

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2022-533801

(P2022-533801A)

(43)公表日 令和4年7月25日(2022.7.25)

(51)国際特許分類	F I	テーマコード (参考)
C 1 2 Q 1/6869(2018.01)	C 1 2 Q 1/6869	4 B 0 6 3
C 1 2 N 15/09 (2006.01)	C 1 2 N 15/09	Z

審査請求 未請求 予備審査請求 未請求 (全113頁)

(21)出願番号	特願2022-512704(P2022-512704)	(71)出願人	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
(85)翻訳文提出日	令和3年12月28日(2021.12.28)		レイテッド
(86)国際出願番号	PCT/US2020/031163		アメリカ合衆国 カリフォルニア 9 4 5
(87)国際公開番号	WO2020/227143		6 0, ニューアーク, ゲートウェイ
(87)国際公開日	令和2年11月12日(2020.11.12)		ブルバード 7 9 7 9, スイート 1
(31)優先権主張番号	62/842,534		0 1
(32)優先日	令和1年5月3日(2019.5.3)	(74)代理人	100078282
(33)優先権主張国・地域又は機関	米国(US)		弁理士 山本 秀策
(31)優先権主張番号	62/904,274	(74)代理人	100113413
(32)優先日	令和1年9月23日(2019.9.23)		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
	最終頁に続く	(74)代理人	100181641
			弁理士 石川 大輔
			最終頁に続く

(54)【発明の名称】 合成による高速フォワードシーケンシング

(57)【要約】

ポリヌクレオチドについてのカップリングされたシーケンシングリードペアを生成する方法、およびカップリングされたシーケンシングリードペアを解析する方法が、本明細書に記載される。カップリングされたシーケンシングリードペアを解析して、カップリングされたシーケンシングリードペアの中の直接シーケンシングされない遺伝子座にあるものを含むポリヌクレオチドバリエーションを検出することができる。他の解析方法は、カップリングされたシーケンシングリードペアを使用してコンセンサス配列を構築または検証するステップを含むことができる。

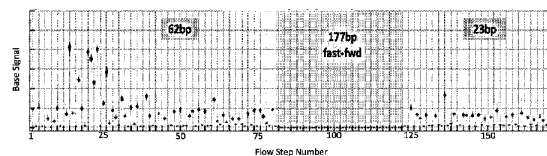


FIG. 19A

【特許請求の範囲】

【請求項 1】

カップリングされたシークエンシングリードペアをポリヌクレオチドから生成する方法であって、

(a) 前記ポリヌクレオチド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；

(b) 前記ポリヌクレオチドの第 1 の領域の配列に関連するシークエンシングデータを、標識ヌクレオチドを使用して前記プライマーを伸長すること、および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により、生成するステップ；

(c) ステップ (b) で伸長された前記プライマーを、第 2 の領域のフロー順序で提供されるヌクレオチドを使用して第 2 の領域を通してさらに伸長するステップであって、(i) 前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、前記第 2 の領域を通して伸長されるか、(ii) 少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、前記第 2 の領域のフロー順序の少なくとも 1 ステップで使用されるか、または (iii) 前記第 2 の領域を通じた前記プライマーの伸長が、ステップ (b) における前記プライマーの伸長よりも速く進行する、ステップ；および

(d) 前記ポリヌクレオチドの第 3 の領域の配列に関連するシークエンシングデータを、標識ヌクレオチドを使用してステップ (c) で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む方法。

【請求項 2】

前記第 2 の領域を通じた前記プライマーの伸長が、前記第 1 の領域を通じた前記プライマーの伸長より速く進行する、請求項 1 に記載の方法。

【請求項 3】

前記第 1 の領域の前記シークエンシングデータを前記第 3 の前記シークエンシングデータと関連付けるステップをさらに含む、請求項 1 または 2 に記載の方法。

【請求項 4】

カップリングされたシークエンシングリードペアをポリヌクレオチドから生成する方法であって、

(a) プライマーを前記ポリヌクレオチドの第 1 の領域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；

(b) 前記プライマーを、第 2 の領域のフロー順序で提供されるヌクレオチドを使用して第 2 の領域を通して伸長するステップであって、(i) 前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、前記第 2 の領域を通して伸長される、または (ii) 少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、前記第 2 の領域のフロー順序の少なくとも 1 ステップで使用される、ステップ；および

(c) 前記ポリヌクレオチドの第 3 の領域の配列に関連するシークエンシングデータを、標識ヌクレオチドを使用してステップ (b) で伸長された前記プライマーをさらに伸長すること、および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により、生成するステップを含む方法。

【請求項 5】

前記第 1 の領域が、前記プライマーの標的にされる天然に存在する配列を含む、請求項 4 に記載の方法。

【請求項 6】

前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、前記第 2 の領域を通して伸長される、請求項 1 から 5 のい

10

20

30

40

50

れか一項に記載の方法。

【請求項 7】

前記第 2 の領域を通して前記プライマーを伸長するために使用される前記ヌクレオチドの少なくとも一部分が、非標識ヌクレオチドである、請求項 1 から 6 のいずれか一項に記載の方法。

【請求項 8】

前記第 2 の領域を通して前記プライマーを伸長するために使用される前記ヌクレオチドが、非標識ヌクレオチドである、請求項 1 から 6 のいずれか一項に記載の方法。

【請求項 9】

少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、前記第 2 の領域のフロー順序の少なくとも 1 ステップで使用される、請求項 1 から 8 のいずれか一項に記載の方法。

10

【請求項 10】

前記第 2 の領域のフロー順序が、5 つまたはそれより多くのヌクレオチドフローを含む、請求項 1 から 9 のいずれか一項に記載の方法。

【請求項 11】

前記ヌクレオチドフローの各々が、単一のヌクレオチド塩基を含む、請求項 10 に記載の方法。

【請求項 12】

前記第 2 の領域のフロー順序が、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 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、請求項 10 または 11 に記載の方法。

20

【請求項 13】

前記第 2 の領域のフロー順序が、1 フロー当たり 0.6 のまたはそれを超える塩基組込みの効率を有する、請求項 10 から 12 のいずれか一項に記載の方法。

【請求項 14】

参照配列および前記第 2 の領域のフロー順序を使用して前記第 2 の領域についての予想シーケンシングデータを決定するステップをさらに含む、請求項 1 から 13 のいずれか一項に記載の方法。

30

【請求項 15】

前記プライマーが、第 3 の領域のフロー順序で提供されるヌクレオチドを使用して前記第 3 の領域を通して伸長され、前記方法が、前記第 2 の領域についての参照配列、前記第 2 の領域のフロー順序、前記第 3 の領域のフロー順序、および前記第 3 の領域についての参照配列を使用して、前記第 3 の領域についての予想シーケンシングデータを決定するステップをさらに含む、請求項 1 から 14 のいずれか一項に記載の方法。

【請求項 16】

前記第 3 の領域のフロー順序が、5 つまたはそれより多くのヌクレオチドフローを含む、請求項 15 に記載の方法。

【請求項 17】

前記ヌクレオチドフローの各々が、単一のヌクレオチド塩基を含む、請求項 16 に記載の方法。

40

【請求項 18】

前記第 3 の領域のフロー順序が、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 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、請求項 16 または 17 に記載の方法。

【請求項 19】

前記第 3 の領域のフロー順序が、1 フロー当たり 0.6 のまたはそれを超える塩基組込みの効率を有する、請求項 16 から 18 のいずれか一項に記載の方法。

50

【請求項 20】

前記プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して前記第3の領域を通して伸長され、前記方法が、前記第2の領域についての参照配列、前記第2の領域のフロー順序、前記第3の領域のフロー順序、および前記第3の領域の配列に関連するシーケンシングデータを使用して、前記第3の領域についての予想シーケンシングデータを決定するステップをさらに含み、前記第3の領域の配列に関連する前記シーケンシングデータが、前記第3の領域について生成される同じまたは異なるシーケンシングデータである、請求項1から19のいずれか一項に記載の方法。

【請求項 21】

前記第2の領域または前記第3の領域についての前記予想参照データが、バイナリまたは非バイナリプログラムを含む、請求項14から20のいずれか一項に記載の方法。

10

【請求項 22】

前記第2の領域のフロー順序、および前記第2の領域についての第2の参照配列を使用して、前記第2の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、前記第2の参照配列が、前記試験バリエーションを含む、請求項14から21のいずれか一項に記載の方法。

【請求項 23】

前記プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して前記第3の領域を通して伸長され、前記方法が、前記第2の領域についての第2の参照配列、前記第2の領域のフロー順序、前記第3の領域のフロー順序、および前記第3の領域についての参照配列を使用して、前記第3の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、請求項22に記載の方法。

20

【請求項 24】

前記プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して前記第3の領域を通して伸長され、前記方法が、前記第2の領域についての第2の参照配列、前記第2の領域のフロー順序、前記第3の領域のフロー順序、および前記第3の領域の配列に関連するシーケンシングデータを使用して、前記第3の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、前記第3の領域の配列に関連する前記シーケンシングデータが、前記第3の領域について生成される同じまたは異なるシーケンシングデータである、請求項22に記載の方法。

30

【請求項 25】

前記第2の領域または前記第3の領域についての予想参照シーケンシングデータが、バイナリまたは非バイナリプログラムを含む、請求項22から24のいずれか一項に記載の方法。

【請求項 26】

カップリングされたシーケンシングリードペアを参照配列にマッピングする方法であって、

請求項1から25のいずれか一項に記載の方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分または第3の領域もしくはその一部分を、参照配列にマッピングするステップ；および

40

マッピングされなかった第1の領域もしくはその一部分、またはマッピングされなかった第3の領域もしくはその一部分を、前記第2の領域の長さを示す距離情報を使用して、前記参照配列にマッピングするステップを含む、方法。

【請求項 27】

構造バリエーションを検出する方法であって、

請求項1から25のいずれか一項に記載の方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分または第3の領域もしくはその一部分を、参照配列にマッピングするステップ；

マッピングされなかった第1の領域もしくはその一部分またはマッピングされなかった第

50

3の領域もしくはその一部分についての参照配列内の予想遺伝子座を、前記第2の領域の長さを示す距離情報を使用して決定するステップ；

前記参照配列に基づいて前記予想遺伝子座における配列についての予想シーケンシングデータを決定するステップ；および

前記マッピングされなかった第1の領域もしくはその一部分または前記マッピングされなかった第3の領域もしくはその一部分を前記予想シーケンシングデータと比較することにより構造バリエーションを検出するステップであって、前記マッピングされなかった第1の領域もしくはその一部分または前記マッピングされなかった第3の領域もしくはその一部分と前記予想シーケンシングデータとの差が、前記構造バリエーションを示す、ステップを含む方法。

10

【請求項28】

構造バリエーションを検出する方法であって、

請求項1から25のいずれか一項に記載の方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分または第3の領域もしくはその一部分を、参照配列にマッピングするステップを含み、マッピングされなかった第1の領域、またはマッピングされなかった第3の領域が、前記参照配列内にマッピング不可能である、方法。

【請求項29】

構造バリエーションを検出する方法であって、

請求項1から25のいずれか一項に記載の方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分および第3の領域もしくはその一部分を、参照配列にマッピングするステップ；

20

マッピングされた第1の領域とマッピングされた第3の領域の間のマッピング距離情報を決定するステップ；および

前記マッピング距離情報を前記第2の領域の予想距離情報と比較することにより前記構造バリエーションを検出するステップであって、前記マッピング距離情報と前記予想距離情報との差が前記構造バリエーションを示す、ステップを含む方法。

【請求項30】

前記構造バリエーションが、染色体融合、逆位、挿入、または欠失である、請求項27から29のいずれか一項に記載の方法。

30

【請求項31】

前記バリエーションが、前記第2の領域内の挿入または欠失である、請求項27から30のいずれか一項に記載の方法。

【請求項32】

カップリングされたシーケンシングリードペアを参照配列にマッピングする方法であって、

請求項1から25のいずれか一項に記載の方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分および第3の領域もしくはその一部分を、参照配列の第1の位置および第2の位置を含む2つまたはそれより多くの異なる位置ペアにマッピングするステップ；および

40

前記第2の領域の長さを示す第1の距離情報、および前記2つまたはそれより多くの位置ペアについての前記第1の位置と前記第2の位置の間の距離を示す第2の距離情報を使用して、正しい位置を選択するステップを含む方法。

【請求項33】

伸長された前記プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して前記第3の領域を通して伸長される、請求項1から25のいずれか一項に従って生成された、カップリングされたシーケンシングリードペアの2つのシーケンシングされた領域間のバリエーションを検出する方法であって、

50

前記第 1 の領域またはその一部分を参照配列にマッピングするステップ；

(1) 前記第 2 の領域についての参照配列、前記第 2 の領域のフロー順序、前記第 3 の領域のフロー順序、および前記第 3 の領域についての参照配列、または (2) 前記第 2 の領域についての参照配列、前記第 2 の領域のフロー順序、前記第 3 の領域のフロー順序、および前記第 3 の領域の配列に関連する生成シーケンシングデータを使用して、前記第 3 の領域またはその一部分についての予想シーケンシングデータを決定するステップであって、前記第 3 の領域の配列に関連する生成配列データが、前記第 3 の領域について生成された同じまたは異なる配列データである、ステップ；および
前記第 3 の領域についての前記予想シーケンシングデータを前記第 3 の領域の配列に関連する前記生成シーケンシングデータと比較することによりバリエーションの存在を検出するステップを含む方法。

10

【請求項 3 4】

前記バリエーションが、構造バリエーションである、請求項 3 3 に記載の方法。

【請求項 3 5】

前記構造バリエーションが、染色体融合、逆位、挿入、または欠失である、請求項 3 4 に記載の方法。

【請求項 3 6】

前記バリエーションが、一塩基多型 (SNP) である、請求項 3 3 に記載の方法。

【請求項 3 7】

20

試験バリエーションを検出するために使用され、前記参照配列が前記試験バリエーションを含む、請求項 3 3 から 3 6 のいずれか一項に記載の方法。

【請求項 3 8】

前記試験バリエーションが、第 2 のポリヌクレオチド中の試験バリエーション 4 を同定することにより選択される、請求項 3 7 に記載の方法。

【請求項 3 9】

検出された試験バリエーションと前記ポリヌクレオチドの前記第 1 の領域または第 3 の領域におけるシーケンシングされた対立遺伝子とを関連付けるステップを含む、請求項 3 7 または 3 8 に記載の方法。

【請求項 4 0】

30

ポリヌクレオチドのシーケンシングされなかった領域における塩基トランスバージョンの存在を検出するためのカップリングされたシーケンシングリードペアを生成する方法であって、

(a) ポリヌクレオチド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) 前記ポリヌクレオチドの第 1 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；

(c) ステップ (b) において伸長された前記プライマーを、第 2 の領域を通して、(1) シトシンおよびチミンと (2) アデニンおよびグアニンとの交互ヌクレオチドペアを含むフロー順序を使用してさらに伸長するステップ；および

40

(d) 前記ポリヌクレオチドの第 3 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ (c) で伸長された前記プライマーをさらに伸長すること、および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により、生成するステップ

を含む方法。

【請求項 4 1】

カップリングされたシーケンシングリードペアをポリヌクレオチドから生成する方法であって、

(a) プライマーを前記ポリヌクレオチドの第 1 の領域とハイブリダイズしてハイブリダ

50

イズされた鋳型を形成するステップ；

(b) 前記プライマーを、第2の領域を通して、(1) シトシンおよびチミンと(2) アデニンおよびグアニンとの交互ヌクレオチドペアを含むフロー順序を使用して伸長するステップ；および

(c) 前記ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(b)で伸長された前記プライマーをさらに伸長すること、および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により、生成するステップを含む方法。

【請求項42】

前記第1の領域が、前記プライマーの標的にされる天然に存在する配列を含む、請求項41に記載の方法。

【請求項43】

前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、前記第2の領域を通して伸長される、請求項41または42に記載の方法。

【請求項44】

ポリヌクレオチドのシーケンシングされなかった領域における塩基トランスバージョンの存在を検出する方法であって、

前記プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して前記第3の領域を通して伸長される、請求項41から43のいずれか一項に従って生成された、カップリングされたシーケンシングリードペアの第1の領域またはその一部分および第3の領域またはその一部分を、参照配列にマッピングするステップ；

第2の領域のフロー順序、前記第3の領域のフロー順序、および前記参照配列を使用して、前記第3の領域についての予想シーケンシングデータを決定するステップ；および前記第3の領域についての予想シーケンシングデータと前記第3の領域についての生成シーケンシングデータとの差に基づいて前記塩基トランスバージョンの存在を検出するステップを含む方法。

【請求項45】

前記第3の領域についての前記予想シーケンシングデータが、前記第2の領域のフロー順序、前記第3の領域のフロー順序、前記第2の領域についての参照配列、および前記第3の領域についての参照配列を使用して決定される、請求項44に記載の方法。

【請求項46】

前記第3の領域についての前記予想シーケンシングデータが、前記第2の領域のフロー順序、前記第3の領域のフロー順序、前記第2の領域についての参照配列、および前記第3の領域の配列に関連する生成配列データを使用して決定され、前記第3の領域の配列に関連する前記生成配列データが、前記第3の領域について生成された同じまたは異なる配列データである、請求項44に記載の方法。

【請求項47】

前記第3の領域についての前記予想シーケンシングデータが、バイナリまたは非バイナリプログラムを含む、請求項44から46のいずれか一項に記載の方法。

【請求項48】

1つまたは複数のコンセンサス配列を生成する方法であって、請求項1から25のいずれか一項に従って生成された複数のカップリングされたシーケンシングリードペアをアセンブルするステップを含む、方法。

【請求項49】

1つまたは複数のコンセンサス配列から選択されたコンセンサス配列の一部分を、前記選択されたコンセンサス配列の一部分に関連する、選択された、カップリングされたシーケンシングリードを使用して検証するステップをさらに含み、前記選択された、カップリ

10

20

30

40

50

ングされたシーケンシングリードを生成する際に前記第3の領域を通して伸長されるプライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長され、前記検証するステップが、

前記第2の領域のフロー順序、前記第3の領域のフロー順序、および前記選択されたコンセンサス配列の一部分を使用して、前記選択された、カップリングされたシーケンシングリードの第3の領域についての予想シーケンシングデータを決定すること；および前記選択された、カップリングされたシーケンシングリードの前記第3の領域についての前記予想シーケンシングデータを前記第3の領域の前記生成シーケンシングデータと比較することにより、前記選択されたコンセンサス配列の一部分を検証することを含む、請求項48に記載の方法。

10

【請求項50】

試験バリエーションのステータスを検証する方法であって、

請求項1から25のいずれか一項に従って生成された複数の重複するカップリングされたシーケンシングリードペアに亘る前記バリエーションのステータスを比較するステップであって、前記複数の重複するカップリングされたシーケンシングリードペアが、前記試験バリエーションの遺伝子座に対応する遺伝子座を含む、ステップ；

前記比較に基づいて前記バリエーションのステータスを検証するステップを含む方法。

【請求項51】

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、

請求項1から25のいずれか一項に記載のカップリングされたシーケンシングリードペアを生成するステップ；

前記ポリヌクレオチドの第3の領域の配列に関連する前記シーケンシングデータを前記ポリヌクレオチドの前記第3の領域の予想配列についての予想シーケンシングデータと比較するステップ；および

前記ポリヌクレオチドの前記第2の領域における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。

【請求項52】

前記第1の領域の前記配列に関連する前記シーケンシングデータ、または前記第3の領域の前記配列に関連する前記シーケンシングデータが、複数のフロー位置の中の各フロー位置に取り込まれた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、請求項1から51のいずれか一項に記載の方法。

20

30

【請求項53】

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52のいずれか一項に記載の方法。

【請求項54】

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52のいずれか一項に記載の方法。

【請求項55】

カップリングされたシーケンシングリードペアを生成するステップが、

前記プライマーを、第4の領域のフロー順序で提供されるヌクレオチドを使用して第4の領域を通して伸長すること、ここで、(i) 前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、前記第4の領域を通して伸長されるか、(ii) 少なくとも2つの異なるタイプのヌクレオチド塩基の混合物が、前記第4の領域のフロー順序の少なくとも1ステップで使用されるか、または(iii) 前記第4の領域を通じた前記プライマーの伸長が、前記第1の領域または前記第3の領域を通じた前記プライマーの伸長よりも速く進行する；および

前記ポリヌクレオチドの第5の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用する第4を通して伸長された前記プライマーのさらなる伸長、および取

40

50

り込まれた標識ヌクレオチドの存在または非存在の検出により、生成することを含む、請求項 1 から 5 4 のいずれか一項に記載の方法。

【請求項 5 6】

前記第 5 の領域の前記シーケンシングデータを前記第 1 の領域の前記シーケンシングデータまたは前記第 3 の領域の前記シーケンシングデータと関連付けるステップをさらに含む、請求項 5 5 に記載の方法。

【請求項 5 7】

前記ポリヌクレオチドが、ローリングサークル増幅を使用して増幅される、請求項 1 から 5 6 のいずれか一項に記載の方法。

【請求項 5 8】

試験試料中の短い遺伝子バリエーションを検出する方法であって、

(a) ローリングサークル増幅 (RCA) を使用してポリヌクレオチドを増幅させて、前記ポリヌクレオチドの第 1 のコピーと前記ポリヌクレオチドの第 2 のコピーとを少なくとも含む RCA 増幅ポリヌクレオチドを生成するステップ；

(b) 前記 RCA 増幅ポリヌクレオチド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；

(c) 前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの第 1 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用して前記プライマーを伸長すること、および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により、生成するステップ；

(d) 前記プライマーを、第 2 の領域のフロー順序で提供されるヌクレオチドを使用して前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの第 2 の領域を通してさらに伸長するステップであって、(i) 前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域を通して伸長されるか、(ii) 少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、前記第 2 の領域のフロー順序の少なくとも 1 ステップで使用されるか、または (iii) 前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域を通じた前記プライマーの伸長が、前記第 1 の領域を通じた前記プライマーの伸長よりも速く進行する、ステップ；

(e) 前記ポリヌクレオチドの第 3 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーをさらに伸長すること、および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により、生成するステップ；

(f) 前記ポリヌクレオチドの前記第 3 の領域について生成された前記シーケンシングデータを前記ポリヌクレオチドの前記第 3 の領域の予想配列についての予想シーケンシングデータと比較するステップ；

(g) 前記ポリヌクレオチドの前記第 2 の領域における前記短い遺伝子バリエーションの存在をコールするステップ；

(h) 前記ポリヌクレオチドの前記第 2 のコピーの中の前記ポリヌクレオチドの第 2 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用して前記プライマーを伸長すること、および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により、生成するステップ；および

(i) 前記ポリヌクレオチドの前記第 2 の領域における前記短い遺伝子バリエーションのアイデンティティをコールするステップを含む方法。

【請求項 5 9】

前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域を通じた前記プライマーの伸長が、前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 1 の領域を通じた前記プライマーの伸長よりも速く進行する、請求項 5 8 に記載の方法。

10

20

30

40

50

【請求項 60】

試験試料中の短い遺伝子バリエーションを検出する方法であって、

(a) ローリングサークル増幅 (RCA) を使用してポリヌクレオチドを増幅させて、前記ポリヌクレオチドの第 1 のコピーと前記ポリヌクレオチドの第 2 のコピーとを少なくとも含む RCA 増幅ポリヌクレオチドを生成するステップ；

(b) プライマーを前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの第 1 の領域に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；

(c) 前記プライマーを、第 2 の領域のフロー順序で提供されるヌクレオチドを使用して前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの第 2 の領域を通して伸長するステップであって、(i) 前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域を通して伸長される、ステップ；

(d) 前記ポリヌクレオチドの第 3 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーをさらに伸長すること、および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により、生成するステップ；

(e) 前記ポリヌクレオチドの前記第 3 の領域について生成された前記シーケンシングデータを前記ポリヌクレオチドの前記第 3 の領域の予想配列についての予想シーケンシングデータと比較するステップ；

(f) 前記ポリヌクレオチドの前記第 2 の領域における前記短い遺伝子バリエーションの存在をコールするステップ；

(g) 前記ポリヌクレオチドの前記第 2 のコピーの中の前記ポリヌクレオチドの第 2 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用して前記プライマーを伸長すること、および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により、生成するステップ；および

(h) 前記ポリヌクレオチドの前記第 2 の領域における前記短い遺伝子バリエーションのアイデンティティをコールするステップを含む方法。

【請求項 61】

前記第 1 の領域が、前記プライマーの標的にされる天然に存在する配列を含む、請求項 60 に記載の方法。

【請求項 62】

前記ポリヌクレオチドの前記第 2 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域の前記配列に関連する前記シーケンシングデータが、前記ポリヌクレオチドの前記第 2 の領域における前記短い遺伝子バリエーションの存在をコールするステップに基づいて動的に生成される、請求項 58 から 61 のいずれか一項に記載の方法。

【請求項 63】

前記プライマーが、前記伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域を通して伸長される、請求項 58 から 62 のいずれか一項に記載の方法。

【請求項 64】

前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域を通して前記プライマーを伸長するために使用される前記ヌクレオチドの少なくとも一部分が、非標識ヌクレオチドである、請求項 58 から 63 のいずれか一項に記載の方法。

【請求項 65】

前記ポリヌクレオチドの前記第 1 のコピーの中の前記ポリヌクレオチドの前記第 2 の領域を通して前記プライマーを伸長するために使用される前記ヌクレオチドが、非標識ヌクレオチドである、請求項 58 から 64 のいずれか一項に記載の方法。

【請求項 66】

シーケンシングクラスター内のシーケンシングプライマーを同期化する方法であって、

(a) プライマーをシーケンシングクラスター内のポリヌクレオチドコピーとハイブリダイズするステップ；

(b) 前記プライマーを、第1の領域のフローサイクルに従って標識ヌクレオチドを使用して前記ポリヌクレオチドコピーの第1の領域を通して伸長するステップ；

(c) 前記プライマーを、1つまたは複数の再位相化フローを使用して前記ポリヌクレオチドコピーの第2の領域を通して伸長するステップであって、少なくとも2つの異なるタイプのヌクレオチドの混合物が、前記1つまたは複数の再位相化フローの各々において使用される、ステップ；および

(d) 前記プライマーを、第3の領域のフローサイクルに従って標識ヌクレオチドを使用して前記ポリヌクレオチドコピーの第3の領域を通して伸長するステップを含む方法。

【請求項67】

3つの異なるタイプのヌクレオチド塩基の混合物が、前記1つまたは複数の再位相化フローのうちの少なくとも1つにおいて使用される、請求項66に記載の方法。

【請求項68】

前記1つまたは複数の再位相化フローが、4つまたはそれより多くのフローステップを含む、請求項66または67に記載の方法。

【請求項69】

前記1つまたは複数の再位相化フローが、任意の順序で：

(i) A、CおよびGヌクレオチドを含むがTヌクレオチドを含まない混合物を含む第1のフロー；

(ii) T、CおよびGヌクレオチドを含むがAヌクレオチドを含まない混合物を含む第2のフロー；

(iii) T、AおよびGヌクレオチドを含むがCヌクレオチドを含まない混合物を含む第3のフロー；および

(iv) T、AおよびCヌクレオチドを含むがGヌクレオチドを含まない混合物を含む第4のフロー

を含む、請求項68に記載の方法。

【請求項70】

前記第1の領域の配列に関連するシーケンシングデータを、前記プライマーを前記第1の領域を通して伸長しながら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により生成するステップを含む、請求項66から69のいずれか一項に記載の方法。

【請求項71】

前記第3の領域の配列に関連するシーケンシングデータを、前記プライマーを前記第3の領域を通して伸長しながら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により生成するステップを含む、請求項66から70のいずれか一項に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

本願は、2019年3月3日に提出した米国特許仮出願第62/842,534号、2019年9月23日に提出した米国特許仮出願第62/904,274号および2020年2月7日に提出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】

A S C I Iテキストファイルでの配列表の提出

A S C I I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（C R F）の配列表（ファイル名：1 6 5 2 7 2 0 0 0 4 4 0 S E Q L I S T . T X T、記録日：2 0 2 0 年 4 月 2 7 日、サイズ：5 K B）。

【0 0 0 3】

発明の分野

カップリングされたシークエンシングリードペアを生成するための方法、およびシークエンシング法から得られたシークエンシングデータを解析する方法を含む、ポリヌクレオチドをシークエンシングする方法が、本明細書で開示される。

【背景技術】

10

【0 0 0 4】

背景

ペアエンドシークエンシング法は、ポリヌクレオチド分子の3'および5'末端についてのシークエンシングデータを得るために使用されている。一般に、シークエンシングプライマーが、シークエンシングすべきDNAポリヌクレオチドとハイブリダイズされ、いくつかの塩基がシークエンシングされて、ポリヌクレオチドの第1の末端についてのシークエンシングデータが得られる。次いで、第2のシークエンシングプライマーが、ポリヌクレオチドの他端付近の相補鎖とハイブリダイズされ、シークエンシングされて、ポリヌクレオチドの他端のシークエンシングデータが決定される。ポリヌクレオチドの3'および5'末端についてのシークエンシングデータは、シークエンシングデータが同じシークエンシングクラスターから得られたという事実に基づいて、カップリングされる。ペアエンドシークエンシング法は、次世代シークエンシング（N G S）プロトコルで使用されることが多い。

20

【0 0 0 5】

しかし、旧来のペアエンドシークエンシングを使用しても、ポリヌクレオチドの3'末端と5'末端の間の領域に関する情報は得られない（またはほとんど得られない）。ある特定の解析を目的としてペアエンドシークエンシングデータを使用することができるが、ポリヌクレオチドのシークエンシングされなかった領域における特定のバリエーションを検出するためにペアエンドシークエンシングデータを使用することはできない。旧来のペアエンドシークエンシング法を使用して一般に見逃されるポリヌクレオチドの領域をシークエンシングするために、ある特定のロングレンジシークエンシング技法が開発された。しかし、ロングレンジシークエンシングは、比較的遅く、かなりのシークエンシングエラーを起こしやすい。

30

【発明の概要】

【課題を解決するための手段】

【0 0 0 6】

発明の簡単な要旨

カップリングされたシークエンシングリードペアを生成するための方法、およびシークエンシング法から得られたシークエンシングデータを解析する方法を含む、ポリヌクレオチドをシークエンシングする方法が、本明細書で開示される。

40

【0 0 0 7】

カップリングされたシークエンシングリードペアをポリヌクレオチドから生成するための方法は、（a）ポリヌクレオチド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；（b）ポリヌクレオチドの第1の領域の配列に関連するシークエンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；（c）ステップ（b）で伸長されたプライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用して第2の領域を通してさらに伸長するステップであって、（i）プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第2の領域を通して伸長されるか、（i i）少なくとも2つ

50

の異なるタイプのヌクレオチド塩基の混合物が、第2の領域のフロー順序の少なくとも1ステップで使用されるか、または(iii)第2の領域を通したプライマーの伸長が、ステップ(b)におけるプライマーの伸長よりも速く進行する、ステップ；および(d)ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(c)で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む。一部の実施形態では、第2の領域を通したプライマーの伸長が、ステップ(b)におけるプライマーの伸長よりも速く進行する。一部の実施形態では、カップリングされたシーケンシングリードペアを生成する方法は、第1の領域のシーケンシングデータを第3のシーケンシングデータと関連付けるステップを含む。

10

【0008】

一部の実施形態では、カップリングされたシーケンシングリードペアをポリヌクレオチドから生成する方法は、(a)プライマーをポリヌクレオチドの第1の領域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；(b)プライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用して第2の領域を通して伸長するステップであって、(i)プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第2の領域を通して伸長される、または(ii)少なくとも2つの異なるタイプのヌクレオチド塩基の混合物が、第2の領域のフロー順序の少なくとも1ステップで使用される、ステップ；および(c)ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(b)で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む。一部の実施形態では、第1の領域が、プライマーの標的にされる天然に存在する配列を含む。

20

【0009】

一部の実施形態では、プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、第2の領域を通して伸長される。一部の実施形態では、第2の領域を通してプライマーを伸長するために使用されるヌクレオチドの少なくとも一部分は、非標識ヌクレオチドである。一部の実施形態では、第2の領域を通してプライマーを伸長するために使用されるヌクレオチドは、非標識ヌクレオチドである。

【0010】

一部の実施形態では、少なくとも2つの異なるタイプのヌクレオチド塩基の混合物は、第2の領域のフロー順序の少なくとも1ステップで使用される。

30

【0011】

カップリングされたシーケンシングリードペアを生成する方法の一部の実施形態では、第2の領域のフロー順序は、5つまたはそれより多くのヌクレオチドフローを含む。一部の実施形態では、ヌクレオチドフローの各々は、単一のヌクレオチド塩基を含む。一部の実施形態では、第2の領域のフロー順序は、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、第2の領域のフロー順序は、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。

40

【0012】

一部の実施形態では、カップリングされたシーケンシングリードペアを生成する方法は、参照配列および第2の領域のフロー順序を使用して第2の領域についての予想シーケンシングデータを決定するステップを含む。一部の実施形態では、プライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長され、方

50

法は、第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列を使用して、第3の領域についての予想シーケンシングデータを決定するステップをさらに含む。一部の実施形態では、第3の領域のフロー順序は、5つまたはそれより多くのヌクレオチドフローを含む。一部の実施形態では、ヌクレオチドフローの各々は、単一のヌクレオチド塩基を含む。一部の実施形態では、第3の領域のフロー順序は、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、第3の領域のフロー順序は、1フロー当たり0.6のまたはそれを超える塩基組込みの効率を有する。

10

【0013】

カップリングされたシーケンシングリードペアを生成する方法の一部の実施形態では、プライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長され、方法は、第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域の配列に関連するシーケンシングデータを使用して、第3の領域についての予想シーケンシングデータを決定するステップをさらに含み、第3の領域の配列に関連するシーケンシングデータは、第3の領域について生成される同じまたは異なるシーケンシングデータである。一部の実施形態では、第2の領域または第3の領域についての予想参照データは、バイナリまたは非バイナリフローグラムを含む。一部の実施形態では、方法は、第2の領域のフロー順序、および第2の領域についての第2の参照配列を使用して、第2の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、第2の参照配列は、試験バリエーションを含む。一部の実施形態では、プライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長され、方法は、第2の領域についての第2の参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列を使用して、第3の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、第3の領域の配列に関連するシーケンシングデータは、第3の領域について生成される同じまたは異なるシーケンシングデータである。一部の実施形態では、第2の領域または第3の領域についての予想参照シーケンシングデータは、バイナリまたは非バイナリフローグラムを含む。

20

30

【0014】

一部の実施形態では、カップリングされたシーケンシングリードペアを生成する方法は、参照配列および第2の領域のフロー順序を使用して第2の領域についての予想シーケンシングデータを決定するステップを含む。一部の実施形態では、ステップ(d)で伸長されるプライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長され、方法は、第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列を使用して、第3の領域についての予想シーケンシングデータを決定するステップをさらに含む。一部の実施形態では、ステップ(d)で伸長されるプライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長され、方法は、第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域の配列に関連するシーケンシングデー

40

50

タを使用して、第3の領域についての予想シーケンシングデータを決定するステップをさらに含み、第3の領域の配列に関連するシーケンシングデータは、ステップ(d)で生成される同じまたは異なるシーケンシングデータである。一部の実施形態では、第2の領域または第3の領域についての予想参照データは、バイナリまたは非バイナリフローグラムを含む。一部の実施形態では、方法は、第2の領域のフロー順序、および第2の領域についての第2の参照配列を使用して、第2の領域についての予想試験バリエーションシーケンシングデータを決定するステップを含み、第2の参照配列は、試験バリエーションを含む。一部の実施形態では、ステップ(d)で伸長されるプライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長され、方法は、第2の領域についての第2の参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列を使用して、第3の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含む。一部の実施形態では、ステップ(d)で伸長されるプライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長され、方法は、第2の領域についての第2の参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域の配列に関連するシーケンシングデータを使用して、第3の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、第3の領域の配列に関連するシーケンシングデータは、ステップ(d)で生成される同じまたは異なるシーケンシングデータである。一部の実施形態では、第2の領域または第3の領域についての予想参照シーケンシングデータは、バイナリまたは非バイナリフローグラムを含む。

10

20

【0015】

実施形態では、カップリングされたシーケンシングリードペアを生成するステップは、(e)ステップ(d)で伸張したプライマーを、第4の領域のフロー順序で提供されるヌクレオチドを使用して第4の領域を通して伸長すること、ここで、(i)プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第4の領域を通して伸長されるか、(ii)少なくとも2つの異なるタイプのヌクレオチド塩基の混合物は、第4の領域のフロー順序の少なくとも1ステップで使用されるか、または(iii)第4の領域を通したプライマーの伸長は、ステップ(b)またはステップ(d)においてプライマーの伸長よりも速く進行する；および(f)ポリヌクレオチドの第5の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用する(e)において伸長されたプライマーのさらなる伸長、および取り込まれた標識ヌクレオチドの存在または非存在の検出により、生成することを含む。一部の実施形態では、方法は、第5の領域のシーケンシングデータを第1の領域のシーケンシングデータまたは第3の領域のシーケンシングデータと関連付けるステップをさらに含む。

30

【0016】

カップリングされたシーケンシングリードペアを参照配列にマッピングする方法であって、カップリングされたシーケンシングリードの第1の領域もしくはその一部分または第3の領域もしくはその一部分を、参照配列にマッピングするステップ；およびマッピングされなかった第1の領域もしくはその一部分、またはマッピングされなかった第3の領域もしくはその一部分を、第2の領域の長さを示す距離情報を使用して、参照配列にマッピングするステップを含む、方法も本明細書に記載される。

40

【0017】

さらに、構造バリエーションを検出する方法であって、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分または第3の領域もしくはその一部分を、参照配列にマッピングするステップ；マッピングされなかった第1の領域もしくはその一部分またはマッピングされなかった第3の領域もしくはその一部分についての参照配列内の予想遺伝子座を、第2の領域の長さを示す距離情報を使用して決定するステップ；参照配列に基づいて予想遺伝子座における配列についての予想シーケンシングデータを決定するステップ；およびマッピングされなかった第1の領域もしくはその一部分またはマッピングされなかった第3の領域もしくはその一部分を予想シーケンシングデータと比較す

50

ることにより構造バリエーションを検出するステップであって、マッピングされなかった第1の領域もしくはその一部分またはマッピングされなかった第3の領域もしくはその一部分と予想シーケンシングデータとの差は、構造バリエーションを示す、ステップを含む方法が提供される。

【0018】

また、構造バリエーションを検出する方法であって、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分または第3の領域もしくはその一部分を、参照配列にマッピングするステップを含み、マッピングされなかった第1の領域、またはマッピングされなかった第3の領域は、参照配列内にマッピング不可能である、方法が本明細書に提供される。一部の実施形態では、方法は、第2の領域の長さを示す予想距離情報に基づいて参照配列内の構造バリエーションの遺伝子座を決定するステップをさらに含む。

10

【0019】

一部の実施形態では、マッピングされなかった第1の領域もしくはその一部分、またはマッピングされなかった第3の領域もしくはその一部分は、参照配列と比較して挿入の中に存在する。一部の実施形態では、マッピングされなかった第1の領域もしくはその一部分、またはマッピングされなかった第3の領域もしくはその一部分は、参照配列と比較して挿入の始点または終点にまたがる。

【0020】

さらに、構造バリエーションを検出する方法であって、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分および第3の領域もしくはその一部分を、参照配列にマッピングするステップ；マッピングされた第1の領域とマッピングされた第3の領域の間のマッピング距離情報を決定するステップ；およびマッピング距離情報を第2の領域の予想距離情報と比較することにより構造バリエーションを検出するステップであって、マッピング距離情報と予想距離情報との差が構造バリエーションを示す、ステップを含む方法が本明細書に提供される。一部の実施形態では、構造バリエーションは、染色体融合、逆位、挿入、または欠失である。一部の実施形態では、バリエーションは、第2の領域内の挿入または欠失である。

20

【0021】

本明細書に記載の方法の一部の実施形態では、距離情報は、第2の領域のフロー順序に関連する情報および第2の領域における塩基の確率分布を使用して決定される。一部の実施形態では、第2の領域のフロー順序に関連する情報は、ステップ(c)においてプライマーを伸長するために同時に使用されるヌクレオチド塩基の異なるタイプの数である。一部の実施形態では、第2の領域における塩基の確率分布は、ゲノム内の塩基の分布から決定される。

30

【0022】

本明細書に記載の方法の一部の実施形態では、距離情報は、参照配列および第2の領域のフロー順序を使用して決定された第2の領域についての予想シーケンシングデータから導出される。一部の実施形態では、予想シーケンシングデータは、バイナリまたは非バイナリフローグラムを含む。

【0023】

さらに、カップリングされたシーケンシングリードペアを参照配列にマッピングする方法であって、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分および第3の領域もしくはその一部分を、参照配列の第1の位置および第2の位置を含む2つまたはそれより多くの異なる位置ペアにマッピングするステップ；および第2の領域の長さを示す第1の距離情報、および2つまたはそれより多くの位置ペアについての第1の位置と第2の位置の間の距離を示す第2の距離情報を使用して、正しい位置を選択するステップを含む方法が本明細書に記載される。一部の実施形態では、距離情報は、第2の領域のフロー順序に関連する情報および第2の領域における塩基の確率分布を使用して決定される。一部の実施形態では、第2の領域のフロー順序に関連する情報は、ステップ(c)においてプライマーを伸長するために同時に使用されるヌクレオチド塩基の異

40

50

なるタイプの数である。一部の実施形態では、第2の領域における塩基の確率分布は、ゲノム内の塩基の分布から決定される。一部の実施形態では、距離情報は、参照配列および第2の領域のフロー順序を使用して決定された第2の領域についての予想シーケンシングデータから導出される。一部の実施形態では、予想参照シーケンシングデータは、バイナリまたは非バイナリフロープログラムを含む。

【0024】

また、ステップ(d)において伸長されたプライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長される、上記方法のいずれかに従って生成された、カップリングされたシーケンシングリードペアの2つのシーケンシングされた領域間のバリエーションを検出する方法であって、第1の領域またはその一部分を参照配列にマッピングするステップ；(1)第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列、または(2)第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域の配列に関連する生成シーケンシングデータを使用して、第3の領域またはその一部分についての予想シーケンシングデータを決定するステップであって、第3の領域の配列に関連する生成配列データが、ステップ(d)において生成された同じまたは異なる配列データである、ステップ；および第3の領域についての予想シーケンシングデータを第3の領域の配列に関連する生成シーケンシングデータと比較することによりバリエーションの存在を検出するステップを含む方法が本明細書に記載される。一部の実施形態では、バリエーションは、構造バリエーションである。一部の実施形態では、構造バリエーションは、染色体融合、逆位、挿入、または欠失である。一部の実施形態では、バリエーションは、一塩基多型(SNP)である。一部の実施形態では、試験バリエーションを検出するために使用され、参照配列が試験バリエーションを含む。一部の実施形態では、試験バリエーションは、第2のポリヌクレオチド中の試験バリエーションを同定することにより選択される。一部の実施形態では、方法は、検出された試験バリエーションとポリヌクレオチドの第1の領域または第3の領域におけるシーケンシングされた対立遺伝子とを関連付けるステップをさらに含む。

【0025】

また、第3の領域を通して伸長されたプライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長される、上記の方法のいずれかに従って生成された、カップリングされたシーケンシングリードペアの2つのシーケンシングされた領域間のバリエーションを検出する方法であって、第1の領域またはその一部分を参照配列にマッピングするステップ；(1)第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列、または(2)第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域の配列に関連する生成シーケンシングデータを使用して、第3の領域またはその一部分についての予想シーケンシングデータを決定するステップであって、第3の領域の配列に関連する生成配列データが、第3の領域について生成された同じまたは異なる配列データである、ステップ；および第3の領域についての予想シーケンシングデータを第3の領域の配列に関連する生成シーケンシングデータと比較することによりバリエーションの存在を検出するステップを含む方法が本明細書に記載される。一部の実施形態では、バリエーションは、構造バリエーションである。一部の実施形態では、構造バリエーションは、染色体融合、逆位、挿入、または欠失である。一部の実施形態では、バリエーションは、一塩基多型(SNP)である。一部の実施形態では、試験バリエーションを検出するために使用され、参照配列が試験バリエーションを含む。一部の実施形態では、試験バリエーションは、第2のポリヌクレオチド中の試験バリエーションを同定することにより選択される。一部の実施形態では、方法は、検出された試験バリエーションとポリヌクレオチドの第1の領域または第3の領域におけるシーケンシングされた対立遺伝子とを関連付けるステップを含む。

【0026】

さらに、ポリヌクレオチドのシーケンシングされなかった領域における塩基トランスバージョンの存在を検出するためのカップリングされたシーケンシングリードペアを生成

する方法であって、(a) ポリヌクレオチド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；(b) ポリヌクレオチドの第1の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；(c) ステップ(b)において伸長されたプライマーを、第2の領域を通して、(1) シトシンおよびチミンと(2) アデニンおよびグアニンとの交互ヌクレオチドペアを含むフロー順序を使用してさらに伸長するステップ；および(d) ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(c)で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む方法が本明細書に記載される。一部の実施形態では、プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、第2の領域を通して伸長される。

10

【0027】

また、カップリングされたシーケンシングリードペアをポリヌクレオチドから生成する方法であって、(a) プライマーをポリヌクレオチドの第1の領域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；(b) プライマーを、第2の領域を通して、(1) シトシンおよびチミンと(2) アデニンおよびグアニンとの交互ヌクレオチドペアを含むフロー順序を使用して伸長するステップ；および(c) ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(b)で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む方法が本明細書に記載される。一部の実施形態では、第1の領域は、プライマーの標的にされる天然に存在する配列を含む。一部の実施形態では、プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、第2の領域を通して伸長される。

20

【0028】

一部の実施形態では、ポリヌクレオチドのシーケンシングされなかった領域における塩基トランスバージョンの存在を検出する方法は、ステップ(d)で伸長されたプライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長される、上記の方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域またはその一部分および第3の領域またはその一部分を、参照配列にマッピングするステップ；第2の領域のフロー順序、第3の領域のフロー順序、および参照配列を使用して、第3の領域についての予想シーケンシングデータを決定するステップ；および第3の領域についての予想シーケンシングデータと第3の領域についての生成シーケンシングデータとの差に基づいて塩基トランスバージョンの存在を検出するステップを含む。一部の実施形態では、第3の領域についての予想シーケンシングデータは、第2の領域のフロー順序、第3の領域のフロー順序、第2の領域についての参照配列、および第3の領域についての参照配列を使用して決定される。一部の実施形態では、第3の領域についての予想シーケンシングデータは、第2の領域のフロー順序、第3の領域のフロー順序、第2の領域についての参照配列、および第3の領域の配列に関連する生成配列データを使用して決定され、第3の領域の配列に関連する生成配列データは、ステップ(d)で生成された同じまたは異なる配列データである。一部の実施形態では、第3の領域についての予想シーケンシングデータは、バイナリまたは非バイナリフローグラムを含む。

30

40

【0029】

さらに、1つまたは複数のコンセンサス配列を生成する方法であって、複数のカップリングされたシーケンシングリードペアをアSEMBルするステップを含む、方法が本明細書に記載される。一部の実施形態では、1つまたは複数のコンセンサス配列は、複数のカップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用してアSEMBルされる。一部の実施形態では、距離情報は、第2の領域のフロー順序に関連

50

する情報および第2の領域における塩基の確率分布を使用して決定される。一部の実施形態では、第2の領域のフロー順序に関連する情報は、ステップ(c)においてプライマーを伸長するために同時に使用されるヌクレオチド塩基の異なるタイプの数である。一部の実施形態では、第2の領域における塩基の確率分布は、ゲノム内の塩基の分布から決定される。一部の実施形態では、距離情報は、参照配列および第2の領域のフロー順序を使用して決定された第2の領域についての予想参照シーケンシングデータから導出される。一部の実施形態では、予想参照シーケンシングデータは、バイナリまたは非バイナリフローグラムを含む。

【0030】

一部の実施形態では、1つまたは複数のコンセンサス配列を生成する方法は、1つまたは複数のコンセンサス配列から選択されたコンセンサス配列の一部分を、選択されたコンセンサス配列の一部分に関連する、選択された、カップリングされたシーケンシングリードを使用して検証するステップをさらに含み、選択された、カップリングされたシーケンシングリードを生成する際にステップ(d)で伸長されるプライマーは、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長され、検証するステップは、第2の領域のフロー順序、第3の領域のフロー順序、および選択されたコンセンサス配列の一部分を使用して、選択された、カップリングされたシーケンシングリードの第3の領域についての予想シーケンシングデータを決定すること；および選択された、カップリングされたシーケンシングリードの第3の領域についての予想シーケンシングデータを第3の領域の生成シーケンシングデータと比較することにより、選択されたコンセンサス配列の一部分を検証することを含む。

【0031】

また、試験バリエーションのステータスを検証する方法であって、複数の重複するカップリングされたシーケンシングリードペアに亘るバリエーションのステータスを比較するステップであって、複数の重複するカップリングされたシーケンシングリードペアは、試験バリエーションの遺伝子座に対応する遺伝子座を含む、ステップ；比較に基づいてバリエーションのステータスを検証するステップを含む方法が記載される。一部の実施形態では、選択された、カップリングされたシーケンシングリードペアの、第1の領域または第3の領域は、複数の重複するカップリングされたシーケンシングリードのうちの他のカップリングされたシーケンシングリードの少なくとも一部分の第2の領域と重複する。一部の実施形態では、選択された、カップリングされたシーケンシングリードの、バリエーションステータスは、選択された、カップリングされたシーケンシングリードの、第1の領域または第3の領域におけるバリエーションを示す。一部の実施形態では、選択された、カップリングされたシーケンシングリードの、第2の領域は、複数の重複するカップリングされたシーケンシングリードのうちの他のカップリングされたシーケンシングリードの少なくとも一部分の第2の領域と重複する。一部の実施形態では、選択された、カップリングされたシーケンシングリードの、バリエーションステータスは、選択された、カップリングされたシーケンシングリードの、第2の領域におけるバリエーションを示す。

【0032】

さらに、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、上記方法のいずれかに従ってのカップリングされたシーケンシングリードペアを生成するステップ；ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータをポリヌクレオチドの第3の領域の予想配列についての予想シーケンシングデータと比較するステップ；およびポリヌクレオチドの第2の領域における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が本明細書に記載される。一部の実施形態では、ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータをポリヌクレオチドの第3の領域についての予想シーケンシングデータと比較するステップは、ポリヌクレオチドの第3の領域について生成されたシーケンシングデータがポリヌクレオチドの第3の領域についての予想シーケンシングデータにマッチする尤度を示すマッチスコアを決定すること；および決定されたマッチスコアを使用してポリヌクレオチドの第2の領

域における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を含む。一部の実施形態では、ポリヌクレオチドの第3の領域についての予想シーケンシングデータは、ポリヌクレオチドの第3の領域の予想配列を *in silico* でシーケンシングすることにより得られる。一部の実施形態では、第1の領域の配列に関連するシーケンシングデータ、または第3の領域の配列に関連するシーケンシングデータは、複数のフロー位置の中の各フロー位置に取り込まれた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。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つの塩基カウントについての塩基カウント尤度を示す統計パラメータを含む。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複数の塩基カウントについての塩基カウント尤度を示す統計パラメータを含む。一部の実施形態では、第3の領域の配列に関連するシーケンシングデータは、複数のフロー位置の中の各フロー位置に取り込まれた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み、フローシグナルは、複数の塩基カウントについての塩基カウント尤度を示す統計パラメータを含み；方法は、シーケンシングデータ中の各フロー位置における統計パラメータであって、そのフロー位置における予想配列の塩基カウントと一致する統計パラメータを選択するステップ、およびシーケンシングデータセットが予想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をさらに含む。一部の実施形態では、マッチスコアは、シーケンシングデータ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の組み合わせられた値である。

10

【0033】

20

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。

【0034】

上記方法の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。

【0035】

上記方法の一部の実施形態では、カップリングされたシーケンシングリードペアを生成するステップは、プライマーを、第4の領域のフロー順序で提供されるヌクレオチドを使用して第4の領域を通して伸長すること、ここで、(i) プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第4の領域を通して伸長されるか、(ii) 少なくとも2つの異なるタイプのヌクレオチド塩基の混合物は、第4の領域のフロー順序の少なくとも1ステップで使用されるか、または(iii) 第4の領域を通したプライマーの伸長は、第1の領域または第3の領域を通したプライマーの伸長よりも速く進行する；およびポリヌクレオチドの第5の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用する第4を通して伸長されたプライマーのさらなる伸長、および取り込まれた標識ヌクレオチドの存在または非存在の検出により、生成することを含む。一部の実施形態では、方法は、第5の領域のシーケンシングデータを第1の領域のシーケンシングデータまたは第3の領域のシーケンシングデータと関連付けるステップをさらに含む。

30

【0036】

40

上記方法の一部の実施形態では、ポリヌクレオチドは、ローリングサークル増幅を使用して増幅される。

【0037】

また、試験試料中の短い遺伝子バリエーションを検出する方法であって、(a) ローリングサークル増幅(RCA)を使用してポリヌクレオチドを増幅させて、ポリヌクレオチドの第1のコピーとポリヌクレオチドの第2のコピーとを少なくとも含むRCA増幅ポリヌクレオチドを生成するステップ；(b) RCA増幅ポリヌクレオチド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；(c) ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第1の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および取り込まれた標

50

識ヌクレオチドの存在または非存在を検出することにより、生成するステップ；（d）プライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用してポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通してさらに伸長するステップであって、（i）プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長されるか、（ii）少なくとも2つの異なるタイプのヌクレオチド塩基の混合物は、第2の領域のフロー順序の少なくとも1ステップで使用されるか、または（iii）ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通したプライマーの伸長は、第1の領域を通したプライマーの伸長よりも速く進行する、ステップ；（e）ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーをさらに伸長すること、および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により、生成するステップ；（f）ポリヌクレオチドの第3の領域について生成されたシーケンシングデータをポリヌクレオチドの第3の領域の予想配列についての予想シーケンシングデータと比較するステップ；（g）ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップ；（h）ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；および（i）ポリヌクレオチドの第2の領域における短い遺伝子バリエーションのアイデンティティをコールするステップを含む方法が本明細書に記載される。一部の実施形態では、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通したプライマーの伸長は、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第1の領域を通したプライマーの伸長よりも速く進行する。一部の実施形態では、ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータは、ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップに基づいて動的に生成される。一部の実施形態では、プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長される。一部の実施形態では、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通してプライマーを伸長するために使用されるヌクレオチドの少なくとも一部分は、非標識ヌクレオチドである。一部の実施形態では、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通してプライマーを伸長するために使用されるヌクレオチドは、非標識ヌクレオチドである。一部の実施形態では、少なくとも2つの異なるタイプのヌクレオチド塩基の混合物は、第2の領域のフロー順序の少なくとも1ステップで使用される。一部の実施形態では、3つの異なるタイプのヌクレオチド塩基の混合物は、第2の領域のフロー順序の少なくとも1ステップで使用される。

10

20

30

40

50

【0038】

さらに、試験試料中の短い遺伝子バリエーションを検出する方法であって、（a）ローリングサークル増幅（RCA）を使用してポリヌクレオチドを増幅させて、ポリヌクレオチドの第1のコピーとポリヌクレオチドの第2のコピーとを少なくとも含むRCA増幅ポリヌクレオチドを生成するステップ；（b）プライマーをポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第1の領域に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；（c）プライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用してポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長するステップであって、（i）プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長される、ステップ；（d）ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーをさらに伸長すること、および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により、生成するステップ；（e）ポリヌクレオチドの第3の領域について生成されたシーケンシングデータをポリヌクレオチドの第3の領域の予想配列についての予想シーケンシングデータと比較するステップ；（f）ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップ；（g）ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；および（h）ポリヌクレオチドの第2の領域における短い遺伝子バリエーションのアイデンティティをコールするステップを含む方法が本明細書に記載される。一部の実施形態では、第1の領域は、プライマーの標的にされる天然に存在する配列を含む。一部の実施形態では、ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータは、ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップに基づいて動的に生成される。一部の実施形態では、プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長される。一部の実施形態では、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通してプライマーを伸長するために使用されるヌクレオチドは、非標識ヌクレオチドである。一部の実施形態では、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通してプライマーを伸長するために使用されるヌクレオチドは、非標識ヌクレオチドである。一部の実施形態では、少なくとも2つの異なるタイプのヌクレオチド塩基の混合物は、第2の領域のフロー順序の少なくとも1ステップで使用される。一部の実施形態では、3つの異なるタイプのヌクレオチド塩基の混合物は、第2の領域のフロー順序の少なくとも1ステップで使用される。

10

20

30

40

50

【0039】

また、シーケンシングクラスター内のシーケンシングプライマーを同期化する方法であって、（a）プライマーをシーケンシングクラスター内のポリヌクレオチドコピーとハイブリダイズするステップ；（b）プライマーを、第1の領域のフローサイクルに従って標識ヌクレオチドを使用してポリヌクレオチドコピーの第1の領域を通して伸長するステップ；（c）プライマーを、1つまたは複数の再位相化フローを使用してポリヌクレオチドコピーの第2の領域を通して伸長するステップであって、少なくとも2つの異なるタイプのヌクレオチドの混合物は、1つまたは複数の再位相化フローの少なくとも1つにおいて使用される、ステップ；および（d）プライマーを、第3の領域のフローサイクルに従って標識ヌクレオチドを使用してポリヌクレオチドコピーの第3の領域を通して伸長するステップを含む方法が本明細書に記載される。一部の実施形態では、3つの異なるタイプのヌクレオチド塩基の混合物は、1つまたは複数の再位相化フローのうちの少なくとも1つにおいて使用される。一部の実施形態では、1つまたは複数の再位相化フローは、4つまたはそれより多くのフローステップを含む。一部の実施形態では、1つまたは複数の再位相化フローは、任意の順序で：（i）A、CおよびGヌクレオチドを含むがTヌクレオチドを含まない混合物を含む第1のフロー；（ii）T、CおよびGヌクレオチドを含むがAヌクレオチドを含まない混合物を含む第2のフロー；（iii）T、AおよびGヌクレオチドを含むがCヌクレオチドを含まない混合物を含む第3のフロー；および（iv）T、AおよびCヌクレオチドを含むがGヌクレオチドを含まない混合物を含む第4のフローを含む。一部の実施形態では、方法は、第1の領域の配列に関連するシーケンシングデータを、プライマーを第1の領域を通して伸長しながら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により生成するステップを含む。一部の実施形態では、方法は、第3の領域の配列に関連するシーケンシングデータを、プライマーを第3の領域を通して伸長しながら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により生成するステップを含む。

【0040】

また、1または複数台のプロセッサと、非一過性記憶媒体であって、1つまたは複数の

カップリングされたシーケンシングリードに関する情報を受信する、および上記方法のいずれかの1つまたは複数を遂行するための、1または複数台のプロセッサにより実行可能な1つまたは複数のプログラムを含む非一過性記憶媒体とを含むシステムが本明細書に記載される。

【図面の簡単な説明】

【0041】

【図1】図1は、カップリングされたシーケンシングリードペアをポリヌクレオチドから生成するための例示的方法の概略図を示す。

【0042】

【図2】図2は、参照配列を使用して予想シーケンシングデータを生成するための例示的方法の概略図を示す。

【0043】

【図3】図3は、カップリングされたシーケンシングリードペアの第3の領域が2つの異なる遺伝子座に位置する場合、カップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用して、カップリングされたシーケンシングリードペアを参照配列にマッピングする方法を示す。

【0044】

【図4】図4は、カップリングされたシーケンシングリードペアの第3の領域が反復領域に位置する場合、カップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用して、カップリングされたシーケンシングリードペアを参照配列にマッピングする方法を示す。

【0045】

【図5】図5は、カップリングされたシーケンシングリードペアを使用して対象のゲノム内への挿入を検出する方法についての概略図を示す。

【0046】

【図6】図6は、カップリングされたシーケンシングリードペアを使用して構造バリエーションを検出するための例示的方法を示す。

【0047】

【図7】図7は、対象のゲノムにおける構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは挿入である。

【0048】

【図8】図8は、対象のゲノムにおける構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは欠失である。

【0049】

【図9】図9は、対象のゲノムにおける構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは逆位である。

【0050】

【図10】図10は、対象のゲノムにおける構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは染色体融合である。

【0051】

【図11】図11は、カップリングされたシーケンシングリードペアを使用して構造バリエーションを検出する例示的方法を示す。

【0052】

【図12】図12は、カップリングされたシーケンシングリードペアを使用して、カップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用して構造バリエーションを検出することができる方法の一例を実証する概略図を示す。

10

20

30

40

50

【0053】

【図13】図13は、カップリングされたシークエンシングリードペアの2つのシークエンシングされた領域間のバリエーションを検出する例示的方法を示す。

【0054】

【図14A】図14Aは、T-A-C-Gの反復フローサイクル順序を使用して5'-TATGGTCGTCGA-3'（配列番号15）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クエンシングデータを示す。このシークエンシングデータは、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クエンシング情報は、実効的に等価である。

【図14B】図14B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、シークエ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14Aに示されているシークエンシングデータを示す。

【図14C】図14Cは、2つの異なる候補配列（それらの相補配列により各々表される）：TATGGTCATCGA（配列番号16）（黒塗りの丸印）およびTATGGTCGTCGA（配列番号15）（白抜きの丸印）を表すトレースを伴う、図14Aに示されているシークエンシングデータを示す。シークエ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。

【0055】

【図15-1】図15Aは、2つの候補配列H1（配列番号19）およびH2（配列番号20）（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、シークエンシングリードR1（配列番号15）、R2（配列番号17）およびR3（配列番号18）（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15Bは、H1（黒塗りの丸印）およびH2（白抜きの丸印）を表すトレースを伴う、R1に対応するシークエンシングデータを示す。図15Cは、H1（黒塗りの丸印）およびH2（白抜きの丸印）を表すトレースを伴う、R2に対応するシークエンシングデータを示す。図15Dは、H1（黒塗りの丸印）およびH2（白抜きの丸印）を表すトレースを伴う、R3に対応するシークエンシングデータを示す。

【図15-2】図15Aは、2つの候補配列H1（配列番号19）およびH2（配列番号20）（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、シークエンシングリードR1（配列番号15）、R2（配列番号17）およびR3（配列番号18）（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15Bは、H1（黒塗りの丸印）およびH2（白抜きの丸印）を表すトレースを伴う、R1に対応するシークエンシングデータを示す。図15Cは、H1（黒塗りの丸印）およびH2（白抜きの丸印）を表すトレースを伴う、R2に対応するシークエンシングデータを示す。図15Dは、H1（黒塗りの丸印）およびH2（白抜きの丸印）を表すトレースを伴う、R3に対応するシークエンシングデータを示す。

【0056】

【図16】図16は、A-T-G-Cフローサイクル順序を使用してシークエンシングされた仮想核酸分子からのシークエンシングデータを示す。可能性のあるハプロタイプ配列（それらの相補配列により各々表される）TATGGTCG-TCGA（配列番号21）（H1）およびTATGGTCGATCG（配列番号22）（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。シークエンシングデータは、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。

【0057】

【図17】図17は、試験バリエーションのステータスを決定するためのカップリングされたシークエンシングリードペアの比較についての例示的概略図を示す。

【0058】

【図18】図18は、本明細書に記載の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示す。

10

20

30

40

50

【0059】

【図19A】図19Aは、ポリヌクレオチドを通したシーケンシングプライマーの伸長時に第1および第3の領域において各々のフローシーケンシングサイクル後に取り込まれた塩基から生じるシグナルを示す。第2の領域内のデータは、この領域を通してプライマー伸長が加速され、塩基組込みが検出されないため、収集されなかった。

【0060】

【図19B】図19Bは、ポリヌクレオチドを通したシーケンシングプライマーの伸長時に第1および第3の領域において各々のフローシーケンシングサイクル後に取り込まれた塩基から生じるシグナルを示す。第2の領域を通してデータは収集されたが、図のサイズを縮小するために示されていない。

10

【0061】

【図20-1】図20A~20Eは、100ヌクレオチドフロー（図20A）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フロー順序は、ヌクレオチドフロー101（図20B）、フロー102（図20C）、フロー103（図20D）およびフロー104（図20E）を含む、4ステップ順序である。

【図20-2】図20A~20Eは、100ヌクレオチドフロー（図20A）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フロー順序は、ヌクレオチドフロー101（図20B）、フロー102（図20C）、フロー103（図20D）およびフロー104（図20E）を含む、4ステップ順序である。

20

【図20-3】図20A~20Eは、100ヌクレオチドフロー（図20A）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フロー順序は、ヌクレオチドフロー101（図20B）、フロー102（図20C）、フロー103（図20D）およびフロー104（図20E）を含む、4ステップ順序である。

【0062】

30

【図21-1】図21A~21Eは、100ヌクレオチドフロー（図21A）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、別の例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フロー順序は、ヌクレオチドフロー101（図21B）、フロー102（図21C）、フロー103（図21D）およびフロー104（図21E）を含む、4ステップ順序である。

【図21-2】図21A~21Eは、100ヌクレオチドフロー（図21A）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、別の例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フロー順序は、ヌクレオチドフロー101（図21B）、フロー102（図21C）、フロー103（図21D）およびフロー104（図21E）を含む、4ステップ順序である。

40

【図21-3】図21A~21Eは、100ヌクレオチドフロー（図21A）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、別の例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フロー順序は、ヌクレオチドフロー101（図21B）、フロー102（図21C）、フロー103（図21D）およびフロー104（図21E）を含む、4ステップ順序である。

【0063】

【図22-1】図22A~22Eは、100ヌクレオチドフロー（図22A）およびシー

50

クエンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、別の例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フローサイクルは、ヌクレオチドフロー１０１（図２２Ｂ）、フロー１０２（図２２Ｃ）、フロー１０３（図２２Ｄ）およびフロー１０４（図２２Ｅ）を含む、４ステップ順序である。

【図２２－２】図２２Ａ～２２Ｅは、１００ヌクレオチドフロー（図２２Ａ）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、別の例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フローサイクルは、ヌクレオチドフロー１０１（図２２Ｂ）、フロー１０２（図２２Ｃ）、フロー１０３（図２２Ｄ）およびフロー１０４（図２２Ｅ）を含む、４ステップ順序である。

10

【図２２－３】図２２Ａ～２２Ｅは、１００ヌクレオチドフロー（図２２Ａ）およびシーケンシングクラスター内のプライマーを同期化するように設計された再位相化フロー後の、別の例示的な模擬シーケンシングプロトコルで同一のポリヌクレオチド鑄型に対して伸長されたプライマーの数を示す。示されている再位相化フローサイクルは、ヌクレオチドフロー１０１（図２２Ｂ）、フロー１０２（図２２Ｃ）、フロー１０３（図２２Ｄ）およびフロー１０４（図２２Ｅ）を含む、４ステップ順序である。

【００６４】

【図２３】図２３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２３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（すなわち、新しいゼロまたは新しい非ゼロシグナル）を誘導したＳＮＰパーミュテーションの分率を示す。

20

【００６５】

【図２４】図２４は、各フローが、単一のヌクレオチド塩基を各々が有する、反復４ステップフローサイクルを使用して合成ポリヌクレオチドの第２の領域がシーケンシングされる、模擬高速フォワードシーケンシングプロトコルを使用して検出された様々なＳＮＰについての塩基検出感度を示す行列を示す。

【００６６】

30

【図２５－１】図２５Ａは、各フローが３つの異なるヌクレオチド塩基の混合物を含む、反復４ステップフローサイクルを使用する模擬高速フォワードシーケンシングプロトコルについての、第１、第２および第３の領域におけるフローにわたっての平均塩基組込みを示す。参照塩基検出感度に対するバリエーション塩基の行列が、図２５Ｂに示されている。図２５Ｃは、合成リードにわたっての塩基カバレッジの分布を示す。

【図２５－２】図２５Ａは、各フローが３つの異なるヌクレオチド塩基の混合物を含む、反復４ステップフローサイクルを使用する模擬高速フォワードシーケンシングプロトコルについての、第１、第２および第３の領域におけるフローにわたっての平均塩基組込みを示す。参照塩基検出感度に対するバリエーション塩基の行列が、図２５Ｂに示されている。図２５Ｃは、合成リードにわたっての塩基カバレッジの分布を示す。

40

【図２５－３】図２５Ａは、各フローが３つの異なるヌクレオチド塩基の混合物を含む、反復４ステップフローサイクルを使用する模擬高速フォワードシーケンシングプロトコルについての、第１、第２および第３の領域におけるフローにわたっての平均塩基組込みを示す。参照塩基検出感度に対するバリエーション塩基の行列が、図２５Ｂに示されている。図２５Ｃは、合成リードにわたっての塩基カバレッジの分布を示す。

【００６７】

【図２６Ａ】図２６Ａは、対照プロトコル（１０５ラウンドのＴ－Ｇ－Ｃ－Ａフローサイクル）または再位相化プロトコル（１０５ラウンドのＴ－Ｇ－Ｃ－Ａフローサイクル、この場合、ＣとＧの混合物を含有する再位相化フローが、２４フローを終えるたびに使用された）について１０，０００模擬フローグラムにわたって累積した全位相誤差（遅れ位相

50

誤差＋進み位相誤差)の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0068】

【図26B】図26Bは、対照プロトコル(105ラウンドのT-G-C-Aフローサイクル)または再位相化プロトコル(105ラウンドのT-G-C-Aフローサイクル、この場合、CとGの混合物を含有する再位相化フローが、48フローを終えるたびに使用された)について10,000模擬フロープログラムにわたって累積した全位相誤差(遅れ位相誤差＋進み位相誤差)の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0069】

【図26C】図26Cは、対照プロトコル(105ラウンドのT-G-C-Aフローサイクル)または再位相化プロトコル(105ラウンドのT-G-C-Aフローサイクル、この場合、CとGの混合物を含有する再位相化フローが、96フローを終えるたびに使用された)について10,000模擬フロープログラムにわたって累積した全位相誤差(遅れ位相誤差＋進み位相誤差)の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0070】

【図26D】図26Dは、対照プロトコル(105ラウンドのT-G-C-Aフローサイクル)または再位相化プロトコル(105ラウンドのT-G-C-Aフローサイクル、この場合、CとGの混合物を含有する再位相化フローが、192フローを終えるたびに使用された)について10,000模擬フロープログラムにわたって累積した全位相誤差(遅れ位相誤差＋進み位相誤差)の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0071】

【図26E】図26Eは、対照プロトコル(105ラウンドのT-G-C-Aフローサイクル)または再位相化プロトコル(105ラウンドのT-G-C-Aフローサイクル、この場合、C、GおよびTの混合物を含有する再位相化フローが、48フローを終えるたびに使用された)について10,000模擬フロープログラムにわたって累積した全位相誤差(遅れ位相誤差＋進み位相誤差)の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0072】

【図26F】図26Fは、対照プロトコル(105ラウンドのT-G-C-Aフローサイクル)または再位相化プロトコル(105ラウンドのT-G-C-Aフローサイクル、この場合、C、GおよびTの混合物を含有する再位相化フローが、96フローを終えるたびに使用された)について10,000模擬フロープログラムにわたって累積した全位相誤差(遅れ位相誤差＋進み位相誤差)の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0073】

【図26G】図26Gは、対照プロトコル(105ラウンドのT-G-C-Aフローサイクル)または再位相化プロトコル(105ラウンドのT-G-C-Aフローサイクル、この場合、C、GおよびTの混合物を含有する第1の再位相化フローおよびA、CおよびGの混合物を含有する第2の再位相化フローが、96フローを終えるたびに使用された)について10,000模擬フロープログラムにわたって累積した全位相誤差(遅れ位相誤差＋進み位相誤差)の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0074】

【図26H】図26Hは、対照プロトコル(105ラウンドのT-G-C-Aフローサイクル)または再位相化プロトコル(105ラウンドのT-G-C-Aフローサイクル、この場合、C、GおよびTの混合物を含有する第1の再位相化フローおよびA、CおよびGの混合物を含有する第2の再位相化フローが、192フローを終えるたびに使用された)

10

20

30

40

50

について10,000模擬フローグラムにわたって累積した全位相誤差（遅れ位相誤差＋進み位相誤差）の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

【0075】

【図26I】図26Iは、対照プロトコル（105ラウンドのT-G-C-Aフローサイクル）または再位相化プロトコル（105ラウンドのT-G-C-Aフローサイクル、この場合、C、GおよびTの混合物を含有する第1の再位相化フロー、A、CおよびTの混合物を含有する第2の再位相化フロー、A、GおよびTの混合物を含有する第3の再位相化フロー、ならびにA、CおよびGの混合物を含有する第4の再位相化フローが、96フローを終えるたびに使用された）について10,000模擬フローグラムにわたって累積した全位相誤差（遅れ位相誤差＋進み位相誤差）の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

10

【0076】

【図26J】図26Jは、対照プロトコル（105ラウンドのT-G-C-Aフローサイクル）または再位相化プロトコル（105ラウンドのT-G-C-Aフローサイクル、この場合、C、GおよびTの混合物を含有する第1の再位相化フロー、A、CおよびTの混合物を含有する第2の再位相化フロー、A、GおよびTの混合物を含有する第3の再位相化フロー、ならびにA、CおよびGの混合物を含有する第4の再位相化フローが、192フローを終えるたびに使用された）について10,000模擬フローグラムにわたって累積した全位相誤差（遅れ位相誤差＋進み位相誤差）の和の分布を示す。平均および標準偏差が凡例に示されている。対照および再位相化プロトコルについての分布の積分も示されている。

20

【発明を実施するための形態】

【0077】

カップリングされたシークエンシングリードペアをポリヌクレオチドから生成する方法、およびそのようなカップリングされたシークエンシングリードペアを解析する方法が、本明細書に記載される。カップリングされたシークエンシングリードペアを解析して、例えば、カップリングされたシークエンシングリードペアを参照配列にマッピングすること、構造バリエーションを検出すること、ポリヌクレオチドのカップリングされたペアエンド間の領域におけるバリエーション（例えば、SNP）を検出すること、トランスバージョンを検出すること、またはコンセンサス配列を決定もしくは検証することができる。

30

【0078】

ポリヌクレオチドをシークエンシングプライマーとハイブリダイズすることができ、このプライマーが、ポリヌクレオチドの第1の領域（すなわち、3'末端）を通して伸長されて、第1の領域がシークエンシングされる。次いで、プライマーは、ポリヌクレオチドの第2の領域を通して伸長され、この伸長は、第1の領域によるプライマーの伸長よりも速い速度で起こり得る。第2の領域を通じた加速プライマー伸長は、「高速フォワードシークエンシング」と呼ばれることがある。本明細書中でさらに論じられるように、プライマーは、第2の領域を通して伸長される（より旧来のペアエンドシークエンシングで起こるもののよう、プライマーが第2の領域を完全にスキップするのではなく）ため、たとえ第2の領域が第1の領域と同様にシークエンシングされなかったとしても、何らかの情報（一部のシークエンシングデータを含む可能性がある）を第2の領域について得ることができる。例えば、プライマーを、非標識ヌクレオチドのみを使用してプライマーを第2の領域を通して伸長することができる。シークエンシングプライマーが第2の領域を通して伸長されると、プライマーは、ポリヌクレオチドの第3の領域（すなわち、5'末端）へと伸長されて、第3の領域がシークエンシングされる。領域および第3の領域のシークエンシングデータをカップリングし、その結果、ポリヌクレオチドのカップリングされたシークエンシングリードペアを得ることができ、本明細書中でさらに説明されるように、追加のシークエンシングデータを第2の領域から導出することができる。

40

50

【0079】

一例において、カップリングされたシークエンシングリードペアをポリヌクレオチドは、
（a）ポリヌクレオチド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；（b）ポリヌクレオチドの第1の領域の配列に関連するシークエン
シングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および取り込
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により、生成するステップ；
（c）ステップ（b）で伸長されたプライマーを、第2の領域のフロー順序で提供される
ヌクレオチドを使用して第2の領域を通してさらに伸長するステップであって、第2の領
域を通したプライマーの伸長が、ステップ（b）におけるプライマーの伸長よりも速く進
行する、ステップ；および（d）ポリヌクレオチドの第3の領域の配列に関連するシーク
エンシングデータを、標識ヌクレオチドを使用してステップ（c）で伸長されたプライマ
ーをさらに伸長すること、および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
出することにより、生成するステップによって生成され得る。第1の領域のシークエンシ
ングデータを第3の領域のシークエンシングデータと関連付けることができ、これが、カ
ップリングされたシークエンシングリードペアを示す。第2の領域を通してプライマーを
伸長するために使用されたヌクレオチドは、非標識であり得る。

10

【0080】

一部の実施形態では、ポリヌクレオチドからのカップリングされたシークエンシングリー
ドペアを、（a）ポリヌクレオチド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；（b）ポリヌクレオチドの第1の領域の配列に関連する
シークエンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、お
よび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により、生成する
ステップ；（c）ステップ（b）で伸長されたプライマーを、第2の領域のフロー順序で
提供されるヌクレオチドを使用して、第2の領域を通してさらに伸長するステップであ
って、プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存
在を検出することなく、第2の領域を通して伸長される、ステップ；および（d）ポリヌ
クレオチドの第3の領域の配列に関連するシークエンシングデータを、標識ヌクレオチド
を使用してステップ（c）で伸長された前記プライマーをさらに伸長することおよび取り
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により生成するステップに
より、生成することができる。第1の領域のシークエンシングデータを第3の領域のシー
クエンシングデータと関連付けることができ、これが、カップリングされたシークエンシ
ングリードペアを示す。第2の領域を通してプライマーを伸長するために使用されたヌク
レオチドは、非標識であり得る。

20

30

【0081】

一部の実施形態では、ポリヌクレオチドからのカップリングされたシークエンシングリー
ドペアを、（a）ポリヌクレオチド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；（b）ポリヌクレオチドの第1の領域の配列に関連する
シークエンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、お
よび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により、生成する
ステップ；（c）ステップ（b）で伸長されたプライマーを、第2の領域のフロー順序で
提供されたヌクレオチドを使用して、第2の領域を通してさらに伸長するステップであ
って、少なくとも2つの異なるタイプのヌクレオチド塩基の混合物が、フロー順序の少な
くとも1ステップで使用される、ステップ；および（d）ポリヌクレオチドの第3の領域の
配列に関連するシークエンシングデータを、標識ヌクレオチドを使用してステップ（c）
で伸長された前記プライマーをさらに伸長することおよび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により生成するステップにより、生成することができ
る。第1の領域のシークエンシングデータを第3の領域のシークエンシングデータと関連
付けることができ、これが、カップリングされたシークエンシングリードペアを示す。第
2の領域を通してプライマーを伸長するために使用されたヌクレオチドは、非標識であり
得る。

40

50

【0082】

一部の実施形態では、プライマーは、シーケンシングクラスター内の複数のシーケンシング反応を再位相化する（すなわち、同期化する）ために第2の領域を通して伸長される。伸長プライマーにヌクレオチドを組み込む化学的プロセスは、不完全であることが多く、そのためシーケンシングクラスター内の鎖間の非同期化が生じる。非同期化の結果として、リード長が増加するにつれて伸長プライマーへのヌクレオチド組込みの存在または非存在を検出する際に、シグナルの劣化、したがって、精度の低下が生じ得る。再同期化は、シグナル損失の相殺をもたらすことができ、それによって、より長い有効リード長が可能になる。シーケンシング反応を再位相化するために、少なくとも2つ（例えば、2つまたは3つ）の異なるタイプのヌクレオチド塩基の混合物が、第2の領域のフロー順序の複数のステップで使用される、再位相化サイクルを使用して、プライマーが第2の領域を通して伸長される。再位相化サイクル中に組み込まれるヌクレオチドを一部の実施形態では検出することができず、その結果、得られるリードにギャップが生じることになる。しかし、配列が参照または他の配列にアラインメントされる場合、このリードギャップを管理することができる。

10

【0083】

たとえ第2の領域を直接または完全にシーケンシングすることができなかったとしても、参照配列を使用して第2の領域についてのシーケンシングデータを抽出することができる。例えば、シーケンシングデータを、ポリヌクレオチドの第1の領域および／または第3の領域から、伸長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により得ることができる。しかし、非標識ヌクレオチドを使用して、または取り込まれたヌクレオチドの存在もしくは非存在を検出せずに、プライマーを第2の領域を通して伸長することができる。非標識ヌクレオチドを使用することにより（または取り込まれた標識を検出するための時間をシーケンシングシステムに与えないことにより）、第2の領域を通したより迅速なプライマー伸長が可能になるが、シーケンシングデータの直接決定は可能にならない。しかし、プライマーは、所定のフロー順序で提供されるヌクレオチドを使用して第2の領域を通して伸長されるため、第2の領域におけるバリエーションは、第3の領域内で決定されるシーケンシングデータに影響を与え得る。参照配列を使用して予想シーケンシングデータ（例えば、予想フローグラム）を決定ことができ、このデータが生成シーケンシングデータ（例えば、検出フローグラム）と比較されて、第2の領域内のバリエーションを含む、バリエーションが検出される。予想シーケンシング情報（例えば、予想フローグラム）と生成シーケンシングデータ（例えば、生成フローグラム）の比較を第3の領域において（第2の領域におけるバリエーションを検出するために）行なうことができる。この方法論は、旧来のペアエンドシーケンシング法に勝る有意な利点を提供し、この利点のため、ポリヌクレオチドの3'末端または5'末端についてのシーケンシングデータは、ポリヌクレオチドの3'末端と5'末端の間のポリヌクレオチドにおけるバリエーションによる影響を受けない。

20

30

定義

【0084】

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。

40

【0085】

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。

【0086】

「予想シーケンシングデータ」は、カップリングされたシーケンシングリードペアを生成するために使用されるポリヌクレオチドの配列、または前記ポリヌクレオチドの領域の配列が、参照配列にマッチするかどうかを予想する、シーケンシングデータを指す。

【0087】

50

「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クエ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、シークエ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。

【0088】

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。

【0089】

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（TCEP）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。

【0090】

「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。

【0091】

「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。

【0092】

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。

【0093】

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。

【0094】

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。

【0095】

本明細書で使用する節の見出しは、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

10

20

30

40

50

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。

【0096】

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られて遂行され得る。したがって、例証される（および下記で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。

10

【0097】

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

【0098】

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鑄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、鑄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。

20

30

【0099】

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鑄型と混合して、相補的塩基が鑄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鑄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キング基は除去される。相補的塩基が鑄型鎖内に存在しない場合、鑄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。

40

50

【0100】

プライマー伸長の過程でヌクレオチドを決定された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。ヌクレオチドが段階的に付加され、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組み込むことが可能になる。サイクルは、ヌクレオチドの同じ順序および異なる塩基タイプの数、またはヌクレオチドの異なる順序、および／または異なる数の異なる塩基タイプを有し得る。しかし、所与のフローステップに対応する塩基のセット（すなわち、単一フローステップで同時に使用される1つまたは複数の異なる塩基）は、この用語が本明細書で使用される場合、同じサイクルで反復されず、異なるサイクルを区別するためのマーカーとなることができる。単に例として、第1のサイクルの順序は、A-T-G-Cであり得、第2のサイクルの順序は、A-T-C-Gであり得る。さらに、1つまたは複数のサイクルが、1つまたは複数のヌクレオチドを含まないこともある。単に例として、第1のサイクルの順序は、A-T-G-Cであり得、第2のサイクルの順序は、A-T-Cであり得る。代替順序を当業者は容易に企図することができる。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。

10

【0101】

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SeqAmp DNAポリメラーゼが、挙げられる。

20

30

【0102】

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。

40

【0103】

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

50

、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。

10

20

30

40

50

【0104】

フローグラムなどの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。以下の鑄型配列：CTGおよびCAG、ならびにT-A-C-Gの反復フローサイクル（つまり、相補的塩基が鑄型ポリヌクレオチドに存在する場合にのみプライマーに組み込まれることになる、T、A、CおよびGヌクレオチドの逐次的付加）を例にとる。結果として生じるフローグラムが表1に示され、この表中の1は、導入されたヌクレオチドが組み込まれること示し、0は、導入されたヌクレオチドが組み込まれないことを示す。フローグラムを使用して、鑄型鎖の配列を決定することができる。

【表1】

表 1

配列	サイクル1				サイクル2			
	T	A	C	G	T	A	C	G
CTG	0	0	0	1	0	1	1	0
CAG	0	0	0	1	1	0	1	0
CCG	0	0	0	2	0	0	1	0

【0105】

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、C C Gの配列は、2つのG塩基を組み込むことになり、標識塩基により放出されるいずれのシグナルも、単一塩基組込みの場合よりも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すが、所与のステップで組み込まれる塩基の数を含む追加情報を提供することができる。

【0106】

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。

10

【0107】

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。

20

【0108】

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、ポリヌクレオチドの第1の領域、第2の領域、および第3の領域を通して伸長される。第1の領域および／または第3の領域内の配列に関連するシーケンシングデータを、上記で論じられたように生成することができる。しかし、プライマーは、加速された「高速フォワード」プロセスを使用する第2の領域（これは、第1の領域と第3の領域の間にある）を通して伸長される。つまり、ポリヌクレオチドの第1の領域と第3の領域の間の第2の領域を通したプライマーの伸長は、第1の領域および／または第3の領域を通したプライマーの伸長よりも速く進行し得る。例えば、第2の領域を通したプライマーの伸長は、伸長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せずに、伸長プライマーを伸張させることによって進行し得る。フローシーケンシング中に、上記で論じられたように、標識ヌクレオチドが伸長プライマーに組み込まれ、ハイブリダイズされた鋳型が洗浄され、検出器が使用されてヌクレオチドの標識からのシグナルが検出され、このシグナルが、伸長されたプライマーにヌクレオチドが取り込まれたかどうかを示す。しかし、この検出プロセスは時間がかかるため、この検出プロセスをスキップすることにより第2の領域を通したプライマーの伸長は加速され得る。一部の実施形態では、プライマーは、非標識ヌクレオチドを使用して（または非標識ヌクレオチドのみを使用して）第2の領域を通して伸長され、このことによりプライマー伸長速度はさらに加速され得る。

30

40

【0109】

代替的に、または加えて、第2の領域を通したプライマーの伸長は、第2の領域を通したプライマーの伸長中に使用されるフロー順序の少なくとも1ステップで少なくとも2つの異なるタイプのヌクレオチドの混合物を使用することにより加速され得る。例えば、GおよびCなどの2つの異なる塩基を同じステップで、同時に使用してもよく、このステップ

50

は、相補的 C または G 塩基が存在する場合にプライマーを伸長する。これは、たとえこれらの塩基が異なる塩基タイプであったとしてもプライマーへの連続塩基組込みにより、プライマーの伸長を加速させる。一部の実施形態では、フロー順序の少なくとも 1 ステップは、2 つの異なる塩基を含む。一部の実施形態では、フロー順序の少なくとも 1 ステップは、3 つの異なる塩基を含む。例として、表 2 に示されている配列番号 1 ならびに対応するフロー順序およびフロープログラムを考える。配列番号 1 を含有するポリヌクレオチドとハイブリダイズされたシーケンシングプライマーを伸長するためのフロー順序プロセスは、5 サイクルを含み、サイクル 1、4 および 5 は、互いに同じであり、サイクル 2 および 3 は、互いに同じである（サイクル 1、4 および 5 は、サイクル 2 および 3 とは異なる）。この例では、各サイクルは、4 ステップを有し、サイクル 1、4 および 5 は、単一塩基タイプが各サイクルステップで付加される、A-C-T-G ヌクレオチドの逐次的な独立した付加を含む。サイクル 2 および 3 は、4 つのサイクルステップを含み、ステップ 1 は、A ヌクレオチドを含まず（すなわち、C、T および G を含み）、ステップ 2 は、C ヌクレオチドを含まず（すなわち、A、T および G を含み）、ステップ 3 は、T ヌクレオチドを含まず（すなわち、A、C および G を含み）、ステップ 4 は、G ヌクレオチドを含まない（すなわち、A、C および T を含む）。サイクル 2 および 3 は、プライマー伸長中に同時に複数の異なるヌクレオチド塩基タイプを含むため、プライマーは、いずれかの所与のステップで単一の塩基タイプしか使用されなかった場合よりも速く伸長される。このフロー順序を使用する配列番号 1 鑄型に対するプライマーの伸長については、表 2 に示されているフロープログラムから、6 以下の塩基がプライマー伸長の高速フォワード部分の間に付加されることになる（サイクル 3、ステップ 3）。対照的に、表 3 は、単一ヌクレオチドが各ステップで使用される（表 2 のサイクル 1、4 および 5 と同様の）A-C-T-G サイクルを使用する同じ配列番号 1 のフロープログラムを示す。表 3 に示されているプライマーを伸長するために使用されたフロー順序は、ポリヌクレオチドを通してプライマーを伸長するために 10 サイクルの 4 ステップサイクルを必要とし、これは、表 2 で提供されるフロー順序を使用してポリヌクレオチドを通してプライマーを伸長するために使用される 5 サイクルの 4 ステップサイクルよりもかなり遅い。

10

20

30

40

50

【表 2】

表 2

サイクル	1				2				3				4				5			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
サイクルステップ	A	C	T	G	C/T/ G	A/T/ G	A/C/ G	A/C/ T	C/T/G	A/T/ G	A/C/G	A/C/T	A	C	T	G	A	C	T	G
取り込まれた塩基の数	1	1	1	1	0	2	1	3	4	3	6	2	0	0	0	1	1	1	1	1
取り込まれた塩基	A	C	T	G	-	AA	C	TTA	GGCT (配列番号 2)	ATA	CGGACG (配列番号 3)	TC	-	-	-	G	A	C	T	G

配列番号 1: 3'-TGACTTGAATCCGATATGCCCTGCAGCTGAC-5' についてのフローグラム

10

20

30

40

50

【表 3】

表 3

	1				2				3				4				5				
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
サイクル																					
サイクルステップ	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
フロー塩基	A	C	T	G	A	C	T	G	A	C	T	G	A	C	T	G	A	C	T	G	
取り込まれた塩基の数	1	1	1	1	2	1	2	0	1	0	0	2	0	1	1	0	1	0	1	0	
取り込まれた塩基	A	C	T	G	AA	C	TT	-	A	-	-	CC	-	C	T	-	A	-	T	-	
	6				7				8				9				10				
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
サイクル																					
サイクルステップ	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
フロー塩基	A	C	T	G	A	C	T	G	A	C	T	G	A	C	T	G	A	C	T	G	
取り込まれた塩基の数	1	1	0	2	1	1	0	1	0	0	1	0	0	1	0	1	1	1	1	1	
取り込まれた塩基	A	C	-	GG	A	C	-	G	-	G	-	T	-	-	C	-	G	A	C	T	G

配列番号 1: 3'-TGACTTGAATCCCGATATGCCTGCAGCTGAC-5'についてのフローグラム

【0110】

高速フォワード法は、直接シーケンシングされない領域を通したプライマー伸長を加速させるのに特に有用である。例えば、表 2 を参照して、サイクル 1、4 および 5 は、標識ヌクレオチドを段階的に使用して、第 1 の領域（サイクル 1）および第 3 の領域（サイクル 4 および 5）に関連するシーケンシングデータを生成したが、プライマーは、第 1 の領域と第 3 の領域の間の第 2 の領域（サイクル 2 およびサイクル 3）を通して急速に伸長された。

【0111】

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

10

20

30

40

50

ど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。第1の領域または第3の領域におけ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、第1の領域におけるプライマー伸長または第3の領域におけるプライマー伸長は、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。第1の領域または第3の領域においてプライマーに取り込まれる塩基の数は、第1の領域または第3の領域の配列、およびプライマーを伸長するために使用される第1の領域または第3の領域におけるフロー順序に依存する。一部の実施形態では、第1のまたは第3の領域は、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。

10

20

30

40

50

【0112】

第2の領域を通したプライマー伸長は、任意の数のフローステップを通して進行し得る。一部の実施形態では、第2の領域を通したプライマーの伸長は、標識ヌクレオチドを含まず、このことにより、ポリメラーゼが失速せずにプライマーの実行可能な伸長距離がさらに増加される。一部の実施形態では、第2の領域によるプライマー伸長は、1～約10,000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、約500～約1000ステップの間のフローステップ、約1000フローステップ～約2500フローステップの間、約2500フローステップ～約5000フローステップの間、または約5000フローステップ～約10,000フローステップの間のフローステップを含む。一部の実施形態では、第2の領域を通したプライマーの伸長は、約10,000ステップより多くのフローステップを含む。第2の領域においてプライマーに組み込まれる塩基の数は、第2の領域の配列、および第2の領域において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、第2の領域は、長さ約1塩基～約50,000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、長さ約2000塩基～約2500塩基、長さ約2500～約5000塩基、長さ約5000～約10,000塩基、長さ約10,000～約25,000塩基、または長さ約25,000～約50,000塩基である。一部の実施形態では、第2の領域の長さは、長さ約50,000塩基を超える。

【0113】

プライマーの伸長は、第1の領域、第2の領域、および第3の領域を通して進行し、プライマーは、標識ヌクレオチドを使用して第1の領域および第3の領域を通して伸長される。伸長プライマーに組み込まれるヌクレオチドを検出してシーケンシングデータを生成することができる。第2の領域を通したプライマーの伸長は、例えば、伸長プライマーに組み込まれるヌクレオチドの標識の存在もしくは非存在を検出することなく、またはプライマーを伸長するために少なくとも2つの異なるタイプのヌクレオチド塩基の混合物を含むことにより（第1および／または第3を通したプライマーの伸長のほうが少ない異なる

タイプのヌクレオチド塩基に頼る)、第1および/または第3の領域を通したプライマーの伸長よりも速い速度で起こり得る。プライマーの伸長は、交互パターンでさらに延長され得る。例えば、プライマーは、第3の領域を通して伸長された後、第4の領域へとさらに伸長され得る。第4の領域を通したプライマーの伸長は、例えば、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、またはプライマーを伸長するために少なくとも2つの異なるタイプのヌクレオチド塩基の混合物を含むことにより、第1および/または第3の領域を通したプライマーの伸長よりも速い速度で起こり得る。次いで、プライマーは、標識ヌクレオチドを使用して第5の領域へと伸長され得、伸長されたプライマーに取り込まれたヌクレオチドを検出することにより第5の領域についてのシーケンシングデータを生成することができる。このプロセスを何交互サイクルでも好きなだけ反復することができる。任意の2領域からのシーケンシングデータを関連付けてカップリングされたシーケンシングリードペアを生成することができ、カップリングされたシーケンシングリードペアを、本明細書中で説明されるように(例えば、選択領域間の領域を、本明細書で提供される解析方法について説明されるような「第2の領域」である考えることにより)解析することができる。

10

20

30

40

50

【0114】

図1は、カップリングされたシーケンシングリードペアをポリヌクレオチド(例えば、DNA)から生成するための例示的方法の概略図を示す。102で、ポリヌクレオチド104がプライマー106とハイブリダイズされて、ハイブリダイズされた鑄型が形成される。一部の実施形態では、ポリヌクレオチドは、シーケンシングライブラリー調製中に標的ポリヌクレオチドの3'にライゲーションされ得るアダプター領域108を含む。アダプター領域108は、ハイブリダイゼーション領域を含むことができ、プライマー106は、アダプター領域108のハイブリダイゼーション領域とハイブリダイズすることができる。ステップ110で、ポリヌクレオチド104の第1の領域112についてのシーケンシングデータが、標識ヌクレオチドを使用してプライマー106を伸長すること、および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により、生成される。プライマーを伸長するために使用されるヌクレオチドは、非標識ヌクレオチドをさらに含むことがあるが、シーケンシングデータを生成するために、ヌクレオチド組込みを検出するために標識ヌクレオチドが使用される。一部の実施形態では、ヌクレオチドは、第1の領域112を通してプライマー106を伸長するために第1の領域のフロー順序に従って1または複数のサイクルで段階的に付加され、ハイブリダイズされた鑄型をサイクルステップに従って洗浄して、組み込まれなかったヌクレオチドを、取り込まれた標識ヌクレオチドの存在または非存在の検出の前に除去することができる。ステップ114で、プライマー106は、第2の領域のフロー順序に従ってポリヌクレオチド104の第2の領域116を通して伸長される。プライマー106は、第2の領域116を通して、ステップ110でのプライマーの伸長よりも速い速度で伸長され得る。この加速されたプライマー伸長は、この方法の「高速フォワード」部分と呼ばれることがある。ヌクレオチド(これらは、一部の実施形態では、非標識である)は、ハイブリダイズされた鑄型に第2の領域のフロー順序に従って1または複数のサイクルで段階的に付加される。一部の実施形態では、1つより多くの(例えば、2つまたは3つの)異なる塩基タイプが所与のサイクルステップで同時に使用され、このことによりプライマー伸長が加速される。一部の実施形態では、ヌクレオチドは非標識であり、このことにより、標識ヌクレオチドよりも速いプライマー伸長が可能になる。一部の実施形態では、プライマーは、ヌクレオチドの標識の存在または非存在を検出することなく、伸長される。ステップ118で、ポリヌクレオチド104の第3の領域118についてのシーケンシングデータが、標識ヌクレオチドを使用してプライマー106を伸長すること、および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により、生成される。第3の領域118についてのシーケンシングデータの生成は、第1の領域112についてのシーケンシングデータの生成について説明したのと同様の方法で進行し得る。ステップ122で、第1の領域112について生成されたシーケンシングデータは、第3の領域120について生成されたシ

ークエンシングデータと関連付けられ、その結果、ポリヌクレオチド 104 についてのカップリングされたシークエンシングリードペア 124 が得られる。第 1 の領域と第 3 の領域とを関連付けたシークエンシングデータは、第 1 および第 3 の領域の配列を含み得る。カップリングされたシークエンシングリードペア 124 は、第 1 の領域 112 および第 3 の領域 120 についてのシークエンシングデータを含み、これらの領域は、必ずしもシークエンシングデータが分かるとは限らない第 2 の領域 116 により隔てられている。

【0115】

ポリヌクレオチドの第 1 の領域についてのシークエンシングデータの生成は、本明細書に記載される実施形態の一部に従って生成されなくてもよい。例えば、シークエンシングプライマーを、標的領域とハイブリダイズすることにより、標的シークエンシングに使用することができる。標的シークエンシングでは、ポリヌクレオチドの第 1 の領域は既知であり、プライマーは、第 1 の領域に特異的に結合するように設計される。次いで、プライマーを、説明されたように第 2 の領域および第 3 の領域を通して伸長することができ、第 3 の領域についてのシークエンシングデータが生成されることになる。一部の実施形態では、カップリングされたシークエンシングリードペアをポリヌクレオチドから生成する方法は、(a) プライマーをポリヌクレオチドの第 1 の領域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；(b) プライマーを、第 2 の領域のフロー順序で提供されるヌクレオチドを使用して第 2 の領域を通して伸長するステップであって、(i) プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第 2 の領域を通して伸長される、または (ii) 少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、第 2 の領域のフロー順序の少なくとも 1 ステップで使用される、ステップ；および (c) ポリヌクレオチドの第 3 の領域の配列に関連するシークエンシングデータを、標識ヌクレオチドを使用してステップ (b) で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む。

【0116】

参照配列を使用して、第 1 の領域、第 2 の領域および／または第 3 の領域についての予想シークエンシングデータ（例えば、フローグラム）を決定することができる。第 1 および第 3 の領域についての配列を、それらの領域について生成されたシークエンシングデータから決定することができる。例えば、表 2 を参照して、サイクル 1 は、第 1 の領域に関連付けられ、この配列は、塩基に対する相補配列として容易に決定され（すなわち、塩基フロー A-C-T-G は、T-G-A-C の配列に対応する）、サイクル 4 および 5 は、第 3 の領域に関連付けられ、この配列は、C-T-G-A-C（すなわち、G-A-C-T-G の相補配列）として決定される。したがって、第 1 の領域および／または第 3 の領域から生成されたシークエンシングデータを使用して、第 1 の領域および／または第 3 の領域（または第 1 の領域および／もしくは第 3 の領域の少なくとも一部分）を参照配列にマッピングすることができる。参照配列にマッピングされると、第 2 の領域および参照配列を通してプライマーを伸長するために使用されたフロー順序を使用して、第 2 の領域についての予想シークエンシングデータを生成することができる。

【0117】

第 2 の領域についての参照配列、第 2 の領域についてのフロー順序、第 3 の領域についてのフロー順序、および第 3 の領域の配列についての情報を使用して、第 3 の領域についての予想シークエンシングデータを決定することもできる。同様に、第 2 の領域についての参照配列、第 2 の領域についてのフロー順序、第 1 の領域についてのフロー順序、および第 1 の領域の配列についての情報を使用して、第 1 の領域についての予想シークエンシングデータを決定することができる。第 3 の領域（または第 1 の領域）の配列についての情報を、例えば、参照配列（もしくは異なる参照配列）、または生成されたシークエンシングデータ、例えば、標識ヌクレオチドを使用してプライマーを伸長し、取り込まれた標識ヌクレオチドの存在もしくは非存在を検出することにより生成されたシークエンシングデータ、または他の方法（例えば、ポリヌクレオチドの第 3 の領域の第 3 の領域を独立して

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から、得ることができる。

【0118】

例として、第3の領域についての予想シーケンシングデータを、第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列を使用して決定することができる。第1の領域(またはその一部分)を参照配列にマッピングすることができ、第2の領域に対応する参照配列、および第2の領域のフロー順序を使用して、第2の領域についての予想参照シーケンシングデータを決定することができる。同様に、第3の領域についての参照配列を、第3の領域のフロー順序とともに使用して、第3の領域についての予想参照シーケンシングデータを決定することができる。同様の方法を使用して、第1の領域についての予想シーケンシングデータを決定することができる。例えば、第1の領域についての予想シーケンシングデータを、第2の領域についての参照配列、第2の領域のフロー順序、第1の領域のフロー順序、および第1の領域についての参照配列を使用して決定することができる。第3の領域(またはその一部分)を参照配列にマッピングすることができ、第2の領域に対応する参照配列、および第2の領域のフロー順序を使用して、第2の領域についての予想参照シーケンシングデータを決定することができる。同様に、第1の領域についての参照配列を、第1の領域のフロー順序とともに使用して、第1の領域についての予想参照シーケンシングデータを決定することができる。

10

【0119】

別の例では、第3の領域についての予想シーケンシングデータを、第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および前述の通り生成されたシーケンシングデータと同じこともあり、または異なることもある、第3の領域の配列に関連するシーケンシングデータを使用して決定することができる。第1の領域(またはその一部分)を参照配列にマッピングすることができ、第2の領域に対応する参照配列、および第2の領域のフロー順序を使用して、第2の領域についての予想参照シーケンシングデータを決定することができる。第3の領域についてのシーケンシングデータを使用して、第3の領域の配列を決定することができる。さらに、第3の領域の配列を、第3の領域のフロー順序とともに使用して、第3の領域についての予想シーケンシングデータを決定することができる。

20

30

【0120】

図2は、予想シーケンシングデータを生成するための例示的方法の概略図を示す。ステップ202で、カップリングされたシーケンシングリードペアが参照配列にマッピングされる。カップリングされたシーケンシングリードペアのマッピングは、カップリングされたシーケンシングリードペア(もしくはその一部分)の第1の領域(もしくはその一部分)の参照配列へのマッピング、カップリングされたシーケンシングリードペアの第3の領域(もしくはその一部分)の参照配列への、または第1の領域(もしくはその一部分)と第3の領域(もしくはその一部分)両方の参照配列へのマッピングを含み得る。ステップ204で、第2の領域についての予想シーケンシングデータ(例えば、予想フローグラム)が、第2の領域のフロー順序および参照配列を使用して決定される。既知のフロー順序および参照シーケンシングを用いて、予想シーケンシングデータ(つまり、ポリヌクレオチドの第2の領域が参照配列にマッチするかどうかを予想されるシーケンシングデータ)の決定を容易に達成することができる。さらに、第2の領域についての予想シーケンシングデータを使用して、第2の領域の予想5'末端を決定することができる。第2の領域の5'末端は、その領域についてのフロー順序、および第2の領域の配列によって変わり得る。したがって、第3の領域の3'末端も、第2の領域のフロー順序、および第2の領域の配列に基づいて変わり得る。なぜなら、第3の領域の3'末端は、第2の領域の5'末端に隣接しているからである。第3の領域の3'末端が(例えば、第2の領域についての予想シーケンシングデータを使用して決定して)確立されると、ステップ206に示されているように、第3の領域についての予想シーケンシングデータを

40

50

決定することができる。本明細書中でさらに説明されるように、第3の領域についての予想シーケンシングデータを使用して、ポリヌクレオチドの第2の領域内のバリエーションなどの、バリエーションを決定することができる。

【0121】

ポリヌクレオチドが第2の領域内にバリエーションを含む場合、第3の領域に関連する生成シーケンシングデータ（例えば、フローグラム）と第3の領域に関連する予想シーケンシングデータとは（配列コンテキスト、およびバリエーションのサイズに依存して）異なり得る。したがって、一部の実施形態では、バリエーションは、予想シーケンシングデータと生成シーケンシングデータとの差に基づいて検出される。

【0122】

参照配列は、ポリヌクレオチドと同じ種の任意の好適な配列であり得、参照配列とポリヌクレオチドの配列の間には多少の差があり得る。本明細書に記載される方法の一部の実施形態では、これらの差、すなわちバリエーション、を検出することができる。一部の実施形態では、試験バリエーション（すなわち、目的のバリエーション）は、参照配列に含まれており、他の実施形態では、試験バリエーションは、参照配列には含まれていない。一部の実施形態では、一方が、試験バリエーションを含む参照配列であり、他方が、試験バリエーションを含まない参照配列である、2つの異なる参照配列を用いて、解析を遂行することができる。一部の実施形態では、2つの参照配列間の唯一の差は、試験バリエーションの存在または非存在である。

【0123】

本明細書に記載されるバリエーション検出方法の感度は、第1、第2および／または第3の領域においてプライマーを伸長するために使用されるバリエーションのコンテキストおよび／またはフロー順序に依存し得る。所与のフロー順序で見逃されるバリエーションは、第1、第2および／または第3の領域において異なるフロー順序を使用することにより検出可能になり得る。したがって、本明細書に記載される方法の一部の実施形態では、ポリヌクレオチドの第1、第2および／または第3の領域のうちの1つまたは複数を通してプライマーを伸長するために、1つより多くのカップリングされたシーケンシングリードペアが異なるフロー順序を使用して生成される。

【0124】

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。

【0125】

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。

【0126】

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

10

20

30

40

50

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。

【0127】

一部の実施形態では、フローシーケンシング法は、ローリングサークル増幅（RCA）シーケンシングとともに使用される。RCAは、線状配列で共有結合された核酸分子の複数のコピーの形成を可能にする。例えば、Dean et al., Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification, Genome Research, vol. 11, pp. 1095-1099 (2000 1) ; および米国特許第5,714,320号を参照されたく、これら参考文献の各々についての内容は、参照により本明細書に組み込まれる。核酸分子の複数のコピーを直線的にシーケンシングすることができるため、所与の領域は、シーケンシングが進行するにつれて「暗」モードまたは「明」モードで交互にシーケンシングされ得る。一部の実施形態では、シーケンシングモード切り替えを、動的に（および必要に応じて、自動的に）決定することができる。例えば、バリエーションは、「暗」領域内で検出され得るが、わずかな情報しか生成されないため、特定のバリエーションのコーリングが阻まれる。それ故、シーケンシングフローを動的に調整して、明モードでバリエーションを含有する核酸の領域をシーケンシングすることができる。例えば、試験試料中の短い遺伝子バリエーションを検出する方法は、（a）ローリングサークル増幅（RCA）を使用してポリヌクレオチドを増幅させて、ポリヌクレオチドの第1のコピーとポリヌクレオチドの第2のコピーとを少なくとも含むRCA増幅ポリヌクレオチドを生成するステップ；（b）RCA増幅ポリヌクレオチド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；（c）ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第1の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；（d）プライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用してポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通してさらに伸長するステップであって、（i）プライマーは、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長されるか、（ii）少なくとも2つの異なるタイプのヌクレオチド塩基の混合物は、第2の領域のフロー順序の少なくとも1ステップで使用されるか、または（iii）ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通したプライマーの伸長は、第1の領域を通したプライマーの伸長よりも速く進行する、ステップ；（e）ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーをさらに伸長すること、および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により、生成するステップ；（f）ポリヌクレオチドの第3の領域について生成されたシーケンシングデータをポリヌクレオチドの第3の領域の予想配列についての予想シーケンシングデータと比較するステップ；（g）ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップ；（h）ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；および（i）ポリヌクレオチドの第2の領域における短い遺伝子バリエーションのアイデンティティをコールするステップを含み得る。一部の実施形態では、ポリヌクレオチドの第2の

10

20

30

40

50

コピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータは、ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップに基づいて動的に生成される。

伸張されたフローサイクル

【0128】

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張することができる。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。伸張されたフローサイクルは、例えば、本明細書に記載の方法に従い第2の領域を通してプライマーを伸張するために使用され得る。

10

【0129】

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）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。

20

30

【0130】

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張された場合、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。

40

50

【0131】

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4に示されている。

【0132】

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。

【0133】

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

10

20

30

40

50

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。

【0134】

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。

10

20

【0135】

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。

30

40

【0136】

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4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

50

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。

【表4-1】

表4

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-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 4 - 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-G-T-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 4 - 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-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-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-C-T-A-G	64.2%	87.5% @ 9%	87.5%	86.5%	79.2%	70.8%

10

20

30

40

【0137】

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%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のまたはそれを越える塩基取り込

50

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 4 に記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。

再位相化フロー

【0138】

1 つまたは複数の再位相化フローを第 2 の領域としてまたは第 2 の領域内で使用して、シーケンシングクラスター内の並列シーケンシング反応を再位相化する（すなわち、同期化する）ことができる。シーケンシングクラスターは、共通の表面（例えば、ビーズまたはフローセル）にしっかりと結合された複数のポリヌクレオチドコピーを含む。クラスターは、例えば、ポリヌクレオチドを表面に結合させること、および結合されたポリヌクレオチドを（例えば、ブリッジ増幅により）増幅させることにより、形成することができる。ポリヌクレオチドの各々とハイブリダイズされたプライマーが、同一の鑄型に基づいてヌクレオチドを組み込むことにより同時に伸長されるので、シーケンシングデータをシーケンシングクラスターからまとめて収集することができる。しかし、伸長プライマーにヌクレオチドを組み込む化学的プロセスは、不完全であることが多く、そのためシーケンシングクラスター内の鎖間の非同期化が生じる。つまり、ある特定のプライマーは、クラスター内の他の伸長されたプライマーと比較して進行が遅いことがある。非同期化の結果として、リード長が増加するにつれて伸長プライマーへのヌクレオチド組込みの存在または非存在を検出する際に、シグナルの劣化、したがって、精度の低下が生じ得る。再同期化は、シグナル損失の相殺をもたらすことができ、それによって、より長い有効リード長が可能になる。シーケンシング反応を再位相化するために、少なくとも 2 つ（例えば、2 つまたは 3 つ）の異なるタイプのヌクレオチド塩基の混合物が、第 2 の領域のフロー順序の複数のステップで使用される、1 つまたは複数の再位相化フローを使用して、プライマーが第 2 の領域を通して伸長される。再位相化フロー中に組み込まれるヌクレオチドを一部の実施形態では検出することができず、その結果、得られるリードにギャップが生じることになる。しかし、配列が参照または他の配列にアラインメントされる場合、このリードギャップを管理することができる。そのような「キャッチアップフロー」を含むことにより、遅れプライマーは、クラスター内の他の伸長されたプライマーに対する遅れを取り戻すことができる。

【0139】

複数のポリヌクレオチドコピーを含むシーケンシングクラスターを（例えば、シーケンシングクラスター内で）再同期化する方法は、少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、再位相化フロー順序の少なくとも 1 ステップで使用される、再位相化フロー順序を使用して、ポリヌクレオチドコピーとハイブリダイズされたプライマーを伸長するステップを含み得る。一部の実施形態では、シーケンシングクラスター内のシーケンシングプライマーを同期化する方法は、（a）プライマーをシーケンシングクラスター内のポリヌクレオチドコピーとハイブリダイズするステップ；（b）プライマーを、第 1 の領域のフローサイクルに従って標識ヌクレオチドを使用してポリヌクレオチドコピーの第 1 の領域を通して伸長するステップ；（c）プライマーを、1 つまたは複数の再位相化フローを使用してポリヌクレオチドコピーの第 2 の領域を通して伸長するステップであって、少なくとも 2 つの異なるタイプのヌクレオチドの混合物は、1 つまたは複数の再位相化フローの各々において使用される、ステップ；および（d）プライマーを、第 3 の領域のフローサイクルに従って標識ヌクレオチドを使用してポリヌクレオチドコピーの第 3 の領域を通して伸長するステップを含む。

【0140】

複数のポリヌクレオチドコピー（例えば、シーケンシングクラスター内の）からシーケンシングリードを生成する方法は、再同期化方法を含み得る。例えば、シーケンシングリードペアを複数のポリヌクレオチドコピーから生成するための方法は、（a）ポリヌクレオチドコピーをプライマーとハイブリダイズして、ハイブリダイゼーション鑄型を形成するステップ；（b）ポリヌクレオチドコピーの第 1 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および取り

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により、生成するステップ；（c）ステップ（b）で伸長されたプライマーを、1つまたは複数の再位相化フローで提供されるヌクレオチドを使用して第2の領域を通してさらに伸長するステップであって、少なくとも2つの異なるタイプのヌクレオチド塩基の混合物が1つまたは複数の再位相化フローの各々で使用される、ステップ；および（d）ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ（c）で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む。

【0141】

再位相化フロー順序（または再位相化フローサイクル）は、シーケンシングクラスターにおいて遅れプライマーが先導プライマーに対する遅れを取り戻すことを可能にする1つまたは複数のステップを含む。再位相化フロー順序でのステップのうちの少なくとも1つ（例えば、1つ、2つ、3つ、4つ、またはそれより多く）は、2つまたはそれより多く（例えば3つ）の異なるタイプのヌクレオチド塩基の混合物を含む。一部の実施形態では、再位相化フロー順序は、2つまたは3つの異なるタイプのヌクレオチド塩基を各々が含む、1つ、2つ、3つ、4つ、5つ、またはそれより多くのフローを含む。

【0142】

再位相化フロー順序は、再位相化フロー順序後の同期化された伸長プライマーの部分を増加させるように構成される。一部の実施形態では、再位相化フロー順序は、任意の順序で、（i）A、CおよびGヌクレオチドを含むがT（および／またはU）ヌクレオチドを含まない混合物を含むフローステップ（「非T」（および／または「非U」）ステップとも呼ばれる）；（ii）T（および／またはU）、CおよびGヌクレオチドを含むがAヌクレオチドを含まない混合物を含むフローステップ（「非A」ステップとも呼ばれる）；（iii）T（および／またはU）、AおよびGヌクレオチドを含むがCヌクレオチドを含まない混合物を含むフローステップ（「非C」ステップとも呼ばれる）；および（iv）T（および／またはU）、AおよびCヌクレオチドを含むがGヌクレオチドを含まない混合物を含むフローステップ（「非G」ステップとも呼ばれる）を含む。

【0143】

他の再位相化フローを決定することができる。例として、一部の実施形態では、再位相化フロー（再位相化フロー順序での）は、任意の順序で、（i）AおよびCヌクレオチドを含むがGおよびT（および／またはU）ヌクレオチドを含まない混合物を含むフローステップ；（ii）T（および／またはU）およびGヌクレオチドを含むがAおよびCヌクレオチドを含まない混合物を含むフローステップ；（iii）AおよびGヌクレオチドを含むがT（および／またはU）およびCヌクレオチドを含まない混合物を含むフローステップ；（iv）T（および／またはU）およびCヌクレオチドを含むがAおよびGヌクレオチドを含まない混合物を含むフローステップ；（v）AおよびT（および／またはU）を含むがGおよびCヌクレオチドを含まない混合物を含むフローステップ；（vi）CおよびGヌクレオチドを含むがAおよびT（および／またはU）を含まない混合物を含むフローステップ；（vii）A、GおよびCヌクレオチドを含むがTヌクレオチドを含まない混合物を含むフローステップ；（viii）T（および／またはU）、AおよびGヌクレオチドを含むがCヌクレオチドを含まない混合物を含むフローステップ；（ix）C、T（および／またはU）およびAヌクレオチドを含むがGヌクレオチドを含まない混合物を含むフローステップ；および／または（x）G、CおよびT（および／またはU）ヌクレオチドを含むがAヌクレオチドを含まない混合物を含むフローステップ、のうちの1つまたは複数を含む。

【0144】

4つすべてのタイプの非終結ヌクレオチドの混合物（すなわち、A、C、GおよびT（および／またはU）を含む混合物）を含むことは、無制御プライマー伸長を生じさせる結果となり得る。しかし、3つの塩基タイプが非終結ヌクレオチドであり、1つの塩基タイプが可逆的ターミネーターを含む、4つすべてのタイプのヌクレオチドの混合物を、再位相

10

20

30

40

50

化フロー順序で 사용할 ことができる。例えば、一部の実施形態では、再位相化フロー順序は、(i) 非終結 Aヌクレオチドと、非終結 Cヌクレオチドと、非終結 Gヌクレオチドと、可逆的ターミネーターを含む T (および/または U)ヌクレオチドとを含む (またはそれらからなる) 混合物を含むフローステップ; または (ii) 非終結 T (および/または U)ヌクレオチドと、非終結 Aヌクレオチドと、非終結 Cヌクレオチドと、可逆的ターミネーターを含む Gヌクレオチドとを含む (またはそれらからなる) 混合物を含むフローステップ; または (iii) 非終結 Gヌクレオチドと、非終結 T (および/または U)ヌクレオチドと、非終結 Aヌクレオチドと、可逆的ターミネーターを含む Cヌクレオチドとを含む (またはそれらからなる) 混合物を含むフローステップ; または (iv) 非終結 Cヌクレオチドと、非終結 Gヌクレオチドと、非終結 T (および/または U)ヌクレオチドと、可逆的ターミネーターを含む Aヌクレオチドとを含む (またはそれらからなる) 混合物を含むフローステップを含む。プライマーは、可逆的ターミネーターを含むヌクレオチドが組み込まれるまで、鋳型鎖に基づいてヌクレオチドを組み込むことにより伸長され、これにより、シーケンシングクラスター内の可逆的ターミネーターを有する塩基の位置で伸長プライマーが同期化される。次いで、可逆的ターミネーターを除去することができ、次いで、シーケンシングプロセスは、同期化されたプライマーを用いて進行することができる。

10

【0145】

一部の実施形態では、再位相化フロー順序は、(i) C、Gおよび T (および/または U)塩基 (A塩基を含まない) の混合物を含む第1の再位相化フロー、ならびに A、Cおよび G塩基 (Tおよび/または U塩基を含まない) の混合物を含む第2の再位相化フローを、任意の順序で含む。

20

【0146】

シーケンシングクラスター内の伸長プライマーを同期化するための本明細書に記載される方法を、プライマーを伸長するために非終結ヌクレオチドを使用する、合成によるシーケンシングで 사용할 ことができる。一部の実施形態では、方法は、本明細書に記載される他の方法、例えば、本明細書に記載される高速フォワードシーケンシング法 (例えば、「暗」領域を生成するシーケンシング法) と組み合わせて使用される。

カップリングされたシーケンシングリードペアの参照配列へのマッピング

【0147】

カップリングされたシーケンシングリードペアを参照配列にマッピングすることができ、この参照配列は、目的の試験バリエーションを含むこともあり、または含まないこともある。第1の領域または第3の領域についてのシーケンシングデータを使用して、第1の領域または第3の領域の配列をそれぞれ導出することができる。第1の領域、もしくは第1の領域の一部分、または第3の領域、もしくは第3の領域の一部分を、参照配列にマッピングすることができる。第1の領域と第3の領域の間の距離 (すなわち、第2の領域の長さ) を決定または推定することができ、それによって、マッピングされなかった第3または第1の領域についてのおおよその遺伝子座を得ることができる。すると、このおおよその遺伝子座を使用して、マッピングされなかった第1または第3の領域を参照配列に容易にマッピングすることができる。

30

40

【0148】

マッピングされた配列は、ある配列 (例えば、領域の配列またはその一部分) の別の配列 (例えば、参照配列) へのアラインメントを指す。マッピング可能な配列は、選択されたマッピング閾値 (すなわち、マッピングスコア) に従って、別の配列 (例えば、参照配列) にマッピングすることができる配列 (例えば、領域の配列またはその一部分) である。したがって、マッピング不可能な配列は、選択されたマッピング閾値 (マッピングスコア) に従って他の配列にマッピング可能でない配列である。スコアを、エラーリスク許容度に基づいて事前に決定する (すなわち、マッピングの前に選択する) ことができる。例えば、ある配列を別の配列にマッピングする際にスミス・ウォーターマンアルゴリズムを使用することができ、マッピング閾値を、「マッピング可能な」配列と「マッピング不可能

50

な」配列を区別するように選択することができる。例として、マッピングスコアが+1であり、ミスマッチスコアが-1であり、ギャップ開始スコアが-2であり、ギャップ伸長スコアが-2である場合、マッピングスコア閾値は、+5であるかもしくはそれより大きく、+6であるかもしくはそれより大きく、+8であるかもしくはそれより大きく、+10であるかもしくはそれより大きく、+12であるかもしくはそれより大きく、+14であるかもしくはそれより大きく、+16であるかもしくはそれより大きく、+18であるかもしくはそれより大きく、または+20であるかもしくはそれより大きい。他のスコアまたはペナルティスコアが当業者により選択されることもある。

【0149】

カップリングされたシーケンシングリードペアの1つまたは複数の領域などの、配列を、任意の好適なマッピングソフトウェア、例えば、GATK、Bowtie、Bowtie2、BWA、BWA-MEM、Novoalign、SOAP2、SOAP3、およびその他、例えば、バローズ・ホイラー変換（BWT）に基づくアライナーなどを用いてマッピングすることができる。例えば、Miller et al., *Assembly algorithms for next-generation sequencing data*, Genomics, vol. 95, pp. 315-327 (2010); Chaisson et al., *De novo fragment assembly with short mate-paired reads: Does the read length matter?* Genome Research, vol. 19, pp. 336-346 (2009); Mielczarek et al., *Review of alignment and SNP calling algorithms for next-generation sequencing data*, J. Appl. Genetics, vol. 57, pp. 71-79 (2016); Nielsen et al., *Genotype and SNP calling from next-generation sequencing data*, Nature Reviews Genetics, vol. 2, pp. 443-451 (2011); および Hwang et al., *Systematic comparison of variant calling pipelines using gold standard personal exome variants*, Sci Rep., vol. 5, 17875 (2015) を参照されたく、これらの参考文献の各々は、あらゆる目的で参照により本明細書に組み込まれる。

【0150】

参照配列に対するポリヌクレオチドの領域の遺伝子座を概算するための距離情報の使用は、ポリヌクレオチドの第2の領域内の構造バリエーション（例えば、挿入または欠失）の検出に、またはゲノム内の複数のマッピング可能な遺伝子座の分解（例えば、第1の領域もしくは第3の領域が、反復領域もしくは他の非固有配列を含む場合）に、有用である。距離情報は、本明細書で論じられる場合、2点（例えば、領域の始点と終点）間の空間の量に関し、種々の基準枠で考慮され得る。例えば、物理的空間の距離情報は、塩基数または物理的距離（例えば、ポリヌクレオチドが線状に配置された場合、一次元空間でのマイクロメートル数）を指すことがある。シーケンシングデータ空間（例えば、フローグラム空間）の距離情報は、所与のフロー順序を用いて空間内のプライマーを伸長するために使用されたフローステップの数を指すこともある。物理的空間の距離情報およびシーケンシングデータ空間の距離情報は、配列（または参照配列）およびフロー順序が既知である場合、分析的に代替可能である。

【0151】

距離情報は、第2の領域の長さを示すが、必ずしも第2の領域の正確な長さではない。なぜなら、マッピングされなかった領域が距離情報により概算された位置内に最終的にマッピングされるからである。一例では、距離情報は、第2の領域のフロー順序（または第2の領域のフロー順序に関連する情報）および第2の領域における塩基の確率分布を使用し

10

20

30

40

50

て決定される。第2の領域における塩基の確率分布は、例えば、ゲノム全体にわたっての塩基の仮定分布であることがあり、または第1の領域もしくは第3の領域のマッピングされた遺伝子座に基づく、より局所的な確率であることもある。第2の領域のフロー順序に関連する情報は、例えば、第2の領域を通してプライマーを伸長するために同時に使用されるヌクレオチド塩基の異なるタイプの数であり得る。例として、第2の領域内のプライマーを伸長するために反復サイクルで3塩基フローステップを使用して（例えば、各サイクルステップが3つの他の塩基を含む、（非A）－（非C）－（非T）－（非G）のサイクルステップを使用して）、および第2の領域における塩基の分布を全体としてゲノムとほぼ同じと仮定して、プライマーは、サイクルにおけるステップごとにおおよそ4.7塩基ずつ伸長されると予想される。したがって、第2の領域の長さは、第2の領域のフロー順序でのステップの数の4.7倍と概算することができる。

10

【0152】

一部の実施形態では、距離情報は、第2の領域についての予想参照シーケンシングデータから導出される。本明細書中で論じられるように、第2の領域についての予想参照シーケンシングデータを、参照配列および第2の領域のフロー順序を使用して決定することができる。ポリヌクレオチドの第1または第3の領域が参照配列にマッピングされると、予想配列長を含む、予想配列情報が決定され、これにより、ポリヌクレオチドの第1の領域と第3の領域の間の長さが得られる。

【0153】

1つより多くのマッピング可能な位置が参照配列内で入手可能な場合、距離情報を使用して、カップリングされたシーケンシングリードペアを参照配列にマッピングすることができる。例えば、一部の実施形態では、第1の領域を高い信頼度で参照配列にマッピングすることができるが、第3の領域は、参照配列内の複数の異なる位置に位置し得る。一部の実施形態では、第3の領域を高い信頼度で参照配列にマッピングすることができるが、第1の領域は、参照配列内の複数の異なる位置に位置し得る。一部の実施形態では、第1の領域と第3の領域両方を、参照配列内の複数の異なる位置にマッピングすることができる。参照配列にマッピングされた第1の領域および第2の領域についての正確な位置のペアを、第2の領域についての距離情報を使用して選択することができる。例えば、カップリングされたシーケンシングリードペアを参照配列にマッピングする方法は、カップリングされたシーケンシングリードペアの第1の領域（またはその一部分）および第3の領域（またはその一部分）を、参照配列の第1の位置および第2の位置を含む2つまたはそれより多くの異なる位置ペアにマッピングするステップを含み得る。次いで、ポリヌクレオチドの第2の領域の長さを示す距離情報を、第1の位置と第2の位置の間の長さを示す位置情報と比較することができる。比較された距離情報が、互いに近似しているかまたはマッチする場合には、正しい位置ペアが選択され得る。しかし、第2の領域の長さが、第1の位置と第2の位置の間の距離と有意に異なる場合、その位置ペアは、却下され得る。

20

30

【0154】

図3は、カップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用して、カップリングされたシーケンシングリードペアを参照配列にマッピングする方法を示す。カップリングされたシーケンシングリードペア304は、第1の領域306、第2の領域308および第3の領域310を含む。第1の領域306は、参照配列302の第1の参照領域312に位置し得るが、第3の領域310は、第3の参照領域、選択肢A、314と、第3の参照領域、選択肢B、316の両方に位置し得る。第1の参照領域312の終点と、第3の参照領域、選択肢A、314の始点との間の距離は、長さn塩基であり（参照配列に基づいて）、第1の参照領域312の終点と、第3の参照領域、選択肢B、316の始点との間の距離は、長さm塩基である（参照配列に基づいて）。第2の領域についての距離情報は、第2の領域の長さが、長さおおよそn塩基であることを示す。したがって、第3の領域310は、第3の参照領域、選択肢A、314に正確に位置すると、結論付けることができる。たとえ、第1の領域について複数のマッピン

40

50

グ可能な遺伝子座および／または第3の領域についての複数のマッピング可能な遺伝子座があったとしても、同様の解析を遂行することができる。

【0155】

さらに、第1の領域または第3の領域の遺伝子座における反復領域のため、第1の領域または第3の領域を正確な位置に最終的にマッピングすることができない場合、距離情報を使用して、カップリングされたシーケンシングリードペアを参照配列にマッピングすることができる。図4は、カップリングされたシーケンシングリードペアの第3の領域が反復領域に位置する場合、カップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用して、カップリングされたシーケンシングリードペアを参照配列にマッピングする方法を示す。図4は、参照配列402、およびカップリングされたシーケンシングリードペア404を示す。カップリングされたシーケンシングリードペアは、第1の領域406、第2の領域408および第3の領域410を含む。第1の領域406は、第1の参照領域412内の特定の遺伝子座に位置し得るが、第3の領域410は、反復領域414内のどこかに位置し得る。第2の領域408の長さを知ることにより、第3の領域410をより正確に反復領域414内にマッピングすることができる。例えば、第1の領域406がマッピングされてしまえば、第2の領域408の長さが、長さおよそn塩基である場合、この距離情報を使用して、第3の領域410の位置を決めることができる。同様に、第3の領域を正確にマッピングすることができるが、第1の領域が反復領域内に位置する場合、この方法を使用することができる。

構造バリエーションの検出

【0156】

ゲノムに由来するポリヌクレオチドから生成された、カップリングされたシーケンシングリードペアを使用して、ゲノム内の構造バリエーションなどの、バリエーションを検出することができる。構造バリエーションとしては、挿入、欠失、逆位および染色体融合バリエーションを挙げることができ、これらは、ポリヌクレオチドの第1、第2もしくは第3の領域内にあることもあり、またはポリヌクレオチドの第1、第2もしくは第3の領域にまたがる位置にあることもある。

【0157】

ゲノム内への挿入は、任意のサイズ、例えば、長さ1塩基から長さ数百もしくは数千キロ塩基またはそれを超える塩基数の間のものであり得る。さらに、挿入は、内因性挿入（つまり、対象のゲノム内の他の場所が起源である、遺伝子座に挿入された配列）であることもあり、外来性挿入（例えば、対象のゲノムに挿入されたウイルスゲノムなどの、対象ゲノム以外の供給源が起源である、遺伝子座に挿入された配列）であることもある。外来性挿入は、参照配列内に存在しない核酸配列であり、したがって、対象のゲノム内の外来性挿入バリエーションの検出または位置決めについてのさらなる課題をもたらす。本明細書に記載される方法を使用して、数ある構造バリエーションの中でも特に、外来性挿入を検出および／または位置決めすることができる。

【0158】

一例では、カップリングされたシーケンシングリードペアを使用してゲノム内の構造バリエーション（例えば、外来性挿入）を検出する方法は、カップリングされたシーケンシングリードペアの第1の領域（またはその一部分）を参照配列にマッピングするステップ、および第3の領域（またはその一部分）を参照配列にマッピングすることを試みるステップを含む。第3の領域（またはその一部分）がマッピング不可能であった場合には、外来性挿入の存在を同定することができる。これは、参照配列が、第3の領域に対応する配列を含まないからである。同様に、カップリングされたシーケンシングリードペアを使用してゲノム内の外来性挿入を検出する方法は、カップリングされたシーケンシングリードペアの第3の領域（またはその一部分）を参照配列にマッピングするステップ、および第1の領域（またはその一部分）を参照配列にマッピングすることを試みるステップを含むこともある。第1の領域（またはその一部分）がマッピング不可能であった場合には、外来性挿入の存在を同定することができる。これは、参照配列が、第1の領域に対応する

配列を含まないからである。さらに（およびどちらの例でも）、参照配列内の外来性挿入の遺伝子座を、第2の領域の長さを示す予想距離情報に基づいて決定することができる。図5は、外来性挿入を検出する例示的方法についての概略図を示す。カップリングされたシーケンシングリードペア502は、第1の領域504、第2の領域506、および第3の領域508を含み、第2の領域506は、第1の領域504と第3の領域508の間にある。第3の領域508は、対象のゲノム512に存在する外来性挿入エレメント510を含むが、参照配列514には存在しない。参照エレメント516は、対象のゲノム512と参照配列514の両方に存在するが、第1の参照領域518からの間隔が異なる。第1の領域504は、参照配列内の第1の参照領域518に位置する。しかし、第3の領域508には、それが位置する対応する領域が参照配列514内にない（すなわち、それはマッピング不可能である）。これは、第3の領域508の配列が、対象のゲノム内への外来性挿入の結果であることを示す。第2の領域506についての距離情報を使用して、第1の参照領域518に対する外来性ゲノムの遺伝子座を決定することもできる。つまり、第2の領域506が、長さおよそ n 塩基である場合、外来性挿入物は、第1の領域504の末端からおおよそ n 塩基に位置する。

10

20

30

40

50

【0159】

別の例では、カップリングされたシーケンシングリードペアは、予想シーケンシングデータを使用して、および生成シーケンシングデータを予想シーケンシングデータと比較して、構造バリエーション（例えば、挿入、欠失、逆位、または染色体融合）を検出するために、使用することができる。例えば、カップリングされたシーケンシングリードペアの第1の領域（もしくはその一部分）または第3の領域（もしくはその一部分）のうちの一方を参照配列にマッピングすることができる。マッピングされなかった第1の領域（もしくはその一部分）またはマッピングされなかった第3の領域（もしくはその一部分）についての参照配列内の遺伝子座を、第2の領域の長さを示す距離情報を使用して決定することができる。この距離情報は、例えば、本明細書で説明されるように、決定することができる。マッピングされなかった第1の領域（もしくはその一部分）またはマッピングされなかった第3の領域（もしくはその一部分）の遺伝子座が決定されると、その遺伝子座における予想シーケンシングデータ参照配列を決定することができる。例えば、予想配列データを、第2の領域の配列、第2の領域のフロー順序、マッピングされなかった領域の配列に関する情報、およびマッピングされなかった領域のフロー順序に基づいて決定することができる。次いで、予想シーケンシングデータを、マッピングされなかった領域の生成シーケンシングデータと比較することができる。マッピングされなかった領域のシーケンシングデータと予想シーケンシングデータとの間の距離は、その遺伝子座における構造バリエーションを示す。

【0160】

図6は、カップリングされたシーケンシングリードペアを使用して構造バリエーションを検出するための例示的方法を示す。ステップ602で、第1の領域またはその一部分（または第3の領域もしくはその一部分）のうちの一方が参照配列にマッピングされる。ステップ604で、参照配列内の予想遺伝子座が、第3の領域またはその一部分（または第1の領域もしくはその一部分）について決定される。つまり、第1の領域またはその一部分がステップ602中にマッピングされた場合、第3の領域またはその一部分についての予想遺伝子座は、ステップ604で決定され、第3の領域またはその一部分がステップ602中にマッピングされた場合、第1の領域またはその一部分についての予想遺伝子座は、ステップ604で決定される。ステップ606で、第3の領域またはその一部分（または第1の領域もしくはその一部分）についての決定された予想遺伝子座における予想シーケンシングデータが決定される。ステップ608で、第3の領域またはその一部分（または第1の領域もしくはその一部分）についての予想シーケンシングデータが、第3の領域またはその一部分（または第1の領域もしくはその一部分）についての決定シーケンシングデータと比較され、決定シーケンシングデータと予想シーケンシングデータとの差が、構造バリエーションを示す。

【0161】

図7は、対象のゲノムにおける構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは挿入である。対象のゲノム702は、第1の領域704を含み、かつ第1の参照領域708と第2の参照領域710の間に挿入706を含む。参照配列712は、第1の領域704、第1の参照領域708、および第2の参照領域710を含むが、第1の参照領域708と第2の参照領域710の間に挿入706を含まない（挿入は、参照領域の別の部分に見られる領域に対応することもあり、または完全に外来性の配列であることもある）。カップリングされたシーケンシングリードペア714は、第1の領域716（第1の領域704に対応する）および第3の領域718（挿入706に対応する）を含み、これらの領域が第2の領域720を隔てている。カップリングされたシーケンシングリードペア714の第1の領域716は、参照配列712の第1の領域704に位置する。距離情報は、カップリングされたシーケンシングリードペア714の第2の領域720の長さを、長さおおよそn塩基と示す。したがって、第3の領域718についての予想遺伝子座722の始点は、第1の領域704の終点からおおよそn塩基の位置において開始するように決定される。次いで、予想遺伝子座についての予想シーケンシングデータを、本明細書で説明されるように決定することができる。例えば、参照配列712（例えば、第1の領域704から予想遺伝子座の間の、および／または予想遺伝子座を含む、参照配列）、第2の領域についてのフロー順序、および第3の領域についてのフロー順序を使用して、予想遺伝子座についての予想シーケンシングデータを決定することができる。図7に示されている例では、予想シーケンシングデータは、第3の領域718が第2の参照領域710であったなら、第2の参照領域710が予想遺伝子座にあるので得られたであろうシーケンシングデータに対応する。予想遺伝子座についての予想シーケンシングデータが、カップリングされたシーケンシングリードペア714の第3の領域718についての生成シーケンシングデータと異なる場合（これは図7に示されている例の状況である）、構造バリエーションが検出される。

10

20

【0162】

図8は、対象のゲノムにおける構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは欠失である。対象のゲノム802は、第1の領域804、第1の参照領域806および第2の参照領域808を含む。参照配列810は、第1の領域804、第1の参照領域806、および第2の参照領域808を第1の参照領域806と第2の参照領域808の間に位置づけられる追加領域812とともに含む。追加領域812は参照配列810中に存在するが、追加領域812は対象のゲノム802から欠失している。カップリングされたシーケンシングリードペア814は、第1の領域816（第1の領域804に対応する）および第3の領域818（第2の参照領域808に対応する）を含み、これらの領域が第2の領域820を隔てている。カップリングされたシーケンシングリードペア814の第1の領域816は、参照配列810の第1の領域804に位置する。距離情報は、カップリングされたシーケンシングリードペア814の第2の領域820の長さを、長さおおよそn塩基と示す。したがって、第3の領域818についての予想遺伝子座822の始点は、第1の領域804の終点からおおよそn塩基の位置において開始するように決定される。次いで、予想遺伝子座についての予想シーケンシングデータを、本明細書で説明されるように決定することができる。例えば、参照配列812（例えば、第1の領域804から予想遺伝子座の間の、および／または予想遺伝子座を含む、参照配列）、第2の領域についてのフロー順序、および第3の領域についてのフロー順序を使用して、予想遺伝子座についての予想シーケンシングデータを決定することができる。図8に示されている例では、予想シーケンシングデータは、第3の領域818が追加領域812（対象のゲノムにおいて欠失されている）であったなら、追加領域812が予想遺伝子座にあるので得られたであろうシーケンシングデータに対応する。予想遺伝子座についての予想シーケンシングデータが、カップリングされたシーケンシングリードペア814の第3の領

30

40

50

域 8 1 8 についての生成シーケンシングデータと異なる場合（これは図 8 に示されている例の状況である）、構造バリエーションが検出される。

【0163】

図 9 は、対象のゲノム内の構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは逆位である。対象のゲノム 9 0 2 は、第 1 のセグメント 9 0 4、第 2 のセグメント 9 0 6、および第 3 のセグメント 9 0 8 を含む。参照配列 9 1 0 も、第 1 のセグメント 9 0 4、第 2 のセグメント 9 0 6、および第 3 のセグメント 9 0 8 を含む。しかし、参照配列 9 1 0 では、第 2 のセグメント 9 0 6 は、第 3 のセグメント 9 0 8 と比較して 5' 末端の近位にあるが、対象のゲノム 9 0 2 では、第 2 のセグメント 9 0 6 は、第 3 のセグメント 9 0 8 と比較して 3' 末端の近位にある。したがって、対象のゲノム 9 0 2 における第 2 のセグメント 9 0 6 および第 3 のセグメント 9 0 8 は、参照配列 9 1 0 と比較して逆である。カップリングされたシーケンシングリードペア 9 1 2 は、第 1 の領域 9 1 4（第 1 のセグメント 9 0 4 に対応する）および第 3 の領域 9 1 6（第 3 のセグメント 9 0 8 に対応する）を含み、これらの領域が第 2 の領域 9 1 8 を隔てている。カップリングされたシーケンシングリードペア 9 1 2 の第 1 の領域 9 1 4 は、参照配列 9 1 0 の第 1 のセグメント 9 0 4 に位置する。距離情報は、カップリングされたシーケンシングリードペア 9 1 2 の第 2 の領域 9 1 8 の長さを、長さおおよそ n 塩基と示す。したがって、第 3 のセグメント 9 0 8 についての予想遺伝子座 9 2 0 の始点は、第 1 のセグメント 9 0 4 の終点からおおよそ n 塩基の位置において開始するように決定される。次いで、予想遺伝子座についての予想シーケンシングデータを、本明細書で説明されるように決定することができる。例えば、参照配列 9 1 0（例えば、第 1 のセグメント 9 0 4 から予想遺伝子座の間の、および／または予想遺伝子座を含む、参照配列）、第 2 の領域についてのフロー順序、および第 3 の領域についてのフロー順序を使用して、予想遺伝子座についての予想シーケンシングデータを決定することができる。図 9 に示されている例では、予想シーケンシングデータは、第 3 の領域 9 1 6 が第 2 のセグメント 9 0 6 と一致したなら、第 2 のセグメント 9 0 6 が参照配列 9 1 0 における予想遺伝子座にある（第 3 のセグメント 9 0 8 は、そこにない）ので得られたであろうシーケンシングデータに対応する。予想遺伝子座についての予想シーケンシングデータが、カップリングされたシーケンシングリードペア 9 1 2 の第 3 の領域 9 1 6 についての生成シーケンシングデータと異なる場合（これは図 9 に示されている例の状況である）、構造バリエーションが検出される。

【0164】

図 10 は、対象のゲノムにおける構造バリエーションを検出するためのカップリングされたシーケンシングリードペアの使用についての概略図を示し、この場合の構造バリエーションは染色体融合である。染色体融合は、染色体の一部が染色体（同じ染色体または異なる染色体のどちらか）の別の部分に融合する、染色体再編成事象の結果として生じる。参照配列 1 0 0 2 は、第 1 のセグメント 1 0 0 4 および第 2 のセグメント 1 0 0 6 を含む染色体 A と、第 3 のセグメント 1 0 0 8 を含む染色体 B とを含む。対象のゲノム 1 0 1 0 は、参照ゲノム 1 0 0 2 の地点 1 0 1 2 および 1 0 1 4 に染色体 A と染色体 B の染色体融合を含む。この融合により、染色体 A の 3' 末端と染色体 B の 5' 末端とを含む染色体 A/B、および染色体 B の 3' 末端と染色体 A の 5' 末端とを含む染色体 B/A が生じることになる。したがって、染色体 A/B は、第 1 のセグメント 1 0 0 4 および第 3 のセグメント 1 0 0 8 を含み、染色体 B/A は、第 2 のセグメント 1 0 0 6 を含む。カップリングされたシーケンシングリードペア 1 0 1 6 は、対象のゲノム 1 0 1 0 の染色体 A/B から導入され、第 1 の領域 1 0 1 8（第 1 のセグメント 1 0 0 4 に対応する）および第 3 の領域 1 0 2 0（第 3 のセグメント 1 0 0 8 に対応する）を含み、これらの領域が第 2 の領域 1 0 2 2 を隔てている。カップリングされたシーケンシングリードペア 1 0 1 6 の第 1 の領域 1 0 1 8 は、参照配列 1 0 0 2 の第 1 のセグメント 1 0 0 4 に位置する。距離情報は、カップリングされたシーケンシングリードペア 1 0 1 6 の第 2 の領域 1 0 2 2 の長さを、長さおおよそ n 塩基と示す。したがって、第 3 のセグメント 1 0 2 0 についての予想遺伝子

座 1 0 2 4 の始点は、第 1 のセグメント 1 0 0 4 の終点からおおよそ n 塩基の位置において開始するように決定される。次いで、予想遺伝子座についての予想シーケンシングデータを、本明細書で説明されるように決定することができる。例えば、参照配列 1 0 0 2（例えば、第 1 のセグメント 1 0 0 4 から予想遺伝子座、第 2 のセグメント 1 0 0 6 の間の、および／または予想遺伝子座、第 2 のセグメント 1 0 0 6 を含む、参照配列）の染色体 A、第 2 の領域 1 0 2 2 についてのフロー順序、および第 3 の領域 1 0 2 0 についてのフロー順序を使用して、予想遺伝子座についての予想シーケンシングデータを決定することができる。図 1 0 に示されている例では、予想シーケンシングデータは、第 3 の領域 1 0 2 0 が第 2 のセグメント 1 0 0 6 と一致したなら、第 2 のセグメント 1 0 0 6 が参照配列 1 0 0 2 における予想遺伝子座にある（第 3 のセグメント 1 0 0 8 は、そこにはないので得られたであろうシーケンシングデータに対応する。予想遺伝子座についての予想シーケンシングデータが、カップリングされたシーケンシングリードペア 1 0 1 6 の第 3 の 1 0 2 0 についての生成シーケンシングデータとは異なる場合（これは図 1 0 に示されている例の状況である）、構造バリエーションが検出される。

10

【 0 1 6 5 】

参照配列に対して構造バリエーション（例えば、挿入、欠失、染色体融合、または逆位）の接合箇所は、カップリングされたシーケンシングリードペアの第 1 の領域または第 3 の領域全体に必ずしも渡らない。一部の実施形態では、構造バリエーションの少なくとも一部分は、カップリングされたシーケンシングリードペアの第 1 の領域または第 3 の領域内で終わる。それでもやはり、予想シーケンシングデータは、第 1 または第 3 の領域についての決定シーケンシングデータと異なることになる。

20

第 2 の領域内のバリエーションの欠失

【 0 1 6 6 】

一部の実施形態では、カップリングされたシーケンシングリードペアは、たとえ第 2 の領域を通して伸長されたプライマーへのヌクレオチドの組込みを検出する必要がなかったとしても、第 2 の領域内のバリエーションを検出するために使用される。検出可能なバリエーションには、構造バリエーション（例えば、挿入、欠失、逆位、もしくは染色体融合）または一塩基多型（SNP）が含まれる。

【 0 1 6 7 】

構造バリエーション（例えば、染色体融合、逆位、挿入、または欠失）を検出する方法は、カップリングされたシーケンシングリードペアの第 1 の領域（またはその一部分）と第 3 の領域（またはその一部分）の両方を参照配列にマッピングするステップを含み得る。完全に第 2 の領域内で起こる逆位についての距離情報は、一般に、第 2 の領域のフロー順序（例えば、フローグラム空間における）に関して考慮されるが、完全に第 2 の領域では（例えば、少なくとも一部は第 1 の領域でも第 3 の領域でも）起こらない染色体融合、挿入または欠失についての距離情報は、物理的空間または第 2 の領域のフロー順序に関して考慮され得る。参照配列にマッピングされた第 1 の領域と参照配列にマッピングされた第 3 の領域との間の距離情報（すなわち、マッピング距離情報）を決定することができる。マッピング距離情報は、参照配列にマッピングされた第 1 の領域のマッピング位置と、参照配列にマッピングされた第 3 の領域のマッピング位置との間の距離、例えば、第 1 のマッピング領域と第 3 のマッピング領域の間の塩基数を示す。カップリングされたシーケンシングリードペアの第 2 の領域の長さを示す予想距離情報も（例えば、第 2 の領域についてフロー順序および参照配列を使用して、または本明細書に別様に記載されるように）決定することができる。予想距離情報とマッピング距離情報との比較を使用して構造バリエーションを検出することができる。例えば、予想距離がマッピング距離よりも短い場合には、対象のゲノム内の挿入または染色体融合バリエーションなどの構造バリエーションが示される。予想距離がマッピング距離よりも長い場合には、対象のゲノム内の欠失バリエーションが示される。

30

40

【 0 1 6 8 】

図 1 1 は、構造バリエーションを検出する例示的な方法であって、ステップ 1 1 0 2 で、カッ

50

ブリッジされたシーケンシングリードペアの第1の領域（またはその一部分）および第3の領域（またはその一部分）を参照配列にマッピングすることを含む、方法を示す。ステップ1104：参照領域にマッピングされた第1の領域と参照領域にマッピングされた第3の領域の間の距離を示す、マッピング配列距離情報が、決定される。ステップ1106で、第2の領域の予想距離情報が、第2の領域の配列（例えば、参照配列からの第2の領域の配列）についての配列領域のフロー順序および情報に基づいて決定される。ステップ1108で、構造バリエーションは、予想距離情報をマッピング距離情報と比較することにより同定され、マッピング距離情報と予想距離情報との差が構造バリエーションを示す。

【0169】

図12は、カップリングされたシーケンシングリードペアを使用して構造バリエーションを検出することができる方法の一例を実証する概略図を示す。示されている例は、対象のゲノムにおける挿入を示すが、この方法論は、他の構造バリエーション（例えば、欠失または染色体融合）に同様に応用される。参照配列1202は、第1のセグメント1204および第2のセグメント1206を含む。対象のゲノム1208も、第1のセグメント1204および第2のセグメント1206を含むが、第1のセグメント1204と第2のセグメント1206の間に挿入物1210をさらに含む。対象のゲノム1208から生成された、カップリングされたシーケンシングリードペア1212は、第1のセグメント1204に対応する第1の領域1214、および第2のセグメント1206に対応する第3の領域1216を含む。第2の領域1218が第1の領域1214と第3の領域1216を隔てている。第1の領域1214および第3の領域1216の配列を、参照配列1202の第1のセグメント1204および第2のセグメント1206にそれぞれマッピングすることができる。マッピングされると、参照配列1202にマッピングされた第1の領域1214と第3の領域1216の間の距離（すなわち、参照配列1202の第1のセグメント1204と第2のセグメント1206の間の距離）を示すマッピング距離情報が、距離nとして決定される。第2の領域1218の長さについての予想距離情報も、mとして決定することができる。次いで、マッピング距離情報nを予想距離情報mと比較することにより、構造バリエーションを決定することができる。

【0170】

第2の領域内のバリエーション（例えば、構造バリエーションまたはSNP）を検出する別の方法では、予想シーケンシングデータが決定シーケンシングデータと比較される。例えば、一部の実施形態では、カップリングされたシーケンシングリードペアの2つのシーケンシングされた領域間のバリエーションを検出する方法（第1の領域のフロー順序で提供されるヌクレオチドを使用して第1の領域を通して伸長されたプライマー、および／または第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長されたプライマーを用いる）は、第1の領域（もしくはその一部分）および／または第3の領域（もしくはその一部分）を参照配列にマッピングするステップを含む。次いで、他の領域またはその一部分についての予想参照シーケンシングデータ（すなわち、第1の領域またはその一部分がマッピングされる場合、他の領域は、第3の領域またはその一部分を指し、第3の領域またはその一部分がマッピングされる場合、他の領域は、第1の領域またはその一部分を指す）が、決定される。予想シーケンシングデータは、例えば、第2の領域についての参照配列、第2の領域のフロー順序、他の領域またはその一部分（例えば、第1の領域またはその一部分がマッピングされる領域である場合、第3の領域またはその一部分、および第3の領域またはその一部分がマッピングされる領域である場合、第1の領域またはその一部分）についての参照配列、および他の領域またはその一部分についてのフロー順序を使用して、決定することができる。別の例では、予想シーケンシングデータは、第2の領域についての参照配列、第2の領域のフロー順序、他の領域についてのフロー順序、および他の領域の配列に関連するシーケンシングデータ（これは、カップリングされたシーケンシングリードペアを生成する際に生成された同じシーケンシングデータであることもあり、または他の手段により生成されたシーケンシングデータであることもある）を使用して、決定される。他の領域についての決定された予想シ

10

20

30

40

50

ークエンシングデータを、他の領域についての生成シークエンシングデータと比較することができる。予想シークエンシングデータと生成シークエンシングデータとの差は、バリエーションの存在を示す。

【0171】

一部の実施形態では、プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して伸長される、カップリングされたシークエンシングリードペアの2つのシークエンシングされた領域間のバリエーション（例えば、構造バリエーション（例えば、染色体融合、逆位、挿入、もしくは欠失）またはSNP）を検出する方法は、第1の領域またはその一部分を参照配列にマッピングするステップ；（1）第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列、または（2）第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域の配列に関連する生成シークエンシングデータを使用して、第3の領域またはその一部分についての予想シークエンシングデータを決定するステップ；および第3の領域についての予想シークエンシングデータを第3の領域の配列に関連する生成シークエンシングデータと比較することによりバリエーションの存在を検出するステップを含む。一部の実施形態では、プライマーが、第1の領域のフロー順序で提供されるヌクレオチドを使用して伸長される、カップリングされたシークエンシングリードペアの2つのシークエンシングされた領域間のバリエーション（例えば、構造バリエーション（例えば、染色体融合、逆位、挿入、もしくは欠失）またはSNP）を検出する方法は、第3の領域またはその一部分を参照配列にマッピングするステップ；（1）第2の領域についての参照配列、第2の領域のフロー順序、第1の領域のフロー順序、および第1の領域についての参照配列、または（2）第2の領域についての参照配列、第2の領域のフロー順序、第1の領域のフロー順序、および第1の領域の配列に関連する生成シークエンシングデータを使用して、第1の領域またはその一部分についての予想シークエンシングデータを決定するステップ；および第1の領域についての予想シークエンシングデータを第1の領域の配列に関連する生成シークエンシングデータと比較することによりバリエーションの存在を検出するステップを含む。

10

20

【0172】

図13は、カップリングされたシークエンシングリードペアの2つのシークエンシングされた領域間のバリエーションを検出する例示的方法を示す。ステップ1302で、カップリングされたシークエンシングリードペアの第1の領域もしくはその一部分または第3の領域もしくはその一部分が参照配列にマッピングされる。ステップ1304で、第3の領域もしくはその一部分、または第1の領域もしくはその一部分についての予想シークエンシングデータが決定される。ステップ1306で、第1の領域または第3の領域についての予想シークエンシングデータを第1の領域または第3の領域の配列に関連する生成シークエンシングデータと比較することにより、バリエーションの存在が検出される。例示的なバリエーション検出方法は、実施例で提供される。

30

【0173】

バリエーションを検出する方法は、参照配列を使用することができ、この参照配列は、試験バリエーションを含むこともあり、または含まないこともある。試験バリエーションを、例えば、第2のポリヌクレオチド内のまたはバイオマーカーパネルからの試験バリエーションを同定することによって、選択することができる。例として、試験バリエーションを使用して、ポリヌクレオチドのハプロタイプを決定することができる。対立遺伝子またはバリエーションをポリヌクレオチドにおいて同定することができ、本明細書に記載される方法を使用して、カップリングされたシークエンシングリードペアを生じさせたポリヌクレオチドが、同定された対立遺伝子またはバリエーションを有するポリヌクレオチドと同じハプロタイプのものであるのか、異なるハプロタイプのものであるのかを判定することができる。カップリングされたシークエンシングリードペアにおいて検出された試験バリエーションを、ポリヌクレオチドの第1の領域または第3の領域内のシークエンシングされた対立遺伝子と関連付けることができる。

40

50

【0174】

試験バリエーションの存在を検出する場合、参照配列は、試験バリエーションを含むことができ、対象のゲノム内の試験バリエーションの存在を、第3の領域またはその一部分についての予想試験バリエーションシーケンシングデータを第3の領域またはその一部分についての決定シーケンシングデータと比較することにより決定することができる。予想試験バリエーションシーケンシングデータが決定シーケンシングデータにマッチする場合には、試験バリエーションは、参照配列内で検出される。例えば、一部の実施形態では、カップリングされたシーケンシングリードペアの2つのシーケンシングされた領域間の試験バリエーションを検出する方法（第1の領域のフロー順序で提供されるヌクレオチドを使用して第1の領域を通して伸長されたプライマー、および／または第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長されたプライマーを用いる）は、第1の領域またはその一部分を、試験バリエーションを含む参照配列にマッピングするステップを含む。次いで、他の領域またはその一部分（すなわち、第1の領域またはその一部分がマッピングされる場合、他の領域は、第3の領域またはその一部分を指す）についての試験バリエーション予想参照シーケンシングデータが決定される。試験バリエーション予想シーケンシングデータは、例えば、第2の領域についての試験バリエーションを含む参照配列、第2の領域のフロー順序、他の領域またはその一部分についての参照配列、および他の領域またはその一部分についてのフロー順序を使用して、決定することができる。別の例では、予想シーケンシングデータは、第2の領域についての試験バリエーションを有する参照配列、第2の領域のフロー順序、他の領域のフロー順序、および他の領域の配列に関連するシーケンシングデータ（これは、カップリングされたシーケンシングリードペアを生成する際に生成された同じシーケンシングデータであることもあり、または他の手段により生成されたシーケンシングデータであることもある）を使用して、決定される。他の領域についての決定された試験バリエーション予想シーケンシングデータを、他の領域についての生成シーケンシングデータと比較することができる。予想シーケンシングデータと生成シーケンシングデータとのマッチが試験バリエーションの存在を示す。

10

20

短い遺伝子バリエーションの検出

【0175】

本明細書に記載される方法を使用して、第2の領域内の短い遺伝子バリエーション（例えば、SNPまたは短いインデル（長さが、連続する10塩基未満）を（例えば、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、またはプライマーを伸長するために少なくとも2つの異なるタイプのヌクレオチド塩基の混合物を含むことにより、プライマーが第2の領域を通して伸長されると）検出することができる。第2の領域内の短い遺伝子バリエーションを、下流の（例えば、第3の）領域内へのヌクレオチドの組み込みを検出する際に得られるシグナルを解析することにより、検出することができる。短い遺伝子バリエーションは、例えば、個体の亜集団内に見られるバリエーション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ーションもしくは突然変異であることもある。短い遺伝子バリエーションは、生殖細胞系バリエーションであることもあり、または体細胞バリエーションであることもある。

30

【0176】

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5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

40

50

ば、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 5 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 5】

表 5

	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	*	*	*

10

【0 1 7 7】

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フローの中の（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5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。

20

【0 1 7 8】

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5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5において、T-A-C-Gフローサイクル順序を使用してCCG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CCG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。

30

【0 1 7 9】

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WO2019084158A1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、

40

50

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5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15）の配列で伸長されたプライマーは、図14A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。

10

【0180】

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14Bに（図14A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15）。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15）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。

20

【0181】

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。

30

40

【0182】

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

50

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14A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16）の候補プライマー伸長配列を仮定する。図14C（図14A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15）配列のトレース（図14Bを参照されたい）が、図14C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16）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15）にマッチする尤度を示すマッチスコアとには、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14C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。

10

20

【0183】

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。

【0184】

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、バリエーションを有さないものとして分類され得る。

30

【0185】

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

40

50

。例示的な候補シーケンシングデータセットは、図 1 4 C 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 1 6)] 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 5)] 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ーションを有するもの、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ーション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ーションを示した場合）。

10

【0186】

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ーションを有する候補配列（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。短い遺伝子バリエーション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、短い遺伝子バリエーション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、短い遺伝子バリエーションを含まない。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 カ所も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 カ所も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 カ所もしくはそれより多くの、4 カ所もしくはそれより多くの、5 カ所もしくはそれより多くの、6 カ所もしくはそれより多くの、7 カ所もしくはそれより多くの、8 カ所もしくはそれより多くの、9 カ所もしくはそれより多くの、または 10 カ所も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。

20

30

【0187】

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。

40

【0188】

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

50

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15A～15Dに示されている。図15A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15B、図15Cおよび図15D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。

10

【0189】

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。

20

30

【0190】

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、公知のバイズ法を使用して、例えば、ゲノム解析ツールキット（GATK）に含まれているHaplotypeCaller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, 201

40

50

8) ; 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。

【0191】

仮想実施例 1 – SNP 検出

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 14A 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 15) 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 15) (H1) および T A T G G T C A T C G A (配列番号 16) (H2) 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。H1 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、H2 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、H1 の配列をこの核酸分子に選択する。

【0192】

仮想実施例 2 – インデル検出

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 16 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 22) 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 21) (H1) および T A T G G T C G A T C G (配列番号 22) (H2) 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。H1 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、H2 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、H2 の配列をこの核酸分子に選択する。

【0193】

第 2 の (すなわち、「暗」) 領域におけるバリエーションに起因するシグナルの相違が、第 3 の領域 (すなわち、ヌクレオチドの組込みが検出される領域) に伝播した場合、第 2 の領域におけるバリエーションの結果として生じるフローシフトを、第 3 の領域において検出することができる。上記で論じられた仮想例では、例えば、サイクル 3 を「暗」または第 2 の領域 (これは、任意のサイクル数であり得る) と考えることができ、サイクル 4 およびサイクル 5 を第 3 の領域 (これもまた、任意のサイクル数であり得る) と考えることができるだろう。

トランスバージョンの検出

【0194】

トランスバージョンは、プリンをピリミジンと交換する、または逆にピリミジンをプリン

ト交換する、SNPである。本明細書に記載される方法を、カップリングされたシーケンシングリードペアの第2の領域内のトランスバージョンの検出に対して特に感度が良いように、実行することができる。例えば、ピリミジン（C+T）とプリン（A+G）の交互ヌクレオチドペアを含む第2の領域のフロー順序を使用する第2の領域を通したプライマー伸長は、トランスバージョンに対して高感度であろう。

【0195】

例えば、ポリヌクレオチドにおける塩基トランスバージョンの存在を検出するためのカップリングされたシーケンシングリードペアは、（a）ポリヌクレオチド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；（b）ポリヌクレオチドの第1の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；（c）ステップ（b）において伸長されたプライマーを、第2の領域を通して、（1）シトシンおよびチミンと（2）アデニンおよびグアニンとの交互ヌクレオチドペアを含むフロー順序を使用してさらに伸長するステップ；および（d）ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ（c）で伸長されたプライマーをさらに伸長すること、および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により、生成するステップによって生成され得る。トランスバージョンは、第2の領域を通して伸長されたプライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出せずとも検出することができる。

【0196】

トランスバージョン検出のために生成された、カップリングされたシーケンシングリードペアを使用して、カップリングされたシーケンシングリードペアの第1の領域またはその一部分（または第3の領域もしくはその一部分）をマッピングすること；第2の領域のフロー順序、第3の領域のフロー順序、および参照配列を使用して第3の領域またはその一部分（または第1の領域もしくはその一部分）についての予想シーケンシングデータを決定すること；および第3の領域についての予想参照シーケンシングデータと第3の領域についての生成シーケンシングデータの差に基づいて塩基トランスバージョンの存在を検出することにより、トランスバージョンを検出することができる。

【0197】

第3の領域またはその一部分（または第1の領域もしくはその一部分）についての予想参照シーケンシングデータは、例えば、第2の領域のフロー順序、第3の領域のフロー順序、第2の領域についての参照配列、および第3の領域についての参照配列を使用することにより、決定することができる。一部の実施形態では、第3の領域についての予想参照シーケンシングデータは、第2の領域のフロー順序、第3の領域のフロー順序、第2の領域についての参照配列、および第3の領域の配列に関連する生成配列データを使用して決定され、第3の領域の配列に関連する生成配列データは、カップリングされたシーケンシングリードペアを生成した際に生成された同じまたは異なる配列データである。

バリエーション検証

【0198】

複数の少なくとも部分的に重複するカップリングされたシーケンシングリードペアを使用して、バリエーションステータスを検証することができる。シーケンシングエラーは、伸長プライマーへのヌクレオチド組込みの通常の過程で（例えば、ポリメラーゼエラーまたはリードエラーに起因して）起こることがたまにあるので、バリエーション検証は、偽陽性または偽陰性の報告を最小限に抑えるのに有用であり得る。加えて、本明細書に記載される方法の感度は、第2の領域を通してプライマーを伸長する際に使用されるバリエーションのコンテキストおよびフロー順序によって変わり得る。したがって、偽陽性または偽陰性エラーを最小限に抑えるために、重複するまたは少なくとも部分的に重複するカップリングされたシーケンシングリードペアを比較して、バリエーションを検証することができる。バリエーションを検証するために使用される複数のカップリングされたシーケンシングリードペアは

、異なる始点（例えば、異なる第1の領域の始点、異なる第2の領域の始点、および／もしくは異なる第3の領域の始点）を含むことができ、または異なる第2の領域のフロー順序を使用して生成されることもある。

【0199】

目的の試験バリエーションを選択することができ、複数の重複するカップリングされたシーケンシングリードペアが解析されて、カップリングされたシーケンシングリードペアの中の試験バリエーションのステータス（例えば、バリエーションが存在するのか、または非存在であるのか）が判定される。重複するカップリングされたシーケンシングリードペアは、試験バリエーションの遺伝子座に対応する遺伝子座を含む。一部の実施形態では、試験バリエーションは、カップリングされたシーケンシングリードペアの少なくとも一部分の第1の領域内にある。一部の実施形態では、試験バリエーションは、カップリングされたシーケンシングリードペアの少なくとも一部分の第2の領域内にある。一部の実施形態では、試験バリエーションは、カップリングされたシーケンシングリードペアの少なくとも一部分の第3の領域内にある。

10

【0200】

試験バリエーションが遺伝子座に存在するのかまたは非存在であるのかに関するコールを行うための許容度閾値を選択することができる。例えば、所定の閾値が試験バリエーションを同定するよりも、上記複数の重複するカップリングされたシーケンシングリードペアが試験バリエーションを陽性同定する場合、試験バリエーションが陽性コールされる。この閾値は、リスク許容度により所望通りに設定することができる。例えば、許容度閾値は、試験バリエーションを同定するカップリングされたシーケンシングリードペアの60%であるかもしくはそれより高い、70%であるかもしくはそれより高い、80%であるかもしくはそれより高い、90%であるかもしくはそれより高い、または95%であるかもしくはそれより高いことがある。

20

【0201】

図17は、試験バリエーションのステータスを決定するためのカップリングされたシーケンシングリードペアの比較についての例示的概略図を示す。複数の重複するカップリングされたシーケンシングリードペア1402が参照配列1404とアラインメントされている。遺伝子座1406において、5つの重複するカップリングされたシーケンシングリードペアのうちの4つにより、これらのカップリングされたシーケンシングリードペアの1つでは同定されないバリエーションの同定が可能であった。具体的には、カップリングされたシーケンシングリードペア1408、1410、1414および1416は、それぞれ、遺伝子座1418、1420、1424および1426において同定されたバリエーションを含む。各々のカップリングされたシーケンシングリードペアにおけるバリエーションの遺伝子座は、参照配列1404と遺伝子座1406でアラインする。カップリングされたシーケンシングリードペア1412は、遺伝子座1422では（例えば、シーケンシングリードエラーに起因して、またはカップリングされたシーケンシングリードペア1412を生成するために使用された、第2の領域に伴うバリエーションのコンテキスト、およびフロー順序のため）バリエーションを同定しなかった。

30

コンセンサス配列の構築または検証

40

【0202】

本明細書に記載される方法に従って生成された、カップリングされたシーケンシングリードペアを使用して、カップリングされたシーケンシングリードペアをアセンブルすることにより1つまたは複数のコンセンサス配列を生成することができる。コンセンサス配列をアセンブルするためにペアエンドシーケンシングがこれまで使用されてきたが、ポリヌクレオチドのシーケンシングされた末端間の領域について入手可能な情報が限られていることから、誤ってアラインメントされる配列を頻繁に伴う、より低い品質のコンセンサス配列が生じることになる。例えば、あらゆる目的で参照により本明細書に組み込まれる、—および Zerbino et al., Velvet: Algorithm for de novo short read assembly using d

50

e Bruinn graphs, Genome Research, vol. 18, pp. 821-820 (2008)を参照されたい。本明細書に記載される方法は、シーケンシングされた第1の領域とシーケンシングされた第3の領域の間のシーケンシングされなかった第2の領域から、実質的により多くの情報を抽出することを可能にする。この追加情報がより堅牢かつ正確なコンセンサス配列を可能にする。

【0203】

一例では、1つまたは複数のコンセンサス配列は、カップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用してアセンブルされる。この距離情報は、本明細書で説明されるように決定することができる。一例では、距離情報は、第2の領域のフロー順序（または第2の領域のフロー順序に関連する情報）および第2の領域における塩基の確率分布を使用して決定される。第2の領域における塩基の確率分布は、例えば、ゲノム全体にわたっての塩基の仮定分布であることがあり、または第1の領域もしくは第3の領域のマッピングされた遺伝子座に基づく、より局所的な確率であることもある。第2の領域のフロー順序に関連する情報は、例えば、第2の領域を通してプライマーを伸長するために同時に使用されるヌクレオチド塩基の異なるタイプの数であり得る。例として、第2の領域内のプライマーを伸長するために反復サイクルで3塩基フローステップを使用して（例えば、各サイクルステップが3つの他の塩基を含む、（非A）-（非C）-（非T）-（非G）のサイクルステップを使用して）、および第2の領域における塩基の分布を全体としてゲノムとほぼ同じと仮定して、プライマーは、サイクルにおけるステップごとにおおよそ4.7塩基ずつ伸長されると予想される。したがって、第2の領域の長さは、第2の領域のフロー順序でのステップの数の4.7倍と概算することができる。

10

20

【0204】

一部の実施形態では、距離情報は、第2の領域についての予想参照シーケンシングデータから導出される。本明細書中で論じられるように、第2の領域についての予想参照シーケンシングデータを、参照配列および第2の領域のフロー順序を使用して決定することができる。ポリヌクレオチドの第1または第3の領域が参照配列にマッピングされると、予想配列長を含む、予想配列情報が決定され、これにより、ポリヌクレオチドの第1の領域と第3の領域の間の長さが得られる。

【0205】

カップリングされたシーケンシングリードペアを使用して、1つもしくは複数のコンセンサス配列、または1つもしくは複数のコンセンサス配列の一部を検証することができる。コンセンサス配列アセンブリーの結果として、利用可能なデータが得られる可能性のある複数の配列アセンブリーを得ることができるが、旧来のペアエンドシーケンシングデータを使用してこれらの可能な配列のどれが正しいコンセンサス配列であるのかを選択することは、困難であり得る。カップリングされたシーケンシングリードペアの第2の領域から追加情報を抽出できるため、本明細書に記載される方法を使用するとコンセンサス配列検証がより頑強になる。コンセンサス配列を検証するために、第1の領域またはその一部分（または第3の領域もしくはその一部分）を選択された参照配列にマッピングすることができる。他の領域またはその一部分についての予想シーケンシングデータ（すなわち、第1の領域もしくはその一部分がマッピングされる場合、第3の領域もしくはその一部分、または第3の領域もしくはその一部分がマッピングされる場合、第1の領域もしくはその一部分）。予想シーケンは、例えば、本明細書で説明されるように、決定することができる。一例では、予想シーケンシングデータは、第2の領域のフロー順序、選択されたコンセンサス配列、および第1の領域のフロー順序（予想シーケンシングデータが第1の領域もしくはその一部分についてのものである場合）または第3の領域のフロー順序（予想シーケンシングデータが第3の領域もしくはその一部分についてのものである場合）を使用して、決定される。次いで、予想シーケンシングデータを、対応する領域におけるカップリングされたシーケンシングリードペアについての生成シーケンシングデータと比較して、コンセンサス配列部分を検証することができる。生成シ

30

40

50

ークエンシングデータにマッチする予想シーケンシングデータは、コンセンサス配列部分が正しくアセンブルされることを示す。生成シーケンシングデータにマッチしない予想シーケンシングデータは、コンセンサス配列部分が誤ってアセンブルされることを示す。

【0206】

一部の実施形態では、1つより多くのコンセンサス配列が構築または検証される。例えば、ある特定の生物は、倍数体生物（例えば、健常なヒトは、二倍体生物であり、各染色体（ヒト男性における性染色体を除く）の2コピーを有する。1つまたは複数の染色体コピーに対応するコンセンサス配列をアセンブルすることができる（例えば、ヒト配列における染色体ペアごとにコンセンサス配列をアセンブルしてもよい）。カップリングされたシーケンシングリードペアを倍数体生物の対応する染色体に割り当てるプロセスは、ハプロタイプ判定と呼ばれることがある。本明細書に記載される方法を使用して、ハプロタイプ判定の精度または効率を向上させることができる。例えば、本明細書に記載されるカップリングされたシーケンシングリードペアの第2の領域からの情報を使用して、試験バリエーションを第1の染色体または第2の染色体（または倍数体生物からの他のさらなる染色体）と関連付けることができる。

10

システム、デバイスおよびレポート

【0207】

図1～17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18に描かれている構成要素により、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18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18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。

20

【0208】

図18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1800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1800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18に示されているように、デバイス1800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1810、入力デバイス1820、出力デバイス1830、記憶装置1840、および通信デバイス1860のうちの1つまたは複数を含み得る。入力デバイス1820および出力デバイス1830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。

30

【0209】

入力デバイス1820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1830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。

40

【0210】

記憶装置1840は、RAM、キャッシュメモリ、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1860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。

50

【0211】

記憶装置1840に記憶され、プロセッサ1810により実行され得る、ソフトウェア1850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ような）プログラミングを含むことができる。

【0212】

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1850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1840であり得る。

10

【0213】

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1850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。

20

【0214】

デバイス1800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。

30

【0215】

デバイス1800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1850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なウェブブラウザによって、展開することができる。

【0216】

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。

40

例示的实施形態

50

【0217】

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。

【0218】

実施形態1．カップリングされたシーケンシングリードペアをポリヌクレオチドから生成する方法であって、

(a) ポリヌクレオチド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；

(b) ポリヌクレオチドの第1の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；

(c) ステップ(b)で伸長されたプライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用して第2の領域を通してさらに伸長するステップであって、(i) プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第2の領域を通して伸長されるか、(ii) 少なくとも2つの異なるタイプのヌクレオチド塩基の混合物が、第2の領域のフロー順序の少なくとも1ステップで使用されるか、または(iii) 第2の領域を通したプライマーの伸長が、ステップ(b)におけるプライマーの伸長よりも速く進行する、ステップ；および

(d) ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(c)で伸長されたプライマーをさらに伸長すること、および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により、生成するステップ

を含む方法。

【0219】

実施形態2．第2の領域を通したプライマーの伸長が、第1の領域を通したプライマーの伸長より速く進行する、実施形態1の方法。

【0220】

実施形態3．第1の領域のシーケンシングデータを第3のシーケンシングデータと関連付けるステップをさらに含む、実施形態1または2の方法。

【0221】

実施形態4．カップリングされたシーケンシングリードペアをポリヌクレオチドから生成する方法であって、

(a) プライマーをポリヌクレオチドの第1の領域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；

(b) プライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用して第2の領域を通して伸長するステップであって、(i) プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第2の領域を通して伸長される、または(ii) 少なくとも2つの異なるタイプのヌクレオチド塩基の混合物が、第2の領域のフロー順序の少なくとも1ステップで使用される、ステップ；および

(c) ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(b)で伸長されたプライマーをさらに伸長すること、および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により、生成するステップ

を含む方法。

【0222】

実施形態5．第1の領域が、プライマーの標的にされる天然に存在する配列を含む、実施形態4の方法。

【0223】

実施形態6．プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在ま

10

20

30

40

50

たは非存在を検出することなく、第2の領域を通して伸長される、実施形態1から5のいずれか1つの方法。

【0224】

実施形態7. 第2の領域を通してプライマーを伸長するために使用されるヌクレオチドの少なくとも一部分が、非標識ヌクレオチドである、実施形態1から6のいずれか1つの方法。

【0225】

実施形態8. 第2の領域を通してプライマーを伸長するために使用されるヌクレオチドが、非標識ヌクレオチドである、実施形態1から6のいずれか1つの方法。

【0226】

実施形態9. 少なくとも2つの異なるタイプのヌクレオチド塩基の混合物が、第2の領域のフロー順序の少なくとも1ステップで使用される、実施形態1から8のいずれか1つの方法。

【0227】

実施形態10. 第2の領域のフロー順序が、5つまたはそれより多くのヌクレオチドフローを含む、実施形態1から9のいずれか1つの方法。

【0228】

実施形態11. ヌクレオチドフローの各々が、単一のヌクレオチド塩基を含む、実施形態10の方法。

【0229】

実施形態12. 第2の領域のフロー順序が、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10または11の方法。

【0230】

実施形態13. 第2の領域のフロー順序が、1フロー当たり0.6のまたはそれを超える塩基組込みの効率を有する、実施形態10から12のいずれか1つの方法。

【0231】

実施形態14. 参照配列および第2の領域のフロー順序を使用して第2の領域についての予想シーケンシングデータを決定するステップをさらに含む、実施形態1から13のいずれか1つの方法。

【0232】

実施形態15. プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長され、方法が、第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列を使用して、第3の領域についての予想シーケンシングデータを決定するステップをさらに含む、実施形態1から14のいずれか1つの方法。

【0233】

実施形態16. 第3の領域のフロー順序が、5つまたはそれより多くのヌクレオチドフローを含む、実施形態15の方法。

【0234】

実施形態17. ヌクレオチドフローの各々が、単一のヌクレオチド塩基を含む、実施形態16の方法。

【0235】

実施形態18. 第3の領域のフロー順序が、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16または17の方法。

【0236】

実施形態19. 第3の領域のフロー順序が、1フロー当たり0.6のまたはそれを超える塩

10

20

30

40

50

基組込みの効率を有する、実施形態 16 から 18 のいずれか 1 つの方法。

【0237】

実施形態 20 . プライマーが、第 3 の領域のフロー順序で提供されるヌクレオチドを使用して第 3 の領域を通して伸長され、方法が、第 2 の領域についての参照配列、第 2 の領域のフロー順序、第 3 の領域のフロー順序、および第 3 の領域の配列に関連するシーケンシングデータを使用して、第 3 の領域についての予想シーケンシングデータを決定するステップをさらに含み、第 3 の領域の配列に関連するシーケンシングデータが、第 3 の領域について生成される同じまたは異なるシーケンシングデータである、実施形態 1 から 19 のいずれか 1 つの方法。

【0238】

実施形態 21 . 第 2 の領域または第 3 の領域についての予想参照データが、バイナリまたは非バイナリプログラムを含む、実施形態 14 から 20 のいずれか 1 つの方法。

【0239】

実施形態 22 . 第 2 の領域のフロー順序、および第 2 の領域についての第 2 の参照配列を使用して、第 2 の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、第 2 の参照配列が、試験バリエーションを含む、実施形態 14 から 21 のいずれか 1 つの方法。

【0240】

実施形態 23 . プライマーが、第 3 の領域のフロー順序で提供されるヌクレオチドを使用して第 3 の領域を通して伸長され、方法が、第 2 の領域についての第 2 の参照配列、第 2 の領域のフロー順序、第 3 の領域のフロー順序、および第 3 の領域についての参照配列を使用して、第 3 の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含む、実施形態 22 の方法。

【0241】

実施形態 24 . プライマーが、第 3 の領域のフロー順序で提供されるヌクレオチドを使用して第 3 の領域を通して伸長され、方法が、第 2 の領域についての第 2 の参照配列、第 2 の領域のフロー順序、第 3 の領域のフロー順序、および第 3 の領域の配列に関連するシーケンシングデータを使用して、第 3 の領域についての予想試験バリエーションシーケンシングデータを決定するステップをさらに含み、第 3 の領域の配列に関連するシーケンシングデータが、第 3 の領域について生成される同じまたは異なるシーケンシングデータである、実施形態 22 の方法。

【0242】

実施形態 25 . 第 2 の領域または第 3 の領域についての予想参照シーケンシングデータが、バイナリまたは非バイナリプログラムを含む、実施形態 22 から 24 のいずれか 1 つの方法。

【0243】

実施形態 26 . カップリングされたシーケンシングリードペアを参照配列にマッピングする方法であって、

実施形態 1 から 25 のいずれか 1 つの方法に従って生成された、カップリングされたシーケンシングリードペアの第 1 の領域もしくはその一部分または第 3 の領域もしくはその一部分を、参照配列にマッピングするステップ；および

マッピングされなかった第 1 の領域もしくはその一部分、またはマッピングされなかった第 3 の領域もしくはその一部分を、第 2 の領域の長さを示す距離情報を使用して、参照配列にマッピングするステップを含む、方法。

【0244】

実施形態 27 . 構造バリエーションを検出する方法であって、

実施形態 1 から 25 のいずれか一項に記載の方法に従って生成された、カップリングされたシーケンシングリードペアの第 1 の領域もしくはその一部分または第 3 の領域もしくはその一部分を、参照配列にマッピングするステップ；

10

20

30

40

50

マッピングされなかった第1の領域もしくはその一部分またはマッピングされなかった第3の領域もしくはその一部分についての参照配列内の予想遺伝子座を、第2の領域の長さを示す距離情報を使用して決定するステップ；

参照配列に基づいて予想遺伝子座における配列についての予想シーケンシングデータを決定するステップ；および

マッピングされなかった第1の領域もしくはその一部分またはマッピングされなかった第3の領域もしくはその一部分を予想シーケンシングデータと比較することにより構造バリエーションを検出するステップであって、マッピングされなかった第1の領域もしくはその一部分またはマッピングされなかった第3の領域もしくはその一部分と予想シーケンシングデータとの差が、構造バリエーションを示す、ステップを含む方法。

10

【0245】

実施形態28．構造バリエーションを検出する方法であって、

実施形態1から25のいずれか1つの方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分または第3の領域もしくはその一部分を、参照配列にマッピングするステップを含み、マッピングされなかった第1の領域、またはマッピングされなかった第3の領域が、参照配列内にマッピング不可能である、方法。

【0246】

実施形態29．第2の領域の長さを示す予想距離情報に基づいて参照配列内の構造バリエーションの遺伝子座を決定するステップをさらに含む、実施形態28の方法。

20

【0247】

実施形態30．マッピングされなかった第1の領域もしくはその一部分、またはマッピングされなかった第3の領域もしくはその一部分が、参照配列と比較して挿入の中に存在する、実施形態27から29のいずれか1つの方法。

【0248】

実施形態31．マッピングされなかった第1の領域もしくはその一部分、またはマッピングされなかった第3の領域もしくはその一部分が、参照配列と比較して挿入の始点または終点にまたがる、実施形態27から29のいずれか1つの方法。

【0249】

実施形態32．構造バリエーションを検出する方法であって、

実施形態1から25のいずれか1つの方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分および第3の領域もしくはその一部分を、参照配列にマッピングするステップ；

マッピングされた第1の領域とマッピングされた第3の領域の間のマッピング距離情報を決定するステップ；および

マッピング距離情報を第2の領域の予想距離情報と比較することにより構造バリエーションを検出するステップであって、マッピング距離情報と予想距離情報との差が構造バリエーションを示す、ステップ

を含む方法。

40

【0250】

実施形態33．構造バリエーションが、染色体融合、逆位、挿入、または欠失である、実施形態27から32のいずれか1つの方法。

【0251】

実施形態34．バリエーションが、第2の領域内の挿入または欠失である、実施形態27から32のいずれか1つの方法。

【0252】

実施形態35．距離情報が、第2の領域のフロー順序に関連する情報および第2の領域における塩基の確率分布を使用して決定される、実施形態26から32のいずれか1つの方法。

50

【0253】

実施形態36．第2の領域のフロー順序に関連する情報が、第2の領域を通してプライマーを伸長するために同時に使用されるヌクレオチド塩基の異なるタイプの数である、実施形態35の方法。

【0254】

実施形態37．第2の領域における塩基の確率分布が、ゲノム内の塩基の分布から決定される、実施形態35または36の方法。

【0255】

実施形態38．距離情報が、参照配列および第2の領域のフロー順序を使用して決定された第2の領域についての予想シーケンシングデータから導出される、実施形態26～35のいずれか1つの方法。

【0256】

実施形態39．予想シーケンシングデータが、バイナリまたは非バイナリフローグラムを含む、実施形態38の方法。

【0257】

実施形態40．カップリングされたシーケンシングリードペアを参照配列にマッピングする方法であって、

実施形態1から25のいずれか1つの方法に従って生成された、カップリングされたシーケンシングリードペアの第1の領域もしくはその一部分および第3の領域もしくはその一部分を、参照配列の第1の位置および第2の位置を含む2つまたはそれより多くの異なる位置ペアにマッピングするステップ；および

第2の領域の長さを示す第1の距離情報、および2つまたはそれより多くの位置ペアについての第1の位置と第2の位置の間の距離を示す第2の距離情報を使用して、正しい位置を選択するステップ

を含む方法。

【0258】

実施形態41．距離情報が、第2の領域のフロー順序に関連する情報および第2の領域における塩基の確率分布を使用して決定される、実施形態40の方法。

【0259】

実施形態42．第2の領域のフロー順序に関連する情報が、第2の領域を通してプライマーを伸長するために同時に使用されるヌクレオチド塩基の異なるタイプの数である、実施形態41の方法。

【0260】

実施形態43．第2の領域における塩基の確率分布が、ゲノム内の塩基の分布から決定される、実施形態41または42の方法。

【0261】

実施形態44．距離情報が、参照配列および第2の領域のフロー順序を使用して決定された第2の領域についての予想シーケンシングデータから導出される、実施形態40の方法。

【0262】

実施形態45．予想参照シーケンシングデータが、バイナリまたは非バイナリフローグラムを含む、実施形態44の方法。

【0263】

実施形態46．伸長されたプライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長される、実施形態1から25のいずれか1つに従って生成された、カップリングされたシーケンシングリードペアの2つのシーケンシングされた領域間のバリエーションを検出する方法であって、

第1の領域またはその一部分を参照配列にマッピングするステップ；

(1) 第2の領域についての参照配列、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域についての参照配列、または(2) 第2の領域についての参照配列

10

20

30

40

50

、第2の領域のフロー順序、第3の領域のフロー順序、および第3の領域の配列に関連する生成シーケンシングデータを使用して、第3の領域またはその一部分についての予想シーケンシングデータを決定するステップであって、第3の領域の配列に関連する生成配列データが、第3の領域について生成された同じまたは異なる配列データである、ステップ；および

第3の領域についての予想シーケンシングデータを第3の領域の配列に関連する生成シーケンシングデータと比較することによりバリエーションの存在を検出するステップを含む方法。

【0264】

実施形態47．バリエーションが、構造バリエーションである、実施形態46の方法。

【0265】

実施形態48．構造バリエーションが、染色体融合、逆位、挿入、または欠失である、実施形態47の方法。

【0266】

実施形態49．バリエーションが、一塩基多型（SNP）である、実施形態46の方法。

【0267】

実施形態50．試験バリエーションを検出するために使用され、参照配列が試験バリエーションを含む、実施形態46から49のいずれか1つの方法。

【0268】

実施形態51．試験バリエーションが、第2のポリヌクレオチド中の試験バリエーションを同定することにより選択される、実施形態50の方法。

【0269】

実施形態52．検出された試験バリエーションとポリヌクレオチドの第1の領域または第3の領域におけるシーケンシングされた対立遺伝子とを関連付けるステップを含む、実施形態50または51の方法。

【0270】

実施形態53．ポリヌクレオチドのシーケンシングされなかった領域における塩基トランスバージョンの存在を検出するためのカップリングされたシーケンシングリードペアを生成する方法であって、

（a）ポリヌクレオチド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）ポリヌクレオチドの第1の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；

（c）ステップ（b）において伸長されたプライマーを、第2の領域を通して、（1）シトシンおよびチミンと（2）アデニンおよびグアニンとの交互ヌクレオチドペアを含むフロー順序を使用してさらに伸長するステップ；および

（d）ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ（c）で伸長されたプライマーをさらに伸長すること、および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により、生成するステップ

を含む方法。

【0271】

実施形態54．カップリングされたシーケンシングリードペアをポリヌクレオチドから生成する方法であって、

（a）プライマーをポリヌクレオチドの第1の領域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）プライマーを、第2の領域を通して、（1）シトシンおよびチミンと（2）アデニンおよびグアニンとの交互ヌクレオチドペアを含むフロー順序を使用して伸長するステップ；および

10

20

30

40

50

(c) ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してステップ(b)で伸長されたプライマーをさらに伸長すること、および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により、生成するステップを含む方法。

【0272】

実施形態55. 第1の領域が、プライマーの標的にされる天然に存在する配列を含む、実施形態54の方法。

【0273】

実施形態56. プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、第2の領域を通して伸長される、実施形態54または55の方法。

【0274】

実施形態57. ポリヌクレオチドのシーケンシングされなかった領域における塩基トランスバージョンの存在を検出する方法であって、

プライマーが、第3の領域のフロー順序で提供されるヌクレオチドを使用して第3の領域を通して伸長される、実施形態54から56のいずれか1つに従って生成された、カップリングされたシーケンシングリードペアの第1の領域またはその一部分および第3の領域またはその一部分を、参照配列にマッピングするステップ；

第2の領域のフロー順序、第3の領域のフロー順序、および参照配列を使用して、第3の領域についての予想シーケンシングデータを決定するステップ；および

第3の領域についての予想シーケンシングデータと第3の領域についての生成シーケンシングデータとの差に基づいて塩基トランスバージョンの存在を検出するステップを含む方法。

【0275】

実施形態58. 第3の領域についての予想シーケンシングデータが、第2の領域のフロー順序、第3の領域のフロー順序、第2の領域についての参照配列、および第3の領域についての参照配列を使用して決定される、実施形態57の方法。

【0276】

実施形態59. 第3の領域についての予想シーケンシングデータが、第2の領域のフロー順序、第3の領域のフロー順序、第2の領域についての参照配列、および第3の領域の配列に関連する生成配列データを使用して決定され、第3の領域の配列に関連する生成配列データが、第3の領域について生成された同じまたは異なる配列データである、実施形態57の方法。

【0277】

実施形態60. 第3の領域についての予想シーケンシングデータが、バイナリまたは非バイナリプログラムを含む、実施形態57から59のいずれか1つの方法。

【0278】

実施形態61. 1つまたは複数のコンセンサス配列を生成する方法であって、実施形態1から25のいずれか1つに従って生成された複数のカップリングされたシーケンシングリードペアをアセンブルするステップを含む、方法。

【0279】

実施形態62. 1つまたは複数のコンセンサス配列が、複数のカップリングされたシーケンシングリードペアの第2の領域の長さを示す距離情報を使用してアセンブルされる、実施形態61の方法。

【0280】

実施形態63. 距離情報が、第2の領域のフロー順序に関連する情報および第2の領域における塩基の確率分布を使用して決定される、実施形態61の方法。

【0281】

実施形態64. 第2の領域のフロー順序に関連する情報が、第3の領域を通してプライマ

10

20

30

40

50

ーを伸長するために同時に使用されるヌクレオチド塩基の異なるタイプの数である、実施形態 6 3 の方法。

【0282】

実施形態 6 5 . 第 2 の領域における塩基の確率分布が、ゲノム内の塩基の分布から決定される、実施形態 6 3 または 6 4 の方法。

【0283】

実施形態 6 6 . 距離情報が、参照配列および第 2 の領域のフロー順序を使用して決定された第 2 の領域についての予想参照シーケンシングデータから導出される、実施形態 6 2 の方法。

【0284】

実施形態 6 7 . 予想参照シーケンシングデータが、バイナリまたは非バイナリフローグラムを含む、実施形態 6 6 の方法。

【0285】

実施形態 6 8 . 1 つまたは複数のコンセンサス配列から選択されたコンセンサス配列の一部を、選択されたコンセンサス配列の一部に関連する、選択された、カップリングされたシーケンシングリードを使用して検証するステップをさらに含み、選択された、カップリングされたシーケンシングリードを生成する際に第 3 の領域を通して伸長されるプライマーが、第 3 の領域のフロー順序で提供されるヌクレオチドを使用して伸長され、検証するステップが、

第 2 の領域のフロー順序、第 3 の領域のフロー順序、および選択されたコンセンサス配列の一部を使用して、選択された、カップリングされたシーケンシングリードの第 3 の領域についての予想シーケンシングデータを決定すること；および

選択された、カップリングされたシーケンシングリードの第 3 の領域についての予想シーケンシングデータを第 3 の領域の生成シーケンシングデータと比較することにより、選択されたコンセンサス配列の一部を検証することを含む、実施形態 6 1 から 6 7 のいずれか 1 つの方法。

【0286】

実施形態 6 9 . 試験バリエーションのステータスを検証する方法であって、

実施形態 1 から 2 5 のいずれか 1 つに従って生成された複数の重複するカップリングされたシーケンシングリードペアに亘るバリエーションのステータスを比較するステップであって、複数の重複するカップリングされたシーケンシングリードペアが、試験バリエーションの遺伝子座に対応する遺伝子座を含む、ステップ；

比較に基づいてバリエーションのステータスを検証するステップを含む方法。

【0287】

実施形態 7 0 . 選択された、カップリングされたシーケンシングリードペアの、第 1 の領域または第 3 の領域が、複数の重複するカップリングされたシーケンシングリードのうち他のカップリングされたシーケンシングリードの少なくとも一部分の第 2 の領域と重複する、実施形態 6 9 の方法。

【0288】

実施形態 7 1 . 選択された、カップリングされたシーケンシングリードの、バリエーションステータスが、選択された、カップリングされたシーケンシングリードの、第 1 の領域または第 3 の領域におけるバリエーションを示す、実施形態 6 9 または 7 0 の方法。

【0289】

実施形態 7 2 . 選択された、カップリングされたシーケンシングリードの、第 2 の領域が、複数の重複するカップリングされたシーケンシングリードのうち他のカップリングされたシーケンシングリードの少なくとも一部分の第 2 の領域と重複する、実施形態 7 1 の方法。

【0290】

実施形態 7 3 . 選択された、カップリングされたシーケンシングリードの、バリエーション

10

20

30

40

50

ステータスが、選択された、カップリングされたシーケンシングリードの、第 2 の領域におけるバリエーションを示す、実施形態 7 1 または 7 2 の方法。

【0291】

実施形態 7 4 . 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、実施形態 1 から 2 5 のいずれか 1 つのカップリングされたシーケンシングリードペアを生成するステップ；

ポリヌクレオチドの第 3 の領域の配列に関連するシーケンシングデータをポリヌクレオチドの第 3 の領域の予想配列についての予想シーケンシングデータと比較するステップ；および

ポリヌクレオチドの第 2 の領域における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。

【0292】

実施形態 7 5 .

ポリヌクレオチドの第 3 の領域の配列に関連するシーケンシングデータをポリヌクレオチドの第 3 の領域についての予想シーケンシングデータと比較するステップが、ポリヌクレオチドの第 3 の領域について生成されたシーケンシングデータがポリヌクレオチドの第 3 の領域についての予想シーケンシングデータにマッチする尤度を示すマッチスコアを決定すること；および

決定されたマッチスコアを使用してポリヌクレオチドの第 2 の領域における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こと

を含む、実施形態 7 4 の方法。

【0293】

実施形態 7 6 . ポリヌクレオチドの第 3 の領域についての予想シーケンシングデータが、ポリヌクレオチドの第 3 の領域の予想配列を *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 4 または 7 5 の方法。

【0294】

実施形態 7 7 . 第 1 の領域の配列に関連するシーケンシングデータ、または第 3 の領域の配列に関連するシーケンシングデータが、複数のフロー位置の中の各フロー位置に取り込まれた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、実施形態 1 から 7 6 のいずれか 1 つの方法。

【0295】

実施形態 7 8 . フローシグナルが、各フロー位置における少なくとも 1 つの塩基カウントについての塩基カウント尤度を示す統計パラメーターを含む、実施形態 7 7 の方法。

【0296】

実施形態 7 9 . 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塩基カウント尤度を示す統計パラメーターを含む、実施形態 7 8 の方法。

【0297】

実施形態 8 0 .

第 3 の領域の配列に関連するシーケンシングデータが、複数のフロー位置の中の各フロー位置に取り込まれた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み、フローシグナルが、複数の塩基カウントについての塩基カウント尤度を示す統計パラメーターを含み；

方法が、シーケンシングデータ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予想配列の塩基カウントと一致する統計パラメーターを選択するステップ、およびシーケンシングデータセットが予想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をさらに含む、

実施形態 7 5 または 7 6 の方法。

【0298】

実施形態 8 1 . マッチスコアが、シーケンシングデータ内のフロー位置にわたっての選

10

20

30

40

50

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 80 の方法。

【0299】

実施形態 82 . 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 81 のいずれか 1 つの方法。

【0300】

実施形態 83 . 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 81 のいずれか 1 つの方法。

【0301】

実施形態 84 . カップリングされたシーケンシングリードペアを生成するステップが、プライマーを、第 4 の領域のフロー順序で提供されるヌクレオチドを使用して第 4 の領域を通して伸長すること、ここで、(i) プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、第 4 の領域を通して伸長されるか、(ii) 少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、第 4 の領域のフロー順序の少なくとも 1 ステップで使用されるか、または (iii) 第 4 の領域を通したプライマーの伸長が、第 1 の領域または第 3 の領域を通したプライマーの伸長よりも速く進行する；および

10

ポリヌクレオチドの第 5 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用する第 4 を通して伸長されたプライマーのさらなる伸長、および取り込まれた標識ヌクレオチドの存在または非存在の検出により、生成すること

20

を含む、実施形態 1 から 83 のいずれか 1 つの方法。

【0302】

実施形態 85 . 第 5 の領域のシーケンシングデータを第 1 の領域のシーケンシングデータまたは第 3 の領域のシーケンシングデータと関連付けるステップをさらに含む、実施形態 84 の方法。

【0303】

実施形態 86 . ポリヌクレオチドが、ローリングサークル増幅を使用して増幅される、実施形態 1 から 85 のいずれか 1 つの方法。

【0304】

実施形態 87 . 試験試料中の短い遺伝子バリエーションを検出する方法であって、

30

(a) ローリングサークル増幅 (RCA) を使用してポリヌクレオチドを増幅させて、ポリヌクレオチドの第 1 のコピーとポリヌクレオチドの第 2 のコピーとを少なくとも含む RCA 増幅ポリヌクレオチドを生成するステップ；

(b) RCA 増幅ポリヌクレオチド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；

(c) ポリヌクレオチドの第 1 のコピーの中のポリヌクレオチドの第 1 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；

40

(d) プライマーを、第 2 の領域のフロー順序で提供されるヌクレオチドを使用してポリヌクレオチドの第 1 のコピーの中のポリヌクレオチドの第 2 の領域を通してさらに伸長するステップであって、(i) プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、ポリヌクレオチドの第 1 のコピーの中のポリヌクレオチドの第 2 の領域を通して伸長されるか、(ii) 少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、第 2 の領域のフロー順序の少なくとも 1 ステップで使用されるか、または (iii) ポリヌクレオチドの第 1 のコピーの中のポリヌクレオチドの第 2 の領域を通したプライマーの伸長が、第 1 の領域を通したプライマーの伸長よりも速く進行する、ステップ；

(e) ポリヌクレオチドの第 3 の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーをさらに伸長すること、および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により、生成するステップ；

50

(f) ポリヌクレオチドの第3の領域について生成されたシーケンシングデータをポリヌクレオチドの第3の領域の予想配列についての予想シーケンシングデータと比較するステップ；

(g) ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップ；

(h) ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；および

(i) ポリヌクレオチドの第2の領域における短い遺伝子バリエーションのアイデンティティをコールするステップを含む方法。

【0305】

実施形態88．ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通したプライマーの伸長が、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第1の領域を通したプライマーの伸長よりも速く進行する、実施形態87の方法。

【0306】

実施形態89．試験試料中の短い遺伝子バリエーションを検出する方法であって、

(a) ローリングサークル増幅(RCA)を使用してポリヌクレオチドを増幅させて、ポリヌクレオチドの第1のコピーとポリヌクレオチドの第2のコピーとを少なくとも含むRCA増幅ポリヌクレオチドを生成するステップ；

(b) プライマーをポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第1の領域に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；

(c) プライマーを、第2の領域のフロー順序で提供されるヌクレオチドを使用してポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長するステップであって、(i) プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在もしくは非存在を検出することなく、ポリヌクレオチドの第1のコピーの中のポリヌクレオチドの第2の領域を通して伸長される、ステップ；

(d) ポリヌクレオチドの第3の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーをさらに伸長すること、および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により、生成するステップ；

(e) ポリヌクレオチドの第3の領域について生成されたシーケンシングデータをポリヌクレオチドの第3の領域の予想配列についての予想シーケンシングデータと比較するステップ；

(f) ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップ；

(g) ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータを、標識ヌクレオチドを使用してプライマーを伸長すること、および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により、生成するステップ；および

(h) ポリヌクレオチドの第2の領域における短い遺伝子バリエーションのアイデンティティをコールするステップを含む方法。

【0307】

実施形態90．第1の領域が、プライマーの標的にされる天然に存在する配列を含む、実施形態89の方法。

【0308】

実施形態91．ポリヌクレオチドの第2のコピーの中のポリヌクレオチドの第2の領域の配列に関連するシーケンシングデータが、ポリヌクレオチドの第2の領域における短い遺伝子バリエーションの存在をコールするステップに基づいて動的に生成される、実施形態8

10

20

30

40

50

7 から 90 のいずれか 1 つの方法。

【0309】

実施形態 92 . プライマーが、伸長プライマーに取り込まれたヌクレオチドの標識の存在または非存在を検出することなく、ポリヌクレオチドの第 1 のコピーの中のポリヌクレオチドの第 2 の領域を通して伸長される、実施形態 87 から 91 のいずれか 1 つの方法。

【0310】

実施形態 93 . ポリヌクレオチドの第 1 のコピーの中のポリヌクレオチドの第 2 の領域を通してプライマーを伸長するために使用されるヌクレオチドの少なくとも一部分が、非標識ヌクレオチドである、実施形態 87 から 92 のいずれか 1 つの方法。

【0311】

実施形態 94 . ポリヌクレオチドの第 1 のコピーの中のポリヌクレオチドの第 2 の領域を通してプライマーを伸長するために使用されるヌクレオチドが、非標識ヌクレオチドである、実施形態 87 から 92 のいずれか 1 つの方法。

【0312】

実施形態 95 . 少なくとも 2 つの異なるタイプのヌクレオチド塩基の混合物が、第 2 の領域のフロー順序の少なくとも 1 ステップで使用される、実施形態 87 から 94 のいずれか 1 つの方法。

【0313】

実施形態 96 . 3 つの異なるタイプのヌクレオチド塩基の混合物が、第 2 の領域のフロー順序の少なくとも 1 ステップで使用される、実施形態 87 から 95 のいずれか 1 つの方法。

【0314】

実施形態 97 . シークエンシングクラスター内のシークエンシングプライマーを同期化する方法であって、

(a) プライマーをシークエンシングクラスター内のポリヌクレオチドコピーとハイブリダイズするステップ；

(b) プライマーを、第 1 の領域のフローサイクルに従って標識ヌクレオチドを使用してポリヌクレオチドコピーの第 1 の領域を通して伸長するステップ；

(c) プライマーを、1 つまたは複数の再位相化フローを使用してポリヌクレオチドコピーの第 2 の領域を通して伸長するステップであって、少なくとも 2 つの異なるタイプのヌクレオチドの混合物が、1 つまたは複数の再位相化フローの少なくとも 1 つにおいて使用される、ステップ；および

(d) プライマーを、第 3 の領域のフローサイクルに従って標識ヌクレオチドを使用してポリヌクレオチドコピーの第 3 の領域を通して伸長するステップ

を含む方法。

【0315】

実施形態 98 . 3 つの異なるタイプのヌクレオチド塩基の混合物が、1 つまたは複数の再位相化フローのうちの少なくとも 1 つにおいて使用される、実施形態 97 の方法。

【0316】

実施形態 99 . 1 つまたは複数の再位相化フローが、4 つまたはそれより多くのフローステップを含む、実施形態 97 または 98 の方法。

【0317】

実施形態 100 . 1 つまたは複数の再位相フローが、任意の順序で：

(i) A、C および G ヌクレオチドを含むが T ヌクレオチドを含まない混合物を含む第 1 のフロー；

(i i) T、C および G ヌクレオチドを含むが A ヌクレオチドを含まない混合物を含む第 2 のフロー；

(i i i) T、A および G ヌクレオチドを含むが C ヌクレオチドを含まない混合物を含む第 3 のフロー；および

(i v) T、A および C ヌクレオチドを含むが G ヌクレオチドを含まない混合物を含む第

10

20

30

40

50

4 のフロー

を含む、実施形態 99 の方法。

【0318】

実施形態 101. 第 1 の領域の配列に関連するシーケンシングデータを、プライマーを第 1 の領域を通して伸長しながら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により生成するステップを含む、実施形態 97 から 100 のいずれか 1 つの方法。

【0319】

実施形態 102. 第 3 の領域の配列に関連するシーケンシングデータを、プライマーを第 3 の領域を通して伸長しながら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により生成するステップを含む、実施形態 97 から 101 のいずれか 1 つの方法。

10

【0320】

実施形態 103.

1 または複数台のプロセッサと、
非一過性記憶媒体であって、

1 つまたは複数のカップリングされたシーケンシングリードに関する情報を受信する、
および

実施形態 26 ~ 52 および 57 ~ 86 のいずれか 1 つの方法を遂行する

ための、1 または複数台のプロセッサにより実行可能な 1 つまたは複数のプログラムを
含む非一過性記憶媒体と
を含むシステム。

20

【0321】

実施形態 104. 1 つまたは複数のカップリングされたシーケンシングリードが、実施形態 1 ~ 25、53 ~ 56、および 87 ~ 96 のいずれか 1 つの方法に従って生成される、実施形態 103 のシステム。

【実施例】

【0322】

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。

30

（実施例 1）

【0323】

高速フォワード領域を含むフローシーケンシング法を使用して、および標準フローシーケンシング法（すなわち、高速フォワード領域を含まないフローシーケンシング法）をさらに使用して、262 塩基を有する核酸構築物をシーケンシングした。ポリヌクレオチドをアダプター配列にライゲーションレベーズに繋留し、これを増幅しシーケンシング表面と会合させた。シーケンシングプライマーをアダプター配列内のハイブリダイゼーション領域とハイブリダイズし、フローシーケンシング法の開始を可能にした。第 1 の方法では、単一タイプの蛍光標識された非終結ヌクレオチドの交互フローを使用してシーケンシングプライマーを伸長することにより 62 塩基をシーケンシングし、各ステップ後のヌクレオチド組込みを、蛍光検出器を使用して判定した。次の 177 塩基を、各フローが、存在する 4 つのヌクレオチドのうちの 3 つを有する、未標識の非終結ヌクレオチドの交互フロー（すなわち「高速フォワード」モード）に曝露して、プライマーを第 2 の領域を通して伸長させた。「暗」（すなわち、取り込まれたヌクレオチドの検出なし

40

50

）の第2の領域を通したプライマーの伸長後、別の23塩基は、単一タイプの蛍光標識された非終結ヌクレオチドのシーケンシングされた交互フローであり、各ステップ後のヌクレオチド組込みを、蛍光検出器を使用して判定した。結果を図19Aに示し、この図は、横アクセス上にフローステップ数、および縦アクセス上にシーケンシングシグナル（すなわち、正規化蛍光シグナル）の測度を示す。この方法は、高速フォワードレジメンに従って高品質シーケンシングデータをもたらす。

【0324】

同じ262塩基構築物を、介在高速フォワードレジメンを用いずに全面的に標準フローシーケンシング法でシーケンシングした。つまり、全262塩基は、単一タイプの蛍光標識された非終結ヌクレオチドのシーケンシングされた交互フローであり、各ステップ後のヌクレオチド組込みを、蛍光検出器を使用して判定した。結果を図19Bに示し、この図は、図を圧縮するために、対応する177塩基領域からのデータを含まない。

【0325】

シーケンシング構築物は、高速フォワードフローシーケンシング法を使用すると、標準フローシーケンシング法よりも迅速に前進する。ポリヌクレオチドの両端からのシーケンシングデータを関連付けてカップリングされたシーケンシングリードペアを生成し、解析することができる。

（実施例2）

【0326】

配列番号4内のバリエーション（配列番号1の参照配列と比較して塩基位置15にC→G単一ヌクレオチド多型バリエーションがある）の検出をこの実施例で説明する。プライマーを配列番号4の5'末端のハイブリダイゼーション配列とハイブリダイズすること、およびフローシーケンシング法を使用してプライマーを伸長することにより、配列番号4についてのカップリングされたシーケンシングリードペアを生成することができる。この実施例では、5サイクルを使用し、サイクル1を使用して、第1の領域を通してプライマーを伸長し、サイクル2およびサイクル3を使用して、第2の領域を通してプライマーを伸長し、サイクル4およびサイクル5を使用して、第3の領域を通してプライマーを伸長する。サイクル1、サイクル4およびサイクル5は、標識ヌクレオチドを使用してプライマーを伸長し、プライマーへのヌクレオチドの組込みを、サイクルステップを終えるたびに検出する。対照的に、サイクル2およびサイクル3中のプライマーへのヌクレオチドの組込みをスキップすることができる。各サイクルは、4つのステップを有し、サイクル1、4および5が、A-C-T-G標識ヌクレオチドの逐次的な独立した付加を含み、単一塩基タイプを各サイクルステップで付加し、標識ヌクレオチドの組込みを各ステップ後に検出する。サイクル2およびサイクル3を「高速フォワード」モードで実行し、これらのサイクルは、4つのサイクルステップを含み、ステップ1は、Aヌクレオチドを含まず（すなわち、C、TおよびGを含み）、ステップ2は、Cヌクレオチドを含まず（すなわち、A、TおよびGを含み）、ステップ3は、Tヌクレオチドを含まず（すなわち、A、CおよびGを含み）、ステップ4は、Gヌクレオチドを含まない（すなわち、A、CおよびTを含む）。ヌクレオチド組込みをサイクル2およびサイクル3の高速フォワードモード中に検出しない。サイクル2および3は、プライマー伸長中に同時に複数の異なるヌクレオチド塩基タイプを含むため、いずれかの所与のステップで単一の塩基タイプしか使用されなかった場合よりも速くプライマーを伸長する。配列番号1（参照配列）および配列番号4（SNP配列）についてのフローグラムを表6に示す。シーケンシングデータは、配列番号1の第3の領域（サイクル4およびサイクル5）が3'-CTGAC-5'（配列番号5）であること、および配列番号4の第3の領域（サイクル4およびサイクル5）が3'-CC TGC-5'（配列番号7）であることを示す。配列番号1と配列番号4の間のシーケンシングデータ間の差は、第2の領域内のバリエーションの存在を示す。

10

20

30

40

50

【表 6】

表 6

	サイクル	1				2				3				4				5			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
		A	C	T	G	C/T/G	A/T/G	ACG	ACT	C/T/G	A/T/G	ACG	ACT	A	C	T	G	A	C	T	G
配列番号 4	サイクルステップ	第1の領域				第2の領域(高速フォワード領域)				第3の領域											
	フロア塩基																				
	取り込まれた塩基の数	1	1	1	1	0	2	1	3	1	1	1	5	0	0	0	2	1	1	0	1
	取り込まれた塩基	A	C	T	G	-	AA	C	TTA	G	A	C	TATAC (配列番号 ⁶⁾)	-	-	-	GG	A	C	-	G
配列番号 1	サイクルステップ	1	1	1	1	0	2	1	3	4	3	6	2	0	0	0	1	1	1	1	1
	フロア塩基	A	C	T	G	-	AA	C	TTA	GGCT (配列番号 ²⁾)	ATA	CGGACG (配列番号 ³⁾)	TC	-	-	-	G	A	C	T	G
	取り込まれた塩基の数	1	1	1	1	0	2	1	3	4	3	6	2	0	0	0	1	1	1	1	1
	取り込まれた塩基	A	C	T	G	-	AA	C	TTA	GGCT (配列番号 ²⁾)	ATA	CGGACG (配列番号 ³⁾)	TC	-	-	-	G	A	C	T	G

配列番号 4: 3'-TGACTTGAATCTGATATGCCTGCAGCTGAC-5'

配列番号 1: 3'-TGACTTGAATCCGATATGCCTGCAGCTGAC-5'

10

20

30

40

50

【表 7】

表 7

	サイクル サイクルステップ フロー塩基	1				2				3				4				5			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
		A	C	T	G	C/T/G	A/T/G	A/C/G	A/C/T	C/T/G	A/T/G	A/C/G	A/C/T	A	C	T	G	A	C	T	G
配列 番号 8	取り込まれた 塩基の数	第1の領域				第2の領域(高速フォワード領域)				第3の領域				第3の領域				第3の領域			
		1	1	1	1	0	2	1	3	4	3	6	2	0	0	0	1	0	0	1	0
		A	C	T	G	-	AA	C	TAA	GGCT (配列番号 2)	ATA	CGGACG (配列番号 3)	TA	-	-	-	G	-	-	T	-
配列 番号 1	取り込まれた 塩基の数	1	1	1	1	0	2	1	3	4	3	6	2	0	0	0	1	1	1	1	1
		A	C	T	G	-	AA	C	TAA	GGCT (配列番号 2)	ATA	CGGACG (配列番号 3)	TC	-	-	-	G	A	C	T	G
		A	C	T	G	-	AA	C	TAA	GGCT (配列番号 2)	ATA	CGGACG (配列番号 3)	TC	-	-	-	G	A	C	T	G

配列番号 8: 3'-TGACTTGAATCCGATATGCCCTGCATCAGCTGAC-5'

配列番号 1: 3'-TGACTTGAATCCGATATGCCCTGCAGCTGAC-5'

10

20

30

40

50

【表 8】

表 8

	サイクル サイクルステップ フロー塩基	1				2				3				4				5			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
		A	C	T	G	C/TG	A/TG	ACG	AC/T	C/TG	A/TG	ACG	AC/T	A	C	T	G	A	C	T	G
配列 番号 9	取り込まれた 塩基の数	第1の領域				第2の領域(高速フォワード領域)				第3の領域											
		1	1	1	1	0	2	1	3	4	3	4	0	0	0	1	1	0	0	0	0
		Λ	C	T	G	-	ΛΛ	C	TTA	GGCT (配列番号 2)	ATA (配列番号 11)	CGAC	-	-	-	T	G	-	-	-	-
配列 番号 1	取り込まれた 塩基の数	1	1	1	1	0	2	1	3	4	3	6	2	0	0	0	1	1	1	1	1
		A	C	T	G	-	AA	C	TTA	GGCT (配列番号 2)	ATA (配列番号 3)	CGGACG	TC	-	-	-	G	A	C	T	G
		A	C	T	G	-	AA	C	TTA	GGCT (配列番号 2)	ATA (配列番号 3)	CGGACG	TC	-	-	-	G	A	C	T	G

配列番号 9: 3'-TGACTTGAATCCGATATGCTGAC-5'
配列番号 1: 3'-TGACTTGAATCCGATATGCTGAC-5'

10

20

30

40

50

【表 9】

	1				2				3				4				5			
	第1の領域				第2の領域(高速フォワード領域)				第3の領域				第4の領域				第5の領域			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
サイクル サイクルステップ	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
フロー塩基	A	C	T	G	C/T	A/T	A/C	T	C/T/G	A/T/G	A/C/G	A/C/T	A	C	T	G	A	C	T	G
配列 番号 12	取り込まれた塩基の数	1	1	1	1	2	1	3	4	5	9	1	0	0	0	1	0	0	0	0
	取り込まれた塩基	A	C	T	G	AA	C	TTA	GGCT (配列番号 2)	ATATG (配列番号 10)	CAGGCGAC (配列番号 14)	T	-	-	-	G	-	-	-	-
配列 番号 1	取り込まれた塩基の数	1	1	1	1	2	1	3	4	3	6	2	0	0	0	1	1	1	1	1
	取り込まれた塩基	A	C	T	G	AA	C	TTA	GGCT (配列番号 2)	ATA (配列番号 3)	CGGACG (配列番号 3)	TC	-	-	-	G	A	C	T	G

配列番号 12: 3'-TGACTTGAATCCGATATACGTCCGGCTGAC-5'

配列番号 1: 3'-TGACTTGAATCCGATATGCTGACGTGAC-5'

(実施例 3)

【0327】

配列番号 8 内のバリエーション（これは、配列番号 1 の参照配列と比較して塩基位置 2 3 の後に A T C 挿入物を含む）の検出をこの実施例で説明する。第 2 の領域を通しての高速フォワード部分を含むフローシーケンシング法を使用して、配列番号 1 および配列番号 8 についてのカップリングされたシーケンシングリードを生成することができる。この実施例では、5 サイクルを使用し、サイクル 1 を使用して、第 1 の領域を通してプライマーを伸長し、サイクル 2 およびサイクル 3 を使用して、第 2 の領域を通してプライマーを伸長し、サイクル 4 およびサイクル 5 を使用して、第 3 の領域を通してプライマーを伸長する

10

20

30

40

50

。サイクル 1、サイクル 4 およびサイクル 5 は、標識ヌクレオチドを使用してプライマーを伸長し、プライマーへのヌクレオチドの組込みを、サイクルステップを終えるたびに検出する。対照的に、サイクル 2 およびサイクル 3 中のプライマーへのヌクレオチドの組込みをスキップすることができる。各サイクルは、4つのステップを有し、サイクル 1、4 および 5 は、A-C-T-G 標識ヌクレオチドの逐次的な独立した付加を含み、単一塩基タイプを各サイクルステップで付加し、標識ヌクレオチドの組込みを各ステップ後に検出する。サイクル 2 およびサイクル 3 を「高速フォワード」モードで実行し、これらのサイクルは、4つのサイクルステップを含み、ステップ 1 は、Aヌクレオチドを含まず（すなわち、C、T および G を含み）、ステップ 2 は、Cヌクレオチドを含まず（すなわち、A、T および G を含み）、ステップ 3 は、Tヌクレオチドを含まず（すなわち、A、C および G を含み）、ステップ 4 は、Gヌクレオチドを含まない（すなわち、A、C および T を含む）。ヌクレオチド組込みをサイクル 2 およびサイクル 3 の高速フォワードモード中に検出しない。サイクル 2 および 3 は、プライマー伸長中に同時に複数の異なるヌクレオチド塩基タイプを含むため、いずれかの所与のステップで単一の塩基タイプしか使用されなかった場合よりも速くプライマーを伸長する。配列番号 1（参照配列）および配列番号 8 についてのフローグラムを表 7 に示す。シーケンシングデータは、配列番号 1 の第 3 の領域（サイクル 4 およびサイクル 5）が 3'-CTGAC-5'（配列番号 5）であること、および配列番号 8 の第 3 の領域（サイクル 4 およびサイクル 5）が 3'-AC-5'であることを示す。配列番号 1 と配列番号 8 の間のシーケンシングデータ間の差は、第 2 の領域内のバリエーションの存在を示す。

（実施例 4）

【0328】

配列番号 9 内のバリエーション（これは、配列番号 1 の参照配列と比較して塩基位置 17 の後に G C C T G C A（配列番号 13）塩基の欠失を含む）の検出をこの実施例で説明する。第 2 の領域を通しての高速フォワード部分を含むフローシーケンシング法を使用して、配列番号 1 および配列番号 9 についてのカップリングされたシーケンシングリードを生成することができる。この実施例では、5 サイクルを使用し、サイクル 1 を使用して、第 1 の領域を通してプライマーを伸長し、サイクル 2 およびサイクル 3 を使用して、第 2 の領域を通してプライマーを伸長し、サイクル 4 およびサイクル 5 を使用して、第 3 の領域を通してプライマーを伸長する。サイクル 1、サイクル 4 およびサイクル 5 は、標識ヌクレオチドを使用してプライマーを伸長し、プライマーへのヌクレオチドの組込みを、サイクルステップを終えるたびに検出する。対照的に、サイクル 2 およびサイクル 3 中のプライマーへのヌクレオチドの組込みをスキップすることができる。各サイクルは、4つのステップを有し、サイクル 1、4 および 5 は、A-C-T-G 標識ヌクレオチドの逐次的な独立した付加を含み、単一塩基タイプを各サイクルステップで付加し、標識ヌクレオチドの組込みを各ステップ後に検出する。サイクル 2 およびサイクル 3 を「高速フォワード」モードで実行し、これらのサイクルは、4つのサイクルステップを含み、ステップ 1 は、Aヌクレオチドを含まず（すなわち、C、T および G を含み）、ステップ 2 は、Cヌクレオチドを含まず（すなわち、A、T および G を含み）、ステップ 3 は、Tヌクレオチドを含まず（すなわち、A、C および G を含み）、ステップ 4 は、Gヌクレオチドを含まない（すなわち、A、C および T を含む）。ヌクレオチド組込みをサイクル 2 およびサイクル 3 の高速フォワードモード中に検出しない。サイクル 2 および 3 は、プライマー伸長中に同時に複数の異なるヌクレオチド塩基タイプを含むため、いずれかの所与のステップで単一の塩基タイプしか使用されなかった場合よりも速くプライマーを伸長する。配列番号 1（参照配列）および配列番号 9 についてのフローグラムを表 8 に示す。シーケンシングデータは、配列番号 1 の第 3 の領域（サイクル 4 およびサイクル 5）が 3'-CTGAC-5'（配列番号 5）であること、および配列番号 9 の第 3 の領域（サイクル 4 およびサイクル 5）が 3'-AC-5'であることを示す。配列番号 1 と配列番号 8 の間のシーケンシングデータ間の差は、第 2 の領域内のバリエーションの存在を示す。

（実施例 5）

【0329】

配列番号12内のバリエーション（これは、配列番号1の参照配列と比較して塩基位置17の後にG C C T G C A（配列番号13）塩基の逆位を含む）の検出をこの実施例で説明する。第2の領域を通しての高速フォワード部分を含むフローシーケンシング法を使用して、配列番号1および配列番号12についてのカップリングされたシーケンシングリードを生成することができる。この実施例では、5サイクルを使用し、サイクル1を使用して、第1の領域を通してプライマーを伸長し、サイクル2およびサイクル3を使用して、第2の領域を通してプライマーを伸長し、サイクル4およびサイクル5を使用して、第3の領域を通してプライマーを伸長する。サイクル1、サイクル4およびサイクル5は、標識ヌクレオチドを使用してプライマーを伸長し、プライマーへのヌクレオチドの組込みを、サイクルステップを終えるたびに検出する。対照的に、サイクル2およびサイクル3中のプライマーへのヌクレオチドの組込みをスキップすることができる。各サイクルは、4つのステップを有し、サイクル1、4および5は、A-C-T-G標識ヌクレオチドの逐次的な独立した付加を含み、単一塩基タイプを各サイクルステップで付加し、標識ヌクレオチドの組込みを各ステップ後に検出する。サイクル2およびサイクル3を「高速フォワード」モードで実行し、これらのサイクルは、4つのサイクルステップを含み、ステップ1は、Aヌクレオチドを含まず（すなわち、C、TおよびGを含み）、ステップ2は、Cヌクレオチドを含まず（すなわち、A、TおよびGを含み）、ステップ3は、Tヌクレオチドを含まず（すなわち、A、CおよびGを含み）、ステップ4は、Gヌクレオチドを含まない（すなわち、A、CおよびTを含む）。ヌクレオチド組込みをサイクル2およびサイクル3の高速フォワードモード中に検出しない。サイクル2および3は、プライマー伸長中に同時に複数の異なるヌクレオチド塩基タイプを含むため、いずれかの所与のステップで単一の塩基タイプしか使用されなかった場合よりも速くプライマーを伸長する。配列番号1（参照配列）および配列番号12についてのフローグラムを表9に示す。シーケンシングデータは、配列番号1の第3の領域（サイクル4およびサイクル5）が3'-CTGAC-5'（配列番号5）であること、および配列番号12の第3の領域（サイクル4およびサイクル5）が3'-G-5'であることを示す。配列番号1と配列番号12の間のシーケンシングデータ間の差は、第2の領域内のバリエーションの存在を示す。

（実施例6）

【0330】

合成によるシーケンシングは、一般に、ヌクレオチドの伸長プライマーへの不完全な組込みを有する。経時的に、シーケンシングクラスター内で、プライマーは非同期化されることがあり、その結果、シグナルが分解され、塩基組込みコールを行なう信頼度が低下することになる。シーケンシングクラスター内のプライマー非同期化を、10,000の同一の鑄型鎖を有するシーケンシングクラスターを仮定すること、および各フローが単一のヌクレオチドを有する、A-C-T-Gのフロー順序を仮定して、非終結ヌクレオチドを使用して鑄型鎖をシーケンシングすることにより模擬した。組込み失敗（すなわち、ヌクレオチドが組み込まれているはずであることを鑄型が示したときにヌクレオチドが伸長プライマー鎖に組み込まれなかった）の確率を0.5%に設定した。図20Aは、100番目のフローがG非終結ヌクレオチドを有する100フロー後に各リード塩基において伸長されたプライマー（鎖）の数を示す。シーケンシングクラスターは、次に予想される組み込まれるヌクレオチドがAであるようなGヌクレオチドが伸長プライマーに取り込まれた先導シーケンシングプライマーとハイブリダイズされた鑄型、次に予想される組み込まれるヌクレオチドがCであるようなGヌクレオチドが伸長プライマーに取り込まれた第1の遅れプライマーとハイブリダイズされた鑄型、およびヌクレオチドが100番目のフローからの伸長プライマーに組み込まれなかった第2の遅れプライマーとハイブリダイズされた鑄型を含む。第1の遅れプライマーおよび第2の遅れプライマーは、伸長プライマーへの予想ヌクレオチドの組込みがシーケンシングプロセス中のある地点で失敗した、プライマーを表す。再位相化フロー順序を使用する伸長プライマーの同期化を、同期化フロー順序を使用して模擬した。フロー101で、G、CおよびA非終結ヌクレオ

チドの混合物を使用してプライマーを伸長し（図20B）、先導プライマーと同期化されるまで第1および第2の遅れプライマーを伸長した。フロー101は、Tヌクレオチドを含まなかったため、さらに伸長されなかった。模擬同期化フロー順序は、G、CおよびT非終結ヌクレオチドの混合物を有するフロー102（図20C）、G、TおよびA非終結ヌクレオチドの混合物を有するフロー103（図20D）、およびT、AおよびC非終結ヌクレオチドの混合物を有するフロー104（図20E）に継続した。

【0331】

図21A～21Eおよび図22A～22Eに見られるように追加の配列を使用して模擬同期化フロー順序を試験した。同期化フロー順序および異なる鑄型配列を使用して、他の良好な模擬実験を行なった。

（実施例7）

【0332】

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23および表4は、この解析の例示的結果を示す。図23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100万リード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8）

【0333】

SNPを検出するための高速フォワードシーケンシングの感度を試験するために、シーケンシング法を*in silico*で模擬して、hg38参照ゲノム内のおおよそ1,140,000の合成核酸分子をシーケンシングし、各々の合成核酸分子は、2キロ塩基セグメントであり、参照ゲノム内にランダムに始点があった。各々の合成シーケンシングリードから502bpセグメントを生成し、この約502bpセグメントの中の各塩基に対してクエリが実行される3つ可能な単一塩基突然変異（すなわち、合計500×約1.14M×3の可能なバリエーション（すなわち、ABC→ADC、ここで $B \neq D$ ））すべてに対してSNP検出についてのクエリを実行した。各SNPバリエーションABC→ADCについて、（A=BかつD=C）または（A=DかつB=C）の場合、いずれのSNPもフローグラムにおいて新しいゼロシグナルも新しい非ゼロシグナルも生成しないことになるので、SNPを検出不能と考えた。参照塩基検出感度に対するバリエーション塩基の行列が、図24に示されている。

【0334】

次いで、各フローが、中央（第2の）領域に3つのヌクレオチドの混合物を含む、4ステップフローサイクルを使用して、合成核酸分子を*in silico*でシーケンシングした。各ステップが単一のヌクレオチド塩基タイプを含む、4ステップフローサイクルに従って、80のヌクレオチドフローを使用して、合成核酸分子の第1の領域をシーケンシングした。シーケンシングプライマーは、第1の領域において80フローで 54 ± 7 塩基（1フロー当たり約0.675塩基）にわたって伸長した。各ステップが3つのヌクレ

10

20

30

40

50

オチド塩基タイプを含み1つのヌクレオチド塩基タイプを含まない（すなわち、（i）A、C、T、および非G；（ii）G、A、C、および非T；（iii）T、G、A、および非C；および（iv）C、T、G、および非A）、4ステップフローサイクルに従って、200ヌクレオチドを使用して、合成核酸分子の第2の領域をシーケンシングした。シーケンシングプライマーは、第2の領域において200フローで 915 ± 89 塩基（1フロー当たり約4.575塩基）にわたって伸長した。各ステップが単一のヌクレオチド塩基タイプを含む、4ステップフローサイクルに従って、80のヌクレオチドフローを使用して、合成核酸分子の第3の領域をシーケンシングした。シーケンシングプライマーは、第3の領域において80フローで 54 ± 7 塩基（1フロー当たり約0.675塩基）にわたって伸長した。各々の合成バリエーション核酸分子についての第3の（下流の）領域のフローグラムを、対応する合成野生型核酸分子についての第3の領域のフローグラムと比較した。対応する合成野生型核酸分子と比較して、合成バリエーション核酸分子の第3の領域における新しい非ゼロフローグラムエントリおよび／または新しいゼロフローグラムエントリは、第2の領域に導入されたSNPの検出を示した。図25Aは、第1、第2および第3の領域におけるフローにわたっての平均塩基組込みを示す。参照塩基検出感度に対するバリエーション塩基の行列を図25Bに示す。図25Cは、合成リードにわたっての塩基カバレッジの分布を示す。

（実施例9）

【0335】

2つまたは3つの異なるヌクレオチド塩基の混合物を有する再位相化フローステップを使用する再位相化の効果、模擬シーケンシング方法論を使用して研究した。各々が長さ600bpのおおよそ10,000の合成シーケンシングリードを、ヒトゲノムからランダム開始部位選択により合成した。対照群では、模擬フローグラムを、105ラウンドのT-G-C-Aフローサイクル（合計420フロー）を使用する合成シーケンシングリードのin silicoシーケンシングにより生成した。遅れ位相の確率（すなわち、鋳型が、ヌクレオチドが組み込まれているはずであることを示すときに伸長プライマー鎖に組み込まれなかったヌクレオチドの、正しく取り込まれたヌクレオチドに対する分率）を0.2%に設定し、進み位相の確率（すなわち、追加のヌクレオチドが各フロー後に伸長プライマーに取り込まれた、シーケンシングリードの分率）を0.5%に設定した。対照群についての平均リード長は、 $322 \text{ bp} \pm 18 \text{ bp}$ であった。

【0336】

一連の試験群では、模擬フローグラムを、以下の条件のうちの1つを除いて、105ラウンドのT-G-C-Aフローサイクル（合計420フロー）を使用する合成シーケンシングリードのin silicoシーケンシングにより生成した：（1）24フローを終えるたびに、CとGの混合物を含有する再位相化フローを挿入した（図26A）；（2）48フローを終えるたびに、CとGの混合物を含有する再位相化フローを挿入した（図26B）；（3）96フローを終えるたびに、CとGの混合物を含有する再位相化フローを挿入した（図26C）；（4）192フローを終えるたびに、CとGの混合物を含有する再位相化フローを挿入した（図26D）；（5）48フローを終えるたびに、CとGとTの混合物を含有する再位相化フローを挿入し、その後、単一のAフローを（冗長フローを回避するために）挿入した後、対照プロトコルによるT-G-C-Aサイクルに戻った（図26E）；（6）96フローを終えるたびに、CとGとTの混合物を含有する再位相化フローを挿入し、その後、単一のAフローを（冗長フローを回避するために）挿入した後、対照プロトコルによるT-G-C-Aサイクルに戻った（図26F）；（7）96フローを終えるたびに、CとGとTの混合物を含有する再位相化フローを挿入し、その後、AとCとGの混合物を含有する再位相化フローを挿入した（図26G）；（8）192フローを終えるたびに、CとGとTの混合物を含有する再位相化フローを挿入し、その後、AとCとGの混合物を含有する再位相化フローを挿入した（図26H）；（9）96フローを終えるたびに、CとGとTの混合物を含有する再位相化フローを挿入し、その後、AとCとTの混合物を含有する再位相化フローを挿入し、その後、AとGとTの混合物を含

有する再位相化フローを挿入し、その後、AとCとGの混合物を含有する再位相化フローを挿入した（図26I）；または（10）192フローを終えるたびに、CとGとTの混合物を含有する再位相化フローを挿入し、その後、AとCとTの混合物を含有する再位相化フローを挿入し、その後、AとGとTの混合物を含有する再位相化フローを挿入し、その後、AとCとGの混合物を含有する再位相化フローを挿入した（図26J）。

【0337】

試験した再位相化フローのいずれについての使用も、最小限のシーケンシングデータ損失で、*in silico*シーケンシングの全ラウンド後に対照と比較して全位相誤差（すなわち、遅れ誤差も進み誤差も導入されていない名目上シーケンシングされた鎖に対する、遅れ位相誤差を有する鎖の分率と進み位相誤差を有する鎖の分率の合計）の実質的減少をもたらした。図26A～26Jは、対照プロトコルおよびそれぞれの再位相化フロープロトコル各々についての全位相誤差の和の分布を示す。CとGの混合物を含有する再位相化フローの使用は、再位相化フローごとに約1bpの平均プライマー伸長（すなわち、シーケンシングギャップ）しか生成せずに、24フローを終えるたびに、平均総累計位相誤差を $31.2 \pm 9.6\%$ に低減させ（ $51.5 \pm 1.3\%$ 対照と比較して）（図26A）、48フローを終えるたびに、平均総累計位相誤差を $36.9 \pm 9.7\%$ に低減させ（図26B）、96フローを終えるたびに、平均総累計位相誤差を $40.2 \pm 10.1\%$ に低減させ（図26C）、192フローを終えるたびに、平均総累計位相誤差を $42.8 \pm 10.4\%$ に低減させた（図26D）。CとGとTの混合物を含有する再位相化フローの使用は、再位相化フローごとに約5bpの平均プライマー伸長しか生成せずに、48フローを終えるたびに、平均総累計位相誤差を $28.5 \pm 10.6\%$ に低減させ（図26E）、96フローを終えるたびに、平均総累計位相誤差を $31.1 \pm 12.2\%$ に低減させた（図26F）。CとGとTの混合物を含有する第1の再位相化フロー、およびAとCとGの混合物を含有する第2の再位相化フローの使用は、再位相化二重フローごとに約9bpの平均プライマー伸長しか生成せずに、96フローを終えるたびに、平均総累計位相誤差を $25.3 \pm 10.6\%$ に低減させ（図26G）、192フローを終えるたびに、平均総累計位相誤差を $26.6 \pm 12.6\%$ に低減させた（図26H）。CとGとTの混合物を含有する第1の再位相化フロー、AとCとTの混合物を含有する第2の再位相化フロー、およびAとGとTの混合物を含有する第3の再位相化フロー、およびAとCとGの混合物を含有する第4の再位相化フローの使用は、再位相化四重フローごとに約18bpの平均プライマー伸長しか生成せずに、96フローを終えるたびに、平均総累計位相誤差を $20.6 \pm 9.4\%$ に低減させ（図26I）、および192フローを終えるたびに、平均総累計位相誤差を $20.9 \pm 11.2\%$ に低減させた（図26J）。

10

20

30

40

50

【図面】

【図 1】

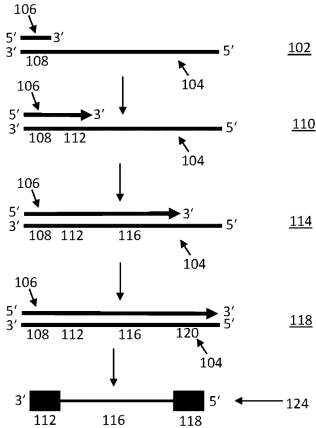
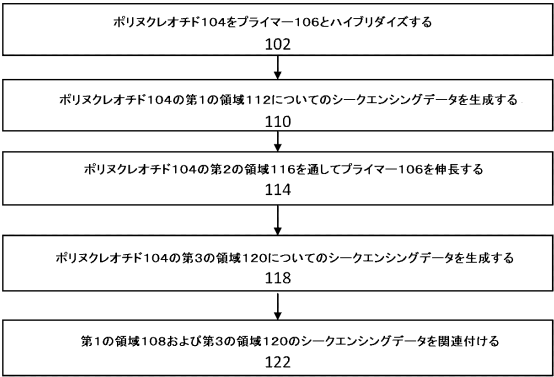


FIG. 1

【図 2】

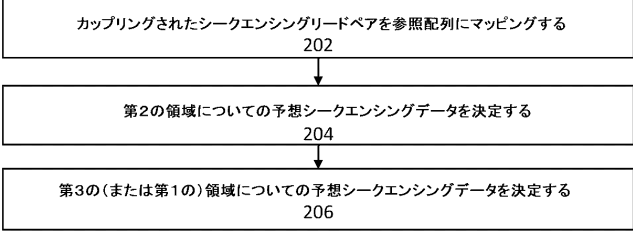


FIG. 2

【図 3】

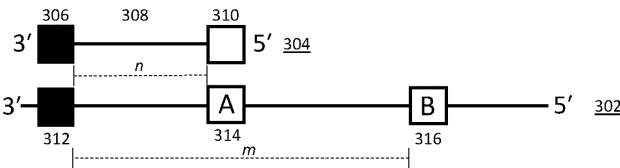


FIG. 3

【図 4】

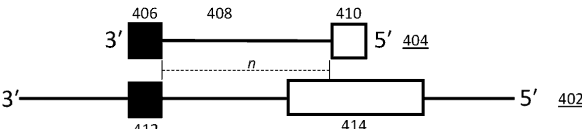


FIG. 4

【図 5】

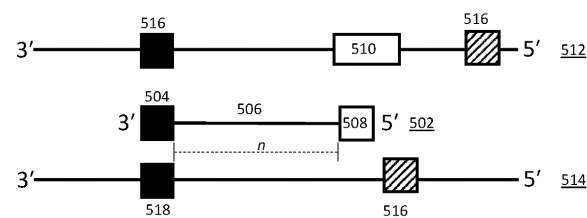


FIG. 5

【図 6】

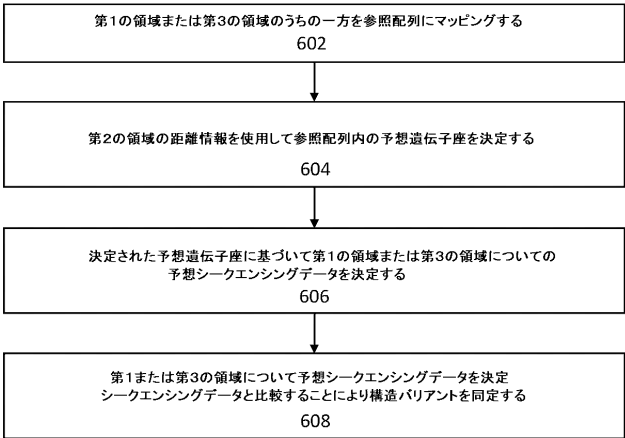


FIG. 6

【図 7】

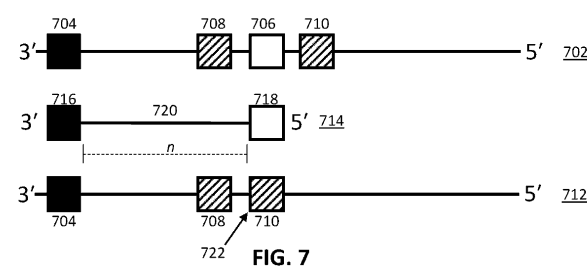


FIG. 7

【図 8】

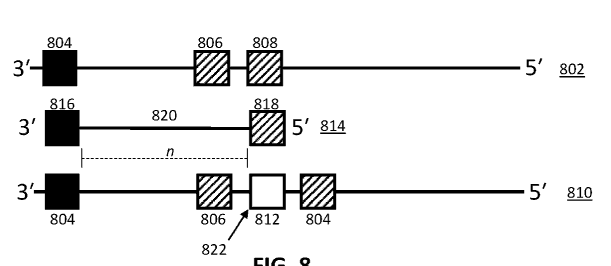


FIG. 8

10

20

30

40

50

【図 9】

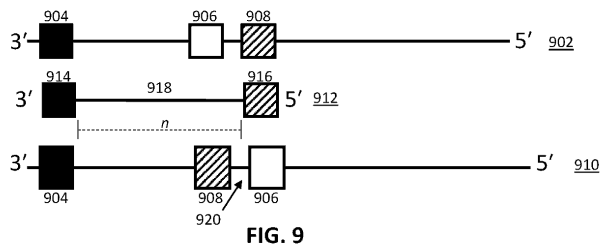


FIG. 9

【図 10】

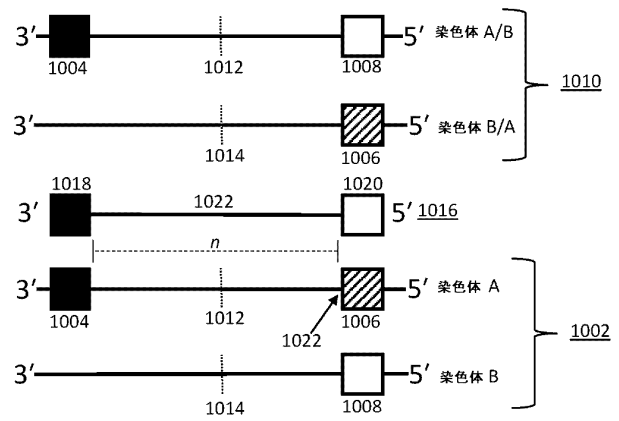


FIG. 10

【図 11】

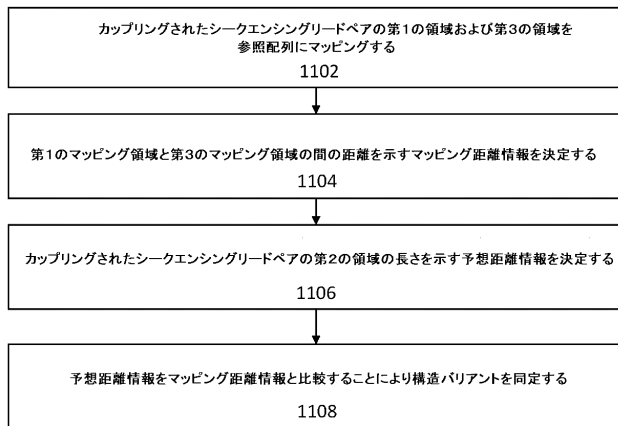


FIG. 11

【図 12】

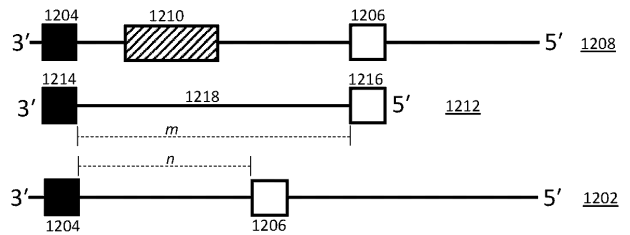


FIG. 12

10

20

30

40

50

【図 1 3】

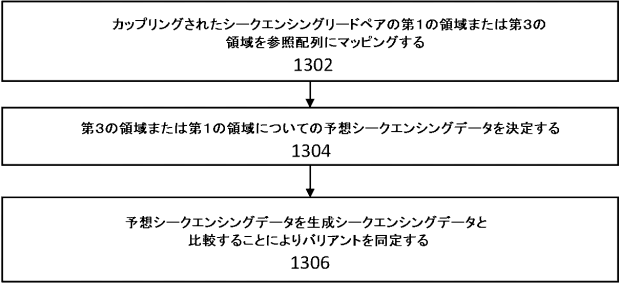


FIG. 13

【図 1 4 A】

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
1	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979
2	0.010	0.010	0.001	0.001	0.010	0.001	0.001	0.979	0.010	0.001	0.001	0.010	0.001	0.001	0.010	0.001	0.001	0.001	0.001	0.001
3	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.010	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

FIG. 14A

【図 1 4 B】

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
1	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979
2	0.010	0.010	0.001	0.001	0.010	0.001	0.001	0.979	0.010	0.001	0.001	0.010	0.001	0.001	0.010	0.001	0.001	0.001	0.001	0.001
3	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.010	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

FIG. 14B

【図 1 4 C】

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
1	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979	0.988	0.010	0.979
2	0.010	0.010	0.001	0.001	0.010	0.001	0.001	0.979	0.010	0.001	0.001	0.010	0.001	0.001	0.010	0.001	0.001	0.001	0.001	0.001
3	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.010	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

第1の 候補配列	T	A	・	・	T	・	・	GG	T	・	C	・	・	A	・	・	T	・	C	G
第2の 候補配列	T	A	・	・	T	・	・	GG	T	・	C	G	T	・	C	G	・	A	・	・

FIG. 14C

10

20

30

40

50

【図 15 - 1】

R1
TATGGTCGTCGA
R2
TGGTCGTCGAGC
R3
TATATGGTCGTC
H1
TATATGGTCATCGAGCTAT
H2
TATATGGTCGTCGAGCTAT

FIG. 15A

R1シーケエンシングデータセット

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					
1	0.9979	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979					
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					
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001					

FIG. 15B

【図 15 - 2】

R2シーケエンシングデータセット

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					
1	0.9979	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979					
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					
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001					

FIG. 15C

R3シーケエンシングデータセット

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					
1	0.9979	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979					
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					
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001					

FIG. 15D

【図 16】

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					
1	0.9979	0.9979	0.0010	0.0010	0.9979	0.0010	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979	0.0010	0.9979					
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					
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001					

FIG. 16

【図 17】

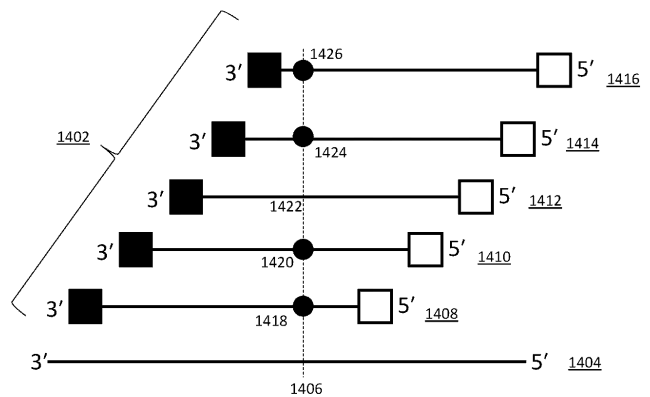


FIG. 17

【図 18】

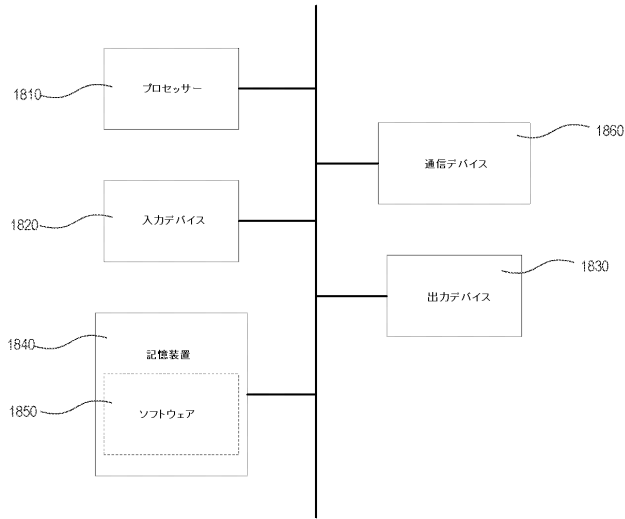


FIG. 18

【図 19 A】

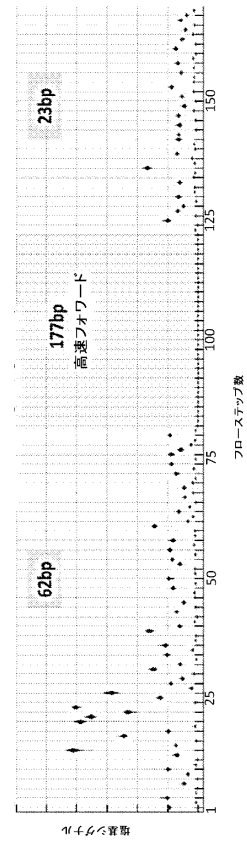


FIG. 19A

10

20

【図 19 B】

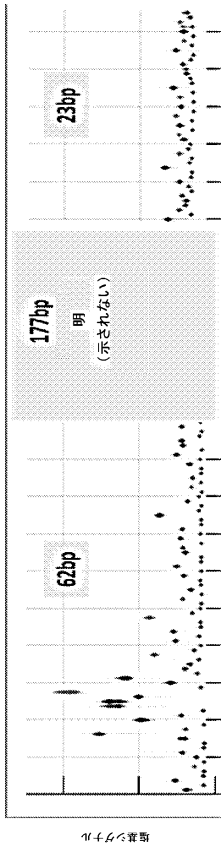


FIG. 19B

【図 20 - 1】

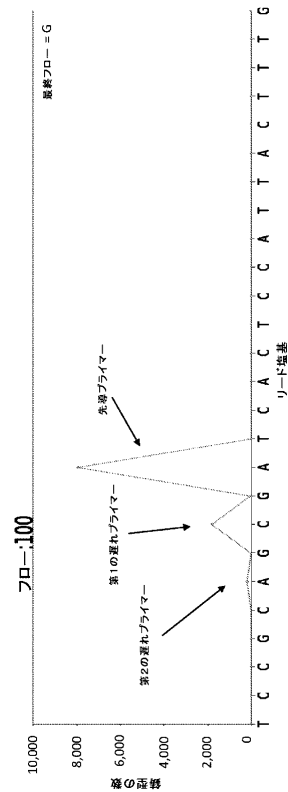
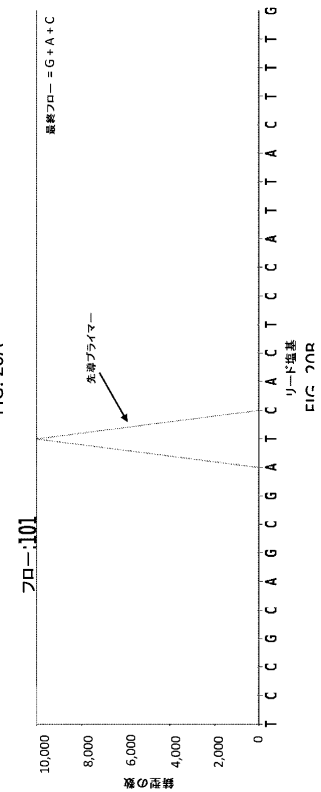


FIG. 20A

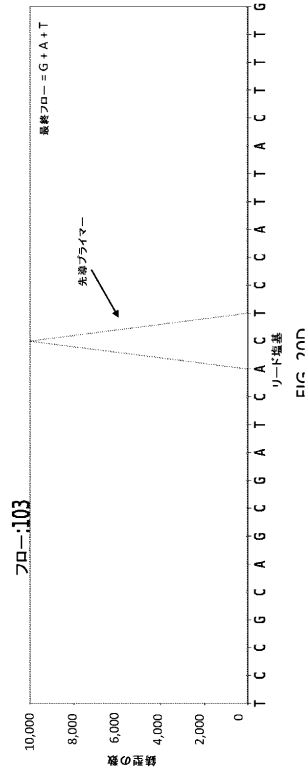
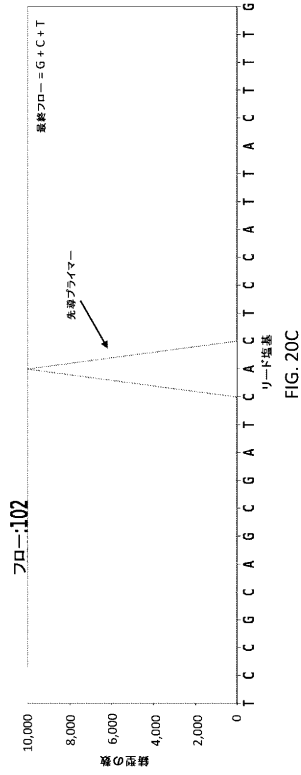
30

40

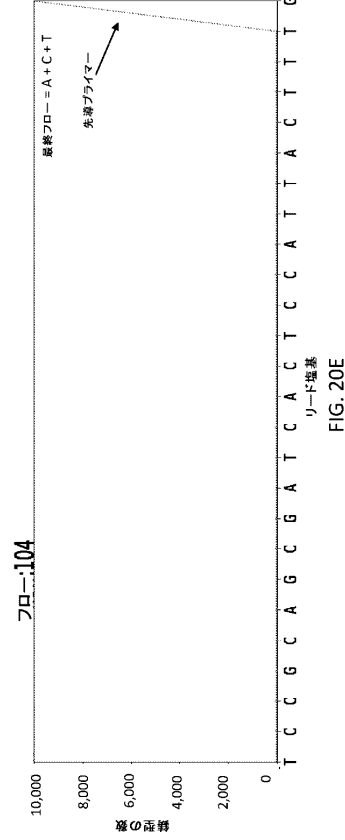


50

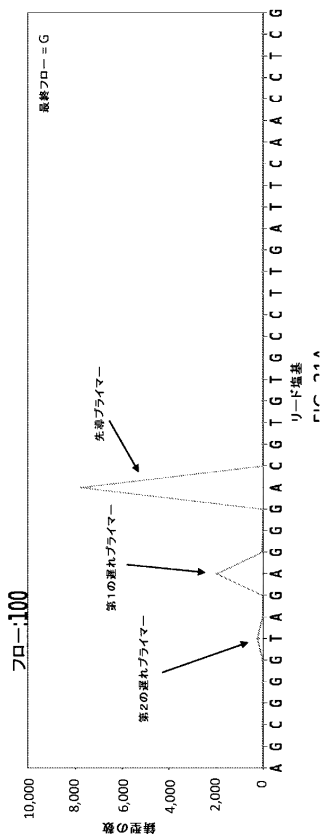
【図 20-2】



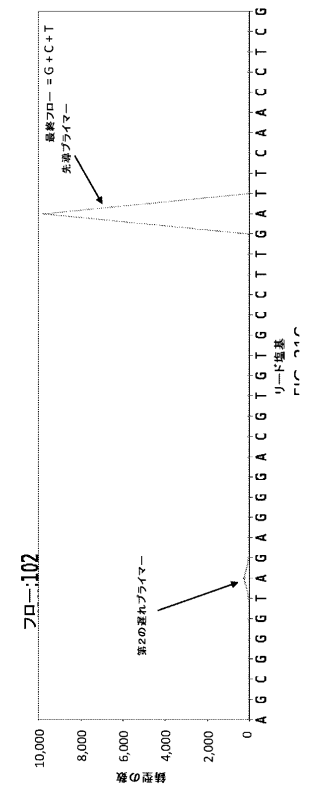
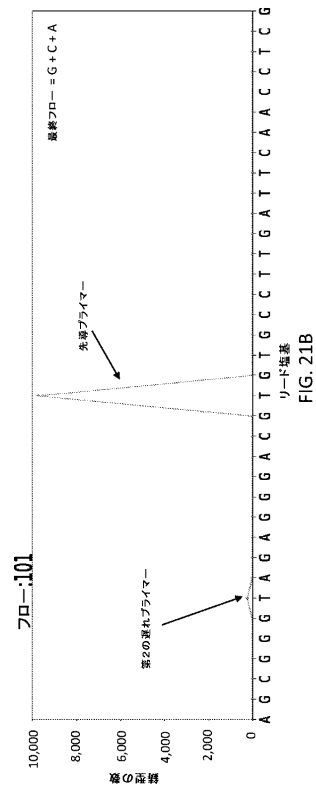
【図 20-3】



【図 21-1】



【図 21-2】



10

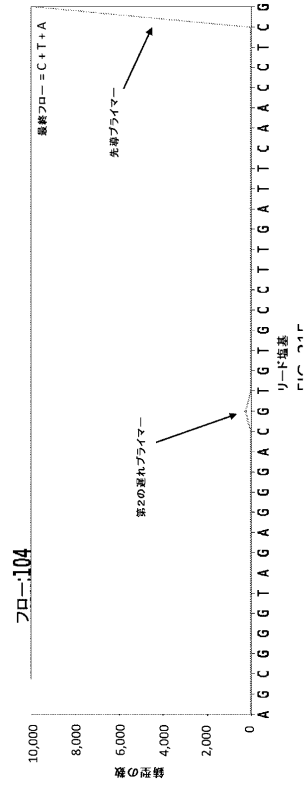
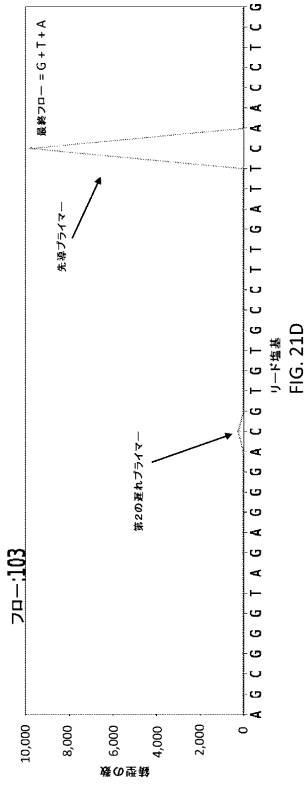
20

30

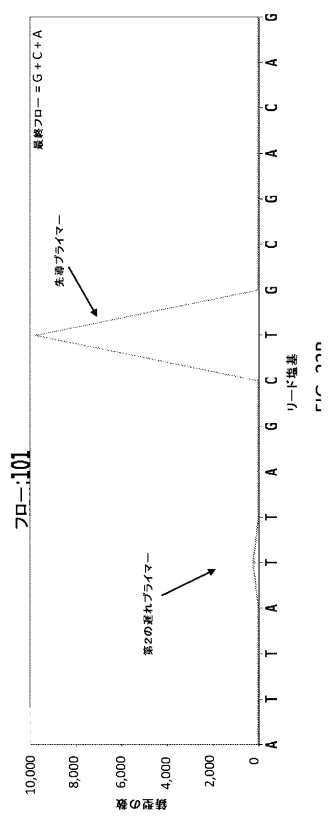
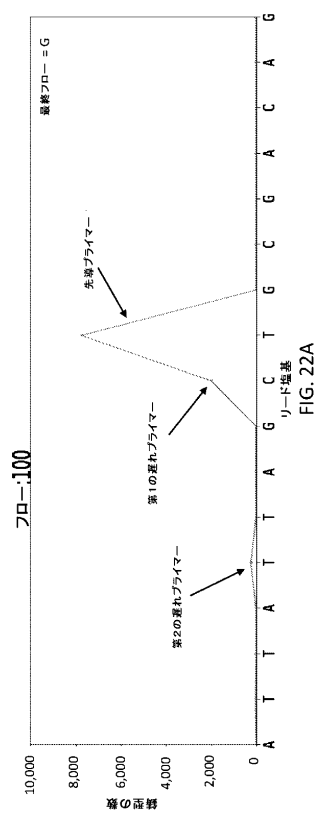
40

50

【図 21-3】



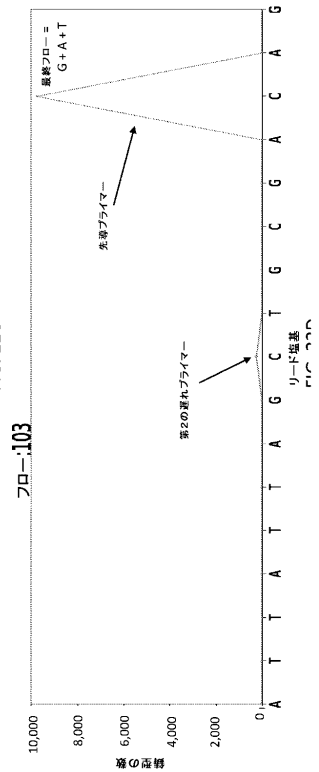
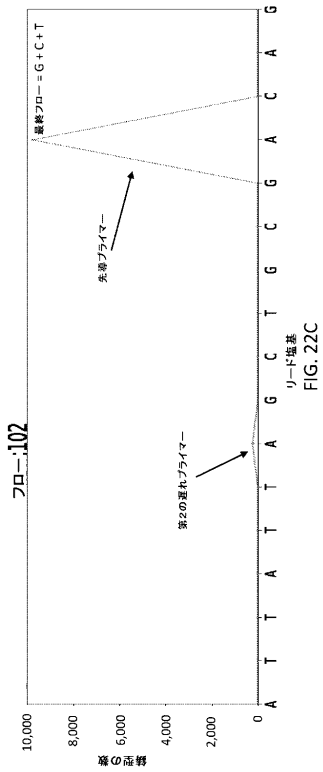
【図 22-1】



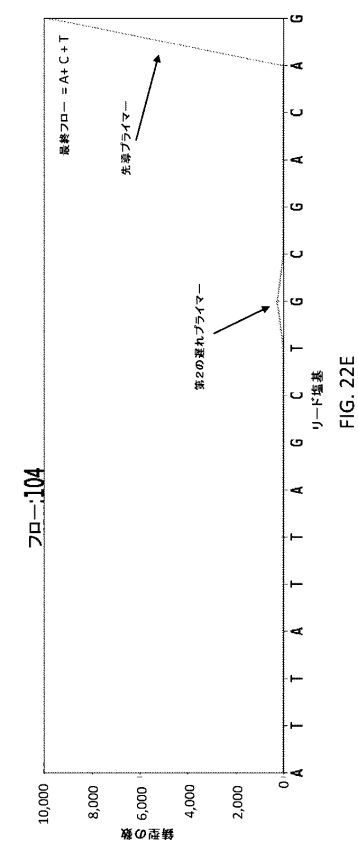
10

20

【図 22-2】



【図 22-3】

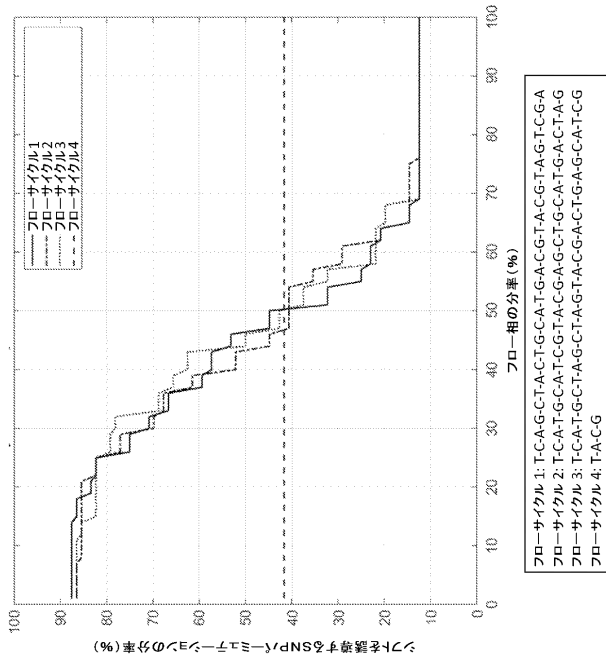


30

40

50

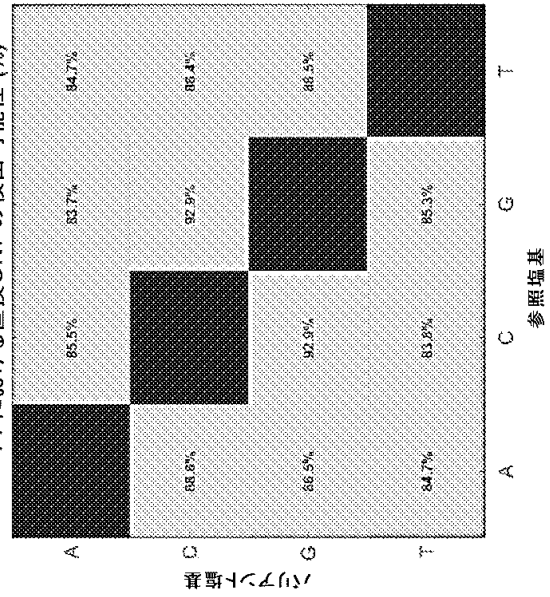
【図 2 3】



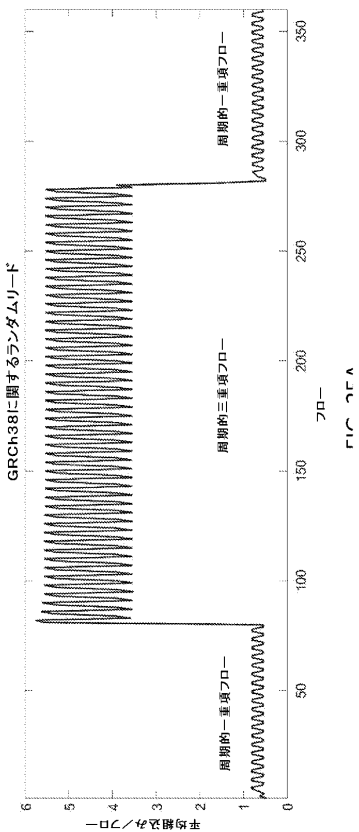
【図 2 4】

FIG. 23

FFにおける置換SNPの検出可能性 (%)



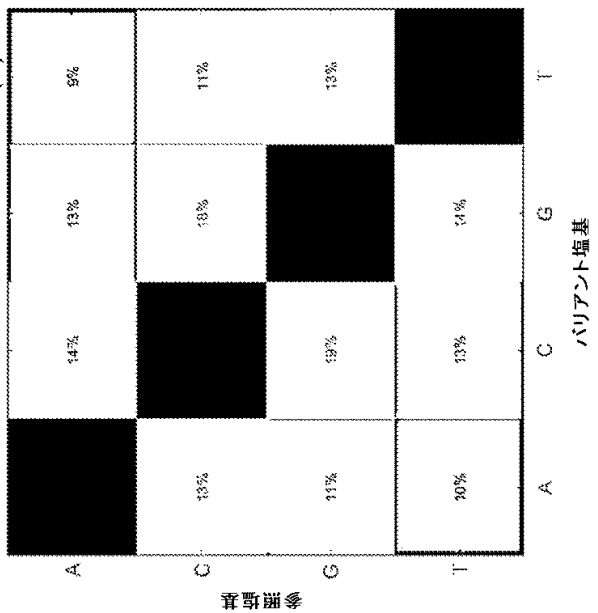
【図 2 5 - 1】



【図 2 5 - 2】

FIG. 25B

三重項FFにおけるサイクルシフトSNP検出 (%)



10

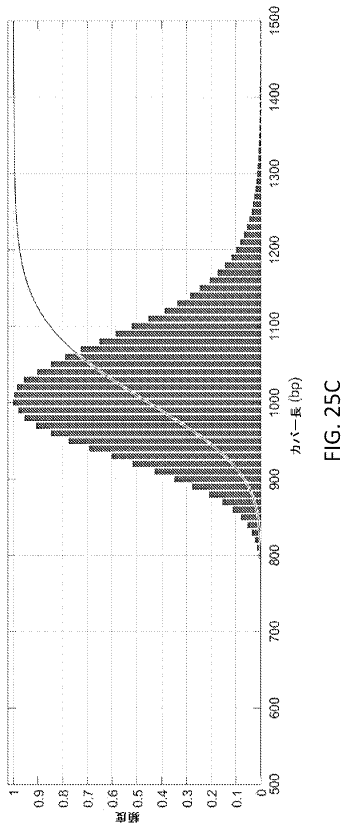
20

30

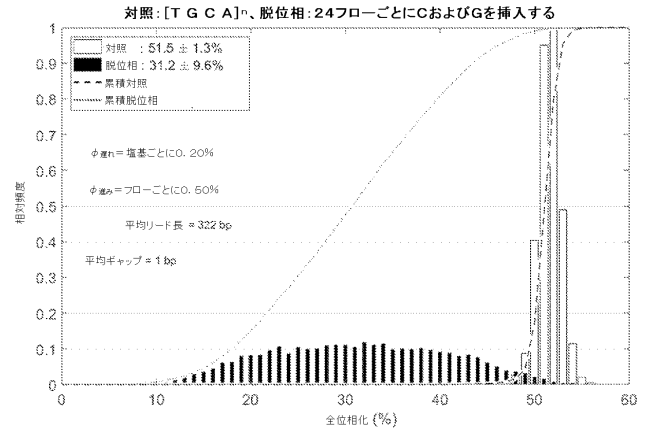
40

50

【図 25-3】



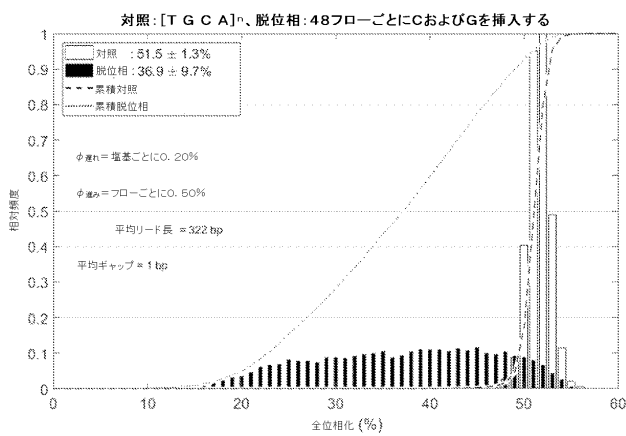
【図 26A】



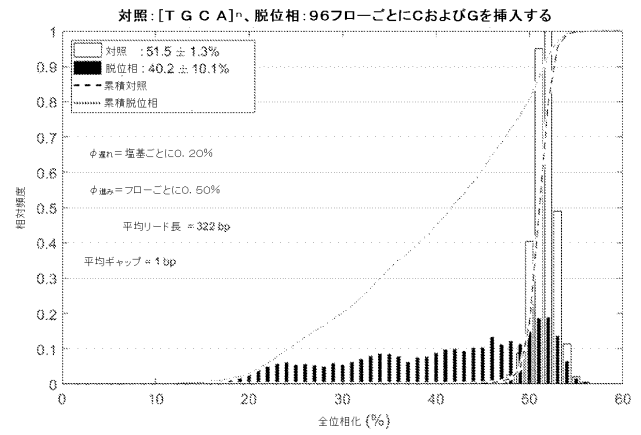
10

20

【図 26B】



【図 26C】



30

40

50

【図 26 D】

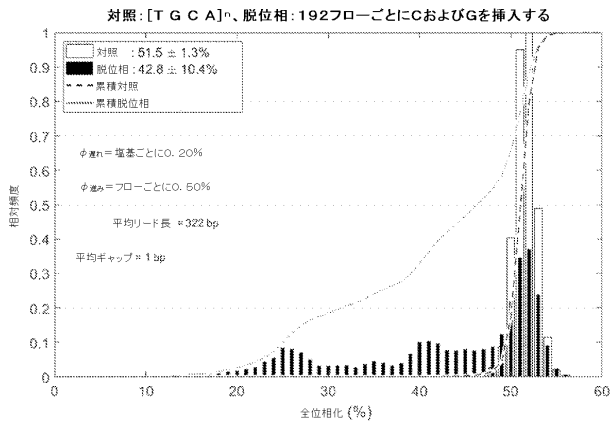


FIG. 26D

【図 26 E】

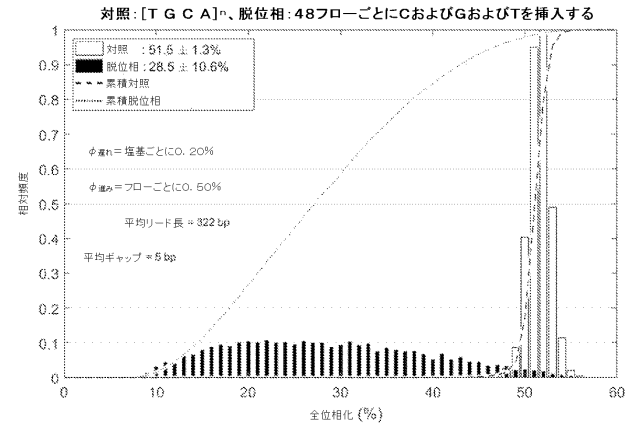


FIG. 26E

【図 26 F】

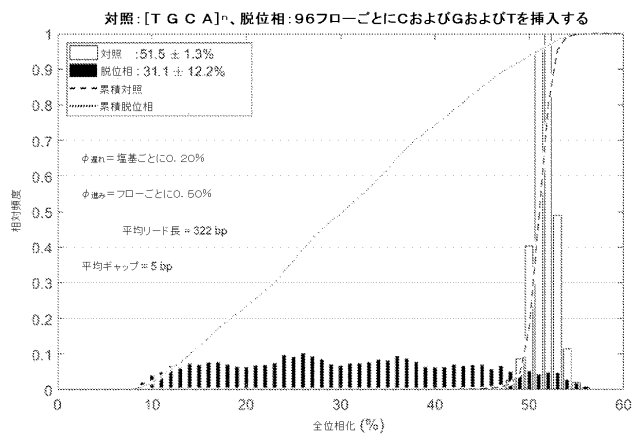


FIG. 26F

【図 26 G】

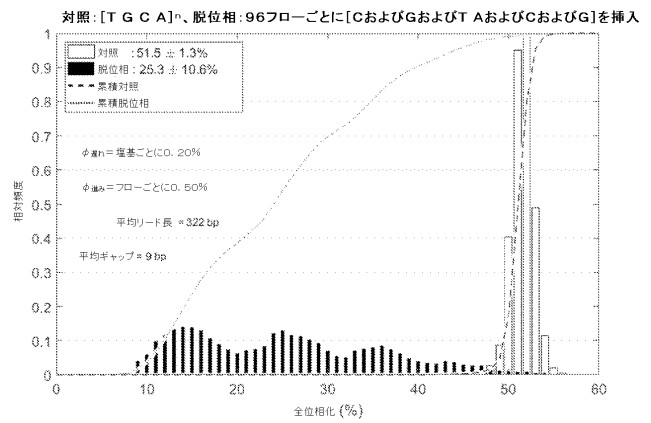


FIG. 26G

10

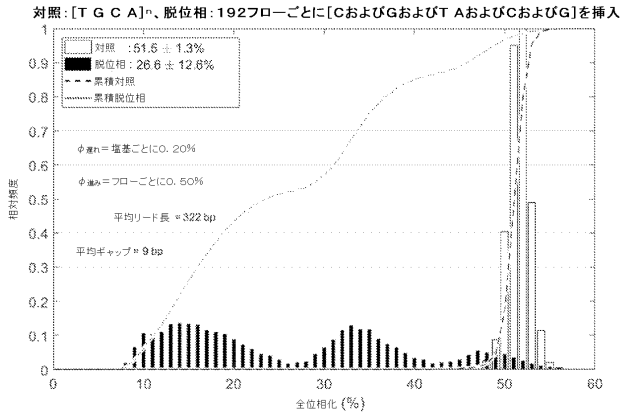
20

30

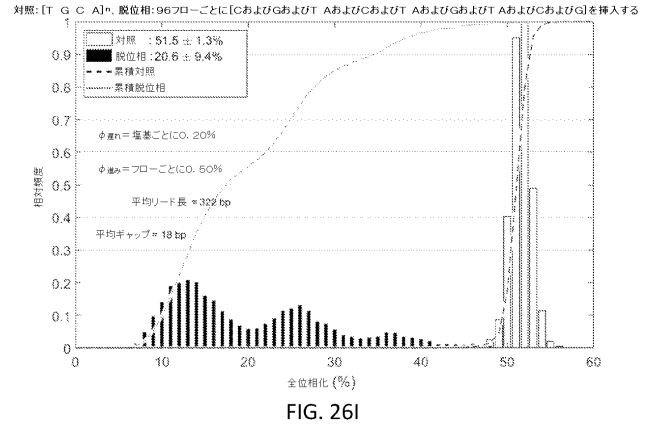
40

50

【図 26 H】

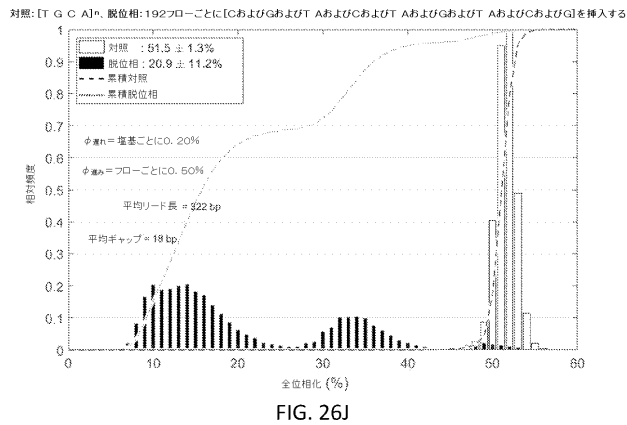


【図 26 I】



10

【図 26 J】



20

【配列表】

2022533801000001.app

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20/31163

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12Q 1/6869, 1/6874, 1/6806, 1/68 (2020.01)
CPC - C12Q 1/6869, 1/6874, 1/6806, 1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2018/0363047 A1 (ILLUMINA CAMBRIDGE LIMITED) 20 December 2018; paragraphs [0012]-[0013], [0039], [0047], [0117]-[0118]	4-5
Y		41-43
A		1-3, 40, 58-62
Y	US 2014/0315724 A1 (ZHOU, W et al.) 23 October 2014; paragraphs [0011], [0014], [0016], [0084], [0137], [0220]	41-43
A		1-3, 40, 58-62
A	(PENG, Q et al.) Targeted Single Primer Enrichment Sequencing with Single End Duplex-UMI. Scientific Reports. 18 March 2019, Vol. 9, No. 4810; abstract; DOI: 10.1038/s41598-019-41215-z	1-3, 40, 58-62

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

18 September 2020 (18.09.2020)

Date of mailing of the international search report

09 OCT 2020

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US20/31163

Box No. II	Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons: 1. ☐ Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: 2. ☐ Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically: 3. ☒ Claims Nos.: 6-39, 44-57, 63-65, 70-71 because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).	
Box No. III	Observations where unity of invention is lacking (Continuation of item 3 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows: This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid. Group I, Claims 1-5, 40-43, and 58-62 are directed toward methods of generating sequence reads comprising extension of a primer through a first, second, and third region. Group II, Claims 66-69 are directed toward a method of synchronizing sequencing primers within a sequencing cluster. ***-Continued on supplemental page-*** 1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims. 2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees. 3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-5, 40-43, and 58-62 Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee. ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation. ☐ No protest accompanied the payment of additional search fees.	

Form PCT/ISA/210 (continuation of first sheet (2)) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US20/31163

Continued from Box No. III: Observations where unity of invention is lacking

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include wherein extension of the primer through the second region of the polynucleotide within the first copy of the polynucleotide proceeds faster than the extension of the primer through the first region, not present in Group II; the special technical features of Group II include synchronizing sequencing primers, not present in Group I.

Groups I and II share the technical features including: hybridizing primers to polynucleotide copies; (b) extending the primers through a first region of the polynucleotide copies using labeled nucleotides; (c) extending the primers through a second region of the polynucleotide copies, wherein a mixture of at least two different types of nucleotide bases are used; and (d) extending the primers through a third region of the polynucleotide copies using labeled nucleotides.

However, these shared technical features are previously disclosed by US 2018/0363047 A1 to Illumina Cambridge Limited (hereinafter "Illumina").

Illumina discloses hybridizing primers to polynucleotide copies (hybridizing primers to polynucleotide copies; paragraph [0176]); (b) extending the primers through a first region of the polynucleotide copies (extending the primers through a first region of the polynucleotide copies paragraphs [0012], [0177]) using labeled nucleotides (paragraphs [0190], [0192]); (c) extending the primers through a second region of the polynucleotide copies (extending the primers through a second region of the polynucleotide copies; paragraph [0013]), wherein a mixture of at least two different types of nucleotide bases are used (wherein a mixture of at least two different types of nucleotide bases are used; paragraphs [0190], [0192]).

Illumina does not disclose extending the primers through a third region of the polynucleotide copies using labeled nucleotides. However, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to have to have modified the disclosure of Illumina to have repeated the steps required to extend a primer hybridized to a second region of a polynucleotide copy on a third region using an appropriate primer, in order to obtain a sequence read for a third region of the same polynucleotide copy.

Since none of the special technical features of the Groups I and II inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Illumina reference, unity of invention is lacking.

フロントページの続き

(32)優先日 令和2年2月7日(2020.2.7)

(33)優先権主張国・地域又は機関
米国(US)

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

(74)代理人 230113332

弁護士 山本 健策

(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 7979, 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 7979, 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 ブリンザ, ドミトル

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 7979, 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 トレパグニア, エリアン

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 7979, 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 バラド, オマー

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 7979, 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 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 7979, 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 7979, 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

F ターム (参考) 4B063 QA13 QQ42 QQ52 QR32 QR35