



República Federativa do Brasil  
Ministério da Economia  
Instituto Nacional da Propriedade Industrial

**(21) BR 112021002189-2 A2**



**(22) Data do Depósito: 02/08/2019**

**(43) Data da Publicação Nacional: 04/05/2021**

**(54) Título:** MÉTODO DE SEQUENCIAMENTO QUE USA PCR MULTIPLEX DE REPLICAÇÃO VARIÁVEL

**(51) Int. Cl.:** C12Q 1/6869; G16B 30/00.

**(30) Prioridade Unionista:** 08/08/2018 US 62/716,082.

**(71) Depositante(es):** INIVATA LTD..

**(72) Inventor(es):** VINCENT PLAGNOL; TIM FORSHEW; SAMUEL WOODHOUSE; ANDREW LAWSON; MATTHEW SMITH.

**(86) Pedido PCT:** PCT IB2019056625 de 02/08/2019

**(87) Publicação PCT:** WO 2020/031048 de 13/02/2020

**(85) Data da Fase Nacional:** 05/02/2021

**(57) Resumo:** MÉTODO DE SEQUENCIAMENTO QUE USA PCR MULTIPLEX DE REPLICAÇÃO VARIÁVEL. A presente invenção refere-se a um método para análise de sequência que compreende a análise de reações de PCR em que cada uma contém diferentes porções da mesma amostra, em que pelo menos alguns dos pares de iniciadores estão em mais de uma reação de PCR e pelo menos uma das reações de PCR contém alguns, mas não todos os pares de iniciadores da(s) outra(s) reação (reações).

R1	R2	R3	R4	
—	—	—	—	A1
—		—		A2
—	—	—	—	A3
	—		—	A4
—	—	—	—	A5
—	—	—		A6
—	—	—	—	A7
	—	—		A8

## **"MÉTODO DE SEQUENCIAMENTO QUE USA PCR MULTIPLEX DE REPLICAÇÃO VARIÁVEL".**

### REFERÊNCIA CRUZADA

[0001] Este pedido reivindica o benefício do pedido provisório nº de série U.S. 62/716.082, depositado em 8 de agosto de 2018, cujo pedido é incorporado ao presente documento a título de referência.

### ANTECEDENTES

[0002] Muitas doenças são causadas por variações genéticas, por exemplo, mutações somáticas. Como as variações genéticas geralmente ocorrem apenas em uma fração das células do corpo, pode ser um desafio detectá-las por meio do sequenciamento de próxima geração (NGS). Um problema é que cada método de preparação de biblioteca e plataforma de sequenciamento resulta em leituras de sequência que contêm erros, por exemplo, erros de PCR e erros de sequenciamento. Embora às vezes seja possível corrigir erros sistemáticos (por exemplo, aqueles que estão correlacionados com parâmetros conhecidos, incluindo número de ciclo de sequenciamento, filamento, contexto de sequência e probabilidades de substituição de base), muitas vezes é impossível constatar com certeza se uma variação em uma sequência é causada por um erro ou se é uma variação genética "real". Este problema é exacerbado se a quantidade de amostra for limitada e os polinucleotídeos que contêm mutação estiverem presentes apenas em níveis relativamente baixos, por exemplo, menos de 5%, na amostra como é tipicamente o caso para DNA livre de células isolado de sangue. Por exemplo, se uma amostra contiver apenas uma cópia de um polinucleotídeo que contém mutação em um fundo de cem polinucleotídeos que são idênticos ao polinucleotídeo que contém mutação, com exceção de que os mesmos não contêm a mutação, então, após esses polinucleotídeos terem sido sequenciados, muitas vezes poderá ser impossível dizer se a variação

(que só pode ser observada em cerca de 1/100 das leituras de sequência) é um erro que ocorreu durante a amplificação ou sequenciamento. Assim, a detecção de mutações somáticas que causam doenças pode ser extremamente difícil de detectar com alguma certeza.

### SUMÁRIO

[0003] É descrito abaixo um fluxo de trabalho que facilita a identificação de variações de sequência de baixa frequência, por exemplo, DNA livre de células do sangue. Em algumas modalidades, o método pode compreender a análise de reações de PCR em que cada uma contém diferentes porções da mesma amostra, em que pelo menos alguns dos pares de iniciadores estão em mais de uma reação de PCR e pelo menos uma das reações de PCR contém alguns, mas não todos os pares de iniciadores da(s) outra(s) reação (reações). Neste método, alguns pares de iniciadores estão em mais reações do que outros, dependendo de uma série de fatores.

[0004] Em algumas modalidades, o método pode compreender:

- (a) obter múltiplos pares de iniciadores que são compatíveis em uma reação de PCR multiplex;
- (b) configurar pelo menos duas reações de PCR multiplex, em que cada uma contém diferentes porções da mesma amostra, em que pelo menos alguns dos pares de iniciadores estão em mais de uma reação de PCR e pelo menos uma das reações de PCR contém alguns, mas não todos os pares de iniciadores da(s) outra(s) reação (reações), em que, para pelo menos alguns dos pares de iniciadores que não estão em todas as reações de PCR, o número de reações que compreendem um par de iniciadores depende da importância percebida de, a probabilidade de e/ou o tipo de uma ou mais variações de sequência esperadas no amplicon amplificado pelo par de iniciadores;
- (c) termociclar as reações de PCR multiplex para produzir vários

amplicons replicados;

(d) sequenciar os amplicons para produzir leituras de sequência;

(e) analisar as leituras de sequência de amplicons replicados para uma variação de sequência selecionada para produzir uma pontuação para a variação de sequência selecionada, em que a pontuação: i. é baseada no número de amplicons replicados que compreendem uma variação de sequência que tem uma frequência acima de um corte (cut-off); ou ii. indica a força da evidência combinada para a variação de sequência entre as réplicas; e

(f) denominar uma variação de sequência com base na pontuação.

[0005] Dependendo de como o método é implementado, o método pode ter determinadas vantagens sobre os métodos convencionais. Por exemplo, o presente método pode fornecer uma maior probabilidade de identificar variações genéticas consideradas mais importantes pelos usuários do método, sem simplesmente aumentar o número de reações de PCR multiplex.

#### BREVE DESCRIÇÃO DAS FIGURAS

[0006] O versado na técnica entenderá que os desenhos, descritos abaixo, são apenas para fins ilustrativos. Os desenhos não se destinam a limitar o escopo dos presentes ensinamentos de forma alguma.

[0007] A Figura 1 ilustra esquematicamente um exemplo de um conjunto de reações de PCR multiplex que podem ser produzidas no método reivindicado. Este exemplo simplesmente ilustra alguns dos princípios do método e não deve limitar o método de forma alguma.

[0008] A Figura 2 ilustra como uma variação genética pode ser denominada com o uso do número de réplicas que têm a variação de sequência selecionada acima de uma frequência de corte.

[0009] A Figura 3 ilustra como uma variação genética pode ser denominada utilizando-se a evidência combinada para a variação genética em múltiplas réplicas.

## DEFINIÇÕES

[0010] A menos que definido de outra forma, todos os termos técnicos e científicos usados no presente documento têm a mesma relevância como comumente entendido por alguém de habilidade comum na técnica à qual esta invenção pertence. Ainda assim, determinados elementos são definidos para fins de clareza e facilidade de referência.

[0011] Os termos e símbolos da química, bioquímica, genética e biologia molecular de ácido nucleico usados no presente documento seguem os dos tratados e textos padrão do campo, por exemplo, Kornberg and Baker, DNA Replication, Segunda Edição (W.H. Freeman, New York, 1992); Lehninger, Biochemistry, Segunda Edição (Worth Publishers, New York, 1975); Strachan and Read, Human Molecular Genetics, Segunda Edição (Wiley-Liss, New York, 1999); Eckstein, editor, Oligonucleotides and Analogs: A Practical Approach (Oxford University Press, New York, 1991); Gait, editor, Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford, 1984); e similares.

[0012] O termo "nucleotídeo" se destina a incluir aquelas porções que contêm não apenas as bases purina e pirimidina conhecidas, mas também outras bases heterocíclicas que foram modificadas. Essas modificações incluem purinas ou pirimidinas metiladas, purinas ou pirimidinas aciladas, riboses alquiladas ou outros heterociclos. Além disso, o termo "nucleotídeo" inclui aquelas porções que contêm hapteno ou marcadores fluorescentes e podem conter não apenas açúcares ribose e desoxirribose convencionais, mas também outros açúcares. Os nucleosídeos ou nucleotídeos modificados também incluem modificações na porção de açúcar, por exemplo, em que um ou mais dos grupos hidroxila são substituídos por átomos de halogênio ou grupos alifáticos, ou são funcionalizados como éteres, amins ou similares.

[0013] Os termos "ácido nucleico" e "polinucleotídeo" são usados no presente documento de forma intercambiável para descrever um polímero de qualquer comprimento, por exemplo, maior do que cerca de 2 bases, maior do que cerca de 10 bases, maior do que cerca de 100 bases, maior do que cerca de 500 bases, maior do que 1.000 bases, maior do que 10.000 bases, maior do que 100.000 bases, maior do que cerca de 1.000.000, até cerca de  $10^{10}$  ou mais bases compostas por nucleotídeos, por exemplo, desoxirribonucleotídeos ou ribonucleotídeos, e podem ser produzidos enzimaticamente ou sinteticamente (por exemplo, PNA, conforme descrito na Patente nº U.S. 5.948.902 e as referências citadas na mesma) que podem hibridizar com ácidos nucleicos de ocorrência natural de uma maneira específica de sequência análoga à de dois ácidos nucleicos de ocorrência natural, por exemplo, podem participar de interações de emparelhamento de bases Watson-Crick. Os nucleotídeos de ocorrência natural incluem guanina, citosina, adenina, timina, uracila (G, C, A, T e U, respectivamente). DNA e RNA têm uma estrutura principal de açúcar de desoxirribose e ribose, respectivamente, enquanto a estrutura principal do PNA é composta por repetições de unidades de N-(2-aminoetil)-glicina ligadas por ligações peptídicas. Em PNA várias bases purina e pirimidina estão ligadas à estrutura principal por ligações de metileno carbonila. Um ácido nucleico bloqueado (LNA), muitas vezes referido como RNA inacessível, é um nucleotídeo de RNA modificado. A porção química de ribose de um nucleotídeo de LNA é modificada com uma ponte extra que conecta o oxigênio 2' e o carbono 4'. A ponte "bloqueia" a ribose na conformação 3'-endo (Norte), que é frequentemente encontrada nos duplexes de forma A. Os nucleotídeos de LNA podem ser misturados com resíduos de DNA ou RNA no oligonucleotídeo sempre que desejado. O termo "ácido nucleico não estruturado" ou "UNA" é um ácido nucleico que contém nucleotídeos não naturais que

se ligam uns aos outros com estabilidade reduzida. Por exemplo, um ácido nucleico não estruturado pode conter um resíduo G' e um resíduo C', em que esses resíduos correspondem a formas de ocorrência não natural, isto é, análogos, de G e C que se emparelham entre si com estabilidade reduzida, mas retêm uma capacidade de emparelhar pares de base com resíduos C e G de ocorrência natural, respectivamente. O ácido nucleico não estruturado é descrito no documento US20050233340, o qual é incorporado ao presente documento a título de referência para divulgação de UNA.

[0014] O termo "amostra de ácido nucleico", conforme usado no presente documento, denota uma amostra que contém ácidos nucleicos. As amostras de ácido nucleico usadas no presente documento podem ser complexas por conterem várias moléculas diferentes que contêm sequências. Amostras de DNA genômico de um mamífero (por exemplo, camundongo ou humano) são tipos de amostras complexas. As amostras complexas podem ter mais do que cerca de  $10^4$ ,  $10^5$ ,  $10^6$  ou  $10^7$ ,  $10^8$ ,  $10^9$  ou  $10^{10}$  moléculas de ácido nucleico diferentes. Qualquer amostra que contém ácido nucleico, por exemplo, DNA genômico de células de cultura de tecidos ou uma amostra de tecido, pode ser empregada no presente documento.

[0015] O termo "oligonucleotídeo", conforme usado no presente documento, denota um multímero de filamento simples de nucleotídeo de cerca de 2 a 200 nucleotídeos, até 500 nucleotídeos de comprimento. Os oligonucleotídeos podem ser sintéticos ou podem ser produzidos enzimaticamente e, em algumas modalidades, têm de 30 a 150 nucleotídeos de comprimento. Os oligonucleotídeos podem conter monômeros de ribonucleotídeos (isto é, podem ser oligorribonucleotídeos) ou monômeros de desoxirribonucleotídeos, ou tanto monômeros de ribonucleotídeos quanto monômeros de desoxirribonucleotídeos. Um oligonucleotídeo pode ter de 10 a 20, 21 a

30, 31 a 40, 41 a 50, 51 a 60, 61 a 70, 71 a 80, 80 a 100, 100 a 150 ou 150 a 200 nucleotídeos de comprimento, por exemplo.

[0016] "Iniciador" significa um oligonucleotídeo, natural ou sintético, que é capaz, ao formar um duplex com um modelo de polinucleotídeo, de atuar como um ponto de iniciação da síntese de ácido nucleico e ser estendido a partir de sua extremidade 3' ao longo do modelo de modo que um duplex estendido seja formado. A sequência de nucleotídeos adicionados durante o processo de extensão é determinada pela sequência do polinucleotídeo modelo. Os iniciadores são estendidos por uma DNA polimerase. Os iniciadores têm geralmente um comprimento compatível com o seu uso na síntese de produtos de extensão de iniciadores e estão normalmente na faixa de 8 a 200 nucleotídeos de comprimento, tal como de 10 a 100 ou de 15 a 80 nucleotídeos de comprimento. Um iniciador pode conter uma cauda 5' que não hibridiza com o modelo.

[0017] Os iniciadores são geralmente de filamento simples para máxima eficiência na amplificação, mas podem, alternativamente, ser de filamento duplo ou parcialmente de filamento duplo. Também incluídos nesta definição estão iniciadores de troca de suporte, conforme descrito em Zhang et al (Nature Chemistry 2012 4: 208 a 214), que é incorporado ao presente documento a título de referência.

[0018] Assim, um "iniciador" é complementar a um modelo, e se complexam por ligação de hidrogênio ou hibridização com o modelo para gerar um complexo de iniciador/modelo para o início da síntese por uma polimerase, o que é estendido pela adição de bases covalentemente ligadas em sua extremidade 3' complementar ao molde no processo de síntese de DNA.

[0019] O termo "hibridização" ou "hibridiza" refere-se a um processo no qual uma região de filamento de ácido nucleico passa por anelamento e forma um duplex estável, um homoduplex ou um

heteroduplex, sob condições normais de hibridização com um segundo filamento de ácido nucleico complementar, e não forma um duplex estável com moléculas de ácido nucleico não relacionadas sob as mesmas condições normais de hibridização. A formação de um duplex é realizada anelando-se duas regiões de filamento de ácido nucleico complementares em uma reação de hibridização. A reação de hibridização pode ser feita de modo a ser altamente específica pelo ajuste das condições de hibridização sob as quais a reação de hibridização ocorre, de modo que dois filamentos de ácido nucleico não formem um duplex estável, por exemplo, um duplex que retém uma região de filamento duplo sob condições normais de estringência, a menos que os dois filamentos de ácido nucleico contenham um determinado número de nucleotídeos em sequências específicas que são substancialmente ou completamente complementares. "Condições de hibridização normal ou estringência normal" são prontamente determinadas para qualquer reação de hibridização. Consultar, por exemplo, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., New York, ou Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press. Conforme usado no presente documento, o termo "hibridizar" ou "hibridização" refere-se a qualquer processo pelo qual um filamento de ácido nucleico se liga a um filamento complementar por meio do emparelhamento de bases.

[0020] Um ácido nucleico é considerado "seletivamente hibridizável" com uma sequência de ácidos nucleicos de referência se as duas sequências hibridizarem especificamente uma com a outra sob condições de hibridização de estringência moderada a alta. Condições de hibridização de estringência moderada e alta (consultar, por exemplo, Ausubel, et al., *Short Protocols in Molecular Biology*, 3ª ed., Wiley & Sons 1995 e Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Terceira Edição, 2001 Cold Spring Harbor, N.Y.).

[0021] O termo "duplex" ou "duplexado", conforme usado no presente documento, descreve duas regiões polinucleotídicas complementares que são submetidas a emparelhamento de bases, isto é, hibridizadas juntas.

[0022] "Lócus genético", "lócus", "lócus de interesse", "região" ou "segmento" em referência a um genoma ou polinucleotídeo-alvo significa uma sub-região ou segmento contíguo do genoma ou polinucleotídeo-alvo. Conforme usado no presente documento, o lócus genético, lócus ou lócus de interesse pode se referir à posição de um nucleotídeo, um gene ou uma porção de um gene em um genoma ou pode se referir a qualquer porção contígua da sequência genômica esteja ou não dentro, ou associada a, um gene, por exemplo, uma sequência de codificação. Um lócus genético, lócus ou lócus de interesse pode ser de um único nucleotídeo a um segmento de algumas centenas ou alguns milhares de nucleotídeos de comprimento ou mais. Em geral, um lócus de interesse terá uma sequência de referência associada ao mesmo (consultar a descrição de "sequência de referência" abaixo).

[0023] O termo "sequência de referência", conforme usado no presente documento, refere-se a uma sequência de nucleotídeos conhecida, por exemplo, uma região cromossômica cuja sequência é depositada no banco de dados Genbank do NCBI ou em outros bancos de dados, por exemplo. Uma sequência de referência pode ser uma sequência do tipo selvagem.

[0024] Os termos "pluralidade", "população" e "coleção" são usados de forma intercambiável de modo a se referir a algo que contém pelo menos 2 membros. Em determinados casos, uma pluralidade, população ou coleção pode ter pelo menos 10, pelo menos 100, pelo menos 1.000, pelo menos 10.000, pelo menos 100.000, pelo menos  $10^6$ , pelo menos  $10^7$ , pelo menos  $10^8$  ou pelo menos  $10^9$  ou mais membros.

[0025] O termo "sequência de identificadores de amostra", "índice de amostra", "identificador de multiplex" ou "MID" é uma sequência de nucleotídeos que é anexada a um polinucleotídeo-alvo, em que a sequência identifica a fonte do polinucleotídeo-alvo (isto é, a amostra de cuja amostra o polinucleotídeo-alvo é derivado). Durante o uso, cada amostra é marcada com uma sequência de identificadores de amostra diferente (por exemplo, uma sequência é anexada a cada amostra, em que as diferentes amostras são anexadas a diferentes sequências) e as amostras marcadas são agrupadas. Depois que a amostra combinada é sequenciada, a sequência de identificadores de amostra pode ser usada para identificar a fonte das sequências. Uma sequência de identificadores de amostra pode ser adicionada à extremidade 5' de um polinucleotídeo ou à extremidade 3' de um polinucleotídeo. Em certos casos, parte da sequência de identificadores de amostra pode estar na extremidade 5' de um polinucleotídeo e o restante da sequência de identificadores de amostra pode estar na extremidade 3' do polinucleotídeo. Quando os elementos do identificador de amostra têm sequência em cada extremidade, juntos, as sequências de identificadores de amostra 3' e 5' identificam a amostra. Em muitos exemplos, a sequência de identificadores de amostra é apenas um subconjunto das bases que são anexadas a um oligonucleotídeo-alvo.

[0026] O termo "sequência de identificadores de replicação" refere-se a uma sequência anexada que permite que leituras de sequência de diferentes réplicas sejam distinguidas umas das outras. As sequências de identificadores de réplica funcionam da mesma maneira que as sequências de identificadores de amostra descritas acima, com exceção de que são usadas em réplicas de uma amostra, em vez de amostras diferentes.

[0027] O termo "variável", no contexto de duas ou mais sequências de ácidos nucleicos que são variáveis, refere-se a dois ou mais ácidos

nucleicos que têm diferentes sequências de nucleotídeos um em relação ao outro. Em outras palavras, se os polinucleotídeos de uma população tiverem uma sequência variável, então, a sequência de nucleotídeos das moléculas de polinucleotídeos da população pode variar de molécula para molécula. O termo "variável" não deve ser lido de modo a exigir que cada molécula em uma população tenha uma sequência diferente das outras moléculas em uma população.

[0028] O termo "substancialmente" refere-se a sequências que são quase duplicadas, conforme medido por uma função de similaridade, incluindo, mas sem limitação, a uma distância de Hamming, distância de Levenshtein, distância de Jaccard, distância de cosseno etc. (consultar, de modo geral, Kemena et al, Bioinformatics 2009 25: 2.455 a 2.465). O limiar exato depende da taxa de erro da preparação de amostra e do sequenciamento usado para realizar a análise, com taxas de erro mais altas exigindo limiares mais baixos de similaridade. Em determinados casos, as sequências substancialmente idênticas têm pelo menos 98% ou pelo menos 99% de identidade de sequência.

[0029] O termo "variação de sequência", conforme usado no presente documento, é uma variante que está presente em uma frequência inferior a 50%, em relação a outras moléculas na amostra, em que as outras moléculas na amostra são substancialmente idênticas às moléculas que contêm a variação de sequência. Em alguns casos, uma variação de sequência específica pode estar presente em uma amostra a uma frequência inferior a 20%, inferior a 10%, inferior a 5%, inferior a 1% ou inferior a 0,5%.

[0030] O termo "modelo de ácido nucleico" se destina a se referir à molécula de ácido nucleico inicial que é copiada durante a amplificação. Copiar neste contexto pode incluir a formação do complemento de um ácido nucleico de filamento simples particular. O ácido nucleico "inicial" pode compreender ácidos nucleicos que já foram processados, por

exemplo, amplificados, estendidos, marcados com adaptadores, etc.

[0031] O termo "cauda", no contexto de um iniciador com cauda ou um iniciador que tem uma cauda 5', refere-se a um iniciador que tem uma região (por exemplo, uma região de pelo menos 12 a 50 nucleotídeos) em sua extremidade 5' que não hibridiza ou hibridiza parcialmente com o mesmo alvo que a extremidade 3' do iniciador.

[0032] O termo "modelo inicial" refere-se a uma amostra que contém uma sequência-alvo a ser amplificada. O termo "amplificação", conforme usado no presente documento, refere-se à geração de uma ou mais cópias de um ácido nucleico-alvo, utilizando o ácido nucleico-alvo como um modelo.

[0033] O termo "amplicon", conforme usado no presente documento, refere-se ao produto (ou "banda") amplificado por um par particular de iniciadores em uma reação de PCR.

[0034] O "amplicon replicado", conforme usado no presente documento, refere-se ao mesmo amplicon amplificado utilizando diferentes porções de uma amostra. Os amplicons replicados típicos têm sequências quase idênticas, exceto por variações de sequência no modelo, erros de PCR e diferenças nas sequências dos iniciadores usados para cada réplica (por exemplo, diferenças nas extremidades 5' dos iniciadores, tal como na sequência de identificadores de réplica, etc.).

[0035] Uma "reação em cadeia da polimerase" ou "PCR" é uma reação enzimática na qual um DNA modelo específico é amplificado com o uso de um ou mais pares de iniciadores específicos de sequência.

[0036] "Condições de PCR" são as condições em que a PCR é realizada e incluem a presença de reagentes (por exemplo, nucleotídeos, tampão, polimerase, etc.), bem como ciclicização de temperatura (por exemplo, através de ciclos de temperaturas

adequadas para desnaturação, renaturação e extensão), como é conhecido na técnica.

[0037] Uma "reação em cadeia da polimerase multiplex" ou "PCR multiplex" é uma reação enzimática que emprega dois ou mais pares de iniciadores para diferentes modelos de alvos. Se os modelos alvo estiverem presentes na reação, uma reação em cadeia da polimerase multiplex resulta em dois ou mais produtos de DNA amplificados que são coamplificados em uma única reação com o uso de um número correspondente de pares de iniciadores específicos de sequência.

[0038] O termo "iniciador específico de sequência", conforme usado no presente documento, refere-se a um iniciador que se liga apenas e se estende em um único sítio em uma amostra em estudo. Em determinadas modalidades, um oligonucleotídeo "específico de sequência" pode hibridizar com uma sequência de nucleotídeos complementar que é única em uma amostra em estudo.

[0039] O termo "sequenciamento de próxima geração" refere-se aos chamados métodos altamente paralelizados de execução de sequenciamento de ácido nucleico e compreende as plataformas de sequenciamento por síntese ou sequenciamento por ligação atualmente empregadas pela Illumina, Life Technologies, Pacific Biosciences e Roche, etc. Os métodos de sequenciamento de próxima geração também podem incluir, mas sem limitação, métodos de sequenciamento nanopore, como os oferecidos pela Oxford Nanopore, ou métodos baseados em detecção eletrônica, como a tecnologia Ion Torrent comercializada pela Life Technologies.

[0040] O termo "leitura de sequência" refere-se à saída de um sequenciador. Uma sequência lida normalmente contém uma sequência de Gs, As, Ts e Cs, de 50 a 1.000 ou mais bases de comprimento e, em muitos casos, cada base de uma sequência lida pode ser associada a uma pontuação indicando a qualidade da denominação de base.

[0041] Os termos "verificação da presença de" e "avaliação da presença de" incluem qualquer forma de medição, incluindo determinar se um elemento está presente e estimar a quantidade do elemento. Os termos "determinação", "medição", "avaliação", "verificação" e "análise" são usados indistintamente e incluem determinações quantitativas e qualitativas. A verificação pode ser relativa ou absoluta. "Verificar a presença de" inclui determinar a quantidade de algo presente e/ou determinar se está presente ou ausente.

[0042] Se dois ácidos nucléicos forem "complementares", os mesmos hibridizarão um com o outro sob condições de alta estringência. O termo "perfeitamente complementar" é usado para descrever um duplex em que cada base de um ácidos nucleicos realiza pareamento de base com um nucleotídeo complementar no outro ácido nucleico. Em muitos casos, duas sequências que são complementares têm pelo menos 10, por exemplo, pelo menos 12 ou 15 nucleotídeos de complementaridade.

[0043] Um "sítio de ligação de oligonucleotídeo" refere-se a um sítio ao qual um oligonucleotídeo hibridiza em um polinucleotídeo-alvo. Se um oligonucleotídeo "fornece" um sítio de ligação para um iniciador, então, o iniciador pode hibridizar com esse oligonucleotídeo ou seu complemento.

[0044] O termo "filamento", conforme usado no presente documento, refere-se a um ácido nucleico constituído por nucleotídeos covalentemente ligados entre si por ligações covalentes, por exemplo, ligações de fosfodiéster. Em uma célula, o DNA geralmente existe em uma forma de filamento duplo e, sendo assim, tem dois filamentos complementares de ácido nucleico referidos aqui como os filamentos "superior" e "inferior". Em determinados casos, os filamentos complementares de uma região cromossômica podem ser referidos como filamentos "mais" e "menos", o "primeiro" e o "segundo"

filamentos, os filamentos "codificante" e "não codificante", os filamentos "Watson" e "Crick" ou os filamentos "senso" e "antissenso". A atribuição de um filamento como sendo um filamento superior ou inferior é arbitrária e não implica em qualquer orientação, função ou estrutura particular. As sequências de nucleotídeos do primeiro filamento de várias regiões cromossômicas exemplificativas de mamíferos (por exemplo, BACs, conjuntos, cromossomos, etc.) são conhecidas e podem ser encontradas no banco de dados Genbank do NCBI, por exemplo.

[0045] O termo "estendendo", conforme usado no presente documento, refere-se à extensão de um iniciador pela adição de nucleotídeos com o uso de uma polimerase. Se um iniciador que é anelado com um ácido nucleico é estendido, o ácido nucleico atua como um modelo para a reação de extensão.

[0046] O termo "sequenciamento", conforme usado no presente documento, refere-se a um método pelo qual é obtida a identidade de pelo menos 10 nucleotídeos consecutivos (por exemplo, a identidade de pelo menos 20, pelo menos 50, pelo menos 100 ou pelo menos 200 ou mais nucleotídeos consecutivos) de um polinucleotídeo.

[0047] O termo "agrupamento", conforme usado no presente documento, refere-se à combinação, por exemplo, mistura, de duas ou mais amostras ou réplicas de uma amostra de modo que as moléculas dentro dessas amostras ou réplicas se tornem intercaladas entre si em solução.

[0048] O termo "amostra agrupada", conforme usado no presente documento, refere-se ao produto de agrupamento.

[0049] O termo "porção", conforme usado no presente documento, no contexto de diferentes porções da mesma amostra, refere-se a uma alíquota ou parte de uma amostra. Por exemplo, se um microlitro de 100 µl de amostra for adicionado a cada uma das 10 reações de PCR

diferentes, então, cada uma dessas reações conterà diferentes porções da mesma amostra.

[0050] Conforme usado no presente documento, os termos "DNA livre de células da corrente sanguínea", "DNA livre de células circulante" e "DNA livre de células" ("cfDNA") referem-se ao DNA que está circulando no sangue periférico de um paciente. As moléculas de DNA no DNA livre de células podem ter um tamanho médio abaixo de 1 kb (por exemplo, na faixa de 50 pb a 500 pb, 80 pb a 400 pb ou 100 a 1.000 pb), embora fragmentos que têm um tamanho médio fora desta faixa possam estar presentes. O DNA livre de células pode conter DNA tumoral circulante (ctDNA), ou seja, DNA tumoral circulando livremente no sangue de um paciente com câncer, ou DNA fetal circulante (se o sujeito for uma mulher grávida). O cfDNA pode ser obtido centrifugando-se o sangue total para remover todas as células e, em seguida, isolando-se o DNA do plasma ou soro remanescente. Esses métodos são bem-conhecidos (consultar, por exemplo, Lo et al., *Am J Hum Genet* 1998; 62:768 a 775). O DNA livre de células em circulação pode ser de filamento duplo ou filamento simples. Esse termo se destina a abranger moléculas de DNA livres que estão circulando na corrente sanguínea, bem como moléculas de DNA que estão presentes em vesículas extracelulares (tais como exossomos) que estão circulando na corrente sanguínea.

[0051] Conforme usado no presente documento, o termo "DNA tumoral circulante" (ou "ctDNA") é DNA derivado de tumor que está circulando no sangue periférico de um paciente. O ctDNA é de origem tumoral e se origina diretamente do tumor ou de células tumorais circulantes (CTCs), que são células tumorais viáveis e intactas que se desprendem de tumores primários e entram na corrente sanguínea ou sistema linfático. O mecanismo preciso de liberação de ctDNA não é claro, embora seja postulado que envolva apoptose e necrose de

células mortas ou liberação ativa de células tumorais viáveis. O ctDNA pode ser altamente fragmentado e, em alguns casos, pode ter um tamanho médio de fragmento de cerca de 100 a 250 pb, por exemplo, 150 a 200 pb de comprimento. A quantidade de ctDNA em uma amostra de DNA livre de células circulante isolada de um paciente com câncer varia muito: as amostras típicas contêm menos de 10% de ctDNA, embora muitas amostras tenham menos de 1% de ctDNA e algumas amostras tenham mais de 10% de ctDNA. As moléculas de ctDNA podem ser frequentemente identificadas porque contêm mutações tumorigênicas.

[0052] Conforme usado no presente documento, os termos "RNA livre de células da corrente sanguínea", "RNA livre de células circulante" e "RNA livre de células" ("cfRNA") referem-se ao RNA que está circulando no sangue periférico de um paciente. O RNA livre de células pode conter RNA tumoral circulante (ctRNA), ou seja, RNA tumoral que circula livremente no sangue de um paciente com câncer, ou RNA fetal circulante (se o sujeito for uma mulher grávida). Esse termo se destina a abranger moléculas de RNA livres que estão circulando na corrente sanguínea, bem como moléculas de RNA que estão presentes em vesículas extracelulares (tais como exossomos) que estão circulando na corrente sanguínea.

[0053] Conforme usado no presente documento, o termo "variação de sequência" refere-se à combinação de uma posição e tipo de alteração de sequência. Por exemplo, uma variação de sequência pode ser referida pela posição da variação e pelo tipo de substituição (por exemplo, G para A, G para T, G para C, A para G, etc. ou inserção/deleção de um G, A, T ou C, etc.) que está presente na posição. Uma variação de sequência pode ser uma substituição, deleção, inserção ou rearranjo de um ou mais nucleotídeos. No contexto do presente método, uma variação de sequência pode ser gerada por,

por exemplo, um erro de PCR, um erro de sequenciamento ou uma variação genética.

[0054] Conforme usado no presente documento, o termo "variação genética" refere-se a uma variação (por exemplo, uma substituição de nucleotídeo, uma indel ou um rearranjo) que está presente ou considerada como tendo probabilidade de estar presente em uma amostra de ácido nucleico. Uma variação genética pode ser de qualquer fonte. Por exemplo, uma variação genética pode ser gerada por uma mutação (por exemplo, uma mutação somática), um transplante de órgão ou gravidez. Se a variação de sequência for denominada variação genética, a denominação indica que a amostra provavelmente contém a variação; em alguns casos, uma "denominação" pode estar incorreta. Em muitos casos, o termo "variação genética" pode ser substituído pelo termo "mutação". Por exemplo, se o método estiver sendo usado para detectar variações de sequência associadas ao câncer ou outras doenças causadas por mutações, então "variação genética" pode ser substituída pelo termo "mutação".

[0055] Conforme usado no presente documento, o termo "denominação" significa indicar se uma variação de sequência particular está presente em uma amostra. Isso pode envolver, por exemplo, fornecer uma sequência que contém a variação de sequência e/ou anotar uma sequência que tem a variação de sequência, indicando que a sequência tem uma variação de A para T em uma posição específica.

[0056] Conforme usado no presente documento, o termo "limiar" refere-se a um nível de evidência que é necessário para fazer uma denominação. Um limiar i. pode variar de uma variação de sequência para outra e ii. em alguns casos, pode ser aumentado ou diminuído independentemente de outros limiares, conforme desejado, dependendo de uma variedade de fatores.

[0057] Conforme usado no presente documento, o termo "corte"

refere-se a uma frequência de leituras de sequência igual ou superior à qual uma réplica pode ser identificada como estatisticamente provável de conter uma variação de sequência com base nos controles. Conforme será explicado em mais detalhes abaixo, no sequenciamento de um produto de PCR que contém uma variação de sequência que está presente em uma minoria das moléculas, algumas das leituras de sequência serão das moléculas variantes, enquanto outras não (por exemplo, serão da sequência do "tipo selvagem"). A frequência de leituras provenientes das moléculas variantes pode ser calculada, por exemplo, dividindo o número de leituras das moléculas variantes pelo número total de leituras. A frequência de corte pode ser estabelecida pelo sequenciamento de várias amostras de controle (por exemplo, amostras que não contêm a variação de sequência). Um corte i. pode variar de uma variação de sequência para outra e ii. em alguns casos, pode ser aumentado ou diminuído independentemente de outros cortes, conforme desejado, dependendo de uma variedade de fatores.

[0058] Conforme usado no presente documento, o termo "valor" refere-se a um número, letra, palavra (por exemplo, "alto", "médio" ou "baixo") ou descritor (por exemplo, "+++" ou "++") que pode indicar a força de evidência. Um valor pode conter um componente (por exemplo, um único número) ou mais de um componente, dependendo de como um valor é analisado.

[0059] Outras definições de termos podem aparecer ao longo do relatório descritivo. É observado ainda que as reivindicações podem ser elaboradas para excluir qualquer elemento opcional. Sendo assim, esta declaração se destina a servir como base antecedente para o uso de terminologia exclusiva como "somente", "apenas" e similares em conexão com a recitação de elementos de reivindicação ou o uso de uma limitação "negativa".

#### DESCRIÇÃO DETALHADA

[0060] Antes de a presente invenção ser descrita em mais detalhes, deve ser entendido que esta invenção não está limitada a modalidades particulares descritas, pois as mesmas podem, certamente, variar. Também deve ser entendido que a terminologia usada no presente documento serve para o propósito de descrever apenas modalidades particulares da invenção, e não se destina a ser limitante, visto que o escopo da presente invenção será limitado apenas pelas reivindicações anexas.

[0061] Quando uma faixa de valores é fornecida, entende-se que cada valor interveniente, até o décimo da unidade do limite inferior, a menos que o contexto indique claramente o contrário, entre o limite superior e inferior dessa faixa e qualquer outro valor declarado ou interveniente nessa faixa declarada, é englobado dentro da invenção.

[0062] A menos que definido de outro modo, todos os termos técnicos e científicos usados no presente documento têm o mesmo significado que o normalmente entendido por uma pessoa de habilidade comum técnica à qual esta invenção pertence. Embora quaisquer métodos e materiais similares ou equivalentes aos descritos neste documento também possam ser usados na prática ou no teste da presente invenção, os métodos e materiais preferenciais são agora descritos.

[0063] Todas as publicações e patentes citadas neste relatório descritivo são incorporadas ao presente documento a título de referência, como se cada publicação ou patente individual fosse especificamente e individualmente indicada para ser incorporada a título de referência e são incorporadas ao presente documento a título de referência para divulgar e descrever os métodos e/ou materiais em conexão com os quais as publicações são citadas. A citação de qualquer publicação é para sua divulgação antes da data de depósito e não deve ser interpretada como uma admissão de que a presente

invenção não tem o direito de anteceder tal publicação em virtude de uma invenção anterior. Além disso, as datas de publicação fornecidas podem ser diferentes das datas de publicação reais que podem precisar ser confirmadas independentemente.

[0064] Deve-se verificar que, conforme usado no presente documento e nas reivindicações anexas, as formas singulares "um", "uma" e "o/a" incluem referentes plurais, a menos que o contexto dite claramente o contrário. É observado ainda que as reivindicações podem ser elaboradas para excluir qualquer elemento opcional. Sendo assim, essa declaração se destina a servir como base antecedente para o uso de tal terminologia exclusiva como "somente", "apenas" e similares em conexão com a citação de elementos de reivindicação, ou uso de uma limitação "negativa".

[0065] Como será evidente para os versados na técnica ao ler esta divulgação, cada uma das modalidades individuais descritas e ilustradas no presente documento tem componentes e características distintas que podem ser prontamente separadas ou combinadas com as características de qualquer uma das outras várias modalidades sem se afastar do escopo ou da essência da presente invenção. Qualquer método recitado pode ser realizado na ordem dos eventos recitada ou em qualquer outra ordem que seja logicamente possível.

[0066] É fornecido no presente documento um método para análise de sequência que emprega vários pares de iniciadores que são compatíveis em uma reação de PCR multiplex. Neste contexto, uma reação de PCR multiplex que contém iniciadores "compatíveis" é aquela em que os pares de iniciadores são projetados para amplificar especificamente regiões de interesse que produzem amplicons que correspondem aos pares de iniciadores de PCR, minimizando a produção de dímeros de iniciadores, quando a reação é submetida a condições de termociclagem adequadas com um modelo apropriado

para os iniciadores. Normalmente, embora nem sempre, cada par de iniciadores amplifica uma única região de interesse em uma reação de PCR multiplex. As condições para realizar PCR multiplex e programas para projetar iniciadores compatíveis são bem-conhecidas (consultar, por exemplo, Sint et al, *Methods Ecol Evol.* 2012 3: 898 a 890 e Shen et al *BMC Bioinformatics* 2010 11: 143). Os pares de iniciadores compatíveis podem ser projetados com o uso de qualquer um de uma série de programas diferentes especificamente projetados para projetar pares de iniciadores para métodos de PCR multiplex. Por exemplo, os pares de iniciadores podem ser projetados com o uso dos métodos de Yamada et al. (*Nucleic Acids Res.* 2006 34:W665-9), Lee et al. (*Appl. Bioinformatics* 2006 5: 99 a 109), Vallone et al. (*Biotechniques.* 2004 37: 226 a 231), Rachlin et al. *BMC Genomics.* 2005 6:102 ou Gorelenkov et al. (*Biotechniques.* 2001 31: 1.326 a 1.330), por exemplo. Em algumas modalidades, o método pode empregar pelo menos 5 pares de iniciadores compatíveis, por exemplo, pelo menos 10, pelo menos 50, pelo menos 100 ou pelo menos 1.000 pares de iniciadores compatíveis. Em algumas modalidades, o método pode empregar pelo menos 10 e até 50.000 pares de iniciadores, pelo menos 10 e até 10.000 pares de iniciadores, pelo menos 10 e até 5.000 pares de iniciadores, pelo menos 10 e até 1.000 pares de iniciadores ou pelo menos 10 e até 500 pares de iniciadores, em que cada par de iniciadores é projetado para amplificar um amplicon diferente. Os amplicons amplificados podem ter qualquer comprimento adequado e podem variar quanto ao comprimento. Em algumas modalidades, o comprimento de cada amplicon, independentemente, pode estar na faixa de 50 pb a 500 pb, embora amplicons mais longos ou mais curtos possam ser usados em algumas implementações.

[0067] Após os pares de iniciadores terem sido obtidos, o método pode compreender configurar pelo menos duas reações de PCR

multiplex (por exemplo, até 10 reações de PCR multiplex, tal como 2, 3, 4, 5, 6, 7, 8, 9 ou 10 reações de PCR multiplex), em que cada uma contém porções diferentes da mesma amostra (isto é, diferentes alíquotas da mesma amostra). Nesta etapa, as reações de PCR multiplex não são idênticas umas às outras, pois alguns pares de iniciadores podem estar em todas as reações, enquanto outros estão em apenas uma ou algumas das (mas não todas as) reações. Por exemplo, se houver três reações de PCR multiplex, então alguns pares de iniciadores podem estar em uma única reação, alguns pares de iniciadores podem estar em duas reações e alguns pares de iniciadores podem estar em todas as três reações. Da mesma forma, se houver quatro reações de PCR multiplex, então alguns pares de iniciadores podem estar em uma única reação, alguns pares de iniciadores podem estar em duas reações, alguns pares de iniciadores podem estar em três reações e alguns pares de iniciadores podem estar em todas as quatro reações. Nessas modalidades, pelo menos alguns dos pares de iniciadores estão em mais de uma reação de PCR e pelo menos uma das reações de PCR contém alguns, mas não todos os pares de iniciadores da(s) outra(s) reação (reações). Quantas reações de PCR contêm um par de iniciadores específico é determinado por uma variedade de fatores, incluindo, mas sem limitação, a: i. a probabilidade de encontrar uma variação genética no amplicon amplificado pelo par de iniciadores selecionado, ii. a probabilidade de encontrar uma variação genética que se correlaciona com um câncer particular de interesse no amplicon amplificado pelo par de iniciadores selecionado, iii. o histórico de tratamento do paciente do qual a amostra foi obtida, iv. a probabilidade de encontrar uma variação genética clinicamente significativa no amplicon amplificado pelo par de iniciadores selecionado, v. testes anteriores realizados no paciente do qual a amostra foi obtida, vi. o perfil de erro de uma variação genética que se

espera que seja encontrada no amplicon amplificado pelo par de iniciadores selecionado (em que o termo "perfil de erro" indica a frequência em que uma variação particular não é devido a uma variação genética), e/ou vii. o comprimento do amplicon amplificado pelo par de iniciadores selecionado, ou qualquer combinação dos mesmos.

[0068] Por exemplo, se a probabilidade de detectar uma variação genética no amplicon amplificado por um par de iniciadores selecionado for alta em relação aos amplicons amplificados por outros pares de iniciadores (conforme previsto por experimentos anteriores e em andamento), então, esse par de iniciadores poderá estar em mais reações (por exemplo, todas as reações). Por outro lado, se a probabilidade de detectar uma variação genética no amplicon amplificado por um par de iniciadores selecionado for baixa em relação aos amplicons amplificados por outros pares de iniciadores (conforme previsto por experimentos anteriores e em andamento), então, esse par de iniciadores poderá estar em menos reações (por exemplo, uma ou duas reações). Em outro exemplo, se a probabilidade de encontrar uma variação genética que se correlaciona com uma doença ou condições específicas (por exemplo, um câncer de interesse) for alta em um amplicon em comparação com outros amplicons, então, o par de iniciadores poderá estar em mais reações (por exemplo, todas as reações). Por exemplo, se houver mais interesse em testar as mutações associadas ao câncer de pulmão de células não pequenas, os pares de iniciadores que amplificam as sequências que contêm potencialmente essas mutações poderão estar em mais reações. Por outro lado, pares de iniciadores que amplificam fragmentos que potencialmente contêm variações genéticas que se correlacionam com doenças ou condições que não são de interesse para o pesquisador podem estar em menos reações (por exemplo, uma ou duas reações). Em outro exemplo, o histórico de tratamento do paciente do qual a amostra foi obtida pode

ser usado para determinar quantas reações contêm um par de iniciadores específico. Neste exemplo, pares de iniciadores que amplificam sequências que podem abrigar variações genéticas associadas à resistência ao tratamento podem estar em mais reações (por exemplo, todas as reações), enquanto pares de iniciadores que amplificam sequências que podem abrigar variações genéticas não associadas à resistência ao tratamento podem estar em menos reações, por exemplo, uma ou duas reações. Em outro exemplo, pares de iniciadores que amplificam sequências que podem abrigar variações genéticas clinicamente acionáveis (isto é, variações genéticas que se correlacionam com um tratamento bem-sucedido) podem estar em mais reações (por exemplo, todas as reações), enquanto os pares de iniciadores que amplificam sequências que não abrigam variações genéticas clinicamente acionáveis podem estar em menos reações (por exemplo, uma ou duas reações). Em outro exemplo, o número de reações que contêm um par de iniciadores específico pode ser determinado por testes anteriores realizados no paciente do qual a amostra foi obtida. Por exemplo, se o paciente já for conhecido por ter uma variação genética particular, um par de iniciadores que amplifica um amplicon que contém potencialmente essa variação genética pode estar em mais (por exemplo, todas as) reações e um par de iniciadores que não amplifica um amplicon que contém potencialmente a variação genética pode estar em menos (por exemplo, uma ou duas) reações. Em outro exemplo, o número de reações que contêm um par de iniciadores particular pode ser determinado pelo tipo de variação genética encontrada no amplicon amplificado pelo par de iniciadores. Há probabilidade de que determinados tipos de variações de sequência (por exemplo, indels e rearranjos) não tenham sido gerados por uma PCR e/ou erro de sequenciamento e, sendo assim, os pares de iniciadores que têm como alvo as indels podem estar em menos reações

(por exemplo, uma ou duas reações). Os pares de iniciadores que visam variações que têm um fundo mais alto (por exemplo, substituições de nucleotídeos) podem estar em mais reações (por exemplo, todas as reações). Em outro exemplo, pares de iniciadores que amplificam produtos mais longos podem estar em mais reações do que pares de iniciadores que amplificam produtos mais curtos porque, quando o DNA de interesse é fragmentado como é o caso do DNA livre de células, os pares de iniciadores que amplificam produtos mais longos falharão com mais frequência ao amplificar o DNA disponível do que os pares de iniciadores que amplificam produtos mais curtos.

[0069] Uma ilustração esquemática de quatro reações de PCR multiplex (R1, R2, R3 e R4) que foram configuradas de acordo com o princípio descrito acima é mostrada na Figura 1. Neste exemplo, o amplicon A1 tem uma alta probabilidade de conter uma variação genética em relação a outros amplicons e, sendo assim, o par de iniciadores de PCR que produz este amplicon está em todas as reações; o amplicon A2 tem uma baixa probabilidade de conter uma variação genética em relação a outros amplicons e, sendo assim, o par de iniciadores de PCR que produz esse amplicon está em duas reações; o amplicon A3 tem uma probabilidade maior de conter uma variação genética que está associada a um câncer de interesse específico, por exemplo, câncer de pulmão de células não pequenas, em relação a outros amplicons e, sendo assim, o par de iniciadores de PCR que produz este amplicon está em todas as reações; o amplicon A4 tem uma probabilidade menor de conter uma variação genética associada a um câncer de interesse específico em relação a outros amplicons e, sendo assim, o par de iniciadores de PCR que produz este amplicon está em dois amplicons; o amplicon A5 tem maior probabilidade de conter uma variação genética clinicamente acionável e, sendo assim, o par de iniciadores de PCR que produz esse amplicon está em todas as

reações; o amplicon A6 tem menor probabilidade de conter uma variação genética clinicamente acionável e, sendo assim, o par de iniciadores de PCR que produz esse amplicon está em apenas três reações; o amplicon A7 tem maior probabilidade de conter uma variação genética de fundo alto e, sendo assim, o par de iniciadores de PCR que produz este amplicon está em todas as reações; e o amplicon A8 tem maior probabilidade de conter uma variação genética de fundo baixo (por exemplo, uma indel ou uma translocação) e, sendo assim, o par de iniciadores de PCR que produz esse amplicon está em duas reações. Em algumas modalidades, os pares de iniciadores de PCR que estão em menos reações podem ser espalhados entre as reações de modo que cada uma das reações de PCR multiplex contenha aproximadamente o mesmo número de pares de iniciadores.

[0070] Em algumas modalidades, os pares de iniciadores de PCR que produzem amplicons que têm uma probabilidade maior de conter uma variação genética podem estar em mais reações do que pares de iniciadores de PCR que produzem amplicons que têm uma probabilidade menor de conter uma variação genética; pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter uma variação genética associada a um câncer de interesse específico podem estar em mais reações do que pares de iniciadores de PCR que produzem amplicons com menor probabilidade de conter uma variação genética que está associada ao câncer de interesse específico; pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter uma variação genética que torna um paciente resistente a uma terapia podem estar em mais reações do que pares de iniciadores de PCR que produzem amplicons com menor probabilidade de conter uma variação genética que torna um paciente resistente à terapia; pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter variações genéticas

cl clinicamente acionáveis podem estar em mais reações do que pares de iniciadores de PCR que produzem amplicons com menor probabilidade de conter variações genéticas clinicamente acionáveis; e/ou pares de iniciadores de PCR que produzem amplicons que têm uma probabilidade maior de conter uma variação genética de fundo alto podem estar em mais reações do que pares de iniciadores de PCR que produzem amplicons que têm uma probabilidade maior de conter uma variação genética de fundo baixo.

[0071] Após as reações terem sido configuradas, o método compreende colocar as reações de PCR multiplex sob condições adequadas para amplificação (por exemplo, termociclagem) para produzir múltiplos amplicons replicados, em que os amplicons "replicados" são amplicons que são amplificados pelos mesmos iniciadores em duas ou mais reações. Os amplicons replicados geralmente têm a mesma sequência (exceto por erros de PCR, variações correspondentes a variações genéticas na amostra e quaisquer variações nos iniciadores de PCR). Ilustrados por exemplo, todos os amplicons mostrados na Figura 1 têm réplicas: amplicon A1 tem quatro réplicas, amplicon A2 tem duas réplicas e amplicon A3 tem quatro réplicas, etc. Os amplicons são, então, sequenciados para produzir leituras de sequência.

[0072] No sequenciamento dos amplicons, os amplicons derivados de cada reação de PCR multiplex diferente podem ser sequenciados separadamente um do outro ou os amplicons podem ser codificados em barras com um identificador de réplica e, em seguida, agrupados antes do sequenciamento. Em algumas modalidades, os iniciadores nas reações de PCR multiplex podem ter uma cauda 5' que contém o identificador de réplica de modo que, após as reações de PCR terem sido concluídas, a sequência da cauda 5' dos iniciadores está presente nos amplicons. Em outras modalidades, as reações de PCR multiplex

podem ser realizadas sem o uso de iniciadores que têm uma cauda 5' que contém um identificador de réplica. Nessas modalidades, os produtos de PCR podem ser codificados em barras com um identificador de réplica em uma segunda rodada de amplificação que usa iniciadores de PCR que têm uma cauda 5' que contém um identificador de réplica. De qualquer forma, os amplicons podem ser amplificados antes do sequenciamento, utilizando iniciadores que têm uma cauda 5' que fornece compatibilidade com uma plataforma de sequenciamento específica. Em determinadas modalidades, além de um identificador de réplica, um ou mais dos iniciadores usados nesta etapa podem conter adicionalmente um identificador de amostra. Se os iniciadores tiverem um identificador de amostra, os produtos derivados de diferentes amostras podem ser agrupados antes do sequenciamento. Em algumas modalidades, os iniciadores de alvo específico contêm de 5' a 3' uma sequência de "marcação" universal, uma sequência de código de barras de réplica opcional seguida por uma sequência projetada para o alvo de interesse. Os iniciadores usados para amplificar ainda mais o multiplex inicial contêm de 5' a 3' uma cauda que fornece compatibilidade com uma plataforma de sequenciamento específica, um código de barras de amostra e, opcionalmente, um código de barras de réplica e uma sequência que pode se ligar a qualquer parte ou todo o complemento inverso da sequência de marcação presente nos iniciadores de alvo específico. Normalmente, os iniciadores diretos e inversos terão sequências de marcação diferentes.

[0073] Os iniciadores usados para a etapa de amplificação podem ser compatíveis com o uso em qualquer plataforma de sequenciamento de próxima geração em que a extensão do iniciador é usada, por exemplo, método de terminador reversível da Illumina, método de pirosequenciamento da Roche (454), sequenciamento por ligação da Life Technologies (a plataforma SOLiD), Plataforma Ion Torrent da Life

Technologies ou método de clivagem por base fluorescente da Pacific Biosciences. Exemplos de tais métodos são descritos nas seguintes referências: Margulies et al (Nature 2005 437: 376 a 380); Ronaghi et al (Analytical Biochemistry 1996 242: 84 a 89); Shendure (Science 2005 309: 1.728); Imelfort et al (Brief Bioinform. 2009 10:609 a 618); Fox et al (Methods Mol Biol. 2009;553:79 a 108); Appleby et al (Methods Mol Biol. 2009;513:19 a 39) English (PLoS One. 2012 7: e47768) e Morozova (Genomics. 2008 92: 255 a 264), que são incorporados a título de referência para as descrições gerais dos métodos e das etapas específicas dos métodos, incluindo todos os produtos de partida, reagentes e produtos finais para cada uma das etapas.

[0074] A etapa de sequenciamento pode ser realizada com o uso de qualquer método de sequenciamento de próxima geração conveniente e pode resultar em pelo menos 10.000, pelo menos 50.000, pelo menos 100.000, pelo menos 500.000, pelo menos 1 M, pelo menos 10 M, pelo menos 100 M, pelo menos 1 B ou pelo menos 10 B de leituras de sequência. Em alguns casos, as leituras podem ser leituras emparelhadas.

[0075] As leituras de sequência são, então, processadas computacionalmente. As etapas de processamento inicial podem incluir a identificação de códigos de barras (incluindo identificadores de amostra ou sequências de identificadores de réplica) e corte de leituras para remover sequências de adaptação ou de baixa qualidade. Além disso, as métricas de verificação de qualidade podem ser executadas para garantir que o conjunto de dados tenha uma qualidade aceitável.

[0076] Depois que as leituras de sequência passam pelo processamento inicial, as mesmas são analisadas para identificar variações genéticas. Denominar variações genéticas no DNA livre de células pode ser desafiador, pois as sequências variantes geralmente estão em minoria (por exemplo, menos de 10% da sequência). Sendo

assim, se uma estratégia de sequenciação de amplicon for empregada, as sequências para cada amplicon podem ser principalmente sequências do tipo selvagem. Variantes minoritárias, que podem ser representadas por menos de 10% das sequências, são difíceis de distinguir de artefatos, por exemplo, sequenciamento e/ou erros de PCR. No presente método, os amplicons são analisados de modo a produzir uma pontuação que, para cada variação de sequência, indica se a variação de sequência tem probabilidade de representar uma variação genética (por exemplo, uma mutação no DNA da amostra), em oposição a um artefato de sequenciamento ou erro de PCR. Nessas modalidades, o método pode incluir analisar as leituras de sequência de amplicons replicados para uma variação de sequência selecionada para produzir uma pontuação para a variação de sequência selecionada. Nessas modalidades, a pontuação pode ser baseada no número de amplicons replicados que compreendem uma variação de sequência que tem uma frequência acima de um corte ou pode indicar a força da evidência combinada para a variação de sequência entre as réplicas. Uma variação de sequência pode ser denominada variação genética com base na pontuação. Em algumas modalidades, a variação genética pode ser denominada comparando-se a pontuação com um limiar. A variação genética pode ser denominada se a pontuação estiver no limiar ou acima.

[0077] Em modalidades em que a pontuação é baseada no número de amplicons replicados que compreendem uma variação de sequência que tem uma frequência acima de um corte, o corte pode ser baseado em uma distribuição de erro que indica com que frequência uma variação de sequência é gerada por um erro de amplificação e/ou sequenciamento. Essa distribuição de erro pode ser estabelecida com o uso de amostras de controle que podem ter variações genéticas ou não. Em algumas modalidades, um corte pode ser determinado com o

uso de um modelo de distribuição de probabilidade binomial, binomial superdisperso, beta, normal, exponencial ou gama com base no sequenciamento de amostras de controle. Em algumas modalidades, uma distribuição de erros pode mostrar com que frequência os erros de amplificação e/ou sequenciamento ocorrem em diferentes profundidades de sequenciamento. Um exemplo de tal distribuição de erro é mostrado na Figura 2. No exemplo mostrado na Figura 2, a frequência de uma variação de sequência em cada posição em um amplicon (isto é, o número de leituras de sequência que contêm uma variação de sequência em uma posição em relação ao número total de leituras de sequência para essa posição) pode ser plotada contra a profundidade de sequenciamento (isto é, número total de leituras de sequência) para um número de amostras de controle a fim de estabelecer o nível de fundo de variação de sequência para cada posição (cujo fundo é presumivelmente devido a artefatos de sequenciamento, ao invés de uma variação genética). Neste exemplo, o "corte" estabelece uma linha de base para identificar variações que são estatisticamente improváveis de serem de fundo. Nessas modalidades, o número de amplicons replicados que compreendem uma variação de sequência que tem uma frequência que está acima de um corte fornece uma pontuação que pode ser usada para determinar se uma variação é uma variação genética. Por exemplo, no exemplo mostrado na Figura 2, a frequência da variante está acima do corte em três das quatro réplicas. Neste exemplo, a pontuação pode ser "3 de 4", 0,75 ou simplesmente "3", indicando que a variação foi identificada positivamente em três réplicas. Essa pontuação é, então, comparada a um limiar para determinar se a variação é provavelmente o resultado de uma variação genética. Esse limiar pode variar de posição para posição e não precisa ser o mesmo para todas as variações genéticas potenciais. Por exemplo, no exemplo mostrado na Figura 2, o limiar

pode ser, por exemplo, 2 ou 3, em cujo caso a variante cujos dados são mostrados na Figura 2 é provavelmente devido a uma variação genética devido ao fato de o número de réplicas em que a variação é encontrada ser igual ou superior ao corte. Se o limiar for 4 neste exemplo, a variação não poderá ser denominada variação genética porque a pontuação está abaixo do limiar. Conforme será observado, um limiar pode ser aumentado ou diminuído dependendo de quantas réplicas de um amplicon são sequenciadas e uma série de outros fatores. O corte também pode ser aumentado ou diminuído com base em vários fatores. Em algumas modalidades, esse método pode compreender (i) para cada posição de nucleotídeo de um amplicon particular, a determinação, por exemplo, realização de plotagem, de uma distribuição de erro que mostra com que frequência os erros de amplificação e/ou sequenciamento ocorrem em diferentes profundidades de sequenciamento; (ii) com base na distribuição para cada posição da sequência, a determinação de um corte para cada profundidade de sequenciamento diferente em ou acima da qual uma variação genética pode ser detectada; (iii) o sequenciamento de múltiplos amplicons replicados da mesma amostra para obter uma pluralidade de leituras para os amplicons replicados; e (iv) a determinação, para cada posição de um amplicon, de se a frequência de uma variação de sequência nas leituras de sequência está acima ou abaixo do corte. O número de amplicons no corte ou acima do mesmo fornece a pontuação. Nessas modalidades, o termo "plotagem" pode ser realizada computacionalmente e, sendo assim, o método pode ser realizado sem desenhar fisicamente um gráfico.

[0078] Em modalidades em que a pontuação indica a força da evidência combinada para a variação de sequência entre as réplicas, os dados podem ser submetidos a procedimentos estatísticos, tanto de frequência quanto bayesianos, e a evidência para a variação pode ser

resumida como um valor de probabilidade ou, alternativamente um fator de Bayes ou uma probabilidade posterior no contexto de uma análise Bayesiana. Nessas modalidades, essa pontuação estatística pode ser alterada por outros dados à medida que se acumulam. Por exemplo, a evidência combinada para uma variação de sequência (cuja evidência pode incluir, por exemplo, o número de réplicas em que as leituras de sequência com a variação foram identificadas e, para cada amplicon: i. o número de leituras de sequência com a variação, ii. o número total de leituras de sequência para o amplicon, iii. a frequência da variação genética nas leituras de sequência e, iv. outras métricas) pode ser resumida como uma pontuação (por exemplo, um valor P ou similares), e a pontuação pode ser comparada a um limiar para determinar se a variação pode ser denominada variação genética. Por exemplo, se a pontuação que resume a evidência combinada é 0,91 e o limiar de probabilidade para denominar uma variação genética é 0,95, então, a variação genética não pode ser denominada. Por outro lado, se a pontuação que resume a evidência combinada for 0,98 e o limiar de probabilidade para denominar uma variação genética for 0,95, então, a variação genética deverá ser denominada. Esses métodos de análise, bem como o limiar, podem ser realizados por aprendizagem por máquina, se desejado.

[0079] Independentemente da forma em que a etapa de análise de sequência é implementada, o limiar ou corte usado pode, por si só, ser aumentado ou diminuído para cada variação conforme os dados se acumulam e/ou outros fatores. Por exemplo, o próprio limiar e/ou corte pode ser aumentado ou diminuído usando fatores similares aos descritos acima. Por exemplo, o limiar e/ou corte pode ser aumentado ou diminuído com base na frequência esperada de uma variação genética específica em pacientes com câncer (caso em que o limiar e/ou corte pode ser menor para mutações mais comuns), no tipo de câncer

do paciente a partir do qual a amostra foi obtida (caso em que o limiar e/ou corte pode ser inferior para mutações associadas a um câncer de interesse, tal como câncer de pulmão de células não pequenas), no histórico de tratamento do paciente do qual a amostra foi obtida (caso em que o limiar e/ou corte pode ser inferior para variações genéticas associadas à resistência a um tratamento), na relevância clínica das variações genéticas (neste caso, o limiar e/ou corte pode ser menor para variações genéticas associadas a um tratamento para um câncer), em testes anteriores realizados no paciente (neste caso, o limiar e/ou corte pode ser menor para variações genéticas que já foram identificadas no paciente), no perfil de erro de uma variação (em cujo caso o limiar e/ou corte pode ser menor para variações genéticas com menor taxa de erro), em outras variações genéticas que são encontradas na amostra (neste caso, o limiar e/ou corte pode ser menor para variações genéticas que não são comumente encontradas juntas em uma amostra) e/ou na taxa de erro geral do sequenciamento.

[0080] Em algumas modalidades, a amostra pode ser cfDNA e o método pode compreender ainda o sequenciamento de pelo menos algumas das mesmas regiões amplificadas com o uso de cfRNA do mesmo sujeito (através de RT-PCR). Isso pode ser realizado com o uso dos mesmos amplicons ou diferentes amplicons. Nesta implementação, o método pode envolver a comparação das variações genéticas denominadas com o uso de cfDNA com as variações genéticas denominadas com o uso de cfRNA. Se uma variação for identificada em ambas as amostras, a mesma poderá ser identificada como sendo uma variação genética com uma confiança mais alta.

[0081] Em algumas modalidades, a amostra pode ser cfDNA e o método pode compreender ainda o sequenciamento de pelo menos alguns dos mesmos amplicons amplificados do DNA de glóbulos vermelhos do mesmo sujeito. Nessas modalidades, o método pode

envolver a comparação das variações genéticas denominadas com o uso de cfDNA com as variações genéticas denominadas usando o DNA de glóbulos vermelhos. Se uma variação for identificada em ambas as amostras, a mesma poderá ser identificada como sendo uma variação genética com uma confiança inferior ou não toda. Esta modalidade fornece uma maneira para identificar variações que podem ser potencialmente devido à hematopoiese clonal de potencial indeterminado (CHIP) (consultar, de modo geral, Funari et al, Blood 2016 128:3.176 e Heuser et al, Dtsch Arztebl Int. 2016 113: 317 a 322), ou podem ser variantes de linhagem germinativa, por exemplo.

[0082] Em modalidades alternativas, o método pode ser realizado aumentando-se ou diminuindo-se o limiar e/ou corte para uma variação de sequência particular, sem variar o número de reações de PCR de réplica que amplificam a variação. Essas modalidades podem compreender: (a) obter múltiplos pares de iniciadores que são compatíveis em uma reação de PCR multiplex; (b) configurar pelo menos duas reações de PCR multiplex, em que cada uma contém porções diferentes da mesma amostra, em que as diferentes reações contêm os mesmos iniciadores; (c) termociclar as reações de PCR multiplex para produzir múltiplos amplicons replicados; (d) sequenciar os amplicons para produzir leituras de sequência; (e) analisar as leituras de sequência de amplicons replicados para uma variação de sequência selecionada para produzir uma pontuação para a variação de sequência selecionada, em que a pontuação: i. é baseada no número de amplicons replicados que compreendem uma variação de sequência que tem uma frequência acima de um corte; ou ii. indica a força da evidência combinada para a variação de sequência entre as réplicas; e (f) denominar a variação de sequência como uma variação genética com base na pontuação, em que a pontuação e/ou o corte usado para cada variação de sequência selecionada é baseado em parte: i. na frequência

esperada da variação genética, ii. no tipo de câncer do paciente do qual a amostra foi obtida, iii. no histórico de tratamento do paciente do qual a amostra foi obtida, iv. na relevância clínica da variação genética, v. em testes anteriores realizados no paciente de onde foi obtida a amostra, vi. no perfil de erro da variação genética, vi. em outras variações genéticas encontradas na amostra e/ou vii. na taxa de erro geral do sequenciamento, ou qualquer combinação dos mesmos. Detalhes de como esse método alternativo pode ser realizado podem ser adaptados a partir de outras partes desta divulgação.

[0083] Em algumas modalidades, o método pode compreender o fornecimento de um relatório que indica se há variações genéticas na amostra, no tipo de variação genética e/ou uma substituição de aminoácido causada pela variação genética. Em algumas modalidades, um relatório pode listar adicionalmente terapias aprovadas (por exemplo, aprovadas pela FDA) para cânceres que estão associados com a variação genética identificada na amostra. Essas informações podem ajudar no diagnóstico de uma doença (por exemplo, se o paciente tem câncer) e/ou nas decisões de tratamento feitas por um médico.

[0084] Em algumas modalidades, o relatório pode estar em um formato eletrônico e o método compreende o envio do relatório a um local remoto, por exemplo, para um médico ou outro profissional da saúde para ajudar a identificar um curso de ação adequado, por exemplo, para diagnosticar um sujeito ou para identificar uma terapia adequada para o sujeito. O relatório pode ser usado junto com outras métricas para determinar se o sujeito é suscetível a uma terapia, por exemplo.

[0085] Em qualquer modalidade, um relatório pode ser encaminhado para um "local remoto", em que "local remoto" significa um local diferente do local em que as sequências são analisadas. Por

exemplo, um local remoto pode ser outro local (por exemplo, escritório, laboratório, etc.) na mesma cidade, outro local em uma cidade diferente, outro local em um estado diferente, outro local em um país diferente, etc. Sendo assim, quando um item é indicado como sendo "remoto" a outro, o que se quer dizer é que os dois itens podem estar na mesma sala, mas separados, ou pelo menos em salas diferentes ou edifícios diferentes, e podem estar no mínimo uma milha, dez milhas, ou pelo menos cem milhas de distância. "Comunicação" de informação refere-se à transmissão dos dados que representam essas informações como sinais elétricos por meio de um canal de comunicação adequado (por exemplo, uma rede privada ou pública). "Encaminhamento" de um item refere-se a qualquer meio para levar esse item de um local para o próximo, seja transportando fisicamente esse item ou de outra forma (onde for possível) e inclui, pelo menos no caso de dados, o transporte físico de um meio que transporta os dados ou a comunicação dos dados. Exemplos de meios de comunicação incluem canais de transmissão de rádio ou infravermelho, bem como uma conexão de rede a outro computador ou dispositivo em rede e a internet, incluindo transmissão de e-mail e informações gravadas em sites e similares. Em determinadas modalidades, o relatório pode ser analisado por um MD ou outro profissional da saúde qualificado e um relatório com base nos resultados da análise das sequências pode ser encaminhado ao paciente do qual a amostra foi obtida.

[0086] Em algumas modalidades, uma amostra biológica pode ser obtida de um paciente e a amostra pode ser analisada usando o método. Em modalidades particulares, o método pode ser empregado para identificar e/ou estimar a quantidade de cópias variantes de um locus genômico que estão em uma amostra biológica que contém cópias do tipo selvagem de um locus genômico e cópias variantes do locus genômico, em que as cópias variantes têm uma variação de sequência

em relação às cópias do tipo selvagem do locus genômico. Neste exemplo, a amostra pode conter pelo menos 2 vezes, (por exemplo, pelo menos 5 vezes, pelo menos 10 vezes, pelo menos 50 vezes, pelo menos 100 vezes, pelo menos 500 vezes, pelo menos 1.000 vezes, pelo menos 5.000 vezes ou pelo menos 10.000) mais cópias do tipo selvagem do locus genômico do que cópias variantes do locus genômico.

[0087] Em algumas modalidades, o método não envolve o sequenciamento shotgun de uma amostra não enriquecida/não amplificada ou o sequenciamento de todo o exoma. Em vez disso, o sequenciamento pode ser feito como parte de um esforço de sequenciamento maior que visa pelo menos parte das sequências de codificação para até 200, por exemplo, até 100 ou até 50 genes, com foco nas sequências de codificação de AKT1, ALK, BRAF, CCND1, CDKN2A, CTNNB1, EGFR, ERBB2, ESR1, FGFR1, FGFR2, FGFR3, GATA3, GNA11, GNAQ, GNAS, HRAS, IDH1, IDH2, KIT, KRAS, MAP2K1, MET, MYC, NFE2L2, NRAS, NTRK1, NTRK3, PDGFRA, PIK3CA, PPP2R1A, PTEN, ROS1, STK11, TP53 e U2AF1, bem como nas sequências de codificação de outros genes, mutações ou que estão associadas ao câncer de pulmão de células não pequenas. Em modalidades alternativas, o método pode ser empregado para a detecção de mutações oncogênicas em, por exemplo, PIK3CA, NRAS, KRAS, JAK2, HRAS, FGFR3, FGFR1, EGFR, CDK4, BRAF, RET, PDGFRA, KIT ou ERBB2, que podem estar associados com câncer de mama, melanoma, câncer renal, câncer endometrial, câncer de ovário, câncer pancreático, leucemia, câncer colorretal, câncer de próstata, mesotelioma, glioma, meduloblastoma, policitemia, linfoma, sarcoma ou mieloma múltiplo (consultar, por exemplo, Chial 2008 Proto-oncogenes to oncogenes to cancer. Nature Education 1:1).

[0088] Em algumas modalidades, uma amostra pode ser coletada

de um paciente em um primeiro local, por exemplo, em um ambiente clínico, tal como em um hospital ou em um consultório médico, e a amostra pode ser encaminhada para um segundo local, por exemplo, um laboratório, onde a mesma é processada e o método descrito acima é realizado para gerar um relatório. Um "relatório", conforme descrito neste documento, é um documento eletrônico ou tangível que inclui elementos de relatório que fornecem resultados de teste que podem indicar a presença e/ou quantidade de variante (ou variantes) minoritária na amostra. Depois de gerado, o relatório pode ser encaminhado para outro local (que pode ser o mesmo local que o primeiro local), onde pode ser interpretado por um profissional de saúde (por exemplo, um clínico, um técnico de laboratório ou um médico, tal como um oncologista, cirurgião, patologista ou virologista), como parte de uma decisão clínica.

[0089] As variações genéticas identificadas por este método podem ser diagnósticas, prognósticas ou teranósticas.

[0090] Em algumas modalidades, o método pode ser usado para orientar as decisões de tratamento. Nessas modalidades, o método pode ser um método de tratamento que compreende realizar ou ter realizado o método descrito acima e administrar um tratamento ao paciente se um tratamento acionável for identificado. Mutações acionáveis incluem, mas sem limitação, mutações de ativação em EGFR e BRAF, tais como: G719X, deleções de exon19, V765A, T783A, V774A, S784P, L858R, S768I e L861X em EGFR e V600E; L601G; K601E; L597V/Q/R e G469V/S/R/E/A em BRAF. Mutações acionáveis também incluem rearranjos em ALK e ROS1, por exemplo, fusões de EML4-ALK, TFG-ALK, STRN-ALK, KIF5B-ALK, CD74-ROS1, SLC34A2-ROS1, SDC4-ROS1 e EZR-ROS1. Por exemplo, erlotinibe (Tarceva), afatinibe (Gilotrif), gefitinibe (Iressa) ou osimertinibe (Tagrisso) podem ser administrados a pacientes com mutação de ativação em EGFR, crizotinibe (Xalkori), ceritinibe (Zykadia), alectinibe

(Alecensa) ou brigatinibe (Alecensa) podem ser administrados a pacientes com uma fusão de ALK, crizotinibe (Xalkori), entrectinibe (RXDX-101), lorlatinibe (PF-06463922), crizotinibe (Xalkori), entrectinibe (RXDX-101), lorlatinibe (PF-06463922), ropotrectinibe (TPX-0005), DS-6051b, ceritinibe, ensartinibe ou cabozantinibe podem ser administrados a pacientes com uma fusão de ROS1, e dabrafenibe (Tafinlar) ou trametinibe (Mekinist) podem ser administrados a pacientes com uma mutação de ativação em BRAF. Muitas outras mutações acionáveis, incluindo mutações que podem ser usadas para orientar o tratamento de um paciente com um inibidor de ponto de controle imunológico, também são conhecidas.

[0091] Em outras modalidades, o método pode ser usado para monitorar um tratamento. Por exemplo, o método pode compreender a análise de uma amostra obtida em um primeiro ponto de tempo com o uso do método e da análise de uma amostra obtida em um segundo ponto de tempo pelo método, e a comparação dos resultados, isto é, a comparação de quais variações são denominadas nas amostras e das frequências alélicas das mesmas. O primeiro e o segundo pontos de tempo podem ser antes e depois de um tratamento, ou dois pontos de tempo após o tratamento. Por exemplo, comparando-se os resultados obtidos de um ponto de tempo para outro, o método pode ser usado para identificar novas variações (por exemplo, mutações) que apareceram durante o curso de um tratamento, ou para determinar se uma variação identificada anteriormente não está mais presente no sujeito durante o curso de um tratamento. O método pode ser usado para determinar se a frequência alélica de qualquer mutação alterou (aumentou ou diminuiu) durante o curso do tratamento. A resposta de um paciente à terapia pode ser monitorada pela detecção de uma alteração na frequência alélica de mutações ou na presença de mutações. Se múltiplas mutações estiverem presentes, a frequência

alélica e a alteração da frequência alélica podem ser determinadas combinando-se as diferentes mutações e réplicas igualmente ou, alternativamente, podem ser ponderadas, por exemplo, com base na provável clonalidade, relevância clínica, probabilidade de ser uma alteração somática dentro do câncer em oposição à linha germinativa ou CHIP e acionabilidade. Se for determinado que um paciente provavelmente está respondendo à terapia, o mesmo pode ser mantido nessa terapia, enquanto se for determinado que provavelmente não está respondendo, o mesmo pode ser alterado para uma terapia alternativa.

[0092] Este método também pode ser usado para determinar se um sujeito está livre de doença ou se uma doença é recorrente.

[0093] Em algumas modalidades, o método pode ser usado para a análise de doença residual mínima. Nessas modalidades, os pares de iniciadores usados no método podem ser projetados para amplificar sequências que contêm variações que foram previamente identificadas em um tumor de paciente por meio de sequenciamento de material de tumor, cfDNA em um ponto de tempo anterior ou sequenciamento de outra amostra adequada. O número de reações que contêm cada par de iniciadores pode ser variado dependendo de, por exemplo, se a variante é prevista para ser uma mutação acionadora, a confiança com a qual a variante foi identificada no câncer, se a variante é prevista para ser clonal ou subclonal no câncer, se a variante está localizada em uma base que é tipicamente ruidosa para sequenciar ou não, se a variante está em uma região do genoma que se espera ser mais ou menos fragmentada (por exemplo, cromatina aberta ou fechada), o nível de confiança de que a variante é uma alteração somática presente dentro do câncer, em vez de CHIP ou uma alteração da linhagem germinativa, se o tipo de variante é uma mutação pontual ou indel, e se uma indel é curta ou longa. Em algumas modalidades, o limiar para denominar cada variante pode ser aumentado ou diminuído com base em se a variante

é prevista para ser uma mutação acionadora, se a variante é prevista para ser clonal ou subclonal no câncer, se a variante está localizada em uma base que é tipicamente ruidosa para a sequência ou não, por exemplo. Em algumas modalidades, a evidência para todas as variantes específicas do paciente pode ser combinada para determinar se o paciente ainda tem doença residual ou pode estar livre de doença. A importância de cada variante pode ser ajustada, conforme descrito acima.

[0094] Conforme seria prontamente observado, muitas etapas do método, por exemplo, as etapas de processamento de sequência e a geração de um relatório indicando uma variação genética, podem ser implementadas em um computador. Sendo assim, em algumas modalidades, o método pode compreender a execução de um algoritmo que calcula a probabilidade de um paciente ter uma variação genética com base na análise das leituras de sequência e no resultado da probabilidade. Em algumas modalidades, este método pode compreender inserir as sequências em um computador e executar um algoritmo que pode calcular a probabilidade com o uso das medições de entrada.

[0095] Conforme seria evidente, as etapas computacionais descritas podem ser implementadas por computador e, sendo assim, as instruções para realizar as etapas podem ser estabelecidas como uma programação que pode ser gravada em um meio de armazenamento legível por computador físico adequado. As leituras de sequenciamento podem ser analisadas computacionalmente.

[0096] Qualquer modalidade do método descrito no presente documento pode ser adaptada à análise de DNA tratado com bissulfito. Por exemplo, o método pode ser adaptado para detectar variações epigenéticas por meio de sequenciamento de bissulfito em vez de variações genéticas. Em tal modalidade, o DNA tratado com bissulfito é

analisado em replicado. Os iniciadores de PCR seriam concebidos para amplificar uma gama de sítios de interesse que contêm CpG. O número de réplicas para cada amplicon que contém diferentes sítios de CpG pode ser priorizado com base em muitos critérios, tais como a frequência com que um determinado sítio de CpG deve ser hipermetilado ou hipometilado na amostra de interesse, a relevância de tal hipo ou hipermetilação e o nível de ruído esperado ao ler um determinado sítio de CpG. Novamente, como ocorre com a variante, os limites e cortes também podem ser ajustados para cada sítio de CpG com base em fatores como esses, a fim de denominar os sítios de CpG metilados ou não metilados e para determinar o grau de metilação do DNA.

## REIVINDICAÇÕES

1. Método para análise de sequências, caracterizado pelo fato de que compreende:

(a) obter múltiplos pares de iniciadores que são compatíveis em uma reação de PCR multiplex;

(b) configurar pelo menos duas reações de PCR multiplex, em que cada uma contém porções diferentes da mesma amostra, em que pelo menos alguns dos pares de iniciadores estão em mais de uma reação de PCR e pelo menos uma das reações de PCR contém alguns, mas não todos os pares de iniciadores da(s) outra(s) reação (reações);

(c) termociclar as reações de PCR multiplex para produzir vários amplicons replicados;

(d) sequenciar os amplicons para produzir leituras de sequência;

(e) analisar as leituras de sequência de amplicons replicados para uma variação de sequência selecionada para produzir uma pontuação para a variação de sequência selecionada, em que a pontuação:

i. é baseada no número de amplicons replicados que compreendem uma variação de sequência que tem uma frequência acima de um corte; ou

ii. indica a força da evidência combinada para a variação de sequência entre as réplicas;

(f) denominar a variação de sequência como uma variação genética baseada na pontuação.

2. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que, para pelo menos alguns dos pares de iniciadores, o número de reações que compreendem um par de iniciadores selecionado depende:

i. da frequência esperada de uma ou mais variações

genéticas encontradas no amplicon amplificado pelo par de iniciadores selecionado,

ii. do tipo de câncer do paciente do qual a amostra foi obtida,

iii. do histórico de tratamento do paciente do qual a amostra foi obtida,

iv. da relevância clínica das variações genéticas que se espera encontrar no amplicon amplificado pelo par de iniciadores selecionado,

v. de testes anteriores realizados no paciente do qual a amostra foi obtida,

vi. do perfil de erro de uma ou mais variações genéticas que se espera serem encontradas no amplicon amplificado pelo par de iniciadores selecionado e/ou

vii. do comprimento do amplicon amplificado pelo par de iniciadores selecionado;

ou qualquer combinação dos mesmos.

3. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que;

pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter uma variação genética estão em mais reações do que pares de iniciadores de PCR que produzem amplicons com menor probabilidade de conter uma variação genética;

pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter uma variação genética associada a um câncer específico de interesse estão em mais reações do que pares de iniciadores de PCR que produzem amplicons com menor probabilidade de conter uma variação genética associada ao câncer de interesse específico;

pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter uma variação genética que torna um

paciente resistente a uma terapia estão em mais reações do que pares de iniciadores de PCR que produzem amplicons com menor probabilidade de conter uma variação genética que torna um paciente resistente à terapia;

pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter variações genéticas clinicamente acionáveis estão em mais reações do que pares de iniciadores de PCR que produzem amplicons com menor probabilidade de conter variações genéticas clinicamente acionáveis;

pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter uma variação genética de fundo alto estão em mais reações do que pares de iniciadores de PCR que produzem amplicons com maior probabilidade de conter uma variação genética de fundo baixo; e/ou

pares de iniciadores de PCR que produzem amplicons mais longos estão em mais reações do que pares de iniciadores de PCR que produzem amplicons mais curtos.

4. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que, na etapa (f), a denominação é feita comparando-se a pontuação a um limiar em ou acima do qual uma variação genética pode ser denominada.

5. Método, de acordo com a reivindicação 4, caracterizado pelo fato de que o limiar é:

(i) o número de réplicas que têm a variação de sequência selecionada acima de uma frequência de corte; e/ou

(ii) um valor que indica a força necessária da evidência combinada para a variação de sequência em múltiplas réplicas.

6. Método, de acordo com qualquer uma das reivindicações 4 e 5, caracterizado pelo fato de que o corte é baseado em uma distribuição de erro que indica com que frequência uma variação de

sequência é gerada por um erro de amplificação e/ou sequenciamento.

7. Método, de acordo com a reivindicação 6, caracterizado pelo fato de que a distribuição de erro é estimada por meio de amostras de controle de sequenciamento.

8. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que compreende o aumento ou a diminuição do limiar ou do corte com base

i. na frequência esperada de uma ou mais variações genéticas encontradas no amplicon amplificado pelo par de iniciadores selecionado,

ii. no tipo de câncer do paciente do qual a amostra foi obtida,

iii. no histórico de tratamento do paciente do qual a amostra foi obtida,

iv. na relevância clínica das variações genéticas que se espera encontrar no amplicon amplificado pelo par de iniciadores selecionado,

v. em testes anteriores realizados no paciente do qual a amostra foi obtida,

vi. no perfil de erro de uma variação genética que se espera encontrar no amplicon amplificado pelo par de iniciadores selecionado,

vii. em outras variações genéticas encontradas na amostra, e/ou

viii. na taxa de erro geral do sequenciamento,

ou qualquer combinação dos mesmos.

9. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que cada reação estabelecida em (b) contém pelo menos 5 pares de iniciadores.

10. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a etapa (b) compreende a configuração de pelo menos três e menos de 10 reações de PCR

multiplex.

11. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a variação de sequência é uma substituição, inserção, deleção, rearranjo ou uma combinação de múltiplas variantes.

12. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a amostra é cfDNA.

13. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que os amplicons replicados são marcados com identificadores de réplica durante a amplificação, e o método compreende reunir as diferentes reações de amplificação antes do sequenciamento.

14. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que o comprimento de cada amplicon está, independentemente, na faixa de 50 pb a 500 pb.

15. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a evidência combinada para a variação de sequência é resumida com o uso de um valor de probabilidade e o limiar é um limiar de probabilidade.

16. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a evidência combinada para a variação de sequência é resumida com o uso de estatísticas Bayesianas e o limiar é um fator de Bayes que pode ser alterado por distribuições anteriores.

17. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que o limiar é estabelecido com o uso de aprendizado de máquina.

18. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a amostra é cfDNA e o método compreende ainda a análise de cfRNA do mesmo sujeito.

19. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a amostra é cfDNA e o método compreende ainda a análise de DNA de glóbulos brancos do mesmo sujeito.

20. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que a variação de sequência é indicativa de uma doença, condição ou tratamento específico.

21. Método, de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo fato de que compreende ainda (g) encaminhar um relatório que compreende informações sobre a variação de sequência a terceiros.

R1	R2	R3	R4	
—	—	—	—	A1
—		—		A2
—	—	—	—	A3
	—		—	A4
—	—	—	—	A5
—	—	—		A6
—	—	—	—	A7
	—	—		A8

**FIG. 1**

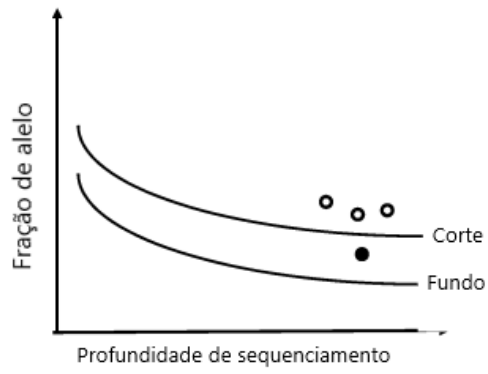


FIG. 2

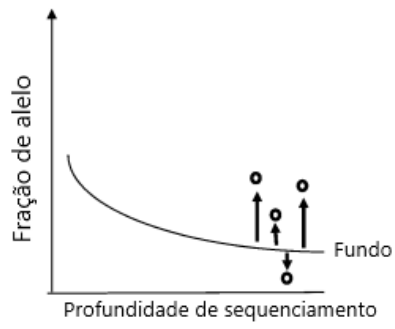


FIG. 3

## **RESUMO**

### **"MÉTODO DE SEQUENCIAMENTO QUE USA PCR MULTIPLEX DE REPLICAÇÃO VARIÁVEL".**

A presente invenção refere-se a um método para análise de sequência que compreende a análise de reações de PCR em que cada uma contém diferentes porções da mesma amostra, em que pelo menos alguns dos pares de iniciadores estão em mais de uma reação de PCR e pelo menos uma das reações de PCR contém alguns, mas não todos os pares de iniciadores da(s) outra(s) reação (reações).