

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号
特許第6556705号
(P6556705)

(45) 発行日 令和1年8月7日(2019. 8. 7)

(24) 登録日 令和1年7月19日(2019. 7. 19)

(51) Int.Cl.

F I

C 1 2 N 15/09 (2006. 01)

C 1 2 N 15/09 Z N A Z

C 1 2 Q 1/6813 (2018. 01)

C 1 2 Q 1/6813 Z

請求項の数 15 (全 43 頁)

(21) 出願番号	特願2016-519575 (P2016-519575)	(73) 特許権者	515343247
(86) (22) 出願日	平成26年6月9日(2014. 6. 9)		バイオナノ ジェノミクス、 インコーポ
(65) 公表番号	特表2016-523533 (P2016-523533A)		レイテッド
(43) 公表日	平成28年8月12日(2016. 8. 12)		アメリカ合衆国 9 2 1 2 1 カリフォル
(86) 国際出願番号	PCT/US2014/041568		ニア州 サン ディエゴ タウン センタ
(87) 国際公開番号	W02014/200926		ー ドライブ 9 6 4 0 スイート 1 0
(87) 国際公開日	平成26年12月18日(2014. 12. 18)		0
審査請求日	平成29年6月9日(2017. 6. 9)	(74) 代理人	100079049
(31) 優先権主張番号	61/833, 378		弁理士 中島 淳
(32) 優先日	平成25年6月10日(2013. 6. 10)	(74) 代理人	100084995
(33) 優先権主張国・地域又は機関	米国 (US)		弁理士 加藤 和詳
		(72) 発明者	カオ、 ハン
			アメリカ合衆国 9 2 1 3 0 カリフォル
			ニア州 サン ディエゴ カーメル カン
			トリー ロード 1 2 6 6 8
			最終頁に続く

(54) 【発明の名称】 ポリヌクレオチドの分析

(57) 【特許請求の範囲】

【請求項 1】

宿主生物のゲノムに結合している組み合わせポリヌクレオチドと一緒に形成する第2のポリヌクレオチドに組み込まれた第1のポリヌクレオチドを新規に特性解析する方法であって、

前記第1のポリヌクレオチドを配列特異的に標識化して前記第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意し、前記第2のポリヌクレオチドを配列特異的に標識化して前記第2のポリヌクレオチド上の配列特異的な第2の標識のパターンを用意する工程であって、ここで前記第1のポリヌクレオチドは、前記宿主生物のゲノムを起源としない核酸配列を含み、前記標識化された第1のポリヌクレオチド中の核酸配列の起源は未知である、工程；

前記組み合わせポリヌクレオチドを標識化後に直線化する工程；

前記第1のポリヌクレオチドおよび前記第2のポリヌクレオチド上の配列特異的標識のパターンを検出する工程；

前記第1のポリヌクレオチドに特徴的な標識のパターンを宿主ゲノムの参照パターンとアラインすることにより、前記宿主ゲノムを起源としない前記第1のポリヌクレオチドを含むものとして、前記組み合わせポリヌクレオチドの第1の領域を同定する工程；および

前記第1のポリヌクレオチドに近接する前記第2のポリヌクレオチドの標識のパターンを特性解析して、前記第2のポリヌクレオチドへの前記第1のポリヌクレオチドの組込

みの位置および方向を同定する工程を含む方法。

【請求項 2】

前記宿主ゲノムの参照パターンが、インシリコ参照を含む、請求項 1 に記載の方法。

【請求項 3】

前記宿主ゲノムの参照パターンが、標識化された分子または分子のコレクションとして用意された参照マップである、請求項 1 に記載の方法。

【請求項 4】

前記宿主ゲノムの参照パターンが、電子的または光学的に保存された参照マップとして用意されたものである、請求項 1 に記載の方法。

10

【請求項 5】

前記第 1 のポリヌクレオチドが、前記宿主ゲノムとは異なる因子を含む、請求項 1 に記載の方法。

【請求項 6】

前記第 2 のポリヌクレオチドが宿主生物のゲノムの配列を含む、請求項 1 に記載の方法。

【請求項 7】

生物学的サンプルにおいて 1 以上の外来性因子の有無を決定する方法であって、
前記生物学的サンプルにおいて、ポリヌクレオチドを配列特異的に標識化して第 1 のポリヌクレオチド上の配列特異的な第 1 の標識のパターンを用意する工程；
前記ポリヌクレオチドを標識化後に、ナノチャンネル内で直線化する工程；
前記ポリヌクレオチド上の配列特異的な第 1 の標識のパターンを検出する工程；

20

前記ポリヌクレオチド上の配列特異的な第 1 の標識のパターンを、参照マップのパターンにアラインする工程；

宿主ゲノムにアラインしたパターンを除外する工程；

前記生物学的サンプル由来のポリヌクレオチドの配列特異的な第 1 の標識のパターンを、少なくとも 1 つの外来性因子核酸の参照配列のインシリコパターンとアラインする工程；および

前記外来性因子核酸のインシリコパターンとアラインするポリヌクレオチドの量が検出限界を超えた場合に対象における外来性因子が存在すると決定するか、または、前記外来性因子核酸のインシリコパターンとアラインするポリヌクレオチドの量が検出限界を超えない場合に対象における外来性因子が存在しないと決定する工程；

30

を含む方法。

【請求項 8】

前記少なくとも 1 つの配列特異的な標識のパターンが、宿主核酸配列から外来性因子核酸配列が区別されるように選択される、請求項 7 に記載の方法。

【請求項 9】

前記ポリヌクレオチド上の配列特異的な標識のパターンを参照宿主ゲノムのパターンとアラインする工程；および

40

前記参照宿主ゲノムと完全にアラインされる任意のポリペプチドが 1 以上の外来性因子に特徴的なパターンを含まないことを決定する工程

を更に含む、請求項 7 に記載の方法。

【請求項 10】

前記生物学的サンプル由来のポリヌクレオチドの配列特異的な第 1 の標識の配列パターンを分析することが、前記生物学的サンプル由来のポリヌクレオチドの配列特異的な第 1 の標識の配列パターンを少なくとも 3 つの異なる外来性因子の核酸のインシリコ参照パターンのパターンとアラインすることを含む、請求項 7 に記載の方法。

【請求項 11】

宿主ゲノムに共有結合または非共有結合で結合している第 1 のポリヌクレオチドを新規

50

に特性解析する方法であって、

前記第 1 のポリヌクレオチドおよび前記第 1 のポリヌクレオチドに隣接する配列を配列特異的に標識化して前記第 1 のポリヌクレオチドおよび隣接配列上の配列特異的な第 1 の標識のパターンを用意する工程であって、ここで前記第 1 のポリヌクレオチドは、宿主生物のゲノムを起源とし、ない核酸配列を含み、前記配列特異的な第 1 の標識のパターンを含む前記第 1 のポリヌクレオチドについては起源が未知である、工程；

前記第 1 のポリヌクレオチドを標識化後に直線化する工程；

前記第 1 のポリヌクレオチド上の配列特異的な第 1 の標識のパターンを検出する工程；および

前記第 1 のポリヌクレオチドに隣接する配列上の前記宿主ゲノムに特徴的なパターンの有無に基づいて、前記宿主ゲノム中における前記第 1 のポリヌクレオチドの有無を決定する工程

を含む、方法。

【請求項 1 2】

配列特異的な標識化が、100 kb 当たり 5 ~ 20 標識の標識密度である、請求項 1 ~ 請求項 1 1 のいずれか一項に記載の方法。

【請求項 1 3】

前記第 1 のポリヌクレオチドのコピー数を決定することを更に含む、請求項 1 ~ 請求項 1 2 のいずれか一項に記載の方法。

【請求項 1 4】

前記第 1 のポリヌクレオチドが、ウイルス配列、細菌配列、ミトコンドリア配列、葉緑体配列、エピソーム配列、ミニ染色体配列、転位因子配列、およびファージ配列の少なくとも 1 つを含む、請求項 1 ~ 請求項 1 3 のいずれか一項に記載の方法。

【請求項 1 5】

直線化が、少なくとも 1 つのナノチャネルを用いて行われる、請求項 1 ~ 請求項 1 4 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】

【0001】

関連出願の相互参照

本願は、その全体を参照により本明細書に援用する 2013 年 6 月 10 日付で出願された「ANALYSIS OF POLYNUCLEOTIDES」というタイトルの米国仮特許出願第 61 / 833, 378 号に基づく優先権を主張するものである。

【背景技術】

【0002】

DNA または RNA 等のポリヌクレオチドは、その直線的配列が宿主生物のゲノムおよびポストゲノム遺伝子発現情報に直接関係する、ヌクレオチドの長いポリマー鎖である。ポリヌクレオチドは一本鎖または二本鎖であり得る。

【0003】

配列領域、モチーフ、および機能的単位、例えばオープンリーディングフレーム (ORF)、非翻訳領域 (UTR)、エキソン、イントロン、タンパク質因子結合部位、エピゲノミックな部位、例えば CpG クラスター、マイクロ RNA 部位、トランスポゾン、逆トランスポゾン、ならびに他の構造的および機能的単位の直接的シーケンシングおよびマッピングがゲノム組成の評価、生物学的サンプル中のゲノム外 (extragenomic) ポリヌクレオチドの存在の検出、およびヒトを含む個々の生物の「健康プロフィール」の評価に有用であり得る。

【0004】

場合によっては、個々の生存期間中の、セグメントの重複、挿入、欠失、逆位、および転座等のヌクレオチド配列の複雑な再配列は、遺伝子異常または細胞悪性腫瘍を含む疾患状態につながる。場合によっては、配列差、コピー数多型 (copy number v

10

20

30

40

50

ariation: CNV)、および異なる個体の遺伝的構成 (genetic makeup) 間のその他の差は、集団の遺伝的構成の多様性を反映しており、環境刺激および、例えば薬物処置などその他の外部の影響、に対する種々の反応につながる。ゲノム構造多型 (structural variation: SV) は、健康な個体の中でも、ゲノム中に一般的に見られる。

【0005】

ゲノム外物質はしばしば、宿主生物のゲノムに結合 (associate) し得、更には、宿主生物ゲノムに組み込まれ得る。例えば、種々のウイルスが、宿主ゲノムへの外来性遺伝子物質の組み込みを媒介することができる。例えば、レトロウイルスおよびレトロウイルス様因子 (例えば、ウイルス機能のための1以上の因子、例えばウイルス外被タンパク質をコードする遺伝子、レトロトランスポゾン因子、例えばLTR、非LTRレトロトランスポゾン、例えばLINEおよびSINE等を失ったディジェネレート・レトロウイルス) は、逆転写されて宿主DNAゲノムに組み込まれ得るRNAゲノムを含む。宿主ゲノムに組み込まれた後、レトロウイルスまたはレトロウイルス様因子は転写され得、したがって、逆転写されて宿主ゲノムの更なる部位に組み込まれ得る、より多くのRNAウイルスゲノムを産生し、したがって、宿主ゲノム中で急速に増殖することができる。したがって、レトロウイルスおよびレトロウイルス様因子は、例えば宿主遺伝子中に挿入されることにより、したがって宿主遺伝子を破壊することにより、宿主ゲノムの構造に影響を与え得、1以上の宿主遺伝子の機能に影響を与え得る。更に、組み込み後、レトロウイルスおよびレトロウイルス様因子は増殖し得、宿主ゲノム中の更なる部位に組み込まれ得るが、容易に切り出されず、したがって、蓄積されたレトロウイルスおよびレトロウイルス様因子のパターンが細胞系統 (cell lineage) を示し得る。他にも多くのウイルスが知られている。例えば、DNAは、宿主ゲノムに組み込まれ得、宿主ゲノムから切り出され得る。二本鎖DNAゲノムを有し得るDNAウイルスの例としてバキュロウイルスが挙げられ、昆虫宿主に感染することができる。一本鎖DNAゲノムを有し得るDNAウイルスの例としては特定のバクテリオファージが挙げられ、これは細菌に感染することができる。

【0006】

一部の癌は、宿主ゲノムへのゲノム外物質の送達に関連し得る。例えば、宿主細胞ゲノムへの1以上のオンコジーンのウイルス送達の蓄積は、宿主ゲノム中へのウイルス配列の挿入による宿主細胞ゲノム中の1以上の腫瘍抑制因子遺伝子の破壊と同様に、その宿主細胞中での癌表現型の発症を促進し得る。例えば、ヒトTリンパ球向性ウイルスは、ヒトの癌に関連するウイルスであり、成人T細胞リンパ腫ウイルスI型に関連している。HTLVゲノムは、2コピーの一本鎖RNAウイルスで構成され、ゲノムは宿主細胞ゲノムに組み込まれる二本鎖DNAへとコピーされる。したがって、ウイルスコピーの数および位置のいずれかまたは両方により、癌表現型または癌表現型発症の素質が示されるまたは予測することができる。その全体を参照により援用するRubin, H. (2012) Proc. Nat. Acad. Sci 108: 14389-14396参照。

【0007】

ゲノム外物質は、他の機構によっても宿主ゲノムに結合し得る。例えば、水平遺伝子伝播が起こり得る。水平遺伝子伝播は、親から子供への遺伝子伝播とは異なる機構 (例えば、有性生殖または無性生殖を含まない機構) による第1の生物からの遺伝情報の伝播を含む。例えば、レトロウイルスは、第1の宿主ゲノムに挿入され得、その転写産物の形で第1の宿主ゲノムの一部に組み込まれ得、これにより、第2の宿主ゲノムに感染すると、第1の宿主ゲノムの一部を第2の宿主ゲノムへと組み込むことができるウイルスが産生され、例えば、種間遺伝子伝播 (interspecies gene transfer) が、宿主に感染する病原性細菌により媒介され得る。例えば、原生動物T. cruzi によるヒトへの感染が報告されており、しばしば、種々の宿主染色体のレトロ転移因子LINE-1因子へのミトコンドリアミニサークルの組み込みを含む。T. cruzi 仲介DNA伝播は、ヒト患者におけるシャーガス心臓病を引き起こし得る。例えば液胞内病原

体サルモネラ・エンテリカ (*Salmonella enterica*) によるおよび細胞質内病原体リステリア・モノサイトゲネス (*Listeria monocytogenes*) による、細菌からヒト細胞への種間遺伝子伝播の更なる例が、それぞれの全体を参照により援用するLlosa et al. (2012)、Trends Microbiol20: 355-9、およびHecht et al. (2010)、PLOS ONE 5: e 9181に記載されている。

【0008】

多くの種類の癌、細菌、寄生虫、およびウイルス性の疾患が、ゲノム中の異常および多型により生じ得る。疾患を検出するための多くの種類のゲノムワイドなアプローチにも関わらず、多くは、宿主ゲノム内の病原性ポリヌクレオチドの特異的コピー数の定量、宿主ゲノム内の病原性ヌクレオチドの位置マッピング、およびゲノム外物質の位置をマッピングするための大きく複雑なゲノムのアセンブリにおいて間接的でもあり、現在の技術およびアプローチは精度が非常に低い。ポリヌクレオチドを特性解析するためならびに生物学的サンプル中および対照中における病原体のポリヌクレオチド構造パターンの有無 (presence or absence) を決定するための方法およびキットに対するニーズが存在する。

【発明の概要】

【発明が解決しようとする課題】

【0009】

本明細書に記載の実施形態は全般的には、核酸のナノテクノロジーおよび分析に関する。本明細書のいくつかの実施形態は単一分子ゲノム分析に関する。

【課題を解決するための手段】

【0010】

いくつかの態様では、組み合わせ (combined) ポリヌクレオチドと一緒に形成する、第2のポリヌクレオチドに組み込まれた第1のポリヌクレオチドを特性解析する方法が提供される。方法は、第1のポリヌクレオチドを配列特異的に標識化して第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意 (provide) する工程を含み得る。方法は、第2のポリヌクレオチドを配列特異的に標識化して第2のポリヌクレオチド上の第2の標識の配列特異的なパターンを用意する工程を含み得る。いくつかの実施形態では、第1および第2の標識は同じであっても異なってもよい。方法は、組み合わせポリヌクレオチドを標識化後に直線化 (linearize) する工程を含み得る。方法は、第1および第2のポリヌクレオチド上の配列特異的な標識のパターンを検出する工程を含み得る。方法は、第1のポリヌクレオチドに特徴的な標識化パターンに基づいて、第1のポリヌクレオチドを含むとして、組み合わせポリヌクレオチドの第1の領域を同定する工程を含み得る。方法は、第1のポリヌクレオチドに近接 (adjacent) する第2のポリヌクレオチドの標識化パターンを特性解析して、前記第2のポリヌクレオチドへの前記第1のポリヌクレオチドの組込みの位置および方向を同定する工程を含み得る。いくつかの実施形態では、方法は、第1のポリヌクレオチドのコピー数を決定することを更に含む。いくつかの実施形態では、コピー数の決定は、第1のポリヌクレオチドの塩基当たりのカバレッジ深度と第2のポリヌクレオチドの塩基当たりのカバレッジ深度の比率を決定することを含む。いくつかの実施形態では、第1のポリヌクレオチドは、一本鎖DNA、一本鎖RNA、二本鎖DNA、および二本鎖RNAのうち少なくとも1つを含む。いくつかの実施形態では、第1のポリヌクレオチドは、ウイルス配列、細菌配列、ミトコンドリア配列、葉緑体配列、エピソーム配列、ミニ染色体配列、転位因子配列、およびファージ配列からなる群から選択される。いくつかの実施形態では、第2のポリヌクレオチドは宿主ゲノム配列を含む。いくつかの実施形態では、宿主はヒトを含む。いくつかの実施形態では、第2のポリヌクレオチドは、疾患状態の細胞のゲノム配列を含む。いくつかの実施形態では、疾患状態は癌を含む。いくつかの実施形態では、方法は、少なくとも第3のポリヌクレオチドの配列特異的な標識のパターンを用意する工程、第3のポリヌクレオチドの少なくとも領域内で第1のポリヌクレオチドに特徴的な標識化パターンに基づいて第1のポリヌクレオチドと少なくとも類似した第3のポリヌクレオチドの領域を同

10

20

30

40

50

定する工程、および前記第1のポリヌクレオチドに特徴的な標識化パターンに基づいて前記第1のポリヌクレオチドと前記第3のポリヌクレオチドとの間の系統関係を決定する工程を更に含む。いくつかの実施形態では、第3のポリヌクレオチドの配列特異的標識のパターンは、参照配列上のインシリコ (in silico) パターンを含む。いくつかの実施形態では、方法は、第3のポリヌクレオチドを配列特異的に標識化して第3のポリヌクレオチド上の第3の標識の配列特異的パターンを用意する工程、第3のポリヌクレオチドを標識化後に直線化する工程、および第3のポリヌクレオチド上の配列特異的標識のパターンを検出する工程を更に含む。いくつかの実施形態では、第1の標識、第2の標識、および第3の標識の任意の2つは同じであっても異なってもよい。いくつかの実施形態では、系統関係は、種間ポリヌクレオチド伝播イベントを含む。いくつかの実施形態では、系統関係はDNA水平伝播イベントを含む。いくつかの実施形態では、方法は、第4のポリヌクレオチド中の前記第1のポリヌクレオチドの組込みの位置を少なくとも1つ用意する工程ならびに前記第2のポリヌクレオチド中の前記第1のポリヌクレオチドの組込みの少なくとも1つの位置および前記第4のポリヌクレオチド中の前記第1のポリヌクレオチドの組込みの少なくとも1つの位置に基づいて前記第2のポリヌクレオチドと前記第4のポリヌクレオチドとの間の系統関係を決定する工程を更に含む。いくつかの実施形態では、前記第4のポリヌクレオチド中の前記第1のポリヌクレオチドの組込みの少なくとも1つの位置はインシリコで用意される。いくつかの実施形態では、直線化は少なくとも1つのナノチャンネルを用いて行われる。いくつかの実施形態では、少なくとも1つの配列特異的標識化のパターンは、宿主核酸配列からゲノム外核酸配列が区別されるように選択される。いくつかの実施形態では、宿主核酸配列のパターンからゲノム外核酸配列のパターンが区別されるように標識密度が選択される。いくつかの実施形態では、配列特異的標識化の少なくとも1つの参照パターンは、100 kb 当たり約5 ~ 約20 標識の標識密度を有するように選択される。いくつかの実施形態では、配列特異的標識化の少なくとも1つの参照パターンは、100 kb 当たり約5 標識、約10 標識、約15 標識、約20 標識の標識密度からこれらの値の間のおよそ任意の量の標識密度を有するように選択される。

【0011】

いくつかの態様は、少なくとも宿主ゲノムに結合している第1のポリヌクレオチドを特性解析する方法を含む。方法は、第1のポリヌクレオチドを配列特異的に標識化して第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意する工程、第1のポリヌクレオチドを標識化後に直線化する工程、第1のポリヌクレオチド上の配列特異的標識のパターンを検出する工程、および第1のポリヌクレオチド上の宿主ゲノムに特徴的なパターンの非存在 (absence) に基づいて宿主ゲノム中における第1のポリヌクレオチドの有無を決定する工程を含み得る。いくつかの実施形態では、方法は、宿主ゲノムに結合している複数のポリヌクレオチドを配列特異的に標識化して、複数のポリヌクレオチドのそれぞれの上の配列特異的標識のパターンを用意する工程、複数のポリヌクレオチドのそれぞれの上の配列特異的標識のパターンを検出する工程、複数のポリヌクレオチドの少なくともいくつかの上の第1のポリヌクレオチドに特徴的なパターンの有無を検出する工程、および第1のポリヌクレオチドに特徴的なパターンと複数のポリヌクレオチドの少なくともいくつかのパターンとの間の類似性に基づいてコンセンサスパターンを同定する工程を更に含む。いくつかの実施形態では、第1のポリヌクレオチドは、第1のポリヌクレオチドの第1の末端および第2の末端のそれぞれの上のパターンを含む。いくつかの実施形態では、コンセンサスパターンに特徴的であるが宿主ゲノムの如何なる領域にも特徴的でない第1および第2の末端上のパターンにより、第1のポリヌクレオチドが宿主ゲノムに組み込まれていないことが示される。いくつかの実施形態では、方法は、第1のポリヌクレオチドのコピー数を決定することを更に含む。いくつかの実施形態では、方法は、第1のポリヌクレオチドのコピー数を決定することを更に含む。いくつかの実施形態では、コピー数の決定は、コンセンサスパターンを実質的に含むポリヌクレオチドの塩基当たりのカバレッジ深度と宿主ゲノムの塩基当たりのカバレッジ深度の比率を決定することを包含する。いくつかの実施形態では、宿主ゲノムに特徴的なパターンは、参照配列上のインシ

10

20

30

40

50

リコパターンを含む。いくつかの実施形態では、方法は、宿主ゲノムのポリヌクレオチドを配列特異的に標識化して宿主ゲノムのポリヌクレオチド配列上の配列特異的な第2の標識のパターンを用意する工程であって、第1および第2の標識は同じであっても異なってもよい工程、宿主ゲノムのポリヌクレオチドを標識化後に直線化する工程、および宿主ゲノム上の配列特異的標識のパターンを検出することにより、少なくとも宿主ゲノムに特徴的なパターンを用意する工程を更に含む。いくつかの実施形態では、第1のポリヌクレオチドは、一本鎖DNA、一本鎖RNA、二本鎖DNA、および二本鎖RNAのうちの少なくとも1つを含む。いくつかの実施形態では、第1のポリヌクレオチドは、ウイルス配列、細菌配列、ミトコンドリア配列、葉緑体配列、エピソーム配列、ミニ染色体配列、転位因子配列、およびファージ配列からなる群から選択される。いくつかの実施形態では、宿主はヒトを含む。いくつかの実施形態では、直線化は少なくとも1つのナノチャンネル中で行われる。

10

【0012】

いくつかの態様では、生物学的サンプル中の1以上の外来性因子の有無を決定する方法が提供される。方法は、外来性因子核酸の配列特異的標識化の少なくとも1つの参照パターンを用意する工程、複数のポリヌクレオチドを含む生物学的サンプルを用意する工程、生物学的サンプルのポリヌクレオチドを配列特異的に標識化して第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意する工程、ポリヌクレオチドを標識化後に直線化する工程、ポリヌクレオチド上の配列特異的標識のパターンを検出する工程、およびポリヌクレオチド上の1以上の外来性因子に特徴的な1以上のパターンの有無を決定することにより、生物学的サンプル中における1以上の外来性因子の有無を決定する工程を含み得る。いくつかの実施形態では、少なくとも1つの配列特異的標識化のパターンは、宿主核酸配列から外来性因子核酸配列が区別されるように選択される。いくつかの実施形態では、配列特異的標識化の少なくとも1つの参照パターンは、100 kb 当たり約5～約20標識の標識密度を有するように選択される。いくつかの実施形態では、生物学的サンプルは、存在可能性のある外来性因子核酸について濃縮される。いくつかの実施形態では、濃縮は、サンプルから宿主核酸を抽出することを含む。いくつかの実施形態では、濃縮は、外来性因子核酸について濃縮されたサンプルを用意するために外来性因子核酸を抽出することを含む。いくつかの実施形態では、濃縮は、サンプルから宿主ポリヌクレオチドを実質的に除去することを含む。いくつかの実施形態では、方法は、ポリヌクレオチドの配列特異的標識のパターンを参照宿主ゲノムのパターンとアラインする工程および参照宿主ゲノムと完全にアラインされる任意のポリヌクレオチドが1以上の外来性因子に特徴的なパターンを含まないことを決定する工程を更に含む。いくつかの実施形態では、検出されたパターンの全てを、参照宿主ゲノムと外来性因子核酸の配列特異的標識化の参照パターンとを含む複合参照のパターンとアラインする工程；および外来性因子核酸の配列特異的標識化の参照パターンとアラインされるのポリヌクレオチドが1以上の外来性因子に特徴的なパターンを含むと決定する工程。いくつかの実施形態では、方法は、外来性因子核酸を含むポリヌクレオチドの割合を推定することを更に含む。いくつかの実施形態では、サンプルは血液サンプルを含む。いくつかの実施形態では、宿主は哺乳動物を含む。いくつかの実施形態では、宿主はヒトを含む。いくつかの実施形態では、外来性因子核酸の配列特異的標識化の少なくとも1つの参照パターンは、ヒト免疫不全ウイルス1、ヒト免疫不全ウイルス2、ヒトTリンパ球向性ウイルスI、ヒトTリンパ球向性ウイルスII、B型肝炎、C型肝炎、梅毒トレポネーマ（梅毒）、ウエストナイルウイルス、トリパノソーマ・クルージ（シャーガス病）、サイトメガロウイルス（CMV）、またはマラリアに特異的な、少なくとも1つのパターンを含む。いくつかの実施形態では、外来性因子核酸の配列特異的標識化の少なくとも1つの参照パターンを用意する工程は、少なくとも3つの異なる外来性因子について核酸の参照パターンを用意することを含む。いくつかの実施形態では、外来性因子核酸の配列特異的標識化の少なくとも1つの参照パターンを用意する工程は、少なくとも10の異なる外来性因子について核酸の参照パターンを用意することを含む。いくつかの実施形態では、少なくとも1つの配列特異的標識化の参照パターン

20

30

40

50

は、参照配列上のインシリコパターンを含む。いくつかの実施形態では、ポリヌクレオチドは、一本鎖DNA、一本鎖RNA、二本鎖DNA、および二本鎖RNAのうちの少なくとも1つを含む。いくつかの実施形態では、直線化は少なくとも1つのナノチャンネル中で行われる。いくつかの実施形態では、標識化は、ニックング酵素を用いた標識化を含む。いくつかの実施形態では、標識化は、非切断標識を用いたタグ化を含む。いくつかの実施形態では、非切断標識は、メチルトランスフェラーゼ、非切断制限酵素、ジンクフィンガータンパク質、抗体、転写因子、DNA結合タンパク質、ヘアピンポリアミド、三本鎖形成性オリゴデオキシヌクレオチド、ペプチド核酸の少なくとも1つによって適用される。いくつかの実施形態では、標識化は光学的標識化を含む。いくつかの実施形態では、標識化は非光学的標識を含む。いくつかの実施形態では、外来性因子は病原性である。いくつかの実施形態では、外来性因子は細菌性の外来性因子である。いくつかの実施形態では、外来性因子はウイルス性の外来性因子である。いくつかの実施形態では、外来性因子は寄生虫の外来性因子である。いくつかの実施形態では、生物学的サンプルは供与のためのものである。

【0013】

いくつかの態様では、対象中における1以上の外来性因子の有無を検出する方法が提供される。方法は、外来性因子核酸の配列特異的標識化の少なくとも1つの参照パターンを用意する工程；複数のポリヌクレオチドを含む生物学的サンプルを対象から採取する工程；生物学的サンプルのポリヌクレオチドを配列特異的に標識化して、第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意する工程；ポリヌクレオチドを標識化後に直線化する工程；ポリヌクレオチド上の配列特異的標識のパターンを検出する工程；ポリヌクレオチド上における1以上の外来性因子に特徴的な1以上のパターンの有無を決定することにより、生物学的サンプル中における1以上の外来性因子の有無を決定する工程；生物学的サンプルに由来するポリヌクレオチドの配列特異的標識のパターンを少なくとも1つの外来性因子核酸の参照配列上のインシリコパターンのパターンとアラインすることにより、生物学的サンプルに由来するポリヌクレオチドの配列特異的標識の配列パターンを分析する工程；および少なくとも1つの外来性因子核酸の参照配列と類似した生物学的サンプルに由来するポリヌクレオチドに特徴的な1以上のパターンの有無に基づいて対象中における外来性因子の有無を決定することにより、対象中における外来性因子の有無を決定する工程を含み得る。いくつかの実施形態では、方法はインビトロの方法を含む。いくつかの実施形態では、外来性因子は、ヒト免疫不全ウイルス1、ヒト免疫不全ウイルス2、ヒトTリンパ球向性ウイルスI、ヒトTリンパ球向性ウイルスII、B型肝炎、C型肝炎、ウエストナイルウイルス、およびサイトメガロウイルス(CMV)のうちの1以上を含む。いくつかの実施形態では、少なくとも1つの配列特異的標識化のパターンは、宿主配列から外来性因子核酸配列が区別されるように選択される。いくつかの実施形態では、少なくとも1つの配列特異的標識化のパターンは、宿主核酸配列から外来性因子核酸配列が区別されるような標識密度を有する。いくつかの実施形態では、配列特異的標識化の少なくとも1つの参照パターンは、100kb当たり約5～約20標識の標識密度を有するように選択される。いくつかの実施形態では、配列特異的標識化の少なくとも1つの参照パターンは、100kb当たり約10標識の標識密度を有するように選択される。いくつかの実施形態では、ポリヌクレオチドは、存在可能性のある外来性因子核酸について濃縮される。いくつかの実施形態では、濃縮は、サンプルから宿主ポリヌクレオチドを実質的に除去することを含む。方法は、ポリヌクレオチド上の配列特異的標識のパターンを参照宿主ゲノムのパターンとアラインする工程および参照宿主ゲノムと完全にアラインされる任意のポリヌクレオチドが1以上の外来性因子に特徴的なパターンを含まないことを決定する工程を更に含み得る。方法は、検出されたパターンの全てを、参照宿主ゲノムと外来性因子核酸の配列特異的標識化の参照パターンとを含む複合参照のパターンとアラインする工程および外来性因子核酸の配列特異的標識化の参照パターンとアラインされる任意のポリヌクレオチドが1以上の外来性因子に特徴的なパターンを含むことを決定する工程を更に含み得る。方法は、検出されたパターンの全てを、参照宿主ゲノムと外来性因

10

20

30

40

50

子核酸の配列特異的標識化の参照パターンとを含む複合参照のパターンとアラインする工程および外来性因子核酸の配列特異的標識化の参照パターンとアラインされた任意のポリヌクレオチドが1以上の外来性因子に特徴的なパターンを含むことを決定する工程を更に含み得る。方法は、外来性因子核酸を含むポリヌクレオチドの割合を推定することを更に含み得る。いくつかの実施形態では、サンプルは血液サンプルを含む。いくつかの実施形態では、対象は哺乳動物を含む。いくつかの実施形態では、対象はヒトを含む。いくつかの実施形態では、外来性因子核酸の配列特異的な標識化の少なくとも1つの参照パターンは、ヒト免疫不全ウイルス1、ヒト免疫不全ウイルス2、ヒトTリンパ球向性ウイルスI、ヒトTリンパ球向性ウイルスII、B型肝炎、C型肝炎、梅毒トレポネーマ（梅毒）、ウエストナイルウイルス、トリパノソーマ・クルージ（シャーガス病）、サイトメガロウイルス（CMV）、またはマラリアに特異的な、少なくとも1つのパターンを含む。いくつかの実施形態では、方法は、生物学的サンプル由来のポリヌクレオチドの配列特異的標識の配列パターンを分析することを含み生物学的サンプル由来のポリヌクレオチドの配列特異的標識の配列パターンを少なくとも3つの異なる外来性因子についての核酸のインシリコ参照パターンのパターンとアラインすることを含む。いくつかの実施形態では、分析は、生物学的サンプル由来のポリヌクレオチドの配列特異的標識の配列パターンを分析することを含み生物学的サンプル由来のポリヌクレオチドの配列特異的標識の配列パターンを少なくとも10の異なる外来性因子についての核酸のインシリコ参照パターンのパターンとアラインすることを含む。いくつかの実施形態では、配列特異的標識化の少なくとも1つの参照パターンは、参照配列上のインシリコパターンを含む。いくつかの実施形態では、ポリヌクレオチドは、一本鎖DNA、一本鎖RNA、二本鎖DNA、および二本鎖RNAのうちの少なくとも1つを含む。いくつかの実施形態では、直線化は、少なくとも1つのナノチャンネル中で行われる。いくつかの実施形態では、標識化は、ニックング酵素を用いた標識化を含む。いくつかの実施形態では、標識化は、非切断標識を用いたタグ化を含む。標識化が非切断標識を用いたタグ化を含むいくつかの実施形態では、非切断標識は、トランスフェラーゼ、非切断制限酵素、ジンクフィンガータンパク質、抗体、転写因子、DNA結合タンパク質、ヘアピンポリアミド、三本鎖形成オリゴデオキシヌクレオチド、ペプチド核酸の少なくとも1つによって適用される。いくつかの実施形態では、外来性因子は病原性である。いくつかの実施形態では、外来性因子はウイルスを含む。いくつかの実施形態では、外来性因子は細菌を含む。いくつかの実施形態では、外来性因子は寄生虫を含む。いくつかの実施形態では、外来性因子は病原体を含む。いくつかの実施形態では、1以上の外来性因子の有無の検出は、疾患進行を決定するために生物学的サンプル由来のポリヌクレオチドの標識密度を分析することを更に含む。いくつかの実施形態では、生物学的サンプルは供与のためのものである。

【0014】

いくつかの態様では、対象中における外来性因子の有無を決定するためのキットが提供される。キットは、宿主核酸配列から外来性因子核酸配列が区別されるような標識密度で外来性因子の核酸を標識化するための試薬を含み得る。いくつかの実施形態では、キットはインビトロでの検出のためのものである。いくつかの実施形態では、キットは、対象のゲノムの核酸配列から外来性因子核酸配列が区別されるような標識密度で複数の外来性因子ゲノムを標識化するための標識化試薬を含む。いくつかの実施形態では、試薬は、100kb当たり約1標識～100kb当たり約50標識の標識密度で1以上の外来性因子のゲノムを標識化する。いくつかの実施形態では、試薬は、100kb当たり約5～約20標識の標識密度で1以上の外来性因子のゲノムを標識化する。いくつかの実施形態では、キットは、ヒト免疫不全ウイルス1、ヒト免疫不全ウイルス2、ヒトTリンパ球向性ウイルスI、ヒトTリンパ球向性ウイルスII、B型肝炎、C型肝炎、梅毒トレポネーマ（梅毒）、ウエストナイルウイルス、トリパノソーマ・クルージ（シャーガス病）、サイトメガロウイルス（CMV）、およびマラリアのうちの1以上を含む1以上の外来性因子を含む。いくつかの実施形態では、キットは、核酸を単離するための試薬を更に含む。いくつかの実施形態では、キットは、外来性因子核酸について対象由来のサンプルを濃縮するた

10

20

30

40

50

めの試薬を更に含む。いくつかの実施形態では、キットは、流体チャネル、例えばナノチャネル中で核酸を直線化するための試薬を更に含む。

【図面の簡単な説明】

【0015】

【図1A】いくつかの実施形態に係るポリヌクレオチドを標識化する方法を示す模式図である。図1Aは、いくつかの実施形態に係るフラップ生成を示している。

【図1B】図1Bは、いくつかの実施形態に係るフラップの標識化を示している。

【図2】いくつかの実施形態に係るポリヌクレオチドを標識化する方法を示す模式図である。タグ化剤A、B、またはCが、組込みまたはハイブリダイゼーション中に、伸長したDNAと結合し得る。特異的FRETシグナルにより共局在イベントを検出することができる。

10

【図3】本明細書のいくつかの実施形態に係る、新規に(*de novo*)作成された、hg19参照インシリコマップに重ねたヒトゲノムマップアセンブリを示す図である。白色セクションは、セントロメアまたはテロメア近くの参照中のN塩基領域である。

【図4】本明細書のいくつかの実施形態に係る参照ゲノム配列32と比較した、アセンブルされたヒトゲノムマップ30中の部位特異的標識のアラインされたパターンを示す模式図である。強力な単一分子サポートを用いて、大きなタンデムリピート(反復の各ユニット34は約80kbである)が5番染色体に同定された。

【図5】本明細書のいくつかの実施形態に係る、エプスタイン・バーウイルス配列の配列特異的標識のアラインされたパターンおよび宿主ゲノムに挿入されたエプスタイン・バーウイルス配列に隣接する(*flanking*)宿主ゲノム配列の配列特異的標識を示す模式図である。隣接配列を破線の四角中に示す。ウイルス配列の内部パターン変化を点線の四角中に示す。

20

【図6A】本明細書のいくつかの実施形態に係る宿主ゲノム中の新規な(*de novo*)ウイルス構成要素の発見の模式図である。図6Aに示されるように、宿主ゲノム中のレトロウイルスのサイン(*signature*)が、部位特異的標識化により同定された。レトロウイルスのサインを含む配列を互いにアラインし、インシリコによる決定でレトロウイルスのサインに対する高度の類似性が示された。

【図6B】図6Bに示されるように、宿主ゲノム中のレトロウイルスコピー数が決定された。

30

【図7】本明細書のいくつかの実施形態に係る、第2のポリヌクレオチド中に組み込まれた第1のポリヌクレオチドを特性解析する方法を示すフローダイアグラムである。

【図8】本明細書のいくつかの実施形態に係る、宿主ゲノムに結合している少なくとも第1のポリヌクレオチドを特性解析する方法を示すフローダイアグラムである。

【図9】本明細書のいくつかの実施形態に係る、1以上の外来性因子の有無について生物学的サンプルをスクリーニングする方法を示すフローダイアグラムである。

【図10】本明細書のいくつかの実施形態に係る、対象における1以上の外来性因子の有無を決定する方法を示すフローダイアグラムである。

【図11】図14に示されているように、本明細書のいくつかの実施形態に係るインシリコ実験の結果から得られたサンプル中の外来性因子の量との関係における信頼値を示す図である。

40

【図12】本明細書のいくつかの実施形態に係る、生物学的サンプル中の外来性因子の有無を分析するためのサンプル調製の例を示すフローダイアグラムである。

【図13】本明細書のいくつかの実施形態に係る外来性因子ゲノムのインシリコ実験の結果を示す表である(外来性因子ゲノムの特徴は図14にも示されている)。FP = 偽陽性、FN = 偽陰性、TP = 真陽性、TN = 真陰性

【図14】本明細書のいくつかの実施形態に係る実験で使用された外来性因子ゲノムの表である。

【発明を実施するための形態】

【0016】

50

宿主生物のゲノムを起源としなかったポリヌクレオチド配列が、種々の機構により宿主生物のゲノムに結合し得る。本明細書に開示されているいくつかの実施形態では、少なくとも部分的にゲノム外因子を含む宿主生物ゲノムのポリヌクレオチドの1以上の長い断片が、特異的部位で標識されて直線化され得、部位特異的標識化のパターンが同定され得る。そのようなゲノム外因子を含む宿主ポリヌクレオチド中の部位特異的標識化のパターンが、レトロウイルス、病原性細菌、および病原性寄生虫の配列等の既知のゲノム外配列の部位特異的標識化のパターンと比較され得る。宿主ポリヌクレオチド中のゲノム外配列に特徴的なパターンは、宿主生物に結合したゲノム外因子の存在および宿主ゲノム中のどこにどのようにそれらが挿入されているかを示し得る。いくつかの実施形態では、宿主ゲノムに結合したゲノム外配列は新規に (*de novo*) 同定される。いくつかの実施形態では、宿主ゲノム結合ゲノム外因子は宿主ゲノムに組み込まれている。宿主ゲノムに結合したゲノム外配列に隣接する配列の部位特異的標識化により、これらの隣接配列上のパターンを作成することができる。宿主ゲノムに特徴的な隣接配列上のパターンの存在は、ゲノム外配列が宿主ゲノムに組み込まれていることを示し得、更に、組込み部位の同定に用いることができる。いくつかの実施形態では、ゲノム外因子は宿主ゲノムに組み込まれていないが、その代わりに、宿主ゲノムから離れて、例えばオルガネラに埋め込まれてまたはミニ染色体、エピソーム、もしくはプラスミド中に存在する。宿主結合ゲノム外因子の末端配列の部位特異的標識化により、宿主ゲノム中に存在しないが既知のゲノム外配列に特徴的なパターンおよび/または宿主結合ゲノム外配列のコンセンサスもしくは「サイン」パターンを生成する場合、宿主結合ゲノム外配列は宿主ゲノムから離れて存在すると決定することができる。いくつかの実施形態では、その宿主生物ゲノムの量 (*quantity count*) を参照して、少なくとも1つのゲノム外配列因子の物理的数が定量的に測定される。いくつかの実施形態では、そのようなゲノム外配列因子は、異常な宿主ゲノム増殖の結果および指標として、宿主ゲノム自体に由来する。

【0017】

本明細書において、「外来性因子 (*foreign element*) 」とは、宿主生物のゲノムを起源としない、ゲノムを構成する因子または自己複製生物を指す。典型的な外来性因子としては、限定されるものではないが、微生物、細菌、真菌、ウイルス、ウイルス性因子、転位因子等が含まれる。外来性因子は、具体的な外来性因子に応じて、宿主生物のゲノムに結合して見られてもよく、宿主ゲノムから遊離して存在してもよく、且つ/または宿主ゲノムに組み込まれていてもよい。いくつかの実施形態では、外来性因子は寄生虫のものである。いくつかの実施形態では、外来性因子は病原性である。いくつかの実施形態では、外来性因子は宿主にとって良性または有益である。従来、サンプルからのウイルス、細菌、酵母、および/または寄生虫の核酸等の外来性因子核酸の検出は、PCRに基づく選別のための免疫学的試験および質量分析法に基づくペプチドフィンガープリント法等の戦略に頼ってきた。しかし、PCRに基づく方法の落とし穴として、PCRが阻害剤、汚染、および他の実験条件の影響を受け易いことが含まれ得る。これは、標的遺伝子、プライマー、PCR技術、およびサンプルの純度に大きく依存するPCRアッセイの感度および特異性に考慮すべきことである。PCR法は更に、時間がかかり、高価であり、標的に対する特異的プローブ設計およびPCR最適化技術が必要である。質量分析法に基づくフィンガープリント法に関して、当業者には、典型的な質量分析法が、タンパク質混合物の分析について精密度が高くないことが理解されよう。したがって、いくつかの実施形態では、サンプルはタンパク質について精製される。

【0018】

標識化および標識

いくつかの実施形態では、ポリヌクレオチド、例えばRNAまたはDNAが、ポリヌクレオチドの一本鎖にプローブをハイブリダイズさせることにより標識化される。プローブは、RNAまたはDNAの鎖、またはその一部に相補的であり得る。いくつかの実施形態では、プローブは特異的配列モチーフに相補的である。いくつかの実施形態では、複数の特異的モチーフに相補的であるように、複数のプローブ、例えば少なくとも2、3、4、

5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、5、000、または10、000（記載値の任意の2つの間の範囲を含む）のプロープが用意される。いくつかの実施形態では、プロープはランダム配列を有する。いくつかの実施形態では、複数のランダム配列を有するプロープが用意される。

【0019】

いくつかの実施形態では、第1のDNA鎖および第2のDNA鎖を含む二本鎖DNAを処理して第1のDNA鎖のハイブリダイズしていない約1～約1000塩基を含むフラップおよび第2のDNA鎖上の対応する領域を生じさせ；第1のDNA鎖を第2のDNA鎖の対応する領域に沿って伸長させ；ハイブリダイズしていないフラップの少なくとも一部、伸長した第1のDNA鎖の一部を標識化することにより、二本鎖DNAが標識化される。標識化は好ましくは、(a)ハイブリダイズしていないフラップの少なくとも一部に、1以上のタグを含む少なくとも1つの相補的プロープを結合させること；(b)第2のDNA鎖の対応する領域に沿って伸長させる第1のDNA鎖の一部である置換塩基として、1以上のタグを含むヌクレオチドを用いること；または(a)および(b)の任意の組合せによって達成することができる。このようにすることで、フラップ、ギャップを埋める塩基、または両方が標識化され得る。

【0020】

いくつかの実施形態では、最初に二本鎖DNAの第1の鎖にニックを入れることにより二本鎖DNAが標識化され得る。ニックングはランダムまたは非特異的位置を含む1以上の非特異的位置でなされ得るが、このニックングは、好ましくは1以上の配列特異的位置でなされ得る。ニックングは、二本鎖DNAポリマーをニックングエンドヌクレアーゼ、またはニッカーゼに曝露することにより適切に達成され得る。ニッカーゼは好ましくは高度に配列特異的であり、このことは、ニッカーゼが高度の特異性で特定の配列の塩基（モチーフ）に結合することを意味する。ニッカーゼは、例えばニュー・イングランド・バイオラボ社（www.neb.comからアクセス可能）から入手可能である。典型的なニッカーゼとしては、限定されるものではないが、Nb.BbvCI；Nb.BsmI；Nb.BsrDI；Nb.BtsI；Nt.AlwI；Nt.BbvCI；Nt.BspQI；Nt.BstNBI；Nt.CviPII、およびこれらの組合せが含まれる。ニックングは、DNAの鎖の破断または切断に影響を与える他の酵素によって達成されてもよい。そのような破断またはニックは、電磁放射（例えば、UV光）、1以上のフリーラジカル等への曝露によって達成されてもよい。ニックは、これらの技術の1以上の影響を受け得る。いくつかの実施形態では、二本鎖DNAの第1の鎖（すなわち、ニックの入った鎖）への置換塩基の取込みは、好ましくは、DNAをポリメラーゼ、1もしくは複数のヌクレオチド、リガーゼ、またはこれらの任意の組合せと接触させることを含む。いくつかの実施形態では、ニックを入れたDNAの標識化後の、リガーゼを用いた処理により、dsDNAの完全性を回復することができ、得られる鎖の強度を有意に強化することができる。

【0021】

標識化は更に、メチルトランスフェラーゼ標識化を含み得る。例えば、M.BseCI（メチルトランスフェラーゼ）は、ヒトゲノム中に低頻度で存在し、微生物および他の下等病原性真核生物中で許容可能な頻度で存在するので、有用であり得る。いくつかの実施形態では、標識化のためのポリヌクレオチドは、一本鎖DNA、一本鎖RNA、二本鎖DNA、または二本鎖RNAの少なくとも1つを含む。

【0022】

いくつかの実施形態では、最初に昇温または有機溶媒を用いた操作により特定のゲノム領域の二本鎖間の水素結合を融解していわゆるDループを開き、次いで、相対的に安定な形態へと再びアニリングする前に一本鎖領域への等しいまたはより高い親和性を有する少なくとも1つの特異的プロープにハイブリダイズさせることにより、二本鎖DNAが標識化され得る。このように、いくつかの実施形態では、いずれかの鎖にニックを入れるこ

10

20

30

40

50

となく、本明細書に記載されているプローブによって二本鎖DNAを標識化することができる。いくつかの実施形態では、一本鎖上で複数のDループを開いてもよい。したがって、複数のプローブが特定の二本鎖DNAにアニーリングし得る。

【0023】

本明細書のいくつかの実施形態では、各標識は、フルオロフォア、量子ドット、デンドリマー、ナノワイヤー、ビーズ、ハプテン、ストレプトアビジン、アビジン、ニュートラアビジン、ピオチン、および反応性基からなる群から独立して選択される。本明細書のいくつかの実施形態では、第1および第2の標識は、フルオロフォアまたは量子ドットからなる群から独立して選択される。本明細書のいくつかの実施形態では、標識の少なくとも1つは非光学的標識を含む。いくつかの実施形態では、2つ以上の標識が同じである。例えば、第1のDNAが標識化および特性解析され、第2のDNAが標識化および特性解析される場合、第1のDNAおよび第2のDNAは、同じ種類の標識、例えば同じフルオロフォア、同じ量子ドット、または同じ非光学的標識で標識化され得る。例えば、第1のDNAを第1のナノチャンネル中で特性解析、第2のDNAを第2のナノチャンネル中で特性解析することができるので、各DNAを同じ標識化部分で標識化したとしても、2つのDNAの標識化パターンを区別することができる。いくつかの実施形態では、例えば、単一DNAが2つ以上の異なるモチーフで標識化される場合、第1の標識および第2の標識は異なる。

【0024】

いくつかの実施形態では、可逆的ターミネーター(reversible terminator)を含む標識化されたヌクレオチドを用いてニック標識化が行われる。単一の可逆的ターミネーターを含む標識化されたヌクレオチドが、ニックに取り込まれ得、その結果、各ニックに1を超えない標識が取り込まれる。例えば、ヌクレオチドをターミネーターに連結するリンカーが切断され得る。ターミネーターの破棄(reverseal)後、ニックが修復され得る。その後、DNA上の第1の標識のパターンが検出されるように標識が検出され得る。

【0025】

いくつかの実施形態では、本明細書に記載の少なくとも1つの標識は非光学的標識を含む。種々の非光学的標識を本明細書に記載の実施形態と組み合わせて用いることができる。いくつかの実施形態では、非光学的標識は電子標識(electronic label)を含む。典型的な電子標識としては、限定されるものではないが、強力な電荷を有する分子、例えば金属イオン、荷電アミノ酸側鎖、または他の陽イオンもしくは陰イオン等のイオンが含まれる。電子標識は、例えば、標識が検出器の中に配置された時に伝導率(または抵抗率)によって検出することができる。いくつかの実施形態では、ナノチャンネルが、チャンネル中に配置された物質の伝導率または抵抗率を決定することにより電子標識の有無を決定するように構成された電極を含む。いくつかの実施形態では、非光学的標識は、金属、金属酸化物(例えば、金属酸化物)、または酸化ケイ素部分を含む。いくつかの実施形態では、非光学的標識は、金属、金属酸化物、または他の酸化物を含む部分(例えばナノ粒子)を含む。特定の金属または酸化物部分の存在は、例えば核磁気共鳴によって検出することができる。いくつかの実施形態では、標識は、特定の条件(例えば、pH変化)で部分、例えばプロトンまたは陰イオンを放出するように構成され、放出される部分の有無が検出される。

【0026】

可逆的ターミネーターを有するヌクレオチドは、第1のホスホジエステル結合を形成し得るが、ターミネーターの破棄の前には、第2のホスホジエステル結合をすることはできない(または、形成能が限定されている)。したがって、可逆的ターミネーターを有するヌクレオチドはポリヌクレオチド(例えば、ニック部位)に取り込まれ得るが、このヌクレオチドは、ターミネーターが破棄されるまで下流のホスホジエステル結合を形成することができない。破棄は当業者に公知の技術を用いて行われ得る。例えば、ターミネーターは、例えば電磁放射によって切断できる切断可能なリンカーを介してヌクレオチドに結合

していよい。3'可逆的ターミネーターを含む標識化されたヌクレオチドを用いてニック修復が行われた場合、単一の標識化されたヌクレオチドがニックに取り込まれ得るが、ターミネーターは、更なる標識化されたヌクレオチドがニックに取り込まれるのを防ぐことができる。したがって、ニック標識化は、ニック当たり1個の標識化ヌクレオチドに限定することができる。ニック当たり1個の標識部分にニック標識化を限定することで、同じニックに複数の標識が取り込まれることによる潜在的偏りを最小限に抑えることができる。例えば、標識化がニック当たり1個の標識部分に限定されるようなアプローチを取る場合、標識からの比較的強いシグナルに基づいて非常に密接する2個のニックを分離することができる(すなわち、単に同じニックに2個の標識が取り込まれた可能性を排除することができる)。例えば、ニック数の定量的推定が望まれる場合、ニック当たり1標識のアプローチにより、標識シグナル強度とニック数との間の直接的相関が容易になり得る。可逆的ターミネーターを含むヌクレオチド上の標識は、本明細書に記載されている通りであり得る。いくつかの実施形態では、可逆的ターミネーターを含むヌクレオチドは量子ドットを含む。いくつかの実施形態では、可逆的ターミネーターを含むヌクレオチドはフルオロフォアを含む。いくつかの実施形態では、可逆的ターミネーターを含むヌクレオチドは非光学的標識を含む。

10

【0027】

本明細書に記載の方法において使用できる様々な種がタグとして機能し得る。タグとしては、例えばフルオロフォア、量子ドット、デンドリマー、ナノワイヤー、ビーズ、ペプチド、タンパク質、磁性ビーズ、または放射標識が含まれる。ポリヌクレオチドを切断しない方法でポリヌクレオチドを標識化するために、種々の他の物質を用いることができ、そのような物質としては、メチルトランスフェラーゼと適用される標識化メチル基、非切断制限酵素、ジンクフィンガータンパク質、抗体、転写因子、DNA結合タンパク質、ヘアピンポリアミド、三本鎖形成オリゴデオキシヌクレオチド、ペプチド核酸等が含まれる。方法は、2種以上の異なるタグの使用を含んでよく、したがって、単一分子が複数のタグを含んでもよい。

20

【0028】

プローブは、本明細書に記載されているように、タグを含む核酸(単一または複数)を含み得る。いくつかの実施形態では、プローブは配列特異的である(例えば、AGGCT、または何らかの他の特定の塩基配列)。いくつかの実施形態では、プローブはランダムに作製される。本明細書に記載されているように、プローブは、目的の配列にプローブを結合させたい、あるいは、特定のDNAポリマー上の目的の配列またはその他の領域の上流または下流の配列にプローブを結合させたい(すなわち、目的の領域に隣接するまたはこれを挟むプローブ)というユーザーの望みに基づいて選択または構築することができる。プローブはフラップと同じ長さであってよい(すなわち、最大約1000塩基)。プローブは好ましくは、1~約500塩基の長さ、約1~100塩基、約3~50塩基、または約5~約20塩基の長さである。プローブの融解温度は66~75であり得る。プローブのアニリング温度は融解温度より低くてよく、例えば融解温度より少なくとも約1、例えば1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20(記載値の任意の2つの間の範囲を含む)低くてよい。

30

40

【0029】

いくつかの実施形態では、プローブは、有機フルオロフォア、量子ドット、デンドリマー、ナノワイヤー、ビーズ、Auビーズ、常磁性ビーズ、磁性ビーズ、放射標識、ポリスチレンビーズ、ポリエチレンビーズ、ペプチド、タンパク質、ハプテン、抗体、抗原、ストレプトアビジン、アビジン、ニュートラアビジン、ビオチン、ヌクレオチド、オリゴヌクレオチド、例えば改変された制限酵素、メチルトランスフェラーゼ、ジンクフィンガー結合タンパク質等の配列特異的結合因子、の1以上を含む。いくつかの実施形態では、プローブはフルオロフォア-クエンチャー対を含む。プローブの形態の1つは、プローブの第1の末端に結合したフルオロフォアおよびプローブの第2の末端に繋ぎ止められた適切

50

なクエンチャーを含み得る。したがって、プローブがハイブリダイズしていない時、クエンチャーはフルオフォアの蛍光発光を防ぎ得、プローブが標的配列にハイブリダイズしている時、プローブは直線化されるので、クエンチャーがフルオロフォアから遠ざかり、適切な波長の電磁放射によって励起された時にフルオロフォアが蛍光発光することができる。いくつかの実施形態では、第1のプローブが、FRET対である第1のフルオロフォアを含み、第2のプローブが、FRET対である第2のフルオロフォアを含む。したがって、第1のプローブおよび第2のプローブの単一フラップへのまたは互いにFRET半径内にある1対のフラップへのハイブリダイゼーションにより、FRETによるエネルギー移動が可能になり得る。いくつかの実施形態では、第1のプローブが、FRET対である第1のフルオロフォアを含み、対応するギャップを埋めるために取り込まれるヌクレオチド上のタグが、FRET対である第2のフルオロフォアを含み得る。したがって、第1のプローブのフラップへのハイブリダイゼーションおよび対応するギャップ中に取り込まれるタグ化ヌクレオチドにより、FRETによるエネルギー移動が可能になり得る。図2を参照すると、タグ化薬剤AまたはBが、取込みまたはハイブリダイゼーション中に、伸長した「フラップ」DNAと結合し得る。タグ化薬剤Cが、伸長しているDNA鎖と結合し得る。こうして、特異的FRETシグナルにより共局在イベントを検出することができる。

【0030】

典型的な標識化方法の模式図を図1Aおよび1Bに示す。図1Aに、フラップの作成および生じたギャップの埋め戻し(back-filling)が示されている。いくつかの実施形態では、ギャップは、放射標識されたヌクレオチド、いわゆる「ホット」または標識塩基で埋め戻され、フラップは、フラップの少なくとも一部に相補的な1以上のプローブと接触させることができる。配列特異的なニッキングエンドヌクレアーゼ、またはニッカーゼは、二本鎖DNA上に一本鎖切断ギャップを形成し、ポリメラーゼがニック形成部位に結合し、置換された鎖またはいわゆる「剥がれたフラップ(peeled flap)」を同時に作りながら、鎖の伸長を開始する。その後、剥がれたフラップは、標識プローブとの特異的ハイブリダイゼーションをシーケンシングして検出可能且つ同定可能なシグナルを生成するために利用可能な領域(すなわち、核酸ポリマーの第2のDNA鎖上のハイブリダイズしていない対応する領域)を作る。

【0031】

図1Bは、ナノチャンネル内で直線状にほどかれている標識化された大きなゲノムDNAを示している。図の下に示されているように、蛍光標識化されたフラップにより、ユーザーは、より大きな巨大分子中でのプローブの位置を可視化することができる。図示されているように、ニック標識化された巨大分子はナノチャンネル内で直線化され得る。タグからシグナルまでの空間的距離は一定であり、その後定量化され、これが、解析中の領域に関する特異的ゲノム配列情報を反映したユニークな「バーコード化」サインパターンを提供する。ラムダdsDNA(全長48.5kbp)上の複数のニッキング部位が、特異的酵素により作成される例として示されており、酵素にはNb.BbvCI; Nb.BsmI; Nb.BsrDI; Nb.BtsI; Nt.AlwI; Nt.BbvCI; Nt.BspQI; Nt.BstNBI; Nt.CviPII、および上記の任意の組合せ消化が含まれ得るが、これに限定されない。予想されるニッカーゼにより形成される位置にハイブリダイズした蛍光標識されたオリゴヌクレオチドプローブを示すために、直線化した単一ラムダDNA画像を含める。長いパイオポリマーに沿ったこのような記録された実際のバーコードは本明細書の別の場所で、観察されたバーコードとして記載されている。いくつかの実施形態では、標識化されたフラップ、標識化されたギャップ、または両方を有する巨大分子を直線化することにより、ユーザーは、標識の互いに対する相対的位置を決定することができる。分析中のゲノム配列に沿ったバーコードサインパターンを示す図4、5、および6を参照する。

【0032】

いくつかの実施形態では、脆弱(fragile)な部位に基づく断片化を軽減する方法が提供される。いくつかの実施形態では、標識の取込み率を制限するため、低減された

10

20

30

40

50

駆動条件 (driving condition) が用いられ、したがって、脆弱部位での断片化を最小限に抑える。いくつかの実施形態では、低減された駆動条件は、DNA伸長に付随する剪断応力を最小限に抑えることができる。いくつかの実施形態では、駆動は、dNTPの濃度の低減、反応温度の低下、補因子濃度の低下、バッファおよび塩濃度の調整、またはその組合せにより低減される。高濃度のdTNPでポリメラーゼエキソヌクレアーゼ活性を刺激した後、少なくとも1つのヌクレオチドを制限または省略することで伸長を制限することにより、修復レベルで駆動を低減することもできる(「チョークド修復(choked repair)」と呼ばれることもある)。いくつかの実施形態では、単一種のdNTP(例えば、dATP)がニック部位で取り込まれ、伸長なしにフラップヌクレアーゼによりフラップが除去され、ライゲーションが行われる。

10

【0033】

いくつかの実施形態では、好熱性ポリメラーゼの最適以下の温度を用いて、駆動条件を低減する。いくつかの実施形態では、反応温度は、約35～約75、例えば35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、または75である。いくつかの実施形態では、温度は約50～約55、約55～約60、約60～約65、または約50～約65である。

【0034】

20

いくつかの実施形態では、本明細書で使用されるポリメラーゼは耐熱性である。いくつかの実施形態では、ポリメラーゼは中温性である。いくつかの実施形態では、ポリメラーゼは校正能を有さない。いくつかの実施形態では、ポリメラーゼは鎖置換能を有する。いくつかの実施形態では、ポリメラーゼは5'-3'エキソヌクレアーゼ活性を有する。いくつかの実施形態では、ポリメラーゼは校正能を有さないが、鎖置換能および5'-3'エキソヌクレアーゼ活性を有する。

【0035】

いくつかの実施形態では、同じ配列モチーフを標的とするが、逆鎖にニックを入れるニックアーゼを用いて特異的DNA鎖を標的化して、脆弱な部位の形成を最小限に抑える。いくつかの実施形態では、ニックアーゼは、二本鎖DNAの一本鎖にのみ結合するように修飾されている。いくつかの実施形態では、ニックアーゼは、第1のDNA分子由来の一本鎖および第2のDNA分子由来の一本鎖を標的とするために用いられる。これらの実施形態のいくつかでは、第1のDNA由来の一本鎖が第1のニックアーゼにより標的化され、第2のDNA分子由来の相補鎖が、第1のニックアーゼと同じ配列モチーフを認識する第2のニックアーゼにより標的化される。いくつかの実施形態では、伸長の方向は鎖の一方について逆転される。例えば、いくつかの実施形態では、ニックング部位からの伸長は第1のDNA分子で1つの方向に起こり、第2のDNA分子では反対方向に起こる。いくつかの実施形態では、ニックング部位からの伸長は、DNA分子の上の鎖(top strand)で1つの方向に起こり、同じDNA分子の下鎖(bottom strand)では反対方向に起こる。

30

40

【0036】

いくつかの実施形態では、目的のポリヌクレオチド中において第2のモチーフが調査される。いくつかの実施形態では、第2のモチーフは、非切断制限酵素、ジンクフィンガータンパク質、抗体、転写因子、転写活性化因子様ドメイン、DNA結合タンパク質、ポリアミド、三重らせん形成オリゴヌクレオチド、およびペプチド核酸から選択される結合実体の少なくとも1つの結合部位を含む。いくつかの実施形態では、第2の標識を含む結合実体を用いて第2のモチーフのマーキングまたはタグ化がなされる。いくつかの実施形態では、マーキングは、DNAを切断しないまたはDNAにニックを入れない標識を用いて行われる。いくつかの実施形態では、タグ化は、DNAを切断しないまたはDNAにニックを入れない標識を用いて行われる。

50

【 0 0 3 7 】

いくつかの実施形態では、第 1 のポリヌクレオチドが第 1 の標識で標識化され、第 2 のポリヌクレオチドが第 2 の標識で標識化される。いくつかの実施形態では、第 1 および第 2 の標識は同じである。いくつかの実施形態では、第 1 および第 2 の標識は異なる。いくつかの実施形態では、第 3 のポリヌクレオチドが第 3 の標識で標識化される。第 3 の標識は、第 1 および第 2 の標識のいずれかまたは両方と同じであってもよく、異なってもよい。

【 0 0 3 8 】

ポリヌクレオチドの直線化

いくつかの実施形態では、ポリヌクレオチドは直線化される。ポリヌクレオチドを直線化する手段は、液体流の剪断力、キャピラリーフロー、対流、電場、誘電場、熱勾配、磁場、その組合せ（例えば、物理的閉じ込めおよび電場の使用）の使用、または当業者に公知の任意の他の方法を含み得る。いくつかの実施形態では、本明細書に記載のチャンネルの断面寸法はマイクロメートルの範囲である。いくつかの実施形態では、チャンネルの断面寸法はナノメートルの範囲である。ナノチャンネルおよびナノチャンネルの使用を組み込んだ方法の例が、その全体を参照により援用する米国特許出願公開第 2 0 1 1 / 0 1 7 1 6 3 4 号および同第 2 0 1 2 / 0 2 3 7 9 3 6 号に記載されている。

【 0 0 3 9 】

いくつかの実施形態では、ナノチャンネルは、断面直径が約 1 ~ 約 5 0 0 ナノメートルである少なくとも 1 個のチャンネルを含む基質を含む。チャンネルの断面直径は、好ましくは、約 1 0 ~ 約 2 0 0 nm、約 2 0 ~ 約 1 0 0 nm、または約 5 0 nm である。特定のチャンネルの幅および深さは同じでなくてよいが、チャンネルの深さはその幅と同じ範囲であり得る。チャンネルは、1 0 nm からセンチメートルまでの事実上任意の長さのものであり得る。所与の応用に最適な長さは当業者には明らかであると考えられるが、そのようなチャンネルの長さは好ましくはミリメートルの範囲である。

【 0 0 4 0 】

いくつかの実施形態では、ナノチャンネルは固定化領域 (immobilization region) を含む。固定化領域は巨大分子を固定化することができる。いくつかの実施形態では、基質は固定化領域を含む。いくつかの実施形態では、固定化領域は、例えば磁場により、少なくとも 1 個のポリヌクレオチドと可逆的に結合するように構成されている。いくつかの実施形態では、固定化領域は、固定化領域中に少なくとも 1 個のポリヌクレオチドが不可逆的に固定化されるように構成されている。如何なる特定の理論にも限定されないが、ポリヌクレオチドが移動性を維持しながら「直線状固定化 (linear immobilizing) 」されるように、ナノチャンネル中でポリヌクレオチドについてエントロピー平衡が維持され得る。したがって、いくつかの実施形態では、ナノチャンネルは、ナノチャンネル中でポリヌクレオチドの移動性を維持しながら少なくとも 1 個のポリヌクレオチドが「直線状固定化」されるように構成されている。いくつかの実施形態では、ナノチャンネルは固定化領域を含まない。いくつかの実施形態では、ナノチャンネルは固定化領域を含まず、少なくとも 1 個のポリヌクレオチドが「直線状固定化」されるように構成されている。いくつかの実施形態では、ポリヌクレオチドは 1 以上の修飾を含み、これは固定化を容易化でき、フラップ、ビーズ、誘電修飾、磁性粒子等を含み得る。システムおよび巨大分子修飾は、互いに対するそれらの親和性と協調しておよびそれに基づいて選択され得る。典型的な固定化領域としては、磁気領域、化学的活性領域、狭窄部 (constriction) 等が含まれる。

【 0 0 4 1 】

本明細書に記載のポリヌクレオチドは、任意の長さ、例えば少なくとも約 0 . 1 K b、例えば 0 . 1 K b、0 . 2 K b、0 . 3 K b、0 . 4 K b、0 . 5 K b、0 . 6 K b、0 . 7 K b、0 . 8 K b、0 . 9 K b、1 K b、2 K b、3 K b、4 K b、5 K b、6 K b、7 K b、8 K b、9 K b、1 0 K b、2 0 K b、3 0 K b、4 0 K b、5 0 K b、6 0 K b、7 0 K b、8 0 K b、9 0 k b、1 0 0 K b、2 0 0 K b、3 0 0 K b、4 0 0 K

10

20

30

40

50

b、500 Kb、600 Kb、700 Kb、800 Kb、900 Kb、または1000 Kb（列挙されている値の任意の2つの間の範囲を含む）のものであり得る。いくつかの実施形態では、本明細書に記載のポリヌクレオチドの長さは約0.1 Kb～1000 Kb、0.1 Kb～100 Kb、0.1 Kb～10 Kb；0.5 Kb～1000 Kb、0.5 Kb～100 Kb、0.5 Kb～10 Kb、1 Kb～1000 Kb、1 Kb～100 Kb、1 Kb～10 Kb、10 Kb～1000 Kb、または10 Kb～100 Kbである。いくつかの実施形態では、本明細書に記載のポリヌクレオチドの長さは約3.3 Kb～170 Kbである。

【0042】

マップ

いくつかの実施形態では、高分解能の物理地図を構築する。物理地図は、特性解析されているポリヌクレオチド上の標識の位置および標識間の距離（相対または絶対）を含み得る。物理地図は、SNaPshotフィンガープリント法等の別の方法を用いて作成される物理地図の検証または補正に用いることができる。いくつかの実施形態では、物理地図は、アセンブルされた領域を検証するためおよび配列スキャップフォルド中の誤りを補正するために用いられる。いくつかの実施形態では、物理地図は、配列スキャップフォルドを固定することにより領域の新規な配列アセンブリを容易化するために用いられる。いくつかの実施形態では、物理地図は、非常に正確で完全な配列アセンブリを生成するために用いられる。いくつかの実施形態では、物理地図は、異なる個体、個体の異なる部分、または薬物処置等の介入プロセスもしくは生物の異なる時間もしくは段階に由来する配列構造多型を明らかにするために複数のゲノムを比較するために用いられる。いくつかの実施形態では、物理地図は、生物学的サンプル中の外来性因子の有無を検出するために用いられる。いくつかの実施形態では、物理地図は、第2のポリヌクレオチド中への第1のポリヌクレオチドの組込みを同定するために用いられる。いくつかの実施形態では、物理地図は、宿主ポリヌクレオチド中へのウイルスポリヌクレオチドの組込みまたは存在を同定するために用いられる。いくつかの実施形態では、物理地図は、生物学的サンプル中の外来性因子の有無を検出するために用いられる。いくつかの実施形態では、物理地図は、生物学的サンプル中の細菌の存在を検出するために用いられる。いくつかの実施形態では、物理地図は、生物学的サンプル中のウイルスの有無を検出するために用いられる。いくつかの実施形態では、物理地図は、生物学的サンプル中の寄生虫の有無を検出するために用いられる。いくつかの実施形態では、物理地図は、対象中の外来性因子の有無をインビトロで検出するために用いられる。いくつかの実施形態では、物理地図は、対象中の細菌の存在をインビトロで検出するために用いられる。いくつかの実施形態では、物理地図は、対象中のウイルスの有無をインビトロで検出するために用いられる。いくつかの実施形態では、物理地図は、対象中の寄生虫の有無をインビトロで検出するために用いられる。

【0043】

いくつかの実施形態では、宿主ゲノム中に存在しないポリヌクレオチド配列の同定に参照マップが用いられる。いくつかの実施形態では、本明細書に記載されているように少なくとも1つの参照ポリヌクレオチド（例えば、健康な生物もしくは細胞、非処置の生物もしくは生物、またはアウトグループに由来するゲノム配列等の対照ゲノム配列）の少なくとも1つのモチーフを標識化し、ポリヌクレオチド上の標識化された部位を同定することにより参照マップが用意される。したがって、いくつかの実施形態では、参照マップは、標識化された分子または分子のコレクションとして用意される。いくつかの実施形態では、参照マップは、電子的または光学的に保存される参照マップ、例えば画像および/またはデータセットとして用意される。いくつかの実施形態では、特性解析されるポリヌクレオチドの標識化に使用される1以上のモチーフを参照配列上で同定することによりインシリコで参照マップが用意される。したがって、参照配列上で生成される標識化パターンは、特性解析されるポリヌクレオチドのプロセスと合致するプロセスにより参照配列が標識化された場合、予測することができる。いくつかの実施形態では、参照マップのコレクション、例えば、種間または水平遺伝子伝播に結びつけられるウイルス、細菌、および/ま

10

20

30

40

50

たは真菌等の既知の外來性因子の参照マップのコレクションが用意される。いくつかの実施形態では、検出対象である外來性因子の参照ゲノムは、利用可能なシーケンシングされた外來性因子ゲノムから得られ、例えばサンガーセンター (Sanger Institute) の原性生物ゲノムデータベース、全米バイオテクノロジー情報センターのゲノムデータベース、TriTrypDBデータベース、またはデータベースの組合せから得られる。1以上の参照ゲノムが得られれば、標識化部位、例えばニックング部位、メチルトランスフェラーゼ標的モチーフ、またはプローブアニーリング部位を当業者に公知のプログラムを使って予測することができる。標識化部位は、ゲノムマッピングのためにDNAの標識化に現在使用されている全ての酵素について、関連する外來性因子ゲノム内で予測することができ、宿主物質をフィルタリングアウトできる標識化酵素および/または
10
プローブ配列が存在するかどうかを決定することができる。好ましくは、標識化酵素またはプローブの候補は、宿主ゲノムでの標識密度が低く、一方、外來性因子ゲノム内で100kbp当たり約2~約50標識、例えば、外來性因子ゲノム内で100kbp当たり約5~約20標識、100kbp当たり約5~約10標識、100kbp当たり約10~約20標識の標識密度である。いくつかの実施形態では、酵素またはプローブは、外來性因子ゲノム内で100kbp当たり約10標識の標識密度である。いくつかの実施形態では、標識密度は、外來性因子ゲノム内で100kbp当たり約5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20標識である。そのような酵素および/またはプローブは、ヒトと外來性因子物質を区別しないものより好ましい。各標識化酵素および/またはプローブについて、1個の外來性因子ゲノムからなる以上
20
の外來性因子ゲノム由来のコンティグのセットからなる参照マップが作成され得る。いくつかの実施形態では、参照マップは、細菌の外來性因子に特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、真菌外來性因子に特異的なパターンを含む、外來性因子核酸の配列特異的標識化のパターンである。いくつかの実施形態では、参照マップは、ウイルスに特異的なパターンを含む、外來性因子核酸の配列特異的標識化のパターンである。いくつかの実施形態では、参照マップは、寄生虫の外來性因子に特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、ヒト免疫不全ウイルスIに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、ヒト免疫不全ウイルスIIに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、ヒトTリンパ球向性ウイルスIIに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、B型肝炎に特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、C型肝炎に特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、梅毒トレポネーマに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、ウエストナイルウイルスに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、トリパノソーマ・クルージに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、サイトメガロウイルスに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、マラリアに特異的なパターンを含む、外來性因子核酸の配列特異的標識化パターンである。
30
40

【0044】

本明細書に記載のいくつかの実施形態に従って対象中の外來性因子の有無を決定するための基本的配列アラインメントにおいて参照マップを用いることができる。いくつかの実施形態では、宿主のマップとアラインされないポリヌクレオチド標識化パターンの存在の検出が、対象中における外來性因子の存在の決定に用いられる。標識化されたサンプルポリヌクレオチドまたはポリヌクレオチドのコレクションに由来する標識化パターンが宿主
50

参照マップの如何なる標識化パターンともアラインされないアラインメントが行われることもあり、これは、第2の標識化されたポリヌクレオチドまたはポリヌクレオチドのコレクションが宿主ゲノムのマーカーを有さないことを示している。あるいは、またはこれに加えて、いくつかの実施形態では、サンプルポリヌクレオチドまたはポリヌクレオチドのコレクションの標識化パターンを外来性因子マップのものとアラインすることにより外来性因子の存在を決定することができ、外来性因子のポリヌクレオチドマップのマッチは、外来性因子の存在を示し得る。アラインメントは、当業者に周知のアラインメントプログラムによって行うことができ、例えば、D I A L I G N、D I A L I G N - T X、K a l i g n、K a l i g n - S B C、M A F F T - C B R C、M A F F T - E B I、M a x A l l i g n、M u l t i a l i n、M U S C L E、K B L A S T、B L A T、B W A、B W A - S W、B W A - M E M、B F A S T、B o w t i e 2、N o v o a l l i g n、G S N A P、S H R i M P 2、C L U S T A L W、C L U S T A L W - 2、C L U S T A L W - P B I L、R e f A l i g n e r、およびS T A Rが含まれ得る。いくつかの実施形態では、外来性因子の有無について生物学的サンプルをスクリーニングするために参照マップのコレクションが用意される。いくつかの実施形態では、病原体はウイルスである。いくつかの実施形態では、外来性因子は細菌である。いくつかの実施形態では、外来性因子は真菌である。いくつかの実施形態では、外来性因子は寄生虫である。いくつかの実施形態では、外来性因子は病原性である。いくつかの実施形態では、外来性因子はプラスモディウムの種である。いくつかの実施形態では、参照マップは、ヒト免疫不全ウイルスIの1以上の病原性核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マ
ップは、ヒト免疫不全ウイルス2の外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、ヒトTリンパ球向性ウイルスIの外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、ヒトTリンパ球向性ウイルスIIの外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、B型肝炎の外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、C型肝炎の外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、梅毒トレポネーマ（梅毒）の外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップはウエストナイルウイルスの外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、トリパノソーマ・クルージ（シャーガス病）の外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、サイトメガロウイルス（CMV）の外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、参照マップは、プラスモディウム（マラリア）の外来性因子核酸の配列特異的標識化パターンである。いくつかの実施形態では、外来性因子は、ヒト免疫不全ウイルス1、ヒト免疫不全ウイルス2、ヒトTリンパ球向性ウイルスI、ヒトTリンパ球向性ウイルスII、B型肝炎、C型肝炎、梅毒トレポネーマ（梅毒）、ウエストナイルウイルス、トリパノソーマ・クルージ（シャーガス病）、サイトメガロウイルス（CMV）、またはプラスモディウム（マラリア）、または上記の任意の組合せから選択される。

【0045】

いくつかの実施形態では、（例えば、メチルトランスフェラーゼ等の複数の異なる標識化酵素を用いることにより）複数の異なるモチーフが標識化される。いくつかの実施形態では、複数の標識化酵素（例えば、ニックアーゼ）によってニックを入れられた分子が、参照マップを用いてアセンブルされる。いくつかの実施形態では、2以上の標識化工程を用いて情報密度が最大化される。いくつかの実施形態では、2以上の標識化工程に供される1以上の分子が、参照マップを用いてアセンブルされる。

【0046】

いくつかの実施形態では、ポリヌクレオチドはDNAを含む。いくつかの実施形態では、ポリヌクレオチドはRNAを含む。いくつかの実施形態では、ポリヌクレオチドは二本鎖である。いくつかの実施形態では、ポリヌクレオチドは一本鎖である。いくつかの実施

10

20

30

40

50

形態では、ポリヌクレオチドはDNA-RNAハイブリッドを含む。

【0047】

ポリヌクレオチドは、高度に純粋な調製物、クルード、またはセミクルード材料であり得る。ポリヌクレオチドは他のポリヌクレオチドから単離され得る。ポリヌクレオチドは、任意の生物学的ソースに由来してよく、合成であってもよい。いくつかの実施形態では、ポリヌクレオチドは生物学的サンプルから単離される。いくつかの実施形態では、単離されたポリヌクレオチドは、非宿主ポリヌクレオチド、例えば外来性因子ポリヌクレオチドについて濃縮される。いくつかの実施形態では、濃縮は、生物学的サンプル由来の宿主ポリヌクレオチドを除去し、一方、少なくとも宿主に結合している外来性因子ポリヌクレオチドおよび/または宿主から遊離した外来性因子ポリヌクレオチドは保持することを含む。いくつかの実施形態では、宿主核酸が、外来性因子ポリヌクレオチドを濃縮するために、外来性因子核酸を更に含むまたは含み得るサンプルから抽出される。いくつかの実施形態では、外来性因子核酸が、外来性因子ポリヌクレオチドを濃縮するために、宿主核酸を更に含むまたは含み得るサンプルから抽出される。いくつかの実施形態では、宿主核酸が分解、切断、または中和され、一方、外来性因子ポリヌクレオチドは選択的に、分解、切断、または中和されない。いくつかの実施形態（例えば、外来性因子がRNAゲノムを含み、宿主がDNAゲノムを含む場合）では、宿主核酸に特異的なエキソヌクレアーゼが宿主核酸を分解する。いくつかの実施形態では、宿主ゲノムのモチーフまたはエピジェネティックな特徴に特異的なエキソヌクレアーゼが宿主核酸を分解する。

【0048】

本明細書において、「ポリメラーゼ」という用語は、3'ヒドロキシル末端から鋳型依存的に天然または修飾ヌクレオチドを取り込むことができる天然のまたは改変された任意の酵素を指す。

【0049】

本明細書において、「ニッキングエンドヌクレアーゼ」という用語は、規定の配列で単一DNA鎖上のホスホジエステル結合を破壊して3'-ヒドロキシルを残すことができる天然のまたは改変された任意の酵素を指す。ニッキングエンドヌクレアーゼは、1つのDNA鎖についての切断活性が消失するように制限酵素を修飾することにより改変することができ、あるいは、DNA結合ドメイン、例えばジンクフィンガーおよび転写活性化因子様エフェクター由来のDNA認識ドメインにニッキングサブユニットを融合させることにより生成することができる。

【0050】

ゲノム外因子の同定および特性解析

いくつかの実施形態は、ゲノム外因子を同定および特性解析する方法を含む。本明細書において「ゲノム外因子 (extragenomic element)」とは、宿主ゲノム以外のソースに由来するポリヌクレオチド配列を指す。典型的なゲノム外因子としては、宿主ゲノム中へのウイルス性挿入物、水平または種間遺伝子伝播により宿主ゲノムに挿入された他の生物のゲノム物質、オルガネラ、ミニ染色体、プラスミド等のゲノム物質が含まれる。いくつかの実施形態では、ゲノム外因子は、外来性因子のゲノム物質を含む。いくつかの実施形態では、ゲノム外因子は、外来性因子の完全ゲノムを含む。いくつかの実施形態では、ゲノム外因子は、外来性因子のゲノムの一部を含む。

【0051】

いくつかの実施形態では、方法は、既知のゲノム外因子を同定する工程を含む。方法は、宿主生物由来の少なくとも1つのポリヌクレオチドを部位特異的に標識化する工程を含み得る。方法は、1以上のポリヌクレオチドを直線化する工程を含み得る。方法は、ポリヌクレオチド上または複数のポリヌクレオチドのそれぞれの上の部位特異的標識化のパターンを同定する工程を含み得る。方法は、宿主生物由来のポリヌクレオチドの部位特異的標識化パターンを、複数の既知のゲノム外ポリヌクレオチドの参照配列上の部位特異的標識化パターンと比較する工程を含み得る。いくつかの実施形態では、参照配列のポリヌクレオチドは、宿主生物由来のポリヌクレオチド(1以上)と同じ特異的部位の部位特異的

標識化を受けて部位特異的標識を生成する。いくつかの実施形態では、宿主生物由来のポリヌクレオチド（１以上）の部位特異的標識化のモチーフに基づいて参照配列のポリヌクレオチド標識化パターンがインシリコで用意される。いくつかの実施形態では、参照配列は宿主生物に由来しない。例えば、宿主生物がヒトの場合、参照配列は、細菌、真菌、寄生虫、および／またはウイルスの配列を含み得る。いくつかの実施形態では、参照配列は、ヒト免疫不全ウイルス１、ヒト免疫不全ウイルス２、ヒトＴリンパ球向性ウイルスⅠ、ヒトＴリンパ球向性ウイルスⅡ、Ｂ型肝炎、Ｃ型肝炎、梅毒トレポネーマ（梅毒）、ウエストナイルウイルス、トリパノソーマ・クルージ（シャーガス病）、サイトメガロウイルス（ＣＭＶ）、またはマラリアに特異的なパターンを含む。同一性（例えば、部位特異的標識化パターンのアラインメントにより決定される）により、宿主生物由来のポリヌクレオチド（１以上）中におけるゲノム外物質の存在が示され得る。異なるポリヌクレオチドのゲノム外配列間の同一性により、存在可能性のあるゲノム外物質のサブバリエーション、例えば一部が変異を獲得している複数コピーのレトロウイルスが示され得る。ゲノム外因子の存在量は、宿主配列とのアラインメントの数（塩基当たりのカバレッジ深度）と比較したゲノム外配列とのアラインメントの数により推測することができる。いくつかの実施形態では、宿主生物中のウイルスコピー数が決定される。いくつかの実施形態では、宿主ゲノム中のウイルス性因子がマッピングされる。いくつかの実施形態では、ウイルス系統分類（strain typing）がなされる。

10

【００５２】

いくつかの実施形態では、宿主ゲノム中におけるゲノム外配列の挿入位置が同定される。ゲノム外配列に隣接する配列の部位特異的標識化パターンが生成され得、宿主ゲノム由来の参照配列の部位特異的標識化パターンとアラインされ得る。宿主ゲノムの領域と隣接配列のパターンとの間の同一性により、隣接配列に隣接されるゲノム外配列の挿入部位が示され得る。いくつかの実施形態では、宿主ゲノム中の複数のゲノム外因子の挿入部位、例えば少なくとも約２、３、４、５、６、７、８、９、１０、２０、３０、４０、５０、６０、７０、８０、９０、１００、１５０、２００、２５０、３００、３５０、４００、４５０、５００、５５０、６００、６５０、７００、７５０、８００、８５０、９００、９５０、または１０００個（上記値の任意の２つの間の範囲を含む）のゲノム外因子の挿入部位が同定される。

20

【００５３】

いくつかの実施形態では、宿主生物は哺乳動物である。いくつかの実施形態では、宿主生物はマウスである。いくつかの実施形態では、宿主生物はヒトである。いくつかの実施形態では、宿主生物は植物である。いくつかの実施形態では、宿主生物は細菌である。いくつかの実施形態では、ゲノム外配列は、宿主種に感染することが知られている外来性因子のものである。例えば、ファージが細菌に感染することができる。例えば、*A. tumefaciens* が双子葉植物に感染することができる。

30

【００５４】

いくつかの実施形態では、方法は、宿主ゲノムに結合しているゲノム外配列を新規に同定することを含む。宿主ゲノムに結合しているポリヌクレオチドは、本明細書に記載されているように部位特異的標識化および直線化され得る。いくつかの実施形態では、宿主ゲノムに結合している複数のポリヌクレオチドの部位特異的標識化パターンが、宿主ゲノムの参照配列の部位特異的標識化パターンと比較され、類似性の非存在が決定される。したがって、ポリヌクレオチドは、宿主ゲノムにとってゲノム外の物質を含んでいると決定され得る。いくつかの実施形態では、宿主ゲノムの参照配列のパターンは、例えば、部位特異的標識化によって標識化されたであろうモチーフをインシリコで同定することにより、インシリコで生成される。いくつかの実施形態では、本明細書に記載されているように、参照配列を含むポリヌクレオチドを標識化および直線化し、パターンを検出する。方法は、ゲノム外配列を含むポリヌクレオチドの部位特異的標識化パターンをアラインしてゲノム外配列に特徴的な１以上のコンセンサス「サイン」パターンを同定することを含む得る。いくつかの実施形態では、ゲノム外配列を含む単一ポリヌクレオチドの部位特異的標識

40

50

化パターンをアラインメントと比較して、コンセンサス配列を示し得るコンセンサスパターンの全部または一部をポリヌクレオチドが含むかどうかを決定する（例えば、宿主ゲノム中の複数コピーのレトロトランスポゾン）。いくつかの実施形態では、サインパターン（例えば隣接配列）により特性解析されない物質をポリヌクレオチドが含むかどうかの決定がなされる。いくつかの実施形態では、隣接配列は宿主生物に由来し、したがって、サインパターンを含まない。その代わり、隣接配列は、宿主ゲノムの領域に特徴的なパターンを含み得る。したがって、隣接配列上の部位特異的標識化パターンのアラインメントを、ゲノム外配列が宿主ゲノムに組み込まれた部位の決定に用いることができる。比較的長いポリヌクレオチドが宿主ゲノム中の位置に挿入されていた場合、隣接配列の同定は、挿入されたポリヌクレオチドおよび少なくとも一部の隣接配列を含む単一ポリヌクレオチドのパターンの分析を含み得る。したがって、いくつかの実施形態では、各末端上に隣接配列を含む単一ポリヌクレオチドが分析される。いくつかの実施形態では、単一ポリヌクレオチドは、例えば各末端上の隣接配列および挿入されたポリヌクレオチドを含む単一分子を維持するために、断片化、剪断、または切断を受けない。いくつかの実施形態では、単一ポリヌクレオチドの長さは少なくとも約 3 . 3 K b、例えば少なくとも約 3 . 3 K b、3 . 5 K b、4 K b、5 K b、6 K b、7 K b、8 K b、9 K b、10 K b、15 K、20 K b、25 K、30 K b、35 K b、40 k b、45 K b、50 k b、60 K b、70 K b、80 K b、90 K b、100 K b、150 k b、200 K b、250 K b、300 K b、350 K b、400 K b、450 K b、500 K b、600 K b、700 K b、800 K b、900 K b、1000 K b、1500 K b、または 2000 K b（上記値の任意の 2 つの間の範囲を含む）である。いくつかの実施形態では、挿入されたポリヌクレオチドと宿主ゲノム由来の隣接配列の間の接合部が、分析されているポリヌクレオチド中に保存されている。

【 0 0 5 5 】

いくつかの実施形態では、ゲノム外配列、例えばウイルスの配列、寄生虫の配列、または細菌の配列の組込み部位が同定される。いくつかの実施形態では、同定される配列は、ヒト免疫不全ウイルス 1、ヒト免疫不全ウイルス 2、ヒト T リンパ球向性ウイルス I、ヒト T リンパ球向性ウイルス II、B 型肝炎、C 型肝炎、梅毒トレポネーマ（梅毒）、ウエストナイルウイルス、トリパノソーマ・クルージ（シャーガス病）、サイトメガロウイルス（CMV）、またはマラリアに由来する。したがって、1 以上の宿主ゲノム座位の破壊が推測され得る。いくつかの実施形態では、宿主ゲノム中の組込み部位は、標的組込みのための配列を表す。いくつかの実施形態では、各ゲノム外配列の挿入部位は本明細書に記載されているように決定される。いくつかの実施形態では、本明細書に記載されているように宿主ゲノムの長分子ゲノムマップが作成され、宿主ゲノム中の種々の位置でゲノム外配列のサインが同定される。いくつかの実施形態では、部位特異的に標識化された宿主ゲノム配列を、宿主の種の参照ゲノム配列と比較する（例えばインシリコでのゲノムアセンブリ）。

【 0 0 5 6 】

いくつかの実施形態では、本明細書に記載されているように特定のゲノム外配列またはゲノム外配列ファミリーの挿入部位のパターンが決定される。いくつかの実施形態では、疾患細胞の起源が決定される。例えば、レトロウイルス配列が宿主ゲノムに組み込まれて宿主ゲノム中の更なる位置へと増殖し得るが、宿主ゲノムから容易に切り出すことはできない。したがって、レトロウイルス挿入部位の蓄積は、細胞系統およびそのゲノム機能変化または形質転換の状態を示し得る。いくつかの実施形態では、2 つ以上の異なる細胞における蓄積された挿入部位のパターンは、これらの細胞間の系統関係の同定に用いることができる。いくつかの実施形態では、第 3 の細胞、例えばアウトグループのパターンとの比較がなされる。いくつかの実施形態では、本明細書に記載されているように部位特異的標識化、直線化、および検出により 1 以上のパターンが同定される。いくつかの実施形態では、パターンの 1 以上がインシリコで検出される。いくつかの実施形態では、2 つの細胞は、同じ種の 2 つの異なる生物に由来する。いくつかの実施形態では、2 つの細胞は、

同じ種の生物の2つの細胞である。1つの細胞は第1の系統 (l i n e a g e) に由来し得、1つの細胞は宿主生物の第2の系統に由来し得る。

【 0 0 5 7 】

いくつかの実施形態では、細胞は疾患状態にある。いくつかの実施形態では、細胞は癌細胞である。細胞へのオンコジーンのウイルス性送達、および/またはウイルス挿入による細胞中での腫瘍抑制因子遺伝子の破壊が癌疾患状態の一因であり得る。したがって、いくつかの実施形態では、癌細胞中でのレトロウイルス組込み部位のパターンは、細胞系統の同定に用いることができるだけでなく、その癌表現型の一因となる細胞への機能的変化または癌表現型を発症する可能性を同定することができる。いくつかの実施形態では、健康細胞またはその系統が癌性になる可能性を推定するために、同じ宿主の癌細胞中および健康細胞中の組込み部位のパターンが同定される。いくつかの実施形態では、癌細胞は、1以上のゲノム再配列、例えば欠失、複製、挿入、逆位、および/または転座を含む。いくつかの実施形態では、ゲノム再配列に特徴的なパターンを用いて癌細胞の細胞系統が決定される。いくつかの実施形態では、ゲノム再配列に特徴的なパターンを用いて、健康細胞と癌細胞の間の系統関係が決定され、例えば、健康細胞が癌を発症するリスクが決定される。

10

【 0 0 5 8 】

いくつかの実施形態では、特定のゲノム再配列イベントに特徴的な少なくとも1つのパターンが、不均一集団中の細胞集団のサブセットに関連付けられる。そのような特異的模式は、癌の進行およびクローン増殖中に少なくとも1つのサブセットの細胞が増殖優位性を得る前の一次癌発生 (p r i m a r y c a n c e r o g e n e s i s) の少なくとも1つの初期段階を示し得る。したがって、そのような特異的模式は、癌ゲノムに特徴的な、根底にある少なくとも1つの分子機構を描写し得る。いくつかの実施形態では、一次癌発生の初期段階の少なくとも1つのプロフィールが用意される。如何なる特定の理論にも限定されないが、一次癌発生を生じる細胞集団は高度に不均一であり得る。したがって、いくつかの実施形態では、初期癌発生の1以上のプロフィールは、癌を生じる細胞集団中における高度な不均一性を示し得る。いくつかの実施形態では、ゲノム外配列の末端がサインパターンとアラインされる。したがって、サインパターンとのゲノム外配列の末端のアラインメントにより、ゲノム外配列の末端が宿主ゲノムから遊離していることが示され得る。いくつかの実施形態では、ヒト配列とのアラインメントの数 (塩基当たりのカバレッジ深度) と比べた非ヒト配列とのアラインメントの数により、ゲノム外配列の存在量 (相対および/または絶対コピー数) が推定される。

20

30

【 0 0 5 9 】

いくつかの実施形態では、複数のポリヌクレオチドを含む生物学的サンプル中の特定のゲノム外配列またはゲノム外配列ファミリーの挿入部位のパターンの有無を決定することにより宿主の疾患状態が決定される。いくつかの実施形態では、宿主は哺乳動物である。いくつかの実施形態では、宿主はヒトである。いくつかの実施形態では、生物学的サンプルは血液である。いくつかの実施形態では、外来性因子のゲノム外配列に特徴的なパターンの非存在が疾患の非存在を示す。いくつかの実施形態では、外来性因子のゲノム外配列に特徴的なパターンの存在が疾患状態の存在を示す。いくつかの実施形態では、疾患は細菌性疾患である。いくつかの実施形態では、疾患はウイルス性疾患である。いくつかの実施形態では、疾患は寄生虫疾患である。いくつかの実施形態では、1以上の外来性因子の有無についてスクリーニングするために生物学的サンプルが提供される。いくつかの実施形態では、特定のゲノム再配列イベントに特徴的な少なくとも1つのパターンが細菌性疾患に関連付けられる。いくつかの実施形態では、特定のゲノム再配列イベントに特徴的な少なくとも1つのパターンがウイルス性疾患に関連付けられる。いくつかの実施形態では、特定のゲノム再配列イベントに特徴的な少なくとも1つのパターンが寄生虫疾患に関連付けられる。いくつかの実施形態では、ゲノム再配列イベントに特徴的な少なくとも1つのパターンが、ヒト免疫不全ウイルス1、ヒト免疫不全ウイルス2、ヒトTリンパ球向性ウイルスI、ヒトTリンパ球向性ウイルスII、B型肝炎、C型肝炎、梅毒トレポネマ

40

50

(梅毒)、ウエストナイルウイルス、トリパノソーマ・クルージ(シャーガス病)、サイトメガロウイルス(CMV)、またはマラリアに関連付けられる。したがって、いくつかの実施形態では、宿主の健康プロフィールが決定され得る。いくつかの実施形態では、ゲノム外配列の末端がサインパターンとアラインされる。したがって、サインパターンとのゲノム外配列の末端のアラインメントが、ゲノム外配列の末端が宿主ゲノムから遊離していることを示し得る。いくつかの実施形態では、ヒト配列へのアラインメントの数(塩基当たりのカバレッジ深度)と比べた非ヒト配列へのアラインメントの数により、ゲノム外配列の存在量(相対および/または絶対コピー数)が推測される。

【0060】

図7は、第2のポリヌクレオチドに組み込まれた第1のポリヌクレオチドを特性解析する方法を示すフローダイアグラムである。方法は、第1のポリヌクレオチドを配列特異的に標識化して第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意する工程を含み得る(110)。方法は、第2のポリヌクレオチドを配列特異的に標識化して第2のポリヌクレオチド上の第2の標識の配列特異的なパターンを用意する工程を含み得る、第1および第2の標識は同じであってもよく、異なってもよい(120)。方法は、組み合わせポリヌクレオチドを標識化後に直線化する工程を含み得る(130)。方法は、第1および第2のポリヌクレオチド上の配列特異的な標識のパターンを検出する工程を含み得る(140)。方法は、第1のポリヌクレオチドに特徴的な標識化パターンに基づいて、第1のポリヌクレオチドを含むとして、組み合わせポリヌクレオチドの第1の領域を同定する工程を含み得る(150)。方法は、第1のポリヌクレオチドに近接する第2のポリヌクレオチドの標識化パターンを特性解析して、第1のポリヌクレオチドに特徴的な前記第2のポリヌクレオチドの標識化パターンへの前記第1のポリヌクレオチドの組み込みの位置および方向を同定する工程を含み得る(160)。当業者には、本明細書に開示されているこのおよび他のプロセスおよび方法のために、プロセスおよび方法中で行われる機能が異なる順番で実施されてもよいことが理解されよう。更に、概説されている工程は単に例として提供するものであり、本開示の実施形態の本質を減じることなく、工程および操作のいくつかは、必要に応じて行われるもの(optional)であってもよく、より少ない工程および操作へと組み合わされてもよく、追加的な工程および操作へと拡大されてもよい。

【0061】

ゲノム外配列の同定に加え、本明細書に記載の方法は、宿主ゲノムの配列の重複、再配列、欠失、および/または欠失を同定することができる。したがって、いくつかの実施形態は、宿主の種のゲノム物質の再配列、逆位、重複、または欠失等の宿主ゲノム中の染色体異常を同定する方法を含む。方法は、本明細書に記載されているように宿主生物に由来する少なくとも1つのポリヌクレオチドの部位特異的な標識化およびポリヌクレオチドの直線化を含み得る。方法は、ポリヌクレオチド上の部位特異的な標識化パターンの同定を含み得る。方法は、参照宿主ゲノムの少なくとも一部とポリヌクレオチドのパターンを比較することを含み得る。いくつかの実施形態では、ブレイクポイントに特徴的なポリヌクレオチド上のパターン(例えば、参照宿主ゲノム中で互いに近接していない、ポリヌクレオチド上の宿主ゲノムの第2のパターンに近接する宿主ゲノムの第1のパターン)により、宿主ゲノムの一部の転座、逆位、欠失、または重複が示される。いくつかの実施形態では、宿主ゲノム上の対応する配列のコピー数と比較した、ポリヌクレオチドの少なくとも一部のコピー数の存在量(相対または絶対)を定量化することにより、欠失または重複が同定される。いくつかの実施形態では、例えば欠失または重複のサイズを決定するためおよび/または重複が縦列重複であるかどうかもしくは重複がゲノム中の別の場所に位置しているかどうかを決定するために、重複または欠失のブレイクポイントを決定することにより、欠失または重複が更に特性解析される。典型的な縦列重複の同定が実施例2に記載されており、図4に示されている。

【0062】

いくつかの実施形態では、個別化された(personalized)癌の診断および

ノまたは予測方法が提供される。腫瘍DNA異常を1分子感度で特性解析することができる。ナノチャンネルチップおよび分析装置を用いた、複数の異なる癌系、例えば3つの異なる癌細胞系に由来する1DNA分子の動的イメージングにより、腫瘍細胞の構造多型を決定することができる。いくつかの実施形態では、大きな集団の癌ゲノムの構造多型のゲノムワイドな調査が行われる。

【0063】

図8は、宿主ゲノムに結合する少なくとも第1のポリヌクレオチドを特性解析する方法を示すフローダイアグラムである。方法は、配列特異的に第1のポリヌクレオチドを標識化して第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意する工程を含み得る(210)。方法は、第1のポリヌクレオチドを標識化後に直線化する工程を含み得る(220)。方法は、第1のポリヌクレオチド上の配列特異的標識のパターンを検出する工程を含み得る(230)。方法は、第1のポリヌクレオチド上の宿主ゲノムに特徴的なパターンの非存在に基づいて宿主ゲノム中の第1のポリヌクレオチドの有無を決定する工程を含み得る(240)。当業者には、本明細書に開示されているこのおよび他のプロセスおよび方法のために、プロセスおよび方法で行われる機能が異なる順番で実施されてもよいことが理解されよう。更に、概説されている工程は単に例として提供するものであり、本開示の実施形態の本質を減じることなく、工程および操作の一部が、必要に応じて行われてもよく、より少ない工程および操作へと組み合わされてもよく、追加的な工程および操作へと拡大されてもよい。

【0064】

いくつかの実施形態では、1以上の外来性因子の有無について生物学的サンプルをスクリーニングする方法が提供される。いくつかの実施形態では、方法は、外来性因子核酸の配列特異的標識化の参照パターンを少なくとも1つ用意することを含む。いくつかの実施形態では、方法は、外来性因子核酸配列の配列特異的標識化参照パターンの外来性因子参照ライブラリーを提供することを含む。

【0065】

いくつかの実施形態では、ウイルス感染、細菌感染、真菌感染、寄生虫感染、種間遺伝子伝播、遺伝子治療ベクターの送達等の診断またはモニタリングのための方法が提供される。方法は、宿主生物由来の少なくとも1つのポリヌクレオチドの部位特異的標識化を行う工程を含み得る。方法は、1以上のポリヌクレオチドを直線化する工程を含み得る。方法は、ポリヌクレオチド上またはポリヌクレオチドのそれぞれの上の部位特異的標識化パターンを同定する工程を含み得る。方法は、部位特異的標識化のパターンを、既知のゲノム外ポリヌクレオチドの複数の参照配列上の部位特異的標識化パターンと比較する工程を含み得る。いくつかの実施形態では、参照配列のポリヌクレオチドは、宿主生物由来の1以上のポリヌクレオチドと同じ特異的部位の部位特異的標識化を受けて部位特異的標識を生成する。いくつかの実施形態では、宿主ゲノム中における既知ウイルスに対応する配列の有無によってウイルス感染が診断される。いくつかの実施形態では、ウイルス挿入部位が決定される。いくつかの実施形態では、所望の配列、例えばRNAiコンストラクトまたは遺伝子治療コンストラクト等の治療薬のウイルス送達がモニタリングされる。いくつかの実施形態では、種間遺伝子伝播(例えば、T. cruziからヒトまたはサルモネラからヒト)がモニタリングされる。

【0066】

宿主ゲノム中のゲノム外配列の同定は、ゲノム外配列の起源の特性解析を更に含み得る。いくつかの実施形態では、本明細書に記載されているような宿主ゲノムに結合しているゲノム外配列の同定は、種間DNA伝播イベントの同定を含む。宿主生物から単離されたポリヌクレオチドは、本明細書に記載されているように部位特異的に標識化されて直線化され得る。ゲノム外因子に対応するポリヌクレオチド内のパターンが本明細書に記載されているように決定され得る。いくつかの実施形態では、ゲノム外因子が宿主ゲノムに挿入されている。ゲノム外配列に隣接する配列のパターンおよび宿主ゲノム参照配列のパターンの比較に基づいて、特定のゲノム外配列の挿入部位が同定され得る。いくつかの実施形

態では、ゲノム外因子は宿主ゲノムから遊離している。各ポリヌクレオチドの末端配列のパターンとコンセンサスゲノム外配列のパターンとの比較に基づき、特定のゲノム外配列が宿主ゲノムから遊離しているものとして同定され得る。宿主ゲノムの一部に対応する隣接配列パターンの非存在によっても、宿主ゲノムから遊離しているものとしてポリヌクレオチドが同定され得る。いくつかの実施形態では、ゲノム外配列は、侵入種 (invasive species) の配列、例えば細菌の配列を含む。いくつかの実施形態では、ゲノム外配列は、独立したゲノム成分として残っており、宿主ゲノムに結合しているが、組み込まれていない (例えば、共有結合していない)。例えば、ゲノム外配列は、プラスミドまたはエピソームの一部であってよく、オルガネラ (例えば、ミトコンドリアまたは葉緑体) に埋め込まれていてもよい。いくつかの実施形態では、ゲノム外配列のパターンは、1以上の参照配列のパターンと比較され、ゲノム外配列の起源が同定される。いくつかの実施形態では、ゲノム外配列は水平遺伝子伝播を起源とする。いくつかの実施形態では、ゲノム外配列は種間遺伝子伝播を起源とする。

【0067】

生物学的サンプルのスクリーニング

生物学的サンプルとしては、全血、骨髓、器官組織、血漿、体液、尿、唾液等が含まれるが、これに限定されない。体液には、リンパ液、脳脊髄液、尿、唾液、および当業者によって同定可能な更なる流体が含まれ得る。生物学的サンプルのスクリーニングを用いて、例えば、ドナー組織、器官、および血液が微生物、寄生虫、およびウイルスのゲノム外物質を含まないことを決定することができる。生物学的サンプルのスクリーニングは、ブラッドバンクへの生物学的サンプルの外来性因子汚染を決定するために用いることができる。検出を用いて、輸血、疾患の広がりや伝染され得るおおよび場合によっては汚染された供与材料による重症敗血症、敗血症性ショックにつながり得る、可能性のある感染性物質を防ぐことができる。外来性因子の検出により、供与された生物学的サンプルの使用およびその生物学的サンプル中に外来性因子の存在が陽性検出された対象による更なる供与を防ぐことができる。いくつかの実施形態では、1以上の外来性因子の有無について生物学的サンプルがスクリーニングされる。いくつかの実施形態では、1以上の病原性外来性因子の有無について生物学的サンプルがスクリーニングされる。いくつかの実施形態では、細菌性外来性因子の有無について生物学的サンプルがスクリーニングされる。いくつかの実施形態では、ウイルス性外来性因子の有無について生物学的サンプルがスクリーニングされる。いくつかの実施形態では、寄生虫外来性因子の有無について生物学的サンプルがスクリーニングされる。いくつかの実施形態では、生物学的サンプルは血液を含む。いくつかの実施形態では、サンプルは供与のためのものである。いくつかの実施形態では、方法は、外来性因子核酸の配列特異的標識化の参照パターンを少なくとも1つ用意する工程、複数のポリヌクレオチドを含む生物学的サンプルを用意する工程、複数のポリヌクレオチドを単離する工程、生物学的サンプルのポリヌクレオチドを配列特異的に標識化して、第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意する工程、ポリヌクレオチドを標識化後に直線化する工程、ポリヌクレオチド上の配列特異的標識のパターンを検出する工程、およびポリヌクレオチド中の1以上の外来性因子に特徴的な1以上のパターンの有無を決定することにより生物学的サンプル中の1以上の外来性因子の有無を決定する工程を含む。いくつかの実施形態では、宿主核酸配列から外来性因子核酸配列が区別されるように少なくとも1つの配列特異的標識化パターンが選択される。いくつかの実施形態では、宿主核酸配列のパターンから外来性因子核酸配列のパターンが区別されるように標識密度が選択される。いくつかの実施形態では、少なくとも1つの配列特異的標識化参照パターンは標識密度が100kb当たり約5～約20標識を含むように選択される。いくつかの実施形態では、少なくとも1つの配列特異的標識化参照パターンは、100kb当たり約5標識、約10標識、約15標識、約20標識の標識密度からこれらの値の間のおよそ任意の量の標識密度を有するように選択される。

【0068】

図9は、1以上の外来性因子の有無について生物学的サンプルをスクリーニングする方

10

20

30

40

50

法を示すフローダイアグラムである。方法は、外来性因子核酸の配列特異的標識化の参照パターンを少なくとも1つ用意する工程を含み得る(310)。方法は、複数のポリヌクレオチドを含む生物学的サンプルを用意する工程を含み得る(320)。方法は、複数のポリヌクレオチドを単離する工程を含み得る(330)。方法は、生物学的サンプルのポリヌクレオチドを配列特異的に標識化して、第1のポリヌクレオチド上の配列特異的な第1の標識のパターンを用意する工程を含み得る(340)。方法は、標識化後にポリヌクレオチドを直線化する工程を含み得る(350)。方法は、ポリヌクレオチド上の配列特異的標識のパターンを検出する工程を含み得る(360)。方法は、ポリヌクレオチド上の1以上の外来性因子に特徴的な1以上のパターンの有無を決定することにより、生物学的サンプル中の1以上の外来性因子の有無を決定する工程を含み得る(370)。当業者には、本明細書に開示されているこのおよび他のプロセスおよび方法のために、プロセスおよび方法で行われる機能が異なる順番で行われてもよいことが理解されよう。更に、概説されている工程は単に例として提供されるものであり、本開示の実施形態の本質を減じることなく、工程および操作の一部が、必要に応じて行われるものであってもよく、より少ない工程および操作へと組み合わされてもよく、追加的な工程および操作へと拡大されてもよい。

10

【0069】

核酸サンプルの濃縮

宿主に由来する核酸サンプル、例えばDNAサンプルは、宿主核酸および微生物核酸の両方を含み得る。ウイルス感染の場合、宿主に由来する核酸サンプルは、宿主DNAおよびウイルスDNAの両方を含み得る。細菌感染の場合、宿主に由来する核酸サンプルは、宿主核酸および細菌核酸の両方を含み得る。マイクロビームサンプルまたはウイルス核酸配列サンプルを得るための核酸分析では、宿主ゲノム由来のリードが非常に過剰になり得る。この場合、効率的な分析または診断により適した微生物またはウイルス核酸を得るために、微生物またはウイルス核酸を宿主核酸から分離することがある。

20

【0070】

外来性因子の核酸を検出する方法は大量の宿主ゲノムに由来する読取りによって妨害され得る。外来性因子核酸の存在を決定するための典型的な方法は、それらの微生物叢(microbial community)の同一性、存在量、および機能的特徴を確立するためのDNAシーケンシングおよび分析を含み得る。しかし、生物学的サンプルから作製されたライブラリーは、細菌、ウイルス、および大量の宿主ゲノム核酸の混合物を含み得る。大量の宿主ゲノムDNAの存在は、ウイルスまたは細菌のDNAの種の同定を限定し得る。宿主サンプル中の微生物またはウイルスのDNAを決定するために、サンプルを微生物またはウイルスのDNAについて濃縮してよい。外来性因子は血液中で低頻度で存在すると予想されるため(血液1ml当たり数個(few-several))、濃縮は、例えばヒトDNAの持ち越しによる汚染を最小限に抑えることにより外来性因子検出の感度を増大させるための有用な工程であり得る。したがって、外来性因子を高い信頼性で検出するためにはミリリットル量の血液が処理され得る。健康な血液1ml当たりの平均ヒトDNA含有量は約35μg、範囲は20~70μgであり、実質的に除去されなければ、数個(few/several)の外来性因子ゲノム(ピコグラム量)は容易に埋もれてしまう。

30

40

【0071】

CpGメチル化密度の差で細菌DNAから宿主DNAを除去する方法により、メチル化宿主DNAから微生物核酸を選択的に濃縮するアプローチが行われ得る。この場合、プロテインA磁性ビーズに強力に結合できるヒト抗体のFc領域に融合させたメチルCpG結合ドメイン(MBD)を用いて微生物DNAからメチル化された宿主DNAを分離するために、MBDが用いられ得る。この技術は、分析のために細菌DNAから宿主DNAを除去するために用いることができる効果的な1工程濃縮法であり、その全体を参照により援用するFeehery, G.R. (2013) PLOS ONE. Vol 8: 1-13に説明されている。

【0072】

50

いくつかの実施形態では、生物学的サンプルのスクリーニング方法は、存在可能性のある外来性因子核酸についてサンプルを濃縮することを更に含む。いくつかの実施形態では、濃縮は、サンプルから宿主ポリヌクレオチドを実質的に除去することを含む。その後、標識化が行われ得、これはニック標識化、非切断標識を用いたタグ化（例えば、メチルトランスフェラーゼ、非切断制限酵素、ジンクフィンガータンパク質、抗体、転写因子、DNA結合タンパク質、ヘアピンポリアミド、三本鎖形成オリゴデオキシヌクレオチド、またはペプチド核酸を用いた標識化）、および/またはプローブを用いた標識化または1以上を含む。

【0073】

対象中における外来性因子の有無の決定

10

いくつかの実施形態では、1以上の外来性因子について生物学的サンプルをスクリーニングする方法が提供される。外来性因子の有無についてのマーカーは、将来の宿主の個別化されたケアに用いることができ、あるいは、疾患状態に罹患している対象からの生物学的サンプルの供与の防止につながり得る。例えば、ブラッドバンク等の供与場所で、外来性因子の有無について生物学的サンプルを調べることができる。典型的な寄生虫性外来性因子としては、トリパノソーマ・クルージ、トラパノソーマ・ブルセイ、梅毒トレペネマ、リーシュマニア、プラスモディウム、アピコンプレクサ、およびTrypanosoma speciesが含まれるが、これに限定されない。いくつかの実施形態では、疾患はウイルス性疾患である。いくつかの実施形態では、疾患は細菌性疾患である。いくつかの実施形態では、疾患は寄生虫疾患である。例えば、限定されるものではないが、疾患は、ヒト免疫不全ウイルス(HIV-1、HIV-2)、B型肝炎、C型肝炎、ヒトTリンパ球向性ウイルス(HTLV-1/II)、サイトメガロウイルス、梅毒、ウエストナイルウイルス、マラリア、およびシャーガス病を含み得る。

20

【0074】

図12は、本明細書のいくつかの実施形態に係る、サンプル中の1以上の外来性因子の存在を決定する方法を示すフローダイアグラムである。外来性因子の参照ゲノムのマップが得られた(510)後、宿主哺乳動物中における外来性因子の有無を検出するためにヒトまたは他の哺乳動物ドナーから生物学的サンプルが得られ得る(520)。生物学的サンプルからポリヌクレオチドが単離され得、干渉する宿主ゲノム物質を除去することにより外来性因子結合ポリヌクレオチドが濃縮され得る(530)。その後、濃縮された配列は、例えばニックに基づく標識化、メチルトランスフェラーゼに基づく標識化等により、またはフルオロフォア、量子ドット、デンドリマー、ナノワイヤー、ビーズ、ハプテン、ストレプトアビジン、アビジン、ニュートラアビジン、ビオチン、反応性基を含むプローブの使用により、または当業者に公知のマッピング目的の標識化法のその他の手段により、マッピング目的のために標識化され得る(540)。その後、骨格が、例えばYOYO-1、YO-PRO-1、TOTO-1、TO-PRO-1、PicoGreen(商標)色素、またはSYBR(商標)green等の核酸用の挿入(intercalating)蛍光色素で染色され得る(550)。その後、核酸は、ナノチャンネル内、例えばIRYS(登録商標)システム等のシステム内で直線化され得る。特に、チャンネルの入口で核酸を濃縮するために核酸に電圧が印加されてよく、ナノチャンネル中へと核酸を移動させ、チャンネル中へと核酸を分散させるために、電圧を増大させてよい。ナノチャンネル中に入ると、核酸は静止して均一に伸び(stretch)得る(560)。例えばIrys View等のナノチャンネルに付随するプログラムを用いて、核酸の画像が処理され得る(570)。次いで、単一連続分子中の数百キロベースから1キロベース超に沿った配列モチーフを含む標識化された核酸の高分解能単一分子画像が生成され得る。その後、画像は処理され得、観察された核酸中の相対的標識位置のデジタル表示が生成され得る。次いで、生成されたマップは、生成されたマップが宿主に属さない配列であることを決定するために、ヒトまたは哺乳動物の宿主ゲノムマッチ中のマーカーが、濃縮配列の生成マップとのアラインメント中でマッチし得るかを決定するために用いられ得る。例えば、サンプルの分子の全てが、宿主ゲノム参照マップに対してアラインされ得る。いくつかの実施形態で

30

40

50

は、宿主ゲノム参照マップは、宿主ゲノムの1以上の標識化された分子を含む。いくつかの実施形態では、宿主ゲノム参照マップは、インシリコ生成マップを含む。いくつかの実施形態では、宿主ゲノム参照マップは光学的または電子的に保存される。宿主参照とのアラインメントに成功した分子をフィルタリングアウトし、理論により限定されないが、おそらく外来性因子に由来する、外れ値 (outlier) は維持する (580)。フィルタリング工程を行った後、残った、フィルタリングにかからなかった分子が、例えば RefAligner 等のアラインメント用プログラムを用いて外来性因子参照マップに対してアラインされ得る (590)。しかし、フィルタリングが行われない場合、分子は、宿主ゲノム全体および外来性因子ゲノム混合物の両方を含む複合参照に対してアラインされ得る (560)。いくつかの実施形態では、任意の分子が外来性因子参照とのアラインメントに成功した場合、これは、サンプル中に外来性因子DNAが存在する可能性を示す。必要に応じて、サンプル中の外来性因子物質の割合の推定値を得るために、宿主参照に対してアラインされた総リード数に関して標準化技術が行われ得る。複数の対照で観察されたばらつきならびに種々の濃度の宿主および外来性因子核酸を含むサンプルのバリエーションに基づき、検出限界、定量限界、および外来性因子DNAの割合を推定するための校正パラメーターが決定され得る。外来性因子参照に対して観察されたDNAマッピングの量が検出限界を有意に超えている場合、サンプルは、感染しているまたは外来性因子を含まないとして分類され得る。しかし、固有のばらつきに応じて、感染しているまたは外来性因子を含まないとしてサンプルを分類するためにロジスティック回帰分析が行われ得る。固有のばらつきに応じて、分類にグレーゾーンが設けられてもよい。分類のグレーゾーンのサンプルについては、サンプルが外来性因子を含まないまたは外来性因子を含むとして決定される等のコールはなされない。必要に応じて、グレーゾーンのサンプルを再度ランしてよく、更なるサンプルを収集してもよい。いくつかの実施形態では、検討されている各外来性因子種がサンプル中に存在する確率が関連付けられる。確率が所定の信頼水準を超えるかどうかに応じて、サンプルは、感染しているまたは外来性因子を含まないとして分類され得る。

【0075】

いくつかの実施形態では、複数の外来性因子生物用の複数試験を含む単一パネルが提供される。いくつかの実施形態では、パネルは、少なくとも2つの外来性因子、例えば2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、150、200、250、300、350、400、450、または500個の外来性因子 (記載値の任意の2つの間の範囲を含む) のための試験を含む。単一のニッキング酵素で調製された単一の混合物中で複数の試験を行うことにより、速度、サンプル調製の容易さ、およびアッセイのコストを改善することができ、更に、異なる外来性因子に異なるニッキング酵素を用いることもできる。

【0076】

ゲノム外配列上の部位特異的標識化パターンおよび/または宿主ゲノム中の挿入部位を種々の応用に用いることができる。いくつかの実施形態は、ゲノム外配列のコピー数を決定することを含む。いくつかの実施形態では、宿主ゲノム中のウイルス量が測定される。いくつかの実施形態は、宿主ゲノム中へのウイルスベクター組込みの状態をモニタリングすることを含む。いくつかの実施形態では、遺伝子移入処置理および/または治療薬のウイルス送達がモニタリングされる。いくつかの実施形態は、遺伝子治療 (例えば、遺伝子置換療法) またはアンチセンス療法と組み合わせて、例えば宿主ゲノムの標的領域を同定することによりまたは所望の形質転換体を選択することにより、例えば治療の精度もしくは効率を向上させるためにまたはモニタリングツールとして用いられる。いくつかの実施形態では、遺伝子調節エレメントの機能、例えば遺伝子産物、プロモーター、または siRNA の機能が特性解析される。いくつかの実施形態では、作用可能な診断または治療用のバイオマーカーまたは薬物標的領域として用いるためにゲノム外因子の組込み部位が同定される。いくつかの実施形態では、ウイルス感染が診断される。いくつかの実施形態で

は、感染ウイルスの系統 (s t r a i n) および / または亜系統 (s u b s t r a i n) が同定される。いくつかの実施形態では、1 以上の系統および / または亜系統のウイルス量が測定される。いくつかの実施形態では、細菌感染が診断される。いくつかの実施形態では、寄生虫感染が診断される。いくつかの実施形態では、長分子ゲノムマップにより、癌細胞の直接的構造マッピングが行われる。いくつかの実施形態では、既知配列またはマップ参照との比較アラインメントがなされる。いくつかの実施形態では、例えばウイルス性癌における組込み部位の分析により、疾患起源が決定される。いくつかの実施形態では、標的組込みに関わる配列が同定され、薬物療法の標的として用いられ得る。いくつかの実施形態では、2 つのゲノム外配列間の系統関係が決定され得る。いくつかの実施形態では、2 つの異なる宿主細胞間の系統関係が決定され得る。いくつかの実施形態では、ウイルス感染、癌の発症、または患者ゲノムへの遺伝子治療配列の組込みがモニタリングされ得る。いくつかの実施形態では、細菌感染がモニタリングされ得る。いくつかの実施形態では、寄生虫感染がモニタリングされ得る。

【 0 0 7 7 】

図 1 0 は、1 以上の外来性因子の有無について生物学的サンプルをスクリーニングする方法を示すフローダイアグラムである。方法は、対象から生物学的サンプルを採取する工程を含み得る (4 1 0) 。方法は、1 以上の外来性因子の有無について生物学的サンプルをスクリーニングする工程を含み得る (4 2 0) 。方法は、生物学的サンプル由来のポリヌクレオチドの配列特異的標識のパターンを、少なくとも 1 つの外来性因子核酸の参照配列のインシリコパターンのパターンとアラインすることにより、生物学的サンプル由来のポリヌクレオチドの配列特異的標識の配列パターンを分析する工程を含み得る (4 3 0) 。方法は、パターンの有無に基づいてサンプル中における 1 以上の外来性因子の非存在の存在を決定する工程を含み得る (4 4 0) 。当業者には、本明細書に開示されているこのおよび他のプロセスおよび方法のために、プロセスおよび方法で行われる機能が異なる順番で行われてもよいことが理解されよう。更に、概説されている工程は単に例として提供されるものであり、本開示の実施形態の本質を減じることなく、工程および操作の一部が、必要に応じて行われるものであってもよく、より少ない工程および操作へと組み合わされてもよく、追加的な工程および操作へと拡大されてもよい。

【 0 0 7 8 】

コンピューターシステムおよびコンピューターで実行される技術

本明細書のいくつかの実施形態に従って記載されている技術またはこれらの技術の一部は、ハードウェア、ソフトウェア、ファームウェア、またはその組合せを用いて実行され得る。技術は、ソフトウェアで実行される場合、実行された時に上記方法の 1 以上を行う命令を含むコンピューターにより読取り可能な媒体によって少なくとも一部実現され得る。コンピューター読取り可能な媒体は、コンピュータープログラム製品の一部を形成し得、製品はパッケージング材料を含み得る。コンピューター読取り可能な媒体は、ランダムアクセスメモリー (R A M) 、例えばシンクロナスダイナミックランダムアクセスメモリー (S D R A M) 、リードオンリーメモリー (R O M) 、不揮発性ランダムアクセスメモリー (N V R A M) 、電氣的消去可能プログラマブルリードオンリーメモリー (E E P R O M) 、フラッシュメモリー、磁氣的または光学的データ保存媒体等を含み得る。あるいは、または更に、技術は、命令またはデータ構造の形態のコードを運ぶまたは伝達する、コンピューターによってアクセス、読取り、および / または実行が可能な、コンピューター読取り可能な通信媒体によって少なくとも一部が実現され得る。

【 0 0 7 9 】

本明細書のいくつかの実施形態に係る方法またはこれらの方法の一部は、I n t e l (登録商標) または A M D (登録商標) マイクロプロセッサに基づき M i c r o s o f t W i n d o w s オペレーティングシステムを実行するもの等の任意の従来のホストコンピューターシステム上で実行され得る。U N I X または L I N U X オペレーティングシステムを用いて、I B M (登録商標) 、D E C (登録商標) 、または M o t o r o l a (登録商標) マイクロプロセッサに基づくシステム等のその他のシステムも考えられる。本

明細書に記載のシステムおよび方法は、クライアントサーバーシステムおよび広域ネットワーク、例えばインターネット上で実行されるように実施されてもよい。

【0080】

本明細書のいくつかの実施形態に係る方法、モデル、またはその一部を実施するためのソフトウェアは、任意の周知のコンピューター言語、例えばJava、C、C++、Visual Basic、FORTRAN、またはCOBOLで書かれ得る、任意の周知の互換性のあるコンパイラを用いてコンパイルされ得る。いくつかの実施形態に係るソフトウェアは通常、ホストコンピューターシステムのメモリーに保存された命令から実行される。メモリーまたはコンピューター読取り可能媒体は、ハードディスク、フロッピーディスク、コンパクトディスク、DVD、光磁気ディスク、ランダムアクセスメモリー、リードオンリーメモリー、またはフラッシュメモリーであり得る。本明細書の実施形態に従って使用されるメモリーまたはコンピューター読取り可能媒体は、単一コンピューター内に含まれてもよく、ネットワーク中に分散してもよい。ネットワークは、ローカルエリアネットワーク(LAN)または広域ネットワーク(WAN)等の当該技術分野で公知の複数の従来のネットワークシステムのいずれであってもよい。本明細書の実施形態に従って使用され得るクライアントサーバー環境、データベースサーバー、およびネットワークは当該技術分野で周知である。例えば、データベースサーバーは、リレーショナルデータベース管理システム、ワールドワイドウェブアプリケーション、およびワールドワイドウェブサーバーを実行するUNIX等のオペレーティングシステム上で実行されてよい。その他の種類のメモリーおよびコンピューター読取り可能媒体も本明細書のいくつかの実施形態の範囲内で機能すると考えられる。

【0081】

本明細書のいくつかの実施形態に係る方法またはその一部によって構築されるデータマトリックスは、例えば汎用マークアップ言語(SGML)、ハイパーテキストマークアップ言語(HTML)、または拡張可能なマークアップ言語(XML)を含む、マークアップ言語フォーマットで表され得る。マークアップ言語を用いて、いくつかの実施形態のデータベースまたはデータ構造に保存された情報にタグを付けることができ、これにより、データベースとデータ構造との間でのデータの移動および便利なアノテーションが提供される。特に、XMLフォーマットが反応、反応物質、およびそのアノテーションのデータ表現の構築；例えばネットワークまたはインターネット上での、データベースコンテンツの交換；ドキュメントオブジェクトモデルを用いた個々の要素のアップデート；またはいくつかの実施形態のデータベースもしくはデータ構造の異なる情報コンテンツについて異なるアクセスを複数のユーザーに付与することに有用であり得る。XMLプログラミングの方法およびXMLコードを記述するためのエディターは、例えばRay, Learning XML O'Reilly and Associates(カリフォルニア州セバストポリ；2001)に記載されているように当該技術分野で公知である。

【0082】

本明細書のいくつかの実施形態に係るコンピューターシステムは更に、1以上の反応の表示を受けることができるユーザーインターフェイスを含み得る。本明細書のいくつかの実施形態に係るユーザーインターフェイスは、データ構造を変更するための少なくとも1つのコマンド、制約集合(constraint set)、またはデータ表現に制約集合を適用するためのコマンド、またはその組合せを送ることもできる。インターフェイスは、メニューまたはダイアログボックス等の選択を行うためのグラフィカルな手段を有するグラフィックユーザーインターフェイスであってもよい。インターフェイスは、メインスクリーンからの選択によりアクセス可能な階層化されたスクリーン(layered screen)を用いてアレンジされ得る。ユーザーインターフェイスは、本明細書のいくつかの実施形態に係る有用な他のデータベース、例えば1もしくは複数の核酸もしくはその複数、例えば宿主ゲノム、外来性因子ゲノム、もしくは他のゲノム外配列の配列特異的標識化パターンのインシリコデータベースへのアクセスまたは種々の核酸の配列特異的標識化のパターンに関係する情報を有する他のデータベースへのリンクを提供し得る。更

に、ユーザーインターフェイスは、ゲノム標識化パターンマップのグラフィカル表示を示すことができる。本明細書のいくつかの実施形態に係る標識化のパターンのマップ、アラインメント、または他の情報、例えば表、グラフ、反応ネットワーク、フラックス分布マップ、またはモデルマトリックスとして。

【実施例】

【0083】

以下の実施例は、本発明を説明することを意図しており、明示的にも暗示的にも、本発明を如何なる様式、形状、または形態にも限定することを意図しない。これらは、使用され得るもののうち、典型的なものであるが、当業者に公知の他の手順、方法、または技術が代わりに用いられてもよい。

【0084】

実施例1：ゲノムマップの新規 (de novo) 作成

ヒトゲノムマップを新規アセンブルした。カリフォルニア州サンディエゴの Bionano Genomics 社から入手可能な自動 I R Y S (登録商標) ナノチャンネルチップおよび自動イメージングシステムを用いて、ヒト細胞株由来のゲノム DNA を抽出し、部位特異的標識化した後、直線化し、処理量が 5 ~ 10 Gb / チップラン (3 ~ 4 時間)、更に将来は 50 ~ 200 Gb / チップランとなるナノチャンネル上で、50x のカバレッジで分析した。部位特異的標識を検出し、標識間の距離を決定し、ペアワイズパターンマッチングおよび新規 (de novo) アセンブリプロセスに生分子 (> 20 kb) を用いた。図3に、アセンブルされたヒトゲノムマップを示す。これは巨大で複雑なゲノムからのゲノムマップの作成し易さを示している。このように、本明細書に記載の方法を用いて、巨大で複雑なゲノムからのゲノムマップを作成することができる。

【0085】

実施例2：参照配列と比較した構造多型の同定

実施例1のアセンブルされたゲノム中、数百の大小の構造的バリエーションが検出された。染色体5上の縦列重複の例を図4に示す。部位特異的標識化パターンに基づき、ゲノム配列をアセンブルし(30)、参照ゲノム配列のパターン(32)と比較した。縦列している、縦列する3つの更なるコピーを含むゲノム領域(34)が同定された。このように、部位特異的標識化パターンに基づいて新規ゲノムアセンブリおよび構造バリエーション分析を行うことができる。

【0086】

実施例3：標識化されたDNAの調製

プラグ溶解 (plug lysis) プロトコルを用いて DNA を単離した。当業者には、本明細書に記載の典型的なプロトコルに加えて、当該技術分野で公知の種々の方法を用いて、単離された DNA の好適なサンプルを調製できることが理解されよう。エプスタイン・バーウイルス (EBV) 感染細胞を対数増殖期まで増殖させ、回収した。2% アガロースを沸騰水に完全に溶解するまで 10 ~ 15 分間浸し、使用前に 43 °C のウォーターバス中にアガロースを少なくとも 30 分間置き、50 ml チューブ THERMOMIXER 装置を 50 °C に温め、10 ~ 20 ml の水をウェルに満たしてウォーターバスをシミュレートすることにより、2% アガロースおよびウォーターバスを用意した。4 x 10⁶ 細胞を 15 ml チューブ中にペレット化し、2200 rpm (1000 rcf) で 5 分間スピンしてペレット化することにより細胞を調製した。細胞を、1 x PBS で 2 回洗浄し、750 µl の懸濁バッファーに再懸濁し、43 °C に温めた。250 µl の溶融アガロースを細胞懸濁液に加え、混合し、各プラグモールドを約 95 µl 容積で満たした。プラグを氷上で冷却し、100 µl の Proteinase K を含む 2.5 ml の Proteinase K 溶液を加え、50 °C で 4 時間インキュベートした。2 時間後に更に 100 µl の Proteinase K を加えた。プラグを洗浄バッファーで 4 回、各 15 分間洗浄し、TE (pH 8) で 4 回、各 15 分間、200 rpm 回転、25 °C で洗浄した。Proteinase K / バッファーを除去し、プラグを 70 °C で 2 分間溶融し、43 °C で 5 分間平衡化させ、2 µl の GELase ストック酵素 (0.2 U / µl) を各チュ

ープに加え、混合し、43 で45分間インキュベートした。一度に15 μ lのサンプルを、p10ピペットチップの約5回のストロークで剪断した。次いで、サンプルを、湿らせた透析膜上に滴下し、4時間透析し、蒸発により約40 μ lのサンプルを得た。

【0087】

DNAの部位特異的標識化を以下の通り行った：反応当たり1 μ lの10 \times NEB Bufferと、0.7 μ lの10U/ μ L Nt.BspQと、10 μ lになるようなdH₂Oとを含むNickingマスターミックスを調製した。300ngの調製DNAサンプルをマスターミックスに加え、短時間遠心し、サーマルサイクラー中で37 で2時間インキュベートし、80 およびブロック温度より約10 高い温度に加熱した蓋で20分間熱失活させた後、標識化の用意ができるまで4 に保持した。反応当たり、反応当たり1.5 μ lの1%トリトンX100と、1.5 μ lの10 \times Labeling Mixと、1 μ lの5U/ μ l Taqポリメラーゼと、1 μ lのdH₂Oとを含むLabelingマスターミックスを調製した。5 μ lのLabelingマスターミックスを10 μ lのサンプルに加え、サーマルサイクラー中で72 にて、蓋はブロック温度より約10 高く加熱して、60分間インキュベートした。反応当たり0.2 μ lの10mM dNTPと、0.2 μ lの50mM NAD⁺と、0.75 μ lのPreCRミックスと、0.5 μ lの10 \times Thermo Polと、および3.35 μ lのdH₂Oとを含みリペアマスターミックスを調製した。5 μ lのリペアマスターミックスを15 μ lのサンプルに加え、サーマルサイクラー中で37 にて、蓋はブロックより約10 高い設定で加熱して、18分間インキュベートした。その後、インキュベートしたサンプルを氷上に置き、1 μ lの0.5M EDTAを加え、混合し、短時間(2~3秒)遠心した。

【0088】

サンプル当たり15 μ lの4 \times Flow Bufferと、12 μ lの1M DTTと、2 μ lのDNA Stainと、11 μ lのdH₂Oとを含む、静置マスターミックスを調製した。40 μ lの静置マスターミックスを20 μ lの修復されたDNAサンプルと合わせ、大口径チップで5回混合し、4 で一晩インキュベートした。次いで、小口径P10チップを用いて、Bionano IRYs(商標)システムナノチャネルチップにローディングし、5.5 μ lの調製サンプルを入口(フロント)ウェルに入れ、2分間待った後、5.0 μ lを出口(バック)ウェルへと通した。

【0089】

上記プロトコールで用いた試薬を表1に示す。

【0090】

10

20

30

【表 1】

表 1：試薬リスト

成分	ソース	Cat#	保存	取扱いに関する注意点
10U/ul Nt.BspQI	NEB	R0644S	-20°C	使用前に、3 回反転させ、短時間遠心する。使用時まで氷上またはフリーザーブロック中に維持する。
10x Buffer 3	NEB	B7003S	-20°C	R Tで解凍する。ボルテックスし、短時間遠心する。使用時まで氷上に維持する。
10x Tbermopol Buffer	NEB	B9004S	-20°C	R Tで解凍する。ボルテックスし、短時間遠心する。使用時まで氷上に維持する。
5U/ul Taq	NEB	M0267S	-20°C	使用前に、3 回反転させ、短時間遠心する。使用時まで氷上またはフリーザーブロック中に維持する。
PreCR Kit	NEB	M0309S	-20°C	使用前に、3 回反転させ、短時間遠心する。使用時まで氷上またはフリーザーブロック中に維持する。
50 mM NAD+	NEB	B9007S	-20°C	R Tで解凍する。ボルテックスし、短時間遠心する。使用時まで氷上に維持する。凍結・解凍サイクルの繰り返しを避ける。
10 mM dNTP Mix	BNG	N0447S	-20°C	R Tで解凍する。ボルテックスし、短時間遠心する。使用時まで氷上に維持する。
1 % トリトン X-100	BNG		4°C	-
10X Labeling Mix	BNG		-20°C	R Tで解凍する。ボルテックスし、短時間遠心する。使用時まで氷上に維持する。
4x Flow Buffer	BNG		4°C	ボルテックスし、短時間遠心する。使用時まで R T に維持する。
1M DTT	BNG		-20°C	R Tで解凍する。ボルテックスし、短時間遠心する。使用時まで R T に維持する。
DNA Stain	BNG		-20°C	R Tで解凍する。ボルテックスし、短時間遠心する。氷上では DMSO が結晶化するので、使用時まで R T に維持する。
0.5 M EDTA	BNG		RT	-
Ultrapure dH2O	BNG		RT	-

【 0 0 9 1 】

実施例 4：挿入されたウイルス配列を同定するためのヒトゲノムアセンブリの作成

実施例 3 と同様に調製した E B V 感染リンパ球サンプルから、1 0 0 k b 超の分子について 3 0 0 G b のデータを収集し、新規 (d e n o v o) ヒトゲノムアセンブリに用いた。部位特異的標識化のパターンをアラインし、コンセンサスマップを作成した。いくつかの配列が参照ヒトゲノムアセンブリと対応した。図 5 に示すように、ヒト参照にマッチ

10

20

30

40

50

しない更なるコンセンサスマップが同定された（例えば「ウイルス配列」）。コンセンサスマップの1つがエプスタイン・バーウイルス（EBV）として同定された。ウイルスコンセンサス配列内の差異を系統分類に用いた（図5、点線の四角を参照）。

【0092】

ウイルスコンセンサス配列とアラインしなかった、個々のウイルス配列に隣接する少なくともいくつかの配列（図5、破線の四角参照）が、ウイルス配列のゲノム挿入部位を表し、隣接されたウイルス配列が挿入されたヒトゲノムの位置の決定に用いることができた。

【0093】

このように、実施例4に記載されているリンパ球サンプルのポリヌクレオチドを標識化し、それらを直線化してナノチャネル上で50×のカバレッジで分析する方法は、ウイルスサイン配列の宿主ゲノム組込みを示すゲノムマップを提供し得、系統分類の決定に用いることができる。

【0094】

実施例5：宿主ゲノム中のウイルス量の検出

実施例5と同様にウイルス配列に特徴的であると同定されたパターンのコンセンサスパターンに基づき、コンセンサスウイルスサインパターンが同定された（図6A）。コンセンサスウイルスサインパターンを、参照EBV配列に基づいてインシリコで同定されたパターンと比較した（図6A参照）。

【0095】

ゲノム全体のウイルスサインパターンのコピー数に基づいてヒト宿主ゲノム中のウイルス量を計算した（図6B）。ハプロイドヒト宿主ゲノム当たり約42コピーのレトロウイルスが存在すると決定された。ゲノムをシーケンシングし（50×/ハプロイド）、このウイルス量の計算が確認された。

【0096】

このように、本方法および技術を用いて、ヒト宿主ゲノム中に組み込まれたウイルスDNAのコピー数を決定することができる。

【0097】

実施例6：癌細胞系譜（cell lineage）の分析

1,000名の異なる各患者のリンパ性白血病細胞に由来するゲノムDNAを用意する。実施例3に記載されている全般的手順に従って、各患者の細胞に由来するゲノムDNAを用意し、調製し、標識化し、直線化する。各患者の単一白血病細胞からゲノムパターンをアセンブルし、健康な個体の参照ゲノムパターンと比較する。白血病細胞のゲノム物質の再配列、挿入、および欠失のパターンが同定される。白血病細胞のゲノム中に見られるが、参照ゲノム中には見られないパターンは、癌細胞中のゲノム外因子である可能性があることが決定される。ゲノム外配列のパターンをアラインし、少なくとも1つのゲノム外配列のサインパターンを同定する。この分析は、1,000名の異なる患者それぞれに対して行われる。リンパ性白血病細胞の亜集団のゲノムに共通する構造的特徴が同定される。ウイルス挿入の「ホットスポット」が同定される。再配列の「ホットスポット」が同定される。種々の白血病に関連する構造的特徴に関連するパターンが同定される。

【0098】

各患者に由来する種々の白血病細胞を、構造的特徴の蓄積に基づいてステージ分けする。単一の系譜（lineage）に由来する白血病細胞を、レトロウイルス挿入等の構造的特徴の蓄積に基づいて特性解析する。白血病細胞の系譜がステージ分けされる。

【0099】

これらの健康な個体の造血系譜（hematopoietic lineage）細胞における白血病に関連する構造的特徴の蓄積に基づいて、健康な個体がリンパ性白血病を発症する個別化された素因が推測される。

【0100】

このように、新規（de novo）ナノチャネルゲノムマッピングアセンブリは、特

10

20

30

40

50

に大きなゲノムを調べるために行うことができ、白血病患者のゲノム中の白血病サインパターンの決定に用いられた。

【0101】

実施例7：ヒトサンプル中の外来性因子の検出

ヒトサンプル中の外来性因子を検出するために、シミュレーションに用いた実現可能性研究に用いた外来性因子ゲノムを、トリパノソーマの種々の種およびプラスモディウムを用いて試験した。自動IRYS（登録商標）プラットフォームを用いて、ヒトおよび外来性因子のデータを収集し、RefAlignerを用いて分析のために処理した。図14の表に示されているように、T. cruzi、T. brucei、T. pallidum、およびP. falciparumからマップを作成した。

10

【0102】

サンプル由来の外来性因子を決定することの実行可能性を調べるために、外来性因子およびヒト血液の混合物のデータセットをシミュレートした。サンプルには、混合物中0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、および50%の外来性因子核酸を含めた。各酵素について、ヒトゲノムおよび全外来性因子ゲノムからなる全コンティグの参照マップを作成した。次いで、「bestRef」オプションを用いて、分子をインシリコ参照ゲノムの種々のデータセットとアラインした。これを、外来性因子が存在してもプログラムが外来性因子を検出できない偽陰性および外来性因子物質の非存在下でヒト物質のみが存在する時に外来性因子を検出し得る偽陽性の検出に用いた。

20

【0103】

ヒト女性サンプル（IPS__sample__control__122313）に由来する実験的に観察された分子を、トリパノソーマに由来する実験的に観察された分子と種々の割合でインシリコで混合した。各分子を、アラインされた外来性因子およびヒトDNA分子の起源ならびにアラインメントの正確さを可能にする固有IDでラベルした。感度は $TP / (TP + FN)$ として定義される。特異性は $TN / (TN + FP)$ として定義される。適中率（precision）は $TP / (TP + FP)$ として定義される。検出率（recall）は $TP / (TP + FN)$ 。正確度（accuracy）は $(TP + TN) / (TP + TN + FP + FN)$ として定義される。F値は、 $2(適中率 - 検出率) / (適中率 + 検出率)$ として定義される。結果を図13に示す。図13に関して、FP = 偽陽性、FN = 偽陰性、TP = 真陽性、TN = 真陰性。したがって、インシリコ分析は、本明細書のいくつかの実施形態の通りに標識化パターンに基づいて外来性因子ゲノムの有無をヒト宿主ゲノムから区別できることを示している。

30

【0104】

図13の結果のグラフ表示を示している図11に示されているように、外来性因子DNAの量が2%を超える場合、外来性因子検出の信頼度は高い（適中率95%および検出率70%）。図示されているように、外来性因子の量が2%を超える場合、それぞれ95%および70%である適中率および検出率に見られるように、比較的信頼度が高い。

【0105】

このように、この技術を用いて、外来性因子が2%の場合に95%の適中率および70%の検出率で外来性因子が決定された。

40

【0106】

本明細書の実質的に全ての複数形および/または単数形の用語の使用に関して、当業者は、本文および/または応用に適するように複数から単数および/または単数から複数へと言い換えることができる。明確にするために、本明細書で種々の単数/複数の交換が明確に記載されていることもある。

【0107】

当業者には、一般的に、本明細書および特に添付の特許請求の範囲（例えば、添付の特許請求の範囲の本文）で使用される用語が一般的にオープンな用語を意図していると理解

50

される（例えば、「含む」という用語は、「含むが、これに限定されない」と解釈されるべきであり、「有する」という用語は、「少なくとも～を有する」と解釈されるべきであり、「含む」という用語は、「含むが、これに限定されない」と解釈されるべきである等）。更に、導入される請求項記載において具体的な数が意図される場合、そのような意図は当該請求項中に明確に記載され、そのような記載がない場合、そのような意図はないことが当業者には理解されよう。例えば、理解の助けとして、以下の添付の特許請求の範囲は、請求項の記載を導入するために「少なくとも１つ」および「１以上の」という導入句の使用を含み得る。しかし、そのような句の使用は、不定冠詞「a」または「an」により請求項の記載が導入された場合に、例え同じ請求項が導入句「１以上の」または「少なくとも１つの」および「a」または「an」等の不定冠詞を含む場合でも、そのように導入された請求項の記載を含む特定の請求項をそのような記載事項を１つだけ含む実施形態に限定することを意味すると解釈されるべきではなく（例えば、「a」および／または「an」は、「少なくとも１つの」または「１以上の」を意味すると解釈されるべきである）、請求項の記載を導入するための定冠詞の使用にも同じことが当てはまる。更に、導入された請求項記載の具体的な数が明示されている場合でも、そのような記載は、少なくとも記載されている数を意味すると解釈されるべきであることが当業者には理解されよう（例えば、他に修飾語のない単なる「２つの記載事項」という記載は、少なくとも２つの記載事項または２つ以上の記載事項を意味する）。更に、「A、B、およびCの少なくとも１つ等」に類する慣用句が使用されている場合、一般にそのような構造は、当業者がその慣用句を理解するであろう意味を意図する（例えば、「A、B、およびCの少なくとも１つを有するシステム」は、Aのみ、Bのみ、Cのみ、AおよびBの両方、AおよびCの両方、BおよびCの両方、ならびに／またはA、B、およびCの全てを有するシステムを含むが、これに限定されない等）。「A、B、またはCの少なくとも１つ等」に類する慣用句が使用されている場合、一般にそのような構造は、当業者がその慣用句を理解するであろう意味を意図する（例えば、「A、B、またはCの少なくとも１つを有するシステム」は、Aのみ、Bのみ、Cのみ、AおよびBの両方、AおよびCの両方、BおよびCの両方、ならびに／またはA、B、およびCの全てを有するシステムを含むが、これに限定されない等）。更に、２つ以上の選択的用語（*alternative term*）を表す事実上全ての選言的な語および／または句は、明細書、特許請求の範囲、または図面中に関わらず、それらの用語のうちの１つ、用語のいずれか、または用語の両方を含む可能性が企図されると理解すべきであることが当業者には理解されよう。例えば、「AまたはB」という句は、「A」、「B」、または「AおよびB」の可能性を含むと理解される。

【0108】

更に、本開示の構成または態様がマーカッシュグループの言葉で記載されている場合、当業者には、これにより、マーカッシュグループの任意の個々のメンバーまたはメンバーのサブグループの言葉でも開示が記載されていると理解される。

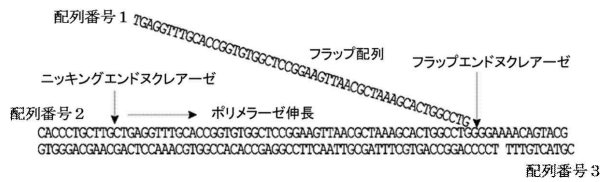
10

20

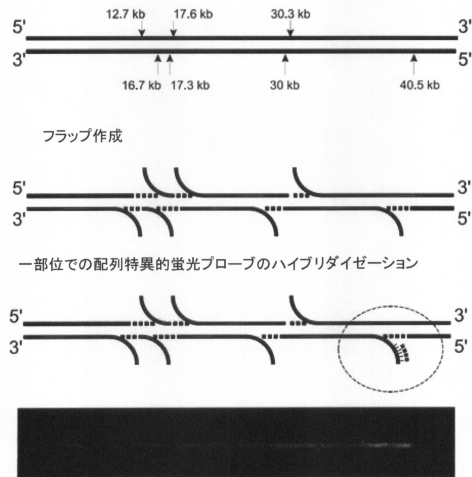
30

【図 1 A】

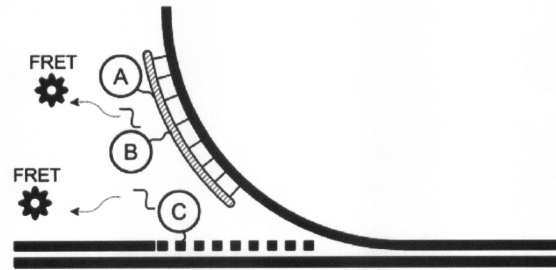
最初に、下流鎖にニックを入れて置換することによりフラップ配列を作成する



【図 1 B】

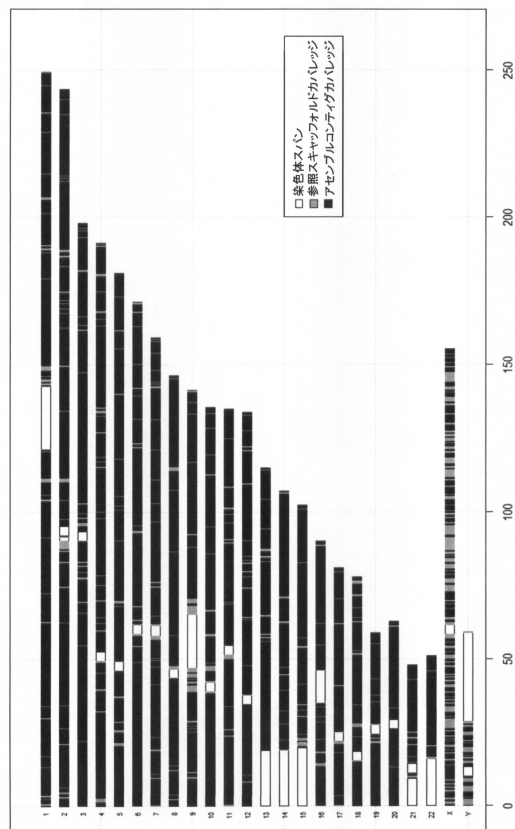


【図 2】

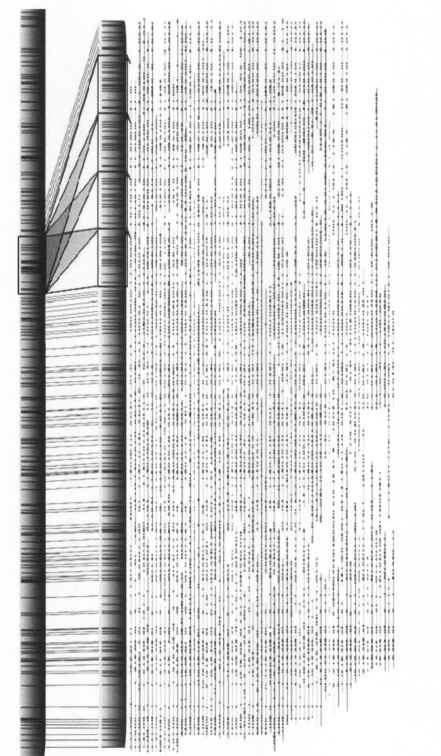


タグ化剤A、B、またはCが、取込みまたはハイブリダイゼーション中に、伸長したDNAと結合する。特異的蛍光共鳴エネルギー移動(FRET)シグナルにより共局在イベントを検出することができる。

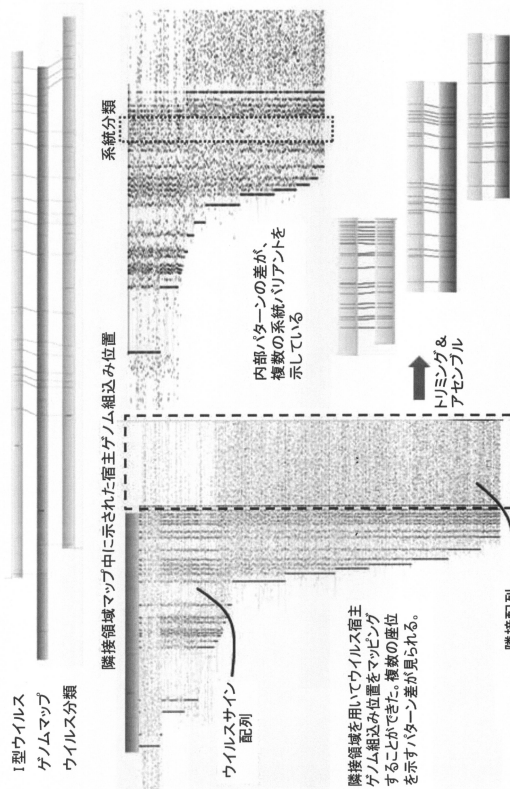
【図 3】



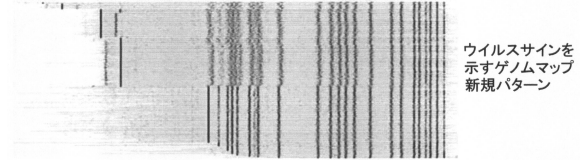
【図 4】



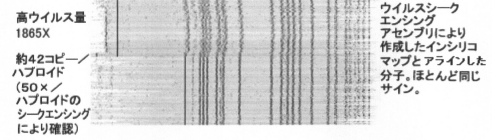
【図 5】



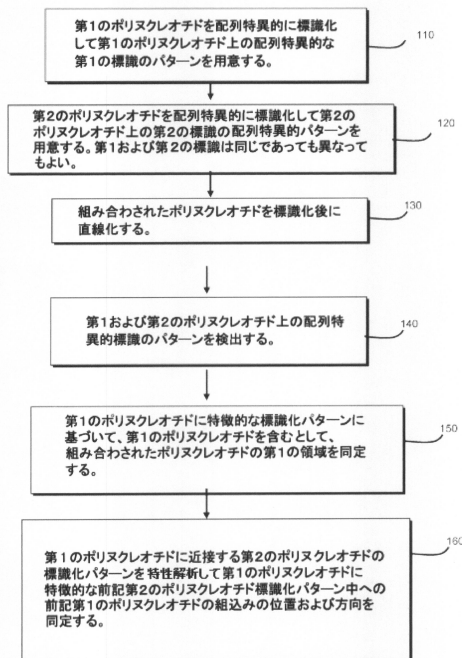
【図 6 A】



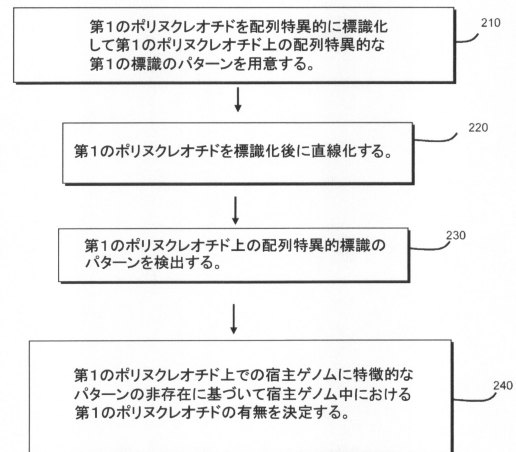
【図 6 B】



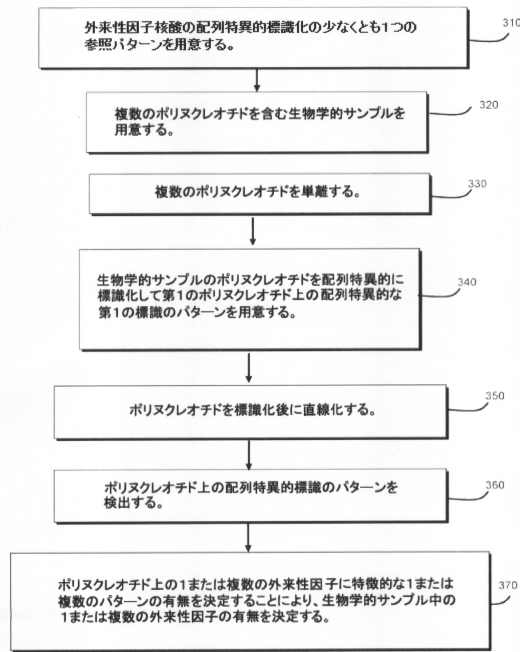
【図 7】



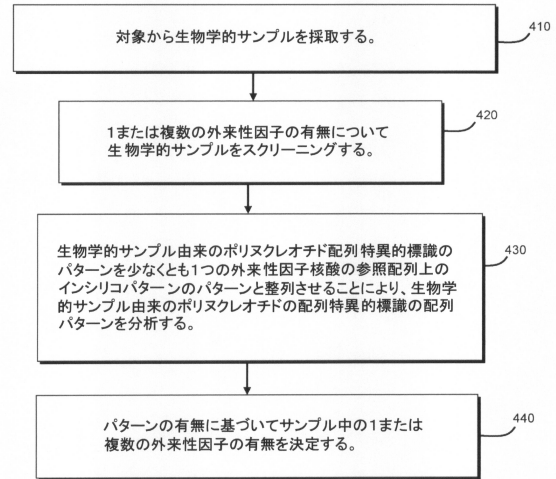
【図 8】



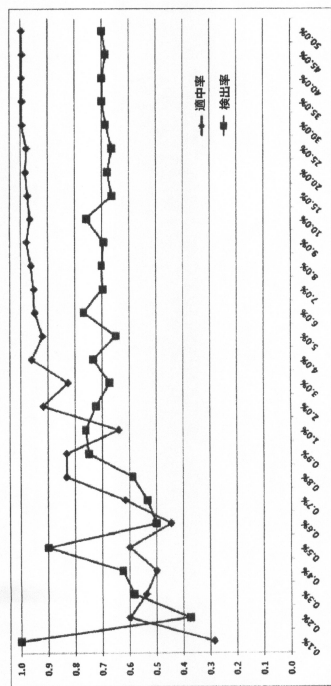
【図 9】



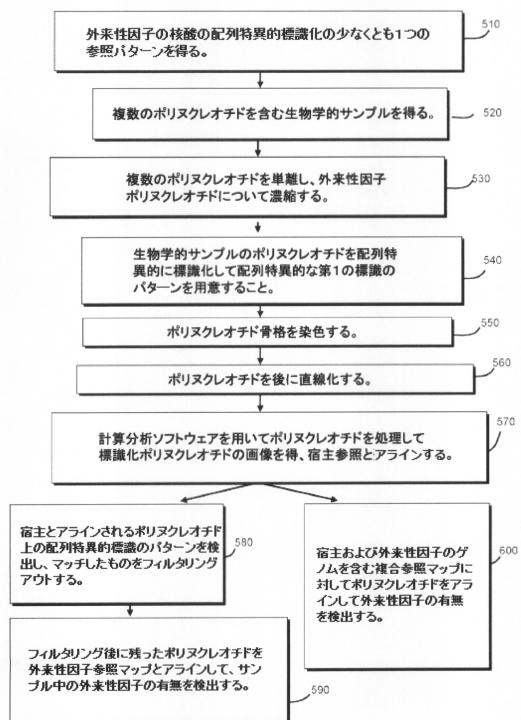
【図 10】



【図 11】



【図 12】



標本体の%	cruz0101_bupol			IP5_sample_control_122113			マージ			特性性			通中率検出率正確度			F値		
	分子	MB	MB(字相)	分子	MB	MB(字相)	分子	MB	MB(字相)	FP	TN	密度	FP	TN	密度	FP	TN	密度
6.1%	21	3,172	3	14524	2997,621	2997	14545	3000,234	3000	5	0	1.8301	1.000	0.999	0.999	0.999	0.999	0.999
6.2%	38	6,006	6	14472	2994,251	2994	14510	3000,258	3000	2	5	3.8201	0.375	1.000	0.600	0.375	0.999	0.444
6.3%	63	9,108	9	14377	2991,177	2991	14435	3000,384	3000	6	5	7.8307	0.583	0.999	0.538	0.999	0.583	0.560
6.4%	86	12,119	12	14347	2988,178	2988	14422	3000,677	3000	5	3	6.8111	0.625	0.999	0.500	0.825	0.999	0.556
6.5%	92	13,097	15	14474	2985,133	2985	14506	3000,140	3000	6	1	9.8246	0.900	0.999	0.600	0.900	0.999	0.720
6.6%	124	18,096	18	14477	2982,108	2982	14509	3000,204	3000	5	4	8.8246	0.500	0.999	0.444	0.900	0.999	0.471
6.7%	126	21,099	21	14359	2979,396	2979	14518	3000,455	3000	5	7	8.8246	0.533	0.999	0.615	0.533	0.999	0.571
6.8%	162	24,004	24	14354	2976,173	2976	14518	3000,177	3000	2	7	10.8138	0.588	1.000	0.833	0.588	0.999	0.690
6.9%	191	27,099	27	14364	2973,060	2973	14553	3000,159	3000	3	5	15.8205	0.750	1.000	0.833	0.750	0.999	0.789
7.0%	200	30,098	30	14424	2970,226	2970	14620	3000,324	3000	9	5	10.8294	0.762	0.999	0.600	0.762	0.999	0.896
7.1%	278	378	63	14258	2960,083	2960	14582	3000,218	3000	3	13	34.8140	0.723	1.000	0.919	0.723	0.998	0.810
7.2%	668	93,087	96	14079	2910,488	2910	14685	3000,455	3000	9	21	43.8695	0.672	0.999	0.827	0.672	0.998	0.741
7.3%	791	120,004	120	13868	2880,093	2880	14601	3000,171	3000	3	28	78.9251	0.735	1.000	0.900	0.735	0.999	0.852
7.4%	1027	150,001	150	13799	2850,128	2850	14708	3000,171	3000	4	28	78.9251	0.735	1.000	0.900	0.735	0.999	0.852
7.5%	1111	160,000	160	13697	2820,161	2820	14618	3000,204	3000	3	34	133.7834	0.699	0.999	0.869	0.699	0.995	0.850
7.6%	1411	210,009	210	13487	2780,244	2780	14620	3000,254	3000	6	53	123.7677	0.699	0.999	0.953	0.699	0.992	0.807
7.7%	1625	240,007	240	13481	2760,169	2760	15113	3000,178	3000	5	55	129.7669	0.701	0.999	0.963	0.701	0.992	0.811
7.8%	1763	270,021	270	13181	2730,196	2730	14944	3000,267	3000	3	60	130.7206	0.694	1.000	0.978	0.694	0.992	0.812
7.9%	1969	300,032	300	12979	2700,309	2700	14944	3000,342	3000	6	54	170.7408	0.759	0.999	0.906	0.759	0.992	0.850
8.0%	2056	420,044	420	12299	2550,256	2550	15538	3000,299	3000	6	114	224.7033	0.663	0.999	0.974	0.663	0.994	0.789
8.1%	4045	600,026	600	11593	2400,222	2400	15564	3000,246	3000	5	134	285.6339	0.680	0.999	0.983	0.680	0.994	0.804
25.0%	6912	700,022	700	10803	2250,087	2250	15725	3000,199	3000	8	178	353.6235	0.664	0.999	0.978	0.664	0.973	0.791
30.0%	5996	900,134	900	10109	2100,068	2100	16105	3000,202	3000	2	288	452.5843	0.685	1.000	0.996	0.685	0.988	0.812
35.0%	6051	1050,143	1050	9552	1950,084	1950	16501	3000,227	3000	3	220	514.5294	0.700	0.999	0.994	0.700	0.993	0.822
40.0%	8811	1200,207	1200	8747	1800,167	1800	16756	3000,174	3000	2	279	648.4355	0.699	1.000	0.997	0.699	0.992	0.827
45.0%	8923	1350,229	1350	8018	1650,059	1650	16941	3000,289	3000	4	310	676.4653	0.686	0.999	0.994	0.686	0.994	0.832
50.0%	9877	1500,444	1500	7322	1500,139	1500	17199	3000,383	3000	2	321	747.4188	0.699	1.000	0.997	0.699	0.999	0.822
計算マージナル	0	0.000	0	14556	3000,001	3000	14556	3000,001	3000	4	0	0.684N/A	1.000	0.000N/A	1.000	0.000N/A	1.000N/A	

名称	ゲノムサイズ	合計 コンティグ 数	Mappability (50-mer)	GC%	N%	ニック密度 (BspQ1/ 100Kb)	ニック密度 (BbvC1/ (100Kb)
<i>T. Cruzi</i> 8つの バージョン	279,187,253	76427	0.946	51.11	10.7	15.410	19.078
<i>T. brucei</i>	35,826,294	131	0.828	45.49	0.03	17.669	16.854
<i>T. pallidum</i>	1,138,011	1	0.987	52.78	0.00	18.19	14.060
<i>P. falciparum</i>	23,328,019	16	0.989	19.34	0.00	3.459	0.836

フロントページの続き

- (72)発明者 ハスティ 、 アレックス、 アール、
 アメリカ合衆国 9 2 1 2 6 カリフォルニア州 サン ディエゴ クレーター ドライブ 1 1
 1 2 5
- (72)発明者 ラム、 アーネスト、 チ ツン
 アメリカ合衆国 9 2 1 2 2 カリフォルニア州 サン ディエゴ マハリア アベニュー 3 9
 5 0 ナンバージェイ 3 2
- (72)発明者 ジャクラ、 ジェリコ
 アメリカ合衆国 9 2 1 2 9 カリフォルニア州 サン ディエゴ サンドانس アベニュー 1
 2 8 3 0

審査官 野村 英雄

- (56)参考文献 BEN-DAVID, Y. et al., "Identification and mapping of a common proviral integration site Fli-1 in erythroleukemia cells induced by Friend murine leukemia virus.", PROC. NATL. ACAD. SCI. USA, 1 9 9 0 年 2 月, Vol.87, No.4, pp.1332-1336
- HASTIE, A.R. et al., "Rapid genome mapping in nanochannel arrays for highly complete and accurate De Novo sequence assembly of the complex Aegilops tauschii genome.", PLOS ONE, 2 0 1 3 年 2 月 6 日, Vol.8, No.2, e55864(pp.1-10), doi: 10.1371/journal.pone.0055864
- LAM, E.T. et al., "Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly.", NATURE BIOTECHNOLOGY, 2 0 1 2 年 8 月, Vol.30, No.8, pp.771-776 + ONLINE METHODS, doi: 10.1038/nbt.2303

(58)調査した分野(Int.Cl., DB名)

C 1 2 N 1 5 / 0 9

C 1 2 Q 1 / 6 8 1 3

P u b M e d

J S T P l u s / J M E D P l u s / J S T 7 5 8 0 (J D r e a m I I I)