

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7577067号
(P7577067)

(45)発行日 令和6年11月1日(2024.11.1)

(24)登録日 令和6年10月24日(2024.10.24)

(51)国際特許分類	F I	
G 1 6 B 20/20 (2019.01)	G 1 6 B 20/20	
C 1 2 Q 1/68 (2018.01)	C 1 2 Q 1/68	
C 1 2 M 1/00 (2006.01)	C 1 2 M 1/00	A

請求項の数 8 (全51頁)

(21)出願番号	特願2021-549786(P2021-549786)	(73)特許権者	515059083
(86)(22)出願日	令和2年2月27日(2020.2.27)		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2022-521777(P2022-521777 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4年4月12日(2022.4.12)	(74)代理人	100078282
(86)国際出願番号	PCT/US2020/020174		弁理士 山本 秀策
(87)国際公開番号	WO2020/176775	(74)代理人	100113413
(87)国際公開日	令和2年9月3日(2020.9.3)		弁理士 森下 夏樹
審査請求日	令和5年2月27日(2023.2.27)	(74)代理人	100181674
(31)優先権主張番号	62/811,159		弁理士 飯田 貴敏
(32)優先日	平成31年2月27日(2019.2.27)	(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
(31)優先権主張番号	62/823,585	(74)代理人	230113332
(32)優先日	平成31年3月25日(2019.3.25)		
最終頁に続く		最終頁に続く	

(54)【発明の名称】 対立遺伝子頻度に基づく機能喪失のコンピューターモデリング

(57)【特許請求の範囲】
【請求項1】

遺伝子の生殖細胞系列欠失を示さない試料中の前記遺伝子の体細胞ホモ接合欠失と体細胞ヘテロ接合欠失とを識別するためのコンピューターシステムであって、

第1の確率分布により、試料のトレーニングセットにおける前記遺伝子に関連する1つまたは複数の生殖細胞系列一塩基多型（SNP）のヘテロ接合性の出現率に基づいて、および、試料の前記トレーニングセットにおける前記1つまたは複数の生殖細胞系列SNPと関連するマイナー対立遺伝子頻度（MAF）の標準偏差に基づいて、前記試料が前記体細胞ホモ接合欠失を含む第1の確率を表す、対立遺伝子数の第1のモデルを作成し、ここで、前記第1のモデルを作成するために、前記体細胞ホモ接合欠失の存在を支持する前記試料中の分子の数と前記試料中の分子の総数とが判定され、試料の前記トレーニングセットが、腫瘍が検出されない（TND）複数の試料を含み、

第2の確率分布により、試料の前記トレーニングセットにおける前記1つまたは複数の生殖細胞系列SNPのヘテロ接合性の出現率に基づいて、および、試料の前記トレーニングセットにおける前記1つまたは複数の生殖細胞系列SNPと関連するマイナー対立遺伝子頻度（MAF）の標準偏差に基づいて、前記試料が前記体細胞ヘテロ接合欠失を含む第2の確率を表す、前記試料中の対立遺伝子数の第2のモデルを作成し、ここで、前記第2のモデルを作成するために、前記試料から推定された腫瘍含有率の平均と前記試料から推定された腫瘍含有率の標準偏差とが判定され、

前記第1のモデルの第1の出力と前記第2のモデルの第2の出力とを比較し、ここで

、前記第1の出力が、前記体細胞ホモ接合欠失の存在についての前記第1の確率を含み、前記第2の出力が、前記体細胞ヘテロ接合欠失の存在についての前記第2の確率を含み、

前記第1の確率と前記第2の確率の間の前記比較に基づいて、前記遺伝子に関する前記体細胞ホモ接合欠失が前記試料中に存在するという予測を作成し、ここで、前記第1の確率が前記第2の確率よりも高い場合に、前記遺伝子に関する前記体細胞ホモ接合欠失が前記試料中に存在すると予測される

ようにプログラミングされたプロセッサを含む、コンピュータシステム。

【請求項2】

(a) 前記第1の確率分布が、前記第2の確率分布と同じ種類の確率分布である、

(b) 前記第1のモデルを作成するために、前記プロセッサが、前記第1の確率分布に対する入力のための1つまたは複数のパラメータを判定するようにプログラミングされている、かつ／または

(c) 前記第1の確率分布が、ベータ二項分布、二項分布、または正規分布のうちの1つを含む1種の確率分布を含む、
請求項1に記載のコンピュータシステム。

【請求項3】

前記第1のモデルを作成するために、前記プロセッサが、
体細胞ホモ接合欠失と関連する分子カバレッジに基づいて、前記体細胞ホモ接合欠失を仮定して、前記試料において、前記1つまたは複数の生殖細胞系列SNPの位置の前記対立遺伝子数の第1の尤度を計算する
ようにさらにプログラミングされている、請求項1および2のいずれか一項に記載のコンピュータシステム。

【請求項4】

前記第2のモデルを作成するために、前記プロセッサが、
体細胞ヘテロ接合欠失と関連する分子カバレッジに基づいて、前記体細胞ヘテロ接合欠失を仮定して、前記試料において、前記1つまたは複数の生殖細胞系列SNPの位置の前記対立遺伝子数の第2の尤度を計算する
ようにさらにプログラミングされている、請求項1～3のいずれか一項に記載のコンピュータシステム。

【請求項5】

前記腫瘍含有率が、配列カバレッジ情報に基づいて推定される、請求項1～4のいずれか一項に記載のコンピュータシステム。

【請求項6】

前記プロセッサが、
複数の試料にアクセスし、
生殖細胞系列欠失を含む前記複数の試料の中から試料のセットを特定し、
前記複数の試料から試料の前記セットをフィルタリングして出力し、
前記フィルタリングされた複数の試料の中から、前記体細胞ホモ接合欠失または前記体細胞ヘテロ接合欠失の存在を特定する
ようにさらにプログラミングされている、請求項1～5のいずれか一項に記載のコンピュータシステム。

【請求項7】

前記プロセッサが、
前記第1の出力と前記第2の出力とに基づいて、対数尤度関数を実行する
ようにさらにプログラミングされている、請求項1～6のいずれか一項に記載のコンピュータシステム。

【請求項8】

前記遺伝子が、BRCA1、BRCA2、およびATMのうちの1つを含む、請求項1～7のいずれか一項に記載のコンピュータシステム。

10

20

30

40

50

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

本出願は、すべての目的に関して参照により本明細書に組み込まれる、2019年2月27日に提出した米国仮出願第62/811,159号、および2019年3月25日に提出した米国仮出願第62/823,585号に基づく優先権の利益を主張する。

【背景技術】

【0002】

背景

腫瘍は、細胞の異常な成長である。断片化したDNAは、細胞、例えば腫瘍細胞が死滅すると、体液中に放出されることが多い。よって、体液中の無細胞DNAの一部は、腫瘍のDNAである。腫瘍は、良性であっても悪性であってもよい。悪性腫瘍は、がんと呼ばれることが多い。

【0003】

がんは、世界中で主要な病因である。各年、世界中で数千万人の人々ががんを有すると診断され、その半分より多くが、最終的にがんが原因で死亡する。多くの国では、がんは、心血管疾患に続く2番目に多い死因として位置付けられる。早期検出は、多くのがんにとって転帰の改善に関連する。

【0004】

がんは、個体の正常細胞内での突然変異および／またはエピジェネティック変異の蓄積によって引き起こされ、少なくともその一部は不適切に調節された細胞分裂をもたらす。このような突然変異、または遺伝物質の状態は、一般的に、コピー数変異(CNV)、コピー数異常(CNA)、一塩基変異(SNV)、遺伝子融合およびインデルを含み、エピジェネティック変異は、シトシンの6原子環の5番目の原子に対する改変ならびにDNAのクロマチンおよび転写因子との会合を含む。

【0005】

特定の例では、相同組換え修復(HRR)遺伝子(BRCA1/2)のヘテロ接合性の喪失(LOH)および二対立遺伝子のコピー数喪失は、腫瘍抑制機能の喪失に関連し、がんをもたらす。多くの場合には、目的の遺伝子の特定の状態によって、処置の種類についての情報を得ることができる。例えば、遺伝子の1つの状態は一組の薬物に応答することができるが、一方、その遺伝子の別の状態は応答することができない。よって、がんおよび他の疾患を単に診断することができるだけでなく、疾患の根本原因を特徴付けることができることが、益々重要である。

【0006】

がんは、腫瘍の生検と、その後の細胞、マーカーまたは細胞から抽出されたDNAの解析によって検出されることが多い。体液の解析に基づいてがんを検出するための研究は、進行中である。成功すれば、これらの検査は、これらが非侵襲的であり、生検によって疑わしいがん細胞を特定することなく実施することができるという利点を有する。しかし、体液中の核酸量が非常に少ないという事実によって、これらの種類の検査を首尾よく完了することは難解である。さらに、体液中の検出可能な腫瘍関連無細胞核酸の量は、無細胞DNAにおけるがんの解析および検出をさらに困難にする可能性がある。言い換えれば、体液中の腫瘍DNAには正常なDNAが夾雑している場合があり、無細胞DNAの試料中の腫瘍の特定原因のコンピューター解析および検出を困難にする。

【発明の概要】

【課題を解決するための手段】

【0007】

要旨

本開示は、試料中の無細胞DNAから配列決定された遺伝子などの遺伝物質の様々な状態について、精度の高い診断をもたらすコンピューター技術に関する。この状態は、限定

10

20

30

40

50

されないが、体細胞ホモ接合欠失、体細胞ヘテロ接合欠失、コピー数変異（「CNV」）（特定コピー数の野生型、増幅、または喪失を含む）などの遺伝子の突然変異状態、および／または他の状態を含み得る。精度の高い診断は、これらの状態の1つまたは複数の確率モデルに基づいてもよい。例えば、コンピューターシステムは、遺伝物質がある特定の状態にある確率を各々出力する競合モデルを作成することができる。

【0008】

各モデルは、トレーニング試料のセットに関してトレーニングされ、遺伝物質がそれぞれの状態にある確率を出力することができる。例えば、第1のモデルは、遺伝物質が特定の遺伝子の対立遺伝子の体細胞ホモ接合欠失を含む第1の確率に関し、これを出力することができる。第2のモデルは、遺伝物質が特定の遺伝子の対立遺伝子の体細胞ヘテロ接合欠失を含む第2の確率に関し、これを出力することができる。他のモデルは、遺伝物質のCNVなどの他の種類の状態の確率に関し、これを出力することができる。コンピューターシステムは、各競合モデルの出力を比較して、どのモデルがより可能性が高いかを判定することができる。例えば、コンピューターシステムは、競合する第1および第2の確率の対数尤度比を使用して、遺伝物質が体細胞ホモ接合欠失または体細胞ヘテロ接合欠失を含むか否かを判定することができる。

【0009】

一部の実施形態では、コンピューターシステムは、様々な確率分布を使用して、モデルを作成することができる。例えば、コンピューターシステムは、ベータ二項分布、二項分布、正規分布（「ガウス分布」とも称される）、および／または他の種類の確率モデリング技法を使用することができる。コンピューターシステムは、トレーニングデータセットに基づいて、状態（特定の状態を支持する対立遺伝子数など）をモデリングして、非正常、または腫瘍の状態のベースライン期待値を設定することができる。例えば、コンピューターシステムは、「正常な」または非腫瘍試料、例えば、体細胞バリエーションが観察されない試料において観察された生殖細胞系列塩基多型（SNP）の位置を特定することができる。これらの試料は、腫瘍非検出（tumor not detected）（TND）試料とも称されるであろう。

【0010】

TND試料は正常であるため、コンピューターシステムは、生殖細胞系列SNPの位置が非正常状態の原因ではないと仮定することができる。このように、コンピューターシステムは、これらのSNP部位を活用して、この状態の確率的判定のために対立遺伝子数をモデリングするための参照期待値とすることができる。例えば、各SNPの位置において観察されたヌクレオチドコールからの偏差は、このような偏差が、腫瘍または他の非正常状態などの特定の状態の原因である確率を示し得る。したがって、コンピューターシステムは、TND試料の生殖細胞系列SNPからのデータに関する計算結果に由来する期待値に基づいて、モデルをトレーニングすることができる。このような計算されたデータは、各SNP部位に対して：ヘテロ接合性の出現率、MAFの標準偏差、遺伝子型、生殖細胞系列の出現率（事前の）、および／または個体の試料の解析に対して情報を与え得る他のデータを含み得る。

【0011】

計算された期待値を用いて、コンピューターシステムは、目的の領域、例えば、目的の遺伝子上流の、下流の、およびそれを含む領域に対してアラインする、検査される個体の試料の配列読み取りデータに基づく状態をモデリングすることができる。一部の実施形態では、個体の試料から得られた分子の配列読み取りデータは、参照ゲノムに対してアラインされ、元の分子が支持する対立遺伝子（突然変異体または野生型）を特定することができる。個体の試料から得られた配列読み取りデータのアライメントに基づき、コンピューターシステムは、代替対立遺伝子を支持する分子の数を特定し、分子の総数を計算することができる。コンピューターシステムは、目的の領域における各生殖細胞系列SNPから計算された期待値データを用いて、個体の試料に由来するこれらのおよび／または他のデータをモデリングすることができる。一部の例では、配列決定は、血漿中無細胞DNA

10

20

30

40

50

(c f D N A) の標的化配列決定に基づいてもよい。

【0012】

一態様では、本開示は、遺伝子の生殖細胞系列欠失を示さない試料中の遺伝子の体細胞ホモ接合欠失と体細胞ヘテロ接合欠失とを識別するために改良されたコンピューターシステムに関する。このコンピューターシステムは、第1の確率分布により、遺伝子に関連する1つまたは複数の生殖細胞系列一塩基多型 (S N P) の位置に基づいて、体細胞ホモ接合欠失を表す、対立遺伝子数の第1のモデルを作成するようにプログラミングされたプロセッサを含むことができる。プロセッサは、第2の確率分布により、1つまたは複数の生殖細胞系列 S N P の位置に基づいて、体細胞ヘテロ接合欠失を表す、試料中の対立遺伝子数の第2のモデルをさらに作成することができる。プロセッサは、第1のモデルの第1の出力と第2のモデルの第2の出力とを比較することができる。プロセッサは、比較に基づいて、遺伝子に関する体細胞ホモ接合欠失が試料中に存在するという予測を作成することができる。

10

【0013】

一部の実施形態では、第1のモデルは、試料が体細胞ホモ接合欠失を含む第1の確率を表すことができ、第2のモデルは、試料が体細胞ヘテロ接合欠失を含む第2の確率を表す。

【0014】

一部の実施形態では、第1の確率分布は、第2の確率分布と同じ種類の確率分布である。

【0015】

一部の実施形態では、第1のモデルを作成するために、プロセッサは、第1の確率分布に対する入力のための1つまたは複数のパラメーターを判定するようにプログラミングされている。

20

【0016】

一部の実施形態では、第1の確率分布は、ベーター二項分布、二項分布、または正規分布を含む。

【0017】

対立遺伝子数の第1のモデルを作成するための一部の実施形態では、プロセッサは、第1の確率分布に対する入力のために、試料のトレーニングセットにおいて1つまたは複数の生殖細胞系列 S N P のヘテロ接合性の出現率 (p r e v a l e n c e) をさらに判定することができる。

30

【0018】

一部の実施形態では、試料のトレーニングセットは、腫瘍が検出されない (T N D) 複数の試料を含んでもよい。

【0019】

一部の実施形態では、対立遺伝子数の第1のモデルを作成するために、プロセッサは、第1の確率分布に対する入力のために、試料のトレーニングセットにおいて1つまたは複数の生殖細胞系列 S N P のそれぞれに関連するマイナー対立遺伝子頻度 (M A F) の標準偏差をさらに判定することができる。

【0020】

一部の実施形態では、第1のモデルを作成するために、プロセッサは、第1の確率分布に対する入力のために、変異型対立遺伝子を支持する試料中の分子の数をさらに判定することができる。

40

【0021】

一部の実施形態では、第1のモデルを作成するために、プロセッサは、第1の確率分布に対する入力のために、試料中の分子の総数をさらに判定することができる。

【0022】

一部の実施形態では、第1のモデルを作成するために、プロセッサは、体細胞ホモ接合欠失と関連する配列読み取りデータカバレッジに基づいて、体細胞ホモ接合欠失を仮定して、試料中の1つまたは複数の生殖細胞系列 S N P の位置の対立遺伝子数の第1の尤度をさらに計算することができる。

50

【0023】

一部の実施形態では、第2のモデルを作成するために、プロセッサは、体細胞ヘテロ接合欠失と関連する配列読み取りデータカバレッジに基づいて、体細胞ヘテロ接合欠失を仮定して、試料中の1つまたは複数の生殖細胞系列SNPの位置の対立遺伝子数の第2の尤度をさらに計算することができる。

【0024】

一部の実施形態では、第2のモデルを作成するために、プロセッサは、第2のモデルに関する第2の確率分布に対する入力のために、試料から推定された腫瘍含有率 (tumor fraction) の平均をさらに判定することができる。

【0025】

一部の実施形態では、腫瘍含有率は、配列カバレッジ情報に基づいて推定され得る。

【0026】

一部の実施形態では、第2のモデルを作成するために、プロセッサは、第2のモデルに関する第2の確率分布に対する入力のために、試料から推定された腫瘍含有率の標準偏差をさらに判定することができる。

【0027】

一部の実施形態では、プロセッサは、さらに、複数の試料にアクセスし、生殖細胞系列欠失を含む複数の試料の中から試料のセットを特定し、複数の試料から試料のセットをフィルタリングして出力し、フィルタリングされた複数の試料の中から、体細胞ホモ接合欠失または体細胞ヘテロ接合欠失の存在を特定することができる。

【0028】

一部の実施形態では、第1の出力は、体細胞ホモ接合欠失の存在についての第1の確率を含んでもよく、第2の出力は、体細胞ヘテロ接合欠失の存在についての第2の確率を含んでもよい。

【0029】

一部の実施形態では、第1のモデルの第1の出力と第2のモデルの第2の出力とを比較するために、プロセッサは、第1の出力と第2の出力とに基づいて、対数尤度関数をさらに実行することができる。

【0030】

一部の実施形態では、遺伝子は、BRCA1、BRCA2、またはATMを含んでもよい。

【0031】

別の態様では、本開示は、システムに関する。システムは、試料中の遺伝子が体細胞ホモ接合欠失を含む第1の確率を作成し、試料中の遺伝子が体細胞ヘテロ接合欠失を含む第2の確率を作成し、第1の確率と第2の確率とを比較し、試料が体細胞ホモ接合欠失または体細胞ヘテロ接合欠失を含むか否かの予測を作成するようにプログラミングされたプロセッサを含んでもよい。

【0032】

別の態様では、本開示は、システムに関する。システムは、試料中の遺伝物質が第1の状態を含む第1の確率を作成し、試料中の遺伝物質が第2の状態を含む第2の確率を作成し、第1の確率と第2の確率とを比較し、試料が第1の状態または第2の状態を含むか否かの予測を作成するようにプログラミングされたプロセッサを含んでもよい。

【0033】

一部の実施形態では、第1の状態は体細胞ホモ接合欠失を含み、第2の状態は体細胞ヘテロ接合欠失を含む。

【0034】

一部の実施形態では、第1の状態は、第1のコピー数バリエーション (CNV) を含んでもよく、第2の状態は、第1のCNVとは異なる第2のCNVを含んでもよい。

【0035】

一部の実施形態では、第1のCNVおよび／または第2のCNVは、有害な状態に関連

10

20

30

40

50

し得る。

【0036】

一部の実施形態では、第1の確率を作成するために、プロセッサは、さらに、遺伝子に関連する1つまたは複数の生殖細胞系列一塩基多型（SNP）の位置にアクセスし、試料のトレーニングセットにおいて1つまたは複数の生殖細胞系列SNPのそれぞれに関連するマイナー対立遺伝子頻度（MAF）の標準偏差を判定することができる。

【0037】

一部の実施形態では、第1の確率を作成するために、プロセッサは、確率分布に対する入力のために、試料のトレーニングセットにおいて1つまたは複数の生殖細胞系列SNPのそれぞれと関連するマイナー対立遺伝子頻度（MAF）の標準偏差をさらに判定することができる。

10

【0038】

別の態様では、本開示は、プロセッサによって実装される方法に関する。本方法は、プロセッサによって、第1の確率分布により、遺伝子に関連する1つまたは複数の生殖細胞系列一塩基多型（SNP）の位置に基づいて、体細胞ホモ接合欠失を表す、対立遺伝子数の第1のモデルを作成するステップを含んでもよい。本方法は、プロセッサによって、第2の確率分布により、1つまたは複数の生殖細胞系列SNPの位置に基づいて、体細胞ヘテロ接合欠失を表す、試料中の対立遺伝子数の第2のモデルを作成するステップをさらに含んでもよい。本方法は、プロセッサによって、第1のモデルの第1の出力と第2のモデルの第2の出力とを比較するステップを含んでもよい。本方法は、プロセッサによって、比較に基づいて、遺伝子に関する体細胞ホモ接合欠失が試料中に存在するという予測を作成するステップをさらに含んでもよい。

20

【0039】

別の態様では、本開示は、プロセッサによって実装される別の方法に関する。本方法は、プロセッサによって、試料中の遺伝子が体細胞ホモ接合欠失を含む第1の確率を作成するステップを含んでもよい。本方法は、プロセッサによって、試料中の遺伝子が体細胞ヘテロ接合欠失を含む第2の確率を作成するステップをさらに含んでもよい。本方法は、プロセッサによって、第1の確率と第2の確率とを比較するステップをさらに含んでもよい。本方法は、プロセッサによって、試料が体細胞ホモ接合欠失または体細胞ヘテロ接合欠失を含むか否かの予測を作成するステップをさらに含んでもよい。

30

【0040】

別の態様では、本開示は、プロセッサによって実装される別の方法に関する。

【0041】

本方法は、プロセッサによって、試料中の遺伝物質が第1の状態を含む第1の確率を作成するステップを含んでもよい。本方法は、プロセッサによって、試料中の遺伝物質が第2の状態を含む第2の確率を作成するステップをさらに含んでもよい。本方法は、プロセッサによって、第1の確率と第2の確率とを比較するステップをさらに含んでもよい。本方法は、プロセッサによって、試料が第1の状態または第2の状態を含むか否かの予測を作成するステップをさらに含んでもよい。

【0042】

40

別の態様では、本開示は、本明細書の開示に基づいて、体細胞ホモ接合欠失を有すると判定された対象に、体細胞ホモ接合欠失に関連するがんを処置するのに有効な治療介入を施与するための方法に関する。

【0043】

一部の実施形態では、治療介入は、ポリADPリボースポリメラーゼ（PARP）阻害剤を含んでもよい。PARP阻害剤の例としては、とりわけ、OLAPARIB、TALAZOPARIB、RUCAPARIB、NIRAPARIB（商標名ZEUJULA）が挙げられる。

【0044】

一部の実施形態では、治療介入は、塩基除去修復（BER）阻害剤を含んでもよい。例

50

えば、OLAPARIBは、BERを阻害し得る。

【0045】

別の態様では、本開示は、本明細書の開示に基づいて、特定の状態の遺伝物質を有すると判定された対象に、遺伝物質の状態に関連する疾患を処置するのに有効な治療介入を施与するための方法に関する。

【0046】

別の態様では、本開示は、本明細書の開示に基づいて、体細胞ホモ接合欠失を有しないと判定された対象に、PARP阻害剤を除外するための治療介入を施与するための方法に関する。

【0047】

本開示のすべての態様の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／または方法の結果は、レポートを作成するための入力として使用される。レポートは、紙であっても電子形式であってもよい。例えば、本明細書に開示される方法またはシステムによって判定される遺伝子および／または遺伝物質の欠失または他の状態に関する情報、および／またはそれに由来する情報を、このようなレポートにおいて示すことができる。本明細書に開示される方法またはシステムは、レポートを、試料が由来する対象または医療従事者などの第3のパーティーに通信するステップをさらに含んでもよい。

本明細書に開示される方法の様々な操作、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操作は、同時にもしくは異なる時間に、および／または同じ地理的場所もしくは異なる地理的場所、例えば、国において実行されてもよい。本明細書に開示される方法の様々なステップは、同じ人間または異なる人々によって実施され得る。

【図面の簡単な説明】

【0048】

【図1】図1は、本開示の実施形態に従って、各状態の確率に基づいて、遺伝物質の状態を予測するためにモデルをトレーニングするためのシステムの例を例示する。

【0049】

【図2】図2は、本開示の実施形態に従って、生殖細胞系列SNPに関する対立遺伝子数を判定して、遺伝子の状態を予測する概略図を例示する。

【0050】

【図3】図3は、本開示の実施形態に従って、トレーニングされたモデルに基づいて、体細胞ホモ接合またはヘテロ接合欠失を予測するプロセスを例示する。

【0051】

【図4】図4は、本開示の実施形態に従って、トレーニングされたモデルに基づいて、遺伝物質の状態を予測するプロセスを例示する。

【0052】

【図5】図5は、本開示の実施形態に従って、体細胞欠失の種類を例示する。

【0053】

【図6A】図6Aは、本開示の実施形態に従って、BRCA1ホモ接合欠失の例示的プロットを例示する。

【0054】

【図6B】図6Bは、本開示の実施形態に従って、BRCA2ヘテロ接合欠失の例示的プロットを例示する。

【0055】

【図7A】図7Aは、本開示の実施形態に従って、TND試料中のhet出現率の例示的プロットを例示する。

【0056】

【図7B】図7Bは、本開示の実施形態に従って、TND試料にわたるMAFの例示的プロットを例示する。

【0057】

【図8A】図8Aは、本開示の実施形態に従って、BRCA1に関するMAF値の例示的

10

20

30

40

50

プロットを例示する。

【0058】

【図8B】図8Bは、本開示の実施形態に従って、BRCA2に関するMAF値の例示的プロットを例示する。

【0059】

【図9A】図9Aは、本開示の実施形態に従って、BRCA2パネルに関するベーター二項モデルと二項モデルのスコア比較の例示的プロットを例示する。

【0060】

【図9B】図9Bは、本開示の実施形態に従って、BRCA2パネルに関するベーター二項モデルとガウス分布モデルのスコア比較の例示的プロットを例示する。

10

【0061】

【図10A】図10Aは、本開示の実施形態に従って、BRCA1陰性試料に関するLLRスコア分布の例示的プロットを例示する。

【0062】

【図10B】図10Bは、本開示の実施形態に従って、BRCA2陰性試料に関するLLRスコア分布の例示的プロットを例示する。

【0063】

【図11A】図11Aは、本開示の実施形態に従って、BRCA1に関する検出限界（LOD）の欠失の例示的プロットを例示する。

【0064】

【図11B】図11Bは、本開示の実施形態に従って、BRCA1に関するLODのHRRホモ接合欠失の例示的プロットを例示する。

20

【0065】

【図12A】図12Aは、本開示の実施形態に従って、BRCA2に関するLODの欠失の例示的プロットを例示する。

【0066】

【図12B】図12Bは、本開示の実施形態に従って、BRCA2に関するLODのHRRホモ接合欠失の例示的プロットを例示する。

【0067】

【図13】図13は、本開示の実施形態に従って、がんの種類に対するTFの出現率の例示的プロットを例示する。

30

【0068】

【図14】図14は、本開示の実施形態に従って、BRCA1およびBRCA2に関するLLRスコアの密度の例示的プロットを例示する。

【0069】

【図15】図15は、本開示の実施形態に従って、BRCA2ホモ接合欠失の出現率の例示的チャートを例示する。

【0070】

【図16】図16は、本開示の実施形態に従って、BRCA1ホモ接合欠失の出現率の例示的チャートを例示する。

40

【0071】

【図17】図17は、本開示の実施形態に従って、BRCA2のホモ接合欠失と潜在的な臨床的対処可能性（clinical actionability）の例を例示する。

【0072】

【図18A】図18Aは、本開示の実施形態に従って、BRCA1ホモ接合欠失の例示的プロットを例示する。

【0073】

【図18B】図18Bは、本開示の実施形態に従って、BRCA1ホモ接合欠失の例示的プロットを例示する。

【0074】

50

【図19A】図19Aは、本開示の実施形態に従って、BRCA2ホモ接合欠失の例示的プロットを例示する。

【0075】

【図19B】図19Bは、本開示の実施形態に従って、BRCA2ホモ接合欠失の例示的プロットを例示する。

【0076】

【図20A】図20Aは、本開示の実施形態に従って、BRCA1二対立遺伝子体細胞コピー数の喪失の例示的プロットを例示する。

【0077】

【図20B】図20Bは、本開示の実施形態に従って、BRCA1 LOHの例示的プロットを例示する。

【0078】

【図21A】図21Aは、本開示の実施形態に従って、BRCA2二対立遺伝子体細胞コピー数の喪失の例示的プロットを例示する。

【0079】

【図21B】図21Bは、本開示の実施形態に従って、BRCA2 LOHの例示的プロットを例示する。

【0080】

【図22】図22は、本開示の実施形態に従って、BRCA1およびBRCA2体細胞欠失の出現率のプロットを例示する。

【発明を実施するための形態】

【0081】

定義

対象は、動物、例えば、哺乳類種（好ましくはヒト）または鳥類（例えば、トリ）種、または他の生物、例えば、植物を指す。より詳細には、対象は、脊椎動物、例えば、マウス、霊長類、類人猿またはヒトなどの哺乳類であってもよい。動物は、飼育動物、狩猟動物、および愛玩動物を含む。対象は、健康な個体、症状もしくは兆候を有するかまたは疾患もしくは疾患に罹り易い体質を有することが疑われる個体、あるいは治療を必要としているかまたは治療を必要とすることが疑われる個体であってもよい。

【0082】

遺伝子バリエーションは、対象の核酸試料またはゲノムにおける変更、バリエーションまたは多型を指す。このような変更、バリエーションまたは多型は、種（例えば、ヒトでは、hg19またはhg38）、対象または他の個体の参照ゲノムであり得る参照ゲノムに関するものであってもよい。変異は、1つまたは複数の塩基変異（SNV）、挿入、欠失、反復、小規模の挿入、小規模の欠失、小規模の反復、構造的バリエーション接合、可変長タンDEM反復、および／または隣接配列を含み、コピー数バリエーション（CNV）、トランスバージョン、遺伝子融合および他の再配列は遺伝的変異の形態でもある。変異は、塩基の変化、挿入、欠失、反復、コピー数変異、トランスバージョン、またはこれらの組合せであってもよい。

【0083】

がんマーカーは、がんの存在またはがんを発症するリスクに関連する遺伝子バリエーションである。がんマーカーは、対象が、がんを有するか、またはがんマーカーを有さない同種の年齢および性別の一致した対象よりもがんを発症するより高いリスクを有することの指標を提供し得る。がんマーカーは、がんの原因であってもなくてもよい。

【0084】

本明細書で使用される場合、「核酸タグ」は、異なる種類の、または異なる処理を受けた、異なる試料（例えば、試料インデックスを表す）、または同じ試料における異なる核酸分子（例えば、分子バーコードを表す）から、核酸を識別するために使用される短い核酸（例えば、約500ヌクレオチド長、約100ヌクレオチド長、約50ヌクレオチド長、または約10ヌクレオチド長未満）を指す。核酸タグは、所定の、固定された、非ラン

10

20

30

40

50

ダム、ランダムまたはセミランダムオリゴヌクレオチド配列を含む。このような核酸タグを使用して、異なる核酸分子または異なる核酸試料もしくはサブ試料を標識することができる。核酸タグは、一本鎖、二本鎖、または少なくとも部分的に二本鎖であってもよい。核酸タグは、必要に応じて、同じ長さまたは変動した長さを有する。核酸タグは、1つもしくは複数の平滑末端を有する二本鎖分子を含んでもよい、5'もしくは3'一本鎖領域（例えば、オーバーハング）を含んでもよい、および／または所与の分子内の他の場所に1つもしくは複数の他の一本鎖領域を含んでもよい。核酸タグは、他の核酸（例えば、増幅および／または配列決定される試料核酸）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、所与の核酸の起源の試料、形態、またはプロセッシングなどの情報を明らかにするために解釈され得る。例えば、その後、核酸タグを検出する（例えば、読み取る）ことによって核酸がデコンボリューションされる、異なる分子バーコードおよび／または試料インデックスを保有する核酸を含む複数の試料のプールおよび／または並行処理を可能とするために核酸タグを使用することもできる。核酸タグを、識別子（例えば、分子識別子、試料識別子）とも称することができる。さらに、またはあるいは、核酸タグは、分子識別子として使用することができる（例えば、同じ試料またはサブ試料中の異なる分子または異なる親分子のアンプリコン間を識別するために）。これは、例えば、所与の試料中の異なる核酸分子を固有にタグ付けすること、またはこのような分子を非固有にタグ付けすることを含む。非固有のタグ付けを適用する事例では、異なる分子を、少なくとも1種の分子バーコードと組み合わせて、それらの内在性配列情報（例えば、選択された参照ゲノムに対してマッピングされる開始および／もしくは停止の位置、配列の一方もしくは両方の末端のサブ配列、ならびに／または配列の長さ）に基づいて識別することができるように、限定数のタグ（すなわち、分子バーコード）を使用して、各核酸分子をタグ付けすることができる。典型的には、いずれか2つの分子が、同じ内在性配列情報（例えば、開始および／もしくは停止の位置、配列の一方もしくは両方の末端のサブ配列、ならびに／または長さ）を有する可能性があり、かつ同じ分子バーコードも有する確率が低くなるように（例えば、約10%未満、約5%未満、約1%未満、または約0.1%未満の見込み）、十分な数の異なる分子バーコードが使用される。

【0085】

アダプターは、試料核酸分子のいずれかまたは両方の末端（各末端に1つのアダプター）への連結のために、通常、少なくとも部分的に二本鎖の短い核酸（例えば、500、1000または50ヌクレオチド長未満）である。アダプターは、両末端がアダプターに隣接する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／または次世代配列決定（NGS）のためのプライマー結合部位を含む配列決定プライマー結合部位を含むことができる。アダプターはまた、捕捉プローブ、例えば、フローセルの支持体に結合したオリゴヌクレオチドに対する結合部位を含むことができる。アダプターは、上述のようなバーコードも含み得る。バーコードは、好ましくは、バーコードが、核酸分子のアンプリコンおよび配列決定読み取りデータに含まれるように、プライマーおよび配列決定プライマー結合部位に対して位置付けられる。同じまたは異なる配列のアダプターは、核酸分子の各末端に連結され得る。同じアダプターは、バーコードが異なることを除いて、各末端に連結されることが多い。好ましいアダプターは、核酸分子（これ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でテール処理されている）への接合のために、一方の末端が本明細書に記載されるように平滑末端であるかまたはテール処理されているY形状アダプターであり、Y形状アダプターの他方の末端は、二本鎖を形成するためにハイブリダイズしない非相補配列を含む。別の好ましいアダプターは、同様に、解析される核酸への接合のための平滑末端またはテール処理された末端を有するベル形状アダプターである。

【0086】

本明細書で使用される場合、用語「配列決定」は、生体分子、例えば、DNAまたはRNAなどの核酸の配列を決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。例示的な配列決定方法としては、以下に限定されないが、標的化配列決定、単一分子

リアルタイム配列決定、エクソン配列決定、電子顕微鏡に基づく配列決定、パネル配列決定、トランジスタ媒介性配列決定、直接的配列決定、ランダムショットガン配列決定、サンガーデオキシ停止配列決定、全ゲノム配列決定、ハイブリダイゼーションによる配列決定、ピロ配列決定、キャピラリー電気泳動、二重鎖配列決定、サイクル配列決定、単一塩基伸長配列決定、固相配列決定、ハイスループット配列決定、超並列シグネチャー配列決定、エマルジョンPCR、低変性温度PCR（COLD-PCR）における同時増幅、マルチプレックスPCR、可逆的ダイターミネーターによる配列決定、ペアードエンド配列決定、ニアターム（near-term）配列決定、エキソヌクレアーゼ配列決定、ライゲーションによる配列決定、ショートリード配列決定、単一分子配列決定、合成による配列決定、リアルタイム配列決定、リバースターミネーター配列決定、ナノポア配列決定、454配列決定、Solexa Genome Analyzer配列決定、SOLID（商標）配列決定、MS-PET配列決定、およびこれらの組合せが挙げられる。一部の実施形態では、配列決定は、例えば、IlluminaまたはApplied Biosystemsから市販されている遺伝子解析器などの遺伝子解析器によって実施することができる。

【0087】

語句「次世代配列決定」またはNGSは、例えば、一度に数十万の比較的小さな配列読み取りデータを生成する能力を有する、古典的なサンガー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する配列決定技術を指す。次世代配列決定技法のいくつかの例として、以下に限定されないが、合成による配列決定、ライゲーションによる配列決定、およびハイブリダイゼーションによる配列決定が挙げられる。

【0088】

DNA（デオキシリボ核酸）は、アデニン（A）、チミン（T）、シトシン（C）、およびグアニン（G）に基づく4種類のヌクレオチドを含むヌクレオチド鎖である。RNA（リボ核酸）は、A、ウラシル（U）、G、およびCに基づく4種類のヌクレオチドを含むヌクレオチド鎖である。ある特定のヌクレオチドの対は、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では、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では、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、その第1の鎖におけるヌクレオチドに対して相補的であるヌクレオチドから構成される第2の核酸鎖に結合する場合、この2本の鎖が結合して二本鎖を形成する。本明細書で使用される場合、「核酸配列決定データ」、「核酸配列決定情報」、「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「核酸配列決定読み取りデータ」は、DNAまたはRNAなどの核酸の分子（例えば、全ゲノム、全トランスクリプトーム、エキソーム、オリゴヌクレオチド、ポリヌクレオチド、または断片）中のヌクレオチド塩基（例えば、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）の順序を示すいずれかの情報またはデータを示す。本発明の教示が、以下に限定されないが：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特定システム、ピロ配列決定、イオンまたはpHに基づく検出システム、および電氣的なサインに基づくシステムを含む、すべての利用可能な多種の技法、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が理解されるべきである。

【0089】

「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、ヌクレオシド間の連結によって接合したヌクレオシド（デオキシリボヌクレオシド、リボヌクレオシド、またはこれらのアナログを含む）の線状ポリマーを指す。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、数個のモノマー単位、例えば、3～4個から数百個のモノマー単位のサイズの範囲に及ぶこと

度を指す。

【0098】

「腫瘍含有率」(TF)は、所与の試料中の腫瘍に関連するDNA分子の割合を指す。TFは、腫瘍細胞におけるバリエーション対立遺伝子のカバレッジの低下を検出することに基づいて導出され得る。所与の試料中のTFがより低いことは、所与の試料中の所与のバリエーション対立遺伝子のMAF、したがって、所与のバリエーション対立遺伝子の検出能に影響を及ぼす可能性がある。

【0099】

用語「腫瘍非検出」または「TND」は、体細胞一塩基バリエーション、挿入-欠失、コピー数バリエーション、および融合のいずれもが検出されなかった試料を指す。

【0100】

用語「処理すること」、「計算すること」、および「比較すること」は、交換可能に使用することができる。この用語は、差異、例えば、数または配列における差異を判定することを指し得る。例えば、遺伝子発現、コピー数変異(CNV)、インデル、および/または一塩基バリエーション(SNV)の値または配列を処理することができる。

【0101】

アダプターは、ライゲーション、ハイブリダイゼーション、および/または増幅を含む任意のアプローチによって、核酸分子またはポリヌクレオチド配列に連結され得る、人工的に合成された配列である。アダプターは、試料核酸分子のいずれかまたは両方の末端への連結のために、通常、少なくとも部分的に二本鎖の短い核酸(例えば、500、100または50ヌクレオチド長未満)である。アダプターは、両末端がアダプターに隣接する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/または次世代配列決定(NGS)のためのプライマー結合部位を含む配列決定プライマー結合部位を含むことができる。アダプターはまた、捕捉プローブ、例えば、フローセルの支持体に結合したオリゴヌクレオチドに対する結合部位を含むことができる。アダプターは、上述のようなバーコードも含み得る。タグは、好ましくは、タグが核酸分子のアンプリコンおよび配列決定読み取りデータに含まれるように、プライマーおよび配列決定プライマー結合部位に対する位置である。同じかまたは異なるアダプターは、核酸分子の各末端に連結され得る。同じアダプターは、タグが異なることを除いて、各末端に連結されることが多い。好ましいアダプターは、核酸分子(これ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でテール処理されている)への接合のために、一方の末端が本明細書に記載される平滑末端であるかまたはテール処理されているY形状アダプターである。別の好ましいアダプターは、同様に、解析される核酸への接合のための平滑末端またはテール処理された末端を有するベル形状アダプターである。

詳細な説明

【0102】

図1は、本開示の実施形態に従って、各状態の確率に基づいて、遺伝物質の状態を予測するためにコンピューターモデルをトレーニングおよび使用するためのシステム100の例を示す。システムは、試料101を処理して、1つまたは複数のモデル140(モデル140a...nとして例示される)をトレーニングすることができ、モデル140はそれぞれ、検査(IUT)111の下で、個体由来の試料などの遺伝物質が特定の状態にある確率を出力する。一部の例では、試料101は、研究される目的の様々な遺伝子のパネルを含むことができる。

【0103】

例えば、システムは、モデル140aを使用して、IUT111からの試料が、遺伝子に関連する体細胞ホモ接合欠失を含む確率を判定することができる。システムは、別のモデル140bを使用して、IUT111からの試料が、遺伝子に関連する体細胞ヘテロ接合欠失を含む確率を判定することができる。次いで、システムは、確率を互いに比較して、体細胞ホモ接合欠失または体細胞ヘテロ接合欠失のどちらがより起こりやすいかを判定することができる。システムは、同様に、競合する確率に基づいて、他の種類の精度の高

10

20

30

40

50

い診断を提供することができる。例えば、システムは、異なるコピー数の確率をモデリングすることによって、CNVをモデリングすることができる。各モデルの出力確率（それぞれ、異なるコピー数予測に対応し得る）の比較に基づいて、システムは、IUT111の試料中のCNVを判定することができる。

【0104】

システム100は、配列決定システム102、コンピューターシステム110、および／または他のコンポーネントを含むことができる。配列決定システム102とコンピューターシステム110とは、互いに隔離されていてもよく、コンピューターネットワーク（例示されていない）を介して互いに接続されていてもよいことに留意されたい。配列決定システム102は、試料採取および調製パイプライン103、配列決定パイプライン105、および配列決定読み取りデータ記憶109、および／または他のコンポーネントを含むことができる。配列決定パイプライン105は、1つまたは複数の配列決定デバイス107（配列決定デバイス107a・・・nとして図1において例示されている）を含むことができる。

【0105】

コンピューターシステム110は、配列解析パイプライン112、プロセッサ120、ストレージデバイス122、データ前処理サブシステム124、分類装置130、モデルバリデーター132、および／または他のコンポーネントを含むことができる。

【0106】

配列解析パイプライン112は、配列品質管理（QC）コンポーネント113、アライメントコンポーネント114、他の解析コンポーネント115、および解析QCコンポーネント116を含むことができる。配列解析パイプライン112からの出力は、解析データ記憶117に記憶され得る。データ前処理サブシステム124は、配列解析パイプライン112からのデータを前処理して、トレーニングデータセット125を作成することができる。例えば、トレーニングデータセット125は、腫瘍が検出されなかったデータ（「TND」）（がんが診断される場合）またはそうでなければ、試料101の中からの正常な試料（他の種類の疾患または状態が診断される場合）のデータを含むことができる。全体を通して開示された例は、例示のためのTND試料を指し得る。

【0107】

一部の実施形態では、トレーニングデータセット125は、トレーニングデータ記憶126に記憶され得る。プロセッサ120の例示的操作を示すために、図2が参照されよう。図2は、本開示の実施形態に従って、生殖細胞系列SNPに関する対立遺伝子数を判定して、遺伝子201の状態を予測する概略図200を示す。一部の例では、プロセッサ120は、遺伝子201周囲の目的の領域201に関するTND試料の生殖細胞系列SNPを特定することができる。1つのトレーニング例では、生殖細胞系列SNPを28,199個の試料から選択した。これらの試料のうち、5105個の試料（18%）がTNDを有するものとして特定され、これらを、集団対立遺伝子／遺伝子型頻度に対して使用した。以下の条件を満たす生殖細胞系列SNPが選択された：（1）選択された遺伝子（BRCA1、BRCA2、ATMなど）から3Mb以内にある、（2）5105個のTND試料にわたるヘテロ接合性コールの頻度（MAF>25%およびMAF<75%）が5%から95%の間である、および（3）28,199個の試料すべてにおいて、バリエーションが体細胞でコールされない。目的の領域203は、遺伝子201の開始から上流のN個の塩基および遺伝子201の末端から下流のM個の塩基を含むことができる。NおよびMの値は、同じであっても異なってもよい。一部の例では、NおよびMは、それぞれ、3,000,000ヌクレオチド（3Mb）であってもよい。

【0108】

図2の例示された例では、SNP部位（i）（SNP（i）として例示される）における参照野生型ヌクレオチドは、「G」であってもよい。TND試料にわたって、その位置でコールされるヌクレオチドは、互いに異なってもよい。TND試料が正常であるため、プロセッサ120は、TND試料のSNP（i）と他のSNPの部位は、腫瘍また

10

20

30

40

50

は非正常状態の原因ではないと仮定し得る。このように、これらのSNP部位は、それぞれ、遺伝子の状態の確率的判定のために対立遺伝子数をモデリングするための参照期待値としての役割を果たすことができる。例えば、各SNPの位置において観察されたヌクレオチドコールからの偏差は、このような偏差が、遺伝子201の腫瘍または他の非正常状態のような特定の状態の原因である確率を示し得る。したがって、プロセッサ120は、TND試料の生殖細胞系列SNPからのデータに関する計算に由来する期待値に基づいて、モデル140をトレーニングすることができる。このような計算されたデータは、各SNP部位について、ヘテロ接合性の出現率、マイナー対立遺伝子頻度(MAF)の標準偏差、遺伝子型、生殖細胞系列の出現率(以前の)、および／または他のデータを含むことができる。

10

【0109】

計算された期待値を用いて、プロセッサ120は、目的の領域203に対してアラインするIUT111の試料の配列読み取りデータに基づいて、遺伝子201の状態をモデリングすることができる。例えば、プロセッサ120は、遺伝子201が特定の状態にある確率を表す各スコアをそれぞれ出力する競合モデル140を作成することができる。プロセッサ120は、各スコアを比較して予測スコアを計算し、これを閾値スコアと比較して、遺伝子201の状態を判定することができる。プロセッサ120は、以下にさらに記載されるように、トレーニング試料から観察されたデータに基づいて、閾値スコアを計算することができる。

【0110】

20

一部の実施形態では、IUT111の試料から作成した分子の配列読み取りデータを参照ゲノムに対してアラインし、元の分子が支持する対立遺伝子(突然変異体または野生型)を特定することができる。IUT111の試料は、試料採取および調製パイプライン103で調製され、配列決定パイプライン105で配列決定され得る。各分子は、配列読み取りデータに関連し得る。IUT111からの試料の分子の数からの配列決定読み取りデータの数、所与の生殖細胞系列SNP部位をカバーすることができる。

【0111】

IUT111の試料から作成した配列読み取りデータのアライメントに基づき、プロセッサ120は、SNP対立遺伝子を支持する分子の数を特定し、分子の総数を計算することができる。プロセッサ120は、目的の領域203における生殖細胞系列SNPのそれぞれから計算された期待値データを用いて、IUT111の試料からの、これらのおよび／または他のデータをモデリングすることができる。例えば、プロセッサ120は、遺伝子201の第1の状態の確率を表す対立遺伝子数のモデル140aの第1の出力および遺伝子201の第2の状態の確率を表す対立遺伝子数のモデル140bの第2の出力を作成することができる。

30

【0112】

プロセッサ120は、異なる種類の確率分布を実装して、モデル140を作成することができる。さらに、モデル140は、遺伝子201の様々な種類の状態、またはより一般的には、遺伝物質の様々な種類の状態をモデリングすることができる。ここで、モデリングの例およびプロセッサ120によってモデリングされる状態の種類に注意が向けられる。

40

【0113】

一般的に言えば、プロセッサ120は、分類装置130を実装し得る(それによってプログラミングされ得る)。あるいは、分類装置130はハードウェアモジュールを含むことができることに留意されたい。いずれかの事象では、分類装置130(プロセッサ120をプログラムすることができる)は、遺伝子に関連する目的の領域において検出された対立遺伝子に基づいて、遺伝子(図2に例示された遺伝子201など)の状態をモデリングすることができる。より詳細には、トレーニングデータセット125に基づいて、分類装置130は、目的の領域(図2に例示された目的の領域203など)における生殖細胞系列一塩基多型(SNP)の位置に基づいて、遺伝子の特定の状態についての1つま

50

たは複数の確率モデル 140（モデル 140 a、140 b、・・・、140 nとして例示される）を判定することができる。各モデル 140 は、遺伝子の状態の各確率に対応し得る。SNP の位置は、様々な試料 101 から配列決定システム 102 によって作成した配列決定読み取りデータに基づいてもよい。状態は、限定されないが、体細胞ホモ接合欠失、体細胞ヘテロ接合欠失、コピー数変異（「CNV」）（特定コピー数の野生型、獲得、または喪失を含む）、および／または遺伝子の他の状態などの遺伝子の突然変異状態を含むことができる。

【0114】

様々な実施形態では、分類装置 130 は、トレーニングデータセット 125 に基づいて作成され得るモデルを適用して、個体の試料中の遺伝子の状態を判定することができる。例えば、分類装置 130 は、個体由来の cfDNA の試料中の遺伝子が、体細胞ホモ接合欠失、体細胞ヘテロ接合欠失、および／またはがんなどの疾患もしくは他の健康状態と相関し得る他の状態を含む確率を判定することができる。予測に基づいて、精度の高い処置を個体に適合させることができる。このように、コンピューターシステム 110 を改良して、cfDNA などの遺伝物質の非侵襲的解析に基づいて、進歩した診断能力を提供することができる。

【0115】

本明細書に記載の例は、遺伝子状態を判定することに関係し得るが、染色体、エキソーム、および／または他の遺伝物質などの他の遺伝物質の状態も同様に判定することができることに留意されたい。例えば、CNV は、染色体、エキソーム、および／または他の遺伝物質に対して判定され得る。分類装置 130 の機能性についての説明を提供したが、ここで、様々なモデル 140 をトレーニングし、モデル 140 を使用して、検査下の特定の試料が特定の遺伝子状態を示す確率を予測することによって、遺伝子の状態を判定するより詳細な例に注意が向けられる。

【0116】

TND 試料に基づくモデルのトレーニング

【0117】

一部の実施形態では、分類装置 130 は、試料 101 からのデータを使用する。データは、腫瘍が検出されない試料（「TND」試料）のセットを含むことができる。分類装置 130 は、TND 試料を使用して、TND 試料中の各生殖細胞系列 SNP のヘテロ接合性の出現率および TND 試料中の各生殖細胞系列 SNP のマイナー対立遺伝子頻度（MAF）の標準偏差を判定することができる。本開示全体を通して記載される式および計算において、標準偏差の代わりに分散を使用するこのような計算に対して適当な調整がなされさえすれば、標準偏差の代わりに分散を使用することができることに留意されたい。ヘテロ接合性の出現率および標準偏差は、「正常な」試料、すなわち、疾患状態を示さない試料のベースライン期待値を与え得る。分類装置 130 は、各部位 i に関する生殖細胞系列の出現率（以前の） g_i を推定することもできる。各生殖細胞系列 SNP のヘテロ接合性の出現率の例示的計算は、等式（1）によって与えられ得る：

【数 1】

$$p_a(g_i) = P(b_{ij} = a | g_i) \quad (1),$$

（式中、

- $p_a(g_i)$ は、各生殖細胞系列 SNP のヘテロ接合性の出現率を表し、
- b_{ij} は、SNP 部位 i において観察された塩基のセットを表し、
- g_i は、SNP 部位 i における遺伝子型（AA / Aa / aa）を表す）。

【0118】

体細胞ホモ接合欠失のモデリング

【0119】

分類装置 130 は、確率分布により、遺伝子に関連する 1 つまたは複数の生殖細胞系列

一塩基多型（SNP）の位置に基づいて、対立遺伝子数についての第1のモデル140aを作成することができる。第1のモデルは、遺伝子の体細胞ホモ接合欠失を表す（例えば、モデリングする）ことができる。例えば、TND試料中の各生殖細胞系列SNPのヘテロ接合性の出現率（式（1））およびTND試料中の各生殖細胞系列SNPのMAFの標準偏差を考慮すると、分類装置130は、個体由来の特定の試料の遺伝子が、体細胞ホモ接合欠失に関連する確率をモデリングすることができる。そうするために、分類装置130は、IUT111の試料の遺伝子における体細胞ホモ接合欠失の存在を支持する分子の数およびIUT111の試料中の分子の総数にアクセスすることができる。例えば、分類装置130は、試料中の遺伝子が体細胞ホモ接合欠失を有する確率を表す、モデル140aなどのモデルを作成することができる。一部の実施形態では、分類装置130は、ベータ二項確率分布を使用して、モデル140aを作成することができるが、二項確率分布、正規分布（ガウス分布）、および／または他の確率モデリングなどの他の確率分布を使用してもよい。

10

【0120】

ベータ二項分布は、各試行における成功確率が固定されるが、ベータ分布から無作為に導かれるnベルヌーイ試行の二項分布である。ベータ二項分布は2つのパラメーター： α および β （分布の平均／標準偏差によって固有に判定される）を使用することができる。n=1である場合、分布はベルヌーイ分布に対して低下する。 $\alpha = \beta = 1$ では、0からnまでの別々の一様な分布である。

【0121】

二項分布は、二項ランダム変数の確率分布である。二項ランダム変数は、二項実験のN反復試行における成功回数である。二項分布は、以下の特性を有する：分布の平均（ μ_x ）は $n * P$ に等しい；分散は $n * P * (1 - P)$ によって与えられる；および標準偏差（ σ_x ）は式（2）によって与えられる：

20

【数2】

$$\sqrt{n * P * (1 - P)} \quad (2).$$

【0122】

正規分布は、正規方程式によって定義され得る：

30

【数3】

$$Y = \{ 1 / [\sigma * \text{sqrt}(2\pi)] \} * e^{-(x - \mu)^2 / 2\sigma^2} \quad (3),$$

（式中、

Xは、正規ランダム変数であり、

μ は、平均であり、

σ は、標準偏差であり、

π は、およそ3.14159であり、

eは、およそ2.71828である）。

40

【0123】

ベータ二項確率分布を適用する例は、ここで、例示目的で記載されることになる。当業者であれば、二項、正規、および／または他の確率分布を、本明細書の開示に基づいて同様に使用することができることを理解するであろう。ベータ二項確率分布では、分類装置130は、等式（4）に従って、R ProjectのVGAMパッケージにおけるdbetabinom関数を使用することができる：

【数4】

$$P(b_{ij} | g_i | TF) = P(b_{ij} | g_i) = \text{VGAM} :: \text{dbetabinom}(m_i, R_i, p_a(g_i), \text{sd}(g_i)) \quad (4),$$

50

(式中、

m_i は、SNP 部位 i において SNP 対立遺伝子を支持する分子の数を表し、

R_i は、分子の総数を表し、

$P_a(g_i)$ は、SNP 部位 i におけるヘテロ接合性の出現率を表し、

$sd(g_i)$ は、MAF の標準偏差を表す)。

【0124】

分類装置 130 は、等式 (5) に従って、第 1 のモデル 140 a の第 1 の確率出力 (L_1) を作成することができる：

【数 5】

$$L_1 = \text{prod}(\text{sum}_{g_i}(\text{prior}_{g_i} * \text{dbtabinom}(m_i, R_i, p_a(g_i), sd(g_i)))) \quad (5).$$

10

【0125】

体細胞ヘテロ接合欠失のモデリング

【0126】

分類装置 130 は、確率分布により、1 つまたは複数の生殖細胞系列 SNP の位置に基づいて、試料中の対立遺伝子数の第 2 のモデル 140 b を作成することができる。第 2 のモデル 140 b は、遺伝子の体細胞ヘテロ接合欠失を表す (例えば、モデリングする) ことができる。ヘテロ接合欠失の検出は TF によって影響を及ぼされる場合があるため、分類装置 130 は、IUT111 の試料中の読み取りデータのカバレッジ (配列読み取りデータカバレッジ) に基づいて、TF の平均、 $\mu.tf$ ($\mu.tf$ と表されてもよい)、および標準偏差、 $sd.tf$ ($\sigma.tf$ と表されてもよい) を判定することができる。

20

【0127】

一部の実施形態では、分類装置 130 は、ベータ二項確率分布を使用して、モデル 140 b を作成することができるが、二項確率分布、ガウス分布、および／または他の確率モデリングなどの他の確率分布を使用してもよい。

【0128】

ベータ二項確率分布では、分類装置 130 は、等式 (6) に従って、R Project の VGAM パッケージにおける dbetabinom 関数を使用してもよい：

【数 6】

$$P(b_{ij} | g_i | TF) = P(b_{ij} | g_i) = \text{VGAM} :: \text{dbetabinom}(m_i, R_i, \mu_i, sd_i) \quad (6),$$

30

(式中、

m_i は、SNP 部位 i において SNP 対立遺伝子を支持する分子の数を表し、

R_i は、分子の総数を表し、

μ_i は、IUT111 の試料について計算された TF の平均を表し、

sd_i は、IUT111 の試料について計算された TF の標準偏差を表す)。

【0129】

分類装置 130 は、等式 (5) に従って、第 2 のモデル 140 b の第 2 の確率出力 L_0 を作成することができる：

40

【数 7-1】

$$L_0 = \text{prod}(\text{sum}_{g_i}(\text{prior}_{g_i} * \text{dbtabinom}(m_i, R_i, m_i(g_i), sd_i(g_i)))) \quad (7),$$

式中、

【数 7-2】

50

$m_i(g_i) = p_a(g_i)$ ($g_i = AA$ または aa の場合);

$1 - m_{tf} * p_a(g_i = Aa) + m_{tf} * \max(p_a(g_i = aa), p_a(g_i = AA))$,

$sd_i(g_i) = sd(g_i)$, ($g_i = AA$ または aa の場合);

$var_prod(sd_{tf}, 1 - m_{tf}, sd(g_i = Aa), p_a(g_i = Aa)) +$

$var_prof(sd_{tf}, m_{tf}, sd(g_i = aa), p_a(g_i = aa))$

ここで、 $var_prod(va, ma, vb, mb) = va * vb + va * mb^2 + vb * ma^2$

10

【0130】

第1のモデル140aおよび第2のモデル140bは、第1のモデル140aおよび第2のモデル140bがそれぞれ確率を出力しさえすれば、同じ確率分布を使用する必要はないことに留意されたい。

【0131】

分類装置130は、第1のモデル140aの第1の確率出力と第2のモデル140bの第2の確率出力とを比較して、どの確率出力がより可能性が高いかを判定することができる。例えば、分類装置130は、体細胞ホモ接合欠失または体細胞ヘテロ接合欠失がより可能性が高いか否かを判定することができる。特定の例では、分類装置130は、対数尤度比(「LLR」)を使用して、第1の確率出力(体細胞ホモ接合欠失の確率)および第2の確率出力(体細胞ヘテロ接合欠失の確率)に基づいて、LLRスコアを作成することができる。一部の実施形態では、LLRスコアが閾値カットオフスコアを超えなければ、ヌル確率が拒絶されるように、第1または第2の確率出力のうちの一方をヌル確率として使用することができる。例えば、分類装置130は、LLRスコアを閾値カットオフスコアと比較して、第2の確率出力が拒絶されるべきか否かを判定することができる。言い換えれば、LLRスコアが閾値カットオフスコアを超える場合、分類装置130は、第1の確率出力が選択されるべきであることを判定することができる。この例では、分類装置130は、比較に基づいて、遺伝子に関する体細胞ホモ接合欠失が、IUT111の試料中に存在するという予測を作成することができる。

20

【0132】

一部の例では、エラーを軽減するために、モデル140Aまたは140Bを使用して、所与の遺伝子と重複する各SNPに関する試料の遺伝子型を判定することができる。生殖細胞系列SNPがヘテロ接合性であることが判定されなければ、所与の遺伝子は「コールなし」と標識されてもよく、体細胞ホモ接合またはヘテロ接合欠失は所与の遺伝子に関連しない。

30

【0133】

閾値スコアカットオフの学習

【0134】

一部の実施形態では、閾値カットオフスコアを、アッセイされる様々な遺伝子または他の遺伝物質に対してカスタマイズすることができる。例えば、BRCA1遺伝子は、BRCA2遺伝子と異なる閾値カットオフスコアに関連し得る。他の遺伝子も同様に、カスタム閾値カットオフスコアに関連し得る。これらの実施形態では、分類装置130をトレーニングして、閾値カットオフスコアを判定することができる。これらの実施形態の一部では、分類装置130をトレーニングして、特定の遺伝子に関する閾値カットオフスコアを判定することができる。例えば、TND試料から開始する体細胞ヘテロ接合欠失のシミュレーションを使用して、分類装置130をトレーニングすることができる。例えば、図10Aおよび10Bを参照して、ブランク限界(LoB)、または最大LLRスコアは、ホモ接合欠失が、BRCA1およびBRCA2陰性試料に対して存在しない場合に見出されることが期待される。図10Aおよび10Bを参照して、TND試料から開始する体細胞ヘテロ接合欠失の100,000の事例がシミュレーションされた。28,000個の試

40

50

料において観察されたTF分布を、BRCA1およびBRCA2に関するLOBを判定するためのTFとして使用した。例示したように、LLRスコアと比較するための閾値カットオフスコアは、BRCA1およびBRCA2に関して、それぞれ、20.1および0である。よって、BRCA1の体細胞欠失がIUT111の試料において観察され、IUT111の試料中のBRCA1に関するLLRスコアが20.1を超える場合、分類装置130は、体細胞欠失が体細胞ホモ接合欠失であることを予測することができる。同様に、BRCA1の体細胞欠失がIUT111の試料において観察され、IUT111の試料中のBRCA2に関するLLRスコアが0を超える場合、分類装置130は、体細胞欠失が体細胞ホモ接合欠失であることを予測することができる。他の遺伝子を同様にシミュレーションして、閾値カットオフスコアを判定することができることに留意されたい。

10

【0135】

一部の実施形態では、モデルバリデーター132は、シミュレーションされたおよび／または臨床データを使用して、モデル140の結果を検証することができる。例えば、モデルバリデーター132は、診断結果データ記憶150および／または臨床結果データ記憶160を調べて、予測を検証することができる。シミュレーションされた結果では、公知の試料のパネルをモデリングして、これらの試料の遺伝物質の状態の予測を作成することができる。これらの結果を使用して、以前の予測および／または将来の予測の結果を検証することができる。

【0136】

図3は、本開示の実施形態に従って、トレーニングされたモデルに基づいて体細胞ホモ接合またはヘテロ接合欠失を予測するプロセス300を例示する。本明細書に記載の方法を実行するためには種々の方法が存在するため、プロセス300は、例として提供される。プロセス300は、主に、図1に例示されるコンピューターシステム110（プロセッサ120を介する）によって実施されるものとして記載されているが、プロセス300は、他のシステム、またはシステムの組合せによって実行またはそうでなければ実施されてもよい。図3に示される各ブロックは、1つまたは複数のプロセス、方法、またはサブルーチンをさらに表してもよく、ブロックの1つまたは複数のは、非一時的なコンピューター可読媒体に保存され、プロセッサまたは他の種類の処理回路によって実行される機械可読命令を含み、本明細書に記載の1つまたは複数の操作を実施することができる。本明細書に開示されるプロセス300の様々な操作、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ブロックは、同一もしくは異なる地理的場所、例えば、国において、および／または同一もしくは異なる人々によって、同時または異なる時間に実行され得る。

20

30

【0137】

操作302では、プロセッサ120は、TND試料を含む試料のセットからの生殖細胞系列SNPデータにアクセスすることができる。操作304では、プロセッサ120は、生殖細胞系列SNPデータに基づいて、ヘテロ接合性の出現率およびMAFのSDを判定することができる。操作306では、プロセッサ120は、個体由来の試料が遺伝子において体細胞ホモ接合欠失を含むという判定を支持する、各生殖細胞系列SNP部位に関する読み取りデータの第1の数を判定することができる。

【0138】

操作308では、プロセッサ120は、ヘテロ接合性の出現率、MAFの標準偏差(s d)、生殖細胞系列SNPデータ、および読み取りデータの数に基づいて、遺伝子が、体細胞ホモ接合欠失に関連する確率の第1のモデルの第1の出力を作成することができる。操作310では、プロセッサ120は、個体由来の試料に基づいて、TFの平均および標準偏差を判定することができる。操作312では、プロセッサ120は、個体由来の試料が、遺伝子において体細胞ヘテロ接合欠失を含むという判定を支持する、各生殖細胞系列SNP部位に関する読み取りデータの第2の数を判定することができる。操作314では、プロセッサ120は、TFの平均およびSD、生殖細胞系列SNPデータ、および読み取りデータの第2の数に基づいて、遺伝子が体細胞ヘテロ接合欠失に関連する確率の第2のモデルの第2の出力を作成することができる。操作316では、プロセッサ

40

50

120は、第1の出力と第2の出力とを比較することができる。操作318では、プロセッサ120は、比較に基づいて、第1の出力が選択されるべきか否かを判定することができる。操作320では、プロセッサ120は、第1の出力が選択されるべきか否かの判定に基づいて、遺伝子が体細胞ホモ接合欠失を含む確率を作成することができる。

【0139】

分類装置130は、遺伝子の状態を予測するために、様々なモデリング技法を適用することができる。分類装置130は、同様に、他のモデリング技法を使用することができる。例えば、図9Aおよび9Bは、異なるモデリング技法の結果の比較を例示する。他の確率技法も同様に使用することができる。

【0140】

遺伝物質の他の種類の状態のモデリング

【0141】

分類装置130は、遺伝物質の他の種類の状態をモデリングすることができる。例えば、分類装置130は、CNVなどの遺伝物質の様々な種類の状態を予測することができる。ここで、遺伝物質の状態を予測するプロセス400を例示する図4に対して言及されることになる。本明細書に記載の方法を実行するためには種々の方法が存在し得るため、プロセス400は、例として提供される。方法400は、主に、図1に例示されるコンピューターシステム110（プロセッサ120を介して）によって実施されるものとして記載されるが、プロセス400は、他のシステム、またはシステムの組合せによって実行またはそうでなければ実施され得る。図4に示される各ブロックは、1つまたは複数のプロセス、方法、またはサブルーチンをさらに表してもよく、ブロックの1つまたは複数の、非一時的なコンピューター可読媒体に保存され、プロセッサまたは他の種類の処理回路によって実行される機械可読命令を含み、本明細書に記載の1つまたは複数の操作を実施することができる。本明細書に開示されるプロセス400の様々な操作、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ブロックは、同一もしくは異なる地理的場所、例えば、国において、および／または同一もしくは異なる人々によって、同時または異なる時間に実行され得る。

【0142】

図4に関して記載される例は、IUT111の試料中のCNVを判定することを含む。より詳細には、例を使用して、IUT111の試料由来の遺伝物質におけるコピー数分散（copy number variance）（増幅など）を判定することができる。しかし、異なる状態の代替の（競合する）確率を使用し、最も可能性の高い確率を選択して、遺伝物質の他の種類の状態を同様にして判定することができる。

【0143】

操作402では、プロセッサ120は、遺伝物質の第1の状態をモデリングする第1のモデルを作成することができる。第1の状態は、第1のCNVまたは他の状態を含んでもよい。操作404では、プロセッサ120は、遺伝物質の第2の状態をモデリングする第2のモデルを作成することができる。第2の状態は、第2のCNVまたは他の状態を含んでもよい。操作406では、プロセッサ120は、第1のモデルに基づいて、第1のスコアを作成することができる。第1のスコアは、遺伝物質が第1の状態にある確率を示すことができる。

【0144】

操作408では、プロセッサ120は、第2のモデルに基づいて、第2のスコアを作成することができる。第2のスコアは、遺伝物質が第2の状態にある確率を示すことができる。操作410では、プロセッサ120は、第1のスコアと第2のスコアとを比較することができる。操作412では、プロセッサ120は、比較に基づいて、遺伝物質が第1の状態または第2の状態にあるという予測を作成することができる。

【0145】

生殖細胞系列SNPに関するMAFを分類装置130によって使用して、体細胞ヘテロ接合欠失および体細胞ホモ接合欠失の確率を作成する方法と同様に、MAFを使用して、

10

20

30

40

50

CNVの確率を解明することができる。例えば、CNVが検出されなかった試料中の生殖細胞系列SNPのMAFを使用して、試料中の読み取りデータが特定の増幅を支持するか否かを判定することができる。

【0146】

図5は、本開示の実施形態に従って、体細胞欠失の種類を例示する。体細胞ホモ接合欠失は、2つの方法で生じ得る：(1)生殖系細胞は遺伝子の単一のコピーを有し、体細胞は第2の欠失を獲得する(単一コピー増幅の検出に類似するLOD)。これらは、カバレッジ+重複するヘテロ接合性SNPなしに基づいて検出され得る。一部の例では、これらが観察されなかったとしても、生殖系細胞は遺伝子のコピーを有することができない。(2)体細胞ホモ接合欠失が生じ得る第2の方法は、生殖系細胞が遺伝子の2つのコピーを有し、体細胞が両コピーを失うことである(このシナリオがより高い出現率で観察された)。一部の実施形態では、二対立遺伝子体細胞コピー数の喪失では、生殖系細胞と体細胞の混合物において、生殖細胞系列ヘテロ接合性SNPの参照対立遺伝子頻度は0.5である。体細胞LOHの場合には、参照対立遺伝子ががん細胞において喪失しているかまたは保持されているかどうかに応じて、参照対立遺伝子頻度は $0.5 - 0.5 * TF$ (腫瘍含有率) または $0.5 + 0.5 * TF$ である。一部の実施形態では、LOHでは、期待された対立遺伝子頻度は、腫瘍細胞の割合に応じて変わり得る。よって、システムは、期待された対立遺伝子頻度0.5と比較して、計算された対立遺伝子頻度に基づいて、LOHと二対立遺伝子コピー数喪失とを識別することができる。

【0147】

図6Aは、本開示の実施形態に従って、BRCA1ホモ接合欠失の例示的プロット600(A)(1)および600(A)(2)を例示する。図6Bは、本開示の実施形態に従って、BRCA2ヘテロ接合欠失の例示的プロット600(A)(1)および600(A)(2)を示す。プロット600(A)(1)および600(B)(1)を参照して、所与のcfDNA試料に関して、正規化された分子カバレッジ(y軸)は、ゲノムの場所によって選別された標的化プローブ(x軸)にわたって表される。染色体分離は、縦線とプロットのボトムラインに提示した識別子によって表される。体細胞コピー数の変化を伴わない領域は、2に近い分子カバレッジを示すが、一方、体細胞欠失は、2未満の分子カバレッジレベルによって特定され得る。プロット600(B)(1)および600(B)(2)を参照して、同じ試料では、公知の生殖細胞系列SNPのMAF(y軸)は、それらのゲノムの場所(x軸)に対して表される。体細胞欠失は、上の行のカバレッジプロットにおいて観察されるように、50%に近い生殖細胞系列バリエーションMAFを明示するが(図6Aに例示されるプロットを参照されたい)、一方、ヘテロ接合性欠失は不均衡な生殖細胞系列バリエーションMAFを生じる(図6Bに例示されるプロットを参照されたい)。

【0148】

図7Aは、本開示の実施形態に従って、TND試料において観察される、ATM、BRCA1およびBRCA2遺伝子と重複する公知の生殖細胞系列SNPに関して、ヘテロ接合性遺伝子型の出現率の例示的プロットを例示する。図7Bは、本開示の実施形態に従って、TND試料にわたるMAFの例示的プロットを例示する。

【0149】

図8Aは、本開示の実施形態に従って、BRCA1に関するMAF値の例示的プロットを例示する。図8Bは、本開示の実施形態に従って、BRCA2に関するMAF値の例示的プロットを例示する。図8Aおよび8Bは、各SNPの3つの可能な遺伝子型(ホモ接合性代替対立遺伝子/ヘテロ接合性/ホモ接合性参照対立遺伝子)(x軸)に関する9つの公知の生殖細胞系列SNVのMAF(y軸)の例を示す。図9Aは、本開示の実施形態に従って、BRCA2パネルに関するベーター二項モデルと二項モデルのスコア比較の例示的プロットを例示する。図9Bは、本開示の実施形態に従って、BRCA2パネルに関するベーター二項モデルとガウス分布モデルのスコア比較の例示的プロットを例示する。図10Aは、本開示の実施形態に従って、BRCA1陰性試料に関するLLRスコア分布の例示的プロットを例示する。図10Bは、本開示の実施形態に従って、BRCA2陰性

試料に関するLLRスコア分布の例示的プロットを例示する。

【0150】

図11Aは、本開示の実施形態に従って、BRCA1に関する欠失のLODの例示的プロットを例示する。図11Bは、本開示の実施形態に従って、BRCA1に関するヘテロ接合性の喪失(LOH)(本明細書において、交換可能に「ヘテロ接合欠失」と称される)のLODの例示的プロットを示す。シミュレーション：TND試料から開始するホモ接合体細胞欠失の100kの事例。

【0151】

使用されたTF=28, 199個の試料において観察されたTF分布。LODは、2つの因子に依存する(2段階アルゴリズム)：(1)欠失検出感度(カバレッジのみに基づく)：BRCA1増幅/欠失平均カットオフ=0.05；および(2)ホモ接合体細胞欠失とヘテロ接合体細胞欠失とを区別する能力(LLR検査)。

【0152】

図12Aは、本開示の実施形態に従って、BRCA2に関する欠失のLODの例示的プロットを例示する。図12Bは、本開示の実施形態に従って、BRCA2に関するLOHのLODの例示的プロットを例示する。シミュレーション：TND試料から開始するホモ接合体細胞欠失の100kの事例。

【0153】

使用されたTF=28, 199個の試料において観察されたTF分布。

【0154】

LODは、2つの因子に依存する(2段階アルゴリズム)：(1)欠失検出感度(カバレッジのみに基づく)：BRCA2増幅/欠失平均カットオフ=0.09；および(2)ホモ接合体細胞欠失とヘテロ接合体細胞欠失とを区別する能力(LLR検査)。

【0155】

図13は、本開示の実施形態に従って、がんの種類に対するTFの出現率の例示的プロットを例示する。

【0156】

図14は、本開示の実施形態に従って、BRCA1およびBRCA2に関するLLRスコアの密度の例示的プロットを例示する。28,000個のトレーニング試料のセットは、ランダムに選択され、2.5および0のカットオフを有し(LOB部分において判定される)、BRCA1/2ホモ接合欠失を有する試料をコールした。387個および994個の試料は、それぞれ、BRCA1およびBRCA2に関する体細胞欠失を示した。これらの試料のうち、49個および60個が、それぞれ、BRCA1およびBRCA2のホモ接合欠失を有することがコールされた。

【0157】

図15は、本開示の実施形態に従って、複数のがんの種類の集団において観察されたBRCA2ホモ接合欠失の出現率の例示的チャートを例示する。図16は、本開示の実施形態に従って、複数のがんの種類の集団において観察されたBRCA1ホモ接合欠失の出現率の例示的チャートを例示する。図17は、本開示の実施形態に従って、BRCA2のホモ接合欠失と潜在的な臨床的対処可能性の例を例示する。図17に例示したプロットは、参照によりその全体が本明細書に組み込まれる、Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. (「Robinson」)による”Integrative clinical genomics of advanced prostate cancer,” Cell 161:1215-1228 (2015)からのものである。Robinsonは、BRCA2における体細胞の変異と病理性生殖細胞系列の変異の両方についての統合解析によって、BRCA2を喪失した19/150(12.7%)の事例を特定し、そのおよそ90%が二対立遺伝子の喪失を示したことを示す。これは、概して、体細胞点突然変異およびヘテロ接合性の喪失、ならびにホモ接合欠失の結

10

20

30

40

50

果であった。非選択のmC R P Cに罹患した個体におけるポリ（A D Pーリボース）ポリメラーゼ（P A R P）阻害を評価する臨床検査は、臨床利益を経験した、この検査における複数の罹患した個体が、二対立遺伝子B R C A 2の喪失を保有したことを示しており、臨床的対処可能性のさらなるエビデンスを提供する。

【0158】

図18Aは、本開示の実施形態に従って、B R C A 1ホモ接合欠失の例示的プロットを例示する。図18Bは、本開示の実施形態に従って、B R C A 1ホモ接合欠失の例示的プロットを例示する。図19Aは、本開示の実施形態に従って、B R C A 2ホモ接合欠失の例示的プロットを例示する。図19Bは、本開示の実施形態に従って、B R C A 2ホモ接合欠失の例示的プロットを例示する。図18A、18B、19A、および19Bは、ヒトゲノムに基づくプロットである。

10

【0159】

図20Aは、本開示の実施形態に従って、B R C A 1二対立遺伝子体細胞コピー数の喪失の例示的プロットを例示する。この開示の目的として、用語「二対立遺伝子体細胞コピー数の喪失」は、「ホモ接合欠失」と交換可能に使用されるであろう。図20Bは、本開示の実施形態に従って、B R C A 1 L O Hの例示的プロットを例示する。この開示の目的として、用語「L O H」は、「ヘテロ接合欠失」と交換可能に使用されるであろう。図21Aは、本開示の実施形態に従って、B R C A 2二対立遺伝子体細胞コピー数の喪失の例示的プロットを例示する。図21Bは、本開示の実施形態に従って、B R C A 2 L O Hの例示的プロットを例示する。図20A、20B、21A、および21Bは、3つの（ヒト）染色体に基づくプロットである。図22は、本開示の実施形態に従って、B R C A 1およびB R C A 2体細胞欠失の出現率のプロットを例示する。

20

【0160】

コンピューターによる実装

【0161】

本方法は、湿式化学ステップ以外の本明細書または添付の特許請求の範囲に記載されるステップのいずれかまたはすべてが、好適なプログラミングされたコンピューターにおいて実施され得るように、コンピューターに実装され得る。コンピューターは、メインフレーム、パーソナルコンピューター、タブレット、スマートフォン、クラウド、オンラインデータストレージ、リモートデータストレージなどであってもよい。コンピューターは、1つまたは複数の場所で操作され得る。

30

【0162】

本方法の様々な操作は、情報および／またはプログラムを利用することができ、コンピューター可読媒体（例えば、ハードドライブ、補助メモリー、外部メモリー、サーバー；データベース、ポータブルメモリーデバイス（例えば、C DーR、D V D、Z I Pディスク、フラッシュメモリーカード）など）に記憶される結果を生じる。

【0163】

本開示は、実行した場合に、本発明の方法のステップを実装する1つまたは複数のプログラムを含有する機械可読媒体を含む、核酸集団を解析するための製品も含む。

【0164】

40

本開示は、ハードウェアおよび／またはソフトウェアに実装され得る。例えば、本開示の異なる態様は、クライアントサイドロジックまたはサーバーサイドロジックのいずれかで実装され得る。本開示またはそのコンポーネントは、適当に設定されたコンピューティングデバイスへとロードされた場合に、そのデバイスに本開示に従って実施させるロジック命令および／またはデータを含有する固定媒体プログラムコンポーネントにおいて具現化され得る。ロジック命令を含有する固定媒体は、ビューアーコンピューターへの物理的ローディングのために、固定媒体上のビューアーに送達されてもよく、またはロジック命令を含有する固定媒体は、ビューアーが、通信媒体を介してアクセスして、プログラムコンポーネントをダウンロードするリモートサーバーに存在してもよい。

【0165】

50

本開示は、本開示の方法を実装するようにプログラミングされたコンピューター制御システムを提供する。プロセッサ１２０は、シングルコアもしくはマルチコアプロセッサ、または並行処理のための複数のプロセッサを含むことができる。ストレージデバイス１２２は、ランダムアクセスメモリー、リードオンリーメモリー、フラッシュメモリー、ハードディスク、および／または他の種類のストレージを含んでもよい。コンピューターシステム１１０は、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（例えば、ネットワークアダプター）、ならびにキャッシュ、他のメモリー、データストレージおよび／または電子ディスプレイアダプターなどの周辺デバイスを含んでもよい。コンピューターシステム１１０のコンポーネントは、内部通信バス、例えば、マザーボードを介して互いに通信することができる。ストレージデバイス１２２は、データを記憶するためのデータストレージユニット（またはデータリポジトリ）であってもよい。コンピューターシステム１１０は、通信インターフェースの助けを借りて、コンピューターネットワーク（「ネットワーク」）に作動可能に連結されていてもよい。ネットワークは、インターネット、インターネットおよび／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／もしくはエクストラネットであってもよい。ネットワークは、一部の場合には、遠距離通信および／またはデータネットワークである。ネットワークは、ローカルエリアネットワークを含んでもよい。ネットワークは、クラウドコンピューティングなどの分散コンピューティングを可能にし得る、１つまたは複数のコンピューターサーバーを含んでもよい。ネットワークは、コンピューターシステム１１０の助けを借りる一部の場合には、デバイスを、クライアントまたはサーバーとして機能するコンピューターシステム１２０に連結可能にし得る、ピアツーピアネットワークを実装し得る。

【０１６６】

プロセッサ１２０は、プログラムまたはソフトウェアにおいて具現化され得る、機械可読命令のシーケンスを実行することができる。命令は、ストレージデバイス１２２などのメモリーの場所に記憶され得る。命令はプロセッサ１２０を対象とすることができ、次に、プロセッサ１２０が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設定することができる。プロセッサ１２０によって実施された操作の例としては、フェッチ、デコード、実行、およびライトバックが挙げられ得る。

【０１６７】

プロセッサ１２０は、集積回路などの回路の一部であってもよい。システム１００の１つまたは複数の他のコンポーネントは、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）を含んでもよい。

【０１６８】

ストレージデバイス１２２は、ファイル、例えば、ドライバー、ライブラリーおよび保存されたプログラムを記憶し得る。ストレージデバイス１２２は、ユーザーのデータ、例えば、ユーザーの好みの設定およびユーザーのプログラムを記憶し得る。コンピューターシステム１１０は、一部の事例では、コンピューターシステム１１０の外部にある、例えば、イントラネットまたはインターネットを介してコンピューターシステム１１０と通信するリモートサーバーに設置された、１つもしくは複数の追加のデータストレージユニットを含んでもよい。

【０１６９】

コンピューターシステム１１０は、ネットワークを介して１つまたは複数のリモートコンピューターシステムと通信することができる。例えば、コンピューターシステム１１０は、ユーザーのリモートコンピューターシステムと通信することができる。リモートコンピューターシステムの例としては、パーソナルコンピューター（例えば、ポータブルPC）、スレートまたはタブレットPC（例えば、Apple（登録商標）のiPad（登録商標）、Samsung（登録商標）のGalaxy Tab）、電話機、スマートフォン（例えば、Apple（登録商標）のiPhone（登録商標）、Android（登録商標）対応機種、Blackberry（登録商標））、または携帯情報端末が挙げら

10

20

30

40

50

れる。ユーザーは、ネットワークを介してコンピューターシステム 110 にアクセスすることができる。

【0170】

本明細書に記載される方法は、例えば、ストレージデバイス 122 などのコンピューターシステム 110 の電子ストレージの場所に記憶された機械（例えば、コンピュータープロセッサ）実行可能コードによって実装され得る。機械実行可能または機械可読コードは、ソフトウェアの形態で提供され得る。使用中、コードは、プロセッサ 905 によって実行され得る。一部の場合には、コードは、ストレージユニット 915 から検索され、プロセッサ 120 によって容易にアクセスするためにストレージデバイス 122 に記憶され得る。

10

【0171】

コードは、コードを実行するために適合させ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および設定されてもよく、または実行時間中にコンパイルされてもよい。コードは、プリコンパイルまたはコンパイルされた様式で、コードを実行可能にするために選択され得るプログラム言語で供給され得る。

【0172】

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、例えば、コンピューターシステム 110 は、プログラミングによって具現化され得る。技術の様々な態様は、典型的には、ある種の機械可読媒体で実行されるかまたはそこで具現化される機械（またはプロセッサ）実行可能コードおよび／または関連するデータの形態の「プロダクト」または「製品」として考えられ得る。機械実行可能コードは、メモリー（例えば、リードオンリーメモリー、ランダムアクセスメモリー、フラッシュメモリー）またはハードディスクのような電子ストレージユニットに記憶され得る。

20

【0173】

「ストレージ」型媒体としては、コンピューター、プロセッサなど、またはそれらの関連モジュールの有形メモリー、例えばソフトウェアプログラミングのいずれかの時点で非一時的ストレージを提供することができる様々な半導体メモリー、テープドライブ、ディスクドライブなどのいずれかまたはすべてを挙げることができる。ソフトウェアのすべてまたは一部は、インターネットまたは様々な他の遠距離通信ネットワークを介して通信可能であることがある。このような通信は、例えば、一方のコンピューターまたはプロセッサから別のものへの、例えば、管理サーバーまたはホストコンピューターからアプリケーションサーバーのコンピュータープラットフォームへのソフトウェアのローディングを可能にし得る。よって、ソフトウェアエレメントを保有することができる別の種類の媒体は、ローカルデバイス間の物理的インターフェースを横切って、有線または光地上通信線ネットワークを介しておよび様々なエアリンクを介して使用されるような光波、電波および電磁波を含む。このような波を搬送する物理的エレメント、例えば、有線リンクまたは無線リンク、光リンクなどもソフトウェアを保有する媒体とみなされ得る。本明細書で使用される場合、非一時的な、有形の

30

【0174】

「ストレージ」媒体に限定されなければ、コンピューターまたは機械「可読媒体」などの用語は、実行のためにプロセッサへの命令の提供に関与する任意の媒体を指す。

40

【0175】

したがって、コンピューター実行可能コードなどの機械可読媒体は、以下に限定されないが、有形ストレージ媒体、搬送波媒体または物理的な伝送媒体を含む多くの形態をとることができる。不揮発性ストレージ媒体としては、例えば、図面に示され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ような、任意のコンピューターなどにおけるストレージデバイスのいずれかなどの光または磁気ディスクが挙げられる。揮発性ストレージ媒体は、このようなコンピュータープラットフォームのメインメモリーなどの動的メモリーを含む。有形伝送媒体は、同軸ケーブル；コンピューターシステム内にバスを含むワイヤを含む、銅線および光学ファイバーを含む。搬送波伝送媒体は、電気もしくは電磁

50

シグナル、または高周波（RF）赤外（IR）データ通信中に生成されるものなどの音波もしくは光波の形態を取ることができる。したがって、コンピューター可読媒体の一般的な形態は、例えば、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、穴のパターンを有する任意の他の物理的ストレージ媒体、RAM、ROM、PROMおよびEPROM、FLASH-EPROM、任意の他のメモリーチップもしくはカートリッジ、搬送波輸送データもしくは命令、このような搬送波を輸送するケーブルもしくはリンク、またはコンピューターがプログラミングコードおよび／もしくはデータを読むことができる任意の他の媒体を含む。コンピューター可読媒体のこれらの形態の多くは、実行のためにプロセッサに対する1つまたは複数の命令の1つまたは複数の配列の保有に關与し得る。

10

【0176】

コンピューターシステム110は、例えば、レポートを提供するためのユーザーインターフェース（UI）を含む電子ディスプレイ935を含むかまたはそれと通信し得る。UIの例としては、限定されないが、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のユーザーインターフェースが挙げられる。

【0177】

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として実装され得る。アルゴリズムは、プロセッサ120による実行の際に、ソフトウェアとして実装され得る。

20

【0178】

試料採取および解析パイプライン

【0179】

試料101は、対象から単離される任意の生体試料であってもよい。試料としては、身体組織、例えば、公知のまたは疑わしい固形腫瘍、全血、血小板、血清、血漿、便、赤血球細胞、白血球細胞または白血球、内皮細胞、組織生検、脳脊髄液、滑液、リンパ液、腹水液、間質内または細胞外液、歯肉溝液、骨髓、胸水、脳脊髄液、唾液、粘液、痰、精液、汗、尿を含む細胞間の空間における液体を挙げることができる。試料は、好ましくは体液、特に血液およびその画分、および尿である。このような試料は、核酸を含み得る。このような試料は、核酸試料とも称することができる。これらの試料の一部では、核酸は、腫瘍から流出され得る。核酸は、DNAおよびRNAを含んでもよく、二本鎖および／または一本鎖形態であってもよい。核酸がRNAを含む例では、本明細書に記載のシステムおよび方法は、参照遺伝子（GAPDHのような内在性制御遺伝子など）に対する目的の遺伝子の遺伝子発現を、正常な試料から計算されたトレーニングされた閾値と比較することによって、RNAによってコードされた目的の遺伝子における体細胞欠失を判定することができる。試料は、対象から元々単離された形態であってもよく、またはさらなる処理に供されて、細胞などの構成成分を除去または付加するか、1つの構成成分を別のものに対して濃縮するか、または1つの形態の核酸を別のものに、例えば、RNAをDNAに、もしくは一本鎖核酸を二本鎖に変換してもよい。よって、例えば、解析のための体液は、無細胞核酸、例えば、無細胞DNA（cfDNA）を含有する血漿または血清である。

30

40

【0180】

血漿の体積は、配列決定された領域に関する所望の読み深度に応じて変化し得る。例示的体積は、0.4~40ml、5~20ml、10~20mlである。例えば、体積は、0.5ml、1ml、5ml、10ml、20ml、30ml、または40mlであってもよい。サンプリングされた血漿の体積は、5から20mlであってもよい。

【0181】

試料は、ゲノム等価物を含有する様々な量の核酸を含むことができる。例えば、約30ngのDNAの試料は、約10,000（ 10^4 ）個の半数体ヒトゲノム等価物、cfDNAの場合には、約2000億（ 2×10^{11} ）個の個々のポリヌクレオチド分子を含有し得る。同様に、約100ngのDNAの試料は、約30,000個の半数体ヒトゲノム

50

等価物、c f D N A の場合には、約 6 0 0 0 億個の個々の分子を含有し得る。

【0182】

試料は、異なる供給源由来の、例えば、細胞および無細胞由来の核酸を含んでもよい。試料は、突然変異を有する核酸を含んでもよい。例えば、試料は、生殖細胞系列突然変異および／または体細胞突然変異を有する D N A を含んでもよい。試料は、がん関連突然変異（例えば、がん関連体細胞突然変異）を有する D N A を含んでもよい。

【0183】

増幅前の試料中の無細胞核酸の例示的な量は、約 1 f g から約 1 μ g、例えば、1 p g から 2 0 0 n g、1 n g から 1 0 0 n g、1 0 n g から 1 0 0 0 n g の範囲に及ぶ。例えば、量は、最大約 6 0 0 n g、最大約 5 0 0 n g、最大約 4 0 0 n g、最大約 3 0 0 n g、最大約 2 0 0 n g、最大約 1 0 0 n g、最大約 5 0 n g、または最大約 2 0 n g の無細胞核酸分子であり得る。量は、少なくとも 1 f g、少なくとも 1 0 f g、少なくとも 1 0 0 f g、少なくとも 1 p g、少なくとも 1 0 p g、少なくとも 1 0 0 p g、少なくとも 1 n g、少なくとも 1 0 n g、少なくとも 1 0 0 n g、少なくとも 1 5 0 n g、または少なくとも 2 0 0 n g 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 (f g)、1 0 f g、1 0 0 f g、1 ピコグラム (p g)、1 0 p g、1 0 0 p g、1 n g、1 0 n g、1 0 0 n g、1 5 0 n g、または 2 0 0 n g の無細胞核酸分子であり得る。方法は、1 フェムトグラム (f g) から 2 0 0 n g を得るステップを含むことができる。

【0184】

無細胞核酸試料は、無細胞核酸を含有する対象由来の試料を指す。無細胞核酸は、細胞内に含有されない核酸またはそうでなければ、細胞に結合していない核酸である。例えば、無細胞核酸試料は、無傷細胞を除去した後に試料中に残っている核酸を含むことができる。無細胞核酸は、対象由来の体液（例えば、血液、尿、C S F など）を供給源とするすべての封入されていない核酸を指し得る。無細胞核酸は、ゲノム D N A、ミトコンドリア D N A、循環 D N A、s i R N A、m i R N A、循環 R N A (c R N A)、t R N A、r R N A、核小体低分子 R N A (s n o R N A)、P i w i 相互作用 R N A (p i R N A)、長鎖ノンコーディング R N A (長鎖 n c R N A) を含む D N A (c f D N A)、R N A (c f R N A)、およびそれらのハイブリッド、またはこれらのいずれかの断片を含む。無細胞核酸は、二本鎖、一本鎖、またはこのハイブリッドであってもよい。無細胞核酸は、分泌または細胞死プロセス、例えば、細胞のネクロシスおよびアポトーシスによって体液に放出され得る。一部の無細胞核酸は、がん細胞、例えば、循環腫瘍 D N A (c t D N A) から体液に放出される。他のものは、健康な細胞から放出される。c t D N A は、封入されていない腫瘍由来の断片化 D N A であり得る。無細胞胎児 D N A (c f f D N A) は、母体の血流を自由に循環する胎児の D N A である。

【0185】

無細胞核酸またはそれに関連するタンパク質は、1 つまたは複数のエピジェネティック改変を有してもよく、例えば、無細胞核酸は、アセチル化、5-メチル化、ユビキチン化、リン酸化、S U M O 化、リボシル化、および／またはシトルリン化されてもよい。

【0186】

無細胞核酸は、約 1 0 0 ~ 5 0 0 ヌクレオチドの例示的なサイズ分布を有し、1 1 0 から約 2 3 0 ヌクレオチドの分子が、分子の約 9 0 % に相当し、ヒトにおける最頻値は約 1 6 8 ヌクレオチドであり、第 2 のマイナーピークは、2 4 0 から 4 4 0 ヌクレオチドの範囲にある。無細胞核酸は、約 1 6 0 から約 1 8 0 ヌクレオチド、または約 3 2 0 から約 3 6 0 ヌクレオチド、または約 4 4 0 から約 4 8 0 ヌクレオチドであってもよい。

【0187】

無細胞核酸は、溶液中に見出される無細胞核酸を、無傷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と一緒に処理してもよい。一般に、緩衝液の添加および洗浄ステップの後、無細胞核酸は、アルコールで沈降させることができる。夾雑物または塩を除去するた

10

20

30

40

50

めに、シリカベースカラムなどのさらなる清澄化ステップを使用してもよい。手順のある特定の態様、例えば、収率を最適化するために、例えば、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。

【0188】

そのような処理の後、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むことができる。必要に応じて、一本鎖DNAおよびRNAは、後続の処理および解析ステップに含められるように、二本鎖形態に変換されてもよい。

【0189】

タグ

【0190】

一部の実施形態では、核酸分子（ポリヌクレオチドの試料由来）は、試料インデックスおよび／または分子バーコード（一般的に「タグ」と称される）でタグ付けされてもよい。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（例えば、平滑末端ライゲーションまたは付着末端ライゲーション）、またはオーバーラップ伸長ポリメラーゼ連鎖反応（PCR）によって、アダプターに組み込まれてもよく、またはそうでなければ、アダプターに接合されてもよい。このようなアダプターは、最終的に、標的核酸分子に接合され得る。他の実施形態では、従来の核酸増幅方法を使用して、試料インデックスを核酸分子に導入するために、1回または複数の増幅サイクル（例えば、PCR増幅）が一般的に適用される。増幅は、1種または複数の反応混合物（例えば、アレイにおける複数のマイクロウェル）中で行われてもよい。分子バーコードおよび／または試料インデックスは、同時に、または任意の連続的順序で導入されてもよい。一部の実施形態では、分子バーコードおよび／または試料インデックスは、配列捕捉ステップが実施される前および／または後に導入される。一部の実施形態では、分子バーコードのみがプローブ捕捉の前に導入され、試料インデックスは、配列捕捉ステップが実施された後に導入される。一部の実施形態では、分子バーコードと試料インデックスの両方が、プローブに基づく捕捉ステップを実施する前に導入される。一部の実施形態では、試料インデックスは、配列捕捉ステップが実施された後に導入される。一部の実施形態では、分子バーコードは、ライゲーション（例えば、平滑末端ライゲーションまたは粘着末端ライゲーション）により、アダプターを介して試料中の核酸分子（例えば、cfDNA分子）に組み込まれる。一部の実施形態では、試料インデックスは、オーバーラップ伸長ポリメラーゼ連鎖反応（PCR）によって、試料中の核酸分子（例えば、cfDNA分子）に組み込まれる。典型的には、配列捕捉プロトコールは、標的化核酸配列に対して相補的な一本鎖核酸分子を導入することに関与し、例えば、ゲノム領域およびこのような領域の突然変異のコード配列は、がんの種類に関連する。

【0191】

一部の実施形態では、タグは、試料核酸分子の一方の末端または両方の末端に位置付けられ得る。一部の実施形態では、タグは、所定のまたはランダムもしくはセミランダム配列のオリゴヌクレオチドである。一部の実施形態では、タグは、約500、200、100、50、20、10、9、8、7、6、5、4、3、2、または1ヌクレオチド長未満であってもよい。タグは、ランダムまたは非ランダムに試料核酸に連結され得る。

【0192】

一部の実施形態では、各試料は、試料インデックスまたは試料インデックスの組合せで固有にタグ付けされる。一部の実施形態では、試料またはサブ試料の各核酸分子は、分子バーコードまたは分子バーコードの組合せで固有にタグ付けされる。他の実施形態では、分子バーコードが複数（例えば、非固有分子バーコード）の中で互いに必ずしも固有ではないように、複数の分子バーコードを使用することができる。これらの実施形態では、一般的に、分子バーコードとそれが結合し得る配列の組合せが個々に追跡され得る固有の配列を生じるように、分子バーコードは個々の分子に結合される（例えば、ライゲーションによって）。内在性配列情報（例えば、試料中の元の核酸分子の配列に対応する始まり（開始）および／もしくは末端（停止）部分、一方もしくは両方の末端の配列読み取りデー

10

20

30

40

50

タのサブ配列、配列読み取りデータの長さ、ならびに／または試料中の元の核酸分子の長さ）と組み合わせて非固有にタグ付けされた分子バーコードの検出により、典型的には、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る。個々の配列読み取りデータの長さ、または塩基対の数も必要に応じて使用して、所与の分子に固有の同一性を割り当てる。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に由来する断片は、それによって、それに続く親鎖に由来する断片、および／または相補鎖の特定を可能にし得る。

【0193】

一部の実施形態では、分子バーコードは、識別子のセット（例えば、固有または非固有の分子バーコードの組合せ）の予測される比で、試料中の分子に導入される。ある例示的な形式は、標的分子の両末端にライゲーションされた、約2から約1,000,000個の異なる分子バーコード、または約5から約150個の異なる分子バーコード、または約20から約50個の異なる分子バーコードを使用する。あるいは、約25から約1,000,000個の異なる分子バーコードが使用されてもよい。例えば、20～50×20～50個の分子バーコードが使用されてもよい。このような識別子の数は、典型的には、同じ開始点および停止点を有する異なる分子が、異なる組合せの識別子を受容する高い確率（例えば、少なくとも94%、99.5%、99.99%、または99.999%）を有するのに十分である。一部の実施形態では、約80%、約90%、約95%、または約99%の分子が、同じ組合せの分子バーコードを有する。

【0194】

一部の実施形態では、反応における固有または非固有の分子バーコードの割り当ては、例えば、それぞれが参照によりその全体が本明細書に組み込まれる、米国特許出願第20010053519号、同第20030152490号、および同第20110160078号、ならびに米国特許第6,582,908号、同第7,537,898号、同第9,598,731号、および同第9,902,992号に記載された方法およびシステムを使用して実施される。あるいは、一部の実施形態では、試料の異なる核酸分子は、内在性配列情報（例えば、開始および／もしくは停止位置、配列の一方もしくは両方の末端のサブ配列、ならびに／または長さ）のみを使用して特定されてもよい。

【0195】

増幅

【0196】

アダプターが隣接する試料核酸を、PCR、および典型的に、増幅させようとするDNA分子に隣接するアダプターにおけるプライマー結合部位にプライマーが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することができる。増幅方法は、サーマル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に関与してもよく、または転写に媒介される増幅におけ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、核酸配列ベースの増幅、および自己持続性配列ベースの複製が挙げられる。

【0197】

従来の核酸増幅方法を使用して、バーコードを核酸分子に導入するために、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種または複数の反応混合物中で行われてもよい。分子タグおよび試料インデックス／タグは、同時に、または任意の連続的順序で導入されてもよい。分子タグおよび試料インデックス／タグは、配列捕捉の前および／または後に導入されてもよい。一部の場合には、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、配列捕捉後に導入される。一部の場合には、分子タグと試料インデックス／タグの両方が、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／タグは、配列捕捉の後に導入される。通常、配列捕捉は、標的化配列、例えば、ゲノム領域のコード配列に相補的な一本鎖核酸分子を導入することに関与し、このような領域の突然変異は、がんの種類と関連している。典型的には、増幅により、200ntから700nt、250ntから350nt、または320ntから5

500 ntの範囲のサイズで、分子タグおよび試料インデックス／タグ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。

【0198】

濃縮

【0199】

一部の実施形態では、配列は、核酸を配列決定する前に濃縮される。濃縮は、特定の標的領域に対して、または非特異的に、必要に応じて実施される（「標的配列」）。一部の実施形態では、目的の標的化領域は、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関して選択される核酸捕捉プローブ（「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一般的に、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（例えば、異なる「分解能」で）、拘束セット（例えば、シーケンサー拘束、例えば、配列決定ロード、それぞれのベイトの有用性など）に供され、下流配列決定に所望のレベルで、標的化核酸を捕捉する。これらの目的の標的化ゲノム領域は、核酸構築物の天然または合成のヌクレオチド配列を必要に応じて含む。一部の実施形態では、1つまたは複数の目的の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、その後、必要に応じて、目的の領域の濃縮のためにこれらの領域を増幅させることができる。

【0200】

配列捕捉は、典型的には、標的核酸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に参与する。ある特定の実施形態では、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に関連する。このようなプローブは、例えば、約60から約120ヌクレオチド長であり得る。セットは、約2×、3×、4×、5×、6×、8×、9×、10×、15×、20×、50×、またはそれを上回る深度を有し得る。一般的に、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。一部の実施形態では、濃縮された集団は、配列決定前に増幅されてもよい。

【0201】

配列決定パイプライン

【0202】

事前の増幅を用いてまたは用いなく、アダプターに隣接する試料核酸を、例えば、1つまたは複数の配列決定デバイス107による、配列決定に供することができる。配列決定方法としては、例えば、サンガー配列決定、ハイスループット配列決定、ピロ配列決定、合成による配列決定、単一分子配列決定、ナノポア配列決定、半導体配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、RNA-Seq (Illumina)、Digital Gene Expression (Helicos)、次世代配列決定、Single Molecule Sequencing by Synthesis (SMSS) (Helicos)、大規模並列配列決定、Clonal Single Molecule Array (Solexa)、ショットガン配列決定、Ion Torrent、Oxford Nanopore、Roche Genia、Maxim-Gilbert配列決定、プライマーウォーキング、PacBio、SOLiD、Ion Torrent、またはNanoporeプラットフォームを使用する配列決定が挙げられる。配列決定の反応は、複数のレーン、複数のチャンネル、複数のウェル、または複数の試料セットを実質的に同時に処理する他の手段であり得る、種々の試料処理ユニットにおいて実施することができる。試料処理ユニットは、複数のランを同時に処理することが可能な複数の試料チャンバーも含んでもよい。

【0203】

配列決定反応は、がんまたは他の疾患のマーカーを含有することが公知の1つまたは複

数の断片の種類に関して実施されてもよい。配列決定反応は、試料中に存在する任意の核酸断片に関して実施されてもよい。配列反応は、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%未満であり得る。

【0204】

同時の配列決定反応は、マルチプレックス配列決定を使用して実施されてもよい。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、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000、100,000回の配列決定反応で配列決定され得る。他の場合には、無細胞ポリヌクレオチドは、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000、100,000回未満の配列決定反応で配列決定され得る。配列決定反応は、逐次的にまたは同時に実施されてもよい。次のデータ解析は、配列決定反応のすべてまたは一部に関して実施されてもよい。一部の場合には、データ解析は、少なくとも1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000、100,000回の配列決定反応で実施されてもよい。他の場合には、データ解析は、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000、100,000回未満の配列決定反応で実施されてもよい。例示的な読み深度は、遺伝子座（塩基）当たり1000~80000個の読み取りデータである。

【0205】

配列解析パイプライン

【0206】

一部の実施形態では、試料中の核酸を、同じ核酸分子のいずれか2つのコピーが、両末端に連結したアダプターから分子バーコードの同じ組合せを受容する確率が低くなる（例えば、1または0.1%未満）ように、分子バーコードを含む十分な数のアダプターと接触させることができる。このようなアダプターの使用により、所与の核酸分子から生じる核酸配列（配列読み取りデータ）のファミリーの特定が可能になる。例えば、参照配列に同じ開始および停止点を有し、分子バーコードの同じ組合せに連結した核酸配列は、ファミリーの一部であると考えられ得る。このように、ファミリーは、試料中の所与の核酸分子の増幅産物の配列を表し、ここで、ファミリーメンバーは、増幅産物から得られた配列読み取りデータである。ファミリーメンバーの配列は、平滑末端形成およびアダプター結合によって改変されたように、元の試料中の核酸分子に関するコンセンサヌクレオチドまたは完全コンセンサヌクレオチドを導出するようにコンパイルされ得る。言い換えれば、試料中の核酸の特定の位置を占有するヌクレオチドは、ファミリーメンバーの配列における対応する位置を占有するヌクレオチドのコンセンサスであることが判定される。ファミリーは、二本鎖核酸の一方または両方の鎖の配列を含むことができる。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、一方の鎖の配列は、コンセンサヌクレオチドまたは配列を導出するためにすべての配列をコンパイルする目的で、それらの相補体に変換される。一部のファミリーは、単一のメンバーの配列のみを含む。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得ることができる。あるいは、単一のメンバーの配列のみを有するファミリーは、次の解析から排除されてもよい。

【0207】

配列決定された核酸におけるヌクレオチドの変異は、配列決定された核酸を参照配列と比較することによって、判定することができる。参照配列は、公知の配列、例えば、対象由来の公知の全ゲノム配列または部分的ゲノム配列（例えば、ヒト対象の全ゲノム配列）である場合が多い。参照配列は、例えば、hg19またはhg38であってもよい。配列決定された核酸は、試料中の核酸について直接的に判定された配列、または上記のように

、このような核酸の増幅産物の配列のコンセンサスを表し得る。参照配列の1つまたは複数の指定の位置で比較を実施してもよい。各配列が最大限にアラインされた場合に、参照配列の指定の位置に対応する位置を含む、配列決定された核酸のサブセットを特定することができる。このようなサブセット内では、配列決定された核酸が、もしあれば、指定の位置にヌクレオチド変異を含み、必要に応じて、もしあれば、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）を含むことが判定され得る。ヌクレオチドバリエーションを含むサブセットにおける配列決定された核酸の数が選択された閾値を超える場合、バリエーションヌクレオチドは指定の位置でコールされ得る。閾値は、単一の数値、例えば、ヌクレオチドバリエーションを含むサブセット内の少なくとも1、2、3、4、5、6、7、9、または10個の配列決定された核酸であってもよく、または閾値は、比率、例えば、他の確率の中でもとりわけ、ヌクレオチドバリエーションを含むサブセット内の配列決定された核酸の少なくとも0.5、1、2、3、4、5、10、15、または20であってもよい。比較は、参照配列における目的のいずれかの指定の位置について反復することができる。時には、比較は、参照配列における少なくとも約20、100、200、または300の連続する位置、例えば、約20〜500、または約50〜300の連続する位置を占有する指定の位置について実施されてもよい。

【0208】

本方法を使用して、対象における状態、特にがんを引き起こし得る遺伝的事象の存在または非存在を特定し、状態を特徴付け（例えば、がんの病期分類またはがんの不均一性の判定）、状態の処置に対する応答をモニターし、状態の発症または状態のその後の経過の予後リスクをもたらすことができる。

【0209】

本発明の方法を使用して、様々ながんを検出することができる。がん細胞は、ほとんどの細胞と同様に、古い細胞が死滅して新しい細胞に置き換わるターンオーバーの速度によって特徴付けることができる。一般的に、所与の対象において脈管構造と接触している死細胞は、DNAまたはDNAの断片を血流中に放出し得る。これは、疾患の様々なステージのがん細胞にも当てはまる。がん細胞は、その疾患のステージに依存して、コピー数変異ならびに稀な突然変異などの様々な遺伝子異常によって、特徴付けることもできる。この現象を使用して、本明細書に記載の方法およびシステムを使用して、個体におけるがんの存在または非存在を検出することができる。

【0210】

検出され得るがんの種類および数としては、血液のがん、脳のがん、肺がん、皮膚がん、鼻のがん、咽喉がん、肝臓がん、骨がん、リンパ腫、膵がん、皮膚がん、腸がん、直腸がん、甲状腺がん、膀胱がん、腎臓がん、口腔がん、胃がん、固形腫瘍、不均一な腫瘍、均一な腫瘍などを挙げるることができる。

【0211】

がんは、突然変異、稀な突然変異、インデル、コピー数変異、トランスバージョン、転座、逆位、欠失、異数性、部分異数性、倍数性、染色体不安定性、染色体の構造変化、遺伝子融合、染色体融合、遺伝子切断、遺伝子増幅、遺伝子重複、染色体損傷、DNA損傷、核酸化学修飾の異常な変化、エピジェネティックパターンの異常な変化を含む遺伝的変異から検出することができる。

【0212】

がんの特異的形態を特徴付けるために、遺伝子データを使用することもできる。がんは、組成と病期分類の両方において不均一であることが多い。遺伝子プロファイルデータによって、がんの特異的サブタイプの診断または処置において重要であり得る、その特異的サブタイプの特徴付けが可能になり得る。この情報は、対象または開業医に、特定の種類のがんの予後に関する手掛かりを与え、対象または開業医のいずれかが、疾患の進行と一致した処置選択肢を採用することを可能にもし得る。一部のがんは進行し、より攻撃的かつ遺伝的に不安定になる。他のがんは、良性、不活性または休止状態のままであり得る。本開示のシステムおよび方法は、疾患進行を判定する際に有用であり得る。

【0213】

本発明の解析は、特定の処置選択肢の有効性を判定する際にも有用である。処置が成功する場合、より多くのがんが死滅し、DNAを流出し得るため、成功する処置選択肢は、対象の血中で検出されるコピー数変異または稀な突然変異の量を増加させ得る。他の例では、これは生じ得ない。別の例では、おそらく、ある特定の処置選択肢は、がんの遺伝子プロファイルと経時的に相関し得る。この相関は、治療を選択する際に有用であり得る。さらに、がんが、処置後に寛解することが観察される場合、本発明の方法を使用して、残りの疾患または疾患の再発をモニターすることができる。

【0214】

がん以外の状態における遺伝的変異を検出するために本発明の方法を使用することもできる。B細胞などの免疫細胞は、ある特定の疾患が存在すると、急速なクローン増殖を受け得る。クローン増殖は、コピー数変異の検出を使用してモニターすることができ、ある特定の免疫状態がモニターされ得る。この例では、コピー数変異の解析を経時的に実施して、特定の疾患がどのように進行し得るかというプロファイルを作成することができる。コピー数変異またはさらに稀な突然変異の検出を使用して、病原体の集団が、感染の経過中にどのように変化しているかを判定することができる。これは、ウイルスが感染の経過中に生活環状態を変化させ得る、および／またはより毒性の高い形に突然変異し得る、HIV/AIDSまたは肝炎感染などの慢性感染中に特に重要であり得る。本発明の方法は、免疫細胞が移植組織を破壊しようとする場合に、宿主の身体の拒絶活性を判定またはプロファイルするために使用され、移植組織の状態をモニターし、および処置の経過または拒絶の予防を変更することができる。

【0215】

さらに、本開示の方法を使用して、対象における異常な状態の不均一性を特徴付けることができ、本方法は、対象における細胞外ポリヌクレオチドの遺伝子プロファイルを作成するステップであって、遺伝子プロファイルが、コピー数変異および稀な突然変異の解析から得られる複数のデータを含む、ステップを含む。以下に限定されないが、がんを含む一部の場合には、疾患は不均一である場合がある。疾患細胞は、同一でない場合がある。がんの例では、一部の腫瘍は、異なる種類の腫瘍細胞、がんの異なるステージにある一部の細胞を含むことが公知である。他の例では、不均一性は、疾患の複数の病巣を含む場合がある。さらに、がんの例では、1つまたは複数の病巣が、おそらく原発部位から拡がった転移の結果である、複数の腫瘍病巣が存在し得る。

【0216】

本方法は、不均一な疾患における異なる細胞由来の遺伝子情報の総和であるプロファイル、フィンガープリントまたはデータのセットを作成するために使用することができる。このデータのセットは、コピー数変異および稀な突然変異の解析を単独または組合せで含んでもよい。

【0217】

本方法は、がんまたは胎児期由来の他の疾患を診断、予測、モニターまたは観察するために使用することができる。すなわち、これらの方法論は、そのDNAおよび他のポリヌクレオチドが母体の分子と共に循環している可能性がある出生前の対象におけるがんまたは他の疾患を診断、予測、モニターまたは観察するために、妊娠中の対象において用いることができる。

【0218】

精度の高い処置の例

【0219】

改良されたコンピューターシステム110によって提供される精度の高い診断によって、精度の高い処置計画を得ることができ、これらは、コンピューターシステム110によって特定され得る（および／または医療従事者によって精選され得る）。例えば、1種の精度の高い診断および処置は、相同組換え修復（HRR）経路の遺伝子に関する場合がある。

【0220】

相同組換えは、ヌクレオチド配列が、2つの、類似するかまたは同一のDNA分子の間で交換される遺伝子組換えの1種である。これは、二本鎖切断（DSB）として公知の、DNAの両方の鎖で起こる有害な切断を正確に修復するために、細胞によって最も広く使用される。HRRは、細胞分裂が起こる前の染色体切断を排除するために、複製された（SおよびG2期）DNAに存在する損傷のエラーを含まない除去に関するメカニズムを提供する。相同組換えが、DNAにおける二本鎖切断をどのように修復するかについての主要なモデルは、二本鎖切断修復（DSBR）経路および合成依存鎖アニーリング（SDSA）経路を媒介する相同組換え修復経路である。相同組換え遺伝子における生殖細胞系列および体細胞の欠陥は、乳がん、卵巣がんおよび前立腺がんに関与している。

10

【0221】

試料中のバリエーションヌクレオチドの数および種類は、試料を提供する対象の処置、すなわち治療介入の受け易さの指標を提供し得る。例えば、様々なポリADPリボースポリメラーゼ（PARP）阻害剤は、BRCA1またはBRCA2遺伝子における遺伝的突然変異によって引き起こされる乳がん、卵巣がんおよび前立腺がん由来の腫瘍の成長を停止させることが示されてきた。これらの治療剤のいくつかは、HRRの欠陥を補うことができる、塩基除去修復（BER）を阻害することができる。

【0222】

一方、ある特定のBRCAおよびHRR野生型患者は、PARP阻害剤による処置から臨床利益を達成することができない。さらに、BRCA突然変異を有する卵巣がん患者のすべてが、PARP阻害剤に対して応答する訳ではない。さらに、異なる種類の突然変異は、異なる治療を示し得る。例えば、HRR遺伝子における体細胞ヘテロ接合欠失は、体細胞ホモ接合欠失とは異なる治療を示し得る。よって、遺伝物質の状態は、治療に影響を及ぼし得る。一例では、PARP阻害剤は、HRR遺伝子に体細胞ホモ接合欠失を保有する個体に投与することができるが、HRR遺伝子に野生型対立遺伝子または体細胞ヘテロ接合欠失を保有する個体には投与することができない。

20

【0223】

配列決定された核酸におけるヌクレオチド変異は、配列決定された核酸を参照配列と比較することによって判定することができる。参照配列は、公知の配列、例えば、対象由来の公知の全ゲノム配列または部分的ゲノム配列、ヒト対象の全ゲノム配列であることが多い。参照配列は、hg19であってもよい。配列決定された核酸は、試料中の核酸について直接的に判定された配列、または上記のように、このような核酸の増幅産物の配列のコンセンサスを表し得る。参照配列の1つまたは複数の指定の位置で比較を実施してもよい。各配列が最大限にアラインされた場合に、参照配列の指定の位置に対応する位置を含む、配列決定された核酸のサブセットを特定することができる。このようなサブセット内では、配列決定された核酸が、もしあれば、指定の位置にヌクレオチド変異を含み、必要に応じて、もしあれば、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）を含むことが判定され得る。ヌクレオチドバリエーションを含むサブセットにおける配列決定された核酸の数が閾値を超える場合、バリエーションヌクレオチドは、指定の位置でコールされ得る。閾値は、単一の数値、例えば、ヌクレオチドバリエーションを含むサブセット内の少なくとも1、2、3、4、5、6、7、8、9、または10個の配列決定された核酸であってもよく、または閾値は、比率、例えば、他の確率の中でもとりわけ、ヌクレオチドバリエーションを含むサブセット内の配列決定された核酸の少なくとも0.5、1、2、3、4、5、10、15、または20であってもよい。比較は、参照配列における目的のいずれかの指定の位置について反復することができる。時には、比較は、参照配列における少なくとも20、100、200、または300の連続する位置、例えば、20～500、または50～300の連続する位置を占有する指定の位置について実施されてもよい。

30

40

【0224】

【実施例】

【0225】

50

Guardant Health, Inc. からの 73 遺伝子次世代配列決定 c t DNA パネルを使用して、配列決定された進行した固形腫瘍を有する 28, 199 名の患者由来の血漿試料に本明細書に記載のモデリングを適用した。

【0226】

結果の例は、9%～11%の腫瘍含有率を示す試料に対して、BRCA1/2 遺伝子の欠失を検出する際に 95%の感度を示した。LOH および二対立遺伝子のコピー数喪失に関する検出限界は、11%～13%である。観察された BRCA1 体細胞欠失の出現率は、乳がん、結腸直腸がん、前立腺がんおよび子宮内膜がんにおける 3%より高い。観察された BRCA2 体細胞欠失の出現率は、乳癌、肺癌、前立腺癌、頭頸部癌 (HNSCC) および肝細胞癌における 6%より高い。

10

【0227】

古典的 HRD に関連するがんを有する 5, 568 名の患者のコホートでは、体細胞 LOH および二対立遺伝子体細胞コピー数の喪失は、BRCA1 では試料の 2.7%で、および BRCA2 では試料の 8.0%で検出され、以前に報告した組織出現率とアラインされる。BRCA1 および BRCA2 の LOH は、乳がん、卵巣がん、前立腺がん、および膀胱がんを含む古典的相同組換え欠陥 (HRD) がんの 2.4% (134/5568) および 7.4% (415/5568) で観察された。BRCA1 および BRCA2 二対立遺伝子体細胞コピー数の喪失は、HRD がんのこの同じ群の 0.3% (19/5568) および 0.5% (31/5568) において観察された。BRCA1/2 体細胞 LOH および二対立遺伝子体細胞コピー数の喪失は、本明細書に記載のモデルの適用に基づいて、c t DNA において正確に検出することができる。非侵襲的 c t DNA 評価による、この治療的に標的可能なゲノム変更を特定する能力は、深部の内臓の場所、主に、乳がんおよび前立腺がんなどの骨および脳への転移のために、特に、その疾患が組織検査を要求する患者において、重要な臨床的意味を有する。

20

【0228】

上記または下記において引用したすべての特許出願、ウェブサイト、他の刊行物、受託番号などは、各個別の項目が参照によって組み込まれることが具体的かつ個別に示されたのと同じ程度に、すべての目的に関して参照によりその全体が組み込まれる。配列の異なるバージョンが異なる時間に受託番号に関連付けられている場合、本出願の有効な出願日において受託番号に関連付けられているバージョンを意味する。有効出願日とは、該当する場合、受託番号に関する実際の出願日または優先出願の出願日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間に公開されている場合、別段の指定のない限り、本出願の有効出願日の直近に公開されたバージョンを意味する。特に別段の指定のない限り、本開示の任意の特徴、ステップ、要素、実施形態、または態様は、任意の他のものと組み合わせて使用され得る。本開示は、明瞭さおよび理解のための例示および例によって、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にある特定の変化および改変を実践することができることが明らかになる。

30

特定の実施形態では、例えば以下の項目が提供される。

(項目 1)

遺伝子の生殖細胞系列欠失を示さない試料中の前記遺伝子の体細胞ホモ接合欠失と体細胞ヘテロ接合欠失とを識別するためのコンピューターシステムであって、

40

第 1 の確率分布により、前記遺伝子に関連する 1 つまたは複数の生殖細胞系列一塩基多型 (SNP) の位置に基づいて、前記体細胞ホモ接合欠失を表す、対立遺伝子数の第 1 のモデルを作成し、

第 2 の確率分布により、前記 1 つまたは複数の生殖細胞系列 SNP の位置に基づいて、前記体細胞ヘテロ接合欠失を表す、前記試料中の対立遺伝子数の第 2 のモデルを作成し、

前記第 1 のモデルの第 1 の出力と前記第 2 のモデルの第 2 の出力とを比較し、

前記比較に基づいて、前記遺伝子に関する前記体細胞ホモ接合欠失が前記試料中に存在するという予測を作成する

ようにプログラミングされたプロセッサ

50

を含む、コンピューターシステム。

(項目 2)

前記第 1 のモデルが、前記試料が前記体細胞ホモ接合欠失を含む第 1 の確率を表し、前記第 2 のモデルが、前記試料が前記体細胞ヘテロ接合欠失を含む第 2 の確率を表す、項目 1 に記載のコンピューターシステム。

(項目 3)

前記第 1 の確率分布が、前記第 2 の確率分布と同じ種類の確率分布である、項目 1 に記載のコンピューターシステム。

(項目 4)

前記第 1 のモデルを作成するために、前記プロセッサが、前記第 1 の確率分布に対する入力のための 1 つまたは複数のパラメータを判定するようにプログラミングされている、項目 1 に記載のコンピューターシステム。

(項目 5)

前記第 1 の確率分布が、ベータ-二項分布、二項分布、または正規分布のうちの 1 つを含む 1 種の確率分布を含む、項目 4 に記載のコンピューターシステム。

(項目 6)

対立遺伝子数の前記第 1 のモデルを作成するために、前記プロセッサが、前記第 1 の確率分布に対する入力のために、試料のトレーニングセットにおいて、前記 1 つまたは複数の生殖細胞系列 SNP のヘテロ接合性の出現率を判定するようにさらにプログラミングされている、項目 4 に記載のコンピューターシステム。

(項目 7)

試料の前記トレーニングセットが、腫瘍が検出されない (TND) 複数の試料を含む、項目 6 に記載のコンピューターシステム。

(項目 8)

対立遺伝子数の前記第 1 のモデルを作成するために、前記プロセッサが、前記第 1 の確率分布に対する入力のために、試料の前記トレーニングセットにおいて、前記 1 つまたは複数の生殖細胞系列 SNP と関連するマイナー対立遺伝子頻度 (MAF) の標準偏差を判定するようにさらにプログラミングされている、項目 6 に記載のコンピューターシステム。

(項目 9)

前記第 1 のモデルを作成するために、前記プロセッサが、前記第 1 の確率分布に対する入力のために、変異型対立遺伝子を支持する前記試料中の分子の数を判定するようにさらにプログラミングされている、項目 8 に記載のコンピューターシステム。

(項目 10)

前記第 1 のモデルを作成するために、前記プロセッサが、前記第 1 の確率分布に対する入力のために、前記試料中の分子の総数を判定するようにさらにプログラミングされている、項目 9 に記載のコンピューターシステム。

(項目 11)

前記第 1 のモデルを作成するために、前記プロセッサが、体細胞ホモ接合欠失と関連する分子カバレッジに基づいて、前記体細胞ホモ接合欠失を仮定して、前記試料において、前記 1 つまたは複数の生殖細胞系列 SNP の位置の前記対立遺伝子数の第 1 の尤度を計算するようにさらにプログラミングされている、項目 10 に記載のコンピューターシステム。

(項目 12)

前記第 2 のモデルを作成するために、前記プロセッサが、体細胞ヘテロ接合欠失と関連する分子カバレッジに基づいて、前記体細胞ヘテロ接合欠失を仮定して、前記試料において、前記 1 つまたは複数の生殖細胞系列 SNP の位置の前記対立遺伝子数の第 2 の尤度を計算するようにさらにプログラミングされている、項目 11 に記載のコンピューターシステム。

10

20

30

40

50

(項目 1 3)

前記第 2 のモデルを作成するために、前記プロセッサが、

前記第 2 のモデルに関する前記第 2 の確率分布に対する入力のために、前記試料から推定された腫瘍含有率の平均を判定する

ようにさらにプログラミングされている、項目 4 に記載のコンピューターシステム。

(項目 1 4)

前記腫瘍含有率が、配列カバレッジ情報に基づいて推定される、項目 1 3 に記載のコンピューターシステム。

(項目 1 5)

前記第 2 のモデルを作成するために、前記プロセッサが、

前記第 2 のモデルに関する前記第 2 の確率分布に対する入力のために、前記試料から推定される腫瘍含有率の標準偏差を判定する

ようにさらにプログラミングされている、項目 1 3 に記載のコンピューターシステム。

(項目 1 6)

前記プロセッサが、

複数の試料にアクセスし、

生殖細胞系列欠失を含む前記複数の試料の中から試料のセットを特定し、

前記複数の試料から試料の前記セットをフィルタリングして出力し、

前記フィルタリングされた複数の試料の中から、前記体細胞ホモ接合欠失または前記体細胞ヘテロ接合欠失の存在を特定する

ようにさらにプログラミングされている、項目 1 に記載のコンピューターシステム。

(項目 1 7)

前記第 1 の出力が、前記体細胞ホモ接合欠失の存在についての第 1 の確率を含み、前記第 2 の出力が、前記体細胞ヘテロ接合欠失の存在についての第 2 の確率を含む、項目 1 に記載のコンピューターシステム。

(項目 1 8)

前記第 1 のモデルの前記第 1 の出力と前記第 2 のモデルの前記第 2 の出力とを比較するために、前記プロセッサが、

前記第 1 の出力と前記第 2 の出力とに基づいて、対数尤度関数を実行する

ようにさらにプログラミングされている、項目 1 4 に記載のコンピューターシステム。

(項目 1 9)

前記遺伝子が、BRCA1、BRCA2、およびATMのうちの1つを含む、項目 1 に記載のコンピューターシステム。

(項目 2 0)

試料中の遺伝子が体細胞ホモ接合欠失を含む第 1 の確率を作成し、

前記試料中の前記遺伝子が体細胞ヘテロ接合欠失を含む第 2 の確率を作成し、

前記第 1 の確率と前記第 2 の確率とを比較し、

前記試料が前記体細胞ホモ接合欠失または前記体細胞ヘテロ接合欠失を含むか否かの予測を作成する

ようにプログラミングされたプロセッサ

を含むシステム。

(項目 2 1)

試料中の遺伝物質が第 1 の状態を含む第 1 の確率を作成し、

前記試料中の遺伝物質が第 2 の状態を含む第 2 の確率を作成し、

前記第 1 の確率と前記第 2 の確率とを比較し、

前記試料が前記第 1 の状態または前記第 2 の状態を含むか否かの予測を作成する

ようにプログラミングされたプロセッサ

を含むシステム。

(項目 2 2)

前記第 1 の状態が、第 1 のモデルに基づいて判定される体細胞ホモ接合欠失を含み、前

10

20

30

40

50

記第2の状態が、第2のモデルに基づいて判定される体細胞ヘテロ接合欠失を含む、項目21に記載のシステム。

(項目23)

前記第1の状態が、第1のコピー数バリエーション(CNV)を含み、前記第2の状態が、前記第1のCNVとは異なる第2のCNVを含む、項目21に記載のシステム。

(項目24)

前記第1のCNVおよび／または前記第2のCNVが、有害な状態に関連する、項目23に記載のシステム。

(項目25)

前記第1の確率を作成するために、前記プロセッサが、
遺伝子に関連する1つまたは複数の生殖細胞系列塩基多型(SNP)の位置にアクセスし、

10

試料のトレーニングセットにおいて前記1つまたは複数の生殖細胞系列SNPに関連するマイナー対立遺伝子頻度(MAF)の標準偏差を判定するようにプログラミングされている、項目23に記載のシステム。

(項目26)

前記第1の確率を作成するために、前記プロセッサが、
試料の前記トレーニングセットにおいて前記1つまたは複数の生殖細胞系列SNPと関連するマイナー対立遺伝子頻度(MAF)の標準偏差を判定するようにプログラミングされている、項目23に記載のシステム。

20

(項目27)

プロセッサによって実装される方法であって、
前記プロセッサによって、第1の確率分布により、遺伝子に関連する1つまたは複数の生殖細胞系列塩基多型(SNP)の位置に基づいて、体細胞ホモ接合欠失を表す、対立遺伝子数の第1のモデルを作成するステップと、
前記プロセッサによって、第2の確率分布により、前記1つまたは複数の生殖細胞系列SNPの位置に基づいて、体細胞ヘテロ接合欠失を表す、試料中の対立遺伝子数の第2のモデルを作成するステップと、
前記プロセッサによって、前記第1のモデルの第1の出力と前記第2のモデルの第2の出力とを比較するステップと、
前記プロセッサによって、前記比較に基づいて、前記遺伝子に関する前記体細胞ホモ接合欠失が前記試料中に存在するという予測を作成するステップとを含む、方法。

30

(項目28)

プロセッサによって実装される方法であって、
前記プロセッサによって、試料中の遺伝子が体細胞ホモ接合欠失を含む第1の確率を作成するステップと、
前記プロセッサによって、前記試料中の前記遺伝子が体細胞ヘテロ接合欠失を含む第2の確率を作成するステップと、
前記プロセッサによって、前記第1の確率と前記第2の確率とを比較するステップと、
前記プロセッサによって、前記試料が前記体細胞ホモ接合欠失または前記体細胞ヘテロ接合欠失を含むか否かの予測を、前記比較に基づいて作成するステップとを含む、方法。

40

(項目29)

プロセッサによって実装される方法であって、
前記プロセッサによって、試料中の遺伝物質が第1の状態を含む第1の確率を作成するステップと、
前記プロセッサによって、前記試料中の遺伝物質が第2の状態を含む第2の確率を作成するステップと、
前記プロセッサによって、前記第1の確率と前記第2の確率とを比較するステップと、

50

前記試料が前記第 1 の状態または前記第 2 の状態を含むか否かの予測を、前記比較に基づいて作成するステップと
を含む、方法。

(項目 3 0)

いずれかの先行する項目に基づいて、体細胞ホモ接合欠失を有すると判定された対象に、前記体細胞ホモ接合欠失に関連するがんを処置するのに有効な治療介入を施与するための方法。

(項目 3 1)

前記治療介入が、ポリ A D P リボースポリメラーゼ (P A R P) 阻害剤を含む、項目 3 0 に記載の方法。

(項目 3 2)

前記治療介入が、塩基除去修復 (B E R) 阻害剤を含む、項目 3 0 に記載の方法。

(項目 3 3)

いずれかの先行する項目に基づいて、特定の状態の遺伝物質を有すると判定された対象に、前記遺伝物質の状態に関連する疾患を処置するのに有効な治療介入を施与するための方法。

(項目 3 4)

いずれかの先行する項目に基づいて、体細胞ホモ接合欠失を有しないと判定された対象に、ポリ A D P リボースポリメラーゼ (P A R P) 阻害剤を除外するための治療介入を施与するための方法。

(項目 3 5)

前記試料中の前記遺伝子および／または遺伝物質の状態に関する情報、および／またはそれに由来する情報を必要に応じて含むレポートを作成するステップをさらに含む、項目 1 から 2 9 のいずれか一項に記載の方法またはシステム。

(項目 3 6)

前記レポートを、前記試料が由来する前記対象または医療従事者などの第 3 のパーティーに通信するステップをさらに含む、項目 3 5 に記載の方法またはシステム。

(項目 3 7)

試料中の遺伝子の体細胞欠失が、体細胞ホモ接合欠失または体細胞ヘテロ接合欠失であるか否かを判定する方法であって、前記試料が、前記遺伝子の生殖細胞系列欠失を示さず、前記方法が、

a) 前記試料中の前記遺伝子における生殖細胞系列一塩基多型 (S N P) の参照対立遺伝子頻度を判定するステップと、

b) 前記体細胞欠失を

(i) 前記試料中の前記生殖細胞系列 S N P の前記参照対立遺伝子頻度が、S N P の遺伝子型に応じて、腫瘍非検出 (T N D) 試料において観察される、所定の範囲内にある場合は、体細胞ホモ接合欠失、または

(i i) 前記試料中の前記生殖細胞系列 S N P の前記参照対立遺伝子頻度が、前記所定の範囲から規定量、逸脱する場合は、体細胞ヘテロ接合欠失
として分類するステップと
を含む、方法。

(項目 3 8)

前記所定の範囲が、0 ~ 0 . 0 1、0 . 4 5 ~ 0 . 5 5、または 0 . 9 9 ~ 1 である、項目 3 7 に記載の方法。

(項目 3 9)

前記試料中の推定腫瘍含有率を判定するステップと、
前記推定腫瘍含有率に基づいて、前記規定量を計算するステップと
をさらに含む、項目 3 7 に記載の方法。

10

20

30

40

50

【図面】

【図 1】

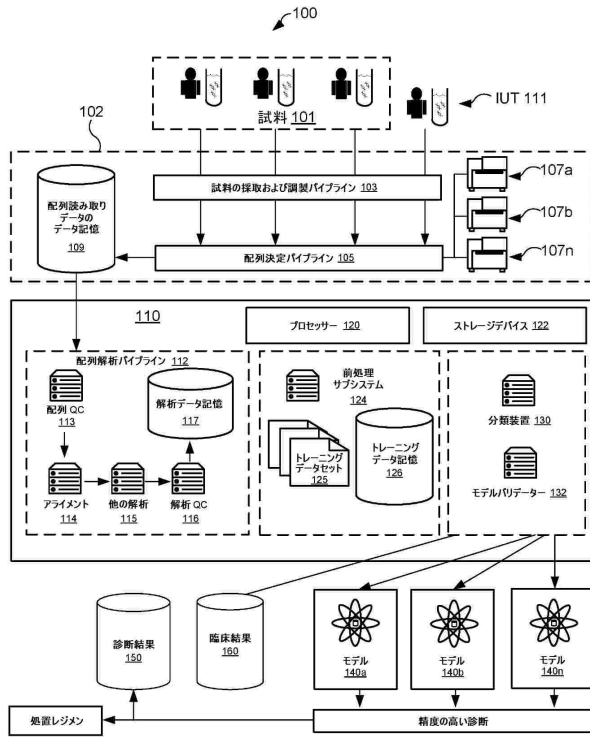


FIG. 1

【図 2】

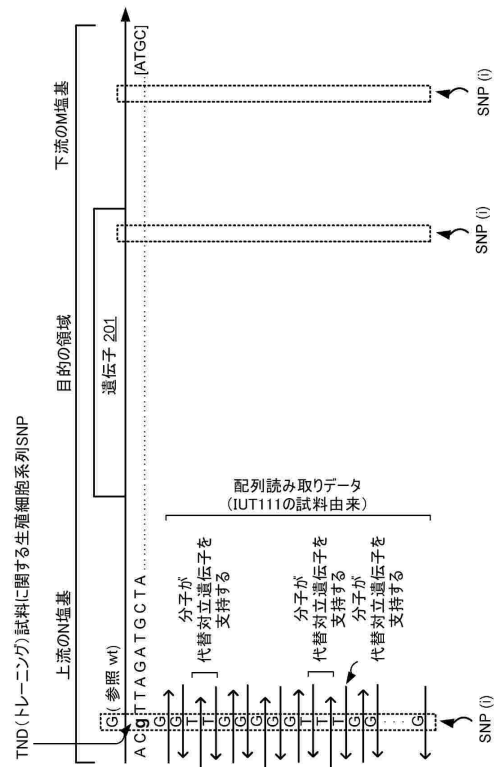


FIG. 2

【図 3】

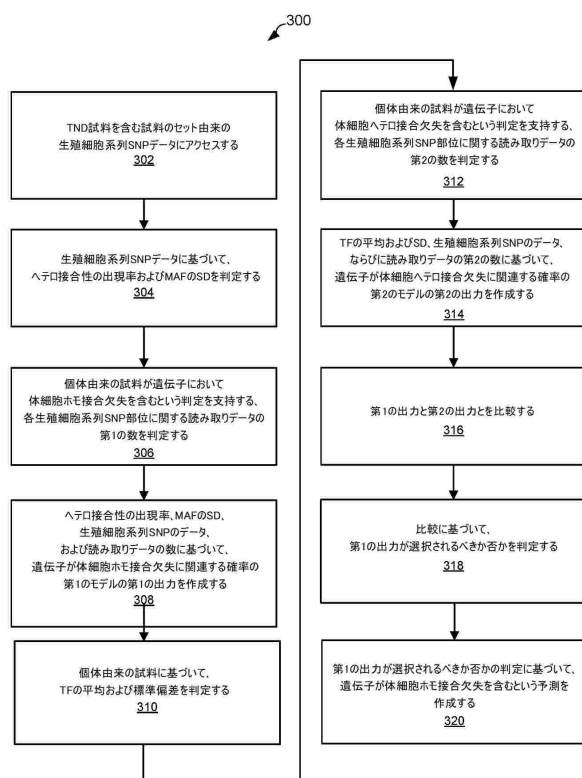


FIG. 3

【図 4】

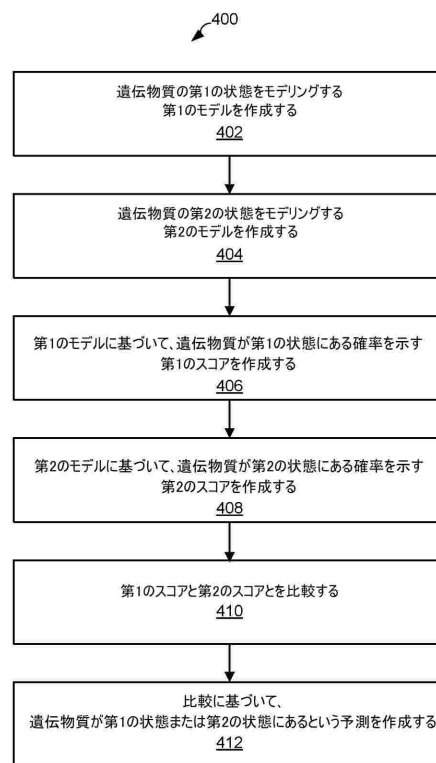


FIG. 4

【図 5】

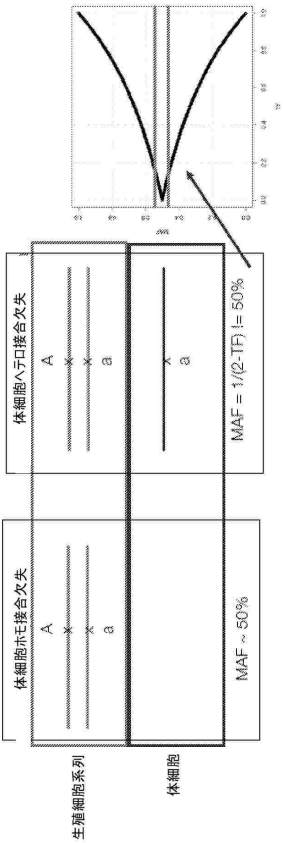


FIG. 5

【図 6 A】

BRCA1ホモ接合欠失の例

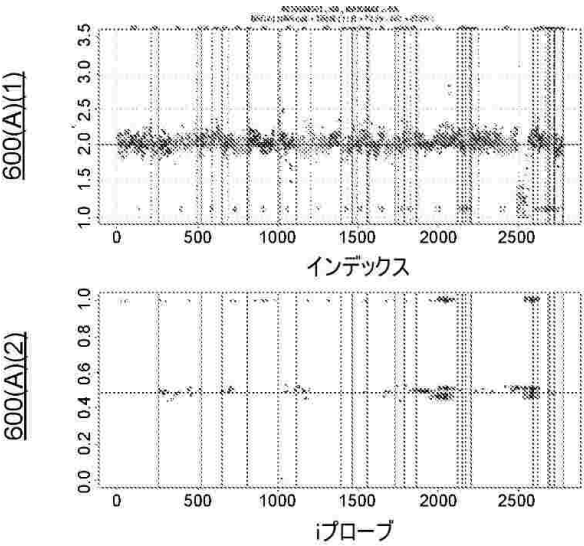


FIG. 6A

【図 6 B】

BRCA2ヘテロ接合欠失の例

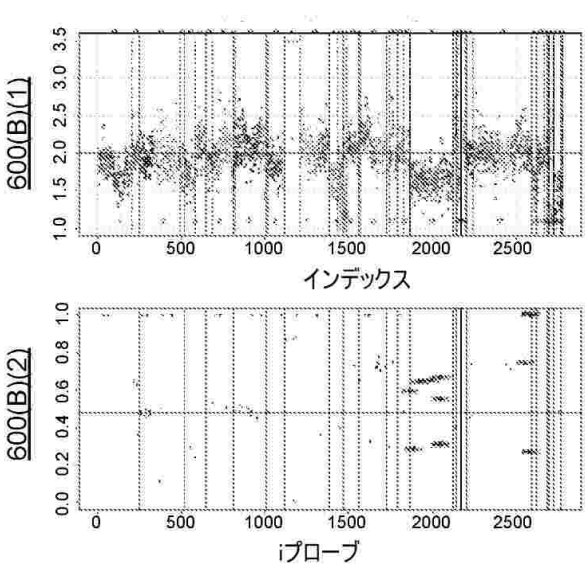


FIG. 6B

【図 7 A】

TND試料にわたるHet出現率

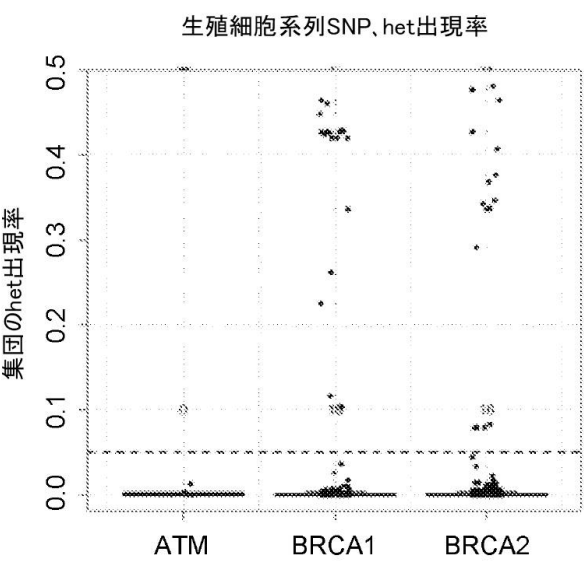


FIG. 7A

10

20

30

40

50

【図 7 B】

TND試料にわたるMAF

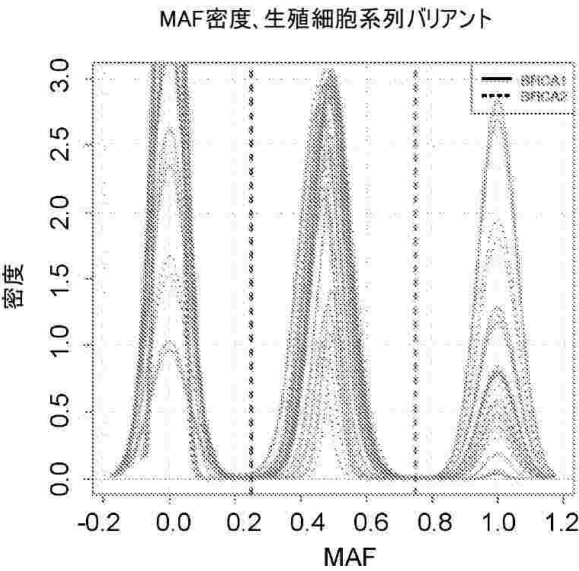


FIG. 7B

【図 8 A】

BRCA1

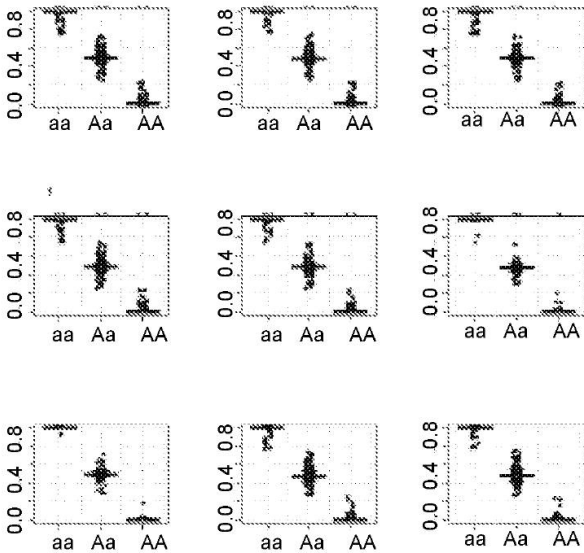


FIG. 8A

【図 8 B】

BRCA2

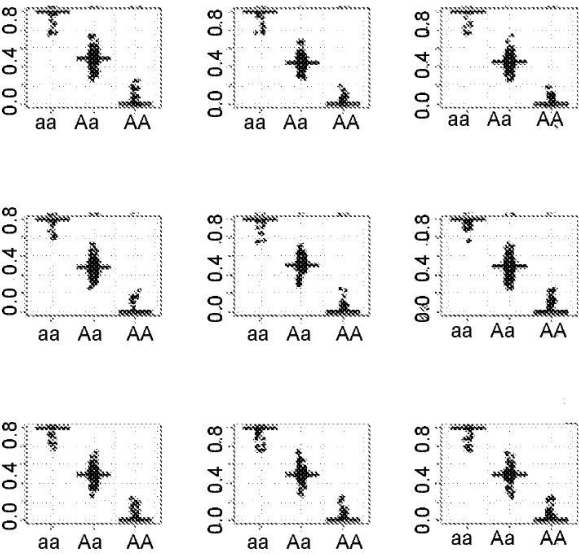


FIG. 8B

【図 9 A】

BRCA2 LLR スコアの比較

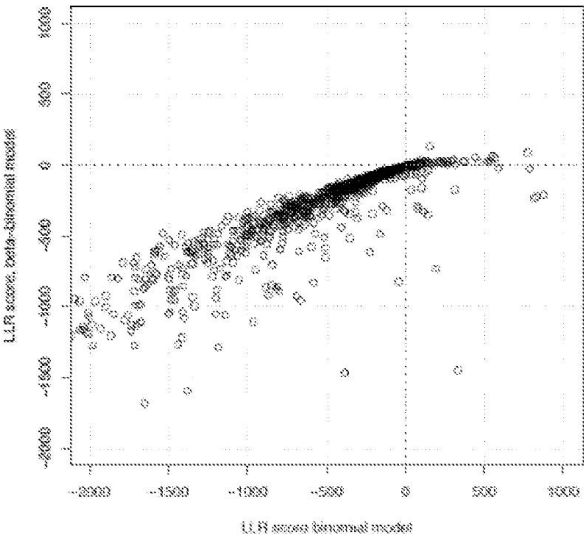


FIG. 9A

10

20

30

40

50

【図 9 B】

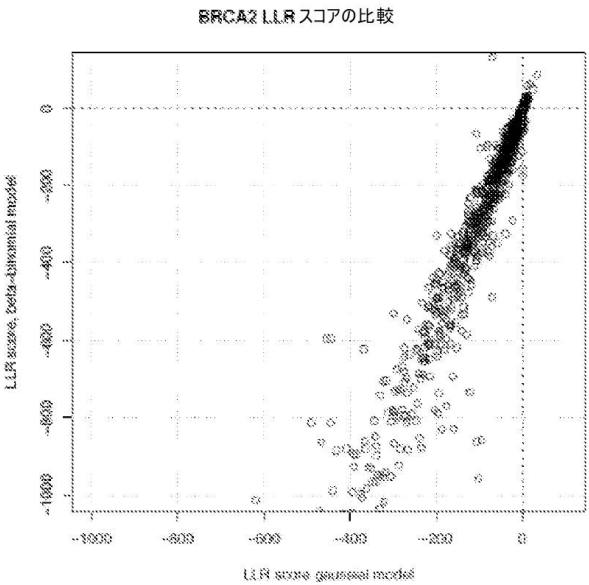


FIG. 9B

【図 10 A】

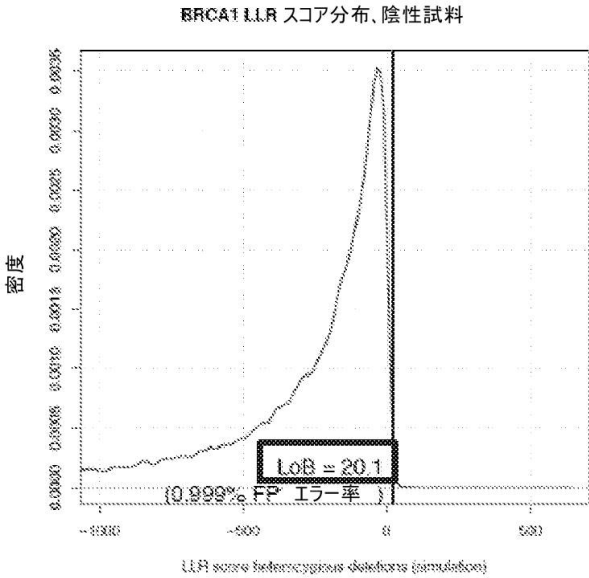


FIG. 10A

【図 10 B】

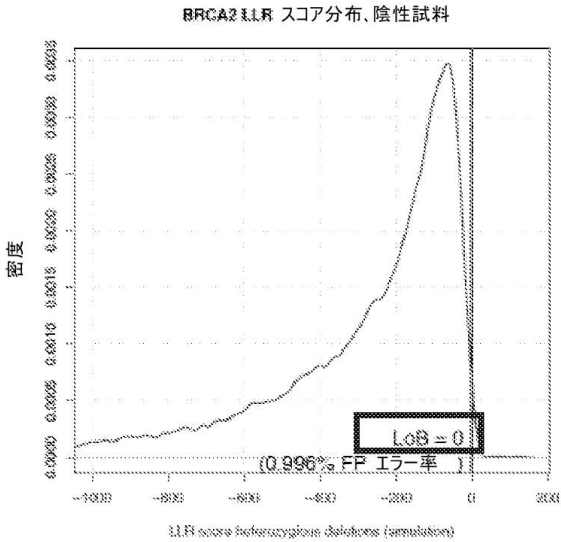


FIG. 10B

【図 11 A】

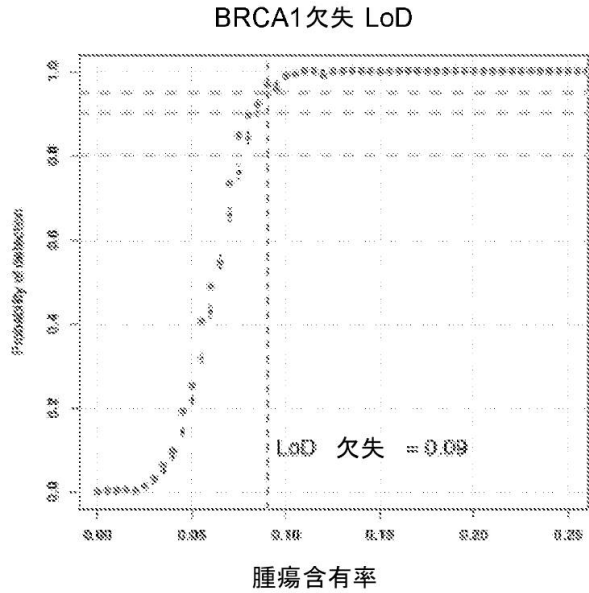


FIG. 11A

【図 1 1 B】

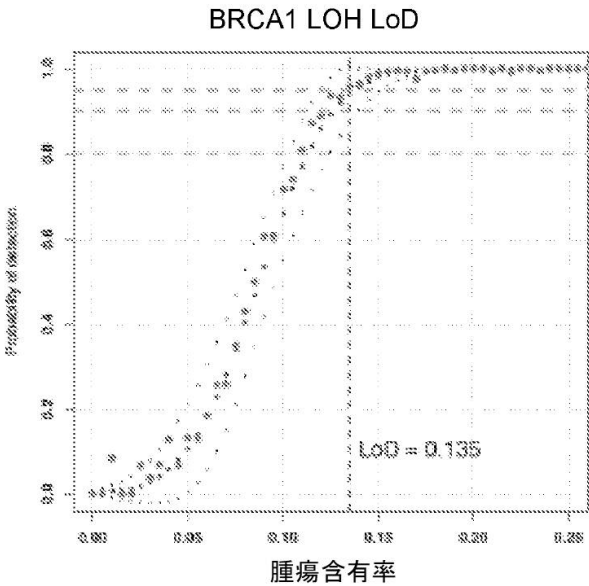


FIG. 11B

【図 1 2 A】

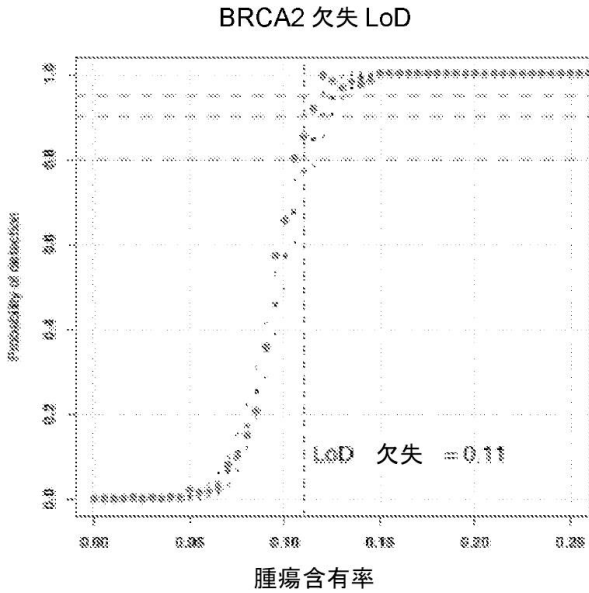


FIG. 12A

【図 1 2 B】

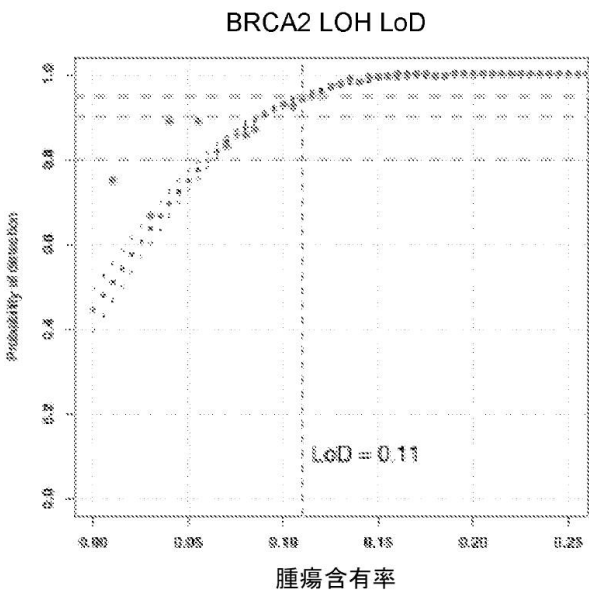
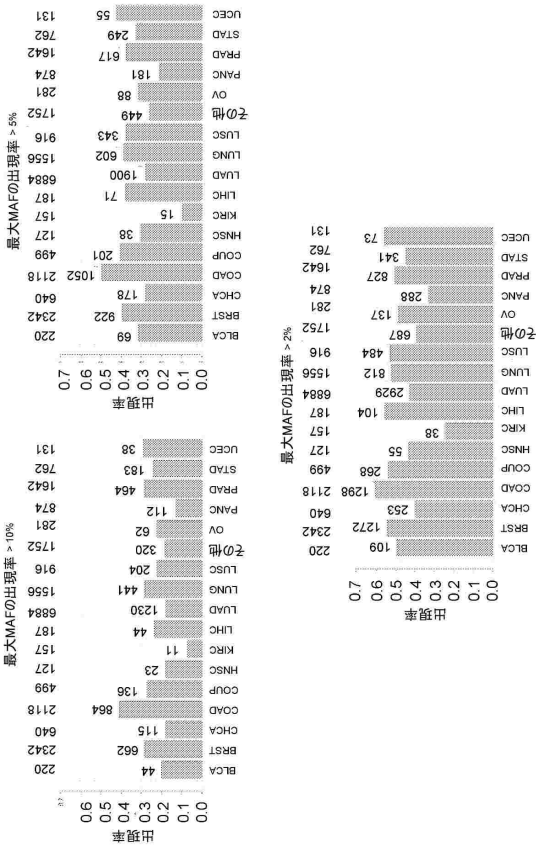


FIG. 12B

【図 1 3】



10

20

30

40

50

【図 1 4】

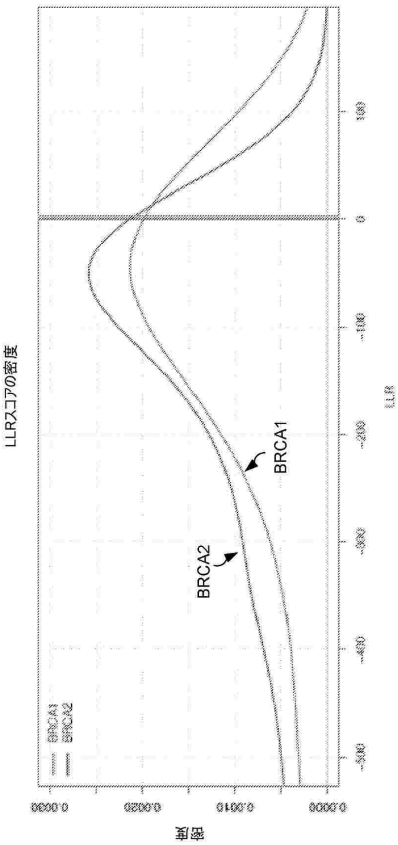


FIG. 14

【図 1 5】

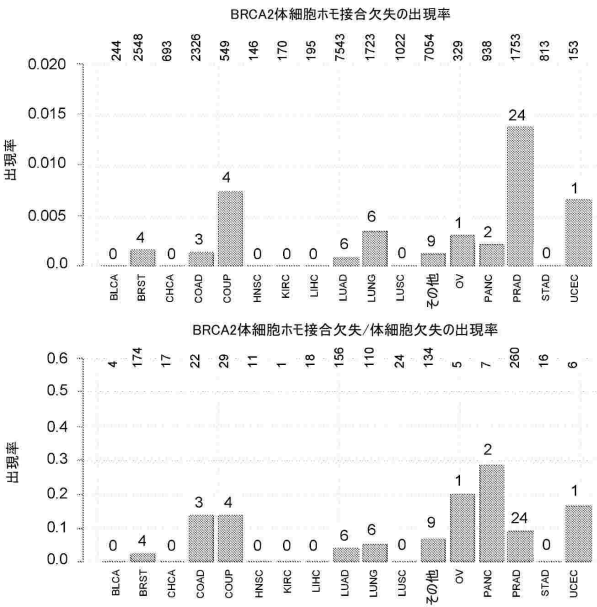


FIG. 15

【図 1 6】

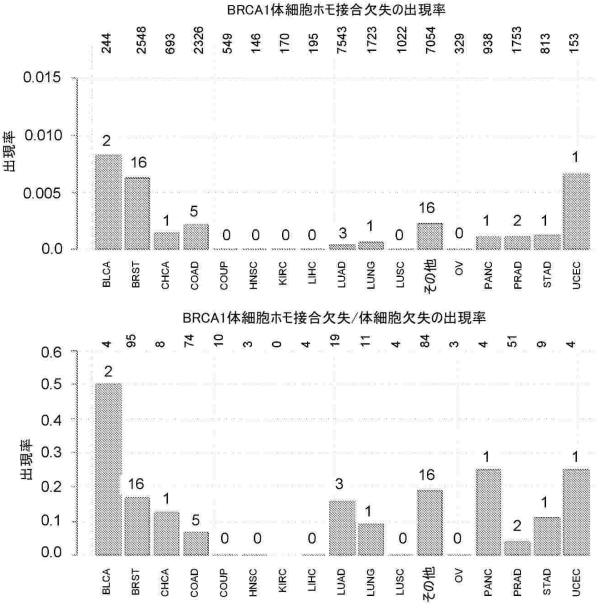


FIG. 16

【図 1 7】

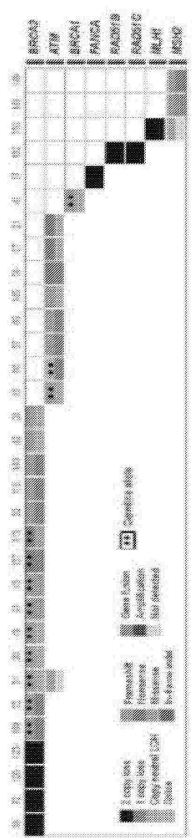


FIG. 17

10

20

30

40

50

【図 18 A】

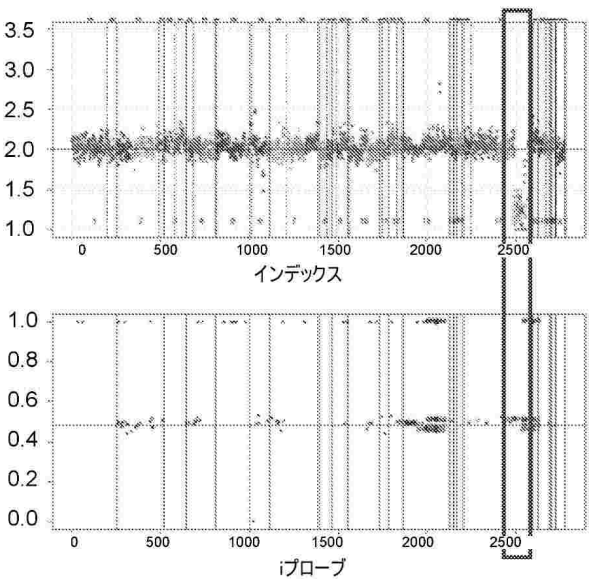


FIG. 18A

【図 18 B】

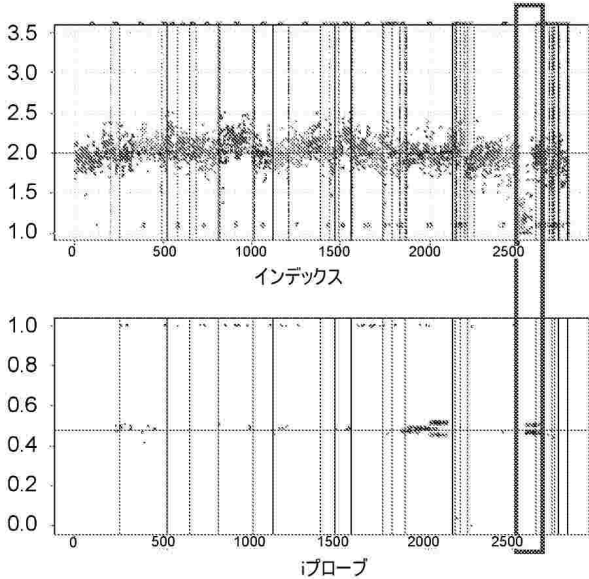


FIG. 18B

10

20

【図 19 A】

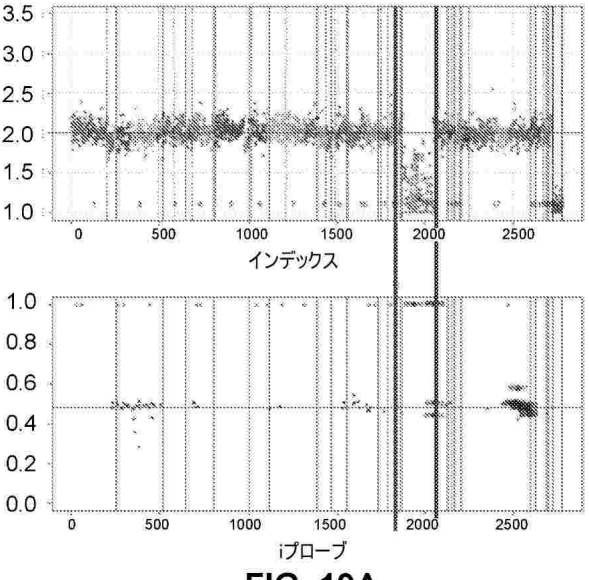


FIG. 19A

【図 19 B】

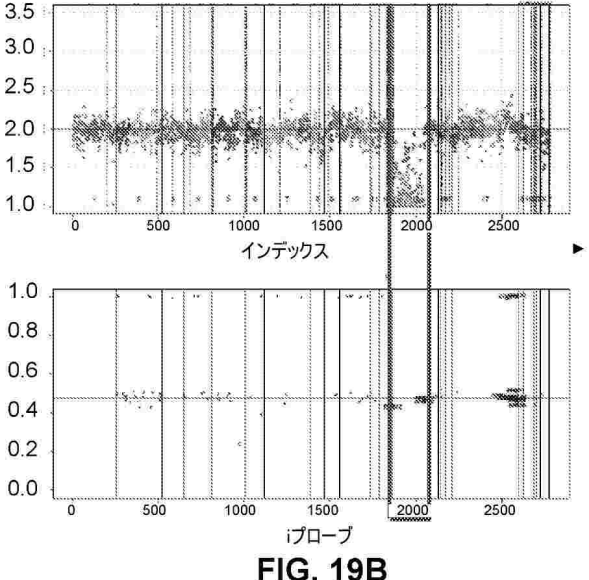


FIG. 19B

30

40

50

【図 20 A】

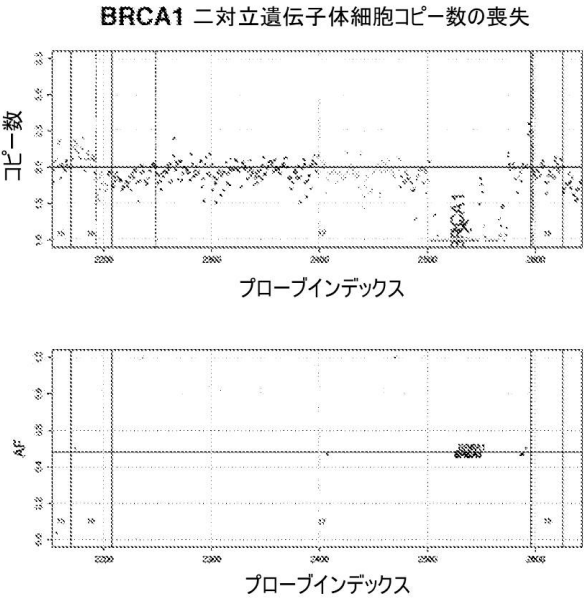


FIG. 20A

【図 20 B】

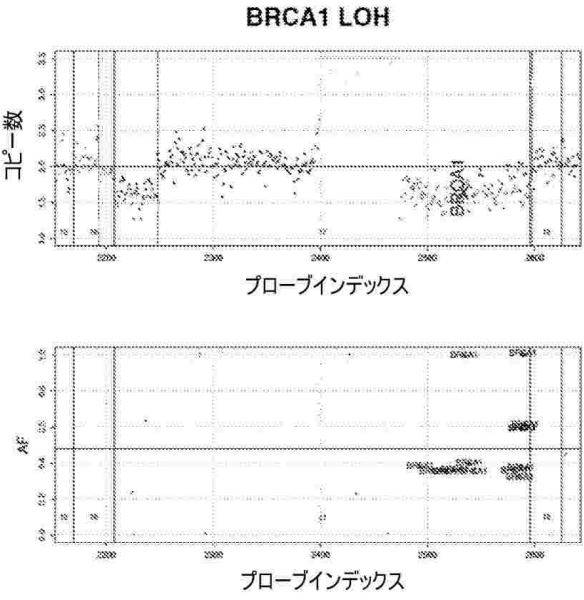


FIG. 20B

10

【図 21 A】

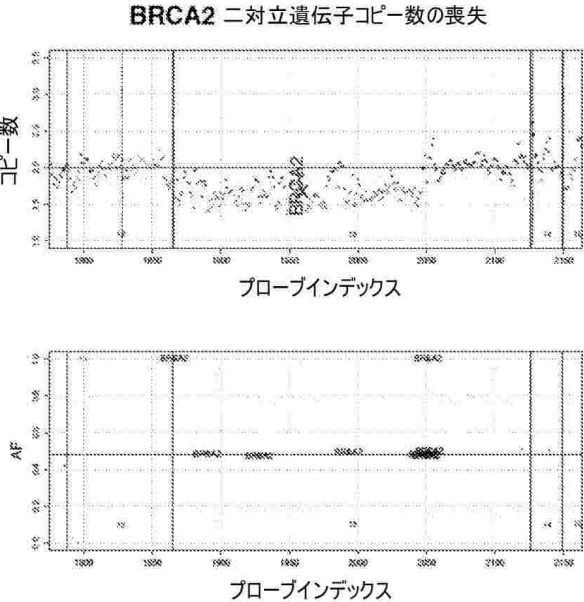


FIG. 21A

【図 21 B】

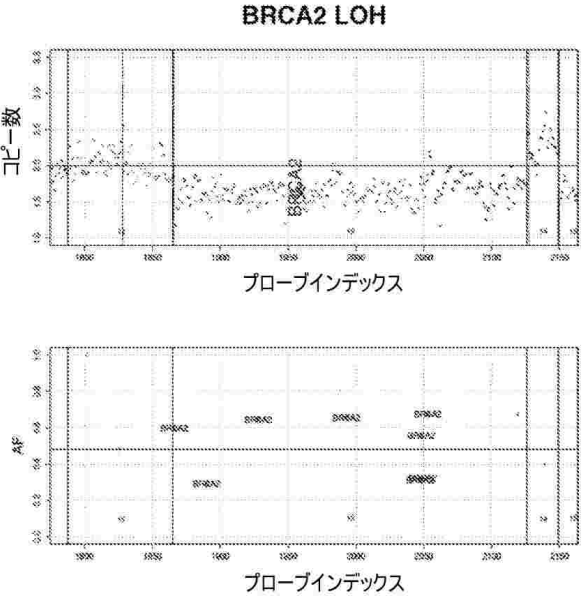


FIG. 21B

20

30

40

50

【図 22】

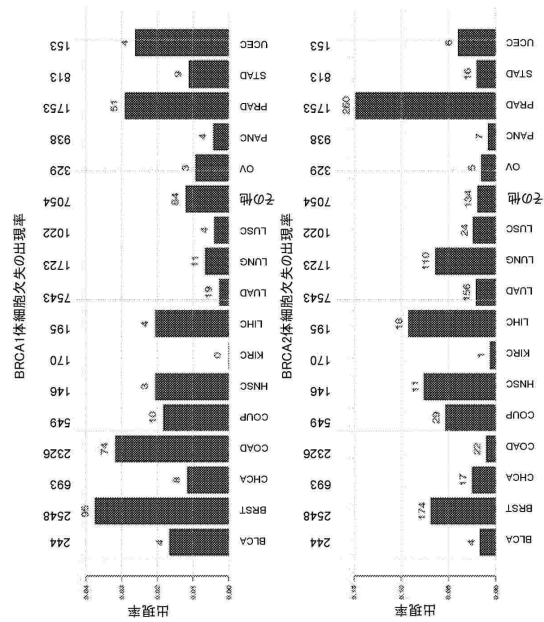


FIG. 22

10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

弁護士 山本 健策

(72)発明者 バルバチオル, カタリン

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

(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

(72)発明者 チュドヴァ, ダーリヤ

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 鈴木 和樹

(56)参考文献 米国特許出願公開第2018/0218113 (US, A1)

米国特許出願公開第2018/0373832 (US, A1)

国際公開第2018/144782 (WO, A1)

Xi Chen, 外5名, "PSSV: a novel pattern-based probabilistic approach for somatic structural variation identification", [online], 米国, 2016年09月21日, [令和5年11月20日検索], インターネット, <URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5254081/pdf/btw605.pdf>>

(58)調査した分野 (Int.Cl., DB名)

G16B 5/00-99/00

C12Q 1/68

C12M 1/00