

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2024-513781

(P2024-513781A)

(43)公表日 令和6年3月27日(2024.3.27)

(51)国際特許分類

G 1 6 B 20/20 (2019.01)

F I

G 1 6 B 20/20

審査請求 未請求 予備審査請求 未請求 (全55頁)

(21)出願番号	特願2023-558897(P2023-558897)	(71)出願人	522232488
(86)(22)出願日	令和4年4月1日(2022.4.1)		アキレス セラピューティクス ユーケー
(85)翻訳文提出日	令和5年11月21日(2023.11.21)		リミテッド
(86)国際出願番号	PCT/EP2022/058793		イギリス国 ダブリュ6 8 ピーダブリュ
(87)国際公開番号	WO2022/207925		ロンドン, ハマースミス ロード 2 4 5
(87)国際公開日	令和4年10月6日(2022.10.6)	(74)代理人	100079108
(31)優先権主張番号	2104715.4		弁理士 稲葉 良幸
(32)優先日	令和3年4月1日(2021.4.1)	(74)代理人	100109346
(33)優先権主張国・地域又は機関			弁理士 大貫 敏史
	英国(GB)	(74)代理人	100117189
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA		弁理士 江口 昭彦
	,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(	(74)代理人	100134120
	AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A		弁理士 内藤 和彦
	T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR	(72)発明者	ロス, アンドリュー
	,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,		イギリス国, オーエックス2 0ジェイ
	最終頁に続く		最終頁に続く

(54)【発明の名称】 クローナルなネオ抗原の同定及びその使用

(57)【要約】

【課題】対象の腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法が提供される。

【解決手段】この方法は、腫瘍遺伝子材料を含む対象からの1つ以上の試料からのシーケンスデータであって、1つ以上の試料の各々について含むシーケンスデータを提供すること、並びに腫瘍特異的突然変異がクローナルである尤度を、突然変異がクローナルである事前確率と、1つ以上の試料の各々についての腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したとき、腫瘍特異的突然変異が(i)クローナルである、及び(ii)非クローナルである場合にシーケンスデータが観察される確率とに依存する事後確率として決定することを含む。関連する方法、システム及び製品もまた記載される。

【選択図】図1

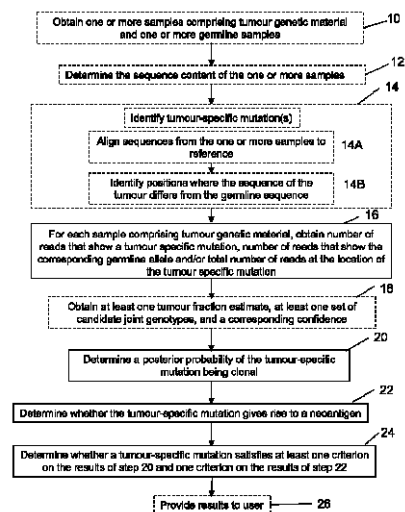


Fig. 1

【特許請求の範囲】

【請求項 1】

対象の腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法であって、

プロセッサに、腫瘍遺伝子材料を含む前記対象からの 1 つ以上の試料からのシーケンスデータであって、前記 1 つ以上の試料の各々について、前記試料中にある前記腫瘍特異的突然変異を示すリードの数 (d_b)、前記試料中にある対応する生殖細胞系列アレルを示すリードの数、及び前記腫瘍特異的突然変異の位置にある全リード数 (d) のうちの少なくとも 2 つを含むシーケンスデータを提供すること、並びに

前記プロセッサにより、前記腫瘍特異的突然変異がクローナルである尤度を、

10

前記突然変異がクローナルである事前確率と、

前記 1 つ以上の試料の各々についての腫瘍割合と、正常集団、前記腫瘍特異的突然変異を含まない参照腫瘍集団、及び前記腫瘍特異的突然変異を含む変異体腫瘍細胞集団についての、ある遺伝子型を前記腫瘍特異的突然変異の位置に各々含む 1 つ以上の候補結合遺伝子型と、を考慮したとき、前記腫瘍特異的突然変異が (i) クローナルである、及び (ii) 非クローナルである場合に前記シーケンスデータが観察される確率と

に依存する事後確率として決定すること

を含む方法。

【請求項 2】

各試料の腫瘍割合及び 1 つ以上の候補結合遺伝子型を考慮したときの前記シーケンスデータが観察される前記確率が、腫瘍割合、がん細胞割合、及び 1 つ以上の候補結合遺伝子型を考慮したときの前記シーケンスデータが観察される確率 ($Pr(d, d_b | \pi, \phi, t)$) に依存し、任意選択で腫瘍割合、がん細胞割合、及び 1 つ以上の候補結合遺伝子型を考慮したときの前記シーケンスデータが観察される前記確率が、腫瘍割合、がん細胞割合、及び前記 1 つ以上の候補結合遺伝子型の各々を考慮したときの前記シーケンスデータが観察される確率の加重和である、請求項 1 に記載の方法。

20

【請求項 3】

各試料の腫瘍割合及び 1 つ以上の候補結合遺伝子型を考慮したときの前記シーケンスデータが観察される前記確率が、各試料の前記がん細胞割合の可能な全ての値に関する積分として求められ、前記がん細胞割合が、前記腫瘍特異的突然変異を含む腫瘍細胞の比率である、請求項 1 又は 2 に記載の方法。

30

【請求項 4】

複数の試料からのシーケンスデータが入手され、及び前記複数の試料の各々についての腫瘍割合及び 1 つ以上の候補結合遺伝子型を考慮したときの前記シーケンスデータが観察される確率が、それぞれの試料中の前記腫瘍割合及び前記 1 つ以上の候補結合遺伝子型を考慮したときの各試料の前記シーケンスデータが観察される前記確率の積として求められる、請求項 1 ～ 3 のいずれか一項に記載の方法。

【請求項 5】

前記プロセッサにより、試料毎に、前記腫瘍割合の少なくとも 1 つの推定値、及び 1 つ以上の候補結合遺伝子型の少なくとも 1 つの対応する集合を求めることをさらに含む、請求項 1 ～ 4 のいずれか一項に記載の方法。

40

【請求項 6】

前記シーケンスデータが観察される前記確率が、前記 1 つ以上の試料の少なくとも 1 つについて、それぞれの腫瘍割合及び 1 つ以上の候補結合遺伝子型の対応する集合を考慮したときの、複数の前記シーケンスデータが観察される確率を組み合わせ、任意選択で前記方法が、前記プロセッサにより、前記 1 つ以上の試料の少なくとも 1 つについて、前記腫瘍割合の複数の推定値、及び 1 つ以上の候補結合遺伝子型の複数の対応する集合を取得することを含む、請求項 1 ～ 5 のいずれか一項に記載の方法。

【請求項 7】

前記突然変異がクローナルである前記事前確率が中立的事前値、又は事前データ及び／

50

又は専門知識から導出される値に設定される、請求項 1～6 のいずれか一項に記載の方法。

【請求項 8】

前記対象において同定された複数の腫瘍特異的突然変異に関して前記方法を繰り返すことをさらに含み、及び任意選択で、前記対象においてクローナルであるというそれらの決定された尤度に少なくとも一部には基づいて前記複数の腫瘍特異的突然変異をランク付けすること、又はその他何らかの優先順位を付けることをさらに含む、請求項 1～7 のいずれか一項に記載の方法。

【請求項 9】

前記プロセッサにより、腫瘍遺伝子材料を含む前記対象からの 1 つ以上の試料及び前記対象からの 1 つ以上の生殖細胞系列試料からのシーケンスデータを任意選択で使用して、前記対象の 1 つ以上の腫瘍特異的突然変異を同定することをさらに含む、請求項 1～8 のいずれか一項に記載の方法。

【請求項 10】

前記決定された前記腫瘍特異的突然変異がクローナルである確率及び／又はそれから導出される若しくはそれに関連する値を、例えばユーザインタフェースを介して、使用者に提供することをさらに含む、請求項 1～9 のいずれか一項に記載の方法。

【請求項 11】

対象の 1 つ以上のクローナルなネオ抗原を同定する方法であって、
前記対象の複数の腫瘍特異的突然変異を同定すること；
請求項 1～10 のいずれか一項に記載の方法を用いて、1 つ以上の前記腫瘍特異的突然変異が前記対象においてクローナルである可能性があるかどうかを決定すること；及び
前記 1 つ以上の前記腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかを決定すること
を含む方法において、
クローナルなネオ抗原が、前記腫瘍特異的突然変異がクローナルである可能性があるかどうかに関する 1 つ以上の所定の基準及び前記腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかに関する 1 つ以上の基準を満たす腫瘍特異的突然変異である、方法。

【請求項 12】

クローナルなネオ抗原が、所定の閾値を上回るクローナルである確率を有すること、確率を決定した際の対象であった前記腫瘍特異的突然変異の中でクローナルである確率が最も高い所定の数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである確率を有すること、及び確率を決定した際の対象であった前記腫瘍特異的突然変異の中で所定の最上位の百分位数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである確率を有すること、から選択される基準を少なくとも満たす腫瘍特異的突然変異である、請求項 11 に記載の方法。

【請求項 13】

クローナルなネオ抗原が、腫瘍細胞に発現する発現産物に関連していること、前記対象の前記正常細胞に発現しないタンパク質又はペプチドを生じさせると予測されること、MHC 分子、好ましくは前記対象に存在することが既知の MHC アレルによって提示される可能性がある少なくとも 1 つのペプチドを生じさせると予測されること、及び免疫原性のタンパク質又はペプチドを生じさせると予測されること、から選択される基準を少なくとも満たす腫瘍特異的突然変異である、請求項 11 又は 12 に記載の方法。

【請求項 14】

がんであると診断された対象に予後を提供する方法であって、前記対象からの 1 つ以上の試料に複数の腫瘍特異的突然変異を同定すること、及び請求項 1～9 のいずれか一項に記載の方法を用いて前記腫瘍特異的突然変異の各々がクローナルである尤度を決定すること、を含む方法。

【請求項 15】

がんであると診断された対象に免疫療法を提供する方法であって、

請求項 1 1 ～ 1 3 のいずれか一項に記載の方法を用いて 1 つ以上のクローナルなネオ抗原を同定すること；及び

同定された前記 1 つ以上の前記クローナルなネオ抗原を標的化する免疫療法を設計すること

を含む方法。

【請求項 1 6】

前記 1 つ以上の前記クローナルなネオ抗原を標的化する前記免疫療法が、免疫原性組成物、免疫細胞を含む組成物又は治療用抗体である、請求項 1 5 に記載の方法。

【請求項 1 7】

前記免疫療法が、同定された前記 1 つ以上の前記クローナルなネオ抗原のうちの少なくとも 1 つを認識する T 細胞を含む組成物である、請求項 1 6 に記載の方法。

【請求項 1 8】

前記組成物が、同定された前記 1 つ以上の前記クローナルなネオ抗原のうちの少なくとも 1 つを標的化する T 細胞に関して濃縮され、任意選択で前記方法が、T 細胞集団を入手すること、及び同定された前記 1 つ以上の前記クローナルなネオ抗原のうちの少なくとも 1 つを標的化する T 細胞の数又は相対的比率が増加するように前記 T 細胞集団を拡大することを含む、請求項 1 7 に記載の方法。

【請求項 1 9】

請求項 1 8 に記載の方法によって入手される又は入手可能な T 細胞の集団を含む組成物。

【請求項 2 0】

がんであると診断された対象を治療する方法であって、請求項 1 5 ～ 1 8 のいずれか一項に記載の方法を使用して提供された免疫療法、又は請求項 1 9 に記載の組成物を投与することを含む方法。

【請求項 2 1】

プロセッサ；及び

前記プロセッサによる実行時に、請求項 1 ～ 1 8 のいずれか一項に記載の方法のステップを前記プロセッサに実施させる命令を含むコンピュータ可読媒体を含むシステム。

【請求項 2 2】

1 つ以上のプロセッサによる実行時に、請求項 1 ～ 1 8 のいずれか一項に記載の方法のステップを前記 1 つ以上のプロセッサに実施させる命令を含む 1 つ以上の非一時的コンピュータ可読媒体。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

開示の分野

本開示は、腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法及び腫瘍に存在する腫瘍特異的突然変異に由来するクローナルなネオ抗原を同定する方法に関する。本開示はまた、同定されたクローナルなネオ抗原を利用する又はそれを標的化するがんの治療のための方法及び組成物にも関する。

【背景技術】

【0 0 0 2】

背景

がん細胞は突然変異を獲得することが知られ、そうした突然変異の一部は、進化上の利点を付与し得る。結果として、腫瘍は多くの場合に、遺伝子型が明白に異なる複数の関連性のある集団（即ちクローン）を含む。腫瘍のクローナル組成を特徴付けることは、治療の文脈では特に重要である。実際、腫瘍細胞集団のうち一部の集合のみに存在する突然変異（「サブクローナル」な突然変異とも称される）を標的化することに伴う臨床的有益性

10

20

30

40

50

は、この療法が集団の一部しか標的化しないため限られたものであり、影響を受けないクローンは依然として増殖能があるため、再発又は転移の可能性が高いことになり得る。代わりに、腫瘍を有効に管理するには、クローナルなネオ抗原（あらゆる腫瘍細胞に存在する突然変異が存在する結果として発現する抗原）を標的化するか、又は複数の標的化療法を組み合わせる必要があり得るという考えが増えつつある（McGranahan et al., 2015）。加えて、クローナルなネオ抗原の負荷量は、少なくとも一部のがんで予後に関連し、及びチェックポイント阻害薬による治療への感受性に関連することが知られている（McGranahan et al., 2016; Litchfield et al., 2021）。

【発明の概要】

【課題を解決するための手段】

【0003】

概要

本発明者らは、腫瘍細胞又はそれに由来する遺伝子材料を含む1つ以上の試料からのシーケンスデータを使用して腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定するための、先行技術の手法の問題点の1つ以上に対処する新規方法を開発した。この方法は、例えばがん療法又は予後判定を目的とした、クローナルなネオ抗原の同定に特に利用が見出される。本方法は、厳密な統計的フレームワークを使用して個別の突然変異をクローナルに分類し、その割り当ての信頼度を反映する確率を提供する。本方法は、高速、柔軟、ロバストで、且つ再現性があり、解釈可能な仮定に依存し、体細胞性コピー数異常データを柔軟に取り入れることができ、その予測を提供する際に倍数性／純度の解が複数あることに起因するコピー数コールの不確かさを考慮に入れることができる。

【0004】

このように、第1の態様によれば、対象の腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法であって、腫瘍遺伝子材料を含む対象からの1つ以上の試料からシーケンスデータを提供すること、又は入手することであって、シーケンスデータが、1つ以上の試料の各々について、試料中にある腫瘍特異的突然変異を示すリードの数（ d_b ）、試料中にある対応する生殖細胞系列アレルを示すリードの数、及び腫瘍特異的突然変異の位置にある全リード数（ d ）のうちの少なくとも2つを含むこと；及び腫瘍特異的突然変異がクローナルである尤度を、突然変異がクローナルである事前確率と、1つ以上の試料の各々についての腫瘍割合、並びに正常集団、腫瘍特異的突然変異を含まない参照腫瘍集団及び腫瘍特異的突然変異を含む変異体腫瘍細胞集団についての、ある遺伝子型を腫瘍特異的突然変異の位置に各々含む1つ以上の候補結合遺伝子型と、を考慮したときの、腫瘍特異的突然変異が（ i ）クローナルである、及び（ i ）非クローナルである場合にシーケンスデータが観察される確率とに依存する事後確率として決定することを含む方法が提供される。

【0005】

本方法では、突然変異がクローナルである確率（ $P(Z=1)$ ）を、突然変異がクローナルである事前確率（ p ）と、シーケンスデータが観察される確率（シーケンスデータが観察される「尤度」、又は単純にシーケンスデータの「尤度」とも称される）とに依存する事後確率（ $p(Z=1 | d_b, d, \pi, t, p)$ ）として求める。従ってこれは、単純に調査中のその突然変異に関係するシーケンスデータを使用して、任意の突然変異について別個に求めることができ、明示的な仮定（即ち、突然変異がクローナルである事前確率、及び集団構造の明示的なモデルを考慮したときのシーケンスデータが観察される尤度）に依存する容易に解釈可能な出力が生み出される。換言すれば、かかる確率は、検証することのできる厳密な一組の仮定によって強調されるベイズのフレームワークを通じて、腫瘍遺伝子材料を含む1つ以上の試料から利用可能なデータ、及び利用可能な任意の事前知識の両方に依存する。この出力を使用すると、任意の数の突然変異を比較すること、例えば優先順位を付けることができ、それが全ゲノムを網羅する必要はない。これはまた、腫瘍遺伝子材料を含む複数の試料からのエビデンスを組み合わせる能力も有するが、遺伝子材料を含む単一の試料を使用して決定することも等しくできる。

10

20

30

40

50

【0006】

本態様の方法は、以下の特徴のうちの1つ以上を有し得る。

【0007】

本方法は、コンピュータ実装することができる。従って、シーケンスデータを入手するステップはプロセッサによって実施されてもよく、及び腫瘍特異的突然変異がクローナルである尤度を決定するステップは前記プロセッサによって実施されてもよい。シーケンスデータを入手するステップは、対象からの1つ以上の試料からシーケンスリードを含むシーケンスデータを受け取ること、及び前記シーケンスリードから、試料中にある腫瘍特異的突然変異を示すリードの数 (d_b)、試料中にある対応する生殖細胞系列アレルを示すリードの数、及び腫瘍特異的突然変異の位置にある全リード数 (d) のうちの少なくとも2つを決定することを含み得る。少なくとも腫瘍特異的突然変異がクローナルである尤度を決定するステップは、コンピュータ実装することができる。腫瘍特異的突然変異がクローナルである尤度を決定するステップは、事後確率を求めるための数値積分ステップを含み得る。詳細には、このステップは、突然変異がクローナルである事前確率と、腫瘍特異的突然変異が (i) クローナルである、及び (ii) 非クローナルである場合にシーケンスデータが観察される確率とを考慮したときの、0~1の間の可能な全てのがん細胞割合に関してシーケンスデータが観察される確率を積分する複数の一次元積分 (例えば、それぞれ突然変異がクローナルである、及び非クローナルであるという仮定を表す各試料についての一対の積分など) を解くことにより、突然変異がクローナルであるという事後確率を決定することを含み得る。これらの数値積分は、試料毎及び突然変異毎に独立に (例えば並行してなど) 解くことができる。提供するステップは1つ以上のステップを含んでもよく、その全て又は一部がコンピュータ実装される。

10

20

【0008】

腫瘍特異的突然変異がクローナルである確率は、突然変異がクローナルである事前確率を所与としたときの突然変異がクローナルのカテゴリに割り当てられる事前確率 ($P(Z=1 | p) = p$) と; 突然変異がクローナルである事前確率を所与としたときの突然変異が非クローナルのカテゴリに割り当てられる事前確率 ($P(Z=0 | p) = (1-p)$) とによる、突然変異がクローナルである事前確率 (p) に依存し得る。腫瘍特異的突然変異がクローナルである場合に (1つ以上の試料の各々についての腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときの) シーケンスデータが観察される確率は、がん細胞割合に関して周辺化されてもよい。同様に、1つ以上の試料の各々についての腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときの、腫瘍特異的突然変異がクローナルでない場合にシーケンスデータが観察される確率は、がん細胞割合に関して周辺化されてもよい。

30

【0009】

腫瘍特異的突然変異がクローナルである確率は、突然変異がクローナルである事前確率を所与としたときの突然変異がクローナルのカテゴリに割り当てられる事前確率 ($P(Z=1 | p) = p$) に、突然変異がクローナルである場合に、腫瘍割合、及び1つ以上の候補結合遺伝子型を考慮したときの各試料のシーケンスデータが観察される確率 (これは、がん細胞割合に関して周辺化した、各試料のシーケンスデータの尤度、 ψ_1 として計算することができる) を乗じたもの; 及び突然変異がサブクローナルである事前確率を所与としたときの突然変異が非クローナルのカテゴリに割り当てられる事前確率 ($P(Z=0 | p) = (1-p)$) に、突然変異が非クローナルである場合に、腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときの各試料のシーケンスデータが観察される確率 (これは、がん細胞割合に関して周辺化した、各試料のシーケンスデータの尤度、 ψ_0 として計算することができる) を乗じたものに依存し得る。

40

【0010】

腫瘍特異的突然変異がクローナルである確率は、(i) 突然変異がクローナルである事前確率を所与としたときの突然変異がクローナルのカテゴリに割り当てられる事前確率に、突然変異がクローナルである場合に、腫瘍割合及び1つ以上の候補結合遺伝子型を考

50

慮したときの各試料のシーケンスデータが観察される確率を乗じたもの ($p(d_b, d, Z=1 | \pi, t, \rho)$ 、これは $p\psi_1$ と表すことができる) を、 (i, i) (i) (即ち $p(d_b, d, Z=1 | \pi, t, \rho)$) と、突然変異がサブクロナルである事前確率を所与としたときの突然変異が非クロナルのカテゴリーに割り当てられる事前確率に、突然変異が非クロナルである場合に、各試料の腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときの各試料中にシーケンスデータが観察される確率を乗じたもの ($p(d_b, d, Z=0 | \pi, t, \rho)$ 、これは $(1-\rho)\psi_0$ と表すことができる) との和で除した比として求めることができる。

【0011】

突然変異がクロナルである確率は、式 (11a) を用いて求めることができる。式 (11a) では、項 $Pr(d_b, d | \pi, \phi, t)$ は、式 (3)、式 (4)、式 (3a)、式 (4a)、式 (3b) 又は式 (4b) のいずれかによって与えられてもよい。式 (11) では、項 $p(\phi | Z=0)$ 及び $p(\phi | Z=1)$ は、式 (6) によって与えられてもよい。

【0012】

クロナルな突然変異とは、腫瘍遺伝子材料を含む対象からの1つ以上の試料中にある全て又は本質的に全ての腫瘍細胞に (又は1つ以上の試料中の腫瘍遺伝子材料の全てに) 存在する突然変異であり得る。かかる突然変異は、対象の全ての腫瘍細胞に存在し得るか、又は存在すると仮定し得る (これに関して完全に確実とするには、対象の全ての腫瘍細胞のシーケンシングを伴い得るが、1つ以上の試料中の本質的に全ての細胞に存在することを、このことの指標として用い得るため)。

【0013】

各試料の腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率は、腫瘍割合、がん細胞割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率 ($Pr(d, d_b | \pi, \phi, t)$) に依存し得る。腫瘍割合、がん細胞割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率は、腫瘍割合、がん細胞割合及び1つ以上の候補結合遺伝子型の各々を考慮したときのシーケンスデータが観察される確率の加重和であってもよい。

【0014】

有利には、シーケンスデータが観察される確率 (シーケンスデータの尤度) は、複数の候補遺伝子型に関して計算されてもよく (例えば各候補遺伝子型についての項を含む確率の和として計算される、例えば式 (3a)、式 (3b) を参照のこと)、その寄与には、例えば候補遺伝子型の相対的確率に関する事前知識 (例えば、一部の遺伝子型が現れる可能性が他よりも高いかどうかに関する任意の事前知識) を反映して重みが付けられてもよい。かかる事前知識が利用可能でない、又は望ましくないときは、候補遺伝子型の各々の確率には等しい重みが付けられてもよい。それぞれの候補遺伝子型の重みは、好適には、考察される異なる候補結合遺伝子型の相対的寄与を反映した全確率となるように、合計が1になるように考察した。単一の候補結合遺伝子型が使用されるとき、それには1の重みを割り当てることができる (即ち、和は求めないことになり得る)。

【0015】

腫瘍割合、がん細胞割合及び特定の候補結合遺伝子型 (G_i) を考慮したときのシーケンスデータが観察される確率 (これは、がん細胞割合に関して周辺化した、各試料のシーケンスデータの尤度、 ψ_z として計算することができる) は、母数 d_b 及び $\xi(G_i, \phi, t)$ の二項分布を用いて求めることができる。或いは、腫瘍割合、がん細胞割合及び特定の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率は、母数 d_b 、 $\xi(G_i, \phi, t)$ 、及び γ のベータ二項分布を用いて求めることができる。両方の場合とも (即ち二項分布又はベータ二項分布のいずれが用いられるにしろ)、 $\xi(G_i, \phi, t)$ は、特定の遺伝子型 G_i 、がん細胞割合 ϕ 及び腫瘍純度 t を仮定したときの変異体アレルを含むリードが抽出される確率を表し得る。確率 $\xi(G_i, \phi, t)$ は、正常、変異体及び参照遺伝子型の各々についての全コピー数、その遺伝子座における遺伝子型及び

10

20

30

40

50

シーケンシングエラー率の点で変異体であるアレルの比率を考慮したときの遺伝子型 G_i の集団から変異体を含むリードが抽出される確率、試料中の腫瘍割合及びその突然変異についてのがん細胞割合の関数として求めることができる。

【0016】

各試料の腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率は、各試料のがん細胞割合の可能な全ての値に関する積分として求めることができ、ここでがん細胞割合とは、腫瘍特異的突然変異を含む腫瘍細胞の比率である。従って、腫瘍特異的突然変異がクローナルである尤度を決定するステップは、プロセッサを使用して前記積分を数値積分することを含み得る。

【0017】

がん細胞割合 (ϕ) は、0～1の間の値をとり得る。換言すれば、突然変異がクローナル又は非クローナルである場合に、各試料の腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率は、がん細胞割合の可能な全ての値に関してがん細胞割合に依存する値を積分する（即ちがん細胞割合に関して周辺化する）ことによって求めることができる。がん細胞割合に依存する値は、 $\Pr(d_b, d | \pi, \phi, t) p(\phi | Z = z)$ と表すことができ、式中、第1項は、腫瘍割合、がん細胞割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率であり、第2項は、突然変異がクローナル又は非クローナルに分類される場合（それぞれ、 $Z = 1$ 又は $Z = 0$ ）のがん細胞割合の事前確率（即ち、クローナル／非クローナルな突然変異についてがん細胞割合がどのように挙動すべきかの仮定に基づく確率）である。このように、各試料の腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率は、

【数1】

$$\int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = z) d\phi$$

として求めることができる。

【0018】

突然変異がクローナルに分類される場合の特定のがん細胞割合の事前確率は、母数 α （1より大きい値、例えば、99に設定され、但し、いずれか他の値が使用されてもよい）及び $\beta = 1$ のベータ分布（ベータ ($\phi | \alpha, 1$)）として定義することができる。突然変異が非クローナルに分類される場合の特定のがん細胞割合の事前確率は、母数 $\alpha = 1$ 及び $\beta = 1$ のベータ分布（ベータ ($\phi | 1, 1$)）として定義することができる。

【0019】

複数の試料からシーケンスデータを入手してもよく、複数の試料の各々についての腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときのシーケンスデータが観察される確率は、それぞれの試料の腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときの各試料のシーケンスデータが観察される確率の積として求めることができる。

【0020】

有利には、本方法は、複数の試料が利用可能な場合に、それらから入手された突然変異のクローナリティーを裏付ける／否定する証拠をシームレスに統合することが可能である。多領域シーケンシングデータを利用することのできる腫瘍細胞集団のクローナルな構造を推測する方法は、ベンチマーク試験で性能が特に良好であることが示されているため（Farahani et al., 2017）、これは特に有利である。

【0021】

本方法は、試料毎に、腫瘍割合の少なくとも1つの推定値、及び1つ以上の候補結合遺伝子型の少なくとも1つの対応する集合を求めること又はそれを提供することをさらに含み得る。腫瘍割合推定値は、腫瘍細胞と正常細胞との混合物を含む試料におけるアレル特異的コピー数プロファイルを決定する方法を用いて求めることができる。シーケンシング又はアレイデータを用いてこれを行う方法は、当該技術分野において公知であり、例えば

10

20

30

40

50

、アレルト異的データをアレルト異的コピー数、腫瘍異数性及び腫瘍細胞割合を含む媒介変数の関数として表し、及び全てのデータにベストフィットするそれらの媒介変数の値を同定することによる。かかる方法の例としては、例えば、数ある中でも、ASCAT (Van Loo et al., 2010) が挙げられる。或いは、腫瘍割合推定値は実験的に決定されてもよい。このように、本方法は、1つ以上の試料の各々について腫瘍割合推定値を求めることをさらに含み得る。詳細には、本方法は、プロセッサにより、試料毎に、腫瘍割合の少なくとも1つの推定値を求めることであって、シーケンスデータを使用して腫瘍割合及びアレルト異性コピー数の推定値を決定するプロセッサを含む、及び前記プロセッサにより、前記アレルト異的コピー数に関連する1つ以上の候補結合遺伝子型の集合を決定することを含み得る。

10

【0022】

1つ以上の候補遺伝子型の集合は、混合試料中の腫瘍細胞についてのアレルト異的コピー数又はそれから導出される変数（又は逆に、Bアレルト割合及び対数Rなど、そこからかかるアレルト異的コピー数が導出されてもよい）を用いて求めることができる。混合試料中の腫瘍細胞についてのアレルト異的コピー数は、例えば、数ある中でも、ASCAT (Van Loo et al., 2010)、又はascatNgs (Raine et al., 2016) など、腫瘍細胞と正常細胞との混合物を含む試料でアレルト異的コピー数プロファイルを決定する方法を用いて求めることができる。

【0023】

このように、本方法は、1つ以上の試料の各々について、試料中の腫瘍細胞におけるメジャーアレルのコピー数、試料中の腫瘍細胞におけるマイナーアレルのコピー数、及び試料中の腫瘍細胞における腫瘍特異的突然変異の位置にある全コピー数のうちの少なくとも2つについての推定値を求めることをさらに含み得る。試料中の腫瘍細胞におけるコピー数の推定値は、試料中の腫瘍細胞の集団全体に関する要約された（例えば平均）推定値を表し得る。

20

【0024】

1つ以上の候補結合遺伝子型の集合は、正常集団のみが1つ又は複数の正常アレルAを含む（即ち $G_H = A A$ 又は A 、例えばその遺伝子座が性染色体上にある場合）；参照集団は変異体アレルBを含まない（即ち $G_R = (A) * n$ ）；及び変異体集団は変異体アレルBを少なくとも1コピー含む（即ち $G_V = (A) * m (B) * l$ ）という仮定に適合する候補結合遺伝子型として求めることができる。

30

【0025】

有利には、候補遺伝子型の集合は、(i) 参照集団遺伝子型が正常集団遺伝子型と一致し、変異体集団が、その位置にある全コピー数に等しく、最大で変異体アレルのメジャーコピー数に至るまでのコピー数を有するか；又は(ii) 参照集団は、その位置にある全コピー数に等しいコピー数を有し、変異体集団は1の変異体アレル及びその位置にある全コピー数に等しいコピー数を有するか、いずれかの仮定にさらに適合する候補結合遺伝子型を含み得る（「メジャーコピー数事前」）。この手法は、有利には、集団の遺伝子型の不確かさを考慮することと、一方で多くの状態を考慮しすぎないようにすることとの間を上手く両立させる。

40

【0026】

代わりに、又はそれに加えて、1つ以上の候補結合遺伝子型の集合は、各突然変異が二倍体且つヘテロ接合体である（即ち $G_V = A B$ 、 $G_R = A A$ ）（「AB事前」）という仮定に適合する候補結合遺伝子型のいずれかを含んでもよい。代わりに、又はそれに加えて、1つ以上の候補結合遺伝子型の集合は、各突然変異が二倍体且つホモ接合体である（即ち $G_V = B B$ 、 $G_R = A A$ ）（「BB事前」）という仮定に適合する候補結合遺伝子型のいずれかを含んでもよい。代わりに、又はそれに加えて、1つ以上の候補結合遺伝子型の集合は、変異体集団の遺伝子型が突然変異の領域に予測される全コピー数を有し、厳密に1つが突然変異体アレルである（即ち $G_V = (A) * m B$ 、ここで $m = \text{全コピー数} - 1$ ）（「接合性なし事前」）という仮定に適合する候補結合遺伝子型のいずれかを含んでもよ

50

い。代わりに、又はそれに加えて、1つ以上の候補結合遺伝子型の集合は、変異体集団の遺伝子型が、突然変異の領域に予測される全コピー数を有し、少なくとも一方が突然変異体アレルであり、及び参照集団がAAであるか、又は予測される全コピー数に等しいコピー数の、且つ変異体アレルがない遺伝子型であるかのいずれかである（即ち $G_R = (A)^n$ 、ここで n は全コピー数である、 $G_V = (A)^m (B)^l$ 、ここで $m + l = n$ 及び $l > 1$ ）（「全コピー数事前」という仮定に適合する候補結合遺伝子型のいずれかを含んでもよい。代わりに、又はそれに加えて、1つ以上の候補結合遺伝子型の集合は、変異体集団の遺伝子型が、メジャーコピー数又はマイナーコピー数のいずれか一方に対応する数の突然変異体アレルを有する（「親モード」という仮定に適合する候補結合遺伝子型のいずれかを含んでもよい。

10

【0027】

シーケンスデータが観察される確率は、1つ以上の試料のうちの少なくとも1つについてのそれぞれの腫瘍割合及び対応する1つ以上の候補結合遺伝子型の集合を考慮したときのシーケンスデータが観察される確率を複数組み合わせてもよく、任意選択で、この方法は、1つ以上の試料のうちの少なくとも1つについて、腫瘍割合の複数の推定値、及び1つ以上の候補結合遺伝子型の複数の対応する集合を求めることを含む。このように、本方法は、少なくとも1つの試料について、腫瘍割合の複数の推定値を求めることを含み得る。これは、プロセッサにより、シーケンスデータに適合する腫瘍割合の複数の推定値及び対応する複数のアレル特異的コピー数を決定すること、及びプロセッサにより、前記複数のアレル特異的コピー数に関連する1つ以上の候補結合遺伝子型の複数の集合を決定することを含み得る。

20

【0028】

本方法は、有利には、突然変異がクローナルである確率であって、複数の可能な腫瘍割合及び対応する候補結合遺伝子型集合を考慮に入れている確率を決定することが可能である。換言すれば、本方法は、突然変異がクローナルである確率であって、腫瘍割合及び候補結合遺伝子型を求めることのできる元となる複数のコピー数解に関して積分する確率を求めることが可能である。対照的に、先行技術の手法は、典型的には、腫瘍純度及びアレル特異的コピー数（これから候補結合遺伝子型を求めることができる）の単一の推定値に依存するものであり、これは多くの場合に、専門家が定義した最適性基準に基づき手作業で選択される。最適と見なされるコピー数解を選択するステップは、高度に誤りの発生し易いステップであり、単一の解に依存する方法の出力は、その解に応じて大幅に変わる可能性がある。

30

【0029】

このように、有利には、シーケンスデータが観察される確率（シーケンスデータの尤度）は、候補遺伝子型及び対応する腫瘍割合推定値の複数の集合に関して計算されてもよく（例えば、各コピー数解についての項を含む確率の和として、式（3b）、式（4b）を参照のこと）、その寄与には、例えば腫瘍割合推定値及び候補遺伝子型集合を求めた際の元となったコピー数解における信頼度を反映した重みが付けられてもよい。コピー数解の寄与の重みは、好適には、考慮される異なるコピー数解の相対的寄与を反映した全確率となるように、合計が1になるように考察した。単一のコピー数解が使用されるとき、それには1の重みを割り当てることができる（即ち、和は求めないことになり得る）。

40

【0030】

突然変異がクローナルである事前確率は、中立的事前値、又は事前データ及び／又は専門知識から導出される値に設定されてもよい。有利には、本明細書に記載される方法は、突然変異がクローナルである確率の決定に使用するベイズのフレームワークにおいて、突然変異に関する事前知識を考慮に入れることが、かかる知識が利用可能な場合には可能である。しかしながら、本方法は、かかる知識がない場合であってもまた、突然変異がクローナルである尤度を提供することが可能である。

【0031】

突然変異がクローナルである事前確率の値は、対象、腫瘍、突然変異、又はこれらの組

50

み合わせに依存し得る。例えば、値は、例えば、同じ型又は亜型のがんに罹患している患者など、関連性のある患者コホートに関して予め収集されたデータを使用して決定されてもよい。例えば、かかるコホートにおけるクローナル／サブクローナルである突然変異の比率の知識を使用して、本明細書で使用される事前確率を設定し得る。或いは、値は、がん型又は突然変異に関する事前知識に基づいて任意に設定されてもよい。例えば、複数のがん試料に見出されている、且つそれらの試料において多くの場合にクローナルであると同定されている特定の突然変異には、0.5より高い確率を割り当てることができる。

【0032】

本方法は、対象において同定された複数の腫瘍特異的突然変異に関して本方法を繰り返すことをさらに含み得る。本方法は、対象においてクローナルであるというそれらの決定された尤度に少なくとも一部には基づいて複数の腫瘍特異的突然変異をランク付けすること、又はその他何らかの優先順位を付けることをさらに含み得る。

10

【0033】

本方法は、対象の1つ以上の腫瘍特異的突然変異を同定することをさらに含み得る。対象の1つ以上の腫瘍特異的突然変異を同定することは、腫瘍遺伝子材料を含む対象からの1つ以上の試料からのシーケンスデータ及び対象からの1つ以上の生殖細胞系列試料からのシーケンスデータを使用して、前記シーケンスデータを比較することによるなどして実施されてもよい。対象の1つ以上の腫瘍特異的突然変異を同定することは、腫瘍遺伝子材料を含む少なくとも1つの試料からのシーケンスデータを参照配列とアラインメントすること、及び試料の配列が参照配列と異なる位置を同定することを含んでもよい。本方法は、少なくとも1つの生殖細胞系列試料からのシーケンスデータを参照配列とアラインメントすること、及び腫瘍遺伝子材料を含む試料の配列が生殖細胞系列試料と異なる位置を同定することをさらに含み得る。

20

【0034】

対象からの1つ以上の試料からのシーケンスデータを提供するステップは、使用者から（例えばユーザインタフェースを介して）、1つ以上の計算装置から、又は1つ以上のデータストア若しくはデータベースからシーケンスデータを受け取ることを含み得るか、又はそれからなり得る。

【0035】

シーケンスデータを提供するステップは、腫瘍遺伝子材料を含む対象からの1つ以上の試料をシーケンシングすること（又は他の方法で試料に存在するゲノム材料の配列組成を決定すること）をさらに含み得る。

30

【0036】

本方法は、対象からの1つ以上の生殖細胞系列試料をシーケンシングすること（又は他の方法で試料に存在するゲノム材料の配列組成を決定すること）をさらに含み得る。

【0037】

本方法は、対象から、腫瘍遺伝子材料を含む1つ以上の試料及び任意選択で1つ以上の生殖細胞系列試料を入手することをさらに含み得る。

【0038】

本方法は、決定された腫瘍特異的突然変異がクローナルである確率及び／又はそれから導出される若しくはそれに関連する値を、例えばユーザインタフェースを介して使用者に提供することをさらに含み得る。例えば、本方法は、決定された腫瘍特異的突然変異がクローナルである確率に基づいて「クローナル状態」のフラグ又は値を提供することを含み得る。別の例として、本方法は、突然変異を同定する情報（例えば突然変異の配列及びそのゲノム位置など）を提供することを含み得る。

40

【0039】

さらなる態様によれば、対象の1つ以上のクローナルなネオ抗原を同定する方法であって、対象の複数の腫瘍特異的突然変異を同定すること；前出の態様のいずれかの実施形態の方法を用いて、腫瘍特異的突然変異のうちの1つ以上が対象においてクローナルである可能性があるかどうかを決定すること；及び腫瘍特異的突然変異のうちの1つ以上がネオ

50

抗原を生じさせる可能性があるかどうかを決定することを含む方法が提供され、ここでクローナルなネオ抗原は、腫瘍特異的突然変異がクローナルである可能性があるかどうかに関する1つ以上の所定の基準及び腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかに関する1つ以上の基準を満たす腫瘍特異的突然変異である。また、本態様によれば、対象の1つ以上のクローナルなネオ抗原を同定する方法であって、プロセッサにより前記対象からの1つ以上の試料からのシーケンスデータを使用して、対象の複数の腫瘍特異的突然変異を同定すること；プロセッサにより、前出のいずれかの請求項の方法を用いて腫瘍特異的突然変異のうちの1つ以上が対象においてクローナルである可能性があるかどうかを決定すること；及び前記プロセッサにより、腫瘍特異的突然変異のうちの1つ以上をクローナルなネオ抗原候補として選択することを含む方法も記載され、ここでクローナルなネオ抗原候補は、腫瘍特異的突然変異がクローナルである可能性があるかどうかに関する少なくとも1つ又は複数の所定の基準及び任意選択で腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかに関する1つ以上の基準を満たす腫瘍特異的突然変異である。

10

【0040】

本態様の方法は、以下の特徴のうちのいずれか1つ以上を有し得る。

【0041】

クローナルなネオ抗原は、所定の閾値を上回るクローナルである確率を有すること、確率を決定した際の対象であった腫瘍特異的突然変異の中でクローナルである確率が最も高い所定の数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである確率を有すること、及び確率を決定した際の対象であった腫瘍特異的突然変異の中で所定の最上位の百分位数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである確率を有すること、から選択される基準を少なくとも満たす腫瘍特異的突然変異であり得る。このように、腫瘍特異的突然変異がクローナルである可能性があるかどうかに関する1つ以上の所定の基準は、所定の閾値を上回るクローナルである尤度をその突然変異が有すること、尤度を決定した際の対象であった腫瘍特異的突然変異の中でクローナルである尤度が最も高い所定の数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである尤度をその突然変異が有すること、及び尤度を決定した際の対象であった腫瘍特異的突然変異の中で所定の最上位の百分位数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである尤度を有すること、から選択することができる。

20

30

【0042】

クローナルなネオ抗原は、腫瘍細胞に発現する発現産物に関連していること、対象の正常細胞に発現しないタンパク質又はペプチドを生じさせると予測されること、MHC分子によって提示される可能性がある少なくとも1つのペプチドを生じさせると予測されること、対象に存在することが既知のMHCアレルによって提示される可能性がある少なくとも1つのペプチドを生じさせると予測されること、及び免疫原性のタンパク質又はペプチドを生じさせると予測されることから選択される基準を少なくとも満たす腫瘍特異的突然変異であり得る。例えば、クローナルなネオ抗原は、タンパク質の配列の変化を（例えば、それがコード配列であるため、それがスプライス部位に影響を及ぼすため、それがランケット型のペプチドを生じさせるため等の理由で）生じさせて、ひいては対象の正常細胞には発現しないものであり得るタンパク質又はペプチドを生じさせることが予測されるという基準を満たす腫瘍特異的突然変異であり得る。これが該当するか否かは、例えば、対象の予測される正常なプロテオームとの比較によってさらに確認することができる。このように、腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかに関する1つ以上の基準は、腫瘍細胞に発現する発現産物にその突然変異が関連していること、対象の正常細胞に発現しないタンパク質又はペプチドがその突然変異の結果として生じると予測されること、MHC分子によって提示される可能性がある少なくとも1つのペプチドがその突然変異の結果として生じると予測されること、対象に存在することが既知のMHCアレルによって提示される可能性がある少なくとも1つのペプチドがその突然変異の結果

40

50

として生じると予測されること、及び免疫原性のタンパク質又はペプチドがその突然変異の結果として生じると予測されること、から選択することができる。

【0043】

本方法は、1つ以上のクローナルなネオ抗原に関連する1つ以上のペプチド（即ち、腫瘍特異的突然変異が存在する結果として腫瘍細胞に存在すると予測される1つ以上のペプチド配列）を同定することをさらに含んでもよく、ここで腫瘍特異的突然変異は、上記に記載したとおりの（クロナリティーの尤度及びクローナルなネオ抗原を生じさせる尤度に関係する）1つ以上の基準を満たしている。

【0044】

当業者が理解するとおり、本明細書に記載される操作の複雑さ（少なくとも、本明細書に記載されるとおりの数値積分が必要な事後確率を求めることの複雑さ、及びゲノムDNAのシーケンシングによって典型的に発生するデータの量に起因する）は、精神活動の範囲を超えるものである。従って、文脈上特に指示されない限り（例えば試料調製又は収集ステップが記載されている場合）、本明細書に記載される方法のステップは全て、コンピュータ実装される。

【0045】

さらなる態様によれば、がんであると診断された対象に予後を提供する方法であって、対象からの1つ以上の試料に複数の腫瘍特異的突然変異を同定すること、及び第1の態様の任意の実施形態の方法を用いて腫瘍特異的突然変異の各々がクローナルである尤度を決定することを含む方法が提供される。

【0046】

本方法は、クローナルである確率が所定の閾値を上回る腫瘍特異的突然変異の比率に少なくとも一部は依存して、高負荷量のクローナルなネオ抗原を有するか、低負荷量のクローナルなネオ抗原を有するかに対象を分類することをさらに含んでもよく、ここで高負荷量のクローナルなネオ抗原の対象は、低負荷量のクローナルなネオ抗原の対象と比較して向上した予後を有する。

【0047】

さらなる態様によれば、がんであると診断された対象に免疫療法を提供する方法であって、第2の態様の任意の実施形態に係る方法など、本明細書に記載されるとおりの方法を用いて1つ以上のクローナルなネオ抗原を同定すること；及び同定されたクローナルなネオ抗原の1つ以上を標的化する免疫療法を設計することを含む方法が提供される。

【0048】

本方法は、以下の特徴のうちのいずれか1つ以上を有し得る。

【0049】

クローナルなネオ抗原の1つ以上を標的化する免疫療法は、免疫原性組成物、免疫細胞又は治療用抗体を含む組成物であり得る。免疫原性組成物は、同定されたクローナルなネオ抗原の1つ以上のクローナル（例えばネオ抗原ペプチド若しくはタンパク質又はネオ抗原を提示する細胞など）、又は同定されたクローナルなネオ抗原の1つ以上の発現に十分な材料（例えばネオ抗原をコードするDNA又はRNA分子）を含み得る。免疫細胞を含む組成物は、T細胞、B細胞及び／又は樹状細胞を含み得る。治療用抗体を含む組成物は、同定されたクローナルなネオ抗原の1つ以上のうちの少なくとも1つを認識する1つ以上の抗体を含み得る。抗体はモノクローナル抗体であってもよい。

【0050】

任意の態様の任意の実施形態において、がんは、膀胱がん、胃がん、食道がん、乳がん、結腸直腸がん、子宮頸がん、卵巣がん、子宮内膜がん、腎がん（腎細胞）、肺がん（小細胞、非小細胞及び中皮腫）、脳がん（神経膠腫、星状細胞腫、膠芽腫）、黒色腫、リンパ腫、小腸がん（十二指腸がん及び空腸がん）、白血病、膵がん、肝胆道腫瘍、胚細胞がん、前立腺がん、頭頸部がん、甲状腺がん及び肉腫から選択することができる。がんは、肺がんであってもよい。がんは、黒色腫であってもよい。がんは、膀胱がんであってもよい。がんは、頭頸部がんであってもよい。

10

20

30

40

50

【0051】

任意の態様の任意の実施形態において、対象はヒトであってもよい。

【0052】

同定されたクローナルなネオ抗原の1つ以上を標的化する免疫療法を設計することは、標的化される1つ以上のクローナルなネオ抗原の各々に対する1つ以上の候補ペプチドであって、標的化されるクローナルなネオ抗原の少なくとも一部分を各々が含むペプチドを設計することを含み得る。

【0053】

本方法は、1つ以上の候補ペプチドを入手することをさらに含み得る。本方法は、1つ以上の候補ペプチドを1つ以上の特性に関して試験することをさらに含み得る。試験はインビトロ又はインシリコで実施されてもよい。例えば、1つ以上のペプチドは、免疫原性、MHC分子によって（任意選択で、特異的MHC分子アレルであって、対象が発現するMHCアレルに応じて選択されたものであり得るアレルによって）提示される傾向、免疫細胞集団の増殖を誘発する能力等に関して試験されてもよい。

10

【0054】

本方法は、免疫療法を作製することをさらに含み得る。本方法は、候補ペプチドの1つ以上でパルスした樹状細胞の集団を入手することをさらに含み得る。免疫療法は、同定されたクローナルなネオ抗原の1つ以上のうちの少なくとも1つを認識するT細胞を含む組成物であり得る。組成物は、同定されたクローナルなネオ抗原の1つ以上のうちの少なくとも1つを標的化するT細胞に関して濃縮されてもよい。本方法は、T細胞集団を入手すること、及び同定されたクローナルなネオ抗原の1つ以上のうちの少なくとも1つを標的化するT細胞の数又は相対的比率が増加するようにT細胞集団を拡大することを含み得る。

20

【0055】

本方法は、T細胞集団を入手することをさらに含み得る。T細胞集団は、対象から、例えば、対象から入手した1つ以上の腫瘍試料から、又は対象の末梢血試料若しくは他の組織からの試料から単離されてもよい。T細胞集団は、腫瘍浸潤性リンパ球を含み得る。T細胞は、当該技術分野において周知の方法を用いて単離されてもよい。例えば、T細胞は、CD3、CD4又はCD8の発現を基準として試料から作成された単一細胞懸濁液から精製されてもよい。T細胞は、Ficoll-paque勾配に通すことにより試料から濃縮されてもよい。

30

【0056】

本方法は、T細胞集団を拡大することをさらに含み得る。例えば、T細胞は、T細胞に細胞分裂刺激を与えることが公知の条件におけるエキソピボ培養によって拡大されてもよい。例として、T細胞をIL-2などのサイトカインと共に、又は抗CD3及び／又はCD28などのマイトジェン抗体と共に培養してもよい。T細胞は、照射されたものであってもよい抗原提示細胞（APC）と共培養されてもよい。APCは、樹状細胞又はB細胞であってもよい。樹状細胞は、同定されたネオ抗原の1つ以上を単一の刺激物質として又は刺激性ネオ抗原ペプチドのプールとして含有するペプチドでパルスされたものであってもよい。T細胞の拡大は、例えば、追加的な共刺激シグナルを提供する人工抗原提示細胞（aAPC）、及び適切なペプチドを提示する自家PBMCの使用を含め、当該技術分野において公知の方法を用いて実施することができる。自家PBMCは、本明細書で考察するとおりのネオ抗原を単一の刺激物質として、又は代わりに刺激性ネオ抗原のプールとして含有するペプチドでパルスされてもよい。

40

【0057】

さらなる態様によれば、対象のがんの治療における使用のためのT細胞集団を拡大する方法であって、第2の態様の任意の実施形態に係る方法など、本明細書に記載されるとおりの方法を用いて1つ以上のクローナルなネオ抗原を同定すること；同定されたクローナルなネオ抗原のうちの1つを特異的に認識する能力のあるT細胞を含むT細胞集団を取得すること；及び同定されたクローナルなネオ抗原を含む組成物とT細胞集団を共培養する

50

ことを含む方法が提供される。

【0058】

本方法は、以下の特徴のうちの1つ以上を有し得る。

【0059】

入手されるT細胞集団は、同定されたクローナルなネオ抗原のうちの1つを特異的に認識する能力のあるT細胞を含むと仮定し得る。本方法は好ましくは、複数のクローナルなネオ抗原を同定することを含む。T細胞集団は複数のT細胞を含んでもよく、その各々が、複数の同定されたクローナルなネオ抗原のうちの1つを特異的に認識する能力、及び複数の同定されたクローナルなネオ抗原を含む組成物とT細胞集団を共培養する能力を有する。共培養する結果、1つ以上のネオ抗原を特異的に認識するT細胞集団の拡大が生じ得る。拡大は、ネオ抗原及び抗原提示細胞とのT細胞の共培養によって実施されてもよい。抗原提示細胞は樹状細胞であってもよい。従って、この拡大は、ネオ抗原に特異的なT細胞の選択的な拡大であり得る。拡大は、1つ以上の非選択的な拡大ステップをさらに含み得る。

10

【0060】

さらなる態様によれば、前出の態様のいずれかの実施形態に係る方法によって取得されるか、又は取得可能なT細胞集団を含む組成物が提供される。

【0061】

さらなる態様によれば、対象のがんの治療又は予防における使用のための、ネオ抗原、ネオ抗原特異的免疫細胞、又はネオ抗原を認識する抗体を含む組成物が提供され、ここで前記ネオ抗原は、本明細書に記載される方法を用いてクローナルなネオ抗原と同定されたものである。

20

【0062】

さらなる態様によれば、ネオ抗原、ネオ抗原特異的免疫細胞、又はネオ抗原を認識する抗体を含む組成物が提供され、ここで前記ネオ抗原は、本明細書に記載される方法を用いてクローナルなネオ抗原と同定されたものである。

【0063】

さらなる態様によれば、その表面にネオ抗原を発現する細胞又は細胞集団が提供され、ここで前記ネオ抗原は、本明細書に記載される方法を用いてクローナルなネオ抗原と同定されたものである。

30

【0064】

さらなる態様によれば、対象のがんの治療又は予防における使用のための、ネオ抗原、ネオ抗原を認識する免疫細胞、又はネオ抗原を認識する抗体が提供され、ここで前記ネオ抗原は、本明細書に記載される方法を用いてクローナルなネオ抗原と同定されたものである。

【0065】

さらなる態様によれば、対象のがんの治療又は予防における使用のための医薬の製造における、ネオ抗原、ネオ抗原を認識する免疫細胞、又はネオ抗原を認識する抗体の使用が提供され、ここで前記ネオ抗原は、本明細書に記載される方法を用いてクローナルなネオ抗原と同定されたものである。

40

【0066】

さらなる態様によれば、がんであると診断された対象を治療する方法であって、本明細書に記載される方法を用いて提供された免疫療法、又は本明細書に記載されるとおりの組成物を投与することを含む方法が提供される。

【0067】

さらなる態様によれば、プロセッサと；プロセッサによる実行時に、上記の第1、第2、第3又は第4の態様の任意の実施形態に係る方法など、本明細書に記載される任意の方法のステップをプロセッサに実施させる命令を含むコンピュータ可読媒体とを含むシステムが提供される。

【0068】

50

さらなる態様によれば、1つ以上のプロセッサによる実行時に、上記の第1、第2、第3又は第4の態様の任意の実施形態に係る方法など、本明細書に記載される任意の方法のステップを1つ以上のプロセッサに実施させる命令を含む、又はより非一時的なコンピュータ可読媒体が提供される。

【0069】

さらなる態様によれば、コンピュータ上で実行されると、上記の第1、第2、第3又は第4の態様の任意の実施形態に係る方法など、本明細書に記載される任意の方法のステップをコンピュータに実施させるコードを含むコンピュータプログラムが提供される。

【図面の簡単な説明】

【0070】

図面の簡単な説明

【図1】腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法、及びクローナルなネオ抗原の同定におけるその使用を概略的に説明するフローチャートである。

【図2】免疫療法を提供する方法を概略的に説明するフローチャートである。

【図3】腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定するための、及び／又はクローナルなネオ抗原を同定するための、及び／又は免疫療法を提供するためのシステムの実施形態を示す。

【図4】本明細書に開示される方法によって仮定される構造化された集団を概略的に説明する。

【図5】合成データを使用した、本明細書に記載されるクローナルな突然変異を同定する方法の結果を示す。各プロットは、異なる最大コピー数及び試料数についての偽陽性率 ($FPR = FP / (FP + TN)$)、式中、 FP = 偽陽性数、 TN = 真陰性数；1 - 特異度) の関数としてのROC曲線 (受信者動作特性、真陽性率 ($TPR = TP / (TP + FN)$) 式中、 TP = 真陽性数、 FN = 真陰性数；感度) を示し、3つの別個の曲線は、各々が異なるシミュレーションのリード深度を示す。A. シミュレーション及び同定に二項エミッションモデルを使用。B. シミュレーションにベータ二項エミッションモデルを使用、同定に二項モデルを使用。C. シミュレーション及び同定にベータ二項エミッションモデルを使用。

【図6】二倍体細胞株 (A) 及び異数体細胞株 (B) との細胞株混合実験からのデータを使用した、本明細書に記載されるクローナルな突然変異を同定する方法の結果を示す。各プロットの表題は、分析にどの試料を含めたかを指示している。各プロットは、本明細書に記載されるとおりの方法 (「ACE」と表示する) を使用したROC曲線、並びに同じデータに対してPyClone (Roth et al., 2014) によるランを行ったときに入手されたTPR及びFPRを示す単一の点を示す。

【図7】TRACERxプロジェクトからのデータを使用した、本明細書に記載されるクローナルな突然変異を同定する方法の結果を示す。A. 分析したデータの概要。試料毎に、TRACERxパイプラインを用いて同定されたクローナル及びサブクローナルなSNVの数 (手作業でキュレートしたPyClone結果) を示す。B. 本明細書に記載されるとおりの方法から予測される遍在性の確率の分布。(左) 全てのASCAT (Van Loo et al., 2010) 倍数性／純度の解を確率によって重み付けして考えたときの結果。(右) ASCATからの最も可能性の高い解を選択したときの結果。C. 手作業でキュレートしたPyClone (Roth et al., 2014) コールをグラウンドトゥールズとして使用した、本明細書に記載されるとおりの方法の予測性能を説明する (確率によって重み付けしたASCATからの可能な全ての解を考慮に入れた場合又は入れない場合の) ROC曲線。

【図8】本明細書に記載されるとおりの方法と比較対照方法との間の実行時間比較の結果を示す。比較対照方法 (PyClone-VI) は、計算効率に関して嚴重に最適化されているが、並列処理はできない。本明細書に記載される方法 (ACE) は、大々的な計算効率最適化には供されていないが、仕様上一層効率が高く、並列処理が可能である (2、3又は4個のCPUによる実装を示す)。

10

20

30

40

50

【発明を実施するための形態】

【0071】

詳細な説明

がん療法及び予後判定におけるその重要性を考えると、本方法ががん患者のクローナルな突然変異及びネオ抗原を同定することには、明らかな臨床適用がある。腫瘍試料のクローナルな構造を再構築しようと試みる方法が複数提案されている (Schwartz and Schaeffer, 2017)。しかしながら、それらの手法はいずれも様々な欠点を抱えているため、その臨床的有用性は限られている。このように、本発明者らは、腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定するための改良された方法がなお必要とされていることを突き止めた。がんにおけるクローナルな突然変異 (がんの全ての細胞に存在する突然変異) を同定するというこの問題は、些細と言うには程遠い。実際、一次試料には細胞の混合物 (正常細胞及びがん細胞を含む) が含まれ、がんゲノムは多くの場合、広範囲にわたる不均一なコピー数変異を持つ。このように、突然変異のアレル保有率は、正常細胞の比率、突然変異を有する腫瘍細胞の比率、及びそれらの腫瘍細胞の各々における突然変異のアレルコピー数を含め、諸要因が組み合わさった結果である。これは、任意の実験プロトコルに付随する技術的ノイズ (例えばシーケンシングエラー)、それらのプロトコルの限界 (例えばショートリードアラインメントの不確かさ)、及び一次試料ががんの状態の部分的なスナップショットを表すという固有の限界によって一層ひどくなる。Roth et al. (2014) は、ディープシーケンシングした体細胞突然変異の集合を、同じ進化史を共有するクラスターに分類するベイジアンクラスタリング法 (「PyClone」と呼ばれる) を提案することにより、この課題に取り組むことを試みた。これは、その細胞保有率を推定し、セグメント毎のコピー数変化及び正常細胞の混入によって導入されるアレル不均衡を考慮することにより行われる。

【0072】

この手法については有利には、統計的フレームワークの明確さが際立つが、本発明者らは、複数の理由から、それが非研究環境での使用にはあまり適さないことを突き止めた。第一に、この手法は比較的遅く、個別の突然変異というよりむしろ、突然変異の集合 (ディープシーケンシングした突然変異の大規模な集合、又は最近では、完全ゲノムなど) のレベルで動作する。これらの欠点は、両方とも、この方法が突然変異の集合全体に関してクラスター構造を推測するという事実の帰結である。そのため、この手法は、がんからの複数の試料に同定された全ての突然変異の集合内にある各突然変異の細胞保有率を含めたモデル母数についての事後密度、並びにこの突然変異集合に関するクラスター構造の計算を必要とする複雑な問題を解く。本発明者らは、少なくとも2つの理由でこれが望ましくないことを突き止めた：この手法は、突然変異集合全体に関するクラスター構造が目的でない場合には不必要に遅く、及びこの手法では、同時に分析される他の突然変異と独立に任意の突然変異を臨床的パイプラインを通じて追跡することは不可能になる。実際、任意の突然変異をクローナルである可能性があると同定することは、一緒に分析された他の突然変異と固有に結び付いている。このように、分析しようとする突然変異の集合に変化があれば、パイプラインの再実行が必要となり、潜在的に異なる結果が生み出される。対照的に、本発明者らが提案する手法は各突然変異を独立に分析し、突然変異の集団に依存するいかなる構造も推測しない。これにより、本発明者らが開発したアルゴリズムの計算の複雑さは従来技術 (PyClone) よりも低いため、ひいては必要となる計算資源 (プロセッサ、メモリ及び／又はネットワーク資源) が全体的に及び各突然変異について少なくなり、コンピュータ技術の明確な改良がもたらされる。さらに、突然変異間に依存関係がないため、突然変異の集合の処理を完全に並列処理化することが可能である。この結果、ある実行に単一の突然変異に関する問題があっても、他の突然変異の実行に影響が及ばないため、複雑さのさらなる低下が生じ、処理速度が増し、及び取り扱い易さが向上することになる。

【0073】

加えて、先行技術の手法は、出力をクロナリティーの指標に変換するために、クラスタ

一構造を分析し、クローナルであると仮定される突然変異のクラスターを選び、並びに選んだクラスター内にある突然変異に優先順位を付けることによる専門家の手作業による介入が必要である。最後に、この手法はまた、誤ったコピー数解がクラスタリング結果に大きい影響を与えることに伴い、適切なコピー数推定値を選ぶために専門家が手作業で介入する必要もあり、要求されるレベルの信頼性をもってコピー数を自動で設定する手法は欠けている。こうした複数レベルの手作業での介入には、それが機械によったならばもっと効率的に行われたであろう作業に（人間の）専門家の存在を必要とするという理由に限らず、問題がある。実際、クローナルであると仮定される突然変異のクラスターを選び、及びそのクラスター内にある突然変異に優先順位を付けるという、コピー数解の選択に専門家が適用する処理は、単純に、信頼性をもって自動化することができない。この処理には専門知識が必要であり、専門家がその判断を用いて選択を行うことに頼るものであるが、そうした選択には、最終的には少なくともある程度主観が入る。これは自動化することができず、また、その過程に主観性が関わるため、信頼性をもって再現することができない。対照的に、本発明の方法によれば、これらの手作業での介入はいずれも必要ない。複数のコピー数解及びそれらの信頼度をシームレスに統合することができ（従って手作業又は自動のいずれであれ、何かを選択する必要がなくなる）、クロナリティーの尤度が突然変異毎に個別に予測される（従って突然変異のクラスターを選び出し又はそのクラスター内で優先順位を付ける必要がなく、予測された尤度に基づいて、自然の、予測可能な、且つ再現性のある優先順位を定義することができる）。このように、クローナルな突然変異が、これまで可能であったものよりも高い信頼性をもって同定されるため、純粋な研究を越えて治療薬開発の分野にまで及ぶその使用が可能となる。がん治療、並びにがんのモニタリング及び管理の分野でかかるクローナルな突然変異を使用できる可能性は想定されていたが、臨床的パイプラインに組み込むためそれらの突然変異を信頼性をもって同定する手段がなかったことにより、実際にはその現実での使用が阻まれていたため、個別の突然変異についてクロナリティーの尤度の信頼性のある推定値を利用可能であることは、そうした分野にとっての明確な改良に相当する。

10

20

【0074】

本開示では、以下の用語を用いることになり、それらは以下に指示するとおり定義されることが意図される。

【0075】

30

「試料」は、本明細書で使用されるとき、ゲノムシーケンシング（例えば全ゲノムシーケンシング、全エクソームシーケンシング）など、ゲノム解析用のゲノム材料をそこから入手することのできる細胞又は組織試料、生体液、抽出物（例えば対象から入手したDNA抽出物）であり得る。試料は、対象から入手される細胞、組織又は生体液試料（例えば生検）であってもよい。かかる試料は、「対象試料」と称することができる。詳細には、試料は、血液試料、又は腫瘍試料、又はそれらに由来する試料であってもよい。試料は、対象から入手したばかりの新鮮なものであってもよく、又はゲノム解析前に処理され、及び／又は保存されていたもの（例えば、凍結され、固定され、又は1つ以上の精製、濃縮又は抽出ステップに供されたもの）であってもよい。試料は、細胞又は組織培養試料であってもよい。このように、本明細書に記載されるとおりの試料は、対象から入手された生体試料からであれ、又は例えば細胞株から入手された試料からであれ、細胞又はそれに由来するゲノム材料を含む任意の種類の試料を指し得る。実施形態において、試料は、ヒト対象など、対象から入手される試料である。試料は、好ましくは、哺乳類からのもの（例えば、哺乳類細胞試料又はネコ、イヌ、ウマ、ロバ、ヒツジ、ブタ、ヤギ、ウシ、マウス、ラット、ウサギ若しくはモルモットなど、哺乳類対象からの試料など）、好ましくはヒトからのもの（例えば、ヒト細胞試料又はヒト対象からの試料など）である。さらに、試料は輸送及び／又は保存されてもよく、採取は、ゲノムシーケンスデータ収集（例えばシーケンシング）場所から離れた場所で行われてもよく、及び／又は本明細書に記載される任意のコンピュータ実装方法ステップは、試料採取場所から離れた、及び／又はゲノムデータ収集（例えばシーケンシング）場所から離れた場所で行われてもよい（例えばコンピ

40

50

ュータ実装方法ステップは、「クラウド」プロバイダーを用いるなど、ネットワーク化されたコンピュータを用いて実施されてもよい）。

【0076】

「混合試料」は、複数の細胞型又は複数の細胞型に由来する遺伝子材料を含むと仮定される試料を指す。本開示の文脈の中では、混合試料とは、典型的には、腫瘍細胞、又は腫瘍細胞に由来する遺伝子材料を含むもの、又はそれを含むと仮定される（予想される）ものである。対象から入手される試料、例えば、腫瘍試料などは、典型的には（それが1つ以上の精製及び／又は分離ステップに供されない限りは）混合試料である。典型的には、試料は、腫瘍細胞及び少なくとも1つの他の細胞型（及び／又はそれに由来する遺伝子材料）を含む。例えば、混合試料は腫瘍試料であってもよい。「腫瘍試料」は、腫瘍に由来する、又は腫瘍から入手される試料を指す。かかる試料は、腫瘍細胞と正常な（非腫瘍）細胞とを含んでもよい。正常細胞は、免疫細胞（例えばリンパ球など）、及び／又は他の正常な（非腫瘍）細胞を含んでもよい。かかる混合試料中のリンパ球は、「腫瘍浸潤性リンパ球」（TIL）と称することができる。腫瘍は、固形腫瘍又は非固形若しくは血液学的腫瘍であってもよい。腫瘍試料は、対象からの原発腫瘍試料、腫瘍関連リンパ節試料、又は転移部位からの試料であってもよい。腫瘍細胞又は腫瘍細胞に由来する遺伝子材料を含む試料は、体液試料であってもよい。従って、腫瘍細胞に由来する遺伝子材料は、循環腫瘍DNA又はエキソソームにある腫瘍DNAであってもよい。代わりに、又はそれに加えて、試料は、循環腫瘍細胞を含んでもよい。混合試料は、遺伝子材料を抽出するための処理を受けた細胞、組織又は体液の試料であってもよい。生体試料から遺伝子材料を抽出する方法は、当該技術分野において公知である。混合試料は、試料中にある複数の細胞型又は複数の細胞型に由来する遺伝子材料の比率を変更し得る1つ以上の処理ステップに供されたものであってもよい。例えば、腫瘍細胞を含む混合試料は、腫瘍細胞に関して試料を濃縮する処理を受けたものであってもよい。従って、精製された腫瘍細胞の試料は、特定の目的上、その試料が純粋である（即ち1又は100%の腫瘍割合を有する）と仮定し得るとしても、少量の他の種類の細胞が存在し得ることに基づいて、「混合試料」と称することができる。

10

20

【0077】

用語「腫瘍割合」（また、「腫瘍純度」又は単純に「純度」、又は異常細胞割合（ACF）と称されることもある）は、混合試料中にあるDNA含有細胞のうち腫瘍細胞であるものの比率、又は試料中の腫瘍細胞及び非腫瘍細胞からの遺伝子材料の特定の混合物を生じさせると仮定される等価な比率を指す。試料中の腫瘍割合を決定する方法は、当該技術分野において公知である。例えば、細胞又は組織試料の文脈では、腫瘍割合は、病理スライドを分析することによるか（例えばヘマトキシリン及びエオシン（H&E）染色したスライド又は他の組織化学若しくは免疫組織化学スライド、試料の1つ以上の代表的な範囲にある腫瘍細胞をカウントすることによる）、又はフローサイトメトリーなどのハイスループットアッセイを用いることにより推定されてもよい。遺伝子材料を含む試料の文脈では、腫瘍割合は、例えば、ASCAT (Van Loo et al., 2010)、ABSOLUTE (Carter et al., 2012)、又はichorCNA (Adalsteinsson et al., 2017) など、腫瘍及び生殖細胞系列ゲノムのデコンボリューションを試みる配列解析プロセスを用いて推定されてもよい。

30

40

【0078】

「正常試料」又は「生殖細胞系列試料」は、腫瘍細胞又は腫瘍細胞に由来する遺伝子材料を含まないと仮定される試料を指す。生殖細胞系列試料は、血液試料、組織試料、又は精製された試料、例えば、対象からの末梢血単核球の試料などであってもよい。同様に、用語「正常」、「生殖細胞系列」又は「野生型」は、配列又は遺伝子型に関するとき、腫瘍細胞以外の細胞の配列／遺伝子型を指す。生殖細胞系列試料は、僅かな比率の腫瘍細胞又はそれに由来する遺伝子材料を含み得るが、それにも関わらず、実際上は、前記細胞又は遺伝子材料を含まないと仮定することができる。換言すれば、細胞又は遺伝子材料は全て、正常であると仮定することができ、及び／又はこの仮定に適合しないシーケンスデー

50

タは無視してもよい。

【0079】

用語「シーケンスデータ」は、試料中にある特定の配列を有するゲノム材料の存在及び好ましくはまた量を指示するものである情報を指す。かかる情報は、シーケンシング技術、例えば、次世代シーケンシング（NGS）、例えば、全エクソームシーケンシング（WES）、全ゲノムシーケンシング（WGS）、又は捕捉されたゲノム遺伝子座のシーケンシング（標的化した又はパネルシーケンシング）などを用いるか、又はアレイ技術、例えば、コピー数変異アレイ、又は他の分子カウントアッセイを用いて入手することができる。NGS技術が用いられるとき、シーケンスデータは、特定の配列を有するシーケンシングリードの数のカウントを含み得る。アレイ技術など、非デジタル技術が用いられるとき、シーケンスデータは、試料中にある特定の配列を有する配列の数を例えば適切な対照との比較によって指示するシグナル（例えば強度値）を含み得る。シーケンスデータは、当該技術分野において公知の方法を用いて（例えば、Bowtie（Langmead et al., 2009）など）、参照配列、例えば参照ゲノムにマッピングされてもよい。このように、シーケンシングリードのカウント又は等価な非デジタルシグナルは、特定のゲノム位置に結び付けられることになり得る（ここで「ゲノム位置」とは、シーケンスデータをマッピングした元の参照ゲノムにおける位置を指す）。さらに、ゲノム位置には突然変異が含まれてもよく、その場合、シーケンシングリードのカウント又は等価な非デジタルシグナルは、特定のゲノム位置における可能な変異体（「アレル」とも称される）の各々に結び付けられることになり得る。試料中の特定の位置における突然変異の存在を同定するプロセスは、「バリエーションコーリング」と称され、当該技術分野において公知の方法を用いて実施することができる（例えば、GATK HaplotypeCaller、<https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller>など）。例えば、シーケンスデータは、特定のゲノム位置において生殖細胞系列（また、「参照」と称されることもある）アレルとマッチするリード数のカウント（又は等価な非デジタルシグナル）、及びゲノム位置において突然変異した（また、「代替の」と称されることもある）アレルとマッチするリード数のカウント（又は等価な非デジタルシグナル）を含み得る。

【0080】

さらに、シーケンスデータを使用することにより、当該技術分野において公知の方法を用いて、ゲノムに沿ったコピー数プロファイルを推測することができる。コピー数プロファイルは、アレル特異的であってもよい。本開示の文脈では、コピー数プロファイルは、好ましくはアレル特異的及び腫瘍／正常試料特異的である。換言すれば、本開示において使用されるコピー数プロファイルは、好ましくは、腫瘍細胞と正常細胞との混合物を含む試料を分析して、試料中の腫瘍細胞及び正常細胞についてのアレル特異的コピー数プロファイルを作製するように設計された方法を用いて入手される。混合試料のアレル特異的コピー数プロファイルは、例えばASCAT（Van Loo et al., 2010）を用いて、シーケンスデータから（例えば、上記に記載したとおりのリードカウントを使用して）入手することができる。他の方法が公知であり、等しく好適である。好ましくは、本開示の文脈の中では、アレル特異的コピー数プロファイルを入手するために用いられる方法は、複数の可能なコピー数解及び関連する品質／信頼性メトリクスを報告するものである。例えば、ASCATは、倍数性（セグメント特異的でなく、全腫瘍試料の倍数性）の値の組み合わせ毎の当てはまりの良さのメトリクス及び対応するアレル特異的コピー数プロファイルを評価した際の純度を出力する。かかる方法によって生成された腫瘍特異的コピー数プロファイルは、腫瘍細胞集団全体の平均又は要約を表すことに留意されたい（即ちこれは、本明細書に記載される新規開発の対象である腫瘍集団内における不均一性を考慮しない）。

【0081】

用語「全コピー数」は、試料中にあるゲノム領域のコピーの全数を指す。用語「メジャーコピー数」は、試料中にある最も保有率の高いアレルのコピーの数を指す。逆に、用語「マイナーコピー数」は、試料中にある最も保有率の高いアレル以外のアレルのコピーの数を指す。特に指示がない限り、これらの用語は、推測される腫瘍コピー数プロファイル

についての推測されるメジャー及びメジャーコピー数（及び全コピー数）を指す。用語「正常コピー数」又は「正常全コピー数」は、試料中にある正常細胞のゲノム領域のコピーの数を指す。正常細胞は、典型的には各染色体を2コピー有し（細胞が遺伝的に男性であり、且つ染色体が性染色体である場合を除く）、従って正常コピー数とは、実施形態において、2に等しいと仮定し得る（ゲノム領域がX又はY染色体上にあり、且つ分析下の試料が男性対象のものである場合を除く、その場合、正常コピー数は1に等しいと仮定し得る）。或いは、特定のゲノム領域の正常コピー数は、正常試料を使用して決定されてもよい。

【0082】

用語「対数R値」（「log R」、「log RR」、「LLR」と称されることもある）は、あるゲノム遺伝子座における全コピー数を定量化するものである正規化後の全シグナル強度の尺度を指す。本開示の文脈では、この用語は、典型的には腫瘍遺伝子材料を含む試料についての対数R値を指し、正規化は、典型的には正常試料を基準に実施される（これは好ましくは、対応する正常試料であるが、またプロセス対応正常試料又は他の好適な正常参照試料であってもよい）。例えば、NGSが用いられるとき、log Rは、リード深度の正規化後の対数変換（log（リード深度腫瘍／リード深度正常））として求めることができる。用語「平均Bアレル頻度」（MBAF、また「Bアレル頻度」（BAF）と称されることもある）は、あるゲノム位置における正規化後のアレル強度比の尺度である。本開示の文脈では、この用語は、典型的には腫瘍遺伝子材料を含む試料についてのBAF値を指し、正規化は、典型的には正常試料を基準に実施される（これは好ましくは、対応する正常試料であるが、またプロセス対応正常試料又は他の好適な正常参照試料であってもよい）。例えば、BAFは、腫瘍アレル対正常アレルについてのアレル頻度の比として求めることができる。コピー数プロファイルは、典型的には、「セグメント」と呼ばれるゲノム領域に関するコピー数推定値を含む。従って、あるゲノム位置に関連するBAF及びlog Rとは、ある特定のゲノム位置（例えば突然変異のゲノム位置など）に重なるセグメントのBAF及びlog Rを指し得る。さらに、BAF及びlog Rを使用して、対応するメジャー及びマイナーコピー数を求めることができる。実施形態において、コピー数メトリクスの値は、たとえ腫瘍コピー数プロファイル値のみを使用する場合であっても、腫瘍コピー数プロファイル推定値及び正常コピー数プロファイル推定値の両方について提供され得る。

【0083】

用語「腫瘍特異的突然変異」、「体細胞突然変異」又は単純に「突然変異」は、同義的に使用され、同じ対象からの健常細胞と比較した腫瘍細胞におけるヌクレオチド配列（例えばDNA又はRNA）の違いを指す。ヌクレオチド配列のこの違いにより、同じ対象からの健常細胞には発現しないタンパク質の発現が生じ得る。例えば、突然変異は、一塩基変異（SNV）、多塩基変異（MNV）、欠失突然変異、挿入突然変異、転座、ミスセンス突然変異、転座、融合、スプライス部位突然変異、又は腫瘍細胞の遺伝子材料の任意の他の変化であり得る。突然変異により、同じ対象からの健常細胞に存在しないタンパク質又はペプチドの発現が生じ得る。突然変異は、エクソームシーケンシング、RNAシーケンシング、全ゲノムシーケンシング及び／又は標的遺伝子パネルシーケンシング及び／又はルーチンの単一遺伝子サンガーシーケンシングと、続く配列アラインメント及び腫瘍試料からのDNA及び／又はRNA配列を参照試料又は参照配列からのDNA及び／又はRNA（例えば生殖細胞系列DNA及び／又はRNA配列、又はデータベースからの参照配列）と比較することによって同定されてもよい。好適な方法は、当該技術分野において公知である。

【0084】

「インデル突然変異」は、生物のヌクレオチド配列（例えばDNA又はRNA）における塩基の挿入及び／又は欠失を指す。典型的には、インデル突然変異は、生物のDNA、好ましくはゲノムDNAに現れる。実施形態において、インデルは、1～100塩基、例えば1～90、1～50、1～23又は1～10塩基であり得る。インデル突然変異は、

フレームシフトインデル突然変異であってもよい。フレームシフトインデル突然変異は、1つ以上のヌクレオチドの挿入又は欠失によって引き起こされるヌクレオチド配列のリーディングフレームの変化である。かかるフレームシフトインデル突然変異によって新規オープンリーディングフレームが発生してもよく、これは典型的には、対象の対応する健常細胞における非突然変異DNA/RNAによってコードされるポリペプチドとは極めて明白な差異がある。

【0085】

「ネオ抗原」（又は「ネオ抗原」）とは、がん細胞内での突然変異の結果として生じる抗原である。このように、ネオ抗原は正常（即ち非腫瘍）細胞からは発現しない（又は大幅に低いレベルで発現する）。ネオ抗原がプロセッシングされると、明白な差異のあるペプチドが発生してもよく、これは、MHC分子のコンテキストで提示されたとき、T細胞によって認識されることができる。本明細書に記載されるとおり、ネオ抗原は、がん免疫療法の基礎として使用し得る。本明細書で「ネオ抗原」と言うとき、それには、ネオ抗原に由来するペプチドもまた含まれることが意図される。用語「ネオ抗原」は、本明細書で使用されるとき、ネオ抗原のうち免疫原性であるいずれか一部を包含することが意図される。「抗原性」分子は、本明細書において参照されるとき、適切な方法で免疫系又は免疫細胞に提示されたとき、それ自体、又はその一部が免疫応答を刺激する能力を有する分子である。ネオ抗原による特定のMHC分子（特定のHLAアレルによってコードされる）への結合は、当該技術分野において公知の方法を用いて予測することができる。MHC結合を予測する方法の例としては、Lundegaard et al.、O'Donnel et al.、及びBullik-Sullivan et al.によって記載されるものが挙げられる。例えば、ネオ抗原のMHC結合は、netMHC-3（Lundegaard et al.）及びnetMHCpan4（Jurtz et al.）アルゴリズムを使用して予測し得る。特定のMHC分子に結合すると予測されたネオ抗原は、その結果、細胞表面上の前記MHC分子によって提示されることが予測される。

【0086】

「クローナルなネオ抗原」（また、「トランカルな（truncal）ネオ抗原」と称されることもある）は、対象からの1つ以上の試料中にある腫瘍細胞の本質的に全てに存在する（又は1つ又は複数の試料中の腫瘍遺伝子材料の由来である元の腫瘍細胞の本質的に全てに存在すると仮定することのできる）突然変異によって生じるネオ抗原である。同様に、「クローナルな突然変異」（「トランカルな（truncal）突然変異」と称されることもある）は、対象からの1つ以上の試料中にある腫瘍細胞の本質的に全てに存在する（又は1つ又は複数の試料中の腫瘍遺伝子材料の由来である元の腫瘍細胞の本質的に全てに存在すると仮定することのできる）突然変異である。このように、クローナルな突然変異とは、対象からの1つ以上の試料中にある全ての腫瘍細胞に存在する突然変異であり得る。「サブクローナルな」ネオ抗原は、対象からの1つ以上の腫瘍試料中にある細胞の一部の集合又は一定比率に存在する（又は1つ又は複数の試料中の腫瘍遺伝子材料の由来である腫瘍細胞の一部の集合に存在すると仮定することのできる）突然変異によって生じるネオ抗原である。同様に、「サブクローナルな」突然変異は、対象からの1つ以上の腫瘍試料中にある細胞の一部の集合又は一定比率に存在する（又は1つ又は複数の試料中の腫瘍遺伝子材料の由来である腫瘍細胞の一部の集合に存在すると仮定することのできる）突然変異である。ネオ抗原又は突然変異は、対象からの1つ以上の試料のコンテキストではクローナルであり得るが、対象に存在し得る（例えば原発腫瘍及び転移の全ての領域を含む）腫瘍細胞の集団全体のコンテキストでは、真にクローナルではないこともあり得る。このように、クローナルな突然変異とは、それが対象の腫瘍細胞の本質的に全てに（即ちあらゆる腫瘍細胞に）存在する突然変異であるという意味で「真にクローナル」であり得る。これは、1つ以上の試料が、対象に存在する細胞のありとあらゆる部分集合を代表するわけではないこともあり得るためである。このように、本開示の文脈の中では、「クローナルなネオ抗原」又は「クローナルな突然変異」はまた、分析した本質的にあらゆる腫瘍細胞にネオ抗原が存在するが、対象に存在し得るあらゆる腫瘍細胞に存在するわけではないこともあり得ることを指示して、「遍在性ネオ抗原」又は「遍在性突然変異」とも称すること

ができる。用語「クローナル」及び「遍在性」は、文脈上「真のクロナリティー」を指す意図があったと指示されない限り、同義的に使用される。表現「本質的に全ての腫瘍細胞」は、1つ以上の試料又は対象との関係では、1つ以上の試料又は対象における腫瘍細胞の少なくとも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%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、又は少なくとも99%を指し得る。

【0087】

それにもかかわらず、本明細書に記載されるとおりのクローナルである（又は「遍在性」である）可能性があると同定されるネオ抗原／突然変異は、真にクローナルである可能性はあるか、又は少なくとも、クローナルである可能性がないと同定されるネオ抗原／突然変異と比べて真にクローナルである可能性がより高い。さらに、クローナルなネオ抗原／突然変異の同定に使用した1つ又は複数の試料が腫瘍の一層複雑な遺伝的多様性像を捉えているとき（例えば、対象からの複数の試料、例えば腫瘍の異なる領域からの試料などを含めることによる、及び／又は例えばc tDNA試料など、腫瘍細胞の多様性を固有に捉える試料を含めることによる）、本明細書において同定されるクローナルなネオ抗原／突然変異が真にクローナルである尤度の信頼度は増加する。逆に、本明細書に記載されるとおりのクローナルである可能性がないと同定されるネオ抗原／突然変異については、ネオ抗原／突然変異がクローナルである可能性がないという同定が、試料抽出過程でもたらされる限られた見方であっても、そのネオ抗原／突然変異が全ての腫瘍細胞に存在しないというエビデンスがあることを指示しているため、真にクローナルである可能性は低い。このように、クローナルなネオ抗原／突然変異を同定する過程は、1つ以上の試料から利用可能な対象の腫瘍のクローナルな構造の限られた見方に基づいて、どの候補ネオ抗原／突然変異がクローナルである可能性が最も高いかについての優先順位を付けるものとすることができる。

【0088】

用語「がん細胞割合」（又は「CCF」）は、突然変異、例えば、特定のネオ抗原を生じさせる突然変異などを含有する腫瘍細胞の比率を指す。本開示の文脈の中では、がん細胞割合は、1つ以上の試料に基づいて推定されてもよく、そのため対象における真のがん細胞割合（上記に説明したとおり）と等しくないこともあり得る。それにもかかわらず、1つ以上の試料に基づいて推定されるがん細胞割合は、真である可能性のあるがん細胞割合の有用な指標を提供し得る。さらに、上記に説明したとおり、かかる推定値の正確さは、がん細胞割合の推定に使用される1つ又は複数の試料が腫瘍の遺伝的多様性像を一層完全に捉えているとき、増加し得る。ゲノムデータにおける追加的な雑音源及び交絡因子とは、1つ以上の試料から決定されるがん細胞割合が推定値に相当することを意味する。そのため、真にクローナルな突然変異／ネオ抗原は $CCF = 1$ を有するはずであるが、実際には、クローナルである可能性がより高い突然変異／ネオ抗原は、より低いCCF推定値に関連すると予想される突然変異であって、クローナルである可能性が低い突然変異よりも、高いCCF推定値（それは1に等しくないものであり得る）に関連するものと予想される。

【0089】

例えば、がん細胞割合推定値は、Landau et al. (2013)により記載されるとおり、変異体アレル頻度をコピー数及び純度推定値と統合することによって求めることができる。かかるCCF推定値はまた、クローナルである可能性がある突然変異を同定するためにも使用することができる。例えば、クローナルな突然変異は、推定されるがん細胞割合（ CCF ） ≥ 0.75 、例えば、 $CCF \geq 0.80$ 、 0.85 、 0.90 、 0.95 又は1.0などを有する突然変異として定義することができる。サブクローナルな突然変異は、 $CCF < 0.95$ 、 0.90 、 0.85 、 0.80 、又は 0.75 を有する突然変異として定義することができる。さらに、CCF推定値は、確率を0～1の間のCCFの複数の可能な値の各々と関連付ける分布に関連してもよく（例えばそれに由来してもよく）、そ

こから信頼度の統計的推定値を求めることができる。例えば、突然変異は、95% CCF 信頼区間が ≥ 0.75 である、即ち推定される CCF の95%信頼区間の上限が 0.75 以上である場合に、クローナルな突然変異である可能性があるとして定義することができる。換言すれば、突然変異は、下限 L 及び上限 H の CCF の区間が、 $P(L < CCF < H) = 95\%$ 、 $H \geq 0.75$ となるような場合に、クローナルな突然変異である可能性があるとして定義することができる。或いは、突然変異は、そのがん細胞割合 (CCF) が上記に定義するのとりの要求される値、例えば 0.75 又は 0.95 に達するか、又はそれを超える機会又は確率が、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%又はそれ以上の機会又は確率など、50%より高い場合に、クローナルであると同定することができる。換言すれば、突然変異は、 $P(CCF > 0.75) \geq 0.5$ の場合にクローナルであると同定することができる。例えば、突然変異は、その CCF が 0.95 (又は 0.75 、又は任意の他の選ばれた閾値) を超える事後確率が、それぞれ、 0.5 より大きい又は小さいかどうかに基づき、クローナル又はサブクローナルである可能性があるとして分類されてもよい。

10

20

30

40

50

【0090】

本開示の方法によれば、以下にさらに説明するとおり、突然変異がクローナルである尤度が求められる。これは、 $P(CCF = 1)$ と等価である。これに関連して、以下でさらに説明するとおり、突然変異は、 $P(CCF = 1)$ が閾値を超える場合に、クローナルである可能性があるとして同定されてもよい。閾値は固定であってもよい。例えば、突然変異は、 $P(CCF = 1) > 0.05$ の場合に、クローナルである可能性があるとして同定されてもよい。或いは、閾値は、調査下の突然変異の特定の集合について決定されてもよい。実施形態において、閾値は、公知のクローナル／非クローナル状態を含むベンチマークデータセットに基づいて、所定の精度及び／又は再現率に達するように設定されてもよい。ベンチマークデータセットは、合成データを使用して、及び／又はクロナリティー構造が既知である集団から入手されるデータセット (例えば細胞株混合物データ) を使用して求めることができる。例えば、突然変異は、 $P(CCF = 1) > t$ である場合に、クローナルである可能性があるとして同定されてもよく、式中、 t は、ベンチマークデータセット中の真にクローナルな突然変異の95% (又は任意の他の値、例えば、50%、55%、60%、65%、70%、75%、80%、85%、90%など) が同定される (即ち、高くても5%の偽陰性率) ような最大値である。別の例として、突然変異は、 $P(CCF = 1) > t$ である場合に、クローナルである可能性があるとして同定されてもよく、式中、 t は、ベンチマークデータセット中の閾値を超える突然変異の少なくとも50% (又は任意の他の値、例えば、50%、55%、60%、65%、70%、75%、80%、85%、90%など) が真にクローナルな突然変異となる (即ち少なくとも50%の真陽性率) ような最小値である。或いは、閾値は、上記に記載される基準を満たす信頼区間を有する推定される CCF (例えばそれは、推定される CCF の95%信頼区間の上限が 0.75 以上となるようなものである) に関連する任意の突然変異 (又は突然変異のある特定の%) が、クローナルである可能性があるとして選択されるように設定されてもよい。或いは、閾値は、上記に記載される基準を満たす事後確率分布を有する (例えばその CCF が 0.95 (又は 0.75 、又は任意の他の選ばれた閾値) を超える事後確率が 0.5 より高い) 推定される CCF に関連する任意の突然変異 (又は突然変異のある特定の%) が、クローナルである可能性があるとして選択されるように設定されてもよい。

【0091】

がん免疫療法 (又は単純に「免疫療法」) は、免疫原性組成物 (例えばワクチン)、免疫細胞を含む組成物、又は免疫活性薬物、例えば治療用抗体などの対象への投与を含む治療手法を指す。用語「免疫療法」はまた、治療用組成物それ自体も指し得る。本開示の文脈では、免疫療法は、典型的にはネオ抗原を標的化する。例えば、免疫原性組成物又はワクチンは、ネオ抗原、ネオ抗原提示細胞又はネオ抗原の発現に必要な材料を含み得る。別の例として、免疫細胞を含む組成物は、ネオ抗原を認識する T 及び／又は B 細胞を含み得る。免疫細胞は、腫瘍又は他の組織 (限定はされないが、リンパ節、血液又は腹水を含む

）から単離され、エキソピボ又はインピトロで拡大され、及び対象に再投与されてもよい（「養子細胞療法」と称されるプロセス）。代わりに、又はそれに加えて、T細胞は、対象から単離し、及びネオ抗原を標的化するように操作し（例えば、ネオ抗原に結合するキメラ抗原受容体の挿入により）、及び対象に再投与することができる。別の例として、治療用抗体は、ネオ抗原を認識する抗体であってもよい。当業者は、ネオ抗原が細胞表面抗原である場合に、本明細書において言及されるとおりの抗体は、ネオ抗原を認識することになると理解するであろう。ネオ抗原が細胞内抗原である場合、抗体は、ネオ抗原ペプチド-MHC複合体を認識することになる。本明細書において参照されるとき、ネオ抗原を「認識」する抗体には、これらの可能性の両方が包含される。さらに、免疫療法は複数のネオ抗原を標的化し得る。例えば、免疫原性組成物は、複数のネオ抗原、複数のネオ抗原を提示する細胞又は複数のネオ抗原の発現に必要な材料を含んでもよい。別の例として、組成物は、複数のネオ抗原を認識する免疫細胞を含んでもよい。同様に、組成物は、同じネオ抗原を認識する複数の免疫細胞を含んでもよい。別の例として、組成物は、複数のネオ抗原を認識する複数の治療用抗体を含んでもよい。同様に、組成物は、同じネオ抗原を認識する複数の治療用抗体を含んでもよい。

10

【0092】

本明細書に記載されるとおりの組成物は、薬学的に許容可能な担体、希釈剤又は賦形剤を加えて含む医薬組成物であってもよい。医薬組成物は、任意選択で、1つ以上のさらなる薬学的に活性なポリペプチド及び／又は化合物を含んでもよい。かかる製剤は、例えば、静脈内注入に好適な形態であり得る。

20

【0093】

「免疫細胞」と言うとき、それには、免疫系の細胞、例えば、T細胞、NK細胞、NK T細胞、B細胞及び樹状細胞が包含されることが意図される。好ましい実施形態において、免疫細胞はT細胞である。ネオ抗原を認識する免疫細胞は、操作されたT細胞であってもよい。ネオ抗原特異的T細胞は、ネオ抗原又はネオ抗原ペプチドに特異的に結合するキメラ抗原受容体（CAR）又はT細胞受容体（TCR）、又はネオ抗原又はネオ抗原ペプチドに特異的に結合する親和性増強T細胞受容体（TCR）（本明細書において以下でさらに考察されるとおり）を発現し得る。例えば、T細胞は、ネオ抗原又はネオ抗原ペプチドに特異的に結合するキメラ抗原受容体（CAR）又はT細胞受容体（TCR）（例えば、ネオ抗原又はネオ抗原ペプチドに特異的に結合する親和性増強T細胞受容体（TCR））を発現し得る。或いは、ネオ抗原を認識する免疫細胞の集団は、腫瘍を有する対象から単離されたT細胞の集団であってもよい。例えば、T細胞集団は、例えば、対象の腫瘍試料、末梢血試料又は他の組織からの試料など、対象から単離された試料中にあるT細胞から作成されてもよい。T細胞集団は、ネオ抗原が同定される腫瘍からの試料から作成されてもよい。換言すれば、T細胞集団は、治療しようとする患者の腫瘍に由来する試料から単離されてもよく、ここでネオ抗原もまた、前記腫瘍からの試料から同定されている。T細胞集団は、腫瘍浸潤性リンパ球（TIL）を含んでもよい。

30

【0094】

用語「抗体」（Ab）には、モノクローナル抗体、ポリクローナル抗体、多重特異性抗体（例えば、二重特異性抗体）、及び所望の生物学的活性を呈する抗体断片が含まれる。用語「免疫グロブリン」（Ig）は、「抗体」と同義的に使用されてもよい。好適なネオ抗原が、例えば本開示に係る方法によって単離された後、当該技術分野において公知の方法を用いて抗体を作成することができる。

40

【0095】

「免疫原性組成物」は、対象において免疫応答を誘導する能力を有する組成物である。この用語は、用語「ワクチン」と同義的に使用される。本明細書に記載される免疫原性組成物又はワクチンは、対象における免疫応答の発生につながり得る。発生させることのできる「免疫応答」は、体液性及び／又は細胞媒介性免疫、例えば抗体産生の刺激、又はワクチン中の抗原に対応する抗原をその表面上に発現する細胞を認識して破壊し得る（又は他の方法で排除し得る）細胞傷害性又はキラー細胞の刺激であってもよい。免疫原性組成

50

物は、1つ以上のネオ抗原、又は1つ以上のネオ抗原の発現に必要な材料を含んでもよい。加えて、ネオ抗原は、抗原提示細胞、例えば樹状細胞など、細胞の形態で送達されてもよい。樹状細胞などの抗原提示細胞は、ネオ抗原又はネオ抗原ペプチドでパルスされるか、若しくはそれが負荷されてもよく、又は1つ、2つ以上のネオ抗原又はネオ抗原ペプチド、例えば、2、3、4、5、6、7、8、9又は10個のネオ抗原又はネオ抗原ペプチドを発現するように（DNA又はRNA移入によって）遺伝子修飾されてもよい。樹状細胞免疫原性組成物又はワクチンの調製方法は、当該技術分野において公知である。

【0096】

ネオ抗原ペプチドは、当該技術分野において公知の方法を用いて合成されてもよい。用語「ペプチド」は、通常の意味で、典型的にはα-アミノ基と隣接アミノ酸のカルボキシル基との間のペプチド結合によって互いにつながった一連の残基、典型的にはL-アミノ酸を意味して使用される。この用語には、修飾されたペプチド及び合成ペプチド類似体が含まれる。ネオ抗原ペプチドは、ペプチド内の任意の残基位置にがん細胞特異的突然変異（例えば、一塩基変異（SNV）によってコードされる非サイレントなアミノ酸置換）を含んでもよい。例として、MHCクラスI分子への結合能を有するペプチドは、典型的には7～13アミノ酸長である。そのため、アミノ酸置換は、13アミノ酸を含むペプチドでは位置1、2、3、4、5、6、7、8、9、10、11、12又は13に存在し得る。実施形態において、より長いペプチド、例えば21～31merが使用されてもよく、任意の位置、例えばペプチドの中央、例えば位置10、11、12、13、14、15又は16に突然変異があってもよい。かかるペプチドはまた、CD4及びCD8の両方の細胞を刺激してネオ抗原を認識するために使用することもできる。

【0097】

本明細書で使用されるとき「治療」は、治療下の疾患の1つ以上の症状を治療前の症状と比べて低減し、軽減し、又は消失させることを指す。「予防（prevention）」（又は予防（prophylaxis））は、疾患の症状の発生を遅延させる又は予防することを指す。予防は、絶対的なもの（疾患が起こらないようにする）であってもよく、又は一部の個体においてのみ、若しくは限られた時間にわたってのみ有効なものであってもよい。

【0098】

本明細書で使用されるとき、用語「コンピュータシステム」には、上述の実施形態に係るシステムを具体化し又は方法を実行するためのハードウェア、ソフトウェア及びデータストレージ装置が含まれる。例えば、コンピュータシステムは、中央演算処理装置（CPU）、入力手段、出力手段及びデータストレージを含んでもよく、これらは1つ以上の接続された計算装置として具体化されてもよい。好ましくはコンピュータシステムはディスプレイを有するか、又はディスプレイを有する計算装置を含むことにより、視覚的出力表示を提供する（例えばビジネスプロセスの設計において）。データストレージは、RAM、ディスクドライブ又は他の非一時的コンピュータ可読媒体を含んでもよい。コンピュータシステムには、ネットワークによって接続された、ネットワーク上で互いに通信することが可能な複数の計算装置が含まれてもよい。コンピュータシステムがクラウドコンピュータからなり得るか、又はそれを含み得ることは、明示的に想定される。

【0099】

本明細書で使用されるとき、用語「コンピュータ可読媒体」には、限定なしに、コンピュータ又はコンピュータシステムが直接読み取り、アクセスすることのできる1つ又は複数の任意の非一時的媒体が含まれる。媒体としては、限定はされないが、フロッピーディスク、ハードディスク記憶媒体及び磁気テープなど、磁気記憶媒体；光ディスク又はCD-ROMなど、光記憶媒体；RAM、ROM及びフラッシュメモリを含めたメモリなど、電氣的記憶媒体；及び磁気／光記憶媒体など、上記のハイブリッド及び組み合わせを挙げることができる。

【0100】

クローナルな突然変異の同定

本開示は、腫瘍細胞又はそれに由来する遺伝子材料を含む1つ以上の試料からのシーケ

ンスデータを使用して腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法を提供する。本開示はまた、クローナルなネオ抗原を同定する方法であって、1つ以上の腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定することを含む方法も提供する。図1を参照することにより、例示的方法を記載することにする。任意選択のステップ10において、腫瘍からのゲノム材料を含む試料を対象から入手し得る。試料は、典型的には、腫瘍細胞を含めた複数の細胞型からの混合ゲノム材料を含む試料である。好ましくは、腫瘍細胞からのゲノム材料を含まない、又はそこから正常細胞からのゲノム材料を抽出し得る対応する試料が入手されてもよく、又は予め入手されていてもよい。対応する試料とは、腫瘍試料と同じ対象から入手された試料である。対応する正常試料を使用することにより、腫瘍試料に同定される任意の変異体位置に対応する正常試料の変異体位置と比較して生殖細胞系列変異体を除外することができるため、体細胞性（腫瘍特異的）突然変異のコーリングの正確さが向上する。同じ対応する正常試料を使用して、対象からの複数の腫瘍試料を分析してもよい。さらに、対応する試料及び1つ以上の腫瘍試料は、異なる時点で入手されたものであってもよい。例えば、第1の腫瘍試料及び対応する試料は、診断時又は腫瘍の切除時に入手されたものであってもよく、さらなる腫瘍試料を入手して、後の時点で最初の対応する試料と併せて分析することができる。対応する試料が利用可能でないときは、参照試料又は共通の体細胞変異体を含むゲノムが使用されてもよい。或いは、処理済みの対応する正常試料を使用してもよく、これは同じ対象から入手されたものでなくてもよく、又は対象のプールから入手されたものであってもよい。

10

20

【0101】

任意選択のステップ12において、例えば、全エクソームシーケンシングの1つ、又は全ゲノムシーケンシングを用いて試料中のゲノム材料をシーケンシングすることにより、1つ以上の混合試料及び任意選択で対応する試料の配列内容を決定し得る。例えばアレル特異的コピー数アレイなどの代替的方法が使用されてもよく、しかしシーケンシング方法は試料中の各特定の配列の数を代表するデジタル出力を生成するため、シーケンシング方法が好ましい。任意選択のステップ14において、シーケンスデータを分析して、腫瘍細胞には存在するが、非がん性細胞には存在しない可能性がある1つ以上の突然変異を同定し得る。これらは腫瘍特異的突然変異に相当し、候補ネオ抗原として使用することができる。これは、1つ以上の試料（即ち、利用可能な場合、1つ又は複数の混合試料及び1つ又は複数の生殖細胞系列試料）からの配列をアラインメントするステップ、及び腫瘍の配列が生殖細胞系列配列と異なる、又は生殖細胞系列配列と異なると仮定することのできる（例えば対象について生殖細胞系列配列が利用可能でない場合）ゲノム位置を同定するステップを含んでもよい。

30

【0102】

ステップ16において、突然変異アレル（「非参照アレル」とも称される）を支持するリードのカウント、そのゲノム位置にある1つ又は複数の生殖細胞系列アレルを支持するリードのカウント（A、まとめて「生殖細胞系列アレル」と称され、生殖細胞系列集団中で遺伝子座がヘテロ接合である場合には、「参照」、「野生型」又は「正常」アレルとも称される）、及び／又は候補腫瘍特異的突然変異のゲノム位置にある全リードカウントを含む、候補腫瘍特異的突然変異のゲノム位置における混合試料のシーケンスデータを入手する。これらのメトリクスのうち2つのみを入手すればよく、このとき3つ目は、これらのうちのいずれか2つから推論することができる。シーケンスデータには、代わりに、又はそれに加えて、リードデータ又は強度データが含まれてもよく、そこからカウントを入手することができる。任意選択のステップ18において、腫瘍遺伝子材料を含む各試料に適合する少なくとも1つのコピー数解に関する情報を入手し得る。この情報は、メジャーコピー数、マイナーコピー数、全コピー数、平均Bアレル頻度、対数R値及び腫瘍倍数性、及び正常コピー数から選択される試料の腫瘍割合についてのアレル特異的コピー数メトリクス、又はこれらのアレル特異的コピー数メトリクスに適合する候補結合遺伝子型の集合など、これらのメトリクスに由来する情報を含み得る。一部は重複する情報を含み、及

40

50

び／又は好適なデフォルト値に関連し得るため、かかるアレルト異的コピー数メトリクスの全てが必要なわけではない。例えば、正常コピー数は、上記に説明したとおり好適なデフォルト値に関連し得る。さらに、メジャーコピー数、全コピー数及びマイナーコピー数のうちの2つのみがあれば、3つ目を推測するのに十分である。同様に、これらの3つの値はMBAF値及びlogR値から推測することができる（及び逆もまた同様である）。任意選択で、コピー数解は、対応する信頼度メトリクスに関連し得る。かかるメトリクスが利用可能でないとき、各コピー数解は、等しく可能性があるものと仮定してもよい。各候補結合遺伝子型は、正常集団、腫瘍特異的突然変異を含まない参照腫瘍集団及び腫瘍特異的突然変異を含む変異体腫瘍細胞集団についての腫瘍特異的突然変異の位置にある遺伝子型を含む。

10

【0103】

ステップ20において、腫瘍特異的突然変異がクローナルである確率を、突然変異がクローナルである事前確率と、1つ以上の試料の各々についての腫瘍割合及び1つ以上の候補結合遺伝子型を考慮したときの、腫瘍特異的突然変異が（i）クローナルである、及び（ii）非クローナルである場合にシーケンスデータが観察される確率と、に依存する、事後確率として決定する。事前確率は、何らかの証拠が考慮に入れられる前の量についての信念に相当する確率である。本文脈では、突然変異がクローナルである事前確率とは、腫瘍において突然変異がクローナルである確率であって、事前知識又は仮定に基づいた、混合試料からのシーケンスデータは考慮に入れない確率に相当し得る。ステップ22において、腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかを決定する。例えば、その突然変異が、生殖系列細胞（これらのゲノムは突然変異を持たない）からは発現しないペプチド又はタンパク質を生じさせる可能性があるかどうかを決定し得る。このステップは、ステップ14の後の任意の時点で実施することができ、詳細には、ステップ16～20の後に実施する必要はない。例えば、候補腫瘍特異的突然変異は、腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する前に、それがネオ抗原を生じさせる可能性があるかどうかに応じてフィルタリングし得る。ステップ24において、ステップ20の結果に適用される1つ以上の基準及びステップ22の結果に適用される1つ以上の基準を満たす腫瘍特異的突然変異を同定し得る。これらを、クローナルなネオ抗原候補に相当すると見なすことができる。任意選択のステップ26において、前出のステップ（及び詳細にはステップ20～24）のいずれかの結果を使用者へと、例えばユーザインタフェースを介して提供し得る。これらの結果は、以下にさらに説明するとおり、例えば対象に免疫療法又は予後を提供するために用いられ得る。

20

30

【0104】

適用

上記の方法は、がん診断、予後判定及び治療手法の文脈で適用が見出される。詳細には、上記の方法は、クローナルなネオ抗原を標的化する免疫療法を提供するために用いられ得る。このように、本明細書にはまた、対象に免疫療法を提供する方法であって、対象からの1つ以上の試料から1つ以上のクローナルなネオ抗原を同定することを含む方法も記載される。

【0105】

図2は、例示的免疫療法を提供する方法を概略的に説明する。任意選択のステップ210において、腫瘍遺伝子材料を含む1つ以上の試料及び1つ以上の生殖細胞系列試料を対象から入手する。対象は、がんであると診断された対象であってもよく、免疫療法が提供されるのと同じ対象であってもよい（しかしそうである必要はない）。ステップ212において、本明細書に記載される方法を、例えば図1を参照することにより用いて、クローナルなネオ抗原候補のリストを入手する。リストは、単一のネオ抗原、又は複数のネオ抗原を含み得る。好ましくは、リストは複数のネオ抗原を含む。ステップ214において、少なくとも1つの（及び任意選択で複数の）候補ネオ抗原を標的化する免疫療法を設計する。かかる免疫療法の設計は、クローナルなネオ抗原候補の各々について1つ以上の候補ペプチドを同定すること（ステップ214A）を含み得る。例えば、クローナルなネオ抗

40

50

原候補の少なくとも1つについて、対応する生殖細胞系列ペプチドと比較してネオ抗原を特徴付けるその長さ及び／又は配列変異の位置が異なる複数のペプチドを設計し得る。ステップ214Bにおいて、同定された1つ以上のペプチドをインビトロ及び／又はインシリコで試験し、その免疫原性、MHC分子によって提示される可能性等などの1つ以上の特性を評価し得る。任意選択のステップ214Cにおいて、例えばステップ214Bの結果に基づき、ペプチドの1つ以上を選択し得る。

【0106】

ステップ216において、選択されたペプチドを入手し得る。選択の配列を有するペプチドは、例えば発現系を使用するか、又は直接的な合成によるなど、当該技術分野において公知の任意の方法を用いて入手することができる。ステップ218において、1つ以上の候補ペプチドを使用して免疫療法を作製し得る。免疫療法は、1つ以上の候補ペプチド又はその発現に十分な材料を含んでもよく（例えば免疫原性組成物又はワクチンの場合）、又は候補ペプチドを使用して入手した分子又は細胞を含んでもよい（例えば候補ペプチドに選択的に結合する治療用抗体、又は候補ペプチドを特異的に認識する免疫細胞の場合）。任意選択のステップ220において、免疫療法を対象に投与することができ、この対象は、好ましくはクローナルなネオ抗原の同定に使用した試料の入手元となった対象である。1つ以上のクローナルなネオ抗原を認識するT細胞が濃縮されたT細胞集団を含む免疫療法を作製する例を記載する。ステップ218Aにおいて、T細胞集団を入手し得る。T細胞は、治療しようとする対象から入手してもよいが、そうすることが必要というわけではない。T細胞は、腫瘍試料から、血液試料から、又は任意の他の組織試料から入手されてもよい。ステップ218Bにおいて、樹状細胞集団を入手し得る。例えば、樹状細胞集団は、治療しようとする対象からの単核球（例えば末梢血単核球、PBMC）に由来してもよい。ステップ218Cにおいて、樹状細胞集団を候補ペプチドでパルスし得る。ステップ218Dにおいて、パルスした樹状細胞の集団を使用してT細胞集団を選択的に拡大し得る。例えばサイトカイン又は刺激性抗体などの追加の拡大因子が使用されてもよい。

【0107】

このように、本開示はまた、1つ以上のクローナルなネオ抗原を認識するT細胞が選択的に濃縮されたT細胞集団を含むT細胞組成物も提供し、ここで1つ以上のクローナルなネオ抗原は、本明細書に記載される方法のいずれかを用いて同定されたものである。

【0108】

本明細書に記載されるとおりのT細胞組成物において、ネオ抗原応答性T細胞の拡大された集団は、ネオ抗原ペプチドによる再刺激へのT細胞集団の応答によって測定したとき、拡大されていないT細胞集団よりも高い活性を有し得る。活性は、サイトカイン産生によって測定されてもよく、ここでより高い活性とは、5～10倍又はそれ以上の活性の増加である。

【0109】

複数のクローナルなネオ抗原と言うとき、それは、ネオ抗原を生じさせる異なる腫瘍特異的突然変異を各々が含む複数のペプチド又はタンパク質を指し得る。前記複数の、2～250、3～200、4～150、又は5～100個の腫瘍特異的突然変異、例えば、5～75又は10～50個の腫瘍特異的突然変異であり得る。各腫瘍特異的突然変異は、1つ以上のクローナルなネオ抗原ペプチドによって表されてもよい。換言すれば、複数のクローナルなネオ抗原は複数の異なるペプチドを含んでもよく、その一部が、同じ腫瘍特異的突然変異を（例えば、ペプチドの配列内の異なる位置に、又は様々な長さのペプチド内に）含む配列を含む。

【0110】

本開示に従い作製されるT細胞集団は、クローナルであると予測される1つ以上のネオ抗原を標的化するT細胞の数又は比率の増加を呈することになる。すなわち、T細胞集団の組成物は、クローナルであると予測されるネオ抗原を標的化するT細胞の割合又は比率が増加していることになる点で、「天然」のT細胞集団（即ち本明細書で考察される拡大

ステップを受けていない集団)と異なることになる。本開示に係るT細胞集団は、クローナルであると予測されるネオ抗原を標的化するT細胞を少なくとも約0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95又は100%有し得る。

【0111】

本明細書に記載される免疫療法は、がんの治療に使用し得る。従って、本開示はまた、対象のがんを治療する方法であって、本明細書に記載されるとおりの免疫治療用組成物を対象に投与することを含む方法も提供する。

【0112】

加えて、クローナルなネオ抗原の存在は、がんの予後の改善に関連することが示されている。従って、本明細書にはまた、がんであると診断された対象に予後を提供する方法であって、対象からの1つ以上の腫瘍試料中にあるクローナルなネオ抗原の割合及び／又は数を決定することを含む方法も記載される。

【0113】

好適には、本明細書に記載される任意の態様の任意の実施形態において、がんは、卵巣がん、乳がん、子宮内膜がん、腎がん(腎細胞)、肺がん(小細胞、非小細胞及び中皮腫)、脳がん(神経膠腫、星状細胞腫、膠芽腫)、黒色腫、メルケル細胞がん、淡明細胞腎細胞がん(ccRCC)、リンパ腫、小腸がん(十二指腸がん及び空腸がん)、白血病、膵がん、肝胆道腫瘍、胚細胞がん、前立腺がん、頭頸部がん、甲状腺がん及び肉腫であってもよい。例えば、がんは、肺腺がん又は肺扁平上皮がんなどの肺がんであってもよい。別の例として、がんは、黒色腫であってもよい。実施形態において、がんは、黒色腫、メルケル細胞がん、腎がん、非小細胞肺がん(NSCLC)、膀胱尿路上皮がん(BLAC)及び頭頸部扁平上皮がん(HNSC)及びマイクロサテライト不安定性(MSI)高値がんから選択することができる。一部の実施形態において、がんは、非小細胞肺がん(NSCLC)である。他の実施形態において、がんは黒色腫である。

【0114】

本開示の組成物及び方法を使用した治療はまた、腫瘍に由来する循環腫瘍細胞及び／又は転移を標的化することも包含し得る。1つ以上のネオ抗原を標的化する本開示に係る治療は、化学療法、放射線療法、又は非特異的免疫療法などの標準的な手法で起こり得る抗治療性腫瘍細胞の進化を防ぐ助けとなり得る。本明細書に記載されるがんを標的化するための方法及び使用は、追加的ながん療法と組み合わせて実施されてもよい。詳細には、本明細書に記載されるT細胞組成物は、免疫チェックポイント介入、共刺激抗体、化学療法及び／又は放射線療法、標的化した療法又はモノクローナル抗体療法と組み合わせて投与されてもよい。「組み合わせて」とは、追加的な療法を本明細書に記載されるとおりのT細胞組成物の投与の前、それと同時に又はその後に投与することを指し得る。

【0115】

本開示はまた、免疫治療用組成物の作製方法であって、ネオ抗原をクローナルである可能性があると同定すること、及びネオ抗原を標的化する免疫治療用組成物を作製することを含む方法も提供する。

【0116】

また、本明細書には、がんであると診断された対象を治療する方法であって、対象の複数の腫瘍特異的突然変異を同定すること；腫瘍特異的突然変異のうちの1つ以上が対象においてクローナルである可能性があるかどうかを決定すること；腫瘍特異的突然変異のうちの1つ以上を候補のクローナルなネオ抗原として選択することであって、クローナルなネオ抗原候補が、腫瘍特異的突然変異がクローナルである可能性があるかどうかに関する少なくとも1つ又は複数の所定の基準を満たす腫瘍特異的突然変異であること；及び選択されたクローナルなネオ抗原候補の1つ以上を標的化する免疫療法で対象を治療することにより、1つ以上のクローナルなネオ抗原を同定することを含む方法も記載され；ここで

10

20

30

40

50

対象の腫瘍特異的突然変異がクローナルである可能性があるかどうかの決定は、本明細書に記載される方法を用いて実施される。詳細には、対象の腫瘍特異的突然変異がクローナルである可能性があるかどうかの決定は、プロセッサにより、腫瘍遺伝子材料を含む対象からの1つ以上の試料からのシーケンスデータを入力することであって、シーケンスデータが、1つ以上の試料の各々について、試料中にある腫瘍特異的突然変異を示すリードの数（ d_b ）、試料中にある対応する生殖細胞系列アレルを示すリードの数、及び腫瘍特異的突然変異の位置にある全リード数（ d ）のうちの少なくとも2つを含むこと、及びプロセッサにより、腫瘍特異的突然変異がクローナルである尤度を、突然変異がクローナルである事前確率と、1つ以上の試料の各々についての腫瘍割合、並びに正常集団、腫瘍特異的突然変異を含まない参照腫瘍集団及び腫瘍特異的突然変異を含む変異体腫瘍細胞集団について、ある遺伝子型を腫瘍特異的突然変異の位置に各々含む1つ以上の候補結合遺伝子型を考慮したときの、腫瘍特異的突然変異が（i）クローナルである、及び（ii）非クローナルである場合にシーケンスデータが観察される確率とに依存する事後確率として決定することを含み得る。

10

【0117】

クローナルなネオ抗原候補は、腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかに関する少なくとも1つ又はそれ以上の所定の基準をさらに満たす腫瘍特異的突然変異として選択されてもよい。クローナルなネオ抗原候補として腫瘍特異的突然変異のうちの1つ以上を前記プロセッサにより選択するステップは、腫瘍細胞に発現する発現産物にその突然変異が関連していること、対象の正常細胞に発現しないタンパク質又はペプチドがその突然変異の結果として生じると予測されること、MHC分子によって提示される可能性がある少なくとも1つのペプチドがその突然変異の結果として生じると予測されること、対象に存在することが既知のMHCアレルによって提示される可能性がある少なくとも1つのペプチドがその突然変異の結果として生じると予測されること、及び免疫原性のタンパク質又はペプチドがその突然変異の結果として生じると予測されること、から選択される、腫瘍特異的突然変異がネオ抗原を生じさせる可能性があるかどうかに関する1つ以上の基準を、1つ以上の腫瘍特異的突然変異が満たすかどうかを決定することを含み得る。クローナルなネオ抗原候補として腫瘍特異的突然変異のうちの1つ以上を前記プロセッサにより選択するステップは、所定の閾値を上回るクローナルである尤度をその突然変異が有すること、尤度を決定した際の対象であった腫瘍特異的突然変異の中でクローナルである尤度が最も高い所定の数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである尤度をその突然変異が有すること、及び尤度を決定した際の対象であった腫瘍特異的突然変異の中で所定の最上位の百分位数の腫瘍特異的突然変異が選択されるように適応的に設定された閾値を上回るクローナルである尤度を有すること、から選択される、腫瘍特異的突然変異がクローナルである可能性があるかどうかに関する1つ以上の所定の基準を、1つ以上の腫瘍特異的突然変異が満たすかどうかを前記プロセッサにより決定することを含み得る。

20

30

【0118】

選択されたクローナルなネオ抗原の1つ以上を標的化する免疫療法は、免疫原性組成物、免疫細胞を含む組成物又は治療用抗体であってもよい。免疫療法は、同定された選択のクローナルなネオ抗原の1つ以上のうちの少なくとも1つを認識するT細胞を含む組成物であってもよい。組成物は、同定された選択のクローナルなネオ抗原の1つ以上のうちの少なくとも1つを標的化するT細胞に関して濃縮されてもよい。本方法は、T細胞集団を入力すること、及び同定された選択のクローナルなネオ抗原の1つ以上のうちの少なくとも1つを標的化するT細胞の数又は相対的比率が増加するようにT細胞集団を拡大することを含んでもよい。

40

【0119】

システム

図3は、本開示に係る、腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定するための、及び／又はクローナルなネオ抗原を同定するための、及び／又は予後

50

を提供するための、又は少なくとも一部には同定されたクローナルなネオ抗原をベースとする免疫療法を提供するためのシステムの実施形態を示す。本システムは計算装置1を含み、これは、プロセッサ101とコンピュータ可読メモリ102とを含む。図示される実施形態において、計算装置1はまたユーザインタフェース103も含み、これは、画面として例示されるが、例えば聴覚的又は視覚的信号によるなど、使用者に情報を伝える任意の他の手段も含み得る。計算装置1は、例えばネットワーク6経由などにより、シーケンシング機などのシーケンスデータ収集手段3、及び／又はシーケンスデータを格納する1つ以上のデータベース2に通信可能に接続されている。1つ以上のデータベースは、例えば、参照配列、母数など、計算装置1が使用し得る他の種類の情報を追加的に保存し得る。計算装置は、スマートフォン、タブレット、パーソナルコンピュータ又は他の計算装置であってもよい。計算装置は、本明細書に記載されるとおりの、腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法を実装するように構成される。代替的实施形態において、計算装置1は遠隔計算装置（図示せず）と通信するように構成され、遠隔計算装置は、それ自体が、本明細書に記載されるとおりの、腫瘍特異的突然変異がクローナルである可能性があるかどうかを決定する方法を実装するように構成される。そのような場合、遠隔計算装置はまた、計算装置に方法の結果を送信するように構成されてもよい。計算装置1と遠隔計算装置との間の通信は、有線又は無線接続を通じたものであってもよく、ローカル又はパブリックネットワーク経由、例えばパブリックインターネット経由又はWi-Fi（登録商標）経由などで行われてもよい。

10

20

30

40

50

【0120】

シーケンスデータ収集3手段は、計算装置1と有線接続されていてもよく、又は図示されるとおり、例えばネットワーク6を通じてなど、無線接続を通じて通信することが可能であってもよい。計算装置1とシーケンスデータ収集手段3との間の接続は、直接又は間接的（例えば遠隔コンピュータを通じてなど）であってもよい。シーケンスデータ収集手段3は、細胞及び／又は組織試料から抽出された核酸試料、例えばゲノムDNA試料からシーケンスデータを取得するように構成される。一部の实施形態において、試料は、DNA精製、断片化、ライブラリ調製、標的配列キャプチャ（例えばエクソンキャプチャ及び／又はパネル配列キャプチャなど）など、1つ以上の前処理ステップに供されたものであってもよい。好ましくは、試料は増幅に供されたものであるか、又はそれが増幅に供されたものであるとき、それは、例えば一意の分子識別子など、増幅バイアス制御手段の存在下で行われた。本開示の文脈の中では、ゲノムコピー数プロファイルの決定における使用に好適な任意の試料調製プロセス（全ゲノム又は配列特異的のいずれであれ）を使用し得る。シーケンスデータ収集手段3は、好ましくは次世代シーケンサーである。シーケンスデータ収集手段3は、シーケンスデータ（未加工のもの、又は部分的に処理したもの）を保存し得る1つ以上のデータベース2と直接又は間接的に接続されていてもよい。

【0121】

以下は例として提示され、特許請求の範囲を限定するものと解釈されてはならない。

【実施例】

【0122】

実施例

これらの例は、本開示に係るクローナルな突然変異を同定する方法について説明し、シミュレートしたデータ及び複数の種類の実験データを使用してその使用を実証する。

【0123】

方法

突然変異遺伝子型モデル

このモデルのデータは、 S 個の試料（ $s = 1, \dots, S$ ）からの N 個の突然変異（ $n = 1, \dots, N$ ）からのアレルカウントである。簡単にするため、及びこの方法は単一の試料及び突然変異を分析することができるため、本節で使用する表記法において、突然変異についての添え字 n 及び試料についての s は明示的に含めないこととする。図4に図示されるとおり、このモデルは、シーケンシングした細胞の集合が各突然変異によって3つの

サブ集団：（i）健常な生殖細胞系列ゲノムを有する細胞からなる正常細胞集団（突然変異の領域では二倍体である可能性がある）；（ii）問題の突然変異のないがん細胞からなる参照細胞集団（問題の突然変異の領域では異数体であり得る）；及び（iii）問題の突然変異を有するがん細胞からなる変異体細胞集団（問題の突然変異の領域では異数体であり得る、前記領域に参照集団と同じコピー数を有しないものであり得る）に分かれることを仮定する。用語「突然変異」は、本明細書では、その広義の意味において、シーケンスデータ、特にゲノムシーケンスデータに検出可能な任意の遺伝子変化を指すことが意図される。これには、詳細には、一塩基変異（SNV）、多塩基変異（MNV）、インデル等が含まれる。

【0124】

$G = (A, B, AA, AB, AAA, AAB, \dots)$ を、全ての遺伝子型の集合とする（ここでA及びBは、それぞれ参照アレル及び変異体アレルを表す）。例えば、ABであれば、全コピー数が2のヘテロ接合変異体（1個の参照／正常アレルAと1個の変異体アレルBとを含む）を表すことになる。この表記法に基づけば、図4において、正常集団は遺伝子型AAを有し（ここで両方のAは同じであっても、又は異なってもよく、即ち正常集団はホモ接合又はヘテロ接合であってもよく、但し両方のアレルとも正常である）、参照集団は遺伝子型AAAを有し（ここでAアレルは正常集団のAアレルから選択される）、及び変異体集団は遺伝子型AABを有する（ここでAアレルは正常集団のAアレルから選択され、Bアレルは任意の非参照アレルである）。本発明者らは、各サブ集団内の全ての細胞の遺伝子型が一定である（即ち、図4を参照することにより、正常集団中の全ての細胞が遺伝子型AAを有し、参照集団中の全ての細胞が遺伝子型AAAを有し、及び変異体集団中の全ての細胞が遺伝子型AABを有する）と仮定する。 $G = (G_H; G_R; G_V) \in G^3$ を、それぞれ正常（健常）、参照及び変異体集団の遺伝子型を成分とするベクトルとする（これらの個別の遺伝子型の各々は、以下では総称的に「G」と称することとする）。 t を、試料中のがん細胞の比率とする。これは、多くの場合に試料の腫瘍含量、腫瘍純度又は細胞充実度と称される。 ϕ を、試料中の突然変異を内包するがん細胞の比率、即ち、変異体集団におけるがん細胞の相対的比率とする。これは、多くの場合にがん細胞割合（CCF）又は突然変異の細胞保有率と称される。 ϵ を、仮定したシーケンシングエラー率とする。以下の関数が定義される：

【数2】

$$a(G): G \rightarrow \mathbb{N}$$

は、遺伝子型をAアレルの数に写像する関数である（例えば、GがAAの場合、 $a(G) = 2$ ）

【数3】

$$b(G): G \rightarrow \mathbb{N}$$

は、遺伝子型をBアレルの数に写像する関数である（例えば、GがAAの場合、 $b(G) = 0$ ）

【数4】

$$c(G): G \rightarrow \mathbb{N}$$

は、遺伝子型を遺伝子座における全コピー数に写像する関数である（即ち、 $c(G) = a(G) + b(G)$ ；例えばGがAAの場合、 $c(G) = 2$ ）

【数5】

$$\mu(G): G \rightarrow \mathbb{N}$$

10

20

30

40

50

は、遺伝子型を値 $\mu(G) = \min \{ \max \{ (b(G) / c(G)), \varepsilon \}, (1 - \varepsilon) \}$ (これは、遺伝子型 G の集団から突然変異を含むリードが抽出される確率と解釈することができる) に写像する関数である。

【0125】

$\xi(G, \phi, t)$ を、変異体アレルを含むリードが抽出される確率とする。シーケンシング時に抽出される細胞の初期集団は無限にあるものと仮定して、変異体アレルを含むリードが抽出される確率は、DNAの入力プール中の変異体アレルのコピー数におおよそ比例する。より正式には、シーケンシングエラーを考慮すると、変異体アレルが抽出される確率(遺伝子型 G の集合、腫瘍含量 t 及びがん細胞割合 ϕ を所与としたときの)は、以下の式(式(1)):

10

【数6】

$$\xi(G, \phi, t) = \frac{1}{T} (1-t)c(G_H)\mu(G_H) + \frac{1}{T} t(1-\phi)c(G_R)\mu(G_R) + \frac{1}{T} t\phi c(G_V)\mu(G_V) \quad (1)$$

$$\text{式中、 } T = (1-t)c(G_H) + t(1-\phi)c(G_R) + t\phi c(G_V) \quad (2)$$

によって与えられる。

【0126】

変数 $\xi(G, \phi, t)$ は、各遺伝子型を起源とする変異体アレルのコピー数に、その遺伝子型から突然変異を含むリードが抽出される確率を乗じたものを、各遺伝子型を起源とする両方のアレルの全コピー数の和で正規化したものの和を捉える。

20

【0127】

変数 d は、試料中の突然変異をカバーするリードの全数であり、そのうち d_b が突然変異体アレルを持つ。このように、これらのリード数 d 、 d_b が観察される確率 ($P(d, d_b | G, \phi, t)$) は、母数 d_b 及び $\xi(G, \phi, t)$ の二項モデルで表すことができる(式(3))。これは、母数 p の m 個のベルヌーイ確率変数の和が、母数 m 、 p^2 の二項分布に従うためである。代わりに、例えば、データが二項モデルによって説明し得るよりも大きい分散を有する場合には、平均 $\xi(G, \phi, t)$ 及び精度(分散の逆数) γ のベータ二項モデル(式(4)):

30

$$P(d, d_b | G, \phi, t) = \text{二項}(d_b | d, \xi(G, \phi, t)) \quad (3)$$

$$P(d, d_b | G, \phi, t, \gamma) = \text{ベータ二項}(d_b | d, \xi(G, \phi, t), \gamma) \quad (4)$$

を使用することができる。

【0128】

母数 γ は、以下の例では200に設定されるが、他の値も可能である。これまでのところ、本発明者らは、サブ集団の遺伝子型が既知であると仮定している。一般に、これは健康集団(例えば対応する生殖細胞系列試料からの)については正しいこともあり得るが、参照集団及び変異体集団については正しくない。代わりに、突然変異に重なる領域については、アレル特異的コピー数推定値が観察されることが典型的である。この情報を使用して、もっともらしい遺伝子型の集合に関する事前値を引き出すことができる。本発明者らは、これをどのように行うかを次節で説明する。今のところ、ベクトル π の事前確率があるものと仮定しておき、ここで π_i は、集団の i 番目のもっともらしい結合遺伝子型 G_i の事前確率である。もっともらしい遺伝子型全てに関して周辺化する観察データの確率は、以下(式(3a)、式(4a)):

40

$$P(d, d_b | \pi, \phi, t) = \sum_i \pi_i \text{二項}(d_b | d, \xi(G, \phi, t)) \quad (3a)$$

$$P(d, d_b | \pi, \phi, t, \gamma) = \sum_i \pi_i \text{ベータ二項}(d_b | d, \xi(G, \phi, t), \gamma) \quad (4a)$$

のとおり書くことができる。

【0129】

50

後節では、表記法 $Pr(d, d_b | \pi, \phi, t)$ を、式 (3a) 及び式 (4a) の表現を等しく指して使用するものとする。 ϕ 及び t は個別の試料に関連するため、上記の表記法は、それぞれ ϕ_s 及び t_s についての省略形であることに留意されたい。

【0130】

突然変異遺伝子型事前値を引き出す

上記のモデルは、既知の結合遺伝子型、又は事前確率 π のいずれかを使用し、ここで π_i は、集団の i 番目のもっともらしい結合遺伝子型 G_i の事前確率である（即ち G_i は、健常、変異体及び参照集団についての遺伝子型の1つの可能な組み合わせである）。様々な方法を用いて、可能性のある遺伝子型事前値を設定することができる。

【0131】

例えば、1つの可能な方法は、「メジャーコピー数」方法と称することができる。 c_{major} 及び c_{minor} は、腫瘍試料中の突然変異に重なる領域についてのメジャー及びマイナーアレルコピー数を意味するものとする。この方法「メジャーコピー数方法」では、2つの場合を考える：

(a) 第1の場合では、突然変異がコピー数イベントの前に起こる。この場合、参照集団遺伝子型が正常集団と一致する。 c_{major} の変異体を持つ染色体に至るまでの変異体集団について可能な全ての突然変異遺伝子型を考える。

(b) 第2の場合では、突然変異がコピー数イベントの後に起こる。この場合、参照集団は、 $c_{minor} + c_{minor}$ の参照アレルを有する。変異体集団は、1の変異体アレル及び $c_{major} + c_{minor} - 1$ の参照アレルを有する。

【0132】

本発明者らは、事前重みを可能な全ての突然変異遺伝子型に等しく設定する。例えば、 $c_{minor} = 2$ 及び $c_{minor} = 1$ 及び正常コピー数が2であるとする。可能な遺伝子型は以下：

$G_1 = (AA, AA, AAB)$

$G_2 = (AA, AA, ABB)$

$G_3 = (AA, AAA, AAB)$

となり、各々が $1/3$ の事前確率を有する。アレル特異的コピー数が利用可能でない場合には、 c_{major} を全コピー数に設定し、 c_{minor} を0に設定し得ることに留意されたい。この手法は、突然変異が1回のみ起こることを仮定しており、従って変異体集団に2コピー以上の突然変異体アレルが存在する場合には、突然変異が遺伝子座でコピー数変化に先行して起こり、続いて増幅されたため、これが起こったということになる。この手法は、集団の遺伝子型における不確かさを説明することと、一方で考慮する状態が多過ぎないこととの間を上手く両立させる。

【0133】

突然変異遺伝子型事前値の設定には、代替的な手法が用いられてもよい。別の可能な手法は、単純に各突然変異が二倍体且つヘテロ接合体であると仮定することである（即ち、変異体集団中の変異体が2本の染色体のうち一方にのみ現れる、 $G = (G_H = AA, G_R = AA, G_V = AB)$ ）。これは、「AB事前値」と称することができる。なおも別の単純化した手法は、各突然変異が二倍体且つホモ接合体であると仮定することである（即ち、変異体集団中の変異体が2本の染色体の両方に現れる、 $G = (G_H = AA, G_R = AA, G_V = BB)$ ）。これは、「BB事前値」と称することができる。なおも別の可能な単純な手法は、変異体集団の遺伝子型が突然変異の領域に予測される全コピー数を有し、厳密に1つが突然変異体アレルであると仮定することである（即ち、全コピー数が3であると仮定する、 $G = (G_H = AA, G_R = AA, G_V = AAB)$ 、即ち、この結果、上記の「メジャーコピー数」方法において G_1 のみを考えることになる）。これは、「接合性なし事前値」と称することができる。これらの手法は、本質的に単一の可能な遺伝子型を考えるものであるため、多くの場合に単純化し過ぎであり得る。

【0134】

別の可能な手法は、変異体集団の遺伝子型が、少なくとも一方が突然変異体アレルであ

る突然変異の領域に予測される全コピー数を有し、及び参照集団が A A 又は予測される全コピー数に等しいコピー数の遺伝子型のいずれかであり、且つ変異体アレルがない（確率が等しい）と仮定することである。これは、「全コピー数事前値」と称することができ、直感的には、遺伝子座にある変異体集団の遺伝子型が予測される全コピー数を有し、任意のコピー数（ > 0 ）の突然変異体アレルを有し得ることを意味する（即ち、全コピー数が 3 であると仮定すれば、可能な遺伝子型は、確率が等しいとき、 $G_1 = (G_H = A A, G_R = A A, G_V = A A B)$ 、 $G_2 = (G_H = A A, G_R = A A, G_V = A B B)$ 、 $G_3 = (G_H = A A, G_R = A A, G_V = B B B)$ 、 $G_4 = (G_H = A A, G_R = A A A, G_V = A A B)$ ）であり、即ち、これは本質的にメジャー及びマイナーコピー数値を無視し、 n コピーを有する可能な全ての遺伝子型を考える—上記の「メジャーコピー数」方法と比較して、追加的な遺伝子型が考えられることにつながる）。用いることのできるなおも別の手法は、コピー数コーラーから予測されるメジャー及びマイナーアレル数を「信用」することであり、従ってメジャーコピー数又はマイナーコピー数のいずれかに対応する数の突然変異体アレルを有する遺伝子型のみを考えることになる。これは、「親」モードと称することができる。例えば、メジャーコピー数 = 3、マイナーコピー数 = 1 の場合、この手法によれば、確率の等しい、以下の可能な遺伝子型を考えることになる： $G_1 = (A A, A A, A A A B)$ 、 $G_2 = (A A, A A, A B B B)$ 、 $G_3 = (A A, A A A A, A A A B)$ （即ち、変異体集団中の突然変異アレルは 1 個又は 3 個のいずれか）。対照的に、「メジャーコピー数」手法は、1 と予測メジャーコピー数との間にある全ての値を考えることにより、可能なメジャーコピーの範囲を「信頼」するが、その絶対値ではない。メジャーコピー数 = 3、マイナーコピー数 = 1 という上記の例であれば、これにより、「親」モードと比較してさらにもう 1 つの遺伝子型を考えることにつながり得る、即ち、 $G_1 = (A A, A A, A A A B)$ 、 $G_2 = (A A, A A, A A B B)$ 、 $G_3 = (A A, A A, A B B B)$ 、 $G_4 = (A A, A A A A, A A A B)$ 。従って、「メジャーコピー数」手法は、不確かさを考え過ぎることなく（「全コピー数」手法と比較して）、コピー数コールから追加的な不確かさを説明することの間を上手く両立させる（「親の」手法と比較して）。

10

20

30

40

50

【0135】

クロナリティー推定モデル

この節は、遍在性突然変異を同定するための階層ベイズモデルについて概説する。 Z をベルヌーイ変数とし、これは、突然変異が遍在性であるとき（クローナルであると仮定される）1 であり、他の場合には 0 である。 p を、突然変異が遍在性である事前確率とする。これは、以下の例では 0.5 に設定される。上記のとおり、 ϕ は、試料中にある突然変異を内包するがん細胞の比率である。従って、このモデルは、

$$Z \mid p \sim \text{ベルヌーイ}(Z \mid p) \quad (5)$$

$$\phi \mid Z \sim \text{ベータ}(\phi \mid \alpha = 1, \beta = 1), Z = 0 \text{ について};$$

$$\text{ベータ}(\phi \mid \alpha, \beta = 1), Z = 1 \text{ について} \quad (6)$$

$$d_b, d \mid \pi, \phi, t \sim \text{Pr}(d, d_b \mid \pi, \phi, t) \quad (7)$$

（ここで α は、 $\phi \mid Z = 1$ の分布における母数 > 1 である）と表すことができる。これは、以下の例では $\alpha = 99$ に設定される。母数 $\alpha = 99$ 及び $\beta = 1$ のベータ分布は、1 に向かって歪んでおり、クローナルな突然変異であれば、より高いがん細胞割合 ϕ について濃縮されるはずであるという仮定を捉えている。母数の他の値が可能であり、しかしこの仮定を取り込む値が好ましい。上述のとおり、式 (7) の確率は、式 (3) / 式 (3a) 又は式 (4) / 式 (4a) によって与えられる。

【0136】

結合分布は、以下の式で表すことができる（式 (8)）：

1 つの試料について、

$$p(d_b, d, \phi, Z = z \mid \pi, t, p) = p(Z = z \mid p) \text{Pr}(d_b, d \mid \pi, \phi, t) p(\phi \mid Z = z) \quad (8)$$

又は複数の試料について：

【数 7】

$$p(d_b, d, \phi, Z = z | \pi, t, \rho) = p(Z = z | \rho) \prod_{s=1}^S \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = z) \quad (8a)$$

【 0 1 3 7 】

突然変異を内包するがん細胞の比率（ ϕ ）は不明である。しかしながら、1つの試料について、

【数 8】

$$\begin{aligned} p(d_b, d, Z = z | \pi, t, \rho) &= \int_0^1 P(d_b, d, \phi, Z = z | \pi, t, \rho) d\phi \\ &= P(Z = z | \rho) \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = z) d\phi \end{aligned} \quad (9)$$

10

又は複数の試料について：

【数 9】

$$p(d_b, d, Z = z | \pi, t, \rho) = P(Z = z | \rho) \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = z) d\phi \quad (9a)$$

と表すことができる。

【 0 1 3 8 】

20

量

【数 1 0】

$$\prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = z) d\phi$$

は、 ψ_z と称することができる（即ち、それぞれ、突然変異がクローナルでない場合及び突然変異がクローナルである場合のデータの尤度を指す、それぞれ、 ψ_0 及び ψ_1 ）。 $z = 0$ について $p(Z = z | \rho) = (1 - \rho)$ （即ち、突然変異がクローナルである事前確率 ρ が、突然変異がクローナルでない事前確率に等しいことを所与としたときの、 $Z = 0$ の、即ち突然変異が非クローナルに分類される事前確率）、及び $z = 1$ について $p(Z = z | \rho) = \rho$ （即ち、突然変異がクローナルである事前確率の ρ が、突然変異がクローナルである事前確率に等しいことを所与としたときの、 $Z = 1$ の、即ち突然変異がクローナルに分類される事前確率）であるため、従って、複数の試料について（単一の試料についての試料に関する積のない）：

30

【数 1 1】

$$\begin{aligned} p(d_b, d | \pi, t, \rho) &= \sum_{z=0}^1 p(d_b, d, Z = z | \pi, t, \rho) \\ &= (1 - \rho) \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 0) d\phi + \\ &\quad \rho \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 1) d\phi \end{aligned} \quad (10)$$

40

ということになる。

【 0 1 3 9 】

最終的に、推定しようとする量は、観察されるリード（ d_b, d ）、遺伝子型事前値（ π ）、腫瘍割合推定値（ t ）、及び突然変異がクローナルである事前確率（ ρ 、即ち、 $p(Z = 1 | d_b, d, \pi, t, \rho)$ を推定しようとする）を考慮したときの、突然変異がクローナルである確率（ $Z = 1$ の確率）である。上記を考慮すると、これは以下と表すことができる：

【数 1 2】

50

$$p(Z = z | d_b, d, \pi, t, \rho) = \frac{p(d_b, d, Z = z | \pi, t, \rho)}{p(d_b, d | \pi, t, \rho)} \quad (11)$$

ここで $p(d_b, d | \pi, t, \rho)$ は、式 (10) によって与えられ、 $p(d_b, d | Z = z | \pi, t, \rho)$ は、式 (9) / 式 (9a) によって与えられる。従って、式 (11) は、 $Z = 1$ について、以下の式 (11a) と書くことができる：

【数 13】

$$p(Z = 1 | d_b, d, \pi, t, \rho) = \frac{\rho \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 1) d\phi}{(1-\rho) \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 0) d\phi + \rho \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 1) d\phi} \quad (11a)$$

10

ここで ρ は母数であり（以下の例では 0.5 に設定される）、 $p(\phi | Z = z)$ は、式 (6) のベータ分布によって与えられ、及び $\Pr(d_b, d | \pi, t, \rho)$ は、式 (3) / 式 (4)（1つの結合遺伝子型）又は式 (3a) / 式 (4a)（事前確率 π の複数の候補結合遺伝子型）によって与えられる。

【0140】

このように、 $z = 1$ についての式 (11)（即ち式 (11a)）を推定することにより、突然変異が遍在性である（即ち、利用可能な試料を 1 つ以上考慮したときクローナルであると仮定される）確率が与えられる。これには、 S 個の一次元積分（各試料につき 1 つ、式 (9)、式 (10) にある）の評価が必要であり、これは公知の数値積分を用いて効率的に行うことができる。当該技術分野において公知の任意の数値積分アルゴリズムを、この目的に用いることができる。例えば、格子近似を用いてもよい。これは、有利には単純であり、積分する単一の母数（ ϕ ）があることを十分に考える。

20

【0141】

これにより、利用可能なデータを考慮したときの、突然変異がクローナルである確率の推定値が提供され、これは効率的に計算することができ、容易に解釈可能であり（明示的な明確な仮定を利用する厳密な統計モデルを考慮したとき）、手作業での入力なしに任意の突然変異に関して求めることができ、分析される任意の他の突然変異と独立して、突然変異に関する事前知識を厳密に含めることができ、及び検定及び／又は使用のための突然変異のリスト（付随する確率を含む）について客観的且つ自動的に優先順位を付けるのに使用することができる。

30

【0142】

コピー数予測において不確かさを考慮する

上記に記載されるモデルは既に数多くの利点を提供するが、モデルに使用されるコピー数推定値の予測における不確かさを考慮に入れることにより、さらに増強することができる。実際、上記のモデルは、コピー数（例えば、遺伝子型事前値を引き出すために使用されるメジャー／マイナー／全／コピー数）が正確に予測されたと仮定する。実際には、それらの値には、幾らかの不確かさがあり得る。実際、腫瘍のアレル特異的コピー数分析の問題は複雑であり、そのための多くの解決法が提案されている。一般的に使用されるアプローチの一つは、ASCAT (allele-specific copy number analysis of tumors, Van Loo et al., 2010) であり、バルクコピー数プロファイルを解釈するにおいて腫瘍細胞の異数性及び非異常細胞浸潤の両方を考慮に入れるものであり、推定されるアレル特異的コピー数プロファイル及び付随する腫瘍純度推定値を出力する。簡単に言えば、ASCATは、関連するアレル特異的コピー数コールは生殖細胞系列ヘテロ接合一塩基変異多型 (SNP) について可能な限り非負整数に近くなければならないという仮定に基づいて、腫瘍倍数性及び腫瘍割合の可能な複数の組み合わせを評価する。次に、最適と見なされる解（試料の腫瘍部分及び正常部分についての推定される腫瘍倍数性、腫瘍純度及びアレル特異的コピー数コール）が、その適合度（上記の仮定に基づいて）と共に報告される。

40

50

【0143】

上記に提供されるモデルは、予測される各コピー数状態（例えば、メジャー及びマイナーコピー状態を含む提案される各解）からの遺伝子型についての成分が π に含まれるように修正することにより、複数のコピー数解及びそれらの不確かさに適応させて調整し、その状態に関連する確率によって重み付けすることができる。加えて、腫瘍純度推定値が、それらのコピー数状態と共に推定され得るため（例えば、ASCATのようなアプローチが用いられるときの場合のように）、関連する腫瘍純度推定値もまた、考慮に入れることができる。例えば腫瘍純度が別個に推定又は測定され、且つ本質的にコピー数状態推定値に関連しないときには、これは必要でないこともあり得ることに留意されたい。それにもかかわらず、一般化するため、 C 個の可能なコピー数／腫瘍含量状態の集合（例えば、 c_{major} 、 c_{minor} 、及び t の推定値の C 個の可能な集合）を有すると仮定する。 π_C を、各成分が推定値の可能な各集合についての確率であるベクトルとする。各状態 C について、上記に説明したとおりの可能な遺伝子型のベクトル π_{CG} を計算することが可能である。ひいては、 π_{CG} に π_C の状態 C についての成分を乗じることにより、最終的な遺伝子型ベクトルを求めることができる。これにより、僅かに修正された以下の式：

$$P(d, d_b | \pi, \phi) = \sum_i \pi_i \text{二項}(d_b | d, \xi(G_i, \phi, t_i)) \quad (3b)$$

$$P(d, d_b | \pi, \phi, \gamma) = \sum_i \pi_i \text{ベータ二項}(d_b | d, \xi(G_i, \phi, t_i), \gamma) \quad (4b)$$

が生じる。ここで腫瘍含量 t_i は、今や特定の状態に依存する（及び π_i は、 π_{CG} に π_C の状態 C についての成分を乗じることにより求められるベクトル π の要素である）。これらの新しい密度を、上記の関連性のある式に代入することができる。詳細には、解かれる問題は、次には解の式(11a)と表されてもよく、ここで $Pr(d_b, d | \pi, \phi, t)$ は、式(3b)又は式(4b)によって与えられる。 t_i 、 c_{major} 、 c_{minor} の値（ひいては、使用されるモデルに従う適合性のある π_{CG} ）及び π_C は、限定はされないが、上記に説明したとおりのASCATを含め、腫瘍のアレル特異的コピー数分析を実施する多くの方法の出力として提供される。いかなる誤解も避けるために言えば、複数の解の互いに対する相対的な重み付けに使用することのできる信頼度又は他のメトリクスと共に、アレル特異的コピー数状態推定値（典型的には腫瘍純度推定値に関連する）を生成する任意の手法を、この目的に用いることができる。

【0144】

実装

本明細書に記載される方法は、当該技術分野において公知の任意のプログラミング言語を使用して実装することができる。以下の例では、上記の方法を実装するPythonスクリプトを使用した。これは、突然変異毎に、入力として以下を取った：突然変異識別名、試料識別名、突然変異位置で参照アレルと一致するリード数のカウント、突然変異位置でもう一方のアレルと一致するリード数のカウント、及び、1つ以上のコピー数解の各々について：指定したコピー数解について突然変異に重なるメジャーコピー数（腫瘍について）、指定したコピー数解について突然変異に重なるマイナーコピー数（腫瘍について）、突然変異における正常細胞のコピー数（常染色体についてはデフォルト＝2に、又は男性対象の性染色体については1に設定されてもよい）、及び指定したコピー数解についての腫瘍純度値（これはまた、例えばASCATの出力として求めることもでき、又は別個に求めることもできる）。指定したコピー数解について、腫瘍集団の突然変異に重なるメジャー及びマイナーコピー数は、ASCATから直接求めることができ（例えば、ascatNgs、Raine et al., 2016を使用して）、又は突然変異に重なるコピー数セグメントの平均Bアレル頻度、突然変異に重なるコピー数セグメントの対数R値、及び解の倍数性を使用するなど、例えばASCATの出力から導出することができる。例えば、腫瘍についての位置 i におけるアレル特異的コピー数推定値

【数14】

$$(\hat{n}_A, \hat{n}_B)$$

は、

【数 1 5】

$$\hat{n}_A = (\rho - 1 + 2^{\frac{r}{\psi}}(1 - b)(2(1 - \rho) + \rho\psi))/\rho \quad \text{及び} \quad \hat{n}_B = (\rho - 1 + 2^{\frac{r}{\psi}}b(2(1 - \rho) + \rho\psi))/\rho$$

を使用して、位置 i における対数 R 値 r 、位置 i における B アレル割合値 b 、倍数性推定値 ψ 、腫瘍細胞割合推定値 ρ 、及びプラットフォーム依存的「技術」パラメータ t （これは、 WES などの次世代シーケンシングデータについて、 $t = 1$ に設定することができる）の関数として表すことができる。正常集団におけるメジャー及びマイナーコピー数は 1 及び 1 であると仮定することができ、但し性染色体上の突然変異は別とし、これは、対象の性別に応じて取り扱うことができる。複数のコピー数解が提供される場合、各解の確率が任意選択で提供されてもよい（これはまた、ある解について負の対数尤度を証明する例えば $ASCAT$ の出力からも求めることができる）。これが提供されない場合、次には複数の解の全てを同様に確からしいとして処理し、等しい重みを与えることができる。このスクリプトにより、出力として、突然変異識別名及び突然変異が遍在性である事後確率が生じた。

10

【0 1 4 5】

以下の例では、コピー数解を推定したときはいつも、それを $ASCAT$ (Van Loo et al., 2010) を用いて行った。

20

【0 1 4 6】

結果

合成データ

上記に記載される手法について、初めに、式 (3) (二項モデル) 及び式 (4) (ベータ二項モデル) が捉えたとおりの図 4 を参照することにより上記に記載した集団モデルを使用してシミュレートした合成データを使用して試験した。全ての合成データセットについて、合計 1000 個の突然変異をシミュレートした。カバレッジ深度は、現実的なカバレッジレジームに対応する平均値 50、100、1000 のポアソン分布からシミュレートした。コピー数プロファイルは、最大全コピーを 2 又は 6 としてランダムにシミュレートした。これらのセッティングは、ゲノム的に安定したゲノム (最大全コピー数が 2) 及び極めて異数性の高いゲノム (最大全コピー数が 6) に対応する。1 つ、2 つ又は 4 つの試料を含むデータセットをシミュレートした。全例とも、腫瘍含量は 1.0 に固定した。

30

【0 1 4 7】

図 5 A は、データを二項モデルからシミュレートし、上記の方法を二項モデルを用いて実行したときの結果を示す。コピー数 2 のランについては、深度又は試料数に関係なく、性能はほぼ完璧である (曲線下面積、 $AUC = 0.997 \sim 1.0$ 、以下の表 1 を参照のこと)。コピー数 6 のランについては、深度及び試料数の増加に伴い性能はかなり向上している ($AUC = 0.853 \sim 1.0$ 、以下の表 1 を参照のこと)。

【0 1 4 8】

40

50

【表 1】

表1. シミュレーションに二項分布モデルを使用した合成データ結果。BB- 200=ベータ二項, $\gamma=200$ 。

分析に使用した モデル	最大コピー数	平均深度	S (試料数)	AUC
BB-200	2	50	1	0.9975431554
二項	2	50	1	0.9975431554
BB-200	2	100	1	0.9981312525
二項	2	100	1	0.9981312525
BB-200	2	1000	1	1
二項	2	1000	1	1
BB-200	2	50	2	1
二項	2	50	2	1
BB-200	2	100	2	1
二項	2	100	2	1
BB-200	2	1000	2	1
二項	2	1000	2	1
BB-200	2	50	4	1
二項	2	50	4	1
BB-200	2	100	4	1
二項	2	100	4	1
BB-200	2	1000	4	1
二項	2	1000	4	1
BB-200	6	50	1	0.852432188
二項	6	50	1	0.852540223
BB-200	6	100	1	0.9012136777
二項	6	100	1	0.90078165
BB-200	6	1000	1	0.961997085
二項	6	1000	1	0.9687399898
BB-200	6	50	2	0.9570949962
二項	6	50	2	0.9567547662
BB-200	6	100	2	0.9732051405
二項	6	100	2	0.9741773894
BB-200	6	1000	2	0.9962227762
二項	6	1000	2	0.9977032559
BB-200	6	50	4	0.9923878205
二項	6	50	4	0.9924078526
BB-200	6	100	4	0.9974953588
二項	6	100	4	0.9978354459
BB-200	6	1000	4	0.9995103528
二項	6	1000	4	0.999959865

【0149】

図5Bは、精度を100とし、及び上記のモデルランについて二項エミッションモデルを使用してベータ二項分布からシミュレートすることにより行った同じ分析を示す。これは、観察されるデータが予想よりも大きく変化し易い場合をシミュレートする。結果は、コピー数2のシミュレーションについては依然としてほぼ完璧である（AUC 0.997~1.0、表2）。しかしながら、コピー数6のシミュレーションについては、結果は（依然として良好であるものの）前出の分析ほど良くはない（AUC 0.846~0.996、表2）。カバレッジ深度が増加しても、性能は以前ほど大きくは向上しない。対照的に、試料数を増加させることには、実質的な効果がある。

【0150】

図5Cは、精度100のベータ二項分布を用いてデータをシミュレートし、精度200

のベータ二項エミッションモデルを使用して上記の方法で分析したときの結果を示す。性能は、二項エミッションによる上記の方法を用いた前出の分析と同様である（表2）。超過変動データに対してベータ二項モデルを使用すると、同じ超過変動データに対して二項モデルを使用するのと比較して、特に低いシーケンシング深度及び／又は高いコピー数で実に僅かに良好な性能が示された。しかしながら、この例では、性能の向上は有意でなく、従って、有利には、いずれの場合にもより単純な二項モデル（これは、精度の高い超母数を設定する必要がないため、母数化がより単純である）を使用し得るであろう。

【0151】

【表2】

表2. シミュレーションに精度100のベータ二項エミッション分布を使用した合成データ結果。BB=ベータ二項

分析に使用したモデル	最大コピー数	平均深度	S (試料数)	AUC
BB-200	2	50	1	0.9975810324
二項	2	50	1	0.9975810324
BB-200	2	100	1	0.9935050647
二項	2	100	1	0.9935050647
BB-200	2	1000	1	0.9983754737
二項	2	1000	1	0.9983754737
BB-200	2	50	2	1
二項	2	50	2	1
BB-200	2	100	2	0.9999919783
二項	2	100	2	0.9999919783
BB-200	2	1000	2	1
二項	2	1000	2	1
BB-200	2	50	4	1
二項	2	50	4	1
BB-200	2	100	4	1
二項	2	100	4	1
BB-200	2	1000	4	1
二項	2	1000	4	1
BB-200	6	50	1	0.8460939595
二項	6	50	1	0.8455535942
BB-200	6	100	1	0.8628064402
二項	6	100	1	0.859359738
BB-200	6	1000	1	0.908828000
二項	6	1000	1	0.899468000
BB-200	6	50	2	0.955679568
二項	6	50	2	0.9547434743
BB-200	6	100	2	0.9603051358
二項	6	100	2	0.9590724257
BB-200	6	1000	2	0.9685954382
二項	6	1000	2	0.9525490196
BB-200	6	50	4	0.9944876472
二項	6	50	4	0.9943236367
BB-200	6	100	4	0.9937650015
二項	6	100	4	0.9932320614
BB-200	6	1000	4	0.9990039641
二項	6	1000	4	0.9961278606

【0152】

細胞株混合データ

次に、細胞株を混合することにより生成されたデータを使用してこの方法を試験した（

10

20

30

40

50

Farahani et al., 2017)。2つの実験からデータを入力し、ここでは、1) 二倍体細胞株を混合した、及び2) 異数体細胞株を混合した。複数の混合物を表3に記載されたとおりの様々な比率で形成した。各実験に、1) 細胞株Aにユニークな突然変異；2) 細胞株Bにユニークな突然変異及び3) 両方の細胞株が共有する突然変異、の3つの突然変異クラスが存在する。本明細書に記載される方法を試験する目的上、共有された突然変異は遍在性突然変異として扱い、細胞株特異的突然変異はサブクロナルな突然変異として扱った。発行者のウェブサイトからダウンロードしたFarahani et al.の補足表S 1 A及びS 1 Bから、アレルカウントデータ及びコピー数情報を抽出した。予測されたメジャー／マイナーコピーのみが各突然変異に与えられるため、このデータでコピー数の不確かさを考慮することは不可能であったことに留意されたい。

10

【0153】

二項エミッションモデルによるランである上記に記載される方法を、1, 000反復は燃焼として破棄した10, 000反復についてのPyClone (Roth et al., 2014) バージョン0.13.1ランと比較した。PyCloneで遍在性突然変異を同定するため、初めに全ての試料で最も高いCCFのクラスターを同定した。このクラスターにある突然変異は全て、遍在性として扱い、他の全ての突然変異はサブクロナルとして扱った。この戦略は信頼度値を割り当てて遍在性を予測するものではないため、従ってPyCloneの結果はROC曲線上に単一の点としてプロットされることに留意されたい。対照的に、本明細書に記載される方法は各突然変異に遍在性の確率を割り当てるものであり、そのためこれは、曲線をプロットすることが可能である。従って、突然変異をクロナル／サブクロナルであると見なすために使用される閾値を、精度及び再現率の所望の均衡に応じて調整することもまた可能である。

20

【0154】

【表3】

表3. 細胞株混合比率

混合物ID	細胞株A	細胞株B
2	0	1
14	0	1
7	0.1	0.9
19	0.1	0.9
6	0.25	0.75
18	0.25	0.75
5	0.5	0.5
17	0.5	0.5
4	0.75	0.25
16	0.75	0.25
3	0.9	0.1
15	0.9	0.1
1	1	0
13	1	0

30

【0155】

図6Aは、二倍体細胞株混合物の結果を示す。図6Aの左側のプロットは、単一の試料のみを使用したランからのものである。これらは、両方の手法について、90%の細胞株A及び10%の細胞株Bである混合物S3を使用すると最も悪かった。細胞株Aからの突然変異は共有される突然変異に近いCCFを有し得るため、これは意味が通っている。混合物が一層均衡のとれたものになり、細胞株における突然変異のCCFが1から離れるに従い、性能は向上する。図6Aの右側は、複数の試料を使用したランからの結果を示す。図6Bは、異数体細胞株の混合物からの結果を示す。これらの結果は、概して、二倍体の場合と一致しているが、予想どおり、性能は両手法とも一様に悪くなっている。

40

【0156】

50

このデータは、予想どおり、より多くの試料を加えるほど性能が向上する可能性があることを指示している。シミュレーションデータ及びここで分析した人工的な試料という特定のセットアップでは、一部の試料について本明細書に記載される方法及びPyCloneの性能は同様であるが、一部の複数試料の例では、本明細書に記載される方法の方がPyCloneよりも優れていた。

【0157】

各方法の非公式な測時試験により、本明細書に記載される方法が1回の分析毎に約10秒かかったことが明らかになった（ここで分析は、患者に同定される全ての突然変異について、クロナリティーの尤度を決定することを含む）。PyCloneは、同じ分析に対し、実施したMCMC反復回数に応じて300～3000秒程度と、大幅に長い時間がかかった。従って、患者に同定される全ての突然変異を同じ分析で評価するときであっても（これらの評価は完全に独立しているため、これは本方法の要件ではない）、本方法は先行技術の方法よりも少なくとも30倍速い。突然変異の評価を順番に実施するのではなく、並行して実施したならば（これは先行技術の方法を用いては可能でない）、この差は尚更大きくなり得る。

【0158】

このように、このデータは、本明細書に記載される手法が、複数試料の場合に、比較となる現在の技術水準の手法に少なくとも匹敵し、及びそれより優れていることを実証している。可能な場合に複数の試料を収集することは、本明細書に記載される方法など、それを生かすことのできる方法を使用したとき、真にクローナルな変異体が同定される尤度を増加させる可能性があると考えられるため、臨床的及び実践的意味は大きい。本明細書に記載される方法と対照的に、PyCloneは、突然変異特異的予測を提供する能力、かかる各予測についての信頼度推定値を提供する能力、又はコピー数解の不確かさを考慮に入れる能力がなく、手作業での入力並びに同じデータを分析するにはるかに長い計算時間が必要である。

【0159】

このように、本明細書に記載される方法は、先行技術の方法と比較して先行技術の方法を同様の性能で用いることができるであろう特殊な場合（上記のシミュレーション及び人工的試料など）に加え、実践的臨床的文脈で特に意味のある幾つかの利点を有する。例えば、少なくとも単一のコピー数解のキュレーション及び選択に、大がかりな手作業での入力を要するというのは、かかる専門家による入力がない場合には、本明細書に記載される方法は、単一試料の状況であっても、先行技術の方法よりもはるかに良好な性能を示すであろうことを意味する。さらに、専門家による手作業での入力を利用可能であっても、上記で研究した人工的な制御された状況と比べて正しいコピー数解の同定がはるかに誤りの発生し易いものとなる現実の（即ち患者からの実際の腫瘍試料を使用した）状況では、本明細書に記載される方法は、先行技術の方法よりもはるかに良好な性能を示すであろう。

【0160】

TRACERxデータ

本明細書に記載される方法を、CRUK TRACERx研究 (<http://tracex.co.uk/>) からのデータに対してさらに試験した。このデータは、以前にTRACERxパイプラインを用いて分析されており (McGranahan et al., 2016を参照)、PyClone方法に基づくクローナルのコールが利用可能であった。加えて、全ての試料について、ASCATの倍数性／純度の解のデータが利用可能であった。データの概要は図7Aに示す (PyCloneを使用して入手した、利用可能なクローナルのコールに基づく)。このように、コピー数の導出に使用した選択の倍数性／純度の解に関して不確かさを考慮して本方法を実行することが可能であった。

【0161】

図7Bは、1) 全てのASCAT純度／倍数性の解を考えたとき、及び2) 最大限に可能なASCATの解を考えたときの、本明細書に記載される方法からの予測された遍在性の確率の分布を示す。本方法では、純度／倍数性の不確かさを考えたとき、より高い確率が遍

10

20

30

40

50

在性突然変異に系統的に割り当てられたことが認められる。図7Cは、2つの手法を比較するROC曲線を示す。倍数性及び純度の不確かさを考慮すると、本方法の性能は大幅に向上し、AUCは0.76から0.91になった。この分析の一つの注意点は、キュレートしたPyClone結果が真であると仮定されたことである（上記で分析したシミュレーションデータ及び人工的試料に反して、グラウンドトゥールズを得ることができないため）。前節は、本明細書に記載される方法が、現実の腫瘍試料及び複数の試料の状況についてPyCloneよりも優れていることを示唆しているため、本明細書に記載される方法の実際の性能は、ここに報告されるよりも良好である可能性がある。

【0162】

先行技術の方法との実行時間比較

本明細書に記載される方法を、最新版のPyClone (PyClone-VI, Gills & Roth, BMC Bioinformatics volume 21, Article number: 571 (2020)) と並べて公式に時間を測定した。この分析の結果は図8に示す（ここで「ACE」は、本明細書に記載されるとおりの方法を意味する）。このバージョンのPyCloneは大幅に最適化されているため、従って上記の非公式の測時試験で使用した元のバージョンよりも速い。しかしながら、これは必然的にがんと同定される全ての腫瘍特異的突然変異を一緒に分析するため、突然変異数の増加に伴い実行時間が大幅に増加し、これは並列処理化によって対策を打てるものではない。対照的に、本開示の方法は各突然変異を独立に分析するため、完全な並列処理化が可能である。結果として、本方法に関連する実行時間は、いずれの場合にも大幅に短くなり、ここでは単一のがんについての分析を複数のCPU上で実行することができる（図8を参照のこと、これは2、3又は4個のCPUについての結果を示す）。例として、200～1000個の突然変異（これは臨床的に現実的な範囲である）を持つ単一の患者について、本方法を用いて各突然変異についてクローナルである確率が同定されるまでの時間は、12.8秒～36.5秒（それぞれ、200個及び1000個の突然変異について）という短さであり得る。対照的にPyClone-VIの実行にかかる時間は、16.8～99.1秒である（それぞれ、200個及び1000個の突然変異について）。加えて、この時間推定値には、PyCloneによって生み出されたクラスター構造を手作業で分析することによる、クローナルであると仮定される突然変異のクラスターを選ぶ時間、並びに選んだクラスター内にある突然変異に優先順位を付ける時間は含まれない。対照的に、本明細書に記載される方法は、各突然変異がクローナルであるという直接解釈可能な独立した確率を生み出し、突然変異の優先順位付けに直接（即ちクロナリティーに関するさらなる評価ステップなしに）使用することができる。これは、大幅な省力化並びにプロセスの取り扱い易さの増加に相当する。

【0163】

考察

これらの例は、患者のがん細胞集団全体に遍在的に存在する突然変異、即ちクローナルな突然変異を同定するための新規モデルの開発及び評価を実証している。この研究には、以前の解決法と比較して性能を大幅に向上させる幾つかの技術革新がある。詳細には、本発明者らは、この問題を分類問題として書き直し、予測のための確率を提供する。これは、バイナリな答えを返す手法を提供する代わりに、突然変異をランク付けする能力を提供する。さらには、本発明者らは、重複するコピー数に関する不確かさをどのように考慮することができるかを示す。現実世界のデータでは、これはクローナルな突然変異の同定精度の点で性能の大幅な増加につながる。これは実際には、本明細書に記載される方法によって同定されるクローナルな突然変異は、先行技術の方法を用いて同定された突然変異よりも真にクローナルである可能性が一層高いことを意味し、より高い検証率及び最終的にはより高い臨床的成功率につながる。本明細書に記載される方法の他の利益としては、他の突然変異のスコアを変えずに新規突然変異を追加する能力が挙げられる。さらには、本方法のオーダーは、PyClone及び同様の他のクラスタリングベースの手法よりも速い。

【0164】

参考文献

- Adalsteinsson, V.A., Ha, G., Freeman, S.S. et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* 8, 1324 (2017).
- Bulik-Sullivan B, Busby J, Palmer CD, Davis MJ, Murphy T, Clark A, Busby M, Duke F, Yang A, Young L, Ojo NC, Caldwell K, Abhyankar J, Boucher T, Hart MG, Makarov V, Montpreville VT, Mercier O, Chan TA, Scagliotti G, Bironzo P, Novello S, Karachaliou N, Rosell R, Anderson I, Gabra il N, Hrom J, Limvarapuss C, Choquette K, Spira A, Rousseau R, Voong C, Rizvi NA, Fadel E, Frattini M, Jooss K, Skoberne M, Francis J, Yelensky R. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nat Biotechnol*. 2018 Dec 17.
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhir R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012 May;30(5):413-21.
- Hosseini Farahani, Camila PE de Souza, Raewyn Billings, Damian Yap, Karey Shumansky, Adrian Wan, Daniel Lai, Anne-Marie Mes-Masson, Samuel Aparicio, and Sohrab P Shah. Engineered in-vitro cell line mixtures and robust evaluation of computational methods for clonal decomposition and longitudinal dynamics in cancer. *Scientific Reports*, 7(1): 13467, 2017.
- Vanessa Jurtz, Sinu Paul, Massimo Andreatta, Paolo Marcatili, Bjoern Peters and Morten Nielsen. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* November 1, 2017, 199 (9) 3360-3368.
- Langmead, B., Trapnell, C., Pop, M. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).
- Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res*. 2008 Jul 1;36(Web Server issue):W509-12.
- Nicholas McGranahan, Francesco Favero, Elza C de Bruin, Nicolai Juul Birkbak, Zoltan Szallasi, and Charles Swanton. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Science translational medicine*, 7(283):283ra54- 283ra54, 2015.
- McGranahan, N., Furness, A. J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., Jamal-Hanjani, M., Wilson, G. A., Birkbak, N. J., Hiley, C. T., Watkins, T. B., Shafi, S., Murugaesu, N., Mitter, R., Akarca, A. U., Linares, J., Marafioti, T., Henry, J. Y., Van Allen, E. M., Miao, D., ... Swanton, C. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (New York, N.Y.)*, 351 (6280), 1463-1469.
- Timothy J. O'Donnell, Alex Rubinsteyn, Maria Bonsack, Angelika B. Riemer, Uri Laserson, Jeff Hammerbacher. MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Systems* Vol. 7, Issue 1, 129-132, July 25, 2018.
- Andrew Roth, Jaswinder Khattra, Damian Yap, Adrian Wan, Emma Laks, Justina Biele, Gavin Ha, Samuel Aparicio, Alexandre Bouchard-Cote, and Sohrab P Shah. PyClone: statistical inference of clonal population struc

ture in cancer. *Nature methods*, 11(4):396, 2014.

Russell Schwartz and Alejandro A Schaffer. The evolution of tumour phylogenetics: principles and practice. *Nature Reviews Genetics*, 18(4):213, 2017.

Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Borresen-Dale AL, Kristensen VN. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA*. 2010 Sep 28;107(39):16910-5.

Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, Wan Y, Zhang W, Shukla SA, Vartanov A, Fernandes SM, Saksena G, Cibulskis K, Tesar B, Gabriel S, Hacohen N, Meyerson M, Lander ES, Neuberg D, Brown JR, Getz G, Wu CJ. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013 Feb 14;152(4):714-26. doi: 10.1016/j.cell.2013.01.019.

Raine KM, Van Loo P, Wedge DC, Jones D, Menzies A, Butler AP, Teague JW, Tarpey P, Nik-Zainal S, Campbell PJ. ascatNgs: Identifying Somatic Acquired Copy-Number Alterations from Whole-Genome Sequencing Data. *Curr Protoc Bioinformatics*. 2016 Dec 8;56:15.9.1-15.9.17. doi: 10.1002/cpbi.17.

Litchfield K, Reading JL, Puttick C, Thakkar K, Abbosh C, Bentham R, Watkins TBK, Rosenthal R, Biswas D, Rowan A, Lim E, Al Bakir M, Turati V, Guerra-Assungao JA, Conde L, Furness AJS, Saini SK, Hadrup SR, Herrero J, Lee SH, Van Loo P, Enver T, Larkin J, Hellmann MD, Turajlic S, Quezada SA, McGranahan N, Swanton C. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell*. 2021 Feb 4;184(3):596-614.e14.

【0165】

本明細書に引用する全ての参考文献は、各個別の刊行物又は特許又は特許出願が全体として参照により援用されることが具体的且つ個別的に指示されたものとみなすのと同程度に、全体として及びあらゆる目的から参照により本明細書に援用される。

【0166】

本明細書に記載される具体的な実施形態は、限定としてではなく、例として提供される。当業者には、記載されるとおりの本技術の範囲及び趣旨から逸脱することなく、本技術についての記載される組成物、方法、及び使用の様々な改良例及び変形例が明らかであろう。本明細書にあるいずれの小見出しも、便宜上含まれるに過ぎず、本開示を何らかの形で限定するものと解釈されてはならない。

【0167】

本明細書に記載される任意の実施形態の方法は、コンピュータ上で実行されると、上記に記載される1つ又は複数の方法を実施するように構成されるコンピュータプログラムとして、又はコンピュータプログラムを搭載したコンピュータプログラム製品若しくはコンピュータ可読媒体として提供されてもよい。

【0168】

文脈上特に指示がない限り、上記に示す特徴の記載及び定義は、本発明のいかなる特定の態様又は実施形態にも限定されず、記載されている全ての態様及び実施形態に等しく適用される。

【0169】

本明細書及び特許請求の範囲全体を通じて、以下の用語は、文脈上特に明確に指示されない限り、本明細書において明示的に関連する意味と解釈する。語句「一実施形態において」は、本明細書で使用されるとき、必ずしも同じ実施形態を指すわけではなく、但しそ

うであってもよい。さらに、語句「別の実施形態において」は、本明細書で使用されるとき、必ずしも異なる実施形態を指すわけではなく、但しそうであってもよい。従って、以下に記載するとおり、本発明の様々な実施形態は、本発明の範囲又は趣旨から逸脱することなく容易に組み合わせることができる。

【0170】

本明細書及び添付の特許請求の範囲で使用されるとき、単数形「ある(a)」、「ある(an)」、及び「その(the)」は、文脈上特に明確に指示されない限り複数の指示対象を含むことに留意しなければならない。範囲は、本明細書では、「約」ある特定の値から、及び／又は「約」別の特定の値までと表現され得る。かかる範囲が表現されるとき、別の実施形態には、そのある特定の値から、及び／又は他の特定の値までが含まれる。同様に、値が先行する語句「約」の使用によって近似で表現されるとき、その特定の値が別の実施形態を形成することは理解されるであろう。数値との関連において用語「約」は、任意選択であり、例えば±10%を意味する。

10

【0171】

以下の特許請求の範囲を含め、本明細書全体を通じて、文脈上特に必須でない限り、語句「～を含む(comprise)」及び「～を含む(include)」、並びに変化形、例えば、「～を含む(comprises)」、「～を含んでいる(comprising)」、及び「～を含んでいる(including)」などは、記載されている完全体又はステップ又は完全体若しくはステップの群の包含を含意するが、任意の他の完全体又はステップ又は完全体若しくはステップの群の除外は含意しないと理解されることになる。

20

【0172】

本発明の他の態様及び実施形態は、文脈上特に指示がない限り、上記に記載される態様及び実施形態であって、用語「～を含んでいる(comprising)」が用語「～からなる(consisting of)」又は「～から本質的になる(consisting essentially of)」によって置き換えられたものを提供する。

【0173】

「及び／又は」は、本明細書で使用される場合に、2つの指定される特徴又は成分の各々の他方を伴う又は伴わない具体的な開示と解釈されるべきである。例えば「A及び／又はB」は、(i) A、(ii) B及び(iii) A及びBの各々の、あたかも各々が個別に本明細書に示されたかのような具体的な開示と解釈されるべきである。

30

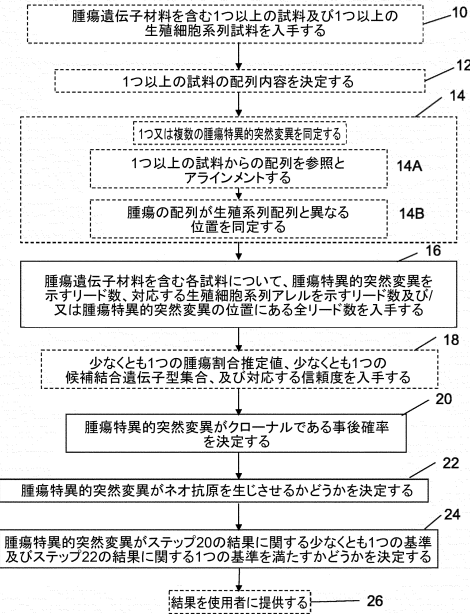
【0174】

適切な場合には、その具体的な形態で表した、又は開示される機能を実施する手段、又は開示される結果を達成するための方法若しくはプロセスの観点で表した前述の記載、又は以下の特許請求の範囲、又は添付の図面に開示される特徴は、本発明をその多様な形態で実現するために、別個に、又はかかる特徴の任意の組み合わせで利用されてもよい。

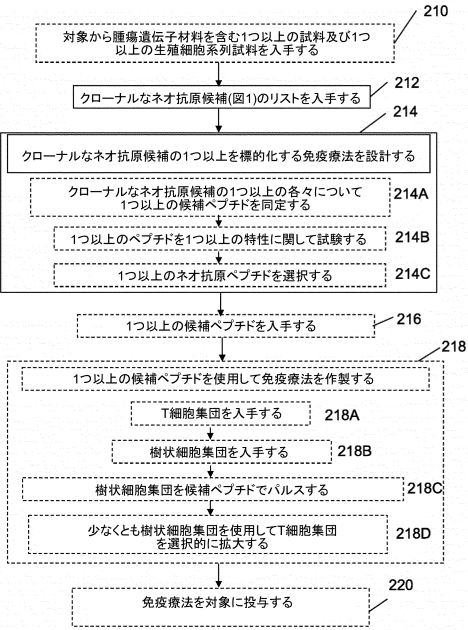
40

50

【図面】
【図 1】



【図 2】



【図 3】

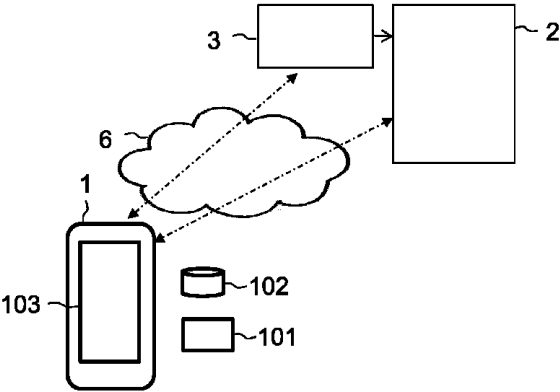
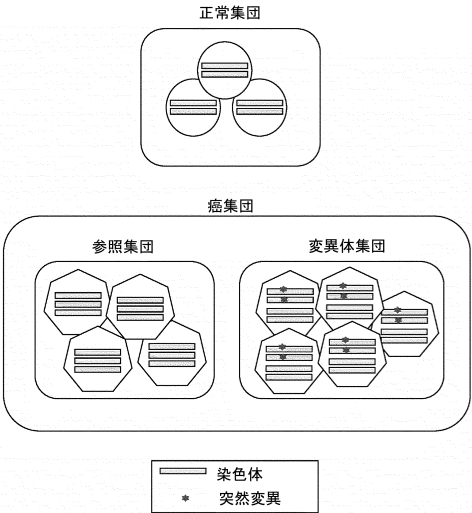


Fig. 3

【図 4】



10

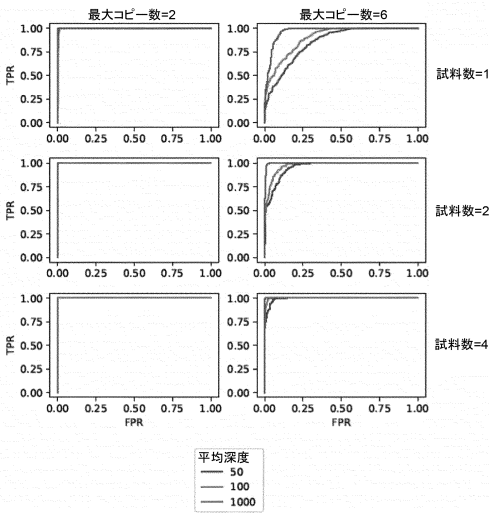
20

30

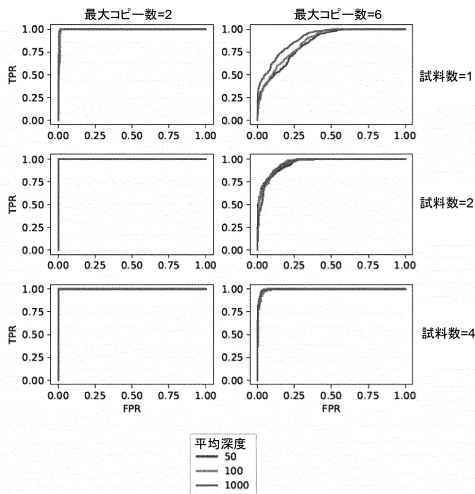
40

50

【図 5 A】



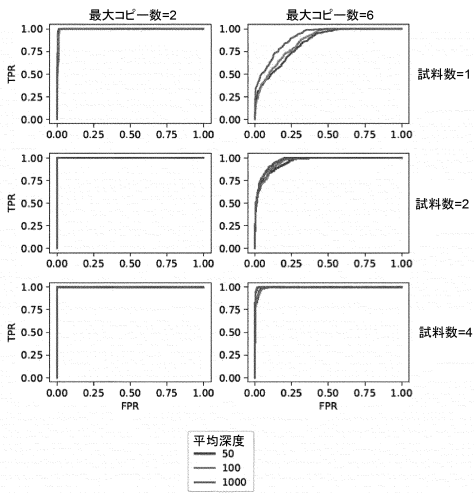
【図 5 B】



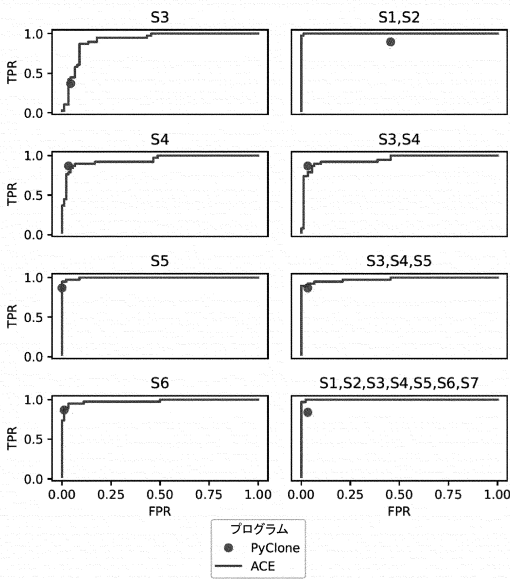
10

20

【図 5 C】



【図 6 A】

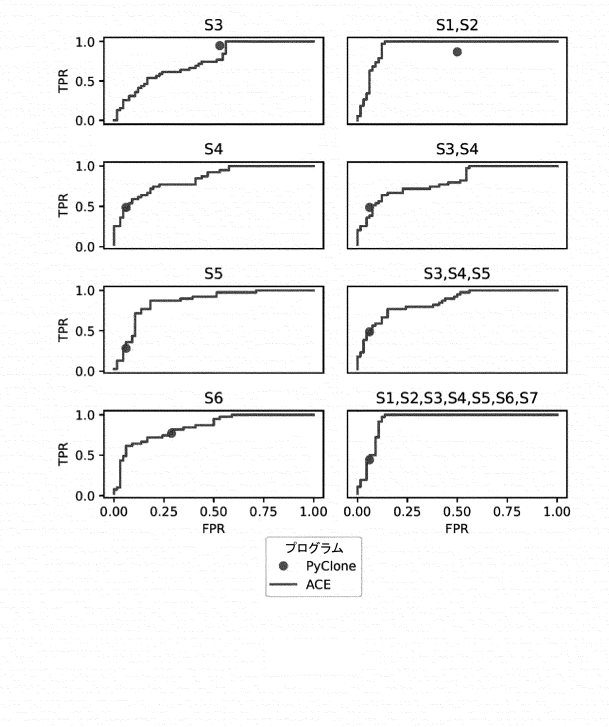


30

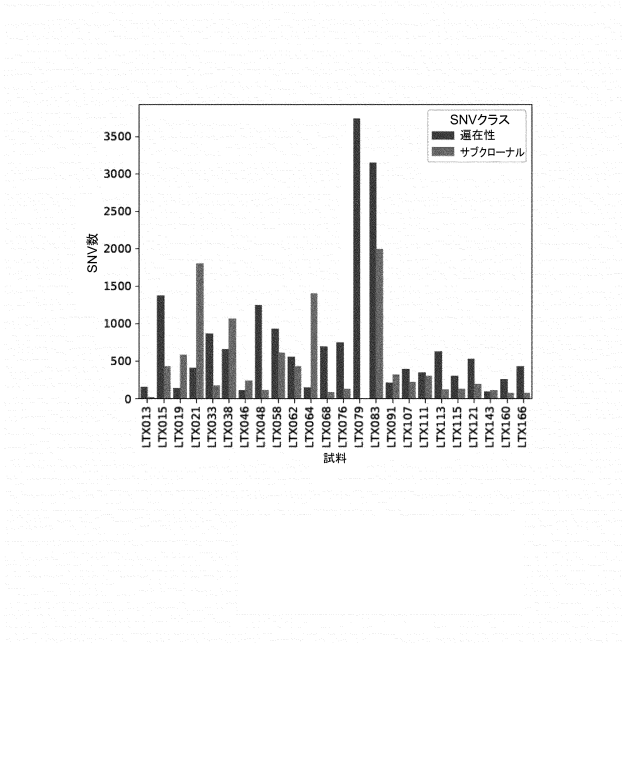
40

50

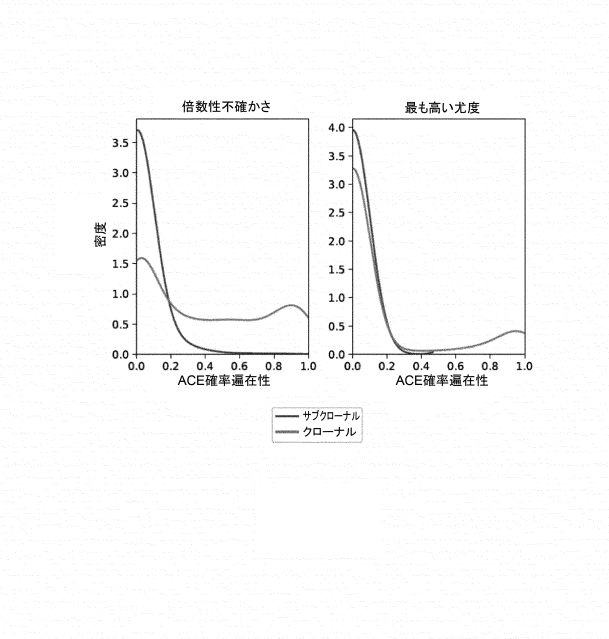
【図 6 B】



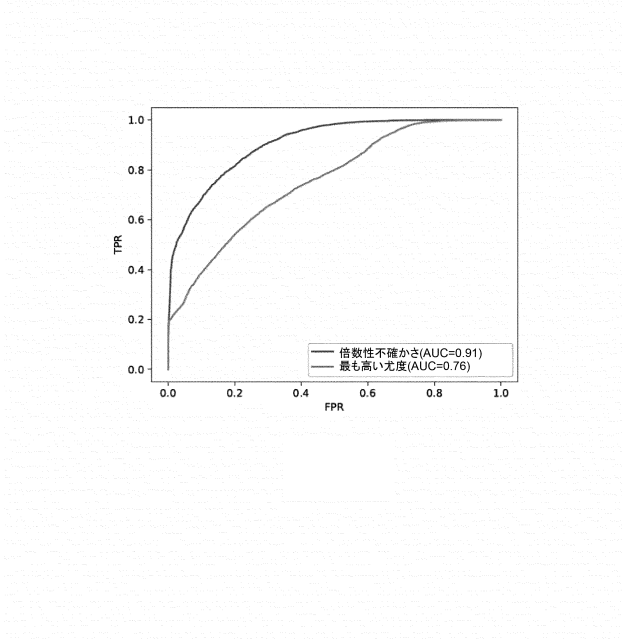
【図 7 A】



【図 7 B】



【図 7 C】



10

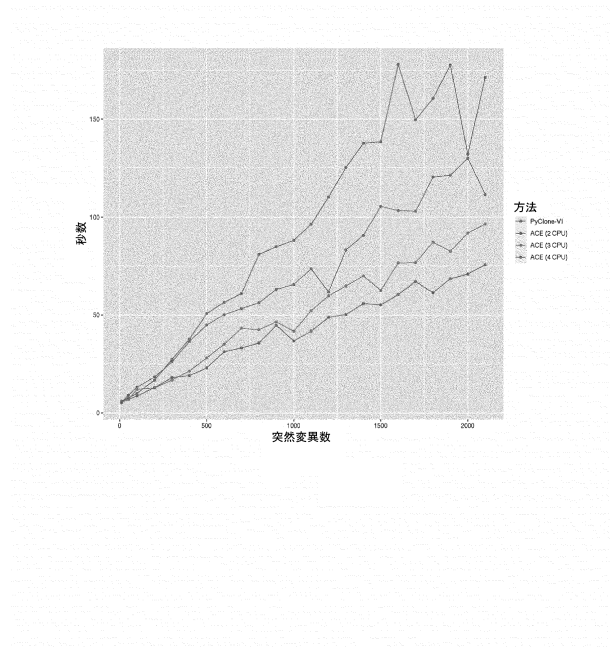
20

30

40

50

【図 8】



10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2022/058793

A. CLASSIFICATION OF SUBJECT MATTER INV. A61K39/00 C12Q1/6869 G16B20/20 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06F A61K C12Q G16B		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, CHEM ABS Data, EMBASE, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 111 785 322 A (HANGZHOU MUGU GENE TECH CO LTD) 16 October 2020 (2020-10-16) the whole document -----	1-22
X	WO 2019/012296 A1 (THE FRANCIS CRICK INSTITUTE LTD [GB]; UNIV LONDON [GB] ET AL.) 17 January 2019 (2019-01-17) abstract; claim 12; examples 1,2 page 14, line 24 - page 16, line 4 page 16, line 6 - line 20 page 22, line 24 - page 24, line 4 -----	1-16, 19-22
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 21 July 2022		Date of mailing of the international search report 29/07/2022
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Celler, Jakub

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No
PCT/EP2022/058793

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
CN 111785322	A	16-10-2020	NONE	
WO 2019012296	A1	17-01-2019	AU 2018298971 A1	16-01-2020
			CA 3068203 A1	17-01-2019
			CN 110913896 A	24-03-2020
			EP 3651791 A1	20-05-2020
			JP 2020526209 A	31-08-2020
			US 2021147934 A1	20-05-2021
			WO 2019012296 A1	17-01-2019

10

20

30

40

50

フロントページの続き

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

(特許庁注：以下のものは登録商標)

1. PYTHON

ビー オックスフォードシャー オックスフォード ウェスト ウェイ 3 バクストン コート オッ
クスフォード ユニバーシティ イノベーション リミテッド内

(72)発明者 サルムーホルストマール, マクシミリアン プリンツ ツー
イギリス国, ダブリュ7 8ピーダブリュ グレーター ロンドン ロンドン ハマースミス ロード
2 4 5 アキレス セラピューティクス ユーケー リミテッド内

(72)発明者 チャン, フォン チュン
イギリス国, ダブリュ7 8ピーダブリュ グレーター ロンドン ロンドン ハマースミス ロード
2 4 5 アキレス セラピューティクス ユーケー リミテッド内