(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2025-513786
(P2025-513786A)
(43)公表日　令和7年4月30日(2025.4.30)

(51)国際特許分類		FI			テーマコード（参考）
Ｃ１２Ｑ	*1/6827(2018.01)*	Ｃ１２Ｑ	1/6827	Ｚ	４Ｂ０６３
Ｇ１６Ｂ	*40/00 (2019.01)*	Ｇ１６Ｂ	40/00		

審査請求　未請求　予備審査請求　未請求　（全156頁）

(21)出願番号　特願2024-559196(P2024-559196)
(86)(22)出願日　令和5年4月7日(2023.4.7)
(85)翻訳文提出日　令和6年12月3日(2024.12.3)
(86)国際出願番号　PCT/US2023/065560
(87)国際公開番号　WO2023/197004
(87)国際公開日　令和5年10月12日(2023.10.12)
(31)優先権主張番号　63/328,602
(32)優先日　令和4年4月7日(2022.4.7)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　63/336,852
(32)優先日　令和4年4月29日(2022.4.29)
(33)優先権主張国・地域又は機関
米国(US)
(81)指定国・地域　AP(BW,CV,GH,GM,KE,LR,LS,MW,MZ
最終頁に続く

(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 ９４３０４， パロ アルト， ハノーバー ストリート ３１００
(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
(74)代理人　230113332
最終頁に続く

(54)【発明の名称】　無細胞核酸分子のメチル化状態に基づいた腫瘍の存在の検出

(57)【要約】
１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む。
【選択図】図２

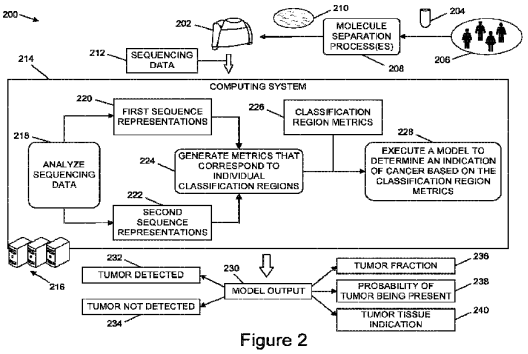


Figure 2

【特許請求の範囲】
【請求項１】
１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、前記複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有する前記ヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと、
前記コンピューティングシステムにより、前記訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する前記訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、前記複数の分類領域の前記個々の分類領域の少なくとも一部分が、がんが検出される対象において前記閾値量のメチル化シトシンを有し、かつ少なくとも前記閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、前記訓練用配列決定リードを解析して、複数の対照領域に対応する前記訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、前記複数の対照領域の個々の対照領域が、少なくとも前記閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも前記閾値量のメチル化シトシンを有する前記参照ゲノムの追加のゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、前記複数の分類領域の前記個々の分類領域についてのメトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、
前記コンピューティングシステムにより、訓練用対象の試料由来の訓練用配列リードについての前記複数の分類領域の前記個々の分類領域についての前記メトリクスを含む訓練用データを生成するステップと、
前記コンピューティングシステムにより、前記訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、前記複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて対象にがんが存在する指標を決定するためのモデルを生成するステップであって、前記モデルが、前記複数の分類領域の個々の分類領域に対する重みを含み、前記個々の分類領域の前記重みの少なくとも一部分が、互いに異なる、ステップと
を含む方法。
【請求項２】
前記コンピューティングシステムにより、前記複数の対象に含まれない追加の対象由来のテスト用配列データを得るステップであって、前記テスト用配列データが、前記追加の対象の試料に由来するテスト用配列決定リードを含み、個々のテスト用配列決定リードが、前記追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々のテスト用配列決定リードが、少なくとも前記閾値シトシン－グアニン含量を有する前記ヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと、
前記モデルおよび前記追加の配列データを使用して、前記追加の対象にがんが存在する前記指標を決定するステップと
を含む、請求項１に記載の方法。
【請求項３】
前記コンピューティングシステムにより、前記テスト用配列決定リードを解析して、前記複数の分類領域の前記個々の分類領域に対応する前記テスト用配列決定リードに由来する第１の定量的尺度を決定するステップと、
前記コンピューティングシステムにより、前記テスト用配列決定リードを解析して、前

記複数の対照領域の前記個々の対照領域に対応する前記テスト用配列決定リードに由来する第２の定量的尺度を決定するステップと、
前記コンピューティングシステムにより、前記個々の分類領域についての前記メトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、
前記コンピューティングシステムにより、前記個々の分類領域についての前記メトリクスを含む入力ベクトルを生成するステップと
を含み、
前記モデルで前記入力ベクトルを使用して、前記追加の対象にがんが存在する前記指標を決定する、
請求項２に記載の方法。
【請求項４】
前記１つまたは複数の機械学習アルゴリズムが、１つまたは複数の分類アルゴリズムを含み、
がんが存在する前記指標が、前記追加の対象にがんが存在する確率に対応する、
請求項２または３に記載の方法。
【請求項５】
前記１つまたは複数の機械学習アルゴリズムが、１つまたは複数の回帰アルゴリズムを含み、
前記指標が、前記追加の試料の腫瘍画分の推定値に対応する、
請求項２または３に記載の方法。
【請求項６】
前記訓練用配列決定リードが、前記訓練用配列データの第１の一部分を含み、追加の訓練用配列決定リードが、前記訓練用配列データの第２の一部分を含み、前記追加の訓練用配列決定リードが、前記訓練用配列決定リードとは異なり、
前記方法が、
前記コンピューティングシステムにより、前記訓練用配列データの前記第１の一部分または前記訓練用配列データの前記第２の一部分の少なくとも一方を解析して、前記複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定するステップと、
前記コンピューティングシステムにより、前記個々の試料について、個々の試料に由来する個々の頻度の中で最大値を有する前記個々の頻度に対応する最高頻度を有する、前記複数のバリアントのうちの１つのバリアントを決定するステップと、
前記コンピューティングシステムにより、個々の試料についての腫瘍画分の個々の尺度を、前記個々の試料に由来する前記個々の頻度の前記最大値に基づいて決定するステップと、
を含む、
請求項５に記載の方法。
【請求項７】
前記訓練用データが、前記複数の試料の前記個々の試料についての腫瘍画分の前記個々の尺度を含み、
前記モデルが、前記複数の試料の前記個々の試料についての腫瘍画分の前記個々の尺度に基づいて生成される、
請求項６に記載の方法。
【請求項８】
前記個々の分類領域についての前記メトリクスが、スケーリング係数および誤差補正係数に基づいて決定される、請求項１～７のいずれか一項に記載の方法。
【請求項９】
前記複数の分類領域が、がんが存在する対象から得られた細胞に由来する核酸のゲノム領域におけるシトシンのメチル化率が、がんが存在しない対象から得られた細胞に由来する核酸の前記ゲノム領域におけるシトシンのメチル化率と異なる前記ゲノム領域に個々に

対応する、請求項１～８のいずれか一項に記載の方法。
【請求項１０】
前記複数の分類領域が、第１のがんタイプに対する第１の複数の分類領域に対応し、前記第１の複数の分類領域とは異なる第２の複数の分類領域に基づいて、第２のがんタイプに対する前記モデルを生成することができる、請求項９に記載の方法。
【請求項１１】
前記複数の試料および前記追加の試料が、無細胞核酸を含む、請求項１～１０のいずれか一項に記載の方法。
【請求項１２】
前記コンピューティングシステムにより、前記訓練用データを使用して訓練プロセスを実施して、前記モデルを生成するステップであって、前記訓練プロセスが、
前記コンピューティングシステムにより、前記訓練用データ内に含まれる個々の試料の１つまたは複数の追加の重みを、閾値信頼水準の範囲内にある前記個々の試料についてのがんの前記指標に基づいて決定すること
を含む、ステップを含む、請求項１～１１のいずれか一項に記載の方法。
【請求項１３】
個々の試料についてのがんの前記指標が、前記閾値信頼水準の範囲外であり、前記方法が、
前記コンピューティングシステムにより、前記訓練プロセス中に前記個々の試料の重みに対してペナルティを適用するステップを含む、
請求項１２に記載の方法。
【請求項１４】
前記コンピューティングシステムにより、前記１つまたは複数の機械学習アルゴリズムを使用して、前記モデルに対して、前記訓練用データの一部分を使用して前記訓練プロセスの１回または複数の第１の反復を実施するステップと、
前記コンピューティングシステムにより、前記訓練プロセスの前記１回または複数の第１の反復に基づいて、前記モデルについての第１の出力データを生成するステップであって、前記第１の出力データが、前記複数の対象のうちの第１の個々の対象にがんが存在する１つまたは複数の第１の追加の指標に対応し、前記第１の個々の対象が、前記訓練用データの前記一部分に対応する、
ステップと
を含む、請求項１２または１３に記載の方法。
【請求項１５】
前記コンピューティングシステムにより、前記第１の出力データと前記訓練用データを組み合わせて、追加の訓練用データを産生するステップと、
前記コンピューティングシステムにより、前記モデルに対して、前記追加の訓練用データの一部分を使用して前記訓練プロセスの１回または複数の第２の反復を実施するステップと、
前記コンピューティングシステムにより、前記訓練プロセスの前記１回または複数の第２の反復に基づいて前記モデルについての第２の出力データを生成するステップであって、前記第２の出力データが、前記複数の対象のうちの第２の個々の対象にがんが存在する１つまたは複数の第２の追加の指標を示し、前記第２の個々の対象が、前記追加の訓練用データの前記一部分に対応する、
ステップと
を含む、請求項１４に記載の方法。
【請求項１６】
前記複数の分類領域の前記個々の分類領域に対する前記重みが、前記第１の出力データおよび前記第２の出力データに基づいて決定される、請求項１５に記載の方法。
【請求項１７】
前記コンピューティングシステムにより、前記訓練プロセスの１回または複数の反復の

間に決定された、いくつかの、がんが存在する指標が、少なくとも、前記訓練用データに含められた１つまたは複数の試料についての閾値であることを決定するステップと、
前記コンピューティングシステムにより、前記モデルの１つまたは複数の重みに対する修正が、修正されないかまたは最小量で修正されることを決定するステップと
を含む、請求項１２～１６のいずれか一項に記載の方法。
【請求項１８】
前記コンピューティングシステムにより、前記訓練プロセスの前記１回または複数の反復の間に決定された、がんが存在する追加の指標の数が、前記訓練用データに含められた１つまたは複数の追加の試料についての前記閾値未満であることを決定するステップと、
前記コンピューティングシステムにより、前記モデルの１つまたは複数の追加の重みに対する修正が、前記最小量よりも多く修正されることを決定するステップと
を含む、請求項１７に記載の方法。
【請求項１９】
対象の血液または組織の少なくとも一方に由来する複数の核酸を、ある１つの量のメチル結合ドメイン（ＭＢＤ）タンパク質を含む溶液と組み合わせて、核酸－ＭＢＤタンパク質溶液を産生するステップと、
塩類溶液を用いた前記核酸－ＭＢＤタンパク質溶液の複数の洗浄を実施して、いくつかの核酸画分を産生するステップであって、個々の核酸画分が、少なくとも前記閾値シトシン－グアニン含量を有する前記複数の核酸の領域内に閾値数のメチル化シトシンを有する、ステップと
を含む、請求項１～１８のいずれか一項に記載の方法。
【請求項２０】
前記複数の洗浄のうちの１つの洗浄を、１つの濃度の塩化ナトリウム（ＮａＣｌ）を有する溶液を用いて実施し、ＭＢＤタンパク質に対する一連の結合強度を有する前記いくつかの核酸画分のうちの１つの核酸画分を産生する、請求項１９に記載の方法。
【請求項２１】
第１の核酸画分が核酸の複数の分配のうちの第１の分配と関連することを決定するステップであって、前記第１の分配が、ＭＢＤタンパク質に対する結合強度の第１の範囲に対応する、ステップと、
第１の分子バーコードを前記第１の核酸画分の核酸に結合させるステップであって、前記第１の分子バーコードが、前記第１の分配と関連する分子バーコードの第１のセット内に含まれる、ステップと、
第２の核酸画分が核酸の前記複数の分配のうちの第２の分配と関連することを決定するステップであって、前記第２の分配が、ＭＢＤタンパク質に対する結合強度の前記第１の範囲とは異なる、ＭＢＤタンパク質に対する結合エネルギーの第２の範囲に対応する、ステップと、
第２の分子バーコードを前記第２の核酸画分の核酸に結合させるステップであって、前記第２の分子バーコードが、前記第２の分配と関連する分子バーコードの第２のセット内に含まれる、ステップと
を含む、請求項１９または２０に記載の方法。
【請求項２２】
前記いくつかの核酸画分の少なくとも一部分を、ある１つの量の、１つまたは複数の非メチル化シトシンを有する分子を切断する制限酵素と組み合わせて、前記配列決定リードを産生するために使用される前記複数の試料の少なくとも一部分を産生するステップであって、
前記閾値量のメチル化シトシンが、少なくとも前記閾値シトシン－グアニン含量を有する領域内のメチル化シトシンの最低頻度に対応する、ステップ
を含む、請求項１９～２１のいずれか一項に記載の方法。
【請求項２３】
前記いくつかの核酸画分の少なくとも一部分を、ある１つの量の、１つまたは複数のメ

チル化シトシンを有する分子を切断する制限酵素と組み合わせて、前記配列決定リードを産生するために使用される前記複数の試料の少なくとも一部分を産生するステップであって、

前記閾値量の非メチル化シトシンが、少なくとも前記閾値シトシン－グアニン含量を有する領域内の切断されないメチル化シトシンの最高頻度に対応する、ステップ

を含む、請求項１９～２１のいずれか一項に記載の方法。

【請求項２４】

試料の腫瘍画分を決定するための前記モデルの検出限界が、０．０５％以下である、請求項１～２３のいずれか一項に記載の方法。

【請求項２５】

１つまたは複数のハードウェアプロセッサーと、

前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、

複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、前記複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有する前記ヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと、

前記訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する前記訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、前記複数の分類領域の前記個々の分類領域の少なくとも一部分が、がんが検出される対象において前記閾値量のメチル化シトシンを有し、かつ少なくとも前記閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップと、

前記訓練用配列決定リードを解析して、複数の対照領域に対応する前記訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、前記複数の対照領域の個々の対照領域が、少なくとも前記閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも前記閾値量のメチル化シトシンを有する前記参照ゲノムの追加のゲノム領域に対応する、ステップと、

前記複数の分類領域の前記個々の分類領域についてのメトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、

訓練用対象の試料由来の訓練用配列リードについての前記複数の分類領域の個々の前記分類領域についての前記メトリクスを含む訓練用データを生成するステップと、

前記訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、前記複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて、対象にがんが存在する指標を決定するためのモデルを生成するステップであって、前記モデルが、前記複数の分類領域の個々の分類領域に対する重みを含み、前記個々の分類領域の前記重みの少なくとも一部分が、互いに異なる、ステップと

を含む操作を実施させるコンピュータ可読命令を含む１つまたは複数の非一時的コンピュータ可読ストレージ媒体と

を含むコンピューティングシステム。

【請求項２６】

前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、

前記複数の対象に含まれない追加の対象由来のテスト用配列データを得るステップであって、前記テスト用配列データが、前記追加の対象の試料に由来するテスト用配列決定リードを含み、個々のテスト用配列決定リードが、前記追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々のテスト用配列決定リードが、少なくとも前記閾

値シトシン－グアニン含量を有する前記ヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと、
前記モデルおよび前記追加の配列データを使用して、前記追加の対象にがんが存在する前記指標を決定するステップと
を含む追加の操作を実施させる追加のコンピュータ可読命令を記憶している、請求項２５に記載のコンピューティングシステム。
【請求項２７】
前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
前記テスト用配列決定リードを解析して、前記複数の分類領域の前記個々の分類領域に対応する前記テスト用配列決定リードに由来する第１の定量的尺度を決定するステップと、
前記テスト用配列決定リードを解析して、前記複数の対照領域の前記個々の対照領域に対応する前記テスト用配列決定リードに由来する第２の定量的尺度を決定するステップと、
前記個々の分類領域についての前記メトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、
前記個々の分類領域についての前記メトリクスを含む入力ベクトルを生成するステップと
を含む追加の操作を実施させる追加のコンピュータ可読命令を記憶しており、
前記モデルで前記入力ベクトルを使用して、前記追加の対象にがんが存在する前記指標を決定する、
請求項２６に記載のコンピューティングシステム。
【請求項２８】
前記１つまたは複数の機械学習アルゴリズムが、１つまたは複数の分類アルゴリズムを含み、
がんが存在する前記指標が、前記追加の対象にがんが存在する確率に対応する、
請求項２６または２７に記載のコンピューティングシステム。
【請求項２９】
前記１つまたは複数の機械学習アルゴリズムが、１つまたは複数の回帰アルゴリズムを含み、
指標が、前記追加の試料の腫瘍画分の推定値に対応する、
請求項２６または２７に記載のコンピューティングシステム。
【請求項３０】
前記訓練用配列決定リードが、前記訓練用配列データの第１の一部分を含み、前記訓練用配列データの第２の一部分が、前記訓練用配列決定リードとは異なる追加の訓練用配列決定リードを含み、
前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
前記訓練用配列データの前記第１の一部分または前記訓練用配列データの前記第２の一部分の少なくとも一方を解析して、前記複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定するステップと、
前記個々の試料について、個々の試料に由来する個々の頻度の中で最大値を有する前記個々の頻度に対応する最高頻度を有する、前記複数のバリアントのうちの１つのバリアントを決定するステップと、
個々の試料についての腫瘍画分の個々の尺度を、前記個々の試料に由来する前記個々の頻度の前記最大値に基づいて決定するステップと

を含む追加の操作を実施させる追加のコンピュータ可読命令を記憶している、
請求項２９に記載のコンピューティングシステム。
【請求項３１】
前記訓練用データが、前記複数の試料の前記個々の試料についての腫瘍画分の前記個々の尺度を含み、
前記モデルが、前記複数の試料の前記個々の試料についての腫瘍画分の前記個々の尺度に基づいて生成される、
請求項３０に記載のコンピューティングシステム。
【請求項３２】
前記個々の分類領域についての前記メトリクスが、スケーリング係数および誤差補正係数に基づいて決定される、請求項２５～３１のいずれか一項に記載のコンピューティングシステム。
【請求項３３】
前記複数の分類領域が、がんが存在する対象から得られた細胞に由来する核酸のゲノム領域におけるシトシンのメチル化率が、がんが存在しない対象から得られた細胞に由来する核酸のゲノム領域におけるシトシンのメチル化率と異なるゲノム領域に個々に対応する、請求項２５～３２のいずれか一項に記載のコンピューティングシステム。
【請求項３４】
前記複数の分類領域が、第１のがんタイプに対する第１の複数の分類領域に対応し、前記第１の複数の分類領域とは異なる第２の複数の分類領域に基づいて、第２のがんタイプに対する前記モデルを生成することができる、請求項３３に記載のコンピューティングシステム。
【請求項３５】
前記複数の試料および前記追加の試料が、無細胞核酸を含む、請求項２５～３４のいずれか一項に記載のコンピューティングシステム。
【請求項３６】
前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
前記訓練用データを使用して訓練プロセスを実施して、前記モデルを生成するステップであって、前記訓練プロセスが、
前記訓練用データ内に含まれる個々の試料の１つまたは複数の追加の重みを、閾値信頼水準の範囲内にある前記個々の試料についてのがんの前記指標に基づいて決定すること
を含む、ステップを含む追加の操作を実施させる追加のコンピュータ可読命令を記憶している、
請求項２５～３５のいずれか一項に記載のコンピューティングシステム。
【請求項３７】
個々の試料についてのがんの前記指標が、前記閾値信頼水準の範囲外にあり、
前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
前記訓練プロセス中に前記個々の試料の重みに対してペナルティを適用するステップ
を含む追加の操作を実施させる追加のコンピュータ可読命令を記憶している、
請求項３６に記載のコンピューティングシステム。
【請求項３８】
前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
前記１つまたは複数の機械学習アルゴリズムを使用して、前記モデルに対して、前記訓練用データの一部分を使用して前記訓練プロセスの１回または複数の第１の反復を実施す

るステップと、
前記モデルについての第１の出力データを、前記訓練プロセスの前記１回または複数の第１の反復に基づいて生成するステップであって、前記第１の出力データが、前記複数の対象のうちの第１の個々の対象にがんが存在する１つまたは複数の第１の追加の指標に対応し、前記第１の個々の対象が、前記訓練用データの前記一部分に対応する、ステップと
を含む追加の操作を実施させる追加のコンピュータ可読命令を記憶している、
請求項２６または３７に記載のコンピューティングシステム。
【請求項３９】
前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
前記第１の出力データと前記訓練用データを組み合わせて、追加の訓練用データを産生するステップと、
前記モデルに対して、前記追加の訓練用データの一部分を使用して前記訓練プロセスの１回または複数の第２の反復を実施するステップと、
前記モデルについての第２の出力データを、前記訓練プロセスの前記１回または複数の第２の反復に基づいて生成するステップであって、前記第２の出力データが、前記複数の対象のうちの第２の個々の対象にがんが存在する１つまたは複数の第２の追加の指標を示し、前記第２の個々の対象が、前記追加の訓練用データの前記一部分に対応する、ステップと
を含む追加の操作を実施させる追加のコンピュータ可読命令を記憶している、
請求項３８に記載のコンピューティングシステム。
【請求項４０】
前記複数の分類領域の前記個々の分類領域に対する前記重みが、前記第１の出力データおよび前記第２の出力データに基づいて決定される、請求項３９に記載のコンピューティングシステム。
【請求項４１】
前記１つまたは複数の非一時的コンピュータ可読ストレージ媒体が、前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
前記訓練プロセスの１回または複数の反復の間に決定された、いくつかの、がんが存在する指標が、少なくとも、前記訓練用データに含められた１つまたは複数の試料についての閾値であることを決定するステップと、
前記モデルの１つまたは複数の重みに対する修正が、修正されないかまたは最小量で修正されることを決定するステップと
を含む追加の操作を実施させる追加のコンピュータ可読命令を記憶している、
請求項３６～４０のいずれか一項に記載のコンピューティングシステム。
【請求項４２】
前記訓練プロセスの前記１回または複数の反復の間に決定された、がんが存在する追加の指標の数が、前記訓練用データに含められた１つまたは複数の追加の試料についての前記閾値未満であることを決定するステップと、
前記モデルの１つまたは複数の追加の重みに対する修正が、前記最小量よりも多く修正されることを決定するステップと
を含む、請求項４１に記載のコンピューティングシステム。
【請求項４３】
１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、対象から得られた試料に由来する配列決定リードを得るステップであって、個々の配列決定リードが、前記試料中に含まれる核酸の断片に対応するヌクレオチド配列を含み、かつ、少なくとも閾値シトシン－グアニン含量を有する前記ヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと

、
前記コンピューティングシステムにより、複数の分類領域の個々の分類領域に対応する配列決定リードに由来する第１の定量的尺度を決定するステップであって、前記複数の分類領域の前記個々の分類領域の少なくとも一部分が、がんが検出される対象において前記閾値量のメチル化シトシンを有し、かつ少なくとも前記閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、前記配列決定リードを解析して、複数の対照領域に対応する配列決定リードに由来する第２の定量的尺度を決定するステップであって、前記複数の対照領域の個々の対照領域が、少なくとも前記閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも前記閾値量のメチル化シトシンを有する前記参照ゲノムの追加のゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、複数のメトリクスのうちの個々のメトリクスが前記複数の分類領域の個々の分類領域に対応する、前記複数のメトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、
前記コンピューティングシステムにより、前記対象にがんが存在する指標を、前記複数のメトリクスの少なくとも一部分に基づいて決定するステップと
を含む方法。
【請求項４４】
前記コンピューティングシステムにより、前記配列決定データを使用して、差次的にメチル化された領域についての配列表示の分布を決定するステップと、
前記コンピューティングシステムにより、前記分布に含まれる少なくとも閾値量の前記配列表示が、前記差次的にメチル化された領域の部分領域と重複することを決定するステップと、
前記コンピューティングシステムにより、前記差次的にメチル化された領域の前記部分領域が、前記複数の分類領域のうちの１つの分類領域であることを決定するステップと
を含む、請求項４３に記載の方法。
【請求項４５】
前記閾値量の配列表示が、前記分布に含まれる前記配列表示の少なくとも約７０％である、請求項４４に記載の方法。
【請求項４６】
前記コンピューティングシステムにより、前記複数のメトリクスの値の順序を決定するステップと、
前記コンピューティングシステムにより、前記複数の分類領域の中からの分類領域のサブセットを前記順序に基づいて決定するステップと、
を含み、
前記分類領域の前記サブセットに対応する前記複数のメトリクスの一部分を使用して、前記対象にがんが存在する前記指標を決定する、
請求項４３～４５のいずれか一項に記載の方法。
【請求項４７】
前記対象にがんが存在する前記指標が、前記対象にがんが存在する最初の指標であり、前記方法が、
前記コンピューティングシステムにより、前記対象にがんが存在する前記最初の指標にスケーリング係数を適用して、前記対象にがんが存在する改変された指標を決定するステップ
を含む、請求項４３～４６のいずれか一項に記載の方法。
【請求項４８】
前記対象にがんが存在する前記指標が、腫瘍画分に対応する、請求項４３～４７のいずれか一項に記載の方法。

【請求項４９】
前記試料が、がんの処置の開始前または開始時の少なくとも一方において収集された第１の試料であり、前記方法が、
１つまたは複数のハードウェアプロセッサーおよびメモリーを有する前記コンピューティングシステムにより、前記対象から得られた第２の試料に由来する追加の配列決定リードを得るステップであって、個々の追加の配列決定リードが、前記第２の試料中に含まれる核酸の断片に対応する追加のヌクレオチド配列を含み、かつ、少なくとも前記閾値シトシン－グアニン含量を有する前記追加のヌクレオチド配列の領域内に含まれる前記閾値量のメチル化シトシンを有する追加の分子に対応する、ステップと、
前記コンピューティングシステムにより、前記複数の分類領域の前記個々の分類領域に対応する前記追加の配列決定リードに由来する追加の第１の定量的尺度を決定するステップであって、前記複数の分類領域の前記個々の分類領域の少なくとも一部分が、がんが検出される対象において前記閾値量のメチル化シトシンを有し、かつ少なくとも前記閾値シトシン－グアニン含量を有する前記参照ゲノムのゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、前記追加の配列決定リードを解析して、複数の対照領域に対応する前記追加の配列決定リードに由来する追加の第２の定量的尺度を決定するステップであって、前記複数の対照領域の個々の対照領域が、少なくとも前記閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも前記閾値量のメチル化シトシンを有する前記参照ゲノムの追加のゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、複数の追加のメトリクスのうちの個々の追加のメトリクスが前記複数の分類領域の個々の分類領域に対応する、前記複数の追加のメトリクスを、前記個々の分類領域についての前記追加の第１の定量的尺度および前記複数の対照領域についての前記追加の第２の定量的尺度に基づいて決定するステップと、
前記コンピューティングシステムにより、前記対象にがんが存在する追加の指標を、前記複数の追加のメトリクスの少なくとも一部分に基づいて決定するステップと
を含む、請求項４３～４８のいずれか一項に記載の方法。
【請求項５０】
前記第２の試料が、前記対象にがんの前記処置が投与されてから少なくとも１週間後に得られる、請求項４９に記載の方法。
【請求項５１】
前記コンピューティングシステムにより、前記対象にがんが存在する前記指標を前記対象にがんが存在する前記追加の指標と関連付けて解析して、前記対象についての前記処置に対する応答を決定するステップを含む、
請求項４９または５０に記載の方法。
【請求項５２】
１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、対象からテスト用配列データを得るステップであって、前記テスト用配列データが、前記対象の試料に由来するテスト用配列決定リードを含み、個々のテスト用配列決定リードが、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々のテスト用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有する前記ヌクレオチド配列の領域内に含まれる、閾値量のメチル化シトシンを有する分子に対応する、ステップと、
前記コンピューティングシステムにより、前記テスト用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する前記テスト用配列決定リードに由来する第１の定量的尺度を決定するステップであって、前記複数の分類領域の前記個々の分類領域の少なくとも一部分が、がんが検出される対象において前記閾値量のメチル化シトシンを有し、かつ少なくとも前記閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、前記テスト用配列決定リードを解析して、複

数の対照領域の個々の対照領域に対応する前記テスト用配列決定リードに由来する第２の定量的尺度を決定するステップであって、前記複数の対照領域の個々の対照領域が、少なくとも前記閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも前記閾値量のメチル化シトシンを有する前記参照ゲノムの追加のゲノム領域に対応する、ステップと、

前記コンピューティングシステムにより、前記個々の分類領域についてのメトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、

前記コンピューティングシステムにより、前記個々の分類領域についての前記メトリクスを含む入力ベクトルを生成するステップと、

前記コンピューティングシステムにより、前記対象にがんが存在する指標を、１つまたは複数の機械学習技術をインプリメントするモデルに前記入力ベクトルを提供して、対象にがんが存在する指標を生成することによって決定するステップであって、前記モデルが、前記複数の分類領域の個々の分類領域に対する重みを含み、前記個々の分類領域に対する前記重みの少なくとも一部分が、互いに異なる、ステップと

を含む方法。

【請求項５３】

１つまたは複数のハードウェアプロセッサーおよびメモリーを有する前記コンピューティングシステムにより、複数の訓練用対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、前記複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有する前記ヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと、

前記コンピューティングシステムにより、前記訓練用配列決定リードを解析して、前記複数の分類領域の個々の分類領域に対応する前記訓練用配列決定リードに由来する追加の第１の定量的尺度を決定するステップと、

前記コンピューティングシステムにより、前記訓練用配列決定リードを解析して、複数の対照領域に対応する前記訓練用配列決定リードに由来する追加の第２の定量的尺度を決定するステップと、

前記コンピューティングシステムにより、前記複数の分類領域の前記個々の分類領域についての追加のメトリクスを、前記個々の分類領域についての前記追加の第１の定量的尺度および前記複数の対照領域についての前記追加の第２の定量的尺度に基づいて決定するステップと、

コンピューティングデバイスにより、前記複数の訓練用対象の試料由来の訓練用配列リードについての前記複数の分類領域の前記個々の分類領域についての前記追加のメトリクスを含む訓練用データを生成するステップと、

前記コンピューティングシステムにより、前記訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、対象にがんが存在する前記指標を、前記複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて決定するための前記モデルを生成するステップと

を含む、請求項５２に記載の方法。

【請求項５４】

前記１つまたは複数の機械学習アルゴリズムが、１つまたは複数の分類アルゴリズムを含み、

がんが存在する前記指標が、前記対象にがんが存在する確率に対応する、

請求項５２または５３に記載の方法。

【請求項５５】

前記１つまたは複数の機械学習アルゴリズムが、１つまたは複数の回帰アルゴリズムを含み、

前記指標が、前記試料の腫瘍画分の推定値に対応する、
請求項５２または５３に記載の方法。
【請求項５６】
前記訓練用配列決定リードが、前記訓練用配列データの第１の一部分を含み、追加の訓練用配列決定リードが、前記訓練用配列データの第２の一部分を含み、前記追加の訓練用配列決定リードが、前記訓練用配列決定リードとは異なり、
前記方法が、
前記コンピューティングシステムにより、前記訓練用配列データの前記第１の一部分または前記訓練用配列データの前記第２の一部分の少なくとも一方を解析して、前記複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定するステップと、
前記コンピューティングシステムにより、前記個々の試料について、個々の試料に由来する個々の頻度の中で最大値を有する前記個々の頻度に対応する最高頻度を有する、前記複数のバリアントのうちの１つのバリアントを決定するステップと、
前記コンピューティングシステムにより、個々の試料についての腫瘍画分の個々の尺度を、前記個々の試料に由来する前記個々の頻度の前記最大値に基づいて決定するステップと
を含む、
請求項５５に記載の方法。
【請求項５７】
前記訓練用データが、前記複数の試料の前記個々の試料についての腫瘍画分の前記個々の尺度を含み、
前記モデルが、前記複数の試料の前記個々の試料についての腫瘍画分の前記個々の尺度に基づいて生成される、
請求項５６に記載の方法。
【請求項５８】
前記対象の前記試料および前記複数の訓練用対象の前記複数の試料が、無細胞核酸を含む、請求項５２～５７のいずれか一項に記載の方法。
【請求項５９】
１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、複数の対象由来の配列決定データを得るステップであって、前記配列決定データが、前記複数の対象の複数の試料に由来する配列決定リードを含み、個々の配列決定リードが、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の配列決定リードが、少なくとも閾値シトシン－グアニン含量を有する前記ヌクレオチド配列のプロモーター領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと、
前記コンピューティングシステムにより、前記配列決定リードを解析して、前記プロモーター領域に対応する前記配列決定リードに由来する第１の定量的尺度を決定するステップと、
前記コンピューティングシステムにより、前記配列決定リードを解析して、複数の対照領域の個々の対照領域に対応する前記配列決定リードに由来する第２の定量的尺度を決定するステップであって、前記複数の対照領域の個々の対照領域が、少なくとも前記閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも前記閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップと、
前記コンピューティングシステムにより、前記プロモーター領域についてのメトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、
前記コンピューティングシステムにより、前記プロモーター領域のメチル化状態の指標を、少なくとも閾値を有する前記メトリクスに基づいて生成するステップと
を含む方法。

【請求項６０】
１つまたは複数のハードウェアプロセッサーと、
前記１つまたは複数のハードウェアプロセッサーによって実行されるときに、前記１つまたは複数のハードウェアプロセッサーに、
複数の対象由来の配列決定データを得るステップであって、前記配列決定データが、前記複数の対象の複数の試料に由来する配列決定リードを含み、個々の配列決定リードが、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の配列決定リードが、少なくとも閾値シトシン－グアニン含量を有する前記ヌクレオチド配列のプロモーター領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップと、
前記配列決定リードを解析して、前記プロモーター領域に対応する前記配列決定リードに由来する第１の定量的尺度を決定するステップと、
前記配列決定リードを解析して、複数の対照領域の個々の対照領域に対応する前記配列決定リードに由来する第２の定量的尺度を決定するステップであって、前記複数の対照領域の個々の対照領域が、少なくとも前記閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも前記閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップと、
前記プロモーター領域についてのメトリクスを、前記個々の分類領域についての前記第１の定量的尺度および前記複数の対照領域についての前記第２の定量的尺度に基づいて決定するステップと、
前記プロモーター領域のメチル化状態の指標を、少なくとも閾値を有する前記メトリクスに基づいて生成するステップと
を含む操作を実施させるコンピュータ可読命令を含む１つまたは複数の非一時的コンピュータ可読ストレージ媒体と
を含むコンピューティングシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願への相互参照
本出願は、２０２２年４月７日に出願された米国仮特許出願第６３／３２８，６０２号、および２０２２年４月２９日に出願された米国仮特許出願第６３／３３６，８５２号に基づく利益を主張する。どちらもその全体が全ての目的に関して参照により組み込まれる。
【背景技術】
【０００２】
背景
がんは、世界的に疾患の主な原因である。毎年、世界中で数千万人ががんの診断を受け、半数よりも多くが最終的にがんによって死亡している。多くの国において、がんは心血管疾患に次いで２番目に多い死因となっている。早期検出は多くがんの転帰の改善に関連付けられる。
【０００３】
がんは、個体の正常細胞内における、その少なくとも一部に不適切に調節された細胞分裂が生じる、遺伝的バリエーションの蓄積によって引き起こされ得る。そのようなバリエーションには、一般に、コピー数バリエーション（ＣＮＶ）、一塩基バリエーション（ＳＮＶ）、遺伝子融合、挿入および／または欠失（インデル）、シトシンの５－メチル化（５－メチルシトシン）を含むエピジェネティックバリエーション、ならびにＤＮＡのクロマチンおよび転写因子との会合が含まれる。
【０００４】
がんは、多くの場合、腫瘍の生検、その後の、細胞、マーカーまたは細胞から抽出され

たＤＮＡの解析によって検出される。しかし、つい最近、血液または尿などの体液中の無細胞核酸からもがんを検出することができることが提唱された。そのような検査には、非侵襲性であり、生検での疑わしいがん細胞の同定を行わずに実施することができるという利点がある。しかしながら、そのような検査は、体液中の核酸の量が非常に少なく、核酸が不均一な形態（例えば、ＲＮＡおよびＤＮＡ、一本鎖および二本鎖、ならびに、複製後改変およびヒストンなどのタンパク質との会合の様々な状態）で存在する事実によって複雑なものになっている。

【０００５】

したがって、液体生検アッセイを使用する改善されたがん検出のための改善されたシステムおよび方法が必要とされている。したがって、本開示の目的は、試料を、腫瘍由来ＤＮＡを含有すると分類する改善された能力を有する、感度が高められた、コンピュータによりインプリメントされるシステムおよび方法を提供することである。

【０００６】

本明細書に組み込まれ、その一部を構成する付属の図面は、ある特定のインプリメンテーションを例示するものであり、書面による記載（written description）とともに、本明細書に開示される方法、コンピュータ可読媒体、およびシステムのある特定の原理を説明する機能を果たす。本明細書に提供される説明は、限定するものではなく例として含められている付属の図面と併せて読むと、よりよく理解される。文脈からそうでないことが示されない限り、図面全体を通して、同様の参照数字により、同様のコンポーネントが識別されることが理解されよう。図の一部または全部は、例証のために略図の場合があり、必ずしも示されている要素の実際の相対的サイズまたは場所を示すものではないことも理解されよう。

【図面の簡単な説明】

【０００７】

【図１】図１は、参照配列の分類領域に対応する核酸を同定する環境１００の例の図式表示であり、ここで、分類領域は、少なくとも閾値数のＣｐＧを有する。

【０００８】

【図２】図２は、１つまたは複数のインプリメンテーションによる、無細胞核酸分子のメチル化状態を解析する１つまたは複数のモデルに基づいて腫瘍メトリクスを決定するためのアーキテクチャの例の図式表示である。

【０００９】

【図３】図３は、１つまたは複数のインプリメンテーションによる、無細胞核酸分子のメチル化状態に基づいてがんメトリクスを決定するために１つまたは複数の機械学習モデルを訓練するためのアーキテクチャの例の図式表示である。

【００１０】

【図４】図４は、１つまたは複数のインプリメンテーションによる、参照配列の分類領域のメチル化のレベルに関連する腫瘍メトリクスを決定するためのプロセスの例の流れ図である。

【００１１】

【図５】図５は、１つまたは複数のインプリメンテーションの例による、１つまたは複数の機械可読媒体からの命令を読み取り、実行して、本明細書に記載される任意の１つまたは複数の方法論を実施することができる、コンピュータシステムの形態の機械のコンポーネントを例示するブロック図である。

【００１２】

【図６】図６は、１つまたは複数のインプリメンテーションの例による、本明細書に記載される１つまたは複数のハードウェアアーキテクチャと併せて使用することができる代表的なソフトウェアアーキテクチャを例示するブロック図である。

【００１３】

【図７】図７Ａ、７Ｂ、および７Ｃは、８８種のＴＳＧ＋ＨＲＤ遺伝子の間での、訓練用試料および試験試料におけるプロモーター－メチル化コールを示すグラフ表示である。

【００１４】
【図８】図８Ａは、＞１のコールを有するがんのない試料におけるがん予測スコアを示すグラフ表示である。図８Ｂは、がんのないドナーにおいてプロモーターメチル化について最も頻繁にコールされた遺伝子を示す表である。図８Ｃは、細胞株ＫＭ１２由来の選択された遺伝子におけるｉｎ－ｓｉｌｉｃｏ　ＬｏＤ推定値を示す表である。
【００１５】
【図９】図９は、本発明者らのテストデータセット（総Ｎ＝５５９）、およびＴＣＧＡ公開データ（総Ｎ＝２，３８０）における、異なるタイプのがんにおけるＴＳＧ遺伝子およびＨＲＤ遺伝子のプロモーターメチル化の存在率を示すグラフ表示である。ＴＣＧＡにおいてはプロモーターメチル化を有する遺伝子に限定した。
【００１６】
【図１０】図１０Ａは、がんのないドナーおよびＣＲＣ患者におけるＭＬＨ１プロモーターメチル化（ＭＳＩ－ＨおよびＭＳＳ）を示すグラフ表示である。図１０Ｂは、ＣＲＣ患者におけるＭＬＨ１プロモーターメチル化およびＢＲＡＦ－Ｖ６００Ｅのコールを示す表である。
【００１７】
【図１１】図１１は、実施例５の訓練用データセットおよびテストデータセットの概要を示す表である。
【００１８】
【図１２】図１２Ａは、訓練用セットにおけるＣＲＣ／がんのない状態の予測についてのモデル性能を示すグラフのグラフ表示である。影は、反復における変動を示す。図１２Ｂは、独立したテストデータセットに対するがん予測モデルの性能を示す表である。
【００１９】
【図１３】図１３Ａは、ｉｎ－ｖｉｔｒｏデータセットにおけるゲノムコールおよびメチル化からのＴＦ推定値のＣＶを示す表である。図１３Ｂは、ＣＲＣ試料およびがんのない試料の訓練用セットにおけるＴＦモデル性能（斜めの黒線）（交差検証）を示すグラフ表示である。図１３Ｃは、より低い真のＴＦについてのｉｎ－ｓｉｌｉｃｏ　データセットを示すグラフ表示である。
【００２０】
【図１４】図１４は、訓練用セットにおけるドライバー変異を有するおよび有さないＣＲＣ患者についての予測ＴＦの分布を示すグラフ表示である。
【００２１】
【図１５】図１５Ａは、ステージＩ／ＩＩ患者およびステージＩＩＩ／ＩＶ患者における肺がん検出についての、個体における陽性率を示すグラフ表示である。
【００２２】
図１５Ｂは、ステージＩ／ＩＩ患者およびステージＩＩＩ／ＩＶ患者における多発がん検出（膀胱がん、胃がん、卵巣がん、膵臓がん、および肝臓がん）についての、個体における陽性率を示すグラフ表示である。
【００２３】
【図１６】図１６は、ステージＩ患者、ステージＩＩ患者、ステージＩＩＩ患者、およびステージＩＶ患者における多発がん検出（膀胱がん、胃がん、卵巣がん、膵臓がん、および肝臓がん）についての、個体における陽性率を示すグラフ表示である。
【００２４】
【図１７】図１７は、エピゲノムＭＡＦの、結腸直腸がん、肺がん、および乳がんについての標的ＭＡＦとの関連のグラフ表示である。
【００２５】
【図１８】図１８は、エピゲノムｃＴＦの定量的精度が、ＣＲＣ、肺および乳房の臨床試料において０．１％未満のＬｏＱに達することが可能であることを示す表である。
【００２６】
【図１９】図１９Ａは、体細胞変異に基づくｃＴＦが、特に０．５％またはそれよりも高

いｃＴＦレベルにおいて、同じｃＴＦレベル内での反復実験についてロバストであることを示すグラフ表示である。

【００２７】

図１９Ｂは、エピゲノムｃＴＦが、１００％の評価率を維持することができ、０．１％のｃＴＦに至るまでのＬｏＱを有することを示すグラフ表示である。

【００２８】

【図２０】図２０Ａは、臨床的用量設定の第１の反復実験（first replicate）についてのメチル化シグナルおよび体細胞変異のグラフ表示である。

【００２９】

図２０Ｂは、臨床的用量設定の第２の反復実験についてのメチル化シグナルおよび体細胞変異のグラフ表示である。

【００３０】

【図２１】図２１は、ゲノムのみの方法およびメチル化方法を使用して計算した、第１の反復実験および第２の反復実験についてのｃｔＤＮＡレベルの変化を示す表である。

【００３１】

【図２２】図２２は、臨床試料に関するエピゲノムｃＴＦ対ゲノムｃＴＦのグラフ表示である（１つの試料について１つの点）。

【００３２】

【図２３】図２３は、乳がん、結腸直腸がん、肺がん、および他がんの群についての初期および末期がん患者におけるエピＭＡＦ分布のグラフ表示である。

【００３３】

【図２４】図２４は、３つの分配に含まれるメチル化シトシンの数を示す確率分布を示すグラフ表示である。

【００３４】

【図２５】図２５Ａは、試薬の第１のセットを使用してＭＢＤで処理した試料の第１の群および試薬の第２のセットを使用してＭＢＤで処理した試料の第２の群のための第１の分類領域に関するメトリクスの変化を示すグラフ表示を含む。

【００３５】

図２５Ｂは、試薬の第１のセットを使用してＭＢＤで処理した試料の第１の群および試薬の第２のセットを使用してＭＢＤで処理した試料の第２の群の第１の分類領域に関する第２の分類領域に関するメトリクスの変化を示すグラフ表示を含む。

【発明を実施するための形態】

【００３６】

概要

１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む。方法は、コンピューティングシステムにより、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含む。方法は、コンピューティングシステムにより、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参

照ゲノムの追加のゲノム領域に対応する、ステップも含む。方法は、コンピューティングシステムにより、複数の分類領域の個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含む。方法は、コンピューティングデバイスにより、訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についてのメトリクスを含む訓練用データを生成するステップも含む。方法は、コンピューティングシステムにより、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて対象にがんが存在する指標（ｉｎｄｉｃａｔｉｏｎ）を決定するためのモデルを生成するステップであって、モデルが、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分が、互いに異なる、ステップも含む。

【００３７】

１つまたは複数の態様では、方法は、コンピューティングシステムにより、複数の対象に含まれない追加の対象由来のテスト用配列データを得るステップであって、テスト用配列データが、追加の対象の試料に由来するテスト用配列決定リードを含み、個々のテスト用配列決定リードが、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々のテスト用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる少なくとも閾値量のメチル化シトシンを有する分子に対応する、ステップと、モデルおよび追加の配列データを使用して、追加の対象にがんが存在する指標を決定するステップとを含む。

【００３８】

１つまたは複数の態様では、方法は、コンピューティングシステムにより、テスト用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応するテスト用配列決定リードに由来する第１の定量的尺度を決定するステップと、コンピューティングシステムにより、テスト用配列決定リードを解析して、複数の対照領域の個々の対照領域に対応するテスト用配列決定リードに由来する第２の定量的尺度を決定するステップと、コンピューティングシステムにより、個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップと、コンピューティングシステムにより、個々の分類領域についてのメトリクスを含む入力ベクトルを生成するステップとを含み、ここで、モデルで入力ベクトルを使用して、追加の対象にがんが存在する指標を決定する。

【００３９】

１つまたは複数の態様では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の分類アルゴリズムを含み、がんが存在する指標は、追加の対象にがんが存在する確率に対応する。

【００４０】

１つまたは複数の態様では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の回帰アルゴリズムを含み、指標（ｉｎｄｉｃａｔｏｒ）は、追加の試料の腫瘍画分の推定値に対応する。

【００４１】

１つまたは複数の態様では、訓練用配列決定リードは、訓練用配列データの第１の一部分を含み、追加の訓練用配列決定リードは、訓練用配列データの第２の一部分を含み、ここで、追加の訓練用配列決定リードは、訓練用配列決定リードとは異なり、方法は、コンピューティングシステムにより、訓練用配列データの第１の一部分または訓練用配列データの第２の一部分の少なくとも一方を解析して、複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定するステップと、コンピューティングシステムにより、個々の試料について、個々の試料に由来する個々の頻度の中で最大値を有する個々の頻度に対応する最高頻度を有する、複数のバリアントのうちの１つのバリアントを決定するステップと、コンピューティングシステムにより、個々の試料についての腫瘍画分の個々の尺度を、個々の試料に由来する個々の頻度の最大値に基づいて決定するステップとを含

む。
【００４２】
１つまたは複数の態様では、訓練用データは、複数の試料の個々の試料についての腫瘍画分の個々の尺度を含み、モデルは、複数の試料の個々の試料についての腫瘍画分の個々の尺度に基づいて生成される。
【００４３】
１つまたは複数の態様では、個々の分類領域についてのメトリクスは、スケーリング係数および誤差補正係数に基づいて決定される。
【００４４】
１つまたは複数の態様では、複数の分類領域は、がんが存在する対象から得られた細胞に由来する核酸のゲノム領域のメチル化率が、がんが存在しない対象から得られた細胞に由来する核酸のゲノム領域のメチル化率と異なるゲノム領域に個々に対応する。
【００４５】
１つまたは複数の態様では、複数の分類領域は、第１のがんタイプに対する第１の複数の分類領域に対応し、第１の複数の分類領域とは異なる第２の複数の分類領域に基づいて、第２のがんタイプに対するモデルを生成することができる。
【００４６】
１つまたは複数の態様では、複数の試料および追加の試料は、無細胞核酸を含む。
【００４７】
１つまたは複数の態様では、方法は、コンピューティングシステムにより、訓練用データを使用して訓練プロセスを実施して、モデルを生成するステップであって、訓練プロセスが、コンピューティングシステムにより、訓練用データ内に含まれる個々の試料の１つまたは複数の追加の重みを、閾値信頼水準の範囲内にある個々の試料についてのがんの指標に基づいて決定することを含む、ステップを含む。
【００４８】
１つまたは複数の態様では、個々の試料についてのがんの指標は、閾値信頼水準の範囲外であり、方法は、コンピューティングシステムにより、訓練プロセス中に個々の試料の重みに対してペナルティを適用するステップを含む。
【００４９】
１つまたは複数の態様では、方法は、コンピューティングシステムにより、１つまたは複数の機械学習アルゴリズムを使用して、モデルに対して、訓練用データの一部分を使用して訓練プロセスの１回または複数の第１の反復を実施するステップと、コンピューティングシステムにより、訓練プロセスの１回または複数の第１の反復に基づいて、モデルについての第１の出力データを生成するステップであって、第１の出力データが、複数の対象のうちの第１の個々の対象にがんが存在する１つまたは複数の第１の追加の指標に対応し、第１の個々の対象が、訓練用データの一部分に対応する、ステップとを含む。
【００５０】
１つまたは複数の態様では、方法は、コンピューティングシステムにより、第１の出力データと訓練用データを組み合わせて、追加の訓練用データを産生するステップと、コンピューティングシステムにより、モデルに対して、追加の訓練用データの一部分を使用して訓練プロセスの１回または複数の第２の反復を実施するステップと、コンピューティングシステムにより、訓練プロセスの１回または複数の第２の反復に基づいてモデルについての第２の出力データを生成するステップであって、第２の出力データが、複数の対象のうちの第２の個々の対象にがんが存在する１つまたは複数の第２の追加の指標を示し、第２の個々の対象が、追加の訓練用データの一部分に対応する、ステップとを含む。
【００５１】
１つまたは複数の態様では、複数の分類領域の個々の分類領域に対する重みは、第１の出力データおよび第２の出力データに基づいて決定される。
【００５２】
１つまたは複数の態様では、方法は、コンピューティングシステムにより、訓練プロセ

スの１回または複数の反復の間に決定された、いくつかの、がんが存在する指標が、少なくとも、訓練用データに含められた１つまたは複数の試料についての閾値であることを決定するステップと、コンピューティングシステムにより、モデルの１つまたは複数の重みに対する修正が、修正されないかまたは最小量で修正されることを決定するステップとを含む。

【００５３】

１つまたは複数の態様では、方法は、コンピューティングシステムにより、訓練プロセスの１回または複数の反復の間に決定された、がんが存在する追加の指標の数が、訓練用データに含められた１つまたは複数の追加の試料についての閾値未満であることを決定するステップと、コンピューティングシステムにより、モデルの１つまたは複数の追加の重みに対する修正が、最小量よりも多く修正されることを決定するステップとを含む。

【００５４】

１つまたは複数の態様では、方法は、対象の血液または組織の少なくとも一方に由来する複数の核酸を、ある１つの量のメチル結合ドメイン（ＭＢＤ）タンパク質を含む溶液と組み合わせて、核酸－ＭＢＤタンパク質溶液を産生するステップと、塩類溶液を用いた核酸－ＭＢＤタンパク質溶液の複数の洗浄を実施して、いくつかの核酸画分を産生するステップであって、個々の核酸画分が、少なくとも閾値シトシン－グアニン含量を有する複数の核酸の領域内に閾値数のメチル化シトシンを有する、ステップとを含む。

【００５５】

１つまたは複数の態様では、複数の洗浄のうちのある洗浄は、ある濃度の塩化ナトリウム（ＮａＣｌ）を有する溶液を用いて実施し、ＭＢＤタンパク質に対する一連の結合強度を有するいくつかの核酸画分のうちの１つの核酸画分を産生する。

【００５６】

１つまたは複数の態様では、方法は、第１の核酸画分が核酸の複数の分配のうちの第１の分配と関連することを決定するステップであって、第１の分配が、ＭＢＤタンパク質に対する結合強度の第１の範囲に対応する、ステップと、第１の分子バーコードを第１の核酸画分の核酸に結合させるステップであって、第１の分子バーコードが、第１の分配と関連する分子バーコードの第１のセット内に含まれる、ステップと、第２の核酸画分が、核酸の複数の分配のうちの第２の分配と関連することを決定するステップであって、第２の分配が、ＭＢＤタンパク質に対する結合強度の第１の範囲とは異なる、ＭＢＤタンパク質に対する結合強度の第２の範囲に対応する、ステップと、第２の分子バーコードを第２の核酸画分の核酸に結合させるステップであって、第２の分子バーコードが、第２の分配と関連する分子バーコードの第２のセット内に含まれる、ステップとを含む。

【００５７】

１つまたは複数の態様では、方法は、いくつかの核酸画分の少なくとも一部分を、ある１つの量の、１つまたは複数の非メチル化シトシンを有する分子を切断する１つまたは複数のメチル化感受性制限酵素と組み合わせて、配列決定リードを産生するために使用される複数の試料の少なくとも一部分を産生するステップを含む。

【００５８】

１つまたは複数の態様では、方法は、いくつかの核酸画分の少なくとも一部分を、ある１つの量の、１つまたは複数のメチル化シトシンを有する分子を切断する１つまたは複数のメチル化依存性制限酵素と組み合わせて、配列決定リードを産生するために使用される複数の試料の少なくとも一部分を産生するステップを含む。

【００５９】

１つまたは複数の態様では、試料の腫瘍画分を決定するためのモデルの検出限界は、０．０５％以下である。

【００６０】

１つまたは複数の態様では、コンピューティングシステムは、１つまたは複数のハードウェアプロセッサーと、１つまたは複数のハードウェアプロセッサーによって実行されるときに、１つまたは複数のハードウェアプロセッサーに、複数の対象の複数の試料に由来

する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む操作を実施させるコンピュータ可読命令を含む１つまたは複数の非一時的コンピュータ可読ストレージ媒体とを含む。操作は、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含む。操作は、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含む。操作は、複数の分類領域の個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含む。操作は、訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についてのメトリクスを含む訓練用データを生成するステップも含む。操作は、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて、対象にがんが存在する指標を決定するためのモデルを生成するステップであって、モデルが、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分が、互いに異なる、ステップも含む。

【００６１】

１つまたは複数の態様では、コンピューティングシステムは、複数の対象に含まれない追加の対象由来のテスト用配列データを得ることができる。テスト用配列データは、追加の対象の試料に由来するテスト用配列決定リードを含み得、個々のテスト用配列決定リードは、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々のテスト用配列決定リードは、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる少なくとも閾値量のメチル化シトシンを有する分子に対応する。操作は、モデルおよび追加の配列データを使用して、追加の対象にがんが存在する指標を決定するステップも含み得る。

【００６２】

１つまたは複数の態様では、コンピューティングシステムは、テスト用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応するテスト用配列決定リードに由来する第１の定量的尺度を決定することができる。コンピューティングシステムは、テスト用配列決定リードを解析して、複数の対照領域の個々の対照領域に対応するテスト用配列決定リードに由来する第２の定量的尺度を決定することもできる。コンピューティングシステムは、個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定することもできる。コンピューティングシステムは、個々の分類領域についてのメトリクスを含む入力ベクトルを生成することもでき、ここで、モデルで入力ベクトルを使用して、追加の対象にがんが存在する指標を決定する。

【００６３】

コンピューティングシステムの１つまたは複数の態様では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の分類アルゴリズムを含み、がんが存在する指標は、追加の対象にがんが存在する確率に対応する。

【００６４】

コンピューティングシステムの１つまたは複数の態様では、１つまたは複数の機械学習

アルゴリズムは、１つまたは複数の回帰アルゴリズムを含み、指標は、追加の試料の腫瘍画分の推定値に対応する。

【００６５】

コンピューティングシステムの１つまたは複数の態様では、訓練用配列決定リードは、訓練用配列データの第１の一部分を含み、追加の訓練用配列決定リードは、訓練用配列データの第２の一部分を含み、ここで、追加の訓練用配列決定リードは、訓練用配列決定リードとは異なり、コンピューティングシステムは、訓練用配列データの第１の一部分または訓練用配列データの第２の一部分の少なくとも一方を解析して、複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定することができる。コンピューティングシステムは、個々の試料について、個々の試料に由来する個々の頻度の中で最大値を有する個々の頻度に対応する最高頻度を有する、複数のバリアントのうちの１つのバリアントを決定することもできる。コンピューティングシステムは、個々の試料についての腫瘍画分の個々の尺度を、個々の試料に由来する個々の頻度の最大値に基づいて決定することもできる。

【００６６】

コンピューティングシステムの１つまたは複数の態様では、訓練用データは、複数の試料の個々の試料についての腫瘍画分の個々の尺度を含み、モデルは、複数の試料の個々の試料についての腫瘍画分の個々の尺度に基づいて生成される。

【００６７】

コンピューティングシステムの１つまたは複数の態様では、個々の分類領域についてのメトリクスは、スケーリング係数および誤差補正係数に基づいて決定される。

【００６８】

コンピューティングシステムの１つまたは複数の態様では、複数の分類領域は、がんが存在する対象から得られた細胞に由来する核酸のゲノム領域のメチル化率が、がんが存在しない対象から得られた細胞に由来する核酸のゲノム領域のメチル化率と異なるゲノム領域に個々に対応する。

【００６９】

コンピューティングシステムの１つまたは複数の態様では、複数の分類領域は、第１のがんタイプに対する第１の複数の分類領域に対応し、第１の複数の分類領域とは異なる第２の複数の分類領域に基づいて、第２のがんタイプに対するモデルを生成することができる。

【００７０】

１つまたは複数の態様では、コンピューティングシステムは、訓練用データを使用して訓練プロセスを実施して、モデルを生成することができ、ここで、訓練プロセスは、訓練用データ内に含まれる個々の試料の１つまたは複数の追加の重みを、閾値信頼水準の範囲内にある個々の試料についてのがんの指標に基づいて決定するステップを含む。

【００７１】

コンピューティングシステムの１つまたは複数の態様では、個々の試料についてのがんの指標は、閾値信頼水準の範囲外であり、方法は、コンピューティングシステムにより、訓練プロセス中に個々の試料の重みに対してペナルティを適用するステップを含む。

【００７２】

１つまたは複数の態様では、コンピューティングシステムは、１つまたは複数の機械学習アルゴリズムを使用して、モデルに対して、訓練用データの一部分を使用して訓練プロセスの１回または複数の第１の反復を実施することができる。コンピューティングシステムは、モデルについての第１の出力データを訓練プロセスの１回または複数の第１の反復に基づいて生成することもでき、第１の出力データは、複数の対象のうちの第１の個々の対象にがんが存在する１つまたは複数の第１の追加の指標に対応し、第１の個々の対象は、訓練用データの一部分に対応する。

【００７３】

１つまたは複数の態様では、コンピューティングシステムは、第１の出力データと訓練

用データを組み合わせて、追加の訓練用データを産生し、モデルに対して追加の訓練用データの一部分を使用して訓練プロセスの１回または複数の第２の反復を実施し、モデルについての第２の出力データを訓練プロセスの１回または複数の第２の反復に基づいて生成することができ、第２の出力データは、複数の対象のうちの第２の個々の対象にがんが存在する１つまたは複数の第２の追加の指標を示し、第２の個々の対象は、追加の訓練用データの一部分に対応する。

【００７４】

コンピューティングシステムの１つまたは複数の態様では、複数の分類領域の個々の分類領域に対する重みは、第１の出力データおよび第２の出力データに基づいて決定される。

【００７５】

１つまたは複数の態様では、コンピューティングシステムは、訓練プロセスの１回または複数の反復の間に決定された、いくつかの、がんが存在する指標が、少なくとも、訓練用データに含められた１つまたは複数の試料についての閾値であることを決定することができる。コンピューティングシステムは、モデルの１つまたは複数の重みに対する修正が、修正されないかまたは最小量で修正されることを決定することもできる。

【００７６】

１つまたは複数の態様では、コンピューティングシステムは、訓練プロセスの１回または複数の反復の間に決定された、がんが存在する追加の指標の数が、訓練用データに含められた１つまたは複数の追加の試料についての閾値未満であることを決定すること、およびモデルの１つまたは複数の追加の重みに対する修正が最小量よりも多く修正されることを決定することができる。

【００７７】

コンピューティングシステムの１つまたは複数の態様では、試料の腫瘍画分を決定するためのモデルの検出限界は０．０５％以下である。

【００７８】

１つまたは複数の態様では、１つまたは複数のコンピュータ可読ストレージ媒体は、コンピューティングシステムの１つまたは複数のプロセッサーによって実行されるときに、コンピューティングシステムに、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む操作を実施させるコンピュータ可読命令を含む。操作は、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含む。操作は、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含む。操作は、複数の分類領域の個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含む。操作は、訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についてのメトリクスを含む訓練用データを生成するステップも含む。操作は、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて、対象にがんが存在する指標を決定するためのモデルを生成するステ

ップであって、モデルが、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分が、互いに異なる、ステップも含む。

【００７９】

１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、対象から得られた試料に由来する配列決定リードを得るステップであって、個々の配列決定リードが、試料中に含まれる核酸の断片に対応するヌクレオチド配列を含み、かつ、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む。方法は、コンピューティングシステムにより、複数の分類領域の個々の分類領域に対応する配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、配列決定リードを解析して、複数の対照領域に対応する配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、複数のメトリクスのうちの個々のメトリクスが複数の分類領域の個々の分類領域に対応する、その複数のメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含み得る。方法は、コンピューティングシステムにより、対象にがんが存在する指標を、複数のメトリクスの少なくとも一部分に基づいて決定するステップも含み得る。

【００８０】

１つまたは複数の態様では、方法は、また、コンピューティングシステムにより、配列決定データを使用して、差次的にメチル化された領域についての配列表示の分布を決定するステップと、コンピューティングシステムにより、分布に含まれる少なくとも閾値量の配列表示が、差次的にメチル化された領域の部分領域と重複することを決定するステップとを含み得る。方法は、コンピューティングシステムにより、差次的にメチル化された領域の部分領域が、複数の分類領域のうちの１つの分類領域であることを決定するステップも含み得る。

【００８１】

１つまたは複数の態様では、方法は、また、コンピューティングシステムにより、複数のメトリクスの値の順序を決定するステップと、コンピューティングシステムにより、複数の分類領域の中からの分類領域のサブセットを順序に基づいて決定するステップとを含み得、ここで、分類領域のサブセットに対応する複数のメトリクスの一部分を使用して、対象にがんが存在する指標を決定する。

【００８２】

１つまたは複数の態様では、対象にがんが存在する指標は、対象にがんが存在する最初の指標であり、方法は、コンピューティングシステムにより、対象にがんが存在する最初の指標にスケーリング係数を適用して、対象にがんが存在する改変された指標を決定するステップも含み得る。

【００８３】

１つまたは複数の態様では、対象にがんが存在する指標は、腫瘍画分に対応する。

【００８４】

１つまたは複数の態様では、試料は、がんの処置の開始前または開始時の少なくとも一方において収集された第１の試料であり、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、対象から得られた第２の試料に由来する追加の配列決定リードを得るステップであって、個々の追加の配列決

定リードが、第２の試料中に含まれる核酸の断片に対応する追加のヌクレオチド配列を含み、かつ、少なくとも閾値シトシン－グアニン含量を有する追加のヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する追加の分子に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、複数の分類領域の個々の分類領域に対応する追加の配列決定リードに由来する追加の第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、追加の配列決定リードを解析して、複数の対照領域に対応する追加の配列決定リードに由来する追加の第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、複数の追加のメトリクスのうちの個々の追加のメトリクスが複数の分類領域の個々の分類領域に対応する、その複数の追加のメトリクスを、個々の分類領域についての追加の第１の定量的尺度および複数の対照領域についての追加の第２の定量的尺度に基づいて決定するステップも含み得る。方法は、コンピューティングシステムにより、対象にがんが存在する追加の指標を、複数の追加のメトリクスの少なくとも一部分に基づいて決定するステップも含み得る。

【００８５】

１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、対象からテスト用配列データを得るステップであって、テスト用配列データが、対象の試料に由来するテスト用配列決定リードを含み、個々のテスト用配列決定リードが、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々のテスト用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む。方法は、コンピューティングシステムにより、テスト用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応するテスト用配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、テスト用配列決定リードを解析して、複数の対照領域の個々の対照領域に対応するテスト用配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含み得る。方法は、コンピューティングシステムにより、個々の分類領域についてのメトリクスを含む入力ベクトルを生成するステップも含み得る。方法は、コンピューティングシステムにより、対象にがんが存在する指標を、１つまたは複数の機械学習技術をインプリメントするモデルに入力ベクトルを提供して、対象にがんが存在する指標を生成することによって決定するステップであって、モデルが、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分が互いに異なる、ステップも含む。

【００８６】

１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、複数の訓練用対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々

の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する追加の第１の定量的尺度を決定するステップも含み得る。方法は、コンピューティングシステムにより、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する追加の第２の定量的尺度を決定するステップも含み得る。方法は、コンピューティングシステムにより、複数の分類領域の個々の分類領域についての追加のメトリクスを、個々の分類領域についての追加の第１の定量的尺度および複数の対照領域についての追加の第２の定量的尺度に基づいて決定するステップも含み得る。方法は、コンピューティングデバイスにより、複数の訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についての追加のメトリクスを含む訓練用データを生成するステップも含み得る。方法は、コンピューティングシステムにより、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、対象にがんが存在する指標を、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて決定するためのモデルを生成するステップも含み得る。

【００８７】

１つまたは複数の態様では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の分類アルゴリズムを含み、がんが存在する指標は、対象にがんが存在する確率に対応する。

【００８８】

１つまたは複数の態様では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の回帰アルゴリズムを含み、指標は、試料の腫瘍画分の推定値に対応する。

【００８９】

１つまたは複数の態様では、対象の試料および複数の訓練用対象の複数の試料は、無細胞核酸を含む。

【００９０】

１つまたは複数の態様では、閾値量の配列表示は、分布に含まれる配列表示の少なくとも約７０％である。

【００９１】

１つまたは複数の態様では、第２の試料は、対象にがんの処置が投与されてから少なくとも１週間後に得られる。

【００９２】

１つまたは複数の態様では、方法は、コンピューティングシステムにより、対象にがんが存在する指標を対象にがんが存在する追加の指標と関連付けて解析して、対象についての処置に対する応答を決定するステップも含み得る。

【００９３】

１つまたは複数の態様では、訓練用配列決定リードは、訓練用配列データの第１の一部分を含み、追加の訓練用配列決定リードは、訓練用配列データの第２の一部分を含み、追加の訓練用配列決定リードは、訓練用配列決定リードとは異なり、方法は、また、コンピューティングシステムにより、訓練用配列データの第１の一部分または訓練用配列データの第２の一部分の少なくとも一方を解析して、複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定するステップと、コンピューティングシステムにより、個々の試料について、個々の試料に由来する個々の頻度の中で最大値を有する個々の頻度に対応する最高頻度を有する、複数のバリアントのうちの１つのバリアントを決定するステップとを含み得る。方法は、コンピューティングシステムにより、個々の試料についての腫瘍画分の個々の尺度を、個々の試料に由来する個々の頻度の最大値に基づいて決定するステップも含み得る。

【００９４】
１つまたは複数の態様では、訓練用データは、複数の試料の個々の試料についての腫瘍画分の個々の尺度を含み、モデルは、複数の試料の個々の試料についての腫瘍画分の個々の尺度に基づいて生成される。
【００９５】
１つまたは複数の態様では、コンピューティングシステムは、プロセッサーと、プロセッサーによって実行されるときに、対象から得られた試料に由来する配列決定リードを得るように、コンピューティングシステムを構成する命令が記憶されたメモリーとを含み、個々の配列決定リードは、試料中に含まれる核酸の断片に対応するヌクレオチド配列を含み、かつ、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる、閾値量のメチル化シトシンを有する分子に対応する。コンピューティングシステムは、複数の分類領域の個々の分類領域に対応する配列決定リードに由来する第１の定量的尺度を決定することもでき、複数の分類領域の個々の分類領域の少なくとも一部分は、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する。コンピューティングシステムは、配列決定リードを解析して、複数の対照領域に対応する配列決定リードに由来する第２の定量的尺度を決定することもでき、複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する。コンピューティングシステムは、複数のメトリクスのうちの個々のメトリクスが複数の分類領域の個々の分類領域に対応する、その複数のメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定することもできる。コンピューティングシステムは、対象にがんが存在する指標を、複数のメトリクスの少なくとも一部分に基づいて決定することもできる。
【００９６】
コンピューティングシステムの１つまたは複数の態様では、閾値量の配列表示は、分布に含まれる配列表示の少なくとも約７０％である。
【００９７】
１つまたは複数の態様では、コンピューティングシステムは、配列決定データを使用して、差次的にメチル化された領域についての配列表示の分布を決定することができる。コンピューティングシステムは、分布に含まれる少なくとも閾値量の配列表示が、差次的にメチル化された領域の部分領域と重複することを決定することもできる。コンピューティングシステムは、コンピューティングシステムにより、差次的にメチル化された領域の部分領域が、複数の分類領域のうちの１つの分類領域であることを決定することもできる。
【００９８】
１つまたは複数の態様では、コンピューティングシステムは、複数のメトリクスの値の順序を決定すること、および順序に基づいて複数の分類領域の中からの分類領域のサブセットを決定することができ、ここで、分類領域のサブセットに対応する複数のメトリクスの一部分を使用して、対象にがんが存在する指標を決定する。
【００９９】
コンピューティングシステムの１つまたは複数の態様では、対象にがんが存在する指標は、対象にがんが存在する最初の指標であり、コンピューティングシステムは、対象にがんが存在する最初の指標にスケーリング係数を適用して、対象にがんが存在する改変された指標を決定することができる。
【０１００】
コンピューティングシステムの１つまたは複数の態様では、対象にがんが存在する指標は、腫瘍画分に対応する。
【０１０１】
コンピューティングシステムの１つまたは複数の態様では、試料は、がんの処置の開始

前または開始時の少なくとも一方において収集された第１の試料であり、コンピューティングシステムは、対象から得られた第２の試料に由来する追加の配列決定リードを得ることができ、個々の追加の配列決定リードは、第２の試料中に含まれる核酸の断片に対応する追加のヌクレオチド配列を含み、かつ、少なくとも閾値シトシン－グアニン含量を有する追加のヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する追加の分子に対応する。コンピューティングシステムは、複数の分類領域の個々の分類領域に対応する追加の配列決定リードに由来する追加の第１の定量的尺度を決定することもでき、複数の分類領域の個々の分類領域の少なくとも一部分は、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する。コンピューティングシステムは、追加の配列決定リードを解析して、複数の対照領域に対応する追加の配列決定リードに由来する追加の第２の定量的尺度を決定することもでき、複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する。コンピューティングシステムは、複数の追加のメトリクスのうちの個々の追加のメトリクスが複数の分類領域の個々の分類領域に対応する、その複数の追加のメトリクスを、個々の分類領域についての追加の第１の定量的尺度および複数の対照領域についての追加の第２の定量的尺度に基づいて決定することもできる。コンピューティングシステムは、対象にがんが存在する追加の指標を、複数の追加のメトリクスの少なくとも一部分に基づいて決定することもできる。

【０１０２】

コンピューティングシステムの１つまたは複数の態様では、第２の試料は、対象にがんの処置が投与されてから少なくとも１週間後に得られる。

【０１０３】

１つまたは複数の態様では、コンピューティングシステムは、対象にがんが存在する指標を対象にがんが存在する追加の指標と関連付けて解析して、対象についての処置に対する応答を決定することができる。

【０１０４】

１つまたは複数の態様では、コンピューティングシステムは、プロセッサーと、プロセッサーによって実行されるときに、対象からテスト用配列データを得るようにシステムを構成する命令が記憶されたメモリーとを含み、テスト用配列データは、対象の試料に由来するテスト用配列決定リード、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含む個々のテスト用配列決定リードおよび少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する個々のテスト用配列決定リードを含む。コンピューティングシステムは、テスト用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応するテスト用配列決定リードに由来する第１の定量的尺度を決定することもでき、複数の分類領域の個々の分類領域の少なくとも一部分は、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する。コンピューティングシステムは、テスト用配列決定リードを解析して、複数の対照領域の個々の対照領域に対応するテスト用配列決定リードに由来する第２の定量的尺度を決定することもでき、複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する。コンピューティングシステムは、個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定することもできる。コンピューティングシステムは、個々の分類領域についてのメトリクスを含む入力ベクトルを生成することもできる。コンピューティングシステムは、対象にがんが存在する指標を、１つまたは複数の機械学習技術をインプリメントするモデルに入力ベクトルを提供して、対象にがんが存在する指標を生成

することによって決定することもでき、モデルは、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分は互いに異なる。

【０１０５】

１つまたは複数の態様では、コンピューティングシステムは、複数の訓練用対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得ることができ、個々の訓練用配列決定リードは、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用シーケンシングリードは、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する。コンピューティングシステムは、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する追加の第１の定量的尺度を決定することもできる。コンピューティングシステムは、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する追加の第２の定量的尺度を決定することもできる。コンピューティングシステムは、複数の分類領域の個々の分類領域についての追加のメトリクスを、個々の分類領域についての追加の第１の定量的尺度および複数の対照領域についての追加の第２の定量的尺度に基づいて決定することもできる。コンピューティングシステムは、複数の訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についての追加のメトリクスを含む訓練用データを生成することもできる。コンピューティングシステムは、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、対象にがんが存在する指標を、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて決定するためのモデルを生成することもできる。

【０１０６】

コンピューティングシステムの１つまたは複数の態様では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の分類アルゴリズムを含み、がんが存在する指標は、対象にがんが存在する確率に対応する。

【０１０７】

コンピューティングシステムの１つまたは複数の態様では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の回帰アルゴリズムを含み、指標は、試料の腫瘍画分の推定値に対応する。

【０１０８】

コンピューティングシステムの１つまたは複数の態様では、訓練用配列リードは、訓練用配列データの第１の一部分を含み、追加の訓練用配列決定リードは、訓練用配列データの第２の一部分を含み、追加の訓練用配列決定リードは、訓練用配列決定リードとは異なり、コンピューティングシステムは、訓練用配列データの第１の一部分または訓練用配列データの第２の一部分の少なくとも一方を解析して、複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定することができる。コンピューティングシステムは、個々の試料について、個々の試料に由来する個々の頻度の中で最大値を有する個々の頻度に対応する最高頻度を有する、複数のバリアントのうちの１つのバリアントを決定することもできる。コンピューティングシステムは、個々の試料についての腫瘍画分の個々の尺度を、個々の試料に由来する個々の頻度の最大値に基づいて決定することもできる。

【０１０９】

コンピューティング装置の１つまたは複数の態様では、訓練用データは、複数の試料の個々の試料についての腫瘍画分の個々の尺度を含み、モデルは、複数の試料の個々の試料についての腫瘍画分の個々の尺度に基づいて生成される。

【０１１０】

１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、複数の対象由来の配列決定データを得るステップであって、配列決定データが、複数の対象の複数の試料に由来する配列決定リードを含み、個々の配列決定リードが、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の配列決定リードが、少なくとも閾値シトシン－グアニ

ン含量を有するヌクレオチド配列のプロモーター領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む。方法は、コンピューティングシステムにより、配列決定リードを解析して、プロモーター領域に対応する配列決定リードに由来する第１の定量的尺度を決定するステップも含み得る。方法は、コンピューティングシステムにより、配列決定リードを解析して、複数の対照領域の個々の対照領域に対応する配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、プロモーター領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含み得る。方法は、コンピューティングシステムにより、プロモーター領域のメチル化状態の指標を、少なくとも閾値を有するメトリクスに基づいて生成するステップも含み得る。

【０１１１】

１つまたは複数の態様では、コンピューティングシステムは、１つまたは複数のハードウェアプロセッサーと、１つまたは複数のハードウェアプロセッサーによって実行されるときに、１つまたは複数のハードウェアプロセッサーに、複数の対象由来の配列決定データを得るステップであって、配列決定データが、複数の対象の複数の試料に由来する配列決定リードを含み、個々の配列決定リードが、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列のプロモーター領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む操作を実施させるコンピュータ可読命令を含む１つまたは複数の非一時的コンピュータ可読ストレージ媒体とを含み得る。システムは、配列決定リードを解析して、プロモーター領域に対応する配列決定リードに由来する第１の定量的尺度を決定することもできる。システムは、配列決定リードを解析して、複数の対照領域の個々の対照領域に対応する配列決定リードに由来する第２の定量的尺度を決定することもでき、複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する。システムは、配列決定リードを解析して、複数の対照領域の個々の対照領域に対応する配列決定リードに由来する第２の定量的尺度を決定することもでき、複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する。システムは、プロモーター領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定することもできる。システムは、プロモーター領域のメチル化状態の指標を、少なくとも閾値を有するメトリクスに基づいて生成することもできる。

（発明を実施するための形態）

【０１１２】

定義

本開示をより容易に理解するために、最初にある特定の用語を以下に定義する。以下の用語および他の用語についてのさらなる定義が本明細書を通じて記載される場合がある。下記の用語の定義が参照により組み込まれる出願または特許における定義と矛盾する場合は、本出願に記載される定義を使用して用語の意味を理解すべきである。

【０１１３】

本明細書および添付の特許請求の範囲で使用される場合、単数形の「１つの（ａ）」、「１つの（ａｎ）」、および「その（ｔｈｅ）」は、文脈により別途明確に示されない限り、複数形の参照物を含む。したがって、例えば、「１つの（ａ）方法」への言及は、本

明細書に記載されるタイプのおよび／または本開示を読めば当業者には明らかになる１つまたは複数の方法および／またはステップなどを含む。

【０１１４】

本明細書において使用される専門用語は、単に特定のインプリメンテーションを説明するためのものであり、制限することを意図するものではないことも理解されたい。さらに、別段の定義のない限り、本明細書において使用される全ての科学技術用語は、本開示が関係する技術分野の当業者に一般に理解されるものと同じ意味を有する。方法、コンピュータ可読媒体、およびシステムの記載および特許請求に関しては、以下の専門用語およびその文法上の変形が下記の定義に従って使用される。

【０１１５】

約：本明細書で使用される場合、「約」または「およそ」とは、目的の１つまたは複数の値または要素に適用される場合、指定された参照値または要素と同様の値または要素を指す。ある特定のインプリメンテーションでは、「約」または「およそ」という用語は、別段の指定のない限り、または文脈からそうでないことが明らかでない限り、指定された参照値または要素の、いずれかの方向に（より大きいまたはより小さい）、２５％以内、２０％以内、１９％以内、１８％以内、１７％以内、１６％以内、１５％以内、１４％以内、１３％以内、１２％以内、１１％以内、１０％以内、９％以内、８％以内、７％以内、６％以内、５％以内、４％以内、３％以内、２％以内、１％以内、またはそれ未満の％以内に入る一連の値または要素を指す（そのような数が可能性のある値または要素の１００％を超える場合を除く）。

【０１１６】

投与する：本明細書で使用される場合、治療剤（例えば、免疫学的治療剤）を対象に「投与する」または「投与すること」とは、対象に組成物を与える、適用するまたは接触させることを意味する。投与は、例えば、局所、経口、皮下、筋肉内、腹腔内、静脈内、髄腔内および皮内を含む、いくつかの経路のいずれかによって実現されてもよい。

【０１１７】

アダプター：本明細書で使用される場合、「アダプター」とは、少なくとも部分的に二本鎖である、所与の試料核酸分子の一方または両方の末端を連結するために使用され得る短い核酸（例えば、長さが約５００ヌクレオチド未満、約１００ヌクレオチド未満、または約５０ヌクレオチド未満）を指す。アダプターは、両末端でアダプターに挟まれた核酸分子の増幅を可能にするための核酸プライマー結合部位、および／または、様々な次世代配列決定（ＮＧＳ）アプリケーションなどの配列決定アプリケーションのためのプライマー結合部位を含む配列決定プライマー結合部位を含み得る。アダプターは、フローセル支持体などに結合したオリゴヌクレオチドなどの捕捉プローブの結合部位も含み得る。アダプターは、本明細書に記載される核酸タグも含み得る。核酸タグは、増幅プライマーおよび配列決定プライマー結合部位に対して、核酸タグが所与の核酸分子のアンプリコンおよび配列リードに含まれるように位置付けることができる。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を連結することができる。一部のインプリメンテーションでは、核酸タグが異なることを除いて、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。一部のインプリメンテーションでは、アダプターは、本明細書に記載されるように、１つまたは複数の相補的なヌクレオチドを有する同様に平滑末端化されたまたはテールが付された核酸分子への接合のために一方の末端が平滑末端化されたまたはテールが付されたＹ形状アダプターである。さらに他のインプリメンテーションの例では、アダプターは、解析される核酸分子への接合のために平滑末端またはテールが付された末端を含むベル形状アダプターである。アダプターの他の例としては、Ｔ尾部を有するアダプターおよびＣ尾部を有するアダプターが挙げられる。

【０１１８】

アライメント：本明細書で使用される場合、「アライメント」または「整列する」とは、少なくとも２つの配列表示が少なくとも閾値量の相同性を有するかどうかを決定することを指す。１つまたは複数の例では、相同性の閾値量は、少なくとも約９０％、少なくと

も約９１％、少なくとも約９２％、少なくとも約９３％、少なくとも約９４％、少なくとも約９５％、少なくとも約９６％、少なくとも約９７％、少なくとも約９８％、少なくとも約９９％、少なくとも約９９．５％、または少なくとも約９９．９％であり得る。２つの配列表示が少なくとも閾値量の相同性を有する状況では、２つの配列表示は、「整列された」と称され得る。

【０１１９】

増幅させる：本明細書で使用される場合、核酸の文脈における「増幅させる」または「増幅」とは、少量のポリヌクレオチド（例えば、単一のポリヌクレオチド分子）から出発した、ポリヌクレオチド、またはポリヌクレオチドの一部分の複数のコピーの産生を指し、ここで、増幅産物またはアンプリコンが一般に検出可能である。ポリヌクレオチドの増幅は、様々な化学的および酵素的プロセスを包含する。

【０１２０】

バーコード：本明細書で使用される場合、核酸の文脈における「バーコード」または「分子バーコード」とは、分子識別子としての機能を果たし得る配列を含む核酸分子を指す。例えば、次世代配列決定（ＮＧＳ）ライブラリー調製の間に各ＤＮＡ断片に個々の「バーコード」配列を付加することができ、したがって、最終的なデータ解析の前に各リードを同定し、選別することができる。

【０１２１】

がんタイプ：本明細書で使用される場合、「がんタイプ」とは、例えば組織病理学によって定義されるがんのタイプまたはサブタイプを指す。がんタイプは、任意の従来の基準によって、例えば、所与の組織における発生（例えば、血液のがん、中枢神経系（ＣＮＳ）、脳のがん、肺がん（小細胞および非小細胞）、皮膚がん、鼻のがん、喉のがん、肝臓がん、骨がん、リンパ腫、膵臓がん、腸がん、直腸がん、甲状腺がん、膀胱がん、腎がん、口腔がん、胃がん、乳がん、前立腺がん、卵巣がん、肺がん、腸管がん、軟部組織がん、神経内分泌がん、胃食道がん、頭頸部がん、婦人科のがん、結腸直腸がん、尿路上皮がん、固形状態のがん、不均一ながん、均一ながん）、原発不明など、および／もしくは同じ細胞系列に基づいて（例えば、癌癌腫、肉腫、リンパ腫、胆管癌、白血病、中皮腫、黒色腫、または神経膠芽腫）、ならびに／または、がんマーカー、例えば、Ｈｅｒ２、ＣＡ１５－３、ＣＡ１９－９、ＣＡ－１２５、ＣＥＡ、ＡＦＰ、ＰＳＡ、ＨＣＧ、ホルモン受容体およびＮＭＰ－２２を示すがんとして定義することができる。がんは、ステージ（例えば、ステージ１、２、３、または４）および原発であるか二次性であるかによっても分類され得る。

【０１２２】

担体シグナル：本明細書で使用される場合、「担体シグナル」とは、機械５００による実行のための一時的または非一時的命令５０２を記憶する、符号化する、または保有することが可能な任意の無形媒体を指し、そのような命令５０２の通信を容易にするためのデジタルもしくはアナログ通信シグナルまたは他の無形媒体を含む。命令５０２は、ネットワーク５３４を通じて、一時的または非一時的送信媒体を使用し、ネットワークインターフェースデバイスを介して、およびいくつかの周知の転送プロトコールのうちのいずれか１つを使用して、送信または受信することができる。

【０１２３】

無細胞核酸：本明細書で使用される場合、「無細胞核酸」とは、細胞内に含有されてもそれ以外の方法で細胞と結合してもいない核酸、または、一部のインプリメンテーションでは、インタクトな細胞の除去後に試料中に残留している核酸を指す。無細胞核酸としては、例えば、対象由来の体液（例えば、血液、血漿、血清、尿、脳脊髄液（ＣＳＦ）など）を起源とする全ての非封入核酸を挙げることができる。無細胞核酸は、ゲノムＤＮＡ、ミトコンドリアＤＮＡ、循環ＤＮＡ、ｓｉＲＮＡ、ｍｉＲＮＡ、循環ＲＮＡ（ｃＲＮＡ）、ｔＲＮＡ、ｒＲＮＡ、核小体低分子ＲＮＡ（ｓｎｏＲＮＡ）、Ｐｉｗｉ相互作用ＲＮＡ（ｐｉＲＮＡ）、長鎖非コードＲＮＡ（長鎖ｎｃＲＮＡ）、および／またはこれらのうちのいずれかの断片を含む、ＤＮＡ（ｃｆＤＮＡ）、ＲＮＡ（ｃｆＲＮＡ）、およびそれら

のハイブリッドを含む。無細胞核酸は、二本鎖、一本鎖、またはそれらのハイブリッドであり得る。無細胞核酸は、分泌または細胞死プロセス、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得る。一部の無細胞核酸、例えば循環腫瘍ＤＮＡ（ｃｔＤＮＡ）は、がん細胞から体液中に放出される。その他のものは、健康な細胞から放出される。ｃｔＤＮＡは、非封入腫瘍由来断片化ＤＮＡであり得る。無細胞核酸は、１つまたは複数のエピジェネティック改変を有し得、例えば、無細胞核酸は、アセチル化、５－メチル化、ユビキチン化、リン酸化、ＳＵＭＯ化、リボシル化、および／またはシトルリン化され得る。

【０１２４】

細胞核酸：本明細書で使用される場合、「細胞核酸」とは、核酸が後に所与の解析プロセスの一部分として取り出されるとしても、少なくとも試料が対象から採取または収集された時点において、１つまたは複数の細胞内に配置されている核酸を意味する。

【０１２５】

分類領域：本明細書で使用される場合、「分類領域」とは、新生物細胞（例えば、腫瘍細胞およびがん細胞）における配列とは独立した変化を示し得る、またはがんを有する対象由来のｃｆＤＮＡにおける、がんが存在しない対象由来のｃｆＤＮＡと比べた、配列とは独立した変化を示し得る、ゲノム領域を指す。配列とは独立した変化の例としては、メチル化率（上昇または低下）、ヌクレオソーム分布、ＣＴＣＦ結合、転写開始部位、および調節タンパク質結合領域における変化が挙げられるが、これらに限定されない。１つまたは複数の例では、分類領域における配列とは独立した変化により、対象における単一の形態のがんの存在が示され得る。１つまたは複数の追加の例では、分類領域における配列とは独立した変化は、対象における複数の形態の存在に対応し得る。分類領域は、１つまたは複数のプローブによって富化させることができる。さらに、分類領域は、プライマー結合部位の対によって定義することができる。さらに、分類領域は、所定の開始ゲノム座および所定の終止ゲノム座によって定義することができる。分類領域は、約２５ヌクレオチド～約２５０ヌクレオチド、約５０ヌクレオチド～約２００ヌクレオチド、または約７５ヌクレオチド～約１５０ヌクレオチドを含み得る。例えば、分類領域は、差次的にメチル化された領域であり得る。「差次的にメチル化された領域」または「ＤＭＲ」とは、少なくとも１つの細胞もしくは組織タイプにおいて、少なくとも１つの他の細胞もしくは組織タイプ由来のＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能に異なる程度のメチル化を有するか、または疾患もしくは障害を有する対象から得られた少なくとも１つの細胞もしくは組織タイプにおいて、健康な対象から得られた同じ細胞もしくは組織タイプにおけるＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能に異なる程度のメチル化を有する、ＤＮＡの領域を指す。一部の実施形態では、差次的にメチル化された領域は、少なくとも１つの細胞または組織タイプにおいて、健康な個体におけるｃｆＤＮＡに寄与する少なくとも１つの他の細胞もしくは組織タイプからの、または健康な対象由来の同じ細胞もしくは組織タイプからのＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能により高い程度のメチル化（例えば、高メチル化領域／高メチル化標的領域）を有する。一部の実施形態では、差次的にメチル化された領域は、少なくとも１つの細胞または組織タイプにおいて、少なくとも１つの他の細胞または組織タイプ、例えば、健康な個体におけるｃｆＤＮＡに寄与する他の免疫細胞タイプおよび／もしくは細胞型からの、または健康な対象由来の同じ細胞もしくは組織タイプからのＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能により低い程度のメチル化（例えば、低メチル化領域／低メチル化標的領域）を有する。一部の実施形態では、分類領域は、高メチル化標的領域および／または低メチル化標的領域を含む。

【０１２６】

通信ネットワーク：本明細書で使用される場合、「通信ネットワーク」とは、アドホックネットワーク、イントラネット、エクストラネット、仮想プライベートネットワーク（ＶＰＮ）、ローカルエリアネットワーク（ＬＡＮ）、無線ＬＡＮ（ＷＬＡＮ）、ワイドエリアネットワーク（ＷＡＮ）、無線ＷＡＮ（ＷＷＡＮ）、メトロポリタンエリアネットワ

ーク（ＭＡＮ）、Ｉｎｔｅｒｎｅｔ、Ｉｎｔｅｒｎｅｔの一部分、公衆交換電話網（Ｐｕｂｌｉｃ Ｓｗｉｔｃｈｅｄ Ｔｅｌｅｐｈｏｎｅ Ｎｅｔｗｏｒｋ）（ＰＳＴＮ）の一部分、基本電話サービス（ｐｌａｉｎ ｏｌｄ ｔｅｌｅｐｈｏｎｅ ｓｅｒｖｉｃｅ）（ＰＯＴＳ）ネットワーク、携帯電話ネットワーク、無線ネットワーク、Ｗｉ－Ｆｉ（登録商標）ネットワーク、別のタイプのネットワーク、または２つまたはそれよりも多くのそのようなネットワークの組合せであってよい、ネットワーク１１４、１０３４の１つまたは複数の一部分を指す。例えば、ネットワーク１１４、１０３４またはネットワークの一部分は、無線またはセルラーネットワークを含んでよく、カップリングは、符号分割多元接続（Ｃｏｄｅ Ｄｉｖｉｓｉｏｎ Ｍｕｌｔｉｐｌｅ Ａｃｃｅｓｓ）（ＣＤＭＡ）接続、グローバルモバイル通信システム（Ｇｌｏｂａｌ Ｓｙｓｔｅｍ ｆｏｒ Ｍｏｂｉｌｅ ｃｏｍｍｕｎｉｃａｔｉｏｎｓ）（ＧＳＭ（登録商標））接続、または他のタイプのセルラーまたは無線カップリングであってよい。本例では、カップリングにより、Ｓｉｎｇｌｅ Ｃａｒｒｉｅｒ Ｒａｄｉｏ Ｔｒａｎｓｍｉｓｓｉｏｎ Ｔｅｃｈｎｏｌｏｇｙ（１ｘＲＴＴ）、Ｅｖｏｌｕｔｉｏｎ－Ｄａｔａ Ｏｐｔｉｍｉｚｅｄ（ＥＶＤＯ）テクノロジー、汎用パケット無線サービス（Ｇｅｎｅｒａｌ Ｐａｃｋｅｔ Ｒａｄｉｏ Ｓｅｒｖｉｃｅ）（ＧＰＲＳ）テクノロジー、Ｅｎｈａｎｃｅｄ Ｄａｔａ ｒａｔｅｓ ｆｏｒ ＧＳＭ（登録商標） Ｅｖｏｌｕｔｉｏｎ（ＥＤＧＥ）テクノロジー、３Ｇを含む、第三世代パートナーシッププロジェクト（ｔｈｉｒｄ Ｇｅｎｅｒａｔｉｏｎ Ｐａｒｔｎｅｒｓｈｉｐ Ｐｒｏｊｅｃｔ）（３ＧＰＰ（登録商標））、第４世代無線（４Ｇ）ネットワーク、ユニバーサルモバイル通信システム（Ｕｎｉｖｅｒｓａｌ Ｍｏｂｉｌｅ Ｔｅｌｅｃｏｍｍｕｎｉｃａｔｉｏｎｓ Ｓｙｓｔｅｍ）（ＵＭＴＳ）、高速パケットアクセス（Ｈｉｇｈ Ｓｐｅｅｄ Ｐａｃｋｅｔ Ａｃｃｅｓｓ）（ＨＳＰＡ）、Ｗｏｒｌｄｗｉｄｅ Ｉｎｔｅｒｏｐｅｒａｂｉｌｉｔｙ ｆｏｒ Ｍｉｃｒｏｗａｖｅ Ａｃｃｅｓｓ（ＷｉＭＡＸ）、Ｌｏｎｇ Ｔｅｒｍ Ｅｖｏｌｕｔｉｏｎ（ＬＴＥ）標準、様々な標準設定組織によって定義される他のテクノロジー、他の長距離プロトコル、または他のデータ転送テクノロジーなどの様々なタイプのデータ転送テクノロジーのいずれかをインプリメントすることができる。

【０１２７】

信頼区間：本明細書で使用される場合、「信頼区間」とは、所与のパラメーターの値がその値の範囲内に入る特定の確率が存在すると定義される一連の値を意味する。

【０１２８】

対照試料：本明細書で使用される場合、「対照試料」または「参照試料」とは、既知のコピー数バリエーションを有さない個体から得られた試料を指す。

【０１２９】

カバレッジ：本明細書で使用される場合、「カバレッジ」または「カバレッジメトリクス」とは、参照配列の特定のゲノム領域に対応する核酸分子または配列決定リードの数を指す。

【０１３０】

デオキシリボ核酸またはリボ核酸：本明細書で使用される場合、「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または改変ヌクレオチドを指す。ＤＮＡは、４つのタイプのヌクレオチド塩基：アデニン（Ａ）、チミン（Ｔ）、シトシン（Ｃ）、およびグアニン（Ｇ）を含むヌクレオチドの鎖を含み得る。本明細書で使用される場合、「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または改変ヌクレオチドを指す。ＲＮＡは、４つのタイプのヌクレオチド：Ａ、ウラシル（Ｕ）、Ｇ、およびＣを含むヌクレオチドの鎖を含み得る。本明細書で使用される場合、「ヌクレオチド」という用語は、天然ヌクレオチドまたは改変ヌクレオチドを指す。ヌクレオチドのある特定の対は、相補的に互いに特異的に結合する（相補塩基対形成と呼ばれる）。ＤＮＡでは、アデニン（Ａ）はチミン（Ｔ）と対を形成し、シトシン（Ｃ）はグアニン（Ｇ）と対を形成する。ＲＮＡでは、アデニン（Ａ）はウラシル（Ｕ）と対を形成し、シトシン（Ｃ）はグアニン（Ｇ）と対を形成する。第１の核酸鎖が、第１の

鎖のヌクレオチドと相補的であるヌクレオチドで構成される第２の核酸鎖に結合する場合、２つの鎖は結合して二本鎖を形成する。本明細書で使用される場合、「核酸配列決定データ」、「核酸配列決定情報」、「配列情報」、「配列表示」、「核酸配列」、「ヌクレオチド配列」、「ゲノム配列」、「遺伝子配列」、「断片配列」、「配列決定リード」、または「核酸配列決定リード」とは、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、または断片）中のヌクレオチド塩基（例えば、アデニン、グアニン、シトシン、およびチミンまたはウラシル）の順序および同一性を示す、任意の情報またはデータを示す。本教示は、これだけに限定されないが、キャピラリー電気泳動、マイクロアレイ、ライゲーションに基づくシステム、ポリメラーゼに基づくシステム、ハイブリダイゼーションに基づくシステム、直接的または間接的なヌクレオチド同定システム、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、および電子シグネチャーに基づくシステムを含む、全ての利用可能な様々な技術、プラットフォームまたはテクノロジーを使用し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。

【０１３１】

差次的にメチル化された領域：本明細書で使用される場合、差次的にメチル化された領域」とは、少なくとも１つの細胞もしくは組織タイプにおいて、少なくとも１つの他の細胞もしくは組織タイプ由来のＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能に異なる程度のメチル化を有する、または疾患もしくは障害を有する対象から得られた少なくとも１つの細胞もしくは組織タイプにおいて、健康な対象から得られた同じ細胞もしくは組織タイプにおけるＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能に異なる程度のメチル化を有するＤＮＡの領域を指す。一部の実施形態では、差次的にメチル化された領域は、少なくとも１つの細胞または組織タイプ、例えば、少なくとも１つの免疫細胞タイプにおいて、少なくとも１つの他の細胞もしくは組織タイプ、例えば、健康な個体におけるｃｆＤＮＡに寄与する他の免疫細胞タイプおよび／もしくは細胞型からの、または健康な対象由来の同じ細胞もしくは組織タイプからのＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能により高い程度のメチル化（例えば、高メチル化領域）を有する。一部の実施形態では、差次的にメチル化された領域は、少なくとも１つの細胞または組織タイプ、例えば、少なくとも１つの免疫細胞タイプにおいて、少なくとも１つの他の細胞または組織タイプ、例えば、健康な個体におけるｃｆＤＮＡに寄与する他の免疫細胞タイプおよび／もしくは細胞型からの、または健康な対象由来の同じ細胞もしくは組織タイプからのＤＮＡの同じ領域におけるメチル化の程度と比べて検出可能により低い程度のメチル化（例えば、低メチル化領域）を有する。

【０１３２】

ドライバー変異：本明細書で使用される場合、「ドライバー変異」とは、がんの進行を駆動する変異を意味する。

【０１３３】

エピジェネティック標的領域：本明細書で使用される場合、「エピジェネティック標的領域」とは、異なる細胞もしくは組織タイプ（例えば、異なるタイプの免疫細胞）において、もしくは新生物細胞（例えば、腫瘍細胞およびがん細胞）において正常細胞と比較して、配列とは独立した相違を示し得る、または、異なる細胞型由来もしくはがんを有する対象由来のＤＮＡ、例えばｃｆＤＮＡにおいて、健康な対象由来のＤＮＡ、例えばｃｆＤＮＡと比べて、もしくは、通常ｃｆＤＮＡには実質的に寄与しない異なる細胞もしくは組織タイプ（例えば、免疫、肺、結腸など）を起源とするｃｆＤＮＡにおいて、バックグラウンドｃｆＤＮＡ（例えば、造血細胞を起源とするｃｆＤＮＡ）と比べて、配列とは独立した相違（すなわち、ヌクレオチド配列に変化がない相違、例えば、メチル化、ヌクレオソーム分布、もしくは他のエピジェネティック特色の相違）を示し得る標的領域を指す。配列とは独立した変化の例としては、メチル化（増加または低減）、ヌクレオソーム分布、ｃｆＤＮＡ断片化パターン、ＣＣＣＴＣ結合因子（「ＣＴＣＦ」）結合、転写開始部位（例えば、ＲＮＡポリメラーゼ構成成分の結合、調節タンパク質の結合、断片化特徴、お

よびヌクレオソーム分布のいずれか１つまたは複数（any one of more of）に関して）、および調節タンパク質結合領域における変化が挙げられるが、これらに限定されない。したがって、エピジェネティック標的領域セットとしては、高メチル化標的領域セット、低メチル化標的領域セット、および断片化可変標的領域セット、例えば、ＣＴＣＦ結合部位および転写開始部位が挙げられるが、これらに限定されない。本目的のために、新生物、腫瘍、またはがんに関連する局所増幅および／または遺伝子融合を受けやすい座もまた、エピジェネティック標的領域セットに含まれ得るが、これは、シーケンシングによるコピー数変化の検出または参照ゲノム内の１つを上回る座にマッピングする融合配列が、例えば、局所増幅および／または遺伝子融合の検出は１つまたは数個の個別の位置における塩基コールの正確度に依存しないため比較的浅いシーケンシング深度で検出することができるという点において、ヌクレオチド置換、挿入、または欠失の検出よりも、上記で考察された例示的なエピジェネティック変化の検出に類似する傾向にあるためである。エピジェネティック標的領域セットとは、エピジェネティック標的領域のセットである。

【０１３４】

高メチル化：本明細書で使用される場合、「高メチル化」とは、同じゲノム座由来の核酸分子の集団（例えば、試料）内の他の核酸分子と比べて、増加したレベルまたは程度の核酸分子（複数可）のメチル化を指す。一部の実施形態では、高メチル化ＤＮＡは、少なくとも１個のメチル化シトシン、少なくとも２個のメチル化シトシン、少なくとも３個のメチル化シトシン、少なくとも５個のメチル化シトシン、または少なくとも１０個のメチル化シトシンを含むＤＮＡ分子を含み得る。

【０１３５】

低メチル化：本明細書で使用される場合、「低メチル化」とは、同じゲノム座由来の核酸分子の集団（例えば、試料）内の他の核酸分子と比べて、減少したレベルまたは程度の核酸分子（複数可）のメチル化を指す。一部の実施形態では、低メチル化ＤＮＡは、非メチル化ＤＮＡ分子を含む。一部の実施形態では、低メチル化ＤＮＡは、０個のメチル化シトシン、多くとも１個のメチル化シトシン、多くとも２個のメチル化シトシン、多くとも３個のメチル化シトシン、多くとも４個のメチル化シトシン、または多くとも５個のメチル化シトシンを含むＤＮＡ分子を含み得る。

【０１３６】

免疫療法：本明細書で使用される場合、「免疫療法」とは、がん細胞を死滅させるもしくは少なくともがん細胞の成長を阻害するため、好ましくはがんのさらなる成長を低減させるため、がんのサイズを縮小させるため、および／またはがんを排除するために、免疫系を刺激するように作用する１つまたは複数の作用剤（agent）を用いた処置を指す。そのような作用剤の一部は、がん細胞に存在する標的に結合し、一部は、免疫細胞に存在し、がん細胞には存在しない標的に結合し、一部は、がん細胞と免疫細胞の両方に存在する標的に結合する。そのような作用剤としては、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的免疫応答の持続時間および大きさをモジュレートして、二次的組織損傷を最小限にする、免疫系経路の阻害剤である（例えば、Pardoll, Nature Reviews Cancer 12, 252-264 (2012)を参照されたい）。作用剤の例としては、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。作用剤の他の例としては、ＩＬ－１β、ＩＬ－６、およびＴＮＦ－αなどの炎症促進サイトカインが挙げられる。作用剤の他の例は、腫瘍に対して活性化されたＴ細胞、例えば、Ｔ細胞によって認識される腫瘍抗原を標的とするキメラ抗原を発現させることによって活性化されたＴ細胞である。

【０１３７】

インデル：本明細書で使用される場合、「インデル」とは、対象のゲノムにおけるヌクレオチドの挿入または欠失を伴う変異を指す。

【０１３８】

検出限界（ＬｏＤ）：本明細書で使用される場合、「検出限界」とは、所与のアッセイまたは解析アプローチによって測定することができる、試料中の物質（例えば、核酸）の最小量を意味する。

【０１３９】

機械可読媒体：本明細書で使用される場合、「機械可読媒体」とは、命令５０２およびデータを一時的にまたは永久に記憶させることができるコンポーネント、デバイス、または他の有形媒体を指し、それらとしては、ランダムアクセスメモリー（ＲＡＭ）、読み取り専用メモリー（ＲＯＭ）、バッファーメモリー、フラッシュメモリー、光学媒体、磁気媒体、キャッシュメモリー、他のタイプのストレージ（例えば、消去可能プログラム可能読み取り専用メモリー（ＥＥＰＲＯＭ））および／またはこれらの任意の好適な組合せが挙げられ得るが、これらに限定されない。「機械可読媒体」という用語は、命令５０２を記憶させることができる単一の媒体または複数の媒体（例えば、集中型もしくは分散型データベース、または関連するキャッシュおよびサーバー）を含むと考えることができる。「機械可読媒体」という用語はまた、機械５００による実行のための命令５０２（例えば、コード）を、命令５０２が、機械５００の１つまたは複数のプロセッサー５０４によって実行されるときに機械５００に本明細書に記載される方法論のいずれか１つまたは複数を実施させるように、記憶させることが可能である任意の媒体または複数の媒体の組合せを含むと考えるものとする。したがって、「機械可読媒体」とは、単一のストレージ装置またはデバイス、ならびに、複数のストレージ装置またはデバイスを含む「クラウドに基づく」ストレージシステムまたはストレージネットワークを指す。シグナル自体は「機械可読媒体」という用語から除外される。

【０１４０】

最大ＭＡＦ：本明細書で使用される場合、「最大（ｍａｘｉｍｕｍ）ＭＡＦ」または「最大（ｍａｘ）ＭＡＦ」とは、試料中の全ての体細胞バリアントの最大ＭＡＦ（変異型アレル画分）を指す。

【０１４１】

メチル化：本明細書で使用される場合、「メチル化」または「ＤＮＡメチル化」という用語は、核酸分子におけるヌクレオチド塩基へのメチル基の付加を指す。一部の実施形態では、メチル化は、ＣｐＧ部位（シトシン－リン酸－グアニン部位（すなわち、核酸配列の５’→３’方向でシトシンの後にグアニンがある）におけるシトシンへのメチル基の付加を指す。一部の実施形態では、ＤＮＡメチル化は、アデニンへのメチル基の付加、例えば、N^6－メチルアデニンにおけるものを指す。一部の実施形態では、ＤＮＡメチル化は、５－メチル化（シトシンの６炭素環の５番目の炭素の改変）である。一部の実施形態では、５－メチル化は、５－メチルシトシン（５ｍＣ）を作製するためのシトシンの５Ｃ位へのメチル基の付加を指す。一部の実施形態では、メチル化は、５ｍＣの誘導体を含む。５ｍＣの誘導体としては、５－ヒドロキシメチルシトシン（５－ｈｍＣ）、５－ホルミルシトシン（５－ｆＣ）、および５－カルボキシルシトシン（caryboxylcytosine）（５－ｃａＣ）が挙げられるが、これらに限定されない。一部の実施形態では、ＤＮＡメチル化は、３Ｃメチル化（シトシンの６炭素環の３番目の炭素の改変）である。一部の実施形態では、３Ｃメチル化は、３－メチルシトシン（３ｍＣ）を生成するためのシトシンの３Ｃ位へのメチル基の付加を含む。メチル化はまた、非ＣｐＧ部位でも生じ得、例えば、メチル化は、ＣｐＡ、ＣｐＴ、またはＣｐＣ部位で生じ得る。ＤＮＡメチル化は、メチル化ＤＮＡ領域の活性を変化させ得る。例えば、プロモーター領域内のＤＮＡがメチル化された場合、遺伝子の転写が抑制され得る。ＤＮＡメチル化は、正常な発達に極めて重要であり、メチル化の異常は、エピジェネティック調節を破壊し得る。エピジェネティック調節における破壊、例えば、抑制は、疾患、例えば、がんを引き起こし得る。ＤＮＡにおけるプロモーターメチル化は、がんを示し得る。

【０１４２】

メチル化依存性ヌクレアーゼ：本明細書で使用される場合、「メチル化依存性ヌクレアーゼ」という用語は、非メチル化ＤＮＡと比べて、メチル化ＤＮＡを優先的に切断する、

ヌクレアーゼを指す。例えば、メチル化依存性ヌクレアーゼは、認識配列内の核酸塩基のうちの少なくとも１つ、例えば、シトシンのメチル化に依存する様式で、制限部位などの認識配列においてまたはその近傍で、切断し得る。一部の実施形態では、メチル化依存性ヌクレアーゼの核酸溶解活性は、標準的な核酸溶解アッセイにおいて、メチル化認識部位に対して、非メチル化対照と比べて少なくとも１０倍、２０倍、５０倍、または１００倍高い。メチル化依存性ヌクレアーゼは、メチル化依存性制限酵素を含む。

【０１４３】

メチル化依存性制限酵素：本明細書で使用される場合、「メチル化依存性制限酵素」または「ＭＤＲＥ」は、ＤＮＡのメチル化（例えば、シトシンメチル化）に依存する、制限酵素を指し、すなわち、ヌクレオチド塩基におけるメチル基の存在または非存在により、酵素が標的ＤＮＡを切断する割合が変化する。一部の実施形態では、メチル化依存性制限酵素は、特定のヌクレオチド塩基が、認識配列において非メチル化である場合、ＤＮＡを切断しない。例えば、ＭｓｐＪＩは、認識配列「ｍＣＮＮＲ（Ｎ９）」を有するメチル化依存性制限酵素であり、認識配列内にメチル化シトシン（ｍＣ）が不在である場合には、ＤＮＡを切断しない。

【０１４４】

メチル化感受性ヌクレアーゼ：本明細書で使用される場合、「メチル化感受性ヌクレアーゼ」という用語は、メチル化ＤＮＡと比べて、非メチル化ＤＮＡを優先的に切断する、ヌクレアーゼを指す。例えば、メチル化感受性ヌクレアーゼは、認識配列内の核酸塩基のうちの少なくとも１つ、例えば、シトシンのメチル化の欠如に依存する様式で、制限部位などの認識配列においてまたはその近傍で、切断し得る。一部の実施形態では、メチル化感受性ヌクレアーゼの核酸溶解活性は、標準的な核酸溶解アッセイにおいて、非メチル化認識配列に対して、メチル化対照と比べて少なくとも１０倍、２０倍、５０倍、または１００倍高い。メチル化感受性ヌクレアーゼは、メチル化感受性制限酵素を含む。

【０１４５】

メチル化感受性制限酵素：本明細書で使用される場合、「メチル化感受性制限酵素」または「ＭＳＲＥ」は、ＤＮＡのメチル化状態（例えば、シトシンメチル化）に感受性である、制限酵素を指し、すなわち、ヌクレオチド塩基におけるメチル基の存在または非存在により、酵素が標的ＤＮＡを切断する割合が変化する。一部の実施形態では、メチル化感受性制限酵素は、特定のヌクレオチド塩基が、認識配列においてメチル化されている場合、ＤＮＡを切断しない。例えば、ＨｐａＩＩは、認識配列「ＣＣＧＧ」を有するメチル化感受性制限酵素であり、認識配列内の第２のシトシンがメチル化されている場合、ＤＮＡを切断しない。

【０１４６】

メチル化率：本明細書で使用される場合、「メチル化率」とは、試料中の解析される特定のゲノム領域におけるＤＮＡ分子上の所与の塩基（例えば、ＣｐＧのシトシン残基）がメチル化されている確率、尤度、またはパーセンテージを指す。一部の実施形態では、メチル化率を、１つまたは複数の潜在的にメチル化された塩基を含む定義された領域に適用することができる。一部の実施形態では、メチル化率は、ＤＮＡ分子におけるメチル化されたＣｐＧ残基のパーセンテージを指す。一部の実施形態では、メチル化率は、特定のゲノム位置またはゲノム領域に対して整列された分子におけるメチル化されたＣｐＧ残基のパーセンテージを指す。メチル化率は、これだけに限定されないが、バイサルファイトシーケンシング（ＴＡＰＳ、ＥＭ－ＳＥＱなどの任意の単一塩基分解能）の使用またはパーティショニング（ＤＮＡ分子分解能）の使用のいずれかを含む、様々な方法によって測定することができる。メチル化率は、種々の方法で測定することができる。１つの推定は、最後にはどれだけのＤＮＡ断片が各メチル化依存性分配に入るかを計数することによるもの、または、バイサルファイトシーケンシングもしくは任意の他の塩基レベルの分解能の配列決定方法の場合では断片あたりの変換されたＣｐＧの数を計数することによるものであり得る。さらに、メチル化依存性パーティショニングの場合では、率の計算を、既知のメチル化状態を有する予め定義された領域（すなわち、陽性対照領域および／または陰性

対照領域）のセットまたは既知のメチル化状態を有するスパイクイン合成ＤＮＡを使用して正規化し、率をパラメーター化した分配分布を導き出し、最尤アプローチを使用して率を推定することができる。１つまたは複数の例では、メチル化率を、ゲノム領域の一部分に対応する配列決定リードの存在量を決定することによって決定することができる。ゲノム領域の一部分は、少なくとも閾値数の配列決定リードが重複する、ゲノム領域のいくつかのゲノム位置を含み得る。

【０１４７】

メチル化状態：本明細書で使用される場合、「メチル化状態」または「メチル化状態（ｍｅｔｈｙｌａｔｉｏｎ　ｓｔａｔｅ）」は、核酸分子中の特定のゲノム位置のＤＮＡ塩基（例えばシトシン）上のメチル基の存在または非存在を指すことができる。これはまた、核酸配列中のメチル化の程度（例えば、高メチル化、低メチル化、中間メチル化、または非メチル化核酸分子）も指し得る。メチル化状態はまた、特定の核酸分子においてメチル化されたヌクレオチドの数も指し得る。

【０１４８】

改変ヌクレオチド特異的結合試薬：本明細書で使用される場合、改変ヌクレオチドに特異的である、または改変ヌクレオチドを標的とする結合試薬を指す。例えば、改変ヌクレオチドは、メチル化されているヌクレオチドであり得、したがって、結合試薬は、メチル化ヌクレオチドに対して特異的であり得る。結合試薬の例としては、メチル化結合タンパク質（「ＭＢＰ」）もしくはそのバリアントのメチル結合ドメイン（ＭＢＤ）、抗体（および抗体バリアント、例えば単鎖抗体）、アプタマー、またはこれらの組合せが挙げられるが、これらに限定されない。したがって、全体を通して開示されるように、ＭＢＤの使用は、任意の他の改変ヌクレオチド特異的結合試薬と交換することができるが、ただし、改変ヌクレオチド特異的結合試薬は、選択されたインプリメンテーションにおいて目的の特定の改変塩基に対する所望の特異性および親和性を有するものとする。

【０１４９】

変異型アレル画分：本明細書で使用される場合、「変異型アレル画分」、「変異用量」、または「ＭＡＦ」とは、所与の試料における所与のゲノム位置にアレルの変更または変異を有する核酸分子の画分を指す。ＭＡＦは、一般に、画分またはパーセンテージとして表される。例えば、ＭＡＦは、所与の座に存在する全ての体細胞バリアントまたはアレルの約０．５未満、０．１未満、０．０５未満、または０．０１未満（すなわち、約５０％未満、１０％未満、５％未満、または１％未満）であり得る。

【０１５０】

変異：本明細書で使用される場合、「変異」とは、既知の参照配列からのバリエーションを指し、例えば、一塩基バリアント（ＳＮＶ）、コピー数バリアントまたはバリエーション（ＣＮＶ）／異常、挿入または欠失（インデル）、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、反復拡大、およびエピジェネティックバリアントなどの変異を含む。変異は、生殖系列変異または体細胞変異であり得る。一部の例では、比較するための参照配列は、試験試料を提供する対象の種の野生型ゲノム配列、典型的には、ヒトゲノムである。

【０１５１】

変異コーラー：本明細書で使用される場合、「変異コーラー」とは、試験試料データ（例えば、対象から得られた配列情報）における変異を同定するために使用されるアルゴリズム（ソフトウェアに具体化されるまたは他の方法でコンピュータによりインプリメントされる）を意味する。

【０１５２】

変異数：本明細書で使用される場合、「変異数（ｍｕｔａｔｉｏｎ　ｃｏｕｎｔ）」または「変異数（ｍｕｔａｔｉｏｎａｌ　ｃｏｕｎｔ）」とは、核酸試料の全ゲノムまたはエクソームまたは標的化領域における体細胞変異の数を指す。

【０１５３】

陰性対照領域：本明細書で使用される場合、「陰性対照領域」とは、ＤＮＡががん細胞

に由来するのかまたは正常細胞に由来するのかにかかわらず、本質的に全ての試料においてメチル化されていないまたは低メチル化されていることが予想されるゲノム領域を指す。

【０１５４】

新生物：本明細書で使用される場合、「新生物」および「腫瘍」という用語は、互換的に使用される。それらは、対象における細胞の異常な成長を指す。新生物または腫瘍は、良性、おそらく悪性、または悪性であり得る。悪性腫瘍は、がんまたはがん様腫瘍と呼ばれる。

【０１５５】

次世代配列決定：本明細書で使用される場合、「次世代シーケンシング」または「ＮＧＳ」は、従来のサンガーおよびキャピラリー電気泳動に基づくアプローチと比較して増加した処理能力を有する、例えば一度に何十万もの比較的小さいシーケンシング読み取りデータを生成する能力を有するシーケンシングテクノロジーを指す。次世代シーケンシング技術の一部の例としては、これらに限定されないが、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。

【０１５６】

核酸タグ：本明細書で使用される場合、「核酸タグ」は、異なる試料からの核酸（例えば、試料インデックスを表す）を識別するため、または異なるタイプのもしくは異なるプロセシングを受けている同じ試料中の異なる核酸分子（例えば、分子バーコードを表す）を識別するために使用される短い核酸（例えば、長さが約５００ヌクレオチド未満、約１００ヌクレオチド未満、約５０ヌクレオチド未満、または約１０ヌクレオチド未満）を指す。核酸タグは、既定の、固定された、非ランダム、ランダム、または半ランダムオリゴヌクレオチド配列を含む。そのような核酸タグは、異なる核酸分子または異なる核酸試料もしくは部分試料を標識するために使用され得る。核酸タグは、一本鎖、二本鎖、または少なくとも部分的に二本鎖であり得る。核酸タグは、必要に応じて同じ長さまたは多様な長さを有する。核酸タグはまた、１つもしくは複数の平滑末端を有する二本鎖分子を含み、５’もしくは３’一本鎖領域（例えば、オーバーハング）を含み、および／または所定の分子内の他の場所で１つもしくは複数の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試料核酸）の１つの末端または両方の末端に付着させることができる。核酸タグは、所定の核酸の起源の試料、形態、またはプロセシングなどの情報を明らかにするために復号することができる。例えば、核酸タグはまた、核酸タグを検出する（例えば、読み取る）ことによって核酸がその後にデコンボリュートされる、異なる分子バーコードおよび／または試料インデックスを有する核酸を含む複数の試料のプールおよび／または並列プロセシングを可能にするためにも使用することができる。核酸タグはまた、識別子（例えば、分子識別子、試料識別子）とも呼ばれ得る。さらに、またはあるいは、核酸タグは分子識別子（例えば、異なる分子、または同じ試料もしくは部分試料中の異なる親分子のアンプリコンの間を識別するために）として使用することができる。これは、例えば、所定の試料中の異なる核酸分子を一意的にタグ付けすること、またはそのような分子を非一意的にタグ付けすることを含む。非一意的タグ付け応用の場合、限定数のタグ（すなわち、分子バーコード）を使用して、異なる分子を、少なくとも１つの分子バーコードと組み合わせてその内因性の配列情報（例えば、それらが選択された参照配列にマッピングされる開始および／もしくは終止位置、配列の一方もしくは両方の末端の部分配列、ならびに／または配列の長さ）に基づいて識別することができるように、各核酸分子をタグ付けしてもよい。任意の２つの分子が同じ内因性の配列情報（例えば、開始および／もしくは終止位置、配列の一方もしくは両方の末端の部分配列、ならびに／または長さ）を有し得る確率、同様に同じ分子バーコードを有し得る確率が低くなるように（例えば、約１０％未満、約５％未満、約１％未満、または約０．１％未満の見込み）、十分数の異なる分子バーコードを使用する。

【０１５７】

分配すること：本明細書で使用される場合、「分配すること」とは、核酸分子の特徴に基づいて、試料中の核酸分子の混合物を物理的に分離または分画することを指す。分配は、分子の物理的分配であり得る。分配は、核酸分子を、エピジェネティック特色（例えば、メチル化に関する）のレベルに基づいて、群またはセットに分離することを含み得る。例えば、核酸分子は、核酸分子のメチル化のレベルに基づいて、分配することができる。一部の実施形態では、分配に使用される方法およびシステムは、ＰＣＴ特許出願第ＰＣＴ／ＵＳ２０１７／０６８３２９号に見出すことができ、これは、参照によりその全体が本明細書に組み込まれる。

【０１５８】

分配セット：本明細書で使用される場合、「分配セット」または「分配」は、核酸分子または核酸分子と会合したタンパク質の結合剤への差次的結合親和性に基づいてセットまたは群に分配された、核酸分子のセットを指す。分配セットはまた、部分試料とも称され得る。結合剤は、エピジェネティック改変を有するヌクレオチドを含む核酸分子に優先的に結合する。例えば、エピジェネティック改変がメチル化である場合、結合剤は、メチル結合ドメイン（ＭＢＤ）タンパク質であり得る。一部の実施形態では、分配セットは、特定のエピジェネティック特色（例えば、メチル化）のレベルまたは程度に属する核酸分子を含み得る。例えば、核酸分子は、高度メチル化核酸分子の１つのセット（第１の部分試料、高分配、高分配セット、または高メチル化分配セット）、低メチル化核酸分子の第２のセット（第２の部分試料、低分配、低分配セット、または低メチル化分配セット）、および中間メチル化核酸分子の第３のセット（第３の部分試料、中間分配セット、中間メチル化分配セット、残留分配、または残留分配セット）の３つのセットに分配され得る。別の例では、核酸分子は、メチル化ヌクレオチドの数に基づいて、分配され得、１つの分配セットは、９個のメチル化ヌクレオチドを有する核酸分子を有し得、別の分配セットは、非メチル化核酸分子（０個のメチル化ヌクレオチド）を有し得る。

【０１５９】

ポリヌクレオチド：本明細書で使用される場合、「ポリヌクレオチド」、「核酸」、「核酸分子」、「ポリヌクレオチド分子」、または「オリゴヌクレオチド」とは、ヌクレオシド間連結によって接合したヌクレオシド（デオキシリボヌクレオシド、リボヌクレオシド、またはそのアナログを含む）の線形ポリマーを指す。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み得る。オリゴヌクレオチドは、多くの場合、少数の単量体単位、例えば３～４から数百もの単量体単位のサイズの範囲である。ポリヌクレオチドが文字の配列、例えば「ＡＴＧＣＣＴＧ」によって表される場合は必ず、ヌクレオチドは、左から右に５’→３’の順であること、および、ＤＮＡの場合では、特に断りのない限り、「Ａ」はデオキシアデノシンを示し、「Ｃ」はデオキシシチジンを示し、「Ｇ」はデオキシグアノシンを示し、「Ｔ」はデオキシチミジンを示すことが理解されよう。Ａ、Ｃ、Ｇ、およびＴという文字は、当技術分野における標準と同様に、塩基自体、ヌクレオシド、または塩基を含むヌクレオチドを指すために使用され得る。

【０１６０】

陽性対照領域：本明細書で使用される場合、本明細書で使用される場合、「陽性対照領域」とは、ＤＮＡががん細胞に由来するのかまたは正常細胞に由来するのかにかかわらず、本質的に全ての試料においてメチル化または高メチル化されていることが予想されるゲノム領域を指す。

【０１６１】

プローブ：本明細書で使用される場合、「プローブ」とは、機能を含むポリヌクレオチドを指す。機能は、検出可能な標識（蛍光）、結合部分（ビオチン）、または固体支持体（磁気吸着可能な粒子またはチップ）であってよい。プローブは、標的核酸配列にハイブリダイズする一本鎖ＤＮＡ／ＲＮＡポリヌクレオチドまたは二本鎖ＤＮＡポリヌクレオチド（例えば、ＳｕｒｅＳｅｌｅｃｔ（登録商標）プローブ、Ａｇｉｌｅｎｔ　Ｔｅｃｈｎｏｌｏｇｉｅｓ）を含み得る。プローブを使用した配列捕捉は、一般に、プローブの配列と相補的な（またはほぼ相補的な）標的核酸配列の少なくとも一部分における連続したヌ

クレオチドの数に部分的に依存する。一部の例では、プローブは、ドライバー変異に対応し得る。

【０１６２】

プロセシング：本明細書で使用される場合、「プロセシング」、「計算（calculating）」、および「比較（comparing）」という用語は、互換的に使用され得る。ある特定の適用では、用語は、相違、例えば、数または配列の相違を決定することを指す。例えば、遺伝子発現、コピー数バリエーション（ＣＮＶ）、インデル、および／または一塩基バリアント（ＳＮＶ）値もしくは配列をプロセシングすることができる。

【０１６３】

プロセッサー：本明細書で使用される場合、「プロセッサー」とは、データ値を制御信号（例えば、「コマンド」、「ｏｐコード」、「機械コード」など）に従って操作し、機械を作動させるために適用される対応する出力信号を生じさせる、任意の回路または仮想回路（実際のプロセッサーでのロジック実行によってエミュレートされる物理的回路）を指す。プロセッサーは、例えば、ＣＰＵ、ＲＩＳＣプロセッサー、ＣＩＳＣプロセッサー、ＧＰＵ、ＤＳＰ、ＡＳＩＣ、ＲＦＩＣまたはこれらの任意の組合せであってよい。プロセッサーはさらに、命令を同時に実行することができる、２つまたはそれよりも多くの独立したプロセッサー（時には「コア」と称される）を有するマルチコアプロセッサーであってよい。

【０１６４】

プロモーター領域：本明細書で使用される場合、「プロモーター領域」とは、遺伝子の特異的な転写を開始するために必要である、細胞の合成機構、または導入された合成機構によって認識されるＤＮＡ配列を指す。

【０１６５】

本明細書で使用される場合、「定量的測定」は、絶対的または相対的測定を指す。定量的測定は、これらに限定されないが、数、統計学的測定（例えば、頻度、平均値、中央値、標準偏差、または分位数）、または程度もしくは相対的数量（例えば、高、中、および低）であり得る。定量的測定は、２つの定量的測定の比であり得る。定量的測定は、定量的測定の線形の組合せであり得る。定量的測定は正規化した測定であり得る。

【０１６６】

本明細書で使用される場合、「参照配列」は、実験によって決定した配列と比較する目的のために使用される公知の配列を指す。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、少なくとも約２０個、少なくとも約５０個、少なくとも約１００個、少なくとも約２００個、少なくとも約２５０個、少なくとも約３００個、少なくとも約３５０個、少なくとも約４００個、少なくとも約４５０個、少なくとも約５００個、少なくとも約１０００個、またはそれよりも多くのヌクレオチドを含み得る。参照配列は、ゲノムもしくは染色体の単一の連続する配列と整列し得る、またはゲノムもしくは染色体の異なる領域と整列する連続していないセグメントを含み得る。参照配列の例としては、例えば、ヒトゲノム参照配列、例えばｈＧ１９およびｈＧ３８が挙げられる。

【０１６７】

本明細書で使用される場合、「試料」は、本明細書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任意のものを意味する。

【０１６８】

感度：本明細書で使用される場合、「感度」とは、所与のＭＡＦおよびカバレッジにおける一塩基バリアント、挿入、および欠失の存在を検出する確率、ならびに所与の腫瘍画分およびカバレッジにおけるコピー数バリアントの存在を検出する確率を意味する。

【０１６９】

本明細書で使用される場合、「シーケンシング」は、生体分子、例えばＤＮＡまたはＲＮＡなどの核酸の配列（例えば、同一性および単量体単位の順序）を決定するために使用されるいくつかのテクノロジーのいずれかを指す。シーケンシング方法の例としては、こ

れらに限定されないが、標的化シーケンシング、一分子リアルタイムシーケンシング、エクソン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、トランジスタ媒介シーケンシング、直接シーケンシング、ランダムショットガンシーケンシング、サンガージデオキシターミネーションシーケンシング、全ゲノムシーケンシング、ハイブリダイゼーションによるシーケンシング、パイロシーケンシング、キャピラリー電気泳動、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、超並列シグネチャーシーケンシング、エマルジョンＰＣＲ、より低い変性温度での同時増幅－ＰＣＲ（ＣＯＬＤ－ＰＣＲ）、マルチプレックスＰＣＲ、可逆的色素ターミネーターによるシーケンシング、ペアードエンドシーケンシング、ニアターム（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、ＳＯＬｉＤ（商標）シーケンシング、ＭＳ－ＰＥＴシーケンシング、およびそれらの組合せが挙げられる。一部のインプリメンテーションでは、シーケンシングは、遺伝子アナライザー、数ある中でも例えばＩｌｌｕｍｉｎａ，Ｉｎｃ．、Ｐａｃｉｆｉｃ Ｂｉｏｓｃｉｅｎｃｅｓ，Ｉｎｃ．、またはＡｐｐｌｉｅｄ Ｂｉｏｓｙｓｔｅｍｓ／Ｔｈｅｒｍｏ Ｆｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃなどから市販されている遺伝子アナライザーによって実施され得る。

【０１７０】

一塩基バリアント：本明細書で使用される場合、「一塩基バリアント」または「ＳＮＶ」とは、ゲノム内の特定の位置で起こる、一塩基における変異またはバリエーションを意味する。

【０１７１】

体細胞突然変異：本明細書で使用される場合、「体細胞突然変異」または「体細胞変異」という用語は、互換的に使用される。それらは、受胎後に起こるゲノムの突然変異を指す。体細胞突然変異は、生殖細胞を除く体の任意の細胞で起こり得、したがって子孫に受け継がれない。

【０１７２】

特異的に結合する：本明細書で使用される場合、プローブまたは他のオリゴヌクレオチドおよび標的配列の文脈における「特異的に結合する」は、適切なハイブリダイゼーション条件下で、オリゴヌクレオチドまたはプローブがその標的配列またはその複製物にハイブリダイズして安定なプローブ：標的ハイブリッドを形成するが、同時に安定なプローブ：非標的ハイブリッドの形成は最小限であることを意味する。このように、プローブは、非標的配列に対してよりも十分に大きい程度に標的配列またはその複製物にハイブリダイズして、標的配列の捕捉または検出を可能にする。適切なハイブリダイゼーション条件は、当技術分野で周知であり、配列組成に基づいて予測され得るか、または通常の試験方法を使用することにより決定され得る（例えば、参照により本明細書に組み込まれる、Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd ed.(Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)§§ 1.90-1.91, 7.37-7.57, 9.47-9.51および11.47-11.57、特に§§ 9.50-9.51, 11.12-11.13, 11.45-11.47および11.55-11.57を参照されたい）。

【０１７３】

本明細書で使用される場合、「対象」は、動物、例えば哺乳動物種（例えば、ヒト）、または鳥類（例えば、鳥）種、または他の生物、例えば植物を指す。より具体的には、対象は脊椎動物、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。動物は、農場動物（例えば、肉牛、乳牛、家禽、ウマ、ブタなど）、競技用動物、およびコンパニオン動物（例えば、ペットまたは介助動物）を含む。対象は、健康な個体、疾患もしくは疾患に対する素因を有するもしくは有することが疑われる個体、または治療を必

要とするもしくは治療を必要とすることが疑われる個体であり得る。「個体」または「患者」という用語は、「対象」と互換的であると意図される。

【０１７４】

例えば、対象は、がんを有すると診断されている、がん治療を受ける予定である、および／または少なくとも１つのがん治療を受けたことがある個体であり得る。対象はがんの寛解期にあり得る。別の例として、対象は、自己免疫疾患を有すると診断された個体であり得る。別の例として、対象は、疾患、例えばがん、自己免疫疾患を有すると診断されているまたは有することが疑われ得る、妊娠中であるまたは妊娠を計画している女性個体であり得る。

【０１７５】

標的領域：本明細書で使用される場合、「標的領域」とは、例えばプローブを使用することによる（例えば、配列相補性による）同定および／または捕捉の標的とされるゲノム座を指す。「標的領域セット」または「標的領域のセット」とは、例えばプローブのセットを使用することによる（例えば、配列相補性による）同定および／または捕捉の標的とされる複数のゲノム座を指す。

【０１７６】

閾値：本明細書で使用される場合、「閾値」とは、異なる試料の同じパラメーターに関して実験的に決定された値を、それらの閾値との関連に応じて特徴付けるために使用される所定の値を指す。

【０１７７】

腫瘍画分：本明細書で使用される場合、「腫瘍画分」とは、所与の試料中の腫瘍に由来する核酸分子の画分の推定値を指す。例えば、試料の腫瘍画分は、試料の最大（ｍａｘ）ＭＡＦまたは試料の配列決定カバレッジのパターンまたは試料中のｃｆＤＮＡ断片の長さまたは試料の任意の他の選択された特色から導き出される尺度であり得る。一部の場合では、試料の腫瘍画分は、試料の最大（ｍａｘ）ＭＡＦと等しい。

【０１７８】

バリアント：本明細書で使用される場合、「バリアント」は、アレルと称され得る。バリアントは、通常、アレルがヘテロ接合性であるかまたはホモ接合性であるかに応じて５０％（０．５）または１００％（１）の頻度で提示される。例えば、生殖系列バリアントは、遺伝性であり、通常、０．５または１の頻度を有する。しかし、体細胞バリアントは、後天的バリアントであり、通常、＜０．５の頻度を有する。遺伝子座（genetic locus）のメジャーアレルおよびマイナーアレルとは、参照配列のヌクレオチドによって座が占有されている座を有する核酸を指し、バリアントヌクレオチドはそれぞれ参照配列と異なる。座における測定値は、試料においてアレルが観察される頻度を測定するアレルの画分（ＡＦ）の形態で取得することができる。

（発明を実施するための形態）

【０１７９】

詳細な説明

がんは通常、個体の細胞の遺伝子内の変異の蓄積によって引き起こされ、その少なくとも一部は、不適切に調節された細胞分裂をもたらす。そのような変異は、一塩基バリエーション（ＳＮＶ）、遺伝子融合、挿入、トランスバージョン、転座、および逆位を含み得る。これらの変異は、個体の非がん性細胞と比べた、腫瘍ゲノム内の遺伝子のコピー数の増加または減少に対応するコピー数バリエーションも含み得る。無細胞核酸に存在する変異の程度および試料の変異した無細胞核酸の量を、腫瘍進行を決定する、患者転帰を予測する、および処置の選択を改善するためのバイオマーカーとして使用することができる。様々な例では、無細胞核酸に存在する変異の程度は、所与の試料についての腫瘍細胞コピー数および腫瘍画分によって示され得る。

【０１８０】

さらに、がんは、メチル化などの非配列改変によって示され得る。がんにおけるメチル化変化の例としては、正常な成長の制御、ＤＮＡ修復、細胞周期調節、および／または細

胞分化に関係する遺伝子のＴＳＳでＣｐＧアイランドにおけるＤＮＡメチル化の局所増加が挙げられる。このメチル化量の増加は、関係する遺伝子の転写能の異常な喪失に関連し得、変更された遺伝子発現の原因としての点突然変異および欠失と少なくとも同程度の頻度で起こる。

【０１８１】

したがって、ＤＮＡメチル化プロファイリングを使用して、試料のＤＮＡにおける異常なメチル化を検出することができる。ＤＮＡは、所与の試料タイプ（例えば、血流に由来するｃｆＤＮＡ）において通常高メチル化または低メチル化されているが、例えば、その試料タイプへの組織の寄与が異常に増加しているために（例えば、新生物もしくはがんにおけるまたはその周囲のＤＮＡの脱落の増加に起因して）、および／または発達中に変更されるか、もしくは疾患、例えば、がんもしくは任意のがん関連疾患によって乱れる、ゲノムのメチル化の程度から、新生物またはがんと相関する異常なメチル化の程度を示し得る、ある特定のゲノム領域（「差次的にメチル化された領域」または「ＤＭＲ」）に対応し得る。

【０１８２】

ＤＮＡメチル化を測定する一部の方法では、ＤＮＡのメチル化の量を正確に決定することが難しくなる場合がある。ＤＮＡメチル化が決定される正確度は、試料についての腫瘍画分の推定値の正確度に影響を及ぼし得る。腫瘍画分は、試料が、腫瘍が存在する対象に由来するか否かを決定するために使用され得るので、腫瘍画分推定値の決定の正確度は、個体についての診断および／または処置の決定に影響を及ぼし得る。

【０１８３】

本明細書に記載される方法およびシステムは、核酸のメチル結合ドメイン（ＭＢＤ）への結合の量を示すデータを使用して、核酸のメチル化の複数の量を示す情報を正確に生成することを対象とする。様々な例では、本出願は、試料の腫瘍画分の推定値を決定するためのシステムおよびプロセスを対象とする。１つまたは複数の例では、核酸のメチル化の量を、核酸によるメチル結合ドメイン（ＭＢＤ）への結合の強度に基づいて決定することができる。核酸をＭＢＤへの結合の強度に従って分配することができる。さらに、核酸についてシトシン－グアニン（ＣＧ）領域の数を決定することができる。核酸の分類領域のメチル化の量を、核酸に関連する分配情報および核酸のシトシン－グアニン領域の数に基づいて決定することができる。分類領域は、腫瘍細胞と非腫瘍細胞とで異なる量のメチル化を有し得る。試料の腫瘍画分の推定値は、分類領域のメチル化の量に従って決定することができる。

【０１８４】

少なくとも一部のインプリメンテーションでは、方法、システム、技術、およびアーキテクチャにより、モデルに提供されたメチル化データに、より正確に適合するように修正することができるパラメーターまたは重みの少なくとも１つを有するように構成されたモデルをインプリメントすることができる。方法、システム、技術、およびアーキテクチャはまた、モデルの訓練中にいくつかの最適化手順をインプリメントして、腫瘍の存在または非存在を示すメトリクスを他のシステム、方法、技術、およびアーキテクチャよりも正確に予測するモデルを生成することも対象とする。さらに、メチル化データを産生するために使用される情報を生成するために使用される方法、技術、およびプロセスでは、メチル化データに存在するノイズの量が減少し、それによって、腫瘍の存在または非存在を示すメトリクスの、他の方法、技術、およびプロセスよりも正確な予測がもたらされる。

【０１８５】

図１は、参照配列の分類領域に対応する核酸を同定する環境１００の例の図式表示であり、ここで、分類領域は、１つまたは複数のインプリメンテーションによる少なくとも閾値数のＣｐＧを有する。１つまたは複数の例では、検討される疾患はあるタイプのがんである。そのようながんの非限定的な例としては、胆管がん、膀胱がん、移行上皮癌、尿路上皮癌、脳のがん、神経膠腫、星細胞腫、乳癌、化生癌、子宮頸がん、子宮頸部扁平上皮癌、直腸がん、結腸直腸癌、結腸がん、遺伝性非ポリポーシス結腸直腸がん、結腸直腸腺

癌、消化管間質腫瘍（ＧＩＳＴ）、子宮内膜癌、子宮内膜間質肉腫、食道がん、食道扁平上皮癌、食道腺癌、眼内黒色腫、ぶどう膜黒色腫、胆嚢癌、胆嚢腺癌、腎細胞癌、明細胞腎細胞癌、移行上皮癌、尿路上皮癌、ウィルムス腫瘍、白血病、急性リンパ球性白血病（ＡＬＬ）、急性骨髄性白血病（ＡＭＬ）、慢性リンパ球性白血病（ＣＬＬ）、慢性骨髄性白血病（ＣＭＬ）、慢性骨髄単球性白血病（ＣＭＭＬ）、肝臓がん、肝臓癌、ヘパトーマ、肝細胞癌、胆管癌、肝芽腫、肺がん、非小細胞肺がん（ＮＳＣＬＣ）、中皮腫、Ｂ細胞リンパ腫、非ホジキンリンパ腫、びまん性大細胞型Ｂ細胞リンパ腫、マントル細胞リンパ腫、Ｔ細胞リンパ腫、非ホジキンリンパ腫、前駆Ｔリンパ芽球性リンパ腫／白血病、末梢性Ｔ細胞リンパ腫、多発性骨髄腫、鼻咽頭癌（ＮＰＣ）、神経芽腫、中咽頭がん、口腔扁平上皮癌、骨肉腫、卵巣癌、膵臓がん、膵管腺癌、偽乳頭状腫瘍、腺房細胞癌、前立腺がん、前立腺腺癌、皮膚がん、黒色腫、悪性黒色腫、皮膚黒色腫、小腸癌、胃がん（stomach cancer）、胃癌（gastric carcinoma）、消化管間質腫瘍（ＧＩＳＴ）、子宮がん、または子宮肉腫が挙げられる。

【０１８６】

環境１００は、試料１０２を含み得る。試料１０２は、対象から得られた生体液に由来し得る。例えば、試料１０２は、対象から得られた血液に由来し得る。１つまたは複数の追加の例では、試料１０２は、対象の組織に由来し得る。様々な例では、試料１０２は、複数の供給源に由来し得る。例として、試料１０２は、対象の１つもしくは複数の体液および／または対象の組織に由来し得る。１つまたは複数の実例では、対象は哺乳動物であり得る。１つまたは複数の追加の実例では、対象はヒトであり得る。１つまたは複数のさらなる実例では、対象は非ヒト哺乳動物であり得る。

【０１８７】

試料１０２は、いくつかの核酸１０４を含み得る。個々の核酸１０４は、少なくとも閾値数のシトシン分子およびグアニン分子を有するいくつかの領域を含み得る。１つまたは複数の例では、個々の核酸１０４は、少なくとも閾値数のシトシン－グアニンジヌクレオチドを有する領域を含み得る。様々な例では、領域内に含まれるシトシン－グアニン対の少なくとも一部分は、核酸１０４の配列内で順次位置決めされ得る。１つまたは複数の実例では、少なくとも閾値量のシトシン－グアニン対を有する核酸の領域は、本明細書では、「ＣＧ領域」または「ＣｐＧ領域」と称され得る。１つまたは複数の例では、ＣＧ領域は、少なくとも２００個のＣｐＧジヌクレオチドを含み得る。１つまたは複数の実例では、ＣＧ領域は、２００個のＣｐＧジヌクレオチド～５０００個のＣｐＧジヌクレオチド、３００個のＣｐＧジヌクレオチド～３０００個のＣｐＧジヌクレオチド、２００個のＣｐＧジヌクレオチド～２５００個のＣｐＧジヌクレオチド、または５００個のＣｐＧジヌクレオチド～１５００個のＣｐＧジヌクレオチドを含み得る。さらに、ＣＧ領域は、少なくとも５０％のＧＣパーセンテージおよび少なくとも６０％の予測されるＣｐＧ対する観察されるＣｐＧの比を有し得る。予測されるＣｐＧに対する観察されるＣｐＧの比は、観察されるＣｐＧを所与のゲノム領域内で同定されたＣｐＧの数とし、予測されるＣｐＧを、ゲノム領域内のシトシンの数にグアニンの数を掛けて塩基の数で割ったものとして計算することができる。予測されるＣｐＧは、

（（シトシンの数＋グアニンの数）／２）^{2}／ゲノム領域の長さ

によって計算することもできる。例えば、ＣＧ領域は、Gardiner-Garden M, Frommer M (1987). "CpG islands in vertebrate genomes". Journal of Molecular Biology. 196 (2): 261-282. および／またはSaxonov S, Berg P, Brutlag DL (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters". Proc Natl Acad Sci USA. 103 (5): 1412-1417によって記載されている技術を使用して決定することができる。

【０１８８】

図１の実例では、例としての核酸１０４の配列の一部分は、第１のＣＧ領域１０６、第２のＣＧ領域１０８、および第３のＣＧ領域１１０を含み得る。図１の実例では、３つの

ＣＧ領域を有する核酸１０４の配列の一部分が例示されているが、試料１０２に含まれる核酸１０４は、異なる数のＣＧ領域を有し得る。例えば、試料１０２に含まれる個々の核酸１０４は、少なくとも１個のＣＧ領域、少なくとも５個のＣＧ領域、少なくとも１０個のＣＧ領域、少なくとも２５個のＣＧ領域、少なくとも５０個のＣＧ領域、少なくとも１００個のＣＧ領域、少なくとも２５０個のＣＧ領域、少なくとも５００個のＣＧ領域、または少なくとも１０００個のＣＧ領域を含み得る。

【０１８９】

個々のＣＧ領域は、１つまたは複数のメチル化シトシンを有するいくつかの分子に対応し得る。図１の実例では、ＣＧ領域１０６は、メチル化シトシン１１２を有する分子を含み得る。図１の実例では、メチル化シトシン１１２を有する分子は、５－メチルシトシンである。個々のＣＧ領域は、非メチル化シトシンを有するいくつかの分子にも対応し得る。例えば、ＣＧ領域１０６は、非メチル化シトシン１１６を有する分子を含み得る。様々な例では、核酸１０４のＣＧ領域の少なくとも一部分は、参照ゲノムの分類領域に対応し得る。分類領域は、１つまたは複数の生物学的状態、例えば１つまたは複数のタイプのがんと一致する非配列相違に対応する参照ゲノムのゲノム領域に対応し得る。少なくとも一部の例では、非配列相違は、１つまたは複数の生物学的状態と一致する１つまたは複数の変異を含み得る。１つまたは複数の例では、分類領域は、少なくとも１つの形態のがんを有する対象に由来する分子に対する参照配列のゲノム領域に対応し得る。少なくとも一部の例では、少なくとも１つのＣＧ領域内に少なくとも閾値量のメチル化シトシンを有する核酸分子（例えば、高メチル化分子）は、がんが存在する対象に由来し得、分類に対応し得る。１つまたは複数の追加の例では、少なくとも１つのＣＧ領域内に閾値量未満のメチル化シトシンを有する核酸分子（例えば、低メチル化分子）は、がんが存在する対象に由来し得、分類領域に対応し得る。

【０１９０】

分類領域に加えて、ＣＧ領域は、１つまたは複数の陽性対照領域、例えば陽性対照領域１１８を含み得る。陽性対照領域１０８は、少なくとも１つのＣＧ領域内に少なくとも閾値数のメチル化シトシン分子を有し、がんのない対象に由来する、およびがんが存在する対象に由来する核酸分子にマッピングされ得る。様々な例では、陽性対照領域１０６は、がんのない対象に由来する細胞において、およびがんが存在する対象に由来する細胞においても、高メチル化され得る。ＣＧ領域は、１つまたは複数の陰性対照領域、例えば陰性対照領域１２０も含み得る。陰性対照領域１２０は、少なくとも１つのＣＧ領域内に閾値数未満のメチル化シトシン分子を有し、がんのない対象に由来し、がんが存在する対象にも由来する核酸分子にマッピングされ得る。１つまたは複数の実例では、陰性対照領域１２０は、がんのない対象において、およびがんが存在する対象においても低メチル化され得る。様々な例では、陽性対照領域および陰性対照領域を使用して、正規化計算を実施することができる。正規化計算を実施して、所与の試料１０２についての腫瘍メトリクスを決定するためにインプリメントされる１つまたは複数のモデルに対する入力データを生成することができる。

【０１９１】

第１の分子分離プロセス１２２を実施することができる。第１の分子分離プロセス１２２により、試料１０２に含まれる核酸１０４を個々の核酸１０４のメチル化シトシンの量に基づいて分離することができる。１つまたは複数の例では、第１の分子分離プロセスにより、試料１０２に含まれる核酸１０４を個々の核酸１０４のＣＧ領域に含まれるメチル化シトシンの量に基づいて分離することができる。様々な例では、第１の分子分離プロセス１２２により、核酸１０４を複数の群に分離することができ、個々の群は、核酸１０４のメチル化シトシンのそれぞれの量に対応する。

【０１９２】

図１の実例では、第１の分子分離プロセス１２２を第１のメチル化閾値１２４に関連して実施することができる。第１の分子分離プロセス１２２を第１のメチル化閾値１２４に関して実施することにより、核酸の第１の分配１２６を生じさせることができる。１つま

たは複数の例では、第１のメチル化閾値１２４は、核酸１０４のＣＧ領域内で位置を特定されたメチル化シトシンを有する分子の第１の閾値数を示し得る。第１の分子分離プロセス１２２により、ＣＧ領域内のメチル化シトシンを有する分子が第１のメチル化閾値１２４よりも少ないいくつかの核酸１０４を同定することができる。様々な例では、第１のメチル化閾値１２４は、第１のメチル化率に対応し得る。

【０１９３】

第１の分子分離プロセス１２２を第２のメチル化閾値１２８に関しても実施することができる。第２のメチル化閾値１２８は、第１のメチル化閾値１２４に対応する１つまたは複数の領域内のメチル化シトシンの量よりも多い、核酸１０４の１つまたは複数のゲノム領域におけるメチル化シトシンの量を示し得る。第２のメチル化閾値１２４は、核酸の数に応じて、いくつかの、メチル化シトシンを有する分子を示し得る。１つまたは複数の追加の例では、第２のメチル化閾値１２４は、第１のメチル化閾値１２４に対応するメチル化の率よりも高い核酸のメチル化の率に対応し得る。第１の分子分離プロセス１２２を第２のメチル化閾値１２８に関して実施することにより、核酸の第２の分配１３０を生じさせることができる。１つまたは複数の例では、第１の分子分離プロセス１２２により、メチル化シトシンの量が第１のメチル化閾値１２４よりも多く、メチル化シトシンの量が第２のメチル化閾値１２８よりも少ない核酸１０４を同定して、核酸の第２の分配１３０を生じさせることができる。

【０１９４】

さらに、第１の分子分離プロセス１２２を第３のメチル化閾値１３２に関しても実施することできる。第３のメチル化閾値１３２は、第１のメチル化閾値１２４に対応する１つまたは複数の領域内のメチル化シトシンの量よりも多く、かつ第２のメチル化閾値１２８に対応する１つまたは複数の領域内のメチル化シトシンの量よりも多い、核酸１０４の１つまたは複数のゲノム領域におけるメチル化シトシンの量を示し得る。第３のメチル化閾値１３２は、核酸の数に応じて、いくつかの、メチル化シトシンを有する分子を示し得る。１つまたは複数の追加の例では、第３のメチル化閾値１３２は、第１のメチル化閾値１２４に対応するメチル化の率よりも高く、かつ第２のメチル化閾値１２８に対応するメチル化の率よりも高いメチル化シトシンの率に対応し得る。第１の分子分離プロセス１２２を第３のメチル化閾値１３２に関して実施することにより、核酸の第３の分配１３４を生じさせることができる。１つまたは複数の例では、第１の分子分離プロセス１２２により、核酸の第２の分配１２８に含まれる核酸１０４よりもメチル化シトシンの量が多い核酸１０４を同定することができる。このように、第１の分配１２２、第２の分配１２６、および第３の分配１３０に含まれる核酸のメチル化シトシンの量は、第１の分配１２２から第２の分配１２６までで増加し、第２の分配１２６から第３の分配１３０までで増加する。１つまたは複数の実例では、核酸の第１の分配１２６は低メチル化分配と称され得、核酸の第２の分配１３０は中間分配と称され得、核酸の第３の分配１３４は高メチル化分配と称され得る。

【０１９５】

１つまたは複数の例では、核酸のメチル化シトシンの量は、メチル結合ドメイン（ＭＢＤ）への結合の強度に対応し得る。これらのシナリオでは、第１の分配１２６、第２の分配１３０、および第３の分配１３４を、異なるメチル化シトシンの量を有するヌクレオチドについてのＭＢＤへの異なる結合の強度に基づいて生じさせることができる。１つまたは複数の例では、第１の分子分離プロセス１２２は、核酸１０４を異なる濃度の塩化ナトリウム（ＮａＣｌ）を有する溶液と接触させる一連の洗浄を含み得る。

【０１９６】

核酸の分配は、核酸を改変ヌクレオチド特異的結合試薬、例えばＭＢＰのＭＢＤと接触させることによって実施することができる。改変ヌクレオチド特異的結合試薬は、５－メチルシトシン（５ｍＣ）に結合し得る。改変ヌクレオチド特異的結合試薬、例えばＭＢＤを、Ｄｙｎａｂｅａｄｓ（登録商標）Ｍ－２８０ストレプトアビジンなどの常磁性ビーズにビオチンリンカーを介してカップリングすることができる。異なる程度のメチル化を有

する画分への分配を、一連の洗浄におけるＮａＣｌ濃度を上昇させることによって実施することができる。改変ヌクレオチド特異的結合試薬から溶離した配列は、いずれの洗浄（例えば、ＮａＣｌ濃度）で配列が溶離したかに応じて２つまたはそれよりも多くの画分（例えば、低、高）に分配される。得られた分配は、以下の核酸形態：二本鎖ＤＮＡ（ｄｓＤＮＡ）、より短いＤＮＡ断片およびより長いＤＮＡ断片の１つまたは複数を含み得る。

【０１９７】

改変ヌクレオチド特異的結合試薬を用いた核酸の結合は、分子あたりのメチル化された（または改変された）部位の数の関数であり得、より多くのメチル化を有する分子は塩濃度を増加させることにより溶離する。ＤＮＡをメチル化の程度に基づいて別個の集団に溶離させるために、ＮａＣｌ濃度を増加させる一連の溶離緩衝液を使用することができる。塩濃度は、１つまたは複数のインプリメンテーションでは、約１００ｎＭ～約２５００ｍＭのＮａＣｌの範囲である。様々なインプリメンテーションでは、プロセスにより、３つの分配がもたらされる。分子を、第１の塩濃度の、メチル結合ドメインを含む分子を含む溶液と接触させ、分子を捕捉部分、例えばストレプトアビジンに結合させてもよい。第１の塩濃度において、分子のある１つの集団はＭＢＤに結合し、ある１つの集団は結合しないままである。結合しなかった集団を「低メチル化」集団（低分配）として分離することができる。例えば、第１の分配１２６は、低塩濃度において結合しないままである、ＤＮＡの低メチル化形態を表し得る。１つまたは複数の実例では、第１の分配１２６を生じさせるために使用される溶液のＮａＣｌの濃度は、約１００ｎＭ、約１２０ｎＭ、約１４０ｎＭ、約１６０ｎＭ、約１８０ｎＭ、約２００ｎＭまたは約２５０ｎＭであり得る。第２の分配１３０は、「残留分配」または「中間分配」と称され得、中間の塩濃度、例えば、１００ｍＭ～２０００ｍＭの間の濃度を使用して溶離される中間メチル化ＤＮＡを表し得る。１つまたは複数の追加の実例では、第２の分配１３０を生じさせるために使用される溶液のＮａＣｌの濃度は、約１００ｍＭ～約５００ｍＭ、約１００ｍＭ～約１０００ｍＭ、約１００ｍＭ～約１５００ｍＭ、約２５０ｍＭ～約１０００ｍＭ、約２５０ｍＭ～約１５００ｍＭ、約５００ｍＭ～約１５００ｍＭ、約２５０ｍＭ～約２０００ｍＭ、約５００ｍＭ～約２０００ｍＭ、または約１０００ｍＭ～約２０００ｍＭであり得る。これも試料から分離することができる。第３の分配１３４は、高メチル化形態のＤＮＡ（高分配）を表し得、高塩濃度、例えば、少なくとも約２０００ｍＭを使用して溶離される。１つまたは複数のさらなる実例では、第３の分配１３４を生じさせるために使用される溶液のＮａＣｌの濃度は、約２０００ｍＭ～約５０００ｍＭ、約２０００ｍＭ～約４０００ｍＭ、約２０００ｍＭ～約３５００ｍＭ、約２０００ｍＭ～約３０００ｍＭ、または約２５００ｍＭ～約４０００ｍＭであり得る。

【０１９８】

様々な例では、第１の分配１２６は、核酸のＭＢＤへの結合強度の第１の範囲およびメチル化ＣＧ領域の第１の範囲に対応し得、第２の分配１３０は、核酸のＭＢＤへの結合強度の第２の範囲およびメチル化ＣＧ領域の第２の範囲に対応し得る。結合強度の第１の範囲は、結合強度の第２の範囲未満であってよい。１つまたは複数のシナリオでは、第１のＮａＣｌ濃度を有する第１の溶液により、結合強度の第１の範囲を有する核酸の第１の群をＭＢＤから分離することができ、第２のＮａＣｌ濃度を有する第２の溶液により、第１のＮａＣｌ濃度よりも高い第２のＮａＣｌ濃度で、結合強度の第２の範囲を有する核酸の第２の群をＭＢＤから分離することができる。さらに、第３の分配１３４は、結合強度の第３の範囲およびメチル化ＣＧ領域の第３の範囲に対応し得る。結合強度の第３の範囲は、結合強度の第１の範囲および結合強度の第２の範囲よりも大きくてよい。１つまたは複数の場合では、第３のＮａＣｌ濃度を有する第３の溶液により、結合強度の第３の範囲を有する核酸の第３の群をＮａＣｌから分離することができる。第３のＮａＣｌ濃度は、第１のＮａＣｌ濃度および第２のＮａＣｌ濃度よりも高くてよい。

【０１９９】

１つまたは複数の実例では、対象の血液または組織の少なくとも一方に由来する複数の核酸を、ある１つの量のＭＢＤを含む溶液と組み合わせて、核酸－ＭＢＤ溶液を産生する

ことができる。第１のＮａＣｌ濃度を含む第１の溶液を用いて核酸－ＭＢＤ溶液の第１の洗浄を実施して、第１の核酸画分および第１の残留溶液を産生することができる。第１の核酸画分は、複数の核酸のうちの第１の一部分を含み得、第１の残留溶液は、複数の核酸のうちの第２の一部分を含み得る。１つまたは複数の例では、複数の核酸のうちの第１の一部分は、複数の核酸のうちの第２の一部分のＭＢＤへの結合強度の第２の範囲未満である、ＭＢＤへの結合強度の第１の範囲を有し得る。

【０２００】

さらに、第１のＮａＣｌの濃度よりも高い第２のＮａＣｌの濃度を含む第２の溶液を用いて第１の残留溶液の第２の洗浄を実施して、第２の核酸画分および第２の残留溶液を産生することができる。第２の核酸画分は、複数の核酸のうちの第２の一部分の第１のサブセットを含み得、第２の残留溶液は、複数の核酸のうちの第２の一部分の第２のサブセットを含み得る。複数の核酸のうちの第２の一部分の第１のサブセットは、複数の核酸のうちの第２の一部分の第２のサブセットのＭＢＤへの結合強度の第４の範囲未満である、ＭＢＤへの結合強度の第３の範囲を有し得る。さらに、第２のＮａＣｌの濃度よりも高い第３のＮａＣｌの濃度を含む第３の溶液を用いて第２の残留溶液の第３の洗浄を実施して、複数の核酸のうちの第２の一部分の第２のサブセットを含む第３の核酸画分を産生することができる。

【０２０１】

第１の洗浄、第２の洗浄、および第３の洗浄の後、複数の核酸のうちの第１の一部分が第１の分配１２６に関連付けられることの決定を行うことができる。複数の核酸のうちの第１の一部分に、第１の分配１２６を示す分子バーコードの第１のセット由来の分子バーコードを結合させることができる。このように、第１の分配１２６に対応する配列決定リードを、配列決定リードが第１の分子バーコードを含むことを決定することに基づいて同定することができる。さらに、複数の核酸のうちの第２の一部分の第１のサブセットが複数の分配のうちの追加の分配に関連付けられることの決定を行うことができる。これらの状況では、分子バーコードの第１のセットとは異なる分子バーコードの第２のセットを複数の核酸のうちの第２の一部分に結合させることができ、第２の分子バーコードは追加の分配を示す。結果として、追加の分配に対応する配列決定リードを、配列決定リードが分子バーコードの第２のセットの中からの１つまたは複数の分子バーコードを含むことを決定することに基づいて同定することができる。さらに、複数の核酸のうちの第２の一部分の第２のサブセットが第２の分配１３０に関連付けられることの決定を行うことができる。次いで、分子バーコードの第１のセットおよび分子バーコードの第２のセットとは異なる分子バーコードの第３のセットを複数の核酸のうちの第２の一部分の第２のサブセットに結合させることができ、ここで、分子バーコードの第３のセットは第２の分配１３０を示す。これらの場合では、第２の分配１３０に対応する配列決定リードを、配列決定リードが分子バーコードの第３のセットの中からの第３の分子バーコードを含むことを決定することに基づいて同定することができる。

【０２０２】

少なくとも一部の例では、第１の分子分離プロセス１２２により、第１の分配１２６、第２の分配１３０、または第３の分配１３４のうちの少なくとも１つに存在する、それぞれの分配中の他の核酸のメチル化の量とは異なる量のメチル化を有する核酸がもたらされ得る。例えば、第１の分配１２６は、第２の分配１３０または第３の分配１３４のうちの少なくとも１つに含まれる核酸のメチル化の量に対応するメチル化の量を有するいくつかの核酸を含み得る。さらに、第２の分配１３０または第３の分配１３４のうちの少なくとも１つは、第１の分配１２６に含まれる核酸のメチル化の量に対応するメチル化の量を有する核酸を含み得る。第１の分配１２６、第２の分配１３０、または第３の分配１３４のうちの少なくとも１つにおける、それぞれの分配中に含まれる他の核酸の少なくとも大多数のメチル化の量に対応しない核酸の存在は、第１の分配１２６、第２の分配１３０、および第３の分配１３４に含まれる核酸から産生される配列リードに関してコンピュータによる操作を実施する際にデータノイズを引き起こし得る。データノイズは、第１の分配１

２６、第２の分配１３０、および第３の分配１３４に含まれる核酸に由来する配列リードに基づいてなされた計算に関して不正確さをもたらし得る。

【０２０３】

第１の分配１２６、第２の分配１３０、または第３の分配１３４のうちの少なくとも１つに存在する、それぞれの分配中に含まれる他の分子の少なくとも大多数のメチル化の量と一致しないメチル化の量を有する核酸に関連するデータノイズを減少させるまたは排除するために、第１の分子分離プロセス１２２の後に第２の分子分離プロセス１３６を実施することができる。第２の分子分離プロセス１３６は、第１の分配１２６に含まれる核酸、第２の分配１３０に含まれる核酸、および第３の分配１３４に含まれる核酸に関して実施することができる。１つまたは複数の例では、第２の分子分離プロセス１３６は、メチル化依存性制限酵素（ＭＤＲＥ）を使用した第１の分配１２６中に含まれる核酸の消化を実施するステップを含んでよく、第２の分配１３０および第３の分配１３４中に含まれる核酸は、メチル化感受性制限酵素（ＭＳＲＥ）を使用して消化することができる。第１の分配１２６中に含まれる核酸のＭＤＲＥを用いた消化により、第１の分配中に含まれる、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有する核酸の、第１の分配に対応するメチル化の量を有する核酸からの分離をもたらすことができる。さらに、第２の分配１３０および第３の分配１３４中に含まれる核酸のＭＳＲＥを用いた消化により、第１の分配１２６に対応するメチル化の量を有する核酸の、第２の分配１３０の核酸および第３の分配１３４の核酸からの分離をもたらすことができる。第１の分配１２６から第２の分配１３０および第３の分配１３４に対応するメチル化の量を有する核酸を除去することにより、かつ第２の分配１３０および第３の分配１３４から第１の分配１２６に対応するメチル化の量を有する核酸を除去することにより、核酸の追加の群１３８を産生することができる。核酸の追加の群１３８は、第２の分配１３０および第３の分配１３４のメチル化量に対応し、第１の分配１２６に対応するメチル化の量を有する核酸を最小量で有するかまたは有さない、核酸を含み得る。例えば、追加の群１３８に含まれる核酸の５０％未満が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも５０％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも６０％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも７０％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも９０％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも９５％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも９７％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも９９％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、追加の群１３８に含まれる核酸の少なくとも９９．５％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得、または追加の群１３８に含まれる核酸の少なくとも９９．９％が、第２の分配１３０および第３の分配１３４に対応するメチル化の量を有し得る。

【０２０４】

アーキテクチャ１００は、配列決定機械１４０を含み得る。１つまたは複数の例では、配列決定機械１４０は、試料１０４中に存在する核酸を増幅させる１つまたは複数の配列決定操作を実施することができるいくつかの配列決定機械のいずれかであり得る。様々な例では、配列決定機械１４０は、次世代配列決定操作を実施することができる。１つまたは複数の例では、試料１０４は、対象から抽出されたある１つの量の少なくとも１つの体液を含み得る。１つまたは複数の追加の例では、試料１０４は、対象から得られた組織試料を含み得る。

【０２０５】

１つまたは複数の例では、配列決定の前に、抽出されたポリヌクレオチドを、ポリヌク

レオチドのＭＢＤへの結合強度に基づいて（based on the binding strength of the of binding strengths of polynucleotides to MBD）２つまたはそれよりも多くの分配に分配することができる。分配されたポリヌクレオチドおよびアダプターに対して平滑末端ライゲーションを実施することができるとともに、分配されたポリヌクレオチドにタグ（例えば、分子バーコード）を付加することができる。１つまたは複数の分配（例えば、高分配および／または中間分配）中のタグ付けされたポリヌクレオチドを１つまたは複数のメチル化感受性制限酵素（ＭＳＲＥ）で処理することができる。一部の例では、低分配を１つまたは複数のメチル化依存性制限酵素（ＭＤＲＥ）で処理することができる。ＭＳＲＥおよび／またはＭＤＲＥ処理後、抽出されたポリヌクレオチドと、参照配列の標的領域に対応するプローブとのハイブリダイゼーションを引き起こすことによって分子を富化することもできる。富化プロセスにより、プローブに関連するオンターゲット領域に対応する数千個、数十万個、最大数百万個のポリヌクレオチドを同定することができる。

【０２０６】

富化プロセスの後および／または前に、分子を１つまたは複数の増幅プロセスに従って増幅させることができる。１つまたは複数の増幅プロセスにより、数千個、最大数百万個の個々の核酸分子のコピーを産生することができる。１つまたは複数の例では、富化されていないポリヌクレオチドの一部分を増幅させることができるが、一部の場合では、富化されたポリヌクレオチドが増幅される程度までは増幅されない。１つまたは複数の増幅プロセスにより、１つまたは複数の配列決定操作を受ける増幅産物を生成することができる。試料１０４に関して１つまたは複数の配列決定操作を実施した後、配列決定機械１４０により、配列決定データ１４２を産生することができる。

【０２０７】

配列決定データ１４２は、増幅産物に含まれる核酸の英数字による表示を含み得る。例えば、配列決定データ１４２は、増幅産物の個々の核酸について、個々の核酸に対応するヌクレオチドのそれぞれの鎖を表す文字列に対応するデータを含み得る。

【０２０８】

配列決定データ１４２は、１つまたは複数のデータファイルに記憶させることができる。例えば、配列決定データ１４２は、生の配列データおよび品質スコアを記憶するテキストに基づく配列決定データファイルフォーマットを含むＦＡＳＴＱファイルに記憶させることができる。１つまたは複数の追加の例では、配列決定データ１４２は、バイナリベースコール（ＢＣＬ）配列ファイルフォーマットに従ったデータファイルに記憶させることができる。１つまたは複数のさらなる例では、配列決定データ１４２は、ＢＡＭファイルに記憶させることができる。１つまたは複数の例では、配列決定データ１４２は、少なくとも約１ギガバイト（ＧＢ）、少なくとも約２ＧＢ、少なくとも約３ＧＢ、少なくとも約４ＧＢ、少なくとも約５ＧＢ、少なくとも約８ＧＢ、または少なくとも約１０ＧＢを含み得る。配列決定データ１０６に含まれる個々の配列表示は、本明細書では、「リード」または「配列決定リード」と称され得る。様々な例では、プール１３８に含まれる個々の第１の核酸は、個々の第１の核酸が増幅された結果として配列決定データ１４２に含まれる複数の配列表示に対応し得る。１つまたは複数の追加の例では、プール１３８に含まれる個々の第２の核酸は、個々の第２の核酸が増幅されなかった結果として配列決定データ１４２に含まれる単一の配列表示に対応し得る。

【０２０９】

図２は、１つまたは複数のインプリメンテーションによる、配列決定データを解析して、対象における腫瘍の存在を示す１つまたは複数のメトリクスを決定するためのアーキテクチャ２００の例である。アーキテクチャ２００は、いくつかの試料２０４に関して１つまたは複数の配列決定操作を実施する１つまたは複数の配列決定機械２０２を含み得る。１つまたは複数の試料２０４は対象２０６から得ることができる。１つまたは複数の実例では、対象２０６の第１の一部分はがんを有さなくてよい。すなわち、対象２０６の第１の一部分において、腫瘍は検出されない。さらに、腫瘍は、対象２０６の第２の一部分に

存在してよい。

【０２１０】

１つまたは複数の分子分離プロセス２０８を試料２０４に関して実施することができる。１つまたは複数の分離プロセス２０８は、核酸分子の特徴に基づいて核酸分子をいくつかの分配に分離することに対応し得る。核酸分子を分配するために使用されることができる特徴の例としては、複数の異なるヌクレオチド改変、メチル化レベル、ヌクレオソーム結合、配列ミスマッチ、免疫沈降、および／またはＤＮＡに結合するタンパク質が挙げられる。１つまたは複数の実例では、核酸分子の不均一な集団を、１つまたは複数のエピジェネティック改変を有する核酸分子と１つまたは複数のエピジェネティック改変を有さない核酸分子とに分配することができる。エピジェネティック改変の例としては、メチル化の存在または非存在；メチル化のレベル、ヒドロキシメチル化、およびメチル化のタイプ（５’シトシンまたは６メチルアデニン）が挙げられるが、これらに限定されない。

【０２１１】

１つまたは複数の分子分離プロセス２０８の前に、核酸分子を試料２０４から抽出することができる。１つまたは複数のインプリメンテーションでは、核酸分子は、無細胞核酸（例えば、無細胞ＤＮＡ）を含む。様々なインプリメンテーションでは、試料２０４は、血液、血漿、血清、尿、糞便、唾液試料、これらの組合せなどの１つまたは複数から選択される試料であってよい。１つまたは複数の追加の例では、試料２０４は、全血、血液分画物、組織生検、胸膜液、心膜液、脳脊髄液、および腹腔液の１つまたは複数から選択される試料を含み得る。１つまたは複数の実例では、無細胞核酸分子を試料２０４から抽出することができ、ここで、試料２０４はがんを有することが分かっている対象２０６（例えば、がん患者）、またはがんを有することが疑われる対象２０６から得られる。

【０２１２】

試料２０４からの核酸分子の抽出は、１つまたは複数の細胞溶解技術をインプリメントして、試料２０４に含まれる細胞の膜を切断すること、および１つまたは複数のプロテアーゼを適用して、試料２０４に含まれるタンパク質を分解することを含み得る。試料２０４からの核酸分子の抽出は、核酸分子を試料２０４に含まれる他の構成成分から分離するためのいくつかの洗浄および／または溶離技術も含み得る。様々な例では、数千個、最大数百万個、最大数十億個の核酸分子を試料２０４から抽出した後に、１つまたは複数の分離プロセス２０８に供することができる。

【０２１３】

試料２０４から抽出された核酸分子は、様々なメチル化のレベルを有する分子を含み得る。メチル化は、任意の１つまたは複数の複製後改変または転写後改変によって生じ得る。複製後改変としては、これだけに限定されないが、５－メチルシトシン、５－ヒドロキシメチルシトシン、５－ホルミルシトシンおよび５－カルボキシルシトシンを含む、ヌクレオチドシトシンの改変が挙げられる。１つまたは複数の分子分離プロセス２０８により、試料２０４から抽出された核酸分子をいくつかの分配に分離することができ、個々の分配は異なるメチル化のレベルに対応する。例えば、分子分離プロセス２０８により、第１のメチル化のレベルを有する核酸分子の第１の分配、第２のメチル化のレベルを有する核酸分子の第２の分配、および第３のメチル化のレベルを有する核酸分子の第３の分配を生じさせることができる。様々な例では、第２のメチル化のレベルは、第１のメチル化のレベルよりも高くてよく、第３のメチル化のレベルは、第１のメチル化のレベルおよび第２のメチル化のレベルよりも高くてよい。１つまたは複数の実例では、１つまたは複数の分子分離プロセス２０８は、図１の第１の分子分離プロセス１２２および第２の分子分離プロセス１３６を含み得る。

【０２１４】

１つまたは複数の分子分離プロセス２０８により、１つまたは複数の試料２０４から抽出された核酸分子の一部分を含み、１つまたは複数の分子分離プロセス２０８に供されるプール２１０を生じさせることができる。例えば、プール２１０は、第２のメチル化のレベルを有するいくつかの核酸分子および第３のメチル化のレベルを有するいくつかの核酸

分子を含み得る。したがって、プール２１０に含まれる核酸分子は、少なくとも閾値量のメチル化を有し得る。１つまたは複数の実例では、プール２１０に含まれる核酸分子は、核酸分子のＣＧ領域内に少なくとも閾値量のメチル化を有し得る。

【０２１５】

１つまたは複数の配列決定機械２０２により、１つまたは複数の配列決定操作を実施して、プール２１０に対応する配列決定データ２１２を産生することができる。アーキテクチャ２００は、１つまたは複数の配列決定機械２０２から配列決定データ２１２を得、配列決定データ２１２を解析するコンピューティングシステム２１４を含み得る。例えば、コンピューティングシステム２１４により、配列決定データ２１２を解析して、少なくとも１つの試料２０４を提供した対象２０６に腫瘍が存在し得ることを示す１つまたは複数のメトリクスを決定することができる。コンピューティングシステム２１４は、１つまたは複数のコンピューティングデバイス２１６を含み得る。１つまたは複数のコンピューティングデバイス２１６は、１つもしくは複数のデスクトップコンピューティングデバイス、１つもしくは複数のモバイルコンピューティングデバイス、または１つもしくは複数のサーバーコンピューティングデバイスのうちの少なくとも１つを含み得る。様々な例では、１つまたは複数のコンピューティングデバイス２１６の少なくとも一部分は、クラウドコンピューティング環境などのリモートコンピューティング環境に含まれていてもよい。１つまたは複数の例では、コンピューティングシステム２１４および配列決定機械２０２は、単一の組織によって所有、操作、維持、および／または制御されていてよい。１つまたは複数の追加の例では、コンピューティングシステム２１４および配列決定機械２０２は、複数の組織によって所有、操作、維持、および／または制御されていてよい。

【０２１６】

操作２１８において、コンピューティングシステム２１４により、配列決定データ２１２を解析することができる。配列決定データ２１２を解析することは、参照配列の１つまたは複数の分類領域に対応する、配列決定データ２１２に含まれる１つまたは複数の第１の配列表示２２０を決定することを含み得る。１つまたは複数の分類領域は、腫瘍が存在しない対象から得られたｃｆＤＮＡにおける参照配列の同じゲノム領域にマッピングされる分子のある１つの量のメチル化に対して、がんが存在する対象から得られたｃｆＤＮＡにおいてある１つの量のメチル化を有する核酸分子にマッピングされる参照配列のゲノム領域に対応し得る。少なくとも一部の例では、分類領域にマッピングされ、がんが存在する対象に由来する核酸分子に存在するメチル化の量は、分類領域にマッピングされ、がんが存在しない対象に由来する核酸分子に存在するメチル化の量よりも少ない。１つまたは複数の追加の例では、分類領域にマッピングされ、がんが存在する対象に由来する核酸分子に存在するメチル化の量は、分類領域にマッピングされ、がんが存在しない対象に由来する核酸分子に存在するメチル化の量よりも多い。１つまたは複数の分類領域は、少なくとも閾値量のシトシン－グアニン含量も含み得る。様々な例では、１つまたは複数の分類領域は、５’→３’方向に一連のシトシン－グアニン（ＣＧ）対（ＣｐＧ部位）、例えば、少なくとも３個のＣｐＧ部位、少なくとも５個のＣｐＧ部位、少なくとも８個のＣｐＧ部位、少なくとも１０個のＣｐＧ部位、少なくとも１２個のＣｐＧ部位、少なくとも１５個のＣｐＧ部位、少なくとも１８個のＣｐＧ部位、または少なくとも２０個のＣｐＧ部位を含み得る。

【０２１７】

さらに、コンピューティングシステム２１４により、配列決定データ２１２を解析して、参照配列の１つまたは複数の対照領域に対応する１つまたは複数の第２の配列表示２２２を決定することができる。１つまたは複数の対照領域は、１つもしくは複数の陽性対照領域および／または１つもしくは複数の陰性対照領域を含み得る。様々な例では、陽性対照領域は、少なくとも閾値量のメチル化シトシンを有する分子を有し、少なくとも閾値数のＣｐＧ部位を含む参照配列のゲノム領域を含み得る。陽性対照領域は、がんが存在する対象から得られる、および腫瘍が存在しない対象から得られた試料における、１つまたは複数のＣＧ領域内に少なくとも閾値量のメチル化を有する核酸分子に対応し得る。少なく

とも一部の例では、閾値量のメチル化は、核酸分子において少なくとも５個、少なくとも６個、少なくとも７個、少なくとも８個、少なくとも９個、少なくとも１０個、少なくとも１１個、少なくとも１２個、少なくとも１５個またはそれよりも多くのＣｐＧがメチル化されていることに対応し得る。１つまたは複数の実例では、陽性対照領域は、１つまたは複数のＣＧ領域において高メチル化であり、がんが存在する対象およびがんが存在しない対象のどちらから得られた試料にも由来する核酸分子にマッピングされ得る。１つまたは複数の例では、陰性対照領域は、閾値量未満の、メチル化シトシンを有する分子、および少なくとも閾値数のＣｐＧ部位を有する参照配列のゲノム領域を含み得る。陰性対照領域は、がんが存在する対象から得られる、および腫瘍が存在しない対象から得られた試料における、１つまたは複数のＣＧ領域内に追加の閾値量未満のメチル化を有する核酸分子に対応し得る。様々な例では、追加の閾値量のメチル化は、核酸分子において１個以下、２個以下、３個以下、４個以下、５個以下、６個以下、または７個以下のＣｐＧがメチル化されていることに対応し得る。１つまたは複数の追加の実例では、陰性対照領域は、１つまたは複数のＣＧ領域において低メチル化であり、がんが存在する対象およびがんが存在しない対象のどちらから得られた試料にも由来する核酸分子にマッピングされ得る。

【０２１８】

１つまたは複数の実例では、第１の配列表示２２０は、配列決定データ２１２に含まれる配列表示を参照配列の１つまたは複数の分類領域と整列することによって決定することができる。さらに、第２の配列表示２２２は、配列決定データ２１２に含まれる配列表示を参照配列の１つまたは複数の対照領域と整列することによって決定することができる。アライメントプロセスでは、参照配列の１つまたは複数の分類領域に対応する、配列決定データ２１２に含まれる配列表示の数を決定することによって第１の配列表示２２０を同定することができる。さらに、アライメントプロセスでは、参照配列の１つまたは複数の対照領域に対応する配列表示の数を決定することによって第２の配列表示２２２を同定することができる。

【０２１９】

１つまたは複数の実例では、アライメントプロセスにより、配列データ２１２に含まれる個々の配列表示と参照配列の一部分との相同性の量を決定することができる。所与の配列表示と参照配列との相同性の量は、所与の配列表示の対応する位置と同じヌクレオチドを有する参照配列の位置の数を示し得る。コンピューティングシステム２１４により、配列表示が参照配列の一部分と整列されることを、配列表示と参照配列の一部分が少なくとも閾値量の相同性を有することを決定することに基づいて決定することができる。配列表示が参照配列の複数の一部分に対して少なくとも閾値量の相同性を有するシナリオでは、配列表示と最大量の相同性を有する参照配列の一部分を、配列表示と整列していると決定することができる。

【０２２０】

所与の配列表示と参照配列の一部分との相同性の量は、ＢＬＡＳＴプログラム（基本的なローカルアライメント検索ツール）およびＰｏｗｅｒＢＬＡＳＴプログラム（Altschul et al., J. Mol. Biol., 1990, 215, 403-410；Zhang and Madden, Genome Res., 1997, 7, 649-656）を使用して、または、Ｎｅｅｄｌｅｍａｎ およびＷｕｎｓｃｈのアルゴリズム（J. Mol. Biol. 48; 443-453 (1970)）を使用するＧａｐプログラム（Wisconsin Sequence Analysis Package, Genetics Computer Group, University Research Park, Madison Wis.）を初期設定を用いて使用することによって、決定することができる。配列表示と参照配列の一部分との相同性の量は、Ｂｕｒｒｏｗｓ－Ｗｈｅｅｌｅｒアライナー（Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754-1760）を使用して決定することもできる。

【０２２１】

１つまたは複数の例では、配列決定データ２１２に含まれる配列表示を参照配列と整列

した後、整列された配列表示を解析して、配列表示の１つまたは複数の群を同定することができる。例えば、個々の整列された配列表示は、配列決定データ２１２に含まれる個々の配列決定リードに対応し得る。これらのシナリオでは、整列された配列表示は、試料プール２１０に含まれる単一の核酸分子に対応する複数のリードを含み得る。１つまたは複数の追加の例では、整列された配列表示は、プール２１０に含まれる個々の核酸分子に対応し得る。これらの状況では、コンピューティングシステムにより、プール２１０に含まれる個々の核酸分子に対応する、配列データ２１２に含まれるリードの群を、配列決定リードの各群に共通する分子バーコードに基づいて決定することができる。すなわち、プール２１０に含まれる個々の核酸分子を、個々の核酸分子を一意的に同定する分子バーコードで符号化することができ、少なくとも一部の場合では、個々の核酸分子を、配列決定データ２１２に含まれる複数の配列決定リードによって表すことができる。したがって、プール２１０に含まれる単一の核酸分子に対応する複数の配列表示が配列決定データ２１２内に存在する場合、コンピューティングシステム２１４により、複数の配列表示を一緒に群分けすることができる。様々な例では、プール２１０に含まれる単一の核酸分子に対応する配列表示の群は、本明細書では、「ファミリー」と称され得る。さらに、共通の分子バーコードを有する整列された配列表示の参照配列に関する開始および終止位置を使用して、プール２１０に含まれる個々の核酸に対応する配列表示を群分けすることができる。１つまたは複数の実例では、プール２１０に含まれる単一の核酸分子に対応する配列表示のファミリーを表す個々の配列表示は、本明細書では、「コンセンサス配列表示」と称され得る。

【０２２２】

操作２２４において、コンピューティングシステム２１４により、第１の配列表示および第２の配列表示２２２を解析して、個々の分類領域に対応するメトリクスを生成することができる。図２の実例では、コンピューティングシステム２１４により、第１の配列表示２２０および第２の配列表示２２２を解析して、分類領域メトリクス２２６を生成することができる。分類領域メトリクス２２６は、少なくとも閾値量のメチル化シトシンを有する第１の配列表示２２０の数に基づいて決定された定量的尺度を含み得る。１つまたは複数の実例では、分類領域メトリクス２２６は、位置特定された少なくとも閾値量のメチル化シトシンを有する第１の配列表示２２０の数に対応する配列決定リードの数に基づいて決定された定量的尺度を含み得る。１つまたは複数の追加の実例では、分類領域メトリクス２２６は、第１の配列表示２２０の数に対応する核酸分子の数に基づいて決定された定量的尺度を含み得る。様々な例では、分類領域メトリクス２２６は、少なくとも閾値量のメチル化シトシンを有する第１の配列表示２２０の数および参照配列の対照領域に対応する第２の配列表示２２２の数に基づいて決定された定量的尺度を含み得る。１つまたは複数のさらなる実例では、分類領域メトリクス２２６は、少なくとも閾値量のメチル化シトシンを有する第１の配列表示２２０の数と第２の配列表示の数に対する比に関連する定量的尺度を含み得る。少なくとも一部の例では、分類領域メトリクス２２６に含まれる定量的尺度を生成するためにコンピューティングシステム２１４によって使用される第２の配列表示２２２の配列表示は、参照配列の陽性対照領域に対応する配列表示を含み得る。

【０２２３】

分類領域メトリクス２２６は、コンピューティングシステム２１４により第１の配列表示２２０および第２の配列表示２２２のうちの少なくとも一方を使用して生成された定量的尺度に関して１つまたは複数の正規化操作を実施することによって決定することもできる。例えば、コンピューティングシステム２１４により、第１の配列表示２２０または第２の配列表示２２２のうちの少なくとも一方を使用して生成された定量的尺度に関して対数計算を実施することができる。さらに、分類領域メトリクス２２６は、コンピューティングシステム２１４により第１の配列表示２２０または第２の配列表示２２２のうちの少なくとも一方を使用して決定された定量的尺度に疑似カウント（ｐｓｅｕｄｏｃｏｕｎｔ）を付加することによって決定することができる。１つまたは複数の実例では、１つまたは複数の正規化操作は、個々の分類領域についての第１の配列表示２２０の参照配列の陽

性対照領域に対応する第２の配列表示２２２の数に対する比に対応する定量的尺度を決定することを含み得る。

【０２２４】

１つまたは複数の実例では、コンピューティングシステム２１４により、参照配列の個々の分類領域に対応し、個々の分類領域内に位置する少なくとも閾値量のメチル化シトシンを有する第１の配列表示２２０の数を決定することができる。これらのシナリオでは、コンピューティングシステム２１４により、個々の分類領域についての個々の分類領域メトリクス２２６を決定することができる。さらに、コンピューティングシステム２１４により、陽性対照領域に対応する第２の配列表示２２２の数を決定することができる。少なくとも一部の例では、コンピューティングシステム２１４により、個々の分類領域について、参照配列の陽性対照領域に対応する（the correspond to）第２の配列表示２２２の総数に関連して、個々の分類領域に対応し、分類領域内に少なくとも閾値量のメチル化シトシンを有する分子を有するいくつかの第１の配列表示２２０を含む、比を決定することができる。１つまたは複数の例では、コンピューティングシステム２１４により、比に疑似カウントの値を付加して、個々の分類領域についての分類領域メトリクス２２６を決定することができる。疑似カウントの値は、少なくとも１、少なくとも１．２、少なくとも１．４、少なくとも１．６、少なくとも１．８、または少なくとも２であり得る。さらに、コンピューティングシステム２１４により、比と疑似カウントの組合せに関して常用対数操作を実施して、個々の分類領域についての分類領域メトリクス２２６を決定することができる。少なくとも一部の実例では、コンピューティングシステム２１４により、分類領域メトリクスの少なくとも一部分を以下の方程式：

【数１】

$$\text{領域iのスコア} = log_{10}\left(\frac{x_i}{x_{\text{陽性}_{\text{対照}}}} + \text{疑似カウント}\right) \quad (1),$$

＃１　領域ｉのスコア

＃２　陽性対照

＃３　疑似カウント

（式中、x_iは、領域Ｉに含まれる少なくとも閾値量のメチル化シトシンを有する、個々の分類領域ｉについての第１の配列表示２２０の総数であり、$x_{\text{陽性_対照}}$は、参照配列の陽性対照領域に対応する第２の配列表示２２２の総数である）に従って決定することができる。

【０２２５】

操作２２８において、コンピューティングシステム２１４により、モデルを実行して、がんの指標を分類領域メトリクス２２６に基づいて決定することができる。図２の実例では、コンピューティングシステム２１４により、分類領域メトリクス２２６を使用してモデルを実行して、モデル出力２３０を生成することができる。１つまたは複数の例では、モデル出力２３０は、対象２０６によって提供された試料２０４に関して腫瘍検出２３２の状態または腫瘍非検出２３４の状態を示し得る。１つまたは複数の追加の例では、コンピューティングシステム２１４により、モデルを実行して、試料２０４についての腫瘍画分２３６の推定値を決定することができる。１つまたは複数のさらなる例では、コンピューティングシステム２１４により、モデルを実行して、試料２０４を提供した対象２０６における腫瘍が存在する確率を決定することができる。

【０２２６】

１つまたは複数の例では、モデルは、１つまたは複数の機械学習技術をインプリメントする分類モデルを含み得る。１つまたは複数の実例では、モデルは、線形回帰モデルを含み得る。様々な例では、モデルを実行して、試料２０４を提供した対象２０６における腫瘍が存在する確率２３８を分類領域メトリクス２２６に基づいて決定することができる。

１つまたは複数の実例では、コンピューティングシステム２１４により、モデルを実行して、個々の分類領域に対する重みを決定することができる。個々の分類領域に対する重みは異なり得る。例えば、コンピューティングシステム２１４により、第１の分類領域についての第１の分類領域メトリクス２２６の第１の重みが第２の分類領域についての第２の分類領域メトリクス２２６の第２の重みとは異なることを決定することができる。少なくとも一部の実例では、試料２０４を提供した対象２０６における腫瘍が存在する確率２３８をコンピューティングシステム２１４により、以下の方程式：

【数２】

$$P(\text{がん} \mid \text{領域スコア}) = \frac{1}{1 + e^{-\sum_i w_i(\text{領域iのスコア}) + b}} \quad (2)$$

（式中、w_iは個々の分類領域の重みであり、領域ｉのスコアは方程式（１）を使用して計算され、ｂは線形回帰モデルに対応する傾きである）に対応するモデルを実行することによって決定することができる。少なくとも一部の例では、腫瘍が存在する確率２３８を使用して、腫瘍検出２３２の状態または腫瘍非検出２３４の状態を生成することができる。１つまたは複数のさらなる実例では、コンピューティングシステム２１４により、腫瘍が存在する確率２３８を閾値確率に対して解析して、試料２０４についての腫瘍検出２３２の状態または腫瘍非検出２３４の状態を決定することができる。コンピューティングシステム２１４により、試料２０４についての腫瘍が存在する確率２３８が少なくとも閾値確率であることを決定することにより、試料２０４が腫瘍検出２３２の状態に対応することを決定することができる。さらに、コンピューティングシステム２１４により、試料２０４についての腫瘍が存在する確率２３８が閾値確率未満であることを決定することにより、試料２０４が腫瘍非検出２３４の状態に対応することを決定することができる。

【０２２７】

１つまたは複数の追加の例では、コンピューティングシステム２１４により、最大変異型アレル画分（ＭＡＦ）を決定するモデルを実行することができる。様々な例では、コンピューティングシステム２１４により、最大ＭＡＦ値を使用してモデルを実行して、試料２０４について腫瘍画分２３６を決定することができる。１つまたは複数の実例では、コンピューティングシステム２１４により、分類領域メトリクス２２６を使用してモデルを実行して、ロジット変換された最大ＭＡＦ値を決定することができ、次いでそれをコンピューティングシステム２１４によって使用して、試料２０４についての腫瘍画分を推定することができる。様々な例では、コンピューティングシステム２１４により、最大ＭＡＦ値を解析して、試料２０４を提供した対象２０６におけるがんが存在する確率２３８を決定することができる。様々な例では、フーバー回帰（Huber, P. J. 1964. "Robust Estimation of a Location Parameter." Annals of Mathematical Statistics 35 (1): 73-101）を実施して、分類領域メトリクス２２６に基づいて最大ＭＡＦ値を決定することができる。

【０２２８】

様々な例では、モデル出力２３０は、腫瘍組織指標２４０も含み得る。腫瘍組織指標２４０は、試料２０４において検出されたゲノム材料を産生したがん細胞の起源である１つまたは複数の組織を示し得る。１つまたは複数の例では、腫瘍組織指標２４０は、試料２０４において検出されたゲノム材料を産生したがん細胞の１つまたは複数の起源組織に対応し得る。これらのシナリオでは、コンピューティングシステム２１４により、所与の組織タイプに対応する個々のモデルを用いた多重モデルを生成することができる。個々のモデルからの出力を解析して、１つまたは複数の試料において検出されたゲノム材料を産生したがん細胞の起源である組織を示す追加のメトリクスを決定することができる。少なくとも一部の例では、個々のモデルについての出力は、腫瘍画分２３６または腫瘍が存在する確率２３８のうちの少なくとも一方を示し得る。コンピューティングシステム２１４に

より、それぞれのモデル出力を解析して、腫瘍画分の最大値またはがんが存在する最大確率のうちの少なくとも一方を有するモデルを決定することができる。次いで、コンピューティングシステム２１４により、腫瘍画分の最大値および／またはがんが存在する最大確率を有するモデルに対応する腫瘍組織指標２４０を生成することができる。

【０２２９】

例えば、試料２０４は、異なるタイプのがんが存在する対象２０６から得ることができる。例として、第１の試料を第１の分類のがんが存在する対象の第１の群から得ることができ、第２の試料を第２の分類のがんが存在する対象の第２の群から得ることができる。第１の試料から生成された配列決定データをコンピューティングシステム２１４により解析して、第１の分類のがんについての分類領域に対応する第１のメトリクスを生成することができ、第１のメトリクスを使用して、第１の分類のがんに対応する第１のモデルを生成することができる。さらに、第２の試料から生成された配列決定データをコンピューティングシステム２１４により解析して、第１の分類のがんについての分類領域に対応する第２のメトリクスを生成することができ、第２のメトリクスを使用して、第２の分類のがんに対応する第２のモデルを生成することができる。モデルを訓練した後、コンピューティングシステム２１４により、訓練用対象に含めなかった１つまたは複数の追加の対象から得られた配列決定データを解析して、１つまたは複数の追加の対象についての分類領域メトリクスを決定することができる。次いで、分類領域メトリクスを、異なる腫瘍分類モデルを使用して解析して、モデル出力を生成することができる。モデル出力をコンピューティングシステムによって解析して、それぞれのモデル出力について最大値を有するモデルを決定し、モデルに対応する腫瘍組織分類を決定することができる。

【０２３０】

１つまたは複数の追加の例示的なインプリメンテーションでは、モデル出力２３０は、参照配列の１つまたは複数のゲノム領域のメチル化状態も示し得る。例えば、コンピューティングシステム２１４により、分類領域メトリクス２２６を解析して、参照配列の１つまたは複数のプロモーター領域のメチル化状態を決定することができる。様々な例では、１つまたは複数のプロモーター領域は、対象における腫瘍の存在に関する少なくとも１つのプロモーター領域を含み得る。１つまたは複数の実例では、分類領域メトリクス２２６は、１つまたは複数のプロモーター領域に関して少なくとも閾値量のメチル化を有する配列表示の数を示し得る。これらのシナリオでは、コンピューティングシステム２１４により、プロモーター領域内の少なくとも閾値量のメチル化シトシンを有する分子を有する配列表示の数が閾値数よりも多いことを決定することにより、プロモーター領域がメチル化されていることを決定することができる。

【０２３１】

さらに他の例示的なインプリメンテーションでは、コンピューティングシステム２１４により、配列決定データ２１２を解析して、プロモーター領域に対応する配列表示の数に対応する定量的尺度を決定し、陽性対照領域の数に対応する追加の配列表示の数に対応する追加の定量的尺度を決定することができる。正規化されたメトリクスを定量的尺度および追加の定量的尺度に基づいて決定することができる。正規化されたメトリクスを閾値に関して解析して、プロモーター領域のメチル化状態を決定することができる。様々な例では、閾値は、異なるプロモーター領域について異なり得る。閾値は、がんのないデータセットを使用して決定することができ、閾値は、偽陽性率が５％以下、４％以下、３％以下、２％以下、または１％以下かつ特異度が少なくとも８０％、少なくとも９０％、少なくとも９５％、少なくとも９８％、または少なくとも９９％になるように設定する。少なくとも一部の例では、コンピューティングシステム２１４により、プロモーター領域を、少なくとも４個の配列表示がプロモーター領域に対応する、少なくとも５個の配列表示がプロモーター領域に対応する、少なくとも６個の配列表示がプロモーター領域に対応する、少なくとも７個の配列表示がプロモーター領域に対応する、少なくとも８個の配列表示がプロモーター領域に対応する、少なくとも９個の配列表示がプロモーター領域に対応する、少なくとも１０個の配列表示がプロモーター領域に対応する、少なくとも１１個の配列

表示がプロモーター領域に対応する、または少なくとも１２個の配列表示がプロモーター領域に対応する状況での閾値に関して解析することができる。

【０２３２】

１つまたは複数のさらなる実例では、コンピューティングシステム２１４により、多重モデルからの結果を組み合わせて、モデル出力２３０を決定することができる。例えば、コンピューティングシステム２１４により、モデルを、分類領域のメチル化などの１つまたは複数のエピジェネティックシグナルに関して実行して、１つまたは複数の第１の腫瘍メトリクスを決定することができる。メチル化に関して、コンピューティングシステム２１４により、試料を提供する対象にがんが存在する指標を生じさせる分類モデル、例えばロジスティック回帰モデル、および試料についての腫瘍画分を予測する追加のモデルの両方を実行することができる。１つまたは複数の例では、エピジェネティックシグナルは、試料から生成される配列表示の断片長にも対応し得る。さらに、コンピューティングシステム２１４により、１つまたは複数の追加のモデルをゲノムシグナルに関して実行して、試料に関するさらなる腫瘍メトリクスを生成することができる。様々な例では、ゲノムシグナルは、参照配列の１つまたは複数のゲノム領域における１つもしくは複数の一塩基バリアント（ＳＮＶ）の存在および／または挿入もしくは欠失の存在に対応し得る。少なくとも一部の例では、コンピューティングシステムは、いくつかのモデルをゲノムシグナルに対応するデータに関して実行することによって生成された腫瘍メトリクスおよびエピジェネティックシグナルを組み合わせて、所与の試料についての集計された腫瘍メトリクスを産生することができる統合システムを含み得る。一部の実施形態では、モデル出力２３０から得られた定量的尺度を使用して閾値に関して解析することができる。モデル出力から得られた定量的尺度が少なくとも閾値である状況では、コンピューティングシステム２１４により、対象にがんが存在する指標を決定することができる。モデル出力から得られた定量的尺度が閾値よりも小さい状況では、コンピューティングシステム２１４により、対象にがんが存在しないまたは検出されないという指標を決定することができる。一部の実施形態では、がんが存在するかどうかの指標を決定するために使用される閾値は、正常試料のセットを使用して計算し、高い特異度がもたらされる特定の値に設定する。

【０２３３】

様々な追加のインプリメンテーションでは、コンピューティングシステム２１４により、個々のゲノム領域のメチル化状態を決定することができる。１つまたは複数の実例では、コンピューティングシステム２１４により、１つまたは複数のプロモーター領域のメチル化状態を決定することができる。１つまたは複数の例では、配列決定データ２１２を解析して、１つまたは複数のゲノム領域に対応する配列表示を決定することができる。例えば、配列決定データ２１２を解析して、１つまたは複数のプロモーター領域に対応する配列表示の数を決定することができる。少なくとも一部の例では、コンピューティングシステム２１４により、少なくとも閾値量のメチル化シトシンを有する個々のプロモーター領域に対応する配列表示の数を決定することができる。

【０２３４】

各ゲノム領域について、および個々の試料について、コンピューティングシステム２１４により、ゲノム領域内に少なくとも閾値数のメチル化シトシンを有するポリヌクレオチド分子に対応する配列表示の数を決定することができる。コンピューティングシステム２１４により、ゲノム領域に対応し、少なくとも閾値数のメチル化シトシンを有するポリヌクレオチド分子または配列リードの計数を使用して１つまたは複数の正規化操作を実施して、正規化されたメトリクスを生成することができる。例として、コンピューティングシステム２１４により、ゲノム領域に対応し、少なくとも閾値数のメチル化シトシンを有するポリヌクレオチド分子またはリードの計数を、対照領域、例えば陽性対照領域に対応する分子または配列決定リードの数で割ることができる。別の場合では、コンピューティングシステム２１４により、ゲノム領域に対応し、少なくとも閾値数のメチル化シトシンを有するポリヌクレオチド分子またはリードの計数を、同じゲノム領域に対応し、少なくとも同じ閾値数のメチル化シトシンを有する、対照データセット（すなわち、腫瘍が検出さ

れない試料で構成される対照データセット）内の分子または配列決定リードの数で割ることによって正規化されたメトリクスを実施することができる。

【０２３５】

正規化されたメトリクスを閾値に関して解析することができる。閾値は、所与のゲノム領域、例えば所与のプロモーター領域に対応し得る。様々な例では、閾値は、異なるプロモーター領域について異なり得る。これらのシナリオでは、第１のプロモーター領域は第１の閾値を有し得、第２のプロモーター領域は第２の閾値を有し得る。正規化メトリクスが少なくとも閾値である状況では、コンピューティングシステム２１４により、ゲノム領域が第１のメチル化状態を有することを決定することができる。正規化メトリクスが閾値未満のシナリオでは、コンピューティングシステム２１４により、ゲノム領域が第２のメチル化状態を有することを決定することができる。１つまたは複数の実例では、第１のメチル化状態を「メチル化」と標識することができ、第２のメチル化状態を「非メチル化」と標識することができる。

【０２３６】

所与のゲノム領域についての閾値を、がんが検出されない個体の試料から得られた訓練用データに基づいて決定することができる。１つまたは複数の例では、訓練試料から得られた配列表示を解析して、ゲノム領域に対応し、少なくとも閾値量のメチル化シトシンを有するポリヌクレオチド分子の数に関するｚスコアを決定することができる。１つまたは複数の実例では、プロモーター領域についての正規化メトリクスを決定するために使用されるプロモーター領域についての閾値は、プロモーター領域に関して訓練試料に基づいて計算されたｚスコアから得ることができる。

【０２３７】

図２の実例では、モデルを生成して、所与の対象におけるがんの存在または非存在に関する指標の数を決定することができることが記載されているが、少なくとも一部の追加の例では、配列決定データ２１２をコンピューティングシステム２１４によって解析して、特定のモデルの訓練を行わずに、がんの存在の指標を決定することができる。１つまたは複数の例では、コンピューティングシステム２１４により、腫瘍画分値を、対象にがんが存在するか否か不明である単一の対象から得られた１つまたは複数の試料から生成された配列決定データ２１２に基づいて決定することができる。１つまたは複数の例では、コンピューティングシステム２１４により、腫瘍画分値の変化を、単一の対象から２つまたはそれよりも多くの時点で得られた１つまたは複数の試料から生成された配列決定データ２１２に基づいて決定することができる。腫瘍画分値の変化を使用して、処置に対する対象の応答をモニタリングすることができる。１つまたは複数の追加の例では、第１の試料を、対象から、がんに関連する処置または手順の少なくとも１回の投与前またはその開始時に得ることができ、１つまたは複数の第２の試料を、対象から、がんに関連する処置または手順の少なくとも１回の投与後に得ることができる。１つまたは複数の実例では、１つまたは複数の第２の試料を、処置または手順の投与の少なくとも１週間、少なくとも２週間、少なくとも３週間、少なくとも４週間、少なくとも５週間、少なくとも６週間、少なくとも８週間、または少なくとも１０週間後に得ることができる。少なくとも一部の例では、第１の試料および第２の試料は、対象から得られた体液または対象から得られた組織のうちの少なくとも一方に由来し得る。

【０２３８】

１つまたは複数の例では、１つまたは複数の試料を所与の対象から得ることができる。１つまたは複数の試料から生成された配列決定データ２１２をコンピューティングシステムによって解析して、いくつかの分類領域についての定量的尺度を決定することができる。１つまたは複数の例では、定量的尺度は、１つまたは複数の分類領域との少なくとも閾値量の重複を有する配列表示の量に対応し得る。１つまたは複数の追加の例では、定量的尺度は、少なくとも閾値量のＣＧ含量を有するＣｐＧ領域内に少なくとも閾値量のメチル化シトシンを有する配列表示に対応し得る。様々な例では、対象にがんが存在する指標は、腫瘍画分を含み得る。１つまたは複数の追加の例では、対象にがんが存在する指標は、

変異型アレル画分を含み得る。少なくとも一部の例では、定量的尺度は、複数の陽性対照領域にわたる配列決定リードの総数に関連して、所与の分類領域に対応する配列決定リードの数に対応し得る。１つまたは複数のさらなる例では、がんが存在する指標を使用して、１つまたは複数のがんが存在する指標を１つまたは複数の閾値に関して解析することにより、所与の個体にがんが存在するまたは存在しないことに対応する出力を決定することができる。１つまたは複数の実例では、対象から得られた１つまたは複数の試料から決定された腫瘍画分を１つまたは複数の閾値に関して解析することができる。腫瘍画分が閾値レベルよりも高い場合、コンピューティングシステム２１４により、対象にがんが存在する確率が少なくとも８０％、少なくとも８５％、少なくとも９０％、少なくとも９５％、または少なくとも９９％であることを決定することができる。さらに、複数の試料が対象から得られた状況では、対象から得られた第１の試料から生成された第１の定量的尺度を、対象から得られた第２の試料から生成された第２の定量的尺度に関して解析することができる。少なくとも一部の例では、第１の定量的尺度と第２の定量的尺度の差を解析して、対象における処置応答の指標を決定することができる。

【０２３９】

対象にがんが存在する指標を決定するために使用される定量的尺度は、分類領域のサブセットの定量的尺度を解析することによって決定することができる。少なくとも一部の例では、分類領域のサブセットは、異なる対象について異なり得る。１つまたは複数の実例では、いくつかの分類領域についての定量的尺度の値を互いに対して解析し、定量的尺度の値の大きさに従って順位付けすることができる。様々な例では、所与の試料についての分類領域を、定量的尺度の最大値を有する１つまたは複数の分類領域から定量的尺度の最小値を有する１つまたは複数の分類領域まで降順に順位付けすることができる。

【０２４０】

様々な例では、分類領域の定量的尺度を順位付けした後、対象にがんが存在する指標を決定する前に、分類領域の群に対応する定量的尺度を除去することができる。例えば、対象にがんが存在する指標を決定するために使用されない分類領域の群は、最大定量的尺度値を有する分類領域の１％、最大定量的尺度値を有する分類領域の２％、最大定量的尺度値を有する分類領域の３％、最大定量的尺度値を有する分類領域の４％、最大定量的尺度値を有する分類領域の５％、または最大定量的尺度値を有する分類領域の６％を含み得る。少なくとも一部の例では、比較的高い定量的尺度値を有するいくつかの分類領域を、対象にがんが存在する指標を決定するために使用される分類領域の群から除外することができ、これは、少なくとも一部の場合では、順位付けられた一覧の最上位またはその付近の定量的尺度値に対応する分類領域は、非腫瘍起源を有する、および／または配列決定アーチファクトに関連する可能性があるからである。したがって、これらの分類領域に対応する定量的尺度を対象にがんが存在する指標を決定するために使用される解析から除去することにより、対象にがんが存在する指標の正確度を増加させることができる。

【０２４１】

１つまたは複数の例では、対象にがんが存在する指標を決定するために使用される分類領域の群を決定した後、次いで、群の分類領域のサブセットを、それぞれの定量的尺度の最大値を有する、群の少なくとも１０の分類領域、群の少なくとも２５の分類領域、群の少なくとも５０の分類領域、群の少なくとも７５の分類領域、群の少なくとも１００の分類領域、群の少なくとも１５０の分類領域、群の少なくとも２００の分類領域、群の少なくとも２５０の分類領域、群の少なくとも３００の分類領域、群の少なくとも３５０の分類領域、群の少なくとも４００の分類領域、群の少なくとも４５０の分類領域、または群の少なくとも５００の分類領域を同定することによって決定することができる。

【０２４２】

少なくとも一部の例では、１つまたは複数の統計学的尺度、例えば、平均値、中央値、または最頻値のうちの少なくとも１つを群の分類領域の定量的尺度のサブセットに適用して、対象にがんが存在する最初の指標を生成することができる。様々な例では、最初のがんの指標を、スケーリング係数に応じて修正することができる。対象にがんが存在する最

初の指標にスケーリング係数を適用することができ、これは、少なくとも一部のシナリオでは、陽性対照領域が異なるメチル化ＣｐＧの量を有し得るからである。例えば、陽性対照領域の少なくとも一部分は完全にメチル化されたＣｐＧを有し得、一方で、他の陽性対照領域は完全にはメチル化されていない可能性がある。さらに、様々な状況では、一部の分類領域は、高値の対象にがんが存在する指標、例えば、９０％腫瘍画分、９５％腫瘍画分、９９％腫瘍画分、または１００％腫瘍画分に対応し得るが、これらの分類領域に対応する核酸分子は完全にはメチル化されていない可能性がある。これらの場合を説明するために、対象にがんが存在する最初の指標にスケーリング係数を適用して、指標のより正確な決定をもたらすことができる。１つまたは複数の実例では、スケーリング係数は、決定された対象にがんが存在する指標を、対象にがんが存在する追加の指標に対応する追加のデータに関する本明細書に記載される１つまたは複数の技術、例えば、検証データ、または本明細書に記載される対象に腫瘍が存在する指標と直交するデータを生成する他の技術を使用して解析することによって決定することができる。

【０２４３】

様々な例では、定量的尺度を決定するために使用される分類領域は、差次的にメチル化された領域の１つまたは複数の一部分に対応する分類領域に対応し得る。１つまたは複数の例では、差次的にメチル化された領域は、がんの１つまたは複数の分類に対応するプロモーター領域を含み得る。例えば、分類領域は、差次的にメチル化された領域にわたるいくつかの配列決定表示を解析することによって決定することができる。これらのシナリオでは、少なくとも閾値数の配列決定表示と重複する差次的にメチル化された領域の１つまたは複数の一部分を分類領域に含めることができる。１つまたは複数の例では、差次的にメチル化された領域の１つまたは複数の一部分の定量的尺度を、差次的にメチル化された領域の分子数分布に基づいて決定することができる。例えば、定量的尺度を、差次的にメチル化された領域の分子分布の１つまたは複数のピーク内の分子数に基づいて決定することができる。例として、様々な例では、差次的にメチル化された領域にわたる分子の分布は、差次的にメチル化された領域内の１つまたは複数の部分領域と重複する分子の量がより多い、１つまたは複数のピークを示し得る。様々な例では、試料について、配列表示の最大量に対応する差次的にメチル化された領域の１つまたは複数の部分領域に対応する１つまたは複数のゲノム領域を、分類領域と定義することができる。少なくとも一部の例では、配列表示の分布は、差次的にメチル化された領域の他の部分領域よりも配列表示の数が多い差次的にメチル化された領域の部分領域に対応するピークを有し得る。これらのシナリオでは、部分領域を分類領域として同定することができる。対象にがんが存在する指標を決定するために使用される差次的にメチル化された領域の少なくとも一部分の部分領域を決定することにより、対象にがんが存在する指標を決定するために使用されるコンピューティングリソースおよびメモリーリソースの量を減少させることができる。

【０２４４】

例として、分類領域は、配列決定表示の少なくとも５０％が試料重複から得られる、配列決定表示の少なくとも５５％が試料重複から得られる、配列決定表示の少なくとも６０％が試料重複から得られる、配列決定表示の少なくとも６５％が試料重複から得られる、配列決定表示の少なくとも７０％が試料重複から得られる、配列決定表示の少なくとも７５％が試料重複から得られる、配列決定表示の少なくとも８０％が試料重複から得られる、配列決定表示の少なくとも８５％が試料重複から得られる、配列決定表示の少なくとも９０％が試料重複から得られる、配列決定表示の少なくとも９５％が試料重複から得られる、または配列決定表示の少なくとも９９％が試料重複から得られる、差次的にメチル化された領域の１つまたは複数の一部分を含み得る。１つまたは複数の実例では、分類領域を含む差次的にメチル化された領域の１つまたは複数の一部分は、参照配列に関して連続している可能性がある。

【０２４５】

図３は、１つまたは複数のインプリメンテーションによる、試料に関して１つまたは複数の腫瘍メトリクスを決定するために計算モデル３０２を訓練するフレームワーク３００

の例の図式表示である。フレームワーク３００は、コンピューティングシステム２１４を含み得る。コンピューティングシステム２１４により、計算モデル３０２を実行して、１つまたは複数のモデル出力３０４を生成することができる。１つまたは複数の例では、計算モデル３０２は、機械学習モデルであってよい。モデル出力３０４は、試料を提供した対象における腫瘍の存在または非存在に対応する指標を含み得る。１つまたは複数の実例では、モデル出力３０４は、腫瘍画分を含み得る。１つまたは複数の追加の実例では、モデル出力３０４は、対象にがんが存在する確率を含み得る。１つまたは複数のさらなる実例では、モデル出力は、対象にがんが存在する指標または対象にがんが存在しないという指標を含み得る。さらに他の実例では、モデル出力３０４は、核酸分子の１つまたは複数の領域のメチル化状態を示し得る。例として、コンピューティングシステム２１４により、プロモーター領域に対応する定量的尺度に関して計算モデル３０２を実行して、プロモーター領域のメチル化の量を決定することができる。他の実例では、モデル出力３０４は、試料の腫瘍組織指標を含み得る。

【０２４６】

フレームワーク３００は、配列表示３０６も含み得る。１つまたは複数の例では、配列表示３０６を、対象により提供された試料に由来する核酸分子を解析することに基づいて生成することができる。配列表示３０６は、目的の領域の数に対応するヌクレオチドの数を有するゲノム領域を含み得る。例えば、配列表示３０６は、第１の分類領域３０８に対応するヌクレオチドの配列を含み得る。さらに、配列表示３０６は、第２の分類領域３１０に対応するヌクレオチドの配列を含み得る。さらに、配列表示３０６は、第３の分類領域３１２に対応するヌクレオチドの配列を含み得る。様々な例では、配列表示３０６の第１の分類領域３０８、第２の分類領域３１０、および第３の分類領域３１２は、それぞれの分類領域３０８、３１０、３１２に含まれる、異なる量のメチル化シトシンを有し得る。１つまたは複数の追加の例では、配列表示３０６は、陽性対照領域３１４に対応するヌクレオチドの配列および陰性対照領域３１６に対応するヌクレオチドの配列を含み得る。

【０２４７】

計算モデル３０２は、個々の分類領域に対応するいくつかのコンポーネントを含み得る。１つまたは複数の例では、計算モデル３０２のコンポーネントは、それぞれの分類領域の定量的メトリクスに対応するそれぞれの値を有し得る。定量的メトリクスは、それぞれの分類領域に対応する配列表示の数を示し得る。１つまたは複数の例では、計算モデル３０２は、計算モデル３０２のそれぞれのコンポーネントに関連するいくつかの重みを含み得る。例えば、計算モデル３０２は、第１の重み３２０を有する第１のモデルコンポーネント３１８を含み得る。第１のモデルコンポーネント３１８は、第１の分類領域３０８に対応し得る。さらに、計算モデル３０２は、第２の重み３２４を有する第２のモデルコンポーネント３２２を含み得る。第２のモデルコンポーネント３２２は、第２の分類領域３１０に対応し得る。様々な例では、第１の重み３２０、第２の重み３２４、または第３の重み３２８のうちの少なくとも１つは、第１の重み３２０、第２の重み３２４、または第３の重み３２８のうちの少なくとも別の１つと異なり得る。

【０２４８】

１つまたは複数の実例では、第１のモデルコンポーネント３１８、第２のモデルコンポーネント３２２、および第３のモデルコンポーネント３２６についての値を試料ごとに決定することができる。例として、異なる試料について、計算モデル３０２により、第１のモデルコンポーネント３１８、第２のモデルコンポーネント３２２、または第３のモデルコンポーネント３２６のうちの少なくとも１つについて異なる値を決定することができる。様々な例では、コンピューティングシステム２１４により、第１の分類領域３０８についての第１の定量的尺度を、試料についての配列決定データに基づいて決定することができる。コンピューティングシステム２１４により、計算モデル３０２を実行して、第１のモデルコンポーネント３１８についての値を第１の定量的尺度に基づいて決定することができる。さらに、コンピューティングシステム２１４により、第２の分類領域３１０についての第２の定量的尺度を試料についての配列決定データに基づいて決定することができ

る。コンピューティングシステム２１４により、計算モデル３０２を実行して、第２のモデルコンポーネント３２２についての値を第２の定量的尺度に基づいて決定することができる。さらに、コンピューティングシステム２１４により、第３の分類領域３１２についての第３の定量的尺度を試料についての配列決定データに基づいて決定することができる。コンピューティングシステム２１４により、計算モデル３０２を実行して、第３のモデルコンポーネント３２６についての値を第３の定量的尺度に基づいて決定することができる。第１の定量的尺度、第２の定量的尺度、および第３の定量的尺度を、それぞれ第１の分類領域３０８、第２の分類領域３１０、および第３の分類領域３１２に対応するＣＧ領域内に少なくとも閾値量のメチル化を有する配列表示の数に基づいて決定することができる。１つまたは複数の追加の実例では、第１の重み３２０に対する値、第２の重み３２４に対する値、および第３の重み３２８に対する値を試料ごとに決定することができる。例えば、異なる試料について、計算モデル３０２により、第１の重み３２０、第２の重み３２４、または第３の重み３２８のうちの少なくとも１つについて異なる値を決定することができる。

【０２４９】

１つまたは複数の例では、コンピューティングシステム２１４により、訓練プロセスを実施して、計算モデル３０２を生成することができる。様々な例では、訓練プロセスにより、モデル出力３０４を決定するために使用することができる分類領域メトリクスに関連する１つまたは複数の特色を決定することができる。さらに、訓練プロセスにより、モデル出力３０４を決定するために使用することができる分類領域メトリクスに関連する１つまたは複数のパラメーターを決定することができる。例えば、訓練プロセスを使用して、計算モデル３０２に含められるモデルコンポーネントおよびモデルコンポーネントの対応する重みを決定することができる。

【０２５０】

図３の実例では、訓練用データ３３０を使用して訓練プロセスを実施することができる。訓練用データ３３０は、少なくとも対象３３２の第１の群に関して得られた情報および少なくとも対象３３４の第２の群に関して得られた情報を含み得る。１つまたは複数の例では、対象３３２の第１の群は、腫瘍が検出されない対象を含み得、対象３３４の第２の群は、腫瘍が検出される対象を含み得る。様々な例では、訓練用データ３３０は、対象３３２の第１の群および対象３３４の第２の群についての参照配列３０６の分類領域のメチル化の量に関連する特徴を含み得る。例えば、訓練用データ３３０は、対象３３２の第１の群および対象３３４の第２の群について分類領域３０８、３１０、３１２について少なくとも閾値レベルのメチル化を有する配列表示の数に対応する定量的尺度を示し得る。訓練用データ３３０は、対象３３２の第１の群および対象３３４の第２の群の配列決定データの解析に基づくモデルコンポーネントに対する重みも含み得る。１つまたは複数の実例では、訓練用データ３３０は、対象３３２の第１の群および対象３３４の第２の群により提供された試料から得られた配列決定データから決定された分類領域メトリクスに基づく第１の重み３２０に対する値、第２の重み３２４対する値、および第３の重み３２８に対する値を含み得る。

【０２５１】

訓練用データ３３０は、対象３３２の第１の群および対象３３４の第２の群の追加の特徴に対応する情報も含み得る。例として、訓練用データ３３０は、診療記録情報、病歴情報、がん処置歴情報、人口統計情報、ゲノミクス情報、これらの１つまたは複数の組合せなどを含み得る。

【０２５２】

１つまたは複数の例では、コンピューティングシステム２１４により、計算モデル３０２を訓練して、個体に存在する１つまたは複数のタイプのがんに関連する指標を決定することができる。さらに、様々な例では、計算モデル３０２は、複数の異なるモデルを含んでもよく、したがって、計算モデル３０２はアンサンブルモデルである。これらの状況では、コンピューティングシステム２１４により、１つまたは複数の訓練プロセスをアンサ

ンブルモデルの個々のモデルに関して実施することができる。１つまたは複数の実例では、計算モデル３０２は、個々のゲノム領域、例えば遺伝子についてまたは遺伝子の特定の群についてモデル出力を決定することにそれぞれが対応するいくつかの個々のモデルを含み得る。例えば、計算モデル３０２は、個々の遺伝子についてまたは遺伝子の特定の群について最大ＭＡＦ値を生成するためのいくつかの個々のモデルを含み得る。これらのシナリオでは、計算モデル３０２の訓練は、訓練プロセスにおけるゲノム情報の使用に起因して、個体にがんが存在する指標を決定するために使用される他のモデルよりも大きな制約を有し得る。結果として、ゲノム情報を使用して計算モデル３０２を訓練する状況では、計算モデル３０２の出力の正確度を増加させることができる。さらに、計算モデル３０２の訓練では、ゲノム情報を組み入れて最大ＭＡＦ値を決定することができるが、訓練後のテスト対象についてがんの指標を決定するための計算モデル３０２の使用は、ゲノム情報の使用を伴わずに実施することができ、テスト対象から得られた配列決定データから決定された定量的尺度に対応する入力ベクトルに基づくことができる。

【０２５３】

コンピューティングシステム２１４により、第１の訓練用データセット３３６から第Ｎの訓練用データセット３３８までを得て、計算モデル３０２を生成するための訓練プロセスを実施することができる。１つまたは複数の例では、第１の訓練用データセット３３６は、計算モデル３０２を訓練するために使用される対象３３２の第１の群および対象３３４の第２の群に対応する訓練用データ３３０の第１の一部分を含み得、第Ｎの訓練用データセット３３８は、対象３３２の第１の群および対象３３４である第２の群に対応する訓練用データ３３０の第２の一部分を計算モデル３０２に対する検証プロセスの一部として含み得る。様々な例では、計算モデル３０２を経時的に更新し、複数の訓練プロセスを行うことができる。これらのシナリオでは、第１の訓練用データセット３３６は、第１の期間に対応する、対象３３２の第１の群および対象３３４である第２の群についての訓練用データ３３０の一部分を含み得、第Ｎの訓練用データセット３３８は、第２の期間に対応する、対象３３２の第１の群および対象３３４である第２の群についての訓練用データ３３０の一部分を含み得る。

【０２５４】

１つまたは複数の例では、計算モデル３０２に対する訓練プロセス中に、コンピューティングシステム２１４により、１つまたは複数の最適化操作を実施することができる。１つまたは複数の実例では、コンピューティングシステム２１４により、計算モデル３０２に対する訓練プロセス中に、対象３３２の第１の群または対象３３４である第２の群のうちの少なくとも一方に含まれる他の対象から得られた試料に関して外れ値である対象３３２の第１の群または対象３３４の第２の群のうちの少なくとも一方から得られた１つまたは複数の試料を同定することができる。例として、コンピューティングシステム２１４により、対象３３２の第１の群または対象３３４の第２の群のうちの少なくとも一方に含まれる１つまたは複数の対象について生成されたモデル出力３０４が、対象３３２の第１の群または対象３３４の第２の群のうちの少なくとも一方に含まれる１つまたは複数の追加の対象について生成されたモデル出力３０４との少なくとも閾値量の相違を有することを決定することができる。１つまたは複数の例では、コンピューティングシステム２１４により、対象３３２の第１の群または対象３３４の第２の群のうちの少なくとも一方の追加の群について決定された平均モデル出力３０４と少なくとも１標準偏差、少なくとも１．５標準偏差、少なくとも２標準偏差、少なくとも２．５標準偏差、または少なくとも３標準偏差分異なるモデル出力３０４を有する１つまたは複数の第１の対象３３２または１つまたは複数の第２の対象３３４のうちの少なくとも一方を同定することができる。様々な例では、コンピューティングシステム２１４により、追加の対象に対応する試料から生成された情報に対して外れ値である対象に対応する試料から生成された情報にペナルティを適用することができる。

【０２５５】

１つまたは複数の追加の例では、計算モデル３０２の訓練においてコンピューティング

システム２１４によってインプリメントされる１つまたは複数の最適化プロセスは、訓練プロセス中に実施される個々の訓練サイクルについての訓練サイクルの数および／または反復の数に対応し得る。

【０２５６】

１つまたは複数の実例では、コンピューティングシステム２１４により、訓練プロセスの少なくとも１０００回の反復を実施して、計算モデル３０２を生成すること、訓練プロセスの少なくとも３０００回の反復を実施して、計算モデル３０２を生成すること、訓練プロセスの少なくとも５０００回の反復を実施して、計算モデル３０２を生成すること、訓練プロセスの少なくとも８０００回の反復を実施して、計算モデル３０２を生成すること、訓練プロセスの少なくとも１０，０００回の反復を実施して、計算モデル３０２を生成すること、訓練プロセスの少なくとも１２，０００回の反復を実施して、計算モデル３０２を生成すること、または訓練プロセスの少なくとも１５，０００回の反復を実施して、計算モデル３０２を生成することができる。様々な例では、コンピューティングシステム２１４により、計算モデル３０２に関連する損失関数の収束前に訓練プロセスを終了することができる。１つまたは複数の例では、計算モデル３０２を生じさせるための訓練プロセスの反復の数は、訓練プロセスが停止される前、および損失関数が収束する前に実施される訓練プロセスの反復の数に対応し得る。

【０２５７】

１つまたは複数の例では、計算モデル３０２を生成するためにコンピューティングシステム２１４によってインプリメントされる訓練プロセスの第１の段階は、１つまたは複数のタイプのがんを示す体細胞変異を含む訓練用データ３３０に含まれる試料を、１つまたは複数のタイプのがんを示す体細胞変異を含まない訓練用データ３３０に含まれる試料に対して決定することを含み得る。次いで、コンピューティングシステム２１４により、計算モデル３０２に対する訓練プロセスを、１つまたは複数のタイプのがんを示す１つまたは複数の体細胞変異を含む、訓練用データ３３０の試料を使用して、および、腫瘍が検出されない対象から得られたいくつかの試料を使用して実施することができる。様々な例では、訓練プロセスの第１の段階の少なくとも１００回の反復を実施することができる。

【０２５８】

さらに、コンピューティングシステム２１４によって実施される訓練プロセスは、体細胞変異を含まない試料の腫瘍メトリクスの値を、１つまたは複数のタイプのがんに対して予測することを含む第２の段階を含み得る。コンピューティングシステム２１４により、訓練プロセスの第２の段階の少なくとも１００回の追加の反復を実施して、計算モデル３０２を生成することができる。計算モデル３０２を生成するためにコンピューティングシステム２１４によって実施される訓練プロセスの第２の段階は、１つまたは複数のタイプのがんを示す体細胞変異を有する試料に対応する訓練用データ３３０の一部分を使用し、１つまたは複数のタイプのがんを示す体細胞変異を含まない試料の予測値、および腫瘍が検出されない対象から得られた試料に対応する訓練用データ３３０の一部分を使用して、計算モデル３０２を訓練することも含み得る。様々な例では、計算モデル３０２を生成するためにコンピューティングシステム２１４によって実施される訓練プロセスの第２の段階を少なくともさらに２回、少なくともさらに３回、少なくともさらに４回、少なくともさらに５回、または少なくともさらに６回実施することができる。訓練プロセスの第１段階および訓練プロセスの第２の段階が完了したら、コンピューティングシステム２１４により、計算モデル３０２に対する検証プロセスを、訓練用データ３３０に含まれる異なる試料から得られた情報を使用して実施することができる。

【０２５９】

１つまたは複数の実例では、コンピューティングシステム２１４により、複数の計算モデル３０２について訓練プロセスを実施することができる。これらのシナリオでは、コンピューティングシステム２１４によって訓練される個々の計算モデル３０２は、訓練用データ３３０に含まれる、対象から得られたゲノム材料の供給源である異なる組織タイプに対応し得る。１つまたは複数の例では、コンピューティングシステム２１４によって訓練

される個々の計算モデル３０２は、異なる分類のがん、例えば、結腸直腸がん、肺がん、膵臓がん、膀胱がん、乳がん、肝臓がん、皮膚がん、または１つもしくは複数の追加の分類のがんに対応し得る。コンピューティングシステム２１４により、異なる分類のがんに対応する複数の計算モデル３０２を訓練する状況では、個々の計算モデル３０２からの出力をコンピューティングシステム２１４によって集計し、解析して、対象について起源組織を決定することができる。

【０２６０】

様々な例では、試料に含まれるゲノム材料が由来する所与の組織に対応する個々の計算モデル３０２は、異なるモデルコンポーネントを有し得る。例えば、第１の組織タイプに対応する、コンピューティングシステム２１４によって生成された第１の計算モデルは、分類領域の第１のセットに対応する第１のモデルコンポーネントを有し得る。さらに、第２の組織タイプに対応する、コンピューティングシステム２１４によって生成された第２の計算モデルは、分類領域の第１のセットとは異なる少なくとも１つの分類領域を有する分類領域の第２のセットに対応する第２のモデルコンポーネントを有し得る。さらに、異なる組織タイプに対応する計算モデルの個々のコンポーネントに対する重みは異なり得る。すなわち、第１の計算モデルの分類領域の第１のセットおよび第２の計算モデルの分類領域の第２のセットが少なくとも１つの分類領域を共通して有する状況では、少なくとも１つの共通の分類領域に対応するモデルコンポーネントに対する重みは、第１の計算モデルおよび第２の計算モデルに関して異なり得る。

【０２６１】

さらに、計算モデル３０２を生成する際にコンピューティングシステムによって１つまたは複数の追加の正規化プロセスを実施することができる。例えば、少なくとも一部のシナリオでは、ＭＢＤで処理した分子は、異なる試料にわたって異なって分配され得る。１つまたは複数の例では、分子は、分子をＭＢＤで処理するために使用された試薬の組成の相違に起因して、異なる試料にわたって異なって分配され得る。１つまたは複数の追加の例では、分子は、分子をＭＢＤで処理するために使用された機器の相違またはプロセス条件のうちの少なくとも一方に起因して、異なる試料にわたって異なって分配され得る。

【０２６２】

例として、１つまたは複数の第１の試料について、ＭＢＤを用いた処理により、第１の分配に分離される、第１のＣＧ含量を有する領域を有する第１の分子、および、第２の分配に分離される、第２のＣＧ含量を有する領域を有する第２の分子がもたらされ得る。さらに、１つまたは複数の第２の試料について、ＭＢＤを用いた処理により、第１の分配に分離される、第１のＣＧ含量とは異なる第３のＣＧ含量を有する第３の分子、および、第２の分配に分離される、第２のＣＧ含量とは異なる第４のＣＧ含量を有する領域を有する第４の分子がもたらされ得る。様々な例では、ＣＧ含量の第１のカットオフ範囲にわたって第１の分子をＭＢＤで処理し、第１の分配に分離することができ、第２の分子をＭＢＤで処理し、第２の分配に分離することができる。さらに、第１のカットオフ範囲とは異なるＣＧ含量の第２のカットオフ範囲にわたって第３の分子をＭＢＤで処理し、第１の分配に分離することができ、第４の分子をＭＢＤで処理し、第２の分配に分離することができる。

【０２６３】

１つまたは複数の実例では、ＣＧ含量の第１のカットオフ範囲は、３～１０個の、メチル化シトシンを有するＣｐＧを含み得、第２のカットオフ範囲は、６～１４個の、メチル化シトシンを有するＣｐＧを含み得る。１つまたは複数の追加の実例では、ＣＧ含量の第１のカットオフ範囲は、４～９個の、メチル化シトシンを有するＣｐＧを含み得、第２のカットオフ範囲は、７～１３個の、メチル化シトシンを有するＣｐＧを含み得る。１つまたは複数のさらなる実例では、ＣＧ含量の第１のカットオフ範囲は、５～８個の、メチル化シトシンを有するＣｐＧを含み得、第２のカットオフ範囲は、８～１２個のＣｐＧを含み得る。さらに他の実例では、ＣＧ含量の第１のカットオフ範囲は、４～７個のＣｐＧを含み得、第２のカットオフ範囲は、６～１０個のＣｐＧを含み得る。様々な例では、ＣＧ

含量の第１のカットオフ範囲およびＣＧ含量の第２のカットオフ範囲を使用して、訓練用配列決定リードまたはテスト用配列決定リードのうちの少なくとも一方を決定するために使用されるメチル化シトシンの閾値量を決定することができる。少なくとも一部の例では、メチル化シトシンの閾値量は、ＭＢＤで処理した個々の分子の、例えば、少なくとも約８０％、少なくとも約８５％、少なくとも約９０％、少なくとも約９５％、または少なくとも約９９％が所与の分配に分離される、確率に対応するカットオフ数を含み得る。１つまたは複数の例では、メチル化シトシンの閾値量は、５個のメチル化シトシン、６個のメチル化シトシン、７個のメチル化シトシン、８個のメチル化シトシン、９個のメチル化シトシン、１０個のメチル化シトシン、１１個のメチル化シトシン、１２個のメチル化シトシン、１３個のメチル化シトシン、または１４個のメチル化シトシンに対応し得る。

【０２６４】

１つまたは複数の例では、コンピューティングシステム２１４により、個々の分類領域についてのメトリクスを、配列決定リードの第１の数を解析して、第１の量のＣＧ含量を有する第１の数の核酸分子を同定することによって、および配列決定リードの第２の数を解析して、第２の量のＣＧ含量を有する第２の数の核酸分子を同定することによって決定される定量的尺度に基づいて生成することができる。少なくとも一部の例では、第２の数の核酸分子を使用して、異なる試料についてのＭＢＤを使用して処理した分子の分離の変動を説明するために、第１の数の核酸分子を使用して決定されたメトリクスを修正することができる。様々な例では、個々の分類領域について、第１のメトリクスを、所与の試料について、個々の分類領域に対応する１つまたは複数の分配（例えば、第２の分配１３０および／または第３の分配１３４）中の閾値量のメチル化シトシンを有し、第１の量のシトシン－グアニン含量を有する分子の数に対応する第１の定量的尺度を決定することによって決定することができる。一部の実施形態では、第１の量のＣＧ含量は、核酸分子中、少なくとも５個、少なくとも６個、少なくとも７個、少なくとも８個、少なくとも９個、少なくとも１０個、少なくとも１５個、少なくとも２０個、少なくとも２５個、または少なくとも３０個のＣｐＧであり得る。一部の実施形態では、第１の量のＣＧ含量は、核酸分子中、５～１０個、５～１５個、５～２０個、５～３０個、１０～１５個、１０～２０個、１０～３０個、１５～２０個、１５～３０個または２０～４０個の間のＣｐＧであり得る。

【０２６５】

第１のメトリクスを、所与の試料について、複数の対照領域（例えば、陽性対照領域）に対応する１つまたは複数の分配（例えば、第２の分配１３０および／または第３の分配１３４）中の閾値量のメチル化シトシンを有し、第１の量のシトシン－グアニン含量を有する分子の数に対応する第２の定量的尺度を決定することによって決定することもできる。例として、個々の分類領域について、第１のメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域に対応する第２の定量的尺度を使用して決定することができる。

【０２６６】

正規化プロセスは、所与の試料について、所与の試料についての第２のメトリクスを、１つまたは複数の分配（例えば、第２の分配１３０および／または第３の分配１３４）中の少なくとも閾値量のメチル化シトシンおよび複数の対照領域に対応する第２の量のシトシン－グアニン含量を有する分子の数に基づいて１つまたは複数の追加の定量的尺度を決定することによって決定することも含み得、ここで、第２の量のシトシン－グアニン含量は、第１の量のシトシン－グアニン含量よりも少ない。一部の実施形態では、第２の量のＣＧ含量は、核酸分子中、５～１０個、５～１５個、１０～１５個、１０～２０個または１５～２０個の間のＣｐＧであり得る。１つまたは複数の例では、複数の対照領域は、陽性対照領域および／または陰性対照領域であり得る。１つまたは複数の例では、第２のメトリクスを、追加の定量的尺度および第２の定量的尺度を使用して決定することができる。少なくとも一部の例では、第２のメトリクスを、所与の試料について、１つまたは複数の追加の定量的尺度の第２の定量的尺度に対する比を決定することによって決定すること

ができる。１つまたは複数の追加の例では、第２のメトリクスを、所与の試料について、１つまたは複数の追加の定量的尺度の第２の定量的尺度に対する比の対数、例えば１０を底とする対数を決定することによって決定することができる。

【０２６７】

１つまたは複数の実例では、所与の試料についての第２のメトリクスは、値の組合せを含み得、ここで、個々の値は、複数の対照領域についての少なくとも閾値量のメチル化シトシンおよび所与の数のＣｐＧを有する分子の数に基づく追加の定量的尺度、ならびに第２の定量的尺度に対応する。例えば、第１の追加の定量的尺度を、第１の数、例えば６個のＣｐＧを有する対照領域内に少なくとも閾値量のメチル化シトシンを有する分子の第１の数に基づいて決定することができ、第２の追加の定量的尺度を、第２の数、例えば７個のＣｐＧを有する対照領域内に少なくとも閾値量のメチル化シトシンを有する分子の第２の数に基づいて決定することができる。少なくとも一部の例では、より多くの追加の定量的尺度を、追加の数のＣｐＧ、例えば、８個のＣｐＧ、９個のＣｐＧ、１０個のＣｐＧなど、最大でＣｐＧの上限閾値、例えば１２個のＣｐＧ、１３個のＣｐＧ、または１４個のＣｐＧを有する対照領域内に閾値量のメチル化シトシンを有する分子の追加の数に基づいて決定することができる。追加の定量的尺度を使用して第２の定量的尺度に関して決定される比を決定し、合計して、第２のメトリクスを決定することができる。

【０２６８】

様々な例では、個々の分類領域について、第２のメトリクスを決定するために使用することができる異なる量のＣｐＧに関して相関係数を決定することもできる。１つまたは複数の例では、相関係数により個々の追加の定量的尺度を修正することができ、次いで、修正された個々の追加の定量的尺度を集計して、第２のメトリクスを決定することができる。１つまたは複数の追加の例では、第１のメトリクスおよび第２のメトリクスを組み合わせて、所与の分類領域に対応する正規化されたメトリクスを決定することができる。１つまたは複数の実例では、第２のメトリクスを第１のメトリクスから差し引いて、正規化されたメトリクスを決定することができる。

【０２６９】

１つまたは複数の追加の実例では、複数の異なる量のシトシン－グアニン含量のそれぞれについての所与の分類領域についての相関係数、例えば、６個のＣｐＧについての第１の相関係数、７個のＣｐＧについての第２の相関係数、８個のＣｐＧについての第３の相関係数など、最大で閾値量のＣＧ含量を決定することができる。少なくとも一部の例では、相関係数は、訓練用データを１つまたは複数の線形回帰技術を使用して解析することによって決定することができる。例えば、訓練用データ３３０を個々の分類領域について線形回帰モデルに当てはめて、相関係数を決定することができる。様々な例では、訓練用データ３３０の少なくとも一部分の線形回帰モデルへの当てはめを、所与の分類領域についての追加の定量的尺度を様々なＣＧ含量、例えば６個のＣｐＧ、７個のＣｐＧ、最大で閾値数のＣｐＧにわたって集計し、定量的尺度の平均値を決定することによって実施することができる。

【０２７０】

１つまたは複数の例では、正規化されたメトリクスにより、個々の試料について決定された定量的尺度の変動を減少させることができる。少なくとも一部の例では、変動の減少により、正規化されたメトリクスを決定するための追加の正規化プロセスをインプリメントせずに決定された少なくともいくつかのモデル出力３０４に関して、モデル出力３０４の正確度の増加をもたらすことができる。

【０２７１】

図４は、１つまたは複数のインプリメンテーションによる、対象における腫瘍メトリクスを分類領域のメチル化のレベルに基づいて決定するための方法４００の例のフローチャートである。操作４０２において、方法４００は、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップを含み得る。個々の訓練用配列決定リードは、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレ

オチド配列を含み得る。個々の訓練用配列決定リードは、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれるメチル化シトシンを有する閾値量の分子を有し得る。１つまたは複数の実例では、複数の試料は、無細胞核酸を含み得る。１つまたは複数の例では、メチル化シトシンは、バイサルファイト変換および配列決定、Ｔｅｔ補助バイサルファイトシーケンシング（ＴＡＢ－Ｓｅｑ）、差次的酵素的切断、ＭＳＲＥおよび／もしくはＭＤＲＥを用いた処理、またはＭＢＤ分配のうちの少なくとも１つを使用して決定することができる。１つまたは複数の追加の例では、メチル化シトシンは、１つまたは複数の一分子配列決定方法、例えば、ナノポアＤＮＡ配列決定またはEid, J., et al. (2009) Real-time DNA sequencing from single polymerase molecules. Science, 323 (5910), 133-138に記載されているものを使用して決定することができる。

【０２７２】

１つまたは複数の例では、訓練プロセスは、コンピューティングシステムにより、複数の対象に含まれない追加の対象由来のテスト用配列データを得るステップを含み得る。テスト用配列データは、追加の対象の試料に由来するテスト用配列決定リードを含み得る。個々のテスト用配列決定リードは、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み得る。さらに、個々のテスト用配列決定リードは、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれるメチル化シトシンを有する少なくとも閾値量の分子を有し得る。追加の配列データに基づいて、モデルを実行して、追加の対象にがんが存在する指標を決定することができる。次いで、テスト用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応するテスト用配列決定リードに由来する第１の定量的尺度を決定することができる。さらに、テスト用配列決定リードを解析して、複数の対照領域の個々の対照領域に対応するテスト用配列決定リードに由来する第２の定量的尺度を決定することができる。次いで、個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定することができる。その後、個々の分類領域についてのメトリクスを含む入力ベクトルを生成することができる。モデルで入力ベクトルを使用して、追加の対象にがんが存在する指標を決定することができる。

【０２７３】

腫瘍画分の推定値を決定するためにモデルを訓練する状況では、訓練用配列決定リードは、訓練用配列データの第１の一部分を含み得、訓練用配列データの第２の一部分は、訓練用配列決定リードとは異なる追加の訓練用配列決定リードを含む。これらのシナリオでは、訓練用配列データの第１の一部分または訓練用配列データの第２の一部分の少なくとも一方を解析して、複数の試料の個々の試料中に存在する複数のバリアントの個々の頻度を決定することができる。次いで、個々の試料に関して、個々の試料に由来する個々の頻度の中で最大値を有する個々の頻度に対応する最高頻度を有する、複数のバリアントのうちの１つのバリアントを決定することができる。１つまたは複数の実例では、個々の試料について最大変異型アレル頻度を決定することができる。様々な例では、次いで、個々の試料についての腫瘍画分の個々の尺度を個々の試料に由来する個々の頻度の最大値に基づいて決定することができる。

【０２７４】

少なくとも一部の例では、モデルに対する訓練プロセスは、１つまたは複数の最適化操作を含み得る。例えば、訓練プロセスは、訓練用データ内に含まれる個々の試料の１つまたは複数の追加の重みを、閾値信頼水準の範囲内にある個々の試料についてのがんの指標に基づいて決定するステップを含み得る。個々の試料についてのがんの指標が閾値信頼水準の範囲外にあることを決定することにより、訓練プロセス中に個々の試料にペナルティを適用することができる（a penalty to can be applied to the individual sample）。

【０２７５】

１つまたは複数の訓練最適化操作は、１つまたは複数の機械学習アルゴリズムを使用し

て、モデルに対して、訓練用データの一部分を使用して訓練プロセスの１回または複数の第１の反復を実施するステップも含み得る。さらに、モデルについての第１の出力データを訓練プロセスの１回または複数の第１の反復に基づいて生成することができる。第１の出力データは、複数の対象のうちの第１の個々の対象にがんが存在する１つまたは複数の第１の追加の指標に対応し得、第１の個々の対象は、訓練用データの一部分に対応し得る。さらに、訓練プロセスは、第１の出力データと訓練用データとを組み合わせて、追加の訓練用データを産生するステップと、モデルに対して、追加の訓練用データの一部分を使用して訓練プロセスの１回または複数の第２の反復を実施するステップとを含み得る。次いで、モデルについての第２の出力データを訓練プロセスの１回または複数の第２の反復に基づいて生成することができる。第２の出力データは、複数の対象のうちの第２の個々の対象にがんが存在する１つまたは複数の第２の追加の指標を示し得、ここで、第２の個々の対象は、追加の訓練用データの一部分に対応する。１つまたは複数の実例では、複数の分類領域の個々の分類領域に対する重みを第１の出力データおよび第２の出力データに基づいて決定することができる。

【０２７６】

さらに、訓練プロセスは、訓練プロセスの１回または複数の反復の間に決定された、訓練用データに含められた１つまたは複数の試料についての少なくとも閾値を有するいくつかのがんの指標が存在することを決定するステップを含み得る。これらのシナリオでは、モデルの１つまたは複数の重みに対する修正が修正されないかまたは最小量で修正される。さらに、訓練プロセスの１回または複数の反復の間に決定された、訓練用データに含められた１つまたは複数の追加の試料についての閾値未満である、がんが存在する追加の指標の数を決定することができる。これらのシナリオでは、モデルの１つまたは複数の追加の重みに対する修正を決定することができ、１つまたは複数の追加の重みは最小量よりも多く修正される。

【０２７７】

さらに、プロセス４００は、操作４０４において、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する第１の定量的尺度を決定するステップを含み得る。１つまたは複数の例では、第１の定量的尺度を訓練用配列決定リードの数に基づいて決定することができる。１つまたは複数の追加の例では、第１の定量的尺度を訓練用配列決定リードに対応するポリヌクレオチド分子の数に基づいて決定することができる。複数の分類領域の個々の分類領域の少なくとも一部分は、がんが検出される対象において閾値量のメチル化シトシンを有する分子を有し、少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する。様々な例では、複数の分類領域は、がんが検出される患者において少なくとも１つの変異が存在するゲノム領域に対応し得る。さらに、複数の分類領域は、第１のがんタイプに対する第１の複数の分類領域に対応し得、第１の複数の分類領域とは異なる第２の複数の分類領域に基づいて、第２のがんタイプに対するモデルを生成することができる。

【０２７８】

プロセス４００は、操作４０６において、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リード由来の第２の定量的尺度を決定するステップを含み得る。１つまたは複数の例では、第２の定量的尺度を訓練用配列決定リードの数に基づいて決定することができる。１つまたは複数の追加の例では、第２の定量的尺度を訓練用配列決定リードに対応するポリヌクレオチド分子の数に基づいて決定することができる。複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有する参照ゲノムの追加のゲノム領域に対応する。さらに、個々の対照領域は、がんが検出される対象において、およびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する分子を有し得る。

【０２７９】

さらに、プロセス４００は、操作４０８において、複数の分類領域の個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領

域についての第２の定量的尺度に基づいて決定するステップを含み得る。１つまたは複数の例では、個々の分類領域についてのメトリクスは、スケーリング係数および誤差補正係数に基づいて決定される。１つまたは複数の実例では、スケーリング係数は対数関数を含み得、誤差補正係数は疑似カウントを含み得る。

【０２８０】

プロセス４００は、操作４１０において、コンピューティングデバイスにより、訓練用配列リードについての複数の分類領域の個々の分類領域についてのメトリクスを含む訓練用データを生成するステップを含み得る。がんの指標が腫瘍画分であるインプリメンテーションでは、訓練用データは、複数の試料の個々の試料についての腫瘍画分の個々の尺度を含み得、モデルを、複数の試料の個々の試料についての腫瘍画分の個々の尺度に関して実行することができる。

【０２８１】

プロセス４００は、操作４１２において、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、対象にがんが存在する指標を複数の分類領域の少なくとも一部分におけるメチル化シトシンを有する分子の量に基づいて決定するためのモデルを生成するステップも含み得る。モデルにより、複数の分類領域の個々の分類領域に対する重みを決定することができ、個々の分類領域の重みの少なくとも一部分は互いに異なり得る。様々な例では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の分類アルゴリズムを含み得、がんが存在する指標は、追加の対象にがんが存在する確率に対応する。１つまたは複数の追加の例では、１つまたは複数の機械学習アルゴリズムは、１つまたは複数の回帰アルゴリズムを含み、指標は、追加の試料の腫瘍画分の推定値に対応する。１つまたは複数の実例では、試料の腫瘍画分を決定するためのモデルの検出限界は、９５％の感度を前提として０．０１％以下、９５％の感度を前提として０．０５％以下、９５％の感度を前提として０．１％以下、９５％の感度を前提として０．１５％以下、９５％の感度を前提として０．２％以下、９５％の感度を前提として０．２５％以下、または９５％の感度を前提として０．３％以下であり得る。

【０２８２】

様々な例では、訓練プロセス中にまたは訓練プロセス後にモデルに提供される配列リードは、分類領域内に少なくとも閾値量のメチル化シトシンを有する。メチル化レベルを満たす配列リードを、少なくとも部分的に、１つまたは複数の分子分離プロセスを使用して産生することができる。分子分離プロセスは、対象の血液または組織の少なくとも一方に由来する複数の核酸を、ある１つの量のメチル結合ドメイン（ＭＢＤ）タンパク質を含む溶液と組み合わせて、核酸－ＭＢＤタンパク質溶液を産生するステップを含み得る。次いで、塩類溶液を用いた核酸－ＭＢＤタンパク質溶液の複数の洗浄を実施して、いくつかの核酸画分を産生することができる。個々の核酸画分は、少なくとも閾値シトシン－グアニン含量を有する複数の核酸の領域内のメチル化シトシンを有する閾値数の分子を有し得る。１つまたは複数の実例では、複数の洗浄のうちの１つの洗浄を１つの濃度の塩化ナトリウム（ＮａＣｌ）を有する溶液を用いて実施することができ、ＭＢＤタンパク質に対する一連の結合強度を有する、いくつかの核酸画分のうちの１つの核酸画分を産生することができる。

【０２８３】

１つまたは複数の例では、核酸の複数の分配のうちの第１の分配に関連付けられる第１の核酸画分を決定することができる。第１の分配は、ＭＢＤタンパク質に対する結合強度の第１の範囲に対応する。さらに、第１の分子バーコードを第１の核酸画分の核酸に結合させることができる。第１の分子バーコードを第１の分配に関連付けることができる。さらに、核酸の複数の分配のうちの第２の分配に関連付けられる第２の核酸画分を決定することができる。第２の分配は、ＭＢＤタンパク質に対する結合強度の第１の範囲とは異なる、ＭＢＤタンパク質に対する結合強度の第２の範囲に対応し得る。第２の分子バーコードを第２の核酸画分の核酸に結合させることができる。第２の分子バーコードは第２の分配に関連付けられる。

【０２８４】
１つまたは複数の追加の例では、いくつかの核酸画分の少なくとも一部分を、ある１つの量の、１つまたは複数の非メチル化シトシンを有する分子を切断する制限酵素と組み合わせて、配列決定リードを産生するために使用される複数の試料の少なくとも一部分を産生することができる。これらのシナリオでは、閾値量のメチル化シトシンを有する分子は、少なくとも閾値シトシン－グアニン含量を有する領域内のメチル化シトシンを有する分子の最低頻度に対応する。１つまたは複数のさらなる例では、いくつかの核酸画分の少なくとも一部分を、ある１つの量の、メチル化シトシンを有する分子を切断する制限酵素と組み合わせて、配列決定リードを産生するために使用される複数の試料の少なくとも一部分を産生する。これらの状況では、閾値量のメチル化シトシンを有する分子は、少なくとも閾値シトシン－グアニン含量を有する領域内のメチル化シトシンを有する分子の最高頻度に対応する。
例示的な方法
Ａ．試料中のがんの指標を決定する
【０２８５】
一部の実施形態では、本明細書に開示される方法は、試料由来のｃｆＤＮＡを配列決定するステップと、差次的にメチル化された領域および対照領域である、ＤＮＡ配列を含む複数の標的領域についてメチル化レベルを決定するステップとを含む。一部の実施形態では、本明細書に開示される方法は、ｃｆＤＮＡまたはその部分試料由来の少なくともある１つのエピジェネティック標的領域セットを捕捉するステップであって、ｃｆＤＮＡまたはその部分試料を少なくとも１つのエピジェネティック標的領域セットに対して特異的な標的特異的プローブと接触させることを含む、ステップと、標的領域についてメチル化レベルを決定するステップと、対象から得られた試料中にがんの指標が存在するか否かを決定するステップとを含む。これらの実施形態のいずれにおいても、核酸分子のメチル化レベル／状態は、改変シトシンを認識する結合剤を使用した分子の分配、ＭＳＲＥ／ＭＤＲＥ消化、バイサルファイト変換などのメチル化感受性変換、配列決定中の直接検出、または任意の他の好適なアプローチを含む方法のうちの１つまたは複数を使用して決定することができる。様々なアプローチが本明細書に記載されている。
【０２８６】
一部の実施形態では、本明細書に開示される方法は、ＤＮＡを含む試料を、ＤＮＡをＤＮＡ内の改変シトシンを認識する作用剤と接触させることによって分配するステップと、ＤＮＡを配列決定するステップと、複数の領域内の核酸の定量的尺度を決定するステップとを含む。
【０２８７】
１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む。方法は、コンピューティングシステムにより、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含む。方法は、コンピューティングシステムにより、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参

照ゲノムの追加のゲノム領域に対応する、ステップも含む。方法は、コンピューティングシステムにより、複数の分類領域の個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含む。方法は、コンピューティングデバイスにより、訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についてのメトリクスを含む訓練用データを生成するステップも含む。方法は、コンピューティングシステムにより、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて対象にがんが存在する指標を決定するためのモデルを生成するステップであって、モデルが、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分が、互いに異なる、ステップも含む。

【０２８８】

１つまたは複数の態様では、コンピューティングシステムは、１つまたは複数のハードウェアプロセッサーと、１つまたは複数のハードウェアプロセッサーによって実行されるときに、１つまたは複数のハードウェアプロセッサーに、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む操作を実施させるコンピュータ可読命令を含む１つまたは複数の非一時的コンピュータ可読ストレージ媒体とを含む。操作は、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含む。操作は、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含む。操作は、複数の分類領域の個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含む。操作は、訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についてのメトリクスを含む訓練用データを生成するステップも含む。操作は、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて、対象にがんが存在する指標を決定するためのモデルを生成するステップであって、モデルが、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分が、互いに異なる、ステップも含む。

【０２８９】

１つまたは複数の態様では、１つまたは複数のコンピュータ可読ストレージ媒体は、コンピューティングシステムの１つまたは複数のプロセッサーによって実行されるときに、コンピューティングシステムに、複数の対象の複数の試料に由来する訓練用配列決定リードを含む訓練用配列データを得るステップであって、個々の訓練用配列決定リードが、複数の試料のうちの１つの試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々の訓練用配列決定リードが、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む操作を実施させるコンピュータ可読命令を含む。操作は、訓練用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応する訓練用配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の

少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含む。操作は、訓練用配列決定リードを解析して、複数の対照領域に対応する訓練用配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含む。操作は、複数の分類領域の個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含む。操作は、訓練用対象の試料由来の訓練用配列リードについての複数の分類領域の個々の分類領域についてのメトリクスを含む訓練用データを生成するステップも含む。操作は、訓練用データを使用して、１つまたは複数の機械学習アルゴリズムをインプリメントして、複数の分類領域の少なくとも一部分におけるメチル化シトシンの量に基づいて、対象にがんが存在する指標を決定するためのモデルを生成するステップであって、モデルが、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分が、互いに異なる、ステップも含む。

【０２９０】

１つまたは複数の態様では、方法は、１つまたは複数のハードウェアプロセッサーおよびメモリーを有するコンピューティングシステムにより、対象から得られた試料に由来する配列決定リードを得るステップであって、個々の配列決定リードが、試料中に含まれる核酸の断片に対応するヌクレオチド配列を含み、かつ、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する、ステップを含む。方法は、コンピューティングシステムにより、複数の分類領域の個々の分類領域に対応する配列決定リードに由来する第１の定量的尺度を決定するステップであって、複数の分類領域の個々の分類領域の少なくとも一部分が、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、配列決定リードを解析して、複数の対照領域に対応する配列決定リードに由来する第２の定量的尺度を決定するステップであって、複数の対照領域の個々の対照領域が、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する、ステップも含み得る。方法は、コンピューティングシステムにより、複数のメトリクスのうちの個々のメトリクスが複数の分類領域の個々の分類領域に対応する、その複数のメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定するステップも含み得る。方法は、コンピューティングシステムにより、対象にがんが存在する指標を、複数のメトリクスの少なくとも一部分に基づいて決定するステップも含み得る。

【０２９１】

１つまたは複数の態様では、コンピューティングシステムは、プロセッサーと、プロセッサーによって実行されるときに、対象から得られた試料に由来する配列決定リードを得るように、コンピューティングシステムを構成する命令が記憶されたメモリーとを含み、個々の配列決定リードは、試料中に含まれる核酸の断片に対応するヌクレオチド配列を含み、かつ、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる、閾値量のメチル化シトシンを有する分子に対応する。コンピューティングシステムは、複数の分類領域の個々の分類領域に対応する配列決定リードに由来する第１の定量的尺度を決定することもでき、複数の分類領域の個々の分類領域の少なくとも一部分は、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する。コンピューティングシステムは、配列決定リードを解析して、複数の対照領域に対応する配列決定リードに

由来する第２の定量的尺度を決定することもでき、複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する。コンピューティングシステムは、複数のメトリクスのうちの個々のメトリクスが複数の分類領域の個々の分類領域に対応する、その複数のメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定することもできる。コンピューティングシステムは、対象にがんが存在する指標を、複数のメトリクスの少なくとも一部分に基づいて決定することもできる。

【０２９２】

コンピューティングシステムは、プロセッサーと、プロセッサーによって実行されるときに、対象からテスト用配列データを得るようにシステムを構成する命令が記憶されたメモリーとを含み、テスト用配列データは、対象の試料に由来するテスト用配列決定リードを含み、個々のテスト用配列決定リードは、追加の試料に含まれる核酸の断片に対応するヌクレオチド配列を含み、個々のテスト用配列決定リードは、少なくとも閾値シトシン－グアニン含量を有するヌクレオチド配列の領域内に含まれる閾値量のメチル化シトシンを有する分子に対応する。コンピューティングシステムは、テスト用配列決定リードを解析して、複数の分類領域の個々の分類領域に対応するテスト用配列決定リードに由来する第１の定量的尺度を決定することもでき、複数の分類領域の個々の分類領域の少なくとも一部分は、がんが検出される対象において閾値量のメチル化シトシンを有し、かつ少なくとも閾値シトシン－グアニン含量を有する参照ゲノムのゲノム領域に対応する。コンピューティングシステムは、テスト用配列決定リードを解析して、複数の対照領域の個々の対照領域に対応するテスト用配列決定リードに由来する第２の定量的尺度を決定することもでき、複数の対照領域の個々の対照領域は、少なくとも閾値シトシン－グアニン含量を有し、かつがんが検出される対象においておよびがんが検出されない追加の対象において少なくとも閾値量のメチル化シトシンを有する参照ゲノムの追加のゲノム領域に対応する。コンピューティングシステムは、個々の分類領域についてのメトリクスを、個々の分類領域についての第１の定量的尺度および複数の対照領域についての第２の定量的尺度に基づいて決定することもできる。コンピューティングシステムは、個々の分類領域についてのメトリクスを含む入力ベクトルを生成することもできる。コンピューティングシステムは、対象にがんが存在する指標を、１つまたは複数の機械学習技術をインプリメントするモデルに入力ベクトルを提供して、対象にがんが存在する指標を生成することによって決定することもでき、モデルは、複数の分類領域の個々の分類領域に対する重みを含み、個々の分類領域の重みの少なくとも一部分は互いに異なる。

Ｂ．試料を複数の部分試料に分配する

【０２９３】

本明細書に記載される一部の実施形態では、異なる形態のＤＮＡ（例えば、高メチル化ＤＮＡおよび低メチル化ＤＮＡ）を、ＤＮＡの１つまたは複数の特徴に基づいて物理的に分配する。このアプローチを使用して、例えば、ある特定の部位または領域が高メチル化されているかまたは低メチル化されているかを決定することができる。分配は、例えば、分配タグをアダプターに含めることが容易になるように、試料中のＤＮＡ分子にアダプターを結合させる前に実施することができる。分配タグを使用して、分子がどの分配に見出されるかを同定することができる。分配（および該当する場合にはアダプターの結合）後、増幅、標的捕捉、および配列決定などのさらなるステップを実施することができる。

【０２９４】

メチル化プロファイリングは、ゲノムの異なる領域にわたるメチル化パターンを決定することを含み得る。例えば、メチル化の程度（例えば、分子あたりのメチル化核酸塩基の相対数）に基づく分子の分配および配列決定を含む、上記で考察したさらなるステップ後に、異なる分配中の分子の配列を参照ゲノムにマッピングすることができる。これは、他の領域と比較してより高度にメチル化されるか、またはあまり高度にメチル化されないゲ

ノムの領域を示すことができる。このようにして、ゲノム領域は、個々の分子と対比するとメチル化の程度が異なり得る。

【０２９５】

試料中の核酸分子を分配することによって、例えば、試料の１つの分配において優勢である希少な核酸分子を富化することにより、希少なシグナルを増加させることができる。例えば、高メチル化ＤＮＡには存在するが、低メチル化ＤＮＡにはあまり（またはまったく）存在しない遺伝的バリエーションを、試料を高メチル化核酸分子と低メチル化核酸分子とに分配することによって、より容易に検出することができる。試料の複数の分配を解析することにより、単一分子の多次元解析を実施することができ、したがって、より高い感度を実現することができる。分配は、核酸分子を、１つまたは複数のメチル化核酸塩基の存在または非存在に基づいて、分配または部分試料に物理的に分配することを含み得る。試料を、差次的遺伝子発現または疾患状態を示す特徴に基づいて、分配または部分試料に分配することができる。試料を、核酸、例えば、無細胞ＤＮＡ（ｃｆＤＮＡ）、非ｃｆＤＮＡ、腫瘍ＤＮＡ、循環腫瘍ＤＮＡ（ｃｔＤＮＡ）および無細胞核酸（ｃｆＮＡ）の解析中に、正常状態と罹患状態との間でシグナルに相違をもたらす特徴またはそれらの組合せに基づいて分配することができる。

【０２９６】

一部の実施形態では、高メチル化および／または低メチル化可変エピジェネティック標的領域を解析して、それらが、特定の免疫細胞タイプ、例えば希少な免疫細胞タイプ、腫瘍細胞または通常は解析されるＤＮＡ試料（例えば、ｃｆＤＮＡなど）に寄与しないタイプの細胞の差次的メチル化特徴を示すかどうかを決定する。

【０２９７】

一部の場合では、試料中の不均一なＤＮＡは、２つまたはそれよりも多くの分配（例えば、少なくとも３つ、４つ、５つ、６つまたは７つの分配）に分配される。一部の実施形態では、各分配は、差次的にタグ付けされる。タグ付けされた分配は、次いで、集合的試料調製および／またはシーケンシングのために、一緒にプールされ得る。分配－タグ付け－プーリングのステップは、１回を上回って行われ得、各回の分配は、異なる特徴（例は本明細書に提供されている）に基づいて行われ、他の分配および分配手段と識別される差次的タグを使用してタグ付けされる。他の場合では、差次的にタグ付けされた分配は別個に配列決定される。

【０２９８】

一部の実施形態では、差次的にタグ付けされ、プールされたＤＮＡから配列リードを得、ｉｎ　ｓｉｌｉｃｏで解析する。タグを使用して、異なる分配由来のリードを選別することができる。遺伝的バリアントを検出するための解析を分配ごとのレベル、ならびに全核酸集団レベルで実施することができる。例えば、解析は、各分配中の核酸における遺伝的バリアント、例えば、ＣＮＶ、ＳＮＶ、インデル、融合体を決定するためのｉｎ　ｓｉｌｉｃｏ解析を含み得る。一部の場合では、ｉｎ　ｓｉｌｉｃｏ解析は、クロマチン構造を決定することを含み得る。例えば、配列リードのカバレッジを使用して、クロマチンにおけるヌクレオソームの位置を決定することができる。より高いカバレッジは、ゲノム領域におけるより高いヌクレオソーム占有率と相関し得、一方で、より低いカバレッジは、より低いヌクレオソーム占有率またはヌクレオソーム枯渇領域（ＮＤＲ）と相関し得る。

【０２９９】

一部の実施形態では、分配は、メチル化などの１つまたは複数の特徴に基づく。適用可能な場合、データ解析または分配の一部として、適切な技術を使用して、他の特徴、例えば、配列長、ヌクレオソーム結合、配列ミスマッチ、免疫沈降、および／またはＤＮＡに結合するタンパク質に従って分子を選別することができる。結果として得られる分配は、以下の核酸形態のうちの１つまたは複数を含み得る：一本鎖ＤＮＡ（ｓｓＤＮＡ）、二本鎖ＤＮＡ（ｄｓＤＮＡ）、より短いＤＮＡ断片、およびより長いＤＮＡ断片。一部の実施形態では、シトシン改変（例えば、シトシンメチル化）またはメチル化に基づく分配が一般に行われ、必要に応じて、前述のＤＮＡの特徴または形態のうちのいずれかに基づき得

る少なくとも１つの追加の分配するステップと組み合わされる。一部の実施形態では、不均一な核酸集団は、１つまたは複数のエピジェネティック改変を有する核酸および１つまたは複数のエピジェネティック改変を有さない核酸に分配される。エピジェネティック改変の例としては、メチル化の存在または非存在；メチル化のレベル；メチル化のタイプ（例えば、５－メチルシトシン対他のタイプのメチル化、例えば、アデニンメチル化および／またはシトシンヒドロキシメチル化）、ならびに１つまたは複数のタンパク質、例えばヒストンとの会合および会合のレベルが挙げられる。あるいはまたはさらに、不均一な核酸集団は、ヌクレオソームと会合した核酸分子およびヌクレオソームが欠如した核酸分子に分配され得る。あるいはまたはさらに、不均一な核酸集団は、一本鎖ＤＮＡ（ｓｓＤＮＡ）および二本鎖ＤＮＡ（ｄｓＤＮＡ）に分配してもよい。あるいはまたはさらに、不均一な核酸集団は、核酸の長さ（例えば、最大で１６０ｂｐの分子および１６０ｂｐより長い長さを有する分子）に基づいて分配してもよい。

【０３００】

試料内の核酸の集団を分配するために使用される作用剤は、所望の特異性を有する抗体、天然の結合パートナーもしくはそのバリアントなどの親和性剤（Bock et al., Nat Biotech 28: 1106-1114 (2010)；Song et al., Nat Biotech 29: 68-72 (2011)）、または、例えばファージディスプレイによって所与の標的に対する特異性を有するように選択された人工ペプチドであり得る。一部の実施形態では、分配に使用される作用剤は、改変核酸塩基を認識する作用剤である。一部の実施形態では、作用剤によって認識される改変核酸塩基は、改変シトシン、例えばメチルシトシン（例えば５－メチルシトシン）である。一部の実施形態では、作用剤によって認識される改変核酸塩基は、試料のＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順の産物である。一部の実施形態では、改変核酸塩基は、「変換された核酸塩基」、つまり、その塩基対形成特異性が手順によって変化したものであり得る。例えば、ある特定の手順では、非メチル化または非改変シトシンがジヒドロウラシルに変換される、またはより一般的には、少なくとも１つの改変または非改変形態のシトシンは脱アミノ化を受け、それによって、ウラシル（ＤＮＡに関して改変核酸塩基とみなされる）またはさらに改変された形態のウラシルが生じる。分配剤の例としては、抗体、例えば、メチルシトシン（例えば、５－メチルシトシン）などの改変シトシンであり得る改変核酸塩基を認識する抗体が挙げられる。一部の実施形態では、分配剤は、５－メチルシトシン以外の改変シトシン、例えば５－カルボキシルシトシン（５ｃａＣ）を認識する抗体である。代替の分配剤としては、ＭｅＣＰ２などのタンパク質を含む、本明細書に記載されるメチル結合ドメイン（ＭＢＤ）およびメチル結合タンパク質（ＭＢＰ）が挙げられる。

【０３０１】

分配剤の追加の非限定的な例は、ヒストンと結合した核酸を遊離のまたは結合していない核酸から分離することができるヒストン結合タンパク質である。本明細書に開示される方法に使用することができるヒストン結合タンパク質の例としては、ＲＢＢＰ４、ＲｂＡｐ４８およびＳＡＮＴドメインペプチドが挙げられる。

【０３０２】

分配剤の特定の核酸への結合および核酸の部分試料への分配は、ある特定の程度まで行うこともでき、本質的にバイナリ様式で行うこともできる。一部の場合では、ある特定の改変をより大きな割合で含む核酸は、改変をより少ない割合で含む核酸よりも大きな程度で作用剤に結合する。同様に、分配により、ある特定の改変を含む核酸をより大きな割合で含む部分試料およびより少ない割合で含む部分試料を産生することができる。あるいは、分配により、改変を含む核酸の本質的に全てを含むまたは改変を含む核酸をまったく含まない部分試料を産生することができる。全ての場合に、様々なレベルの改変を逐次的に分配剤から溶離させることができる。

【０３０３】

一部の実施形態では、分配は、バイナリ分配および改変の程度／レベルに基づく分配の両方を含み得る。例えば、メチル化断片をメチル化ＤＮＡ免疫沈降（ＭｅＤＩＰ）によっ

て分配することもでき、メチル結合ドメインタンパク質（例えば、ＭｅｔｈｙｌＭｉｎｄｅｒ Ｍｅｔｈｙｌａｔｅｄ ＤＮＡ Ｅｎｒｉｃｈｍｅｎｔ Ｋｉｔ（ＴｈｅｒｍｏＦｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ）を使用して全てのメチル化断片を非メチル化断片から分配することもできる。続いて、さらなる分配には、メチル結合ドメインおよび結合断片を有する溶液中の塩濃度を調整することによって、異なるレベルのメチル化を有する断片を溶出することが含まれ得る。塩濃度が増加するにつれ、より高いメチル化レベルを有する断片が溶出される。

【０３０４】

一部の場合では、最終的な分配は、異なる程度の改変（改変の過剰発現または過少発現）を有する核酸が富化されたものである。過剰発現および過少発現は、集団において鎖あたりの改変の中央値数と比べた、核酸が有する改変の数によって定義され得る。例えば、試料中の核酸における５－メチルシトシン残基の中央値数が２である場合、２つより多くの５－メチルシトシン残基を含む核酸は、この改変が過剰発現しており、１または０個の５－メチルシトシン残基を有する核酸は、過少発現している。親和性分離の効果は、結合相で改変が過剰発現している核酸および非結合相（すなわち、溶液中）で改変が過少発現している核酸を富化することである。結合相内の核酸は、後続のプロセシングの前に溶出することができる。

【０３０５】

ＭｅＤＩＰまたはＭｅｔｈｙｌＭｉｎｅｒ（登録商標） Ｍｅｔｈｙｌａｔｅｄ ＤＮＡ Ｅｎｒｉｃｈｍｅｎｔ Ｋｉｔ（ＴｈｅｒｍｏＦｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ）を使用する場合、様々なレベルのメチル化を、逐次溶出を使用して分配することができる。例えば、低メチル化分配（メチル化なし）は、核酸集団を、磁気ビーズに付着した、キットからのＭＢＤと接触させることによって、メチル化分配から分離することができる。ビーズを使用して、メチル化核酸を、非メチル化核酸から分離する。続いて、１つまたは複数の溶出ステップを逐次的に行って、異なるレベルのメチル化を有する核酸を溶出させる。例えば、メチル化核酸の第１のセットを、１６０ｍＭまたはそれよりも高い、例えば、少なくとも１５０ｍＭ、少なくとも２００ｍＭ、３００ｍＭ、４００ｍＭ、５００ｍＭ、６００ｍＭ、７００ｍＭ、８００ｍＭ、９００ｍＭ、１０００ｍＭ、または２０００ｍＭの塩濃度で溶出させることができる。そのようなメチル化核酸を溶出させた後、磁気分離を、もう一度使用して、より高いレベルのメチル化核酸を、より低いレベルのメチル化を有するものから分離する。溶離および磁気分離ステップを繰り返して、低メチル化分配（メチル化を含まない核酸が富化されたもの）、メチル化分配（低レベルのメチル化を含む核酸が富化されたもの）、および高メチル化分配（高レベルのメチル化を含む核酸が富化されたもの）などの様々な分配を作成することができる。

【０３０６】

一部の方法では、親和性分離に基づく分配のために使用される作用剤に結合した核酸を洗浄ステップに供する。洗浄ステップは、親和性剤に弱く結合した核酸を洗い流す。そのような核酸は、平均または中央値に近い程度（すなわち、試料を作用剤と最初に接触させたときに固相に結合したままの核酸と、固相に結合していない核酸との中間値）に近い程度に改変を有する核酸を富化することができる。

【０３０７】

親和性分離は、異なる程度の改変を有する核酸の少なくとも２つ、ときには３つまたはそれよりも多くの分配をもたらす。分配は、依然として、別個であるが、少なくとも１つの分配の核酸、および通常は２つまたは３つ（またはそれよりも多く）の分配を、通常はアダプターの構成要素として提供される核酸タグに連結させ、異なる分配中の核酸は、１つ分配のメンバーを別のものと識別する異なるタグを受容する。同じ分配の核酸分子に連結されるタグは、互いに同じであっても異なっていてもよい。しかし、互いに異なる場合、タグは、それらが付着した分子を、特定の分配のものであるとして同定するように、共通したコードの部分を有し得る。

【０３０８】

メチル化などの特徴に基づく核酸試料の分配（portioning）に関するさらなる詳細については、参照により本明細書に組み込まれるＷＯ２０１８／１１９４５２を参照されたい。

【０３０９】

一部の実施形態では、核酸分子は、特定のタンパク質またはその断片に結合している核酸分子およびその特定のタンパク質またはその断片に結合していないものに基づいて、異なる分配に分画され得る。

【０３１０】

核酸分子をＤＮＡ－タンパク質結合に基づいて分画することができる。タンパク質－ＤＮＡ複合体をタンパク質の特定の特性に基づいて分画することができる。そのような特性の例としては、様々なエピトープ、改変（例えば、ヒストンメチル化もしくはアセチル化）または酵素活性が挙げられる。ＤＮＡに結合し、分画のための基準として機能し得るタンパク質の例としては、プロテインＡおよびプロテインＧを挙げることができるが、これらに限定されない。任意の好適な方法を使用して、核酸分子をタンパク質が結合した領域に基づいて分画することができる。核酸分子をタンパク質が結合した領域に基づいて分画するために使用される方法の例としては、ＳＤＳ－ＰＡＧＥ、クロマチン免疫沈澱（ＣｈＩＰ）、ヘパリンクロマトグラフィー、および非対称流動場分離法（ＡＦ４）が挙げられるが、これらに限定されない。

【０３１１】

一部の実施形態では、試料の複数の部分試料への分配を、核酸を、ＤＮＡ内の改変核酸塩基を認識する抗体と接触させることによって実施し、改変核酸塩基は、改変シトシン、または試料のＤＮＡ内の第２の核酸塩基とは異なってＤＮＡ内の第１の核酸塩基に影響を及ぼす手順の産物であり得る。一部の実施形態では、改変核酸塩基は５ｍＣである。一部の実施形態では、改変核酸塩基は５ｃａＣである。一部の実施形態では、改変核酸塩基はジヒドロウラシル（ＤＨＵ）である。一部の実施形態では、ＤＮＡ内の改変核酸塩基を認識する抗体を使用して、一本鎖ＤＮＡを分配する。

【０３１２】

一部の実施形態では、分配を、核酸をメチル結合タンパク質（「ＭＢＰ」）のメチル結合ドメイン（「ＭＢＤ」）と接触させることによって実施する。一部のそのような実施形態では、核酸をＭＢＰ全体と接触させる。一部の実施形態では、ＭＢＤは５－メチルシトシン（５ｍＣ）に結合し、ＭＢＰはＭＢＤを含み、本明細書ではメチル結合タンパク質またはメチル結合ドメインタンパク質と互換的に称される。一部の実施形態では、ＭＢＤは５ｍＣおよび５ｈｍＣに結合する。一部の実施形態では、ＭＢＤを、Ｄｙｎａｂｅａｄｓ（登録商標）Ｍ－２８０ストレプトアビジンなどの常磁性ビーズにビオチンリンカーを介してカップリングする。異なる程度のメチル化を有する画分への分配は、ＮａＣｌ濃度を上昇させることによって画分を溶離させることにより実施することができる。

【０３１３】

一部の実施形態では、結合したＤＮＡを、抗体またはＭＢＤをプロテイナーゼＫなどのプロテアーゼと接触させることによって溶離させる。これは、上記で考察したＮａＣｌを使用する溶離ステップの代わりにまたはそれに加えて実施することができる。

【０３１４】

本明細書で企図される改変核酸塩基を認識する作用剤の例としては、以下のものが挙げられるが、これらに限定されない：

（ａ）ＭｅＣＰ２は、非改変シトシンよりも５－メチル－シトシンに優先的に結合するタンパク質である。

（ｂ）ＲＰＬ２６、ＰＲＰ８およびＤＮＡミスマッチ修復タンパク質ＭＨＳ６は、非改変シトシンよりも５－ヒドロキシメチル－シトシンに優先的に結合する。

（ｃ）ＦＯＸＫ１、ＦＯＸＫ２、ＦＯＸＰ１、ＦＯＸＰ４およびＦＯＸＩ３は、好ましくは、非改変シトシンよりも５－ホルミルシトシンに結合する（Iurlaro et al., Genome Biol. 14: R119 (2013)）。

（ｄ）１つまたは複数のメチル化もしくは改変核酸塩基またはその変換産物、例えば５ｍＣ、５ｃａＣ、またはＤＨＵに対して特異的な抗体。

【０３１５】

一般に、溶離は、分子あたりの改変の数、例えばメチル化部位の数の関数であり、高塩濃度下ではより多くのメチル化を有する分子が溶離する。ＤＮＡをメチル化の程度に基づいて別個の集団に溶離させるために、ＮａＣｌ濃度を漸増させた一連の溶離緩衝液を使用することができる。塩濃度は、約１００ｎｍ～約２５００ｍＭのＮａＣｌの範囲であり得る。一実施形態では、プロセスにより、３つの分配がもたらされる。分子を、第１の塩濃度の、改変核酸塩基を認識する作用剤を含む分子を含む溶液と接触させ、その分子を捕捉部分、例えばストレプトアビジンに結合させてもよい。第１の塩濃度において、分子のある１つの集団は作用剤に結合し、ある１つの集団は結合しないままになる。結合しなかった集団を「低メチル化」集団として分離することができる。例えば、第１の分配により、低塩濃度、例えば１００ｍＭまたは１６０ｍＭにおいて結合しないままである、ＤＮＡの低メチル化形態が富化される。中間メチル化ＤＮＡが富化された第２の分配は、中間の塩濃度、例えば、１００ｍＭ～２０００ｍＭの間の濃度を使用して溶離される。この分配も試料から分離される。高メチル化形態のＤＮＡが富化された第３の分配は、高塩濃度を使用して、例えば、少なくとも約２０００ｍＭで溶離される。

【０３１６】

一部の実施形態では、５－メチルシチジン（５ｍＣ）に対して生じたモノクローナル抗体を使用して、メチル化ＤＮＡを精製する。一本鎖ＤＮＡ断片を得るために、ＤＮＡを、例えば９５℃で変性させる。標準的なまたは磁気ビーズとカップリングしたプロテインＧ、ならびに抗５ｍＣ抗体とのインキュベーション後の洗浄を使用して、抗体に結合したＤＮＡを免疫沈降させる。次いで、そのようなＤＮＡを溶離させることができる。分配は、沈降しなかったＤＮＡおよびビーズから溶離した１つまたは複数の分配を含み得る。

【０３１７】

一部の実施形態では、試料ＤＮＡ（例えば、５～２００ｎｇの間）を、メチル結合ドメイン（ＭＢＤ）緩衝液と混合し、磁気ビーズをＭＢＤタンパク質とコンジュゲートさせ、一晩インキュベートする。メチル化ＤＮＡ（高メチル化ＤＮＡ）は、このインキュベーション中に磁気ビーズ上のＭＢＤタンパク質に結合する。非メチル化である（低メチル化ＤＮＡ）またはあまり非メチル化であるＤＮＡ（中間メチル化）は、漸増濃度の塩を含有する緩衝液で、ビーズから洗い流される。例えば、非メチル化、低メチル化、および／または中間メチル化ＤＮＡを含有する、１つ、２つ、またはそれよりも多くの分画が、そのような洗浄から得られ得る。最終的に、高塩緩衝液を使用して、ＭＢＤタンパク質から、高度メチル化ＤＮＡ（高メチル化ＤＮＡ）を溶出させる。一部の実施形態では、これらの洗浄により、漸増メチル化レベルを有するＤＮＡの３つの分配（低メチル化分配、中間メチル化分画、および高メチル化分配）が得られる。

【０３１８】

一部の実施形態では、分配手順により、部分試料の間でＤＮＡ分子の不完全な選別がもたらされる可能性がある。例えば、非メチル化もしくは低メチル化部分試料中の分子うちの少数が高度に改変されている可能性があり（例えば、高メチル化）、かつ／または、高メチル化部分試料中の分子うちの少数が非改変もしくは大部分が非改変である可能性がある（例えば、非メチル化もしくは大部分が非メチル化）。そのような分子は非特異的に分配されたとみなされる。

【０３１９】

一部の実施形態では、非特異的に分配された分子を、メチル化依存性ヌクレアーゼ、例えば、制限酵素（ＲＥ）認識部位がメチル化ヌクレオチドを含有するＤＮＡを消化／切断するが、制限酵素（ＲＥ）認識部位が非メチル化ヌクレオチドを含有するＤＮＡは切断しない、メチル化依存性制限酵素（ＭＤＲＥ）を使用して除去する。一部の実施形態では、非特異的に分配された分子を、メチル化感受性ヌクレアーゼ、例えば、制限酵素（ＲＥ）認識部位が非メチル化ヌクレオチドを含有するＤＮＡを消化／切断するが、制限酵素（Ｒ

Ｅ）認識部位がメチル化ヌクレオチドを含有するＤＮＡは切断しない、メチル化感受性制限酵素（ＭＳＲＥ）を使用して除去する。例えば、一部の実施形態では、低メチル化部分試料を、メチル化依存性ヌクレアーゼ、例えばメチル化依存性制限酵素と接触させ、それによって、部分試料中の非特異的に分配されたＤＮＡ、例えばメチル化ＤＮＡを分解する。その代わりに、またはそれに加えて、高メチル化部分試料を、メチル化感受性エンドヌクレアーゼ、例えばメチル化感受性制限酵素と接触させ、それによって、部分試料中の非特異的に分配されたＤＮＡを分解する。

【０３２０】

１つまたは複数の分配された部分試料中の非特異的に分配されたＤＮＡの分解により、シトシン改変に基づくＤＮＡの正確な分配に依存する方法の性能を改善することができる。例えば、そのような分解により、改善された感度をもたらすこと、および／または下流の解析を単純化することができる。一部の実施形態では、ＤＮＡをメチル化などの改変に基づいて分配し、次いで、非特異的に分配されたＤＮＡを本明細書に記載されるＭＤＲＥおよび／またはＭＳＲＥを使用して除去することにより、第１の核酸塩基に対して第２の核酸塩基とは異なって影響を及ぼす手順を含むＤＮＡ解析法、例えばバイサルファイトシーケンシングまたはバイサルファイト変換と比べて改善された効率および／または費用をもたらす。

【０３２１】

一部の実施形態では、１つまたは複数のヌクレアーゼを使用して、非特異的に分配されたＤＮＡ分子を分解する。一部の実施形態では、部分試料を複数のヌクレアーゼと接触させる。部分試料をヌクレアーゼと逐次的にまたは同時に接触させることができる。ヌクレアーゼの同時使用は、ヌクレアーゼが同様の条件（例えば、緩衝液の組成）下で活性である場合、不必要な試料の操作を回避するために有利であり得る。部分試料を１種よりも多くのメチル化依存性制限酵素と接触させることにより、非特異的に分配された高メチル化ＤＮＡをより徹底的に分解することができる。部分試料を１種よりも多くのメチル化感受性制限酵素に接触させることにより、非特異的に分配された低メチル化および／または非メチル化ＤＮＡをより徹底的に分解することができる。

【０３２２】

一部の実施形態では、メチル化依存性ヌクレアーゼは、ＭｓｐＪＩ、ＬｐｎＰＩ、ＦｓｐＥＩ、またはＭｃｒＢＣのうちの１つまたは複数を含む。一部の実施形態では、少なくとも２種のメチル化依存性ヌクレアーゼを使用する。一部の実施形態では、少なくとも３種のメチル化依存性ヌクレアーゼを使用する。

【０３２３】

一部の実施形態では、メチル化感受性ヌクレアーゼは、ＡａｔＩＩ、ＡｃｃＩＩ、ＡｃｉＩ、Ａｏｒ１３ＨＩ、Ａｏｒ１５ＨＩ、ＢｓｐＴ１０４Ｉ、ＢｓｓＨＩＩ、ＢｓｔＵＩ、Ｃｆｒ１０Ｉ、ＣｌａＩ、ＣｐｏＩ、Ｅｃｏ５２Ｉ、ＨａｅＩＩ、ＨａｐＩＩ、ＨｈａＩ、Ｈｉｎ６Ｉ、ＨｐａＩＩ、ＨｐｙＣＨ４ＩＶ、ＭｌｕＩ、ＭｓｐＩ、ＮａｅＩ、ＮｏｔＩ、ＮｒｕＩ、ＮｓｂＩ、ＰｍａＣＩ、Ｐｓｐ１４０６Ｉ、ＰｖｕＩ、ＳａｃＩＩ、ＳａｌＩ、ＳｍａＩ、およびＳｎａＢＩのうちの１つまたは複数を含む。一部の実施形態では、少なくとも２種のメチル化感受性ヌクレアーゼを使用する。一部の実施形態では、少なくとも３種のメチル化感受性ヌクレアーゼを使用する。一部の実施形態では、メチル化感受性ヌクレアーゼは、ＢｓｔＵＩおよびＨｐａＩＩを含む。一部の実施形態では、２種のメチル化感受性ヌクレアーゼは、ＨｈａＩおよびＡｃｃＩＩを含む。一部の実施形態では、メチル化感受性ヌクレアーゼは、ＢｓｔＵＩ、ＨｐａＩＩおよびＨｉｎ６Ｉを含む。

【０３２４】

一部の実施形態では、ライブラリー調製の酵素ステップに備えて、ＤＮＡの分配を脱塩し、濃縮する。

Ｃ．アダプターライゲーション

【０３２５】

一部の実施形態では、アダプターをＤＮＡに付加する。これは、増幅手順と併せて、例

えば、アダプターをプライマーの５’部分に提供することによって行うことができる（ＰＣＲを使用する場合、これはライブラリープレップ－ＰＣＲまたはＬＰ－ＰＣＲと称され得る）。一部の実施形態では、アダプターをライゲーションなどの他のアプローチによって付加する。一部のそのような方法では、分配前または捕捉前に、第１のアダプターを核酸に、その３’末端へのライゲーションによって付加し、これは、一本鎖ＤＮＡへのライゲーションを含み得る。アダプターを、例えばユニバーサルプライマーおよびＤＮＡポリメラーゼを使用した第２の鎖の合成のためのプライミング部位として使用することができる。次いで、第２のアダプターを少なくとも今の二本鎖分子（the now double-stranded molecule）の第２の鎖の３’末端にライゲーションすることができる。一部の実施形態では、第１のアダプターはビオチンなどの親和性タグを含み、第１のアダプターにライゲーションされる核酸は、ストレプトアビジンなどの親和性タグの結合パートナーを含み得る固体支持体（例えば、ビーズ）に結合している。関連する手順のさらなる考察については、Gansauge et al., Nature Protocols 8: 737-748 (2013)を参照されたい。一本鎖核酸と適合する配列決定ライブラリー調製のための市販のキット、例えば、Ｓｗｉｆｔ ＢｉｏｓｃｉｅｎｃｅｓからのＡｃｃｅｌ－ＮＧＳ（登録商標）Ｍｅｔｈｙｌ－Ｓｅｑ ＤＮＡ Ｌｉｂｒａｒｙ Ｋｉｔが入手可能である。一部の実施形態では、アダプターライゲーション後、核酸を増幅させる。

【０３２６】

好ましくは、アダプターは、タグの組合せの数が、同じ開始および終止点を有する２つの核酸が同じタグの組合せを受容する低い確率、例えば、９５、９９または９９．９％をもたらす、十分な数の異なるタグを含む。アダプターは、同じタグを有するかまたは異なるタグを有するかにかかわらず、同じまたは異なるプライマー結合部位を含んでよいが、好ましくは、アダプターは、同じプライマー結合部位を含む。

【０３２７】

一部の実施形態では、アダプターの結合後、核酸を増幅に供する。増幅には、例えば、アダプター内のプライマー結合部位を認識するユニバーサルプライマーを使用することができる。

【０３２８】

一部の実施形態では、アダプターの結合後、ＤＮＡを分配し、これは、ＤＮＡを、エピジェネティック改変を有する核酸に優先的に結合する作用剤と接触させることを含む。核酸を、作用剤との結合から、核酸が改変を有する程度が異なる少なくとも２つの部分試料に分配する。例えば、作用剤が、改変を有する核酸に対する親和性を有する場合、改変が過剰発現した核酸（集団における中央値表示と比較して）は作用剤と優先的に結合し、一方で、改変が過少発現した核酸は作用剤と結合しないか、または作用剤からより容易に溶離される。次いで、核酸を、アダプター内のプライマー結合部位に結合するプライマーから増幅させることができる。その代わりに、分配をアダプター結合前に実施することができ、その場合、アダプターは、どの分配に分子が存在するかを同定する構成成分を含む差次的タグを含み得る。［０２１４］一部の実施形態では、核酸を、両方の末端において、プライマー結合部位およびタグを含むＹ形状アダプターに連結する。分子を増幅させる。

Ｄ．タグ付け

【０３２９】

ＤＮＡ分子の「タグ付け」は、ＤＮＡ分子にタグを結合させるまたは会合される手順である。タグは、タグが会合される分子の特色を示す情報を含有する分子、例えば核酸であり得る。例えば、分子は、試料タグ（１つの試料中の分子を異なる試料中の分子と識別する）または分子タグ／分子バーコード／バーコード（異なる分子を互いに識別する（一意的および非一意的タグ付けシナリオの両方において）を有し得る。分配ステップを伴う方法に関しては、分配タグ（１つの分配中の分子を異なる分配中の分子と識別する）を含めることができる。一部の実施形態では、ＤＮＡ分子に付加されるアダプターはタグを含む。ある特定の実施形態では、タグは、１つのバーコードまたはバーコードの組合せを含み得る。本明細書で使用される場合、「バーコード」という用語は、文脈に応じて、特定の

ヌクレオチド配列を有する核酸分子、またはヌクレオチド配列自体を指す。バーコードは、例えば、１０～１００個の間のヌクレオチドを有し得る。バーコードのコレクションは、具体的な目的に所望されるように、縮重配列を有し得るか、またはある特定のハミング距離を有する配列を有し得る。そのため、例えば、分子バーコードは、１つのバーコード、または各々が分子の異なる末端に付着する２つのバーコードの組合せから構成され得る。さらにまたはあるいは、異なる分配および／または試料について、バーコードが、個々の配列を通じて分子タグとしての機能を果たし、それらがメンバーであるセットに基づいて対応する分配および／または試料を同定する機能も果たすように、異なるセットの分子バーコード、または分子タグが、使用され得る。

【０３３０】

一部の実施形態では、２つまたはそれよりも多くの分配、例えば、各分配に差次的にタグ付けする。タグ（単数または複数）を特定の分配と相関付けるために、タグを使用して、個々のポリヌクレオチド集団分配を標識することができる。あるいは、タグは、分配ステップを用いない実施形態において使用することができる。一部の実施形態では、単一のタグを使用して、特定の分配を標識することができる。一部の実施形態では、複数の異なるタグを使用して、特定の分配を標識することができる。複数のタグを用いて特定の分配を標識する実施形態では、１つの分配を標識するために使用されるタグのセットは、他の分配を標識するために使用されるタグのセットと容易に区別することができる。一部の実施形態では、タグは、追加の機能を有してもよく、例えば、タグを、試料の起源に索引を付けるために使用するか、または一意的分子識別子（例えば、Kinde et al., Proc Nat'l Acad Sci USA 108: 9530-9535 (2011)、Kou et al., PLoS ONE, 11 : e0146638 (2016)と同様に、配列決定エラーを変異と区別することによって配列決定データの品質を改善するために使用することができる）として使用するか、または、例えば米国特許第９，５９８，７３１号に記載されるように、非一意的分子識別子として使用することができる。同様に、一部の実施形態では、タグは、追加の機能を有してもよく、例えば、タグを、試料の起源に索引を付けるために使用するか、または非一意的分子識別子（配列決定エラーを変異と区別することによって配列決定データの品質を改善するために使用することができる）として使用することができる。［０２１７］一部の実施形態では、分配タグ付けは、各分配中の分子を分配タグでタグ付けすることを含む。分配を再度組み合わせ（例えば、必要とされる配列決定実行数を減少させ、不必要な費用を回避するため）、分子を配列決定した後に、分配タグにより、起源の分配を同定する。一部の実施形態では、分配タグは、起源の分配および分子の識別子としての機能を果たし得る、すなわち、異なる分配は、例えば、バーコード対で構成される、分子タグの異なるセットでタグ付けされる。このように、分子に結合した１つまたは複数の分子バーコードは、起源の分配を示すだけでなく、分配内の分子を識別するためにも有用である。例えば、３５個のバーコードの第１のセットを使用して、第１の分配中の分子をタグ付けすることができ、一方で、３５個のバーコードの第２のセットを使用して、第２の分配中の分子をタグ付けすることができる。

【０３３１】

一部の実施形態では、分配および分配タグでのタグ付けの後に、分子は、単回でのシーケンシングのためにプールされ得る。一部の実施形態では、試料タグは、例えば、分配タグの付加およびプールの後のステップにおいて、分子に付加される。試料タグは、単回シーケンシングでのシーケンシングのために複数の試料から生成された材料をプールすることを容易にし得る。

【０３３２】

あるいは、一部の実施形態では、分配タグは、試料ならびに分配と相関付けられてもよい。単純な例として、第１のタグは、第１の試料の第１の分配を示し得、第２のタグは、第１の試料の第２の分配を示し得、第３のタグは、第２の試料の第１の分配を示し得、第４のタグは、第２の試料の第２の分配を示し得る。

【０３３３】

タグは、１つまたは複数の特徴に基づいて、既に分配された分子に付着され得るが、ライブラリー内の最終的なタグ付けされた分子は、もはやその特徴を有さない場合がある。例えば、一本鎖ＤＮＡ分子が、分配されタグ付けされ得るが、ライブラリー内の最終的なタグ付けされた分子は、二本鎖である可能性が高い。同様に、ＤＮＡは、異なるレベルのメチル化に基づく分配に供され得るが、最終的なライブラリーにおいて、これらの分子に由来するタグ付けされた分子は、非メチル化である可能性が高い。したがって、ライブラリー内の分子に付着したタグは、典型的には、必ずしもタグ付けされた分子自体の特徴ではなく、最終的なタグ付けされた分子が由来する「親分子」の特徴を示す。

【０３３４】

一例として、バーコード１、２、３、４などを使用して、第１の分配中の分子をタグ付けおよび標識し、バーコードＡ、Ｂ、Ｃ、Ｄなどを使用して、第２の分配中の分子をタグ付けおよび標識し、バーコードａ、ｂ、ｃ、ｄなどを使用して、第３の分配中の分子をタグ付けおよび標識する。差次的にタグ付けされた分配は、シーケンシングの前にプールされ得る。差次的にタグ付けされた分配は、別個にシーケンシングされるか、または例えば、Ｉｌｌｕｍｉｎａシーケンサーの同じフローセルにおいて、一緒に同時にシーケンシングされてもよい。

【０３３５】

配列決定後、リードの解析を分配ごとのレベル、ならびに全ＤＮＡ集団レベルで実施することができる。タグを使用して、異なる分配に由来する読み取りデータを選別する。解析には、配列情報、ゲノム座標の長さ、カバレッジ、および／またはコピー数を使用して、遺伝的変異およびエピジェネティック変異（メチル化、クロマチン構造などのうちの１つまたは複数）を決定するためのｉｎ　ｓｉｌｉｃｏ解析が含まれ得る。一部の実施形態では、より高いカバレッジは、ゲノム領域におけるより高いヌクレオソーム占有率と相関し得、一方でより低いカバレッジは、より低いヌクレオソーム占有率またはヌクレオソーム枯渇領域（ＮＤＲ）と相関し得る。

Ｅ．富化する／捕捉するステップ；増幅

【０３３６】

本明細書に開示される方法は、ＤＮＡ、例えばｃｆＤＮＡ標的領域を捕捉するステップを含み得る。一部の実施形態では、捕捉するステップは、ＤＮＡを標的領域に対して特異的なプローブ（例えば、オリゴヌクレオチド）と接触させることを含む。富化または捕捉は、本明細書に記載される任意の試料または部分試料に対して、当技術分野において公知の任意の好適なアプローチを使用して実施することができる。

【０３３７】

一部の実施形態では、富化または捕捉を、試料分子へのアダプターの結合後に実施する。一部の実施形態では、富化または捕捉を、分配ステップの後に実施する。一部の実施形態では、富化または捕捉を、増幅ステップの後に実施する。一部の実施形態では、試料分子を分配し、次いでアダプターを結合させ、次いで試料分子を増幅させ、次いで増幅させた分子を富化または捕捉に供する。次いで、富化または捕捉された分子を別の増幅に供し、次いで配列決定することができる。

【０３３８】

一部の実施形態では、標的領域に対して特異的なプローブは、プローブとハイブリダイズしたＤＮＡの富化または捕捉を容易にする捕捉部分を含む。一部の実施形態では、捕捉部分はビオチンである。一部のそのような実施形態では、磁気ビーズなどの固体支持体に結合させたストレプトアビジンを、ビオチンに結合させるために使用する。標的領域を含まない、非特異的に結合したＤＮＡを、捕捉されたＤＮＡから洗い流す。一部の実施形態では、次いで、ＤＮＡをプローブから解離させ、塩洗浄または別のＤＮＡ変性剤を含む緩衝液を使用して固体支持体から溶離させる。一部の実施形態では、例えば、ビオチン－ストレプトアビジン相互作用を破壊することによっても同様にプローブが固体支持体から溶離する。一部の実施形態では、捕捉されたＤＮＡを、固体支持体からの溶離後に増幅させる。一部のそのような実施形態では、アダプターを含むＤＮＡを、アダプターとアニーリ

ングするＰＣＲプライマーを使用して増幅させる。一部の実施形態では、捕捉されたＤＮＡを、固体支持体に結合させている間に増幅させる。一部のそのような実施形態では、増幅は、アダプター内の配列とアニーリングするＰＣＲプライマー、およびＤＮＡの標的領域とアニーリングしたプローブ内の配列とアニーリングするＰＣＲプライマーの使用を含む。

【０３３９】

一部の実施形態では、本明細書の方法は、エピジェネティックおよび／または配列可変標的領域を含むＤＮＡを富化させるまたは捕捉するステップを含む。そのような領域を試料（例えば、アダプターの結合および増幅を受けた試料）のアリコートから捕捉することができ、一方で、試料の別々のアリコートに対して、改変シトシン、例えばメチルシトシンを認識する作用剤を用いてＤＮＡを分配するステップを実施する。エピジェネティックおよび／または配列可変標的領域を含むＤＮＡを富化させるまたは捕捉するステップは、ＤＮＡを、標的特異的プローブの第１のまたは第２のセットと接触させることを含み得る。そのような標的特異的プローブは、これだけに限定されないが、上記の実施形態および下記のプローブに関する節を含め、本明細書に記載される、標的特異的プローブのセットについての特色のいずれかを有し得る。捕捉するステップは、本明細書に開示される方法の間に調製された１つまたは複数の部分試料に対して実施することができる。一部の実施形態では、ＤＮＡを、第１の部分試料または第２の部分試料、例えば、第１の部分試料と第２の部分試料から捕捉する。一部の実施形態では、部分試料を差次的にタグ付けし（例えば、本明細書に記載されるように）、次いで、プールした後、捕捉に供する。エピジェネティックおよび／または配列可変標的領域を含むＤＮＡを捕捉するための例示的な方法は、例えば、これにより参照により本明細書に組み込まれるＷＯ２０２０／１６０４１４に見出すことができる。

【０３４０】

捕捉するステップは、一般的に長さ、塩基組成などのプローブの特色にある程度依存する特異的核酸ハイブリダイゼーションにとって好適な条件を使用して実施され得る。当業者は、核酸ハイブリダイゼーションに関して当技術分野で一般的な知識を考慮して適切な条件を熟知している。一部の実施形態では、標的特異的プローブおよびＤＮＡの複合体が形成される。

【０３４１】

一部の実施形態では、本明細書に記載される方法は、対象から得られたｃｆＤＮＡの標的領域の複数のセットを捕捉するステップを含む。標的領域は、それらが腫瘍を起源とするかもしくは健康な細胞を起源とするか、またはある特定の細胞型を起源とするのかに応じて相違を含み得る。捕捉するステップにより、捕捉されたｃｆＤＮＡ分子のセットが産生される。一部の実施形態では、配列可変標的領域セットに対応するｃｆＤＮＡ分子は、捕捉されたｃｆＤＮＡ分子のセットでは、エピジェネティック標的領域セットに対応するｃｆＤＮＡ分子よりも高い捕捉収率で捕捉される。一部の実施形態では、本明細書に記載される方法は、対象から得られたｃｆＤＮＡを標的特異的プローブのセットと接触させるステップを含み、ここで、標的特異的プローブのセットは、配列可変標的領域セットに対応するｃｆＤＮＡを、エピジェネティック標的領域セットに対応するｃｆＤＮＡよりも高い捕捉収率で捕捉するように構成されている。捕捉するステップ、捕捉収率、および関連する態様の追加の考察については、あらゆる目的に関して参照により本明細書に組み込まれるＷＯ２０２０／１６０４１４を参照されたい。

【０３４２】

配列可変標的領域を十分な信頼度または精度で解析するためには、エピジェネティック標的領域を解析するために必要であり得るよりも大きい深度のシーケンシングが必要であり得ることから、配列可変標的領域セットに対応するｃｆＤＮＡを、エピジェネティック標的領域セットに対応するｃｆＤＮＡよりも高い捕捉収率で捕捉することは有益であり得る。断片化パターン（例えば、転写開始部位またはＣＴＣＦ結合部位の乱れ（ｐｅｒｔｕｒｂａｔｉｏｎ）に関して試験するために）または断片の存在量（例えば、高メチル化分配および

低メチル化分配における）を決定するために必要なデータ量は、一般に、がん関連配列変異の存在または非存在を決定するために必要なデータ量よりも少ない。標的領域セットを異なる収率で捕捉することにより、同じ配列決定実行において（例えば、プールされた混合物を使用するおよび／または同じ配列決定セル中で）、配列決定の異なる深度まで標的領域を配列決定することが容易になり得る。

【０３４３】

一部の実施形態では、ＤＮＡを増幅させる。一部の実施形態では、増幅を、捕捉するステップの前に実施する。一部の実施形態では、増幅を、捕捉するステップの後に実施する。一部の実施形態では、増幅を、捕捉するステップの前後に実施する。様々な実施形態では、方法は、本明細書における考察と一貫して、捕捉されたＤＮＡを、例えば、エピジェネティックおよび配列可変標的領域セットに関して種々の程度の配列決定深度まで配列決定するステップをさらに含む。

【０３４４】

一部の実施形態では、捕捉するステップは、同じ容器中で同時に、配列可変標的領域セットのプローブおよびエピジェネティック標的領域セットのプローブについて実施され、例えば配列可変標的領域セットおよびエピジェネティック標的領域セットのプローブは、同じ組成物中にある。このアプローチにより、比較的合理化されたワークフローがもたらされる。一部の実施形態では、配列可変標的領域セットに対するプローブの濃度は、エピジェネティック標的領域セットに対するプローブの濃度よりも高い。

【０３４５】

あるいは、捕捉するステップを、第１の容器中の配列可変標的領域プローブセットおよび第２の容器中のエピジェネティック標的領域プローブセットを用いて実施するか、または接触させるステップを、配列可変標的領域プローブセットを第１の時間および第１の容器で用い、エピジェネティック標的領域プローブセットを第１の時間の前またはその後の第２の時間で用いて実施する。このアプローチにより、別々の、配列可変標的領域セットに対応する捕捉されたＤＮＡを含む第１の組成物とエピジェネティック標的領域セットに対応する捕捉されたＤＮＡを含む第２の組成物を調製することが可能になる。組成物を別個に所望の通り処理し（例えば、本明細書に記載されるようにメチル化に基づいて分配するために）、さらなる処理および解析、例えば配列決定のための材料をもたらすために適した割合でプールすることができる。

【０３４６】

一部の実施形態では、アダプターは本明細書に記載されるＤＮＡに含まれる。一部の実施形態では、バーコードであってもバーコードを含んでいてもよいタグはＤＮＡに含まれる。一部の実施形態では、そのようなタグはアダプターに含まれる。タグは、核酸の起源の同定を容易にし得る。例えば、並列シーケンシングのために複数の試料をプールした後に、バーコードを使用して、ＤＮＡが由来する起源（例えば、対象）を同定することが可能であり得る。これは、例えば、本明細書に記載のように、例えば、プライマーの５’部分にバーコードを提供することによって、増幅手順と同時に行うことができる。一部の実施形態では、アダプターおよびタグ／バーコードは、同じプライマーまたはプライマーセットによって提供される。例えば、バーコードは、アダプターの３’およびプライマーの標的にハイブリダイズする部分の５’に位置付けられ得る。あるいは、バーコードは、他のアプローチ、例えば、ライゲーションによって、必要に応じて同じライゲーション基質においてアダプターと一緒に、付加され得る。

【０３４７】

増幅、タグ、およびバーコードに関する追加の詳細が本明細書で考察されており、これらの実施形態のいずれかと実行可能な程度に組み合わせることができる。

Ｆ．ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順またはメチル化感受性変換方法

【０３４８】

一部の実施形態では、本明細書に開示される方法は、ＤＮＡを、ＤＮＡ内の第１の核酸

塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順に供するステップを含み、ここで、第１の核酸塩基は、改変または非改変核酸塩基であり、第２の核酸塩基は、第１の核酸塩基とは異なる改変または非改変核酸塩基であり、第１の核酸塩基と第２の核酸塩基は同じ塩基対形成特異性を有する。一部の実施形態では、手順は、第１の核酸塩基または第２の核酸塩基を化学的に変換し、その結果、変換された核酸塩基の塩基対形成特異性が変更される。一部の実施形態では、第１の核酸塩基が改変または非改変アデニンである場合には、第２の核酸塩基は改変または非改変アデニンであり、第１の核酸塩基が改変または非改変シトシンである場合には、第２の核酸塩基は改変または非改変シトシンであり、第１の核酸塩基が改変または非改変グアニンである場合には、第２の核酸塩基は改変または非改変グアニンであり、第１の核酸塩基が改変または非改変チミンである場合には、第２の核酸塩基は改変または非改変チミンである（このステップの目的に関しては、改変および非改変ウラシルは改変チミンに包含される）。

【０３４９】

一部の実施形態では、第１の核酸塩基は改変または非改変シトシンであり、そうすると第２の核酸塩基は改変または非改変シトシンである。例えば、第１の核酸塩基は非改変シトシン（Ｃ）を含んでよく、第２の核酸塩基は、５－メチルシトシン（ｍＣ）および５－ヒドロキシメチルシトシン（ｈｍＣ）のうちの１つまたは複数を含んでよい。あるいは、第２の核酸塩基は、Ｃを含み得、第１の核酸塩基は、ｍＣおよびｈｍＣのうちの１つまたは複数を含み得る。例えば、上記の発明の概要および以下の考察において示されるように、第１および第２の核酸塩基のうちの一方がｍＣを含み、他方がｈｍＣを含む場合など、他の組合せもまた可能である。

【０３５０】

一部の実施形態では、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順は、バイサルファイト変換を含む。バイサルファイトを用いた処理により、非改変シトシンおよびある特定の改変シトシン（例えば、５－ホルミルシトシン（ｆＣ）または５－カルボキシルシトシン（ｃａＣ））がウラシルに変換され、一方で、他の改変シトシン（例えば、５－メチルシトシン、５－ヒドロキシルメチルシトシン）は変換されない。バイサルファイト変換を実施することで、配列リードを使用して、ｍＣまたはｈｍＣを含有する位置を同定することを容易にすることができる。バイサルファイト変換の例示的な説明については、例えば、Moss et al., Nat Commun. 2018; 9: 5068を参照されたい。［０２３８］一部の実施形態では、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順は、酸化的バイサルファイト（Ｏｘ－ＢＳ）変換を含む。Ｏｘ－ＢＳ変換を実施することで、配列リードを使用して、ｍＣを含有する位置を同定することを容易にすることができる。酸化的バイサルファイト変換の例示的な説明については、例えば、Booth et al., Science 2012; 336: 934-937を参照されたい。

【０３５１】

一部の実施形態では、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順は、Ｔｅｔ補助バイサルファイト（ＴＡＢ）変換を含む。例えば、Yu et al., Cell 2012; 149: 1368-80に記載されているように、ｂ－グルコシルトランスフェラーゼを使用して、ｈｍＣを保護することができ（５－グルコシルヒドロキシメチルシトシン（ｇｈｍＣ）を形成する）、次いで、ｍＴｅｔｌなどのＴＥＴタンパク質を使用して、ｍＣをｃａＣに変換することができ、次いで、バイサルファイト処理を使用して、ＣおよびｃａＣをＵに変換する一方でｇｈｍＣには影響を及ぼさないままにすることができる。したがって、ＴＡＢ変換を使用する場合、第１の核酸塩基は、非改変シトシン、ｆＣ、ｃａＣ、ｍＣ、または他のバイサルファイトによる影響を受けるシトシン形態のうちの１つまたは複数を含み、第２の核酸塩基は、ｈｍＣを含む。ＴＡＢ変換を実施することで、配列リードを使用して、ｈｍＣを含有する位置を同定することを容易にすることができる。［０２４０］一部の実施形態では、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順は、置換ボラン還元剤を

用いたＴｅｔ補助変換を含み、必要に応じて、置換ボラン還元剤は、２－ピコリンボラン、ボランピリジン、ｔｅｒｔ－ブチルアミンボラン、またはアンモニアボランである。例えば、Liu et al., Nature Biotechnology 2019; 37:424-429（例えば、補足の図１および補足説明７）を参照されたい。ＴＡＰ変換を実施することで、配列リードを使用して、非改変Ｃを含有する位置を同定することを容易にすることができる。この手順は、Liu et al. 2019、上記においてさらに詳細に記載されているＴｅｔ補助ピリジンボラン配列決定（ＴＡＰＳ）を包含する。

【０３５２】

あるいは、ｈｍＣの保護（例えば、ｂＯＴを使用する）を、置換ボラン還元剤を用いたＴｅｔ補助変換と組み合わせることができる。そのようなＴＡＰＳＰ変換を実施することで、配列リードを使用して、一方で非改変ＣまたはｈｍＣを含有する位置とｍＣを含有する位置とを識別することを容易にすることができる。このタイプの変換の例示的な説明については、例えば、Liu et al., Nature Biotechnology 2019; 37: 424-429を参照されたい。

【０３５３】

一部の実施形態では、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順は、ＡＰＯＢＥＣカップリングエピジェネティック（ＡＣＥ）変換を含む。ＡＣＥ変換を実施することで、配列リードを使用して、ｈｍＣを含有する位置とｍＣまたは非改変Ｃを含有する位置とを識別することを容易にすることができる。ＡＣＥ変換の例示的な説明については、例えば、Schutsky et al., Nature Biotechnology 2018; 36: 1083-1090を参照されたい。

【０３５４】

一部の実施形態では、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順は、例えば、ＥＭ－Ｓｅｑと同様に第１の核酸塩基の酵素的変換を含む。例えば、Vaisvila R, et al. (2019) EM-seq: Detection of DNA methylation at single base resolution from picograms of DNA. bioRxiv; DOI [0, available at www.biorxiv.org/content/10.1101/2019.12.20.884692vlを参照されたい。

【０３５５】

一部の実施形態では、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順は、５ｍＣを検出するための直接メチル化配列決定（ＤＭ－Ｓｅｑ）を伴う方法を含む。ＤＭ－Ｓｅｑに関連する例示的な方法および組成物は、例えば、これにより参照により本明細書に組み込まれる、ＷＯ２０２３／２８８２２２およびＷＯ２０２１／２３６７７８に見出すことができる。

【０３５６】

一部の実施形態では、第１の核酸塩基は、改変または非改変アデニンであり、第２の核酸塩基は、改変または非改変アデニンである。一部の実施形態では、改変アデニンは、N^6－メチルアデニン（ｍＡ）である。一部の実施形態では、改変アデニンは、N^6－メチルアデニン（ｍＡ）、N^6－ヒドロキシメチルアデニン（ｈｍＡ）、またはN^6－ホルミルアデニン（ｆＡ）のうちの１つまたは複数である。

Ｇ．捕捉されたセット；標的領域

【０３５７】

一部の実施形態では、本明細書に記載される方法を使用して捕捉または富化された核酸は、捕捉されたＤＮＡ、例えば１つまたは複数の捕捉されたＤＮＡのセットを含む。一部の実施形態では、捕捉されたＤＮＡは、異なる免疫細胞タイプにおいて差次的にメチル化されている標的領域を含む。一部の実施形態では、免疫細胞タイプは、希少なまたは密接に関連する免疫細胞タイプ、例えば、異なる分化段階にある活性化されたおよびナイーブなリンパ球または骨髄性細胞を含む。

【０３５８】

一部の実施形態では、１つの試料または第１の部分試料から捕捉された、捕捉されたエ

ピジェネティック標的領域セットは、高メチル化標的領域を含む。一部の実施形態では、高メチル化標的領域は、１つの細胞型においてまたは１つの免疫細胞タイプにおいて、またはある１つのクラスター内の１つの免疫細胞タイプにおいて差次的にまたは排他的に高メチル化されている。一部の実施形態では、高メチル化標的領域は、１つの細胞型または１つの免疫細胞タイプまたはある１つのクラスター内の１つの免疫細胞タイプにおいて識別可能により高いまたは排他的に存在する程度まで高メチル化されている。そのような高メチル化標的領域は、他の細胞型においても高メチル化され得るが、１つの細胞型において観察される程度までではない。一部の実施形態では、高メチル化標的領域は、健康なｃｆＤＮＡにおいて、少なくとも１つの他の組織タイプにおけるよりも低いメチル化を示す。

【０３５９】

一部の実施形態では、１つの試料または第２の部分試料から捕捉された、捕捉されたエピジェネティック標的領域セットは、低メチル化標的領域を含む。一部の実施形態では、低メチル化標的領域は、１つの細胞型においてまたは１つの免疫細胞タイプにおいてまたはある１つのクラスター内の１つの免疫細胞タイプにおいて排他的に低メチル化されている。一部の実施形態では、低メチル化標的領域は、１つの細胞型または１つの免疫細胞タイプにおいて、またはある１つのクラスター内の１つの免疫細胞タイプにおいて排他的に存在する程度まで低メチル化されている。

【０３６０】

そのような低メチル化標的領域は、他の細胞型においても低メチル化され得るが、１つの細胞型において観察される程度までではない。一部の実施形態では、低メチル化標的領域は、健康なｃｆＤＮＡにおいて少なくとも１つの他の組織タイプにおけるよりも高いメチル化を示す。［０２４８］いずれの特定の理論によっても束縛されることを望むものではないが、がん、増殖しているまたは活性化された免疫細胞（および潜在的にがん細胞も）を有する個体は、健康な個体における免疫細胞（およびそれぞれ同じ組織タイプの健康な細胞）よりも多くのＤＮＡを血流中に脱落させ得る。したがって、ｃｆＤＮＡの細胞型および／または起源組織の分布は、発癌すると変化し得る。例えば、がん、前がん状態、感染、移植片拒絶、または免疫系に直接的もしくは間接的に影響を及ぼす他の疾患もしくは障害を有する対象では、起源の免疫細胞タイプの分布が変化し得る。ある特定の免疫細胞タイプのエピジェネティック標的領域の状態も同様に、そのような疾患を有する対象では健康な対象と比べてまたは疾患もしくは障害を有する前の同じ対象と比べて変化し得る。したがって、高メチル化および／または低メチル化の変動は、疾患の指標になり得る。例えば、分配ステップ後の部分試料における高メチル化標的領域および／または低メチル化標的領域のレベルの上昇は、がんの存在（または対象の既往に応じて、再発）の指標になり得る。

【０３６１】

これだけに限定されないが、免疫細胞タイプを含む、様々な細胞型間を識別するために有用な例示的な高メチル化標的領域および低メチル化標的領域が、例えば、Stunnenberg, H. G. et. al., "The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery," Cell 167, 1145 (2016)(doi.org/10.1186/sl3059-020-02065-5)に記載されているように、全ゲノムバイサルファイトシーケンシングによって様々な細胞型から得られたＤＮＡを解析することによって同定されている。全ゲノムバイサルファイトシーケンシングデータは、インターネットでdcc.blueprint-epigenome.euにおいて入手可能なＢｌｕｅｐｒｉｎｔ　ｃｏｎｓｏｒｔｉｕｍから入手可能である。

【０３６２】

一部の実施形態では、例えば、ＷＯ２０２０／１６０４１４に記載されているように、第１の捕捉された標的領域セットおよび第２の捕捉された標的領域セットは、それぞれ、配列可変標的領域セットに対応するＤＮＡおよびエピジェネティック標的領域セットに対応するＤＮＡを含む。第１の捕捉されたセットと第２の捕捉されたセットを組み合わせて

、組合せの捕捉されたセットをもたらすことができる。
【０３６３】
ＤＮＡ（例えば、試料または部分試料）が、バイサルファイト変換、デアミナーゼを用いた処理、またはある特定の塩基の塩基対形成特異性を変更する本明細書で言及されている他のそのような手順のいずれかなどの手順に供されている場合、富化または捕捉には、所望の通り、変更されたまたは変更されていない配列に対して特異的なオリゴヌクレオチド（例えば、プライマーまたはプローブ）を使用することができる。
【０３６４】
配列可変標的領域セットに対応するＤＮＡおよびエピジェネティック標的領域セットに対応するＤＮＡを含む捕捉されたセットが、上記で考察された組合せの捕捉されたセットを含む一部の実施形態では、配列可変標的領域セットに対応するＤＮＡは、エピジェネティック標的領域セットに対応するＤＮＡよりも高い濃度、例えば、１．１倍～１．２倍高い濃度、１．２倍～１．４倍高い濃度、１．４倍～１．６倍高い濃度、１．６倍～１．８倍高い濃度、１．８倍～２．０倍高い濃度、２．０倍～２．２倍高い濃度、２．２倍～２．４倍高い濃度、２．４倍～２．６倍高い濃度、２．６倍～２．８倍高い濃度、２．８倍～３．０倍高い濃度、３．０倍～３．５倍高い濃度、３．５倍～４．０、４．０倍～４．５倍高い濃度、４．５倍～５．０倍高い濃度、５．０倍～５．５倍高い濃度、５．５倍～６．０倍高い濃度、６．０倍～６．５倍高い濃度、６．５倍～７．０倍高い濃度、７．０倍～７．５倍高い濃度、７．５倍～８．０倍高い濃度、８．０倍～８．５倍高い濃度、８．５倍～９．０倍高い濃度、９．０倍～９．５倍高い濃度、９．５倍～１０．０倍高い濃度、１０倍～１１倍高い濃度、１１倍～１２倍高い濃度、１２倍～１３倍高い濃度、１３倍～１４倍高い濃度、１４倍～１５倍高い濃度、１５倍～１６倍高い濃度、１６倍～１７倍高い濃度、１７倍～１８倍高い濃度、１８倍～１９倍高い濃度、１９倍～２０倍高い濃度、２０倍～３０倍高い濃度、３０倍～４０倍高い濃度、４０倍～５０倍高い濃度、５０倍～６０倍高い濃度、６０倍～７０倍高い濃度、７０倍～８０倍高い濃度、８０倍～９０倍高い濃度、または９０倍～１００倍高い濃度で存在し得る。濃度の差の程度により、定義の節において考察されるように、標的領域のフットプリントサイズに関する正規化が説明される。
１．エピジェネティック標的領域セット
【０３６５】
一部の実施形態では、エピジェネティック標的領域セットは、種々の免疫細胞タイプに由来するＤＮＡと他の非免疫細胞タイプに由来するＤＮＡとを区別する可能性がある、および／または、新生物（例えば、腫瘍またはがん）細胞に由来するＤＮＡと、健康な細胞、例えば非新生物循環細胞に由来するＤＮＡとを区別する可能性がある、１つまたは複数のタイプの標的領域を含み得る。そのような領域の例示的なタイプは、本明細書において詳細に考察されている。エピジェネティック標的領域セットはまた、例えば、本明細書に記載される１つまたは複数の対照領域も含み得る。
【０３６６】
一部の実施形態では、エピジェネティック標的領域セットは、少なくとも１００ｋｂ、例えば、少なくとも２００ｋｂ、少なくとも３００ｋｂ、または少なくとも４００ｋｂのフットプリントを有する。一部の実施形態では、エピジェネティック標的領域セットは、１００～１０００ｋｂ、例えば、１００～２００ｋｂ、２００～３００ｋｂ、３００～４００ｋｂ、４００～５００ｋｂ、５００～６００ｋｂ、６００～７００ｋｂ、７００～８００ｋｂ、８００～９００ｋｂ、および９００～１，０００ｋｂの範囲のフットプリントを有する。
ａ．高メチル化標的領域
【０３６７】
一部の実施形態では、エピジェネティック標的領域セットは、１つまたは複数の高メチル化標的領域を含む。一部の実施形態では、高メチル化標的領域は、１つの免疫細胞タイプにおいて排他的に高メチル化されているか、または、１つの免疫細胞タイプにおいて、

任意の他の免疫細胞タイプにおけるよりも、もしくは同じ免疫細胞クラスター内の任意の他の免疫細胞タイプにおけるよりも大きな程度まで高メチル化されている。一部のそのような実施形態では、高メチル化標的領域により、活性化Ｂ細胞（メモリーＢ細胞および形質細胞を含む）、活性化Ｔ細胞（調節性Ｔ細胞（Ｔｒｅｇ）、ＣＤ４エフェクターメモリーＴ細胞、ＣＤ４セントラルメモリーＴ細胞、ＣＤ８エフェクターメモリーＴ細胞、およびＣＤ８セントラルメモリーＴ細胞を含む）、ならびにナチュラルキラー（ＮＫ）細胞などの希少な免疫細胞タイプを含む、ＤＮＡの起源である特定の免疫細胞タイプのレベルが示される。免疫細胞タイプをデコンボリューションするために有用な高メチル化標的領域のメチル化パターンは、ある特定の疾患状態、例えばがんにおいてはさらに変化し得る。したがって、一部の実施形態では、免疫細胞タイプをデコンボリューションするために有用な高メチル化標的領域は、試料を得たところの対象が、がんまたは前がん状態を有する可能性を決定するためにも有用である。一部のそのような実施形態では、高メチル化標的領域は、特定の免疫細胞タイプのレベルが異常であるかどうか、およびそのような異常なレベルが、がんの存在もしくは前がん状態に関連する可能性があるかどうか、またはそのような異常なレベルが、がんもしくは前がん状態以外の種々の疾患もしくは状態に関連するかどうかを決定するために有用である。

【０３６８】

一部の実施形態では、ある特定の高メチル化標的領域は、腫瘍またはがん細胞などの新生物細胞によって産生されるＤＮＡにおいて、観察されるメチル化のレベルの上昇を示す、例えば、高メチル化されている。そのような高メチル化標的領域の検出により、例えば、免疫細胞タイプを示す高メチル化標的領域の検出と併せて、本明細書に記載される方法の特異度および／または感度をさらに増加させることができる。一部の実施形態では、そのような、高メチル化標的領域において観察されるメチル化の増加により、試料（例えばｃｆＤＮＡのもの）が、がんを有する対象から得られた可能性の増加が示される。例えば、腫瘍抑制因子遺伝子のプロモーターの高メチル化が繰り返し観察されている。例えば、Kang et ah, Genome Biol. 18: 53 (2017)およびそこで引用されている参考文献を参照されたい。別の例では、上記で考察されたように、高メチル化標的領域は、がん性組織において、同じタイプの健康な組織に由来するＤＮＡと比べて必ずしもメチル化が異なるわけではないが、健康な対象において典型的であるｃｆＤＮＡと比べてメチル化が異なる（例えば、より多くのメチル化を有する）領域を含み得る。例えば、がんの存在により、細胞死、例えば、がんに対応する組織タイプの細胞のアポトーシスの増加がもたらされる場合、そのようながんは、少なくとも部分的に、そのような高メチル化標的領域を使用して検出することができる。一部の実施形態では、対象ががんを有する可能性を決定するために有用な高メチル化標的領域は、特定の免疫細胞タイプのレベルを決定するために有用な高メチル化標的領域とは異なる。一部の実施形態では、対象ががんを有する可能性を決定するために有用な高メチル化標的領域の少なくとも一部は、特定の免疫細胞タイプのレベルを決定するために有用な高メチル化標的領域と同じである。

【０３６９】

結腸直腸がんにおけるメチル化可変標的領域に関する広範な考察が、Lam et al., Biochim Biophys Acta. 1866: 106-20 (2016)に提供されている。これらには、ＶＩＭ、ＳＥＰＴ９、ＩＴＧＡ４、ＯＳＭ４、ＧＡＴＡ４およびＮＤＲＧ４が含まれる。結腸直腸がん（ＣＲＣ）研究に基づく例示的な高メチル化標的領域のセットを表１に提供する。これらの遺伝子の多くは、結腸直腸がん以外のがんとも関連性を有する可能性がある。例えば、ＴＰ５３は、極めて重要な腫瘍抑制因子として広範に理解されており、この遺伝子の高メチル化に基づく不活化は、一般的な発癌機序であり得る。

表１．ＣＲＣ研究に基づく例示的な高メチル化標的領域。

【表１】

遺伝子名	さらなる遺伝子名	染色体
VIM		chr10
SEPT9		chr17
CYCD2	CCND2	chr12
TFPI2		chr7
GATA4		chr8
RARB2	RARB	chr3
p16INK4a	CDKN2A	chr9
MGMT		chr10
APC		chr5
NDRG4		chr16
HLTF		chr3
HPP1	TMEFF2	chr2
hMLH1	MLH1	chr3
RASSF1A	RASSF1	chr3
CDH13		chr16
IGFBP3		chr7
ITGA4		chr2

【０３７０】

一部の実施形態では、高メチル化標的領域は、表１に列挙されている複数の座、例えば、表１に列挙されている座の少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、または１００％を含む。例えば、標的領域として含まれる各座に関して、遺伝子の転写開始部位と終止コドン（選択的にスプライシングされる遺伝子の最後の終止コドン）の間、または遺伝子のプロモーター領域に結合するハイブリダイゼーション部位を有する１つまたは複数のプローブが存在し得る。一部の実施形態では、１つまたは複数のプローブは、表１の遺伝子の転写開始部位から３００ｂｐ以内、例えば、２００ｂｐまたは１００ｂｐ以内に結合する。

【０３７１】

様々なタイプの肺がんにおけるメチル化可変標的領域が、例えば、Ooki et al., Clin. Cancer Res. 23: 7141-52 (2017)；Belinksy, Annu. Rev. Physiol. 77: 453-74 (2015)；Hulbert et al., Clin. Cancer Res. 23: 1998-2005 (2017)；Shi et al., BMC Genomics 18: 901 (2017)；Schneider et al., BMC Cancer. 11: 102 (2011)；Lissa et al., Transl Lung Cancer Res 5 (5): 492-504 (2016)；Skvortsova et al., Br. J. Cancer. 94 (10): 1492-1495 (2006)；Kim et al., Cancer Res. 61: 3419-3424 (2001)；Furonaka et al., Pathology International 55: 303-309 (2005)；Gomes et al., Rev. Port. Pneumol. 20: 20-30 (2014)

；Kim et al., Oncogene. 20: 1765-70 (2001)；Hopkins－Donaldson et al., Cell Death Differ. 10: 356-64 (2003)；Kikuchi et al., Clin. Cancer Res. 11: 2954-61 (2005)；Heller et al., Oncogene 25: 959-968 (2006)；Licchesi et al., Carcinogenesis. 29: 895-904 (2008)；Guo et al., Clin. Cancer Res. 10: 7917-24 (2004)；Palmisano et al., Cancer Res. 63: 4620-4625 (2003)；およびToyooka et al., Cancer Res. 61: 4556-4560 (2001)において詳細に考察されている。

【０３７２】

肺がん研究に基づく例示的な高メチル化標的領域のセットを表２に提供する。これらの遺伝子の多くは、肺がん以外のがんとも関連性を有する可能性がある。例えば、Ｃａｓｐ８（カスパーゼ８）は、プログラム細胞死における重要な酵素であり、この遺伝子の高メチル化に基づく不活化は、肺がんに限らず一般的な発癌機序であり得る。さらに、いくつかの遺伝子が表１および２の両方に出現し、一般性が示される。

表２．肺がん研究に基づく例示的な高メチル化標的領域

【表２】

遺伝子名	染色体
MARCH11	chr5
TAC1	chr7
TCF21	chr6
SHOX2	chr3
p16	chr3
Casp8	chr2
CDH13	chr16
MGMT	chr10
MLH1	chr3
MSH2	chr2
TSLC1	chr11
APC	chr5
DKK1	chr10
DKK3	chr11
LKB1	chr11
WIF1	chr12
RUNX3	chr1
GATA4	chr8
GATA5	chr20
PAX5	chr9
E-Cadherin	chr16
H-Cadherin	chr16

【０３７３】

表２において同定される標的領域に関する上記の実施形態のいずれかを、表１において同定される標的領域に関する上記の実施形態のいずれかと組み合わせることができる。一部の実施形態では、高メチル化標的領域は、表１または表２に列挙されている複数の座、例えば、表１または表２に列挙されている座の少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、または１００％を含む。

【０３７４】

一部の実施形態では、高メチル化標的領域は、表２ｂに列挙されている１つまたは複数の遺伝子、例えば、表３に列挙されている少なくとも５種、１０種、２０種、３０種、４０種、５０種、６０種、７０種、８０種、９０種、１００種、１５０種、２００種、２５０種、３００種、３５０種、４００種、４５０種、５００種、５５０種、６００種、６５０種、７００種、７５０種、８００種、８５０種、９００種、９５０種、１０００種、１０５０種、

【０３７５】
１０００種、１１００種、１１５０種または１２００種の遺伝子の領域を含む。これらの遺伝子の高メチル化は、ＤＮＡ試料に対する免疫細胞からの寄与を検出するために有用であり得る。一部の実施形態では、高メチル化標的領域は、表２ｂに列挙されている複数の遺伝子、例えば、表３に列挙されている遺伝子の少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、または１００％の領域を含む。一部の実施形態では、高メチル化標的領域は、表３に列挙されている全ての遺伝子の領域を含む。

表３．例示的な高メチル化標的領域を含む例示的な遺伝子

【表３－１】

遺伝子						
CCL5	RSAD2	DBNDD1	CD80	CD2	ITPK1	PRR5
CD19	ISG15	SNRPB	CEACAM4	G0S2	JAG1	PTGS2
CD37	MMP3	MNAT1	SH2D1A	MT1G	MMRN2	PTPN11
CD3D	MARCKSL1	NUP35	SAA2-SAA4	CARD16	NFIL3	PVRIG
CD3E	IGF2R	TCEB1	TMEM150B	MT1H	PDGFB	RELA
CD3G	LAMB1	HSPB11	PYHIN1	SERPINA1	PGPEP1	RICTOR
CD3Z	SPARC	C19orf48	CCL7	ACP5	RPL7A	RIPK2
CD79A	FN1	ID3	PDCD1	CXCL5	SERPINB5	SPP1
CD79B	ITGA4	IPO4	BNC1	CD74	SERPINH1	STAT4
CD8A	SMO	FARSB	ICOS	C1QB	STC1	TNFRSF14
CD8B1	MMP19	EIF4G1	ITK	HLA-DRB5	TNFAIP6	TNFRSF25
IGHG3	ITGB8	SKA1	CLEC4E	FADS3	VCAN	TNFSF9
IGJ	ITGA5	MFSD11	GZMK	IL15RA	VEGFA	TRAT1
IGLC1	NID1	PLAUR	UBASH3A	MFI2	VEGFB	TSLP
CD14	TIMP1	MARVELD2	CRTAM	IL4I1	VEGFC	VTCN1
LCK	SEMA3F	MCM3	EOMES	HLA-A	ATF3	CCL13
LTB	RHOQ	DHFR	P2RY10	TM4SF19	BATF3	CCL14
MS4A1	CTNNB1	RNF41	NCF1	MMP9	CD1C	CCL18
CORO1A	MMP2	ID2	CARD17	ACY3	CD36	CCL19
MNDA	SERPINE1	H2AFZ	ZBP1	KRT14	CD4	CCL20
CCRL2	EPHB2	CDK2	TNFRSF9	MT1X	CDC20	CCL21
SLC7A7	COL16A1	NCLN	PDCD1LG2	MT1M	CTSW	CCL22
HLA-DMA	EPHA2	ZWILCH	IL12RB1	IFITM1	CXCL1	CCL3/L1
FYB	TNC	DYNLT1	LILRA6	IFI35	DTX3L	CSF1
RNASE6	JUP	C16orf61	CXCR3	BST2	HLA-DMB	CSF2
TLR2	ITGA3	SLC25A40	SLAMF1	CD8B	HLA-DOA	CSF2RB
CTSC	TCF7L2	RHOC	SLAMF6	BATF2	HLA-DOB	CSF3R
LILRB4	COL3A1	CCT5	RHOH	CTSZ	HLA-DPA1	CX3CL1
PSCDBP	CDH6	PDIA4	GPR114	HAPLN3	HLA-DPB1	CX3CR1
CTSS	WNT2B	SNRPA	SLA2	TAP2	HLA-DQA1	CXCL12
RASSF4	ADAM9	RBM14	IL18RAP	ODF3B	HLA-DQA2	CXCL14
MSN	DSP	PDLIM7	GPR171	SQRDL	IRF8	CXCL16

【表３－２】

CYBB	HSPG2	PITPNC1	SCIMP	DOK2	KIF2C	CXCL2
LAPTM5	ARHGAP1	TPM3	P2RY13	UBE2L6	KIR2DL3	CXCL3
DOCK2	ITGB5	CORO1C	LAX1	SLC15A3	KIR3DL1	CXCL6
FCGR1A	IGFBP5	ERLIN1	AOAH	SECTM1	KIR3DL2	CXCL8
EVI2B	ARHGDIA	PAICS	CXCL11	GAL	KLRD1	CXCR2
ADCY7	LRP1	TPRKB	CD38	CXCL13	MRC1	GNG4
CD48	IGFBP2	SKA2	GRAP2	RAC2	PSMB10	IL10
ARHGAP15	CTNNA1	MYBL1	WDFY4	IL1R2	PSMB5	IL11
ARRB2	LRRC17	SH3BP5L	LTA	LRFN4	PSMB8	IL11RA
SYK	MMP14	BRCA2	GZMH	CLEC4A	PSMC4	IL12RB2
BTK	NEO1	SAR1A	PTPN22	SLAMF8	THBD	IL16
TNFAIP3	EFNA5	POLR3K	IRF4	SH2D2A	TNF	IL1A
FCGR2A	ITGB3	MRPS28	GNLY	CDC25B	TRIM21	IL1B
VSIG4	EPHB3	NUP107	CCR5	PSME2	UBA7	IL1RN
FPRL2	CD44	TUBG1	SNX20	GNGT2	UBB	IL22RA1
IL10RA	IGFBP4	PNN	LILRB5	TYROBP	UBE2C	IL24
IFI16	TNFRSF1A	FAM167A	RGS18	AIF1	ULBP2	IL33
ITGB2	RAC1	RFC3	EMR1	WARS	VHL	IL34
IL7R	PXN	MYL6	TIGIT	CD300A	AKT1	IL6
TBXAS1	PLAT	MCM7	SIGLEC7	C1QC	APC	IL6R
FMNL1	COL8A1	MAGOHB	GPR65	ARPC1B	BAD	JAK2
FLI1	WNT8B	FAM89B	TNFRSF17	CIITA	BAX	JAK3
RASSF2	IGFBP3	TOMM40	ADAMDEC1	HAVCR2	BBC3	PF4
LYZ	RHOA	CDCA4	HK3	SPI1	BCL2	PIK3R5
CD163	EPHB4	MT3	LILRB2	MIR3123	BCL2L1	SHC2
CD97	MMP1	MTHFD1	CTLA4	ACAA2	BCL6B	STAT2
CCL2	PAK1	PSMD12	SIGLEC12	KRTAP19-8	BID	STAT3
FCGR2B	MTA1	MYBL2	CXorf21	LPCAT1	BIRC3	CBLC
MERTK	THBS2	CKLF	SPIB	CARTPT	BIRC5	CD47
CD84	CSPG2	NRIP3	CLEC4G	ARHGAP26	BLM	CNTFR
CD53	MMP17	EZR	CD300C	RPA3	CASP3	GHR
CD86	CD59	C12orf24	CD300E	MIR129-1	CASP8	GZMM
HMHA1	DVL3	GPLD1	LILRB1	FAM150A	CASP9	IFITM2

【表３－３】

CTSL	RHOB	SRM	LAG3	KCNB2	CDH1	IRF9
EVI2A	COL6A3	RAB3B	HMGA2	TERF1	FADD	JAK1
TNFRSF1B	NOTCH2	NLN	VNN1	KCNK9	HMGB1	KLRB1
CXCR4	BSG	MT1F	IL2RA	PTK2	LY96	KLRK1
LCP1	MMP11	TNFRSF12A	NCR3	GPR20	RIPK1	LIF
SAMHD1	COL1A2	TPI1	CCR8	TSNARE1	RIPK3	PRLR
CPVL	ZYX	HAS2	TFEC	MIR4475	ROCK1	SIRPA
HLA-DRB1	RND3	APOO	CCR2	IGFBPL1	TICAM1	SPRY4
C13orf18	THBS1	FBXO41	STAP1	LOC100129316	TLR3	DDB2
GIMAP4	RHOG	MRPL37	IDO1	QRFP	TLR4	ERCC3
SAMSN1	ICAM1	GSTCD	MYO1G	CAMSAP1	TNFRSF10B	FANCA
PLCG2	LAMA4	SDC1	SIT1	GPR37L1	TNFSF10	MGMT
OSBPL3	DVL1	WDR54	CD209	CCDC85C	TP53	NEIL1
RUNX3	PAK2	RNF138	LILRA5	GACAT2	BNIP3	PARP4
FCGR3A	COL6A1	APITD1	SIGLEC1	LINC01443	DEPTOR	POLR2A
AMPD3	FGD1	RMND5B	UBD	CDH2	HIF1A	TNKS
MYO1F	CEP78	ENO1	SPN	MIR4436A	HRAS	UBE2T
CECR1	LSM3	MAP3K8	CXCL9	INSM1	KRAS	XCL1
LYN	LRRC40	TMEM130	CD70	LOC101929625	MAP3K7	XCL2
MPP1	STK17A	SNX17	GZMB	COL18A1	MAPK10	ARID1A
LRMP	RPN1	KRR1	SIGLEC10	OPA1-AS1	MTOR	BRD3
FGL2	JUNB	TAGLN	CD300LF	C9	NRAS	BRD4
NCKAP1L	NUP85	PA2G4	CCL8	MIR585	PIK3CA	DNMT1
HCLS1	FLNC	RUVBL1	PTPN7	MAD1L1	PIK3CD	HDAC11
SELL	HMGN2	SNRPD1	CCL4	CPA2	PIK3R1	HDAC3
CASP1	RPP40	LCXL2	CMKLR1	LOC100130298	PIK3R2	HDAC4
SELPLG	UQCR10	POLE2	ICAM3	C8orf34	PRKAA2	HDAC5
CD33	AIMP2	MAPRE1	SIRPG	OC90	PRKACB	HELLS
GPNMB	CHEK1	IMP4	NLRP3	TRAPPC9	PTEN	HMGA1
NCF2	VTA1	EMP2	MARCO	MED27	RPS6KB1	KAT2B
FNBP1	EXOSC8	PSMD2	CCL4L1	MLH1	RPTOR	MAP3K12
IL18	CENPO	MET	CD180	MSH2	ANLN	BMP2
B2M	PNO1	IFRD2	SIGLEC9	MSH6	ATM	GAS1
SP140	SLC16A1	LMNB2	KRT6B	PMS2	CCNA1	GLI1

【表３－４】

FCER1G	WDR77	PLOD2	NFAM1	BRCA1	CCNB1	IHH
LCP2	UBE2J1	NCEH1	GPR84	TERT	CCND1	PRKX
LY86	NOP16	NME1	KLK5	EphA2	CCND3	WNT10A
LAIR1	NUDT1	STRA13	CCL23	cGAS	CCNO	WNT11
IFI30	SMC2	ACTL6A	ZBED2	STING	CDC25C	WNT2
TNFSF13B	SLC25A5	DLEU1	CD28	CD8	CDK6	WNT3A
LST1	NUPL1	SNRPA1	CEACAM21	PDL2	CDKN1A	WNT4
FGR	DLEU2	CBX1	PRF1	ASXL3	CDKN1C	WNT5A
NPL	PDAP1	LYAR	CSF3	SPEDF	CDKN2A	WNT5B
PLEK	CCBL2	PTPLB	KIAA0226L	PSMA	CDKN2B	WNT7B
PTPRC	COX17	PFN1	CXCR6	GPC3	CENPF	ALDOA
GNPTAB	BCCIP	CENPJ	CCL24	RAD51	CEP55	EGF
SLC1A3	PLG	COTL1	ARHGAP9	ACVR1	EXO1	EGFR
HCK	RGS8	SPRYD7	LILRB3	APLNR	H2AFX	ERBB2
NPC2	SNRPC	USPL1	VNN2	CDC27	MELK	HK1
C3AR1	PLK4	MRPL12	ZAP70	CDK4	MKI67	HK2
PIK3CG	NUTF2	ADAMTS1	TNFSF8	HLA-C	MYC	IFNGR1
DAPK1	LSM4	GLRX3	HEPHL1	HLA-E	NBN	IFNGR2
ALOX5AP	SMS	WSB2	CCL3	HLA-G	POLD1	LDHA
CSF1R	EBNA1BP2	MRPS16	CLEC10A	IL1R	PRKCA	NOS2
CUGBP2	C13orf27	DCLRE1B	CCL3L3	IL2RB	RAD50	PDK1
APOE	VDAC1	MKKS	CCR4	INFGR1	RAD51C	PFKFB3
APOC1	PSMD14	C3orf26	ITGAL	INFGR2	RB1	SLC2A1
CD52	MYCBP	CPEB4	SAA2	IRF1	RBL2	TFRC
LHFPL2	SMURF2	SPAG17	GZMA	IRF2	RRM2	CD276
C1orf54	GNG11	MLF1IP	BCL2A1	KIR2DS2	TYMS	CD6
IKZF1	F3	UAP1	RNASE2	KIR2DS3	ADORA2A	CDH5
SH2B3	BRIP1	COQ2	SAA1	PIAS4	BTLA	CLEC14A
WIPF1	HNRNPA2B1	WDHD1	NKG7	PTPN2	CD247	CLEC5A
IFI44	DCK	DCBLD2	MT2A	SERPINB3	CD27	CLEC7A
IFI44L	ALKBH7	KIAA0090	FAM26F	SERPINB4	CD274	CLECL1
DDX58	HN1L	SAR1B	PSMB9	SPOP	CD40	ICAM2
IFI6	TPM1	PSMA7	TYMP	TAPBP	CD40LG	ICAM5
IFI27	HYLS1	PSMC3	CD7	TGFBR1	CD69	ITGA1

【表３－５】

IFIT2	HAUS1	COPS6	SYT12	CD58	CHUK	ITGA2
IFIT1	NUP93	DUT	S100A7	ANGPT1	DPP4	ITGAE
IFIT3	SNRPE	PPIH	HLA-F	ANGPT2	EGR1	ITGAM
CXCL10	ITGA6	PHF19	MT1E	ANGPTL4	FYN	ITGAX
MX1	CENPN	TPM2	ISG20	CCND2	ICOSLG	NCAM1
OAS1	C11orf24	MCTS1	C1QA	CCNE1	IKBKB	NECTIN1
OAS2	GGH	EIF4EBP1	TNFRSF18	CES3	IKBKG	PECAM1
OAS3	PFKP	HNRNPR	IL32	DLL4	IL15	PVR
HERC5	FARSA	IFNG	SOCS1	E2F3	IL17A	SELE
SAMD9	EIF2AK1	TLR8	HLA-DQB1	EDN1	IL18R1	SELP
HERC6	CENPW	FASLG	IFITM3	EZH2	IL2	THY1
DDX60	TUBA4A	KIR2DL4	PI3	FGF18	IL4	VCAM1
RTP4	TRA2B	IL21R	HCST	FGFR1	LY9	EIF2AK2
IFIH1	UMPS	GBP5	SUSD3	FLT1	NECTIN2	FLNB
STAT1	MRTO4	ZNF683	HLA-B	FSTL3	NFATC2	GBP1
TAP1	NUDT15	PLA2G2D	IL2RG	HEY1	NFKB1	GBP2
OASL	PGM2	TBX21	HLA-DRA	ITGAV	NFKBIA	GBP4
TPSD1	GATA2	SERPINF1	FCER1A	DAPP1	CCR7	EREG
SLC18A2	HPGDS	PLAC8	CD1A	C10ORF54	XCR1	CD81
IL1RL1	TCF4	JCHAIN	AXL	ANPEP	MTSS1	PRDM1
VWA5A	MZB1	SLC15A4	CD1E	CST7	MSA4A7	ABL2
TPSAB2	IL3RA	ITM2C	PKIB	FLT3	LILRA1	SDC4
IFNA1	PDGFA	MMP7	CRABP2	TNFRSF4	FZD9	PNOC
IFNAR1	PDGFRB	NID2	DAB2	TNFSF12	GPC4	HLA-DR
IRF3	PLA1A	RELN	FCAR	TNFSF13	SFRP1	LIN
IRF5	PLA2G2A	AQP9	FCN1	TNFSF18	SFRP4	CD11b
IRF7	RASAL1	CEBPB	FOSL1	TNFSF4	SOX11	CD15
BLK	RASGRF1	ERO1A	FPR1	APH1B	SOX2	CD11c
CD5	RELB	FBP1	FPR3	DLL1	FOXP3	CD34
CD96	TGFB1	G6PD	IER3	DTX4	NCR1	CD83
F2RL1	TGFB2	GLS	OLR1	HES1	CEACAM3	PD-L1
SLAMF7	TGFB3	GOT1	PDZK1IP1	JAG2	FCGR3A/B	CD45
DUSP1	TGFBR2	GOT2	S100A12	MAML2	SIGLEC5	CD124
DUSP2	A2M	LDHB	S100A8	MFNG	CPA3	CD66

【表３－６】

DUSP5	COL11A1	PC	S100A9	NOTCH1	HDC	CD3
FAS	COL11A2	PCK2	SIGLEC8	TCL1A	MS4A2	CD13
FGF13	COL17A1	PFKM	SIRPB2	ACVR1C	TPSAB1/B2	TPSAB1
FGF9	COL4A5	PKM	SLC11A1	BAMBI	CD68	LILRA4
KDR	COL5A1	SGK1	TLR1	ID4	MS4A4A	CD16
KIT	COMP	SLC1A5	TREM1	INHBA	CD244	PLTP
MAP3K5	LAMA1	SLC7A5	TREM2	SMAD5	PTGER4	ADM
NF1	LAMB3	AREG	NFKBIE	AXIN1	HSD11B1	CSTB
NFKB2	LAMC2	ARG1	TNFRSF11A	DKK1	FAM30A	PPARG
NGFR	LTBP1	C5AR1	TNFRSF11B	FZD8	FCRL2	RGCC

【０３７６】

さらなる高メチル化標的領域は、例えば、Ｃａｎｃｅｒ Ｇｅｎｏｍｅ Ａｔｌａｓから得ることができる。Kang et al., Genome Biology 18:53 (2017)は、乳房、結腸、腎臓、肝臓、および肺に由来する高メチル化標的領域を使用したＣａｎｃｅｒＬｏｃａｔｏｒと称される確率的方法の構築について記載している。一部の実施形態では、高メチル化標的領域は、がんの１つまたは複数のタイプに特異的であり得る。したがって、一部の実施形態では、高メチル化標的領域は、乳がん、結腸がん、腎臓がん、肝臓がん、および肺がんのうちの１つ、２つ、３つ、４つ、または５つにおいて高メチル化を集合的に示す高メチル化標的領域の１つ、２つ、３つ、４つ、または５つのサブセットを含む。

【０３７７】

一部の実施形態では、異なるエピジェネティック標的領域を第１の部分試料および第２の部分試料から捕捉する場合、第１の部分試料から捕捉されたエピジェネティック標的領域は高メチル化標的領域を含む。

ｂ．低メチル化標的領域

【０３７８】

一部の実施形態では、エピジェネティック標的領域セットは、１つまたは複数の低メチル化標的領域を含む。一部の実施形態では、低メチル化標的領域は、１つの免疫細胞タイプにおいて排他的に低メチル化されているか、または１つの免疫細胞タイプにおいて、任意の他の免疫細胞タイプにおけるもしくは同じ免疫細胞クラスター内の任意の他の免疫細胞タイプにおけるよりも大きな程度に低メチル化されている。一部のそのような実施形態では、低メチル化標的領域により、活性化Ｂ細胞（メモリーＢ細胞および形質細胞を含む）、活性化Ｔ細胞（調節性Ｔ細胞（Ｔｒｅｇ）、ＣＤ４エフェクターメモリーＴ細胞、ＣＤ４セントラルメモリーＴ細胞、ＣＤ８エフェクターメモリーＴ細胞、およびＣＤ８セントラルメモリーＴ細胞を含む）、ならびにナチュラルキラー（ＮＫ）細胞などの希少な免疫細胞タイプを含む、ＤＮＡの起源である特定の免疫細胞タイプのレベルが示される。免疫細胞タイプをデコンボリューションするために有用な低メチル化標的領域のメチル化パターンは、ある特定の疾患状態、例えばがんにおいてはさらに変化し得る。したがって、一部の実施形態では、免疫細胞タイプをデコンボリューションするために有用な低メチル化標的領域は、試料を得たところの対象が、がんまたは前がん状態を有する可能性を決定するためにも有用である。一部のそのような実施形態では、低メチル化標的領域は、特定の免疫細胞タイプのレベルが異常であるかどうか、およびそのような異常なレベルががんの存在もしくは前がん状態に関連する可能性があるかどうか、またはそのような異常なレベルが、がんもしくは前がん状態以外の種々の疾患もしくは状態に関連するかどうかを決

定するために有用である。

【０３７９】

さらに、全体的な低メチル化は、様々ながんにおいて一般に観察される現象である。Hon et al., Genome Res. 22:246-258 (2012)（乳がん）、Ehrlich, Epigenomics 1:239-259 (2009)（結腸がん、卵巣がん、前立腺がん、白血病、肝細胞がん、および子宮頸がんにおける低メチル化の観察に言及した総説論文）を参照されたい。例えば、反復エレメント、例えば、ＬＩＮＥ１エレメント、Ａｌｕエレメント、セントロメアタンデムリピート、ペリセントロメアタンデムリピート、およびサテライトＤＮＡなどの領域、ならびに健康な細胞において通常メチル化されている遺伝子間領域は、腫瘍細胞において、低減されたメチル化を示し得る。したがって、一部の実施形態では、エピジェネティック標的領域セットは、観察されるメチル化のレベルの低下により、がんの存在の可能性の増加が示される、低メチル化標的領域を含む。そのような低メチル化標的領域の検出により、例えば、免疫細胞タイプを示す低メチル化標的領域の検出と併せて、本明細書に記載される方法の特異度および／または感度をさらに増加させることができる。別の例では、上記で考察されたように、低メチル化標的領域は、がん性組織において、同じタイプの健康な組織に由来するＤＮＡと比べて必ずしもメチル化が異なるわけではないが、健康な対象において典型的であるｃｆＤＮＡと比べてメチル化が異なる（例えば、メチル化がより少ない）領域を含み得る。例えば、がんの存在により、細胞死、例えば、がんに対応する組織タイプの細胞のアポトーシスの増加がもたらされる場合、そのようながんを、少なくとも部分的にそのような低メチル化標的領域を使用して検出することができる。一部の実施形態では、対象ががんを有する可能性を決定するために有用な低メチル化標的領域は、特定の免疫細胞タイプのレベルを決定するために有用な低メチル化標的領域とは異なる。一部の実施形態では、対象ががんを有する可能性を決定するために有用な低メチル化標的領域の少なくとも一部は、特定の免疫細胞タイプのレベルを決定するために有用な低エチル化可変標的領域と同じである。

【０３８０】

一部の実施形態では、低メチル化標的領域は、反復エレメントおよび／または遺伝子間領域を含む。一部の実施形態では、反復エレメントは、ＬＩＮＥ１エレメント、Ａｌｕエレメント、セントロメアタンデムリピート、ペリセントロメアタンデムリピート、および／またはサテライトＤＮＡのうちの１つ、２つ、３つ、４つ、または５つを含む。

【０３８１】

がん関連低メチル化を示す例示的な特定のゲノム領域としては、ヒト第１染色体のヌクレオチド８４０３５６５～８９５３７０８および１５１１０４７０１～１５１１０６０３５が挙げられる。一部の実施形態では、低メチル化標的領域は、これらの領域の一方または両方と重複するまたはこれらの領域の一方または両方を含む。

【０３８２】

さらに、低メチル化標的領域は、例えば、Fox-Fisher et al., ElifeNov 29; 10 (2021)、ＥｐｉＤＩＳＨ Ｒパッケージ、Moss et al., Nat Commun 9:1 (2018)、およびLoyfer et al. bioRxiv https://doi.org/10.1101/2022.01.24.477547 (2022)から得ることができる。一部の実施形態では、低メチル化標的領域は、１つまたは複数のタイプの免疫細胞に特異的であり得る。

【０３８３】

一部の実施形態では、異なるエピジェネティック標的領域を第１の部分試料および第２の部分試料から捕捉する場合、第２の部分試料から捕捉されたエピジェネティック標的領域は、低メチル化標的領域を含む。一部の実施形態では、第２の部分試料から捕捉されたエピジェネティック標的領域は低メチル化標的領域を含み、第１の部分試料から捕捉されたエピジェネティック標的領域は高メチル化標的領域を含む。

ｃ．ＣＴＣＦ結合領域

【０３８４】

ＣＴＣＦは、クロマチン構成に寄与し、しばしば、コヒーシンと共局在化する、ＤＮＡ

結合タンパク質である。ＣＴＣＦ結合部位の乱れが、様々な異なるがんにおいて報告されている。例えば、Katainen et al., Nature Genetics, doi:10.1038/ng.3335, published online 8 June 2015、Guo et al., Nat. Commun. 9:1520 (2018)を参照されたい。ＣＴＣＦ結合は、シーケンシングによって、例えば、断片長解析を通じて検出することができる、ｃｆＤＮＡにおける認識可能なパターンをもたらす。シーケンシングに基づく断片長解析に関する詳細は、Snyder et al., Cell 164:57-68 (2016)、ＷＯ２０１８／００９７２３、およびＵＳ２０１７０２１１１４３Ａ１において提供されており、これらのそれぞれは、参照により本明細書に組み込まれる。

【０３８５】

したがって、ＣＴＣＦ結合の乱れにより、ｃｆＤＮＡの断片化パターンに変動が生じる。したがって、ＣＴＣＦ結合部位は、断片化可変標的領域の１つのタイプである。

【０３８６】

多数の公知のＣＴＣＦ結合部位が存在する。例えば、例えばその各々が参照により組み込まれる、insulatordb.uthsc.edu/においてインターネット上で利用可能なＣＴＣＦＢＳＤＢ（ＣＴＣＦ結合部位データベース）、Cuddapah et al., Genome Res. 19:24-32 (2009)、Martin et al., Nat. Struct. Mol. Biol. 18:708-14 (2011)、Rhee et al., Cell. 147:1408-19 (2011)を参照されたい。例示的なＣＴＣＦ結合部位は、第８染色体のヌクレオチド５６０１４９５５～５６０１６１６１、および第１３染色体のヌクレオチド９５３５９１６９～９５３６０４７３である。

【０３８７】

したがって、一部の実施形態では、エピジェネティック標的領域セットは、ＣＴＣＦ結合領域を含む。一部の実施形態では、ＣＴＣＦ結合領域は、少なくとも１０個、２０個、５０個、１００個、２００個、もしくは５００個のＣＴＣＦ結合領域、または１０～２０個、２０～５０個、５０～１００個、１００～２００個、２００～５００個、もしくは５００～１０００個のＣＴＣＦ結合領域、例えば、上述または上記に列挙されたＣＴＣＦＢＳＤＢもしくはＣｕｄｄａｐａｈら、Ｍａｒｔｉｎら、もしくはＲｈｅｅらの論文のうちの１つもしくは複数に記載されるＣＴＣＦ結合領域などを含む。

【０３８８】

一部の実施形態では、ＣＴＣＦ部位のうちの少なくともいくつかは、メチル化されていても、メチル化されていなくてもよく、ここで、メチル化状態は、細胞ががん細胞であるかそうでないかと相関する。一部の実施形態では、エピジェネティック標的領域セットは、ＣＴＣＦ結合部位の少なくとも１００ｂｐ、少なくとも２００ｂｐ、少なくとも３００ｂｐ、少なくとも４００ｂｐ、少なくとも５００ｂｐ、少なくとも７５０ｂｐ、少なくとも１０００ｂｐ上流および下流の領域を含む。

ｄ．転写開始部位

【０３８９】

転写開始部位もまた、新生物細胞において、乱れを示し得る。例えば、造血系列の健康な細胞における様々な転写開始部位におけるヌクレオソーム構成（これは、健康な個体におけるｃｆＤＮＡに実質的に寄与する）は、新生物細胞におけるそれらの転写開始部位では、ヌクレオソーム構成が異なり得る。これは、Snyder et al., Cell 164:57-68 (2016)、ＷＯ２０１８／００９７２３、およびＵＳ２０１７０２１１１４３Ａ１において一般に考察されているように、シーケンシングによって検出することができる異なるｃｆＤＮＡパターンをもたらす。別の例では、転写開始部位は、がん性組織において、同じタイプの健康な組織由来のＤＮＡと比べて必ずしもエピジェネティックに異なるわけではない場合があるが、健康な対象において典型的であるｃｆＤＮＡと比べてエピジェネティックに異なる（例えば、ヌクレオソーム構成に関して）。例えば、がんの存在により、細胞死、例えば、がんに対応する組織タイプの細胞のアポトーシスの増加がもたらされる場合、そのようながんを、少なくとも部分的に、そのような転写開始部位の相違を使用して検出することができる。

【０３９０】
したがって、転写開始部位の乱れによっても、ｃｆＤＮＡの断片化パターンに変動が生じる。したがって、転写開始部位もまた、断片化可変標的領域の１つのタイプである。
【０３９１】
ヒト転写開始部位は、btss.hgc.jpにおいてインターネット上で利用可能なＤＢＴＳＳ（ＤａｔａＢａｓｅ ｏｆ Ｈｕｍａｎ Ｔｒａｎｓｃｒｉｐｔｉｏｎ Ｓｔａｒｔ Ｓｉｔｅｓ）から利用可能であり、Yamashita et al., Nucleic Acids Res. 34(Database issue): D86-D89 (2006)において記載されており、これは、参照により本明細書に組み込まれる。
【０３９２】
したがって、一部の実施形態では、エピジェネティック標的領域セットは、転写開始部位を含む。一部の実施形態では、転写開始部位は、少なくとも１０個、２０個、５０個、１００個、２００個、もしくは５００個の転写開始部位、または１０～２０個、２０～５０個、５０～１００個、１００～２００個、２００～５００個、もしくは５００～１０００個の転写開始部位、例えば、ＤＢＴＳＳにおいて列挙されている転写開始部位などを含む。一部の実施形態では、転写開始部位のうちの少なくともいくつかは、メチル化されていても、メチル化されていなくてもよく、ここで、メチル化状態は、細胞ががん細胞であるかそうでないかと相関する。一部の実施形態では、エピジェネティック標的領域セットは、転写開始部位の少なくとも１００ｂｐ、少なくとも２００ｂｐ、少なくとも３００ｂｐ、少なくとも４００ｂｐ、少なくとも５００ｂｐ、少なくとも７５０ｂｐ、少なくとも１０００ｂｐ上流および下流の領域を含む。

ｅ．局所増幅

【０３９３】
局所増幅は、体細胞突然変異であるが、それらは、ある特定のエピジェネティック変化、例えば、メチル化の変化を検出するためのアプローチと類似の様式で、読み取りデータの頻度に基づいてシーケンシングすることによって検出することができる。そのため、がんにおいて局所増幅を示し得る領域は、エピジェネティック標的領域セットに含まれ得、ＡＲ、ＢＲＡＦ、ＣＣＮＤ１、ＣＣＮＤ２、ＣＣＮＥ１、ＣＤＫ４、ＣＤＫ６、ＥＧＦＲ、ＥＲＢＢ２、ＦＧＦＲ１、ＦＧＦＲ２、ＫＩＴ、ＫＲＡＳ、ＭＥＴ、ＭＹＣ、ＰＤＧＦＲＡ、ＰＩＫ３ＣＡ、およびＲＡＦ１のうちの１つまたは複数を含み得る。例えば、一部の実施形態では、エピジェネティック標的領域セットは、前述の標的のうちの少なくとも２個、３個、４個、５個、６個、７個、８個、９個、１０個、１１個、１２個、１３個、１４個、１５個、１６個、１７個、または１８個を含む。

ｆ．メチル化対照領域または対照領域

【０３９４】
データの検証を容易にするために対照領域を含めることは有用であり得る。一部の実施形態では、エピジェネティック標的領域セットは、ＤＮＡががん細胞または正常細胞に由来するか否かによらず、本質的に全ての試料においてメチル化または非メチル化であると予想される対照領域を含む。一部の実施形態では、エピジェネティック標的領域セットは、本質的に全ての試料において低メチル化されているまたはメチル化されていないと予想される陰性対照領域を含む。一部の実施形態では、エピジェネティック標的領域セットは、本質的に全ての試料において高メチル化されていることが予想される陽性対照領域を含む。

２．配列可変標的領域セット

【０３９５】
一部の実施形態では、配列可変標的領域セットは、がんにおいて体細胞変異（例えば、一塩基バリエーションおよび／またはインデル）を受けることが分かっている複数の領域を含む。一塩基バリエーションおよび／またはインデルは、参照配列、例えば、ＧＲＣｈ３８ヒトゲノムアセンブリなどの公開されたヒトゲノム配列と比べたものであってよい。
【０３９６】

一部の態様では、配列可変標的領域セットは、がんを有する対象のうちの決定された割合が、パネル内の１つまたは複数の異なる遺伝子またはゲノム領域において遺伝的変異体または腫瘍マーカーを示すように選択された、複数の異なる遺伝子またはゲノム領域（「パネル」）を標的とする。パネルは、シーケンシングのための領域を、固定数の塩基対に限定するように選択され得る。パネルは、本明細書の他の箇所に記載されるように、例えば、プローブの親和性および／または量を調整することによって、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。パネルは、さらに、所望の配列読み取りデータ深度を達成するように選択され得る。パネルは、シーケンシングされる塩基対の量に所望される配列読み取りデータ深度または配列読み取りデータカバレッジを達成するように選択され得る。パネルは、試料中の１つまたは複数の遺伝的変異体を検出するための理論上の感度、理論上の特異度、および／または理論上の正確度を達成するように選択され得る。

［０２８４］領域のパネルを検出するためのプローブとしては、目的のゲノム領域（ホットスポット領域）を検出するためのものを挙げることができる。プローブの設計にクロマチン構造に関する情報を考慮することができ、かつ／または、プローブを、特定の部位（例えば、ＫＲＡＳコドン１２および１３）を捕捉することができる可能性が最大になるように設計することができ、また、ヌクレオソーム結合パターンおよびＧＣ配列組成の影響を受けるｃｆＤＮＡカバレッジおよび断片サイズ変動の解析に基づく捕捉が最適化されるように設計することができる。本明細書で使用される領域には、ヌクレオソーム位置およびＧＣモデルに基づいて最適化された非ホットスポット領域を含めることもできる。

【０３９７】

目的のゲノムの場所の一覧の例は、表４および表５に見出すことができる。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表３の遺伝子のうちの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、または７０種の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表４のＳＮＶのうちの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、または７０種を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表４の融合体のうちの少なくとも１種、少なくとも２種、少なくとも３種、少なくとも４種、少なくとも５種、または６種を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表４のインデルのうちの少なくとも１種、少なくとも２種、または３種の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表５の遺伝子のうちの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、少なくとも７０種、または７３種の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表５のＳＮＶのうちの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、少なくとも７０種、または７３種を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表５の融合体のうちの少なくとも１種、少なくとも２種、少なくとも３種、少なくとも４種、少なくとも５種、または６種を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表５のインデルのうちの少なくとも１種、少なくとも２種、少なくとも３種、少なくとも４種、少なくとも５種、少なくとも６種、少なくとも７種、少なくとも８種、少なくとも９種、少なくとも１０種、少なくとも１１種、少なくとも１２種、少なくとも１３種

、少なくとも１４種、少なくとも１５種、少なくとも１６種、少なくとも１７種、または１８種の少なくとも一部分を含む。これらの目的のゲノムの場所のそれぞれを、所与のパネルについて骨格領域またはホットスポット領域として同定することができる。目的のホットスポットゲノムの場所の一覧の例は、表６に見出すことができる。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表６の遺伝子のうちの少なくとも１種、少なくとも２種、少なくとも３種、少なくとも４種、少なくとも５種、少なくとも６種、少なくとも７種、少なくとも８種、少なくとも９種、少なくとも１０種、少なくとも１１種、少なくとも１２種、少なくとも１３種、少なくとも１４種、少なくとも１５種、少なくとも１６種、少なくとも１７種、少なくとも１８種、少なくとも１９種、または少なくとも２０種の少なくとも一部分を含む。各ホットスポットゲノム領域が、関連する遺伝子、それが存在する染色体、遺伝子の座を表すゲノムの開始位置および終止位置、塩基対における遺伝子の座の長さ、遺伝子によってカバーされるエクソン、および所与の目的のゲノム領域が捕捉しようとし得る重要な特色（例えば、変異のタイプ）を含む、いくつかの特徴とともに列挙されている。

表４

【表４】

点変異 (SNVs)						融合体
AKT1	ALK	APC	AR	ARAF	ARID1A	ALK
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	FGFR2
CCNE1	CDH1	CDK4	CDK6	CDKN2A	CDKN2B	FGFR3
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	NTRK1
FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	RET

GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	ROS1
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	
PTPN11	RAF1	RB1	RET	RHEB	RHOA	
RIT1	ROS1	SMAD4	SMO	SRC	STK11	
TERT	TP53	TSC1	VHL			

表５

【表５】

点変異 (SNVs)						融合体
AKT1	ALK	APC	AR	ARAF	ARID1A	ALK
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	FGFR2
CCNE1	CDH1	CDK4	CDK6	CDKN2A	DDR2	FGFR3
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	NTRK1
FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	RET
GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	ROS1
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	
PTPN11	RAF1	RB1	RET	RHEB	RHOA	
RIT1	ROS1	SMAD4	SMO	MAPK1	STK11	
TERT	TP53	TSC1	VHL	MAPK3	MTOR	
NTRK3						

表６

【表６－１】

遺伝子	染色体	開始位置	終止位置	長さ (bp)	カバーされるエクソン	重要な特色
ALK	chr2	29446405	29446655	250	イントロン19	融合体
ALK	chr2	29446062	29446197	135	イントロン20	融合体
ALK	chr2	29446198	29446404	206	20	融合体
ALK	chr2	29447353	29447473	120	イントロン19	融合体

【表６－２】

ALK	chr2	29447614	29448316	702	イントロン19	融合体
ALK	chr2	29448317	29448441	124	19	融合体
ALK	chr2	29449366	29449777	411	イントロン18	融合体
ALK	chr2	29449778	29449950	172	18	融合体
BRAF	chr7	140453064	140453203	139	15	BRAF V600
CTNNB1	chr3	41266007	41266254	247	3	S37
EGFR	chr7	55240528	55240827	299	18および19	G719および欠失
EGFR	chr7	55241603	55241746	143	20	挿入　　/T790M
EGFR	chr7	55242404	55242523	119	21	L858R
ERBB2	chr17	37880952	37881174	222	20	挿入
ESR1	chr6	152419857	152420111	254	10	V534, P535, L536, Y537, D538
FGFR2	chr10	123279482	123279693	211	6	S252
GATA3	chr10	8111426	8111571	145	5	SS /インデル
GATA3	chr10	8115692	8116002	310	6	SS / インデル
GNAS	chr20	57484395	57484488	93	8	R844
IDH1	chr2	209113083	209113394	311	4	R132
IDH2	chr15	90631809	90631989	180	4	R140, R172
KIT	chr4	55524171	55524258	87	1	
KIT	chr4	55561667	55561957	290	2	
KIT	chr4	55564439	55564741	302	3	
KIT	chr4	55565785	55565942	157	4	
KIT	chr4	55569879	55570068	189	5	
KIT	chr4	55573253	55573463	210	6	
KIT	chr4	55575579	55575719	140	7	
KIT	chr4	55589739	55589874	135	8	
KIT	chr4	55592012	55592226	214	9	
KIT	chr4	55593373	55593718	345	10および11	557, 559, 560, 576
KIT	chr4	55593978	55594297	319	12および13	V654
KIT	chr4	55595490	55595661	171	14	T670, S709
KIT	chr4	55597483	55597595	112	15	D716
KIT	chr4	55598026	55598174	148	16	L783
KIT	chr4	55599225	55599368	143	17	C809, R815, D816, L818, D820, S821F, N822, Y823

【表６－３】

KIT	chr4	55602653	55602785	132	18	A829P
KIT	chr4	55602876	55602996	120	19	
KIT	chr4	55603330	55603456	126	20	
KIT	chr4	55604584	55604733	149	21	
KRAS	chr12	25378537	25378717	180	4	A146
KRAS	chr12	25380157	25380356	199	3	Q61
KRAS	chr12	25398197	25398328	131	2	G12/G13
MET	chr7	116411535	116412255	720	13, 14, イントロン13, イントロン14	MET エクソン 14 SS
NRAS	chr1	115256410	115256609	199	3	Q61
NRAS	chr1	115258660	115258791	131	2	G12/G13
PIK3CA	chr3	178935987	178936132	145	10	E545K
PIK3CA	chr3	178951871	178952162	291	21	H1047R
PTEN	chr10	89692759	89693018	259	5	R130
SMAD4	chr18	48604616	48604849	233	12	D537
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)		

【０３９８】

それに加えて、またはその代わりに、好適な標的領域セットは、文献から入手可能である。例えば、参照により本明細書に組み込まれる、Gale et al., PLoS one 13: e0194630 (2018)には、配列可変標的領域セットの一部または全てとして使用することができる３５種のがん関連遺伝子標的のパネルが記載されている。これらの３５種の標的は、ＡＫＴｌ、ＡＬＫ、ＢＲＡＦ、ＣＣＮＤ１、ＣＤＫ２Ａ、ＣＴＮＮＢ１、ＥＧＦＲ、ＥＲＢＢ２、ＥＳＲ１、ＦＧＦＲ１、ＦＧＦＲ２、ＦＧＦＲ３、ＦＯＸＬ２、ＧＡＴＡ３、ＧＮＡ１１、ＧＮＡＱ、ＧＮＡＳ、ＨＲＡＳ、ＩＤＨ１、ＩＤＨ２、ＫＩＴ、ＫＲＡＳ、ＭＥＤ １２、ＭＥＴ、ＭＹＣ、ＮＦＥ２Ｌ２、ＮＲＡＳ、ＰＤＧＦＲＡ、ＰＩＫ３ＣＡ、ＰＰＰ２Ｒ１Ａ、ＰＴＥＮ、ＲＥＴ、ＳＴＫ１１、ＴＰ５３、およびＵ２ＡＦ１

である。

【０３９９】

一部の実施形態では、配列可変標的領域セットは、上に列挙されているがん関連遺伝子などの少なくとも１０種、２０種、３０種、または３５種のがん関連遺伝子由来の標的領域を含む。

Ｈ．対象

【０４００】

一部の実施形態では、ＤＮＡ（例えば、ｃｆＤＮＡ）は、がんもしくは前がん状態、感染、移植片拒絶、または免疫系に直接的もしくは間接的に影響を及ぼす他の疾患を有する対象から得られる。一部の実施形態では、ＤＮＡ（例えば、ｃｆＤＮＡ）は、がんもしくは前がん状態、感染、移植片拒絶、または免疫系に直接的もしくは間接的に影響を及ぼす他の疾患を有することが疑われる対象から得られる。一部の実施形態では、ＤＮＡ（例えば、ｃｆＤＮＡ）は、腫瘍を有する対象から得られる。一部の実施形態では、ＤＮＡ（例えば、ｃｆＤＮＡ）は、腫瘍を有することが疑われる対象から得られる。一部の実施形態では、ＤＮＡ（例えば、ｃｆＤＮＡ）は、新生物を有する対象から得られる。一部の実施形態では、ＤＮＡ（例えば、ｃｆＤＮＡ）は、新生物を有することが疑われる対象から得られる。一部の実施形態では、ＤＮＡ（例えば、ｃｆＤＮＡ）は、（例えば、化学療法、外科的切除、放射線照射、またはこれらの組合せの後に）腫瘍、がん、または新生物から寛解した対象から得られる。前述の実施形態のいずれかでは、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、肺、結腸、直腸、腎臓、乳房、前立腺、または肝臓のものであり得る。一部の実施形態では、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、肺のものである。一部の実施形態では、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、結腸または直腸のものである。一部の実施形態では、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、乳房のものである。一部の実施形態では、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、前立腺のものである。前述の実施形態のいずれかでは、対象は、ヒト対象であり得る。

Ｉ．試料または部分試料またはその一部分由来のＤＮＡのプール

【０４０１】

一部の実施形態では、本明細書の方法は、複数の分配された部分試料からのタグ付けされたＤＮＡを含む１つまたは複数のプールを調製するステップを含む。一部の実施形態では、プールは、低メチル化分配のＤＮＡの少なくとも一部分および高メチル化分配のＤＮＡの少なくとも一部分を含む。例えばエピジェネティック標的領域および／または配列可変標的領域を含む、標的領域をプールから捕捉することができる。本明細書の他の箇所に記載される試料または部分試料の少なくともある１つのアリコートまたは一部分から標的領域セットを捕捉するステップは、第１の部分試料および第２の部分試料に由来するＤＮＡを含むプールに対して実施される捕捉ステップを包含する。プールから標的領域を捕捉する前に、プール内のＤＮＡを増幅させるステップを実施することができる。捕捉するステップは、本明細書の他の箇所で捕捉するステップについて記載されている特色のいずれかを有し得る。

【０４０２】

一部の実施形態では、方法は、低メチル化分配のＤＮＡの少なくとも一部分を含む第１のプールを調製するステップを含む。一部の実施形態では、方法は、高メチル化分配のＤＮＡの少なくとも一部分を含む第２のプールを調製するステップを含む。一部の実施形態では、方法は、第１のプールから少なくとも標的領域の第１のセットを捕捉するステップを含み、ここで、第１のセットは、配列可変標的領域を含む。この捕捉ステップの前に、第１のプール内のＤＮＡを増幅させるステップを実施することができる。一部の実施形態では、第１のプールから標的領域の第１のセットを捕捉するステップは、第１のプールのＤＮＡを標的特異的プローブの第１のセットと接触させることを含み、ここで、標的特異的プローブの第１のセットは、配列可変標的領域に対して特異的な標的結合プローブを含

む。一部の実施形態では、方法は、第２のプールから標的領域の第２の複数のセットを捕捉するステップを含み、ここで、第２の複数物は、配列可変標的領域およびエピジェネティック標的領域を含む。この捕捉ステップの前に、第２のプール内のＤＮＡを増幅させるステップを実施することができる。一部の実施形態では、第２のプールから標的領域の第２の複数のセットを捕捉するステップは、第１のプールのＤＮＡを標的特異的プローブの第２のセットと接触させることを含み、ここで、標的特異的プローブの第２のセットは、配列可変標的領域に対して特異的な標的結合プローブおよびエピジェネティック標的領域に対して特異的な標的結合プローブを含む。

【０４０３】

一部の実施形態では、分配された部分試料の第２の一部分から配列可変標的領域を捕捉する。第２の一部分は、プールに含められなかった部分試料のＤＮＡの一部、大多数、実質的に全て、または全てを含み得る。プールから捕捉された領域と部分試料から捕捉された領域を組み合わせ、並行して解析してもよい。

【０４０４】

エピジェネティック標的領域は、本明細書の他の箇所で考察されるように、それらが特定の細胞もしくは組織タイプを起源とするか、または腫瘍を起源とするか、または健康な細胞を起源とするかに応じて、メチル化レベルおよび／または断片化パターンに相違を示し得る。配列可変標的領域は、それらが腫瘍を起源とするかまたは健康な細胞を起源とするかに応じて配列に相違を示し得る。［０２９３］一部の適用では、低メチル化分配に由来するエピジェネティック標的領域の解析は、高メチル化分配および低メチル化分配に由来する配列可変標的領域ならびに高メチル化分配に由来するエピジェネティック標的領域の解析よりも情報量が少ない場合がある。したがって、配列可変標的領域およびエピジェネティック標的領域を捕捉する方法では、後者は、高メチル化分配および低メチル化分配から配列可変標的領域のうちの１つもしくは複数が捕捉されるよりも低い程度に、ならびに／または、高メチル化分配からエピジェネティック標的領域が捕捉されるよりも低い程度で捕捉され得る。例えば、配列可変標的領域は、高メチル化分配とともにはプールされていない低メチル化分配の一部分から捕捉され得、プールは、高メチル化分配に由来するＤＮＡの一部（例えば、大多数、実質的に全て、または全て）、および低メチル化分配に由来するＤＮＡなしでまたはその一部（例えば、少数）を用いて調製することができる。そのようなアプローチにより、低メチル化分配に由来するエピジェネティック標的領域の配列決定を減少または排除し、それによって、さらなる解析のために十分である配列決定データの量を減少させることができる。

【０４０５】

一部の実施形態では、プール内に低メチル化分配のＤＮＡのうちの少数を含めることにより、１つまたは複数のエピジェネティック特色（例えば、メチル化または本明細書の他の箇所に詳細に考察されている他のエピジェネティック特色（複数可））の定量が、例えば相対的に、容易になる。

【０４０６】

一部の実施形態では、プールは、例えば、低メチル化分配のＤＮＡの約５０％未満、例えば、低メチル化分配のＤＮＡのうちの約４５％未満もしくはそれに等しい、４０％未満もしくはそれに等しい、３５％未満もしくはそれに等しい、３０％未満もしくはそれに等しい、２５％未満もしくはそれに等しい、２０％未満もしくはそれに等しい、１５％未満もしくはそれに等しい、１０％未満もしくはそれに等しい、または５％未満もしくはそれに等しい、低メチル化分配のＤＮＡのうちの少数を含む。一部の実施形態では、プールは、低メチル化分配のＤＮＡのうちの約５％～２５％を含む。一部の実施形態では、プールは、低メチル化分配のＤＮＡのうちの約１０％～２０％を含む。一部の実施形態では、プールは、低メチル化分配のＤＮＡのうちの約１０％を含む。一部の実施形態では、プールは、低メチル化分配のＤＮＡのうちの約１５％を含む。一部の実施形態では、プールは、低メチル化分配のＤＮＡのうちの約２０％を含む。

【０４０７】

一部の実施形態では、プールは、高メチル化分配の一部分を含み、これは、高メチル化分配のＤＮＡのうちの少なくとも約５０％であり得る。例えば、プールは、高メチル化分配のＤＮＡのうちの少なくとも約５５％、６０％、６５％、７０％、７５％、８０％、８５％、９０％、または９５％を含み得る。一部の実施形態では、プールは、高メチル化分配のＤＮＡのうちの５０～５５％、５５～６０％、６０～６５％、６５～７０％、７０～７５％、７５～８０％、８０～８５％、８５～９０％、９０～９５％、または９５～１００％を含む。一部の実施形態では、第２のプールは、高メチル化分配のＤＮＡの全てまたは実質的に全てを含む。

【０４０８】

一部の実施形態では、第１のプールは、低メチル化分配のＤＮＡの実質的に全てまたは全てを含む（例えば、第２のプールが低メチル化分配のＤＮＡを含まない場合）。一部の実施形態では、第２のプールは、低メチル化分配のＤＮＡを含まない（例えば、第１のプールが低メチル化分配のＤＮＡの実質的に全てまたは全てを含む場合）。

【０４０９】

一部の実施形態では、第２のプールは、高メチル化分配の一部分を含み、これは、低メチル化分配に関して上述の値および範囲のいずれかであり得る。一部の実施形態では、第２のプールは、高メチル化分配のＤＮＡの全てまたは実質的に全てを含む。

【０４１０】

例示的な実施形態では、分配後、分配は、別個に、末端修復および分子バーコードを含むアダプターへのライゲーションを受け、次いで、別個に増幅される。増幅後、増幅された分子を富化する（それでもなお分配は別個に保持される）。富化後、富化されたＤＮＡを、本明細書に記載される実施形態のいずれかに従ってプールし、次いで、再度増幅させる。増幅後、分子を配列決定する。

【０４１１】

様々な実施形態では、上述の考察と一貫して、方法は、捕捉されたＤＮＡを、例えば、エピジェネティックおよび配列可変標的領域セットに対して異なる程度の配列決定深度まで配列決定するステップをさらに含む。

Ｊ．配列決定

【０４１２】

一般に、以前の増幅の有無にかかわらず、アダプターに挟まれている核酸を含む、試料核酸は、配列決定に供することができる。配列決定方法としては、例えば、サンガー配列決定、ハイスループット配列決定、パイロシーケンシング、合成による配列決定、一分子配列決定、ナノポア配列決定、半導体配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、デジタル遺伝子発現（Ｈｅｌｉｃｏｓ）、次世代配列決定（ＮＧＳ）、合成による一分子配列決定（ＳＭＳＳ）（Ｈｅｌｉｃｏｓ）、大規模並列配列決定、クローン単一分子アレイ（Ｓｏｌｅｘａ）、ショットガン配列決定、Ｉｏｎ Ｔｏｒｒｅｎｔ、Ｏｘｆｏｒｄ Ｎａｎｏｐｏｒｅ、Ｒｏｃｈｅ Ｇｅｎｉａ、マキサム－ギルバート配列決定、プライマーウォーキング、およびＰａｃＢｉｏ、ＳＯＬｉＤ、Ｉｏｎ Ｔｏｒｒｅｎｔ、またはＮａｎｏｐｏｒｅプラットフォームを使用した配列決定が挙げられる。

【０４１３】

一部の実施形態では、配列決定は、非改変および改変核酸塩基を検出および／または識別することを含む。例えば、ＰａｃＢｉｏ配列決定（例えば、一分子リアルタイム（ＳＭＲＴ）配列決定）により、例えば、５－メチルシトシンおよび５－ヒドロキシメチルシトシンならびに非改変シトシンを直接検出する能力がもたらされる。例えば、Schatz., Nature Methods. 14 (4): 347-348 (2017)；およびＵＳ９，１５０，９１８を参照されたい。また、ＤＮＡのメチル化（例えば：５－メチルシトシンおよび５－ヒドロキシメチルシトシン）を直接検出することができるＯｘｆｏｒｄナノポア配列決定システム（例えば、ＭｉｎＩＯＮシーケンサー）を本実施形態で使用することができる。シーケンシング反応は、様々な試料プロセシングユニットにおいて行うことができ、これは、複

数のレーン、複数のチャネル、複数のウェル、または複数の試料セットを実質的に同時にプロセシングする他の手段を含み得る。試料プロセシングユニットにはまた、複数の実行のプロセシングを同時に可能にする複数の試料チャンバーが含まれ得る。同様に、Ｉｏｎ Ｔｏｒｒｅｎｔ配列決定を使用して、メチル化を直接検出することもできる。したがって、一部の実施形態では、メチル化状態を、配列決定中に、例えば、分配ステップもしくはバイサルファイト処置などの変換手順なしで、またはそれとは独立して、決定することができる。

【０４１４】

配列決定反応を、核酸の１つまたは複数の形態、例えば、がんまたは他の疾患のマーカーを含有することが分かっているものに対して実施することができる。シーケンシング反応はまた、試料中に存在する任意の核酸断片に行うこともできる。一部の実施形態では、ゲノムの配列カバレッジは、５％を下回る、１０％を下回る、１５％を下回る、２０％を下回る、２５％を下回る、３０％を下回る、４０％を下回る、５０％を下回る、６０％を下回る、７０％を下回る、８０％を下回る、９０％を下回る、９５％を下回る、９９％を下回る、９９．９％を下回る、または１００％を下回り得る。一部の実施形態では、配列反応は、ゲノムの少なくとも５％、１０％、１５％、２０％、２５％、３０％、４０％、５０％、６０％、７０％、または８０％の配列カバレッジを提供し得る。配列カバレッジは、少なくとも５個、１０個、２０個、７０個、１００個、２００個、もしくは５００個の異なる遺伝子、または多くとも５０００個、２５００個、１０００個、５００個、もしくは１００個の異なる遺伝子に行うことができる。［０３０４］　多重シーケンシングを使用して、同時のシーケンシング反応を行ってもよい。一部の場合には、無細胞核酸は、少なくとも１０００回、２０００回、３０００回、４０００回、５０００回、６０００回、７０００回、８０００回、９０００回、１００００回、５００００回、１００，０００回のシーケンシング反応でシーケンシングされ得る。他の場合には、無細胞核酸は、１０００回を下回る、２０００回を下回る、３０００回を下回る、４０００回を下回る、５０００回を下回る、６０００回を下回る、７０００回を下回る、８０００回を下回る、９０００回を下回る、１００００回を下回る、５００００回を下回る、１００，０００回を下回るシーケンシング反応でシーケンシングされ得る。シーケンシング反応は、逐次的に、または同時に行われてもよい。後続のデータ解析は、シーケンシング反応の全てまたは一部で行われ得る。一部の場合には、データ解析は、少なくとも１０００回、２０００回、３０００回、４０００回、５０００回、６０００回、７０００回、８０００回、９０００回、１００００回、５００００回、１００，０００回のシーケンシング反応で行われ得る。他の場合には、データ解析は、１０００回を下回る、２０００回を下回る、３０００回を下回る、４０００回を下回る、５０００回を下回る、６０００回を下回る、７０００回を下回る、８０００回を下回る、９０００回を下回る、１００００回を下回る、５００００回を下回る、１００，０００回を下回るシーケンシング反応で行われ得る。例示的なリード深度は、座（塩基）あたり１０００～５００００のリードである。１。

１．配列決定の差次的深度

【０４１５】

一部の実施形態では、配列可変標的領域セットに対応する核酸は、エピジェネティック標的領域セットに対応する核酸よりも大きい配列決定深度まで配列決定される。例えば、配列バリアント標的領域セットに対応する核酸の配列決定深度は、エピジェネティック標的領域セットに対応する核酸に対する配列決定深度よりも少なくとも１．２５倍、１．５倍、１．７５倍、２倍、２．２５倍、２．５倍、２．７５倍、３倍、３．５倍、４倍、４．５倍、５倍、６倍、７倍、８倍、９倍、１０倍、１１倍、１２倍、１３倍、１４倍、もしくは１５倍大きいか、または１．２５倍～１．５倍、１．５倍～１．７５倍、１．７５倍～２倍、２倍～２．２５倍、２．２５倍～２．５倍、２．５倍～２．７５倍、２．７５倍～３倍、３倍～３．５倍、３．５倍～４倍、４倍～４．５倍、４．５倍～５倍、５倍～５．５倍、５．５倍～６倍、６倍～７倍、７倍～８倍、８倍～９倍、９倍～１０倍、１０倍～１１倍、１１倍～１２倍、１３倍～１４倍、１４倍～１５倍、もしくは１５倍～１０

０倍大きくてよい。一部の実施形態では、前記シーケンシング深度は、少なくとも２倍大きい。一部の実施形態では、前記シーケンシング深度は、少なくとも５倍大きい。一部の実施形態では、前記シーケンシング深度は、少なくとも１０倍大きい。一部の実施形態では、前記シーケンシング深度は、４～１０倍大きい。一部の実施形態では、前記シーケンシング深度は、４～１００倍大きい。

【０４１６】

一部の実施形態では、配列可変標的領域セットおよび／またはエピジェネティック標的領域セットに対応するＤＮＡは、併せて、例えば、同じ配列決定セル（例えば、Ｉｌｌｕｍｉｎａシーケンサーのフローセル）において、ならびに／あるいは、別個に捕捉されたセットを組換えることに起因する組み合わされたもしくはプールされた組成物、または、例えば、配列可変標的領域セットに対応するｃｆＤＮＡおよび／もしくはエピジェネティック標的領域セットに対応する捕捉されたｃｆＤＮＡを同じ容器中に捕捉することによって得られた組成物であり得る同じ組成物において、配列決定される。

Ｋ．解析

【０４１７】

一部の実施形態では、本明細書に開示される方法のいずれかは、ＤＮＡを採取した対象が、免疫系に関連する疾患または障害、例えば、感染、移植片拒絶、またはがんもしくは前がん状態を有する可能性を決定するステップを含む。

がんの指標

【０４１８】

一部の実施形態では、本明細書に開示される方法のいずれかは、腫瘍（もしくは新生物細胞、もしくはがん細胞）によって、または前がん状態細胞によって産生されるＤＮＡの存在を同定するステップを含む。一部の実施形態では、本明細書に記載される方法は、対象におけるがんの指標を決定するステップを含む。一部のそのような実施形態では、がんの指標の決定により、がんもしくは前がん状態の（ｏｒ）検出もしくは診断、またはがんの予後もしくはがん処置選択肢の決定が容易になる。例えば、１つまたは複数の分類領域および１つまたは複数の対照領域からメトリクスを決定することは、がんの指標を決定することに役立ち得る。一部の実施形態では、メトリクスを使用して、試料の腫瘍画分を決定することができる。

【０４１９】

本方法は、対象における状態、特にがんまたは前がん状態の存在を診断するため、状態を特徴付ける（例えば、がんをステージ分類またはがんの不均一性を決定する）ため、状態の処置に対する応答をモニタリングする、状態を発症するリスクまたは状態のその後の経過の予後判定を行うために、使用することができる。本開示はまた、特定の処置の選択肢の有効性を決定するのにも有用であり得る。例えば、腫瘍画分の変化、または１つもしくは複数の領域（ｏｎｅ ｏｒ ｒｅｇｉｏｎｓ）のメチル化状態の決定は、患者が処置に応答しているか否かを決定することに関して有用であり得る。別の例では、おそらく、ある特定の処置選択肢は、経時的に、がんのメチル化プロファイルと相関付けられ得る。この相関は、治療を選択するのに有用であり得る。

【０４２０】

さらに、がんが、処置後に寛解期にあることが観察された場合、本方法を使用して、残存疾患または疾患の再発をモニタリングすることができる。

【０４２１】

検出することができるがんのタイプおよび数としては、血液のがん、脳のがん、肺がん、皮膚がん、鼻のがん、喉のがん、肝臓がん、骨がん、リンパ腫、膵臓がん、皮膚がん、腸がん、直腸がん、甲状腺がん、膀胱がん、腎臓がん、口腔がん、胃がん、固形状態の腫瘍、不均一な腫瘍、均一な腫瘍などを挙げることができる。がんのタイプおよび／またはステージは、変異、希少な変異、インデル、コピー数バリエーション、トランスバージョン、転座、組換え、逆位、欠失、異数性、部分異数性、倍数性、染色体不安定性、染色体構造の変更、遺伝子融合、染色体融合、遺伝子トランケーション、遺伝子増幅、遺伝子重

複、染色体病変、ＤＮＡ病変、核酸化学修飾の異常な変化、エピジェネティックパターンの異常な変化、および核酸５－メチルシトシンの異常な変化を含む、遺伝的バリエーションから検出することができる。

【０４２２】

遺伝子データはまた、がんの特定の形態を特徴付けるためにも使用することができる。がんは、組成およびステージ分類の両方において、不均一であることが多い。遺伝子プロファイルデータにより、その特定のサブタイプの診断または処置に重要であり得る特定のサブタイプのがんの特徴付けが可能となり得る。この情報はまた、対象または従事者に特定のがんタイプの予後に関する手がかりを提供し、対象または従事者のいずれかが、疾患の進行に合わせて処置選択肢を適合させることを可能にし得る。一部のがんは、進行すると、より侵襲性かつ遺伝子的に不安定になり得る。他のがんは、良性、不活性、または休眠中のままであり得る。本開示のシステムおよび方法は、疾患の進行を決定するのに有用であり得る。

【０４２３】

さらに、本開示の方法は、対象における異常な状態の不均一性を特徴付けるために使用され得る。そのような方法には、例えば、対象に由来する細胞外ポリヌクレオチドの遺伝子プロファイルを生成するステップが含まれ得、ここで、遺伝子プロファイルは、コピー数変異および希少な突然変異の解析により得られる複数のデータを含む。一部の実施形態では、異常な状態は、がんまたは前がんである。一部の実施形態では、異常な状態は、不均一なゲノム集団をもたらすものであり得る。がんの例では、一部の腫瘍は、がんの異なるステージにある腫瘍細胞を含むことが公知である。他の例では、不均一性は、複数の疾患病巣を構成し得る。さらに、がんの例では、おそらくは１つまたは複数の病巣が原発部位から拡がった転移の結果である複数の腫瘍病巣が、存在し得る。

【０４２４】

本方法は、不均一な疾患における異なる細胞に由来する遺伝情報をまとめたものであるデータのフィンガープリントまたはセットを生成またはプロファイリングするために使用することができる。このデータのセットは、コピー数バリエーション、エピジェネティックバリエーション、または他の変異の解析を、単独または組合せで含み得る。

【０４２５】

本方法を使用して、がんまたは他の疾患を診断、予後判定、モニタリング、または観察することができる。一部の実施形態では、本明細書の方法は、胎児を診断することも、予後判定することも、モニタリングすることも含まず、そのため、非侵襲性の出生前検査を対象としない。他の実施形態では、これらの方法論は、ＤＮＡおよび他のポリヌクレオチドが母体分子と共循環し得る、出生前の対象におけるがんまたは他の疾患を診断、予後判定、モニタリング、または観察するために、妊娠している対象において用いられ得る。

【０４２６】

ＮＧＳによってがんの指標を決定するための例示的な方法は、以下のステップを含む：

１．血液試料からｃｆＤＮＡを抽出するステップ

２．ＤＮＡを、ＤＮＡ内の改変シトシン、例えばメチルシトシンを認識する作用剤と接触させることにより、ｃｆＤＮＡを複数の分配に分配するステップ

３．分配を、分子バーコードを含むアダプターとライゲーションするステップ

４．高分配および／もしくは中間分配を１つもしくは複数のＭＳＲＥで処理する、ならびに／または低分配を１つもしくは複数のＭＤＲＥで処理するステップ

５．消化後の分配をＰＣＲ増幅によって増幅させるステップ

６．高メチル化標的領域、低メチル化標的領域および対照領域を含むＤＮＡを、標的特異的プローブを使用して捕捉するステップ。

７．捕捉されたＤＮＡを増幅させ、ＮＧＳ機器で多重にアッセイするステップ。

８．本明細書に開示される１つまたは複数の方法を使用してＮＧＳデータを解析して、がんの指標を決定するステップ。

【０４２７】

ＮＧＳによってがんの指標を決定するための別の例示的な方法は、以下のステップを含む：
１．血液試料からｃｆＤＮＡを抽出するステップ。
３．ｃｆＤＮＡを、分子バーコードを含むアダプターとライゲーションするステップ
４．ライゲーションさせたｃｆＤＮＡをメチル化感受性変換方法に供するステップ
６．高メチル化標的領域、低メチル化標的領域および対照領域を含むＤＮＡを、標的特異的プローブを使用して捕捉するステップであって、プローブが、それらを、変換されたまたは変換されていない分子をメチル化感受性変換方法のタイプおよび捕捉される標的領域（高メチル化標的領域であるかまたは低メチル化標的領域であるか）に応じて捕捉するための標的とすることができるように設計される、ステップ。
７．捕捉されたＤＮＡを増幅させ、ＮＧＳ機器で多重にアッセイするステップ。
８．本明細書に開示される１つまたは複数の方法を使用してＮＧＳデータを解析して、がんの指標を決定するステップ。

【０４２８】

標的領域（例えば、プロモーター領域）のメチル化状態をＮＧＳによって決定するための別の例示的な方法は、以下のステップを含む：
１．血液試料からｃｆＤＮＡを抽出するステップ
２．ＤＮＡを、ＤＮＡ内の改変シトシン、例えばメチルシトシンを認識する作用剤と接触させることにより、ｃｆＤＮＡを複数の分配に分配するステップ
３．分配を、分子バーコードを含むアダプターとライゲーションするステップ
４．高分配および／もしくは中間分配を１つもしくは複数のＭＳＲＥで処理する、ならびに／または低分配を１つもしくは複数のＭＤＲＥで処理するステップ
５．消化後の分配をＰＣＲ増幅によって増幅させるステップ
６．高メチル化標的領域、低メチル化標的領域および対照領域を含むＤＮＡを、標的特異的プローブを使用して捕捉するステップ。
７．捕捉されたＤＮＡを増幅させ、ＮＧＳ機器で多重にアッセイするステップ。
８．本明細書に開示される１つまたは複数の方法を使用してＮＧＳデータを解析して、標的領域（例えば、プロモーター領域）がメチル化されているか否かを決定するステップ。

【０４２９】

本明細書に記載される方法の一部の実施形態では、分子バーコードは、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順によって変更されないヌクレオチド、例えば、本明細書に記載されるもののうちのいずれか（例えば、手順がバイサルファイト変換またはｍＣに影響を及ぼさない任意の他の変換である場合、Ａ、Ｔ、およびＧとともにｍＣ；手順がｈｍＣに影響を及ぼさない変換である場合、Ａ、Ｔ、およびＧとともにｈｍＣなど）からなる。本明細書に記載される方法の一部の実施形態では、分子タグは、ＤＮＡ内の第１の核酸塩基に対してＤＮＡ内の第２の核酸塩基とは異なって影響を及ぼす手順によって変更されるヌクレオチド、例えば、本明細書に記載されるもののうちのいずれも含まない（例えば、手順がバイサルファイト変換またはＣに影響を及ぼす任意の他の変換である場合、タグは、非改変Ｃを含まず、手順がｍＣに影響を及ぼす変換である場合、タグは、ｍＣを含まず、手順がｈｍＣに影響を及ぼす変換である場合、タグは、ｈｍＣを含まないなど）。

ある特定の開示される方法の追加の特色

Ａ．試料

【０４３０】

試料は、対象から単離された任意の生体試料であり得る。試料は、身体試料であり得る。試料としては、体組織、例えば、判明しているかまたは疑われる固形腫瘍、全血、血小板、血清、血漿、糞便、赤血球、白血球（white blood cell）または白血球（leucocyte）、内皮細胞、組織生検、脳脊髄液　滑液、リンパ液、腹水、間質液または細胞外液、細胞間の空間の流体が挙げられ得、歯肉溝滲出液、骨髄、胸水、脳脊髄液、唾液、粘

液、喀痰、精液、汗、尿が挙げられる。試料は、好ましくは、体液、特に、血液およびその分画、ならびに尿である。試料は、対象からもともと単離されている形態であってもよく、または成分、例えば、細胞を除去もしくは追加するため、または１つの構成要素を別の構成要素と比べて富化するために、さらなるプロセシングに供されていてもよい。したがって、解析のための好ましい体液は、無細胞核酸を含有する血漿または血清である。

【０４３１】

一部の実施形態では、核酸の集団は、新生物、腫瘍、前がん状態、もしくはがんを有することが疑われるか、または新生物、腫瘍、前がん状態、もしくはがんを有すると以前に診断されている対象に由来する血清、血漿または血液試料から得られる。集団は、様々なレベルの配列バリエーション、エピジェネティックバリエーション、および／または複製もしくは転写後改変を有する核酸を含む。複製後改変としては、シトシンの改変、特に核酸塩基の５位におけるもの、例えば、５－メチルシトシン、５－ヒドロキシメチルシトシン、５－ホルミルシトシンおよび５－カルボキシルシトシンが挙げられる。

【０４３２】

試料は、対象から単離または取得することができ、試料解析の現場へと輸送され得る。試料は、望ましい温度、例えば、室温、４℃、－２０℃、および／または－８０℃で保存または発送され得る。試料は、試料解析の現場で、対象から単離または取得することができる。対象は、ヒト、哺乳動物、動物、コンパニオン動物、介助動物、または愛玩動物であり得る。対象は、がん、前がん状態、感染、移植片拒絶、または免疫系の変化に関連する他の疾患もしくは障害を有し得る。対象は、がんまたは検出可能ながんの症状を有さない場合がある。対象は、１つまたは複数のがん治療、例えば、化学療法、抗体、ワクチン、または生物学のうちのいずれか１つまたは複数で処置されていてもよい。対象は、寛解期にあってもよい。対象は、がんまたは任意のがん関連遺伝子突然変異／障害に罹患しやすいと診断されていてもよく、またはされていなくてもよい。

【０４３３】

一部の実施形態では、試料は、血漿を含む。得られる血漿の体積は、配列決定される領域の所望のリード深度に依存し得る。例示的な体積は、０．４～４０ｍｌ、５～２０ｍｌ、１０～２０ｍｌである。例えば、体積は、０．５ｍＬ、１ｍＬ、５ｍＬ　１０ｍＬ、２０ｍＬ、３０ｍＬ、または４０ｍＬであり得る。試料採取される血漿の体積は、５～２０ｍＬであり得る。

【０４３４】

試料は、ゲノム等価物を含有する様々な量の核酸を含み得る。例えば、約３０ｎｇのＤＮＡの試料は、約１０，０００（10^4）半数体ヒトゲノム等価物を含有し、ｃｆＤＮＡの場合、約２０００億（2×10^n）個の個々のポリヌクレオチド分子を含有し得る。同様に、約１００ｎｇのＤＮＡの試料は、約３０，０００個の半数体ヒトゲノム等価物を含有し、ｃｆＤＮＡの場合、約６０００億個の個々の分子を含有し得る。

【０４３５】

試料は、異なる起源、例えば、同じ対象の細胞および無細胞、異なる対象の細胞および無細胞に由来する核酸を含み得る。試料は、突然変異を有する核酸を含み得る。例えば、試料は、生殖系列変異および／または体細胞変異を有するＤＮＡを含み得る。生殖系列変異とは、対象の生殖系列ＤＮＡに存在する変異を指す。体細胞変異とは、対象の体細胞、例えば、前がん状態細胞またはがん細胞を起源とする変異を指す。試料は、がん関連突然変異（例えば、がん関連体細胞突然変異）を有するＤＮＡを含み得る。試料は、エピジェネティック変異体（すなわち、化学的またはタンパク質改変）を含み得、ここで、エピジェネティック変異体は、遺伝的変異体、例えば、がん関連突然変異の存在と関連付けられる。一部の実施形態では、試料は、遺伝的変異体の存在と関連するエピジェネティック変異体を含み、ここで、試料は、遺伝的変異体を含まない。

【０４３６】

増幅前の試料中の無細胞核酸の例示的な量は、約１ｆｇ～約１ｐｇ、例えば、１ｐｇ～２００ｎｇ、１ｎｇ～１００ｎｇ、１０ｎｇ～１０００ｎｇの範囲である。例えば、量は

、最大約６００ｎｇ、最大約５００ｎｇ、最大約４００ｎｇ、最大約３００ｎｇ、最大約２００ｎｇ、最大約１００ｎｇ、最大約５０ｎｇ、または最大約２０ｎｇの無細胞核酸分子であり得る。量は、少なくとも１ｆｇ、少なくとも１０ｆｇ、少なくとも１００ｆｇ、少なくとも１ｐｇ、少なくとも１０ｐｇ、少なくとも１００ｐｇ、少なくとも１ｎｇ、少なくとも１０ｎｇ、少なくとも１００ｎｇ、少なくとも１５０ｎｇ、または少なくとも２００ｎｇの無細胞核酸分子であり得る。量は、最大１フェムトグラム（ｆｇ）、１０ｆｇ、１００ｆｇ、１ピコグラム（ｐｇ）、１０ｐｇ、１００ｐｇ、１ｎｇ、１０ｎｇ、１００ｎｇ、１５０ｎｇ、または２００ｎｇの無細胞核酸分子であり得る。方法は、１フェムトグラム（ｆｇ）～２００ｎｇを得ることを含み得る－［０３２６］無細胞核酸は、細胞に含有されていないもしくは他の方法で結合していない核酸、または言い換えれば、インタクトな細胞を除去した後に試料中に残留している核酸である。無細胞核酸としては、ＤＮＡ、ＲＮＡ、およびそれらのハイブリッドが挙げられ、これには、ゲノムＤＮＡ、ミトコンドリアＤＮＡ、ｓｉＲＮＡ、ｍｉＲＮＡ、循環ＲＮＡ（ｃＲＮＡ）、ｔＲＮＡ、ｒＲＮＡ、核小体低分子ＲＮＡ（ｓｎｏＲＮＡ）、Ｐｉｗｉ相互作用ＲＮＡ（ｐｉＲＮＡ）、長鎖非コードＲＮＡ（長鎖ｎｃＲＮＡ）、またはこれらのうちのいずれかの断片が含まれる。無細胞核酸は、二本鎖、一本鎖、またはそれらのハイブリッドであり得る。無細胞核酸は、分泌または細胞死プロセス、例えば、細胞壊死およびアポトーシスを通じて、体液中に放出され得る。一部の無細胞核酸は、がん細胞、例えば、循環腫瘍ＤＮＡ（ｃｔＤＮＡ）から、体液中に放出される。その他のものは、健康な細胞から放出される。一部の実施形態では、ｃｆＤＮＡは、無細胞胎児ＤＮＡ（ｃｆｆＤＮＡ）である。一部の実施形態では、無細胞核酸は、腫瘍細胞によって産生される。一部の実施形態では、無細胞核酸は、腫瘍細胞と非腫瘍細胞との混合物によって産生される。

【０４３７】

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

【０４３８】

無細胞核酸は、溶液中に見出される無細胞核酸を、インタクトな細胞および体液の他の不溶性構成成分から分離する、分画ステップによって、体液から単離することができる。分配するステップには、遠心分離または濾過などの技術が含まれ得る。あるいは、体液中の細胞を、溶解し、無細胞核酸および細胞核酸を、一緒にプロセシングしてもよい。一般に、緩衝液の添加および洗浄ステップの後に、核酸は、アルコールで沈殿させることができる。シリカに基づくカラムなどのさらなる洗浄ステップを使用して、夾雑物または塩を除去してもよい。バイサルファイトシーケンシング、ハイブリダイゼーション、および／またはライゲーションのための非特異的バルク担体核酸、例えば、Ｃ　１　ＤＮＡ、ＤＮＡ、またはタンパク質を、手順のある特定の態様、例えば、収量を最適化するために、反応全体にわたって添加してもよい。［０３２９］そのようなプロセシングの後、試料は、二本鎖ＤＮＡ、一本鎖ＤＮＡ、および一本鎖ＲＮＡを含む、様々な形態の核酸を含み得る。一部の実施形態では、一本鎖ＤＮＡおよびＲＮＡは、後続のプロセシングおよび解析ステップに含められるように、二本鎖形態に変換してもよい。［０３３０］ＤＮＡ分子は、一方の末端または両方の末端において、アダプターに連結され得る。典型的には、二本鎖分子は、４つ全ての標準的なヌクレオチドの存在下で、５’－３’ポリメラーゼおよび３’－５’エキソヌクレアーゼ（またはプルーフリーディング機能）を有するポリメラーゼを用いた処理によって平滑末端化される。クレノウラージ断片およびＴ４ポリメラーゼが、好適なポリメラーゼの例である。平滑末端化されたＤＮＡ分子は、少なくとも部分的に二本鎖のアダプター（例えば、Ｙ字形状またはベル形状のアダプター）とライゲーションされ得る。あるいは、ライゲーションを容易にするために、相補的なヌクレオチドが、試料核酸およびアダプターの平滑末端に付加されてもよい。平滑末端ライゲーションおよび付着末端ライゲーションの両方が、本明細書において企図される。平滑末端ライゲーションでは、核酸分子およびアダプタータグの両方が、平滑末端を有する。付着末端ライゲーシ

ョンでは、典型的には、核酸分子は、「Ａ」オーバーハングを有し、アダプターは、「Ｔ」オーバーハングを有する。

Ｂ．タグ

【０４３９】

バーコードを含むタグは、アダプターに組み込まれ得るか、またはそうでなければ、アダプターに接合され得る。タグは、他の方法の中でもとりわけ、ライゲーション、オーバーラップ伸長ＰＣＲによって、組み込まれ得る。

ｉ）分子タグ付け戦略

【０４４０】

分子タグ付けとは、配列リードの起源であるＤＮＡ分子の間を区別することを可能にするタグ付け実施を指す。タグ付け戦略は、一意的タグ付け戦略および非一意的タグ付け戦略に分けることができる。一意的タグ付けでは、試料中の全てまたは実質的に全ての分子が、異なるタグを有し、そのため、読み取りデータは、タグ情報単独に基づいて、もともとの分子に割り当てることができる。そのような方法において使用されるタグは、「一意的タグ」と称される場合がある。非一意的タグ付けでは、同じ試料中の異なる分子が同じタグを有し得、そのため、タグ情報に加えて他の情報を使用して、配列読み取りデータがもともとの分子に割り当てられる。そのような情報としては、開始および終止座標、分子がマッピングされる座標、開始または終止座標単独などを挙げることができる。そのような方法において使用されるタグは、「非一意的タグ」と称される場合がある。したがって、必ずしも、試料中の全ての分子を一意的にタグ付けするわけではない。試料内の同定可能なクラス内に含まれる分子を一意的にタグ付けすることで十分である。したがって、異なる同定可能なファミリー内の分子は、タグ付けされた分子の同一性に関する情報を失うことなく、同じタグを有し得る。

【０４４１】

非一意的タグ付けのある特定の実施形態では、使用される異なるタグの数は、特定の群の全てのＤＮＡ分子が異なるタグを有する非常に高い可能性（例えば、少なくとも９９％、少なくとも９９．９％、少なくとも９９．９９％、または少なくとも９９．９９９％）であれば、十分であり得る。バーコードをタグとして使用する場合、およびバーコードが、例えば、ランダムに分子の両方の末端に付着している場合、バーコードの組合せが、一緒に、タグを構成し得ることに留意されたい。この数は、明確にコールに含まれる分子の数の関数である。例えば、クラスは、参照ゲノム上の同じ開始－終止位置にマッピングされる全ての分子であり得る。クラスは、特定の遺伝子座、例えば、特定の塩基または特定の領域（例えば、最大１００個の塩基または遺伝子または遺伝子のエクソン）にわたってマッピングする全ての分子であり得る。ある特定の実施形態では、クラス内の分子の数ｚを一意的に同定するために使用される異なるタグの数は、２＊ｚ、３＊ｚ、４＊ｚ、５＊ｚ、６＊ｚ、７＊ｚ、８＊ｚ、９＊ｚ、１０＊ｚ、１１＊ｚ、１２＊ｚ、１３＊ｚ、１４＊ｚ、１５＊ｚ、１６＊ｚ、１７＊ｚ、１８＊ｚ、１９＊ｚ、２０＊ｚ、または１００＊ｚのうちのいずれか（例えば、下限）と、１００，０００＊ｚ、１０，０００＊ｚ、１０００＊ｚ、または１００＊ｚのうちのいずれか（例えば、上限）との間であり得る。

【０４４２】

例えば、約５ｎｇ～３０ｎｇの無細胞ＤＮＡの試料において、およそ３０００個の分子が、特定のヌクレオチド座標にマッピングされ、任意の開始座標を有する約３～１０個の間の分子が、同じ終止座標を共有することが予測される。したがって、約５０～約５０，０００個の異なるタグ（例えば、約６～２２０個の間のバーコードの組合せ）が、全てのそのような分子を一意的にタグ付けするのに十分であり得る。ヌクレオチド座標全体にマッピングされる３０００個全ての分子を一意的にタグ付けするためには、約１００万個～約２０００万個の異なるタグが必要となるであろう。

［０３３６］

一般に、反応における一意的または非一意的タグバーコードの割り当ては、米国特許出願第２００１００５３５１９号、同第２００３０１５２４９０号、同第２０１１０１６０

０７８号、ならびに米国特許第６，５８２，９０８号および米国特許第７，５３７，８９８号および米国特許第９，５９８，７３１号によって記載される方法およびシステムに従う。タグは、ランダムまたは非ランダムに、試料核酸に連結され得る。

［０３３７］

一意的タグは、ゲノム試料あたり約１個を上回る、２個を上回る、３個を上回る、４個を上回る、５個を上回る、６個を上回る、７個を上回る、８個を上回る、９個を上回る、１０個を上回る、２０個を上回る、５０個を上回る、１００個を上回る、５００個を上回る、１０００個を上回る、５０００個を上回る、１００００個を上回る、５０，０００個を上回る、１００，０００個を上回る、５００，０００個を上回る、１，０００，０００個を上回る、１０，０００，０００個を上回る、５０，０００，０００個を上回る、または１，０００，０００，０００個を上回る一意的タグがロードされるように、ロードされ得る。一部の場合には、一意的タグは、ゲノム試料あたり約２個を下回る、３個を下回る、４個を下回る、５個を下回る、６個を下回る、７個を下回る、８個を下回る、９個を下回る、１０個を下回る、２０個を下回る、５０個を下回る、１００個を下回る、５００個を下回る、１０００個を下回る、５０００個を下回る、１００００個を下回る、５０，０００個を下回る、１００，０００個を下回る、５００，０００個を下回る、１，０００，０００個を下回る、１０，０００，０００個を下回る、５０，０００，０００個を下回る、または１，０００，０００，０００個を下回る一意的タグがロードされるように、ロードされ得る。一部の場合には、試料ゲノムあたりのロードされる一意的タグの平均数は、ゲノム試料あたり約１個未満もしくはそれより多くの、２個未満もしくはそれより多くの、３個未満もしくはそれより多くの、４個未満もしくはそれより多くの、５個未満もしくはそれより多くの、６個未満もしくはそれより多くの、７個未満もしくはそれより多くの、８個未満もしくはそれより多くの、９個未満もしくはそれより多くの、１０個未満もしくはそれより多くの、２０個未満もしくはそれより多くの、５０個未満もしくはそれより多くの、１００個未満もしくはそれより多くの、５００個未満もしくはそれより多くの、１０００個未満もしくはそれより多くの、５０００個未満もしくはそれより多くの、１００００個未満もしくはそれより多くの、５０，０００個未満もしくはそれより多くの、１００，０００個未満もしくはそれより多くの、５００，０００個未満もしくはそれより多くの、１，０００，０００個未満もしくはそれより多くの、１０，０００，０００個未満もしくはそれより多くの、５０，０００，０００個未満もしくはそれより多くの、または１，０００，０００，０００個未満もしくはそれより多くの一意的タグである。

【０４４３】

好ましいフォーマットは、標的核酸の両方の末端にライゲーションされた２０～５０個の異なるタグ（例えば、バーコード）を使用する。例えば、標的分子の両方の末端に、３５個の異なるタグ（例えば、バーコード）をライゲーションすると、３５×３５の順列を作製し、これは、３５個のタグについて１２２５個に等しい。そのような数のタグは、同じ開始点および終止点を有する異なる分子が、高い確率（例えば、少なくとも９４％、９９．５％、９９．９９％、９９．９９９％）で異なる組合せのタグを受容するのに十分である。他のバーコードの組合せとしては、１０～５００の間の任意の数、例えば、約１５×１５、約３５×３５、約７５×７５、約１００×１００、約２５０×２５０、約５００×５００が挙げられる。

【０４４４】

一部の場合には、一意的タグは、所定のまたはランダムなまたは半ランダムな配列のオリゴヌクレオチドであり得る。他の場合には、バーコードが複数の中で必ずしも互いに一意的とならないような、複数のバーコードが使用されてもよい。この例では、バーコードは、バーコードおよびそれがライゲーションされ得る配列の組合せが、個個々に追跡され得る一意的配列を作製するように、個々の分子にライゲーションされ得る。本明細書に記載されるように、配列読み取りデータの始まり（開始）および終わり（終止）の部分の配列データと組み合わせた非一意的バーコードの検出により、特定の分子に対する一意的同一性の割り当てが可能となり得る。個々の配列読み取りデータの長さまたは塩基対の数も

また、そのような分子に一意的同一性を割り当てるために使用することができる。本明細書に記載されるように、一意的同一性が割り当てられた核酸の一本鎖に由来する断片は、それによって、後続の親鎖に由来する断片の同定を可能にし得る。

Ｃ．増幅

【０４４５】

アダプターに挟まれている試料核酸は、ＰＣＲおよび他の増幅方法によって増幅され得る。増幅は、典型的には、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とアニーリングまたは結合するプライマーによってプライミングされる。増幅方法は、サーモサイクリングの結果としての変性、アニーリング、および伸長のサイクルを含み得るか、または転写媒介増幅におけるような等温のものであってもよい。他の増幅方法としては、リガーゼ連鎖反応、鎖置換増幅、核酸配列に基づく増幅、および自家持続性配列に基づく複製法が挙げられる。

【０４４６】

一部の実施形態では、本方法は、ＴテールおよびＣテールアダプターを用いてｄｓＤＮＡライゲーションを行い、これにより、アダプターへ連結する前に、二本鎖核酸のうちの少なくとも５０、６０、７０、または８０％の増幅がもたらされる。好ましくは、本方法は、Ｔテールアダプターを単独で用いて行われる対照方法と比べて、増幅される分子の量または数を、少なくとも１０、１５、または２０％増加させる。

Ｄ．捕捉部分

【０４４７】

上記で考察されたように、試料中の核酸は、捕捉ステップに供することができ、ここで、標的領域を有する分子は、後続の解析のために捕捉される。標的の捕捉は、捕捉部分、例えば、ビオチンで標識されたプローブ（例えば、オリゴヌクレオチド）、および捕捉部分に結合する第２の部分または結合パートナー、例えばストレプトアビジンの使用を伴い得る。一部の実施形態では、捕捉部分および結合パートナーは、本明細書の他の箇所で考察されるように、標的領域の異なるセット、例えば、それぞれ、配列可変標的領域セットおよびエピジェネティック標的領域セットのものに関して、より高い捕捉収率およびより低い捕捉収率を有し得る。捕捉部分を含む方法は、例えば、参照により本明細書に組み込まれる、２０１７年１２月２６日に発行された米国特許９，８５０，５２３においてさらに記載されている。

【０４４８】

捕捉部分としては、限定することなく、ビオチン、アビジン、ストレプトアビジン、特定のヌクレオチド配列を含む核酸、抗体によって認識されるハプテン、および磁気吸着可能な粒子が挙げられる。抽出部分は、結合ペア、例えば、ビオチン／ストレプトアビジンまたはハプテン／抗体のメンバーであり得る。一部の実施形態では、検体に付着した捕捉部分は、単離可能な部分、例えば、磁気吸着可能な粒子または遠心分離によって沈降させることができる大きい粒子に付着した、その結合ペアによって捕捉される。捕捉部分は、捕捉部分を有する核酸の、捕捉部分が欠如した核酸からの親和性分離を可能にする、任意のタイプの分子であり得る。例示的な捕捉部分は、固相に連結しているかもしくは連結可能なストレプトアビジンへの結合による親和性分離を可能にするビオチン、または固相に連結しているかもしくは連結可能な相補的なオリゴヌクレオチドへの結合を通じた親和性分離を可能にするオリゴヌクレオチドである。

Ｅ．標的特異的プローブのコレクション

【０４４９】

一部の実施形態では、標的特異的プローブのコレクションを、本明細書に記載されるエピジェネティック標的領域セットおよび／または配列可変標的領域セットを含む方法において使用する。一部の実施形態では、標的特異的プローブのコレクションは、配列可変標的領域セットに対して特異的な標的結合プローブおよびエピジェネティック標的領域セットに対して特異的な標的結合プローブを含む。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの捕捉収率は、エピジェネティック標的領域セッ

トに対して特異的な標的結合プローブの捕捉収率よりも高い（例えば、少なくとも２倍高い）。一部の実施形態では、標的特異的プローブのコレクションは、エピジェネティック標的領域セットに対して特異的なその捕捉収率よりも高い（例えば、少なくとも２倍高い）配列可変標的領域セットに対して特異的な捕捉収率を有するように構成される。

【０４５０】

一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの捕捉収率は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの捕捉収率よりも少なくとも１．２５倍、１．５倍、１．７５倍、２倍、２．２５倍、２．５倍、２．７５倍、３倍、３．５倍、４倍、４．５倍、５倍、６倍、７倍、８倍、９倍、１０倍、１１倍、１２倍、１３倍、１４倍、または１５倍高い。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの捕捉収率は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの捕捉収率よりも１．２５倍～１．５倍、１．５倍～１．７５倍、１．７５倍～２倍、２倍～２．２５倍、２．２５倍～２．５倍、２．５倍～２．７５倍、２．７５倍～３倍、３倍～３．５倍、３．５倍～４倍、４倍～４．５倍、４．５倍～５倍、５倍～５．５倍、５．５倍～６倍、６倍～７倍、７倍～８倍、８倍～９倍、９倍～１０倍、１０倍～１１倍、１１倍～１２倍、１３倍～１４倍、または１４倍～１５倍高い。

【０４５１】

一部の実施形態では、標的特異的プローブのコレクションは、エピジェネティック標的領域セットに対するその捕捉収率よりも少なくとも１．２５倍、１．５倍、１．７５倍、２倍、２．２５倍、２．５倍、２．７５倍、３倍、３．５倍、４倍、４．５倍、５倍、６倍、７倍、８倍、９倍、１０倍、１１倍、１２倍、１３倍、１４倍、または１５倍高い、配列可変標的領域セットに対して特異的な捕捉収率を有するように構成される。一部の実施形態では、標的特異的プローブのコレクションは、エピジェネティック標的領域セットに対して特異的なその捕捉収率よりも１．２５倍～１．５倍、１．５倍～１．７５倍、１．７５倍～２倍、２倍～２．２５倍、２．２５倍～２．５倍、２．５倍～２．７５倍、２．７５倍～３倍、３倍～３．５倍、３．５倍～４倍、４倍～４．５倍、４．５倍～５倍、５倍～５．５倍、５．５倍～６倍、６倍～７倍、７倍～８倍、８倍～９倍、９倍～１０倍、１０倍～１１倍、１１倍～１２倍、１３倍～１４倍、または１４倍～１５倍高い、配列可変標的領域セットに対して特異的な捕捉収率を有するように構成される。

【０４５２】

プローブのコレクションは、濃度、様々な長さ、および／または化学（例えば、親和性に影響を及ぼす）、ならびにそれらの組合せを含む様々な方法で配列可変標的領域セットに対してより高い捕捉収率を提供するように構成され得る。親和性は、以下に考察するように、プローブの長さを調整する、および／またはヌクレオチド改変を含めることによってモジュレートすることができる。

【０４５３】

一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、エピジェネティック標的領域セットに対して特異的な標的特異的プローブよりも高い濃度で存在する。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの濃度は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの濃度よりも少なくとも１．２５倍、１．５倍、１．７５倍、２倍、２．２５倍、２．５倍、２．７５倍、３倍、３．５倍、４倍、４．５倍、５倍、６倍、７倍、８倍、９倍、１０倍、１１倍、１２倍、１３倍、１４倍、または１５倍高い。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの濃度は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの濃度よりも１．２５倍～１．５倍、１．５倍～１．７５倍、１．７５倍～２倍、２倍～２．２５倍、２．２５倍～２．５倍、２．５倍～２．７５倍、２．７５倍～３倍、３倍～３．５倍、３．５倍～４倍、４倍～４．５倍、４．５倍～５倍、５倍～５．５倍、５．５倍～６倍、６倍～７倍、７倍～８倍、８倍～９倍、９倍～１０倍、１０倍～１１倍、１１倍～１２倍、１３倍～１４倍、または

１４倍～１５倍高い。そのような実施形態では、濃度は、各セット中の個々のプローブの体積濃度あたりの平均質量を指し得る。

【０４５４】

一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、エピジェネティック標的領域セットに対して特異的な標的特異的プローブよりもそれらの標的に対する高い親和性を有する。親和性は、異なるプローブ化学を使用することによってを含む、当業者に公知の任意の方法でモジュレートすることができる。ある特定のヌクレオチド改変、例えばシトシン５－メチル化（ある特定の配列の文脈において）、Ｔ糖のヘテロ原子を提供する改変、およびＬＮＡヌクレオチドは、二本鎖核酸の安定性を増加させることができ、そのような改変を有するオリゴヌクレオチドが、その相補的配列に関して比較的高い親和性を有することを示している。例えば、Severin et al., Nucleic Acids Res. 39: 8740-8751 (2011); Freier et al., Nucleic Acids Res. 25: 4429-4443 (1997)；米国特許第９，７３８，８９４号を参照されたい。同様に、より長い配列長は、一般的に親和性の増加を提供する。他のヌクレオチド改変、例えば、核酸塩基ヒポキサンチンをグアニンの代わりに置換すると、オリゴヌクレオチドとその相補配列の間の水素結合の量を低減させることによって親和性を低減する。一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、その標的に対するその親和性を増加させる改変を有する。一部の実施形態では、あるいはまたはさらに、エピジェネティック標的領域セットに対して特異的な標的特異的プローブは、その標的に対するその親和性を減少させる改変を有する。一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、エピジェネティック標的領域セットに対して特異的な標的特異的プローブよりも長い平均長および／または高い平均融解温度を有する。これらの実施形態を互いに、および／または上記で考察した濃度の差と組み合わせて、上記の任意の倍率の差またはその範囲などの、捕捉収率の所望の倍率の差を達成してもよい。

【０４５５】

一部の実施形態では、標的特異的プローブは、捕捉部分を含む。捕捉部分は、本明細書に記載される捕捉分子のいずれか、例えばビオチンであってもよい。一部の実施形態では、標的特異的プローブは、固体支持体に、例えば捕捉部分の結合対の相互作用などを通して共有結合または非共有結合により連結される。一部の実施形態では、固体支持体はビーズ、例えば磁気ビーズである。

【０４５６】

一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブおよび／またはエピジェネティック標的領域セットに対して特異的な標的特異的プローブは、上記で考察されたような捕捉部分、例えば、捕捉部分および遺伝子などの領域のパネルにわたってタイリングするために選択された配列を含むプローブを含む。

【０４５７】

一部の実施形態では、標的特異的プローブは、単一の組成物中で提供される。

【０４５８】

単一の組成物は溶液（液体または凍結）であり得る。あるいは、これは凍結乾燥物であり得る。

【０４５９】

あるいは、標的特異的プローブは、例えば、エピジェネティック標的領域セットに対して特異的なプローブを含む第１の組成物、および配列可変標的領域セットに対して特異的なプローブを含む第２の組成物を含む、複数の組成物として提供され得る。これらのプローブを適切な割合で混合して、濃度および／または捕捉収率の前述の倍率の差のいずれかを有する組み合わせたプローブ組成物を提供してもよい。あるいは、それらを個別の捕捉手順（例えば、試料のアリコートを用いて、または同じ試料で逐次的に）で使用して、捕捉されたエピジェネティック標的領域および配列可変標的領域をそれぞれ含む第１および第２の組成物を提供してもよい。

ｉ）エピジェネティック標的領域に対して特異的なプローブ

【０４６０】

エピジェネティック標的領域セットに対するプローブは、希少な免疫細胞タイプを含む、異なるタイプの免疫細胞を起源とするＤＮＡを区別する、および／または、前がん性もしくは新生物（例えば、腫瘍もしくはがん）細胞由来のＤＮＡを、健康な細胞、例えば、非新生物の循環細胞と区別する可能性がある１つまたは複数のタイプの標的領域に対して特異的なプローブを含み得る。そのような領域の例示的なタイプは、本明細書において詳細に考察されている。エピジェネティック標的領域セットのプローブはまた、例えば本明細書に記載される１つまたは複数の対照領域のプローブも含み得る

【０４６１】

一部の実施形態では、エピジェネティック標的領域プローブセットに対するプローブは、少なくとも１００ｋｂ、例えば、少なくとも２００ｋｂ、少なくとも３００ｋｂ、または少なくとも４００ｋｂのフットプリントを有する。一部の実施形態では、エピジェネティック標的領域セットに対するプローブは、１００～１０００ｋｂ、例えば、１００～２００ｋｂ、２００～３００ｋｂ、３００～４００ｋｂ、４００～５００ｋｂ、５００～６００ｋｂ、６００～７００ｋｂ、７００～８００ｋｂ、８００～９００ｋｂ、および９００～１，０００ｋｂの範囲のフットプリントを有する。一部の実施形態では、エピジェネティック標的領域プローブセットに対するプローブは、少なくとも５ｋｂ、例えば、少なくとも１０ｋｂ、２０ｋｂ、または５０ｋｂのフットプリントを有する。

ａ．高メチル化標的領域。

【０４６２】

一部の実施形態では、メチル化感受性変換（例えば、バイサルファイトまたはＥＭ－ｓｅｑ）を使用する方法に関して、プローブは、メチル化感受性変換のタイプおよび富化される標的領域に応じて、変換された分子または変換されなかった分子のいずれかを標的とするように設計することができる。例えば、バイサルファイト処理を使用する場合、ＤＮＡ分子内の非メチル化シトシンはジヒドロウラシルに変換され、メチル化シトシンは変換されずにシトシンのままになる。高メチル化標的領域（調査中のがんまたは任意の他の疾患に対する目的の分子が高メチル化されている）内のＤＮＡ分子を捕捉するために、プローブを、変換されなかった分子を捕捉するように設計することができ、一方で、低メチル化標的領域（調査中のがんまたは任意の他の疾患に対する目的の分子が低メチル化されているまたはメチル化されていない）内の分子を捕捉するために、プローブを、変換された分子を捕捉するように設計することができる。

【０４６３】

一部の実施形態では、エピジェネティック標的領域セットに対するプローブは、１つまたは複数の高メチル化標的領域に対して特異的なプローブを含む。高メチル化標的領域は、上記のもののいずれかであってよい。例えば、一部の実施形態では、高メチル化標的領域に対して特異的なプローブは、異なる免疫細胞タイプにおいて差次的にメチル化される複数の座に対して特異的なプローブを含む。一部の実施形態では、各免疫細胞タイプ特異的高メチル化標的領域は、１つの免疫細胞タイプにおいて０．３、０．４、０．５、または０．６よりも高いまたはそれに等しい頻度で、および他の全ての免疫細胞タイプにおいて０．１、０．２、または０．３よりも低いまたはそれに等しい頻度でメチル化される少なくとも１つのＣｐＧ部位を含む。一部の実施形態では、各免疫細胞タイプ特異的高メチル化標的領域は、互いから１００塩基対以内に少なくとも２つのＣｐＧ部位を含み、１つの免疫細胞タイプにおいて０．３、０．４、０．５、または０．６よりも高いまたはそれに等しい頻度で、他の全ての免疫細胞タイプにおいて０．１、０．２、または０．３よりも低いまたはそれに等しい頻度でそれぞれメチル化されている。一部のそのような実施形態では、各免疫細胞タイプ特異的高メチル化標的領域は、１５０塩基対以内または２００塩基対以内に合計で少なくとも２、３、４、５、６、７、８、９、または１０個のＣｐＧ部位を含み、ここで、任意の正常な組織タイプにおいて、少なくとも２、３、４、５、６、７、８、９、または１０個のＣｐＧ部位のうちの３つより少数が、０．１、０．２、ま

たは０．３よりも高い頻度でメチル化されている。一部の実施形態では、各免疫細胞タイプ特異的エピジェネティック標的領域セットは、方法において同定される免疫細胞タイプのひとつひとつにおいて一意的に高メチル化された少なくとも３、少なくとも５、少なくとも１０、少なくとも２０、または少なくとも３０の高メチル化標的領域を含む。

【０４６４】

一部の実施形態では、高メチル化標的領域に対して特異的なプローブは、表１に列挙されている複数の座、例えば、表１に列挙されている座の少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、または１００％に対して特異的なプローブを含む。一部の実施形態では、高メチル化標的領域に対して特異的なプローブは、表２に列挙される複数の座、例えば表２に列挙される座の少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、または１００％に対して特異的なプローブを含む。一部の実施形態では、高メチル化標的領域に対して特異的なプローブは、表１または表２に列挙される複数の座、例えば表１または表２に列挙される座の少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、または１００％に対して特異的なプローブを含む。

【０４６５】

一部の実施形態では、標的領域として含まれる各座に関して、遺伝子の転写開始部位と終止コドン（選択的にスプライシングされる遺伝子の場合は最後の終止コドン）の間に結合するハイブリダイゼーション部位を有する１つまたは複数のプローブが存在し得る。一部の実施形態では、１つまたは複数のプローブは、列挙された位置の３００ｂｐ以内、例えば２００または１００ｂｐ以内に結合する。一部の実施形態では、プローブは、上に列挙された位置と重複するハイブリダイゼーション部位を有する。一部の実施形態では、高メチル化標的領域に対して特異的なプローブは、乳がん、結腸がん、腎臓がん、肝臓がん、および肺がんのうちの１、２、３、４、または５つにおいて集合的に高メチル化を示す高メチル化標的領域の１、２、３、４、または５つのサブセットに対して特異的なプローブを含む。

ｂ．低メチル化標的領域。

【０４６６】

一部の実施形態では、エピジェネティック標的領域セットに対するプローブは、１つまたは複数の低メチル化標的領域に対して特異的なプローブを含む。低メチル化標的領域は、上記のもののいずれかであってよい。例えば、一部の実施形態では、低メチル化標的領域に対して特異的なプローブは、異なる免疫細胞タイプにおいて差次的にメチル化される複数の座に対して特異的なプローブを含む。一部の実施形態では、各免疫細胞タイプ特異的低メチル化標的領域は、１つの免疫細胞タイプにおいて０．１、０．２、または０．３よりも低いまたはそれに等しい頻度で、および他の全ての免疫細胞タイプにおいて０．３、０．４、０．５、または０．６よりも高いまたはそれに等しい頻度でメチル化されている少なくとも１つのＣｐＧ部位を含む。一部の実施形態では、各免疫細胞タイプ特異的低メチル化標的領域は、互いから１００塩基対以内に少なくとも２つのＣｐＧ部位を含み、１つの免疫細胞タイプにおいて０．１、０．２、または０．３よりも低いまたはそれに等しい頻度で、他の全ての免疫細胞タイプにおいて０．３、０．４、０．５、または０．６よりも高いまたはそれに等しい頻度でそれぞれメチル化されている。一部のそのような実施形態では、各免疫細胞タイプ特異的低メチル化標的領域は、１５０塩基対以内または２００塩基対以内に合計で少なくとも２、３、４、５、６、７、８、９、または１０個のＣｐＧ部位を含み、ここで、任意の正常な組織タイプにおいて、少なくとも２、３、４、５、６、７、８、９、または１０個のＣｐＧ部位のうちの３つよりも少数が、０．１、０．２、または０．３未満の頻度でメチル化されている。一部の実施形態では、各免疫細胞タイプ特異的エピジェネティック標的領域セットは、方法において同定される免疫細胞タイプのひとつひとつにおいて一意的に低メチル化された、少なくとも３、少なくとも５、少なくとも１０、少なくとも２０、または少なくとも３０の低メチル化標的領域を含む。

【０４６７】

一部の実施形態では、１つまたは複数の低メチル化標的領域に対して特異的なプローブは、反復エレメント、例えば、ＬＩＮＥ１エレメント、Ａｌｕエレメント、セントロメアタンデムリピート、ペリセントロメアタンデムリピート、およびサテライトＤＮＡなどの領域に対するプローブを含んでもよく、健康な細胞において通常メチル化されている遺伝子間領域は、腫瘍細胞において低減されたメチル化を示し得る。

【０４６８】

一部の実施形態では、低メチル化標的領域に対して特異的なプローブは、反復エレメントおよび／または遺伝子間領域に対して特異的なプローブを含む。一部の実施形態では、反復エレメントに対して特異的なプローブは、ＬＩＮＥ１エレメント、Ａｌｕエレメント、セントロメアタンデムリピート、ペリセントロメアタンデムリピート、および／またはサテライトＤＮＡのうちの１、２、３、４、または５個に対して特異的なプローブを含む。

【０４６９】

がん関連低メチル化を示すゲノム領域に対して特異的な例示的なプローブは、ヒト第１染色体のヌクレオチド８４０３５６５～８９５３７０８および／または１５１１０４７０１～１５１１０６０３５に対して特異的なプローブを含む。一部の実施形態では、低メチル化標的領域に対して特異的なプローブは、ヒト第１染色体のヌクレオチド８４０３５６５～８９５３７０８および／または１５１１０４７０１～１５１１０６０３５と重複するまたはそれらを含む領域に対して特異的なプローブを含む。

【０４７０】

一部の実施形態では、エピジェネティック標的領域セットのプローブは、ＣＴＣＦ結合領域に対して特異的なプローブを含む。一部の実施形態では、ＣＴＣＦ結合領域に対して特異的なプローブは、少なくとも１０、２０、５０、１００、２００、もしくは５００個のＣＴＣＦ結合領域、または１０～２０、２０～５０、５０～１００、１００～２００、２００～５００、もしくは５００～１０００個のＣＴＣＦ結合領域、例えば上記の、またはＣＴＣＦＢＳＤＢの１つもしくは複数、または上記で引用したＣｕｄｄａｐａｈら、Ｍａｒｔｉｎら、もしくはＲｈｅｅらの論文におけるＣＴＣＦ結合領域などに対して特異的なプローブを含む。一部の実施形態では、エピジェネティック標的領域セットのプローブは、ＣＴＣＦ結合部位の少なくとも１００ｂｐ、少なくとも２００ｂｐ、少なくとも３００ｂｐ、少なくとも４００ｂｐ、少なくとも５００ｂｐ、少なくとも７５０ｂｐ、または少なくとも１０００ｂｐ上流および下流の領域を含む。

ｄ．転写開始部位。

【０４７１】

一部の実施形態では、エピジェネティック標的領域セットのプローブは、転写開始部位に対して特異的なプローブを含む。一部の実施形態では、転写開始部位に対して特異的なプローブは、少なくとも１０、２０、５０、１００、２００、もしくは５００個の転写開始部位、または１０～２０、２０～５０、５０～１００、１００～２００、２００～５００、もしくは５００～１０００個の転写開始部位、例えばＤＢＴＳＳに列挙される転写開始部位などに対して特異的なプローブを含む。一部の実施形態では、エピジェネティック標的領域セットのプローブは、転写開始部位の少なくとも１００ｂｐ、少なくとも２００ｂｐ、少なくとも３００ｂｐ、少なくとも４００ｂｐ、少なくとも５００ｂｐ、少なくとも７５０ｂｐ、または少なくとも１０００ｂｐ上流および下流の配列のプローブを含む。

【０４７２】

上記のように、局所増幅が体細胞突然変異であるが、それらは、メチル化の変化などのある特定のエピジェネティック変化を検出するためのアプローチと類似の方法で読み取りデータの頻度に基づいてシーケンシングすることによって検出することができる。そのため、がんにおいて局所増幅を示し得る領域を、上記で考察したようにエピジェネティック標的領域セットに含めることができる。一部の実施形態では、エピジェネティック標的領域セットに対して特異的なプローブは、局所増幅に対して特異的なプローブを含む。一部の実施形態では、局所増幅に対して特異的なプローブは、ＡＲ、ＢＲＡＦ、ＣＣＮＤ１、

ＣＣＮＤ２、ＣＣＮＥ１、ＣＤＫ４、ＣＤＫ６、ＥＧＦＲ、ＥＲＢＢ２、ＦＧＦＲ１、ＦＧＦＲ２、ＫＩＴ、ＫＲＡＳ、ＭＥＴ、ＭＹＣ、ＰＤＧＦＲＡ、ＰＩＫ３ＣＡ、およびＲＡＦ１のうちの１つまたは複数に対して特異的なプローブを含む。例えば、一部の実施形態では、局所増幅に対して特異的なプローブは、前述の標的の少なくとも２、３、４、５、６、７、８、９、１０、１１、１２、１３、１４、１５、１６、１７、または１８個のうちの１つまたは複数に対して特異的なプローブを含む。

対照領域

【０４７３】

一部の実施形態では、エピジェネティック標的領域セットに対して特異的なプローブは、本質的に全ての試料においてメチル化されると予想される陽性対照領域に対して特異的なプローブを含む。一部の実施形態では、エピジェネティック標的領域セットに対して特異的なプローブは、本質的に全ての試料において低メチル化されているまたはメチル化されていないと予想される陰性対照領域に対して特異的なプローブを含む。

２）配列可変標的領域に対して特異的なプローブ

【０４７４】

配列可変標的領域セットのプローブは、がんにおいて体細胞突然変異を受けることが公知である複数の領域に対して特異的なプローブを含み得る。プローブは、本明細書に記載される任意の配列可変標的領域セットに対して特異的であり得る。例示的な配列可変標的領域セットは、本明細書において、例えば捕捉されたセットに関する上記の節で詳細に考察される。［０３６６］一部の実施形態では、配列可変標的領域プローブセットは、少なくとも０．５ｋｂ、例えば、少なくとも１ｋｂ、少なくとも２ｋｂ、少なくとも５ｋｂ、少なくとも１０ｋｂ、少なくとも２０ｋｂ、少なくとも３０ｋｂ、または少なくとも４０ｋｂのフットプリントを有する。一部の実施形態では、エピジェネティック標的領域プローブセットは、０．５～１００ｋｂ、例えば、０．５～２ｋｂ、２～１０ｋｂ、１０～２０ｋｂ、２０～３０ｋｂ、３０～４０ｋｂ、４０～５０ｋｂ、５０～６０ｋｂ、６０～７０ｋｂ、７０～８０ｋｂ、８０～９０ｋｂ、および９０～１００ｋｂの範囲のフットプリントを有する。

【０４７５】

一部の実施形態では、配列可変標的領域セットに対して特異的なプローブは、表４の遺伝子のうちの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、または７０種（ｏｒ　ａｔ　７０）の少なくとも一部分に対して特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに対して特異的なプローブは、表３のＳＮＶのうちの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、または７０種（ｏｒ　７０）に対して特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに対して特異的なプローブは、表３の融合体のうちの少なくとも１種、少なくとも２種、少なくとも３種、少なくとも４種、少なくとも５種、または６種に対して特異的なプローブを含む。一部の実施形態では種、配列可変標的領域セットに対して特異的なプローブは、表４のインデルの少なくとも１種、少なくとも２種、または３種の少なくとも一部分に対して特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに対して特異的なプローブは、表５の遺伝子のうちの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、少なくとも７０種、または７３種の少なくとも一部分に対して特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに対して特異的なプローブは、表５のＳＮＶの少なくとも５種、少なくとも１０種、少なくとも１５種、少なくとも２０種、少なくとも２５種、少なくとも３０

種、少なくとも３５種、少なくとも４０種、少なくとも４５種、少なくとも５０種、少なくとも５５種、少なくとも６０種、少なくとも６５種、少なくとも７０種、または７３種に対して特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表５の融合体のうちの少なくとも１個、少なくとも２個、少なくとも３個、少なくとも４個、少なくとも５個、または６個に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表５のインデルのうちの少なくとも１個、少なくとも２個、少なくとも３個、少なくとも４個、少なくとも５個、少なくとも６個、少なくとも７個、少なくとも８個、少なくとも９個、少なくとも１０個、少なくとも１１個、少なくとも１２個、少なくとも１３個、少なくとも１４個、少なくとも１５個、少なくとも１６個、少なくとも１７個、または１８個の少なくとも一部分に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表６の遺伝子のうちの少なくとも１個、少なくとも２個、少なくとも３個、少なくとも４個、少なくとも５個、少なくとも６個、少なくとも７個、少なくとも８個、少なくとも９個、少なくとも１０個、少なくとも１１個、少なくとも１２個、少なくとも１３個、少なくとも１４個、少なくとも１５個、少なくとも１６個、少なくとも１７個、少なくとも１８個、少なくとも１９個、または少なくとも２０個の少なくとも一部分に特異的なプローブを含む。

【０４７６】

一部の実施形態では、配列可変標的領域セットに対して特異的なプローブは、少なくとも１０、２０、３０、または３５個のがん関連遺伝子、例えばＡＫＴ１、ＡＬＫ、ＢＲＡＦ、ＣＣＮＤ１、ＣＤＫ２Ａ、ＣＴＮＮＢ１、ＥＧＦＲ、ＥＲＢＢ２、ＥＳＲ１、ＦＧＦＲ１、ＦＧＦＲ２、ＦＧＦＲ３、ＦＯＸＬ２、ＧＡＴＡ３、ＧＮＡ１１、ＧＮＡＱ、ＧＮＡＳ、ＨＲＡＳ、ＩＤＨ１、ＩＤＨ２、ＫＩＴ、ＫＲＡＳ、ＭＥＤ１２、ＭＥＴ、ＭＹＣ、ＮＦＥ２Ｌ２、ＮＲＡＳ、ＰＤＧＦＲＡ、ＰＩＫ３ＣＡ、ＰＰＰ２Ｒ１Ａ、ＰＴＥＮ、ＲＥＴ、ＳＴＫ１１、ＴＰ５３、およびＵ２ＡＦ１からの標的領域に対して特異的なプローブを含む。

正確な処置

【０４７７】

改善されたコンピュータシステム１１０によって提供される正確な診断により、正確な処置計画をもたらすことができ、これは、コンピュータシステム１１０によって同定すること（および／または医療従事者がキュレートすること）ができる。例えば、１つのタイプの正確な診断および処置は、相同組換え修復（ＨＲＲ）経路内の遺伝子に関し得る。

【０４７８】

相同組換えは、２つの類似したまたは同一のＤＮＡ分子間でヌクレオチド配列が交換される、遺伝子組換えの１つのタイプである。相同組換えは、二本鎖切断（ＤＳＢ）として公知である、ＤＮＡの両方の鎖に生じる有害な切断を正確に修復するために細胞によって最も広く使用される。ＨＲＲは、細胞分裂が起こる前に染色体切断を排除するために、複製されたＤＮＡ（Ｓ期およびＧ２期）に存在する損傷に対するエラーのない除去のための機構を提供する。相同組換えによりＤＮＡの二本鎖切断がどのように修復されるかについての主要なモデルは、二本鎖切断修復（ＤＳＢＲ）経路および合成依存性鎖アニーリング（ＳＤＳＡ）経路を媒介する相同組換え修復経路である。相同組換え遺伝子の生殖系列および体細胞性の欠損は、乳がん、卵巣がんおよび前立腺がんに強く関連付けられている。

【０４７９】

試料中のバリアントヌクレオチドの数およびタイプにより、試料を提供する対象の、処置、すなわち治療介入に対する適応性の指標がもたらされ得る。例えば、様々なポリＡＤＰリボースポリメラーゼ（ＰＡＲＰ）阻害剤が、ＢＲＣＡ１またはＢＲＣＡ２遺伝子における遺伝性変異によって引き起こされる乳がん、卵巣がんおよび前立腺がんに由来する腫瘍の成長を停止させることが示されている。これらの治療剤のいくつかは、ＨＲＲの欠陥を補い得る塩基除去修復（ＢＥＲ）を阻害し得る。

【０４８０】

他方では、ある特定のＢＲＣＡおよびＨＲＲ野生型患者は、ＰＡＲＰ阻害剤による処置からの臨床的利益が実現されない可能性がある。さらに、ＢＲＣＡ変異を有する卵巣がん患者の全てがＰＡＲＰ阻害剤に応答するとは限らない。さらに、異なるタイプの変異は、異なる治療を示し得る。例えば、ＨＲＲ遺伝子における体細胞性ヘテロ接合性欠失は、体細胞性ホモ接合性欠失とは異なる治療を示し得る。したがって、遺伝物質の状態が治療に影響を及ぼし得る。一例では、ＰＡＲＰ阻害剤をＨＲＲ遺伝子に体細胞性ホモ接合性欠失を有する個体には投与することができるが、ＨＲＲ遺伝子に野生型アレルまたは体細胞性ヘテロ接合性欠失を有する個体には投与することができない。

【０４８１】

一部のインプリメンテーションでは、開示される方法のいずれかによって決定された、ＨＲＤを有する対象に、標的化治療を投与することができる。標的化治療は、ＰＡＲＰ阻害剤を含み得る。投与することができるＰＡＲＰ阻害剤の例としては、ベリパリブ、オラパリブ、タラゾパリブ、ルカパリブ、ニラパリブ、パミパリブ、ＣＥＰ　９７２２（Ｃｅｐｈａｌｏｎ）、Ｅ７０１６（Ｅｉｓａｉ）、Ｅ７４４９（Ｅｉｓａｉ、ＰＡＲＰ　１／２およびタンキラーゼ１／２阻害剤）、または３－アミノベンズアミドのうちの１つまたは複数が挙げられる。一部のインプリメンテーションでは、標的化治療は、少なくとも１つの塩基除去修復（ＢＥＲ）阻害剤を含み得る。例えば、オラパリブは、ＢＥＲを阻害し得る。ある特定のインプリメンテーションでは、標的化治療は、ＰＡＲＰ阻害剤と放射線療法の組合せを含み得る。あるインプリメンテーションでは、ＰＡＲＰ阻害剤と放射線療法の組合せは、ＰＡＲＰ阻害剤が、腫瘍組織（例えば、ＢＲＣＡ１／ＢＲＣＡ２変異を有する組織）において放射線療法によって生じた一本鎖切断から二本鎖切断の形成をもたらすことを可能にし得る。この組合せにより、放射線線量ごとにより強力な治療をもたらすことができる。

カスタマイズされた治療および関連する投与

【０４８２】

一部のインプリメンテーションでは、本明細書に開示される方法は、所与の疾患、障害または状態を有する患者に対して、治療を同定し、投与することに関する。本質的に任意のがん治療（例えば、外科療法、放射線療法、化学療法など）がこれらの方法の一部分として含められる。ある特定のインプリメンテーションでは、対象に投与される治療は、少なくとも１つの化学療法薬を含み得る。一部のインプリメンテーションでは、化学療法薬は、アルキル化剤（例えば、これだけに限定されないが、クロラムブシル、シクロホスファミド、シスプラチンおよびカルボプラチン）、ニトロソ尿素（例えば、これだけに限定されないが、カルムスチンおよびロムスチン）、代謝拮抗薬（例えば、これだけに限定されないが、フルオロウラシル（Ｆｌｕｏｒａｕｒａｃｉｌ）、メトトレキサートおよびフルダラビン）、植物アルカロイドおよび天然物（例えば、これだけに限定されないが、ビンクリスチン、パクリタキセルおよびトポテカン）、抗腫瘍抗生物質（例えば、これだけに限定されないが、ブレオマイシン、ドキソルビシンおよびミトキサントロン）、ホルモン剤（例えば、これだけに限定されないが、プレドニゾン、デキサメタゾン、タモキシフェンおよびロイプロリド）ならびに生物学的応答改変剤（例えば、これだけに限定されないが、ハーセプチンおよびアバスチン、アービタックスおよびリツキサン）を含み得る。一部のインプリメンテーションでは、対象に投与される化学療法は、ＦＯＬＦＯＸまたはＦＯＬＦＩＲＩを含み得る。典型的には、治療は、少なくとも１つの免疫療法（または免疫療法剤）を含む。免疫療法とは、一般に、所与のがんタイプに対する免疫応答を増強する方法を指す。ある特定のインプリメンテーションでは、免疫療法とは、腫瘍またはがんに対するＴ細胞応答を増強する方法を指す。

【０４８３】

一部のインプリメンテーションでは、免疫療法または免疫療法剤は、免疫チェックポイント分子を標的とする。ある特定の腫瘍は、免疫チェックポイント経路を利用することによって免疫系から逃れることができる。したがって、腫瘍の免疫系から逃れる能力を阻止するため、およびある特定のがんに対する抗腫瘍免疫を活性化するための有効なアプロー

チとして免疫チェックポイントを標的とすることが出現した。Pardoll, Nature Reviews Cancer, 2012, 12: 252-264。

【０４８４】

ある特定のインプリメンテーションでは、免疫チェックポイント分子は、抗原に対するＴ細胞応答に関与するシグナルを低減させる阻害性分子である。例えば、ＣＴＬＡ４は、Ｔ細胞上に発現され、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細胞活性化を下方調節することにおいて役割を果たす。ＰＤ－１は、Ｔ細胞上に発現される別の阻害性チェックポイント分子である。ＰＤ－１は、炎症応答中、末梢組織におけるＴ細胞の活性を限定する。さらに、ＰＤ－１のリガンド（ＰＤ－Ｌ１またはＰＤ－Ｌ２）は、一般に、多くの異なる腫瘍の表面上で上方調節されており、その結果、腫瘍微小環境における抗腫瘍免疫応答の下方調節がもたらされる。ある特定のインプリメンテーションでは、阻害性免疫チェックポイント分子は、ＣＴＬＡ４またはＰＤ－１である。他のインプリメンテーションでは、阻害性免疫チェックポイント分子は、ＰＤ－１のリガンド、例えば、ＰＤ－Ｌ１またはＰＤ－Ｌ２である。他のインプリメンテーションでは、阻害性免疫チェックポイント分子は、ＣＴＬＡ４のリガンド、例えば、ＣＤ８０またはＣＤ８６である。他のインプリメンテーションでは、阻害性免疫チェックポイント分子は、リンパ球活性化遺伝子３（ＬＡＧ３）、キラー細胞免疫グロブリン様受容体（ＫＩＲ）、Ｔ細胞膜タンパク質３（ＴＩＭ３）、ガレクチン９（ＧＡＬ９）、またはアデノシンＡ２ａ受容体（Ａ２ａＲ）である。

【０４８５】

これらの免疫チェックポイント分子を標的とするアンタゴニストを使用して、ある特定のがんに対する抗原特異的Ｔ細胞応答を増強することができる。したがって、ある特定のインプリメンテーションでは、免疫療法または免疫療法剤は、阻害性免疫チェックポイント分子のアンタゴニストである。ある特定のインプリメンテーションでは、阻害性免疫チェックポイント分子は、ＰＤ－１である。ある特定のインプリメンテーションでは、阻害性免疫チェックポイント分子は、ＰＤ－Ｌ１である。ある特定のインプリメンテーションでは、阻害性免疫チェックポイント分子のアンタゴニストは、抗体（例えば、モノクローナル抗体）である。ある特定のインプリメンテーションでは、抗体またはモノクローナル抗体は、抗ＣＴＬＡ４、抗ＰＤ－１、抗ＰＤ－Ｌ１、または抗ＰＤ－Ｌ２抗体である。ある特定のインプリメンテーションでは、抗体は、モノクローナル抗ＰＤ－１抗体である。一部のインプリメンテーションでは、抗体は、モノクローナル抗ＰＤ－Ｌ１抗体である。ある特定のインプリメンテーションでは、モノクローナル抗体は、抗ＣＴＬＡ４抗体と抗ＰＤ－１抗体、抗ＣＴＬＡ４抗体と抗ＰＤ－Ｌ１抗体、または抗ＰＤ－Ｌ１抗体と抗ＰＤ－１抗体の組合せである。ある特定のインプリメンテーションでは、抗ＰＤ－１抗体は、ペムブロリズマブ（Ｋｅｙｔｒｕｄａ（登録商標））またはニボルマブ（Ｏｐｄｉｖｏ（登録商標））のうちの１つまたは複数である。ある特定のインプリメンテーションでは、抗ＣＴＬＡ４抗体は、イピリムマブ（Ｙｅｒｖｏｙ（登録商標））である。ある特定のインプリメンテーションでは、抗ＰＤ－Ｌ１抗体は、アテゾリズマブ（Ｔｅｃｅｎｔｒｉｑ（登録商標））、アベルマブ（Ｂａｖｅｎｃｉｏ（登録商標））、またはデュルバルマブ（Ｉｍｆｉｎｚｉ（登録商標））のうちの１つまたは複数である。

【０４８６】

ある特定のインプリメンテーションでは、免疫療法または免疫療法剤は、ＣＤ８０、ＣＤ８６、ＬＡＧ３、ＫＩＲ、ＴＩＭ３、ＧＡＬ９、またはＡ２ａＲに対するアンタゴニスト（例えば、抗体）である。他のインプリメンテーションでは、アンタゴニストは、可溶性型の阻害性免疫チェックポイント分子、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と抗体のＦｃドメインとを含む可溶性融合体タンパク質である。ある特定のインプリメンテーションでは、可溶性融合体タンパク質は、ＣＴＬＡ４、ＰＤ－１、ＰＤ－Ｌ１、またはＰＤ－Ｌ２の細胞外ドメインを含む。一部のインプリメンテーションでは、可溶性融合体タンパク質は、ＣＤ８０、ＣＤ８６、ＬＡＧ３、ＫＩＲ、ＴＩＭ３、ＧＡＬ９、またはＡ２ａＲの細胞外ドメインを含む。１つのインプリメンテーションでは、可

溶性融合体タンパク質は、ＰＤ－Ｌ２またはＬＡＧ３の細胞外ドメインを含む。

【０４８７】

ある特定のインプリメンテーションでは、免疫チェックポイント分子は、抗原に対するＴ細胞応答に関与するシグナルを増幅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受容体である。Ｔ細胞がそのＴ細胞受容体を通じて抗原に結合すると、ＣＤ２８が、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細胞受容体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により、ＣＤ２８によって媒介される共刺激シグナル伝達を打ち消すまたは調節することができる。ある特定のインプリメンテーションでは、免疫チェックポイント分子は、ＣＤ２８、誘導性Ｔ細胞共刺激物質（ＩＣＯＳ）、ＣＤ１３７、ＯＸ４０、またはＣＤ２７から選択される共刺激分子である。他のインプリメンテーションでは、免疫チェックポイント分子は、例えば、ＣＤ８０、ＣＤ８６、Ｂ７ＲＰ１、Ｂ７－Ｈ３、Ｂ７－Ｈ４、ＣＤ１３７Ｌ、ＯＸ４０Ｌ、またはＣＤ７０を含む、共刺激分子のリガンドである。

【０４８８】

これらの共刺激チェックポイント分子を標的とするアゴニストを使用して、ある特定のがんに対する抗原特異的Ｔ細胞応答を増強することができる。したがって、ある特定のインプリメンテーションでは、免疫療法または免疫療法剤は、共刺激チェックポイント分子のアゴニストである。ある特定のインプリメンテーションでは、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。ある特定のインプリメンテーションでは、アゴニスト抗体またはモノクローナル抗体は、抗ＣＤ２８抗体である。他のインプリメンテーションでは、アゴニスト抗体またはモノクローナル抗体は、抗ＩＣＯＳ、抗ＣＤ１３７、抗ＯＸ４０、または抗ＣＤ２７抗体である。他のインプリメンテーションでは、アゴニスト抗体またはモノクローナル抗体は、抗ＣＤ８０、抗ＣＤ８６、抗Ｂ７ＲＰ１、抗Ｂ７－Ｈ３、抗Ｂ７－Ｈ４、抗ＣＤ１３７Ｌ、抗ＯＸ４０Ｌ、または抗ＣＤ７０抗体である。

【０４８９】

がん以外の、特定の遺伝子に基づく疾患、障害、または状態を処置するための治療選択肢は、一般に、当業者には周知であり、検討されている特定の疾患、障害、または状態を考慮すれば明らかである。

【０４９０】

ある特定のインプリメンテーションでは、本明細書に記載されるカスタマイズされた治療は、典型的に非経口（例えば、静脈内または皮下）投与される。免疫療法剤を含有する医薬組成物は、典型的に静脈内投与される。ある特定の治療剤は経口投与される。しかし、カスタマイズされた治療（例えば、免疫療法剤など）はまた、例えば頬側、舌下、直腸、膣、尿道内、局所、眼内、鼻腔内、および／または耳介内を含む、当技術分野において公知の任意の方法によって投与されてもよく、投与は、錠剤、カプセル剤、顆粒剤、水性懸濁剤、ゲル剤、スプレー剤、坐剤、塗擦剤（ｓａｌｖｅ）、軟膏剤などを含み得る。

【０４９１】

図５は、インプリメンテーションの一部の例による、機械可読媒体（例えば、機械可読ストレージ媒体）からの命令を読み取り、本明細書で考察される方法論のうちのいずれか１つまたは複数を実施することができる機械５００のコンポーネントを例示するブロック図である。具体的には、図５は、機械５００に本明細書で考察される方法論のうちのいずれか１つまたは複数を実施させるための命令５０２（例えば、ソフトウェア、プログラム、アプリケーション、アプレット、ａｐｐ、または他の実行可能なコード）を実行することができる、コンピュータシステムの機械５００の形態の例の図式表示を示す。したがって、命令５０２を使用して、本明細書に記載されるモジュールまたはコンポーネントをインプリメントすることができる。命令５０２により、一般的な、プログラムされていない機械５００が、記載および例示される機能を記載の様式で実施するようにプログラムされた特定の機械５００に変換される。代替のインプリメンテーションでは、機械５００は、

独立型デバイスとして動作する、または他の機械にカップリングされ（例えば、ネットワークで結ばれ）てもよい。ネットワーク配置（networked deployment）では、機械５００は、サーバー－クライアントネットワーク環境におけるサーバーマシンもしくはクライアントマシンとして、または、ピアツーピア（または分散）ネットワーク環境におけるピアマシンとして、動作し得る。機械５００は、サーバーコンピュータ、クライアントコンピュータ、パーソナルコンピュータ（ＰＣ）、タブレットコンピュータ、ラップトップコンピュータ、ネットブック、セットトップボックス（ＳＴＢ）、パーソナルデジタルアシスタント（ＰＤＡ）、エンターテイメントメディアシステム、携帯電話、スマートフォン、モバイルデバイス、ウェアラブルデバイス（例えば、スマートウォッチ）、スマートホームデバイス（例えば、スマート家電）、他のスマートデバイス、ウェブアプライアンス、ネットワークルーター、ネットワークスイッチ、ネットワークブリッジ、または、機械５００により行われる動作を指定する、命令５０２を逐次的にもしくはそれ以外で実行することが可能な任意の機械を含み得るが、これらに限定されない。さらに、単一の機械５００のみが例示されているが、「機械」という用語は、命令５０２を個々にまたは共同で実行して、本明細書で考察される方法論のうちのいずれか１つまたは複数を実施する機械のコレクションも含むとみなされるものとする。

【０４９２】

機械５００は、プロセッサー５０４、メモリー／ストレージ５０６、およびＩ／Ｏコンポーネント５０８を含み得、これらは、例えばバス５１０を介して互いに通信するように構成されていてよい。インプリメンテーションのある例では、プロセッサー５０４（例えば、中央処理装置（ＣＰＵ）、縮小命令セットコンピューティング（ＲＩＳＣ）プロセッサー、複合命令セットコンピューティング（ＣＩＳＣ）プロセッサー、グラフィックスプロセシングユニット（ＧＰＵ）、デジタルシグナルプロセッサー（ＤＳＰ）、特定用途向け集積回路（ＡＳＩＣ）、無線周波数集積回路（ＲＦＩＣ）、別のプロセッサー、またはこれらの任意の好適な組合せ）は、例えば、命令５０２を実行することができるプロセッサー５１２およびプロセッサー５１４を含み得る。「プロセッサー」という用語は、命令５０２を同時に実行することができる２つまたはそれよりも多くの独立したプロセッサー（時には「コア」と称される）を含み得るマルチコアプロセッサー５０４を含むことが意図される。図５には複数のプロセッサー５０４が示されているが、機械５００は、単一のコアを有する単一のプロセッサー５１２、複数のコアを有する単一のプロセッサー５１２（例えば、マルチコアプロセッサー）、単一のコアを有する複数のプロセッサー５１２、５１４、複数のコアを有する複数のプロセッサー５１２、５１４、またはこれらの任意の組合せを含んでもよい。

【０４９３】

メモリー／ストレージ５０６は、メインメモリー５１６または他のメモリーストレージなどのメモリー、およびストレージユニット５１８を含んでよく、これらはどちらも例えばバス５１０を介してプロセッサー５０４にアクセス可能である。ストレージユニット５１８およびメインメモリー５１６は、本明細書に記載される方法論または機能のうちのいずれか１つまたは複数を具体化する命令５０２を記憶する。命令５０２は、機械５００によるその実行中、完全にまたは部分的に、メインメモリー５１６内、ストレージユニット５１８内、プロセッサー５０４のうちの少なくとも１つの内（例えば、プロセッサーのキャッシュメモリー内）、またはこれらの任意の好適な組合せに存在してもよい。したがって、メインメモリー５１６、ストレージユニット５１８、およびプロセッサー５０４のメモリーは、機械可読媒体の例である。

【０４９４】

Ｉ／Ｏコンポーネント５０８コンポーネント５０８は、入力を受け取るため、出力を提供する、出力を生じさせる、情報を送信する、情報を交換する、測定値を捕捉するなどのための多種多様なコンポーネントを含み得る。特定の機械５００に含められる特定のＩ／Ｏコンポーネント５０８は、機械のタイプに依存する。例えば、携帯電話などの携帯型機械は、タッチ入力デバイスまたは他のそのような入力機構を含む可能性があり、一方で、

ヘッドレスサーバーマシンは、そのようなタッチ入力デバイスを含まない可能性がある。Ｉ／Ｏコンポーネント５０８コンポーネント５０８は、図５に示されていない多くの他のコンポーネントを含み得ることが理解されよう。Ｉ／Ｏコンポーネント５０８は、ただ単に以下の考察を単純化するために、機能に応じて群分けされており、群分けは決して限定的なものではない。インプリメンテーションの様々な例では、Ｉ／Ｏコンポーネント５０８コンポーネント５０８は、ユーザー出力コンポーネント５２０およびユーザー入力コンポーネント５２２を含み得る。ユーザー出力コンポーネント５２０は、視覚的コンポーネント（例えば、プラズマディスプレイパネル（ＰＤＰ）、発光ダイオード（ＬＥＤ）ディスプレイ、液晶ディスプレイ（ＬＣＤ）、プロジェクター、またはブラウン管（ＣＲＴ）などのディスプレイ）、聴覚的コンポーネント（例えば、スピーカー）、触覚的コンポーネント（例えば、振動モーター、抵抗機構）、他のシグナル発生装置などを含み得る。ユーザー入力コンポーネント５２２は、英数字入力コンポーネント（例えば、キーボード、英数字入力を受け取るように構成されたタッチスクリーン、写真－光学キーボード、または他の英数字入力コンポーネント）、ポイントに基づく入力コンポーネント（例えば、マウス、タッチパッド、トラックボール、ジョイスティック、動作センサー、または他のポインティング機器）、触知的入力コンポーネント（例えば、物理ボタン、タッチもしくはタッチジェスチャーの場所もしくは力を提供するタッチスクリーン、または他の触知的入力コンポーネント）、オーディオ入力コンポーネント（例えば、マイクロホン）などを含み得る。

【０４９５】

インプリメンテーションのさらなる例では、Ｉ／Ｏコンポーネント５０８コンポーネント５０８は、幅広い他のコンポーネントの中でも、バイオメトリクスコンポーネント５２４、動作コンポーネント５２６、環境コンポーネント５２８、または位置コンポーネント５３０を含み得る。例えば、バイオメトリクスコンポーネント５２４は、表現（例えば、手による表現、顔による表現、声による表現、身ぶり、または視線追跡）を検出する、生体シグナル（例えば、血圧、心拍数、体温、発汗、または脳波）を測定する、人を識別する（例えば、音声識別、網膜識別、顔識別、指紋識別、または脳波に基づく識別）などのためのコンポーネントを含み得る。動作コンポーネント５２６は、加速センサーコンポーネント（例えば、加速度計）、重力センサーコンポーネント、回転センサーコンポーネント（例えば、ジャイロスコープ）などを含み得る。環境コンポーネント５２８は、例えば、照明センサーコンポーネント（例えば、光度計）、温度センサーコンポーネント（例えば、外界温度を検出する１つまたは複数の温度計）、湿度センサーコンポーネント、圧力センサーコンポーネント（例えば、気圧計）、音響センサーコンポーネント（例えば、バックグラウンドノイズを検出する１つまたは複数のマイクロホン）、近接センサーコンポーネント（例えば、近くの物体を検出する赤外センサー）、ガスセンサー（例えば、安全性のために有害ガスの濃度を検出するため、もしくは大気中の汚染物質を測定するためのガス検知センサー）、または周囲の物理的環境に対応する指標、測定値、もしくはシグナルを提供し得る他のコンポーネントを含み得る。位置コンポーネント５３０は、位置センサーコンポーネント（例えば、ＧＰＳレシーバーコンポーネント）、高度センサーコンポーネント（例えば、高度計または高度を算出することができる気圧を検出する気圧計）、方位センサーコンポーネント（例えば、磁気計）などを含み得る。

【０４９６】

通信は、多種多様なテクノロジーを使用してインプリメントすることができる。Ｉ／Ｏコンポーネント５０８は、機械５００をネットワーク５３４またはデバイス５３６にカップリングするために動作可能な通信コンポーネント５３２を含み得る。例えば、通信コンポーネント５３２は、ネットワーク５３４とインターフェース接続するためのネットワークインターフェースコンポーネントまたは他の好適なデバイスを含み得る。さらなる例では、通信コンポーネント５３２は、有線通信コンポーネント、無線通信コンポーネント、セルラー方式の通信コンポーネント、近距離無線通信（ＮＦＣ）コンポーネント、Ｂｌｕｅｔｏｏｔｈ（登録商標）コンポーネント（例えば、Ｂｌｕｅｔｏｏｔｈ（登録商標）Ｌ

ｏｗ　Ｅｎｅｒｇｙ）、Ｗｉ－Ｆｉ（登録商標）コンポーネント、および他のモダリティを介した通信を提供する他の通信コンポーネントを含み得る。デバイス５３６は、別の機械５００または多種多様な周辺デバイスのいずれか（例えば、ＵＳＢを介してカップリングした周辺デバイス）であってよい。

【０４９７】

さらに、通信コンポーネント５３２は、識別子を検出するかまたは識別子を検出するように動作可能なコンポーネントを含み得る。例えば、通信コンポーネント５３２は、無線周波数識別（ＲＦＩＤ）タグリーダーコンポーネント、ＮＦＣスマートタグ検出コンポーネント、光学リーダーコンポーネント（例えば、Ｕｎｉｖｅｒｓａｌ　Ｐｒｏｄｕｃｔ　Ｃｏｄｅ（ＵＰＣ）バーコードなどの一次元バーコード、Ｑｕｉｃｋ　Ｒｅｓｐｏｎｓｅ（ＱＲ）コードなどの多次元バーコード、Ａｚｔｅｃコード、Ｄａｔａ　Ｍａｔｒｉｘ、Ｄａｔａｇｌｙｐｈ、ＭａｘｉＣｏｄｅ、ＰＤＦ４１７、Ｕｌｔｒａ　Ｃｏｄｅ、ＵＣＣ　ＲＳＳ－２Ｄバーコード、および他の光学的なコードを検出するための光学センサー）、または音響検出コンポーネント（例えば、標識音声信号（tagged audio signal）を同定するためのマイクロホン）を含み得る。さらに、例えば、Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ（ＩＰ）ジオロケーションを介した場所、Ｗｉ－Ｆｉ（登録商標）シグナル三角測量を介した場所、特定の場所を示し得るＮＦＣビーコンシグナルの検出を介した場所など、通信コンポーネント５３２を介して様々な情報を得ることができる。

【０４９８】

本明細書で使用される場合、「コンポーネント」とは、関数もしくはサブルーチンコール、分岐点、ＡＰＩ、または特定の処理もしくは制御機能の分割もしくはモジュール化をもたらす他のテクノロジーによって規定される境界を有するデバイス、物理的実体、またはロジックを指す。コンポーネントをそれらのインターフェースを介して他のコンポーネントと組み合わせて、機械処理を行うことができる。コンポーネントは、他のコンポーネントおよび関連する機能の特定の機能を通常実施するプログラムの一部とともに使用するために設計されたパッケージングされた機能性ハードウェアユニットであってもよい。コンポーネントは、ソフトウェアコンポーネント（例えば、機械可読媒体上に具体化されるコード）またはハードウェアコンポーネントのいずれかを構成してもよい。「ハードウェアコンポーネント」は、ある特定の操作を実施することができ、ある特定の物理的様式で構成または配置することができる有形単位である。インプリメンテーションの様々な例では、１つまたは複数のコンピュータシステム（例えば、独立型コンピュータシステム、クライアントコンピュータシステム、もしくはサーバーコンピュータシステム）またはコンピュータシステムの１つもしくは複数のハードウェアコンポーネント（例えば、プロセッサーもしくはプロセッサーの群）は、ソフトウェア（例えば、アプリケーションまたはアプリケーション部分）によって、本明細書に記載されるある特定の操作を実施するように動作するハードウェアコンポーネントとして構成されていてよい。

【０４９９】

ハードウェアコンポーネントは、機械的に、電子的に、またはこれらの任意の好適な組合せでインプリメントされてもよい。例えば、ハードウェアコンポーネントは、ある特定の操作を実施するように恒久的に構成された専用回路網またはロジックを含み得る。ハードウェアコンポーネントは、フィールドプログラマブルゲートアレイ（ＦＰＧＡ）またはＡＳＩＣなどの特殊用途プロセッサーであってもよい。ハードウェアコンポーネントは、ある特定の操作を実施するためにソフトウェアによって一時的に構成されるプログラム可能なロジックまたは回路網も含んでもよい。例えば、ハードウェアコンポーネントは、汎用プロセッサー５０４または他のプログラム可能なプロセッサーによって実行されるソフトウェアを含み得る。そのようなソフトウェアによって構成されたら、ハードウェアコンポーネントは、構成された機能を実施するように独特に調整された特定の機械（または機械５００の特定のコンポーネント）になり、もはや汎用プロセッサー５０４ではなくなる。ハードウェアコンポーネントを、専用かつ恒久的に構成された回路網において、または一時的に構成された回路網（例えば、ソフトウェアによって構成される）において機械的

にインプリメントする決定は、費用および時間の検討によって推進されてもよいことが理解されよう。したがって、「ハードウェアコンポーネント」（または「ハードウェアによりインプリメントされるコンポーネント」）という句は、ある特定の様式で動作するように、または本明細書に記載されるある特定の操作を実施するように、物理的に構築された、恒久的に構成された（例えば、ハードウェアに組み込まれた）、または一時的に構成された（例えば、プログラムされた）実体である有形実体を包含するものと理解されるべきである。ハードウェアコンポーネントが一時的に構成された（例えば、プログラムされた）インプリメンテーションを考えると、ハードウェアコンポーネントのそれぞれが、いかなる時にも構成されているまたはインスタンス化されている必要はない。例えば、ハードウェアコンポーネントが、ソフトウェアによって特殊用途プロセッサーになるように構成された汎用プロセッサー５０４を含む場合、汎用プロセッサー５０４は、異なる時間において、それぞれ異なる特殊用途プロセッサー（例えば、異なるハードウェアコンポーネントを含む）として構成されてもよい。ソフトウェアは、それに応じて、例えば、１つの時点で特定のハードウェアコンポーネントを構築するように、および異なる時点では異なるハードウェアコンポーネントを構築するように、特定のプロセッサー５１２プロセッサー５１２、５１４またはプロセッサー５０４を構成する。

【０５００】

ハードウェアコンポーネントは、他のハードウェアコンポーネントに情報を提供することおよび他のハードウェアコンポーネントから情報を受け取ることができる。したがって、記載のハードウェアコンポーネントは、通信可能にカップリングしているとみなすことができる。複数のハードウェアコンポーネントが同時に存在する場合、通信は、ハードウェアコンポーネントの２つまたはそれよりも多くの間での信号送信（例えば、適切な回路およびバスを通じて）によって実現することができる。複数のハードウェアコンポーネントが異なる時間に構成またはインスタンス化されるインプリメンテーションでは、そのようなハードウェアコンポーネント間の通信は、例えば、複数のハードウェアコンポーネントがアクセスできるメモリー構造における情報のストレージおよび検索によって実現することができる。例えば、１つのハードウェアコンポーネントにより操作を実施し、その操作の出力を、通信可能にカップリングしたメモリーデバイスに記憶することができる。次いで、後の時点で、さらなるハードウェアコンポーネントがメモリーデバイスにアクセスして、記憶された出力を検索し、処理する。

【０５０１】

ハードウェアコンポーネントはまた、入力または出力デバイスとの通信を開始してもよく、リソース（例えば、情報のコレクション）に対して動作することができる。本明細書に記載される方法の例の様々な操作を、少なくとも部分的に、関連する操作を実施するように一時的に構成された（例えば、ソフトウェアによって）または恒久的に構成された１つまたは複数のプロセッサー５０４によって実施することができる。一時的に構成されるかまたは恒久的に構成されるかにかかわらず、そのようなプロセッサー５０４は、本明細書に記載される１つまたは複数の操作または機能を実施するように動作するプロセッサーによりインプリメントされるコンポーネントを構築し得る。本明細書で使用される場合、「プロセッサーによりインプリメントされるコンポーネント」とは、１つまたは複数のプロセッサー５０４を使用してインプリメントされるハードウェアコンポーネントを指す。同様に、本明細書に記載される方法は、少なくとも部分的にプロセッサーによりインプリメントされてもよく、特定のプロセッサー５１２プロセッサー５１２、５１４またはプロセッサー５０４がハードウェアの例である。例えば、方法の操作の少なくとも一部を、１つまたは複数のプロセッサー５０４またはプロセッサーによりインプリメントされるコンポーネントによって実施することができる。さらに、１つまたは複数のプロセッサー５０４は、「クラウドコンピューティング」環境でまたは「サービスとしてのソフトウェア」（ＳａａＳ）として、関連する操作の実行を支持するように動作してもよい。例えば、操作の少なくとも一部をコンピュータの群（プロセッサー５０４を含む、機械１０００の例として）によって実施することができ、これらの操作は、ネットワーク５３４（例えば、

Ｉｎｔｅｒｎｅｔ）を介して、および１つまたは複数の適切なインターフェース（例えば、ＡＰＩ）を介してアクセス可能である。操作のある特定の実行をプロセッサー間に分散させ、単一の機械５００内に存在させるだけではなく、いくつかの機械にわたって配置することができる。インプリメンテーションの一部の例では、プロセッサー５０４またはプロセッサーによりインプリメントされるコンポーネントは、単一の地理的場所（例えば、家庭環境、オフィス環境、またはサーバーファーム内）に配置されていてよい。インプリメンテーションの他の例では、プロセッサー５０４またはプロセッサーによりインプリメントされるコンポーネントはいくつかの地理的場所にわたって分散していてもよい。

【０５０２】

図６は、本明細書に記載される様々なハードウェアアーキテクチャと併せて使用することができるソフトウェアアーキテクチャ６０２の例を含むシステム６００を例示するブロック図である。図６は、ソフトウェアアーキテクチャの非限定的な例であり、本明細書に記載される機能を容易にするために、多くの他のアーキテクチャをインプリメントすることができることが理解されよう。ソフトウェアアーキテクチャ６０２は、ハードウェア、例えば、とりわけ、プロセッサー５０４、メモリー／ストレージ５０６、および入力／出力（Ｉ／Ｏ）コンポーネント５０８を含む図５の機械５００上で実行し得る。代表的なハードウェアレイヤー６０４が例示されており、これは、例えば、図５の機械５００を表し得る。代表的なハードウェアレイヤー６０４は、実行可能命令６０８が付随するプロセシングユニット６０６を含む。実行可能命令６０８は、本明細書に記載される方法のインプリメンテーション、コンポーネントなどを含む、ソフトウェアアーキテクチャ６０２の実行可能命令を表す。ハードウェアレイヤー６０４はまた、同様に実行可能命令６０８を有するメモリーまたはストレージモジュールメモリー／ストレージ６１０のうちの少なくとも１つを含む。ハードウェアレイヤー６０４は、他のハードウェア６１２も含んでもよい。

【０５０３】

図６のアーキテクチャの例では、ソフトウェアアーキテクチャ６０２は、各レイヤーにより特定の機能が提供されるレイヤーのスタックとして概念化することができる。例えば、ソフトウェアアーキテクチャ６０２は、オペレーティングシステム６１４、ライブラリー６１６、フレームワーク／ミドルウェア６１８、アプリケーション６２０、およびプレゼンテーションレイヤー６２２などのレイヤーを含み得る。操作上、アプリケーション６２０またはレイヤー内の他のコンポーネントは、ソフトウェアスタックを通じてＡＰＩコール６２４を呼び出すこと、およびＡＰＩコール６２４に応答してメッセージ６２６を受け取ることができる。例示されているレイヤーは、実際の代表的なものであり、全てのソフトウェアアーキテクチャが全てのレイヤーを有するとは限らない。例えば、一部のモバイルまたは特殊用途オペレーティングシステムはフレームワーク／ミドルウェア６１８を提供しない場合があるが、一方で、他のものはそのようなレイヤーを提供する場合がある。他のソフトウェアアーキテクチャは、追加のまたは異なるレイヤーを含み得る。

【０５０４】

オペレーティングシステム６１４は、ハードウェアリソースを管理し、共通サービスを提供することができる。オペレーティングシステム６１４は、例えば、カーネル６２８、サービス６３０、およびドライバー６３２を含み得る。カーネル６２８は、ハードウェアと他のソフトウェアレイヤーの間の抽象化レイヤーとしての機能を果たし得る。例えば、カーネル６２８は、メモリー管理、プロセッサー管理（例えば、スケジューリング）、コンポーネント管理、ネットワーキング、セキュリティ設定などを担い得る。サービス６３０は、他のソフトウェアレイヤーに対して他の共通サービスを提供し得る。ドライバー６３２は、基礎をなすハードウェアの制御またはそれとのインターフェース接続を担う。例えば、ドライバー６３２は、ハードウェア構成に応じて、ディスプレイドライバー、カメラドライバー、Ｂｌｕｅｔｏｏｔｈ（登録商標）ドライバー、フラッシュメモリードライバー、シリアル通信ドライバー（例えば、Ｕｎｉｖｅｒｓａｌ Ｓｅｒｉａｌ Ｂｕｓ（ＵＳＢ）ドライバー）、Ｗｉ－Ｆｉ（登録商標）ドライバー、オーディオドライバー、電

源管理ドライバーなどを含む。

【０５０５】

ライブラリー６１６は、アプリケーション６２０、他のコンポーネント、またはレイヤーのうちの少なくとも１つによって使用される共通基盤を提供する。ライブラリー６１６は、他のソフトウェアコンポーネントが、基礎をなすオペレーティングシステム６１４機能（例えば、カーネル６２８、サービス６３０、ドライバー６３２）と直接インターフェース接続した場合よりも容易にタスクを実施することを可能にする機能を提供する。ライブラリー６１６は、例えば、メモリー割り当て関数、文字列操作関数、数学関数などの関数を提供し得るシステムライブラリー６３４（例えば、標準Ｃライブラリー）を含み得る。さらに、ライブラリー６１６は、ＡＰＩライブラリー６３６、例えば、メディアライブラリー（例えば、ＭＰＥＧ４、Ｈ．２６４、ＭＰ３、ＡＡＣ、ＡＭＲ、ＪＰＧ、ＰＮＧなどの様々なメディアフォーマットのプレゼンテーションおよび操作を支持するためのライブラリー）、グラフィックスライブラリー（例えば、ディスプレイ上のグラフィックコンテンツに２次元および３次元をレンダリングするために使用することができるＯｐｅｎＧＬフレームワーク）、データベースライブラリー（例えば、様々なリレーショナルデータベース機能を提供し得るＳＱＬｉｔｅ）、ウェブライブラリー（例えば、ウェブブラウジング機能を提供し得るＷｅｂＫｉｔ）などを含み得る。ライブラリー６１６はまた、アプリケーション６２０および他のソフトウェアコンポーネント／モジュールに多くの他のＡＰＩを提供するために多種多様な他のライブラリー６３８を含んでもよい。

【０５０６】

フレームワーク／ミドルウェア６１８（時にはミドルウェアとも称される）は、アプリケーション６２０または他のソフトウェアコンポーネント／モジュールによって使用され得るより高レベルの共通基盤を提供する。例えば、フレームワーク／ミドルウェア６１８は、様々なグラフィカルユーザーインターフェース機能、高レベルのリソース管理、高レベルの位置情報サービスなどを提供し得る。フレームワーク／ミドルウェア６１８は、アプリケーション６２０または他のソフトウェアコンポーネント／モジュールによって利用され得る幅広い他のＡＰＩを提供し得、その一部は、特定のオペレーティングシステム６１４またはプラットフォームに特異的なものであってもよい。

【０５０７】

アプリケーション６２０は、内蔵アプリケーション６４０およびサードパーティーアプリケーション６４２を含む。代表的な内蔵アプリケーション６４０の例としては、連絡先アプリケーション、ブラウザアプリケーション、ブックリーダーアプリケーション、位置情報アプリケーション、メディアアプリケーション、メッセージングアプリケーション、またはゲームアプリケーションを挙げることができるが、これらに限定されない。サードパーティーアプリケーション６４２は、特定のプラットフォームのベンダー以外の実体によってＡＮＤＲＯＩＤ（登録商標）またはＩＯＳ（商標）ソフトウェア開発キット（ＳＤＫ）を使用して開発されたアプリケーションを含んでもよく、ＩＯＳ（商標）、ＡＮＤＲＯＩＤ（登録商標）、ＷＩＮＤＯＷＳ（登録商標）Ｐｈｏｎｅ、または他のモバイルオペレーティングシステムなどのモバイルオペレーティングシステムで実行されるモバイルソフトウェアであってよい。サードパーティーアプリケーション６４２は、モバイルオペレーティングシステム（例えば、オペレーティングシステム６１４）によって提供されるＡＰＩコール６２４を呼び出して、本明細書に記載される機能を容易にすることができる。

【０５０８】

アプリケーション６２０は、システムのユーザーとの対話のために、内蔵オペレーティングシステム機能（例えば、カーネル６２８、サービス６３０、ドライバー６３２）、ライブラリー６１６、およびフレームワーク／ミドルウェア６１８を使用して、ＵＩを作成し得る。その代わりにまたはそれに加えて、一部のシステムでは、ユーザーとの対話は、プレゼンテーションレイヤー６２２などのプレゼンテーションレイヤーによって行うことができる。これらのシステムでは、アプリケーション／コンポーネント「ロジック」をユーザーと対話するアプリケーション／コンポーネントの態様と切り離すことができる。

【０５０９】
本明細書に記載されるプロセスの少なくとも一部は、１つまたは複数のプロセッサーによる実行のためのコンピュータ可読命令に具体化することができ、したがって、プロセスの操作を、部分的にまたは全体として、１つまたは複数のコンピュータシステムの機能的コンポーネントによって実施することができる。したがって、本明細書に記載されるコンピュータによりインプリメントされるプロセスは、一部の状況では、例として参照されるものである。しかし、他のインプリメンテーションでは、本明細書に記載されるコンピュータによりインプリメントされるプロセスの操作の少なくとも一部を様々な他のハードウェア構成に配置することができる。したがって、本明細書に記載されるコンピュータによりインプリメントされるプロセスは、図５および６に関して記載されているシステムおよび構成に限定することを意図せず、全体として、または部分的に、１つまたは複数の追加のシステムおよび／またはコンポーネントによってインプリメントすることができる。
【０５１０】
本明細書に記載されるフローチャートでは、操作が逐次的なプロセスとして示され得るが、操作の多くは、並行してまたは併せて実施されてもよい。さらに、操作の順序を再配置することができる。プロセスは、その操作が完了したときに終了する。プロセスは、方法、手順、アルゴリズムなどに対応し得る。方法の操作は、全体的にまたは部分的に実施することができ、他の方法の操作の一部または全部と併せて実施することができ、また、本明細書に記載されるシステム、またはその任意の一部分、例えばシステムのいずれかに含まれるプロセッサーなど、任意の数の異なるシステムによって実施することができる。
【実施例】
【０５１１】
（実施例１）
本解析では、本発明者らの、無細胞ＤＮＡ（ｃｆＤＮＡ）におけるゲノムの変更とメチル化シグネチャーの両方を同時に捕捉する高感度の標的化アッセイを実証した。本発明者らのアッセイは、がんと健康なドナーとを分類する差次的メチル化を検出すること、ならびにプロモーターメチル化を定量することができる。腫瘍関連メチル化ｃｆＤＮＡシグナルを捕捉するために、健康な個体に由来する血漿ｃｆＤＮＡにおける非メチル化領域を標的とする広範なゲノムパネル（約１５．２Ｍｂ）に対してカスタムアッセイを開発した。このパネルは、周知の腫瘍抑制因子遺伝子（ＴＳＧ）、例えばＰＴＥＮ、ＴＰ５３、および相同組換え修復（ＨＲＲ）遺伝子、例えばＢＲＣＡ１、ＲＡＤ５１を含む、１１，９５０種の遺伝子のプロモーター領域を網羅する。本実施例において使用した、定量的メトリクス／尺度を生成するための配列決定データの解析は、本明細書に記載される１つまたは複数の技術を使用して実施した。
【０５１２】
アッセイを行う各試料について、網羅される遺伝子それぞれの定義済みプロモーター領域を用い、観察されたメチル化シグネチャーからメチル化スコアを得る。プロモーター高メチル化コールを行う方法を訓練するために、まず、１３１例のがんのないドナーの血液試料に対してモデルを訓練し、その特異度を評価した。次いで、５５９例のがん患者（２０３例の肺がん、１４６例の乳がん、１５１例の膀胱がん、３２例のＣＲＣおよび２７例の他のがんタイプ）ならびに２，６３１例のがんのないドナーに由来する血液試料の検証データセットに対する性能を試験した。
【０５１３】
これらの１３１例のがんのないドナーの血液試料において、プロモーター－メチル化コールを行う閾値を確立した。８８種のがん関連遺伝子に解析の焦点を当てる。
【０５１４】
確立した閾値の下で、１３１例の訓練試料全てにわたって１２件のプロモーターメチル化（８８種の遺伝子のうちの少なくとも１つにおいて）が観察された（図７Ａ）。２，６１２例のがんのないドナーのテストデータセットでは、これらの８８種の遺伝子における合計２５１件のプロモーター－メチル化コールが観察され（図７Ｃ）、そのことから特異

度は＞９９％と推定される。５５９例の末期がん患者のテストデータでは、３３４例（６０％）が、これらの遺伝子に少なくとも１つのプロモーターメチル化を有するとコールされた。プロモーター－メチル化コールがなされた試料では、試料あたりのコール中央値は３である（図７Ｂ）。

【０５１５】

テストデータから、がん患者および健康なドナーにおけるメチル化シグナルの分布を手作業で調査した。メチル化コールを有するがん患者では、シグナルは主にＣｐＧクラスター内で観察された。がんのないドナーの大多数では、このプロモーター領域においてメチル化に関連するシグナルは観察されなかった。例えば、訓練用セットからの１３１例の健康なドナーでは、ＭＬＨ１プロモーター領域においていかなるメチル化シグナルも観察されなかった。

【０５１６】

想定される偽陽性（ＦＰ）をさらに調査するために、がんのないについて、１件よりも多くのメチル化コールを有するがん予測スコアを調査した。４０％が、高いがん予測スコアを有することが見出された（がん－正常カットオフの＞１０倍の標準偏差、図８Ａ）。本コホートにおける「がんのない」ドナーは、自己同定されたので、これらの２例のドナーはすでにがんを発症していた可能性がある。図８Ｂは、がんのないドナーにおいて最も頻繁にプロモーターメチル化がコールされる遺伝子を示す表である。

【０５１７】

低がん予測スコアを有する試料に関しても、数種のがん関連遺伝子、例えばＢＲＣＡ１についておそらく真のメチル化が見出された。刊行物からは、一般集団ではプロモーターメチル化が非常に低い率で散発的に起こり得ることが示されている。

【０５１８】

本発明者らの方法のＬｏＤを、ｉｎ－ｓｉｌｉｃｏにおいてＫＭ１２細胞株とがんのないドナーを０．１％、０．３％、０．５％、０．７％および１％のレベルで混合することによって計算した。細胞株ＫＭ１２において高いシグナルを有する、プロモーター高メチル化されているとコールされた６種のがん関連遺伝子についてＬｏＤを計算した（図８Ｃ）。

【０５１９】

各遺伝子について、そのＬｏＤは、健康な集団ではメチル化レベルおよびそのバックグラウンドノイズの影響を受ける。ＭＬＨ１はＫＭ１２において完全なメチル化を有し（直交性実験によって検証された）、一般集団においては非常に低いメチル化率を有し、全ての遺伝子の中で最低ＬｏＤを有する。他の５種の遺伝子に関しては、それらのＬｏＤは０．１％～０．５％の間の範囲である。

【０５２０】

プロモーター高メチル化をさらに検証するため、および本発明者らの発見を検証するために、ＴＣＧＡデータベースからのがん試料におけるＴＳＧおよびＨＲＲ遺伝子を調査した。４５０Ｋアレイからメチル化データが入手可能な２，３８０例の原発腫瘍がん試料に解析の焦点を当てる。各遺伝子に関して、ＴＣＧＡデータおよび本発明者らのテストデータセットの両方におけるその存在率（プロモーターメチル化が観察される試料のパーセンテージ）を計算した（図９）。

【０５２１】

大多数の遺伝子に関して、ＴＣＧＡと本発明者らのメチル化検出方法の間で同様の存在率が観察された。ＴＣＧＡにおける試料は組織試料であり、それらの多くがまだ初期段階にあるが、本発明者らのテストデータセットは、主に末期患者の血液由来のｃｆＤＮＡであることに注目することに価値がある。

【０５２２】

本発明者らのＣＲＣ試料では高い存在率を有するが、ＴＣＧＡ　ＣＲＣデータではそれを有さない２種の遺伝子（ＣＤ８ＡおよびＣＤＫＮ２Ａ）をさらに調査した。ＣＤ８Ａプロモーター領域は、５種のアレイプローブによってカバーされ、そのうち２種は＞０．５

の存在率を示すが、この高い存在率は、メチル化シグナルを示さない他の３種のアレイプローブによって低下した。同様の状況がＣＤＫＮ２Ａで起こり、５種のアレイプローブのうち１種が０．１の存在率を示すが、他の４種はメチル化シグナルの非常に低いレベルを有する。

【０５２３】

本発明者らのアッセイの実現性を明白に示すために、ＭＬＨ１におけるプロモーターメチル化をさらにまとめた。以前の研究から、マイクロサテライト不安定性（ＭＳＩ－Ｈ）腫瘍を有するＣＲＣ患者の５４％～１００％がＭＬＨ１プロモーターメチル化を有することが示されている。

【０５２４】

１，９６６例のＣＲＣ患者において、ＭＳＩ－Ｈ群（フィッシャーのｐ＜０．０５）では、マイクロサテライト安定（ＭＳＳ）状態の患者およびがんのない集団と比較して有意に高いＭＬＨ１プロモーターメチル化が検出された（図１０Ａ）。

【０５２５】

さらに、ＣＲＣの共通のドライバー変異である変異ＢＲＡＦ－Ｖ６００Ｅの存在率を試験した。本発明者らの所見から、ＭＬＨ１プロモーター高メチル化とＢＲＡＦ－Ｖ６００Ｅの強力な共存が示され（図１０Ｂ）、これは以前の所見と一致する。

【０５２６】

（実施例２）

液体生検は、バイオマーカーを同定するための、組織生検に対する迅速かつ非侵襲性の代替を提供する。つい最近、その適用が、治療に対する初期応答（すなわち分子応答）の評価、および初期の状況では、微小残存病変（ＭＲＤ）および初期疾患再発の検出が含まれるまでに拡張された[1]。体細胞変異によって推定される循環腫瘍画分（ｃＴＦ）は腫瘍進行および予後に十分に関連付けられるが、未確定の潜在能を持つクローン性造血（ＣＨＩＰ）による干渉が生じ得、また、検出可能な体細胞変異を欠く無細胞ＤＮＡ（ｃｆＤＮＡ）試料に関しては、体細胞性腫瘍画分を推定することができない。この解析では、エピゲノムシグネチャーにより、体細胞変異に対して直交分析物（orthogonal analyte）を使用してｃＴＦが正確に測定され、検出可能な腫瘍ドライバーバリアントがない場合であってもｃＴＦ推定が可能になることを実証する。

【０５２７】

腫瘍関連メチル化ｃｆＤＮＡを捕捉するために、健康な個体に由来する血漿ｃｆＤＮＡにおける非メチル化領域を標的とする広範なゲノムパネル（１５．２Ｍｂ）に対してカスタムアッセイを設計した。メチル化を支持するＤＮＡ分子を本発明者らのアッセイによって富化し、この情報を後処理して、本発明者らの機械学習モデルに入れた。

【０５２８】

このパネルを用い、約２，０００例のがん患者および約２，６００例のがんのないドナーの訓練用セットに由来する血漿試料をプロファイリングした（図１１）。がん／がんのない状態の予測のために、ロジスティック回帰モデルを訓練した。腫瘍画分予測のために、ゲノムバリアントコールのアレル頻度を基礎をなす真理値（underlying truth）として利用して線形モデルを訓練した。訓練性能を５重の交差検証によって検証した。

【０５２９】

テストデータセットに対して、５５９例のがん患者、１３１例のがんのないドナーをプロファイリングした。全てのがん特異的予測モデルをテストデータセットに適用して、１）単一のモデルおよび集約されたモデルのがん／がんのない分類性能；２）腫瘍画分予測性能を推定した。

【０５３０】

本発明者らのメチル化モデルの正確度をさらにベンチマークテストするために、ｉｎ－ｖｉｔｒｏおよびｉｎ－ｓｉｌｉｃｏタイトレーションデータセットを構築した。ｉｎ－ｖｉｔｒｏタイトレーションデータセットは、結腸直腸がん（ＣＲＣ）を有する患者に由来するｃｆＤＮＡを、がんのないドナーに由来する血漿中に実験的タイトレーションによ

って混合することによって生成した。ｉｎ－ｓｉｌｉｃｏタイトレーションデータセットは、ＣＲＣ患者に由来する配列決定リードを、がんのないドナーに由来する配列決定リードとコンピュータにより混合することによって生成した。

【０５３１】

メチル化ｃｆＤＮＡからのｃＴＦにより、体細胞変異に基づく方法の現行の限界を克服することができる。本発明者らのメチル化アプローチは、腫瘍ドライバー陽性症例および陰性症例においてｃＴＦを正確に検出することが可能であることを実証する。本発明者らのアッセイは、がん患者における差次的にメチル化されたシグナル中のメチル化シグナルを有する分子を確実に富化させることができる。本発明者らのがん特異的モデルは、特異度９５％を維持すると同時に、末期がん患者について＞９０％の検出率を実現する。ゲノムドライバー変異コールを真の腫瘍画分の代理として用い、本発明者らのメチル化に基づく予測は、この代理を用いてＣＲＣに関して０．８５の相関を有する。

【０５３２】

腫瘍陰性症例は、ステージＩ～ＩＩＩがんを有する患者の３０～５０％およびステージＩＶがんを有する患者の１５～２０％であると推定されるので、本発明者らのメチル化アプローチにより患者のケアおよび管理のためのより良好な評価が提供される見込みがあり得る。

【０５３３】

２，０００例のがん患者および２，６１４例のがんのないドナーの訓練用データセットに対するがん／がんのない分類および腫瘍画分予測のための本発明者らの機械学習モデルを構築する。まず、本発明者らの方法を訓練用データセットに対して、５重の交差検証によって評価した－このプロセスを１０回繰り返した。特異度９５％において、がん／がんのない状態についての本発明者らの予測モデルは、全てのステージにわたる試料について９３％の平均検出率を有する。腫瘍画分予測モデルは、状態予測モデルと同様の性能を有する（図１２Ａ）。この実施例において使用した、定量的メトリクス／尺度を生成するための配列決定データの解析は、本明細書に記載される１つまたは複数の技術を使用して実施した。

【０５３４】

訓練プロセス後、訓練されたモデルおよびそれらの特異度９５％カットオフを、５５９例のがん患者および１３１例のがんのないドナーの独立したテストデータセットに適用した。テストデータセットに関して、３つのモデル全てにわたって、特異度９７％が観察され、合計４例の偽陽性（ＦＰ）が観察された。ＣＲＣモデルにおけるＦＰは肺モデルにおける４例のＦＰに含まれる（図１２Ｂ）。

【０５３５】

テストデータセットは全て末期患者であり、乳がんモデルについての感度（９５％）は、その乳がん患者の大多数（＞７０％）が初期である訓練用データセットにおける全ステージ感度（６２％）よりも高い。

【０５３６】

手作業での調査により、がんのないドナーにおいて、ＦＰが腫瘍－正常カットオフをわずかに上回ること、ならびに、一部の強力なシグナルが、散発的な変異を有する領域に由来することが明らかになった。高度な領域定義およびモデルの微調整により、ＦＰを排除することができる可能性がある。他方では、差次的にメチル化された領域において少数のＦＮが比較的弱いシグナルを有し、それにより、これらの試料が訓練用データではあまり表されておらず、さらなる試料の追加が検出力を増加させることに役立ち得る。

【０５３７】

本発明者らのＣＲＣ腫瘍画分予測モデルを、１，０００例のＣＲＣ患者試料由来のリードを１，０００例のがんのないドナーのリードとランダムに混合することによって生成した１，０００例の試料のｉｎ－ｓｉｌｉｃｏタイトレーションデータセットに対して適用した。この方法により、これらの試料の＞９９％において０．１％を超えるｃＴＦが定量された。対照的に、２，０３７例のがんのないドナーに適用した場合、試料の＜５％によ

り、推定ｃＴＦ＞０．１％がもたらされた。
【０５３８】
本発明者らのＣＲＣ腫瘍画分予測モデルを、本発明者らのＣＲＣドライバー変異のゲノムコーラーとともに、がんのないドナーにＣＲＣ患者由来の血漿を異なるレベルで実験的にタイトレーションすることによって生成された２７０例の試料で構成されるｉｎ－ｖｉｔｒｏデータセットに適用した。
【０５３９】
低ＴＦの全ての試料において、メチル化の変動係数（ＣＶ）は、ゲノム変異によって推定されるｃＴＦからのＣＶよりもロバストであった：０．３～１％のＴＦで、メチル化に基づくＴＦはゲノムＴＦの５分の１である（図１３Ａ）。
【０５４０】
以前の研究からの予め定義されたＣＲＣのドライバー遺伝子のセットを用い、ゲノムデータからコールされたドライバー変異のアレル頻度を、基礎をなす真の腫瘍画分の概数として使用した。まず、交差検証から予測された腫瘍画分を訓練用セットにおける真の腫瘍画分と比較した（図１３Ｂ）。ピアソン相関０．８５で、メチル化に基づく腫瘍画分の大多数が基礎をなす真理値と一致する。
【０５４１】
メチル化により予測されたＴＦが真理値ＴＦよりも有意に高い上位１０例の外れ値について、６例は、本発明者らの包括的なパネル全体にわたって、ＣＲＣとの高い類似性を有する、がんのない試料である。メチル化により予測されたＴＦが真理値ＴＦよりも有意に低い１９例の外れ値について、それらの非常に高いゲノムＴＦ（＞１０％）は、本発明者らのドライバー変異フィルターを逃れたＣＨＩＰに由来する可能性がある。
【０５４２】
メチル化により、ＴＦ＜０．１％の試料が解明され得る可能性があるが、ｃＴＦｘ＜０．１％は現行のコンパニオン診断製品の大多数についての推定される検出限界（ＬｏＤ）を下回るので、この可能性を評価することは難しい。これを試すために、１００例のＣＲＣ患者に由来する分子を１００例のがんのないドナー由来の分子に０．００５％～０．０１％の間のレベルでｉｎ－ｓｉｌｉｃｏでタイトレーションし、本発明者らの訓練されたモデルをこのデータセットに適用した（図１３Ｃ）。
【０５４３】
０．０３％～０．０１％の範囲内のシミュレートされた試料のｃＴＦはメチル化に基づくｃＴＦと一致したが、推定値は０．０１％を下回った。このレベルを下回ると、ロバストなｃＴＦ予測を可能にするための、ノイズシグナルを上回る十分な真のメチル化シグナルが存在しないという仮説を立てる。
【０５４４】
以前の研究により、ステージＩ～ＩＩＩがんを有する患者の３０～５０％、およびステージＩＶがんを有する患者の１５～２０％が、検出可能なドライバー変異を欠き得ることが示されている。現行の方法はドライバー変異に依拠するものであり、したがって、これらの患者については腫瘍画分を定量することができず、アンメットニーズを有する大きな集団が残されている。ここで、本発明者らは、メチル化に基づくｃＴＦが変異に基づくｃＴＦと強力に相関付けられ、それによって、患者のこの１５～５０％について腫瘍画分の予測が可能になることを実証した（図１４）。
【０５４５】
予測通り、これらの患者における予測ＴＦの中央値は、検出可能なドライバー変異を有する患者よりも小さい（０．３％対０．０２％）。
【０５４６】
（実施例３）
多発がんの血液に基づく試験は、以下の２つの方法で臨床的利益をもたらし得る：より簡便であり、投与が容易であり、患者にやさしいモダリティを用いて、ガイドラインにより推奨されるがんスクリーニングに対するアドヒアランスを改善することによって、なら

びに、早期検出および介入により命が救われ得るにもかかわらずスクリーニング推奨を欠くがんタイプにおける初期（ステージＩ／ＩＩ）腫瘍の検出によって。両方の機会に対処する臨床的に意味のある性能を有する単一の試験は、まだ開発されていない。

【０５４７】

ｃｆＤＮＡデバイスを、結腸直腸がん（ＣＲＣ）、肺がん、および複数の追加の固形腫瘍がんタイプの検出を可能にするＣｐＧのメチル化解析に基づき、特異度の閾値をがんクラスごとに現行のスクリーニング推奨および臨床診断経路に基づいて調整して評価した。この実施例において使用した定量的メトリクスを生成するための配列決定データの解析は、本明細書に記載される１つまたは複数の技術を使用して実施した。

【０５４８】

方法。血液試料を、結腸直腸がん（Ｎ＞２，０００）、肺がん（Ｎ＞３００）、および他の固形腫瘍がん（膀胱がん、胃がん、肝臓がん、卵巣がん、膵臓がん（Ｎ＞３００））を有する個体ならびにがんのない個体（Ｎ＞３，０００）の複数のコホートから得た。

【０５４９】

がんの検出のための、ゲノミクス、エピゲノミクス、およびプロテオミクスが統合された液体生検ＣＲＣスクリーニング試験がデバイスの骨格である。高いＣｐＧ密度および高いメチル化の程度を有する断片のアッセイにおいて富化される血漿由来ｃｆＤＮＡのエピゲノムパターンのさらなる解析により、肺がんおよび多発がん検出アルゴリズムを開発した。この実施例において使用した、定量的メトリクス／尺度を生成するための配列決定データの解析は、本明細書に記載される１つまたは複数の技術を使用して実施した。

【０５５０】

ＣＲＣおよび肺がん検出の感度は、９０％標的特異度閾値で計算する。多発がん特異度は９８％閾値を目標に設定した。肺がんおよび多発がん検出性能を、開発セットの交差検証からの結果と、検証セットに対する単一の事前ロックモデルからの結果を組み合わせることによって得る。図１５Ａは、ステージＩ／ＩＩ患者およびステージＩＩＩ／ＩＶ（page III/IV）患者における肺がん検出についての個体における陽性率を示すグラフ表示である。図１５Ｂは、ステージＩ／ＩＩ患者におけるおよびステージＩＩＩ／ＩＶ患者における多発がん検出（膀胱がん、胃がん、卵巣がん、膵臓がん、および肝臓がん）についての個体における陽性率を示すグラフ表示である。さらに、図１６は、ステージＩ患者、ステージＩＩ患者、ステージＩＩＩ患者、およびステージＩＶ患者における多発がん検出（膀胱がん、胃がん、卵巣がん、膵臓がん、および肝臓がん）についての個体における陽性率のより詳細な内訳を示すグラフ表示である。ＣＲＣ、肺がん、および多発がんについての特異度閾値は、がんタイプおよび臨床診断経路について調整されたアッセイ性能がもたらされるように選択する。

【０５５１】

結果。閾値が調整されたこの単一デバイスにより、特異度９０％において、ＣＲＣ感度９１％（ステージＩ／ＩＩ：９３％）および肺がん感度８５％（ステージＩ／ＩＩ：７５％）がもたらされた。多発がん感度は特異度９８％において７５％（ステージＩ／ＩＩ：６６％）であった。

【０５５２】

結論。現行のがんスクリーニング推奨および臨床診断経路に対して調整された性能を有するこの血液に基づくがんスクリーニングおよび検出デバイスにより、スクリーニングガイドラインを有するがん（ＣＲＣおよび肺）に対する、現在利用可能なスクリーニング試験と同等の性能、ならびに、初期介入により臨床的利益がもたらされ得る、スクリーニングガイドラインがないがんタイプにおいて、臨床的に意味のある初期段階での検出がもたらされる。このプロファイルによると、この多発がんの血液に基づく試験は、ＳＥＥＲ推定値に従って、２０２２年に予測されたがん診断の３２％を網羅し得、全体的な感度は８０％（ステージＩ／ＩＩ：７８％）であり、このテクノロジーの、初期がんの検出に関して臨床的に意味のある結果をもたらす能力が強調される。

【０５５３】

（実施例４）
循環腫瘍ＤＮＡ（ｃｔＤＮＡ）レベルおよび後続の時点におけるレベルの変化（例えば、処置中のベースラインからの変化または経時的な術後変化）は、患者の予後および治療に対する応答を予測するための有望なツールである。既存の方法では、一般に、体細胞変異のマイナーアレル頻度（ＭＡＦ）を使用して、循環腫瘍画分（ｃＴＦ）を定量する。それらの性能は、検出可能な体細胞変異の数および関連する検出限界（ＬｏＤ）、ならびにコピー数バリエーションおよび非腫瘍変更、例えばクローン性造血（ＣＨＩＰ）の干渉によって限定され得る。ここで、８００種を超える遺伝子を網羅する次世代配列決定パネルをゲノムワイドなメチル化検出とともに使用したｃＴＦレベルのＬｏＤ、精度および定量限界（ＬｏＱ）ならびに変化を記載する。この実施例において使用した定量的メトリクスを生成するための配列決定データの解析は、本明細書に記載される１つまたは複数の技術を使用して実施した。
【０５５４】
単一の試料のエピゲノムｃＴＦ（エピＭＡＦによって表される）を、がんのないドナーおよび混合がんタイプを有する患者を含む、５，０００例を超える個体の臨床採血試料を有する、本発明者らの内部訓練用データを使用して較正したメチル化パネルの標的化領域にわたるメチル化シグナルから推定する。患者特異的なメチル化領域を同定し、対形成した領域のメチル化シグナルを比較するために、エピゲノミクスｃＴＦ変化を、同じ患者に由来する２つまたはそれよりも多くの試料で比較する。体細胞変異もゲノムパネルを通して検出した。
【０５５５】
ＬｏＤを、反復実験の９５％以上で腫瘍に由来するメチル化シグナルが示される最低ｃＴＦレベルとして決定した。ＬｏＱを、反復実験にわたる変動係数（ＣＶ）が３０％未満である最低ｃＴＦレベルと定義した。５，０４５例のがん患者の臨床試料に関して、メチル化に基づくｃＴＦの正確度を、体細胞変異の最大ＭＡＦから得られたｃＴＦと比較した。
【０５５６】
結腸直腸がん試料１例、乳がん試料１例、肺がん試料１例、および細胞株試料１例を、０．１％～０．５％ＭＡＦの範囲の目標レベルでがんのないバックグラウンドにタイトレーションした。＞９５％の正確度で検出可能な腫瘍由来ＤＮＡの最低濃度と定義したメチル化ＬｏＤはおよそ０．０５％であると推定された。
【０５５７】
図１７は、エピゲノムＭＡＦの、結腸直腸がん、肺がん、および乳がんについての標的ＭＡＦとの関連のグラフ表示であり、臨床濾過におけるエピゲノムｃＴＦの正確度を示す。臨床試料のエピゲノムｃＴＦは、ピアソン－ｒが０．９よりも大きく、直線性エラーが５％未満であることによって示されるように、基礎をなすタイトレーションレベルと高度の一貫性を示し、異なるタイトレーションレベル間で強力な直線性を維持する。
【０５５８】
図１８は、技術的反復実験におけるエピゲノムｃＴＦバリエーションを示す表であり、エピゲノムｃＴＦの定量的精度が、ＣＲＣ、肺および乳がん臨床試料において０．１％未満のＬｏＱに達することが可能であることを示す。図１９Ａは、体細胞変異に基づくｃＴＦが、特に０．５％またはそれよりも高いｃＴＦレベルにおいて、同じｃＴＦレベル内での反復実験についてロバストであることを示すグラフ表示である。しかし、エピゲノムｃＴＦは、低タイトレーションレベルにおいてより安定である。図１９Ｂは、エピゲノムｃＴＦが、１００％の評価率を維持することができ、０．１％のｃＴＦに至るまでのＬｏＱを有することを示すグラフ表示である。
【０５５９】
低ｃＴＦにおけるエピＭＡＦのロバストネスは、パネル内の「評価可能な」領域の数が多いことに起因する。具体的には、臨床的な結腸直腸がん（ＣＲＣ）試料の、０．５％および０．３％ｃＴＦのタイトレーションレベルを用いた２回の技術的反復実験では、エピ

ＭＡＦは数千もの領域に基づいて推定されたが、一方で、ゲノムｃＴＦは、３種の検出可能な体細胞変異からのみ推定することができた。図２０Ａは、臨床的用量設定の第１の反復実験についてのメチル化シグナルおよび体細胞変異のグラフ表示である。図２０Ｂは、臨床的用量設定の第２の反復実験についてのメチル化シグナルおよび体細胞変異のグラフ表示である。図２１は、ゲノムのみの方法およびメチル化方法を使用して計算した、第１の反復実験および第２の反復実験についてのｃｔＤＮＡレベルの変化を示す表である。

【０５６０】

５，０４５例の臨床試料（ＣＲＣ患者、肺がん患者、および乳がん患者（ステージＩ～ＩＶについてＮ＝５２２、９０９、６９６および７８４、ステージ不明２，６５６例を合わせて）のコホートでは、メチル化解析に基づいて、６４％が体細胞変異を有し、９０％がｃｔＤＮＡの存在の証拠を示した。注目すべきことに、エピＭＡＦ方法では、真のｃＴＦのより正確な表示である可能性がある「ドライバー」体細胞変異を有する患者について体細胞性ＭＡＦと高度に一貫した推定値がもたらされる。図２２は、臨床試料に関するエピゲノムｃＴＦ対ゲノムｃＴＦのグラフ表示である（１つの試料について１つの点）。図２３は、乳がん、結腸直腸がん、肺がん、および他のがんの群についての初期および末期がん患者におけるエピＭＡＦ分布のグラフ表示である。

【０５６１】

様々ながんタイプからの２３１例の追加の試料を解析したところ、検出可能な体細胞変異を欠く大多数の試料が０．１％未満の腫瘍画分を有することが見出された（図２２のサイドバーに示されているように）。図２３に示されているように、初期がんを有する患者は、そのような低腫瘍画分を有する傾向がより多く、その場合、現行の体細胞変異に基づく方法では変異を検出することも定量することもできない。

【０５６２】

メチローム配列決定は、液体のみのアプローチでｃｔＤＮＡレベルの正確な定量を可能にし、利用しやすい縦断的ｃｔＤＮＡモニタリングをもたらす。以前の研究から、ステージＩ～ＩＩＩがんを有する患者の３０～５０％、およびステージＩＶがんを有する患者の１５～２０％が検出可能な体細胞変異を欠くことが示されている。本明細書に記載される方法論では、これらの患者におけるｃＴＦが正確に検出および定量され、患者の評価および疾患管理が改善される。

【０５６３】

（実施例５）

いくつかの対象から試料が提供され、それらを無細胞ＤＮＡアッセイに従って解析した。この実施例において使用した定量的メトリクスを生成するための配列決定データの解析は、本明細書に記載される１つまたは複数の技術を使用して実施した。ＭＢＤを含むいくつかの溶液で試料を処理して、試料の分子を第１の分配、第２の分配、および第３の分配に分離することにより、試料から情報を得た。ここで、各分配は、ある１つの量のＣＧ含量を有するゲノム領域内の異なるメチル化シトシンの量を表すものであった。各分配について、所与の数のメチル化ＣｐＧを有する分子を決定した。図２４は、３つの分配に含まれるメチル化シトシンの数を示す確率分布を示すグラフ表示である。

【０５６４】

各試料について、ＭＢＤ分配プロファイルを、所与の数のＣｐＧを有する陽性対照領域からのメチル化分子の数を、ある特定の閾値を超えるＣｐＧ含量を有する陽性対照領域からのメチル化分子の総数で割ったものとして計算した。ある特定の範囲内のＣｐＧ含量についての比または比の対数変換を使用して、領域レベルのメチル化測定値を修正する。調整された測定値は、無調整の測定値からＭＢＤ分配オフセットを差し引いたものと等しい。オフセットは、比または比の対数変換の加重和と推定される。一部の例では、ＣｐＧ閾値を１２、１３、１４および１５に設定し、ＣｐＧ範囲は１～３０個のＣｐＧ、４～８個のＣｐＧおよび６～１２個のＣｐＧである。

【０５６５】

図２５Ａは、試薬の第１のセットを使用してＭＢＤで処理した試料の第１の群および試

薬の第２のセットを使用してＭＢＤで処理した試料の第２の群についての、第１の分類領域についての領域レベルでのメチル化測定値の変化を示すグラフ表示を含む。領域スコアは、正規化方法を適用する前の、無調整の領域レベルのメチル化測定値を示し、一方で、正規化された領域スコアは、調整された領域レベルのメチル化測定値を示す。さらに、図２５Ｂは、試薬の第１のセットを使用してＭＢＤで処理した試料の第１の群および試薬の第２のセットを使用してＭＢＤで処理した試料の第２の群についての第１の分類領域についての第２の分類領域についての（for a second classification region for a first classification region）領域レベルでのメチル化測定値の変化を示すグラフ表示を含む。領域スコアは、正規化方法を適用する前の、無調整の領域レベルのメチル化測定値を示し、一方で、正規化された領域スコアは、調整された領域レベルのメチル化測定値を示す。

【図面】

【図１】

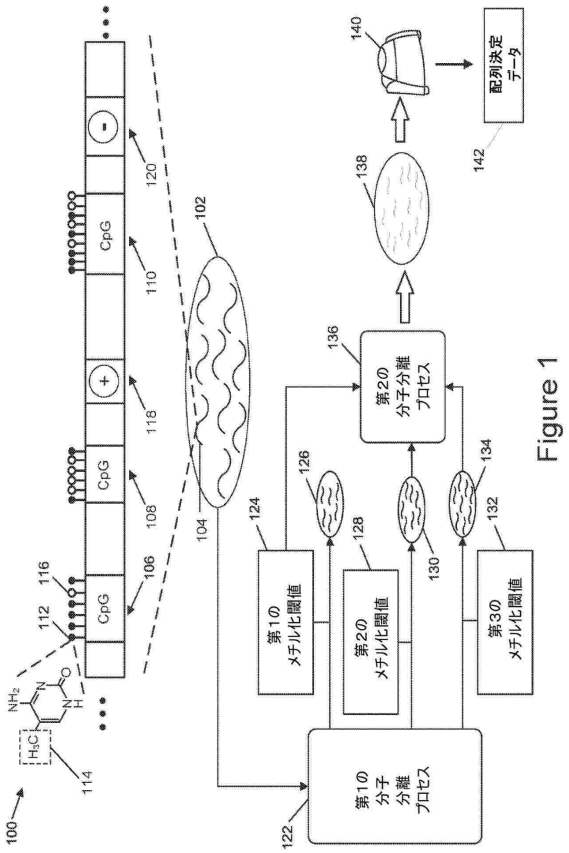


Figure 1

【図２】

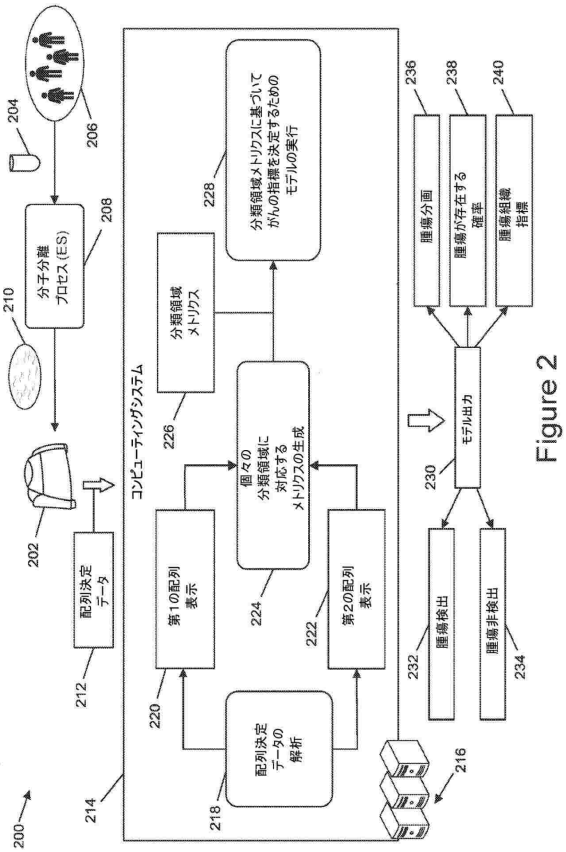


Figure 2

【図３】

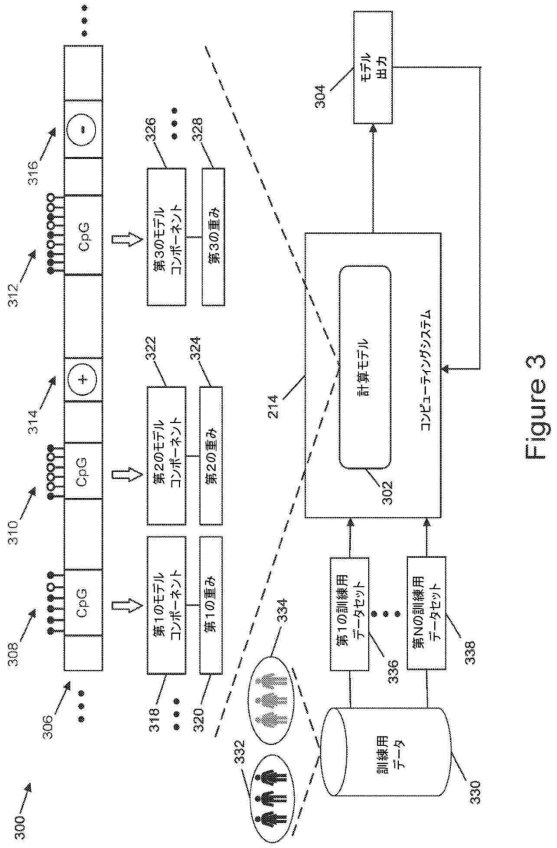


Figure 3

【図４】

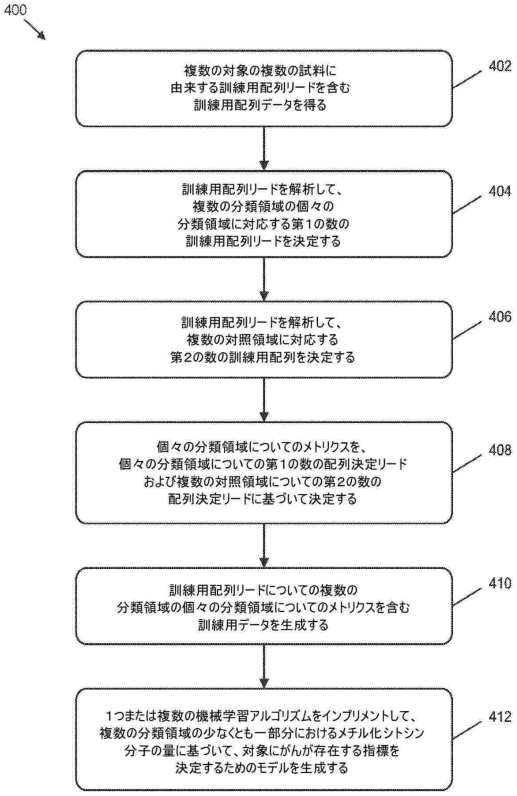


Figure 4

【図５】

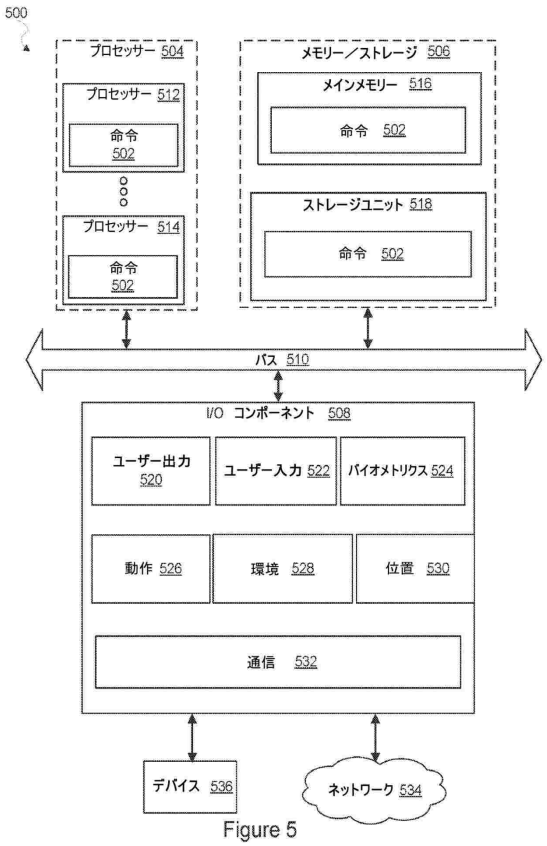


Figure 5

【図６】

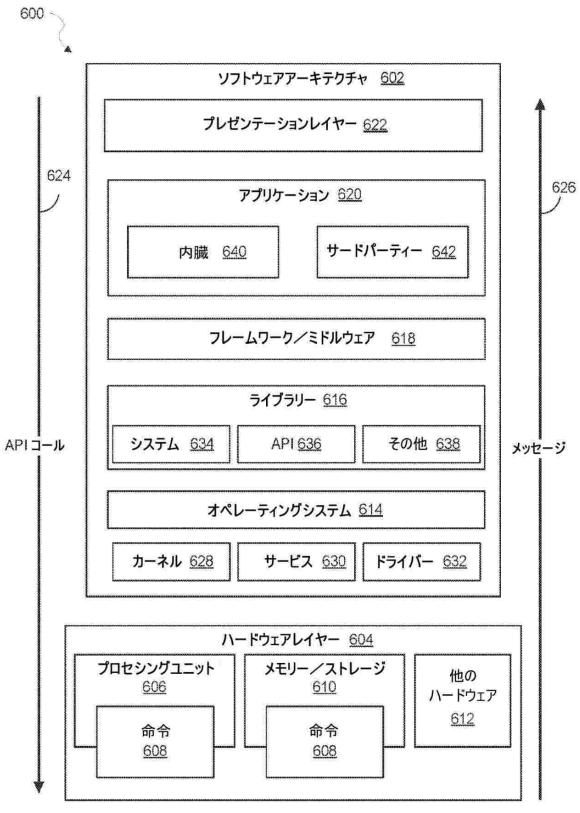


Figure 6

【図７】

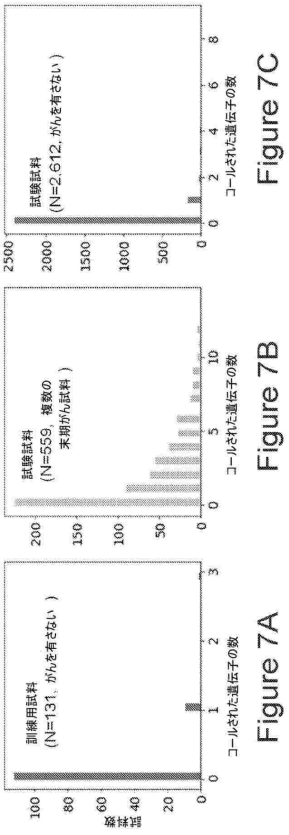


Figure 7A

Figure 7B

Figure 7C

【図８】

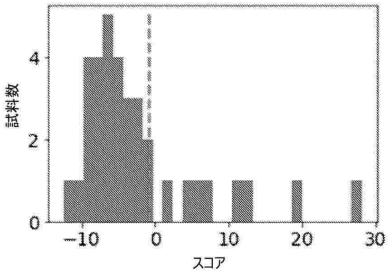


Figure 8A

遺伝子	試料数 (%)
FAT1	56 (2.1%)
RAD51C	41 (1.6%)
IFNL2	35 (1.3%)
BRCA1	30 (1.1%)
CDH1	24 (0.92%)

Figure 8B

遺伝子	推定 LoD
CD8A	0.50%
CDKN2A	0.30%
HLA-F	0.30%
HLA-L	0.30%
IFNL2	0.10%
MLH1	0.08%

Figure 8C

【図９】

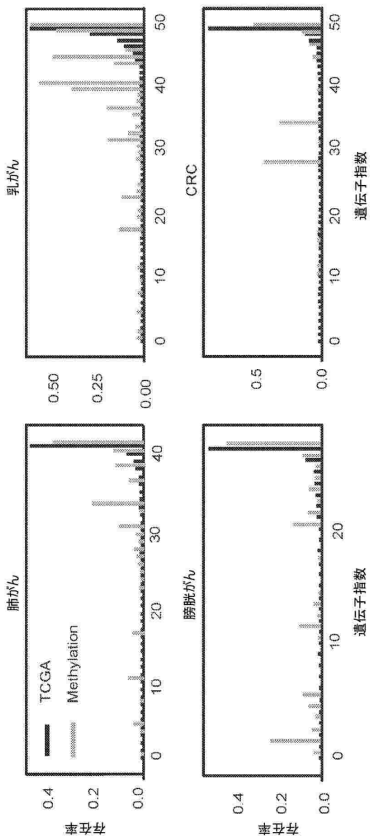


Figure 9

【図１０】

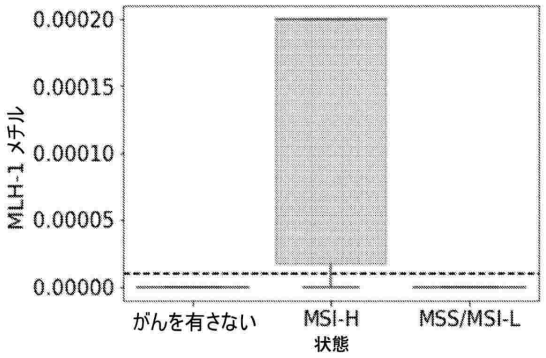


Figure 10A

	BRAF WT	BRAF V600E
MLH1 プロモーター 高メチル化 検出	118 (6.2%)	34 (61%)
MLH1 プロモーター 高メチル化 非検出	1792 (93.8%)	22 (39%)
総計	1910	56

Figure 10B

【図１１】

タイプ	訓練用試料	試験試料
がんを有さない	2,614	131
CRC	2,000	32
肺がん	262	203
乳がん	253	146
他のがん	300	27

Figure 11

【図１２】

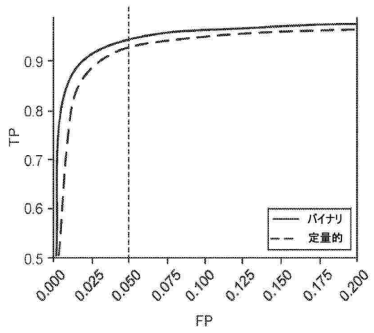


Figure 12A

モデル	特異度 (#TN)	感度 (#TP)
肺がん	96.9% (127)	90% (182)
乳がん	100% (131)	95.3% (139)
CRC	99.2% (130)	84% (26)
集計	96.9% (127)	91.3% (409)

Figure 12B

【図１３】

TF	試料数	CVゲノム	CVメチル化
<0.3%	154	0.72	0.17
0.3%-1%	62	0.24	0.17
>1%	54	0.03	0.05

Figure 13A

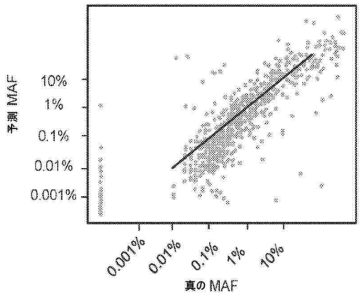


Figure 13B

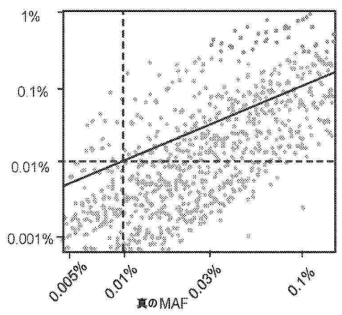


Figure 13C

【図１４】

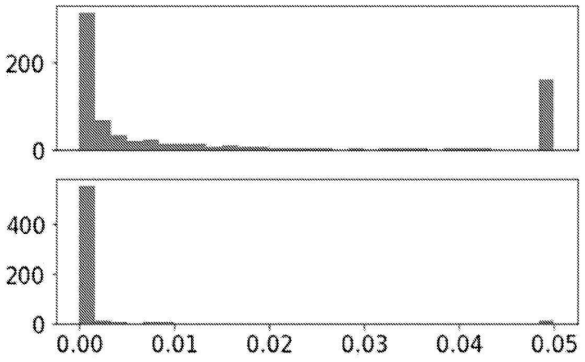


Figure 14

【図１５】

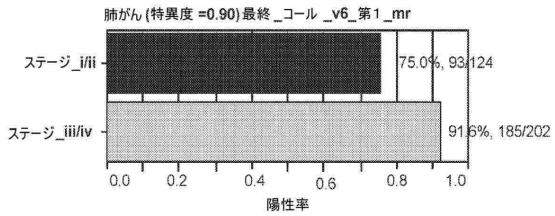


Figure 15A

【図１６】

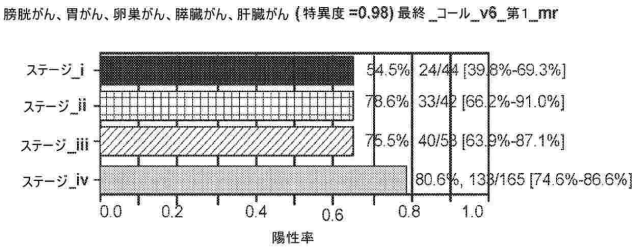


Figure 16

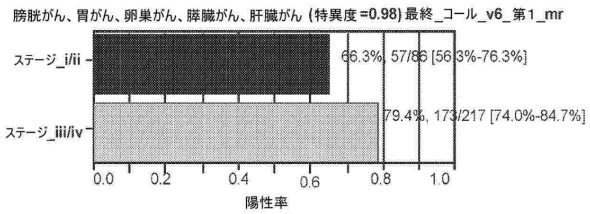


Figure 15B

【図１７】

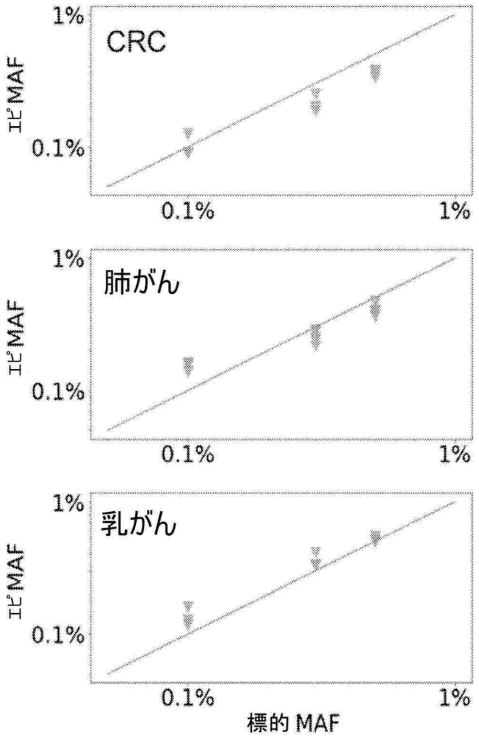


Figure 17

【図１８】

標的 MAF	変動係数 (CV)		
	CRC	肺がん	乳がん
0.5%	4.7%	9.2%	2.3%
0.3%	6.3%	8.2%	4.9%
0.1%	9.0%	9.6%	5.5%

Figure 18

【図１９】

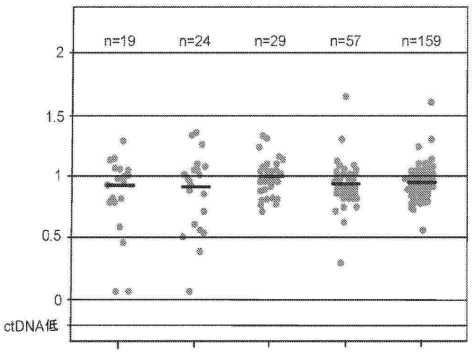


Figure 19A

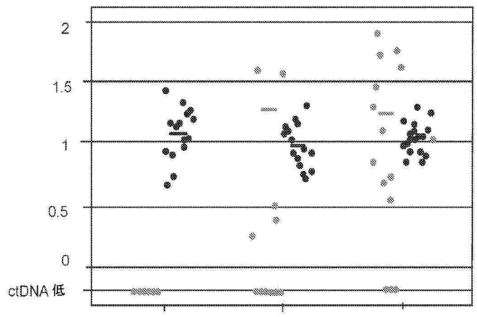


Figure 19B

【図２０】

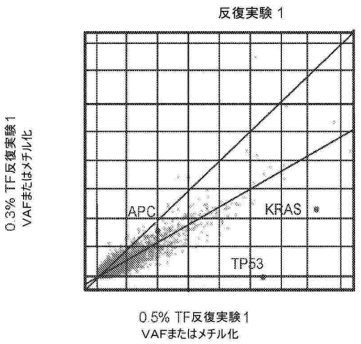


Figure 20A

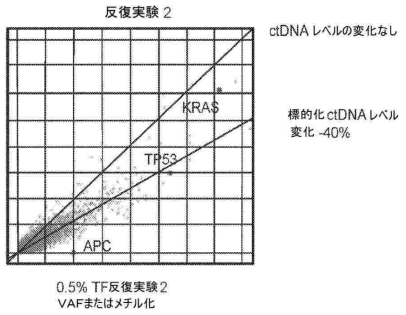


Figure 20B

【図２１】

方法	標的化 ctDNA レベル変化	追跡した特色数	算出された腫瘍分子数	ctDNA レベル変化
ゲノムのみ	-40%	3つの体細胞変異体	約 100	反復実験 1: -84%, 反復実験 2: -30%
メチル化	-40%	>>1000 のメチル化領域	約 100,000	反復実験 1: -39%, 反復実験 2: -39%

Figure 21

【図２２】

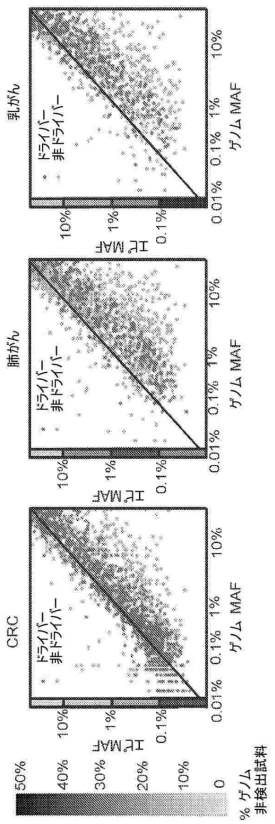


Figure 22

【図２３】

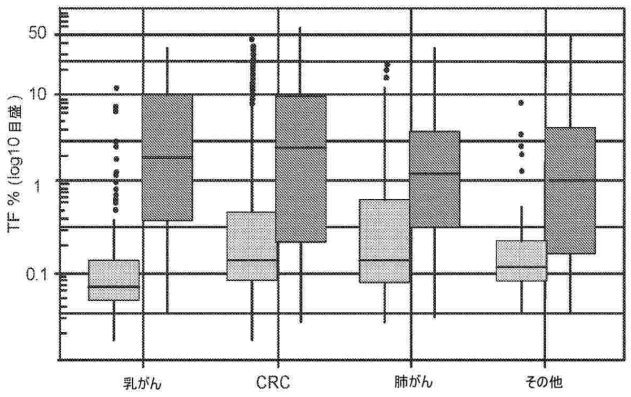


Figure 23

【図２４】

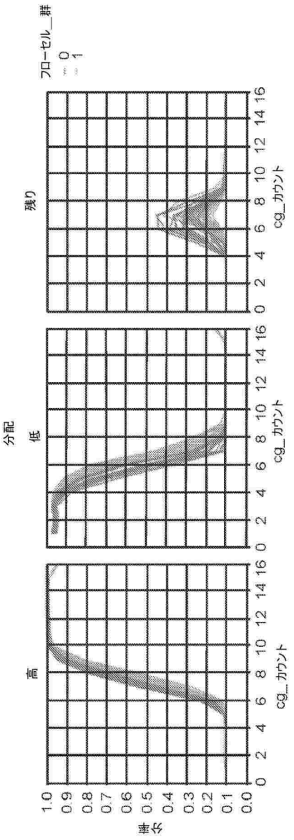


Figure 24

【図２５】

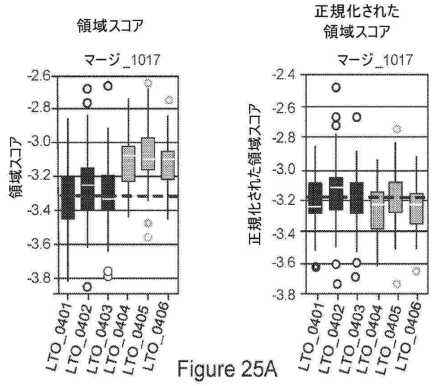


Figure 25A

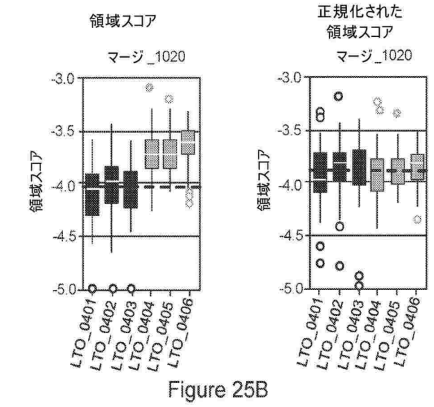


Figure 25B

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/065560

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6869 C12Q1/6886 G16B20/20 G16B30/00 G16B40/10
G16B40/20 G16H50/20
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)
C12Q G16B G16H

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2021/407623 A1 (WESTESSON OSCAR [US] ET AL) 30 December 2021 (2021-12-30) [0048][0075][0090][0108][0123] [0145] [0150] [0118][0119] [0154] [0165][0199]-[0201][0205]; claims 1,11,12; figure 9	1-60
Y	US 2011/028333 A1 (CHRISTENSEN BROCK C [US] ET AL) 3 February 2011 (2011-02-03) [0004],[0010], [0051] - [0155], [0074], [0075]	1-60
A	US 2022/098672 A1 (VENN OLIVER CLAUDE [US] ET AL) 31 March 2022 (2022-03-31) [0073][0074][0125]-[0127] [187]	1-60

☐ 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	Date of mailing of the international search report
7 July 2023	18/07/2023

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Eberhardt, Anja

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

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2023/065560

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2021407623 A1	30-12-2021	EP 4128244 A1	08-02-2023
		US 2021407623 A1	30-12-2021
		WO 2021202752 A1	07-10-2021
US 2011028333 A1	03-02-2011	NONE	
US 2022098672 A1	31-03-2022	AU 2020217719 A1	19-08-2021
		AU 2020219853 A1	26-08-2021
		CA 3127894 A1	13-08-2020
		CA 3129043 A1	13-08-2020
		CN 114026254 A	08-02-2022
		CN 114026255 A	08-02-2022
		EP 3921444 A1	15-12-2021
		EP 3921445 A1	15-12-2021
		IL 285310 A	30-09-2021
		IL 285316 A	30-09-2021
		US 2022064737 A1	03-03-2022
		US 2022098672 A1	31-03-2022
		WO 2020163403 A1	13-08-2020

Form PCT/ISA/210 (patent family annex) (April 2005)

フロントページの続き

,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,ME,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,CV,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,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,MG,MK,MN,MU,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,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

弁護士　山本 健策

(72)発明者　へ， ユペン
アメリカ合衆国 カリフォルニア ９４３０４， パロ アルト， ハノーバー ストリート ３１００

(72)発明者　へ， ジャオレン
アメリカ合衆国 カリフォルニア ９４３０４， パロ アルト， ハノーバー ストリート ３１００

(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 ９４３０４， パロ アルト， ハノーバー ストリート ３１００

(72)発明者　ヤング グリーンウォルド， ウィリアム ウォルター
アメリカ合衆国 カリフォルニア ９４３０４， パロ アルト， ハノーバー ストリート ３１００

(72)発明者　グリーンリーフ， ウィリアム ジェイ.
アメリカ合衆国 カリフォルニア ９４３０４， パロ アルト， ハノーバー ストリート ３１００

(72)発明者　チェン， サイ
アメリカ合衆国 カリフォルニア ９４３０４， パロ アルト， ハノーバー ストリート ３１００

(72)発明者　ジアン， ティンティン
アメリカ合衆国 カリフォルニア ９４３０４， パロ アルト， ハノーバー ストリート ３１００

(72)発明者　ジャオ， ジュン
アメリカ合衆国 カリフォルニア ９４３０４， パロ アルト， ハノーバー ストリート ３１００

Ｆターム（参考）　4B063　QA01 QA17 QA19 QQ03 QQ42 QR55 QS34