

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2008.07.31	(73) Titular(es): DANA FARBER CANCER INSTITUTE 44 BINNEY STREET BOSTON, MA 02155 US
(30) Prioridade(s): 2007.08.01 US 962838 P	
(43) Data de publicação do pedido: 2010.05.12	(72) Inventor(es): GERASSIMOS MAKRIGIORGOS US
(45) Data e BPI da concessão: 2015.04.15 188/2015	(74) Mandatário: NUNO MIGUEL OLIVEIRA LOURENÇO RUA CASTILHO, Nº 50 - 9º 1269-163 LISBOA PT

(54) Epígrafe: **ENRIQUECIMENTO DE UMA SEQUÊNCIA ALVO**

(57) Resumo:

A PRESENTE INVENÇÃO É DIRIGIDA A MÉTODOS, COMPOSIÇÕES, SOFTWARE E DISPOSITIVOS DE ENRIQUECIMENTO DE ALELOS EM POUCA ABUNDÂNCIA NUMA AMOSTRA. O MÉTODO É BASEADO EM PARTE NA MODIFICAÇÃO DE UM PROTOCOLO DE AMPLIFICAÇÃO DE ÁCIDOS NUCLEICOS, QUE INCLUI A INCUBAÇÃO DE UMA MISTURA REACIONAL A UMA TEMPERATURA CRÍTICA DE DESNATURAÇÃO OU T_{TC} . ATRAVÉS DA APLICAÇÃO DA PRESENTE INVENÇÃO, OS LIMITES DE DETECÇÃO ATUAIS DE TODAS AS TECNOLOGIAS COM BASE EM RCP SÃO FORTEMENTE MELHORADOS.

RESUMO

"ENRIQUECIMENTO DE UMA SEQUÊNCIA ALVO"

A presente invenção é dirigida a métodos, composições, *software* e dispositivos de enriquecimento de alelos em pouca abundância numa amostra. O método é baseado em parte na modificação de um protocolo de amplificação de ácidos nucleicos, que inclui a incubação de uma mistura reacional a uma temperatura crítica de desnaturação ou "Tc". Através da aplicação da presente invenção, os limites de detecção atuais de todas as tecnologias com base em RCP são fortemente melhorados.

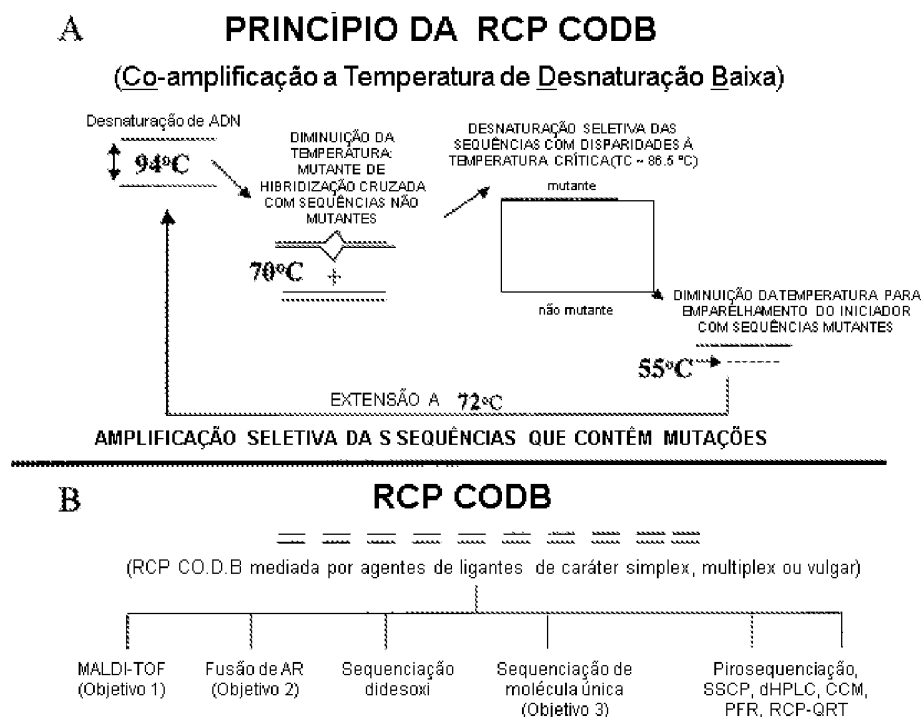


Figura 1. A. Ilustração do protocolo de enriquecimento. É apresentado um exemplo para a sequência do exão 8 de 167 pb p153. A formação de disparidades em qualquer posição ao longo da sequência durante a RCP, permite a desnaturação preferencial e amplificação dos alelos menores (mutantes) a cada ciclo da RCP. **B.** Substituição da RCP pelo método de enriquecimento: Todos os ensaios de teste genéticos com base em RCP, beneficiam do enriquecimento das mutações durante a etapa de RCP que os antecede.

DESCRIÇÃO

"ENRIQUECIMENTO DE UMA SEQUÊNCIA ALVO"

Descrição

Área da invenção

Uma situação comumente encontrada em análises genéticas, compreende a necessidade de identificar uma baixa percentagem de sequências de ADN variante ("sequências alvo") na presença de um grande excesso de sequências não variantes ("sequências de referência"). Exemplos de tais situações incluem: (a) a identificação e sequenciação de alguns alelos metilados, na presença de um grande excesso de alelos normais; (b) a identificação de alguns alelos metilados, na presença de um grande excesso de alelos não metilados (ou vice-versa) em análises epigenéticas; (c) a identificação e genotipagem de algumas sequências de ADN fetal em circulação no sangue da mãe, onde também está presente um grande excesso de sequências de ADN da mãe, e (d) a identificação de ADN referente a eventos tumorais em circulação no sangue de pacientes de cancro (ou pessoas suspeitas de ter cancro), na presença de um grande excesso de alelos de tipo não mutante.

Enquanto métodos fiáveis de rastreio de alta eficiência para mutações em células germinativas ou mutações somáticas de alta prevalência foram recentemente descritos (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), a deteção de mutações somáticas de baixa prevalência em tumores com heterogeneidade, na contaminação do estroma ou em fluidos corporais ainda é problemática. E ainda, o significado

clínico da identificação destas mutações é importante em várias situações. Por exemplo: (a) no adenocarcinoma do pulmão, as mutações EGFR de baixo nível que não podem ser identificadas por sequenciação tradicional, podem conferir uma resposta positiva a inibidores da tirosina-quinase (Paez, J.G., et al. (2004) *Science*, 304; 1497-1500) ou à resistência de fármacos (Janne, P.A., et al. (2006) *Clin Cancer Res*, 12; 751-758); (b) as mutações no plasma úteis como biomarcadores para a detecção precoce (Diehl, F., et al. (2005) *Proc Natl Acad Sci U S A*, 102; 16368-16373) ou para a resposta de tumores ao tratamento (Kimura, T., et al. (2004) *Ann N Y Acad Sci*, 1022; 55-60), não podem ser sequenciadas utilizando métodos convencionais; e (c) as mutações em tumores com a contaminação frequente do estroma, tais como no pâncreas ou na próstata, podem ser 'mascaradas' pela presença de alelos de tipo não mutante, exigindo portanto uma microdissecção laboriosa ou causando o despercebimento completo das mutações.

O documento WO 03/072809 (ver exemplos 1 e 2) divulga os protocolos para uma RCP que utiliza uma temperatura de desnaturação 95 °C, que vai provocar a desnaturação tanto da sequência alvo como da sequência de referência. Numa fase posterior, a temperatura é inicialmente diminuída para 58 °C, seguindo-se de um aumento para 72 °C. Subsequentemente, a temperatura é aumentada para uma temperatura entre a T_f do alvo e a T_f das sequências de referência. Nos próximos passos, permite-se que os iniciadores se liguem à sequência alvo e que se estendam. As sequências diferencialmente metiladas podem ser enriquecidas durante a amplificação.

Sumário da invenção

A presente invenção é dirigida a métodos para o enriquecimento de alelos em pouca abundância numa amostra, tal como definido nas reivindicações.

O método baseia-se em parte num protocolo modificado para a amplificação de ácidos nucleicos, incluindo a incubação da mistura de reação a uma temperatura de desnaturação crítica ou "Tc". Ao empregar a presente invenção, os limites atuais de detecção de todas as tecnologias baseadas na RCP são significativamente melhorados.

A "temperatura crítica" ou "Tc" refere-se a uma temperatura abaixo da temperatura de fusão "T_f" da sequência de referência. Em algumas formas de realização, a Tc está abaixo da T_f tanto da sequência alvo como da sequência de referência. A temperatura crítica retira vantagens de uma T_f inferior referente a uma sequência alvo de cadeia dupla, ou referente a um dúplex de ADN de cadeia dupla hibridizado cruzadamente com uma sequência de referência, de modo a desnaturar preferencialmente estes dúplexes em detrimento dos homodúplexes referência/referência. Quando a sequência alvo e a sequência de referência hibridizam cruzadamente, pequenas diferenças na sequência de um ou mais nucleotídeos referentes a disparidades ao longo de uma sequência de ADN de cadeia dupla de curta dimensão (por exemplo, < 200 pb), irão gerar uma pequena alteração, mas previsível, da temperatura de fusão (T_f) para essa sequência (Lipsky, R.H., *et al.* (2001) Clin Chem, 47, 635-644; Liew, M., *et al.* (2004) Clin Chem, 50, 1156-1164). Dependendo do contexto exato da sequenciação e da posição de disparidade, estão contempladas mudanças de temperatura de fusão de 0,1-20 °C.

A temperatura de desnaturação crítica (T_c), é a temperatura cujos valores inferiores de temperatura causam a queda abrupta da eficiência da RCP para uma sequência de ácidos nucleicos de referência.

Por exemplo, uma sequência 167 pb p53 é amplificada corretamente se a temperatura de desnaturação da RCP é definida a 87 °C, é amplificada razoavelmente a 86,5 °C e não produz nenhum produto detetável se a temperatura de desnaturação da RCP é definida a 86 °C ou menos. Portanto, neste exemplo, $T_c \sim 86,5$ °C.

Num primeiro aspeto, a invenção é dirigida a um método para o enriquecimento de uma sequência alvo numa amostra de ácidos nucleicos, que é suspeita de possuir sequências alvo e de referência. O método inclui a sujeição da mistura reacional de amplificação a uma primeira temperatura de desnaturação, que está acima da temperatura de fusão " T_f " de uma sequência de referência. Posteriormente, a temperatura da mistura reacional de amplificação é reduzida, permitindo que as sequências alvo de cadeia simples e as sequências de referência possam hibridizar de forma a gerar moléculas de cadeia dupla. Assim, a reação inclui homodúplexes alvo/alvo, homodúplexes referência/referência e heterodúplexes de cadeias referência-alvo. Um heterodúplex por definição é um dúplex imperfeito com disparidades, que apesar de tudo contém homologia suficiente entre as cadeias para se manter numa forma dúplex na mistura reacional.

Um homodúplex por definição é um dúplex perfeitamente compatível. A temperatura da mistura reacional é então aumentada para a T_c , o que resulta na desnaturação preferencial dos dúplexes hibridados de sequência alvo/referência. A T_c , ou temperatura crítica, está abaixo

da T_f da sequência de referência, e pode ser determinada pelos métodos aqui descritos. Na T_c , os dúplexes de sequência alvo/referência (e apenas os dúplexes de sequência alvo/alvo que têm uma T_f mais baixa do que a T_f referente à sequência de referência) estão substancialmente desnaturados, ao passo que os dúplexes alvo/alvo (se tiverem uma T_f igual ou superior à T_f da sequência de referência) e os dúplexes de sequência referência/referência, estão substancialmente em formas não desnaturadas. O termo "substancialmente" significa pelo menos 60 %, preferencialmente pelo menos 70 %, mais preferencialmente pelo menos 80 %, ainda mais preferencialmente pelo menos 90 % e sobretudo preferencialmente pelo menos 98 %, numa determinada forma desnaturada ou não desnaturada. Após a desnaturação preferencial dos dúplexes de referência-alvo e dos dúplexes de alvo/alvo (aqueles que têm uma T_f inferior à T_f da sequência de referência), é aplicada uma temperatura reduzida à mistura reacional, de modo a permitir que um par de iniciadores emparelhe com a sequência alvo. Os iniciadores emparelhados são então estendidos, enriquecendo assim a sequência alvo em relação à sequência de referência na amostra.

É divulgado outro método de enriquecimento de uma sequência alvo. Neste método, uma amostra de ácidos nucleicos que é suspeita de conter cada uma das sequência alvo e sequências de referência, é desnaturada pela aplicação de uma primeira temperatura de desnaturação que é acima da T_f da sequência de referência. De seguida, as cadeias alvo e de referência são emparelhadas uma na outra de modo a formar dúplexes de sequência alvo/referência de cadeia dupla. As formas dúplex de sequência alvo/referência estão presentes na mistura reacional juntamente com as cadeias duplas alvo/alvo e dúplexes de sequência referência/referência. Os dúplexes de

sequência alvo/referência e de sequência alvo/alvo de cadeia dupla, são preferencialmente desnaturados pela aplicação da T_c na amostra. Na T_c , os dúplexes de sequência alvo/referência (e apenas os dúplexes de sequência alvo/alvo que têm uma T_f mais baixa do que a T_f referente à sequência de referência) estão substancialmente desnaturados, ao passo que os dúplexes alvo/alvo (se tiverem uma T_f igual ou superior à T_f da sequência de referência) e os dúplexes de sequência referência/referência, estão substancialmente em formas não desnaturadas. O termo "substancialmente" significa pelo menos 60 %, preferencialmente pelo menos 70 %, mais preferencialmente pelo menos 80 %, ainda mais preferencialmente pelo menos 90 % e sobretudo preferencialmente pelo menos 98 %, numa determinada forma desnaturada ou não desnaturada. Um par de iniciadores é então emparelhado com a sequência alvo e é estendido, aumentando assim a concentração da sequência alvo em relação à sequência de referência na amostra.

É ainda divulgado um método para o enriquecimento de uma sequência alvo, segundo a realização de um protocolo para uma reação de amplificação de ácidos nucleicos. O protocolo para a reação de amplificação inclui uma primeira temperatura de desnaturação e uma segunda temperatura de desnaturação. A primeira temperatura de desnaturação é superior à T_f da sequência de referência, e a segunda temperatura de desnaturação é inferior à T_f da sequência de referência.

Mesmo assim, é ainda divulgado um método de enriquecimento de uma sequência alvo, que tem uma T_f inferior à T_f da sequência de referência correspondente, através da sujeição de uma mistura reacional de amplificação a uma T_c , da redução da temperatura da mistura reacional e da extensão

de um par de iniciadores. A mistura reacional de amplificação é suspeita de conter cada uma das sequências alvo e sequências de referência. A T_c é inferior à T_f da sequência de referência, permitindo assim a desnaturação preferencial da sequência alvo com a menor T_f . Na T_c , os dúplexes de sequência alvo/referência e os dúplexes de sequência alvo/alvo estão substancialmente desnaturados, enquanto os dúplexes de sequência referência/referência estão substancialmente em formas não desnaturadas. O termo "substancialmente" significa pelo menos 60 %, preferencialmente pelo menos 70 %, mais preferencialmente pelo menos 80 %, ainda mais preferencialmente pelo menos 90 % e sobretudo preferencialmente pelo menos 98 %, numa determinada forma desnaturada ou não desnaturada. A etapa de redução da temperatura da mistura reacional permite que o par de iniciadores emparelhe com a sequência alvo. Estes iniciadores emparelhados são posteriormente estendidos por uma polimerase, aumentando a quantidade de sequência alvo na amostra relativamente à sequência de referência.

É também divulgado um método de enriquecimento de uma sequência alvo, por sujeição de uma mistura reacional de amplificação a múltiplos ciclos de condições de emparelhamento e a T_c . A mistura reacional de amplificação, que é suspeita de conter cada uma das sequências alvo e sequências de referência, é inicialmente submetida a uma primeira temperatura de desnaturação superior à T_f da sequência de referência. De seguida, a amostra percorre um ciclo repetitivo entre duas etapas de incubação a temperatura diferentes. Na primeira etapa de incubação, a temperatura é reduzida de modo a permitir a hibridização da sequência alvo com a sequência de referência, formando um dúplex. Na segunda etapa de incubação, a temperatura é aumentada para a T_c , a qual é inferior à T_f da sequência de referência. Na T_c , os dúplexes de sequência alvo/referência

(e apenas os dúplexes de sequência alvo/alvo que têm uma T_f mais baixa do que a T_f referente à sequência de referência) estão substancialmente desnaturados, ao passo que os dúplexes alvo/alvo (se tiverem uma T_f igual ou superior à T_f da sequência de referência) e os dúplexes de sequência referência/referência, estão substancialmente em formas não desnaturadas. O termo "substancialmente" significa pelo menos 60 %, preferencialmente pelo menos 70 %, mais preferencialmente pelo menos 80 %, ainda mais preferencialmente pelo menos 90 % e sobretudo preferencialmente pelo menos 98 %, numa determinada forma desnaturada ou não desnaturada. Estas primeiras e segundas etapas são então repetidas uma ou mais vezes. Uma vez que a etapa do ciclo de incubação esteja completa, a temperatura da mistura reacional é reduzida de modo a permitir que um par de iniciadores emparelhe com a sequência alvo. Estes iniciadores são posteriormente estendidos por uma polimerase, aumentando então a quantidade de sequência alvo na amostra relativamente à sequência de referência.

Breve descrição das figuras

A Figura 1 ilustra uma forma de realização do processo de enriquecimento da sequência alvo referente à invenção, que utiliza uma primeira temperatura de desnaturação e a T_c .

A Figura 2 ilustra uma forma de realização do processo de enriquecimento da sequência alvo referente à invenção, que utiliza uma etapa de emparelhamento oscilante/ T_c .

A Figura 3 ilustra as mutações p53 e Kras testadas para o enriquecimento através de um protocolo de enriquecimento da sequência alvo.

A Figura 4 ilustra o enriquecimento de alelos mutantes de p53 com um protocolo de enriquecimento da sequência alvo.

A Figura 5 ilustra o enriquecimento de alelos mutantes de Kras com um protocolo de enriquecimento da sequência alvo.

A Figura 6 ilustra o enriquecimento de alelos mutantes em amostras clínicas de tumores do pulmão e do cólon.

A Figura 7 ilustra o enriquecimento de alelos mutantes através de uma forma de realização do protocolo de enriquecimento a que falta a etapa de formação de disparidades.

A Figura 8A-8D ilustra os gráficos da amplificação do exão 8 de p53 não mutante e mutante, segundo RCP em tempo real convencional ou através de um método de enriquecimento da invenção aplicado num formato de tempo real: (a) apresenta os gráficos de amplificação referentes a RCP em tempo real convencional, para linhas celulares serialmente diluídas que contêm uma mutação do exão 8 de p53 mutante; (b) apresenta os gráficos de amplificação referentes a um processo de enriquecimento da invenção, aplicado num formato de RCP em tempo real, para linhas celulares serialmente diluídas que contêm uma mutação do exão 8 de p53 mutante; (c) apresenta os gráficos de amplificação referentes a RCP em tempo real convencional para quatro amostras clínicas de tumores, um dos quais é conhecido por conter uma mutação do exão 8 de p53 (CT20); (d) apresenta os gráficos de amplificação referentes a um processo de enriquecimento da invenção, aplicado num formato de RCP em tempo real, para as quatro amostras clínicas de tumores.

A Figura 9A-D ilustra os gráficos da amplificação por RCP em tempo real convencional e por RCP em tempo real

enriquecida, para o exão 8 de p53 não mutante e mutante, através da utilização de um corante de detecção de ADN (LCGreen): (a) apresenta os gráficos de amplificação para RCP em tempo real convencional referentes a p53 mutante (SW480, TL6, CT20) e referentes a p53 não mutante (R27, TL8, TL18, TL81, TL82), em amostras que contêm o exão 8; (b) apresenta os gráficos de amplificação para um processo de enriquecimento da invenção, aplicado num formato de tempo real, referentes a p53 mutante (SW480, TL6, CT20) e referentes a p53 não mutante (R27, TL8, TL18, TL81, TL82), em amostras que contêm o exão 8; (c) apresenta os gráficos de amplificação para RCP em tempo real convencional referentes a p53 mutante (SW480,CT7,HCC,CT20) e referentes a p53 não mutante, em amostras que contêm o exão 8; (d) apresenta os gráficos de amplificação para um processo de enriquecimento da invenção, aplicado num formato de tempo real, referentes a p53 mutante (SW480,CT7,HCC, CT20) e referentes a p53 não mutante, em amostras que contêm o exão 8.

A Figura 10A-B ilustra o impacto de solventes orgânicos (por exemplo, 3 % de DMSO) no enriquecimento de alvos: (a) gráficos de amplificação para RCP em tempo real referentes ao exão 8 de p53 não mutante e mutante; (b) gráficos de amplificação para um processo de enriquecimento da invenção, aplicado num formato de tempo real, referentes ao exão 8 de p53 não mutante e mutante.

A Figura 11 ilustra a melhoria da detecção por Polimorfismo dos Fragmentos de Restrição (PFR), através da utilização de um processo de enriquecimento da invenção após a digestão TaqI da mutação AA761 EGFR; (a) ilustra a detecção de (AA761) EGFR não mutante e mutante aquando uma diluição de 1:10000 (mutante:genoma), após RCP regular ou um

procedimento de enriquecimento; (b) ilustra a detecção de EGFR não mutante, tal como testado por RCP normal.

Descrição detalhada

A presente invenção é dirigida a métodos para o enriquecimento de alelos em pouca abundância (por exemplo, sequências alvo) a partir de uma amostra. A presente invenção é dirigida, em parte, ao enriquecimento seletivo de sequências alvo, através da aplicação de uma temperatura crítica de desnaturação a uma mistura reacional. A temperatura de desnaturação crítica, ou T_c , é a temperatura abaixo da temperatura de fusão T_f da sequência de referência. A temperatura crítica retira vantagens de uma T_f inferior referente a um dúplex de ADN de cadeia dupla hibridizado com uma sequência de referência, de modo a desnaturar preferencialmente estes dúplexes em detrimento dos homodúplexes referência/referência.

Vários dos métodos que são conhecidos para a detecção de mutações podem beneficiar da presente invenção, dado que esta utiliza um protocolo de amplificação que enriquece seletivamente a sequência alvo. A utilização de RCP, como a etapa inicial para a análise genética, é empregue na maioria das reações de sequenciação e mutação. Geralmente, uma sequência de ácidos nucleicos (por exemplo, ADN genómico/ADNc) é amplificada numa primeira etapa, seguindo-se um método de sequenciação didesoxi ou um método de rastreio de mutações (por exemplo, SSCP, dHPLC, MALDI-TOF, pirosequenciação, fusão de alta resolução). Deste modo, os limites de praticamente todos os métodos de detecção de mutações seriam beneficiados uniformemente, caso o enriquecimento das sequências que contêm a mutação (por exemplo, sequências alvo) fosse realizado durante a etapa de RCP precedente ao rastreio.

Definições

Tal como aqui utilizado, o termo "enriquecimento de uma sequência alvo" refere-se ao aumento da quantidade de uma sequência alvo, e ao aumento do rácio da sequência alvo em relação à sequência de referência correspondente numa amostra. Por exemplo, quando o rácio da sequência alvo em relação à sequência de referência é inicialmente de 5 % a 95 % numa amostra, a sequência alvo pode ser preferencialmente amplificada numa reação de amplificação, de modo a produzir um rácio de 70 % de sequência alvo face a 30 % de sequência de referência. Assim, há um enriquecimento de 14 vezes da sequência alvo em relação à sequência de referência. O enriquecimento de uma sequência alvo resulta num aumento de 2× a 200× da sequência alvo em relação à sequência de referência, face ao seu valor anterior ao enriquecimento. O enriquecimento do alvo é, pelo menos, 2×, 3×, 4×, 5×, 6×, 7×, 8×, 9×, 10×, 15×, 20×, 25×, 30×, 35×, 40×, 45×, 50×, 60×, 70×, 80×, 90×, 100×, 150×, 200× ou mais, o enriquecimento da sequência de referência. O enriquecimento de uma sequência alvo resulta numa amostra que possui 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 80 %, 90 %, 95 % ou mais, de sequência alvo comparativamente à sequência de referência (por exemplo, 10 % de sequência alvo:90 % de sequência de referência a 95 % de sequência alvo:5 % sequência de referência).

Tal como aqui utilizado, o termo "sequência alvo" refere-se a um ácido nucleico que é menos prevalente numa amostra de ácidos nucleicos, em relação a uma sequência de referência correspondente. A sequência alvo constitui menos de 50 % da quantidade total de sequenciação (sequência de referência + sequência alvo) de uma amostra. Preferencialmente, a sequência alvo é expressa ao nível do ARN e/ou ADN em 1:10,

1:15, 1:20, 1:25×, 1:30, 1:35, 1:40, 1:45, 1:50, 1:60, 1:70, 1:80, 1:90, 1:100, 1:150, 1:200× ou menos, face à sequência de referência. Numa forma de realização, a sequência alvo é um alelo mutante. A título de exemplo, uma amostra (por exemplo, uma amostra de sangue) pode conter várias células normais e algumas células cancerígenas. As células normais contêm alelos não mutantes, enquanto um pequeno número de células cancerígenas contém mutações somáticas. Neste caso, a sequência mutante é a sequência alvo, enquanto a sequência de tipo não mutante é a sequência de referência. Numa outra forma de realização, a invenção é dirigida à detecção de ADN fetal numa amostra de ácidos nucleicos obtida a partir de uma mãe. Nesta forma de realização, a sequência alvo está presente no ADN fetal, enquanto a sequência de referência está presente no ADN materno que é mais prevalente. Tal como aqui utilizado, uma sequência alvo pretende incluir ADN fetal obtido a partir de uma mãe grávida. Tal como aqui utilizado, uma "cadeia alvo" refere-se a uma cadeia simples de ácido nucleico de uma sequência alvo.

A sequência alvo tem um comprimento de cerca de 17-2000 nucleótidos. Numa forma de realização, a sequência alvo tem um comprimento 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 ou mais nucleótidos. As sequências alvo partilham pelo menos 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % ou mais, de homologia com a sequência de referência correspondente, mas diferem pelo menos num nucleótido face à sequência de referência. As sequências alvo podem ser amplificadas por RCP, utilizando o mesmo par de iniciadores que foi usado para a sequência de referência.

Tal como aqui utilizado, o termo "sequência de referência" refere-se a um ácido nucleico que é mais prevalente numa amostra de ácidos nucleicos, em relação a uma sequência alvo correspondente (por exemplo, o mesmo gene mas com uma sequência de ácidos nucleicos diferente). A sequência de referência constitui mais de 50 % da quantidade total de sequenciação (sequência de referência + sequência alvo) de uma amostra. Preferencialmente, a sequência de referência é expressa ao nível do ARN e/ou ADN em 10×, 15×, 20×, 25×, 30×, 35×, 40×, 45×, 50×, 60×, 70×, 80×, 90×, 100×, 150×, 200× ou mais, face à sequência alvo. Numa forma de realização, a sequência de referência é um alelo não mutante. A título de exemplo, uma amostra (por exemplo, uma amostra de sangue) pode conter várias células normais e algumas células cancerígenas. As células normais contêm alelos não mutantes, enquanto um pequeno número de células cancerígenas contém mutações somáticas. Neste caso, a sequência não mutante é a sequência de referência, enquanto a sequência mutante é a sequência alvo (por exemplo, um alelo mutante). Tal como aqui utilizado, uma "cadeia de referência" refere-se a uma cadeia simples de ácido nucleico de uma sequência de referência.

A sequência de referência tem um comprimento de cerca de 17-2000 nucleótidos. Numa forma de realização, a sequência de referência tem um comprimento 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 ou mais nucleótidos. As sequências de referência partilham pelo menos 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % ou mais, de homologia com a sequência alvo correspondente, mas diferem pelo menos num nucleótido face à sequência alvo. As sequências de referência podem ser amplificadas por RCP,

utilizando o mesmo par de iniciadores que foi usado para a sequência de referência.

O termo "alelo" refere-se a formas alternativas de um gene, de uma porção sua ou de uma região de não codificação de ADN, que ocupa o mesmo *locus* ou posição em cromossomas homólogos, possuindo esta região pelo menos uma diferença na sequência nucleotídica. O termo alelo pode ser usado para descrever o ADN obtido a partir de qualquer organismo, incluindo, mas não limitado a, bactérias, vírus, fungos, protozoários, bolor, leveduras, plantas, seres humanos, seres não humanos, animais e arqueobactérias. Os alelos podem ser encontrados numa única célula (por exemplo, dois alelos, um herdado do pai e um herdado da mãe) ou numa população celular (por exemplo, um alelo não mutante obtido a partir de tecido normal e um alelo mutante somático obtido a partir de tecido afetado por doença).

Um alelo tem um comprimento de cerca de 17-2000 nucleótidos. Numa forma de realização, o alelo tem um comprimento 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 ou mais nucleótidos. Os alelos partilham pelo menos 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % ou mais, de homologia entre eles. Os alelos podem ser amplificados por RCP com o mesmo par de iniciadores.

Numa forma de realização, a presente invenção é utilizada para enriquecer um polimorfismo. Qualquer gene que é dado pode ter nenhuma, uma ou várias formas alélicas (polimorfismo). As alterações mutacionais comuns que originam alelos podem resultar de eliminações, adições ou substituições nucleotídicas, de carácter natural ou artificial (por exemplo, agentes químicos cancerígenos).

Cada um destes tipos de alterações pode ocorrer sozinho, ou em combinação com os outros tipos de alterações, uma ou mais vezes numa determinada sequência.

Foram relatados vários tipos de polimorfismos diferentes. Um polimorfismo dos fragmentos de restrição (PFR) é uma variação na sequência de ADN que altera o comprimento de um fragmento de restrição (Botstein *et al.*, *Am. J. Hum. Genet.* 32:314-331 (1980)). O polimorfismo dos fragmentos de restrição pode criar ou eliminar um local de restrição, alterando assim o comprimento do fragmento de restrição. Os PFRs têm sido amplamente utilizados em análises genéticas de humanos e de animais (ver documentos WO 90/13668; WO 90/11369; Donis-Keller, *Cell* 51:319-337 (1987); Lander *et al.*, *Genetics* 121:85-99 (1989)). Quando uma característica hereditária pode ser ligada a um PFR em particular, a presença do PFR num indivíduo pode ser utilizada para prever a probabilidade do indivíduo de exibir também essa característica. O PFR pode ser utilizado em combinação com os métodos de enriquecimento aqui descritos, de modo a melhorar a deteção de polimorfismos. Tais métodos são aqui descritos no Exemplo 9.

Outros polimorfismos assumem a forma de repetições curtas em tandem (RCTs), que incluem repetições de motivos em tandem di, tri e tetra-nucleotídicos. Estas repetições em tandem são também referidas como polimorfismos de repetições em tandem de número variável (RTNV). As RTNV têm sido utilizadas na análise de identidade e de paternidade (Patente E.U.A. No. 5,075,217; Armour *et al.*, *FEBS Lett.* 307:113-115 (1992); Horn *et al.*, WO 91/14003; Jeffreys, EP 370,719) e num grande número de estudos de mapeamento genético.

Outros polimorfismos assumem a forma de variações nucleotídicas simples entre os indivíduos da mesma espécie. Tais polimorfismos são muito mais frequentes do que PFRs, RCTs (repetições curtas em tandem) e RTNVs (repetições em tandem de número variável). Alguns polimorfismos nucleotídicos simples ocorrem em sequências de codificação de proteínas, cujo neste caso, uma das formas polimórficas pode dar origem à expressão de uma proteína defeituosa ou de outra variante, e, potencialmente, podem dar origem a uma doença genética. Outros polimorfismos nucleotídicos simples ocorrem em regiões de não codificação. Alguns destes polimorfismos podem também resultar na expressão de proteínas defeituosas (por exemplo, como resultado de *splicing* defeituoso). Outros polimorfismos nucleotídicos simples não têm efeitos fenotípicos.

No entanto, outras mutações incluem as mutações somáticas. As mutações somáticas são alterações que ocorrem no ADN após a sua concepção. As mutações somáticas podem ocorrer em qualquer uma das células do corpo, com exceção das células germinativas (espermatozoides e óvulos), e conseqüentemente não são transferidas para crianças. Estas alterações podem (mas nem sempre) conduzir a cancro ou outras doenças.

Tal como aqui utilizado, o termo "não mutante" refere-se à sequência polinucleotídica ou alelo mais comum para um certo gene numa população. Geralmente, o alelo não mutante será obtido a partir de células normais.

Tal como aqui utilizado, o termo "mutante" refere-se a uma alteração nucleotídica (isto é, uma substituição nucleotídica única ou múltipla, eliminação ou adição) numa sequência de ácidos nucleicos. Um ácido nucleico que possui uma mutação, tem uma sequência de ácidos nucleicos (alelo mutante) que difere na sua sequência face à sequência

polinucleotídica não mutante correspondente. Os métodos da invenção são especialmente úteis para enriquecer seletivamente um alelo mutante que contém entre 1 a 500 alterações na sequência nucleotídica. Um alelo mutante pode ter 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 ou 500 alterações na sequência nucleotídica em relação a um alelo não mutante correspondente. Preferencialmente, uma mutação num alelo mutante irá conter entre 1 a 10 alterações na sequência nucleotídica, e mais preferencialmente entre 1 a 5 alterações na sequência nucleotídica. O alelo mutante terá 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % ou mais, de homologia com o alelo não mutante. Geralmente, o alelo mutante irá ser obtido a partir de células ou tecidos afetados por doença e está associado com um estado da doença.

Tal como aqui utilizado, o termo "temperatura de fusão" ou " T_f " refere-se à temperatura à qual um polinucleico dissocia-se da sua sequência complementar. Geralmente, a T_f pode ser definida como a temperatura à qual metade dos pares de bases de Watson-Crick, numa molécula de ácidos nucleicos de cadeia dupla, é quebrada ou dissociada (isto é, são "fundida"), enquanto a outra metade dos pares de bases de Watson-Crick permanece intacta numa conformação de cadeia dupla. Em outras palavras, a T_f é definida como a temperatura à qual 50 % dos nucleotídeos de duas sequências complementares estão emparelhadas (cadeias duplas) e 50 % dos nucleotídeos estão desnaturados (cadeias simples). Consequentemente, a T_f define um ponto médio da transição de cadeia dupla para cadeia simples em moléculas de ácidos nucleicos (ou, inversamente, a transição de cadeia simples para cadeia dupla em moléculas de ácidos nucleicos).

A T_f pode ser calculada por uma série de métodos, por exemplo, através da interpolação pelo vizinho mais próximo como em Wetmur 1991 (Wetmur, J. G. 1991. DNA probes: applications of the principles of nucleic acid hybridization. Crit Rev Biochem Mol Biol 26: 227-259, aqui incorporada como referência) e por *softwares* comerciais, incluindo OligoTM Primer Design, e outros *softwares* disponíveis na Internet. Alternativamente, a T_f pode ser determinada por experimentação. Por exemplo, ADN de cadeia dupla de ligação ou corantes intercalantes, tais como brometo de etídio ou verde-SYBR (Molecular Probes), podem ser utilizados num ensaio de curva de fusão para a determinação da T_f efetiva do ácido nucleico. Métodos adicionais para a determinação da T_f referente a um ácido nucleico são conhecidos na especialidade e são aqui descritos.

Tal como aqui utilizado, o termo "temperatura crítica" ou " T_c " refere-se a uma temperatura inferior à T_f da sequência de referência. A T_c é aplicada para desnaturar preferencialmente dúplexes de cadeia dupla de sequência alvo ou dúplexes de cadeia dupla de sequência alvo/referência, de modo a permitir o enriquecimento seletivo da sequência alvo durante uma reação de amplificação. A temperatura de desnaturação crítica (T_c), é a temperatura cujos valores inferiores de temperatura causam a queda abrupta da eficiência da RCP para uma dada sequência de ácidos nucleicos. Por exemplo, uma sequência 167 pb p53 é amplificada corretamente se a temperatura de desnaturação da RCP é definida a 87 °C, é amplificada razoavelmente a 86,5 °C e não produz nenhum produto detetável se a temperatura de desnaturação da RCP é definida a 86 °C ou menos. Portanto, neste exemplo, $T_c \sim 86,5$ °C. A T_c é de cerca de 0,1-20 °C abaixo da T_f da sequência de referência. Mais preferencialmente, a T_c é de

cerca de 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, 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, abaixo da T_f da sequência de referência. Em algumas formas de realização, a T_c é inferior à T_f de ambas as sequências de referência e alvo. Por exemplo, a T_c pode ser de cerca de 0,1-10 °C, 0,1-9 °C, 0,1-8 °C, 0,1-7 °C, 1-6 °C, 0,2-5 °C, 0,3-4,5 °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 abaixo da T_f da sequência alvo.

Tal como aqui utilizado, o termo "desnaturação seletiva" ou "desnaturação preferencial" refere-se à quebra preferencial de pontes de hidrogénio entre os pares de bases presentes numa molécula de ácido nucleico de cadeia dupla, referente a uma sequência alvo ou a um dúplex de sequência alvo/referência, para a produção de uma sequência alvo de cadeia simples. A desnaturação seletiva da sequência alvo é conseguida através da aplicação da temperatura crítica a uma amostra que contém as sequências alvo e de referência.

Tal como aqui utilizado, "par de iniciadores" refere-se a dois iniciadores que emparelham com cadeias opostas de uma sequência alvo e de referência, de modo a formar um produto amplificado durante a reação RCP. O par de iniciadores é concebido de modo a ter uma T_f inferior à T_c da reação.

Tal como aqui utilizado, o termo "hibridização cruzada" refere-se a um dúplex de cadeia dupla formado entre duas sequências de ácidos nucleicos que diferem num ou mais nucleicos, em virtude da formação de pontes de hidrogénio entre as bases complementares G e C, as bases complementares A e T ou as bases complementares A e U. As duas sequências de ácidos nucleicos complementares e ligadas por pontes de hidrogénio, apresentam uma

configuração antiparalela. O ADN hibridado cruzadamente conterá disparidades nas posições de diferença entre as sequências de referência e sequências alvo. As disparidades podem incluir polimorfismos, mutações, adições, eliminações e outras modificações que resultem em tais diferenças. Por exemplo, a metilação. Por exemplo, laços e/ou regiões de cadeia simples de um ou mais nucleotídeos que ocorrem em locais de eliminação/adicação.

A hibridização cruzada tipicamente envolve a desnaturação das sequências alvo e de referência, por exemplo, por aquecimento, seguindo-se a renaturação sob condições (tais como a temperatura) que permitem a ocorrência da hibridização e formação do dúplex.

Tal como aqui utilizado, uma "mistura reacional" é uma mistura suspeita de conter um dúplex de sequência alvo que compreende um tampão adequado para permitir a desnaturação de uma sequência alvo.

Tal como aqui utilizado, "identidade" ou "homologia" refere-se à semelhança de sequências, ao nível das suas subunidades, entre duas moléculas poliméricas, por exemplo, dois polinucleicos ou dois polipéptidos. Quando uma posição da subunidade está ocupada pela mesma subunidade monomérica em ambas as moléculas, por exemplo, se uma posição em cada um dos dois péptidos é ocupada por serina, então estas são idênticas naquela posição. A identidade entre duas sequências é uma função direta do número de posições coincidentes ou idênticas, por exemplo, se metade (por exemplo, 5 posições num polímero de 10 subunidades de comprimento) das posições de duas sequências ou compostos peptídicos é idêntica, então as duas sequências são 50 % idênticas; se 90 % das posições, por exemplo, 9 de 10

subunidades, são coincidentes, as duas sequências partilham 90 % de identidade de sequência.

A percentagem de identidade de nucleotídeos pode ser determinada pelos parâmetros pré-definidos de BLAST. Para comparação de sequências, tipicamente uma sequência atua como uma sequência de referência com a qual as sequências de teste são comparadas. Quando se utiliza um algoritmo de comparação de sequências, as sequências de teste e de referência são introduzidas num computador, as coordenadas de subsequência são designadas se for necessário, e os parâmetros do programa do algoritmo de sequenciação são designados. Podem ser utilizados parâmetros pré-definidos do programa, ou podem ser designados parâmetros alternativos. O algoritmo de comparação de sequência calcula então a identidade de sequência percentual para as sequências de teste, face à sequência de referência e com base nos parâmetros do programa.

A janela de comparação inclui a referência a um segmento de qualquer uma das posições contíguas, selecionadas a partir do grupo que consiste de 20 a 600, usualmente de cerca de 50 a cerca de 200, mais usualmente de cerca de 100 a cerca de 150, posições contíguas, nas quais uma sequência pode ser comparada a uma sequência de referência com o mesmo número de posições contíguas, após as duas sequências serem otimamente alinhadas. Os métodos de alinhamento de sequências para comparação são bem conhecidos na especialidade. O alinhamento ótimo de sequências para comparação pode ser realizado, por exemplo, pelo algoritmo de homologia local de Smith & Waterman, Adv. Appl. Math. 2: 482 (1981), pelo algoritmo de alinhamento de homologia de Needleman & Wunsch, J. Mol. Biol. 48: 443 (1970), pelo método de pesquisa de similaridade de Pearson & Lipman, Proc. Nat'l. Acad. Sci. EUA 85: 2444 (1988), pelas

implementações computacionais destes algoritmos (GAP, BESTFIT, FASTA e TFASTA presentes no pacote de *software* Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, Wis.) ou pelo alinhamento manual e inspeção visual (ver, por exemplo, Current Protocols in Molecular Biology (Ausubel *et al.*, eds., 1995 suplemento)).

Exemplos de algoritmos que são adequados para determinar a percentagem de identidade de sequência e semelhança de sequências são os algoritmos BLAST e BLAST 2.0, que são descritos em Altschul *et al.*, Nuc. Acids Res. 25: 3389-3402 (1977) e em Altschul *et al.*, J. Mol. Biol. Biol. 215: 403-410 (1990), respetivamente. O *software* para a realização de análises segundo o algoritmo BLAST está disponível publicamente através do Centro Nacional para a Informação Biotecnológica (<http://www.ncbi.nlm.nih.gov/>). Inicialmente, este algoritmo envolve a identificação de pares de sequências de pontuação elevada (PSPEs) através da identificação de palavras curtas de comprimento W na sequência de procura, que correspondem ou satisfazem uma pontuação limite T de valor positivo, quando estas palavras são alinhadas face a uma palavra com o mesmo comprimento e pertencente a uma sequência da base de dados. T é referido como a pontuação limite da palavra vizinha (Altschul *et al.*, supra). Estes *hits* iniciais da palavra vizinha atuam como sementes para iniciar as pesquisas de PSPEs mais longos, que estejam contidos na sequência de procura. Os *hits* das palavras são estendidos em ambas as direções ao longo de cada sequência, tão longe quanto a pontuação de alinhamento cumulativa pode ser aumentada. As pontuações cumulativas são calculadas utilizando, para sequências nucleotídicas, os parâmetros M (pontuação de recompensa por um par de resíduos correspondente; sempre > 0) e N (pontuação de penalização para disparidade dos resíduos;

sempre < 0). Para sequências de aminoácidos, uma matriz de pontuação é usada para calcular a pontuação cumulativa. A extensão dos *hits* das palavras em cada direção é interrompida quando: a pontuação de alinhamento cumulativa diminui pela quantidade X face ao seu valor máximo alcançado; a pontuação cumulativa tende para zero ou inferior, devido à acumulação de um ou mais alinhamentos de resíduos com pontuações negativas; ou a extremidade de qualquer uma das sequências é atingida. Os parâmetros W , T e X do algoritmo BLAST determinam a sensibilidade e velocidade do alinhamento. O programa BLASTN (para sequências nucleotídicas) utiliza por defeito um comprimento de palavra (W) de 11, uma expectativa (E) de 10, $M=5$, $N=-4$ e uma comparação de ambas as cadeias. Para sequências de aminoácidos, o programa BLASTP utiliza por defeito um comprimento de palavra de 3, uma expectativa (E) de 10 e a matriz de pontuação BLOSUM62 (ver Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)), alinhamentos (B) de 50, expectativa (E) de 10, $M=5$, $N=-4$ e uma comparação de ambas as cadeias.

O algoritmo BLAST também realiza uma análise estatística da similaridade entre duas sequências (ver, por exemplo, Karlin & Altschul, Proc. Nat'l Acad. Sci. USA 90:5873-5787 (1993)). Uma medida da semelhança, fornecida pelo algoritmo BLAST, é a probabilidade da menor soma ($P(N)$), que proporciona uma indicação da probabilidade de correspondência, caso ocorresse por acaso, entre duas sequências nucleotídicas ou de aminoácidos. Por exemplo, um ácido nucleico é considerado semelhante a uma sequência de referência, se a probabilidade da menor soma, face à comparação entre o ácido nucleico em teste com o ácido nucleico de referência, é inferior a cerca de 0,2, mais preferencialmente inferior a cerca de 0,01 e ainda mais preferencialmente inferior a cerca de 0,001.

Num primeiro aspeto, a invenção é dirigida a um método para o enriquecimento de uma sequência alvo numa amostra de ácidos nucleicos, que é suspeita de conter sequências alvo e de referência. As sequências de referência e alvo podem ser amplificadas anteriormente à sua utilização no presente método. Isto é, as sequências de referência e alvo de interesse podem ser amplificadas a partir de um *template* genómico numa reação RCP, antes da sua utilização no presente método. Uma alíquota desta reação RCP é então transferida para utilização no método de enriquecimento seletivo. Alternativamente, as sequências de referência e alvo não necessitam de ser submetidas a uma primeira reação de RCP, mas podem ser utilizadas na sua forma nativa (por exemplo, ADN genómico) no método de enriquecimento seletivo. As sequências alvo e de referência podem ser obtidas a partir de qualquer sequência de ácidos nucleicos incluindo, ADN genómico, ADNc, ADN viral, ADN de mamíferos, ADN fetal ou ADN bacteriano. Embora a sequência de referência seja usualmente um alelo não mutante e a sequência alvo seja um alelo mutante, o contrário também pode ser verdadeiro. O alelo mutante pode incluir qualquer uma ou mais das eliminações, adições ou alterações nucleotídicas. Em algumas formas de realização, o alelo mutante é uma mutação somática. Em outras formas de realização, a sequência alvo é ADN metilado, enquanto a sequência de ADN de referência é um ADN não metilado. Alternativamente, a sequência alvo é um ADN não metilado, enquanto a sequência de referência é ADN metilado. Os iniciadores utilizados no presente método são geralmente concebidos, de modo a gerarem produtos de amplificação de sequências de referência e alvo com tamanhos de cerca de 17-1000 bases, mais preferencialmente de cerca de 25 a 500 bases, e ainda mais preferencialmente de cerca de 50 a 100 bases.

O método inclui a sujeição da mistura reacional de amplificação, a uma primeira temperatura de desnaturação que é superior à temperatura de fusão " T_f " de uma sequência de referência. A T_f de um ácido nucleico pode ser determinada por experimentação ou estimada por cálculo. O perito na especialidade é consciente dos vários métodos conhecidos para a determinação da T_f de um ácido nucleico, alguns dos quais são aqui descritos. A primeira temperatura de desnaturação é definida de acordo com processos convencionais usados na RCP. Assim, a primeira temperatura de desnaturação deve ser suficientemente elevada de modo a permitir a desnaturação completa das sequências alvo e de referência (por exemplo, 96 °C). Numa forma de realização, a primeira temperatura de desnaturação é de cerca de 1 °C a 30 °C superior à T_f da sequência de referência, mais preferencialmente, a T_f da sequência de referência é cerca de 5 °C a 20 °C superior à T_f da sequência de referência.

De seguida, a temperatura da mistura reacional de amplificação é reduzida permitindo que as sequências alvo e as sequências de referência possam hibridizar.

Esta temperatura de hibridização ou temperatura intermédia (sendo a temperatura inferior à primeira temperatura de desnaturação e T_c , mas superior à temperatura de emparelhamento/extensão do iniciador, por exemplo, cerca de 60 °C a 80 °C) é superior à T_f do par de iniciadores, e portanto permite que as sequências alvo e de referência possam hibridizar, enquanto é evitada a ligação do par de iniciadores às sequências alvo e/ou de referência. Esta etapa de emparelhamento, resulta na formação de dúplices híbridos de cadeia dupla de sequências alvo/alvo, referência/referência e de alvo/referência. Os dúplices híbridos de cadeia dupla de sequência referência-alvo são

então preferencialmente desnaturados, através do aumento da temperatura da mistura reacional para a T_c . A T_c ou temperatura crítica é inferior à T_f da sequência de referência, e pode ser determinada pelos métodos aqui descritos. Numa forma de realização, a T_c é de cerca de 0,3 °C a 5 °C inferior, e mais preferencialmente de cerca de 0,5 °C a 1,5 °C inferior, face à T_f da sequência de referência. Geralmente, a T_c será de cerca de 70-90 °C. Os dúplexes hibridados de sequência alvo/alvo também podem ser preferencialmente desnaturados, caso a sequência alvo possua uma sequência nucleotídica que resulte numa T_f inferior em comparação com a T_f da sequência de referência. Na T_c , os dúplexes de sequência alvo/referência (e apenas os dúplexes de sequência alvo/alvo que têm uma T_f mais baixa do que a T_f referente à sequência de referência) estão substancialmente desnaturados, ao passo que os dúplexes alvo/alvo (se tiverem uma T_f igual ou superior à T_f da sequência de referência) e os dúplexes de sequência referência/referência, estão substancialmente em formas não desnaturadas. O termo "substancialmente" significa pelo menos 60 %, preferencialmente pelo menos 70 %, mais preferencialmente pelo menos 80 %, ainda mais preferencialmente pelo menos 90 % e sobretudo preferencialmente pelo menos 98 %, numa determinada forma desnaturada ou não desnaturada. A T_c é geralmente aplicada entre cerca de 1 segundo a 5 minutos, mais preferencialmente entre 2 segundos a 1 minuto e ainda mais preferencialmente entre 5 segundos a 30 segundos.

Após a desnaturação preferencial dos dúplexes hibridados de sequência referência-alvo e/ou alvo/alvo, a temperatura da mistura reacional é reduzida de modo a permitir que um par de iniciadores possa emparelhar com a sequência alvo. Os iniciadores emparelhados são então estendidos por uma polimerase de ácidos nucleicos, enriquecendo assim a

sequência alvo em relação à sequência de referência da amostra.

As etapas do método são geralmente repetidas em ciclos múltiplos, de forma a obter uma amplificação suficiente das sequências alvo e de referência. Numa forma de realização, as etapas do método são repetidas em 5-40 ciclos, e mais preferencialmente em 10-30 ciclos. O número ótimo de ciclos pode ser determinado por um perito comum na especialidade. Preferencialmente, os métodos presentes são realizados num dispositivo de RCP, mais preferencialmente sob condições de tempo real num dispositivo de RCP de detecção em tempo real, tal como o dispositivo de RCP em tempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) e o dispositivo de RCP em tempo real Mx3005P (Stratagene, La Jolla, CA). Nesta forma de realização, a mistura reacional pode incluir um agente de detecção de ácidos nucleicos (por exemplo, um corante de detecção de ácidos nucleicos, tal como o corante verde-SYBR ou o corante verde-LC ou uma sonda acoplada operativamente a um marcador fluorescente), para a quantificação e/ou monitorização dos produtos de amplificação da reação. Uma vez que o enriquecimento da sequência alvo esteja completo, a amostra pode ser adicionalmente processada (por exemplo, para a identificação de quaisquer alterações genéticas enriquecidas pelo método, por exemplo, sujeitadas a uma reação de sequenciação).

As sequências de referência enriquecidas podem adicionalmente ser processadas, segundo uma variedade de procedimentos que incluem: MALDI-TOF, fusão de alta resolução, sequenciação didesoxi, sequenciação de molécula única, pirosequenciação, PFR, RCP digital e RCP quantitativo.

É ainda divulgado um método para o enriquecimento de uma sequência alvo, segundo a realização de um protocolo para uma reação de amplificação de ácidos nucleicos. O protocolo para a reação de amplificação inclui uma primeira temperatura de desnaturação e uma segunda temperatura de desnaturação. A primeira temperatura de desnaturação é superior à T_f da sequência de referência, e a segunda temperatura de desnaturação é inferior à T_f da sequência de referência. O método inclui a sujeição da mistura reacional de amplificação a uma primeira temperatura de desnaturação, que está acima da temperatura de fusão " T_f " de uma sequência de referência. O perito na especialidade é consciente dos vários métodos conhecidos para a determinação da T_f de um ácido nucleico. A primeira temperatura de desnaturação é geralmente selecionada, da mesma forma que alguém selecionaria usualmente a temperatura de desnaturação de uma reação RCP, e deve ser suficientemente elevada de modo a permitir a desnaturação das sequências alvo e de referência. Numa forma de realização, a primeira temperatura de desnaturação é de cerca de 1 °C a 30 °C superior à T_f da sequência de referência, mais preferencialmente, a T_f da sequência de referência é cerca de 5 °C a 20 °C superior à T_f da sequência de referência.

A segunda temperatura de desnaturação é inferior à T_f da sequência de referência, e pode ser determinada pelos métodos aqui descritos. Numa forma de realização, a T_c é cerca de 0,3 °C a 5 °C inferior à T_f da sequência de referência, e mais preferencialmente cerca de 0,5 °C a 1,5 °C inferior à T_f da sequência de referência. Geralmente, a T_c estará entre cerca de 70-90 °C. A segunda temperatura de desnaturação é geralmente aplicada a partir de cerca de 1 segundo a 5 minutos, mais preferencialmente entre 2

segundos a 1 minuto e ainda mais preferencialmente entre 5 segundos a 30 segundos.

É ainda descrito um método de enriquecimento de uma sequência alvo através da sujeição de uma mistura reacional de amplificação a uma T_c , da redução da temperatura da mistura reacional e da extensão de um par de iniciadores. A mistura reacional de amplificação é suspeita de conter uma sequência alvo e uma sequência de referência. Neste aspeto, a sequência alvo tem uma T_f inferior à T_f da sequência de referência. A T_c é inferior à T_f da sequência de referência, permitindo assim a desnaturação preferencial da sequência alvo que possui uma T_f inferior à T_f da sequência de referência, como resultado da sua composição nucleotídica (por exemplo, supressão). Com respeito aos outros aspetos da invenção, as sequências de referência e alvo podem ser amplificadas anteriormente à sua utilização no presente método. Isto é, as sequências de referência e alvo de interesse podem ser amplificadas a partir de um *template* genómico numa reação RCP, antes da sua utilização no presente método. Uma alíquota desta reação RCP é então transferida para utilização no método de enriquecimento seletivo. Alternativamente, as sequências de referência e alvo não necessitam de ser submetidas a uma primeira reação de RCP, mas podem ser utilizadas na sua forma nativa no método de enriquecimento seletivo, por exemplo, ADN genómico. As sequências alvo e de referência podem ser obtidas a partir de qualquer sequência de ácidos nucleicos incluindo, ADN genómico, ADNc, ADN viral, ADN de mamíferos, ADN fetal ou ADN bacteriano. Embora a sequência de referência seja usualmente um alelo não mutante e a sequência alvo seja um alelo mutante, o contrário também pode ser verdadeiro. O alelo mutante pode incluir qualquer uma ou mais das eliminações, adições ou alterações nucleotídicas. Em algumas formas de realização, o alelo

mutante é uma mutação somática. Em outras formas de realização, a sequência alvo é ADN metilado, enquanto a sequência de ADN de referência é um ADN não metilado. Alternativamente, a sequência alvo é um ADN não metilado, enquanto a sequência de referência é ADN metilado. Os iniciadores utilizados no presente método são geralmente concebidos, de modo a gerarem produtos de amplificação de sequências de referência e alvo com tamanhos de cerca de 15 a 1000 bases, mais preferencialmente de cerca de 25 a 500 bases, e ainda mais preferencialmente de cerca de 50 a 100 bases.

Os dúplexes hibridizados alvo/alvo são preferencialmente desnaturados pelo aumento da temperatura da mistura reacional para a T_c . A T_c ou temperatura crítica é inferior à T_f da sequência de referência, e pode ser determinada pelos métodos aqui descritos. Numa forma de realização, a T_c é de cerca de 0,3 °C a 5 °C inferior face à T_f da sequência de referência, e mais preferencialmente de cerca de 0,5 °C a 1,5 °C inferior, face à T_f da sequência de referência. Geralmente, a T_c será de cerca de 70-90 °C. A T_c é geralmente aplicada entre cerca de 1 segundo a 5 minutos, mais preferencialmente entre 2 segundos a 1 minuto e ainda mais preferencialmente entre 5 segundos a 30 segundos. Na T_c , os dúplexes de sequência alvo/referência e os dúplexes de sequência alvo/alvo estão substancialmente desnaturados, enquanto os dúplexes de sequência referência/referência estão substancialmente em formas não desnaturadas. O termo "substancialmente" significa pelo menos 60 %, preferencialmente pelo menos 70 %, mais preferencialmente pelo menos 80 %, ainda mais preferencialmente pelo menos 90 % e sobretudo preferencialmente pelo menos 98 %, numa determinada forma desnaturada ou não desnaturada.

A etapa de redução da temperatura da mistura reacional permite que o par de iniciadores emparelhe com a sequência alvo. Estes iniciadores emparelhados são posteriormente estendidos por uma polimerase, aumentando a quantidade de sequência alvo na amostra relativamente à sequência de referência. As etapas do método são geralmente repetidas em ciclos múltiplos, de forma a obter uma amplificação suficiente das sequências alvo e de referência. Numa forma de realização, as etapas do método são repetidas em 5-40 ciclos, e mais preferencialmente em 10-30 ciclos. O número ótimo de ciclos pode ser determinado por um perito comum na especialidade. Preferencialmente, os métodos presentes são realizados num dispositivo de RCP, mais preferencialmente sob condições de tempo real num dispositivo de RCP de detecção em tempo real, tal como o dispositivo de RCP em tempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) e o dispositivo de RCP em tempo real Mx3005P (Stratagene, La Jolla, CA). Nesta forma de realização, a mistura reacional pode incluir um agente de detecção de ácidos nucleicos (por exemplo, um corante de detecção de ácidos nucleicos, tal como o corante verde-SYBR ou o corante verde-LC ou uma sonda acoplada operativamente a um marcador fluorescente), para a quantificação e/ou monitorização dos produtos de amplificação da reação. Uma vez que o enriquecimento da sequência alvo esteja completo, a amostra pode ser adicionalmente processada, por exemplo, sujeitada a uma reação de sequenciação. Os alelos enriquecidos podem ser adicionalmente processados por uma variedade de procedimentos que incluem: MALDI-TOF, fusão de alta resolução, sequenciação didesoxi, sequenciação de molécula única, pirosequenciação, PFR, RCP digital e RCP quantitativo.

Ainda, é adicionalmente divulgado um método de enriquecimento de uma sequência alvo, submetendo uma

mistura reacional de amplificação a etapas alternadas de condições de emparelhamento e de condições de desnaturação por aplicação da T_c . A mistura reacional de amplificação, possuindo uma sequência alvo e de referência, é inicialmente submetida a uma primeira temperatura de desnaturação, que é superior à T_f da sequência de referência. Tal como nos outros aspetos, as sequências de referência e alvo podem ser amplificadas anteriormente à sua utilização no presente método. Isto é, as sequências de referência e alvo de interesse podem ser amplificadas a partir de um *template* genómico numa reação RCP, antes da sua utilização no presente método. Uma alíquota desta reação RCP é então transferida para utilização no método de enriquecimento seletivo. Alternativamente, as sequências de referência e alvo não necessitam de ser submetidas a uma primeira reação de RCP, mas podem ser utilizadas na sua forma nativa no método de enriquecimento seletivo, por exemplo, ADN genómico. As sequências alvo e de referência podem ser obtidas a partir de qualquer sequência de ácidos nucleicos incluindo, ADN genómico, ADNc, ADN viral, ADN de mamíferos, ADN fetal ou ADN bacteriano. Embora a sequência de referência seja usualmente um alelo não mutante e a sequência alvo seja um alelo mutante, o contrário também pode ser verdadeiro. O alelo mutante pode incluir qualquer uma ou mais das eliminações, adições ou alterações nucleotídicas. Em algumas formas de realização, o alelo mutante é uma mutação somática. Em outras formas de realização, a sequência alvo é ADN metilado, enquanto a sequência de ADN de referência é um ADN não metilado. Alternativamente, a sequência alvo é um ADN não metilado, enquanto a sequência de referência é ADN metilado. Os iniciadores utilizados no presente método são geralmente concebidos, de modo a gerarem produtos de amplificação de sequências de referência e alvo com tamanhos de cerca de 1 a 1000 bases, mais preferencialmente de cerca de 25 a 500

bases, e ainda mais preferencialmente de cerca de 50 a 100 bases.

A T_f de um ácido nucleico pode ser determinada por experimentação ou estimada por cálculo. O perito na especialidade é consciente dos vários métodos conhecidos para a determinação da T_f de um ácido nucleico. A primeira temperatura de desnaturação é geralmente selecionada, da mesma forma que alguém selecionaria usualmente a temperatura de desnaturação de uma reação RCP, e deve ser suficientemente elevada de modo a permitir a desnaturação das sequências alvo e de referência. Numa forma de realização, a primeira temperatura de desnaturação é de cerca de 1 °C a 30 °C superior à T_f da sequência de referência, mais preferencialmente, a T_f da sequência de referência é cerca de 5 °C a 20 °C superior à T_f da sequência de referência.

De seguida, a amostra percorre um ciclo repetitivo entre duas etapas de incubação a temperatura diferentes. Na primeira etapa de incubação, a temperatura é reduzida de modo a permitir a hibridização da sequência alvo com a sequência de referência, formando um dúplex. Na segunda etapa de incubação, a temperatura é aumentada para a T_c , a qual é inferior à T_f da sequência de referência. Estas primeiras e segundas etapas são então repetidas uma ou mais vezes, mais preferencialmente 3-20 vezes e ainda mais preferencialmente 5-10 vezes.

A primeira etapa de incubação resulta na formação de dúplexes hibridados de sequências alvo/alvo, referência/referência e alvo/referência. Numa forma de realização preferencial, esta temperatura de hibridização ou temperatura intermédia (sendo a temperatura inferior à primeira temperatura de desnaturação e T_c , mas superior à

temperatura de emparelhamento/extensão do iniciador, por exemplo, cerca de 60 °C a 80 °C) é superior à T_f do par de iniciadores, e portanto permite que as sequências alvo e de referência possam hibridizar, enquanto é evitada a ligação do par de iniciadores às sequências alvo e/ou de referência. Os dúplexes hibridados referência-alvo e alvo/alvo (enquanto o alvo tiver uma T_f inferior à T_f da sequência de referência) são então preferencialmente desnaturados, pelo aumento da temperatura da mistura reacional para o valor da T_c , durante a segunda etapa de incubação. A T_c ou temperatura crítica é inferior à T_f da sequência de referência, e pode ser determinada pelos métodos aqui descritos. Numa forma de realização, a T_c é de cerca de 0,3 °C a 5 °C inferior face à T_f da sequência de referência, e mais preferencialmente de cerca de 0,5 °C a 1,5 °C inferior, face à T_f da sequência de referência. Geralmente, a T_c será de cerca de 70-90 °C. A T_c é geralmente aplicada entre cerca de 1 segundo a 5 minutos, mais preferencialmente entre 2 segundos a 1 minuto e ainda mais preferencialmente entre 5 segundos a 30 segundos. Uma vez que a etapa do ciclo de incubação esteja completa, a temperatura da mistura reacional é diminuída de modo a permitir que, um ou mais iniciadores possam emparelhar com a sequência alvo. Estes iniciadores são então estendidos por uma polimerase, enriquecendo assim a sequência alvo.

Uma vez que cada etapa esteja completa, a reação pode ser repetida em múltiplos ciclos, de forma a obter uma amplificação suficiente das sequências alvo e de referência. Numa forma de realização, as etapas do método são repetidas em 5-40 ciclos, e mais preferencialmente em 10-30 ciclos. O número ótimo de ciclos pode ser determinado por um perito comum na especialidade. Preferencialmente, os métodos presentes são realizados num dispositivo de RCP, mais preferencialmente sob condições de tempo real num

dispositivo de RCP de detecção em tempo real, tal como o dispositivo de RCP em tempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) e o dispositivo de RCP em tempo real Mx3005P (Stratagene, La Jolla, CA). Nesta forma de realização, a mistura reacional pode incluir um agente de detecção de ácidos nucleicos (por exemplo, um corante de detecção de ácidos nucleicos, tal como o corante verde-SYBR ou o corante verde-LC ou uma sonda acoplada operativamente a um marcador fluorescente), para a quantificação e/ou monitorização dos produtos de amplificação da reação. Uma vez que o enriquecimento da sequência alvo esteja completo, a amostra pode ser adicionalmente processada. Formas de processamento adicionais incluem: MALDI-TOF, fusão de alta resolução, sequenciação didesoxi, sequenciação de molécula única, pirosequenciação, PFR, RCP digital e RCP quantitativo.

Numa outra forma de realização, o método da presente invenção pode ser utilizado para detetar se ocorreu a metilação numa sequência alvo ou numa sequência de referência. Numa outra forma de realização, o método utilizado é ADN genómico para testar a metilação.

O método de detecção da metilação compreende uma abordagem química ou enzimática para o tratamento de sensibilidade à metilação de ADN. Os tratamentos químicos incluem a incubação de ADN com bissulfito de sódio, o qual converte seletivamente citosinas não metiladas para uracilos. O ADN é inicialmente desnaturado por choque térmico, e posteriormente pode ser tratado com 5 M de bissulfito, pH 5-7. O pré-tratamento de ADN genómico, para a remoção de uracilos pré-existent, é realizado anteriormente ao tratamento com bissulfito. Este pré-tratamento consiste no tratamento com uracil glicosilase na presença de 5 mM de

hidroxilamina, pH 7. O ADN modificado pode então ser utilizado nos métodos da invenção.

Dado que as citosinas metiladas, presentes nas sequência de referência ou alvo, são convertidas em uracilos, estas irão formar disparidades quando dúplexadas com a cadeia oposta (alvo ou de referência), a qual não está metilada, durante a etapa de hibridização cruzada da reação.

Ainda outro aspeto, qualquer um dos métodos da invenção é usado para enriquecer múltiplas sequências alvo diferentes segundo uma reação multi-plex. Nesta forma de realização, o método inclui conjuntos adicionais de pares de iniciadores para as sequências alvo adicionais.

Num outro aspeto, a invenção é dirigida a um formato legível informaticamente que contém instruções de um *software*, para a realização de qualquer um dos métodos da invenção. É ainda descrito um sistema de RCP para o enriquecimento da sequência alvo. O sistema inclui uma memória para a execução das instruções do *software* do formato legível informaticamente.

As Figuras 1 e 2 ilustram dois aspetos diferentes da invenção. A Figura 1 ilustra um aspeto da invenção em que o método utiliza uma reação de amplificação, que tem uma primeira temperatura de desnaturação e uma temperatura de desnaturação crítica ou T_c . A Figura 2 também ilustra uma reação de amplificação que tem uma primeira temperatura de desnaturação e uma T_c , mas inclui ainda etapas de temperaturas de desnaturação ou de emparelhamento oscilantes, ou repetitivas, que ocorrem múltiplas vezes, anteriormente às etapas de emparelhamento dos iniciadores e de extensão da reação.

A Figura 1 demonstra o processo para o enriquecimento de uma sequência alvo numa amostra de ácidos nucleicos com sequências alvo e de referência. As sequências alvo e de referência podem ser obtidas a partir de qualquer sequência de ácidos nucleicos incluindo, ADN genómico, ADNc, ADN viral, ADN de mamíferos, ADN fetal ou ADN bacteriano. Embora a sequência de referência seja usualmente o alelo e a sequência alvo seja um alelo mutante, o contrário também pode ser verdadeiro. O alelo mutante pode incluir qualquer uma ou mais das eliminações, adições ou alterações nucleotídicas. Em algumas formas de realização, o alelo mutante é uma mutação somática. Em outras formas de realização, a sequência alvo é ADN não metilado, enquanto a sequência de referência é um ADN metilado.

O método inclui a sujeição da mistura reacional de amplificação a uma primeira temperatura de desnaturação (Fig. 1A), que é superior à temperatura de fusão " T_f " de uma sequência de referência. O perito na especialidade é consciente dos vários métodos conhecidos para a determinação da T_f de um ácido nucleico. A primeira temperatura de desnaturação é geralmente selecionada, da mesma forma que alguém selecionaria usualmente a temperatura de desnaturação de uma reação RCP, e deve ser suficientemente elevada de modo a permitir a desnaturação das sequências alvo e de referência (por exemplo, 94 °C). Numa forma de realização, a primeira temperatura de desnaturação é de cerca de 1 °C a 30 °C superior à T_f da sequência de referência, mais preferencialmente, a T_f da sequência de referência é cerca de 5 °C a 20 °C superior à T_f da sequência de referência.

De seguida, a temperatura da mistura reacional de amplificação é reduzida permitindo que as sequências alvo e as sequências de referência possam hibridizar (Fig. 1B).

Esta etapa de emparelhamento resulta na formação de dúplexes hibridados de sequências alvo/alvo, referência/referência e alvo/referência. A determinação da temperatura de emparelhamento é conhecida pelos peritos na especialidade. Os iniciadores de RCP utilizados no método, são concebidos para terem uma T_f que os impeça de se ligarem a sequências alvo e de referência a esta temperatura intermédia, de modo a que estes não interfiram com a hibridização cruzada das sequências mutante (alvo) e não mutante (referência). Devido às mutações na sequência alvo, a maioria das sequências alvo acaba numa estrutura de disparidades face à sequência de referência, e consequentemente tem uma temperatura de fusão mais baixa, quando heterodúplexada com a sequência de referência, do que a temperatura de homodúplexes de referência/referência totalmente emparelhados.

Os dúplexes hibridados alvo/referência são então preferencialmente desnaturados pelo aumento da temperatura da mistura reacional para a T_c (Fig. 1C). A T_c ou temperatura crítica é inferior à T_f da sequência de referência, e pode ser determinada pelos métodos aqui descritos. Numa forma de realização, a T_c é de cerca de 0,3 °C a 5 °C inferior face à T_f da sequência de referência, e mais preferencialmente de cerca de 0,5 °C a 1,5 °C inferior, face à T_f da sequência de referência. Geralmente, a T_c será de cerca de 70-90 °C. Os dúplexes hibridados alvo/alvo também podem ser preferencialmente desnaturados, caso a sequência alvo possua uma sequência nucleotídica que resulte numa T_f inferior em comparação com a sequência de referência. A T_c é geralmente aplicada entre cerca de 1 segundo a 5 minutos, mais preferencialmente entre 2 segundos a 1 minuto e ainda mais preferencialmente entre 5 segundos a 30 segundos.

Após a desnaturação preferencial dos dúplexes hibridados de sequência alvo/referência e/ou alvo/alvo, a temperatura da mistura reacional é reduzida de modo a permitir que um ou mais iniciadores possam emparelhar com a sequência alvo (Fig. 1D). Os iniciadores emparelhados são então estendidos por uma polimerase de ácidos nucleicos, enriquecendo assim a sequência alvo em relação à sequência de referência da amostra.

As etapas do método são geralmente repetidas em ciclos múltiplos, de forma a obter uma amplificação suficiente das sequências alvo e de referência. Numa forma de realização, as etapas do método são repetidas em 5-40 ciclos, e mais preferencialmente em 10-30 ciclos. O número ótimo de ciclos pode ser determinado por um perito comum na especialidade. Preferencialmente, os métodos presentes são realizados num dispositivo de RCP, mais preferencialmente sob condições de tempo real num dispositivo de RCP de detecção em tempo real, tal como o dispositivo de RCP em tempo real SMARTCYCLER (Cepheid, Sunnyvale, CA) e o dispositivo de RCP em tempo real Mx3005P (Stratagene, La Jolla, CA). Nesta forma de realização, a mistura reacional pode incluir um agente de detecção de ácidos nucleicos (por exemplo, um corante de detecção de ácidos nucleicos, tal como o corante verde-SYBR ou o corante verde-LC ou uma sonda acoplada operativamente a um marcador fluorescente), para a quantificação e/ou monitorização dos produtos de amplificação da reação. Uma vez que o enriquecimento da sequência alvo esteja completo, a amostra pode ser adicionalmente processada, por exemplo, sujeitada a uma reação de sequenciação. Os alelos enriquecidos podem ser adicionalmente processados por uma variedade de procedimentos que incluem: MALDI-TOF, fusão de alta resolução, sequenciação didesoxi, sequenciação de molécula única, pirosequenciação, PFR, RCP digital e RCP quantitativo.

Ao realizar o método de enriquecimento em cada ciclo RCP, a quantidade de sequências mutantes (sequência alvo) é enriquecida gradualmente ao longo das sequências (sequência de referência). Ambas as mutações homozigóticas e heterozigóticas são enriquecidas através do método. É frequente o enriquecimento em 10-60 vezes das sequências que contêm mutações, em relação à realização de RCP convencional para uma temperatura de desnaturação de 94 °C. A uma dada temperatura crítica de desnaturação (T_c), o enriquecimento das mutações ocorre simultaneamente em todas as posições da sequência, embora com uma eficiência diferente que depende do contexto da sequência e do tamanho total do produto de amplificação da RCP. Tanto a temperatura de desnaturação crítica T_c , como o enriquecimento antecipado em qualquer posição, são previsíveis através da utilização de um *software* de fusão de ADN apropriado e podem ser verificados experimentalmente. Assim, dependendo do local onde a mutação ocorre, um enriquecimento ligeiramente diferente pode ser antecipado, contudo em todos os casos é atingido um enriquecimento substancial, e consequentemente os limites de detecção dos ensaios são melhorados, por exemplo, nas reações de sequenciação.

A Figura 2 ilustra uma forma de realização do método de enriquecimento de uma sequência alvo, submetendo uma mistura reacional de amplificação a múltiplas etapas alternadas de condições de emparelhamento e de condições de temperatura de desnaturação crítica. Esta forma de realização retira vantagens tanto da desnaturação preferencial à temperatura crítica, bem como da hibridização cruzada preferencial à temperatura de hibridização.

A mistura reacional de amplificação, com uma sequência alvo e de referência, é inicialmente submetida a uma primeira temperatura de desnaturação que é superior à T_f da sequência de referência (Fig. 2A).

De seguida, a amostra percorre um ciclo repetitivo entre duas etapas de incubação a temperatura diferentes. Na primeira etapa de incubação, a temperatura é reduzida de modo a permitir a hibridização da sequência alvo com a sequência de referência (Fig. 2B). Na segunda etapa de incubação, a temperatura é aumentada para a T_c (Fig. 2C). Estas primeiras e segundas etapas são então repetidas uma ou mais vezes, mais preferencialmente 3-20 vezes e ainda mais preferencialmente 5-10 vezes.

A temperatura de hibridização preferencial para uma dada sequência é a temperatura em que os alelos não mutantes hibridizam de volta a si mesmo, verificando uma velocidade mais rápida do que no caso de alelos que contêm mutações. Porque os alelos mutados estão numa prevalência muito menor do que os alelos não mutados, a hibridização cruzada dos alelos não mutantes com eles próprios ocorre mais rapidamente do que hibridização cruzada dos alelos mutantes com alelos mutantes, ou de alelos mutantes com alelos (os últimos formando uma disparidade). Como resultado, quando a temperatura da RCP é reduzida para uma temperatura de hibridização cruzada, os alelos mutantes não hibridizam cruzadamente com a mesma extensão que aquela verificada em alelos não mutantes. Esta temperatura de hibridização ou temperatura intermédia (sendo a temperatura inferior à primeira temperatura de desnaturação e T_c , mas superior à temperatura de emparelhamento/extensão do iniciador, por exemplo, cerca de 60 °C a 80 °C) é superior à T_f do par de iniciadores, e portanto permite que as sequências alvo e de referência possam hibridizar, enquanto é evitada a ligação

do par de iniciadores às sequências alvo e/ou de referência.

De seguida, a temperatura da reação é aumentada para a T_c , resultando em dúplexes hibridados de sequência alvo/referência e alvo/alvo preferencialmente desnaturados (Fig. 2C). A T_c ou temperatura crítica é inferior à T_f da sequência de referência, e pode ser determinada pelos métodos aqui descritos. Numa forma de realização, a T_c é de cerca de 0,3 °C a 5 °C inferior face à T_f da sequência de referência, e mais preferencialmente de cerca de 0,5 °C a 1,5 °C inferior, face à T_f da sequência de referência. Geralmente, a T_c será de cerca de 70-90 °C. Os dúplexes hibridados alvo/alvo também podem ser preferencialmente desnaturados, caso a sequência alvo possua uma sequência nucleotídica que resulte numa T_f inferior em comparação com a sequência de referência. A T_c é geralmente aplicada entre cerca de 1 segundo a 5 minutos, mais preferencialmente entre 2 segundos a 1 minuto e ainda mais preferencialmente entre 5 segundos a 30 segundos. Este processo é repetido por diversas vezes, oscilando entre a temperatura de emparelhamento (Fig. 2B) e a temperatura crítica (Fig. 2C), o qual gera seletivamente a cada vez, mais sequências mutantes na forma de cadeia simples comparativamente com as sequências na forma de cadeia simples.

Uma vez que a etapa do ciclo de incubação esteja completa, a temperatura da mistura reacional é diminuída de modo a permitir que, um ou mais iniciadores possam emparelhar com a sequência alvo (Fig. 2D). Estes iniciadores são então estendidos por uma polimerase, enriquecendo assim a sequência alvo. As etapas do método são geralmente repetidas em ciclos múltiplos, de forma a obter uma amplificação suficiente das sequências alvo e de referência. Numa forma de realização, as etapas do método

são repetidas em 5-40 ciclos, e mais preferencialmente em 10-30 ciclos. O número ótimo de ciclos pode ser determinado por um perito comum na especialidade. Preferencialmente, os métodos presentes são realizados num dispositivo de RCP, mais preferencialmente sob condições reacionais de tempo real num dispositivo de RCP de detecção em tempo real.

Uma vez que o enriquecimento da sequência alvo esteja completo, a amostra pode ser adicionalmente processada. Formas de processamento adicionais incluem: MALDI-TOF, fusão de alta resolução, sequenciação didesoxi, sequenciação de molécula única, pirosequenciação, PFR, RCP digital e RCP quantitativo.

A RCP digital pode ser utilizada na detecção de mutações de com níveis de detecção ultrabaixos em combinação com um procedimento de enriquecimento. Na RCP digital, a amostra de ADN é diluída até à obtenção de moléculas individuais, de modo que em cada reação RCP o material de partida seja não mutante ou mutante. Após um grande número de reações RCP partindo do mesmo material de partida, as moléculas mutantes são isoladas e detetadas. A empresa Fluidigm (South San Francisco, CA) vende uma variedade de sistemas baseados em RCP digitais, que podem ser utilizados com a presente invenção. Deste modo, o enriquecimento em tempo real pode ser realizado a partir de moléculas individuais em milhares de reações de enriquecimento paralelas e simultâneas, permitindo a identificação da reação RCP que originou o ADN mutante. Tal sistema é particularmente útil para a detecção de mutações com níveis de expressão ultrabaixos em genomas cancerígenos. A combinação de RCP digital com o processo de enriquecimento, pode beneficiar de igual forma aplicações de sequenciação de moléculas individuais.

Reação de amplificação de ácido nucleico

Numa forma de realização, uma amostra de ácidos nucleicos utilizada no método da presente invenção, compreende ADN genómico que possui uma sequência alvo e uma sequência de referência. Numa outra forma de realização, a amostra de ácidos nucleicos do método da invenção compreende uma sequência alvo e de referência, que foram previamente amplificadas numa reação de amplificação de ácidos nucleicos. O perito na especialidade irá apreciar que existem vários métodos disponíveis para amplificar um ácido nucleico. Talvez o método mais popular seja a reação em cadeia da polimerase (RCP; por exemplo, ver, Patentes E.U.A. N°s 4,683,195 e 4,683,202, bem como Saiki *et al.*, Science 230:1350-1354 (1985) e Gyllensten *et al.*, PNAS (USA) 85:7652-7656 (1985)). Uma variação preferencial do método da RCP é a RCP assimétrica (por exemplo, ver 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); e Patente E.U.A. N° 6,197,499). Outros métodos de amplificação incluem, mas não estão limitados a, amplificação por deslocamento de cadeia (ADC) (ver, Walker *et al.*, Nuc. Acids Res. 20(7):1691-1696 (1992), bem com as Patentes E.U.A. N°s. 5,744,311, 5,648,211 e 5,631,147), amplificação por círculo rolante (ACR) (ver Patente PCT WO 97/19193), amplificação com base em sequência de ácidos nucleicos (ABSAN) (ver Compton, Nature 350:91-92 (1991); bem como as Patentes E.U.A. N°s. 5,409,818 e 5,554,527), amplificação mediada por transcrição (AMT) (ver Kwoh *et al.*, PNAS (USA) 86:1173-1177 (1989), bem com a Patente E.U.A N° 5,399,491), replicação de sequenciação autossustentada (3SR) (ver Guatelli *et al.*, PNAS (USA) 87:1874-1879 (1990)) e reação

em cadeia de ligase (RCL) (ver Patentes E.U.A. N°s. 5,427,930 e 5,792,607).

O presente método utiliza uma RCP modificada. A RCP é efetuada tal como descrito em Mullis e Faloona, 1987, *Methods Enzymol.*, 155: 335. A técnica de reação em cadeia da polimerase (RCP), é divulgada nas Patentes E.U.A. N°s. 4,683,202, 4,683,195 e 4,800,159. Na sua forma mais simples, a RCP é um método *in vitro* para a síntese enzimática de sequências específicas de ADN, utilizando dois iniciadores oligonucleicos que hibridizam com cadeias opostas e flanqueiam a região de interesse no ADN alvo. Uma série repetitiva de etapas reacionais que envolvem a desnaturação do *template*, o emparelhamento dos iniciadores e a extensão dos iniciadores emparelhados pela polimerase do ADN, resulta na acumulação exponencial de um fragmento específico, cujos terminais são definidos pelas extremidades 5' dos iniciadores. A RCP é reportada como sendo capaz de produzir um enriquecimento seletivo referente a uma sequência específica de ADN, segundo um fator de 10^9 . O método da RCP também está descrito em Saiki *et al.*, 1985, *Science* 230:1350.

A RCP é realizada utilizando ADN *template* (sequências alvo e de referência) (pelo menos 1 fg; mais utilmente 1-1000 ng) e pelo menos 25 pmol de iniciadores oligonucleicos. Uma mistura reacional típica inclui: 2 µL de ADN, 25 pmol de iniciador oligonucleico, 2,5 µL de um tampão adequado, 0,4 µL de 1,25 µM de dente, 2,5 unidades de Taq ADN polimerase (Stratagene) e água desionizada, perfazendo um volume total de 25 µL. A RCP é realizada utilizando um termociclador programável.

O comprimento e temperatura de cada etapa de um ciclo de RCP, bem como o número de ciclos, são ajustados de acordo com os requisitos de rigor para o efeito. A temperatura de emparelhamento e temporização são determinadas tanto pela eficácia com a qual um iniciador é esperado hibridizar com um *template*, bem como pelo grau de disparidade que deve ser tolerado. A capacidade para otimizar o rigor das condições de hibridização do iniciador, está dentro dos conhecimentos de uma pessoa com capacidades moderadas na técnica. É usada uma temperatura de emparelhamento entre 30 °C a 72 °C. A desnaturação inicial das moléculas do *template* ocorre normalmente entre 92 °C a 99 °C durante 4 minutos, seguida de 20-40 ciclos que consistem na desnaturação (94-99 °C durante 15 segundos a 1 minuto), emparelhamento (temperatura determinada como discutido acima; 1-2 minutos) e extensão (72 °C durante 1 minuto). A etapa final da extensão é geralmente realizada durante 4 minutos a 72 °C, e pode ser seguida por uma etapa indefinida (0-24 horas) a 4 °C.

A RCP utiliza uma polimerase de ácidos nucleicos ou uma enzima, que catalisa a polimerização dos nucleosídeos trifosfato. Usualmente, a enzima irá iniciar a síntese na extremidade 3' do iniciador emparelhado com a sequência alvo, e irá prosseguir na direção 5' ao longo do *template*. ADN polimerases conhecidas incluem, por exemplo, *E. coli* ADN polimerase I, ADN polimerase T7, *Thermus thermophilus* (Tth) ADN polimerase, *Bacillus stearothermophilus* ADN polimerase, *Thermococcus litoralis* ADN polimerase, *Thermus aquaticus* (Taq) ADN polimerase e *Pyrococcus furiosus* (Pfu) ADN polimerase. O termo "polimerase de ácidos nucleicos" também abrange as polimerases de ARN. Se o *template* de ácido nucleico é ARN, então consequentemente "polimerase de ácidos nucleicos" refere-se a uma atividade de

polimerização dependente de ARN, tal como uma transcriptase reversa.

Nos métodos da invenção, o protocolo da RCP inclui adicionalmente uma etapa de desnaturação crítica. Estes protocolos são descritos aqui na sua totalidade e são ilustrados nas Figuras 1 e 2.

Preferencialmente, os métodos presentes são realizados num dispositivo de RCP, mais preferencialmente sob condições reacionais de tempo real num dispositivo de RCP de detecção em tempo real. As condições reacionais de tempo real utilizam adicionalmente um agente de detecção de ácidos nucleicos (por exemplo, um corante ou uma sonda) para medir/detetar o produto da RCP do modo como é produzido.

Numa forma de realização, os métodos de enriquecimento são praticados num formato multi-plex. A RCP multi-plex é quando mais do que um par de iniciadores é usado numa reação RCP, de modo a detetar mais do que uma sequência alvo. O objetivo da RCP multi-plex é amplificar mais do que uma sequência alvo em simultâneo, e consequentemente poupar tempo e minimizar despesas. A RCP multi-plex permite a amplificação de múltiplas sequências alvo num ensaio apenas. Usualmente, a utilização da RCP multi-plex em conjunção com a presente invenção, com a finalidade de enriquecer em simultâneo os alelos em menores quantidades para todas as sequências amplificadas, implica que os iniciadores sejam concebidos de modo a que os produtos de amplificação resultantes da RCP partilhem a mesma T_c (temperatura de desnaturação crítica).

A determinação da T_f e T_c

A T_f pode ser definida como a temperatura à qual metade dos pares de bases de Watson-Crick, numa molécula dúplice de ácidos nucleicos de cadeia dupla, é quebrada ou dissociada (isto é, são "fundida"), enquanto a outra metade dos pares de bases de Watson-Crick permanece intacta numa conformação de cadeia dupla. A "temperatura crítica", "temperatura crítica de desnaturação" ou " T_c " referem-se a uma temperatura inferior à T_f da sequência de referência. A T_c é aplicada para desnaturar preferencialmente dúplices de cadeia dupla de sequência alvo ou dúplices de cadeia dupla de sequência alvo/referência, numa amostra de ácidos nucleicos, de modo a permitir o enriquecimento seletivo da sequência alvo durante uma reação de amplificação.

A T_f de um dado par de cadeias de ácidos nucleicos, é indicativa da estabilidade das cadeias face à ligação das cadeias, e depende da complementaridade das cadeias, do comprimento da sequência, do teor de GC, da presença ou ausência de disparidades dentro da região de cadeia dupla e de outros factores de menor importância, por exemplo, a concentração de sal da amostra (Lewin, Genes V, Capítulo 5, Oxford University Press & Cell Press: New York, (1994) páginas 109-126; SantaLucia, 1998).

A temperatura do ponto de fusão é geralmente determinada experimentalmente pela submissão da amostra a um aumento de temperatura constitutivo, e pela medição contínua da dissociação do dúplice de hibridização em cadeias simples. A dissociação pode ser detetada por uma variedade de diferentes métodos, por exemplo, por uma mudança na absorvância de UV, por fluorescência de corantes ligados a ADN de cadeia dupla, por ressonância plasmónica superficial ou preferencialmente por fluorescência. Neste último caso,

a sonda de hibridização é normalmente marcada com uma entidade fluorescente, e a geração de um sinal de fluorescência depende de alguma forma da formação do dúplice de hibridização.

A T_f pode ser determinada experimentalmente ou calculada com base em métodos bem definidos e conhecidos dos peritos comuns da especialidade. Os métodos de observação e análise das transições de desnaturação de ácidos nucleicos incluem: a medição da variação entálpica de uma amostra por calorimetria diferencial de rastreamento (CDV) dada que esta desnatura (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)), a medição da fluorescência de pares de fluoróforos covalentemente ligados (Vamosi e Clegg, *Biochemistry* 37:14300-14316 (1998)), e a monitorização da mudança de hipercromicidade de ácidos nucleicos (Haugland, "In Vitro Applications for Nucleic Acid Stains and Probes", em *Handbook of Fluorescent Probes and Research Chemicals*, 6ª Ed., Molecular Probes Inc, Eugene OR (1996) páginas 161-174).

Os valores da T_f de ácidos nucleicos de cadeia dupla, também podem ser observados pela monitorização da fluorescência de corantes específicos para ADN de cadeia dupla em combinação com ácidos nucleicos (Wittwer *et al.*, 1996). Corantes específicos para cadeias duplas são fluoróforos de ligação a ácidos nucleicos. Tipicamente, a fluorescência dos corantes em questão aumenta quando estes estão ligados a dúplexes de ácidos nucleicos (Wittwer *et al.*, *BioTechniques* 22:176-181 (1997)). Ririe *et al.* (1997) demonstraram que os produtos de pós-RCP podem ser diferenciados pelas suas análises de curva de fusão, através da utilização do corante de ligação a ácidos

nucleicos de cadeia dupla verde-SYBR[®] I. O verde-SYBR[®] I liga-se preferencialmente a ácidos nucleicos de cadeia dupla (Haugland, 1996). Outros corantes adequados para a determinação da T_f de um ácido nucleico de cadeia dupla incluem ouro-SYBR[®], brometo de etídio, acridina laranja, brometo de propídio, PicoGreen[®], Hoechst 33258, Hoechst 33342, Hoechst 34580, YO-PRO[®]-1 e YOYO[®]-1. Cada um destes corantes está comercialmente disponível. Por exemplo, o Capítulo 8 do catálogo *Molecular Probes* (Eugene, Oreg.) Handbook of Fluorescent Probes and Research Products, 8^a Edição (no CD-ROM, maio de 2001), enumera uma série de corantes que podem ser utilizados na presente invenção.

Como um perito na técnica irá aperceber-se, a fusão de qualquer estrutura de ácidos nucleicos de cadeia dupla, ocorre geralmente numa percentagem substancial de uma população de ácidos nucleicos semelhantes, ao longo de um intervalo limitado de temperatura, e terá tipicamente um pico de fusão (transição mais rápida) próximo da T_f para esse ácido nucleico. Portanto, tais picos na mudança da emissão de fluorescência, podem ser utilizados para calcular a T_f para ácidos nucleicos de cadeia dupla.

Um perfil de temperatura de fusão pode ser representado graficamente, ao representar $-dF/dT$ contra T , onde dF é a alteração na emissão de fluorescência medida, dT é a alteração na temperatura do ácido nucleico e T é a temperatura do ácido nucleico. Tal representação gráfica apresentará picos para temperaturas às quais ocorrem as mudanças de fluorescência mais rápidas, indicando assim as temperaturas de fusão.

Outros métodos para a determinação da T_f de um ácido nucleico de cadeia dupla são conhecidos na especialidade,

incluindo aqueles descritos nas Patentes E.U.A. N°s. 7,226,736 e 6,030,115.

A T_f pode ser prevista a partir de equações empíricas. A T_f é conhecida por depender do comprimento da sequência de um ácido nucleico que forma pares de bases complementares (n), do teor de G e C na sequência, das concentrações de sal (μ) e do agente de desnaturação (% FA) na solução da amostra, e usualmente segue a equação empírica $T_f = 81.5 + 16.6 \log(\mu) + 0.41(\% \text{ GC}) - 500/n - 0.61 (\% \text{ FA})$. A T_f pode ser estimada por uma série de métodos, por exemplo, através da interpolação pelo vizinho mais próximo como em Wetmur 1991 (Wetmur, J. G. 1991. DNA probes: applications of the principles of nucleic acid hybridization. Crit Rev Biochem Mol Biol 26: 227-259) e por *softwares* comerciais, incluindo OligoTM Primer Design e outros *softwares* disponíveis na Internet.

A "temperatura crítica" ou " T_c " refere-se a uma temperatura abaixo da temperatura de fusão " T_f " da sequência de referência. A T_c é aplicada para desnaturar preferencialmente dúplexes de cadeia dupla de sequência alvo ou dúplexes de cadeia dupla de sequência alvo/referência, de modo a permitir o enriquecimento seletivo da sequência alvo durante uma reação de amplificação. A temperatura crítica retira vantagens de uma T_f inferior referente a uma sequência alvo de cadeia dupla, ou referente a um dúplex de ADN de cadeia dupla hibridizado cruzadamente com uma sequência de referência. Quando a sequência alvo e a sequência de referência hibridizam cruzadamente, pequenas diferenças na sequência de um ou mais nucleotídeos referentes a disparidades ao longo de uma sequência de ADN de cadeia dupla de curta dimensão (por exemplo, < 200 pb), irão gerar uma pequena alteração, mas previsível, da temperatura de fusão (T_f) para essa

sequência (Lipsky, R.H., *et al.* (2001) Clin Chem, 47, 635-644; Liew, M., *et al.* (2004) Clin Chem, 50, 1156-1164). Dependendo do contexto exato da sequenciação e da posição de disparidade, variações da temperatura de fusão entre 0,5-1,5 °C são comuns para sequência até 200 pb. Assim, o emparelhamento da sequência alvo/referência, irá ter essencialmente uma T_f inferior do que aquela do alelo conhecido (por exemplo, sequência de referência) devido às disparidades. Pelo menos em parte, a presente invenção explora as pequenas diferenças na T_f entre sequências totalmente coincidentes e com disparidades. Dado que a desnaturação crítica é realizada em cada ciclo da RCP, o enriquecimento diferencial dos alelos que contêm mutações cresce exponencialmente e resulta em grandes diferenças na eficiência global da amplificação entre alelos mutantes e não mutantes. A T_c é cerca de 0,1-20 °C inferior à T_f da sequência de referência. Mais preferencialmente, a T_c é cerca de 0,1-15 °C, 0,1-10 °C, 0,1-9 °C, 0,1-8 °C, 0,1-7 °C, 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 inferior à T_f da sequência de referência.

Em algumas formas de realização, a T_c pode ser inferior à T_f de ambas as sequências de referência e alvo. Por exemplo, num dado caso a T_f de uma sequência não mutante era 84 °C e a T_f da sequência mutante era 83,8 °C. Quando uma versão rápida de um processo de enriquecimento foi realizada, a T_c ideal foi 83,5 °C.

Em algumas formas de realização preferenciais, a T_c é escolhida de tal modo que seja inferior à T_f das sequências de referência e de todas as sequências alvo possíveis. Por exemplo, num dado caso a T_f de uma sequência não mutante era 84 °C e as T_f das sequências mutantes em diferentes posições (mutações de ponto único) eram 83,8 °C, 83,7 °C,

83,9 °C, 83,6 °C e 83,75 °C. Quando uma versão rápida de um processo de enriquecimento foi realizada, a T_c ideal foi 83,5 °C.

A temperatura de desnaturação crítica (T_c), é a temperatura cujos valores inferiores de temperatura causam a queda abrupta da eficiência da RCP para uma dada sequência de ácidos nucleicos. Por exemplo, uma sequência 167 pb p53 é amplificada corretamente se a temperatura de desnaturação da RCP é definida a 87 °C, é amplificada razoavelmente a 86,5 °C e não produz nenhum produto detetável se a temperatura de desnaturação da RCP é definida a 86 °C ou menos.

De modo semelhante à T_f , a T_c para uma dada sequência pode ser determinada experimentalmente ou por cálculos. Para identificar a T_c experimentalmente para um dado produto de RCP, é realizada uma curva de fusão em tempo real na presença de um corante de intercalação (LC-VERDE ou verde-SYBR), de forma a obter a temperatura de fusão T_f média da sequência (ver acima em relação à determinação de T_f). Temperaturas de cerca de 0,5-1,5 °C inferiores face à T_f , são geralmente temperaturas de desnaturação críticas T_c adequadas, resultando no enriquecimento de sequências alvo.

Pode-se também estimar a T_c pela determinação de ΔT_f para uma sequência de ADN, devido à existência de disparidades de pares de bases, como aqueles que estariam presentes em dúplexes de hibridização cruzada alvo/referência. Esta diferença é de cerca de 0,1 °C a cerca de 12,5 °C em comparação com uma sequência perfeitamente coincidente referência-referência. Assim, numa forma de realização, a T_c pode ser representada pela equação $T_c = T_f - \Delta T_f$, em que T_f é a temperatura de fusão do dúplex de sequência referência/referência e ΔT_f é a alteração na T_f da

sequência de referência, que resulta de uma ou mais disparidades dos pares de bases formadas durante a hibridização cruzada das sequências alvo/referência. O ΔT_f foi encontrado como sendo dependente do comprimento do dúplex de sequência alvo/referência, do teor percentual de guanina-citosina (% GC) e da localização da mutação pontual ou disparidade dos pares de bases no dúplex. Por exemplo, se a disparidade encontra-se em qualquer uma das extremidades do dúplex, o ΔT_f será geralmente mais baixo, tipicamente na gama de cerca de 0,1 °C a cerca de 8 °C. Se a disparidade encontra-se na direção do centro do dúplex, o ΔT_f será relativamente maior, tipicamente na gama de cerca de 0,2 °C a cerca de 11 °C. O ΔT_f é geralmente equivalente a cerca de 0,5 °C a cerca de 1,5 °C por percentagem de disparidade de base no dúplex hibridado cruzadamente alvo/referência. Deverá ser considerado que o ΔT_f irá variar em função, não só do comprimento do dúplex e da localização da mutação, mas também irá variar dependendo da ordem específica da sequência. Consequentemente, todas estas variações estão dentro do âmbito da divulgação.

Sequências de Ácidos Nucleicos Úteis na Invenção

A invenção proporciona métodos para o enriquecimento de uma sequência alvo numa amostra de ácidos nucleicos e também utiliza iniciadores para a amplificação de uma sequência de ácido nucleico *template*.

Tal como aqui utilizados, os termos "ácido nucleico", "polinucleico" e "oligonucleico" referem-se a iniciadores, sondas e fragmentos de oligómeros a serem detetados, e devem ser genéricos para polidesoxirribonucleicos (contendo 2-desoxi-D-ribose), para os polirribonucleicos (contendo D-ribose) e para qualquer outro tipo de polinucleico que seja um N-glicósido de uma base de purina ou pirimidina, ou

bases modificadas de purina ou pirimidina (incluindo locais abásicos). Não se pretende nenhuma distinção de acordo com o comprimento entre o termo "ácido nucleico", "polinucleico" e "oligonucleico", e estes termos serão utilizados alternadamente de forma indiferenciada. Estes termos referem-se apenas à estrutura primária da molécula. Assim, estes termos incluem ADN de cadeia dupla e simples, bem como ARN de cadeia dupla e simples.

Não é necessário que o oligonucleico derive fisicamente de qualquer sequência existente ou natural, mas pode ser gerado a partir de qualquer método que inclua síntese química, replicação de ADN, transcrição reversa ou uma combinação das mesmas. Os termos "oligonucleicos" ou "ácido nucleico" pretendem a referência a um polinucleico de ADN ou ARN genómico ou ADNc, de origem semissintética ou sintética, que por virtude da sua origem sintética ou manipulação: (1) não está associado à totalidade ou a uma porção do polinucleico com o qual está associado na natureza; e/ou (2) está ligado a um polinucleico diferente daquele a que está ligado na natureza.

Numa forma de realização, os ácidos nucleicos utilizados no método contêm nucleótidos modificados de modo a aumentar as diferenças da temperatura de desnaturação entre homodúplexes de sequência referência/referência e heterodúplexes de sequência alvo/referência. Tais modificações iriam aumentar o enriquecimento da sequência alvo. Os nucleotídeos modificados ou não naturais podem ser incorporados antes ou durante o procedimento de enriquecimento. Os nucleotídeos modificados contemplados para utilização nos métodos da invenção incluem: análogos diamino-purina (por exemplo, 2'-O-metil-2,6-diaminopurina), uracilo, análogos peptídicos de ácidos nucleicos, análogos dos anteriores modificados com biotina, análogos dos

anteriores modificados com fluoróforo, inosina, 7-desazaguanina, ácidos nucleicos 2'-desoxi-2'-fluoro-β-D-arabinonucleico (2'F-ANA), ácidos nucleicos bloqueados (LNAs), ENAs: ácidos nucleicos com ligações de 2'-O,4'-C-etileno e outros. Os nucleotídeos modificados podem ser incorporados em qualquer um dos ácidos nucleicos da invenção, incluindo o *template*, iniciador e sonda de ácidos nucleicos.

Estas modificações podem aumentar a diferença das T_f entre bases coincidentes e díspares, aumentando deste modo o enriquecimento obtido com os presentes métodos. Por exemplo, os ácidos nucleicos bloqueados, que representam uma classe de análogos de nucleótidos conformacionalmente restringidos, classe esta descrita, por exemplo, no documento WO 99/1 4226, aumentam a temperatura de fusão de um ácido nucleico. Os oligonucleicos que contêm os nucleótidos bloqueados estão descritos em Koshkin, A.A., *et al.*, Tetrahedron (1998), 54: 3607-3630 e em Obika, S. *et al.*, Tetrahedron Lett. (1998), 39: 5401-5404, ambos incorporados nas referências. A introdução de um nucleótido bloqueado num oligonucleico aumenta a afinidade para sequências complementares, e aumenta a temperatura de fusão em vários graus (Braasch, D.A. e D.R. Corey, Chem. Biol. (2001), 8:1-7). A invenção pode ser realizada com um qualquer dos LNAs conhecidos na especialidade, por exemplo, aqueles divulgados no documento WO 99/14226 e em Latorra D, *et al.*, 2003. Hum. Mutat. 22: 79-85.

Não é necessário que a complementaridade seja perfeita; dúplexes estáveis podem conter pares de bases díspares ou bases não coincidentes. Os peritos na especialidade da tecnologia de ácidos nucleicos podem determinar empiricamente a estabilidade do dúplex, através da

consideração de um certo número de variáveis que incluem, por exemplo, o comprimento do oligonucleico, a composição das bases e a sequência do oligonucleico, a força iónica e a incidência de pares de bases díspares. A estabilidade de um dúplex de ácidos nucleicos é medida pela temperatura de fusão ou " T_f ".

A invenção utiliza iniciadores oligonucleicos para a amplificação de uma sequência de ácido nucleico *template*.

O termo "iniciador" pode referir-se a mais do que um iniciador e refere-se a um oligonucleico, quer ocorra naturalmente, tal como um produto purificado de uma digestão de restrição, quer seja produzido sinteticamente, sendo capaz de atuar como um ponto de iniciação para a síntese ao longo de uma cadeia complementar, quando colocado sob condições nas quais a síntese de um produto de extensão do iniciador, que é complementar a uma cadeia de ácido nucleico, é catalisada. Tais condições incluem a presença de quatro desoxirribonucleotídeos trifosfatados diferentes e um agente indutor de polimerização, tal como a ADN polimerase ou a transcriptase reversa num tampão apropriado ("tampão" inclui substituintes que são cofatores, ou que afetam o pH, força iónica, etc.), e uma temperatura adequada. O iniciador é preferencialmente de cadeia simples para uma eficiência máxima da amplificação.

De acordo com a invenção, os iniciadores oligonucleicos úteis são as moléculas de ADN ou ARN de cadeia simples que são capazes de hibridizar com uma sequência de ácido nucleico *template*, e passíveis de síntese enzimática despoletada por iniciadores para a formação de uma cadeia de ácidos nucleicos. O iniciador é complementar a uma porção de uma molécula alvo que está presente numa conjunto de moléculas de ácidos nucleicos. É observado que os

iniciadores oligonucleicos são preparados por métodos sintéticos, quer químicos ou enzimáticos. Alternativamente, tal molécula ou um fragmento resultante é de ocorrência natural, sendo isolada a partir da sua fonte natural ou adquirida através de um fornecedor comercial.

Os iniciadores oligonucleicos têm um comprimento de 5 a 100 nucleotídeos, idealmente de 17-40 nucleotídeos, embora iniciadores de diferentes comprimentos também sejam usados. Os iniciadores para a amplificação têm preferencialmente cerca de 17-25 nucleótidos. Os iniciadores úteis de acordo com a invenção, podem também ser concebidos para ter uma temperatura de fusão específica (T_f) segundo o método de estimativa da temperatura de fusão. Os *softwares* comerciais, incluindo OligoTM Primer Design e *softwares* disponíveis na Internet, incluindo Primer3 e Oligo Calculator, podem ser utilizados para calcular uma T_f de uma sequência de ácidos nucleicos útil de acordo com a invenção. Preferencialmente, a T_f de um iniciador de amplificação que é útil de acordo com a invenção, tal como calculada por exemplo pelo Oligo Calculator, está preferencialmente entre cerca de 45 a 65 °C, e mais preferencialmente entre cerca de 50 a 60 °C.

Normalmente, a hibridização seletiva ocorre quando duas sequências de ácidos nucleicos são substancialmente complementares (pelo menos cerca de 65 % de complementaridade ao longo de um segmento com pelo menos 14 a 25 nucleótidos, preferencialmente pelo menos cerca de 75 %, mais preferivelmente pelo menos cerca de 90 % de complementaridade). Ver Kanehisa, M., 1984, *Nucleic Acids Res.* 12: 203.

Como um resultado, é esperado que seja tolerado um certo grau de disparidade no local de iniciação. Tal disparidade

poder ser pequena, como por exemplo um mono-, di- ou tri-nucleótido. Alternativamente, uma região de disparidade pode conter laços, que são definidos como regiões em que existe uma disparidade numa série ininterrupta de quatro ou mais nucleótidos.

É divulgado que os procedimentos de enriquecimento podem ser utilizados em combinação com iniciadores ácido nucleico peptídicos (ANP), de modo a aumentar a sensibilidade de enriquecimento da mutação. Os ANP são utilizados para suprimir apenas a amplificação das sequências não mutantes (de referência). Os iniciadores podem ser sintetizados de modo a distinguir entre uma sequência alvo (mutante) e uma sequência de referência. Os iniciadores com base em ANP reconhecem e ligam sequências não mutantes complementares com uma maior estabilidade térmica e especificidade, em comparação com os iniciadores que ligam sequências mutantes. Isto não só aumenta as diferenças das T_f , entre os ácidos nucleicos de referência com iniciação através de ANP e os ácidos nucleicos alvo com iniciação normal, mas também evita que os iniciadores baseados em ANP possam ser estendidos por uma ADN polimerase, o que resulta num enriquecimento ainda maior da sequência alvo. Tais ensaios são conhecidos na especialidade e descritos em Orum *et al.* Nucleic Acids Research, 21(23): 5332-5336 (1993).

Os iniciadores oligonucleicos podem ser concebidos com estas considerações em mente, e sintetizado de acordo com os seguintes métodos.

Estratégia de Conceção de Iniciadores Oligonucleicos

A conceção de um iniciador oligonucleico específico para a finalidade de sequenciação ou RCP, envolve a seleção de uma sequência que é capaz de reconhecer a sequência alvo, mas

que tem uma estrutura secundária minimamente previsível. A sequência oligonucleica pode, ou não, ligar-se apenas a um único local no ácido nucleico alvo. Além disso, a T_f do oligonucleico é otimizada pela análise do seu comprimento e teor em GC do oligonucleico. Ainda, ao conceber um iniciador de RCP útil para a amplificação de ADN genómico, a sequência de iniciador selecionada não demonstra similaridade significativa face às sequências presentes na base de dados GenBank (ou outras bases de dados disponíveis).

A conceção de um iniciador que é útil de acordo com a invenção, é facilitada através da utilização de *softwares* prontamente disponíveis, desenvolvidos para facilitar a avaliação dos vários parâmetros descritos anteriormente e para otimizar as sequências dos iniciadores. Exemplos de tais *softwares* são "PrimerSelect" do pacote de *software* DNASTarTM (DNASTar, Inc.; Madison, WI), OLIGO 4.0 (National Biosciences, Inc.), PRIMER, *Oligonucleotide Selection Program*, PGEN e Amplify (descrito em Ausubel *et al.*, 1995, *Short Protocols in Molecular Biology*, 3^a Edição, John Wiley & Sons).

Os iniciadores são concebidos de modo a ter uma T_f inferior à temperatura aplicada durante a etapa de hibridização cruzada da sequência alvo/referência. Assim, os iniciadores não emparelham com as sequências alvo ou de referência durante esta etapa de hibridização (ver Fig. 1). Numa forma de realização, a T_f dos iniciadores é entre 5-10 °C inferior face à temperatura de emparelhamento da hibridização cruzada.

Síntese

Os próprios iniciadores são sintetizados através da utilização de técnicas que são bem conhecidas na especialidade. Os métodos para a preparação de oligonucleicos de sequência específica, são conhecidos na especialidade, e incluem, por exemplo, síntese química direta, clonagem e análise da digestão restritiva de sequências adequadas. Uma vez concebidos, os oligonucleicos são preparados por um método adequado de síntese química, que inclui, por exemplo, o método do fosfotriéster descrito em Narang *et al.*, 1979, *Methods in Enzymology*, 68:90, o método do fosfodiéster divulgado em Brown *et al.*, 1979, *Methods in Enzymology*, 68:109, o método do dietilfosforamidato divulgado em Beaucage *et al.*, 1981, *Tetrahedron Letters*, 22:1859, e o método de suporte sólido divulgado na Patente E.U.A Nº 4,458,066, ou por outros métodos químicos que utilizam quer um sintetizador automático comercial de oligonucleicos (que está disponível comercialmente) ou tecnologia VLSIPSTM. Os iniciadores podem também ser sintetizados com ácidos nucleicos modificados, segundo métodos bem conhecidos na especialidade.

Amostras

Tal como aqui utilizado, "amostra" refere-se a qualquer substância que contém, ou que se presume que contém, um ácido nucleico de interesse (sequências alvo e de referência), ou que é por si só um ácido nucleico que contém, ou que se presume que contém, um ácido nucleico alvo de interesse. O termo "amostra" inclui assim uma amostra de ácidos nucleicos (ADN genómico, ADNc, ARN), células, organismos, tecidos, fluidos ou substâncias,

incluindo mas não limitando-se a, por exemplo, plasma, soro, fluido espinal, fluido linfático, fluido sinovial, urina, lágrimas, fezes, secreções externas da pele, tratos respiratório, intestinal e genito-urinário, saliva, células sanguíneas, tumores, órgãos, tecido, amostras de constituintes de culturas celulares *in vitro*, isolados naturais (tais como água potável, água do mar, materiais sólidos), amostras microbianas e objetos ou espécimes que foram "marcados" com moléculas de ácido nucleico para detecção.

As sequências de ácidos nucleicos podem ser amplificadas a partir de ADN genómico. O ADN genómico pode ser isolado a partir de tecidos ou células de acordo com o seguinte método. Alternativamente, as sequências de ácidos nucleicos podem ser isoladas a partir de sangue segundo métodos bem conhecidos na especialidade.

Para facilitar a detecção de uma forma variante de um gene pertencente a um tecido em particular, o tecido é isolado. Para isolar o ADN genómico a partir de tecido de mamíferos, o tecido é picado e congelado em azoto líquido. O tecido congelado é moído até à obtenção de um pó fino utilizando um almofariz e um pilão previamente arrefecidos, e suspende-se este pó num tampão de digestão (100 mM de NaCl, 10 mM de Tris-HCl, pH 8,0, 25 mM de EDTA, pH 8,0, 0,5 % (m/v) de SDS, 0,1 mg/mL de proteinase K) num volume de tampão de digestão de 1,2 mL por 100 mg de tecido. Para isolar o ADN genómico a partir de culturas celulares de tecido de mamíferos, as células são sedimentadas por centrifugação durante 5 min a $500 \times g$, ressuspensas em 1-10 mL de PBS gelado, sedimentadas durante 5 min a $500 \times g$ e ressuspensas em 1 volume de tampão de digestão.

As amostras em tampão de digestão são incubadas (com agitação) durante 12-18 horas a 50 °C, e posteriormente extraídas com um volume igual de fenol/clorofórmio/álcool isoamílico. Se as fases não são separadas após um passo de centrifugação (10 min a 1700 × g), é adicionado um outro volume de tampão de digestão (sem proteinase K) e a etapa de centrifugação é repetida. Se for evidente a presença de um material branco espesso na interface das duas fases, a fase de extração orgânica é repetida. Após a extração da camada aquosa superior, esta é transferida para um novo tubo ao qual vai ser adicionado 1/2 do volume de acetato de amónio a 7,5 M e 2 volumes de etanol a 100 %. O ácido nucleico é sedimentado por centrifugação durante 2 min a 1700 × g, lavado com etanol a 70 %, seco ao ar e ressuspenso em tampão TE (10 mM de Tris-HCl, pH 8,0, 1 mM de EDTA, pH 8,0) a 1 mg/mL. ARN residual é removido através da incubação da amostra durante 1 hora a 37 °C na presença de 0,1 % de SDS e 1 µg/mL de ARNase sem ADNase, repetindo-se as etapas de extração e precipitação de etanol. É esperado que o rendimento de ADN genómico, de acordo com este método, seja cerca de 2 mg ADN/1 g de células ou tecido (Ausubel *et al.*, *supra*). O ADN genómico isolado de acordo com este método pode ser utilizado de acordo com a invenção.

O ADN alvo pode também ser extraído a partir de sangue não fracionado. Por exemplo, o sangue pode ser colhido por métodos de referência para um tubo de recolha, que preferencialmente compreende vidro siliconizado, quer sem anticoagulante para a preparação do soro, quer com EDTA, citrato de sódio, heparina ou anticoagulantes semelhantes, mais preferencialmente EDTA, para a preparação do plasma. O método preferencial, embora não seja absolutamente necessário, é o método que consiste no facto do plasma ou soro serem fracionados a partir de sangue não fracionado. O

plasma ou soro podem ser fracionados a partir do sangue não fracionado, por centrifugação, preferencialmente uma centrifugação suave a $300-800 \times g$ durante 5-10 minutos, ou fracionados por outros métodos convencionais. Uma vez que a heparina pode interferir com a RCP, o uso de sangue heparinizado pode exigir um pré-tratamento com heparinase. Assim, o EDTA é o anticoagulante preferencial para amostras de sangue. Quer plasma sanguíneo ou soro recentemente colhidos, quer plasma ou soro congelados (armazenados) e subsequentemente descongelados, podem ser usados nos métodos da invenção. O plasma ou soro armazenados devem ser mantidos à temperatura de $-20\text{ }^{\circ}\text{C}$ a $-70\text{ }^{\circ}\text{C}$, e plasma ou soro recentemente colhidos devem ser mantidos em refrigeração, ou mantidos em gelo, até à sua utilização. O ADN pode então ser posteriormente extraído segundo métodos bem conhecidos na especialidade, bem como aqueles aqui descritos.

Ensaaios de diagnóstico

A invenção proporciona um método para o enriquecimento de sequências alvo a partir de uma amostra do paciente, para diagnóstico, detecção, controlo, avaliação ou tratamento de uma doença, em especial uma doença neoplásica ou proliferativa, num animal ou num ser humano. Numa forma de realização preferencial, o ácido nucleico é derivado de um ácido nucleico que codifica um oncogene ou um outro ADN associado a um tumor.

O enriquecimento preferencial da sequência alvo permite a posterior análise ou manipulação do ADN. Por exemplo, os alelos enriquecidos podem ser analisados para determinar as características da célula a partir do qual se origina o ADN. Qualquer um de vários métodos pode ser usado, de acordo com a informação desejada, incluindo a sequenciação de ácidos nucleicos, PFR, RCP digital, espectroscopia

incluindo espectroscopia de RMN de prótons, análise bioquímica e análise imunológica. Numa forma de realização, o ADN amplificado é isolado através da excisão de bandas de ADN mutante a partir de um gel de agarose, sendo novamente amplificado, clonado num vetor de plasmídeo, por exemplo, o vetor plasmídico pGEM-T (Promega) e sequenciado utilizando um *kit* comercial, tal como o *kit* Sequenase 2.0 (USB). A análise para definir as características do ADN alvo, e consequentemente por exemplo um tumor, proporciona uma grande variedade de informação clínica útil, incluindo a descrição, a caracterização ou a classificação da célula (por exemplo, tumor), quer seja conhecida ou desconhecida, de acordo com o tecido de origem, tipo (tal como pré-maligna ou maligna), fenótipo, genótipo e de acordo com a descrição ou caracterização do comportamento do tumor, fisiologia e bioquímica, de forma a ganhar compreensão da capacidade de invasão tumoral, propensão à metástase e sensibilidade ou resistência a várias terapias, permitindo então a previsão da resposta quer face à terapia em curso ou a terapias planeadas, e ainda, permitindo a avaliação do prognóstico. A comparação das características do ADN alvo com biópsias ou espécimes cirúrgicos anteriores, permite a consequente avaliação da semelhança ou heterogeneidade do tumor face a esse espécime, e assim a avaliação da recorrência do tumor.

Seguindo também o enriquecimento seletivo da sequência alvo, o ácido ribonucleico complementar (ARN) pode ser transcrito ou fabricado a partir do ADN. Numa forma de realização preferencial, a transcrição de ARN é realizada através da aplicação de um iniciador com uma região do promotor da ARN polimerase, ligada à sequência de iniciador padrão para o ADN de interesse, na reação de amplificação (etapa três). O ARN complementar ao ADN é então transcrito a partir da região do promotor que está ligado. Num método

alternativo, o ADN do alelo amplificado é clonado num vetor de expressão, e o ARN complementar ao ADN é transcrito. Além disso, como uma forma de realização preferencialmente opcional, o ARN complementar é utilizado numa reação de tradução *in vitro*, para a produção de proteínas associadas a um tumor ou específicas de tumores.

A caracterização do alelo, a amplificação de ADN derivado de tumores ou associado a tumores, a caracterização e transcrição de ARN complementar, e a tradução de proteínas associadas a um tumor ou específicas de tumores, é de utilidade significativa, tanto na atribuição da terapia, bem como no desenvolvimento de terapias específicas para o tumor. A sequenciação de ADN extracelular ou a transcrição de ARN complementar, permite a atribuição ou desenvolvimento de compostos *antisense*, incluindo oligonucleicos sintéticos e outras construções *antisense* adequadas de forma específica para o ADN extracelular, quer pela construção de um plasmídeo de expressão, quer pela adaptação do método de Aoki *et al.* (1995, Cancer Res. 55: 3810-3816). Semelhantemente, a definição das características tumorais permite a atribuição de terapias específicas de anticorpos monoclonais ou de vacinas, adequadas de forma específica para o ADN amplificado. A produção da proteína correspondente imunológica, pode então ser utilizada no desenvolvimento de anticorpos monoclonais específicos para os tumores. Analogamente, as proteínas traduzidas podem ser utilizadas no desenvolvimento de vacinas específicas para os tumores.

De particular valor, a invenção permite o desenvolvimento e aplicação destas terapias ou diagnósticos específicos para tumores, mesmo quando apenas tumores pré-malignos, cancros precoces ou cancros ocultos estão presentes. Assim, a

invenção permite a intervenção terapêutica quando a carga tumoral é baixa, a função imunológica está relativamente intacta e o paciente não está comprometido, factos que contribuem para o aumentando do potencial de cura.

Processamento Adicional das Sequências Enriquecidas

A combinação dos atuais métodos com MALDI-TOF, fusão de alta resolução ou sequenciação de molécula única, iria abordar três necessidades distintas na deteção de mutações: a deteção rápida de mutações somáticas conhecidas ou suspeitas de correlacionarem-se com a evolução do estado clínico (MALDI-TOF); o rastreio rápido de amostras individuais de pacientes para mutações somáticas desconhecidas (fusão-AR), seguindo-se da sequenciação seletiva de alguns exões; e sequenciação massivamente em paralelo (Sequenciação de Moléculas Única, SMS) de vários genes em "amostras difíceis", isto é, em amostras de tumores com heterogeneidade, de contaminações do estroma ou de fluidos corporais, onde mutações clinicamente relevantes podem estar ao nível de 0,5-5 %.

Espectrometria De Massa

Numa forma de realização, uma sequência alvo enriquecida é sujeita a sequenciação por MALDI-TOF. A espectrometria de massa (MS) emergiu como uma ferramenta poderosa na sequenciação de ADN. Os espectrofotómetros de massa produzem uma medição de massa direta, que pode ser adquirida em segundos ou minutos na gama femtomolar a picomolar. A ionização e dessorção a laser assistida por matriz (MALDI) com analisador de tempo de voo (TOF) MS, tem sido utilizada com sucesso para a sequenciação rápida de ADN e eficiente determinação do tamanho das moléculas de ADN. O advento de MALDI-TOF MS tornou mais fácil a

ionização de grandes moléculas de ADN intacto, e a medição dos seus rácios de massa-carga. Os produtos da reação em cadeia da polimerase (RCP®) de cadeia simples e cadeia dupla com 500 nucleotídeos (nt) de comprimento, foram detetados por MALDI-TOF MS. Utilizando combinações otimizadas matriz-laser que reduzem a fragmentação do ADN, os espectros de massa de MALDI infravermelhos de ADN sintético, os fragmentos de enzimas de restrição do ADN plasmídeo e os transcritos de ARN até a um tamanho de 2180 nt, foram reportados com uma precisão de $\pm 0,5-1 \%$. Apesar de grandes oligómeros terem sido detetados por MALDI-TOF MS, é geralmente aceite que até um 100-mero é presentemente rotineiro.

Para vários genes, mutações clinicamente importantes não ocorrem aleatoriamente no genoma (por exemplo, mutações pontuais de aquisição de funcionalidade que ocorrem em oncogenes mais conhecidos). Em vez de as alterações afetarem um número relativamente pequeno de codões, estas são frequentemente responsáveis pela maior parte das mutações somáticas. Desta forma, e em princípio, um número limitado de ensaios genéticos judiciosamente concebidos, deve ser eficaz para analisar uma grande proporção das mutações clinicamente relevantes. Por exemplo, Garraway e os seus colegas (Thomas, R.K., *et al.* (2007) *Nat Genet*, 39, 347-351) demonstraram que 16-44 ensaios MALDI-TOF por gene Ras, EGFR e BRAF, capturaram 90 % -99 % da prevalência da mutação observada, até ao momento para estes genes em tumores malignos humanos. Assim, é proposto que genotipagem de alto rendimento pode proporcionar um meio eficaz para a deteção de mutações críticas e/ou alvos de base cancerígena, para grande escala de amostras clínicas. O MALDI-TOF é idealmente adequado para a deteção de mutações anteriormente identificadas em amostras tumorais não heterogéneas. No entanto, enquanto não há dúvida sobre a

fiabilidade de MALDI-TOF para mutações germinativas ou identificação-SNP, a experiência na detecção de mutações somáticas é relativamente recente. Assim, a fiabilidade de MALDI-TOF diminui substancialmente quando as amostras heterogêneas com < 10 % de células mutadas são utilizadas (por exemplo, cancros do pâncreas, do pulmão ou da próstata) ou quando o ADN a partir de fluidos corporais deve ser rastreado. Ao melhorar a sensibilidade, o presente método de enriquecimento permite à técnica de MALDI-TOF detetar mutações somáticas com níveis de expressão baixos, e também irá fornecer a fiabilidade necessária que é imperativa para o rastreio de amostras de tumores cirúrgicos convencionais.

Fusão de Alta Resolução

Numa outra forma de realização, uma sequência alvo enriquecida é submetida a fusão de alta resolução. Os genes que contêm mutações clinicamente relevantes em numerosas posições ao longo exões, tais como p53, são mais fáceis rastreadas através de mutação de varrimento em vez de genotipagem mutação individual. O HR-fusão é uma tecnologia de digitalização mutação de alto rendimento introduzida nos últimos anos, com excelente capacidade para descobrir SNP ou mutações da linha germinativa (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. and Wittwer, C.T. (2004); Clin Chem, 50, 1748-1754).

Imediatamente após a amplificação por RCP de uma região genómica de interesse na presença de um corante fluorescente intercalante (LC ou verde SYBR verde) a presença de uma mutação é identificada em tempo real por análise de curva de fusão e cuidadosa comparação com o tipo

não mutante sem qualquer tratamentos pós amplificação. Além disso, a análise de alta resolução de varrimento de fusão realiza simultânea e genotipagem do gene de mutação (isto é, identificação de SNP) numa fracção do tempo necessário quando se usa os métodos tradicionais, enquanto se mantém um ambiente de tubo fechado. A RCP exige <30 min (capilares) ou 1,5 h (placas de 96 ou 384 poços de fusão) e aquisição leva 1-2 min por capilar ou 5 min por placa.

No entanto, como com MALDI-TOF, as vantagens de usar FC de ponto de fusão não pode ser exploradas para as mutações somáticas abaixo de aproximadamente -20 % mutante de tipo não mutante proporções, assim, várias classes de amostras clínicas não podem ser rastreados de forma fiável através HR-fusão. Ao aumentar os limites de detecção, a presente invenção permite a conveniência e taxa de transferência de HR-fusão a ser aplicado para o rastreio convencional amostra de tumor cirúrgico e também para a detecção de baixo nível mutações somáticas em amostras clínicas "difícil" com a contaminação do estroma ou ADN a partir de fluidos corporais.

Sequenciação de Molécula Única

A sequência alvo enriquecido pode ser submetido a sequenciação de molécula única (Thomas, RK, *et al.* (2006);. Nat Med, 12, 852-855). As capacidades da única molécula de sequenciamento também seriam beneficiadas com a incorporação da presente invenção. Por exemplo, para a mutação rastreio em amostras de doentes, a RCP dos exões selecionados de uma máquina normal de RCP a partir de ADN genómico é ainda necessária antes de iniciar a segunda geração de sequenciação. Além disso, a detecção de mutações ao nível de 1-5 % para rácio mutante tipo não mutante em amostras clínicas requer sequenciação de numerosos eventos

individuais repetido de modo a obter estatísticas aceitáveis. Esta última análise, reduz as capacidades de rendimento, e por mutações ao nível de 1 % apenas 10-20 sequências podem ser rastreados simultaneamente, em oposição a ~4,000 sequências se as mutações foram predominantes (por 454 Life Sciences, o serviço técnico). Ao realizar a presente invenção, antes da sequenciação, única molécula a prevalência de mutações vai aumentar em 1-2 ordens de grandeza como uma fracção do número total ou alelos, aumentando assim o rendimento de sequenciação única molécula a um grau equivalente.

O método de enriquecimento seletivo da presente invenção pode ser aplicada durante a fase em emulsão de uma única molécula de reação de sequenciação. As sequências alvo enriquecidas são então submetidas a pirosequenciação.

Extensão do iniciador

A sequência alvo enriquecida pode ser submetida a uma reação de extensão do iniciador de sequenciação. Na extensão do iniciador, os oligonucleicos são usados para avaliar a variação na sequência numa posição predeterminada em relação a um ácido nucleico, cuja sequência é conhecida. Um oligonucleico de amostra é fornecida como uma molécula de cadeia única, a molécula de cadeia única é misturada com um agente indutor, um nucleótido marcado, e um iniciador que possui uma sequência idêntica a uma região flanqueadora a posição predeterminada de modo a formar uma mistura, tendo a mistura um elemento essencial ausência de nucleotídeos constituída por outras do que a base do qual o nucleótido marcado é constituído bases. A mistura é então sujeita a condições propícias para o emparelhamento do iniciador com a molécula de cadeia simples e a formação de um produto de extensão do iniciador incorporando o

nucleótido marcado, e a mistura é analisada para a presença de produto da extensão do iniciador que contém nucleótido marcado (US Pat. No. 5.846.710).

Tampões

A inclusão de solventes orgânicos para aumentar as diferenças entre T_f entre dúplexes incompatíveis e combinados no método de amplificação da invenção corrente é contemplado. A inclusão de certos solventes orgânicos podem melhorar o enriquecimento da sequência alvo, por exemplo, a inclusão de solventes orgânicos pode aumentar a diferença entre a temperatura de desnaturação de referência e as sequências alvo de ADN e, assim, ajudar a amplificação preferencial da sequência alvo. Os solventes orgânicos, tais como DMSO, formamida, betaina glicerol (aparato, D. e Medrano, JF, Biotechniques, 10, 58-59 (1991)) pode aumentar a diferença T_f entre combinados (referência/referência) e incompatível (alvo/referência) sequências. Por conseguinte, uma vez que a etapa de hibridização intermédio (hibridização cruzada) no método da presente invenção forma sequências alvo de referência contendo desemparelhamentos, a inclusão de solventes orgânicos na medida em que eles não inibem a ação da polimerase é vantajosa. Portanto, em algumas formas de realização, a mistura reacional é suplementado com DMSO, formamida, betaina, glicerol ou uma combinação dos mesmos, a níveis de volume de 1-10 % do volume ou o volume Preferencialmente 3-8 % do volume, preferencialmente 5 ou mais -6 % de volume para volume. O Exemplo 8 ilustra o uso de DMSO num método de enriquecimento.

Outra vantagem prática de utilização de solventes orgânicos é que o TC, que é apropriado para uma dada sequência de alterações na utilização de um solvente orgânico na reação.

Assim, enquanto que na ausência de DMSO a T_c para uma sequência é 83,5 °C, na presença de 3 % de DMSO a T_c é 80,5 °C. Como resultado, através da adição de uma quantidade diferente de DMSO ou outro solvente para sequências diferentes, pode se garantir que o T_c é a mesma para todas as sequências. Isto é útil para a execução de numerosas reações de enriquecimento para uma variedade de sequências com uma única máquina de PCR-prazo, uma vez que a temperatura de desnaturação é, em seguida, o mesmo para todos.

EXEMPLOS

Exemplo 1. Materiais e métodos para o enriquecimento de uma sequência alvo

Sequências utilizadas para a validação da RCP-CODB: Para validar a presente invenção, foram usadas uma série de ADN genómico e de linhas celulares que contêm mutações em diferentes posições do exão 8 de p53 e exão 2 de Kras (codões 12-13) (Figura 3). As mutações do exão 8 de p53 correlacionam-se com um fraco prognóstico do cancro de pulmão e são mutações de baixa prevalência no plasma de pacientes com cancro. Da mesma forma, mutações no Kras têm um prognóstico importante no adenocarcinoma de pulmão.

Protocolo de enriquecimento e iniciadores: a RCP foi realizada na presença do corante de intercalação 0,1 x verde-LC e seguida em tempo real num dispositivo Cepheid. O acompanhamento em tempo real da RCP não é necessário, mas é conveniente, por isso foi adotado para todas as experiências.

Para a sequência de 167 pb de p53 a RCP normal foi realizada primeiramente durante 10 ciclos, a fim de gerar produto suficiente para ser utilizado no protocolo de enriquecimento. Uma máquina de Cepheid foi programada com os seguintes parâmetros de ciclo:

95 °C, 120 segundos; (95 °C, 15 segundos/55 °C leitura de fluorescência ON, 30 segundos/72 °C, 1 min de extensão) x 10 ciclos.

O produto da RCP resultante foi, em seguida, diluído de 1:1000 e submeteu-se ao protocolo de enriquecimento abaixo, do qual também é ilustrado na Figura 1:

95 °C, 15 segundos; 70 °C durante 120 segundos; desnaturação a $T_c = 86,5$ °C durante 3 segundos; 55 °C leitura de fluorescência ON durante 30 segundos; de seguida, 72 °C, 1 min durante 30 ciclos de extensão.

A fim de preparar o produto da RCP para a didesoxi-sequenciação de Sanger, o produto foi tratado com exonuclease I e fosfatase alcalina de camarão. Os seguintes iniciadores foram utilizados no método de sequenciação:

167 pb fragmento: 5'-GCT TCT CTT TTC CTA TCC TG - 3' sentido direto; 5'- ACC CTT CTT TCG AGT GCT - 3' sentido reverso

Para os fragmentos 87 pb e 210 pb de p53, e o fragmento 135 pb de Kras o protocolo de enriquecimento foi como descrito acima, contudo as temperaturas críticas de desnaturação foram fixadas em $T_c = 83,5$, $87,5$ e 80 °C, respetivamente. Os iniciadores utilizados na reação de sequenciação foram os seguintes:

5'- TGG TAA TCT ACT GGG ACG-3' sentido direto;

5'-CGG AGA TCT TTC TCC TCT - 3' sentido reverso (fragmento de 87 pb do exão 8 de p53)

5'- GCT TCT CTT TTC CTA TCC TG - 3' sentido direto;
 5'- TAA CTG CAC CCT TGG TC - 3' sentido reverso (fragmento de 210 pb do exão 8 de p53)
 5'-AACTTGTGGTAGTTGGACCT-3' sentido direto;
 5'-CTCTATTGTTGGATCATATT-3' sentido reverso (fragmento Kras do exão 2). A reprodutibilidade de todos os protocolos de enriquecimento foi testada em 3-6 experiências independentes.

Resultados

Mutações no exão 8 de p53: quando o protocolo de enriquecimento utilizando uma temperatura de desnaturação crítica $T_c = 86,5^\circ\text{C}$ foi aplicado ao fragmento 167 pb do exão 8, o enriquecimento era evidente para todas as mutações testadas. Figura 4 ilustra resultados representativos. Por exemplo, o ADN a partir de células do carcinoma hepatocelular inicialmente diluídos em células não mutantes até à proporção abaixo de 5 % de mutante para não mutante, torna-se numa proporção ~ 70 % mutante para não mutante através do protocolo de enriquecimento, como estimado por observação da sequenciação no cromatograma (isto é, enriquecimento por um fator de ~14). Do mesmo modo, o ADN a partir das células SW480 (mutação homozigótica G>A no codão 273) diluída por um fator de 10 em células não mutante é enriquecido por um fator de ~ 7, através do procedimento de enriquecimento. Enriquecimento por um fator de 12 para a amostra CT7 (mutação heterozigótica C>A) e por um fator de 6 para a amostra de MDA-MB231 (mutação heterozigótica C>T) também foram observados. A amostra não mutante de p53, amplificado pelo método de enriquecimento não indicou qualquer mutação (Figura 4). No geral, todas as mutação de p53 que foram estudadas para o fragmento de 167 pb, conforme listado na Figura 2, o enriquecimento variou de 5-14 vezes. Assim, o

processo de enriquecimento aumentou a prevalência de todas as sequências contendo mutações independentemente do local onde se encontra a mutação.

Mutações do codão 12/13 de Kras: Figura 5 ilustra os resultados do fragmento de 135 pb a partir de Kras. Os resultados foram comparados com aglomerados da RCP normal realizada a uma temperatura de desnaturação de 94 °C e seguida pela sequenciação de Sanger. Figura 5 ilustra que as mutações abaixo dos 33 % de mutante para não mutante, podem ser claramente detetadas usando a sequenciação de Sanger.

Exemplo 2: Sequenciamento de Sanger de amostras clínicas

Para aplicar a presente invenção em análises de amostras clínicas, 20 amostras clínicas do tumor do cólon e do adenocarcinoma do pulmão previamente sequenciadas seguindo a RCP convencional foram submetidas ao protocolo de enriquecimento e de sequenciação de Sanger, como descrito no Exemplo 1. Os resultados indicam que todas as mutações identificadas através da sequenciação RCP-Sanger convencional também foram identificadas através do procedimento de enriquecimento seguido por sequenciação. No entanto, o processo de enriquecimento também identificou mutações que foram perdidas pela sequenciação convencional. A Figura 6 demonstra duas amostras clínicas, TL64 e CT20, onde mutações de baixa prevalência G>A foram detetadas através do protocolo de enriquecimento de sequenciamento Sanger do codão 273 do exão 8 de p53, mas não foram detetados por sequenciação da RCP convencional. Uma verificação independente da presença de mutações foi conduzida usando a sequenciação baseada em RFLP.

Além disso, a mutação (G>A) no exão 8 de p53 foi detetada no ADN do plasma circulante de 5 pacientes com cancro do cólon, através do procedimento de enriquecimento - sequenciação de Sanger, e não por sequenciação por RCP-Sanger convencional. Em seguida, mutações no p53 (C>T) que foram perdidas por sequenciação convencional foram também reveladas utilizando o ADN obtido a partir de uma espécime fixado em formalina (FFPE) obtido a partir de uma célula não pequena de um paciente do cancro do pulmão (NSCLC) (Figura 6). A cromatografia inferior na Figura 6 demonstra a deteção de mutações do codão 12 de Kras numa outra amostra de FFPE, obtida a partir de um paciente NSCLC. As mutações identificadas através do procedimento de enriquecimento foram subsequentemente verificadas de forma independente a partir de ADN genómico através de métodos de RFLP. Assim, mutações relevantes não encontradas pelo sequenciamento-RCP convencional, tornaram-se facilmente detetáveis utilizando a sequenciação RCP-CODB.

Exemplo 3. Mutações que reduzem a T_m podem ser enriquecidas sem a etapa de desnaturação emparelhada.

A dependência da RCP na sequência de nucleótidos é tão pronunciada, que quando a temperatura de desnaturação é ajustada à temperatura crítica (T_c) mesmo sem a formação de uma disparidade durante a RCP, existe enriquecimento dessas mutações que diminuem a T_f . Assim, quando um alelo G e um A estão presentes, o alelo A será enriquecido durante RCP-CODB uma vez que este diminui a T_m do alelo. Para demonstrar este ponto, e também para examinar a dependência do enriquecimento no tamanho da sequência examinada, um fragmento de 87 pb e um fragmento de 210 pb contendo as mesmas mutações que o fragmento de 167 pb do exão 8 de p53, foram examinados (Figura 3). Tal como aconteceu com o

fragmento de 167 pb, estes dois fragmentos foram amplificados a partir do produto de amplificação do exão 8 de p53 inicial através da RCP aglomerada seguido pelo protocolo de enriquecimento. No entanto, neste caso, foi utilizada uma versão truncada do protocolo de amplificação do Exemplo 1, mas a etapa de formação de disparidades a 70 °C, bem como a etapa 94 °C foram ambas omitidas (temperaturas de desnaturação crítica $T_c = 83,5$ °C para o fragmento 87 pb e $T_c = 87,5$ para o fragmento 210 fragmento). Portanto, nesta versão do protocolo de enriquecimento dos ciclos da RCP e apenas entre a temperatura de desnaturação crítica (T_c), a etapa de ligação do iniciador (por exemplo, 55 °C) e a etapa de síntese do iniciador (por exemplo, 72 °C).

Figura 7A mostra cromatografias de sequenciamento para o fragmento de 87 pb (sequenciamentos, quer no sentido direto ou reverso, foram realizados dependente do local onde a mutação se encontra na sequência de 87 pb). Os dados indicam que, utilizando esta versão modificada do protocolo de enriquecimento, enriquece o fragmento de 87 pb em 20-50 vezes. Por exemplo, uma diluição inicial de 1 % de ADN mutante para não mutante do ADN SW480 resultou em 50 % do mutante para o não mutante seguindo o processo de enriquecimento, ou seja enriquecimento ~ 50 vezes. O subsequente sequenciamento de Sanger revelou uma sequência "heterozigota". O efeito do tamanho no enriquecimento através do protocolo de enriquecimento abreviado o é demonstrado na Figura 7B. Estes dados ilustram que o enriquecimento é mais elevado para os fragmentos <100 pb, mas ainda claramente evidente (~8-10 vezes) para os fragmentos de até 210 pb.

Exemplo 4- Mutações que aumentam ou diminuem a T_f podem ser enriquecidas através do método de enriquecimento completo.

Embora a maioria (~ 70 %) das mutações encontradas em diversas amostras de cancro diminui a T_f , ~15 % das mutações aumentaram a T_f (por exemplo, A>G), enquanto 15 % retiveram a T_f (por exemplo, G>C). A fim de ser capaz de enriquecer todas as mutações possíveis, incluindo ambas as mutações e eliminações de G>A e A>G, o programa de enriquecimento completo é preferível (Figura 1). Para demonstrar a capacidade de enriquecer sequências alvo, tendo mutações que aumentam ou diminuem a T_m , o fragmento de 167 pb do exão 8 de p53 com nucleótidos C ou T (de linha celular do tipo não mutante vs HCC) foram amplificadas por meio do protocolo de enriquecimento da Fig. 1.

Duas misturas foram formadas, uma com o alelo C em minoria (C: T 1:10) e outro com o alelo T em minoria (T: C 1:10). Seguindo o procedimento de enriquecimento, ou alternativamente a RCP convencional, os produtos foram sequenciados. Em ambos os casos, o alelo menor foi enriquecido, isto é, quer o C ou T, consoante o que estava mais diluído antes da amplificação. Presumivelmente, as sequências incompatíveis têm uma temperatura de fusão mais baixa do que o alelo C ou o T, por conseguinte, ao realizar o protocolo da Figura 1, as sequências incompatíveis são sempre preferencialmente desnaturadas. Por conseguinte, através da formação de uma incompatibilidade a uma temperatura intermédia (~ 70 °C) durante o protocolo de enriquecimento há sempre o enriquecimento do alelo menor, mesmo se a alteração de nucleótido específico tende a aumentar a T_m local.

Exemplo 5- Sequenciação de MALDI-TOF

A presente invenção também está prevista para melhorar outras tecnologias baseadas em RCP, incluindo MALDI-TOF para a detecção de mutações somáticas. Para demonstrar este ponto, o mesmo modelo aplicado no Exemplo 1, através da sequenciação de Sanger para identificar mutações específicas na p53 exão 8 utilizando uma série de diluições de linhas de células que contêm mutuações em amostras de tipo não mutante, foi utilizado para comparar o protocolo da RCP de enriquecimento contra o convencional seguido por MALDI-TOF.

Seguindo o protocolo da RCP de enriquecimento ou convencional, os dNTP em excesso foram removidos a partir da reação por incubação com 0,3 U da fosfatase alcalina de camarão (USB) a 37 °C durante 20 minutos, seguido de uma incubação de 5 minutos a 85 °C para desativar a enzima. As extensões do iniciador único ao longo do SNP ou de adição/supressão foram realizadas a uma concentração final de: 600 nM de cada iniciador extensão, 50µM de d/ddNTP e 0,126 U de Termosequenase (*Solis Biodyne*) e foram incubadas a 94°C durante 2 minutos, seguido por 45 ciclos de 94 °C durante 5 segundos, 52°C durante 5 segundos, e 72°C durante 5 segundos. Os iniciadores de extensão usados foram desenhados pelas instalações de Harvard MALDI-TOF Harvard para cada mutação estudada p53, utilizando o *software* de desenho de ensaios MassARRAY versão 3.1.2.2. Os iniciadores utilizados para cada mutação foram:

p53_sw480: CAGGACAGGCACAAACA;
 p53_CT7: AGGACAGGCACAAACAC;
 p53_DU145: ACAGCTTTGAGGTGCGT;
 KRAS_SW480: TGTGGTAGTTGGACCTG;
 KRAS_A549: ACTCTTGCCTACGCCAC.

A reação foi então dessalinizada por adição de uma resina de troca catiónica, seguido por mistura e centrifugação para assentar o conteúdo do tubo. O produto de extensão foi espalhado em 384 poços do spectroCHIP antes de ser executado no espectrômetro de massa MALDI-TOF (*Sequenom*).

Os resultados estão ilustrados na Tabela I. O fator de enriquecimento da mutação encontra-se na terceira coluna da Tabela I. O enriquecimento é calculada por comparação com os valores obtidos quando a RCP convencional MALDI-TOF é aplicado. Factores de enriquecimento de 10-60 são obtidos para a maioria das mutações estudadas. Os enriquecimentos aumentam com a diminuição da proporção de mutantes para não mutantes, indicando uma dependência não linear do fator de enriquecimento na concentração inicial de mutações.

TABELA I

São indicadas as % de mutações de alelos de p53 mutantes ou não mutantes para várias mutações do exão 8 de p53			
RESULTADO MALDI-TOF			RCP- ENRIQUECIMENTO
mutação =14484G>A exão 8 de p53			
LINHAS CELULARES DILUÍDAS RASTREADAS	% de mutação A	% G não mutante	ENRIQUECIMENTO
SW480- não mutante rácio 1:5 RCP convencional	15	85	
SW480- não mutante rácio 1:10 RCP convencional	5 (Limite MassSpec)	95	
SW480- não mutante rácio 1:33 RCP convencional	0	100	

São indicadas as % de mutações de alelos de p53 mutantes ou não mutantes para várias mutações do exão 8 de p53

RESULTADO MALDI-TOF			
SW480- não mutante rácio 1:10 RCP CO.D.B	45	55	- 9
SW480- não mutante rácio 1:100 RCP CO.D.B	33	67	> 10
SW480- não mutante rácio 1:300 RCP CO.D.B	31	69	>30
APENAS NÃO MUTANTE, RCP CO.D.B	0	100	Nenhuma
RESULTADO MALDI-TOF			
mutação =14483C>T exão 8 de p53			RCP-CO.D.B
LINHAS CELULARES DILUÍDAS RASTREADAS	% de mutação T	% G não mutante	ENRIQUECIMENTO
CT7- não mutante rácio 1:5 RCP convencional	28	72	N/A
CT7- não mutante rácio 1:10 RCP convencional	18	82	N/A
CT7- não mutante rácio 1:33 RCP convencional	5 (Limite MassSpec)	95	N/A
CT7- não mutante rácio 1:100 RCP CO.D.B	45	54	-18
CT7- não mutante rácio 1:200 RCP CO.D.B	28	72	-40
CT7- não mutante rácio 1:300 RCP CO.D.B	27	73	~60
APENAS NÃO MUTANTE, RCP CO.D.B	0	100	Nenhuma
RESULTADO MALDI-TOF			
mutação =GGT>AGT codão 12 de Kras			RCP-CO.D.B
LINHAS CELULARES DILUÍDAS RASTREADAS	% de mutação G	% T não mutante	ENRIQUECIMENTO
DU145- não mutante rácio 1:5 RCP convencional	28	72	N/A
DU145- não mutante rácio 1:10 RCP convencional	18	82	N/A
DU145- não mutante rácio 1:33 RCP convencional	0	100	N/A
DU145- não mutante rácio 1:33 RCP CO.D.B	27	73	-7
DU145- não mutante rácio 1:100 RCP CO.D.B	12	88	>10

RESULTADO MALDI-TOF			
DU145- não mutante rácio 1:300 RCP CO.D.B	0	100	Não detetável
APENAS NÃO MUTANTE, RCP CO.D.B	0	100	Nenhuma
Kras mutations			
São indicadas % de alelos mutantes ou não mutante para codão 12 de Kras			
RESULTADO MALDI-TOF			
mutação=GGT>AGT do codão 12 de Kras			RCP-CO.D.B
LINHAS CELULARES DILUÍDAS RASTREADAS	% de mutação A	% G não mutante	ENRIQUECIMENTO
A549- não mutante rácio 1:10 RCP convencional	25	75	N/A
A549- não mutante rácio 1:33 RCP CO.D.B	40	60	- 5,5
A549- não mutante rácio 1:100 RCP CO.D.B	25	75	- 10
A549- não mutante rácio 1:200 RCP CO.D.B	12	88	- 10
mutação=GGT>GTT do codão 12 de Kras			RCP-CO.D.B
LINHAS CELULARES DILUÍDAS RASTREADAS	% de mutação G	% T não mutante	ENRIQUECIMENTO
SW48- não mutante rácio 1:10 RCP convencional	15	85	N/A
SW48- não mutante rácio 1:33 RCP CO.D.B	15	85	-3
SW48- não mutante rácio 1:100 RCP CO.D.B	7	93	-5
SW48- não mutante rácio 1:200 RCP CO.D.B	5	95	-7
APENAS NÃO MUTANTE, RCP CO.D.B	0	100	Nenhuma
Comparação de RCP-MALDI-TOF COM ENRIQUECIMENTO e RCP-MALDI-TOG convencional.			
O enriquecimento da mutação conseguido é listado na última coluna.			

Exemplo 6-Comparação de TAQMAN baseada em RCP em tempo real convencional e RCP em tempo real enriquecida.

Reações de amplificação de ácidos nucleicos foram realizadas para comparação com os ensaios de sondas TAQMAN em tempo real convencional e em RCP em tempo real

enriquecido. Para comparar os dois métodos de detecção em tempo real tanto o do ADN genómico e o das amostras de tumores clínicos com uma mutação G>A na p53 no exão 8 foram testadas em várias diluições em sequência de tipo não mutante. Uma série de diluições (1:3,1:10,1:30,1:100 e 1:300) de ADN genómico foram feitas a partir de SW480 em ADN de tipo não mutante. Especificamente, as reações da RCP em tempo real foram realizadas diretamente a partir de 20 ng de ADN genómico na presença de 0,2 μ M da sonda Taqman 5'-6-Fam-TTT GAG GTG CAT TTG TGT CCG-BHQ_1-3' que coincidem, na totalidade à sequência que contém a mutação na p53 no ADN das células SW480. Concentrações finais de outros reagentes foram: 1X GoTaq Flexi tampão (Promega), 1X GoTaq polimerase (Promega) 0,2 mM de cada dNTP, 0,2 μ M de iniciadores de sentido direto, 5'-TGG TCT AAT ACG ACT GGG-3', 0,2 μ M iniciador de sentido reverso, 5'-CGG AGA TCC TCT TCT TTC-3', 3 mM de $MgCl_2$, mais ADN. O tamanho do produto amplificado da RCP foi de 87 pb e $T_c = 83,5$ °C. O ciclo rápido RCP-CODB foi: 95 °C, 120 s; (95 °C, 15 segundos, 58 °C leitura de fluorescência ON, 60 segundos) x 25 ciclos; (83,5 °C 15 segundos; 58 °C leitura de fluorescência ON, 60 segundos) x 25 ciclos. Para os ciclos da RCP convencional, foi utilizado o mesmo programa, mas a temperatura de desnaturação durante toda RCP foi de 95 °C. As experiências foram repetidas, pelo menos 5 vezes em experiências independentes.

Gráficos da amplificação, ilustrando a sensibilidade da RCP em tempo real convencional e enriquecida, aplicados ao ADN genómico da linha celular de cancro do cólon SW480 estão apresentados na FIG 8A e 8. A Fig. 8B demonstra que RCP em tempo real enriquecido pode detetar a presença da mutação na proporção de 1:300 rácio entre o alelo mutante e o não mutante. Em contraste, a RCP em tempo real convencional conduzida sob condições idênticas, com a exceção do passo

de Tc, pode detetar apenas o mutante na diluição máxima de 1:10 (FIG. 8A). Por conseguinte, a sensibilidade do ensaio é de 30 vezes melhor utilizando o procedimento de enriquecimento.

Gráficos da amplificação, comparando a sensibilidade da RCP em tempo real convencional e enriquecida em amostras de tumor clínicos contendo a mutação da p53 exão 8 (uma das quais se sabe que contém uma mutação de baixo nível (5 % de mutante para não mutante) na p53 exão 8, CT20 estão ilustradas na FIG. 8C e 8D. RCP em tempo real enriquecida foi capaz de detetar facilmente a mutação (Fig. 8D), enquanto RCP convencional não (Fig. 8C). O restante da amostra (TL6, TL8 e TL18), que foram conhecidos como sendo amostras de tipo não mutante, não amplificaram (Fig. 8C) sob as mesmas condições.

Exemplo 7-Rastreo de mutações através de RCP em tempo real

Misturas reacionais de amplificação de ácidos nucleicos foram criados para comparar a capacidade da RCP em tempo real convencional e enriquecido, utilizando corantes de detecção de ADN para detetar amostras que contêm mutações em qualquer lugar ao longo do exão 8 de p53. O método proporciona um método rápido e conveniente para identificar mutações heterozigóticas desconhecidas ou SNPs. Assim, o presente método pode ser adaptado por um especialista para examinar grandes números de genes para a linha germinativa ou mutações somáticas com uma variedade de aplicações diferentes (por exemplo, a verificação das mutações BRCA1/2 em populações de alto risco para o desenvolvimento do cancro mama/ovário, a rastreio de redes de transcrição genética inteiros para identificar mutações, etc.).

FIG. 9A e 9B ilustram gráficos da amplificação, comparando ambas as RCP em tempo real convencional e enriquecida, utilizando corantes LC-verde numa variedade de linhas celulares e amostras clínicas que são conhecidas por conterem mutações na p53 exão 8. Os dados mostram que a RCP em tempo real convencional (Fig. 9A) é incapaz de distinguir os mutante (SW480, TL6 e CT20) e as amostras não mutantes (R27, TL8, TL18, TL81 e TL82), quanto que a RCP em tempo real enriquecida é capaz de distinguir (FIG. 9B). O método enriquecido proporciona antecipadamente limites de detecção para amostras que contém mutações, face a amostras não mutantes.

FIG. 9C e 9D ilustram os resultados das reações de amplificação preparadas nas condições de reação idênticas às ilustradas na FIG.9A e 9B, contudo aqui as amostras são todas de tumores pulmonares. Os dados mostram que a RCP em tempo real convencional (FIG. 9C) não consegue diferenciar amostras mutante de amostras não mutante, enquanto que a RCP em tempo real enriquecida consegue (FIG. 9D). O método enriquecido proporciona antecipadamente limites de detecção para amostras que contém mutações, face a amostras não mutantes.

Além disso, o método de detecção enriquecido foi capaz de identificar uma mutação C>T, anteriormente desconhecida, na amostra TL6.

Exemplo 8-Solventes orgânicos aumentam o enriquecimento de mutações durante as RCP em tempo real convencional e melhorada

Misturas reacionais de ácidos nucleicos de amplificação foram preparadas para avaliar o impacto de solventes

orgânicos nas RCP em tempo real convencional e enriquecida. As reações foram realizadas, quer na presença, quer na ausência, de um solvente orgânico (3 % de DMSO). O procedimento usado foi o mesmo que o descrito no Exemplo 7, e representado na FIG. 9C e 9D, exceto para a adição de 3 % de DMSO.

A presença do solvente orgânico reforçou a discriminação entre gráficos de amplificação, entre as amostras que continham sequências mutadas e as amostras que continham sequências não mutadas (Fig. 10). Por exemplo, a diferença do limite entre as amostras não mutantes e as amostras mutantes aumentou de ~ 5 ciclos (enriquecimento em tempo real sem DMSO, ver Fig. 9A) para mais de 10 ciclos (enriquecimento em tempo real com DMSO, ver a Fig. 10A).

Exemplo 9- Detecção de mutações com níveis de detecção ultrabaixos usando PFR combinado com RCP enriquecida

RCP enriquecida combinada com PFR-RCP pode ser usada para melhorar a identificação de mutações com níveis de detecção ultrabaixos, por exemplo, para a identificação de mutações aleatórias num genoma de cancro ou mutações de resistência em amostras de cancro, que se encontram num estado muito precoce, isto é, antes do tratamento.

Por exemplo, as amostras que contêm o exão 19 de EGFR não mutante foram digeridas seletivamente com a enzima TaqI. Diluições até 1:10000 de mutante para ADN genómico, foram então sujeito a RCP, quer num formato de enriquecimento ($T_c = 81,5$ ou 81°C), quer num formato de RCP convencional (95°C). O enriquecimento da mutação foi quantificado pela digestão com TaqI, seguida por dHPLC (FIG. 11). A quantidade da mutação presente foi medida pela presença de

um pico de mutação separado, que ocorreu a cerca de 7 minutos do tempo de retenção. Após o processo de enriquecimento, o pico da mutação é muito mais evidente do que o pico aquando RCP convencional. Em conclusão, o processo de enriquecimento melhorou substancialmente a deteção de mutações com níveis de deteção muito baixos, através da identificação por PCR-RFLP.

Embora esta invenção tenha sido particularmente apresentada e descrita com referências a formas de realização preferenciais, deverá ser compreendido pelos peritos na especialidade, que podem ser feitas várias alterações na forma e detalhes, sem que se afastasse do âmbito da invenção abrangida pelas reivindicações em anexo.

Lisboa, 14 de Agosto de 2015

REIVINDICAÇÕES

1. Um método para o enriquecimento de uma sequência alvo numa mistura reacional, em que o referido método compreende:

a. a sujeição de uma mistura reacional suspeita de conter um dúplex de sequência alvo e um dúplex de sequência de referência, a uma primeira temperatura de desnaturação que é superior à temperatura de fusão (T_f) do dúplex de sequência de referência e à T_f do dúplex de sequência alvo, de modo a permitir a desnaturação do dúplex de sequência alvo referido e do dúplex de sequência de referência referido, em que o dúplex de sequência alvo referido compreende uma ou mais adições, eliminações ou alterações, e difere pelo menos num nucleótido face ao dúplex de sequência de referência referido, e em que a sequência alvo referida é pelo menos 50 % homóloga face ao dúplex de sequência de referência referido, sendo amplificável pelo mesmo par de iniciadores usado para o dúplex de sequência de referência referido, e em que a sequência alvo é menos prevalente na mistura reacional do que a sequência de referência;

b. a redução da temperatura da mistura reacional para uma temperatura de hibridização, de modo a permitir a formação dos dúplexes de cadeia de referência/cadeia alvo e impedir a ligação do par de iniciadores referido às cadeias desnaturadas de referência e alvo na mistura reacional, na qual a temperatura de hibridização é inferior à temperatura crítica (T_c) e superior à temperatura de emparelhamento do iniciador;

c. aumento da temperatura da mistura reacional de amplificação referida para uma temperatura crítica (T_c), que é inferior à T_f do dúplex de sequência de referência referido, de modo a permitir a desnaturação preferencial dos dúplexes referidos na etapa (b) para formar cadeias desnaturadas alvo e de referência;

d. redução da temperatura da mistura reacional para uma temperatura de emparelhamento do iniciador, de forma a permitir que o par de iniciadores referido emparelhe com as cadeias alvo e de referência referidas;

e. extensão do par de iniciadores referido de modo a enriquecer a sequência alvo referida, em relação à sequência de referência referida;

em que o método é repetido em dois ou mais ciclos.

2. O método da reivindicação 1 em que as sequências alvo e de referência referidas são amplificadas inicialmente pela sujeição da mistura reacional a RCP.

3. O método da reivindicação 1 em que a sequência alvo referida é um alelo mutante que difere da sequência de referência entre 1 e 10 nucleótidos.

4. O método de qualquer uma das reivindicações 1 a 3, em que:

(i) as sequências alvo e de referência referidas compreendem 25 a 500 bases, e/ou

(ii) a T_c referida é de 0,3 °C a 5 °C inferior face à T_f da referida sequência de referência, e/ou

(iii) a T_c referida é inferior à T_f da sequência alvo e/ou

(iv) o método referido é repetido em 5 e 40 ciclos e/ou

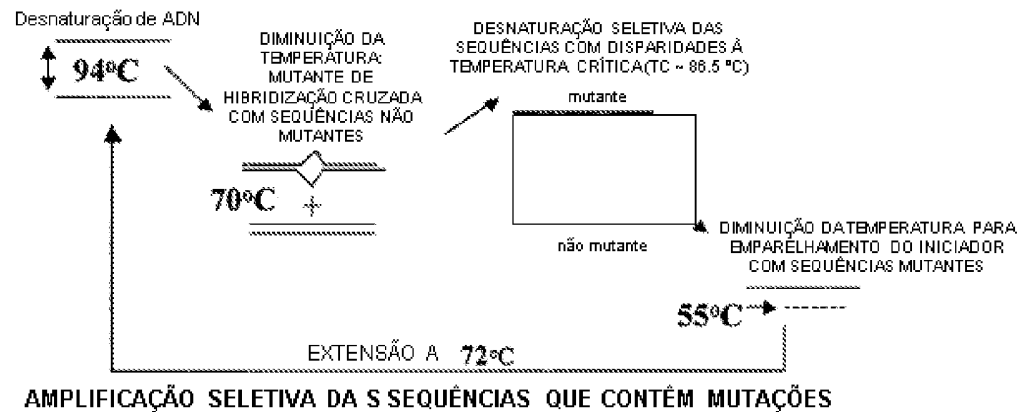
(v) o método referido é repetido em 10 e 30 ciclos.

5. O método de qualquer uma das reivindicações 1 a 4, que compreende ainda a etapa de análise da mistura reacional referida com sequência alvo enriquecida, em que a análise referida utiliza opcionalmente um ou mais dos métodos selecionados a partir do grupo que consiste em: MALDI-TOF, fusão-AR, sequenciação didesoxi, sequenciação de molécula única, pirosequenciação, SSCP, PFR, dHPLC, RCP digital e RCP-quantitativo.
6. O método da reivindicação 1 em que a Tc referida é aplicada entre 1 segundo e 5 minutos.
7. O método de qualquer uma das reivindicações anteriores, em que a sequência alvo referida está diferencialmente metilada a partir da sequência de referência, e anteriormente à implementação do método de acordo com a reivindicação 1 na mistura reacional, a mistura reacional é opcionalmente tratada com bissulfito de sódio.
8. O método de qualquer uma das reivindicações anteriores em que a mistura reacional referida contém um corante de deteção de ácidos nucleicos, e é opcionalmente realizado num dispositivo de RCP em tempo real.
9. O método da reivindicação 1 realizado sob condições reacionais em tempo real, utilizando um agente de deteção de ácidos nucleicos.

10. O método da reivindicação 1 em que o referido método é utilizado para enriquecer duas ou mais sequências alvo diferentes, e em que o método referido compreende ainda um ou mais pares de iniciadores específicos para as sequências alvo referidas.
11. O método da reivindicação 1 em que o par de iniciadores referido tem uma temperatura de fusão que é inferior à temperatura aplicada na etapa (b), opcionalmente pelo menos 5 °C abaixo da temperatura da etapa (b).
12. O método da reivindicação 1, em que a mistura reacional inclui um ácido nucleico modificado.
13. Um método de enriquecimento da sequência alvo, tal como descrito na reivindicação 1, que compreende ainda: a repetição das etapas b) e c) mais de uma vez por ciclo, antes da execução da etapa d).
14. O método da reivindicação 13, em que a sequência alvo é pelo menos 70 % homóloga face à sequência de referência, e preferencialmente pelo menos 80 % homóloga face à sequência de referência.
15. Um formato legível informaticamente que compreende instruções de *software* adaptadas para executar o método de acordo com a reivindicação 1.

Lisboa, 14 de Agosto de 2015

A PRINCÍPIO DA RCP CODB (Co-amplificação a Temperatura de Desnaturação Baixa)



B RCP CODB

(RCP CODB mediada por agentes de ligantes de caráter simplex, multiplex ou vulgar)

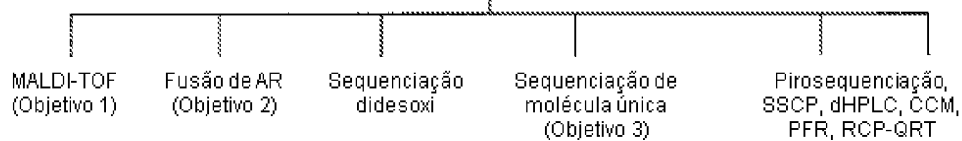
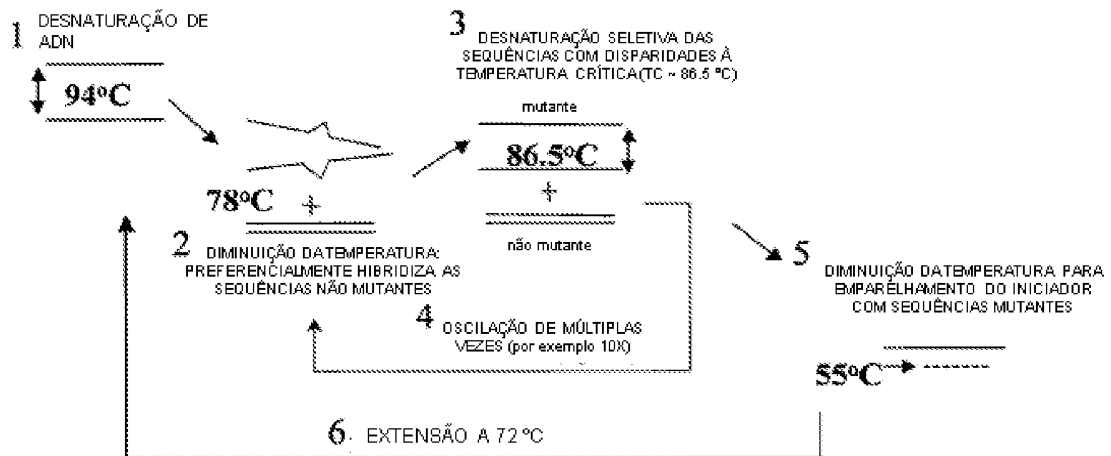


Figura 1. A. Ilustração do protocolo de enriquecimento. É apresentado um exemplo para a sequência do exão 8 de 167 pb p153. A formação de disparidades em qualquer posição ao longo da sequência durante a RCP, permite a desnaturação preferencial e amplificação dos alelos menores (mutantes) a cada ciclo da RCP. **B.** Substituição da RCP pelo método de enriquecimento: Todos os ensaios de teste genéticos com base em RCP, beneficiam do enriquecimento das mutações durante a etapa de RCP que os antecede.

PRINCÍPIO DA RCP CODB OSCILANTE

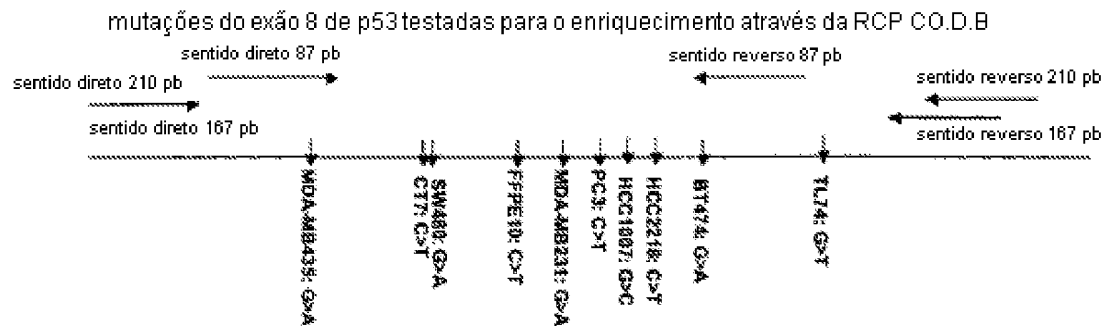
(é apresentado um exemplo da sequência do exão 8 de 87bp p53 com $T_c = 83,5$)



REPETIÇÃO DE TODA A ROTINA POR VÁRIOS CICLOS:

AMPLIFICAÇÃO SELETIVA DE SEQUÊNCIAS QUE CONTÊM MUTAÇÕES EM CADA CICLO

Figura 2: PRINCÍPIO DO PROTOCOLO DE ENRIQUECIMENTO OSCILANTE

FIGURA 3

mutações dos códons 12/13 de Kras testadas para o enriquecimento através da RCP CO.D.B

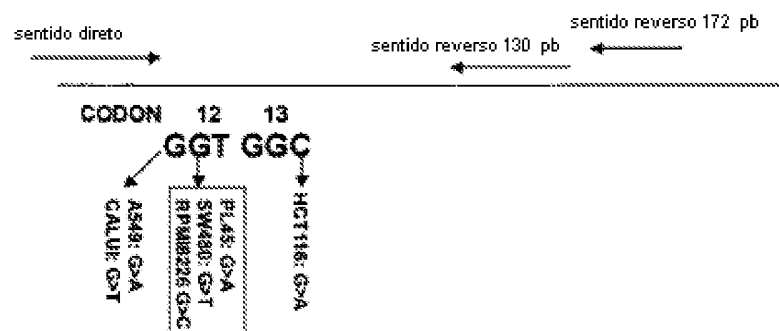


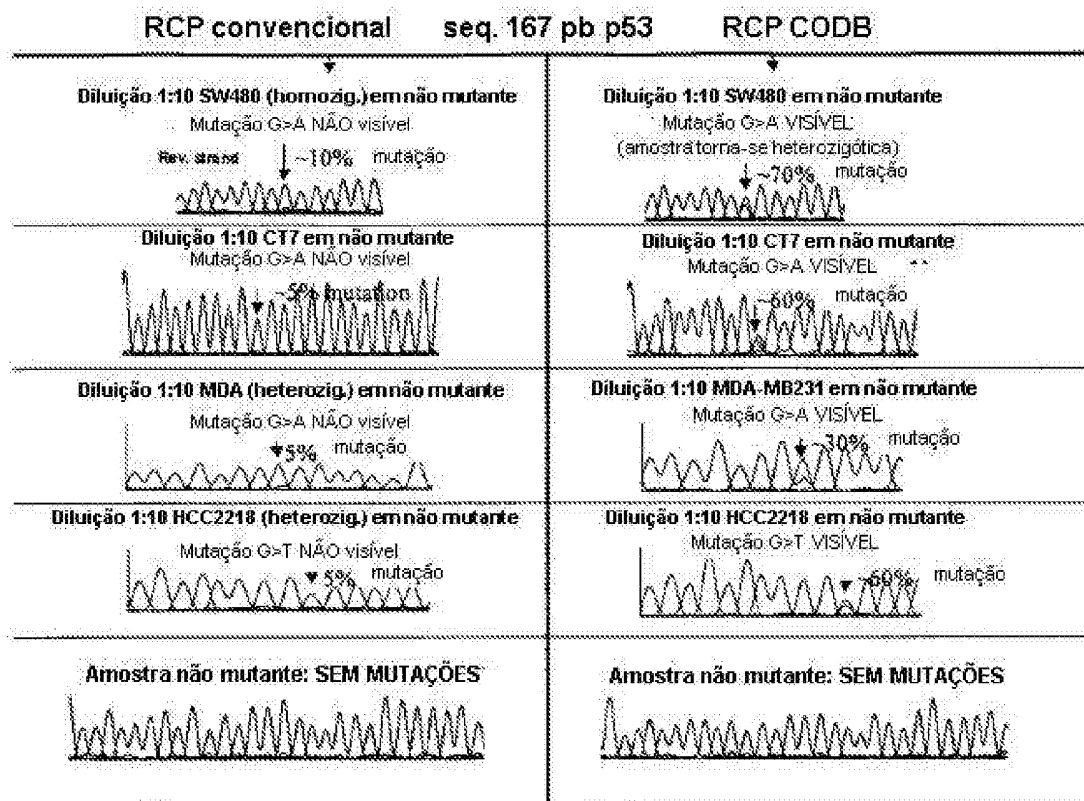
FIGURA 4

Figura 4: Comparação do enriquecimento da RCP com RCP convencional seguida da sequencição didesoxi de Sanger. As diluições de ADN proveniente de linhas celulares que contêm mutações p53 em não mutantes, revelam o enriquecimento da mutação em várias posições da sequência através do procedimento de enriquecimento.

FIGURA 5

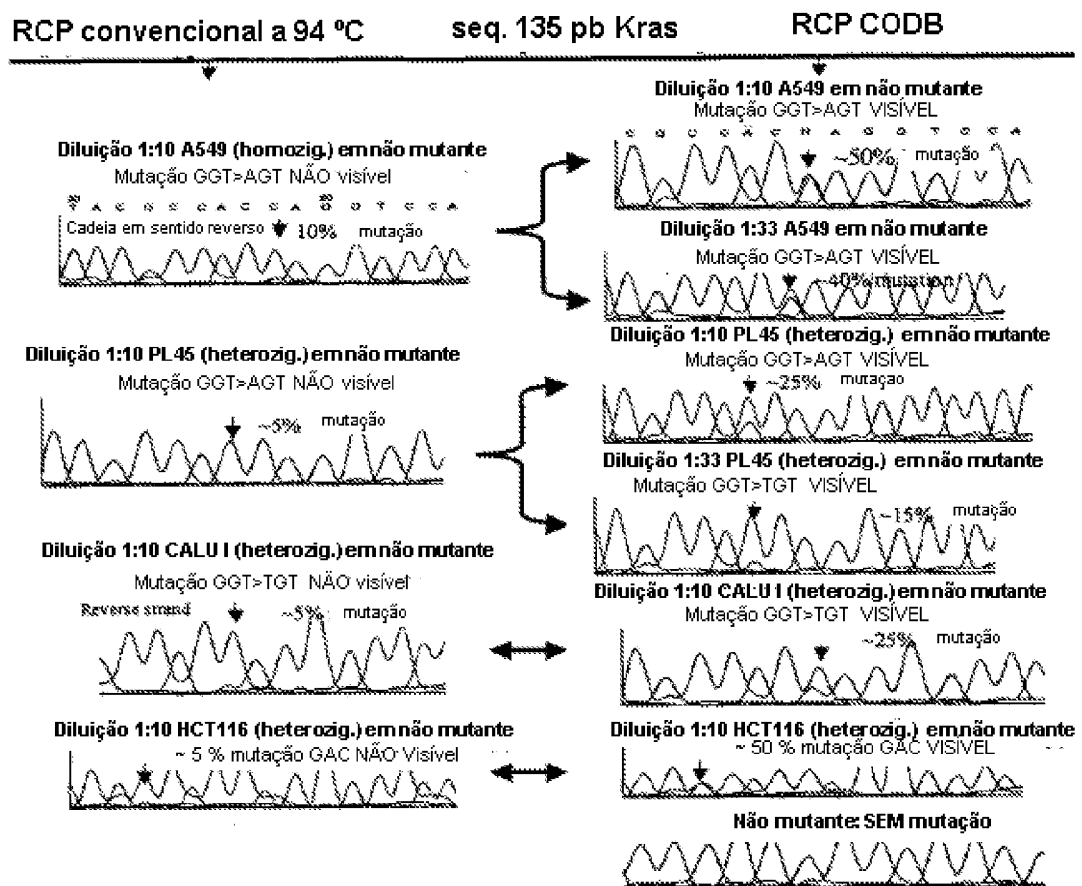


Figura 5: Comparação do enriquecimento com RCP convencional seguida da sequenciação didesoxi de Sanger. As diluições de ADN proveniente de linhas celulares que contém mutações Kras em não mutantes, revelam o enriquecimento da mutação em várias posições da sequência através de RCP CO.D.B.

FIGURA 6

EXEMPLOS DE MUTAÇÃO COM NÍVEIS DE EXPRESSÃO BAIXOS EM AMOSTRAS CLÍNICAS DE TUMORES SÓLIDOS, ANTERIORMENTE 'INVISÍVEIS' ATRAVÉS DE SEQUENCIÇÃO DIDESOXI DE SANGER, QUE AGORAM SE TORNAM VISÍVEIS ATRAVÉS DA RCP CODB

Sequencição didesoxi de Sanger de amostra CLÍNICAS de tumores para mutações do exão 8 de p53

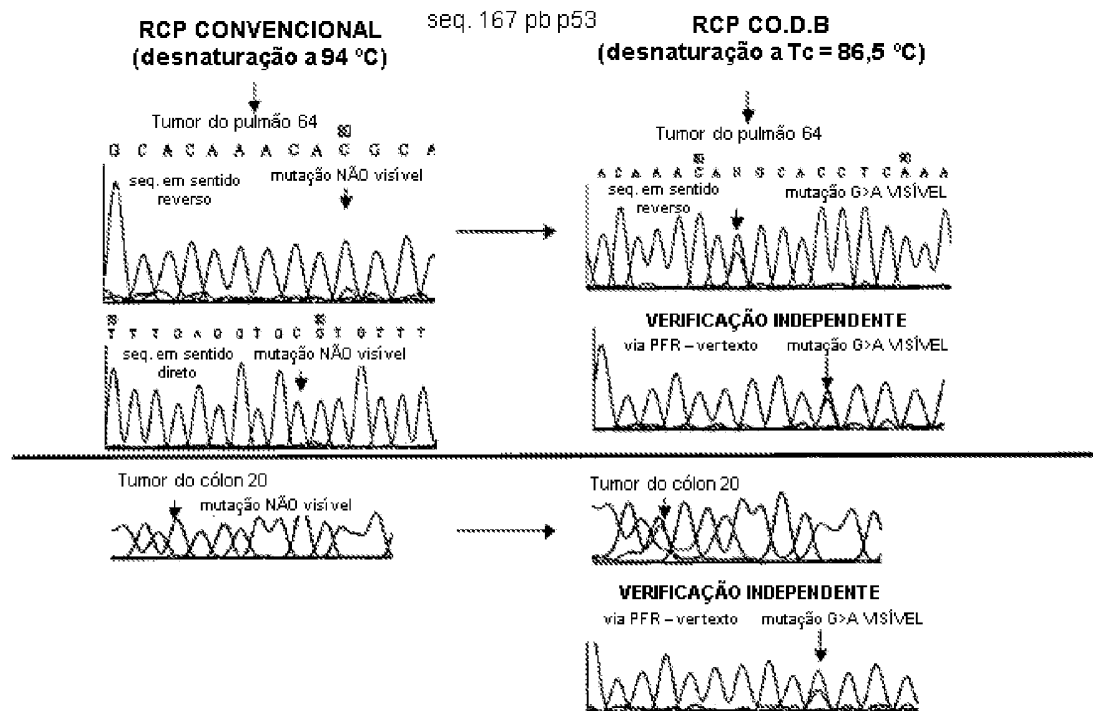


Figura 6: Amostra clínicas de tumores com baixa prevalência de mutações do exão 8 de p53: Comparação de enriquecimento vs. RCP convencional seguindo-se pela sequencição didesoxi de Sanger. O enriquecimento de mutações permite a detecção das mutações.

FIGURA 7

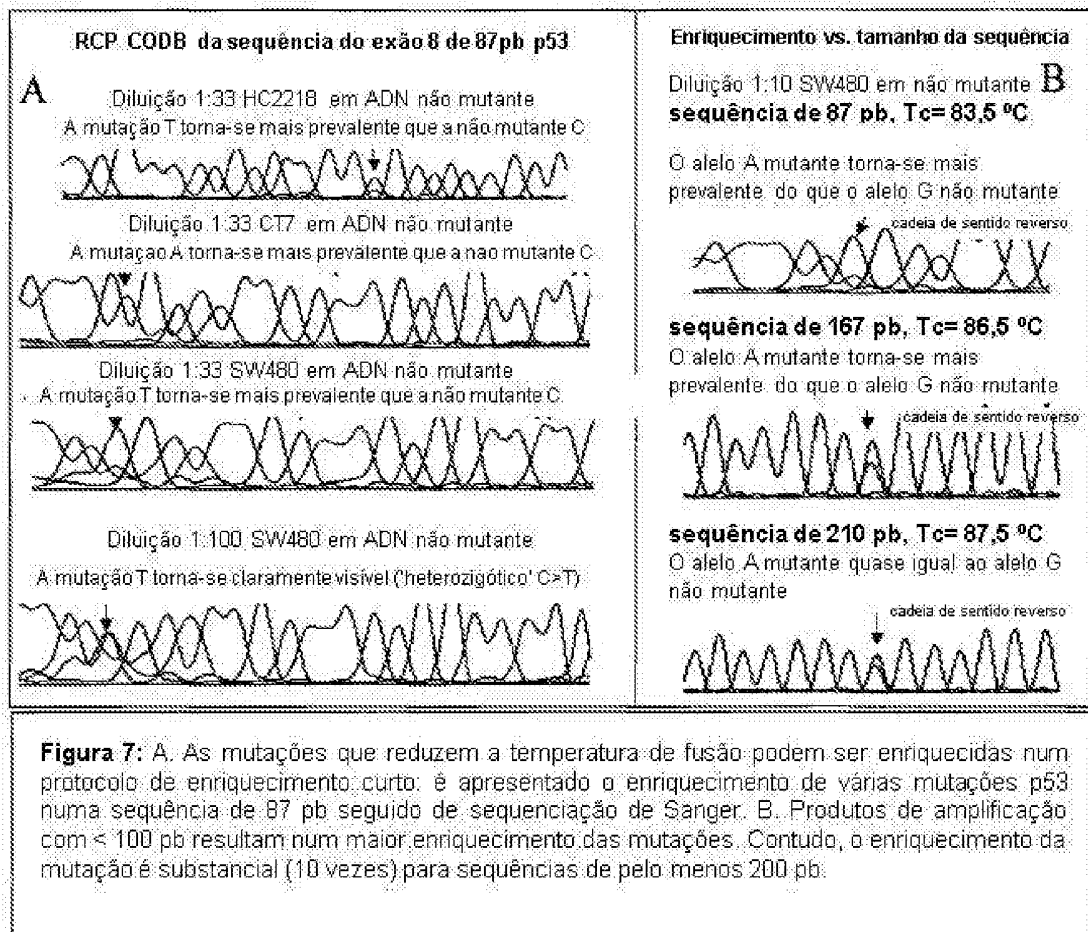


FIG. 8
DETEÇÃO DE MUTAÇÕES DO EXÃO 8 DE p53: COMPARAÇÃO
DE RCP CONVENCIONAL EM TEMPO REAL vs. RCP CODB EM
TEMPOREAL

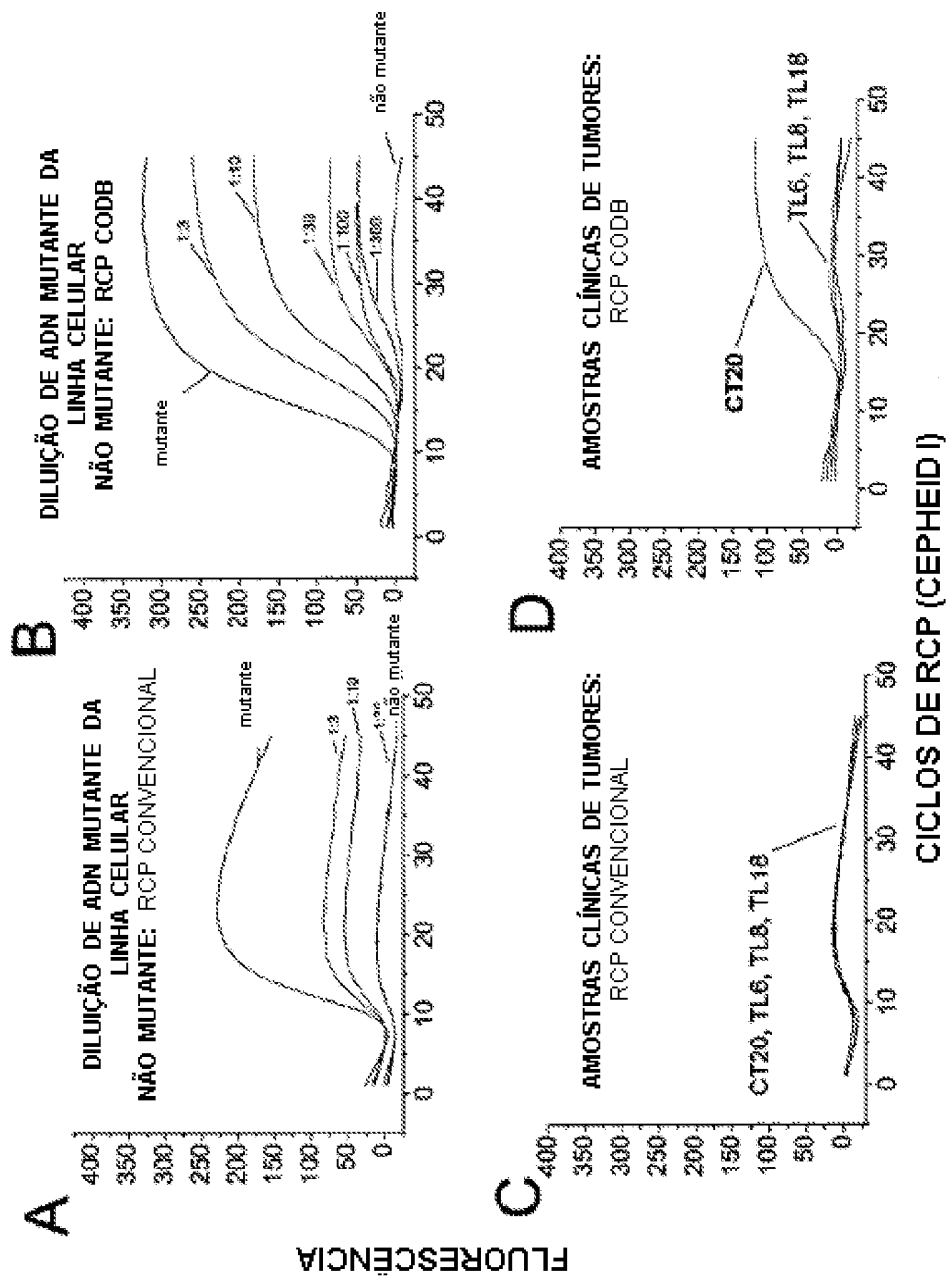


Fig 9

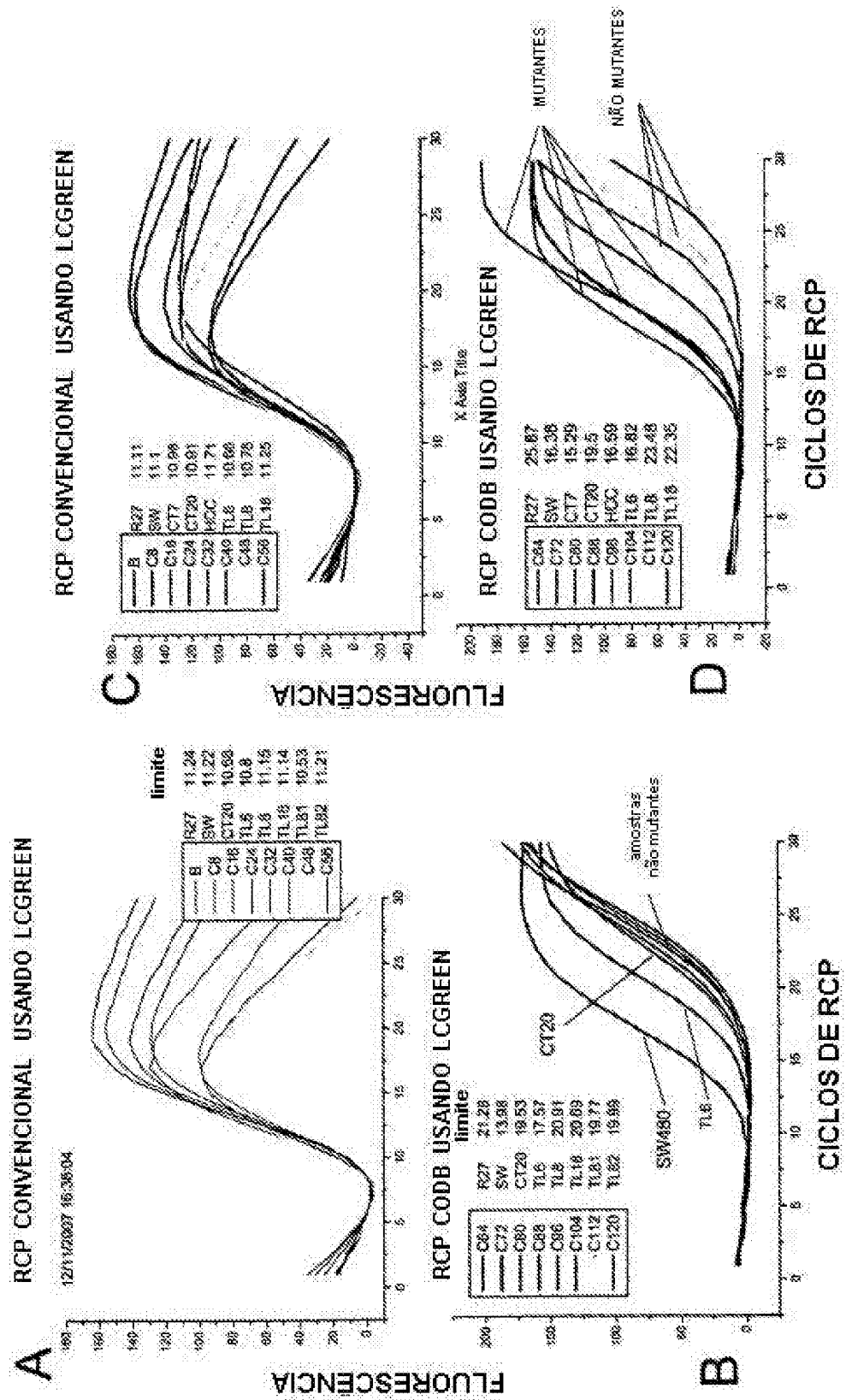


Fig 10

3 % de DMSO num tampão para o PCR: Tc varia para 80,5 °C

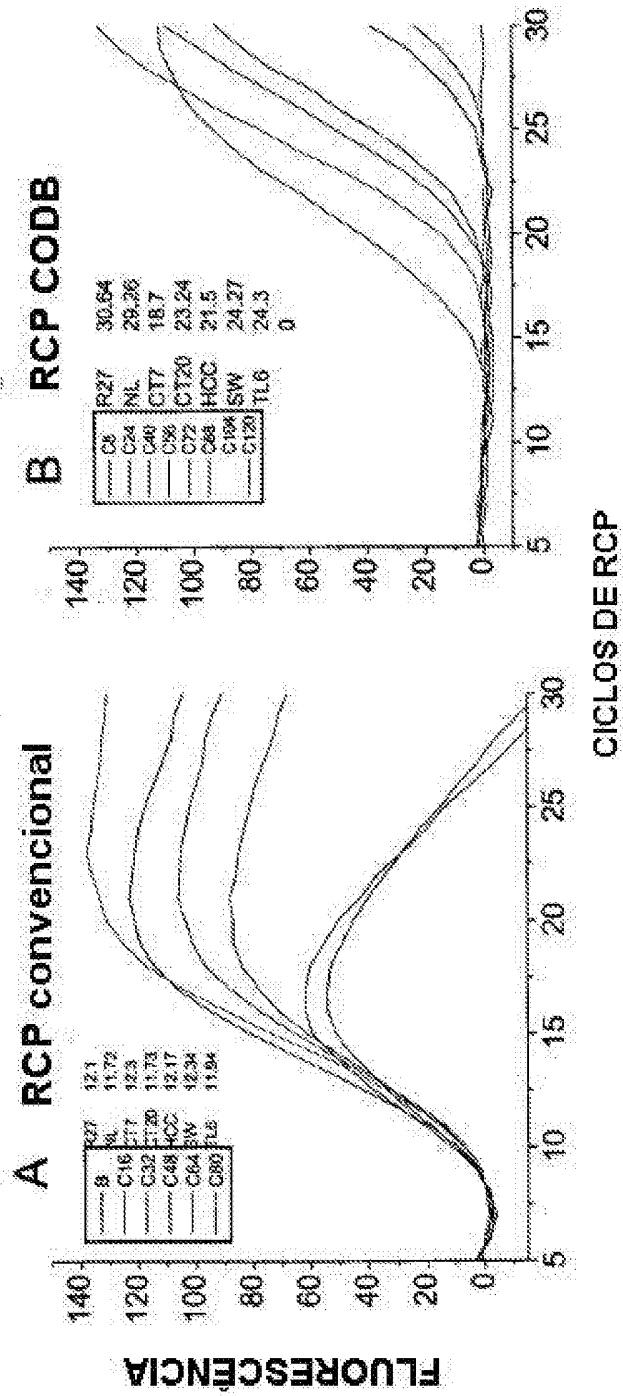


Fig 11

Melhoria da Detecção do Local de Restrição da Mutação (PFR) através da utilização da RCP CODB seguida de uma digestão seletiva com TaqI da mutação AA761 EGFR (mutação resistente): comparação de RCP convencional vs. RCP CODB

