

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7605763号
(P7605763)

(45)発行日 令和6年12月24日(2024.12.24)

(24)登録日 令和6年12月16日(2024.12.16)

(51)国際特許分類

F I

C 1 2 Q 1/6869(2018.01)

C 1 2 Q 1/6869

Z Z N A

G 1 6 B 20/20 (2019.01)

G 1 6 B 20/20

請求項の数 18 (全84頁)

(21)出願番号	特願2021-565779(P2021-565779)	(73)特許権者	521480341
(86)(22)出願日	令和2年5月1日(2020.5.1)		ウルティマ ジェノミクス, インコーポ
(65)公表番号	特表2022-531017(P2022-531017		レイテッド
	A)		アメリカ合衆国 カリフォルニア 9 4 5
(43)公表日	令和4年7月5日(2022.7.5)		6 0, ニューアーク, ゲートウェイ
(86)国際出願番号	PCT/US2020/031147		ブールバード 7 9 7 9, スイート 1
(87)国際公開番号	WO2020/227137		0 1
(87)国際公開日	令和2年11月12日(2020.11.12)	(74)代理人	100078282
審査請求日	令和5年4月28日(2023.4.28)		弁理士 山本 秀策
(31)優先権主張番号	62/842,534	(74)代理人	100113413
(32)優先日	令和1年5月3日(2019.5.3)		弁理士 森下 夏樹
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100181674
(31)優先権主張番号	62/971,530		弁理士 飯田 貴敏
(32)優先日	令和2年2月7日(2020.2.7)	(74)代理人	100181641
	最終頁に続く		弁理士 石川 大輔
			最終頁に続く

(54)【発明の名称】 核酸バリエーションを検出するための方法

(57)【特許請求の範囲】

【請求項1】

試験試料中の短い遺伝子バリエーションを検出するための方法であって、
疾患に関連する短い標的遺伝子バリエーションを選択するステップであって、前記短い標的
遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットおよび参
照配列に関連する参照シーケンシングデータセットが、フローサイクル順序に従って別
々のフロー位置で提供される非終結ヌクレオチドを使用してそれぞれ前記標的配列および
前記参照配列をシーケンシングすることにより得られた場合、前記短い標的遺伝子バリエ
ーションを含む前記標的配列に関連する前記標的シーケンシングデータセットと、前記参
照配列に関連する前記参照シーケンシングデータセットとが、4またはそれより多くの
連続したフロー位置において異なるように、前記短い標的遺伝子バリエーションを選択する、
ステップ；

1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験
シーケンシングデータセットが、試験核酸配列に関連し、各試験核酸配列が、前記短い
標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ前記試験試
料に由来し、前記1つまたは複数の試験シーケンシングデータセットが、前記フローサ
イクル順序に従って別々のフロー位置で提供される非終結ヌクレオチドを使用して前記試
験試料をシーケンシングすることにより決定されたものである、ステップ；

試験核酸配列の各々について、前記試験シーケンシングデータセットが前記標的シー
ケンシングデータセットにマッチする尤度を示すそれぞれのマッチスコア、または前記

試験シーケンシングデータセットが前記参照シーケンシングデータセットにマッチする尤度を示すそれぞれのマッチスコアを決定するステップ；および

1つまたは複数の前記それぞれのマッチスコアを使用して、前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップを含み、

前記短い標的遺伝子バリエーションが、前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップの前に選択される、方法。

【請求項2】

前記得るステップが、前記フローサイクル順序に従って別々のフロー位置で提供される非終結ヌクレオチドを使用して前記試験試料に由来する試験核酸分子をシーケンシングすることを含み、各々の試験核酸分子はそれぞれの試験シーケンシングデータセットに関連する、請求項1に記載の方法。

【請求項3】

前記試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、前記個別化バイオマーカーパネルが、前記短い標的遺伝子バリエーションを含む、請求項1または2に記載の方法。

【請求項4】

前記フローサイクル順序を選択するステップを含む、請求項1から3のいずれか一項に記載の方法。

【請求項5】

前記標的シーケンシングデータセットおよび前記参照シーケンシングデータセットが、前記標的配列および前記参照配列を *in silico* でシーケンシングすることにより得られる、請求項1から4のいずれか一項に記載の方法。

【請求項6】

前記1つまたは複数の試験シーケンシングデータセットの各々が、前記別々のフロー位置のフローシグナルを含み、前記フローシグナルが、各フロー位置においてシーケンシングされた前記試験核酸分子のそれぞれの塩基の数を示す塩基カウントを含む、請求項2に記載の方法。

【請求項7】

前記フローシグナルが、各フロー位置における少なくとも1塩基カウントの尤度を示す統計パラメーターを含み、前記塩基カウントが、前記フロー位置においてシーケンシングされた前記試験核酸分子のそれぞれの塩基の数を示す、請求項6に記載の方法。

【請求項8】

前記フローシグナルが、各フロー位置における複数の塩基カウントの尤度を示す統計パラメーターを含み、各塩基カウントが、前記フロー位置においてシーケンシングされた前記試験核酸分子のそれぞれの塩基の数を示す、請求項6または請求項7に記載の方法。

【請求項9】

前記決定するステップが、前記試験シーケンシングデータセット内の各フロー位置における統計パラメーターであって、そのフロー位置における前記標的配列の塩基カウントに対応する統計パラメーターを選択すること、および前記試験シーケンシングデータセットが前記標的シーケンシングデータセットにマッチする尤度を示す前記それぞれのマッチスコアを決定すること；または

前記試験シーケンシングデータセット中の各フロー位置における統計パラメーターであって、そのフロー位置における前記参照配列の塩基カウントに対応する統計パラメーターを選択すること、および前記試験シーケンシングデータセットが前記参照シーケンシングデータセットにマッチする尤度を示す前記それぞれのマッチスコアを決定することを含む、請求項8に記載の方法。

【請求項10】

前記決定するステップで決定される前記それぞれのマッチスコアが、前記試験シーケ

10

20

30

40

50

ンシングデータセット内の前記フロー位置にわたっての選択された統計パラメーターの組み合わせられた値である、請求項 9 に記載の方法。

【請求項 1 1】

前記決定するステップが、前記試験シーケンシングデータセットが前記標的シーケンシングデータセットにマッチする尤度を示す前記それぞれのマッチスコアを決定することを含む、請求項 1 から 1 0 のいずれか一項に記載の方法。

【請求項 1 2】

前記決定するステップが、前記試験シーケンシングデータセットが前記参照シーケンシングデータセットにマッチする尤度を示す前記それぞれのマッチスコアを決定することを含む、請求項 1 から 1 1 のいずれか一項に記載の方法。

10

【請求項 1 3】

前記 1 つまたは複数の試験シーケンシングデータセットが、複数の試験シーケンシングデータセットを含む、請求項 1 から 1 2 のいずれか一項に記載の方法。

【請求項 1 4】

前記短い標的遺伝子バリエーションの存在または非存在が、前記 1 つまたは複数の試験シーケンシングデータセットの各々について別々にコールされる、請求項 1 3 に記載の方法。

【請求項 1 5】

前記複数の試験シーケンシングデータセットの少なくとも一部分が、異なる試験シーケンシング開始位置を有する異なる試験核酸分子に関連する、請求項 1 3 または 1 4 に記載の方法。

20

【請求項 1 6】

前記フローサイクル順序が、同じ順序で反復される 4 つの別々のフローを含む、請求項 1 から 1 5 のいずれか一項に記載の方法。

【請求項 1 7】

前記フローサイクル順序が、5 つまたはそれより多くの別々のフローを含む、請求項 1 から 1 6 のいずれか一項に記載の方法。

【請求項 1 8】

試験試料中の疾患に関連する短い遺伝子バリエーションを検出するための方法であって、

1 つまたは複数の第 1 の試験シーケンシングデータセットを得るステップであって、各第 1 の試験シーケンシングデータセットが、前記試験試料に由来する異なる試験核酸分子に関連し、前記第 1 の試験シーケンシングデータセットが、第 1 のフローサイクル順序に従って別々のフロー位置で提供される非終結ヌクレオチドを使用して 1 つまたは複数の試験核酸分子をシーケンシングすることにより決定されたものであり、前記 1 つまたは複数の第 1 の試験シーケンシングデータセットが、前記別々のフロー位置におけるフローシグナルを含む、ステップ；

30

1 つまたは複数の第 2 の試験シーケンシングデータセットを得るステップであって、各第 2 の試験シーケンシングデータセットが、対応する第 1 の試験シーケンシングデータセットと同じそれぞれの試験核酸分子に関連し、前記第 2 の試験シーケンシングデータセットが、第 2 のフローサイクル順序に従って別々のフロー位置で提供される非終結ヌクレオチドを使用して前記 1 つまたは複数の試験核酸分子をシーケンシングすることにより決定されたものであり、前記第 1 のフローサイクル順序と前記第 2 のフローサイクル順序が異なり、前記 1 つまたは複数の第 2 の試験シーケンシングデータセットが、前記別々のフロー位置におけるフローシグナルを含む、ステップ；

40

第 1 のシーケンシングデータセットおよび対応する第 2 のシーケンシングデータセットの各々について、1 つまたは複数の候補配列とのマッチスコアを決定するステップであって、それぞれの前記マッチスコアが、前記第 1 の試験シーケンシングデータセット、前記第 2 の試験シーケンシングデータセット、または両方が前記 1 つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および

決定されたマッチスコアを使用して、前記試験試料中の前記短い遺伝子バリエーションの存在または非存在をコールするステップ

50

を含む方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

本願は、2019年3月3日に出願した米国特許仮出願第62／842，534号および2020年2月7日に出願した米国特許仮出願第62／971，530号に基づく優先権の利益を主張しており、前記仮出願の各々の内容は、それら全体が参照により本明細書に援用される。

【0002】

ASCIIテキストファイルでの配列表の提出

ASCIIテキストファイルでの以下の提出内容は、その全体が参照により本明細書に取り込まれる：コンピュータ可読形式（CRF）の配列表（ファイル名：165272000540SEQLIST.TXT、記録日：2020年4月27日、サイズ：5KB）。

【0003】

発明の分野

遺伝子バリエーションの検出を含む、シーケンシングデータを生成および／または解析する方法を含む、ポリヌクレオチドをシーケンシングする方法が、本明細書に記載される。

【背景技術】

【0004】

背景

DNA試料中の遺伝子バリエーションは、試料中のDNAをシーケンシングすること、配列を参照配列とアラインメントすること、および差を評価することにより、検出することができる。シーケンシングされたDNAと参照配列との信頼度の高い差は、DNA試料が由来する生物のバリエーションとしてコールされる。次世代シーケンシングは、単一試料中の多くの異なる核酸分子の同時シーケンシングに必要なツールを、研究および臨床研究室にもたらし、解析すべき大量のデータを生じさせた。

【0005】

加えて、可逆的ターミネーターを用いた合成によるシーケンシング（例えば、可逆的に終結される、色素標識シーケンシング法）は、各塩基について単一の異なるシグナルを提供し、したがって、単一シグナルシーケンシングエラーは、エラーのバリエーションコールを生じさせる結果となり得る。一部のケースでは、これは、真の陽性シグナルを以て、エラーのコールを有効に圧倒する高深度シーケンシングにより克服することができるが、そのような高深度でのシーケンシングは、費用がかかり、時間がかかる。

【発明の概要】

【課題を解決するための手段】

【0006】

非常に効率的で正確な塩基コーリングおよびバリエーションコーリングプロトコールへの要求が、当技術分野には依然としてある。

【0007】

簡単な要旨

核酸分子を含有する試験試料中の短い遺伝子バリエーションを検出するための方法であって、ある特定の実施形態では、コンピュータ実装方法であり得る方法が、本明細書に記載される。そのような方法を遂行するためのシステムも本明細書に記載される。さらに、核酸分子をシーケンシングする方法が記載される。

【0008】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シ

10

20

30

40

50

ークエンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。

10

【0009】

上記方法の一部の実施形態では、得るステップは、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることを含む。

20

【0010】

上記方法の一部の実施形態では、短い標的遺伝子バリエーションは、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの前に事前選択される。一部の実施形態では、短い標的遺伝子バリエーションは、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの後にコールの信頼度に基づいて選択される。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップを含み、バイオマーカーパネルは、短い標的遺伝子バリエーションを含む。

【0011】

上記方法の一部の実施形態では、方法は、フローサイクル順序を選択するステップをさらに含む。

30

【0012】

一部の実施形態では、標的シーケンシングデータセットは、予想標的シーケンシングデータセットであり、または参照シーケンシングデータセットは、予想参照シーケンシングデータセットである。一部の実施形態では、標的シーケンシングデータセットおよび予想参照シーケンシングデータセットは、標的配列および参照配列を *in silico* でシーケンシングすることにより得られる。

【0013】

上記方法の一部の実施形態では、標的シーケンシングデータセットと参照シーケンシングデータとは、2カ所より多くの非連続のフロー位置において異なる。一部の実施形態では、標的シーケンシングデータセットと参照シーケンシングデータとは、2カ所より多くの連続するフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、Xカ所の塩基位置において異なり、標的シーケンシングデータセットと参照シーケンシングデータとは、(X+2)カ所またはそれより多くの連続するフロー位置において異なる。一部の実施形態では、(X+2)カ所のフロー位置の相違は、実質的にゼロに等しい値と実質的にゼロより大きい値の間の相違を含む。一部の実施形態では、標的シーケンシングデータセットと参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。一部の実施形態では、フローシグナルは、各フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す塩基カウントを含む。

40

【0014】

50

上記方法の一部の実施形態では、フローシグナルは、各フロー位置における少なくとも1塩基カウントの尤度を示す統計パラメーターを含み、塩基カウントは、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す。一部の実施形態では、フローシグナルは、各フロー位置における複数の塩基カウントの尤度を示す統計パラメーターを含み、各塩基カウントは、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す。

【0015】

上記方法の一部の実施形態では、ステップ(c)は、(i)試験シーケンシングデータセット内の各フロー位置における統計パラメーターであって、そのフロー位置における標的配列の塩基カウントに対応する統計パラメーターを選択すること、および試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコアを決定すること；または(ii)試験シーケンシングデータセット中の各フロー位置における統計パラメーターであって、そのフロー位置における参照配列の塩基カウントに対応する統計パラメーターを選択すること、および試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定することを含む。一部の実施形態では、ステップ(c)で決定されるマッチスコアは、試験シーケンシングデータセット内のフロー位置にわたっての選択された統計パラメーターの組み合わせされた値である。一部の実施形態では、ステップ(c)は、試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコアを決定することを含む。一部の実施形態では、ステップ(c)は、試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定することを含む。

【0016】

上記方法の一部の実施形態では、1つまたは複数の試験シーケンシングデータセットは、複数の試験シーケンシングデータセットを含む。一部の実施形態では、短い標的遺伝子バリエーションの存在または非存在は、1つまたは複数の試験シーケンシングデータセットの各々について別々にコールされる。一部の実施形態では、複数の試験シーケンシングデータセットの少なくとも一部分は、異なる試験シーケンシング開始位置を有する異なる試験核酸分子に関連する。

【0017】

上記方法の一部の実施形態では、フローサイクル順序は、同じ順序で反復される4つの別々のフローを含む。一部の実施形態では、フローサイクル順序は、5つまたはそれより多くの別々のフローを含む。

【0018】

上記方法の一部の実施形態では、方法は、コンピュータ実装方法である。例えば、一部の実施形態では、コンピュータ実装方法は、1または複数台のプロセッサを使用して短い標的遺伝子バリエーションを選択するステップ；1つまたは複数の試験シーケンシングデータセットを1または複数台のプロセッサで受信することにより、1つまたは複数の試験シーケンシングデータセットを得るステップ；1または複数台のプロセッサを使用して1つまたは複数のマッチスコアを決定するステップ；および1または複数台のプロセッサを使用して試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0019】

1または複数台のプロセッサと、上記の方法を実行するための命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムも本明細書に提供される。

【0020】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、(a) 1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に

従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(b) 1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、第1のフローサイクル順序と第2のフローサイクル順序が異なり、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(c) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットの各々について、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列のマッチする尤度を示す、ステップ；および(d) 決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。

10

【0021】

上記方法の一部の実施形態では、方法は、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングするステップ、および第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングするステップを含む。

20

【0022】

上記方法の一部の実施形態では、マッチスコアは、第1の試験シーケンシングデータセットが候補配列にマッチする尤度、または第2の試験シーケンシングデータセットが候補配列にマッチする尤度を示す。一部の実施形態では、マッチスコアは、第1の試験シーケンシングデータセットと第2のシーケンシングデータセット両方が候補配列にマッチする尤度を示す。

【0023】

上記方法の一部の実施形態では、1つまたは複数の候補配列は、2つまたはそれより多くの異なる候補配列を含み、方法は、第1のシーケンシングデータセットおよび第2のシーケンシングデータセットに関連する核酸分子の各々について：2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列は、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くの非連続のフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くの非連続のフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配

30

40

50

列と、選択候補配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くの連続のフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くの連続のフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、3カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、3カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、Xカ所の塩基位置において異なり、試験核酸分子に関連する試験シーケンシングデータセットと、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、 $(X + 2)$ カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、Xカ所の塩基位置において異なり、試験核酸分子に関連する試験シーケンシングデータセットと、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、 $(X + 2)$ カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、 $(X + 2)$ カ所のフロー位置の相違は、実質的にゼロに等しい値と実質的にゼロより大きい値の間の相違を含む。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、1または複数のフローサイクルにわたって異なる。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、1または複数のフローサイクルにわたって異なる。

【0024】

上記方法の一部の実施形態では、フローシグナルは、各フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す塩基カウントを含む。一部の実施形態では、フローシグナルは、各フロー位置における少なくとも1塩基カウントの尤度を示す統計パラメーターを含み、塩基カウントは、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す。一部の実施形態では、フローシグナルは、各フロー位置における複数の塩基カウントの尤度を示す統計パラメーターを含み、各塩基カウントは、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す。一部の実施形態では、マッチスコアを決定するステップは、第1の試験シーケンシングデータセットおよび第2の試験シーケンシングデータセットにおける各フロー位置における統計パラメーターであって、候補配列のそのフロー位置における塩基カウントに対応する統計パラメーターを、1つまたは複数の異なる候補配列の各々について選択することを含む。上記方法の一部の実施形態では、方法は、1つまたは複数の異なる候補配列について、各フロー位置における候補配列の塩基カウントを含む候補シーケンシングデータセットを生成するステップを含む。一部の実施形態では、候補シーケンシングデータセットは、`in silico`で生成される。一部の実施形態では、マッチスコアは、第1の試験シーケンシングデータセットおよび第2の試験シーケンシングデータセットにおけるフロー位置にわたって選択された統計パラメーターの組み合わせられた値である。

【0025】

上記方法の一部の実施形態では、試験核酸分子の少なくとも一部分は、異なるシーケンシング開始位置を有する。

【0026】

上記方法の一部の実施形態では、方法は、短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとは、標的シーケンシングデータセットおよび参照シーケンシングデータセットは、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップをさらに含み、1つまたは複数の候補配列は、標的配列および参照配列を含む。一部の実施形態では、短い標的遺伝子バリエーションは、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの前に事前選択される。一部の実施形態では、短い標的遺伝子バリエーションは、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの後にコールの信頼度に基づいて選択される。一部の実施形態では、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップを含み、バイオマーカーパネルは、試験試料中に存在する短い標的遺伝子バリエーションを含む。一部の実施形態では、参照シーケンシングデータセットは、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のフローで提供される非終結ヌクレオチドを使用して参照配列がシーケンシングされた場合の予想参照シーケンシングデータセットを決定することにより得られる。一部の実施形態では、参照シーケンシングデータセットは、第1のフローサイクル順序および第2のフローサイクル順序の両方に従って別々のフローで提供される非終結ヌクレオチドを使用して参照配列がシーケンシングされた場合の予想参照シーケンシングデータセットを決定することにより得られる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くの非連続のフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くの非連続のフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くの連続であるフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くの連続であるフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、3カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、3カ所またはそれより多くのフロー位置において異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くのフローサイクルにわたって異なる。一部の実施形態では、標的配列と参照配列とは、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くのフローサイクルにわたって異なる。

【0027】

上記方法の一部の実施形態では、第1のフローサイクル順序または第2のフローサイクル順序は、同じ順序で反復される4つの別々のフローを含む。一部の実施形態では、第1のフローサイクル順序または第2のフローサイクル順序は、同じ順序で反復される5つまたはそれより多くの別々のフローを含む。

【0028】

上記方法の一部の実施形態では、方法は、試験核酸分子をシーケンシングするステップであって、第1のフローサイクル順序に従って別々のヌクレオチドフローで非終結ヌク

10

20

30

40

50

レオチドを提供することと、シークエンシングプライマーを伸長することと、各ヌクレオチドフローの後にシークエンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して第1の試験シークエンシングデータセットを生成することを含むステップ；伸長されたシークエンシングプライマーを除去するステップ；および同じ試験核酸分子をシークエンシングするステップであって、第2のフローサイクル順序に従って別々のヌクレオチドフローで非終結ヌクレオチドを提供することと、シークエンシングプライマーを伸長することと、各ヌクレオチドフローの後にシークエンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して第2の試験シークエンシングデータセットを生成することを含むステップを含む。

【0029】

上記方法の一部の実施形態では、方法は、コンピュータ実装方法である。例えば、一部の実施形態では、コンピュータ実装方法は、1つまたは複数の第1のシークエンシングデータセットを1または複数台のプロセッサで受信するステップ；1つまたは複数の第1のシークエンシングデータセットを1または複数台のプロセッサで受信するステップ；1または複数台のプロセッサを使用してマッチスコアを決定するステップ；および1または複数台のプロセッサを使用して試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0030】

1または複数台のプロセッサと、上記の方法のいずれかを実行するための命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムも本明細書に記載される。

【0031】

上記の方法またはシステムのいずれかの一部の実施形態では、別々のフローは、単一の塩基タイプを含む。

【0032】

上記の方法またはシステムのいずれかの一部の実施形態では、別々のフローのうちの少なくとも1つは、2つまたは3つの異なる塩基タイプを含む。

【0033】

上記の方法またはシステムのいずれかの一部の実施形態では、方法は、試験試料中の短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示すバリエーションコールファイルを生成またはアップデートするステップを含む。

【0034】

上記の方法またはシステムのいずれかの一部の実施形態では、方法は、試験試料中の短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示すレポートを生成するステップを含む。一部の実施形態では、レポートは、試験試料中の短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示す、テキスト出力、確率的出力、数値出力またはグラフィック出力を含む。一部の実施形態では、方法は、レポートを患者にまたは患者の医療担当者に提供するステップを含む。

【0035】

上記の方法またはシステムのいずれかの一部の実施形態では、短い遺伝子バリエーションは、一塩基多型またはインデルを含む。

【0036】

上記の方法またはシステムのいずれかの一部の実施形態では、短い遺伝子バリエーションは、インデルを含む。

【0037】

上記の方法またはシステムのいずれかの一部の実施形態では、試験試料は、断片化DNAを含む。

【0038】

上記の方法またはシステムのいずれかの一部の実施形態では、試験試料は、無細胞DNAを含む。一部の実施形態では、無細胞DNAは、循環腫瘍DNA (ctDNA) を含む。

10

20

30

40

50

【0039】

一部の実施形態では、核酸分子をシーケンシングする方法は、核酸分子をプライマーとハイブリダイズして、ハイブリダイズされた鑄型を形成するステップ；5つまたはそれより多くの別々のヌクレオチドフローを含む反復フローサイクル順序に従って別々のヌクレオチドフローで提供される標識された非終結ヌクレオチドを使用してプライマーを伸長するステップ；および取り込まれた標識ヌクレオチドからのシグナルを検出する、またはプライマーがヌクレオチドフローにより伸長された場合のシグナルの非存在を検出するステップを含む。一部の実施形態では、方法は、シグナル、またはシグナルの非存在を、各ヌクレオチドフロー後に検出するステップを含む。一部の実施形態では、方法は、複数の核酸分子をシーケンシングするステップを含む。一部の実施形態では、複数のうちの核酸分子は、ある遺伝子座に対して異なるシーケンシング開始位置を有する。一部の実施形態では、試験試料は、無細胞DNAである。一部の実施形態では、無細胞DNAは、循環腫瘍DNA (ctDNA) を含む。一部の実施形態では、フローサイクル順序は、2カ所より多くのフロー位置において、ランダムシーケンシング開始位置の5%またはそれより多くにおける可能なSNPパーミュテーションの50%またはそれより多くについてのシグナル変化を誘導する。一部の実施形態では、誘導されるシグナル変化は、シグナル強度の変化、または新しい実質的にゼロ（もしくは新しいゼロ）もしくは新しい実質的に非ゼロ（もしくは新しい非ゼロ）シグナルである。一部の実施形態では、誘導されるシグナル変化は、新しい実質的にゼロ（もしくは新しいゼロ）または新しい実質的に非ゼロ（もしくは新しい非ゼロ）シグナルである。一部の実施形態では、前記フローサイクル順序は、1フロー当たり0.6のまたはそれを超える塩基取り込みの効率を有する。

本発明の実施形態において、例えば以下の項目が提供される。

(項目1)

試験試料中の短い遺伝子バリエーションを検出するための方法であって、

(a) 短い標的遺伝子バリエーションを選択するステップであって、前記短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、前記標的シーケンシングデータセットおよび前記参照シーケンシングデータセットが複数のフロー位置を含むフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して前記標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、前記フロー位置が、前記ヌクレオチドフローに対応する、ステップ；

(b) 1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、前記短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ前記試験試料に由来し、前記1つまたは複数の試験シーケンシングデータセットが、前記フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用してそれぞれの試験核酸分子をシーケンシングすることにより決定され、各試験シーケンシングデータセットが、前記複数のフロー位置におけるフローシグナルを含む、ステップ；

(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、前記核酸分子に関連する前記試験シーケンシングデータセットが前記標的配列にマッチする尤度を示すマッチスコア、または前記核酸分子に関連する前記試験シーケンシングデータセットが前記参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および

(d) 1つまたは複数の決定されたマッチスコアを使用して、前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む方法。

(項目2)

前記得るステップが、前記フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して前記試験核酸分子をシーケンシングすることを

含む、項目 1 に記載の方法。

(項目 3)

前記短い標的遺伝子バリエーションが、前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップの前に事前選択される、項目 1 または項目 2 に記載の方法。

(項目 4)

前記短い標的遺伝子バリエーションが、前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップの後に前記コールの信頼度に基づいて選択される、項目 1 または項目 2 に記載の方法。

(項目 5)

前記試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップを含み、前記バイオマーカーパネルが、前記短い標的遺伝子バリエーションを含む、項目 1 から 4 のいずれか一項に記載の方法。

(項目 6)

前記フローサイクル順序を選択するステップを含む、項目 1 から 5 のいずれか一項に記載の方法。

(項目 7)

前記標的シーケンシングデータセットおよび前記参照シーケンシングデータセットが、前記標的配列および前記参照配列を *in silico* でシーケンシングすることにより得られる、項目 1 から 6 のいずれか一項に記載の方法。

(項目 8)

前記標的シーケンシングデータセットと前記参照シーケンシングデータとが、2 カ所より多くの非連続のフロー位置において異なる、項目 1 から 7 のいずれか一項に記載の方法。

(項目 9)

前記標的シーケンシングデータセットと前記参照シーケンシングデータとが、2 カ所より多くの連続するフロー位置において異なる、項目 1 から 8 のいずれか一項に記載の方法。

(項目 10)

前記標的配列と前記参照配列とが、X カ所の塩基位置において異なり、前記標的シーケンシングデータセットと前記参照シーケンシングデータとが、(X + 2) カ所またはそれより多くの連続するフロー位置において異なる、項目 1 から 9 のいずれか一項に記載の方法。

(項目 11)

前記標的シーケンシングデータセットと前記参照シーケンシングデータセットとが、1 または複数のフローサイクルにわたって異なる、項目 1 から 10 のいずれか一項に記載の方法。

(項目 12)

前記フローシグナルが、各フロー位置におけるシーケンシングされた前記試験核酸分子の塩基の数を示す塩基カウントを含む、項目 1 から 11 のいずれか一項に記載の方法。

(項目 13)

前記フローシグナルが、各フロー位置における少なくとも 1 塩基カウントの尤度を示す統計パラメーターを含み、前記塩基カウントが、前記フロー位置におけるシーケンシングされた前記試験核酸分子の塩基の数を示す、項目 1 から 12 のいずれか一項に記載の方法。

(項目 14)

前記フローシグナルが、各フロー位置における複数の塩基カウントの尤度を示す統計パラメーターを含み、各塩基カウントが、前記フロー位置におけるシーケンシングされた前記試験核酸分子の塩基の数を示す、項目 1 から 13 のいずれか一項に記載の方法。

(項目 15)

10

20

30

40

50

ステップ（c）が、

前記試験シークエンシングデータセット内の各フロー位置における統計パラメーターであって、そのフロー位置における前記標的配列の塩基カウントに対応する統計パラメーターを選択すること、および前記試験シークエンシングデータセットが前記標的配列にマッチする尤度を示す前記マッチスコアを決定すること；または

前記試験シークエンシングデータセット中の各フロー位置における統計パラメーターであって、そのフロー位置における前記参照配列の塩基カウントに対応する統計パラメーターを選択すること、および前記試験シークエンシングデータセットが前記参照配列にマッチする尤度を示す前記マッチスコアを決定すること

を含む、項目 14 に記載の方法。

10

（項目 16）

ステップ（c）で決定される前記マッチスコアが、前記試験シークエンシングデータセット内の前記フロー位置にわたっての選択された統計パラメーターの組み合わせられた値である、項目 15 に記載の方法。

（項目 17）

ステップ（c）が、前記試験シークエンシングデータセットが前記標的配列にマッチする尤度を示す前記マッチスコアを決定することを含む、項目 1 から 16 のいずれか一項に記載の方法。

（項目 18）

ステップ（c）が、前記試験シークエンシングデータセットが前記参照配列にマッチする尤度を示す前記マッチスコアを決定することを含む、項目 1 から 17 のいずれか一項に記載の方法。

20

（項目 19）

前記 1 つまたは複数の試験シークエンシングデータセットが、複数の試験シークエンシングデータセットを含む、項目 1 から 18 のいずれか一項に記載の方法。

（項目 20）

前記短い標的遺伝子バリエーションの存在または非存在が、前記 1 つまたは複数の試験シークエンシングデータセットの各々について別々にコールされる、項目 19 に記載の方法。

（項目 21）

前記複数の試験シークエンシングデータセットの少なくとも一部分が、異なる試験シークエンシング開始位置を有する異なる試験核酸分子に関連する、項目 19 または 20 に記載の方法。

30

（項目 22）

前記フローサイクル順序が、同じ順序で反復される 4 つの別々のフローを含む、項目 1 から 21 のいずれか一項に記載の方法。

（項目 23）

前記フローサイクル順序が、5 つまたはそれより多くの別々のフローを含む、項目 1 から 22 のいずれか一項に記載の方法。

（項目 24）

1 または複数台のプロセッサを使用して前記短い標的遺伝子バリエーションを選択するステップ；

40

前記 1 つまたは複数の試験シークエンシングデータセットを前記 1 または複数台のプロセッサで受信することにより、前記 1 つまたは複数の試験シークエンシングデータセットを得るステップ；

前記 1 または複数台のプロセッサを使用して前記 1 つまたは複数のマッチスコアを決定するステップ；および

前記 1 または複数台のプロセッサを使用して前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップ

を含むコンピュータ実装方法である、項目 1 から 23 のいずれか一項に記載の方法。

（項目 25）

50

試験試料中の短い遺伝子バリエーションを検出するための方法であって、

(a) 1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、前記試験試料に由来する異なる試験核酸分子に関連し、前記第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、前記1つまたは複数の第1の試験シーケンシングデータセットが、前記ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；

(b) 1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、前記第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して前記1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、前記第1のフローサイクル順序と前記第2のフローサイクル順序が異なり、前記試験シーケンシングデータセットが、前記ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；

(c) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットの各々について、1つまたは複数の候補配列とのマッチスコアを決定するステップであって、前記マッチスコアが、前記第1の試験シーケンシングデータセット、前記第2の試験シーケンシングデータセット、または両方が前記1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および

(d) 決定されたマッチスコアを使用して、前記試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む方法。

(項目26)

前記第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して前記試験核酸分子をシーケンシングするステップ、および前記第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して前記試験核酸分子をシーケンシングするステップを含む、項目25に記載の方法。

(項目27)

前記マッチスコアが、前記第1の試験シーケンシングデータセットが前記候補配列にマッチする尤度、または前記第2の試験シーケンシングデータセットが前記候補配列にマッチする尤度を示す、項目25または26に記載の方法。

(項目28)

前記マッチスコアが、前記第1の試験シーケンシングデータセットと前記第2のシーケンシングデータセット両方が前記候補配列にマッチする尤度を示す、項目25または26に記載の方法。

(項目29)

前記1つまたは複数の候補配列が、2つまたはそれより多くの異なる候補配列を含み、前記方法が、第1のシーケンシングデータセットおよび第2のシーケンシングデータセットに関連する核酸分子の各々について：

前記2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列が、前記第1の試験シーケンシングデータセット、前記第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および前記選択候補配列を使用して、前記試験試料中の前記短い遺伝子バリエーションの存在または非存在をコールするステップを含む、項目25から28のいずれか一項に記載の方法。

(項目30)

前記2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列

10

20

30

40

50

と、前記選択候補配列とが、前記第1のフローサイクル順序または前記第2のフローサイクル順序によれば、2カ所またはそれより多くのフロー位置において異なる、項目29に記載の方法。

(項目31)

前記2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、前記選択候補配列とが、前記第1のフローサイクル順序または前記第2のフローサイクル順序によれば、2カ所またはそれより多くの非連続のフロー位置において異なる、項目29に記載の方法。

(項目32)

前記2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、前記選択候補配列とが、前記第1のフローサイクル順序または前記第2のフローサイクル順序によれば、3カ所またはそれより多くのフロー位置において異なる、項目29に記載の方法。

(項目33)

前記2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、前記選択候補配列とが、Xカ所の塩基位置において異なり、前記試験核酸分子に関連する試験シーケンシングデータセットと、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列とが、前記第1のフローサイクル順序または前記第2のフローサイクル順序によれば、(X+2)カ所またはそれより多くのフロー位置において異なる、項目29に記載の方法。

(項目34)

前記2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、前記選択候補配列とが、前記第1のフローサイクル順序または前記第2のフローサイクル順序によれば、1または複数のフローサイクルにわたって異なる、項目29に記載の方法。

(項目35)

前記フローシグナルが、各フロー位置におけるシーケンシングされた前記試験核酸分子の塩基の数を示す塩基カウントを含む、項目22から34のいずれか一項に記載の方法。

(項目36)

前記フローシグナルが、各フロー位置における少なくとも1塩基カウントの尤度を示す統計パラメータを含み、前記塩基カウントが、前記フロー位置におけるシーケンシングされた前記試験核酸分子の塩基の数を示す、項目25から35のいずれか一項に記載の方法。

(項目37)

前記フローシグナルが、各フロー位置における複数の塩基カウントの尤度を示す統計パラメータを含み、各塩基カウントが、前記フロー位置におけるシーケンシングされた前記試験核酸分子の塩基の数を示す、項目25から36のいずれか一項に記載の方法。

(項目38)

前記マッチスコアを決定するステップが、前記第1の試験シーケンシングデータセットおよび前記第2の試験シーケンシングデータセットにおける各フロー位置における統計パラメータであって、前記候補配列のそのフロー位置における塩基カウントに対応する統計パラメータを、前記1つまたは複数の異なる候補配列の各々について選択することを含む、項目37に記載の方法。

(項目39)

前記1つまたは複数の異なる候補配列について、各フロー位置における前記候補配列の塩基カウントを含む候補シーケンシングデータセットを生成するステップを含む、項目37に記載の方法。

(項目40)

前記候補シーケンシングデータセットが、*in silico*で生成される、項目39に記載の方法。

10

20

30

40

50

(項目 4 1)

前記マッチスコアが、前記第 1 の試験シーケンシングデータセットおよび前記第 2 の試験シーケンシングデータセットにおけるフロー位置にわたって選択された統計パラメーターの組み合わせられた値である、項目 3 7 から 4 0 のいずれか一項に記載の方法。

(項目 4 2)

前記試験核酸分子の少なくとも一部分が、異なるシーケンシング開始位置を有する、項目 2 5 から 4 1 のいずれか一項に記載の方法。

(項目 4 3)

短い標的遺伝子バリエーションを選択するステップであって、前記短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、前記標的シーケンシングデータセットおよび前記参照シーケンシングデータセットが、前記第 1 のフローサイクル順序または前記第 2 のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して前記標的配列をシーケンシングすることにより得られた場合、2 カ所またはそれより多くのフロー位置において異なり、前記第 1 のフローサイクル順序と前記第 2 のフローサイクル順序とが異なり、前記フロー位置が前記ヌクレオチドフローに対応する、ステップを含み、

前記 1 つまたは複数の候補配列が、前記標的配列および前記参照配列を含む、項目 2 5 から 4 2 のいずれか一項に記載の方法。

(項目 4 4)

前記短い標的遺伝子バリエーションが、前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップの前に事前選択される、項目 4 3 に記載の方法。

(項目 4 5)

前記短い標的遺伝子バリエーションが、前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップの後に前記コールの信頼度に基づいて選択される、項目 4 3 に記載の方法。

(項目 4 6)

前記試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップを含み、前記バイオマーカーパネルが、前記試験試料中に存在する前記短い標的遺伝子バリエーションを含む、項目 4 3 から 4 5 に記載の方法。

(項目 4 7)

前記参照シーケンシングデータセットが、前記第 1 のフローサイクル順序または前記第 2 のフローサイクル順序に従って別々のフローで提供される非終結ヌクレオチドを使用して前記参照配列がシーケンシングされた場合の予想参照シーケンシングデータセットを決定することにより得られる、項目 4 3 から 4 6 のいずれか一項に記載の方法。

(項目 4 8)

前記標的配列と前記参照配列とが、前記第 1 のフローサイクル順序および前記第 2 のフローサイクル順序の両方によれば、2 カ所またはそれより多くのフロー位置において異なる、項目 4 3 から 4 6 のいずれか一項に記載の方法。

(項目 4 9)

前記標的配列と前記参照配列とが、前記第 1 のフローのフローサイクル順序によれば、2 カ所またはそれより多くの非連続のフロー位置において異なる、項目 4 3 から 4 6 のいずれか一項に記載の方法。

(項目 5 0)

前記標的配列と前記参照配列とが、前記第 1 のフローサイクル順序または前記第 2 のフローサイクル順序によれば、3 カ所またはそれより多くのフロー位置において異なる、項目 4 3 から 4 6 のいずれか一項に記載の方法。

(項目 5 1)

前記標的配列と、前記参照配列とが、前記第 1 のフローサイクル順序または前記第 2 のフローサイクル順序によれば、1 または複数のフローサイクルにわたって異なる、項目 4

10

20

30

40

50

3 から 4 6 のいずれか一項に記載の方法。

(項目 5 2)

前記第 1 のフローサイクル順序または前記第 2 のフローサイクル順序が、同じ順序で反復される 4 つの別々のフローを含む、項目 4 3 から 5 1 のいずれか一項に記載の方法。

(項目 5 3)

前記第 1 のフローサイクル順序または前記第 2 のフローサイクル順序が、同じ順序で反復される 5 つまたはそれより多くの別々のフローを含む、項目 4 3 から 5 1 のいずれか一項に記載の方法。

(項目 5 4)

前記試験核酸分子をシークエンシングするステップであって、前記第 1 のフローサイクル順序に従って別々のヌクレオチドフローで非終結ヌクレオチドを提供することと、シークエンシングプライマーを伸長することと、各ヌクレオチドフローの後に前記シークエンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して前記第 1 の試験シークエンシングデータセットを生成することを含むステップ；

伸長されたシークエンシングプライマーを除去するステップ；および

同じ試験核酸分子をシークエンシングするステップであって、前記第 2 のフローサイクル順序に従って別々のヌクレオチドフローで非終結ヌクレオチドを提供することと、シークエンシングプライマーを伸長することと、各ヌクレオチドフローの後に前記シークエンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して前記第 2 の試験シークエンシングデータセットを生成することを含むステップ

を含む、項目 2 5 から 5 3 のいずれか一項に記載の方法。

(項目 5 5)

前記 1 つまたは複数の第 1 のシークエンシングデータセットを 1 または複数台のプロセッサで受信するステップ；

前記 1 つまたは複数の第 1 のシークエンシングデータセットを前記 1 または複数台のプロセッサで受信するステップ；

前記 1 または複数台のプロセッサを使用して前記マッチスコアを決定するステップ；および

前記 1 または複数台のプロセッサを使用して前記試験試料中の前記短い標的遺伝子バリエーションの存在または非存在をコールするステップ

を含むコンピュータ実装方法である、項目 2 5 から 5 4 のいずれか一項に記載の方法。

(項目 5 6)

前記別々のフローが、単一の塩基タイプを含む、項目 1 から 5 5 のいずれか一項に記載の方法またはシステム。

(項目 5 7)

前記別々のフローのうちの少なくとも 1 つが、2 つまたは 3 つの異なる塩基タイプを含む、項目 1 から 5 6 のいずれか一項に記載の方法またはシステム。

(項目 5 8)

前記試験試料中の前記短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示すバリエーションコールファイルを生成またはアップデートするステップを含む、項目 1 から 5 7 のいずれか一項に記載の方法またはシステム。

(項目 5 9)

前記試験試料中の前記短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示すレポートを生成するステップを含む、項目 1 から 5 8 のいずれか一項に記載の方法またはシステム。

(項目 6 0)

前記レポートが、前記試験試料中の前記短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示す、テキスト出力、確率的出力、数値出力またはグラフィック出力を含む、項目 5 9 に記載の方法またはシステム。

(項目 6 1)

前記レポートを患者にまたは前記患者の医療担当者に提供するステップを含む、項目 5 9 または 6 0 に記載の方法またはシステム。

(項目 6 2)

前記短い遺伝子バリエーションが、一塩基多型またはインデルを含む、項目 1 から 6 1 のいずれか一項に記載の方法またはシステム。

(項目 6 3)

核酸分子をシーケンシングする方法であって、

前記核酸分子をプライマーとハイブリダイズして、ハイブリダイズされた鋳型を形成するステップ；

5 つまたはそれより多くの別々のヌクレオチドフローを含む反復フローサイクル順序に従って別々のヌクレオチドフローで提供される標識された非終結ヌクレオチドを使用して前記プライマーを伸長するステップ；および

取り込まれた標識ヌクレオチドからのシグナルを検出する、または前記プライマーが前記ヌクレオチドフローにより伸長された場合のシグナルの非存在を検出するステップを含む、方法。

(項目 6 4)

前記シグナル、または前記シグナルの非存在を、各ヌクレオチドフロー後に検出するステップを含む、項目 6 3 に記載の方法。

(項目 6 5)

複数の核酸分子をシーケンシングするステップを含む、項目 6 3 または 6 4 に記載の方法。

(項目 6 6)

前記複数のうちの前記核酸分子が、ある遺伝子座に対して異なるシーケンシング開始位置を有する、項目 6 5 に記載の方法。

(項目 6 7)

前記フローサイクル順序が、2 カ所より多くのフロー位置において、ランダムシーケンシング開始位置の 5 % またはそれより多くにおける可能な SNP パーミュテーションの 5 0 % またはそれより多くについてのシグナル変化を誘導する、項目 6 3 から 6 6 のいずれか一項に記載の方法。

(項目 6 8)

前記フローサイクル順序が、フロー当たり 0 . 6 のまたはそれを超える塩基取り込みの効率を有する、項目 6 3 から 6 7 のいずれか一項に記載の方法。

(項目 6 9)

1 または複数台のプロセッサと、

項目 1 から 6 2 のいずれか一項に記載の方法を実行するための命令を含む 1 つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステム。

【図面の簡単な説明】

【0040】

【図 1 A】図 1 A は、T - A - C - G の反復フローサイクル順序を使用して T A T G G T C G T C G A (配列番号 1) の配列でプライマーを伸長させることにより得られたシーケンシングデータを示す。このシーケンシングデータは、伸長されたプライマー鎖を代表しており、容易に決定され得る相補鋳型鎖のシーケンシング情報は、実効的に等価である。

【0041】

【図 1 B】図 1 B は、各フロー位置における最高尤度に基づいて選択された、最も可能性が高い配列であって、シーケンシングデータが得られた配列 (星印により示されている通り) を伴う、図 1 A に示されているシーケンシングデータを示す。

【0042】

【図 1 C】図 1 C は、2 つの異なる候補配列：T A T G G T C A T C G A (配列番号 2)

10

20

30

40

50

(黒塗りの丸印) および T A T G G T C G T C G A (配列番号 1) (白抜きの丸印) を表すトレースを伴う、図 1 A に示されているシーケンシングデータを示す。シーケンシングデータが所与の配列にマッチする尤度は、各フロー位置が候補配列にマッチする尤度の積として決定することができる。

【0043】

【図 2-1】図 2 A は、2 つの候補配列 H 1 (配列番号 5) および H 2 (配列番号 6) (それらの相補配列により各々表されている) とアラインされた、シーケンシングリード R 1 (配列番号 1)、R 2 (配列番号 3) および R 3 (配列番号 4) (伸長されたプライマーの配列により各々表されている) のアラインメントを示す。図 2 B は、H 1 (黒塗りの丸印) および H 2 (白抜きの丸印) を表すトレースを伴う、R 1 に対応するシーケンシングデータを示す。図 2 C は、H 1 (黒塗りの丸印) および H 2 (白抜きの丸印) を表すトレースを伴う、R 2 に対応するシーケンシングデータを示す。図 2 D は、H 1 (黒塗りの丸印) および H 2 (白抜きの丸印) を表すトレースを伴う、R 3 に対応するシーケンシングデータを示す。

10

【図 2-2】図 2 A は、2 つの候補配列 H 1 (配列番号 5) および H 2 (配列番号 6) (それらの相補配列により各々表されている) とアラインされた、シーケンシングリード R 1 (配列番号 1)、R 2 (配列番号 3) および R 3 (配列番号 4) (伸長されたプライマーの配列により各々表されている) のアラインメントを示す。図 2 B は、H 1 (黒塗りの丸印) および H 2 (白抜きの丸印) を表すトレースを伴う、R 1 に対応するシーケンシングデータを示す。図 2 C は、H 1 (黒塗りの丸印) および H 2 (白抜きの丸印) を表すトレースを伴う、R 2 に対応するシーケンシングデータを示す。図 2 D は、H 1 (黒塗りの丸印) および H 2 (白抜きの丸印) を表すトレースを伴う、R 3 に対応するシーケンシングデータを示す。

20

【0044】

【図 3】図 3 は、試験試料中の短い遺伝子バリエーションを検出するための例示的方法のフローチャートを示す。

【0045】

【図 4】図 4 A は、第 1 のフローサイクル順序 (T-A-C-G) を使用して核酸分子をシーケンシングすることにより得られた、T A T G G T C G T C G A (配列番号 1) の伸長されたプライマー配列を有する核酸分子からのシーケンシングデータを示し、図 4 B は、第 2 のフローサイクル順序 (A-G-C-T) を使用して同じ核酸分子をシーケンシングすることにより得られたシーケンシングデータを示す。さらに、図 4 A および図 4 B 各々は、第 1 の候補配列 T A T G G T C G T C G A (配列番号 1) (黒塗りの丸印) および第 2 の候補配列 T A T G G T C A T C G A (配列番号 2) (白抜きの丸印) からのトレースを示す。図 4 A および図 4 B に示されているように、フローサイクル順序の差は、所与のフロー位置において検出されるシグナルを大きく変化させ得るので、バリエーションのコンテキストにとってより良いフローサイクルを使用すると、より有意なシグナルの相違を検出することができる。

30

【0046】

【図 5】図 5 は、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための別の例示的方法を示す。

40

【0047】

【図 6】図 6 は、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための別の例示的方法を示す。

【0048】

【図 7】図 7 は、本明細書に記載される方法を実行するために使用することができる、一実施形態によるコンピュータデバイスの例を説明する。

【0049】

【図 8】図 8 は、A-T-G-C フローサイクル順序を使用してシーケンシングされた仮想核酸分子からのシーケンシングデータを示す。可能性のあるハプロタイプ配列 T A

50

T G G T C G - T C G A (配列番号7) (H1) および T A T G G T C G A T C G (配列番号8) (H2) を使用して、トレースを生成することができ、H1は、H2と比較して1塩基欠失を有する。シーケンシングデータは、H2候補配列とのより良いマッチを有し、インデルは、この配列ではコールされない。

【0050】

【図9】図9は、4つの例示的フローサイクル順序（これらのうちの、延長フローサイクル順序である、3つを含む）について、ランダムシーケンシング開始位置が得られるSNPパーミュテーションの検出の感度を示す。図9中、x軸は、フロー相（または断片化開始位置）の分率を示し、y軸は、2カ所より多くのフロー位置においてシグナル変化を誘導したSNPパーミュテーションの分率を示す。

【発明を実施するための形態】

【0051】

対象に由来する試験試料中の1つまたは複数の短い遺伝子バリエーション、例えば、一塩基多型(SNP)、多塩基多型(MNP)またはインデルを検出するための方法が、本明細書に記載される。試験試料からの試験核酸分子に関連する試験シーケンシングデータは、試験シーケンシングデータと別の配列[例えば、試験配列、候補配列(または候補ハプロタイプ配列および/もしくは参照配列)]とのマッチを判定するために解析され、マッチの近似性を示すマッチスコア(例えば、試験シーケンシングデータを考えると、試験シーケンシングデータが比較配列の核酸分子から生じる尤度)を決定することにより表され得る。したがって、マッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在もしくはアイデンティティ、または非存在をコールすることができる。

【0052】

試験シーケンシングデータセットは、計算効率の良い解析をもたらすように独特に構造化される。例えば、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより、試験シーケンシングデータセットを生成することができる。したがって、核酸分子の試験シーケンシングデータセットは、特定のヌクレオチドのフローに各々が対応するフロー位置にあるフローシグナルを含む。この独特に構造化されたデータセットを使用して、核酸分子(単数)[または分子(複数)]を、「塩基空間」(「ヌクレオチド空間」または「配列空間」とも呼ばれる)ではなく「フロー空間」で解析することができる。フロー空間データは、塩基空間データが保有しない、フローサイクル順序に関する追加情報に依存する。フロー空間で収集されるデータの解析は、塩基空間に変換されるデータまたは塩基空間で収集されるデータの解析に勝る利点を少なくとも2つ提供する。第1に、試験核酸分子中に最も多く見られるバリエーションタイプ(置換SNP)は、フロー空間で参照配列と比較すると2つまたはそれより多くの明確に異なるフローシグナル(これらは全フローサイクルにわたって、またはそれを超えて、伝播し得る)を生じさせる結果となるが、塩基空間で配列を解析するとデータシグナルを1つだけしか入手できない。つまり、塩基空間では、各塩基位置が単一シグナルと関連付けられ、バリエーション塩基は、バリエーション塩基のシグナルのみに影響を与え、隣接シグナルには影響を与えない。フロー空間では、バリエーションは、複数のフロー位置に影響を与えることができ、ある特定のバリエーションについては、バリエーションは、参照配列と比較してその後のフローグラムシグナルのシフトを誘導することができ、それによって実質的にバリエーション検出を継続的に強化し得る。第2に、試験核酸分子の配列と1つまたは複数の候補配列とを直接アラインメントしなくても、フロー空間データを解析して1つまたは複数の候補フロー空間配列とのマッチを判定することができる。配列アラインメントは計算コストが高いが、本明細書に記載されるマッチ解析を使用してそれを単純化することができる。

【0053】

所与の遺伝子バリエーションについてのフロー空間における多重シグナルインジケータは、塩基空間解析で同定され得る単一シグナルインジケータと比べてバリエーションコール正確度を増加させる。さらに、フローシグナルの相違の数が多いほど、バリエーションコールが

検出されることになる尤度が増加する。本明細書中でさらに論じられるように、ある特定の状況では、信頼度が高い事前選択されたバリエーションをコールすることが望ましく、それらのバリエーションおよび／またはフロー順序を、所望の数のフローシグナルの相違が確実に生成されるように選択して、遺伝子バリエーションを高い信頼性でコールすることができる。核酸分子のシーケンシングデータセットを候補配列と比較して、試験シーケンシングデータセットが候補配列にマッチする尤度を示すマッチスコアを決定することができる。

【0054】

決定された配列と候補配列（例えば、候補ハプロタイプ配列）の塩基空間におけるアライメントは、計算コストが高く、現在のところ、ゲノム解析ツールキット（GATK）HaplotypeCallerでのほぼ計算集約的なステップである。HaplotypeCallerの中のPairHMMは、各シーケンシングリードと各ハプロタイプのアライメントを行い、塩基の品質をエラーの推定値として使用して、シーケンシングリードが得られハプロタイプの尤度を判定する。しかし、本明細書に記載される方法で使用されるデータセットの構造は、エラーモード尤度を保持し、それによってバリエーションコーリングの計算効率がより良くなる。例えば、所与の遺伝子型の尤度を、その遺伝子型を有する配列とアラインする各フロー位置における尤度の積として、簡単に決定することができる。このフロー空間決定尤度は、HaplotypeCallerのPairHMMモジュールを、より計算効率の良いバリエーションコールに置き換えることができる。

【0055】

シーケンシングデータセット内の任意のフロー位置のフローシグナルは、任意の塩基位置の核酸分子をシーケンシングするために使用されるフロー順序がその位置におけるフローシグナルに影響を与え得ることから、フロー順序依存性である。本明細書中でさらに説明されるように、この発見を1つまたは複数の方法で利用することができる。第1に、同じ遺伝子座で重複する核酸分子のランダム断片化（in vivo断片化、例えば無細胞DNA、あるいはin vitro断片化、例えば超音波処理もしくは酵素的消化によるもの）は、それらの核酸分子についての複数の異なるシーケンシング開始部位（遺伝子座に関連して）を生じさせる結果となる。一部のケースでは、異なるフローコンテキストが遺伝子座で入手可能である（例えば、異なるフロー順序で再シーケンシングした場合、または準周期的なフロー順序を使用した場合）。したがって、たとえ他の核酸分子が信頼度のより低いシグナル（例えば、単一フローシグナル変化）を生じさせる結果となっても、バリエーションについての高感度フローシグナルを用いて（例えば、参照または非選択候補配列と比較して2つまたはそれより多くのフローシグナルの相違を用いて）単一核酸分子に基づいて、遺伝子座にあるバリエーションを正確に検出することができる。第2に、所与の核酸分子を、第1のフロー順序を使用してシーケンシングし、第2の（異なる）フロー順序を使用して再シーケンシングすることができ、この結果として、その核酸分子にわたって異なるフロー配列コンテキストが得られる。1つのフロー順序を使用して、バリエーションを有する核酸分子とそのバリエーションを有する候補配列との尤度マッチが低かったとしても、第2のフロー順序を使用して、その核酸分子とその候補配列との尤度マッチが高いことがある。第3に、フロー順序は、延長フローサイクル（例えば、サイクル内に4つより多くの塩基タイプがある）であることがあり、これは、そのフロー順序が、単に、4つの塩基タイプA、C、TおよびGの周期的な4フロー反復でないことを意味する。一部のケースでは、反復単位は、4塩基より長く、例えば、すべての可能な2塩基フロー配列（すなわち、すべてのX-Yペアが反復単位内にあり、ここでのXは、4つすべての塩基であり、Yは、非X塩基の各々である）または3塩基フロー配列（すなわち、すべての可能なX-Y-Zパーミュテーションが反復単位内にある）を含むパターンである。第4に、フローシーケンシング順序を、特異的な遺伝子バリエーションを標的とするように選択することができる。

【0056】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試

10

20

30

40

50

験シーケンシングデータセットが、試験試料に由来する試験核酸分子に関連し、試験シーケンシングデータセットが、フロー順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより生成され、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（b）試験シーケンシングデータセットに関連する試験核酸分子の各々について、試験シーケンシングデータセットが1つまたは複数の候補配列にマッチする尤度を示すマッチスコアを決定するステップ；および（c）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0057】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；（c）試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0058】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（b）1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、第1のフローサイクル順序と第2のフローサイクル順序が異なり、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（c）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットの各々について、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列のマッチする尤度を示す、ステップ

；および（d）決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0059】

本明細書に記載される方法は、コンピュータ実装方法であり得、方法の1つまたは複数のステップを、例えば、1または複数台のコンピュータプロセッサを使用して遂行することができる。

【0060】

1つまたは複数のプログラムを記憶する非一過性コンピュータ可読記憶媒体であって、1つまたは複数のプログラムが、電子デバイスの1または複数台のプロセッサにより実行されたとき、電子デバイスに、本明細書に記載される方法のいずれか1つまたは複数を実行させる命令を含む、非一過性コンピュータ可読記憶媒体も、本明細書で提供される。

【0061】

1または複数台のプロセッサと、メモリーと、そのメモリーに記憶された1つまたは複数のプログラムとを含む、電子デバイスであって、1つまたは複数のプログラムが、1または複数台のプロセッサにより実行されるように構成されている、電子デバイスが、本明細書にさらに記載される。1つまたは複数のプログラムは、本明細書に記載される方法のいずれか1つまたは複数を実行するための命令を含み得る。

【0062】

核酸分子をシーケンシングする方法も、本明細書に記載される。例えば、核酸分子をシーケンシングする方法は、核酸分子をプライマーとハイブリダイズして、ハイブリダイズされた鑄型を形成するステップ；5つまたはそれより多くの別々のヌクレオチドフローを含む反復フローサイクル順序に従って別々のヌクレオチドフローで提供される標識された非終結ヌクレオチドを使用してプライマーを伸長するステップ；および取り込まれた標識ヌクレオチドからのシグナルを検出する、またはプライマーがヌクレオチドフローにより伸長された場合のシグナルの非存在を検出するステップを含み得る。

定義

【0063】

本明細書で使用される場合、単数形「1つの（a）」、「1つの（an）」および「その（the）」は、文脈による別段の明白な指示がない限り、複数形の言及対象を含む。

【0064】

本明細書での「約」ある値またはパラメーターへの言及は、その値またはパラメーター自体に関する変動を含む（および記載する）。例えば、「約X」に言及する記載は、「X」の記載を含む。

【0065】

所与の配列についての「予想シーケンシングデータ」または「予想シーケンシングデータセット」は、フロー順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して配列がシーケンシングされた場合に生成されることになる、計算されたシーケンシングデータを指す。予想シーケンシングデータセットまたは予想シーケンシングデータセットを、例えばコンピュータモデリングにより（すなわち、*in silico*）、決定することができる。

【0066】

「フロー順序」は、非終結ヌクレオチドを使用して核酸分子をシーケンシングするために使用される別々のヌクレオチドフローの順序を指す。フロー順序を反復単位のサイクルに分けることができ、反復単位のフロー順序は、「フローサイクル順序」と呼ばれる。

「フロー位置」は、シーケンシングプロセス中の所与の別々のヌクレオチドフローの逐次的位置を指す。

【0067】

用語「個体」、「患者」および「対象」は、同義語として使用され、ヒトを含む動物を指す。

【0068】

10

20

30

40

50

用語「標識」は、本明細書で使用される場合、別の部分、例えばヌクレオチドまたはヌクレオチドアナログ、とカップリングされる、またはカップリングされ得る、検出可能な部分を指す。標識は、標識に送達されたシグナルを放出することまたはシグナルを変更することができ、したがって、標識の存在または非存在を検出することができる。一部のケースでは、カップリングは、リンカーを介してのカップリングであり得、リンカーは、切断可能、例えば、光切断可能（例えば、紫外線下で切断可能）、化学的に切断可能（例えば、ジチオトレイトール（DTT）、トリス（2-カルボキシエチル）ホスフィン（TCPEP）などの、還元剤によって）、または酵素的に切断可能（例えば、エステラーゼ、リパーゼ、ペプチダーゼもしくはプロテアーゼによって）であり得る。一部の実施形態では、標識は、フルオロフォアである。

10

【0069】

「非終結ヌクレオチド」は、ポリメラーゼまたはトランスクリプターゼを使用してポリヌクレオチドの3'末端に結合させることができる、およびポリメラーゼまたはトランスクリプターゼを使用してそれに結合された別の非終結核酸を有することができ、そのヌクレオチドから保護基または可逆的ターミネーターを除去する必要がない、核酸部分である。天然に存在する核酸は、非終結核酸の一種である。非終結核酸は、標識されていることがあり、または未標識であることもある。

【0070】

「ヌクレオチドフロー」は、一連の1つまたは複数の非終結ヌクレオチド（標識されていることがあり、またはその一部分が標識されていることもある）を指す。

20

【0071】

「短い遺伝子バリエーション」は、長さが連続する10塩基またはそれ未満（すなわち、長さが10、9、8、7、6、5、4、3、2または1塩基）である遺伝的多型（すなわち、突然変異）を指す。この用語は、長さが連続する10塩基またはそれ未満である、一塩基多型（SNP）、多塩基多型（MNP）およびインデルを含む。

【0072】

本明細書に記載される本発明の態様および変形形態が、態様および変形形態「からなること」および／または「から本質的になること」を含むことは理解されよう。

【0073】

値の範囲が提供される場合、その範囲の上限値と下限値の間に介在する各々の値、およびその述べられている範囲内の、任意の他の述べられているまたは介在する値が、本開示の範囲内に包含されることは、理解されるはずである。述べられている範囲が上限値または下限値を含む場合、これらの含まれる限界値のどちらかを含まない範囲もまた、本開示に含まれる。

30

【0074】

本明細書に記載される解析方法の一部は、配列を参照配列にマッピングすること、配列情報を判定すること、および／または配列情報を解析することを含む。相補配列を容易に決定および／または解析することができること、ならびに本明細書で提供される説明が、相補配列に関して遂行される解析方法を包含することは、当技術分野では十分に理解されよう。

40

【0075】

本明細書で使用される節の見出しは、単に構成のためのものであり、記載される主題を限定するものと解釈すべきでない。この説明は、当業者による本発明の実施および使用を可能にするために提供され、特許出願およびその要件に関連して提供される。記載される実施形態の様々な修飾形態が当業者には容易に分かることになり、本明細書における一般原理を他の実施形態に応用することができる。したがって、本発明は、示される実施形態に限定されるように意図されたものではなく、本発明には、本明細書に記載される原理および特徴に対応する最も広い範囲が与えられる。

【0076】

図は、様々な実施形態によるプロセスを例証する。これらの例示的プロセスでは、一部

50

のブロックは、必要に応じて組み合わせられ、一部のブロックの順序は、必要に応じて変更され、一部のブロックは、必要に応じて割愛される。一部の例では、追加のステップが例示的プロセスと組み合わせて遂行され得る。したがって、例証される（および下記により詳細に説明される）ような操作は、本質的に例示的なものであり、したがって、限定と見なすべきではない。

【0077】

本明細書で言及されるすべての公表文献、特許および特許出願の開示は、これにより各々その全体が参照により本明細書に取り込まれる。参照により取り込まれるいずれかの参考文献が本開示と矛盾する場合には、本開示が優先されるものとする。

フローシーケンシング法

【0078】

任意の所与のフロー位置において単一のタイプのヌクレオチドが伸長プライマーに到達できる所定のフローサイクルに従って鋳型ポリヌクレオチド分子に結合されたプライマーを伸長するステップを含む、フローシーケンシング法を使用して、シーケンシングデータを生成することができる。一部の実施形態では、特定のタイプのヌクレオチドの少なくとも一部は、標識を含み、標識されたヌクレオチドが伸長プライマーに取り込まれると、この標識が検出可能なシグナルをもたらす。そのようなヌクレオチドが伸長されたプライマーに取り込まれることにより得られる配列は、鋳型ポリヌクレオチド分子の配列の逆相補配列であるはずである。一部の実施形態では、例えば、シーケンシングデータは、標識されたヌクレオチドを使用してプライマーを伸長するステップ、および伸長プライマーに取り込まれた標識されたヌクレオチドの存在または非存在を検出するステップを含むフローシーケンシング法を使用して生成される。フローシーケンシング法は、「自然な合成によるシーケンシング」または「非終結型の合成によるシーケンシング」方法と呼ばれることもある。例示的な方法は、その全体が参照により本明細書に取り込まれる米国特許第8,772,473号に記載されている。以下の説明は、フローシーケンシング法に関して提供されるが、シーケンシングされる領域のすべてまたは一部分をシーケンシングするために他のシーケンシング法が使用され得ることは、理解されよう。例えば、本明細書で論じられるシーケンシングデータを、パイロシーケンシング法を使用して生成することができる。

【0079】

フローシーケンシングは、ポリヌクレオチドとハイブリダイズされたプライマーを伸長するためのヌクレオチドの使用を含む。所与の塩基タイプのヌクレオチド（例えば、A、C、G、T、Uなど）をハイブリダイズされた鋳型と混合して、相補的塩基が鋳型鎖内に存在する場合には、プライマーを伸長することができる。ヌクレオチドは、例えば、非終結ヌクレオチドであり得る。ヌクレオチドが、非終結ヌクレオチドであるとき、1つより多くの連続する相補的塩基が鋳型鎖内に存在する場合には、1つより多くの連続する塩基を伸長プライマー鎖に取り込むことができる。非終結ヌクレオチドは、3'可逆的ターミネーターを有するヌクレオチドと対照をなし、一般に、連続ヌクレオチドが結合される前にブロック基は除去される。相補的塩基が鋳型鎖内に存在しない場合、鋳型鎖内の次の塩基と相補的であるヌクレオチドが導入されるまで、プライマー伸長は停止する。ヌクレオチドの少なくとも一部分に標識することができ、その結果、取り込みを検出することができる。最も一般的には、単一のヌクレオチドタイプのみが一度に導入される（すなわち、個々に付加される）が、ある特定の実施形態では、2つまたは3つの異なるタイプのヌクレオチドが同時に導入されることもある。この方法論は、あらゆる単一塩基の伸長後、ターミネーターが反転されて次に続く塩基の取り込みが可能になるまで、プライマー伸長が停止される、可逆的ターミネーターを使用するシーケンシング法と対比され得る。

【0080】

プライマー伸長の過程でヌクレオチドをフロー順序で導入することができ、この過程をフローサイクルにさらに分けることができる。フローサイクルは、反復されるヌクレオチドフロー順序であり、任意の長さのものであり得る。ヌクレオチドが段階的に付加され、

10

20

30

40

50

これにより、付加されたヌクレオチドを鋳型鎖内に存在する相補的塩基のシーケンシングプライマーの末端に取り込むことが可能になる。単に例として、フローサイクルのフロー順序は、A-T-G-Cであることもあり、またはフローサイクル順序は、A-T-C-Gであることもある。代替順序を当業者は容易に企図することができる。フローサイクル順序は、いずれの長さのものであってもよいが、4つの固有の塩基タイプ（任意の順序でA、T、CおよびG）を含有するフローサイクルが最も一般的である。一部の実施形態では、フローサイクルは、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20またはそれより多くの別々のヌクレオチドフローをフローサイクル順序で含む。単に例として、フローサイクル順序は、T-C-A-C-G-A-A-T-G-C-A-T-G-C-T-A-Gであり得、これら16の別々に提供されるヌクレオチドが数サイクルにわたってこのフローサイクル順序で提供される。異なるヌクレオチドの導入と導入の間に、例えば洗浄液でシーケンシングプラットフォームを洗浄することにより、取り込まれていないヌクレオチドを除去することができる。

【0081】

ポリメラーゼを使用して、1つまたは複数のヌクレオチドをプライマーの末端に鋳型依存的に取り込むことによりシーケンシングプライマーを伸長させることができる。一部の実施形態では、ポリメラーゼは、DNAポリメラーゼである。ポリメラーゼは、天然に存在するポリメラーゼであることもあり、または合成（例えば、突然変異型）ポリメラーゼであることもある。ポリメラーゼをプライマー伸長の最初のステップで付加させることができるが、補足ポリメラーゼを、必要に応じて、シーケンシング中に、例えば、ヌクレオチドの段階的付加を用いて、またはいくつかのフローサイクル後に、付加させることができる。例示的なポリメラーゼとしては、DNAポリメラーゼ、RNAポリメラーゼ、熱安定性ポリメラーゼ、野生型ポリメラーゼ、改変ポリメラーゼ、Bst DNAポリメラーゼ、Bst 2.0 DNAポリメラーゼ、Bst 3.0 DNAポリメラーゼ、Bsu DNAポリメラーゼ、E.coli DNAポリメラーゼI、T7 DNAポリメラーゼ、バクテリオファージT4 DNAポリメラーゼ Φ29（ファイ29）DNAポリメラーゼ、Taqポリメラーゼ、Tthポリメラーゼ、Tliポリメラーゼ、Pfuポリメラーゼ、およびSeq Amp DNAポリメラーゼが、挙げられる。

【0082】

導入されるヌクレオチドは、鋳型鎖の配列を決定する場合、標識ヌクレオチドを含むことができ、取り込まれた標識核酸の存在または非存在を検出して配列を決定することができる。標識は、例えば、光学活性標識（例えば、蛍光標識）または放射性標識であることがあり、標識により放出または変更されたシグナルを、検出器を使用して検出することができる。鋳型ポリヌクレオチドとハイブリダイズされたプライマーに取り込まれた標識ヌクレオチドの存在または非存在を検出することができ、このことによって配列の決定が（例えば、フローグラムを生成することにより）可能になる。一部の実施形態では、標識ヌクレオチドは、蛍光部分、発光部分、または他の光出射部分で標識される。一部の実施形態では、標識は、リンカーを介してヌクレオチドに結合される。一部の実施形態では、リンカーは、例えば、光化学的または化学的切断反応によって、切断可能である。例えば、標識を、検出後かつ連続ヌクレオチドの取り込み前に切断することができる。一部の実施形態では、標識（またはリンカー）は、ヌクレオチド塩基に結合されるか、または新生DNA鎖の延長に干渉しないヌクレオチド上の別の部位に結合される。一部の実施形態では、リンカーは、ジスルフィドまたはPEG含有部分を含む。

【0083】

一部の実施形態では、導入されるヌクレオチドは、非標識ヌクレオチドのみを含み、一部の実施形態では、ヌクレオチドは、標識ヌクレオチドと非標識ヌクレオチドの混合物を含む。例えば、一部の実施形態では、全ヌクレオチドと比較して標識ヌクレオチドの部分は、約90%もしくはそれ未満、約80%もしくはそれ未満、約70%もしくはそれ未満、約60%もしくはそれ未満、約50%もしくはそれ未満、約40%もしくはそれ未満、約30%もしくはそれ未満、約20%もしくはそれ未満、約10%もしくはそれ未満、約

5%もしくはそれ未満、約4%もしくはそれ未満、約3%もしくはそれ未満、約2.5%もしくはそれ未満、約2%もしくはそれ未満、約1.5%もしくはそれ未満、約1%もしくはそれ未満、約0.5%もしくはそれ未満、約0.25%もしくはそれ未満、約0.1%もしくはそれ未満、約0.05%もしくはそれ未満、約0.025%もしくはそれ未満、または約0.01%もしくはそれ未満である。一部の実施形態では、全ヌクレオチドと比較して標識ヌクレオチドの部分は、約100%であり、約95%であるかもしくはそれより多く、約90%であるかもしくはそれより多く、約80%であるかもしくはそれより多く、約70%であるかもしくはそれより多く、約60%であるかもしくはそれより多く、約50%であるかもしくはそれより多く、約40%であるかもしくはそれより多く、約30%であるかもしくはそれより多く、約20%であるかもしくはそれより多く、約10%であるかもしくはそれより多く、約5%であるかもしくはそれより多く、約4%であるかもしくはそれより多く、約3%であるかもしくはそれより多く、約2.5%であるかもしくはそれより多く、約2%であるかもしくはそれより多く、約1.5%であるかもしくはそれより多く、約1%であるかもしくはそれより多く、約0.5%であるかもしくはそれより多く、約0.25%であるかもしくはそれより多く、約0.1%であるかもしくはそれより多く、約0.05%であるかもしくはそれより多く、約0.025%であるかもしくはそれより多く、または約0.01%であるかまたはそれより多い。一部の実施形態では、全ヌクレオチドと比較して標識ヌクレオチドの部分は、約0.01%～約100%、例えば、約0.01%～約0.025%、約0.025%～約0.05%、約0.05%～約0.1%、約0.1%～約0.25%、約0.25%～約0.5%、約0.5%～約1%、約1%～約1.5%、約1.5%～約2%、約2%～約2.5%、約2.5%～約3%、約3%～約4%、約4%～約5%、約5%～約10%、約10%～約20%、約20%～約30%、約30%～約40%、約40%～約50%、約50%～約60%、約60%～約70%、約70%～約80%、約80%～約90%、約90%～100%未満、または約90%～約100%である。

【0084】

シーケンシングデータを生成する前に、ポリヌクレオチドは、ハイブリダイズされた鋳型を生成するためにシーケンシングプライマーとハイブリダイズされる。ポリヌクレオチドをシーケンシングライブラリー調製中にアダプターにライゲーションすることができる。アダプターは、シーケンシングプライマーとハイブリダイズするハイブリダイゼーション配列を含むことができる。例えば、アダプターのハイブリダイゼーション配列は、複数の異なるポリヌクレオチドにわたって一様な配列であることがあり、シーケンシングプライマーは、一様なシーケンシングプライマーであることがある。これは、シーケンシングライブラリー内の異なるポリヌクレオチドの多重シーケンシングを可能にする。

【0085】

ポリヌクレオチドをシーケンシングのために表面（例えば、固体支持体）に結合させることができる。ポリヌクレオチドを（例えば、ブリッジ増幅または他の増幅技法により）増幅させて、ポリヌクレオチドシーケンシングコロニーを生成することができる。クラスター内の増幅されたポリヌクレオチドは、実質的に同一または相補的である（増幅プロセス中に多少のエラーが導入されることがあり、その結果、ポリヌクレオチドの一部分は、元のポリヌクレオチドと必ずしも同一でないことがある）。コロニー形成により、検出器が標識ヌクレオチド取り込みをコロニーごとに正確に検出することができるようなシグナル増幅が可能になる。一部のケースでは、コロニーは、エマルジョンPCRを使用してビーズ上に形成され、ビーズがシーケンシング面全体に分配される。シーケンシングのためのシステムおよび方法の例は、その全体が参照により本明細書に取り込まれる米国特許出願第10,344,328号において見つけることができる。

【0086】

ポリヌクレオチドとハイブリダイズされたプライマーは、フロー順序に従って別々のヌクレオチドフロー（これらは、フローサイクル順序に従って周期的であり得る）を使用し

て核酸分子を通して伸長され、ヌクレオチドの取り込みを上記の通り検出することができ、それによって、核酸分子についてのシーケンシングデータセットを生成することができる。

【0087】

フローシーケンシングを使用するプライマー伸長は、長さが数百またはさらには数千ほともの塩基のロングレンジシーケンシングを可能にする。フローステップまたはサイクルの数を増加または減少させて、所望のシーケンシング長を得ることができる。プライマーの伸長は、1つまたは複数の異なる塩基タイプを有するヌクレオチドを使用するプライマーの段階的伸長のための1つまたは複数のフローステップを含むことができる。一部の実施形態では、プライマー伸長は、1～約1000ステップの間のフローステップ、例えば、1～約10ステップの間のフローステップ、約10～約20ステップの間のフローステップ、約20～約50ステップの間のフローステップ、約50～約100ステップの間のフローステップ、約100～約250ステップの間のフローステップ、約250～約500ステップの間のフローステップ、または約500～約1000ステップの間のフローステップを含む。フローステップを同一のまたは異なるフローサイクルに分割することができる。プライマーに取り込まれる塩基の数は、シーケンシングされる領域の配列、およびプライマーを伸長するために使用されるフロー順序に依存する。一部の実施形態では、シーケンシングされる領域は、長さ約1塩基～約4000塩基、例えば、長さ約1塩基～約10塩基、長さ約10塩基～約20塩基、長さ約20塩基～約50塩基、長さ約50塩基～約100塩基、長さ約100塩基～約250塩基、長さ約250塩基～約500塩基、長さ約500塩基～約1000塩基、長さ約1000塩基～約2000塩基、または長さ約2000塩基～約4000塩基である。

【0088】

本明細書に記載される方法で使用されるポリヌクレオチドは、任意の好適な生物源、例えば、組織試料、血液試料、血漿試料、唾液試料、糞便試料、または尿試料から得ることができる。ポリヌクレオチドは、DNAポリヌクレオチドであることもあり、またはRNAポリヌクレオチドであることもある。一部の実施形態では、RNAポリヌクレオチドは、ポリヌクレオチドをシーケンシングプライマーとハイブリダイズする前にDNAポリヌクレオチドに逆転写される。一部の実施形態では、ポリヌクレオチドは、無細胞DNA (cfDNA)、例えば、循環腫瘍DNA (ctDNA) または胎児無細胞DNAである。核酸分子を、例えば *in vivo* で (例えば、cfDNAにおけるように) または *in vitro* で (例えば、超音波処理または酵素的断片化により)、ランダムに断片化することができる。

【0089】

ポリヌクレオチドのライブラリーを公知の方法によって調製することができる。一部の実施形態では、ポリヌクレオチドをアダプター配列にライゲーションすることができる。アダプター配列は、カップリングされたシーケンシングリードペアの生成中に伸長されたプライマーとハイブリダイズしたハイブリダイゼーション配列を含み得る。

【0090】

一部の実施形態では、シーケンシングデータは、シーケンシングコロニー (シーケンシングクラスターとも呼ばれる) を確立する前に核酸分子を増幅することなく得られる。シーケンシングコロニーを生成するための方法としては、ブリッジ増幅またはエマルジョンPCRが挙げられる。ショットガンシーケンシング、およびコンセンサス配列のコーリングに頼る方法は、一般に、固有分子識別子 (UMI) を使用して核酸分子を標識し、その核酸分子を増幅させて、独立してシーケンシングされる同じ核酸分子の非常に多数のコピーを生成する。次いで、増幅された核酸分子を表面に結合させ、ブリッジ増幅させて、独立してシーケンシングされるシーケンシングクラスターを生成し得る。次いで、UMIを使用して、独立してシーケンシングされた核酸分子を関連付けることができる。しかし、増幅プロセスは、例えばDNAポリメラーゼの限られた忠実度に起因して、核酸分子にエラーを導入し得る。一部の実施形態では、核酸分子は、シーケンシ

ングデータを得るためのコロニーを生成するための増幅の前に増幅されない。一部の実施形態では、核酸シーケンシングデータは、固有分子識別子（UMI）を使用せずに得られる。

シーケンシングデータセットおよびバリエーション検出

【0091】

シーケンシングデータを、取り込まれたヌクレオチドの検出およびヌクレオチド導入の順序に基づいて生成することができる。以下の伸長される配列（すなわち、対応する鋳型配列の各逆相補配列）：CTG、CAG、CCG、CGT、およびCAT（先行する配列も後続の配列もシーケンシング法に供されないと仮定して）、ならびにT-A-C-Gの反復フローサイクル（つまり、反復サイクル中のT、A、CおよびGヌクレオチドの逐次的付加）を例にとる。所与のフロー位置における特定のタイプのヌクレオチドは、相補的塩基が鋳型ポリヌクレオチド中に存在する場合にのみプライマーに取り込まれることになる。結果として生じる例示的なフローグラムが表1に示され、この表中の1は、導入されたヌクレオチドが取り込まれること示し、0は、導入されたヌクレオチドが取り込まれないことを示す。フローグラムを使用して、鋳型鎖の配列を導出することができる。例えば、本明細書で論じられるシーケンシングデータ（例えば、フローグラム）は、伸長されたプライマー鎖およびその逆相補鎖を表し、この逆相補鎖は、鋳型鎖の配列を表すために容易に決定され得る。表1中のアスタリスク（*）は、伸長されたシーケンシング鎖（例えば、より長い鋳型鎖）に追加のヌクレオチドが取り込まれた場合にシグナルがシーケンシングデータ中に存在し得ることを示す。

【表1】

表1

	サイクル1				サイクル2				サイクル3			
フロー位置	1	2	3	4	5	6	7	8	9	10	11	12
フロー中の塩基	T	A	C	G	T	A	C	G	T	A	C	G
伸長された配列: CTG	0	0	1	0	1	0	0	1	*	*	*	*
伸長された配列: CAG	0	0	1	0	0	1	0	1	*	*	*	*
伸長された配列: CCG	0	0	2	1	*	*	*	*	*	*	*	*
伸長された配列: CGT	0	0	1	1	1	*	*	*	*	*	*	*
伸長された配列: CAT	0	0	1	0	0	1	0	0	1	*	*	*

【0092】

フローグラムは、バイナリであることもあり、ノンバイナリであることもある。バイナリフローグラムは、取り込まれたヌクレオチドの存在（1）または非存在（0）を検出する。ノンバイナリフローグラムは、各々の段階的導入から取り込まれたヌクレオチドの数をより定量的に決定することができる。例えば、CCGの伸長された配列は、同じCフローの中（例えば、フロー位置3における）伸長プライマー内への2つのC塩基の取り込みを含むことになり、標識された塩基により放出されるシグナルは、単一塩基取り込みに相当する強度レベルより高い強度を有することになる。このことが表1に示されている。ノンバイナリフローグラムはまた、塩基の存在または非存在を示し、所与のフロー位置における各伸長プライマーに取り込まれる可能性が高い塩基の数を含む追加情報を提供することができる。値が整数である必要はない。一部のケースでは、値は、所与のフロー位置に取り込まれる塩基の数の不確実性および／または確率を反映していることもある。

【0093】

一部の実施形態では、シーケンシングデータセットは、各フロー位置に取り込まれているシーケンシングされた核酸分子中の塩基の数を示す塩基カウントを表すフローシグナルを含む。例えば、表1に示されているように、T-A-C-Gフローサイクル順序を使用してCTG配列で伸長されたプライマーは、位置3に1の値を有し、これは、その位置における1の塩基カウントを示す（この1塩基は、シーケンシングされた鋳型鎖内のGと相補的であるCである）。また表1において、T-A-C-Gフローサイクル順序を

使用して C C G 配列で伸長されたプライマーは、位置 3 に 2 の値を有し、これは、このフロー位置にある間の伸長プライマーのその位置における 2 の塩基カウントを示す。ここで、2 塩基は、伸長プライマー配列内の C C G 配列の最初の C - C 配列を指し、この配列は、鋳型鎖内の G - G 配列と相補的である。

【0094】

シーケンシングデータセット内のフローシグナルは、各フロー位置における 1 または複数の塩基カウントについての尤度または信頼区間を示す 1 つまたは複数の統計パラメーターを含み得る。一部の実施形態では、フローシグナルは、シーケンシング中にシーケンシングプライマーに取り込まれる 1 つまたは複数の塩基の蛍光シグナルなどの、シーケンシングプロセス中に検出されるアナログシグナルから決定される。一部のケースでは、アナログシグナルを処理して統計パラメーターを生成することができる。例えば、その全体が参照により本明細書に取り込まれる公開国際特許出願 WO 2019084158 A1 に記載されているように、機械学習アルゴリズムを使用してアナログシーケンシングシグナルのコンテキスト効果について補正することができる。ゼロまたはそれを超える整数の塩基がいずれかの所与のフロー位置に取り込まれるが、所与のアナログシグナルは、そのアナログシグナルと完全にマッチしないことがある。したがって、検出されたシグナルを考えれば、フロー位置に取り込まれる塩基の数の尤度を示す統計パラメーターを決定することができる。単に例として、表 1 の C C G 配列について、フローシグナルがフロー位置 3 に取り込まれた 2 塩基を示す尤度は、0.999 であり得、フローシグナルがフロー位置 3 に取り込まれた 1 塩基を示す尤度は、0.001 であり得る。フローシグナルが、各フロー位置における複数の塩基カウントについての尤度を示す統計パラメーターを含む場合、シーケンシングデータセットを疎行列としてフォーマットすることができる。単に例として、T - A - C - G の反復フローサイクル順序を使用して T A T G G T C G T C G A (配列番号 1) の配列で伸長されたプライマーは、図 1 A に示されているシーケンシングデータセットを生じさせる結果となり得る。統計パラメーターまたは尤度値は、例えば、シーケンシング中のアナログシグナルの検出中に存在するノイズまたは他のアーチファクトによって、異なり得る。一部の実施形態では、統計パラメーターまたは尤度が所定の閾値よりも下であった場合、実質的にゼロである所定の非ゼロ値（すなわち、何らかの非常に小さい値または無視できる値）にパラメーターを設定して、真のゼロ値を用いると計算誤差が生じるか、または可能性の低さのレベル同士、例えば、非常に可能性の低いレベル (0.0001) とあり得ないレベル (0) とが十分に区別されなくなる可能性がある、本明細書でさらに論じられる統計解析を補助することができる。

【0095】

所与の配列についてのシーケンシングデータセットの尤度を示す値を、配列アラインメントなしにシーケンシングデータセットから決定することができる。例えば、データが得られる可能性の最も高い配列を、図 1 B に (図 1 A に示されているのと同じデータを使用して) 星印により示されているように、各フロー位置において最高尤度を有する塩基カウントを選択することにより決定することができる。したがって、プライマー伸長の配列を、各フロー位置において可能性の最も高い塩基カウントに従って決定することができる：T A T G G T C G T C G A (配列番号 1)。このことから、逆相補配列（すなわち、鋳型鎖）を容易に決定することができる。さらに、T A T G G T C G T C G A (配列番号 1) 配列 (または逆相補配列) が得られる、このシーケンシングデータセットの尤度を、各フロー位置における選択尤度の積として決定することができる。

【0096】

核酸分子に関連するシーケンシングデータセットを 1 つまたは複数の (例えば、2、3、4、5、6 もしくはそれより多くの) 可能性のある候補配列と比較することができる。シーケンシングデータセットと候補配列との (下記で論じられるような、マッチスコアに基づく) 近似マッチは、そのシーケンシングデータセットが、近似マッチする候補配列と同じ配列を有する核酸分子から生じた可能性が高いことを示す。一部の実施形態では、シーケンシングされた核酸分子の配列を、参照配列に (例えば、バローズ・ホイー

ラーアラインメント（BWA）アルゴリズムまたは他の好適なアラインメントアルゴリズムを使用して）マッピングして、その配列についての遺伝子座（または1つもしくは複数の遺伝子座）を決定することができる。上記で論じられたように、フロー空間におけるシーケンシングデータセットを塩基空間に（またはフロー順序が既知である場合には、その逆に）容易に変換することができ、マッピングをフロー空間または塩基空間において行うことができる。マッピングされた配列に対応する遺伝子座（単数）〔または遺伝子座（複数）〕を、本明細書に記載される解析方法のための候補配列（またはハプロタイプ配列）として動作することができる1つまたは複数のバリエーション配列と、関連付けることができる。本明細書に記載される方法の1つの利点は、一部のケースではアラインメントアルゴリズムを使用するシーケンシングされた核酸分子の配列と各候補配列との一般に計算コストの高いアラインメントを必要としない点である。その代わりに、フロー空間におけるシーケンシングデータを使用して候補配列の各々についてマッチスコアを決定することができ、この操作のほうが、計算効率が良い。

【0097】

マッチスコアは、シーケンシングデータセットがいかに良く候補配列を支持するかを示す。例えば、シーケンシングデータセットが候補配列にマッチする尤度を示すマッチスコアは、各フロー位置における統計パラメーター（例えば、尤度）であって、候補配列についての予想シーケンシングデータが得られたそのフロー位置における塩基カウントに対応する統計パラメーターを選択することにより、決定することができる。選択された統計パラメーターの積によりマッチスコアを得ることができる。例えば、伸長されたプライマーについて図1Aに示されているシーケンシングデータセット、およびT A T G G T C A T C G A（配列番号2）の候補プライマー伸長配列を仮定する。図1C（図1Aにおける同じシーケンシングデータセットを示す）は、候補配列（塗りつぶした丸印）についてのトレースを示す。比較として、T A T G G T C G T C G A（配列番号1）配列のトレース（図1Bを参照されたい）が、図1Cに白抜きの丸印を使用して示されている。シーケンシングデータが第1の候補配列T A T G G T C A T C G A（配列番号2）に対応する尤度を示すマッチスコアと、シーケンシングデータが第2の候補配列T A T G G T C G T C G A（配列番号1）にマッチする尤度を示すマッチスコアとは、たとえこれらの配列が単一塩基変動分しか変わらなかったとしても、大きな差がある。図1Cで見られるように、トレース間の差は、フロー位置12に見られ、少なくとも9フロー位置（およびシーケンシングデータがさらなるフロー位置にわたって伸長する場合にはより長い可能性がある）にわたって伝播する。1または複数のフローサイクルにわたって継続するこの伝播は、「フローシフト」または「サイクルシフト」と呼ばれることがあり、シーケンシングデータセットが候補配列にマッチする場合、一般に、非常に可能性の低い事象である。

【0098】

したがって、各シーケンシングデータセットと候補配列（または各候補配列）とのマッチスコアを決定することができる。例えば、シーケンシングデータセットが、所与の候補配列に一致する尤度 $L(R_j | H_i)$ は、所与の候補配列についての各フロー位置における選択された塩基カウントの尤度（例えば、その積）を使用して、決定することができる。

【0099】

マッチスコアを使用して、試験シーケンシングデータ、および／または試験シーケンシングデータに関連する核酸分子を分類することができる。分類子は、核酸分子がバリエーション（例えば、候補配列に含まれるバリエーション）を含むことを示すこともあり、核酸分子が、バリエーションを含まないことを示すこともあり、またはヌルコールを示すこともある。ヌルコールは、試験シーケンシングデータに関連する核酸分子におけるバリエーションの存在も非存在も示さず、それどころか、マッチスコアを使用して所望の統計的信頼度でコールを行なうことができないことを示す。試験シーケンシングデータまたは核酸分子は、例えば、マッチスコアが所望の信頼度閾値よりも上であった場合、バリエーションを有するも

10

20

30

40

50

のとして分類され得る。逆に、試験シーケンシングデータまたは核酸分子は、例えば、マッチスコアが所望の信頼度閾値よりも下であった場合、バリエントを有さないものとして分類され得る。

【0100】

上記の解析を応用して、2つまたはそれより多くの異なる候補配列から候補配列を選択することができる。シーケンシングデータセットが各候補配列にマッチする尤度を示すマッチスコアを決定することができる。例えば、シーケンシングデータセット内の各フロー位置における統計パラメーターであって、そのフロー位置における候補配列の塩基カウントに対応する統計パラメーターを、候補配列ごとに選択することができる。一部の実施形態では、この解析は、シーケンシングされた試験核酸分子についてのシーケンシングデータセットを生成するために使用されたのと同じフロー順序を使用して候補配列がシーケンシングされることを仮定して、候補シーケンシングについての予測シーケンシングデータを生成することを含む。このデータは、候補配列を有する核酸分子をシーケンシングすることにより、または候補配列およびフロー順序に基づいて候補シーケンシングデータセットを *in silico* で生成することにより、生成することができる。例示的な候補シーケンシングデータセットは、図1Cの試験データシーケンシングデータセットの下に示されており、第1の候補配列 [T A T G G T C A T C G A (配列番号2)] は塗りつぶした丸印トレースに対応し、第2の候補配列 [T A T G G T C G T C G A (配列番号1)] は白抜きの丸印トレースに対応する。一部の実施形態では、例えば、マッチスコアが2つまたはそれより多くの異なる候補配列について決定される場合、試験シーケンシングデータまたは核酸分子は、2つもしくはそれより多くの候補配列のうちの1つについてのバリエントを有するもの、2つしくはそれより多くの候補配列のうちの1つについてのバリエントを有さないものとして分類されることがあり、またはヌルコールが2つしくはそれより多くの候補配列間で得られることもある（例えば、候補配列のいずれについてのコールも得ることができなかった場合、もしくはマッチスコアが、同じ遺伝子座で2つしくはそれより多くの異なるバリエントを示した場合）。

【0101】

シーケンシングデータセットのマッチスコアが、候補配列について決定されると、マッチスコアに基づいて短い遺伝子バリエントを有する候補配列（例えば、2つまたはそれより多くの候補配列の中から最高尤度マッチを有するマッチスコアをもたらす候補配列）を選択することができる。短い遺伝子バリエントは、例えば、個体の亜集団内に見られるバリエントもしくは突然変異であることもあり、または単一もしくは特異的個体に固有のバリエントもしくは突然変異であることもある。短い遺伝子バリエントは、生殖細胞系列バリエントであることもあり、または体細胞バリエントであることもある。短い遺伝子バリエントを有する核酸分子の配列から生じるシーケンシングデータは、短い遺伝子バリエントを有する候補配列にマッチすることになり、その候補配列を選択することができるが、棄却された（または非選択）候補配列は、より小さい尤度マッチ（これらの候補配列について決定されたマッチスコアに基づいて）により示されるように、短い遺伝子バリエントを含まない。非選択候補配列と、選択候補配列（シーケンシングされた核酸分子シーケンシングデータセットに最も良くマッチする）とは、2カ所またはそれより多くのフロー位置において異なることがあり、これらの位置は、2カ所しくはそれより多くの連続するフロー位置であることもあり、または2カ所しくはそれより多くの非連続のフロー位置であることもある。一部の実施形態では、非選択候補配列と、選択候補配列とは、3カ所しくはそれより多くの、4カ所しくはそれより多くの、5カ所しくはそれより多くの、6カ所しくはそれより多くの、7カ所しくはそれより多くの、8カ所しくはそれより多くの、9カ所しくはそれより多くの、または10カ所しくはそれより多くのフロー位置において異なる。一部の実施形態では、非選択候補配列と、選択候補配列とは、1もしくは複数の、2もしくはそれより多くの、3もしくはそれより多くの、4もしくはそれより多くの、または5もしくはそれより多くのフローサイクルにわたって異なる。一部の実施形態では、非選択候補配列と選択候補配列とは、Xカ所の塩基位置に

10

20

30

40

50

において異なり、この場合、核酸分子の配列に関連するシーケンシングデータセットと、非選択候補配列とは、 $(X+2)$ カ所またはそれより多くのフロー位置において異なる。シーケンシングされた核酸分子シーケンシングデータセットが選択候補配列に最も良くマッチする、選択候補配列と非選択候補配列とで異なるフロー位置の数の増加は、シーケンシングされた核酸分子シーケンシングデータセットが、非選択候補配列を有する核酸分子のシーケンシングから得られた尤度を低下させる。

【0102】

シーケンシングされた核酸分子のシーケンシングデータセットが非選択候補配列にマッチする尤度は、好ましくは低く、例えば、0.05未満、0.04未満、0.03未満、0.02未満、0.01未満、0.005未満、0.001未満、0.0005未満、または0.0001未満である。シーケンシングされた核酸分子のシーケンシングデータセットが選択候補配列にマッチする尤度は、好ましくは高く、例えば、0.95より高く、0.96より高く、0.97より高く、0.98より高く、0.99より高く、0.995より高く、または0.999より高い。

【0103】

試験試料中の短い遺伝子バリエーションを検出するための方法は、一部の実施形態では、各試験シーケンシングデータセットが試験試料中の別々の試験核酸分子に関連している、複数の試験シーケンシングデータセットを解析するステップを含み得る。核酸分子は、例えば、核酸分子の配列が参照配列とアラインメントされた場合、ある遺伝子座において少なくとも部分的に重複する。核酸分子の少なくとも一部分は、異なるシーケンシング開始位置（ある遺伝子座に対して）を有することができ、その結果、配列内の所与の塩基について異なるフロー位置、および／または異なるフロー順序コンテキストが生じることになる。このようにして、同じ候補配列を使用して、上記複数のうちの試験シーケンシングデータセットを解析することができる。候補配列ごとに、複数の試験シーケンシングデータセットが候補配列にマッチする尤度を示すマッチスコアを決定することができ、最高尤度マッチを有する（したがって、短い遺伝子バリエーションを含む）候補配列を選択することができる。複数の試験シーケンシングデータセットを使用して短い遺伝子バリエーションを検出するための例示的解析が、図2A～2Dに示されている。図2Aでは、3つのシーケンシングされた試験核酸分子に対応する配列（伸長されたプライマーの配列により各々表されている、R1、R2およびR3）が、2つの候補配列（H1およびH2）に関連する重複遺伝子座において参照配列とアラインメントされている。図2B、図2Cおよび図2Dは、それぞれ、R1、R2およびR3についての例示的シーケンシングデータセットを、H1の塩基（黒塗りの丸印）またはH2の塩基（白抜きの丸印）に対応するシーケンシングデータセット内の各フロー位置における選択統計パラメータとともに示す。

【0104】

1つまたは複数の決定されたマッチスコアを使用して、試験試料についての短い遺伝子バリエーションの存在（もしくはアイデンティティ）または非存在をコールすることができる。一部の実施形態では、例えば、バリエーションを有するものとして分類される単一核酸分子（または関連試験シーケンシングデータセット）は、例えば、マッチスコアが、所望の信頼度または事前設定された信頼度で候補配列とのマッチを示す場合、バリエーションの存在、アイデンティティまたは非存在をコールするのに十分なものであり得る。一部の実施形態では、所定数の（例えば、1つまたは複数の、2つまたはそれより多くの、3つまたはそれより多くの、4つまたはそれより多くの、5つまたはそれより多くの、などの）核酸分子（または核酸分子に関連する試験シーケンシングデータセット）は、バリエーションが試験試料についてコールされる前にバリエーションを有するものとして分類される。一部の実施形態では、核酸分子（または核酸分子に関連する試験シーケンシングデータセット）の数は、マッチスコアに依存して動的に選択され、例えば、高い信頼度マッチスコアでバリエーションを有するものとして分類された単一核酸分子を使用してバリエーションをコールすることもあり、またはより低い信頼度マッチスコアでバリエーションを有するものとして分

類された2つまたはそれより多くの核酸分子を使用してバリエーションをコールすることもある。

【0105】

必要に応じて、シーケンシングデータセットについての別々のマッチスコアは、複数の試験シーケンシングデータセットのマッチスコアを決定するためにまとめて解析される。例えば、候補配列ごとに各試験シーケンシングデータセットについてのマッチスコアが、本明細書に記載される方法を使用して決定されると、公知のベイズ法を使用して、例えば、ゲノム解析ツールキット (GATK) に含まれている Haplotype Caller アルゴリズムを使用して、複数の試験シーケンシングデータセットが候補配列にマッチする尤度を示すマッチスコアを決定することができ、最高尤度マッチを有する候補配列を選択することができる。例えば、DePristo et al., A framework for variation discovery and genotyping using next-generation DNA sequencing data, Nature Genetics 43, 491-498 (2011); および Poplin et al., Scaling accurate genetic variant discovery to tens of thousands of samples, bioRxiv, www.biorxiv.org/content/10.1101/201178v3 (July 24, 2018); Hwang et al., Systematic comparison of variant calling pipelines using gold standard personal exome variants, Scientific Reports, vol. 5, no. 17875 (2015) を参照されたく、これらの各々の内容は、本明細書に取り込まれる。

標的バリエーションおよび／またはフローサイクル順序の選択

【0106】

短い標的遺伝子バリエーションを、例えば、フロー順序および／もしくは候補配列を選択するための基礎として (すなわち、短い標的遺伝子バリエーションを事前選択することにより)、または下流の解析のための基礎として役立つように、選択することができる。下流の解析は、例えば、同定された短い遺伝子バリエーションを含むバイオマーカーパネルをアセンブルすることを含み得る。バイオマーカーパネルを試験試料に関連する個々の対象用に個別化することができる。例として、バイオマーカーパネルは、疾患 (例えば、がん) に関連する1つまたは複数の短い遺伝子バリエーション、例えば、バリエーションシグネチャーを含み得る。別の例では、バイオマーカーパネルは、対象用に個別化され、対象からの試料において以前に検出された1つまたは複数の短い遺伝子バリエーションを含み、このバリエーションは、対象における疾患 (例えば、がん) に起因すると考えられ得る。

【0107】

本明細書に記載の短い遺伝子バリエーションを同定するための方法は、1つまたは複数の短い標的遺伝子バリエーションが事前選択される場合、特に有用であり得る。所与の短い遺伝子バリエーションの検出限界 (LOD) は、核酸分子をシーケンシングするために、および核酸分子のシーケンシングデータセットを生成するために使用される、短い遺伝子バリエーションの配列コンテキスト (例えば、短い標的遺伝子バリエーション遺伝子座に隣接する核酸分子の配列)、およびフロー順序 (またはフローサイクル順序) に依存し得る。つまり、所与のフロー順序、短い遺伝子バリエーション、および短い遺伝子バリエーションコンテキストを使用して、短い遺伝子バリエーションを有する核酸分子および短い遺伝子バリエーションを有さない核酸分子 (例えば、参照配列) の間の、フロー空間におけるフロー位置のバリエーションの数を決定することができる。これにより、特に感度の良いバリエーションの選択、または高感度で特定のバリエーションを検出することができるフロー順序の選択が、可能になる。短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットを、短い標的遺伝子バリエーションを有さない参照配列に関連する参照シーケンシングデータセットと比較して、標的配列と参照配列間に存在するフロー位置の相違の数を決定することがで

きる。つまり、参照配列は、短い標的遺伝子バリエーションを除いて標的配列と同一である。より多いフロー位置の相違数は、そのバリエーションに対してより高い感度（すなわち、より低い検出限界）を示す。標的および参照シーケンシングデータセットは、標的配列を有する核酸分子および／もしくは参照配列を有する核酸分子を実際にシーケンシングすることにより決定されることがあり、またはデータセットは、予想シーケンシングデータセット（例えば、*in silico*で決定されるような）であることがある。

【0108】

一例では、特定の対象またはがんの遺伝子指紋が所望されることがあるが、対象のまたはがんのゲノムにおける各々のおよびあらゆる短い遺伝子バリエーションを検出する必要はない。その代わりに、所与のフロー順序に対して特に高感度である1つまたは複数の短い遺伝子バリエーションを事前選択することができる。感度の良いバリエーションを事前選択することにより、試験試料に対してより低いシーケンシング深度を使用してバリエーションを高い信頼性でコールすることができる。

【0109】

一部の実施形態では、試験試料中の短い標的遺伝子バリエーションを検出するための方法は、短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、フロー位置がヌクレオチドフローに対応する、ステップを含み得る。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータとは、2カ所またはそれより多くの非連続のフロー位置において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータとは、3カ所またはそれより多くの連続するフロー位置において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータとは、Xカ所の塩基位置において異なり、標的シーケンシングデータセットと参照シーケンシングデータとは、 $(X+2)$ カ所またはそれより多くの連続するフロー位置において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータとは、1または複数のフローサイクルにわたって異なる。

【0110】

一部の実施形態では、試験試料中の短い標的遺伝子バリエーションを検出するための方法は、短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列および参照配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、フロー位置がヌクレオチドフローに対応する、ステップを含み得る。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータとは、2カ所またはそれより多くの非連続のフロー位置において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータとは、2カ所またはそれより多くの連続するフロー位置において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータとは、3カ所またはそれより多くのフロー位置において異なり、これらのフロー位置は、連続していることもあり、または非連続であることもある。一部の実施形態では、標的配列と参照配列とは、Xカ所の塩基位置において異なり、標的シーケンシングデータセットと参照シーケンシングデータとは、 $(X+2)$ カ所またはそれより多くの連続するフロー位置において異なる。一部の実施形態では

、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0111】

選択された、標的となる短い遺伝子バリエーションの検出は、一般に上記で論じられたように進行し得る。例えば、一部の実施形態では、短い標的遺伝子バリエーションの遺伝子座を有する試験核酸分子に関連する試験シーケンシングデータセットを得ることができる。シーケンシングデータは、標的および参照シーケンシングデータセットを生成するために使用されたのと同じフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより生成される。試験シーケンシングデータセットが短い遺伝子バリエーションを有する標的配列にマ

10

【0112】

一部の実施形態では、各試験シーケンシングデータセットが試験試料中の異なる試験核酸分子に関連している、複数の試験シーケンシングデータセットを使用して、試験試料中の短い標的遺伝子バリエーションが検出される。解析される試験核酸分子は、短い標的遺伝子バリエーション遺伝子座において重複しており、データセットは、短い標的遺伝子バリエーションを選択するために使用されたのと同じフローサイクル順序を使用して標的核酸分子をシーケンシングすることにより生成される。複数の試験シーケンシングデータセットが短い遺伝子バリエーションを有する標的配列にマッチする尤度を示すマッチスコア（または、代替的にもしくは加えて、複数の試験シーケンシングデータセットが参照配列にマ

20

【0113】

一部の実施形態では、シーケンシングデータを生成するために使用されるフロー順序またはフローサイクル順序は、事前選択される。本明細書で論じられるように、フロー順序におけるバリエーションのコンテキストは、バリエーション配列と比較（例えば、参照）配列とのシグナルの相違に影響を与え得る。選択された標的バリエーションを検出する尤度を増加させるように、フロー順序またはフローサイクル順序を事前選択することができる。

30

【0114】

図3は、試験試料中の短い遺伝子バリエーションを検出するための例示的方法のフローチャートを示す。ステップ302で、短い標的遺伝子バリエーションが選択される。短い標的遺伝子バリエーションは、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なるように選択され、フロー位置は、ヌクレオチドフローに対応する。ステップ304で、1つまたは複数の試験シーケンシングデータセットは、例えば、1つもしくは複数の試験核酸分子をシーケンシングして1つもしくは複数の試験シーケンシングデータセットを得ることにより、または1つもしくは複数の試験シーケンシングデータセットを受信することにより、得られる。試験シーケンシングデータセットの各々は、試験試料に由来する試験核酸分子に関連している。選択された短い標的遺伝子バリエーションの解析のために、試験核酸分子は、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なる。試験シーケンシングデータセットが複数のフロー位置におけるフローシグナルを含む場合、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより、シーケンシングデータセットを決定することができる（または前以て決定しておいてもよ

40

50

い)。ステップ306で、試験シーケンシングデータセットに関連する試験核酸分子ごとに、マッチスコアが決定される。マッチスコアは、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示す。あるいは、マッチスコアは、核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すこともある。ステップ308で、1つまたは複数の決定されたマッチスコアは、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするために使用される。

【0115】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0116】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングして、複数のフロー位置におけるフローシグナルを含む1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来する、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチス

コア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

10

【0117】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）短い標的遺伝子バリエーションを事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；（c）試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

20

30

【0118】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）短い標的遺伝子バリエーションを事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングして、複数のフロー位置におけるフローシグナルを含む1

40

50

つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来する、ステップ；（c）試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0119】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）短い標的遺伝子バリエーションおよびフローサイクル順序を事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；（c）試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0120】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するための方法は、（a）短い標的遺伝子バリエーションおよびフローサイクル順序を事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標

的シーケンシングデータセットおよび参照シーケンシングデータセットが事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングして、複数のフロー位置におけるフローシグナルを含む1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来する、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

標的バリエーションおよび／またはフローサイクル順序の選択

【0121】

フローサイクル順序は、必ずしも4塩基フローサイクル（例えば、A、G、CおよびTを1つずつ、任意の反復順序で）に限定される必要はなく、サイクル内の塩基タイプが4タイプより多い延長フローサイクルであってもよい。延長サイクル順序を所望のサイクル数、反復して、シーケンシングプライマーを伸長することができる。例として、一部の実施形態では、延長フロー順序は、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20またはそれより多くの別々のヌクレオチドフローをフローサイクル順序で含む。サイクルは、A、G、CおよびTを少なくとも1つずつ含むことができるが、サイクルが反復される前にサイクル内の1または複数の塩基タイプを反復する。

【0122】

延長フローサイクル順序は、4塩基が反復されるフローサイクル順序よりも小さいゲノムバリエーション（例えばSNP）のより大きな割合を検出するのに有用であり得る。例えば、形式XYZ→XQZ [ここで、Q≠Y（およびQ、X、YおよびZは、各々、A、C、GおよびTのいずれかを1つである）] の置換SNPについて192の有効な配置がある。これらの中で、168は、シーケンシングデータセット（例えば、フローグラム）において新しいシグナル（すなわち、新しい非ゼロシグナルまたは新しいゼロシグナル）を生じさせることができる。感度の良いフロー順序と組み合わせられた新しいゼロまたは非ゼロシグナルは、複数のフロー位置にわたって伝播されるシグナル（例えば、サイクルの長さよりも長く延長することができる、フローシフトまたはサイクルシフト）を生じさせることができ、このことにより、バリエーションにおいて参照と比較して同一の末尾の配列が得られる。ホモポリマー長変化ではなく、ホモポリマーの挿入または欠失が、シグナルの相違の伝播を生じさせる結果となり得ることは注目される。残りの24のバリエーションは、影響を受けたフロー位置においてホモポリマー長変化を引き起こすが、そのような変化は、シグナル変化の伝播を引き起こさない。したがって、SNPの理論上最大87.5%は、2カ所より多くのフロー位置について参照（または候補）配列とは異なる新しいシグナ

ルを生じさせる結果となり得る。上記で論じられたように、シグナルの相違の伝播は、試験シーケンシングデータセットと誤ってマッチした候補配列との尤度差を増加させる。さらに、シグナル変化の伝播は、バリエーションに及ぶフロー順序に依存する。

【0123】

ランダムに断片化された試験試料中の核酸分子のシーケンシングは、シーケンシングプライマーがフロー順序を使用して伸長された場合、バリエーションのフロー順序コンテキストのランダムシフトを生じさせる結果となる。つまり、バリエーションのフロー位置は、シーケンシングされる核酸分子の開始位置によって変わり得る。たとえ核酸分子配列中のすべてのシーケンシング開始位置が用いられたとしても、SNPの87.5%すべてについて、すべてのフローサイクルの組合せが2カ所より多くのフロー位置においてシグナル変化を検出できるとは限らない。例えば、4塩基フローサイクル順序T-A-C-Gは、SNPの41.7%について2カ所より多くのフロー位置において参照シーケンシングデータセットと異なる試験シーケンシングデータセットを生じさせる結果となり得る。本明細書中でさらに論じられるように、延長フローサイクル順序は、SNPの理論上最大量（すなわち、可能なSNPの87.5%、またはホモポリマー長の変化を生じさせる結果となるもの以外のすべてのSNP）のすべてが、2カ所より多くの位置において試験シーケンシングデータセットと参照シーケンシングデータセットとの差を生じさせることができるように設計されており、このことにより十分な高さのシーケンシング深度（すなわち、十分多い数の出発位置のサンプリング）が得られる。

【0124】

延長シーケンシングフロー順序は、異なる効率（すなわち、ヒト参照ゲノム配列に使用された場合のフローごとの平均取り込み数）を有し得る。一部の実施形態では、フロー順序は、約0.6のまたはそれを超える（例えば、約0.62もしくはそれを超える、約0.64もしくはそれを超える、約0.65もしくはそれを超える、約0.66もしくはそれを超える、または約0.67もしくはそれを超える）効率を有する。一部の実施形態では、フロー順序は、約0.6～約0.7の効率を有する。フローサイクル順序および対応する推定効率の例は、表2に示されている。

【0125】

一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも5%についてのSNPパーミュテーションの約50%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置（すなわち、「フロー相」）の少なくとも5%についてのSNPパーミュテーションの約60%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも5%についてのSNPパーミュテーションの約70%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所またはそれより多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも5%についてのSNPパーミュテーションの約80%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。

【0126】

一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも10%についてのSNPパーミュテーションの約50%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも10%についてのSNPパーミュテーションの約60%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも10%についてのSNPパーミュテーションの約70%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所またはそれより多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも10%についてのSNPパーミュテーションの約80%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。

10

20

【0127】

一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも20%についてのSNPパーミュテーションの約50%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも20%についてのSNPパーミュテーションの約60%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも20%についてのSNPパーミュテーションの約70%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所またはそれより多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも20%についてのSNPパーミュテーションの約80%～87.5%がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。

30

40

【0128】

一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも30%についてのSNPパーミュテーションの約50%～87.5%（または約50%～約80%）がSNPによって異なる、核酸分子に関連する2つのシ

50

ークエンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも30%についてのSNPパーミュテーションの約60%~87.5%（または約60%~約80%）がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所より多くのフロー位置において生じさせるように選択される。一部の実施形態では、延長シーケンシングフロー順序は、ランダムシーケンシング開始位置の少なくとも30%についてのSNPパーミュテーションの約70%~87.5%（または約70%~約80%）がSNPによって異なる、核酸分子に関連する2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を、2カ所またはそれより多くのフロー位置において生じさせるように選択される。

【0129】

一部の実施形態では、延長シーケンシングフロー順序は、表2における延長シーケンシングフロー順序のうちのいずれか1つである。「シフト感度」は、すべての可能なSNPパーミュテーションにわたって2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間のシグナルの相違を2カ所より多くのフロー位置において生じさせる最大感度を指す。「最大シフト感度」は、その感度が維持されるフロー相の最高分率ですべての可能なSNPパーミュテーションにわたって2つのシーケンシングデータセット（例えば、試験または標的シーケンシングデータセットと候補または参照シーケンシングデータセット）間でシグナルの相違を2カ所より多くのフロー位置において生じさせる最大感度を指す。

10

20

30

40

50

【表 2 - 1】

表2

フローサイクル順序	推定効率	最大 シフト感度	フロー相 の5%での シフト感度	フロー相 の10%での シフト感度	フロー相 の20%での シフト感度	フロー相 の30%での シフト感度
T-C-A-G-A-T-G-C-A-T-G-C-T-A-C-G	67.5%	82.3% @ 19%	82.3%	82.3%	75.0%	66.7%
T-C-A-C-G-A-T-G-C-A-T-G-C-T-A-G	67.5%	83.3% @ 12%	83.3%	83.3%	72.9%	62.5%
T-C-A-T-G-C-A-T-G-C-T-A-C-G-A-G	67.3%	82.3% @ 12%	82.3%	82.3%	72.9%	67.7%
T-C-A-G-T-A-C-G-A-T-G-C-A-T-G-C	67.3%	82.3% @ 12%	82.3%	82.3%	75.0%	63.5%
T-C-A-G-T-C-G-A-T-G-A-C-T-A-G-C	67.2%	81.3% @ 12%	81.3%	81.3%	74.0%	69.8%
T-C-A-T-C-G-A-C-T-G-A-G-C-T-A-G	67.2%	81.3% @ 12%	81.3%	81.3%	74.0%	69.8%
T-C-G-T-A-G-C-T-G-A-C-A-T-G-C-A	67.2%	83.3% @ 12%	83.3%	83.3%	75.0%	67.7%
T-C-G-T-A-G-C-A-T-G-C-T-A-C-G-A	67.0%	79.2% @ 25%	79.2%	79.2%	79.2%	75.0%
T-C-A-T-G-C-A-G-T-C-G-A-C-T-A-G	66.9%	83.3% @ 19%	83.3%	83.3%	75.0%	68.8%
T-C-A-T-G-C-A-T-C-G-T-A-C-G-A-G-C-T-G-C-A-T- G-A-C-T-A-G	66.7%	86.5% @ 7%	86.5%	85.4%	85.4%	69.8%
T-C-G-A-C-T-G-T-A-G-C-T-A-G-C-A	66.7%	82.3% @ 19%	82.3%	82.3%	75.0%	66.7%
T-C-A-C-G-A-T-G-C-T-A-G-C-T-A-G	66.5%	82.3% @ 12%	82.3%	82.3%	75.0%	67.7%
T-C-A-G-T-A-C-G-A-T-G-C-T-A-C-G	66.4%	83.3% @ 19%	83.3%	83.3%	75.0%	68.8%

10

20

30

40

50

【表 2 - 2】

フローサイクル順序	推定効率	最大 シフト感度	フロー相 の5%での シフト感度	フロー相 の10%での シフト感度	フロー相 の20%での シフト感度	フロー相 の30%での シフト感度
T-C-G-A-C-T-A-G-C-A-T-G-C-A-T-G	66.0%	81.3% @ 12%	81.3%	81.3%	70.8%	62.5%
T-A-C-G	66.0%	41.7% @ 100%	41.7%	41.7%	41.7%	41.7%
T-C-A-G-C-T-G-A-C-T-A-G-T-C-A-T-G-A-C-T-A-G-C-G-A-T-C-G	65.7%	87.5% @ 11%	87.5%	87.5%	82.3%	75.0%
T-C-T-A-G-C-A-T-G-A-C-T-G-A-C-G	65.7%	83.3% @ 12%	83.3%	83.3%	71.9%	63.5%
T-C-G-A-C-T-A-T-G-C-A-T-G-C-A-G	65.5%	81.3% @ 19%	81.3%	81.3%	71.9%	63.5%
T-C-G-A-C-T-G-C-A-T-C-G-A-T-G-C-A-T-G-A-C-T-A-G	65.4%	87.5% @ 12%	87.5%	87.5%	82.3%	74.0%
T-C-A-C-T-G-A-C-G-T-A-G-C-T-A-T-G-C-A-T-C-G-A-G	65.3%	84.4% @ 17%	84.4%	84.4%	83.3%	76.0%
T-C-A-T-G-C-T-A-G-C-T-A-G-T-A-C-G-A-C-T-G-A-G-C-A-T-C-G	65.2%	86.5% @ 11%	86.5%	86.5%	82.3%	78.1%
T-C-G-A-T-G-C-A-T-C-G-T-A-C-T-A-G-C-A-G-T-G-A-C	65.2%	87.5% @ 8%	87.5%	86.5%	84.4%	71.9%
T-C-A-T-G-A-G-C-T-A-G-C-A-T-C-G-T-A-C-T-G-A-C-G	65.2%	87.5% @ 8%	87.5%	86.5%	81.3%	70.8%
T-C-A-G-C-A-T-G-T-A-C-T-G-A-T-G-C-A-T-C-G-A-G-C-T-A-C-G	65.0%	87.5% @ 11%	87.5%	87.5%	82.3%	77.1%
T-C-A-G-T-A-C-T-A-G-C-A-T-G-C-G-A-T-C-G-T-A-G-C-T-G-A-C	65.0%	86.5% @ 11%	86.5%	86.5%	78.1%	74.0%
T-C-A-C-G-T-A-G-C-T-A-T-G-C-T-G-A-C-T-G-A-C-A-T-G-A-C-G	64.6%	85.4% @ 9%	85.4%	84.4%	76.0%	61.5%

10

20

30

40

50

【表 2 - 3】

フローサイクル順序	推定効率	最大 シフト感度	フロー相 の5%での シフト感度	フロー相 の10%での シフト感度	フロー相 の20%での シフト感度	フロー相 の30%での シフト感度
T-C-A-G-C-T-A-T-G-A-C-T-G-A-C-G-C-A-T-C-G-T-A-C-G	64.5%	85.4% @ 12%	85.4%	85.4%	77.1%	74.0%
T-C-A-G-C-T-A-C-T-G-C-A-T-G-A-C-G-T-A-C-G-T-A-G-T-C-G-A	64.5%	87.5% @ 14%	87.5%	87.5%	83.3%	70.8%
T-C-A-G-A-C-T-A-G-C-G-A-T-G-C-A-T-G-T-C-T-A-G-T-C-A-C-G	64.5%	86.5% @ 11%	86.5%	86.5%	83.3%	62.5%
T-C-A-T-C-G-A-C-T-G-C-G-A-T-G-C-T-A-G-T-A-C-A-G	64.4%	85.4% @ 17%	85.4%	85.4%	83.3%	72.9%
T-C-A-C-G-T-A-C-T-G-A-C-A-T-G-C-A-T-G-C-T-A-G-T-A-G-C-G-A-T-C-G	64.4%	85.4% @ 9%	85.4%	84.4%	83.3%	72.9%
T-C-A-G-T-G-C-T-A-C-G-T-C-A-C-G-A-T-C-A-G-A-T-G-C-T-A-G	64.4%	86.5% @ 11%	86.5%	86.5%	71.9%	67.7%
T-C-A-G-C-G-A-T-G-A-C-T-A-G-C-T-A-C-G-T-C-A-T-G	64.4%	85.4% @ 17%	85.4%	85.4%	84.4%	66.7%
T-C-A-T-G-C-T-A-C-G-A-G	64.4%	81.3% @ 17%	81.3%	81.3%	80.2%	66.7%
T-C-A-T-G-A-C-G-T-A-C-G-A-C-T-C-A-T-G-C-A-G-T-G-C-T-A-G	64.3%	85.4% @ 11%	85.4%	85.4%	82.3%	75.0%
T-C-A-G-T-C-G-A-T-G-C-T-A-C-T-G-C-A-T-A-C-G-T-C-G-A-T-G-A-T-G-A-C-A-G	64.3%	87.5% @ 9%	87.5%	86.5%	83.3%	74.0%
T-C-G-A-T-G-A-C-A-G	64.3%	81.3% @ 17%	81.3%	81.3%	80.2%	66.7%
T-C-A-G-T-C-G-A-C-A-T-G-C-A-T-C-G-A-T-A-C-G-T-G-C-T-A-G	64.2%	87.5% @ 9%	87.5%	86.5%	79.2%	70.8%

【0130】

一部の実施形態では、核酸分子をシークエンシングする方法は、（a）核酸分子をプライマーとハイブリダイズして、ハイブリダイズされた鋳型を形成するステップ；（b）5つまたはそれより多くの別々のヌクレオチドフローを含む反復フローサイクル順序に従って別々のヌクレオチドフローで提供される標識された非終結ヌクレオチドを使用してプライマーを伸長するステップ；および（c）取り込まれた標識ヌクレオチドからのシグナルを検出する、またはプライマーがヌクレオチドフローにより伸長された場合のシグナルの非存在を検出するステップを含む。一部の実施形態では、フローサイクル順序は、2カ所より多くのフロー位置において、ランダムシークエンシング開始位置の5%における可能

10

20

30

40

50

なSNPパーミュテーションの50%またはそれより多くについてのシグナル変化を誘導する。一部の実施形態では、誘導されるシグナル変化は、シグナル強度の変化、または新しい実質的にゼロ（もしくは新しいゼロ）もしくは新しい実質的に非ゼロ（もしくは新しい非ゼロ）シグナルである。一部の実施形態では、誘導されるシグナル変化は、新しい実質的にゼロ（もしくは新しいゼロ）または新しい実質的に非ゼロ（もしくは新しい非ゼロ）シグナルである。一部の実施形態では、前記フローサイクル順序は、1フロー当たり0.6のまたはそれを超える塩基取り込みの効率を有する。一部の実施形態では、フローサイクルは、表2に収載されているフローサイクル順序のうちのいずれか1つである。

異なるフロー順序での再シーケンシング

【0131】

検出される短い遺伝子バリエーションの感度は、核酸分子のシーケンシングに使用されるフローサイクル順序に依存するので、本明細書に記載される方法を、2つまたはそれより多くの異なるフローサイクル順序を使用してシーケンシングされた試験核酸分子（または遺伝子座が重複している複数の核酸分子）の解析に適応させることができる。マッチスコアは、2つまたはそれより多くの異なるシーケンシングデータセット（異なるフローサイクル順序の結果として生じる）と1つまたは複数の候補配列とのマッチに基づいて決定することができる。上記で論じられたようなマッチスコアに基づいて、バリエーションの存在または非存在をコールすることおよび／または選択候補配列をコールすることができる。

【0132】

方法は、第1のフローサイクル順序を使用してシーケンシングされた試験試料に由来する試験核酸分子に関連する第1の試験シーケンシングデータセット、および第2のフローサイクル順序を使用してシーケンシングされた同じ試験核酸分子に関連する第2の試験シーケンシングデータセットを得るステップを含み得る。例えば、試験核酸分子を、第1のフローサイクル順序に従って別々のヌクレオチドフローで非終結核酸分子を提供すること、シーケンシングプライマーを伸長すること、および各ヌクレオチドフローの後にシーケンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して第1の試験シーケンシングデータセットを生成することにより、シーケンシングし；伸長されたシーケンシングプライマーを除去し；同じ試験核酸分子を、第2のフローサイクル順序に従って別々のヌクレオチドフローで非終結ヌクレオチドを提供すること、シーケンシングプライマーを伸長すること、および各ヌクレオチドフローの後にシーケンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して第2の試験シーケンシングデータセットを生成することにより、シーケンシングすることができる。

【0133】

核酸分子は、異なるフローサイクル順序を使用してシーケンシングされるため、シーケンシングデータセットは異なる。図4Aおよび図4Bは、第1のフローサイクル順序（T-A-C-G）（図4A）および第2のフローサイクル順序（A-G-C-T）（図4B）を使用して決定されたTATGGTCGTCGA（配列番号1）の伸長されたプライマー配列を有する核酸分子についての例示的シーケンシングデータセットを示す。見られる通り、図4Aおよび図4Bのシーケンシングデータセットは、たとえ核酸分子配列が変化しなかったとしても、フローサイクル順序の差に起因して異なる。シーケンシングデータセット内で、第1の候補の伸長されたプライマー配列TATGGTCGTCGA（配列番号1）（黒塗りの丸印）および第2の候補の伸長されたプライマー配列TATGGTCATCGA（配列番号2）（白抜きの丸印）の塩基カウントに対応する各フロー位置における統計パラメータを選択することができる。図4Aおよび図4Bは、フローサイクル順序がバリエーション検出感度の有意な変化を来すことを実証する。例えば、第1のフローサイクル順序を使用する第1の候補配列と第2の候補配列間の差がフロー位置12～20で明らかである（図4A）一方で、第1のフローサイクル順序を使用する第1の候補配列と第2の候補配列間の差が位置17および18でのみ明らかである（図4B）。

【0134】

第1のシーケンシングデータセットおよび第2のシーケンシングデータセットが1つまたは複数の候補配列〔例えば、事前選択された短い標的遺伝子バリエーションを有する標的配列、事前選択された短い標的遺伝子バリエーションのない配列を有する参照配列、または他の可能性のある候補配列（例えば、ハプロタイプ）〕にマッチする尤度を示すマッチスコアを決定することができ、短い標的遺伝子バリエーションの存在または非存在をコールすることまたは選択候補配列をコールすることができる。

【0135】

本明細書中で論じられるように、このプロセスは、共通の遺伝子座で重複している複数の異なる試験核酸分子をシーケンシングする場合に使用することができる。例えば、各々が、第1のフローサイクル順序を使用してシーケンシングされた試験核酸分子に関連する試験シーケンシングデータセットである、複数の第1の試験シーケンシングデータセットを得ることができ、各々が、第2のフローサイクル順序を使用してシーケンシングされた同じ核酸分子に関連する試験シーケンシングデータセットである、複数の第2の試験シーケンシングデータセットを得ることができる。第1のフローサイクル順序と第2のフローサイクル順序は、異なる。複数の第1のシーケンシングデータセットおよび複数の第2のシーケンシングデータセットが1つまたは複数の候補配列〔例えば、事前選択された短い標的遺伝子バリエーションを有する標的配列、事前選択された短い標的遺伝子バリエーションのない配列を有する参照配列、または他の可能性のある候補配列（例えば、ハプロタイプ）〕にマッチする尤度を示すマッチスコアを決定することができ、短い標的遺伝子バリエーションの存在または非存在をコールすることまたは選択候補配列をコールすることができる。

【0136】

図5は、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための例示的方法を示す。ステップ502で、1つまたは複数の第1の試験シーケンシングデータセットが得られる。1つまたは複数の第1の試験シーケンシングデータセットは、例えば、1つもしくは複数の第1のシーケンシングデータセットを受信することにより、または1つもしくは複数の核酸分子をシーケンシングすることにより、得ることができる。第1の試験シーケンシングデータセットの各々は、試験試料に由来する異なる核酸分子に関連している。第1のシーケンシングデータセットは、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定される。得られた1つまたは複数の第1の試験シーケンシングデータセットの各々は、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む。ステップ504で、1つまたは複数の第2の試験シーケンシングデータセットが得られる。1つまたは複数の第2の試験シーケンシングデータセットは、例えば、1つもしくは複数の第2の試験シーケンシングデータセットを受信することにより、または1つもしくは複数の核酸分子をシーケンシングすることにより、得ることができる。第2の試験シーケンシングデータセットの各々は、第1の試験シーケンシングデータセットと同じ核酸分子に関連している。つまり、核酸分子は、第1のシーケンシングデータセットと第2のシーケンシングデータセットの両方に関連している。第2のシーケンシングデータセットは、第1のフローサイクル順序とは異なる第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定される。得られた1つまたは複数の第2の試験シーケンシングデータセットの各々は、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む。ステップ506で、第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、マッチスコアが決定される。マッチスコアは、第1の試験シーケンシングデータセット、試験シーケンシングデータセット、または両方が、1つまたは複数の候補配列からの候補配列にマッチすることを示す。ステップ508で、決定されたマッチスコアを使用して試験試料中の短い遺伝子バリエーションの存在または非存在がコールされる。

【0137】

図6は、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための別の例示的方法を示す。ステップ602で、短い標的遺伝子バリエーションが選択される。短い標的遺伝子バリエーションは、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータと、参照配列に関連するシーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序または両方に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なるように選択され、第1のフローサイクル順序と第2のフローサイクル順序は異なり、フロー位置はヌクレオチドフローに対応する。ステップ604で、1つまたは複数の第1の試験シーケンシングデータセットが得られる。1つまたは複数の第1の試験シーケンシングデータセットは、例えば、1つもしくは複数の第1のシーケンシングデータセットを受信することにより、または1つもしくは複数の核酸分子をシーケンシングすることにより、得ることができる。第1の試験シーケンシングデータセットの各々は、試験試料に由来する異なる核酸分子に関連している。第1のシーケンシングデータセットは、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定される。得られた1つまたは複数の第1の試験シーケンシングデータセットの各々は、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む。ステップ606で、1つまたは複数の第2の試験シーケンシングデータセットが得られる。1つまたは複数の第2の試験シーケンシングデータセットは、例えば、1つもしくは複数の第2の試験シーケンシングデータセットを受信することにより、または1つもしくは複数の核酸分子をシーケンシングすることにより、得ることができる。第2の試験シーケンシングデータセットの各々は、第1の試験シーケンシングデータセットと同じ核酸分子に関連している。つまり、核酸分子は、第1のシーケンシングデータセットと第2のシーケンシングデータセットの両方に関連している。第2のシーケンシングデータセットは、第1のフローサイクル順序とは異なる第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定される。得られた1つまたは複数の第2の試験シーケンシングデータセットの各々は、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む。ステップ608で、第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、マッチスコアが決定される。マッチスコアは、第1の試験シーケンシングデータセット、試験シーケンシングデータセット、または両方が、1つまたは複数の候補配列（これは参照配列を含み得る）からの候補配列にマッチすることを示す。ステップ610で、決定されたマッチスコアを使用して試験試料中の短い遺伝子バリエーションの存在または非存在がコールされる。

【0138】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、(a) 1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(b) 1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数

の試験核酸分子をシーケンシングすることにより決定され、第1のフローサイクル順序と第2のフローサイクル順序が異なり、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（c）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットの各々について、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列のマッチする尤度を示す、ステップ；および（d）決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0139】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、（a）第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各々の第1の試験シーケンシングデータセットが、異なる試験核酸分子に関連している、ステップ；（b）第1のフローサイクル順序と異なる第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各々の第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；（c）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および（d）決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0140】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、（a）1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（b）1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、第1のフローサイクル順序および第2のフローサイクル順序は異なり、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（c）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；（d）2つまたはそれより多くの異なる候補配列から候補配列を選択す

るステップであって、選択候補配列が、第1の試験シークエンシングデータセット、第2の試験シークエンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および（e）選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそれより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

【0141】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、（a）第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシークエンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シークエンシングデータセットを得るステップであって、各々の第1の試験シークエンシングデータセットが、異なる試験核酸分子に関連している、ステップ；（b）第1のフローサイクル順序と異なる第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシークエンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シークエンシングデータセットを得るステップであって、各々の第2の試験シークエンシングデータセットが、第1の試験シークエンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；（c）第1のシークエンシングデータセットおよび第2のシークエンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シークエンシングデータセット、第2の試験シークエンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；（d）2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列が、第1の試験シークエンシングデータセット、第2の試験シークエンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および（e）選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそれより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

【0142】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、（a）短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シークエンシングデータセットと、参照配列に関連する参照シークエンシングデータセットとが、標的シークエンシングデータセットおよび参照シークエンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシークエンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；（b）1つまたは複数の第1の試験シークエンシングデータセットを得るステップであって、各第1の試験シークエンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シークエンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数

10

20

30

40

50

の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(c) 1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(d) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および(e) 決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0143】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；(b) 第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各々の第1の試験シーケンシングデータセットが、異なる試験核酸分子に関連している、ステップ；(c) 第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各々の第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；(d) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および(e) 決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。

【0144】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所または

それより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；(b) 1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(c) 1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(d) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列（これは参照配列を含み得る）についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；(e) 2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列が、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および(f) 選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそれより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

【0145】

一部の実施形態では、試験試料中の短い遺伝子バリエーションの存在または非存在を検出するための方法は、(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；(b) 第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各々の第1の試験シーケンシングデータセットが、異なる試験核酸分子に関連している、ステップ；(c) 第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各々の第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；(d) 第1のシーケン

10

20

30

40

50

シングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；（e）2つまたはそれより多くの異なる候補配列（これらは参照配列を含み得る）から候補配列を選択するステップであって、選択候補配列が、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および（f）選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそれより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

システム、デバイスおよびレポート

【0146】

図に関連して説明されたものを含む、上記で説明された操作は、図7に描かれている1つまたは複数の構成要素により、必要に応じて実行される。どのようにすれば他のプロセス、例えば、上記で説明された操作のすべてまたは一部の組合せまたは部分的組合せを図7に描かれている構成要素に基づいて実行することができるのかは、当業者には明らかであろう。どのようにすれば本明細書に記載される方法、技法、システムおよびデバイスを互いに、全体として、または部分的に組み合わせることができるのかもまた、それらの方法、技法、システムおよび／またはデバイスが、図7に描かれている構成要素により実行されるか否か、および／または提供されるか否かを問わず、当業者には明らかであろう。

【0147】

図7は、一実施形態に従ってコンピュータデバイスの例を説明する。デバイス700は、ネットワークに接続されたホストコンピュータであることがある。デバイス700は、クライアントコンピュータまたはサーバーであることもある。図7に示されているように、デバイス700は、任意の好適なタイプのマイクロプロセッサベースのデバイス、例えば、パーソナルコンピュータ、ワークステーション、サーバー、またはハンドヘルドコンピュータデバイス（携帯用電子デバイス）、例えば電話機もしくはタブレットであり得る。デバイスは、例えば、プロセッサ710、入力デバイス720、出力デバイス730、記憶装置740、および通信デバイス760のうちの1つまたは複数を含み得る。入力デバイス720および出力デバイス730は、一般に、上記のものに対応することができ、コンピュータと接続可能または一体型のどちらかであり得る。

【0148】

入力デバイス720は、入力を行なう任意の好適なデバイス、例えば、タッチスクリーン、キーボードもしくはキーパッド、マウス、または音声認識デバイスであり得る。出力デバイス730は、出力を行なう任意の好適なデバイス、例えば、タッチパネル、触覚デバイス、またはスピーカーであり得る。

【0149】

記憶装置740は、RAM、キャッシュメモリー、ハードドライブまたは脱着式保存ディスクを含む、電子、磁気または光メモリーなどの、記憶域を提供する任意の好適なデバイスであり得る。通信デバイス760は、ネットワークを用いてシグナルを送信および受信することができる任意の好適なデバイス、例えば、ネットワークインターフェースチップまたはデバイスを含み得る。コンピュータの構成要素を、任意の好適な方法で、例えば物理的バスを介してまたは無線で、接続することができる。

【0150】

記憶装置740に記憶され、プロセッサ710により実行され得る、ソフトウェア750は、例えば、本開示の機能性を具現化する（例えば、上記のデバイスで具現化される

10

20

30

40

50

ような) プログラミングを含むことができる。

【0151】

上記のものなどの命令実行システム、装置もしくはデバイスで使用するための、またはそれと接続している、任意の非一過性コンピュータ可読記憶媒体であって、ソフトウェアに関連する命令を命令実行システム、装置またはデバイスから取り出し、命令を実行することができる可読記憶媒体の中に、ソフトウェア750を記憶および／またはトランスポートすることもできる。本開示に関して、コンピュータ可読記憶媒体は、命令実行システム、装置もしくはデバイスで使用するための、またはそれと接続している、プログラミングを収容または記憶することができる任意の媒体、例えば、記憶装置740であり得る。

【0152】

上記のものなどの命令実行システム、装置もしくはデバイスで使用するための、またはそれと接続している、任意のトランスポート媒体であって、ソフトウェアに関連する命令を命令実行システム、装置またはデバイスから取り出し、命令を実行することができるトランスポート媒体の中に、ソフトウェア750を伝播することもできる。本開示に関して、トランスポート媒体は、命令実行システム、装置もしくはデバイスで使用するための、またはそれと接続している、プログラミングを伝える、伝播するまたはトランスポートすることができる、任意の媒体であり得る。トランスポート可読媒体としては、電子、磁気、光、電磁または赤外有線もしくは無線伝播媒体を挙げることができるが、これらに限定されない。

【0153】

デバイス700をネットワークに接続することができ、これは任意の好適なタイプの相互接続通信システムであり得る。ネットワークは、任意の好適な通信プロトコルを実行することができ、ネットワークを任意の好適なセキュリティープロトコルにより保護することができる。ネットワークは、ネットワークシグナルの通信および受信を実行することができる任意の好適な構成のネットワークリンク、例えば、無線ネットワーク接続、T1もしくはT3ライン、ケーブルネットワーク、DSL、または電話線を含むことができる。

【0154】

デバイス700は、ネットワークでの操作に好適な任意の操作システムを実装することができる。ソフトウェア750を任意の好適なプログラミング言語、例えば、C、C++、Java（登録商標）またはPythonで書くことができる。様々な実施形態では、本開示の機能性を具現化するアプリケーションソフトウェアを、例えば、異なる配置で、例えばクライアント／サーバー構成で、またはウェブベースのアプリケーションもしくはウェブサービスのようウェブブラウザによって、展開することができる。

【0155】

本明細書に記載される方法は、解析方法を使用して決定された情報を報告するステップ、および／または解析方法を使用して決定された情報を含むレポートを生成するステップを、必要に応じてさらに含む。例えば、一部の実施形態では、方法は、対象に由来する（例えば、対象のゲノム内の）ポリヌクレオチドにおけるバリエーションの同定に関する___を報告するステップまたはそれを含むレポートを生成するステップをさらに含む。報告される情報、またはレポートの中の情報は、例えば、参照配列にマッピングされたカップリングされたシーケンシングリードペアの遺伝子座、検出されたバリエーション（例えば、検出された構造バリエーションまたは検出されたSNP）、1つもしくは複数のアセンブルされたコンセンサス配列、および／または1つもしくは複数のアセンブルされたコンセンサス配列についての検証統計量に関連し得る。受信者、例えば、臨床医、対象または研究者に、レポートを配布することができ、または情報を報告することができる。

【0156】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：(a) 1または複数台のプロセッサで、短い標的遺伝子バリエーションを選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連す

10

20

30

40

50

る標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0157】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングして、複数のフロー位置におけるフローシグナルを含む1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来する、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデ

10

20

30

40

50

ータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0158】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：(a) 短い標的遺伝子バリエーションを事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0159】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：(a) 短い標的遺伝子バリエーションを事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングして、複数のフロー位置におけるフローシグナルを含む1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来する、ステップ；(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；およ

び（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0160】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：（a）短い標的遺伝子バリエーションおよびフローサイクル順序を事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；（c）試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0161】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：（a）短い標的遺伝子バリエーションおよびフローサイクル順序を事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）事前選択されたフローサイクル順序に従って別々のヌクレオチド

フローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングして、複数のフロー位置におけるフローシグナルを含む1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来する、ステップ；（c）試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0162】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：（a）1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（b）1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、第1のフローサイクル順序および第2のフローサイクル順序は異なり、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（c）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および（d）決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップ。

【0163】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：（a）第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各々の第1の試験シーケンシングデータセットが、異なる試験核酸分子に関連している、ステップ；（b）第1のフローサイクル順序と異なる第2のフローサイクル順序

に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各々の第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；（c）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および（d）決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップ。

10

【0164】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：（a）1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（b）1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、第1のフローサイクル順序および第2のフローサイクル順序は異なり、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（c）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；（d）2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列が、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および（e）選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそれより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

20

30

40

【0165】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：（a）第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シーケンシングデータセットを得るステップであ

50

って、各々の第1の試験シーケンシングデータセットが、異なる試験核酸分子に関連している、ステップ；(b) 第1のフローサイクル順序と異なる第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各々の第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；(c) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；(d) 2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列が、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および(e) 選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそれより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

【0166】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；(b) 1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(c) 1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(d) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および(e) 決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーション

トの存在または非存在をコールするステップ。

【0167】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；(b) 第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各々の第1の試験シーケンシングデータセットが、異なる試験核酸分子に関連している、ステップ；(c) 第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各々の第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；(d) 第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；および(e) 決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップ。

【0168】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；(b) 1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各第1の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第1の試験シーケンシングデータセットが、第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1つまたは複数の第1の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；(c) 1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第2の試験シーケンシングデータセットが、第2のフローサイクル順序に従って別々のヌクレオチドフローで

提供される非終結ヌクレオチドを使用して1つまたは複数の試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；（d）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列（これは参照配列を含み得る）についてのマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；（e）2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列が、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および（f）選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそれより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

【0169】

一部の実施形態では、1または複数台のプロセッサと、以下についての命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステムが存在する：（a）短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第1のフローサイクル順序または第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所またはそれより多くのフロー位置において異なり、第1のフローサイクル順序と第2のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップ；（b）第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第1の試験シーケンシングデータセットを得るステップであって、各々の第1の試験シーケンシングデータセットが、異なる試験核酸分子に関連している、ステップ；（c）第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験試料に由来する同じ1つまたは複数の試験核酸分子をシーケンシングして、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む1つまたは複数の第2の試験シーケンシングデータセットを得るステップであって、各々の第2の試験シーケンシングデータセットが、第1の試験シーケンシングデータセットの1つと同じ試験核酸分子に関連している、ステップ；（d）第1のシーケンシングデータセットおよび第2のシーケンシングデータセットごとに、1つまたは複数の候補配列に対するマッチスコアを決定するステップであって、マッチスコアが、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方が1つまたは複数の候補配列からの候補配列にマッチする尤度を示す、ステップ；（e）2つまたはそれより多くの異なる候補配列（これらは参照配列を含み得る）から候補配列を選択するステップであって、選択候補配列が、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および（f）選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップ。一部の実施形態では、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とは、第1のフローサイクル順序および／または第2のフローサイクル順序によれば、2カ所またはそ

10

20

30

40

50

れより多くの（または3カ所もしくはそれより多くの、または1もしくは複数のフローサイクルにわたっての）フロー位置（これらの位置は、連続的であることもあり、または非連続であることもある）において異なる。

【0170】

一部の実施形態では、本明細書に記載される方法は、図7で説明される構成要素のうちの1つまたは複数を使用して遂行され得る、コンピュータ実装方法である。例えば、一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するためのコンピュータ実装方法は、（a）1または複数台のプロセッサを使用して、短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）1または複数台のプロセッサで、1つまたは複数の試験シーケンシングデータセットを受信するステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；（c）1または複数台のプロセッサを使用して、試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1または複数台のプロセッサおよび1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0171】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するためのコンピュータ実装方法は、（a）1または複数台のプロセッサを使用して、短い標的遺伝子バリエーションを事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）1または複数台のプロセッサで、1つまたは複数の試験シーケンシングデータセットを受信するステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシー

クエンシングすることにより決定され、試験シークエンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；（c）1または複数台のプロセッサで、試験シークエンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シークエンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シークエンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1または複数台のプロセッサで、1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シークエンシングデータセットと、参照シークエンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シークエンシングデータセットと、参照シークエンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0172】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するためのコンピュータ実装方法は、（a）1または複数台のプロセッサを使用して、短い標的遺伝子バリエーションおよびフローサイクル順序を事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シークエンシングデータセットと、参照配列に関連する参照シークエンシングデータセットとが、標的シークエンシングデータセットおよび参照シークエンシングデータセットが事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシークエンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；（b）1または複数台のプロセッサで、1つまたは複数の試験シークエンシングデータセットを受信するステップであって、各試験シークエンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シークエンシングデータセットが、事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシークエンシングすることにより決定され、試験シークエンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；（c）1または複数台のプロセッサで、試験シークエンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シークエンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シークエンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および（d）1または複数台のプロセッサで、1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シークエンシングデータセットと、参照シークエンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シークエンシングデータセットと、参照シークエンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0173】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するためのコンピュータ実装方法は、（a）1または複数台のプロセッサで、短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シー

クエンシングデータセットと、参照配列に関連する参照シークエンシングデータセットとが、標的シークエンシングデータセットおよび参照シークエンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシークエンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 1または複数台のプロセッサで、1つまたは複数の試験シークエンシングデータセットを受信するステップであって、各試験シークエンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シークエンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシークエンシングすることにより決定され、試験シークエンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；(c) 1または複数台のプロセッサで、試験シークエンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シークエンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シークエンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1または複数台のプロセッサで、1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シークエンシングデータセットと、参照シークエンシングデータセットとは、2カ所より多くのフロー位置(例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置)において異なる。一部の実施形態では、標的シークエンシングデータセットと、参照シークエンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0174】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するためのコンピュータ実装方法は、(a) 1または複数台のプロセッサで、短い標的遺伝子バリエーションを事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シークエンシングデータセットと、参照配列に関連する参照シークエンシングデータセットとが、標的シークエンシングデータセットおよび参照シークエンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシークエンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 1または複数台のプロセッサで、1つまたは複数の試験シークエンシングデータセットを受信するステップであって、各試験シークエンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シークエンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシークエンシングすることにより決定され、試験シークエンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；(c) 1または複数台のプロセッサで、試験シークエンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シークエンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シークエンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1または複数台のプロセッサで、1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネル

10

20

30

40

50

を生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

【0175】

一部の実施形態では、試験試料中の短い遺伝子バリエーションを検出するためのコンピュータ実装方法は、(a) 1または複数台のプロセッサで、短い標的遺伝子バリエーションおよびフローサイクル順序を事前選択するステップであって、事前選択された短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；(b) 1または複数台のプロセッサで、1つまたは複数の試験シーケンシングデータセットを受信するステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、事前選択された短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、事前選択されたフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；(c) 1または複数台のプロセッサで、試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および(d) 1または複数台のプロセッサで、1つまたは複数の決定されたマッチスコアを使用して、試験試料中の事前選択された短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む。一部の実施形態では、方法は、試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップをさらに含み、このバイオマーカーパネルは、短い標的遺伝子バリエーションを含む。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、2カ所より多くのフロー位置（例えば、2カ所より多くの連続するフロー位置、または2カ所より多くの非連続のフロー位置）において異なる。一部の実施形態では、標的シーケンシングデータセットと、参照シーケンシングデータセットとは、1または複数のフローサイクルにわたって異なる。

例示の実施形態

【0176】

以下の実施形態は、例示的なものであり、請求項記載の本発明の範囲を限定するように意図されたものではない。

【0177】

実施形態1. 試験試料中の短い遺伝子バリエーションを検出するための方法であって、

(a) 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットがフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2カ所より多くのフロー位置において異なり、フロー位置が、ヌクレオチドフローに対応する、ステップ；

(b) 1つまたは複数の試験シーケンシングデータセットを得るステップであって、各試験シーケンシングデータセットが、試験核酸分子に関連し、各試験核酸分子が、短い標的遺伝子バリエーションに関連する遺伝子座に少なくとも部分的に重なり、かつ試験試料に由来し、1つまたは複数の試験シーケンシングデータセットが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることにより決定され、試験シーケンシングデータセットが、複数のフロー位置におけるフローシグナルを含む、ステップ；

(c) 試験シーケンシングデータセットに関連する試験核酸分子の各々について、核酸分子に関連する試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコア、または核酸分子に関連する試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定するステップ；および

10

(d) 1つまたは複数の決定されたマッチスコアを使用して、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含む方法。

【0178】

実施形態2. 得るステップが、フローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングすることを含む、実施形態1の方法。

【0179】

実施形態3. 短い標的遺伝子バリエーションが、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの前に事前選択される、実施形態1または実施形態2の方法。

20

【0180】

実施形態4. 短い標的遺伝子バリエーションが、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの後にコールの信頼度に基づいて選択される、実施形態1または実施形態2に記載の方法。

【0181】

実施形態5. 試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップを含み、バイオマーカーパネルが、短い標的遺伝子バリエーションを含む、実施形態1から4のいずれか1つの方法。

30

【0182】

実施形態6. フローサイクル順序を選択するステップを含む、実施形態1から5のいずれか1つの方法。

【0183】

実施形態7. 標的シーケンシングデータセットが、予想標的シーケンシングデータセットであり、または参照シーケンシングデータセットが、予想参照シーケンシングデータセットである、実施形態1から6のいずれか1つの方法。

【0184】

実施形態8. 標的シーケンシングデータセットおよび予想参照シーケンシングデータセットが、標的配列および参照配列を *in silico* でシーケンシングすることにより得られる、実施形態7の方法。

40

【0185】

実施形態9. 標的シーケンシングデータセットと参照シーケンシングデータとが、2カ所より多くの非連続のフロー位置において異なる、実施形態1から8のいずれか1つの方法。

【0186】

実施形態10. 標的シーケンシングデータセットと参照シーケンシングデータとが、2カ所より多くの連続するフロー位置において異なる、実施形態1から9のいずれか1つの方法。

【0187】

50

実施形態 11. 標的配列と参照配列とが、 X カ所の塩基位置において異なり、標的シーケンシングデータセットと参照シーケンシングデータとが、 $(X+2)$ カ所またはそれより多くの連続するフロー位置において異なる、実施形態 1 から 10 のいずれか 1 つの方法。

【0188】

実施形態 12. $(X+2)$ カ所のフロー位置の相違が、実質的にゼロに等しい値と実質的にゼロより大きい値の間の相違を含む、実施形態 11 の方法。

【0189】

実施形態 13. 標的シーケンシングデータセットと参照シーケンシングデータセットとが、1 または複数のフローサイクルにわたって異なる、実施形態 1 から 12 のいずれか 1 つの方法。

10

【0190】

実施形態 14. フローシグナルが、各フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す塩基カウントを含む、実施形態 1 から 13 のいずれか 1 つの方法。

【0191】

実施形態 15. フローシグナルが、各フロー位置における少なくとも 1 塩基カウントの尤度を示す統計パラメーターを含み、塩基カウントが、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す、実施形態 1 から 14 のいずれか 1 つの方法。

【0192】

20

実施形態 16. フローシグナルが、各フロー位置における複数の塩基カウントの尤度を示す統計パラメーターを含み、各塩基カウントが、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す、実施形態 1 から 15 のいずれか 1 つの方法。

【0193】

実施形態 17. ステップ (c) が、試験シーケンシングデータセット内の各フロー位置における統計パラメーターであって、そのフロー位置における標的配列の塩基カウントに対応する統計パラメーターを選択すること、および試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコアを決定すること；または

試験シーケンシングデータセット中の各フロー位置における統計パラメーターであって、そのフロー位置における参照配列の塩基カウントに対応する統計パラメーターを選択すること、および試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定することを含む、実施形態 16 の方法。

30

【0194】

実施形態 18. ステップ (c) で決定されるマッチスコアが、試験シーケンシングデータセット内のフロー位置にわたっての選択された統計パラメーターの組み合わせられた値である、実施形態 17 の方法。

【0195】

実施形態 19. ステップ (c) が、試験シーケンシングデータセットが標的配列にマッチする尤度を示すマッチスコアを決定することを含む、実施形態 1 から 18 のいずれか 1 つの方法。

40

【0196】

実施形態 20. ステップ (c) が、試験シーケンシングデータセットが参照配列にマッチする尤度を示すマッチスコアを決定することを含む、実施形態 1 から 19 のいずれか 1 つの方法。

【0197】

実施形態 21. 1 つまたは複数の試験シーケンシングデータセットが、複数の試験シーケンシングデータセットを含む、実施形態 1 から 20 のいずれか 1 つの方法。

【0198】

50

実施形態 22. 短い標的遺伝子バリエーションの存在または非存在が、1つまたは複数の試験シーケンシングデータセットの各々について別々にコールされる、実施形態 21 の方法。

【0199】

実施形態 23. 複数の試験シーケンシングデータセットの少なくとも一部分が、異なる試験シーケンシング開始位置を有する異なる試験核酸分子に関連する、実施形態 21 または 22 の方法。

【0200】

実施形態 24. フローサイクル順序が、同じ順序で反復される 4 つの別々のフローを含む、実施形態 1 から 23 のいずれか 1 つの方法。

【0201】

実施形態 25. フローサイクル順序が、5 つまたはそれより多くの別々のフローを含む、実施形態 1 から 24 のいずれか 1 つの方法。

【0202】

実施形態 26. 1 または複数台のプロセッサを使用して短い標的遺伝子バリエーションを選択するステップ；

1 つまたは複数の試験シーケンシングデータセットを 1 または複数台のプロセッサで受信することにより、1 つまたは複数の試験シーケンシングデータセットを得るステップ；

1 または複数台のプロセッサを使用して 1 つまたは複数のマッチスコアを決定するステップ；および

1 または複数台のプロセッサを使用して試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップ
を含むコンピュータ実装方法である、実施形態 1 から 25 のいずれか 1 つの方法。

【0203】

実施形態 27. 1 または複数台のプロセッサと、
実施形態 1 から 26 のいずれか 1 つの方法を実行するための命令を含む 1 つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体と
を含むシステム。

【0204】

実施形態 28. 試験試料中の短い遺伝子バリエーションを検出するための方法であって、

(a) 1 つまたは複数の第 1 の試験シーケンシングデータセットを得るステップであって、各第 1 の試験シーケンシングデータセットが、試験試料に由来する異なる試験核酸分子に関連し、第 1 の試験シーケンシングデータセットが、第 1 のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して 1 つまたは複数の試験核酸分子をシーケンシングすることにより決定され、1 つまたは複数の第 1 の試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；

(b) 1 つまたは複数の第 2 の試験シーケンシングデータセットを得るステップであって、各第 2 の試験シーケンシングデータセットが、第 1 の試験シーケンシングデータセットと同じ試験核酸分子に関連し、第 2 の試験シーケンシングデータセットが、第 2 のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して 1 つまたは複数の試験核酸分子をシーケンシングすることにより決定され、第 1 のフローサイクル順序と第 2 のフローサイクル順序が異なり、試験シーケンシングデータセットが、ヌクレオチドフローに対応するフロー位置におけるフローシグナルを含む、ステップ；

(c) 第 1 のシーケンシングデータセットおよび第 2 のシーケンシングデータセットの各々について、1 つまたは複数の候補配列についてのマッチスコアを決定するステップであって、マッチスコアが、第 1 の試験シーケンシングデータセット、第 2 の試験シーケンシングデータセット、または両方が 1 つまたは複数の候補配列からの候補配列の

10

20

30

40

50

マッチする尤度を示す、ステップ；および

(d) 決定されたマッチスコアを使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む方法。

【0205】

実施形態29. 第1のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングするステップ、および第2のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して試験核酸分子をシーケンシングするステップを含む、実施形態28の方法。

【0206】

実施形態30. マッチスコアが、第1の試験シーケンシングデータセットが候補配列にマッチする尤度、または第2の試験シーケンシングデータセットが候補配列にマッチする尤度を示す、実施形態28または29の方法。

【0207】

実施形態31. マッチスコアが、第1の試験シーケンシングデータセットと第2のシーケンシングデータセット両方が候補配列にマッチする尤度を示す、実施形態28または29の方法。

【0208】

実施形態32. 1つまたは複数の候補配列が、2つまたはそれより多くの異なる候補配列を含み、方法が、第1のシーケンシングデータセットおよび第2のシーケンシングデータセットに関連する核酸分子の各々について：

2つまたはそれより多くの異なる候補配列から候補配列を選択するステップであって、選択候補配列が、第1の試験シーケンシングデータセット、第2の試験シーケンシングデータセット、または両方との最高尤度マッチを有する、ステップ；および選択候補配列を使用して、試験試料中の短い遺伝子バリエーションの存在または非存在をコールするステップを含む、実施形態28から31のいずれか1つの方法。

【0209】

実施形態33. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くのフロー位置において異なる、実施形態32の方法。

【0210】

実施形態34. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くのフロー位置において異なる、実施形態32の方法。

【0211】

実施形態35. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くの非連続のフロー位置において異なる、実施形態32の方法。

【0212】

実施形態36. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くの非連続のフロー位置において異なる、実施形態32の方法。

【0213】

実施形態37. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非

10

20

30

40

50

選択候補配列と、選択候補配列とが、第1のフローサイクル順序または第2のフローサイクル順序によれば、2カ所またはそれより多くの連続のフロー位置において異なる、実施形態32の方法。

【0214】

実施形態38. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、2カ所またはそれより多くの連続のフロー位置において異なる、実施形態32の方法。

【0215】

実施形態39. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序または第2のフローサイクル順序によれば、3カ所またはそれより多くのフロー位置において異なる、実施形態32の方法。

10

【0216】

実施形態40. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、3カ所またはそれより多くのフロー位置において異なる、実施形態32の方法。

【0217】

実施形態41. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、Xカ所の塩基位置において異なり、試験核酸分子に関連する試験シーケンシングデータセットと、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列とが、第1のフローサイクル順序または第2のフローサイクル順序によれば、 $(X+2)$ カ所またはそれより多くのフロー位置において異なる、実施形態32の方法。

20

【0218】

実施形態42. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、Xカ所の塩基位置において異なり、試験核酸分子に関連する試験シーケンシングデータセットと、2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列とが、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、 $(X+2)$ カ所またはそれより多くのフロー位置において異なる、実施形態32の方法。

30

【0219】

実施形態43. $(X+2)$ カ所のフロー位置の相違が、実質的にゼロに等しい値と実質的にゼロより大きい値の間の相違を含む、実施形態41または42の方法。

【0220】

実施形態44. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序または第2のフローサイクル順序によれば、1または複数のフローサイクルにわたって異なる、実施形態32の方法。

40

【0221】

実施形態45. 2つまたはそれより多くの異なる候補配列からの少なくとも1つの非選択候補配列と、選択候補配列とが、第1のフローサイクル順序および第2のフローサイクル順序の両方によれば、1または複数のフローサイクルにわたって異なる、実施形態32の方法。

【0222】

実施形態46. フローシグナルが、各フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す塩基カウントを含む、実施形態28から45のいずれか1つの方法。

【0223】

50

実施形態 47. フローシグナルが、各フロー位置における少なくとも 1 塩基カウントの尤度を示す統計パラメーターを含み、塩基カウントが、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す、実施形態 28 から 46 のいずれか 1 つの方法。

【0224】

実施形態 48. フローシグナルが、各フロー位置における複数の塩基カウントの尤度を示す統計パラメーターを含み、各塩基カウントが、フロー位置におけるシーケンシングされた試験核酸分子の塩基の数を示す、実施形態 28 から 47 のいずれか 1 つの方法。

【0225】

実施形態 49. マッチスコアを決定するステップが、第 1 の試験シーケンシングデータセットおよび第 2 の試験シーケンシングデータセットにおける各フロー位置における統計パラメーターであって、候補配列のそのフロー位置における塩基カウントに対応する統計パラメーターを、1 つまたは複数の異なる候補配列の各々について選択することを含む、実施形態 48 の方法。

10

【0226】

実施形態 50. 1 つまたは複数の異なる候補配列について、各フロー位置における候補配列の塩基カウントを含む候補シーケンシングデータセットを生成するステップを含む、実施形態 49 の方法。

【0227】

実施形態 51. 候補シーケンシングデータセットが、*in silico* で生成される、実施形態 50 の方法。

20

【0228】

実施形態 52. マッチスコアが、第 1 の試験シーケンシングデータセットおよび第 2 の試験シーケンシングデータセットにおけるフロー位置にわたって選択された統計パラメーターの組み合わせられた値である、実施形態 49 から 51 のいずれか 1 つの方法。

【0229】

実施形態 53. 試験核酸分子の少なくとも一部分が、異なるシーケンシング開始位置を有する、実施形態 28 から 52 のいずれか 1 つの方法。

【0230】

実施形態 54. 短い標的遺伝子バリエーションを選択するステップであって、短い標的遺伝子バリエーションを含む標的配列に関連する標的シーケンシングデータセットと、参照配列に関連する参照シーケンシングデータセットとが、標的シーケンシングデータセットおよび参照シーケンシングデータセットが、第 1 のフローサイクル順序または第 2 のフローサイクル順序に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して標的配列をシーケンシングすることにより得られた場合、2 カ所またはそれより多くのフロー位置において異なり、第 1 のフローサイクル順序と第 2 のフローサイクル順序とが異なり、フロー位置がヌクレオチドフローに対応する、ステップを含む、

30

1 つまたは複数の候補配列が、標的配列および参照配列を含む、実施形態 28 から 52 のいずれか 1 つの方法。

40

【0231】

実施形態 55. 短い標的遺伝子バリエーションが、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの前に事前選択される、実施形態 54 の方法。

【0232】

実施形態 56. 短い標的遺伝子バリエーションが、試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップの後にコールの信頼度に基づいて選択される、実施形態 54 の方法。

【0233】

実施形態 57. 試験試料に関連する対象についての個別化バイオマーカーパネルを生成するステップを含み、バイオマーカーパネルが、試験試料中に存在する短い標的遺伝子

50

バリエーションを含む、実施形態 56 の方法。

【0234】

実施形態 58. 参照シーケンシングデータセットが、第 1 のフローサイクル順序または第 2 のフローサイクル順序に従って別々のフローで提供される非終結ヌクレオチドを使用して参照配列がシーケンシングされた場合の予想参照シーケンシングデータセットを決定することにより得られる、実施形態 54 から 57 のいずれか 1 つの方法。

【0235】

実施形態 59. 参照シーケンシングデータセットが、第 1 のフローサイクル順序および第 2 のフローサイクル順序の両方に従って別々のフローで提供される非終結ヌクレオチドを使用して参照配列がシーケンシングされた場合の予想参照シーケンシングデータセットを決定することにより得られる、実施形態 54 から 57 のいずれか 1 つの方法。

【0236】

実施形態 60. 標的配列と参照配列とが、第 1 のフローサイクル順序および第 2 のフローサイクル順序の両方によれば、2 カ所またはそれより多くのフロー位置において異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0237】

実施形態 61. 標的配列と参照配列とが、第 1 のフローサイクル順序または第 2 のフローサイクル順序によれば、2 カ所またはそれより多くの非連続のフロー位置において異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0238】

実施形態 62. 標的配列と参照配列とが、第 1 のフローサイクル順序および第 2 のフローサイクル順序の両方によれば、2 カ所またはそれより多くの非連続のフロー位置において異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0239】

実施形態 63. 標的配列と参照配列とが、第 1 のフローサイクル順序または第 2 のフローサイクル順序によれば、2 カ所またはそれより多くの連続であるフロー位置において異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0240】

実施形態 64. 標的配列と参照配列とが、第 1 のフローサイクル順序および第 2 のフローサイクル順序の両方によれば、2 カ所またはそれより多くの連続であるフロー位置において異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0241】

実施形態 65. 標的配列と参照配列とが、第 1 のフローサイクル順序または第 2 のフローサイクル順序によれば、3 カ所またはそれより多くのフロー位置において異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0242】

実施形態 66. 標的配列と参照配列とが、第 1 のフローサイクル順序および第 2 のフローサイクル順序の両方によれば、3 カ所またはそれより多くのフロー位置において異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0243】

実施形態 67. 標的配列と参照配列とが、第 1 のフローサイクル順序または第 2 のフローサイクル順序によれば、2 カ所またはそれより多くのフローサイクルにわたって異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0244】

実施形態 68. 標的配列と参照配列とが、第 1 のフローサイクル順序および第 2 のフローサイクル順序の両方によれば、2 カ所またはそれより多くのフローサイクルにわたって異なる、実施形態 54 から 57 のいずれか 1 つの方法。

【0245】

実施形態 69. 第 1 のフローサイクル順序または第 2 のフローサイクル順序が、同じ順序で反復される 4 つの別々のフローを含む、実施形態 28 から 68 のいずれか 1 つの方

10

20

30

40

50

法。

【0246】

実施形態70. 第1のフローサイクル順序または第2のフローサイクル順序が、同じ順序で反復される5つまたはそれより多くの別々のフローを含む、実施形態28から68のいずれか1つの方法。

【0247】

実施形態71. 試験核酸分子をシークエンシングするステップであって、第1のフローサイクル順序に従って別々のヌクレオチドフローで非終結ヌクレオチドを提供することと、シークエンシングプライマーを伸長することと、各ヌクレオチドフローの後にシークエンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して第1の試験シークエンシングデータセットを生成することを含むステップ；

伸長されたシークエンシングプライマーを除去するステップ；および

同じ試験核酸分子をシークエンシングするステップであって、第2のフローサイクル順序に従って別々のヌクレオチドフローで非終結ヌクレオチドを提供することと、シークエンシングプライマーを伸長することと、各ヌクレオチドフローの後にシークエンシングプライマーへのヌクレオチド取り込みの存在または非存在を検出して第2の試験シークエンシングデータセットを生成することを含むステップを含む、実施形態28から70のいずれか1つの方法。

【0248】

実施形態72. 1つまたは複数の第1のシークエンシングデータセットを1または複数台のプロセッサで受信するステップ；

1つまたは複数の第1のシークエンシングデータセットを1または複数台のプロセッサで受信するステップ；

1または複数台のプロセッサを使用してマッチスコアを決定するステップ；および

1または複数台のプロセッサを使用して試験試料中の短い標的遺伝子バリエーションの存在または非存在をコールするステップを含むコンピュータ実装方法である、実施形態28から71のいずれか1つの方法。

【0249】

実施形態73. 1または複数台のプロセッサと、

実施形態28から72のいずれか1つの方法を実行するための命令を含む1つまたは複数のプログラムを記憶する非一過性コンピュータ可読媒体とを含むシステム。

【0250】

実施形態74. 別々のフローが、単一の塩基タイプを含む、実施形態1から73のいずれか1つの方法またはシステム。

【0251】

実施形態75. 別々のフローのうちの少なくとも1つが、2つまたは3つの異なる塩基タイプを含む、実施形態1から74のいずれか1つの方法またはシステム。

【0252】

実施形態76. 試験試料中の短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示すバリエーションコールファイルを生成またはアップデートするステップを含む、実施形態1から75のいずれか1つの方法またはシステム。

【0253】

実施形態77. 試験試料中の短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示すレポートを生成するステップを含む、実施形態1から76のいずれか1つの方法またはシステム。

【0254】

実施形態78. レポートが、試験試料中の短い遺伝子バリエーションの存在、アイデンティティまたは非存在を示す、テキスト出力、確率的出力、数値出力またはグラフィック出力を含む、実施形態77の方法またはシステム。

10

20

30

40

50

【0255】

実施形態79. レポートを患者にまたは患者の医療担当者に提供するステップを含む、実施形態77または78の方法またはシステム。

【0256】

実施形態78. 短い遺伝子バリエーションが、一塩基多型またはインデルを含む、実施形態1から77のいずれか1つの方法またはシステム。

【0257】

実施形態79. 短い遺伝子バリエーションが、インデルを含む、実施形態1から77のいずれか1つの方法またはシステム。

【0258】

実施形態80. 試験試料が、断片化DNAを含む、実施形態1から79のいずれか1つの方法またはシステム。

【0259】

実施形態81. 試験試料が、無細胞DNAを含む、実施形態1から80のいずれか1つの方法またはシステム。

【0260】

実施形態82. 無細胞DNAが、循環腫瘍DNA (ctDNA) を含む、実施形態81の方法またはシステム。

【0261】

実施形態83. 核酸分子をシーケンシングする方法であって、
核酸分子をプライマーとハイブリダイズして、ハイブリダイズされた鋳型を形成するステップ；

5つまたはそれより多くの別々のヌクレオチドフローを含む反復フローサイクル順序に従って別々のヌクレオチドフローで提供される標識された非終結ヌクレオチドを使用してプライマーを伸長するステップ；および

取り込まれた標識ヌクレオチドからのシグナルを検出する、またはプライマーがヌクレオチドフローにより伸長された場合のシグナルの非存在を検出するステップを含む、方法。

【0262】

実施形態84. シグナル、またはシグナルの非存在を、各ヌクレオチドフロー後に検出するステップを含む、実施形態83の方法。

【0263】

実施形態85. 複数の核酸分子をシーケンシングするステップを含む、実施形態83または84の方法。

【0264】

実施形態86. 複数のうちの核酸分子が、ある遺伝子座に対して異なるシーケンシング開始位置を有する、実施形態85の方法。

【0265】

実施形態87. 試験試料が、無細胞DNAである、実施形態83から86のいずれか1つの方法。

【0266】

実施形態88. 無細胞DNAが、循環腫瘍DNA (ctDNA) を含む、実施形態83から86のいずれか1つの方法。

【0267】

実施形態89. フローサイクル順序が、2カ所より多くのフロー位置において、ランダムシーケンシング開始位置の5%またはそれより多くにおける可能なSNPパーミューテーションの50%またはそれより多くについてのシグナル変化を誘導する、実施形態83から86のいずれか1つの方法。

【0268】

実施形態90. フローサイクル順序が、フロー当たり0.6のまたはそれを超える塩基

10

20

30

40

50

取り込みの効率を有する、実施形態 8 3 から 8 6 のいずれか 1 つの方法。

【実施例】

【0269】

本願の例示的实施形態として提供する以下の非限定的実施例を参照することにより、本願をよりよく理解することができる。以下の実施例を、実施形態をより十分に説明するために提示するが、いかなる点においても本願の広い範囲を限定するものと解釈すべきでない。本願のある特定の实施形態を本明細書で示し、説明したが、このような实施形態を単なる例として提供することは明らかであろう。本発明の趣旨および範囲から逸脱しない非常に多くの変形形態、変更形態および置換形態に当業者なら想到するであろう。本明細書に記載する实施形態の様々な代替形態を、本明細書に記載する方法を実施する際に利用することができることは、理解されるはずである。

10

(実施例 1)

SNP 検出

【0270】

フローサイクル順序 A-T-G-C に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して仮想核酸分子をシーケンシングし、その結果、図 1 A に示す試験シーケンシングデータセットを得た。シーケンシングデータセットにおける各値は、各フロー位置における示されている塩基カウントが正しい尤度を示す。シーケンシングデータセットに基づいて、予備配列を T A T G G T C G T C G A (配列番号 1) として決定し、これを参照ゲノムの遺伝子座にマッピングする。参照ゲノムに遺伝子座を、可能性のあるハプロタイプ配列 T A T G G T C G T C G A (配列番号 1) (H 1) および T A T G G T C A T C G A (配列番号 2) (H 2) と関連付ける。各フロー位置についてのハプロタイプ配列の塩基カウントに関連する尤度値を、ハプロタイプごとに選択する。各ハプロタイプに付与されるシーケンシングデータセットの尤度を、各フロー位置についてのハプロタイプ配列の塩基カウントに関連する尤度値を乗じることにより決定する。H 1 が正しい配列である場合のシーケンシングデータセットの log 尤度は、-0.015 であり、H 2 が正しい配列である場合のシーケンシングデータセットの log 尤度は、-27.008 である。したがって、H 1 の配列をこの核酸分子に選択する。

20

(実施例 2)

インデル検出

【0271】

フローサイクル順序 A-T-G-C に従って別々のヌクレオチドフローで提供される非終結ヌクレオチドを使用して仮想核酸分子をシーケンシングし、その結果、図 8 に示す試験シーケンシングデータセットを得た。シーケンシングデータセットにおける各値は、各フロー位置における示されている塩基カウントが正しい尤度を示す。シーケンシングデータセットに基づいて(すなわち、各フロー位置において最も可能性の高い塩基カウントを選択することにより)、予備配列を T A T G G T C G A T C G (配列番号 8) として決定し、これを参照ゲノムの遺伝子座にマッピングする。参照ゲノムの遺伝子座を、可能性のあるハプロタイプ配列 T A T G G T C G - T C G A (配列番号 7) (H 1) および T A T G G T C G A T C G (配列番号 8) (H 2) と関連付ける。各フロー位置についてのハプロタイプ配列の塩基カウントに関連する尤度値を、ハプロタイプごとに選択する。各ハプロタイプに付与されるシーケンシングデータセットの尤度を、各フロー位置についてのハプロタイプ配列の塩基カウントに関連する尤度値を乗じることにより決定する。H 1 が正しい配列である場合のシーケンシングデータセットの log 尤度は、-24.009 であり、H 2 が正しい配列である場合のシーケンシングデータセットの log 尤度は、-0.015 である。したがって、H 2 の配列をこの核酸分子に選択する。

40

(実施例 3)

延長シーケンシングフロー順序

【0272】

100 万を超える延長シーケンシングフロー順序を、すべての可能な SNP のセット

50

にわたって2カ所より多くのフロー位置においてシグナル変化〔XYZ→XQZ、ここで、 $Q \neq Y$ （およびQ、X、YおよびZは、各々、A、C、GおよびTのいずれかを1つである）〕を誘導するそれらの尤度について、*in silico*で検定した。すべての有効な2塩基フローパーミュテーションで最低12塩基の配列を有するように延長フロー順序を設計し、逐次的塩基反復を有するフロー順序を除去した。フロー順序のすべての可能な開始位置を検定して、2カ所より多くのフロー位置においてシグナル変化を誘導する延長フロー順序の感度を評定した。図9および表2は、この解析の例示的結果を示す。図9中のx軸は、フロー相（または断片化開始位置）の分率を示し、y軸は、2カ所より多くのフロー位置においてシグナル変化を誘導したSNPパーミュテーションの分率を示す。いくつかのフロー順序は、リード（またはフロー開始位置）のおおよそ10%についてすべての可能な（87.5%）SNPパーミュテーションで2つまたはそれより多くのシグナルの相違を誘導する。4塩基周期的フローは、可能なSNPのたった42%でサイクルシフトを誘導するだけであるが、すべてのリードまたはフロー相で行なう。効率の最終評価をヒト参照ゲノムの100万塩基サブセットに対して遂行して、実行可能性を確立した。これは、現実の生物にはパターンおよび偏りがある配列をフロー順序がいかに効率的に伸長するかの実際的な評価基準である。

（実施例4）

SNP検出精度

【0273】

DNA試料NA12878（コリエル医学研究所（Coriell Institute for Medical Research）から入手可能な試料）のゲノムを、4フローサイクル（T-A-C-G）に従って非終結蛍光標識ヌクレオチドを使用してシーケンシングした。シーケンシング実行により、平均長が176塩基である415,900,002のリードが生成された。399,804,925リードをhg38参照ゲノムと（BWA、バージョン0.7.17-r1188で）アラインメントした。

【0274】

アラインメント後、参照ゲノムと完全にアラインしたリード（178,634,625リード）、または参照ゲノムとの単一ミスマッチを有し、20のもしくはそれを超えるマッピング品質スコアでアラインしたリード（27,265,661リード）を選択した。つまり、193,904,639は、例えば、インデル、複数のミスマッチ、または参照ゲノムとの誤っている（アーチファクトの）可能性のあるアラインメントを有するため、さらなる解析に含めなかった。したがって、27,265,661リードは、真の陽性NA12878 SNPはもちろん、シーケンシングエラーから生じるあらゆる偽陽性SNPも含むと推定した。27,265,661リードのこのプールから、真の陽性NA12878 SNPバリエーションの効果を低下させる1回より多くミスマッチ遺伝子座に及んだシーケンシングリードを除去し、その結果、深度1のミスマッチを有する合計3,413,700リードを得た。

【0275】

残りの3,413,700リード各々は、（1）フローグラムフローシグナルがフローサイクル順序に基づいて参照に対して1フルサイクル（例えば、4フロー位置）シフトした場合、サイクルシフトを誘導すると予想されるミスマッチ、（2）異なるフローサイクルを使用した場合、サイクルシフトを誘導し得る（例えば、それが、フローグラムで新しいゼロもしくは新しい非ゼロシグナルを生成する）可能性のあるミスマッチ、または（3）フローサイクル順序に関係なくサイクルシフトを誘導することができないであろうミスマッチを含んだ。3,413,700ミスマッチのうち、1,184,954（34%）は、サイクルシフトを誘導したが、1,546,588（43%）は、異なるフロー順序でサイクルシフト（すなわち、「可能性のあるサイクルシフト」）を誘導することがあった。比較して、ランダムミスマッチの理論的予想は、名目上、サイクルシフト42%および可能性のあるサイクルシフトミスマッチ46%を示唆した。全体的に見て、サイクルシフトを誘導するミスマッチ率は、 3.7×10^{-5} 事象／塩基であり、可能性のあるサイ

クルシフトを誘導するミスマッチ率は、 4.8×10^{-5} 事象／塩基であった。表 3 は、サイクルシフトを誘導する 10 の最高頻度単一ミスマッチ、および発生率の相対パーセンテージを示す。

【表 3】

表3

参照	リード	ケース%
TTT	TCT	7.18
AAA	AGA	7.18
GAG	GGG	4.63
CTC	CCC	4.62
CAG	CGG	4.12
CTG	CCG	4.09
AAC	AGC	3.86
GTT	GCT	3.83
CAT	CGT	3.63
GAT	GGT	3.62

【0276】

次いで、3つの異なるクラス（すなわち、サイクルシフトを誘導する、サイクルシフトを誘導する可能性がある、またはサイクルシフトを誘導しないおよび誘導することができない）の各々におけるミスマッチに基づくバリエーションコーリングの性能を評価した。BWAを用いてリードを参照ゲノムとアラインメントし、GATK（バージョン4）のHaplotypeCallerツールを使用してバリエーションコーリングを遂行した。得られたミスマッチコールを、10塩基より長いホモポリマー内のバリエーションコール、または10塩基もしくはそれを超える長さを有するホモポリマーに隣接する10塩基以内のバリエーションコールを捨てることにより、フィルター処理した。

【0277】

ミスマッチコールを、genome-in-the-bottle (GIAB) プロジェクトによって同じNA12878について生成されたコールと比較して、ミスマッチのクラスごとに精度 $\#TP / (\#FP + \#FN + \#TP)$ を決定した。シーケンシングデータを、示した平均ゲノム深度にランダムにダウンサンプリングした。サイクルシフトを誘導するミスマッチ、およびサイクルシフトを誘導する可能性のあるミスマッチは、表4で実証されるように、サイクルシフトを誘導しないミスマッチよりも高い精度を有した。

【表 4】

表4

ミスマッチのタイプ	30x	22x	15x	8x
サイクルシフト	0.9834	0.981	0.981	0.9772
サイクルシフトなし	0.9799	0.9759	0.9775	0.9696
可能性のあるサイクルシフト	0.9826	0.9808	0.9795	0.9767

10

20

30

40

50

【図面】

【図 1 A】

番号 カラム	サイクル 1				サイクル 2				サイクル 3				サイクル 4				サイクル 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
1	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
2	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
3	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010

FIG. 1A

【図 1 B】

番号 カラム	サイクル 1				サイクル 2				サイクル 3				サイクル 4				サイクル 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
1	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
2	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
3	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010

FIG. 1B

【図 1 C】

番号 カラム	サイクル 1				サイクル 2				サイクル 3				サイクル 4				サイクル 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
1	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
2	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
3	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010

FIG. 1C

【図 2 - 1】

TATGGTCGTCGA	R1
TGGTCGTCGAGC	R2
TATATGGTCGTC	R3
TATATGGTCATCGAGCTAT	H1
TATATGGTCGTCGAGCTAT	H2

FIG. 2A

R1シーケンシングデータセット

番号 カラム	サイクル 1				サイクル 2				サイクル 3				サイクル 4				サイクル 5			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
1	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
2	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010
3	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010	0010

FIG. 2B

【図 2 - 2】

塩基 カウント	サイクル 1					サイクル 2					サイクル 3					サイクル 4					サイクル 5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20					
0	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
1	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
2	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
3	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010

FIG. 2C

R3シナエエンシグナータセット																									
塩基 カウント	サイクル 1					サイクル 2					サイクル 3					サイクル 4					サイクル 5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20					
0	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G					
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010					
1	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A					
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010					
2	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C					
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010					
3	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T					
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010					

FIG. 2D

【図 3】

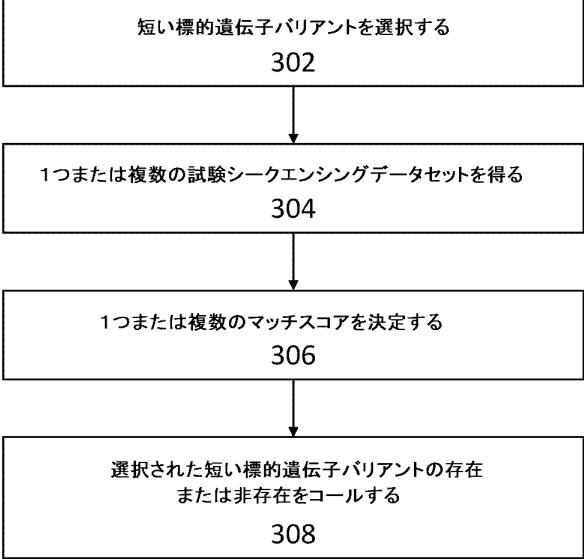


FIG. 3

【図 4】

塩基 カウント	サイクル1				サイクル2				サイクル3				サイクル4					サイクル5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
0	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G		
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010		
1	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010		
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010		
2	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010		
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010		
3	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010		
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010		

FIG. 4A

塩基 カウント	サイクル1					サイクル2					サイクル3					サイクル4					サイクル5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20					
0	A	G	C	T	A	G	C	T	A	G	C	T	A	G	C	T	A	G	C	T					
	0.9979	0.9979	0.9988	0.0010	0.0010	0.9988	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9988	0.0010	0.9979	0.9988	0.0010	0.9988	0.0010					
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
	0.0010	0.0010	0.0010	0.9988	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.0010	0.9979	0.0010	0.0010	0.9979	0.0010	0.9988					
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.9979	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010					
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.9979	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010					

FIG. 4B

【図 5】

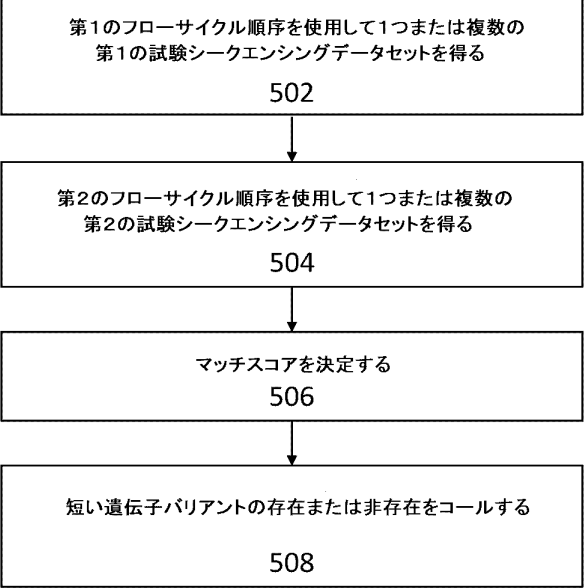


FIG. 5

10

20

30

40

50

【図 6】

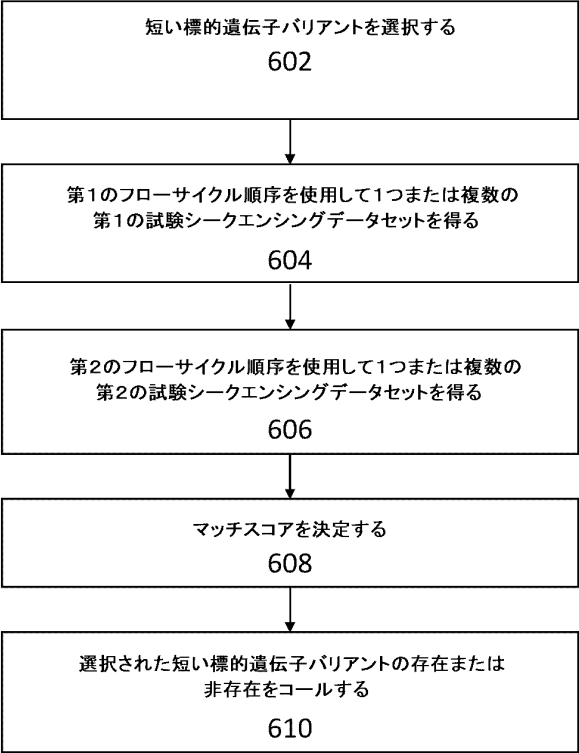


FIG. 6

【図 7】

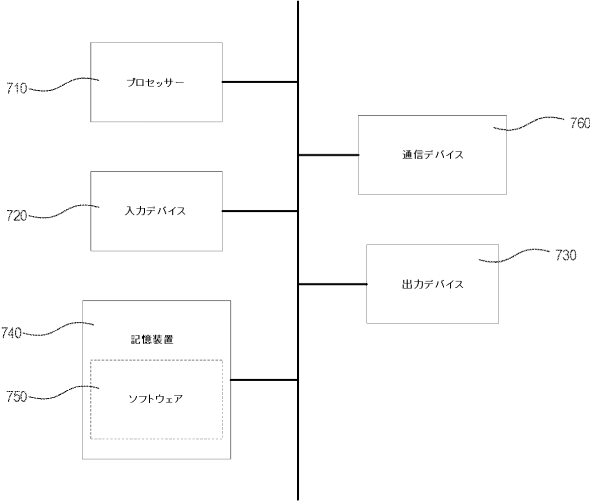


FIG. 7

【図 8】

基準 アライメント	サイクル 1					サイクル 2					サイクル 3					サイクル 4					サイクル 5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
1	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
2	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
3	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010

FIG. 8

【図 9】

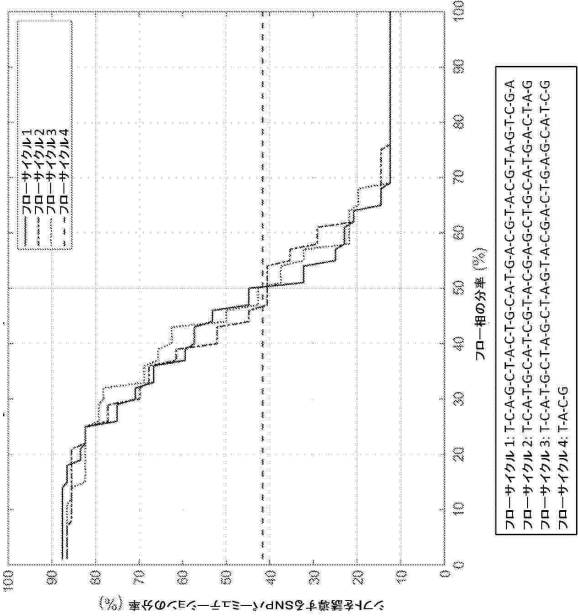


FIG. 9

【配列表】
0007605763000001.app

10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

(74)代理人 230113332

弁護士 山本 健策

(72)発明者 エツィオーニ, ヨアヴ

アメリカ合衆国 カリフォルニア 94560, ニューアーク, ゲートウェイ ブールバード 7
979, スイート 101, ウルティマ ジェノミクス, インコーポレイテッド 気付

(72)発明者 フェイグラ, シムチョン

アメリカ合衆国 カリフォルニア 94560, ニューアーク, ゲートウェイ ブールバード 7
979, スイート 101, ウルティマ ジェノミクス, インコーポレイテッド 気付

(72)発明者 アルモジー, ギラッド

アメリカ合衆国 カリフォルニア 94560, ニューアーク, ゲートウェイ ブールバード 7
979, スイート 101, ウルティマ ジェノミクス, インコーポレイテッド 気付

(72)発明者 プラット, マーク

アメリカ合衆国 カリフォルニア 94560, ニューアーク, ゲートウェイ ブールバード 7
979, スイート 101, ウルティマ ジェノミクス, インコーポレイテッド 気付

(72)発明者 オーバーストラス, フロリアン

アメリカ合衆国 カリフォルニア 94560, ニューアーク, ゲートウェイ ブールバード 7
979, スイート 101, ウルティマ ジェノミクス, インコーポレイテッド 気付

審査官 西 賢二

(56)参考文献 国際公開第2010/117804 (WO, A2)

米国特許出願公開第2013/073214 (US, A1)

米国特許出願公開第2013/0288902 (US, A1)

(58)調査した分野 (Int.Cl., DB名)

C12Q 1/6869

CAplus/MEDLINE/EMBASE/BIOSIS (STN)