

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号
特表2024-528489
(P2024-528489A)

(43)公表日 令和6年7月30日(2024.7.30)

(51)国際特許分類
G 1 6 B 40/00 (2019.01)

F I
G 1 6 B 40/00

審査請求 未請求 予備審査請求 未請求 (全54頁)

(21)出願番号	特願2023-579476(P2023-579476)	(71)出願人	517192663 ファウンデーション・メディシン・イン コーポレイテッド アメリカ合衆国 マサチューセッツ 0 2 2 1 0, ボストン, サマー ストリート 4 0 0
(86)(22)出願日	令和4年6月24日(2022.6.24)	(74)代理人	110002077 園田・小林弁理士法人
(85)翻訳文提出日	令和6年1月9日(2024.1.9)	(72)発明者	ソコル, イーサン アメリカ合衆国 0 2 1 4 1 マサチュー セッツ, ケンブリッジ, セカンド ス トリート 1 5 0, シー／オー ファウ ンデーション メディシン インコーポレ イテッド
(86)国際出願番号	PCT/US2022/073167	(72)発明者	ムーア, ジェイ
(87)国際公開番号	WO2022/272310		
(87)国際公開日	令和4年12月29日(2022.12.29)		
(31)優先権主張番号	63/215,281		
(32)優先日	令和3年6月25日(2021.6.25)		
(33)優先権主張国・地域又は機関	米国(US)		
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA ,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR ,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC, 最終頁に続く		最終頁に続く

(54)【発明の名称】 相同修復欠損を分類するシステムおよび方法

(57)【要約】

相同修復欠損（HRD）分類モデルを訓練および使用する
ための1つまたは複数の特徴の重要度のメトリック
を使用して、複数の特徴のサブセットを識別するための
方法、デバイス、およびシステムが本明細書に記載され
る。さらに、膵臓がんなどのがんの腫瘍を、HRD陽性
である可能性が高いまたはHRD陰性である可能性が高
いものと分類し、腫瘍をHRD陽性またはHRD陰性と
みなすための方法、デバイス、およびシステムが記載さ
れる。分類に基づいて、膵臓がんなどのがんの腫瘍を治
療する方法も本明細書に記載される。

【選択図】 図 7

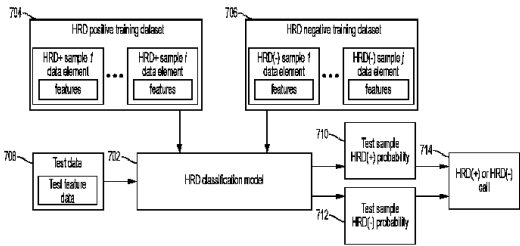


FIG. 7

【特許請求の範囲】

【請求項 1】

対象の腫瘍から得られるゲノムを準備するステップ、
任意選択で、1つまたは複数のアダプターを前記ゲノム上にライゲーションするステップ、

前記ゲノムから核酸分子を増幅するステップ、

前記増幅したゲノムから核酸分子を捕捉するステップであって、1つ以上のベイト分子へのハイブリダイゼーションによって、前記捕捉された核酸分子が捕捉される、核酸分子を捕捉するステップ、

前記捕捉された核酸分子から、入力特徴のセットを誘導するステップ、

1つまたは複数のプロセッサによって、前記入力特徴のセットを訓練された相同組換え欠損（HRD）モデルに入力し、前記訓練されたHRDモデルを使用して前記腫瘍をHRD陽性またはHRD陰性として識別するステップであって、前記モデルは、

複数の特徴の各特徴と関連付けられた1つまたは複数の特徴の重要度のメトリックを判定すること、

前記1つまたは複数の特徴の重要度のメトリックを使用して前記複数の特徴のうちの特徴のサブセットを識別すること、および

前記1つまたは複数のプロセッサによって、前記識別された特徴のサブセットに基づいて前記HRDモデルを訓練すること

により訓練される、前記入力特徴のセットを訓練された相同組換え欠損（HRD）モデルに入力し、前記訓練されたHRDモデルを使用して前記腫瘍をHRD陽性またはHRD陰性として識別するステップ、ならびに

前記訓練されたHRDモデルを使用して、前記1つまたは複数のプロセッサによって、前記腫瘍をHRD陽性またはHRD陰性として分類するステップを含む、方法。

【請求項 2】

1つまたは複数のプロセッサによって、複数の特徴を受信するステップ、

前記1つまたは複数のプロセッサによって、1つまたは複数の特徴の重要度のメトリックを使用して前記複数の特徴のうちの特徴のサブセットを識別するステップ、および

前記1つまたは複数のプロセッサによって、前記複数の特徴の前記識別されたサブセットに基づいて相同組換え欠損（HRD）モデルを訓練するステップであって、

前記HRDモデルが、対象の腫瘍のゲノムと関連付けられたサンプルデータを受信し、前記サンプルデータを使用して、前記対象の前記腫瘍をHRD陽性またはHRD陰性として識別するように構成される、HRDモデルを訓練するステップを含む、方法。

【請求項 3】

1つまたは複数のプロセッサによって、対象における腫瘍のゲノムに関連するサンプルデータを受信するステップ、

前記1つまたは複数のプロセッサによって、前記サンプルデータを訓練された相同組換え欠損（HRD）モデルに入力するステップであって、前記HRDモデルは、

複数の特徴の各特徴と関連付けられた1つまたは複数の特徴の重要度のメトリックを判定すること、

前記1つまたは複数の特徴の重要度のメトリックを使用して前記複数の特徴のうちの特徴のサブセットを識別すること、および

前記1つまたは複数のプロセッサによって、前記識別された特徴のサブセットに基づいて前記HRDモデルを訓練すること

により訓練される、前記サンプルデータを訓練された相同組換え欠損（HRD）モデルに入力するステップ、ならびに

前記訓練されたHRDモデルを使用して、前記1つまたは複数のプロセッサによって、前記腫瘍をHRD陽性またはHRD陰性として分類するステップ

10

20

30

40

50

を含む、方法。

【請求項 4】

前記複数の特徴が、1つもしくは複数のコピー数の特徴、1つもしくは複数の短い変異体の特徴、またはそれらの組み合わせを含む、請求項 1～3 のいずれか一項に記載の方法。

【請求項 5】

前記 1 つまたは複数の特徴の重要度のメトリックは、カイ二乗検定、分散分析（ANOVA）、ランダムフォレスト、または勾配ブースティングのうちの 1 つまたは複数を含む、請求項 1～4 のいずれか一項に記載の方法。

【請求項 6】

前記複数の特徴のうちの前記特徴のサブセットを識別するステップが、
前記 1 つまたは複数のプロセッサによって、前記 1 つまたは複数の特徴の重要度のメトリックに従って 1 つまたは複数の特徴ランキングを取得するステップ、および
前記 1 つまたは複数のプロセッサによって、1 つまたは複数の特徴ランキングに基づいて前記複数の特徴の前記サブセットを選択するステップ
を含む、請求項 1～5 のいずれか一項に記載の方法。

【請求項 7】

前記複数の特徴のうちの前記サブセットを識別するステップが、
（a）1 つまたは複数のプロセッサによって、特徴の重要度のメトリックに従って前記複数の特徴の特徴ランキングを取得するステップ、
（b）前記 1 つ以上のプロセッサによって、前記特徴ランキングに基づいて、前記複数の特徴からの 1 つ以上の特徴を既存の特徴セットに追加するステップによって新しい特徴セットを取得するステップ、
（c）前記 1 つまたは複数のプロセッサによって、前記新しい特徴セットを使用して新しい HRD モデルを訓練するステップ、
（d）前記 1 つまたは複数のプロセッサによって、前記訓練された新しい HRD モデルを評価して、評価結果を取得するステップ、
（e）前記 1 つまたは複数のプロセッサによって、前記新しい HRD モデルおよび前記新しい特徴セットに関連する前記評価結果を記憶するステップ、
（f）条件が満たされるまで、複数の評価結果を得るために、前記 1 つまたは複数のプロセッサによって、ステップ（b）～（e）を繰り返すステップ、ならびに
（g）前記 1 つまたは複数のプロセッサによって、前記複数の評価結果に基づいて前記複数の特徴の前記サブセットを選択するステップ
を含む、請求項 1～5 のいずれか一項に記載の方法。

【請求項 8】

前記訓練された HRD モデルが分類モデルであり、前記方法が、
新しい対象における腫瘍のゲノムと関連付けられた新しいサンプルデータを受信するステップであって、前記新しいサンプルデータは、前記複数の特徴の前記サブセットに関連する、新しいサンプルデータを受信するステップ、
前記新しいサンプルデータを前記訓練された HRD 分類モデルに与えて、HRD 陽性または HRD 陰性の分類結果を生成するステップ、および
前記分類結果を出力するステップ
をさらに含む、請求項 1～7 のいずれか一項に記載の方法。

【請求項 9】

前記分類結果が、HRD 陽性尤度スコアおよび HRD 陰性尤度スコアのうちの少なくとも 1 つを含む、請求項 8 に記載の方法。

【請求項 10】

前記 HRD モデルが、分類モデル、回帰モデル、ニューラルネットワーク、またはそれらの任意の組み合わせである、請求項 1～9 のいずれか一項に記載の方法。

【請求項 11】

10

20

30

40

50

前記新しい対象に関連付けられたデジタル電子ファイルに、前記HRD陽性尤度スコアおよび前記HRD陰性尤度スコアの少なくとも一方を記録するステップを含む、請求項9または10に記載の方法。

【請求項12】

前記新しい対象に関連するデジタル電子ファイルに、前記HRD陽性尤度スコアに基づいて前記腫瘍がHRD陽性であること、または前記HRD陰性尤度スコアに基づいて前記腫瘍がHRD陰性であるという指定を記録するステップを含む、請求項9～11のいずれか一項に記載の方法。

【請求項13】

前記複数の特徴が、セグメントマイナーアリアル頻度 (segMAF) の特徴、シーケンシングリードの数の特徴、セグメントサイズの特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つを含む、請求項1～12のいずれか一項に記載の方法。

【請求項14】

前記複数の特徴のうちの少なくとも1つが、前記ゲノムのセントロメア部分にわたって評価される、請求項1～13のいずれか一項に記載の方法。

【請求項15】

前記複数の特徴のうちの少なくとも1つが、前記ゲノムのテロメア部分にわたって評価される、請求項1～14のいずれか一項に記載の方法。

【請求項16】

前記複数の特徴のうちの少なくとも1つが、前記ゲノムの前記セントロメア部分およびテロメア部分の両方にわたって評価される、請求項1～15のいずれか一項に記載の方法。

【請求項17】

前記複数の特徴が、xメガベースあたりのブレイクポイント数の特徴を含み、前記xメガベースあたりのブレイクポイント数の特徴が、前記ゲノム全体にわたって長さがxメガベースのウィンドウに現れる前記ブレイクポイントの数に基づく、請求項1～16のいずれか一項に記載の方法。

【請求項18】

xメガベースあたりのブレイクポイント数の特徴が、(i) 前記ゲノムの前記テロメア部分、(ii) 前記ゲノムの前記セントロメア部分、または(iii) 前記ゲノムの前記テロメア部分と前記セントロメア部分の両方にわたって評価される、請求項17に記載の方法。

【請求項19】

xが約1～約100メガベースである、請求項17または請求項18に記載の方法。

【請求項20】

xが、約10メガベース、約25メガベース、約50メガベース、または約100メガベースである、請求項17～19のいずれか一項に記載の方法。

【請求項21】

xメガベースあたりのブレイクポイント数の特徴が、ビニングされた特徴である、請求項17～20のいずれか一項に記載の方法。

【請求項22】

前記複数の特徴が変化点コピー数の特徴を含み、前記変化点コピー数が、前記対象の前記腫瘍の前記ゲノムにわたる隣接ゲノムセグメント間のコピー数の絶対的な差に基づく、請求項1～21のいずれか一項に記載の方法。

【請求項23】

前記変化点コピー数の特徴が、倍数正規化コピー数データから誘導される、請求項22に記載の方法。

【請求項24】

10

20

30

40

50

変化点コピー数の特徴が、(i) 前記ゲノムの前記テロメア部分、(ii) 前記ゲノムの前記セントロメア部分、または (iii) 前記ゲノムの前記テロメア部分と前記セントロメア部分の両方にわたって評価される、請求項 22 または 23 に記載の方法。

【請求項 25】

前記変化点コピー数の特徴がビニングされた特徴である、請求項 22～24 のいずれか一項に記載の方法。

【請求項 26】

前記複数の特徴がセグメントコピー数の特徴を含み、セグメントコピー数が各ゲノムセグメントの前記コピー数に基づく、請求項 1～25 のいずれか一項に記載の方法。

【請求項 27】

前記セグメントコピー数の特徴が、(i) 前記ゲノムの前記テロメア部分、(ii) 前記ゲノムの前記セントロメア部分、または (iii) 前記ゲノムの前記テロメア部分と前記セントロメア部分の両方にわたって評価される、請求項 26 に記載の方法。

【請求項 28】

前記セグメントコピー数の特徴が、倍数正規化コピー数データから誘導される、請求項 26 または請求項 27 に記載の方法。

【請求項 29】

前記セグメントコピー数の特徴がビニングされた特徴である、請求項 26～28 のいずれか一項に記載の方法。

【請求項 30】

前記複数の特徴が、前記対象の前記腫瘍のゲノムの染色体アームあたりのブレイクポイント数の特徴を含む、請求項 1～29 のいずれか一項に記載の方法。

【請求項 31】

前記染色体アームあたりのブレイクポイント数の特徴が、(i) 前記ゲノムの前記テロメア部分、(ii) 前記ゲノムの前記セントロメア部分、または (iii) 前記ゲノムの前記テロメア部分と前記セントロメア部分の両方にわたって評価される、請求項 30 に記載の方法。

【請求項 32】

前記染色体アームあたりのブレイクポイント数の特徴がビニングされた特徴である、請求項 30 または請求項 31 に記載の方法。

【請求項 33】

前記複数の特徴が、振動コピー数のセグメントの数の特徴を含む、請求項 1～32 のいずれか一項に記載の方法。

【請求項 34】

前記振動コピー数のセグメントの数の特徴が、前記対象の前記腫瘍の前記ゲノムにわたる 2 つのコピー数の間の繰り返される交互のセグメントの前記数に基づく、請求項 33 に記載の方法。

【請求項 35】

振動コピー数のセグメントの数の特徴が、(i) 前記ゲノムの前記テロメア部分、(ii) 前記ゲノムの前記セントロメア部分、または (iii) 前記ゲノムの前記テロメア部分と前記セントロメア部分の両方にわたって評価される、請求項 33 または請求項 34 に記載の方法。

【請求項 36】

前記振動コピー数のセグメントの数の特徴がビニングされた特徴である、請求項 33～35 のいずれか一項に記載の方法。

【請求項 37】

前記 1 つまたは複数のコピー数の特徴が、セグメントマイナーアリル頻度 (segMAF) 特徴を含み、segMAF が、ヘテロ接合一塩基多型における前記マイナーアリル頻度に基づく、請求項 1～36 のいずれか一項に記載の方法。

【請求項 38】

10

20

30

40

50

s e g M A F が、(i) 前記ゲノムの前記テロメア部分、(i i) 前記ゲノムの前記セントロメア部分、または(i i i) 前記ゲノムの前記テロメア部分と前記セントロメア部分の両方にわたって評価される、請求項 3 7 に記載の方法。

【請求項 3 9】

前記セグメントマイナーアリアル頻度の特徴がビニングされた特徴である、請求項 3 7 または請求項 3 8 に記載の方法。

【請求項 4 0】

前記 1 つまたは複数のコピー数の特徴が、シーケンシングリードの数の特徴を含む、請求項 1 ～ 3 9 のいずれか一項に記載の方法。

【請求項 4 1】

前記シーケンシングリードの数の特徴がビニングされた特徴である、請求項 4 0 に記載の方法。

【請求項 4 2】

前記複数の特徴が、前記対象の前記腫瘍の前記ゲノムのヘテロ接合性のゲノム全体の喪失の尺度をさらに含む、請求項 1 ～ 4 1 のいずれか一項に記載の方法。

【請求項 4 3】

前記複数の特徴が 1 つまたは複数の短い変異体の特徴を含む、請求項 1 ～ 4 2 のいずれか一項に記載の方法。

【請求項 4 4】

前記 1 つまたは複数の短い変異体の特徴が、ミクロ相同性または反復領域特徴の欠失および 2 つ以上の短い変異体の特徴に由来する変異シグネチャの少なくとも 1 つを含む、請求項 4 3 に記載の方法。

【請求項 4 5】

ミクロ相同性または反復領域特徴の前記欠失が、少なくとも 5 塩基対の欠失である、請求項 4 4 に記載の方法。

【請求項 4 6】

前記 H R D モデルを訓練するステップは、

前記 1 つまたは複数のプロセッサによって、H R D 陽性訓練データセットを受信するステップであって、前記 H R D 陽性訓練データセットが、H R D 陽性腫瘍および H R D 陽性標識に関連する複数の特徴を含む、H R D 陽性訓練データセットを受信するステップ、

前記 1 つまたは複数のプロセッサによって、H R D 陰性訓練データセットを受信するステップであって、前記 H R D 陰性訓練データセットが、H R D 陰性腫瘍および H R D 陰性標識に関連する複数の特徴を含む、H R D 陰性訓練データセットを受信するステップ、

前記 1 つまたは複数のプロセッサによって、前記 H R D 陽性訓練データセットおよび前記 H R D 陰性訓練データセットを使用して前記 H R D モデルを訓練するステップを含む、請求項 1 ～ 4 5 のいずれか一項に記載の方法。

【請求項 4 7】

前記 1 つまたは複数のプロセッサによって、B R C A 1、B R C A 2、B R C A 1 と B R C A 2 の両方における機能喪失変異、または B R C A 1 と B R C A 2 の二対立遺伝子変異を含むゲノム配列に由来する H R D 陽性対照を含む H R D 陽性試験データセットを使用して、前記訓練されたモデルを試験するステップをさらに含む、請求項 1 ～ 4 6 のいずれか一項に記載の方法。

【請求項 4 8】

前記 1 つまたは複数のプロセッサによって、A T M、B A R D 1、B R I P 1、C D K 1 2、C H E K 1、C H E K 2、F A N C L、P A L B 2、R A D 5 1 B、R A D 5 1 C、R A D 5 1 D、または R A D 4 5 L のうちの少なくとも 1 つに機能喪失変異を含むゲノム配列に由来する H R D 陽性対照を含む H R D 陽性試験データセットを使用して、前記訓練されたモデルを試験するステップをさらに含む、請求項 1 ～ 4 7 のいずれか一項に記載の方法。

【請求項 4 9】

10

20

30

40

50

前記 1 つまたは複数のプロセッサによって、コンセンサスなヒトゲノム配列に由来する H R D 陰性対照を含む H R D 陰性訓練データセットを含む H R D 陰性試験データセットを使用して前記訓練されたモデルを試験するステップをさらに含む、請求項 1 ～ 4 8 のいずれか一項に記載の方法。

【請求項 5 0】

訓練するステップが、H R D 陽性訓練データセットおよび H R D 陰性訓練データセットを使用するステップを含む、請求項 4 6 ～ 4 9 のいずれか一項に記載の方法。

【請求項 5 1】

前記 H R D モデルを訓練する前に、前記 1 つまたは複数のプロセッサによって、前記 H R D 陽性訓練データセットと前記 H R D 陰性訓練データセットとのバランスをとるステップを含む、請求項 5 0 に記載の方法。

10

【請求項 5 2】

前記対象における前記腫瘍が、前立腺がん、卵巣がん、乳がん、非小細胞肺癌（N S C L C）、結腸直腸がん（C R C）、または膵臓がんである、請求項 1 ～ 5 1 のいずれか一項に記載の方法。

【請求項 5 3】

前記 H R D モデルを訓練するステップが、卵巣がん、非小細胞肺癌（N S C L C）、結腸直腸がん（C R C）、乳がん、膵臓がん、または前立腺がんに関連するサンプルデータに前記 H R D モデルを適合させることを含み、前記サンプルデータが前記複数の特徴の前記サブセットを含む、請求項 1 ～ 5 2 のいずれか一項に記載の方法。

20

【請求項 5 4】

前記腫瘍が、固形組織生検サンプルであるサンプルから得られる、請求項 1 ～ 5 3 のいずれか一項に記載の方法。

【請求項 5 5】

前記固形組織生検サンプルがホルマリン固定パラフィン包埋（F F P E）サンプルである、請求項 5 4 に記載の方法。

【請求項 5 6】

前記腫瘍が、循環腫瘍 D N A（c t D N A）を含む液体生検サンプルであるサンプルから得られる、請求項 1 ～ 5 3 のいずれか一項に記載の方法。

【請求項 5 7】

前記腫瘍が、無細胞 D N A（c f D N A）を含む液体生検サンプルであるサンプルから得られる、請求項 1 ～ 5 3 のいずれか一項に記載の方法。

30

【請求項 5 8】

前記患者に関連する診断値として前記腫瘍の前記出力を H R D 陽性または H R D 陰性と判定、識別、または適用するステップをさらに含む、請求項 1 ～ 5 7 のいずれか一項に記載の方法。

【請求項 5 9】

前記腫瘍の H R D 陽性または H R D 陰性としての前記出力に基づいて、前記対象のゲノムプロファイルを生成するステップをさらに含む、請求項 1 ～ 5 8 のいずれか一項に記載の方法。

40

【請求項 6 0】

前記生成されたゲノムプロファイルに基づいて、抗がん剤を投与するステップ、または抗がん治療を前記対象に適用するステップをさらに含む、請求項 5 9 に記載の方法。

【請求項 6 1】

前記腫瘍の H R D 陽性または H R D 陰性としての前記出力が、前記対象のゲノムプロファイルを生成するのに使用される、請求項 1 ～ 6 0 のいずれか一項に記載の方法。

【請求項 6 2】

前記腫瘍の H R D 陽性または H R D 陰性としての前記出力が、前記対象に対する提案された治療の決定を行う際に使用される、請求項 1 ～ 6 1 のいずれか一項に記載の方法。

【請求項 6 3】

50

前記腫瘍のHRD陽性またはHRD陰性としての前記出力が、前記対象に治療を適用または投与するのに使用される、請求項1～62のいずれか一項に記載の方法。

【請求項64】

前記HRDモデルが、機械学習モデルである、請求項1～63のいずれか一項に記載の方法。

【請求項65】

前記対象が、がんを有するか、がんを有するリスクがあるか、またはがんを有する疑いがある、請求項1～64のいずれか一項に記載の方法。

【請求項66】

対象のがんを治療する方法であって、

(a) 請求項1～65のいずれか一項に記載の方法に従って前記腫瘍をHRD陽性またはHRD陰性として識別するステップ、

(b) 前記がんの前記腫瘍がHRD陽性と評価された場合、治療有効量のHRD陽性腫瘍に有効な薬物を前記対象に投与するステップ、を含む、方法。

【請求項67】

HRD陽性腫瘍に有効な前記薬物が白金系薬物またはPARP阻害剤である、請求項66に記載の方法。

【請求項68】

前記腫瘍がHRD陰性と評価された場合、白金系薬物でもPARP阻害剤でもない治療有効量の薬物を前記対象に投与するステップを含む、請求項66に記載の方法。

【請求項69】

対象のがんの治療法を選択するための方法であって、

(a) 請求項1～65のいずれか一項に記載の方法に従って、前記がんの腫瘍をHRD陽性またはHRD陰性として評価するステップ、

(b) 前記がんがHRD陽性と評価された場合、HRD陽性腫瘍において有効な治療を選択するステップを含む、方法。

【請求項70】

前記腫瘍がHRD陰性と評価された場合、白金系薬物でもPARP阻害剤でもない治療法を選択するステップを含む、請求項69に記載の方法。

【請求項71】

HRD陽性腫瘍に有効な前記治療法が白金系薬物またはPARP阻害剤である、請求項70に記載の方法。

【請求項72】

コンピュータシステムであって、

1つまたは複数のプロセッサと、

メモリと、

1つまたは複数のプログラムであって、前記1つまたは複数のプログラムが前記メモリに記憶され、前記1つまたは複数のプロセッサによって実行されるように構成され、前記1つまたは複数のプログラムが、請求項1から65のいずれか一項に記載の方法を実行するための指示を含む、1つまたは複数のプログラムとを含む、コンピュータシステム。

【請求項73】

1つまたは複数のプログラムを記憶する非一時的コンピュータ可読記憶媒体であって、前記1つまたは複数のプログラムが命令を含み、前記命令が、電子デバイスの1つ以上のプロセッサによって実行されると、前記電子デバイスに、請求項1から65のいずれか一項に記載の方法を実行させる、非一時的コンピュータ可読記憶媒体。

【発明の詳細な説明】

【関連出願の相互参照】

【0001】

10

20

30

40

50

本願は、2021年6月25日に出願された「SYSTEM AND METHOD OF CLASSIFYING HOMOLOGOUS REPAIR DEFICIENCY」と題する米国仮出願第63/215,281号の優先権の利益を主張し、その内容は、すべての目的のために参照により本明細書に組み込まれる。

【技術分野】

【0002】

発明の分野

相同修復欠損（HRD）モデルの特徴を選択し、HRDモデルを使用して腫瘍を評価し、評価に基づいて腫瘍を治療するための方法、デバイス、およびシステムが本明細書に記載される。

【背景技術】

【0003】

発明の背景

コピー数の異常は、ゲノムの大きな連続セグメントの欠失または増幅を伴い、がんにおける一般的な変異である。特定のコピー数の異常は、相同修復欠損（HRD）と呼ばれる相同組換え修復機構によってゲノムを修復することができないことに関連する。HRDを有するいくつかの腫瘍を識別するために、相同修復経路に関与する遺伝子の変異を配列決定することが可能である。あるいは、その原因にかかわらず、HRDの物理的結果であるゲノム瘢痕化を検出することが可能である。

【0004】

HRDを示す腫瘍ゲノムは、白金化学療法またはポリ（ADP）リボースポリメラーゼ（PARP）阻害剤などの特定の薬物に対する感受性に関連する。しかしながら、特定の腫瘍をHRD陽性として分類することは依然として困難である。したがって、適切な治療を選択して対象に投与することができるように、特に重要である膀胱がん、乳がん、または前立腺がんなどのがんの腫瘍を、HRD陽性またはHRD陰性として分類する必要性が、依然として存在する。過去において、HRDを識別するための技術は、実際にそれらを使用することができない不正確さおよび非効率性に悩まされてきた。この1つの理由は、例えば、オーバーフィッティングに起因して、前記腫瘍をHRD陽性またはHRD陰性として効率的かつ正確に識別する（例えば、分類する）ために、サンプルのHRD状態を正確に判定することができるようにするには、特徴選択技術が現在不十分であることである。この別の理由は、HRDの状態を正確に判定するためにどの特徴を識別するかを判定することも、困難であり得るということである。したがって、前記識別を実行するためのモデルを訓練するために使用することができる、複数の特徴から特徴のサブセットを正確かつ効率的に選択する技術およびシステムが必要とされている。

【発明の概要】

【0005】

対象の腫瘍から得られるゲノムを準備するステップ、任意選択で、1つまたは複数のアダプターをゲノム上にライゲーションするステップ、ゲノムから核酸分子を増幅するステップ、増幅したゲノムから核酸分子を捕捉するステップであって、1つ以上のベイト分子へのハイブリダイゼーションによって、捕捉された核酸分子が捕捉される、核酸分子を捕捉するステップ、捕捉された核酸分子から、入力特徴のセットを誘導するステップ、1つまたは複数のプロセッサによって、入力特徴のセットを訓練された相同組換え欠損（HRD）モデルに入力し、訓練されたHRDモデルを使用して腫瘍をHRD陽性またはHRD陰性として識別するステップであって、モデルは、複数の特徴の各特徴と関連付けられた1つまたは複数の特徴の重要度のメトリックを判定すること、1つまたは複数の特徴の重要度のメトリックを使用して複数の特徴のうちの特徴のサブセットを識別すること、および1つまたは複数のプロセッサによって、識別された特徴のサブセットに基づいてHRDモデルを訓練すること、により訓練される、入力特徴のセットを訓練された相同組換え欠損（HRD）モデルに入力し、訓練されたHRDモデルを使用して腫瘍をHRD陽性またはHRD陰性として識別するステップ、ならびに訓練されたHRDモデルを使用して、1

10

20

30

40

50

つまたは複数のプロセッサによって、腫瘍をHRD陽性またはHRD陰性として分類するステップ、を含む方法が、本明細書に記載される。

【0006】

1つまたは複数のプロセッサによって、複数の特徴を受信するステップ、1つまたは複数のプロセッサによって、1つまたは複数の特徴の重要度のメトリックを使用して複数の特徴のうちの特徴のサブセットを識別するステップ、および1つまたは複数のプロセッサによって、複数の特徴の識別されたサブセットに基づいて相同組換え欠損（HRD）モデルを訓練するステップであって、HRDモデルが、対象の腫瘍のゲノムと関連付けられたサンプルデータを受信し、サンプルデータを使用して、対象の腫瘍をHRD陽性またはHRD陰性として識別するように構成される、HRDモデルを訓練するステップ、を含む方法が、本明細書にさらに記載される。

10

【0007】

1つまたは複数のプロセッサによって、対象における腫瘍のゲノムに関連するサンプルデータを受信するステップ、1つまたは複数のプロセッサによって、サンプルデータを訓練された相同組換え欠損（HRD）モデルに入力するステップであって、HRDモデルは、複数の特徴の各特徴と関連付けられた1つまたは複数の特徴の重要度のメトリックを判定すること、1つまたは複数の特徴の重要度のメトリックを使用して複数の特徴のうちの特徴のサブセットを識別すること、および1つまたは複数のプロセッサによって、識別された特徴のサブセットに基づいてHRDモデルを訓練すること、により訓練される、サンプルデータを訓練された相同組換え欠損（HRD）モデルに入力するステップ、ならびに訓練されたHRDモデルを使用して、1つまたは複数のプロセッサによって、腫瘍をHRD陽性またはHRD陰性として分類するステップ、を含む方法が、本明細書にさらに記載される。

20

【0008】

記載された方法のいくつかの実施形態では、複数の特徴は、1つもしくは複数のコピー数の特徴、1つもしくは複数の短い変異体の特徴、またはそれらの組み合わせを含む。記載された方法のいくつかの実施形態では、1つまたは複数の特徴の重要度のメトリックは、カイ二乗検定、分散分析（ANOVA）、ランダムフォレスト、または勾配ブースティングのうちの1つまたは複数を含む。

【0009】

記載された方法のいくつかの実施形態では、複数の特徴のうちの特徴のサブセットを識別するステップが、1つまたは複数のプロセッサによって、1つまたは複数の特徴の重要度のメトリックに従って1つまたは複数の特徴ランキングを取得するステップ、および1つまたは複数のプロセッサによって、1つまたは複数の特徴ランキングに基づいて複数の特徴のサブセットを選択するステップを含む。

30

【0010】

記載された方法のいくつかの実施形態では、複数の特徴のサブセットを識別することは、（a）1つまたは複数のプロセッサによって、特徴の重要度のメトリックに従って複数の特徴の特徴ランキングを取得するステップ、（b）1つ以上のプロセッサによって、特徴ランキングに基づいて、複数の特徴からの1つ以上の特徴を既存の特徴セットに追加するステップによって新しい特徴セットを取得するステップ、（c）1つまたは複数のプロセッサによって、新しい特徴セットを使用して新しいHRDモデルを訓練するステップ、（d）1つまたは複数のプロセッサによって、訓練された新しいHRDモデルを評価して、評価結果を取得するステップ、（e）1つまたは複数のプロセッサによって、新しいHRDモデルおよび新しい特徴セットに関連する評価結果を記憶するステップ、（f）条件が満たされるまで、複数の評価結果を得るために、1つまたは複数のプロセッサによって、ステップ（b）～（e）を繰り返すステップ、ならびに（g）1つまたは複数のプロセッサによって、複数の評価結果に基づいて複数の特徴のサブセットを選択するステップ、を含む。

40

【0011】

50

記載された方法のいくつかの実施形態では、訓練されたHRDモデルが分類モデルであり、方法が、新しい対象における腫瘍のゲノムと関連付けられた新しいサンプルデータを受信するステップであって、新しいサンプルデータは、複数の特徴のサブセットに関連する、新しいサンプルデータを受信するステップ、新しいサンプルデータを訓練されたHRD分類モデルに与えて、HRD陽性またはHRD陰性の分類結果を生成するステップ、および分類結果を出力するステップ、をさらに含む。いくつかの実施形態では、分類結果は、HRD陽性尤度スコアおよびHRD陰性尤度スコアの少なくとも1つを含む。いくつかの実施形態では、方法は、新しい対象に関連付けられたデジタル電子ファイルに、HRD陽性尤度スコアおよびHRD陰性尤度スコアの少なくとも1つを記録するステップを含む。いくつかの実施形態では、方法は、新しい対象に関連するデジタル電子ファイルに、HRD陽性尤度スコアに基づいて腫瘍がHRD陽性であること、またはHRD陰性尤度スコアに基づいて腫瘍がHRD陰性であるという指定を記録するステップを含む。

10

【0012】

記載された方法のいくつかの実施形態では、HRDモデルは、分類モデル、回帰モデル、ニューラルネットワーク、またはそれらの任意の組み合わせである。いくつかの実施形態では、方法は、新しい対象に関連付けられたデジタル電子ファイルに、HRD陽性尤度スコアおよびHRD陰性尤度スコアの少なくとも1つを記録するステップを含む。いくつかの実施形態では、方法は、新しい対象に関連するデジタル電子ファイルに、HRD陽性尤度スコアに基づいて腫瘍がHRD陽性であること、またはHRD陰性尤度スコアに基づいて腫瘍がHRD陰性であるという指定を記録するステップを含む。

20

【0013】

記載された方法のいくつかの実施形態では、複数の特徴が、セグメントマイナーアリアル頻度 (segMAF) の特徴、シーケンシングリードの数の特徴、セグメントサイズの特徴、xメガベースあたりのブレークポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレークポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つを含む。記載された方法のいくつかの実施形態では、複数の特徴のうちの少なくとも1つが、ゲノムのセントロメア部分にわたって評価される。記載された方法のいくつかの実施形態では、複数の特徴のうちの少なくとも1つが、ゲノムのテロメア部分にわたって評価される。

【0014】

記載された方法のいくつかの実施形態では、複数の特徴のうちの少なくとも1つが、ゲノムのセントロメア部分およびテロメア部分の両方にわたって評価される。

30

【0015】

記載された方法のいくつかの実施形態では、複数の特徴が、xメガベースあたりのブレークポイント数の特徴を含み、xメガベースあたりのブレークポイント数の特徴が、ゲノム全体にわたって長さがxメガベースのウィンドウに現れるブレークポイントの数に基づく。いくつかの実施形態においてxメガベースあたりのブレークポイント数の特徴が、(i) ゲノムのテロメア部分、(ii) ゲノムのセントロメア部分、または(iii) ゲノムのテロメア部分とセントロメア部分の両方にわたって評価される。いくつかの実施形態において、xは、約1～約100メガベースである。いくつかの実施形態では、xは、約10メガベース、約25メガベース、約50メガベース、または約100メガベースである。いくつかの実施形態において、xメガベースあたりのブレークポイント数の特徴が、ビンニングされた特徴である。

40

【0016】

記載された方法のいくつかの実施形態では、複数の特徴が変化点コピー数の特徴を含み、変化点コピー数が、対象の腫瘍のゲノムにわたる隣接ゲノムセグメント間のコピー数の絶対的な差に基づく。いくつかの実施形態において、変化点コピー数の特徴は、倍数性正規化コピー数データに由来する。いくつかの態様では、変化点コピー数の特徴は、(i) ゲノムのテロメア部分、(ii) ゲノムのセントロメア部分、または(iii) ゲノムのテロメア部分とセントロメア部分の両方にわたって評価される。いくつかの実施形態では

50

、変化点コピー数の特徴は、ビンニングされた特徴である。

【0017】

記載された方法のいくつかの実施形態では、複数の特徴は、セグメントコピー数の特徴を含み、セグメントコピー数が各ゲノムセグメントのコピー数に基づく。いくつかの態様において、セグメントコピー数の特徴は、(i)ゲノムのテロメア部分、(ii)ゲノムのセントロメア部分、または(iii)ゲノムのテロメア部分とセントロメア部分の両方にわたって評価される。いくつかの実施形態において、セグメントコピー数の特徴は、倍数性正規化コピー数データに由来する。いくつかの実施形態において、セグメントコピー数の特徴は、ビンニングされた特徴である。

【0018】

記載された方法のいくつかの実施形態では、複数の特徴が、対象の腫瘍のゲノムの染色体アームあたりのブレイクポイント数の特徴を含む。いくつかの実施形態において、染色体アームあたりのブレイクポイント数の特徴は、(i)ゲノムのテロメア部分、(ii)ゲノムのセントロメア部分、または(iii)ゲノムのテロメア部分とセントロメア部分の両方にわたって評価される。いくつかの実施形態において、染色体アームあたりのブレイクポイント数の特徴が、ビンニングされた特徴である。

【0019】

記載された方法のいくつかの実施形態では、複数の特徴が、振動コピー数のセグメントの数の特徴を含む。いくつかの実施形態において、振動コピー数のセグメントの数の特徴が、対象の腫瘍のゲノムにわたる2つのコピー数の間の繰り返される交互のセグメントの数の特徴に基づく。いくつかの実施形態において、振動コピー数のセグメントの数の特徴が、(i)ゲノムのテロメア部分、(ii)ゲノムのセントロメア部分、または(iii)ゲノムのテロメア部分とセントロメア部分の両方にわたって評価される。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、ビンニングされた特徴である。

【0020】

記載された方法のいくつかの実施形態では、1つまたは複数のコピー数の特徴が、セグメントマイナーアリル頻度(segMAF)特徴を含み、segMAFが、ヘテロ接合一塩基多型におけるマイナーアリル頻度に基づく。いくつかの態様では、segMAFは、(i)ゲノムのテロメア部分、(ii)ゲノムのセントロメア部分、または(iii)ゲノムのテロメア部分とセントロメア部分の両方にわたって評価される。いくつかの実施形態では、segMAF特徴はビンニングされた特徴である。

【0021】

記載された方法のいくつかの実施形態では、1つまたは複数のコピー数の特徴が、シーケンシングリードの数の特徴を含む。いくつかの態様において、シーケンシングリードの数の特徴はビンニングされた特徴である。

【0022】

記載された方法のいくつかの実施形態では、複数の特徴が、対象の腫瘍のゲノムのヘテロ接合性のゲノム全体の喪失の尺度をさらに含む。

【0023】

記載された方法のいくつかの実施形態では、複数の特徴は、1つまたは複数の短い変異体の特徴を含む。いくつかの実施形態において、1つまたは複数の短い変異体の特徴が、マイクロ相同性または反復領域特徴の欠失および2つ以上の短い変異体の特徴に由来する変異シグネチャの少なくとも1つを含む。いくつかの実施形態において、マイクロ相同性領域または反復領域特徴における欠失は、少なくとも5塩基対の欠失である。

【0024】

記載された方法のいくつかの実施形態では、HRDモデルを訓練するステップが、1つまたは複数のプロセッサによって、HRD陽性訓練データセットを受信するステップであって、HRD陽性訓練データセットが、HRD陽性腫瘍およびHRD陽性標識に関連する複数の特徴を含む、HRD陽性訓練データセットを受信するステップ、1つまたは複数のプロセッサによって、HRD陰性訓練データセットを受信するステップであって、HRD

10

20

30

40

50

陰性訓練データセットが、HRD陰性腫瘍およびHRD陰性標識に関連する複数の特徴を含む、HRD陰性訓練データセットを受信するステップ、1つまたは複数のプロセッサによって、HRD陽性訓練データセットおよびHRD陰性訓練データセットを使用してHRDモデルを訓練するステップを含む。いくつかの実施形態では、訓練するステップは、HRD陽性訓練データセットおよびHRD陰性訓練データセットを使用するステップを含む。いくつかの実施形態では、方法は、HRDモデルを訓練する前に、1つまたは複数のプロセッサによって、HRD陽性訓練データセットとHRD陰性訓練データセットとのバランスをとるステップを含む。

【0025】

記載された方法のいくつかの実施形態では、方法は、1つまたは複数のプロセッサによって、BRCA1、BRCA2、BRCA1とBRCA2の両方における機能喪失変異、またはBRCA1とBRCA2の二対立遺伝子変異を含むゲノム配列に由来するHRD陽性対照を含むHRD陽性試験データセットを使用して、訓練されたモデルを試験するステップをさらに含む。いくつかの実施形態では、訓練するステップは、HRD陽性訓練データセットおよびHRD陰性訓練データセットを使用するステップを含む。いくつかの実施形態では、方法は、HRDモデルを訓練する前に、1つまたは複数のプロセッサによって、HRD陽性訓練データセットとHRD陰性訓練データセットとのバランスをとるステップを含む。

10

【0026】

記載された方法のいくつかの実施形態では、方法は、1つまたは複数のプロセッサによって、ATM、BARD1、BRIP1、CDK12、CHEK1、CHEK2、FANCL、PALB2、RAD51B、RAD51C、RAD51D、またはRAD51Lのうちの少なくとも1つに機能喪失変異を含むゲノム配列に由来するHRD陽性対照を含むHRD陽性試験データセットを使用して、訓練されたモデルを試験するステップをさらに含む。いくつかの実施形態では、訓練するステップは、HRD陽性訓練データセットおよびHRD陰性訓練データセットを使用するステップを含む。いくつかの実施形態では、方法は、HRDモデルを訓練する前に、1つまたは複数のプロセッサによって、HRD陽性訓練データセットとHRD陰性訓練データセットとのバランスをとるステップを含む。

20

【0027】

記載された方法のいくつかの実施形態では、方法は、1つまたは複数のプロセッサによって、コンセンサスなヒトゲノム配列に由来するHRD陰性対照を含むHRD陰性訓練データセットを含むHRD陰性試験データセットを使用して訓練されたモデルを試験するステップをさらに含む。いくつかの実施形態では、訓練するステップは、HRD陽性訓練データセットおよびHRD陰性訓練データセットを使用するステップを含む。いくつかの実施形態では、方法は、HRDモデルを訓練する前に、1つまたは複数のプロセッサによって、HRD陽性訓練データセットとHRD陰性訓練データセットとのバランスをとるステップを含む。

30

【0028】

記載された方法のいくつかの実施形態では、対象における腫瘍が、前立腺がん、非小細胞肺癌（NSCLC）、結腸直腸がん（CRC）、卵巣がん、乳がん、または膵臓がんである。

40

【0029】

記載された方法のいくつかの実施形態では、HRDモデルを訓練するステップが、卵巣がん、非小細胞肺癌（NSCLC）、結腸直腸がん（CRC）、乳がん、膵臓がん、または前立腺がんに関連するサンプルデータにHRDモデルを適合させることを含み、サンプルデータが複数の特徴のサブセットを含む。

【0030】

記載された方法のいくつかの実施形態では、腫瘍は、固形組織生検サンプルであるサンプルから得られる。いくつかの実施形態では、固形組織生検サンプルがホルマリン固定パラフィン包埋（FFPE）サンプルである。記載された方法のいくつかの実施形態では、

50

腫瘍は、循環腫瘍DNA（ctDNA）を含む液体生検サンプルであるサンプルから得られる。記載された方法のいくつかの実施形態では、腫瘍は、無細胞DNA（cfDNA）を含む液体生検サンプルであるサンプルから得られる。

【0031】

記載された方法のいくつかの実施形態では、方法は、患者に関連する診断値として腫瘍の出力をHRD陽性またはHRD陰性と判定、識別、または適用するステップをさらに含む。記載された方法のいくつかの実施形態では、方法は、腫瘍のHRD陽性またはHRD陰性としての出力に基づいて、対象のゲノムプロファイルを生成するステップをさらに含む。いくつかの実施形態において、方法は、生成されたゲノムプロファイルに基づいて、抗がん剤を投与するステップ、または抗がん治療を対象に適用するステップをさらに含む。記載された方法のいくつかの実施形態では、腫瘍のHRD陽性またはHRD陰性としての出力が、対象のゲノムプロファイルを生成するのに使用される。記載された方法のいくつかの実施形態では、腫瘍のHRD陽性またはHRD陰性としての出力が、対象に対する提案された治療の決定を行う際に使用される。記載された方法のいくつかの実施形態では、腫瘍のHRD陽性またはHRD陰性としての出力が、対象に治療を適用または投与するのに使用される。

10

【0032】

記載された方法のいくつかの実施形態では、HRDモデルは機械学習モデルである。

【0033】

記載された方法のいくつかの実施形態では、対象は、がんを有するか、がんを有するリスクがあるか、またはがんを有する疑いがある。

20

【0034】

対象のがんを治療する方法であって、（a）上記の任意の方法に従って腫瘍をHRD陽性またはHRD陰性として識別するステップ、（b）がんの腫瘍がHRD陽性と評価された場合、治療有効量のHRD陽性腫瘍に有効な薬物を対象に投与するステップ、を含む方法が、本明細書にさらに記載される。いくつかの実施形態では、HRD陽性腫瘍に有効な薬物は、白金系薬物またはPARP阻害剤である。いくつかの態様において、方法は、腫瘍がHRD陰性と評価された場合、白金系薬物でもPARP阻害剤でもない治療有効量の薬物を対象に投与するステップを含む。

【0035】

対象のがんの治療法を選択する方法であって、（a）上記の任意の方法に従ってHRD陽性またはHRD陰性としてがんの腫瘍を評価するステップ、（b）がんがHRD陽性と評価された場合、HRD陽性腫瘍に有効な療法を選択するステップ、を含む方法が、本明細書にさらに記載される。いくつかの態様では、方法は、腫瘍がHRD陰性と評価された場合、白金系薬物でもPARP阻害剤でもない治療法を選択するステップを含む。いくつかの実施形態では、HRD陽性腫瘍に有効な治療法は、白金系薬物またはPARP阻害剤である。

30

【0036】

コンピュータシステムであって、1つまたは複数のプロセッサと、メモリと、1つまたは複数のプログラムであって、1つまたは複数のプログラムがメモリに記憶され、1つまたは複数のプロセッサによって実行されるように構成され、1つまたは複数のプログラムが、請求項1から65のいずれか一項に記載の方法を実行するための指示を含む、1つまたは複数のプログラムと、を含む、コンピュータシステムが、本明細書にさらに記載される。

40

【0037】

1つまたは複数のプログラムを記憶する非一時的コンピュータ可読記憶媒体であって、1つまたは複数のプログラムが命令を含み、命令が、電子デバイスの1つ以上のプロセッサによって実行されると、電子デバイスに、上述の方法のいずれか1つを実行させる、非一時的コンピュータ可読記憶媒体が、本明細書にさらに記載される。

【図面の簡単な説明】

50

【0038】

【図1】対象におけるがんの腫瘍をHRD陽性（HRD（+））またはHRD陰性（HRD（-））として分類するための例示的なプロセスを示す。

【0039】

【図2】ANOVA、ランダムフォレスト、勾配ブースティング（例えば、XGB）、およびカイ二乗などの異なる特徴の重要度のメトリックを使用して評価され得る異なるタイプの特徴を示す。

【0040】

【図3A】例示的な特徴重複分析を示す。

【0041】

【図3B】例示的な特徴重複分析を示す。

【0042】

【図4】例示的な反復特徴選択プロセスを示す。

【0043】

【図5】例示的な反復特徴選択プロセスから得られたモデルのパフォーマンスの例示的なプロットを示す。

【0044】

【図6A】モデルのパフォーマンスを評価および調整するために使用され得る例示的な交差検証プロセスを示す。

【0045】

【図6B】複数のデータ要素の等しいサイズのサブセットへの例示的な分割を示す。

【0046】

【図7】対象のがんの腫瘍をHRD陽性（HRD（+））またはHRD陰性（HRD（-））として分類するように構成されたHRD分類モデルを訓練および操作するための例示的な方法を示す。

【0047】

【図8】ロジスティック回帰、勾配ブースティング（例えば、XGB）、およびランダムフォレストを使用した異なる機械学習モデルに対するHRDスコア分布の例を示す。

【0048】

【図9】HRDおよび／またはBRCA1／2の変異の状態によって層別化されたサンプルにおける例示的なモデルのパフォーマンスを示す。左側は、「HRD野生型：真」（N＝245，050；図の右側の-1）、「HRD野生型：偽」（N＝30，799；図の右側の0）および真のHRD陽性サンプル（二対立遺伝子BRCA変異；N＝6，851；図の右側の1）と呼ばれるサンプル腫瘍のプールを示す。

【0049】

【図10】異なる腫瘍タイプ（乳がん、卵巣がん、膵がんおよび前立腺がん）における図9のサブセットからの例示的なモデルのパフォーマンスを示す。各腫瘍タイプについて、サブセットは、図9のサブセット-1、0、および1に対応する（すなわち、各がんに対するHRD野生型：真、HRD野生型：偽、および二対立遺伝子BRCA変異）。

【0050】

【図11】本明細書に記載の特定の方法で 사용할 ことができる、一実施形態によるコンピューティングデバイスの例を示す。

【発明を実施するための形態】

【0051】

発明の詳細な説明

相同組換え欠損（HRD）モデル（例えば、分類モデル）を訓練するための1つまたは複数の特徴の重要度のメトリックを使用して、複数の特徴のサブセットを識別するコンピュータ実装方法が本明細書に記載される。モデルは、対象における腫瘍のゲノムに関連する複数の特徴のサブセットに関する試験サンプルデータを受け取り、腫瘍をHRD陽性である可能性が高い、またはHRD陰性である可能性が高いものと識別（例えば、分類する

10

20

30

40

50

）するように構成される。HRD陽性である可能性が高い（HRD（＋））、またはHRD陰性である可能性が高い（HRD（－））ものとして、腫瘍、例えば、前立腺がん、卵巣がん、乳がん、結腸直腸がん、NSCLC、または膵臓がんの腫瘍を識別する（例えば、分類する）方法を、本明細書においてさらに記載する。HRD陽性（またはHRD陽性である可能性が高い）またはHRD陰性（またはHRD陰性である可能性が高い）としての腫瘍の識別に基づいて、膵臓がん、前立腺がん、卵巣がん、乳がん、非小細胞肺癌（NSCLC）、または結腸直腸がん（CRC）などであるがこれらに限定されないがんを治療する方法を、本明細書においてさらに記載する。

【0052】

特徴のサブセットを選択すると、モデルのオーバーフィッティングを減らすことができる。オーバーフィッティングは、モデルのスケラビリティを低減し、モデルがモデルを訓練するために使用されるデータの範囲外にあるシナリオを無視するので、不正確な分類（例えば、不正確なHRDの状態）をもたらす可能性があるため、問題がある。さらに、より高い特徴の重要度を有する特徴のサブセットを選択することによって、分類モデルは、より少ない訓練データで訓練することができ、より少ない入力データを必要とする。これは、より効率的なモデリングプロセスを可能にするだけでなく、モデルからのより広い範囲のサンプルからのより正確な分類も可能にする。さらに、入力される特徴の低減されたセットを有するモデルは、訓練するため、また分類タスクを実行するために、より少ない処理パワーを必要とし得る。したがって、特徴選択プロセスは、処理速度を改善し、コンピュータメモリおよび処理パワーの効率的な使用を可能にすることによって、コンピュータシステムの機能を改善する。さらに、特定の派生するコピー数の特徴および／または短い変異体の特徴から選択することによって、訓練されたモデルは、以前の方法と比較して、腫瘍をHRD陽性またはHRD陰性として識別するときに、より高い効率および精度（例えば、偽陽性／偽陰性が少ない）をもたらす。ヘテロ接合性の喪失、テロメア対立遺伝子不均衡、および大規模の移行などのHRDを評価する以前の方法は、本明細書に記載されている派生するコピー数の特徴および／または短い変異体の特徴の評価と比較して、ノイズおよび誤差にさらされやすい。腫瘍の適切な識別は、患者（対象）の治療を適切に選択できることに不可欠である。

【0053】

腫瘍形成は、部分的には、細胞のゲノムの体細胞変化の蓄積によって推進される。これらの変化の中には、多くのがんにおいて一般的であるコピー数の変化が含まれる。相同修復欠損経路に関与する特定の遺伝子における機能の喪失、機能の獲得、または遺伝子調節変異は、これらのコピー数の変化の蓄積をもたらす。しかしながら、BRCA1およびBRCA2などの特定の重要な遺伝子における変異以外に、HRD陽性状態をもたらす変異の正確な組み合わせは不明である。いくつかの腫瘍は、非ゲノム手段を介して、例えばBRCA1などのHRD関連遺伝子のプロモーターメチル化を介してHRD陽性となる。HRD関連遺伝子を配列決定する代わりに、代替的なアプローチは、特定のコピー数の変化の特徴、またはヘテロ接合性の喪失の特徴などのHRDの結果を識別および評価することである。しかしながら、HRD陽性ゲノムおよびHRD陰性ゲノムの両方がコピー数の変化を示し得るが、HRDの存在を示す正確な値および特徴の組み合わせは不明である。

【0054】

したがって、一態様では、本発明の方法は、HRD分類器プロセスを訓練および操作するために使用することができる特徴のサブセットを（より大きな複数の潜在的な特徴から）選択することに関する。別の態様では、本発明の方法は、一般に、コピー数の収差に対応する特徴などの特徴の評価に、少なくとも部分的に基づいて、HRD陽性である可能性が高い（HRD（＋））またはHRD陰性である可能性が高い（HRD（－））腫瘍を識別（例えば、分類）する手段に関する。この分類は、一般に、腫瘍がHRD陽性またはHRD陰性である可能性の評価に基づく。この評価に基づいて、HRD分類器プロセスはさらに、腫瘍をHRD陽性またはHRD陰性とみなすことができる。このように分類することおよび／またはみなすことは、腫瘍を有する患者の診断の値として使用され得る。

【0055】

腫瘍をHRD陽性である可能性が高いまたはHRD陰性である可能性が高いものと分類するための既存の方法は、特に野生型BRCA1およびBRCA2を有するHRD陽性腫瘍（「BRCAness」プロファイルを有する腫瘍、すなわち、関連するBRCA1／2変異を有せずにBRCA1／2変異腫瘍との類似性を示す腫瘍として記載されることがある）では、信頼性が低い、または不正確であることが多い。あるいは、すべての変異が、BRCA1／2の変化などの病理性変異でさえ、HRDをもたらすわけではない（例えば、いくつかの変異は、単一对立遺伝子の同乗者であり得る）。がんに関連する相同修復欠損は、腫瘍細胞ゲノムを傷つけ、コピー数（すなわち、コピー数の異常）および／またはインデルパターンの検出可能な変化をもたらす。これらのコピー数の異常および／またはインデルパターンの特定のパターン、分布および形態を使用して、腫瘍をHRD表現型クラスに分類することができる。本願は、様々な実施形態において、HRD陽性腫瘍を識別するために使用することができる他の潜在的特徴（本明細書で他の様式で記載されるような基本的特徴など）の中から、これらのパターン（すなわち、コピー数の特徴）およびインデルパターン（すなわち、短い変異体の特徴）に関連する特徴を選択する手段を提供する。

10

【0056】

本願はさらに、腫瘍を、HRD陽性である可能性が高いまたはHRD陰性である可能性が高いものとしてより確実に識別（例えば、分類する）することができ、任意選択で腫瘍をHRD陽性またはHRD陰性とみなすことができる、対象におけるがん性腫瘍のゲノムに関連する1つまたは複数のデータ特徴（1つまたは複数のコピー数の特徴および／または1つまたは複数の短い変異体の特徴など）に基づく具体的に構成されたモデルを提供する。対象におけるがんの腫瘍の識別（例えば、分類）は、腫瘍がどのように治療されるべきかを示す。例えば、セグメントサイズの特徴、シーケンシングリードの特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、振動コピー数のセグメントの数の特徴、およびセグメントマイナーアレル頻度の特徴のうちの1つまたは複数を含む、少なくとも1つまたは複数のコピー数の特徴を含む試験データを使用する訓練されたHRDモデルを使用して、試験用腫瘍をHRD陽性である可能性が高いまたはHRD陰性である可能性が高いと識別し（例えば、分類する）、また、尤度スコアに基づいて腫瘍をHRD陽性またはHRD陰性とみなすことができる。コピー数の特徴のこれらのカテゴリは、この識別に有用であると識別されている。特定のカテゴリの短い変異体の特徴がまた、この識別に有用であると識別されており、これには、例えば、ミクロ相同性または反復領域特徴における欠失（例えば、少なくとも5塩基対）および／または2つ以上の短い変異体の特徴を組み込んだ変異シグネチャが含まれるが、これらに限定されない。

20

30

【0057】

これらのコピー数の特徴の1つもしくは複数、および／またはこれらの短い変異体の特徴の1つもしくは複数と組み合わせ、他の特徴または測定が記載された方法において有用であり得、これらには、限定されないが、対象の年齢、がんの型、がんの病期、腫瘍の純度、腫瘍のゲノム倍数性、および／またはヘテロ接合性の腫瘍ゲノム喪失などの特定の基本的特徴が含まれる。

40

【0058】

対象におけるがんの腫瘍が、HRD陽性である可能性が高いまたはHRD陰性である可能性が高い、またはHRD陽性またはHRD陰性とみなされるものと識別され（例えば、分類され）たら、それを適切な治療で治療することができる。例えば、腫瘍がHRD陽性である可能性が高いと識別された場合、それはHRD陽性がんにも有効な薬物、例えば白金系薬物またはPARP阻害剤で治療することができる。

【0059】

定義

50

本明細書において使用される場合、単数形「1つの(a)」、「1つの(an)」、および「その(the)」は、文脈上他に明確に指示されない限り、複数の言及を含む。

【0060】

本明細書における値またはパラメータ「約」への言及は、その値またはパラメータ自体を対象とする変動を含む（および説明する）。例えば、「約X」という記載は、「X」という記載を含む。

【0061】

「がん」および「がん性」という用語は、典型的には調節されていない細胞の増殖を特徴とする哺乳動物における生理学的状態を指すまたは記載する。この定義には、良性および悪性のがんが含まれる。「早期がん」または「初期の腫瘍」とは、浸潤性でも転移性でもないか、ステージ0、1または2のがんとして分類されるがんを意味する。がんの例には、肺がん（例えば、非小細胞肺癌（NSCLC））、腎臓がん（例えば、腎尿路上皮癌）、膀胱がん（例えば、膀胱尿路上皮（移行細胞）癌）、乳がん、結腸直腸がん（例えば、結腸腺癌）、卵巣がん、膵臓がん、胃癌、食道がん、中皮腫、黒色腫（例えば、皮膚黒色腫）、頭頸部がん（例えば、頭頸部扁平上皮癌（HNSCC））、甲状腺がん、肉腫（例えば、軟部組織肉腫、線維肉腫、粘液肉腫、脂肪肉腫、骨肉腫（osteogenic sarcoma）、骨肉腫（osteosarcoma）、軟骨肉腫、血管肉腫、内皮肉腫、リンパ管肉腫、リンパ血管内皮肉腫、平滑筋肉腫または横紋筋肉腫）、前立腺がん、神経膠芽腫、子宮頸がん、胸腺癌、白血病（例えば、急性リンパ性白血病（ALL）、急性骨髄球性白血病（AML）、慢性骨髄性白血病（CML）、慢性好酸球性白血病、または慢性リンパ性白血病（CLL））、リンパ腫（例えば、ホジキンリンパ腫または非ホジキンリンパ腫（NHL））、骨髄腫（例えば、多発性骨髄腫（MM））、菌状息肉症、メルケル細胞がん、血液悪性腫瘍、血液組織のがん、B細胞がん、気管支がん、胃がん、脳または中枢神経系がん、末梢神経系がん、子宮または子宮内膜がん、口腔または咽頭のがん、肝臓がん、精巣がん、胆道がん、小腸または虫垂がん、唾液腺がん、副腎がん、腺がん、炎症性筋線維芽細胞性腫瘍、消化管間質腫瘍（GIST）、結腸がん、骨髄異形成症候群（MDS）、骨髄増殖性障害（MPD）、真性赤血球増加症、脊索腫、滑膜腫、ユーイング腫瘍、扁平上皮癌、基底細胞癌、腺癌、汗腺癌、脂腺癌、乳頭状癌、乳頭状腺癌、髄様癌、気管支原性癌、腎細胞癌、肝癌、胆管癌、絨毛癌、セミノーマ、胚性癌、ウィルムス腫瘍、膀胱癌、上皮癌、神経膠腫、星状細胞腫、髄芽細胞腫、頭蓋咽頭腫、上衣腫、髄膜腫、血管芽腫、聴神経腫瘍、乏突起膠腫、髄膜腫、神経芽腫、網膜芽細胞腫、濾胞性リンパ腫、びまん性大細胞型B細胞リンパ腫、マントル細胞リンパ腫、肝細胞癌、甲状腺がん、小細胞がん、本態性血小板血症、無色素性骨髄性化生、好酸球増加症候群、全身性肥満細胞症、よく知られている好酸球増加症、神経内分泌がん、またはカルチノイド腫瘍が含まれるが、これらに限定されない。

【0062】

本明細書で使用される腫瘍「腫瘍」は、悪性または良性にかかわらず、すべての新生物細胞の成長および増殖、ならびにすべての前がん性およびがん性の細胞および組織を指す。「がん」、「がん性」、および「腫瘍」という用語は、本明細書で言及される場合、相互に排他的ではない。

【0063】

「個体」、「患者」、および「対象」という用語は同義的に使用され、哺乳動物を指し、ヒト、ウシ、ウマ、ネコ、イヌ、げっ歯類、または霊長類が含まれるが、これらに限定されない。いくつかの実施形態では、対象は、ヒトである。

【0064】

本明細書で使用される場合、「有効量」または「治療有効量」という用語は、特定の障害、状態または疾患を治療する、例えばその症状の1つまたは複数を改善、緩和、軽減、および／または遅延させるのに十分な化合物、薬物、または組成物の量を指す。がんに関して、有効量は、対象に存在するがん細胞の数を、数および／またはサイズにおいて減少する、および／またはがん細胞の成長速度を遅くするのに十分な量を含む。いくつかの実

10

20

30

40

50

施形態では、有効量は、疾患の再発を予防する、または遅らせるのに十分な量である。がんの場合、化合物または組成物の有効量は、(i)がん細胞の数を減少させる；(ii)がん細胞の増殖を阻害する、遅延させる、ある程度遅らせる、好ましくは停止させる；(iii)がんの発生および／または再発を予防または遅延させる；および／または(iv)がんに関連する1つまたは複数の症状をある程度軽減することができる。

【0065】

本明細書で使用される場合、「治療」または「治療すること」は、臨床結果を含む有益なまたは所望の結果を得るためのアプローチである。本発明の目的のために、有益なまたは所望の臨床結果には、以下のうちの1つ以上が含まれるが、これらに限定されない：疾患に起因する1つ以上の症候の緩和、疾患の程度の縮小、疾患の安定化（例えば、疾患の悪化の予防または遅延）、疾患の蔓延（例えば転移）の予防または遅延、疾患の再発の予防または遅延、疾患の進行の遅延または遅延、疾患状態の改善、疾患の寛解（部分的または全体）の提供、疾患の処置に必要な1つ以上の他の薬物の用量の減少、疾患の進行の遅延、生活の質の向上、および／または生存期間の延長。がんに関して、対象に存在するがん細胞の数は、数および／またはサイズが減少し得る、および／またはがん細胞の成長速度が遅くなり得る。いくつかの実施形態では、治療は、疾患の再発を予防または遅延させ得る。がんの場合、治療は以下の通りであり得る：(i)がん細胞の数を減少させる；(ii)がん細胞の増殖を阻害する、遅延させる、ある程度遅らせる、好ましくは停止させる；(iii)がんの発生および／または再発を予防または遅延させる；および／または(iv)がんに関連する1つ以上の症状をある程度軽減する。本発明の方法は、これらの治療の態様のいずれか1つ以上を企図する。

【0066】

本明細書に記載される本発明の態様および変形は、態様および変形「からなる」および／または「から本質的になる」を含むことが理解される。

【0067】

値の範囲が提供される場合、その範囲の上限と下限との間の各介在値、およびその状態範囲内の任意の他の記載値または介在値は、本開示の範囲内に含まれることを理解されたい。記載された範囲が上限または下限を含む場合、それらの含まれる限界のいずれかを除外した範囲も本開示に含まれる。

【0068】

本明細書において使用されるセクションの見出しは、編成のみを目的としており、記載された主題を限定するものと解釈されるべきではない。説明は、当業者が本発明を製造および使用することを可能にするために提示され、特許出願およびその要件の文脈で提供される。記載された実施形態に対する様々な変更は、当業者にとって容易に明らかであり、本明細書の一般的な原理は、他の実施形態に適用されてもよい。したがって、本発明は、示された実施形態に限定されることを意図するものではなく、本明細書に記載された原理および特徴と一致する最も広い範囲が与えられるべきである。

【0069】

図は、様々な実施形態にかかるプロセスを示している。例示的なプロセスでは、いくつかのブロックは、任意に組み合わせられ、いくつかのブロックの順序は、任意に変更され、いくつかのブロックは、任意に省略される。いくつかの例では、例示的なプロセスと組み合わせて追加のステップが実行されることができる。したがって、図示されている（および以下により詳細に説明されている）動作は、本質的に例示的なものであり、したがって、限定するものとみなされるべきではない。

【0070】

本明細書で参照されるすべての刊行物、特許、および特許出願の開示は、それぞれ、参照によりその全体が本明細書に援用される。参照により援用される参考文献が本開示と矛盾する限り、本開示が優先するものとする。

【0071】

特徴の選択

本明細書で別途説明されるものを含む複数の特徴から開始して、複数の特徴のサブセットは、1つまたは複数の特徴の重要度のメトリックを使用して識別することができる。一般に、特徴の重要度のメトリックは、個々の特徴の評価を可能にし、どの特徴がHRDの評価に最も関連し得るかを判定する。例示的な特徴の重要度のメトリックには、勾配ブースティング（例えばXGB o o s t、XGBとしても知られる）、分散分析（AN O V A）、カイ二乗分析、およびランダムフォレストが含まれるが、これらに限定されない。個々の特徴は、これらの特徴の重要度のメトリックに基づいて値を割り当てることができ、特徴は、HRDモデルのパフォーマンスに対する寄与の増加に基づいて、重要度が増加して割り当てられる（例えば、腫瘍をHRD陽性またはHRD陰性として分類する際のモデルのパフォーマンスの改善）。次いで、閾値を超える特徴（複数の特徴のうちの中央値を超える特徴など）などのより重要な特徴を、HRDモデルを訓練または実行することを使用するために、選択することができる。特徴のサブセットが識別されると、特徴のサブセットを使用して、HRDモデル（例えば、分類モデル）を訓練することができる。次いで、HRDモデルを使用して、腫瘍から得られ、特徴を選択している間に識別された特徴の少なくとも一部を含む試験データを使用して、対象の腫瘍を識別（例えば、分類する）することができる。

10

【0072】

より高い特徴の重要度を有する特徴のこのサブセットを選択することにより、モデルは、より少ない訓練データで訓練することができ、より少ない入力データを必要とするようになり、したがってメモリの使用量および管理を改善する。さらに、入力される特徴の低減されたセットを有するモデルは、訓練するため、また識別（例えば、分類）タスクを実行するために、より少ない処理パワーを必要とする。したがって、特徴選択プロセスは、処理速度を改善し、コンピュータメモリおよび処理パワーの効率的な使用を可能にすることによって、コンピュータシステムの機能を改善する。

20

【0073】

図1は、いくつかの実施形態による、複数の特徴のサブセットを識別するためのブロックを含む、対象におけるがんの腫瘍をHRD陽性またはHRD陰性として分類するための例示的なプロセスを示す。いくつかの実施形態において、プロセス100は、例えば、ソフトウェアプログラムを実装する1つ以上の電子デバイスを使用して実行される。一部の実施例では、プロセス100は、クライアント-サーバシステムを使用して実行され、プロセス100のブロックは、サーバとクライアントデバイスとの間で、任意の方法で分けられる。他の実施例では、プロセス100は、クライアントデバイスのみ、または複数のクライアントデバイスのみを使用して実行される。プロセス100では、いくつかのブロックが、任意選択的に結合され、いくつかのブロックの順序が、任意選択的に変更され、いくつかのブロックが、任意選択的に省略される。一部の実施例では、プロセス100と組み合わせて追加のステップを実行することができる。したがって、図示されている（および以下により詳細に説明されている）動作は、本質的に例示的なものであり、したがって、限定するものとみなされるべきではない。

30

【0074】

図1のブロック102において、例示的なシステム（例えば、1つまたは複数の電子デバイス）は、複数の特徴を受信する。いくつかの実施形態では、システムは、複数のデータ要素を含むデータセットを受信する。データ要素は、複数の特徴および関連する分類標識（例えば、HRD陽性またはHRD陰性）に関するデータを含むことができる。例えば、データ要素は、特定の対象由来のサンプルの複数の特徴に関するデータ、およびサンプルがHRD陽性およびHRD陰性であるかどうかを示す関連する分類標識を含むことができる。特徴は、基本的特徴、コピー数の特徴、および／または短い変異体の特徴（例えば、塩基置換またはインデル（挿入または欠失）に対応する特徴）として分類された特徴を含むことができる。基本的特徴には、限定するものではないが、データが得られた患者の年齢、がんの型、がんの病期、腫瘍の純度、腫瘍のゲノム倍数性、およびヘテロ接合性の腫瘍ゲノム喪失（ヘテロ接合性喪失下のゲノムの割合など）に関する特徴が含まれ得る。

40

50

コピー数の特徴には、セグメントサイズの特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレークポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレークポイント数の特徴、振動コピー数のセグメントの数、およびセグメンマイナーアリアル頻度の特徴が含まれ得るが、これらに限定されない。短い変異体の特徴は、例えば、ホモポリマーまたは反復領域特徴での欠失（例えば、少なくとも5塩基対）、および／または2つ以上の短い変異体の特徴を組み込んだ変異シグネチャを含み得るが、これらに限定されない。いくつかの実施形態では、特徴のうちの1つまたは複数のビニングされた特徴であり、値は、2分位数、3分位数、4分位数、5分位数、6分位数、7分位数、または任意の他の適切なビニング構成などのビンにソートされる。

10

【0075】

図1のブロック104において、システムおよび方法は、複数の特徴（すなわち、基本的特徴、コピー数の特徴、および／または短い変異体の特徴）から特徴のサブセットを選択する。選択された特徴のサブセットは、対象のがんの腫瘍を、HRD陽性またはHRD陰性として分類するための比較的高い予測値を有し得る。いくつかの実施形態では、比較的低い予測値を有するおよび／または冗長である特徴は、ブロック104において特徴のサブセットから除外することができる。いくつかの実施形態では、特徴の予測値は、特徴の重要度のメトリックを使用して定量化することができる。いくつかの実施形態では、特徴の重要度のメトリックを適用して、複数の特徴の各特徴の特徴重要度スコアを取得することができる。特徴の特徴重要度スコアは、特徴と分類標識との統計的相関から得られる（例えば、HRD陽性またはHRD陰性）。特徴と分類標識との間の統計的相関は、特徴が分類タスクに対してどれだけの予測値を有するかに基づいて解釈され得る。換言すれば、例えば、特徴と分類標識との間のより高い統計的相関を有することによって、より高い特徴重要度スコアを達成することができ、これは、特徴が分類標識を予測する際に、より重要な役割を果たすことを示すことができる。より高い特徴の重要度を有する特徴を使用することにより、分類モデルをより少ないデータで訓練することができ、したがって訓練プロセスに多大な程度の有効性をもたらす、コンピュータリソースに対する制約（例えば、メモリ使用量、処理速度など）を少なくする。例えば、入力される特徴の低減されたセットを有するモデルは、訓練するため、また分類タスクを実行するために、より少数の処理リソースを必要とし得る。最後に、入力される特徴のセットが低減したモデルは、より少ないノイズを示し、オーバーフィッティングを回避することができる。したがって、特徴選択プロセスは、訓練プロセスの全体的な有効性を改善し、処理速度を改善し、コンピュータメモリおよび処理パワーの効率的な使用を可能にすることによって、コンピュータシステムの機能性を改善する。

20

30

【0076】

いくつかの実施形態では、システムは、ブロック104aによって示されるように、特徴重複分析を実行することによって、図1のブロック102で受信された複数の特徴から特徴のサブセットを選択する。ブロック104aにおいて、各々の特徴の重要度のメトリックは、ブロック102から受信した複数の特徴の特徴重要度スコアを計算するために使用される。各々の特徴の重要度のメトリックについて、システムは、それらの特徴重要度スコアに従って複数の特徴をランク付けすることができる。したがって、システムは、複数の特徴重要度特徴に対応する複数の特徴ランキングを取得することができる。そのとき、システムは、複数のランキングに基づいて特徴のサブセットを識別することができる。特徴をランク付けし、特徴のサブセットを識別するプロセスは、以下でより詳細に説明される。

40

【0077】

いくつかの実施形態では、異なるタイプの特徴は、異なる特徴の重要度のメトリックを使用して評価することができる。図2は、いくつかの実施形態による、ブロック104aにおいて複数の特徴をランク付けするために使用され得る複数の特徴の重要度のメトリックを示す。図示された例示的な特徴の重要度のメトリックは、ANOVA、ランダムフォ

50

レスト、勾配ブースティング（例えば、XGB）、およびカイ二乗を含む。さらに、ANOVAを使用して、複数の特徴の数値的な特徴を評価し、数値的な特徴のランキングを得ることができる。カイ二乗を使用して、複数の特徴のカテゴリ的な特徴を評価し、カテゴリ的な特徴のランキングを得ることができる。ランダムフォレストを使用して、複数の特徴のすべてを評価し、すべての特徴をランク付けすることができる。同様に、勾配ブースティング（XGBなど）を使用して、複数の特徴のすべてを評価し、すべての特徴をランク付けすることができる。

【0078】

いくつかの実施形態では、特徴の重要度のメトリックは、分散分析（ANOVA）モデルを含む。ANOVAは、数値入力変数が分類対象変数と比較されるときに群間で等しい分散があるかどうか（すなわち、HRD陽性またはHRD陰性）を評価する。群間に等しい分散がある場合、特徴は応答に影響せず、モデルの訓練のために考慮されなくてもよい。分散の値（f値）に基づいて、特徴をランク付けすることができ、例えば中央値を超える特徴を、モデルの有用な特徴として選択することができる。

10

【0079】

いくつかの実施形態では、特徴の重要度のメトリックは、カイ二乗分析を含む。特徴選択のために、カイ二乗解析は、予想される計数（すなわち、特徴が出力から独立している場合）と観測された計数とがどれほど互いに逸脱するかを検証する。特徴のより高いカイ二乗値は、それが応答変数にいっそう依存し、したがってより重要であることを示す。カイ二乗解析を使用して、特徴をランク付けすることができ、例えば中央値を上回る特徴を、モデルの有用な特徴として選択することができる。

20

【0080】

いくつかの実施形態では、特徴の重要度のメトリックは、ランダムフォレスト分析を含む。特徴選択の間、各ツリーについて、データの袋外の部分の予測精度が記録される。このプロセスは、各予測子変数を並べ替えた後に繰り返される。次いで、2つの精度の差がすべてのツリーにわたって平均化され、標準誤差によって正規化される。

【0081】

いくつかの実施形態では、特徴の重要度のメトリックは、勾配ブースティング分析（例えば、`extreme gradient boosting`（XGB）分析）を含む。XGBなどの勾配ブースティングは、モデルに対する各特徴の利得の寄与を検証する。ブー

30

ーストされたツリーモデルでは、各ツリーの各特徴の各利得が考慮され、次いで特徴寄与ごとの平均が評価される。次いで、最も高いパーセンテージの誘因特徴を選択することができる。

【0082】

図1のブロック104aにおいて、複数の特徴が特徴の重要度のメトリックに従ってランク付けされた後、システムは、複数のランキングを使用して特徴のサブセットを選択する。特徴のサブセットを選択する例示的なプロセスは、以下の図3Aおよび図3Bでさらに詳細に説明される。

【0083】

図3Aは、いくつかの実施形態による例示的な特徴重複分析を示す。図2で上述したように、複数の特徴をランク付けするために、複数の特徴の重要度のメトリックが使用されてもよい。図3Aの例では、例示的なプロセスは、特徴をランク付けするためにANOVA、ランダムフォレスト、および勾配ブースティング分析を使用する。しかしながら、当業者は、当技術分野で知られている他の学習技術も同様に使用することができることを理解するであろう。しかしながら、図3Aの例示的な目的のために、ANOVA特徴ランキング302は、最高ランクの特徴として特徴1、4、5、および8を含む。ランダムフォレストランキング304は、最高ランクの特徴として特徴8、2、3、および1を含む。勾配ブースティングランキング306は、最高ランクの特徴として特徴6、1、4、および2を含む。いくつかの実施形態では、他の特徴の重要度のメトリックを使用して、特徴を評価することができる。いくつかの実施形態では、特徴を評価するために、3つよりも

40

50

少ないまたは多いメトリックを使用することができる。いくつかの実施形態では、4つを超える特徴が上位の特徴とみなされてもよく、例えば、5つを超える、6つを超える、7つを超える、8つを超える、9つを超える、10を超える、11を超える、12を超える、13を超える、14を超える、15を超える、16を超える、17を超える、18を超える、19を超える、20を超える、21を超える、22を超える、23を超える、24を超える、または25を超える特徴のいずれかが上位の特徴とみなされてもよい。

【0084】

特徴がランク付けされると、システムは、特徴重複分析を実行して、1つまたは複数のメトリックが上位の特徴として識別された特徴を判定することができる。図3Aの例では、特徴重複分析308は、特徴1を、ANOVA特徴ランキング302、ランダムフォレストランキング304、および勾配ブースティングランキング306において識別された上位の特徴として識別する。特徴重複分析308はまた、特徴2、4、および8を、2つのメトリックによって識別された上位の特徴として識別する。いくつかの実施形態では、特徴重複分析308は、すべてのメトリックによって上位と識別された特徴を出力することによって、特徴のサブセットを出力することができる。いくつかの実施形態では、特徴重複分析308は、1つまたは複数のメトリックによって上位であると識別された特徴を出力することによって、特徴のサブセットを出力することができる。いくつかの実施形態では、特徴重複分析308は、グラフに表されてもよい。いくつかの実施形態では、特徴重複分析308は、特徴のサブセットを含むリストを出力することができる。

【0085】

図3Bは、いくつかの実施形態による、対象のがんの腫瘍をHRD陽性またはHRD陰性として分類するために使用される特徴の特徴選択プロセスの例示的な出力310を示す。特徴の重要度ランキング312がグラフで示されており、各グラフは、特定の特徴の重要度のメトリックによる特徴のランキングを示している。各グラフ（ANOVA、ランダムフォレスト、および勾配ブースティング）において、各ドットは特徴を表し、そのy軸の値は、特徴の重要度のメトリックによって計算されたその特徴の重要度に対応する。図3Bの例では、特徴重複分析314は、各特徴の重要度のメトリックに従って上位の特徴を含むことができる。図示のように、特徴重複分析は、メトリックのすべておよび／またはメトリックの一部によって高くランク付けされた特徴を識別することができる。

【0086】

図1に戻ると、いくつかの実施形態では、システムおよび方法は、プロセス104aに加えて、またはプロセス104aの代わりに、反復特徴選択プロセス104bを使用して複数の特徴のサブセットを判定することができる。ブロック104bで、システムは、図4で後述するように、1つまたは複数の特徴の重要度のメトリック（例えば、勾配ブースティング）を使用して特徴を評価し、次いで反復特徴選択プロセスを実行して特徴セットを徐々に拡張する。

【0087】

図4は、いくつかの実施形態による、図1のブロック104bによって使用され得る反復特徴選択プロセスを示す。ブロック402において、システムは、複数の特徴（例えば、図1のブロック102で受信された複数の特徴）を有するデータセットを受信する。

【0088】

図4のブロック404において、システムは、ブロック402において受信された特徴を、1つまたは複数の特徴の重要度のメトリック（例えば、勾配ブースティング）を使用して評価する。次いで、システムは、それらの対応する特徴の重要度のメトリックのスコアに従って特徴をランク付けすることができる。

【0089】

図4のブロック408において、システムおよび方法は、新しい特徴セットを取得する。最初の反復では、システムは、ブロック404によって判定された最高ランクの特徴を特徴セットに含めることによって、新しい特徴セットを取得することができる。後続の反復では、システムは、新しい特徴セットを取得するために、ブロック404によって判定

された次の最高ランクの特徴を追加することによって既存の特徴セットを拡張することができる。システムは、新しい特徴セットに基づいて訓練データセットをさらに取得する。訓練データセットは、複数のデータ要素を含むことができ、各データ要素は、新しい特徴セットおよび対応する分類標識（例えば、HRD陽性またはHRD陰性）に関連するデータを含む。例えば、データ要素は、サンプルからの新しい特徴セットの特徴およびサンプルの対応する分類標識（例えば、HRD陽性またはHRD陰性）に関するデータを含むことができる。

【0090】

図4のブロック410において、システムおよび方法は、ブロック408からの訓練データセットを使用して新しい分類モデルを訓練および評価する。システムは、モデルの訓練および評価で使用された特徴のリストに関連してモデルのパフォーマンスを記録する。いくつかの実施形態では、分類モデルの訓練および評価は、図6Aおよび図6Bによって以下でさらに説明するように、交差検証方法を使用して実行することができる。いくつかの実施形態では、分類モデルの訓練および評価は、ブロック408からのデータセットの別個のサブセットを使用することができる。

10

【0091】

いくつかの実施形態では、図4のブロック408および410は、ブロック402で受信されたすべての特徴がデータに含まれるまで反復される。各反復において、ブロック408は、次の最高ランクの特徴をデータセットに追加する。例えば、第1の反復では、ブロック408は、最高ランクの特徴を含む特徴セットと、対応する訓練セットとを出力する。第2の反復では、ブロック408は、2つの最高ランクの特徴を含む特徴セットと、対応する訓練セットとを出力する。第3の反復において、ブロック408は、3つの最高ランクの特徴を含む特徴セットおよび対応する訓練セットを出力する、などである。各反復において、ブロック410はそのとき、ブロック406からの訓練データセットを使用して、新しい分類モデルを訓練および評価する。システムは、条件が満たされるまでブロック408および410を繰り返す。いくつかの実施形態では、条件はブロック412を含み、システムは、追加される特徴がこれ以上ないと判定する（例えば、ブロック402で受信されたすべての特徴は、ブロック410で分類モデルを訓練および評価するために使用されるデータセットに含まれる）。いくつかの実施形態では、条件は、新しい分類モデルのパフォーマンスが閾値を超えるという判定を含む。この反復プロセスにより、システムは、ブロック402で受信されたすべての特徴が分類モデルを訓練し、パフォーマンスを評価するために使用されるまで、最高ランクの特徴、上位2つの最高ランクの特徴、上位3つの最高ランクの特徴などで訓練および評価されたときに分類モデルのパフォーマンスを記録することができる。記録されたパフォーマンスデータの例を図5に示す。

20

30

【0092】

図4のブロック414において、システムおよび方法は、ブロック410からの記録されたモデルのパフォーマンスを利用して、分類モデルのパフォーマンスを最適化する特徴の最小サブセットを判定する。いくつかの実施形態では、システムは、特徴の最小サブセットを判定して、追加の特徴の追加がモデルのパフォーマンスを実質的に改善しないようにすることができる。いくつかの実施形態では、システムは、分類モデルのパフォーマンスが特定の所定の閾値を超えるように、特徴の最小サブセットを決定することができる。特徴のサブセットは、ブロック414において出力される。

40

【0093】

図5は、図4のブロック410で判定されたモデルのパフォーマンスの例示的なプロットを示す。図5に示す例では、横軸は、分類モデルの訓練および評価に使用されるデータに含まれる上位の特徴の数を示す。縦軸はモデルのパフォーマンスを示す。いくつかの実施形態では、モデルのパフォーマンスは、受信者動作特性（ROC）曲線下面積（AUC）を使用して評価することができる。図5の例では、ブロック416において、26個の最高ランクの特徴が特徴のサブセットとして出力されると判定することができるが、各々の追加された特徴によるモデルのパフォーマンスの相対的な増加の変化に基づいて、より

50

少ない数の特徴を選択することができる。

【0094】

図6Aは、いくつかの実施形態による、モデルのパフォーマンスを評価するために使用され得る例示的な交差検証プロセスを示す。いくつかの実施形態では、図4のブロック410でプロセス600を使用して、モデルのパフォーマンスを評価することができる。ブロック602において、システムは、複数のデータ要素を受信することができる。複数のデータ要素の各々は、1つまたは複数の特徴および既知の分類標識を含むことができる。ブロック604において、システムは、ブロック602からの複数のデータ要素を、 n 個の等しいサイズのサブセットに分ける。ブロック606において、システムは、ブロック604からのサブセットのうちの1つを「ホールドアウト」セットとしてホールドアウトする。ブロック608において、システムは、ホールドアウトされていないすべてのデータ要素（例えば、「ホールドアウト」セットではない $n-1$ 個のサブセットからのデータ要素）についてモデルを訓練する。ブロック610において、システムは、ブロック608からのモデルへの入力として設定された「ホールドアウト」からのデータ要素の特徴を使用する。モデルは、データ要素の特徴に対応する複数の予測分類標識を生成する。次いで、予測された分類標識は、「ホールドアウト」セットの既知の分類標識と比較され、「ホールドアウト」セットに対するモデルのパフォーマンスを評価する。ブロック606、608、および610は、ブロック604からのすべての n 個のサブセットが「ホールドアウト」セットとして一度使用されるまで反復される。すなわち、ブロック606、608、および610は n 回反復され、異なるサブセットが各反復の「ホールドアウト」セットとして使用される。最後に、ステップ612において、ブロック610のすべての n 回の反復からのパフォーマンスが平均されて、パフォーマンスの平均が出力される。

10

20

【0095】

図6Bは、いくつかの実施形態による、複数のデータ要素の5つの等しいサイズのサブセットへの例示的な分割を示す。図6Bは、 $n=5$ である図6Aの例であり得る。複数のデータ要素622は、図6Aのブロック602からの複数のデータ要素の例であり得る。図6Bの例では、複数のデータ要素622は、セット1、セット2、セット3、セット4、セット5に分けられる。反復1の623では、複数のデータ要素622において、セット1が、ブロック606によって説明されるような「ホールドアウト」のデータセットとして使用され得る。モデルは、ブロック608によって説明されるように、セット2、セット3、セット4、およびセット5で訓練され得る。次いで、モデルのパフォーマンスを「ホールドアウト」データセット1で評価することができる。次いで、このプロセスはさらに4回繰り返される。反復2の624では、セット2が「ホールドアウト」セットであり、モデルはセット1、セット3、セット4、およびセット5で訓練され、モデルのパフォーマンスはセット2で評価される。反復3の626では、セット3が「ホールドアウト」セットであり、モデルはセット1、セット2、セット4、およびセット5で訓練され、モデルのパフォーマンスはセット3で評価される。反復4の628では、セット4が「ホールドアウト」セットであり、モデルはセット1、セット2、セット3、およびセット5で訓練され、モデルのパフォーマンスはセット4で評価される。反復5の630では、セット5が「ホールドアウト」セットであり、モデルはセット1、セット2、セット3、およびセット4で訓練され、モデルのパフォーマンスはセット5で評価される。図6Bの例では、平均のパフォーマンスは、反復1の622、反復2の624、反復3の626、反復4の628、および反復5の630からのモデルのパフォーマンスの平均であり得る。

30

40

【0096】

図1に戻ると、ブロック106において、システムは、ブロック104の特徴の選択によって判定されるように、選択された特徴のサブセットを取得する。分類モデル108は、選択された特徴106および標識された訓練データ110からの情報を使用して訓練される。いくつかの実施形態では、特徴の選択104に使用されるデータセットは、標識訓練データ110である同じデータセットである。いくつかの実施形態では、特徴の選択104に使用されるデータセットは、標識訓練データ110とは異なるデータセットである

50

。分類モデルを訓練するプロセスは、以下のセクションおよび図7で後述される。分類モデル108が訓練されると、対象のがんの未知の腫瘍由来の特徴（例えば、ブロック102で受信されたデータに含まれず、既知の分類標識に関連付けられていないデータ要素）をモデル108に入力して、対象のがんの腫瘍がHRD陽性である可能性が高いまたはHRD陰性である可能性が高いものかどうかを予測することができる。

【0097】

データの特徴

識別されている（例えば、分類されている）腫瘍からの試験サンプルを対象から得ることができる。試験サンプルに関連する基本的特徴、コピー数の特徴、および／または短い変異体の特徴などの特徴は、HRD分類モデルの入力として使用することができる1つまたは複数の特徴を含む。HRD分類モデルは、HRD陽性サンプル（腫瘍サンプルなど）に関連するHRD陽性データ、およびHRD陰性サンプル（腫瘍サンプルなど）に関連するHRD陰性データからの対応する特徴（例えば、基本的特徴、コピー数の特徴、および／または短い変異体の特徴）に基づいて訓練される。特徴は、HRDに関連する「BRCAness」プロファイルを有する腫瘍を識別するのに助けることができるHRDの機能的読み出しとして、使用することができる。そのようなHRD陽性表現型を有する腫瘍は、HRD陰性表現型では有効ではない（または、えてして有効ではない）特定の薬物療法の適切な候補であり得る。

【0098】

コピー数の特徴は、セグメントサイズの特徴、シーケンシングリードの数の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数を含み得るが、これらに限定されない。Macintyre et al., Copy-number signatures and mutational processes in ovarian carcinoma, Nat. Genet. 2018 Sep; 50(9): 1262-1270を参照されたい。混合モデル化を適用して、各特徴分布をガウス分布の混合またはポアソン分布の混合に分け、浮動またはバイナリ成分特徴を達成することができる。コピー数の特徴はまた、セグメントにおける生殖系列SNPのAおよびBのアリル頻度に基づくセグメントマイナーアリル頻度の特徴を含み得る。

【0099】

いくつかの実施形態では、HRDモデル（例えば、HRD分類器モデル）は、入力として使用されるよりも多くの特徴を使用して訓練することができる。例えば、HRD分類モデルは、それぞれがHRD陽性腫瘍および／またはHRD陰性腫瘍に関連する、とある数の特徴を含むHRD陽性データおよびHRD陰性データに基づいて、訓練することができる。そのとき、HRD分類モデルに入力されるデータは、より少ない特徴を含むことができる。HRD分類器モデルは、一例では、訓練されたHRD分類器モデルに入力されるサンプルデータから省略されたデータの特徴の重みを調整することができる。さらに、HRD分類器モデルは、追加のデータの特徴（例えば、それぞれ本明細書に記載されるような、ヘテロ接合性および／または1つもしくは複数の短い変異体の特徴のゲノム全体の喪失の尺度など）を使用して訓練することができるが、いくつかの実施形態では、データの入力は、対象のがんに関連する腫瘍のゲノムに関連する1つまたは複数のコピー数の特徴のみを含み得る。

【0100】

コピー数の特徴、gLOHおよび腫瘍ゲノム倍数性の測定値を含む基本的特徴、および／または短い変異体の特徴を含むゲノムデータの特徴を得るために、腫瘍の少なくとも1つのゲノムの少なくとも一部の配列決定によって、配列決定データが収集される。絶対コピー数または相対コピー数およびセグメント化は、その後、シャローホールゲノムシーケンシング（sWGS）データなどの全ゲノムシーケンシングデータから導き出すことができる。円形バイナリセグメンテーション（CBS）はまた、DNAマイクロアレイデータ

に基づいてゲノムを一定の総コピー数のセグメントに分けるために使用され得、そこからコピー数の特徴が導出され得る。あるいは、絶対コピー数およびセグメント化は、エクソームシーケンシング（ES）またはSNPアレイを含むがこれらに限定されない、当技術分野で公知の任意の技術から誘導することができる。コピー数の特徴の分布は、WGSデータなどの絶対コピー数データから計算することができる。混合モデル化を適用して、各特徴分布をガウス分布の混合またはポアソン分布の混合に分け、浮動またはバイナリ成分特徴を達成することができる。したがって、HRD分類モデルを訓練するために、または訓練されたHRD分類モデルに入力されるために使用される特定の「コピー数の特徴」は、その構成要素の特徴として表現される。例えば、セグメントサイズのコピー数の特徴の場合、 z 個の構成要素に分けられると、そのとき、HRD分類モデルを訓練するために使用され得る、またはHRD分類モデルを実行するために使用され得る z 個の可能な特徴が存在する。言い換えれば、特定の試験サンプルについて、「セグメントサイズ」のカテゴリでの「コピー数の特徴」（セグメントサイズが z 個の構成要素に分けられたと仮定する）は、HRD分類モデルを訓練するためであるか実行するためであるかにかかわらず、 z 個の可能な入力を有する。 z が 3 に等しい場合、3つのセグメントサイズの特徴のうちの少なくとも1つがHRD分類モデルに入力され得る、すなわち、`seg size 1`、`seg size 2`、または`seg size 3`である。最適なモデルのパフォーマンスは、部分的には、特徴の各々の特定のカテゴリに対して選択された構成要素の特徴の数に依存し得る。しかしながら、特徴の特定のカテゴリは、任意の適切な数の構成要素の特徴に分けることができ、必ずしも特定の確率分布に対応するものではない。したがって、モデルは、パフォーマンスが最適ではなくても、より多くのまたはより少ない数の構成要素の特徴で良好に機能し、効率的に検証することができる。

10

20

【0101】

コピー数の特徴を導出するとき、絶対コピー数データは、最初に、コピー数変異イベントをコールするベースラインレベルを判定するために、通常のデータセットとのマッチングによって正規化され得る。正常のパネルは、典型的には、健康な組織サンプルに由来する（腫瘍が由来するのと同じ個体に由来し得る）。健康な組織サンプルの分析は、本明細書に記載されているコピー数の特徴を導き出すためのベースラインコピー数を設定することを可能にする。

【0102】

記載されたコピー数の特徴のいくつかは、ゲノムのサブ領域にわたって評価され得る。例えば、特定のコピー数の特徴は、ゲノムのセントロメア部分にわたって評価され得る。別の例において、コピー数の特徴がゲノムのテロメア部分にわたって評価され得る。またさらなる例では、コピー数の特徴は、ゲノムのテロメア部分およびセントロメア部分の両方にわたって評価され得る。例示的な方法では、ゲノムのテロメアおよびセントロメア部分を定めるために、hg19などのヒト参照配列ゲノムを使用して、各染色体アームの開始部および終了部を定めることができる。次に、特定のアームの長さを2で除算して中間点を定める。コピー数の特徴について解析された各領域について、この中間点のセントロメア側にあるセグメントを、セントロメアセグメントと定める。この中間点のテロメア側にあるセグメントをテロメアセグメントと定める。セグメントが中間点（例えば、中間点のセントロメア側で始まりテロメア側で終わるセグメント）に及ぶ場合、そのセグメントは、セントロメアおよびテロメアの両方と呼称され得て、テロメアおよびセントロメアの両方のコピー数の特徴の評価に使用され得る。したがって、本明細書に記載されるデータの特徴のいずれも、必要に応じて、ゲノムのテロメア領域、ゲノムのセントロメア領域、またはゲノムのテロメア領域とセントロメア領域の両方にわたって評価され得る。

30

40

【0103】

コピー数のモデル化は、評価されているゲノムの推定された基本倍数性によって影響され得る。基本倍数性がより高く推定される場合、浮動小数点のコピー数の特徴が右にシフトされ得、ゆがんだ成分スコアおよび最終的に誤った分類をもたらし得る。コピー数のデータを基本倍数性に正規化することは、コピー数のデータを、評価されているゲノムの平

50

均倍数性で割ることを含む。したがって、記載されたコピー数の特徴のいずれかは、倍数性正規化コピー数データに由来し得、絶対コピー数は、腫瘍のゲノムの平均倍数性に対して正規化されている。平均倍数性を計算するための例示的な方法は、サンプルのすべてのセグメントについて加重平均コピー数を取得することである。平均倍数性を計算する例示的な方法については、Sun et al., A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal, PLoS Comput. Biol. 2018 Feb 7; 14 (2) : e1005965を参照されたい。

10

【0104】

本明細書に記載の特徴は、いくつかの実施形態では、ビンングされた特徴であってもよい。特徴ビンングは、特定の値を特定のカテゴリビンに編成することを含む。例えば、0から10の範囲の値を有する特徴の場合、四分位ビンングは、0から10までのこれらの値の各々を4つのビンのうちの1つに編成することができ、より低い値はより低いビンに編成することができ、より高い値はより高いビンに編成することができる。いくつかの実施形態では、ビンングは監視されていない。いくつかの実施形態では、ビンングは監視される。いくつかの実施形態では、ビンングは、等しい幅のビンングである。等しい幅のビンングでは、ピンはほぼ同じ幅の範囲を有する。例えば、1～8の値を有する特徴の場合、4つのピンを有する等しい幅のビンングは、1および2の値を第1のビンに編成し、3および4の値を第2のビンに編成し、以下同様である。いくつかの実施形態では、ビンングは等しい頻度のビンングである。等しい頻度のビンングでは、ピンは、各ピンがほぼ同じ数の値を有するように編成され、値はピンにほぼ等しく分配される。例えば、より低い値がはるかに高い頻度である1から10の値を有する特徴の場合、ビンングは、第1のビンに1、第2のビンに2、および第3のビンに3から10を編成することができる。ビンングは、2分位数、3分位数、4分位数、5分位数、6分位数、7分位数、または任意の他の適切なビンング編成であり得る。

20

【0105】

記載される方法のいずれかのいくつかの実施形態において、コピー数の特徴は、セグメントサイズの特徴を含む。セグメントサイズは、ゲノム全体にわたる各コピー数のセグメントのゲノム塩基における長さに由来する。例えば、セグメントがxのコピー数を有し、次のセグメントがyのコピー数を有する場合、コピー数xを有するセグメントの長さおよびコピー数yを有するセグメントの長さは、セグメントサイズのコピー数のカテゴリの因子である。例示的な実施形態では、セグメントサイズの分布は、10個の構成要素の特徴に分けられる。より小さい番号のセグメントサイズの特徴は、より小さいセグメントサイズ（例えば、segsize1）を表し、一方でより大きい番号のセグメントサイズの特徴は、より大きいセグメントサイズ（例えば、segsize10）を表す。いくつかの実施形態では、セグメントサイズの分布は、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、または少なくとも11の構成要素の特徴など、少なくとも5つの構成要素の特徴に分けられる。いくつかの実施形態では、セグメントサイズの分布は、5、6、7、8、9、10、または11個の構成要素の特徴のいずれかに分けられる。いくつかの実施形態において、セグメントサイズの特徴は、ゲノムのテロメア部分にわたって評価される。いくつかの実施形態において、セグメントサイズの特徴は、ゲノムのセントロメア部分にわたって評価される。いくつかの実施形態において、セグメントサイズの特徴は、ゲノムのテロメア部分およびセントロメア部分の両方にわたって評価される。いくつかの実施形態において、セグメントサイズの特徴は、ゲノム全体にわたって評価される。いくつかの実施形態において、セグメントサイズの特徴は、倍数性正規化コピー数データに由来する。いくつかの実施形態では、セグメントサイズの特徴はビンングされた特徴である。

30

40

【0106】

50

記載される方法のいずれかのいくつかの態様において、コピー数の特徴は、xメガベースあたりのブレイクポイント数の特徴を含む。いくつかの実施形態において、xは、約1メガベース(MB)～約150メガベースである。いくつかの実施形態において、xは、約10MB、約25MB、約50MB、約100MB、および約150MBのいずれかである。1セクションあたりのブレイクポイントの数は、ゲノムまたはゲノムの一部にわたる1セクションあたりのブレイクポイントの数を表す。例えば、10MBあたりのブレイクポイント数について、10MBの処理隣接ウィンドウ(または代替的に、スライドウィンドウ)がゲノム全体にわたって分析され、その後、スライディングウィンドウの各フレームについてのブレイクポイントの数が評価され得る。このアプローチでは、隣接ウィンドウが使用されたが、ブレイクポイント数を評価するのに適したスライドウィンドウまたはその他任意の技術が使用され得ることに留意されるべきである。それにもかかわらず、いくつかの例示的な実施形態では、1xメガベースあたりのブレイクポイント数は、3つの構成要素の特徴に分けられる。より小さい番号のブレイクポイント数の特徴は、より少ないブレイクポイントを表し(例えば、10MBあたりのブレイクポイント数: bp 10 MB 1の場合、10MBスライディングウィンドウの1フレームあたり、または10MB処理隣接ウィンドウの1フレームあたりのブレイクポイントが少ないことを示す)、一方、より大きい番号の特徴は、セクションあたりより多くのブレイクポイント(例えば、10MBあたりのブレイクポイント数 bp 10 MB 3の場合、bp 10 MB 1などのより低い番号の特徴と比較して、10MBのスライディングウィンドウのフレームあたりのより多くのブレイクポイントを示す)を表す。いくつかの実施形態では、ブレイクポイント数の分布は、少なくとも3つまたは少なくとも4つの構成要素の特徴などの、少なくとも2つの構成要素の特徴に分けられる。いくつかの実施形態では、セクションあたりのブレイクポイント数は、2、3、4、または5つの構成要素の特徴のいずれかに分けられる。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ゲノムのテロメアの部分にわたって評価される。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ゲノムのセントロメアの部分にわたって評価される。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ゲノム全体にわたって評価される。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴は、倍数性正規化コピー数データから導かれる。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ビニングされた特徴である。

10

20

30

【0107】

記載される方法のいずれかのいくつかの実施形態において、コピー数の特徴は、ゲノムセグメントの配列決定から得られるシーケンシングリードの数の特徴を含む。特定のゲノムセグメントについて、この値は、配列決定されたセグメントにアラインメントする(すなわち、「カバー」)シーケンシングリードの平均数を指す。異常に高いコピー数を有するゲノムセグメントの場合、シーケンシングリードの数が増加する。対照的に、コピー数を失ったゲノムセグメント(ホモ接合欠失など)については、シーケンシングリードが少なくなる。シーケンシングリードの特徴は、リードの実際の数(分析された各セグメントに対するリードの平均など)またはシーケンシングリードのビンとして表現され得る。より低い番号のシーケンシングリードの特徴はより低い絶対シーケンシングリードを表し、一方、より高い番号のシーケンシングリードの特徴はより高い絶対シーケンシングリードを表す。いくつかの実施形態において、シーケンシングリードの特徴は、ゲノムのテロメア部分にわたって評価される。いくつかの実施形態において、シーケンシングリードの特徴は、ゲノムのセントロメア部分にわたって評価される。いくつかの実施形態において、シーケンシングリードの特徴は、ゲノムのテロメア部分およびセントロメア部分の両方にわたって評価される。いくつかの実施形態において、シーケンシングリードの特徴は、倍数性正規化データに由来する。いくつかの実施形態において、シーケンシングリードの特徴はビニングされた特徴である。いくつかの実施形態において、シーケンシングリードの数の特徴は、次世代配列決定(NGS)からのリードの数の測定値である。いくつかの実

40

50

施形態において、シーケンシングリードの数の特徴は、対照におけるそのゲノムセグメントに対するシーケンシングリードの数と比較した、腫瘍サンプルのゲノムセグメントに対するシーケンシングリードの比として表される。

【0108】

記載される方法のいずれかのいくつかの実施形態において、コピー数の特徴は、絶対コピー数の特徴を含む。絶対コピー数は、各ゲノムセグメントについて計算され、値が割り当てられ得る。例えば、割り当てられた値は、0（ホモ接合性欠失を示す）、1（ヘテロ接合性欠失を示し得る）、2（正常なカウントであり得る）またはそれより多く（コピー数の増幅を示し得る）を含み得る。絶対コピー数の特徴は、実際のコピー数カウント（分析された各セグメントについてのコピー数の平均など）またはコピー数値のピンを表し得る。例えば、少なくとも6のコピー数は、セグメントについての高いコピー数を表すものとしてビンニングされ得る。3～5のコピー数は、中程度に増加したコピー数を表すとしてビンニングされ得る。1および2のコピー数は正常であり得、0のコピー数はホモ接合欠失としてビンニングされ得る。低い番号の絶対コピー数の特徴は低い絶対コピー数を表し、高い番号の絶対コピー数の特徴は高い絶対コピー数を表す。いくつかの実施形態において、絶対コピー数は、3、4、5、6、7、8または9個の構成要素の特徴のいずれかに分けられる。いくつかの実施形態において、絶対コピー数の特徴が、ゲノムのテロメア部分にわたって評価される。いくつかの実施形態において、絶対コピー数の特徴が、ゲノムのセントロメア部分にわたって評価される。いくつかの実施形態において、絶対コピー数の特徴が、ゲノムのテロメア部分およびセントロメア部分の両方にわたって評価される。いくつかの実施形態において、絶対コピー数の特徴は、倍数正規化データに由来する。いくつかの実施形態において、絶対コピー数の特徴は、ビンニングされた特徴である。

【0109】

記載された方法のいずれかのいくつかの実施形態において、コピー数の特徴は、変化点コピー数の特徴を含む。変化点コピー数は、ゲノム全体にわたるゲノムセグメント間のコピー数の絶対的な差を指す。例えば、コピー数7および2でモデル化された隣接するセグメントは、5の絶対差を有する。例示的な実施形態では、変化点コピー数の分布は、7つの構成要素の特徴に分けられる。より低い番号の変化点コピー数の特徴は、コピー数変化のより小さい絶対的な差（例えば、変化点1）を表し、一方、より高い番号の特徴は、コピー数の変化のより大きい絶対差（例えば、変化点7）を表す。いくつかの実施形態では、変化点コピー数の分布は、少なくとも5つ、少なくとも6つ、少なくとも7つ、または少なくとも8つの構成要素の特徴など、少なくとも4つの構成要素の特徴に分けられる。いくつかの実施形態では、変化点コピー数は、3、4、5、6、7、8または9個の構成要素の特徴のいずれかに分けられる。いくつかの実施形態において、変化点コピー数の特徴が、ゲノムのテロメア部分にわたって評価される。いくつかの実施形態において、変化点コピー数の特徴は、ゲノムのセントロメア部分にわたって評価される。いくつかの実施形態において、変化点コピー数の特徴が、ゲノムのテロメア部分およびセントロメア部分の両方にわたって評価される。いくつかの実施形態において、変化点コピー数の特徴は、倍数性正規化コピー数データに由来する。いくつかの実施形態では、変化点コピー数の特徴は、ビンニングされた特徴である。

【0110】

記載される方法のいずれかのいくつかの実施形態において、コピー数の特徴は、セグメントコピー数の特徴を含む。セグメントコピー数は、ゲノムまたはゲノムの一部にわたる各セグメントのコピー数から導かれる。例示的な実施形態では、セグメントコピー数の分布は、8つの構成要素の特徴に分けられる。より低い番号のセグメントコピー数の特徴は、より低いコピー数を表し（例えば、コピー数1は、0もしくは1、または0から1のコピー数レベルを表すことができる）、一方、より高い番号のコピー数の特徴は、より高いコピー数（例えば、コピー数8）を表す。いくつかの実施形態において、セグメントコピー数の分布は、少なくとも5つ、少なくとも6つ、少なくとも7つ、少なくとも8つまたは少なくとも9つの構成要素の特徴など、少なくとも4つの構成要素の特徴に分けられる

。いくつかの実施形態において、セグメントコピー数の分布は、4、5、6、7、8、9 または10個の構成要素の特徴のいずれかに分けられる。いくつかの実施形態において、セグメントコピー数の特徴が、ゲノムのテロメア部分にわたって評価される。いくつかの実施形態において、セグメントコピー数の特徴は、ゲノムのセントロメア部分にわたって評価される。いくつかの実施形態において、セグメントコピー数の特徴は、ゲノム全体にわたって評価される。いくつかの実施形態において、セグメントコピー数の特徴は、倍数性正規化コピー数データに由来する。いくつかの実施形態において、セグメントコピー数の特徴は、ビンニングされた特徴である。

【0111】

記載される方法のいずれかのいくつかの態様において、コピー数の特徴は、染色体アームあたりのブレークポイント数の特徴を含む。例示的な実施形態において、染色体アームあたりのブレークポイント数の分布は、5つの構成要素の特徴に分けられる。より低い番号の染色体アームあたりのブレークポイント数の特徴は、アームあたりのより少ないブレークポイントを表し（例えば、b p c h r a r m 1）、一方、より高い番号の染色体アームあたりのブレークポイント数の特徴は、染色体アームあたりより多いブレークポイントを表す（例えば、b p c h r a r m 5）。いくつかの実施形態において、染色体アームあたりのブレークポイント数の分布は、少なくとも4つの構成要素の特徴、少なくとも5つの構成要素の特徴、少なくとも6つの構成要素の特徴または少なくとも7つの構成要素の特徴などの少なくとも3つの構成要素の特徴に分けられる。いくつかの実施形態において、染色体アームあたりのブレークポイント数の分布は、4つ、5つ、6つ、7つまたは8つの構成要素の特徴のいずれかに分けられる。いくつかの実施形態において、染色体アームあたりのブレークポイント数は、倍数性正規化コピー数データから導かれる。いくつかの実施形態において、染色体アームあたりのブレークポイント数の特徴が、ビンニングされた特徴である。

【0112】

いくつかの実施形態において、コピー数の特徴は、振動コピー数のセグメントの数（o s C N）の特徴を有するいくつかのセグメントを含む。振動コピー数のセグメントの数は、2つのコピー数の間で繰り返される交互のセグメントの数をカウントするゲノムまたはゲノムの一部の横断を表す。例示的な実施形態において、振動コピー数のセグメントの数の分布は、3つの構成要素の特徴に分けられる。より低い番号の振動コピー数のセグメントの数の特徴は、2つのコピー数の間におけるより少ない反復された変化を表し（例えば、o s C N 1）、一方、より高い番号の振動コピー数のセグメントの数の特徴は、2つのコピー数の間におけるより多くの反復された変化を表す（例えば、o s C N 3）。いくつかの実施形態において、振動コピー数のセグメントの数の分布は、少なくとも2つ、例えば、少なくとも3つ、または少なくとも4つの構成要素の特徴に分けられる。いくつかの実施形態において、振動コピー数のセグメントの数の分布は、2つ、3つ、4つまたは5つの構成要素の特徴のいずれかに分けられる。いくつかの実施形態において、振動コピー数のセグメントの数の特徴が、ゲノムのテロメア部分にわたって評価される。いくつかの実施形態において、振動コピー数のセグメントの数の特徴が、ゲノムのセントロメア部分にわたって評価される。いくつかの実施形態において、振動コピー数のセグメントの数の特徴が、ゲノム全体にわたって評価される。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、倍数性正規化コピー数データに由来する。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、ビンニングされた特徴である。

【0113】

いくつかの実施形態において、コピー数の特徴は、セグメントマイナーアリル頻度（s e g M A F）の特徴を含む。s e g M A F 特徴は、腫瘍ゲノムの平均の s e g M A F、または中央値の s e g M A F のいずれかに由来し得る。ヘテロ接合アリル部位の正常なゲノムでは、各アリルの予想コピー数は1.0である。H R D は、アリルの完全な喪失（ヘテロ接合性の喪失）または一方のアリルの他方に対するコピー数の増加に関連する。したがって、s e g M A F は、メジャーアリルに対するマイナーアリルの比を比較する、セグメ

ントごとのゲノムの横断である。具体的には、各ヘテロ接合SNPをAアリルおよびBアリル頻度について分析する。マイナーアリルの頻度はマイナーアリル画分として取得される。平衡遺伝子座は約0.5:0.5の比を有し、マイナーアリル頻度は0.5である。ヘテロ接合性の喪失事象は、不均衡およびマイナーアリル頻度の約0.5未満へのゆがみをマイナーアリル画分について引き起こす。いくつかの実施形態において、segMAFの特徴は、ゲノムのテロメア部分にわたって評価される。いくつかの実施形態において、segMAFの特徴は、ゲノムのセントロメア部分にわたって評価される。いくつかの実施形態において、segMAFの特徴は、ゲノム全体にわたって評価される。いくつかの実施形態において、セグメントマイナーアリル頻度の特徴は、ビンニングされた特徴である。

10

【0114】

HRD分類モデルは、複数のHRD陽性腫瘍の各HRD陽性腫瘍について、HRD陽性腫瘍に関連する1つまたは複数の特徴およびHRD陽性標識を含むHRD陽性データと、複数のHRD陰性訓練腫瘍の各HRD陰性腫瘍について、HRD陰性腫瘍に関連する1つまたは複数のコピー数の特徴およびHRD陰性標識を含むHRD陰性データとによって訓練される。HRD分類モデルはまた、他の特徴または尺度に基づいて訓練されてもよい。したがって、これらの他の特徴または尺度を含む試験データを（1つまたはコピー数の特徴と組み合わせて含む）HRD分類モデルに入力することができる。例えば、ヘテロ接合性のゲノム損失の尺度、および／または1つもしくは複数の短い変異体の特徴を含む基本的特徴をHRD分類モデルで（HRD分類モデルを訓練するため、またはHRD分類モデルに入力される試験データとして）使用することができる。

20

【0115】

いくつかの実施形態では、基本的特徴は、腫瘍が得られた対象の年齢を含む。患者は、少なくとも5、少なくとも10、少なくとも15、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50、少なくとも55、少なくとも60、少なくとも65、少なくとも70、少なくとも75、または少なくとも80才のいずれかを含む任意の年齢であり得る。年齢の特徴は、対象の整数の値であり得る。あるいは、年齢の特徴は、乳児、若年、小児、若年成人、または高齢者の対象のいずれかなどの定性的特徴であってもよい。いくつかの実施形態では、年齢の特徴はビンニングされた特徴である。

30

【0116】

いくつかの実施形態では、基本的特徴はがんの型の特徴を含む。がんの型の特徴は、腫瘍の起源を指す。がんの型には、例えば、副腎、胆管、骨／軟部組織、乳房、結腸／直腸、食道、眼、頭頸部、腎臓、肝臓、肺、リンパ系、髄芽腫、中皮腫、骨髄系、神経系、神経内分泌、卵巣、膵臓、前立腺、皮膚、胃、精巣、胸腺、甲状腺、尿路、子宮、または外陰がんの1つが含まれ得る。いくつかの実施形態では、がんの型の特徴はビンニングされた特徴である。

【0117】

いくつかの実施形態では、基本的特徴は、がんの病期の特徴を含む。がんの病期分類は、多くの場合、がんの型（例えば、膵臓がんの病期分類、前立腺がんの病期分類、乳がんの病期分類、卵巣がんの病期分類など）に基づくが、普遍的な病期分類システムも当技術分野で知られている。任意の適切ながん病期分類システムを使用してもよく、例えば、腫瘍の位置、細胞型、腫瘍のサイズ、腫瘍の広がりおよび分布、腫瘍の転移、および腫瘍のグレードに依存し得る。データの特徴として、がんの病期は、典型的には、より重症度の低い病期からより重症度の高い病期までの範囲として表される。例えば、4つの構成要素の特徴を含むがんの病期の特徴の場合、ステージ1は早期段階のがんを示し得、ステージ4は後期がんを示し得る。いくつかの実施形態では、がんの病期の特徴はビンニングされた特徴である。

40

【0118】

HRD陽性データおよびHRD陰性データは、通常、訓練データセット、検証データセ

50

ット、および／または試験データセットに分けられる。訓練中、HRD分類モデルには訓練セットのみが提供される。任意選択で、訓練セットはバランスされ得る。訓練されると、モデルは、検証セットに対するパフォーマンスによって検証され、調整され得る。モデルが検証セットに対して過適合を示す場合、訓練を調整し、繰り返すことができる。訓練されると、また任意選択的に検証された後、訓練されたモデルは、試験データセットを使用して評価され得る。

【0119】

ヘテロ接合性のゲノム喪失 (gLOH) (例えば、ゲノム全体でのヘテロ接合性の喪失またはエクソーム全体でのヘテロ接合性の喪失) の尺度が、いくつかの実施形態において基本的特徴として含まれ得る。ゲノムの十分に大きな部分にわたる全エクソームシーケンシングまたは標的化シーケンシングは、ヘテロ接合性のゲノム損失の代理として解釈され得るので、ヘテロ接合性のゲノム損失を判定するために全ゲノムを分析する必要はない。いくつかの実施形態では、gLOHは、連続数値特徴として符号化される。いくつかの実施形態では、例えば、gLOHが所定の閾値より上または下である場合、gLOHはカテゴリ特徴として符号化される。所定の閾値は、例えば、約10%以上、約12%以上、約14%以上、または約16%以上に設定されてもよい。所定の閾値は、例えば16%程度に設定されてもよい。gLOHは、例えば、Swisher et al., Rucap ar i b i n r e l a p s e d , p l a t i n u m - s e n s i t i v e h i g h - g r a d e o v a r i a n c a r c i n o m a (A R I E L 2 P a r t 1) : a n i n t e r n a t i o n a l , m u l t i c e n t e r , o p e n - l a b e l , p h a s e 2 t r i a l , L a n c e t O n c o l o g y , v o l . 1 8 , n o . 1 , p p . 7 5 - 8 7 (2 0 1 7) に記載の方法を用いて判定され得る。

10

20

【0120】

1つまたは複数の短い変異体の特徴を、HRD分類モデルで (HRD分類モデルを訓練するため、および／またはHRD分類モデルに入力される試験データとして) 使用することができる。これらの短い変異体の特徴には、限定されないが、例えば、反復またはミクロ相同性領域の特徴における1つまたは複数の欠失 (例えば、少なくとも5塩基対の欠失など) および／または2つもしくはそれを超える短い変異体の特徴を組み込んだ変異シグネチャが含まれ得る。これらの短い変異体の特徴は、例示的な方法において、腫瘍サンプルに対応する配列決定データを、コンセンサスなヒトゲノム配列 (hg19など) と比較することによって識別され得る。いくつかの実施形態では、短い変異体の特徴はビニングされた特徴である。

30

【0121】

複数の短い変異体の特徴を組み合わせ、変異シグネチャスコアとして表現することができる。例えば、1つまたは複数の短い変異体の特徴は、COSMICがんデータベースからのものなどの変異のプロファイルを含み得る。一例では、1つまたは複数の短い変異体の特徴は、COSMICがんデータベースのCOSMIC ID6またはCOSMIC ID8インデルシグネチャなどのインデルベースのシグネチャを含む。サンプルプロファイルは、例えば、NNMF方法論を使用して、これらのCOSMICプロファイルにマッピングすることができる。別の例では、1つまたは複数の短い変異体の特徴は、COSMICがんデータベースのCOSMIC ID8を含む。さらに別の例では、1つまたは複数の短い変異体の特徴は、COSMICがんデータベースのSBS3変異シグネチャを含む。例示的なCOSMIC IDシグネチャの概要については、Alexandrov et al., The repertoire of mutational signatures in human cancer, Nature 2020; 578 (7793): 94-101を参照されたい。Forbes et al., COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer, Nuc. Acids Res. 2011 Jan; 39: D945-D950も参照されたい。

40

50

【0122】

いくつかの実施形態では、1つまたは複数の短い変異体の特徴は、ミクロ相同性または反復領域特徴の欠失を含む。いくつかの実施形態において、欠失は、少なくとも1塩基対である。いくつかの実施形態において、欠失は、少なくとも5塩基対である。ミクロ相同性領域における欠失は、相同組換えの非存在下で起こるミクロ相同性媒介末端結合（MMEJ）の特徴的な結果である。このプロセスでは、ゲノムでの二本鎖切断の修復を誘導するために、短い類似領域（ミクロ相同性）が使用される。これらの欠失を識別させる特性は、欠失配列の3'末端が欠失の上流の状況と類似性を共有することである。したがって、ミクロ相同性領域における欠失の特徴は、この挙動を示す欠失の数の尺度であり、また、ミクロ相同性の長さ（すなわち、より長い長さを有する多数の欠失対、より短い長さを有するより少ない欠失）に基づき得る。

10

【0123】

例示的な実施形態において、試験データは、セグメントマイナーアリル頻度の特徴およびセグメントサイズの特徴を含む。いくつかの実施形態において、セグメントマイナーアリル頻度の特徴は、ビニングされた特徴である。いくつかの実施形態では、セグメントサイズの特徴はビニングされた特徴である。試験データは、xメガベース特徴あたりのブレークポイント数の特徴、変化点コピー数の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレークポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

20

【0124】

別の例示的な実施形態において、試験データは、セグメントマイナーアリル頻度の特徴およびxメガベースあたりのブレークポイント数の特徴を含む。いくつかの実施形態において、セグメントマイナーアリル頻度の特徴は、ビニングされた特徴である。いくつかの実施形態において、xメガベースあたりのブレークポイント数の特徴が、ビニングされた特徴である。試験データは、セグメントサイズの特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレークポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

30

【0125】

別の例示的な実施形態において、試験データは、セグメントマイナーアリル頻度の特徴および変化点コピー数の特徴を含む。いくつかの実施形態において、セグメントマイナーアリル頻度の特徴は、ビニングされた特徴である。いくつかの実施形態では、変化点コピー数の特徴は、ビニングされた特徴である。試験データは、セグメントサイズの特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレークポイント数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレークポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

40

【0126】

別の例示的な実施形態において、試験データは、セグメントマイナーアリル頻度の特徴およびセグメントコピー数の特徴を含む。いくつかの実施形態において、セグメントマイ

50

ナーアリアル頻度の特徴は、ビニングされた特徴である。いくつかの実施形態において、セグメントコピー数の特徴は、ビニングされた特徴である。試験データは、セグメントサイズの特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数のさらに含み得る。

【0127】

10

別の例示的な実施形態において、試験データは、セグメントマイナーアリアル頻度の特徴および染色体アームあたりのブレイクポイント数の特徴を含む。いくつかの実施形態において、セグメントマイナーアリアル頻度の特徴は、ビニングされた特徴である。いくつかの実施形態において、染色体アームあたりのブレイクポイント数の特徴が、ビニングされた特徴である。試験データは、セグメントサイズの特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数のさらに含み得る。

20

【0128】

別の例示的な実施形態において、試験データは、セグメントマイナーアリアル頻度の特徴および振動コピー数のセグメントの数の特徴を含む。いくつかの実施形態において、セグメントマイナーアリアル頻度の特徴は、ビニングされた特徴である。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、ビニングされた特徴である。試験データは、セグメントサイズの特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、および染色体アームあたりのブレイクポイント数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数のさらに含み得る。

30

【0129】

別の例示的な実施形態において、試験データは、セグメントサイズの特徴およびxメガベースあたりのブレイクポイント数の特徴を含む。いくつかの実施形態では、セグメントサイズの特徴はビニングされた特徴である。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ビニングされた特徴である。試験データは、セグメントマイナーアリアル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数のさらに含み得る。

40

【0130】

別の例示的な実施形態において、試験データは、セグメントサイズの特徴および変化点コピー数の特徴を含む。いくつかの実施形態では、セグメントサイズの特徴はビニングされた特徴である。いくつかの実施形態では、変化点コピー数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリアル頻度 (s e g M A F) の特徴、シー

50

ケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴の特徴、セグメントコピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

【0131】

別の例示的な実施形態において、試験データは、セグメントサイズの特徴およびセグメントコピー数の特徴を含む。いくつかの実施形態では、セグメントサイズの特徴はビニングされた特徴である。いくつかの実施形態において、セグメントコピー数は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

【0132】

別の例示的な実施形態において、試験データは、セグメントサイズの特徴および染色体アームあたりのブレイクポイント数の特徴を含む。いくつかの実施形態では、セグメントサイズの特徴はビニングされた特徴である。いくつかの実施形態において、染色体アームあたりのブレイクポイント数の特徴が、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

【0133】

別の例示的な実施形態において、試験データは、セグメントサイズの特徴および振動コピー数のセグメントの数の特徴を含む。いくつかの実施形態では、セグメントサイズの特徴はビニングされた特徴である。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、セグメントコピー数の特徴、および染色体アームあたりのブレイクポイント数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

【0134】

別の例示的な実施形態では、試験データは、xメガベースあたりのブレイクポイント数の特徴および変化点コピー数の特徴を含む。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ビニングされた特徴である。いくつかの実施形態では、変化点コピー数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、セグメントコピー数の特徴、染色体アームあた

りのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数の特徴をさらに含み得る。

【0135】

別の例示的な実施形態では、試験データは、xメガベースあたりのブレイクポイント数の特徴およびセグメントコピー数の特徴を含む。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ビニングされた特徴である。いくつかの実施形態において、セグメントコピー数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、変化点コピー数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数の特徴をさらに含み得る。

【0136】

別の例示的な実施形態では、試験データは、xメガベースあたりのブレイクポイント数の特徴および染色体アームあたりのブレイクポイント数の特徴を含む。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ビニングされた特徴である。いくつかの実施形態において、染色体アームあたりのブレイクポイント数の特徴が、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、変化点コピー数の特徴、セグメントコピー数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数の特徴をさらに含み得る。

【0137】

別の例示的な実施形態では、試験データは、xメガベースあたりのブレイクポイント数の特徴および振動コピー数のセグメントの数の特徴を含む。いくつかの実施形態において、xメガベースあたりのブレイクポイント数の特徴が、ビニングされた特徴である。いくつかの実施形態において、振動コピー数のセグメントの数は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、変化点コピー数の特徴、セグメントコピー数の特徴、および染色体アームあたりのブレイクポイント数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数の特徴をさらに含み得る。

【0138】

別の例示的な実施形態では、試験データは、変化点コピー数の特徴およびセグメントコピー数の特徴を含む。いくつかの実施形態では、変化点コピー数の特徴は、ビニングされた特徴である。いくつかの実施形態において、セグメントコピー数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (s e g M A F) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、xメガベースあたりのブレイクポイント数の特徴、染色体アームあたりのブレイクポイント数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、g L O Hの尺度および／または1つもしくは複数の短い

変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

【0139】

別の例示的な実施形態において、試験データは、変化点コピー数の特徴および染色体アームあたりのブレークポイント数の特徴を含む。いくつかの実施形態では、変化点数の特徴は、ビニングされた特徴である。いくつかの実施形態において、染色体アームあたりのブレークポイント数の特徴が、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (segMAF) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、xメガベースあたりのブレークポイント数の特徴、セグメントコピー数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

10

【0140】

別の例示的な実施形態において、試験データは、変化点コピー数の特徴および振動コピー数のセグメントの数の特徴を含む。いくつかの実施形態では、変化点コピー数の特徴は、ビニングされた特徴である。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (segMAF) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、xメガベースあたりのブレークポイント数の特徴、セグメントコピー数の特徴、および染色体アームあたりのブレークポイント数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

20

【0141】

別の例示的な実施形態において、試験データは、セグメントコピー数の特徴および染色体アームあたりのブレークポイント数の特徴を含む。いくつかの実施形態において、セグメントコピー数の特徴は、ビニングされた特徴である。いくつかの実施形態において、染色体アームあたりのブレークポイント数の特徴が、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (segMAF) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、xメガベースあたりのブレークポイント数の特徴、変化点コピー数の特徴、および振動コピー数のセグメントの数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

30

【0142】

別の例示的な実施形態において、試験データは、セグメントコピー数の特徴および振動コピー数のセグメントの数の特徴を含む。いくつかの実施形態において、セグメントコピー数の特徴は、ビニングされた特徴である。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (segMAF) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、xメガベースあたりのブレークポイント数の特徴、変化点コピー数の特徴、および染色体アームあたりのブレークポイント数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍

40

50

数性の特徴の1つまたは複数をさらに含み得る。

【0143】

別の例示的な実施形態では、試験データは、染色体アームあたりのブレイクポイント数の特徴および振動コピー数のセグメントの数の特徴を含む。いくつかの実施形態において、染色体アームあたりのブレイクポイント数の特徴が、ビニングされた特徴である。いくつかの実施形態において、振動コピー数のセグメントの数の特徴は、ビニングされた特徴である。試験データは、セグメントマイナーアリル頻度 (segMAF) の特徴、シーケンシングリードの数の特徴、絶対コピー数の特徴、セグメントサイズの特徴、xメガベースあたりのブレイクポイント数の特徴、変化点コピー数の特徴、およびセグメントコピー数の特徴のうちの少なくとも1つをさらに含み得る。試験データは、gLOHの尺度および／または1つもしくは複数の短い変異体の特徴をさらに含み得る。試験データは、試験データが得られた対象の年齢、がんの型の特徴、がんの病期の特徴、腫瘍純度の特徴、および腫瘍ゲノム倍数性の特徴の1つまたは複数をさらに含み得る。

10

【0144】

HRDモデル

対象のがんの腫瘍は、腫瘍をHRD陽性（またはHRD陽性である可能性が高い）またはHRD陰性（またはHRD陰性である可能性が高い）として分類するように構成された、訓練されたHRD分類モデルを使用して分類される。HRD分類モデルは、複数のHRD陽性腫瘍の各HRD陽性腫瘍について、HRD陽性腫瘍に関連する1つまたは複数のデータ特徴（例えば、他の可能な特徴の中でも、1つまたは複数のコピー数の特徴および／または1つまたは複数の短い変異体の特徴）、およびHRD陽性標識を含むHRD陽性データを使用して訓練される。HRD分類モデルは、さらに、複数のHRD陰性腫瘍の各HRD陰性腫瘍について、HRD陰性腫瘍に関連する1つまたは複数のデータ特徴（例えば、他の可能な特徴の中でも、1つまたは複数のコピー数の特徴および／または1つまたは複数の短い変異体の特徴）、およびHRD陰性標識を含むHRD陰性データを使用して訓練される。対象の腫瘍のゲノムに関連する1つまたは複数のデータ特徴（例えば、他の可能な特徴の中でも、1つまたは複数のコピー数の特徴および／または1つまたは複数の短い変異体の特徴）を含む試験データを、訓練されたHRD分類モデルに入力し、次いで、試験データに基づいて腫瘍をHRD陽性（またはHRD陽性である可能性が高い）またはHRD陰性（またはHRD陰性である可能性が高い）として分類する。

20

30

【0145】

本明細書に記載のモデルは、1つまたは複数の機械学習モデル、1つまたは複数の非機械学習モデル、またはそれらの任意の組み合わせを含むことができる。本明細書に記載の機械学習モデルは、経験を通じておよびデータの使用によって自動的に改善する任意のコンピュータアルゴリズムを含む。機械学習モデルは、教師ありモデル、教師なしモデル、半教師ありモデル、自己教師ありモデルなどを含むことができる。例示的な機械学習モデルには、線形回帰、ロジスティック回帰、決定木、SVM、ナイーブベイズ、ニューラルネットワーク、K平均、分散分析 (ANOVA)、カイ二乗分析、ランダムフォレスト、次元削減アルゴリズム、および勾配ブースティングアルゴリズム (XGBなど) が含まれるが、これらに限定されない。非機械学習モデルは、必ずしも訓練および再訓練を必要としない任意のコンピュータアルゴリズムを含むことができる。

40

【0146】

HRD分類器は、勾配ブースティングモデルなどの確率的分類器であってもよい。確率的分類器は、HRD陽性尤度スコアまたはHRD陰性尤度スコアを出力することなどによって、腫瘍がHRD陽性またはHRD陰性である確率を計算するように構成することができる。HRD分類モデルから出力された確率に基づいて、腫瘍はHRD陽性またはHRD陰性であるとみなすことができる。任意選択で、例えば、腫瘍がHRD陽性である確率も腫瘍がHRD陰性である確率も所定の確率閾値を超えない場合、腫瘍は曖昧とみなし得る。HRD陽性データおよびHRD陰性データは、本明細書に記載のコピー数の特徴および／または短い変異体の特徴を含み得る。

50

【0147】

HRD陰性データは、特定のHRD関連遺伝子に野生型アリル（すなわち、HRDに関連しないアリル）を有するゲノムを含み得る。例えば、いくつかの実施形態では、HRD陰性データは、BRCA1、BRCA2、ATM、BARD1、BRIP1、CDK12、CHEK1、CHEK2、FANCL、PALB2、RAD51B、RAD51C、RAD51D、および／またはRAD45Lを含むがこれらに限定されない、HRDに関連する遺伝子の1つまたは複数の野生型アリルを有するゲノムに関連するデータを含む。いくつかの実施形態では、HRD陰性データは、BRCA1、BRCA2、ATM、BARD1、BRIP1、CDK12、CHEK1、CHEK2、FANCL、PALB2、RAD51B、RAD51C、RAD51D、および／またはRAD45Lを含むがこれらに限定されない、HRDに関連する遺伝子の1つまたは複数のプロモータメチル化データを含む。いくつかの実施形態では、HRD陰性データは、BRCA1、BRCA2、ATM、BARD1、BRIP1、CDK12、CHEK1、CHEK2、FANCL、PALB2、RAD51B、RAD51C、RAD51D、および／またはRAD45Lを含むがこれらに限定されない、HRDに関連する遺伝子の1つまたは複数のRNA発現データを含む。いくつかの実施形態では、HRD陰性データは、白金系薬物（例えば、化学療法）および／またはPARP阻害剤に耐性があることが分かった腫瘍に関連するゲノムに関連するデータを含む。いくつかの実施形態では、HRD陰性データは、以前にHRD陰性と分類された腫瘍に関連するゲノムに関連するデータを含む。いくつかの実施形態では、HRD陰性データは、少なくとも部分的に、コンセンサスなヒトゲノム配列またはその一部に由来する。

10

20

【0148】

HRD陽性データは、特定のHRD関連遺伝子においてHRD関連アリルを有するゲノムに関連するデータを含み得る。例えば、いくつかの実施形態では、HRD陽性データは、BRCA1、BRCA2、ATM、BARD1、BRIP1、CDK12、CHEK1、CHEK2、FANCL、PALB2、RAD51B、RAD51C、RAD51D、および／またはRAD45L、特にその二対立遺伝子変異を含むがこれらに限定されない、HRDに関連する遺伝子の1つまたは複数の変異を有するゲノムに関連するデータを含む。いくつかの実施形態では、HRD陽性データは、BRCA1、BRCA2、ATM、BARD1、BRIP1、CDK12、CHEK1、CHEK2、FANCL、PALB2、RAD51B、RAD51C、RAD51D、および／またはRAD45Lを含むがこれらに限定されない、HRDに関連する遺伝子の1つまたは複数のプロモータメチル化データを含む。いくつかの実施形態では、HRD陽性データは、BRCA1、BRCA2、ATM、BARD1、BRIP1、CDK12、CHEK1、CHEK2、FANCL、PALB2、RAD51B、RAD51C、RAD51D、および／またはRAD45Lを含むがこれらに限定されない、HRDに関連する遺伝子の1つまたは複数のRNA発現データを含む。いくつかの実施形態では、HRD陽性データは、白金系薬物および／またはPARP阻害剤に感受性があることが分かった腫瘍に関連するゲノムに関連するデータを含む。いくつかの実施形態では、HRD陽性データは、以前にHRD陽性と分類された腫瘍に関連するゲノムに関連するデータを含む。いくつかの実施形態では、HRD陽性データは、HRDに関連する二対立遺伝子BRCA1およびBRCA2変異を有する腫瘍に関連するデータを含む。

30

40

【0149】

HRD陽性データは、HRD陰性データとバランスをとることができる。例えば、不均衡な訓練データセットでは、HRD陽性訓練腫瘍の数がHRD陰性腫瘍の数を上回り（またはその逆）得る。データのバランスをとることにより、1つのラベルへ標識のバイアスを回避するのに十分な数の各標識をモデルが有することが保証される。バランスがとれている場合、HRD陽性腫瘍の数またはHRD陰性腫瘍の数は、それらの間の比が所望のレベル（約1：1または任意の他の所望の比など）になるように調整される。バランスのとれたデータセットを使用して、HRD分類器を訓練し、次いで、HRD陽性腫瘍およびH

50

R D 陰性腫瘍を含む試験データセットに対して試験することができる。

【0150】

H R D 分類器を訓練するために使用される腫瘍はそれぞれ、H R D 陽性標識またはH R D 陰性標識を含む。任意の適切な方法論を使用して、腫瘍を（例えば、腫瘍にメタデータタグを適用して）H R D 陽性またはH R D 陰性として計算で標識することができる。H R D 陽性標識は、H R D 関連遺伝子の1つ、例えば、B R C A 1、B R C A 2、A T M、B A R D 1、B R I P 1、C D K 1 2、C H E K 1、C H E K 2、F A N C L、P A L B 2、R A D 5 1 B、R A D 5 1 C、R A D 5 1 D、および／またはR A D 4 5 Lを含むがこれらに限定されないH R Dに関連する遺伝子の1つの変化、特にその二対立遺伝子変化の存在によって割り当てられ得る。B R C A 1およびB R C A 2の一方または両方における変異は、H R D 陽性、特に二対立遺伝子B R C A 1／B R C A 2変異を特に示す。腫瘍はまた、臨床歴に基づいてH R D 陽性として標識され得る。例えば、腫瘍がP A R P阻害剤または白金系薬物レジメンに感受性であった場合、腫瘍はH R D 陽性である可能性が高い。H R D 陰性標識は、H R D 関連遺伝子の1つ、例えば、B R C A 1、B R C A 2、A T M、B A R D 1、B R I P 1、C D K 1 2、C H E K 1、C H E K 2、F A N C L、P A L B 2、R A D 5 1 B、R A D 5 1 C、R A D 5 1 D、および／またはR A D 4 5 Lを含むがこれらに限定されないH R Dに関連する遺伝子の1つの変化、特にその二対立遺伝子変化の非存在に基づいて割り当てられ得る。H R D 関連遺伝子の変異は、遺伝子配列と参照ゲノム、例えばコンセンサスなヒトゲノム配列、例えばh g 1 9との比較によって検出され得る。同様に、腫瘍はまた、臨床歴に基づいてH R D 陰性として標識され得る。例えば、腫瘍がP A R P阻害剤または白金系薬物レジメンに耐性がある場合、腫瘍はH R D 陰性である可能性が高い。H R D 陽性腫瘍は、何回かの治療後にこれらの薬物に対する耐性を生じ得るので、これは、腫瘍がP A R P阻害剤または白金系薬物レジメンによる治療前に治療ナীবであった場合に特に当てはまる。各腫瘍はH R D 陽性またはH R D 陰性の標識を含み得るが、この標識は、腫瘍がH R D 陽性またはH R D 陰性であるという絶対的な確実性を必要としない。代わりに、多数のH R D 陽性腫瘍および多数のH R D 陰性腫瘍を含む堅牢な訓練データセットが与えられ、当技術分野で知られているようにこれらのデータの過剰適合を回避することによって、偽陽性および偽陰性の寄与がモデルにおいて平均化される。さらに、より大きな訓練データセット、特にバランスのとれた訓練データセット、ならびに明確に定義された陽性および陰性のラベルを有するデータセット（例えば、H R D 陰性標識のために検証されたコンセンサスゲノムを使用することによって、H R D 陽性標識のために、検証された二対立遺伝子B R C A 1／2変異体または検証された十分に特徴付けられたB R C A n e s sサンプルを使用することによって）を使用することにより、モデルは、H R D 陰性表現型とH R D 痕跡を示す表現型（すなわち、H R D 陽性表現型）との間の微妙な差を適切に評価することができる。

【0151】

分類方法は、コンピュータ実装方法である。この分類は、訓練されたH R D 分類器モデルを実行するためのプログラム命令を含む具体的に構成された機械またはシステム上で実行することができる。これはコンピュータまたはシステムの非一時的コンピュータ可読メモリに記憶することができる。コンピュータは、一般に、メモリにアクセスすることができる1つまたは複数のプロセッサを含む。1つまたは複数のプロセッサは、メモリに記憶されてもよいデータ（例えば、対象における腫瘍のゲノムに関連する1つまたは複数のコピー数の特徴および／または1つまたは複数の短い変異体の特徴などの試験データ、ならびにいくつかの実施形態では、他の特徴および測定値）を受信することができる。1つまたは複数のプロセッサは、訓練されたH R D 分類器モデルにアクセスすることができ、テストデータをモデルに入力することができる。次いで、1つまたは複数のプロセッサおよび訓練されたH R D 分類器モデルは、がんを、H R D 陽性である可能性が高いまたはH R D 陰性である可能性が高いものとして分類することができる。

【0152】

H R D 分類器モデルは、がんの腫瘍をH R D 陽性またはH R D 陰性として分類すること

ができる。いくつかの実施形態では、HRD分類器モデルは、腫瘍をHRD陽性である可能性が高い、HRD陰性である可能性が高い、または曖昧と分類することができる。例えば、HRD分類器モデルは、十分に高い信頼性または確率で腫瘍をHRD陽性である可能性が高いまたはHRD陰性である可能性が高いと分類することができない場合、腫瘍を曖昧と分類することができる。信頼度または確率の閾値は、不正確な分類に対する許容度を考慮して、必要に応じてユーザによって設定されてもよい。一例では、ユーザは、HRD陽性尤度スコアの閾値を0.8に設定し、HRD陰性尤度スコアの閾値を0.2に設定することができる。HRD陽性尤度スコアが0.8未満である場合、および／またはHRD陰性尤度スコアが0.2を超える場合、HRDモデルは腫瘍をHRD陽性と分類することができず、腫瘍をHRD陰性と分類するか（HRD陽性尤度スコアがどの程度低いか、およびHRD陰性尤度スコアがどの程度高いか）に依存する）、または曖昧であると分類する。

10

【0153】

いくつかの実施形態では、HRD分類器は、腫瘍がHRD陽性である尤度スコアを出力する。いくつかの実施形態では、HRD分類器は、腫瘍がHRD陰性である尤度スコアを出力する。HRD分類器は、HRD陽性尤度スコアおよびHRD陰性尤度スコアのいずれかまたは両方を出力するように構成されてもよい。HRD分類器はまた、HRD陽性尤度スコアとHRD陰性尤度スコアとの比および／またはHRD陰性尤度スコアとHRD陽性尤度スコアとの比を出力するように構成されてもよい。尤度スコアは、0.0（腫瘍がHRD陽性でもHRD陰性でもない確実性を示す）～1.0（腫瘍がHRD陽性またはHRD陰性である確実性を示す）の値として表すことができる。例えば、訓練されたHRD分類器は、対象のがんの腫瘍に関連する複数のデータ特徴を含む試験サンプルデータを受信し、0.8のHRD陽性尤度スコアおよび0.15のHRD陰性尤度スコアを出力することができる。HRD分類器は、1つまたは複数の尤度スコアに基づいて腫瘍をHRD陽性またはHRD陰性とみなすように構成することができる。前述の例では、HRD陽性尤度スコア0.8およびHRD陰性尤度スコア0.15に基づいて、HRD分類器は腫瘍をHRD陽性とみなすことができる。いくつかの実施形態では、HRD分類器は、HRD陽性尤度スコアが少なくとも0.4、例えば少なくとも0.45、少なくとも0.5、少なくとも0.55、少なくとも0.6、少なくとも0.65、少なくとも0.70、少なくとも0.75、少なくとも0.80、少なくとも0.85、少なくとも0.90、少なくとも0.95、または少なくとも0.99である場合、腫瘍をHRD陽性とみなす。いくつかの実施形態では、HRD分類器は、HRD陽性尤度スコアが少なくとも0.7である場合、腫瘍をHRD陽性とみなす。いくつかの実施形態では、HRD分類器は、HRD陽性尤度スコアが少なくとも0.8である場合、腫瘍をHRD陽性とみなす。いくつかの実施形態では、HRD分類器は、HRD陽性尤度スコアが少なくとも0.9である場合、腫瘍をHRD陽性とみなす。いくつかの実施形態では、HRD分類器は、HRD陰性尤度スコアが少なくとも0.4、例えば少なくとも0.5、少なくとも0.6、少なくとも0.65、少なくとも0.70、少なくとも0.75、少なくとも0.80、少なくとも0.85、少なくとも0.90、少なくとも0.95、または少なくとも0.99である場合、腫瘍をHRD陰性とみなす。いくつかの実施形態では、HRD分類器は、HRD陰性尤度スコアが少なくとも0.7である場合、腫瘍をHRD陰性とみなす。いくつかの実施形態では、HRD分類器は、HRD陰性尤度スコアが少なくとも0.8である場合、腫瘍をHRD陰性とみなす。いくつかの実施形態では、HRD分類器は、HRD陰性尤度スコアが少なくとも0.9である場合、腫瘍をHRD陰性とみなす。いくつかの実施形態では、HRD分類器は、HRD陰性尤度スコアが0.5未満、例えば0.45未満、0.40未満、0.35未満、0.30未満、0.30未満、0.25未満、0.20未満、0.15未満、0.10未満、または0.05未満である場合、腫瘍をHRD陽性とみなす。いくつかの実施形態では、HRD分類器は、HRD陽性尤度スコアが0.5未満、例えば0.45未満、0.40未満、0.35未満、0.30未満、0.30未満、0.25未満、0.20未満、0.15未満、0.10未満、または0.05未満である場合、腫瘍をH

20

30

40

50

R D 陰性とみなす。いくつかの実施形態では、H R D 分類器は、H R D 陽性尤度スコアが特定の閾値（少なくとも 0.80 など）を上回り、H R D 陰性尤度スコアが特定の閾値（0.25 未満など）を下回る場合、腫瘍を H R D 陽性とみなす。いくつかの実施形態では、H R D 分類器は、H R D 陰性尤度スコアが特定の閾値（少なくとも 0.80 など）を上回り、H R D 陽性尤度スコアが特定の閾値（0.25 未満など）を下回る場合、腫瘍を H R D 陰性とみなす。いくつかの実施形態では、H R D 分類器は、H R D 陽性尤度スコアが特定の閾値を下回り、H R D 陰性尤度スコアが閾値を下回る場合、または尤度スコアの絶対値が閾値類似性パーセント内にある場合、腫瘍を曖昧とみなす。

【0154】

がんを、H R D 陽性である可能性が高いまたは H R D 陰性である可能性が高い（または曖昧）と識別する報告が生成され得る。報告は、例えば、電子医療記録または印刷された報告であってもよく、これは、対象または対象に関連付けられたヘルスケア提供者（例えば、医師、看護師、診療所など）に送信することができる。報告は、がんの腫瘍を治療する方法または薬物などの医療での決定を行うために使用され得る。

【0155】

報告は、電子ディスプレイまたはカスタマイズされたインターフェースに表示されてもよい。例えば、いくつかの実施形態では、コンピュータ実装方法は、報告を自動的に生成することができ、生成された報告を電子ディスプレイまたはカスタマイズされたインターフェースに自動的に表示することができる。

【0156】

図 7 は、対象のがんの腫瘍を H R D 陽性または H R D 陰性として分類するように構成された H R D 分類モデル 702 を訓練および操作するための例示的な方法を示す。H R D 分類モデル 702 は、H R D 陽性訓練データセット 704 および H R D 陰性訓練データセット 706 を含むデータセットを使用して訓練される。H R D 陽性訓練データセット 704 は、1 つまたは複数の H R D 陽性サンプルデータ要素（すなわち、H R D 陽性サンプル i までの H R D 陽性サンプル 1 のデータ）を含む。各 H R D 陽性サンプルデータ要素は、H R D 陽性腫瘍の特徴（例えば、コピー数の特徴、基本的特徴、短い変異体の特徴などである）と関連付けられている。H R D 陽性サンプルデータ要素はまた、g L O H の尺度および／または短い変異体の特徴（図示せず）などの他のデータ特徴も含み得る。特徴は、H R D 陽性ラベルに関連するものとして標識される。同様に、H R D 陰性訓練データセット 706 は、1 つまたは複数の H R D 陰性訓練サンプルデータ要素（すなわち、H R D (－) サンプル j までの H R D (－) サンプル 1）を含む。各 H R D 陰性サンプルデータ要素は、H R D 陰性腫瘍の特徴（例えば、コピー数の特徴、基本的特徴、短い変異体の特徴などである）と関連付けられている。H R D 陰性サンプルデータ要素はまた、g L O H の尺度および／または短い変異体の特徴（図示せず）などの他のデータ特徴も含み得る。H R D 陰性サンプルは、H R D 陰性標識に関連するものとして標識される。

【0157】

いくつかの実施形態では、H R D 分類モデル 702 は、ツリーベースの勾配ブースティングモデル（X G B o o s t など）である。このモデルでは、すべてのモデルを互いに分離して訓練するのではなく（例えば、ランダムフォレストによって）、モデルは、各新しいモデルが前のモデルからの残差に適合するように連続して訓練される。したがって、モデルは、順次接続された多くのより弱い分類器から強い分類器を達成する。H R D 分類モデルのパフォーマンスを推定するための訓練データでは、反復交差検証を使用することができる。

【0158】

分類モデル 702 が訓練データセットで訓練された後、分類モデル 702 を使用して、対象のがんの腫瘍を H R D 陽性または H R D 陰性として分類することができる。対象のがんの腫瘍を H R D 陽性または H R D 陰性として分類するために、分類モデル 702 は、分類される腫瘍に関連する試験特徴データを含む試験データ 708 を受信する。テストデータ 708 は、1 つまたは複数のコピー数の特徴を含み、1 つまたは複数の基本的特徴、1

10

20

30

40

50

つまたは複数の短い変異体の特徴などを含んでもよい。分類モデル 7 0 2 は、腫瘍が H R D 陽性である確率 7 1 0 および／または腫瘍が H R D 陰性である確率 7 1 2 を判定することができる。確率 7 1 0 および 7 1 2 は、任意選択的に、H R D コールモジュール 7 1 4 に入力される。H R D コールモジュール 7 1 4 は、がんを H R D 陽性または H R D 陰性とみなすことができる。例えば、腫瘍試験サンプルが H R D 陽性である確率 7 1 0 が、腫瘍試験サンプルが H R D 陰性である確率 7 1 2 より大きい場合、腫瘍試験サンプルは H R D 陽性とみなすことができる。腫瘍試験サンプルが H R D 陰性である確率 7 1 2 が、腫瘍試験サンプルが H R D 陽性である確率 7 1 0 より大きい場合、腫瘍試験サンプルは H R D 陰性とみなすことができる。任意選択で、確率 7 1 0 および 7 1 2 のいずれも所定の閾値を上回っていない場合、腫瘍試験サンプルは曖昧とみなすことができる。

10

【0159】

本明細書に記載の方法は、1つまたは複数のコンピュータシステムを使用して実装されることができる。そのようなコンピュータシステムは、コンピュータシステムがそのような方法を実行するために1つまたは複数のプロセッサを実行するように構成された1つまたは複数のプログラムを含むことができる。コンピュータ実装方法の1つまたは複数のステップは、自動的に実行されてもよい。コンピュータシステムは、1つまたは複数のコンピューティングノードを含むことができる。例えば、システムは、ネットワークの1つまたは複数のコンピューティングノード上で前記ネットワークを介して方法を通信および実行するように接続および構成することができる2つ以上のコンピューティングノード（例えば、サーバ、コンピュータ、ルータ、またはネットワークインターフェースを含む他の種類の電子デバイス）を含むことができる。

20

【0160】

図8は、一実施形態によるコンピューティングデバイスの例を示している。デバイス 1 1 0 0 は、ネットワークに接続されたホストコンピュータとすることができる。デバイス 1 1 0 0 は、クライアントコンピュータまたはサーバとすることができる。図8に示されるように、デバイス 1 1 0 0 は、パーソナルコンピュータ、ワークステーション、サーバ、またはハンドヘルド計算デバイス（携帯電子デバイス、例えば、電話またはタブレット）などの任意の好適なタイプのマイクロプロセッサベースのデバイスであり得る。デバイスは、例えば、プロセッサ 1 1 1 0、入力デバイス 1 1 2 0、出力デバイス 1 1 3 0、ストレージ 1 1 4 0、および通信デバイス 1 1 6 0 のうちの1つまたは複数を含むことができる。入力デバイス 1 1 2 0 および出力デバイス 1 1 3 0 は、一般に、上述したものに对应することができ、コンピュータと接続可能とすることができるかまたは一体化されることができる。

30

【0161】

入力デバイス 1 1 2 0 は、タッチスクリーン、キーボードもしくはキーパッド、マウス、または音声認識デバイスなどの入力を提供する任意の好適なデバイスとすることができる。出力デバイス 1 1 3 0 は、ディスプレイ、タッチスクリーン、触覚デバイス、またはスピーカなど、出力を提供する任意の好適なデバイスであり得る。

【0162】

ストレージ 1 1 4 0 は、R A M、キャッシュ、ハードドライブ、またはリムーバブルストレージディスクを含む、電気、磁気、または光学メモリなどのストレージを備える任意の好適なデバイスであり得る。通信デバイス 1 1 6 0 は、ネットワークインターフェースチップまたはデバイスなどの、ネットワークを介してシグナルを送受信することができる任意の好適なデバイスを含むことができる。コンピュータの構成要素は、物理的バスまたは無線などの任意の好適な様式で接続することができる。

40

【0163】

H R D 分類モジュール 1 1 5 0 は、ストレージ 1 1 4 0 に記憶することができ、プロセッサ 1 1 1 0 によって実行することができ、例えば、H R D モデルに関連する方法およびプロセス（例えば、上記のようなデバイスで実施されるように）を実行および実施するための1つまたは複数のプログラム命令を含むことができる。

50

【0164】

HRDモジュール1150はまた、命令実行システム、装置、もしくはデバイス（例えば、上記のもの）によって、またはそれらと接続して使用するための任意の非一時的コンピュータ可読記憶媒体内に記憶および／または転送することができ、命令実行システム、装置、もしくはデバイスからの、ソフトウェアに関連付けられた命令をフェッチし、命令を実行することができる。本開示の文脈において、コンピュータ可読記憶媒体は、ストレージ1140などの任意の媒体とすることができ、命令実行システム、装置、もしくはデバイスによって、またはそれらと接続して使用するためのプログラミングを含むもしくは記憶することができる。

【0165】

HRDモジュール1150はまた、命令実行システム、装置、もしくはデバイス（例えば、上記のもの）によって、またはそれらと接続して使用するための任意の伝送媒体内に伝播することができ、命令実行システム、装置、もしくはデバイスからの、ソフトウェアに関連付けられた命令をフェッチし、命令を実行することができる。本開示の文脈において、伝送媒体は、任意の媒体とすることができ、命令実行システム、装置、もしくはデバイスによって、またはそれらと接続して使用するための伝送プログラミングを通信、伝播、または伝送することができる。伝送可読媒体は、電子、磁気、光学、電磁気、もしくは赤外線の有線または無線伝播媒体を含むことができるが、これらに限定されない。

【0166】

デバイス1100は、任意の好適な種類の相互接続された通信システムとすることができるネットワークへと接続されることができる。ネットワークは、任意の好適な通信プロトコルを実装することができ、任意の好適なセキュリティプロトコルによって保護されることができる。ネットワークは、無線ネットワーク接続（T1またはT3回線）、ケーブルネットワーク、DSL、または電話回線などの、ネットワーク信号の送受信を実装することができる任意の好適な構成のネットワークリンクを含むことができる。

【0167】

デバイス1100は、ネットワークで動作するのに好適な任意のオペレーティングシステムを実装することができる。ソフトウェア350は、C、C++、Java、またはPythonなどの任意の好適なプログラミング言語で書かれることができる。様々な実施形態では、本開示の機能を具現化するアプリケーションソフトウェアは、異なる構成で、例えば、クライアント／サーバ構成で、またはウェブベースのアプリケーションもしくはウェブサービスとしてのウェブブラウザを介してなど、展開されることができる。

【0168】

治療方法

腫瘍をHRD陽性またはHRD陰性（またはHRD陽性である可能性が高いまたはHRD陰性である可能性が高い）として特徴付けることは、腫瘍を有する対象に有効な治療を選択するのに特に有用である。HRD陽性として分類される腫瘍は、多くの場合、HRD陰性腫瘍が抵抗性であり得る特定の薬物および療法に対してより感受性である。腫瘍のHRD陽性、HRD陽性である可能性が高い、HRD陰性、またはHRD陰性である可能性が高いものとしての分類に基づいて、異なる薬物または療法を選択することができる。したがって、対象のがんを治療する方法は、本明細書に記載の方法に従って、がんの腫瘍をHRD陽性またはHRD陰性である可能性が高いと評価すること（またはがんの腫瘍をHRD陽性またはHRD陰性とみなすこと）、次いで、腫瘍をHRD陽性である可能性が高い、またはHRD陰性である可能性が高いと分類することに基づいて（または腫瘍をHRD陽性またはHRD陰性とみなすことに基づいて）、治療有効量の薬物を対象に投与することを含み得る。

【0169】

対象のがんを治療する方法は、対象のがんの腫瘍の分類をHRD陽性である可能性が高い、またはHRD陰性である可能性が高いものとして得ることを含むことができる。この分類を取得するために、本明細書に記載のHRD分類モデルを使用することができる。が

10

20

30

40

50

んの腫瘍のゲノムに関連する1つまたは複数のコピー数の特徴は、対象の腫瘍のゲノムに関連する1つまたは複数のコピー数の特徴に基づいて、腫瘍をHRD陽性である可能性が高いまたはHRD陰性である可能性が高いものと分類するように構成されたHRD分類モデルに入力され得る。HRD分類モデルは、複数のHRD陽性腫瘍からのHRD陽性データ、および複数のHRD陰性腫瘍からのHRD陰性データを使用して訓練される。分類は、例えば、HRD分類モデルを操作することによって、またはHRD分類モデルを操作した別のものからの結果を受信することによって取得することができる。

【0170】

1つまたは複数の基本的特徴および／または1つまたは複数の短い変異体の特徴は、1つまたは複数の基本的特徴および／または1つまたは複数の短い変異体の特徴に基づいて腫瘍をHRD陽性である可能性が高いまたはHRD陰性である可能性が高いと分類するように構成されたHRD分類モデルに入力することができる。1つまたは複数の短い変異体の特徴および1つまたは複数の基本的特徴は、1つまたは複数のコピー数の特徴に加えて、またはその代わりに、1つまたは複数のコピー数の特徴であり得る。

【0171】

いくつかの実施形態において、治療方法は、1つまたは複数のコピー数の特徴を含む試験サンプルデータを得ることを含み得る。いくつかの実施形態では、治療方法は、1つ以上の基本的特徴を得ることを含み得る。いくつかの態様において、治療方法は、ヘテロ接合性のゲノム規模の喪失の尺度を得ることを含み得る。いくつかの実施形態では、治療方法は、1つまたは複数の短い変異体の特徴を得ることを含み得る。対象から試験サンプルを得てもよく、核酸分子は試験サンプルに由来し得る。試験サンプルは、例えば、がんの固形組織生検であってもよく、固形組織サンプルから核酸を単離し得る。任意選択で、試験サンプルは、例えば、核酸分子を単離する前に試験サンプルを凍結するか、またはサンプルを固定することによって（例えば、ホルマリン固定パラフィン包埋（FFPE）サンプルを形成することによって）、保存され得る。あるいは、試験サンプルは液体生検サンプル（例えば、対象からの血液、血漿、または他の液体サンプル）であり、循環腫瘍DNA（ctDNA）を含む核酸を液体サンプルから得ることができる。サンプル由来の核酸をアッセイし、次いで分析して、1つ以上のコピー数の特徴、1つ以上の基本的特徴、または1つまたは複数の短い変異体の特徴のいずれかを生成することができる。

【0172】

HRD陽性である可能性が高いまたはHRD陰性である可能性が高いものとしての腫瘍の分類を得ることは、記載された特徴および／または尺度をHRD分類モデルに入力すること、および特徴および／または尺度を使用して、HRD分類モデルに入力されたデータに基づいて、HRD陽性である可能性が高いまたはHRD陰性である可能性が高いものとしてがんを分類することを含むことができる。あるいは、腫瘍をHRD陽性である可能性が高いまたはHRD陰性である可能性が高いものとする分類を取得することは、別のエンティティからの報告を受信することを含み得る。報告は、他のエンティティによって生成されてもよく、報告は、HRD陽性である可能性が高いまたはHRD陰性である可能性が高い腫瘍の分類を含むことができ、分類は、本明細書に記載のHRD分類モデルを使用して生成される。いくつかの態様では、報告は、腫瘍がHRD陽性である尤度スコアおよび／または腫瘍がHRD陰性である尤度スコアを含み、尤度スコアに基づいて最終的な分類を行うことができる。

【0173】

HRD陽性である可能性が高いまたはHRD陰性である可能性が高い腫瘍の分類が行われると、分類に基づいて治療を選択することができる。腫瘍がHRD陽性である可能性が高いと分類された場合、HRD陽性腫瘍に有効な治療が選択される。次いで、選択された治療を対象に投与して、HRD陽性である可能性が高いと分類される腫瘍を治療することができる。腫瘍がHRD陰性である可能性が高いと分類された場合、白金系薬物でもPARP阻害剤でもない治療が選択され得る。次いで、選択された治療を対象に投与して、HRD陰性である可能性が高いと分類される腫瘍を治療することができる。

【0174】

HRD陽性腫瘍に有効な処置は、1つ以上のPARP阻害剤および／または1つ以上の白金系薬剤を含むことができる。PARP阻害剤には、ベリパリブ、オラパリブ、タラゾパリブ、イニパリブ、ルカパリブ、およびニラパリブが含まれ得るが、これらに限定されない。PARP阻害剤は、Murphy and Muggia, PARP inhibitors: clinical development, emerging differences, and the current therapeutic issues, Cancer Drug Resist 2019; 2: 665-79に記載されている。白金系薬物としては、シスプラチン、オキサリプラチンおよびカルボプラチンが挙げられ得るが、これらに限定されない。白金系薬物は、Rottenberg et al., The rediscovery of platinum-based cancer therapy, Nat. Rev. Cancer 2021 Jan; 21 (1): 37-50に記載されている。

10

【0175】

治療される腫瘍は、対象の腫瘍である。一実施形態では、腫瘍は膵臓がんである。別の実施形態では、腫瘍は前立腺がんである。いくつかの実施形態において、腫瘍は、卵巣がん、乳がんまたは前立腺がんである。いくつかの実施形態では、腫瘍は、HRDに関連する腫瘍であり、限定されないが、副腎、胆管、骨／軟部組織、乳房、結腸／直腸、食道、眼、頭頸部、腎臓、肝臓、肺、リンパ系、髄芽腫、中皮腫、骨髄系、神経系、神経内分泌、卵巣、膵臓、前立腺、皮膚、胃、精巣、胸腺、甲状腺、尿路、子宮、または外陰がんの1つを含み得る。Nguyen et al., Pan-cancer landscape of homologous recombination deficiency, Nat. Commun. 2020 Nov 4; 11 (1): 5584を参照されたい。

20

【0176】

本開示は、添付の図面を参照して十分に説明されているが、様々な変形および変更が当業者にとって明らかになることに留意されたい。そのような変形および変更は、特許請求の範囲によって定義される本開示の範囲の中に含まれると理解されるべきである。

【0177】

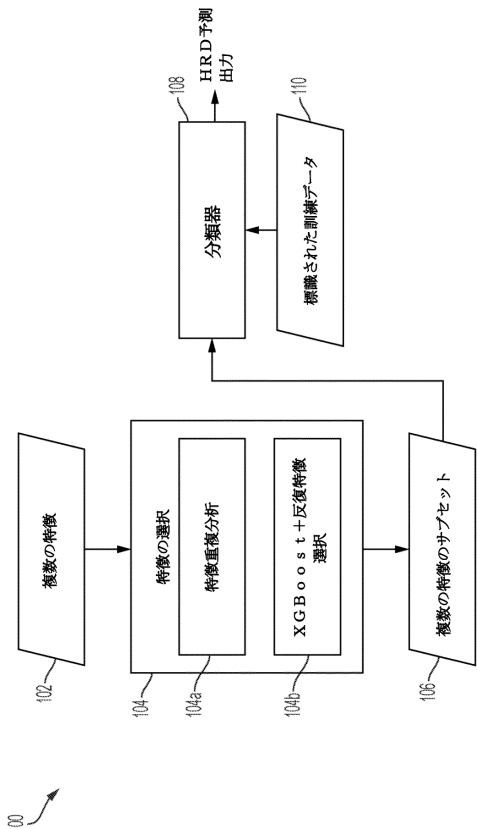
上記の説明は、説明を目的として、特定の実施形態を参照して説明されている。しかしながら、上記の例示的な説明は、網羅的であること、または本発明を開示された正確な形態に限定することを意図するものではない。上記の教示を考慮して、多くの変更および変形が可能である。実施形態は、技術の原理およびそれらの実際の用途を最もよく説明するために選択および記載された。それにより、他の当業者は、企図される特定の使用に適した様々な変更を伴う技術および様々な実施形態を最良に利用することが可能になる。

30

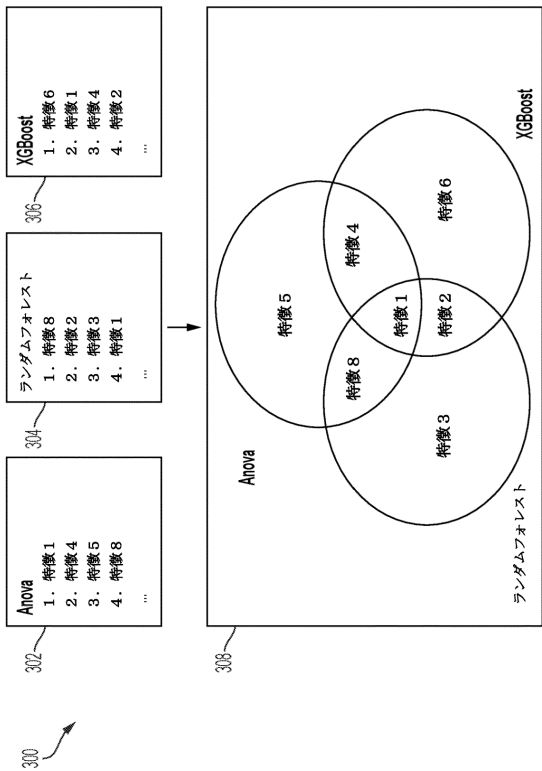
40

50

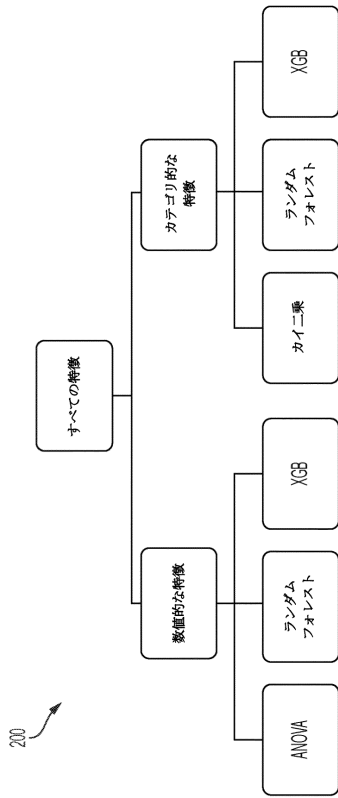
【図面】
【図 1】



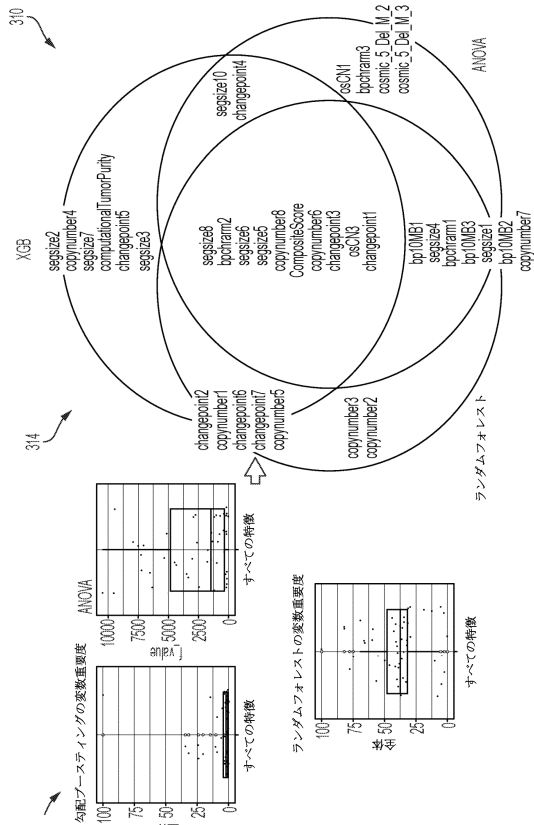
【図 3 A】



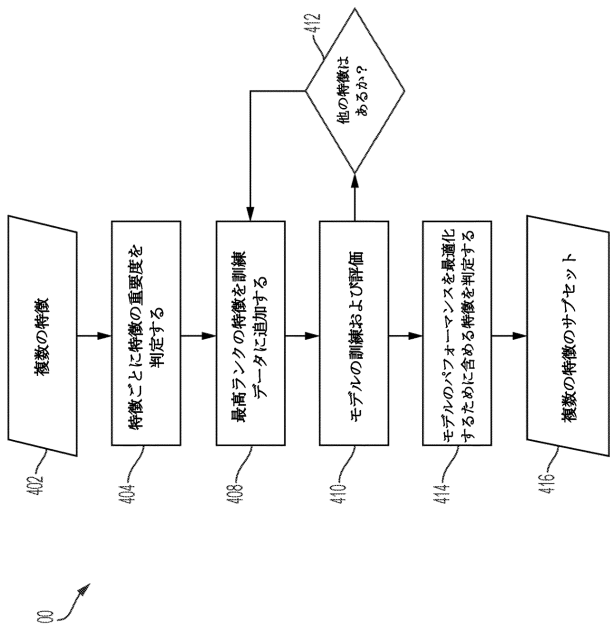
【図 2】



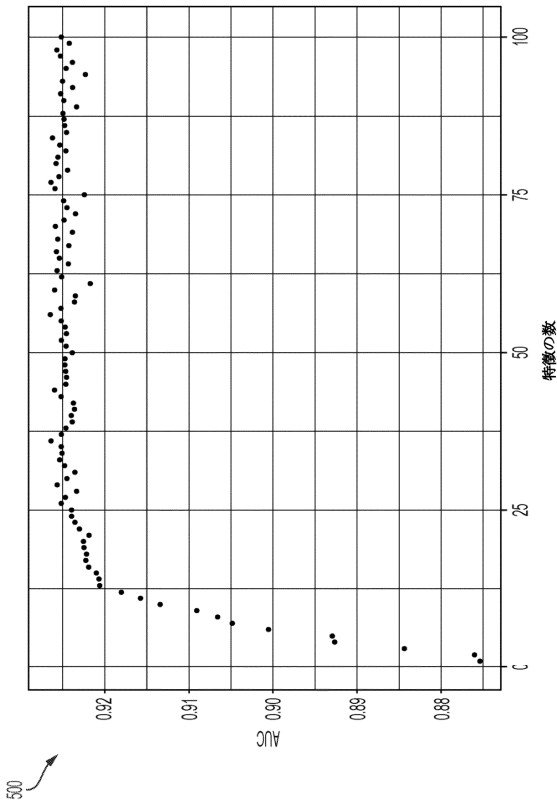
【図 3 B】



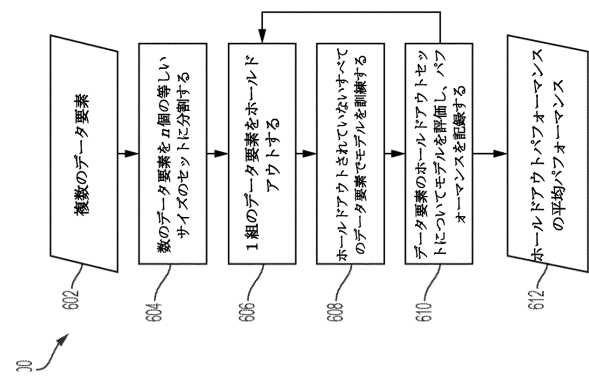
【図 4】



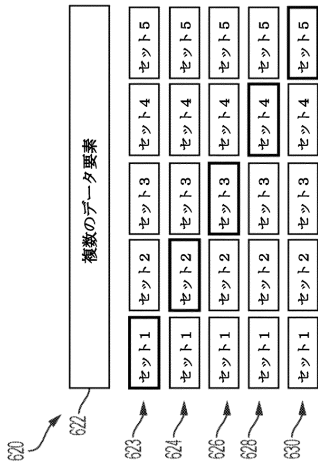
【図 5】



【図 6 A】



【図 6 B】



10

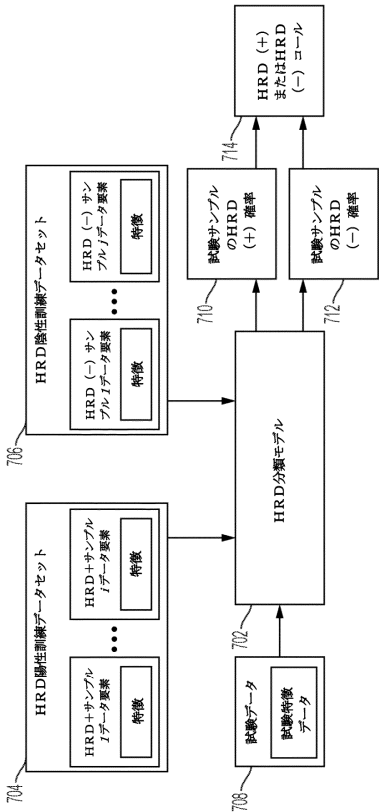
20

30

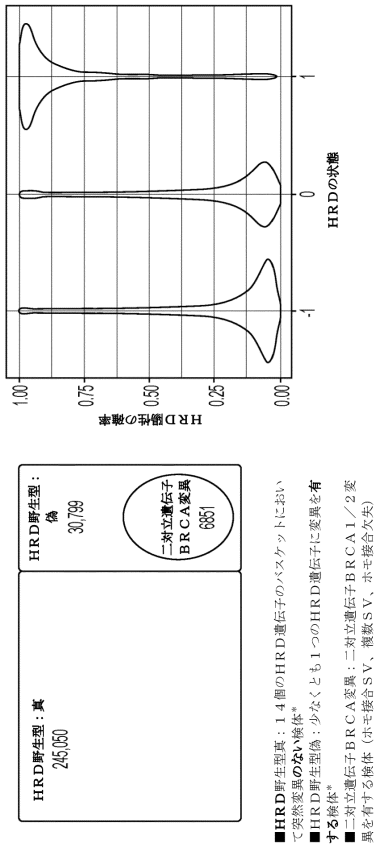
40

50

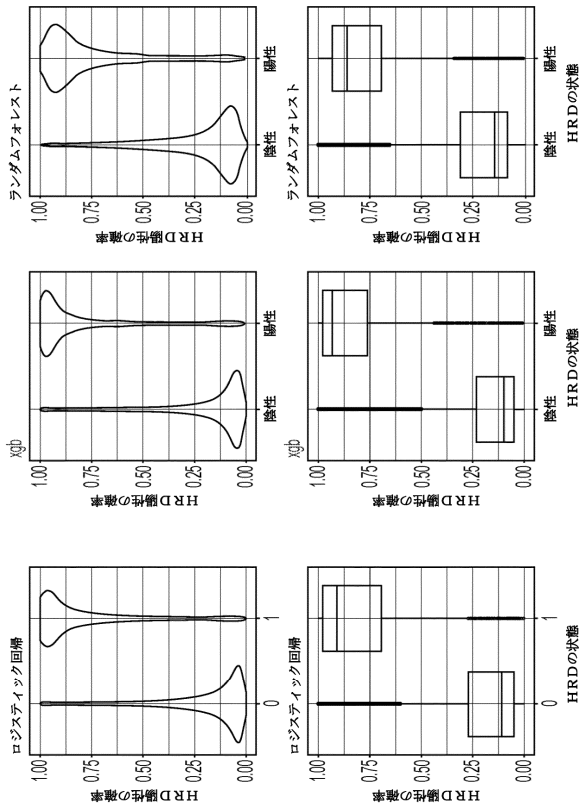
【図 7】



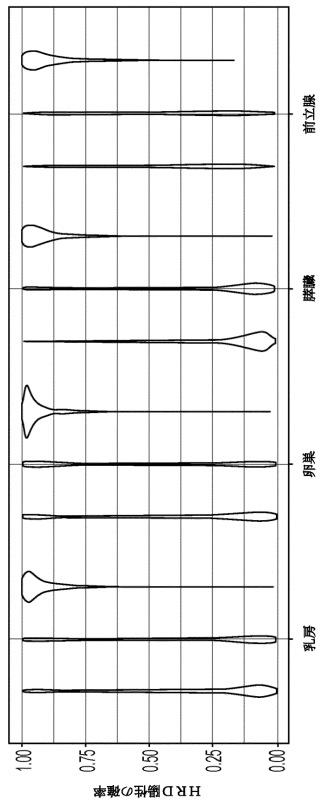
【図 9】



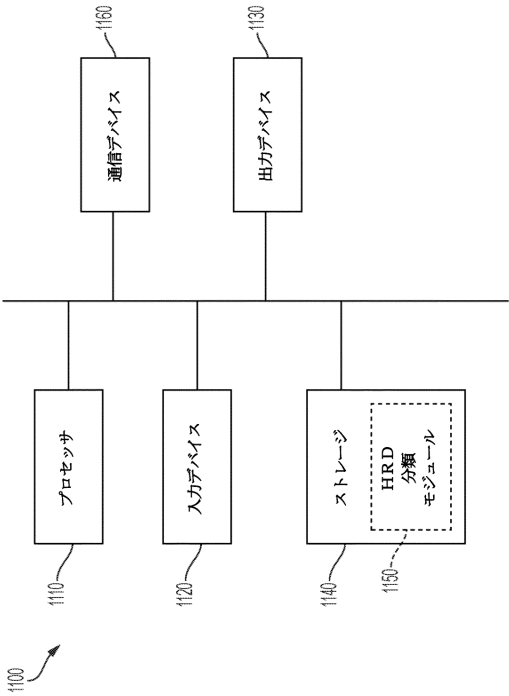
【図 8】



【図 10】



【図 11】



10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US22/73167
A. CLASSIFICATION OF SUBJECT MATTER IPC - INV. G16B 20/10; C12Q 1/6886; G06F 17/18; G16B 20/20; G16B 40/20 (2022.01) ADD. CPC - INV. G16B 20/10; C12Q 1/6886; G06F 17/18; G16B 20/20; G16B 40/20 ADD. According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) See Search History document Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched See Search History document Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) See Search History document		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2020/0255909 A1 (TEMPUS) 13 August 2020; figures 1A-1B and 2; paragraphs [0047, 0051, 0057, 0061, 0064, 0068, 0069, 0072, 0081-0082, 0092, 0113-0115, 0127, 0147, 0149].	1-4
A	US 2021/0172024 A1 (TEMPUS) 10 June 2021; entire document.	1-4
A	US 2018/0030544 A1 (MYRIAD GENETICS INC) 01 February 2018; entire document	1-4
A	US 2021/0166785 A1 (TEMPUS LABS, INC . .) 03 June 2021; entire document	1-4
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 18 August 2022 (18.08.2022)		Date of mailing of the international search report SEP 06 2022
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300		Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

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

INTERNATIONAL SEARCH REPORT		International application No. PCT/US22/73167
Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)		
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	☐	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	☐	Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.	☒	Claims Nos.: 5-73 because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows:		
1.	☐	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.	☐	As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3.	☐	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.	☐	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest	☐	The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
	☐	The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
	☐	No protest accompanied the payment of additional search fees.

10

20

30

40

50

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,N
E,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,
CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,IT,JM,J
O,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,M
Z,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,
TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

(特許庁注：以下のものは登録商標)

1. J A V A

2. P Y T H O N

アメリカ合衆国 0 2 1 4 1 マサチューセッツ， ケンブリッジ， セカンド ストリート 1 5 0，
シー／オー ファウンデーション メディシン インコーポレイテッド

(72)発明者 ニューバーグ， ジャスティン

アメリカ合衆国 0 2 1 4 1 マサチューセッツ， ケンブリッジ， セカンド ストリート 1 5 0，
シー／オー ファウンデーション メディシン インコーポレイテッド

(72)発明者 ジン， デクスター

アメリカ合衆国 0 2 1 4 1 マサチューセッツ， ケンブリッジ， セカンド ストリート 1 5 0，
シー／オー ファウンデーション メディシン インコーポレイテッド

(72)発明者 チェン， クイーティン

アメリカ合衆国 0 2 1 4 1 マサチューセッツ， ケンブリッジ， セカンド ストリート 1 5 0，
シー／オー ファウンデーション メディシン インコーポレイテッド