

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6383789号  
(P6383789)

(45) 発行日 平成30年8月29日 (2018. 8. 29)

(24) 登録日 平成30年8月10日 (2018. 8. 10)

(51) Int. Cl.	F I
C 1 2 Q 1/686 (2018. 01)	C 1 2 Q 1/686 Z N A Z
C 1 2 Q 1/6869 (2018. 01)	C 1 2 Q 1/6869 Z
C 1 2 N 15/09 (2006. 01)	C 1 2 N 15/09 Z

請求項の数 31 (全 63 頁)

(21) 出願番号	特願2016-524301 (P2016-524301)	(73) 特許権者	516376592
(86) (22) 出願日	平成26年6月30日 (2014. 6. 30)		アダプティブ バイオテクノロジーズ コーポレーション
(65) 公表番号	特表2016-523547 (P2016-523547A)		アメリカ合衆国, ワシントン州 98102, シアトル, スイート 200, 1551 イーストレイク アベニュー イー.
(43) 公表日	平成28年8月12日 (2016. 8. 12)		
(86) 国際出願番号	PCT/US2014/044971	(74) 代理人	100114775
(87) 国際公開番号	W02015/002908		弁理士 高岡 亮一
(87) 国際公開日	平成27年1月8日 (2015. 1. 8)	(74) 代理人	100121511
審査請求日	平成29年6月8日 (2017. 6. 8)		弁理士 小田 直
(31) 優先権主張番号	61/841, 878	(74) 代理人	100202751
(32) 優先日	平成25年7月1日 (2013. 7. 1)		弁理士 岩堀 明代
(33) 優先権主張国	米国 (US)	(74) 代理人	100191086
(31) 優先権主張番号	62/001, 580		弁理士 高橋 香元
(32) 優先日	平成26年5月21日 (2014. 5. 21)		
(33) 優先権主張国	米国 (US)		
早期審査対象出願		最終頁に続く	

(54) 【発明の名称】 配列タグによる大規模生体分子解析

(57) 【特許請求の範囲】

【請求項 1】

複数の免疫細胞受容体鎖からクロノタイププロファイルを生成する方法であって、

(a) 反応混合物中、プライマー伸長条件下で、第1セットのプライマーとB細胞及び/又はT細胞由来の再構成された核酸サンプル及び/又は無細胞系DNAサンプルとを組み合わせる工程であって、前記第1セットのそれぞれのプライマーは受容体特異的な部分、第1のプライマー結合部位を含有する5' - 非相補的末端、及び前記受容体特異的な部分と前記第1のプライマー結合部位との間に配置された配列タグを含み、前記受容体特異的な部分は、第1の所定の位置にて異なる再構成された核酸にアニーリングし、伸長して第1の伸長産物を生成する、工程と；

(b) 前記第1セットのプライマーのうち伸長していないプライマーを前記反応混合物より除去する、工程と；

(c) プライマー伸長条件下で、前記反応混合物に第2セットのプライマーを添加する工程であって、前記第2セットのそれぞれのプライマーは、受容体特異的な部分と第2のプライマー結合部位を含む5' 非相補的末端とを有し、前記受容体特異的な部分は、第2の所定の位置にて前記第1の伸長産物にアニーリングし、前記第2セットのそれぞれのプライマーは、伸長して第2の伸長産物を生成し、それぞれの第2の伸長産物は、第1のプライマー結合部位、第2のプライマー結合部位、配列タグ、及びT細胞受容体鎖又はB細胞受容体鎖の部分を含み、工程と；

(d) 前記反応混合物においてポリメラーゼ連鎖反応を実施して、アンプリコンを生成

10

20

する工程であって、前記ポリメラーゼ連鎖反応は、前記第 1 のプライマー結合部位に特異的な順方向プライマーと、前記第 2 のプライマー結合部位に特異的な逆方向プライマーとを使用する、工程と；

( e ) 前記アンプリコンを配列決定して、複数の T 細胞受容体鎖及び / 又は B 細胞受容体鎖のクロノタイププロファイルを生成する工程と、を含む、複数の免疫細胞受容体鎖からクロノタイププロファイルを生成する方法。

【請求項 2】

前記複数の T 細胞受容体鎖が T C R 、 T C R 及び T C R を含み、前記第 1 セットのプライマー及び前記第 2 セットのプライマーが、 T C R 及び T C R の V D J 領域をコードしている再構成された核酸領域に隣接するプライマー及び T C R の V J 領域に隣接するプライマーを含む、請求項 1 に記載の方法。

10

【請求項 3】

前記配列決定する工程が；

それぞれのエラー率を有し、それぞれのタグ配列と、再構成された核酸配列とを含む配列リードを生成することと；

同一のタグ配列を有する配列リードをアライメントして、同一の配列タグを有する配列リードのグループを形成することであって、配列リードのグループのそれぞれが候補のクロノタイプである、形成することと；

異なるクロノタイプを決定することであって、比較する前記配列リードのグループ中の相対頻度またはリードカウントが予め定められた閾値を超え、比較する前記配列リードのグループ中の相違するヌクレオチドの数が予め定められた閾値を超えときに、それぞれの候補のクロノタイプが、他の候補のクロノタイプと比較して尤度が明らかに少なくとも 9 5 % である場合、前記配列リードを異なるクロノタイプに結合させると決定することと

20

、  
を含む、請求項 1 に記載の方法。

【請求項 4】

前記第 2 の伸長産物の生成後に、伸長していない第 2 セットのプライマーを除去する工程を更に含む、請求項 1 に記載の方法。

【請求項 5】

前記第 1 セットの前記プライマーのアニーリング及び伸長が、前記第 1 の伸長産物の融解後に繰り返される、請求項 1 に記載の方法。

30

【請求項 6】

前記第 2 セットの前記プライマーのアニーリング及び伸長が、前記第 2 の伸長産物の融解後に繰り返される、請求項 1 に記載の方法。

【請求項 7】

前記除去する工程が、前記反応混合物にエキソヌクレアーゼ活性を有する酵素を添加することを含む、請求項 4 に記載の方法。

【請求項 8】

前記複数の B 細胞受容体鎖が I g H 及び I g K を含み、前記第 1 セットの前記プライマー及び前記第 2 セットの前記プライマーが、 I g H の V D J 領域、 I g H の D J 領域、及び I g K の V J 領域をコードしている再構成された核酸領域に隣接するプライマーを含む、請求項 1 に記載の方法。

40

【請求項 9】

サンプル中の、複数の免疫受容体鎖をコードしている再構成された核酸のクロノタイププロファイルを生成する方法であって、

( a ) T 細胞及び / 又は B 細胞及び / 又は無細胞系 D N A を含む個体由来のサンプルの T 細胞受容体遺伝子又はイムノグロブリン遺伝子の再構成された核酸分子に配列タグを付加させてタグ - 核酸複合体を生成する工程であって、少なくとも 1 つのタグ - 核酸複合体又はそれらのコピーには異なる配列タグが付加されている、工程と；

( b ) 前記タグ - 核酸複合体を増幅させる工程と；

50

(c) 前記タグ - 核酸複合体を配列決定して、それぞれのエラー率を有し、それぞれの配列タグと、再構成された核酸配列とを含む配列リードを提供する工程と；

(d) 同一の配列タグを有する配列リードをアライメントして、同一の配列タグを有する配列リードのグループを形成する工程であって、配列リードのグループのそれぞれが候補のクロノタイプである、形成する工程と；

(f) サンプル中の異なるクロノタイプを決定することであって、相対頻度またはリードカウントが予め定められた閾値を超え、相違するヌクレオチドの数が予め定められた閾値を超えときに、それぞれの候補のクロノタイプが、他の候補のクロノタイプと比較して尤度が明らかに少なくとも95%である場合、前記配列リードを異なるクロノタイプに結合させると決定することにより前記サンプルのクロノタイププロファイルを生成する工程と、  
を含み、

前記付加させる工程が、

反応混合物中、プライマー伸長条件下で、第1セットのプライマーとサンプルとを組み合わせることとであって、前記第1セットのそれぞれのプライマーは受容体特異的な部分、第1のプライマー結合部位を含有する5' - 非相補的末端、及び前記受容体特異的な部分と前記第1のプライマー結合部位との間に配置された配列タグを含み、前記受容体特異的な部分は、第1の所定の位置にて異なる再構成された核酸にアニーリングし、伸長して第1の伸長産物を生成する、組み合わせることと；

前記第1セットのプライマーのうち伸長していないプライマーを前記反応混合物より除去することと；

プライマー伸長条件下で、前記反応混合物に第2のプライマーセットを添加することとであって、前記第2セットのそれぞれのプライマーは、受容体特異的な部分と第2のプライマー結合部位を含む5' 非相補的末端とを有し、前記受容体特異的な部分は、第2の所定の位置にて前記第1の伸長産物にアニーリングし、前記第2セットのそれぞれのプライマーは、伸長して第2の伸長産物を生成し、それぞれの第2の伸長産物は、第1のプライマー結合部位、第2のプライマー結合部位、配列タグ、及びT細胞受容体鎖又はB細胞受容体鎖の部分をコードしている再構成された核酸を含む、添加することと、を含む、  
サンプル中の、複数の免疫受容体鎖をコードしている再構成された核酸のクロノタイププロファイルを生成する方法。

【請求項10】

前記反応混合物から、伸長していない前記第2セットのプライマーを除去する工程を更に含む、請求項9に記載の方法。

【請求項11】

前記第1セットの前記プライマーのアニーリング及び伸長が、前記第1の伸長産物の融解後に繰り返される、請求項9に記載の方法。

【請求項12】

前記第2セットの前記プライマーのアニーリング及び伸長が、前記第2の伸長産物の融解後に繰り返される、請求項9に記載の方法。

【請求項13】

前記除去する工程が、前記反応混合物にエキソヌクレアーゼ活性を有する酵素を添加することを含む、請求項10に記載の方法。

【請求項14】

前記付加する工程及び前記増幅させる工程がタグ - 核酸複合体を生成し、少なくとも1つの再構成された核酸及びそのそれぞれのコピーには複数の異なる配列タグが付加されている、請求項9に記載の方法。

【請求項15】

前記再構成された核酸が、TCR、TCR及びTCR鎖をコードしている再構成された核酸を含み、前記第1セットのプライマー及び前記第2セットのプライマーが、TCR及びTCRのVDJ領域をコードしている再構成された核酸領域に隣接するブラ

10

20

30

40

50

イマー及びTCRのVJ領域に隣接するプライマーを含む、請求項14に記載の方法。

【請求項16】

前記再構成された核酸が、IGH及びIGKをコードしている再構成された核酸を含み、前記第1セットのプライマー及び前記第2セットのプライマーが、IGHのVDJ領域、IGHのDJ領域、及びIGKのVJ領域をコードしている再構成された核酸領域に隣接するプライマーを含む、請求項14に記載の方法。

【請求項17】

前記配列タグがモザイクタグであり、前記モザイクタグが、交互する定常領域及び可変領域を含む、請求項1に記載の方法。

【請求項18】

前記順方向プライマー及び/又は逆方向プライマーが、サンプルタグを含有する5'-非相補的末端を含む、請求項1に記載の方法。

【請求項19】

前記第1の所定の位置が、V領域である、請求項1に記載の方法。

【請求項20】

前記第1セットのそれぞれのプライマーの受容体特異的な部分が、V領域の重複していない異なる部分にアニーリングする、請求項19に記載の方法。

【請求項21】

前記第2の所定の位置が、J領域又はC領域である、請求項1に記載の方法。

【請求項22】

前記第2セットのそれぞれのプライマーの受容体特異的な部分が、J領域の重複していない異なる部分にアニーリングする、請求項21に記載の方法。

【請求項23】

前記第2セットのそれぞれのプライマーの受容体特異的な部分が、J領域の単一のプライマー結合部位にアニーリングする、請求項21に記載の方法。

【請求項24】

前記添加する工程が、前記反応混合物においてポリメラーゼ連鎖反応を実施して、アンプリコンを生成することを含み、前記ポリメラーゼ連鎖反応が、前記第1のプライマー結合部位に特異的な順方向プライマーと、前記第2のプライマー結合部位に特異的な逆方向プライマーとを使用する、請求項9に記載の方法。

【請求項25】

前記順方向プライマー及び/又は逆方向プライマーが、サンプルタグを含有する5'-非相補的末端を含む、請求項24に記載の方法。

【請求項26】

前記配列タグがモザイクタグであり、前記モザイクタグが、交互する定常領域及び可変領域を含む、請求項9に記載の方法。

【請求項27】

前記第1の所定の位置が、V領域である、請求項9に記載の方法。

【請求項28】

前記第1セットのそれぞれのプライマーの受容体特異的な部分が、V領域の重複していない異なる部分にアニーリングする、請求項27に記載の方法。

【請求項29】

前記第2の所定の位置が、J領域又はC領域である、請求項9に記載の方法。

【請求項30】

前記第2セットのそれぞれのプライマーの受容体特異的な部分が、J領域の重複していない異なる部分にアニーリングする、請求項29に記載の方法。

【請求項31】

前記第2セットのそれぞれのプライマーの受容体特異的な部分が、J領域の単一のプライマー結合部位にアニーリングする、請求項29に記載の方法。

【発明の詳細な説明】

10

20

30

40

50

## 【技術分野】

## 【0001】

(関連出願の相互参照)

本出願は、2013年7月1日に出願された米国仮出願第61/841,878号、及び2014年5月21日に出願された同第62/001,580号、に対する優先権を主張するものであり、これらの仮出願は、参照によりその全体が本願に援用される。

## 【0002】

(配列表)

本願はEFS-Webによって提出されたASCIIフォーマットの配列表を含み、配列表の全内容を参照により本願に援用する。かかるASCIIのコピーは、2014年6月27日に作成されたものであり、ファイル名は848US00-SL-ST25.txtであり、ファイルサイズは2キロバイトである。新規内容は追加されていない。

## 【背景技術】

## 【0003】

速度及び利便性が増し、塩基当たりのコストが低減されるに伴い、診断及び予測用途において、大規模DNA配列決定が急速に発展している[例えば、Ding et al, Nature, 481(7382):506~510(2012); Chiu et al, Brit. Med. J., 342:c7401(2011); 及びKu et al, Annals of Neurology, 71(1):5~14(2012)など]。特に、T細胞又はB細胞受容体又はそれらの構成分子などの免疫分子をコードしている核酸のプロファイルは、生命体の健康状態又は疾患状態について豊富な情報を有していることから、このプロファイルを様々な状態の診断又は予測の指標として利用することが提案されている[例えば、Faham及びWillis、米国特許第8,236,503号及び同第8,628,927号; Freeman et al, Genome Research, 19:1817~1824(2009); Han et al, J. Immunol., 182(1001):42.6(2009); Boyd et al, Sci. Transl. Med., 1(12):12ra23(2009); He et al, Oncotarget(March 8, 2011)]。

## 【0004】

例えば、多くの癌に関し治療を受けた患者はしばしば癌に関連する微小残存病変(MRD)を保有している。すなわち、臨床的な尺度では、患者の疾患が治療により完全に寛解しているとされる場合であっても、何らかの理由により、癌細胞の小集団が排除から逃れ、残存し得るのである。この残存している集団の種類及び大きさは、患者に継続される治療に関係する重要な予後因子である[例えば、Campana, Hematol. Oncol. Clin. North Am., 23(5):1083~1098(2009); Buccisano et al, Blood, 119(2):332~341(2012)]。その結果、フローサイトメトリー、in situハイブリダイゼーション、細胞遺伝学、及び核酸マーカーの増幅による手法など、この集団を評価する手法が複数開発されている[例えば、Buccisano et al, Current Opinion in Oncology, 21:582~588(2009); 並びにvan Dongen et al, Leukemia, 17(12):2257~2317(2003)など]。T細胞及び/又はB細胞の免疫受容体断片(すなわち、クロナタイプ)をコードしている再構成された核酸の増幅は、このようなクロナタイプが、典型的には、癌細胞に関連付けられる分子タグとして提供され得る特有の配列を有することから、白血病及びリンパ腫におけるMRDの評価に特に有用である。増幅が高度に多重化されており、開発が困難であることに一部起因し、測定は、通常、受容体の短鎖をコードしている核酸の増幅及び配列決定によりなされる。多重化の規模が大きくなるにつれ、ミスハイブリダイゼーションによる擬陽性増幅、プライマー二量体の形成、及び増幅速度が異なることによる配列の偏った提示などの確率が増大するなどいくつかの問題に直面するようになっている[例えば、Elmifro et al, Clinical Microbiolog

10

20

30

40

50

y Reviews, 13(4):559~570(2000)]。更に、配列解析、サンプルの追跡、又は夾雑物の検出などのいずれに関しても、標的配列の類似性及び増幅される配列への配列タグの組み込みが、上記の大規模増幅に関連する課題に悪影響を及ぼしている。これらの課題は、微小病変に相関する核酸配列を評価するのに必要とされる別個のアッセイ数を減らすのに非常に有効である、複数の免疫受容体鎖の大規模一反応増幅の開発により解決されている。

【0005】

これらの観点から、免疫受容体鎖のセットをコードしている癌遺伝子のエキソン又はクロノタイプなどといった選別された核酸を、1回の反応で評価する、より効率的な方法が利用可能になったことは非常に有利である。

【発明の概要】

【発明が解決しようとする課題】

【0006】

本発明は、単一反応での大規模増幅法に関し、特に、免疫受容体鎖をコードしている再構成された核酸などの標的ポリヌクレオチド集団のポリメラーゼ連鎖反応(PCR)後に、大規模DNA配列決定法を利用してそれらを同定する方法に関する。本発明は、前述の方法を応用した、癌の微小残存病変のモニタリングを包含する。本発明は数多くの実施及び用途が例示され、それらの一部を、以下、本明細書を通して要約する。

【課題を解決するための手段】

【0007】

いくつかの実施形態では、本発明は、免疫受容体分子など対象とする生体分子群をコードしている核酸のプロファイルを生成する方法を目的とする。一態様では、本発明の方法は、サンプル中の選別された核酸群に配列タグを付加して、タグ-核酸複合体を生成することと、タグ-核酸複合体を増幅させることと、増幅させたタグ-核酸複合体を配列決定して、核酸プロファイルを生成するタグ配列と核酸配列の両方をそれぞれが含む配列リードを提供することと、を含む。いくつかの実施形態では、配列タグの付加は、プライマーの伸長及びプライマーの除去に係する1工程以上の連続する工程により成すことができ、これらの工程後、得られた生成物を、順方向及び逆方向プライマーにより偏りなく更に増幅することもできる。

【0008】

いくつかの実施形態では、本発明は、異なるサンプルに由来する材料によるサンプルの汚染、例えば、キャリーオーバー汚染を、検出及び測定するための方法を目的とする。一実施形態では、微小残存病変をモニターされる個人において、汚染を検出するための、このような方法は：(a)個人から組織サンプルを得ることと；(b)癌遺伝子分子又は再構成された核酸に配列タグを付加してタグ-核酸複合体を生成することと(少なくとも1つの核酸又はそれらのコピーは、付加された異なる配列タグを有し、癌遺伝子分子は、個人の癌に特徴的なものである)；(c)タグ-核酸複合体を増幅させることと；(d)タグ-核酸複合体サンプルを配列決定して、エラー率を有し、タグ配列と、癌遺伝子配列又は再構成された核酸配列とを含む配列リードを提供することと；(e)タグ配列を、他の組織サンプルから別途決定されたタグ配列と比較することと；(f)1つ以上のタグ配列のアイデンティティの存在、非存在、及び/又はそれによる汚染の程度を、その他の組織サンプルから別途決定されたタグ配列により評価することと、を含む。

【0009】

別の態様では、本発明は、少なくとも2つのB細胞受容体鎖に基づきクロノタイププロファイルを生成するための上記の通りの方法を目的とし、この方法は、2つ以上のB細胞受容体鎖をコードしている標的核酸を1回の反応で増幅させることを含む。別の態様では、このような方法は、B細胞癌において微小残存病変をモニターするのに使用される。

【0010】

別の態様では、本発明は、少なくとも2つのT細胞受容体鎖に基づきクロノタイププロファイルを生成するための上記の通りの方法を目的とし、この方法は、2つ以上のT細胞

10

20

30

40

50

受容体鎖をコードしている標的核酸を1回の反応で増幅させることを含む。別の態様では、このような方法は、T細胞癌において微小残存病変をモニターするのに使用される。

【0011】

本発明の、これらの上記に特徴づけられる態様、並びに他の態様は、数多くの実施例及び用途により例示される。そのうちいくつかを図に示し、以下の特許請求の範囲において特徴づける。しかしながら、上記要約はそれぞれ本発明に記載される実施形態又はすべての実施を記載することを意図するものではない。

【図面の簡単な説明】

【0012】

添付の特許請求の範囲に本発明の新規特徴を詳細に示す。本発明の原理を利用する代表的な実施形態について示す、以下の詳細な説明、並びに添付の図面を参照することにより、本発明の特徴及び利点が良好に理解される。

【図1A】本発明の様々な実施形態を模式図で示す。

【図1B】本発明の様々な実施形態を模式図で示す。

【図1C】本発明の様々な実施形態を模式図で示す。

【図1D】所定の長さを有する再構成された核酸のテンプレートを生成する方法（配列タグを利用する場合又は利用しない場合）を示す。

【図2A】固有の配列タグをサンプル中の実質的にすべての標的配列に付加するための様々な方法を示す。

【図2B】固有の配列タグをサンプル中の実質的にすべての標的配列に付加するための様々な方法を示す。

【図2C】固有の配列タグをサンプル中の実質的にすべての標的配列に付加するための様々な方法を示す。

【図2D】固有の配列タグをサンプル中の実質的にすべての標的配列に付加するための様々な方法を示す。

【図2E】固有の配列タグをサンプル中の実質的にすべての標的配列に付加するための様々な方法を示す。

【図2F】固有の配列タグをサンプル中の実質的にすべての標的配列に付加するための様々な方法を示す。

【図2G】固有の配列タグをサンプル中の実質的にすべての標的配列に付加するための様々な方法を示す。

【図3A】IGH鎖をコードしている核酸配列からクロノタイププロファイルを生成するための本発明の態様の模式図を示す。

【図3B】IGH鎖をコードしている核酸配列からクロノタイププロファイルを生成するための本発明の態様の模式図を示す。

【図4A】配列リードからクロノタイプ配列を決定するための配列タグの使用を示す。

【図4B】複数の異なる配列タグが同じ標的ポリヌクレオチド又はそれらのコピーに付加される実施形態における、配列タグの使用を例示する。

【図5A】緊密に係るクロノタイプ間の配列スペース及び距離におけるクロノタイプの概念を示す。

【図5B】単に配列誤差により異なる（結合させるべき）クロノタイプから本当に異なるクロノタイプを識別するための方法の一実施形態を示すフローチャートである。

【図5C】関連するクロノタイプを結合させるべきか否か判定するための一実施形態において使用される、数多くの関数の形態を示す。

【図5D】配列リードを結合させる方法における配列ツリーの使用を示す。

【図5E】配列リードを結合させる方法における配列ツリーの使用を示す。

【発明を実施するための形態】

【0013】

本発明の実施には、特に記載のない限り、従来法、並びに当業界の範囲内のものである分子生物学（組み換え法など）、バイオインフォマティクス、細胞生物学、及び生化学に

10

20

30

40

50

についての記載を使用することができる。このような従来法としては、限定するものではないが、血液細胞のサンプル収集及び解析、並びに核酸配列決定及び解析などが挙げられる。好適な手法についての具体的な例は、以下の実施例を参照することにより得ることができる。しかしながら、その他の均等な従来法も勿論使用することができる。このような従来法及び記載は、標準的な実験室マニュアル、例えば、Genome Analysis : A Laboratory Manual Series (Vols. I ~ IV) ; PCR Primer : A Laboratory Manual ; and Molecular Cloning : 及び A Laboratory Manual (すべて、Cold Spring Harbor Laboratory Press から出版) などに見ることができる。

10

#### 【0014】

一態様では、本発明は、複数の免疫受容体鎖をコードしている核酸の大規模多重増幅後に、増幅産物又はアンプリコンに対しハイスループットな配列決定を行うことにより、このような免疫受容体鎖のクロノタイププロファイルを生成する方法を目的とする。いくつかの実施形態では、本発明は、一連の、プライマーを伸長させる工程、伸長されていない又は組み込まれていないプライマーを除去する工程、並びに増幅（例えば、PCRによる）又は更なるプライマー伸長のいずれかのために新しいプライマーを添加する工程、を包含させることにより、多重増幅に一般的な欠点を克服する。このような工程により、配列タグを使用することも可能になる。配列タグを使用しない場合、非特異的な又は偽陽性増幅につながる。別の態様では、配列タグは、臨床用途に関係する実施形態に利用され、特に、微小残存病変（MRD）の解析、例えば、癌のための処置を受ける患者のサンプルのMRD解析に利用される。配列リードに組み込まれた配列タグは、クロノタイプの決定に有効な手段を提供し、それと同時に、以前のアッセイから持ち込まれた、同じ患者のサンプルに由来する又は同じ実験室で検査した異なる患者のサンプルに由来する、配列タグの有無を検出することでキャリアオーバー汚染を検出するに当たって便利な手段を提供する。増幅反応を1回用いた後、ハイスループットな次世代配列決定法を使用することにより、複数のB細胞受容体（BCR）鎖をコードしている再構成された核酸の配列に基づいたクロノタイププロファイルを生成する方法を特に対象とする。同様に、増幅反応を1回用いた後、ハイスループットな次世代配列決定法を使用することにより、複数のT細胞受容体（TCR）鎖をコードしている再構成された核酸配列に基づいたクロノタイププロファイル

20

30

#### 【0015】

本発明の一実施形態を、図1Aに例示する。反応混合物において、第1セットのプライマー（22）[第1セットのそれぞれのプライマーは、受容体特異的な部分（16）と、第1のプライマー結合部位を含む5'-非相補的部分（15）とを有する]は標的ポリヌクレオチド（10）の一端にアニーリングし[標的ポリヌクレオチド（10）の融解後]、第2セットのプライマー（24）[第2セットのそれぞれのプライマーは、受容体特異的な部分（20）と、配列タグ（14）及び第2のプライマー結合部位（12）を含む5'-非相補的部分とを有する]は、標的ポリヌクレオチド（10）の別の末端にアニーリングする。いくつかの実施形態では、以下に記載の通り、プライマー（22）の非相補的部分（15）は、配列タグも含み得る。いくつかの状況では、2種類の短い配列タグの方が、多様性が同程度の1種類のより長い配列タグよりも都合のよい場合もある。したがっ

40

50



て、例えば、２種類の 8 - m e r ランダムヌクレオチド配列タグは、１種類の 16 - m e r ランダムヌクレオチド配列タグよりも、偽陽性及びプライマー二量体の形成などを生じさせる恐れが低い。標的ポリヌクレオチド（１０）は、典型的には、Ｔ細胞受容体（ＴＣＲ）又はＢ細胞受容体（例えば、ＩｇＨ鎖又はＩｇＫ鎖の部分）の鎖又は鎖の部分にコードされており、Ｔ細胞又はＢ細胞に由来する、体細胞的に再構成された核酸である。したがって、いくつかの実施形態では、プライマー（２２）及び（２４）の受容体特異的な部分は、それぞれＶ領域の配列及びＪ領域の配列に特異的であってよく、あるいは他の実施形態ではその逆であってもよい。

#### 【００１６】

いくつかの実施形態では、標的ポリヌクレオチド（１０）は、配列のプロファイルが望まれる、限定するものではないが、免疫受容体分子の部分をコードしている再構成された核酸、微生物群の 16 S r D N A、工業的に又は医薬的に重要なタンパク質（酵素など）をコードしている遺伝子のメタゲノムの増幅産物、ヒト若しくは動物の遺伝子、及び／又は癌若しくは感染症などの特定の疾患に係るエキソンなどの、核酸の複合混合物を含み得る。免疫受容体をコードしている再構成された核酸に係る実施形態では、通常、少なくともＶ、Ｄ、又はＪ領域の部分が、第１及び第２のプライマーセットの２つの結合位置の間に存在する。いくつかの実施形態では、第１及び第２のプライマーセットの２つの結合位置の間には、少なくとも、ＩｇＨのＶＤＪ再構成配列、ＩｇＨのＤＪ再構成配列、ＩｇＫのＶＪ再構成配列、ＩｇＬのＶＪ再構成配列、ＴＣＲ のＶＤＪ再構成配列、ＴＣＲ のＤＪ再構成配列、ＴＣＲ のＶＪ再構成配列、ＴＣＲ のＶＪ再構成配列、ＴＣＲ のＶＤＪ再構成配列、又はＴＣＲ のＶＤ再構成配列の部分が存在する。いくつかの実施形態では、第１及び第２のプライマーセットの２つの結合位置の間には、少なくとも、ＩｇＨのＶＤＪ再構成配列、ＩｇＨのＤＪ再構成配列、ＩｇＫのＶＪ再構成配列、又はＩｇＬのＶＪ再構成配列の部分が存在する。いくつかの実施形態では、第１及び第２のプライマーセットの２つの結合位置の間には、少なくとも、ＴＣＲ のＶＤＪ再構成配列、ＴＣＲ のＤＪ再構成配列、ＴＣＲ のＶＪ再構成配列、ＴＣＲ のＶＪ再構成配列、ＴＣＲ のＶＤＪ再構成配列、又はＴＣＲ のＶＤ再構成配列の部分が存在する。更に他の実施形態では、第１及び第２のプライマーセットの２つの結合位置の間には、少なくともＩｇＨのＶＤＪ再構成配列、ＩｇＨのＤＪ再構成配列、及びＩｇＫのＶＪ再構成配列の部分が存在する。並びに、その他の実施形態では、第１及び第２のプライマーセットの２つの結合位置の間には、少なくとも、ＴＣＲ のＶＤＪ再構成配列、ＴＣＲ のＶＪ再構成配列、及びＴＣＲ のＶＤＪ再構成配列、又はＴＣＲ のＶＤ再構成配列の部分が存在する。いくつかの実施形態では、少なくとも、ＶＤＪ再構成部分は、完全なＤ又はＮＤＮ部分と、それらの同定に十分なＶ及びＪセグメント部分とを含む。いくつかの実施形態では、少なくとも、ＶＤＪ再構成部分は、完全なＤ又はＮＤＮ部分と、Ｖ及びＪセグメントの部分とを含む少なくとも５０ヌクレオチドの断片を含む。いくつかの実施形態では、少なくとも、ＶＤＪ再構成部分は、完全なＤ又はＮＤＮ部分と、Ｖ及びＪセグメントの部分とを含む少なくとも７０ヌクレオチドの断片を含む。

#### 【００１７】

いくつかの実施形態では、第１セットは、それぞれＪ断片又はＣ断片に特異的な１種以上のプライマーを含む。このような第１セットのプライマーがそれらの標的配列にアニーリングし、伸長した後、第１セットのうち伸長していないプライマーは除去される。それぞれＶ断片に特異的なものである第２セットのプライマーは、それらの標的配列にアニーリングし、伸長する。他の実施形態では、第１セットは、それぞれＶ断片に特異的なプライマーを含み、このような第１セットのプライマーがそれらの標的配列にアニーリングし、伸長した後、第１セットのうち伸長していないプライマーは除去され、それぞれＪ断片又はＣ断片に特異的なものである第２のプライマーが、それらの標的配列にアニーリングし、伸長する。これらの実施形態のいずれもの代替において、第１及び第２セットは、それぞれ複数のプライマーを含有してよく、それぞれのプライマーは異なる免疫受容体断片に特異的なものであってよい。

10

20

30

40

50

## 【0018】

図1Aに戻ると、いくつかの実施形態では、第1及び第2セットのプライマーは、代替的な実施形態において、融解、アニーリング、及び伸長サイクルを1~10、又は2~10、又は3~10、又は4~10、又は5~10サイクル実施した後、従来法を利用して、反応混合物から伸長していないプライマーを除去することにより、伸長される(5)。その他の実施形態では、第1及び第2セットのプライマーは、代替的な実施形態において、融解、アニーリング、及び伸長サイクルを2~5、又は3~5、又は4~5サイクル実施した後、従来法を利用して、反応混合物から伸長していないプライマーを除去することにより伸長される(5)。更に別の実施形態では、第1及び第2セットのプライマーは、融解、アニーリング、及び伸長を2サイクル実施することにより伸長される。例えば、伸長されていないプライマーは、エキソヌクレアーゼによる消化、磁性ビーズ上の相補的配列に対するハイブリダイゼーション、又は分子ふるいクロマトグラフィー、市販のスピンカラム(例えば、Qiagen QIAquick PCR精製キット)などにより除去することができる。一実施形態では、伸長されていない又は組み込まれていないプライマーは、例えば、エキソヌクレアーゼIにより消化させることにより除去する。伸長(5)の産物である二本鎖DNA(18)は、それぞれの末端に共通の第1及び第2プライマー結合部位を有しており、(いくつかの実施形態では)これらの結合部位に対する相補的配列を有する順方向及び逆方向プライマー(6及び11)を、以降のブリッジPCRによるクラスターの生成に加えてもよい。いくつかの実施形態では、二本鎖DNAは、配列タグ(19)も有し、順方向又は逆方向プライマーは、サンプル又は患者のDNA(18)を同定する又は追跡する又は結合するためのサンプルタグ(2)を含み得る。いくつかの実施形態では、配列タグ(19)は、サンプル中のそれぞれの異なる再構成核酸に実質的に固有のものである。以下に十分に説明される通り、配列リードをクロノタイプに結合させるために、並びにサンプルの汚染を検出及び追跡するために配列タグ(19)を使用することができる。順方向及び逆方向プライマーには、ある種の配列決定プロトコル、例えば、Genome Analyzer(Illumina, San Diego)(17)でブリッジPCR(13)を実施するため、プライマー結合部位(4)及び(8)を含有させることもできる。配列タグを含有しているプライマーにより1回以上の伸長が実施されるその他の実施形態では、サンプル中のそれぞれの異なる再構成核酸は、異なる配列タグを付加されたコピーを有してもよく、したがって、例えば、図1Aに記載の実施形態に準拠して、標的ポリヌクレオチドに対し、融解、アニーリング、及び伸長、のサイクルを4回実施した場合、及びサンプルが再構成された核酸、S<sub>1</sub>を含有する場合、共通のプライマーによる増幅(13)の完了時に、S<sub>1</sub>のコピーは最大で4つの異なる配列タグを有する。したがって、S<sub>1</sub>の配列リードは最大で4つの異なる配列タグを有し得る。以下に十分に説明される通り、このような実施形態では、クロノタイプは、配列タグをアライメントすることと、共通の配列タグにより定義されるそれぞれのサブセットに含まれる配列リードを結合させることを組み合わせることにより決定することができる。

## 【0019】

別の実施形態では、少なくとも2回の伸長工程、及び組み込まれていないプライマーの2回の除去工程が実施された後、共通のプライマーを用いPCRが実施される。図1Bに図示される通り、プライマー(101)は免疫受容体鎖をコードしている再構成された核酸などの標的ポリヌクレオチド(100)の一端にアニーリングし、例えば、DNAポリメラーゼにより伸長される。プライマー(101)は、受容体特異的な部分(103)及び5'非相補的部分(105)をそれぞれ含み得るものであり、同様にして、配列タグ(104)及び第1のプライマー結合部位(102)を含む。上記の通り、組み込まれていないプライマー(130)の伸長及び除去後、反応混合物中の第1の伸長産物(109)に、(a)プライマー(125)[ここで、それぞれのプライマーは、受容体特異的な部分(106)及び5'-非相補的部分(115)(プライマー結合部位を含有する)を含む]と、(b)第1のプライマー結合部位(102)及び5'非相補的部分(117)に特異的な部分(108)を含むプライマー(127)とが添加される。プライマー(125

）及び（１２７）をそれらのプライマー結合部位にアニーリングさせ、伸長させて（１０７）、第２の伸長産物（１１８）を生成した後、伸長していないプライマーを除去する。第２の伸長産物（１１８）に対し、共通の順方向プライマー（１１２）及び逆方向プライマー（１１０）を添加し、ＰＣＲを実施（１１１）した後、得られたアンプリコンを配列決定する（１２０）。上記の図１Ａの実施形態の通り、配列タグを含有しているプライマー〔（１０１）など〕の存在下で１回以上の伸長工程が実施される場合には、常に、同じ標的ポリヌクレオチド（１００）のコピーを複数の異なる配列タグで標識してもよい。

#### 【００２０】

図１Ｃには、明示のＶ、Ｄ及びＪ領域を用いる別の実施形態を例示する。プライマーをアニーリングさせる条件下で、反応混合物中の、ＴＣＲなどの免疫受容体をコードしている再構成された核酸（１２００）に対し、プライマー（１２１２）、すなわちＶ領域（１２２６）に特異的な第１セットのプライマーを添加する。それぞれの第１セットのプライマー（１２１２）は、受容体特異的な部分と、５′-非相補的部分とを含み、同様にして、所望により、配列タグと第１のプライマー結合部位（例えば、図１Ｂ中、１０２、１０３、及び１０４）を含む。第１セットのプライマー（１２１２）は、再構成された核酸（１２００）のＶ領域（１２２６）にアニーリングし、第１セットのプライマー（１２１２）は、Ｄ領域（１２２４）から少なくともＪ領域（１２２２）に、及び場合によりＣ領域（１２２０）に伸長して（１２０２）、場合により配列タグ（１２２８）及び第１のプライマー結合部位（１２３０）を含む第１の伸長産物（１２１６）を生成する。第１セットの伸長していないプライマー（１２１２）の除去後、アニーリング条件下で第２セットのプライマー（１２４０）を反応混合物に添加して、それらのそれぞれの標的Ｊ領域（１２２２）にアニーリングさせた後、それらのプライマーを伸長させて（１２０４）、それぞれ配列タグ（１２３６）（場合による）及び第２のプライマー結合部位（１２３４）を含む第２の伸長産物（１２３２）を生成する。第２の伸長産物（１２３２）は、例えば、配列タグ（１２２８）に示される通りＶ領域（１２２６）に隣接させて、又は配列タグ（１２３６）により示される通りＪ領域（１２２２）に隣接させて配置される単一の配列タグを含んでよく、あるいは、第２の伸長産物（１２３２）は、両方の部分に２つの配列タグを配置されていてもよい。一実施形態では、第２の伸長産物（１２３２）は、Ｖ領域（１２２６）に隣接した単一の配列タグ（１２２８）を含む。

#### 【００２１】

別の実施形態では、第２の伸長産物（１２３２）は、Ｊ領域（１２２２）に隣接した配列タグ（１２３６）を含む。いくつかの実施形態では、配列タグ（１２２８）及び／又は（１２３６）は、下記のコモリタグである。第２セットのプライマー（１２４０）のうち伸長していないプライマーを除去した後、それぞれ第１及び第２のプライマー結合部位（１２３０）及び（１２３４）に特異的な共通の順方向及び逆方向プライマーを添加し、ＰＣＲを実施した（１２０６）。得られたアンプリコンサンプルを配列決定し（１２０８）、クロノタイプ及びクロノタイププロファイルを構築するための配列リードを生成する。

#### 【００２２】

図１Ｄは、所定の長さのテンプレートを生成する方法、及びテンプレートに１つ又は２つの配列タグを付加する方法を例示する。図１Ｄの実施形態には、出発材料としてメッセンジャーＲＮＡ（ｍＲＮＡ）を示すものの、方法にはＤＮＡ又はＲＮＡサンプルのいずれを使用してもよい。ＶＤＪ領域を含むｍＲＮＡ（１３００）に対し、Ｃ領域（１３０８）に特異的な１つ以上のプライマー（１３１２）（「Ｃプライマー」）をアニーリングさせる。通常は単一のＣプライマーのみが使用される。あるいは、Ｊ領域に特異的な１種以上のプライマー（同様の構造を有する）を使用してもよい。Ｃプライマー（１３１２）は、標的特異的な断片（１３１３）、配列タグ断片（１３１４）、及び共通のプライマー結合部位（１３１５）を含む。同様に、標的ｍＲＮＡ（１３００）には、Ｖ領域（１３０２）のオリゴヌクレオチドに特異的なものであり得るポリメラーゼブロッカー（１３１０）がアニーリングする。いくつかの実施形態では、プライマー（１３１２）の伸長に使用する

ポリメラーゼが、鎖置換活性又は5' - 3'エキソヌクレアーゼ活性のいずれをも有していない場合に限り、並びにオリゴヌクレオチドが非伸張性のものであり、例えば、3' - ジデオキシヌクレオチドを有している場合に限り、ブロッカー(1310)は天然のオリゴヌクレオチドであってよい。通常、ブロッカー(1310)は、結合活性及びヌクレアーゼ耐性が増強された、アンチセンス化合物などのオリゴヌクレオチド類似体である。いくつかの実施形態では、ブロッカー(1310)は、ロックド核酸(LNA)又はペプチド核酸(PNA)又は架橋化核酸(BNA)とすることができ、これらの核酸は、次の参考文献、Wengelら、米国特許第6,794,499号；同第7,572,582号；Vesterら、Biochemistry, 43(42):13233~13241(2004)など；並びにKazuyukiら、Chem. Comm., 3765~3767(2007)；Nielson et al, Chem. Soc. Rev., 26:73~78(1997)などに開示されている。ブロッカー(1310)の配列は、プライマー(複数可)(1312)の伸長がV領域(1302)上の所定の位置で停止するように選択される。いくつかの実施形態では、ブロッカー(1310)は、十分なV領域(1302)のみが伸長工程で複製され、V領域が複製された配列から識別され得るように設計される。いくつかの実施形態では、ある程度のミスマッチが許容されるようコンセンサス配列を選択できることから、ポリメラーゼの進行が停止されさえすれば、それぞれのV領域に特異的なブロッカー(1310)を得る必要はない。ブロッカー(1310)の長さは、使用するオリゴヌクレオチド又は類似体の種類に応じ様々に変更することができる。いくつかの実施形態では、ブロッカー(1310)は、10~25モノマーの範囲の長さを有する。いくつかの実施形態では、ブロッカー(1310)は、異なるV領域配列上の異なる位置にアニーリングさせることができる。

#### 【0023】

図1Dに戻ると、プライマー(1312)がブロッカー(1310)まで伸長して、所定の長さを有する標的(1300)のVDJ部分のcDNAコピーが生成される。いくつかの実施形態では、所定の長さ(すなわちブロッカー(1310)の結合部位に相当)は、方法に使用する配列決定法の1つ以上の配列リードにより、VDJ領域の所望の部分がカバーされ得るよう、選択される。伸長の完了後、従来法、例えば、RNA分解酵素(リボヌクレアーゼH及び/又はリボヌクレアーゼAなど)を利用してRNAテンプレート(1300)を消化し(1325)、一本鎖cDNA(1326)を得る。このcDNAには、一般的なプロトコールで末端デオキシヌクレオチド転移酵素(TdT)を使用して、3'モノヌクレオチドテール、例えば、polyCテールを付加する。テールを付加したcDNA(1331)のモノヌクレオチドテールに対し、相補性を有するアダプター(1336)をオーバーハングさせた後、このアダプターを伸長させて二本鎖DNA(1340)を生成し、この二本鎖DNAを、例えば、PCR(1337)により増幅させて、得られたアンプリコンの配列決定(1338)することができる。

#### 【0024】

IgHをコードする核酸など、超変異を受ける再構成された核酸は、同じ再構成された核酸上の異なるプライマー結合部位に結合するプライマーを含むプライマーセットを使用し、増幅させることができる。すなわち、このようなセットは、受容体鎖をコードしている同じ再構成された核酸上の、オーバーラップしていない1つ以上のプライマー結合部位に結合するプライマーを包含し得る。このようなセットは、第1セットのプライマー及び第2セットのプライマーのいずれか又は両方を含んでよい。いくつかの実施形態では、超変異を受ける再構成された核酸は、第1セットのプライマー及び第2セットのプライマーにより増幅され、ここで、2セットのうち少なくとも1セットは、オーバーラップしていない複数のプライマー結合部位に特異的なプライマーを含み、例えば、あるセットは、それぞれ異なるV断片を含有し、複数のプライマーは、異なるV断片に結合するものでありオーバーラップしていない異なるプライマーにそれぞれ特異的である。超変異を受ける再構成された核酸の増幅に応用可能な実施形態を図3A~3Bに示す。図中、ネステッドプライマーセットは、例えば、体細胞超変異、又はクローン進化などの条件下で、サンプル

中の異なる再構成核酸を確実に増幅させるべく利用する。再構成された核酸（例えば、I g Hをコードしている分子）を、アニーリング条件下で、プライマーの第1のネステッドセット（302）〔本例では、再構成された核酸（300）のV領域（316）の異なる部位に対し特異的な群（304）、（306）及び（308）を含む〕と合わせる。本実施形態では、第1のネステッドセットは、V領域の異なる部位又は位置にそれぞれ特異的な複数のプライマー群を含み、群中の異なるメンバーは、部位が異なるV領域変異体に特異的である。いくつかの実施形態では、複数のグループとは2～4であり、他の実施形態では、複数とは2又は3である。いくつかの実施形態では、第1のネステッドセット（302）のそれぞれのプライマーは、固有の配列タグ（314）に加え、5'非相補的テールに第1のプライマー結合部位（312）を有し得る。第1のネステッドセット（302）のプライマーは、それらの標的である再構成された核酸にアニーリングし、D領域（318）と、少なくともJ領域（320）の部分とにわたって伸長して、第1のアンプリコン（323）を生成する。この第1のアンプリコンは、それぞれ3つのサブプライマーセット（304）、（306）及び（308）に対応する3つの生成物（330）、（332）及び（334）を含む。第1のアンプリコン（323）のそれぞれのメンバーには、配列タグ（324）及びプライマー結合部位（326）が組み込まれている。

#### 【0025】

伸長していないプライマーの除去（322）後、アニーリング条件下で、プライマーの第2のネステッドセット（340）を反応混合物に添加する。図3Aに図示される通り、第2のネステッドセットのプライマー（340）は、プライマーのサブセット（336）及び（338）を含み、これらのサブセットは、第1のアンプリコン（323）メンバーのJ領域（320）上の、重複していない異なる部分にアニーリングする。いくつかの実施形態では、プライマーの第2のネステッドセットは、単一のプライマーグループのみを含有してよい。第2のネステッドセット（340）のプライマーは、伸長して、サブセット（350）、（352）及び（354）を含む第2の伸長産物（360）を生成し、今度はこの伸長産物のそれぞれが、プライマー（336）及び（338）に対応するサブセット（サブサブセット）を更に2つ含む。いくつかの実施形態では、プライマーの第2のネステッドセット（340）は、単一のプライマー結合部位にのみ特異的なプライマーを含有し、プライマーの第1のネステッドセット（302）は、オーバーラップしていない少なくとも2つのプライマー結合部位に特異的なプライマーを含有する。伸長していないプライマーの除去（342）後、共通の順方向及び逆方向プライマーを添加してPCR（356）を実施し、得られるアンプリコンサンプルを配列決定（358）することもある。様々な実施形態において、第1のネステッドセット及び第2のネステッドセットの両方のプライマーが配列タグ（339）を含有してよく；第1のネステッドセットのプライマーが配列タグを含有し、第2のネステッドセットは配列タグを含有しなくてよく；並びに第2のネステッドセットのプライマーが配列タグを含有し、第1のネステッドセットは配列タグを含有しなくてよい。いくつかの実施形態では、第1のネステッドセットのプライマーを最初に伸長させ、その後、伸長していないプライマーを除去又は破壊し、第2のネステッドセットのプライマーをアニーリング及び伸長させる（図3A～3Bに示す通り）。他の実施形態では、アニーリング工程、伸長工程、及び除去工程の順番を入れ替え、すなわち、最初に第2のネステッドセットのプライマーを伸長させた後、伸長していないプライマーを除去又は破壊し、第1のネステッドセットのプライマーをアニーリング及び伸長させる。

#### 【0026】

上記方法のいくつかの実施形態では、例えば、サンプル中の標的ポリヌクレオチドの大部分に配列タグを付加することを目的として、1つ以上の伸長工程（322）又は（342）のいずれかが実施される。このような実施形態では、標的とするポリヌクレオチド及び/又はそれらのコピーに、異なる配列タグを1つ以上付加することもある。すなわち、標的ポリヌクレオチド及びPCRなどの増幅反応に由来するそれらの複製分子には、異なる配列タグを複数付加することができ、したがって、もとの標的ポリヌクレオチドのコ

10

20

30

40

50

ピーは1つ以上の配列タグで標識され得る。以下に十分に説明される通り、このような複数の配列タグは、キャリアオーバー汚染を追跡するのに、及び標的ポリヌクレオチド配列をより高感度に決定するのに尚も有用である。

#### 【0027】

上記の実施形態のうちいくつかのものを、以下の工程を用い実施することもできる。例えば、多数の又は複数のT細胞受容体鎖からクロノタイププロファイルを生成する方法は、次の工程：(a) 反応混合物中、プライマー伸長条件下で、第1セットのプライマーとT細胞由来の再構成された核酸サンプルとを組み合わせる工程[ここで、第1セットのそれぞれのプライマーは受容体特異的な部分を有し、このとき受容体特異的な部分は、標的再構成された核酸の所定の位置又は部位にて異なる再構成核酸にアニーリングし、伸長して第1伸長産物を生成するような長さを備え、第1セットのそれぞれのプライマーは、第1のプライマー結合部位を含有する5'-非相補的末端を有する]；(b) 反応混合物から、伸長していない第1セットのプライマーを除去する工程；(c) プライマー伸長条件下で、反応混合物に第2のプライマーセットを添加する工程[ここで、第2セットのそれぞれのプライマーは、受容体特異的な部分を有し、このとき受容体特異的な部分は、所定の位置又は部位にて第1の伸長産物にアニーリングし、第2のプライマー結合部位を含有する5'-非相補的末端を有し、第1セットのプライマー及び/又は第2セットのプライマーは、それぞれ、受容体特異的な部分と第1又は第2のプライマー結合部位との間に配置された配列タグを含み、第2セットのそれぞれのプライマーは、伸長して第2の伸長産物を生成し、このとき、それぞれの第2の伸長産物は、第1のプライマー結合部位と、第2のプライマー結合部位と、少なくとも1つの配列タグと、(i) T細胞受容体鎖のVセグメント部分及びJセグメント部分、(ii) T細胞受容体鎖のVセグメント部分及びJセグメント部分、又は(iii) T細胞受容体鎖のVセグメント部分及びJセグメント部分のいずれかと、を含む]；(d) 反応混合物においてポリメラーゼ連鎖反応を実施して、アンプリコンを生成する工程[このポリメラーゼ連鎖反応は、第1のプライマー結合部位に特異的な順方向プライマーと、第2のプライマー結合部位に特異的な逆方向プライマーとを使用する]；並びに(e) アンプリコンの核酸を配列決定して、多数のT細胞受容体鎖のクロノタイププロファイルを生成する工程、を含む。本明細書で使用する  
とき、反応混合物中での「プライマー伸長条件」としては、実質的にすべてのプライマー結合部位が一本鎖の状態であるような条件が挙げられる。いくつかの実施形態では、このような条件は、二本鎖の標的核酸を融解させることで得られ、融解させた結果、プライマー結合部位は一本鎖の状態になり、それらの標的核酸にプライマーをアニーリングさせて、ポリメラーゼによる伸長のための足がかりを形成させることができるようになる。

#### 【0028】

第1及び第2セットのプライマーに結合させる所定の位置又は部位は、以下に引用する参照により例示される通りの多重核酸増幅法(多重PCRなど)の従事者により知られる従来法により決定してもよい。例えば、標的ポリヌクレオチドが免疫受容体分子をコードしている再構成された核酸である場合、Faham and Willis (cited above), Van Dongen et al, Leukemia, 17:2257~2317(2003)並びに同様の参照文献が、標的ポリヌクレオチドを多重増幅する際にプライマー結合部位を選別する指針を提供する。いくつかの実施形態では、このような所定の位置又は部位の選択は、(i) 増幅効率に及ぼす影響(忠実なアンプリコン中の異なるコピーの頻度が、サンプル中の標的ポリヌクレオチドの頻度を表わすことが望ましい)、(ii) DNAの配列決定に利用する化学的手法において、アンプリコン中のコピーに必要とされる長さに及ぼす影響、(iii) 選択したプライマーが、所望の多様性を備える再構成された核酸の部分(例えば、VDJ領域にまたがるものである)、などのいくつかの要素に基づくものである。本態様に関し、本発明では、類似体シグナルではなく配列セットが読み出されることから、プライマーの、異なる標的ポリヌクレオチドとの交差反応性は、本発明の方法による結果に影響を与えない(例えば、PCR増幅産物及びスペクトルの入力(spectratyping)などのアナログ読み出しにのみ基づく方法と比較し

10

20

30

40

50

て与えない)とする識別及び認識を含む。

【0029】

いくつかの実施形態では、配列決定をする工程は、次の工程：(i)それぞれエラー率を有し、それぞれヌクレオチド配列と、タグ配列を含む複数の配列リードを提供することと、(ii)同様のタグ配列を有する配列リード群をアライメントさせた後、グループの配列リードをもとにベースコールを作成してヌクレオチド配列を決定することと、を含む。次に、このようなグループレベルのヌクレオチド配列を以下に記載のものと同じ又は異なるクロノタイプに結合させることもできる。いくつかの実施形態では、PCR工程において、第1及び第2セットのプライマーの受容体特異的な部分の長さは、アンプリコン中の異なる再構成核酸の相対的なレベルが、サンプル中の再構成された核酸のレベルと実質的に同じになるように選択される。このようにプライマーを選択するに当たり、それぞれの標的ポリヌクレオチドに対するプライマーの結合部位の位置及び長さを変更してもよい。いくつかの実施形態では、配列タグは、サンプル中の異なる標的ポリヌクレオチドの数よりも十分大きくなるように配列タグのセットから選択され、その結果、サンプル中の実質的にすべての別個の標的ポリヌクレオチド及びそれらのコピーは異なる配列タグを有することになる(例えば、Brenner、米国特許第7,537,897号に記載の方法、「サンプリングによる標識(labeling by sampling)」による)。いくつかの実施形態では、このようなセット中の配列タグ数は、サンプル中の標的ポリヌクレオチド集団の総数の少なくとも100倍である。更に、もとの標的ポリヌクレオチド及びそれらのコピーの実質的にすべてが同じ固有の配列タグで標識されているいくつかの実施形態では、配列決定する工程は、アンプリコンの核酸の配列リードを生成する工程と、同じ配列タグを有する配列リードをアライメントして、サンプルの同じクロノタイプに対応する配列リードを決定する工程と、を含む。更に、いくつかの実施形態では、アライメントする工程は、同じ配列タグを有する配列リードのそれぞれのヌクレオチド位置に主要なヌクレオチドを決定することにより、それぞれのクロノタイプのヌクレオチド配列を決定することを更に含む。更に、いくつかの実施形態では、伸長していないプライマーを除去する工程は、3' 5'一本鎖エキソヌクレアーゼ活性を有するヌクレアーゼを使用して、反応混合物中の一本鎖核酸を消化することにより実施することもできる(消化は、例えば、大腸菌(E. coli)エキソヌクレアーゼIにより行うことができ、このエキソヌクレアーゼは、熱により簡単に失活させることができる)。更なる実施形態では、上記の方法を用い、骨髄腫、リンパ腫、又は白血病患者などの癌患者の微小残存病変を診断及び/又はモニタリングするためのクロノタイププロファイルを生成することもできる。このような診断及び/又はモニタリングは、上記方法の工程後に、次の追加の工程：クロノタイププロファイルから、癌に関連する1つ以上の患者特異的なクロノタイプの存在、非存在、及び/又はレベルを決定すること、により実施することもできる。この実施形態の方法には、1つ以上の配列タグのそれぞれの配列を決定する工程、及びこのような配列を、それまでに決定されたクロノタイププロファイルの配列タグの配列と比較して、汚染配列の存在、非存在、及び/又はレベルを決定する工程、を更に含んでもよい。いくつかの実施形態では、このような比較工程は、1つ以上(one of more)の配列タグの配列を、患者以外の少なくとも1名の個体由来のクロノタイプを含有しているクロノタイプデータベースの配列タグと比較することを包含する。

【0030】

更に別の実施形態では、T細胞受容体構成成分、及びをコードしている複数の再構成された核酸を1回の反応で増幅させる方法には、次の工程：(a)反応混合物中、プライマー伸長条件下で、第1セットのプライマーとT細胞由来の再構成された核酸サンプルとを組み合わせる工程[ここで、異なる再構成核酸は、第1末端に、少なくとも、T細胞受容体のJ、J、又はJ断片の部分を含み、第1セットのそれぞれのプライマーは、それぞれ、受容体特異的な部分を有し、このとき、受容体特異的部分は、異なる再構成核酸の第1末端にアニーリングし、伸長して、第1伸長産物を生成するような長さを備え、第1セットのそれぞれのプライマーは、3' 5'の方向で配列タグ及び第1のブラ

10

20

30

40

50

イマー結合部位を含有する5' - 非相補的末端を有し、配列タグは、実質的に第1セットのプライマー毎に異なるものである] ; (b) 反応混合物から、伸長していない第1セットのプライマーを除去する工程 ; (c) プライマー伸長条件下で、反応混合物に第2のプライマーセットを添加する工程 [ 第2セットのそれぞれのプライマーは受容体特異的部分を有し、このとき、受容体特異的部分は、第1の伸長産物にアニーリングし、伸長して、第2の伸長産物を生成するような長さを備え、それぞれの第2の伸長産物は、少なくとも、T細胞受容体のV<sub>H</sub>、V<sub>L</sub>、又はV<sub>H</sub>断片の部分を含み、第2セットのそれぞれのプライマーは、第2のプライマー結合部位を含有する5' - 非相補的末端を有する] ; 並びに (d) 反応混合物にポリメラーゼ連鎖反応を実施して、アンプリコンを生成する工程 [ このポリメラーゼ連鎖反応は、第1のプライマー結合部位に特異的な順方向プライマーと、第2のプライマー結合部位に特異的な逆方向プライマーとを使用する]、を含む。上記の方法には、アンプリコンの配列サンプルを配列決定する工程を更に包含させてもよい。典型的には、このようなサンプルは「代表的なサンプル」であり、サンプル中に、もとの生体試料におけるものとおおよそ同じ頻度で異なるクロノタイプを存在させるのに十分大きなものである。いくつかの実施形態では、配列決定する工程は、それぞれエラー率を有し、それぞれヌクレオチド配列と、タグ配列とを含む複数の配列リードを提供することと、同様のタグ配列を有する配列リードをアライメントさせて、同じクロノタイプに対応する配列リードを決定することと、を含む。もとの標的ポリヌクレオチド又はそれらのコピーに複数の配列タグが付加されている場合は常に、配列リードは、以下に更に十分に記載する結合工程でプロセッシングすることができる。

#### 【0031】

別の実施形態では、多数のT細胞受容体鎖からクロノタイププロファイルを生成する方法は、次の工程 : (a) 反応混合物中、プライマー伸長条件下で、第1セットのプライマーとT細胞由来の再構成された核酸サンプルとを組み合わせる工程 [ ここで、それぞれの第1セットのプライマーは受容体特異的な部分を有し、このとき受容体特異的な部分は、所定の位置にて異なる再構成核酸にアニーリングし、伸長して第1伸長産物を生成するような長さを備え、それぞれの第1セットのプライマーは、第1プライマー結合部位を含有する5' - 非相補的末端を有する] ; (b) 反応混合物から、伸長していない第1セットのプライマーを除去する工程 ; (c) 反応混合物に第2セットのプライマーを添加する工程 [ ここで、第2セットのそれぞれのプライマーは、ある程度の長さを備える受容体特異的部分を有し、受容体特異的部分は、所定の位置にて第1の伸長産物に特異的なものであり、第2のプライマー結合部位を含有する5' - 非相補的末端を有し、第1セットのプライマー及び/又は第2セットのプライマーは、それぞれ、受容体特異的な部分と第1又は第2のプライマー結合部位との間に配置された配列タグを含む] ; (d) 第1のポリメラーゼ連鎖反応を実施して、第1のアンプリコンを生成する工程 [ この第1のポリメラーゼ連鎖反応は、第1のプライマー結合部位に特異的な順方向プライマーと、第2セットのプライマーとを使用し、それぞれの第1のアンプリコンの核酸配列は、第1のプライマー結合部位と、第2のプライマー結合部位と、少なくとも1つの配列タグと、T細胞受容体鎖のV<sub>H</sub>セグメント部分及びJ<sub>H</sub>断片部分、T細胞受容体鎖のV<sub>L</sub>セグメント部分及びJ<sub>L</sub>断片、又はT細胞受容体鎖のV<sub>H</sub>セグメント部分及びJ<sub>H</sub>セグメント部分のいずれかとを含み、第1及び第2セットのプライマーの受容体特異的部分の長さは、アンプリコンにおいて、異なる再構成核酸の相対的なレベルが、サンプル中の異なる再構成核酸のレベルと実質的に同じになるよう選択される] ; (e) 第2のプライマー結合部位について特異的な逆方向プライマーを添加する工程 ; (f) 反応混合物に第2のポリメラーゼ連鎖反応を実施して、第2のアンプリコンを生成する工程 [ このポリメラーゼ連鎖反応は、第1のプライマー結合部位に特異的な順方向プライマーと、第2のプライマー結合部位に特異的な逆方向プライマーとを使用する] ; 並びに (g) 第2のアンプリコンの核酸を配列決定して、T細胞受容体鎖のクロノタイププロファイルを生成する工程、を含み得る。いくつかの実施形態では、配列決定する工程は、それぞれエラー率を有し、それぞれヌクレオチド配列と、タグ配列とを含む複数の配列リードを提供することと、同様のタグ配列を有する



配列リードをアライメントさせて、同じクロノタイプに対応する配列リードを決定することと、を含む。標的ポリヌクレオチド及び／又はそれらのコピーが1つ以上の配列タグで標識される更なる実施形態では、同様の配列タグのアライメント後、配列リードは、以下により十分に記載する通りの更なる結合工程においてプロセッシングすることもできる。

#### 【0032】

別の実施例では、多数のB細胞受容体鎖からクロノタイププロファイルを生成する方法は、次の工程：(a) 反応混合物中、プライマー伸長条件下で、プライマーの第1のネステッドセットと、B細胞由来の再構成された核酸サンプルとを組み合わせる工程[ここで、第1のネステッドセットは1つ以上のプライマー群を含み、それぞれのグループのそれぞれのプライマーは受容体特異的部分を有し、このとき受容体特異的部分は、異なる群由来のそれぞれのプライマーの受容体特異的部分が所定の部位にて異なる再構成核酸にアニーリングするような長さを備え、この所定の部位は、第1のネステッドセットの任意のその他のプライマーの所定の部位とオーバーラップせず、それぞれのグループのそれぞれのプライマーは、第1のプライマー結合部位を含有している5'-非相補的末端を有する]；(b) 第1のネステッドセットのプライマーを伸長させて、第1の伸長産物を生成する工程；(c) 第1のネステッドセットの反応混合物から、伸長していないプライマーを除去する工程；(d) プライマー伸長条件下で、反応混合物に、第2のネステッドセットのプライマーを添加する工程[ここで、第2のネステッドセットは1つ以上のプライマー群を含み、それぞれのグループのそれぞれのプライマーは受容体特異的部分を有し、このとき、異なる群由来のそれぞれのプライマーの受容体特異的部分は、所定の部位にて第1の伸長産物にアニーリングするような長さを備え、この所定の部位は、第2のネステッドセットの任意のその他のプライマーの所定の部位とはオーバーラップせず、それぞれのグループのそれぞれのプライマーは、第2のプライマー結合部位を含有する5'-非相補的末端を有し、第1のネステッドセットのプライマー及び／又は第2のネステッドセットのプライマーは、それぞれ受容体特異的部分と第1又は第2のプライマー結合部位との間に配置された配列タグを含む]；(e) 第2のネステッドセットのプライマーを伸長させて第2の伸長産物を生成する工程[このとき、それぞれの第2の伸長産物は、第1のプライマー結合部位と、第2のプライマー結合部位と、少なくとも1つの配列タグと、(i) B細胞受容体重鎖のVセグメント部分及びJセグメント部分又は(ii) B細胞受容体軽鎖のVセグメント部分及びJセグメント部分のいずれかとを含む]；(f) 反応混合物においてポリメラーゼ連鎖反応を実施して、アンプリコンを生成する工程[このポリメラーゼ連鎖反応は、第1のプライマー結合部位に特異的な順方向プライマーと、第2のプライマー結合部位に特異的な逆方向プライマーとを使用する]；並びに(g) アンプリコンの核酸を配列決定して、多数のB細胞受容体鎖のクロノタイププロファイルを生成する工程、により実施することもできる。

#### 【0033】

いくつかの実施形態では、(伸長産物の融解後)プライマーをアニーリング及び伸長させる1回以上のサイクルは、工程(b)及び／又は(e)で実施することができ、この場合、サンプル中のもともとの再構成された核酸のコピーを1つ以上の配列タグにより標識してもよい。これらの実施形態では、配列決定する工程(g)には、クロノタイプ及びクロノタイププロファイルを決定するための下記の通りのアライメント及び結合工程を更に包含させてもよい。いくつかの実施形態では、例えば、工程(b)及び(e)において単回の伸長のみがなされる場合、配列決定する工程は、それぞれエラー率を有し、それぞれヌクレオチド配列と、タグ配列とを含む複数の配列リードを提供することと、同様のタグ配列を有する配列リードをアライメントさせて、同じクロノタイプに対応する配列リードを決定することと、を含む。上記の通り、いくつかの実施形態では、PCR工程において、第1及び第2セットのプライマーの受容体特異的部分の位置及び長さは、アンプリコン中の異なる再構成核酸の相対的なレベルが、サンプル中の再構成された核酸のものと実質的に同じになるように選択される。

#### 【0034】

いくつかの実施形態では、配列タグは、プライマー伸長工程において標的ポリヌクレオチド又はそれらのコピーに付加され、実質的にすべての異なる標的ポリヌクレオチド及びそれらのコピーが同じ配列タグにより標識される。他の実施形態では、サンプル又はそれらのコピーの標的ポリヌクレオチドは、1つ以上の異なる配列タグにより標識することもできる。以下に更に例示される通り、いくつかの実施形態では、配列タグを含有しているプライマー（第1セットのプライマー又は第2セットのプライマーのいずれか）の存在下で複数回の伸長又は複数回のPCRのサイクルを実施することができ、これにより、結果として、同じ標的ポリヌクレオチド及び/又はそのコピーに異なる配列タグが付加され得る。

#### 【0035】

クロノタイプ解析における配列タグ

一態様では、本発明は、T細胞受容体（TCR）又はB細胞受容体（BCR）又はこれらの所定の断片などの免疫分子のレパトアから、配列データを取得及び解析して、迅速かつ効率的にクロノタイププロファイルを決定するための方法を目的とする。配列データは、典型的には、配列リードの大規模コレクションを含み、すなわち、DNAシーケンサ由来のベースコール配列及び関連するクオリティスコアを使用して、免疫分子を解析する。クロノタイププロファイルの構築に当たり、抽出工程、配列決定をするための化学的手法、又は増幅を行うための化学的手法などといった生体以外に由来するエラーを含有している配列リードから、本当に違いを含んでいる配列リードを迅速かつ正確に識別することが重要な課題である。本発明の態様は、複合体の配列リードが同じもとの標的ポリヌクレオチドに由来するものかどうかの判定を補助するため、それぞれの標的ポリヌクレオチド（例えば、サンプル中の再構成された核酸）に、固有の配列タグを付加することを包含する。本発明の一態様によると、タグ-分子複合体を生成するため、配列タグが体細胞性の再構成された核酸分子に付加され、このような複合体の再構成核酸はそれぞれ固有の配列タグを有する。通常、このような付加は、T細胞及び/又はB細胞及び/又は無細胞系DNAを含有しているサンプルから核酸分子を抽出した後になされる。好ましくは、このような固有の配列タグは、ハミング距離などの従来の配列の距離測定法により評価される通り、可能な限り互いに大きく異なるものである。タグ分子複合体中の配列タグ間の距離を最大化することで、配列決定エラー及び増幅エラーが高率の場合でさえ、複合体の配列タグが、異なる複合体の任意のその他のタグ配列よりも、もともとのタグ配列にはるかに近いまま保持される。例えば、16-mer配列タグを利用し、かつクロノタイプのセットのそれぞれのこのようなタグが、クロノタイプのすべてのその他の配列タグから少なくとも50%の距離又は8ヌクレオチドの距離でハミング距離を有する場合、タグを別のものに交換して配列タグのミスリードを生じさせるには少なくとも8つの配列決定エラー若しくは増幅エラーが必要になる（及び誤った配列タグを備えたクロノタイプの配列リードが不適切にグループ分けされる必要がある）。一実施形態では、配列タグは、再構成された核酸分子に付加してタグ-分子複合体を生成した後に選別され、タグ-分子複合体のタグ間のハミング距離は、配列タグの全長の少なくとも25%の数値であり（すなわち、それぞれの配列タグは、少なくとも25%のヌクレオチドが、その他のすべてのタグの配列とは異なっている）；別の実施形態では、このような配列タグ間のハミング距離は、配列タグの全長の少なくとも50%の数値である。

#### 【0036】

一態様では、本発明は、次の工程：（a）個人から、T細胞及び/又はB細胞及び/又は無細胞系DNAを含むサンプルを得る工程；（b）サンプル中のT細胞受容体遺伝子又は免疫グロブリン遺伝子の再構成された核酸分子に配列タグを付加して、タグ-分子複合体を生成する工程（ここで、実質的にすべてのタグ-分子複合体は固有の配列タグを有する）；（c）タグ-分子複合体を増幅させる工程；（d）タグ-分子複合体を配列決定する工程；並びに（e）同様の配列タグの配列リードをアライニングして、サンプル中の同じ再構成された核酸に対応する配列リードを決定する工程と、を含む。B細胞又はT細胞を含有しているサンプルは、従来法を利用して得る。配列タグを付加する工程において

10

20

30

40

50

、好ましくは、配列タグは、固有なだけでなく、別のタグと十分異なるものでもあり、すなわち、配列決定エラー又は増幅エラーがかなり大きい場合でさえ、配列タグが別のタグに変更される尤度はゼロに近い。大抵の配列決定技術では、配列タグの付加後、タグ-分子複合体を増幅させる必要があるものの、単一分子配列決定法を利用する場合、増幅工程は任意選択になる。1分子配列決定法としては、限定するものではないが、1分子リアルタイム(SMRT)配列決定法、又はナノポア配列決定法など、例えば、米国特許第7,313,308号;同第8,153,375号;同第7,907,800号;同第7,960,116号;同第8,137,569号;並びにManrao et al, Nature Biotechnology, 4(8):2685~2693(2012)が挙げられる。

10

#### 【0037】

別の態様では、本発明は、固有の配列タグをカウントすることにより、サンプル中のリンパ細胞数を求める方法を包含する。配列タグを用いずとも、TCR又はIGH遺伝子のクロノタイプ、特に、V(D)J領域を含むクロノタイプにより、リンパ細胞及びそのクローンに固有のマーカーを提供する。再構成された核酸がゲノムDNAから得られる場合、サンプル中のリンパ細胞数は、配列決定後にカウントされる固有のクロノタイプの数により評価することもできる。このアプローチは、同じクロノタイプに関連する同一のリンパ細胞の実質的にクローン性の集団が存在する場合は常に(あるいは、再構成された核酸がサンプルのmRNAから得られ、かつそれぞれの配列のクオリティが発現率並びに細胞数を反映し得る又はこれらに応じる場合)、破綻する。配列タグの使用により、この問題(short coming)は克服され、かつ特に、リンパ腫又は白血病などの様々なリンパ性疾患に罹患している患者においてリンパ細胞を計数するのに特に有用である。本発明の一態様によると、配列タグを使用して、白血病などの多量のクローンが存在するかどうかに関係なくサンプル中のリンパ細胞の絶対数を得ることもできる。このような方法は、工程:(a)個人からリンパ球を含むサンプルを得る工程;(b)T細胞受容体遺伝子又はリンパ球のイムノグロブリン遺伝子の再構成された核酸分子に配列タグを付加して、タグ-分子複合体を生成する工程(ここで、実質的にすべての分子のタグ-分子複合体は固有の配列タグを有する);(c)タグ-分子複合体を増幅させる工程;(d)タグ-分子複合体を配列決定する工程;並びに(e)異なる配列タグの数を計数して、サンプル中のリンパ球の数を決定する工程、により実施することもできる。いくつかの実施形態では、再構成された核酸分子はゲノムDNAに由来する。

20

30

#### 【0038】

本発明の一実施形態では、配列タグは、例えば、Brennerら、米国特許第5,846,719号;Brennerら、米国特許第7,537,897号;Macevitz、国際公開第WO 2005/111242号などに開示されるものなどの、サンプリングによる標識により、サンプルの再構成された核酸分子に付加される;これらの文献は参照により本願に援用される。サンプリングによるラベリングでは、標識される(すなわち固有のタグを付加される)一群のポリヌクレオチドを、より多量のサンプル配列タグ群に対し使用する(付加、結合、又は同様の方法による)。すなわち、ポリヌクレオチド集団がKメンバーを有し(同じポリヌクレオチドの複製を含む)、配列タグ集団がNメンバーを含む場合、 $N \gg K$ になる。一実施形態では、本発明で使用する配列タグ集団の大きさは、サンプル中のクロノタイプ集団の大きさの少なくとも10倍である;別の実施形態では、本発明で使用する配列タグ集団の大きさは、サンプル中のクロノタイプ集団の大きさの少なくとも100倍である;別の実施形態では、本発明で使用する配列タグ集団の大きさは、サンプル中のクロノタイプ集団の大きさの少なくとも1000倍である。他の実施形態では、配列タグ群の大きさは、クロノタイプを配列タグ群と組み合わせたときに、例えば、ライゲーション反応、又は増幅反応などといった、付加する反応において組み合わせたときには常に、サンプル中の実質的にすべてのクロノタイプが固有の配列タグを有するように、選択される。いくつかの実施形態では、実質的にすべてのクロノタイプは、このようなクロノタイプのうち少なくとも90%が固有の配列タグを有し得ることを意味

40

50

する；他の実施形態では、実質的にすべてのクロノタイプは、このようなクロノタイプのうち少なくとも99%が固有の配列タグを有し得ることを意味する；他の実施形態では、実質的にすべてのクロノタイプは、このようなクロノタイプのうち少なくとも99.9%が固有の配列タグを有し得ることを意味する。多くの組織サンプル又は生検において、T細胞又はB細胞の数は、最大又は約 $1 \times 10^6$ 個である；したがって、このようなサンプルを用いる本発明のいくつかの実施形態では、サンプリングによる標識において用いられる固有の配列タグ数は、少なくとも $10^8$ 個であり、又は他の実施形態では少なくとも $10^9$ 個である。

#### 【0039】

このような実施形態では、最大で $1 \times 10^6$ 個のクロノタイプがサンプリングにより標識され、全4つのヌクレオチド前駆体混合物を、合成反応のそれぞれの付加工程において反応させることによるコンビナトリアルな合成により、配列タグの大型のセットが効率的に生成され得る（例えば、参照により援用されるChurch, 米国特許第5,149,625号に開示されている）。その結果、配列タグのセットは構造「 $N_1 N_2 \dots N_k$ 」を有するようになり、ここで、それぞれの $N_i = A, C, G$ 又は $T$ であり、かつ $k$ はタグ中のヌクレオチド数である。このようなコンビナトリアルな合成により作製される配列タグセット中の配列タグ数は $4^k$ 個である。したがって、サンプリングによる標識により $10^6$ 個の分子集団に配列タグを付加するには、 $k$ が少なくとも14である、又は $k$ が約14~18の範囲である配列タグセットが適当である。上記構造を有する配列タグセットは、本発明の方法を実施する間、差異又はエラーを導入し得る数多くの配列を含む。例えば、上記のコンビナトリアルに合成した配列タグセットには、ホモポリマー断片を含む数多くのメンバータグが包含され、この場合、合成時解読法（sequencing-by-synthesis）などのいくつかの配列決定アプローチでは、ある程度の長さを超えると正確に測定することが難しくなる。したがって、本発明は、配列決定などの特定の工程について効率的な構造を有するコンビナトリアルに合成した配列タグを含む。例えば、天然の4種のヌクレオチドを、コンビナトリアルな合成において代替的に使用されるばらばらのサブセットに分割して、合成時解読に効率のよいいくつかの配列タグ構造を作製することにより、ホモポリマーが所定の長さを超えることを防ぐことができる。例えば、 $z$ を $A$ 又は $C$ のいずれかにして、 $x$ を $G$ 又は $T$ のいずれかにすることで、

$[(z)_1 (z)_2 \dots (z)_i] [(x)_1 (x)_2 \dots (x)_j] \dots$  が得られ、

ここで、 $i$ 及び $j$ は同じであっても異なってもよく、任意のホモポリマー断片の大きさを制限するよう選択される。一実施形態では、 $i$ 及び $j$ は1~6の範囲である。このような実施形態では、配列タグは、12~36ヌクレオチドの範囲の長さを有してよく；並びに、その他の実施形態では、このような配列タグは、12~24ヌクレオチドの範囲の長さを有してよい。他の実施形態では、その他のヌクレオチドをペアとして使用することもでき、例えば、 $z$ を $A$ 又は $T$ とし、かつ $x$ を $G$ 又は $C$ とすることができ；あるいは $z$ を $A$ 又は $G$ とし、かつ $x$ を $T$ 又は $C$ とすることができ。あるいは、 $z'$ を4つの天然のヌクレオチドのうち3つの組み合わせとし、 $x'$ を $z'$ 以外のいずれかのヌクレオチドにする（例えば、 $z'$ を $A, C$ 又は $G$ とすると、 $x'$ は $T$ になる）。これにより、次の通りの配列タグ構造が与えられる：

$[(z')_1 (z')_2 \dots (z')_i] x' [(z')_1 (z')_2 \dots (z')_i] x' \dots$

ここで、 $i$ は上記の通り選択され、かつ $x'$ の出現は、中断部分として提供され、任意の望ましくないホモポリマーが終結させられる。

#### 【0040】

##### 更なる配列タグ

本発明は、ゲノムDNA断片などの核酸を、増幅及び配列決定より前に、「モザイクタグ」を含み得る固有の配列タグにより標識する方法を用いるものである。このような配列タグは、増幅エラー及び配列決定エラーを識別するのに有用である。モザイクタグは、従



る位置と、1～10塩基の長さでヌクレオチドとを有し、かつそれぞれの可変領域は、モザイクタグにおける位置と、1～10塩基の長さでランダムに選択されたヌクレオチドとを有し、同じ位置の定常領域は同じ長さを有し、かつ同じ位置の可変領域は同じ長さを有する；(i i i) タグ-テンプレート複合体の多重度を増幅させる工程と；(i v) 増幅させたそれぞれのタグ-テンプレート複合体の配列リードを複数生成する工程と；(v) 同一のモザイクタグを有するそれぞれの複数の配列リードのそれぞれのヌクレオチド位置にて、一致するヌクレオチドを決定することによりそれぞれの核酸のヌクレオチド配列を決定する工程；で使用することもできる。別の態様では、モザイクタグを以下の工程：(a) サンプル中の核酸から一本鎖核酸DNAテンプレートを調製する工程；(b) 一本鎖核酸DNAテンプレートをサンプリングしてタグ-テンプレート複合体を形成させることにより標識する工程、ここで、タグ-テンプレート複合体の実質的にすべての一本鎖核酸DNAテンプレートは、少なくとも15塩基の長さのヌクレオチドを有し、かつ次の形態

$$[(N_1 N_2 \dots N_{K_j})(b_1 b_2 \dots b_{L_j})]_M$$

を有する固有の配列タグ(すなわち、モザイクタグ)を有し、ここで、 $i = 1, 2, \dots, K_j$ であるそれぞれの $N_i$ は、A、C、G及びTからなる群からランダムに選択されたヌクレオチドであり； $K_j$ は1～10の範囲の整数であり、それぞれの $j$ はM以下であり(すなわち、領域 $N_1 N_2 \dots N_{K_j}$ は可変領域である)； $i = 1, 2, \dots, L_j$ であるそれぞれの $b_i$ はヌクレオチドであり； $L_j$ は1～10の範囲の整数であり、それぞれの $j$ はM以下であり、すべての配列タグ(i)はすべての $j$ について同じ $K_j$ を有し、かつ(i i)すべての $j$ について同じ配列 $b_1 b_2 \dots b_{L_j}$ を有し(すなわち、領域 $b_1 b_2 \dots b_{L_j}$ は定常領域である)；かつMは2以上の整数である；(c) タグ-テンプレート複合体を増幅させる工程；(d) 増幅させたそれぞれのタグ-テンプレート複合体の配列リードを複数生成する工程；(e) 同一の配列タグを有する複数の配列リードのそれぞれの核酸位置において一致するヌクレオチドを決定することによりそれぞれの核酸のヌクレオチド配列を決定する工程、で使用することもできる。いくつかの実施形態では複数の配列リードは少なくとも $10^4$ 個あり；他の実施形態では、複数の配列リードは少なくとも $10^5$ 個あり；更に他の実施形態では、複数の配列リードは少なくとも $10^6$ 個ある。いくつかの実施形態では、上記の配列タグの全長は、ヌクレオチド15塩基～80塩基の範囲である。

#### 【0046】

##### 配列タグの付加

様々な異なる付加反応を利用して、上記のものに加え、サンプル中の実質的にすべてのクロノタイプに対し、固有のタグを付加することができる。配列タグを再構成核酸に付加するため、本発明の常軌(routine)改変には、例えば、マイクロアレイ法又はゲノム配列決定法においてサンプルの複雑さを低減するためのものなどの、サンプル核酸のサブセットを捕捉する多くの手法を使用することができる。以降の操作のため、配列タグの付加、配列決定、及び同様の方法など、標的とする核酸の様々なセットを捕捉させる手法の例としては、Willisら、米国特許第7,700,323号；Jonesら、米国特許出願公開第2005/0142577号；Gullbergら、米国特許出願公開第2005/0037356号；Porrecaら、Nature Methods, 4(11): 931～936(2007)；Turner et al, Nature Methods, 6(5): 315～316(2009)；Church, 米国特許第5,149,625号；及びMacevicz, 米国特許第8,137,936号が挙げられる。

#### 【0047】

一実施形態では、このような付加は、再構成された核酸分子(同様にして、クロノタイプ配列を含むもの)を含有しているサンプルを、配列タグ群又は配列タグのライブラリと組み合わせて、2群の分子をランダムに合わせかつ会合又は結合(例えば、共有結合)させることで成すことができる。例えば、このようなランダムな組み合わせは、タグを含有しているプライマーを標的核酸にアニーリングさせ、伸長させる、又はタグを含有してい

るアダプターを標的核酸の末端にライゲーションさせる生体分子反応において生じ得る。いくつかの実施形態では、タグを付加する方法は、DNA配列決定法のアプローチにある程度依存し得る。例えば、454社の配列決定法などの比較的長く正確な配列リードを生成する配列決定法において、例えば、Freeman et al, Genome Research, 19:1817~1824(2009)に開示される5'-RACEなどの従来法を用い、再構成核酸を含むmRNAからcDNAライブラリを作製した後、配列タグを含有しているアダプターを一端又は両端にライゲーションさせることにより、配列タグを付加することができる。他の実施形態では、比較的短く誤りが生じやすい配列リードを生成する「Illumina」社の配列決定又は「Ion Torrent」配列決定などの配列決定法を使用するとき、配列決定するアンプリコンが、その手法により生成される配列リードによりカバーされる長さを有するよう更なる工程が必要とされる場合もある。このようなタグを付加する反応において、クロノタイプの配列は直鎖一本鎖又は二本鎖ポリヌクレオチドを含み、配列タグは、PCRプライマーなどの増幅プライマー、ライゲーションアダプター、環状化可能なプローブ、プラスミド、又は均等物などの試薬により輸送することができる。配列タグ群の輸送を可能にするいくつかのこのような試薬はMacevicz, 米国特許第8,137,936号; Faham et al, 米国特許第7,862,999号; Landegren et al, 米国特許第8,053,188号; Unrau and Deugau, Gene, 145:163~169(1994); 並びにChurch, 米国特許第5,149,625号に開示されており、これらは参照により本願に援用される。

#### 【0048】

図2A及び2Bは、PCRを含む付加反応を例示する。配列タグ群( $T_1$ 、 $T_2$ 、 $T_3$ 、 $\dots$ 、 $T_j$ 、 $T_{j+i}$ 、 $\dots$ 、 $T_k$ 、 $T_{k+1}$ 、 $\dots$ 、 $T_{n-1}$ 、 $T_n$ )がプライマー(2100)に組み込まれる。配列タグ群は、再構成された核酸分子(2102)よりも十分大きなサイズを有する。配列タグは、核酸分子に対しプライマーをアニーリングさせること、及びPCRの第1サイクルにおいてDNAポリメラーゼによりプライマーを伸長させることにより再構成された核酸分子に付加される。図は、再構成された核酸分子の選別方法を示し、すなわち、共通のプライマー結合領域(2104)、例えばV領域(2108)をもとに、配列タグの全集団のうちの小画分を、ランダムにプライマーにアニーリングさせることによる方法例を示す。プライマー(したがって配列タグ)は、再構成された核酸配列分子に対し無作為に合わせられることから、同じ配列タグが異なる核酸分子に付加される可能性が僅かに存在するものの、配列タグ群が本明細書で教示される通りに大きいものである場合、この可能性は、無視可能な程度にまで小さくなり、実質的にすべての核酸分子に固有の配列タグが付加されることになる。順方向及び逆方向プライマー対の、その他のプライマー(2106)は、C領域(2110)にアニーリングし、複数回のアニーリング、伸長、及び融解サイクル後にアンプリコン(2112)が形成され、これらのサイクルにより、群のクロノタイプ(2102)を含むV(D)J領域に固有の配列タグが付加される。すなわち、アンプリコン(2112)は、付加反応由来のタグ-分子複合体を含む。

#### 【0049】

図2C及び2Dは、サンプル中のそれぞれの又は実質的に異なる再構成核酸に一对の配列タグを付加する方法を例示する。図2A及び2Bの方法と同じくして、配列タグ( $T_1$ 、 $T_2$ 、 $T_3$ 、 $\dots$ 、 $T_j$ 、 $T_{j+i}$ 、 $\dots$ 、 $T_k$ 、 $T_{k+1}$ 、 $\dots$ 、 $T_{n-1}$ 、 $T_n$ )を輸送するプライマー(2200)を下流プライマーとして使用し、更に、共通のプライマー(2106)を配列タグ( $T_m$ 、 $T_{m+1}$ 、 $T_{m+2}$ 、 $\dots$ 、 $T_q$ 、 $T_{q+1}$ 、 $T_{q+2}$ 、 $\dots$ 、 $T_r$ 、 $T_{r+1}$ 、 $T_{r+2}$ 、 $\dots$ 、 $T_s$ 、 $T_{s+1}$ 、 $T_{s+2}$ 、 $\dots$ )を輸送するプライマー(2206)に置き換え上流プライマーとして使用する。下流のプライマーセット同様、上流プライマー(2206)により輸送される異なる配列タグの数は、増幅後に実質的にすべての再構成された核酸(2202)が固有のタグを有するよう、再構成された核酸分子(2202)の数と比較して大きくしてもよい。いくつかの実施形態では、

プライマー（２２０６）及び（２２００）におけるそれぞれの配列タグのセットは、図２Ａ及び２Ｂの実施形態における配列タグのセットほど大きくする必要はない。異なる再構成核酸が配列タグ対により固有に標識されることから、異なる再構成核酸が対のうちの１つの配列タグを共有する場合にも、単一の再構成された核酸を標識する配列タグ対の実質的な固有性は損なわれない。したがって、図２Ｃ及び２Ｄの実施形態において、それぞれのプライマーセット（２２００）及び（２２０６）の配列タグは、プライマーセット（２１００）の配列タグよりも種類が少ない。例えば、ランダム配列タグを利用し、かつプライマー（２１００）に１６-mer配列タグを含有させる場合、プライマー（２２００）及び（２２０６）にはそれぞれ８-merの配列タグを含有させて、配列タグの総合的な多様性を同程度にすることができる。あるいは、図２Ｃ及び２Ｄの実施形態は、図２Ａ及び２Ｂと同様に実施する。プライマーを核酸分子にアニーリングさせ、かつＰＣＲの第１サイクルにおいてＤＮＡポリメラーゼによりプライマーを伸長させることにより、再構成された核酸分子に配列タグが付加される。上記の通り、図２Ｃは、再構成された核酸分子の選別方法を示し、すなわち、共通のプライマー結合部位（２２０４）及び（２２０５）、例えばそれぞれＶ領域（２２０８）及びＣ領域（２２１０）をもとに、配列タグ対の全集団のうちの小画分を、ランダムにプライマーにアニーリングさせることによる方法例を示す。プライマー（したがって配列タグ）は再構成された核酸配列分子に無作為に合わせられることから、同じ配列タグ対が異なる核酸分子に付加される可能性が僅かに存在するものの、配列タグ群が本明細書で教示される通りに大きいものである場合、この可能性は、無視可能な程度にまで小さくなり、実質的にすべての再構成された核酸分子に固有の配列タグが付加されることになる。複数回のアニーリング、伸長、及び融解サイクルの後アンプリコン（２２１２）が形成され、これらのサイクルにより、群のクロノタイプ（２２０２）を含むＶ（Ｄ）Ｊ領域に固有の配列タグが付加される。すなわち、アンプリコン（２２１２）は、付加反応に由来するタグ-分子複合体を含む。

#### 【００５０】

いくつかの実施形態では、例えば、Porreca et al（上掲）；Willis et al（上掲）；又は同様の参照文献により開示される技術的な常軌改変により、配列タグを捕捉し、かつ所望の再構成された核酸に付加するために、環状化可能なプローブを使用することができる。図２Ｅ及び２Ｆに示す通り、次の構成要素：上流の標的結合断片（２３０４）；５'-リン酸化末端（２３０５）を有する下流の標的結合断片（２３０６）；配列タグ（２３１０）；第２の共通のプライマー結合部位（２３１４）；任意選択的な切断部位（２３０８）；及び第１の共通のプライマー結合部位（２３１２）を含む環状化可能なプローブ（２３０２）が提供される。アニーリング条件下で、環状化可能なプローブ（２３０２）を、標的ポリヌクレオチド（例えば、一般的な手法によりｍＲＮＡから調製された一本鎖又は二本鎖ｃＤＮＡなどであってよい）を含有しているサンプル（２３００）を含む反応混合物と組み合わせる。図示する通り、標的ポリヌクレオチドは、IGH又はTCR鎖をコードしている再構成された核酸のＶ、NDN、Ｊ及びＣ領域を含む。いくつかの実施形態では、上流及び下流の標的結合断片（２３０４）及び（２３０６）の配列は、それぞれ、標的ポリヌクレオチドのＶＤＪ領域の部分にまたがるよう選択される。環状化可能なプローブ（２３０２）及び標的ポリヌクレオチド（２３００）は、上流及び下流の標的結合断片（２３０４及び２３０６）のアニーリングにより、反応混合物中で複合体（２３３０）を形成する。ＤＮＡポリメラーゼ及びdNTPの存在下で、上流の標的結合断片（２３０４）を、標的ポリヌクレオチドのＶＤＪ領域の部分を複製している（及びそれにより捕捉している）下流の標的結合断片（２３０６）まで伸長（２３４０）させる。伸長させた、上流の標的結合断片を、リガーゼ活性の存在下で下流の標的結合断片（２３０６）にライゲーションさせて、閉環した一本鎖ＤＮＡ環（２３４２）を形成させる。次に、任意選択的に、反応混合物をエキソヌクレアーゼにより処理（２３４４）して、未反応のプローブ及び標的ポリヌクレオチドを除去してもよい。いくつかの実施形態では、切断部位（２３０８）を切断して、一本鎖核酸環（２３４２）を開環した後（この切断部位は、例えば、稀な位置で切断するエンドヌクレアーゼの認識部位であって



よい)、又はRNAモノマーをプローブに挿入し、RNA分解酵素Hなどにより切断させた後、開環させたプローブ(2348)のVDJ-タグ挿入部分をプライマー(2350)及び(2352)により増幅させてもよい。プライマー(2350)及び(2352)には、以降のDNA配列決定(2354)を可能にする要素を追加するため、非相補的な領域が含まれていてもよい。あるいは、一本鎖核酸環を使用して、直接配列決定法、例えば、Drmanac et al, Science, 327(5961):78~81(2010);及び米国特許第8,445,196号などのためのナノボールテンプレートを生成してもよい。

#### 【0051】

図2Gは、免疫受容体分子をコードしている再構成された核酸に配列タグを付加するための別の実施形態を例示する。本実施形態を実施する際の指針は、参照により本願に援用されるFaham and Zheng, 米国特許第7,208,295号に見ることができる。反応混合物中、アニーリング条件下で、再構成された核酸(2450)をプローブ(2454)及びアダプター(2456)と組み合わせる。プローブ(2454)は、受容体特異的な部分(2455)及びアダプター特異的な部分(2457)を含む。例えば、プローブ(2454)がプローブ混合物を含んでもよく、ここで、異なるプローブは、異なるJ領域に特異的な、受容体特異的な部分を有し、又は他の実施形態では、異なるV領域に特異的な、受容体特異的な部分を有する。5'-リン酸化されているアダプター(2456)は、5'末端にプローブ特異的な部分(2458)と、配列タグ(2460)と、第1のプライマー結合部位(2462)とを含む。プローブ(2454)の、受容体特異的な部分(2455)及びアダプター特異的な部分(2457)と、プローブ特異的な部分(2458)の、位置、配列、及び長さは、それらが互いにハイブリダイゼーションして構造体(2452)を形成するよう選択する。構造体(2452)の形成後、再構成された核酸(2450)から一本鎖核酸部分(2461)を切断し、再構成された核酸(2450)の切断された3'末端をアダプター(2456)の5'リン酸化末端とライゲーションさせて、第1の伸長産物(2459)を形成させた後、プローブ(2454)を除去する(2474)。(2461)の切断は、Faham及びZhengにより記載される通り、一本鎖核酸ヌクレアーゼにより実施することもできる。一実施形態では、例えば、参照により本願に援用されるFaham et al, 米国特許第7,208,295号による教示の通り、例えば、dTTPをdUTPに置き換えたPCRにおいて、チミジンをウラシルに置き換えて、プローブ(2454)を合成し、並びにウラシル-DNAグリコシラーゼ(UDG)で処理してこのプローブを除去する。UDG処理により、ウラシルでプローブ(2454)が切断されて断片(2455)が得られる。プローブの切断後、アダプター及び突出部を除去し(2476)、伸長産物(2464)に順方向プライマー(2466)及び逆方向プライマー(2468)を加え、PCR(2470)を実施した後、得られるアンプリコンサンプルを配列決定する(2472)。

#### 【0052】

図2Gと類似の実施形態では、類似のプローブ及びアダプターを使用して、標的とするポリヌクレオチドの所定の部位に配列タグを付加することができる。ここでは、(2461)に対応する一本鎖核酸部分の開裂にはFEN-1などのフラップエンドヌクレアーゼが使用される。本実施形態において、更に異なるヌクレアーゼを使用する場合、複数のプローブ及びアダプター配列を逆転させる;すなわち、フラップエンドヌクレアーゼの基質には、標的配列(2450)に対するアニーリングに対応するアダプターの3'末端(2454)と、標的配列の5'末端(2452)に対応する一本鎖核酸部分とが必要とされる。プローブ配列の切断及び除去後、以降の工程は実質的に同時に行われる。検出アッセイにおいてフラップエンドヌクレアーゼを利用するための指針は、以下の参考文献に見ることができる:Lyamichiev et al, Nature Biotechnology, 17:292~296(1999);及びEis et al, Nature Biotechnology, 19:673~676(2001)など。

#### 【0053】

いくつかの実施形態では、再構成された核酸は免疫受容体分子鎖をコードし、この分子鎖は、典型的には、非常に類似したポリヌクレオチドの非常に大規模（例えば、1000超10,000未満、更に、通常、100,000~1,000,000、又はそれ以上）のセットを含み得る免疫レパトアを形成し、かつこの分子鎖は、ヌクレオチド500塩基未満の長さを有してよく、又は他の実施形態では、ヌクレオチド400塩基未満、又は更に他の実施形態では、ヌクレオチド300塩基未満の長さを有してよい。本発明の一態様では、本発明者らは、これらの特徴こそが、非常に異なる配列タグを利用し、非常に類似するクロノタイプの配列リードを効率的に比較して、それらが同じ配列に由来するものかどうかを判定することを可能にしたと認識しており、かつそのように理解している。

【0054】

10

サンプル：

用語「サンプル（sample）」は、ある程度の量の生体材料を指し、いくつかの実施形態では、この生体材料は患者から得られ、細胞及び/又は無細胞系DNAを含有する；すなわち、本用語は「検体」又は「組織サンプル」とも言い換えられる。用語「標本（sample）」は、例えば、再構成された核酸の大規模セット又は多量のサブセット又は部分を得るに当たって統計的な観点から使用される場合もある；特に、用語「標本」の統計上の利用は、「代表標本」を意味するものと理解することもでき、このような標本は、（例えば）組織中の異なる核酸の相対頻度を反映する又は概算するものと理解される。当業者であれば、文脈からこの用語の適切な用法を認識することができるであろう。

【0055】

20

クロノタイププロファイルは、免疫細胞、又は免疫受容体鎖をコードしている無細胞核酸を含有している血液などの体液サンプルから得ることができる。例えば、免疫細胞は、T細胞及び/又はB細胞を含み得る。T細胞（Tリンパ球）としては、例えば、T細胞受容体を発現している細胞が挙げられる。T細胞としては、ヘルパーT細胞（エフェクターT細胞又はTh細胞）、細胞傷害性T細胞（CTL）、記憶T細胞、及び制御性T細胞が挙げられる。一態様では、T細胞サンプルには、少なくとも1,000個のT細胞が含有され；より典型的には、サンプルには少なくとも10,000個のT細胞が含有され、更に典型的には、少なくとも100,000個のT細胞が含有される。別の態様では、サンプルは、1000~1,000,000個の範囲で多数のT細胞を含有する。免疫細胞サンプルは、B細胞も含み得る。B細胞としては、例えば、形質B細胞、メモリーB細胞、B1細胞、B2細胞、辺縁帯B細胞、及び濾胞性B細胞が挙げられる。B細胞は、免疫グロブリン（抗体、B細胞受容体）を発現することができるものである。上記の通り一態様では、B細胞サンプルには、少なくとも1,000個のB細胞が含有され；より典型的には、サンプルには少なくとも10,000個のB細胞が含有され、更に典型的には、少なくとも100,000個のB細胞が含有される。別の態様では、サンプルは、1000~1,000,000個の範囲で多数のB細胞を含有する。

30

【0056】

本発明の方法で使用するサンプルは、例えば、腫瘍組織、血液、及び血漿、リンパ液、脳及び脊髄を取り囲む脳脊髄液、及び関節を取り囲む滑液などの、様々な組織より採取することができる。一実施形態では、サンプルは血液サンプルである。血液サンプルは、約0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.5、2.0、2.5、3.0、3.5、4.0、4.5、又は5.0mLとすることができる。サンプルは腫瘍生検とすることができる。生検は、例えば、脳、肝臓、肺、心臓、大腸、腎臓、又は骨髄に由来するものとして行うことができる。対象者からサンプルを単離するのに当たり、当業者により用いられる任意の生検法を使用することができる。例えば、生検は、全身麻酔医が行う直視下生検とすることができる。生検は、直視下生検で用いるものよりも小さな切開で行う経皮生検とすることもできる。生検は、組織が部分的に切除されたコア生検又は切開生検とすることもできる。生検は、病変部位の完全な切除を試みる切除生検とすることもできる。生検は、組織又は体液サンプルを針により回収する微細針吸引生検とすることもできる。

40

50

## 【0057】

いくつかの実施形態では、本発明の方法のクロノタイププロファイルは、診断用試料の場合には腫瘍又は末梢血から生成され、あるいは残存病変をモニタリングするためのサンプルの場合には末梢血から生成される。リンパ増殖性疾患又は骨髄増殖性疾患などの疾患と関連する1つ以上のクロノタイプは、診断用試料から診断する。通常、リンパ増殖性疾患又は骨髄増殖性疾患に関連する1つ以上のクロノタイプは、クロノタイププロファイルにおいて最も高頻度なものとして存在する。いくつかの場合では、単一に相関するクロノタイプが存在し、他の場合では、リンパ増殖性疾患又は骨髄増殖性疾患と相関する複数のクロノタイプが存在し得る。腫瘍サンプルは、このような疾患から影響を受けている、リンパ管又はリンパ系外部のその他の組織を包含する任意の組織から採取することができる。上記の通り、残存病変をモニタリングするためのクロノタイププロファイルは、末梢血から抽出した核酸サンプルから生成することもできる。サンプルの核酸は、末梢血の細胞含有画分のB細胞に由来するものであってよく、あるいは末梢血の無細胞核分、例えば、血漿又は血清に由来するものであってもよい。一実施形態では、末梢血サンプルには、少なくとも1,000個のB細胞が含有され；より典型的には、かかるサンプルには少なくとも10,000個のB細胞が含有され、更に典型的には、少なくとも100,000個のB細胞が含有される。別の態様では、サンプルは、1000~1,000,000個の範囲で多数のB細胞を含有する。いくつかの実施形態では、サンプル中の細胞数が測定感度の限界を設定する。すなわち、多量の末梢血サンプルを使用することで、残渣病変の検出感度がより高くなる。例えば、1,000個のB細胞を含有するサンプルの場合、このような細胞のDNAを配列決定により解析するときに取得される配列決定リード数とは関係なく、クロノタイプの低頻度検出限界は1/1000すなわち0.001である。サンプルの核酸は、末梢血の細胞含有画分のT細胞に由来するものであってよく、あるいは末梢血の無細胞核分、例えば、血漿又は血清に由来するものであってもよい。一実施形態では、末梢血サンプルには、少なくとも1,000個のT細胞が含有され；より典型的には、かかるサンプルには少なくとも10,000個のT細胞が含有され、更に典型的には、少なくとも100,000個のT細胞が含有される。別の態様では、サンプルは、1000~1,000,000個の範囲で多数のT細胞を含有する。いくつかの実施形態では、サンプル中の細胞数が測定感度の限界を設定する。すなわち、多量の末梢血サンプルを使用することで、残渣病変の検出感度がより高くなる。例えば、1,000個のT細胞を含有するサンプルの場合、このような細胞のDNAを配列決定により解析するときに取得される配列決定リード数とは関係なく、クロノタイプの低頻度検出限界は1/1000すなわち0.001である。

## 【0058】

本発明に使用するサンプルは、DNA（例えば、ゲノムDNA）又はRNA（例えば、メッセンジャーRNA）を含み得る。核酸は無細胞系DNA又はRNAであってよく、例えば、循環系から抽出される[Vlassov et al, CUIT. Mol. Med., 10:142~165(2010); Swarup et al, FEBS Lett., 581:795~799(2007)]。本発明で提供される方法では、対象に由来し、解析され得るRNA又はDNAの量は、例えば、いくつかの用途では、可能な限り少量で細胞1個であり（例えば、その他の細胞選別基準、例えば、形態学的基準による校正試験）、並びに細胞 $10^6$ 個以上もの量であり、これは、DNA量に換算すると6pg~60ugの範囲であり、RNA量に換算すると1pg~10ugの範囲である。いくつかの実施形態では、核酸サンプルは、6pg~60ugのDNAサンプルである。他の実施形態では、核酸サンプルは、100μL~10mLの末梢血由来のDNAサンプルである；他の実施形態では、核酸サンプルは、100μL~10mLの末梢血の無細胞核分由来のDNAサンプルである。

## 【0059】

いくつかの実施形態では、リンパ球又は無細胞核酸のサンプルは、異なるクロノタイプを有する実質的にすべてのB細胞又はT細胞を提示させ、それをもとにクロノタイプの「

レパトア」を生成するべく、十分に多量である。一実施形態では、異なるすべてのクロノタイプを実質的に表現させるため、0.001%以上の頻度で存在しているそれぞれの集団のクロノタイプが99%の確率で含有されるようサンプルを採取する。別の実施形態では、0.0001%以上の頻度で存在しているそれぞれの集団のクロノタイプが99%の確率で含有されるようサンプルを採取する。並びに、別の実施形態では、0.00001%以上の頻度で存在しているそれぞれの集団のクロノタイプが99%の確率で含有されるようサンプルを採取する。一実施形態では、B細胞又はT細胞サンプルは、少なくとも $5 \times 10^5$ 個の細胞を含有し、別の実施形態では、かかるサンプルは少なくとも $1 \times 10^6$ 個の細胞を含有する。

#### 【0060】

核酸サンプルは、従来法、例えば、Innisら編、PCRプロトコール (Academic Press, 1990) などを用い、末梢血から得ることができる。例えば、白血球細胞は、従来法、例えば、RosetteSepキット (Stem Cell Technologies, Vancouver, Canada) を用い、血液サンプルから分離することができる。血液サンプルは100 $\mu$ L~10mLの範囲の用量とすることができ；一態様では、血液サンプル用量は、100 $\mu$ L~2mLの範囲である。次に、本発明の方法に使用するため、例えば、DNeasy Blood & Tissueキット (Qiagen, Valencia, CA) などの従来法を用い、このような血液サンプルからDNA及び/又はRNAを抽出することができる。任意選択的に、例えば、蛍光表示式細胞分取 (FACS) (Becton Dickinson, San Jose, CA)、磁気細胞分離装置 (MACS) (Miltenyi Biotec, Auburn, CA) などの従来法を用い、白血球細胞、例えば、リンパ球のサブセットを更に単離することもできる。例えば、メモリーB細胞は、表面マーカーCD19及びCD27をもとに単離することもできる。

#### 【0061】

無細胞系DNAは、従来法、例えば、参照により本願に援用される、Lo et al, 米国特許第6,258,540号；Huang et al, Methods Mol. Biol., 444:203~208 (2008) などを用い末梢血サンプルから抽出することもできる。非限定例として、EDTAを入れた採血管に末梢血を採取した後、遠心分離により、血漿成分、白血球細胞成分、及び赤血球細胞成分に分画する。無細胞血漿画分 (例えば、0.5~2.0mL) のDNAは、QIAamp DNA Blood Miniキット (Qiagen, Valencia, CA) などのキットを用い、製造元の使用説明書にしたがって抽出することができる。

#### 【0062】

一態様では、クロノタイププロファイルを生成するためのリンパ球サンプルは、異なるクロノタイプを有する実質的にすべてのT細胞又はB細胞を提示させるべく十分に多量である。一実施形態では、0.001%以上の頻度で存在しているそれぞれの集団のクロノタイプが99%の確率で含有されるようサンプルを採取する。別の実施形態では、0.0001%以上の頻度で存在しているそれぞれの集団のクロノタイプが99%の確率で含有されるようサンプルを採取する。別の実施形態では、0.00001%以上の頻度で存在しているそれぞれの集団のクロノタイプが99%の確率で含有されるようサンプルを採取する。他の実施形態では、0.001%以上の頻度で存在しているそれぞれの集団のクロノタイプが95%の確率で含有されるようサンプルを採取する。別の実施形態では、0.0001%以上の頻度で存在しているそれぞれの集団のクロノタイプが95%の確率で含有されるようサンプルを採取する。別の実施形態では、0.00001%以上の頻度で存在しているそれぞれの集団のクロノタイプが95%の確率で含有されるようサンプルを採取する。更に別の実施形態では、B細胞又はT細胞サンプルは、少なくとも $5 \times 10^5$ 個の細胞を含有し、別の実施形態では、かかるサンプルは少なくとも $1 \times 10^6$ 個の細胞を含有する。

#### 【0063】

サンプルが回収される材料の供給源が乏しい場合、例えば、臨床検体などの場合、全ゲノム増幅 (WGA)、多置換増幅 (MDA)；又は同様の手法、例えば、Hawkins et al, Curr. Opin. Biotech., 13: 65~67 (2002)；Dean et al, Genome Research, 11: 1095~1099 (2001)；Wang et al, Nucleic Acids Research, 32: e76 (2004)；並びに Hosono et al, Genome Research, 13: 954~964 (2003) など、バイアスのかからない手法により材料由来のDNAを増幅してもよい。

#### 【0064】

血液サンプルは、特定の対象のものであり、従来法、例えば、Innisら編, PCR Protocols (Academic Press, 1990) などを用い、得ることができる。例えば、白血球細胞は、従来法、例えば、RosetteSepキット (Stem Cell Technologies, Vancouver, Canada) を用い、血液サンプルから分離することができる。血液サンプルは100 $\mu$ L~10mLの範囲の用量とすることができ；一態様では、血液サンプル用量は、100 $\mu$ L~2mLの範囲である。次に、本発明の方法に使用するため、例えば、DNeasy Blood & Tissueキット (Qiagen, Valencia, CA) などの従来法を用い、このような血液サンプルからDNA及び/又はRNAを抽出することができる。任意選択的に、例えば、蛍光表示式細胞分取 (FACS) (Becton Dickinson, San Jose, CA)、磁気細胞分離装置 (MACS) (Miltenyi Biotec, Auburn, CA) などの従来法を用い、白血球細胞、例えば、リンパ球のサブセットを更に単離することもできる。

#### 【0065】

それぞれの個体の獲得免疫細胞のDNA並びにそれらに関連するRNA転写物には再構成が存在することが特定されており、RNA又はDNAのいずれの場合も、本発明に提供する方法で配列決定することができる。T細胞受容体鎖若しくは免疫グロブリン分子鎖、又はそれらの部分をコードしているT細胞又はB細胞由来の再構成配列は、クロノタイプとして参照される。DNA又はRNAは、抗体をコードするT細胞受容体 (TCR) 遺伝子又は免疫グロブリン (Ig) 遺伝子由来の配列に対応し得る。例えば、DNA及びRNAは、TCRの 、 、 又は 鎖をコードし得る配列に対応し得る。T細胞のほとんどでは、TCRは 鎖及び 鎖から構成されるヘテロ二量体である。TCR 鎖はVJ再構成により生成され、 鎖受容体はV(D)J再構成により生成される。TCR 鎖に関し、ヒトではV断片が48、D断片が2、及びJ断片が13存在する。2つのジャンクションのそれぞれにおいて、いくつかの塩基を欠失させ、かつその他の塩基を追加する (N及びPヌクレオチドと呼ばれる) こともできる。T細胞のうち、小数のものでは、TCRは 鎖及び 鎖から構成される。TCR 鎖はVJ再構成により生成され、TCR 鎖はV(D)J再構成により生成される (Kenneth Murphy, Paul Travers, and Mark Walport, Janeway's Immunology 7th edition, Garland Science, 2007、該文献は参照によりその全体が本願に援用される)。

#### 【0066】

本発明の方法において解析するDNA及びRNAは、定常領域 ( 、 、 、 又は  $\mu$ ) を有する重鎖免疫グロブリン (IgH) 又は定常領域 又はKを有する軽鎖免疫グロブリン (IgK又はIgL) をコードしている配列に対応し得る。それぞれの抗体が、2つの同一の軽鎖及び2つの同一の重鎖を有する。それぞれの鎖は、定常 (C) 領域及び可変領域から構成される。重鎖に関し、可変領域は可変 (V) セグメント、多様性 (D) セグメント、及び連結 (J) セグメントから構成される。ゲノム中には、これらのセグメントのうち各種類のものをコードしている別個の複数の配列が存在する。B細胞の分化中に特異的なVDJ再構成イベントが生じることにより、細胞が特異的な重鎖を生成するようになる。D領域が存在せず、VJ再構成のみが存在することを除き、同様の方法で軽鎖に

10

20

30

40

50

多様性が生じる。多くの場合、再構成部位に近接して体細胞変異が生じ、これによりいくつかのヌクレオチドが付加又は欠失され、B細胞により生成される重鎖及び軽鎖の多様性が更に増す。また、B細胞により生成される抗体の多様性は、重鎖及び軽鎖が異なることによっても生じ得る。重鎖及び軽鎖の可変領域が、抗原認識（又は結合）領域又は部位の形成に参与する。この多様性に加え、いくつかのエピトープに対し生じる特定の応答後に、体細胞超変異のプロセスが生じ得る。

#### 【0067】

上記の通り、本発明に従い、リンパ球に由来する再構成された核酸又は血液などの組織に由来する無細胞核酸部分を含有するアンプリコンを生成するようプライマーを選択することもできる。本明細書では、このような部分を、「体細胞変異により再構成された領域」としても参照し得る。体細胞変異により再構成された領域が、分化中のリンパ球又は完全に分化したリンパ球由来の核酸を含む場合もあり、ここで、分化中のリンパ球は、分子（例えば完全なV(D)J領域を有するもの）を形成する免疫遺伝子の再構成が完了していない細胞を指す。例えば、体細胞変異が完了していない再構成領域は、不完全なIgH分子（D-J領域のみを含有する分子など）、不完全なTCR分子（D-J領域のみを含有する分子など）、及び不活性なIgK（例えば、Kdε-V領域を含む）を包含する。

#### 【0068】

##### 核酸群の増幅

いくつかの実施形態では、第1及び第2のプライマーセットのプライマー配列は、従来の多重ポリメラーゼ連鎖反応（PCR）にしたがって選択することもできる。例えば、プライマーの選択、及び様々な免疫受容体鎖をコードしている核酸の多重PCRの実施についての指針は、以下の参照文献に見ることができ、これらの文献は参照により援用される：Faham及びWillis、米国特許第8,236,503号及び同第8,628,927号；Morley、米国特許第5,296,351号；Gorski、米国特許第5,837,447号；Dau、米国特許第6,087,096号；Van Dongen et al、米国特許広報第2006/0234234号；欧州特許公報第1544308B1号；並びにVan Dongen et al, Leukemia, 17:2257~2317（2003）など。多重PCRについての指針は、Henegariu et al, BioTechniques, 23:504~511（1997）などの参照文献に見ることがもできる。いくつかの実施形態では、最終産物において増幅されている配列頻度が、出発反応混合物中の配列頻度と実質的に同じになるよう、プライマーが選択される。このようなプライマーの選択には、プライマー長、プライマー結合部位、及びプライマー濃度の選択も包含され得る。上記の通り、配列リードの作成に選択される方法及び付加される配列タグに応じ、多重レベルは多様に変化し得る。

#### 【0069】

いくつかの実施形態では、標的核酸を増幅させる工程は、例えば、1セットのプライマー（例えば、第1セットの「上流」又は「下流」プライマー）をアニーリングさせ、このプライマーを伸長させ、伸長させた核酸鎖をテンプレートから融解させるサイクルを繰り返すことにより、多量の伸長鎖をサイクル数の一次関数として増幅させるなどといった、標的核酸の線形的増幅を含む。言い換えると、増幅工程は、1セットのプライマーを繰り返し伸長させることにより、標的ポリヌクレオチド（すなわち少なくとも1つの標的ポリヌクレオチド鎖）を複製することを含む。いくつかの実施形態では、このような一方向に単回又は繰り返し伸長させる工程後に、伸長していないプライマーを除去する工程と、別のセットのプライマー（例えば、第2セットの「下流」又は「逆転」プライマー）を別方向に単回又は繰り返し伸長させる工程とを行うこともできる。

#### 【0070】

第1セットのプライマー及び第2セットのプライマー中のプライマー数は、アッセイにおいて増幅される免疫受容体鎖核酸の数及び種類に応じて様々に変更することができる。いくつかの実施形態では、様々な鎖に対してコンセンサスプライマーを使用してもよい。

10

20

30

40

50

他の実施形態では、特異的なプライマーは、増幅させるそれぞれの異なる標的ポリヌクレオチド用に設計することもできる。通常、プライマーの第1セット及び第2セットは、両方共、それぞれ複数のプライマーを含む。いくつかの実施形態では、第1セット又は第2セットのプライマー中の複数のプライマーは、少なくとも50プライマーであり；他の実施形態では、第1セット又は第2セットのプライマー中の複数のプライマーは、少なくとも100プライマーであり；他の実施形態では、第1セット又は第2セットのプライマー中の複数のプライマーは、少なくとも150プライマーであり；他の実施形態では、第1セット又は第2セットのプライマー中の複数のプライマーは、少なくとも200プライマーであり；他の実施形態では、第1セット又は第2セットのプライマー中の複数のプライマーは、少なくとも250プライマーである。第1セットのプライマー中のプライマー数は、第2セットのプライマー中のプライマー数と同じであっても異なってもよい。

10

#### 【0071】

いくつかの実施形態では、第1セットのプライマー及び第2セットのプライマーは、クロノタイプが少なくともヌクレオチド30塩基の長さになるよう選択され；他の実施形態では、第1セットのプライマー及び第2セットのプライマーは、クロノタイプがヌクレオチド30～500塩基の範囲の長さになるよう選択され；他の実施形態では、第1セットのプライマー及び第2セットのプライマーは、クロノタイプがヌクレオチド30～400塩基の範囲の長さになるよう選択され；他の実施形態では、第1セットのプライマー及び第2セットのプライマーは、クロノタイプがヌクレオチド30～300塩基の範囲の長さになるよう選択され；他の実施形態では、第1セットのプライマー及び第2セットのプライマーは、クロノタイプがヌクレオチド30～200塩基の範囲の長さになるよう選択される。

20

#### 【0072】

PCR増幅プロトコルの例は、van Dongen et al, Leukemia, 17:2257～2317(2003)又はvan Dongen et al、米国特許公報2006/0234234号に見ることができ、これらの文献は参照により援用される。簡単に、プロトコル例としては以下のものがある：反応緩衝液：ABI Buffer II又はABI Gold Buffer(Life Technologies, San Diego, CA)；最終用量50μL；サンプルDNA 100ng；各プライマー10pmol(以下の通り、増幅を調節するため調整する)；dNTP最終濃度200μM；MgCl<sub>2</sub>最終濃度1.5mM(標的配列及びポリメラーゼに応じ最適化させる)；Taqポリメラーゼ(1～2U/チューブ)；サイクル条件：95で7分の熱変性；60でアニーリング；サイクル数：変性30秒；アニーリング30秒；伸長30秒。本発明の方法における、増幅に使用することのできるポリメラーゼは、市販のものであってよく、例えば、Taqポリメラーゼ、AccuPrimeポリメラーゼ、又はPfuが挙げられる。使用するポリメラーゼは、忠実度又は効率の好ましさをもとに選択することができる。

30

#### 【0073】

初期工程において、リアルタイムPCR、ピコグリーン染色、ナノスケールで流体を用いる電気泳動(例えば、LabChip)又は紫外線吸収測定を用いることで、サンプル中の増幅可能な材料の機能量を判断することができる。

40

#### 【0074】

一態様では、初期群中の配列の相対量が、増幅させた群又はアンプリコンにおけるものと実質的に等しくなるよう、本発明の多重増幅を実施する。すなわち、サンプル群に含まれるメンバー配列間の増幅バイアスが最小限に抑えられるよう多重増幅を実施する。一実施形態では、アンプリコン中のそれぞれの相対量が、初期サンプルにおける値の5倍以内に収まる場合、このような相対量は実質的に等しい。別の実施形態では、アンプリコン中のそれぞれの相対量が、初期サンプルにおける値の2倍以内に収まる場合、このような相対量は実質的に等しい。以下に詳細に開示される通り、PCRにおける増幅バイアスは、従来法を用い検出及び補正することができ、そのため、PCRプライマーのセットは、任

50

意のサンプルについてバイアスのかかっていない増幅をもたらすよう、所定のレパトアに関し選択することができる。

【 0 0 7 5 】

いくつかの実施形態では、増幅バイアスは、2段階増幅（例えば、上記に引用した、Faham及びWillisに記載の通り）を実施することにより回避することができ、ここで、第1又は一次段階では、標的配列に対し非相補的なテールを有するプライマーを用い、増幅サイクルを小数回（例えば、2～5、又は2～10、又は2～15サイクル）実施する。テールは、一次アンプリコンの配列末端に付加されるプライマー結合部位を含むものであり、かかる結合部位を、単一の順方向プライマー及び単一の逆方向プライマーのみを用いる第2段階の増幅に用いることにより、増幅バイアスの主因が排除される。第2段階の増幅を開始する前に、第1段階で伸長していないプライマーを反応混合物から除去し、あるいは失活させる。いくつかの実施形態では、一次PCRは、プライマーが異なることにより増幅に生じる差を最小限に抑えるべく、サイクル数が十分に少ない（例えば、2～10サイクル）。次に、差次的な増幅の発生を排除すべく、一对のプライマーにより二次増幅を実施する。いくつかの実施形態では、反应用量の数%又は数部、例えば、一次PCRの反應用量の1%を、二次PCRの反応混合物に直接用いる。いくつかの実施形態では、第一段階の増幅及び第二段階の増幅に、合計少なくとも35サイクルが割り当てられる。

【 0 0 7 6 】

いくつかの実施形態では、内部標準を組み合わせ、再構成された核酸のサンプルと同じ反応で増幅させることができる。内部標準は、配列が既知でありかつ濃度が既知である核酸とする。例えば、内部標準は、免疫受容体鎖の部分をコードしている天然の核酸をクローン化した複製物であってよく、又は合成核酸であってよい。いくつかの実施形態では、内部標準の長さ及び塩基組成は、増幅させる具体的な免疫受容体鎖を再現するよう選択する。増幅後の内部標準の相対濃度における変化をモニタリングすることにより増幅バイアスを検出することができ、バイアスのない増幅条件を決定することができる。例えば、プライマー長、位置、及び濃度を変更させて、増幅産物のバイアスを最小限に抑えることもできる。いくつかの実施形態では、複数の内部標準を反応に使用し；いくつかの実施形態では、2～50種の異なる内部標準を反応に使用し；他の実施形態では、2～25種の異なる内部標準を反応に使用し；いくつかの実施形態では、2～10種の異なる内部標準を反応に使用する。いくつかの実施形態では、増幅産物中の、異なる標的ヌクレオチド（例えば、すべての又は抜粋したクロノタイプ又は内部標準）の配列の相対頻度を測定することにより増幅バイアスを判断する。他の実施形態では、2種以上の内部標準など、選択した核酸を定量的リアルタイムPCRすることにより、増幅バイアスの存在、非存在、又はレベルを求める。内部標準を使用して、元のサンプルに含まれる異なるクロノタイプ数を定量することもできる。このような分子定量法はよく知られており、例えば、Brennert et al、米国特許第7,537,897号が挙げられる。この文献は、参照により本明細書に援用される。

【 0 0 7 7 】

配列リードの生成

本発明の方法には、任意のハイスループットな核酸配列決定法を使用することができる。好ましくは、このような方法には、費用効率が高い方法で大量の配列データを生成する能力があり、これにより、少なくとも1000種のクロノタイプを決定することができ、好ましくは、少なくとも10,000～1,000,000種のクロノタイプを生成することができる。DNA配列決定法としては、標識した末端又はプライマーと、スラブゲル及びキャピラリーゲル分離を用いる古典的なジデオキシ法（サンガー法）、可逆的に末端標識したヌクレオチドを用いる合成による配列決定法、ピロシーケンス、454シーケンス、標識したオリゴヌクレオチドプローブのライブラリに対する対立遺伝子特異的なハイブリダイゼーション、標識したクローンライブラリに対する対立遺伝子特異的なハイブリダイゼーション及びそれに続くライゲーションを用いる合成による配列決定、重合工程中



に組み込まれる標識ヌクレオチドのリアルタイムモニタリング、ポロニー (polony) 配列決定法、及び SOLID 配列決定法が挙げられる。近年では、ポリメラーゼ又はリガーゼを用いる連続的又は単回の伸長反応、並びにブロープライブラリを用いる単回又は連続的な差次的ハイブリダイゼーションにより、分離した分子を配列決定することが実証されている。これらの反応は、これまでも数多くのクローン配列に対して同時に実施されており、現行の商業用途では、平行して  $100 \times 10^6$  の配列で実証されているものなどが挙げられる。したがって、T細胞受容体 (TCR) 及び/又はB細胞受容体 (BCR) のレパトアを研究するに当たって、これらの配列決定アプローチを使用することができる。本発明の一態様では、同時に配列決定される固相表面上で個々の分子を空間的に分離する工程を含む、ハイスループットな配列決定法を利用する。このような固相表面は、非多孔性表面 [Solexa シーケンシング、例えば、Bentley et al, Nature, 456: 53 ~ 59 (2008) 又は Complete Genomics シーケンシング、例えば、Drmanac et al, Science, 327: 78 ~ 81 (2010)]、ビーズ又は粒子を結合させたテンプレートを含み得る、ウェルのアレイ [454 社によるものなど、例えば、Margulies et al, Nature, 437: 376 ~ 380 (2005) 又は Ion Torrent シーケンシング、米国特許公報第 2010/0137143 号又は同第 2010/0304982 号]、微細加工膜 [SMRT シーケンシング、例えば、Eid et al, Science, 323: 133 ~ 138 (2009)]、又はビーズアレイ (SOLID シーケンシング又はポロニーシーケンシング、例えば、Kim et al, Science, 316: 1481 ~ 1414 (2007)] を含み得る。別の態様では、このような方法は、固相表面でそれぞれを空間的に分離する前又はその後に単離分子を増幅させることを含む。増幅の前には、エマルジョン PCR などのエマルジョン系増幅、又はローリングサークル型の増幅を含ませることもできる。

#### 【0078】

個々のテンプレート分子が固相表面上で空間的に分離されている Solexa 系シーケンシングなど、可逆化ターミネーターを用いる合成によるシーケンシング後、それらの分子をブリッジ PCR により同時に増幅させて、別個のクローン群、すなわちクラスターを形成させた後、配列決定する、Bentley et al (上掲) 及び製造元の使用説明書 (例えば、TruSeq (商標) サンプル調製キット及びデータシート [Illumina 社, San Diego, CA, 2010] における記載; 並びに、以下の参照: 米国特許第 6,090,592 号; 同第 6,300,070 号; 同第 7,115,400 号; 及び欧州特許第 EP0972081B1 号 (これらの文献は参照により援用する)) における記載を用いるアプローチを特に対象とする。一実施形態では、固相表面に配置及び増幅させた各分子は、少なくとも  $10^5 / \text{cm}^2$  の密度; 又は少なくとも  $5 \times 10^5 / \text{cm}^2$  の密度; 又は少なくとも  $10^6 / \text{cm}^2$  の密度でクラスターを形成する。Solexa 系シーケンシングは、クラスター中の同じ標的配列 (又はテンプレート) から 2 つの配列リードを生成し、それぞれの標的配列の反対側の末端から 1 つの配列リードを生成する性能も提供する。いくつかの実施形態では、このような配列リード対を組み合わせて、一連の解析において単一の配列リードとして処理することもでき、又はこのような対を別個に、ただし、それらが同じクラスターに由来することを考慮に入れて、処理することもできる。場合により、同じテンプレートに由来する配列リード対は「メートペア (mate pair)」として言及し、テンプレートの両端からの配列決定プロセスは「二方向」配列決定法として言及する。いくつかの実施形態では、可逆的に末端標識したヌクレオチドを用いる合成により配列決定する工程には、テンプレートのそれぞれのクラスター又はクローン群に対する単一の配列リード、及びテンプレートのそれぞれのクラスター又はクローン群に対する複数の配列リード (限定するものではないが、メートペアが挙げられる) の生成が包含される。尚更なる実施形態では、テンプレートのそれぞれのクラスター又はクローン群について複数の配列リードを生成するとき、このような複数の配列リードを組み合わせて、結合工程などの以降の解析で使用するのに有効な、単一の配列リードを形成するこ

10

20

30

40

50

ともできる。

【0079】

一態様では、個人に由来するサンプルの配列に基づくクロノタイププロファイルは、以下の工程：(a) 個人のT細胞及び/又はB細胞に由来する核酸サンプルを得る工程；(b) このような核酸サンプルに由来する個々の分子を空間的に分離する工程（個々の分子は、サンプル中の核酸から生成された少なくとも1つのテンプレートを含み、このテンプレートは、体細胞変異により再構成された領域又はそれらの部分を含み、それぞれの個々の分子は、少なくとも1つの配列リードを生成し得るものである）；(c) この空間的に分離させた個々の分子を配列決定する工程；並びに(d) 核酸サンプル由来の各配列の核酸分子の豊富さを評価して、クロノタイププロファイルを生成する工程、を用い得られる。一実施形態では、体細胞変異により再構成された領域は、それぞれ、V領域及びJ領域を含む。別の実施形態では、配列決定する工程は、求められたそれぞれのクロノタイプについて複数の配列リードを生成する工程を包含する。更に他の実施形態では、配列決定する工程は、複数の配列リード由来の情報又はデータを組み合わせて、それぞれのクロノタイプを生成する工程を包含する。いくつかの実施形態では、このような組み合わせる工程は、Faham及びWillis、米国特許第8,628,927号（この教示は参照により本願に援用される）に記載の通りに配列リードを結合させることにより、あるいはFaham et al, 米国特許公報2013/0236895A1号（この教示は参照により本願に援用される）に記載の通りに配列タグを用いることにより実施することもできる。別の実施形態では、配列決定する工程は、空間的に分離させた個々の分子のそれぞれを二方向的に配列決定して、少なくとも1つの順方向配列リード及び少なくとも1つの逆方向配列リードを生成することを含む。

【0080】

更に、後者の実施形態では、少なくとも1つの順方向配列リード及び少なくとも1つの逆方向配列リードは重複領域を有し、このような重複領域の塩基は、このような配列リード間の逆方向に相補的な関係をもとに決定される。更に別の実施形態では、体細胞変異により再構成された領域のそれぞれは、V領域及びJ領域を含み、かつ配列決定する工程は、J領域中の部分から開始し、会合するV領域の方向に伸長する1つ以上の順方向配列リード及び少なくとも1つの逆方向配列リードをもとに、個々の核酸分子のそれぞれの配列を決定する工程を更に含む。別の実施形態では、個々の分子は、完全なIGH分子、不完全なIGH分子、完全なIGK分子（complete）、失活しているIGK分子、TCR分子、TCR分子、完全なTCR分子、及び不完全なTCR分子からなる群から選択される核酸を含む。別の実施形態では、配列決定する工程は、クオリティスコアが単調減少する配列リードを生成する工程を含む。別の実施形態では、上記の方法は、以下の工程：(a) 個人のT細胞及び/又はB細胞に由来する核酸サンプルを得る工程；(b) このような核酸サンプルに由来する個々の分子を空間的に分離する工程（個々の分子は、サンプル中の核酸からそれぞれ生成されるテンプレートネステッドセットを含み、かつ体細胞変異により再構成された領域又はそれらの部分をそれぞれ含み、それぞれのネステッドセットは、それぞれ同じ方向に伸長しかつネステッドセットが生成された核酸上の異なる部分から始まる複数の配列リードを産生し得るものである）；(c) この空間的に分離させた個々の分子を配列決定する工程；並びに(d) 核酸サンプル由来の各配列の核酸分子の豊富さを評価して、クロノタイププロファイルを生成する工程、を含む。別の実施形態では、配列決定する工程はネステッドセットのそれぞれについて複数の配列リードを生成する工程を包含する。別の実施形態では、体細胞変異により再構成されたそれぞれの領域は、V領域及びJ領域を含み、複数の配列リードのそれぞれは、V領域の異なる部分から開始し、会合する（associated）J領域の方向に伸長する。

【0081】

一態様では、個人に由来するそれぞれのサンプルに関し、本発明の方法に使用される配列決定法は、1回のラン当たり少なくとも1000のクロノタイプ配列を生成し；別の態様では、かかる手法は、1回のラン当たり少なくとも10,000のクロノタイプ配列を

生成し；別の態様では、かかる手法は、1回のラン当たり少なくとも100, 000のクロノタイプ配列を生成し；別の態様では、かかる手法は、1回のラン当たり少なくとも500, 000のクロノタイプ配列を生成し；並びに別の態様では、かかる手法は、1回のラン当たり少なくとも1, 000, 000のクロノタイプ配列を生成する。更に別の態様では、このような手法は、個々のサンプルの1回のラン当たり100, 000~1, 000, 000のクロノタイプの配列を生成する。前述のそれぞれにおいて、1回のラン当たりのそれぞれのクロノタイプは、少なくとも10個の配列リードから求められる。

#### 【0082】

提供する本発明の方法で使用する配列決定法により、1リード当たり、約30bp、約40bp、約50bp、約60bp、約70bp、約80bp、約90bp、約100bp、約110、約120bp、約150bp、約200bp、約250bp、約300bp、約350bp、約400bp、約450bp、約500bp、約550bp、又は約600bpを生成することができる。

#### 【0083】

##### 配列データからのクロノタイプの決定

本発明のいくつかの実施形態では、配列タグを使用してクロノタイプを決定し、その他の実施形態では、配列タグを配列リードと組み合わせて結合させる工程を用い、クロノタイプを決定する。単一の固有の配列タグが実質的にすべての異なる標的ポリヌクレオチドに付加される実施形態では、配列タグによるクロノタイプの決定は直接的 (straightforward) なものである。このような実施形態では、最初に配列タグに基づき配列リードをグループ化して、サンプルのクロノタイプを決定する。このようなグループ化は、従来の配列アライメント法により実施することもできる。アライメント法を選択するための指針として、参照により本願に援用される Batzoglou Briefings in Bioinformatics 6:6~22 (2005) に記載のものを使用可能である。配列リードを固有の配列タグに応じたグループに構築した後、関連するクロノタイプの配列を解析して、サンプル由来のクロノタイプの配列を決定することができる。図4Aには、固有の配列タグに関連するクロノタイプの配列 (配列番号: 2) を決定することによるアライメント及び方法の例を例示する。本例では、それぞれの配列タグ (4302) をもとに11の配列リードをアライメントさせた後、それぞれの配列リードのクロノタイプ部分 (4304) のそれぞれの部分のヌクレオチドを1、2、3、4、... nとして示し、比較する。例えば、部分6 (4306) のヌクレオチドはt、t、g、t、t、t、t、c、tである；すなわち、9つのベースコールがtであり、1つは「g」 (4308) であり、1つは「c」 (4310) である (配列番号: 3 及び配列番号: 4)。一実施形態では、クロノタイプ配列の各位置の正しい塩基の呼称は、主要な塩基がどう同定されるかによる。部分6 (4306) の例では、配列リードのほとんどでこの部分のヌクレオチドがtであることから、この塩基の呼称は「t」である。他の実施形態では、配列リードのベースコールのクオリティスコア又は隣接する塩基の同定などの他の因子を考慮に入れて、クロノタイプ配列について正確なベースコールを決定することもできる。クロノタイプを上記の通り決定したならば、サンプルのそれぞれの異なるクロノタイプの豊富さ又は頻度を含むクロノタイププロファイルを構築することもできる。

#### 【0084】

いくつかの実施形態では、増幅前に配列タグにより標識されている、サンプル中の標的ポリヌクレオチド画分を増加させる目的で、配列タグを含有しているプライマーを利用して、1工程以上の伸長工程を実施することもできる。このような実施形態では、配列タグを含有しているプライマーの存在下での1工程以上の伸長工程により、標的ポリヌクレオチド及び/又はそのコピーが複数の異なる配列タグにより標識される。「複数」の指す数は、配列タグを含有しているプライマーの存在下で実施する伸長工程の数、増幅反応の効率、並びに順方向及び逆方向プライマーのどちらが配列タグを有するか又は両方が配列タグを有するかなどに応じて異なる。いくつかのこのような実施形態では、「複数」は、2~15の範囲、又は2~10の範囲、又は2~5の範囲である。いくつかのこのような実

施形態では、増幅後、サンプルのそれぞれの標的ポリヌクレオチドのコピーを複数のグループ又はサブセットに分類してもよく、ここで、それぞれのグループ又はサブセットのメンバーは、同じ配列タグにより標識されており、それぞれの異なるグループ又はサブセットのメンバーは、異なる配列タグにより標識されており；すなわち、同じグループのメンバーは同じ配列タグを有し、異なるグループのメンバーは異なる配列タグを有する。言い換えると、いくつかの実施形態では、増幅後、サンプル由来の標的ポリヌクレオチドのそれぞれのコピーは、異なる2つの配列タグのうちの1つで標識されることになり；又は他の実施形態では、サンプル由来の標的ポリヌクレオチドのそれぞれのコピーは、異なる3つの配列タグのうちの1つで標識されることになり；又は他の実施形態では、サンプル由来の標的ポリヌクレオチドのそれぞれのコピーは、異なる4つの配列タグのうちの1つで標識されることになり、これが以降同様に順次繰り返される。これらの実施形態では、クロノタイプは、配列タグの組み合わせをアライメントした後、共通の起点 (common origin) がエラー率及び相対頻度の関数として真であるという尤度に基づき、同じ親配列に由来するものとしてグループ内の配列リードを処理するために、結合させる工程により決定することもできる。図4Bは、このような実施形態に由来する配列リードを例示する。あるアプローチでは、配列リードは、最初に共通の配列タグ (4402) によりグループ化され、この図の場合、結果として3つのグループ (4420)、(4422)、及び (4424) が形成されている。いくつかの実施形態では、それぞれのグループ内で、配列 (4404) を解析してグループのコンセンサス配列を決定する；例えば、上記の通り、それぞれのヌクレオチド部分において、塩基は、主要な塩基、又は最も高頻度な塩基などとして参照することができる。次に、コンセンサス配列群を互いに結合させてクロノタイプを決定することもできる。

#### 【0085】

いくつかの実施形態では、本発明の上記態様は、サンプル中の実質的に任意の核酸群をプロファイリングするための方法において実施することもできる。このような方法は、工程：(a) 核酸群を含むサンプルを得る工程；(b) 群の核酸に対し配列タグを付加してタグ-核酸複合体を生成する工程（ここで、少なくとも1つの核酸群又はそれらのコピーは異なる配列タグを付加される）；(c) タグ-核酸複合体を増幅させる工程；(d) タグ-核酸複合体を配列決定して、エラー率を有し核酸配列及びタグ配列を含む配列リードを生成する工程；(e) 同様のタグ配列を有する配列リードをアライメントして、同じ配列タグを有する配列リードのグループを形成する工程；(f) グループの配列リードを結合させて核酸配列を決定する工程（ここで、配列リードのグループの尤度が明らかに少なくとも25%であるときは常に、配列リードのグループを、異なる配列に結合させる）；並びに (g) 配列のレベルを決定することにより群の配列プロファイル決定する工程、を含む。再構成された核酸群のプロファイリングに利用するとき、このような方法は、次の工程：(a) 個人から、T細胞及び/又はB細胞及び/又は無細胞系DNAを含むサンプルを得る工程；(b) サンプル由来のT細胞受容体遺伝子又はイムノグロブリン遺伝子の再構成された核酸分子に配列タグを付加してタグ-核酸複合体を生成する工程（ここで、サンプル由来の少なくとも1つの再構成された核酸又はそれらのコピーには異なる配列タグが付加される）；(c) タグ-核酸複合体を増幅させる工程；(d) タグ-核酸複合体を配列決定して、エラー率を有しタグ配列及び再構成された核酸配列を含む配列リードを生成する工程；(e) 同様のタグ配列を有する配列リードをアライメントして、同じ配列タグを有する配列リードのグループを形成する工程；(f) グループの配列リードを結合させてクロノタイプを決定する工程（ここで、配列リードのグループの尤度が明らかに少なくとも25%であるとき、配列リードのグループを異なる配列に結合させる）；並びに (g) クロノタイプのレベルを決定することによりサンプルのクロノタイププロファイル決定する工程、により実施することもできる。

#### 【0086】

上記の実施形態及び本明細書に開示されるその他の実施形態において、タグ-核酸複合体を配列決定する工程は、アンブリコンに由来するタグ-核酸複合体のサンプルを配列決

10

20

30

40

50

定することを含む。通常、このようなサンプルは、もとのサンプル（すなわち、組織サンプル又は血液サンプルなど）における標的ポリヌクレオチドの相対頻度が、増幅反応産物に由来するタグ-核酸複合体サンプルにおいても維持されている、代表的なサンプルである。いくつかの実施形態では、免疫受容体分子をコードしている再構成された核酸群が解析され、タグ-核酸複合体サンプルは、少なくとも  $10^4$  個のタグ-核酸複合体を含み；他の実施形態では、このようなサンプルは少なくとも  $10^5$  個のタグ-核酸複合体を含み；他の実施形態では、このようなサンプルは、少なくとも  $10^6$  個のタグ-核酸複合体を含み；他の実施形態では、このようなサンプルは、少なくとも  $10^7$  個のタグ-核酸複合体を含む。

#### 【0087】

##### 配列リードの結合

複数の配列タグがもとの再構成された核酸又はそれらのコピーに付加される実施形態では、クロノタイプを決定するために配列リード（又はグループのコンセンサス配列リード）を結合させる工程を実施してもよい。配列決定法にエラーが含まれない場合、所定のサンプルの配列リードのセットを減少させて、異なるクロノタイプのセットにすること、及びそれぞれのクロノタイプのリード数を記録することは自明である。しかしながら、配列決定エラーが存在する場合には、それぞれの本当のクロノタイプは、その配列に様々な数のエラーを含む配列リードの「集団」に埋没されることになる。配列決定エラーの「集団」は、配列空間のクロノタイプからの距離を延長させて、密度を低下させる。配列リードをクロノタイプに変換するに当たって、様々なアルゴリズムを利用可能である。一態様では、配列リードの結合（すなわち、1つ以上の配列エラーを有することが確認されているクロノタイプ候補を組み合わせること）は、少なくとも3つの因子：比較するそれぞれのクロノタイプについて得られる配列数；異なる塩基の数；及び不一致の部分の配列決定クオリティスコアに依存する。いくつかの実施形態では、予想されるエラー率及びエラーの二項分布をもとに尤度の比を構築し、評価してもよい。例えば、配列決定クオリティの低い領域に違いが1つ含まれる、150リードを有するクロノタイプ及び2リードを有するクロノタイプの2つのクロノタイプは、これらのクロノタイプが配列決定エラーにより生成されている可能性があるものとして、結合される傾向がある。一方で、違いが2つ含まれる、100リードを有するクロノタイプ及び50リードを有するクロノタイプの2つのクロノタイプは、これらのクロノタイプが配列決定エラーにより生成されている可能性は低いものとみなされ結合されない。いくつかの実施形態では、配列リードに由来するクロノタイプを決定するに当たり下記のアルゴリズムを使用することもできる。これらの概念のうちいくつかを図5Aに例示する。結合工程のいくつかの実施形態では、配列リードは、最初に候補のクロノタイプに変換される。このような変換は、利用する配列決定プラットフォームに依存する。Qスコアが高く長い配列リードを生成するプラットフォームに関しては、クロノタイプ候補として配列リード又はそれらの部分を直接採用することもできる。Qスコアが低く短い配列リードを生成するプラットフォームに関しては、関連する配列リードのセットをクロノタイプ候補に変換するにはいくつかのアラインメント及び構築工程が必要とされ得る。例えば、いくつかの実施形態では、Solexa系のプラットフォームに関しては、前述の通り、例えば、10以上の複数のクラスターから対としたリードのコレクションをもとにクロノタイプ候補が生成される。

#### 【0088】

図5Aに例示する通り、配列空間にクロノタイプ候補の頻度をプロットすることができる。図5Aでは、かかる空間は、例示の都合により次元（横軸）に修正している。縦軸により、それぞれのクロノタイプ候補の頻度、 $\log$ （リードカウント）、又は同様の測定値の規模を与える。図中、様々な記号（530）によりクロノタイプ候補を表す。本発明の一実施形態によると、候補となる2つのクロノタイプを、それらのそれぞれの頻度又はリードカウント（前述の通り）、それらの間で異なっている塩基の数（異なっている数が大きくなるほど結合させる可能性は低くなる）、並びにそれぞれの配列が異なっている位置の塩基のクオリティスコア（クオリティスコアが高くなるほど、結合させる可能性は

10

20

30

40

50

低くなる)に応じ結合させる。クロノタイプ候補は、それらのそれぞれの頻度の順に検討することができる。図5Aは、頻度が最も高い3つの候補として、クロノタイプ候補1(532)、クロノタイプ候補7(534)、及びクロノタイプ候補11(536)を示す。それぞれのこのようなクロノタイプ候補に関連して、配列は近いものの頻度が劣る、その他のクロノタイプ候補が存在し、例えば、(i)クロノタイプ候補1(532)に関しては、クロノタイプ候補2(538)と、コーン(540)の内側に含まれる、クロノタイプ候補3、4、5、及び6とが存在し；クロノタイプ候補7(534)に関しては、コーン(542)の内側に含まれる、クロノタイプ候補8、9及び10が存在し；(iii)クロノタイプ候補11に関しては、コーン(544)の内側に含まれる、クロノタイプ候補12が存在する。このコーンは尤度の境界を示し、境界内部に含まれるクロノタイプ候補は、より頻度の高いクロノタイプ候補1、7又は11と結合されることになる。このような尤度の境界は、近隣のクロノタイプ候補(1については3、4、5及び6；7については8、9及び10；並びに11については12)の頻度と、配列空間における、それぞれの候補の、それぞれのより頻度の高いクロノタイプ候補からの距離とに応じる。クロノタイプ候補2(538)は、コーン(540)の外側にある；したがって、クロノタイプ候補1(532)とは結合されない。候補となるクロノタイプの頻度が高くなるほど、頻度の低いものよりも真に異なるクロノタイプである可能性が高くなり、並びに頻度が低い(複数の)違いは、頻度の高い(複数の)違いよりもエラーである可能性が高くなることから、再度、(結合の)尤度の境界をコーンとして示す。

【0089】

それぞれのクロノタイプ候補を埋没させる配列リード集団は、二項分布及び単一塩基エラーの見込みについての単純モデルを用いモデル化することができる。後者のエラーモデルは、V及びJ断片のマッピングから推察することができ、又はアルゴリズムを見出すクロノタイプそのものから自己整合及び収束性をもとに推察することができる。モデルは、配列空間のこの領域においてXのみが真のクロノタイプであるというヌルモデルの下、リードカウントC2及びEエラー(配列Xに対してのもの)を有する所与の「集団」配列Yが、完全なリードカウントC1を有する真のクロノタイプ配列Xの一部となる確率をもとに構築する。パラメーターC1、C2、及びEに従い、配列YをクロノタイプXと結合させるか否かの判断を行う。任意の所与のC1及びEについて、配列Yの結合を決定するため、最大値C2を予め算出する。YがクロノタイプXに含まれるというヌル仮説の下、隣接する配列XにエラーEを有するすべての可能性のある配列Yをまとめた後に、Yを結合させることに失敗する確率があるP値未満になるよう、C2の最大値を選択する。P値は、アルゴリズムのふるまいを制御し、結合の寛容性を増大させる若しくは低下させる値である。

【0090】

クロノタイプXに結合させるに当たって、そのリードカウントが閾値C2を超えるために、配列YがクロノタイプXに結合されない場合、配列Yは、別のクロノタイプを提供する候補となる(図5A中、クロノタイプ候補2(538)など)。同様に、このような原則を実装するアルゴリズムでは、この配列Y(Xとは独立していると考えられている)に、「より近い」ものである任意のその他の配列Y2、Y3などは、確実にXにまとめられなくなる。この「類似性」という概念には、Y及びXに対するエラーカウントと、X及びYの完全なリードカウントとの両方が包含され、すなわち、クロノタイプX周辺のエラー配列集団についての上記のモデルと同じ方法でモデル化される。この方法において、たまたま1つ以上のクロノタイプに「近い」ものであるとされた場合、「集団」配列は、それらの正確なクロノタイプに適切に所与するものであり得る。したがって、図5Aによると、クロノタイプ候補2がクロノタイプ候補1(532)とは真に異なるものであるとみなされる場合、特有のルーチン又はサブアルゴリズムにより、クロノタイプ候補1(532)及び2(538)のどちらと、1及び2の間にある候補4及び5とを結合すべきか(いずれか)を決定するルールが提供される。

【0091】

ー実施形態では、アルゴリズムは、リードカウントの最も高い配列Xから開始して、トップダウン式に進行する。この配列が、第1のクロノタイプを提供する。それらのカウントが予め算出された閾値を下回る場合（上記を参照のこと）、隣接する配列は第1のクロノタイプに結合させ、又はそれらが閾値を超える場合、若しくは結合させなかった別の配列に「より近い」場合には、そのまま残す。最大エラーカウント内の隣接するすべての配列を検索した後、リードをクロノタイプXに結合させるプロセスを完了する。結合させたリード及びすべてのリードは、その他のクロノタイプ生成に利用可能なリードのリストに加え、及びこのリストから除外する。次に、隣の配列を最も高いリードカウントに移動させる。隣接するリードを上記の通りこのクロノタイプに結合させ、閾値を超えるリードカウントを有する配列がなくなるまで、例えば、1超のカウントを有するすべての配列がクロノタイプのシードとして使用されるまでこのプロセスを継続する。

10

#### 【0092】

上記の通り、上記アルゴリズムの別の実施形態では、関連する配列リードのクオリティスコアを考慮に入れて、候補となる配列Yを既存のクロノタイプXに結合させるかどうかを決定することを試験に更に追加してもよい。配列（複数可）Yについて、平均クオリティスコア（複数可）（配列Yを有するすべてのリード間で平均したもの）を決定する。ここで（were）、配列YとXは異なる。平均スコアが所定値を超える場合、その違いは、真に異なるクロノタイプを示すものである可能性が高いため、結合させるべきではなく、また平均スコアが所定値を下回る場合、配列Yは配列決定エラーにより生じたものである可能性が高いため、Xに結合させるべきである。

20

#### 【0093】

クロノタイプ候補を結合させるために上記アルゴリズムを良好に実装するには、いくつかのインプット配列XよりもEエラーの小さい（すなわち、いくつかの配列距離測定値よりも小さい）配列を効率的にすべて特定する方法が求められる。この課題は、配列ツリーを利用することで解決できる。図5Dに例示する通り、このようなツリーの実装には、クロノタイプ候補のDNA配列の一文字表記にツリーのノードが限定されないという、ある程度独自の特徴がある。ノードは、コンピューターメモリーのより効率的な使用を可能にする任意の長配列を有し得る。

#### 【0094】

所与のサンプルのすべてのリードを配列ツリーに配置する。それぞれの葉ノードは、関連するリードへの分岐点を保有する。クロノタイプ候補に固有の配列は、葉ノードから根ノードへと逆に走査することにより回収される。根ノード1つと、リードの完全な配列を含有する葉ノード1つとを有する単純なツリーに、第1の配列を配置する。配列は、1つずつ隣り合わせて追加する。ツリーが既に配列を含有している場合、追加したそれぞれの配列に関し、リード及び既存のツリーとの間で共通の配列の終点で新しい分岐が形成され、又は既存の葉ノードにリードが追加される。すべてのリードをツリーに配置することで、以下の目的1）、2）でのツリーの使用が容易になる：1）最も高いリードカウントを特定する：リードカウントをもとに葉ノードをソートして、最もリードを有する葉ノード（すなわち、配列）及び連続的により少ないリードを有する葉ノードを特定可能である；2）隣接する葉を特定する：任意の配列について、この配列に関しXよりも少ないエラーを有するツリー中のすべての経路が検索可能である。経路は根から始まり、この経路はツリーに沿って進み別の経路に分岐する。ツリーに沿って進行させるに伴い、それぞれの経路の最新のエラーカウントを記載する。エラーカウントが、エラーに許される最大値を超過したとき、所与の経路は終了する。この方法では、ツリーの大部分が可能な限り早期に切り上げられる。この方法は、任意の所与の配列をもとに、Xのエラー内のすべての経路（すなわち、すべての葉）を効率的に特定する。

30

40

#### 【0095】

上記概念の特徴を、図5Bのフローチャートにより詳細に例示する。クロノタイプ候補のセットは、T細胞又はB細胞サンプルから抽出した再構成された核酸を配列決定することにより得た配列データから得る。一態様では、クロノタイプ候補は、それぞれ、NDN

50

領域と、V領域及びJ領域の部分とを含む。これらの配列をデータ構造(550)に構造化する。このデータ構造は配列ツリーであってよい。図5Bには非掲載であるが、一実施形態では、クロノタイプ候補のセットを生成する際に、既知のV領域及び既知のJ領域について配列ツリーを構築してもよい。配列ツリーをもとに、これらの既知の配列に対してクロノタイプ候補を構成する配列リードをマッピング又はアライメントして、クロノタイプ候補に関し最も可能性の高い、既知のV配列及びJ配列を効率的に決定してもよい。図5Bに戻ると、クロノタイプ候補が生成されたなら、実験誤差又は測定誤差(配列決定誤差など)を含むクロノタイプ候補から真のクロノタイプを識別する方法に使用するため、配列ツリーなどのデータ構造を構築する。データ構造、例えば、配列ツリーから、現在のクロノタイプ候補( $H F C C_k$ )間で出現頻度が最も高い、クロノタイプ候補を選択する(552)；言い換えると、 $H F C C_k$ は、クロノタイプ候補のうち、サイクルkにおいて最もコピー数が大きいものであり、すなわちリードカウントが大きいものである。次に、隣接する、頻度に劣るクロノタイプ候補を特定する( $L F C C_s$ )(554)；すなわち、距離 $D_k$ 以内のクロノタイプ候補を特定する。本発明の一態様では、比較的短い(300bp未満)配列を効率的に配列比較できる、配列ツリーを利用してこの特定を実施する。

#### 【0096】

一実施形態では、動的プログラミング、例えば、Gusfield(上掲)に開示されているものなどを用い、比較又は配列アライメントを実施する。更なる実施形態では、このような動的プログラミングは、帯域を制限した動的プログラミングであり、選択したH F C Cとは所定の距離を超えて異なる配列については検討を行わないため、計算速度が増加する。候補となる $H F C C_k$ 及び $L F C C_j$ は、多くの様々な基準又は属性をもとに比較することができる。一態様では、上記の通り、少なくとも2つの基準：(i)頻度又はリードカウント及び(ii)配列の違い、をもとに、クロノタイプ候補を比較する。一態様では、上記の通り、少なくとも3つの基準：(i)頻度又はリードカウント、(ii)配列の違い、及び(iii)クオリティスコア又は違いの生じている塩基の評価、をもとに、クロノタイプ候補を比較する。一実施形態では、配列の違いは、塩基の置換を含む；別の実施形態では、配列の違いは、塩基の置換、欠失及び挿入を含む。後者の実施形態は、454シーケンサ及びIon Torrentシーケンサなどの、ターミネーターを利用しない合成時解読法により配列データが生成される場合に、特に応用可能である。このような配列決定アプローチは、シグナル振幅をもとに、異なる大きさのホモポリマー伸長を識別する；したがって、ホモポリマーのサイズが大きくなるほど、ホモポリマーの1ヌクレオチドが異なることによるシグナルレベルの差が大幅に減少することから(すなわち、2-merは3-merから容易に識別されるのに対し、8-merはほとんど9-merと識別不能である)、このようなアプローチでは、ベースコールのルーチンにおいて挿入及び欠失エラーが生じる傾向がある。一態様では、意思決定ボックス(558)に示す、上記(i)～(iii)の量に依存する $P(H F C C_k, L F C C_j, D, Q)$ などの関数(本明細書では、「結合尤度関数」として言及する)を用い、H F C C及びL F C Cの比較を実行する。このような関数は、多くの異なる形態を取り得るものの、一般的に、P値は、以下の通り、(i)、(ii)及び(iii)における変化を伴って変化する：P値は、好ましくは、H F C Cの頻度、及びL F C Cの頻度に対するH F C Cの頻度の比によって単調増加し、L F C Cの頻度に対するH F C Cの頻度の比が高くなるほど、L F C CをH F C Cに結合させる尤度は高くなる。同様にして、P値は、H F C C及びL F C Cの配列の異なる度合いによって好ましくは単調減少し、H F C CとL F C Cとの違いが大きくなるほど(例えば、最小数の置換、挿入又は欠失により互いに対し数をもとに測定される通り)、L F C CをH F C Cに結合させる尤度は小さくなる。最終的に、P値は、H F C C及びL F C Cの配列が異なっている位置のクオリティスコアが増加するほど、好ましくは単調減少し、クオリティスコアが高くなるほど、L F C CをH F C Cに結合させる尤度は小さくなる。

#### 【0097】



H F C C 及び L F C C の配列が 1 つ以上の位置で異なるとき、異なる位置のクオリティスコアを様々な異なる手法で組み合わせることもできる。一実施形態では、このような違いが複数存在するときは常に、複数のクオリティスコアを平均値として表す。この平均値は非加重平均又は加重平均のいずれであってもよい。図 5 C は、所与の配列差についてそれぞれのクオリティ値（曲線 a ~ e）を算出した関数例、P を示す。図 5 C に示す通り、H F C C が約 200 リードカウント程度であり（570）、クオリティスコアが曲線（a）により決定される場合、約 50 リードカウント（572）未満の任意の L F C C は H F C C に結合させる。関数 P の独立変数 D は、配列 H F C C<sub>k</sub> 及び L F C C<sub>j</sub> 間の距離の測定値であり、この値は、解析が進むにつれてサイクル間で異なり得る（「k」は、「k」が添えられている定数の値が、計算サイクル k に依存し得るものであることを意味する）。一実施形態では、 $D = D_k$  であり、この値はサイクル数の関数である。別の実施形態では、 $D = D(\text{H F C C 頻度})$  であり、この値は、サイクル数とは独立した H F C C の頻度の関数である。例えば、H F C C の頻度が低下するにつれ、比較する候補の距離 D は減少する。一実施形態では、D は H F C C<sub>k</sub> 及び L F C C<sub>j</sub> 間のハミング距離である；しかしながら、その他の距離測定値を使用してもよい。一実施形態では、 $D_k$  は、k の非増加関数であり；別の実施形態では、 $D_k$  は k の減少関数である。いくつかの実施形態では、サイクル数の増加又は H F C C 頻度の減少に伴い D の規模を縮小させることが有用である。計算が進むにつれてクロノタイプ候補の頻度は低下するため、このような候補のほとんどはシングルトンであり、よって（頻度の違いよりも）配列距離が主要な比較になる。計算が進み D が低下することにより、遠く、頻度の低いクロノタイプ候補に対する、生産性の低い比較が削減され、これにより計算速度が増加する。関数 P は、検討する因子の数に応じ複雑な式になり得る。図 5 C には、一実施形態の P について算出された値を例示する。P は、上記の通り、L F C C を、クオリティスコアの異なる H F C C の所与のリードカウントと結合させる際の、リードカウントの閾値に関係する。曲線「a」~「e」は、異なるクオリティスコアの関係を表す（曲線「a」が最も高いクオリティスコアに相当する）。

#### 【0098】

図 5 B に戻ると、 $P < P_k$  である場合、L F C C<sub>j</sub> を H F C C<sub>k</sub> と結合させず、別の L F C C を選択する（560）。 $P > P_k$  である場合、L F C C<sub>j</sub> を H F C C<sub>k</sub>（562）と結合させ、評価する L F C C が残っていない場合（564）を除き、別の L F C C を選択する（566）。評価する L F C C が残っていない場合（564）、配列ツリーなどのデータ構造から現在の H F C C<sub>k</sub>（結合させたすべての L F C C を含む）を除外する（568）。このような除外を、図 5 D ~ 5 E の単純な配列ツリー（590）に例示する。配列ツリー（590）において、経路（592）（破線により示す）が H F C C（596）に対応しており、この経路には L F C C（598）が結合されている。結合後、影をつけた領域（599）における経路セグメント（592）を配列ツリー（590）から除外し、図 5 E に示す配列ツリー（597）を減らしている。隣接する L F C C を特定する（554）一連の計算にこのツリーを利用するこのような除去後、停止規則（570）が満たされた場合、クロノタイプの決定を完了させる。一実施形態では、停止規則（570）は、最後の非シングルトンなクロノタイプ候補が処理された（552）かどうかである。別の実施形態では、停止規則（570）は、選択された H F C C の頻度又はリードカウントが、単個のリンパ球の場合に相当する頻度又はリードカウントを下回っているかどうかである。本発明の方法の一態様では、増幅工程により、結果的に、サンプル中のそれぞれのリンパ球が同じクロノタイプの複数のコピーにより表され；したがって、一実施形態では、何であろうと、H F C C のリードカウント数が、単個のリンパ球の場合に相当するリードカウント数を下回っていた場合、算出が停止される。いくつかの実施形態では、このようなリードカウント数（又はクロノタイプ候補のコピー）は、少なくとも 10 であり；別の実施形態では、このような数は、少なくとも 20 であり；別の実施形態では、このような数は、少なくとも 30 であり；別の実施形態では、このような数は、少なくとも 40 である。停止規則が満たされない場合、次の H F C C が選択される（572）。図 5 B のフ

ローチャートに要約した解析工程を、C、C++、Java（登録商標）、C#、Fortran、又はPascalなどの任意の好適なプログラミング言語で実装することができる。

#### 【0099】

本発明の一態様に従って、クロノタイプ及び/又はクロノタイププロファイルを決定する上記の方法は、(a)ハイスループットな核酸配列決定により得られた配列リードから組み換え免疫分子のデータ構造を形成する工程、(b)任意の、より劣る方の頻度が、所定の頻度値を下回り、かつそれらの配列の異なる度合いが所定の度合いを下回る場合には常に、最も高頻度のクロノタイプ候補と、より頻度の劣るクロノタイプ候補とを結合させて、クロノタイプを形成する工程、(c)結合させた、クロノタイプ候補をデータ構造から除去する工程、並びに(d)クロノタイププロファイルが形成されるまで工程(b)及び(c)を繰り返す工程、を含む。一実施形態では、データ構造は配列ツリーである。

#### 【0100】

本発明の別の態様によると、クロノタイプを決定する上記の方法は：(a)それぞれV領域、NDN領域、及びJ領域を有する組み換え免疫分子のレパトアから、配列リードのセットを準備する工程、ここで、それぞれのこのような分子に関し、少なくとも1つの配列リードは、このような分子のNDN領域の少なくとも部分を包含する；(b)NDN領域の少なくとも部分を包含する配列リードから、クロノタイプ候補を葉が表す配列ツリーを形成する工程[それぞれの葉及び対応するクロノタイプ候補は頻度を有する]；(c)任意の、より劣る方の頻度が、所定の頻度値を下回り、かつそれらの配列の異なる度合いが所定の度合いを下回る場合には常に、最も高頻度のクロノタイプ候補と、より頻度の劣るクロノタイプ候補とを結合させて、最も高頻度のクロノタイプ候補の配列を有する、クロノタイプを形成する工程；(d)結合させた、クロノタイプ候補に対応する葉を配列ツリーから除去する工程；及び(e)頻度に劣る、候補となるクロノタイプの最高頻度が所定の停止値を下回るまで、工程(c)及び(d)を繰り返す工程、を含む工程により実施することができる。一実施形態では、形成する工程は、最も高頻度のクロノタイプ候補を選択する工程と、異なる度合いが所定の度合いを下回るものであり異なる配列を有する、頻度に劣るクロノタイプ候補をすべて特定して、結合サブセットを形成する工程と、を更に包含する。したがって、このような実施形態では、結合処理のために比較すべきLFCの総数を限定することができる(所定の異なる値内の1つのみが検討される)。かかる値は、用途に応じてプロセスから入力され、例えば、レパトワの規模、及び用いる計算時間などが該当する。上記の通り、HFCCにLFCを結合させるかどうかを決定するのに使用する関数は様々な形態を有し得る。ある概括的な態様では、結合させる工程では、この関数は次の属性を有し得る：HFCCの頻度、LFCの頻度、それらの配列間の違い(ハミング距離などの、文字列の違いを計測する従来法により表すことができる)、及び1つ以上のヌクレオチド位置のクオリティスコア(ここで、HFCC及びLFCは異なる)に応じるものであり；関数は、(i)LFCの頻度に対するHFCCの頻度の比が大きくなるほど単調増加し、(ii)HFCC及びLFCの配列間の違いが大きくなるほど単調減少し、(iii)1つ以上のヌクレオチド位置のクオリティスコアが大きくなるほど単調減少する。すなわち、属性(iii)の観点では、HFCC及びLFCが異なるという確からしさが高いと(例えば、ベースコールの信頼性が非常に高いなど)、それらを結合させる可能性は低くなる。

#### 【0101】

いくつかの実施形態では、配列リードの尤度が明らかに少なくとも95パーセントであるときには常に、このような配列リードを異なるクロノタイプ(又は再構成された核酸などの標的ポリヌクレオチド)に結合させるよう結合尤度関数を選択する；他の実施形態では、配列リードの尤度が明らかに少なくとも99パーセントのであるときは常に、このような配列リードを異なるクロノタイプに結合させるよう結合尤度関数を選択する；他の実施形態では、配列リードの尤度が明らかに少なくとも99.9パーセントであるときは常に、このような配列リードを異なるクロノタイプに結合させるよう結合尤度関数を選択す

る。上記の通り、いくつかの実施形態では、結合尤度関数は、配列決定に使用する化学的手法のエラー率、比較する配列リード中の相違するヌクレオチドの数、並びに比較する配列リードの相対的な頻度に応じて異なり；別の実施形態では、結合尤度関数は、配列決定に使用する化学的手法のエラー率、比較する配列リード中の相違するヌクレオチドの数、比較する配列リードの相対的な頻度、及び相違したヌクレオチドのクオリティスコアに応じて異なる。上記において、所定の頻度及び所定の異なり度合いは、設計の際に具体的な用途に応じて選択される。かかる選択に影響を及ぼす因子としては、生体現象の詳細、及び実施速度などが挙げられる。

#### 【0102】

##### モニタリング用途

一態様では、本発明は、サンプル中の、疾患に特徴的な又は疾患に関連する核酸の存在、非存在、及び/又はレベルを評価することにより微小残存病変をモニタリングする方法を目的とする。いくつかの実施形態では、このような核酸は体細胞的に再構成された核酸、あるいは前癌性の状態又はリンパ増殖性疾患若しくは骨髄増殖性疾患などの癌の状態と相関するクロノタイプであり、これを利用して、疾患の状態又は病態をモニターすることができる。このような核酸、及び特にクロノタイプは、処置後に癌の微小残存病変をモニタリングするのに有用であり、このようなモニタリングの結果は、処置の継続、中断、又は変更を決定するに当たって重要な因子である。多くの悪性リンパ腫瘍及び骨髄腫瘍では、クロノタイププロファイルを生成する（「診断的クロノタイププロファイル」）処理より前に、末梢血サンプル又は骨髄サンプルなどの診断用組織サンプルを得る。リンパ増殖性疾患又は骨髄増殖性疾患に関し、通常、診断試料を採取する前には免疫受容体鎖（複数可）が疾患又は状態に関係するリンパ球又は骨髄のクローンに関連するかは分からない。結果として、現在の慣例では、患者の疾患又は状態に関係するクロノタイプを特定する目的で、異なる候補となる免疫受容体鎖をコードしている異なる再構成核酸に対し、増幅及び配列決定を別々に複数回実施する必要がある。このように手間を掛けた増幅及び配列決定により生成されるクロノタイププロファイルにおいて、疾患に関係する1つ以上のクロノタイプ（すなわち、「関係するクロノタイプ」）が特定される。典型的には、クロノタイププロファイルにおいて頻度が最も高いクロノタイプが、対応するクロノタイプとして扱われる。本発明の一態様では、複数の異なる免疫受容体鎖をコードしている再構成された核酸部分の単回反応での大規模多重増幅を提供することにより、関連するクロノタイプを特定するのに必要とされていた増幅及び配列決定の実施回数を大幅に低減させる。いくつかの実施形態では、「複数」とは、異なる免疫受容体鎖が2～4種の範囲内のものであることを意味し；並びに、その他の実施形態では、「複数」とは、異なる免疫受容体鎖が2～3種の範囲内のものであることを意味する。より具体的には、いくつかの実施形態では、BCR鎖のなかでも、次のもの：VDJ領域の少なくとも部分を含むIGH、DJ領域の少なくとも部分を含むIGH、及びIGK、をコードしている再構成された核酸を、単回多重反応において増幅させる；並びに、その他の実施形態では、TCR鎖のなかでも、次のもの：TCR、TCR及びTCRをコードしている再構成された核酸を、単回多重反応において増幅させる。

#### 【0103】

処置後、及び好ましくは癌の完全な寛解が達成された後、かかる関連するクロノタイプ又は核酸の存在、非存在、又は頻度を定期的に評価して、処置後のクロノタイププロファイル又は核酸プロファイルにおける、関連する核酸又はクロノタイプ（又は関係づけられるクロノタイプ）の存在又は頻度の増大に基づき、寛解の維持、又は腫瘍の再出現若しくは再発を診断する。すなわち、処置後に、関連するクロノタイプ又は特徴的な核酸の存在、非存在、又は頻度に基づき、癌の微小残存病変を評価する。上記の通り、関連するクロノタイプが、再構成された受容体断片に共通する又は相当するものであるとき、十分な多様性が不足することになり（そのため、癌細胞性のものではない細胞もクロノタイプを共有し得る）、処置後のクロノタイププロファイルにこのようなクロノタイプが存在することで、再発を示す擬陽性が生じ得る。

## 【 0 1 0 4 】

本発明の方法は、免疫受容体又はそれらの部分をコードする再構成された核酸を、疾患に  
関与する細胞についてのマーカーとして使用できる、任意の増殖性疾患のモニタリング  
に応用可能である。一態様では、本発明の方法は、リンパ増殖性疾患又は骨髄増殖性疾患  
に応用可能である。別の態様では、本発明の方法は、リンパ腫及び白血病に応用可能であ  
る。別の態様では、本発明の方法は、濾胞性リンパ腫、慢性リンパ性白血病（CLL）、  
急性リンパ性白血病（ALL）、慢性骨髄性白血病（CML）、急性骨髄性白血病（AML）、  
ホジキン及び非ホジキンリンパ腫、多発性骨髄腫（MM）、良性単クローン性 グ  
ロブリン血症（MGUS）、マンツル細胞リンパ腫（MCL）、びまん性大細胞型B細胞  
リンパ腫（DLBCL）、骨髄異形成症候群（MDS）、又はT細胞リンパ腫などにおけるMRD  
のモニタリングに応用可能である。特定の実施形態では、本発明の方法は、ALL、MM、  
又はDLBCLのMRDをモニタリングするのに良好なものである。

10

## 【 0 1 0 5 】

いくつかの実施形態では、血液又は骨髄などの患者サンプルに対して診断検査を行い、  
複数の免疫受容体鎖のうち疾患のクローンにより産生されたクロノタイプ（すなわち、関  
連するクロノタイプ）を含み得るものを特定する。関連するクロノタイプの免疫受容体鎖  
が特定されたならば、以降のモニタリングアッセイを特定の免疫受容体鎖に特異的なもの  
にすることもできる。例えば、いくつかの実施形態では、診断検査は、同じ反応において  
IGH（VDJ）、IGH（DJ）、及びIGKなどの複数のBCR鎖の配列に基づくクロ  
ノタイププロファイルを生成する。関連するクロノタイプがIGH（VDJ）鎖である  
場合、以降のモニタリングアッセイでは、IGH（VDJ）のクロノタイププロファイル  
のみを生成してもよい。いくつかの実施形態では、診断試料における配列決定の深度は、  
モニタリングサンプルのものとは異なるものであり得る。「配列決定の深度」は、クロノ  
タイププロファイルを構築するのに解析する配列リードの総数を意味する。白血病又はリ  
ンパ腫などの癌に関しては、診断検査は処置前の患者サンプルに対し実施され、サンプル  
において関連するクロノタイプの頻度又はレベルは典型的には高く、かつ容易に特定され  
る。例えば、所定のレベルを超える頻度を有する任意のクロノタイプを、関連するクロノ  
タイプとして定義することができる。このような所定のレベルは、その他の患者の指標下  
では変化し得るものである；しかしながら、多くの場合、所定のレベルは、2～5%の範  
囲であってよく；又はいくつかの実施形態では5%であってよい。したがって、いくつか  
の実施形態では、配列決定を実施する深度は、1又は2%以上の頻度で存在するクロノ  
タイプを確実に検出するのに必要なものとする。いくつかの実施形態では、診断試料の配  
列決定の深度は、少なくとも10,000の配列リードをもたらすものであり；又は他の実  
施形態では、少なくとも100,000の配列リードをもたらすものであり；更に他の  
実施形態では、診断試料の配列決定の深度は、少なくとも10<sup>6</sup>個の配列リードをもたら  
すものである。いくつかの実施形態では、モニタリングサンプルの配列決定の深度は、少  
なくとも100,000の配列リードをもたらすものであり；他の実施形態では、モニ  
タリングサンプルの配列決定の深度は、少なくとも10<sup>6</sup>個の配列リードをもたらすも  
のである。

20

30

## 【 0 1 0 6 】

いくつかの実施形態では、患者から継続的に得たサンプル（又は組織サンプル）からク  
ロノタイププロファイルを生成することにより、患者の白血病又はリンパ腫などのリンパ  
増殖性疾患をモニターすることができる。このようなクロノタイププロファイルは上記の  
通りに生成することができる。いくつかの実施形態では、このようなモニタリングは、次  
の工程：（a）個人から、T細胞及び/又はB細胞及び/又は無細胞系DNAを含むサン  
プルを得る工程；（b）サンプルのT細胞受容体遺伝子又はイムノグロブリン遺伝子の再  
構成された核酸分子に配列タグを付加してタグ-核酸複合体を生成する工程（ここで、少  
なくとも1つの再構成された核酸又はそれらのコピーは異なる配列タグが付加されてい  
る）；（c）タグ-核酸複合体を増幅させる工程；（d）タグ-核酸複合体サンプルを配  
列決定して、それぞれエラー率を有しタグ配列及び再構成された核酸配列を含む配列リード

40

50

を生成する工程；（e）同様のタグ配列を有する配列リードをアライメントして、同じ配列タグを有する配列リードのグループを形成する工程；（f）グループの配列リードを結合させてクロノタイプを決定する工程（ここで、配列リードの尤度が明らかに少なくとも95%であるときは常に、配列リードのグループを、異なる再構成核酸配列に結合させる）；（g）クロノタイプレベルを決定することによりサンプルのクロノタイププロファイルを決定する工程；並びに（h）クロノタイププロファイルにおいて関連するクロノタイプのレベルを決定する工程、により実施することもできる。いくつかの実施形態では、患者をモニタリングして、関連するクロノタイプのレベルが疾患の寛解のエビデンスになり得るかどうかを診断するプロセスにおいて、工程（a）～（h）を繰り返してもよい。いくつかの実施形態では、付加及び増幅工程は、次の工程：（a）反応混合物中、プライマー伸長条件下で、第1セットのプライマーと免疫受容体を発現している免疫細胞由来の再構成された核酸サンプル及び／又は無細胞系DNAサンプルとを組み合わせることと〔ここで、第1セットのそれぞれのプライマーは受容体特異的な部分を有し、このとき受容体特異的な部分は、所定の位置にて異なる再構成核酸にアニーリングし、伸長して第1伸長産物を生成し、第1セットのそれぞれのプライマーは、第1のプライマー結合部位を含有する5'-非相補的末端を有する〕；（b）反応混合物から、伸長していない第1セットのプライマーを除去することと；（c）プライマー伸長条件下で、反応混合物に第2セットのプライマーを添加することと〔ここで、第2セットのそれぞれのプライマーは、受容体特異的な部分を有し、このとき受容体特異的な部分は、所定の位置にて第1の伸長産物にアニーリングし、第2のプライマー結合部位を含有する5'-非相補的末端を有し、第1セットのプライマー及び／又は第2セットのプライマーは、それぞれ、受容体特異的な部分と第1又は第2のプライマー結合部位との間に配置された配列タグを含み、第2セットのそれぞれのプライマーは、伸長して第2の伸長産物を生成し、このとき、それぞれの第2の伸長産物は、第1のプライマー結合部位、第2のプライマー結合部位、少なくとも1つの配列タグ、及び免疫細胞受容体鎖の部分をコードしている再構成された核酸を含む〕、を含む。いくつかの実施形態では、再構成された核酸を結合させる工程は、このような配列リードの尤度が明らかに少なくとも99パーセントであるときは常に；並びに、その他の実施形態では、尤度が明らかに少なくとも99.9パーセントであるときには常に、異なる再構成核酸の配列リードを結合させることを含む。

#### 【0107】

本発明の方法は、非リンパ系癌又は非骨髄癌などといった、患者の癌の微小残存病変のモニタリングにも応用可能であり、方法は、例えば、選択された癌遺伝子のセットにおいて、変異のパターンを特定することを含む。このような変異のパターン、すなわち、このような変異を含有している遺伝子の存在、非存在、及び／又はレベルは、疾患再発の尤度を示し得る。いくつかの実施形態では、このようなモニタリングのための標的ポリヌクレオチドは、エキソン、エキソンの部分、選択されたイントロン及び／又は遺伝子発現制御領域、例えば、複数の遺伝子のプロモーター（本明細書において「癌遺伝子分子」として言及する）とすることができる。癌遺伝子分子は、エキソン捕捉法〔例えば、TruSeq（商標）exome enrichment kit（Illumina, San Diego, CA）；Frampton et al, Nature Biotechnology 31（11）：1023～1031（2013）〕などの従来法を用い、組織サンプルから単離することができる。このような癌遺伝子分子を得た後、配列タグを付加してタグ-核酸複合体を生成し、本発明に準拠しタグ-核酸複合体を増幅及び配列決定する。

#### 【0108】

最新の癌ゲノム配列決定研究では、異なる癌、同じ癌に罹患している異なる患者、同じ腫瘍の細胞、及び同じ患者の異なる転移部の細胞同士で、変異パターンが著しく不均質であることが示されている；しかしながら、同じ患者の場合、典型的には、共通の祖先細胞から不均質な癌細胞が進化するため、それらは変異を共有しており、癌細胞同士の進化的な関連性は、経時的に連続的に測定することにより識別することもできる〔例えば、Vo

gelstein et al, Science, 339:1546~1558(2013);及びDing et al, Nature, 481(7382):506~510(2012)など];したがって、診断試料において測定される、癌に関連する変異パターンは、同じ癌の再発又はその癌がクローン的に進化した癌を検出するための手立てを提供する。

#### 【0109】

癌遺伝子分子は、限定するものではないが、表Iに記載の遺伝子などの、様々な遺伝子から選択することができる。

#### 【0110】

#### 【表1】

表I 癌遺伝子の例

ABL1	AKT1	ALK	APC	ATM
BRAF	CDH1	CSF1R	CTNNB1	EGFR
ERBB2	ERBB4	FBXW7	FGFR1	FGFR2
FGFR3	FLT3	GNA11	GNAC	GNAS
HNF1A	HRAS	IDH1	JAK2	JAK3
KDR	KIT	KRAS	MET	MLH1
MPL	NOTCH1	NPM1	NRAS	PGGFRA
PIK3CA	PTEN	PTPN11	RB1	RET
SMAD4	SMO	SRC	STK	TP53
VHL				

#### 【0111】

いくつかの実施形態では、上記の癌の微小残存病変をモニタリングする方法は、次の工程：(a)個人から組織サンプルを得る工程；(b)サンプル中の複数のそれぞれの癌遺伝子分子に配列タグを付加して、タグ-核酸複合体を生成する工程（ここで、少なくとも1つの核酸又はそれらのコピーは異なる配列タグが付加されており、癌遺伝子分子は、個人の癌に特徴的なものである）；(c)タグ-核酸複合体を増幅させる工程；(d)タグ-核酸複合体サンプルを配列決定して、エラー率を有しタグ配列及び癌遺伝子配列を含む配列リードを形成する工程；(e)同様のタグ配列を有する配列リードをアライメントして、同じ配列タグを有する配列リードのグループを形成する工程；(f)グループの癌遺伝子配列を結合させて癌遺伝子分子の配列を決定する工程（ここで、癌遺伝子配列グループの尤度が明らかに少なくとも95%であるときは常に、配列リードのグループを、異なる癌遺伝子分子に結合させる）；並びに(g)癌遺伝子分子のプロファイルにおいて、個人の癌に特徴的な癌遺伝子分子の存在、非存在、及び/又はレベルを検出する工程、を含み得る。いくつかの実施形態では、癌遺伝子配列を結合させる工程は、このような配列リードが尤度少なくとも99パーセントで異なるときは常に；並びに、その他の実施形態では、尤度少なくとも99.9パーセントで異なるときには常に、異なる癌遺伝子分子の配列リードを結合させることを含む。

#### 【0112】

キャリアオーバー汚染を検出するための配列タグの利用

キャリアオーバー汚染は、核酸増幅を含む手法に関し重大な問題である[例えば、Borst et al, Eur. J. Clin. Microbiol. Infect. Dis., 23(4):289~299(2004); Aslantzadeh, Ann. Clin. Lab. Sci., 34(4):389~396(2004)など]。このような汚染は、サンプルのアッセイにおいて、サンプルのものではない微量の核酸が非意図的に増幅されたときに生じ、測定結果に影響を及ぼす。最悪の場合、患者由来の医学的サンプルにおけるキャリアオーバー汚染では、アッセイ結果において疑陽性の解釈がもたらされ

る場合もある。サンプルのものではない核酸は、特定の患者とは関係のない供給源に由来する場合もあり；例えば、別の患者のサンプルに由来する場合もある。あるいは、サンプルのものではない核酸は、患者に関係する供給源に由来する場合があり；例えば、過去に同じ検査室で扱われていた、同じ患者に由来する別のサンプルに由来する場合があり、あるいは過去に同じ検査室で処理していた、同じ患者の別のサンプルに対するアッセイ反応物に由来する場合がある。

#### 【0113】

キャリアオーバー汚染は、臨床の場で、T細胞受容体又は免疫グロブリンなどの免疫分子をコードしている再構成された核酸群などといった、関連する核酸の非常に複雑な集団を評価するときに、特に問題となる。この問題は、配列リード又はクロノタイプが、意図するサンプルに関係する実際の多様性の要素であるのか、あるいはそれらが、サンプルとは異なる核酸源（別の患者のサンプル、又は同じ検査室において、同種のアッセイで処理された、同じ患者の以前のサンプルなど）に由来するものなのかを判断することが難しいために生じる。本発明の一態様では、配列タグを利用して、配列リードをもとにクロノタイプを決定するだけでなく、配列タグが現在のサンプルに由来するものなのか、あるいは別のサンプルに由来するものなのかを評価することにより、このようなキャリアオーバー汚染を検出できる。このような検出は、それぞれの患者サンプルから決定された配列タグの記録を維持した後、以降の測定がなされたときには常に、最新の測定の配列タグを、それまでに測定されたものと比較することにより達成される。それぞれの測定に由来するタグの数が多いこと、並びに従来のアルゴリズムを用いれば電子記録の検索及び比較が容易であることから、このようなクロノタイプに関連する配列タグの記録は、電子記録として大容量記憶装置に便利に維持される。測定に利用する配列タグ群が十分に大規模であるならば、マッチがみられた場合にキャリアオーバー汚染が生じている可能性は非常に高い。上記のサンプリングによる標識のための、クロノタイプ群に対する配列タグ群のサイズ比についての同様の例は、キャリアオーバー汚染の検出に応用可能である。一実施形態では、このような比は100：1以上である。

#### 【0114】

様々な検索方法又はアルゴリズムを利用して、評価するクロノタイプをデータベースのクロノタイプと比較する工程を実施することもできる。従来の多くの配列アライメント法及び検索アルゴリズムを公的に利用可能であり、これらは、参照により援用される以下の参照文献において記載されている：Mount, Bioinformatics Sequence and Genome Analysis, Second Edition (Cold Spring Harbor Press, 2004); Batzoglou, Briefings in Bioinformatics, 6:6~22 (2005); Altschul et al, J. Mol. Biol., 215(3):403~410 (1990); Needleman and Wunsch, J. Mol. Biol., 48:443~453 (1970);並びにSmith and Waterman, Advances in Applied Mathematics, 2:482~489 (1981) など。

#### 【0115】

いくつかの実施形態では、異なるサンプルに由来する材料によるサンプルの汚染、例えば、キャリアオーバー汚染などの汚染を、検出及び測定するための上記方法は、次の工程：(a) 個人の組織サンプルから得る工程；(b) 癌遺伝子分子又は再構成された核酸に配列タグを取り付けてタグ-核酸複合体を生成する工程（少なくとも1つの核酸又はそれらのコピーは、取り付けられた異なる配列タグを有し、癌遺伝子分子は、個人の癌に特徴的なものである）；(c) タグ-核酸複合体を増幅させる工程；(d) タグ-核酸複合体サンプルを配列決定して、それぞれエラー率を有し、それぞれタグ配列と、癌遺伝子配列又は再構成された核酸配列とを含む配列リードを提供する工程；(e) タグ配列を、他の組織サンプルから別途決定されたタグ配列と比較する工程；並びに(f) 1つ以上のタグ配列のアイデンティティの存在、非存在、及び/又はそれによる汚染の程度を、その他の

組織サンプルから別途決定されたタグ配列により評価する工程、を含み得る。アッセイにおいてタグ配列が決定されたならば、それらのタグ配列を、その他の患者に対するアッセイにより記録されたタグ配列のデータベース中のタグ配列と比較することもできる。このような比較工程は、アッセイ時に実施することもでき、あるいはこのような工程は、例えば、アッセイ後にレトロスペクティブに実施することもできる。一実施形態では、配列タグは、リンパ系組織の癌などのリンパ増殖性疾患に罹患している個人に由来する組織サンプル（例えば、血液又は骨髄）において再構成された核酸に付加される。別の実施形態では、上記のものなどの癌遺伝子分子に配列タグを付加する。

#### 【0116】

組織サンプルの相互汚染について再構成された核酸がモニターされる更なる実施形態では、付加及び増幅させる工程は次の通り：（a）反応混合物中、プライマー伸長条件下で、第1セットのプライマーとT細胞由来の再構成された核酸サンプル及び/又は無細胞系DNAサンプルとを組み合わせることと[ここで、第1セットのそれぞれのプライマーは受容体特異的な部分を有し、このとき受容体特異的な部分は、所定の位置にて異なる再構成核酸にアニーリングし、伸長して第1伸長産物を生成し、第1セットのそれぞれのプライマーは、第1のプライマー結合部位を含有する5'-非相補的末端を有する]；（b）反応混合物から、伸長していない第1セットのプライマーを除去することと；（c）プライマー伸長条件下で、反応混合物に第2セットのプライマーを添加することと[ここで、第2セットのそれぞれのプライマーは、受容体特異的な部分を有し、このとき受容体特異的な部分は、所定の位置にて第1の伸長産物にアニーリングし、第2のプライマー結合部位を含有する5'-非相補的末端を有し、第1セットのプライマー及び/又は第2セットのプライマーは、それぞれ、受容体特異的な部分と第1又は第2のプライマー結合部位との間に配置された配列タグを含み、第2セットのそれぞれのプライマーは、伸長して第2の伸長産物を生成し、このとき、それぞれの第2の伸長産物は、第1のプライマー結合部位、第2のプライマー結合部位、少なくとも1つの配列タグ、及び免疫受容体鎖の部分をコードしている再構成された核酸を含む]、（d）反応混合物においてポリメラーゼ連鎖反応を実施して、アンプリコンを生成することと[このポリメラーゼ連鎖反応は、第1のプライマー結合部位に特異的な順方向プライマーと、第2のプライマー結合部位に特異的な逆方向プライマーとを使用する]により実施される。

#### 【0117】

##### キット

本発明は、本発明の方法を実施するための様々なキットを含む。いくつかの実施形態では、キットは、（a）多重PCRにおいて、複数の免疫受容体鎖をコードしている再構成された核酸を増幅させるための順方向プライマーのセット及び逆方向プライマーのセットと[ここで、それぞれの順方向プライマー及び/又は逆方向プライマーは標的特異的な部分、配列タグ、及び共通のプライマー結合部位を有する]、（b）セットのプライマーのうち、少なくとも最初の伸長後に組み込まれていないプライマー（すなわち、伸長していないプライマー）を除去するためのプライマー除去成分と、を含む。いくつかの実施形態では、キットは、共通のプライマー結合部位に特異的な共通のプライマーを更に含む。いくつかの実施形態では、キットは、本発明の方法でキット構成要素を使用するための使用説明書を更に含む。いくつかの実施形態では、キットは、IGH(VDJ)、IGH(DJ)、及びIGKをコードしている再構成された核酸を増幅させるため、順方向及び逆方向に特異的なプライマーを更に含む。いくつかの実施形態では、キットは、TCR、TCR及びTCRをコードしている再構成された核酸を増幅させるため、順方向及び逆方向に特異的なプライマーを更に含む。いくつかの実施形態では、キットは、標的再構成された核酸を再現する長さ及び組成を有する複数の核酸を含む内部標準を更に含み、内部標準は予め判明している濃度で提供される。いくつかの実施形態では、キットは、プライマー除去成分として、大腸菌(E. coli)エキソヌクレアーゼIなどの一本鎖エキソヌクレアーゼを含む。いくつかの実施形態では、キットは、サイズをもとに二本鎖DNAを選別することのできるスピンカラムをプライマー除去成分として含む。



## 【0118】

いくつかの実施形態について具体例を参照し、本発明について記載したが、当業者であれば、本発明の趣旨及び範囲から逸脱せずとも多くの変更を加えることができることを認識されるであろう。本発明は、上記の開示に加えて様々なセンサーの実施及びその他の主題に応用可能である。

## 【0119】

## 定義

別途記載のない限り、本明細書で使用する核酸化学、生化学、遺伝学、及び分子生物学に関する用語及び記号は、当該技術分野の標準的な論文及び文書、例えば、Kornberg and Baker, DNA Replication, Second Edition (W. H. Freeman, New York, 1992); Lehninger, Biochemistry, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human Molecular Genetics, Second Edition (Wiley-Liss, New York, 1999); Abbas et al, Cellular and Molecular Immunology, 6th edition (Saunders, 2007) に従うものである。

## 【0120】

「アライメント」は、配列距離測定法に基づき、配列リードなどの試験配列を1つ以上の参照配列と比較して、どの参照配列が、あるいは参照配列のどの位置が、最も近いのかを決定することを意味する。ヌクレオチド配列をアライメントさせる方法例としては、Smith Waterman アルゴリズムがある。距離測定法としては、ハミング距離、又はレーベンシュタイン距離などを挙げることができる。距離測定法には、比較する配列のヌクレオチドのクオリティスコアに関する要素を含ませることができる。

## 【0121】

「アンプリコン」は、ポリヌクレオチド増幅反応の生成物；すなわち、ポリヌクレオチドのクローン集団を意味し、一本鎖核酸又は二本鎖核酸であってよく、ポリヌクレオチドは1つ以上の出発配列から複製される。1つ以上の出発配列は、同じ配列の1つ以上のコピーであってよく、あるいはそれらは異なる配列の混合物であってよい。アンプリコンは、様々な増幅反応により生成することができ、この反応生成物は、出発核酸又は標的核酸の複製物を1つ以上含む。一態様では、アンプリコンを生成する増幅反応は、「テンプレートをもとに実施」される。この反応では、反応生成物、すなわちヌクレオチド又はオリゴヌクレオチドのいずれかの塩基対形成は、反応生成物の生成に必要とされるテンプレートポリヌクレオチドと相補的なものである。一態様では、テンプレートをもとに実施される反応は、核酸ポリメラーゼによるプライマー伸長、又は核酸リガーゼによるオリゴヌクレオチドライゲーションである。このような反応としては、限定するものではないが、ポリメラーゼ連鎖反応法 (PCR)、線状ポリメラーゼ反応法 (linear polymerase reaction)、核酸配列ベース増幅法 (NASBA)、及びローリングサークル増幅法などが挙げられ、参照により本明細書に援用される以下の参考文献：Mullis et al, 米国特許第4,683,195号；同第4,965,188号；同第4,683,202号；同第4,800,159号 (PCR)；Gelfand et al, 米国特許第5,210,015号 (「taqman」プローブによるリアルタイムPCR)；Wittwer et al, 米国特許第6,174,670号；Kacian et al, 米国特許第5,399,491号 (「NASBA」)；Lizardi, 米国特許第5,854,033号；Aono et al, 特開平第4-262799号 (ローリングサークル増幅) などに開示される。一態様では、本発明のアンプリコンはPCRにより生成される。増幅反応の進行とともに反応生成物を測定することができる検出化学反応が利用可能なものである場合、例えば、下記の「リアルタイムPCR」又はLeone et al, Nucleic Acids Research, 26:2150~2155 (1998) 及び同様の参考文献に記載の通りの「リアルタイムNASBA」などの増幅反応は「

リアルタイム」増幅とすることもできる。本明細書で使用する時、用語「増幅」は、増幅反応を実施することを意味する。「反応混合物」は、反応を実施させるのに必要な反応物をすべて含有している溶液を意味し、このようなものとしては、限定するものではないが、反応中に選択した程度にpHを維持する緩衝剤、塩類、補助因子、及びスカベンジャーなどを挙げるができる。

#### 【0122】

本明細書で使用する時、「クローナリティー」は、レパトアのクロナタイプにおいて、クロナタイプの豊富さについての分布を単一又は少数のクロナタイプに傾斜させる尺度を意味する。おおまかにいうと、クローナリティーは、クロナタイプ多様性とは逆の尺度である。生態学から、種の豊富さの関係性について説明する多くの尺度及び統計が利用可能である。例えば、Pielou著、「数理生態学序論 (An Introduction to Mathematical Ecology)」、第17章及び第18章 (Wiley-Interscience, 1969年)の考案に従い、関係性について説明する尺度及び統計を利用して、クローナリティーを評価することができる。一態様では、本発明で使用されるクローナリティーの尺度はクロナタイププロファイルの関数であり(すなわち、検出されることとなるクロナタイプの数及びそれらの豊富さ)、クロナタイププロファイルの評価後に、クロナタイププロファイルからクローナリティーを算出して、単一の数値を得ることもできる。クローナリティー尺度の1つとしては、単純に、ランダムに選択した2つのクロナタイプが同じである確率を求める、シン普森尺度がある。その他のクローナリティー尺度としては、Pielou(上掲)に開示される情報ベースの尺度及びMcIntoshの多様性指数が挙げられる。

#### 【0123】

「クロナタイプ」は、免疫受容体又はそれらの部分をコードするリンパ球の再構成核酸を意味する。より具体的には、クロナタイプとは、通常は、T細胞又はB細胞から抽出され、場合により、T細胞受容体(TCR)若しくはB細胞受容体(BCR)又はこれらの部分をコードしている無細胞資源に由来するものであり得る、再構成核酸を意味する。様々な実施形態において、クロナタイプは、IGHのVDJ再構成配列、IGHのDJ再構成配列、IGKのVJ再構成配列、IGLのVJ再構成配列、TCRのVDJ再構成配列、TCRのDJ再構成配列、TCRのVJ再構成配列、TCRのVJ再構成配列、TCRのVDJ再構成配列、TCRのVD再構成配列、又はKde-V再構成配列のすべて又は部分をコードし得る。クロナタイプは、Bcl1-IGH又はBel1-IGHなどの免疫受容体遺伝子を含む転座切断領域もコードし得る。一態様では、クロナタイプは、それらの由来する免疫分子の多様性を再現又は反映させるのに十分長い配列を有する；結果的に、クロナタイプの長さは非常に多様になり得る。いくつかの実施形態では、クロナタイプは、25~400ヌクレオチドの範囲の長さを有してよく；並びに、その他の実施形態では、クロナタイプは、25~200ヌクレオチドの範囲の長さを有する。

#### 【0124】

「クロナタイププロファイル」は、リンパ球集団に由来する様々なクロナタイプ及び相対的豊富さを列挙することを意味し、例えば、相対的豊富さは、所与の集団における頻度として表現することができる(すなわち、0~1の間の数)。典型的には、リンパ球集団は組織サンプルから得る。用語「クロナタイププロファイル」は、次の：Arstila et al, Science, 286:958~961(1999)；Yassai et al, Immunogenetics, 61:493~502(2009)；並びにKedzierska et al, Mol. Immunol., 45(3):607~618(2008)などの参考文献に記載の通りの免疫「レパトア」についての免疫概念に関連するが、より一般的なものである。用語「クロナタイププロファイル」には、様々なリスト及び免疫受容体をコードする再構成された核酸の豊富さが含まれ、かかるプロファイルは選択したリンパ球サブセット(例えば、組織浸潤性リンパ球、又は免疫表現性のサブセットなど)に由来し得るものであり、又は完全な免疫受容体と比較して多様性の

低減されている免疫受容体部分をコードし得るものである。いくつかの実施形態では、クロノタイププロファイルは、少なくとも  $10^3$  個の異なるクロノタイプを含み得る；他の実施形態では、クロノタイププロファイルは、少なくとも  $10^4$  個の異なるクロノタイプを含み得る；他の実施形態では、クロノタイププロファイルは、少なくとも  $10^5$  個の異なるクロノタイプを含み得る；他の実施形態では、クロノタイププロファイルは、少なくとも  $10^6$  個の異なるクロノタイプを含み得る。このような実施形態では、このようなクロノタイププロファイルは、異なるそれぞれのクロノタイプについて、豊富さ又は相対頻度を更に含ませることもできる。一態様では、クロノタイププロファイルは、個人のリンパ球集団において、それぞれT細胞受容体（TCR）若しくはB細胞受容体（BCR）又はこれらの断片をコードしている、（豊富さが）異なる組み換えヌクレオチド配列セットであり、ここで、セットのヌクレオチド配列は、集団のリンパ球の実質的にすべてに関し、異なるリンパ球又はそれらのクローン性サブ集団との1対1対応を有する。一態様では、個人の实質的にすべてのT細胞又はB細胞又はそれらのクローンに、このようなレパトアの固有の核酸配列を輸送させるべく、クロノタイプの多様性（すなわち、セット中の異なる核酸配列の数）が十分に大きくなるよう、クロノタイプを定義する核酸セグメントを選択する。すなわち、好ましくは、サンプルのそれぞれの異なるクローンは異なるクロノタイプを有する。本発明のその他の態様では、レパトアに対応するリンパ球集団は循環B細胞であってよく、又は循環T細胞であってよく、又は限定するものではないがCD4 + T細胞若しくはCD8 + T細胞などの前述の集団のいずれかのサブ集団、若しくは細胞表面マーカーにより定義されるその他のサブ集団などであってよい。このようなサブ集団は、特定の組織、例えば、骨髄又はリンパ節などからサンプルを採取することにより獲得することもでき、あるいは1つ以上の細胞表面マーカー、大きさ、又は形態などに基づき、サンプル（末梢血など）から細胞をソートすること又は富化させることにより獲得することもできる。更に他の態様では、レパトアに対応するリンパ球集団は、腫瘍組織又は感染組織などの疾患組織に由来するものであってもよい。一実施形態では、ヒトTCR鎖又はそれらのフラグメントを含むクロノタイププロファイルは、異なるヌクレオチド配列を  $0.1 \times 10^6 \sim 1.8 \times 10^6$  の範囲、又は  $0.5 \times 10^6 \sim 1.5 \times 10^6$  の範囲、又は  $0.8 \times 10^6 \sim 1.2 \times 10^6$  の範囲で数多く含む。別の実施形態では、ヒトIGH鎖又はそれらのフラグメントを含むクロノタイププロファイルは、異なるヌクレオチド配列を  $0.1 \times 10^6 \sim 1.8 \times 10^6$  の範囲、又は  $0.5 \times 10^6 \sim 1.5 \times 10^6$  の範囲、又は  $0.8 \times 10^6 \sim 1.2 \times 10^6$  の範囲で数多く含む。特定の実施形態では、本発明のクロノタイププロファイルは、IGH鎖のV(D)J領域のセグメントを実質的にすべてコードしているヌクレオチド配列のセットを含む。本明細書において使用するとき、一態様では、「実質的にすべて」は、相対的な豊富さが0.001%以上であるセグメントのすべてを意味し；又は本明細書において使用するとき、別の態様では、「実質的にすべて」は、相対的な豊富さが0.0001%以上であるセグメントのすべてを意味する。別の具体的な実施形態では、本発明のクロノタイププロファイルは、TCR鎖のV(D)J領域の実質的にすべてのセグメントをコードしているヌクレオチド配列のセットを含む。別の実施形態では、本発明のクロノタイププロファイルは、長さがヌクレオチド25 ~ 200塩基の範囲であり、かつTCR鎖のV、D、及びJ領域のセグメントを含む、ヌクレオチド配列のセットを含む。別の実施形態では、本発明のクロノタイププロファイルは、長さがヌクレオチド25 ~ 200塩基の範囲であり、かつIGH鎖のV、D、及びJ領域のセグメントを含む、ヌクレオチド配列のセットを含む。別の実施形態では、本発明のクロノタイププロファイルは、異なるヌクレオチド配列を多数含み、この数は、異なるIGH鎖を発現しているリンパ球の数と実質的に等しい。別の実施形態では、本発明のクロノタイププロファイルは、異なるヌクレオチド配列を多数含み、この数は、異なるTCR鎖を発現しているリンパ球の数と実質的に等しい。更に別の実施形態では、「実質的に等しい」は、クロノタイププロファイルが、個人のすべてのリンパ球集団により0.001%以上の頻度で輸送又は発現されるIGH若しくはTCR鎖又はそれらの部分をコードしているヌクレオチド配列を99%の確率で含有していることを意味する。更に

10

20

30

40

50

別の実施形態では、「実質的に等しい」はレパトアのヌクレオチド配列が、0.0001%以上の頻度で存在するすべてのリンパ球細胞により輸送又は発現されるIgH若しくはTCR又はそれらの部分をコードしているヌクレオチド配列を99%の確率で含有していることを意味する。いくつかの実施形態では、クロノタイププロファイルは、 $10^5 \sim 10^7$ 個のリンパ球から構成されるサンプルに由来するものである。このようなリンパ球の数は、1~10mLの末梢血サンプルから得ることができる。

#### 【0125】

「相補性決定領域」(CDR)は、免疫グロブリン(すなわち、抗体)又はT細胞受容体の領域であって、それらの分子が、抗原の立体構造を補完することにより自身の備える特異性を決定し、かつ特異的な抗原と接触する、領域を意味する。T細胞受容体及び免疫グロブリンはそれぞれのCDRを3つ有する: CDR1及びCDR2は、可変(V)ドメインで見られ、CDR3はある程度のVドメインと、すべての多様性(D)ドメイン(重鎖のみ)及び連結ドメイン(J)と、ある程度の定常(C)ドメインと、を含む。

#### 【0126】

「クロノタイプデータベース」は、検索、比較、及び回収を容易にし、かつ速度を上げるためにフォーマット化し、用意したクロノタイプコレクションを意味する。いくつかの実施形態では、クロノタイプデータベースは、免疫受容体の同じ領域又はセグメントをコードしているクロノタイプのコレクションを含む。いくつかの実施形態では、クロノタイプデータベースは、複数の個人に由来するクロノタイプのクロノタイププロファイルを含む。いくつかの実施形態では、クロノタイプデータベースは、少なくとも10名の個人に由来する少なくとも $10^4$ 個のクロノタイプのクロノタイププロファイルを含む。いくつかの実施形態では、クロノタイプデータベースは少なくとも $10^6$ 個のクロノタイプ、又は少なくとも $10^8$ 個のクロノタイプ、又は少なくとも $10^9$ 個のクロノタイプ、又は少なくとも $10^{10}$ 個のクロノタイプを含む。IMGTデータベース([www.imgt.org](http://www.imgt.org))などといった、クロノタイプを含む公共のデータベースをクロノタイプデータベースとしてもよく、例えば、Nucleic Acids Research, 31:307~310(2003)に記載されている。クロノタイプデータベースはFASTAフォーマットであってよく、BLASTアルゴリズム、例えば、Altschul et al., J. Mol. Biol., 215(3):403~410(1990)などのアルゴリズムを利用して、クロノタイプデータベースのエントリーを検索又は比較することもできる。

#### 【0127】

「結合」は、異なる配列を有する2つのクロノタイプ候補を、このような違いが経験誤差及び実験誤差に起因するものであり、本当に生物学的な違いに起因するものではないものとして判断することにより、同じものとして処理することを意味する。一態様では、高頻度のクロノタイプ候補の配列を低頻度のクロノタイプ候補のものと比較して、所定の基準が満たされた場合には、低頻度のクロノタイプ候補の数を高頻度のクロノタイプ候補の数に加え、以降、この低頻度クロノタイプ候補を無視する。すなわち、低頻度クロノタイプ候補に関連付けられるリードカウントを高頻度のクロノタイプ候補のリードカウントに加え、この高頻度のクロノタイプ候補と低頻度クロノタイプ候補を同じものとして扱う; すなわち、これらの間に観察された差異はエラーによるものであると判断する(例えば、配列決定エラー又は増幅エラーなど)。いくつかの実施形態では、所定の基準は、比較するクロノタイプ候補の相対頻度、候補の異なる位置の数、及びその位置のクオリティスコアなどの因子に応じる、尤度関数である。

#### 【0128】

「相補性決定領域」(CDR)は、免疫グロブリン(すなわち、抗体)又はT細胞受容体の領域であって、それらの分子が、抗原の立体構造を補完することにより自身の備える特異性を決定し、かつ特異的な抗原と接触する、領域を意味する。T細胞受容体及び免疫グロブリンはそれぞれのCDRを3つ有する: CDR1及びCDR2は、可変(V)ドメインで見られ、CDR3はある程度のVドメインと、すべての多様性(D)ドメイン(重

10

20

30

40

50

鎖のみ)及び連結ドメイン( J )と、ある程度の定常( C )ドメインと、を含む。

【 0 1 2 9 】

本明細書で使用するとき、「汚染」は、ある個人の組織サンプル中に別の個人由来の核酸が存在していることを意味する。一態様では、「汚染」は、患者のクロノタイププロファイルの解釈に影響を及ぼし得る、患者に由来しない核酸が存在することを意味する。

【 0 1 3 0 】

「遺伝子による識別」は、個人と、個人の1つ以上の遺伝子座に由来するゲノムマーカーの値(又はステータス)のセットとの間に固有の対応関係を意味する。

【 0 1 3 1 】

「遺伝子マーカー」は、個人の特定に使用することのできる遺伝子座におけるDNAの多型セグメントを意味する。遺伝子マーカーは、配列をもとに、又は隣接若しくはフランキング配列を元に識別することもできる。典型的には、遺伝子マーカーは、群の各個人において複数の配列又は値を有し得る。ゲノムマーカーの例としては、限定するものではないが、縦列型反復配列(STR)、及び一塩基多型(SNP)などが挙げられる。DNAの多型断片はゲノムDNAであってよく、又は逆転写RNAであってもよい。一実施形態では、多型断片はゲノムDNAである。一実施形態では、本発明に使用する遺伝子マーカーは、従来法を利用する増幅及び配列決定により特定される。別の実施形態では、遺伝子マーカーは、クロノタイププロファイルの生成のためのプロセス中に免疫分子とともに増幅及び配列決定される。

【 0 1 3 2 】

「内部標準」は、サンプル中の標的ポリヌクレオチドの絶対量又は相対量を定量する目的で、1つ以上の標的ポリヌクレオチドと同じ反応でプロセッシングする核酸配列を意味する。一態様では、反応はPCRなどの増幅反応である。内部標準は、内在性のものでも、加えたものでもよい。すなわち、内部標準は、サンプル中に天然に生じ得るものであり、あるいは反応前にサンプルに加えてもよい。一態様では、加える1つ以上の内部標準配列を、所定の濃度で反応混合物に加えることで、較正が得られ、増幅させた配列をこの較正と比較することで、サンプル中の対応する標的ポリヌクレオチドの量を決定することができる。加える内部標準の数、配列、長さ、及びその他の特徴は、当業者にとっては設計時に日常的な選択肢である。本明細書において、「参照配列」としても言及される内在性内部標準は、サンプルにとって天然の配列であり、なされている制御が最低限のものであり一定レベル及び細胞周期依存的なレベルの転写産物を示す遺伝子が相当する[例えば、Selvey et al, Mol. Cell Probes, 15:307~311(2001)]。内在性内部標準の例としては、限定するものではないが、次の遺伝子: GAPDH、 $\alpha_2$ -ミクログロブリン、18SリボソームRNA、及び  $\beta$ -アクチンに由来する配列が挙げられる。

【 0 1 3 3 】

「キット」は、本発明の方法を実施するために材料又は試薬を送達するための任意の送達システムを指す。本発明の方法の文脈において、このような送達システムは、ある場所から他の場所への反応試薬(例えば、適切な容器内のプライマー、酵素、内部標準など)及び/又は支持材料(例えば、緩衝剤、アッセイを実施するための使用説明書など)の保管、輸送、又は送達を可能にするシステムを含む。例えば、キットは、関連する反応試薬及び/又は指示材料を含む1つ以上の入れ物(例えば、箱)を包含する。このような内容物は、意図される容器と合わせて又は別に提供することもできる。例えば、第1の容器にはアッセイに使用する酵素を含有させ、第2の容器にはプライマーを含有させてもよい。

【 0 1 3 4 】

「微小残存病変」は、処置後に残存している癌細胞を意味する。この用語は、リンパ腫及び白血病と関連させて頻繁に使用される。

【 0 1 3 5 】

「リンパ増殖性疾患又は骨髄増殖性疾患」は、任意の異常増殖性疾患を意味し、このような疾患のモニタリングには、再構成された1つ以上の免疫受容体をコードしている1つ

10

20

30

40

50

以上のヌクレオチド配列をマーカーとして使用することができる。「リンパ腫瘍又は骨髄腫瘍」は、リンパ球又は骨髄細胞の異常増殖を意味し、悪性又は非悪性のものであり得る。リンパ系組織の癌は、悪性リンパ腫瘍である。骨髄の癌は悪性の骨髄腫である。リンパ系腫瘍及び骨髄腫瘍は、リンパ増殖性又は骨髄増殖性疾患の結果生じ、又はこれらに付随するものであり、限定するものではないが、濾胞性リンパ腫、慢性リンパ性白血病（CLL）、急性リンパ性白血病（ALL）、慢性骨髄性白血病（CML）、急性骨髄性白血病（AML）、ホジキン及び非ホジキンリンパ腫、多発性骨髄腫（MM）、良性単クローン性グロブリン血症（MGUS）、マントル細胞リンパ腫（MCL）、びまん性大細胞型B細胞リンパ腫（DLBCL）、骨髄異形成症候群（MDS）、又はT細胞リンパ腫など、例えば、Jaffe et al, Blood, 112: 4384~4399 (2008); Swerdlow et al, WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (e. 4<sup>th</sup>) (IARC Press, 2008) が挙げられる。本明細書で使用する時、「B細胞癌」は、B細胞、又は血漿細胞などのB細胞から分化する細胞に關与するリンパ腫瘍又は骨髄腫瘍を意味する。同様にして、「T細胞癌」は、T細胞、又はT細胞から分化する細胞に關与するリンパ腫瘍又は骨髄腫瘍を意味する。

#### 【0136】

参照配列及び別の配列（「比較配列」）の比較を参照する際に使用される「%相同性」又は「%同一性」などの用語は、2つの配列間の最適アライメントにおけるものを意味し、比較配列は、記載の割合の数だけサブユニット部分が参照配列のものと同一である。サブユニットは、ポリヌクレオチドの比較の際にはヌクレオチドであり、又はポリペプチドの比較の際にはアミノ酸である。本明細書で使用する時、比較する配列の「最適アライメント」は、サブユニット間の一致を最大化させ、かつアライメントの構築に用いられるギャップ数を最小に抑えるものである。%同一性は、Needleman and Wunsch, J. Mol. Biol., 48: 443~453 (1970)（「GAP」program of Wisconsin Sequence Analysis Package, Genetics Computer Group, Madison, WI）などに記載のものなどの公的に利用可能なアルゴリズムを実装して求めることができる。当業界においてアライメントを構築し、%同一性を算出するためのそのほかのソフトウェアパッケージ、又は類似するその他の測定法としては、program, based on the algorithm of Smith and Waterman, Advances in Applied Mathematics, 2: 482~489 (1981)（Wisconsin Sequence Analysis Package, Genetics Computer Group, Madison, WI）のアルゴリズムに基づく「BestFit」プログラムが挙げられる。言い換えると、例えば、参照ヌクレオチド配列に対し、少なくとも95%同一性のヌクレオチド配列を有するポリヌクレオチドを得るには、参照配列中の最大で5%までのヌクレオチドが欠失するか又は別のヌクレオチドにより置換され、あるいは参照配列中のヌクレオチドの総数の最大で5%までの数のヌクレオチドを参照配列に挿入することもできる。

#### 【0137】

「ポリメラーゼ連鎖反応」又は「PCR」は、相補的なDNA鎖の同時のプライマー伸長を促すことにより特異的なDNA配列をin vitro増幅させる反応を意味する。言い換えると、PCRは、プライマー結合部位に隣接する標的核酸のコピー又は複製物を複数製造するための反応であり、かかる反応では、次の工程：（i）標的核酸を編成させることと、（ii）プライマーをプライマー結合部位にアニーリングさせることと、（iii）ヌクレオシド三リン酸の存在下で核酸ポリメラーゼによりプライマーを伸長させることと、を1回以上繰り返すことを含む。本明細書で使用する時、用語「順方向プライマー」及び「上流プライマー」は互換的に使用され、かつ用語「逆方向プライマー」及び「下流プライマー」は互換的に使用される。同様にして、本明細書で使用する時、二本鎖標的ポリヌクレオチドが、センス鎖を左から右へ向けて5' 3'配向で示される場合

、順方向プライマーがアンチセンス鎖の左側に結合して右に向かって伸長し、逆方向プライマーがセンス鎖の右側に結合して左に向かって伸長する。通常、反応では、サーマルサイクラーにおいて、それぞれの工程について最適化させた異なる温度が周期的に繰り返される。具体的な温度、各工程の持続時間、及び各工程に移っていく速度は、当業者により良く知られている数多くの因子、例えば、参考文献：McPherson et al, editors, PCR: A Practical Approach and PCR 2: A Practical Approach (IRL Press, Oxford, それぞれ1991年及び1995年)：に例示されるものに応じ異なる。例えば、Taq DNAポリメラーゼを利用する従来のPCRでは、二本鎖の標的核酸を90 超の温度で変性させて、50～75 の範囲の温度でプライマーをアニーリングさせて、72～78 の範囲の温度でプライマーを伸長させる。用語「PCR」は、限定するものではないが、RT-PCR、リアルタイムPCR、ネステッドPCR、定量的PCR、及び多重化PCRなどといった反応の派生系を包含する。反应用量は、数百ナノリットル、例えば、200nL～数百μL、例えば、200μLの範囲である。「逆転写PCR」又は「RT-PCR」は、標的RNAを相補的一本鎖核酸DNAへと変換した後、増幅させる、逆転写反応により進行するPCRを意味する[例えば、参照により本明細書に援用されるTecott et al、米国特許第5,168,038号]。「リアルタイムPCR」は、反応の進行に伴い、反応生成物の量、すなわち、アンプリコンの量がモニターされるPCRを意味する。反応生成物をモニタリングするために利用する検出法の化学的機序が主な違いである、多くの形態のリアルタイムPCRが存在している[例えば、参照により本明細書に援用される、Gel'fand et al、米国特許第5,210,015号(「taqman」); Wittwer et al、米国特許第6,174,670号及び同第6,569,627号(インターカレート色素); Tyagi et al、米国特許第5,925,517号(分子ビーコン)]。リアルタイムPCRの検出化学反応は、参照により本明細書に援用される、Mackay et al, Nucleic Acids Research, 30:1292～1305(2002)によりレビューされている。「ネステッドPCR」は、2段階PCRを意味し、この場合、第1のPCRのアンプリコンが第2のPCRのサンプルとなり、第2のPCRでは、少なくとも1種のプライマーが第1のアンプリコンの内側の領域に結合する新しいプライマーセットを利用する。本明細書で使用するとき、ネステッド増幅反応への言及において、「最初のプライマー」は、第1のアンプリコンを生成するのに利用するプライマーを意味し、「第2のプライマー」は、第2の、又はネステッドアンプリコンを生成するのに利用する1つ以上のプライマーを意味する。「多重化PCR」は、複数の標的配列(又は単一の標的配列及び1つ以上の参照配列)が同じ反応混合物において同時に用いられるPCRを意味する[例えば、Bernard et al, Anal. Biochem., 273:221～228(1999)(2色リアルタイムPCR)]。通常、増幅させるそれぞれの配列には別個のプライマーセットを利用する。「定量的PCR」は、サンプル又は標本中の1つ以上の特異的な標的配列の豊富さを評価するよう設計されたPCRを意味する。定量的PCRには、標的配列の完全量及び相対量の両方が含まれる。定量的測定は、標的配列とは別個に又は標的配列と一緒にアッセイすることのできる1つ以上の参照配列又は内部標準を利用してなされる。参照配列は、サンプル又は標本に対して内在性のものであっても、他から加えられるものであってもよく、後者の場合、競合するテンプレートを1種以上含ませてもよい。典型的な内在性参照配列としては、次の遺伝子： - アクチン、GAPDH、<sub>2</sub> - ミクログロブリン、及びリボソームRNAなどの転写産物のセグメントが挙げられる。定量的PCRの手法は、参照により本明細書に援用される、次の参考文献：Freeman et al, Biotechniques, 26:112～126(1999); Becker-Andre et al, Nucleic Acids Research, 17:9437～9447(1989); Zimmerman et al, Biotechniques, 21:268～279(1996); Diviacco et al, Gene, 122:3013～3020(1992); Becker-Andre

10

20

30

40

50

et al, Nucleic Acids Research, 17: 9437~9446 (1989)などに例示される通り、当業者にはよく知られている。

【0138】

「プライマー」は、ポリヌクレオチドテンプレートとの二本鎖の形成により、核酸合成の開始点として機能することができ、3'末端からテンプレートに沿って伸長して、伸長された二本鎖が形成される、天然又は合成のオリゴヌクレオチドを意味する。プライマーの伸長は、通常、DNA又はRNAポリメラーゼなどの核酸ポリメラーゼにより実施される。伸長プロセスに加えるヌクレオチド配列は、テンプレートポリヌクレオチドの配列をもとに決定される。通常、プライマーはDNAポリメラーゼにより伸長される。プライマーは、通常、ヌクレオチド14~40塩基の範囲の長さを有し、あるいはヌクレオチド18~36塩基の範囲の長さを有する。プライマーは、様々な核酸増幅反応法（例えば、単一のプライマーを利用する線形的増幅反応、又は2種以上のプライマーを利用するポリメラーゼ連鎖反応）に利用される。特定の用途に関し、プライマーの長さ及び配列を選択する指針は、参照により援用される次の参考文献：Dieffenbach, editor, PCR Primer: A Laboratory Manual, 2<sup>nd</sup> Edition (Cold Spring Harbor Press, New York, 2003)により証明される通り、当業者にはよく知られている。

【0139】

「クオリティスコア」は、特定の配列位置の塩基の割当が適切なものである確率についての尺度を意味する。特定の条件について、例えば、配列決定の化学反応、検出システム、及びベースコールアルゴリズムなどが異なる結果呼び出される塩基について、クオリティスコアを算出する様々な手法（例えば、ベースコールなど）が当業者にはよく知られている。概して、クオリティスコアの値は、ベースコールが正しいものである確率に単調に関連付けられる。例えば、クオリティスコア、又はQが10であると、塩基が正しく呼び出される確率が90%であることが意味され、Qが20であると、塩基が正しく呼び出される確率が99%であることが意味され、これが他でも同様に順次該当していく。いくつかの配列決定プラットフォーム、特に合成時解読法を利用するものについては、平均クオリティスコアは、不完全な伸長、順方向の伸長の保有、テンプレートの目減り、ポリメラーゼの目減り、保護の失敗、及び脱保護の失敗などの現象に起因して、配列リードが長くなるにつれ減少するため、配列リードの開始点のクオリティスコアは配列リードの終了点のクオリティスコアよりも高くなる。

【0140】

「配列リード」は、配列決定法により生成された配列又はデータストリームから決定されたヌクレオチド配列を意味し、この決定は、例えば、手法に関係するベースコールソフトウェア、例えば、DNA配列決定プラットフォームの民間プロバイダによるベースコールソフトウェアを用いる手法によりなされる。配列リードは、通常、配列のそれぞれのヌクレオチドについてクオリティスコアを含む。典型的には、配列リードは、例えば、DNAポリメラーゼ又はDNAリガーゼにより、テンプレート核酸に沿ってプライマーを伸長させることにより作製される。データは、このような伸長に関連する光学シグナル、化学シグナル（例えば、pH変化）、又は電気シグナルなどの記録されたシグナルをもとに生成される。このような初期データを配列リードに変換する。

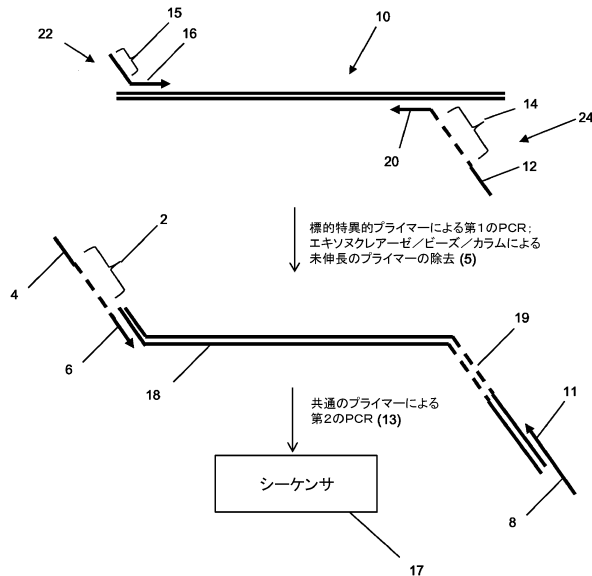
【0141】

「配列タグ」（若しくは「タグ」）又は「バーコード」は、ポリヌクレオチド又はテンプレート分子に付加させて、反応において、又は一連の反応において、ポリヌクレオチド又はテンプレートの識別及び/又は追跡に使用するオリゴヌクレオチドを意味する。それぞれの配列タグは、本明細書においてしばしば「タグ配列」として言及するヌクレオチド配列を有する。配列タグは、ポリヌクレオチド若しくはテンプレートの3'若しくは5'末端に付加することができ、又はポリヌクレオチド又はテンプレート内部に挿入して、本明細書において場合により「タグ化ポリヌクレオチド」又は「タグ化テンプレート」、又は「タグ-ポリヌクレオチド複合体」、又は「タグ-分子複合体」などとして参照される

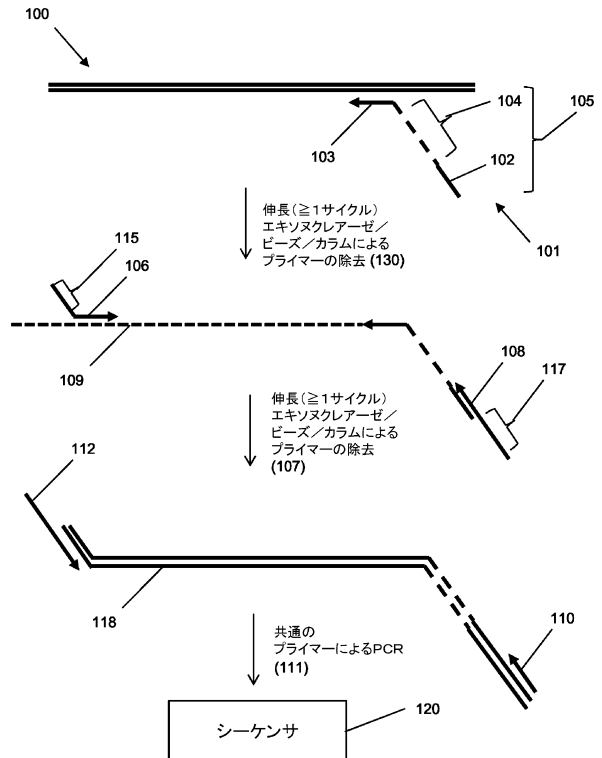


線状又は環状複合体を形成させることもできる。配列タグの大きさ及び組成は非常に様々であり得る；参照により本明細書に援用される以下の参考文献：Brenner、米国特許第5,635,400号；Brenner及びMacevicz、米国特許第7,537,897号；Brenner et al, Proc. Natl. Acad. Sci., 97:1665~1670(2000)；Church et al、欧州特許広報第0303459号；Shoemaker et al, Nature Genetics, 14:450~456(1996)；Morris et al、欧州特許公報第0799897A1号；Wallace、米国特許第5,981,179号などは、具体的な実施形態に適切な配列タグセットの選択の指針を提供する。具体的なタグの長さ及び／又は組成の選択は、限定するものではないが、タグの解読に使用する配列決定法；標的ポリヌクレオチドセットを明確に同定するために必要とされる識別可能なタグ数、信頼できる同定を確実にを行うために、セットの必須タグをどの程度まで異ならせるか、例えば、クロスハイブリダイゼーション又は配列決定エラーによる誤同定からの自由度などのいくつかの因子に依存し得る。いくつかの実施形態では、配列タグは、それぞれ、ヌクレオチド6~100塩基の範囲内、又はヌクレオチド10~100塩基の範囲内、又はヌクレオチド12~50塩基の範囲内、又はヌクレオチド12~25塩基の範囲内の長さを有する。いくつかの実施形態では、セットの各配列タグが同じセットのすべてのその他のタグとは少なくとも4塩基異なる固有のヌクレオチド配列を有する配列タグセットが使用される；他の実施形態では、セットの各配列タグが同じセットのすべてのその他のタグとは少なくとも5塩基異なる配列タグセットが使用される；更に他の実施形態では、セットの各配列タグが同じセットのすべてのその他のタグと塩基が少なくとも10%異なる配列タグセットが使用される；又は他の実施形態では、少なくとも25%の塩基が異なる配列タグセットが使用される；又は他の実施形態では、少なくとも50%の塩基が異なる配列タグセットが使用される。

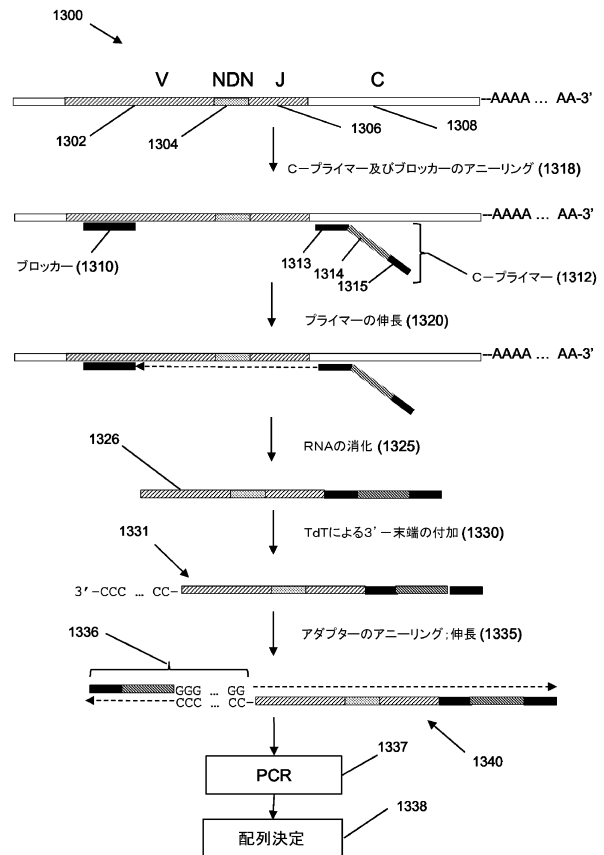
【図1A】



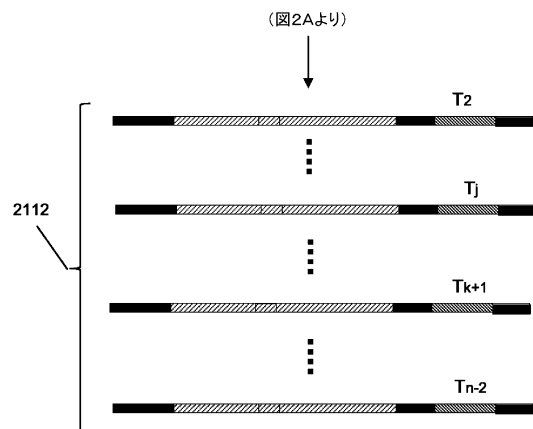
【図1B】



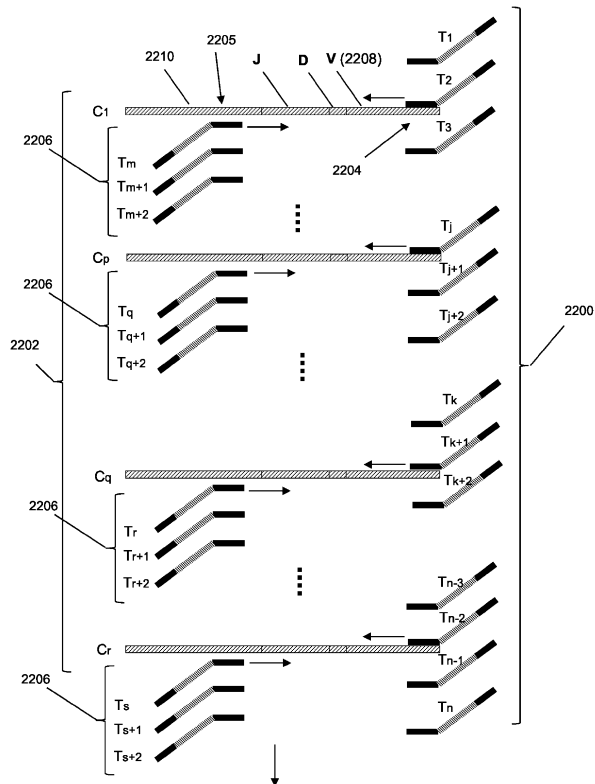
【 図 1 D 】



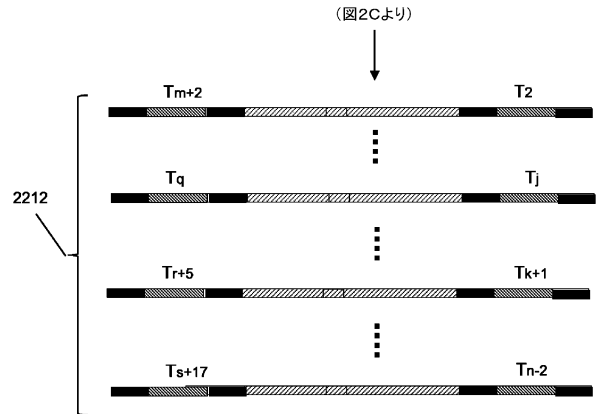
【 図 2 B 】



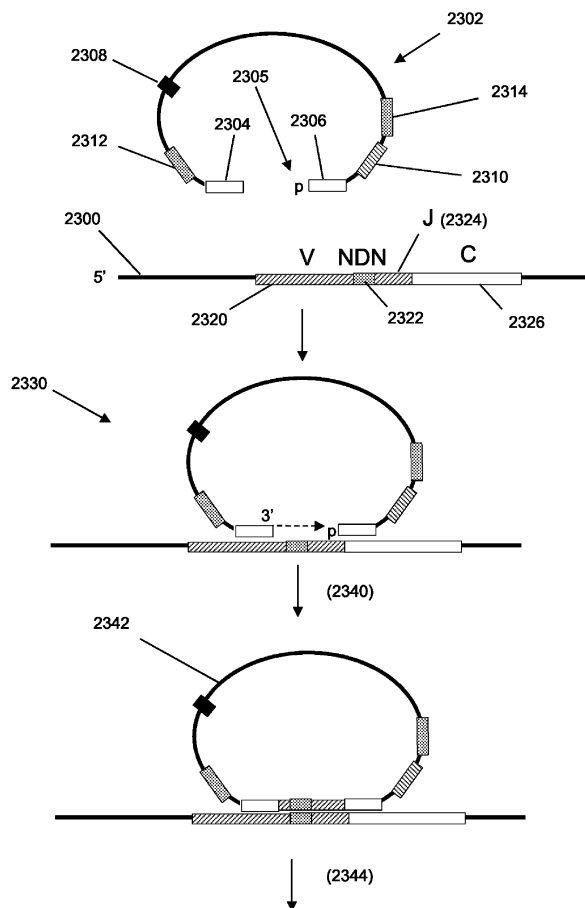
【図 2 C】



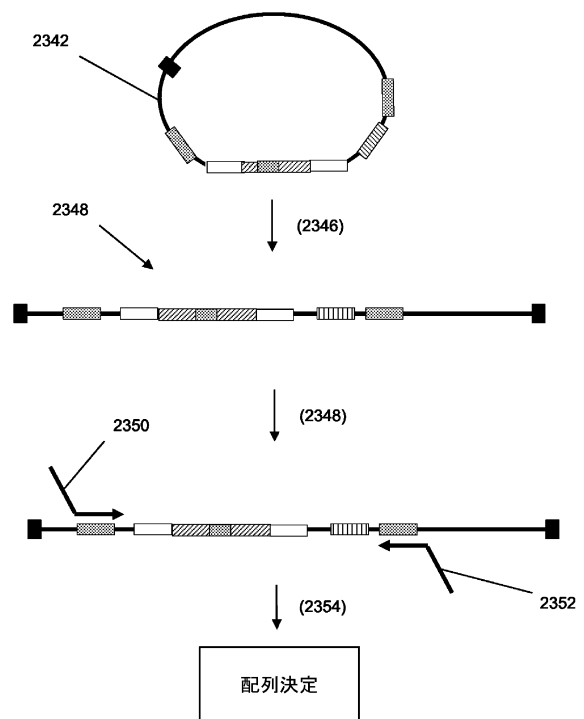
【図 2 D】



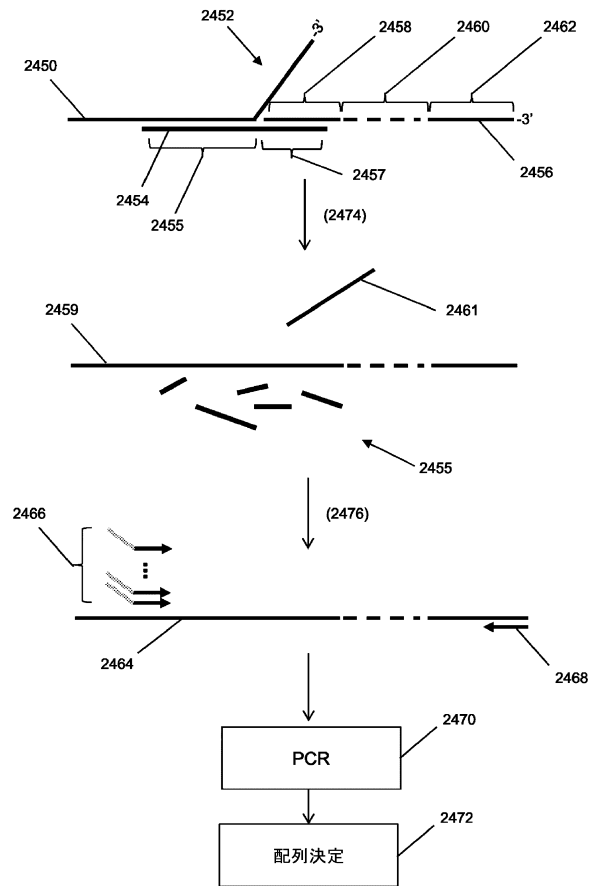
【図 2 E】



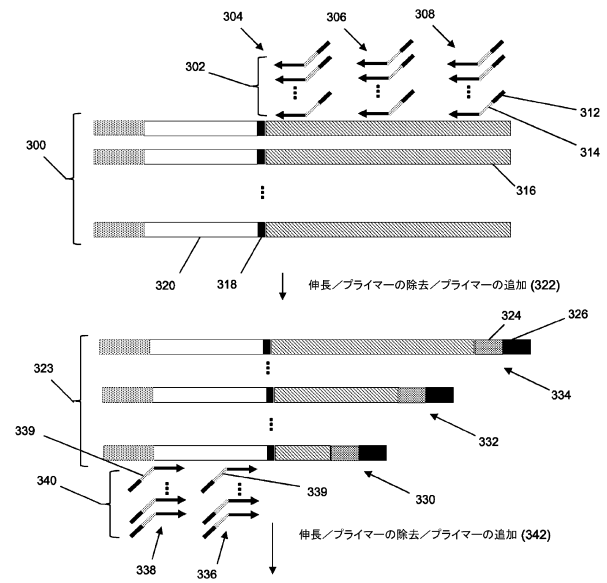
【図 2 F】



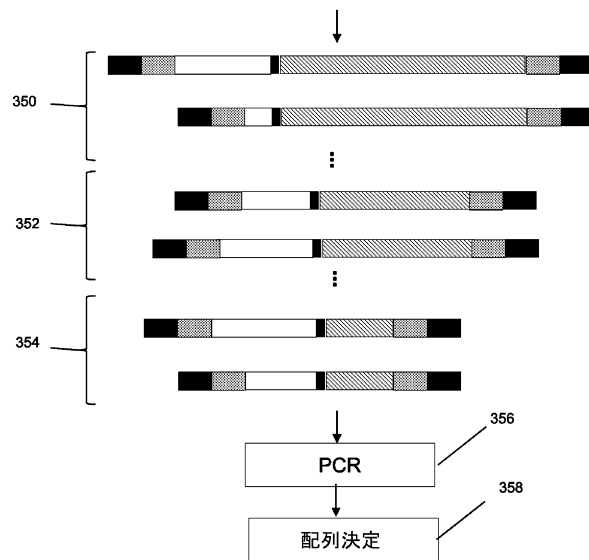
【図 2 G】



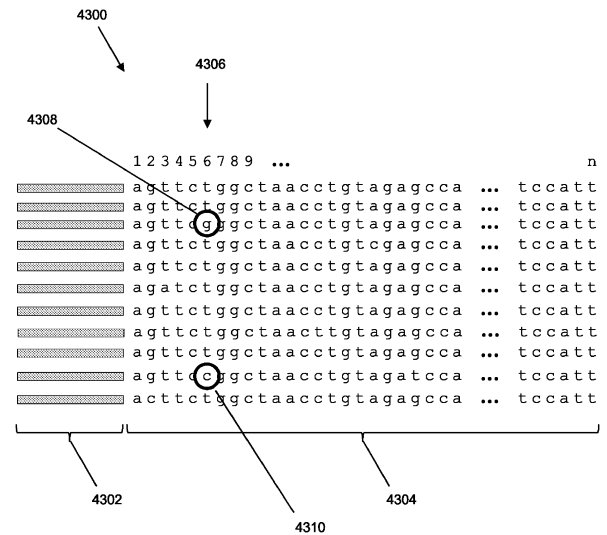
【図 3 A】



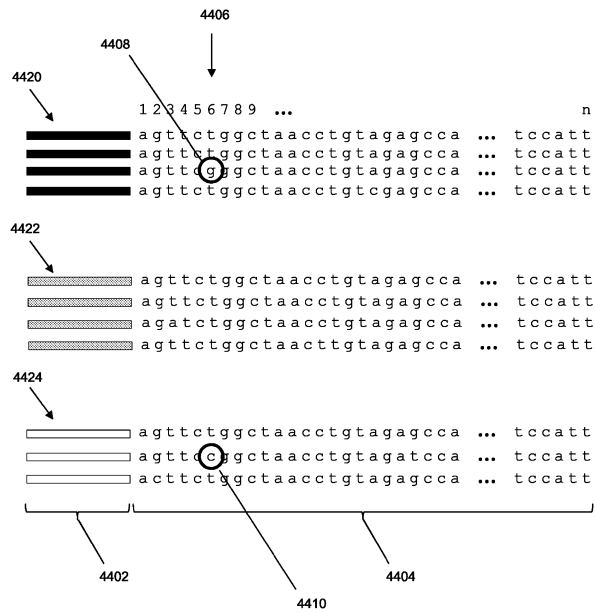
【図 3 B】



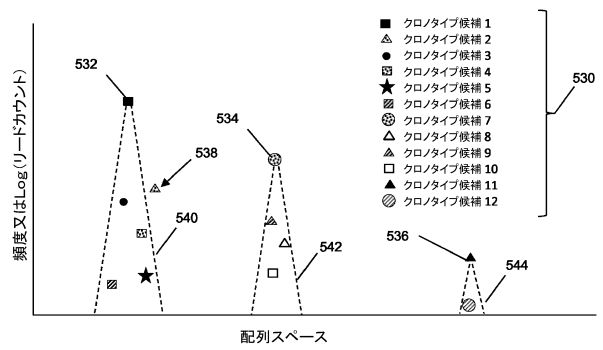
【図 4 A】



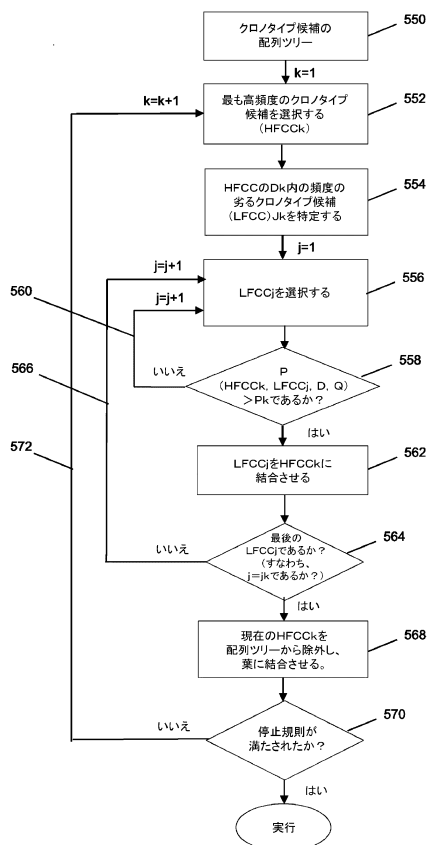
【図 4 B】



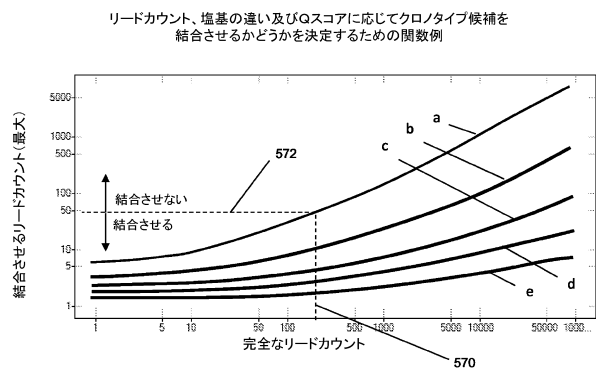
【図 5 A】



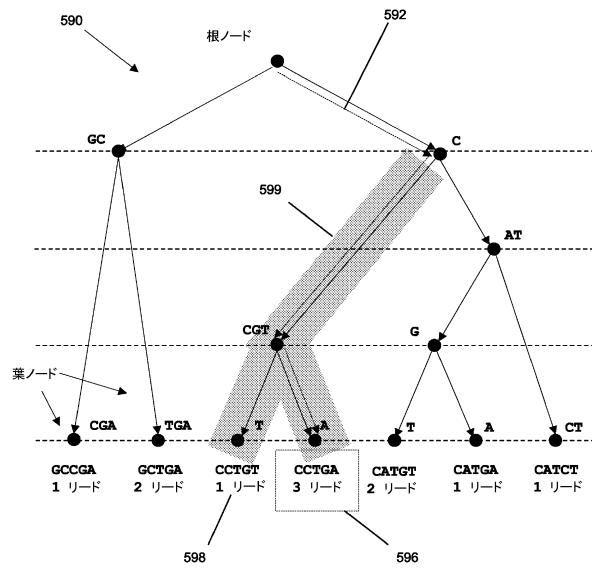
【図 5 B】



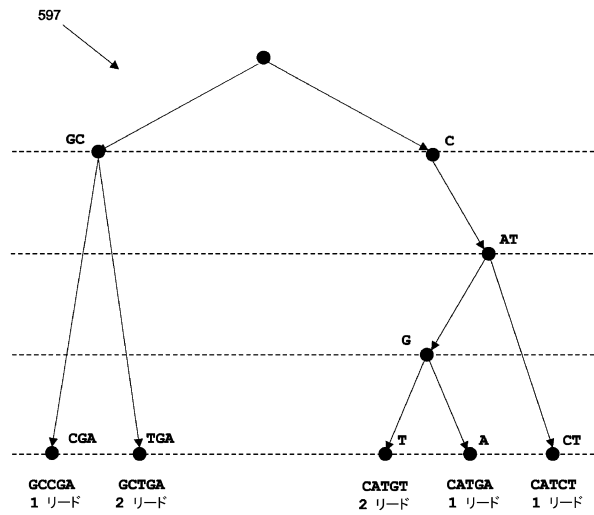
【図 5 C】



【図 5 D】



【図 5 E】



【配列表】

0006383789000001.xml

## フロントページの続き

- (72)発明者 アスブリー, トーマス  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301
- (72)発明者 ハーヴォルド, キエラン  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301
- (72)発明者 コットワリワーレ, チットラ  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301
- (72)発明者 ファハム, マレック  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301
- (72)発明者 ムーアヘッド, マーティン  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301
- (72)発明者 ウェン, リー  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301
- (72)発明者 ヴィットコップ, トバイアス  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301
- (72)発明者 ツェン, ジャンピョウ  
アメリカ合衆国 カリフォルニア州 94080, サウス サンフランシスコ イースト ジェイ  
ミー コート 400 スイート 301

審査官 西 賢二

- (56)参考文献 特表2013-524849(JP, A)  
特表2013-524848(JP, A)  
特表2012-508011(JP, A)  
国際公開第2012/142213(WO, A2)  
国際公開第2013/155119(WO, A1)  
Kinde I. et al., Detection and quantification of rare mutations with massively parallel sequencing, Proc. Natl. Acad. Sci. U.S.A., 2011年, Vol. 108, pp. 9530-9535; Supporting Information
- (58)調査した分野(Int.Cl., DB名)  
C12Q 1/68 - 1/6897  
C12N 15/00 - 15/90  
MEDLINE/EMBASE/BIOSIS(STN)