

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2015-519081

(P2015-519081A)

(43) 公表日 平成27年7月9日 (2015. 7. 9)

(51) Int. Cl.	F I	テーマコード (参考)
C 1 2 Q 1/68 (2006.01)	C 1 2 Q 1/68 Z N A A	4 B 0 2 4
C 1 2 N 15/09 (2006.01)	C 1 2 N 15/00 A	4 B 0 6 3

審査請求 未請求 予備審査請求 未請求 (全 50 頁)

(21) 出願番号	特願2015-517372 (P2015-517372)	(71) 出願人	511110533
(86) (22) 出願日	平成25年6月11日 (2013. 6. 11)		シーケンタ インコーポレイテッド
(85) 翻訳文提出日	平成27年1月29日 (2015. 1. 29)		アメリカ合衆国 94080 カリフォル
(86) 国際出願番号	PCT/US2013/045276		ニア州 サウス サンフランシスコ オイ
(87) 国際公開番号	W02013/188471		スター ポイント ブルバード 329
(87) 国際公開日	平成25年12月19日 (2013. 12. 19)		スイート 100
(31) 優先権主張番号	61/658, 317	(74) 代理人	100105957
(32) 優先日	平成24年6月11日 (2012. 6. 11)		弁理士 恩田 誠
(33) 優先権主張国	米国 (US)	(74) 代理人	100068755
(31) 優先権主張番号	61/776, 647		弁理士 恩田 博宣
(32) 優先日	平成25年3月11日 (2013. 3. 11)	(74) 代理人	100142907
(33) 優先権主張国	米国 (US)		弁理士 本田 淳
(31) 優先権主張番号	61/738, 277		
(32) 優先日	平成24年12月17日 (2012. 12. 17)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 配列タグを用いた配列決定法

(57) 【要約】

本発明は、配列タグを使用して、関連する配列のアンプリコン、特に大きくて複雑なアンプリコン、例えば、免疫受容体分子をコードする組換え核酸を含むアンプリコンの配列決定を改善することに関する。一態様では、同じ配列タグを有する配列リードが整列させられ、次いで、最終ベースコールが、各位置における配列リードのベースコールからの（重み付けされる場合もある）平均ベースコールから決定される。同様に、別の態様では、一連の取り込みシグナルを含む配列リードが、共通の配列タグによって整列させられて、ホモポリマー領域内のベースコールが、各「流れ」位置における取り込みシグナル値の関数として生成される。

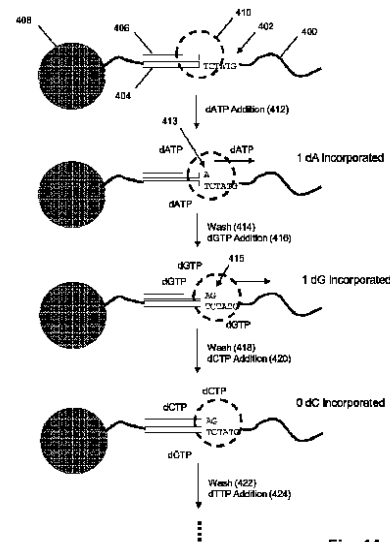


Fig. 4A

【特許請求の範囲】**【請求項 1】**

免疫レパトリのクローン型を決定する方法であって：

- (a) T細胞およびB細胞の少なくとも一方を含む、個体由来サンプルを得るステップ；
- (b) 前記T細胞およびB細胞の少なくとも一方のT細胞受容体遺伝子または免疫グロブリン遺伝子の組換え核酸分子に配列タグを付着させてタグ-分子コンジュゲートを形成するステップであって、前記タグ-分子コンジュゲートの実質的に全ての分子がユニークな配列タグを有する、ステップ；
- (c) 前記タグ-分子コンジュゲートを増幅するステップ；
- (d) 前記タグ-分子コンジュゲートをシークエンシングするステップ；および
- (e) 同様の配列タグの配列リードを整列させて、前記レパトリの同じクローン型に一致する配列リードを決定するステップを含む、方法。

10

【請求項 2】

前記整列させるステップが、前記同様の配列タグの前記クローン型の各ヌクレオチド位置における大部分のヌクレオチドを決定することによって、前記タグ-分子コンジュゲートのそれぞれの前記クローン型のそれぞれのヌクレオチド配列を決定するステップをさらに含む、請求項 1 に記載の方法。

【請求項 3】

前記付着させるステップが、前記組換え核酸分子をサンプリングによって標識化するステップを含む、請求項 1 に記載の方法。

20

【請求項 4】

前記配列タグが前記組換え核酸分子の濃度の少なくとも 100 倍の濃度で反応混合物に存在するように、前記付着させるステップが前記反応混合物で行われる、請求項 3 に記載の方法。

【請求項 5】

前記配列タグが、前記組換え核酸分子に特異的なプライマーに取り込まれる、請求項 4 に記載の方法。

【請求項 6】

前記配列タグがモザイクタグである、請求項 5 に記載の方法。

【請求項 7】

サンプル中のリンパ球の数を決定する方法であって：

- (a) リンパ球を含む、個体由来サンプルを得るステップ；
- (b) 前記リンパ球のT細胞受容体遺伝子または免疫グロブリン遺伝子の組換え核酸分子に配列タグを付着させてタグ-分子コンジュゲートを形成するステップであって、前記タグ-分子コンジュゲートの実質的に全ての分子がユニークな配列タグを有する、ステップ；
- (c) 前記タグ-分子コンジュゲートを増幅するステップ；
- (d) 前記タグ-分子コンジュゲートをシークエンシングするステップ；
- (e) 異なる配列タグの数を計数して前記サンプル中の前記リンパ球の数を決定するステップを含む、方法。

30

40

【請求項 8】

前記組換え核酸がDNAである、請求項 7 に記載の方法。

【請求項 9】

前記リンパ球がT細胞であり、前記組換え核酸が、T細胞受容体遺伝子またはその断片である、請求項 7 に記載の方法。

【請求項 10】

前記リンパ球がB細胞であり、前記組換え核酸が、免疫グロブリン遺伝子またはその断片である、請求項 7 に記載の方法。

【請求項 11】

免疫レパトリのクローン型を決定する方法であって：

50

(a) T細胞およびB細胞の少なくとも一方を含む、個体由来サンプルを得るステップ；
(b) 前記T細胞およびB細胞の少なくとも一方からの分子をサンプリングによって標識化してタグ-分子コンジュゲートを形成するステップであって、各タグが配列を有し、前記各分子が、T細胞受容体遺伝子または免疫グロブリン遺伝子からの組換え核酸を含む、ステップ；

(c) 前記タグ-分子コンジュゲートをシーケンシングするステップ；

(d) 同様のタグの配列リードを整列させて、同様のクローン型を決定するステップを含む、方法。

【請求項12】

前記整列させるステップが、前記同様の配列タグの前記クローン型の各ヌクレオチド位置における大部分のヌクレオチドを決定することによって、前記タグ-分子コンジュゲートのそれぞれの前記クローン型のそれぞれのヌクレオチド配列を決定するステップをさらに含む、請求項11に記載の方法。

【請求項13】

前記配列タグが前記組換え核酸分子の濃度の少なくとも100倍の濃度で反応混合物に存在するように、前記付着させるステップが前記反応混合物で行われる、請求項11に記載の方法。

【請求項14】

微小残留病変について監視される患者におけるクローン型の持越し汚染を検出する方法であって：

(a) 請求項1に記載の方法に従って、患者のクローン型プロファイルを定期的に測定することによって前記患者を微小残留病変について監視するステップ；および

(b) クローン型プロファイルの各測定の前記配列タグのそれぞれの前記配列を記録するステップ；および

(c) 任意の前のクローン型プロファイルの配列タグが後のクローン型プロファイルで検出された場合にクローン型の持越し汚染を検出するステップを含む、方法。

【請求項15】

1つ以上の合成反応によるシーケンシングにおいて1つ以上のポリヌクレオチドのヌクレオチド配列を決定する方法であって：

(a) 配列タグを前記1つ以上のポリヌクレオチドのそれぞれに付着させてタグ-ポリヌクレオチドコンジュゲートを形成するステップであって、前記タグ-ポリヌクレオチドコンジュゲートの実質的に全てのポリヌクレオチドがユニークな配列タグを有する、ステップ；

(b) 前記タグ-ポリヌクレオチドコンジュゲートを増幅するステップ；

(c) 前記増幅されたタグ-ポリヌクレオチドコンジュゲートを合成によってシーケンシングするステップであって、少なくとも1つのdNTPの流れを含む、ステップ；および

(d) 同じ配列タグを有するタグとポリヌクレオチドのそれぞれについて、各dNTPの流れに対するヌクレオチドの取り込み数を、このような各dNTPの流れに対して測定された取り込みシグナルの関数として決定するステップを含む、方法。

【請求項16】

前記1つ以上のポリヌクレオチドが複数のポリヌクレオチドである、請求項15に記載の方法。

【請求項17】

前記複数が少なくとも104である、請求項15に記載の方法。

【請求項18】

前記dNTPの流れのそれぞれについての前記ヌクレオチドの取り込み数が、前記測定された取り込みシグナルの平均に最も近い整数である、請求項15に記載の方法。

【請求項19】

前記合成によるシーケンシングのステップが：

10

20

30

40

50

(a) 前記タグ - ポリヌクレオチドコンジュゲートのそれぞれに対して、シーケンシングプライマー、核酸ポリメラーゼ、およびタグ - ポリヌクレオチドコンジュゲートを含む複合体を、前記シーケンシングプライマーのこのようなタグ - ポリヌクレオチドコンジュゲートに対するアニーリングを可能にし、かつこのようなシーケンシングプライマーの、ヌクレオチド三リン酸の存在下での前記核酸ポリメラーゼによる前記タグ - ポリヌクレオチドコンジュゲートに沿った伸長を可能にする条件下で形成するステップ；

(b) d N T P を d N T P の流れによって前記複合体に導入するステップ；

(c) 取り込みシグナルを測定するステップ；

(d) 前記複合体を洗浄するステップ；および

(e) 前記ステップ (b) から前記ステップ (d) を繰り返すステップを含む、請求項 15 に記載の方法。

10

【請求項 20】

前記 d N T P が、伸長ブロック d N T P であり、前記洗浄するステップが、続く導入するステップで、さらなる伸長ブロック d N T P が取り込まれ得るように、取り込まれた伸長ブロック d N T P のブロックを解除するステップをさらに含む、請求項 19 に記載の方法。

【請求項 21】

異なるタグ - ポリヌクレオチドコンジュゲートがそれぞれ、異なる反応区画にある、請求項 15 に記載の方法。

【請求項 22】

前記付着させるステップが、前記 1 つ以上のポリヌクレオチドをサンプリングによって標識化して前記タグ - ポリヌクレオチドコンジュゲートを形成するステップを含む、請求項 15 に記載の方法。

20

【請求項 23】

前記配列タグが、1 ~ 5 ヌクレオチドの範囲の長さの可変領域を有するモザイクタグである、請求項 15 に記載の方法。

【請求項 24】

前記配列タグが、各交互領域が 1 ~ 5 ヌクレオチドの範囲の長さを有するように、A、C、G、および T の互いに素なサブセットから選択されるヌクレオチドを有する交互領域を含む、請求項 15 に記載の方法。

30

【請求項 25】

前記シーケンシングするステップが、前記増幅されたタグ - ポリヌクレオチドコンジュゲートのサンプルをシーケンシングするステップを含む、請求項 15 に記載の方法。

【請求項 26】

前記 1 つ以上のポリヌクレオチドが、T 細胞受容体遺伝子または免疫グロブリン遺伝子の組み換え核酸分子である、請求項 15 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、配列タグを用いた配列決定法に関する。

40

【背景技術】

【0002】

生物学的サンプルまたは医学的サンプルの分析は、DNA および / または RNA の大きくて複雑な集団の核酸配列の決定を必要とする場合が多い。例えば、(非特許文献 1)；(非特許文献 2)；(非特許文献 3)。特に、免疫分子、例えば、T 細胞受容体もしくは B 細胞受容体、または他の成分をコードする核酸のプロファイルは、生物の健康もしくは疾患の状態についての豊富な情報を含むため、このようなプロファイルの診断指標もしくは予後指標としての使用が様々な状態に対して提案されてきた。例えば、(特許文献 1)；(非特許文献 4)；(非特許文献 5)；(非特許文献 6)。このような配列ベースのプロファイルは、増幅された標的核酸のサイズ分布、マイクロアレイによる逐次サンプリン

50

グ、PCRアンプリコンからのハイブリダイゼーション反応速度曲線に基づいたアプローチ、または他のアプローチよりも格段に高い感度を有し得る。例えば、(特許文献2)；(非特許文献7)；(非特許文献8)；(非特許文献9)；(非特許文献10)。しかしながら、配列データからのクローン型およびクローン型プロファイルの効率的な決定は、分析される集団のサイズ、このような集団における配列の類似性、配列間の自然の多様性の限定された予測可能性、ならびに宿主のサンプルの調製ステップおよび測定ステップによるノイズのデータへの混入のため困難を伴う。例えば、(非特許文献11)。

【0003】

配列タグまたはバーコードは、標識化、汚染の監視、稀な突然変異の検出、物理的分類、および分子の計数などを含む核酸集団の分析に役立てるために様々な方法で使用されている。例えば、(非特許文献12)；(特許文献3)；(特許文献4)；(特許文献5)；(非特許文献13)；(特許文献6)；(非特許文献14)；(特許文献7)；(特許文献8)。近年、(非特許文献12)(前掲)は、シーケンシングエラーおよび増幅エラーと基準配列における稀な突然変異とを区別するために配列タグをどのように使用できるかを示した。

10

【0004】

医学的適用および診断的適用における正確なシーケンシングの重要性から、このような適用における配列決定の効率および精度を高めるために配列タグの使用を拡大できると非常に有利であろう。

20

【先行技術文献】

【特許文献】

【0005】

【特許文献1】ファハム(Faham)およびウィリス(Willis)による米国特許出願公開第2010/0151471号明細書

【特許文献2】モーレイ(Morley)らによる米国特許第5,418,134号明細書

【特許文献3】カスボン(Casbon)らによる米国特許出願公開第2012/0071331号明細書

【特許文献4】ブレンナー(Brenner)による米国特許第5,635,400号明細書

30

【特許文献5】ブレンナー(Brenner)およびマチェビッツ(Macevicz)による米国特許第7,537,897号明細書

【特許文献6】チャーチ(Church)らによる欧州特許第0303459号明細書

【特許文献7】モーリス(Morris)らによる欧州特許出願公開第0799897A1号明細書

【特許文献8】ワレス(Wallace)による米国特許第5,981,179号明細書

【非特許文献】

【0006】

【非特許文献1】グロアー(Gloor)ら著、PLOS ONE、2010年、第5巻、第10号、e15406

40

【非特許文献2】ペトロシノ(Petrosino)ら著、クリニカル・ケミストリー(Clinical Chemistry)、2009年、第55巻、第5号、p.856~866

【非特許文献3】アスティラ(Arstilla)ら著、サイエンス(Science)、1999年、第286巻、p.958~961

【非特許文献4】フリーマン(Freeman)ら著、ゲノム・リサーチ(Genome Research)、2009年、第19巻、p.1817~1824

【非特許文献5】ボイド(Boyd)ら著、サイエンス・トランスレーショナル・メディシン(Sci. Transl. Med.)、2009年、第1巻、第12号、12ra23

50

【非特許文献 6】ヒ (He) ら著、オンコターゲット (Oncotarget) (2011 年 3 月 8 日)

【非特許文献 7】ファン・ドンゲン (van Dongen) ら著、ルーキミア (Leukemia)、2003 年、第 17 巻、p. 2257 ~ 2317

【非特許文献 8】オグル (Ogle) ら著、ヌクレイック・アシッド・リサーチ (Nucleic Acids Research)、2003 年、第 31 巻、p. e139

【非特許文献 9】ワン (Wang) ら著、BMC ゲノミクス (BMC Genomics)、2007 年、第 8 巻、p. 329

【非特許文献 10】ボーム (Baum) ら著、ネイチャー・メソッド (Nature Methods)、2006 年、第 3 巻、第 11 号、p. 895 ~ 901

10

【非特許文献 11】ワーレン (Warren) ら著、ゲノム・リサーチ (Genome Research)、2011 年、第 21 巻、第 5 号、p. 790 ~ 797

【非特許文献 12】キンデ (Kinde) ら著、米国科学アカデミー紀要 (Proc. Natl. Acad. Sci.)、2011 年、第 108 巻、第 23 号、p. 9530 ~ 9535

【非特許文献 13】ブレンナー (Brenner) ら著、国科学アカデミー紀要 (Proc. Natl. Acad. Sci.)、2000 年、第 97 巻、p. 1665 ~ 1670

【非特許文献 14】シューメーカー (Shoemaker) ら著、ネイチャー・ジェネティクス (Nature Genetics)、1996 年、第 14 巻、p. 450 ~ 456

20

【図面の簡単な説明】

【0007】

【図 1 A】サンプリングによってユニークな配列タグを核酸分子に付着させる標識化の例を例示している。

【図 1 B】サンプリングによってユニークな配列タグを核酸分子に付着させる標識化の例を例示している。

【図 1 C】IGH 転写物およびこの IGH 転写物内の自然の多様性の原因を例示している。

【図 2 A】組換え核酸分子を増幅するための 2 段階 PCR スキームを示している。

【図 2 B】組換え核酸分子を増幅するための 2 段階 PCR スキームを示している。

【図 2 C】組換え核酸分子を増幅するための 2 段階 PCR スキームを示している。

30

【図 3】同じ配列タグに関連する配列リードからクローン型の配列を決定するステップの 1 つの実施を例示している。

【図 4 A】合成操作によるシーケンシングでのホモポリマー領域の決定に関する本発明の一実施形態を例示している。

【図 4 B】合成操作によるシーケンシングでのホモポリマー領域の決定に関する本発明の一実施形態を例示している。

【図 4 C】合成操作によるシーケンシングでのホモポリマー領域の決定に関する本発明の一実施形態を例示している。

【図 4 D】合成操作によるシーケンシングでのホモポリマー領域の決定に関する本発明の一実施形態を例示している。

40

【発明を実施するための形態】

【0008】

(発明の概要)

本発明は、複雑な核酸集団、特に免疫分子のレパートリまたはその一部をコードする組換え核酸集団の配列ベースのプロファイルを作成する方法に関する。本発明は、一部が以下で概説される多数の実施および適用例、ならびに本出願全体で例示される。

【0009】

一態様では、本発明は、次のステップを含む、免疫レパートリのクローン型を決定する方法に関する：(a) T 細胞および B 細胞の少なくとも一方を含む個体由来サンプルを得るステップ；(b) T 細胞および B 細胞の少なくとも一方の T 細胞受容体遺伝子または免

50

疫グロブリン遺伝子の組換え核酸分子に配列タグを付着させてタグ - 分子コンジュゲートを形成するステップであって、タグ - 分子コンジュゲートの実質的に全ての分子がユニークな配列タグを有する、ステップ； (c) タグ - 分子コンジュゲートを増幅するステップ； (d) タグ - 分子コンジュゲートをシークエンシングするステップ；および (e) 同様の配列タグの配列リードを整列させて、レパートリの同じクローン型に一致する配列リードを決定するステップ。

【 0 0 1 0 】

別の態様では、本発明は、次のステップを含む、微小残留病変について監視される患者におけるクローン型の持越し汚染を検出する方法に関する： (a) 請求項 A 1 に記載の方法に従って、患者のクローン型プロファイルを定期的に測定することによって患者を微小残留病変について監視するステップ； (b) クローン型プロファイルの各測定の前記配列タグを記録するステップ；および (c) 任意の前のクローン型プロファイルの配列タグが後のクローン型プロファイルで検出された場合にクローン型の持越し汚染を検出するステップ。

10

【 0 0 1 1 】

別の態様では、本発明は、次のステップを含む、サンプル中のリンパ球の数を決定する方法に関する： (a) リンパ球を含む個体由来サンプルを得るステップ； (b) リンパ球の T 細胞受容体遺伝子または免疫グロブリン遺伝子の組換え核酸分子に配列タグを付着させてタグ - 分子コンジュゲートを形成するステップであって、タグ - 分子コンジュゲートの実質的に全ての分子がユニークな配列タグを有する、ステップ； (c) タグ - 分子コンジュゲートを増幅するステップ； (d) タグ - 分子コンジュゲートをシークエンシングするステップ； (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ステップ。

20

【 0 0 1 2 】

別の態様では、本発明は、次のステップを含む、1つ以上の合成反応によるシークエンシングにおいて1つ以上のポリヌクレオチドのヌクレオチド配列を決定する方法に関する： (a) 配列タグを1つ以上のポリヌクレオチドのそれぞれに付着させてタグ - ポリヌクレオチドコンジュゲートを形成するステップであって、タグ - ポリヌクレオチドコンジュゲートの実質的に全てのポリヌクレオチドがユニークな配列タグを有する、ステップ； (b) タグ - ポリヌクレオチドコンジュゲートを増幅するステップ； (c) 増幅されたタグ - ポリヌクレオチドコンジュゲートを合成によってシークエンシングするステップであって、少なくとも1つの d N T P の流れを含む、ステップ；および (d) 同じ配列タグを有するタグとポリヌクレオチドのそれぞれについて、各 d N T P の流れに対するヌクレオチドの取り込み数を、このような各 d N T P の流れに対して測定された取り込みシグナルの関数として決定するステップ。

30

【 0 0 1 3 】

本発明は、免疫分子をコードする体細胞的に組み換えられた核酸の高スループットシークエンシングによって得られる配列データの大きいセットからクローン型およびクローン型プロファイルを決定する方法を提供する。一態様では、本発明は、サンプル中の体細胞的に組み換えられた核酸分子のそれぞれをユニークな配列タグで標識し、これを使用してサンプルの同じクローン型配列のコピーを含む配列リードを分類することによって、上記の方法を実施する。

40

【 0 0 1 4 】

本発明のこれらの上記特徴付けられた態様および他の態様は、一部が図面に示され、添付の特許請求の範囲によって特徴付けられた多数の例示的な実施および適用例で例示される。しかしながら、上記の概要は、本発明の各例示された実施形態または全ての実施を説明することを意図するものではない。

【 0 0 1 5 】

本発明の新規な特徴は、特に添付の特許請求の範囲に記載される。本発明の原理が利用される、例示的な実施形態を説明する以下の詳細な説明および添付の図面を参照すれば、

50

本発明の特徴および利点のより良い理解が得られる。

【0016】

(発明の詳細な説明)

本発明の実施は、特段の記載がない限り、当業者の技術の範囲内である、分子生物学(組み換え技術を含む)、バイオインフォマティクス、細胞生物学、および生物化学の従来の技術および記述を利用することができる。このような従来の技術としては、限定されるものではないが、血液細胞のサンプリングおよび分析、ならびに核酸のシーケンシングおよび分析などが挙げられる。適切な技術の特定の例示では、以下の例を参照しなければならないことがある。しかしながら、他の同等の従来の手順も、もちろん使用することができる。このような従来の技術および記述は、標準的な実習マニュアル、例えば、ゲノム解析(Genome Analysis):実習マニュアルシリーズ(A Laboratory Manual Series)(第I巻~第IV巻);PCRプライマー(PCR Primer):実習マニュアル(A Laboratory Manual);および分子クローニング:実習マニュアル(Molecular Cloning):実習マニュアル(A Laboratory Manual)(全て、コールド・スプリング・ハーバ研究所の出版局(Cold Spring Harbor Laboratory Press)から)などで確認することができる。

10

【0017】

一態様では、本発明は、免疫分子のレパートリ、例えば、T細胞受容体(TCR: T cell receptor)もしくはB細胞受容体(BCR: B cell receptor)またはそれらの定義された断片から配列データを得て分析して、クローン型プロフィールを迅速かつ効率的に決定する方法に関する。配列データは、典型的には、免疫分子を分析するために使用されるDNAシーケンサからの多数の配列リード、即ち、ベースコールの配列および関連する品質スコアを含む。クローン型プロフィールを作成する上での大きな課題は、真の差異を含む配列リードと、非生物学的起源由来のエラー、例えば、抽出ステップ、シーケンシング化学、または増幅化学などに由来のエラーを含む配列リードとを迅速かつ正確に区別することである。本発明の一態様は、サンプル中の各クローン型にユニークな配列タグを付着させて、このようなコンジュゲートの配列リードが同じ起源のクローン型に由来するか否かの決定に役立てることを含む。本発明の一態様によると、配列タグを、体細胞的に組み換えられた核酸分子に付着させてタグ-分子コンジュゲートを形成し、このようなコンジュゲートの各組換え核酸は、ユニークな配列タグを有する。通常、このような付着は、核酸分子がT細胞およびB細胞の少なくとも一方を含むサンプルから抽出されてから行われる。好ましくは、このようなユニークな配列タグは、互いに可能な限り大きく異なり、この差異は、従来の配列の距離指標、例えば、ハミング距離などによって決定される。たとえシーケンシングエラーおよび増幅エラーが高率でも、タグ-分子コンジュゲートの配列タグ間の距離を最大にすることにより、コンジュゲートの配列タグが、異なるコンジュゲートの他のタグ配列よりも、祖先のタグ配列に遥かに近いままである。例えば、16-merの配列タグが利用され、クローン型のセットに対するこのような各タグが、そのクローン型のどの配列タグからも少なくとも50%、または8ヌクレオチドのハミング距離を有する場合は、1つのこのようなタグを、配列タグのミスリード(mis-read)(および、クローン型の配列リードの誤った配列タグでの不正確な分類)によって別のタグに変換するには、少なくとも8つのシーケンシングエラーまたは増幅エラーが必要であろう。一実施形態では、配列タグは、組換え核酸分子に付着させてタグ-分子コンジュゲートを形成した後に、タグ-分子コンジュゲートのタグ間のハミング距離が、このような配列タグの全長の少なくとも25%の数値となる(即ち、各配列タグは、このような他のどのタグの配列ともヌクレオチドが少なくとも25%異なる)ように選択され;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、このような配列タグの全長の少なくとも50%の数値である。

20

30

40

【0018】

一態様では、本発明は、次のステップによって実施される:(a)T細胞およびB細胞

50

の少なくとも一方を含む個体由来サンプルを得るステップ；(b) T細胞およびB細胞の少なくとも一方のT細胞受容体遺伝子または免疫グロブリン遺伝子の組換え核酸分子に配列タグを付着させてタグ-分子コンジュゲートを形成するステップであって、このタグ-分子コンジュゲートの実質的に全ての分子がユニークな配列タグを有する、ステップ；(c) タグ-分子コンジュゲートを増幅するステップ；(d) タグ-分子コンジュゲートをシーケンシングするステップ；および(e) 同様の配列タグの配列リードを整列させて、同じクローン型のレパートリに対応する配列リードを決定するステップ。B細胞またはT細胞を含むサンプルは、以下により十分に説明される従来の技術を用いて得られる。配列タグを付着させるステップでは、好ましくは、配列タグは、単にユニークであるだけではなく、たとえ多数のシーケンシングエラーまたは増幅エラーがあったとしても1つの配列タグを別の配列タグに変換する可能性がほぼゼロであるように互いに十分異なっている。配列タグを付着させた後に、タグ-分子コンジュゲートの増幅が、ほとんどのシーケンシング技術には必要である；しかしながら、1分子シーケンシング技術が利用される場合は常に、増幅ステップは任意選択的である。1分子シーケンシング技術としては、限定されるものではないが、1分子リアルタイム(SMRT: single molecule real-time)シーケンシング、またはナノ細孔シーケンシングなどが挙げられる。例えば、米国特許第7,313,308号明細書；同第8,153,375号明細書；同第7,907,800号明細書；同第7,960,116号明細書；同第8,137,569号明細書；およびマンラオ(Manrao)ら著、ネイチャー・バイオテクノロジー(Nature Biotechnology)、2012年、第4巻、第8号、p.2685~2693など。

10

20

【0019】

別の態様では、本発明は、ユニークな配列タグを計数することによってサンプル中のリンパ球の数を決定する方法を含む。たとえ配列タグがなくても、TCR β またはIGH遺伝子のクローン型、特にV(D)J領域を含むクローン型では、リンパ球およびそのクローンがユニークなマーカとなる。組換え核酸がゲノムDNAから得られる場合は常に、サンプル中のリンパ球の数を、シーケンシング後に計数されるユニークなクローン型の数によって推定することができる。このアプローチは、同じクローン型に関連した同一のリンパ球のかなりのクローン集団が存在する場合(または、組換え核酸が、サンプルのmRNAから得られ、その個々の配列の量が、発現率および細胞数に反映し得る、または発現率および細胞数に依存し得る場合)は常に機能しない。配列タグの使用は、この欠点を克服し、多くのリンパ障害、例えば、リンパ腫または白血病に罹患している患者のリンパ球の数を得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、例えば、白血病でのように大きい優勢クローンが存在するしないにかかわらず、サンプル中のリンパ球の絶対数を得ることができる。このような方法は、次のステップで実施することができる：(a) リンパ球を含む個体由来サンプルを得るステップ；(b) リンパ球のT細胞受容体遺伝子または免疫グロブリン遺伝子の組換え核酸分子に配列タグを付着させてタグ-分子コンジュゲートを形成するステップであって、タグ-分子コンジュゲートの実質的に全ての分子がユニークな配列タグを有する、ステップ；(c) タグ-分子コンジュゲートを増幅するステップ；(d) タグ-分子コンジュゲートをシーケンシングするステップ；および(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ステップ。一部の実施形態では、組換え核酸分子は、ゲノムDNA由来である。

30

40

【0020】

本発明の一実施形態では、例えば、参照により本明細書に組み入れられるブレンナー(Brenner)らによる米国特許第5,846,719号明細書；ブレンナー(Brenner)らによる米国特許第7,537,897号明細書；およびマチェビッツ(MacEvicz)による国際公開第2005/111242号などに開示されているように、配列タグが、サンプリングによる標識化によってサンプルの組換え核酸分子に付着される。サンプリングによる標識化では、標識化されるべき(またはユニークにタグが付けられるべき)ポリヌクレオチドの集団は、より大きい集団の配列タグを(付着または連結な

50

どによって) サンプリングするために使用される。即ち、ポリヌクレオチドの集団がKメンバー(同じポリヌクレオチドの複製を含む)を有し、かつ配列タグの集団がNメンバーを有する場合は、 $N \gg K$ である。一実施形態では、本発明に使用される配列タグの集団のサイズは、サンプル中のクローン型の集団のサイズの少なくとも10倍であり; 別の実施形態では、本発明に使用される配列タグの集団のサイズは、サンプル中のクローン型の集団のサイズの少なくとも100倍であり; 別の実施形態では、本発明に使用される配列タグの集団のサイズは、サンプル中のクローン型の集団のサイズの少なくとも1000倍である。他の実施形態では、配列タグの集団のサイズは、サンプル中の実質的に全てのクローン型が、このようなクローン型が、例えば、付着反応、例えば、ライゲーション反応または増幅反応でこのような配列タグの集団に結合されると必ず、ユニークな配列タグを有するように選択される。一部の実施形態では、実質的に全てのクローン型とは、このようなクローン型の少なくとも90%がユニークな配列タグを有することを意味し; 他の実施形態では、実質的に全てのクローン型とは、このようなクローン型の少なくとも99%がユニークな配列タグを有することを意味し; 他の実施形態では、実質的に全てのクローン型とは、このようなクローン型の少なくとも99.9%がユニークな配列タグを有することを意味する。多くの組織サンプルまたは生検では、T細胞またはB細胞の数は、最大100万または約100万の細胞であり得; 従って、このようなサンプルを利用する本発明の一部の実施形態では、サンプリングによる標識化に利用されるユニークな配列タグの数は、少なくとも 10^8 であり、他の実施形態では、少なくとも 10^9 である。

10

20

【0021】

最大100万のクローン型がサンプリングによって標識化されるこのような実施形態では、配列タグの多くのセットを、例えば、参照により本明細書に組み入れられるチャーチ(Church)による米国特許第5,149,625号明細書に開示されているように、合成反応の各追加のステップで4つ全てのヌクレオチド前駆体の混合物を反応させることにより、コンビナトリアル合成によって効率的に作製することができる。その結果、「 $N_1 N_2 \dots N_k$ 」の構造を有する配列タグのセットが得られ、各 $N_i = A, C, G$ 、またはTであり、kは、タグ中のヌクレオチドの数である。このようなコンビナトリアル合成によって作製される配列タグのセットにおける配列タグの数は 4^k である。従って、kが少なくとも14またはkが14~18の範囲であるこのような配列タグのセットは、サンプリングによる標識化によって配列タグを分子の 10^6 のメンバーの集団に付着させるのに適している。上記の構造を有する配列タグのセットは、本発明の方法の実施中に困難またはエラーをもたらし得る多数の配列を含む。例えば、上記のコンビナトリアル合成された配列タグのセットは、一部のシーケンシングアプローチ、例えば、合成によるシーケンシングアプローチが特定の長さを超えると高精度での決定が困難となる、ホモポリマーセグメントを有する多数のメンバーのタグを含む。従って、本発明は、特定の方法のステップ、例えば、シーケンシングに効率的な構造を有するコンビナトリアル合成された配列を含む。例えば、合成によるシーケンシング化学に効率的ないくつかの配列タグ構造は、4つの天然ヌクレオチドを、コンビナトリアル合成で代替的に使用される互いに素なサブセットに分割することによって作製し、これにより、所与の長さを超えるポリマーセグメントを防止することができる。例えば、zをAまたはCとし、かつxをGまたはTとして、次の配列タグ構造が得られ

30

40

$[(z)_1 (z)_2 \dots (z)_i] [(x)_1 (x)_2 \dots (x)_j] \dots$
 式中、iおよびjは、同じでも異なっても良く、任意のホモポリマーセグメントのサイズを制限するために選択される。一実施形態では、iおよび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得; 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得る。他の実施形態では、ヌクレオチドの他の対を使用することができ、例えば、zはAもしくはTであり、xはGもしくはCであり; または、zはAもしくはGであり、xはTもしくはCである。あるいは、z'を、4つの天然ヌクレオチドの3つの任意の組み合わせとし、x'を、ヌクレオチドがz'のヌクレオチドではないようにする(例えば、z'が

50

A、C、またはGであり、x'がTである)。これにより、次の配列タグ構造が得られる：

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

式中、iは、上記のように選択され、x'の存在は、あらゆる望ましくないホモポリマーを終了させる中断として機能する。

【0022】

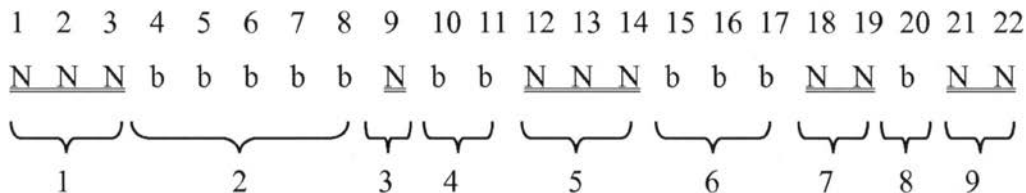
さらなる配列タグ

本発明は、増幅およびシーケンシングの前に、「モザイクタグ」を含み得るユニークな配列タグを用いて核酸、例えば、ゲノムDNAの断片を標識化する方法を使用する。このような配列タグは、増幅エラーおよびシーケンシングエラーの識別に有用である。モザイクタグは、従来技術の完全にランダムな配列タグで起こり得る、不適切なアニーリング、プライミング、またはヘアピン形成などによるシーケンシングアーチファクトおよび増幅アーチファクトを最小限にする。一態様では、モザイクタグは、交互する定常領域と可変領域を有する配列タグであり、各定常領域は、モザイクタグにおける位置を有し、かつ所定のヌクレオチド配列を含み、各可変領域は、モザイクタグにおける位置を有し、かつ所定数のランダムに選択されたヌクレオチドを含む。単なる例示として、22-merのモザイクタグ（配列番号4）は、次の形を有し得る。

ヌクレオチド位置：

【0023】

【化1】



領域位置

9つの定常領域および可変領域が存在し、領域1（ヌクレオチド1～3）、領域3（ヌクレオチド9）、領域5（ヌクレオチド12～14）、領域7（ヌクレオチド18～19）、および領域9（ヌクレオチド21～22）が可変であり（2本の下線が引かれたヌクレオチド）、領域2（ヌクレオチド4～8）、領域4（ヌクレオチド10～11）、領域6（ヌクレオチド15～17）、および領域8（ヌクレオチド20）が定常である。Nは、A、C、G、またはTのセットからランダムに選択されたヌクレオチドを表し；従って、この例のモザイクタグの数は、 $4^1 \cdot 1 = 4$ 、194、304のタグである。bは、指定位置での所定のヌクレオチドを表す。一部の実施形態では、bの配列、「***b b b b b * b b * * * b b b * * b * *」は、サンプルを構成する生物のゲノムで完全な一致を有する可能性を最小限にするように選択される。

【0024】

一態様では、本発明の方法の特定の実施形態のモザイクタグでは、同じ位置を有する全ての定常領域は同じ長さを有し、同じ位置を有する全ての可変領域は同じ長さを有する。これにより、従来の化学および器具を用いて、部分コンビナトリアル合成によりモザイクタグを合成することができる。

【0025】

一態様では、モザイクタグは、10～100ヌクレオチド、12～80ヌクレオチド、または15～60ヌクレオチドを含む。一部の実施形態では、モザイクタグは、ランダムに選択されたヌクレオチドを有する少なくとも8つのヌクレオチド位置を含み；他の実施形態では、モザイクタグが、少なくとも15ヌクレオチドの長さを有する場合は常に、これらのモザイクタグは、ランダムに選択されたヌクレオチドを有する少なくとも12ヌ

レオチド位置を含む。別の態様では、モザイクタグ内の非可変領域は、7ヌクレオチドよりも長い長さを有し得る。

【0026】

別の態様では、モザイクタグは、次のステップで 사용할 ことができる：(i) サンプル中の核酸からDNA鋳型を調製するステップ；(ii) DNA鋳型をサンプリングによって標識化して多数のタグ-鋳型コンジュゲートを形成するステップであって、タグ-鋳型コンジュゲートの実質的に全てのDNA鋳型が、交互する定常領域と可変領域を含むユニークなモザイクタグを有し、各定常領域が、同じ位置を有する定常領域が同じ長さを有するように、モザイクタグにおける位置および所定の配列の1~10ヌクレオチドの長さを有し、かつ各可変領域が、同じ位置を有する可変領域が同じ長さを有するように、モザイクタグにおける位置および1~10のランダムに選択されたヌクレオチドの長さを有する、ステップ；(iii) 多数のタグ-鋳型コンジュゲートを増幅するステップ；(iv) 増幅されたタグ-鋳型コンジュゲートのそれぞれに対して複数の配列リードを形成するステップ；および(v) 同一のモザイクタグを有する複数の配列リードのそれぞれの各ヌクレオチド位置におけるコンセンサヌクレオチドを決定することによって、各核酸のヌクレオチド配列を決定するステップ。別の態様では、モザイクタグは、次のステップに使用することができる：(a) サンプル中の核酸から一本鎖DNA鋳型を調製するステップ；(b) 一本鎖DNA鋳型をサンプリングによって標識化してタグ-鋳型コンジュゲートを形成するステップであって、タグ-鋳型コンジュゲートの実質的に全ての一本鎖DNA鋳型が、少なくとも15ヌクレオチドの長さおよび次の形を有するユニークな配列タグ(即ち、モザイクタグ)を有し：

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

式中、各 N_i は、 $i = 1, 2, \dots, K_j$ で、A、C、G、および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、各 j がM以下の場合には1~10の範囲の整数であり(即ち、領域 $N_1 N_2 \dots N_{K_j}$ が可変領域)； $i = 1, 2, \dots, L_j$ である各 b_i はヌクレオチドであり； L_j は、各 j がM以下の場合には1~10の範囲の整数であり；従って、全ての配列タグが、(i) 全ての j に対して同じ K_j を有し、かつ(ii) 全ての j に対して同じ配列 $b_1 b_2 \dots b_{L_j}$ を有し(即ち、領域 $b_1 b_2 \dots b_{L_j}$ が定常領域である)；そしてMが2以上の整数である、ステップ；(c) タグ-鋳型コンジュゲートを増幅するステップ；(d) 増幅されたタグ-鋳型コンジュゲートのそれぞれに対して複数の配列リードを形成するステップ；および(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各ヌクレオチド位置におけるコンセンサヌクレオチドを決定することによって、核酸のそれぞれのヌクレオチド配列を決定するステップ。一部の実施形態では、複数の配列リードは、少なくとも 10^4 であり；他の実施形態では、複数の配列リードは、少なくとも 10^5 であり；なお他の実施形態では、複数の配列リードは、少なくとも 10^6 である。一部の実施形態では、上記の配列タグの全長は、15~80ヌクレオチドの範囲である。

【0027】

様々な付着反応を使用して、ユニークなタグをサンプル中の実質的に全てのクローン型に付着させることができる。一実施形態では、このような付着は、組換え核酸分子(クローン型配列を含む)を含むサンプルを配列タグの集団またはライブラリと混合することにより、分子の2つの集団の分子のメンバーがランダムに組み合わせられて、例えば、共有結合によって結合または連結されることによって達成される。このようなタグ付着反応では、クローン型配列は、線状の一本鎖または二本鎖ポリヌクレオチドを含み、配列タグは、試薬、例えば、増幅プライマー、例えば、PCRプライマー、ライゲーションアダプター、環状化可能なプロンプ、またはプラスミドなどによって保持される。配列タグ集団を保持できるいくつかのこのような試薬は、参照により本明細書に組み入れられる、マチェビッツ(MacEvicz)による米国特許第8,137,936号明細書；ファハム(Faham)らによる米国特許第7,862,999号明細書；ランデグレン(Landegren)らによる米国特許第8,053,188号明細書；ウンラウ(Unrau)

およびデウガウ (Deugau) 著、ジーン (Gene)、1994年、第145巻、p. 163~169; チャーチ (Church) による米国特許第5,149,625号明細書に開示されている。

【0028】

図1Aおよび図1Bは、配列タグ (T_1 、 T_2 、 $T_3 \dots T_j$ 、 $T_{j+1} \dots T_k$ 、 $T_{k+1} \dots T_{n-1}$ 、 T_n) の集団がプライマー (100) に取り込まれるPCRを含む付着反応を例示している。配列タグの集団は、組換え核酸分子 (102) のサイズよりも遥かに大きいサイズを有する。配列タグは、PCRの最初のサイクルでプライマーを核酸分子にアニーリングして、プライマーをDNAポリメラーゼで伸長させることによって組換え核酸分子に付着される。この図は、組換え核酸分子が、例えば、V領域 (108) にある、それらの共通のプライマー結合領域 (104) によってプライマーにランダムにアニーリングすることにより、配列タグの全集団のほんの一部をどのように選択するか、またはサンプリングするかを示している。プライマー (従って配列タグ) が、組換え核酸配列分子にランダムに結合するため、同じ配列タグが異なる核酸分子に付着し得る可能性は少し存在する; しかしながら、配列タグの集団が、本明細書で教示されるように大きいと、このような可能性は、無視できるほど小さいため、実質的に全ての組換え核酸分子が、付着されたユニークな配列タグを有する。順方向と逆方向のプライマー対の他方のプライマー (106) が、C領域 (110) にアニーリングするため、アニーリング、伸長、および融解の複数のサイクルの後に、アンプリコン (112) が形成され、従って、ユニークな配列タグが、クローン型の集団 (102) を含むV(D)J領域に付着する。即ち、アンプリコン (112) は、付着反応から得られるタグ-分子コンジュゲートを含む。

【0029】

このような免疫分子は、典型的には、免疫レパートリを形成し、この免疫レパートリは、比較的長さが短い (例えば、通常は300bp未満である) 非常に類似したポリヌクレオチドの非常に大きいセット (例えば、1000よりも大きい、より一般的には100,000~1,000,000以上である) を含む。本発明の一態様では、本発明者らは、これらの特徴により、大きく異なる配列タグを使用して非常に類似したクローン型の配列リードを効率的に比較し、これらの配列リードが同じ起源の配列に由来するか否かを決定できること分かり、これを歓迎した。

【0030】

免疫レパートリの複合体は周知である。例えば、アスティラ (Astilla) ら著、サイエンス (Science)、1999年、第286巻、p. 958~961およびワーレン (Warren) ら著 (前掲)。図1Cは、本発明の一部の実施形態に従ってクローン型プロファイルが得られるIGH分子 (120) の典型的な転写物を図式的に例示している。天然の配列の多様性の原因には、ゲノムによって保持される多数のセットからのCセグメント、Dセグメント、Jセグメント、およびVセグメントのモジュール組換え、いわゆる「NDN」領域を形成するDセグメントの末端に対するヌクレオチドの付加および欠失、ならびに、曲線 (128) によって大まかに示されているような相対頻度で転写物 (122) の長さに対して置換がランダムに行われる体細胞超変異が含まれる。本発明の一態様では、このようなIGHおよびTCR転写物の複合集団が増幅され、シーケンシングされる。一態様では、IGH分子に対する一方または両方の操作が、V領域の異なる部位にアニーリングする重複プライマーを用いて行われる (以下により詳細に説明される)。これは、エラー率が比較的高いシーケンシング化学が利用される場合、またはこのような配列の多様性を事前に知ることが困難もしくは不可能な場合に特に有利である。後者の場合、たとえ、(例えば) 1つ以上の体細胞変異によるミスマッチのために1つ以上のプライマー結合部位が機能しない、または実質的に機能しなくても、配列リードの増幅または形成のためのプライマーの伸長は起こる。曲線 (128) によって示されている相対頻度は、プロモータP (122) から始まり、リーダ領域 (124) で上昇し、転写のV(D)J領域 (126) で最大となり、以降は下降してゼロに近づいていく。本発明

の一態様では、ファハム (Faham) およびウィリス (Willis) による米国特許出願公開第 2011/0207134 号明細書に開示されているように、組換え B 細胞核酸のセグメントが、鋳型のネステッドセット (nested set) を得るために複数の順方向プライマーまたは複数の逆方向プライマーを用いて PCR によって増幅される。このようなセットからの鋳型は、ある表面でさらに増幅して別個のアンプリコンを形成することができる (例えば、cBot インストゥルメント (cBot instrument)、イリノイ州、サンディエゴ、カリフォルニア州を用いるブリッジ PCR によって)。同じネステッドセットからの鋳型は、共通の末端で形成される配列リードによって互いに関連し得る。鋳型のネステッドセットにより、エラー率が比較的高いシーケンシング化学を使用して、他の場合には可能であろうより長い配列を分析することができると同時に、配列の全長に対して高い平均品質スコアを維持することができる。また、ネステッドセットにより、少なくとも 1 つの配列リードが、たとえ体細胞超変異を受けたとしても V 領域から得られる。

10

20

30

40

50

【0031】

IgH 分子の体細胞変異は、塩基の変化がシーケンシングエラーもしくは増幅エラーによるものであるか、または自然の変異プロセスによるものであるかの決定が困難であるため、配列リードデータからのクローン型の再構築をより困難にする。本発明に従った配列タグの使用は、IgH をコードする全ての核酸が異なるユニークな配列タグを受け取るため、このような困難さを大幅に緩和する。図 3 に例示されているように、本発明による配列リード (300) はそれぞれ、配列タグ (302) のコピーおよびクローン型 (304) (配列番号 1) のコピーを含む。同じ配列タグを有する全ての配列リードは、クローン型部分における各位置のヌクレオチドを比較できるように構築される (例えば、配列番号 2 および配列番号 3)。従って、たとえ IgH をコードする配列が僅か 1 つの塩基しか異なっていなくても、これらの配列は、異なる配列タグを受け取るため、サンプル中の密接に関連した IgH をコードする核酸は、クローン型決定プロセスで互いに比較されない。上述のように、配列タグにおけるエラーは、クローン型に関連した配列タグの配列が配列空間で非常に離れていて、1 つの配列タグが他のどの配列タグに対しても配列空間で近づくことなく、非常に多数の塩基の変化を維持できるため、それほど重要ではない。

【0032】

配列リードのデータからのクローン型の構築は、異なる方法は異なる予測リード長さおよびデータの質を有するため、このようなデータを生成するために使用されるシーケンシング法に一部依存する。1 つのアプローチでは、Solexa シーケンサが利用されて、分析用の配列リードデータが生成される。一実施形態では、少なくとも 100 万の鋳型分子を作製するために少なくとも $0.5 \sim 1.0 \times 10^6$ のリンパ球を提供するサンプルが得られ、任意選択の増幅後に、対応する鋳型分子 (クラスター) の 100 万以上のクローン集団を作製することができる。Solexa アプローチを含む最高のスループットのシーケンシングアプローチでは、クラスターレベルでのこのような過剰サンプリングは、各鋳型配列がかなりの重複性で決定されて配列決定の精度が上がるため望ましい。Solexa ベースの実施では、好ましくは、独立した各鋳型の配列は、10 回以上決定される。異なる予測リード長さおよびデータの質を用いる他のシーケンシングアプローチでは、異なるレベルの重複性を、配列決定の同等の精度を得るために使用することができる。当業者であれば、上述のパラメータ、例えば、サンプルのサイズおよび重複性などが、特定の適用例に関連する設計上の選択であることを理解されよう。

【0033】

図 2A ~ 図 2C は、2 段階 PCR でユニークな配列タグを組換え核酸分子に付着させる例示的なステップを例示している。T 細胞または B 細胞を含むサンプルからの組換え核酸分子の集団 (250) が、PCR 混合物中で順方向プライマー (202) および逆方向プライマー (262) と混合される。プライマー (262) はそれぞれ、3 つの領域: 標的アニーリング領域 (263) (この例示では、V 領域 (206) である); 配列タグ (264); および 2 段階 PCR の第 2 の段階用のプライマー結合領域 (265) を含む。こ

の例示では、プライマー（２６２）は、Ｖ領域の配列の多様性を考慮した標的アニーリング領域の混合物を含む。従って、全ての異なるプライマーが、配列タグ領域を用いて調製される。あるいは、配列タグ要素は、第２のＰＣＲ段階用のプライマー結合領域と共にＣ領域プライマー（２０２）に付着させることができる。前述のように、組換え核酸分子（２５０）は、定常またはＣ領域（２０３）、Ｊ領域（２１０）、Ｄ領域（２０８）、およびＶ領域（２０６）を含み、これらの領域は、ＴＣＲまたは免疫グロブリンのＣＤＲ３領域をコードするＶ（Ｄ）Ｊセグメントを表し得る。数サイクル、例えば、４～１０回のサイクル後に、第１段階のアンプリコン（２６６）が、配列タグ（２７０）を含む各メンバーのポリヌクレオチドを用いて作製される。第２段階のＰＣＲでは、アンプリコン（２６６）のポリヌクレオチドが、新たな順方向プライマーＰ５（２２２）および逆方向プライマーＰ７（２２０）を用いて再増幅され、これらのプライマーは、Solexa/IlluminaシーケンサでのブリッジＰＣＲを用いたクラスター形成用のさらなるプライマー結合部位（２２４）および（２２３）を追加する。プライマーＰ７は、１回のシーケンシングの実施中にサンプルの任意選択のマルチプレックス化のための二次配列タグ（２２１）も含む。二次ＰＣＲアンプリコン（２８０）が、組み込まれたＰ５およびＰ７で作製されたら、ブリッジＰＣＲを実施することができる。

10

20

30

40

50

【００３４】

配列データからのクローン型の決定

本発明の一態様によると、サンプルのクローン型は、それらの配列タブに基づいた配列リードの第１の分類によって決定される。このような分類は、従来のシーケンシングアラインメント法によって達成することができる。アラインメント法を選択するための案内は、参照により本明細書に組み入れられるバツゾグロウ（Batzoglou）著、ブリーフィング・イン・バイオインフォマティクス（Briefings in Bioinformatics）、２００５年、第６巻、ｐ．６～２２に記載されている。配列リードが、ユニークな配列タグに従って群に分けられたら、関連するクローン型の配列を分析して、サンプル由来のクローン型の配列を決定することができる。図３は、ユニークな配列タグに関連したクローン型の配列（配列番号１）の決定からの例示的なアラインメントおよび方法を例示している。この例では、１１の配列リードが、それぞれの各配列タグ（３０２）によって整列され、この後に、１、２、３、４、．．．ｎとして示されている配列リードのクローン型部分の各位置におけるヌクレオチドが比較される。例えば、位置６（３０６）のヌクレオチドは、ｔ、ｔ、ｇ、ｔ、ｔ、ｔ、ｔ、ｔ、ｔ、ｃ、ｔ；即ち、９つのベースコールがｔであり、１つが「ｇ」（３０８）であり、１つが「ｃ」（３１０）である（それぞれ、配列番号２および配列番号３）。一実施形態では、ある位置におけるクローン型配列の正しいベースコールとは、何であれ大多数の塩基が同じものである。位置６（３０６）の例では、ベースコールは、「ｔ」がこの位置にある配列リードの大多数のヌクレオチドであるため、「ｔ」である。他の実施形態では、クローン型配列の正しいベースコールを決定する際は、他の因子、例えば、配列リードのベースコールの品質スコアまたは隣接する塩基が同じであることを考慮することができる。

【００３５】

クローン型が上記のように決定されたら、サンプルの各異なるクローン型の存在量または頻度を含むクローン型プロファイルを作成することができる。

合成によるシーケンシング法でのホモポリマーの決定

本発明の一態様によると、配列タグを、合成によるシーケンシング（ＳＢＳ）法で生成されるシグナルからホモポリマー領域のヌクレオチド数の決定に使用することができる。即ち、標的ポリヌクレオチドが複製されて、複数の複製が、別々のＳＢＳ反応を受けると、各複製は、同じ配列タグで標識されている可能性があるため、別個の複製の同様の領域への塩基の取り込みから生じるＳＢＳシグナルが、異なる取り込みシグナルを共通の配列タグを通じて連結することによって分析されることができる。ＳＢＳ法は、様々な実施を有する。例えば、フラー（Fuller）ら著、ネイチャー・バイオテクノロジー（Nature Biotechnology）、２００９年、第２７巻、ｐ．１０１３～１

023;ロナギ(Ronaghi)ら著、サイエンス(Science)、1998年、第281巻、p.363~365;マーグリース(Margulies)ら著、ネイチャー(Nature)、2005年、第437巻、p.376~380;ロスバーク(Rothberg)ら著、ネイチャー(Nature)、2011年、第475巻、p.348~352;クマー(Kumar)ら著、サイエンティフィック・リポート(Scientific Reports)、2012年、第2巻、p.684;シム(Sims)ら著、ネイチャー・メソッズ(Nature Methods)、2011年、第8巻、p.575~580;ポアーマンド(Pourmand)ら著、米国科学アカデミー紀要(Proc. Natl. Acad. Sci.)、2006年、第103巻、p.6466~6470;およびセオ(Seo)ら著、米国科学アカデミー紀要(Proc. Natl. Acad. Sci.)、2005年、第102巻、p.5926~5931など。SBSを用いる市販のDNAシーケンサには、454 GSシーケンサ(454 ライフサイエンス(454 Life Sciences))、PGMおよびProton Ion Torrentシーケンサ(ライフテクノロジー(Life Technologies))、およびPyroMarkシーケンサ(キアゲン(Qiagen))などが含まれる。SBS法の多数の変法の共通の特徴は、一連のシグナル(「取り込みシグナル」)の生成であり、各シグナルは、鋳型主導性合成反応における伸長ヌクレオチド鎖に取り込まれたヌクレオチド数に比例する(または、少なくとも単調に関連する)。SBSのこのような変法の共通の課題は、ホモポリマー領域の長さが5~6ヌクレオチドを超える場合に鋳型のホモポリマー領域に組み込まれたヌクレオチド数を決定することである。本発明の一態様では、この問題は、配列タグの使用によって対応する。上記のように、このような実施形態における配列タグの配列は、互いに十分に離れているため、これらの配列タグは、たとえ複数の増幅エラーおよび/またはシーケンシングエラーが起きた後でも、それらの祖先配列または親配列に容易に関連付けられる。このようにして、これらの配列タグは、シーケンシングに頻繁に利用されるタグ、例えば、ベースコーリングソフトウェアを校正するためのタグ(例えば、いわゆる「キー」配列)、サンプルもしくは標本の起源を指定するタグ、または持越し汚染を低減するタグとは異なる。一部の実施形態では、鋳型は、モザイクタグを用いてサンプリングによって標識化されるため、配列タグ自体は、タグの最も長い可変領域の長さよりも長いホモポリマー領域を用いずを選択することができる。一部の実施形態では、モザイクタグは、5ヌクレオチド以下の可変領域と共に使用され;他の実施形態では、モザイクタグは、4ヌクレオチド以下の可変領域と共に使用される。あるいは、SBSに適した配列タグは、参照により本明細書に組み入れられるブレンナー(Brenner)による米国特許第7,537,897号明細書に開示されているように構築することができる。

【0036】

図4Aは、ヌクレオチドの付加および洗浄のサイクルを含むSBS法の一変法を例示し、シグナル生成は典型的には追加されるヌクレオチドの取り込みの最中または後で起こる。言い換えれば、例示されている変法は、異なるヌクレオチド三リン酸のそれぞれを連続的に付加するサイクルおよび洗浄するサイクルを含み、例えば、次の順序である:(a)成長する鋳型結合鎖へのdATPの付加および洗浄;(b)成長する鋳型結合鎖へのdGTPの付加および洗浄;(c)成長する鋳型結合鎖へのdCTPの付加および洗浄;および(d)成長する鋳型結合鎖へのdTTPの付加および洗浄。このようなサイクルの多くの追加の実施形態および変形形態も可能であり、例えば、シュルツ(Schultz)らによる米国特許出願公開第2012/0035062号明細書に開示され、合成によるシーケンシングの範囲内である。典型的には、伸長は、核酸ポリメラーゼ、例えば、DNAポリメラーゼを用いて行われる。図4Aに戻ると、SBSの一変法が示されており、プライマー結合部位(404)に隣接した配列領域(402)を有する一本鎖鋳型(400)が、固相支持体(408)、例えば、Ion Torrent(商標)半導体シーケンシングチップのマイクロウェルのSNAPPに固定される。例えば、参照により本明細書に組み入れられる、ノバイル(Nobile)らによる米国特許出願公開第2010/

10

20

30

40

50

0300895号明細書；ハインツ（Hinz）らによる米国特許出願公開第2011/0195253号明細書。固定された鋳型（400）は、シークエンシングプライマー（406）のプライマー結合部位（404）へのアニーリング、およびヌクレオシド三リン酸の付加時にポリメラーゼ（410）によるシークエンシングプライマー（406）の伸長を可能にする条件下で、シークエンシングプライマー（406）およびポリメラーゼ（410）と混合される。図4Aの変法では、鋳型（400）、シークエンシングプライマー（406）、およびポリメラーゼ（410）を含む複合体が、dATP、dGTP、dCTP、およびdTTPの流れ（フロー）に別々かつ連続的に曝露される。dATPの流れ（412）に曝露されると、シークエンシングプライマー（406）が、1つのdA（413）によって伸長される。残った全てのdATPを除去または破壊する洗浄（414）の後で、dGTPの流れ（416）が導入され、これにより、ポリメラーゼ（410）がdGTPを取り込んで、1つのdG（415）によってシークエンシングプライマー（406）を伸長させる。洗浄（418）後に、dCTPの流れ（420）が導入されるが、鋳型の次の塩基がdCに相補的でないため、ヌクレオチドが取り込まれず、シークエンシングプライマー（406）が伸長しない。このサイクルは、洗浄ステップ（422）およびdTTPの流れ（424）で完了する。

10

20

30

40

50

【0037】

このSBS変法および他のSBS変法における各取り込み事象の最中または後で、取り込みシグナルが生成され、この取り込みシグナルは、特性が非常に多様であり得、同時に2つ以上の物理的または化学的表示、例えば、光および/または電気を含み得る。例示的な取り込みシグナルの例として、限定されるものではないが、pHの変化、蛍光の変化、化学発光の変化、および抵抗の変化などが挙げられる。図4Bは、図4AのdATP、dGTP、およびdCTPの上記の各流れの最中または後で記録される典型的な取り込みシグナル（426、428、430）をそれぞれ例示している。このようなデータを分析して、プライマー（406）に付加されたヌクレオチドの数またはその伸長に比例する、または単調に関連する各ピークの数を得ることができる。このような分析で得られる数は、特定の実施形態によって異なり得、例えば、このような数は、このようなピークの面積、ピーク高さ、半値ピーク幅の関数、または当分野で公知の同様の関数であり得る。図4Cおよび図4Dは、ホモポリマー領域（432）の存在下でのSBSプロセス、および取り込みシグナルを決定するための配列タグの使用を例示している。支持体（436、438、440）はそれぞれ、別々のSBS反応1、SBS反応2、およびSBS反応Kに対応しており、これらは、付着された同じ配列タグ-鋳型コンジュゲートの複製を有する（配列番号5）。プライマー406、406'、および406''はそれぞれ、別々のポリメラーゼ410、410'、および410''によって別々の反応で伸長される。このような別々の反応は、別々の反応区画で行われ、これらの反応区画は、様々な形態、例えば、様々な規模のマイクロウェルのアレイを有する。例えば、参照により本明細書に組み入れられるロスバーク（Rothberg）ら著、ネイチャー（Nature）、2011年、第475巻、p.348~352；リーモン（Leamon）らによる米国特許第8,158,359号明細書；ロスバーク（Rothberg）らによる米国特許第8,313,625号明細書。反応区画は、例えば、参照により本明細書に組み入れられる、ボールス（Boles）らによる米国特許第6,300,070号明細書；バラスブラマニアン（Balasubramanian）らによる米国特許第6,787,308号明細書に開示されているブリッジPCRによって作製される、ビーズもしくは表面上またはゲル内の鋳型核酸のクラスターによって画定することもできる。SBS法の一部の実施形態によると、異なるdNTPの流れが、それぞれの反応に連続的に導入される。必ずしも必要ではないが、典型的には、各異なる反応は、例えば、適切な流動システムと共に使用されるIon Torrent半導体シークエンシングチップの異なるマイクロウェルで行われるSBS反応の場合と同様に、ほぼ同じ時間に同じ流れに曝露することができる。例えば、ノバイル（Nobile）（前掲）；参照により本明細書に組み入れられるデイビー（Daviey）らによる米国特許出願公開第2012/0143531号明細書。各dN

TPの流れの最中または後で、取り込みシグナル(i_1 、 i_2 、 $\dots$ 、 i_k)が得られる。図4Cは、第2のdATPの流れ(442)の取り込みシグナルを例示している。5~6塩基を超えるホモポリマーでは、最後の方に取り込まれるヌクレオチドの全シグナルに対する寄与は次第に小さくなり、従って、取り込まれたヌクレオチドからのシグナルとノイズとを区別することが困難になる。本発明によると、複数の複製鑄型からの取り込みシグナルは、共通の配列タグによって識別できるため一緒に分析することができる。上述のように、配列タグが選択されるため(例えば、サンプリングプロセスによる標識化によって)、たとえ複数の増幅エラーまたはシーケンシングエラーが存在しても、タグとポリヌクレオチドの各コンジュゲートの配列タグおよびその複製は、他のタグ-ポリヌクレオチドコンジュゲートの配列タグおよびその各複製と区別可能である。

10

【0038】

図4Dは、取り込みシグナル(450)の大きさの複数(1、2、 $\dots$ 、K)の配列であり、各配列は、共通の配列タグ(452)および鑄型TCGGGGGACCTT(配列番号5)に結合されている。例示されたデータセットの形式は、図3に例示されている形式に類似し、どのように塩基測定の列を使用して最終的なベースコールを決定できるかを例示している。図4Dは、どのように取り込みシグナルの列を使用してヌクレオチド取り込み数の最終決定をできるかを例示している。このような決定は、別個のモデルを使用して、別個の反応からの取り込みシグナルからヌクレオチド取り込みの測定値を求めた後で行ってもよいし、または、同じ親タグを有するタグ-ポリヌクレオチドコンジュゲートからの全てもしくは複数の取り込みシグナルにモデルを適用して行っても良い。上記のように、図4Dに示されている数字は、ピーク、例えば、(426、428、および430)によって例示されるデータセットの関数の値である(取り込みシグナルの時間に対する生成率を示し、その積分または面積は、ヌクレオチド取り込みの1つの指標である)。「f1a、f1g、f1c、f1t、f2a、f2g」などと標識された図4Dの列は、dNTPの流れのサイクルを表し、第1のサイクル(f1a、f1g、f1c、f1t)は、dATP、dGTP、dCTP、およびdTTPのSBS反応への連続的な導入を含み、以下同様である。

20

【0039】

上述のように、本発明のこの態様は、次のステップを有する方法によって行うことができる：(a)配列タグを1つ以上の各ポリヌクレオチドに付着させてタグ-ポリヌクレオチドコンジュゲートを形成するステップであって、タグ-ポリヌクレオチドコンジュゲートの実質的に全てのポリヌクレオチドがユニークな配列タグを有する、ステップ；(b)このタグ-ポリヌクレオチドコンジュゲートを増幅するステップ；(c)増幅されたタグ-ポリヌクレオチドコンジュゲートを合成によってシーケンシングするステップであって、少なくとも1つのdNTPの流れを含む、ステップ；および(d)同じ配列タグを有するタグ-ポリヌクレオチドのそれぞれについて、各dNTPの流れでのヌクレオチドの取り込み数を、このような各dNTPの流れに対して測定された取り込みシグナルの関数として決定するステップ。典型的には、増幅されたタグ-ポリヌクレオチドコンジュゲートは、様々なステップを行うのに役立つ隣接する配列を含む。例えば、タグ-ポリヌクレオチドコンジュゲートは、二次増幅、シーケンシングプライマーのアニーリングおよび伸長、ならびにハイブリダイゼーションベースの捕捉などの工程が可能となるように、各末端に1つ以上の隣接プライマー結合部位を含み得る。上述のように、dNTPの流れは、様々な流体送達装置および制御システム、例えば、最も市販されているDNAシーケンシングシステムに対して実施されるシステムによって実施することができる。例えば、参照により本明細書に組み入れられるシュルツ(Schultz)らによる米国特許出願公開第2012/0073667号明細書。dNTPの流れは、dNTP含有試薬の反応部位を通る移動である。好ましくは、dNTPの流れは、連続的な流れと流れとの間に流体の明確な境界が存在するように実施される。一部の実施形態では、洗浄ステップは、dNTP含有試薬とdNTPの存在を除き実質的に同一の試薬で実施される。一部の実施形態では、洗浄ステップを実施するための試薬は、前の流れのdNTPを破壊する、または

30

40

50

変性させる試薬、例えば、アピラーゼを含み得る。殆どの実施形態では、d N T Pの流れは、複合体中のポリメラーゼが、鋳型主導性反応における流れの中のd N T Pの取り込みによってシーケンシングプライマーをタグ - ポリヌクレオチドコンジュゲートに沿って伸長させるのに十分な全ての試薬、塩濃度、およびp Hを含む。上述のように、様々な関数またはアルゴリズムを使用して、タグ - ポリヌクレオチドコンジュゲートが同じ配列タグを有するようにすることによって関連した別々の反応からのシグナル（例えば、取り込みシグナル値）をモデル化することができる。一部の実施形態では、同じ配列タグを有するタグ - ポリヌクレオチドコンジュゲートからの取り込みシグナルの関数は、別々のS B S反応から測定されるこのような取り込みシグナルの平均に最も近い整数である。この平均は、算術平均または重み付け平均とすることができ、例えば、重み付け平均は、使用される特定のシーケンシングシステム、例えば、使用される特定のシーケンシング化学、標識が使用されるか否か、どんな種類のシグナルが生成されるか（例えば、光、抵抗パルス、p Hなど）、およびd N T Pの流れにおけるS B S反応部位の位置などの特徴に依存し得る。S B S反応からのシグナルを取り込まれたヌクレオチドに関連付ける代替のモデルの案内は、共に参照により本明細書に組み入れられる、リアーリック（R e a r i c k）による米国特許出願公開第2 0 1 2 / 0 1 7 2 2 4 1号明細書；およびハッベル（H u b b e l l）による米国特許出願公開第2 0 1 2 / 0 1 7 3 1 5 8号明細書に開示されている。

10

【0040】

上述のように、合成によるシーケンシングは、様々なプロセスステップおよび試薬によって行うことができる。一部の実施形態では、合成によるシーケンシングは、次のステップで行うことができる：（a）複数のタグとヌクレオチドのコンジュゲートのそれぞれに対して、シーケンシングプライマー、核酸ポリメラーゼ、およびタグ - ポリヌクレオチドコンジュゲートを含む複合体を、シーケンシングプライマーのこのようなタグ - ポリヌクレオチドコンジュゲートに対するアニーリングを可能にし、かつシーケンシングプライマーの、ヌクレオシド三リン酸の存在下での核酸ポリメラーゼによるこのタグ - ポリヌクレオチドコンジュゲートに沿った伸長を可能にする条件下で形成するステップ；（b）d N T Pをd N T Pの流れによって複合体に導入するステップ；（c）各複合体からの取り込みシグナルを測定するステップ；（d）複合体を洗浄するステップ；および（e）ステップ（b）からステップ（d）を繰り返すステップ。一部の実施形態では、タグ - ポリヌクレオチドコンジュゲートが増幅されたら、得られたアンプリコンのサンプルが、合成ステップによるシーケンシングに使用される。一部の実施形態では、このようなサンプルは、少なくとも 10^2 を超える数のタグ - ポリヌクレオチドコンジュゲートを含み；他の実施形態では、このようなサンプルは、 $10^2 \sim 10^{15}$ の範囲、または他の実施形態では $10^2 \sim 10^{12}$ の範囲、または他の実施形態では $10^2 \sim 10^9$ の範囲の数のタグ - ポリヌクレオチドコンジュゲートを含む。他の実施形態では、このようなサンプルは、0.001以上の頻度を有する全てのポリヌクレオチドが99%の確率で存在する；または0.0001以上の頻度を有する全てのポリヌクレオチドが99%の確率で存在するように十分に大きい。

20

30

【0041】

本発明のこの態様に利用されるd N T Pは、非ブロックd N T Pまたは伸長ブロックd N T Pとすることができる。前者（非ブロックd N T P）の場合は、1回のd N T Pの流れの最中に、鋳型がホモポリマー領域を含む場合は、ホモポリマーの長さに等しい数の取り込みが起こり得る。（取り込みを少なくする因子（例えば、処理能力の低いポリメラーゼ）が存在し得るが、非ブロックd N T Pでは、このような数をホモポリマーの長さに等しくすることが可能である）。例えば、合成化学によるI o n T o r r e n tシーケンシングは、非ブロックd N T Pを利用する。伸長ブロックd N T Pの場合は、通常は、取り込まれるヌクレオチド上に化学ブロック基が存在するため、各流れで1つのd N T Pの取り込みしか起こり得ない。従って、d N T Pの取り込みのサイクルは、伸長されたシーケンシングプライマーの伸長可能な末端を再生するために、ブロックを解除するステ

40

50

ップをさらに含む。例えば、合成化学によるイルミナ (Illumina) シークエンシングは、伸長ブロック dNTP を利用する。例えば、ベントレイ (Bentley) ら (後述)。一部の実施形態では、ブロックを解除するステップは、洗浄ステップの一部として行うことができる。

【0042】

配列タグを使用する持越し汚染の検出

持越し汚染 (キャリアオーバーコンタミネーション) は、核酸の増幅を含む技術では大きな問題である。例えば、ボースト (Borst) ら著、ヨーロピアン・ジャーナル・オブ・クリニカル・マイクロバイオロジー・アンド・インフェクシャス・ディゼイズ (Eur. J. Clin. Microbiol. Infect. Dis.), 2004 年、第 2 3 巻、第 4 号、p. 289 ~ 299; アスランザデ・アン (Aslanzadeh, Ann.), クリニカル・ラボラトリー・サイエンス (Clin. Lab. Sci.), 2004 年、第 34 巻、第 4 号、p. 389 ~ 396 など。このような汚染は、サンプル外の微量の核酸がサンプルのアッセイで意図せずに増幅された場合に起こり、測定結果に変化をもたらす、または悪影響を与える。最悪の場合には、患者由来の医学的サンプル中の持越し汚染により、アッセイの結果が疑陽性に見なされることも起こり得る。外来核酸は、特定の患者に無関係の起源に由来し得る; 例えば、外来核酸は、別の患者のサンプルに由来し得る。あるいは、外来核酸は、患者に関連した起源に由来し得る; 例えば、外来核酸は、過去に同じ研究室で取り扱われた同じ患者からの異なるサンプル、または過去に同じ研究室で処理された同じ患者からの異なるサンプルに対するアッセイ反応に由来し得る。

10

20

【0043】

持越し汚染は、関連する核酸の非常に複雑な集団、例えば、免疫分子、例えば、T 細胞受容体または免疫グロブリンをコードする組換え核酸の集団を測定する場合に臨床現場で特に問題である。この問題は、配列リードもしくはクローン型が、目的のサンプルの真の多様性の一部であるか否か、または配列リードもしくはクローン型が、同じ研究室で同じ種類のアッセイで処理される、核酸の外部起源、例えば、別の患者のサンプルもしくは同じ患者の前のサンプルに由来するか否かを決定するのが困難であるために生じる。本発明の一態様では、このような持越し汚染は、配列タグを用いて、配列リードからクローン型を決定するだけでなく、配列タグが現在のサンプルまたは別のサンプルに由来するか否かを決定することによって検出することができる。これは、各患者のサンプルから決定される配列タグの記録を保存して、次の測定が行われる度に現在の測定の配列タグを前の測定の配列タグと比較することによって達成される。クローン型に関連した配列タグのこのような記録は、各測定からのタグの数が多く、従来のアルゴリズムを用いた電子記録の検索および比較が容易であるため、電子記録として大容量記憶装置に便利に保存される。一致が検出された場合、測定に利用された配列タグの集団が十分に大きいとすると、最も可能性が高い原因は持越し汚染である。上記のサンプリングによる標識化のためのクローン型の集団に対する配列タグの集団のサイズの同じ例示的な比を、持越し汚染の検出に適用可能である。一実施形態では、このような比は 100 : 1 以上である。

30

【0044】

サンプル

40

クローン型プロファイルは、様々な組織に存在する免疫細胞のサンプルから得られる。目的の免疫細胞は、T 細胞および B 細胞の少なくとも一方を含む。T 細胞 (T リンパ球) は、例えば、T 細胞受容体 (TCR) を発現する細胞を含む。B 細胞 (B リンパ球) は、例えば、B 細胞受容体 (BCR) を発現する細胞を含む。T 細胞は、細胞表面マーカーによって区別することができる、ヘルパー T 細胞 (エフェクター T 細胞または Th 細胞)、細胞傷害性 T 細胞 (CTL: cytotoxic T cell)、記憶 T 細胞、および制御性 T 細胞を含む。一態様では、T 細胞のサンプルは、少なくとも 1,000 の T 細胞を含むが; より典型的には、サンプルは、少なくとも 10,000 の T 細胞、より典型的には、少なくとも 100,000 の T 細胞を含む。別の態様では、サンプルは、1000 ~ 1,000,000 細胞の範囲の数の T 細胞を含む。免疫細胞のサンプルは、B 細胞も含

50

み得る。B細胞は、例えば、プラズマB細胞、記憶B細胞、B1細胞、B2細胞、辺縁帯B細胞、および濾胞B細胞を含む。B細胞は、免疫グロブリン（抗体またはB細胞受容体とも呼ばれる）を発現し得る。上記のように、一態様では、B細胞のサンプルは、少なくとも1,000のB細胞を含むが；より典型的には、サンプルは、少なくとも10,000のB細胞、より典型的には、少なくとも100,000のB細胞を含む。別の態様では、サンプルは、1000～1,000,000B細胞の範囲の数のB細胞を含む。

【0045】

本発明の方法に使用されるサンプル（「組織」サンプルとも呼ばれることもある）は、例えば、腫瘍組織含む様々な組織、血液および血漿、リンパ液、脳および脊髄を取り囲む脳脊髄液、ならびに骨関節を取り囲む滑液などに由来し得る。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0.1mL、約0.2mL、約0.3mL、約0.4mL、約0.5mL、約0.6mL、約0.7mL、約0.8mL、約0.9mL、約1.0mL、約1.5mL、約2.0mL、約2.5mL、約3.0mL、約3.5mL、約4.0mL、約4.5mL、または約5.0mLとすることができる。サンプルは、腫瘍生検標本とすることができる。腫瘍生検標本は、例えば、脳、肝臓、肺、心臓、結腸、腎臓、または骨髄の腫瘍に由来し得る。当業者によって使用される任意の生検技術を、対象からサンプルを採取するために使用することができる。例えば、生検は、全身麻酔が使用される直視下生検とすることができる。生検は、直視下生検で形成される切開部よりも小さい切開部が形成される経皮的生検とすることができる。生検は、組織の一部が除去されるコア生検または切開生検とすることができる。生検は、病変全体を除去しようとする切除生検とすることができる。生検は、組織または液体のサンプルが針で除去される微細針吸引生検とすることができる。

【0046】

サンプルまたは組織サンプルは、核酸、例えば、DNA（例えば、ゲノムDNA）またはRNA（例えば、メッセンジャーRNA）を含む。核酸は、例えば、循環系から抽出される無細胞DNAまたはRNAとすることができる。ウラソフ（Vlassov）ら著、カレント・モレキュラー・メディシン（Curr. Mol. Med.）、2010年、第10巻、p. 142～165；スワラップ（Swarup）ら著、FEBSレターズ（FEBS Lett.）、2007年、第581巻、p. 795～799。本発明の方法では、分析することができる対象由来のRNAまたはDNAの量は様々である。例えば、1つの細胞のDNAまたはRNAが、校正試験（即ち、ある疾患の相関するクローン型を決定する最初の測定）に必要な全てであり得る。クローン型プロファイルを作成するためには、個々の免疫受容体レパートリーの有用な表現を得るために十分な核酸がサンプル中に存在しなければならない。より具体的には、ゲノムDNAからクローン型プロファイルを作成するためには、T細胞またはB細胞からの少なくとも1ngの全DNA（即ち、約300の2倍体ゲノム当量）がサンプルから抽出され；別の実施形態では、少なくとも2ngの全DNA（即ち、約600の2倍体ゲノム当量）がサンプルから抽出され；そして別の実施形態では、少なくとも3ngの全DNA（即ち、約900の2倍体ゲノム当量）がサンプルから抽出される。当業者であれば、サンプル中のリンパ球の割合が減少すると、約1000を超える独立したクローン型を含むクローン型プロファイルを作成するために、前述の最小量のDNAを増加させなければならないことを理解されよう。RNAからクローン型プロファイルを作成するためには、一実施形態では、異なるTCR、BCR、またはそれらの断片をコードする少なくとも1000の転写物が得られるように十分な量のRNAが抽出される。この範囲に対応するRNAの量は、サンプル中のリンパ球の割合およびリンパ球の分化段階などによってサンプルごとに大きく異なる。一実施形態では、少なくとも100ngのRNAが、クローン型プロファイルの作成のためのB細胞および/またはT細胞を含む組織サンプルから抽出され；別の実施形態では、少なくとも500ngのRNAが、クローン型プロファイルの作成のためのB細胞および/またはT細胞を含む組織サンプルから抽出される。本発明の方法に使用されるRNAは、組織サンプルから抽出される全RNA、組織サンプルから直接抽出されるポリA RNA、または組織サン

ブルから抽出される全RNAであり得る。上記の核酸の抽出は、例えば、インビトロジェン (Invitrogen) (カルスバッド、カリフォルニア州)、キアゲン (Qiagen) (サンディエゴ、カリフォルニア州)、または同様の業者が市販するキットを用いて行うことができる。RNA抽出の案内は、リートケ (Liedtke) ら著、PCRメソッド・アンド・アプリケーションズ (PCR Methods and Applications)、1994年、第4巻、p. 185~187および同様の文献に記載されている。

【0047】

より完全に後述されるように(定義)、リンパ球を含むサンプルは、異なるクローン型を有する実質的に全てのT細胞またはB細胞がサンプル中に存在し、レパートリ(本明細書で使用される用語として)を形成できるように十分に大きい。一実施形態では、99%の確率で集団のどのクローン型も0.001以上の頻度で存在するサンプルが採取される。別の実施形態では、99%の確率で集団のどのクローン型も0.0001以上の頻度で存在するサンプルが採取される。一実施形態では、B細胞またはT細胞のサンプルは、少なくとも50万の細胞を含み、別の実施形態では、このようなサンプルは、少なくとも100万の細胞を含む。

10

【0048】

サンプルが採取される材料の供給源が十分でない場合、例えば、臨床試験のサンプルなどの場合は、この材料に由来するDNAを、偏りのない技術、例えば、全ゲノム増幅(WGA: whole genome amplification)、複数変位増幅(MDA: multiple displacement amplification)、または同様の技術によって増幅することができる。例えば、ホーキンス(Hawkins)ら著、カレント・オピニオン・イン・バイオテクノロジー(Curr. Opin. Biotech.)、2002年、第13巻、p. 65~67; ディーン(Dean)ら著、ゲノム・リサーチ(Genome Research)、2001年、第11巻、p. 1095~1099; ワン(Wang)ら著、ヌクレイック・アシッド・リサーチ(Nucleic Acids Research)、2004年、第32巻、p. e76; およびホソノ(Hosono)ら著、ゲノム・リサーチ(Genome Research)、2003年、第13巻、p. 954-964など。

20

【0049】

血液サンプルは、特に興味深く、従来の技術を用いて得ることができる。例えば、インニス(Innis)ら編、PCRプロトコル(PCR Protocols)、アカデミックプレス(Academic Press)、1990年など。例えば、白血球は、従来の技術、例えば、RosetteSep(商標)キット(ステムセル・テクノロジーズ(Stem Cell Technologies)、バンクーバーカナダ)を用いて血液サンプルから分離することができる。同様に、全血の他の分画、例えば、末梢血単核細胞(PBMC)は、市販のキット、例えば、ミルテニー・バイオテック(Miltenyi Biotec)、オーバーン、カリフォルニア州などを用いて本発明の方法で使用するために単離することができる。血液サンプルは、100 μ L~10mLの範囲の量とすることができる; 一態様では、血液サンプルの量は、200~1000 μ L~2mLの範囲である。次いで、DNAおよび/またはRNAを、本発明の方法に使用される従来の技術、例えば、DNeasy Blood & Tissueキット(キアゲン(Qiagen)、バレンシア、カリフォルニア州)を用いてこのような血液サンプルから抽出することができる。任意選択で、白血球、例えば、リンパ球のサブセットを、従来の技術、例えば、蛍光活性化細胞分類(FACS)(ベクトン・ディッキンソン(Becton Dickinson)、サンノゼ、カリフォルニア州)、または磁気活性化細胞分類(MACS)(ミルテニー・バイオテック(Miltenyi Biotec)、オーバーン、カリフォルニア州)などを用いてさらに単離することができる。

30

40

【0050】

一部の実施形態では、組換え核酸が、各個人の適応免疫細胞のDNA中およびそれらの

50

関連するRNA転写物中に存在し、これにより、RNAまたはDNAを提供される本発明の方法でシーケンシングすることができる。T細胞受容体または免疫グロブリン分子、またはそれらの一部をコードする、T細胞またはB細胞由来の組換え配列がクローン型と呼ばれる。このDNAまたはRNAは、T細胞受容体(TCR)遺伝子または抗体をコードする免疫グロブリン(Ig: immunoglobulin)遺伝子由来の配列に一致し得る。例えば、このDNAおよびRNAは、TCRの α 鎖、 β 鎖、 γ 鎖、または δ 鎖をコードする配列に一致し得る。大部分のT細胞では、TCRは、 α 鎖および β 鎖からなるヘテロ二量体である。TCR α 鎖は、VJ組換えによって形成され、 β 鎖は、V(D)J組換えによって形成される。TCR β 鎖の場合、ヒトでは、48のVセグメント、2つのDセグメント、および13のJセグメントが存在する。2つの接合部のそれぞれにおいて、いくつかの塩基を欠失させ、他の塩基を付加することができる(NヌクレオチドおよびPヌクレオチドと呼ばれる)。少数のT細胞では、TCRは γ 鎖および δ デルタ鎖からなる。TCR γ 鎖は、VJ組換えによって形成され、TCR δ 鎖は、V(D)J組換えによって形成される(参照によりその全開示内容が本明細書に組み入れられる、ケネス・マーフィ(Kenneth Murphy)、ポール・トラバース(Paul Travers)、およびマーク・ウォルポート(Mark Walport)著、ジェンウェイズ・イムノロジー(Janeway's Immunology)、第7版、ガーランド・サイエンス(Garland Science)、2007年)。

【0051】

本発明の方法で分析されるDNAおよびRNAは、定常領域(α 、 δ 、 ϵ 、 γ 、もしくは μ)を有する重鎖免疫グロブリン(IgH)または定常領域 λ もしくは κ を有する軽鎖免疫グロブリン(IgKもしくはIgL)をコードする配列に一致し得る。各抗体は、2つの同一の軽鎖および2つの同一の重鎖を有する。各鎖は、定常(C)領域および可変領域を有する。重鎖の場合、可変領域は、可変(V)セグメント、多様性(D)セグメント、および接合(J)セグメントから構成されている。これらのセグメントの各タイプをコードするいくつかの異なる配列がゲノムに存在する。特定のVDJ組換え事象が、B細胞の分化中に起こり、細胞が特定の重鎖を形成することを示している。軽鎖の多様性は、D領域が存在せず、VJ組換えのみが存在する点を除き、同様に形成される。体細胞変異は、組換え部位の近傍で起こる場合が多く、いくつかのヌクレオチドの付加または欠失を引き起こし、B細胞によって産生される重鎖および軽鎖の多様性をさらに高める。従って、B細胞によって産生される抗体の可能な多様性は、異なる重鎖と軽鎖の積である。重鎖および軽鎖の可変領域は、抗原認識(または結合)領域または部位の形成に寄与する。この多様性に、特定の応答が一部のエピトープで生じた後に起こり得る体細胞超変異のプロセスが加えられる。

【0052】

本発明によると、プライマーは、リンパ球から抽出された組換え核酸のアンプリコンのサブセットを形成するために選択することができる。このようなサブセットは、本明細書では「体細胞的に再配列された領域」と呼ぶこともある。体細胞的に再配列された領域は、分化中のリンパ球または完全に分化したリンパ球由来の核酸を含み得、分化中のリンパ球は、免疫遺伝子の再配列が完了せず、完全なV(D)J領域を有する分子が形成されていない細胞である。例示的な不完全な体細胞的に再配列された領域は、不完全なIgH分子(例えば、D-J領域のみを含む分子)、不完全なTCR δ 分子(例えば、D-J領域のみを含む分子)、および不活性なIgK(例えば、Kde-V領域を含む)を含む。

【0053】

細胞の十分なサンプリングは、「クローン型」および「レパートリ」の定義で以下にさらに説明されるレパートリデータを解釈する上で重要な点である。例えば、1,000の細胞で開始すると、いくつかのシーケンシングリードが得られるかに関わらずアッセイが高感度である最小頻度を得られる。従って、本発明の一態様は、入力免疫受容体分子の数を定量する方法の開発である。これは、TCR β 配列およびIgH配列に対して実施された。いずれの場合も、異なる配列全てを増幅することができる同じプライマーのセットが

使用される。コピーの絶対数を得るために、複数のプライマーを用いるリアルタイムPCRが、既知の数の免疫受容体コピーで標準に沿って行われる。このリアルタイムPCR測定は、後にシーケンシングされる増幅反応から行っても良いし、または同じサンプルの別個のアリコートに対して行っても良い。DNAの場合は、再配列免疫受容体分子の絶対数を、細胞数に容易に変換することができる（一部の細胞が、評価された特定の免疫受容体の2つの再配列コピーを有し、他が1つを有するため2倍以内）。cDNAの場合は、リアルタイムサンプルにおける再配列分子の測定総数を外挿して、同じサンプルの別の増幅反応に使用されたこれらの分子の総数を決定することができる。加えて、この方法は、RNAの総量を決定する方法と組み合わせて、特定の効率のcDNA合成と仮定して、RNAの単位量（例えば、1 μ g）中の再配列免疫受容体分子の数を決定することができる。cDNAの総量が測定されると、cDNA合成の効率を考慮する必要がない。細胞数も分かる場合は、1細胞当たりの再配列免疫受容体コピーを計算することができる。細胞数が分からない場合は、特定の種類の細胞が通常は同量のRNAを産生するため、全RNAから細胞数を推定することができる。従って、1 μ g当たりの再配列免疫受容体分子のコピーから、1細胞当たりのこれらの分子の数を推定することができる。

10

【0054】

シーケンシングのために実施され得る反応とは別のリアルタイムPCRを行う1つの不都合な点は、異なる酵素、入力DNA、および他の条件を利用し得るため、リアルタイムPCRで他の反応とは異なる阻害効果が存在し得ることである。シーケンシングのためにリアルタイムPCRの産物を処理することにより、この問題を改善することができる。しかしながら、リアルタイムPCRを用いた低いコピー数は、コピーの数が少ないか、または反応中の阻害効果もしくは他の最適以下の条件が原因であろう。

20

【0055】

cDNAまたはゲノムDNAに対する既知の量の1つ以上の内部標準をアッセイ反応に加えて、未知の量のcDNAまたはゲノムDNAサンプルの絶対量または絶対濃度を決定することができる。内部標準の分子数を計数して、これを同じサンプルの残りの配列と比較することによって、最初のcDNAサンプル中の再配列免疫受容体分子の数を推定することができる。（分子を計数するような技術は周知である。例えば、参照により本明細書に組み入れられるブレンナー（Brenner）らによる米国特許第7,537,897号明細書）。

30

【0056】

核酸集団の増幅

核酸の標的集団のアンプリコン、特に組換え免疫分子を、様々な増幅技術によって作製することができる。本発明の一態様では、マルチプレックスPCRを使用して、組換え免疫分子、例えば、T細胞受容体もしくはその一部、またはB細胞受容体もしくはその一部の混合物のメンバーが増幅される。このような免疫分子のマルチプレックスPCRを行うための案内は、参照により本明細書に組み入れられる次の文献に記載されている：ファハム（Faham）およびウィリス（Willis）による米国特許第8,236,503号明細書；モーレイ（Morley）による米国特許第5,296,351号明細書；ゴルスキー（Gorski）による米国特許第5,837,447号明細書；ダウ（Dau）による米国特許第6,087,096号明細書；ヴァン・ドンゲン（Von Dongen）らによる米国特許出願公開第2006/0234234号明細書；欧州特許第1544308B1号明細書など。本発明の一部の実施形態では、クローン型プロファイルを作成するステップは、（a）T細胞受容体遺伝子の一部および/またはB細胞受容体遺伝子の一部を増幅するステップ、および（b）得られるアンプリコンの核酸をシーケンシングするステップを含む。他の部分で説明されるように、シーケンシングされたアンプリコンの核酸の数は、適用例によって異なり得る。例えば、白血病患者がなお寛解期にあるか否かを決定するクローン型プロファイルは、任意の腫瘍クローンの検出限界が非常に低い場合大きくなる。一部の実施形態では、シーケンシングされたアンプリコンの核酸の数は、少なくとも1000であり；他の実施形態では、シーケンシングされたアンプ

40

50

リコンの核酸の数は、少なくとも 10^4 であり；他の実施形態では、シーケンシングされたアンプリコンの核酸の数は、少なくとも 10^5 である。このような作成ステップはまた、配列リードを合体させてクローン型にするステップ、クローン型を列挙する、または表にするステップ、クローン型の度数分布を作成するステップ、クローン型の関連するサブセットを識別するステップ、およびクローン型の頻度の情報を表示するステップなどのさらなるステップも含み得る。

【0057】

ゲノム由来のDNAの増幅（またはRNAの逆転写によるcDNAの形態の核酸の増幅）の後に、個々の核酸分子を単離し、任意選択で再増幅し、そして個々にシーケンシングすることができる。例示的な増幅プロトコルは、参照により本明細書に組み入れられる、バン・ドンゲン（van Dongen）らによるルーキミア（Leukemia）、2003年、第17巻、p. 2257～2317またはバン・ドンゲン（van Dongen）らによる米国特許出願公開第2006/0234234号明細書に記載されている。簡単に述べると、例示的なプロトコルは次の通りである：反応バッファー：ABI Buffer IIまたはABI Gold Buffer（ライフテクノロジーズ（Life Technologies）、サンディエゴ、カリフォルニア州）；50 μ lの最終反応量；100 ngのサンプルDNA；10 pモルの各プライマー（後述されるように増幅をバランスさせるために調整される）；200 μ Mの最終濃度のdNTP；1.5 mMの最終濃度のMgCl₂（標的配列およびポリメラーゼによっては最適化される）；Taqポリメラーゼ（1～2 U/管）；サイクリング条件：95 で7分の前活性化；60 のアニーリング；サイクリング時間：30秒の変性；30秒のアニーリング；30秒の伸長。本発明の方法の増殖に使用できるポリメラーゼは、市販されており、例えば、Taqポリメラーゼ、AccuPrimeポリメラーゼ、またPfuを含む。使用するポリメラーゼの選択は、忠実度または効率のどちらが優先されるかに基づくことができる。

10

20

【0058】

リアルタイムPCR、ピコグリーン（picogreen）染色、ナノ流体電気泳動法（例えば、LabChip）、またはUV吸収測定を最初のステップで使用して、増幅可能な材料の機能的な量を見積もることができる。

【0059】

一態様では、開始集団の配列の相対量が、増幅された集団またはアンプリコンの相対量と実質的に同じであるように、マルチプレックス増幅が行われる。即ち、マルチプレックス増幅は、サンプル集団のメンバーの配列の間で最小の増幅バイアスで行われる。一実施形態では、このような相対量は、アンプリコンの各相対量が、開始サンプルの相対量の値の5倍以内である場合は実質的に同じである。別の実施形態では、このような相対量は、アンプリコンの各相対量が、開始サンプルの相対量の値の2倍以内である場合は実質的に同じである。以下にさらに十分に説明されるように、PCRの増幅バイアスは、従来の技術を用いて検出して修正することができ、これにより、PCRプライマーのセットを所定のレパートリのために選択して、あらゆるサンプルのバイアスのない増幅を実現することができる。

30

40

【0060】

TCR配列またはBCR配列に基づいた多くのレパートリに関して、マルチプレックス増幅は、任意選択で全てのVセグメントを使用する。この反応は、異なるVセグメントプライマーによって増幅される配列の相対存在量を維持する増幅が得られるように最適化される。プライマーの一部は関連し、従って、プライマーの多くはクロストークして、プライマーに完全には一致していない鋳型を増幅し得る。各鋳型が、どのプライマーが鋳型を増幅するかにかかわらず同様の方式で増幅できるように条件が最適化される。言い換えれば、2つの鋳型が存在する場合、1,000倍の増幅後に、両方の鋳型を約1,000倍に増幅することができ、一方の鋳型のために、増幅産物の半分が、クロストークのために異なるプライマーを有していても問題ない。シーケンシングデータの後の分析では、プ

50

ライマー配列が分析から排除されるため、鋳型が等しく増幅されれば、増幅にどのプライマーが使用されるかは問題ではない。

【0061】

一実施形態では、増幅バイアスは、2段階増幅（前掲のファハム（Faham）およびウィリス（Willis）に記載されている）を行うことによって回避することができ、この増幅では、少数の増幅サイクルが、標的配列に相補的ではない尾部を有するプライマーを用いて第1の段階または一次段階で行われる。この尾部は、一次アンプリコンの配列の末端に付加されたプライマー結合部位を含むため、このような部位が、1つの順方向プライマーと1つの逆方向プライマーのみを用いる第2段階増幅で使用され、これにより、増幅バイアスの主な原因が排除される。好ましくは、一次PCRは、異なるプライマーによる差動増幅を最小限にするために十分に少ない数のサイクル（例えば、5～10回）を有する。二次増幅は、一对のプライマーで行われるため、差動増幅の問題は最小限である。一次PCRの1パーセントが、二次PCRに直接使用される。2つの増幅の間に使用される35のサイクル（100倍の希釈ステップを用いない約28のサイクルに等しい）が、サイクルの詳細：1サイクルの一次と34の二次または25の一次と10の二次、にかかわらず強い増幅を示すのに十分であった。理想的には、一次PCRでの1サイクルのみの実施で、増幅バイアスを減少できることであるが、他の考えも存在する。この一態様はレプリゼンテーション（representation）である。これは、開始入力量が、最終的に得られるリード数を超えていない場合に役割を果たす。例えば、1,000,000のリードが得られ、1,000,000の入力分子で開始される場合、二次増幅に対する100,000の分子のレプリゼンテーション（representation）のみをとると、元のサンプルにおける異なる種の相対存在量の推定の精度が低下し得る。2つのステップの間の100倍の希釈とは、一次PCRが100を大幅に超える分子を作製しない限り、レプリゼンテーション（representation）が減少することを意味する。これは、最少で8サイクル（256倍）、より快適には10サイクル（約1,000倍）を使用することができることを示す。これに対する代替は、一次PCRの1%超を二次PCRに使用するが、一次PCRに使用される高濃度のプライマーのため、これらのプライマーが増幅を妨げず、配列間の増幅バイアスを悪化させないように大きい希釈倍数を使用することができる。別の代替は、精製ステップまたは酵素的ステップを加えて一次PCRのプライマーを除去して、より小さい希釈を可能にする。この例では、一次PCRが10サイクル、二次PCRが25サイクルであった。

10

20

30

【0062】

組換え核酸の集団のシーケンシング

核酸シーケンシングの任意の高スループット技術を、本発明の方法に使用することができる。好ましくは、このような技術は、少なくとも1000のクローン型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のクローン型を決定することができる量の配列データをコスト効率の良い方法で作製する能力を有する。DNAシーケンシング技術としては、標識ターミネータまたはプライマーを用いた古典的ジデオキシシーケンシング反応（サンガー法）およびスラブまたは毛细管でのゲル分離、可逆的に終端された標識ヌクレオチドを用いた合成によるシーケンシング、パイロシーケンシング、454シーケンシング、標識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ハイブリダイゼーション、後にライゲーションされる標識クローンのライブラリに対する対立遺伝子特異的ハイブリダイゼーションを用いた合成によるシーケンシング、重合ステップ中の標識ヌクレオチドの取り込みのリアルタイム監視、コロニー（colony）シーケンシング、およびSOLiDシーケンシングが挙げられる。分離された分子のシーケンシングは、最近になって、ポリメラーゼまたはリガーゼを用いた連続または単一伸長反応、およびプローブのライブラリを用いた単一または連続差動ハイブリダイゼーションによって実証された。これらの反応は、1億を超える平行の配列の現在の市販アプリケーションでの実証を含め、多数の平行のクローン配列に対して行われた。従って、これらのシーケンシングアプローチを使用して、T細胞受容

40

50

体 (T C R) および / または B 細胞受容体 (B C R) のレパートリを研究することができる。本発明の一態様では、シーケンシングの高スループット法が利用され、この方法は、個々の分子を固体表面上で空間的に分離して、これらの分子を平行にシーケンシングするステップを含む。このような固体表面としては、無孔表面 (例えば、S o l e x a シーケンシング、例えば、ベントレイ (B e n t l e y) 著、ネイチャー (N a t u r e) 、 2 0 0 8 年、第 4 5 6 巻、p . 5 3 ~ 5 9 または C o m p l e t e G e n o m i c s シーケンシング、例えば、ドルマナック (D r m a n a c) ら著、サイエンス (S c i e n c e) 、 2 0 1 0 年、第 3 2 7 巻、p . 7 8 ~ 8 1) 、ビーズ結合鋳型もしくは粒子結合鋳型を含み得るウェルのアレイ (例えば、4 5 4 シーケンシングを用いる、例えば、マングリース (M a r g u l i e s) ら著、ネイチャー (N a t u r e) 、 2 0 0 5 年、第 4 3 7 巻、p . 3 7 6 ~ 3 8 0 または I o n T o r r e n t シーケンシング、米国特許出願公開第 2 0 1 0 / 0 1 3 7 1 4 3 号明細書または同第 2 0 1 0 / 0 3 0 4 9 8 2 号明細書) 、微細加工膜 (例えば、S M R T シーケンシングと同様、例えば、エイド (E i d) ら著、サイエンス (S c i e n c e) 、 2 0 0 9 年、第 3 2 3 巻、p . 1 3 3 ~ 1 3 8) 、またはビーズアレイ (S O L i D シーケンシングまたはポロニー (p o l o n y) シーケンシングと同様、例えば、キム (K i m) ら著、サイエンス (S c i e n c e) 、 2 0 0 7 年、第 3 1 6 巻、p . 1 4 8 1 ~ 1 4 1 4) を挙げることができる。別の態様では、このような方法は、単離された分子を、これらの分子が固体表面上で空間的に単離される前または後で増幅するステップを含む。前増幅は、エマルジョンをベースとした増幅、例えば、エマルジョン P C R 、またはローリングサークル増幅を含み得る。特に興味深いのは S o l e x a ベースのシーケンシングであり、このシーケンシングでは、個々の鋳型分子が、固体表面上で空間的に分離された後に、これらの分子が、ブリッジ P C R によって平行に増幅されて、別個のクローン集団またはクラスターが形成され、次いで、ベントレイ (B e n t l e y) (前掲) ならびに製造者の取扱説明書 (例えば、T r u S e q (商標) サンプル精製キットおよびデータシート、イルミナ社 (I l l u m i n a , I n c .) 、サンディエゴ、カリフォルニア州、2 0 1 0 年) ; および参照により本明細書に組み入れられる次の文献 : 米国特許第 6 , 0 9 0 , 5 9 2 号明細書 ; 同第 6 , 3 0 0 , 0 7 0 号明細書 ; 同第 7 , 1 1 5 , 4 0 0 号明細書 ; および欧州特許第 0 9 7 2 0 8 1 B 1 号明細書に記載されているようにシーケンシングされる。一実施形態では、個々の分子が、固定表面に配置されて増幅され、少なくとも 10^5 クラスタ / cm^2 の密度 ; または少なくとも 5×10^5 / cm^2 の密度 ; または 10^6 クラスタ / cm^2 の密度でクラスターが形成される。一実施形態では、比較的高いエラー率のシーケンシング化学が利用される。このような実施形態では、このような化学によって生成される平均品質スコアは、配列リードの長さの単調に減少する関数である。一実施形態では、このような減少は、配列リードの 0 . 5 % が、位置 1 ~ 7 5 に少なくとも 1 つのエラーを有すること ; 配列リードの 1 % が、位置 7 6 ~ 1 0 0 に少なくとも 1 つのエラーを有すること ; 配列リードの 2 % が、位置 1 0 1 ~ 1 2 5 に少なくとも 1 つのエラーを有することに一致する。

【 0 0 6 3 】

一態様では、個体の配列ベースのクローン型プロファイルが、次のステップを用いて得られる : (a) 個体の T 細胞および B 細胞の少なくとも一方から核酸サンプルを得るステップ ; (b) このような核酸サンプルから得られる個々の分子を空間的に分離するステップであって、個々の分子が、サンプル中の核酸から作製された少なくとも 1 つの鋳型を含み、この鋳型が、体細胞的に再配列された領域またはその一部を含み、各個々の分子が、少なくとも 1 つの配列リードを形成することができる、ステップ ; (c) 前記空間的に分離された個々の分子をシーケンシングするステップ ; および (d) 核酸サンプルからの核酸分子の異なる配列の存在量を決定してクローン型プロファイルを作成するステップ。一部の実施形態では、この方法は、T 細胞受容体遺伝子または免疫グロブリン遺伝子の組み換え核酸を含む個々の分子を増幅する 1 つ以上のステップをさらに含む。例えば、このような 1 つ以上の増幅ステップは、多段階 P C R を含み得る。別の実施形態では、シーク

エンシングするステップは、空間的に分離された個々の分子のそれぞれを二方向シーケンシングして、少なくとも1つの順方向配列リードおよび少なくとも1つの逆方向配列リードを形成するステップを含む。別の実施形態では、上記の方法は、次のステップを含む：(a) 個体のT細胞およびB細胞の少なくとも一方から核酸サンプルを得るステップ；(b) このような核酸サンプルから得られる個々の分子を空間的に分離するステップであって、個々の分子が、それぞれサンプル中の核酸から作製された、それぞれ体細胞的に再配列された領域またはその一部を有する鋳型のネステッドセットを含み、各ネステッドセットが、それぞれ同じ方向に伸長した、それぞれネステッドセットが形成された核酸の異なる位置から開始する複数の配列リードを形成することができる、ステップ；(c) 前記空間的に分離された個々の分子をシーケンシングするステップ；および(d) 核酸サンプルからの核酸分子の異なる配列の存在量を決定してクローン型プロファイルを作成するステップ。一実施形態では、シーケンシングするステップは、ネステッドセットのそれぞれに対して複数の配列リードを形成するステップを含む。別の実施形態では、体細胞的に再配列された領域はそれぞれ、V領域およびJ領域を含み、複数の配列リードはそれぞれ、V領域の異なる位置から開始し、かつその関連するJ領域の方向に伸長する。

10

20

30

40

50

【0064】

一態様では、個体からの各サンプルに対して、本発明の方法に使用されるシーケンシング技術は、1回に最も少ない1000のクローン型の配列を形成し；別の態様では、このような技術は、1回に少なくとも10,000のクローン型の配列を形成し；別の態様では、このような技術は、1回に少なくとも100,000のクローン型の配列を形成し；別の態様では、このような技術は、1回に少なくとも500,000のクローン型の配列を形成し；別の態様では、このような技術は、1回に少なくとも1,000,000のクローン型の配列を形成する。なお別の態様では、このような技術は、個々のサンプル当たり、1回に100,000～1,000,000のクローン型の配列を形成する。

【0065】

提供される発明の方法に使用されるシーケンシング技術は、1リードに付き約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、または約600bpを形成することができる。

【0066】

本発明は、いくつかの特定の実施形態の例を参照して説明されたが、当業者であれば、本発明の概念および範囲から逸脱することなく本発明の様々な変更が可能であることを理解されよう。本発明は、上記説明されたものに加えて、様々なセンサの実施および他の主題に適用可能である。

【0067】

定義

本明細書に特段の記載がない限り、本明細書で使用される核酸化学、生化学、遺伝学、および分子生物学の用語および記号は、その分野の標準的な論文または教科書の意味である。例えば、コーンバーグ(Kornberg)およびベイカー(Baker)著、DNA複製(DNA Replication)、第2版(W.H.フリーマン(W.H. Freeman)、ニューヨーク、1992年)；レインガー(Lehninger)著、バイオケミストリー(Biochemistry)、第2版(ワース・パブリシャー(Worth Publishers)、ニューヨーク、1975年)；ストラカン(Strachan)およびリード(Read)著、ヒューマン・モレキュラー・ジェネティクス(Human Molecular Genetics)、第2版(ワイリー・リス(Wiley-Liss)、ニューヨーク、1999年)；アバス(Abbas)ら著、セルラー・アンド・モレキュラー・イムノロジー(Cellular and Molecular Immunology)、第6版(サウンダーズ(Saunders)、2007年)。

【0068】

2つ以上の配列、例えば、配列リードについての「アライニング」とは、2つ以上の配列を比較して、ある配列の距離の測定に基づいてこれらの配列がどの程度類似しているかを決定する方法のことである。核酸配列をアライニングする例示的な方法は、スミス・ウォーターマン (Smith Waterman) アルゴリズムである。距離の測定は、ハミング距離またはレーベンシュタイン距離などを含み得る。距離の測定は、比較される配列のヌクレオチドの品質値に関連した要素を含み得る。

【0069】

「アンプリコン」とは、ポリヌクレオチド増幅反応の産物；即ち、一本鎖または二本鎖であり得、かつ1つ以上の開始配列から複製されるポリヌクレオチドのクローン集団のことである。1つ以上の開始配列は、同じ配列の1つ以上のコピーでも良いし、または異なる配列の混合物でも良い。好ましくは、アンプリコンは、1つの開始配列の増幅によって形成される。アンプリコンは、1つ以上の開始核酸または標的核酸の複製を含む産物が作製される様々な増幅反応によって作製することができる。一態様では、アンプリコンを作製する増幅反応は、ヌクレオチドおよび/またはオリゴヌクレオチドである反応物の塩基対が、反応産物の生成に必要な鋳型ポリヌクレオチドに相補体を有するという点で「鋳型主導性」である。一態様では、鋳型主導性反応は、核酸ポリメラーゼを用いるプライマーの伸長、または核酸リガーゼを用いるオリゴヌクレオチドのライゲーションである。このような反応としては、限定されるものではないが、ポリメラーゼ連鎖反応 (PCR: polymerase chain reaction)、線状ポリメラーゼ反応、核酸配列ベース増幅 (NASBA: nucleic acid sequence-based amplification)、およびローリングサークル増幅などが挙げられ、これらは、参照により本明細書に組み入れられる次の参考文献に開示されている：マリス (Mullis) らによる米国特許第4,683,195号明細書；同第4,965,188号明細書；同第4,683,202号明細書；同第4,800,159号明細書 (PCR)；ゲルファント (Gelfand) らによる米国特許第5,210,015号明細書 (「taqman」プローブを用いるリアルタイムPCR)；ウィットワース (Wittwer) らによる米国特許第6,174,670号明細書；カシャン (Kacian) らによる米国特許第5,399,491号明細書 (「NASBA」)；リザルディ (Lizardi) による米国特許第5,854,033号明細書；およびアオノ (Aono) らによる特開平4-262799号公報 (ローリングサークル増幅) など。一態様では、本発明のアンプリコンは、PCRによって作製される。増幅反応は、反応産物の増幅反応の進行としての測定を可能にする検出化学が利用可能である場合は「リアルタイム」増幅とすることができる。例えば、以下に記載される「リアルタイムPCR」、またはレオン (Leone) ら著、ヌクレイック・アシッド・リサーチ (Nucleic Acids Research)、1998年、第26巻、p. 2150~2155などに記載される「リアルタイムNASBA」。本明細書で使用される「増幅する」という語は、増幅反応を行うことを意味する。「反応混合物」とは、限定されるものではないが、反応中にpHを選択されたレベルに維持する緩衝剤、塩、補助因子、およびスカベンジャーなどを含み得る、反応を行うために必要な全ての反応物を含む溶液のことである。

【0070】

本明細書で使用される「クローン性」という語は、レパートリのクローン型の中のクローン型の存在量の分布が1つまたは少数のクローン型に偏っている程度を表す指標のことである。大雑把に言えば、クローン性は、クローン型多様性の逆の指標である。多数の指標または統計を、本発明に従ってクローン性指標として使用することができる種と存在量の関係を説明する生態学から借用することができる。例えば、ピール (Pielou) 著、数理生態学入門 (An Introduction to Mathematical Ecology) の第17章および第18章、(ワイリー・インターサイエンス (Wiley-Interscience)、1969年)。一態様では、本発明に使用されるクローン性指標は、クローン型プロファイル (即ち、検出された異なるクローン型の数お

10

20

30

40

50

よびその存在量)の関数であるため、クローン型プロファイルが測定された後に、クローン性をこの測定値から計算して単数を得ることができる。1つのクローン性指標は、シンプソンの指数(Simpson's measure)であり、この指数は単純に、無作為に選択されたクローン型が同じである確率である。他のクローン性指標としては、情報をベースとする指標およびマッキントッシュの多様性指数が挙げられ、ピール(Pielou)(前掲)に開示されている。

【0071】

「クローン型」とは、免疫受容体またはその一部をコードするリンパ球の組換えヌクレオチド配列のことである。より具体的には、クローン型とは、T細胞受容体(TCR)またはB細胞受容体(BCR)、またはその一部をコードするT細胞またはB細胞の組換えヌクレオチド配列のことである。様々な実施形態では、クローン型は、IGHのVDJ再配列、IGHのDJ再配列、IGKのVJ再配列、IGLのVJ再配列、TCR β のVDJ再配列、TCR β のDJ再配列、TCR α のVJ再配列、TCR γ のVJ再配列、TCR δ のVDJ再配列、TCR δ のVD再配列、またはKde-V再配列などの全てまたは一部をコードし得る。クローン型は、免疫受容体遺伝子、例えば、Bcl1-IGHまたはBcl1-IGHに關与する転座切断点領域もコードし得る。一態様では、クローン型は、クローン型が由来する免疫分子の多様性を表す、または反映するのに十分な長い配列を有し；結果として、クローン型は、様々な長さであり得る。一部の実施形態では、クローン型は、25~400ヌクレオチドの範囲の長さを有し；他の実施形態では、クローン型は、25~200ヌクレオチドの範囲の長さを有する。

【0072】

「クローン型プロファイル」とは、リンパ球の集団に由来する異なるクローン型のリストおよびその相対存在量のことである。典型的には、リンパ球の集団は、組織サンプルから得られる。「クローン型プロファイル」という語は、文献、例えば、次の文献：アスティラ(Arstilla)ら著、サイエンス(Science)、1999年、第286巻、p.958~961；ヤッサイ(Yassai)ら著、免疫遺伝学(Immunogenetics)、2009年、第61巻、p.493~502；ケンジェルスカ(Kedzierska)ら著、モレキュラー・イムノロジー(Mol. Immunol)、2008年、第45巻、第3号、p.607~618などに記載されている免疫「レポートリ」の免疫学的概念に関連するが、この概念よりも一般的である。「クローン型プロファイル」という語は、再配列免疫受容体をコードする核酸の様々なリストおよび存在量を含み、この核酸は、選択されたリンパ球のサブセット(例えば、組織浸潤性リンパ球または免疫表現性サブセットなど)に由来し得る、または完全な免疫受容体と比較すると多様性が低減された免疫受容体の一部をコードし得る。一部の実施形態では、クローン型プロファイルは、少なくとも 10^3 の異なるクローン型を含み得；他の実施形態では、クローン型プロファイルは、少なくとも 10^4 の異なるクローン型を含み得；他の実施形態では、クローン型プロファイルは、少なくとも 10^5 の異なるクローン型を含み得；クローン型プロファイルは、少なくとも 10^6 の異なるクローン型を含み得る。このような実施形態では、このようなクローン型プロファイルは、異なるクローン型のそれぞれの存在量または相対頻度をさらに含み得る。一態様では、クローン型プロファイルは、個体のリンパ球の集団中のT細胞受容体(TCR)またはB細胞受容体(BCR)、またはその断片をそれぞれコードする異なる組換えヌクレオチド配列(その存在量を含む)のセットであり、このセットのヌクレオチド配列は、実質的に全てのリンパ球の集団の異なるリンパ球またはそれらのクローン亜集団と1対1の対応を有する。一態様では、クローン型を決定する核酸セグメントは、それらの多様性(すなわち、セット中の異なる核酸セグメントの数)が、個体における実質的に全てのT細胞またはB細胞、またはそのクローンが、このようなレポートリのユニークな核酸配列を有する十分な長さであるように選択される。即ち、好ましくは、サンプルの各異なるクローンは、異なるクローン型を有する。本発明の他の態様では、レポートリに対応するリンパ球の集団は、循環B細胞であり得る、循環T細胞であり得る、または、限定されるものではないが、CD4+T細胞もしくはCD

8 + T細胞を含むいずれかの上記の集団の亜集団、または細胞表面マーカーなどによって定義される他の亜集団であり得る。このような亜集団は、特定の組織、例えば、骨髄もしくはリンパ節などからサンプルを採取することによって、または1つ以上の細胞の表面マーカー、サイズ、もしくは形態などに基づいてサンプル（例えば、末梢血）から細胞を選別もしくは濃縮することによって得ることができる。なお他の態様では、レパトリに対応するリンパ球の集団は、病変組織、例えば、腫瘍組織または感染組織などから得ることができる。一実施形態では、ヒトTCRβ鎖またはその断片を含むクローン型プロファイルは、 $0.1 \times 10^6 \sim 1.8 \times 10^6$ の範囲、 $0.5 \times 10^6 \sim 1.5 \times 10^6$ の範囲、または $0.8 \times 10^6 \sim 1.2 \times 10^6$ の範囲の数の異なるヌクレオチド配列を含む。別の実施形態では、ヒトIGH鎖またはその断片を含むクローン型プロファイルは、 $0.1 \times 10^6 \sim 1.8 \times 10^6$ の範囲、 $0.5 \times 10^6 \sim 1.5 \times 10^6$ の範囲、または $0.8 \times 10^6 \sim 1.2 \times 10^6$ の範囲の数の異なるヌクレオチド配列を含む。特定の実施形態では、本発明のクローン型プロファイルは、IGH鎖のV(D)J領域の実質的に全てのセグメントをコードするヌクレオチド配列のセットを含む。一態様では、本明細書で使用する「実質的に全ての」とは、0.001%以上の相対存在量を有する全てのセグメントのことである；または、別の態様では、本明細書で使用する「実質的に全ての」とは、0.0001%以上の相対存在量を有する全てのセグメントのことである。別の特定の

実施形態では、本発明のクローン型プロファイルは、TCRβ鎖のV(D)J領域の実質的に全てのセグメントをコードするヌクレオチド配列のセットを含む。別の実施形態では、本発明のクローン型プロファイルは、25～200ヌクレオチドの範囲の長さを有し、かつTCRβ鎖のV領域、D領域、および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、本発明のクローン型プロファイルは、25～200ヌクレオチドの範囲の長さを有し、かつIGH鎖のV領域、D領域、および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、本発明のクローン型プロファイルは、異なるIGH鎖を発現するリンパ球の数に実質的に等しい数の異なるヌクレオチド配列を含む。別の実施形態では、本発明のクローン型プロファイルは、異なるTCRβ鎖を発現するリンパ球の数に実質的に等しい数の異なるヌクレオチド配列を含む。なお別の実施形態では、「実質的に等しい」とは、99%の確率で、クローン型プロファイルが、0.001%以上の頻度で個体のリンパ球の全ての集団によって保持または発現されるIGHまたはTCRβ、またはその一部をコードするヌクレオチド配列を含むこと

である。なお別の実施形態では、「実質的に等しい」とは、99%の確率で、ヌクレオチド配列のレパトリが、0.0001%以上の頻度で存在する全てのリンパ球によって保持または発現されるIGHまたはTCRβ、またはその一部をコードするヌクレオチド配列を含むことである。一部の実施形態では、クローン型プロファイルは、 $10^5 \sim 10^7$ のリンパ球を含むサンプルから得られる。このような数のリンパ球は、1～10mLの末梢血サンプルから得ることができる。

【0073】

「同一にする」とは、配列の差異が、実験誤差または測定誤差によるものであって真の生物学的な差異によるものではないと決定することによって、配列に差異がある2つの候補クローン型を同一として取り扱うことである。一態様では、高頻度の候補クローン型の配列が、低頻度の候補クローン型の配列と比較されて、所定の基準が満たされた場合は、低頻度の候補クローン型の数が、高頻度の候補クローン型の数に加えられ、この後、この低頻度の候補クローン型は無視される。即ち、低頻度候補クローン型に関連したリード数が、高頻度候補クローン型のリード数に加えられる。

【0074】

「相補性決定領域」(CDR: complementarity determining region)とは、分子が抗原の高次構造を補完し、これにより分子の特定の抗原との特異性および接触を決定する免疫グロブリン（即ち、抗体）またはT細胞受容体の領域のことである。T細胞受容体および免疫グロブリンはそれぞれ、3つのCDRを有し、CDR1およびCDR2は、可変(V)ドメインに見られ、CDR3は、Vドメインの

一部、多様性（D）ドメイン（重鎖のみ）および接合（J）ドメインの全て、および定常（C）ドメインの一部を含む。

【0075】

本明細書で使用される「汚染」とは、ある時点で一個体から採取されたサンプルに、別の個体からの核酸、または別の時点で採取された同じ個体のサンプルからの核酸が存在することである。一態様では、「汚染」とは、本発明のクローン型プロファイルの解釈に影響を与え得る所与の患者の組織サンプルに由来しない核酸が存在することである。

【0076】

「内部標準」とは、サンプル中の標的ポリヌクレオチドの絶対定量または相対定量を可能にするために、1つ以上の標的ポリヌクレオチドと同じ反応で処理される核酸配列のことである。一態様では、反応は、増幅反応、例えば、PCRである。内部標準は、内因性または外因性であり得る。即ち、内部標準は、サンプル中で自然に発生するものでも良いし、または反応の前にサンプルに加えられるものでも良い。一態様では、1つ以上の外因性内部標準配列を、所定の濃度で反応混合物に加えて、増幅された配列を比較して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校正を可能にする。外因性内部標準の数、配列、長さ、および他の特性の選択は、当業者にはルーチンの設計上の選択である。外因性内部標準は、本明細書では「基準配列」とも呼ばれ、一定の細胞サイクル依存レベルの転写を示す僅かに調整された遺伝子に対応するサンプルにとって自然な配列である。例えば、セルベイ（Selvey）ら著、モレキュラー・アンド・セルラー・プローブス（Mol. Cell Probes）、2001年、第15巻、p. 307～311。例示的な内部標準としては、限定されるものではないが、次の遺伝子：GAPDH、 β_2 -ミクログロブリン、18SリボソームRNA、および β -アクチンの配列が挙げられる。

10

20

【0077】

「キット」とは、本発明の方法を行うために材料または試薬を送達するための任意の送達システムのことである。本発明の方法との関連では、このような送達システムは、反応試薬（例えば、適切な容器に入れられたプライマー、酵素、内部標準など）および/または支援材料（例えば、緩衝剤、アッセイなどを行うための取扱説明書）の保存、ある位置から別の位置への移送または送達を可能にするシステムを含む。例えば、キットは、関連する反応試薬および/または支援材料を含む1つ以上の容器（例えば、箱）を含む。このような内容物は、目的のレシピエントと一緒にまたは別個に送達することができる。例えば、第1の容器は、アッセイに使用される酵素を含み得、第2の容器はプライマーを含む。

30

【0078】

「リンパ系腫瘍」とは、悪性または非悪性であり得るリンパ球の異常増殖のことである。リンパ癌は、悪性リンパ系腫瘍である。リンパ系腫瘍は、限定されるものではないが、濾胞性リンパ腫、慢性リンパ性白血病（CLL：chronic lymphocytic leukemia）、急性リンパ性白血病（ALL：acute lymphocytic leukemia）、有毛細胞白血病、リンパ腫、多発性骨髄腫、移植後リンパ増殖異常、マントル細胞リンパ腫（MCL：mantle cell lymphoma）、びまん性大細胞型リンパ腫（DLBCL：diffuse large B cell lymphoma）、またはT細胞リンパ腫などを含むリンパ増殖異常の結果であるか、またはこのようなリンパ増殖異常に関連する。例えば、ジャフィ（Jaffe）ら著、ブラッド（Blood）、2008年、第112巻、p. 4384～4399；スウェドレー（Swerdlow）ら著、造血組織およびリンパ系組織の腫瘍のWHO分類（WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues）（e. 第4版）（IARCプレス（IARC Press）、2008年）。

40

【0079】

「微小残留病変」とは、治療後に残る癌細胞のことである。この語は、リンパ腫および

50

白血病の治療に関連して最も頻繁に使用される。

基準配列と別の配列（「比較配列」）の比較に関連して使用される「パーセント相同性」または「パーセント同一」などの語は、2つの配列間の最適なアラインメントでは、比較配列が、示されるパーセンテージに等しい数のサブユニット位置で基準配列と同一であることを意味し、このサブユニットは、ポリヌクレオチドの比較のためのヌクレオチドまたはポリペプチドの比較のためのアミノ酸である。本明細書で使用される、比較される配列の「最適なアラインメント」とは、サブユニット間の一致を最大にし、かつアラインメントの構築に利用されるギャップを最小にするアラインメントのことである。パーセント同一性は、アルゴリズム、例えば、ニードルマン（Needleman）およびブンシュ（Wunsch）著、ジャーナル・オブ・モレキュラー・バイオロジー（J. Mol. Biol.）、1970年、第48巻、p. 443～453（ウイスコンシン配列分析パッケージの「GAP」プログラム（“GAP” program of Wisconsin Sequence Analysis Package）、ジェネティクス・コンピュータ・グループ（Genetics Computer Group）、マディソン、ウイスコンシン州）などに記載されている市販のアルゴリズムの実施で決定することができる。アラインメントの構築、およびパーセンテージ同一性または他の類似性の指標の計算に用いられる当分野の他のソフトウェアパッケージは、スミス（Smith）およびウォーターマン（Waterman）著、アドバンシス・イン・アプライド・マセマティクス（Advances in Applied Mathematics）、1981年、第2巻、p. 482～489（ウイスコンシン配列分析パッケージ（Wisconsin Sequence Analysis Package）、ジェネティクス・コンピュータ・グループ（Genetics Computer Group）、マディソン、ウイスコンシン州）のアルゴリズムに基づいた、「BestFit」プログラムを含む。言い換えれば、例えば、基準ヌクレオチド配列に少なくとも95%同一であるヌクレオチド配列を有するポリヌクレオチドを得るために、基準配列の最大5%のヌクレオチドを欠失させる、もしくは別のヌクレオチドで置換する、または基準配列のヌクレオチドの総数の最大5%の数のヌクレオチドを基準配列に挿入することができる。

【0080】

「ポリメラーゼ連鎖反応」または「PCR」とは、DNAの相補鎖の同時プライマー伸長による特定のDNA配列のin vitro増幅反応のことである。言い換えれば、PCRは、プライマー結合部位に隣接した標的核酸の複数のコピーまたは複製を作製する反応であり、このような反応は、次のステップの1回以上の反復を含む：（i）標的核酸を変性させるステップ、（ii）プライマーをプライマー結合部位にアニーリングするステップ、および（iii）ヌクレオシド三リン酸の存在下でプライマーを核酸ポリメラーゼによって伸長させるステップ。通常は、この反応は、サーマル・サイクラー装置で各ステップに最適な異なる温度で繰り返される。各ステップでの特定の温度、期間、ステップ間の変化の割合は、例えば、参考文献：マクファーソン（McPherson）ら編、PCR：実践的なアプローチ（PCR：A Practical Approach）およびPCR2：実践的なアプローチ（PCR2：A Practical Approach）（IRLプレス（IRL Press）、オックスフォード、それぞれ1991年および1995年）によって例示されている、当業者に周知の多数の因子によって決まる。例えば、Taq DNAポリメラーゼを使用する従来のPCRでは、二本鎖標的核酸を、90を超える温度で変性させ、プライマーを50～75の温度でアニーリングし、そしてプライマーを72～78の温度で伸長させることができる。「PCR」という語は、限定されるものではないが、RT-PCR、リアルタイムPCR、ネステッドPCR、定量PCR、およびマルチプレックスPCRなどを含む反応の派生形を包含する。利用される特定の形式のPCRは、本出願との関連から当業者には認識されよう。反応量は、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」または「RT-PCR」とは、標的RNAを相補的な一本鎖DNAに変換する逆転写反応が先に起こり、次いで、この一本鎖DNAが増幅されるPCRのことである

10

20

30

40

50

。例えば、参照により本明細書に組み入れられるテコット (T e c o t t) らによる米国特許第 5, 1 6 8, 0 3 8 号明細書。「リアルタイム P C R」とは、反応産物、即ち、アンプリコンの量が、反応が進行するときに監視される P C R のことである。様々な形式のリアルタイム P C R が存在し、これらのリアルタイム P C R は主に、反応産物の監視に使用される検出化学が異なる。例えば、参照により本明細書に組み入れられる、ゲルフアント (G e l f a n d) らによる米国特許第 5, 2 1 0, 0 1 5 号明細書 (「t a q m a n」); ウィットワース (W i t t w e r) らによる米国特許第 6, 1 7 4, 6 7 0 号明細書および同第 6, 5 6 9, 6 2 7 号明細書 (挿入染料 (i n t e r c a l a t i n g d y e s)); チャギ (T y a g i) らによる米国特許第 5, 9 2 5, 5 1 7 号明細書 (分子ビーコン (m o l e c u l a r b e a c o n s))。リアルタイム P C R の検出化学は、同様に参照により本明細書に組み入れられるマックイ (M a c k a y) ら著、ヌクレイック・アシッド・リサーチ (N u c l e i c A c i d s R e s e a r c h)、2 0 0 2 年、第 3 0 巻、p. 1 2 9 2 ~ 1 3 0 5 に概説されている。「ネステッド P C R」とは、2 段階 P C R のことであり、第 1 の P C R のアンプリコンが、新たなセットのプライマーを用いる第 2 の P C R のサンプルとなり、このプライマーの少なくとも 1 つが、第 1 のアンプリコンの内部の位置に結合する。ネステッド増幅反応に関連して本明細書で使用される「一次プライマー」とは、第 1 のアンプリコンを作製するために使用されるプライマーのことであり、「二次プライマー」とは、第 2 のアンプリコンまたはネステッドアンプリコンを作製するために使用される 1 つ以上のプライマーのことである。「マルチプレックス P C R」とは、複数の標的配列 (または 1 つの標的配列と 1 つ以上の基準配列) が、同じ反応混合物内で同時に増幅される P C R のことである。例えば、バーナード (B e r n a r d) ら著、アナリティカル・バイオケミストリー (A n a l . B i o c h e m.)、1 9 9 9 年、第 2 7 3 巻、p. 2 2 1 ~ 2 2 8 (2 色リアルタイム P C R (t w o - c o l o r r e a l - t i m e P C R))。通常は、プライマーの異なるセットが、増幅される各配列に利用される。典型的には、マルチプレックス P C R の標的配列の数は、2 ~ 5 0、2 ~ 4 0、または 2 ~ 3 0 の範囲である。「定量 P C R」とは、サンプルまたは標本中の 1 つ以上の特定の標的配列の存在量を測定するように設計された P C R のことである。定量 P C R は、このような標的配列の絶対定量および相対定量の両方を含む。定量測定は、標的配列と別個にまたは一緒にアッセイされ得る 1 つ以上の基準配列または内部標準を用いて行われる。基準配列は、サンプルまたは標本に対して内因性または外因性であり得、後者の場合は、1 つ以上の競合誘型を含み得る。典型的な内因性基準配列は、次の遺伝子の転写物のセグメントを含む: β - アクチン、G A P D H、 β_2 - ミクログロブリン、およびリボソーム R N A など。定量 P C R の技術は、当分野で周知であり、参照により本明細書に組み入れられる次の文献に例示されている: フリーマン (F r e e m a n) ら著、バイオテクニクス (B i o t e c h n i q u e s)、1 9 9 9 年、第 2 6 巻、p. 1 1 2 ~ 1 2 6; ベッカー - アンドレ (B e c k e r - A n d r e) ら著、ヌクレイック・アシッド・リサーチ (N u c l e i c A c i d s R e s e a r c h)、1 9 8 9 年、第 1 7 巻、p. 9 4 3 7 ~ 9 4 4 7; ジーマン (Z i m m e r m a n) ら著、バイオテクニクス (B i o t e c h n i q u e s)、1 9 9 6 年、第 2 1 巻、p. 2 6 8 ~ 2 7 9; ディビアッコ (D i v i a c c o) ら著、ジーン (G e n e)、1 9 9 2 年、第 1 2 2 巻、p. 3 0 1 3 ~ 3 0 2 0; ベッカー - アンドレ (B e c k e r - A n d r e) ら著、ヌクレイック・アシッド・リサーチ (N u c l e i c A c i d s R e s e a r c h)、1 9 8 9 年、第 1 7 巻、p. 9 4 3 7 ~ 9 4 4 6 など。

【0081】

「ポリヌクレオチド」は、ヌクレオチド単量体の線状ポリマーを指し、D N A または R N A であり得る。ポリヌクレオチドを構成する単量体は、単量体と単量体との相互作用の規定パターン、例えば、ワトソン - クリック型の塩基対合、塩基の積み重ね、またはフーグスティーン型もしくは逆フーグスティーン型の塩基対合などによって天然ポリヌクレオチドに特異的に結合することができる。このような単量体およびそのヌクレオチド間結合は、自然発生し得る、またはそれらの類似体、例えば、天然類似体もしくは非天然類似体

であり得る。非天然類似体は、PNA、ホスホロチオエートヌクレオシド間結合、および標識、例えば、フルオロフォアまたはハプテンの付着を可能にする連結基を含む塩基などを含み得る。ポリヌクレオチドは、ホスホジエステル結合によって連結された4つの天然ヌクレオシド（例えば、DNAのデオキシアデノシン、デオキシシチジン、デオキシグアノシン、デオキシチミジン、またはRNAのそのリボース対応物）を含み得る。しかしながら、ポリヌクレオチドは、例えば、修飾塩基、糖、またはヌクレオシド間結合を含む非天然ヌクレオチド類似体も含み得る。酵素が、活性のために特定のオリゴヌクレオチドまたはポリヌクレオチド基質の要件（例えば、一本鎖DNA、またはRNA/DNA二重鎖など）を有することは当業者には明らかであり、オリゴヌクレオチドまたはポリヌクレオチド基質に適切な組成物の選択は、特に、論文、例えば、サンプブルック（Sambrook）ら著、モレキュラー・クローニング（MOLECULAR CLONING）、第2版（コールド・スプリングス・ハーバ研究所（Cold Spring Harbor Laboratory）、ニューヨーク、1989年）および同様の文献の手引きから、十分に当業者の知識の範囲内である。本明細書で使用される「オリゴヌクレオチド」は、例えば、3～60単量体単位、または一部の実施形態では、12～60の単量体単位を有する小さいポリヌクレオチドを指す。様々な実施形態では、ポリヌクレオチドまたはオリゴヌクレオチドは、文字列（大文字または小文字）、例えば、「ATGCCCTG」で表され、ヌクレオチドは、左から右に5'→3'の順であり、特段の記載がなく、文脈から明らかにそうでない場合を除き、「A」はデオキシアデノシンを示し、「C」はデオキシシチジンを示し、「G」はデオキシグアノシンを示し、「T」はチミジンを示し、「I」はデオキシイノシンを示し、「U」はウリジンを示すことが理解されよう。

10

20

30

40

50

【0082】

「プライマー」とは、ポリヌクレオチド鋳型と二重鎖を形成するときに、核酸合成の開始点として機能し、その3'末端から鋳型に沿って伸長して伸長した二重鎖を形成することができる天然または合成のオリゴヌクレオチドのことである。プライマーの伸長は、通常は、核酸ポリメラーゼ、例えば、DNAまたはRNAポリメラーゼで行われる。伸長プロセス中に付加されるヌクレオチドの配列は、鋳型ポリヌクレオチドの配列によって決定される。通常は、プライマーは、DNAポリメラーゼによって伸長される。プライマーは、通常は、14～40ヌクレオチドの範囲、または18～36ヌクレオチド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、例えば、1種類のプライマーを用いる線状増幅反応、または2種類以上のプライマーを用いるポリメラーゼ連鎖反応に利用される。特定の適用例に対してプライマーの長さおよび配列を選択する案内は、参照により本明細書に組み入れられる次の参照文献から分かるように当業者には周知である：ディーフェンバッハ（Dieffenbach）編、PCRプライマー：実習マニュアル（PCR Primer: A Laboratory Manual）、第2版（コールド・スプリング・ハーバ・プレス（Cold Spring Harbor Press）、ニューヨーク、2003年）。

【0083】

「品質スコア」とは、特定の配列位置での塩基の割り当てが正確である確率の指標のことである。特定の状況、例えば、異なるシーケンシング化学、検出システム、およびベースコーリングアルゴリズムなどの結果としてベースコーリングされた塩基に対する品質スコアを計算する様々な方法が当業者に周知である。一般に、品質スコアの値は、正確なベースコーリングの確率に単調に関連する。例えば、10の品質スコアまたはQは、塩基が正確にベースコーリングされる確率が90%であることを意味し、20のQは、塩基が正確にベースコーリングされる確率が99%であることを意味し、以降同様である。一部のシーケンシングプラットフォーム、特に、合成化学によるシーケンシングを用いるシーケンシングプラットフォームでは、平均品質スコアは、配列リードの長さの関数として低下するため、配列リードの最初の品質スコアは、配列リードの末端の品質スコアよりも高く、このような低下は、現象、例えば、不完全な伸長、繰り越し伸長（carry forward extension）、鋳型の減少、ポリメラーゼの減少、キャピ

ングの失敗、および脱保護の失敗などによるものである。

【0084】

「配列リード」とは、シーケンシング技術によって生成される一連のデータまたはデータのストリームから決定されるヌクレオチド配列のことであり、この決定は、例えば、技術に関連したベースコーリングソフトウェア、例えば、DNAシーケンシングプラットフォームの供給業者からのベースコーリングソフトウェアによって行われる。配列リードは、通常は、配列の各ヌクレオチドの品質スコアを含む。典型的には、配列リードは、プライマーを、例えば、DNAポリメラーゼまたはDNAリガーゼを用いて鋳型核酸に沿って伸長させることによって作製される。データは、このような伸長に関連したシグナル、例えば、光シグナル、化学シグナル（例えば、pHの変化）、または電気シグナルを記録することによって生成される。このような初期データが、配列リードに変換される。

10

【0085】

「配列タグ」（もしくは「タグ」）または「バーコード」とは、ポリヌクレオチドまたは鋳型分子に付着されたオリゴヌクレオチドのことであり、1回の反応または一連の反応でポリヌクレオチドまたは鋳型分子を識別し、かつ/または追跡するために使用される。配列タグは、ポリヌクレオチドまたは鋳型の3'末端もしくは5'末端に付着させることができ、または配列タグは、このようなポリヌクレオチドまたは鋳型の内部に挿入して線状コンジュゲートを形成することができ、この線状コンジュゲートは、本明細書では、「タグ付きポリヌクレオチド」、「タグ付き鋳型」、「タグ-ポリヌクレオチドコンジュゲート（タグとポリヌクレオチドのコンジュゲート）」、または「タグ-分子コンジュゲート（タグと分子とのコンジュゲート）」などと呼ばれることもある。配列タグは、サイズおよび組成が非常に多様であり得；参照により本明細書に組み入れられる次の文献が、特定の実施形態に適切な配列タグのセットを選択するための案内を示す：ブレンナー（Brenner）による米国特許第5,635,400号明細書；ブレンナー（Brenner）およびマチェビッツ（Macevitz）による米国特許第7,537,897号明細書；ブレンナー（Brenner）ら著、米国科学アカデミー紀要（Proc. Natl. Acad. Sci.）、2000年、第97巻、p.1665~1670；チャーチ（Church）らによる欧州特許公開第0303459号明細書；シューメーカー（Shoemaker）ら著、ネイチャー・ジェネティクス（Nature Genetics）、1996年、第14巻、p.450~456；モリス（Morris）らによる欧州特許出願公開第0799897A1号明細書；ワレス（Wallace）による米国特許第5,981,179号など。配列タグの長さおよび組成は、非常に多様であり得、特定の長さおよび/または組成の選択は、限定されるものではないが、例えば、ハイブリダイゼーション反応または酵素反応、例えば、シーケンシングによってタグがどのように使用されて読出しが生成されるか；例えば、蛍光染料などで配列タグが標識されているか否か；ポリヌクレオチドのセットなどを一義的に識別するために必要な区別可能なオリゴヌクレオチドタグの数、および、例えば、クロスハイブリダイゼーションまたはシーケンシングエラーによる誤認が存在しない、確実な識別を保障するためにタグのセットがどの程度異ならなければならないかを含め、いくつかの因子によって決まる。一態様では、配列タグはそれぞれ、2~36ヌクレオチド、4~30ヌクレオチド、8~20ヌクレオチド、または6~10ヌクレオチドの範囲内の長さを有し得る。一態様では、配列タグのセットが使用され、各配列タグのセットは、同じセットの他のどのタグのヌクレオチド配列とも少なくとも2つの塩基が異なるユニークなヌクレオチド配列を有し；別の態様では、配列タグのセットが使用され、各配列タグのセットは、同じセットの他のどのタグとも少なくとも3つの塩基が異なる。

20

30

40

【0086】

「配列ツリー」とは、ヌクレオチド配列を表すツリーデータ構造のことである。一態様では、本発明のツリーデータ構造は、循環または循環経路を含まないノードおよびエッジを含む根付き有向木である。本発明のツリーデータ構造のエッジからノードは、通常は順序付けられている。ノードおよび/またはエッジは、値を含み得る構造、または値に関連

50

し得る構造である。木の各ノードは、0 以上の子ノードを有し、子ノードは、慣例により、木の中のノードの下に示されている。子を有するノードは、子の親ノードと呼ばれる。ノードは、多くとも 1 つの親しか有さない。子を一切有していないノードは、葉ノードと呼ばれる。木の最上位ノードは、根ノードと呼ばれる。最上位ノードであると、根ノードは親を有していない。最上位ノードは、木に対する操作が通常始まるノードである（ただし、一部のアルゴリズムは、葉ノードで始まり、根で操作が終了する）。他の全てのノードには、エッジまたはリンクを辿ることによって到達することができる。

【 0 0 8 7 】

配列表

【 0 0 8 8 】

10

【 表 1 】

<110> Sequentia, Inc.
Faham, Malek
Moorhead, Martin
Willis, Thomas

<120> METHOD OF SEQUENCE DETERMINATION USING SEQUENCE TAGS

<130> 838US00 (37623-730.201)

20

<140> 13/835093

<141> 2013-03-15

<150> 13/100365

<151> 2011-05-04

<150> 61/658317

<151> 2012-06-11

30

<150> 61/738277

<151> 2012-12-17

<150> 61/776647

<151> 2013-03-11

<160> 6

<170> PatentIn version 3.5

40

<210> 1
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 1
agttctggct aacctgtaga gcca

10
24

<210> 2
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

20

<400> 2
agttcgggct aacctgtcga gcca

24

<210> 3
<211> 24
<212> DNA
<213> Artificial Sequence

30

<220>
<223> primer

<400> 3
agttccggct aacctgtcga gcca

24

<210> 4
<211> 22
<212> DNA
<213> Artificial Sequence

40

<220>

<223> primer

<220>

<221> misc_feature

<222> (1)..(22)

<223> n is a, c, g, or t

<400> 4

nnnnnnnnnnnn nnnnnnnnnnn nn

22

10

<210> 5

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 5

gtatttttt ct

12

20

<210> 6

<211> 13

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 6

ttcagggggg gct

13

30

【図 1 A】

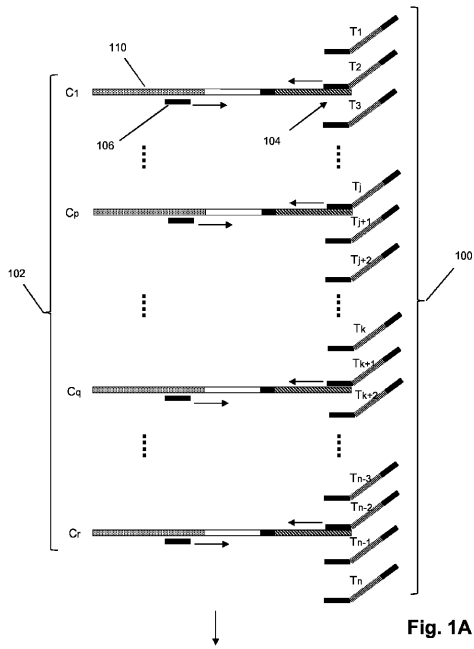


Fig. 1A

【図 3】

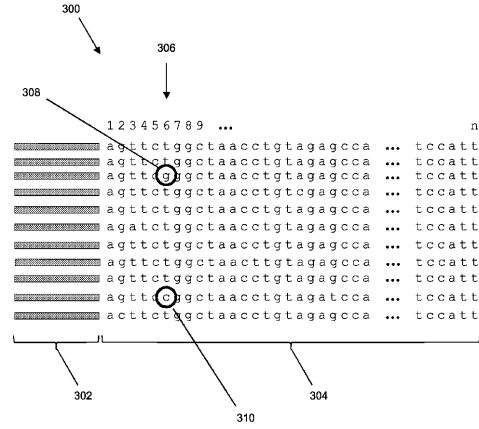
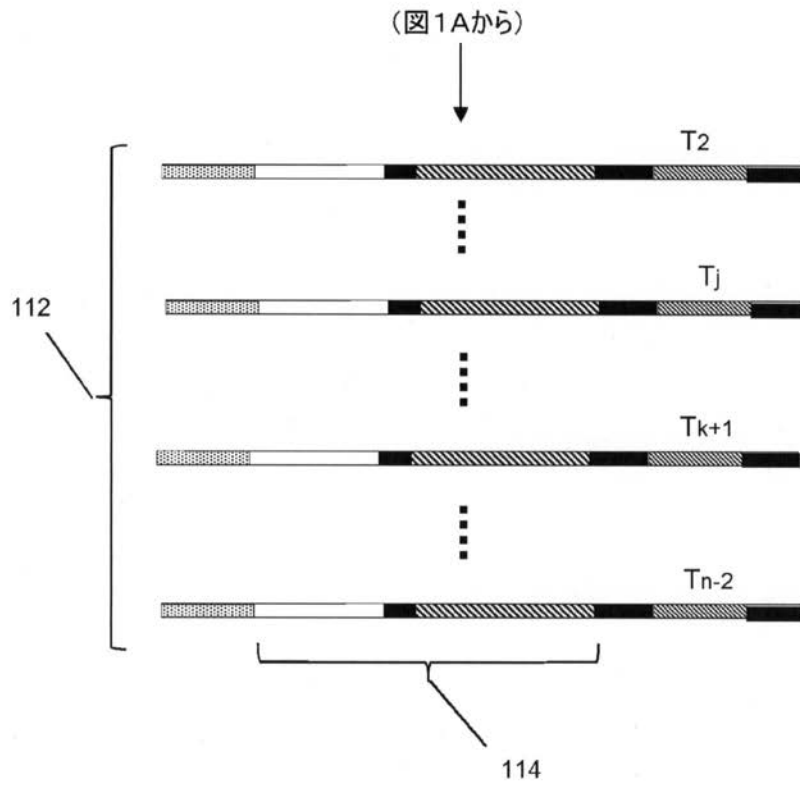
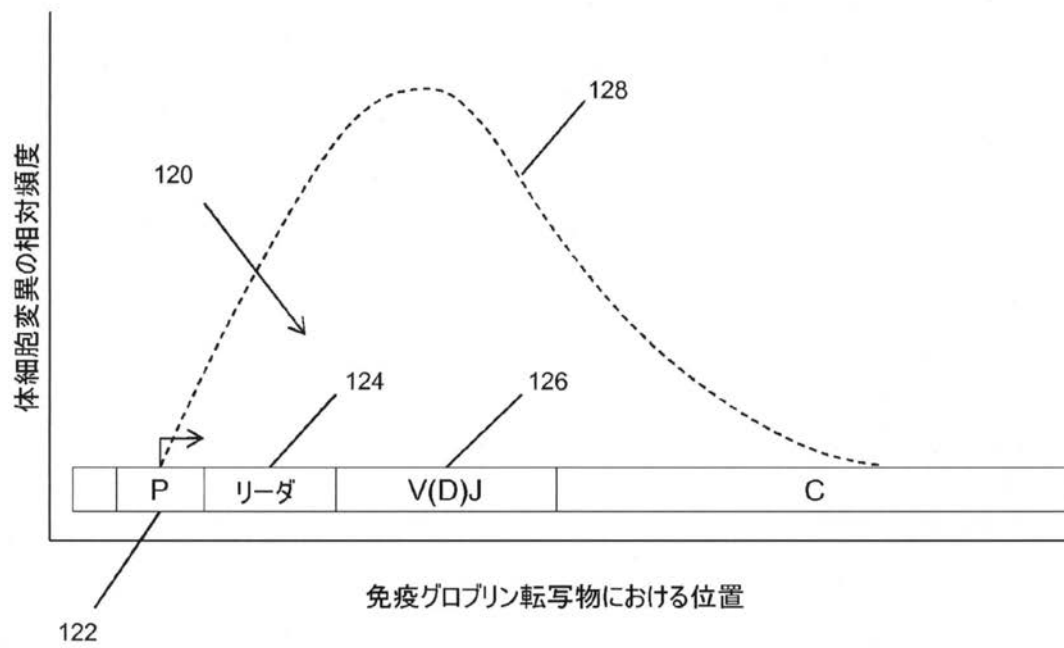


Fig. 3

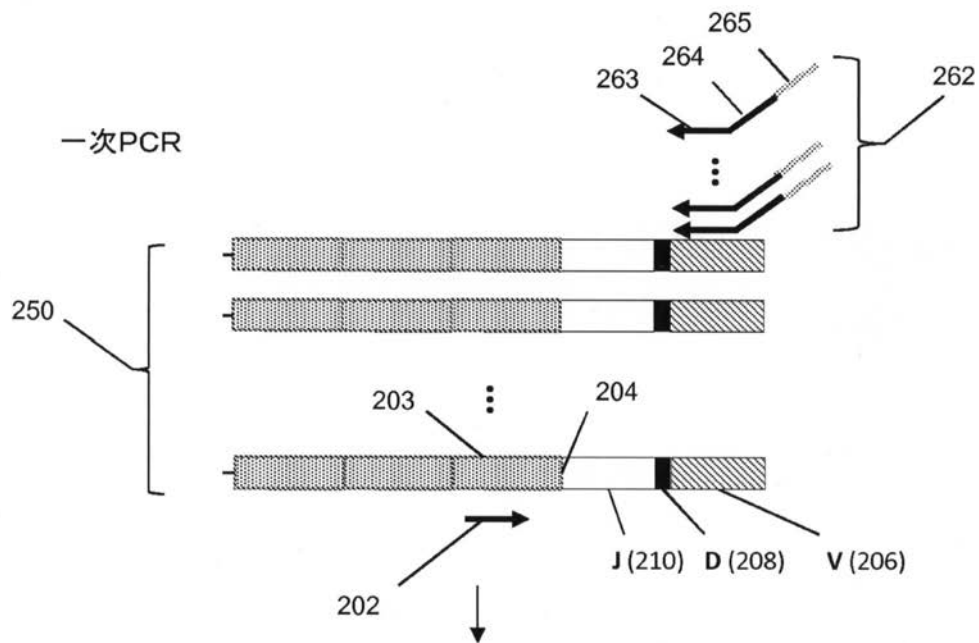
【図 1 B】



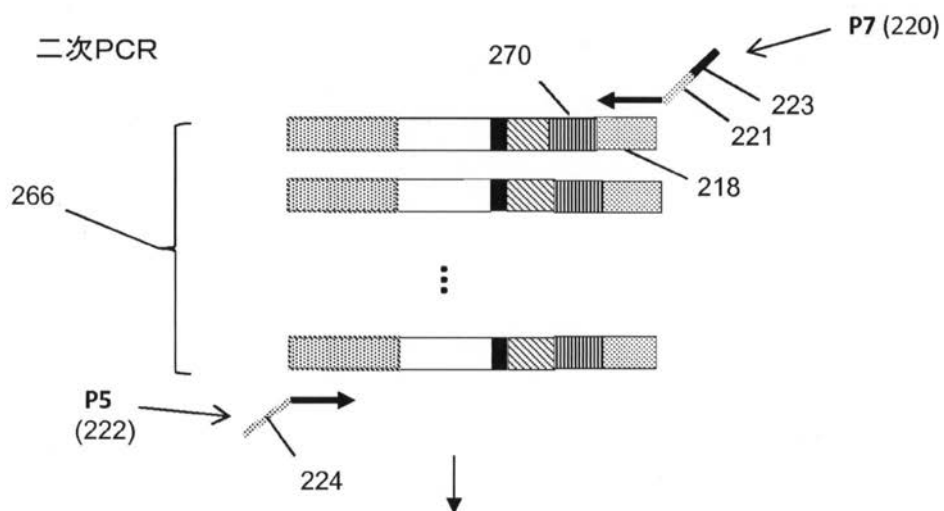
【図 1 C】



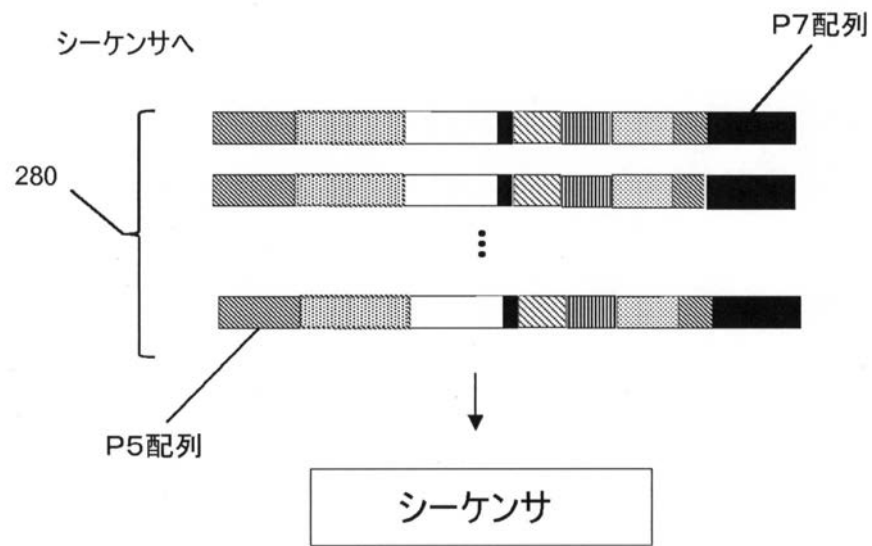
【図 2 A】



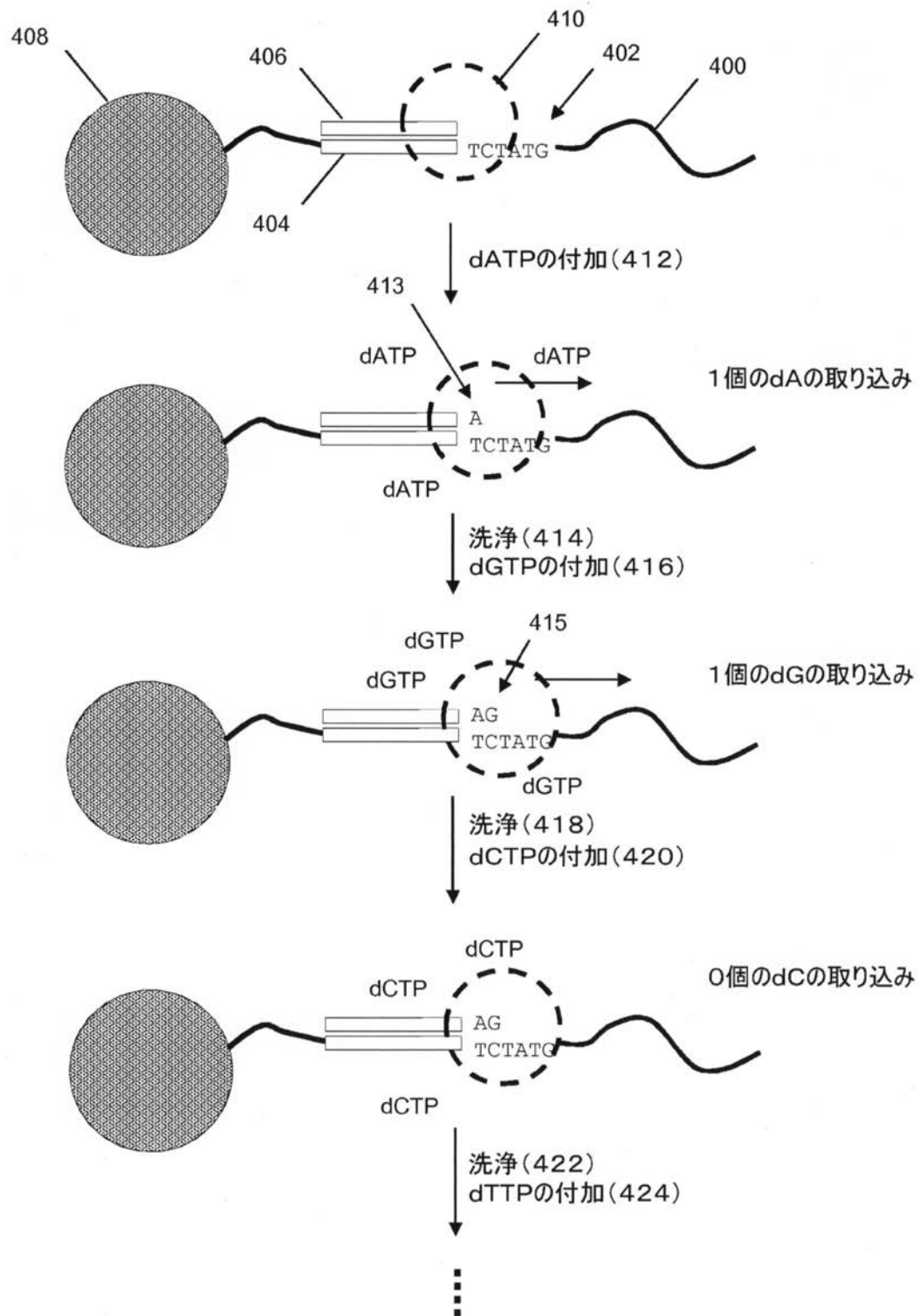
【図 2 B】



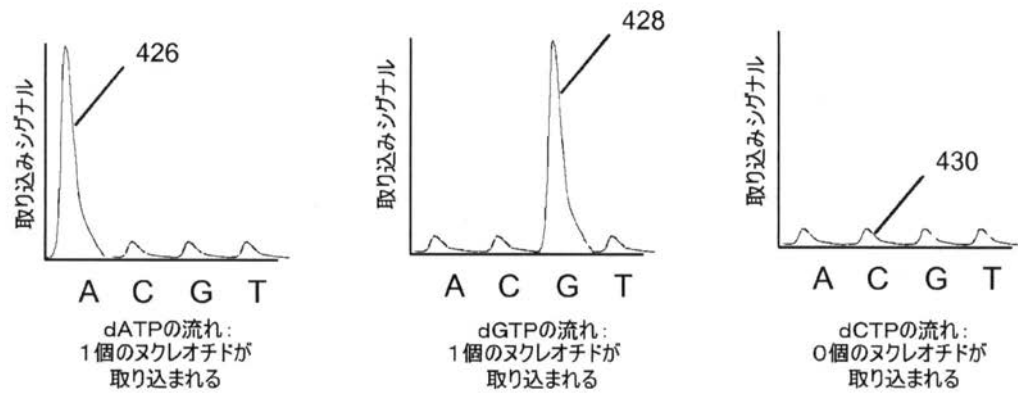
【図 2 C】



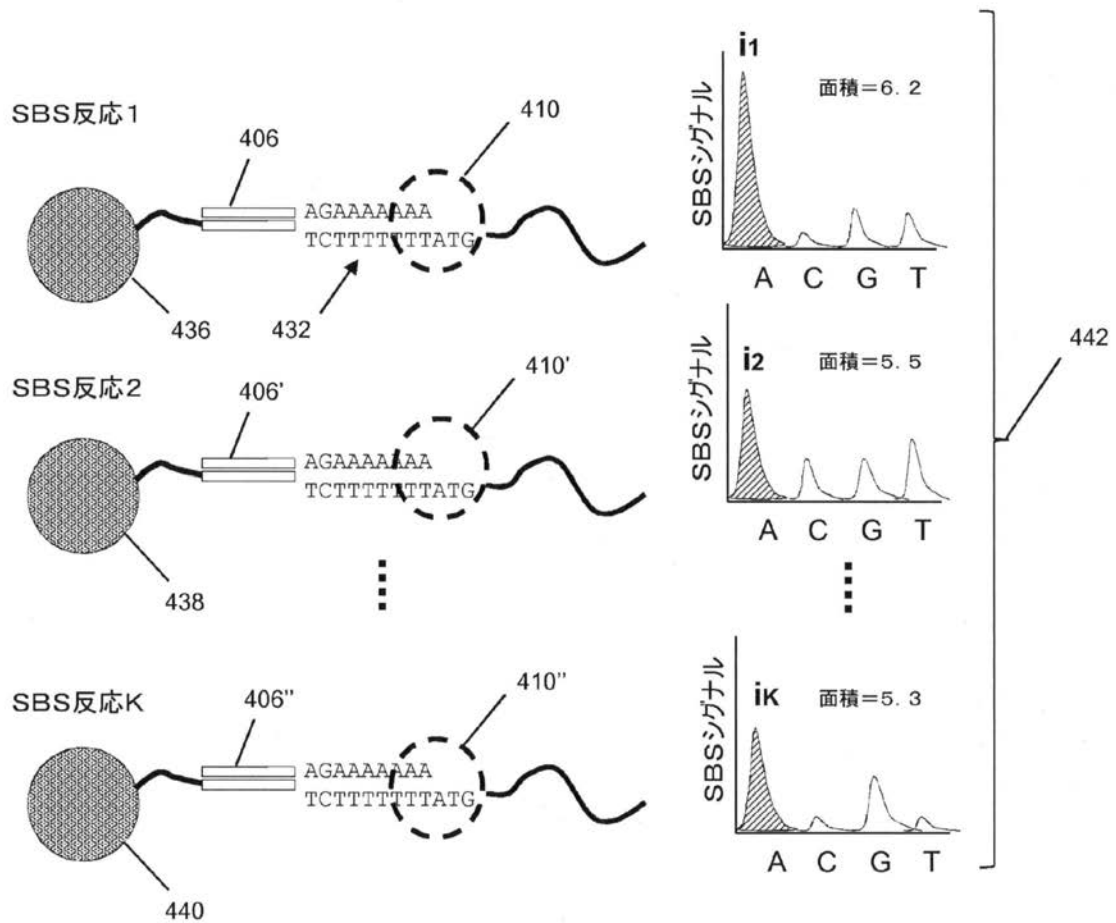
【図 4 A】



【図 4 B】



【図 4 C】



【図 4 D】

鋳型:
TCGGGGGGGACTT

SBS反応		共通の 配列タグ(452)		別個のSBS反応からのdNTPの流れに対する 取り込みシグナル(450)					
				f1a	f1g	f1c	f1t	f2a	f2g ...
1	tct	...	ccct	(1.1)	(1.0)	(6.5)	(0.9)	(1.1)	(2.0) ...
2	tct	...	ccct	(0.9)	(1.1)	(7.3)	(1.9)	(1.0)	(2.3) ...
3	tct	...	ccct	(1.0)	(0.7)	(6.8)	(0.9)	(1.0)	(2.1) ...
4	tct	...	ccct	(1.2)	(0.6)	(6.9)	(0.9)	(1.0)	(1.3) ...
5	tct	...	ccct	(0.8)	(1.2)	(6.8)	(0.9)	(1.2)	(2.0) ...
6	tct	...	ccct	(1.1)	(1.0)	(7.0)	(0.9)	(0.8)	(2.0) ...
7	tct	...	ccct	(1.0)	(0.9)	(6.2)	(1.0)	(0.9)	(1.5) ...
8	tct	...	ccct	(1.0)	(0.7)	(7.2)	(1.1)	(1.0)	(2.0) ...
9	tct	...	ccct	(0.9)	(1.0)	(6.9)	(1.1)	(0.5)	(2.0) ...
.									
.									
.									
K	tct	...	ccct	(1.1)	(1.0)	(7.0)	(1.2)	(1.0)	(1.4) ...

【配列表】

2015519081000001.app

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 13/45276															
A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - C12Q 1/68 (2013.01) USPC - 435/6.11 According to International Patent Classification (IPC) or to both national classification and IPC																	
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC(8): C12Q 1/68 (2013.01) USPC: 435/6.11 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched USPC: 435/6.1, 6.11, 6.12 (text search) Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PatBase, PubWEST, Google Scholar, Google Patents: sequence tag, barcode, clone type, T cell or B cell repertoire, sequence by synthesis, sequence analysis, minimal residual disease, tag*, unique, amplif*, lymphocyte*																	
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>US 2011/0207135 A1 (FAHAM et al.) 25 August 2011 (25.08.2011); para [0015], [0029], [0034], [0040], [0073], [0090], [0096], [0099],</td> <td>1-2, 7-10</td> </tr> <tr> <td>Y</td> <td></td> <td>3-6, 11-14</td> </tr> <tr> <td>Y</td> <td>US 5,846,719 A (BRENNER et al.) 08 December 1998 (08.12.1998); col 7, ln 41-60; col 12, ln 16-20</td> <td>3-6, 11-13</td> </tr> <tr> <td>Y</td> <td>US 2011/0207134 A1 (FAHAM et al.) 25 August 2011 (25.08.2011); para [0021], [0423]-[0431]</td> <td>14</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	US 2011/0207135 A1 (FAHAM et al.) 25 August 2011 (25.08.2011); para [0015], [0029], [0034], [0040], [0073], [0090], [0096], [0099],	1-2, 7-10	Y		3-6, 11-14	Y	US 5,846,719 A (BRENNER et al.) 08 December 1998 (08.12.1998); col 7, ln 41-60; col 12, ln 16-20	3-6, 11-13	Y	US 2011/0207134 A1 (FAHAM et al.) 25 August 2011 (25.08.2011); para [0021], [0423]-[0431]	14
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.															
X	US 2011/0207135 A1 (FAHAM et al.) 25 August 2011 (25.08.2011); para [0015], [0029], [0034], [0040], [0073], [0090], [0096], [0099],	1-2, 7-10															
Y		3-6, 11-14															
Y	US 5,846,719 A (BRENNER et al.) 08 December 1998 (08.12.1998); col 7, ln 41-60; col 12, ln 16-20	3-6, 11-13															
Y	US 2011/0207134 A1 (FAHAM et al.) 25 August 2011 (25.08.2011); para [0021], [0423]-[0431]	14															
☐ Further documents are listed in the continuation of Box C. ☐																	
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family																	
Date of the actual completion of the international search 06 January 2014 (06.01.2014)		Date of mailing of the international search report 29 JAN 2014															
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201		Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774															

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 13/45276

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-14, directed to a method of determining clonotypes of an immune repertoire.

Group II: Claims 15-26, directed to a method of determining nucleotide sequences of one or more polynucleotides in one or more sequencing-by synthesis reactions.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

*****Continued in Supplemental Box*****

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-14

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US 13/45276

Continuation of Box III: (Lack of Unity of Invention)

Special Technical Features:

Group I requires the special technical features of a method of determining clonotypes of an immune repertoire comprising the steps of obtaining a sample from an individual comprising T-cells and/or B-cells, and aligning sequence reads of like sequence tags to determine sequence reads corresponding to the same clonotypes of the repertoire, not required by Group II.

Group II requires the special technical features of a method of determining nucleotide sequences of one or more polynucleotides in one or more sequencing-by synthesis reactions, wherein sequencing by synthesis comprises at least one dNTP flow; and determining for each tag-polynucleotide having the same sequence tag a number of nucleotide incorporations for each dNTP flow as a function of measured incorporation signals for each such dNTP flow, not required by Group I.

Common Technical Feature

The common technical elements shared by the above groups is that they are related to attaching a tag to a polynucleotide molecule to form a tag-molecule conjugate wherein substantially every polynucleotide of the tag-polynucleotide has a unique sequence tag, amplifying the tag-polynucleotide conjugates, sequencing the tag-polynucleotide conjugates. This common technical element does not represent an improvement over the prior art of US 2011/0207135 A1 to Faham et al. (hereinafter "Faham"), which discloses a method of determining clonotypes of an immune repertoire (para [0041]; "the invention may be used to provide simultaneously (this is, based on measurements on a single sample) lymphocyte number and clonality. Such embodiment may be implemented with the following steps: obtaining from an individual a sample comprising T cells and/or B cells"), the method comprising the steps:

(a) obtaining a sample from an individual comprising T-cells and/or B cells (para [0041]; "Such embodiment may be implemented with the following steps: obtaining from an individual a sample comprising T cells and/or B cells");

(b) attaching sequence tags to molecules of recombinant nucleic acids of T cell receptor genes or immunoglobulin genes of the T-cells and/or B-cells to form tag molecule conjugates, wherein substantially every molecule of the tag-molecule conjugates has a unique sequence tag (para [0040]; "clonotypes are readily distinguished in the same sequencing run by the use of labels, particularly oligonucleotide tags that are attached during the sample preparation process; [0096] and Fig. 2B, 'Read2_TagX_P7' comprises a Tag (221); [0099], 'each tag is different from all the other tags');

(c) amplifying the tag-molecule conjugates (para [0096], [0098] 'amplify for the purpose of identifying the sample tag');

(d) sequencing the tag-molecule conjugates (para [0040]); and

(e) aligning sequence reads of like sequence tags to determine sequence reads corresponding to the same clonotypes of the repertoire (para [0073] 'sequences of clonotypes are determined in part by aligning sequence reads', 'sample tag to distinguish between clonotypes originating from different biological samples'; Fig. 3B, sample tag (312)).

As the common technical feature was known in the art at the time of the invention, it cannot be considered a special technical feature that would otherwise unify the groups. The inventions lack unity with one another

Therefore, the inventions of Groups I-II lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.

フロントページの続き

(31)優先権主張番号 61/829,054
 (32)優先日 平成25年5月30日(2013.5.30)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 13/835,093
 (32)優先日 平成25年3月15日(2013.3.15)
 (33)優先権主張国 米国(US)

(81)指定国 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC

(72)発明者 ファハム、マレク
 アメリカ合衆国 94080 カリフォルニア州 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 94080 カリフォルニア州 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 ウィリス、トーマス
 アメリカ合衆国 94080 カリフォルニア州 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 ジェン、ジャンピャオ
 アメリカ合衆国 94080 カリフォルニア州 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

F ターム(参考) 4B024 AA11 CA01 CA04 CA09 CA11 CA20 HA11
 4B063 QA01 QA13 QA18 QA19 QQ02 QQ08 QQ42 QQ52 QR32 QR35
 QR55 QR62 QS16 QS25 QS32 QX01