

(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2015-504661

(P2015-504661A)

(43) 公表日 平成27年2月16日(2015.2.16)

(51) Int.Cl.	F I	テーマコード (参考)
<b>C 1 2 Q 1/68 (2006.01)</b>	C 1 2 Q 1/68 A	2 G 0 4 5
<b>C 1 2 N 15/09 (2006.01)</b>	C 1 2 N 15/00 A	4 B 0 2 4
<b>G 0 1 N 33/50 (2006.01)</b>	G 0 1 N 33/50 P	4 B 0 6 3

審査請求 未請求 予備審査請求 未請求 (全 31 頁)

(21) 出願番号	特願2014-548844 (P2014-548844)	(71) 出願人	511110533
(86) (22) 出願日	平成24年12月19日 (2012.12.19)		シーケンタ インコーポレイテッド
(85) 翻訳文提出日	平成26年8月4日 (2014.8.4)		アメリカ合衆国 カリフォルニア州 サウ
(86) 国際出願番号	PCT/US2012/070674		ス サンフランシスコ イースト ジェイ
(87) 国際公開番号	W02013/096480		ミー コート 4 0 0 スイート 3 0 1
(87) 国際公開日	平成25年6月27日 (2013.6.27)	(74) 代理人	100102978
(31) 優先権主張番号	61/578,175		弁理士 清水 初志
(32) 優先日	平成23年12月20日 (2011.12.20)	(74) 代理人	100102118
(33) 優先権主張国	米国 (US)		弁理士 春名 雅夫
		(74) 代理人	100160923
			弁理士 山口 裕孝
		(74) 代理人	100119507
			弁理士 刑部 俊
		(74) 代理人	100142929
			弁理士 井上 隆一

最終頁に続く

(54) 【発明の名称】 免疫レパートリー解析による、濾胞性リンパ腫からびまん性大細胞型B細胞リンパ腫への形質転換のモニタリング

## (57) 【要約】

本発明は、個体の連続した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特定のグループ（本明細書中で「クラン」と称される）における変化および/または変化の欠如を測定することにより、濾胞性リンパ腫からびまん性大細胞型B細胞リンパ腫（DLBCL）への形質転換を該個体において予後予測する方法に関する。クランは、多くの関連するリンパ球子孫を発生させる単一のリンパ球前駆体から生じてよく、各リンパ球子孫は、体細胞変異、例えば塩基置換、逆位、結果として共通のV(D)J遺伝子セグメントの使用をもたらす関連する再構成などが原因で、わずかに異なる免疫グロブリン受容体を保有および/または発現する。濾胞性リンパ腫からDLBCLへの形質転換の可能性がより高いことは、連続したクロノタイププロファイルにおいて、クランのクロノタイプメンバー構成に経時的な多様化が起こらずに該クランが存続することと相関がある。

**【特許請求の範囲】****【請求項 1】**

以下の工程を含む、個体における濾胞性リンパ腫からびまん性大細胞型B細胞リンパ腫への形質転換を検出する方法：

(a) 個体からBリンパ球を含有する試料を得る工程；

(b) 免疫グロブリン遺伝子の組換えDNAを含む核酸または該組換えDNAよりコピーされた核酸から、クロノタイププロファイルを生成する工程；

(c) クロノタイププロファイルから、クランおよびそのサイズを決定する工程；

(d) 工程(a)～(c)を繰り返す工程；ならびに

(e) 実質的に変化していない多様性を有する1つまたは複数のクランを、該個体において濾胞性リンパ腫がびまん性大細胞型B細胞リンパ腫へと形質転換する高い可能性と関連させる工程。

10

**【請求項 2】**

クランの各々がクロノタイプからなり、該クロノタイプが各々、そのクランの他のすべてのメンバーと少なくとも90パーセント相同である、請求項1に記載の方法。

**【請求項 3】**

各クロノタイプが、同じVおよびJ参照セグメントにマッピングされ、このようなマッピングが、クロノタイプ配列中の同じ相対位置で行われ、かつそれらのNDN領域の各々が実質的に同一である場合には必ず、該クロノタイプは、同じクランのメンバーである、請求項1に記載の方法。

20

**【請求項 4】**

(a) 各クロノタイプのVリードが、同じV領域にマッピングされ、(b) 各クロノタイプのCリードが、同じJ領域にマッピングされ、(c) 各クロノタイプのNDN領域が実質的に同一であり、かつ(d) 各クロノタイプのV-NDN境界とJ-NDN境界との間のNDN領域の位置が同じである場合には必ず、該クロノタイプは、同じクランのメンバーである、請求項1に記載の方法。

**【請求項 5】**

(a) 各クロノタイプのVリードが、同じV領域にマッピングされ、(b) 各クロノタイプのCリードが、同じJ領域にマッピングされ、(c) 各クロノタイプのNDN領域が実質的に同一であり、かつ(d) 各クロノタイプのD領域に付加された下流側の塩基が同じでありかつ各クロノタイプのD領域に付加された上流側の塩基が同じである場合には必ず、該クロノタイプは、同じクランのメンバーである、請求項1に記載の方法。

30

**【請求項 6】**

(a) 各クロノタイプのVリードが同一であり、(b) 各クロノタイプのCリードが同一であり、かつ(c) 各クロノタイプのNDN領域が異なる場合には必ず、該クロノタイプは、同じクランのメンバーである、請求項1に記載の方法。

**【請求項 7】**

(a) 各クロノタイプのCリードが同一であり、(b) 各クロノタイプのND領域が同一であり、かつ(c) 各クロノタイプのV領域が異なる場合には必ず、該クロノタイプは、同じクランのメンバーである、請求項1に記載の方法。

40

**【請求項 8】**

繰り返す工程において、クラン中のクロノタイプの少なくとも80パーセントが、6か月の期間内で少なくとも2か月の間隔をあけたクロノタイププロファイルの少なくとも2回の連続した測定において同じである場合には必ず、該クランは、実質的に変化していない多様性を有する、請求項1～7のいずれか一項に記載の方法。

**【請求項 9】**

繰り返す工程において、第2のクロノタイププロファイルにおけるクラン中のクロノタイプの少なくとも80パーセントが、第1のクロノタイププロファイルにおける該クラン中のクロノタイプと同じであり、第1のクロノタイププロファイルと第2のクロノタイププロファイルとが6か月の期間内で少なくとも2か月の間隔をあけたものである場合には必ず、

50

該クランは、実質的に変化していない多様性を有する、請求項1～7のいずれか一項に記載の方法。

【請求項 1 0】

繰り返す工程において、第2のクロノタイププロファイルにおけるクラン中のクロノタイプの少なくとも90パーセントが、第1のクロノタイププロファイルにおける該クラン中のクロノタイプと同じであり、第1のクロノタイププロファイルと第2のクロノタイププロファイルとが6か月の期間内で少なくとも2か月の間隔をあけたものである場合には必ず、該クランは、実質的に変化していない多様性を有する、請求項1～7のいずれか一項に記載の方法。

【請求項 1 1】

組換え配列が各々、IgH鎖のV(D)J領域の一部を含む、請求項1～10のいずれか一項に記載の方法。

【請求項 1 2】

V(D)J領域の前記部分が、30から200ヌクレオチドの範囲の長さを有するヌクレオチド配列を含む、請求項11に記載の方法。

【請求項 1 3】

クロノタイププロファイルが各々、少なくとも $10^4$ 種のクロノタイプを有する、請求項1～12のいずれか一項に記載の方法。

【請求項 1 4】

生成する工程が、

(a) 免疫グロブリン遺伝子に由来する組換えDNA配列またはそのコピーを含む、B細胞由来の核酸の分子を、増幅すること；および

(b) クロノタイププロファイルを形成するために、増幅された該核酸の分子を配列決定すること

を含む、請求項1～13のいずれか一項に記載の方法。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

相互参照

本出願は、参照により全体として本明細書に組み入れられる、2011年12月20日に出願された米国特許仮出願第61/578,175号の恩典を主張する。

【背景技術】

【0 0 0 2】

発明の背景

濾胞性リンパ腫（FL）は、びまん性大細胞型B細胞リンパ腫（DLBCL）などのより急速に進行するリンパ腫の前段階であることが多く、非ホジキンリンパ腫全体の25～40パーセントを占める；Lossos et al, Proc. Natl. Acad. Sci., 99: 8886-8891 (2002)（非特許文献1）。FLは、初期には緩徐進行性であるが、化学療法に対する感受性の低下と関連する連続的な再発のパターンを示す。より急速に進行する大細胞型リンパ腫への形質転換は、FLを有する患者の25～60パーセントにおいて起こる；Lossos et al（上記で引用、非特許文献1）。この過程において、細胞のより悪性のサブクローンが現れ、処置に対して不応性の臨床経過および短い生存につながる。以下の5つのリスク因子に基づいた国際予後指標（IPI）が、DLBCLについて開発されている：（1）年齢＞60、（2）進行した臨床段階、すなわちIII～IV、（3）正常より高い血清乳酸デヒドロゲナーゼレベル、（4）ECOGパフォーマンスステータス（全身の健康の尺度）、および（5）1つより多い節外疾患部位の存在。これらのリスク因子は、4つの予後カテゴリーの基礎を形成する。

10

20

30

40

患者の全体的なリスク	存在する リスク因子	3年生存率(%)
低	0~1	91
低~中	2	81
高~中	3	65
高	4~5	59

多くの他の癌と同様に、DLBCLへの形質転換の早期検出は、患者の生存に対して明らかな影響を与える。

#### 【0003】

T細胞受容体もしくはB細胞受容体またはそれらの成分などの免疫分子をコードする核酸のプロファイルは、生物の健康または疾患の状態に関する豊富な情報を含み、そのため、そのようなプロファイルを診断または予後予測の指標として使用することが、多種多様な状態に対して提唱されてきている；例えば、Faham and Willis、米国特許出願公開第2010/0151471号（特許文献1）および米国特許出願公開第2011/0207134号（特許文献2）；Freeman et al, Genome Research, 19: 1817-1824 (2009)（非特許文献2）；Boyd et al, Sci. Transl. Med., 1(12): 12ra23 (2009)（非特許文献3）；He et al, Oncotarget (March 8, 2011)（非特許文献4）。このような配列ベースのプロファイルは、免疫レパートリーまたはその成分のクロノタイプを測定するための他のアプローチ、例えば、van Dongen et al, Leukemia, 17: 2257-2317 (2003)（非特許文献5）；Ottensmeier et al, Blood, 91: 4292-4299 (1998)（非特許文献6）よりもはるかに高い感度を有し得る。

#### 【0004】

ハイスループット配列決定などの新たな高感度技法が、緩徐進行期から急速進行期へのFLの形質転換をより早期にモニタリングおよび検出するために利用可能であれば、FLを有する患者にとって有利である。

#### 【先行技術文献】

#### 【特許文献】

#### 【0005】

【特許文献1】米国特許出願公開第2010/0151471号

【特許文献2】米国特許出願公開第2011/0207134号

#### 【非特許文献】

#### 【0006】

【非特許文献1】Lossos et al, Proc. Natl. Acad. Sci., 99: 8886-8891 (2002)

【非特許文献2】Freeman et al, Genome Research, 19: 1817-1824 (2009)

【非特許文献3】Boyd et al, Sci. Transl. Med., 1(12): 12ra23 (2009)

【非特許文献4】He et al, Oncotarget (March 8, 2011)

【非特許文献5】van Dongen et al, Leukemia, 17: 2257-2317 (2003)

【非特許文献6】Ottensmeier et al, Blood, 91: 4292-4299 (1998)

#### 【発明の概要】

#### 【0007】

本発明は、配列ベースの免疫レパートリー解析を用いた、濾胞性リンパ腫からびまん性大細胞型B細胞リンパ腫への形質転換を検出するために個体をモニタリングするための方法に関する。本発明は多くの実行および適用において例証されるが、そのうちのいくつかについて以下におよび本明細書を通して概要を述べる。

#### 【0008】

1つの局面において、本発明は、以下の工程により、濾胞性リンパ腫からびまん性大細胞型B細胞リンパ腫への形質転換を個体において予測する方法を提供する：（a）個体からBリンパ球を含有する試料を得る工程；（b）免疫グロブリン遺伝子の組換えDNAを含む核酸または該組換えDNAよりコピーされた核酸から、クロノタイププロファイルを生成する工程；（c）クロノタイププロファイルから、クランおよびそのサイズを決定する工程；

(d) 工程(a)～(c)を繰り返す工程；ならびに(e)実質的に変化していない多様性を有する1つまたは複数のクランを、該個体において濾胞性リンパ腫がびまん性大細胞型B細胞リンパ腫へと形質転換する高い可能性と関連させる工程。

【0009】

本発明は一部には、免疫活性化を起こしている個体では、B細胞レパトリーのクロノタイププロファイルが、2種またはそれ以上のアイソタイプと関連したクロノタイプ、すなわち異なるアイソタイプを示す重鎖定常領域のセグメントと関連したクロノタイプが高頻度であることによって特徴づけられる、という認識および理解である。本発明の、上記で特徴づけられたこれらの局面およびその他の局面は、説明がなされる多くの実行および適用において例証されるが、そのうちのいくつかを図面で示し、添付の特許請求の範囲において特徴づける。しかしながら、上記の概要は、本発明のそれぞれ説明がなされる態様またはすべての実行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】

【0010】

本発明の新規な特徴は、添付の特許請求の範囲に詳細に記載されている。本発明の特徴および利点のより良い理解は、本発明の原理が利用される例示的な態様を記載している以下の詳細な説明、および添付の図面を参照することによって得られる。

【0011】

【図1】図1Aは、IgH転写物およびその内部の天然の可変性の源を図示する。図1B～1Dは、免疫グロブリン遺伝子を増幅および配列決定するための二段階PCRスキームを示す。

【図2】図2Aは、図1CのPCR産物のヌクレオチド配列を決定する一態様の詳細を図示する。図2Bは、図1CのPCR産物のヌクレオチド配列を決定する別の態様の詳細を図示する。

【図3】図3Aは、単一反応においてIgH鎖から3つの配列決定鋳型を生成するためのPCRスキームを図示する。図3B～3Cは、3つの別々の反応においてIgH鎖から3つの配列決定鋳型を生成し、その後、得られたアンプリコンを混合して、P5プライマー結合部位およびP7プライマー結合部位を付加するための二次PCRを行うためのPCRスキームを図示する。図3Dは、IgH鎖について生成された配列リードの位置を図示する。

【発明を実施するための形態】

【0012】

発明の詳細な説明

本発明の実施は、特記されない限り、当技術分野の技術の範囲内にある、分子生物学(組換え技法を含む)、バイオインフォマティクス、細胞生物学、および生化学の従来の技法および説明を使用することができる。このような従来の技法には、血液細胞のサンプリングおよび解析、核酸の配列決定および解析などが含まれるが、これらに限定されない。適切な技法の具体的な例は、本明細書における以下の実施例を参照することによって知ることができる。しかしながら、その他の同等な従来の手順もまた、当然ながら用いることができる。このような従来の技法および説明は、Genome Analysis: A Laboratory Manual Series (Vols. I-IV) ; PCR Primer: A Laboratory Manual ; およびMolecular Cloning: A Laboratory Manual (いずれもCold Spring Harbor Laboratory Pressによる)などの標準的な実験マニュアルにおいて見出すことができる。

【0013】

本発明は、クロノタイプのクランの構造および変化をモニタリングすることにより、濾胞性リンパ腫からびまん性大細胞型B細胞リンパ腫などのより急速に進行するリンパ腫への形質転換を個体において予測または予後予測するための方法に関する。1つの局面において、形質転換は、多様性が実質的に変化していない少なくとも1つのクロノタイプクランの出現および存続と関連する。換言すると、多様性が実質的に変化していない少なくとも1つのクロノタイプクランが出現および存続する場合にはいつでも、そのような形質転換が起こる高い可能性がある。これは、クロノタイププロファイルにおけるそのようなクランの部分的サイズの増加と関連することもあり、または関連しないこともある。本明細書において用いられる、クランに関する「多様性」とは、クランを作っているクロノタイ

ブのセットが経時的に変化する量または比率を意味する。下記のように、各クランは、関連するクロノタイプからなる。1つの局面において、そのような関係は、B細胞受容体を受ける体細胞超変異の量により決定される。濾胞性リンパ腫は、高い比率の体細胞超変異を特徴とし、そのためクランでは、個体に特徴的な比率で、その過程に基づいて多様化が起こる。本発明に従って、この比率は、変化が比較される参照比率とみなされてもよい。あるいは、濾胞性リンパ腫を有する患者における体細胞超変異の比率の集団平均が用いられてもよい。患者平均に基づく参照比率は、経時的に複数回のクロノタイププロファイル測定を行うこと、および各測定についてクラン、そのサイズの変化を決定し、かつ各クランについて追加された新たなクロノタイプの数および消失した以前のクロノタイプの数を決

定することにより、得ることができる。参照比率または値は、そのような測定された量の関数であってもよい。例えば、参照比率は、所定の期間もしくは所定の測定回数またはその2つの何らかの組み合わせにわたる、所定数のクランから追加されたおよび除去されたクロノタイプの数

の移動平均であってもよい。1つの態様において、測定は約2か月ごとに行われ、参照比率は、複数のクランからの追加および除去についての、複数の測定時点の各々における移動平均とみなされる。1つの態様において、後者の複数とは、2~10の間の数（2および10を含む）であるか；または3~5の間の数（3および5を含む）である。別の態様において、前者の複数とは、1~10の間の数（1および10を含む）であるか；または3~5の間の数（3および5を含む）である。

10

20

30

40

50

#### 【0014】

本発明の別の態様において、クロノタイプの少なくとも80パーセントが、同じ6か月の期間内で少なくとも2か月の間隔をあけた少なくとも2回の、濾胞性リンパ腫を有する個体のクロノタイププロファイルの測定において同じである場合、クランは存続しているまたはクランは実質的に多様化されていない。本発明の別の態様において、クロノタイプの少なくとも90パーセントが、同じ6か月の期間内で少なくとも3か月の間隔をあけた少なくとも2回の、濾胞性リンパ腫を有する個体のクロノタイププロファイルの測定において同じである場合、クランは存続しているまたはクランは実質的に多様化されていない。

#### 【0015】

異なるリンパ球は、様々な配列特性に関して互いに関連するクロノタイプを頻繁に生じる。すなわち、その配列が類似しているクロノタイプを生じる複数のリンパ球が、存在または発達し得る。これは、IgH分子の場合、超変異などの様々な機構によるものであり得る。別の例として、リンパ系新生物などの癌においては、単一のリンパ球前駆体が、多くの関連するリンパ球子孫を発生させ得、各リンパ球子孫は、癌に関連する体細胞変異、例えば塩基置換、異常な再構成などが原因で、わずかに異なるBCR、したがって異なるクロノタイプを保有および/または発現する。そのような関連するクロノタイプのセットは、本明細書において「クラン」と称される。いくつかの場合に、クランのクロノタイプは、別のクランメンバーの変異から生じ得る。そのような「子」クロノタイプは、系統発生学的クロノタイプと称され得る。クラン内のクロノタイプは、親クロノタイプに対する、または互いに対する関連性の1つまたは複数の測定により同定され得る。1つの態様において、クロノタイプは、以下でより十分に記載されるように、%相同性によって同じクランにグループ化され得る。別の態様において、クロノタイプは、V領域、J領域、および/またはNDN領域の共通した使用によって1つのクランに割り当てられ得る。例えば、クランは、共通のJおよびND領域を有するが異なるV領域を有するクロノタイプによって定義されてもよく；または同じVおよびJ領域を有する（同一の塩基置換変異を含む）が異なるNDN領域を有するクロノタイプによって定義されてもよく；または1~10塩基、もしくは1~5塩基、もしくは1~3塩基の、1つもしくは複数の挿入および/もしくは欠失を起こしてクランメンバーを生成したクロノタイプによって定義されてもよい。別の態様において、クランのメンバーは以下の通りに決定される。クロノタイプは、以下の基準を満たす場合に、同じクランに割り当てられる：i）それらが同じVおよびJ参照セグメントにマッピングされ、これらのマッピングがクロノタイプ配列中の同じ相対位置で行われていること（例えば、図3Bにおいて、プライマー(3404)からの配列リードに基づくクロノタイプは、プライマ

ー(3406)からの配列リードに基づくクロノタイプと同じクラン中ではない)、およびii) それらのNDN領域が実質的に同一であること。本明細書において用いられる、クロノタイプおよび配列セグメントに関する「マッピング」、「マッピングする」、または「マッピングされる」とは、クロノタイプが、示される配列セグメントを含むことを意味する。クランメンバー構成に関する「実質的」とは、NDN領域におけるいくつかの小さな違いが、この領域において体細胞変異が起こり得るものであることから許容されることを意味する。好ましくは、1つの態様において、NDN領域における変異を誤ってコールすることを回避するために、塩基置換が癌関連変異として受諾されるか否かは、クランのNDN領域のサイズに直接的に依存する。例えば、方法は、クランNDN配列の長さが $m$ ヌクレオチドもしくはそれより長い、例えば9ヌクレオチドもしくはそれより長い場合に、あるクロノタイプが癌関連変異としてクランNDN配列との1塩基の違いを有するならば、そのクロノタイプをクランメンバーとして受諾し得るが、それ以外は受諾せず、またはクランNDN配列の長さが $n$ ヌクレオチドもしくはそれより長い、例えば20ヌクレオチドもしくはそれより長い場合に、そのクロノタイプが癌関連変異としてクランNDN配列との2塩基の違いを有するならば、そのクロノタイプをクランメンバーとして受諾し得るが、それ以外は受諾しない。別の態様において、クランのメンバーは、以下の基準を用いて決定される：(a) Vリードが同じV領域にマッピングされること、(b) Cリードが同じJ領域にマッピングされること、(c) NDN領域が(上記のように)実質的に同一であること、および(d) V-NDN境界とJ-NDN境界との間のNDN領域の位置が同じであること。換言すると、条件(d)は、Dに対する下流側の塩基付加の数が同じであり、かつDに対する上流側の塩基付加の数が同じであることを意味する。したがって、そのような1つのクランに由来するクロノタイプの2つのNDN領域が「 $n_1D_jn_2$ 」および「 $n_3D_kn_4$ 」(式中、 $n_1$ および $n_3$ は各々、それぞれJ領域と $D_j$ または $D_k$ との間のヌクレオチドの数であり、 $D_j$ および $D_k$ は特定のD領域であり、かつ $n_2$ および $n_4$ は、それぞれV領域と $D_j$ および $D_k$ との間のヌクレオチドの数である)で表わされる場合は、 $n_1 = n_3$ 、 $D_j = D_k$ 、および $n_2 = n_4$ である。本明細書において用いられる、NDN領域に関する「ND」および「DN」という用語は、それぞれ(a) D領域およびD領域とJ領域との間のヌクレオチドを含む、NDN領域の一部分、ならびに(b) D領域およびD領域とV領域との間のヌクレオチドを含む、NDN領域の一部分を意味する。

#### 【0016】

単一試料のクロノタイプはクランにグループ化され得、異なる時点で獲得された連続した試料に由来するクランが互いに比較され得る。特に、本発明の1つの局面において、リンパ系新生物などの疾患と相關するクロノタイプを含有するクランが、各試料のクロノタイプから同定され、寛解の継続、再発の初期段階、さらなるクローン進化の証拠などのような疾患状態を決定するために、直前の試料のものと比較される。本明細書において用いられる、クランに関する「サイズ」とは、クラン中のクロノタイプの数を意味する。いくつかの適用において、クランのサイズは、試料の異なるサイズを考慮して標準化され得る。例えば、第1の試料において $10^6$ 個のリンパ球が得られ得るのに対して、第2の試料において $5 \times 10^5$ 個のリンパ球が得られ得る；したがって、クランサイズが、単にそのクランにおいて数えられるクロノタイプの数に基づいていた場合は、クランサイズは、2回の測定にわたって減少するように見られ得る。いくつかの態様において、クランに関する「サイズ」とは、試料中のリンパ球の数に関して標準化されたクロノタイプの数を意味する。いくつかの態様において、クランのサイズは、クロノタイププロファイルにおけるそのクロノタイプの頻度である。

#### 【0017】

免疫レパートリーの複雑性は、周知である；例えば、Arstila et al, Science, 286: 958-961 (1999)およびWarren et al (上記で引用)。図1Aは、そこから配列リードが生成され、かつクロノタイプが決定されるIgH分子の典型的な転写物(120)を図示する。天然の配列可変性の源は、遺伝子の大きなセットに由来するC、D、J、およびVセグメントの組換え、いわゆる「NDN」領域を生じるためのDセグメントの末端へのヌクレオチドの付加および欠失、ならびに転写物の長さ(122)にわたって大まかに曲線(128)により示される相対頻

10

20

30

40

50

度で置換がランダムに生じる体細胞超変異を含む。いくつかの態様において、活性化された免疫状態に対応する大きなクランは、超変異イベント、例えば、1個または複数個の点変異、または1~5個の点変異、または1~10個の点変異、または1~20個の点変異により関連付けられるクロノタイプを含む。本発明の1つの局面において、そのようなIgHおよびTCRの転写物の複雑な集団を増幅し、かつ配列決定する。1つの局面において、IgH分子についての一方または両方の作業を、V領域中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ることにより実施する(以下でより十分に記載される)。これは、比較的高いエラー率を有する配列決定化学を使用する場合、またはそのような配列可変性を事前に知ることが困難もしくは不可能である場合に、特に有利である。後者の場合には、(例えば)1つまたは複数の体細胞変異により引き起こされたミスマッチのために、1つまたは複数のプライマー結合部位が動作不能であるかまたは実質的に動作不能である場合でさえも、増幅または配列リードの生成のためのプライマー伸長が起こる。曲線(128)により示される相対変異頻度は、プロモーターP(122)から始まり、リーダー領域(124)を経て、転写物のV(D)J領域(126)の上で最大まで上昇し、その後ほぼゼロまで低下する。いくつかの態様において、鋳型の入れ子セットを生成するための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たPCRにより、組換えB細胞核酸のセグメントを増幅する(図3Aおよび図3Bならびにそれらについての以下の説明を参照されたい)。そのようなセットに由来する鋳型を表面上でさらに増幅して、別々のアンプリコンを形成してもよい(例えば、cBot装置、Illumina, San Diego, CAを用いるブリッジPCRにより)。同じ入れ子セットに由来する鋳型は、それらの共通末端で生成された配列リードによって互いに関連付けられ得る。鋳型の入れ子セットにより、比較的高いエラー率を有する配列決定化学を用いて、同時に配列の全長にわたって高い平均品質スコアを維持しながら、そうでない場合に可能である長さよりも長い配列を解析することが可能になる。入れ子セットはまた、V領域が体細胞超変異を受けた場合でさえ、V領域から少なくとも1つの配列リードを得ることを確実にする。1つの態様において、IgH分子などの可変性の高い核酸を解析するために、下記以上のエラー率を有する配列決定化学を使用してもよい: 配列リードの0.2パーセントが、1~50位において少なくとも1つのエラーを含み; 配列リードの0.2~1.0パーセントが、51~75位において少なくとも1つのエラーを含み; 配列リードの0.5~1.5パーセントが、76~100位において少なくとも1つのエラーを含み; および配列リードの1~5パーセントが、101~125位において少なくとも1つのエラーを含む。上記を考慮して、本発明の方法は、密接に関連している、および密接に関連するものとは純粋に異なっている、および配列決定または他のエラーの結果であるクロノタイプ配列を識別するための工程を含む。

#### 【0018】

配列リードデータからのクロノタイプの構築は、一部、そのようなデータを生成するのに使用された配列決定法に依存しており、これは、方法が異なると、予想リード長およびデータ品質が異なるためである。1つのアプローチにおいて、解析用の配列リードデータを生成するためにSolexaシーケンサーが使用される。1つの態様において、少なくとも100万の鋳型分子をもたらすための少なくとも $0.5 \sim 1.0 \times 10^6$ 個のリンパ球を提供する試料が得られ、鋳型分子は、任意の増幅後に、対応する100万またはそれ以上の鋳型分子クローン集団(またはクラスター)をもたらし得る。Solexaアプローチを含む最もハイスループットな配列決定アプローチについては、配列決定の精度を増大させるべく高い冗長度で各鋳型配列が決定されるように、そのようなクラスターレベルでの過剰サンプリングが望ましい。Solexaベースでの実施について、好ましくは、各々独立した鋳型の配列は、10回またはそれ以上決定される。異なる予想リード長およびデータ品質を有する他の配列決定アプローチについては、同等の配列決定精度のために異なる冗長度レベルが使用されてもよい。当業者は、上記のパラメータ、例えば試料サイズ、冗長度などが、特定の適用に関連する設計上の選択事項であることを認識している。

#### 【0019】

本発明に従って、リンパ球由来の組換え核酸の試料からクロノタイププロファイルを生成した後、クランを同定し、そのサイズを決定し、かつその数をカウントするか、または

10

20

30

40

50



いくつかの態様においては、その頻度を決定する。下記のように、それからリンパ球がサンプリングされる組織は、大きく異なってもよい。1つの態様において、リンパ球は、末梢血から；例えば、最初に末梢血単核細胞（PBMC）を単離することによりサンプリングする。そのような単離の後、RNAまたはDNAを、従来の技法を用いて抽出してもよい。クランは、上記のようなクランの定義を用いて同定する。クランのサイズは、クラン（用いられるクランの定義に依存する）中のクロノタイプの数である。1つの局面において、下記のように、クランは、1種のクロノタイプの超変異、または共有する免疫グロブリンセグメントにより関連付けられる1種もしくは複数種のクロノタイプの超変異に関して定義される。そのような場合、B細胞由来の組換え配列を使用して、クロノタイププロファイルを生成する。そのようなデータを得た後、これを、同じ個体に由来するか、またはそのような値の集団平均が決定される個体の集団に由来するかのいずれかの他のクロノタイププロファイルの参照値または参照範囲と比較することができる。

10

#### 【0020】

##### 試料

本発明の方法のためのクロノタイププロファイルは、B細胞を含む試料から抽出された核酸の試料から生成される。B細胞には、例えば、形質B細胞、記憶B細胞、B1細胞、B2細胞、辺縁帯B細胞、および濾胞性B細胞が含まれる。B細胞は免疫グロブリン（抗体、B細胞受容体）を発現し得る。1つの局面において、B細胞の試料は少なくとも1,000個のB細胞を含む；しかしより典型的には、試料は少なくとも10,000個のB細胞を含み、かつ、より典型的には少なくとも100,000個のB細胞を含む。別の局面において、試料は、B細胞1000~1,000,000個の範囲内の数のB細胞を含む。「クロノタイプ」および「レパートリー」の定義において以下にさらに記載されるように、試料の適切なサンプリングは、レパートリーデータを解釈する重要な局面である。試料中の細胞の数は、測定感度に制限を課す。例えば、1,000個のB細胞を含む試料では、そのような細胞のDNAが配列決定によって解析される際にどのくらいの配列リードが得られるかにかかわらず、検出可能なクロノタイプの最低頻度は1/1000または0.001である。

20

#### 【0021】

試料は、核酸、例えばDNA（例えばゲノムDNAもしくはミトコンドリアDNA）またはRNA（例えばメッセンジャーRNAもしくはマイクロRNA）を含み得る。核酸は、無細胞DNAまたはRNA、例えば循環系から抽出されたそれらであり得る。Vlassov et al, Curr. Mol. Med., 10: 142-165 (2010); Swarup et al, FEBS Lett., 581: 795-799 (2007)。提供される発明の方法において、解析され得る対象由来のRNAまたはDNAの量は、例えば、いくつかの用途（例えば較正試験）では単一細胞という少量、ならびに6pg~60ugのDNAおよびおよそ1pg~10ugのRNAの範囲に換算される1000万個の細胞またはそれ以上という多量を含む。

30

#### 【0022】

以下（定義）でより十分な議論がなされているように、リンパ球の試料は、別個のクロノタイプを有する実質的にすべてのB細胞がその中で表現（represent）され、それによって（この用語の本明細書における意味での）レパートリーが形成されるように、十分に多量である。1つの態様においては、ある集団の、0.001パーセントまたはそれ以上の頻度で存在するすべてのクロノタイプを、99パーセントの確率で含む試料が取得される。別の態様においては、ある集団の、0.0001パーセントまたはそれ以上の頻度で存在するすべてのクロノタイプを、99パーセントの確率で含む試料が取得される。1つの態様において、B細胞の試料は少なくとも50万個の細胞を含み、別の態様においてはそのような試料は少なくとも100万個の細胞を含む。

40

#### 【0023】

試料が取得される供給源物質が十分でない場合、例えば臨床研究試料等の場合には、必ず、その物質からDNAが、BCRコード配列の特異的な増幅の前に、バイアスのない技法、例えば総ゲノム増幅（WGA）、多置換増幅（MDA）；または同様の技法、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); Hosono et al,

50

Genome Research, 13: 954-964 (2003)の技法等により増幅され得る。

【0024】

血液試料は、特に注目されており、従来の技法、例えばInnis et al 編, PCR Protocols (Academic Press, 1990)等の技法を用いて得られ得る。例えば、白血球は、従来の技法、例えばRosetteSepキット(Stem Cell Technologies, Vancouver, Canada)を用いて血液試料から分離され得る。血液試料は、100  $\mu$ Lから10 mLの範囲の容量であり得；1つの局面において、血液試料の容量は、200  $\mu$ Lから2 mLまたは100  $\mu$ Lから2 mLの範囲である。DNAおよび/またはRNAは、その後、本発明の方法において使用するために、そのような血液試料から、従来の技法、例えば、DNeasy Blood & Tissueキット(Qiagen, Valencia, CA)を用いて抽出され得る。任意で、白血球のサブセット、例えばリンパ球がさらに、従来の技法、例えば蛍光活性化細胞ソーティング(FACS)(Becton Dickinson, San Jose, CA)、磁気活性化細胞ソーティング(MACS)(Miltenyi Biotec, Auburn, CA)等を用いて単離され得る。例えば、記憶B細胞は、表面マーカーCD19およびCD27によって単離され得る。

【0025】

識別性のある組換えは各個体の適応免疫細胞のDNAおよびそれらに関連するRNA転写物に存在するので、RNAまたはDNAのいずれかが、提供される発明の方法において配列決定され得る。免疫グロブリン分子またはその一部分をコードするB細胞由来の組換え配列は、クロータイプと称される。DNAまたはRNAは、抗体をコードする免疫グロブリン(Ig)遺伝子由来の配列に対応するものであり得る。

【0026】

本発明の方法において解析されるDNAおよびRNAは、重鎖免疫グロブリン(IgH)をコードする配列に対応する。各鎖は、定常(C)領域および可変領域から構成される。重鎖に関して、可変領域は、可変(V)、多様(D)および結合(J)セグメントから構成される。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別個の配列がゲノム中に存在する。B細胞の発達の際に特定のVDJ組換えイベントが起こり、これはその細胞が特定の重鎖を生成することを示すものである。組換え部位の付近では体細胞変異がしばしば起こり、それによっていくつかのヌクレオチドが付加または欠失し、これがB細胞によって生成される重鎖の多様性をさらに増大させる。B細胞により生成される抗体で生じ得る多様性は、異なる重鎖と軽鎖の掛け算である。重鎖および軽鎖の可変領域は、抗原認識(または結合)領域または部位を形成するのに寄与する。この多様性には、あるエピトープに対して特異的な応答が惹起された後に起こり得る体細胞超変異が加わる。

【0027】

本発明に従って、プライマーは、Bリンパ球から抽出された組換え核酸のアンプリコンを生成するように選択することができる。そのような配列は、本明細書において、「体細胞再構成領域」または「体細胞組換え領域」または「組換え配列」と称され得る。体細胞再構成領域は、発達段階のまたは十分に発達したリンパ球に由来する核酸を含んでよく、ここで発達段階のリンパ球とは、免疫遺伝子の再構成が完了しておらず、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再構成領域には、不完全なIgH分子(D-J領域のみを含む分子など)が含まれる。

【0028】

核酸集団の増幅

下記のように、標的核酸集団のアンプリコンは、様々な増幅技法により生成され得る。本発明の1つの局面においては、マルチプレックスPCRが、核酸の混合物、特に組換え免疫分子、例えばT細胞受容体、B細胞受容体またはそれらの一部分を含む混合物のメンバーを増幅するのに使用される。そのような免疫分子のマルチプレックスPCRを実施するための手引きは、参照により組み入れられる以下の参考文献において見出される：Faham et al、米国特許出願公開第2011/0207134号；Lim et al、米国特許出願公開第2008/0166718号等。以下でより十分な記載がなされているように、1つの局面において、個々の核酸分子を空間的に単離する工程は、事前に選択した体細胞再構成領域またはその一部分(すなわ

ち、標的配列)の一次マルチプレックス増幅を、各々が標的配列に非相補的な尾部を有する順方向および逆方向プライマーを用いて実施し、そのメンバー配列が各末端にさらなる操作を可能にする共通配列を有する第1のアンプリコンを作製することによって達成される。例えば、そのような共通末端は、単一の順方向プライマーおよび単一の逆方向プライマーを複数のそれらに代えて使用する連続増幅のためのまたは固相表面上での個々の分子のブリッジ増幅のためのプライマー結合部位等を含み得る。そのような共通末端は、上記のような1回の増幅で付加されることもあり、またはそれらは、長鎖プライマー(例えば、50~70塩基またはそれ以上)の混合物の製造および利用上の品質管理に関する難題を回避するために2工程手順で付加されることもある。そのような2工程プロセス(以下により十分な記載がある)における一次増幅は、第1のアンプリコンの配列の末端に順方向および逆方向プライマー結合部位のみを提供するようプライマーの尾部の長さが制限されることを除いて、上記のようにして実施される。二次増幅は、その後、これらのプライマー結合部位に特異的な二次増幅プライマーを用いて実施され、第2のアンプリコンの末端にさらなる配列が付加される。二次増幅プライマーは、標的配列に非相補的な尾部を有し、この部分が第2のアンプリコンの末端を形成し、かつ第2のアンプリコンのクロノタイプの配列決定に関連して利用され得る。1つの態様において、そのような付加される配列は、配列リードを生成するためのプライマー結合部位、および空間的に単離された個々の分子のクローン集団を生成するため、例えばSolexaベースの配列決定が使用される場合に、固相表面上でブリッジPCRを実施するためのプライマー結合部位を含み得る。この後者のアプローチにおいては、第2のアンプリコン由来の配列の試料が、その試料の配列にアニールすることができ、相補的オリゴヌクレオチドを付着させた固相表面上に配置され、その後、鋳型のクローン集団が形成されるまでプライマー伸長、変性、アニールのサイクルが実施される。好ましくは、試料のサイズは、(i)それが当初試料中のクロノタイプの効果的な表現(representation)を含むように、および(ii)固相表面上のクローン集団の密度がクロノタイプの明確な配列決定を実現する範囲内となるように選択される。

#### 【0029】

増幅される領域は、全クローン配列、または、免疫グロブリン遺伝子のV-D接合部、D-J接合部、免疫グロブリンの全可変領域、抗原認識領域、もしくはCDR、例えば相補性決定領域3(CDR3)を含むクローン配列のサブセットを含み得る。

#### 【0030】

ゲノムからのDNA増幅(またはRNAの逆転写によるcDNA形式での核酸増幅)の後、個々の核酸分子は、単離され、任意で再増幅され、次いで個別に配列決定され得る。例示的な増幅プロトコールは、参照により組み入れられるvan Dongen et al, Leukemia, 17: 2257-2317 (2003)またはvan Dongen et al、米国特許出願公開第2006/0234234号に見出され得る。簡潔に説明すると、例示的なプロトコールは以下のようなものである: 反応緩衝液: ABI Buffer IIまたはABI Gold Buffer (Life Technologies, San Diego, CA); 50  $\mu$ Lの最終反応容量; 100 ngの試料DNA; 10 pmolの各プライマー(以下に記載されるように増幅のバランスをとるために調整される); 終濃度200  $\mu$ MのdNTP; 終濃度1.5 mMのMgCl<sub>2</sub>(標的配列およびポリメラーゼに依存して最適化される); Taqポリメラーゼ(1~2U/チューブ); サイクル条件: 95 °での予備活性化7分間; 60 °でのアニール; サイクル時間: 30秒間の変性; 30秒間のアニール; 30秒間の伸長。本発明の方法における増幅に使用することが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ポリメラーゼまたはPfuを含む。使用するポリメラーゼの選択は、忠実性と効率性のどちらが好ましいかに基づくものであり得る。

#### 【0031】

核酸をプールから単離する方法は、DNAベクターへの核酸のサブクローニングおよび細菌の形質転換(細菌クローニング)、固相基材(例えば、ガラススライド)上での分子の二次元的な空間分離、ミセル内の溶液中での(例えばこれは分子をビーズ等の固相表面上に固定してまたはしないで油エマルジョンを使用することで達成することができる)または例えばマイクロ流体もしくはナノ流体チップ上のマイクロ反応チャンバーを用いる分子の

三次元的な空間分離を含む。希釈は、単一の分子が所定の容量、空間的領域、ビーズまたは反応チャンパー中に平均して存在することを確かめるのに使用することができる。そのような個々の核酸分子の単離方法の手引きは、以下の参考文献において見出される：Sambrook, Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 2001s); Shendure et al, Science, 309: 1728-1732 (補足資料を含む) (2005); 米国特許第6,300,070号; Bentley et al, Nature, 456: 53-59 (補足資料を含む) (2008); 米国特許第7,323,305号; Matsubara et al, Biosensors & Bioelectronics, 20: 1482-1490 (2005); 米国特許第6,753,147号等。

#### 【0032】

リアルタイムPCR、ピコグリーン染色、ナノ流体電気泳動（例えば、LabChip）またはUV吸収測定は、初期工程において、増幅可能な物質の関数的な量を判断するのに使用することができる。

#### 【0033】

1つの局面において、マルチプレックス増幅は、出発集団における配列の相対量が増幅集団またはアンプリコンにおけるそれと実質的に同じになるように実行される。すなわち、マルチプレックス増幅は、試料集団のメンバー配列間の増幅バイアスが最小となるように実行される。1つの態様において、そのような相対量は、アンプリコンにおける各相対量が出発試料におけるその値の5倍以内である場合、実質的に同じとされる。別の態様において、そのような相対量は、アンプリコンにおける各相対量が出発試料におけるその値の2倍以内である場合、実質的に同じとされる。以下でより十分な議論がなされているように、PCRにおける増幅バイアスは、任意の試料においてバイアスのない増幅を提供するPCRプライマーセットが既定のレパートリーのために選択され得る従来の技法を用いて、検出および補正され得る。

#### 【0034】

1つの態様において、増幅のバイアスは、標的配列と非相補的な尾部を有するプライマーを用いて第1または一次段階において少数の増幅サイクルを実施する二段階増幅（上述したような）を実施することによって回避され得る。この尾部は、一次アンプリコンの配列の末端に付加されるプライマー結合部位を含み、そのような部位は、1つのみの順方向プライマーおよび1つのみの逆方向プライマーを用いる第2段階の増幅において使用され、それによって増幅のバイアスの主たる原因が除かれる。好ましくは、一次PCRは、異なるプライマーによる差次的増幅を最小限にするよう十分少ないサイクル数（例えば、5~10）を有する。二次増幅は一对のプライマーを用いて実行され、したがって差次的増幅の問題は非常に小さい。一次PCRの1パーセントが二次PCRに直接利用される。2つの増幅間で使用される35サイクル（100倍希釈工程のない場合の約28サイクルに相当）は、サイクルの内訳によらずに、一次で1サイクル、二次で34サイクルであろうが、一次で25サイクル、二次で10サイクルであろうが、堅調な増幅を示すのに十分であった。理論的には一次PCRにおいて1サイクルのみ実施すれば増幅のバイアスが小さくなる可能性があるが、それ以外にも考慮すべきことがある。この1つの局面は、表現である。これは、最終的に得られるリード数に対して出発投入量が過剰でない場合に役目を果たす。例えば、1,000,000のリードが得られ、1,000,000の投入分子で開始するのであれば、二次増幅に100,000の分子からの表現のみを利用することは、当初試料中の異なる種の相対存在量を概算する精度を下げる。2つの工程の間での100倍の希釈は、一次PCR増幅が100より有意に多い分子を生成していない限り、その表現が縮小することを意味する。これは、最小で8サイクル（256倍）であるが余裕をもって10サイクル（約1,000倍）が使用され得ることを示している。その代案は、一次PCRの1%超を二次に利用することであるが、一次PCRにおいて使用されるプライマー濃度は高いため、これらのプライマーが増幅において干渉して配列間の増幅バイアスを悪化させることがないのを確実にするために、大きな希釈係数が使用され得る。別の代案は、精製または酵素処理工程を追加して一次PCR由来のプライマーを排除し、その希釈率を小さくすることである。この実施例では、一次PCRは10サイクル、二次は25サイクルであった。

10

20

30

40

50

## 【0035】

簡潔に、IgHコード核酸(RNA)を増幅するためのFaham and Willis(上記で引用)のスキームを図1B~1Dに図示する。核酸(1200)を、試料中のリンパ球から抽出し、PCRにおいて、C領域(1203)に特異的なプライマー(1202)および免疫グロブリン遺伝子の様々なV領域(1206)に特異的なプライマー(1212)と組み合わせる。プライマー(1212)はそれぞれ同一の尾部(1214)を有し、これは増幅の第2段階のためのプライマー結合部位を提供する。前述の通り、プライマー(1202)は、C領域(1203)とJ領域(1210)との間の接合部(1204)に隣接して配置される。PCRにおいて、Cコード領域(1203)の一部分、Jコード領域(1210)、Dコード領域(1208)、およびVコード領域(1206)の一部分を含むアンプリコン(1216)が生成される。それぞれIllumina DNAシーケンサーで使用するために設計された尾部(それぞれ1225および1221/1223)を有するプライマーP5(1222)およびプライマーP7(1220)を用いて、第2段階においてアンプリコン(1216)をさらに増幅する。プライマーP7(1220)の尾部(1221/1223)は、配列決定過程において別々の試料を標識するためのタグ(1221)を任意で組み入れる。第2段階の増幅により、Illumina DNAシーケンサーで使うことができるアンプリコン(1230)が作製される。

## 【0036】

配列リードの生成

本発明の方法では、任意のハイスループット核酸配列決定技法を用いることができる。好ましくは、このような技法は、そこから少なくとも1000種のクロノタイプが決定され得る、および好ましくはそこから少なくとも10,000~1,000,000種のクロノタイプが決定され得る大量の配列データを、費用に対して効果の高い方法で生成する能力を有する。DNA配列決定技法は、標識されたターミネーターまたはプライマーおよびスラブまたはキャピラリー中でのゲル分離を用いるジデオキシ配列決定反応(サンガー法)、可逆的に停止される標識ヌクレオチドを用いる合成による配列決定、パイロシークエンシング、454配列決定、標識オリゴヌクレオチドプローブのライブラリーに対するアレルト異的ハイブリダイゼーション、標識クローンのライブラリーに対するアレルト異的ハイブリダイゼーションおよびその後のライゲーションを用いる合成による配列決定、重合工程中における標識ヌクレオチドの組み込みのリアルタイムモニタリング、ポロニーシークエンシング、ならびにSOLiD配列決定を含む。分離された分子の配列決定は、最近になって、ポリメラーゼまたはリガーゼを用いる連続的または単回の伸長反応によって、およびプローブのライブラリーを用いる単回または連続的なディファレンシャルハイブリダイゼーションによって実証されている。これらの反応は、多くのクローン配列に対して並列で実施され、現在の商業利用においては1億超の配列の並列化が実現している。本発明の1つの局面においては、個々の分子を固相表面上で空間的に単離し、その表面上で配列決定を並列で行う工程を含むハイスループットの配列決定法が使用される。そのような固相表面は、非多孔性表面(例えば、Solexa配列決定におけるようなもの、例えばBentley et al, Nature, 456: 53-59 (2008)、またはComplete Genomics配列決定、例えばDrmanac et al, Science, 327: 78-81 (2010))、ビーズまたは粒子に結合された鋳型を含み得るウェルのアレイ(例えば、454と共に用いるもの、例えばMargulies et al, Nature, 437: 376-380 (2005)、またはIon Torrent配列決定、米国特許出願公開第2010/0137143号もしくは第2010/0304982号)、微細加工膜(例えば、SMRT配列決定と共に用いるもの、例えばEid et al, Science, 323: 133-138 (2009))、またはビーズアレイ(SOLiD配列決定またはポロニーシークエンシングと共に用いるもの、例えばKim et al, Science, 316: 1481-1414 (2007))を含み得る。別の局面において、そのような方法は、単離された分子を、それらを固相表面上で空間的に単離する前または後のいずれかにおいて増幅する工程を含む。先行増幅は、エマルジョンベースの増幅、例えばエマルジョンPCR、またはローリングサークル増幅を含み得る。特に関心対象となるものは、参照により組み入れられる、Bentley et al(上記で引用)および製造元の説明書(例えば、TruSeq(商標) Sample Preparation Kit and Data Sheet, Illumina, Inc., San Diego, CA, 2010); さらに以下の参考文献、米国特許第6,090,592号; 同第6,300,070号; 同第7,115,400号; およびEP0972081B1に記載されるよう

な、個々の鋳型分子を固相表面上で空間的に単離し、その後にそれらをブリッジPCRにより並列で増幅して個別のクローン集団またはクラスターを形成し、次いで配列決定する、Solexaベースの配列決定である。1つの態様において、固相表面上に配置され増幅される個々の分子は、1 cm<sup>2</sup>あたり少なくとも10<sup>5</sup>クラスターの密度；または1 cm<sup>2</sup>あたり少なくとも5×10<sup>5</sup>の密度；または1 cm<sup>2</sup>あたり少なくとも10<sup>6</sup>クラスターの密度のクラスターを形成する。1つの態様においては、比較的高いエラー率を有する配列決定化学が使用される。そのような態様において、そのような化学によりもたらされる平均品質スコアは、配列リード長の単調に下降する関数である。1つの態様において、そのような下降は、配列リードの0.5パーセントが1～75位において少なくとも1つのエラーを有すること；配列リードの1パーセントが76～100位において少なくとも1つのエラーを有すること；および配列リードの2パーセントが101～125位において少なくとも1つのエラーを有することに相当する。

10

#### 【0037】

1つの局面において、個体の配列ベースのクロノタイププロファイルは、以下の工程を用いて得られる：(a) 個体のB細胞から核酸試料を得る工程；(b) そのような核酸試料に由来する個々の分子を空間的に単離する工程であって、個々の分子が、該試料中の核酸から生成されかつ体細胞再構成領域またはその一部分を含む少なくとも1つの鋳型を含み、個々の分子が各々、少なくとも1つの配列リードを作成することができる、工程；(c) 空間的に単離された個々の分子を配列決定する工程；ならびに(d) クロノタイププロファイルを生成するために、該核酸試料に由来する核酸分子の異なる配列の存在量を決定する工程。1つの態様において、体細胞再構成領域の各々は、V領域およびJ領域を含む。別の態様において、配列決定する工程は、前記空間的に単離された個々の分子の各々を双方向に配列決定して、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作成することを含む。後者の態様に関してさらに、順方向配列リードの少なくとも1つと逆方向配列リードの少なくとも1つとが重複領域を有し、そのような配列リード間の逆相補関係によってそのような重複領域の塩基が決定されるようにする。さらに別の態様において、体細胞再構成領域の各々は、V領域およびJ領域を含み、配列決定する工程はさらに、その順方向配列リードの1つまたは複数と、J領域内のある位置から始まりそれと連結したV領域の方向へと延びる少なくとも1つの逆方向配列リードとから、個々の核酸分子の各々の配列を決定することを含む。別の態様において、個々の分子は、完全IgH分子および不完全IgH分子からなる群より選択される核酸を含む。別の態様において、配列決定する工程は、単調に下降する品質スコアを有する配列リードを生成することを含む。後者の態様に関してさらに、単調に下降する品質スコアは、その配列リードが下記以上のエラー率を有するものである：配列リードの0.2パーセントが、塩基1～50位において少なくとも1つのエラーを含み、配列リードの0.2から1.0パーセントが、51～75位において少なくとも1つのエラーを含み、配列リードの0.5から1.5パーセントが、76～100位において少なくとも1つのエラーを含む。別の態様において、上記方法は以下の工程を含む：(a) 個体のT細胞および/またはB細胞から核酸試料を得る工程；(b) そのような核酸試料に由来する個々の分子を空間的に単離する工程であって、個々の分子が、試料中の核酸から各々生成されかつ体細胞再構成領域またはその一部分を各々含む鋳型の入れ子セットを含み、各入れ子セットが、同じ方向に各々延びかつ該入れ子セットが生成された核酸上の異なる位置から各々始まる複数の配列リードを作成することができる、工程；(c) 空間的に単離された個々の分子を配列決定する工程；ならびに(d) クロノタイププロファイルを生成するために、該核酸試料に由来する核酸分子の異なる配列の存在量を決定する工程。1つの態様において、配列決定する工程は、入れ子セットの各々について複数の配列リードを作成することを含む。別の態様において、体細胞再構成領域の各々はV領域およびJ領域を含み、かつ複数の配列リードの各々は、該V領域内の異なる位置から始まりそれと連結したJ領域の方向へと延びる。

20

30

40

#### 【0038】

1つの局面において、個体由来の試料ごとに、本発明の方法において使用される配列決

50

定技法は、1回あたり少なくとも1000種のクロノタイプの配列を生成し；別の局面において、そのような技法は、1回あたり少なくとも10,000種のクロノタイプの配列を生成し；別の局面において、そのような技法は、1回あたり少なくとも100,000種のクロノタイプの配列を生成し；別の局面において、そのような技法は、1回あたり少なくとも500,000種のクロノタイプの配列を生成し；そして別の局面において、そのような技法は、1回あたり少なくとも1,000,000種のクロノタイプの配列を生成する。さらに別の局面において、そのような技法は、個々の試料につき、1回あたり100,000から1,000,000種の間のクロノタイプの配列を生成する。

#### 【0039】

提供される発明の方法において使用される配列決定技法は、1リードあたり約30 bp、約40 bp、約50 bp、約60 bp、約70 bp、約80 bp、約90 bp、約100 bp、約110 bp、約120 bp、1リードあたり約150 bp、約200 bp、約250 bp、約300 bp、約350 bp、約400 bp、約450 bp、約500 bp、約550 bpまたは約600 bpを生成することができる。

#### 【0040】

##### 配列データからのクロノタイプの生成

配列リードデータからのクロノタイプの構築は、参照により本明細書に組み入れられる Faham and Willis (上記で引用) に開示されている。簡潔に説明すると、配列リードデータからのクロノタイプの構築は、一部、そのようなデータを生成するのに使用された配列決定法に依存しており、これは、方法が異なると、予想リード長およびデータ品質も異なるためである。1つのアプローチにおいて、解析用の配列リードデータの生成にSolexaシーケンサーが使用される。1つの態様において、少なくとも100万の鋳型分子をもたらすための少なくとも $0.5 \sim 1.0 \times 10^6$ 個のリンパ球を提供する試料が得られ、鋳型分子は、任意の増幅後に、対応する100万またはそれ以上の鋳型分子クローン集団（またはクラスター）をもたらす得る。Solexaアプローチを含む最もハイスループットな配列決定アプローチについては、配列決定の精度を増大させるべく高い冗長度で各鋳型配列が決定されるように、そのようなクラスターレベルでの過剰サンプリングが望ましい。Solexaベースでの実施について、好ましくは、各々独立した鋳型の配列は10回またはそれ以上決定される。異なる予想リード長およびデータ品質を有する他の配列決定アプローチについては、同等の配列決定精度のために異なる冗長度レベルが使用され得る。当業者は、上記のパラメータ、例えば試料サイズ、冗長度などが、特定の用途に関連する設計上の選択事項であることを認識している。

#### 【0041】

1つの局面において、(図2Aに図示される) IgH鎖のクロノタイプは、そのC領域から始まりそれと連結したV領域の方向へと延びる少なくとも1つの配列リード(本明細書において「Cリード」(2304)と称される)およびそのV領域から始まりそれと連結したJ領域の方向へと延びる少なくとも1つの配列リード(本明細書において「Vリード」(2306)と称される)により決定される。そのようなリードは重複領域(2308)を有している場合とない場合があり、そのような重複領域は、図2Aに示されるようにNDN領域(2315)を含んでいる場合とない場合がある。重複領域(2308)は、全体がJ領域内にある、全体がNDN領域内にある、全体がV領域内にある場合もあるし、または、それはJ領域-NDN領域の境界もしくはV領域-NDN領域の境界またはそのような境界の両方を含む場合がある(図2Aに図示されている)。典型的には、そのような配列リードは、合成による配列決定反応においてポリメラーゼにより配列決定プライマー、例えば図2Aの(2302)および(2310)を伸長することによって生成される；例えば、Metzger, Nature Reviews Genetics, 11: 31-46 (2010); Fuller et al, Nature Biotechnology, 27: 1013-1023 (2009)。プライマー(2302)および(2310)の結合部位は、それらが配列リードの初期のアラインメントおよび解析のための出発点または投錨点を提供することができるよう、予め決定されている。1つの態様において、Cリードは、例えば図2Aおよび2Bに図示されているように、それがIgH鎖のDおよび/またはNDN領域を網羅し、かつ隣接するV領域の一部分を含むように位置決めされる。1つの局面において、V領域におけるVリードとCリードの重複は、これらのリード

を互いにアラインするのに使用される。他の態様においては、そのような配列リードのアラインメントは必要ではなく、Vリードがクロノタイプの特定のV領域を同定するのに十分に長いというだけであり得る。この後者の局面は、図2Bに図示されている。配列リード(2330)は、V領域を同定するのに使用され、これは別の配列リードと重複しているまたはしておらず、そして別の配列リード(2332)は、NDN領域にかかるものであり、その配列を決定するのに使用される。配列リード(2332)の、V領域へと延びる部分(2334)は、配列リード(2332)の配列情報を配列リード(2330)のそれと関連付け、クロノタイプを決定するのに使用される。いくつかの配列決定法、例えばSolexa配列決定法のような塩基単位(base-by-base)のアプローチでは、解析における配列決定サイクルの数を最小限にすることによって、配列決定に要する時間および試薬の費用が削減される。任意で、図2Aに示されるように、アンプリコン(2300)は、異なる生物学的試料、例えば異なる患者に由来するクロノタイプを区別するための試料タグ(2312)を含むように作製される。試料タグ(2312)は、プライマーをプライマー結合領域(2316)にアニールさせ、それを伸長させて(2314)、タグ(2312)にかけて配列リードを作成し、そこから試料タグ(2312)をデコードすることによって同定され得る。

10

#### 【0042】

本発明の1つの局面において、クロノタイプの配列は、1つまたは複数の配列リードからの情報を、例えば選択された鎖のV(D)J領域に沿って組み合わせることにより決定され得る。別の局面において、クロノタイプの配列は、複数の配列リードからの情報を組み合わせることにより決定される。このような複数の配列リードは、センス鎖に沿う1つまたは複数の配列リード(すなわち、「順方向」配列リード)およびその相補鎖に沿う1つまたは複数の配列リード(すなわち、「逆方向」配列リード)を含み得る。同じ鎖に沿って複数の配列リードを生成する場合は、最初に、該配列リードの異なる位置に対して選択されたプライマーを用いて試料分子を増幅することにより、別々の鋳型が生成される。このコンセプトは、図3Aに図示されており、そこではプライマー(3404、3406および3408)が、1回の反応でアンプリコン(それぞれ3410、3412および3414)を生成するのに使用されている。このような増幅は、同じ反応において実施してもよく、または別々の反応において実施してもよい。1つの局面において、PCRが使用される場合には必ず、別々の鋳型を生成するために別々の増幅反応が使用され、これらはその後組み合わせられ、同じ鎖に沿う複数の配列リードを生成するのに使用される。この後者のアプローチは、複数の鋳型の均等増幅を実現するためにプライマー濃度(および/またはその他の反応パラメータ)のバランスをとる必要がない点で好ましい(本明細書において、「バランスのとれた増幅」または「非バイアス増幅」と称されることがある)。別々の反応による鋳型の生成については、図3B~3Cに図示されている。その中で、IgHを含む試料(3400)が3等分され(3470、3472および3474)、これらがJ領域プライマー(3401)およびV領域プライマー(それぞれ3404、3406および3408)を用いる別々のPCRに添加され、アンプリコン(それぞれ3420、3422および3424)が作製されている。後者のアンプリコンはその後、P5およびP7プライマーを用いる二次PCR(3480)において組み合わせられ(3478)、ブリッジPCRおよびIllumina GAシーケンサーまたは同等の機器における配列決定のための鋳型(3482)が調製される。

20

30

#### 【0043】

本発明の配列リードは、様々な長さを有するものであり、それは一部、使用される配列決定技法に依存する。例えば、いくつかの技法では、その実施の中でいくつかのトレードオフ、例えば、(i)鋳型あたりの配列リードの数と長さおよび(ii)配列決定作業の費用と時間、が発生し得る。1つの態様において、配列リードは、20から3400ヌクレオチドの範囲であり；別の態様において、配列リードは、30から200ヌクレオチドの範囲であり；さらに別の態様において、配列リードは、30から120ヌクレオチドの範囲である。1つの態様においては、1つから4つの配列リードが、各クロノタイプの配列の決定のために生成され；別の態様においては、2つから4つの配列リードが、各クロノタイプの配列の決定のために生成され；そして別の態様においては、2つから3つの配列リードが、各クロノタイプの配列の決定のために生成される。上記の態様において、示されている数は、異なる個

40

50



体由来の試料を同定するのに使用される配列リードを除いたものである。以下に記載される態様において使用される様々な配列リードの長さもまた、そのリードによって取り込むことが求められる情報により変化し得；例えば、配列リードの出発位置および長さは、NDN領域の長さおよびそのヌクレオチド配列を提供するよう設計され得；したがって全NDN領域に及ぶ配列リードが選択される。他の局面においては、Dおよび/またはNDN領域を組み合わせ（しかし別々にではなく）含む1つまたは複数の配列リードで十分である。

#### 【0044】

本発明の別の局面において、クロノタイプの配列は、一部には、配列リードを1つまたは複数のV領域参照配列および1つまたは複数のJ領域参照配列とアラインすることにより、および、一部には、例えば、高可変性のNDN領域に関しては、参照配列とのアラインメントを用いない塩基決定により、決定される。様々なアラインメントアルゴリズムが、配列リードおよび参照配列に適用され得る。例えば、アラインメント法を選択する手引きは、参照により組み入れられるBatzoglou, Briefings in Bioinformatics, 6: 6-22 (2005)において利用可能である。1つの局面において、（上述のように）VリードまたはCリードがVおよびJ領域参照配列に対してアラインされる場合には必ず、ツリー検索アルゴリズムが使用される；例えば、Gusfield（上記で引用）およびCormen et al, Introduction to Algorithms, Third Edition (The MIT Press, 2009)に概説されている。

#### 【0045】

配列リードからのIgHクロノタイプの構築は、少なくとも2つの要因によって特徴づけられる：i) アラインメントをより困難にしている体細胞変異の存在、およびii) NDN領域が大きいこと。本発明の1つの局面において、この問題は、Vリードを生成するために、V領域に沿う異なる位置に配置される複数のプライマーセットを使用することによって、好ましくは、プライマー結合部位が重複せず間隔をあけて配置され、そして少なくとも1つのプライマー結合部位がNDN領域に隣接するように、例えば、1つの態様ではV-NDN接合部から5から50塩基、または別の態様ではV-NDN接合部から10から50塩基となるよう、複数のプライマーセットを使用することによって解消される。複数のプライマーセットの冗長度は、体細胞変異による影響を受ける結合部位を有する1つまたは2つのプライマーの不具合によるクロノタイプの検出の失敗のリスクを最小限にする。さらに、NDN領域に隣接するプライマー結合部位が少なくとも1つ存在することにより、VリードがCリードと重複する可能性が高くなり、したがってCリードの長さが効果的に伸長される。これにより、すべてのサイズのNDN領域を網羅しかつVおよびJ領域の実質的に全体をそのNDN領域の両方の側にマッピングすることもできる連続配列を生成することが可能となる。そのようなスキームを実施する態様は、図3Aおよび3Dに図示されている。図3Aにおいて、IgH鎖を含む試料（3400）は、単一セットのJ領域プライマー（3401）および複数（示されているのは3）セットのV領域（3402）プライマー（3404、3406、3408）を用いて鎖を増幅し、すべてが同じNDN領域を含みかつV領域（3402）の段階的に大きくなる部分（3411、3413、3415）を含む異なる長さを有する複数の入れ子アンプリコン（例えば3410、3412、3414）を作製することにより、各鎖につき複数のアンプリコンを生成することによって配列決定される。入れ子セットのメンバーは、それら各々のNDN、Jおよび/またはC領域の同一性（実質的同一性）を確認することにより、配列決定後にひとつにグループ化され得、それによって、リード長および/または配列決定品質が限定される他の配列決定プラットフォームの場合よりも長いV(D)Jセグメントの再構築が実現される。1つの態様において、複数のプライマーセットは、2から5の範囲の数であり得る。別の態様において、複数のプライマーは、2～3であり；さらに別の態様では複数のプライマーは3である。複数のプライマーの濃度および位置は、様々な変化し得る。V領域プライマーの濃度は、同じである場合もそうでない場合もある。1つの態様において、NDN領域に最も近いプライマーは、例えばNDN領域を含むアンプリコンが、得られるアンプリコンにおいて表現されることを確実にするために、その複数の他のプライマーよりも高い濃度を有する。複数の3つのプライマーを使用する特定の態様において、60：20：20の濃度比が用いられる。NDN領域（3444）に隣接する1つまたは複数のプライマー（例えば

10

20

30

40

50

、図3Dでは3435および3437)は、J領域プライマー(3432)によって生成される配列リード(3442)と重複する1つまたは複数の配列リード(例えば、3434および3436)を生成するのに使用され得、それによって重複領域(3440)における塩基コールの質が改善される。複数のプライマーからの配列リードは、隣接する下流のプライマー結合部位および/または隣接する下流の配列リードと重複している場合もそうでない場合もある。1つの態様においては、NDN領域に近接する配列リード(例えば、3436および3438)が、クロノタイプに関連する特定のV領域を同定するのに使用され得る。そのような複数のプライマーは、プライマー結合部位の1つが免疫グロブリンの発達時に超変異している場合に増幅が不完全または不成功となる可能性を低下させる。それはまた、V領域の超変異により導入された多様性がクロノタイプ配列に取り込まれる可能性を高める。二次PCRは、配列決定のための入れ子アンプリコンを調製するために、例えば図示されているようなP5(3401)およびP7(3404、3406、3408)プライマーを用いる増幅によりアンプリコン(3420、3422および3424)を作製することによって実施され得、それらは固相表面上に単一分子として配分され得、さらにブリッジPCRまたは同様の技法により増幅される。

#### 【0046】

体細胞超変異。1つの態様において、体細胞超変異を起こしたIgHベースのクロノタイプは、以下のようにして決定される。体細胞変異は、(関連セグメントの、通常はV、JまたはCの)対応する参照配列の塩基と異なり、かつ統計的に有意な数のリードに存在する、配列決定された塩基と定義される。1つの態様においては、Cリードが、マッピングされたJセグメントに関する体細胞変異を発見するのに使用され得、同様に、Vセグメントに関してはVリードが使用され得る。JまたはVセグメントに直接マッピングされるかまたはNDN境界までのクロノタイプ伸長物の内側であるかのいずれかのCおよびVリードのみが使用される。このようにして、NDN領域は回避され、以前にクロノタイプの決定に使用された同じ「配列情報」が、変異の発見に使用されることはない(実際は異なる組換えNDN領域であるにすぎないのに誤って変異ヌクレオチドとして分類されるのを回避するため)。セグメントタイプごとに、マッピングされたセグメント(優性のアレル)がスキャホールドとして使用され、リードのマッピング段階でこのアレルにマッピングされたすべてのリードが考慮される。少なくとも1つのリードがマッピングされている参照配列の各位置が、体細胞変異について解析される。1つの態様において、非参照塩基を有効な変異として受諾する基準は、以下のものを含む:1)所定の変異塩基を有する少なくともN個のリード、2)少なくとも所定の分数 $N/M$ のリード(Mはこの塩基位置にマッピングされたリードの総数である);および3)2項分布、変異塩基におけるN個のリードの平均Qスコアおよび非変異塩基を有するリードの数(M-N)に基づく統計的な切り捨て。好ましくは、上記のパラメータは、クロノタイプ単位での変異の誤発見率が1000中1未満、より好ましくは10000中1未満となるように選択される。

#### 【0047】

PCRのエラーは、PCRの早い段階のサイクルで変異したいくつかの塩基に集中すると考えられる。配列決定のエラーは、全体としてランダムであるにもかかわらず、そのエラーがある程度の体系的バイアスを有している可能性がある場合、多くの塩基に分散すると考えられる。いくつかの塩基は、およそ5%(平均の5倍)の高い比率の配列決定エラーを有すると推測される。これらの推測の下では、配列決定エラーが支配的なエラーのタイプとなる。PCRのエラーと、高度に関連するクロノタイプの発生とを区別することは、解析において重要となる。2つまたはそれ以上の高度に関連するクロノタイプが存在すると判定することには生物学的意義があるため、そのようなコールを生成するために保存的なアプローチが採用される。高い信頼性(およそ99.9%)で2以上のクロノタイプが存在することを確認するために、少数派のクロノタイプの十分な検出が考慮される。100コピー/1,000,000で存在するクロノタイプの例では、少数派のバリエーションが、独立したクロノタイプとして指定されるよう、14回またはそれ以上検出される。同様に、1,000コピー/1,000,000で存在するクロノタイプにおいては、少数派のバリエーションが、独立したクロノタイプとして指定されるよう、74回またはそれ以上検出され得る。このアルゴリズムは、各々の配列決

10

20

30

40

50

定された塩基により得られる塩基品質スコアを使用することによって向上させることができる。品質スコアとエラー率との間の関係が上記のように確認されれば、すべての塩基に対して保存的な5%のエラー率を使用することに代えて、品質スコアを、独立したクロノタイプをコールするために存在する必要のあるリード数を決定するのに使用することができる。すべてのリードにおける特定塩基の品質スコアの中央値を使用することができる、またはより厳密に言えば、エラーである可能性は、各リードにおける特定塩基の品質スコアに基づき計算することができ、そしてその確率は、その塩基の可能性のある配列決定エラーの数を概算するために組み合わせることができる（独立と仮定して）。結果的に、配列決定エラー仮説を拒絶する閾値は、異なる品質スコアを有する異なる塩基ごとに異なる。例えば、1,000,000あたり1,000コピー存在するクロノタイプでは、少数派のバリエーションは、エラーの確率が0.01および0.05として、それぞれ22および74回検出される場合、独立と指定される。

10

#### 【0048】

配列決定エラーが存在する場合、真の各クロノタイプは、その配列に関して異なる数のエラーを有するリードの「クラウド」に取り囲まれている。配列空間においてクロノタイプからの距離が増大するにつれて、配列決定エラーの「クラウド」は密度を低下させる。配列リードをクロノタイプに変換するために、種々のアルゴリズムが利用できる。1つの局面において、配列リードの合体（すなわち、1つまたは複数の配列決定エラーを有すると判定された候補クロノタイプを統合すること）は、少なくとも3つの因子に依存する：比較されるクロノタイプの各々について得られる配列の数；相違する塩基の数；および不一致の位置における配列決定の品質スコア。予想エラー率およびエラーの二項分布に基づく尤度比が構築され、評価され得る。例えば、一方が150個のリードを有し、もう一方が2個のリードを有し、それらの間には配列決定品質の低い領域内に1つの相違がある、2つのクロノタイプは、それらが配列決定エラーにより生成された可能性があるため、合体される可能性がある。それに対して、一方が100個のリードを有し、もう一方が50個のリードを有し、それらの間に2つの相違がある2つのクロノタイプは、それらが配列決定エラーにより生成された可能性は低いと見なされるため、合体されない。本発明の1つの態様において、以下に記載されるアルゴリズムが、配列リードからクロノタイプを決定するために使用され得る。本発明の1つの局面において、配列リードは最初に候補クロノタイプに変換される。そのような変換は、使用される配列決定プラットフォームに依存する。高いQスコアの長い配列リードを生成するプラットフォームでは、配列リードまたはその一部分は候補クロノタイプとして直接採用され得る。より低いQスコアのより短い配列リードを生成するプラットフォームでは、関連配列リードのセットを候補クロノタイプに変換するために、いくつかのアラインメントおよび組み立て工程が必要とされ得る。例えば、Solexaベースのプラットフォームでは、いくつかの態様において、候補クロノタイプは、上記のように、例えば10個またはそれ以上といった複数のクラスターに由来する対をなしたリードの収集物から生成される。

20

30

#### 【0049】

各候補クロノタイプを取り囲んでいる配列リードのクラウドは、2項分布および一塩基エラーの確率についての単純モデルを用いてモデル化することができる。この後者のエラーモデルは、VおよびJセグメントのマッピングからまたはクロノタイプ発見アルゴリズム自体から自己無撞着（self-consistency）または収斂（convergence）を通じて推測することができる。モデルは、（配列Xに関して）リードカウントC2およびE個のエラーを有する所定の「クラウド」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の一部分である確率に関して、Xが配列空間のこの領域における唯一の真のクロノタイプであるというヌル（null）モデルの下で、構築される。判定は、パラメータC1、C2およびEにしたがい配列YをクロノタイプXに合体させるかさせないかに関して為される。任意の所定のC1およびEについて、配列Yを合体させるとの判定のための最大値C2が事前に算出される。C2の最大値は、YがクロノタイプXの一部分であるというヌル仮説の下でYを合体させない確率が、配列Xの近隣にエラーEを有するすべての可能性のある配列Yにわたって積

40

50

分した後に一定値P未満となるように、選択される。値Pは、アルゴリズムの挙動を制御し、合体を多少寛容的にする。

【0050】

配列Yが、そのリードカウントがクロノタイプXへの合体のための閾値C2を上回っているために、クロノタイプXに合体されない場合、それは、別々のクロノタイプをシード (seed) するための候補となる。そのような原則を実施するアルゴリズムは、この(Xに非依存であるとみなされた) 配列Yに「より近い」任意のその他の配列Y2、Y3等がXに集約されないようにする。この「近さ (nearness)」のコンセプトは、YおよびXに関するエラーカウントとXおよびYの絶対リードカウントの両方を含む、すなわち、それは、クロノタイプXの周囲のエラー配列のクラウドに関する上記のモデルと同じ様式でモデル化される。このようにして、「クラウド」配列は、それらが2以上のクロノタイプの「近く」にあった場合に、それらの正確なクロノタイプに適切に帰属化させることができる。

10

【0051】

1つの態様において、アルゴリズムは、最大のリードカウントを有する配列Xから開始することにより、トップダウン式に進行する。この配列は、第1のクロノタイプをシードする。隣接する配列は、それらのカウントが事前に算出された閾値 (上記参照) を下回る場合にこのクロノタイプに合体されるか、またはそれらが閾値を上回るもしくは合体されなかった別の配列に「近い」場合に放置されるかのいずれかである。最大エラーカウント内のすべての隣接する配列を検索した後、リードをクロノタイプXに合体させるプロセスが終了する。そのリードおよびそれに合体されたすべてのリードが報告され、他のクロノタイプを生成するのに利用できるリードのリストから除去される。続いて、次に最大のリードカウントを有する配列に移行する。隣接するリードが上記のようにしてこのクロノタイプに合体され、そしてこのプロセスは、所定の閾値を上回るリードカウントを有する配列がそれ以上存在しなくなるまで、例えば1カウント超の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、継続される。

20

【0052】

上述したように、上記のアルゴリズムの別の態様においては、候補配列Yを既存のクロノタイプXに合体させるかどうかを判定するために、関連する配列リードの品質スコアを考慮するさらなる試験が追加され得る。配列YとXが相違する場合、配列Yについての平均品質スコア (配列Yを有するすべてのリードの平均) が、決定される。その平均スコアが既定の値を上回る場合、その差が合体されるべきでない真に異なるクロノタイプを示している可能性が高く、平均スコアがそのような既定の値を下回る場合、配列Yは配列決定エラーによるものでありしたがってXに合体されるべきである可能性が高い。

30

【0053】

いくつかの特定の態様例を参照して本発明を説明してきたが、本発明の精神および範囲から逸脱することなく、それらに対して多くの変更がなされ得ることを、当業者は認識するであろう。本発明は、上記のものに加えて、様々なセンサー実現およびその他の主題に適用可能である。

【0054】

定義

本明細書において他に具体的に規定されない限り、本明細書において用いられる核酸化学、生化学、遺伝学、および分子生物学の用語および記号は、当分野における標準的な論文および教科書、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, 6<sup>th</sup> edition (Saunders, 2007)の用語および記号に従っている。

40

【0055】

「アラインすること」とは、配列リードなどの試験配列を1つまたは複数の参照配列と比較して、どの参照配列が、または参照配列のどの部分が、何らかの配列距離尺度に基づ

50

いて最も近いのかを判定する方法を意味する。ヌクレオチド配列をアラインする例示的な方法は、Smith Watermanアルゴリズムである。距離尺度には、ハミング距離、レーヴェンシュタイン距離等が含まれ得る。距離尺度は、比較される配列のヌクレオチドの品質値に関連した成分を含み得る。

#### 【0056】

「アンプリコン」とは、ポリヌクレオチド増幅反応の産物；すなわち、一本鎖または二本鎖であってよく、1つまたは複数の出発配列から複製されるポリヌクレオチドのクローン集団を意味する。1つもしくは複数の出発配列は、同じ配列の1つもしくは複数のコピーであってもよいし、またはそれらは異なる配列の混合物であってもよい。好ましくは、アンプリコンは、単一の出発配列の増幅によって形成される。アンプリコンは、その反応産物が1つまたは複数の出発核酸または標的核酸の複製物を含む、様々な増幅反応によって作製され得る。1つの局面において、アンプリコンを作製する増幅反応は、ヌクレオチドまたはオリゴヌクレオチドのいずれかである反応物の塩基対形成が、反応産物の創出に必要な相補体を鋳型ポリヌクレオチドにおいて有するという点で、「鋳型駆動型」である。1つの局面において、鋳型駆動型反応は、核酸ポリメラーゼによるプライマー伸長、または核酸リガーゼによるオリゴヌクレオチド連結である。このような反応には、参照により本明細書に組み入れられる以下の参考文献、Mullis et al、米国特許第4,683,195号；第4,965,188号；第4,683,202号；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4-262799(ローリングサークル増幅)などに開示されているポリメラーゼ鎖反応(PCR)、線状ポリメラーゼ反応、核酸配列ベースの増幅(NASBA)、ローリングサークル増幅などが含まれるが、これらに限定されない。1つの局面において、本発明のアンプリコンはPCRによって作製される。増幅反応の進行に伴った反応産物の測定を可能にする検出化学が利用できるのであれば、増幅反応は「リアルタイム」増幅であってよく、例えば、以下に記載される「リアルタイムPCR」、またはLeon et al、Nucleic Acids Research, 26: 2150-2155 (1988)、および同様の参考文献に記載されているような「リアルタイムNASBA」であってよい。本明細書において用いられる「増幅すること」という用語は、増幅反応を行うことを意味する。「反応混合物」とは、反応を行うために必要な反応物をすべて含む溶液を意味し、この反応物には、反応中にpHを選択されたレベルに維持するための緩衝剤、塩、補因子、スカベンジャーなどが含まれ得るが、これらに限定されない。

#### 【0057】

本明細書において用いられる「クローン性」とは、あるレパートリーのクロノタイプ間のクロノタイプ存在量の分布が1種または数種のクロノタイプにどれだけ偏っているかの尺度を意味する。大まかに言うと、クローン性はクロノタイプ多様性の逆の尺度である。多くの尺度または統計学が、本発明によるクローン性尺度に対して使用され得る、種と存在量の関係を説明する生態学、例えば、Chapters 17 & 18, in Pielou, An Introduction to Mathematical Ecology, (Wiley-Interscience, 1969)から利用できる。1つの局面において、本発明と共に用いられるクローン性尺度はクロノタイププロファイル(すなわち、検出される別個のクロノタイプの数およびそれらの存在量)と関係しており、したがって、クロノタイププロファイルを測定した後、そこからクローン性を計算して、1つの数字をもたらすことができる。1つのクローン性尺度はシンプソンの尺度であり、これは単純に、ランダムに選び出された2つのクロノタイプが同じである確率である。その他のクローン性尺度には、情報ベースの尺度、およびPielou(上記で引用)に開示されているマッキントッシュの多様性指数が含まれる。

#### 【0058】

「クロノタイプ」とは、T細胞受容体(TCR)もしくはB細胞受容体(BCR)またはその一部分をコードする、T細胞またはB細胞の組換えヌクレオチド配列を意味する。1つの局面において、ある個体のあるリンパ球集団の別個のクロノタイプすべての収集物は、そのような

集団の1つのレパートリーである；例えば、Arstila et al, Science, 286: 958-961 (1999)；Yassai et al, Immunogenetics, 61: 493-502 (2009)；Kedzierska et al, Mol. Immunol., 45(3): 607-618 (2008)など。本発明で用いられる「クロノタイププロファイル」または「レパートリープロファイル」とは、レパートリーのクロノタイプおよびそれらの相対存在量の実質的にすべてを含めた、T細胞および/またはB細胞の試料(そのような細胞を含む末梢血試料など)のクロノタイプの集計である。「クロノタイププロファイル」、「レパートリープロファイル」、および「レパートリー」は、本発明において互換的に用いられる。(すなわち、以下にさらに詳述する「レパートリー」という用語は、リンパ球の試料から測定されるレパートリーを意味する)。本発明の1つの局面において、クロノタイプは、免疫グロブリン重鎖(IgH)またはTCR 鎖の一部を含む。本発明の他の局面において、クロノタイプは、免疫グロブリン軽鎖もしくはTCR 鎖またはそれらの一部分などのその他の組換え分子に基づき得る。

10

20

30

40

50

#### 【0059】

「合体させること」とは、配列相違のある2つの候補クロノタイプを、そのような相違が実験的エラーまたは測定エラーによるものであり、真の生物学的相違によるものではないと判定することにより、同じものとして扱うことを意味する。1つの局面では、より高頻度の候補クロノタイプの配列をより低頻度の候補クロノタイプの配列と比較し、所定の基準が満たされる場合に、より低頻度の候補クロノタイプの数をより高頻度の候補クロノタイプの数に加え、より低頻度の候補クロノタイプをその後無視する。すなわち、より低頻度の候補クロノタイプと関連したリードカウントは、より高頻度の候補クロノタイプのリードカウントに加えられる。

#### 【0060】

「相補性決定領域」(CDR)とは、免疫グロブリン(すなわち、抗体)またはT細胞受容体の領域を意味するものであり、この領域において分子は抗原の高次構造を補完し、それによって該分子の特異性を決定し、特定の抗原と接触する。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を有する：CDR1およびCDR2は可変(V)ドメイン中に見出され、CDR3は、Vの一部、多様部(D)(重鎖のみ)および結合部(J)のすべて、ならびに定常(C)ドメインの一部を含む。

#### 【0061】

「免疫活性化」とは、(抗原特異的リンパ球が抗原に結合する)抗原認識期に続く適応免疫応答期を意味し、リンパ球の増殖およびエフェクター細胞への分化によって特徴づけられる；例えば、Abbas et al, Cellular and Molecular Immunology, Fourth Edition, (W.B. Saunders Company, 2000)。

#### 【0062】

「リンパ系新生物」とは、悪性または非悪性であってよいリンパ球の異常な増殖を意味する。リンパ系癌とは、悪性のリンパ系新生物である。リンパ系新生物は、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)、ヘアリー細胞白血病、リンパ腫、多発性骨髄腫、移植後のリンパ球増殖障害、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)、T細胞リンパ腫などを含むがこれらに限定されないリンパ球増殖障害の結果であるか、またはそれに付随する；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)。

#### 【0063】

参照配列と別の配列(「比較配列」と)の比較に関して用いられる「%相同」、「%同一」、または同様の用語は、この2つの配列の間の最適なアラインメントにおいて、比較配列が、示されるパーセンテージに等しいサブユニット位置の数において参照配列と同一であることを意味し、このサブユニットは、ポリヌクレオチド比較に関してはヌクレオチドであり、またはポリペプチド比較に関してはアミノ酸である。本明細書において用いられる、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おいて使用されるギャップの数を最小にするアラインメント

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)などによって記載されているようなアルゴリズムの商用の実行によって決定され得る。アラインメントを構築し、同一性のパーセンテージまたは類似性の他の尺度を算出するための、当技術分野におけるその他のソフトウェアパッケージには、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」プログラムが含まれる。言い換えれば、例えば、参照ヌクレオチド配列と少なくとも95パーセント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は、参照配列中のヌクレオチドの5パーセントまでが欠失されるか、もしくは別のヌクレオチドで置換されてよく、または参照配列中の全ヌクレオチド数の5パーセントまでのヌクレオチド数が参照配列中に挿入されてよい。

10

#### 【0064】

「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配列のインピット増幅のための反応を意味する。言い換えると、PCRは、プライマー結合部位によって挟まれた標的核酸の複数のコピーまたは複製物を作製するための反応であり、このような反応は以下の工程の1回または複数回の反復を含む：(i) 標的核酸を変性させる工程、(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工程、および(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工程。通常、反応は、サーマルサイクラー装置において、各工程に最適化された異なる温度の間で繰り返される。特定の温度、各工程における持続時間、および工程間の変化速度は、例えば参考文献、McPherson et al, editors, PCR: A Practical Approach and PCR2: A Practical Approach (IRL Press, Oxford、それぞれ1991および1995)によって例示される、当業者に周知の多くの要因に依存する。例えば、Taq DNAポリメラーゼを用いる従来のPCRでは、 $>90$  の温度で二本鎖標的核酸が変性され得、 $50 \sim 75$  の範囲の温度でプライマーがアニーリングされ得、そして $72 \sim 78$  の範囲の温度でプライマーが伸長され得る。「PCR」という用語は、RT-PCR、リアルタイムPCR、ネステッドPCR、定量的PCR、マルチプレックスPCRなどを含むがこれらに限定されない、該反応の派生形態を含む。反応容量は、数百ナノリットル、例えば200 nL～数百 $\mu$ L、例えば200 $\mu$ L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へと変換する逆転写反応が先行し、次いで増幅が行われるPCRを意味する；例えば、参照により本明細書に組み入れられるTecott et al、米国特許第5,168,038号。「リアルタイムPCR」とは、反応の進行と共に反応産物、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。リアルタイムPCRの多くの形態が存在し、主に反応産物のモニタリングに用いら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Gelfand et al、米国特許第5,210,015号(「taqman」)；Wittwer et al、米国特許第6,174,670号および第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のための検出化学は、同様に参照により本明細書に組み入れられるMackay et al, Nucleic Acids Research, 30: 1292-1305 (2002)において概説されている。「ネステッドPCR」とは、第1PCRのアンプリコンが、プライマーの新たなセットを用いる第2PCRのための試料となる二段階PCRを意味し、そのセットのうちの少なくとも一方は第1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する、本明細書において用いられる「初期プライマー」とは、第1アンプリコンを生成するために用いられるプライマーを意味し、「二次プライマー」とは、第2または入れ子アンプリコンを生成するために用いられる1つまたは複数のプライマーを意味する。「マルチプレックスPCR」とは、複数の標的配列(または単一の標的配列および1つまたは複数の参照配列)が同じ反応混合物中で同時に行われる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、増幅される各配列について、プライマーの別個のセットが用いられる。「定量的PCR」とは、試料または標本中の1つまたは複数の特定の標的配列の存在量を測定するために設計されたPCRを意味する。定量的PCRは、このような標

20

30

40

50

的配列の絶対的な定量および相対的な定量の両方を含む。定量的な測定は、標的配列と別々にまたは共にアッセイされ得る1つまたは複数の参照配列または内部標準を用いてなされる。参照配列は、試料または標本にとって内因性であっても外因性であってもよく、後者の場合には、1つまたは複数の競合鋳型を含み得る。典型的な内因性の参照配列には、以下の遺伝子の転写物のセグメントが含まれる：アクチン、GAPDH、 $\beta_2$ ミクログロブリン、リボソームRNAなど。定量的PCRのための技法は、参照により組み入れられる以下の参考文献において例証されるように、当業者に周知である：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 et al, Nucleic Acids Research, 17: 9437-9446 (1989)など。

#### 【0065】

「プライマー」とは、ポリヌクレオチド鋳型との二重鎖の形成に際して、核酸合成の開始点として働くことができ、伸長された二重鎖が形成されるように、鋳型に沿ってその3'末端から伸長され得る、天然または合成のいずれか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ポリメラーゼまたはRNAポリメラーゼなどの核酸ポリメラーゼを用いて行われる。伸長過程において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、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、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、単一のプライマーを用いる線状増幅反応、または2種もしくはそれ以上のプライマーを用いるポリメラーゼ連鎖反応において用いられる。特定の適用に関してプライマーの長さおよび配列を選択するための指針は、参照により組み入れられる以下の参考文献によって明らかのように、当業者に周知である：Dieffenbach, editor, PCR Primer: A Laboratory Manual, 2<sup>nd</sup> Edition (Cold Spring Harbor Press, New York, 2003)。

#### 【0066】

「品質スコア」とは、特定の配列位置における塩基の割り当てが正しいという確率についての尺度である。異なる配列決定化学、検出システム、塩基コールアルゴリズムなどの結果としてコールされる塩基に関するような特定の状況について、品質スコアを算出するための様々な方法が当業者に周知である。一般的に、品質スコア値は、正しい塩基コールの確率に単調に関連している。例えば、10という品質スコアまたはQは、塩基が正しくコールされる可能性が90パーセントあることを意味し得、20というQは、塩基が正しくコールされる可能性が99パーセントあることを意味し得るなどである。いくつかの配列決定プラットフォーム、特に合成による配列決定化学を用いる配列決定プラットフォームでは、平均品質スコアは配列リード長に応じて低下し、その結果、配列リードの始めの品質スコアは配列リードの終わりの品質スコアよりも高く、そのような低下は、不完全な伸長、繰り返し伸長、鋳型の減少、ポリメラーゼの減少、キャップ形成の障害、脱保護の障害などのような現象に起因する。

#### 【0067】

本明細書において用いられる「レパートリー」または「免疫レパートリー」とは、ある個体のあるリンパ球集団中の、B細胞受容体(BCR)またはそれらの断片それぞれをコードする別個の組換えヌクレオチド配列のセットを意味し、該セットのヌクレオチド配列は、該集団のリンパ球の実質的にすべてについて、別個の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。1つの局面において、レパートリーが決定されるリンパ球集団は、1つまたは複数の血液試料などの1つまたは複数の組織試料から取得される。レパートリーのメンバーヌクレオチド配列は、本明細書において「クロノタイプ」と称される。1つの局面において、レパートリーのクロノタイプは、BCRの発達中に体細胞組換えを起こしたB細胞集団に共通する核酸の任意のセグメントを含み、これには正常なまたは異常な(例えば、癌に関連する)その前駆体分子が含まれ、以下のもののうちのいずれかが含まれるがこれらに限定されない：免疫グロブリン重鎖(IgH)またはそのサブセット(例えば、Ig



H可変領域、CDR3領域など)、不完全なIgH分子、免疫グロブリン軽鎖またはそのサブセット(例えば、可変領域、CDR領域など)、CDR(BCRのCDR1、CDR2、もしくはCDR3、またはそのようなCDRの組み合わせを含む)、BCRのV(D)J領域、IgH可変領域の超変異領域など。1つの局面において、レパートリーのクロノタイプを規定する核酸セグメントは、それらの多様性(すなわち、セット中の別個の核酸配列の数)が十分に高く、結果として個体中の実質的にすべてのB細胞またはそのクローンがそのようなレパートリーの特有の核酸配列を保持するように、選択される。すなわち、本発明に従って、実施者は、B細胞の集団の完全な多様性を反映しない、BCRをコードする組換え核酸の特定のセグメントまたは領域を、クロノタイプを規定するために選択してもよい;しかしながら、好ましくは、クロノタイプは、それらの由来元であるB細胞の集団の多様性を反映するように規定される。すなわち、好ましくは、試料のそれぞれ異なるクローンは異なるクロノタイプを有する。(当然ながら、いくつかの適用においては、白血病またはリンパ腫患者由来の試料の場合のように、プロファイル内には1つまたは複数の特定のクロノタイプの複数のコピーが存在する)。本発明の他の局面において、レパートリーに相当するリンパ球集団は、循環B細胞、または細胞表面マーカーによって規定されるその他の亜集団などであってよい。そのような亜集団は、特定の組織、例えば骨髄もしくはリンパ節などから試料を取得することによって、または1つもしくは複数の細胞表面マーカー、サイズ、形態などに基づいて試料(末梢血など)から細胞を選別もしくは濃縮することによって獲得され得る。さらなる他の局面において、レパートリーに相当するリンパ球集団は、腫瘍組織、感染組織などの罹患組織に由来し得る。1つの態様において、ヒトIgH鎖またはその断片を含むレパートリーは、 $0.1 \times 10^6 \sim 1.8 \times 10^6$ 種の範囲、または $0.5 \times 10^6 \sim 1.5 \times 10^6$ 種の範囲、または $0.8 \times 10^6 \sim 1.2 \times 10^6$ 種の範囲内の数の別個のヌクレオチド配列を含む。特定の態様において、本発明のレパートリーは、IgH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するヌクレオチド配列のセットを含む。1つの局面において、本明細書において用いられる「実質的にすべての」とは、0.001パーセントもしくはそれ以上の相対存在量を有するすべてのセグメントを意味し;または別の局面において、本明細書において用いられる「実質的にすべての」とは、0.0001パーセントもしくはそれ以上の相対存在量を有するすべてのセグメントを意味する。別の特定の態様において、本発明のレパートリーは、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するヌクレオチド配列のセットを含む。別の態様において、本発明のレパートリーは、25~200ヌクレオチドの範囲の長さを有しかつIgH鎖のV、D、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。別の態様において、本発明のレパートリーは、別個のIgH鎖を発現するリンパ球の数と実質的に等しい多数の別個のヌクレオチド配列を含む。さらなる別の態様において、「実質的に等しい」とは、ヌクレオチド配列のレパートリーが、ある個体のある集団の、0.001パーセントまたはそれ以上の頻度で存在するすべてのリンパ球によって保持または発現されるIgHまたはその一部分をコードするヌクレオチド配列を、99パーセントの確率で含むことを意味する。さらなる別の態様において、「実質的に等しい」とは、ヌクレオチド配列のレパートリーが、0.0001パーセントまたはそれ以上の頻度で存在するすべてのリンパ球によって保持または発現されるIgHまたはその一部分をコードするヌクレオチド配列を、99パーセントの確率で含むことを意味する。前述の2文に記載されるクロノタイプのセットは、本明細書において、IgH配列の「完全なレパートリー」を表現すると見なされる場合がある。上記のように、クロノタイププロファイル(またはレパートリープロファイル)を測定または生成する場合には、そのようなプロファイルが、特定の適用に対してレパートリーのかなり正確な表現を提供するように、十分に多量のリンパ球の試料を得る。1つの局面において、特に1~10 mLの末梢血試料から得られる場合、 $10^5 \sim 10^7$ 個のリンパ球を含む試料が用いられる。

#### 【0068】

「配列リード」とは、配列決定技法によって生成された一連のデータまたはデータストリームから決定されたヌクレオチドの配列を意味し、この決定は、例えば、この技法と関連した塩基コーリングソフトウェア、例えばDNA配列決定プラットフォームの商業的供給

10

20

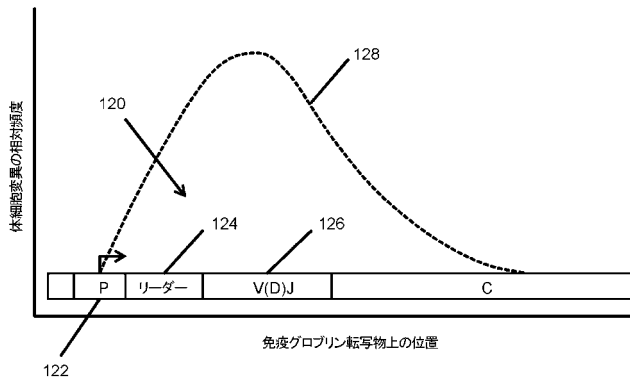
30

40

50

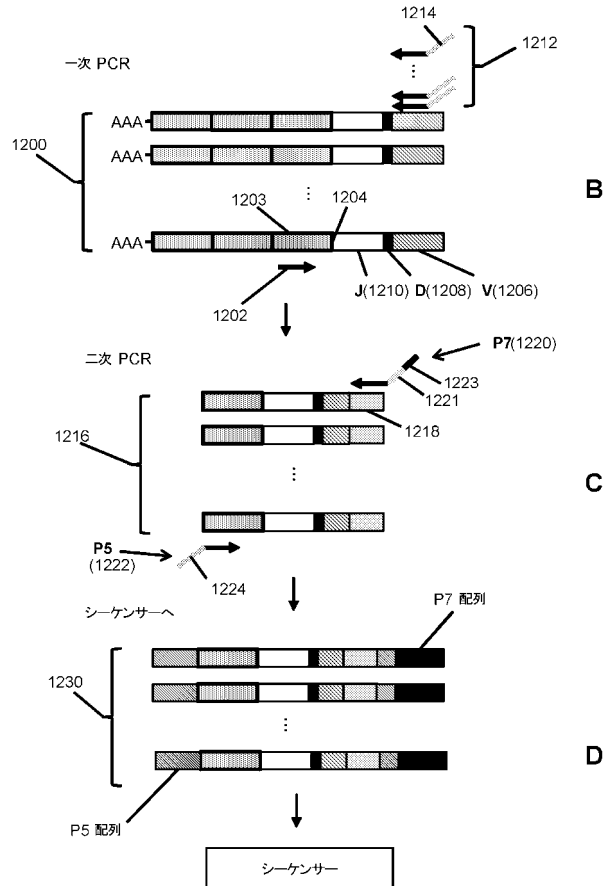
業者からの塩基コーリングソフトウェアを用いてなされる。配列リードは通常、配列内の各ヌクレオチドに関する品質スコアを含む。典型的には、配列リードは、例えばDNAポリメラーゼまたはDNAリガーゼを用いて、鋳型核酸に沿ってプライマーを伸長させることによって作られる。データは、このような伸長に付随する光学的、化学的(例えば、pH変化)、または電氣的シグナルなどのシグナルを記録することによって生成される。このような初期データが配列リードに変換される。

【図 1 - 1】



A

【図 1 - 2】

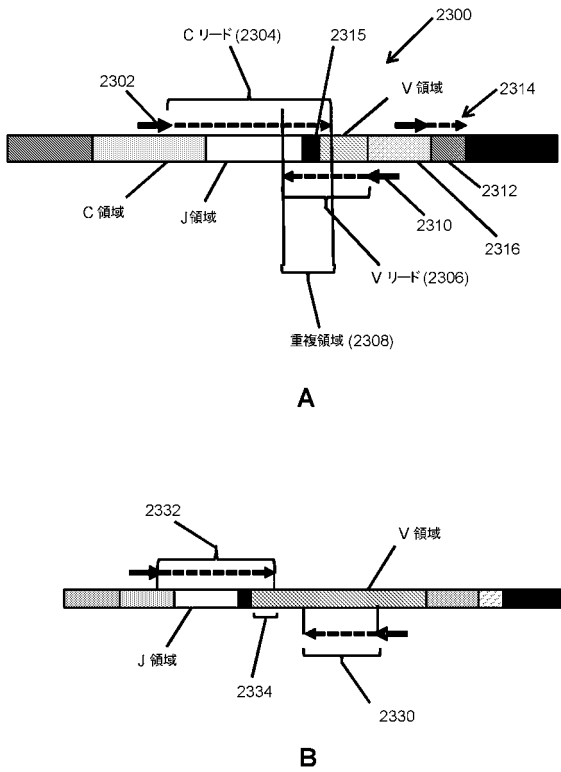


B

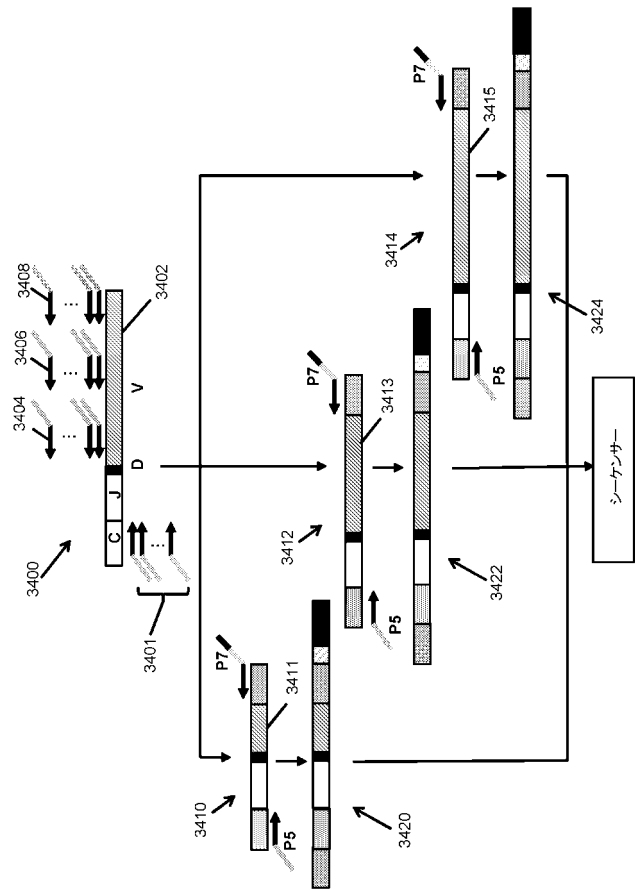
C

D

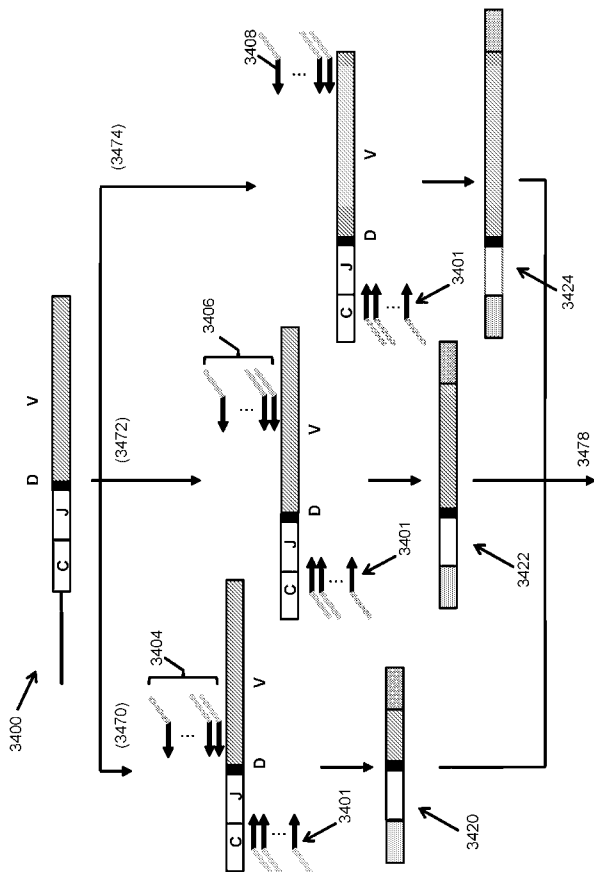
【図 2】



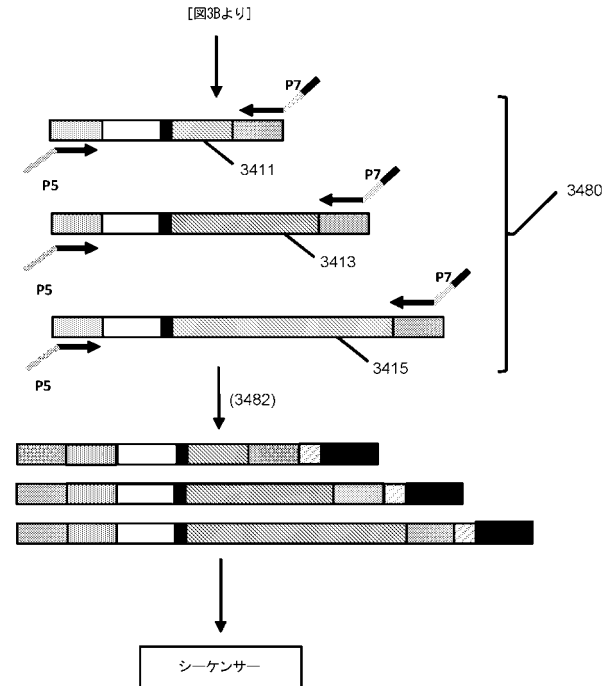
【図 3 A】



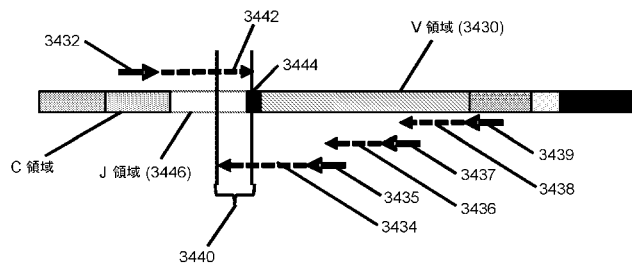
【図 3 B】



【図 3 C】



【図 3 D】



## 【国際調査報告】

61400650011



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 12/70674

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/68 (2013.01)

USPC - 435/6.11; 435/6.14

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, 435/6.14

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC: 435/6.12

(keyword limited; terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase; Google; PubMed

Search terms: B cells, clonotype, profile, clan, V read, C read, V, J, NDN, follicular lymphoma, diffuse large b cell lymphoma

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2011/0207134 A1 (FAHAM et al.) 25 August 2011 (25.08.2011) para [0017], [0020], [0065], [0117]-[0118], [0280], [0285]	1-7
Y	CARLOTTI et al., Transformation of follicular lymphoma to diffuse large B-cell lymphoma may occur by divergent evolution from a common progenitor cell or by direct evolution from the follicular lymphoma clone. Blood. 09 April 2009 (09.04.2009), Vol. 113, No. 15, pages 3553-3557. Especially pg 3553, para 1; pg 3555, para 5	1-7
A	DAVIES et al., Transformation of follicular lymphoma to diffuse large B-cell lymphoma proceeds by distinct oncogenic mechanisms. Br. J. Haematol. January 2007 (01.2007), Vol. 136, No. 2, pages 286-293	1-7

☐ 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

"Z" document member of the same patent family

Date of the actual completion of the international search

28 January 2013 (28.01.2013)

Date of mailing of the international search report

22 FEB 2013

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 QBP: 571-272-7774

30.10.2014

2]

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 12/70674

**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.: 8-14  
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:

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.:

**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.

## フロントページの続き

(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, 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, KM, 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

(74)代理人 100148699  
弁理士 佐藤 利光

(74)代理人 100128048  
弁理士 新見 浩一

(74)代理人 100129506  
弁理士 小林 智彦

(74)代理人 100130845  
弁理士 渡邊 伸一

(74)代理人 100114340  
弁理士 大関 雅人

(74)代理人 100114889  
弁理士 五十嵐 義弘

(74)代理人 100121072  
弁理士 川本 和弥

(72)発明者 ファハム マレク  
アメリカ合衆国 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  
4 0 0 スイート 3 0 1

F ターム(参考) 2G045 AA26 CA19 DA13  
4B024 AA12 CA02 CA09 CA11 HA12  
4B063 QA01 QA19 QQ43 QQ52 QQ58 QR55 QR62 QS25 QS34 QX02