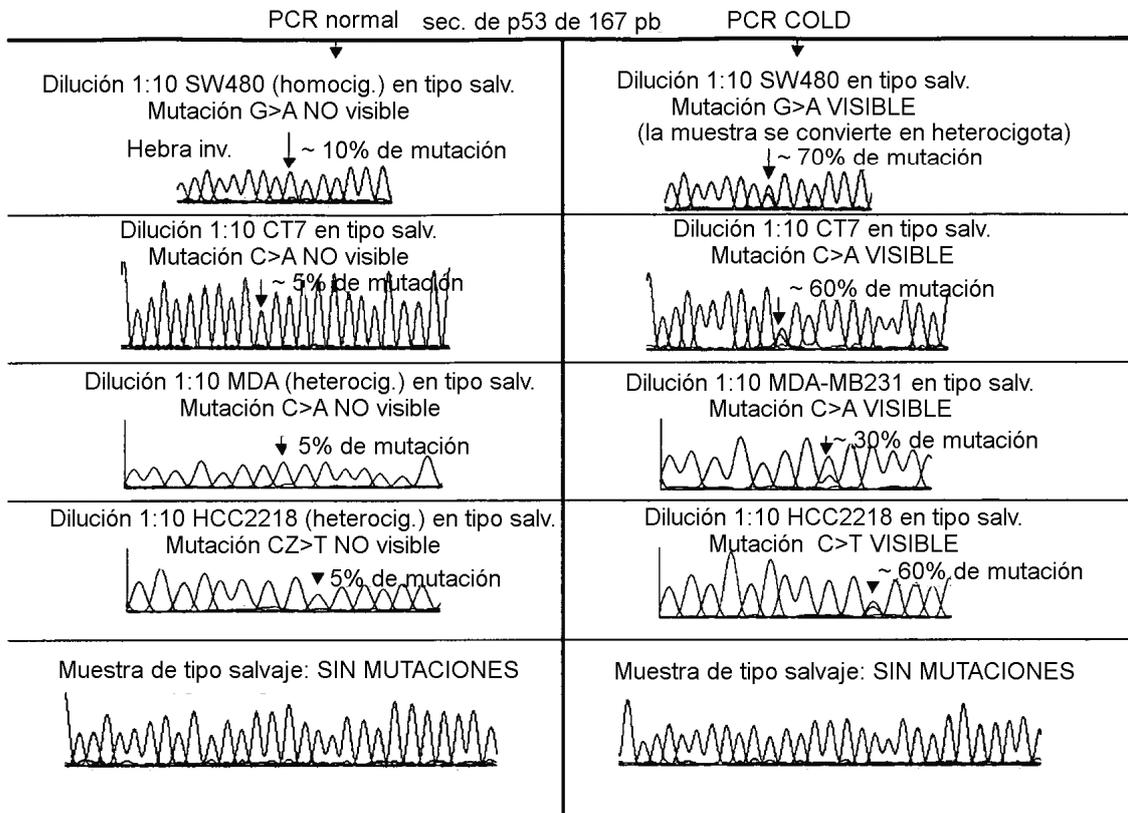


Figura 4: Comparación de la PCR de enriquecimiento con la PCR normal seguida de didesoxisecuenciación de Sanger. Diluciones de ADN procedente de estirpes celulares con mutaciones de p53 en tipo salvaje revelan el enriquecimiento de mutaciones en varias posiciones de la secuencia vía el procedimiento de enriquecimiento.

FIGURA 5

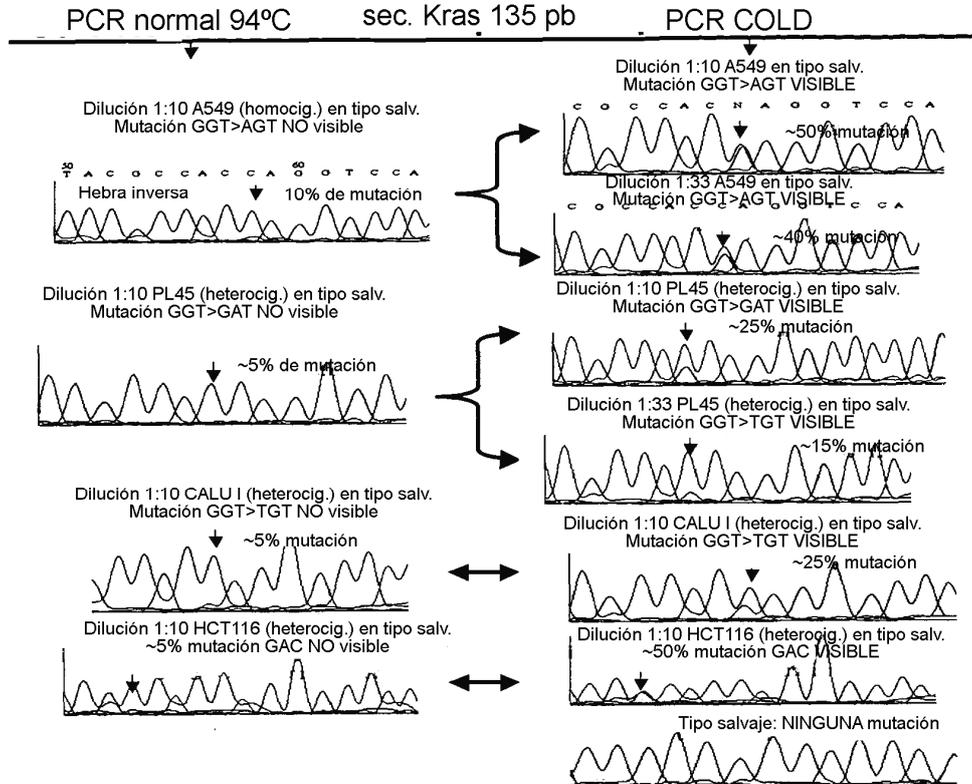


Figura 5: Comparación de PCR enriquecida con PCR normal seguida de didesoxisecuencia de Sanger. Diluciones de ADN procedente de estirpes celulares con mutaciones Kras en tipo salvaje revelan el enriquecimiento de mutaciones en varias posiciones de la secuencia vía PCR COLD.

FIGURA 6

EJEMPLOS DE MUTACIONES DE BAJO NIVEL EN MUESTRAS CLÍNICAS DE TUMORES SÓLIDOS, PREVIAMENTE "INVISIBLES" VÍA DIDESOXISECUENCIACIÓN DE SANGER, QUE AHORA SE HACEN DETECTABLES VÍA PCR COLD

Didesoxisecuenciación de Sanger de muestras tumorales CLÍNICAS para mutaciones del exón 8 de p53

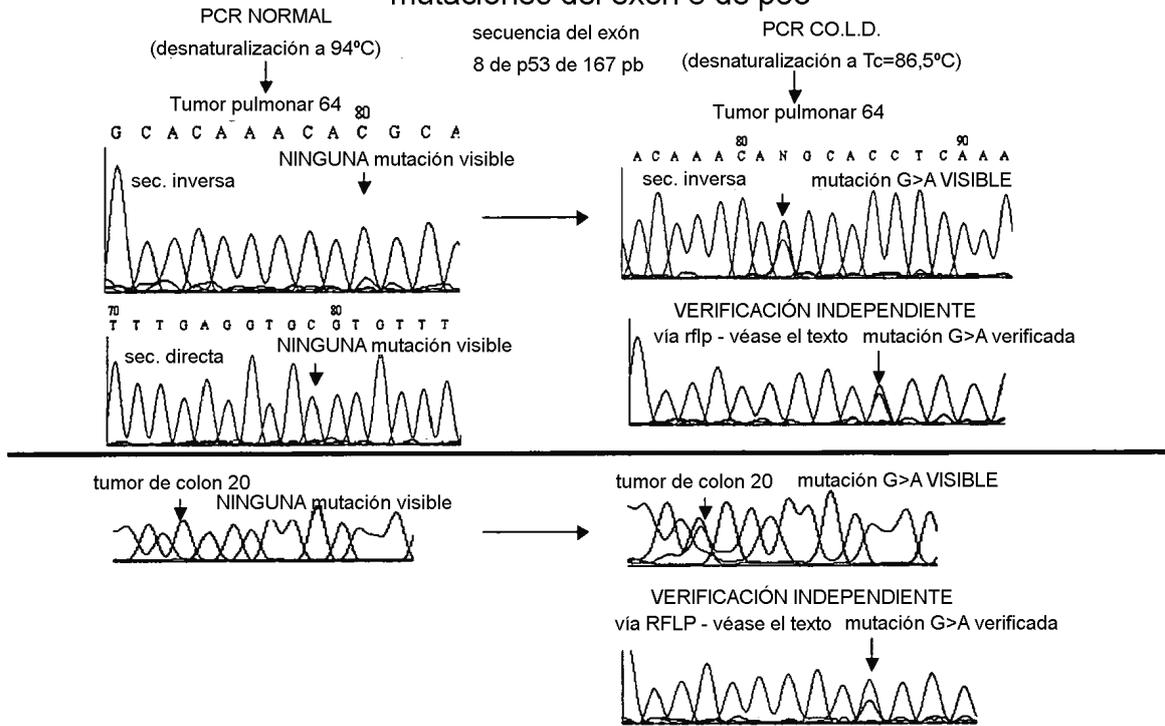


Figura 6: Muestra tumorales clínicas con mutaciones de bajo predominio en el exón 8 de p53: Comparación de PCR enriquecida frente a PCR normal seguida de didesoxisecuenciación de Sanger. El enriquecimiento de mutaciones permite la detección de mutaciones.

FIGURA 7

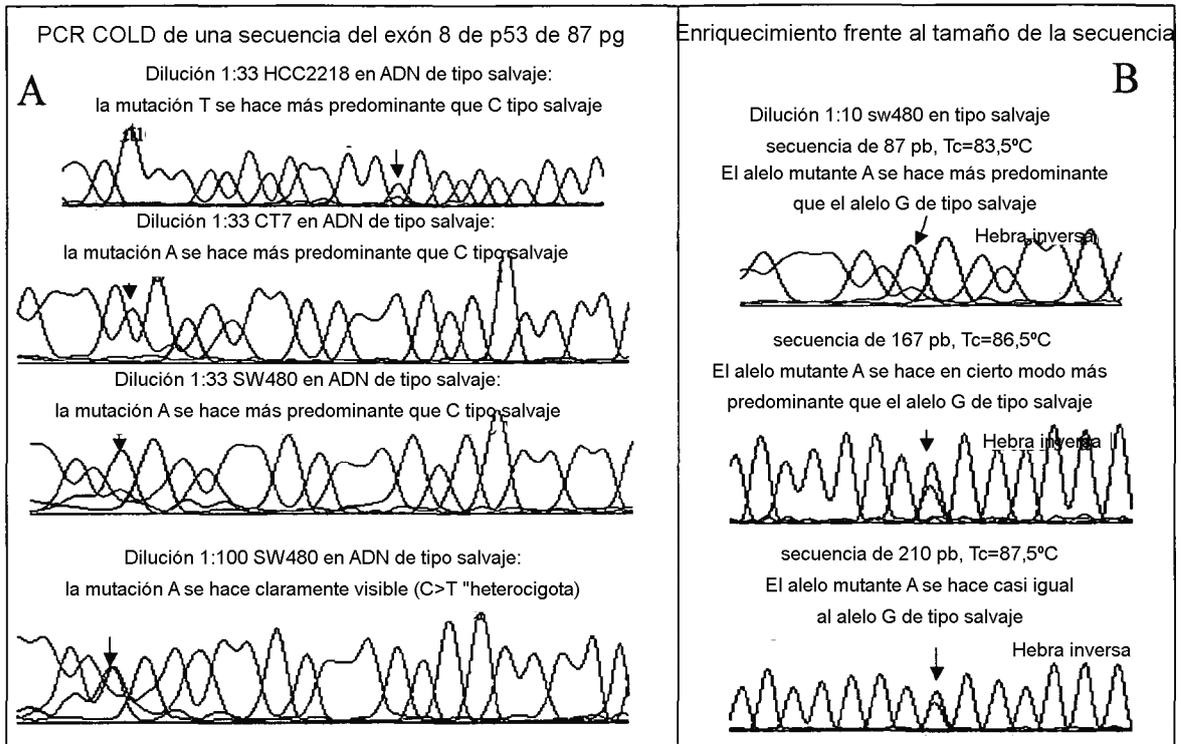


Figura 7: A. Las mutaciones que reducen la temperatura de fusión se pueden enriquecer en un protocolo de enriquecimiento acortado: se representa el enriquecimiento de varias mutaciones de p53 en una secuencia de 87 pb, seguido de la secuenciación de Sanger. B. Los amplicones de <100 pb dan como resultado un mayor enriquecimiento de las mutaciones. Sin embargo, el enriquecimiento de las mutaciones es sustancial (10 veces) para secuencias de al menos 200 pb.

DETECCIÓN DE MUTACIONES DEL EXÓN 8 DE p53: COMPARACIÓN DE PCR EN TIEMPO REAL NORMAL FRENTE A PCR COLD EN TIEMPO REAL

FIG. 8

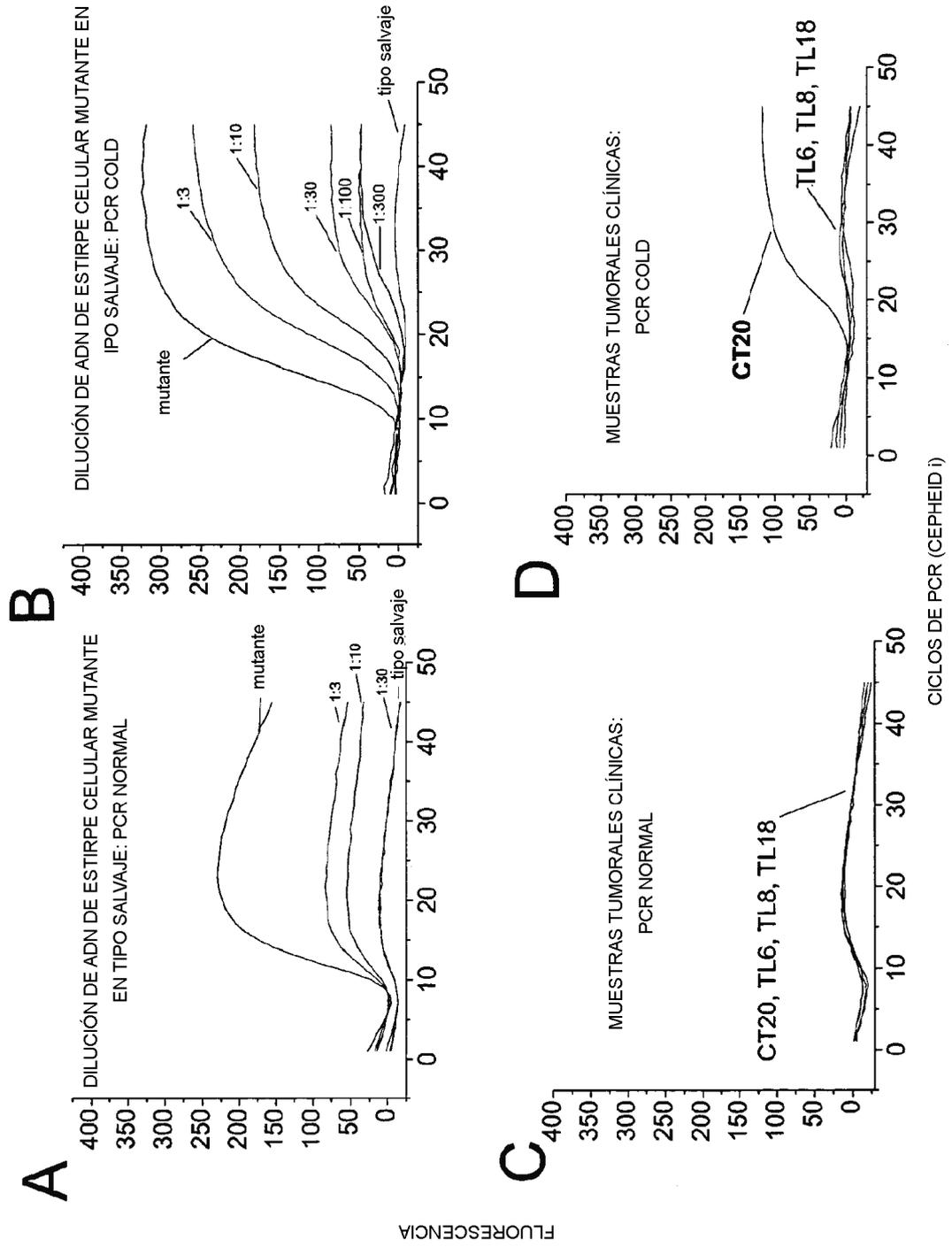


Fig 9

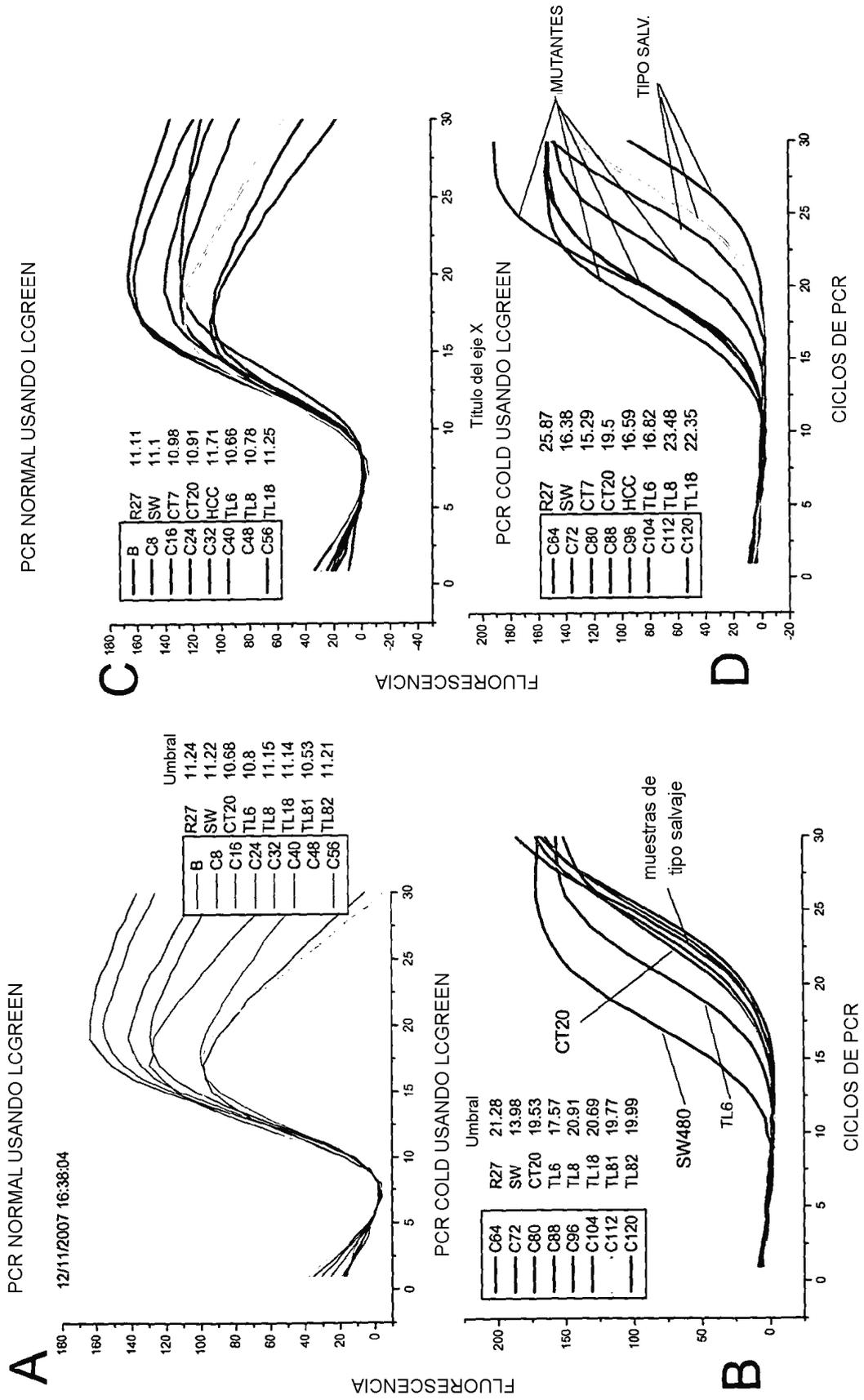
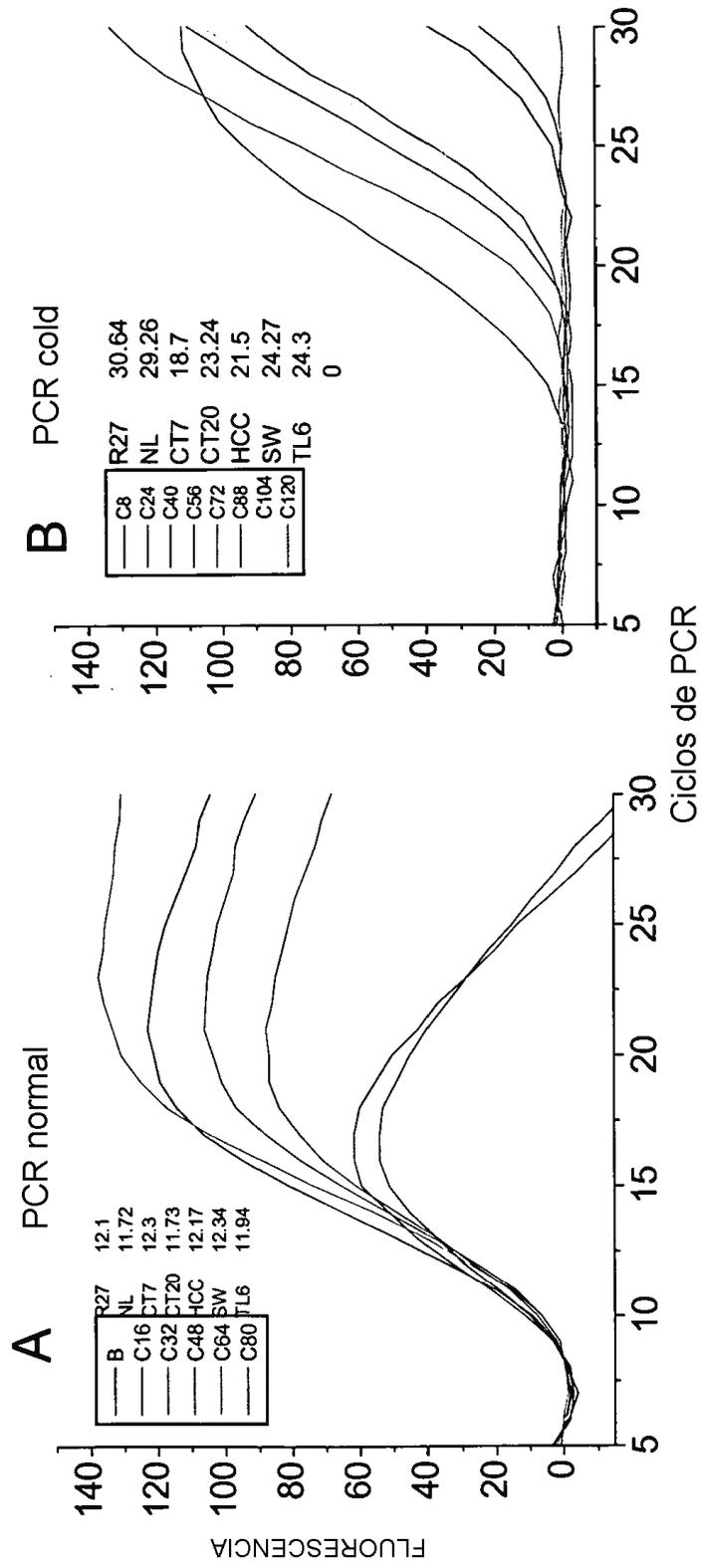


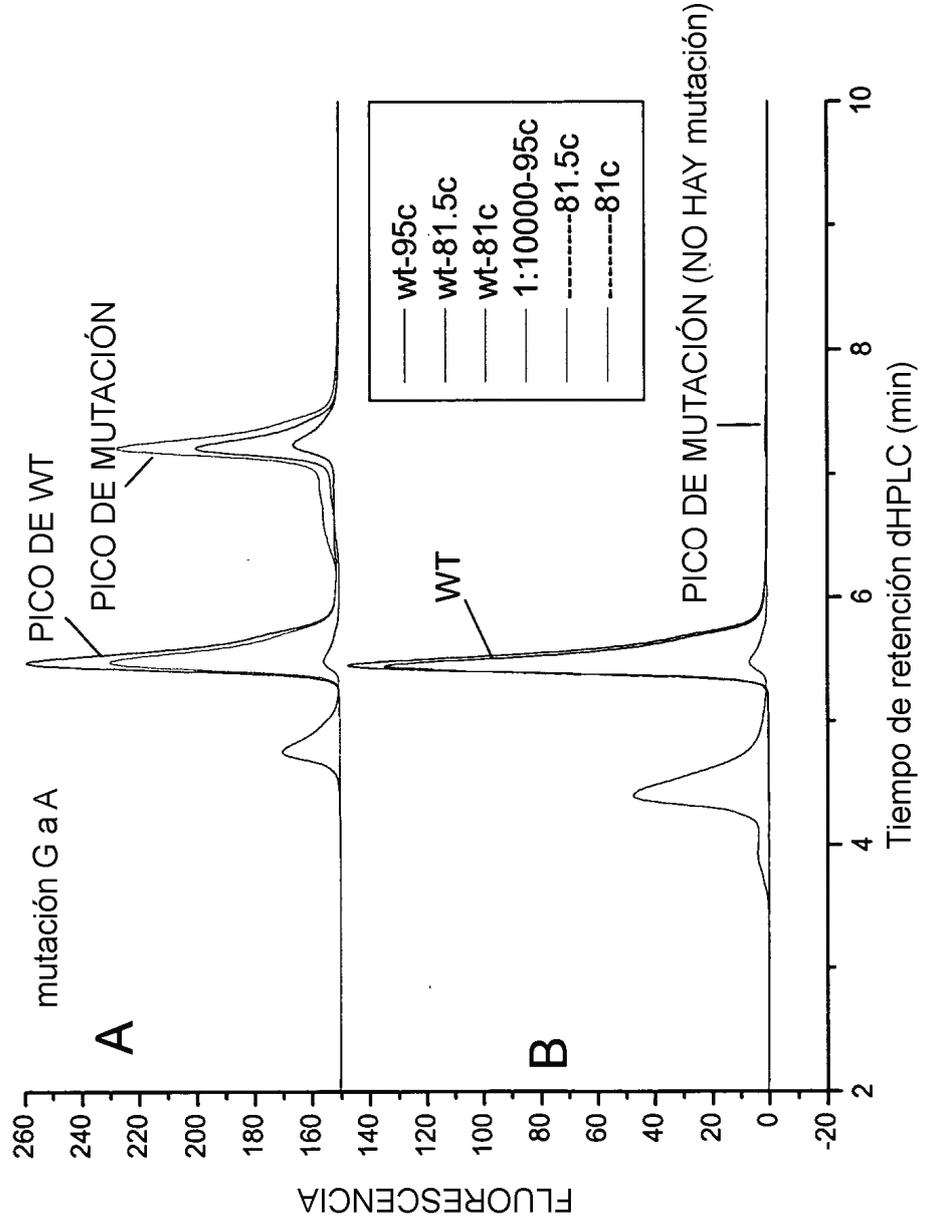
Fig 10

3% DE DMSO en tampón de PCR: la Tc cambia a 80,5°C



**Fig 11**

Mejora de la detección de la mutación del sitio de restricción (RFLP) usando PCR COLD tras la digestión selectiva mediante TaqI de la mutación de EGFR AA761 (mutación de resistencia): comparación de PCR normal frente a PCR COLD



12/4/2007 14:29:11