

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7657149号  
(P7657149)

(45)発行日 令和7年4月4日(2025.4.4)

(24)登録日 令和7年3月27日(2025.3.27)

(51)国際特許分類		F I		
C 1 2 Q	1/6806(2018.01)	C 1 2 Q	1/6806	Z
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z
C 1 2 N	15/09 (2006.01)	C 1 2 N	15/09	Z

請求項の数 14 (全81頁)

(21)出願番号	特願2021-523280(P2021-523280)	(73)特許権者	515059083
(86)(22)出願日	令和1年10月31日(2019.10.31)		ガーダント ヘルス, インコーポレイテッド
(65)公表番号	特表2022-512848(P2022-512848 A)		アメリカ合衆国 カリフォルニア 9 4 3 0 4, パロ アルト, ハノーバー ストリート 3 1 0 0
(43)公表日	令和4年2月7日(2022.2.7)	(74)代理人	100078282
(86)国際出願番号	PCT/US2019/059217		弁理士 山本 秀策
(87)国際公開番号	WO2020/092807	(74)代理人	100113413
(87)国際公開日	令和2年5月7日(2020.5.7)		弁理士 森下 夏樹
審査請求日	令和4年10月31日(2022.10.31)	(74)代理人	100181674
(31)優先権主張番号	62/753,826		弁理士 飯田 貴敏
(32)優先日	平成30年10月31日(2018.10.31)	(74)代理人	100181641
(33)優先権主張国・地域又は機関	米国(US)		弁理士 石川 大輔
		(74)代理人	230113332

最終頁に続く

(54)【発明の名称】 エピジェネティック区画アッセイを較正するための方法、組成物およびシステム

(57)【特許請求の範囲】

【請求項1】

エピジェネティック状態に基づいてポリヌクレオチドの試料中の核酸分子の区画を評価する方法であって、前記エピジェネティック状態はエピジェネティック修飾のレベルまたは度合いであり、前記方法は、

a. 前記ポリヌクレオチドの試料の少なくとも1つのサブセット由来の核酸分子を複数の区画化セットに区画する工程と、

b. 前記複数の区画化セットから特定の標的ゲノム領域を含む分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成する工程であって、前記富化された分子のセットが、前記ポリヌクレオチドの試料由来の核酸分子の群を含み、前記ポリヌクレオチドの試料由来の核酸分子の群が、エピジェネティック状態が不変の少なくとも1つのヒトゲノム領域に対応する内在性対照分子のセットを含む、工程と、

c. 前記富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、複数のシーケンシングリードを作製する工程と、

d. 前記内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するために、前記シーケンシングリードのセットの少なくとも1つのサブセットを解析する工程と、

e. 前記1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較する工程と、

f. (i) 前記エピジェネティック対照核酸分子および/または前記内在性対照分子の

セットの前記1つまたは複数のエピジェネティック区画スコアが、対応するエピジェネティック区画カットオフの範囲内に入る場合には、前記方法が成功と分類するか、または(i i)前記エピジェネティック対照分子および/または前記内在性対照分子のセットの前記1つまたは複数のエピジェネティック区画スコアのうちの少なくとも1つが、対応するエピジェネティック区画カットオフの範囲外である場合には、前記方法が失敗と分類する工程と

を含む、方法。

【請求項2】

エピジェネティック対照核酸分子のセットを前記ポリヌクレオチドの試料中の前記核酸分子に添加し、これにより、スパイクイン試料を作製する工程をさらに含み、

10

工程 a . が、前記スパイクイン試料に対して実行され、

工程 b . において、前記富化された分子のセットが、エピジェネティック対照核酸分子の群をさらに含み、

工程 d . が、前記エピジェネティック対照核酸分子について1つまたは複数のエピジェネティック区画スコアをさらに生成する、請求項1に記載の方法。

【請求項3】

i ) 前記エピジェネティック対照核酸分子のセットが、エピジェネティック対照核酸分子の2つまたはそれ以上のサブセットを含み、前記エピジェネティック対照核酸分子の2つまたはそれ以上のサブセットのうちの1つのサブセットが、エピジェネティック修飾領域を含む複数のエピジェネティック対照核酸分子を含み、

20

i i ) 前記エピジェネティック対照核酸分子が識別子領域をさらに含み、

i i i ) 少なくとも1つのサブセット内の前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、エピジェネティック修飾を有する少なくとも1つのヌクレオチドを含み、および/または

i v ) 前記複数の区画化セットが、核酸分子のメチル化レベルに基づいて区画化された前記スパイクイン試料の核酸分子を含む、

請求項2に記載の方法。

【請求項4】

前記区画が、前記核酸分子を、エピジェネティック修飾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結合性物質に対する前記核酸分子の示差的な結合親和性に基づいて区画することを含み、請求項1~3のいずれか一項に記載の方法。

30

【請求項5】

前記複数の区画化セットのうちの1つの区画化セット内の核酸分子に、タグのセットを用いてタグ付けして、タグ付けされた核酸分子の集団を作製する工程をさらに含み、前記タグ付けされた核酸分子が、1つまたは複数のタグを含む、請求項1~4のいずれか一項に記載の方法。

【請求項6】

i ) 前記エピジェネティック修飾が、DNAメチル化であり、

i i ) 前記エピジェネティック修飾を有するヌクレオチドが、メチル化ヌクレオチドを含み、および/または

40

i i i ) 前記エピジェネティック状態が、前記核酸分子のメチル化レベルである、請求項1~5のいずれか一項に記載の方法。

【請求項7】

前記エピジェネティック区画スコアが、

i ) 区画化セット内の高メチル化エピジェネティック対照核酸分子および/または高メチル化対照分子の数、または

i i ) 区画化セット内の低メチル化エピジェネティック対照核酸分子および/または低メチル化対照分子の数

のフラクションまたはパーセンテージを含み、および/または

前記区画化セットが、

50

a) 高メチル化区画化セット、または  
 b) 低メチル化区画化セット  
 である、請求項 1 ~ 6 のいずれか一項に記載の方法。

【請求項 8】

前記エピジェネティック区画スコアが、

i) 0 CGスコア、

ii) hypoスコア、

iii) methyl-half、または

iv) methyl-5 である、請求項 1 ~ 7 のいずれか一項に記載の方法。

【請求項 9】

前記複数の富化された分子のシーケンシングが核酸シーケンサーによって実施される、請求項 1 ~ 8 のいずれか一項に記載の方法。

【請求項 10】

前記ポリヌクレオチドの試料が、

i) DNA の試料、RNA の試料、ポリヌクレオチドの試料、無細胞 DNA の試料、および無細胞 RNA の試料からなる群から選択され、および / または

ii) 無細胞 DNA である、請求項 1 ~ 9 のいずれか一項に記載の方法。

【請求項 11】

レポートを作成する工程をさらに含む、請求項 1 ~ 10 のいずれか一項に記載の方法。

【請求項 12】

システムであって、少なくとも 1 つの電子プロセッサによって実行されると、少なくとも、

a. 核酸シーケンサーによって生成された試料のシーケンシングリードのセットを得る工程であって、前記シーケンシングリードのセットが、前記試料のポリヌクレオチドから生成されたシーケンシングリードを含み、前記試料が、請求項 1 の工程 a および b に従ってシーケンシングする前に区画化され、かつ富化される、工程と、

b. 前記内在性対照分子の 1 つまたは複数のエピジェネティック区画スコアを生成するために、前記シーケンシングリードのセットの少なくとも 1 つのサブセットを解析する工程と、

c. 前記 1 つまたは複数のエピジェネティック区画スコアを 1 つまたは複数のエピジェネティック区画カットオフと比較する工程と、

d. 前記エピジェネティック区画スコアの比較に基づいて、区画方法の結果の状態を生成する工程であって、前記区画方法の結果の状態が、( i ) 前記内在性対照分子のセットの前記 1 つまたは複数のエピジェネティック区画スコアが、対応するエピジェネティック区画カットオフの範囲内に入る場合には成功に分類されるか、または ( ii ) 前記内在性対照分子の前記 1 つまたは複数のエピジェネティック区画スコアのうちの少なくとも 1 つが、前記対応するエピジェネティック区画カットオフの範囲外である場合には失敗に分類される、工程

を実施する、非一時的なコンピュータで実行可能な命令を含むコンピュータ可読媒体を含む、またはそれにアクセスすることができるコントローラを含む、システム。

【請求項 13】

前記試料がスパイクイン試料であり、請求項 2 の工程 a および b に従って区画化され、かつ富化され、

工程 b が、前記エピジェネティック対照核酸分子について 1 つまたは複数のエピジェネティック区画スコアをさらに生成し、および

工程 d の結果の状態が、( i ) 前記エピジェネティック対照核酸分子および / または前記内在性対照分子のセットの前記 1 つまたは複数のエピジェネティック区画スコアが、対応するエピジェネティック区画カットオフの範囲内に入る場合には成功に分類されるか、または ( ii ) 前記エピジェネティック対照分子および / または前記内在性対照分子の前記 1 つまたは複数のエピジェネティック区画スコアのうちの少なくとも 1 つが、対応する

10

20

30

40

50

エピジェネティック区画カットオフの範囲外である場合には失敗に分類される、請求項 1 2 に記載のシステム。

【請求項 1 4】

前記エピジェネティック区画スコアが、

i) 0 CGスコア、

ii) hypoスコア、

iii) methyl-half、または

iv) methyl-5 である、請求項 1 2 または 1 3 に記載のシステム。

【発明の詳細な説明】

【技術分野】

10

【0001】

関連出願への相互参照

本出願は、全体が参照により本明細書に組み込まれる、2018年10月31日出願の米国仮特許出願第 62 / 753 , 826 号に基づいて優先権を主張するものである。

【背景技術】

【0002】

背景

がんは、世界的に疾患の主要な原因である。毎年世界中で数千万人ががんと診断され、半数よりも多くが最終的にがんにより死亡する。多くの国で、がんは、心血管疾患に次いで2番目に多い死因として位置付けられている。多くのがんについて、早期検出が転帰の改善に関連付けられる。

20

【0003】

がんは、個体の正常な細胞における遺伝子変異の蓄積によって引き起こされる可能性があり、遺伝子変異の少なくとも一部は、細胞分裂の不適切な調節をもたらすものである。そのような変異は、一般に、コピー数変異 (CNV)、一塩基変異 (SNV)、遺伝子融合、挿入および/または欠失 (インデル)、5 - シトシンのメチル化 (5 - メチルシトシン) および DNA とクロマチンおよび転写因子の会合を含めたエピジェネティック変異を含む。

【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 の分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、非侵襲的であり、推測されるがん細胞を生検で識別することを伴わずに実施することができるという利点がある。しかし、そのようなリキッドバイオプシー検査は、体液中の核酸の量が非常に少なく、核酸が存在する形態が不均一である (例えば、RNA および DNA、一本鎖および二本鎖、ならびに複製後修飾およびヒストンなどのタンパク質との会合の種々の状態) という事実によって複雑なものになっている。

30

【発明の概要】

【課題を解決するための手段】

【0005】

40

要旨

ある態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、b) スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、c) 複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含む、ステップと、d) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作

50

製するステップと、e) エピジェネティック対照核酸分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、f) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を提供する。

【0006】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、b) スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、c) 複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、d) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、e) エピジェネティック対照核酸分子および内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、f) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を提供する。

【0007】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) ポリヌクレオチドの試料の少なくとも1つのサブセット由来の核酸分子を複数の区画化セットに区画するステップと、b) 複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、ポリヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、c) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、d) 内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するためにシーケンシングリードのセットのサブセットを解析するステップと、e) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を提供する。

【0008】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、b) スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、c) 区画分子の少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、d) エピジェネティック対照核酸分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、e) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を提供する。一部の実施形態では、方法は、シーケンシングステップの前に、複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含む、ステップをさらに含む。

【0009】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、b) スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、c) 区画分子の少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、d) エピジェネティック対照核酸分子および内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、e) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を提供する。一部の実施形態では、方法は、シーケンシングの前に、複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップをさらに含む。

10

**【0010】**

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) ポリヌクレオチドの試料の少なくとも1つのサブセット由来の分子を複数の区画化セットに区画するステップと、b) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、d) 内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するためにシーケンシングリードのセットのサブセットを解析するステップと、e) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を提供する。一部の実施形態では、方法は、シーケンシングの前に、複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、ポリヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップをさらに含む。

20

30

**【0011】**

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、b) スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、c) 複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含む、ステップと、d) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップとを含む方法を提供する。一部の実施形態では、方法は、e) エピジェネティック対照核酸分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、f) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとをさらに含む。

40

**【0012】**

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、b) スパイクイン試料の少なくとも1つのサブセットの核酸分

50

子を複数の区画化セットに区画するステップと、c) 複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、d) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップとを含む方法を提供する。一部の実施形態では、方法は、e) エピジェネティック対照核酸分子および内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、f) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとをさらに含む。

10

**【0013】**

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、a) ポリヌクレオチドの試料の少なくとも1つのサブセット由来の核酸分子を複数の区画化セットに区画するステップと、b) 複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、ポリヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、c) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップとを含む方法を提供する。一部の実施形態では、方法は、d) 内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するためにシーケンシングリードのセットのサブセットを解析するステップと、e) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとをさらに含む。

20

**【0014】**

一部の実施形態では、解析ステップは、区画化セットのうちの少なくとも1つにおける所与のエピジェネティック状態にあるエピジェネティック対照核酸分子および/または内在性対照分子の数/フラクションを推定することを含む。

**【0015】**

一部の実施形態では、方法は、複数の区画化セットのうちのある1つの区画化セット内の核酸分子に、タグのセットを用いてタグ付けして、タグ付けされた核酸分子の集団を作製するステップであって、タグ付けされた核酸分子が、1つまたは複数のタグを含む、ステップをさらに含む。一部の実施形態では、複数の区画化セットのうち第1の区画化セットに使用されるタグ(分子バーコード)のセットは、複数の区画化セットのうち第2の区画化セットに使用されるタグ(分子バーコード)のセットとは異なる。一部の実施形態では、タグのセットを核酸分子に、アダプターを核酸分子にライゲーションすることによって付着させ、ここで、アダプターは、1つまたは複数のタグ(分子バーコード)を含む。使用されるタグ(分子バーコード)配列は、区画化セットと関連し得る、例えば、1つの区画化セットに使用されるタグ(分子バーコード)は他の区画化セットには使用されない。

30

40

**【0016】**

一部の実施形態では、方法は、g) 区画方法を(i) エピジェネティック対照核酸分子および/または内在性対照分子のセットの1つまたは複数のエピジェネティック区画スコアのそれぞれが対応するエピジェネティック区画カットオフの範囲内に入る場合には、成功に分類するか、または、(ii) エピジェネティック対照分子および/または内在性対照分子のセットの1つまたは複数のエピジェネティック区画スコアのうちの少なくとも1つが対応するエピジェネティック区画カットオフの範囲外である場合には、失敗に分類することをさらに含む。

**【0017】**

50

一部の実施形態では、エピジェネティック対照核酸分子のセットは、エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットを含み、ここで、エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットのうちの1つのサブセットは、エピジェネティック修飾領域を含む複数のエピジェネティック対照核酸分子を含む。

【0018】

一部の実施形態では、複数の富化された分子のシーケンシングを核酸シーケンサーによって実施する。一部の実施形態では、核酸シーケンサーは次世代シーケンサーである。

【0019】

別の態様では、本開示は、エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットを含むエピジェネティック対照核酸分子のセットであって、エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットのうちの1つのサブセットが、エピジェネティック修飾領域を含む複数のエピジェネティック対照核酸分子を含む、エピジェネティック対照核酸分子のセットを提供する。

10

【0020】

別の態様では、本開示は、(i) エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットを含むエピジェネティック対照核酸分子のセットであって、エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットのうちの1つのサブセットが、エピジェネティック修飾領域を含む複数のエピジェネティック対照核酸分子を含む、エピジェネティック対照核酸分子のセットと、(ii) 対象由来のポリヌクレオチドの試料中の核酸分子のセットとを含む核酸の集団を提供する。

20

【0021】

一部の実施形態では、エピジェネティック対照核酸分子は、識別子領域をさらに含む。一部の実施形態では、識別子領域は、エピジェネティック対照核酸分子のエピジェネティック修飾領域の片側または両側に存在する。

【0022】

一部の実施形態では、少なくとも1つのサブセット内のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、エピジェネティック修飾を有するヌクレオチドを少なくとも1つ含む。一部の実施形態では、サブセットは、エピジェネティック修飾を有するヌクレオチドを同数有するエピジェネティック対照核酸分子を含む。一部の実施形態では、第1のサブセット内のエピジェネティック修飾を有するヌクレオチドの数は、第2のサブセット内のエピジェネティック修飾を有するヌクレオチドの数と異なる。一部の実施形態では、エピジェネティック修飾を有するヌクレオチドは、メチル化ヌクレオチドを含む。一部の実施形態では、メチル化ヌクレオチドは、5-メチルシトシンを含む。一部の実施形態では、メチル化ヌクレオチドは、5-ヒドロキシメチルシトシンを含む。

30

【0023】

一部の実施形態では、エピジェネティック対照核酸分子の識別子領域は、分子バーコードを含む。一部の実施形態では、識別子領域は、少なくとも1つのエピジェネティック状態バーコードをさらに含む。一部の実施形態では、識別子領域は、1つまたは複数のプライマー結合部位を含む。

【0024】

一部の実施形態では、2つまたはそれよりも多くのサブセット内の複数のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、同一の核酸配列を含む。

40

【0025】

一部の実施形態では、第1のサブセット内の複数のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、第2のサブセット内の複数のエピジェネティック対照核酸分子のエピジェネティック修飾領域の核酸配列とは区別可能な核酸配列を含む。

【0026】

一部の実施形態では、エピジェネティック修飾は、DNAメチル化である。

【0027】

一部の実施形態では、エピジェネティック対照核酸分子の各サブセットは、等モル濃度

50

で存在する。一部の実施形態では、エピジェネティック対照核酸分子の各サブセットは、非等モル濃度で存在する。

【0028】

一部の実施形態では、サブセットのうちの少なくとも1つにおけるエピジェネティック対照核酸分子内のメチル化ヌクレオチドの数は、0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、11個、少なくとも12個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも40個または少なくとも50個である。

【0029】

一部の実施形態では、エピジェネティック対照核酸分子は、ラムダファージDNA、ヒトゲノム領域またはその両方の組合せに対応する配列を含む。

10

【0030】

一部の実施形態では、エピジェネティック状態は、核酸分子のメチル化レベルである。一部の実施形態では、複数の区画化セットは、核酸分子のメチル化レベルに基づいて区画されたスパイクイン試料の核酸分子を含む。

【0031】

一部の実施形態では、エピジェネティック対照核酸分子のエピジェネティック修飾領域は、約160bpの長さで構成される。

【0032】

一部の実施形態では、エピジェネティック対照核酸分子のエピジェネティック修飾領域は、非ヒトゲノムに対応する核酸配列を含む。

20

【0033】

一部の実施形態では、ポリヌクレオチドの試料は、DNAの試料、RNAの試料、ポリヌクレオチドの試料、無細胞DNAの試料、および無細胞RNAの試料からなる群から選択される。一部の実施形態では、ポリヌクレオチドの試料は、DNAの試料、RNAの試料、ポリヌクレオチドの試料、無細胞DNAの試料、および無細胞RNAの試料からなる群から選択される。一部の実施形態では、無細胞DNAは、1ngから500ngの間である。

【0034】

一部の実施形態では、エピジェネティック対照核酸分子は、1フェムトモルから200フェムトモルの間である。

30

【0035】

一部の実施形態では、区画は、核酸分子を、エピジェネティック修飾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結合性物質に対する核酸分子の示差的な結合親和性に基づいて区画することを含む。

【0036】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子をエピジェネティック状態に基づいて区画する方法を評価するためのシステムであって、通信ネットワークを通じて、核酸シーケンサーによって生成されたスパイクイン試料のシーケンシングリードのセットを受信する通信インターフェースであって、シーケンシングリードのセットが、(i) 試料を起源とするポリヌクレオチドから生成されたシーケンシングリードの少なくともある1つの第1の集団であって、第1の集団由来のシーケンシングリードが、タグ配列および試料を起源とするポリヌクレオチドに由来する配列を含む、第1の集団；および(ii) エピジェネティック対照核酸分子から生成されたシーケンシングリードの少なくともある1つの第2の集団であって、第2の集団から生成されたシーケンシングリードが、エピジェネティック修飾領域および必要に応じて識別子領域を含む、第2の集団を含む、通信インターフェースと、通信インターフェースと通信するコンピュータであって、1つまたは複数のコンピュータプロセッサ、および、1つまたは複数のコンピュータプロセッサにより実行されると、通信ネットワークを通じて、核酸シーケンサーによるシーケンシングリードの第1の集団および第2の集団由来のシーケンシングリードのセットを受信す

40

50

るステップと、エピジェネティック対照核酸分子および/または内在性対照分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法をインプリメントする、機械により実行可能なコードを含むコンピュータ可読媒体を含む、コンピュータとを含むシステムを提供する。

【0037】

別の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)核酸シーケンサーによって生成されたスパイクイン試料のシーケンシングリードのセットを得るステップであって、スパイクイン試料が、試料のポリヌクレオチドおよびエピジェネティック対照核酸分子を含み、シーケンシングリードのセットが、(i)試料のポリヌクレオチドから生成されたシーケンシングリードの第1の集団および(ii)エピジェネティック対照核酸分子から生成されたシーケンシングリードの第2の集団を含む、ステップと、(b)エピジェネティック対照核酸分子および/または内在性対照分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(c)1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを実施する、非一時的なコンピュータで実行可能な命令を含むコンピュータ可読媒体を含む、またはそれにアクセスすることができるコントローラを含むシステムを提供する。

【0038】

別の態様では、本開示は、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)核酸シーケンサーによって生成された試料のシーケンシングリードのセットを得るステップであって、シーケンシングリードのセットが、試料のポリヌクレオチドから生成されたシーケンシングリードを含む、ステップと、(b)内在性対照分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(c)1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを実施する、非一時的なコンピュータで実行可能な命令を含むコンピュータ可読媒体を含む、またはそれにアクセスすることができるコントローラを含むシステムを提供する。

【0039】

一部の実施形態では、システムは、g)区画方法の結果の状態をエピジェネティック区画スコアの比較に基づいて生成するステップをさらに含む。一部の実施形態では、区画方法の結果の状態は、(i)エピジェネティック対照核酸分子および/または内在性対照分子のセットの1つまたは複数のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には成功に分類されるか、または、(ii)エピジェネティック対照核酸分子および/または内在性対照分子の1つまたは複数のエピジェネティック区画スコアのうちの少なくとも1つが対応するエピジェネティック区画カットオフの範囲外である場合には失敗に分類される。

【0040】

一部の実施形態では、エピジェネティック区画スコアは、区画化セット内の高メチル化エピジェネティック対照核酸分子および/または高メチル化対照分子の数のフラクションまたはパーセンテージを含む。一部の実施形態では、エピジェネティック区画スコアは、区画化セット内の低メチル化エピジェネティック対照核酸分子および/または低メチル化対照分子の数のフラクションまたはパーセンテージを含む。一部の実施形態では、区画化セットは、高メチル化区画化セットである。一部の実施形態では、区画化セットは、低メチル化区画化セットである。一部の実施形態では、エピジェネティック区画スコアは、0CGスコアである。一部の実施形態では、エピジェネティック区画スコアは、hypocスコアである。一部の実施形態では、エピジェネティック区画スコアは、methyl-h

a 1 f である。一部の実施形態では、エピジェネティック区画スコアは、m e t h y l - 5 である。

【 0 0 4 1 】

一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0 . 0 1 %、0 . 0 2 %、0 . 0 5 %、0 . 1 %、0 . 2 %、0 . 3 %、0 . 4 %、0 . 5 %、0 . 6 %、0 . 7 %、0 . 8 %、0 . 9 %、1 %、2 %、5 %、少なくとも5 % または少なくとも1 0 % である。一部の実施形態では、h y p oスコアについてのエピジェネティック区画カットオフは、0 . 1 %、0 . 5 %、1 %、2 %、3 %、4 %、5 %、7 % または少なくとも1 0 % である。一部の実施形態では、m e t h y l - h a l f についてのエピジェネティック区画カットオフは、5、1 0、1 5、2 0、2 5、3 0、3 5 または4 0 m C G である。一部の実施形態では、m e t h y l - 5 についてのエピジェネティック区画カットオフは、5、1 0、2 0、3 0、4 0 または5 0 m C G である。

10

【 0 0 4 2 】

本開示の例示的な実施形態のみが示され、説明されている以下の詳細な説明から、本開示の追加的な態様および利点が当業者には容易に明らかになる。理解される通り、本開示は、他のおよび異なる実施形態も可能であり、そのいくつかの詳細は、種々の明白な点で改変することができ、全て本開示から逸脱しない。したがって、図および説明は、例示的な性質のものであり、限定を加えるものではないとみなされるべきである。

【 0 0 4 3 】

一部の実施形態では、本明細書に開示されるシステムおよび/または方法の結果を、レポートを生成するための入力として使用する。レポートは、紙または電子形式であり得る。例えば、本明細書に開示される方法またはシステムによって決定される核酸分子の区画に関する情報および/または核酸分子の区画から導き出された情報をそのようなレポートに提示することができる。本明細書に開示される方法またはシステムは、レポートを、試料が由来する対象または健康管理実践者などの第三者に伝達するステップをさらに含む得る。

20

【 0 0 4 4 】

本明細書に開示される方法の種々のステップ、または本明細書に開示されるシステムによって実施されるステップは、同時に行うこともでき、違う時間に行うこともでき、かつ/または、同じ地理的位置、例えば、国で行うこともでき、異なる地理的位置、例えば、国で行うこともできる。本明細書に開示される方法の種々のステップは、同じ人が行うこともでき、異なる人が行うこともできる。

30

【 0 0 4 5 】

本明細書に組み込まれ、その一部を構成する付属図は、ある特定の実施形態を例示し、明細書と共に、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。本明細書に提示される説明は、例として含まれており、限定するものではない付属図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、図全体を通して同様の参照数字により同様の構成成分が識別されることが理解されよう。図面の一部または全部が、例示目的で略図であり得、必ずしも示される要素の実際の相対的なサイズまた位置が示されているわけではないことも理解されよう。

40

【 図面の簡単な説明 】

【 0 0 4 6 】

【 図 1 】 図 1 A および図 1 B は、二本鎖 D N A における完全にメチル化された C p G 対 ( 図 1 A ) および半分がメチル化された C p G 対 ( 図 1 B ) の概略図である。

【 0 0 4 7 】

【 図 2 】 図 2 は、本開示のある実施形態に従ってポリヌクレオチドの試料の区画を査定するための方法を表すフローチャートである。

【 0 0 4 8 】

【 図 3 】 図 3 は、本開示のある実施形態に従ってポリヌクレオチドの試料の区画を査定す

50

るための方法を表すフローチャートである。

【0049】

【図4】図4は、本開示のある実施形態に従ってポリヌクレオチドの試料の区画を査定するための方法を表すフローチャートである。

【0050】

【図5】図5は、本開示の一部の実施形態での使用に適したエピジェネティック対照核酸分子の略図である。

【0051】

【図6】図6は、本開示の一部の実施形態での使用に適したエピジェネティック対照核酸分子の略図である。

10

【0052】

【図7】図7は、本開示の一部の実施形態での使用に適したエピジェネティック対照核酸分子の略図である。

【0053】

【図8】図8は、本開示の一部の実施形態での使用に適したシステムの例の概略図である。

【0054】

【図9】図9A、図9Bおよび図9Cは、高区画化セット内(図9A)、中区画化セット内(図9B)および低区画化セット内(図9C)のサブセット1、2、3、4、5、および6に属するエピジェネティック対照核酸分子のエピジェネティック区画スコアのグラフ表示である。

20

【0055】

【図10】図10Aおよび図10Bは、高区画化セット内(図10A)および低区画化セット内(図10B)の試料1の高メチル化対照分子のフラクションのグラフ表示である。

【0056】

【図11】図11Aおよび図11Bは、高区画化セット内(図11A)および低区画化セット内(図11B)の試料2の高メチル化対照分子のフラクションのグラフ表示である。

【0057】

#### 定義

本開示の理解を容易にするために、最初に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追加的な定義が本明細書を通して記載されている場合がある。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の定義と相反する場合には、用語の意味を理解するために本出願に記載の定義が使用されるべきである。

30

【0058】

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、「1つの(a)」、「1つの(an)」、および「その(the)」という単数形は、文脈によりそうでないことが明確に規定されない限り、複数の参照対象を含む。したがって、例えば、「1つの(a)方法」への言及は、本明細書に記載の型および/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。

【0059】

本明細書において使用される用語法は、特定の実施形態を説明するためだけのものであり、限定的なものではないことも理解される。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では、以下の用語法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。

40

【0060】

約：本明細書で使用される場合、「約」または「およそ」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、規定された参照値または要素と同様の値またはエレメントを指す。ある特定の実施形態では、「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、規定された参照値また

50

は要素に対して、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、またはそれ未満以内に入る値または要素の範囲を指す（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）。

【0061】

アダプター：本明細書で使用される場合、「アダプター」は、一般には少なくとも部分的に二本鎖であり、所与の試料核酸分子のいずれかの末端または両末端に付着させる短い核酸（例えば、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ヌクレオチド未満の長さ）を指す。アダプターは、両末端にアダプターが隣接する核酸分子の増幅を可能にする核酸プライマー結合部位、および/または、種々の次世代シーケンシング（NGS）アプリケーションなどのシーケンシングアプリケーションのためのプライマー結合部位を含めたシーケンシングプライマー結合部位を含み得る。アダプターは、フローセル支持体などに付着させたオリゴヌクレオチドなどの捕捉用プローブの結合部位も含み得る。アダプターは、本明細書に記載の核酸タグも含み得る。核酸タグは、一般には、増幅プライマーおよびシーケンシング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配置される。同じ配列または異なる配列のアダプターを核酸分子のそれぞれの末端に連結することができる。一部の実施形態では、核酸タグが異なる以外は同じ配列のアダプターを核酸分子のそれぞれの末端に連結する。一部の実施形態では、アダプターは、一方の末端が、核酸分子との接合のために本明細書に記載の通り平滑末端化されたかまたは尾部付加されたY形状アダプターであり、ここで、核酸分子も同様に1つまたは複数の相補的なヌクレオチドで平滑末端化または尾部付加されている。さらに他の実施形態例では、アダプターは、解析される核酸分子との接合のための平滑末端または尾部付加末端を含む鐘形アダプターである。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。

【0062】

増幅させる：本明細書で使用される場合、「増幅させる」または「増幅」は、核酸に関しては、ポリヌクレオチドまたはポリヌクレオチドの一部の多数のコピーを、一般に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）から出発して産生させることを指し、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、種々の化学的および酵素的プロセスを包含する。増幅は、これだけに限定されないが、ポリメラーゼ連鎖反応（PCR）を含む。

【0063】

バーコード：本明細書で使用される場合、「バーコード」または「分子バーコード」は、核酸に関しては、分子識別子としての機能を果し得る配列を含む核酸分子を指す。例えば、一般には、次世代シーケンシング（NGS）ライブラリー調製の際に個々の「バーコード」配列をDNA断片に付加し、したがって、最終的なデータ解析前にシーケンシングリードを識別し、選別することができる。

【0064】

がんの型：本明細書で使用される場合、「がんの型」は、例えば病理組織検査によって定義されるがんの型または亜型を指す。がんの型は、任意の従来基準によって、例えば、所与の組織における発生に基づいて（例えば、血液がん、中枢神経系（CNS）、脳がん、肺がん（小細胞および非小細胞）、皮膚がん、鼻のがん、咽頭がん、肝がん、骨がん、リンパ腫、膵がん、腸がん（bowel cancer）、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、腸がん（intestinal cancer）、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、異種がん、同種がん）、原発不明などに基づいて、および/または同じ細胞系列のものに基づいて（

10

20

30

40

50

例えば、癌腫、肉腫、リンパ腫、胆管細胞癌、白血病、中皮腫、黒色腫、または神経膠芽腫)、および/または、これだけに限定されないが、Her 2、CA15-3、CA19-9、CA-125、CEA、AFP、PSA、HCG、ホルモン受容体およびNMP-22などのがんマーカーを示すがんに基づい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)および原発性であるか続発性であるかによって分類することもできる。

【0065】

無細胞核酸：本明細書で使用される場合、「無細胞核酸」とは、細胞内に含有されないもしくは他のやり方で結合していない核酸、または一部の実施形態では、インタクトな細胞の除去後に試料中に残っている核酸を指す。無細胞核酸は、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)を供給源とする封入されていない核酸全てを含み得る。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(crRNA)、tRNA、rRNA、低分子核小体RNA(snRNA)、piwi相互作用性RNA(piRNA)、長鎖ノンコーディングRNA鎖(長鎖ncRNA)、および/またはこれらのいずれかの断片を含めた、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から体液中に放出される、例えば、循環腫瘍DNA(ctDNA)である。他の無細胞核酸は健康な細胞から放出される。ctDNAは、封入されていない腫瘍由来断片化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修飾を有し得、例えば、無細胞核酸は、アセチル化、5-メチル化、および/またはヒドロキシメチル化されたものであり得る。

【0066】

細胞内核酸：本明細書で使用される場合、「細胞内核酸」とは、後で所与の解析プロセスの一部として除去されたとしても(例えば、細胞溶解によって)、少なくとも対象から試料を取得または採取した時点で、それが由来する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。

【0067】

カバレッジ：本明細書で使用される場合、「カバレッジ」、「総分子計数」または「総対立遺伝子計数」という用語は互換的に使用される。これらの用語は、所与の試料中の特定のゲノム位置におけるDNA分子の総数を指す。

【0068】

CpG対：本明細書で使用される場合、「CpG対」という用語は、二本鎖DNA分子のセンス鎖上のジヌクレオチドCpG(シトシン-リン酸-グアニン(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続く))ジヌクレオチドとアンチセンス鎖上のその相補的CpGを指す(図1に示されている)。

【0069】

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」は、糖部分の2'位に水素基を有する天然のまたは修飾されたヌクレオチドを指す。DNAは、一般には、4つの型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)で構成されるヌクレオチドの鎖を含む。本明細書で使用される場合、「リボ核酸」または「RNA」は、糖部分の2'位にヒドロキシル基を有する天然のまたは修飾されたヌクレオチドを指す。RNAは、一般には、4つの型のヌクレオチド塩基;A、ウラシル(U)、G、およびCで構成されるヌクレオチドの鎖を含む。本明細書で使用される場合、「ヌクレオチド」という用語は、天然のヌクレオチドまたは修飾されたヌクレオチドを指す。ある特定のヌクレオチドの対は、互いに相補的に特異的に結合する(相補的塩基対合と称される)。DNAでは、アデニン(A)がチミン(T)と対合し、シトシン(C)がグアニン(G)と対合する。RNAでは、アデニン(A)がウラシル(U)と対合し、シトシン(C)がグアニン(G)と対合する。第

10

20

30

40

50

1の核酸鎖が、第1の鎖のヌクレオチドと相補的なヌクレオチドで構成される第2の核酸鎖と結合する場合、これらの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、「核酸シーケンシングデータ」、「核酸シーケンシング情報」、「配列情報」、「核酸配列」、「ヌクレオチド配列」、「ゲノム配列」、「遺伝子配列」または「断片配列」または「核酸シーケンシングリード」とは、DNAまたはRNAなどの核酸の分子（例えば、全ゲノムの、全トランスクリプトーム、エキソーム、オリゴヌクレオチド、ポリヌクレオチド、または断片）内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。本教示では、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、直接または間接ヌクレオチド識別システム、パイロシーケンシング、イオン-またはpHに基づく検出システム、および電子シグネチャーに基づくシステムを含めた全ての利用可能な様々な技法、プラットフォームまたは技術を使用して得られる配列情報が意図されていることが理解されるべきである。

10

#### 【0070】

内在性対照分子：本明細書で使用される場合、「内在性対照分子」は、エピジェネティック状態が不変の少なくとも1つのヒトゲノム領域に対応する、ポリヌクレオチドの試料中の核酸分子を指す。一部の実施形態では、内在性対照分子は、組織、対象およびがんにかかわらず一貫して高度にメチル化された状態または低度にメチル化された状態であり得る。一部の実施形態では、一貫して高度にメチル化された状態である領域を有するヒトゲノム領域に対応する内在性対照分子は、「高メチル化対照分子」と称することができる。一部の実施形態では、一貫して低度にメチル化された状態である領域を有するヒトゲノム領域に対応する内在性対照分子は、「低メチル化対照分子」と称することができる。

20

#### 【0071】

エピジェネティック対照核酸分子：本明細書で使用される場合、「エピジェネティック対照核酸分子」とは、試料の区画をエピジェネティック修飾に基づいて評価するためにポリヌクレオチドの試料に添加される核酸分子のセットを指す。例えば、エピジェネティック修飾は、DNAメチル化であり得、エピジェネティック対照核酸分子は、種々のノ区別可能なレベルのメチル化を有し得る。一部の実施形態では、エピジェネティック対照核酸分子は、エピジェネティック修飾領域および必要に応じて識別子領域を含む。一部の実施形態では、エピジェネティック対照核酸分子は、エピジェネティック修飾領域および識別子領域を含む。エピジェネティック対照核酸分子は、合成オリゴヌクレオチドであり得る。一部の実施形態では、エピジェネティック対照核酸分子は、天然には存在しない核酸配列を有し得る。一部の実施形態では、エピジェネティック対照核酸分子は、天然に存在する核酸配列を有し得る。一部の実施形態では、エピジェネティック対照核酸分子は、非ヒトゲノムに対応する核酸配列を有し得る。非限定的な例として、これらの分子は、(i)ラムダファージDNAまたはヒトゲノムの領域に対応する配列、(ii)天然には存在しない配列、および/または(iii)(i)と(ii)の組合せのいずれかを有し得る。

30

#### 【0072】

エピジェネティック修飾：本明細書で使用される場合、「エピジェネティック修飾」は、核酸分子内のヌクレオチド（単数または複数）の塩基の修飾を意味する。修飾は、ヌクレオチドの塩基の化学修飾であり得る。一部のケースでは、修飾は、ヌクレオチドの塩基のメチル化であり得る。例えば、修飾は、5-メチルシトシンが生じるシトシンのメチル化であり得る。

40

#### 【0073】

エピジェネティック修飾領域：本明細書で使用される場合、「エピジェネティック修飾領域」とは、エピジェネティック対照核酸分子のエピジェネティック修飾のレベル/度合いを表すエピジェネティック対照核酸分子の領域を指す。一部の実施形態では、エピジェネティック修飾領域は、エピジェネティック修飾を有するヌクレオチドを含み得る。一部の実施形態では、エピジェネティック修飾は、DNAメチル化である。これらの実施形態

50

では、エピジェネティック対照核酸分子のエピジェネティック修飾領域は、メチル化ヌクレオチドを有し得る。エピジェネティック修飾領域内のメチル化ヌクレオチドの数は、エピジェネティック対照核酸分子の間で変動し得る。一部の実施形態では、エピジェネティック対照核酸分子は、エピジェネティック修飾領域内にメチル化ヌクレオチドを0個、1個、2個、3個、4個、5個、6個、7個、8個、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30個、少なくとも40個または少なくとも50個有し得る。エピジェネティック対照核酸分子は、エピジェネティック修飾領域内のエピジェネティック修飾を有するヌクレオチドの数に基づいてサブセットに群分けすることができる。異なるサブセットの中でのエピジェネティック修飾領域の長さは、同じ、例えば、およそ160bpであり得る。サブセット間でエピジェネティック修飾領域の長さは異なり得る。例えば、エピジェネティック対照核酸分子を、エピジェネティック修飾領域内のメチル化ヌクレオチドの数に基づいて3つのサブセット(サブセットA、BおよびC)に群分けすることができる。サブセットA、BおよびCは、それぞれ、エピジェネティック修飾領域内にメチル化ヌクレオチドを5個、10個および15個有するエピジェネティック対照核酸分子を有し得、サブセットA、BおよびCにおけるエピジェネティック修飾領域の長さは、同じ(例えば、160bp)であり得る、または異なり得る。サブセットA、BおよびCについて、それぞれ100bp、150bpおよび200bp。

10

#### 【0074】

エピジェネティック区画スコア：本明細書で使用される場合、「エピジェネティック区画スコア」は、所与の区画化セット内の特定のエピジェネティック状態に属する核酸分子の区画を表す数値を指す。一部の実施形態では、エピジェネティック状態に属する核酸分子のエピジェネティック区画スコアを各区画化セットについて決定する。例えば、特定のエピジェネティック状態に属するエピジェネティック対照核酸分子および/または内在性対照分子のエピジェネティック区画スコアを決定することができる。エピジェネティック区画スコアは、特定のエピジェネティック状態に属する核酸分子の数(または統計的に推定される数)の尺度であり得る。エピジェネティック区画スコアは、フラクションまたはパーセンテージによるものであり得る。エピジェネティック区画スコアは、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属するエピジェネティック対照核酸分子および/または内在性対照分子の数と、他の残りの区画化セット(単数または複数)内に存在するそのエピジェネティック状態に属するエピジェネティック対照核酸分子および/または内在性対照分子の数の比の尺度であり得る。一部の実施形態では、エピジェネティック区画スコアは、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属するエピジェネティック対照核酸分子および/または内在性対照分子の数の、全ての区画化セット内のそのエピジェネティック状態に属するエピジェネティック対照核酸分子および/または内在性対照分子の総数に対するフラクションまたはパーセンテージであり得る。一部の実施形態では、エピジェネティック区画スコアを、区画化セットのそれぞれにおけるエピジェネティック対照核酸分子および/または内在性対照分子の各エピジェネティック状態に関して決定する。一部の実施形態では、エピジェネティック区画スコアを、1つまたは複数の区画化セット内の1つまたは複数の特定のエピジェネティック状態を有するエピジェネティック対照核酸分子および/または内在性対照分子に関して決定する。一部の実施形態では、エピジェネティック区画スコアを、特定の区画化セット内の特定のエピジェネティック状態を有するエピジェネティック対照核酸分子および/または内在性対照分子に関して決定する。

20

30

40

#### 【0075】

一部の実施形態では、エピジェネティック区画スコアは、CGを有さない(「ゼロ」CG)分子が高区画化セットに区画される有効度を対象とし得る。このスコアは、0CGスコアと称することができる。一部の実施形態では、0CGスコアを高区画化セット内のCGを有さない分子のフラクションまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、高区画化セット内の、以下のうちの少なくとも1つを有するエピジェネティック対照核酸分子のフラクションおよび/または高

50

メチル化対照分子のフラクシオンに対する尺度であり得る：

( i ) 1つのメチルC G ( エピジェネティック区画スコアを1 C Gスコアと称することができる )、

( i i ) 2つのメチルC G ( エピジェネティック区画スコアを2 C Gスコアと称することができる )、

( i i i ) 3つのメチルC G ( エピジェネティック区画スコアを3 C Gスコアと称することができる )、

( i v ) 4つのメチルC G ( エピジェネティック区画スコアを4 C Gスコアと称することができる )、および

( v ) 5つのメチルC G ( エピジェネティック区画スコアを5 C Gスコアと称することができる )。

10

【 0 0 7 6 】

一部の実施形態では、エピジェネティック区画スコアは、低メチル化対照分子または低メチル化エピジェネティック対照核酸分子が高メチル化区画化セットに区画される有効度を対象とし得る。このスコアは、h y p oスコアと称することができる。一部の実施形態では、h y p oスコアを高メチル化区画化セット内の低メチル化対照分子または低メチル化エピジェネティック対照核酸分子のフラクシオンまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、低メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子が指定の値、例えば5%未満になるのに必要なメチル化C Gの数の尺度であり得る。低メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子が5%であることを使用する。すなわち、エピジェネティック区画スコアが、低メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子が5%未満になるのに必要なメチル化C Gの数の尺度である例では、このスコアを、利便性のためにm e t h y l - 5と称することができる。一部の実施形態では、エピジェネティック区画スコアは、高メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子が少なくとも指定の値、例えば50%になるのに必要なメチル化C Gの数の尺度であり得る。高メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子が50%であることを使用する。すなわち、エピジェネティック区画スコアが、高メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子の少なくとも50%になるのに必要なメチル化C Gの数の尺度である例では、このスコアをm e t h y l - h a l fと称することができる。0%から100%までの広範な異なる値( ちょうど5%および50%ではなく )を異なる実施形態において使用することができ、指定の値を指す対応する異なる利便性のよい名称を使用することができる。

20

30

【 0 0 7 7 】

例えば、エピジェネティック対照核酸分子の3つのサブセット( サブセットA、BおよびC )を使用し、各サブセットはメチル化ヌクレオチドの数が異なる。これらの3つのサブセット内のエピジェネティック対照核酸分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P 1、P 2およびP 3に区画することができる。各サブセットについて、エピジェネティック区画スコアを、区画化セットのそれぞれ( P 1、P 2およびP 3 )について決定する。すなわち、サブセットAに属するエピジェネティック対照核酸分子は、3つのエピジェネティック区画スコア、つまり3つの区画化セット、P 1、P 2およびP 3のそれぞれについて1つのエピジェネティック区画スコアを有する。同様に、サブセットBおよびCのそれぞれは、3つのエピジェネティック区画スコア、つまり3つの区画化セットP 1、P 2およびP 3のそれぞれについて1つのエピジェネティック区画スコアを有する。内在性対照分子についてもエピジェネティック区画スコアを決定することができる。

40

【 0 0 7 8 】

別の実施形態では、エピジェネティック対照核酸分子の3つのサブセット( サブセット

50

A、BおよびC)を使用し、各サブセットはメチル化ヌクレオチドの数が異なる(すなわち、各サブセットは異なるエピジェネティック状態を有する)。これらの3つのサブセット内のエピジェネティック対照核酸分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。この実施形態では、P1区画化セット内のサブセットA分子についてのみエピジェネティックスコアを決定する。このエピジェネティックスコアは、P1区画化セット内のサブセットA分子の、サブセットA分子(P1、P2およびP3つの区画化セット内の)の総数に対するフラクションまたはパーセンテージの尺度であり得る。

【0079】

エピジェネティック区画カットオフ：本明細書で使用される場合、「エピジェネティック区画カットオフ」は、特定の区画化セット内の特定のエピジェネティック状態に属する核酸分子の区画を査定するために使用される所定のカットオフ値またはカットオフを指す。一部の実施形態では、エピジェネティック区画カットオフは、社内試料データセットを解析することによって決定する。各区画化セットは、エピジェネティック状態に属する核酸分子についてのエピジェネティック区画カットオフを有し得る。1つまたは複数のエピジェネティック状態に属するエピジェネティック対照核酸分子の1つまたは複数のエピジェネティック区画スコア(区画を査定するために使用されるもの)がそれらの対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法は成功である。そうでなければ、区画方法は失敗である。エピジェネティック区画カットオフは、核酸分子および区画化セットのエピジェネティック状態によって異なる、すなわち、各エピジェネティック状態がそれ自体のエピジェネティック区画カットオフを有し、区画化セットのいずれもがそのエピジェネティック状態についての別々のエピジェネティック区画カットオフを有する。カットオフは、パーセンテージまたはフラクションによるものであり得、カットオフは、特定のカットオフ値ではなくカットオフ範囲であり得る。例えば、特定のエピジェネティック状態に属するエピジェネティック対照核酸分子についてのエピジェネティック区画カットオフは、区画化セットP1、P2およびP3について、それぞれ70%から79%の間、10%から15%の間および5%未満であり得る。そのエピジェネティック状態に属するエピジェネティック対照核酸分子のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法は成功である。

【0080】

エピジェネティック状態：本明細書で使用される場合、「エピジェネティック状態」は、核酸分子のエピジェネティック修飾のレベル/度合いを指す。例えば、エピジェネティック修飾がDNAメチル化(またはヒドロキシメチル化)である場合には、エピジェネティック状態は、DNA塩基(例えば、シトシン)におけるメチル化の存在もしくは非存在、または核酸配列内のメチル化の度合い(例えば、高度にメチル化された、低度にメチル化された、中程度にメチル化されたもしくはメチル化されていない核酸分子)を指し得る。エピジェネティック状態は、エピジェネティック修飾を有するヌクレオチドの数も指し得る。例えば、エピジェネティック修飾がDNAメチル化である場合には、エピジェネティック状態は、核酸分子のメチル化ヌクレオチドの数を指し得る。

【0081】

エピジェネティック状態バーコード：本明細書で使用される場合、「エピジェネティック状態バーコード」は、エピジェネティック対照核酸分子のエピジェネティック状態を識別するために使用される核酸配列を指す。識別は、特定の1つまたは複数のエピジェネティック状態バーコードとエピジェネティック対照核酸分子のエピジェネティック状態の間に所定の相関を持たせることによって実現することができる。これは、エピジェネティック対照核酸分子のエピジェネティック修飾領域内のエピジェネティック修飾を有するヌクレオチドの数を指し得る。一部の実施形態では、エピジェネティック対照核酸分子の識別子領域は、少なくとも1つのエピジェネティック状態バーコードを含む。例えば、エピジェネティック修飾がDNAメチル化であり、エピジェネティック対照核酸分子のサブセットがメチル化ヌクレオチドを5つ有する場合には、そのサブセットに入るエピジェネティ

10

20

30

40

50

ック対照核酸分子の全てが同じエピジェネティック状態バーコードを有することになる。一部の実施形態では、エピジェネティック状態バーコードを使用して、エピジェネティック対照核酸分子のエピジェネティック修飾領域のエピジェネティック修飾のレベル/度合いを識別することができる。エピジェネティック対照核酸分子をエピジェネティック修飾領域内のシトシンまたはCpGヌクレオチドの数に基づいてサブセットに群分けすることができる。一部の実施形態では、各サブセット内でメチル化のレベルが変動し得（例えば、高度にメチル化された、中程度にメチル化された、低度にメチル化されたまたはメチル化されていない）、メチル化の各レベルは別々のエピジェネティック状態バーコードを有し得る。例えば、サブセットA内で、低度にメチル化されたエピジェネティック対照核酸分子は全てがエピジェネティック状態バーコード、例えばESB1を有し、高度にメチル化された状態であるエピジェネティック対照核酸分子は全てが別のエピジェネティック状態バーコード、例えばESB3を有することになる。本実施例では、エピジェネティック状態バーコードを使用して、メチル化のレベル/度合いを識別する。

10

#### 【0082】

エピジェネティック状態が不変のヒトゲノム領域：本明細書で使用される場合、「エピジェネティック状態が不変のヒトゲノム領域」は、特定のエピジェネティック状態を有するヒトゲノム内の領域であって、その領域のエピジェネティック状態が多くの場合に変動/変化せず、常に同じままであるか、または異なる対象および/または異なる疾患の型/病期の場合でも一貫したままである、領域を指す。例えば、エピジェネティック状態が不変のヒトゲノム領域は、大部分がメチル化されているかまたは大部分がメチル化されていない可能性がある。

20

#### 【0083】

識別子領域：本明細書で使用される場合、「識別子領域」とは、エピジェネティック対照核酸分子を他のエピジェネティック対照核酸分子と区別することに使用される、エピジェネティック対照核酸分子の領域を指す。識別子領域は、分子バーコードおよび/またはエピジェネティック状態バーコードを有し得る。識別子領域は、エピジェネティック修飾領域の片側または両側に存在し得る。分子バーコードは、エピジェネティック対照核酸分子の識別子としての機能を果たし、一方、エピジェネティック状態バーコードは、エピジェネティック対照核酸分子のエピジェネティック状態の識別子としての機能を果たす。識別子領域は、1つまたは複数のプライマーの結合を容易にする追加的な領域（プライマー結合部位）を有し得る。

30

#### 【0084】

突然変異体対立遺伝子計数：本明細書で使用される場合、「突然変異体対立遺伝子計数」という用語は、特定のゲノム遺伝子座に突然変異体対立遺伝子を有するDNA分子の数を指す。

#### 【0085】

突然変異体対立遺伝子画分：本明細書で使用される場合、「突然変異体対立遺伝子画分」、「突然変異量」または「MAF」は、所与の試料中の所与のゲノム位置/遺伝子座に對立遺伝子の変更または突然変異を有する核酸分子の画分を指す。MAFは、一般に、フラクションまたはパーセンテージで表される。例えば、体細胞変異体のMAFは、0.15未満であり得る。

40

#### 【0086】

突然変異：本明細書で使用される場合、「突然変異」は、既知の参照配列からの変異を指し、例えば、一塩基変異体(SNV)、および挿入または欠失(インデル)などの突然変異を含む。突然変異は、生殖細胞系列突然変異または体細胞突然変異であり得る。一部の実施形態では、比較を行うための参照配列は、試験試料を提供する対象の種の野生型ゲノム配列、一般にはヒトゲノムである。

#### 【0087】

突然変異コール器：本明細書で使用される場合、「突然変異コール器」とは、試験試料データ（例えば、対象から得た配列情報）における突然変異を識別するために使用される

50

アルゴリズム（一般には、ソフトウェアに組み入れられるか、または他のやり方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。

【0088】

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。これらの用語は、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、潜在的に悪性、または悪性であり得る。悪性腫瘍は、がんまたはがん性腫瘍と称される。

【0089】

次世代シーケンシング：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、従来のサンガー法に基づく手法およびキャピラリー電気泳動に基づく手法と比較して増大したスループットを有し、例えば、何十万もの比較的小さな配列リードを一度に生成する能力を有するシーケンシング技術を指す。次世代シーケンシング技法の一部の例としては、これだけに限定されないが、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。一部の実施形態では、次世代シーケンシングは、単一分子をシーケンシングすることができる機器の使用を含む。

【0090】

核酸タグ：本明細書で使用される場合、「核酸タグ」は、核酸を、型が異なる、または異なる処理に供された、異なる試料（例えば、試料インデックスを表す）または同じ試料中の異なる核酸分子（例えば、分子バーコードを表す）と区別するために使用される短い核酸（例えば、約500ヌクレオチド未満、約100ヌクレオチド、約50ヌクレオチド、または約10ヌクレオチドの長さ）を指す。核酸タグは、所定の、固定された、ランダムでない、ランダムな、またはセミランダムなオリゴヌクレオチド配列を含む。そのような核酸タグを使用して、異なる核酸分子または異なる核酸試料またはサブサンプルを標識することができる。核酸タグは、一本鎖、二本鎖、または少なくとも部分的に二本鎖であり得る。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、1つまたは複数の平滑末端を有する二本鎖分子、5'もしくは3'一本鎖領域（例えば、突出）を含む二本鎖分子、および/または所与の分子内の他の位置に1つもしくは複数の他の一本鎖領域を含む二本鎖分子も含み得る。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）の一方の末端に付着させることもでき、両方の末端に付着させることもできる。核酸タグを復号して、所与の核酸の起源試料、形態、または処理などの情報を明らかにすることができる。例えば、核酸タグを使用して、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多数の試料のプールおよび/または並行処理を可能にすることもでき、この場合、その後、核酸タグを検出すること（例えば、読み取ること）によって核酸をデコンボリューションする。核酸タグは、識別子（例えば、分子識別子、試料識別子）と称することもできる。それに加えて、またはその代わりに、核酸タグを分子識別子として使用することができる（例えば、同じ試料またはサブサンプル中の異なる分子または親分子が異なるアンプリコンを区別するために）。これは、例えば、所与の試料中の異なる核酸分子に一意的にタグ付けすること、またはそのような分子に非一意的にタグ付けすることを含む。非一意的にタグ付けする適用の場合では、限られた数のタグ（すなわち、分子バーコード）を使用して各核酸分子にタグ付けすることができ、したがって、異なる分子をそれらの内在性配列情報（例えば、選択された参照ゲノムにそれらが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、および/または配列の長さ）と少なくとも1つの分子バーコードの組合せに基づいて区別することができる。一般には、任意の2つの分子が同じ内在性配列情報（例えば、開始および/または終止位置、配列の一方または両方の末端の部分配列、および/または長さ）を有し、同じ分子バーコードも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%未満の可能性）、十分な数の異なる分子バーコードを使用する。

【0091】

10

20

30

40

50

区画：本明細書で使用される場合、「区画」および「エピジェネティック区画」は、互換的に使用される。この用語は、核酸分子を、核酸分子の特徴（例えば、エピジェネティック修飾のレベル/度合い）に基づいて分離または分別することを指す。区画は、分子の物理的区画であり得る。区画は、核酸分子をエピジェネティック修飾のレベル（すなわち、エピジェネティック状態）に基づいて群またはセットに分離することを伴い得る。例えば、核酸分子を核酸分子のメチル化のレベルに基づいて区画することができる。一部の実施形態では、区画のために使用される方法およびシステムは、その全体が参照によって組み込まれる P C T 特許出願第 P C T / U S 2 0 1 7 / 0 6 8 3 2 9 号において見ることができる。

#### 【 0 0 9 2 】

区画化セット：本明細書で使用される場合、「区画化セット」は、核酸分子の結合性物質に対する示差的な結合親和性に基づいてセット/群に区画された核酸分子のセットを指す。結合性物質は、エピジェネティック修飾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修飾がメチル化である場合、結合性物質は、メチル結合性ドメイン（ M B D ）タンパク質であり得る。一部の実施形態では、区画化セットは、特定のエピジェネティック修飾のレベル/度合い（すなわち、エピジェネティック状態）に属する核酸分子を含み得る。例えば、核酸分子を3つのセット：高メチル化区画化セットまたは高区画化セットと称することができる、高度にメチル化された核酸分子（または高メチル化核酸分子）の1つのセット、低メチル化区画化セットまたは低区画化セットと称することができる、低度にメチル化された核酸分子（または低メチル化核酸分子）の別のセット、および、中程度にメチル化された区画化セットまたは中区画化セットと称することができる中程度にメチル化された核酸分子の第3のセットに区画することができる。別の例では、核酸分子を、エピジェネティック修飾を有するヌクレオチドの数に基づいて区画することができ、1つの区画化セットは、メチル化ヌクレオチドを9個有する核酸分子を有し得、別の区画化セットは、メチル化されていない核酸分子（ゼロメチル化ヌクレオチド）を有し得る。

#### 【 0 0 9 3 】

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」、または「オリゴヌクレオチド」は、ヌクレオシド間連結によって接合したヌクレオシドの直鎖状ポリマー（デオキシリボヌクレオシド、リボヌクレオシド、またはその類似体を含む）を指す。一般には、ポリヌクレオチドは、少なくとも3つのヌクレオシドを含む。オリゴヌクレオチドは、多くの場合、少数、例えば3～4個の単量体単位から数百の単量体単位までのサイズにわたる。ポリヌクレオチドが「 A T G C C T G 」などの文字列で表されるときはいつでも、特に断りのない限り、ヌクレオチドは左から右に5' 3'の順序であること、および、DNAの場合では、「 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。文字 A、C、G、および T 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を示すために使用され得る。

#### 【 0 0 9 4 】

参照配列：本明細書で使用される場合、「参照配列」は、実験により決定された配列との比較のために使用される既知配列を指す。例えば、既知配列は、ゲノム全体、染色体、またはそれらの任意のセグメントであり得る。参照は、一般には、少なくとも約20、少なくとも約50、少なくとも約100、少なくとも約200、少なくとも約250、少なくとも約300、少なくとも約350、少なくとも約400、少なくとも約450、少なくとも約500、少なくとも約1000、または1000よりも多くのヌクレオチドを含む。参照配列は、ゲノムもしくは染色体の単一の連続した配列とアラインメントすることができるものである、またはゲノムもしくは染色体の異なる領域とアラインメントされ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、h G 1 9 および h G 3 8 などのヒトゲノムが挙げられる。

10

20

30

40

50

## 【0095】

試料：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分析することができる任意のものを意味する。

## 【0096】

シーケンシング：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、単量体単位の同一性および順序）を決定するために使用される多数の技術のいずれかを指す。シーケンシング方法の例としては、これだけに限定されないが、標的化シーケンシング、単一分子リアルタイムシーケンシング、エクソンまたはエキ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、トランジスタ媒介性シーケンシング、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、ゲル電気泳動、2重鎖シーケンシング、サイクルシーケンシング、一塩基伸長シーケンシング、固相シーケンシング、ハイスループットシーケンシング、大規模並列処理シグネチャーシーケンシング、エマルジョンPCR、低変性温度での共増幅-PCR（COLD-PCR）、多重PCR、可逆的ダイターミネーターによるシーケンシング、ペアエンドシーケンシング、短期シーケンシング、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、単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、およびこれらの組合せが挙げられる。一部の実施形態では、シーケンシングは、例えば、多くの他のものの中でも、Illumina、Inc.、Pacific Biosciences、Inc.、またはApplied Biosystems/Thermo Fisher Scientificから市販されている遺伝子解析機器などの遺伝子解析機器によって実施することができる。

## 【0097】

配列情報：本明細書で使用される場合、「配列情報」とは、核酸ポリマーに関しては、そのポリマー内の単量体単位（例えば、ヌクレオチドなど）の順序および同一性を意味する。

## 【0098】

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。これらの用語は、受胎後に生じるゲノム内の突然変異を指す。体細胞突然変異は、生殖細胞以外の体の任意の細胞において生じ得、したがって、後代には受け渡されない。

## 【0099】

スパイクイン試料：本明細書で使用される場合、「スパイクイン試料」とは、対象由来のポリヌクレオチドの試料にエピジェネティック対照核酸分子が添加された試料である。

## 【0100】

対象：本明細書で使用される場合、「対象」とは、哺乳動物種（例えば、ヒト）またはもしくは鳥類（例えば、鳥）の種などの動物、または植物などの他の生物体を指す。より詳細には、対象は、脊椎動物、例えば、マウス、霊長類、サルまたはヒトなどの哺乳動物であり得る。動物は、農場動物（例えば、肉用牛、乳牛、家禽、ウマ、ブタなど）、競技動物、および伴侶動物（例えば、愛玩動物または支援動物）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それを有する疑いがある個体、または治療を必要とするもしくは治療を必要とする疑いがある個体であり得る。「個体」または「患者」という用語は、「対象」と交換ができることが意図されている。

## 【0101】

例えば、対象は、がんを有するという診断を受けた個体、がん治療を受ける予定の個体

10

20

30

40

50

、および/または少なくとも1つのがん治療を受けた個体であり得る。対象は、がんが寛解した状態にある対象であり得る。別の例として、対象は、自己免疫疾患を有するという診断を受けた個体であり得る。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いう診断を受けたかまたはそれを有する疑いがある妊娠中または妊娠する計画がある女性個体であり得る。

【発明を実施するための形態】

【0102】

詳細な説明

I. 概要

ゲノム/エピジェネティック区画に基づく方法により、1つのアッセイでの多数の分析物についての同時シグナル検出が可能になり得る。しかし、区画に基づく分析物の検出されたシグナルは分解能が不十分なものであり得、シグナルの感度および特異度を変更する可変アッセイ条件に供される。リキッドバイオプシーアッセイの感度を上昇させる一方で、プロセス中の循環核酸(元の材料)またはデータの喪失を低減することが望ましい。本明細書に記載の1つまたは複数の対照を使用することによってアッセイ変動性を制御することにより、異なる実験にわたって結果を比較する能力を提供することも望ましい。

10

【0103】

本開示は、エピジェネティック区画アッセイを較正するための方法および組成物を提供する。本発明は、完全に分解されたゲノムの/エピジェネティック特色(例えば、合成オリゴヌクレオチド2重鎖内のメチル化シトシンの数が別個であること)を有するエピジェネティック対照核酸分子のセットを対照または参照として使用して、分析される試料のシグナル感度および特異度を上昇させることを含む。これらの分子を、エピジェネティック修飾に基づいた試料中の核酸分子の区画評価するため、および試料中の核酸分子(単数または複数)のエピジェネティック状態を決定するためにも使用することができる。

20

【0104】

無細胞ポリヌクレオチドなどの核酸分子は、メチル化などのエピジェネティック特徴に基づいて異なり得る。核酸は、異なるヌクレオチド配列、例えば、特定の遺伝子または遺伝子座を有し得る。特徴は、度合いに関して異なり得る。例えば、DNA分子は、それらのエピジェネティック修飾の程度が異なり得る。修飾の程度は、分子が供される修飾事象の数、例えば、メチル化群(メチル化の程度)または他のエピジェネティック変化の数などを指し得る。例えば、メチル化DNAは、低メチル化または高メチル化の状態であり得る。形態は、特徴の組合せ、例えば、一本鎖-メチル化されていないまたは二本鎖-メチル化を特徴とするものであり得る。特徴の1つまたは組合せに基づく分子の分別が単一分子の多次元解析のために有用であり得る。これらの方法を、試料中の核酸の多数の形態および/または修飾に適應させ、したがって、多数の形態に関する配列情報を得ることができる。また当該方法では最初の多数の形態または修飾状態の同一性が処理および解析を通して保存され、したがって、核酸塩基配列の解析をエピジェネティック解析と組み合わせることができる。一部の方法は、異なる形態または修飾状態を分離し、タグ付けし、その後、プールすることを伴い、それにより、試料中に存在する多数の形態を解析するために必要な処理ステップの数が減少する。試料中の多数の形態の核酸を解析することにより、より多くの情報がもたらされ、これは、より多くの分子が解析されること(入手可能な核酸の総量が非常に少ない場合に重要であり得る)に一部起因するだけでなく、異なる形態または修飾状態により異なる情報がもたらされることに起因し(例えば、突然変異がRNAにのみ存在する可能性がある)、また、異なる型の情報(例えば、遺伝学およびエピジェネティック)を互いに相関させ、それにより、より高い正確度、確実性を生じさせること、または医学的状态との新しい相関の発見をもたらすことができることに起因する。

30

40

【0105】

核酸分子の特徴は、修飾であり得、修飾は、種々の化学修飾(すなわち、エピジェネティック修飾)を含み得る。化学修飾の非限定的な例としては、これだけに限定されないが、DNAメチル化を含めた共有結合性DNA修飾を挙げることができる。一部の実施形態

50

では、DNA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続く））におけるシトシンへのメチル基の付加を含む。一部の実施形態では、DNAメチル化は、例えばN<sup>6</sup>-メチルアデニンなど、アデニンへのメチル基の付加を含む。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員炭素環の5番目の炭素の修飾）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（m5c）が創出されるシトシンの5C位へのメチル基の付加を含む。一部の実施形態では、メチル化は、m5cの誘導体を含む。m5cの誘導体としては、これだけに限定されないが、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-カルボキシメチルシトシン（5-caC）が挙げられる。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員炭素環の3番目の炭素の修飾）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）が生成するシトシンの3C位へのメチル基の付加を含む。メチル化はまた、非CpG部位にも生じ得、例えば、メチル化は、CpA、CpT、またはCpC部位に生じ得る。DNAメチル化により、メチル化DNA領域の活性が変化し得る。例えば、プロモーター領域内のDNAがメチル化されている場合、遺伝子の転写が抑止される可能性がある。DNAメチル化は、正常な発生に極めて重要であり、メチル化の異常により、エピジェネティック調節が妨害される恐れがある。エピジェネティック調節の妨害、例えば抑止により、がんなどの疾患が生じ得る。DNAにおけるプロモーターのメチル化によりがんが示され得る。

10

## 【0106】

20

CpG対は、二本鎖DNA分子のセンス鎖上のジヌクレオチドCpG（シトシン - リン酸 - グアニン、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続く）とアンチセンス鎖上のその相補的CpGである。CpG対は、完全にメチル化されているか、または半分がメチル化されている。図1は、二本鎖DNAにおける完全にメチル化されたCpG対および半分がメチル化されたCpG対の概略図である。図1Aは、ここでは、鎖101および102の両方のCpG対のシトシンヌクレオチドがメチル化されている、完全にメチル化されたCpG対103を示す（M - メチルシトシン；G - グアニン）。図1Bは、一方の鎖101上のCpG対のシトシンヌクレオチドはメチル化されているが相補鎖102上のCpG対のシトシンヌクレオチドはメチル化されていない、半分がメチル化されたCpG対104を示す（C - メチル化されていないシトシン；G - グアニン）。

30

## 【0107】

CpGジヌクレオチドは正常なヒトゲノム内では存在量が少なく、大多数のCpGジヌクレオチド配列は転写的に不活性であり（例えば、染色体の動原体周囲部分内およびリピートエレメント内のDNA異質染色質領域）、メチル化されている。しかし、特に転写開始点（TSS）周辺の多くのCpGアイランドはそのようなメチル化から保護される。

## 【0108】

がんは、メチル化などのエピジェネティック変異によって示され得る。がんにおけるメチル化の変化の例としては、正常な成長制御、DNA修復、細胞周期の調節、および/または細胞分化に關与する遺伝子のTSSにおけるCpGアイランドにおけるDNAメチル化の局所的な増加が挙げられる。この高メチル化状態は、關与する遺伝子の転写能力の異常な喪失に關連する可能性があり、また、遺伝子発現の変更の原因として点突然変異および欠失と少なくとも同じくらいの頻度で起こる。DNAメチル化プロファイリングを使用して、ゲノムの、発生の間に変更されるかまたは疾患、例えば、がんまたは任意のがん關連疾患によって攪乱を受けるメチル化の程度が異なる領域（「示差的にメチル化された領域」または「DMR」）を検出することができる。

40

## 【0109】

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ってメチル化パターンを決定することを伴い得る。例えば、分子をメチル化の程度（例えば、分子当たりのメチル化ヌクレオチドの相対数）に基づいて区画し、シーケンシングを行った後、異なる区画における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。これにより、他の領域と比較

50

して、高度にメチル化された状態がより多いかまたは高度にメチル化された状態がより少ないゲノムの領域を示すことができる。このように、ゲノム領域は、個々の分子とは対照的に、それらのメチル化の程度が異なり得る。メチル化に加えて、他のエピジェネティック修飾を同様にプロファイリングすることができる。

#### 【0110】

試料中の核酸分子を1つまたは複数の特徴に基づいて分別または区画することができる。試料中の核酸分子を区画することにより、稀なシグナルを増大させることができる。例えば、試料を高メチル化核酸分子と低メチル化核酸分子に区画することにより、高メチル化DNAには存在するが、低メチル化DNAではより少なく存在する（または存在しない）遺伝子変異をより容易に検出することができる。試料の多数の画分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。区画は、核酸分子をゲノム特徴の存在または非存在に基づいてサブセットまたは群に物理的に区画することを含み得る。分別は、核酸分子を、エピジェネティック修飾などのゲノム特徴が存在する度合いに基づいて区画群に物理的に区画することを含み得る。試料を、示差的な遺伝子発現または病態を示す特徴に基づいて1つまたは複数の群区画に分別または区画することができる。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」）の解析中に正常な状態と疾患状態の間のシグナルの差異をもたらすある1つの特徴、またはそれらの組合せに基づいて分別することができる。

#### 【0111】

本開示は、核酸分子の区画を査定または評価し、核酸分子におけるエピジェネティック状態（例えば、メチル化の状態）およびエピジェネティック修飾されたヌクレオチドの数（例えば、メチル化ヌクレオチドの数）を決定するための方法、組成物およびシステムを提供する。方法は、核酸分子を1つまたは複数のエピジェネティック修飾に基づいて異なる区画化セットに区画し、その後、シーケンシングを行い（単独でまたはまとめて）、各区画内の核酸分子を解析することを含み得る。一部の実施形態では、核酸の区画を、特定の標的ゲノム領域について富化させる。一部の実施形態では、核酸分子の区画を、富化の前および/または後に増幅させる。一部の実施形態では、区画化セットに分子バーコードを示差的にタグ付けし、示差的にタグ付けされた区画化セットの混合物中に再度組み合わせた後に富化を実施することができる。当該方法は、疾患の予後判定、診断および/またはモニタリングなどの種々の適用に使用することができる。一部の実施形態では、疾患はがんである。

#### 【0112】

核酸分子の区画方法を、エピジェネティック対照核酸分子を使用することによって評価することができる。エピジェネティック対照核酸分子は、エピジェネティック修飾されたヌクレオチドを有し得る合成核酸分子である。一部の実施形態では、エピジェネティック対照核酸分子は、異なるエピジェネティック状態を有する核酸分子を含み得る。エピジェネティック状態は、核酸分子のエピジェネティック修飾のレベル/度合いを指し得る。例えば、エピジェネティック修飾がDNAメチル化である場合には、エピジェネティック状態は、高度にメチル化された、低度にメチル化されたまたは中程度にメチル化された核酸分子を指し得る。エピジェネティック状態は、エピジェネティック修飾を有するヌクレオチドの数も指し得る。例えば、エピジェネティック修飾がDNAメチル化である場合には、エピジェネティック状態は、核酸分子のメチル化ヌクレオチドの数を指し得る。エピジェネティック修飾は、核酸分子の配列および/または塩基対合特異性の変化を伴わない、ヌクレオチド（単数または複数）の塩基の任意の修飾であり得る。修飾は、ヌクレオチドの塩基の化学修飾であり得る。一部の場では、修飾は、ヌクレオチドの塩基のメチル化であり得る。例えば、修飾は、5-メチルシトシンがもたらされるシトシンのメチル化であり得る。

#### 【0113】

一部の実施形態では、エピジェネティック対照核酸分子は合成分子であり、エピジェネ

ティック対照核酸分子の配列ならびにエピジェネティック対照核酸分子内のエピジェネティック修飾されたヌクレオチドの位置および数は解析前にすでに分かっている。したがって、エピジェネティック対照核酸分子をポリヌクレオチドの試料に添加することによって、および区画化セット内のエピジェネティック対照核酸分子を追跡することによって、エピジェネティック対照核酸分子の区画の効果を検査することができる。

#### 【0114】

したがって、一態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、(a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それによって、スパイクイン試料を作製するステップと、(b) スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、(c) 複数の区画化セットから分子のサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含む、ステップと、(d) 富化された分子のセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、(e) エピジェネティック対照核酸分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(f) 1つまたは複数のエピジェネティック区画スコアをエピジェネティック区画カットオフの1つまたは複数と比較するステップとを含む方法を提供する。これらの実施形態では、試料の核酸分子およびエピジェネティック対照核酸分子の区画を必ず同時に行う。一部の実施形態では、解析ステップは、区画化セットのうちの少なくとも1つにおける所与のエピジェネティック状態にあるエピジェネティック対照核酸分子の数/フラクションを推定することを含む。

#### 【0115】

図2は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法200の実施形態例を例示する。201では、区画を評価しようとしているエピジェネティック対照核酸分子を試料に添加して、スパイクイン試料を生成する。

#### 【0116】

一部の実施形態では、エピジェネティック対照核酸分子は、エピジェネティック状態のレベルが異なる(すなわち、エピジェネティック修飾されたヌクレオチドの数が異なる)核酸分子の1つまたは複数のサブセットを含み得る。一部の実施形態では、エピジェネティック対照核酸分子は、配列が異なるおよび/または長さが異なる核酸分子を含み得る。他の実施形態では、エピジェネティック対照核酸分子は、配列が同一であるまたは長さが同一である核酸分子を含み得る。

#### 【0117】

202では、エピジェネティック対照核酸分子とポリヌクレオチドの試料由来の核酸分子の両方を含むスパイクイン試料の少なくとも1つのサブセットの核酸分子を分子のエピジェネティック状態に基づいて複数の区画化セットに区画または分別する。区画は、エピジェネティック修飾の存在または非存在に基づき得る、かつ/またはエピジェネティック修飾の度合いに基づき得る。エピジェネティック修飾の例としては、これだけに限定されないが、メチル化の存在または非存在、メチル化のレベルおよびメチル化の型(5'シトシン)が挙げられる。一部の実施形態では、エピジェネティック修飾は、DNAメチル化であり得る。これらの実施形態では、スパイクイン試料の分子を異なるメチル化のレベル(異なるメチル化ヌクレオチドの数)に基づいて区画する。一部の実施形態では、スパイクイン試料を2つまたはそれよりも多くの区画化セット(例えば、少なくとも3つ、4つ、5つ、6つ、または7つの区画化セット)に区画することができる。一部の実施形態では、区画は、核酸分子の結合性物質に対する示差的な結合親和性に基づく。結合性物質の例としては、これだけに限定されないが、メチル結合性ドメイン(MBD)およびメチル結合タンパク質(MBP)が挙げられる。本明細書において意図されているMBPの例とし

10

20

30

40

50

ては、これだけに限定されないが、以下が挙げられる：

( a ) MeCP2 は、5 - メチル - シトシンに、修飾されていないシトシンよりも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、5 - ヒドロキシメチル - シトシンに、修飾されていないシトシンよりも優先的に結合する；

( c ) FOXK1、FOXK2、FOXP1、FOXP4 および FOXI3 は、5 - ホルミル - シトシンに、修飾されていないシトシンよりも好んで結合する ( Iurlaro et al., Genome Biol. 14, R119 (2013)) ; ならびに

( d ) 1 つまたは複数のメチル化ヌクレオチド塩基に特異的な抗体。

#### 【0118】

いくつかの親和性作用物質および修飾に関しては、作用物質への結合は、核酸が修飾を有するかどうかに応じて基本的に全か無かの様式で起こるが、分離には様々な度合いがあり得る。そのような実施形態では、修飾が過剰に存在する核酸は、作用物質に、修飾が過少に存在する核酸よりも大きな度合いで結合する。あるいは、修飾を有する核酸は、全か無かの様式で結合し得る。しかし、次いで、結合性物質から種々のレベルの修飾を逐次的に溶出させることができる。

#### 【0119】

例えば、一部の実施形態では、区画は、バイナリであり得るまたは修飾の度合い / レベルに基づき得る。例えば、メチル結合性ドメインタンパク質 (例えば、MethylMiner Methylated DNA Enrichment Kit (ThermoFisher Scientific)) を使用して全てのメチル化断片をメチル化されていない断片から区画することができる。その後、追加的な区画は、メチル結合性ドメインおよび結合した断片を伴う溶液の塩濃度を調整することにより、異なるメチル化のレベルを有する断片を溶出させることを伴い得る。塩濃度が上昇するにつれ、より高いメチル化レベルを有する断片が溶出する。

#### 【0120】

一部の実施形態では、区画は、核酸分子を、エピジェネティック修飾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結合性物質に対する核酸分子の示差的な結合親和性に基づいて区画することを含む。

#### 【0121】

一部の実施形態では、区画化セットは、異なる修飾の程度を有する核酸を表すものである (修飾の過剰存在または過少存在)。過剰存在および過少存在は、集団内の鎖当たりの修飾の数の中央値と比べた、核酸が有する修飾の数によって定義することができる。例えば、試料中の核酸分子における 5 - メチルシトシンヌクレオチドの数の中央値が 2 である場合、5 - メチルシトシン残基を 2 つよりも多く含む核酸分子はこの修飾が過剰に存在するものであり、5 - メチルシトシン残基が 1 つまたはゼロである核酸が過少に存在するものである。アフィニティー分離の効果は、修飾が過剰に存在する結合相中の核酸について、および修飾が過少に存在する非結合相中 (すなわち、溶液中) の核酸について区画することである。その後の処理前に結合相中の核酸を溶出させることができる。

#### 【0122】

MethylMiner Methylated DNA Enrichment Kit (ThermoFisher Scientific) を使用すると、種々のメチル化のレベルを、逐次的な溶出を使用して区画することができる。例えば、核酸集団を、磁気ビーズに付着させた、キットからの MBD と接触させることによって低メチル化区画 (メチル化なし) をメチル化区画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。その後、1 つまたは複数の溶出ステップを逐次的に実施して、異なるメチル化のレベルを有する核酸を溶出する。例えば、メチル化核酸の第 1 のセットを約 150 mM または約 160 mM またはそれよりも高い濃度、例えば、少なくとも 150 mM、200 mM、300 mM、400 mM、500 mM、600 mM、700 mM、800 mM、900 mM、1000 mM、または 2000 mM の塩濃度で溶出さ

10

20

30

40

50

せることができる。そのようなメチル化核酸の溶出後、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核酸から分離する。溶出および磁気分離ステップ自体を繰り返して、低メチル化区画（メチル化がないことを表す）、メチル化区画（低レベルのメチル化を表す）、および高メチル化区画（高レベルのメチル化を表す）などの種々の区画を創出することができる。

#### 【0123】

一部の方法では、アフィニティー分離に使用した作用物質に結合した核酸を洗浄ステップに供する。洗浄ステップでは、親和性作用物質に弱く結合した核酸が洗い流される。そのような核酸を、平均値または中央値（すなわち、固相に結合したままの核酸と試料を作用物質に最初に接触させた際に固相に結合しなかった核酸の間）にある程度近い修飾を有する核酸を富化させることができる。アフィニティー分離の結果、修飾の程度が異なる核酸の区画が少なくとも2つ、および時には、3つまたはそれよりも多く生じる。

10

#### 【0124】

核酸分子の区画を、区画された核酸分子をシーケンシングすることによって、またはデジタル液滴PCR（ddPCR）によって、または定量的PCR（qPCR）によって解析することができる。区画を解析する前に、区画化セット内の核酸分子を富化させることができ、その結果、目的の核酸分子からのシグナルを増大させ、したがって、感度を改善することができる。203では、複数の区画化セット内の核酸分子の少なくとも1つのサブセットを富化させ、したがって、目的の領域に属するエピジェネティック対照核酸分子およびポリヌクレオチドの試料由来の核酸分子が富化される。

20

#### 【0125】

一部の実施形態では、富化の前に、複数の区画化セットのそれぞれに示差的にタグ付けする。次いで、タグ付けされた区画化セットを、集成的な試料調製および/またはシーケンシングのために一緒にプールする。区画化セットへの示差的なタグ付けは、特定の区画化セットに属する核酸分子の追跡を続けるのに役立つ。タグは通常、アダプターの構成成分として提供される。異なる区画化セット内の核酸分子は、1つの区画化セットのメンバーを別の区画化セットのメンバーと区別することができる異なるタグを受ける。同じ区画化セットの核酸分子と連結したタグは互いに同じであっても異なってもよい。しかし、互いに異なる場合には、タグは、それらのタグが付着した分子が特定の区画化セットに属すると識別されるように、それらの配列の一部が共通するものであり得る。例えば、スパイクイン試料の分子を2つの区画化セット、P1およびP2に区画する場合には、P1内の分子にA1、A2、A3などでタグ付けすることができ、P2の分子にB1、B2、B3などでタグ付けすることができる。そのようなタグ付けシステムにより、区画化セット、および区画化セット内の分子間を区別することが可能になる。

30

#### 【0126】

204では、富化された分子の少なくとも1つのサブセットをシーケンシングする。得られる配列情報は、核酸分子の配列および核酸分子に付着したタグを含む。核酸分子に付着したタグの配列から、タグを核酸分子の区画化セットと関連させることができる。配列情報を使用して、エピジェネティック対照核酸分子およびそれらの対応する区画化セットを識別する。この情報を使用して、エピジェネティック対照核酸分子の区画を解析する。205では、1つまたは複数の区画化セット内の1つまたは複数のエピジェネティック状態に属するエピジェネティック対照核酸分子の1つまたは複数のエピジェネティック区画スコアを決定する。一部の実施形態では、区画方法の感度および/または特異度をエピジェネティック区画スコアによって査定することができる。エピジェネティック区画スコアは、特定のエピジェネティック状態に属する核酸分子の区画を表すスコアである。エピジェネティック状態に属する核酸分子のエピジェネティック区画スコアを各区画化セットについて決定する。例えば、特定のエピジェネティック状態に属するエピジェネティック対照核酸分子のエピジェネティック区画スコアを決定することができる。エピジェネティック区画スコアは、特定のエピジェネティック状態に属する核酸分子の数（または統計学的に推定される数）の尺度であり得る。エピジェネティック区画スコアは、フラクションま

40

50

たはパーセンテージによるものであり得る。エピジェネティック区画スコアは、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属するエピジェネティック対照核酸分子の数と他の残りの区画化セット（単数または複数）内に存在するそのエピジェネティック状態に属するエピジェネティック対照核酸分子の数の比の尺度であり得る。一部の実施形態では、エピジェネティック区画スコアは、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属するエピジェネティック対照核酸分子の数の、全ての区画化セット内のそのエピジェネティック状態に属するエピジェネティック対照核酸分子の総数に対するフラクションまたはパーセンテージであり得る。一部の実施形態では、エピジェネティック区画スコアを、区画化セットのそれぞれにおけるエピジェネティック対照核酸分子各エピジェネティック状態に関して決定する。一部の実施形態では、エピジェネティック区画スコアを、1つまたは複数の区画化セット内の1つまたは複数の特定のエピジェネティック状態を有するエピジェネティック対照核酸分子に関して決定する。一部の実施形態では、エピジェネティック区画スコアを、特定の区画化セット内の特定のエピジェネティック状態を有するエピジェネティック対照核酸分子に関して決定する。

10

#### 【0127】

一部の実施形態では、エピジェネティック区画スコアは、CGを有さない（「ゼロ」CG）分子が高区画化セットに区画される有効度を対象とし得る。このスコアは、0 CGスコアと称することができる。一部の実施形態では、0 CGスコアを高区画化セット内のCGを有さない分子のフラクションまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、以下のうちの少なくとも1つを有するエピジェネティック対照核酸分子のフラクションおよび/または高メチル化対照分子のフラクションの尺度であり得る：

20

- 高メチル化区画化セット（すなわち、高度にメチル化された区画化セット）において、
  - (i) 1つのメチルCG（エピジェネティック区画スコアを1 CGスコアと称することができる）、
  - (ii) 2つのメチルCG（エピジェネティック区画スコアを2 CGスコアと称することができる）、
  - (iii) 3つのメチルCG（エピジェネティック区画スコアを3 CGスコアと称することができる）、
  - (iv) 4つのメチルCG（エピジェネティック区画スコアを4 CGスコアと称することができる）および
  - (v) 5つのメチルCG（エピジェネティック区画スコアを5 CGスコアと称することができる）。

30

#### 【0128】

一部の実施形態では、エピジェネティック区画スコアは、低メチル化（すなわち、低度メチル化された）エピジェネティック対照核酸分子が高メチル化区画化セットに区画される有効度を対象とし得る。このスコアは、hyposコアと称することができる。一部の実施形態では、hyposコアは、高メチル化区画化セット内の低メチル化エピジェネティック対照核酸分子のフラクションまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、低メチル化区画化セット内の高メチル化エピジェネティック対照核酸分子が5%未満になるのに必要なメチル化CGの数の尺度であり得る。このスコアは、methyl-5と称することができる。一部の実施形態では、エピジェネティック区画スコアは、高メチル化区画化セット内の高メチル化エピジェネティック対照核酸分子が少なくとも50%になるのに必要なメチル化CGの数の尺度であり得る。このスコアは、methyl-halfと称することができる。

40

#### 【0129】

例えば、エピジェネティック対照核酸分子の3つのサブセット（サブセットA、BおよびC）を使用し、各サブセットはメチル化ヌクレオチドの数が異なる。これらの3つのサブセット内のエピジェネティック対照核酸分子を、それらのメチル結合タンパク質に対す

50

る結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。各サブセットについて、エピジェネティック区画スコアを、区画化セットのそれぞれ(P1、P2およびP3)について決定する。すなわち、サブセットAに属するエピジェネティック対照核酸分子は、3つのエピジェネティック区画スコア、つまり3つの区画化セット、P1、P2およびP3のそれぞれについて1つのエピジェネティック区画スコアを有する。同様に、サブセットBおよびCのそれぞれは、3つのエピジェネティック区画スコア、つまり3つの区画化セットP1、P2およびP3のそれぞれについて1つのエピジェネティック区画スコアを有する。内在性対照分子についてもエピジェネティック区画スコアを決定することができる。

【0130】

10

別の実施形態では、エピジェネティック対照核酸分子の3つのサブセット(サブセットA、BおよびC)を使用し、各サブセットはメチル化ヌクレオチドの数が異なる(すなわち、各サブセットは異なるエピジェネティック状態を有する)。これらの3つのサブセット内のエピジェネティック対照核酸分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。この実施形態では、P1区画化セット内のサブセットA分子についてのみエピジェネティックスコアを決定する。このエピジェネティックスコアは、P1区画化セット内のサブセットA分子の、サブセットA分子(P1、P2およびP3つの区画化セット内の)の総数に対するフラクションまたはパーセンテージの尺度であり得る。

【0131】

20

エピジェネティック区画スコアは、0~1の間(フラクションで)または0~100%の間(パーセンテージで)の任意の値または範囲であり得る。一部の実施形態では、エピジェネティック区画スコアは、メチル化CGの数によるものであり得る(例えば、methyl-halfおよびmethyl-5について)。

【0132】

206では、エピジェネティック対照核酸分子のエピジェネティック区画スコアをエピジェネティック区画カットオフ(所定のカットオフ)と比較して、区画方法を評価する。エピジェネティック区画カットオフは、特定のエピジェネティック状態に属する核酸分子の区画を評価するために使用される所定のカットオフ値またはカットオフ範囲であり、各区画化セットは、エピジェネティック状態に属する核酸分子についてのエピジェネティック区画カットオフを有する。エピジェネティック区画カットオフは、核酸分子および区画化セットのエピジェネティック状態によって異なる、すなわち、各エピジェネティック状態がそれ自体のエピジェネティック区画カットオフを有し、区画化セットのいずれもがそのエピジェネティック状態についての別々のエピジェネティック区画カットオフを有する。カットオフは、パーセンテージまたはフラクションによるものであり得、カットオフは、特定のカットオフ値ではなくカットオフ範囲であり得る。例えば、特定のエピジェネティック状態に属するエピジェネティック対照核酸分子についてのエピジェネティック区画カットオフは、区画化セットP1、P2およびP3について、それぞれ70%から79%の間、10%から15%の間および5%未満であり得る。そのエピジェネティック状態に属するエピジェネティック対照核酸分子のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法は成功である。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.01%、0.02%、0.05%、0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、5%、少なくとも5%または少なくとも10%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.01%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.02%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.03%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.04%であり得る。一部の実施形態では、0CGスコアについてのエ

30

40

50

ピジェネティック区画カットオフは、0.05%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.1%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.2%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.3%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.4%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.5%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.6%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.7%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.8%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、0.9%であり得る。一部の実施形態では、0 C Gスコアについてのエピジェネティック区画カットオフは、1%であり得る。

10

#### 【0133】

一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.1%、0.5%、1%、2%、3%、4%、5%、7%または少なくとも10%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.1%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.5%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、1%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、2%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、3%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、4%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、5%であり得る。

20

#### 【0134】

一部の実施形態では、カットオフは、メチル化C Gの数によるものであり得る（例えば、methyl-5およびmethyl-halfについて）。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、5、10、20、30、40または50 m C Gであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、5 m C Gであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、10 m C Gであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、20 m C Gであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、30 m C Gであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、40 m C Gであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、50 m C Gであり得る。

30

#### 【0135】

一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、5、10、15、20、25、30、35または40 m C Gであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、5 m C Gであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、10 m C Gであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、15 m C Gであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、20 m C Gであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、25 m C Gであり得る。一部の実施形態では、methyl-halfスコアについてのエ

40

50

ピジェネティック区画カットオフは、30mCGであり得る。一部の実施形態では、methy1-halfスコアについてのエピジェネティック区画カットオフは、35mCGであり得る。一部の実施形態では、methy1-halfスコアについてのエピジェネティック区画カットオフは、40mCGであり得る。

#### 【0136】

一部の実施形態では、1つまたは複数の区画化セット内の1つまたは複数のエピジェネティック状態に属するエピジェネティック対照核酸分子の1つまたは複数のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法を成功に分類することができる。そうでなければ、全ての区画化セットについてエピジェネティック区画スコアがカットオフの外側にある場合、区画方法を失敗に分類することができる。例えば、エピジェネティック対照核酸分子の2つのサブセット、サブセットAおよびBが存在し、各サブセットは、エピジェネティック修飾の度合いが異なる（すなわち、各サブセットはエピジェネティック状態が異なる）。これらのエピジェネティック対照核酸分子を2つの区画化セット、P1およびP2に区画することができる。サブセットAに属する分子について、各区画化セットP1およびP2に対して1つの、2つのエピジェネティック区画スコア（例えば、S1およびS2）をそれらの区画に基づいて決定する。同様に、サブセットBに属する分子について、P1に対して1つ、およびP2に対して1つの、2つのエピジェネティック区画スコア（例えば、S3およびS4）を決定する。特定のエピジェネティック状態を有する分子の各サブセットは、区画化セットのそれぞれについて所定のエピジェネティック区画カットオフを有する。本実施例では、サブセットAのエピジェネティック対照核酸分子は、2つのエピジェネティック区画カットオフ、C1およびC2（2つの区画化セットP1およびP2に対するもの）を有し、同様に、サブセットBのエピジェネティック対照核酸分子は、2つのエピジェネティック区画カットオフ、C3およびC4を有する。両方のサブセットのエピジェネティック区画スコアをそれらの対応するエピジェネティック区画カットオフと比較する。本実施例では、4つのエピジェネティック区画スコアの全てがそれらの対応するエピジェネティック区画カットオフの範囲内に入る場合、すなわち、本実施例では、 $S1 < C1$ および $S2 < C2$ および $S3 < C3$ および $S4 < C4$ の場合にのみ、区画方法を成功とみなす。そうでなければ、全ての区画化セットについてエピジェネティック区画スコアがカットオフの外側にある場合、区画方法を失敗に分類することができる。

#### 【0137】

別の実施形態では、エピジェネティック対照核酸分子の3つのサブセット（サブセットA、BおよびC）を使用し、各サブセットはメチル化ヌクレオチドの数が異なる（すなわち、各サブセットは異なるエピジェネティック状態を有する）。これらの3つのサブセット内のエピジェネティック対照核酸分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。この実施形態では、P1区画化セット内のサブセットA分子についてのみエピジェネティックスコアを決定する。このエピジェネティックスコアは、P1区画化セット内のサブセットA分子の、サブセットA分子（P1、P2およびP3の区画化セット内の）の総数に対するフラクションまたはパーセンテージの尺度であり得る。このエピジェネティック区画スコアがその対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法を成功に分類する。そうでなければ、区画方法を失敗に分類する。

#### 【0138】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、(a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、(b) スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、(c) 複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポ

リヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、(d) 富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、(e) エピジェネティック対照核酸分子および内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(f) 1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を提供する。これらの実施形態では、試料の核酸分子およびエピジェネティック対照核酸分子の区画を必ず同時に行う。一部の実施形態では、解析ステップは、区画化セットのうちの少なくとも1つにおける所与のエピジェネティック状態にあるエピジェネティック対照核酸分子および/または内在性対照分子の数/フラクションを推定することを含む。

10

#### 【0139】

図3は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法300の実施形態例を例示する。この実施形態では、エピジェネティック対照核酸分子および内在性対照分子の両方の区画を解析して、区画方法を評価する。特定のエピジェネティック状態を有するヒトゲノム内の領域が存在し、その領域のエピジェネティック状態は多くの場合に変動/変化せず、常に同じままである/異なる対象および/または異なる疾患の型/病期の場合でも一貫したままである。そのようなエピジェネティック状態が不変のヒトゲノム領域に対応するポリヌクレオチドの試料中の核酸分子は、内在性対照分子と称される。301では、エピジェネティック対照核酸分子を、区画を評価しようとしているポリヌクレオチドの試料に添加して、スパイクイン試料を生成する。

20

#### 【0140】

一部の実施形態では、エピジェネティック対照核酸分子は、エピジェネティック状態のレベルが異なる(すなわち、エピジェネティック修飾されたヌクレオチドの数が異なる)核酸分子の1つまたは複数のサブセットを含み得る。一部の実施形態では、エピジェネティック対照核酸分子は、配列が異なるおよび/または長さが異なる核酸分子を含み得る。他の実施形態では、エピジェネティック対照核酸分子は、配列が同一であるまたは長さが同一である核酸分子を含み得る。

30

#### 【0141】

302では、エピジェネティック対照核酸分子とポリヌクレオチドの試料由来の核酸分子の両方を含むスパイクイン試料の少なくとも1つのサブセットの核酸分子を分子のエピジェネティック状態に基づいて複数の区画化セットに区画または分別する。区画は、エピジェネティック修飾の存在または非存在に基づき得る、かつ/またはエピジェネティック修飾の度合いに基づき得る。エピジェネティック修飾の例としては、これだけに限定されないが、メチル化の存在または非存在、メチル化のレベルおよびメチル化の型(5' シトシン)が挙げられる。一部の実施形態では、エピジェネティック修飾は、DNAメチル化であり得る。これらの実施形態では、スパイクイン試料の分子を異なるメチル化のレベル(異なるメチル化ヌクレオチドの数)に基づいて区画する。一部の実施形態では、スパイクイン試料を2つまたはそれよりも多くの区画化セット(例えば、少なくとも3、4、5、6、または7つの区画化セット)に区画することができる。一部の実施形態では、区画は、核酸分子の結合性物質に対する示差的な結合親和性に基づく。

40

#### 【0142】

核酸分子の区画を、区画された核酸分子をシーケンシングすることによって、デジタル液滴PCR(ddPCR)によって、または定量的PCR(qPCR)によって解析することができる。区画を解析する前に、区画化セット内の核酸分子を富化させることができ、その結果、目的の核酸分子からのシグナルを増大させ、したがって、感度を改善することができる。303では、複数の区画化セット内の核酸分子の少なくとも1つのサブセットを富化させ、したがって、目的の領域に属するエピジェネティック対照核酸分子、内在

50

性対照分子（ポリヌクレオチドの試料由来）および他のポリヌクレオチドの試料由来の核酸分子を富化する。

【0143】

一部の実施形態では、富化の前に、複数の区画化セットのそれぞれに示差的にタグ付ける。次いで、タグ付けされた区画化セットを、集成的な試料調製および/またはシーケンシングのために一緒にプールする。区画化セットへの示差的なタグ付けは、特定の区画化セットに属する核酸分子の追跡を続けるのに役立つ。タグは通常、アダプターの構成成分として提供される。異なる区画化セット内の核酸分子は、1つの区画化セットのメンバーを別の区画化セットのメンバーと区別することができる異なるタグを受ける。同じ区画化セットの核酸分子と連結したタグは互いに同じであっても異なってもよい。しかし、互いに異なる場合には、タグは、それらのタグが付着した分子が特定の区画化セットに属すると識別されるように、それらの配列の一部が共通するものであり得る。

10

【0144】

304では、富化された分子の少なくとも1つのサブセットをシーケンシングする。得られる配列情報は、核酸分子の配列および核酸分子に付着したタグを含む。核酸分子に付着したタグの配列から、タグを核酸分子の区画化セットと関連させることができる。配列情報を使用して、エピジェネティック対照核酸分子および内在性対照分子およびそれらの対応する区画化セットを識別する。この情報を使用して、エピジェネティック対照核酸分子および内在性対照分子の区画を解析する。305では、1つまたは複数の区画化セット内の1つまたは複数のエピジェネティック状態に属するエピジェネティック対照核酸分子および内在性対照分子の1つまたは複数のエピジェネティック区画スコアを決定する。一部の実施形態では、区画方法の感度および/または特異度をエピジェネティック区画スコアによって査定することができる。エピジェネティック区画スコアは、特定のエピジェネティック状態に属する核酸分子の区画を表すスコアである。エピジェネティック状態に属する核酸分子のエピジェネティック区画スコアを各区画化セットについて決定する。例えば、特定のエピジェネティック状態に属するエピジェネティック対照核酸分子および内在性対照分子のエピジェネティック区画スコアを決定することができる。エピジェネティック区画スコアは、特定のエピジェネティック状態に属する核酸分子の数（または統計学的に推定される数）の尺度であり得る。エピジェネティック区画スコアは、フラクションまたはパーセンテージによるものであり得る。エピジェネティック区画スコアは、(i) エピジェネティック対照核酸分子については、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属するエピジェネティック対照核酸分子の数の他の残りの区画化セット（単数または複数）内に存在するそのエピジェネティック状態に属するエピジェネティック対照核酸分子の数に対する比の尺度、および(ii) 内在性対照分子については、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属する内在性対照分子の数の他の残りの区画化セット（単数または複数）内に存在するそのエピジェネティック状態に属する内在性対照分子の数に対する比の尺度であり得る。一部の実施形態では、エピジェネティック区画スコアは、(i) エピジェネティック対照核酸分子については、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属するエピジェネティック対照核酸分子の数の、全ての区画化セット内のそのエピジェネティック状態に属するエピジェネティック対照核酸分子の総数に対するフラクションまたはパーセンテージであり得、(ii) 内在性対照分子については、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属する内在性対照分子の数の、全ての区画化セット内のそのエピジェネティック状態に属する内在性対照分子の総数に対するフラクションまたはパーセンテージであり得る。一部の実施形態では、エピジェネティック区画スコアを、区画化セットのそれぞれにおけるエピジェネティック対照核酸分子および内在性対照分子の各エピジェネティック状態に関して決定する。一部の実施形態では、エピジェネティック区画スコアを、1つまたは複数の区画化セット内の1つまたは複数の特定のエピジェネティック状態を有するエピジェネティック対照核酸分子および内在性対照分子に関して決定する。一部の実施形態では、エピジェネティック区画スコ

20

30

40

50

アを、特定の区画化セット内の特定のエピジェネティック状態を有するエピジェネティック対照核酸分子および内在性対照分子に関して決定する。

【0145】

一部の実施形態では、エピジェネティック区画スコアは、CGを有さない(「ゼロ」CG)分子が高区画化セットに区画される有効度を対象とし得る。このスコアは、0 CGスコアと称することができる。一部の実施形態では、0 CGスコアを高区画化セット内のCGを有さない分子のフラクションまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、以下のうちの少なくとも1つを有する、エピジェネティック対照核酸分子のフラクションおよび/または高メチル化対照分子のフラクションの尺度であり得る：

高メチル化区画化セット(すなわち、高度にメチル化された区画化セット)において、  
(vi) 1つのメチルCG(エピジェネティック区画スコアを1 CGスコアと称することができる)、

(vii) 2つのメチルCG(エピジェネティック区画スコアを2 CGスコアと称することができる)、

(viii) 3つのメチルCG(エピジェネティック区画スコアを3 CGスコアと称することができる)、

(ix) 4つのメチルCG(エピジェネティック区画スコアを4 CGスコアと称することができる)および

(x) 5つのメチルCG(エピジェネティック区画スコアを5 CGスコアと称することができる)。

【0146】

一部の実施形態では、エピジェネティック区画スコアは、低メチル化対照分子または低メチル化エピジェネティック対照核酸分子が高メチル化区画化セットに区画される有効度を対象とし得る。このスコアは、hyposコアと称することができる。一部の実施形態では、hyposコアを高メチル化区画化セット内の低メチル化対照分子または低メチル化エピジェネティック対照核酸分子のフラクションまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、低メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子が5%未満になるのに必要なメチル化CGの数の尺度であり得る。このスコアを、methyl-5と称することができる。一部の実施形態では、エピジェネティック区画スコアは、高メチル化区画化セット内の高メチル化対照分子および/または高メチル化エピジェネティック対照核酸分子が少なくとも50%になるのに必要なメチル化CGの数の尺度であり得る。このスコアを、methyl-halfと称することができる。

【0147】

例えば、エピジェネティック対照核酸分子の3つのサブセット(サブセットA、BおよびC)を使用し、各サブセットはメチル化ヌクレオチドの数が異なる。これらの3つのサブセット内のエピジェネティック対照核酸分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。各サブセットについて、エピジェネティック区画スコアを、区画化セットのそれぞれ(P1、P2およびP3)について決定する。すなわち、サブセットAに属するエピジェネティック対照核酸分子は、3つのエピジェネティック区画スコア、つまり3つの区画化セット、P1、P2およびP3のそれぞれについて1つのエピジェネティック区画スコアを有する。同様に、サブセットBおよびCのそれぞれは、3つのエピジェネティック区画スコア、つまり3つの区画化セットP1、P2およびP3のそれぞれについて1つのエピジェネティック区画スコアを有する。内在性対照分子についてもエピジェネティック区画スコアを決定することができる。

【0148】

別の実施形態では、エピジェネティック対照核酸分子の3つのサブセット(サブセットA、BおよびC)を使用し、各サブセットはメチル化ヌクレオチドの数が異なる(すなわ

10

20

30

40

50

ち、各サブセットは異なるエピジェネティック状態を有する)。これらの3つのサブセット内のエピジェネティック対照核酸分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。この実施形態では、P1区画化セット内のサブセットA分子についてのみエピジェネティックスコアを決定する。このエピジェネティックスコアは、P1区画化セット内のサブセットA分子の、サブセットA分子(P1、P2およびP3つの区画化セット内の)の総数に対するフラクションまたはパーセンテージの尺度であり得る。

【0149】

エピジェネティック区画スコアは、0~1の間(フラクションで)または0~100%の間(パーセンテージで)任意の値または範囲であり得る。一部の実施形態では、エピジェネティック区画スコアは、メチル化CGの数によるものであり得る(例えば、methyl-halfおよびmethyl-5について)。

【0150】

306では、エピジェネティック対照核酸分子および内在性対照分子のエピジェネティック区画スコアを、それらの対応するエピジェネティック区画カットオフ(所定のカットオフ)と比較して、区画方法を評価する。エピジェネティック区画カットオフは、特定のエピジェネティック状態に属する核酸分子の区画を評価するために使用される所定のカットオフ値またはカットオフ範囲であり、各区画化セットは、エピジェネティック状態に属する核酸分子についてのエピジェネティック区画カットオフを有する。エピジェネティック区画カットオフは、核酸分子および区画化セットのエピジェネティック状態によって異なる、すなわち、各エピジェネティック状態がそれ自体のエピジェネティック区画カットオフを有し、区画化セットのいずれもがそのエピジェネティック状態についての別々のエピジェネティック区画カットオフを有する。カットオフは、パーセンテージまたはフラクションによるものであり得、カットオフは、特定のカットオフ値ではなくカットオフ範囲であり得る。例えば、特定のエピジェネティック状態に属するエピジェネティック対照核酸分子についてのエピジェネティック区画カットオフは、区画化セットP1、P2およびP3について、それぞれ70%から79%の間、10%から15%の間および5%未満であり得る。そのエピジェネティック状態に属するエピジェネティック対照核酸分子のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法は成功である。

【0151】

一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.01%、0.02%、0.05%、0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、5%、少なくとも5%または少なくとも10%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.01%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.02%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.03%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.04%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.05%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.1%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.2%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.3%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.4%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.5%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.6%であり得る。一部の実施形態では、0CGスコアについてのエピジェネティック区画カットオフは、0.7%であり得る。一部の実施形態では、0CGスコアに

10

20

30

40

50

ついでのエピジェネティック区画カットオフは、0.8%であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.9%であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、1%であり得る。

【0152】

一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.1%、0.5%、1%、2%、3%、4%、5%、7%または少なくとも10%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.1%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.5%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、1%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、2%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、3%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、4%であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、5%であり得る。

10

【0153】

一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、5、10、20、30、40または50 mCGであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、5 mCGであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、10 mCGであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、20 mCGであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、30 mCGであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、40 mCGであり得る。一部の実施形態では、methyl-5についてのエピジェネティック区画カットオフは、50 mCGであり得る。

20

【0154】

一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、5、10、15、20、25、30、35または40 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、5 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、10 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、15 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、20 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、25 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、30 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、35 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、40 mCGであり得る。

30

40

【0155】

一部の実施形態では、1つまたは複数の区画化セット内の1つまたは複数のエピジェネティック状態に属する1つまたは複数のエピジェネティック対照核酸分子および内在性対照分子のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法を成功に分類することができる。そうでなければ、区画方法を失敗に分類することができる。

【0156】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック

50

状態に基づく区画を評価するための方法であって、(a)ポリヌクレオチドの試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、(c)複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、ポリヌクレオチドの試料由来の核酸分子の群を含み、無細胞ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、(d)富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、(e)内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(f)1つまたは複数のエピジェネティック区画スコアをエピジェネティック区画カットオフの1つまたは複数と比較するステップとを含む方法を提供する。これらの実施形態では、試料の核酸分子およびエピジェネティック対照核酸分子の区画を必ず同時に行う。一部の実施形態では、解析ステップは、区画化セットのうちの少なくとも1つにおける所与のエピジェネティック状態にある内在性対照分子の数/フラクションを推定することを含む。

10

#### 【0157】

図4は、ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法400の実施形態例を例示する。この実施形態では、ポリヌクレオチドの試料中の内在性対照分子の区画を解析して、区画方法を評価する。特定のエピジェネティック状態を有するヒトゲノム内の領域が存在し、その領域のエピジェネティック状態は多くの場合に変動/変化せず、常に同じままである/異なる対象および/または異なる疾患の型/病期の場合でも一貫したままである。そのようなエピジェネティック状態が不変のヒトゲノム領域に対応するポリヌクレオチドの試料中の核酸分子は、内在性対照分子と称される。401では、対象由来のポリヌクレオチドの試料を考慮に入れる。402では、ポリヌクレオチドの試料の少なくとも1つのサブセットの核酸分子を、分子のエピジェネティック状態に基づいて複数の区画化セットに区画または分別する。区画は、エピジェネティック修飾の存在または非存在に基づき得る、かつ/またはエピジェネティック修飾の度合いに基づき得る。エピジェネティック修飾の例としては、これだけに限定されないが、メチル化の存在または非存在、メチル化のレベルおよびメチル化の型(5'シトシン)を挙げることができる。一部の実施形態では、エピジェネティック修飾は、DNAメチル化であり得る。これらの実施形態では、スパイクイン試料の分子を異なるメチル化のレベル(異なるメチル化ヌクレオチドの数)に基づいて区画する。一部の実施形態では、スパイクイン試料を2つまたはそれよりも多くの区画化セット(例えば、少なくとも3、4、5、6、または7つの区画化セット)に区画することができる。一部の実施形態では、区画は、核酸分子の結合性物質に対する示差的な結合親和性に基づく。

20

30

#### 【0158】

核酸分子の区画を、区画された核酸分子をシーケンシングすることによって、またはデジタル液滴PCR(ddPCR)によって解析することができる。区画を解析する前に、区画化セット内の核酸分子を富化させることができ、その結果、目的の核酸分子からのシグナルを増大させ、したがって、感度を改善することができる。403では、複数の区画化セット内の核酸分子の少なくとも1つのサブセットを富化させ、したがって、目的の領域に属する内在性対照分子(ポリヌクレオチドの試料由来)および他のポリヌクレオチドの試料由来の核酸分子を富化させる。

40

#### 【0159】

一部の実施形態では、富化の前に、複数の区画化セットのそれぞれに示差的にタグ付ける。次いで、タグ付けされた区画化セットを、集合的な試料調製および/またはシーケンシングのために一緒にプールする。区画化セットへの示差的なタグ付けは、特定の区画化セットに属する核酸分子の追跡を続けるのに役立つ。タグは通常、アダプターの構成成分として提供される。異なる区画化セット内の核酸分子は、1つの区画化セットのメンバーを別の区画化セットのメンバーと区別することができる異なるタグを受ける。同じ区画

50

セットの核酸分子と連結したタグは互いに同じであっても異なってもよい。しかし、互いに異なる場合には、タグは、それらのタグが付着した分子が特定の区画化セットに属すると識別されるように、それらの配列の一部が共通するものであり得る。

【0160】

404では、富化された分子の少なくとも1つのサブセットをシーケンシングする。得られる配列情報は、核酸分子の配列および核酸分子に付着したタグを含む。核酸分子に付着したタグの配列から、タグを核酸分子の区画化セットと関連させることができる。配列情報を使用して、内在性対照分子およびそれらの対応する区画化セットを識別する。この情報を使用して、内在性対照分子の区画を解析する。405では、1つまたは複数の区画化セットに属する内在性対照分子の1つまたは複数のエピジェネティック区画スコアを決定する。一部の実施形態では、区画方法の感度および/または特異度をエピジェネティック区画スコアによって査定することができる。エピジェネティック区画スコアは、特定のエピジェネティック状態に属する核酸分子の区画を表すスコアである。一部の実施形態では、エピジェネティック状態に属する核酸分子のエピジェネティック区画スコアを各区画化セットについて決定する。例えば、特定のエピジェネティック状態に属する内在性対照分子のエピジェネティック区画スコアを決定することができる。エピジェネティック区画スコアは、特定のエピジェネティック状態に属する核酸分子の数（または統計学的に推定される数）の尺度であり得る。エピジェネティック区画スコアは、フラクションまたはパーセンテージによるものであり得る。エピジェネティック区画スコアは、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属する内在性対照分子の数の他の残りの区画化セット（単数または複数）内に存在するそのエピジェネティック状態に属する内在性対照分子の数に対する比の尺度であり得る。一部の実施形態では、エピジェネティック区画スコアは、少なくとも1つの区画化セットに区画される特定のエピジェネティック状態に属する内在性対照分子の数の、全ての区画化セット内のそのエピジェネティック状態に属する内在性対照分子の総数に対するフラクションまたはパーセンテージであり得る。一部の実施形態では、エピジェネティック区画スコアを、区画化セットのそれぞれにおける内在性対照分子の各エピジェネティック状態に関して決定する。一部の実施形態では、エピジェネティック区画スコアを、1つまたは複数の区画化セット内の1つまたは複数の特定のエピジェネティック状態を有する内在性対照分子に関して決定する。一部の実施形態では、エピジェネティック区画スコアを、特定の区画化セット内の特定の

10

20

30

【0161】

一部の実施形態では、エピジェネティック区画スコアは、CGを有さない（「ゼロ」CG）分子が高区画化セットに区画される有効度を対象とし得る。このスコアは、0 CGスコアと称することができる。一部の実施形態では、0 CGスコアを高区画化セット内のCGを有さない分子のフラクションまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、以下のうちの少なくとも1つを有する、高メチル化対照分子のフラクションの尺度であり得る：

高メチル化区画化セット（すなわち、高度にメチル化された区画化セット）において、  
 (xi) 1つのメチルCG（エピジェネティック区画スコアを1 CGスコアと称することができる）、

40

(xii) 2つのメチルCG（エピジェネティック区画スコアを2 CGスコアと称することができる）、

(xiii) 3つのメチルCG（エピジェネティック区画スコアを3 CGスコアと称することができる）、

(xiv) 4つのメチルCG（エピジェネティック区画スコアを4 CGスコアと称することができる）および

(xv) 5つのメチルCG（エピジェネティック区画スコアを5 CGスコアと称することができる）。

【0162】

50

一部の実施形態では、エピジェネティック区画スコアは、低メチル化対照分子が高メチル化区画化セットに区画される有効度を対象とし得る。このスコアは、*hypo*スコアと称することができる。一部の実施形態では、*hypo*スコアを高メチル化区画化セット内の低メチル化対照分子のフラクションまたはパーセンテージで表すことができる。一部の実施形態では、エピジェネティック区画スコアは、低メチル化区画化セット内の高メチル化対照分子が5%未満になるのに必要なメチル化CGの数の尺度であり得る。このスコアを、*methyl-5*と称することができる。一部の実施形態では、エピジェネティック区画スコアは、高メチル化区画化セット内の高メチル化対照分子が少なくとも50%になるのに必要なメチル化CGの数の尺度であり得る。このスコアを、*methyl-half*と称することができる。

10

**【0163】**

例えば、内在性対照分子の2つのサブセット（サブセットAおよびB）を解析し、各サブセットはメチル化のレベル/度合い（すなわち、各サブセットは異なるエピジェネティック状態を有する）が異なる。これらの2つのサブセット内の内在性対照分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。各サブセットについて、エピジェネティック区画スコアを、区画化セットのそれぞれ（P1、P2およびP3）について決定する。すなわち、サブセットAに属するエピジェネティック対照核酸分子は、3つのエピジェネティック区画スコア、つまり3つの区画化セット、P1、P2およびP3のそれぞれについて1つのエピジェネティック区画スコアを有する。同様に、サブセットBは、3つのエピジェネティック区画スコア、つまり3つの区画化セットP1、P2およびP3のそれぞれについて1つのエピジェネティック区画スコアを有する。

20

**【0164】**

別の実施形態では、内在性対照分子の3つのサブセット（サブセットA、BおよびC）を解析し、各サブセットはメチル化のレベル/度合い（すなわち、各サブセットは異なるエピジェネティック状態を有する）が異なる。これらの3つのサブセット内の内在性対照分子を、それらのメチル結合タンパク質に対する結合親和性に基づいて3つの区画化セット、P1、P2およびP3に区画することができる。この実施形態では、P1区画化セット内のサブセットAの内在性分子についてのみエピジェネティックスコアを決定する。このエピジェネティックスコアは、P1区画化セット内のサブセットAの内在性対照分子のサブセットA内在性対照分子の総数（P1、P2およびP3の区画化セット内の）に対するフラクションまたはパーセンテージの尺度であり得る。

30

**【0165】**

エピジェネティック区画スコアは、0~1の間（フラクションで）または0~100%の間（パーセンテージで）の任意の値または範囲であり得る。一部の実施形態では、エピジェネティック区画スコアは、メチル化CGの数によるものであり得る（例えば、*methyl-half*および*methyl-5*について）

**【0166】**

406では、内在性対照分子のエピジェネティック区画スコアを、それらの対応するエピジェネティック区画カットオフ（所定のカットオフ）と比較して、区画方法を評価する。エピジェネティック区画カットオフは、特定のエピジェネティック状態に属する核酸分子の区画を評価するために使用される所定のカットオフ値またはカットオフ範囲であり、各区画化セットは、エピジェネティック状態に属する核酸分子についてのエピジェネティック区画カットオフを有する。エピジェネティック区画カットオフは、核酸分子および区画化セットのエピジェネティック状態によって異なる、すなわち、各エピジェネティック状態がそれ自体のエピジェネティック区画カットオフを有し、区画化セットのいずれもがそのエピジェネティック状態についての別々のエピジェネティック区画カットオフを有する。カットオフは、パーセンテージまたはフラクションによるものであり得、カットオフは、特定のカットオフ値ではなくカットオフ範囲であり得る。例えば、特定のエピジェネティック状態に属する内在性対照分子についてのエピジェネティック区画カットオフは、

40

50

区画化セット P 1、P 2 および P 3 について、それぞれ 70% から 79% の間、10% から 15% の間および 5% 未満であり得る。そのエピジェネティック状態に属する内在性対照分子のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法は成功である。

【0167】

一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.01%、0.02%、0.05%、0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、5%、少なくとも5% または少なくとも10% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.01% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.02% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.03% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.04% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.05% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.1% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.2% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.3% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.4% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.5% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.6% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.7% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.8% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、0.9% であり得る。一部の実施形態では、0 CGスコアについてのエピジェネティック区画カットオフは、1% であり得る。

【0168】

一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.1%、0.5%、1%、2%、3%、4%、5%、7% または少なくとも10% であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.1% であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、0.5% であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、1% であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、2% であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、3% であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、4% であり得る。一部の実施形態では、hypoスコアについてのエピジェネティック区画カットオフは、5% であり得る。

【0169】

一部の実施形態では、methyl-5 についてのエピジェネティック区画カットオフは、5、10、20、30、40 または 50 mCG であり得る。一部の実施形態では、methyl-5 についてのエピジェネティック区画カットオフは、5 mCG であり得る。一部の実施形態では、methyl-5 についてのエピジェネティック区画カットオフは、10 mCG であり得る。一部の実施形態では、methyl-5 についてのエピジェネティック区画カットオフは、20 mCG であり得る。一部の実施形態では、methyl-5 についてのエピジェネティック区画カットオフは、30 mCG であり得る。一部の実施形態では、methyl-5 についてのエピジェネティック区画カットオフは、40 mCG であり得る。一部の実施形態では、methyl-5 についてのエピジェネティック

10

20

30

40

50

区画カットオフは、50 mCGであり得る。

【0170】

一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、5、10、15、20、25、30、35または40 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、5 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、10 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、15 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、20 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、25 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、30 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、35 mCGであり得る。一部の実施形態では、methyl-halfスコアについてのエピジェネティック区画カットオフは、40 mCGであり得る。

10

【0171】

一部の実施形態では、1つまたは複数の区画化セット内の1つまたは複数のエピジェネティック状態に属する内在性対照分子の1つまたは複数のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には、区画方法を成功に分類することができる。そうでなければ、区画方法を失敗に分類することができる。

20

【0172】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子（単数または複数）のエピジェネティック状態を決定するための方法であって、（a）エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それによって、スパイクイン試料を作製するステップと、（b）スパイクイン試料の少なくとも1つのサブセットの核酸分子を複数の区画化セットに区画するステップと、（c）複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含む、ステップと、（d）富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、（e）配列リードのセットの少なくとも1つのサブセットを解析して、複数の区画化セット内の異なるエピジェネティック状態にあるエピジェネティック対照核酸分子の複数の区画プロファイルを生成するステップと、（f）エピジェネティック対照核酸分子の複数の区画化プロファイルを使用して、試料の核酸分子のエピジェネティック状態に関する確率を推定するステップとを含む方法を提供する。これらの実施形態では、試料の核酸分子およびエピジェネティック対照核酸分子の区画を必ず同時に行う。

30

【0173】

一部の実施形態では、解析ステップは、複数の区画化セット内のエピジェネティック状態当たりのエピジェネティック対照核酸分子の数またはフラクションを決定することを含む。区画プロファイルは、2つまたはそれよりも多くの区画化セット内の各エピジェネティック状態にあるエピジェネティック対照核酸分子のフラクション/数の表現を指し得る。一部の実施形態では、区画プロファイルは、エピジェネティック対照核酸分子内のエピジェネティック修飾を有するヌクレオチドの数、エピジェネティック対照核酸分子内のエピジェネティック修飾を有するヌクレオチドの位置および/またはエピジェネティック対照核酸分子の配列組成に関する情報をさらに含む。この区画プロファイルを試料中の核酸分子のエピジェネティック状態に関する確率の推定に使用することができる。一部の実施形態では、エピジェネティック修飾がメチル化である場合には、区画プロファイルを試料の核酸分子のメチル化の状態（すなわち、メチル化のレベル/度合いまたはメチル化ヌク

40

50

レオチドの数)に関する確率の推定に使用することができる。

【0174】

別の態様では、本開示は、ポリヌクレオチドの試料中の核酸分子(単数または複数)のエピジェネティック状態を決定するための方法であって、(a)試料の少なくとも1つのサブセット由来の核酸分子を複数の区画化セットに区画するステップと、(b)複数の区画化セットから分子の少なくとも1つのサブセットを富化させて、富化された分子のセットを生成するステップであって、富化された分子のセットが、ポリヌクレオチドの試料由来の核酸分子の群を含み、ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、(c)富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、(e)配列リードのセットの少なくとも1つのサブセットを解析して、異なる複数の区画化セット内のエピジェネティック状態にある内在性対照分子の複数の区画プロファイルを生成するステップと、(f)内在性対照分子の複数の区画化プロファイルを使用して、核酸分子のエピジェネティック状態に関する確率を推定するステップとを含む方法を提供する。

10

【0175】

一部の実施形態では、解析ステップは、複数の区画化セット内のエピジェネティック状態当たりの内在性対照分子の数を決定することを含む。区画プロファイルは、2つまたはそれよりも多くの区画化セット内の各エピジェネティック状態にある内在性対照分子のフラクション/数の表現を指し得る。一部の実施形態では、区画プロファイルは、エピジェネティック対照核酸分子内のエピジェネティック修飾を有するヌクレオチドの数、エピジェネティック対照核酸分子内のエピジェネティック修飾を有するヌクレオチドの位置および/またはエピジェネティック対照核酸分子の配列組成に関する情報をさらに含む。一部の実施形態では、内在性対照分子内のメチル化CpGの数を以前の実験データに基づいておよび/または文献から決定する。この区画プロファイルを試料中の核酸分子のエピジェネティック状態に関する確率の推定に使用することができる。一部の実施形態では、エピジェネティック修飾がメチル化である場合には、区画プロファイルを試料の核酸分子のメチル化の状態(すなわち、メチル化のレベル/度合いまたはメチル化ヌクレオチドの数)に関する確率の推定に使用することができる。

20

【0176】

一部の実施形態では、内在性対照分子(例えば、高メチル化対照分子および低メチル化対照分子)を使用して、試料の核酸分子のメチル化の状態を推定することができる。3つの区画化セット、P1、P2およびP3が存在する場合、高メチル化対照分子の区画プロファイルをP1、P2およびP3について、3つの区画化セットのそれぞれにおける高メチル化対照分子のフラクションおよび高メチル化対照分子内に存在するメチル化CpGの数に基づいて生成することができる。同様に、低メチル化対照分子について、低メチル化対照分子の区画プロファイルをP1、P2およびP3について、3つの区画化セットのそれぞれにおける低メチル化対照分子のフラクションおよび低メチル化対照分子内に存在するメチル化されていないCpGの数に基づいて生成することができる。一部の実施形態では、内在性対照分子を使用する場合、内在性対照分子内のメチル化CpGの数を以前の実験データに基づいておよび/または文献から決定する。これらの6つの区画プロファイルを試料の核酸分子内の特定の領域に存在するメチル化のレベル/度合いまたはメチル化ヌクレオチドの数に関する確率の推定に使用することができる。

30

40

【0177】

一部の実施形態では、エピジェネティック対照核酸分子(例えば、高度にメチル化されたエピジェネティック対照核酸分子および低度にメチル化されたエピジェネティック対照核酸分子)を使用して、試料の核酸分子のメチル化の状態を推定することができる。3つの区画化セット、P1、P2およびP3が存在する場合、高度にメチル化されたエピジェネティック対照核酸分子の区画プロファイルをP1、P2およびP3について、3つの区画化セットのそれぞれにおける高度にメチル化されたエピジェネティック対照核酸分子のフラクションおよび高度にメチル化されたエピジェネティック対照核酸分子内に存在する

50

メチル化 CpG の数に基づいて生成することができる。同様に、低度にメチル化されたエピジェネティック対照核酸分子について、低度にメチル化されたエピジェネティック対照核酸分子の区画プロファイルを P 1、P 2 および P 3 について、3 つの区画化セットのそれぞれにおける低度にメチル化されたエピジェネティック対照核酸分子のフラクションおよび低度にメチル化されたエピジェネティック対照核酸分子内に存在するメチル化されていない CpG の数に基づいて生成することができる。これらの 6 つの区画プロファイルを試料の核酸分子内の特定の領域に存在するメチル化のレベル/度合いまたはメチル化ヌクレオチドの数に関する確率の推定に使用することができる。

#### II. エピジェネティック対照核酸分子

##### 【0178】

エピジェネティック対照核酸分子を、エピジェネティック修飾に基づいた試料中の核酸分子の区画を評価するための対照または参照分子として使用する。これらのエピジェネティック対照核酸分子を、試料中の核酸分子（単数または複数）のエピジェネティック状態を決定するために使用することもできる。例えば、エピジェネティック修飾は、DNA メチル化であり得、エピジェネティック対照核酸分子は、異なる/区別可能なメチル化のレベルを有し得る。エピジェネティック対照核酸分子は合成オリゴヌクレオチドであり得る。一部の実施形態では、エピジェネティック対照核酸分子は、天然には存在しない核酸配列を有し得る。一部の実施形態では、エピジェネティック対照核酸分子は、天然に存在する核酸配列を有し得る。一部の実施形態では、エピジェネティック対照核酸分子は、非ヒトゲノムに対応する核酸配列を有し得る。例えば、これらの分子は、(i) ラムダファージ DNA またはヒトゲノムの領域に対応する配列、(ii) 天然には存在しない配列、および/または (iii) (i) と (ii) の組合せのいずれかを有し得る。また、エピジェネティック対照核酸分子をサブセットに群分けすることができ、各サブセットはエピジェネティック修飾の度合いを表す特定の数のヌクレオチドを有し得、その数は、あらゆる他のセットにおけるエピジェネティック修飾の度合いを表すヌクレオチドの数とは異なる。

##### 【0179】

別の態様では、本開示は、1 つまたは複数のエピジェネティック対照核酸分子のサブセットを含むエピジェネティック対照核酸分子のセットであって、各サブセットが複数のエピジェネティック対照核酸分子を含み、各エピジェネティック対照核酸分子がエピジェネティック修飾領域を含む、エピジェネティック対照核酸分子のセットを提供する。エピジェネティック修飾領域は、エピジェネティック対照核酸分子のエピジェネティック状態を表すエピジェネティック対照核酸分子の領域である。エピジェネティック状態は、核酸分子のエピジェネティック修飾のレベル/度合いである。例えば、エピジェネティック修飾が DNA メチル化である場合には、エピジェネティック状態は、高度にメチル化された、低度にメチル化されたまたは中程度にメチル化された核酸分子を指し得る。エピジェネティック状態は、エピジェネティック修飾を有するヌクレオチドの数も指し得る。例えば、エピジェネティック修飾が DNA メチル化である場合には、エピジェネティック状態は、核酸分子のメチル化ヌクレオチドの数を指し得る。

##### 【0180】

一部の実施形態では、エピジェネティック対照核酸分子は、以下のうちの少なくとも 1 つを含む：(i) エピジェネティック修飾領域および(ii) 識別子領域。一部の実施形態では、エピジェネティック修飾領域は、エピジェネティック修飾を有するヌクレオチドを含む。一部の実施形態では、エピジェネティック修飾は、DNA メチル化である。これらの実施形態では、エピジェネティック対照核酸分子のエピジェネティック修飾領域は、メチル化されたヌクレオチドを有し得る。エピジェネティック修飾領域内のメチル化ヌクレオチドの数は、エピジェネティック対照核酸分子の間で変動し得る。一部の実施形態では、エピジェネティック対照核酸分子は、エピジェネティック修飾領域内に 0 個、1 個、2 個、3 個、4 個、5 個、6 個、7 個、8 個、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 30 個、少なくとも 40 個または少なくとも 50 個のメチル化ヌクレオチドを有し得る。エピジェネティック対照核酸分子は、エピジェネティ

10

20

30

40

50

ック修飾領域内のエピジェネティック状態（すなわち、エピジェネティック修飾を有するヌクレオチドの数/エピジェネティック修飾のレベル）に基づいてサブセットに群分けすることができる。異なるサブセットの中でのエピジェネティック修飾領域の長さは、同じ、例えば、およそ160bpであり得る。サブセット間でエピジェネティック修飾領域の長さは異なり得る。例えば、エピジェネティック対照核酸分子を、エピジェネティック修飾領域内のメチル化ヌクレオチドの数に基づいて3つのサブセット（サブセットA、BおよびC）に群分けすることができる。サブセットA、BおよびCは、それぞれ、エピジェネティック修飾領域内にメチル化ヌクレオチドを5個、10個および15個有するエピジェネティック対照核酸分子を有し得、サブセットA、BおよびCにおけるエピジェネティック修飾領域の長さは、同じ（例えば、160bp）であり得る、または異なり得る - サブセットA、BおよびCについて、それぞれ100bp、150bpおよび200bp。

10

#### 【0181】

ある特定の実施形態では、エピジェネティック対照核酸分子をサブセットに群分けすることができ、各サブセットはエピジェネティック修飾の度合いを表し、各サブセット内のポリヌクレオチドの数はあらゆる他のセット内のヌクレオチドの数とは異なる。一部の実施形態では、サブセット内のメチル化ヌクレオチドの数は、0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、11個、少なくとも12個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも40個または少なくとも50個である。一部の実施形態では、少なくとも1つのサブセット内のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、エピジェネティック修飾を有するヌクレオチドを少なくとも1つ含む。一部の実施形態では、少なくとも1つのエピジェネティック対照核酸分子のサブセットは、エピジェネティック対照核酸分子のエピジェネティック修飾領域内にいかなるエピジェネティック修飾（すなわち、エピジェネティック修飾されていないヌクレオチド）も有さないヌクレオチドを含み得る。

20

#### 【0182】

一部の実施形態では、サブセット内のあらゆるエピジェネティック対照核酸分子のエピジェネティック修飾領域は、エピジェネティック修飾を有するヌクレオチドを同数含む。一部の実施形態では、第1のサブセット内のエピジェネティック修飾を有するヌクレオチドの数は、第2のサブセット内のエピジェネティック修飾を有するヌクレオチドの数と異なる。一部の実施形態では、1つまたは複数のサブセット内の複数のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、同一の核酸配列を含む。一部の実施形態では、第1のサブセット内の複数のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、第2のサブセット内の複数のエピジェネティック対照核酸分子のエピジェネティック修飾領域の核酸配列とは区別可能な核酸配列を含む。

30

#### 【0183】

一部の実施形態では、1つまたは複数のサブセット内のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、同じ長さであり、かつ同じ配列組成を有するが、エピジェネティック修飾を有するヌクレオチドの数は、1つまたは複数のサブセットのそれぞれにおいて異なり得る。一部の実施形態では、1つまたは複数のサブセット内のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、同じ長さであり、かつ、エピジェネティック修飾を有するヌクレオチドを同数有するが、エピジェネティック修飾を有するヌクレオチドの位置は、1つまたは複数のサブセットのそれぞれにおいて異なり得る。一部の実施形態では、1つまたは複数のサブセット内のエピジェネティック対照核酸分子のエピジェネティック修飾領域は同じ長さであり、エピジェネティック修飾を有するヌクレオチドを同数有し、かつエピジェネティック修飾を有するヌクレオチドの位置が同じであり得るが、エピジェネティック修飾を有するヌクレオチドの両側に近接するヌクレオチドは、1つまたは複数のサブセットのそれぞれにおいて異なり得る。

40

#### 【0184】

一部の実施形態では、エピジェネティック対照核酸分子の各サブセットは、等モル濃度で存在する。一部の実施形態では、エピジェネティック対照核酸分子の各サブセットは、

50

非等モル濃度で存在する。一部の実施形態では、エピジェネティック修飾は、DNAメチル化である。一部の実施形態では、エピジェネティック修飾を有するヌクレオチドは、メチル化ヌクレオチドを含む。一部の実施形態では、メチル化ヌクレオチドは、5-メチルシトシンを含む。一部の実施形態では、メチル化ヌクレオチドは、5-ヒドロキシメチルシトシンを含む。一部の実施形態では、メチル化ヌクレオチドは、N<sup>6</sup>-メチルアデニンを含む。

#### 【0185】

一部の実施形態では、エピジェネティック対照核酸分子は、識別子領域をさらに含む。識別子領域は、エピジェネティック対照核酸分子を他のエピジェネティック対照核酸分子と区別するために使用される、エピジェネティック対照核酸分子の領域である。識別子領域は、分子バーコードおよび/またはエピジェネティック状態バーコードを有し得る。識別子領域は、エピジェネティック修飾領域の片側または両側に存在し得る。分子バーコードは、エピジェネティック対照核酸分子の識別子としての機能を果たし、一方、エピジェネティック状態バーコードは、エピジェネティック対照核酸分子のエピジェネティック状態の識別子としての機能を果たす。エピジェネティック状態バーコードは、エピジェネティック対照核酸分子のエピジェネティック状態を識別するために使用されるバーコード(核酸配列)の一種である。一部の実施形態では、エピジェネティック状態バーコードは、エピジェネティック対照核酸分子のエピジェネティック修飾領域内のエピジェネティック修飾を有するヌクレオチドの数を識別する(所定の相関によって)ことができるものである。一部の実施形態では、エピジェネティック状態バーコードは、エピジェネティック対照核酸分子のエピジェネティック修飾領域内のエピジェネティック修飾のレベルを識別することができるものである。一部の実施形態では、エピジェネティック対照核酸分子の識別子領域は、エピジェネティック状態バーコードを含む。例えば、エピジェネティック修飾がDNAメチル化であり、エピジェネティック対照核酸分子のサブセットがメチル化ヌクレオチドを5つ有する場合には、そのサブセットに入るエピジェネティック対照核酸分子の全てが同じエピジェネティック状態バーコードを有することになる。一部の実施形態では、エピジェネティック状態バーコードを使用して、エピジェネティック対照核酸分子のエピジェネティック修飾領域のエピジェネティック修飾のレベル/度合いを識別することができる。エピジェネティック対照核酸分子をエピジェネティック修飾領域内のシトシンまたはCpGヌクレオチドの数に基づいてサブセットに群分けすることができる。一部の実施形態では、各サブセット内でメチル化のレベルが変動し得(例えば、高度にメチル化された、中程度にメチル化された、および低度にメチル化された)、メチル化の各レベルは別々のエピジェネティック状態バーコードを有し得る。例えば、サブセットA内で、低メチル化のエピジェネティック対照核酸分子は全てがエピジェネティック状態バーコード、例えばESB1を有し、高度にメチル化されたエピジェネティック対照核酸分子は全てが別のエピジェネティック状態バーコード、例えばESB3を有することになる。本実施例では、エピジェネティック状態バーコードを使用して、メチル化のレベル/度合いを識別する。識別子領域内の分子バーコードは一意的なバーコード(各分子が一意的なバーコードを有する)または非一意的なバーコードであり得る。分子バーコードは、2ヌクレオチドから50ヌクレオチドの間の任意の長さであり得る。一部の実施形態では、分子バーコードは、少なくとも2ヌクレオチド、少なくとも3ヌクレオチド、少なくとも4ヌクレオチド、少なくとも5ヌクレオチド、少なくとも6ヌクレオチド、少なくとも7ヌクレオチド、少なくとも8ヌクレオチド、少なくとも9ヌクレオチド、または少なくとも10ヌクレオチドであり得る。一部の実施形態では、エピジェネティック状態バーコードは、少なくとも2ヌクレオチド、少なくとも3ヌクレオチド、少なくとも4ヌクレオチド、少なくとも5ヌクレオチド、少なくとも6ヌクレオチド、少なくとも7ヌクレオチドまたは少なくとも8ヌクレオチドであり得る。

#### 【0186】

図5は、本開示の一部の実施形態での使用に適したエピジェネティック対照核酸分子の略図である。本明細書に記載のエピジェネティック対照核酸分子は、アッセイされる試料

10

20

30

40

50

のエピジェネティック対照核酸分子と同様の長さを有し、あらゆる配列特異的な区画の影響を低減するためにサブセットの全てが同じ配列組成を有する。図5では、例として、エピジェネティック対照核酸分子が4つのサブセット、サブセット1、2、3および4に群分けされている。図5のエピジェネティック対照核酸分子は二本鎖DNA分子である。例示のために、各サブセット内のエピジェネティック対照核酸分子の1つの代表のみが図に示されている。この実施形態では、エピジェネティック対照核酸分子のエピジェネティック修飾領域の配列はサブセットの全てにおいて同じである。4つのサブセット全てにおいてエピジェネティック対照核酸分子のエピジェネティック修飾領域は5つのCpG対を有する。二本鎖DNA配列内の「- - -」領域はCpG対とは別の任意の他の配列を表し、Mは5-メチルシトシンを表し、Cはシトシンを表し、Gはグアニンを表す。図5では、サブセット内のエピジェネティック対照核酸分子のエピジェネティック状態（メチル化のレベル）は、他のサブセットのエピジェネティック状態とは異なる。サブセット1は、メチル化CpG対がゼロであり、サブセット2は完全にメチル化されたCpG対を1つ有し、サブセット3は完全にメチル化されたCpG対を3つ有し、サブセット4は完全にメチル化されたCpG対を5つ有する。この実施形態では、識別子領域はエピジェネティック修飾領域の両側に存在する。両側の識別子領域はエピジェネティック状態バーコード（ESB）を有し、一方、分子バーコード（MB）は片側のみにある。分子バーコードはエピジェネティック対照核酸分子の識別子として使用され、各エピジェネティック対照核酸分子が一意的な分子バーコードを有する（すなわち、分子1はMB1を有し、分子2はMB2を有し、分子3はMB3を有するなど）。エピジェネティック状態バーコードをエピジェネティック対照核酸分子のエピジェネティック状態の識別子として使用することができる。ここで、エピジェネティック状態バーコードを使用して、エピジェネティック対照核酸分子内の完全にメチル化されたCpG対の数を識別する。サブセット1のエピジェネティック対照核酸分子は全て、メチル化CpG対がゼロであり、したがって、サブセット1のエピジェネティック対照核酸分子は全て、同じエピジェネティック状態バーコード - ESB1を有する。同様に、サブセット2、3および4のエピジェネティック対照核酸分子は全て、完全にメチル化されたCpG対をそれぞれ1つ、3つおよび5つ有する。したがって、サブセット2、3および4のエピジェネティック対照核酸分子は全て、それぞれESB2、ESB3およびESB4であるエピジェネティック状態バーコードを有する。本実施例では、同じエピジェネティック状態バーコードはエピジェネティック修飾領域の両側に存在する。

#### 【0187】

一部の実施形態では、分子バーコードは、エピジェネティック修飾領域の片側または両側にあり得る。一部の実施形態では、エピジェネティック状態バーコードは、エピジェネティック修飾領域の片側または両側にあり得る。一部の実施形態では、エピジェネティック修飾領域の両側のエピジェネティック状態バーコードは同じもしくは異なり得、かつ/またはランダムに付着させることができる。

#### 【0188】

一部の実施形態では、識別子領域は、1つまたは複数のプライマーの結合を容易にする追加的な領域（プライマー結合部位）を有し得る。一部の実施形態では、1つのサブセットにおける識別子領域プライマー結合部位は、他のサブセットにおけるプライマー結合部位とは異なる。一部の実施形態では、サブセット内でエピジェネティック対照核酸分子が異なるエピジェネティック状態を有する場合には、プライマー結合部位は、分子内の各エピジェネティック状態について異なり得る、すなわち、一意的なエピジェネティック状態のそれぞれが一意的なプライマー結合部位を有する。一部の実施形態では、これらのプライマー結合部位をエピジェネティック対照核酸分子の区画の解析に使用する。一部の実施形態では、エピジェネティック対照核酸分子の区画をシーケンシングによって解析する代わりに、エピジェネティック対照核酸分子の区画を、これらのプライマー結合性状態に結合するプライマーを使用したデジタル液滴PCR（ddPCR）によって解析することができる。

10

20

30

40

50

## 【 0 1 8 9 】

一部の実施形態では、エピジェネティック対照核酸分子をサブセットに、各サブセット内のエピジェネティック対照核酸分子が配列を有するが、各サブセット内のエピジェネティック対照核酸分子のエピジェネティック状態は変動し得るように群分けすることができる。

## 【 0 1 9 0 】

図6は、本開示のある特定の実施形態での使用に適し得るエピジェネティック対照核酸分子の略図である。本明細書に記載のエピジェネティック対照核酸分子は、核酸分子の区画の間のC p G対の配列組成および数/完全にメチル化されたC p G対の影響も考慮に入れることができる。図6では、例として、エピジェネティック対照核酸分子が3つのサブセット、サブセット1、2および3に群分けされている。図6のエピジェネティック対照核酸分子は二本鎖DNA分子である。例示のために、各サブセット内のあらゆるエピジェネティック状態についてエピジェネティック対照核酸分子の1つの代表のみが図に示されている。この実施形態では、サブセット1、2および3内のエピジェネティック対照核酸分子のエピジェネティック修飾領域は長さが異なる。サブセット1、2および3内のエピジェネティック対照核酸分子のエピジェネティック修飾領域は、C p G対をそれぞれ1つ、3つおよび5つ有する。二本鎖DNA配列内の‘ - - - ’領域はC p G対とは別の任意の他の配列を表し、Mは5 - メチルシトシンを表し、Cはシトシンを表し、Gはグアニンを表す。図6では、各サブセット内で、エピジェネティック対照核酸分子は、異なるエピジェネティック状態、例えば、低度にメチル化された、中程度にメチル化されたおよび高度にメチル化された状態にある。サブセット1のエピジェネティック対照核酸分子は、2つの異なるエピジェネティック状態、つまり低度にメチル化された状態（メチル化C p G対がゼロ）および高度にメチル化された状態（完全にメチル化されたC p G対が1つ）にある。サブセット2のエピジェネティック対照核酸分子は、3つの異なるエピジェネティック状態、つまり低度にメチル化された状態（ゼロメチル化C p G対）、中程度にメチル化された（完全にメチル化されたC p G対が1つ）および高度にメチル化された状態（完全にメチル化されたC p G対が3つ）にある。サブセット3のエピジェネティック対照核酸分子は、3つの異なるエピジェネティック状態、つまり低度にメチル化された状態（完全にメチル化されたC p G対が1つ）、中程度にメチル化された状態（完全にメチル化されたC p G対が3つ）および高度にメチル化された状態（完全にメチル化されたC p G対が5つ）にある。ここで、識別子領域はエピジェネティック修飾領域の両側に存在する。両側の識別子領域はエピジェネティック状態バーコード（ESB）および分子バーコード（MB）を有する。分子バーコードはエピジェネティック対照核酸分子の識別子として使用され、各エピジェネティック対照核酸分子が一意的な分子バーコードを有する（すなわち、分子1はMB1を有し、分子2はMB2を有し、分子3はMB3を有するなど）。エピジェネティック状態バーコードをエピジェネティック対照核酸分子のエピジェネティック状態の識別子として使用する。ここで、エピジェネティック状態バーコードを使用して、エピジェネティック対照核酸分子の度合い/メチル化のレベル、すなわち、低度にメチル化された状態であるか、中程度にメチル化された状態であるか、または高度にメチル化された状態であるかを識別する。サブセット1、2および3内の低度にメチル化されたエピジェネティック対照核酸分子は全て同じエピジェネティック状態バーコード - ESB1を有する。サブセット2および3は中程度にメチル化されたエピジェネティック対照核酸分子を有し、これらの分子は全てが同じエピジェネティック状態バーコード - ESB2を有する（サブセット1は中程度にメチル化された状態を有さず、したがって、ESB2エピジェネティック状態バーコードを有するエピジェネティック対照核酸分子は存在しない）。したがって、エピジェネティック対照核酸分子の配列およびエピジェネティック状態バーコードの配列から、エピジェネティック対照核酸分子のエピジェネティック状態およびエピジェネティックな核酸分子が属するサブセットを識別することができる。

## 【 0 1 9 1 】

さらに、識別子領域はプライマー結合性部位を有し得る。異なるプライマー結合部位を

10

20

30

40

50

、各サブセット内およびサブセット間で異なるエピジェネティック状態を弁別するために使用することができる。例えば、サブセット1内の低度にメチル化されたエピジェネティック対照核酸分子は、エピジェネティック修飾領域の両側にプライマー結合部位 - Pr 1 および Pr 2 を有し得る。サブセット1内の高度にメチル化されたエピジェネティック対照核酸分子は、エピジェネティック修飾領域の両側にプライマー結合部位 - Pr 3 および Pr 4 を有し得る。同様に、サブセット2では、低度、中程度および高度メチル化エピジェネティック対照核酸分子は、それぞれ、エピジェネティック修飾領域の両側にプライマー結合部位 Pr 5 & Pr 6、Pr 7 & Pr 8 および Pr 9 & Pr 10 を有し得る。同様に、サブセット3では、低度、中程度および高度メチル化エピジェネティック対照核酸分子は、それぞれ、エピジェネティック修飾領域の両側にプライマー結合部位 Pr 11 & Pr 12、Pr 13 & Pr 14 および Pr 15 & Pr 16 を有し得る。また、異なるサブセット内の異なるエピジェネティック状態分子に対して使用した別個のプライマーセットから、特定のサブセット内の特定のエピジェネティック状態に属するエピジェネティック対照核酸分子の数の尺度を ddPCR または定量的 PCR (qPCR) のいずれかによって推定することができる。この実施形態では、エピジェネティック状態バーコード配列およびエピジェネティック修飾領域の配列から、エピジェネティック修飾領域内の CpG 対の数およびエピジェネティック修飾領域内の完全にメチル化された CpG 対の数を決定することができる。

10

#### 【0192】

図7は、本開示の一部の実施形態での使用に適したエピジェネティック対照核酸分子の略図である。本明細書に記載のエピジェネティック対照核酸分子は、核酸分子の区画の間の完全にメチル化された CpG 対の位置特異的影響を考慮に入れることができる。図7では、エピジェネティック対照核酸分子を5つのサブセットに群分けする。エピジェネティック対照核酸分子のエピジェネティック修飾領域の配列の長さおよびの配列組成は全てのサブセットにおいて同じである。各サブセットは、完全にメチル化された CpG 対を2つ有するが、2つの完全にメチル化された CpG 対の位置がサブセットによって変動する(すなわち、2つの完全にメチル化された CpG 対間の距離がサブセットによって変動する)。サブセット1では、2つの完全にメチル化された CpG 対は遠く離れているが、サブセット4では、2つの完全にメチル化された CpG 対は互いに極めて近い。ここで、識別子領域はエピジェネティック修飾領域の両側に存在する。両側の識別子領域はエピジェネティック状態バーコード (ESB) および分子バーコード (MB) を有する。分子バーコードを個々のエピジェネティック対照核酸分子の識別子として使用し、各エピジェネティック対照核酸分子は一意的な分子バーコードを有する、すなわち、分子1はMB1を有し、分子2はMB2を有し、分子3はMB3を有するなどである。これらのサブセットは、完全にメチル化された CpG 対の位置の影響に基づいて異なる結合親和性を有する。ここで、エピジェネティック状態バーコードを使用して、完全にメチル化された CpG 対の位置を識別することができる。サブセット1のエピジェネティック対照核酸分子は全て、2つの完全にメチル化された CpG 対を同じ位置に有し、したがって、サブセット1のエピジェネティック対照核酸分子は同じエピジェネティック状態バーコード - ESB1 を有する。同様に、全てのサブセット2、3、4および5のエピジェネティック対照核酸分子は、それぞれ ESB2、ESB3 および ESB4 であるエピジェネティック状態バーコードを有する。本実施例では、同じエピジェネティック状態バーコードはエピジェネティック修飾領域の両側に存在する。

20

30

40

#### 【0193】

別の態様では、本開示は、エピジェネティック対照核酸分子のセットであって、1つまたは複数のエピジェネティック対照核酸分子のサブセットを含み、各サブセットが複数のエピジェネティック対照核酸分子を含み、各エピジェネティック対照核酸分子がエピジェネティック修飾領域を含む、エピジェネティック対照核酸分子のセットと、対象由来のポリヌクレオチドの試料中の核酸分子のセットとを含む核酸の集団を提供する。

#### 【0194】

50

一部の実施形態では、エピジェネティック対照核酸分子は、( i ) ラムダファージ DNA またはヒトゲノムの領域に対応する配列、( i i ) 天然には存在しない配列、および / または ( i i i ) ( i ) と ( i i ) の組合せのいずれかを有し得る。一部の実施形態では、エピジェネティック対照核酸分子は、天然には存在しない配列を含み得る。

【 0 1 9 5 】

一部の実施形態では、ポリヌクレオチドの試料は、DNA の試料、RNA の試料、無細胞ポリヌクレオチドの試料、無細胞 DNA の試料または無細胞 RNA の試料である。一部の実施形態では、ポリヌクレオチドの試料は無細胞 DNA の試料である。

【 0 1 9 6 】

一部の実施形態では、無細胞 DNA は、少なくとも 1 ng、少なくとも 5 ng、少なくとも 10 ng、少なくとも 15 ng、少なくとも 20 ng、少なくとも 30 ng、少なくとも 50 ng、少なくとも 75 ng、少なくとも 100 ng、少なくとも 150 ng、少なくとも 200 ng、少なくとも 250 ng、少なくとも 300 ng、少なくとも 350 ng、少なくとも 400 ng、少なくとも 450 ng または少なくとも 500 ng である。

10

【 0 1 9 7 】

一部の実施形態では、エピジェネティック対照核酸分子の量は、少なくとも 1 フェムトモル、少なくとも 2 フェムトモル、少なくとも 5 フェムトモル、少なくとも 10 フェムトモル、少なくとも 15 フェムトモル、少なくとも 20 フェムトモル、少なくとも 50 フェムトモル、少なくとも 75 フェムトモル、少なくとも 100 フェムトモル、少なくとも 125 フェムトモル、少なくとも 150 フェムトモルまたは少なくとも 200 フェムトモル

20

III . 方法の一般的な特色

A . 試料

【 0 1 9 8 】

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、体組織、全血、血小板、血清、血漿、便、赤血球、白血球 ( white blood cell or leucocyte )、内皮細胞、組織生検材料 ( 例えば、既知のまたは推測される固形腫瘍からの生検材料 )、脳脊髄液、滑液、リンパ液、腹水、間質液または細胞外液 ( 例えば、細胞間隙からの滲出液 )、歯肉滲出液、歯肉溝滲出液、骨髓、胸水、脳脊髄液、唾液、粘液、喀痰、精液、汗、および尿を含み得る。試料は、血液およびその画分、ならびに尿などの体液であり得る。そのような試料は、腫瘍から放出される核酸を含み得る。核酸は、DNA および RNA を含み得、二重鎖形態および一本鎖形態であり得る。試料は、対象から最初に単離された形態であり得る、または細胞などの構成成分を除去もしくは添加する、1つの構成成分を別の構成成分と相対して富化させるため、もしくは1つの形態の核酸を別の形態の核酸に変換する、例えば、RNA を DNA に変換するもしくは一本鎖核酸を二本鎖に変換するためにさらなる処理に供されたものであり得る。したがって、例えば、分析用の体液は、無細胞核酸、例えば、無細胞 DNA ( cf DNA ) を含有する血漿または血清であり得る。

30

【 0 1 9 9 】

一部の実施形態では、対象から取得した体液の試料体積は、シーケンシングされる領域に対する所望のリードの深さに依存する。体積の例は、約 0 . 4 ~ 40 ミリリットル ( mL )、約 5 ~ 20 mL、約 10 ~ 20 mL である。例えば、体積は、約 0 . 5 mL、約 1 mL、約 5 mL、約 10 mL、約 20 mL、約 30 mL、約 40 mL、またはそれよりも多くのミリリットルであり得る。試料採取された血漿の体積は、一般には、約 5 mL から約 20 mL の間である。

40

【 0 2 0 0 】

試料は、種々の量の核酸を含み得る。一般には、所与の試料中の核酸の量を多数のゲノム等価物と同等とみなす。例えば、DNA 約 30 ナノグラム ( ng ) の試料は一倍体ヒトゲノム等価物を約 10 , 000 (  $10^4$  ) 個含有し得、cf DNA の場合では、個々のポリヌクレオチド分子を約 2000 億 (  $2 \times 10^{11}$  ) 個含有し得る。同様に、DNA 約 1

50

00 ng の試料は、一倍体ヒトゲノム等価物を約 30,000 個含有し得、cfDNA の場合は、個々の分子を約 6000 億個含み得る。

#### 【0201】

一部の実施形態では、試料は、異なる供給源由来の核酸、例えば、細胞由来の核酸および無細胞供給源（例えば、血液試料など）由来の核酸を含む。一般には、試料は、突然変異を保有する核酸を含む。例えば、試料は、必要に応じて、生殖細胞系列突然変異および/または体細胞突然変異を保有するDNAを含む。一般には、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）を保有するDNAを含む。

#### 【0202】

増幅前の試料中の無細胞核酸の量の例は、一般には、約1フェムトグラム（fg）から約1マイクログラム（μg）まで、例えば、約1ピコグラム（pg）～約200ナノグラム（ng）、約1ng～約100ng、約10ng～約1000ngにわたる。一部の実施形態では、試料は、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の無細胞核酸分子を含む。必要に応じて、量は、少なくとも約1fg、少なくとも約10fg、少なくとも約100fg、少なくとも約1pg、少なくとも約10pg、少なくとも約100pg、少なくとも約1ng、少なくとも約10ng、少なくとも約100ng、少なくとも約150ng、または少なくとも約200ngの無細胞核酸分子である。一部の実施形態では、量は、最大約1fg、約10fg、約100fg、約1pg、約10pg、約100pg、約1ng、約10ng、約100ng、約150ng、または約200ngの無細胞核酸分子である。一部の実施形態では、方法は、試料から約1fgから約200ngの間の無細胞核酸分子を得ることを含む。

#### 【0203】

無細胞核酸は、一般には約100ヌクレオチド長から約500ヌクレオチド長の間のサイズ分布を有し、約110ヌクレオチド長～約230ヌクレオチド長の分子が試料中の分子の約90%を占め、最頻値は約168ヌクレオチド長（ヒト対象由来の試料中）であり、第2の副次的なピークが約240ヌクレオチド長から約440ヌクレオチド長の間にわたる。一部の実施形態では、無細胞核酸は約160ヌクレオチド長から約180ヌクレオチド長まで、または約320ヌクレオチド長から約360ヌクレオチド長まで、または約440ヌクレオチド長から約480ヌクレオチド長までである。

#### 【0204】

一部の実施形態では、溶液中に見出される無細胞核酸を体液中のインタクトな細胞および他の不溶性構成成分から分離する区画ステップを通じて無細胞核酸を体液から単離する。一部の実施形態では、区画は、遠心分離または濾過などの技法を含む。あるいは、体液中の細胞を溶解させることができ、無細胞核酸および細胞内核酸を一緒に処理することができる。一般に、緩衝剤の添加および洗浄ステップ後、無細胞核酸を、例えばアルコールを用いて沈殿させることができる。一部の実施形態では、シリカに基づくカラムなどの追加的な清浄化ステップを使用して、夾雑物または塩を除去する。非特異的バルクキャリア核酸を、例えば、必要に応じて、例示的な手順の側面、例えば収量などを最適化するために反応全体を通して添加する。そのような処理後、試料は、一般には、二本鎖DNA、一本鎖DNAおよび/または一本鎖RNAを含めた様々な形態の核酸を含む。必要に応じて、一本鎖DNAおよび/または一本鎖RNAを二本鎖形態に変換し、したがって、その後の処理および解析ステップに含める。

#### B. タグ付け

#### 【0205】

一部の実施形態では、核酸分子（ポリヌクレオチドの試料由来）に、試料インデックスおよび/または分子バーコード（一般に「タグ」と称される）を用いてタグ付けすることができる。タグは、他の方法の中でも、化学合成、ライゲーション（例えば、平滑末端ライゲーションもしくは粘着末端ライゲーション）、またはオーバーラップ伸長ポリメラーゼ連鎖反応（PCR）によってアダプターに組み入れることまたは他のやり方で接合する

ことができる。そのようなアダプターを最終的に標的核酸分子に接合することができる。他の実施形態では、従来の核酸増幅方法を使用して試料インデックスを核酸分子に導入するために、1つまたは複数の増幅サイクル（例えば、PCR増幅）のラウンドを一般に適用する。増幅は、1つまたは複数の反応混合物中（例えば、アレイ中の複数のマイクロウェル）で行うことができる。分子バーコードおよび/または試料インデックスは、同時に、または任意の逐次的順序で導入することができる。一部の実施形態では、分子バーコードおよび/または試料インデックスを、配列捕捉ステップを実施する前および/または実施した後に導入する。一部の実施形態では、分子バーコードのみをプローブによる捕捉前に導入し、試料インデックスを配列捕捉ステップの実施後に導入する。一部の実施形態では、分子バーコードおよび試料インデックスの両方をプローブに基づく捕捉ステップの実施前に導入する。一部の実施形態では、試料インデックスを配列捕捉ステップの実施後に導入する。一部の実施形態では、分子バーコードを試料中の核酸分子（例えば、cfDNA分子）にアダプターを通じて、ライゲーション（例えば、平滑末端ライゲーションまたは粘着末端ライゲーション）によって組み入れる。一部の実施形態では、試料インデックスを試料中の核酸分子（例えば、cfDNA分子）にオーバーラップ伸長ポリメラーゼ連鎖反応（PCR）を通じて組み入れる。一般には、配列捕捉プロトコールは、標的とされる核酸配列、例えばゲノム領域のコード配列に相補的な一本鎖核酸分子を導入することを伴い、そのような領域の突然変異ががんの型に関連付けられる。

#### 【0206】

一部の実施形態では、タグを試料核酸分子の一方の末端または両方の末端に位置させることができる。一部の実施形態では、タグは、所定のまたはランダムもしくはセミランダムな配列のオリゴヌクレオチドである。一部の実施形態では、タグは、約500ヌクレオチド未満、約200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、約20ヌクレオチド未満、約10ヌクレオチド未満、約9ヌクレオチド未満、約8ヌクレオチド未満、約7ヌクレオチド未満、約6ヌクレオチド未満、約5ヌクレオチド未満、約4ヌクレオチド未満、約3ヌクレオチド未満、約2ヌクレオチド未満、または約1ヌクレオチド未満の長さであり得る。タグは、試料核酸にランダムに連結することもでき、非ランダムに連結することもできる。

#### 【0207】

一部の実施形態では、各試料に、試料インデックスまたは試料インデックスの組合せを用いて一意的にタグ付けする。一部の実施形態では、各試料または副試料の核酸分子に、分子バーコードまたは分子バーコードの組合せを用いて一意的にタグ付けする。他の実施形態では、複数の分子バーコードを、分子バーコードが必ずしも複数内で互いに一意的にならないように使用することができる（例えば、非一意的分子バーコード）。これらの実施形態では、一般に、分子バーコードを個々の分子に、分子バーコードとそれを付着させることができる配列の組合せにより、個別に追跡することができる一意的な配列が創出されるように付着させる（例えば、ライゲーションによって）。非一意的にタグ付けされた分子バーコードと内在性配列情報（例えば、試料中の元の核酸分子の配列に対応する最初の（開始）および/または最後の（終止）ゲノム場所/位置、一方の末端もしくは両方の末端における配列リードの部分配列、配列リードの長さ、および/または試料中の元の核酸分子の長さ）の組合せの検出により、一般には、一意的な同一性を特定の分子に割り当てることが可能になる。一部の実施形態では、非一意的にタグ付けされた分子バーコードと内在性配列情報（例えば、配列リードと参照配列のアラインメントの最初の（開始）および/または最後の（終止）領域、一方の末端もしくは両方の末端における配列リードの部分配列、配列リードの長さ、および/または試料中の元の核酸分子の長さ）の組合せの検出により、一般には、一意的な同一性を特定の分子に割り当てることができる。一部の実施形態では、最初の領域は、参照配列に対するアラインメントが開始されるシーケンシングリードの5'末端が決定されるシーケンシングリードのゲノム開始位置を含み、最後の領域は、参照配列に対するアラインメントが終止するシーケンシングリードの3'末端が決定されるシーケンシングリードのゲノム終止位置を含む。一部の実施形態では、最初の領

10

20

30

40

50

域は、参照配列とアラインメントされるシーケンシングリードの5'末端の最初の1塩基位置、最初の2塩基位置、最初の5塩基位置、最初の10塩基位置、最初の15塩基位置、最初の20塩基位置、最初の25塩基位置、最初の30塩基位置または少なくとも最初の30塩基位置を含む。一部の実施形態では、最後の領域は、参照配列とアラインメントされるシーケンシングリードの3'末端の最後の1塩基位置、最後の2塩基位置、最後の5塩基位置、最後の10塩基位置、最後の15塩基位置、最後の20塩基位置、最後の25塩基位置、最後の30塩基位置または少なくとも最後の30塩基位置を含む。

#### 【0208】

必要に応じて、個々の配列リードの塩基対の長さまたは数を使用して一意的な同一性を所与の分子に割り当てる。本明細書に記載の通り、一意的な同一性が割り当てられた核酸の一本鎖由来の断片により、その後の親鎖および/または相補鎖由来の断片の識別が可能になり得る。

10

#### 【0209】

一部の実施形態では、分子バーコードを、識別子のセット（例えば、一意的または非一意的分子バーコードの組合せ）と試料中の分子の予測比率で導入する。フォーマットの1つの例では、約2種から約1,000,000種までの異なる分子バーコード配列、または約5種から約150種までの異なる分子バーコード配列、または約20種から約50種までの異なる分子バーコード配列を標的分子の両末端にライゲーションして使用する。あるいは、約25種から約1,000,000種の異なる分子バーコード配列を使用することができる。例えば、20~50×20~50種の分子バーコード配列（すなわち、20~50種の異なる分子バーコード配列のうちの1種を標的分子の各末端に付着させることができる）を使用することができる。そのような識別子の数は、一般には、同じ開始点および終止点を有する異なる分子が識別子の異なる組合せを有する確率を高くする（例えば、少なくとも94%、99.5%、99.99%、または99.999%）十分なものである。一部の実施形態では、分子の約80%、約90%、約95%、または約99%が分子バーコードの同じ組合せを有する。

20

#### 【0210】

一部の実施形態では、反応における一意的または非一意的分子バーコードの割り当てを、例えば、そのそれぞれの全体が参照により本明細書に組み込まれる米国特許出願第20010053519号、同第20030152490号、および同第20110160078号、および米国特許第6,582,908号、同第7,537,898号、同第9,598,731号、および同第9,902,992号に記載されている方法およびシステムを使用して実施する。あるいは、一部の実施形態では、内在性配列情報（例えば、開始および/または終止位置、配列の一方または両方の末端の部分配列、および/または長さ）だけを使用して試料の異なる核酸分子を識別することができる。

30

#### 【0211】

エピジェネティック状態バーコード(ESB)は、エピジェネティック対照核酸分子のエピジェネティック修飾領域に付着させるタグの一種である。ESBをエピジェネティック対照核酸分子のエピジェネティック状態の識別子として使用することができる。ESBは、エピジェネティック対照核酸分子のエピジェネティック修飾領域内のエピジェネティック修飾を有するヌクレオチドの数を指し得る。一部の実施形態では、エピジェネティック対照核酸分子の識別子領域は、少なくとも1つのエピジェネティック状態バーコードを含む。一部の実施形態では、ESBは、エピジェネティック対照核酸分子の識別子領域の一部である。例えば、エピジェネティック修飾がDNAメチル化であり、エピジェネティック対照核酸分子のサブセットがメチル化ヌクレオチドを5つ有する場合には、そのサブセットに入るエピジェネティック対照核酸分子の全てが同じエピジェネティック状態バーコードを有することになる。一部の実施形態では、エピジェネティック状態バーコードを使用して、エピジェネティック対照核酸分子のエピジェネティック修飾領域のエピジェネティック修飾のレベル/度合いを識別することができる。エピジェネティック対照核酸分子をエピジェネティック修飾領域内のシトシンまたはCpGヌクレオチドの数に基づいて

40

50

サブセットに群分けすることができる。一部の実施形態では、各サブセット内でメチル化のレベルは変動し得（例えば、高度にメチル化された、中程度にメチル化されたおよび低度でメチル化された）、メチル化の各レベルは別々のエピジェネティック状態バーコードを有し得る。例えば、サブセットA内で、低度でメチル化されたエピジェネティック対照核酸分子は全てがエピジェネティック状態バーコード、例えばESB1を有し、高度にメチル化されたエピジェネティック対照核酸分子は全てが別のエピジェネティック状態バーコード、例えばESB3を有することになる。本実施例では、エピジェネティック状態バーコードを使用して、メチル化のレベル/度合いを識別する。

#### 【0212】

一部の実施形態では、反応における一意的なまたは非一意的分子バーコードの割り当てを、例えば、そのそれぞれの全体が参照により本明細書に組み込まれる米国特許出願第20010053519号、同第20030152490号、および同第20110160078号、ならびに米国特許第6,582,908号、同第7,537,898号、同第9,598,731号、および同第9,902,992号に記載されている方法およびシステムを使用して実施する。

#### C. 増幅

#### 【0213】

試料核酸を、アダプターで挟み、増幅されるDNA分子を挟むアダプター内のプライマー結合部位に結合する核酸プライマーを使用してPCRおよび他の増幅方法によって増幅させることができる。一部の実施形態では、増幅方法は、伸長、変性、およびアニーリングのサイクルを伴い、これは、サーモサイクリングによるもの、または、例えば転写増幅法におけるものと同様に等温性であり得る。必要に応じて利用することができる増幅方法の他の例としては、リガーゼ連鎖反応、鎖置換増幅、核酸配列に基づく増幅、および自家持続配列に基づく複製が挙げられる。

#### 【0214】

一般には、増幅反応により、約150ヌクレオチド(nt)から約700ntまで、250ntから約350ntまで、または約320ntから約550ntまでにわたるサイズの、分子バーコードおよび試料インデックスを有する非一意的にまたは一意的にタグ付けされた核酸アンプリコンが複数生成される。一部の実施形態では、アンプリコンのサイズは約180ntである。一部の実施形態では、アンプリコンのサイズは約200ntである。

#### D. 富化

#### 【0215】

一部の実施形態では、核酸をシーケンシングする前に、配列を富化させる。富化は、必要に応じて特異的な標的領域に対して実施するかまたは非特異的に実施する（「標的配列」）。一部の実施形態では、標的とされる目的の領域を、1つまたは複数のベイトセットパネルについて選択された核酸捕捉用プローブ（「ベイト」）を用い、示差的なタイリングおよび捕捉スキームを使用して富化させ得る。示差的なタイリングおよび捕捉スキームでは、一般に、ベイトと会合するゲノム領域にわたって示差的にタイリングするために（例えば、異なる「分解能」で）種々の相対的濃度のベイトセットを使用し、一連の制約（例えば、シーケンシング負荷、各ベイトの有用性などシーケンサー制約）に供し、標的とされる核酸を下流のシーケンシングのために所望のレベルで捕捉する。これらの目的の標的とされるゲノム領域は、必要に応じて核酸構築物の天然または合成ヌクレオチド配列を含む。一部の実施形態では、1つまたは複数の目的の領域に対するプローブを伴うピオチン標識ビーズを使用して、標的配列を捕捉することができ、必要に応じて、これらの領域を増幅させて、目的の領域を富化させる。

#### 【0216】

配列捕捉は、一般には、標的核酸配列とハイブリダイズするオリゴヌクレオチドプローブの使用を伴う。一部の実施形態では、プローブセット戦略は、目的の領域にわたってプローブをタイリングすることを伴う。そのようなプローブは、例えば、約60ヌクレオチ

10

20

30

40

50

ドから約120ヌクレオチドまでの長さであり得る。セットの深さ（例えば、カバレッジの深さ）は、約2x、3x、4x、5x、6x、7x、8x、9x、10x、15x、20x、50xの深さ、または50xよりも深いものであり得る。配列捕捉の効果は、一般に、一部において、プローブの配列と相補的な（またはほぼ相補的な）標的分子内の配列の長さに依存する。

#### E. シーケンシング

##### 【0217】

必要に応じてアダプターに挟まれ、予め増幅させたまたはさせていない試料核酸を、一般に、シーケンシングに供する。必要に応じて利用されるシーケンシング方法または市販のフォーマットとしては、例えば、サンガーシーケンシング、ハイスループットシーケンシング、パイロシーケンシング、合成によるシーケンシング、単一分子シーケンシング、ナノポアに基づくシーケンシング、半導体シーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシング、RNA-Seq (Illumina)、Digital Gene Expression (Helicos)、次世代シーケンシング (NGS)、単一分子合成によるシーケンシング (SMS) (Helicos)、大規模並列シーケンシング、Clonal Single Molecule Array (Solexa)、ショットガンシーケンシング、Ion Torrent、Oxford Nanopore、Roche Genia、Maxim-Gilbertシーケンシング、プライマーウォーキング、PacBio、SOLiD、Ion Torrent、またはNanoporeプラットフォームを使用したシーケンシングが挙げられる。シーケンシング反応は、多数の試料セットを実質的に同時に処理する多数のレーン、多数のチャンネル、多数のウェル、または他の手段を含み得る種々の試料処理装置で実施することができる。試料処理装置は、多数の動作を同時に処理することを可能にするために多数の試料チャンパーも含み得る。

##### 【0218】

シーケンシング反応は、がんまたは他の疾患のマーカーを含有することが分かっている1つまたは複数の核酸断片型または領域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、試料中に存在するあらゆる核酸断片に対して実施することもできる。シーケンシング反応は、ゲノムの少なくとも約5%、10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90%、95%、99%、99.9%、または100%に対して実施することができる。他の場合では、シーケンシング反応をゲノムの約5%未満、約10%未満、約15%未満、約20%未満、約25%未満、約30%未満、約40%未満、約50%未満、約60%未満、約70%未満、約80%未満、約90%未満、約95%未満、約99%未満、約99.9%未満、または約100%未満に対して実施することができる。

##### 【0219】

同時シーケンシング反応を、マルチプレックスシーケンシング技法を使用して実施することができる。一部の実施形態では、無細胞ポリヌクレオチドのシーケンシングを、少なくとも約1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000、または100,000のシーケンシング反応で行う。他の実施形態では、無細胞ポリヌクレオチドのシーケンシングを、約1000未満、約2000未満、約3000未満、約4000未満、約5000未満、約6000未満、約7000未満、約8000未満、約9000未満、約10000未満、約50000未満、または約100,000未満のシーケンシング反応で行う。シーケンシング反応は、一般には、逐次的にまたは同時に実施される。その後のデータ解析を、一般に、シーケンシング反応の全部または一部に対して実施する。一部の実施形態では、データ解析を少なくとも約1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、50000、または100,000のシーケンシング反応に対して実施する。他の実施形態では、データ解析を約1000未満、約2000未満、約3000未満、約4000未満、約5000未満、約6000未満、約7000

10

20

30

40

50

0 未満、約 8 0 0 0 未満、約 9 0 0 0 未満、約 1 0 0 0 0 未満、約 5 0 0 0 0 未満、または約 1 0 0 , 0 0 0 未満のシーケンシング反応に対して実施することができる。リードの深さの例は、遺伝子座（例えば、塩基位置）当たり約 1 0 0 0 リードから約 5 0 0 0 0 リードまでである。

F . 解析

【 0 2 2 0 】

シーケンシングにより、複数のシーケンシングリードまたはリードが生成される。シーケンシングリードまたはリードは、約 1 5 0 塩基未満の長さ、または約 9 0 塩基未満の長さのヌクレオチドデータの配列を含み得る。一部の実施形態では、リードは、約 8 0 塩基から約 9 0 塩基の間、例えば、約 8 5 塩基の長さである。一部の実施形態では、本開示の方法を非常に短いリード、例えば、約 5 0 塩基未満または約 3 0 塩基未満の長さに適用する。シーケンシングリードデータは、配列データならびにメタ情報を含み得る。配列リードデータは、例えば、V C F ファイル、F A S T A ファイル、または F A S T Q ファイルを含めた任意の適切なファイルフォーマットに保存することができる。

10

【 0 2 2 1 】

F A S T A は、配列データベースを検索するためのコンピュータプログラムを指し得、F A S T A という名称は、標準ファイルフォーマットも指し得る。例えば、F A S T A は、例えば、その全体が参照により本明細書に組み込まれる Pearson & Lipman, 1988, Improved tools for biological sequence comparison, PNAS 85:2444-2448 に記載されている。F A S T A フォーマットの配列は、一行の記述で始まり、その後に配列データの行が続く。記述行は最初の列の大なり（「>」）記号によって配列データと区別される。「>」記号に続く単語は配列の識別子であり、その行の残りは説明である（どちらも省略できる）。「>」と識別子の最初の文字の間にスペースを入れるべきではない。文字列の全ての行が 8 0 文字未満であることが推奨される。「>」から始まる別の行が出現すると配列が終わり、これは、別の配列が始まることを示す。

20

【 0 2 2 2 】

F A S T Q フォーマットは、生物学的配列（通常ヌクレオチド配列）およびその対応するクオリティスコアの両方を保存するためのテキストに基づくフォーマットである。F A S T Q フォーマットは F A S T A フォーマットと同様であるが、配列データの次にクオリティスコアが伴う。簡潔にするために配列文字およびクオリティスコアの両方が単一の A S C I I 文字でコードされる。F A S T Q フォーマットは、例えば、その全体が参照により本明細書に組み込まれる Cock et al. ("The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants," Nucleic Acid Res 38 (6): 1767-1771, 2009) に記載されている通り、I l l u m i n a G e n o m e A n a l y z e r などのハイスループットシーケンシング機器の出力を保存するための事実上の標準である。

30

【 0 2 2 3 】

F A S T A および F A S T Q ファイルに関しては、メタ情報は、記述行を含み、配列データの行は含まない。一部の実施形態では、F A S T Q ファイルに関しては、メタ情報は、クオリティスコアを含む。F A S T A および F A S T Q ファイルに関しては、配列データは、記述行の後に始まり、一般には、I U P A C 多義性コードのいくつかのサブセットを使用し、必要に応じて「-」を用いて示される。ある実施形態では、配列データには、A、T、C、G、および N 文字を使用することができ、必要に応じて「-」または必要であれば U を含める（例えば、ギャップまたはウラシルを表すために）。

40

【 0 2 2 4 】

一部の実施形態では、少なくとも 1 つのマスター配列リードファイルおよび出力ファイルをプレーンテキストファイルとして保存する（例えば、A S C I I ; I S O / I E C 6 4 6 ; E B C D I C ; U T F - 8 ; または U T F - 1 6 などのエンコーディングを使用する）。本開示に提示されるコンピュータシステムは、プレーンテキストファイルを開くことができるテキストエディタプログラムを含み得る。テキストエディタプログラムは、

50

テキストファイル（例えば、プレーンテキストファイルなど）の内容をコンピュータスクリーン上に提示し、人がテキストを編集することを可能にする（例えば、モニター、キーボード、およびマウスを使用して）ことができるコンピュータプログラムを指し得る。テキストエディタの例としては、限定することなく、Microsoft Word、emacs、pico、vi、BBEdit、およびTextWranglerが挙げられる。テキストエディタプログラムは、プレーンテキストファイルをコンピュータスクリーン上に表示し、メタ情報および配列リードを人間が読むことが可能なフォーマット（例えば、バイナリコードされていないのではなく、印刷または人が書くことに使用することができる英数字が使用される）を示すことができるものである。

#### 【0225】

FASTAまたはFASTQファイルに関連して方法が考察されているが、本開示の方法およびシステムを、例えば、変異体コールフォーマット（Variant Call Format）（VCF）フォーマットのファイルを含めた任意の適切な配列ファイルフォーマットを圧縮するために使用することができる。典型的なVCFファイルは、ヘッダーセクションおよびデータセクションを含み得る。ヘッダーは、任意の数のメタ情報行を含有し、それぞれが「##」文字で始まり、TABにより単一の「#」文字で始まるフィールド定義行が区切られる。フィールド定義行により8つの必須列が指名され、本文セクションは、フィールド定義行によって定義される列を集合させたデータの行を含有する。VCFフォーマットは、例えば、その全体が参照により本明細書に組み込まれる、Danecek et al. ("The variant call format and VCF tools," *Bioinformatics* 27(15):2156-2158, 2011) に記載されている。ヘッダーセクションは圧縮されたファイルに書き込むためのメタ情報として扱うことができ、データセクションは行として扱うことができ、そのそれぞれが、一意的な場合に限りマスターファイルに保存される。

#### 【0226】

一部の実施形態は、シーケンシングリードのアセンブリを提供する。アラインメントによるアセンブリでは、例えば、シーケンシングリードを互いにアラインメントするかまたは参照配列に対してアラインメントする。各リードを参照ゲノムに対して順番にアラインメントすることにより、リードの全てが互いとの関係で位置付けられて、アセンブリが創出される。さらに、シーケンシングリードを参照配列に対してアラインメントまたはマッピングすることを、シーケンシングリード内の変異体配列を識別するために使用することもできる。疾患もしくは状態の診断もしくは予後判定をさらに補助するため、または処置決定を手引きするために、変異体配列の識別を本明細書に記載の方法およびシステムと組み合わせて使用することができる。

#### 【0227】

一部の実施形態では、ステップのいずれかまたは全てを自動化する。あるいは、本開示の方法は、例えば、それぞれが必要に応じてC++などのコンパイル型言語で書かれた1つまたは複数の専用プログラムに完全にまたは部分的に具体化し、次いで、コンパイルし、バイナリとして配布することができる。本開示の方法は、完全にまたは部分的に、既存の配列解析プラットフォーム内のモジュールとして、または既存の配列解析プラットフォーム内のファンクションを呼び出すことによってインプリメントすることができる。一部の実施形態では、本開示の方法は、多数のステップを含み、それらは全て、単一の開始キュー（例えば、人間による作業、別のコンピュータプログラム、または機械から供給される誘発事象の1つまたは組合せ）に回答して自動的に呼び出される。したがって、本開示は、任意のまたはステップまたはステップの任意の組合せをキューに回答して自動的に行うことができる方法を提供する。「自動的に」とは、一般に、人間による入力、影響、または相互作用を伴わないことを意味する（例えば、元のまたはキュー前の人間による作業にのみ応答する）。

#### 【0228】

本開示の方法は、対象の核酸試料に関する正確かつ高感度の解釈を含む、様々な形態の出力も包含し得る。検索の出力は、コンピュータファイルのフォーマットで提供すること

10

20

30

40

50

ができる。一部の実施形態では、出力は、FASTAファイル、FASTQファイル、またはVCFファイルである。出力を処理して、参照ゲノムの配列に対してアラインメントされた核酸の配列などの配列データを含むテキストファイル、またはXMLファイルを作成することができる。他の実施形態では、処理により、対象核酸における、参照ゲノムと比べた1つまたは複数の突然変異を記述する座標またはストリングを含む出力がもたらされる。アラインメントストリングは、Simple Ungapped Alignment Report (SUGAR)、Verbose Useful Labeled Gapped Alignment Report (VULGAR)、およびCompact Idiosyncratic Gapped Alignment Report (CIGAR) (例えば、その全体が参照により本明細書に組み込まれるNing et al., Genome Research 11 (10): 1725-9, 2001に記載されている)を含み得る。これらのストリングは、例えば、European Bioinformatics Institute (Hinxton, UK)のExonerate配列アラインメントソフトウェアでインプリメントすることができる。

#### 【0229】

一部の実施形態では、例えば、CIGARストリングを含む、配列アラインメントマップ(SAM)またはバイナリアラインメントマップ(BAM)ファイルなどの配列アラインメントを作製する(SAMフォーマットは、例えば、その全体が参照により本明細書に組み込まれるLi et al., "The Sequence Alignment/Map format and SAMtools," Bioinformatics, 25 (16): 2078-9, 2009に記載されている)。一部の実施形態では、CIGARは、行当たり1つのギャップ挿入されたアラインメントを表示するまたは含む。CIGARは、CIGARストリングとしてレポートされる圧縮されたペアワイズアラインメントフォーマットである。CIGARストリングは、長い(例えば、ゲノム)ペアワイズアラインメントを表すために有用であり得る。CIGARストリングをSAMフォーマットで使用して、リードの参照ゲノム配列に対するアラインメントを表すことができる。

#### 【0230】

CIGARストリングは、確立されたモチーフに従い得る。各文字には数字が先行し、それにより、事象の塩基計数が示される。使用される文字は、M、I、D、N、およびSを含み得る(M = マッチ; I = 挿入; D = 欠失; N = ギャップ; S = 置換)。CIGARストリングにより、マッチ/ミスマッチおよび欠失(またはギャップ)の配列が定義される。例えば、CIGARストリング2MD3M2D2Mは、アラインメントが2つのマッチ、1つの欠失(いくらかのスペースを節約するために1という数字は省略される)、3つのマッチ、2つの欠失、および2つのマッチを含むことを示し得る。

#### 【0231】

一部の実施形態では、核酸集団を、シーケンシングのために、一方の末端または両方の末端に一本鎖突出を有する二本鎖核酸上に平滑末端を酵素により形成することによって調製する。これらの実施形態では、集団を、一般には、5' - 3' DNAポリメラーゼ活性および3' - 5' エキソヌクレアーゼ活性を有する酵素を用い、ヌクレオチド(例えば、A、C、G、およびTまたはU)の存在下で処理する。必要に応じて使用することができる酵素またはその触媒断片の例としては、クレノウ大断片およびT4ポリメラーゼが挙げられる。5' 突出では、酵素により、一般には、反対の鎖上の陥凹3'末端が5'末端と揃うまで伸長して、平滑末端が生じる。3' 突出では、酵素により、一般に、3'末端から反対の鎖の5'末端まで、および時にはそれを越えて消化される。この消化が反対の鎖の5'末端を越えて進行する場合、5'突出に対して使用される、同じポリメラーゼ活性を有する酵素によってギャップを埋めることができる。二本鎖核酸における平滑末端の形成により、例えば、アダプターの付着およびその後の増幅が容易になる。

#### 【0232】

一部の実施形態では、核酸集団を、一本鎖核酸から二本鎖核酸への変換および/またはRNAからDNA(例えば、相補DNAまたはcDNA)への変換などの追加的な処理に供する。これらの形態の核酸も必要に応じてアダプターに連結し、増幅させる。

10

20

30

40

50

## 【0233】

核酸を、予め増幅させてまたは増幅させずに上記の平滑末端を形成するプロセスに供し、必要に応じて、試料中の他の核酸をシーケンシングして、シーケンシングされた核酸を生じさせることができる。シーケンシングされた核酸とは、核酸の配列（例えば、配列情報）または配列が決定された核酸のいずれも指し得る。シーケンシングは、試料中の個々の核酸分子の配列データが直接、または試料中の個々の核酸分子の増幅産物のコンセンサス配列から間接的にもたらされるように実施することができる。

## 【0234】

一部の実施形態では、試料中の一本鎖突出を有する二本鎖核酸には、平滑末端形成後、両末端にバーコードを含むアダプターを連結し、シーケンシングにより、核酸配列ならびにアダプターによって導入されたインラインバーコードを決定する。平滑末端DNA分子を必要に応じて少なくとも部分的に二本鎖のアダプター（例えば、Y形または鐘形アダプター）の平滑末端とライゲーションする。あるいは、ライゲーションを容易にするために（例えば、粘着末端ライゲーションに関して）、試料核酸およびアダプターの平滑末端に、相補的なヌクレオチドを用いて尾部付加することができる。

## 【0235】

一般には、同じ核酸の任意の2つのコピーが、両末端に連結したアダプター由来のアダプターバーコードの同じ組合せを受け取る確率が低くなる（例えば、約1%未満または約0.1%未満）ように、核酸試料を十分な数のアダプターと接触させる。アダプターをこのように使用することにより、参照核酸における開始点および終止点が同じであり、バーコードの同じ組合せと連結した核酸配列のファミリーを識別することが可能になり得る。そのようなファミリーは、増幅前の試料中の核酸の増幅産物の配列を表し得る。平滑末端形成およびアダプター付着によって修飾されるにしたがって、元の試料中の核酸分子について、ファミリーメンバーの配列をコンパイルして、コンセンサスヌクレオチド（単数または複数）または完全なコンセンサス配列を導き出すことができる。言い換えれば、試料中の核酸の特定の位置を占有するヌクレオチドを、ファミリーメンバー配列のその対応する位置を占有するヌクレオチドのコンセンサスであると決定することができる。ファミリーは、二本鎖核酸の一方の鎖または両方の鎖の配列を含み得る。ファミリーのメンバーが二本鎖核酸の両方の鎖の配列を含む場合、配列をコンパイルして、コンセンサスヌクレオチド（単数または複数）または配列を導き出す目的で、一方の鎖の配列をそれらの相補物に変換することができる。一部のファミリーは、単一のメンバー配列のみを含む。この場合、この配列を、増幅前の試料中の核酸の配列とみなすことができる。あるいは、単一のメンバー配列のみを伴うファミリーをその後の解析から除外することができる。

## 【0236】

シーケンシングされた核酸におけるヌクレオチド変異（例えば、SNVまたはインデル）は、シーケンシングされた核酸を参照配列と比較することによって決定することができる。参照配列は、多くの場合、既知配列、例えば、対象由来の既知の全ゲノム配列または部分的なゲノム配列（例えば、ヒト対象由来の全ゲノム配列）である。参照配列は、例えば、hg19またはhg38であり得る。シーケンシングされた核酸は、上記の通り、試料中の核酸について直接決定された配列、またはそのような核酸の増幅産物の配列のコンセンサスを表す。比較は、参照配列の1つまたは複数の指定位置で実施することができる。それぞれの配列を最大限にアラインメントした場合に、参照配列の指定位置に対応する位置を含むシーケンシングされた核酸のサブセットを識別することができる。そのようなサブセット内では、もしあれば、どのシーケンシングされた核酸が指定位置にヌクレオチド変異を含むか、および必要に応じて、もしあれば、どれが参照ヌクレオチド（例えば、参照配列におけるものと同じ）を含むかを決定することができる。サブセット内のヌクレオチド変異体を含むシーケンシングされた核酸の数が選択された閾値を超える場合には、その指定位置において変異体ヌクレオチドをコールすることができる。閾値は、他の可能性の中でも、サブセット内のヌクレオチド変異体を含むシーケンシングされた核酸の単純な数、例えば、少なくとも1個、2個、3個、4個、5個、6個、7個、8個、9個、も

10

20

30

40

50

しくは10個などであり得る、または、サブセット内のヌクレオチド変異体を含むシーケンシングされた核酸の比、例えば、少なくとも0.5、1、2、3、4、5、10、15、または20などであり得る。参照配列内の目的の任意の指定位置について比較を繰り返すことができる。時には、参照配列の少なくとも約20、100、200、または300の連続した位置、例えば、約20~500、または約50~300の連続した位置を占有する指定位置について比較を実施することができる。

#### 【0237】

本明細書に記載のフォーマットおよび適用を含めた核酸シーケンシングに関する追加的な詳細は、例えば、Levy et al., *Annual Review of Genomics and Human Genetics*, 17:95-115 (2016)、Liu et al., *J. of Biomedicine and Biotechnology*, Volume 2 012, Article ID 251364: 1-11 (2012)、Voelkerding et al., *Clinical Chem.*, 55: 641-658(2009)、MacLean et al., *Nature Rev. Microbiol.*, 7: 287-296 (2009)、Astier et al., *J Am Chem Soc.*, 128 (5): 1705-10(2006)、米国特許第6,210,891号、米国特許第6,258,568号、米国特許第6,833,246号、米国特許第7,115,400号、米国特許第6,969,488号、米国特許第5,912,148号、米国特許第6,130,073号、米国特許第7,169,560号、米国特許第7,282,337号、米国特許第7,482,120号、米国特許第7,501,245号、米国特許第6,818,395号、米国特許第6,911,345号、米国特許第7,501,245号、米国特許第7,329,492号、米国特許第7,170,050号、米国特許第7,302,146号、米国特許第7,313,308号、および米国特許第7,476,503号にも記載されており、これらはそれぞれの全体が参照により本明細書に組み込まれる。

#### IV. コンピュータシステム

#### 【0238】

本開示の方法は、コンピュータシステムを使用して、またはその助けを借りてインプリメントすることができる。例えば、そのような方法は、(a) エピジェネティック対照核酸分子のセットをポリヌクレオチドの試料中の核酸分子に添加し、それによって、スパイクイン試料を作製するステップと、(b) スパイクイン試料の核酸分子を複数の区画化セットに区画するステップと、(c) 複数の区画化セットから分子のサブセットを富化させて、複数の富化された分子を生成するステップであって、複数の富化された分子が、エピジェネティック対照核酸分子の群およびポリヌクレオチドの試料由来の核酸分子の群を含む、ステップと、(d) 複数の富化された分子をシーケンシングして、複数のシーケンシングリードを作製するステップと、(e) 複数のエピジェネティック対照核酸分子のエピジェネティック区画スコアを生成するために、複数のシーケンシングリードを解析するステップと、(f) 複数のエピジェネティック区画スコアを複数のエピジェネティック区画カットオフとを含み得、コンピュータプロセッサを用いて実施することができる。この実施形態では、システムは、エピジェネティック対照核酸分子の添加、区画、富化、およびシーケンシングのための構成要素を含む。

#### 【0239】

別の実施形態では、ポリヌクレオチドの試料中の核酸分子をエピジェネティック状態に基づいて区画する方法を評価するためのシステムであって、通信ネットワークを通じて、核酸シーケンサーによって生成されたスパイクイン試料のシーケンシングリードのセットを受信する通信インターフェースであって、シーケンシングリードのセットが、(i) 試料を起源とするポリヌクレオチドから生成されたシーケンシングリードの少なくともある1つの第1の集団であって、第1の集団由来のシーケンシングリードが、タグ配列および試料を起源とするポリヌクレオチドに由来する配列を含む、第1の集団；ならびに(ii) エピジェネティック対照核酸分子に由来するシーケンシングリードの少なくともある1つの第2の集団であって、第2の集団から生成されたシーケンシングリードが、エピジェネティック修飾領域および必要に応じて識別子領域を含む、第2の集団を含む、通信インターフェースと、通信インターフェースと通信するコンピュータであって、1つまたは複

数のコンピュータプロセッサ、および、1つまたは複数のコンピュータプロセッサにより実行されると、(i)通信ネットワークを通じて、核酸シーケンサーによるシーケンシングリードの第1の集団および第2の集団由来のシーケンシングリードのセットを受信するステップと、(ii)エピジェネティック対照核酸分子および/または内在性対照分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(iii)1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法をインプリメントする、機械により実行可能なコードを含むコンピュータ可読媒体を含むコンピュータを含む、システム。

**【0240】**

別の実施形態では、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)核酸シーケンサーによって生成されたスパイクイン試料のシーケンシングリードのセットを得るステップであって、スパイクイン試料が、試料のポリヌクレオチドおよびエピジェネティック対照核酸分子を含み、シーケンシングリードのセットが、(i)試料のポリヌクレオチドから生成されたシーケンシングリードの第1の集団および(ii)エピジェネティック対照核酸分子に由来するシーケンシングリードの第2の集団を含む、ステップと、(b)エピジェネティック対照核酸分子および/または内在性対照分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(c)1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを実施する、非一時的なコンピュータで実行可能な命令を含むコンピュータ可読媒体を含む、またはそれにアクセスすることができるコントローラを含む、システム。

**【0241】**

別の実施形態では、少なくとも1つの電子プロセッサによって実行されると、少なくとも、(a)核酸シーケンサーによって生成された試料のシーケンシングリードのセットを得るステップであって、シーケンシングリードのセットが、試料のポリヌクレオチドから生成されたシーケンシングリードを含む、ステップと、(b)内在性対照分子の1つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、(c)1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを実施する、非一時的なコンピュータで実行可能な命令を含むコンピュータ可読媒体を含む、またはそれにアクセスすることができるコントローラを含むシステム。

**【0242】**

一部の実施形態では、システムは、g)区画方法の結果の状態をエピジェネティック区画スコアの比較に基づいて生成するステップをさらに含む。一部の実施形態では、区画方法の結果の状態を、(i)エピジェネティック対照核酸分子および/または内在性対照分子のセットの1つまたは複数のエピジェネティック区画スコアが対応するエピジェネティック区画カットオフの範囲内に入る場合には成功に分類するか、または、(ii)エピジェネティック対照核酸分子および/または内在性対照分子の1つまたは複数のエピジェネティック区画スコアのうちの少なくとも1つが対応するエピジェネティック区画カットオフの範囲外である場合には失敗に分類する。

**【0243】**

図8は、本開示の方法をインプリメントするようにプログラムされたまたは他のやり方で構成されたコンピュータシステム801を示す。コンピュータシステム801は、試料調製、シーケンシング、および/または解析の種々の側面を調節することができるものである。一部の実施例では、コンピュータシステム801は、試料調製および核酸シーケンシングを含めた試料解析を実施するように構成されている。

**【0244】**

コンピュータシステム801は、中央処理装置(CPU、本明細書では「プロセッサ」

10

20

30

40

50

および「コンピュータプロセッサ」とも) 805を含み、中央処理装置805は、シングルコアプロセッサもしくはマルチコアプロセッサ、または並行処理用の複数のプロセッサであり得る。コンピュータシステム801はまた、メモリまたはメモリ位置810(例えば、ランダムアクセスメモリ、リードオンリーメモリ、フラッシュメモリ)、電子ストレージユニット815(例えば、ハードディスク)、1つまたは複数の他のシステムと通信するための通信インターフェース820(例えば、ネットワークアダプター)、ならびにキャッシュ、他のメモリ、データストレージ、および/または電子ディスプレイアダプターなどの周辺機器825も含む。メモリ810、ストレージユニット815、インターフェース820、および周辺機器825は、CPU805と通信ネットワークまたはマザーボードなどのバス(実線)を通じて通信する。ストレージユニット815は、データ記憶用のデータストレージユニット(またはデータリポジトリ)であってよい。コンピュータシステム801は、コンピュータネットワーク430に通信インターフェース820を用いて作動可能にカップリングすることができる。コンピュータネットワーク830は、インターネット、インターネットおよび/もしくはエクストラネット、またはインターネットと通信するイントラネットおよび/もしくはエクストラネットであってよい。コンピュータネットワーク830は、一部の場合では、電気通信および/またはデータネットワークである。コンピュータネットワーク830に1つまたは複数のコンピュータサーバーを含めることができ、それにより、クラウドコンピューティングなどの分散コンピューティングを可能にすることができる。コンピュータネットワーク830は、一部の場合では、コンピュータシステム801を用いてピアツーピアネットワークをインプリメントすることができ、それにより、コンピュータシステム801とカップリングしたデバイスをクライアントまたはサーバーとして動作させることを可能にすることができる。

#### 【0245】

CPU805は、プログラムまたはソフトウェアに具体化することができる一連の機械可読命令を実行することができるものである。命令は、メモリ810などのメモリ位置に記憶させることができる。CPU405によって実施される操作の例は、フェッチ、復号、実行、およびライトバックを含み得る。

#### 【0246】

ストレージユニット815は、ドライバー、ライブラリー、および保存されたプログラムなどのファイルを記憶し得る。ストレージユニット815は、使用者によって作成されたプログラムおよび記録されたセッション、ならびにプログラムに関連する出力(単数または複数)を記憶し得る。ストレージユニット815は、使用者データ、例えば、使用者選好および使用者プログラムを記憶し得る。コンピュータシステム801は、一部の場合では、例えば、コンピュータシステム801とイントラネットまたはインターネットを通じて通信する遠隔サーバー上に位置するものなど、コンピュータシステム801に対して外付けの1つまたは複数の追加的なデータストレージユニットを含み得る。例えば通信ネットワークまたは物理的なデータ移行(例えば、ハードドライブ、サムドライブ、もしくは他のデータストレージ機構を使用する)を使用して、データを1つの場所から別の場所に移行することができる。

#### 【0247】

コンピュータシステム801は、1つまたは複数の遠隔コンピュータシステムとネットワーク830を通じて通信し得る。具体化のために、コンピュータシステム801は、使用者(例えば、オペレーター)の遠隔コンピュータシステムと通信し得る。遠隔コンピュータシステムの例としては、パーソナルコンピュータ(例えば、携帯型PC)、スレートまたはタブレットPC(例えば、Apple(登録商標)iPad(登録商標)、Samsung(登録商標)Galaxy Tab)、電話機、スマートフォン(例えば、Apple(登録商標)iPhone(登録商標)、Android対応デバイス、BlackBerry(登録商標))、または携帯情報端末が挙げられる。使用者は、コンピュータシステム801にネットワーク830を介してアクセスすることができる。

#### 【0248】

10

20

30

40

50

本明細書に記載の方法は、コンピュータシステム 801 の電子ストレージ位置、例えば、メモリ 810 または電子ストレージユニット 815 などに記憶された機械（例えば、コンピュータプロセッサ）により実行可能なコードによってインプリメントすることができる。機械により実行可能なまたは機械可読コードは、ソフトウェアの形態で提供することができる。使用中、コードは、プロセッサ 805 により実行することができる。一部の 경우에는、コードをストレージユニット 815 から検索し、プロセッサ 805 によりいつでもアクセスできるようにメモリ 810 に記憶させることができる。いくつかの状況では、電子ストレージユニット 815 を除外することができ、機械により実行可能な命令をメモリ 810 に記憶させることができる。

**【0249】**

ある態様では、本開示は、少なくとも 1 つの電子プロセッサによって実行されると、(a) 核酸シーケンサーによって生成したシーケンシングリードのセットを得るステップと、(b) エピジェネティック対照核酸分子の 1 つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも 1 つのサブセットを解析するステップと、(f) 1 つまたは複数のエピジェネティック区画スコアを 1 つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法を実施する、コンピュータで実行可能な命令を含む非一時的なコンピュータ可読媒体を提供する。

**【0250】**

コードを、プリコンパイルし、コードを実行するように適合させたプロセッサを有する機械で使用するために構成することができる、または実行時間の間にコンパイルすることができる。コードは、コードがプリコンパイル様式または都度コンパイル様式で実行されることが可能になるように選択することができるプログラミング言語で供給され得る。

**【0251】**

コンピュータシステム 801 などの本明細書に記載のシステムおよび方法の態様は、プログラミングに具体化することができる。当該技術の種々の態様は、一般には、機械可読媒体の一種で行われるまたは具体化される機械（またはプロセッサ）により実行可能なコードおよび/または関連データの形態の「製品」または「製造品」と考えることができる。機械により実行可能なコードは、電子ストレージユニット、そのようなメモリ（例えば、リードオンリーメモリ、ランダムアクセスメモリ、フラッシュメモリ）またはハードディスクに記憶させることができる。「ストレージ」型媒体は、ソフトウェアプログラミングのためにいつでも非一時的ストレージを提供し得るコンピュータ、プロセッサなど任意のまたは全ての有形メモリ、または例えば種々の半導体メモリ、テープドライブ、ディスクドライブなどそれに関連するモジュールを含み得る。

**【0252】**

ソフトウェアの全てまたは一部を、時々、インターネットまたは種々の他の電気通信ネットワークを通じて通信させることができる。そのような通信により、例えば、1 つのコンピュータまたはプロセッサから別のコンピュータまたはプロセッサに、例えば、管理サーバーまたはホストコンピュータからアプリケーションサーバーのコンピュータプラットフォームにソフトウェアをロードすることが可能になり得る。したがって、ソフトウェア要素を有し得る別の型の媒体として、有線および光陸線ネットワークを通じて、ならびに種々のエアリンクによってローカルデバイス間で物理的インターフェースを横切って使用されるものなどの光波、電波、および電磁波が挙げられる。例えば有線または無線リンク、光リンクなどの、そのような波を運ぶ物理的要素もソフトウェアを有する媒体とみなすことができる。本明細書で使用される場合、非一時的なものに制限される場合を除き、有形「ストレージ」媒体、コンピュータまたは機械「可読媒体」などの用語は、実行のためのプロセッサへの命令の提供に関与する任意の媒体を指す。

**【0253】**

したがって、コンピュータで実行可能なコードなどの機械可読媒体は、これだけに限定されないが、有形記憶媒体、搬送波媒体または物理的伝送媒体を含めた多くの形態を取り得る。非揮発性記憶媒体としては、例えば、図に示されているデータベースなどをインプ

10

20

30

40

50

リメントするために使用することができるものなどの任意のコンピュータ（単数または複数）などにおけるストレージデバイスのいずれかなどの光学ディスクまたは磁気ディスクが挙げられる。揮発性記憶媒体としては、そのようなコンピュータプラットフォームメインメモリなどのダイナミックメモリが挙げられる。有形伝達媒体としては、同軸ケーブル；コンピュータシステム内のバスを含む電線を含めた銅線および光ファイバーが挙げられる。搬送波伝送媒体は高周波（RF）および赤外（IR）データ通信中に生成するものなどの、電気シグナルもしくは電磁気シグナル、または音波もしくは光波の形態を取り得る。したがって、コンピュータ可読媒体の一般的な形態としては、例えば、フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、CD-ROM、DVDまたはDVD-ROM、任意の他の光学媒体、パンチカード、紙テープ、穴のパターンを有する任意の他の物理的記憶媒体、RAM、ROM、PROMおよびEPROM、FLASH（登録商標）-EPROM、任意の他のメモリチップもしくはカートリッジ、搬送波伝達データもしくは命令、そのような搬送波を運ぶケーブルもしくはリンク、またはコンピュータがプログラミングコードおよび/またはデータを読み取ることができる任意の他の媒体が挙げられる。これらの形態のコンピュータ可読媒体の多くは、実行のためにプロセッサに1つまたは複数の命令の1つまたは複数のシーケンスを運ぶことに関与し得る。

10

## 【0254】

コンピュータシステム801は、例えば、試料解析の1つまたは複数の結果を提供するためのユーザーインタフェース（UI）を含む電子ディスプレイを含み得るまたはそれと通信し得る。UIの例としては、限定することなく、グラフィカルユーザーインタフェース（GUI）およびウェブに基づくユーザーインタフェースが挙げられる。

20

## 【0255】

コンピュータシステムおよびネットワーク、データベース、およびコンピュータプログラム製品に関する追加的な詳細は、例えば、それぞれの全体が参照により本明細書に組み込まれる、Peterson, Computer Networks: A Systems Approach, Morgan Kaufmann, 5th Ed. (2011)、Kurose, Computer Networking: A Top-Down Approach, Pearson, 7th Ed. (2016)、Elmasri, Fundamentals of Database Systems, Addison Wesley, 6th Ed. (2010)、Coronel, Database Systems: Design, Implementation, & Management, Cengage Learning, 11th Ed. (2014)、Tucker, Programming Languages, McGraw-Hill Science/Engineering/Math, 2nd Ed. (2006)、およびRhoton, Cloud Computing Architected: Solution Design Handbook, Recursive Press (2011)においても提供される。

30

## V. 適用

## A. がんおよび他の疾患

## 【0256】

一部の実施形態では、本明細書に開示される方法およびシステムを使用して、患者における所与の疾患または状態を核酸変異体が体細胞起源であるかまたは生殖細胞系列起源であるかの分類に基づいて処置するためのカスタマイズされたまたは標的化治療を識別することができる。一般には、考慮される疾患は、がんの一種である。そのようながんの非限定的な例としては、胆道がん、膀胱がん、移行上皮癌、尿路上皮癌、脳がん、神経膠腫、星状細胞腫、乳癌、化生性癌、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大腸がん、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、腎明細胞癌、移行上皮癌、尿路上皮癌、ウィルムス腫瘍、白血病、急性リンパ性白血病（ALL）、急性骨髄性白血病（AML）、慢性リンパ性白血病（CLL）、慢性骨髄性白血病（CML）、慢性骨髄単球性白血病（CMML）、肝がん、肝癌、ヘパトーマ、肝細胞癌、胆管細胞癌、肝芽腫、肺がん、非小細胞肺癌（NSCLC）、中皮腫、B細胞リンパ腫、非ホジキンリンパ腫、びまん性大細胞型B細胞性リンパ腫、マントル細胞リンパ腫、T細胞リンパ腫、非ホジキンリン

40

50

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、上咽頭癌（NPC）、神経芽細胞腫、中咽頭がん、口腔扁平上皮癌、骨肉腫、卵巣癌、膵がん、膵管腺癌、偽乳頭状新生物、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、小腸癌腫、胃がん、胃癌、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。

【0257】

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

B. 治療および関連する施行

【0258】

ある特定の実施形態では、本明細書に開示される方法は、核酸変異体が体細胞起源であるかまたは生殖細胞系列起源であるかの状態を考慮して、患者に対するカスタマイズされた治療を識別し、施行することに関する。一部の実施形態では、基本的にあらゆるがん治療（例えば、外科療法、放射線療法、化学療法、および/または同様のもの）をこれらの方法の一部として含めることができる。一般には、カスタマイズされた治療は、少なくとも1つの免疫療法（または免疫療法剤）を含む。免疫療法は、一般に、所与のがんの型に対する免疫応答を増強する方法を指す。ある特定の実施形態では、免疫療法は、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。

【0259】

ある特定の実施形態では、対象由来の試料由来の核酸変異体が体細胞起源であるかまたは生殖細胞系列起源であるかの状態を、参照集団から得られた比較器であるデータベースと比して、その対象に対するカスタマイズされたまたは標的化治療を識別することができる。一般には、参照集団は、試験対象と同じがんもしくは疾患の型を有する患者、および/または、試験対象と同じ治療を受けているもしくは受けた患者を含む。核酸変異体および比較器である結果がある特定の分類基準を満たす（例えば、実質的にまたはほぼマッチする）場合に、カスタマイズされたまたは標的化治療（または複数の治療）を識別することができる。

【0260】

ある特定の実施形態では、本明細書に記載のカスタマイズされた治療は、一般には非経口投与される（例えば、静脈内または皮下）。免疫療法剤を含有する医薬組成物は一般には静脈内に投与される。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）を、例えば、頬側、舌下、直腸、膣、尿道内、局部、眼内、鼻腔内、および/または耳介内を含めた当技術分野で公知の任意の方法によって投与することもでき、投与は、錠剤、カプセル剤、顆粒剤、水性懸濁剤、ゲル剤、噴霧剤、坐剤、膏薬、軟膏剤などを含み得る。

【実施例】

【0261】

（実施例1）

無細胞DNA試料の区画の評価

患者由来の無細胞DNA試料を本実施例において解析する。無細胞DNA試料とエピジ

10

20

30

40

50

エネティック対照核酸分子のセットを組み合わせることによってスパイクイン試料を創出する。本実施例では、エピジェネティック対照核酸分子は二本鎖DNA分子であり、エピジェネティック対照核酸分子のセットはエピジェネティック対照核酸分子の6つの異なるサブセット(サブセット1~サブセット6)のプールである。サブセット1、サブセット2、サブセット3、サブセット4、サブセット5およびサブセット6は、エピジェネティック修飾領域内にメチル化シトシン(5-メチルシトシン)を0個、1個、3個、5個、7個、および9個有するエピジェネティック対照核酸分子を含む。エピジェネティック対照核酸分子はエピジェネティック修飾領域の一方の末端に分子バーコードを有し、エピジェネティック状態バーコードはエピジェネティック修飾領域の両末端に存在する。本実施例で使用する分子バーコードは一意的分子バーコードである、すなわち、各エピジェネティック対照核酸分子が別個の分子バーコードを有する。

10

#### 【0262】

次いで、このスパイクイン試料をメチル結合性ドメイン(MBD)緩衝剤およびMBDタンパク質とコンジュゲートした磁気ビーズと合わせ、終夜インキュベートする。このインキュベーション中にメチル化DNA(無細胞DNA試料中に存在する場合)およびメチル化エピジェネティック対照核酸分子にMBDタンパク質が結合する。メチル化されていないまたはメチル化が少ないDNAを、漸増濃度の塩を含有する緩衝剤を用いてビーズから洗い流す。最後に、高塩濃度の緩衝剤を使用して、重度にメチル化されたDNAをMBDタンパク質から洗い流す。これらの洗浄の結果、メチル化が漸増するDNAの3つの区画(3つの区画化セット、低、中、高)がもたらされる。区画化セット内に存在する区画化DNAは、無細胞DNA試料およびエピジェネティック対照核酸分子に由来するDNAを含む。3つの区画化セット内の区画化DNAを清澄化して塩を除去し、ライブラリー調製の酵素的ステップに備えて濃縮する。

20

#### 【0263】

区画化セット内のDNAを濃縮した後、区画化DNAの末端突出を伸長させ、断片の3'末端にアデノシン残基を付加する。各断片の5'末端をリン酸化する。これらの修飾により、区画化DNAがライゲーション可能になる。DNAリガーゼおよびアダプターを添加して、各区画化DNA分子を各末端のアダプターとライゲーションする。これらのアダプターは非一意的バーコードを含有し、各区画化セットは他の区画化セットに使用されるアダプター内のバーコードと区別可能な非一意的バーコードを有するアダプターとライゲーションする。ライゲーション後、3つの区画化セットを一緒にプールし、PCRによって増幅する。

30

#### 【0264】

PCR後、増幅したDNAを再度清澄化し、濃縮した後、富化させる。濃縮させたら、増幅したDNAを塩緩衝剤および特定の目的の領域を標的とするビオチン化RNAプローブと合わせ、エピジェネティック対照核酸分子およびこの混合物を終夜インキュベートする。ビオチン化RNAプローブをストレプトアビジン磁気ビーズによって捕捉し、一連の塩洗浄によって捕捉されなかった増幅したDNAと分離し、それにより、試料を富化させる。富化後、PCR増幅を介して試料インデックスを富化された分子に組み入れる。PCR増幅後、異なる試料(バッチ内)由来の増幅した分子を一緒にプールし、Illumina NovaSeqシーケンサーを使用してシーケンシングを行う。

40

#### 【0265】

次いで、シーケンサーによって生成された配列リードを、バイオインフォマティクスツール/アルゴリズムを使用して解析して、3つの区画化セットのそれぞれに存在する各サブセットに属するエピジェネティック対照核酸分子のエピジェネティック区画スコアを生成する。図9Aは、高区画化セット内の本実施例に記載の6つのサブセット(サブセット1、サブセット2、サブセット3、サブセット4、サブセット5およびサブセット6)のそれぞれに属するエピジェネティック対照核酸分子のエピジェネティック区画スコアのグラフィカルプロットを示す。図9Bは、中区画化セット内の6つのサブセットのそれぞれに属するエピジェネティック対照核酸分子のエピジェネティック区画スコアのグラフィカ

50

ルプロットを示す。図9Cは、低区画化セット内の6つのサブセットのそれぞれに属するエピジェネティック対照核酸分子のエピジェネティック区画スコアのグラフィカルプロットを示す。図9に示されている通り、高区画化セット、中区画化セットおよび低区画化セット内のサブセット1のエピジェネティック区画スコアは、それぞれ約0.1%、0.3%および99.6%である。高区画化セット、中区画化セットおよび低区画化セット内のサブセット1についての所定のエピジェネティック区画カットオフは、それぞれ0.3%、0.5%および97%である。本明細書では、エピジェネティック区画スコアをパーセンテージで表す。区画化セットのそれぞれにおけるサブセット1のエピジェネティック区画スコアをサブセット1についての対応するエピジェネティックカットオフと比較する。すなわち、高区画化セット内のサブセット1のエピジェネティック区画スコア(0.1%)を高区画化セット内のサブセット1についてのエピジェネティック区画カットオフ(0.3%)と比較する。高区画化セット内のサブセット1のエピジェネティック区画スコア(0.1%)は高区画化セット内のサブセット1についてのエピジェネティック区画カットオフ(0.3%)の範囲内に入る。同様に、中区画化セットおよび低区画化セット内のサブセット1のエピジェネティック区画スコアを中区画化セットおよび低区画化セット内のサブセット1についてのそれぞれのエピジェネティック区画カットオフと比較する。同様に、サブセット2、サブセット3、サブセット4、サブセット5およびサブセット6のエピジェネティック区画スコアを3つ全ての区画化セット内のそれぞれのエピジェネティック区画カットオフと比較する。したがって、合計18のエピジェネティック区画スコア(3つの区画化セット内の6つのサブセットについて、 $6 \times 3 = 18$ )を有し、これらのエピジェネティック区画スコアのそれぞれを対応するエピジェネティック区画カットオフと比較する。18のエピジェネティック区画スコアの全てがそれぞれのエピジェネティック区画カットオフの範囲内に入ることが見出された。したがって、本実施例で解析した無細胞DNA試料に対して実施した区画方法は成功に分類される。

(実施例2)

無細胞DNA試料の区画の評価

【0266】

患者のセット由来の無細胞DNA試料のセットを本実施例において解析する。本実施例では、エピジェネティック対照核酸分子は使用しない。その代わりに、無細胞DNA試料中の内在性対照分子を、無細胞DNA試料の区画を査定するために使用する。各患者の無細胞DNA試料をメチル結合性ドメイン(MBD)緩衝剤およびMBDタンパク質とコンジュゲートした磁気ビーズと合わせ、終夜インキュベートする。このインキュベーション中にメチル化DNA(無細胞DNA試料中に存在する場合)およびメチル化エピジェネティック対照核酸分子にMBDタンパク質が結合する。メチル化されていないまたはメチル化が少ないDNAを、漸増濃度の塩を含有する緩衝剤を用いてビーズから洗い流す。最後に、高塩濃度の緩衝剤を使用して、重度にメチル化されたDNAをMBDタンパク質から洗い流す。これらの洗浄の結果、メチル化が漸増するDNAの3つの区画(3つの区画化セット、低、中、高)がもたらされる。区画化セット内に存在する区画化DNAは、無細胞DNA試料およびエピジェネティック対照核酸分子に由来するDNAを含む。3つの区画化セット内の区画化DNAを清澄化して塩を除去し、ライブラリー調製の酵素的ステップに備えて濃縮する。

【0267】

区画化セット内のDNAを濃縮した後、区画化DNAの末端突出を伸長させ、断片の3'末端にアデノシン残基を付加する。各断片の5'末端をリン酸化する。これらの修飾により、区画化DNAがライゲーション可能になる。DNAリガーゼおよびアダプターを添加して、各区画化DNA分子を各末端のアダプターとライゲーションする。これらのアダプターは非一意的バーコードを含有し、各区画化セットは他の区画化セットに使用されるアダプター内のバーコードと区別可能な非一意的バーコードを有するアダプターとライゲーションする。ライゲーション後、3つの区画化セットを一緒にプールし、PCRによって増幅する。

10

20

30

40

50

## 【0268】

PCR後、増幅したDNAを再度清澄化し、濃縮した後、富化させる。濃縮させたら、増幅したDNAを塩緩衝剤および特定の目的の領域を標的とするビオチン化RNAプローブと合わせ、エピジェネティック対照核酸分子およびこの混合物を終夜インキュベートする。ビオチン化RNAプローブをストレプトアビジン磁気ビーズによって捕捉し、一連の塩洗浄によって捕捉されなかった増幅したDNAと分離し、それにより、試料を付加させる。富化後、PCR増幅を介して試料インデックスを富化された分子に組み入れる。PCR増幅後、異なる試料(バッチ内)由来の増幅した分子と一緒にプールし、Illumina NovaSeqシーケンサーを使用してシーケンシングを行う。

## 【0269】

次いで、シーケンサーによって生成された配列リードを、バイオインフォマティクスツール/アルゴリズムを使用して解析して、1つまたは複数の内在性対照分子のエピジェネティック区画スコアを生成する。本実施例では、methyl-halfおよびmethyl-5をエピジェネティック区画スコアとして使用する。図10Aは、高区画化セット内の試料1の高メチル化対照分子のフラクションのグラフィカルプロットを示し、試料1のmethyl-halfスコアが11である。図10Bは、低区画化セット内の試料1の高メチル化分子のフラクションのグラフィカルプロットを示し、試料1のmethyl-5スコアは13である。図11Aは、高区画化セット内の試料2の高メチル化対照分子のフラクションのグラフィカルプロットを示し、試料2のmethyl-halfスコアは13である。図11Bは、低区画化セット内の試料2の高メチル化分子のフラクションのグラフィカルプロットを示し、試料2のmethyl-5スコアは決定することができない(図11Bに示されている)。本実施例では、methyl-halfおよびmethyl-5についてのエピジェネティック区画カットオフはそれぞれ15メチル化CGおよび20メチル化CGである。試料1のmethyl-halfおよびmethyl-5スコアは対応するエピジェネティック区画カットオフの範囲内に入る。しかし、試料2に関しては、methyl-halfスコアはその対応するエピジェネティック区画カットオフの範囲内に入るが、methyl-5スコアはその対応するエピジェネティック区画カットオフの範囲内に入らない。したがって、試料1の区画方法は成功に分類され、試料2の区画方法は失敗に分類される。

## 【0270】

本発明の好ましい実施形態が本明細書において示され、記載されているが、そのような実施形態が単に例として提供されていることは当業者には明白であろう。本発明は、本明細書の中で提供される特定の実施例によって限定されるものではない。本発明は上記の明細に関連して記載されているが、本明細書の実施形態の説明および図表は、限定の意味で解釈されることを意図するものではない。当業者は、本発明から逸脱することなく多数の変形、変化および置換をすぐに思いつくであろう。さらに、本発明の全ての態様は、種々の条件および変数に左右される本明細書に記載の特定の描写、構成または相対的な割合に限定されないことが理解されるべきである。本明細書に記載の本開示の複数の実施形態に対する種々の代替を本発明の実施に使用することができることが理解されるべきである。したがって、本開示はあらゆるそのような代替、改変、変形または均等物を包含すべきことが意図されている。以下の特許請求の範囲により本発明の範囲が定義されること、ならびにこれらの特許請求の範囲の範囲内に入る方法および構造およびそれらの均等物がそれにより包含されることが意図されている。

## 【0271】

前述の開示は、明瞭さおよび理解のために実例としておよび例として一部の詳細が記載されているが、本開示を読めば、形態および詳細の種々の変化を本開示の真の範囲から逸脱することなく行うことができ、また、添付の特許請求の範囲の範囲内で実施することができることが当業者には明らかになる。例えば、方法、システム、コンピュータ可読媒体、および/または構成要素の特色、ステップ、要素、またはそれらの他の態様は全て種々の組合せで使用することができる。

10

20

30

40

50

## 【 0 2 7 2 】

本明細書において引用されている特許、特許出願、ウェブサイト、他の刊行物または文書、受託番号などは全て、各項目が参照により組み込まれることが具体的にかつ個別に示されたものと同じく、あらゆる目的に関してその全体が参照により組み込まれる。異なるバージョンの配列が違う時間に受託番号に関連付けられている場合、本出願の有効な出願日に受託番号に関連付けられたバージョンが意図される。有効な出願日とは、実際の出願日より前または該当する場合には受託番号を参照する優先出願の出願日を意味する。同様に、異なるバージョンの刊行物、ウェブサイトなどが違う時間に公開されている場合、別段の指定のない限り、本出願の有効な出願日のごく最近に公開されたバージョンが意図される。

10

本発明は、例えば、以下の項目を提供する。

(項目 1)

エピジェネティック対照核酸分子の 2 つまたはそれよりも多くのサブセットを含むエピジェネティック対照核酸分子のセットであって、

前記エピジェネティック対照核酸分子の 2 つまたはそれよりも多くのサブセットのうちの 1 つのサブセットが、エピジェネティック修飾領域を含む複数のエピジェネティック対照核酸分子を含む、

エピジェネティック対照核酸分子のセット。

(項目 2)

前記エピジェネティック対照核酸分子が、識別子領域をさらに含む、項目 1 に記載のエピジェネティック対照核酸分子のセット。

20

(項目 3)

前記識別子領域が、前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域の片側または両側に存在する、項目 2 に記載のエピジェネティック対照核酸分子のセット。

(項目 4)

少なくとも 1 つのサブセット内の前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、エピジェネティック修飾を有するヌクレオチドを少なくとも 1 つ含む、上記項目のいずれかに記載のエピジェネティック対照核酸分子のセット。

(項目 5)

前記サブセットが、エピジェネティック修飾を有するヌクレオチドを同数有するエピジェネティック対照核酸分子を含む、項目 4 に記載のエピジェネティック対照核酸分子のセット。

30

(項目 6)

第 1 のサブセット内のエピジェネティック修飾を有するヌクレオチドの数が、第 2 のサブセット内のエピジェネティック修飾を有するヌクレオチドの数と異なる、項目 4 に記載のエピジェネティック対照核酸分子のセット。

(項目 7)

前記エピジェネティック対照核酸分子の前記識別子領域が、分子バーコードを含む、項目 2 に記載のエピジェネティック対照核酸分子のセット。

40

(項目 8)

前記 2 つまたはそれよりも多くのサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、同一の核酸配列を含む、前記項目のいずれか一項に記載のエピジェネティック対照核酸分子のセット。

(項目 9)

前記識別子領域が、少なくとも 1 つのエピジェネティック状態バーコードをさらに含む、項目 2 に記載のエピジェネティック対照核酸分子のセット。

(項目 10)

前記識別子領域が、1 つまたは複数のプライマー結合部位を含む、項目 2 に記載のエピジェネティック対照核酸分子のセット。

50

## (項目 1 1)

第 1 のサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、第 2 のサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域の核酸配列とは区別可能な核酸配列を含む、項目 1 または 2 に記載のエピジェネティック対照核酸分子のセット。

## (項目 1 2)

前記エピジェネティック修飾が、DNA メチル化である、前記項目のいずれか一項に記載のエピジェネティック対照核酸分子のセット。

## (項目 1 3)

前記エピジェネティック修飾を有するヌクレオチドが、メチル化ヌクレオチドを含む、前記項目のいずれか一項に記載のエピジェネティック対照核酸分子のセット。

10

## (項目 1 4)

前記メチル化ヌクレオチドが、5 - メチルシトシンを含む、項目 1 3 に記載のエピジェネティック対照核酸分子のセット。

## (項目 1 5)

前記メチル化ヌクレオチドが、5 - ヒドロキシメチルシトシンを含む、項目 1 3 に記載のエピジェネティック対照核酸分子のセット。

## (項目 1 6)

エピジェネティック対照核酸分子の各サブセットが、等モル濃度で存在する、項目 1 に記載のエピジェネティック対照核酸分子のセット。

20

## (項目 1 7)

エピジェネティック対照核酸分子の各サブセットが、非等モル濃度で存在する、項目 1 に記載のエピジェネティック対照核酸分子のセット。

## (項目 1 8)

前記サブセットのうちの少なくとも 1 つにおける前記エピジェネティック対照核酸分子内のメチル化ヌクレオチドの数が、0 個、1 個、2 個、3 個、4 個、5 個、6 個、7 個、8 個、9 個、10 個、11 個、少なくとも 12 個、少なくとも 15 個、少なくとも 20 個、少なくとも 25 個、少なくとも 30 個、少なくとも 40 個または少なくとも 50 個である、項目 1 3 に記載のエピジェネティック対照核酸分子のセット。

## (項目 1 9)

前記エピジェネティック対照核酸分子が、ラムダファージ DNA、ヒトゲノム領域またはその両方の組合せに対応する配列を含む、項目 1 に記載のエピジェネティック対照核酸分子のセット。

30

## (項目 2 0)

( i ) エピジェネティック対照核酸分子の 2 つまたはそれよりも多くのサブセットを含むエピジェネティック対照核酸分子のセットであって、

前記エピジェネティック対照核酸分子の 2 つまたはそれよりも多くのサブセットのうちの 1 つのサブセットが、エピジェネティック修飾領域を含む複数のエピジェネティック対照核酸分子を含む、エピジェネティック対照核酸分子のセットと、

( i i ) 対象由来のポリヌクレオチドの試料中の核酸分子のセットとを含む、核酸の集団。

40

## (項目 2 1)

前記エピジェネティック対照核酸分子が、識別子領域をさらに含む、項目 2 0 に記載の核酸の集団。

## (項目 2 2)

前記識別子領域が、前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域の片側または両側に存在する、項目 2 1 に記載の核酸の集団。

## (項目 2 3)

前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、エピジェネティック修飾を有するヌクレオチドを少なくとも 1 つ含む、上記項目のいずれかに記載

50

の核酸の集団。

(項目 2 4)

前記サブセットが、エピジェネティック修飾を有するヌクレオチドを同数有するエピジェネティック対照核酸分子を含む、項目 2 3 に記載の核酸の集団。

(項目 2 5)

第 1 のサブセット内のエピジェネティック修飾を有するヌクレオチドの数が、第 2 のサブセット内のエピジェネティック修飾を有するヌクレオチドの数と異なる、項目 2 3 に記載の核酸の集団。

(項目 2 6)

前記エピジェネティック対照核酸分子の前記識別子領域が、分子バーコードを含む、項目 2 1 に記載の核酸の集団。

10

(項目 2 7)

前記 2 つまたはそれよりも多くのサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、同一の核酸配列を含む、前記項目のいずれか一項に記載の核酸の集団。

(項目 2 8)

前記識別子領域が、少なくとも 1 つのエピジェネティック状態バーコードをさらに含む、項目 2 1 に記載の核酸の集団。

(項目 2 9)

前記識別子領域が、1 つまたは複数のプライマー結合部位を含む、項目 2 1 に記載の核酸の集団。

20

(項目 3 0)

第 1 のサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、第 2 のサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域の核酸配列とは区別可能な核酸配列を含む、項目 2 0 または 2 1 に記載の核酸の集団。

(項目 3 1)

前記エピジェネティック修飾が、DNAメチル化である、前記項目のいずれか一項に記載の核酸の集団。

(項目 3 2)

前記エピジェネティック修飾を有するヌクレオチドが、メチル化ヌクレオチドを含む、前記項目のいずれか一項に記載の核酸の集団。

30

(項目 3 3)

前記メチル化ヌクレオチドが、5 - メチルシトシンを含む、項目 3 2 に記載の核酸の集団。

(項目 3 4)

前記メチル化ヌクレオチドが、5 - ヒドロキシメチルシトシンを含む、項目 3 2 に記載の核酸の集団。

(項目 3 5)

エピジェネティック対照核酸分子の各サブセットが、等モル濃度で存在する、項目 2 0 に記載の核酸の集団。

40

(項目 3 6)

エピジェネティック対照核酸分子の各サブセットが、非等モル濃度で存在する、項目 2 0 に記載の核酸の集団。

(項目 3 7)

前記サブセットのうちの少なくとも 1 つにおける前記エピジェネティック対照核酸分子内のメチル化ヌクレオチドの数が、0 個、1 個、2 個、3 個、4 個、5 個、6 個、7 個、8 個、9 個、1 0 個、1 1 個、少なくとも 1 2 個、少なくとも 1 5 個、少なくとも 2 0 個、少なくとも 2 5 個、少なくとも 3 0 個、少なくとも 4 0 個または少なくとも 5 0 個である、項目 3 2 に記載の核酸の集団。

50

## (項目 3 8)

前記エピジェネティック対照核酸分子が、ラムダファージDNA、ヒトゲノム領域またはその両方の組合せに対応する配列を含む、項目 2 0 に記載の核酸の集団。

## (項目 3 9)

ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、

a . エピジェネティック対照核酸分子のセットを前記ポリヌクレオチドの試料中の前記核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、

b . 前記スパイクイン試料の少なくとも 1 つのサブセットの核酸分子を複数の区画化セットに区画するステップと、

c . 前記複数の区画化セットから分子の少なくとも 1 つのサブセットを富化させて、富化された分子のセットを生成するステップであって、前記富化された分子のセットが、エピジェネティック対照核酸分子の群および前記ポリヌクレオチドの試料由来の核酸分子の群を含む、ステップと、

d . 前記富化された分子のセットの少なくとも 1 つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、

e . 前記エピジェネティック対照核酸分子の 1 つまたは複数のエピジェネティック区画スコアを生成するために、前記シーケンシングリードのセットの少なくとも 1 つのサブセットを解析するステップと、

f . 前記 1 つまたは複数のエピジェネティック区画スコアを 1 つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法。

## (項目 4 0)

ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、

a . エピジェネティック対照核酸分子のセットを前記ポリヌクレオチドの試料中の前記核酸分子に添加し、それにより、スパイクイン試料を作製するステップと、

b . 前記スパイクイン試料の少なくとも 1 つのサブセットの核酸分子を複数の区画化セットに区画するステップと、

c . 前記複数の区画化セットから分子の少なくとも 1 つのサブセットを富化させて、富化された分子のセットを生成するステップであって、前記富化された分子のセットが、エピジェネティック対照核酸分子の群および前記ポリヌクレオチドの試料由来の核酸分子の群を含み、前記ポリヌクレオチドの試料由来の核酸分子の群が、内在性対照分子のセットを含む、ステップと、

d . 前記富化された分子のセットの少なくとも 1 つのサブセットをシーケンシングして、シーケンシングリードのセットを作製するステップと、

e . 前記エピジェネティック対照核酸分子および前記内在性対照分子のセットについて 1 つまたは複数のエピジェネティック区画スコアを生成するために、前記シーケンシングリードのセットの少なくとも 1 つのサブセットを解析するステップと、

f . 前記 1 つまたは複数のエピジェネティック区画スコアを 1 つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法。

## (項目 4 1)

ポリヌクレオチドの試料中の核酸分子のエピジェネティック状態に基づく区画を評価するための方法であって、

a . 前記ポリヌクレオチドの試料の少なくとも 1 つのサブセット由来の核酸分子を複数の区画化セットに区画するステップと、

b . 前記複数の区画化セットから分子の少なくとも 1 つのサブセットを富化させて、富化された分子のセットを生成するステップであって、前記富化された分子のセットが、前記ポリヌクレオチドの試料由来の核酸分子の群を含み、前記ポリヌクレオチドの試料由来

10

20

30

40

50

の核酸分子の群が、内在性対照分子のセットを含む、ステップと、

c. 前記富化された分子のセットの少なくとも1つのサブセットをシーケンシングして、複数のシーケンシングリードを作製するステップと、

d. 前記内在性対照分子のセットについて1つまたは複数のエピジェネティック区画スコアを生成するために、前記シーケンシングリードのセットの少なくとも1つのサブセットを解析するステップと、

e. 前記1つまたは複数のエピジェネティック区画スコアを1つまたは複数のエピジェネティック区画カットオフと比較するステップとを含む方法。

(項目42)

前記複数の区画化セットのうちのある1つの区画化セット内の核酸分子に、タグのセットを用いてタグ付けして、タグ付けされた核酸分子の集団を作製するステップであって、前記タグ付けされた核酸分子が、1つまたは複数のタグを含む、ステップをさらに含む、項目39、40または41に記載の方法。

(項目43)

前記複数の区画化セットのうち第1の区画化セットに使用される前記タグのセットが、前記複数の区画化セットのうち第2の区画化セットに使用される前記タグのセットと異なる、項目42に記載の方法。

(項目44)

前記タグのセットを前記核酸分子に、アダプターを前記核酸分子にライゲーションすることによって付着させ、前記アダプターが、1つまたは複数のタグを含む、項目43に記載の方法。

(項目45)

g) (i) 前記エピジェネティック対照核酸分子および/または前記内在性対照分子のセットの前記1つまたは複数のエピジェネティック区画スコアが前記対応するエピジェネティック区画カットオフの範囲内に入る場合には、成功と分類するか、または、(ii) 前記エピジェネティック対照分子および/または前記内在性対照分子のセットの前記1つまたは複数のエピジェネティック区画スコアのうちの少なくとも1つが前記対応するエピジェネティック区画カットオフの範囲外である場合には、失敗に分類するステップをさらに含む、項目39、40または41に記載の方法。

(項目46)

前記エピジェネティック対照核酸分子のセットが、エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットを含み、

前記エピジェネティック対照核酸分子の2つまたはそれよりも多くのサブセットのうちの1つのサブセットが、エピジェネティック修飾領域を含む複数のエピジェネティック対照核酸分子を含む、項目39または40に記載の方法。

(項目47)

前記エピジェネティック対照核酸分子が、識別子領域をさらに含む、項目46に記載の方法。

(項目48)

前記識別子領域が、前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域の片側または両側に存在する、項目47に記載の方法。

(項目49)

少なくとも1つのサブセット内の前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、エピジェネティック修飾を有する少なくとも1つヌクレオチドを含む、上記項目のいずれかに記載の方法。

(項目50)

前記サブセットが、エピジェネティック修飾を有するヌクレオチドを同数有するエピジェネティック対照核酸分子を含む、項目49に記載の方法。

(項目51)

10

20

30

40

50

第1のサブセット内のエピジェネティック修飾を有するヌクレオチドの数が、第2のサブセット内のエピジェネティック修飾を有するヌクレオチドの数と異なる、項目49に記載の方法。

(項目52)

前記エピジェネティック対照核酸分子の前記識別子領域が、分子バーコードを含む、項目47に記載の方法。

(項目53)

前記2つまたはそれよりも多くのサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、同一の核酸配列を含む、前記項目のいずれか一項に記載の方法。

(項目54)

前記識別子領域が、少なくとも1つのエピジェネティック状態バーコードをさらに含む、項目47に記載の方法。

(項目55)

前記識別子領域が、1つまたは複数のプライマー結合部位を含む、項目47に記載の方法。

(項目56)

第1のサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、第2のサブセット内の前記複数のエピジェネティック対照核酸分子の前記エピジェネティック修飾領域の核酸配列とは区別可能な核酸配列を含む、項目46に記載の方法。

(項目57)

前記エピジェネティック修飾が、DNAメチル化である、前記項目のいずれか一項に記載の方法。

(項目58)

前記エピジェネティック修飾を有するヌクレオチドが、メチル化ヌクレオチドを含む、前記項目のいずれか一項に記載の方法。

(項目59)

前記メチル化ヌクレオチドが、5-メチルシトシンを含む、項目58に記載の方法。

(項目60)

前記メチル化ヌクレオチドが、5-ヒドロキシメチルシトシンを含む、項目58に記載の方法。

(項目61)

エピジェネティック対照核酸分子の各サブセットが、等モル濃度で存在する、項目46に記載の方法。

(項目62)

エピジェネティック対照核酸分子の各サブセットが、非等モル濃度で存在する、項目46に記載の方法。

(項目63)

前記サブセットのうちの少なくとも1つにおける前記エピジェネティック対照核酸分子内のメチル化ヌクレオチドの数が、0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、11個、少なくとも12個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも40個または少なくとも50個である、項目58に記載の方法。

(項目64)

前記エピジェネティック対照核酸分子が、ラムダファージDNAのゲノム領域、ヒトのゲノム領域、またはその両方の組合せに対応する配列を含む、項目46に記載の方法。

(項目65)

前記エピジェネティック状態が、前記核酸分子のメチル化レベルである、上記項目のいずれかに記載の方法。

10

20

30

40

50

## (項目 6 6)

前記複数の区画化セットが、前記核酸分子のメチル化レベルに基づいて区画された前記スパイクイン試料の核酸分子を含む、項目 3 9、4 0 または 4 1 に記載の方法。

## (項目 6 7)

前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、約 1 6 0 b p の長さで構成される、項目 4 6 に記載の方法。

## (項目 6 8)

前記複数の富化された分子のシーケンシングを核酸シーケンサーによって実施する、項目 3 9、4 0 または 4 1 に記載の方法。

## (項目 6 9)

前記核酸シーケンサーが次世代シーケンサーである、項目 6 8 に記載の方法。

10

## (項目 7 0)

前記エピジェネティック対照核酸分子の前記エピジェネティック修飾領域が、非ヒトゲノムに対応する核酸配列を含む、項目 4 6 に記載の方法。

## (項目 7 1)

前記ポリヌクレオチドの試料が、DNA の試料、RNA の試料、ポリヌクレオチドの試料、無細胞 DNA の試料、および無細胞 RNA の試料からなる群から選択される、前記項目のいずれかに記載の方法。

## (項目 7 2)

前記ポリヌクレオチドの試料が、無細胞 DNA である、前記項目のいずれかに記載の方法。

20

## (項目 7 3)

前記サブセットのうちの少なくとも 1 つにおける前記エピジェネティック対照核酸分子内のメチル化ヌクレオチドの数が、0 個、2 個、4 個、6 個、8 個、1 0 個、1 2 個、1 4 個、少なくとも 1 6 個、少なくとも 2 0 個、少なくとも 3 0 個、少なくとも 4 0 個または少なくとも 5 0 個である、項目 7 2 に記載の方法。

## (項目 7 4)

前記無細胞 DNA が、1 n g から 5 0 0 n g の間である、項目 7 2 に記載の方法。

## (項目 7 5)

前記エピジェネティック対照核酸分子が、1 フェムトモルから 2 0 0 フェムトモルの間である、前記項目のいずれかに記載の方法。

30

## (項目 7 6)

前記区画が、前記核酸分子を、エピジェネティック修飾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結合性物質に対する前記核酸分子の示差的な結合親和性に基づいて区画することを含む、前記項目のいずれかに記載の方法。

## (項目 7 7)

ポリヌクレオチドの試料中の核酸分子をエピジェネティック状態に基づいて区画する方法を評価するためのシステムであって、

通信ネットワークを通じて、核酸シーケンサーによって生成されたスパイクイン試料のシーケンシングリードのセットを受信する通信インターフェースであって、

40

前記シーケンシングリードのセットが、( i ) 前記試料を起源とするポリヌクレオチドから生成されたシーケンシングリードの少なくともある 1 つの第 1 の集団であって、前記第 1 の集団由来のシーケンシングリードが、タグ配列および前記試料を起源とするポリヌクレオチドに由来する配列を含む、第 1 の集団；および ( i i ) エピジェネティック対照核酸分子から生成されたシーケンシングリードの少なくともある 1 つの第 2 の集団であって、前記第 2 の集団から生成されたシーケンシングリードが、エピジェネティック修飾領域および必要に応じて識別子領域を含む、第 2 の集団を含む、通信インターフェースと、前記通信インターフェースと通信するコンピュータであって、1 つまたは複数のコンピュータプロセッサ、および、前記 1 つまたは複数のコンピュータプロセッサにより実行されると、

50

— ( i ) 前記通信ネットワークを通じて、前記核酸シーケンサーによるシーケンシングリードの前記第 1 の集団および第 2 の集団由来のシーケンシングリードのセットを受信するステップと、

— ( i i ) 前記エピジェネティック対照核酸分子および / または内在性対照分子の 1 つまたは複数のエピジェネティック区画スコアを生成するために、前記シーケンシングリードのセットの少なくとも 1 つのサブセットを解析するステップと、

— ( i i i ) 前記 1 つまたは複数のエピジェネティック区画スコアを 1 つまたは複数のエピジェネティック区画カットオフと比較するステップと、

を含む方法をインプリメントする、機械により実行可能なコードを含むコンピュータ可読媒体を含む、コンピュータと

を含む、システム。

(項目 7 8)

— 少なくとも 1 つの電子プロセッサによって実行されると、少なくとも、

— ( a ) 核酸シーケンサーによって生成されたスパイクイン試料のシーケンシングリードのセットを得るステップであって、前記スパイクイン試料が、試料のポリヌクレオチドおよびエピジェネティック対照核酸分子を含み、前記シーケンシングリードのセットが、( i ) 試料のポリヌクレオチドから生成されたシーケンシングリードの第 1 の集団および ( i i ) エピジェネティック対照核酸分子から生成されたシーケンシングリードの第 2 の集団を含む、ステップと、

— ( b ) 前記エピジェネティック対照核酸分子および / または内在性対照分子の 1 つまたは複数のエピジェネティック区画スコアを生成するために、前記シーケンシングリードのセットの少なくとも 1 つのサブセットを解析するステップと、

— ( c ) 前記 1 つまたは複数のエピジェネティック区画スコアを 1 つまたは複数のエピジェネティック区画カットオフと比較するステップと、

を実施する、非一時的なコンピュータで実行可能な命令を含むコンピュータ可読媒体を含む、またはそれにアクセスすることができるコントローラを含むシステム。

(項目 7 9)

— 少なくとも 1 つの電子プロセッサによって実行されると、少なくとも、

— a . 核酸シーケンサーによって生成された試料のシーケンシングリードのセットを得るステップであって、前記シーケンシングリードのセットが、前記試料のポリヌクレオチドから生成されたシーケンシングリードを含む、ステップと、

— b . 前記内在性対照分子の 1 つまたは複数のエピジェネティック区画スコアを生成するために、シーケンシングリードのセットの少なくとも 1 つのサブセットを解析するステップと、

— c . 前記 1 つまたは複数のエピジェネティック区画スコアを 1 つまたは複数のエピジェネティック区画カットオフと比較するステップと

を実施する、非一時的なコンピュータで実行可能な命令を含むコンピュータ可読媒体を含む、またはそれにアクセスすることができるコントローラを含むシステム。

(項目 8 0)

— g ) 前記区画方法の結果の状態を前記エピジェネティック区画スコアの比較に基づいて生成するステップをさらに含む、前記項目のいずれか一項に記載のシステム。

(項目 8 1)

— 前記区画方法の結果の状態が、( i ) 前記エピジェネティック対照核酸分子および / または前記内在性対照分子のセットの前記 1 つまたは複数のエピジェネティック区画スコアが前記対応するエピジェネティック区画カットオフの範囲内に入る場合には成功に分類されるか、または、( i i ) 前記エピジェネティック対照核酸分子および / または前記内在性対照分子の前記 1 つまたは複数のエピジェネティック区画スコアのうちの少なくとも 1 つが前記対応するエピジェネティック区画カットオフの範囲外である場合には失敗に分類される、項目 4 に記載のシステム。

(項目 8 2)

10

20

30

40

50

前記エピジェネティック区画スコアが、区画化セット内の高メチル化エピジェネティック対照核酸分子および/または高メチル化対照分子の数のフラクションまたはパーセンテージを含む、前記項目のいずれか一項に記載のシステム。

(項目 8 3)

前記エピジェネティック区画スコアが、区画化セット内の低メチル化エピジェネティック対照核酸分子および/または低メチル化対照分子の数のフラクションまたはパーセンテージを含む、項目 7 8 から 8 0 までのいずれか一項に記載のシステム。

(項目 8 4)

前記区画化セットが、高メチル化区画化セットである、項目 8 2 または 8 3 に記載のシステム。

(項目 8 5)

前記区画化セットが、低メチル化区画化セットである、項目 8 2 または 8 3 に記載のシステム。

(項目 8 6)

前記エピジェネティック区画スコアが、0 C G スコアである、項目 7 8 から 8 0 までのいずれか一項に記載のシステム。

(項目 8 7)

前記エピジェネティック区画スコアが、h y p o スコアである、項目 7 8 から 8 0 までのいずれか一項に記載のシステム。

(項目 8 8)

前記エピジェネティック区画スコアが、m e t h y l - h a l f である、項目 7 8 から 8 0 までのいずれか一項に記載のシステム。

(項目 8 9)

前記エピジェネティック区画スコアが、m e t h y l - 5 である、項目 7 8 から 8 0 までのいずれか一項に記載のシステム。

(項目 9 0)

前記 0 C G スコアについての前記エピジェネティック区画カットオフが、0 . 0 1 %、0 . 0 2 %、0 . 0 5 %、0 . 1 %、0 . 2 %、0 . 3 %、0 . 4 %、0 . 5 %、0 . 6 %、0 . 7 %、0 . 8 %、0 . 9 %、1 %、2 %、5 %、少なくとも 5 % または少なくとも 1 0 % である、項目 8 6 に記載のシステム。

(項目 9 1)

前記 h y p o スコアについての前記エピジェネティック区画カットオフが、0 . 1 %、0 . 5 %、1 %、2 %、3 %、4 %、5 %、7 % または少なくとも 1 0 % である、項目 8 7 に記載のシステム。

(項目 9 2)

前記 m e t h y l - h a l f についての前記エピジェネティック区画カットオフが、5、1 0、1 5、2 0、2 5、3 0、3 5 または 4 0 m C G である、項目 8 8 に記載のシステム。

(項目 9 3)

前記 m e t h y l - 5 についての前記エピジェネティック区画カットオフが、5、1 0、2 0、3 0、4 0 または 5 0 m C G である、項目 8 9 に記載のシステム。

(項目 9 4)

核酸分子の区画に関する情報および/または核酸分子の区画から導き出された情報を必要に応じて含むレポートを作成するステップをさらに含む、項目 3 9 から 9 3 までのいずれか一項に記載の方法またはシステム。

(項目 9 5)

前記レポートを、試料が由来する対象または健康管理実践者などの第三者に伝達するステップをさらに含む、項目 9 4 に記載の方法またはシステム。

10

20

30

40

50

【 図 面 】  
【 図 1 】

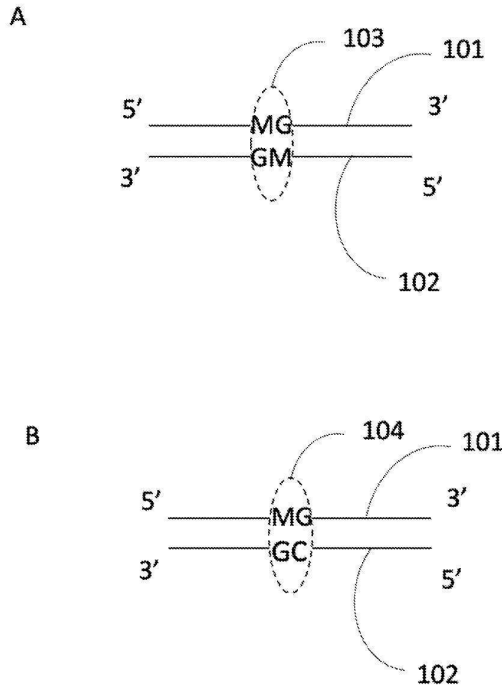


FIG. 1

【 図 2 】

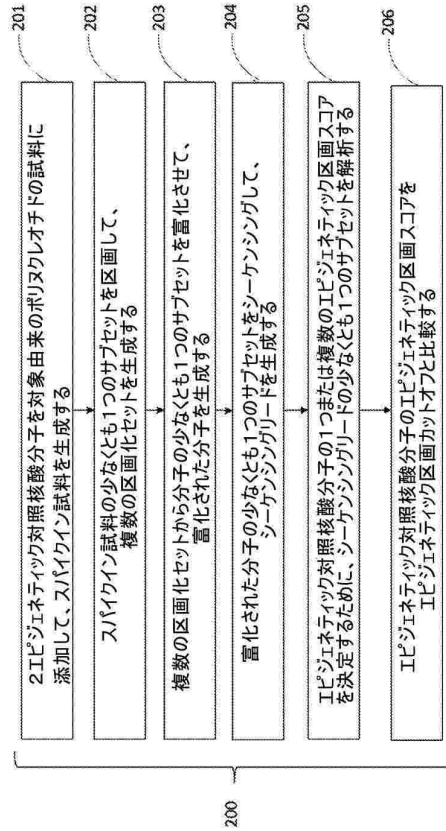


FIG. 2

【 図 3 】

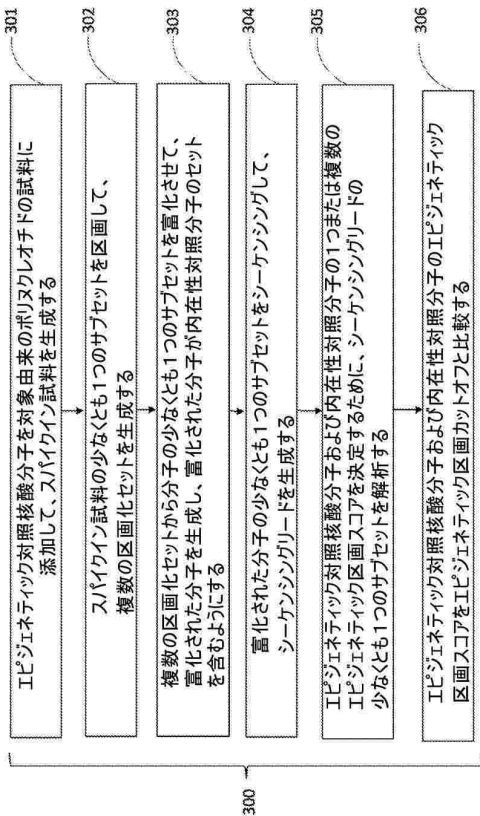


FIG. 3

【 図 4 】

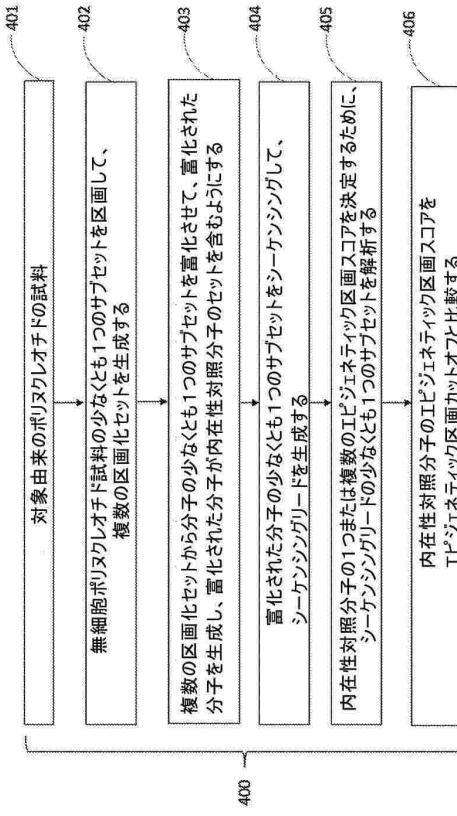


FIG. 4

10

20

30

40

50

【 図 5 】

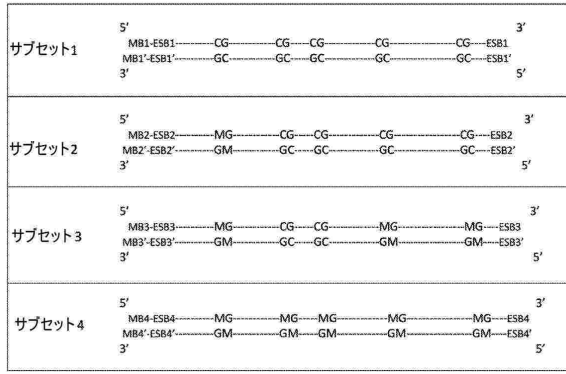


FIG. 5

【 図 6 】

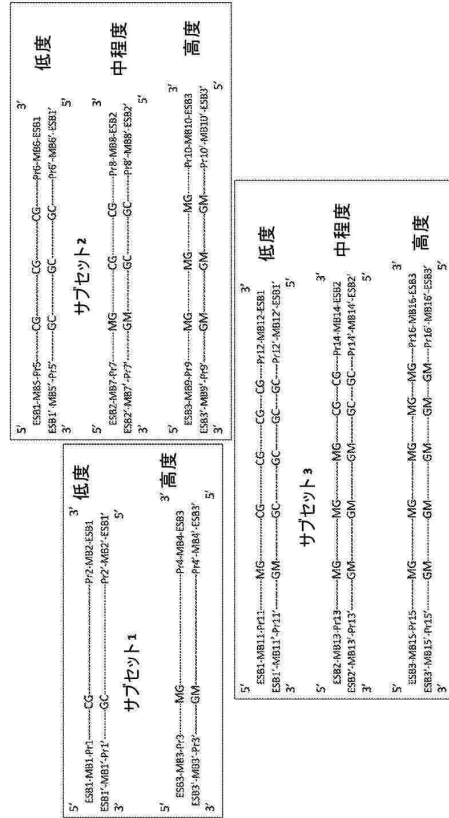


FIG. 6

【 図 7 】

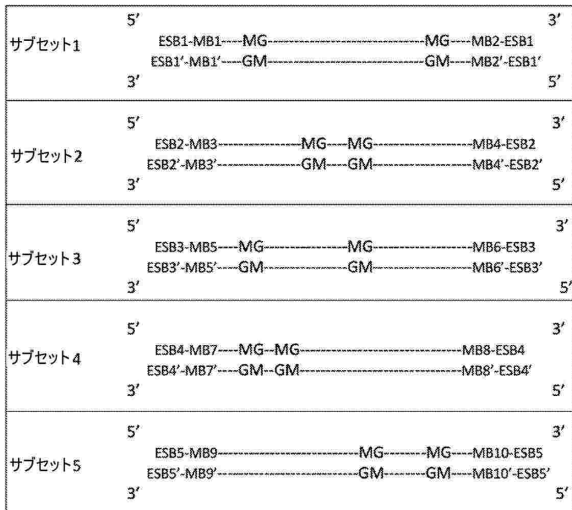


FIG. 7

【 図 8 】

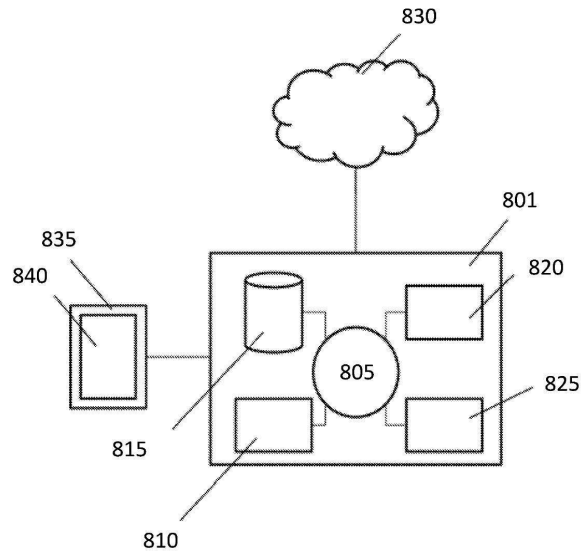


FIG. 8

10

20

30

40

50

【図 9】

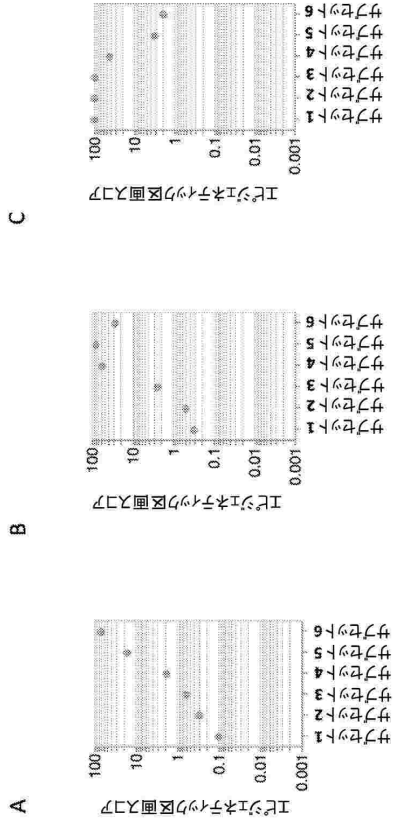


FIG. 9

【図 10】

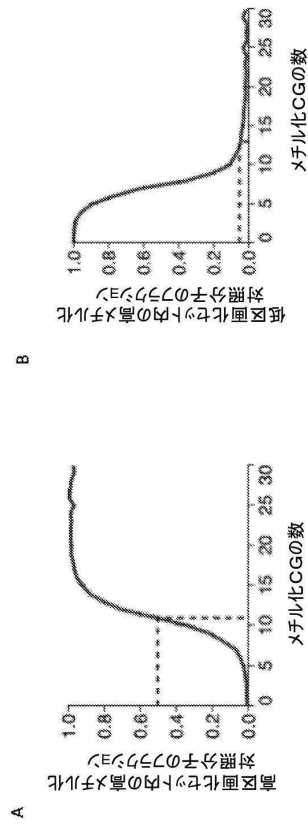


FIG. 10

【図 11】

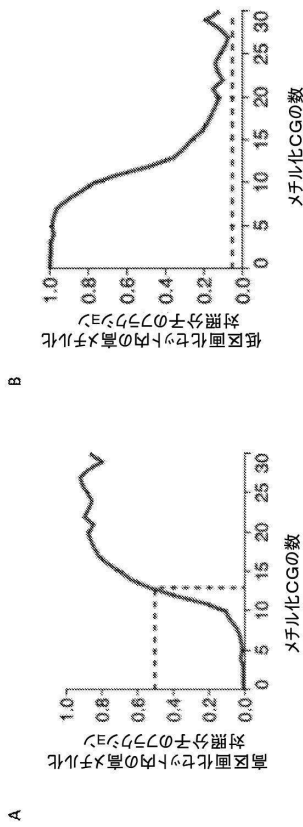


FIG. 11

10

20

30

40

50

## フロントページの続き

- 弁護士 山本 健策
- (72)発明者 ケネディ, アンドリュー  
アメリカ合衆国 カリフォルニア 9 2 0 3 7, ラ ホーヤ, モーニング ウェイ 3 1 2 1
- (72)発明者 ウェステソン, オスカー  
アメリカ合衆国 カリフォルニア 9 4 0 6 3, レッドウッド シティ, ペノブスコット ドライブ  
5 0 5
- (72)発明者 ヘ, ユペン  
アメリカ合衆国 カリフォルニア 9 4 0 6 3, レッドウッド シティ, ペノブスコット ドライブ  
5 0 5
- (72)発明者 シュルツ, マシュー  
アメリカ合衆国 カリフォルニア 9 4 0 6 3, レッドウッド シティ, ペノブスコット ドライブ  
5 0 5
- 審査官 中村 勇介
- (56)参考文献 米国特許出願公開第 2 0 1 0 / 0 1 0 5 0 6 3 ( U S , A 1 )  
Methods, 2010年, Vol.52, No.3, pp.232-236
- (58)調査した分野 (Int.Cl., D B 名)  
C 1 2 Q 1 / 0 0 - 3 / 0 0  
C 1 2 N 1 5 / 0 0 - 1 5 / 9 0  
C A p l u s / M E D L I N E / E M B A S E / B I O S I S ( S T N )