

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7546486号
(P7546486)

(45)発行日 令和6年9月6日(2024.9.6)

(24)登録日 令和6年8月29日(2024.8.29)

(51)国際特許分類		F I		
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z
C 1 2 N	15/09 (2006.01)	C 1 2 N	15/09	
C 1 2 Q	1/6844(2018.01)	C 1 2 Q	1/6844	Z
C 1 2 Q	1/6886(2018.01)	C 1 2 Q	1/6886	Z
請求項の数 11 (全64頁)				
(21)出願番号	特願2020-555454(P2020-555454)	(73)特許権者	515059083	
(86)(22)出願日	平成31年4月12日(2019.4.12)		ガーダント ヘルス, インコーポレイテッド	
(65)公表番号	特表2021-520810(P2021-520810 A)		アメリカ合衆国 カリフォルニア 9 4 3 0 4, パロ アルト, ハノーバー ストリート 3 1 0 0	
(43)公表日	令和3年8月26日(2021.8.26)	(74)代理人	100078282	
(86)国際出願番号	PCT/US2019/027337		弁理士 山本 秀策	
(87)国際公開番号	WO2019/200328	(74)代理人	100113413	
(87)国際公開日	令和1年10月17日(2019.10.17)		弁理士 森下 夏樹	
審査請求日	令和4年4月11日(2022.4.11)	(74)代理人	100181674	
(31)優先権主張番号	62/657,200		弁理士 飯田 貴敏	
(32)優先日	平成30年4月13日(2018.4.13)	(74)代理人	100181641	
(33)優先権主張国・地域又は機関	米国(US)		弁理士 石川 大輔	
前置審査		(74)代理人	230113332	
			最終頁に続く	

(54)【発明の名称】 融合事象によって引き起こされるアライメントエラーを検出および抑制する方法

(57)【特許請求の範囲】

【請求項1】

少なくとも部分的にコンピュータを使用して、対象の生物学的試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、

(a) 前記コンピュータによって、対象からの生物学的試料中の無細胞核酸分子から得られた前記遺伝子配列リードを含む、配列情報を受信するステップと、

(b) 前記遺伝子配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、

(c) 前記アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、

(d) 前記遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、前記遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、前記領域が、前記遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含み、前記遺伝子バリエーションが、前記生物学的試料中の前記ブレイクポイントの突然変異対立遺伝子画分よりも低いか、またはそれと同等である突然変異対立遺伝子画分を含む、ステップと、

(e) 前記1つまたは複数の遺伝子融合リードの前記サブセットにおける前記1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、

(f) 前記参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップと

を含む、方法。

【請求項 2】

前記遺伝子融合リードの前記セットが、1つまたは複数のプロセスされた偽遺伝子 (PPG) に対応し、例えば、

i) 前記 1つまたは複数の PPG が、1つまたは複数の試料特異的 PPG を含み、例えば、前記 1つまたは複数の試料特異的 PPG により、対象の集団において、前記対象が識別されるか、

ii) 前記 1つまたは複数の PPG が、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、および NRAS からなる群に由来するか、

iii) 前記 1つまたは複数の PPG が、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、および NRAS からなる群に由来する 2つまたはそれを上回る PPG を含むか、または

iv) 前記 1つまたは複数の PPG が、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、および NRAS からなる群に由来する 3つまたはそれを上回る PPG を含む、

請求項 1 に記載の方法。

【請求項 3】

前記遺伝子バリエーションまたは真の遺伝子バリエーションが、単一ヌクレオチドバリエーション (SNV) または挿入もしくは欠失 (インデル) を含み、例えば、

i) 前記遺伝子バリエーションが、SNV を含むか、

ii) 前記 SNV が、イントロン-エクソン境界部に位置するか、

iii) 前記 SNV が、遺伝子コーディング配列 (CDS) 内に位置するか、または

iv) 前記遺伝子バリエーションが、インデルを含む、

請求項 1 または 2 に記載の方法。

【請求項 4】

前記領域が、前記遺伝子内融合ブレイクポイントに隣接する約 2、4、6、8、10、15、または 20 個のヌクレオチドを含む、請求項 1 ~ 3 のいずれか一項に記載の方法。

【請求項 5】

前記 1つまたは複数の検出されたアライメントエラーの一部分が、事前に定義された臨床的に対処可能なバリエーションのセットに属さない遺伝子バリエーションを含む前記遺伝子融合リードに基づいて、フィルタリングされ、例えば、前記事前に定義されたセットが、COSMIC、The Cancer Genome Atlas (TCGA)、または Exome Aggregation Consortium (ExAC) において見出されるバリエーションを含む、請求項 1 ~ 4 のいずれか一項に記載の方法。

【請求項 6】

i) 前記試料が、血液、血漿、血清、尿、唾液、粘膜排出物、喀痰、糞便、および涙液からなる群から選択される体液試料であり、および／または

ii) 前記対象が、疾患または障害を有し、例えば、前記疾患が、がんである、

請求項 1 ~ 5 のいずれか一項に記載の方法。

【請求項 7】

前記対象の前記生物学的試料から、無細胞核酸分子を単離するステップを含み、必要に応じて、前記無細胞核酸分子が、DNA、RNA、またはこれらの組合せを含み、例えば、前記無細胞核酸分子が、二本鎖 DNA である、請求項 1 ~ 6 のいずれか一項に記載の方法。

【請求項 8】

シーケンシングの前に、分子バーコードを含む 1つまたは複数のアダプターを、前記無

10

20

30

40

50

細胞核酸分子に結合させて、タグ付けされた親ポリヌクレオチドを生成するステップをさらに含み、必要に応じて、

i) 前記アダプターが、前記無細胞核酸分子の両端に結合されるか、

ii) 前記無細胞核酸分子が、固有にバーコーディングされるか、または

iii) 前記無細胞核酸分子が、非固有にバーコーディングされ、例えば、それぞれのバーコードが、選択された領域からシーケンシングされる多様な分子と組み合わせて、固有な分子の識別を可能にする、固定またはセミランダムなオリゴヌクレオチド配列を含む、請求項 1 ～ 7 のいずれか一項に記載の方法。

【請求項 9】

前記タグ付けされた親ポリヌクレオチドを増幅させて、子孫ポリヌクレオチドを生成するステップをさらに含み、必要に応じて、目的の標的配列に関して、前記子孫ポリヌクレオチドを選択的に濃縮させ、それによって、濃縮された子孫ポリヌクレオチドを生成するステップをさらに含み、さらに必要に応じて、前記濃縮された子孫ポリヌクレオチドを増幅させるステップをさらに含む、請求項 8 に記載の方法。

【請求項 10】

前記子孫ポリヌクレオチドまたは濃縮された子孫ポリヌクレオチドに、試料インデックス配列がタグ付けされる、請求項 9 に記載の方法。

【請求項 11】

i) 前記配列情報が、核酸シーケンサーから得られ、

ii) 前記遺伝子融合リードのセットが、シーケンシングされたペアエンドリードをアライメントおよび接続することによって識別され、および／または

iii) 前記遺伝子融合リードのセットが、イントロン-エクソン境界部にまたがるカバレッジにおける不連続性に基づいて識別される、

請求項 1 ～ 10 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】

相互参照

本国際特許出願は、2018年4月13日に出願された米国仮特許出願第62／657,200号に基づく優先権を主張し、これは、参照によりその全体が本明細書に組み込まれる。

【背景技術】

【0002】

背景

ゲノム再編成事象によって引き起こされる重複したゲノム領域は、重複特異的バリエーションが誤って標的に割り当てられる可能性があるため、臨床的シーケンシング適用における正確なバリエーションコーリングに対する課題を提示し得る。プロセスされた偽遺伝子 (processed pseudogenes) (PPG) は、LINE (長鎖散在エレメント (Long Interspersed Element)) に媒介される逆転写およびプロセッシングされた mRNA のゲノム組み込みにより生じ得る重複したコーディング配列の原因であり、もともとの遺伝子の部分的または完全なコピーをもたらし、イントロン配列が欠如している。参照ゲノムにおいて見出される偽遺伝子により生じる偽陽性バリエーション、たとえば、PIK3CA および PTEN のものは、十分に研究されているが、しかしながら、珍しいもの、およびさらには個体特異的ながん関連の PPG の発見は、試料ごとでの PPG に関連する臨床アーチファクトのより体系的な調査および介在の必要性を示す。

【発明の概要】

【課題を解決するための手段】

【0003】

要旨

ある特定の態様では、本開示は、遺伝子配列リードにおけるアライメントエラーを検出

10

20

30

40

50

するための方法であって、対象の試料からの無細胞デオキシリボ核酸（DNA）分子をシーケンシングするステップであって、無細胞DNA分子のそれぞれが、複数の配列リードを生成する、ステップと、シーケンシングにより導出された配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップとを含む、方法を提供する。

10

【0004】

ある特定の態様では、本開示は、対象の試料からの無細胞DNA分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、対象の試料からの無細胞DNA分子をシーケンシングするステップであって、無細胞DNA分子のそれぞれが、複数の配列リードを生成する、ステップと、シーケンシングにより導出された配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、1つまたは複数の遺伝子融合リードのサブセットにおいて、1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を提供する。

20

【0005】

ある特定の態様では、本開示は、対象の試料からの無細胞DNA分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、対象の試料からの無細胞DNA分子をシーケンシングするステップであって、無細胞DNA分子のそれぞれが、複数の配列リードを生成する、ステップと、シーケンシングにより導出された配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、遺伝子融合リードのうちの1つまたは複数のサブセットが、SMAD4および／またはRAF1に対応する遺伝子配列を含む、ステップと、遺伝子融合リードのうちの1つまたは複数のサブセットにおいて、1つまたは複数の検出されたアライメントエラーの少なくとも一部分を、フィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を提供する。

30

40

【0006】

ある特定の態様では、本開示は、遺伝子配列リードにおけるアライメントエラーを検出するための方法であって、対象の試料からの無細胞DNA分子をシーケンシングするステップであって、無細胞DNA分子のそれぞれが、複数の配列リードを生成する、ステップと、シーケンシングにより導出された配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを決定するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子

50

融合リードのうちの1つまたは複数のサブセットを決定するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、所定の基準を満たす、領域内のそれぞれの遺伝子バリエーションを、アライメントエラーとして識別するステップとを含む、方法を提供する。

【0007】

ある特定の態様では、本開示は、対象の試料からの無細胞DNA分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、対象の試料からの無細胞DNA分子をシーケンシングするステップであって、無細胞DNA分子のそれぞれが、複数の配列リードを生成する、ステップと、シーケンシングにより導出された配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを決定するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを決定するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、所定の基準を満たす、領域内のそれぞれの遺伝子バリエーションを、アライメントエラーとして識別するステップと、1つまたは複数の遺伝子融合リードのサブセットにおいて、1つまたは複数のアライメントエラーをフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を提供する。

【0008】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、遺伝子配列リードにおけるアライメントエラーを検出するための方法であって、コンピュータによって、対象からの生物学的試料中の無細胞核酸分子から得られた遺伝子配列リードを含む、配列情報を受信するステップと、遺伝子配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップとを含む、方法を提供する。

【0009】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、対象の生物学的試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、コンピュータによって、無細胞核酸分子から得られた配列リードを含む、配列情報を受信するステップと、配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、1つまたは複数の遺伝子融合リードのサブセットにおいて、1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を提供する。

【0010】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、対象の試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、コンピュータによって、無細胞核酸分子

から得られたシーケンシングリードを含む、配列情報を受信するステップと、配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、遺伝子融合リードのうちの1つまたは複数のサブセットが、SMAD4、TYRO3、および／またはRAF1に対応する遺伝子配列を含む、ステップと、遺伝子融合リードのうちの1つまたは複数のサブセットにおいて、1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を提供する。

10

【0011】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、遺伝子配列リードにおけるアライメントエラーを検出するための方法であって、コンピュータによって、対象からの生物学的試料中の無細胞核酸分子から得られた遺伝子配列リードを含む、配列情報を受信するステップと、遺伝子配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを決定するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを決定するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、所定の基準を満たす、領域内のそれぞれの遺伝子バリエーションを、アライメントエラーとして識別するステップとを含む、方法を提供する。

20

【0012】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、対象の試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、コンピュータによって、無細胞核酸分子から得られたシーケンシングリードを含む、配列情報を受信するステップと、配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、遺伝子融合リードのうちの1つまたは複数のサブセットが、SMAD4、TYRO3、および／またはRAF1に対応する遺伝子配列を含む、ステップと、遺伝子融合リードのうちの1つまたは複数のサブセットにおいて、1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を提供する。

30

40

【0013】

ある特定の実施形態では、遺伝子融合リードのセットは、1つまたは複数のプロセスされた偽遺伝子（PPG）に対応する。ある特定の実施形態では、1つまたは複数のPPGは、1つまたは複数の試料特異的PPGを含む。ある特定の実施形態では、1つまたは複数のPPGは、参照ゲノムにおけるギャップに起因して、またはPPGが試料特異的PPGであるためのいずれかにより、参照ゲノムには存在しない。ある特定の実施形態では、1つまたは複数の試料特異的PPGにより、対象の集団において対象が識別される。ある特定の実施形態では、1つまたは複数のPPGは、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、ME T、MYC、およびNRASからなる群に由来する遺伝子のエクソン配列に由来する。あ

50

る特定の実施形態では、1つまたは複数のPPGは、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASからなる群に由来する1つまたは複数の配列からの2つまたはそれを上回るPPGを含む。ある特定の実施形態では、1つまたは複数のPPGは、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASからなる群に由来する1つまたは複数の配列からの3つまたはそれを上回るPPGを含む。

【0014】

ある特定の実施形態では、遺伝子バリエーションまたは真の遺伝子バリエーションは、単一ヌクレオチドバリエーション(SNV)または挿入もしくは欠失(インデル)を含む。ある特定の実施形態では、遺伝子バリエーションは、SNVを含む。ある特定の実施形態では、SNVは、イントロン-エクソン境界部に位置する。ある特定の実施形態では、SNVは、遺伝子コーディング配列(CDS)内に位置する。ある特定の実施形態では、遺伝子バリエーションは、インデルを含む。

【0015】

ある特定の実施形態では、領域は、遺伝子内融合ブレイクポイントに隣接する約2個、4個、6個、8個、10個、15個、または20個のヌクレオチドを含む。ある特定の実施形態では、領域は、融合ブレイクポイントから、約100、50、20、15、10、8、6、4、2ヌクレオチド未満である。ある特定の実施形態では、1つまたは複数の検出されたアライメントエラーの一部分は、試料中の突然変異対立遺伝子画分が、試料中の遺伝子内融合ブレイクポイントに対応する遺伝子内融合に対する突然変異対立遺伝子画分よりも低いかまたはそれと同等である、検出されたアライメントエラーに基づいて、フィルタリングされる。ある特定の実施形態では、1つまたは複数の検出されたアライメントエラーの一部分は、事前に定義された臨床的に対処可能なバリエーションのセットに属さない遺伝子バリエーションを含む遺伝子融合リードに基づいて、フィルタリングされる。

【0016】

ある特定の実施形態では、試料は、血液、血漿、血清、尿、唾液、粘膜排出物、喀痰、糞便、および涙液からなる群から選択される、体液試料である。ある特定の実施形態では、対象は、疾患または障害を有する。ある特定の実施形態では、疾患は、がんである。

【0017】

ある特定の実施形態では、本方法は、対象の生物学的試料から、無細胞核酸分子を単離するステップを含む。ある特定の実施形態では、無細胞核酸分子は、DNA、RNA、またはこれらの組合せを含む。ある特定の実施形態では、無細胞核酸分子は、無細胞DNAである。ある特定の実施形態では、無細胞核酸分子は、二本鎖DNAである。

【0018】

ある特定の実施形態では、本方法は、シーケンシングの前に、分子バーコードを含む1つまたは複数のアダプターを、無細胞核酸分子に結合させて、タグ付けされた親ポリヌクレオチドを生成するステップを含む。ある特定の実施形態では、アダプターは、無細胞核酸分子の両端に結合される。ある特定の実施形態では、無細胞核酸分子は、固有にバーコーディングされる。ある特定の実施形態では、無細胞核酸分子は、非固有にバーコーディングされる。ある特定の実施形態では、それぞれのバーコードは、選択された領域からシーケンシングされる多様な分子と組み合わせて、固有な分子の識別を可能にする、固定またはセミランダムなオリゴヌクレオチド配列を含む。

【0019】

ある特定の実施形態では、本方法は、タグ付けされた親ポリヌクレオチドを増幅させて、子孫ポリヌクレオチドを生成するステップを含む。ある特定の実施形態では、本方法は、目的の標的配列に関して、子孫ポリヌクレオチドを選択的に濃縮させ、それによって、濃縮された子孫ポリヌクレオチドを生成するステップを含む。ある特定の実施形態では、本方法は、濃縮された子孫ポリヌクレオチドを増幅させるステップを含む。ある特定の実施形態では、本方法は、子孫ポリヌクレオチドまたは濃縮された子孫ポリヌクレオチドに

10

20

30

40

50

、試料インデックス配列をタグ付けするステップを含む。

【0020】

ある特定の実施形態では、配列情報は、核酸シーケンサーから得られる。ある特定の実施形態では、遺伝子融合リードのセットは、シーケンシングされたペアエンドリードをアライメントおよび接続することによって識別される。ある特定の実施形態では、遺伝子融合リードのセットは、イントロン-エクソン境界部にまたがるカバレッジにおける不連続性に基づいて識別される。ある特定の実施形態では、事前に定義されたセットは、COSMIC、The Cancer Genome Atlas (TCGA)、または Exome Aggregation Consortium (ExAC) において見出されるバリエーションを含む。

10

【0021】

ある特定の実施形態では、本方法は、コンピュータにより実装することができ、その結果、湿式化学反応ステップ以外の本明細書および添付の特許請求の範囲に記載されるステップのうちのいずれかまたはすべては、好適なプログラミングされたコンピュータにおいて実施され得る。

【0022】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、遺伝子配列リードにおけるアライメントエラーを検出するための方法であって、対象からの生物学的試料中の無細胞核酸分子から得られた遺伝子配列リードを含む、配列情報を受信するステップと、遺伝子配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップとを含む、方法を実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

20

【0023】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、対象の生物学的試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、無細胞核酸分子から得られた配列リードを含む、配列情報を受信するステップと、配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む遺伝子融合リードのセットを識別するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、1つまたは複数の遺伝子融合リードのサブセットにおいて、1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

30

40

【0024】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、対象の試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、無細胞核酸分子から得られたシーケンシングリードを含む、配列情報を受信するステップと、配列リードを、参照配

50

列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む遺伝子融合リードのセットを識別するステップと、遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、遺伝子融合リードのうちの1つまたは複数のサブセットが、S M A D 4、T Y R O 3、および／またはR A F 1に対応する遺伝子配列を含む、ステップと、遺伝子融合リードのうちの1つまたは複数のサブセットにおいて、1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

10

【0025】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、遺伝子配列リードにおけるアライメントエラーを検出するための方法であって、対象からの生物学的試料中の無細胞核酸分子から得られた遺伝子配列リードを含む、配列情報を受信するステップと、遺伝子配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む遺伝子融合リードのセットを決定するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを決定するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、所定の基準を満たす、領域内のそれぞれの遺伝子バリエーションを、アライメントエラーとして識別するステップとを含む、方法を実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

20

【0026】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、対象の試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、対象からの生物学的試料中の無細胞核酸分子から得られた配列リードを含む、配列情報を受信するステップと、配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む遺伝子融合リードのセットを決定するステップと、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを決定するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、所定の基準を満たす、領域内のそれぞれの遺伝子バリエーションを、アライメントエラーとして識別するステップと、1つまたは複数の遺伝子融合リードのサブセットにおいて、1つまたは複数のアライメントエラーをフィルタリングして、フィルタリングされた配列リードを生成するステップと、参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法を実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

30

40

【0027】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、フィルタリングされた配列情報データセットを生成する方法であって、(a) 対象から得られた生物学的試料中のc f D N Aから得られた試験配列リードを含む、試験配列情報を受信するステップと、(b) 試験配列リードの中から、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブ

50

レイクポイントを含む、ステップと、(c) 試験配列情報において、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制し、それによって、フィルタリングされた配列情報データセットを生成するステップとを含む、方法を提供する。

【0028】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、フィルタリングされた配列情報データセットを生成する方法であって、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制し、それによって、フィルタリングされた配列情報データセットを生成するステップとを含む、方法を提供する。

10

【0029】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、フィルタリングされた配列情報データセットを生成する方法であって、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制し、それによって、フィルタリングされた配列情報データセットを生成するステップとを含む、方法を提供する。

20

【0030】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、フィルタリングされた配列情報データセットを生成する方法であって、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードを含む、試験配列情報を受信するステップと、(b) 試験配列リードの中から、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(c) 試験配列情報において、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制し、それによって、フィルタリングされた配列情報データセットを生成するステップとを含む、方法を提供する。

30

【0031】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、フィルタリングされた配列情報データセットを生成する方法であって、(a) 対象から得られた生物学的試料中の無細胞核酸(cfNA)から得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制し、それによって、フィルタリングされた配列情報データセットを生成するステップを含む、方法を提供する。

40

【0032】

50

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、フィルタリングされた配列情報データセットを生成する方法であって、(a) 対象から得られた生物学的試料中の無細胞核酸 (c f N A) から得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制し、それによって、フィルタリングされた配列情報データセットを生成するステップを含む、方法を提供する。

10

【0033】

ある特定の態様では、本開示は、フィルタリングされた配列情報データセットを生成する方法であって、(a) 対象から得られた生物学的試料中の無細胞デオキシリボ核酸 (c f D N A) をシーケンシングして、試験配列リードのセットを生成するステップと、(b) 試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(c) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制し、それによって、フィルタリングされた配列情報データセットを生成するステップとを含む、方法を提供する。

20

【0034】

ある特定の態様では、本開示は、少なくとも部分的にコンピュータを使用して、標的配列バリエーションを検出する方法であって、(a) 対象から得られた生物学的試料中の c f D N A から得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの非標的配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制して、フィルタリングされた配列情報データセットを生成するステップと、(c) フィルタリングされた配列情報データセットにおいて、標的配列バリエーションを含む、少なくとも1つの標的試験配列リードを識別し、それによって、標的配列バリエーションを検出するステップとを含む、方法を提供する。

30

【0035】

ある特定の態様では、本開示は、対象における疾患、障害、または状態を処置する方法であって、(a) 対象から得られた生物学的試料中の c f D N A から得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの非標的配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制して、フィルタリングされた配列情報データセットを生成するステップと、(c) フィルタリングされた配列情報データセットにおいて、対象における疾患、障害、または状態の指標である標的配列バリエーションを含む、少なくとも1つの標的試験配列リードを識別するステップと、(d) 疾患、障害、または状態を処置するのに有効な1つまたは複数の治療を、対象に投与し、それによって、対象における疾患、障害、または状態を処置するステップとを含む、方法を提供する。

40

【0036】

ある特定の実施形態では、本方法は、追加の試験配列リードが、S M A D 4、G N A S

50

、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASからなる群から選択される1つまたは複数の遺伝子配列の少なくとも一部分とアライメントする場合に、所与のブレイクポイントから選択されたヌクレオチド数以内にない1つまたは複数の配列バリエーションを含む、1つまたは複数の追加の試験配列リードを抑制するステップを含む。

【0037】

ある特定の実施形態では、所与のスプリット配列リードを識別するステップは、参照配列情報と部分的にしかアライメントしない試験配列リードを識別することを含む。ある特定の実施形態では、所与のスプリット配列リードを識別するステップは、参照配列情報と比べて、試験配列情報における1つまたは複数のゲノム領域を含むスプリット配列リードが欠如した1つまたは複数のゲノム領域のカバレッジの増加を識別することを含む。

10

【0038】

ある特定の実施形態では、1つまたは複数のゲノム領域は、少なくとも1つのコーディング配列(CDS)を含む。ある特定の実施形態では、所与のスプリット配列リードを識別するステップは、互いに異なり、それぞれが同一のブレイクポイントを含む、少なくとも2つのスプリット配列リードを識別することを含む。ある特定の実施形態では、本方法は、フィルタリングされた配列情報データセットにおいて、少なくとも1つの標的試験配列リードを識別するステップを含む。ある特定の実施形態では、標的試験配列リードは、対象における所与の疾患、障害、または状態の指標である標的配列バリエーションを含む。ある特定の実施形態では、本方法は、対象における所与の疾患、障害、または状態を処置するステップを含む。

20

【0039】

ある特定の実施形態では、抑制されたスプリット配列リードのうちの1つまたは複数、プロセスされた偽遺伝子(PPG)の少なくとも一部分を含む。ある特定の実施形態では、本方法は、試験配列情報から、所与のブレイクポイントから選択されたヌクレオチド数以内に配列バリエーションを含む、スプリット配列リードおよび／または試験配列リードを除去するステップを含む。

【0040】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)対象から得られた生物学的試料中のcfDNAから得られた試験配列リードを含む、試験配列情報を受信するステップと、(b)試験配列リードの中から、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(c)試験配列情報において、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

30

【0041】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b)試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

40

50

【0042】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b)試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

10

【0043】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)対象から得られた生物学的試料中のcfDNAから得られた試験配列リードを含む、試験配列情報を受信するステップと、(b)試験配列リードの中から、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(c)試験配列情報において、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

20

【0044】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)対象から得られた生物学的試料中の無細胞核酸(cfNA)から得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b)試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

30

【0045】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)対象から得られた生物学的試料中のcfNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b)試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

40

【0046】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)対象から得られた生物学的試料中のcfDNAをシーケンシングして、試験配列リードのセットを生成するステップと、(b)試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞ

50

れのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、
(c) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含むコンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

【0047】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの非標的配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制して、フィルタリングされた配列情報データセットを生成するステップと、(c) フィルタリングされた配列情報データセットにおいて、標的配列バリエーションを含む、少なくとも1つの標的試験配列リードを識別するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

【0048】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中の無細胞デオキシリボ核酸(cfDNA)から得られた試験配列リードを含む、試験配列情報を受信するステップと、(b) 試験配列リードの中から、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(c) 試験配列情報において、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

【0049】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

【0050】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つも

しくは複数のベースコールを抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

【0051】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードを含む、試験配列情報を受信するステップと、(b) 試験配列リードの中から、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(c) 試験配列情報において、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

10

【0052】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

20

【0053】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

30

【0054】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中の無細胞デオキシリボ核酸(cfdNA)をシーケンシングして、試験配列リードのセットを生成するステップと、(b) 試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(c) 試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

40

【0055】

ある特定の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a) 対象から得られた生物学的試料中のcfDNAから得られた試験配列リードのセットにおいて、1つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップと、(b) 試験配列リードのセットにおいて、所与のブレイクポイント

50

から選択されたヌクレオチド数以内に少なくとも1つの非標的配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分を抑制して、フィルタリングされた配列情報データセットを生成するステップと、(c) フィルタリングされた配列情報データセットにおいて、標的配列バリエーションを含む、少なくとも1つの標的試験配列リードを識別するステップとを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を提供する。

【図面の簡単な説明】

【0056】

【図1】図1は、プロセスされた偽遺伝子の存在に起因するアライメントエラーを検出および抑制するための例示的な方法を示す図である。

10

【0057】

【図2】図2は、本開示の一部の実施形態による、フィルタリングされた配列情報データセットを生成する例示的な方法ステップを概略的に示す、フローチャートである。

【0058】

【図3】図3は、本開示の一部の実施形態による、フィルタリングされた配列情報データセットを生成する例示的な方法ステップを概略的に示す、フローチャートである。

【0059】

【図4】図4Aは、プロセスされた偽遺伝子が生じるプロセスを示す図である。ヒトLINE ELEMENTに存在する非特異的逆転写酵素機序により、プロセシングされた（すなわち、イントロンのない）mRNAのDNAコピーが、作成され、ゲノムに組み込まれる。図4Bは、試料特異的PPGがヒトゲノムアセンブリ（たとえば、hg19）にはないため、偽遺伝子を起源とするリードが、どのようにして、もともとの遺伝子に固有にマッピングされ得るかを示す図である。しかしながら、偽遺伝子の存在は、イントロン-エクソン境界部にまたがるPPG断片を起源とするスプリットリードの存在によって判明し得る。

20

【0060】

【図5】図5は、本明細書に提供される方法を実装するようにプログラミングされるか、またはそうでなければそのように構成される、コンピュータシステムを示す図である。

【0061】

【図6】図6は、SMAD4エクソン11にマッピングされる配列リードを示す図である。単一の分子を起源とするリードは、共通した色（すなわち、灰色スケールの濃淡）およびゲノム座標でグループ分けされる。PPGの存在は、イントロン配列の配列が欠如した複数のソフトクリップリードの存在（リードの右手側の複数色のパターン）、ならびにイントロン-エクソン境界部におけるカバレッジの不連続性（図の上部）の両方によって判明する。矢印によって示される疑似A>C SNVコールは、1.7%の対立遺伝子頻度で観察される。

30

【0062】

【図7】図7Aは、PPGが検出された場合に、HRAS、SMAD4、およびPT53では、PPGを有さない試料において予測されるものよりも高いパーセンテージでスプライスジャンクションにおけるSNVコールが観察されることを示すグラフである。10,000個のランダムなバックグラウンド試料において、これらの同じジャンクション内でコールされるSNVはなく、結果として、灰色のバックグラウンドバーは、同じ高さ、0、となり、したがって、見えていない。図7Bは、PPGが検出された場合に、SNVが、SMAD4およびRAF1のコーディング配列(CDS)内で高いパーセンテージでコールされることを示すグラフである。≥PPGを有する試料を有するすべての遺伝子が示されているが、GNASもTP53も、PPGが存在する場合に、高いパーセンテージのCDS SNVコールを示さなかった。***p<0.01、*p<0.05、n.s. カイ二乗検定に基づいて有意差なし(1 d.f.)。

40

【0063】

【図8】図8は、ヒト第15染色体上のTYRO3にマッピングされる配列リードを示す

50

図である。単一の分子を起源とするリードは、共通した色（すなわち、灰色スケールの濃淡）およびゲノム座標でグループ分けされる。PPGによって生じるエクソン-エクソンジャンクションにまたがるアライメントアーチファクトは、TYRO3遺伝子座の状況で示される。疑似C.T. SNVコール(TYRO3 c. 1422 C>T)は、矢印で示されている。

【発明を実施するための形態】

【0064】

定義

「対象」という用語は、動物、たとえば、哺乳動物種（好ましくは、ヒト）または鳥類（たとえば、鳥）種を指し得る。より具体的には、対象は、脊椎動物、たとえば、哺乳動物、たとえば、マウス、霊長類、サル、またはヒトであり得る。動物には、家畜動物、競技動物、および愛玩動物が含まれる。対象は、健康な個体、症状もしくは徴候を有するかまたは疾患もしくは疾患の素因を有することが疑われる個体、または治療を必要とするかもしくは治療を必要とすることが疑われる個体であり得る。一部の実施形態では、対象は、ヒト、たとえば、がんを有するかまたはがんを有することが疑われるヒトである。

【0065】

「無細胞核酸」という語句は、細胞内に含まれていないかもしくは細胞に結合していない核酸、または換言すると、インタクトな細胞を除去した後に試料中に残存している核酸を指し得る。無細胞核酸は、対象からの体液（たとえば、血液、尿、CSFなど）から得られた非封入核酸と称され得る。無細胞核酸としては、DNA(cfDNA)、RNA(cfRNA)、およびそれらのハイブリッドが挙げられ、ゲノムDNA、ミトコンドリアDNA、循環DNA、siRNA、miRNA、循環RNA(cRNA)、tRNA、rRNA、核小体RNA(snoRNA)、Piwi結合RNA(piRNA)、長い非コーディングRNA(長いncRNA)、またはこれらのうちのいずれかの断片が含まれる。無細胞核酸は、二本鎖であっても、一本鎖であっても、部分的に二本鎖および一本鎖であってもよい。無細胞核酸は、分泌または細胞死プロセス、たとえば、細胞壊死およびアポトーシスを通じて、体液中に放出され得る。一部の無細胞核酸、たとえば、循環腫瘍DNA(ctDNA)は、がん細胞から、体液中に放出される。その他のものは、健常細胞から放出される。ctDNAは、非封入腫瘍由来断片化DNAであり得る。無細胞胎児DNA(cffDNA)は、母体血流中に自由に循環している胎児DNAである。無細胞核酸は、1つまたは複数の関連するエピジェネティックな改変を有し得、たとえば、アセチル化、5-メチル化、ユビキチン化、リン酸化、SUMO化、リボシル化、および／またはシトルリン化され得る。一部の実施形態では、無細胞核酸は、cfDNAであり、これは、通常、二本鎖cfDNAを含む。

【0066】

「核酸タグ」という語句は、核酸を、異なる種類または異なるプロセッシングを受けた、異なる試料（たとえば、試料インデックスを提示する）、または同じ試料中の異なる核酸分子（たとえば、分子バーコードを提示する）と区別するために、核酸分子を標識するのに使用される、短い核酸（たとえば、500、100、50、または10ヌクレオチド長を下回る）を指し得る。タグは、一本鎖であっても、二本鎖であっても、少なくとも部分的に二本鎖であってもよい。タグは、同じ長さを有してもよく、または変動した長さを有してもよい。タグは、平滑末端であってもよく、またはオーバーハングを有してもよい。タグは、核酸の一方の末端または両方の末端に結合され得る。核酸タグは、デコードされると、核酸の起源試料、形態、またはプロセッシングなどの情報を示し得る。タグを使用して、異なる分子タグおよび／または試料インデックスを有する核酸を含む複数の試料のプールおよび並行したプロセッシングを可能にすることができ、核酸は、後で分子タグを読み取ることによりデコンボリューションされる。追加または代替として、核酸タグは、同じ試料中の異なる分子（すなわち、分子バーコード）を区別するために使用することができる。これには、試料中の異なる分子を固有にタグ付けすること、または試料中の分子を非固有にタグ付けすることの両方が含まれる。非固有にタグ付けする場合には、異なる分子

が、少なくとも1つのタグと組み合わせて、参照ゲノム上にマッピングされるそれらの開始位置および／または終止位置（すなわち、ゲノム座標）に基づいて区別され得るように、限られた数の異なるタグを使用して、分子をタグ付けすることができる。そのため、典型的には、同じ開始／終止を有する任意の2つの分子が、同じタグも有する確率が低くなるように（たとえば、10%を下回る、5%を下回る、1%を下回る、または0.1%を下回る）、十分な数の異なるタグが使用される。一部のタグは、複数の試料、1つの試料内の複数の分子形態、ならびに同じ開始点および終止点を有する1つの形態内の複数の分子を標識するための複数の識別子を含む。そのようなタグは、形式A 1 iで存在し得、ここで、文字は、試料の種類を示し、アラビア数字は、試料内の分子の形態を示し、ローマ数字は、形態内の分子を示す。

【0067】

「アダプター」という用語は、通常、試料核酸分子のいずれかの末端または両方の末端への連結のために、少なくとも部分的に二本鎖になっている、短い核酸（たとえば、500、100、または50ヌクレオチド長を下回る）を指す。アダプターは、両方の末端においてアダプターが隣接した核酸分子の増幅を可能にするプライマー結合部位、および／または次世代シーケンシング（NGS）のためのプライマー結合部位を含むシーケンシングプライマー結合部位を含み得る。アダプターはまた、捕捉用プローブ、たとえば、フローセル支持体に結合したオリゴヌクレオチドの結合部位も含み得る。アダプターはまた、上述のようなタグも含み得る。タグがアンプリコンおよび核酸分子のシーケンシングリードに含まれるように、タグは、好ましくは、プライマーおよびシーケンシングプライマー結合部位に対して位置付けられる。同じかまたは異なる配列のアダプターが、核酸分子のそれぞれの末端に連結され得る。同じ配列のアダプターが、それぞれの末端に連結されることがあるが、ただし、バーコードは異なる。好ましいアダプターは、Y字型アダプターであり、その場合、一方の末端が、平滑末端であるかまたは核酸分子への結合のための尾部があり、この核酸分子もまた、平滑末端であるかまたは1つもしくは複数の相補的ヌクレオチドを有する尾部がある。別の好ましいアダプターは、釣鐘型アダプターであり、同様に、平滑末端または分析しようとする核酸への結合のための尾部を有する末端を有する。

【0068】

本明細書において使用される場合、「シーケンシング」または「シーケンサー」という用語は、生体分子、たとえば、核酸、たとえば、DNAまたはRNAの配列を判定するために使用されるいくつかの技術のうちのいずれかを指す。例示的なシーケンシング方法としては、標的化シーケンシング、単一分子リアルタイムシーケンシング、エクソンシーケンシング、電子顕微鏡に基づくシーケンシング、パネルシーケンシング、トランジスタに媒介されるシーケンシング、ダイレクトシーケンシング、ランダムショットガンシーケンシング、サンガージデオキシターミネーションシーケンシング、全ゲノムシーケンシング、ハイブリダイゼーションによるシーケンシング、ピロシーケンシング、二重鎖シーケンシング、サイクルシーケンシング、単一塩基伸長シーケンシング、固相シーケンシング、高スループットシーケンシング、大規模並列シグネチャーシーケンシング、エマルジョンPCR、低変性温度における共増幅PCR（COLD-PCR）、多重PCR、可逆的色素ターミネーターによるシーケンシング、ペアエンドシーケンシング、ニアタームシーケンシング（near-term sequencing）、エクソヌクレアーゼシーケンシング、ライゲーションによるシーケンシング、短いリードのシーケンシング、単一分子シーケンシング、合成によるシーケンシング、リアルタイムシーケンシング、逆ターミネーターシーケンシング、ナノポアシーケンシング、454シーケンシング、Solexa Genome Analyzerシーケンシング、SOLiD（商標）シーケンシング、MS-PETシーケンシング、およびこれらの組合せが挙げられるが、これらに限定されない。一部の実施形態では、シーケンシングは、遺伝子分析装置、たとえば、IlluminaまたはApplied Biosystemsから市販入手可能な遺伝子分析装置などによって実施することができる。

【0069】

「次世代シーケンシング」またはN G Sという語句は、従来のサンガー法およびキャピラリー電気泳動に基づくアプローチと比較して増加したスループットを有し、たとえば、一度に数十万個の比較的小さな配列リードを生成する能力を有する、シーケンシング技術を指す。次世代シーケンシング技法の一部の例としては、合成によるシーケンシング、ライゲーションによるシーケンシング、およびハイブリダイゼーションによるシーケンシングが挙げられるが、これらに限定されない。

【0070】

「DNA（デオキシリボ核酸）」という用語は、それぞれが4つの核酸塩基、すなわち、アデニン（A）、チミン（T）、シトシン（C）、およびグアニン（G）のうちの1つを含む、デオキシリボヌクレオシドを含む、ヌクレオチドの鎖を指す。「RNA（リボ核酸）」という用語は、それぞれが4つの核酸塩基、すなわち、A、ウラシル（U）、G、およびCのうちの1つを含む、4つの種類のリボヌクレオシドを含む、ヌクレオチドの鎖を指す。ある特定のヌクレオチド対は、相補的様式で互いに特異的に結合する（相補的塩基対合と称される）。DNAの場合、アデニン（A）はチミン（T）と対合し、シトシン（C）はグアニン（G）と対合する。RNAの場合、アデニン（A）はウラシル（U）と対合し、シトシン（C）はグアニン（G）と対合する。第1の核酸鎖が、第1の鎖のものに相補的なヌクレオチドでできた第2の核酸鎖に結合すると、2つの鎖は、結合して二本鎖を形成する。本明細書において使用される場合、「核酸シーケンシングデータ」、「核酸シーケンシング情報」、「核酸配列」、「ヌクレオチド配列」、「ゲノム配列」、「遺伝子配列」、または「断片配列」、または「核酸シーケンシングリード」は、核酸、たとえば、DNAまたはRNAの分子（たとえば、ゲノム全体、トランスクリプトーム全体、エクソーム、オリゴヌクレオチド、ポリヌクレオチド、または断片）におけるヌクレオチド塩基（たとえば、アデニン、グアニン、シトシン、およびチミンまたはウラシル）の順序を示す、任意の情報またはデータを指す。本教示は、キャピラリー電気泳動、マイクロアレイ、ライゲーションに基づくシステム、ポリメラーゼに基づくシステム、ハイブリダイゼーションに基づくシステム、直接的または間接的ヌクレオチド識別システム、ピロシーケンシング、イオンまたはpHに基づく検出システム、および電子シグネチャーに基づくシステムを含むがこれらに限定されない、すべての利用可能な種類の技法、プラットフォーム、または技術を使用して得られた配列情報を企図することを理解されたい。

【0071】

「ポリヌクレオチド」、「核酸」、「核酸分子」、または「オリゴヌクレオチド」は、ヌクレオシド間結合によって結合された、ヌクレオシドの線形ポリマー（デオキシリボヌクレオシド、リボヌクレオシド、またはそれらの類似体を含む）を指す。典型的には、ポリヌクレオチドは、少なくとも3つのヌクレオシドを含む。オリゴヌクレオチドは、多くの場合、サイズが、数モノマー単位、たとえば、3〜4個から、数百モノマー単位までの範囲に及ぶ。ポリヌクレオチドが、「A T G C C T G」などの文字列によって表される場合は常に、ヌクレオチドが、左から右に、5'から3'の順序になっており、別途示されない限り、「A」はアデノシンを示し、「C」はシトシンを示し、「G」はグアノシンを示し、「T」はチミジンを示すことが理解されるであろう。A、C、G、およびTという文字は、当該技術分野において標準的なように、塩基自体、ヌクレオシド、または塩基を含むヌクレオチドを指して使用され得る。

【0072】

「参照配列」という語句は、実験的に決定された配列と比較するために使用される公知の配列を指す。たとえば、公知の配列は、ゲノム全体、染色体、またはその任意のセグメントであり得る。参照物は、典型的には、少なくとも20個、50個、100個、200個、250個、300個、350個、400個、450個、500個、1000個、またはそれを上回るヌクレオチドを含む。参照配列は、ゲノムもしくは染色体の単一の連続配列とアライメントし得るか、またはゲノムもしくは染色体の異なる領域とアライメントする非連続的セグメントを含み得る。一部の実施形態では、参照配列は、ヒトゲノムである。参照ヒトゲノムとして、たとえば、h G 19およびh G 38が挙げられる。

【0073】

「偽遺伝子」という用語は、一般に、その遺伝子配列が、対応物である完全な遺伝子に類似であるが、細胞内遺伝子発現またはタンパク質コーディング能力における少なくとも一部の機能性が失われている、ゲノムDNAのセグメントを指す。偽遺伝子は、その対応物である機能性遺伝子に対して、高い程度の相同性または同一性を有し得る。一部の実施形態では、偽遺伝子は、対応物である機能性遺伝子と少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、または少なくとも95%の相同性を共有する。

【0074】

「プロセスされた偽遺伝子」という用語は、一般に、逆転写されたmRNA転写産物である相補的なDNA (cDNA) が、ゲノム内の新しい位置に再度組み込まれる、レトロ転位 (retrotransposition) のプロセスから生じる偽遺伝子を指す。プロセスされた偽遺伝子は、一般に、イントロンが欠如しており、それによって、エクソン-エクソンの遺伝子内 (すなわち、遺伝子の内部での) 融合が生じる。プロセスされた偽遺伝子の他の特徴としては、ポリA尾部、短縮された5'末端 (対応物である完全な遺伝子と比較して)、および転写機序 (たとえば、プロモーター領域) の欠如が挙げられる。

【0075】

本明細書において使用される「生物学的試料」という語句は、一般に、対象からの組織または体液試料を指す。生物学的試料は、対象から直接的に得ることができる。生物学的試料は、1つまたは複数の核酸分子、たとえば、デオキシリボ核酸 (DNA) またはリボ核酸 (RNA) 分子であり得るか、またはそれを含み得る。生物学的試料は、任意の器官、組織、または生物学的流体に由来し得る。生物学的試料は、たとえば、体液または固体組織試料を含み得る。固体組織試料の例は、たとえば、固形腫瘍生検からの腫瘍試料である。体液としては、たとえば、血液、血清、血漿、腫瘍細胞、唾液、尿、リンパ液、前立腺液、精液、乳、喀痰、糞便、涙液、およびこれらの誘導体が挙げられる。一部の実施形態では、生物学的試料は、血液であるか、または血液からである。

【0076】

「突然変異対立遺伝子画分」、「突然変異量 (mutation dose)」、または「MAF」という語句は、所与の試料中の、所与のゲノム位置に対立遺伝子の改変または突然変異を有する核酸分子の画分を指す。MAFは、一般に、画分またはパーセンテージとして表される。たとえば、MAFは、典型的に、所与の遺伝子座に存在するすべての体細胞系バリエーションまたは対立遺伝子のうち、約0.5、0.1、0.05、または0.01を下回る (すなわち約50%、10%、5%、または1%を下回る)。

【0077】

核酸配列情報の文脈における「スプリット配列リード」または「スプリットリード」または「遺伝子融合リード」という語句は、所与の参照配列の異なる非連続的な領域または遺伝子座にマッピングされる部分配列を含む、シーケンシングリードを指す。ある特定の実施形態では、たとえば、所与のスプリット配列リードの第1の部分配列は、参照配列の所与の遺伝子の第1のエクソンにマッピングされ、一方で、その所与のスプリット配列リードの第2の部分配列は、参照配列の同じ遺伝子の第2のエクソンにマッピングされ、この第1および第2のエクソンは、参照配列の同じ遺伝子の介在イントロンによって離間している。これらの実施形態のうちの一部では、そのようなスプリット配列リードは、所与のスプリット配列リードが得られた対象のゲノムにおける遺伝子内融合の存在を示す。他の例示的な実施形態では、所与のスプリット配列リードの第1の部分配列は、参照配列の第1の遺伝子のエクソンにマッピングされ、一方で、その所与のスプリット配列リードの第2の部分配列は、参照配列の異なる第2の遺伝子のエクソンにマッピングされ、これらのエクソンは、参照配列において互いに非連続的である。これらの実施形態のうちの一部では、そのようなスプリット配列リードは、所与のスプリット配列リードが得られた対象のゲノムにおける遺伝子間融合の存在を示す。

10

20

30

40

50

【0078】

核酸融合分子または対応するシーケンシングリードの文脈における「ブレイクポイント」という用語は、核酸融合の、または対応するシーケンシングリードにおいて表される、融合された部分配列間のジャンクションにおける末端ヌクレオチド位置を指す。たとえば、所与のスプリット配列リードは、第1の部分配列を含み得、これは、そのスプリット配列リードにおける第2の部分配列と連続的であり、その5'にあり、ここで、第1の部分配列は、参照配列において、第1の遺伝子座にマッピングされ、第1の遺伝子座は、その参照配列における第2の部分配列がマッピングされる第2の遺伝子座とは非連続的である。この例では、スプリット配列リードの第1の部分配列は、その3'末端ヌクレオチドにブレイクポイントを含むが、一方でスプリット配列リードの第2の部分配列は、その5'末端ヌクレオチドにブレイクポイントを含む。ある特定の適用では、これらのようなブレイクポイントは、「ブレイクポイント対」と称される。

10

【0079】

治療剤（たとえば、治療用核酸構築物）の文脈における「投与する」という語句は、その薬剤を、対象に与えること、対象に適用すること、または対象と接触させることを意味する。投与は、たとえば、局所、経口、皮下、筋肉内、腹腔内、静脈内、髄腔内、および皮内を含む、いくつかの経路のうちのいずれかによって達成することができる。

【0080】

目的とされる1つまたは複数の値または要素に適用される「約」または「およそ」という語句は、言及された参照値または要素に類似する値または要素を指す。ある特定の実施形態では、「約」または「およそ」という用語は、別途示されない限りまたは文脈からそうでないことが明らかでない限り、言及された参照値または要素のいずれかの方向で（より大きいまたは小さい）、25%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%以内、またはそれよりも小さい範囲に入る値または要素の範囲を指す（ただし、そのような数が可能な値または要素の100%を超える場合を除く）。

20

I. 概要

【0081】

臨床診断用シーケンシング試験の主な課題は、短いリードのアーチファクトを受けやすいゲノム領域を識別し、それらの影響を軽減することである。これらの領域の多くは、ヒトゲノムアセンブリの分析により識別されているが、しかしながら、野生型染色体の全体的な構造が同じ染色体上の非隣接ゲノム領域を近接させるように改変される試料特異的融合事象、または逆転写のアーチファクト、たとえば、プロセスされた偽遺伝子（PPG）の存在によって生じるものは、生殖細胞系および体細胞系のいずれも、適切に識別されなかった場合、体細胞系対立遺伝子頻度に、偽陽性バリエーションのコールをもたらす得る。試料ごとに、これらの融合事象によって生じるシグナルを識別することによって、本明細書に開示される方法およびシステムは、臨床的に誤った方向へ導くバリエーションの重要な供給源を識別および排除し、同時に感度に対する最小限の費用で高い特異度を維持することができる。

30

【0082】

本明細書に提供される方法およびシステムは、核酸分子、特に無細胞核酸分子の分析に、特に有用であり得る。一部の場合には、無細胞核酸分子は、対象からの生物学的試料から抽出および単離することができる。生物学的試料としては、血液、血漿、血清、尿、唾液、粘膜排出物、喀痰、糞便、および涙液を含むがこれらに限定されない群から選択される、体液試料を挙げることができる。無細胞核酸分子は、イソプロパノール沈降および／またはシリカに基づく精製を含むがこれらに限定されない、当該技術分野において公知の様々な方法を使用して抽出することができる。

40

【0083】

生物学的試料は、様々な対象、たとえば、疾患を有さない対象、疾患、たとえば、がんもしくはウイルスの危険性にあるか、その症状を示すか、もしくはそれを有する対象、ま

50

たは遺伝性障害の危険性にあるか、その症状を示すか、もしくはそれを有する対象から採取され得る。一部の実施形態では、疾患または障害は、免疫不全障害、血友病、サラセミア、鎌状赤血球症、血液疾患、慢性肉芽腫性障害、先天性盲目、リソソーム蓄積症、筋ジストロフィー、がん、神経変性疾患、またはこれらの組合せからなる群から選択される。一部の実施形態では、疾患は、がんである。

【0084】

無細胞核酸分子の獲得および提供の後に、核酸分子をシーケンシング用に調製するためのいくつかの異なるライブラリー調製手順のうちのいずれかを、無細胞核酸分子に実施することができる。無細胞核酸分子は、シーケンシングの前に、1つまたは複数の試薬（たとえば、酵素、アダプター、タグ（たとえば、バーコード）、プローブなど）で処理されてもよい。タグ付けされた分子は、次いで、下流の適用、たとえば、個々の分子を追跡することができるシーケンシング反応において、使用することができる。

【0085】

一部の実施形態では、本方法は、シーケンシングの前に、タグ付けされた分子の領域が選択的または非選択的に濃縮される、濃縮ステップをさらに含んでもよい。

【0086】

無細胞核酸分子のシーケンシングデータが収集されると、1つまたは複数のバイオフィォマティクスプロセスを、配列データに適用して、アライメントエラー（たとえば、偽陽性配列リード）、たとえば、PPGの存在によって引き起こされるものを検出し、遺伝子シーケンシング試験の結果を提供することにおいて、そのアライメントエラーを抑制または排除することができる。そのようなプロセスには、生殖細胞系および体細胞系遺伝子融合配列リードを識別するステップ、配列リード内の体細胞系単一ヌクレオチドバリエーション（SNV）および／または挿入もしくは欠失（インデル）を識別するステップ、遺伝子融合ブレイクポイント（たとえば、遺伝子内または遺伝子間）の領域内のアライメントエラーを判定するステップ、フィルターを適用して、所定の基準に基づいて、配列リードまたは最終的な検出されたバリエーションのセットからアライメントエラーを除去するステップ、ならびにフィルタリングされた配列リードから、真の遺伝子バリエーションを識別するステップが含まれ得るが、これらに限定されない。

【0087】

一部の場合には、シーケンシング反応により生成された配列リードは、バイオフィォマティクス分析を実行するために、参照配列に対してアライメントすることができる。バイオフィォマティクス分析の様々な態様では、1つまたは複数の閾値が、品質を確保するように設定され得る。たとえば、アライメント閾値は、高度に類似する配列リード（たとえば、参照配列と配列リードとの間のミスマッチが10個またはそれよりも少ない）だけが、参照配列にマッピングされるように、設定され得る。一部の場合には、たとえば、配列リードのクロマトグラムに基づいて、品質閾値を通過できない配列リードは、除去され得る。一部の場合には、所与の配列のコピー数または量は、所与の配列にマッピングまたはアライメントする配列リードの数に基づいて、定量され得る。一部の場合には、配列の過剰出現は、すべての配列リード間で異なる配列のコピー数または量を比較することによって判定され得る。

【0088】

ある特定の実施形態では、試料は、同じ核酸の任意の2つのコピーが、一方の末端または両方の末端に連結されたアダプターから、同じアダプター分子バーコードの組合せを收容する確率が低くなるように（たとえば1%を下回る）十分な数のアダプターと接触され得る。アダプターをこの様式で使用するにより、参照配列にアライメント（またはマッピング）される同じ開始点および終止点を有し、同じバーコードの組合せに連結された配列リードを、同じもとの分子から生成されたリードのファミリーにグループ分けすることが可能となり得る。そのようなファミリーは、増幅前の試料中の核酸の増幅産物の配列を表し得る。

【0089】

一部の実施形態では、ファミリーメンバーの配列をコンパイルして、平滑末端化およびアダプター結合によって改変された、もとの試料中の核酸分子のコンセンサヌクレオチドまたは完全なコンセンサス配列を導出することができる。換言すると、試料中の核酸の特定の位置を占有しているヌクレオチドは、ファミリーメンバー配列においてその対応する位置を占有しているヌクレオチドのコンセンサスであると決定され得る。コンセンサスヌクレオチドは、2つの非限定的で例示的な方法を挙げると、投票または信頼性スコアなどの方法によって、決定することができる。ファミリーには、二本鎖核酸の一方の鎖または両方の鎖の配列が含まれ得る。ファミリーのメンバーが、二本鎖核酸由来の両方の鎖の配列を含む場合、一方の鎖の配列を、すべての配列をコンパイルする目的で相補物に変換して、コンセンサスヌクレオチドまたは配列を導出する。一部のファミリーは、単一メンバー配列のみを含み得る。この場合には、この配列は、増幅前の試料中の核酸の配列として解釈され得る。あるいは、単一メンバー配列のみを有するファミリーは、後続の分析から排除してもよい。

【0090】

参照配列は、1つまたは複数の公知の配列、たとえば、所与の対象の公知の全体的または部分的なゲノム配列、たとえば、ヒト対象の全ゲノム配列であり得る。参照配列は、h G 19であってもよい。シーケンシングされた核酸は、試料中の核酸について直接的に決定された配列、または上述のように、そのような核酸の増幅産物の配列のコンセンサスを提示し得る。比較は、参照配列における1つまたは複数の指定位置で行われ得る。それぞれの配列を最大にアライメントしたときに、参照配列の指定位置に対応する位置を含む、シーケンシングされた核酸のサブセットを、識別することができる。そのようなサブセット内で、あるとすればどのシーケンシングされた核酸が、指定位置にヌクレオチド変動を含むか、および必要に応じて、あるとすればどれが参照ヌクレオチド（すなわち、参照配列におけるものと同じもの）を含むかを、判定することができる。ヌクレオチドバリエーションを含むサブセット内のシーケンシングされた核酸の数が、閾値を超えた場合、バリエーションヌクレオチドが、指定位置においてコールされ得る。閾値は、単純な数字、たとえば、少なくとも1、2、3、4、5、6、7、8、9、もしくは10個の、ヌクレオチドバリエーションを含むサブセット内のシーケンシングされた核酸であってもよく、または閾値は、比、たとえば、他の可能性の中でもとりわけ、少なくとも0.5、1、2、3、4、5、10、15、もしくは20パーセントの、ヌクレオチドバリエーションを含むサブセット内のシーケンシングされた核酸であってもよい。比較は、参照配列における任意の目的とされる指定位置について、繰り返すことができる。しばしば、比較は、参照配列上の少なくとも20個、100個、200個、または300個の連続した位置、たとえば、20～500個または50～300個の連続した位置を占有する指定位置について、行われ得る。

【0091】

図1は、アライメントエラーを検出および抑制するための方法の実施形態を示す。概して、本方法は、バリエーションコーラーおよび／または融合コーラーを使用して、所定の指定閾値セットに従って、可能性のある遺伝子バリエーションのセットを識別することができる。たとえば、バリエーションコーラーを使用して、指定閾値に従って、体細胞系SNVまたはインデルバリエーションのセットを識別することができ、融合コーラーを使用して、特定の閾値に従って、生殖細胞系および体細胞系遺伝子内（遺伝子の内部での）遺伝子融合のセットを識別することができる。そのような可能性のある遺伝子バリエーションのセットは、バリエーションが、プロセスされた偽遺伝子の存在を起源とする場合に、それが誤って遺伝子に割り当てられる可能性のある、1つまたは複数のアライメントエラーを含む場合がある（それによって、偽陽性遺伝子バリエーションの検出が引き起こされる）。そのようなアライメントエラーは、バリエーションコーリングプロセス中に、たとえば、そのような検出されたアライメントエラーを特定からフィルタリングもしくは除去すること、または可能性のある遺伝子バリエーションとしてのさらなる分析などによって、検出および抑制することができる。

【0092】

本明細書に開示される方法の態様をさらに例示するために、図2および3に、少なくとも

も部分的にコンピュータを使用して、フィルタリングされた配列情報データセットを生成するための例示的な方法ステップを概略的に示すフローチャートを提供する。本明細書に開示される方法のいずれも、必要に応じて、少なくとも部分的に、システムまたはコンピュータ可読媒体において実装または具現化され、これらも、本明細書にさらに説明されている。図2および3に示されるように、方法200および300は、いずれも、それぞれステップ202および302において、対象から得られた生物学的試料中のcfDNA分子または断片から得られた試験配列リードのセットにおいて、スプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも1つのブレイクポイントを含む、ステップを含む。典型的に、方法200および300は、それぞれが、ステップ202および302の前に対象から得られた生物学的試料中のcfDNA分子からの試験配列リードを含む、試験配列情報を（たとえば、電気通信ネットワークまたは他の通信もしくは記憶媒体を介して）受容するステップを含む。一部の実施形態では、方法200および300は、それぞれが、ステップ202および302の前に、対象から得られた生物学的試料中のcfDNA断片をシーケンシングして、試験配列リード（すなわち、試験配列情報）のセットを生成するステップを含む。

【0093】

スプリット配列リードまたはアライメントエラーは、必要に応じて、試料から得られた試験配列情報において、様々な技法のうちのいずれか1つまたは複数を使用して識別される。一部の実施形態では、スプリット配列リードは、試験配列情報セットにおいて、所与の参照配列情報セットと部分的にしかアライメントしない試験配列リードを識別することによって識別される。たとえば、スプリット配列リードは、典型的に、所与の参照ゲノム配列の第1の領域にマッピングされる少なくとも第1の部分配列、および所与の参照ゲノム配列の第2の領域にマッピングされる少なくとも第2の部分配列を含み、ここで、所与の参照ゲノム配列の第1および第2の領域は、互いに非連続的であるかまたは隣接していない。これらの実施形態のうちの一部では、本方法は、第1の遺伝子座（たとえば、所与の参照ゲノム配列の遺伝子内または遺伝子間の遺伝子座）にマッピングされる第1のブレイクポイントに隣接する第1の部分配列を識別するステップと、第2の異なる遺伝子座（たとえば、所与の参照ゲノム配列の非連続的遺伝子内または非連続的遺伝子間の遺伝子座）にマッピングされる第2のブレイクポイントに隣接する第2の部分配列を識別するステップとを含む。これらの実施形態では、第1のブレイクポイントおよび第2のブレイクポイントは、ブレイクポイント対を形成する。

【0094】

他の例示的な実施形態では、所与のスプリット配列リードまたはアライメントエラーは、参照配列情報と比べて、試験配列情報において観察される、ゲノム領域を含むスプリット配列リードが欠如しているゲノム領域（たとえば、コーディング配列（CDS）など）のカバレッジの増加を識別することによって識別される。一部の実施形態では、疑わしいスプリット配列または遺伝子融合（たとえば、プロセスされた偽遺伝子（PPG））は、スプリット配列リードが所与の試料中の異なるcfDNA断片を起源とすることを示し得る、それぞれが少なくとも1つの同一なブレイクポイントを含むが、それ以外は所与の特性、たとえば、長さに関して、互いに異なる、少なくとも2つの（たとえば、2つ、3つ、4つ、5つ、6つ、7つ、8つ、9つ、10個、またはそれを上回る）スプリット配列リード以外は、コールされない。これにより、典型的に、真のスプリット配列または遺伝子融合が所与の試料において観察されている、信頼性レベルが増加する。必要に応じて本開示の方法および関連する態様での使用に適合される、スプリット配列リードおよび遺伝子融合を識別することに関するさらなる詳細は、たとえば、WO2017/062970およびWO2018/213814号に提供されており、これらは、それぞれ、参照により本明細書に組み込まれる。

【0095】

また図2および3に示されるように、方法200は、ステップ204において、試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内

に少なくとも1つの配列バリエーションを含む、スプリット配列リードのうちの1つもしくは複数の少なくとも一部分（たとえば、所与のリードおよび／もしくは全リードの少なくとも一部分）および／または試験配列リードのうちの1つもしくは複数の少なくとも一部分（たとえば、所与のリードおよび／もしくは全リードの少なくとも一部分）を抑制するステップを含み、一方で、方法300は、ステップ304において、試験配列リードのセットにおいて、所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも1つの配列バリエーションを含む、スプリット配列リードの1つもしくは複数のベースコールおよび／または試験配列リードの1つもしくは複数のベースコールを抑制して、フィルタリングされた配列情報データセットを生成するステップを含む。配列リード（またはその部分）および／またはベースコールは、典型的に、所与のデータセットからその情報を除去することによって、または単純にデータセットの所与の適用においてその情報を使用しないことによって、「抑制」される。一部の例示的な実施形態では、本明細書に記載されるように、抑制されたスプリット配列リードは、プロセスされた偽遺伝子（PPG）の少なくとも一部分を含む。

【0096】

一部の実施形態では、所与のブレイクポイントから選択されたヌクレオチド数以内の配列バリエーションは、生物学的試料中のブレイクポイントの突然変異対立遺伝子画分（MAF）よりも低いかまたはそれに等しいMAFを含む。必要に応じて、他の数が使用されるが、所与のブレイクポイントからの選択されたヌクレオチドの数は、典型的に、約1個、2個、3個、4個、5個、6個、7個、8個、9個、19個、11個、12個、13個、14個、15個、16個、17個、18個、19個、20個、またはそれを上回るヌクレオチドを含む。他の実施形態では、所与のブレイクポイントからのヌクレオチドの数は、100個、50個、20個、15個、10個、8個、6個、4個、または2個を下回るヌクレオチドを含んでもよい。加えて、所与のブレイクポイントからの選択された数のヌクレオチドは、所与のブレイクポイントに対して5'および／または3'（すなわち、所与のブレイクポイントのいずれかの側または両側）に位置している。本明細書に記載されるように、様々な種類の配列バリエーションが、必要に応じて、本開示の方法を実施することにおいて使用される。これらの実施形態のうちの一部では、たとえば、配列バリエーションは、単一ヌクレオチドバリエーション（SNV）および／または挿入もしくは欠失（インデル）を含む。ある特定の実施形態では、本方法は、追加の試験配列リードが、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASからなる群から選択される1つまたは複数の遺伝子配列の少なくとも一部分とアライメントする場合に、所与のブレイクポイントから選択されたヌクレオチド数以内にない1つまたは複数の配列バリエーションを含む、1つまたは複数の追加の試験配列リードまたはその部分を抑制するステップを含む。

【0097】

本明細書に開示される方法を使用して得られたフィルタリングされた配列情報データセットは、多種多様な用途において使用することができる。典型的に、それらは、対象が、所与の疾患、障害、または状態を有するかどうかを判定するために、対象から得られた試験試料において、臨床的有意性のある配列バリエーションを識別することを容易にするために使用される。ある特定の実施形態では、特定の疾患、障害、または状態が、そのようにして診断されると、本方法は、本明細書にさらに記載されるように、対象におけるその疾患、障害、または状態を処置するために、1つまたは複数の治療を対象に投与するステップをさらに含む。

【0098】

遺伝子融合は、液体生検アッセイを使用して無細胞DNA（たとえば、循環腫瘍DNA、ctDNAを含む）における体細胞系ゲノム改変を識別することによって、対象の試料から識別され得る。そのようなアッセイは、無細胞DNA分子をシーケンシングして、配列リードを生成すること、ならびに遺伝子マーカー（たとえば、ALK、FGFR2、FGFR3、NTRK1、RET、およびROS1）のパネルを使用して配列リードを分析

することを含み得る。

【0099】

PPGは、生殖細胞系起源または体細胞系起源であり得、1つまたは複数の遺伝子座においてゲノム全体にわたる配列リードカバレッジデータを分析することによって識別することができる。たとえば、PPGは、アライメントアーチファクトがエクソン-エクソンジャンクションにまたがって観察される位置に見出され得る。PPGの存在は、イントロン配列が欠如した複数のソフトクリップリード（すなわち、配列リードの一部が参照配列にアライメントしないリード）の存在によって、またはイントロン-エクソン境界部におけるカバレッジの不連続性によって、判明し得る。PPGは、たとえば、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASのエクソン配列に由来し得る。

10

【0100】

1つまたは複数の基準を使用して、可能性のあるアライメントエラーを識別することができる。たとえば、遺伝子内融合ブレイクポイントを含む遺伝子融合に対応する配列リードのセット（遺伝子融合リード）のうち、可能性のあるアライメントエラーは、遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合とオーバーラップするリードのサブセットから識別され得る。この領域は、遺伝子内融合ブレイクポイントに隣接する20個またはそれを下回るヌクレオチド（たとえば、約20個、15個、10個、8個、6個、4個、または2個のヌクレオチド）を含み得る。遺伝子融合リードのセットは、1つまたは複数のプロセスされた偽遺伝子（PPG）、たとえば、試料特異的PPG（所与の試料または対象に特異的であり、参照ヒトゲノム、たとえば、hg19において一般に見出されないもの）に対応し得る。遺伝子バリエーションは、単一ヌクレオチドバリエーション（SNV）または挿入もしくは欠失（インデル）を含み得る。たとえば、SNVは、イントロン-エクソン境界部に位置し得るか、または遺伝子コーディング配列（CDS）内に位置し得る。

20

【0101】

別の例として、遺伝子融合に対応する配列リードのセット（遺伝子融合リード）のうち、可能性のあるアライメントエラーは、SMAD4、TYRO3、および／またはRAF1遺伝子において検出されている遺伝子融合リードのサブセットから識別され得る。

【0102】

識別されている、可能性のあるアライメントエラーは、真の遺伝子バリエーション（たとえば、対象の試料からの無細胞DNA分子からである）を検出することにおいて、抑制され得る。たとえば、そのような識別されている可能性のあるアライメントエラーの少なくとも一部分を、遺伝子融合リードのセットからフィルタリングして、フィルタリングされた配列リードを生成することができる。そのようなフィルタリングされた配列リードを、次いで、プロセッシングまたは分析して、参照配列と比較して、真の遺伝子バリエーション（たとえば、PPGの存在の結果として偽陽性バリエーションによって引き起こされるものではないもの）を検出し、それによって、有利なことに、バリエーションの偽陽性検出の比率を減少させることができる。結果として、バリエーションは、対象から得られた試料の分析により、より高い精度、感度、特異度、陽性的中率（PPV）、陰性的中率（NPV）、または曲線下面積（AUC）で、識別され得る。

30

【0103】

一部の場合には、検出されたアライメントエラーの一部分は、試料中の突然変異対立遺伝子画分（MAF）が、試料中の遺伝子内融合ブレイクポイントに対応する遺伝子内融合のMAFよりも低いかまたはそれと同等である、検出されたアライメントエラーに基づいて、フィルタリングされる。融合に媒介されるエラーは、融合スパニングリードに見出され得るため、偽陽性アライメントエラーは、試料中のMAFが、試料中の遺伝子内融合ブレイクポイントに対応する遺伝子内融合のMAFよりも大きくなる可能性はない。

40

【0104】

一部の場合には、検出されたアライメントエラーの一部分は、事前に定義された臨床的

50

に対処可能なバリエーションのセットに属さない遺伝子バリエーションを含む遺伝子融合リードに基づいて、フィルタリングされる。そのような「ホワイトリスト」のバリエーションは、対象の試料中の存在が、対象における疾患または障害（たとえば、がん）と相関性があるかまたはその指標であることが示されている、バリエーションの様々なデータベースにおいて見出すことができる。そのようなバリエーションのデータベースとしては、たとえば、Catalogue of Somatic Mutations in Cancer (COSMIC)、The Cancer Genome Atlas (TCGA)、および Exome Aggregation Consortium (ExAC) を挙げることができる。事前に定義されたそのようなカタログにあるバリエーションのセットは、臨床決定（たとえば、診断、予後診断、処置選択、標的化処置、処置モニタリング、再発のモニタリングなど）との関連性に起因して、さらなるバイオインフォマティクス分析に指定することができる。そのような事前に定義されたセットは、たとえば、臨床試料（たとえば、疾患または障害の存在または不在が判明している患者コホートのもの）の分析、ならびに公的データベースおよび臨床文献から得られる注釈情報に基づいて、判定することができる。

【0105】

アライメントエラーを特定および抑制した後、フィルタリングされた配列リードのセットは、参照配列と比較して、真の遺伝子バリエーションを検出するために分析することができる。

【0106】

本開示はさらに、本明細書に開示される方法ステップが、必要に応じて、本明細書に開示されるシステムおよび／またはコンピュータ可読媒体を使用した実施に適合されることを提供する。ある特定の態様では、システムは、少なくとも1つの電子プロセッサによって実行されると、本明細書に記載される方法のうちの少なくとも1つを実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含み得る。

【0107】

一部の実施形態では、シーケンサーは、DNAシーケンサーである。一部の実施形態では、シーケンサーは、高スループットシーケンシング、たとえば、次世代シーケンシングを実施するように設計される。一部の実施形態では、システムは、シーケンサーにおいて、アダプターがタグ付けされたcfDNA分子を含む。一部の実施形態では、アダプターがタグ付けされたcfDNA分子は、1つの対象または複数の対象から得られる。一部の実施形態では、試料からのcfDNA分子は、固有または非固有のバーコードを有する。

【0108】

一部の実施形態では、コンピュータプロセッサによって実装される方法は、配列リードをファミリーにグループ分けするステップであって、ファミリーのそれぞれが、同じバーコードを含み同じ開始位置および終止位置を有する配列リードを含み、それによって、ファミリーのそれぞれが、同じもともとのcfDNA分子から増幅された配列リードを含む、ステップをさらに含む。

【0109】

一部の実施形態では、本明細書に記載される方法およびシステムは、デジタル処理デバイスを利用する。さらなる実施形態では、デジタル処理デバイスは、デバイスの機能を実行する1つまたは複数のハードウェア中央処理装置(CPU)または汎用グラフィック処理装置(GPGPU)を含む。なおさらなる実施形態では、デジタル処理デバイスは、実行可能命令を実施するように構成される、オペレーティングシステムをさらに含む。一部の実施形態では、デジタル処理デバイスは、必要に応じて、コンピュータネットワークに接続されている。さらなる実施形態では、デジタル処理デバイスは、必要に応じて、ワールドワイドウェブにアクセスするように、インターネットに接続されている。なおさらなる実施形態では、デジタル処理デバイスは、必要に応じて、クラウドコンピューティングインフラストラクチャに接続されている。他の実施形態では、デジタル処理デバイスは、必要に応じて、イントラネットに接続されている。他の実施形態では、デジタル処理

10

20

30

40

50

デバイスは、必要に応じて、データ記憶デバイスに接続されている。本明細書の説明によると、好適なデジタル処理デバイスとしては、非限定的な例として、サーバーコンピュータ、デスクトップコンピュータ、ラップトップコンピュータ、ノートブックコンピュータ、ハンドヘルドコンピュータ、インターネット機器、モバイルスマートフォン、およびタブレットコンピュータが挙げられる。

【0110】

一部の実施形態では、デジタル処理デバイスは、実行可能命令を実施するように構成される、オペレーティングシステムを含む。オペレーティングシステムは、たとえば、デバイスのハードウェアを管理し、アプリケーションの実行のためのサービスを提供する、プログラムおよびデータを含む、ソフトウェアである。当業者であれば、好適なサーバーオペレーティングシステムとして、非限定的な例として、FreeBSD、OpenBSD、NetBSD（登録商標）、Linux（登録商標）、Apple（登録商標）Mac OS X Server（登録商標）、Oracle（登録商標）Solaris（登録商標）、Windows Server（登録商標）、およびNovell（登録商標）NetWare（登録商標）が挙げられることを認識するであろう。当業者であれば、好適なパーソナルコンピュータオペレーティングシステムとして、非限定的な例として、Microsoft（登録商標）Windows（登録商標）、Apple（登録商標）Mac OS X（登録商標）、UNIX（登録商標）、およびGNU/Linux（登録商標）などのUNIX（登録商標）様オペレーティングシステムが挙げられることを認識するであろう。一部の実施形態では、オペレーティングシステムは、クラウドコンピューティングによって提供される。当業者であれば、好適なモバイルスマートフォンオペレーティングシステムとして、非限定的な例として、Nokia（登録商標）Symbian（登録商標）OS、Apple（登録商標）iOS（登録商標）、Research In Motion（登録商標）BlackBerry OS（登録商標）、Google（登録商標）Android（登録商標）、Microsoft（登録商標）Windows Phone（登録商標）OS、Microsoft（登録商標）Windows Mobile（登録商標）OS、Linux（登録商標）、およびPalm（登録商標）Web OS（登録商標）が挙げられることも認識するであろう。

【0111】

一部の実施形態では、デバイスには、記憶および／またはメモリデバイスが含まれる。記憶および／またはメモリデバイスは、データまたはプログラムを、一時的または恒久的に記憶するために使用される、1つまたは複数の物理的装置である。一部の実施形態では、デバイスは、揮発性メモリであり、記憶された情報を維持するために電力を必要とする。一部の実施形態では、デバイスは、不揮発性メモリであり、デジタル処理デバイスが電力供給を受けていない場合にも記憶された情報を保持する。さらなる実施形態では、不揮発性メモリは、フラッシュメモリを含む。一部の実施形態では、不揮発性メモリは、ダイナミックランダムアクセスメモリ（DRAM）を含む。一部の実施形態では、不揮発性メモリは、強誘電体ランダムアクセスメモリ（FRAM（登録商標））を含む。一部の実施形態では、不揮発性メモリは、相変化ランダムアクセスメモリ（PRAM）を含む。他の実施形態では、デバイスは、記憶デバイスであり、これには、非限定的な例として、CD-ROM、DVD、フラッシュメモリデバイス、磁気ディスクドライブ、磁気テープドライブ、光学ディスクドライブ、およびクラウドコンピューティングに基づく記憶装置が含まれる。さらなる実施形態では、記憶および／またはメモリデバイスは、本明細書に開示されるものなどのデバイスの組合せである。

【0112】

一部の実施形態では、デジタル処理デバイスには、視覚的情報をユーザーに送信するための電子ディスプレイが含まれる。一部の実施形態では、ディスプレイは、液晶ディスプレイ（LCD）である。さらなる実施形態では、ディスプレイは、薄膜トランジスタ液晶ディスプレイ（TFT-LCD）である。一部の実施形態では、ディスプレイは、有機発光ダイオード（OLED）ディスプレイである。様々なさらなる実施形態では、OLED

ディスプレイは、パッシブマトリックスOLED（PMOLED）またはアクティブマトリックスOLED（AMOLED）ディスプレイである。一部の実施形態では、ディスプレイは、プラズマディスプレイである。他の実施形態では、ディスプレイは、ビデオプロジェクターである。なおも他の実施形態では、ディスプレイは、デジタル処理デバイスと通信する頭部装着型ディスプレイ、たとえば、VRヘッドセットである。さらなる実施形態では、好適なVRヘッドセットとしては、非限定的な例として、HTC Vive、Oculus Rift、Samsung Gear VR、Microsoft HoloLens、Razer OSVR、FOVE VR、Zeiss VR One、Avegant Glyph、Freefly VRヘッドセットなどが挙げられる。なおもさらなる実施形態では、ディスプレイは、本明細書に開示されるものなどのデバイスの組合せである。

10

【0113】

一部の実施形態では、デジタル処理デバイスは、ユーザーから情報を受信するための入力デバイスを含む。一部の実施形態では、入力デバイスは、キーボードである。一部の実施形態では、入力デバイスは、非限定的な例として、マウス、トラックボール、トラックパッド、ジョイスティック、ゲームコントローラー、またはスタイラスを含む、ポインティングデバイスである。一部の実施形態では、入力デバイスは、タッチスクリーンまたはマルチタッチスクリーンである。他の実施形態では、入力デバイスは、音声または他の音による入力を捕捉するためのマイクロホンである。他の実施形態では、入力デバイスは、動きまたは視覚的入力を捕捉するためのビデオカメラまたは他のセンサーである。さらなる実施形態では、入力デバイスは、Kinect、Leap Motionなどである。なおもさらなる実施形態では、入力デバイスは、本明細書に開示されるものなどのデバイスの組合せである。

20

【0114】

一部の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、本明細書に提供される方法を実施する、非一過的コンピュータ実行可能命令を含む、コンピュータ可読媒体を含むかまたはそれにアクセスすることができる、コントローラーを含む、システムを提供する。

【0115】

図5は、本明細書に提供される方法を実装するようにプログラミングされるか、またはそうでなければそのように構成される、コンピュータシステム501を示す。

30

【0116】

コンピュータシステム501は、遺伝子配列リードにおけるアライメントエラーを検出および／または抑制するための方法を実装するようにプログラミングされ得るか、またはそうでなければそのように構成され得る。コンピュータシステム501は、本開示の様々な態様、たとえば、（a）生物学的試料中の無細胞核酸分子をシーケンシングして、遺伝子配列リードを生成するステップ、（b）遺伝子配列リードを参照配列に対してアライメントして、アライメントされた配列リードを生成するステップ、（c）アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップ、（d）遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップ、（e）遺伝子融合リードのサブセットにおいて、検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップ、および（f）参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップなどを統制することができる。コンピュータシステム501は、ユーザーの電子デバイスであってもよく、または電子デバイスに対して遠隔に位置するコンピュータシステムであってもよい。電子デバイスは、携帯型電子デバイスであってもよい。

40

【0117】

50

コンピュータシステム501は、中央処理装置（CPU、本明細書において「プロセッサ」および「コンピュータプロセッサ」とも）505を含み、これは、シングルコアもしくはマルチコアプロセッサであり得るか、または並列処理のための複数のプロセッサであり得る。コンピュータシステム501はまた、メモリまたはメモリ位置510（たとえば、ランダムアクセスメモリ、読み出し専用メモリ、フラッシュメモリ）、電子記憶装置515（たとえば、ハードディスク）、1つまたは複数の他のシステムと通信するための通信インターフェース520（たとえば、ネットワークアダプター）、ならびに周辺デバイス525、たとえば、キャッシュ、他のメモリ、データ記憶装置、および／もしくは電子ディスプレイアダプターも含む。メモリ510、記憶装置515、インターフェース520、および周辺デバイス525は、通信バス（実線）、たとえば、マザーボードを通じて、CPU505と通信状態にある。記憶装置515は、データを記憶するためのデータ記憶装置（またはデータリポジトリ）であり得る。コンピュータシステム501は、通信インターフェース520を利用して、コンピュータネットワーク（「ネットワーク」）530に作動可能に連結され得る。ネットワーク530は、インターネット、インターネット（internet）および／もしくはエクストラネット、またはインターネットと通信するイントラネットおよび／またはエクストラネットであり得る。ネットワーク530は、一部の場合には、電気通信および／またはデータネットワークである。ネットワーク530は、クラウドコンピューティングなどの分散型コンピューティングを可能にすることができる、1つまたは複数のコンピュータサーバーを含み得る。ネットワーク530は、一部の場合には、コンピュータシステム501を利用して、コンピュータシステム501に連結されたデバイスが、クライアントまたはサーバーとして挙動することを可能にし得る、ピアトゥピアネットワークを実装することができる。

【0118】

CPU505は、プログラムまたはソフトウェアで具現化され得る、一連の機械可読命令を実行することができる。命令は、メモリ510などのメモリ位置に記憶され得る。命令は、CPU505へと指示され得、これが、続いて、CPU505を、本開示の方法を実装するようにプログラミングするか、またはそうでなければそのように構成し得る。CPU505によって実施される動作の例としては、フェッチ、復号、実行、およびライトバックを挙げることができる。

【0119】

CPU505は、集積回路など、回路の一部であり得る。システム501の1つまたは複数の他の構成要素が、回路に含まれてもよい。一部の場合には、回路は、特定用途向け集積回路（ASIC）である。

【0120】

記憶装置515は、ドライバ、ライブラリー、および保存されたプログラムなど、ファイルを記憶することができる。記憶装置515は、ユーザーデータ、たとえば、ユーザーの選好およびユーザーのプログラムを記憶することができる。コンピュータシステム501は、一部の場合には、コンピュータシステム501に対して外部である、たとえば、イントラネットまたはインターネットを通じてコンピュータシステム501と通信するリモートサーバーに位置する、1つまたは複数の追加のデータ記憶装置を含み得る。

【0121】

コンピュータシステム501は、ネットワーク530を通じて、1つまたは複数のリモートコンピュータシステムと通信することができる。たとえば、コンピュータシステム501は、ユーザーのリモートコンピュータシステムと通信し得る。リモートコンピュータシステムの例としては、パーソナルコンピュータ（たとえば、ポータブルPC）、スレートもしくはタブレットPC（たとえば、Apple（登録商標）のiPad（登録商標）、Samsung（登録商標）のGalaxy Tab）、電話、スマートフォン（たとえば、Apple（登録商標）のiPhone（登録商標）、Android対応デバイス、BlackBerry（登録商標））、または携帯情報端末が挙げられる。ユーザーは、ネットワーク530を介して、コンピュータシステム501にアクセスすることがで

10

20

30

40

50

きる。

【0122】

本明細書に記載される方法は、コンピュータシステム501の電子記憶位置、たとえば、メモリ510または電子記憶装置515などに記憶された、機械（たとえば、コンピュータプロセッサ）により実行可能なコードを用いて実装され得る。機械により実行可能なコードまたは機械により読み取り可能なコードは、ソフトウェアの形態で提供され得る。使用の際、コードが、プロセッサ505によって実行され得る。一部の場合には、コードは、記憶装置515から取り出され、プロセッサ505による即時アクセスのために、メモリ510に記憶され得る。一部の状況では、電子記憶装置515は、除外され得、機械により実行可能な命令は、メモリ510に記憶される。

10

【0123】

コードは、コードを実行するように適合されたプロセッサを有する機械で使用するために、プリコンパイルされ、そのように構成され得るか、または実行時にコンパイルされてもよい。コードは、コードがプリコンパイルまたは即時コンパイルの様式で実行されるのを可能にするように選択され得る、プログラミング言語で供給され得る。

【0124】

コンピュータシステム501など、本明細書において提供されるシステムおよび方法の態様は、プログラミングで具現化され得る。この技術の様々な態様は、典型的には、機械（またはプロセッサ）により実行可能なコードおよび／または機械可読媒体の一種で運搬もしくは具現化される関連データの形態をした、「製品」または「製造品」とであると考えることができる。機械により実行可能なコードは、メモリ（たとえば、読み出し専用メモリ、ランダムアクセスメモリ、フラッシュメモリ）、またはハードディスクなど、電子記憶装置に記憶され得る。「記憶装置」型の媒体としては、コンピュータ、プロセッサなどのありとあらゆる有形メモリ、またはその関連モジュール、たとえば、様々な半導体メモリ、テープドライブ、ディスクドライブなどが含まれ得、これらは、ソフトウェアプログラミングの任意の時点において、非一過的記憶装置を提供し得る。ソフトウェアのすべてまたは一部分は、時折、インターネットまたは様々な他の電気通信ネットワークを通じて通信され得る。そのような通信は、たとえば、ソフトウェアを1つのコンピュータまたはプロセッサから別のものへ、たとえば、管理サーバーまたはホストコンピュータからアプリケーションサーバーのコンピュータプラットフォームへ、ロードすることを可能にし得る。したがって、ソフトウェア要素を保持し得る別の種類の媒体としては、光波、電波、および電磁波、たとえば、ローカルデバイス間の物理的インターフェースで、有線および光学の地上ネットワークを通じて、ならびに様々なエアリンク上で、使用されるものが挙げられる。そのような波を有する物理的要素、たとえば、有線または無線リンク、光学リンクなどもまた、ソフトウェアを保持する媒体と考えることができる。本明細書において使用される場合、非一過的有形「記憶」媒体に限定されない限り、コンピュータまたは機械「可読媒体」などの用語は、実行のために命令をプロセッサに提供することに関する任意の媒体を指す。

20

30

【0125】

したがって、コンピュータにより実行可能なコードなど、機械可読媒体は、有形記憶媒体、搬送波媒体、または物理的伝送媒体を含むがこれらに限定されない、多数の形態を取り得る。不揮発性記憶媒体としては、たとえば、光ディスクまたは磁気ディスク、たとえば、任意のコンピュータなどにおける記憶デバイスのうちのいずれか、たとえば、図面に示されているデータベースなどを実装するために使用することができるものが挙げられる。揮発性記憶媒体としては、ダイナミックメモリ、たとえば、そのようなコンピュータプラットフォームの主メモリが挙げられる。有形伝送媒体としては、同軸ケーブル、銅線、および光ファイバ、たとえば、コンピュータシステム内のバスを含む電線が挙げられる。搬送波伝送媒体は、電気信号もしくは電磁信号または音波もしくは光波、たとえば、無線（RF）および赤外線（IR）データ通信の際に生成されるものの形態を取り得る。したがって、コンピュータ可読媒体の一般的な形態としては、たとえば、次のものが挙げられ

40

50

る：フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、CD-ROM、DVDもしくはDVD-ROM、任意の他の光学媒体、パンチカード紙テープ、任意の他の孔のパターンによる物理的記憶媒体、RAM、ROM、PROMおよびEPROM、FLASH（登録商標）-EPROM、任意の他のメモリチップもしくはカートリッジ、データもしくは命令を送送する搬送波、そのような搬送波を送送するケーブルもしくはリンク、またはコンピュータがプログラミングコードおよび／もしくはデータを読み取ることができる任意の他の媒体。これらのコンピュータ可読媒体の形態の多くは、実行のために1つまたは複数の命令の1つまたは複数のシーケンスをプロセッサに搬送することに関与し得る。

【0126】

コンピュータシステム501は、ユーザーインターフェース（UI）540を含む、電子ディスプレイ535を含み得るか、またはそれと通信し得る。UIの例としては、グラフィカルユーザーインターフェース（GUI）およびウェブベースユーザーインターフェースが挙げられるが、これらに限定されない。

【0127】

本開示の方法およびシステムは、1つまたは複数のアルゴリズムを用いて実施することができる。アルゴリズムは、中央処理装置505により実行すると、ソフトウェアによって実装され得る。アルゴリズムは、たとえば、（a）生物学的試料中の無細胞核酸分子をシーケンシングして、遺伝子配列リードを生成すること、（b）遺伝子配列リードを参照配列に対してアライメントして、アライメントされた配列リードを生成すること、（c）アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別すること、（d）遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出することであって、この領域が、遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含むこと、（e）遺伝子融合リードのサブセットにおいて、検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成すること、および（f）参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出することを行うことができる。

II. 方法の一般的な特徴

A. 試料

【0128】

試料は、対象から単離された任意の生物学的試料であり得る。試料としては、体組織、たとえば、判明しているかまたは疑われる固形腫瘍、全血、血小板、血清、血漿、糞便、赤血球、白血球（white blood cell）もしくは白血球（leucocyte）、内皮細胞、組織生検、脳脊髄液、関節液、リンパ液、腹水、間質液もしくは細胞外液、細胞間の空間の流体が挙げられ、歯肉溝滲出液、骨髓、胸水、脳脊髄液、唾液、粘液、喀痰、精液、汗、尿が挙げられる。試料は、好ましくは、体液、特に、血液およびその画分、ならびに尿である。試料はまた、腫瘍から流れ出た核酸、たとえば、循環腫瘍DNA（ctDNA）も含み得る。核酸としては、DNAおよびRNAを挙げることができ、二本鎖および一本鎖の形態であり得る。試料は、対象からもともと単離した形態であってもよく、または成分、たとえば、細胞を除去もしくは追加するため、1つの成分を別の成分と比べて濃縮するため、もしくは1つの形態の核酸を別のものに、たとえば、RNAからDNAに、もしくは一本鎖核酸から二本鎖に変換するために、さらなるプロセッシングに供されていてもよい。したがって、たとえば、分析のための体液は、無細胞核酸、たとえば、無細胞DNA（cfDNA）を含む、血漿または血清である。

【0129】

血漿の体積は、シーケンシングされる領域の所望されるリード深度に依存し得る。例示的な体積は、0.4~40ml、5~20ml、および10~20mlである。たとえば、体積は、0.5ml、1ml、5ml、10ml、20ml、30ml、または40ml

10

20

30

40

50

しであり得る。サンプリングされる血漿の体積は、5～20mlであり得る。

【0130】

試料は、ゲノム等価物を含む様々な量の核酸を含み得る。たとえば、約30ngのDNAの試料は、約10,000(10⁴)個のハプロイドヒトゲノム等価物を含み得、cfDNAの場合には、約2000億(2×10¹¹)個の個々のポリヌクレオチド分子を含み得る。同様に、約100ngのDNAの試料は、約30,000個のハプロイドヒトゲノム等価物を含み得、cfDNAの場合には、約6000億個の個々の分子を含み得る。

【0131】

試料は、異なる源、たとえば、無細胞または外来物からの核酸を含み得る。試料は、突然変異を有する核酸を含み得る。たとえば、試料は、生殖細胞系突然変異および／または体細胞系突然変異を有するDNAを含み得る。試料は、がん関連突然変異(たとえば、がん関連体細胞系突然変異)を有するDNAを含み得る。

【0132】

増幅前の試料中の無細胞核酸の例示的な量は、約1フェムトグラム(fg)～約1マイクログラム(ug)、たとえば、1ピコグラム(pg)～200ナノグラム(ng)、1ng～100ng、10ng～1000ngの範囲である。たとえば、量は、最大約600ng、最大約500ng、最大約400ng、最大約300ng、最大約200ng、最大約100ng、最大約50ng、または最大約20ngの無細胞核酸分子であり得る。量は、少なくとも1fg、少なくとも10fg、少なくとも100fg、少なくとも1pg、少なくとも10pg、少なくとも100pg、少なくとも1ng、少なくとも10ng、少なくとも100ng、少なくとも150ng、または少なくとも200ngの無細胞核酸分子であり得る。量は、最大1フェムトグラム(fg)、10fg、100fg、1ピコグラム(pg)、10pg、100pg、1ng、10ng、100ng、150ng、または200ngの無細胞核酸分子であり得る。本方法は、1フェムトグラム(fg)～200ngを得ることを含み得る。

【0133】

ある特定の実施形態では、試料中の無細胞核酸の量は、約5ng～300ngである。

【0134】

無細胞核酸は、約100～500ヌクレオチドの例示的なサイズ分布を有し、110～約230ヌクレオチドの分子が、分子の約90%に相当し、最頻値は約168ヌクレオチドであり、第2のマイナーピークは、240～440ヌクレオチドの範囲にある。無細胞核酸は、約160～約180ヌクレオチド、または約320～約360ヌクレオチド、または約440～約480ヌクレオチドであり得る。

【0135】

無細胞核酸は、溶液中に見出される無細胞核酸を、インタクトな細胞および体液の他の不溶性成分から分離する、分割ステップによって、体液から単離することができる。分割には、遠心分離または濾過などの技法が含まれ得る。あるいは、体液中の細胞を、溶解し、無細胞核酸および細胞核酸を、一緒にプロセッシングしてもよい。一般に、緩衝液の添加および洗浄ステップの後に、無細胞核酸は、アルコールで沈降させることができる。混入物質または塩を除去するために、シリカベスカラムなどのさらなる洗浄ステップを使用してもよい。手順のある特定の態様、たとえば、収率を最適化するために、たとえば、非特異的バルク担体核酸を、反応全体にわたり添加してもよい。そのようなプロセッシングの後、試料は、二本鎖DNA、一本鎖DNA、および一本鎖RNAを含む、様々な形態の核酸を含み得る。必要に応じて、一本鎖DNAおよびRNAは、後続のプロセッシングおよび分析ステップに含められるように、二本鎖形態に変換してもよい。

B. タグ

【0136】

試料インデックスおよび／または分子バーコードを提供するタグは、他の方法の中でもとりわけ、化学合成、ライゲーション、オーバーラップ伸長PCRによって、アダプターに組み込まれ得るか、またはそれ以外では結合され得る。一般に、反応における固有また

10

20

30

40

50

は非固有の分子バーコードの割当ては、米国特許出願第20010053519号、同第20110160078号、ならびに米国特許第6,582,908号、同第7,537,898号、および同第9,598,731号に記載されている方法およびシステムに従う。

【0137】

タグは、ランダムまたは非ランダムで、試料核酸に連結され得る。一部の場合には、それらは、予測された比で導入される。バーコード集団は、固有であってもよく、たとえば、すべてのバーコードが、同じヌクレオチド配列を有する。バーコード集団は、非固有であってもよく、たとえば、バーコードのうちの一部が、同じヌクレオチド配列を有し、バーコードのうちの一部が、異なるヌクレオチド配列を有する。たとえば、ゲノム試料当たり1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、20個、50個、100個、500個、1000個、5000個、10000個、50,000個、100,000個、500,000個、1,000,000個、10,000,000個、50,000,000個、または1,000,000,000個を上回る識別子がロードされるように、識別子がロードされ得る。一部の場合には、ゲノム試料当たり2個、3個、4個、5個、6個、7個、8個、9個、10個、20個、50個、100個、500個、1000個、5000個、10000個、50,000個、100,000個、500,000個、1,000,000個、10,000,000個、50,000,000個、または1,000,000,000個を下回る識別子がロードされるように、識別子がロードされ得る。一部の場合には、試料ゲノム当たりのロードされる識別子の平均数は、ゲノム試料当たり約1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、20個、50個、100個、500個、1000個、5000個、10000個、50,000個、100,000個、500,000個、1,000,000個、10,000,000個、50,000,000個、または1,000,000,000個の識別子を下回るか、または上回る。

【0138】

好ましい形式は、標的分子の両方の末端にライゲーションされる20～50個の異なるタグを使用し、20～50×20～50個のタグが作製される。そのような数のタグは、同じ開始点および終止点を有する異なる分子が、異なる組合せのタグを受容する高い確率（たとえば、少なくとも94%、99.5%、99.99%、99.999%）を有するものに十分である。

【0139】

一部の場合には、識別子は、所定のまたはランダムなまたはセミランダムな配列のオリゴヌクレオチドであり得る。他の場合には、バーコードが複数の中で必ずしも互いに固有とならないような、複数のバーコードが使用されてもよい。この例では、バーコードは、バーコードおよびそれが結合され得る配列の組合せが、個別にトラッキングすることができる固有の配列をもたらすように、（たとえば、ライゲーションまたはPCR増幅によって）個々の分子に結合され得る。本明細書に記載されるように、参照配列またはゲノムにマッピングされる配列リードの最初の（開始）点および終わりの（終止）点の配列データと組み合わせた、非固有の分子バーコードの検出により、特定の分子に固有の同一性を割り当てることが可能となり得る。個々の配列リードの長さまたは塩基対の数もまた、そのような分子に固有の同一性を割り当てするために使用することができる。本明細書に記載されるように、固有の同一性が割り当てられた核酸の一本鎖からの断片は、それによって、後続の親鎖からの断片および／または相補鎖の特定を可能にし得る。

【0140】

核酸分子に分子バーコードおよび／または試料インデックスを導入するために、従来の核酸増幅方法を使用して、1回または複数回の増幅を適用することができる。増幅は、1つまたは複数の反応混合物において行うことができる。分子バーコードおよび試料インデックスは、同時、または任意の逐次的順序で導入することができる。分子バーコードおよび試料インデックスは、配列捕捉（たとえば、濃縮）の前および／または後に導入する

ことができる。一部の実施形態では、分子タグのみが、プローブ捕捉の前に導入され、一方で試料インデックス／タグは、配列捕捉の後に導入される。一部の場合には、分子バーコードおよび試料インデックスの両方が、プローブ捕捉の前に導入される。一部の場合には、試料インデックスは、配列捕捉の後に導入される。通常、配列捕捉は、標的化配列に相補的な一本鎖核酸分子を導入することを伴う。典型的に、増幅により、200ヌクレオチド (nt) ~ 700 nt、250 nt ~ 350 nt、または320 nt ~ 550 ntの範囲のサイズで、分子バーコードおよび試料インデックスを有する非固有または固有にタグ付けされた複数の核酸アンプリコンが生成される。一部の実施形態では、アンプリコンは、約300 ntのサイズを有する。一部の実施形態では、アンプリコンは、約500 ntのサイズを有する。

10

C. 増幅

【0141】

アダプターが隣接する試料核酸は、PCR、および典型的にはプライマーが増幅させようとする核酸分子に隣接するアダプターのプライマー結合部位に結合することによりプライミングされる他の増幅方法によって、増幅させることができる。増幅方法は、サーモサイクリングの結果としての伸長、変性、およびアニーリングのサイクルを含み得るか、または転写に媒介される増幅にあるような等温のものであってもよい。他の増幅方法としては、リガーゼ連鎖反応、鎖置換増幅、核酸配列に基づく増幅、および自己持続性配列に基づく複製が挙げられる。

D. 濃縮

20

【0142】

配列は、シーケンシングの前に濃縮させることができる。濃縮は、特定の標的領域に実施してもよく、または非特異的に実施してもよい（「標的配列」）。一部の実施形態では、目的とされる標的化領域は、差次的タイリングおよび捕捉スキームを使用して、1つまたは複数のベイトセットパネルに選択される捕捉プローブ（「ベイト」）を用いて濃縮され得る。差次的タイリングおよび捕捉スキームは、異なる相対濃度のベイトセットを使用して、ベイトと関連するゲノム領域全体に差次的にタイリングし（たとえば、異なる「分解能」で）、拘束セット（たとえば、シーケンサー拘束、たとえば、シーケンシングロード、それぞれのベイトの有用性など）に供され、下流シーケンシングに所望されるレベルで、それらを捕捉する。一部の実施形態では、1つまたは複数の目的とされる領域に対するプローブを有するビオチン標識化ビーズを使用して、標的配列を捕捉し、必要に応じて、続いて、これらの領域を増幅させ、目的の領域を濃縮することができる。

30

【0143】

配列捕捉は、典型的に、標的配列にハイブリダイズするオリゴヌクレオチドプローブの使用を伴う。プローブセット戦略は、目的の領域全体にプローブをタイリングすることを含み得る。そのようなプローブは、たとえば、約60~120塩基の長さであり得る。セットは、約2x、3x、4x、5x、6x、8x、9x、10x、15x、20x、50x、またはそれを上回る深度を有し得る。配列捕捉の有効性は、部分的に、プローブの配列に相補的である（またはほぼ相補的である）標的分子内の配列の長さに依存する。

E. シーケンシング

40

【0144】

前の増幅の有無にかかわらず、アダプターが隣接している試料核酸は、シーケンシングに供することができる。シーケンシング方法としては、たとえば、サンガーシーケンシング、高スループットシーケンシング、ピロシーケンシング、合成によるシーケンシング、単一分子シーケンシング、ナノポアシーケンシング、半導体シーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシング、RNA-Seq (Illumina)、Digital Gene Expression (Helicos)、次世代シーケンシング、合成による単一分子シーケンシング (SMSS) (Helicos)、大規模並列シーケンシング、クローン単一分子アレイ (Solexa)、ショットガンシーケンシング、Ion Torrent、Oxford Nanopore

50

e、Roche Genia、Maxim-Gilbertシーケンシング、プライマーウォーキング、PacBio、SOLiD、Ion Torrent、またはナノポアのプラットフォームを使用したシーケンシングが挙げられる。シーケンシング反応は、様々な試料プロセッシングユニットにおいて行うことができ、これは、複数レーン、複数チャネル、複数ウェル、または複数の試料セットを実質的に同時にプロセッシングする他の手段であり得る。試料プロセッシングユニットにはまた、複数の実行を同時にプロセッシングすることを可能にする複数の試料チャンバが含まれ得る。

【0145】

シーケンシング反応は、他の疾患のがんのマーカーを含むことが公知の1つまたは複数の断片型に行うことができる。シーケンシング反応はまた、試料中に存在する任意の核酸断片に行うこともできる。シーケンシング反応は、少なくとも5%、10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90%、95%、99%、99.9%、または100%のゲノムの配列カバレッジを提供し得る。他の場合には、ゲノムの配列カバレッジは、5%、10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90%、95%、99%、99.9%、または100%を下回り得る。

【0146】

多重シーケンシングを使用して、同時のシーケンシング反応を行ってもよい。一部の場合には、無細胞ポリヌクレオチドは、少なくとも1000回、2000回、3000回、4000回、5000回、6000回、7000回、8000回、9000回、10000回、50000回、または100,000回のシーケンシング反応でシーケンシングされ得る。他の場合では、無細胞ポリヌクレオチドは、1000回、2000回、3000回、4000回、5000回、6000回、7000回、8000回、9000回、10000回、50000回、または100,000回を下回るシーケンシング反応でシーケンシングされ得る。シーケンシング反応は、逐次的に行われてもよく、または同時に行われてもよい。後続のデータ分析は、シーケンシング反応のすべてまたは一部に行われ得る。一部の場合には、データ分析は、少なくとも1000回、2000回、3000回、4000回、5000回、6000回、7000回、8000回、9000回、10000回、50000回、または100,000回のシーケンシング反応に行われ得る。他の場合には、データ分析は、1000回、2000回、3000回、4000回、5000回、6000回、7000回、8000回、9000回、10000回、50000回、または100,000回を下回るシーケンシング反応に行われ得る。例示的なリード深度は、遺伝子座（塩基）1つ当たり1000～50000リードである。

F. 分析

【0147】

本開示の実施形態によるシーケンシングにより、複数の配列リードが生成される。本開示による配列リードは、一般に、約150塩基を下回る長さ、または約90塩基を下回る長さのヌクレオチドデータの配列を含む。ある特定の実施形態では、リードは、約80～約90塩基、たとえば、約85塩基の長さである。一部の実施形態では、本開示の方法は、非常に短いリード、すなわち、約50塩基または約30塩基を下回る長さのリードに、適用される。配列リードデータには、配列データならびにメタ情報が含まれ得る。配列リードデータは、当業者には公知のように、たとえば、VCFファイル、FASTAファイル、またはFASTQファイルを含む、任意の好適なファイル形式で記憶することができる。

【0148】

FASTAは、もともと、配列データベースを検索するためのコンピュータプログラムであり、FASTAという名称は、標準ファイル形式も指すようになっている。Pearson & Lipman, 1988, Improved tools for biological sequence comparison, PNAS 85:2444-2448を参照されたい。FASTA形式の配列は、単一行の説明で開始され、複数行の配列データが続く。説明の行は、1列目における大なり（「>」）の記号に

よって、配列データと区別される。「>」記号に続く言葉は、配列の識別子であり、その行の残りが、説明である（いずれも必要に応じたものである）。「>」と識別子の1文字目との間にスペースは入れないものとする。すべて行のテキストは、80文字よりも短くすることが推奨される。「>」で開始する別の行が出現すると配列は終了し、これは、別の配列の開始を示す。

【0149】

FASTQ形式は、生物学的配列（通常、ヌクレオチド配列）およびその対応する品質スコアの両方を記憶するためのテキストに基づく形式である。これは、FASTA形式に類似であるが、配列データに続いて品質スコアを有する。配列の文字および品質スコアのいずれも、簡潔さのために単一のASCII文字でコードされる。FASTQ形式は、高スループットのシーケンシング機器、たとえば、Illumina Genome Analyzerの出力を記憶するための事実上の標準物である。Cock et al., 2009, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res 38(6):1767-1771.

【0150】

FASTAおよびFASTQのファイルに関して、メタ情報は、説明の行を含み、配列データの行は含まない。一部の実施形態では、FASTQファイルについては、メタ情報は、品質スコアを含む。FASTAおよびFASTQのファイルに関して、配列データは、説明の行の後に開始され、典型的に、IUPACの多義コードのあるサブセットを使用し、必要に応じて「-」を用いて表す。好ましい実施形態では、配列データは、A、T、C、G、およびNの文字を使用し、必要に応じて、「-」またはUを必要な場合には（たとえば、ギャップまたはウラシルを表すために）含む。

【0151】

一部の実施形態では、少なくとも1つのマスター配列リードファイルおよび出力ファイルは、プレーンテキストファイルとして（たとえば、ASCII、ISO/IEC 646、EBCDIC、UTF-8、またはUTF-16などのコーディングを使用して）記憶される。本発明によって提供されるコンピュータシステムは、プレーンテキストファイルを開くことができるテキストエディタプログラムを含み得る。テキストエディタプログラムは、コンピュータスクリーン上にテキストファイル（たとえば、プレーンテキストファイル）の内容を提示することができ、ヒトによるテキストの編集が可能である（たとえば、モニター、キーボード、およびマウスを使用して）、コンピュータプログラムを指し得る。例示的なテキストエディタとしては、限定することなく、Microsoft Word、emacs、pico、vi、BBEdit、およびTextWranglerが挙げられる。好ましくは、テキストエディタプログラムは、コンピュータスクリーン上にプレーンテキストファイルを表示し、ヒト可読形式で（たとえば、バイナリでコードされるのではなく、代わりに、ヒトの活字で使用されるような英数字を使用して）メタ情報および配列リードを示すことができる。

【0152】

方法は、FASTAまたはFASTQファイルに関連して考察されているが、本開示の方法およびシステムを使用して、たとえば、Variant Call Format (VCF) 形式のファイルを含む、任意の好適な配列ファイル形式を圧縮してもよい。典型的なVCFファイルは、ヘッダー部分およびデータ部分を含む。ヘッダーは、それぞれが「##」という文字で開始される任意の数のメタ情報の行、ならびに単一の「#」文字で開始されるタブ区切りのフィールド定義の行を含む。フィールド定義の行には、8つの必須列が示され、本体部分には、フィールド定義の行によって定義される列を埋めるデータの行が含まれる。VCF形式は、Danecek et al., 2011, The variant call format and VCFtools, Bioinformatics 27(15):2156-2158に記載されている。ヘッダー部分は、圧縮ファイルに書き込むためのメタ情報として扱うことができ、データ部分は、行として扱うことができ、これらのそれぞれは、固有である場合にのみマスターファイルに記憶されることになる。

【0153】

本開示のある特定の実施形態は、配列リードのアセンブリを提供する。アライメントごとのアセンブリの場合、たとえば、リードは、互いに対してまたは参照に対してアライメントされる。それぞれのリードを、次に参照ゲノムに対してアライメントすることにより、リードのすべてが、互いに関連して位置付けられて、アセンブリが作成される。加えて、配列リードの参照配列に対するアライメントまたはマッピングは、配列リード内のバリエーション配列を識別するために使用することもできる。バリエーション配列の特定は、本明細書に記載される方法およびシステムと組み合わせて、疾患もしくは状態の診断もしくは予後診断、または処置決定を誘導するのをさらに補助するために使用することができる。

【0154】

一部の実施形態では、ステップのうちのいずれかまたはすべてが、自動化される。あるいは、本発明の方法は、たとえば、それぞれが、必要に応じて、C++などのコンパイル言語で記述され、次いでバイナリとしてコンパイルされ配布される、1つまたは複数の専用プログラムにおいて全体的または部分的に具現化され得る。本発明の方法は、全体的もしくは部分的に既存の配列分析プラットフォーム内のモジュールとして、またはプラットフォーム内の機能性を呼び出すことによって、実装されてもよい。ある特定の実施形態では、本発明の方法は、いずれも単一の開始キュー（たとえば、ヒトの活動、別のコンピュータプログラム、または機械を起源とするトリガー事象のうちの1つまたは組合せ）に回答して自動的に呼び出される、いくつかのステップを含む。したがって、本発明は、ステップのうちのいずれかまたはステップの任意の組合せが、キューに回答して自動的に発生し得る方法を提供する。自動的にとは、一般に、ヒトによる入力、影響、または相互作用が介在しない（すなわち、もともとのまたはキュー前のヒトの活動にのみ応答する）ことを意味する。

【0155】

システムは、対象核酸の正確かつ高感度の解釈を含む、様々な形式の出力も包含する。検索 (retrieval) の出力は、コンピュータファイルの形式で提供され得る。ある特定の実施形態では、出力は、FASTAファイル、FASTQファイル、またはVCFファイルである。出力は、テキストファイル、または参照ゲノムの配列に対してアライメントされた核酸の配列などの配列データを含むXMLファイルが得られるように処理され得る。他の実施形態では、処理により、参照ゲノムと比べた対象核酸における1つまたは複数の突然変異を示す座標またはストリングを含む出力が得られる。当該技術分野において公知のアライメントストリングとしては、Simple UnGapped Alignment Report (SUGAR)、Verbose Useful Labeled Gapped Alignment Report (VULGAR)、およびCompact Idiosyncratic Gapped Alignment Report (CIGAR) (Ning, Z., et al., Genome Research 11(10):1725-9 (2001)) が挙げられる。これらのストリングは、たとえば、European Bioinformatics Institute (Hinxton, UK) からのExonerate配列アライメントソフトウェアにおいて実装される。

【0156】

一部の実施形態では、たとえば、CIGARストリングを含む、配列アライメントマップ (SAM) またはバイナリアライメントマップ (BAM) ファイルなどの配列アライメントが得られる (SAM形式は、たとえば、Li, et al., The Sequence Alignment/Map format and SAMtools, Bioinformatics, 2009, 25(16):2078-9に記載されている)。一部の実施形態では、CIGARは、1行ごとにギャップ付きアライメントを提示するかまたはそれを含む。CIGARは、CIGARストリングとして報告される圧縮されたペアでのアライメント形式である。CIGARストリングは、長い (たとえば、ゲノムの) ペアでのアライメントを表すのに有用である。CIGARストリングは、参照ゲノム配列に対するリードのアライメントを表すためにSAM形式で使用される。

【0157】

C I G A Rストリングは、確立されたモチーフに従う。それぞれの文字には、事象の塩基数を示す数が先行している。使用される文字としては、M、I、D、N、およびSを挙げることができる（M＝マッチ、I＝挿入、D＝欠失、N＝ギャップ、S＝置換）。C I G A Rストリングは、マッチ／ミスマッチおよび欠失（またはギャップ）の配列を定義する。たとえば、C I G A Rストリング2 M D 3 M 2 D 2 Mは、アライメントが、2つのマッチ、1つの欠失（1という数は、スペースを節約するために省略される）、3つのマッチ、2つの欠失、および2つのマッチを含むことを意味する。

【0158】

本発明によって企図されるように、上述の機能は、ソフトウェア、ハードウェア、ファームウェア、ハードワイヤ（hardwiring）、またはこれらの任意の組合せを含むシステムを使用して実装され得る。機能を実装する特徴はまた、機能の部分が、異なる物理位置で実装されるように分配されることを含め、様々な位置に物理的に位置してもよい。

【0159】

システムは、サーバーコンピュータ、端末、シーケンサー、シーケンサーコンピュータ、コンピュータ、またはこれらの任意の組合せのうちの1つまたは複数を含み得る。それぞれのそのようなコンピュータデバイスは、ネットワークを介して通信し得る。シーケンサーは、必要に応じて、たとえば、それ専用のシーケンサーコンピュータ（任意の入力／出力機序（I／O）、プロセッサ、およびメモリを含む）を含むか、またはそれに作動可能に連結され得る。追加または代替として、シーケンサーは、ネットワークを介してサーバーまたはコンピュータ（たとえば、ラップトップ、デスクトップ、またはタブレット）に作動可能に連結されてもよい。コンピュータは、1つまたは複数のプロセッサ、メモリ、およびI／Oを含む。本発明の方法が、クライアント／サーバーアーキテクチャを利用する場合、本発明の方法の任意のステップは、サーバーを使用して行われてもよく、このサーバーには、プロセッサ、メモリ、およびI／Oのうちの1つまたは複数が含まれ、データ、命令などを取得すること、またはインターフェースモジュールを介して結果を提供すること、または結果をファイルとして提供することができる。サーバーは、コンピュータもしくは端末を通じてネットワークにつながっていてもよく、またはサーバーは、端末に直接接続されてもよい。端末は、好ましくは、コンピュータデバイスである。本発明によるコンピュータは、好ましくは、I／O機序およびメモリに連結された1つまたは複数のプロセッサを含む。

【0160】

プロセッサは、たとえば、単一コアまたは複数コアのプロセッサのうちの1つまたは複数を含む、1つまたは複数のプロセッサによって提供され得る。I／O機序としては、ビデオディスプレイユニット（たとえば、液晶ディスプレイ（LCD）もしくは陰極線管（CRT））、英数字入力デバイス（たとえば、キーボード）、カーソルコントロールデバイス（たとえば、マウス）、ディスクドライブユニット、シグナル生成デバイス（たとえば、スピーカー）、加速度計、マイクロホン、セルラー無線周波数アンテナ、およびネットワークインターフェースデバイス（たとえば、ネットワークインターフェースカード（NIC）、Wi-Fiカード、セルラーモデム、データジャック、イーサネットポート、モデムジャック、HDMI（登録商標）ポート、ミニHDMI（登録商標）ポート、USBポート）、タッチスクリーン（たとえば、CRT、LCD、LED、AMOLED、スーパーAMOLED）、ポインティングデバイス、トラックパッド、光（たとえば、LED）、光／画像投射デバイス、またはこれらの組合せを挙げることができる。本発明によるメモリは、好ましくは、本明細書に記載される手法または機能のうちのいずれか1つまたは複数を実現化する1つまたは複数の命令セットが記憶された1つまたは複数の機械可読媒体（たとえば、ソフトウェア）を含む、1つまたは複数の有形デバイスによって提供される、非一過的メモリを指す。ソフトウェアはまた、完全にまたは少なくとも部分的に、システム内のコンピュータによる実行時に主メモリ内、プロセッサ内、またはその両方に存在してもよく、主メモリおよびプロセッサもまた、機械可読媒体を構築している。ソフトウェアは、さらに、ネットワークインターフェースデバイスを介してネ

ットワーク上で送信または受信されてもよい。

【0161】

機械可読媒体は、例示的な実施形態では、単一の媒体であり得るが、「機械可読媒体」という用語は、1つまたは複数の命令セットを記憶する単一の媒体または複数の媒体（たとえば、集中型もしくは分散型データベース、ならびに／または関連するキャッシュおよびサーバー）を含んで解釈されるべきである。「機械可読媒体」という用語はまた、機械によって実行するための命令セットを記憶、コード、または保有することができ、機械に本発明の手法のうちのいずれか1つまたは複数を実行させる、任意の媒体を含んで解釈されるべきである。メモリは、たとえば、ハードディスクドライブ、ソリッドステートドライブ（SSD）、光ディスク、フラッシュメモリ、ジップディスク、テープドライブ、「クラウド」記憶位置、またはこれらの組合せのうちの1つまたは複数であり得る。ある特定の実施形態では、本発明のデバイスは、メモリのための有形の非一過的コンピュータ可読媒体を含む。メモリとして使用するための例示的なデバイスとしては、半導体メモリデバイス（たとえば、EPROM、EEPROM、ソリッドステートドライブ（SSD）、およびフラッシュメモリデバイス、たとえば、SD、マイクロSD、SDXC、SDIO、SDHCカード）、磁気ディスク（たとえば、内蔵ハードディスクまたは取外し可能なディスク）、ならびに光学ディスク（たとえば、CDおよびDVDディスク）が挙げられる。

10

【0162】

一部の実施形態では、本明細書に開示されるシステムおよび方法の結果は、インプットとして使用され、レポートが生成される。レポートは、紙形式であってもよい。たとえば、レポートは、本明細書に開示される方法およびシステムによって識別される、フィルタリングされた配列情報からのデータを含み得る。そのようなデータは、たとえば、識別された配列情報を踏まえた診断情報または治療法の推奨を含み得る。一部の実施形態では、レポートは、本明細書に開示される方法およびシステムによって識別される、1つまたは複数の真の遺伝子バリエーションなどの情報を含み得る。

20

【0163】

本明細書に開示される方法の様々なステップ、または本明細書に開示されるシステムによって実行されるステップは、同じかもしくは異なる時点で、同じかもしくは異なる地理的位置、たとえば、国において、および／または同じかもしくは異なる人々によって実行され得る。

30

III. 例示的な適用

A. シーケンシングパネル

【0164】

突然変異を示す腫瘍を検出する可能性を向上させるために、シーケンシングされるDNAの領域は、遺伝子またはゲノム領域のパネルを含んでもよい。限定的なシーケンシングの領域（たとえば、限定パネル）を選択することにより、必要とされる総シーケンシング（たとえば、シーケンシングされるヌクレオチドの総量を低減することができる。シーケンシングパネルは、単一のがん、がんのセット、またはすべてのがんを検出するために、複数の異なる遺伝子または領域を標的とし得る。あるいは、DNAは、シーケンシングパネルを使用せずに、全ゲノムシーケンシング（WGS）または他の非バイアス型シーケンシング法によってシーケンシングしてもよい。

40

【0165】

一部の態様では、複数の異なる遺伝子またはゲノム領域を標的とするパネルは、がんを有する対象のうちの決定されたパーセンテージが、パネル内の1つまたは複数の異なる遺伝子において遺伝子バリエーションまたは腫瘍マーカーを示すように、選択される。パネルは、シーケンシングの領域を、固定の塩基対数に限定するように選択することができる。パネルは、所望される量のDNAをシーケンシングするように選択することができる。パネルは、さらに、所望される配列リード深度を達成するように選択することができる。パネルは、シーケンシングされる塩基対の量に所望される配列リード深度または配列リードカ

50

バレッジを達成するように選択することができる。パネルは、試料中の1つまたは複数の遺伝子バリエーションを検出するための理論上の感度、理論上の特異度、および／または理論上の精度を達成するように選択することができる。

【0166】

領域のパネルと検出するためのプローブとしては、目的のゲノム領域（ホットスポット領域）を検出するためのもの、ならびにヌクレオソームを意識した（nucleosome-aware）プローブ（たとえば、K R A S コドン12および13）を挙げることができる、c f D N A カバレッジならびにヌクレオソーム結合パターンおよびG C 配列組成によって影響を受ける断片サイズ変動の分析に基づいて、捕捉を最適化するように設計され得る。本明細書で使用される領域としては、ヌクレオソーム位置およびG C モデルに基づいて最適化された非ホットスポット領域も含まれ得る。パネルは、起源組織（たとえば、公開文献を使用して、組織全体でもっとも多様な転写プロファイルを有する遺伝子を示す、50～100個のバイトを定義する（プロモーターとは限らない）、全ゲノムスキファールド（たとえば、コピー数ベースライニング目的で、超保存的ゲノムコンテンツを識別し、一握りのプローブを染色体全体に分散してタイリングするため）、転写開始部位（T S S）／C p G アイランド（たとえば、たとえば、腫瘍抑制遺伝子（たとえば、結腸直腸がんにおけるS E P T 9／V I M）のプロモーターにおける差示的にメチル化された領域（たとえば、差示的メチル化領域（D M R））を捕捉するため）を識別するための部分パネルを含む、複数の部分パネルを含み得る。一部の実施形態では、起源組織のマーカーは、組織特異的エピジェネティックマーカーである。

【0167】

目的のゲノム位置の一覧の一部の例は、表1および表2に見出すことができる。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表1の遺伝子のうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、少なくとも70個、少なくとも75個、少なくとも80個、少なくとも85個、少なくとも90個、少なくとも95個、または97個の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表1のS N Vのうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、または70個を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表1のC N Vのうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくとも16個、少なくとも17個、または18個を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表1の融合のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、または6個を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表1のインデルのうちの少なくとも1個、少なくとも2個、または3個の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表2の遺伝子のうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、少なくとも70個、少なくとも75個、少なくとも80個、少なくとも85個、少なくとも90個、少なくとも95個、少なくとも100個、少なくとも105個、少なくとも110個、または115個の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表2のS N Vのうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくと

も35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、少なくとも70個、または73個を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表2のCNVのうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくとも16個、少なくとも17個、または18個を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表2の融合のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、または6個を含む。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表2のインデルのうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくとも16個、少なくとも17個、または18個の少なくとも一部分を含む。これらの目的とされるゲノム位置のそれぞれは、所与のバインドパネルの骨格領域またはホットスポット領域として識別され得る。目的とされるホットスポットゲノム位置の一覧の例は、表3に見出すことができる。一部の実施形態では、本開示の方法において使用されるゲノム位置は、表3の遺伝子のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくとも16個、少なくとも17個、少なくとも18個、少なくとも19個、または少なくとも20個の少なくとも一部分を含む。それぞれのホットスポットゲノム位置は、関連する遺伝子、それが存在する染色体、遺伝子座を表すゲノムの開始位置および終止位置、塩基対での遺伝子座の長さ、遺伝子によってカバーされるエクソン、および目的とされる所与のゲノム位置が捕捉しようとし得る重要な特性（たとえば、突然変異の種類）を含む、いくつかの特徴とともに列挙されている。

【表1】

表 1

点突然変異(SNV) 増幅						増幅 (CNVs)		融合	インデル
AKT1	ALK	APC	AR	ARAF	ARID1A	AR	BRAF	ALK	EGFR (エクソン 19 & 20)
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	CCND1	CCND2	FGFR2	
CCNE1	CDH1	CDK4	CDK6	CDKN2A	CDKN2B	CCNE1	CDK4	FGFR3	
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	CDK6	EGFR	NTRK1	ERBB2
FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	ERBB2	FGFR1	RET	RET (エクソン 19 & 20)
GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	FGFR2	KIT	ROS1	
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	KRAS	MET		
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	MYC	PDGFRA		MET (エクソン14 スキッピング)
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	PIK3CA	RAF1		
PTPN11	RAF1	RB1	RET	RHEB	RHOA				
RIT1	ROS1	SMAD4	SMO	SRC	STK11				
TERT	TP53	TSC1	VHL						

10

20

30

40

50

【表 2－1】

表 2

点突然変異(SNV)増幅						増幅 (CNVs)		融合	インデル					
AKT1	ALK	APC	AR	ARAF	ARID1A	AR	BRAF	ALK	EGFR					
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	CCND1	CCND2	FGFR2	(エクソン					
CCNE1	CDH1	CDK4	CDK6	CDKN2A	DDR2	CCNE1	CDK4	FGFR3	19 & 20)					
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	CDK6	EGFR	NTRK1	ERBB2					
FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	ERBB2	FGFR1	RET	(エクソン					
GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	FGFR2	KIT	ROS1	19 & 20)					
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	KRAS	MET		MET					
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	MYC	PDGFRA		(エクソン14					
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	PIK3CA	RAF1		スキッピング)					
PTPN11	RAF1	RB1	RET	RHEB	RHOA									
RIT1	ROS1	SMAD4	SMO	MAPK1	STK11									
TERT	TP53	TSC1	VHL	MAPK3	MTOR									
NTRK3														

【表 2－2】

				GATA3
				KIT
				MLH1
				MTOR
				NF1
				PDGFRA
				PTEN
				RB1
				SMAD4
				STK11
				TP53
				TSC1
				VHL

【表 3 - 1】

表 3

遺伝子	染色体	開始位置	終止位置	長さ(bp)	カバーされる エクソン	重要な特性
ALK	chr2	29446405	29446655	250	イントロン 19	融合
ALK	chr2	29446062	29446197	135	イントロン 20	融合
ALK	chr2	29446198	29446404	206	20	融合
ALK	chr2	29447353	29447473	120	イントロン 19	融合
ALK	chr2	29447614	29448316	702	イントロン 19	融合
ALK	chr2	29448317	29448441	124	19	融合
ALK	chr2	29449366	29449777	411	イントロン 18	融合
ALK	chr2	29449778	29449950	172	18	融合
BRAF	chr7	140453064	140453203	139	15	BRAF V600
CTNNB1	chr3	41266007	41266254	247	3	S37
EGFR	chr7	55240528	55240827	299	18 および 19	G719 および欠失
EGFR	chr7	55241603	55241746	143	20	挿入 /T790M
EGFR	chr7	55242404	55242523	119	21	L858R
ERBB2	chr17	37880952	37881174	222	20	挿入
ESR1	chr6	152419857	152420111	254	10	V534, P535, L536, Y537, D538
FGFR2	chr10	123279482	123279693	211	6	S252
GATA3	chr10	8111426	8111571	145	5	SS / インデル
GATA3	chr10	8115692	8116002	310	6	SS / インデル
GNAS	chr20	57484395	57484488	93	8	R844
IDH1	chr2	209113083	209113394	311	4	R132

【表 3 - 2】

IDH2	chr15	90631809	90631989	180	4	R140, R172
KIT	chr4	55524171	55524258	87	1	
KIT	chr4	55561667	55561957	290	2	
KIT	chr4	55564439	55564741	302	3	
KIT	chr4	55565785	55565942	157	4	
KIT	chr4	55569879	55570068	189	5	
KIT	chr4	55573253	55573463	210	6	
KIT	chr4	55575579	55575719	140	7	
KIT	chr4	55589739	55589874	135	8	
KIT	chr4	55592012	55592226	214	9	
KIT	chr4	55593373	55593718	345	10および11	557, 559, 560, 576
KIT	chr4	55593978	55594297	319	12および13	V654
KIT	chr4	55595490	55595661	171	14	T670, S709
KIT	chr4	55597483	55597595	112	15	D716
KIT	chr4	55598026	55598174	148	16	L783
KIT	chr4	55599225	55599368	143	17	C809, R815, D816, L818, D820, S821F, N822, Y823
KIT	chr4	55602653	55602785	132	18	A829P
KIT	chr4	55602876	55602996	120	19	
KIT	chr4	55603330	55603456	126	20	
KIT	chr4	55604584	55604733	149	21	
KRAS	chr12	25378537	25378717	180	4	A146
KRAS	chr12	25380157	25380356	199	3	Q61
KRAS	chr12	25398197	25398328	131	2	G12/G13
MET	chr7	116411535	116412255	720	13, 14, イントロン 13, イントロン 14	MET エクソン 14 SS
NRAS	chr1	115256410	115256609	199	3	Q61
NRAS	chr1	115258660	115258791	131	2	G12/G13
PIK3CA	chr3	178935987	178936132	145	10	E545K
PIK3CA	chr3	178951871	178952162	291	21	H1047R
PTEN	chr10	89692759	89693018	259	5	R130
SMAD4	chr18	48604616	48604849	233	12	D537

10

20

30

40

50

【表 3 - 3】

TERT	chr5	1294841	1295512	671	プロモーター	chr5:1295228
TP53	chr17	7573916	7574043	127	11	Q331, R337, R342
TP53	chr17	7577008	7577165	157	8	R273
TP53	chr17	7577488	7577618	130	7	R248
TP53	chr17	7578127	7578299	172	6	R213/Y220
TP53	chr17	7578360	7578564	204	5	R175 / 欠失
TP53	chr17	7579301	7579600	299	4	
				12574 (合計標的領域)		
				16330 (合計プローブカバレッジ)		

10

【0168】

一部の実施形態では、パネル内の1つまたは複数の領域は、外科手術の後に残留しているがんを検出するための1つまたは複数の遺伝子からの1つまたは複数の遺伝子座を含む。この検出は、既存のがん検出方法で可能であるよりも早期であり得る。一部の実施形態では、パネル内の1つまたは複数のゲノム位置は、高リスク患者集団においてがんを検出するための1つまたは複数の遺伝子からの1つまたは複数の遺伝子座を含む。たとえば、喫煙者は、一般の集団よりもはるかに高い肺がん率を有する。さらに、喫煙者は、肺における不規則な小結節の発生など、がんの検出をより困難にする他の肺状態を発症し得る。一部の実施形態では、本明細書に記載される方法により、既存のがん検出方法で可能であるよりも早期に、高リスク患者におけるがんが検出される。

20

【0169】

ゲノム位置は、その遺伝子または領域に腫瘍マーカーを有する、がんを有する対象の数に基づいて、シーケンシングパネルへの包含に選択され得る。ゲノム位置は、その遺伝子に存在するがんおよび腫瘍マーカーを有する対象の有病率に基づいて、シーケンシングパネルへの包含に選択されてもよい。領域内の腫瘍マーカーの存在は、がんを有する対象を示し得る。

30

【0170】

一部の事例では、パネルは、1つまたは複数のデータベースからの情報を使用して選択され得る。がんに関する情報は、がん腫瘍生検またはcfDNAアッセイに由来してもよい。データベースは、シーケンシングした腫瘍試料の集団を説明する情報を含み得る。データベースは、腫瘍試料におけるmRNAの発現に関する情報を含み得る。データベースは、腫瘍試料における制御エレメントまたはゲノム領域に関する情報を含み得る。シーケンシングした腫瘍試料に関連する情報は、様々な遺伝子バリエーションの頻度を含み得、その遺伝子バリエーションが生じる遺伝子または領域を説明し得る。遺伝子バリエーションは、腫瘍マーカーであり得る。そのようなデータベースの非限定的な例は、COSMICである。COSMICは、様々ながんにおいて見出される体細胞系突然変異のカタログである。特定のがんに関して、COSMICは、突然変異の頻度に基づいて、遺伝子をランク付けしている。遺伝子は、所与の遺伝子内に高頻度の突然変異を有することにより、パネルへの包含に選択され得る。たとえば、COSMICは、シーケンシングした乳がん試料集団のうちの33%が、TP53に突然変異を有し、サンプリングした乳がん集団のうちの22%が、KRASに突然変異を有することを示す。APCを含め、他のランク付けされている遺伝子は、シーケンシングした乳がん試料集団のうちの約4%にしか見出されない突然変異を有する。TP53およびKRASは、サンプリングされた乳がんの中で、比較的高い

40

50

頻度（たとえば、約4%の頻度で生じるAPCと比較して）を有することに基づいて、シーケンシングパネルに包含され得る。COSMICは、非限定的な例として提供されているが、しかしながら、がんを遺伝子または遺伝子領域に位置する腫瘍マーカーと関連付ける任意のデータベースまたは情報セットを、使用してもよい。別の例では、COSMICによって提供されるように、1156個の胆道がん試料のうち、380個の試料（33%）が、TP53に突然変異を有していた。いくつかの他の遺伝子、たとえば、APCは、すべての試料のうちの4~8%において突然変異を有する。したがって、TP53は、胆道がん試料集団における比較的高い頻度に基づいて、パネルへの包含に選択され得る。

【0171】

遺伝子またはゲノム部分は、腫瘍マーカーの頻度が、サンプリングした腫瘍組織または循環腫瘍DNAにおいて、所与のバックグラウンド集団において見出されるものよりも有意に高い場合に、パネルに選択され得る。ゲノム位置の組合せは、がんを有する対象の少なくとも過半数が、パネル内のゲノム位置または遺伝子のうちの少なくとも1つに存在する腫瘍マーカーまたはゲノム領域を有し得るように、パネルの包含に選択され得る。ゲノム位置の組合せは、特定のがんまたはがんのセットについて、対象の過半数が、選択された領域のうちの1つまたは複数の1つまたは複数の腫瘍マーカーを有することを示すデータに基づいて、選択され得る。たとえば、がん1を検出するために、がん1を有する対象のうちの90%がパネルの領域A、B、C、および／またはDに腫瘍マーカーを有することを示すデータに基づいて、領域A、B、C、および／またはDを含むパネルを選択することができる。あるいは、腫瘍マーカーは、がんを有する対象において2つまたはそれを上回る領域に独立して生じることが示されてもよく、結果として、組合せで、2つまたはそれを上回る領域における腫瘍マーカーが、がんを有する対象集団の過半数に存在する。たとえば、がん2を検出するために、対象のうちの90%が1つまたは複数の領域に腫瘍マーカーを有し、そのような対象のうちの30%において、腫瘍マーカーは領域Xにのみ検出され、一方で腫瘍マーカーが検出された対象の残りについては、腫瘍マーカーが領域Yおよび／またはZにのみ検出されることを示すデータに基づいて、領域X、Y、およびZを含むパネルを選択することができる。1つまたは複数のがんに関連することがこれまでに示されている1つまたは複数のゲノム位置に存在する腫瘍マーカーは、腫瘍マーカーがその時点でそれらの領域のうちの1つまたは複数において50%またはそれを上回って検出される場合、がんを有する対象を示すかまたは予測することができる。1つまたは複数の領域内の腫瘍マーカーのセットのがん頻度を踏まえてがんを検出する条件付き確率を利用するモデルなど、コンピュータによるアプローチを使用して、どの領域が、単独または組合せで、がんを予測し得るかを予測することができる。パネル選択の他のアプローチは、大型パネルおよび／または全ゲノムシーケンシング（WGS、RNA-seq、ChIP-seq、重硫酸塩（bisulfate）シーケンシング、ATAC-seq、およびその他）による腫瘍の包括的ゲノムプロファイリングを利用した研究から得られた情報を記載しているデータベースの使用を伴う。文献から収集した情報により、ある特定のがんにおいて共通して影響を受け、突然変異している経路を説明することもできる。パネル選択は、遺伝子情報を説明するオントロジーの使用によってさらに詳しいものとなり得る。

【0172】

シーケンシングのためのパネルに含まれる遺伝子は、完全に転写される領域、プロモーター領域、エンハンサー領域、制御エレメント、および／または下流配列を含み得る。突然変異を示す腫瘍を検出する可能性をさらに増加させるために、エクソンのみを、パネルに含めてもよい。パネルは、選択された遺伝子のすべてのエクソンを含んでもよく、または選択された遺伝子のエクソンのうちの1つまたは複数のみを含んでもよい。パネルは、複数の異なる遺伝子のそれぞれからのエクソンから構成されてもよい。パネルは、複数の異なる遺伝子のそれぞれからの少なくとも1つのエクソンを含み得る。

【0173】

一部の態様では、複数の異なる遺伝子のそれぞれからのエクソンのパネルは、がんを有する対象のうちの決定されたパーセンテージが、エクソンのパネル内の少なくとも1つの

10

20

30

40

50

エクソンに遺伝子バリエーションを示すように、選択される。

【0174】

遺伝子のパネル内のそれぞれの異なる遺伝子からの少なくとも1つの完全なエクソンが、シーケンシングされ得る。シーケンシングされるパネルは、複数の遺伝子からのエクソンを含み得る。パネルは、2～100個の異なる遺伝子、2～70個の遺伝子、2～50個の遺伝子、2～30個の遺伝子、2～15個の遺伝子、または2～10個の遺伝子からのエクソンを含み得る。

【0175】

選択されるパネルは、様々な数のエクソンを含み得る。パネルは、2～3000個のエクソンを含み得る。パネルは、2～1000個のエクソンを含み得る。パネルは、2～500個のエクソンを含み得る。パネルは、2～100個のエクソンを含み得る。パネルは、2～50個のエクソンを含み得る。パネルは、300個以下のエクソンを含み得る。パネルは、200個以下のエクソンを含み得る。パネルは、100個以下のエクソンを含み得る。パネルは、50個以下のエクソンを含み得る。パネルは、40個以下のエクソンを含み得る。パネルは、30個以下のエクソンを含み得る。パネルは、25個以下のエクソンを含み得る。パネルは、20個以下のエクソンを含み得る。パネルは、15個以下のエクソンを含み得る。パネルは、10個以下のエクソンを含み得る。パネルは、9個以下のエクソンを含み得る。パネルは、8個以下のエクソンを含み得る。パネルは、7個以下のエクソンを含み得る。

【0176】

パネルは、複数の異なる遺伝子からの1つまたは複数のエクソンを含み得る。パネルは、あるパーセンテージの複数の異なる遺伝子のそれぞれからの1つまたは複数のエクソンを含み得る。パネルは、異なる遺伝子のうちの少なくとも25%、50%、75%、または90%のそれぞれからの少なくとも2つのエクソンを含み得る。パネルは、異なる遺伝子のうちの少なくとも25%、50%、75%、または90%のそれぞれからの少なくとも3つのエクソンを含み得る。パネルは、異なる遺伝子のうちの少なくとも25%、50%、75%、または90%のそれぞれからの少なくとも4つのエクソンを含み得る。

【0177】

シーケンシングパネルのサイズは、変動し得る。シーケンシングパネルは、たとえば、パネル内の特定の領域のシーケンシングされるヌクレオチドの総量またはシーケンシングされる固有の分子の数を含み、いくつかの因子に応じてより大きくまたはより小さく作製することができる（ヌクレオチドサイズに関して）。シーケンシングパネルは、5 kb～50 kbにサイズ決定され得る。シーケンシングパネルは、サイズが10 kb～30 kbであり得る。シーケンシングパネルは、サイズが12 kb～20 kbであり得る。シーケンシングパネルは、サイズが12 kb～60 kbであり得る。シーケンシングパネルは、サイズが、少なくとも10 kb、12 kb、15 kb、20 kb、25 kb、30 kb、35 kb、40 kb、45 kb、50 kb、60 kb、70 kb、80 kb、90 kb、100 kb、110 kb、120 kb、130 kb、140 kb、または150 kbであり得る。シーケンシングパネルは、サイズが100 kb、90 kb、80 kb、70 kb、60 kb、または50 kbを下回り得る。

【0178】

シーケンシングに選択されるパネルは、少なくとも1個、5個、10個、15個、20個、25個、30個、40個、50個、60個、80個、または100個のゲノム位置（たとえば、それぞれが目的のゲノム領域を含む）を含み得る。一部の場合には、パネル内のゲノム位置は、位置のサイズが比較的小さくなるように選択される。一部の場合には、パネル内の領域は、約10 kbもしくはそれを下回る、約8 kbもしくはそれを下回る、約6 kbもしくはそれを下回る、約5 kbもしくはそれを下回る、約4 kbもしくはそれを下回る、約3 kbもしくはそれを下回る、約2.5 kbもしくはそれを下回る、約2 kbもしくはそれを下回る、約1.5 kbもしくはそれを下回る、または約1 kbもしくはそれを下回るか、またはそれを下回るサイズを有する。一部の場合には、パネル内のゲノ

10

20

30

40

50

ム位置は、約0.5 kb～約10 kb、約0.5 kb～約6 kb、約1 kb～約11 kb、約1 kb～約15 kb、約1 kb～約20 kb、約0.1 kb～約10 kb、または約0.2 kb～約1 kbのサイズを有する。たとえば、パネル内の領域は、約0.1 kb～約5 kbのサイズを有し得る。

【0179】

本明細書において選択されるパネルは、（たとえば、試料から得られた無細胞核酸分子中の）低頻度の遺伝子バリエーションを検出するのに十分であるディープシーケンシングを可能にし得る。試料中の遺伝子バリエーションの量は、所与の遺伝子バリエーションの少数対立遺伝子頻度に関して、言及され得る。少数対立遺伝子頻度は、所与の核酸集団、たとえば、試料において、少数対立遺伝子（たとえば、もっとも一般的な対立遺伝子ではない）が生じる頻度を指し得る。低い少数対立遺伝子頻度の遺伝子バリエーションは、試料における比較的低い存在頻度を有し得る。一部の場合には、パネルは、少なくとも0.0001%、0.001%、0.005%、0.01%、0.05%、0.1%、または0.5%の少数対立遺伝子頻度の遺伝子バリエーションの検出を可能にする。パネルは、0.001%またはそれを上回る少数対立遺伝子頻度の遺伝子バリエーションの検出を可能にし得る。パネルは、0.01%またはそれを上回る少数対立遺伝子頻度の遺伝子バリエーションの検出を可能にし得る。パネルは、0.0001%、0.001%、0.005%、0.01%、0.025%、0.05%、0.075%、0.1%、0.25%、0.5%、0.75%、または1.0%ほどに低い頻度で試料中に存在する遺伝子バリエーションの検出を可能にし得る。パネルは、少なくとも0.0001%、0.001%、0.005%、0.01%、0.025%、0.05%、0.075%、0.1%、0.25%、0.5%、0.75%、または1.0%の頻度で試料中に存在する腫瘍マーカーの検出を可能にし得る。パネルは、試料中、1.0%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.75%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.5%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.25%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.1%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.075%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.05%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.025%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.01%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.005%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.001%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.0001%ほどに低い頻度の腫瘍マーカーの検出を可能にし得る。パネルは、試料中、1.0%～0.0001%ほどに低い頻度のシーケンシングされたcfDNAにおける腫瘍マーカーの検出を可能にし得る。パネルは、試料中、0.01%～0.0001%ほどに低い頻度のシーケンシングされたcfDNAにおける腫瘍マーカーの検出を可能にし得る。

【0180】

遺伝子バリエーションは、疾患（たとえば、がん）を有する対象の集団に対するパーセンテージで、表すことができる。一部の場合には、がんを有する集団のうちの少なくとも1%、2%、3%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、または99%が、パネル内の領域のうちの少なくとも1つに、1つまたは複数の遺伝子バリエーションを呈する。たとえば、がんを有する集団のうちの少なくとも80%が、パネル内のゲノム位置のうちの少なくとも1つに、1つまたは複数の遺伝子バリエーションを呈し得る。

【0181】

パネルは、1つまたは複数の遺伝子のそれぞれからの、目的のゲノム領域を含む1つまたは複数の位置を含み得る。一部の場合には、パネルは、少なくとも1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、15個、20個、25個、30個、40個、50個、または80個の遺伝子のそれぞれからの目的のゲノム領域を含む1つまたは

複数の位置を含み得る。一部の場合には、パネルは、多くとも1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、15個、20個、25個、30個、40個、50個、または80個の遺伝子のそれぞれからの目的のゲノム領域を含む1つまたは複数の位置を含み得る。一部の場合には、パネルは、約1～約80個、1～約50個、約3～約40個、5～約30個、10～約20個の異なる遺伝子のそれぞれからの目的のゲノム領域を含む1つまたは複数の位置を含み得る。

【0182】

パネル内の領域は、1つまたは複数の組織にわたって差示的に転写される配列を含むように、選択することができる。一部の場合には、ゲノム領域を含む位置は、ある特定の組織において、他の組織と比較して高いレベルで転写される配列を含み得る。たとえば、ゲノム領域を含む位置は、ある特定の組織では転写されるが他の組織では転写されない配列を含み得る。

10

【0183】

パネル内のゲノム位置は、コーディング配列および／または非コーディング配列を含み得る。たとえば、パネル内のゲノム位置は、エクソン、イントロン、プロモーター、3'非翻訳領域、5'非翻訳領域、制御エレメント、転写開始部位、および／またはスプライス部位における1つまたは複数の配列を含み得る。一部の場合には、パネル内の領域は、偽遺伝子、反復配列、トランスポゾン、ウイルスエレメント、およびテロメアを含む、他の非コーディング配列を含んでもよい。一部の場合には、パネル内のゲノム位置は、非コーディングRNA、たとえば、リボソームRNA、トランスファーRNA、Piwi結合RNA、およびマイクロRNAにおける配列を含み得る。

20

【0184】

パネル内のゲノム位置は、所望されるレベルの感度で（たとえば、1つまたは複数の遺伝子バリエーションの検出を通じて）がんを検出（診断）するように選択することができる。たとえば、パネル内の領域は、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の感度で（たとえば、1つまたは複数の遺伝子バリエーションの検出を通じて）がんを検出するように選択することができる。パネル内のゲノム位置は、100%の感度でがんを検出するように選択することができる。

【0185】

30

パネル内のゲノム位置は、所望されるレベルの特異度で（たとえば、1つまたは複数の遺伝子バリエーションの検出を通じて）がんを検出（診断）するように選択することができる。たとえば、パネル内のゲノム位置は、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の特異度で（たとえば、1つまたは複数の遺伝子バリエーションの検出を通じて）がんを検出するように選択することができる。パネル内のゲノム位置は、100%の特異度で1つまたは複数の遺伝子バリエーションを検出するように選択することができる。

【0186】

パネル内のゲノム位置は、所望される陽性的中率でがんを検出（診断）するように選択することができる。陽性的中率は、感度（たとえば、実際の陽性が検出される確率）および／または特異度（たとえば、実際の陰性を陽性であると間違わない確率）を増加させることによって、増加させることができる。非限定的な例として、パネル内のゲノム位置は、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の陽性的中率で1つまたは複数の遺伝子バリエーションを検出するように選択することができる。パネル内の領域は、100%の陽性的中率で1つまたは複数の遺伝子バリエーションを検出するように選択することができる。

40

【0187】

パネル内のゲノム位置は、所望される精度でがんを検出（診断）するように選択するこ

50

とができる。本明細書において使用される場合、「精度」という用語は、疾患状態（たとえば、がん）と健常状態とを区別する試験の能力を指し得る。精度は、感度および特異度、的中率、尤度比、ROC曲線下面積、ヨーデン指標、および／または診断オッズ比などの尺度を使用して定量化することができる。

【0188】

精度は、パーセンテージとして表すことができ、これは、正確な結果が得られた試験の回数と試験が行われた合計回数との間の比を指す。パネル内の領域は、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の精度でがんを検出するように選択することができる。パネル内のゲノム位置は、100%の精度でがんを検出するように選択することができる。

10

【0189】

パネルは、感度が高く、低頻度の遺伝子バリエーションを検出するように、選択することができる。たとえば、パネルは、試料中に0.01%、0.05%、または0.001%ほどの低い頻度で存在する遺伝子バリエーションまたは腫瘍マーカーを、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の感度で検出するように、選択することができる。パネル内のゲノム位置は、試料中に1%またはそれを下回る頻度で存在する腫瘍マーカーを、70%またはそれを上回る感度で検出するように選択することができる。パネルは、試料中に0.1%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の感度で検出するように選択することができる。パネルは、試料中に0.01%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の感度で検出するように選択することができる。パネルは、試料中に0.001%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の感度で検出するように選択することができる。

20

30

【0190】

パネルは、特異度が高く、低頻度の遺伝子バリエーションを検出するように、選択することができる。たとえば、パネルは、試料中に0.01%、0.05%、または0.001%ほどの低い頻度で存在する遺伝子バリエーションまたは腫瘍マーカーを、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の特異度で検出するように、選択することができる。パネル内のゲノム位置は、試料中に1%またはそれを下回る頻度で存在する腫瘍マーカーを、70%またはそれを上回る特異度で検出するように選択することができる。パネルは、試料中に0.1%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の特異度で検出するように選択することができる。パネルは、試料中に0.01%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の特異度で検出するように選択することができる。パネルは、試料中に0.001%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の特異度で検出するように選択することができる。

40

【0191】

パネルは、精度が高く、低頻度の遺伝子バリエーションを検出するように、選択することができる。パネルは、試料中に0.01%、0.05%、または0.001%ほどの低い頻

50

度で存在する遺伝子バリエーションまたは腫瘍マーカーを、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の精度で検出することができるように、選択することができる。パネル内のゲノム位置は、試料中に1%またはそれを下回る頻度で存在する腫瘍マーカーを、70%またはそれを上回る精度で検出するように選択することができる。パネルは、試料中に0.1%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の精度で検出するように選択することができる。パネルは、試料中に0.01%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の精度で検出するように選択することができる。パネルは、試料中に0.001%ほどの低い頻度で存在する腫瘍マーカーを、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の精度で検出するように選択することができる。

10

【0192】

パネルは、的中率が高く、低頻度の遺伝子バリエーションを検出するように、選択することができる。パネルは、試料中に0.01%、0.05%、または0.001%ほどの低い頻度で存在する遺伝子バリエーションまたは腫瘍マーカーが、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、または99.9%の陽性的中率を有し得るように、選択することができる。

20

【0193】

パネルにおいて使用されるプローブまたはベイトの濃度を増加させて(2~6 ng/μL)、試料内のより多くの核酸分子を捕捉することができる。パネルにおいて使用されるプローブまたはベイトの濃度は、少なくとも2 ng/μL、3 ng/μL、4 ng/μL、5 ng/μL、6 ng/μL、またはそれを上回り得る。プローブの濃度は、約2 ng/μL~約3 ng/μL、約2 ng/μL~約4 ng/μL、約2 ng/μL~約5 ng/μL、約2 ng/μL~約6 ng/μLであり得る。パネルにおいて使用されるプローブまたはベイトの濃度は、2 ng/μLまたはそれを上回って、6 ng/μLまたはそれを下回り得る。一部の事例では、これにより、生物学的試料内のより多くの分子を分析することが可能となり、それによって、より低頻度の対立遺伝子を検出することが可能となり得る。

30

B. がんおよび他の疾患

【0194】

ある特定の実施形態では、本明細書に開示される方法および態様は、患者における所与の疾患、障害、または状態を診断するために使用される。典型的に、考慮される疾患は、一種のがんである。そのようながんの非限定的な例としては、胆道がん、膀胱がん、移行上皮癌、尿路上皮癌、脳がん、神経膠腫、星状細胞腫、乳癌、化生性癌、子宮頸がん、子宮頸扁平上皮癌、直腸がん、結腸直腸癌、結腸がん、遺伝性非ポリポーシス結腸直腸がん、結腸直腸腺癌、消化管間質腫瘍(GIST)、子宮内膜癌、子宮内膜間質肉腫、食道がん、食道扁平上皮癌、食道腺癌、眼内黒色腫、ブドウ膜黒色腫、胆嚢癌、胆嚢腺癌、腎細胞癌、腎明細胞癌、移行上皮癌、尿路上皮癌、ウイルス腫瘍、白血病、急性リンパ球性白血病(ALL)、急性骨髄性白血病(AML)、慢性リンパ球性(CLL)、慢性骨髄性(CML)、慢性骨髄単球性(CMML)、肝臓がん、肝臓癌、ヘパトーマ、肝細胞癌、胆管癌、肝芽腫、肺がん、非小細胞肺癌(NSCLC)、中皮腫、B細胞リンパ腫、非ホジキンリンパ腫、びまん性大細胞型B細胞リンパ腫、マントル細胞リンパ腫、T細胞リンパ腫、非ホジキンリンパ腫、前駆Tリンパ芽球性リンパ腫/白血病、末梢T細胞リンパ腫、多発性骨髄腫、鼻咽頭癌(NPC)、神経芽細胞腫、口腔咽頭がん、口腔扁平上皮癌、骨肉腫、卵巣癌、膵臓がん、膵管腺癌、偽乳頭新生物、腺房細胞癌、前立腺がん、前立腺の腺癌、皮膚がん、黒色腫、悪性黒色腫、皮膚黒色腫、小腸癌、胃がん、胃癌、消化管間質腫瘍(GIST)、子宮がん、または子宮肉腫が挙げられる。

40

50

【0195】

必要に応じて本明細書に開示される方法およびシステムを使用して評価される、他の遺伝子に基づく疾患、障害、または状態の非限定的な例としては、軟骨形成不全、アルファ-1抗トリプシン欠損症、抗リン脂質症候群、自閉症、常染色体優性多発性嚢胞腎、シャルコー・マリー・トゥース（CMT）、ネコ鳴き症候群、クローン病、嚢胞性線維症、ダーカム病、ダウン症候群、デュアン症候群、デュシェンヌ型筋ジストロフィー、第V因子ライデン血栓形成傾向、家族性高コレステロール血症、家族性地中海熱、脆弱X症候群、ゴーシェ病、ヘモクロマトーシス、血友病、全前脳胞症、ハンチントン病、クラインフェルター症候群、マルファン症候群、筋強直性ジストロフィー、神経線維腫症、ヌーナン症候群、骨形成不全症、パーキンソン病、フェニルケトン尿症、ポーランド奇形、ポルフィリン症、早老症、網膜色素変性、重症複合型免疫不全症（scid）、鎌状赤血球症、脊髄性筋萎縮症、ティサックス病、サラセミア、トリメチルアミン尿症、ターナー症候群、口蓋心臓顔面症候群、WAGR症候群、ウイルソン病などが挙げられる。

10

C. カスタマイズされた治療および関連する投与

【0196】

一部の実施形態では、本明細書に開示される方法は、所与の疾患、障害、または状態を有する患者を識別し、患者に治療を投与することに関する。本質的にあらゆるがん治療（たとえば、外科手術治療、放射線療法、化学療法など）が、これらの方法の一部として含まれる。典型的には、治療には、少なくとも1つの免疫療法（または免疫療法剤）が含まれる。免疫療法は、一般に、所与のがん種に対する免疫応答を増強させる方法を指す。ある特定の実施形態では、免疫療法は、腫瘍またはがんに対するT細胞応答を増強させる方法を指す。

20

【0197】

一部の実施形態では、免疫療法または免疫療法剤は、免疫チェックポイント分子を標的とする。ある特定の腫瘍は、免疫チェックポイント経路を制圧することによって、免疫系を回避することができる。したがって、腫瘍が免疫系を回避する能力に対抗し、ある特定のがんに対する抗腫瘍免疫を活性化するための有効なアプローチとして、免疫チェックポイントの標的化が生じた。Pardoll, Nature Reviews Cancer, 2012, 12:252-264。

【0198】

ある特定の実施形態では、免疫チェックポイント分子は、抗原に対するT細胞応答に関与するシグナルを低減させる阻害性分子である。たとえば、CTLA4は、T細胞上に発現され、抗原提示細胞上のCD80（B7.1としても知られる）またはCD86（B7.2としても知られる）に結合することによって、T細胞活性化を下方制御する役割を果たす。PD-1は、T細胞上に発現される別の阻害性チェックポイント分子である。PD-1は、炎症性応答の際に、末梢組織においてT細胞の活性を制限する。加えて、PD-1のリガンド（PD-L1またはPD-L2）は、一般に、多数の異なる腫瘍の表面において上方制御されており、腫瘍微小環境において抗腫瘍免疫応答の下方制御をもたらしている。ある特定の実施形態では、阻害性免疫チェックポイント分子は、CTLA4またはPD-1である。他の実施形態では、阻害性免疫チェックポイント分子は、PD-1のリガンド、たとえば、PD-L1またはPD-L2である。他の実施形態では、阻害性免疫チェックポイント分子は、CTLA4のリガンド、たとえば、CD80またはCD86である。他の実施形態では、阻害性免疫チェックポイント分子は、リンパ球活性化遺伝子3（LAG3）、キラー細胞免疫グロブリン様受容体（KIR）、T細胞膜タンパク質3（TIM3）、ガレクチン9（GAL9）、またはアデノシンA2a受容体（A2aR）である。

30

40

【0199】

これらの免疫チェックポイント分子を標的とするアンタゴニストを使用して、ある特定のがんに対する抗原特異的T細胞応答を増強させることができる。したがって、ある特定の実施形態では、免疫療法または免疫療法剤は、阻害性免疫チェックポイント分子のアンタゴニストである。ある特定の実施形態では、阻害性免疫チェックポイント分子は、PD

50

ー1である。ある特定の実施形態では、阻害性免疫チェックポイント分子は、PD-L1である。ある特定の実施形態では、阻害性免疫チェックポイント分子のアンタゴニストは、抗体（たとえば、モノクローナル抗体）である。ある特定の実施形態では、抗体またはモノクローナル抗体は、抗CTLA4、抗PD-1、抗PD-L1、または抗PD-L2抗体である。ある特定の実施形態では、抗体は、モノクローナル抗PD-1抗体である。一部の実施形態では、抗体は、モノクローナル抗PD-L1抗体である。ある特定の実施形態では、モノクローナル抗体は、抗CTLA4抗体および抗PD-1抗体の組合せ、抗CTLA4抗体および抗PD-L1抗体の組合せ、または抗PD-L1抗体および抗PD-1抗体の組合せである。ある特定の実施形態では、抗PD-1抗体は、ペンブロリズマブ（Keytruda（登録商標））またはニボルマブ（Opdivo（登録商標））のうちの1つまたは複数である。ある特定の実施形態では、抗CTLA4抗体は、イピリムマブ（Yervoy（登録商標））である。ある特定の実施形態では、抗PD-L1抗体は、アテゾリズマブ（Tecentriq（登録商標））、アベルマブ（Bavencio（登録商標））、またはデュルバルマブ（Imfinzi（登録商標））のうちの1つまたは複数である。

【0200】

ある特定の実施形態では、免疫療法または免疫療法剤は、CD80、CD86、LAG3、KIR、TIM3、GAL9、またはA2aRに対するアンタゴニスト（たとえば、抗体）である。他の実施形態では、アンタゴニストは、阻害性免疫チェックポイント分子の可溶性バージョン、たとえば、阻害性免疫チェックポイント分子の細胞外ドメインおよび抗体のFcドメインを含む、可溶性融合タンパク質である。ある特定の実施形態では、可溶性融合タンパク質は、CTLA4、PD-1、PD-L1、またはPD-L2の細胞外ドメインを含む。一部の実施形態では、可溶性融合タンパク質は、CD80、CD86、LAG3、KIR、TIM3、GAL9、またはA2aRの細胞外ドメインを含む。一実施形態では、可溶性融合タンパク質は、PD-L2またはLAG3の細胞外ドメインを含む。

【0201】

ある特定の実施形態では、免疫チェックポイント分子は、抗原に対するT細胞応答に関与するシグナルを増幅させる共刺激性分子である。たとえば、CD28は、T細胞上に発現される共刺激性受容体である。T細胞が、そのT細胞受容体を通じて抗原に結合する際、CD28は、抗原提示細胞上のCD80（B7.1としても知られる）またはCD86（B7.2としても知られる）に結合して、T細胞受容体シグナル伝達を増幅させ、T細胞活性化を促進する。CD28は、CTLA4と同じリガンド（CD80およびCD86）に結合するため、CTLA4は、CD28によって媒介される共刺激性シグナル伝達に対抗するかまたはそれを制御することができる。ある特定の実施形態では、免疫チェックポイント分子は、CD28、誘導性T細胞共刺激因子（ICOS）、CD137、OX40、またはCD27から選択される共刺激性分子である。他の実施形態では、免疫チェックポイント分子は、たとえば、CD80、CD86、B7RP1、B7-H3、B7-H4、CD137L、OX40L、またはCD70を含む、共刺激性分子のリガンドである。

【0202】

これらの共刺激性チェックポイント分子を標的とするアゴニストを使用して、ある特定のがんに対する抗原特異的T細胞応答を増強させることができる。したがって、ある特定の実施形態では、免疫療法または免疫療法剤は、共刺激性チェックポイント分子のアゴニストである。ある特定の実施形態では、共刺激性チェックポイント分子のアゴニストは、アゴニスト抗体であり、好ましくは、モノクローナル抗体である。ある特定の実施形態では、アゴニスト抗体またはモノクローナル抗体は、抗CD28抗体である。他の実施形態では、アゴニスト抗体またはモノクローナル抗体は、抗ICOS、抗CD137、抗OX40、または抗CD27抗体である。他の実施形態では、アゴニスト抗体またはモノクローナル抗体は、抗CD80、抗CD86、抗B7RP1、抗B7-H3、抗B7-H4、抗CD137L、抗OX40L、または抗CD70抗体である。

【0203】

がん以外の特定の遺伝子に基づく疾患、障害、または状態を処置するための治療選択肢は、概して、当業者に周知であり、具体的な疾患、障害、または状態を考慮すると、明らかであろう。

【0204】

ある特定の実施形態では、本明細書に記載されるカスタマイズされた治療は、典型的に、非経口で（たとえば、静脈内または皮下に）投与される。免疫療法剤を含有する医薬組成物は、典型的に、静脈内投与される。ある特定の治療剤は、経口で投与される。しかしながら、カスタマイズされた治療（たとえば、免疫療法剤など）はまた、たとえば、頬内、舌下、直腸内、膈内、尿道内、局所、眼内、鼻内、および／または心房内を含む、当該技術分野において公知の任意の方法によって、投与され得、この投与には、錠剤、カプセル、顆粒、水性懸濁液、ゲル、スプレー、坐剤、膏薬、軟膏などが含まれ得る。

【実施例】

【0205】

（実施例1）

PPGの検出

17,825個の臨床試料のセットを、Guardant Health, Inc. (Redwood City, CA) から入手した73遺伝子パネルcfDNA試験を使用して、処理し、分析した。このセットの中で、107個の試料を、以下の表4に示されるように、112個の試料特異的PPGを有するとして識別した。これは、試料当たりのPPG率0.6%、または臨床試料167個につき1つの試料特異的PPGに相当する。

【表4-1】

表4

遺伝子	PPGs	遺伝子	PPGs
SMAD4	49	CCND1	2
GNAS	19	HRAS	2
TP53	9	MET	2
RAF1	5	MYC	2

【表4-2】

CDK4	4	NRAS	2
MAPK1	4	シングルトン	9
STK11	3		

【0206】

表4には、試料特異的PPGが少なくとも1つの試料において検出されたすべての遺伝子を示すが、すべてのシングルTONは、「シングルTON」のカテゴリーに集約されている。

【0207】

生殖細胞系および体細胞系の両方の試料特異的PPGによって生じるエクソン-エクソンジャンクションにまたがるアライメントアーチファクトは、図5に示されるように、疑似バリエントコールをもたらす得る。PPGの存在は、イントロン配列が欠如した複数のソフトクリップリードの存在、ならびにイントロン-エクソン境界部におけるカバレッジの不連続性の両方によって判明する。矢印によって示される疑似A.C.SNVコールは、1.7%の対立遺伝子頻度(AF)で観察される。

（実施例2）

PPGの臨床的重要性

【0208】

PPGの存在は、偽陽性バリエーションコールの2つの異なる原因をもたらし得る。第1に、PPGによって生じるPPGエクソン-エクソンジャンクションにまたがるリード間でのアライメントアーチファクトにより、疑似バリエーションコールが生じ得る（図6）。第2に、PPGに存在するSNVは、もともとの遺伝子にマッピングされ得る。

【0209】

PPGが検出されなかった10,000個の臨床試料のランダムなサブセットを使用して、イントロン-エクソン境界部（図7A）およびコーディング配列（CDS）内（図7B）において偶然で予測されるであろうものよりも多くのSNVをもたらすいくつかの遺伝子のPPGコピーの存在が、観察された。

（実施例3）

偽陽性バリエーションの排除

【0210】

合計で、スプライスジャンクションでは48個のSNV、ならびにCDSでは32個のSNVが、HRAS、RAF1、SMAD4、およびTP53からのPPGの存在に起因する可能性があるとして判定された。PPGを意識した偽陽性バリエーションの抑制を行うことにより、表5に示されるように、試料当たりの偽陽性率の増加0.45%（80/17,825）が、回避された。

【表5】

表5

位置	抑制される偽陽性	
	SNVs	試料当たりの率
スプライスジャンクション	48	0.27%
CDS	32	0.18%

（実施例4）

TYRO3 PPGによって生じる偽陽性バリエーションの検出および抑制

【0211】

2,094個の患者試料のセットを、Guardant Health, Inc.（Redwood City, CA）から入手した500遺伝子パネルcfDNA試験を使用して、処理し、分析した。このセットの中で、1,140個の試料を、遺伝子TYRO3に関して、試料特異的PPGを有するとして識別した。これは、試料当たりのPPG率54%、または試料2つにつき1つのPPGに相当する。これらの試料を、偽陽性が疑われる第15染色体41,862,477位のTYRO3遺伝子座におけるCからTの突然変異（TYRO3 c.1422C>Tとして知られている）の存在について評価した。

【表6】

表6

	試料	TYRO3 c.1422C>T が検出されたもの
PPGが検出されたもの	1,140	11
PPGが検出されなかったもの	954	0

【0212】

表 6 において、P P G が検出された 1 1 個の試料において偽陽性が疑われるバリエーションが観察されるが、P P G が検出されていない試料では観察されず、統計学的に有意な差があった（フィッシャー直接検定、 $p = 0.0013$ ）。バリエーションは、P P G の存在下においてのみ確認されるため、T Y R O 3 遺伝子座にアライメントする P P G を起源とするリードのアーチファクトであることが示唆される。

【0213】

P P G によって生じるエクソン-エクソンジャンクションにまたがるアライメントアーチファクトは、図 8 に示されるように、T Y R O 3 遺伝子座の状況で示されている。疑似 C . T . S N V コール (T Y R O 3 c . 1 4 2 2 C > T) は、矢印で示される。

【0214】

本開示の様々な実施形態が本明細書に示され、説明されているが、そのような実施形態が単なる例示として提供されていることが、当業者には理解されるであろう。当業者であれば、本開示から逸脱することなく、多数の変化形、変更、および置換を想起し得る。本明細書に記載される本開示の実施形態に対する様々な代替形を、用いることができることを理解されたい。

【0215】

上記または下記に引用されるすべての特許出願、ウェブサイト、他の刊行物、受託番号などは、それぞれ個々の品目が、具体的かつ個別に、参照により組み込まれると示されるのと同程度に、あらゆる目的で、参照によりその全体が本明細書に組み込まれる。配列の異なるバージョンが、異なる時点で、1つの受託番号と関連付けられている場合、本出願の有効出願日の時点で受託番号と関連付けられているバージョンを意味する。有効出願日は、実際の出願日、または該当する場合には受託番号に言及している優先出願の出願日のうちの早い方を意味する。同様に、刊行物、ウェブサイトなどの異なるバージョンが、異なる時点で公開されている場合、別途示されない限り、本出願の有効出願日の時点でもっとも新しく公開されているバージョンを意味する。本開示の任意の特性、ステップ、要素、実施形態、または態様は、別途具体的に示されない限り、任意の他のものと組み合わせることができる。本開示は、明確さおよび理解の目的で、図示および例を用いていくらか詳細に記載されているが、添付の特許請求の範囲の範囲内で、ある特定の変更および修正が実施されてもよいことは明らかであろう。

本発明は、例えば、以下の項目を提供する。

(項目 1)

少なくとも部分的にコンピュータを使用して、遺伝子配列リードにおけるアライメントエラーを検出するための方法であって、

(a) 前記コンピュータによって、対象からの生物学的試料中の無細胞核酸分子から得られた前記遺伝子配列リードを含む、配列情報を受信するステップと、

(b) 前記遺伝子配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、

(c) 前記アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、

(d) 前記遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、前記遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、前記領域が、前記遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップとを含む、方法。

(項目 2)

少なくとも部分的にコンピュータを使用して、対象の生物学的試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、

(a) 前記コンピュータによって、前記無細胞核酸分子から得られた配列リードを含む、配列情報を受信するステップと、

(b) 前記配列リードを、参照配列に対してアライメントして、アライメントされた配列

10

20

30

40

50

リードを生成するステップと、

(c) 前記アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、

(d) 前記遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、前記遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、前記領域が、前記遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、

(e) 前記1つまたは複数の遺伝子融合リードの前記サブセットにおける前記1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、

(f) 前記参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法。

(項目3)

少なくとも部分的にコンピュータを使用して、対象の試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、

(a) 前記コンピュータによって、前記無細胞核酸分子から得られたシーケンシングリードを含む、配列情報を受信するステップと、

(b) 前記配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、

(c) 前記アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、

(d) 遺伝子バリエーションを含む、前記遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、前記遺伝子融合リードのうちの前記1つまたは複数の前記サブセットが、SMAD4、TYRO3、および／またはRAF1に対応する遺伝子配列を含む、ステップと、

(e) 前記遺伝子融合リードのうちの前記1つまたは複数の前記サブセットにおける前記1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、

(f) 前記参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法。

(項目4)

少なくとも部分的にコンピュータを使用して、遺伝子配列リードにおけるアライメントエラーを検出するための方法であって、

(a) 前記コンピュータによって、対象からの生物学的試料中の無細胞核酸分子から得られた前記遺伝子配列リードを含む、配列情報を受信するステップと、

(b) 前記遺伝子配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、

(c) 前記アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを判定するステップと、

(d) 前記遺伝子内融合ブレイクポイントを含む領域内に遺伝子バリエーションを含む、前記遺伝子融合リードのうちの1つまたは複数のサブセットを判定するステップであって、前記領域が、前記遺伝子内融合ブレイクポイントに隣接する1つまたは複数のヌクレオチドを含む、ステップと、

(e) 所定の基準を満たす前記領域内のそれぞれの遺伝子バリエーションを、アライメントエラーとして識別するステップとを含む、方法。

(項目5)

少なくとも部分的にコンピュータを使用して、対象の試料からの無細胞核酸分子において真の遺伝子バリエーションを検出することにおいて、アライメントエラーを抑制するための方法であって、

- (a) 前記コンピュータによって、前記無細胞核酸分子から得られたシーケンシングリードを含む、配列情報を受信するステップと、
- (b) 前記配列リードを、参照配列に対してアライメントして、アライメントされた配列リードを生成するステップと、
- (c) 前記アライメントされた配列リードから、遺伝子内融合ブレイクポイントを含む、遺伝子融合リードのセットを識別するステップと、
- (d) 遺伝子バリエーションを含む、前記遺伝子融合リードのうちの1つまたは複数のサブセットを識別することによって、アライメントエラーを検出するステップであって、前記遺伝子融合リードのうちの前記1つまたは複数の前記サブセットが、SMAD4、TYRO3、および／またはRAF1に対応する遺伝子配列を含む、ステップと、
- (e) 前記遺伝子融合リードのうちの前記1つまたは複数の前記サブセットにおける前記1つまたは複数の検出されたアライメントエラーの少なくとも一部分をフィルタリングして、フィルタリングされた配列リードを生成するステップと、
- (f) 前記参照配列と比較して、真の遺伝子バリエーションを含む、フィルタリングされた配列リードを検出するステップとを含む、方法。

10

(項目6)

前記遺伝子融合リードの前記セットが、1つまたは複数のプロセスされた偽遺伝子(PPG)に対応する、項目1から5のいずれか一項に記載の方法。

(項目7)

前記1つまたは複数のPPGが、1つまたは複数の試料特異的PPGを含む、項目6に記載の方法。

20

(項目8)

前記1つまたは複数の試料特異的PPGにより、対象の集団において、前記対象が識別される、項目7に記載の方法。

(項目9)

前記1つまたは複数のPPGが、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASからなる群に由来する、項目6に記載の方法。

(項目10)

前記1つまたは複数のPPGが、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASからなる群に由来する2つまたはそれを上回るPPGを含む、項目6に記載の方法。

30

(項目11)

前記1つまたは複数のPPGが、SMAD4、GNAS、TP53、RAF1、CDK4、TYRO3、MAPK1、STK11、CCND1、HRAS、MET、MYC、およびNRASからなる群に由来する3つまたはそれを上回るPPGを含む、項目6に記載の方法。

(項目12)

前記遺伝子バリエーションまたは真の遺伝子バリエーションが、単一ヌクレオチドバリエーション(SNV)または挿入もしくは欠失(インデル)を含む、項目1から11のいずれか一項に記載の方法。

40

(項目13)

前記遺伝子バリエーションが、SNVを含む、項目12に記載の方法。

(項目14)

前記SNVが、イントロン-エクソン境界部に位置する、項目12に記載の方法。

(項目15)

前記SNVが、遺伝子コーディング配列(CDS)内に位置する、項目12に記載の方法。

(項目16)

50

前記遺伝子バリエーションが、インデルを含む、項目 12 に記載の方法。

(項目 17)

前記領域が、前記遺伝子内融合ブレイクポイントに隣接する約 2、4、6、8、10、15、または 20 個のヌクレオチドを含む、項目 1 に記載の方法。

(項目 18)

前記 1 つまたは複数の検出されたアライメントエラーの一部分が、前記試料中の前記遺伝子内融合ブレイクポイントに対応する前記遺伝子内融合の画分よりも低いかまたはそれと同等である前記試料中の突然変異対立遺伝子画分を有する、前記検出されたアライメントエラーに基づいて、フィルタリングされる、先行する項目のいずれかに記載の方法。

(項目 19)

前記 1 つまたは複数の検出されたアライメントエラーの一部分が、事前に定義された臨床的に対処可能なバリエーションのセットに属さない遺伝子バリエーションを含む前記遺伝子融合リードに基づいて、フィルタリングされる、項目 18 に記載の方法。

(項目 20)

前記試料が、血液、血漿、血清、尿、唾液、粘膜排出物、喀痰、糞便、および涙液からなる群から選択される体液試料である、先行する項目のいずれか一項に記載の方法。

(項目 21)

前記対象が、疾患または障害を有する、先行する項目のいずれか一項に記載の方法。

(項目 22)

前記疾患が、がんである、項目 21 に記載の方法。

(項目 23)

前記対象の前記生物学的試料から、無細胞核酸分子を単離するステップを含む、先行する項目のいずれか一項に記載の方法。

(項目 24)

前記無細胞核酸分子が、DNA、RNA、またはこれらの組合せを含む、項目 23 に記載の方法。

(項目 25)

前記無細胞核酸分子が、二本鎖 DNA である、項目 24 に記載の方法。

(項目 26)

シーケンシングの前に、分子バーコードを含む 1 つまたは複数のアダプターを、前記無細胞核酸分子に結合させて、タグ付けされた親ポリヌクレオチドを生成するステップをさらに含む、先行する項目のいずれか一項に記載の方法。

(項目 27)

前記アダプターが、前記無細胞核酸分子の両端に結合される、項目 26 に記載の方法。

(項目 28)

前記無細胞核酸分子が、固有にバーコーディングされる、項目 26 に記載の方法。

(項目 29)

前記無細胞核酸分子が、非固有にバーコーディングされる、項目 26 に記載の方法。

(項目 30)

それぞれのバーコードが、選択された領域からシーケンシングされる多様な分子と組み合わせて、固有な分子の識別を可能にする、固定またはセミランダムなオリゴヌクレオチド配列を含む、項目 29 に記載の方法。

(項目 31)

前記タグ付けされた親ポリヌクレオチドを増幅させて、子孫ポリヌクレオチドを生成するステップをさらに含む、項目 26 に記載の方法。

(項目 32)

目的の標的配列に関して、前記子孫ポリヌクレオチドを選択的に濃縮させ、それによって、濃縮された子孫ポリヌクレオチドを生成するステップをさらに含む、項目 31 に記載の方法。

(項目 33)

10

20

30

40

50

前記濃縮された子孫ポリヌクレオチドを増幅させるステップをさらに含む、項目 3 2 に記載の方法。

(項目 3 4)

前記子孫ポリヌクレオチドまたは濃縮された子孫ポリヌクレオチドに、試料インデックス配列がタグ付けされる、項目 3 1 から 3 3 のいずれか一項に記載の方法。

(項目 3 5)

前記配列情報が、核酸シーケンサーから得られる、先行するいずれかの項目に記載の方法。

(項目 3 6)

前記遺伝子融合リードのセットが、シーケンシングされたペアエンドリードをアライメントおよび接続することによって識別される、先行する項目のいずれか一項に記載の方法。

(項目 3 7)

前記遺伝子融合リードのセットが、イントロン-エクソン境界部にまたがるカバレッジにおける不連続性に基づいて識別される、先行する項目のいずれか一項に記載の方法。

(項目 3 8)

前記事前に定義されたセットが、COSMIC、The Cancer Genome Atlas (TCGA)、または Exome Aggregation Consortium (ExAC) において見出されるバリエーションを含む、項目 1 9 に記載の方法。

(項目 3 9)

少なくとも部分的にコンピュータを使用して、フィルタリングされたリード配列情報データセットを生成するための方法であって、

(a) 対象から得られた生物学的試料中の無細胞核酸 (cfNA) から得られた試験配列リードのセットにおいて、1 つまたは複数のスプリット配列リードを識別するステップであって、それぞれのスプリット配列リードが、少なくとも 1 つのブレイクポイントを含む、ステップと、

(b) 前記試験配列リードのセットにおいて、(i) 所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも 1 つの配列バリエーションを含む、前記スプリット配列リードのうちの 1 つもしくは複数の少なくとも一部分および／または前記試験配列リードのうちの 1 つもしくは複数の少なくとも一部分を抑制し、それによって前記フィルタリングされた配列情報データセットを生成するか、あるいは (ii) 所与のブレイクポイントから選択されたヌクレオチド数以内に少なくとも 1 つの配列バリエーションを含む前記スプリット配列リードの 1 つもしくは複数のベースコールおよび／または前記試験配列リードの 1 つもしくは複数のベースコールを抑制し、それによって、前記フィルタリングされた配列情報データセットを生成するステップとを含む、方法。

10

20

30

40

50

【図面】
【図 1】

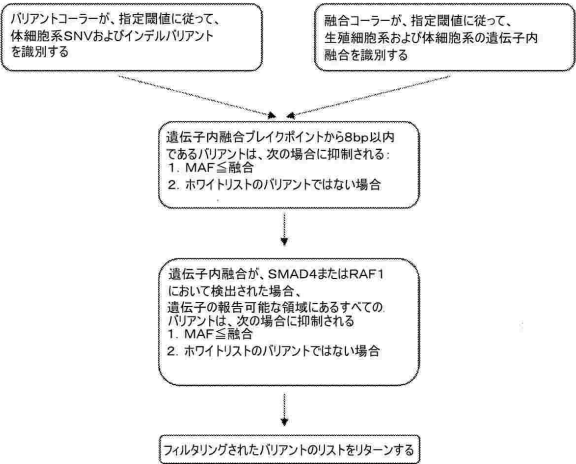


FIG. 1

【図 2】

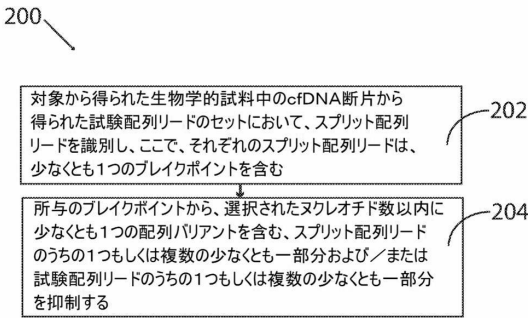


FIG. 2

【図 3】

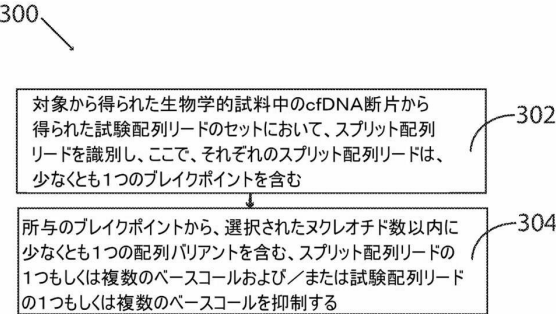


FIG. 3

【図 4】

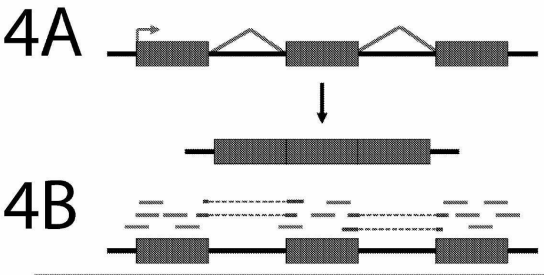


FIG. 4

【図 5】

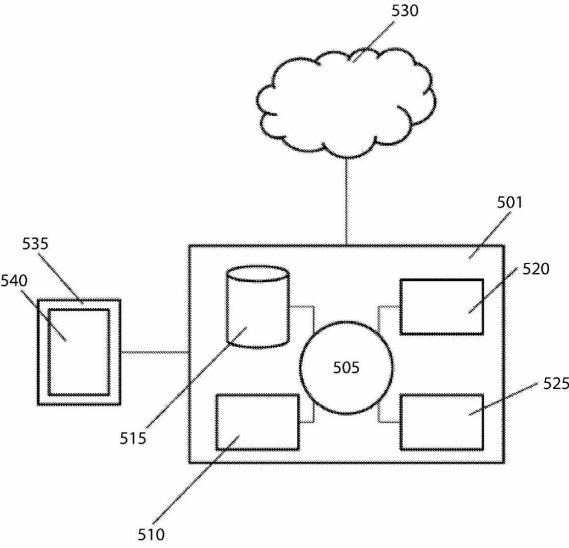


FIG. 5

【図 6】

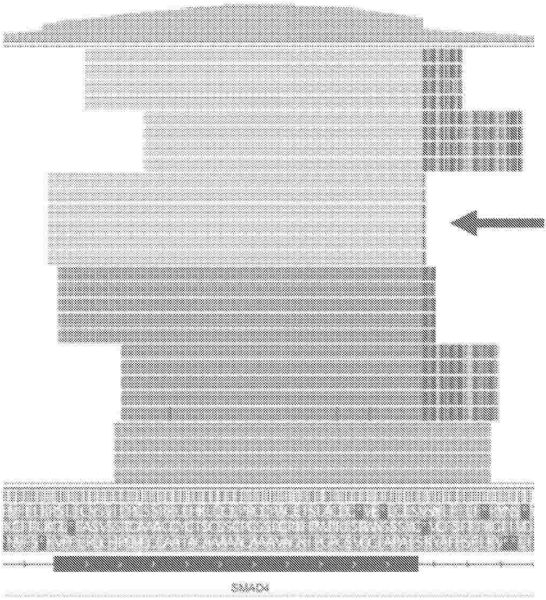


FIG. 6

【図 7】

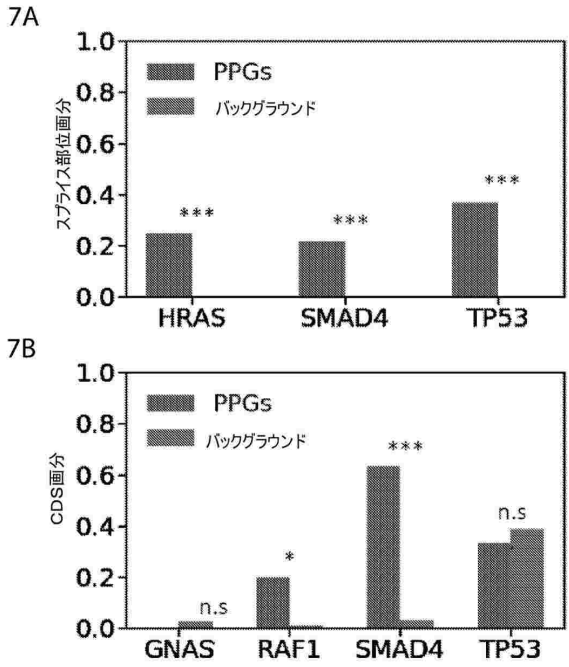


FIG. 7

【図 8】

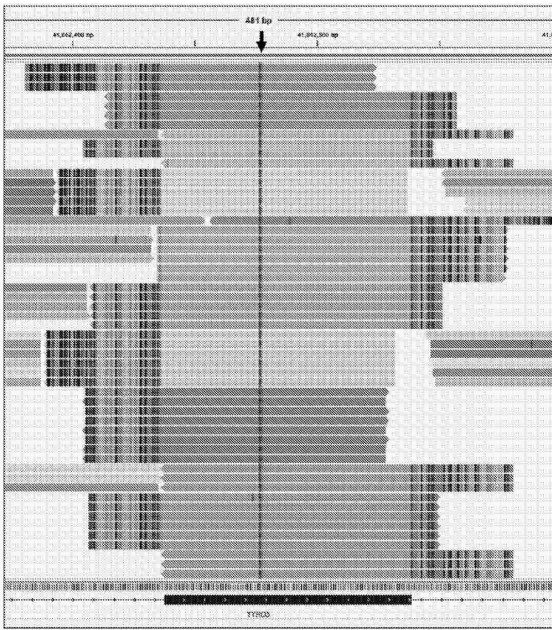


FIG. 8

10

20

30

40

50

フロントページの続き

- 弁護士 山本 健策
- (72)発明者 アルティエリ, カルロ
アメリカ合衆国 カリフォルニア 94063, レッドウッド シティ, ペノブスコット ドライブ
505
- (72)発明者 シコラ, マーシン
アメリカ合衆国 カリフォルニア 94063, レッドウッド シティ, ペノブスコット ドライブ
505
- 審査官 千葉 直紀
- (56)参考文献 特表2016-518123 (JP, A)
Genome Biology, 2013年, Vol. 14, No. 4, R36, pp. 1-13
- (58)調査した分野 (Int.Cl., DB名)
C12Q 1/6869
C12N 15/09
C12Q 1/6844
C12Q 1/6886