

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2021-526791

(P2021-526791A)

(43) 公表日 令和3年10月11日 (2021. 10. 11)

(51) Int. Cl.	F I	テーマコード (参考)
C 1 2 Q 1/02 (2006. 01)	C 1 2 Q 1/02	4 B 0 2 9
C 1 2 Q 1/6813 (2018. 01)	C 1 2 Q 1/6813 Z	4 B 0 6 3
C 1 2 Q 1/686 (2018. 01)	C 1 2 Q 1/686 Z	4 C 0 8 4
C 1 2 Q 1/6869 (2018. 01)	C 1 2 Q 1/6869 Z	5 L 0 9 9
C 1 2 M 1/00 (2006. 01)	C 1 2 M 1/00 A	
審査請求 未請求 予備審査請求 未請求 (全 73 頁) 最終頁に続く		

(21) 出願番号 特願2020-567550 (P2020-567550)
 (86) (22) 出願日 令和1年6月3日 (2019. 6. 3)
 (85) 翻訳文提出日 令和3年1月28日 (2021. 1. 28)
 (86) 国際出願番号 PCT/US2019/035214
 (87) 国際公開番号 W02019/236478
 (87) 国際公開日 令和1年12月12日 (2019. 12. 12)
 (31) 優先権主張番号 62/680, 301
 (32) 優先日 平成30年6月4日 (2018. 6. 4)
 (33) 優先権主張国・地域又は機関
 米国 (US)

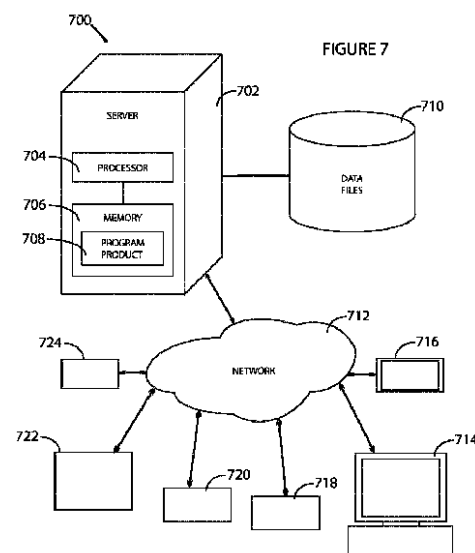
(71) 出願人 515059083
 ガーダント ヘルス、 インコーポレイテッド
 アメリカ合衆国 カリフォルニア 940
 63, レッドウッド シティ, ペノブ
 スコット ドライブ 505
 (74) 代理人 100078282
 弁理士 山本 秀策
 (74) 代理人 100113413
 弁理士 森下 夏樹
 (74) 代理人 100181674
 弁理士 飯田 貴敏
 (74) 代理人 100181641
 弁理士 石川 大輔

最終頁に続く

(54) 【発明の名称】 セルフリー核酸の細胞起源を決定するための方法およびシステム

(57) 【要約】

液体の生検サンプルなどのセルフリー核酸 (c f N A) サンプル由来の c f N A フラグメントの細胞起源を決定する際に有用な方法が本明細書中に提供される。本明細書中に開示される方法は、通常、ある特定の実施形態において造血幹細胞などの非標的細胞によって生成されるバリエーションアレルを識別することによって c f N A サンプル中の罹患細胞核酸 (例えば、癌細胞 DNA) を検出するためのアッセイの特異度および/または感度を改善する。なおも他の態様は、とりわけ、関連するシステムおよびコンピュータ可読媒体を含む。



【特許請求の範囲】**【請求項 1】**

少なくとも部分的にコンピュータを使用して、被験体内の標的細胞を起源とする核酸分子を検出する方法であって、前記方法は、

(a) 前記被験体から得られた試験サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む試験配列情報を前記コンピュータによって受信する工程；および

(b) 前記試験配列情報における少なくとも 1 つのアレルバリエントを識別する工程；

(c) 前記アレルバリエントを標的核酸バリエントフィルターリスト上の少なくとも 1 つの分類アレルにマッピングする工程；

(d) 前記分類アレルのサブクロナリティスコアを識別する工程；および

(e) 前記サブクロナリティスコアを少なくとも 1 つの選択されたカットオフ閾値と比較する工程であって、ここで、前記サブクロナリティスコアが、前記選択されたカットオフ閾値未満であるとき、それは、前記分類アレルが、前記標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それによって、前記被験体内の前記標的細胞を起源とする核酸分子を検出する、工程を含む、方法。

【請求項 2】

少なくとも部分的にコンピュータを使用して、被験体内の腫瘍細胞を起源とする核酸分子を検出する方法であって、前記方法は、

(a) 前記被験体から得られた試験サンプル中のセルフリーデオキシリボ核酸 (c f D N A) フラグメントから得られた配列リードを含む試験配列情報を前記コンピュータによって受信する工程；

(b) 前記コンピュータによって、前記被験体の造血幹細胞を起源とする 1 つまたはそれを超える前記配列リードまたは分類アレルを試験配列情報から除去して、フィルターされた前記試験配列情報を生成する工程；および

(c) 1 つまたはそれを超える参照被験体から得られた参照配列情報と実質的にアラインメントする 1 つまたはそれを超える前記配列リードが前記フィルターされた試験配列情報に存在することを前記コンピュータによって識別する工程であって、その参照配列情報は、前記参照被験体内の 1 つまたはそれを超える腫瘍細胞を起源とし、それによって、前記被験体内の前記腫瘍細胞を起源とする前記核酸分子を検出する、工程を含む、方法。

【請求項 3】

被験体における疾患を処置する方法であって、前記方法は、

(a) 前記被験体から得られた試験サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む試験配列情報を受信する工程；

(b) 標的核酸バリエントフィルターリスト上の少なくとも 1 つの分類アレルと実質的にマッチする少なくとも 1 つのアレルバリエントが前記試験配列情報に存在することを識別する工程であって、その分類アレルは、少なくとも 1 つの選択されたカットオフ閾値未満のサブクロナリティスコアを含み、それによって、前記分類アレルは、罹患細胞を起源とする参照 c f N A フラグメント由来であることを示し、それによって、前記被験体における前記疾患を診断する、工程；および

(c) 前記被験体に 1 つまたはそれを超える療法を施し、それによって、前記被験体における疾患を処置する工程を含む、方法。

【請求項 4】

少なくとも部分的にコンピュータを使用して、分類器を生成する方法であって、前記方法は、

(a) 1 つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対す

10

20

30

40

50

るサブクロナリティスコアを、前記コンピュータによって生成する工程であって、ここで、各分類アレルは、潜在的に臨床的に有意であり、前記参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む、工程；および

(b) 前記コンピュータによって、前記サブクロナリティスコアと少なくとも1つの選択されたカットオフ閾値を比較する工程であって、

ここで、前記選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ／または

前記カットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられ、

それによって、前記分類器を生成する工程

を含む、方法。

【請求項5】

少なくとも部分的にコンピュータを使用して、分類器を生成する方法であって、前記方法は、

(a) 1つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む配列情報から分類アレルセットを前記コンピュータによって識別する工程であって、ここで、各分類アレルは、潜在的に臨床的に有意であり、前記参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む、工程；

(b) 前記配列情報から、前記参照サンプルの各々における各分類アレルに対するマイナーアレル頻度 (M A F) の値を前記コンピュータによって決定する工程；

(c) 前記参照サンプルの各々について、最大マイナーアレル頻度 (m a x M A F) の値を前記コンピュータによって決定する工程；

(d) 所与の参照サンプルにおいて観察された各分類アレルについて、前記参照サンプルの少なくとも一部に対する前記 m a x M A F の値に対する前記 M A F の値の比を前記コンピュータによって計算して、比の値を生成する工程；

(e) 前記参照サンプルの少なくとも一部における所与の分類アレルが、前記参照サンプルの少なくとも一部に出現した前記所与の分類アレルの回数の合計数に対する、少なくとも1つの選択されたクロナリティ境界値未満の比の値を有した回数の比を、前記分類アレルの各々について、前記コンピュータによって計算して、前記参照サンプルの少なくとも一部における前記分類アレルの各々に対するサブクロナリティスコアを生成する工程；および

(f) 前記サブクロナリティスコアと少なくとも1つの選択されたカットオフ閾値を前記コンピュータによって比較する工程であって、

ここで、前記選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ／または

前記カットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられ、

それによって、前記分類器を生成する工程

を含む、方法。

【請求項6】

被験体から得られた試験サンプル中のセルフリー核酸 (c f N A) フラグメントの細胞起源を分類する際に使用するサブクロナリティスコアのデータベースを作成する方法であって、前記方法は、

(a) 1つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメ

10

20

30

40

50

ントから得られた配列リードを含む配列情報から分類アレルセットをコンピュータによって識別する工程であって、ここで、各分類アレルは、潜在的に臨床的に有意であり、前記参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む、工程；

(b) 前記配列情報から、前記参照サンプルの各々における各分類アレルに対するマイナーアレル頻度 (MAF) の値を前記コンピュータによって決定する工程；

(c) 前記参照サンプルの各々について、最大マイナーアレル頻度 (maxMAF) の値を前記コンピュータによって決定する工程；

(d) 所与の参照サンプル中に観察された各分類アレルについて、前記参照サンプルの少なくとも一部に対する前記maxMAFの値に対する前記MAFの値の比を前記コンピュータによって計算して、比の値を生成する工程；

(e) 前記参照サンプルの少なくとも一部における所与の分類アレルが、前記参照サンプルの少なくとも一部に出現した前記所与の分類アレル回数の合計数に対する、少なくとも1つの選択されたクロナリティ境界値未満の比の値を有した回数の比を、前記分類アレルの各々について、前記コンピュータによって計算して、前記参照サンプルの少なくとも一部における前記分類アレルの各々に対するサブクロナリティスコアを生成する工程；
および

(f) 対応する分類アレルにインデックスされた前記サブクロナリティスコアをデータベースシステムに非一時的に保存し、それによって、被験体から得られた試験サンプル中のcfNAフラグメントの前記細胞起源を分類する際に使用する前記サブクロナリティスコアのデータベースを作成する工程
を含む、方法。

【請求項7】

前記分類アレルセットを識別する工程が、前記参照サンプルから得られた前記配列情報から、潜在的に臨床的に有意な標的ゲノム遺伝子座セット内の各遺伝子座における各体細胞性核酸バリエーションに対するMAFの値を決定する工程であって、ここで、前記標的ゲノム遺伝子座セットは、各参照サンプルにおいて同一である、工程、および前記参照サンプルの各々についてmaxMAFの値を決定してアレル情報を生成する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項8】

各分類アレルに対する前記MAFが、約2%未満である、先行する請求項のいずれか1項に記載の方法。

【請求項9】

各分類アレルに対する前記MAFが、約1%未満である、先行する請求項のいずれか1項に記載の方法。

【請求項10】

前記参照サンプルにインデックスされた臨床情報を用いて、前記分類器を生成する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項11】

前記試験サンプルにインデックスされた臨床情報を用いて、前記被験体内の前記標的細胞を起源とする前記核酸分子を検出する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項12】

前記臨床情報が、年齢、性別、人種、体重、ボディマス指数 (BMI)、病歴、喫煙および飲酒からなる群より選択される、先行する請求項のいずれか1項に記載の方法。

【請求項13】

前記分類アレルの各々に対する各MAF/maxMAF値の頻度を用いて、サブクロナリティスコアを決定する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項14】

前記選択されたクロナリティ境界値が、約1%～約99%の範囲内である、先行する請求項のいずれか1項に記載の方法。

10

20

30

40

50

【請求項 15】

前記選択されたクローナリティ境界値が、約10%、約20%、約30%、約40%、約50%、約60%、約70%、約80%または約90%である、先行する請求項のいずれか1項に記載の方法。

【請求項 16】

前記選択されたカットオフ閾値が、約1%～約99%の範囲内である、先行する請求項のいずれか1項に記載の方法。

【請求項 17】

前記選択されたカットオフ閾値が、約10%、約20%、約30%、約40%、約50%、約60%、約70%、約80%または約90%である、先行する請求項のいずれか1項に記載の方法。

10

【請求項 18】

前記サブクローナリティスコアを複数の選択されたカットオフ閾値と比較する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 19】

前記複数の選択されたカットオフ閾値が、第1のカットオフ閾値および第2のカットオフ閾値を含み、その第1のカットオフ閾値は、前記第2のカットオフ閾値より大きく、前記第1のカットオフ閾値を超えるサブクローナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレルは、前記非標的核酸バリエーションフィルタリストに加えられ、かつ／または前記第2のカットオフ閾値未満のサブクローナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレルは、前記標的核酸バリエーションフィルタリストに加えられる、先行する請求項のいずれか1項に記載の方法。

20

【請求項 20】

非標的核酸バリエーションフィルタリスト上の少なくとも1つの分類アレルと実質的にマッチする、前記試験配列情報におけるアレルバリエーションが、約1%を超えるMAFを含むとき、前記アレルバリエーションを、標的細胞を起源とすると分類する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 21】

非標的核酸バリエーションフィルタリスト上の少なくとも1つの分類アレルと実質的にマッチする、前記試験配列情報におけるアレルバリエーションが、トランケーション、インデルおよび／またはスプライス部位バリエーションを含むとき、前記アレルバリエーションを、標的細胞を起源とすると分類する工程を含む、先行する請求項のいずれか1項に記載の方法。

30

【請求項 22】

前記参照サンプルの少なくとも一部における所与の分類アレルに対する各比の値の頻度を決定する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 23】

前記分類器を用いて、被験体から得られた試験サンプルが、前記標的細胞を起源とするcfNAフラグメントを含むか否かを決定する工程を含む、先行する請求項のいずれか1項に記載の方法。

40

【請求項 24】

前記分類器を用いて、被験体から得られた試験サンプルが、前記非標的細胞を起源とするcfNAフラグメントを含むか否かを決定する工程を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 25】

先行する請求項のいずれか1項に記載の標的核酸バリエーションフィルタリストおよび／または非標的核酸バリエーションフィルタリストを含む、データベース。

【請求項 26】

前記非標的細胞が、非罹患細胞を含む、先行する請求項のいずれか1項に記載の方法。

50

【請求項 27】

前記非標的細胞が、造血幹細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 28】

前記非標的細胞が、非腫瘍細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 29】

前記非標的細胞が、母体細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 30】

前記非標的細胞が、移植レシピエント細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 31】

前記標的細胞が、罹患細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 32】

前記標的細胞が、腫瘍細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 33】

前記標的細胞が、胎児細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 34】

前記標的細胞が、移植ドナー細胞を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 35】

前記疾患が、癌を含み、前記療法が、少なくとも1つの免疫療法を含む、先行する請求項のいずれか1項に記載の方法。

【請求項 36】

前記被験体が、哺乳動物被験体である、先行する請求項のいずれか1項に記載の方法。

【請求項 37】

前記哺乳動物被験体が、ヒト被験体である、先行する請求項のいずれか1項に記載の方法。

【請求項 38】

前記被験体から前記試験サンプルを得る工程をさらに含む、先行する請求項のいずれか1項に記載の方法。

【請求項 39】

前記試験サンプルが、血液、血漿、血清、痰、尿、精液、膣液、便、滑液、髄液および唾液からなる群より選択される、先行する請求項のいずれか1項に記載の方法。

【請求項 40】

前記試験サンプル中の前記 c f N A フラグメントから前記試験配列情報を生成する工程をさらに含む、先行する請求項のいずれか1項に記載の方法。

【請求項 41】

標的ゲノム遺伝子座を含む前記 c f N A フラグメントのセグメントを増幅して、増幅核酸を生成する工程をさらに含む、先行する請求項のいずれか1項に記載の方法。

【請求項 42】

前記試験サンプル中の前記 c f N A フラグメントをシーケンシングして、前記試験配列情報を生成する工程をさらに含む、先行する請求項のいずれか1項に記載の方法。

【請求項 43】

前記試験配列情報が、前記試験サンプル中の前記 c f N A フラグメントの標的化セグメントから得られ、前記標的化セグメントは、シーケンシングの前に、前記試験サンプル中の前記 c f N A フラグメントから1つまたはそれを超える領域を選択的に濃縮することによって得られる、先行する請求項のいずれか1項に記載の方法。

【請求項 44】

シーケンシングの前に、前記得られた標的化セグメントを増幅する工程をさらに含む、先行する請求項のいずれか1項に記載の方法。

【請求項 45】

シーケンシングの前に、バーコードを含む1つまたはそれを超えるアダプターを前記 c f

10

20

30

40

50

N Aフラグメントおよび／または前記増幅された標的化セグメントに付着させる工程をさらに含む、先行する請求項のいずれか1項に記載の方法。

【請求項46】

前記シーケンシングが、ターゲットシーケンシング、バイサルファイトシーケンシング、イントロンシーケンシング、エクソームシーケンシングおよびホールゲノムシーケンシングからなる群より選択される、先行する請求項のいずれか1項に記載の方法。

【請求項47】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたは前記コンピュータ可読媒体にアクセスできるコントローラを備えるシステムであって、前記命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、

10

(a) 被験体から得られた試験サンプル由来のセルフリー核酸(c f N A)フラグメントから得られた配列リードを含む試験配列情報を受信すること；および

(b) 標的核酸バリエーションフィルタリスト上の少なくとも1つの分類アレルと実質的にマッチする少なくとも1つのアレルバリエーションが前記試験配列情報に存在することを識別することであって、その分類アレルは、少なくとも1つの選択されたカットオフ閾値未満のサブクロナリティスコアを含み、それによって、前記分類アレルが、標的細胞を起源とする参照c f N Aフラグメント由来であることを示し、それによって、前記試験配列情報における前記アレルバリエーションが、前記被験体内の前記標的細胞を起源とすることを示すこと

を行う、システム。

20

【請求項48】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたは前記コンピュータ可読媒体にアクセスできるコントローラを備えるシステムであって、前記命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、

(a) 前記被験体から得られた試験サンプル中のセルフリーデオキシリボ核酸(c f D N A)フラグメントから得られた配列リードを含む試験配列情報を受信すること；

(b) 前記被験体の造血幹細胞を起源とする1つまたはそれを超える前記配列リードまたは分類アレルを前記試験配列情報から除去して、フィルターされた試験配列情報を生成すること；および

(c) 1またはそれを超える参照被験体から得られた参照配列情報と実質的にアラインメントする1つまたはそれを超える前記配列リードが前記フィルターされた試験配列情報に存在することを識別することであって、それらの参照配列情報は、前記参照被験体内の腫瘍細胞を起源とし、それによって、前記試験サンプルが、前記被験体内の前記腫瘍細胞を起源とする1つまたはそれを超えるc f D N Aフラグメントを含むことを示すこと

30

を行う、システム。

【請求項49】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたは前記コンピュータ可読媒体にアクセスできるコントローラを備えるシステムであって、前記命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、

(a) 1つまたはそれを超える参照サンプル由来のセルフリー核酸(c f N A)フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対するサブクロナリティスコアを生成することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、前記参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと；および

40

(b) 前記サブクロナリティスコアと少なくとも1つの選択されたカットオフ閾値を比較すること

を行い、

ここで、前記選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照c f N Aフラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルタリストに

50

加えられ、かつ／または

前記カットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる、システム。

【請求項 5 0】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたは前記コンピュータ可読媒体にアクセスできるコントローラを備えるシステムであって、前記命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、少なくとも、

(a) 1 つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む配列情報から分類アレルセットを識別することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、前記参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと；

(b) 前記配列情報から前記参照サンプルの各々における各分類アレルに対するマイナーアレル頻度 (M A F) の値を決定すること；

(c) 前記参照サンプルの各々について、最大マイナーアレル頻度 (m a x M A F) の値を決定すること；

(d) 所与の参照サンプルにおいて観察された各分類アレルについて、前記参照サンプルの少なくとも一部に対する前記 m a x M A F の値に対する前記 M A F の値の比を計算して、比の値を生成すること；

(e) 前記参照サンプルの少なくとも一部における所与の分類アレルが、前記参照サンプルの少なくとも一部に出現した前記所与の分類アレルの回数の合計数に対する、少なくとも 1 つの選択されたクロナリティ境界値未満の比の値を有した回数の比を、前記分類アレルの各々について計算して、前記参照サンプルの少なくとも一部における前記分類アレルの各々に対するサブクロナリティスコアを生成すること；および

(f) 前記サブクロナリティスコアと少なくとも 1 つの選択されたカットオフ閾値を比較すること

を行い、

ここで、前記選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ／または

前記カットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる、システム。

【請求項 5 1】

前記コントローラと作動可能に接続された核酸配列決定装置を備え、その核酸配列決定装置は、前記試験サンプル中および／または前記参照サンプル中の前記 c f N A フラグメントから前記配列情報を提供するように構成されている、先行する請求項のいずれか 1 項に記載のシステム。

【請求項 5 2】

前記核酸配列決定装置が、前記核酸に対してパイロシーケンシング、バイサルファイトシーケンシング、一分子シーケンシング、ナノポアシーケンシング、半導体シーケンシング、ライゲーションによるシーケンシングまたはハイブリダイゼーションによるシーケンシングを行って、シーケンシングリードを生成するように構成されている、先行する請求項のいずれか 1 項に記載のシステム。

【請求項 5 3】

前記コントローラに作動可能に接続されたサンプル調製コンポーネントを備え、そのサンプル調製コンポーネントは、核酸配列決定装置によってシーケンシングされる前記 c f N

10

20

30

40

50

Aフラグメントを調製するように構成されている、先行する請求項のいずれか1項に記載のシステム。

【請求項54】

前記サンプル調製コンポーネントが、前記試験サンプル中および／または前記参照サンプル中の前記cfNAフラグメントからの領域を選択的に濃縮するように構成されている、先行する請求項のいずれか1項に記載のシステム。

【請求項55】

前記サンプル調製コンポーネントが、バーコードを含む1つまたは複数のアダプターを前記cfNAフラグメントに付着させるように構成されている、先行する請求項のいずれか1項に記載のシステム。

10

【請求項56】

前記コントローラに作動可能に接続された核酸増幅コンポーネントを備え、前記核酸増幅コンポーネントは、前記試験サンプル中および／または前記参照サンプル中の前記cfNAフラグメントを増幅するように構成されている、先行する請求項のいずれか1項に記載のシステム。

【請求項57】

前記核酸増幅コンポーネントが、前記試験サンプル中および／または前記参照サンプル中の前記cfNAフラグメントから、選択的に濃縮された領域を増幅するように構成されている、先行する請求項のいずれか1項に記載のシステム。

【請求項58】

20

前記コントローラに作動可能に接続された材料移動コンポーネントを備え、その材料移動コンポーネントは、1つまたはそれを超える材料を、核酸配列決定装置、核酸増幅コンポーネントおよび／またはサンプル調製コンポーネントの間を移動させるように構成されている、先行する請求項のいずれか1項に記載のシステム。

【請求項59】

前記コントローラに作動可能に接続されたデータベースを備え、そのデータベースは、前記非標的核酸バリエーションフィルターリストおよび／または前記標的核酸バリエーションフィルターリストを含む、先行する請求項のいずれか1項に記載のシステム。

【請求項60】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体であって、前記命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、

30

(a) 被験体から得られた試験サンプル由来のセルフリー核酸(cfNA)フラグメントから得られた配列リードを含む試験配列情報を受信すること；および

(b) 標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルと実質的にマッチする少なくとも1つのアレルバリエーションが前記試験配列情報に存在することを識別することであって、その分類アレルは、少なくとも1つの選択されたカットオフ閾値未満のサブクロナリティスコアを含み、それによって、前記分類アレルが、標的細胞を起源とする参照cfNAフラグメントに由来することを示し、それによって、前記試験配列情報における前記アレルバリエーションが、前記被験体内の前記標的細胞を起源とすることを示すこと

40

を行う、コンピュータ可読媒体。

【請求項61】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体であって、前記命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、

(a) 前記被験体から得られた試験サンプル中のセルフリーデオキシリボ核酸(cfDNA)フラグメントから得られた配列リードを含む試験配列情報を受信すること；

(b) 前記被験体の造血幹細胞を起源とする1つまたはそれを超える前記配列リードまたは分類アレルを前記試験配列情報から除去して、フィルターされた試験配列情報を生成すること；および

(c) 1またはそれを超える参照被験体から得られた参照配列情報と実質的にアラインメ

50

ントする1つまたはそれを超える前記配列リードが前記フィルターされた試験配列情報に存在することを識別することであって、その参照配列情報は、前記参照被験体内の腫瘍細胞を起源とし、それによって、前記試験サンプルが、前記被験体内の前記腫瘍細胞を起源とする1つまたはそれを超えるc f DNAフラグメントを含むことを示すことを行う、コンピュータ可読媒体。

【請求項62】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体であって、前記命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、

(a) 1つまたはそれを超える参照サンプル由来のセルフリー核酸(c f NA)フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対するサブクロナリティスコアを生成することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、前記参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと；および

(b) 前記サブクロナリティスコアと少なくとも1つの選択されたカットオフ閾値を比較すること

を行い、

ここで、前記選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照c f NAフラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ／または

前記カットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照c f NAフラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる、コンピュータ可読媒体。

【請求項63】

コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体であって、前記命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、

(a) 1つまたはそれを超える参照サンプル由来のセルフリー核酸(c f NA)フラグメントから得られた配列リードを含む配列情報から分類アレルセットを識別することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、前記参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと；

(b) 前記配列情報から前記参照サンプルの各々における各分類アレルに対するマイナーアレル頻度(MAF)の値を決定すること；

(c) 前記参照サンプルの各々について、最大マイナーアレル頻度(maxMAF)の値を決定すること；

(d) 所与の参照サンプルにおいて観察された各分類アレルについて、前記参照サンプルの少なくとも一部に対する前記maxMAFの値に対する前記MAFの値の比を計算して、比の値を生成すること；

(e) 前記参照サンプルの少なくとも一部における所与の分類アレルが、前記参照サンプルの少なくとも一部に出現した前記所与の分類アレルの回数の合計数に対する、少なくとも1つの選択されたクロナリティ境界値未満の比の値を有した回数の比を、前記分類アレルの各々について計算して、前記参照サンプルの少なくとも一部における前記分類アレルの各々に対するサブクロナリティスコアを生成すること；および

(f) 前記サブクロナリティスコアと少なくとも1つの選択されたカットオフ閾値を比較すること

を行い、

ここで、前記選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照c f NAフラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ／または

10

20

30

40

50

前記カットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる、コンピュータ可読媒体。

【請求項 6 4】

コンピュータが実行可能な非一時的な命令が、前記少なくとも 1 つの電子プロセッサによって実行されたとき、前記参照サンプルから得られた前記配列情報から、潜在的に臨床的に有意な、標的ゲノム遺伝子座セット内の各遺伝子座における各体細胞性核酸バリエーションに対する M A F の値を決定することであって、ここで、前記標的ゲノム遺伝子座セットは、各参照サンプルにおいて同一であること、および前記参照サンプルの各々について m a x M A F の値を決定してアレル情報を生成することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体が含む、先行する請求項のいずれか 1 項に記載のシステムまたはコンピュータ可読媒体。

10

【請求項 6 5】

コンピュータが実行可能な非一時的な命令が、前記少なくとも 1 つの電子プロセッサによって実行されたとき、前記参照サンプルにインデックスされた臨床情報を用いて、前記非標的核酸バリエーションフィルターリストおよび／または前記標的核酸バリエーションフィルターリストを生成することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体が含む、先行する請求項のいずれか 1 項に記載のシステムまたはコンピュータ可読媒体。

20

【請求項 6 6】

コンピュータが実行可能な非一時的な命令が、前記少なくとも 1 つの電子プロセッサによって実行されたとき、前記試験サンプルにインデックスされた臨床情報を用いて、前記被験体内の前記標的細胞を起源とする c f N A フラグメントを検出することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体が含む、先行する請求項のいずれか 1 項に記載のシステムまたはコンピュータ可読媒体。

【請求項 6 7】

コンピュータが実行可能な非一時的な命令が、前記少なくとも 1 つの電子プロセッサによって実行されたとき、前記分類アレルの各々に対する各 M A F / m a x - M A F 値の頻度を用いてサブクロナリティスコアを決定することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体が含む、先行する請求項のいずれか 1 項に記載のシステムまたはコンピュータ可読媒体。

30

【請求項 6 8】

コンピュータが実行可能な非一時的な命令が、前記少なくとも 1 つの電子プロセッサによって実行されたとき、前記サブクロナリティスコアを複数の選択されたカットオフ閾値と比較することであって、ここで、前記複数の選択されたカットオフ閾値は、第 1 のカットオフ閾値および第 2 のカットオフ閾値を含み、その第 1 のカットオフ閾値は、前記第 2 のカットオフ閾値より大きく、前記第 1 のカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、前記非標的核酸バリエーションフィルターリストに加えられる、かつ／または前記第 2 のカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、前記標的核酸バリエーションフィルターリストに加えられることを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体が含む、先行する請求項のいずれか 1 項に記載のシステムまたはコンピュータ可読媒体。

40

【請求項 6 9】

コンピュータが実行可能な非一時的な命令が、前記少なくとも 1 つの電子プロセッサによって実行されたとき、非標的核酸バリエーションフィルターリスト上の少なくとも 1 つの分

50

類アレルと実質的にマッチする、前記試験配列情報におけるアレルバリエーションが、約1%を超えるMAFを含むとき、前記アレルバリエーションを、標的細胞を起源とすると分類することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体を含む、先行する請求項のいずれか1項に記載のシステムまたはコンピュータ可読媒体。

【請求項70】

コンピュータが実行可能な非一時的な命令が、前記少なくとも1つの電子プロセッサによって実行されたとき、非標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルと実質的にマッチする、前記試験配列情報におけるアレルバリエーションが、ランケーション、インデルおよび／またはスプライス部位バリエーションを含むとき、前記アレルバリエーションを、標的細胞を起源とすると分類することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体を含む、先行する請求項のいずれか1項に記載のシステムまたはコンピュータ可読媒体。

10

【請求項71】

コンピュータが実行可能な非一時的な命令が、前記少なくとも1つの電子プロセッサによって実行されたとき、前記参照サンプルの少なくとも一部における所与の分類アレルに対する各比の値の頻度を決定することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体を含む、先行する請求項のいずれか1項に記載のシステムまたはコンピュータ可読媒体。

【請求項72】

コンピュータが実行可能な非一時的な命令が、前記少なくとも1つの電子プロセッサによって実行されたとき、前記標的核酸バリエーションフィルターリストを用いて、被験体から得られた試験サンプルが、前記標的細胞を起源とするcfNAフラグメントを含むか否かを決定することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体を含む、先行する請求項のいずれか1項に記載のシステムまたはコンピュータ可読媒体。

20

【請求項73】

コンピュータが実行可能な非一時的な命令が、前記少なくとも1つの電子プロセッサによって実行されたとき、前記非標的核酸バリエーションフィルターリストを用いて、被験体から得られた試験サンプルが、前記非標的細胞を起源とするcfNAフラグメントを含むか否かを決定することを少なくともさらに行う、前記コンピュータが実行可能な非一時的な命令を、前記コンピュータ可読媒体を含む、先行する請求項のいずれか1項に記載のシステムまたはコンピュータ可読媒体。

30

【発明の詳細な説明】

【背景技術】

【0001】

背景

ポリヌクレオチドの検出および定量は、分子生物学、および診断法などの医学的応用にとって重要である。遺伝子検査は、いくつかの診断方法にとって特に有用である。例えば、稀な遺伝子変化（genetic alterations）（例えば、配列バリエーション）またはエピジェネティックマーカーの変化（changes）（例えば、癌および部分異数性または完全異数性）が原因の障害は、DNA配列情報によって検出され得るか、またはDNA配列情報によってより正確に特徴付けられ得る。

40

【0002】

癌などの遺伝性疾患の早期の検出およびモニタリングは、処置の成功または疾患の管理に必要であることが多い。アプローチの1つとしては、種々のタイプの体液中に見られ得るポリヌクレオチド集団であるセルフリー核酸に由来するサンプルをモニタリングすることが挙げられ得る。いくつかの場合において、疾患は、遺伝子異常（例えば、1つまたはそれを超える核酸配列のコピー数バリエーションおよび／または配列バリエーション）の検出または他の遺伝子変化の発生の検出に基づいて特徴付けられ得るか、または検出され

50

得る。セルフリーDNA (cfDNA) は、特定の疾患に関連する遺伝子異常を含み得る。

しかしながら、血液中に存在するcfDNAは、いくつかの細胞起源、癌性細胞と非癌性細胞の両方を起源とし得る。問題になり得るセルフリーDNAの起源の1つは、造血幹細胞であり、この細胞における変異は、血液細胞のクローン集団を拡大させ得る。他の血液悪性腫瘍の徴候無くクローン性増殖を推進するような体細胞変異の獲得は、「未確定の潜在能をもつクローン性造血 (CHIP)」に由来する細胞と称される。Steenmaら、Blood, 126: 9-16 (2015) を参照のこと。70歳を超える高齢者集団の少なくとも10%が、変異した造血幹細胞のオリゴクローン性増殖に起因するCHIPを有する。Jaiswalら、N. Engl. J. Med., 371 (26): 2488-2498 (2014) を参照のこと。造血幹細胞は、その造血幹細胞が非癌性であったとしても、癌に関連するゲノム領域に遺伝的バリエーションを含み得る。したがって、造血幹細胞において優勢に存在するが癌細胞には存在しない、サンプリングされたcfDNA集団に寄与するアレルを識別することは興味深い。

【先行技術文献】

【非特許文献】

【0003】

【非特許文献1】Steenmaら、Blood, 126: 9-16 (2015)

【非特許文献2】Jaiswalら、N. Engl. J. Med., 371 (26): 2488-2498 (2014)

【発明の概要】

【課題を解決するための手段】

【0004】

要旨

本開示は、液体生検サンプルなどのセルフリー核酸 (cfNA) サンプル由来のcfNAフラグメントの細胞起源を決定することにおいて有用な方法、コンピュータ可読媒体およびシステムを提供する。これらの態様は、ある特定の実施形態では、造血幹細胞などの非標的細胞によって生成されたバリエーションアレルを識別することによって、通常、cfNAサンプル中の罹患細胞核酸 (例えば、癌細胞DNA) を検出するためのアッセイの特異度および/または感度を改善する。さらに、本明細書中に開示される方法は、初期癌の腫瘍由来核酸の場合など、cfNAサンプル中に非常に少量で存在することが多い核酸の細胞起源の識別を容易にする。したがって、本明細書中に開示される方法および関連する態様は、他の数多くの用途のなかでも、疾患の早期の検出を促す。

【0005】

1つの態様において、本開示は、少なくとも部分的にコンピュータを使用して、被験体内の標的細胞を起源とする核酸分子を検出する方法を提供する。この方法は、(a) 被験体から得られた試験サンプル由来のセルフリー核酸 (cfNA) フラグメントから得られた配列リードを含む試験配列情報をコンピュータによって受信する工程を含む。この方法は、(b) 標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルと実質的にマッチする少なくとも1つのアレルバリエーションが試験配列情報に存在することを識別する工程も含む。その分類アレルは、少なくとも1つの選択されたカットオフ閾値未満のサブクローナリティスコアを含み、それによって、その分類アレルが、標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それによって、その被験体内の標的細胞を起源とする核酸分子を検出する。いくつかの実施形態において、例えば、(b) は、試験配列情報における少なくとも1つのアレルバリエーションを識別すること；そのアレルバリエーションを標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルにマッピングすること；その分類アレルのサブクローナリティスコアを識別すること；およびそのサブクローナリティスコアを少なくとも1つの選択されたカットオフ閾値と比較することを含み、ここで、そのサブクローナリティスコアが、選択されたカットオフ閾値未満であるとき、それは、その分類アレルが、標的細胞を起源とする参照cfNAフラグメ

ント由来であることを示す。

【0006】

1つの態様において、本開示は、少なくとも部分的にコンピュータを使用して、被験体内の腫瘍細胞を起源とする核酸分子を検出する方法を提供する。その方法は、(a)被験体から得られた試験サンプル中のセルフリーデオキシリボ核酸(c f DNA)フラグメントから得られた配列リードを含む試験配列情報をコンピュータによって受信する工程を含む。その方法は、(b)コンピュータによって、被験体の造血幹細胞を起源とする1つまたはそれを超える上記配列リード(例えば、分類アレルの少なくとも一部を含む)を試験配列情報から除去(例えば、削除、非表示、無視など)して、フィルターされた試験配列情報を生成する工程も含む。さらに、その方法は、(c)1またはそれを超える参照被験体から得られた参照配列情報と実質的にアラインメントする上記フィルターされた試験配列情報に1つまたはそれを超える上記配列リードが存在することをコンピュータによって識別する工程であって、その参照配列情報は、参照被験体内の1つまたはそれを超える腫瘍細胞を起源とし、それによって、その被験体内の腫瘍細胞を起源とする核酸分子を検出する、工程も含む。

10

【0007】

1つの態様において、本開示は、被験体における疾患を処置する方法を提供する。その方法は、(a)被験体から得られた試験サンプル由来のセルフリー核酸(c f NA)フラグメントから得られた配列リードを含む試験配列情報を受信する工程を含む。その方法は、(b)標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルと実質的にマッチする少なくとも1つのアレルバリエーションが試験配列情報に存在することを識別する工程も含む。その分類アレルは、少なくとも1つの選択されたカットオフ閾値未満のサブクロナリティスコアを含み、それによって、その分類アレルが、罹患細胞を起源とする参照c f NAフラグメント由来であることを示し、それによって、その被験体の疾患が診断される。さらに、その方法は、(c)被験体に1つまたはそれを超える療法を施し、それによって、その被験体の疾患を処置する工程も含む。

20

【0008】

別の態様において、本開示は、少なくとも部分的にコンピュータを使用して、分類器またはその少なくとも一部を生成する方法を提供する。その方法は、(a)1つまたはそれを超える参照サンプル由来のセルフリー核酸(c f NA)フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対するサブクロナリティスコアを、コンピュータによって生成する工程を含み、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む。その方法は、(b)上記サブクロナリティスコアと少なくとも1つの選択されたカットオフ閾値をコンピュータによって比較する工程であって、ここで、その選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照c f NAフラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ/またはそのカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照c f NAフラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられ、それによって、分類器を生成する工程も含む。

30

40

【0009】

別の態様において、本開示は、少なくとも部分的にコンピュータを使用して、分類器またはその少なくとも一部を生成する方法を提供する。その方法は、(a)1つまたはそれを超える参照サンプル由来のセルフリー核酸(c f NA)フラグメントから得られた配列リードを含む配列情報から分類アレルセットをコンピュータによって識別する工程を含み、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む。その方法は、(b)上記配列情報から、参照サンプルの各々における各分類アレルに対するマイナーアレル頻度(MAF)の値をコンピ

50

ュータによって決定する工程、および（c）参照サンプルの各々に対して最大マイナーアレレル頻度（maxMAF）の値をコンピュータによって決定する工程も含む。その方法は、（d）所与の参照サンプルにおいて観察された各分類アレレルについて、参照サンプルの少なくとも一部に対するmaxMAFの値に対する上記MAFの値の比をコンピュータによって計算して、比の値を生成する工程も含む。その方法は、（e）参照サンプルの少なくとも一部における所与の分類アレレルが、参照サンプルの少なくとも一部に出現したその所与の分類アレレルの回数の合計数に対する、少なくとも1つの選択されたクローナリティ境界値未満の比の値を有した回数の比を、前記分類アレレルの各々について、コンピュータによって計算して、参照サンプルの少なくとも一部における前記分類アレレルの各々に対するサブクローナリティスコアを生成する工程も含む。さらに、その方法は、（f）上記サブクローナリティスコアと少なくとも1つの選択されたカットオフ閾値をコンピュータによって比較する工程であって、ここで、その選択されたカットオフ閾値を超えるサブクローナリティスコアを有する分類アレレルは、それらの分類アレレルが、非標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレレルは、非標的核酸バリエーションフィルタリストに加えられ、かつ／またはそのカットオフ閾値未満のサブクローナリティスコアを有する分類アレレルは、それらの分類アレレルが、標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレレルは、標的核酸バリエーションフィルタリストに加えられ、それによって、分類器を生成する工程も含む。

10

【0010】

別の態様において、本開示は、被験体から得られた試験サンプル中のセルフリー核酸（cfNA）フラグメントの細胞起源を分類する際に使用するサブクローナリティスコアのデータベースを作成する方法を提供する。その方法は、（a）1つまたはそれを超える参照サンプル由来のセルフリー核酸（cfNA）フラグメントから得られた配列リードを含む配列情報から分類アレレルセットをコンピュータによって識別する工程を含み、ここで、各分類アレレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレレルを含む。その方法は、（b）上記配列情報から、参照サンプルの各々における各分類アレレルに対するマイナーアレレル頻度（MAF）の値をコンピュータによって決定する工程、（c）参照サンプルの各々について、最大マイナーアレレル頻度（maxMAF）の値をコンピュータによって決定する工程、および（d）所与の参照サンプルにおいて観察された各分類アレレルについて、参照サンプルの少なくとも一部に対するmaxMAFの値に対する上記MAFの値の比をコンピュータによって計算して、比の値を生成する工程も含む。その方法は、（e）参照サンプルの少なくとも一部における所与の分類アレレルが、参照サンプルの少なくとも一部に出現したその所与の分類アレレルの回数の合計数に対する、少なくとも1つの選択されたクローナリティ境界値未満の比の値を有した回数の比を、上記分類アレレルの各々について、コンピュータによって計算して、参照サンプルの少なくとも一部における上記分類アレレルの各々に対するサブクローナリティスコアを生成する工程も含む。さらに、その方法は、（f）対応する分類アレレルにインデックスされたサブクローナリティスコアをデータベースシステムに非一時的に保存し、それによって、被験体から得られた試験サンプル中のcfNAフラグメントの細胞起源を分類する際に使用するサブクローナリティスコアのデータベースを作成する工程も含む。

20

30

40

【0011】

いくつかの実施形態において、本明細書中に開示される方法は、上記分類アレレルセットを識別する工程を含み、その工程は、参照サンプルから得られた配列情報から、潜在的に臨床的に有意な標的ゲノム遺伝子座セット内の各遺伝子座における各体細胞性核酸バリエーションに対するMAFの値を決定する工程であって、ここで、その標的ゲノム遺伝子座セットは、各参照サンプルにおいて同一である工程、および参照サンプルの各々についてmaxMAFの値を決定してアレレル情報を生成する工程を含む。ある特定の実施形態において、各分類アレレルに対するMAFは、約2%未満である。いくつかの実施形態において、各分類アレレルに対するMAFは、約1%未満である。

【0012】

50

ある特定の実施形態において、本明細書中に開示される方法は、参照サンプルにインデックスされた臨床情報を用いて、分類器を生成する工程を含む。いくつかの実施形態において、本明細書中に開示される方法は、試験サンプルにインデックスされた臨床情報を用いて、被験体内の標的細胞を起源とする核酸分子を検出する工程を含む。ある特定の実施形態において、その臨床情報は、年齢、性別、人種、体重、ボディマス指数（BMI）、病歴、喫煙、飲酒などからなる群より選択される。他の例示的な実施形態では、サンプルの種々のサブセットに対して、例えば最小maxMAFに基づいて、サブクロナルリスト（例えば、標的核酸バリエーションフィルターリストまたは非標的核酸バリエーションフィルターリスト）が生成され、公知のドライバー変異などに基づいて、maxMAFがコールされる。いくつかの実施形態では、特定の指摘（例えば、所与の癌のタイプ（例えば、肺、結腸直腸など））に基づいて、サブクロナルリストが生成される。ある特定の実施形態において、機械学習分類器は、変異アレル頻度、サブクローン比、遺伝子のタイプ、血液悪性腫瘍に関連するバリエーション、患者の年齢、他のCHIPバリエーションの観察結果、癌のタイプなどを含む1つまたはそれを超える特徴量に基づいて訓練される。

10

【0013】

いくつかの実施形態において、本明細書中に開示される方法は、上記分類アレルの各々に対する各MAF/maxMAF値の頻度を用いて、サブクロナリティスコアを決定する工程を含む。ある特定の実施形態において、選択されたクロナリティ境界値は、約1%～約99%の範囲内である。これらの実施形態のうちのいくつかにおいて、例えば、選択されたクロナリティ境界値は、約10%、約20%、約30%、約40%、約50%、約60%、約70%、約80%または約90%である。いくつかの実施形態において、選択されたカットオフ閾値は、約1%～約99%の範囲内である。これらの実施形態のうちのいくつかにおいて、例えば、選択されたカットオフ閾値は、約10%、約20%、約30%、約40%、約50%、約60%、約70%、約80%または約90%である。

20

【0014】

ある特定の実施形態において、本明細書中に開示される方法は、上記サブクロナリティスコアを複数の選択されたカットオフ閾値と比較する工程を含む。これらの実施形態のうちのいくつかにおいて、例えば、複数の選択されたカットオフ閾値は、第1のカットオフ閾値および第2のカットオフ閾値を含み、その第1のカットオフ閾値は、第2のカットオフ閾値より大きく、第1のカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ/または第2のカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる。

30

【0015】

いくつかの実施形態において、本明細書中に開示される方法は、非標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルと実質的にマッチする、試験配列情報におけるアレルバリエーションが、約1%を超えるMAFを含むとき、そのアレルバリエーションを、標的細胞を起源とすると分類する工程を含む。ある特定の実施形態において、本明細書中に開示される方法は、非標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルと実質的にマッチする、試験配列情報におけるアレルバリエーションが、ランケーション、インデルおよび/またはスプライス部位バリエーションを含むとき、そのアレルバリエーションを、標的細胞を起源とすると分類する工程を含む。

40

【0016】

ある特定の実施形態において、本明細書中に開示される方法は、参照サンプルの少なくとも一部における所与の分類アレルに対する各比の値の頻度を決定する工程を含む。いくつかの実施形態において、本明細書中に開示される方法は、上記分類器を用いて、被験体から得られた試験サンプルが、前記標的細胞を起源とするcfNAフラグメントを含むか

50

否かを決定する工程を含む。ある特定の実施形態において、本明細書中に開示される方法は、上記分類器を用いて、被験体から得られた試験サンプルが、非標的細胞を起源とする c f N A フラグメントを含むか否かを決定する工程を含む。いくつかの実施形態において、データベースは、標的核酸バリエーションフィルタリストおよび／または非標的核酸バリエーションフィルタリストを含む。

【0017】

ある特定の実施形態において、非標的細胞は、非罹患細胞を含む。いくつかの実施形態において、非標的細胞は、造血幹細胞を含む。ある特定の実施形態において、非標的細胞は、非腫瘍細胞を含む。いくつかの実施形態において、非標的細胞は、母体細胞を含む。ある特定の実施形態において、非標的細胞は、移植レシピエント細胞を含む。

10

【0018】

ある特定の実施形態において、標的細胞は、罹患細胞を含む。いくつかの実施形態において、標的細胞は、腫瘍細胞を含む。いくつかの実施形態において、標的細胞は、胎児細胞を含む。ある特定の実施形態において、標的細胞は、移植ドナー細胞を含む。

【0019】

ある特定の実施形態において、本明細書中に開示される方法は、疾患を処置する工程を含む。これらの実施形態のうちのいくつかにおいて、例えば、その疾患は、癌を含み、その療法は、少なくとも1つの免疫療法を含む。通常、被験体は、哺乳動物被験体（例えば、ヒト被験体）である。

【0020】

いくつかの実施形態において、本明細書中に開示される方法は、被験体から試験サンプルを得る工程をさらに含む。その試験サンプルは、通常、血液、血漿、血清、痰、尿、精液、腔液、便、滑液、髄液、唾液などからなる群より選択される。いくつかの実施形態において、本明細書中に開示される方法は、上記試験サンプル中の c f N A フラグメントから試験配列情報を生成する工程をさらに含む。いくつかの実施形態において、本明細書中に開示される方法は、標的ゲノム遺伝子座を含む c f N A フラグメントのセグメントを増幅して、増幅核酸を生成する工程をさらに含む。ある特定の実施形態において、本明細書中に開示される方法は、上記試験サンプル中の c f N A フラグメントをシーケンシングして、試験配列情報を生成する工程をさらに含む。これらの実施形態のうちのいくつかにおいて、その試験配列情報は、試験サンプル中の c f N A フラグメントの標的化セグメントから得られ、その標的化セグメントは、シーケンシングの前に、試験サンプル中の c f N A フラグメントから1つまたはそれを超える領域を選択的に濃縮することによって得られる。ある特定の実施形態において、本明細書中に開示される方法は、シーケンシングの前に、得られた標的化セグメントを増幅する工程をさらに含む。ある特定の実施形態において、本明細書中に開示される方法は、シーケンシングの前に、バーコードを含む1つまたはそれを超えるアダプターを c f N A フラグメントおよび／または増幅された標的化セグメントに付着させる工程をさらに含む。ある特定の実施形態において、そのシーケンシングは、ターゲットシーケンシング、バイサルファイトシーケンシング、イントロンシーケンシング、エクソームシーケンシングおよびホールゲノムシーケンシングからなる群より選択される。

30

40

【0021】

なおも別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたはそのコンピュータ可読媒体にアクセスできるコントローラを備えるシステムを提供し、その命令は、少なくとも1つの電子プロセッサによって実行されたとき、少なくとも、（a）被験体から得られた試験サンプル由来のセルフフリー核酸（c f N A）フラグメントから得られた配列リードを含む試験配列情報を受信すること、および（b）標的核酸バリエーションフィルタリスト上の少なくとも1つの分類アレルと実質的にマッチする少なくとも1つのアレルバリエーションがその試験配列情報に存在することを識別することであって、その分類アレルは、少なくとも1つの選択されたカットオフ閾値未満のサブクロナリティスコアを含み、それによって、その分類アレルが

50

、標的細胞を起源とする参照 c f N A フラグメントに由来することを示し、それによって、その試験配列情報におけるアレルバリエーションが、被験体内の標的細胞を起源とすることを示すことを行う。いくつかの実施形態において、例えば、(b) は、上記試験配列情報における少なくとも 1 つのアレルバリエーションを識別すること；そのアレルバリエーションを標的核酸バリエーションフィルターリスト上の少なくとも 1 つの分類アレルにマッピングすること；その分類アレルのサブクロナリティスコアを識別すること；およびそのサブクロナリティスコアを少なくとも 1 つの選択されたカットオフ閾値と比較することを含み、ここで、そのサブクロナリティスコアが、選択されたカットオフ閾値未満であるとき、それは、その分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示唆する。

10

【0022】

なおも別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたはそのコンピュータ可読媒体にアクセスできるコントローラを備えるシステムを提供し、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、少なくとも、(a) 被験体から得られた試験サンプル中のセルフリーデオキシリボ核酸 (c f D N A) フラグメントから得られた配列リードを含む試験配列情報を受信すること、(b) 被験体の造血幹細胞を起源とする 1 つまたはそれを超える上記配列リード (例えば、分類アレルの少なくとも一部を含む) をその試験配列情報から除去 (例えば、削除、非表示、無視など) して、フィルターされた試験配列情報を生成すること、および (c) 1 つまたはそれを超える参照被験体から得られた参照配列情報と実質的にアラインメントする上記 1 つまたはそれを超える配列リードが、フィルターされた試験配列情報に存在することを識別することであって、その参照配列情報は、参照被験体内の腫瘍細胞を起源とし、それによって、その試験サンプルが、被験体内の腫瘍細胞を起源とする 1 つまたはそれを超える c f D N A フラグメントを含むことを示すことを行う。

20

【0023】

なおも別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたはそのコンピュータ可読媒体にアクセスできるコントローラを備えるシステムを提供し、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、少なくとも、(a) 1 つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対するサブクロナリティスコアを生成することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと、および (b) 上記サブクロナリティスコアと少なくとも 1 つの選択されたカットオフ閾値を比較することを行い、ここで、その選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ/またはそのカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる。

30

40

【0024】

なおも別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を備えるかまたはそのコンピュータ可読媒体にアクセスできるコントローラを備えるシステムを提供し、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、少なくとも、(a) 1 つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む配列情報から分類アレルセットを識別することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと、(b) 上記配列情報から参照サンプルの各々における各分類アレルに対するマイナーアレル頻度 (M A F) の値を決定すること、(c) 参照サンプルの各々について、最大マイナ

50

ーアレル頻度 (maxMAF) の値を決定すること、(d) 所与の参照サンプルにおいて観察された各分類アレルについて、参照サンプルの少なくとも一部に対する maxMAF の値に対する上記 MAF の値の比を計算して、比の値を生成すること、(e) 参照サンプルの少なくとも一部における所与の分類アレルが、参照サンプルの少なくとも一部に出現したその所与の分類アレルの回数の合計数に対する、少なくとも 1 つの選択されたクローナリティ境界値未満の比の値を有した回数の比を、上記分類アレルの各々について計算して、参照サンプルの少なくとも一部における上記分類アレルの各々に対するサブクローナリティスコアを生成すること、および (f) 上記サブクローナリティスコアと少なくとも 1 つの選択されたカットオフ閾値を比較することを行い、ここで、その選択されたカットオフ閾値を超えるサブクローナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 cfNA フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルタリストに加えられ、かつ／またはそのカットオフ閾値未満のサブクローナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 cfNA フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルタリストに加えられる。

10

【0025】

いくつかの実施形態において、本明細書中に開示されるシステムは、コントローラと作動可能に接続された核酸配列決定装置を備え、その核酸配列決定装置は、試験サンプル中および／または参照サンプル中の cfNA フラグメントから配列情報を提供するように構成されている。これらの実施形態のうちのある特定の実施形態において、その核酸配列決定装置は、上記核酸に対してパイロシーケンシング、バイサルファイトシーケンシング、一分子シーケンシング、ナノポアシーケンシング、半導体シーケンシング、ライゲーションによるシーケンシングまたはハイブリダイゼーションによるシーケンシングを行って、シーケンシングリードを生成するように構成されている。

20

【0026】

いくつかの実施形態において、本明細書中に開示されるシステムは、コントローラに作動可能に接続されたサンプル調製コンポーネントを備え、そのサンプル調製コンポーネントは、核酸配列決定装置によってシーケンシングされる cfNA フラグメントを調製するように構成されている。これらの実施形態のうちの一つにおいて、そのサンプル調製コンポーネントは、試験サンプル中および／または参照サンプル中の cfNA フラグメントからの領域を選択的に濃縮するように構成されている。ある特定の実施形態において、サンプル調製コンポーネントは、バーコードを含む 1 つまたは複数のアダプターを cfNA フラグメントに付着させるように構成されている。

30

【0027】

ある特定の実施形態において、本明細書中に開示されるシステムは、コントローラに作動可能に接続された核酸増幅コンポーネントを備え、その核酸増幅コンポーネントは、試験サンプル中および／または参照サンプル中の cfNA フラグメントを増幅するように構成されている。これらの実施形態のうちの一つにおいて、その核酸増幅コンポーネントは、試験サンプル中および／または参照サンプル中の cfNA フラグメントから、選択的に濃縮された領域を増幅するように構成されている。いくつかの実施形態において、本明細書中に開示されるシステムは、コントローラに作動可能に接続された材料移動コンポーネントを備え、その材料移動コンポーネントは、1 つまたはそれを超える材料を、核酸配列決定装置、核酸増幅コンポーネントおよび／またはサンプル調製コンポーネントの間を移動させるように構成されている。ある特定の実施形態において、本明細書中に開示されるシステムは、コントローラに作動可能に接続されたデータベースを備え、そのデータベースは、非標的核酸バリエーションフィルタリストおよび／または標的核酸バリエーションフィルタリストを含む。

40

【0028】

なおも別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を提供し、その命令は、少なくとも 1 つの電子プロセッサによ

50

て実行されたとき、少なくとも、(a) 被験体から得られた試験サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む試験配列情報を受信すること、および (b) 標的核酸バリエーションフィルターリスト上の少なくとも 1 つの分類アレルと実質的にマッチする少なくとも 1 つのアレルバリエーションが試験配列情報に存在することを識別することを行い、その分類アレルは、少なくとも 1 つの選択されたカットオフ閾値未満のサブクロナリティスコアを含み、それによって、その分類アレルが、標的細胞を起源とする参照 c f N A フラグメントに由来することを示し、それによって、その試験配列情報におけるアレルバリエーションが、被験体内の標的細胞を起源とすることを示す。いくつかの実施形態において、例えば、(b) は、試験配列情報における少なくとも 1 つのアレルバリエーションを識別すること；そのアレルバリエーションを標的核酸バリエーションフィルターリスト上の少なくとも 1 つの分類アレルにマッピングすること；その分類アレルのサブクロナリティスコアを識別すること；およびそのサブクロナリティスコアを少なくとも 1 つの選択されたカットオフ閾値と比較することを含み、ここで、そのサブクロナリティスコアが、選択されたカットオフ閾値未満であるとき、それは、その分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示す。

10

【0029】

別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を提供し、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、少なくとも、(a) 被験体から得られた試験サンプル中のセルフリーデオキシリボ核酸 (c f D N A) フラグメントから得られた配列リードを含む試験配列情報を受信すること、(b) 被験体の造血幹細胞を起源とする 1 つまたはそれを超える上記配列リード (例えば、分類アレルの少なくとも一部を含む) をその試験配列情報から除去 (例えば、削除、非表示、無視など) して、フィルターされた試験配列情報を生成すること、および (c) 1 つまたはそれを超える参照被験体から得られた参照配列情報と実質的にアラインメントする上記 1 つまたはそれを超える配列リードが、フィルターされた試験配列情報に存在することを識別することであって、その参照配列情報は、参照被験体内の腫瘍細胞を起源とし、それによって、その試験サンプルが、被験体内の腫瘍細胞を起源とする 1 つまたはそれを超える c f D N A フラグメントを含むことを示すことを行う。

20

【0030】

別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を提供し、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、少なくとも、(a) 1 つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対するサブクロナリティスコアを生成することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと、および (b) 上記サブクロナリティスコアと少なくとも 1 つの選択されたカットオフ閾値を比較することを行い、ここで、その選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられる、かつ／またはそのカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 c f N A フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる。

30

40

【0031】

別の態様において、本開示は、コンピュータが実行可能な非一時的な命令を含むコンピュータ可読媒体を提供し、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、少なくとも、(a) 1 つまたはそれを超える参照サンプル由来のセルフリー核酸 (c f N A) フラグメントから得られた配列リードを含む配列情報から分類アレルセットを識別することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含むこと、(b) 上記

50

配列情報から参照サンプルの各々における各分類アレルに対するマイナーアレル頻度 (MAF) の値を決定すること、(c) 参照サンプルの各々について、最大マイナーアレル頻度 (maxMAF) の値を決定すること、(d) 所与の参照サンプルにおいて観察された各分類アレルについて、参照サンプルの少なくとも一部に対する maxMAF の値に対する上記 MAF の値の比を計算して、比の値を生成すること、(e) 参照サンプルの少なくとも一部における所与の分類アレルが、参照サンプルの少なくとも一部に出現したその所与の分類アレルの回数の合計数に対する、少なくとも 1 つの選択されたクロナリティ境界値未満の比の値を有した回数の比を、上記分類アレルの各々について計算して、参照サンプルの少なくとも一部における上記分類アレルの各々に対するサブクロナリティスコアを生成すること、および (f) 上記サブクロナリティスコアと少なくとも 1 つの選択されたカットオフ閾値を比較することを行い、ここで、その選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 cfNA フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ／またはそのカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 cfNA フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられる。

【0032】

本明細書中に開示されるシステムまたはコンピュータ可読媒体のいくつかの実施形態において、そのコンピュータ可読媒体は、コンピュータが実行可能な非一時的な命令を含み、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、参照サンプルから得られた配列情報から、潜在的に臨床的に有意な、標的ゲノム遺伝子座セット内の各遺伝子座における各体細胞性核酸バリエーションに対する MAF の値を決定することであって、ここで、その標的ゲノム遺伝子座セットは、各参照サンプルにおいて同一であること、および参照サンプルの各々について maxMAF の値を決定してアレル情報を生成することを少なくともさらに行う。

【0033】

本明細書中に開示されるシステムまたはコンピュータ可読媒体のある特定の実施形態において、そのコンピュータ可読媒体は、コンピュータが実行可能な非一時的な命令を含み、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、参照サンプルにインデックスされた臨床情報を用いて、非標的核酸バリエーションフィルターリストおよび／または標的核酸バリエーションフィルターリストを生成することを少なくともさらに行う。本明細書中に開示されるシステムまたはコンピュータ可読媒体のある特定の実施形態において、そのコンピュータ可読媒体は、コンピュータが実行可能な非一時的な命令を含み、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、試験サンプルにインデックスされた臨床情報を用いて、被験体内の標的細胞を起源とする cfNA フラグメントを検出することを少なくともさらに行う。本明細書中に開示されるシステムまたはコンピュータ可読媒体のある特定の実施形態において、そのコンピュータ可読媒体は、コンピュータが実行可能な非一時的な命令を含み、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、分類アレルの各々に対する各 MAF / maxMAF 値の頻度を用いて、サブクロナリティスコアを決定することを少なくともさらに行う。

【0034】

本明細書中に開示されるシステムまたはコンピュータ可読媒体のある特定の実施形態において、そのコンピュータ可読媒体は、コンピュータが実行可能な非一時的な命令を含み、その命令は、少なくとも 1 つの電子プロセッサによって実行されたとき、サブクロナリティスコアを複数の選択されたカットオフ閾値と比較することであって、ここで、その複数の選択されたカットオフ閾値は、第 1 のカットオフ閾値および第 2 のカットオフ閾値を含み、第 1 のカットオフ閾値は、第 2 のカットオフ閾値より大きく、第 1 のカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが

10

20

30

40

50

50

50

【図 1】図 1 A および 1 B は、50 % の閾値に設定されたクローナリティ境界値に基づいて推定されたサブクローナリティスコアによって識別される、2 つのアレル（図 1 A は、分類アレル 1 を示しており、図 1 B は、分類アレル 2 を示している）のヒストグラムである。本発明のいくつかの実施形態によると、アレル 1 は、陰性である（すなわち、癌細胞

起源を示さない（おそらく造血幹細胞））のに対して、アレル 2 は、参照サンプルデータベースにおける被験体の 50 % 超に存在するので、陽性である（すなわち、癌細胞起源を示す）。図 1 A および 1 B の各々において、Y 軸は、記録の数を示し、X 軸は、MAF / max MAF 比の分布を示す。

【0038】

【図 2】図 2 は、本発明のいくつかの実施形態に係る、被験体内の標的細胞を起源とする核酸分子を検出する例示的な方法工程を模式的に表しているフローチャートである。

【0039】

【図 3】図 3 は、本発明のいくつかの実施形態に係る、被験体内の腫瘍細胞を起源とする核酸分子を検出する例示的な方法工程を模式的に表しているフローチャートである。

10

【0040】

【図 4】図 4 は、本発明のいくつかの実施形態に係る、被験体における疾患を処置する例示的な方法工程を模式的に表しているフローチャートである。

【0041】

【図 5】図 5 は、本発明のいくつかの実施形態に係る、分類器を生成する例示的な方法工程を模式的に表しているフローチャートである。

【0042】

【図 6】図 6 は、本発明のいくつかの実施形態に係る、分類器を生成する例示的な方法工程を模式的に表しているフローチャートである。

【0043】

20

【図 7】図 7 は、本発明のある特定の実施形態との使用に適した例示的なシステムの模式図である。

【0044】

【図 8】図 8 A ~ C は、フィルターを用いなかったとき（図 8 A）、組織フィルタリングを用いたとき（図 8 B）、および分類器フィルタリングを用いたとき（図 8 C；すなわち、サブクロナリティスコアの使用）の、患者データに対するカプラン・マイヤープロットを示している。図 8 A ~ C に示されている各プロットにおいて、検出されなかった曲線は、上の曲線であり、検出された曲線は、下の曲線である。

【0045】

【図 9】図 9 は、図 8 A ~ C に示された異なるフィルターシナリオに対して観察されたバリエーションの数（y 軸）に対するアレル頻度の範囲（x 軸）のプロットを示している。

30

【発明を実施するための形態】

【0046】

定義

本開示をより容易に理解するために、まず、ある特定の用語を下記に定義する。以下の用語および他の用語に対するさらなる定義は、本明細書全体に示され得る。下記に示される用語の定義が、参照により援用される特許出願または発行された特許における定義と矛盾する場合、本願に示される定義が、その用語の意味を理解するために使用されるべきである。

【0047】

40

本明細書および添付の請求項において使用されるとき、単数形「a」、「an」および「the」は、文脈が明らかに他のことを指示しない限り、複数の対象を含む。したがって、例えば、「方法（a method）」という言葉及には、1 つまたはそれを超える方法、ならびに／または本明細書中に記載されるタイプの工程および／もしくは本開示を読めば当業者には明らかになるタイプの工程などが含まれる。また、本開示で議論される温度、濃度、時間、塩基または塩基対の数、カバレッジ（coverage）などの前には暗黙の「約」が存在し、わずかで実体のない等価物も本開示の範囲内であることが認識されるだろう。本願において、単数形の使用は、具体的に別段述べられていない限り、複数形を含む。また、「含む（comprise）」、「含む（comprises）」、「含む（comprising）」、「含む（contain）」、「含む（contai

50

ns)」、「含む (containing)」、「含む (include)」、「含む (includes)」および「含む (including)」の使用は、限定を意図していない。

【0048】

本明細書中で使用される術語は、特定の実施形態を説明する目的のみであって、限定を意図していないことも理解されるべきである。さらに、別段定義されない限り、本明細書中で使用されるすべての専門用語および科学用語は、本開示が属する分野の当業者が通常理解する意味と同じ意味を有する。方法、コンピュータ可読媒体およびシステムを説明し、特許請求する際、以下の術語およびその文法上の変化形は、下記に示される定義に従って使用される。

【0049】

約：本明細書中で使用されるとき、「約」または「およそ」は、1つまたはそれを超える目的の値またはエレメントに対して適用されるとき、記載の参照値または参照エレメントと類似の値またはエレメントのことを指す。ある特定の実施形態において、用語「約」または「およそ」とは、別途記載されていないかまたは文脈から明らかでない限り、記載の参照値または参照エレメントの25%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%またはそれ未満の、どちらかの方向（それらを超えるまたは低い）の範囲に収まる値またはエレメントの範囲のことを指す（そのような数値が、考えられる値またはエレメントの100%を超える場合を除く）。

【0050】

アダプター：本明細書中で使用されるとき、「アダプター」とは、通常、少なくとも部分的に二本鎖であり、かつ所与のサンプル核酸分子の片端または両端に連結するように使用される、短い核酸（例えば、約500ヌクレオチド長未満、約100ヌクレオチド長未満または約50ヌクレオチド長未満）のことを指す。アダプターは、両端でアダプターに隣接する核酸分子の増幅を可能にする核酸プライマー結合部位を含み得、かつ／または様々な次世代シーケンシング (NGS) 用途などのシーケンシング用途のプライマー結合部位を含むシーケンシングプライマー結合部位を含み得る。アダプターは、捕捉プローブ（例えば、フローセル支持体などに付着させたオリゴヌクレオチド）に対する結合部位も含み得る。アダプターは、本明細書中に記載されるような核酸タグも含み得る。核酸タグは、通常、核酸タグが所与の核酸分子のアンプリコンおよびシーケンシングリードの中に含まれるように、増幅プライマー結合部位およびシーケンシングプライマー結合部位に対して配置される。ある核酸分子のそれぞれの末端に、同じまたは異なるアダプターが連結され得る。ある特定の実施形態において、核酸タグが異なることを除いては、同じアダプターが、核酸分子のそれぞれの末端に連結される。いくつかの実施形態において、アダプターは、Y形のアダプターであり、その1つの末端は、平滑末端化されているか、またはある核酸分子に結合するために本明細書中に記載されるようにテイル化されており、その核酸分子もまた、平滑末端化されているか、または1つもしくはそれを超える相補的なヌクレオチドでテイル化されている。なおも他の例示的な実施形態では、アダプターは、平滑末端、または解析される核酸分子に結合するためのテイル化末端を含むベル形のアダプターである。他の例示的なアダプターとしては、Tテイル化アダプターおよびCテイル化アダプターが挙げられる。

【0051】

投与する：本明細書中で使用されるとき、療法薬（例えば、免疫学的療法薬）を被験体に「投与する (administer)」または「投与する (administering)」は、組成物を被験体に与える、適用する、または接触させることを意味する。投与は、いくつかの経路（例えば、局所的、経口、皮下、筋肉内、腹腔内、静脈内、髄腔内および皮内を含む）のうちのいずれかによって達成され得る。

【0052】

アレル：本明細書中で使用されるとき、「アレル」または「アレルバリエント」とは、

10

20

30

40

50

規定されたゲノム位置または遺伝子座における特定の遺伝的バリエーションのことを指す。アレルバリエーションは、通常、そのアレルがヘテロ接合性であるかホモ接合性であるかに応じて、50% (0.5) または100%の頻度で示される。例えば、生殖細胞系列バリエーションは、遺伝的、通常、0.5または1の頻度を有する。しかしながら、体細胞性バリエーションは、後天的なバリエーションであり、通常、 <0.5 の頻度を有する。ある遺伝子座のメジャーアレルおよびマイナーアレルとは、その遺伝子座が、それぞれ参照配列のヌクレオチド、および参照配列とは異なるバリエーションヌクレオチドによって占有されている、その遺伝子座を保持している核酸のことを指す。ある遺伝子座における測定値は、サンプル中でアレルが観察される頻度であるアレル率 (allelic fraction) (AF) の形を取り得る。

10

【0053】

増幅する：本明細書中で使用されるとき、核酸の文脈における「増幅する」または「増幅」とは、通常、少量のポリヌクレオチド (例えば、単一ポリヌクレオチド分子) から開始する、複数コピーのポリヌクレオチドまたはそのポリヌクレオチドの一部の生成のことを指し、その増幅産物またはアンプリコンは、通常、検出可能である。ポリヌクレオチドの増幅は、種々の化学的および酵素的プロセスを含む。

【0054】

バーコード：本明細書中で使用されるとき、核酸の文脈における「バーコード」とは、分子識別子として役立ち得る配列を含む核酸分子のことを指す。例えば、個々の「バーコード」配列は、最終的なデータ解析の前に各リードを識別し、ソーティングできるように、通常、次世代シーケンシング (NGS) ライブラリーの調製中に各DNAフラグメントに付加される。

20

【0055】

癌のタイプ：本明細書中で使用されるとき、「癌」、「癌のタイプ」または「腫瘍タイプ」とは、例えば組織病理学によって定義される、癌のタイプまたはサブタイプのことを指す。癌のタイプは、任意の従来の基準によって、例えば、所与の組織における出現 (例えば、血液癌、中枢神経系 (CNS)、脳腫瘍、肺癌 (小細胞および非小細胞)、皮膚癌、鼻癌、咽喉癌、肝臓癌、骨癌、リンパ腫、膵癌、腸 (bowel) 癌、直腸癌、甲状腺癌、膀胱癌、腎臓癌、口腔癌、胃癌、乳癌、前立腺癌、卵巣癌、肺癌、腸 (intestinal) 癌、軟部組織癌、神経内分泌癌、胃食道癌、頭頸部癌、婦人科癌、結腸直腸癌、尿路上皮癌、固形の癌、不均一な癌、均一な癌)、原発源が不明であることなど、ならびに／または細胞系譜が同じであること (例えば、癌腫、肉腫、リンパ腫、胆管癌、白血病、中皮腫、メラノーマまたは神経膠芽腫)、および／もしくは癌マーカー (例えば、Her2、CA15-3、CA19-9、CA-125、CEA、AFP、PSA、HCG、ホルモンレセプターおよびNMP-22) を示す癌に基づいて定義され得る。癌は、ステージ (例えば、ステージ1、2、3または4) によって、および原発源であるか二次源であるかによっても分類され得る。

30

【0056】

セルフリー核酸：本明細書中で使用されるとき、「セルフリー核酸」とは、細胞内に含まれていない核酸もしくは細胞に結合していない核酸、またはいくつかの実施形態では、インタクトな細胞を除去した後のサンプル中に残存している核酸のことを指す。セルフリー核酸には、例えば、被験体由来の体液 (例えば、血液、血漿、血清、尿、脳脊髄液 (CSF) など) を起源とする被包されていないあらゆる核酸が含まれ得る。セルフリー核酸としては、DNA (cfDNA)、RNA (cfRNA) およびそれらのハイブリッドが挙げられ、それらには、ゲノムDNA、ミトコンドリアDNA、循環DNA、siRNA、miRNA、循環RNA (cRNA)、tRNA、rRNA、低分子RNA (snRNA)、Piwiと相互作用するRNA (piRNA)、長い非コードRNA (長いncRNA) および／またはこれらのいずれかのフラグメントが含まれる。セルフリー核酸は、二本鎖、一本鎖またはそれらのハイブリッドであり得る。セルフリー核酸は、分泌または細胞死のプロセス、例えば、細胞壊死、アポトーシスなどによって、体液中に放出され

40

50

得る。セルフリー核酸は、エフェロソーム (e f f e r o s o m e) またはエキソソームに見られ得る。いくつかのセルフリー核酸は、癌細胞から体液中に放出される (例えば、循環腫瘍 DNA (c t D N A))。他のセルフリー核酸は、健康な細胞から放出される。C t D N A は、腫瘍由来の被包されていない断片化された DNA であり得る。セルフリー核酸の別の例は、母体血流中を自由に循環している胎児 DNA であり、セルフリー胎児 DNA (c f f D N A) とも呼ばれる。セルフリー核酸は、1 つまたはそれを超えるエピジェネティック修飾を有し得、例えば、セルフリー核酸は、アセチル化、5-メチル化、ユビキチン化、リン酸化、SUMO 化、リボシル化、および／またはシトルリン化され得る。

【0057】

10

細胞起源：本明細書中で使用されるとき、セルフリー核酸の文脈における「細胞起源」は、所与のセルフリー核酸分子が由来するまたは起源とする (例えば、アポトーシスプロセス、ネクローシスプロセスなどを介して) 細胞型を意味する。ある特定の実施形態において、例えば、所与のセルフリー核酸分子は、腫瘍細胞 (例えば、癌性肺細胞など) または非腫瘍細胞もしくは正常細胞 (例えば、非癌性肺細胞、造血幹細胞など) を起源とし得る。

【0058】

分類アレル：本明細書中で使用されるとき、「分類アレル」とは、所与の核酸分子にそれが存在することが、その核酸分子の起源 (例えば、細胞起源) を識別する、アレルバリエーションのことを指す。ある特定の実施形態では、例えば、ある核酸分子に所与の分類アレルが存在することによって、その核酸分子が、特定の用途に応じて標的細胞 (例えば、罹患細胞、腫瘍細胞、胎児細胞、移植ドナー細胞など) または非標的細胞 (例えば、非罹患細胞、造血幹細胞、母体細胞、移植レシピエント細胞など) を起源とすると識別され得る。通常、所与の分類アレルを標的核酸バリエーションフィルターリストまたは非標的核酸バリエーションフィルターリストに割り当てるために使用され得るサブクローナリティスコアが、所与の用途において使用される選択されたカットオフ閾値未満のか、その閾値であるか、その閾値を超えるかに応じて、そのサブクローナリティスコアと分類アレルが関連付けられる。

20

【0059】

分類器：本明細書中で使用されるとき、「分類器」とは、一般に、試験データを入力として受け取り、その入力データをあるクラスまたは別のクラス (例えば、腫瘍 DNA または非腫瘍 DNA) に属すると分類することを出力として生成するアルゴリズム計算機コードのことを指す。

30

【0060】

臨床情報：本明細書中で使用されるとき、「臨床情報」とは、被験体に対する健康管理の決定を知らせ得る任意の情報のことを指す。臨床情報の例としては、とりわけ、ゲノム情報、年齢、性別、人種、体重、ボディマス指数 (B M I)、病歴、薬物使用、喫煙および飲酒が挙げられるが、これらに限定されない。

【0061】

クローン性造血由来の変異：本明細書中で使用されるとき、「クローン性造血由来の変異」とは、クローン性増殖に至る、造血性幹細胞および／または造血性前駆細胞におけるゲノム変異の体細胞性獲得のことを指す。

40

【0062】

未確定の潜在能をもつクローン性造血：本明細書中で使用されるとき、「未確定の潜在能をもつクローン性造血」または「C H I P」とは、造血幹細胞の拡大を意味する、個体における造血のことを指し、その造血幹細胞は、1 つまたはそれを超える体細胞変異 (例えば、血液悪性腫瘍に関連する変異および／またはそうでない変異) を含むが、血液悪性腫瘍に対する診断基準 (例えば、異形成の決定的な形態的エビデンス) を有しない。C H I P は、造血幹細胞が、遺伝的に異なる血液細胞の部分集団の形成に寄与する、ありふれた加齢性の現象である。

50

【0063】

クローナリティ境界値：本明細書中で使用されるとき、「クローナリティ境界値」とは、所与のサブクローナリティスコアの計算において用いられる、選択された値のことを指す。

【0064】

比較対照結果：本明細書中で使用されるとき、「比較対照結果」または「参照結果」は、所与の試験サンプルまたは試験結果の1つまたはそれを超える有望な特性を識別するため、ならびに／あるいは試験サンプルを採取したもしくは得た被験体に対する1つもしくはそれを超える考えられる予後の結果および／または1つもしくはそれを超えるカスタマイズされた療法を識別するためにその試験サンプルまたは試験結果と比較され得る、1つの結果または結果のセットを意味する。比較対照結果は、通常、参照サンプルのセット（例えば、試験被験体と同じ疾患もしくは癌のタイプを有する被験体、および／または試験被験体と同じ療法を受けているもしくは受けた被験体）から得られる。

【0065】

コントロールサンプル：本明細書中で使用されるとき、「コントロールサンプル」または「コントロールDNAサンプル」とは、解析手順の精度を評価するために、試験サンプルとともにまたは試験サンプルと比較して解析される、組成が既知のサンプルならびに／または特性が既知および／もしくはパラメータが既知（例えば、細胞起源が既知、腫瘍フラクションが既知、カバレッジが既知など）のサンプルのことを指す。コントロールサンプルのデータセットは、通常、少なくとも約25個～少なくとも約30,000個またはそれを超えるコントロールサンプルを含む。いくつかの実施形態において、コントロールサンプルのデータセットは、約50、75、100、150、200、300、400、500、600、700、800、900、1,000、2,500、5,000、7,500、10,000、15,000、20,000、25,000、50,000、100,000、1,000,000個またはそれを超えるコントロールサンプルを含む。

【0066】

カバレッジ：本明細書中で使用されるとき、「カバレッジ」とは、特定の塩基位置を占める核酸分子の数のことを指す。

【0067】

カットオフ閾値：本明細書中で使用されるとき、「カットオフ閾値」とは、サブクローナリティスコアを有する分類アレルを標的核酸バリエーションフィルタリストまたは非標的核酸バリエーションフィルタリストに割り当てるために、そのサブクローナリティスコアと比較される選択された値のことを指す。

【0068】

デオキシリボ核酸またはリボ核酸：本明細書中で使用されるとき、「デオキシリボ核酸」または「DNA」とは、糖部分の2'位に水素基を有する天然のヌクレオチドまたは修飾ヌクレオチドのことを指す。DNAは、通常、デオキシリボヌクレオシドを含むヌクレオチドの鎖を含み、その各デオキシリボヌクレオシドは、4つのタイプの核酸塩基、すなわち、アデニン（A）、チミン（T）、シトシン（C）およびグアニン（G）のうちの1つを含む。本明細書中で使用されるとき、「リボ核酸」または「RNA」とは、糖部分の2'位にヒドロキシル基を有する天然のヌクレオチドまたは修飾ヌクレオチドのことを指す。RNAは、通常、リボヌクレオシドを含むヌクレオチドの鎖を含み、その各リボヌクレオシドは、4つのタイプの核酸塩基、すなわち、A、ウラシル（U）、GおよびCのうちの1つを含む。本明細書中で使用されるとき、用語「ヌクレオチド」とは、天然のヌクレオチドまたは修飾ヌクレオチドのことを指す。ある特定のヌクレオチド対は、相補的な形式（相補的な塩基対形成と呼ばれる）で互いに特異的に結合する。DNAの場合、アデニン（A）は、チミン（T）と対形成し、シトシン（C）は、グアニン（G）と対形成する。RNAの場合、アデニン（A）は、ウラシル（U）と対形成し、シトシン（C）は、グアニン（G）と対形成する。第1の核酸鎖が、その第1の鎖におけるヌクレオチドと相補的なヌクレオチドで構成されている第2の核酸鎖に結合すると、それらの2本の鎖は、

結合して二本鎖を形成する。本明細書中で使用されるとき、「核酸シーケンシングデータ」、「核酸シーケンシング情報」、「配列情報」、「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「フラグメント配列」または「核酸シーケンシングリード」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、エクソーム、オリゴヌクレオチド、ポリヌクレオチドまたはフラグメント）におけるヌクレオチド塩基（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）の順序および同一性を示す任意の情報またはデータを表す。本教示は、利用可能なあらゆる手法、プラットフォームまたは技術（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、直接または間接的なヌクレオチド識別システム、パイロシーケンシング、イオンまたはpHに基づく検出システム、および電子的シグネチャに基づくシステムを含むがこれらに限定されない）を用いて得られる配列情報を企図することが理解されるべきである。

10

【0069】

フラグメント：本明細書中で使用されるとき、セルフリー核酸の文脈における「フラグメント」とは、被験体の身体（または被験体から得られたサンプル）に天然に存在する核酸分子のことを指し、断片化工程をインビトロで行う必要があると解釈されるべきでない。

【0070】

造血幹細胞：本明細書中で使用されるとき、「造血幹細胞」または「HSC」は、造血プロセスによって他の血液細胞を発生させる幹細胞である。

20

【0071】

免疫療法：本明細書中で使用されるとき、「免疫療法」とは、癌細胞を殺滅するようにまたは癌細胞の成長を少なくとも阻害するように、好ましくは、癌のさらなる成長を抑えるように、癌のサイズを縮小するように、および／または癌を排除するように、免疫系を刺激するように作用する1つまたはそれを超える作用物質による処置のことを指す。そのような作用物質のいくつかは、癌細胞上に存在する標的に結合し；いくつかは、免疫細胞上に存在するが癌細胞上に存在しない標的に結合し；いくつかは、癌細胞上と免疫細胞上の両方に存在する標的に結合する。そのような作用物質としては、チェックポイント阻害剤および／または抗体が挙げられるが、これらに限定されない。チェックポイント阻害剤は、自己寛容を維持し、末梢組織における生理学的免疫応答の持続時間および振幅を調節して付随的な組織損傷を最小限に抑える、免疫系の経路の阻害剤である（例えば、Paradoll, Nature Reviews Cancer 12, 252-264 (2012)を参照のこと）。例示的な作用物質としては、PD-1、PD-2、PD-L1、PD-L2、CTLA-40、OX40、B7.1、B7He、LAG3、CD137、KIR、CCR5、CD27またはCD40のいずれかに対する抗体が挙げられる。他の例示的な作用物質としては、炎症促進性サイトカイン、例えば、IL-1 β 、IL-6およびTNF- α が挙げられる。他の例示的な作用物質は、腫瘍に対して活性化されたT細胞、例えば、そのT細胞によって認識される腫瘍抗原を標的化するキメラ抗原を発現することによって活性化されたT細胞である。

30

40

【0072】

インデル：本明細書中で使用されるとき、「インデル」とは、被験体のゲノムにおけるヌクレオチド位置の挿入または欠失を含む変異のことを指す。

【0073】

インデックスされた：本明細書中で使用されるとき、「インデックスされた」とは、第1のエレメント（例えば、臨床情報）が第2のエレメント（例えば、所与のサンプル）に結び付けられることを指す。

【0074】

最大マイナーアレル頻度：本明細書中で使用されるとき、「最大マイナーアレル頻度」、「最大MAF」または「maxMAF」とは、所与のサンプル中に存在するまたは所与

50

のサンプルにおいて観察されるすべての体細胞性バリエーションのうち最大または最高の MAF のことを指す。

【0075】

マイナーアレル頻度：本明細書中で使用されるとき、「マイナーアレル頻度」または「MAF」とは、マイナーアレル（例えば、最も一般的でないアレル）が、所与の核酸集団（例えば、被験体から得られたサンプル）に存在する頻度のことを指す。換言すれば、「マイナーアレル頻度」は、所与のサンプル中の所与の遺伝子座において観察される最も優勢なアレルではない、そのサンプル中のその遺伝子座において観察されるアレルの頻度を意味する。MAFは、一般に、割合またはパーセンテージとして表現される。例えば、MAFは、通常、所与の遺伝子座に存在するすべての体細胞性バリエーションまたは全アレルの約0.5、0.1、0.05または0.01未満（すなわち、約50%、10%、5%または1%未満）である。

10

【0076】

変異：本明細書中で使用されるとき、「変異」または「遺伝子異常」とは、公知の参照配列からのバリエーションのことを指し、それには、変異、例えば、単一ヌクレオチドバリエーション（SNV）、コピー数バリエーションもしくはコピー数バリエーション（CNV）／コピー数異常、挿入または欠失（インデル）、トランケーション、遺伝子融合、トランスバージョン、転座、フレームシフト、重複、反復増殖およびエピジェネティックバリエーションが含まれる。変異は、生殖細胞系変異または体細胞変異であり得る。いくつかの実施形態において、比較目的の参照配列は、試験サンプルを提供する被験体の種の野生型ゲノム配列、通常、ヒトゲノムである。

20

【0077】

新生物：本明細書中で使用されるとき、用語「新生物」と「腫瘍」は、交換可能に使用される。それらは、被験体における細胞の異常な成長のことを指す。新生物または腫瘍は、良性、潜在的に悪性または悪性であり得る。悪性腫瘍は、癌または癌性腫瘍と称される。

【0078】

次世代シーケンシング：本明細書中で使用されるとき、「次世代シーケンシング」または「NGS」とは、従来のSanger法に基づくアプローチおよびキャピラリー電気泳動に基づくアプローチと比べて高スループットである、例えば、数十万個の比較的小さい配列リードを一度に生成できる、シーケンシング技術のことを指す。次世代シーケンシング法のいくつかの例としては、合成によるシーケンシング、ライゲーションによるシーケンシングおよびハイブリダイゼーションによるシーケンシングが挙げられるが、これらに限定されない。

30

【0079】

核酸タグ：本明細書中で使用されるとき、「核酸タグ」とは、異なるサンプル由来の核酸を区別するために核酸分子をラベルするため（例えば、サンプルインデックスに相当する）、または異なるタイプのもしくは異なる処理を経た、同じサンプル中の異なる核酸分子を区別するために核酸分子をラベルするため（例えば、分子タグに相当する）に使用される短い核酸（例えば、約500、約100、約50または約10ヌクレオチド長未満）のことを指す。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り得る。核酸タグは、必要に応じて、同じ長さまたは様々な長さを有する。核酸タグはまた、1つもしくはそれを超える平滑末端を有する二本鎖分子を含み得、5'もしくは3'の一本鎖領域（例えば、オーバーハング）を含み得、かつ／または所与の分子内の他の位置に1つもしくはそれを超える他の一本鎖領域を含み得る。核酸タグは、他の核酸（例えば、増幅および／またはシーケンシングされるサンプル核酸）の片端または両端に付着され得る。核酸タグをデコードすることにより、サンプル起源、所与の核酸の形態または処理などの情報を明らかにすることができる。核酸タグを用いることにより、種々の核酸タグおよび／またはサンプルインデックスを有する核酸を含む複数のサンプルのプーリングおよび／または並列処理も可能にすることができ、それらの核酸は、その後、核酸タグを読み出

40

50

すことによってデコンボリユートされる。核酸タグは、分子識別子もしくは分子タグ、サンプル識別子、インデックスタグおよび／またはバーコードとも称され得る。追加的または選択的に、核酸タグを用いることにより、同じサンプル中の異なる分子を区別することができる。これには、例えば、所与のサンプル中の異なる各核酸分子をユニークにタグ化すること、またはそのような分子を非ユニークにタグ化することが含まれる。非ユニークにタグ化する場合、異なる分子が、例えば、少なくとも1つの核酸タグとともに、選択された参照ゲノムに位置する開始／終止位置に基づいて区別され得るように、限られた数のタグを用いて各核酸分子をタグ化し得る。任意の2つの分子が、同じ開始／終止位置を有し、かつ同じ核酸タグを有する確率が低くなる（例えば、約10%未満、約5%未満、約1%未満または約0.1%未満の確率）ように、通常、十分な数の異なる核酸タグが使用される。いくつかの核酸タグは、サンプル、サンプル内の核酸分子の形、ならびに同じ開始位置および終止位置を有する形の中の核酸分子をラベルする複数の分子識別子を含む。そのような核酸タグは、例示的な形「A 1 i」を用いて言及され得、ここで、大文字は、サンプルタイプを示し、アラビア数字は、サンプル内の分子の形を示し、小文字のローマ数字は、ある形の中の分子を示す。

10

【0080】

ポリヌクレオチド：本明細書中で使用されるとき、「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」とは、ヌクレオシド間結合によってつながったヌクレオシド（デオキシリボヌクレオシド、リボヌクレオシドまたはそれらのアナログを含む）の直鎖状ポリマーのことを指す。通常、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、数モノマー単位、例えば、3～4個から、数百モノマー単位の範囲のサイズであることが多い。ポリヌクレオチドが、「A T G C C T G」などの文字の配列によって表されるときは常に、それらのヌクレオチドは、左から右に向かって5'→3'の順序であること、およびDNAの場合、別段述べられない限り、「A」は、デオキシアデノシンを表し、「C」は、デオキシシチジンを表し、「G」は、デオキシグアノシンを表し、「T」は、デオキシチミジンを表すことが理解される。A、C、GおよびTといった文字は、当該分野において標準的であるように、塩基自体、ヌクレオシド、または塩基を含むヌクレオチドのことを指すために使用され得る。

20

【0081】

潜在的な臨床的有意性：本明細書中で使用されるとき、アレルバリエーションの文脈における「潜在的な臨床的有意性」とは、あるアレルが被験体由来の所与の核酸分子に存在することが、その被験体に対する健康管理の決定に影響を与え得ることを指す。

30

【0082】

参照配列：本明細書中で使用されるとき、「参照配列」または「参照ゲノム」とは、実験によって決定された配列との比較目的で使用される公知の配列のことを指す。例えば、公知の配列は、全ゲノム、染色体またはその任意のセグメントであり得る。参照配列は、通常、少なくとも約20、少なくとも約50、少なくとも約100、少なくとも約200、少なくとも約250、少なくとも約300、少なくとも約350、少なくとも約400、少なくとも約450、少なくとも約500、少なくとも約1000ヌクレオチドまたはそれを超えるヌクレオチドを含む。参照配列は、ゲノムもしくは染色体の単一の連続配列とアラインメントし得るか、またはゲノムもしくは染色体の種々の領域とアラインメントする連続していないセグメントを含み得る。例示的な参照配列としては、例えば、ヒトゲノム、例えば、hg19およびhg38が挙げられる。

40

【0083】

サンプル：本明細書中で使用されるとき、「サンプル」は、本明細書中に開示される方法および／またはシステムによって解析されることが可能なものを意味する。

【0084】

感度：本明細書中で使用されるとき、所与のアッセイまたは方法の文脈における「感度」とは、そのアッセイまたは方法が、標的の被検体（例えば、腫瘍細胞を起源とするcfDNAフラグメント）および標的でない被検体（例えば、非腫瘍細胞を起源とするcfD

50

N Aフラグメント)を検出および区別する能力のことを指す。

【0085】

シーケンシング：本明細書中で使用されるとき、「シーケンシング」とは、生体分子、例えば、DNAまたはRNAなどの核酸の配列（例えば、モノマー単位の同一性および順序）を決定するために用いられるいくつかの技術のうちのいずれかのことを指す。例示的なシーケンシング法としては、ターゲットシーケンシング、一分子リアルタイムシーケンシング、エキソンシーケンシングまたはエクソームシーケンシング、イントロンシーケンシング、電子顕微鏡法に基づくシーケンシング、パネルシーケンシング、トランジスタ媒介性シーケンシング、ダイレクトシーケンシング、ランダムショットガンシーケンシング、Sangerジデオキシ終結シーケンシング、全ゲノムシーケンシング、ハイブリダイゼーションによるシーケンシング、パイロシーケンシング、キャピラリー電気泳動、ゲル電気泳動、二重鎖シーケンシング、サイクルシーケンシング、一塩基伸長シーケンシング、固相シーケンシング、ハイスループットシーケンシング、大規模並列処理シグネチャシーケンシング、エマルジョンPCR、低変性温度PCRでの共増幅（co-amplification at lower denaturation temperature-PCR）（COLD-PCR）、マルチプレックスPCR、リバーシブルダイターミネーターによるシーケンシング、ペアードエンドシーケンシング、短期（near-term）シーケンシング、エキソヌクレアーゼシーケンシング、ライゲーションによるシーケンシング、ショートリードシーケンシング、一分子シーケンシング、合成によるシーケンシング、リアルタイムシーケンシング、リバースターミネーターシーケンシング、ナノポアシーケンシング、454シーケンシング、Solexa Genome Analyzerシーケンシング、SOLiDTMシーケンシング、MS-PETシーケンシングおよびそれらの組み合わせが挙げられるが、これらに限定されない。いくつかの実施形態において、シーケンシングは、遺伝子アナライザー、例えば、とりわけ、Illumina, Inc.、Pacific Biosciences, Inc. またはApplied Biosystems/Thermo Fisher Scientificから商業的に入手可能な遺伝子アナライザーによって行われ（performer）得る。

【0086】

配列情報：本明細書中で使用されるとき、核酸ポリマーの文脈における「配列情報」は、そのポリマーにおけるモノマー単位（例えば、ヌクレオチドなど）の順序および同一性を意味する。

【0087】

体細胞変異：本明細書中で使用されるとき、「体細胞変異」は、受胎後に生じる、ゲノムにおける変異を意味する。体細胞変異は、生殖細胞を除く身体の任意の細胞において生じ得るので、子孫には受け継がれない。

【0088】

スプライス部位バリエーション：本明細書中で使用されるとき、核酸変異の文脈における「スプライス部位バリエーション」とは、エキソンとイントロンの境界線（スプライス部位）において生じる、所与のDNA配列における遺伝子変化のことを指す。この変化は、RNAスプライシングを妨害して、エキソンの喪失またはイントロンの包含およびタンパク質コード配列の変化をもたらす得る。

【0089】

特異度：本明細書中で使用されるとき、診断解析または診断アッセイの文脈における「特異度」とは、その解析またはアッセイが、所与のサンプルの他の構成要素を排除するまで、意図した標的被検体を検出する程度のことを指す。

【0090】

サブクロナリティスコア：本明細書中で使用されるとき、「サブクロナリティスコア」は、所与のアレルが、あるサンプルセットにおいて観察されたまたは出現した回数の合計数に対する（すなわち、その合計数で除算した）、その所与のアレルが、そのサンプルセットにおいてクロナリティ境界値未満のMAF/maxMAF比の値を有すると観

10

20

30

40

50

察された回数の比である。

【0091】

被験体：本明細書中で使用されるとき、「被験体」または「試験被験体」とは、動物、例えば、哺乳動物種（例えば、ヒト）もしくは鳥類（例えば、トリ）の種、または植物などの他の生物のことを指す。より詳細には、被験体は、脊椎動物、例えば、哺乳動物、例えば、マウス、霊長類、サルまたはヒトであり得る。動物には、家畜（例えば、生産用のウシ、乳牛、家禽、ウマ、ブタなど）、競技用動物および伴侶動物（例えば、ペットまたは補助用動物）が含まれる。被験体は、健康な個体、疾患もしくは疾患の素因を有するかもしれないもしくは有すると疑われる個体、または療法を必要とするかもしれないもしくは療法を必要とすると思われる個体であり得る。用語「個体」または「患者」は、「被験体」と相互交換可能であると意図される。いくつかの実施形態において、被験体は、癌を有するまたは癌を有すると疑われるヒトである。例えば、被験体は、癌を有すると診断された個体、癌療法を受けることになっている個体、および／または少なくとも1つの癌療法を受けた個体であり得る。被験体は、癌の緩解状態にあり得る。別の例として、被験体は、自己免疫疾患を有すると診断されている個体であり得る。別の例として、被験体は、疾患、例えば、癌、自己免疫疾患を有すると診断されたかもしれないもしくは有すると疑われたかもしれない、妊娠中または妊娠を計画中の女性個体であり得る。

10

【0092】

実質的なマッチ：本明細書中で使用されるとき、「実質的なマッチ」は、少なくとも1つの第1の値またはエレメントが、少なくとも1つの第2の値またはエレメントと少なくともほぼ等しいことを意味する。ある特定の実施形態において、例えば、c f DNA サンプルの所与のアレルバリエーションの細胞起源は、そのアレルバリエーションと参照サンプルまたは分類アレルとの間に少なくとも1つの実質的なマッチまたはおおよそのマッチ（例えば、配列アラインメントおよび／または他の臨床情報もしくは特性）があるとき、判定される。

20

【0093】

実質的にアラインメントする：本明細書中で使用されるとき、核酸配列のアラインメントの文脈における句「実質的にアラインメントする」は、第1の核酸配列が、第2の核酸配列の少なくとも1つの部分配列と少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%またはさらには100%の配列同一性を有することを意味する。いくつかの実施形態において、例えば、所与の配列リードが、参照配列の少なくとも1つの部分配列もしくは領域または全体と95%、96%、97%、98%、99%または100%の配列同一性を有するとき、その所与の配列リードは、その参照配列と実質的にアラインメントする。

30

【0094】

閾値：本明細書中で使用されるとき、「閾値」とは、実験によって測定された値を特徴付けるまたは分類するために用いられる別々に決定された値のことを指す。

【0095】

トランケーション：本明細書中で使用されるとき、核酸変異の文脈における「トランケーション」とは、発現されたとき所与のDNA配列によってコードされるポリペプチド（例えば、タンパク質）を切断または短縮し得る、そのDNA配列において観察される配列バリエーションのことを指す。

40

【0096】

腫瘍フラクション：本明細書中で使用されるとき、「腫瘍フラクション」とは、所与のサンプル中の、腫瘍に由来する核酸分子の割合の推定値のことを指す。例えば、サンプルの腫瘍フラクションは、そのサンプルの最大マイナーアレル頻度（maxMAF）もしくはそのサンプルのカバレッジ、あるいはそのサンプル中のc f DNAフラグメントの長さ、エピジェネティック状況もしくは他の特性またはそのサンプルの他の任意の選択された特徴から導かれる尺度であり得る。用語「maxMAF」とは、所与のサンプル中に存在するすべての体細胞性バリエーションのうち最大または最高のMAFのことを指す。いくつか

50

の実施形態において、あるサンプルの腫瘍フラクションは、そのサンプルのmaxMAFと等しい。

【0097】

値：本明細書中で使用されるとき、「値」とは、一般に、その値が言及している特徴を特徴付けるものであり得る、データセットへの登録のことを指す。これには、数字、単語または句、符号（例えば、+または-）または程度が含まれるが、これらに限定されない。

詳細な説明

【0098】

緒言

10

【0099】

患者から得られたサンプル中に存在するセルフリー核酸（cfNA）における癌細胞DNAまたは他の標的核酸の検出の改善された感度および／または特異度をもたらすための方法、コンピュータ可読媒体およびシステムが、本明細書中に提供される。その主題の方法、コンピュータ可読媒体およびシステムは、腫瘍のcfDNA解析および他の標的cfNA（例えば、米国特許第9920366（B2）号、米国特許第9840743（B2）号およびPCT公開特許出願WO2017/181146（A1）（これらの各々は、参照により援用される）に記載されている手法）に容易に適用され得る。いくつかの実施形態において、アレルが癌細胞を起源とするかそれとも造血幹細胞を起源とするかを決定するために用いられ得る、それらのアレルを識別する方法が本明細書中に提供される。そのような情報価値のあるアレルは、いったん識別されたら、ある特定の例示的な実施形態において、そのアレルを用いることにより、サンプルを、腫瘍細胞DNAを含むまたは腫瘍細胞DNAを含まないと分類することができる。

20

【0100】

CFNAの細胞起源を決定する方法および関連する態様

【0101】

本願は、セルフリー核酸（cfNA）サンプルが、所与の細胞タイプまたは組織タイプを起源とする核酸分子またはフラグメントを含むか否かを決定することに関連する様々な方法を開示する。いくつかの例示的な実施形態において、それらの方法は、cfNAサンプルが、罹患細胞（例えば、腫瘍細胞など）、胎児細胞、移植ドナー細胞などを起源とする核酸分子（例えば、セルフリーデオキシリボ核酸（cfDNA）フラグメントおよび／またはセルフリーリボ核酸（cfRNA）フラグメント）を含むか否かを決定するために用いられる。しばしば、これらのタイプの核酸分子は、所与のcfNAサンプル中に存在する全核酸分子のほんの一部しか占めず、そのcfNAサンプルは一般に、例えば、非罹患細胞、正常細胞または健康細胞（例えば、造血幹細胞または他の非腫瘍細胞）、母体細胞、移植レシピエント細胞などを起源とする核酸分子の大きなバックグラウンドを含む。多くの既存の分析技術は、cfNAサンプル中にそのような少数で存在する核酸分子を確実に検出し、特徴付けるのに十分な感度を有しない。本明細書中に開示される方法から得られる情報は、通常、cfNAサンプルを得た被験体が所与の疾患、障害または症状を有するか否かを診断するために用いられる。ある特定の実施形態において、それらの方法は、被験体に療法を施す工程、または診断された疾患、障害もしくは症状を処置する工程を含む。本願は、例えば、分類器を生成する関連する方法、ならびに試験サンプル中のcfNAフラグメントの細胞起源を分類する際に有用なサブクローナリティスコアのデータベースを作成する方法も開示する。

30

40

【0102】

主題の方法の様々な実施形態において、複数の遺伝子座のアレルバリエーションおよびそれらの各遺伝子座におけるアレル頻度を検出するために、それらの遺伝子座がシーケンシングされる。上記DNAは、種々の細胞起源（各々がセルフリーDNAを生成する）に由来し得、それによって、同じ遺伝子座について異なるゲノム起源に由来するセルフリーDNAの混合物が生成される。そのDNA起源は、腫瘍細胞（同じ被験体に存在するいくつか

50

のクローン的に異なる腫瘍細胞バリエーションを含む）および非腫瘍細胞（特に、血液細胞）であり得る。いくつかの実施形態では、シーケンシングのためにゲノム領域が標的化される（ホールゲノムシーケンシングとは対照的に）。同じ遺伝子座の複数のアレルを検出するため、およびその遺伝子座におけるアレル頻度を提供するために、ハイスループットDNA配列決定装置を用いることによって、サンプル由来のcfDNAの複数のフラグメントが同時にシーケンシングされ得る。未確定の潜在能をもつクローン性造血（CHIP）は、造血幹細胞が、遺伝的に異なる血液細胞の部分集団の形成に寄与する、ありふれた加齢性の現象である。これらの造血幹細胞は、癌性細胞において生成されたアレルバリエーションと混同され得るセルフリーDNAのアレル情報を生成し得る。

【0103】

参照被験体由来のアレル情報のデータベースを用いることにより、セルフリーDNAサンプルを、腫瘍細胞DNAを含むまたは含まないと分類するために用いることができるアレルが発見され得る。それらのデータベースは、通常、癌を有すると疑われる任意の被験体由来のセルフリーDNA配列情報を含む。一般に、データベースが大きいほど、そのデータベースは、セルフリーDNA内の腫瘍細胞DNAの有無を示すアレルバリエーションを発見するために用いることができるアレルバリエーションの特定にとって有用である。データベースでは、各患者に対して、潜在的に臨床的に有意な複数の遺伝子座がシーケンシングされており、シーケンシングされた各遺伝子座について、その遺伝子座における各アレルの頻度を決定する。各遺伝子座に対するマイナーアレル頻度（MAF）も決定する。所与のcfDNAサンプルにおける遺伝的不均一性のせいで、各MAFは、遺伝子座によって大きく異なり得る。例えば、ある遺伝子座におけるドライバー変異は、腫瘍の進展中に後のクローンにおいて獲得されるパッセンジャー変異よりも高いMAFを有する可能性が高い。所与の患者について、解析されるアレルセットの中で最大MAF（maxMAF）を有するアレルが明らかにされ、maxMAFに対するMAFの値も明らかにされる。データベースは、各患者に対する他の臨床情報も含み得、その他の臨床情報は、各患者の遺伝情報と互いに関係づけられ得る。そのような臨床情報の例としては、腫瘍の検出、患者の生存時間、患者の年齢などが挙げられる。

【0104】

次いで、cfDNAサンプルを、臨床的に有意な腫瘍DNAを含むまたは臨床的に有意な腫瘍DNAを含まないと分類するために用いることができるアレルについて、データベース内のアレル情報をスクリーニングし得る。試験サンプル中の潜在的に臨床的に有意な所与の各アレルバリエーションについて、マイナーアレル頻度（MAF）とmaxMAFとの比を決定する。次いで、目的のアレルに対するMAF/maxMAF比の計算値を、通常、データベース内の多くのサンプルに対して生成する。次いで、データベース（またはデータベースの一部）内の所与のアレルに対する各MAF/maxMAF値の頻度を求め得る。例えば、MAF/maxMAF値のヒストグラムをプロットし得る。次いで、サブクローナリティスコアを算出するためにクローナリティ境界値を設定し得る。そのサブクローナリティスコアは、データベースに集められたサンプルに所与のアレルが観察された場合の総数に対する、データベース内の所与のアレルがクローナリティ境界未満のMAF/maxMAF値を有する場合の数の比である。次いで、所与のアレルが腫瘍DNAの存在を示すか否かを決定するようにカットオフ閾値を設定し得る。その閾値を超えるサブクローナリティスコアを有するアレルを用いることにより、非腫瘍DNAに由来するアレルを識別することができる。逆に、その閾値未満のサブクローナリティスコアを有するアレルを用いることにより、腫瘍DNAに由来するアレルを識別することができる。

【0105】

例えば、50%のクローナリティ境界値は、例えば図1AおよびBに示されるように設定され得る。示されているように、アレル1は、50%に設定されたクローナリティ境界値に基づいて推定された高サブクローナリティスコアを有し、アレル2は、低サブクローナリティスコアを有する。したがって、アレルは、いくつかの実施形態において、非腫瘍DNAを示す（陰性、例えば、図1Aにおけるアレル1）または腫瘍DNAを示す（陽性

10

20

30

40

50

、例えば、図1Bにおけるアレル2)、2つのカテゴリーのうちの1つに該当し得る。図1AおよびBに提供される例において、アレル1およびアレル2は、異なる遺伝子座に存在し、すなわち、同じ遺伝子座のバリエーションアレルではない。この解析は、データベース内の試験された複数のアレルに適用され得、所与のアレルを陽性カテゴリーまたは陰性カテゴリーのいずれかに入れることによって、陽性および陰性のアレルのセットが生成される。必要に応じて、いずれかのカテゴリーからアレルを除外するために、よりストリンジェントな選択閾値を適用してもよく、従って、それらの除外されたアレルは、所与のサンプルに対して分類決定を行うために用いられない。例えば、25%未満のサブクロナリティスコアを有するアレルは、陽性であり得、75%を超えるサブクロナリティスコアを有するアレルは、陰性であり得、除外された範囲内(すなわち、25%~75%)のアレルは、サンプルを、腫瘍DNAを含むまたは含まないと分類するために用いられない。類別されたアレルを用いることにより、被験体から得られた所与の試験サンプルを分類するためのアレルのリストを生成することができる。そのようなリストは、その用語の使用の文脈に応じて、「腫瘍バリエーションフィルターリスト」または「非腫瘍バリエーションフィルターリスト」と称される。また、図1AおよびBは、MAF/maxMAF分布に対するクロナリティ境界値が、データベースに集められたサンプルに対して50%に設定されている例を提供している。

10

【0106】

低サブクロナリティスコアは、通常、観察されたアレルが、腫瘍DNAの存在を示すことを示唆する。例えば、スコアがゼロであることは、アレルが観察されたデータベースに集められたどのサンプルにおいても、MAF/maxMAFが、クロナリティ境界値を超えることを示唆し得、それは、そのアレルが、データベース内の各サンプルにおける優勢なマイナーアレルであったことを示唆し得る。

20

【0107】

サンプル中の陽性(すなわち、腫瘍起源)および陰性(すなわち、非腫瘍起源)のアレルの有無についての試験に加えて、他の分類基準も、必要に応じて用いられる。臨床的に有意な変異をコールするために、MAF/maxMAF比から発見された陽性および陰性のアレルセット内の情報を使用することは、一般に、低MAFを有するアレルにとって最も有用である。いくつかの実施形態において、例えば、所与のバリエーションアレルが、1%を超えるMAFを有すると見出され、そのアレルが、陰性アレルである場合、そのアレルが陰性アレルリストに掲載されていたとしても、そのサンプルは、なおも腫瘍DNAを含むと分類される。別の例において、あるバリエーションアレルが、2%を超えるMAFを有すると見出される場合、ある特定の実施形態において、そのアレルが陰性アレルリストに掲載されていたとしても、そのサンプルは、腫瘍DNAを含むと分類される。他の実施形態では、を超えるMAF閾値が用いられ得る。

30

【0108】

別の例示的な分類基準は、観察されるアレルバリエーションのタイプである。いくつかの実施形態において、例えば、アレルバリエーションが、サブクロナリティスコアによって陰性と分類され、MAFが、選択された値より低かったとしても(例えば、いくつかの実施形態では2%未満、他の実施形態では1%未満)、そのバリエーションは、臨床的有意性を有するとコールされる。トランケーション、インデルまたはスプライス部位バリエーションなどのアレルバリエーションは、癌を示し、通常、造血幹細胞には存在しない。

40

【0109】

いくつかの実施形態において、患者由来のcfDNAサンプルは、以下の基準:(1)トランケーション、インデルまたはスプライス部位バリエーションであるアレルバリエーションを有する、(2)サブクロナリティスコア陽性アレルを有する、または(3)1%を超えるMAFを有するサブクロナリティスコア陰性アレルを有する、のうちのいずれか1つを満たす場合、癌細胞由来のDNAを含むと特徴付けられ得る。いくつかの実施形態において、患者由来のcfDNAサンプルは、以下の基準:(1)トランケーション、インデルまたはスプライス部位バリエーションであるアレルバリエーションを有する、(2)サブクロ

50

ナリティスコア陽性アレルを有する、または(3)2%を超えるMAFを有するサブクロ
ーナリティスコア陰性アレルを有する、のうちのいずれか1つを満たす場合、癌細胞由来
のDNAを含むと特徴付けられ得る。

【0110】

CHIP変異の頻度は、通常、患者の年齢とともに上昇するので、ある特定の実施形態
において、セルフリーDNAサンプルが腫瘍DNAを含むか否かを決定するために、分類
において、患者の年齢および／または他の患者データが利用され得る。他の例示的な実施
形態では、例えば、最小maxMAFに基づいて、サブクロナルリスト（例えば、標的
核酸バリエーションフィルタリストまたは非標的核酸バリエーションフィルタリスト）が、種
々のサンプルサブセットにわたって生成され、公知のドライバー変異に基づいてmaxM
AFなどをコールする。いくつかの実施形態では、特定の指摘（例えば、所与の癌のタイ
プ（例えば、肺、結腸直腸など））に基づいて、サブクロナルリストが生成される。ある
特定の実施形態では、変異アレル頻度、サブクローン比、遺伝子のタイプ、血液悪性腫
瘍に関連するバリエーション、患者の年齢、他のCHIPバリエーションの観察結果、癌のタイプ
などを含む1つまたはそれを超える特徴量に基づいて、機械学習分類器が訓練される。

【0111】

本明細書中に開示される方法の態様をさらに例証するために、図2は、少なくとも部分的
的にコンピュータを使用して、被験体内の標的細胞（例えば、腫瘍細胞など）を起源とす
る核酸分子を検出するための例示的な方法工程を模式的に表しているフローチャートを提
供している。示されているように、方法200は、被験体から得られた試験サンプル由来
のセルフリー核酸（cfNA）フラグメントから得られた配列リードを含む試験配列情報を
コンピュータによって受信する工程を工程202に含む。方法200はまた、標的核酸
バリエーションフィルタリスト上の少なくとも1つの分類アレルと実質的にマッチする少な
くとも1つのアレルバリエーションが試験配列情報に存在することを識別する工程であって、
その分類アレルは、少なくとも1つの選択されたカットオフ閾値（例えば、約10%、約
20%、約30%、約40%、約50%、約60%、約70%、約80%、約90%または別
の値）未満のサブクローナリティスコアを含み、それによって、その分類アレルは、
標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それによって、
その被験体内の標的細胞を起源とする核酸分子を検出する、工程を工程204に含む。い
くつかの実施形態において、例えば、工程204は、試験配列情報における少なくとも1
つのアレルバリエーションを識別する工程；そのアレルバリエーションを標的核酸バリエ
ーションフィルタリスト上の少なくとも1つの分類アレルにマッピングする工程；その分類
アレルのサブクローナリティスコアを識別する工程；およびそのサブクローナリティスコ
アを少なくとも1つの選択されたカットオフ閾値と比較する工程を含み、ここで、そのサ
ブクローナリティスコアが、選択されたカットオフ閾値未満であるとき、それは、その分
類アレルが、標的細胞を起源とする参照cfNAフラグメント由来であることを示す。コン
ピュータおよびコンピュータ可読媒体を含む関連するシステムが、さらに本明細書中に記
載される。

【0112】

図3は、いくつかの実施形態に係る、少なくとも部分的にコンピュータを使用して、被
験体内の腫瘍細胞を起源とする核酸分子を検出するための例示的な方法工程を模式的に表
しているフローチャートを提供している。示されているように、方法300は、被験体か
ら得られた試験サンプル中のセルフリーデオキシリボ核酸（cfDNA）フラグメントか
ら得られた配列リードを含む試験配列情報をコンピュータによって受信する工程を工程
302に含む。方法300はまた、コンピュータによって、被験体の造血幹細胞を起源とす
る1つまたはそれを超える配列リード（例えば、分類アレルの少なくとも一部を含む）を
試験配列情報から除去（例えば、削除、非表示、無視など）して、フィルタされた試験
配列情報を生成する工程を工程304に含む。方法300はさらに、1またはそれを超える
参照被験体から得られた参照配列情報と実質的にアラインメントする1つまたはそれ
を超える上記配列リードが上記フィルタされた試験配列情報に存在することをコンピュ
ー

タによって識別する工程であって、その参照配列情報は、参照被験体内の1つまたはそれを超える腫瘍細胞を起源とし、それによって、被験体内の腫瘍細胞を起源とする核酸分子を検出する、工程を工程306に含む。

【0113】

図4は、被験体における疾患を処置する例示的な方法工程を模式的に表しているフローチャートを提供している。示されているように、方法400は、被験体から得られた試験サンプル由来のセルフリー核酸(c f N A)フラグメントから得られた配列リードを含む試験配列情報を受信する工程を工程402に含む。方法400はさらに、標的核酸バリエーションフィルターリスト上の少なくとも1つの分類アレルと実質的にマッチする少なくとも1つのアレルバリエーションが試験配列情報に存在することを識別する工程であって、その分類アレルは、少なくとも1つの選択されたカットオフ閾値未満のサブクロナリティスコアを含み、それによって、その分類アレルが、罹患細胞を起源とする参照c f N Aフラグメント由来であることを示し、それによって、その被験体における疾患を診断する、工程を工程404に含む。さらに、方法400はまた、被験体に1つまたはそれを超える療法を施し、それによって、被験体における疾患を処置する工程を工程406に含む。例示的な療法は、本明細書中にさらに記載される。

【0114】

図5は、少なくとも部分的にコンピュータを使用して分類器を生成する例示的な方法工程を模式的に表しているフローチャートを提供している。示されているように、方法500は、1つまたはそれを超える参照サンプル由来のセルフリー核酸(c f N A)フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対するサブクロナリティスコアをコンピュータによって生成する工程を工程502に含み、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む。方法500はまた、それらのサブクロナリティスコアと少なくとも1つの選択されたカットオフ閾値(例えば、約10%、約20%、約30%、約40%、約50%、約60%、約70%、約80%、約90%または別の値)をコンピュータによって比較し、ここで、選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、その分類アレルが、非標的細胞を起源とする参照c f N Aフラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ/またはそのカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、その分類アレルが、標的細胞を起源とする参照c f N Aフラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられ、それによって、分類器を生成する工程を工程504に含む。

【0115】

図6は、少なくとも部分的にコンピュータを使用して分類器を生成する例示的な方法工程を模式的に表しているフローチャートを提供している。示されているように、方法600は、1つまたはそれを超える参照サンプル由来のセルフリー核酸(c f N A)フラグメントから得られた配列リードを含む配列情報から分類アレルセットをコンピュータによって識別する工程であって、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む、工程を工程602に含む。方法600はまた、その配列情報から、各参照サンプルにおける各分類アレルに対するマイナーアレル頻度(M A F)の値をコンピュータによって決定する工程を工程604に含み、各参照サンプルに対する最大マイナーアレル頻度(max M A F)の値をコンピュータによって決定する工程を工程606に含む。方法600はまた、所与の参照サンプルにおいて観察された各分類アレルについて、参照サンプルの少なくとも一部に対するmax M A Fの値に対するM A Fの値の比をコンピュータによって計算して、比の値を生成する工程を工程608に含む。方法600はまた、参照サンプルの少なくとも一部における所与の分類アレルが、参照サンプルの少なくとも一部に出現したその所与の分類アレルの回数の合計数に対する、少なくとも1つの選択されたクロナリティ境界値(例えば、約10%、約20%、約30%、約40%、約50%、約60%、約70%、約80%、

約 90 % または別の値) 未満の比の値を有した回数の比を、各分類アレルについて、コンピュータによって計算して、参照サンプルの少なくとも一部における各分類アレルに対するサブクロナリティスコアを生成する工程を工程 610 に含む。さらに、方法 600 は、それらのサブクロナリティスコアと少なくとも 1 つの選択されたカットオフ閾値 (例えば、約 10 %、約 20 %、約 30 %、約 40 %、約 50 %、約 60 %、約 70 %、約 80 %、約 90 % または別の値) をコンピュータによって比較し、ここで、選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照 cfNA フラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルターリストに加えられ、かつ/またはそのカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照 cfNA フラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルターリストに加えられ、それによって、分類器を生成する工程を工程 612 に含む。

10

【0116】

いくつかの実施形態において、上記方法は、被験体から cfDNA サンプルを得る工程を含む。本質的に任意のサンプルタイプが、必要に応じて利用される。ある特定の実施形態において、例えば、cfDNA サンプルは、血液、血漿、血清、痰、尿、精液、髄液、便、滑液、髄液、唾液などである。必要に応じて利用されるさらなる例示的なサンプルタイプは、本明細書中にさらに記載される。通常、被験体は、哺乳動物被験体 (例えば、ヒト被験体) である。本質的に任意のタイプの核酸 (例えば、DNA および/または RNA) が、本願に開示されている方法に従って評価され得る。いくつかの例としては、セルフリー核酸 (例えば、腫瘍起源、胎児起源、母体起源などの cfDNA)、細胞性核酸 (循環腫瘍細胞 (例えば、サンプル中のインタクトな細胞を溶解することによって得られる) を含む)、循環腫瘍核酸などが挙げられる。

20

【0117】

本願に開示される方法は、一般に、被験体から採取されたサンプル中の核酸から配列情報を得る工程を含む。ある特定の実施形態において、その配列情報は、核酸の標的化セグメントから得られる。本質的に任意の数のゲノム領域が、必要に応じて標的化される。標的化セグメントは、少なくとも 10、少なくとも 50、少なくとも 100、少なくとも 500、少なくとも 1000、少なくとも 2000、少なくとも 5000、少なくとも 10000、少なくとも 20000 または少なくとも 50000 (例えば、25、50、75、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10000、15000、20000、30000、40000、50000、60000、70000、80000、90000、100000、150000、200000、300000、400000、500000、600000、700000、800000、900000、1000000、1500000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、15000000、20000000、30000000、40000000、50000000、60000000、70000000、80000000、90000000、100000000、150000000、200000000、300000000、400000000、500000000、600000000、700000000、800000000、900000000、1000000000、1500000000、2000000000、3000000000、4000000000、5000000000、6000000000、7000000000、8000000000、9000000000、10000000000、15000000000、20000000000、30000000000、40000000000、50000000000、60000000000、70000000000、80000000000、90000000000、100000000000、150000000000、200000000000、300000000000、400000000000、500000000000、600000000000、700000000000、800000000000、900000000000、1000000000000、1500000000000、2000000000000、3000000000000、4000000000000、5000000000000、6000000000000、7000000000000、8000000000000、9000000000000、10000000000000、15000000000000、20000000000000、30000000000000、40000000000000、50000000000000、60000000000000、70000000000000、80000000000000、90000000000000、100000000000000、150000000000000、200000000000000、300000000000000、400000000000000、500000000000000、600000000000000、700000000000000、800000000000000、900000000000000、1000000000000000、1500000000000000、2000000000000000、3000000000000000、4000000000000000、5000000000000000、6000000000000000、7000000000000000、8000000000000000、9000000000000000、10000000000000000、15000000000000000、20000000000000000、30000000000000000、40000000000000000、50000000000000000、60000000000000000、70000000000000000、80000000000000000、90000000000000000、100000000000000000、150000000000000000、200000000000000000、300000000000000000、400000000000000000、500000000000000000、600000000000000000、700000000000000000、800000000000000000、900000000000000000、1000000000000000000、1500000000000000000、2000000000000000000、3000000000000000000、4000000000000000000、5000000000000000000、6000000000000000000、7000000000000000000、8000000000000000000、9000000000000000000、10000000000000000000、15000000000000000000、20000000000000000000、30000000000000000000、40000000000000000000、50000000000000000000、60000000000000000000、70000000000000000000、80000000000000000000、90000000000000000000、100000000000000000000、150000000000000000000、200000000000000000000、300000000000000000000、400000000000000000000、500000000000000000000、600000000000000000000、700000000000000000000、800000000000000000000、900000000000000000000、1000000000000000000000、1500000000000000000000、2000000000000000000000、3000000000000000000000、4000000000000000000000、5000000000000000000000、6000000000000000000000、7000000000000000000000、8000000000000000000000、9000000000000000000000、10000000000000000000000、15000000000000000000000、20000000000000000000000、30000000000000000000000、40000000000000000000000、50000000000000000000000、60000000000000000000000、70000000000000000000000、80000000000000000000000、90000000000000000000000、100000000000000000000000、150000000000000000000000、200000000000000000000000、300000000000000000000000、400000000000000000000000、500000000000000000000000、600000000000000000000000、700000000000000000000000、800000000000000000000000、900000000000000000000000、1000000000000000000000000、1500000000000000000000000、2000000000000000000000000、3000000000000000000000000、4000000000000000000000000、5000000000000000000000000、6000000000000000000000000、7000000000000000000000000、8000000000000000000000000、9000000000000000000000000、10000000000000000000000000、15000000000000000000000000、20000000000000000000000000、30000000000000000000000000、40000000000000000000000000、50000000000000000000000000、60000000000000000000000000、70000000000000000000000000、80000000000000000000000000、90000000000000000000000000、100000000000000000000000000、150000000000000000000000000、200000000000000000000000000、300000000000000000000000000、400000000000000000000000000、500000000000000000000000000、600000000000000000000000000、700000000000000000000000000、800000000000000000000000000、900000000000000000000000000、1000000000000000000000000000、1500000000000000000000000000、2000000000000000000000000000、3000000000000000000000000000、4000000000000000000000000000、5000000000000000000000000000、6000000000000000000000000000、7000000000000000000000000000、8000000000000000000000000000、9000000000000000000000000000、10000000000000000000000000000、15000000000000000000000000000、20000000000000000000000000000、30000000000000000000000000000、40000000000000000000000000000、50000000000000000000000000000、60000000000000000000000000000、70000000000000000000000000000、80000000000000000000000000000、90000000000000000000000000000、100000000000000000000000000000、150000000000000000000000000000、200000000000000000000000000000、300000000000000000000000000000、400000000000000000000000000000、500000000000000000000000000000、600000000000000000000000000000、700000000000000000000000000000、800000000000000000000000000000、900000000000000000000000000000、1000000000000000000000000000000、1500000000000000000000000000000、2000000000000000000000000000000、3000000000000000000000000000000、4000000000000000000000000000000、5000000000000000000000000000000、6000000000000000000000000000000、7000000000000000000000000000000、8000000000000000000000000000000、9000000000000000000000000000000、10000000000000000000000000000000、15000000000000000000000000000000、20000000000000000000000000000000、30000000000000000000000000000000、40000000000000000000000000000000、50000000000000000000000000000000、60000000000000000000000000000000、70000000000000000000000000000000、80000000000000000000000000000000、90000000000000000000000000000000、100000000000000000000000000000000、150000000000000000000000000000000、200000000000000000000000000000000、300000000000000000000000000000000、400000000000000000000000000000000、500000000000000000000000000000000、600000000000000000000000000000000、700000000000000000000000000000000、800000000000000000000000000000000、900000000000000000000000000000000、1000000000000000000000000000000000、1500000000000000000000000000000000、2000000000000000000000000000000000、3000000000000000000000000000000000、4000000000000000000000000000000000、5000000000000000000000000000000000、6000000000000000000000000000000000、7000000000000000000000000000000000、8000000000000000000000000000000000、9000000000000000000000000000000000、10000000000000000000000000000000000、15000000000000000000000000000000000、20000000000000000000000000000000000、30000000000000000000000000000000000、40000000000000000000000000000000000、50000000000000000000000000000000000、60000000000000000000000000000000000、70000000000000000000000000000000000、80000000000000000000000000000000000、90000000000000000000000000000000000、100000000000000000000000000000000000、150000000000000000000000000000000000、200000000000000000000000000000000000、300000000000000000000000000000000000、400000000000000000000000000000000000、500000000000000000000000000000000000、600000000000000000000000000000000000、700000000000000000000000000000000000、800000000000000000000000000000000000、900000000000000000000000000000000000、1000000000000000000000000000000000000、1500000000000000000000000000000000000、2000000000000000000000000000000000000、3000000000000000000000000000000000000、4000000000000000000000000000000000000、5000000000000000000000000000000000000、6000000000000000000000000000000000000、7000000000000000000000000000000000000、8000000000000000000000000000000000000、9000000000000000000000000000000000000、10000000000000000000000000000000000000、15000000000000000000000000000000000000、20000000000000000000000000000000000000、30000000000000000000000000000000000000、40000000000000000000000000000000000000、50000000000000000000000000000000000000、60000000000000000000000000000000000000、70000000000000000000000000000000000000、80000000000000000000000000000000000000、90000000000000000000000000000000000000、100000000000000000000000000000000000000、150000000000000000000000000000000000000、200000000000000000000000000000000000000、300000000000000000000000000000000000000、400000000000000000000000000000000000000、500000000000000000000000000000000000000、600000000000000000000000000000000000000、700000000000000000000000000000000000000、800000000000000000000000000000000000000、900000000000000000000000000000000000000、1000000000000000000000000000000000000000、1500000000000000000000000000000000000000、2000000000000000000000000000000000000000、3000000000000000000000000000000000000000、4000000000000000000000000000000000000000、5000000000000000000000000000000000000000、6000000000000000000000000000000000000000、7000000000000000000000000000000000000000、8000000000000000000000000000000000000000、9000000000000000000000000000000000000000、10000000000000000000000000000000000000000、15000000000000000000000000000000000000000、20000000000000000000000000000000000000000、30000000000000000000000000000000000000000、40000000000000000000000000000000000000000、50000000000000000000000000000000000000000、60000000000000000000000000000000000000000、70000000000000000000000000000000000000000、80000000000000000000000000000000000000000、90000000000000000000000000000000000000000、100000000000000000000000000000000000000000、150000000000000000000000000000000000000000、200000000000000000000000000000000000000000、300000000000000000000000000000000000000000、400000000000000000000000000000000000000000、500000000000000000000000000000000000000000、600000000000000000000000000000000000000000、700000000000000000000000000000000000000000、800000000000000000000000000000000000000000、900000000000000000000000000000000000000000、1000000000000000000000000000000000000000000、1500000000000000000000000000000000000000000、2000000000000000000000000000000000000000000、3000000000000000000000000000000000000000000、4000000000000000000000000000000000000000000、5000000000000000000000000000000000000000000、6000000000000000000000000000000000000000000、7000000000000000000000000000000000000000000、8000000000000000000000000000000000000000000、9000000000000000000000000000000000000000000、10000000000000000000000000000000000000000000、15000000000000000000000000000000000000000000、20000000000000000000000000000000000000000000、30000000000000000000000000000000000000000000、40000000000000000000000000000000000000000000、50000000000000000000000000000000000000000000、60000000000000000000000000000000000000000000、70000000000000000000000000000000000000000000、80000000000000000000000000000000000000000000、90000000000000000000000000000000000000000000、100000000000000000000000000000000000000000000、150000000000000000000000000000000000000000000、200000000000000000000000000000000000000000000、300000000000000000000000000000000000000000000、400000000000000000000000000000000000000000000、500000000000000000000000000000000000000000000、600000000000000000000000000000000000000000000、700000000000000000000000000000000000000000000、800000000000000000000000000000000000000000000、900000000000000000000000000000000000000000000、1000000000000000000000000000000000000000000000、1500000000000000000000000000000000000000000000、2000000000000000000000000000000000000000000000、3000000000000000000000000000000000000000000000、4000000000000000000000000000000000000000000000、5000000000000000000000000000000000000000000000、6000000000000000000000000000000000000000000000、7000000000000000000000000000000000000000000000、8000000000000000000000000000000000000000000000、9000000000000000000000000000000000000000000000、10000000000000000000000000000000000000000000000、15000000000000000000000000000000000000000000000、20000000000000000000000000000000000000000000000、30000000

ntal Cell Research, 322 (1) : 12-20 (2014)、Micic (Ed.), Sample Preparation Techniques for Soil, Plant, and Animal Samples (Springer Protocols Handbooks), 1st Ed., Humana Press (2016) および Chiu, Next-Generation Sequencing and Sequence Data Analysis, Bentham Science Publishers (2018) (これらの各々は、その全体が参照により援用される) にも記載されている。

【0119】

本明細書中に開示される方法は、通常、疾患、障害もしくは症状、特に癌が被験体に存在すると診断するため、そのような疾患、障害もしくは症状を特徴付けるため（例えば、所与の癌をステージ分類するため、癌の不均一性を明らかにするためなど）、処置に対する応答をモニターするため、所与の疾患、障害もしくは症状を発症する潜在的风险を評価するため、および／またはその疾患、障害もしくは症状の予後を評価するために、用いられる。本明細書中に開示される方法は、必要に応じて、特定の形態の癌を特徴付けるためにも用いられる。癌は、組成とステージ分類の両方が不均一であることが多いので、本明細書中に開示される方法を用いて生成されるデータは、癌の特定のサブタイプの特徴付けを可能にし、それによって、診断および処置の選択を支援し得る。この情報はまた、特定のタイプの癌の予後に関する手がかりを被験体または医療関係者に提供し得、被験体および／または医療関係者が疾患の進行に従って処置の選択肢を適応させることを可能にし得る。一部の癌は、進行するにつれて、侵襲的になり、遺伝的に不安定になる。他の腫瘍は、良性、不活性または休止状態のままである。

【0120】

サンプル

【0121】

サンプルは、被験体から単離された任意の生物学的サンプルであり得る。サンプルには、体組織、全血、血小板、血清、血漿、便、赤血球、白血球細胞または白血球、内皮細胞、組織生検材料（例えば、既知のまたは疑いのある固形腫瘍由来の生検材料）、脳脊髄液、滑液、リンパ液、腹水、間質液または細胞外液（例えば、細胞間隙由来の体液）、歯肉滲出液、歯肉溝液、骨髄、胸水、脳脊髄液、唾液、粘液、痰、精液、汗、尿が含まれ得る。サンプルは、好ましくは、体液、特に、血液およびその画分ならびに尿である。そのようなサンプルは、腫瘍から脱落した核酸を含む。それらの核酸は、DNA および RNA を含み得、二本鎖および一本鎖の形態であり得る。サンプルは、被験体から単離されたままの形態であり得るか、あるいは細胞などの構成要素を除去もしくは添加する、1つの構成要素を別の構成要素に対して濃縮する、または1つの形態の核酸を別の形態に（例えば、RNA を DNA に、または一本鎖核酸を二本鎖に）変換するさらなる処理に供されたものであり得る。したがって、例えば、解析用の体液サンプルは、セルフリー核酸、例えば、セルフリー DNA (cfDNA) を含む血漿または血清である。

【0122】

いくつかの実施形態において、被験体から採取される体液サンプルの体積は、シーケンシングされる領域に対する所望のリードデプスに依存する。例示的な体積は、約 0.4 ~ 40 ml、約 5 ~ 20 ml、約 10 ~ 20 ml である。例えば、その体積は、約 0.5 ml、約 1 ml、約 5 ml、約 10 ml、約 20 ml、約 30 ml、約 40 ml またはそれを超えるミリリットルであり得る。サンプリングされる血漿の体積は、通常、約 5 ml ~ 約 20 ml である。

【0123】

サンプルは、様々な量の核酸を含み得る。通常、所与のサンプル中の核酸の量は、複数のゲノム当量 (genome equivalents) と同等と見なされる。例えば、約 30 ng の DNA のサンプルは、約 10,000 (10^4) 半数体ヒトゲノム当量、cfDNA の場合は、約 2000 億 (2×10^{11}) 個の個々のポリヌクレオチド分子を含

み得る。同様に、約100 ngのDNAのサンプルは、約30,000半数体ヒトゲノム当量、cfDNAの場合は、約6000億個の個々の分子を含み得る。

【0124】

いくつかの実施形態において、サンプルは、種々の起源、例えば、細胞およびセルフリー起源（例えば、血液サンプルなど）に由来する核酸を含む。通常、サンプルは、変異を有する核酸を含む。例えば、サンプルは、必要に応じて、生殖細胞系列変異および／または体細胞変異を有するDNAを含む。通常、サンプルは、癌に関連する変異（例えば、癌に関連する体細胞変異）を有するDNAを含む。

【0125】

増幅前のサンプル中のセルフリー核酸の例示的な量は、通常、約1フェムトグラム（fg）～約1マイクログラム（μg）、例えば、約1ピコグラム（pg）～約200ナノグラム（ng）、約1 ng～約100 ng、約10 ng～約1000 ngの範囲である。いくつかの実施形態において、サンプルは、最大約600 ng、最大約500 ng、最大約400 ng、最大約300 ng、最大約200 ng、最大約100 ng、最大約50 ngまたは最大約20 ngのセルフリー核酸分子を含む。必要に応じて、その量は、少なくとも約1 fg、少なくとも約10 fg、少なくとも約100 fg、少なくとも約1 pg、少なくとも約10 pg、少なくとも約100 pg、少なくとも約1 ng、少なくとも約10 ng、少なくとも約100 ng、少なくとも約150 ngまたは少なくとも約200 ngのセルフリー核酸分子である。ある特定の実施形態において、その量は、最大約1 fg、約10 fg、約100 fg、約1 pg、約10 pg、約100 pg、約1 ng、約10 ng、約100 ng、約150 ngまたは約200 ngのセルフリー核酸分子である。いくつかの実施形態において、方法は、約1 fg～約200 ngのセルフリー核酸分子をサンプルから得る工程を含む。

【0126】

セルフリー核酸は、通常、約100ヌクレオチド長～約500ヌクレオチド長のサイズ分布を有し、約110ヌクレオチド長～約230ヌクレオチド長の分子が、サンプル中の分子の約90%を占め、約168ヌクレオチド長という最頻値、および約240～約440ヌクレオチド長の範囲に第2の小さなピークを有する。ある特定の実施形態において、セルフリー核酸は、約160～約180ヌクレオチド長または約320～約360ヌクレオチド長または約440～約480ヌクレオチド長である。

【0127】

いくつかの実施形態において、セルフリー核酸は、溶液中に見られるセルフリー核酸をインタクトな細胞および体液の他の不溶性の構成要素から分離する分割工程によって体液から単離される。これらの実施形態のうちのいくつかにおいて、分割は、遠心分離または濾過などの手法を含む。あるいは、体液中の細胞を溶解し、セルフリー核酸および細胞性核酸を共に処理する。一般に、緩衝液の添加および洗浄工程の後、例えばアルコールを用いて、セルフリー核酸を沈殿させる。ある特定の実施形態では、シリカベースのカラムなどのさらなるクリーンアップ工程を用いて、夾雑物または塩を除去する。その反応全体にわたって、例えば、非特異的なバルクのキャリア核酸を必要に応じて加えることにより、その例示的な手順のある特定の面（例えば、収率）が最適化される。そのような処理の後、サンプルは、通常、二本鎖DNA、一本鎖DNAおよび／または一本鎖RNAをはじめとした様々な形態の核酸を含む。必要に応じて、一本鎖DNAおよび／または一本鎖RNAを二本鎖の形態に変換し、それをその後の処理工程および解析工程に含める。

【0128】

核酸タグ

【0129】

ある特定の実施形態において、分子識別子またはバーコードを提供するタグが、とりわけ、化学合成、ライゲーションまたはオーバーラップ伸長PCRによって、アダプターに組み込まれるか、またはその他の方法で結合される。いくつかの実施形態において、反応におけるユニークなもしくはユニークでない識別子、または分子バーコードの割り当ては

、例えば、米国特許出願第20010053519号、同第20030152490号、同第20110160078号ならびに米国特許第6,582,908号、同第7,537,898号および同第9,598,731号（これらの各々が参照により援用される）に記載されている方法に従い、それらに記載されているシステムを利用する。

【0130】

タグは、ランダムにまたは非ランダムにサンプル核酸に連結される。いくつかの実施形態において、タグは、予想される、識別子（例えば、ユニークなバーコードおよび／またはユニークでないバーコードの組み合わせ）とマイクロウェルとの比で導入される。例えば、ゲノムサンプル1つあたり約1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000、50,000、100,000、500,000、1,000,000、10,000,000、50,000,000、100,000,000、500,000,000、1,000,000,000個を超える識別子がロードされるように、識別子はロードされ得る。いくつかの実施形態では、ゲノムサンプル1つあたり約2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000、50,000、100,000、500,000、1,000,000、10,000,000、50,000,000、100,000,000、500,000,000、1,000,000,000個未満の識別子がロードされるように、識別子はロードされる。ある特定の実施形態において、ゲノムサンプル1つあたりにロードされる識別子の平均数は、ゲノムサンプル1つあたり約1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000、50,000、100,000、500,000、1,000,000、10,000,000、50,000,000、100,000,000、500,000,000、1,000,000,000個未満、または約1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000、50,000、100,000、500,000、1,000,000、10,000,000、50,000,000、100,000,000、500,000,000、1,000,000,000個を超える識別子である。それらの識別子は、一般に、ユニークおよび／または非ユニークである。

【0131】

例示的な形式の1つは、標的核酸分子の両端にライゲートされた、約2～約1,000,000個の異なるタグまたは約5～約150個の異なるタグまたは約20～約50個の異なるタグを使用する。20～50×20～50個のタグの場合、合計400～2500個のタグが生成される。そのような数のタグは、通常、同じ開始点および終止点を有する異なる分子が、異なるタグの組み合わせを受け取る確率が高確率（例えば、少なくとも94%、99.5%、99.99%、99.999%）となるのに十分である。

【0132】

いくつかの実施形態において、識別子は、所定の、ランダムな、またはセミランダムな、配列オリゴヌクレオチドである。他の実施形態では、複数のバーコードがその複数の中で互いに必ずしもユニークでないように、複数のバーコードが用いられ得る。これらの実施形態では、バーコードは、そのバーコードとそれが付着し得る配列との組み合わせが、個別に追跡され得るユニークな配列を生成するように、一般に、個々の分子に付着される（例えば、ライゲーションまたはPCR増幅によって）。本明細書中に記載されるように、配列リードの始め（開始）および終わり（終止）の部分の配列データとともに、非ユニークにタグ化されたバーコードを検出することによって、通常、ユニークな同一性を特定の分子に割り当てることが可能になる。個々の配列リードの長さ、すなわち塩基対の数を必要に応じて用いることによって、ユニークな同一性が所与の分子に割り当てられる。本明細書中に記載されるように、それによって、ユニークな同一性を割り当てられた一本鎖の核酸由来のフラグメントが、その後の、親鎖由来のフラグメントおよび／または相補鎖の特定を可能にし得る。

【0133】

核酸増幅

【0134】

アダプターと隣接しているサンプル核酸は、通常、増幅されるDNA分子に隣接するアダプター内のプライマー結合部位に結合する核酸プライマーを用いて、PCRおよび他の増幅方法によって増幅される。いくつかの実施形態において、増幅方法は、サーモサイクリングによってもたらされる、伸長、変性およびアニーリングのサイクルを含むか、または例えば、転写媒介増幅におけるような等温性であり得る。必要に応じて利用される他の例示的な増幅方法としては、とりわけ、リガーゼ連鎖反応、鎖置換増幅、核酸配列に基づく増幅および自家持続配列に基づく複製が挙げられる。

【0135】

従来の核酸増幅方法を用いて核酸分子に分子タグおよび／またはサンプルインデックス／タグを導入するためには、通常、1回またはそれを超える回数の増幅サイクルが適用される。それらの増幅は、通常、1つまたはそれを超える反応混合物において行われる。分子タグおよびサンプルインデックス／タグは、必要に応じて、同時にまたは任意の連続した順序で導入される。いくつかの実施形態において、分子タグおよびサンプルインデックス／タグは、配列捕捉工程が行われる前および／または後に導入される。いくつかの実施形態では、分子タグだけが、プローブ捕捉の前に導入され、配列捕捉工程を行った後に、サンプルインデックス／タグが導入される。ある特定の実施形態では、分子タグとサンプルインデックス／タグの両方が、プローブに基づく捕捉工程を行う前に導入される。いくつかの実施形態において、サンプルインデックス／タグは、配列捕捉工程を行った後に導入される。通常、配列捕捉プロトコルは、標的化された核酸配列、例えば、あるゲノム領域のコード配列、およびある癌のタイプに関連するそのような領域の変異に相補的な一本鎖核酸分子を導入する工程を含む。通常、増幅反応は、約200ヌクレオチド(nt)～約700nt、250nt～約350ntまたは約320nt～約550ntの範囲のサイズの分子タグおよびサンプルインデックス／タグを有する、非ユニークにまたはユニークにタグ化された複数の核酸アンプリコンを生成する。いくつかの実施形態において、それらのアンプリコンは、約300ntのサイズを有する。いくつかの実施形態において、それらのアンプリコンは、約500ntのサイズを有する。

【0136】

核酸濃縮

【0137】

いくつかの実施形態において、核酸をシーケンシングする前に、配列が濃縮される。濃縮は、必要に応じて、特定の標的領域に対して行われるかまたは非特異的に行われる（「標的配列」）。いくつかの実施形態において、標的化された目的の領域は、差次的なタイリングおよび捕捉スキームを用いて、1つまたはそれを超えるベイトセットパネルに対して選択された核酸捕捉プローブ（「ベイト」）を用いて濃縮され得る。差次的なタイリングおよび捕捉スキームは、一般に、異なる相対的濃度のベイトセットを用いて、それらのベイトに関連するゲノム区画にわたって差次的に（例えば、異なる「解像度」で）タイリングし、制約セット（例えば、シーケンシングロード、各ベイトの有用性などの配列分析装置の制約）にかけ、下流のシーケンシングにとって所望のレベルで標的化核酸を捕捉する。これらの標的化された目的のゲノム区画は、必要に応じて、核酸構築物の天然または合成のヌクレオチド配列を含む。いくつかの実施形態において、1つまたはそれを超える目的の区画に対するプローブを有するビオチン標識ビーズを用いることにより、標的配列を捕捉し、必要に応じてその後、それらの区画を増幅して、目的の領域を濃縮することができる。

【0138】

配列捕捉は、通常、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を必要とする。ある特定の実施形態において、プローブセットストラテジーは、目的の区画にわたってプローブをタイリングすることを含む。そのようなプローブは、例えば、約60～約120ヌクレオチド長であり得る。そのセットは、約2x、3x、4x、5x、6x、8x、9x、10x、15x、20x、50xまたはそれを超える深さを有し得る。配列捕捉の有効性は、一般に、プローブの配列と相補的な（またはほぼ相補的な）

、標的分子における配列の長さに部分的に依存する。

【0139】

核酸シーケンシング

【0140】

必要に応じてアダプターと隣接している、サンプル核酸は一般に、事前増幅ありまたはなしで、シーケンシングに供される。必要に応じて利用されるシーケンシングの方法または商業的に利用可能な形式としては、例えば、Sangerシーケンシング、ハイスループットシーケンシング、バイサルファイトシーケンシング、パイロシーケンシング、合成によるシーケンシング、一分子シーケンシング、ナノポアに基づくシーケンシング、半導体シーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシング、RNA-Seq (Illumina)、Digital Gene Expression (Helicos)、次世代シーケンシング (NGS)、Single Molecule Sequencing by Synthesis (SMSS) (Helicos)、超並列シーケンシング、Clonal Single Molecule Array (Solexa)、ショットガンシーケンシング、Ion Torrent、Oxford Nanopore、Roche Genia、マクサム・ギルバートシーケンシング、プライマーウォーキング、PacBioを用いたシーケンシング、SOLiD、Ion Torrentまたはナノポアプラットフォームが挙げられる。シーケンシング反応は、複数のレーン、複数のチャネル、複数のウェル、または複数のサンプルセットを実質的に同時に処理する他の手段を備え得る種々のサンプル処理装置において行われ得る。サンプル処理装置は、複数のランの処理を同時に可能にする複数のサンプルチャンバーも備え得る。

10

20

【0141】

シーケンシング反応は、癌または他の疾患のマーカーを含むと知られている1つまたはそれを超える (one more) 核酸フラグメントタイプまたは区画に対して行われ得る。シーケンシング反応は、サンプル中に存在する任意の核酸フラグメントに対しても行われ得る。シーケンシング反応 (sequence reactions) は、ゲノムの少なくとも約5%、10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90%、95%、99%、99.9%または100%というゲノムの配列カバレッジを提供し得る。他の場合では、ゲノムの配列カバレッジは、ゲノムの約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。

30

【0142】

同時のシーケンシング反応が、マルチプレックスシーケンシング法を用いて行われ得る。いくつかの実施形態において、セルフリーポリヌクレオチドは、少なくとも約1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000または100,000シーケンシング反応によってシーケンシングされる。他の実施形態において、セルフリーポリヌクレオチドは、約1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000または100,000未満のシーケンシング反応によってシーケンシングされる。シーケンシング反応は、通常、連続的にまたは同時に行われる。その後のデータ解析は、一般に、シーケンシング反応の全部または一部に対して行われる。いくつかの実施形態において、データ解析は、少なくとも約1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000または100,000シーケンシング反応に対して行われる。他の実施形態において、データ解析は、約1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000または100,000未満のシーケンシング反応に対して行われ得る。例示的なリードデプスは、1遺伝子座 (塩基位置) あたり約1000~約50000リードである。

40

【0143】

50

いくつかの実施形態において、片端または両端に一本鎖オーバーハングを有する二本鎖核酸に平滑末端を酵素的に形成することによって、シーケンシング用の核酸集団が調製される。これらの実施形態において、その集団は、通常、ヌクレオチド（例えば、A、C、GおよびTまたはU）の存在下において、5' - 3' DNAポリメラーゼ活性および3' - 5' エキソヌクレアーゼ活性を有する酵素で処理される。例示的な酵素または必要に応じて使用されるその触媒フラグメントとしては、クレノウラージフラグメントおよびT4ポリメラーゼが挙げられる。5' オーバーハングでは、上記酵素は、通常、反対鎖の陥凹3'末端を5'末端と揃うまで伸長して、平滑末端を生成する。3' オーバーハングでは、上記酵素は一般に、その3'末端から反対鎖の5'末端まで、時折、反対鎖の5'末端を越えて消化する。この消化が、反対鎖の5'末端を越えて進む場合、5' オーバーハングに対して用いられる同じポリメラーゼ活性を有する酵素によって、ギャップが満たされ得る。二本鎖核酸に対する平滑末端の形成によって、例えば、アダプターの付着およびその後の増幅が容易になる。

10

【0144】

いくつかの実施形態において、核酸集団は、さらなる処理（例えば、一本鎖核酸から二本鎖への変換および／またはRNAからDNAへの変換）に供される。これらの形態の核酸はまた、必要に応じて、アダプターに連結され、増幅される。

【0145】

事前の増幅ありまたはなしで、上に記載された平滑末端を形成するプロセスに供された核酸、および必要に応じてサンプル中の他の核酸をシーケンシングして、シーケンシングされた核酸が生成され得る。シーケンシングされた核酸は、核酸の配列（すなわち、配列情報）または配列が決定された核酸のいずれかのことを指すことができる。シーケンシングは、サンプル中の個々の核酸分子の増幅産物のコンセンサス配列から、サンプル中の個々の核酸分子の配列データを直接または間接的に提供するために行われ得る。

20

【0146】

いくつかの実施形態において、平滑末端形成の後、サンプル中の一本鎖オーバーハングを有する二本鎖核酸は、両端において、バーコードを含むアダプターに連結され、シーケンシングによって、核酸配列、ならびにアダプターによって導入されたインラインバーコードが決定される。それらの平滑末端DNA分子は、必要に応じて、少なくとも部分的に二本鎖のアダプター（例えば、Y形またはベル形のアダプター）の平滑末端にライゲートされる。あるいは、サンプル核酸およびアダプターの平滑末端は、相補的なヌクレオチドでテイル化されて、ライゲーション（例えば、粘着末端ライゲーション）を容易にし得る。

30

【0147】

核酸サンプルは、通常、同じ核酸の任意の2コピーが、両端において連結されたアダプターから同じ組み合わせのアダプターバーコードを受け取る確率が低確率（例えば、<1または0.1%）になるように、十分な数のアダプターと接触される。このようにアダプターを使用することにより、参照核酸上に同じ開始点および終止点を有する核酸配列のファミリーの特定、および同じ組み合わせのバーコードに連結された核酸配列のファミリーの特定が可能となる。そのようなファミリーは、増幅前のサンプル中の核酸の増幅産物の配列に相当する。それらの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付着によって修飾された、元のサンプル中の核酸分子に対するコンセンサスヌクレオチドまたは完全なコンセンサス配列を導き出すためにコンパイルされ得る。換言すれば、そのサンプル中のある核酸の特定の位置を占めるヌクレオチドは、ファミリーメンバー配列内のその対応する位置を占めるヌクレオチドのコンセンサスであると判断される。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み得る。あるファミリーのメンバーが、二本鎖核酸の両方の鎖の配列を含む場合、一方の鎖の配列は、コンセンサスヌクレオチドまたはコンセンサス配列を導き出すためにすべての配列をコンパイルする目的で、それらの相補鎖に変換される。いくつかのファミリーは、ただ1つのメンバー配列だけを含む。この場合、この配列は、増幅前のサンプル中の核酸の配列とみなされ得る。あるいは、

40

50

ただ1つのメンバー配列だけを含むファミリーは、その後の解析から除外され得る。

【0148】

シーケンシングされた核酸におけるヌクレオチドバリエーションは、シーケンシングされた核酸を参照配列と比較することによって明らかにされ得る。参照配列は、既知の配列、例えば、被験体の既知のホールゲノム配列または部分ゲノム配列（例えば、ヒト被験体のホールゲノム配列）であることが多い。参照配列は、例えば、h G 1 9 または h G 3 8 であり得る。シーケンシングされた核酸は、上に記載されたような、サンプル中の核酸に対して直接決定された配列またはそのような核酸の増幅産物の配列のコンセンサスに相当し得る。比較は、参照配列上の1つまたはそれを超える指定の位置において行われ得る。参照配列の指定の位置と対応する位置を含むシーケンシングされた核酸のサブセットは、それぞれの配列が最大限にアラインメントされたとき、識別され得る。そのようなサブセット内において、どのシーケンシングされた核酸が、指定の位置にヌクレオチドバリエーションを含むのか（もしあれば）、その終点（すなわち、5' および 3' 末端のヌクレオチド）が参照配列にマッピングされる場所に基づく所与の c f D N A フラグメントの長さ、所与の c f D N A フラグメントにおけるゲノム領域の中間点からのその c f D N A フラグメントの中間点のオフセット、および必要に応じて、どれが参照ヌクレオチド（すなわち、参照配列中のものと同じヌクレオチド）を含むのか（もしあれば）を明らかにすることができる。ヌクレオチドバリエーションを含むサブセット内のシーケンシングされた核酸の数が、選択された閾値を超える場合、指定の位置においてバリエーションヌクレオチドがコールされ得る。その閾値は、単純な数（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個のシーケンシングされた核酸）であり得るか、または比（例えば、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シーケンシングされた核酸に対する少なくとも（a l e a s t）0.5、1、2、3、4、5、10、15または20）であり得る。比較は、参照配列における任意の目的の指定の位置に対して反復され得る。時折、参照配列上の少なくとも約20、100、200または300個連続した位置、例えば、約20～500または約50～300個連続した位置を占める指定の位置に対して、比較が行われ得る。

【0149】

本明細書中に記載される形式および用途を含む核酸シーケンシングに関するさらなる詳細は、例えば、Levyら、Annual Review of Genomics and Human Genetics, 17:95-115 (2016)、Liuら、J. of Biomedicine and Biotechnology, Volume 2012, Article ID 251364:1-11 (2012)、Voelkerdingら、Clinical Chem., 55:641-658 (2009)、MacLeanら、Nature Rev. Microbiol., 7:287-296 (2009)、Astierら、J Am Chem Soc., 128(5):1705-1710 (2006)、米国特許第6,210,891号、米国特許第6,258,568号、米国特許第6,833,246号、米国特許第7,115,400号、米国特許第6,969,488号、米国特許第5,912,148号、米国特許第6,130,073号、米国特許第7,169,560号、米国特許第7,282,337号、米国特許第7,482,120号、米国特許第7,501,245号、米国特許第6,818,395号、米国特許第6,911,345号、米国特許第7,501,245号、米国特許第7,329,492号、米国特許第7,170,050号、米国特許第7,302,146号、米国特許第7,313,308号および米国特許第7,476,503号（これらの各々は、その全体が参照により援用される）にも提供されている。

【0150】

データ解析

【0151】

いくつかの実施形態において、シーケンシングの生データは、様々なファイル形式（例えば、FASTQ、VCF、CRAMまたはBAM）で提供され得る配列リードセットを

含み得る。シーケンシングの生データを含むファイルは、ペアードエンドリードにおけるように、一方の鎖または両方の鎖に対する配列データを含み得る。一例において、シーケンシングの生データは、両方の鎖、すなわち、ペアードエンドシーケンシング手順から生成されたセンス鎖およびアンチセンス鎖に対するFASTQファイルとして提供される。それらのファイルは、リードの品質に関する情報を提供するさらなる符号を含み得、品質スコアも提供し得る。各ポリヌクレオチド分子のシーケンシングの生データは、ローカルドライブ、クラウドまたはサーバーに保存され得る。

【0152】

いくつかの場合では、シーケンシング反応によって生成された配列リードは、バイオインフォマティクス解析を行うために、参照配列とアラインメントされ得るか、または参照配列にマッピングされ得る。参照配列は、既知の配列、例えば、ある対象の既知のホールゲノム配列または部分ゲノム配列、ヒト被験体のホールゲノム配列であることが多い。参照配列は、hg19であり得る。シーケンシングされた核酸は、上に記載されたように、サンプル中の核酸に対して直接決定された配列またはそのような核酸の増幅産物の配列のコンセンサスに相当し得る。比較は、参照配列上の1つまたはそれを超える指定の位置において行われ得る。

【0153】

配列リードは、マッピングツールを用いて参照配列とアラインメントされ得、そのマッピングツールの非限定的な例としては、Burrow's Wheeler Transform (BWA)、NovoalignおよびBowtieが挙げられ得る。それらのマッピングツールは、使用されるアラインメントパラメータ、参照配列に対する配列リードの位置（例えば、座標）およびマッピングの品質スコアを記述しているアラインメントファイルを生成する。それらのアラインメントパラメータ（例えば、シーケンシングリードと参照配列との間で許容される差異の数、許容されるギャップの数、およびギャップオープニングペナルティ (gap opening penalty)、ギャップ伸長の数など）は、ユーザーによって定義され得る。ある場合では、デフォルトのアラインメントパラメータを用いるBWAマッピングツールを使用することにより、hg19などのヒト参照ゲノムとリードがアラインメントされる。BWAツールは、アラインメントの統計情報を含むBAMファイルである出力ファイルを提供する。アラインメントの統計情報には、処理されたリードがアラインメントされた参照配列の座標が含まれ得る。アラインメントの統計情報は、参照配列にマッピングされたときのリードのユニークさを知らせるMapQスコアも提供し得る。次いで、処理されたリードは、分子バーコードおよび参照配列上の座標を用いて、ソーティングされ得る。

【0154】

参照配列の指定の位置と対応する位置を含むシーケンシングされた核酸のサブセットは、それぞれの配列が最大限にアラインメントされたとき、識別され得る。そのようなサブセット内において、どのシーケンシングされた核酸が、指定の位置にヌクレオチドバリエーションを含むのか（もしあれば）、および必要に応じて、どれが参照ヌクレオチド（すなわち、参照配列中のものと同じヌクレオチド）を含むのか（もしあれば）を明らかにすることができる。比較は、参照配列における任意の目的の指定の位置に対して反復され得る。時折、参照配列上の少なくとも20、100、200または300個連続した位置、例えば、20～500または50～300個連続した位置を占める指定の位置に対して、比較が行われ得る。

【0155】

サンプルは、同じ核酸の任意の2コピーが、片端または両端において連結されたアダプターから、分子バーコードを含む同じ組み合わせのアダプターを受け取る確率が低確率（例えば、<1または0.1%）になるように、十分な数の異なる分子バーコードと接触され得る。このようにアダプターを使用することにより、参照核酸とアラインメントされた（またはマッピングされた）同じ開始点および終止点を有する配列リード、ならびに同じ組み合わせの分子バーコードに連結された配列リードを、同じ元の分子から生成されたり

ードのファミリーにグループ分けすることが可能となり得る。そのようなファミリーは、増幅前のサンプル中の核酸の増幅産物の配列に相当し得る。

【0156】

ファミリーメンバーの配列は、平滑末端化およびアダプター付着によって修飾された、元のサンプル中の核酸分子に対するコンセンサヌクレオチドまたは完全なコンセンサス配列を導き出すためにコンパイルされ得る。換言すれば、そのサンプル中のある核酸の特定の位置を占めるヌクレオチドは、ファミリーメンバー配列内のその対応する位置を占めるヌクレオチドのコンセンサスであると判断され得る。コンセンサスヌクレオチドは、2つの非限定的な例示的方法を挙げるとすると投票または信頼スコアなどの方法によって決定され得る。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み得る。あるファミリーのメンバーが、二本鎖核酸の両方の鎖の配列を含む場合、一方の鎖の配列は、コンセンサスヌクレオチドまたはコンセンサス配列を導き出すためにすべての配列をコンパイルする目的で、それらの相補鎖に変換される。いくつかのファミリーは、ただ1つのメンバー配列だけを含み得る。この場合、この配列は、増幅前のサンプル中の核酸の配列とみなされ得る。あるいは、ただ1つのメンバー配列だけを含むファミリーは、その後の解析から除外され得る。

10

【0157】

いくつかの実施形態において、本明細書中に開示されるシステムおよび方法の結果は、レポートを生成するための入力として用いられる。そのレポートは、紙の形式であり得る。例えば、レポートは、生物学的サンプル中の療法的な核酸構築物の有無を指摘し得る。いくつかの実施形態において、そのレポートは、生物学的サンプル中の療法的な核酸構築物のレベルの指摘を含み得る。

20

【0158】

本明細書中に開示される方法の様々な工程、または本明細書中に開示されるシステムによって行われる工程は、同時にもしくは異なる時点において、同じもしくは異なる地理的位置、例えば、同じ国もしくは異なる国において、および／または同じもしくは異なる人々によって、行われ得る。

【0159】

シーケンシングパネル

【0160】

変異を示す腫瘍を検出する可能性を改善するために、シーケンシングされるDNAの領域は、遺伝子またはゲノム領域のパネルを含み得る。シーケンシングのために、限られた領域（例えば、限られたパネル）を選択することにより、必要な総シーケンシング（例えば、シーケンシングされるヌクレオチドの総量）を減少させることができる。シーケンシングパネルが、複数の異なる遺伝子または領域を標的化することにより、単一の癌、一連の癌またはすべての癌を検出することができる。あるいは、ホールゲノムシーケンシング（WGS）によって、またはシーケンシングパネルを用いない他の偏りのないシーケンシング方法によって、DNAがシーケンシングされ得る。

30

【0161】

いくつかの態様において、複数の異なる遺伝子またはゲノム領域を標的化するパネルは、所定の比率の、癌を有する被験体が、そのパネル内の1つまたはそれを超える異なる遺伝子において遺伝的バリエーションまたは腫瘍マーカーを示すように、選択される。そのパネルは、シーケンシング用の領域を固定数の塩基対に限定するように選択され得る。そのパネルは、所望の量のDNAをシーケンシングするように選択され得る。そのパネルは、ある量のシーケンシングされた塩基対に対して所望の配列リードデプスまたは配列リードカバレッジを達成するように、選択され得る。そのパネルは、サンプル中の1つまたはそれを超える遺伝的バリエーションを検出するための理論的感度、理論的特異度および／または理論的精度を達成するように、選択され得る。

40

【0162】

50

領域のパネルを検出するためのプローブには、目的のゲノム領域（ホットスポット領域）を検出するためのプローブ、ならびにヌクレオソーム認識プローブ（例えば、K R A S コドン 1 2 および 1 3）が含まれ得、それらのプローブは、ヌクレオソーム結合パターンおよび G C 配列組成によって影響される c f D N A カバレッジおよびフラグメントサイズバリエーションの解析に基づいて捕捉を最適化するようにデザインされ得る。本明細書中で使用される領域は、ヌクレオソーム位置および G C モデルに基づいて最適化された非ホットスポット領域も含み得る。そのパネルは、複数のサブパネル（起源組織を識別するためのサブパネル（例えば、組織にわたって最も多様な転写プロファイルを有する遺伝子（必ずしもプロモーターではない）に相当する 5 0 ～ 1 0 0 個のベイトを定義するために既刊の刊行物を使用すること）、ホールゲノムスクャフォールドを識別するためのサブパネル（例えば、超保存的なゲノム含有量を識別するため、およびコピー数塩基をライニングする目的のために、染色体にわたってほんの一握りのプローブで低密度にタイリングするため）、例えば癌抑制遺伝子（例えば、結腸直腸癌における S E P T 9 / V I M）のプロモーターにおける、転写開始部位（T S S）／C p G アイランド（例えば、差次的にメチル化された領域（例えば、差次的メチル化領域（DMR））を捕捉するため）を識別するためのサブパネルを含む）を含み得る。いくつかの実施形態において、起始組織についてのマーカーは、組織特異的エピジェネティックマーカーである。

【0163】

目的のゲノム位置の一覧のいくつかの例が表 1 および表 2 に見られ得る。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 5、少なくとも 1 0、少なくとも 1 5、少なくとも 2 0、少なくとも 2 5、少なくとも 3 0、少なくとも 3 5、少なくとも 4 0、少なくとも 4 5、少なくとも 5 0、少なくとも 5 5、少なくとも 6 0、少なくとも 6 5、少なくとも 7 0、少なくとも 7 5、少なくとも 8 0、少なくとも 8 5、少なくとも 9 0、少なくとも 9 5 または 9 7 個の、表 1 の遺伝子の少なくとも一部を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 5、少なくとも 1 0、少なくとも 1 5、少なくとも 2 0、少なくとも 2 5、少なくとも 3 0、少なくとも 3 5、少なくとも 4 0、少なくとも 4 5、少なくとも 5 0、少なくとも 5 5、少なくとも 6 0、少なくとも 6 5 または 7 0 個の、表 1 の S N V を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 1、少なくとも 2、少なくとも 3、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 1 0、少なくとも 1 1、少なくとも 1 2、少なくとも 1 3、少なくとも 1 4、少なくとも 1 5、少なくとも 1 6、少なくとも 1 7 または 1 8 個の、表 1 の C N V を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 1、少なくとも 2、少なくとも 3、少なくとも 4、少なくとも 5 または 6 つの、表 1 の融合物を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 1 つ、少なくとも 2 つまたは 3 つの、表 1 のインデルの少なくとも一部を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 5、少なくとも 1 0、少なくとも 1 5、少なくとも 2 0、少なくとも 2 5、少なくとも 3 0、少なくとも 3 5、少なくとも 4 0、少なくとも 4 5、少なくとも 5 0、少なくとも 5 5、少なくとも 6 0、少なくとも 6 5、少なくとも 7 0、少なくとも 7 5、少なくとも 8 0、少なくとも 8 5、少なくとも 9 0、少なくとも 9 5、少なくとも 1 0 0、少なくとも 1 0 5、少なくとも 1 1 0 または 1 1 5 個の、表 2 の遺伝子の少なくとも一部を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 5、少なくとも 1 0、少なくとも 1 5、少なくとも 2 0、少なくとも 2 5、少なくとも 3 0、少なくとも 3 5、少なくとも 4 0、少なくとも 4 5、少なくとも 5 0、少なくとも 5 5、少なくとも 6 0、少なくとも 6 5、少なくとも 7 0 または 7 3 個の、表 2 の S N V を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも 1、少なくとも 2、少なくとも 3、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 1 0、少なくとも 1 1、少なくとも 1 2、少なくとも 1 3

10

20

30

40

50

、少なくとも１４、少なくとも１５、少なくとも１６、少なくとも１７または１８個の、表２のＣＮＶを含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも１、少なくとも２、少なくとも３、少なくとも４、少なくとも５つまたは６つの、表２の融合物を含む。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも１、少なくとも２、少なくとも３、少なくとも４、少なくとも５、少なくとも６、少なくとも７、少なくとも８、少なくとも９、少なくとも１０、少なくとも１１、少なくとも１２、少なくとも１３、少なくとも１４、少なくとも１５、少なくとも１６、少なくとも１７または１８個の、表２のインデルの少なくとも一部を含む。これらの目的のゲノム位置の各々は、所与のバインドセットパネルに対する骨格領域またはホットスポット領域と識別され得る。目的のホットスポットゲノム位置の一覧の例は、表３に見られ得る。いくつかの実施形態において、本開示の方法において用いられるゲノム位置は、少なくとも１、少なくとも２、少なくとも３、少なくとも４、少なくとも５、少なくとも６、少なくとも７、少なくとも８、少なくとも９、少なくとも１０、少なくとも１１、少なくとも１２、少なくとも１３、少なくとも１４、少なくとも１５、少なくとも１６、少なくとも１７、少なくとも１８、少なくとも１９または少なくとも２０個の、表３の遺伝子の少なくとも一部を含む。各ホットスポットゲノム位置は、関連する遺伝子、それが存在する染色体、その遺伝子の遺伝子座に相当するゲノムの開始位置および終止位置、塩基対を単位とするその遺伝子の遺伝子座の長さ、その遺伝子によってカバーされるエキソン、および目的の所与のゲノム位置が捕捉しようとし得る重要な特徴（例えば、変異のタイプ）をはじめとしたいくつかの特色とともに列挙される。

【表１－１】

表 1

点変異 (SNVs)						増幅物 (CNVs)		融合物	インデル
AKT1	ALK	APC	AR	ARAF	ARID1A	AR	BRAF	ALK	EGFR (エキソン 19 & 20)
ATM	BRAF	BRCA 1	BRCA2	CCND1	CCND2	CCND 1	CCND2	FGFR2	
CCNE1	CDH1	CDK4	CDK6	CDKN2 A	CDKN2 B	CCNE 1	CDK4	FGFR3	
CTNNB 1	EGFR	ERBB 2	ESR1	EZH2	FBXW7	CDK6	EGFR	NTRK1	ERBB2 (エキソン 19 & 20)
FGFR1	FGFR 2	FGFR 3	GATA3	GNA11	GNAQ	ERBB2	FGFR1	RET	
GNAS	HNF1 A	HRAS	IDH1	IDH2	JAK2	FGFR2	KIT	ROS1	
JAK3	KIT	KRAS	MAP2K 1	MAP2K 2	MET	KRAS	MET		MET (エキソン 14)
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH 1	MYC	PDGFR A		

【表１－２】

NPM1	NRAS	NTRK 1	PDGFR A	PIK3CA	PTEN	PIK3C A	RAF1		スキッピング)
PTPN1 1	RAF1	RB1	RET	RHEB	RHOA				
RIT1	ROS1	SMAD 4	SMO	SRC	STK11				
TERT	TP53	TSC1	VHL						

【表 2】

表 2

点変異 (SNVs)						増幅物 (CNVs)		融合物	インデル
AKT1	ALK	APC	AR	ARAF	ARID1A	AR	BRAF	ALK	EGFR (エクソン 19 & 20)
ATM	BRAF	BRCA 1	BRCA2	CCND1	CCND2	CCND 1	CCND2	FGFR2	
CCNE1	CDH1	CDK4	CDK6	CDKN2 A	DDR2	CCNE 1	CDK4	FGFR3	
CTNNB 1	EGFR	ERBB 2	ESR1	EZH2	FBXW7	CDK6	EGFR	NTRK1	ERBB2 (エクソン 19 & 20)
FGFR1	FGFR 2	FGFR 3	GATA3	GNA11	GNAQ	ERBB2	FGFR1	RET	
GNAS	HNF1 A	HRAS	IDH1	IDH2	JAK2	FGFR2	KIT	ROS1	
JAK3	KIT	KRAS	MAP2K 1	MAP2K 2	MET	KRAS	MET		MET (エクソン 14 スキッピング)
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH 1	MYC	PDGFR A		
NPM1	NRAS	NTRK 1	PDGFR A	PIK3CA	PTEN	PIK3C A	RAF1		
PTPN1 1	RAF1	RB1	RET	RHEB	RHOA				ATM
RIT1	ROS1	SMAD 4	SMO	MAPK1	STK11				
TERT	TP53	TSC1	VHL	MAPK3	MTOR				
NTRK3									

【表 3 - 1】

表 3

遺伝子	染色体	開始位置	終止位置	長さ (bp)	カバー される エキソン	重要な特徴
ALK	chr2	29446405	29446655	250	イントロン 19	融合
ALK	chr2	29446062	29446197	135	イントロン 20	融合
ALK	chr2	29446198	29446404	206	20	融合
ALK	chr2	29447353	29447473	120	イントロン 19	融合
ALK	chr2	29447614	29448316	702	イントロン 19	融合
ALK	chr2	29448317	29448441	124	19	融合
ALK	chr2	29449366	29449777	411	イントロン 18	融合
ALK	chr2	29449778	29449950	172	18	融合
BRAF	chr7	14045306 4	14045320 3	139	15	BRAF V600
CTNNB1	chr3	41266007	41266254	247	3	S37
EGFR	chr7	55240528	55240827	299	18 および 19	G719 および 欠失
EGFR	chr7	55241603	55241746	143	20	挿入 / T790 M
EGFR	chr7	55242404	55242523	119	21	L858R
ERBB2	chr17	37880952	37881174	222	20	挿入
ESR1	chr6	15241985 7	15242011 1	254	10	V534, P535, L536, Y537, D538
FGFR2	chr10	12327948 2	12327969 3	211	6	S252
GATA3	chr10	8111426	8111571	145	5	SS / インデル
GATA3	chr10	8115692	8116002	310	6	SS / インデル
GNAS	chr20	57484395	57484488	93	8	R844
IDH1	chr2	20911308 3	20911339 4	311	4	R132
IDH2	chr15	90631809	90631989	180	4	R140, R172
KIT	chr4	55524171	55524258	87	1	
KIT	chr4	55561667	55561957	290	2	
KIT	chr4	55564439	55564741	302	3	

10

20

30

40

【表 3 - 2】

KIT	chr4	55565785	55565942	157	4	
KIT	chr4	55569879	55570068	189	5	
KIT	chr4	55573253	55573463	210	6	
KIT	chr4	55575579	55575719	140	7	
KIT	chr4	55589739	55589874	135	8	
KIT	chr4	55592012	55592226	214	9	
KIT	chr4	55593373	55593718	345	10 および 11	557, 559, 560, 576
KIT	chr4	55593978	55594297	319	12 および 13	V654
KIT	chr4	55595490	55595661	171	14	T670, S709
KIT	chr4	55597483	55597595	112	15	D716
KIT	chr4	55598026	55598174	148	16	L783
KIT	chr4	55599225	55599368	143	17	C809, R815, D816, L818, D820, S821F, N822, Y823
KIT	chr4	55602653	55602785	132	18	A829P
KIT	chr4	55602876	55602996	120	19	
KIT	chr4	55603330	55603456	126	20	
KIT	chr4	55604584	55604733	149	21	
KRAS	chr12	25378537	25378717	180	4	A146
KRAS	chr12	25380157	25380356	199	3	Q61
KRAS	chr12	25398197	25398328	131	2	G12/G13
MET	chr7	11641153 5	11641225 5	720	13, 14, イントロン 13, イントロン 14	METエキソン14 SS
NRAS	chr1	11525641 0	11525660 9	199	3	Q61
NRAS	chr1	11525866 0	11525879 1	131	2	G12/G13
PIK3CA	chr3	17893598 7	17893613 2	145	10	E545K
PIK3CA	chr3	17895187 1	17895216 2	291	21	H1047R
PTEN	chr10	89692759	89693018	259	5	R130

10

20

30

40

【表 3 - 3】

SMAD4	chr18	48604616	48604849	233	12	D537
TERT	chr5	1294841	1295512	671	プロモーター	chr5:1295228
TP53	chr17	7573916	7574043	127	11	Q331, R337, R342
TP53	chr17	7577008	7577165	157	8	R273
TP53	chr17	7577488	7577618	130	7	R248
TP53	chr17	7578127	7578299	172	6	R213/Y220
TP53	chr17	7578360	7578564	204	5	R175 / 欠失
TP53	chr17	7579301	7579600	299	4	
				12574 (全標的領域)		
				16330 (全プローブカバレッジ)		

10

20

【0164】

いくつかの実施形態において、上記パネル内の1つまたはそれを超える領域は、手術後に残留している癌を検出するための、1つまたは複数の遺伝子の1つまたはそれを超える遺伝子座を含む。この検出は、既存の癌検出方法よりも早く検出できる。いくつかの実施形態において、上記パネル内の1つまたはそれを超えるゲノム位置は、高リスク患者集団において癌を検出するための、1つまたは複数の遺伝子の1つまたはそれを超える遺伝子座を含む。例えば、喫煙者は、一般集団よりもかなり高い肺癌の割合を有する。さらに、喫煙者は、癌検出を難しくする他の肺の症状（例えば、肺における不規則小結節の発症）を発症し得る。いくつかの実施形態において、本明細書中に記載される方法は、既存の癌検出方法より早く、高リスク患者において癌を検出する。

30

【0165】

遺伝子または領域に腫瘍マーカーを有する癌を有するいくつかの被験体に基づいて、あるゲノム位置が、シーケンシングパネルに含めるために選択され得る。癌を有する被験体の有病率およびその遺伝子に存在する腫瘍マーカーに基づいて、あるゲノム位置が、シーケンシングパネルに含めるために選択され得る。ある領域に腫瘍マーカーが存在することは、被験体が癌を有することを示し得る。

【0166】

いくつかの場合では、上記パネルは、1つまたはそれを超えるデータベースからの情報を用いて選択され得る。癌に関する情報は、癌腫瘍生検またはcfDNAアッセイから得られ得る。データベースは、シーケンシングされた腫瘍サンプルの集団を説明する情報を含み得る。データベースは、腫瘍サンプルにおけるmRNA発現に関する情報を含み得る。データベース(databased)は、腫瘍サンプルにおける調節エレメントまたはゲノム領域に関する情報を含み得る。シーケンシングされた腫瘍サンプルに関係する情報は、様々な遺伝的バリエーションの頻度を含み得、それらの遺伝的バリエーションが存在する遺伝子または領域を説明し得る。それらの遺伝的バリエーションは、腫瘍マーカーであり得る。そのようなデータベースの非限定的な例は、COSMICである。COSMICは、様々な癌に見られる体細胞変異のカatalogである。特定の癌について、COSMICは、変異の

40

50

頻度に基づいて遺伝子を格付けしている。ある遺伝子が、所与の遺伝子内に高頻度の変異を有することによって、パネルに含めるために選択され得る。例えば、COSMICは、シーケンシングされた乳癌サンプルの集団の33%が、TP53に変異を有し、サンプリングされた乳癌の集団の22%が、KRASに変異を有することを示唆している。APCをはじめとした他の格付けされた遺伝子は、シーケンシングされた乳癌サンプルの集団の約4%にしか見られない変異を有する。TP53およびKRASは、サンプリングされた乳癌のうち比較的高頻度（例えば約4%の頻度で存在するAPCと比べて）を有することに基づいて、シーケンシングパネルに含められ得る。COSMICを、非限定的な例として提供するが、しかしながら、ある遺伝子または遺伝子領域に位置する腫瘍マーカーと癌を関連付ける任意のデータベースまたは情報セットを使用してもよい。別の例では、COSMICが提供するように、1156個の胆管癌サンプルのうち、380個のサンプル（33%）が、TP53に変異を有した。いくつかの他の遺伝子（例えば、APC）が、全サンプルの4～8%に変異を有する。したがって、TP53は、胆管癌サンプルの集団における比較的高頻度に基づいて、パネルに含めるために選択され得る。

【0167】

サンプリングされた腫瘍組織または循環腫瘍DNAにおける腫瘍マーカーの頻度が、所与のバックグラウンド集団で見られる頻度よりも有意に高い場合、あるパネルのために、ある遺伝子またはゲノム区画が選択され得る。癌を有する被験体の少なくとも大部分が、パネル内のゲノム位置または遺伝子の少なくとも1つに存在する腫瘍マーカーまたはゲノム領域を有し得るように、あるゲノム位置の組み合わせが、あるパネルに含めるために選択され得る。そのゲノム位置の組み合わせは、特定の癌または癌のセットについて、被験体の大部分が、1つまたはそれを超えるその選択された領域に1つまたはそれを超える腫瘍マーカーを有することを示すデータに基づいて選択され得る。例えば、癌1を検出するために、癌1を有する被験体の90%が、パネルの領域A、B、Cおよび／またはDに腫瘍マーカーを有することを示すデータに基づいて、領域A、B、Cおよび／またはDを含むパネルが選択され得る。あるいは、腫瘍マーカーは、癌を有する被験体における2つまたはそれを超える領域に独立して存在すると示され得、それらの2つまたはそれを超える領域における腫瘍マーカーは、合わせると、癌を有する被験体集団の大部分に存在する。例えば、癌2を検出するために、被験体の90%が、1つまたはそれを超える領域に腫瘍マーカーを有し、そのような被験体の30%において、腫瘍マーカーが領域Xでのみ検出されるが、腫瘍マーカーが検出された残りの被験体では、腫瘍マーカーは領域Yおよび／またはZでのみ検出されることを示すデータに基づいて、領域X、YおよびZを含むパネルが選択され得る。1つまたはそれを超える癌と関連すると以前に示された1つまたはそれを超えるゲノム位置に存在する腫瘍マーカーは、1つまたはそれを超えるそれらの領域において、ある腫瘍マーカーが、50%またはそれを超えて検出された場合、被験体が癌を有することを示し得るかまたは予測し得る。コンピュータによるアプローチ（例えば、1つまたはそれを超える領域内の腫瘍マーカーセットに対する癌頻度が与えられた場合に癌を検出する条件付き確率を用いるモデル）を用いることにより、どの領域が、単独でまたは組み合わせられて、癌を予測し得るかが予測され得る。パネル選択のための他のアプローチは、大きなパネルを用いた腫瘍の網羅的なゲノムプロファイリングおよび／またはホールゲノムシーケンシング（WGS、RNA-seq、ChIP-seq、バイサルフェート（bisulfate）シーケンシング、ATAC-seqなど）を用いた研究からの情報を説明するデータベースの使用を含む。文献から収集された情報も、ある特定の癌においてよく影響を受けるおよびよく変異する経路を説明し得る。パネルの選択は、遺伝情報を説明するオントロジーを使用することによって、さらに特徴付けられ得る。

【0168】

シーケンシングのために上記パネルに含められる遺伝子は、完全に転写される領域、プロモーター領域、エンハンサー領域、調節エレメントおよび／または下流の配列を含み得る。変異を示す腫瘍を検出する可能性をさらに高めるために、エキソンだけが、上記パネルに含められ得る。上記パネルは、選択された遺伝子の全エキソン、または選択された遺

10

20

30

40

50

伝子のエキソンのうちのただ1つもしくはそれを超えるエキソンを含み得る。上記パネルは、複数の異なる各遺伝子のエキソンを含み得る。上記パネルは、複数の異なる各遺伝子の少なくとも1つのエキソンを含み得る。

【0169】

いくつかの態様において、複数の異なる各遺伝子のエキソンのパネルは、所定の比率の、癌を有する被験体が、エキソンのパネル内の少なくとも1つのエキソンに遺伝的バリエーションを示すように選択される。

【0170】

遺伝子のパネル内の異なる各遺伝子の少なくとも1つの完全長エキソンが、シーケンシングされ得る。シーケンシングされたパネルは、複数の遺伝子のエキソンを含み得る。そのパネルは、2～100個の異なる遺伝子、2～70個の遺伝子、2～50個の遺伝子、2～30個の遺伝子、2～15個の遺伝子または2～10個の遺伝子のエキソンを含み得る。

10

【0171】

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

20

【0172】

上記パネルは、複数の異なる遺伝子の1つまたはそれを超えるエキソンを含み得る。上記パネルは、ある割合のその複数の異なる遺伝子の各々の1つまたはそれを超えるエキソンを含み得る。上記パネルは、それらの異なる遺伝子の少なくとも25%、50%、75%または90%の各々の少なくとも2つのエキソンを含み得る。上記パネルは、それらの異なる遺伝子の少なくとも25%、50%、75%または90%の各々の少なくとも3つのエキソンを含み得る。上記パネルは、それらの異なる遺伝子の少なくとも25%、50%、75%または90%の各々の少なくとも4つのエキソンを含み得る。

30

【0173】

シーケンシングパネルのサイズは、様々であり得る。シーケンシングパネルは、いくつかの因子（例えばシーケンシングされたヌクレオチドの総量またはパネル内の特定の領域についてシーケンシングされたユニークな分子の数）に応じて、（ヌクレオチドのサイズに関して）大きくまたは小さくしてよい。シーケンシングパネルは、5kb～50kbのサイズにされ得る。シーケンシングパネルは、10kb～30kbのサイズであり得る。シーケンシングパネルは、12kb～20kbのサイズであり得る。シーケンシングパネルは、12kb～60kbのサイズであり得る。シーケンシングパネルは、少なくとも10kb、12kb、15kb、20kb、25kb、30kb、35kb、40kb、45kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kbまたは150kbのサイズであり得る。シーケンシングパネルは、100kb、90kb、80kb、70kb、60kbまたは50kb未満のサイズであり得る。

40

【0174】

シーケンシングのために選択されたパネルは、少なくとも1、5、10、15、20、25、30、40、50、60、80または100個のゲノム位置（例えば、各々が目的

50

のゲノム領域を含む)を含み得る。いくつかの場合では、パネル内のゲノム位置は、それらの位置のサイズが比較的小さくなるように選択される。いくつかの場合では、パネル内の領域は、約10 kbもしくはそれ未満、約8 kbもしくはそれ未満、約6 kbもしくはそれ未満、約5 kbもしくはそれ未満、約4 kbもしくはそれ未満、約3 kbもしくはそれ未満、約2.5 kbもしくはそれ未満、約2 kbもしくはそれ未満、約1.5 kbもしくはそれ未満または約1 kbもしくはそれ未満もしくはそれ未満のサイズを有する。いくつかの場合では、パネル内のゲノム位置は、約0.5 kb～約10 kb、約0.5 kb～約6 kb、約1 kb～約11 kb、約1 kb～約15 kb、約1 kb～約20 kb、約0.1 kb～約10 kbまたは約0.2 kb～約1 kbのサイズを有する。例えば、パネル内の領域は、約0.1 kb～約5 kbのサイズを有し得る。

10

【0175】

本明細書中で選択されるパネルは、低頻度遺伝的バリエーションを検出する(例えば、サンプルから得られたセルフリー核酸分子において)のに十分なディープシーケンシングを可能にし得る。サンプル中の遺伝的バリエーションの量は、所与の遺伝的バリエーションに対するマイナーアレル頻度に関して言及され得る。マイナーアレル頻度とは、マイナーアレル(例えば、最も一般的でないアレル)が所与の核酸集団(例えば、サンプル)中に存在する頻度のことを指し得る。低マイナーアレル頻度の遺伝的バリエーションは、サンプル中において比較的低い存在頻度を有し得る。いくつかの場合では、上記パネルは、少なくとも0.0001%、0.0001%、0.0005%、0.001%、0.005%、0.01%または0.05%のマイナーアレル頻度の遺伝的バリエーションの検出を可能にする。上記パネルは、0.0001%またはそれを超えるマイナーアレル頻度の遺伝的バリエーションの検出を可能にし得る。上記パネルは、0.001%またはそれを超えるマイナーアレル頻度の遺伝的バリエーションの検出を可能にし得る。上記パネルは、0.00001%、0.0001%、0.0005%、0.001%、0.0025%、0.005%、0.0075%、0.01%、0.025%、0.05%、0.075%または1.0%もの低い頻度でサンプル中に存在する遺伝的バリエーションの検出を可能にし得る。上記パネルは、少なくとも0.00001%、0.0001%、0.0005%、0.001%、0.0025%、0.005%、0.0075%、0.01%、0.025%、0.05%、0.075%または1.0%の頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、1.0%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.75%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.5%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.25%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.1%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.075%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.05%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.025%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.01%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.005%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、0.0001%もの低い頻度でサンプル中に存在する腫瘍マーカーの検出を可能にし得る。上記パネルは、1.0%～0.00001%もの低い頻度でサンプル中に存在する、シーケンシングされたcfDNAにおける腫瘍マーカーの検出を可能にし得る。上記パネルは、0.01%～0.00001%もの低い頻度でサンプル中に存在する、シーケンシングされたcfDNAにおける腫瘍マーカーの検出を可能にし得る。

20

30

40

【0176】

遺伝的バリエーションは、疾患(例えば、癌)を有する被験体集団のパーセンテージを単位として提示され得る。いくつかの場合では、癌を有する集団の少なくとも1%、2%、3%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%

50

、95%または99%が、上記パネル内の領域のうちの少なくとも1つにおいて1つまたはそれを超える遺伝的バリエーションを示す。例えば、癌を有する集団の少なくとも80%が、上記パネル内のゲノム位置のうちの少なくとも1つにおいて1つまたはそれを超える遺伝的バリエーションを示し得る。

【0177】

上記パネルは、1つまたはそれを超える各遺伝子由来の目的のゲノム領域を含む1つまたはそれを超える位置を含み得る。いくつかの場合では、上記パネルは、少なくとも1、2、3、4、5、6、7、8、9、10、15、20、25、30、40、50または80個の各遺伝子由来の目的のゲノム領域を含む1つまたはそれを超える位置を含み得る。いくつかの場合では、上記パネルは、多くとも1、2、3、4、5、6、7、8、9、10、15、20、25、30、40、50または80個の各遺伝子由来の目的のゲノム領域を含む1つまたはそれを超える位置を含み得る。いくつかの場合では、上記パネルは、約1～約80、1～約50、約3～約40、5～約30、10～約20個の異なる各遺伝子由来の目的のゲノム領域を含む1つまたはそれを超える位置を含み得る。

【0178】

上記パネル内の領域は、それらが、1つまたはそれを超える組織にわたって差次的に転写される配列を含むように選択され得る。いくつかの場合では、ゲノム領域を含む位置は、他の組織よりも高いレベルである特定の組織において転写される配列を含み得る。例えば、ゲノム領域を含む位置は、ある特定の組織では転写されるが他の組織では転写されない配列を含み得る。

【0179】

上記パネル内のゲノム位置は、コード配列および／または非コード配列を含み得る。例えば、上記パネル内のゲノム位置は、エキソン、イントロン、プロモーター、3'非翻訳領域、5'非翻訳領域、調節エレメント、転写開始部位および／またはスプライス部位における1つまたはそれを超える配列を含み得る。いくつかの場合では、上記パネル内の領域は、偽遺伝子、反復配列、トランスポゾン、ウイルスエレメントおよびテロメアをはじめとした他の非コード配列を含み得る。いくつかの場合では、上記パネル内のゲノム位置は、非コードRNA、例えば、リボソームRNA、転移RNA、P i w i と相互作用するRNAおよびマイクロRNAにおける配列を含み得る。

【0180】

上記パネル内のゲノム位置は、所望の感度レベルで（例えば、1つまたはそれを超える遺伝的バリエーションの検出によって）癌を検出（診断）するように、選択され得る。例えば、上記パネル内の領域は、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の感度で癌を検出する（例えば、1つまたはそれを超える遺伝的バリエーションの検出によって）ように選択され得る。上記パネル内のゲノム位置は、100%の感度で癌を検出するように選択され得る。

【0181】

上記パネル内のゲノム位置は、所望の特異度レベルで（例えば、1つまたはそれを超える遺伝的バリエーションの検出によって）癌を検出（診断）するように、選択され得る。例えば、上記パネル内のゲノム位置は、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の特異度で癌を検出する（例えば、1つまたはそれを超える遺伝的バリエーションの検出によって）ように選択され得る。上記パネル内のゲノム位置は、100%の特異度で1つまたはそれを超える遺伝的バリエーションを検出するように選択され得る。

【0182】

上記パネル内のゲノム位置は、所望の陽性予測値で癌を検出（診断）するように、選択され得る。陽性予測値は、感度（例えば、実際の陽性が検出される確率）および／または特異度（例えば、実際の陰性を陽性と間違えない確率）を高めることによって上昇し得る。非限定的な例として、上記パネル内のゲノム位置は、少なくとも50%、55%、60

%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の陽性予測値で1つまたはそれを超える遺伝的バリエーションを検出するように、選択され得る。上記パネル内の領域は、100%の陽性予測値で1つまたはそれを超える遺伝的バリエーションを検出するように、選択され得る。

【0183】

上記パネル内のゲノム位置は、所望の精度で癌を検出（診断）するように選択され得る。本明細書中で使用されるとき、用語「精度」とは、試験が疾患症状（例えば、癌）と健康症状とを判別する能力のことを指し得る。精度は、尺度（例えば、感度および特異度、予測値、尤度比、ROC曲線下面積、Youden指数ならびに／または診断オッズ比）を用いて定量され得る。10

【0184】

精度は、正しい結果を与えた試験の数と行われた試験の総数との比のことを指すパーセンテージとして提示され得る。上記パネル内の領域は、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の精度で癌を検出するように選択され得る。上記パネル内のゲノム位置は、100%の精度で癌を検出するように選択され得る。

【0185】

パネルは、高感度であるように、かつ低頻度の遺伝的バリエーションを検出するように、選択され得る。例えば、パネルは、0.01%、0.05%または0.001%もの低い頻度でサンプル中に存在する遺伝的バリエーションまたは腫瘍マーカーが、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の感度で検出され得るように、選択され得る。上記パネル内のゲノム位置は、サンプル中に1%またはそれ未満の頻度で存在する腫瘍マーカーを70%またはそれを超える感度で検出するように選択され得る。パネルは、サンプル中に0.1%もの低い頻度で存在する腫瘍マーカーを少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の感度で検出するように選択され得る。パネルは、サンプル中に0.01%もの低い頻度で存在する腫瘍マーカーを少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の感度で検出するように選択され得る。20
30

【0186】

パネルは、高度に特異的であるように、かつ低頻度の遺伝的バリエーションを検出するように、選択され得る。例えば、パネルは、0.01%、0.05%または0.001%もの低い頻度でサンプル中に存在する遺伝的バリエーションまたは腫瘍マーカーが、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の特異度で検出され得るように、選択され得る。上記パネル内のゲノム位置は、サンプル中に1%またはそれ未満の頻度で存在する腫瘍マーカーを70%またはそれを超える特異度で検出するように選択され得る。パネルは、サンプル中に0.1%もの低い頻度で存在する腫瘍マーカーを少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の特異度で検出するように選択され得る。パネルは、サンプル中に0.01%もの低い頻度で存在する腫瘍マーカーを少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の特異度で検出するように選択され得る。40
50

異度で検出するように選択され得る。

【0187】

パネルは、高精度であるように、かつ低頻度の遺伝的バリエーションを検出するように、選択され得る。パネルは、0.01%、0.05%または0.001%もの低い頻度でサンプル中に存在する遺伝的バリエーションまたは腫瘍マーカーが、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の精度で検出され得るように、選択され得る。上記パネル内のゲノム位置は、サンプル中に1%またはそれ未満の頻度で存在する腫瘍マーカーを70%またはそれを超える精度で検出するように選択され得る。パネルは、サンプル中に0.1%もの低い頻度で存在する腫瘍マーカーを少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の精度で検出するように選択され得る。パネルは、サンプル中に0.01%もの低い頻度で存在する腫瘍マーカーを少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の精度で検出するように選択され得る。パネルは、サンプル中に0.001%もの低い頻度で存在する腫瘍マーカーを少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の精度で検出するように選択され得る。

10

【0188】

パネルは、高度に予測的であるように、かつ低頻度の遺伝的バリエーションを検出するように、選択され得る。パネルは、0.01%、0.05%または0.001%もの低い頻度でサンプル中に存在する遺伝的バリエーションまたは腫瘍マーカーが、少なくとも70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%または99.9%の陽性予測値を有し得るように、選択され得る。

20

【0189】

上記パネルにおいて使用されるプローブまたはベイトの濃度は、サンプル内のより多くの核酸分子を捕捉するために、高められ得る(2~6 ng/μL)。上記パネルにおいて使用されるプローブまたはベイトの濃度は、少なくとも2 ng/μL、3 ng/μL、4 ng/μL、5 ng/μL、6 ng/μLまたはそれを超える濃度であり得る。プローブの濃度は、約2 ng/μL~約3 ng/μL、約2 ng/μL~約4 ng/μL、約2 ng/μL~約5 ng/μL、約2 ng/μL~約6 ng/μLであり得る。上記パネルにおいて使用されるプローブまたはベイトの濃度は、2 ng/μLまたはそれを超える濃度から6 ng/μLまたはそれ未満の濃度までであり得る。いくつかの場合では、これによって、生物製剤内のより多くの分子が解析されることが可能となり、それによって、低頻度のアレルを検出することが可能になり得る。

30

【0190】

癌および他の疾患

【0191】

ある特定の実施形態において、本明細書中に開示される方法および態様は、患者における所与の疾患、障害または症状を診断するために用いられる。通常、検討中の疾患は、あるタイプの癌である。そのような癌の非限定的な例としては、胆管癌、膀胱癌、移行上皮癌腫、尿路上皮癌腫、脳腫瘍、グリオーマ、星状細胞腫、乳癌腫、化生性癌腫、子宮頸癌、子宮頸部扁平上皮癌腫、直腸癌、結腸直腸癌腫、結腸癌、遺伝性非ポリポーシス結腸直腸癌、結腸直腸腺癌、消化管間質腫瘍(GIST)、子宮内膜癌腫、子宮内膜間質肉腫、食道癌、食道扁平上皮癌腫、食道腺癌、眼メラノーマ、ブドウ膜黒色腫、胆嚢癌腫、胆嚢腺癌、腎細胞癌腫、腎明細胞癌腫、移行上皮癌腫、尿路上皮癌腫、ウィルムス腫瘍、白血病、急性リンパ性白血病(ALL)、急性骨髄性白血病(AML)、慢性リンパ性(CLL)、慢性骨髄性(CML)、慢性骨髄単球性(CMML)、肝臓癌、肝臓癌腫、ヘパトーマ、肝細胞癌腫、胆管癌腫、肝芽腫、肺癌、非小細胞肺癌(NSCLC)、中皮腫、B細胞リンパ腫、非ホジキンリンパ腫、びまん性大細胞型B細胞型リンパ腫、マントル細胞リンパ腫、T細胞リンパ腫、非ホジキンリンパ腫、前駆Tリンパ芽球性リンパ腫/白血病

40

50

、末梢性T細胞リンパ腫、多発性骨髄腫、鼻咽頭癌腫（NPC）、神経芽細胞腫、中咽頭癌、口腔扁平上皮癌腫、骨肉腫、卵巣癌腫、膵癌、膵管腺癌、偽乳頭状腫瘍、腺房細胞癌腫、前立腺癌、前立腺腺癌、皮膚癌、メラノーマ、悪性黒色腫、皮膚黒色腫、小腸癌腫、胃癌、胃癌腫、消化管間質腫瘍（GIST）、子宮癌または子宮肉腫が挙げられる。

【0192】

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

【0193】

カスタマイズされた療法および関連する投与

【0194】

いくつかの実施形態において、本明細書中に開示される方法は、所与の疾患、障害または症状を有する患者に対する療法の特定およびその投与に関する。本質的に任意の癌療法（例えば、外科的療法、放射線療法、化学療法など）が、これらの方法の一部として含まれる。通常、療法には、少なくとも1つの免疫療法（または免疫療法薬）が含まれる。免疫療法とは、一般に、所与の癌のタイプに対する免疫応答を高める方法のことを指す。ある特定の実施形態において、免疫療法とは、腫瘍または癌に対するT細胞応答を高める方法のことを指す。

【0195】

いくつかの実施形態において、免疫療法または免疫療法薬は、免疫チェックポイント分子を標的化する。ある特定の腫瘍は、免疫チェックポイント経路を利用することによって免疫系を免れることができる。したがって、免疫チェックポイントの標的化は、腫瘍が免疫系を免れる能力に対抗するためおよびある特定の癌に対する抗腫瘍免疫を活性化するための効果的なアプローチとして現れた。Pardoll, Nature Reviews Cancer, 2012, 12: 252-264。

【0196】

ある特定の実施形態において、免疫チェックポイント分子は、抗原に対するT細胞応答に関わるシグナルを減じる阻害性分子である。例えば、CTLA4は、T細胞上に発現されており、抗原提示細胞上のCD80（別名、B7.1）またはCD86（別名、B7.2）に結合することによってT細胞活性化のダウンレギュレーションにおいて役割を果たす。PD-1は、T細胞上に発現されるもう1つの阻害性チェックポイント分子である。PD-1は、炎症反応中の末梢組織においてT細胞の活性を制限する。さらに、PD-1に対するリガンド（PD-L1またはPD-L2）は、通常、多くの種々の腫瘍の表面上でアップレギュレートされ、その結果、腫瘍微小環境において抗腫瘍免疫応答がダウンレギュレーションする。ある特定の実施形態において、阻害性免疫チェックポイント分子は、CTLA4またはPD-1である。他の実施形態において、阻害性免疫チェックポイント分子は、PD-L1またはPD-L2などの、PD-1に対するリガンドである。他の実施形態において、阻害性免疫チェックポイント分子は、CD80またはCD86などの、CTLA4に対するリガンドである。他の実施形態において、阻害性免疫チェックポイント分子は、リンパ球活性化遺伝子3（LAG3）、キラー細胞免疫グロブリン様レセプ

10

20

30

40

50

ター（K I R）、T細胞膜タンパク質3（T I M 3）、ガレクチン9（G A L 9）またはアデノシンA 2 aレセプター（A 2 a R）である。

【0197】

これらの免疫チェックポイント分子を標的化するアンタゴニストを用いることにより、ある特定の癌に対する抗原特異的T細胞応答を高めることができる。したがって、ある特定の実施形態において、免疫療法または免疫療法薬は、阻害性免疫チェックポイント分子のアンタゴニストである。ある特定の実施形態において、阻害性免疫チェックポイント分子は、PD-1である。ある特定の実施形態において、阻害性免疫チェックポイント分子は、PD-L1である。ある特定の実施形態において、阻害性免疫チェックポイント分子のアンタゴニストは、抗体（例えば、モノクローナル抗体）である。ある特定の実施形態において、その抗体またはモノクローナル抗体は、抗CTLA4、抗PD-1、抗PD-L1または抗PD-L2抗体である。ある特定の実施形態において、その抗体は、モノクローナル抗PD-1抗体である。いくつかの実施形態において、その抗体は、モノクローナル抗PD-L1抗体である。ある特定の実施形態において、そのモノクローナル抗体は、抗CTLA4抗体と抗PD-1抗体との組み合わせ、抗CTLA4抗体と抗PD-L1抗体との組み合わせ、または抗PD-L1抗体と抗PD-1抗体との組み合わせである。ある特定の実施形態において、抗PD-1抗体は、ペンブロリズマブ（Keytruda（登録商標））またはニボルマブ（Opdivo（登録商標））のうちの1つまたはそれを超えるものである。ある特定の実施形態において、抗CTLA4抗体は、イピリムマブ（Yervoy（登録商標））である。ある特定の実施形態において、抗PD-L1抗体は、アテゾリズマブ（Tecentriq（登録商標））、アベルマブ（Bavencio（登録商標））またはデュルバルマブ（Imfinzi（登録商標））のうちの1つまたはそれを超えるものである。

【0198】

ある特定の実施形態において、免疫療法または免疫療法薬は、CD80、CD86、LAG3、K I R、T I M 3、G A L 9またはA 2 a Rに対するアンタゴニスト（例えば、抗体）である。他の実施形態において、そのアンタゴニストは、阻害性免疫チェックポイント分子の可溶性バージョン（例えば、阻害性免疫チェックポイント分子の細胞外ドメインおよび抗体のFcドメインを含む可溶性融合タンパク質）である。ある特定の実施形態において、その可溶性融合タンパク質は、CTLA4、PD-1、PD-L1またはPD-L2の細胞外ドメインを含む。いくつかの実施形態において、その可溶性融合タンパク質は、CD80、CD86、LAG3、K I R、T I M 3、G A L 9またはA 2 a Rの細胞外ドメインを含む。1つの実施形態において、その可溶性融合タンパク質は、PD-L2またはLAG3の細胞外ドメインを含む。

【0199】

ある特定の実施形態において、免疫チェックポイント分子は、ある抗原に対するT細胞応答に関わるシグナルを増幅する共刺激分子である。例えば、CD28は、T細胞上に発現される共刺激レセプターである。T細胞が、そのT細胞レセプターを介して抗原に結合するとき、CD28は、抗原提示細胞上のCD80（別名、B7.1）またはCD86（別名、B7.2）に結合して、T細胞レセプターシグナル伝達を増幅し、T細胞活性化を促進する。CD28は、CTLA4と同じリガンド（CD80およびCD86）に結合するので、CTLA4は、CD28によって媒介される共刺激シグナル伝達を相殺または制御することができる。ある特定の実施形態において、免疫チェックポイント分子は、CD28、誘導性T細胞共刺激物質（ICOS）、CD137、OX40またはCD27から選択される共刺激分子である。他の実施形態において、免疫チェックポイント分子は、共刺激分子のリガンドであり、例えば、CD80、CD86、B7RP1、B7-H3、B7-H4、CD137L、OX40LまたはCD70が挙げられる。

【0200】

これらの共刺激チェックポイント分子を標的化するアゴニストを用いることにより、ある特定の癌に対する抗原特異的T細胞応答を高めることができる。したがって、ある特定

の実施形態において、免疫療法または免疫療法薬は、共刺激チェックポイント分子のアゴニストである。ある特定の実施形態において、その共刺激チェックポイント分子のアゴニストは、アゴニスト抗体であり、好ましくは、モノクローナル抗体である。ある特定の実施形態において、そのアゴニスト抗体またはモノクローナル抗体は、抗CD28抗体である。他の実施形態において、そのアゴニスト抗体またはモノクローナル抗体は、抗ICOS、抗CD137、抗OX40または抗CD27抗体である。他の実施形態において、そのアゴニスト抗体またはモノクローナル抗体は、抗CD80、抗CD86、抗B7RP1、抗B7-H3、抗B7-H4、抗CD137L、抗OX40Lまたは抗CD70抗体である。

【0201】

癌以外の、遺伝子に基づく特定の疾患、障害または症状を処置するための療法の選択肢は、当業者に広く周知されており、検討中の特定の疾患、障害または症状を考慮すれば明らかである。

【0202】

ある特定の実施形態において、本明細書中に記載されるカスタマイズされた療法は、通常、非経口的に（例えば、静脈内にまたは皮下に）施される。免疫療法薬を含む薬学的組成物は、通常、静脈内に投与される。ある特定の療法薬は、経口的に投与される。しかしながら、カスタマイズされた療法（例えば、免疫療法薬など）は、当該分野で公知の任意の方法によっても投与され得、その方法としては、例えば、頬側、舌下、直腸、膣、尿道内、局所的、眼内、鼻腔内および／または耳介内が挙げられ、その投与には、錠剤、カプセル剤、顆粒剤、水性懸濁剤、ゲル、スプレー、坐剤、軟膏剤（salves）、軟膏剤（ointments）などが含まれる。

【0203】

システムおよびコンピュータ可読媒体

【0204】

本開示は、様々なシステムおよびコンピュータプログラム製品または機械可読媒体も提供する。いくつかの実施形態において、例えば、本明細書中に記載される方法は、必要に応じて、システム、分散コンピューティングハードウェアおよびアプリケーション（例えば、クラウドコンピューティングサービス）、電子通信ネットワーク、通信用インターフェース、コンピュータプログラム製品、機械可読媒体、電子記憶媒体、ソフトウェア（例えば、機械が実行可能なコードまたは論理命令）などを少なくとも部分的に用いて、行われるかまたは促進される。例えば、図7は、本願に開示されている方法の態様を少なくとも実行して使用するのに適した例示的なシステムの模式図を提供している。示されているように、システム700は、少なくとも1つのコントローラまたはコンピュータ、例えば、プロセッサ704およびメモリ、ストレージデバイスまたはメモリコンポーネント706を備えるサーバー702（例えば、検索エンジンサーバー）、ならびにリモートサーバー702から離れて位置するおよび電子通信ネットワーク712（例えば、インターネットまたは他のインターネットワーク）を介してリモートサーバー702と通信している、1つまたはそれを超える他のコミュニケーションデバイス714および716（例えば、クライアント側コンピュータ端末、電話、タブレット、ラップトップ、他のモバイルデバイスなど）を備える。コミュニケーションデバイス714および716は、通常、例えばサーバー702コンピュータとネットワーク712を介して通信する電子ディスプレイ（例えば、インターネット対応コンピュータなど）を含み、ここで、その電子ディスプレイは、本明細書中に記載される方法を実行したときの結果を表示するための、ユーザーインターフェース（例えば、グラフィカルユーザーインターフェース（GUI）、ウェブベースのユーザーインターフェースなど）を備える。ある特定の実施形態において、通信ネットワークは、例えば、ハードドライブ、サムドライブまたは他のデータストレージメカニズムを用いて、1つの場所から別の場所にデータを物理的に移動させることも包含する。システム700は、例えば、ガイド付き検索アプリケーションまたは1つもしくはそれを超える他のコミュニケーションデバイス（例えば、714（デスクトップまたはパ

10

20

30

40

50

パーソナルコンピュータとして模式的に示されている) および 716 (タブレットコンピュータとして模式的に示されている)) によって実行可能な他のアプリケーションに使用する、コンピュータ可読媒体または機械可読媒体 (例えば、サーバー 702 によって可読な 1 つまたはそれを超える様々なタイプのメモリ、例えば、サーバー 702 のメモリ 706) 上に保存されたプログラム製品 708 も備える。いくつかの実施形態において、システム 700 は、必要に応じて、少なくとも 1 つのデータベースサーバー (例えば、直接または検索エンジンサーバー 702 を介して検索可能なデータ (例えば、分類器スコア、コントロールサンプルの結果または比較対照の結果のデータ、インデックスされたカスタマイズされた療法など) を保存しているオンラインウェブサイトと関連付けられたサーバー 710) も備える。システム 700 は、必要に応じて、サーバー 702 から離れて位置する 1 つまたはそれを超える他のサーバーも備え、それらの各サーバーは、必要に応じて、その他の各サーバーと離れてまたは近くに配置されている 1 つまたはそれを超えるデータベースサーバー 710 と関連付けられている。その他のサーバーは、地理的に離れたユーザーにサービスを有益に提供でき、地理的に分散した操作を強化することができる。

【0205】

当業者が理解するように、サーバー 702 のメモリ 706 は、必要に応じて、例えばとりわけ、RAM、ROM および磁気ディスクまたは光ディスクを含む、揮発性および／または不揮発性のメモリを含む。単一のサーバーとして図示されているが、図示されているサーバー 702 の配置は、単に例として与えられているのであって、他の様々な方法または構造に従って構成された他のタイプのサーバーまたはコンピュータも使用できることも当業者は理解する。図 7 に模式的に示されているサーバー 702 は、サーバーまたはサーバークラスターまたはサーバーファームであり、任意の個々の物理的サーバーに限定されない。サーバーの場所は、サーバーホスティングプロバイダーが管理するサーバーファームまたはサーバークラスターとして配備され得る。サーバーならびにそれらの構造および配置の数は、システム 700 に対する使用法、要求および容量の要件に基づいて増加させてよい。また、当業者が理解するように、これらの実施形態における他のユーザーコミュニケーションデバイス 714 および 716 は、例えば、ラップトップ、デスクトップ、タブレット、携帯情報端末 (PDA)、携帯電話、サーバーまたは他のタイプのコンピュータであり得る。当業者が承知および理解するように、ネットワーク 712 は、通信ネットワークおよび／またはローカルエリアネットワークもしくは他のエリアネットワークの一部を介して 1 つまたはそれを超える他のコンピュータと通信している複数のコンピュータ／サーバーの、インターネット、イントラネット、テレコミュニケーションネットワーク、エクストラネットまたはワールドワイドウェブを含み得る。

【0206】

当業者がさらに理解するように、例示的なプログラム製品または機械可読媒体 708 は、必要に応じて、ハードウェアの機能を制御し、その操作を指示する 1 セットまたはそれを超える順序づけられた操作を提供する、マイクロコード、プログラム、クラウドコンピューティング形式、ルーチンおよび／または記号言語の形である。プログラム製品 708 もまた、例示的な実施形態によると、その全体が揮発性メモリに存在する必要はないが、当業者が承知および理解するように、様々な方法に従って、必要に応じて選択的にロードされ得る。

【0207】

当業者にさらに理解されるように、用語「コンピュータ可読媒体」または「機械可読媒体」とは、実行のためにプロセッサに命令を提供することに関与する任意の媒体のことを指す。例えば、用語「コンピュータ可読媒体」または「機械可読媒体」は、配布媒体、クラウドコンピューティング形式、中間記憶媒体、コンピュータの実行メモリ、および例えばコンピュータによる読み取りのための、本開示の様々な実施形態の機能またはプロセスを実行するプログラム製品 708 を保存することができる他の任意の媒体またはデバイスを包含する。「コンピュータ可読媒体」または「機械可読媒体」は、不揮発性媒体、揮発性媒体および伝送媒体を含むがこれらに限定されない多くの形を取り得る。不揮発性媒

体としては、例えば、光ディスクまたは磁気ディスクが挙げられる。揮発性媒体としては、所与のシステムのメインメモリなどのダイナミックメモリが挙げられる。伝送媒体としては、バスを構成するワイヤを含む、同軸ケーブル、銅線および光ファイバーが挙げられる。伝送媒体は、音波または光波（例えば、とりわけ、電波通信および赤外線通信中に発生するもの）の形も取り得る。コンピュータ可読媒体の例示的な形態としては、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、フラッシュドライブもしくは他の任意の磁気媒体、CD-ROM、他の任意の光学媒体、パンチカード、紙テープ、穴のパターンを有する他の任意の物理的媒体、RAM、PROMおよびEPROM、FLASH（登録商標）-EPROM、他の任意のメモリチップもしくはカートリッジ、搬送波、またはコンピュータが読める他の任意の媒体が挙げられる。

10

【0208】

プログラム製品708は、必要に応じて、コンピュータ可読媒体からハードディスクまたは類似の中間記録媒体にコピーされる。プログラム製品708またはその一部が実行されるとき、それは、必要に応じて、それらの配布媒体、中間記録媒体などから、1つまたはそれを超えるコンピュータの実行メモリにロードされて、それらのコンピュータが、様々な実施形態の機能または方法に従って働くように構成される。そのような操作のすべてが、例えばコンピュータシステムの分野の当業者に周知である。

【0209】

さらなる例証として、ある特定の実施形態において、本願は、1つまたはそれを超えるプロセッサ、およびそのプロセッサと通信する1つまたはそれを超えるメモリコンポーネントを備えるシステムを提供する。そのメモリコンポーネントは、通常、実行されたとき、プロセッサに、配列情報、サブクロナリティスコア、分類器スコア、試験結果、コントロールまたは比較対照の結果、カスタマイズされた療法などを（例えば、コミュニケーションデバイス714、716などを介して）表示させる情報を提供させる、ならびに／または他のシステムコンポーネントおよび／もしくはシステムユーザーから（例えば、コミュニケーションデバイス714、716などを介して）情報を受け取らせる、1つまたはそれを超える命令を含む。

20

【0210】

いくつかの実施形態において、プログラム製品708は、コンピュータが実行可能な非一時的な命令を含み、その命令は、電子プロセッサ704によって実行されたとき、少なくとも、（a）1つまたはそれを超える参照サンプル由来のセルフリー核酸（cfNA）フラグメントから得られた配列リードを含む配列情報から、分類アレルセット内の各アレルに対するサブクロナリティスコアを生成することであって、ここで、各分類アレルは、潜在的に臨床的に有意であり、参照サンプル中の所与の遺伝子座に観察されるマイナーアレルを含む、こと、およびb）少なくとも1つの選択されたカットオフ閾値をサブクロナリティスコアと比較することを行い、ここで、選択されたカットオフ閾値を超えるサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、非標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレルは、非標的核酸バリエーションフィルタリストに加えられ、かつ／またはそのカットオフ閾値未満のサブクロナリティスコアを有する分類アレルは、それらの分類アレルが、標的細胞を起源とする参照cfNAフラグメント由来であることを示し、それらの分類アレルは、標的核酸バリエーションフィルタリストに加えられる。さらなるコンピュータ可読媒体の実施形態が本明細書中に記載される。

30

40

【0211】

システム700は、通常、本明細書中に記載される方法の様々な態様を行うように構成されているさらなるシステムコンポーネントも備える。これらの実施形態のうちのいくつかにおいて、これらのさらなるシステムコンポーネントのうちの1つまたはそれを超えるものが、リモートサーバー702から離れて、かつ電子通信ネットワーク712を介してリモートサーバー702と通信した状態で、配置されるのに対して、他の実施形態では、これらのさらなるシステムコンポーネントのうちの1つまたはそれを超えるものが、ロー

50

カルに、かつサーバー 702 と通信した状態で（すなわち、電子通信ネットワーク 712 の非存在下において）配置されるか、または例えばデスクトップコンピュータ 714 と直接、配置される。

【0212】

いくつかの実施形態において、例えば、さらなるシステムコンポーネントとしては、コントローラ 702 に作動可能に接続された（直接または間接的に（例えば、電子通信ネットワーク 712 を介して））サンプル調製コンポーネント 718 が挙げられる。サンプル調製コンポーネント 718 は、核酸増幅コンポーネント（例えば、サーマルサイクラーなど）および／または核酸配列決定装置によって増幅および／またはシーケンシングされる、サンプル中の核酸を調製する（例えば、核酸のライブラリーを調製する）ように構成されている。これらのうちのある特定の実施形態において、サンプル調製コンポーネント 718 は、サンプル中の他の構成要素から核酸を単離するように、バーコードを含む 1 つまたは複数のアダプターを本明細書中に記載されるような核酸に付着させるように、シーケンシングの前にゲノムまたはトランスクリプトームから 1 つまたはそれを超える領域を選択的に濃縮するように、など、構成されている。

10

【0213】

ある特定の実施形態において、システム 700 は、コントローラ 702 に作動可能に接続された（直接または間接的に（例えば、電子通信ネットワーク 712 を介して））核酸増幅コンポーネント 720（例えば、サーマルサイクラーなど）も備える。核酸増幅コンポーネント 720 は、被験体由来のサンプル中の核酸を増幅するように構成されている。例えば、核酸増幅コンポーネント 720 は、必要に応じて、本明細書中に記載されるようなサンプル中のゲノムまたはトランスクリプトームから、選択的に濃縮された領域を増幅するように構成されている。

20

【0214】

システム 700 はまた、通常、コントローラ 702 に作動可能に接続された（直接または間接的に（例えば、電子通信ネットワーク 712 を介して））少なくとも 1 つの核酸配列決定装置 722 を備える。核酸配列決定装置 722 は、被験体由来のサンプル中の核酸（例えば、増幅核酸）から配列情報を提供するように構成されている。本質的に任意のタイプの核酸配列決定装置が、これらのシステムにおいて使用するために適合され得る。例えば、核酸配列決定装置 722 は、必要に応じて、核酸に対してバイサルファイトシーケンシング、パイロシーケンシング、一分子シーケンシング、ナノポアシーケンシング、半導体シーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシングまたは他の手法を行って、シーケンシングリードを生成するように構成されている。必要に応じて、核酸配列決定装置 722 は、配列リードを配列リードのファミリーにグループ分けするように構成されており、その各ファミリーは、所与のサンプル中の核酸から生成された配列リードを含む。いくつかの実施形態において、核酸配列決定装置 722 は、シーケンシングライブラリーから得られたクロオナル一分子アレイを用いて、シーケンシングリードを生成する。ある特定の実施形態において、核酸配列決定装置 722 は、シーケンシングライブラリーをシーケンシングしてシーケンシングリードを生成するためのマイクロウェルのアレイを有する少なくとも 1 つのチップを備える。

30

40

【0215】

システムの完全なまたは部分的な自動化を促進するために、システム 700 は、通常、コントローラ 702 に作動可能に接続された（直接または間接的に（例えば、電子通信ネットワーク 712 を介して））材料移動コンポーネント 724 も備える。材料移動コンポーネント 724 は、核酸配列決定装置 722、サンプル調製コンポーネント 718 および核酸増幅コンポーネント 720 に、ならびに／またはそれらから、1 つまたはそれを超える材料（例えば、核酸サンプル、アンプリコン、試薬など）を移動させるように構成されている。

【0216】

コンピュータシステムおよびネットワーク、データベースならびにコンピュータプログ

50

ラム製品に関するさらなる詳細は、例えば、Peterson, Computer Networks: A Systems Approach, Morgan Kaufmann, 5th Ed. (2011)、Kurose, Computer Networking: A Top-Down Approach, Pearson, 7th Ed. (2016)、Elmasri, Fundamentals of Database Systems, Addison Wesley, 6th Ed. (2010)、Coronel, Database Systems: Design, Implementation, & Management, Cengage Learning, 11th Ed. (2014)、Tucker, Programming Languages, McGraw-Hill Science/Engineering/Math, 2nd Ed. (2006) および Rhoton, Cloud Computing Architected: Solution Design Handbook, Recursive Press (2011) (これらの各々は、その全体が参照により援用される) にも提供されている。

10

【0217】

【実施例】

【0218】

実施例1：循環腫瘍セルフリーDNA

【0219】

手術後の結腸直腸癌(CRC)患者におけるCtDNA(循環腫瘍セルフリーDNA)は、分子残存疾患(molecular residual disease)と相関し、予後判定に有用であり得、また、補助療法の意思決定を導くのに有用であり得る。

20

【0220】

手術後のctDNAは、根治的手術を受けた転移性CRCを有する患者において疾患の再発と強く関連する($p=0.004$)。Overmanら(2017)を参照のこと。高感度パネルを利用して、循環腫瘍DNA(ctDNA)を見出すことにより、肝切除術後の微小残存病変を検出することおよび疾患の再発を予測することができる。JCO35(suppl)。初期の研究では、個々の患者特異的腫瘍組織由来変異にインデックスされた臨床的に無理のあるアッセイが用いられたか、またはクローン性造血に関係するバリエーションを含む非腫瘍関連の体細胞性変化によって混乱が生じた。Tie J.ら(2016). Sci Transl Med 8(346)を参照のこと。

30

【0221】

高感度のCRC次世代シーケンシング(NGS)パネルを用いることによって、手術後のctDNAの検出には、既知の体細胞性変化の予知が必要ないことを示すデータが提供される。手術後のCRC患者におけるctDNA検出の特異度を高めることを目標に、バリエーション分類器を用いて、腫瘍由来の変化と非腫瘍由来の変化をさらに区別した(図8A~Cにおける分類器フィルターを参照のこと)。

【0222】

肝臓の転移巣切除術を予定されていたCRC患者を、IRB承認済みの治験に前向きに登録した。手術前および手術後の血漿を、CRCに対する理論的感度が96%の38遺伝子NGSパネルを用いて高デプスまでシーケンシングした。事前のctDNA結果と事後のctDNA結果の両方を有する51人の転移性結腸直腸癌患者を単一施設で募集した(表4: コホートの人口統計)。このパネルまたはローカルテストを用いて、腫瘍組織をシーケンシングした。17700 CRC pts (Guardant Health, Redwood City, CA)からのctDNAプロファイルを用いて、非腫瘍由来の変化を除外するようにバリエーション分類器を訓練した。その分類器は、腫瘍を起源とするcfDNA変異を識別するようにデザインされた。

40

【表 4】

表4

ユニークな患者の数	51
診断時の年齢の中央値(範囲)	55歳;範囲(33~76)
性別	60. 8%男性 39. 2%女性
組織学的グレード	98%中分化 2%低分化
原発部位	21. 6%右側 78. 4%左側
発症	15. 7%異時性 84. 3%同時性
ネオアジュバント化学療法	80. 4%
摘出腫瘍数の中央値	2
リンパ節陽性原発	66. 7%
KRAS変異	43%
手術から手術後のサンプリングまでの時間の中央値	18日(13~123日)
追跡調査の中央値	42. 7ヶ月(範囲4. 4. ~59. 4ヶ月)
再発	72. 5%(37人)
再発までの時間(中央値)	7. 8ヶ月(範囲1. 2~34. 5ヶ月)

10

20

【0223】

手術後の体細胞性バリエーション検出のみを用いた再発予測は、高い臨床上の偽陽性率を伴う。非腫瘍起源の変異の多くが、低アレル頻度で生じる。しかしながら、アレル頻度に対する単純な閾値は、多くの臨床的に関連性のある変異を除外し得る。腫瘍組織を用いたフィルタリングは有効であるが、複雑さおよびコストが増すので、臨床的に無理があり得る。腫瘍遺伝子型の予知なしで新規のバリエーション分類器を用いるフィルタリングは、臨床的に許容され得る感度を維持しつつ、偽陽性を無くした。先験的なバリエーション分類は、初期の疾患における補助療法の意思決定のための臨床的に実現可能な c t D N A 診断を可能にし得る。

30

【0224】

前述の開示は、明確化および理解の目的で例証および例示としてある程度詳細に説明してきたが、形態および詳細の様々な変更が、本開示の真の範囲から逸脱することなく可能であること、および添付の請求項の範囲内で行われ得ることが、本開示を読むことで当業者には明らかになるだろう。例えば、すべての方法、システム、コンピュータ可読媒体および／または構成要素の特徴、工程、エレメントまたはそれらの他の態様が、様々な組み合わせで用いることができる。

40

【0225】

本明細書に引用されたすべての特許、特許出願、ウェブサイト、他の刊行物または書類、アクセッション番号などは、各個別の項目が参照により援用されると明確かつ個別に示されたのと同程度に、あらゆる目的のためにそれらの全体が参照により援用される。ある配列の異なるバージョンが、種々の時点におけるアクセッション番号と関連付けられている場合、本出願の有効出願日時点のそのアクセッション番号と関連付けられているバージョンを意味する。有効出願日とは、実際の出願日、または該当する場合、アクセッション番号を参照している優先権出願の出願日のうち、いずれか早い日を意味する。同様に、異なるバージョンの刊行物、ウェブサイトなどが、異なる時点に公開されている場合、別段

50

示されない限り、本出願の有効出願日から最も近く公開されたバージョンを意味する。

【図 1】

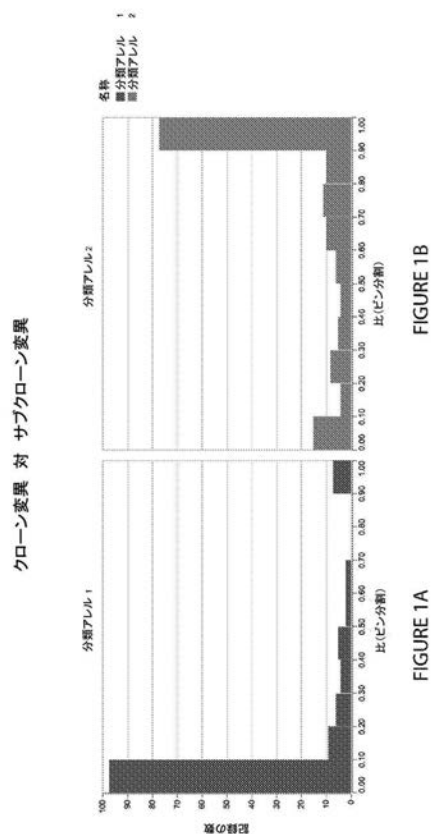
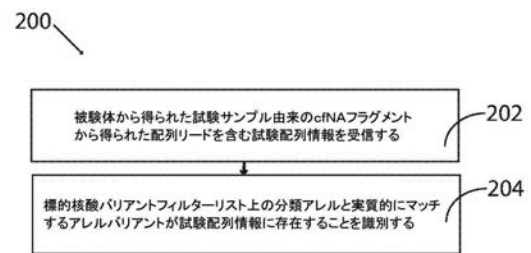
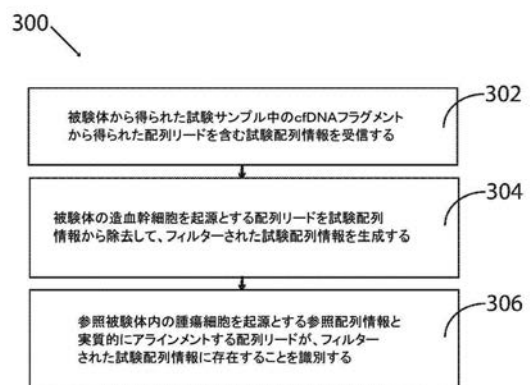


FIGURE 1B

【図 2】



【図 3】



【図 4】

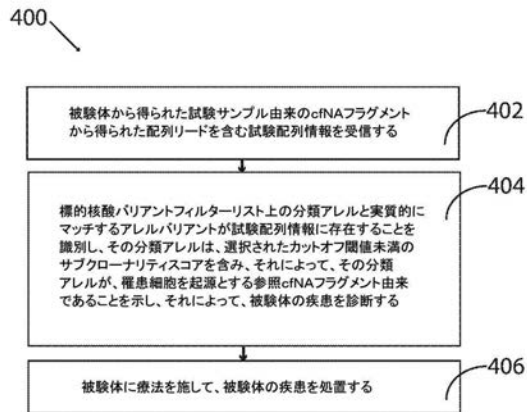


FIGURE 4

【図 5】

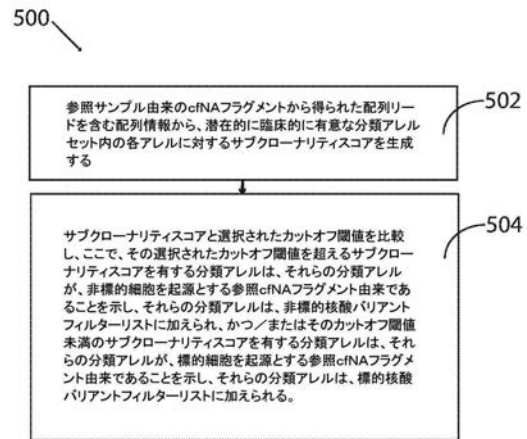


FIGURE 5

【図 6】

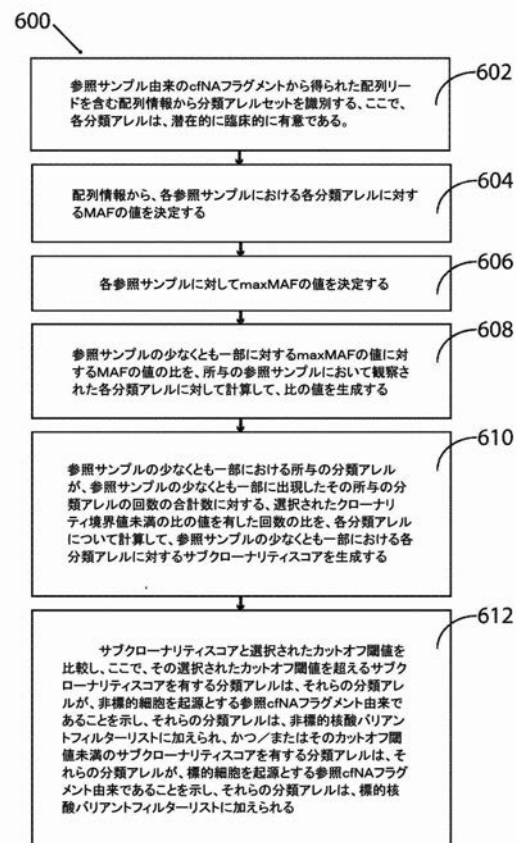


FIGURE 6

【図 7】

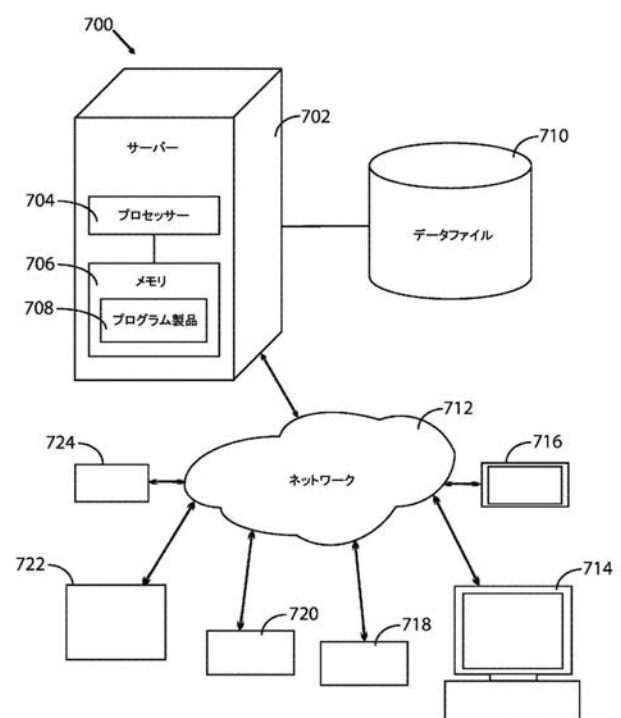


FIGURE 7

【図 8】

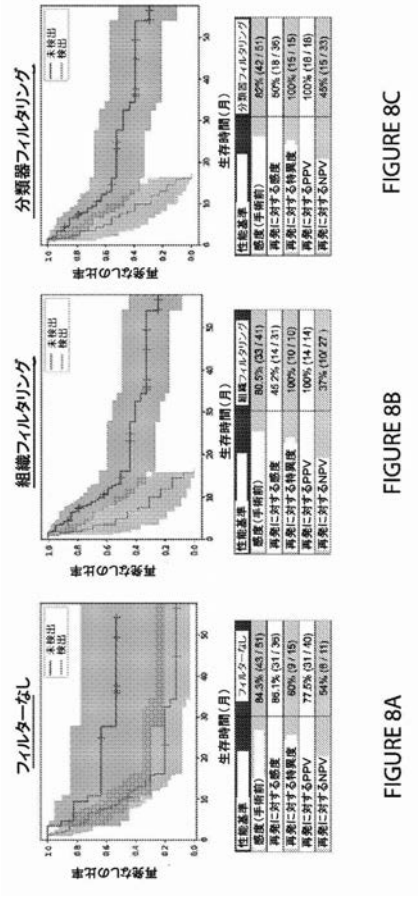


FIGURE 8C

FIGURE 8B

FIGURE 8A

【図 9】

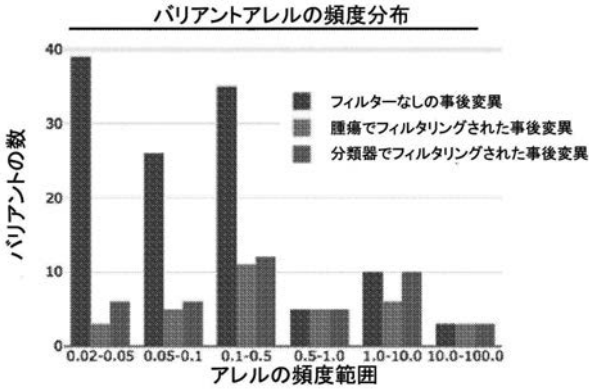


FIGURE 9

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2019/035214

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12Q1/6869 G16B30/00
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 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, BIOSIS, Sequence Search, 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	WO 2016/015058 A2 (UNIV WASHINGTON [US]) 28 January 2016 (2016-01-28) see whole doc. esp. claims -----	1-73
X	WO 2017/181146 A1 (GUARDANT HEALTH INC [US]) 19 October 2017 (2017-10-19) see whole dco. esp. claims and par. 273 ff. -----	1-73
X	WO 2013/190441 A2 (UNIV HONG KONG CHINESE [CN]; CHIU WAI KWUN ROSSA [CN]) 27 December 2013 (2013-12-27) see whole doc. esp. claims and Figures -----	1-73
A	WO 2015/169947 A1 (LIFECODEXX AG [DE]) 12 November 2015 (2015-11-12) the whole document -----	1-73

☐ 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

5 August 2019

Date of mailing of the international search report

14/08/2019

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

Mueller, Frank

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2019/035214

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2016015058 A2	28-01-2016	AU 2015292311 A1	09-03-2017
		CA 2956208 A1	28-01-2016
		CN 107002122 A	01-08-2017
		EP 3172341 A2	31-05-2017
		JP 2017522908 A	17-08-2017
		KR 20170044660 A	25-04-2017
		US 2017211143 A1	27-07-2017
		US 2019127794 A1	02-05-2019
		WO 2016015058 A2	28-01-2016
WO 2017181146 A1	19-10-2017	EP 3443066 A1	20-02-2019
		US 2019085406 A1	21-03-2019
		WO 2017181146 A1	19-10-2017
WO 2013190441 A2	27-12-2013	AU 2013278994 A1	29-01-2015
		AU 2017204558 A1	27-07-2017
		CA 2876327 A1	27-12-2013
		CN 104662168 A	27-05-2015
		CN 107779506 A	09-03-2018
		DK 2864501 T3	05-11-2018
		EA 201500027 A1	29-05-2015
		EP 2864501 A2	29-04-2015
		EP 3456843 A1	20-03-2019
		ES 2687847 T3	29-10-2018
		HK 1204013 A1	06-11-2015
		HK 1205533 A1	18-12-2015
		JP 6371280 B2	08-08-2018
		JP 2015527057 A	17-09-2015
		JP 2018191640 A	06-12-2018
		KR 20150032708 A	27-03-2015
		KR 20180088922 A	07-08-2018
		MX 360264 B	26-10-2018
		SG 10201808217W A	30-10-2018
		SG 112014081130 A	29-01-2015
		TW 201403066 A	16-01-2014
		WO 2013190441 A2	27-12-2013
		ZA 201409281 B	30-08-2017
WO 2015169947 A1	12-11-2015	AU 2015257654 A1	10-11-2016
		CA 2985134 A1	12-11-2015
		CN 107223159 A	29-09-2017
		EP 3140421 A1	15-03-2017
		EP 3521454 A1	07-08-2019
		JP 2017514499 A	08-06-2017
		SG 11201608993R A	29-11-2016
		US 2017314073 A1	02-11-2017
		WO 2015169947 A1	12-11-2015

フロントページの続き

(51)Int.Cl.		F I	テーマコード (参考)
A 6 1 P 35/00	(2006.01)	A 6 1 P 35/00	
A 6 1 P 35/02	(2006.01)	A 6 1 P 35/02	
A 6 1 K 45/00	(2006.01)	A 6 1 K 45/00	
G 1 6 H 10/40	(2018.01)	G 1 6 H 10/40	
G 1 6 H 20/00	(2018.01)	G 1 6 H 20/00	

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

(74)代理人 230113332

弁護士 山本 健策

(72)発明者 ジャイモビッチ, アリエル

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 ケネディ, アンドリュー

アメリカ合衆国 カリフォルニア 92037, ラ ホヤ, モーニング ウェイ 3121

(72)発明者 オーデガード, ジャスティン アイ.

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

F ターム(参考) 4B029 AA07 AA27 BB11 CC01 FA03 FA15

4B063 QA01 QA13 QA18 QA19 QA20 QQ02 QQ08 QQ42 QR08 QR32

QR42 QR56 QR62 QS24 QS25 QS34 QX02 QX10

4C084 AA17 NA10 ZB261 ZB262 ZB271 ZB272

5L099 AA03 AA15