

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2023-544721

(P2023-544721A)

(43)公表日 令和5年10月25日(2023.10.25)

(51)国際特許分類		F I		テーマコード (参考)	
C 1 2 Q	1/44 (2006.01)	C 1 2 Q	1/44	4 B 0 6 3	
C 1 2 Q	1/6806(2018.01)	C 1 2 Q	1/6806	Z	
C 1 2 Q	1/6851(2018.01)	C 1 2 Q	1/6851	Z	
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z	
C 1 2 N	9/16 (2006.01)	C 1 2 N	9/16	Z	
		審査請求	未請求	予備審査請求	未請求 (全106頁) 最終頁に続く
(21)出願番号	特願2023-519658(P2023-519658)	(71)出願人	515059083		
(86)(22)出願日	令和3年9月29日(2021.9.29)		ガーダント ヘルス, インコーポレイテッド		
(85)翻訳文提出日	令和5年5月25日(2023.5.25)		アメリカ合衆国 カリフォルニア 9 4 0		
(86)国際出願番号	PCT/US2021/071649		6 3, レッドウッド シティ, ペノブ		
(87)国際公開番号	WO2022/073012		スコット ドライブ 5 0 5		
(87)国際公開日	令和4年4月7日(2022.4.7)	(74)代理人	100078282		
(31)優先権主張番号	63/086,000		弁理士 山本 秀策		
(32)優先日	令和2年9月30日(2020.9.30)	(74)代理人	100113413		
(33)優先権主張国・地域又は機関	米国(US)		弁理士 森下 夏樹		
(31)優先権主張番号	63/105,183	(74)代理人	100181674		
(32)優先日	令和2年10月23日(2020.10.23)		弁理士 飯田 貴敏		
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100181641		
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA	(74)代理人	弁理士 石川 大輔		
			230113332	最終頁に続く	

(54)【発明の名称】 分配およびメチル化依存性ヌクレアーゼを使用してDNAを解析するための組成物および方法

(57)【要約】

本開示は、DNA、例えば、無細胞DNAを解析することに関連する組成物および方法を提供する。一部の実施形態では、無細胞DNAは、がんを有するかもしれないががんを有することが疑われる対象に由来する、かつ／または無細胞DNAは、がん細胞に由来するDNAを含む。一部の実施形態では、DNAは、第1の部分試料および第2の部分試料に分配され、ここで、第1の部分試料は、第2の部分試料よりも多い割合でヌクレオチド改変（例えば、シトシン改変）を有するDNAを含み、第2の部分試料は、メチル化依存性ヌクレアーゼと接触させられる。

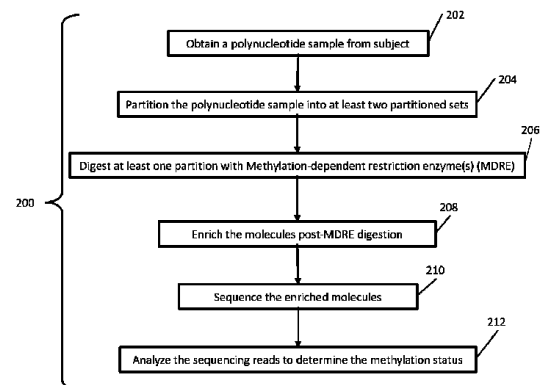


Fig. 2

【特許請求の範囲】**【請求項 1】**

試料中の DNA を解析する方法であって、

- a) 前記試料を、第 1 の部分試料および第 2 の部分試料を含む複数の部分試料に分配するステップであって、前記第 1 の部分試料が、前記第 2 の部分試料よりも多い割合で、シトシン改変を有する DNA を含む、ステップと、
 - b) 前記第 2 の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、前記第 2 の部分試料中の非特異的に分配された DNA を分解して、処置された第 2 の部分試料を産生するステップ、必要に応じて、前記第 1 の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、前記第 1 の部分試料中の非特異的に分配された DNA を分解して、処置された第 1 の部分試料を産生するステップと、
 - c) 前記第 1 の部分試料または前記処置された第 1 の部分試料の少なくとも一部分から、エピジェネティック標的領域を含む第 1 の標的領域セットを捕捉し、前記処置された第 2 の部分試料の少なくとも一部分から、エピジェネティック標的領域を含む第 2 の標的領域セットを捕捉するステップと
- を含む、方法。

10

【請求項 2】

試料中の DNA を解析する方法であって、

- a) 前記試料から、エピジェネティック標的領域を含む第 1 の標的領域セットを捕捉するステップと、
 - b) 前記標的領域セットを、第 1 の部分試料および第 2 の部分試料を含む複数の部分試料に分配するステップであって、前記第 1 の部分試料が、前記第 2 の部分試料よりも多い割合で、シトシン改変を有する DNA を含む、ステップと、
 - c) 前記第 2 の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、前記第 2 の部分試料中の非特異的に分配された DNA を分解して、処置された第 2 の部分試料を産生するステップ、必要に応じて、前記第 1 の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、前記第 1 の部分試料中の非特異的に分配された DNA を分解して、処置された第 1 の部分試料を産生するステップと
- を含む、方法。

20

【請求項 3】

前記第 1 の部分試料、前記処置された第 1 の部分試料、または前記処置された第 2 の部分試料のうちの 1 つまたは複数から捕捉されたか、またはそこに存在する、エピジェネティック標的領域を定量するステップをさらに含む、請求項 1 に記載の方法。

30

【請求項 4】

前記定量するステップが、定量的増幅を含み、必要に応じて、前記定量的増幅が、定量的 PCR である、請求項 2 に記載の方法。

【請求項 5】

前記第 1 の標的領域セットおよび前記第 2 の標的領域セットまたは前記処置された第 2 の部分試料中の DNA をシーケンシングするステップをさらに含む、先行する請求項のいずれか一項に記載の方法。

40

【請求項 6】

前記処置された第 2 の部分試料中の DNA および前記処置された第 1 の部分試料中の DNA が、シーケンシングされる、先行する請求項に記載の方法。

【請求項 7】

前記エピジェネティック標的領域が、低メチル化可変標的領域セットを含む、先行する請求項のいずれか一項に記載の方法。

【請求項 8】

前記低メチル化可変標的領域セットが、少なくとも 1 つのタイプの組織において、健康な対象に由来する無細胞 DNA におけるメチル化の程度よりも低い程度のメチル化を有する領域を含む、先行する請求項に記載の方法。

50

【請求項 9】

前記低メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、がんの存在、非存在、または可能性を決定するステップをさらに含む、先行する請求項に記載の方法。

【請求項 10】

前記低メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、前記試料中の腫瘍 DNA を定量するステップをさらに含む、請求項 7～9 のいずれか一項に記載の方法。

【請求項 11】

試料中の DNA を解析する方法であって、

- a) 前記試料を、第 1 の部分試料および第 2 の部分試料を含む複数の部分試料に分配するステップであって、前記第 1 の部分試料が、前記第 2 の部分試料よりも多い割合で、シトシン改変を有する DNA を含む、ステップと、
 - b) 前記第 2 の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、前記第 2 の部分試料中の非特異的に分配された DNA を分解して、処置された第 2 の部分試料を産生するステップ、必要に応じて、前記第 1 の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、前記第 1 の部分試料中の非特異的に分配された DNA を分解して、処置された第 1 の部分試料を産生するステップと、
 - c) 前記第 1 の部分試料または前記処置された第 1 の部分試料の少なくとも一部分から、エピジェネティック標的領域を含む第 1 の標的領域セットを捕捉するステップと
- を含む、方法。

【請求項 12】

前記第 1 の部分試料もしくは前記処置された第 1 の部分試料のうちの 1 つもしくは複数から捕捉されたか、または前記処置された第 2 の部分試料中に存在する、エピジェネティック標的領域を定量するステップをさらに含む、請求項 11 に記載の方法。

【請求項 13】

前記定量するステップが、定量的増幅を含み、必要に応じて、前記定量的増幅が、定量的 PCR である、請求項 12 に記載の方法。

【請求項 14】

前記第 1 の標的領域セット中の DNA および前記第 2 の部分試料に由来する DNA をシーケンシングするステップをさらに含む、請求項 11～13 のいずれか一項に記載の方法。

【請求項 15】

前記 DNA が、体液から得られた DNA を含み、必要に応じて、前記体液が、血漿、尿、リンパ液、または脊髄液である、先行する請求項のいずれか一項に記載の方法。

【請求項 16】

前記 DNA が、試験対象から得られた無細胞 DNA (cfDNA) を含む、先行する請求項のいずれか一項に記載の方法。

【請求項 17】

前記シトシン改変が、メチル化である、先行する請求項のいずれか一項に記載の方法。

【請求項 18】

前記シトシン改変が、シトシンの 5 位におけるメチル化である、先行する請求項のいずれか一項に記載の方法。

【請求項 19】

前記第 1 の部分試料が、メチル化感受性エンドヌクレアーゼと接触させられる、先行する請求項のいずれか一項に記載の方法。

【請求項 20】

前記メチル化感受性エンドヌクレアーゼが、非メチル化 CpG 配列を切断する、先行する請求項に記載の方法。

【請求項 21】

前記メチル化感受性エンドヌクレアーゼが、AatII、AccII、AciI、Aor13HI、Aor15HI、BspT104I、BssHII、BstUI、Cfr10I、ClaI、CpoI、Eco52I、HaeII、HapII、HhaI、Hin6I、HpaII、HpyCH4IV、MluI、NaeI、NotI、NruI、NsbI、PmaCI、Psp1406I、PvuI、SacII、SalI、SmaI、およびSnaBIのうちの1つまたは複数である、先行する請求項のいずれか一項に記載の方法。

【請求項22】

前記メチル化感受性エンドヌクレアーゼが、BstUI、HpaII、Hin6I、HhaI、またはAccIIのうちの1つまたは複数であり、必要に応じて、前記メチル化感受性エンドヌクレアーゼが、(i) BstUIおよびHpaII、(ii) BstUI、HpaII、およびHin6I、または(iii) HhaIおよびAccIIである、先行する請求項に記載の方法。

10

【請求項23】

前記メチル化依存性エンドヌクレアーゼが、メチル化CpG配列を切断する、先行する請求項のいずれか一項に記載の方法。

【請求項24】

前記メチル化依存性エンドヌクレアーゼが、MspJI、LpnPI、FspEI、またはMcrBCのうちの1つまたは複数である、先行する請求項のいずれか一項に記載の方法。

20

【請求項25】

前記第1の標的領域セットが、高メチル化可変標的領域セットを含む、先行する請求項のいずれか一項に記載の方法。

【請求項26】

前記高メチル化可変標的領域セットが、少なくとも1つのタイプの組織において、健康な対象に由来する無細胞DNAにおけるメチル化の程度よりも高い程度のメチル化を有する領域を含む、先行する請求項に記載の方法。

【請求項27】

前記高メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、がんの存在、非存在、または可能性を決定するステップをさらに含む、請求項25または26に記載の方法。

30

【請求項28】

前記低メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、前記試料中の腫瘍DNAを定量するステップをさらに含む、請求項25～27のいずれか一項に記載の方法。

【請求項29】

前記エピジェネティック標的領域が、メチル化対照標的領域セットを含む、先行する請求項のいずれか一項に記載の方法。

【請求項30】

前記第1および／または第2のエピジェネティック標的領域セットが、断片化可変標的領域セットを含む、先行する請求項のいずれか一項に記載の方法。

40

【請求項31】

前記断片化可変標的領域セットが、転写開始部位領域を含む、先行する請求項に記載の方法。

【請求項32】

前記断片化可変標的領域セットが、CTCF結合領域を含む、請求項30または31に記載の方法。

【請求項33】

前記第1の標的領域セットが、配列可変標的領域をさらに含む、先行する請求項のいずれか一項に記載の方法。

50

【請求項 3 4】

前記第 2 の標的領域セットが、配列可変標的領域をさらに含む、先行する請求項のいずれか一項に記載の方法。

【請求項 3 5】

前記配列可変標的領域セットに対応する DNA 分子が、前記エピジェネティック標的領域セットに対応する DNA 分子よりも高い捕捉収率で捕捉される、請求項 3 3 または 3 4 に記載の方法。

【請求項 3 6】

捕捉するステップが、捕捉しようとする DNA を、標的特異的プローブのセットと接触させることを含み、それによって、標的特異的プローブと DNA との複合体が形成される、先行する請求項のいずれか一項に記載の方法。

10

【請求項 3 7】

捕捉するステップが、前記複合体を、標的特異的プローブに結合していない DNA から分離させ、それによって、捕捉された DNA を得ることをさらに含む、先行する請求項に記載の方法。

【請求項 3 8】

前記標的特異的プローブのセットが、前記エピジェネティック標的領域セットに対応する DNA よりも高い捕捉収率で前記配列可変標的領域セットに対応する DNA を捕捉するように構成される、請求項 3 6 または 3 7 に記載の方法。

【請求項 3 9】

前記エピジェネティック標的領域セットに対応する DNA 分子よりも大きいシーケンシング深度まで、前記配列可変標的領域セットに対応する DNA 分子をシーケンシングするステップを含む、請求項 3 3 ～ 3 8 のいずれか一項に記載の方法。

20

【請求項 4 0】

前記 DNA が、前記シーケンシングするステップの前に増幅されるか、または前記 DNA が、前記捕捉するステップの前に増幅される、先行する請求項のいずれか一項に記載の方法。

【請求項 4 1】

捕捉の前に前記 DNA にアダプターをライゲーションするステップをさらに含み、必要に応じて、前記ライゲーションするステップが、増幅の前または増幅と同時に生じる、先行する請求項のいずれか一項に記載の方法。

30

【請求項 4 2】

前記アダプターが、バーコード含有アダプターである、先行する請求項に記載の方法。

【請求項 4 3】

前記アダプターが、前記第 2 の部分試料をメチル化依存性ヌクレアーゼと接触させる前に前記 DNA にライゲーションされる、請求項 4 1 または 4 2 に記載の方法。

【請求項 4 4】

前記アダプターが、前記メチル化依存性ヌクレアーゼによる消化に対して抵抗性である、先行する請求項に記載の方法。

【請求項 4 5】

前記アダプターが、非メチル化である、先行する請求項に記載の方法。

40

【請求項 4 6】

前記試料を複数の部分試料に分配するステップが、メチル化レベルに基づいて分配することを含む、先行する請求項のいずれか一項に記載の方法。

【請求項 4 7】

前記分配するステップが、収集した cfDNA を、固体支持体に固定化されたメチル結合試薬と接触させることを含む、先行する請求項に記載の方法。

【請求項 4 8】

前記第 1 の部分試料および第 2 の部分試料、前記処置された第 1 の部分試料および前記第 2 の部分試料、前記第 1 の部分試料および前記処置された第 2 の部分試料、または前記

50

処置された第 1 の部分試料および前記処置された第 2 の部分試料を差次的にタグ付けするステップを含む、先行する請求項のいずれか一項に記載の方法。

【請求項 49】

前記処置された第 1 の部分試料および前記第 2 の部分試料、前記第 1 の部分試料および前記処置された第 2 の部分試料、または前記処置された第 1 の部分試料および前記処置された第 2 の部分試料が、前記第 1 の部分試料を前記メチル化感受性エンドヌクレアーゼと接触させた、および／または前記第 2 の部分試料をメチル化依存性ヌクレアーゼと接触させた後に、プールされる、先行する請求項に記載の方法。

【請求項 50】

前記処置された第 1 の部分試料および前記第 2 の部分試料、前記第 1 の部分試料および前記処置された第 2 の部分試料、または前記処置された第 1 の部分試料および前記処置された第 2 の部分試料に由来する DNA が、同じシーケンシングセルにおいてシーケンシングされる、請求項 47～49 のいずれか一項に記載の方法。

10

【請求項 51】

前記複数の部分試料が、前記第 2 の部分試料よりも高い割合であるが前記第 1 の部分試料よりも低い割合でシトシン改変を有する DNA を含む、第 3 の部分試料を含む、先行する請求項のいずれか一項に記載の方法。

【請求項 52】

前記第 3 の部分試料を差次的にタグ付けするステップをさらに含む、先行する請求項に記載の方法。

20

【請求項 53】

前記第 1、第 2、および第 3 の部分試料が、前記第 1 の部分試料を前記メチル化感受性エンドヌクレアーゼと接触させた、および／または前記第 2 の部分試料をメチル化依存性ヌクレアーゼと接触させた後に組み合わせられ、必要に応じて、前記第 1、第 2、および第 3 の部分試料に由来する DNA が、同じシーケンシングセルにおいてシーケンシングされる、先行する請求項に記載の方法。

【請求項 54】

前記第 3 の部分試料が、メチル化感受性エンドヌクレアーゼと接触させられる、請求項 51～53 のいずれか一項に記載の方法。

【請求項 55】

前記第 3 の部分試料が、前記第 1 の部分試料と組み合わせられ、前記組み合わせられた第 1 および第 3 の部分試料が、メチル化感受性エンドヌクレアーゼと接触させられる、請求項 52～53 のいずれか一項に記載の方法。

30

【請求項 56】

組み合わせられた第 1 および第 3 の部分試料が、前記第 1 および第 3 の部分試料を前記メチル化感受性エンドヌクレアーゼと接触させ、前記第 2 の部分試料をメチル化依存性ヌクレアーゼと接触させた後に、前記第 2 の部分試料とさらに組み合わせられ、必要に応じて、前記第 1、第 2、および第 3 の部分試料に由来する DNA が、同じシーケンシングセルにおいてシーケンシングされる、先行する請求項に記載の方法。

【請求項 57】

前記メチル化依存性ヌクレアーゼが、非特異的に分配された DNA を分解した後に、熱不活化される、先行する請求項のいずれか一項に記載の方法。

40

【請求項 58】

前記第 1 の部分試料が、前記第 1 の部分試料の前記 DNA 内の第 1 の核酸塩基に対して前記 DNA 内の第 2 の核酸塩基とは異なって影響を及ぼす手順に供され、前記第 1 の核酸塩基が、改変または非改変核酸塩基であり、前記第 2 の核酸塩基が、前記第 1 の核酸塩基とは異なる改変または非改変核酸塩基であり、前記第 1 の核酸塩基および前記第 2 の核酸塩基が、同じ塩基対形成特異性を有する、先行する請求項のいずれか一項に記載の方法。

【請求項 59】

前記第 1 の部分試料が供される前記手順が、前記第 2 の核酸塩基の塩基対形成特異性を

50

実質的に変化させることなく、前記第 1 の核酸塩基の塩基対形成特異性を変化させる、先行する請求項に記載の方法。

【請求項 6 0】

前記第 1 の核酸塩基が、改変または非改変シトシンであり、前記第 2 の核酸塩基が、改変または非改変シトシンである、請求項 5 8 または 5 9 に記載の方法。

【請求項 6 1】

前記第 1 の核酸塩基が、非改変シトシン (C) を含む、請求項 5 8 ～ 6 0 のいずれか一項に記載の方法。

【請求項 6 2】

前記第 2 の核酸塩基が、5-メチルシトシン (mC) を含む、請求項 5 8 ～ 6 1 のいずれか一項に記載の方法。

【請求項 6 3】

前記第 1 の部分試料が供される前記手順が、バイサルファイト変換を含む、請求項 5 8 ～ 6 2 のいずれか一項に記載の方法。

【請求項 6 4】

前記第 1 の核酸塩基が、mC を含む、請求項 5 8 ～ 6 0 のいずれか一項に記載の方法。

【請求項 6 5】

前記第 2 の核酸塩基が、5-ヒドロキシメチルシトシン (hmC) を含む、請求項 5 8 ～ 6 2 のいずれか一項に記載の方法。

【請求項 6 6】

前記第 1 の部分試料が供される前記手順が、5 hmC の保護を含む、請求項 6 2 に記載の方法。

【請求項 6 7】

前記第 1 の部分試料が供される前記手順が、Tet 補助バイサルファイト変換を含む、請求項 6 5 に記載の方法。

【請求項 6 8】

前記第 1 の部分試料が供される前記手順が、置換ボラン還元剤を用いた Tet 補助変換を含み、必要に応じて、前記置換ボラン還元剤が、2-ピコリンボラン、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランである、請求項 6 5 に記載の方法。

【請求項 6 9】

前記置換ボラン還元剤が、2-ピコリンボランまたはボランピリジンである、請求項 6 8 に記載の方法。

【請求項 7 0】

前記第 2 の核酸塩基が、C を含む、請求項 5 8 ～ 6 0、6 4 ～ 6 6、または 6 8 ～ 6 9 のいずれか一項に記載の方法。

【請求項 7 1】

前記第 1 の部分試料が供される前記手順が、hmC の保護、続いて置換ボラン還元剤を用いた Tet 補助変換を含み、必要に応じて、前記置換ボラン還元剤が、2-ピコリンボラン、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランである、請求項 6 4 ～ 6 6、または 7 0 のいずれか一項に記載の方法。

【請求項 7 2】

前記置換ボラン還元剤が、2-ピコリンボランまたはボランピリジンである、請求項 7 1 に記載の方法。

【請求項 7 3】

前記第 1 の部分試料が供される前記手順が、hmC の保護、続いて mC および／または C の脱アミノ化を含む、請求項 6 1、6 2、6 4 ～ 6 6、または 7 0 のいずれか一項に記載の方法。

【請求項 7 4】

前記 mC および／または C の脱アミノ化が、AID / APOBEC ファミリー DNA デ

10

20

30

40

50

アミナーゼ酵素での処置を含む、請求項 7 3 に記載の方法。

【請求項 7 5】

h m C の保護が、h m C のグルコシル化を含む、請求項 6 6 または 7 0 ～ 7 4 のいずれか一項に記載の方法。

【請求項 7 6】

前記第 1 の部分試料が供される前記手順が、置換ボラン還元剤を用いた化学物質補助変換を含み、必要に応じて、前記置換ボラン還元剤が、2-ピコリンボラン、ボランピリジン、ter t-ブチルアミンボラン、またはアンモニアボランである、請求項 5 8 ～ 6 0 、 6 2 、 6 4 、 または 7 0 のいずれか一項に記載の方法。

【請求項 7 7】

前記置換ボラン還元剤が、2-ピコリンボランまたはボランピリジンである、請求項 7 6 に記載の方法。

【請求項 7 8】

前記第 1 の核酸塩基が、h m C を含む、請求項 5 8 ～ 6 0 、 6 2 、 6 4 、 7 0 、 または 7 6 ～ 7 7 のいずれか一項に記載の方法。

【請求項 7 9】

前記第 1 の部分試料の D N A および前記第 2 の部分試料の D N A が、差次的にタグ付けされ、差次的タグ付けの後に、前記第 2 の部分試料に由来する D N A の一部分が、前記第 1 の部分試料または処置された第 1 の部分試料またはその少なくとも一部分に添加され、それによってプールが形成され、配列可変標的領域およびエピジェネティック標的領域が、前記プールから捕捉される、先行する請求項のいずれか一項に記載の方法。

【請求項 8 0】

前記プールが、前記第 2 の部分試料の D N A のうちの約 4 5 % 未満もしくはそれに等しい、4 0 % 未満もしくはそれに等しい、3 5 % 未満もしくはそれに等しい、3 0 % 未満もしくはそれに等しい、2 5 % 未満もしくはそれに等しい、2 0 % 未満もしくはそれに等しい、1 5 % 未満もしくはそれに等しい、1 0 % 未満もしくはそれに等しい、または 5 % 未満もしくはそれに等しいを含む、先行する請求項に記載の方法。

【請求項 8 1】

前記プールが、前記第 2 の部分試料の D N A のうちの約 7 0 ～ 9 0 % 、 約 7 5 ～ 8 5 % 、 または 約 8 0 % を含む、先行する請求項に記載の方法。

【請求項 8 2】

前記プールが、前記第 1 の部分試料の D N A の実質的に全てを含む、請求項 7 9 ～ 8 1 のいずれか一項に記載の方法。

【請求項 8 3】

前記プールが、前記第 1 の部分試料または処置された第 1 の部分試料の D N A の実質的に全てを含む、請求項 7 9 ～ 8 2 のいずれか一項に記載の方法。

【請求項 8 4】

前記第 1 の標的領域セットが、前記プールの形成後に、前記第 1 の部分試料または処置された第 1 の部分試料の少なくとも一部分から捕捉される、請求項 7 9 ～ 8 3 のいずれか一項に記載の方法。

【請求項 8 5】

前記対象ががんを有する可能性を決定するステップをさらに含む、先行する請求項のいずれか一項に記載の方法。

【請求項 8 6】

前記シーケンシングするステップが、複数のシーケンシング読み取りデータを生成し、前記方法が、前記複数の配列読み取りデータを 1 つまたは複数の参照配列にマッピングして、マッピングされた配列読み取りデータを生成するステップ、ならびに前記配列可変標的領域セットおよび前記エピジェネティック標的領域セットに対応する前記マッピングされた配列読み取りデータをプロセッシングして、前記対象ががんを有する可能性を決定するステップをさらに含む、先行する請求項に記載の方法。

10

20

30

40

50

【請求項 87】

前記試験対象が、以前にがんと診断され、1つまたは複数の以前のがん処置を受けており、必要に応じて、前記 cfDNA が、前記 1つまたは複数の以前のがん処置の後の 1つまたは複数の事前に選択された時点で取得され、前記 cfDNA 分子の捕捉されたセットをシーケンシングし、それによって、配列情報のセットが産生される、請求項 1～85 のいずれか一項に記載の方法。

【請求項 88】

前記配列情報のセットを使用して、事前に選択された時点において、腫瘍細胞を起源とするかまたはそれに由来する DNA の存在または非存在を検出するステップをさらに含む、先行する請求項に記載の方法。

10

【請求項 89】

前記試験対象について、前記腫瘍細胞を起源とするかまたはそれに由来する前記 DNA の存在または非存在を示すがん再発スコアを決定するステップをさらに含み、必要に応じて、前記がん再発スコアに基づいて、がん再発ステータスを決定するステップをさらに含み、前記試験対象の前記がん再発ステータスが、がん再発スコアが所定の閾値であるかもしくはそれを上回ると決定された場合にがん再発のリスクがあると決定されるか、または前記試験対象の前記がん再発ステータスが、前記がん再発スコアが前記所定の閾値よりも低い場合にがん再発のリスクが低いと決定される、先行する請求項に記載の方法。

【請求項 90】

前記試験対象の前記がん再発スコアを所定のがん再発閾値と比較するステップをさらに含み、前記試験対象が、前記がん再発スコアが前記がん再発閾値を上回る場合に後続のがん処置の候補として分類されるか、または前記がん再発スコアが前記がん再発閾値よりも低い場合に後続のがん処置の候補ではないとして分類される、先行する請求項に記載の方法。

20

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

本出願は、その各々の全体が全ての目的に関して参照により本明細書に組み込まれる、2020年9月30日に出願された米国仮特許出願第 63 / 086, 000 号、および 2020年10月23日に出願された米国仮特許出願第 63 / 105, 183 号に基づく優先権の利益を主張する。

30

【0002】

発明の分野

本開示は、DNA、例えば、無細胞 DNA を解析することに関連する組成物および方法を提供する。一部の実施形態では、無細胞 DNA は、がんを有するもしくはがんを有することが疑われる対象に由来する、かつ／または無細胞 DNA は、がん細胞に由来する DNA を含む。一部の実施形態では、DNA は、第 1 の部分試料および第 2 の部分試料に分配され、ここで、第 1 の部分試料は、第 2 の部分試料よりも多い割合でヌクレオチド改変（例えば、シトシン改変）を有する DNA を含み、第 2 の部分試料は、メチル化依存性ヌクレアーゼと接触させられる。

40

【背景技術】

【0003】

序説および概要

がんは、世界中で年間数百万人もの死亡の原因となっている。がんの早期検出は、初期のがんが処置に対してより感受性となる傾向があるため、転帰の改善をもたらす可能性がある。

【0004】

不適切に制御された細胞成長は、遺伝子変化およびエピジェネティック変化、例えば、コピー数変異（CNV）、単一ヌクレオチド変異（SNV）、遺伝子融合、挿入、および

50

／または欠失（インデル）、シトシンの改変を含むエピジェネティック変異（例えば、5-メチルシトシン、5-ヒドロキシメチルシトシン、および他のより酸化した形態）、ならびにDNAのクロマチンタンパク質および転写因子との会合の蓄積により一般的に生じる、がんの顕著な特徴である。

【0005】

生検は、可能性のあるがんの部位から細胞または組織を抽出し、関連する表現型および／または遺伝子型特色について解析する、がんを検出または診断するための従来のアプローチを表す。生検は、侵襲的であるという欠点を有する。

【0006】

体液（「液体生検」）、例えば、血液の解析に基づくがんの検出は、がん細胞に由来するDNAが体液中に放出されるという観察に基づく興味深い代替法である。液体生検は、非侵襲的である（採血だけは必要な場合がある）。無細胞核酸（例えば、無細胞DNAまたは無細胞RNA）のがん診断アッセイの現在の方法は、単一ヌクレオチド変異体（SNV）、コピー数変異（CNV）、融合、およびインデル（すなわち、挿入または欠失）を含む、腫瘍関連の体細胞変異体の検出に焦点を当てている可能性があり、これらは全て、液体生検の主流の標的である。無細胞DNAにおけるメチル化状態およびフラグメントームシグナルなどの非配列改変が、無細胞DNAの起源および疾患レベルに関する情報を提供し得るという根拠が増えてきている。無細胞DNAの非配列改変は、体細胞突然変異のコーリングと組み合わせた場合に、いずれかのアプローチ単独で入手可能なものよりも包括的な腫瘍状態評価を得ることができる。しかしながら、無細胞DNAの低い濃度および不均一性を考えると、核酸塩基改変に関する詳細な情報を提供する液体生検材料を解析するための正確かつ感度の高い方法を開発することは、困難となっている。

【0007】

液体生検手順におけるさらなる解析に有用な無細胞DNAの分画を単離しプロセッシングすることは、これらの方法の重要な部分である。したがって、例えば、液体生検において、無細胞DNAを解析するための改善された方法および組成物が、必要とされている。

【発明の概要】

【課題を解決するための手段】

【0008】

本開示は、無細胞DNAの改善された解析の必要性を満たすこと、および／または他の利点を提供することを目的としている。したがって、以下の例示的な実施形態を提供する。

【0009】

実施形態1は、試料中のDNAを解析する方法であって、

a) 試料を、第1の部分試料および第2の部分試料を含む複数の部分試料に分配するステップであって、第1の部分試料が、第2の部分試料よりも多い割合で、シトシン改変を有するDNAを含む、ステップと、

b) 第2の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、第2の部分試料中の非特異的に分配されたDNAを分解して、処置された第2の部分試料を産生するステップ、必要に応じて、第1の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、第1の部分試料中の非特異的に分配されたDNAを分解して、処置された第1の部分試料を産生するステップと、

c) 第1の部分試料または処置された第1の部分試料の少なくとも一部分から、エピジェネティック標的領域を含む第1の標的領域セットを捕捉し、処置された第2の部分試料の少なくとも一部分から、エピジェネティック標的領域を含む第2の標的領域セットを捕捉するステップと

を含む、方法である。

【0010】

実施形態2は、試料中のDNAを解析する方法であって、

a) 試料から、エピジェネティック標的領域を含む第1の標的領域セットを捕捉するステ

ップと、

b) 標的領域セットを、第1の部分試料および第2の部分試料を含む複数の部分試料に分配するステップであって、第1の部分試料が、第2の部分試料よりも多い割合で、シトシン改変を有するDNAを含む、ステップと、

c) 第2の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、第2の部分試料中の非特異的に分配されたDNAを分解して、処置された第2の部分試料を産生するステップ、必要に応じて、第1の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、第1の部分試料中の非特異的に分配されたDNAを分解して、処置された第1の部分試料を産生するステップと

を含む、方法である。

10

【0011】

実施形態3は、第1の部分試料、処置された第1の部分試料、または処置された第2の部分試料のうちの1つもしくは複数から捕捉されたか、またはそこに存在する、エピジェネティック標的領域を定量するステップをさらに含む、実施形態1に記載の方法である。

【0012】

実施形態4は、定量するステップが、定量的増幅を含み、必要に応じて、定量的増幅が、定量的PCRである、実施形態2に記載の方法である。

【0013】

実施形態5は、第1の標的領域セットおよび第2の標的領域セットまたは処置された第2の部分試料中のDNAをシーケンシングするステップをさらに含む、先行する実施形態のいずれか1つに記載の方法である。

20

【0014】

実施形態6は、処置された第2の部分試料中のDNAおよび処置された第1の部分試料中のDNAが、シーケンシングされる、先行する実施形態に記載の方法である。

【0015】

実施形態7は、エピジェネティック標的領域が、低メチル化可変標的領域セットを含む、先行する実施形態のいずれか1つに記載の方法である。

【0016】

実施形態8は、低メチル化可変標的領域セットが、少なくとも1つのタイプの組織において、健康な対象に由来する無細胞DNAにおけるメチル化の程度よりも低い程度のメチル化を有する領域を含む、先行する実施形態に記載の方法である。

30

【0017】

実施形態9は、低メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、がんの存在、非存在、または可能性を決定するステップをさらに含む、先行する実施形態に記載の方法である。

【0018】

実施形態10は、低メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、試料中の腫瘍DNAを定量するステップをさらに含む、実施形態7~9のいずれか1つに記載の方法である。

【0019】

実施形態11は、試料中のDNAを解析する方法であって、

a) 試料を、第1の部分試料および第2の部分試料を含む複数の部分試料に分配するステップであって、第1の部分試料が、第2の部分試料よりも多い割合で、シトシン改変を有するDNAを含む、ステップと、

b) 第2の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、第2の部分試料中の非特異的に分配されたDNAを分解して、処置された第2の部分試料を産生するステップ、必要に応じて、第1の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、第1の部分試料中の非特異的に分配されたDNAを分解して、処置された第1の部分試料を産生するステップと、

c) 第1の部分試料または処置された第1の部分試料の少なくとも一部分から、エピジェ

50

ネティック標的領域を含む第1の標的領域セットを捕捉するステップとを含む、方法である。

【0020】

実施形態12は、第1の部分試料もしくは処置された第1の部分試料のうちの1つもしくは複数から捕捉されたか、または処置された第2の部分試料中に存在する、エピジェネティック標的領域を定量するステップをさらに含む、実施形態11に記載の方法である。

【0021】

実施形態13は、定量するステップが、定量的増幅を含み、必要に応じて、定量的増幅が、定量的PCRである、実施形態12に記載の方法である。

【0022】

実施形態14は、第1の標的領域セット中のDNAおよび第2の部分試料に由来するDNAをシーケンシングするステップをさらに含む、実施形態11～13のいずれか1つに記載の方法である。

【0023】

実施形態15は、DNAが、体液から得られたDNAを含み、必要に応じて、体液が、血漿、尿、リンパ液、または脊髄液である、先行する実施形態のいずれか1つに記載の方法である。

【0024】

実施形態16は、DNAが、試験対象から得られた無細胞DNA（cfDNA）を含む、先行する実施形態のいずれか1つに記載の方法である。

【0025】

実施形態17は、シトシン改変が、メチル化である、先行する実施形態のいずれか1つに記載の方法である。

【0026】

実施形態18は、シトシン改変が、シトシンの5位におけるメチル化である、先行する実施形態のいずれか1つに記載の方法である。

【0027】

実施形態19は、第1の部分試料が、メチル化感受性エンドヌクレアーゼと接触させられる、先行する実施形態のいずれか1つに記載の方法である。

【0028】

実施形態20は、メチル化感受性エンドヌクレアーゼが、非メチル化CpG配列を切断する、先行する実施形態に記載の方法である。

【0029】

実施形態21は、メチル化感受性エンドヌクレアーゼが、AatII、AccII、AciI、Aor13HI、Aor15HI、BspT104I、BssHII、BstUI、Cfr10I、ClaI、CpoI、Eco52I、HaeII、HapII、HhaI、Hin6I、HpaII、HpyCH4IV、MluI、NaeI、NotI、NruI、NsbI、PmaCI、Psp1406I、PvuI、SacII、SalI、SmaI、およびSnaBIのうちの1つまたは複数である、先行する実施形態のいずれか1つに記載の方法である。

【0030】

実施形態22は、メチル化感受性エンドヌクレアーゼが、BstUI、HpaII、Hin6I、HhaI、またはAccIIのうちの1つまたは複数であり、必要に応じて、メチル化感受性エンドヌクレアーゼが、(i) BstUIおよびHpaII、(ii) BstUI、HpaII、およびHin6I、または(iii) HhaIおよびAccIIである、先行する実施形態に記載の方法である。

【0031】

実施形態23は、メチル化依存性エンドヌクレアーゼが、メチル化CpG配列を切断する、先行する実施形態のいずれか1つに記載の方法である。

【0032】

10

20

30

40

50

実施形態 24 は、メチル化依存性エンドヌクレアーゼが、M s p J I、L p n P I、F s p E I、または M c r B C のうちの 1 つまたは複数である、先行する実施形態のいずれか 1 つに記載の方法である。

【0033】

実施形態 25 は、第 1 の標的領域セットが、高メチル化可変標的領域セットを含む、先行する実施形態のいずれか 1 つに記載の方法である。

【0034】

実施形態 26 は、高メチル化可変標的領域セットが、少なくとも 1 つのタイプの組織において、健康な対象に由来する無細胞 DNA におけるメチル化の程度よりも高い程度のメチル化を有する領域を含む、先行する実施形態に記載の方法である。

10

【0035】

実施形態 27 は、高メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、がんの存在、非存在、または可能性を決定するステップをさらに含む、実施形態 25 または 26 に記載の方法である。

【0036】

実施形態 28 は、低メチル化可変標的領域セット中の領域の配列または数量に少なくとも部分的に基づいて、試料中の腫瘍 DNA を定量するステップをさらに含む、実施形態 25 ～ 27 のいずれか 1 つに記載の方法である。

【0037】

実施形態 29 は、エピジェネティック標的領域が、メチル化対照標的領域セットを含む、先行する実施形態のいずれか 1 つに記載の方法である。

20

【0038】

実施形態 30 は、第 1 および／または第 2 のエピジェネティック標的領域セットが、断片化可変標的領域セットを含む、先行する実施形態のいずれか 1 つに記載の方法である。

【0039】

実施形態 31 は、断片化可変標的領域セットが、転写開始部位領域を含む、先行する実施形態に記載の方法である。

【0040】

実施形態 32 は、断片化可変標的領域セットが、C T C F 結合領域を含む、実施形態 30 または 31 に記載の方法である。

30

【0041】

実施形態 33 は、第 1 の標的領域セットが、配列可変標的領域をさらに含む、先行する実施形態のいずれか 1 つに記載の方法である。

【0042】

実施形態 34 は、第 2 の標的領域セットが、配列可変標的領域をさらに含む、先行する実施形態のいずれか 1 つに記載の方法である。

【0043】

実施形態 35 は、配列可変標的領域セットに対応する DNA 分子が、エピジェネティック標的領域セットに対応する DNA 分子よりも高い捕捉収率で捕捉される、実施形態 33 または 34 に記載の方法である。

40

【0044】

実施形態 36 は、捕捉するステップが、捕捉しようとする DNA を、標的特異的プローブのセットと接触させることを含み、それによって、標的特異的プローブと DNA との複合体が形成される、先行する実施形態のいずれか 1 つに記載の方法である。

【0045】

実施形態 37 は、捕捉するステップが、複合体を、標的特異的プローブに結合していない DNA から分離させ、それによって、捕捉された DNA を得ることをさらに含む、先行する実施形態に記載の方法である。

【0046】

実施形態 38 は、標的特異的プローブのセットが、エピジェネティック標的領域セット

50

に対応するDNAよりも高い捕収率で配列可変標的領域セットに対応するDNAを捕捉するように構成される、実施形態36または37に記載の方法である。

【0047】

実施形態39は、エピジェネティック標的領域セットに対応するDNA分子よりも大きいシーケンシング深度まで、配列可変標的領域セットに対応するDNA分子をシーケンシングするステップを含む、実施形態33～38のいずれか1つに記載の方法である。

【0048】

実施形態40は、DNAが、シーケンシングするステップの前に増幅されるか、またはDNAが、捕捉ステップの前に増幅される、先行する実施形態のいずれか1つに記載の方法である。

【0049】

実施形態41は、捕捉の前にDNAにアダプターをライゲーションするステップをさらに含み、必要に応じて、ライゲーションするステップが、増幅の前または増幅と同時に生じる、先行する実施形態のいずれか1つに記載の方法である。

【0050】

実施形態42は、アダプターが、バーコード含有アダプターである、先行する実施形態に記載の方法である。

【0051】

実施形態43は、アダプターが、第2の部分試料をメチル化依存性ヌクレアーゼと接触させる前にDNAにライゲーションされる、実施形態41または42に記載の方法である。

【0052】

実施形態44は、アダプターが、メチル化依存性ヌクレアーゼによる消化に対して抵抗性である、先行する実施形態に記載の方法である。

【0053】

実施形態45は、アダプターが、非メチル化である、先行する実施形態に記載の方法である。

【0054】

実施形態46は、試料を複数の部分試料に分配するステップが、メチル化レベルに基づいて分配することを含む、先行する実施形態のいずれか1つに記載の方法である。

【0055】

実施形態47は、分配するステップが、収集したcfDNAを、固体支持体に固定化されたメチル結合試薬と接触させることを含む、先行する実施形態に記載の方法。

【0056】

実施形態48は、第1の部分試料および第2の部分試料、処置された第1の部分試料および第2の部分試料、第1の部分試料および処置された第2の部分試料、または処置された第1の部分試料および処置された第2の部分試料を差次的にタグ付けするステップを含む、先行する実施形態のいずれか1つに記載の方法である。

【0057】

実施形態49は、処置された第1の部分試料および第2の部分試料、第1の部分試料および処置された第2の部分試料、または処置された第1の部分試料および処置された第2の部分試料が、第1の部分試料をメチル化感受性エンドヌクレアーゼと接触させた、および／または第2の部分試料をメチル化依存性ヌクレアーゼと接触させた後に、プールされる、先行する実施形態に記載の方法である。

【0058】

実施形態50は、処置された第1の部分試料および第2の部分試料、第1の部分試料および処置された第2の部分試料、または処置された第1の部分試料および処置された第2の部分試料に由来するDNAが、同じシーケンシングセルにおいてシーケンシングされる、実施形態47～49のいずれか1つに記載の方法である。

【0059】

10

20

30

40

50

実施形態 5 1 は、複数の部分試料が、第 2 部分試料よりも高い割合であるが第 1 の部分試料よりも低い割合でシトシン改変を有する DNA を含む、第 3 の部分試料を含む、先行する実施形態のいずれか 1 つに記載の方法である。

【0060】

実施形態 5 2 は、第 3 の部分試料を差次的にタグ付けするステップをさらに含む、先行する実施形態に記載の方法である。

【0061】

実施形態 5 3 は、第 1、第 2、および第 3 の部分試料が、第 1 の部分試料をメチル化感受性エンドヌクレアーゼと接触させた、および／または第 2 の部分試料をメチル化依存性ヌクレアーゼと接触させた後に組み合わせられ、必要に応じて、第 1、第 2、および第 3 の部分試料に由来する DNA が、同じシーケンシングセルにおいてシーケンシングされる、先行する実施形態に記載の方法である。

10

【0062】

実施形態 5 4 は、第 3 の部分試料が、メチル化感受性エンドヌクレアーゼと接触させられる、実施形態 5 1～5 3 のいずれか 1 つに記載の方法である。

【0063】

実施形態 5 5 は、第 3 の部分試料が、第 1 の部分試料と組み合わせられ、組み合わせられた第 1 および第 3 の部分試料が、メチル化感受性エンドヌクレアーゼと接触させられる、実施形態 5 2～5 3 のいずれか 1 つに記載の方法である。

【0064】

実施形態 5 6 は、組み合わせられた第 1 および第 3 の部分試料が、第 1 および第 3 の部分試料をメチル化感受性エンドヌクレアーゼと接触させ、第 2 の部分試料をメチル化依存性ヌクレアーゼと接触させた後に、第 2 の部分試料とさらに組み合わせられ、必要に応じて、第 1、第 2、および第 3 の部分試料に由来する DNA が、同じシーケンシングセルにおいてシーケンシングされる、先行する実施形態に記載の方法である。

20

【0065】

実施形態 5 7 は、メチル化依存性ヌクレアーゼが、非特異的に分配された DNA を分解した後に、熱不活化される、先行する実施形態のいずれか 1 つに記載の方法である。

【0066】

実施形態 5 8 は、第 1 の部分試料が、第 1 の部分試料の DNA 内の第 1 の核酸塩基に対して DNA 内の第 2 の核酸塩基とは異なって影響を及ぼす手順に供され、第 1 の核酸塩基が、改変または非改変核酸塩基であり、第 2 の核酸塩基が、第 1 の核酸塩基とは異なる改変または非改変核酸塩基であり、第 1 の核酸塩基および第 2 の核酸塩基が、同じ塩基対形成特異性を有する、先行する実施形態のいずれか 1 つに記載の方法である。

30

【0067】

実施形態 5 9 は、第 1 の部分試料が供される手順が、第 2 の核酸塩基の塩基対形成特異性を実質的に変化させることなく、第 1 の核酸塩基の塩基対形成特異性を変化させる、先行する実施形態に記載の方法である。

【0068】

実施形態 6 0 は、第 1 の核酸塩基が、改変または非改変シトシンであり、第 2 の核酸塩基が、改変または非改変シトシンである、実施形態 5 8 または 5 9 に記載の方法である。

40

【0069】

実施形態 6 1 は、第 1 の核酸塩基が、非改変シトシン (C) を含む、実施形態 5 8～6 0 のいずれか 1 つに記載の方法である。

【0070】

実施形態 6 2 は、第 2 の核酸塩基が、5-メチルシトシン (mC) を含む、実施形態 5 8～6 1 のいずれか 1 つに記載の方法である。

【0071】

実施形態 6 3 は、第 1 の部分試料が供される手順が、バイサルファイト変換を含む、実施形態 5 8～6 2 のいずれか 1 つに記載の方法である。

50

【0072】

実施形態64は、第1の核酸塩基が、mCを含む、実施形態58～60のいずれか1つに記載の方法である。

【0073】

実施形態65は、第2の核酸塩基が、5-ヒドロキシメチルシトシン(hmC)を含む、実施形態58～62のいずれか1つに記載の方法である。

【0074】

実施形態66は、第1の部分試料が供される手順が、5hmCの保護を含む、実施形態62に記載の方法である。

【0075】

実施形態67は、第1の部分試料が供される手順が、Tet補助バイサルファイト変換を含む、実施形態65に記載の方法である。

【0076】

実施形態68は、第1の部分試料が供される手順が、置換ボラン還元剤を用いたTet補助変換を含み、必要に応じて、置換ボラン還元剤が、2-ピコリンボラン、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランである、実施形態65に記載の方法である。

【0077】

実施形態69は、置換ボラン還元剤が、2-ピコリンボランまたはボランピリジンである、実施形態68に記載の方法である。

【0078】

実施形態70は、第2の核酸塩基が、Cを含む、実施形態58～60、64～66、または68～69のいずれか1つに記載の方法である。

【0079】

実施形態71は、第1の部分試料が供される手順が、hmCの保護、続いて置換ボラン還元剤を用いたTet補助変換を含み、必要に応じて、置換ボラン還元剤が、2-ピコリンボラン、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランである、実施形態64～66、または70のいずれか1つに記載の方法である。

【0080】

実施形態72は、置換ボラン還元剤が、2-ピコリンボランまたはボランピリジンである、実施形態71に記載の方法である。

【0081】

実施形態73は、第1の部分試料が供される手順が、hmCの保護、続いてmCおよび／またはCの脱アミノ化を含む、実施形態61、62、64～66、または70のいずれか1つに記載の方法である。

【0082】

実施形態74は、mCおよび／またはCの脱アミノ化が、AID/APOBECファミリーDNAデアミナーゼ酵素での処置を含む、実施形態73に記載の方法である。

【0083】

実施形態75は、hmCの保護が、hmCのグルコシル化を含む、実施形態66または70～74のいずれか1つに記載の方法である。

【0084】

実施形態76は、第1の部分試料が供される手順が、置換ボラン還元剤を用いた化学物質補助変換を含み、必要に応じて、置換ボラン還元剤が、2-ピコリンボラン、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランである、実施形態58～60、62、64、または70のいずれか1つに記載の方法である。

【0085】

実施形態77は、置換ボラン還元剤が、2-ピコリンボランまたはボランピリジンである、実施形態76に記載の方法である。

【0086】

10

20

30

40

50

実施形態 78 は、第 1 の核酸塩基が、h m C を含む、実施形態 58 ～ 60、62、64、70、または 76 ～ 77 のいずれか 1 つに記載の方法である。

【0087】

実施形態 79 は、第 1 の部分試料の DNA および第 2 の部分試料の DNA が、差次的にタグ付けされ、差次的タグ付けの後に、第 2 の部分試料に由来する DNA の一部分が、第 1 の部分試料または処置された第 1 の部分試料またはその少なくとも一部分に添加され、それによってプールが形成され、配列可変標的領域およびエピジェネティック標的領域が、プールから捕捉される、先行する実施形態のいずれか 1 つに記載の方法である。

【0088】

実施形態 80 は、プールが、第 2 の部分試料の DNA のうちの約 45 % 未満もしくはそれに等しい、40 % 未満もしくはそれに等しい、35 % 未満もしくはそれに等しい、30 % 未満もしくはそれに等しい、25 % 未満もしくはそれに等しい、20 % 未満もしくはそれに等しい、15 % 未満もしくはそれに等しい、10 % 未満もしくはそれに等しい、または 5 % 未満もしくはそれに等しいを含む、先行する実施形態に記載の方法である。

【0089】

実施形態 81 は、プールが、第 2 の部分試料の DNA のうちの約 70 ～ 90 %、約 75 ～ 85 %、または約 80 % を含む、先行する実施形態に記載の方法である。

【0090】

実施形態 82 は、プールが、第 1 の部分試料の DNA の実質的に全てを含む、実施形態 79 ～ 81 のいずれか 1 つに記載の方法である。

【0091】

実施形態 83 は、プールが、第 1 の部分試料または処置された第 1 の部分試料の DNA の実質的に全てを含む、実施形態 79 ～ 82 のいずれか 1 つに記載の方法である。

【0092】

実施形態 84 は、第 1 の標的領域セットが、プールの形成後に、第 1 の部分試料または処置された第 1 の部分試料の少なくとも一部分から捕捉される、実施形態 79 ～ 83 のいずれか 1 つに記載の方法である。

【0093】

実施形態 85 は、対象ががんを有する可能性を決定するステップをさらに含む、先行する実施形態のいずれか 1 つに記載の方法である。

【0094】

実施形態 86 は、シーケンシングするステップが、複数のシーケンシング読み取りデータを生成し、方法が、複数の配列読み取りデータを 1 つまたは複数の参照配列にマッピングして、マッピングされた配列読み取りデータを生成するステップと、配列可変標的領域セットおよびエピジェネティック標的領域セットに対応するマッピングされた配列読み取りデータをプロセッシングして、対象ががんを有する可能性を決定するステップとをさらに含む、先行する実施形態に記載の方法である。

【0095】

実施形態 87 は、試験対象が、以前にがんと診断され、1 つまたは複数の以前のがん処置を受けており、必要に応じて、c f DNA が、1 つまたは複数の以前のがん処置の後の 1 つまたは複数の事前を選択された時点で取得され、c f DNA 分子の捕捉されたセットをシーケンシングし、それによって、配列情報のセットが産生される、実施形態 1 ～ 85 のいずれか 1 つに記載の方法である。

【0096】

実施形態 88 は、配列情報のセットを使用して、事前を選択された時点において、腫瘍細胞を起源とするかまたはそれに由来する DNA の存在または非存在を検出するステップをさらに含む、先行する実施形態に記載の方法である。

【0097】

実施形態 89 は、試験対象について、腫瘍細胞を起源とするかまたはそれに由来する DNA の存在または非存在を示すがん再発スコアを決定するステップをさらに含み、必要に

10

20

30

40

50

応じて、がん再発スコアに基づいて、がん再発ステータスを決定するステップをさらに含み、試験対象のがん再発ステータスが、がん再発スコアが所定の閾値であるかもしくはそれを上回ると決定された場合にがん再発のリスクがあると決定されるか、または試験対象のがん再発ステータスが、がん再発スコアが所定の閾値よりも低い場合にがん再発のリスクが低いと決定される、先行する実施形態に記載の方法である。

【0098】

実施形態90は、試験対象のがん再発スコアを所定のがん再発閾値と比較するステップをさらに含み、試験対象が、がん再発スコアががん再発閾値を上回る場合に後続のがん処置の候補として分類されるか、またはがん再発スコアががん再発閾値よりも低い場合に後続のがん処置の候補ではないとして分類される、先行する実施形態に記載の方法である。

10

I. 図面の簡単な説明

【図面の簡単な説明】

【0099】

【図1】図1Aは、制限酵素（RE）認識部位がメチル化ヌクレオチドを含有する場合にDNAを消化／切断するが、制限酵素（RE）認識部位が非メチル化ヌクレオチドを含有する場合にはDNAを切断しない、メチル化依存性ヌクレアーゼ（例えば、メチル化依存性制限酵素（MDRE））を示す概略図である。図1Bは、制限酵素（RE）認識部位が非メチル化ヌクレオチドを含有する場合にDNAを消化／切断するが、制限酵素（RE）認識部位がメチル化ヌクレオチドを含有する場合にはDNAを切断しない、メチル化感受性ヌクレアーゼ（例えば、メチル化感受性制限酵素（MSRE））の概略図である。

20

【0100】

【図2】図2は、本開示の実施形態による、対象から得られたポリヌクレオチド試料中の核酸分子のメチル化状態を決定するための方法のフローチャート図である。

【0101】

【図3】図3は、本開示の実施形態による、対象から得られたポリヌクレオチド試料中の核酸分子のメチル化状態を決定するための方法のフローチャート図である。

【0102】

【図4-1】図4は、本開示のある特定の実施形態に従う、対象におけるがんの存在または非存在を検出するための方法の概略図である。

【図4-2】図4は、本開示のある特定の実施形態に従う、対象におけるがんの存在または非存在を検出するための方法の概略図である。

30

【0103】

【図5】本開示の一部の実施形態による使用にとって好適なシステムの例の概略図。

【0104】

【図6】図6は、正常試料および希釈したCRC試料におけるMSRE処置ありおよびなしでの3つの分配中の分子数を示す。

【0105】

【図7】図7は、初期結腸直腸がんを有する対象からの3つの試料（「初期CRC」）および3人の健康な対象（「正常」）に関する、実施例2に記載したように得られたCpGメチル化定量結果を示す。初期CRCプロットに関して、MAFは、変異型アレル分画を示す。

40

【0106】

【図8-1】図8A～Dは、実施例4に記載したように、表記の酵素および緩衝液条件に関するFspEIパリンドローム部位を有する陽性および陰性対照分子の数を示す。図8Aおよび8Cは、第1のドナーに対応し、図8Bおよび8Dは第2のドナーに対応する。データ点を、読みやすいように水平軸に沿って分布させる。

【図8-2】図8A～Dは、実施例4に記載したように、表記の酵素および緩衝液条件に関するFspEIパリンドローム部位を有する陽性および陰性対照分子の数を示す。図8Aおよび8Cは、第1のドナーに対応し、図8Bおよび8Dは第2のドナーに対応する。データ点を、読みやすいように水平軸に沿って分布させる。

50

【0107】

【図9-1】図9A～Dは、実施例4に記載されるように消化効率および陽性対照分子数を示す。

【図9-2】図9A～Dは、実施例4に記載されるように消化効率および陽性対照分子数を示す。

【0108】

【図10-1】図10A～Jは、実施例5に記載されるように表記の条件に関する低メチル化可変標的領域（「低VTR」）分子数（10A～E）または低VTR／陰性対照分子比（10F～J）を示す。データ点を、読みやすいように水平軸に沿って分布させる。三角、丸、プラス記号、および四角は、正常なcfDNAの起源が、それぞれ4人の健康なドナーの1番目、2番目、3番目、または4番目の人であったことを示す。

10

【図10-2】図10A～Jは、実施例5に記載されるように表記の条件に関する低メチル化可変標的領域（「低VTR」）分子数（10A～E）または低VTR／陰性対照分子比（10F～J）を示す。データ点を、読みやすいように水平軸に沿って分布させる。三角、丸、プラス記号、および四角は、正常なcfDNAの起源が、それぞれ4人の健康なドナーの1番目、2番目、3番目、または4番目の人であったことを示す。

【発明を実施するための形態】

【0109】

II. ある特定の実施形態の詳細な説明

ここで、本発明のある特定の実施形態への詳細な参照がなされる。本発明は、そのような実施形態と関連して記載されているが、それらが、本発明をそれらの実施形態に制限することを意図するものではないことが理解される。逆に、本発明は、全ての代替形態、改変形態、および等価物を網羅することを意図し、それらは、添付の特許請求の範囲によって定義される本発明内に含まれ得る。

20

【0110】

本教示を詳細に説明する前に、本開示が、特定の組成物またはプロセスステップに限定されるものではなく、それが変動し得ることを理解されたい。本明細書および添付の特許請求の範囲において使用されるとき、単数形の「1つの(a)」、「1つの(an)」、および「その(the)」は、文脈により別途明確に示されない限り、複数形の参照物を含むことに留意すべきである。したがって、例えば、「核酸(a nucleic acid)」への言及は、複数の核酸を含み、「細胞(a cell)」への言及は、複数の細胞を含むなどである。

30

【0111】

数値範囲は、その範囲を定める数を含む。測定値および測定可能な値は、測定と関連する有効桁および誤差を考慮した、近似値であるとして理解される。また、「含む(comprise)」、「含む(comprises)」、「含むこと(comprising)」、「含有する(contain)」、「含有する(contains)」、「含有すること(containing)」、「含む(include)」、「含む(includes)」、および「含むこと(including)」の使用は、制限することを意図するものではない。前述の一般的な説明および詳細な説明は、いずれも、例示的および説明的であるにすぎず、教示を制限するものではないことを理解されたい。

40

【0112】

上記の明細書に具体的に言及されていない限り、様々な構成要素を「含む」と列挙した本明細書における実施形態はまた、列挙された構成要素「からなる」または「から本質的になる」ことも企図され、様々な構成要素「からなる」と列挙した本明細書における実施形態はまた、列挙された構成要素を「含む」またはそれ「から本質的になる」ことも企図され、様々な構成要素「から本質的になる」と列挙した本明細書における実施形態はまた、列挙された構成要素「からなる」またはそれを「含む」ことも企図される（この互換可能性は、特許請求の範囲におけるこれらの用語の使用には適用されない）。

【0113】

50

本明細書において使用される節の見出しは、構成上の目的のためのものであり、決して開示される主題を制限するものと解釈されるべきではない。参照によって組み込まれる任意の文書または他の材料が、本明細書の任意の明示的な内容と相反する場合、定義を含め、本明細書が優先される。

A. 定義

【0114】

「無細胞DNA」、「cfDNA分子」、または単純に「cfDNA」は、対象において細胞外形態で（例えば、血液、血清、血漿、または他の体液、例えば、リンパ液、脳脊髄液、尿、もしくは喀痰中に）天然に存在するDNA分子を含む。cfDNAは、大型の複合生物体、例えば、哺乳動物の細胞（単数または複数）においてもともと存在していたが、細胞から生物において見出される流体中への放出を受けており、*in vitro*細胞溶解ステップを行う必要なしに、流体の試料から得ることができる。

10

【0115】

本明細書で使用される場合、「細胞核酸」とは、核酸が後に所与の解析プロセスの一部として取り出される（例えば、細胞溶解によって）としても、少なくとも試料が対象から採取または収集された時点において、核酸が起源とする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。

【0116】

本明細書で使用される場合、改変または他の特色は、改変または他の特色を有するヌクレオチドの分画が、第1の試料または集団において、第2の集団におけるよりも高い場合に、第1の試料または核酸の集団において第2の試料または集団よりも「高い割合」で存在する。例えば、第1の試料において、ヌクレオチドのうちの10分の1がmCであり、第2の試料において、ヌクレオチドのうちの20分の1がmCである場合には、第1の試料は、第2の試料よりも高い割合で、シトシンの5-メチル化改変を含む。

20

【0117】

本明細書で使用される場合、所与の核酸塩基の「塩基対形成特異性を実質的に変化させることなく」とは、シーケンシングすることができるその核酸塩基を含む分子の大部分が、それがもともと単離された試料中にあったときのその塩基対形成特異性と比べて、第2の核酸塩基の塩基対形成特異性の変更に有さないことを意味する。一部の実施形態では、シーケンシングすることができるその核酸塩基を含む分子のうちの75%、90%、95%、または99%が、それがもともと単離された試料中にあったときのその塩基対形成特異性と比べて、第2の核酸塩基の塩基対形成特異性の変更に有さない。

30

【0118】

本明細書で使用される場合、「塩基対形成特異性」とは、所与の塩基がもっとも優先的に対を形成する標準的なDNA塩基（A、C、G、またはT）を指す。したがって、例えば、非改変シトシンおよび5-メチルシトシンは、同じ塩基対形成特異性（すなわち、Gに対する特異性）を有するが、ウラシルおよびシトシンは、ウラシルがAに対する塩基対形成特異性を有する一方でシトシンがGに対する塩基対形成特異性を有するため、異なる塩基対形成特異性を有する。ウラシルは、いずれにせよ、4つの標準的なDNA塩基の中でもAともっとも優先的に対を形成するため、ウラシルがGと不安定な対を形成する能力は、無関係である。

40

【0119】

本明細書で使用される場合、複数のメンバーを含む「組合せ」は、メンバーを含む単一の組成物、または近傍にある、例えば、より大きな容器、例えば、マルチウェルプレート、チューブラック、冷蔵庫、冷凍庫、インキュベーター、水浴、アイスバケット、機械、もしくは他の保管形態内の別個の容器もしくはコンパートメント内にある組成物のセットのいずれかを指す。

【0120】

所与の標的セットに対するプローブのコレクションの「捕捉収率」とは、典型的な条件下においてプローブのコレクションが捕捉する標的セットに対応する核酸の量（例えば、

50

別の標的セットと比べた量または絶対量)を指す。例示的な典型的な捕捉条件は、試料核酸およびプローブの、ストリンジェントなハイブリダイゼーション緩衝液を含有する小さな反応体積(約20 μL)における、65℃で10~18時間のインキュベーションである。捕捉収率は、絶対値で、または複数のプローブのコレクションについては、相対値で表され得る。複数の標的領域のセットの捕捉収率を比較する場合、それらは、標的領域セットのフットプリントサイズに関して(例えば、キロ塩基に基づいて)正規化される。したがって、例えば、第1および第2の標的領域のフットプリントサイズが、それぞれ、50 kbおよび500 kbである場合(正規化係数0.1となる)には、第1の標的領域セットに対応するDNAは、第1の標的領域セットに対応する捕捉されたDNAの体積濃度あたりの質量が、第2の標的領域セットに対応する捕捉されたDNAの体積濃度あたりの質量の0.1倍を上回る場合に、第2の標的領域セットに対応するDNAよりも高い収率で捕捉される。さらなる例として、同じフットプリントサイズを使用すると、第1の標的領域セットに対応する捕捉されたDNAが、第2の標的領域セットに対応する捕捉されたDNAの体積濃度あたりの質量の0.2倍の体積濃度あたりの質量を有する場合には、第1の標的領域セットに対応するDNAは、第2の標的領域セットに対応するDNAよりも2倍高い捕捉収率で捕捉されている。

10

【0121】

1つまたは複数の標的核酸を「捕捉する」とは、1つまたは複数の標的核酸を、非標的核酸から優先的に単離または分離することを指す。

20

【0122】

核酸の「捕捉されたセット」とは、捕捉を受けた核酸を指す。

【0123】

「標的領域セット」または「標的領域のセット」とは、捕捉の標的とされる、かつ／またはプローブのセットによって標的とされる(例えば、配列相補性を通じて)、複数のゲノム座を指す。

【0124】

「標的領域セットに対応する」とは、核酸、例えば、cfDNAが、標的領域セット中の座を起源としているか、または標的領域セットに対する1つまたは複数のプローブに特異的に結合することを意味する。

【0125】

プローブまたは他のオリゴヌクレオチドおよび標的配列の文脈における「特異的に結合する」とは、適切なハイブリダイゼーション条件下において、オリゴヌクレオチドまたはプローブが、その標的配列またはその複製物にハイブリダイズして、安定なプローブ：標的ハイブリッドを形成し、同時に、安定なプローブ：非標的ハイブリッドの形成を最小限に抑えることを意味する。したがって、プローブは、非標的配列よりも十分に高い程度で、標的配列またはその複製物にハイブリダイズして、標的配列の捕捉または検出を可能にする。適切なハイブリダイゼーション条件は、当技術分野において周知であり、配列組成物に基づいて予測され得るか、または慣例的な試験方法を使用することによって決定することができる(例えば、参照により本明細書に組み込まれる、Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)、§§ 1.90-1.91, 7.37-7.57, 9.47-9.51および11.47-11.57、特に、§§ 9.50-9.51, 11.12-11.13, 11.45-11.47および11.55-11.57を参照されたい)。

30

40

【0126】

「配列可変標的領域セット」とは、新生物細胞(例えば、腫瘍細胞およびがん細胞)において、配列の変化、例えば、ヌクレオチド置換(すなわち、単一ヌクレオチド変異)、挿入、欠失、または遺伝子融合もしくは転位を示し得る、標的領域のセットを指す。

【0127】

「エピジェネティック標的領域セット」とは、新生物細胞(例えば、腫瘍細胞およびがん細胞)において、配列とは独立した変化を示し得るか、または健康な対象に由来するc

50

f DNA と比べて、がんを有する対象に由来する cf DNA において配列とは独立した変化を示し得る、標的領域のセットを指す。配列とは独立した変化の例としては、メチル化（増加または減少）、ヌクレオソーム分布、CTCF 結合、転写開始部位、および制御性タンパク質結合領域における変化が挙げられるが、これらに限定されない。本目的のために、新生物、腫瘍、またはがんに関連する局所増幅および／または遺伝子融合を受けやすい座もまた、エピジェネティック標的領域セットに含まれ得るが、これは、シーケンシングによるコピー数変化の検出または参照ゲノム内の 1 つを上回る座にマッピングする融合配列が、例えば、局所増幅および／または遺伝子融合の検出は 1 つまたは数個の個別の位置における塩基コールの正確度に依存しないため比較的浅いシーケンシング深度で検出することができるという点において、ヌクレオチド置換、挿入、または欠失の検出よりも、上記で考察された例示的なエピジェネティック変化の検出に類似する傾向にあるためである。

10

【0128】

核酸は、腫瘍細胞を起源とする場合、「腫瘍によって産生される」または ct DNA もしくは循環腫瘍 DNA によって産生される。腫瘍細胞は、それらが、腫瘍内にとどまっているか、または腫瘍から分離されているか（例えば、転移がん細胞および循環腫瘍細胞の場合にあるような）にかかわらず、腫瘍を起源とする新生物細胞である。

【0129】

「メチル化」または「DNA メチル化」という用語は、核酸分子におけるヌクレオチド塩基へのメチル基の付加を指す。一部の実施形態では、メチル化は、CpG 部位（シトシン-リン酸-グアニン部位（すなわち、核酸配列の 5' → 3' 方向でシトシンの後にグアニンがある）におけるシトシンへのメチル基の付加を指す。一部の実施形態では、DNA メチル化は、アデニンへのメチル基の付加、例えば、N⁶-メチルアデニンにおけるものを指す。一部の実施形態では、DNA メチル化は、5-メチル化（シトシンの 6 炭素環の 5 番目の炭素の改変）である。一部の実施形態では、5-メチル化は、5-メチルシトシン（5mC）を作製するためのシトシンの 5C 位へのメチル基の付加を指す。一部の実施形態では、メチル化は、5mC の誘導体を含む。5mC の誘導体としては、5-ヒドロキシメチルシトシン（5-hmC）、5-ホルミルシトシン（5-fC）、および 5-カルボキシルシトシン（carboxylcytosine）（5-caC）が挙げられるが、これらに限定されない。一部の実施形態では、DNA メチル化は、3C メチル化（シトシンの 6 炭素環の 3 番目の炭素の改変）である。一部の実施形態では、3C メチル化は、3-メチルシトシン（3mC）を生成するためのシトシンの 3C 位へのメチル基の付加を含む。メチル化はまた、非 CpG 部位でも生じ得、例えば、メチル化は、CpA、CpT、または CpC 部位で生じ得る。DNA メチル化は、メチル化 DNA 領域の活性を変化させ得る。例えば、プロモーター領域内の DNA がメチル化された場合、遺伝子の転写が抑制され得る。DNA メチル化は、正常な発達に極めて重要であり、メチル化の異常は、エピジェネティック調節を破壊し得る。エピジェネティック調節における破壊、例えば、抑制は、疾患、例えば、がんを引き起こし得る。DNA におけるプロモーターメチル化は、がんを示し得る。

20

30

【0130】

「高メチル化」という用語は、核酸分子の集団（例えば、試料）内の他の核酸分子と比べて、増加したレベルまたは程度の核酸分子のメチル化を指す。一部の実施形態では、高メチル化 DNA は、少なくとも 1 個のメチル化された残基、少なくとも 2 個のメチル化された残基、少なくとも 3 個のメチル化された残基、少なくとも 5 個のメチル化された残基、または少なくとも 10 個のメチル化された残基を含む、DNA 分子を含み得る。

40

【0131】

「低メチル化」という用語は、核酸分子の集団（例えば、試料）内の他の核酸分子と比べて、減少したレベルまたは程度の核酸分子のメチル化を指す。一部の実施形態では、低メチル化 DNA には、非メチル化 DNA 分子が含まれる。一部の実施形態では、低メチル化 DNA は、0 個のメチル化された残基、多くとも 1 個のメチル化された残基、多くとも

50

2 個のメチル化された残基、多くとも 3 個のメチル化された残基、多くとも 4 個のメチル化された残基、または多くとも 5 個のメチル化された残基を含む、DNA 分子を含み得る。

【0132】

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

10

【0133】

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

20

【0134】

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

【0135】

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

30

【0136】

本明細書で使用される場合、「消化効率」または「切断効率」とは、制限酵素消化の効率を指す。消化効率は、制限酵素での消化の際に観察される対照分子の数、および制限酵素消化の非存在時に観察される対照分子の数に基づいて計算することができる。MSRE 消化効率は、以下によって計算することができる： $\text{効率} = 1 - (\text{陰性対照分子 [MSRE] の数} / \text{陰性対照分子 [Mock] の数})$ 。MDRE 消化効率は、以下によって計算することができる： $\text{効率} = 1 - (\text{陽性対照分子 [MDRE] の数} / \text{陽性対照分子 [Mock] の数})$ 。

40

【0137】

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

50

てメチル化されたヌクレオチドの数も指し得る。

【0138】

本明細書で使用される場合、「突然変異」は、公知の参照配列からの変異を指し、例えば単一ヌクレオチド変異体（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、生殖系列または体細胞突然変異であり得る。一部の実施形態では、比較目的のための参照配列は、試験試料を提供する対象の種の野生型ゲノム配列、典型的にヒトゲノムである。

【0139】

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

【0140】

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、従来のサングーおよびキャピラリー電気泳動に基づくアプローチと比較して増加した処理能力を有する、例えば一度に何十万もの比較的小さい配列読み取りデータを生成する能力を有するシーケンシングテクノロジーを指す。次世代シーケンシング技術の一部の例としては、これらに限定されないが、合成によるシーケンシング、ライゲーションによるシーケンシング、およびハイブリダイゼーションによるシーケンシングが挙げられる。一部の実施形態では、次世代シーケンシングは、単一の分子をシーケンシングすることが可能な機器の使用を含む。次世代シーケンシングを実施するための市販の機器の例としては、これらに限定されないが、NextSeq、HiSeq、NovaSeq、MiSeq、IonPGM、およびIonGeneStudio S5が挙げられる。

【0141】

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

10

20

30

40

50

に／または配列の長さ)に基づいて識別することができるように、各核酸分子をタグ付けしてもよい。典型的に、任意の2つの分子が同じ内因性の配列情報(例えば、開始および／もしくは終止位置、配列の一方もしくは両方の末端の部分配列、ならびに／または長さ)を有し得る確率、同様に同じ分子バーコードを有し得る確率が低くなるように(例えば、約10%未満、約5%未満、約1%未満、または約0.1%未満の見込み)、十分数の異なる分子バーコードを使用する。

【0142】

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

【0143】

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

【0144】

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、ヌクレオシド間の連結によって接合したヌクレオシド(デオキシリボヌクレオシド、リボヌクレオシド、またはそのアナログを含む)の線形ポリマーを指す。典型的に、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しばしば、少数の単量体単位、例えば3~4から数百もの単量体単位のサイズの範囲である。ポリヌクレオチドが文字の配列、例えば「ATGCCCTG」によって表される場合は必ず、ヌクレオチドは、左から右に5'→3'の順であり、DNAの場合、特に記載されていない限り、「A」はデオキシアデノシンを示し、「C」は、デオキシチジンを示し、「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。文字A、C、G、およびTは、塩基そのもの、ヌクレオシド、または塩基を含むヌクレオチドを指すために使用され得る。

【0145】

本明細書で使用される場合、「プロセッシング」は、シーケンシングにとって好適である核酸のライブラリーを生成するために使用されるステップのセットを指す。ステップのセットは、これらに限定されないが、分配するステップ、末端修復するステップ、シーケンシングアダプターの付加、タグ付けするステップ、および／または核酸のPCR増幅を含み得る。

【0146】

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

【0147】

本明細書で使用される場合、「参照配列」は、実験によって決定した配列と比較する目的のために使用される公知の配列を指す。例えば、公知の配列は、全ゲノム、染色体、またはその任意のセグメントであり得る。参照配列は、ゲノム、もしくは染色体、もしくは染色体アームの単一の連続する配列と整列することができるか、またはゲノムもしくは染色体の異なる領域と整列する非連続セグメントを含み得る。参照配列の例としては、例えばヒトゲノム、例えばhg19およびhg38が挙げられる。

10

【0148】

本明細書で使用される場合、「制限酵素」は、特異的認識部位またはその付近でDNAを認識し、切断する酵素である。

【0149】

本明細書で使用される場合、「試料」は、本明細書に開示される方法および／またはシステムによって解析されることが可能な任意のものを意味する。

【0150】

本明細書で使用される場合、「シーケンシング」は、生体分子、例えばDNAまたはRNAなどの核酸の配列（例えば、同一性および単量体単位の順序）を決定するために使用されるいくつかのテクノロジーのいずれかを指す。シーケンシング方法の例としては、これらに限定されないが、標的化シーケンシング、一分子リアルタイムシーケンシング、エクソンまたはエクソームシーケンシング、イントロンシーケンシング、電子顕微鏡に基づくシーケンシング、パネルシーケンシング、トランジスタ媒介シーケンシング、直接シーケンシング、ランダムショットガンシーケンシング、サンガー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、二重鎖シーケンシング、サイクルシーケンシング、一塩基伸長シーケンシング、固相シーケンシング、ハイスループットシーケンシング、超並列シグネチャーシーケンシング、エマルジョンPCR、より低い変性温度での同時増幅PCR（COLD-PCR）、マルチプレックスPCR、可逆的色素ターミネーターによるシーケンシング、ペアードエンドシーケンシング、ニアターム（near-term）シーケンシング、エキソヌクレアーゼシーケンシング、ライゲーションによるシーケンシング、ショートリードシーケンシング、一分子シーケンシング、合成によるシーケンシング、リアルタイムシーケンシング、リバースターミネーターシーケンシング、ナノポアシーケンシング、454シーケンシング、Solexa Genome Analyzerシーケンシング、SOLiD（商標）シーケンシング、MS-PETシーケンシング、およびそれらの組合せが挙げられる。一部の実施形態では、シーケンシングは、遺伝子アナライザー、数ある中でも例えばIllumina, Inc.、Pacific Biosciences, Inc.、またはApplied Biosystems/Thermo Fisher Scientificなどから市販されている遺伝子アナライザーによって実施され得る。

20

30

40

【0151】

本明細書で使用される場合、核酸ポリマーの文脈における「配列情報」は、そのポリマーにおける単量体単位（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。

【0152】

本明細書で使用される場合、「配列可変標的領域セット」とは、新生物細胞（例えば、腫瘍細胞およびがん細胞）において、配列の変化、例えば、ヌクレオチド置換、挿入、欠失、または遺伝子融合もしくは転位を示し得る、標的領域のセットを指す。

【0153】

50

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

【0154】

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

【0155】

本明細書で使用される場合、「対象」は、動物、例えば哺乳動物種（例えば、ヒト）、または鳥類（例えば、鳥）種、または他の生物、例えば植物を指す。より具体的には、対象は脊椎動物、例えば哺乳動物、例えばマウス、霊長類、サル、またはヒトであり得る。動物は、農場動物（例えば、肉牛、乳牛、家禽、ウマ、ブタなど）、競技用動物、およびコンパニオン動物（例えば、ペットまたは介助動物）を含む。対象は、健康な個体、疾患もしくは疾患に対する素因を有するもしくは有することが疑われる個体、または治療を必要とするもしくは治療を必要とすることが疑われる個体であり得る。「個体」または「患者」という用語は、「対象」と互換的であると意図される。例えば、対象は、がんを有すると診断されている、がん治療を受ける予定である、および／または少なくとも1つのがん治療を受けたことがある個体であり得る。対象はがんの寛解期にあり得る。別の例として、対象は、自己免疫疾患を有すると診断された個体であり得る。別の例として、対象は、疾患、例えばがん、自己免疫疾患を有すると診断されているまたは有することが疑われ得る、妊娠中であるまたは妊娠を計画している女性個体であり得る。

【0156】

本明細書で使用される場合、「標的領域セット」または「標的領域のセット」、または「標的領域」、または「目的の標的領域」、または「目的の領域」、または「目的のゲノム領域」は、捕捉のために標的化されるおよび／またはプローブのセットによって標的化される（例えば、配列相補性を通して）複数のゲノム座または複数のゲノム領域を指す。

【0157】

本明細書で使用される場合、「腫瘍分画」は、所定の試料または試料－領域対に関する、腫瘍細胞を起源とするcfDNA分子の割合を指す。

【0158】

「またはその組合せ」（単数および複数）という用語は、本明細書で使用される場合、用語の前に列挙される用語のあらゆる全ての順列および組合せを指す。例えば、「A、B、C、またはそれらの組合せ」は、A、B、C、AB、AC、BC、またはABCのうちの少なくとも1つを含むと意図され、特定の文脈において順序が重要である場合、BA、CA、CB、ACB、CBA、BCA、BAC、またはCABも含むと意図される。この例を続けると、1つまたは複数の項目または用語の反復、例えばBB、AAA、AAB、BBC、AAABCCCC、CBBAAAA、CABABBBなどを含有する組合せが明白に含まれる。当業者は、文脈から特に明白でない限り、典型的に任意の組合せにおける項目または用語の数に制限はないことを理解するであろう。

【0159】

「または」は、包括的な意味で使用され、すなわち文脈が特に必要としていない限り、「および／または」と等価である。

B. 例示的な方法

1. 概要

【0160】

がんの形成および進行は、デオキシリボ核酸（DNA）の遺伝子改変およびエピジェネティック特色の両方から生じ得る。本開示は、無細胞DNA（cfDNA）などのDNAを解析するための方法およびシステムを提供する。本開示は、メチル化分配アッセイのシグナル対ノイズ比を低減するための方法およびシステムを提供する。

【0161】

いかなる特定の理論にも拘束されることを望まないが、がんまたは新生物におけるまたはその周囲の細胞は、健康な対象における同じ組織タイプの細胞よりも多くのDNAを脱落させ得る。そのため、ある特定のDNA試料、例えばcfDNAの起源の組織の分布は、発癌の際に変化し得る。このため、例えば、少なくとも1つの他の組織タイプよりも健康なcfDNAにおいて低いメチル化を示す高メチル化可変標的領域のレベルの増加は、がんの存在（または対象の既往に応じて再発）の指標であり得る。同様に、試料中の低メチル化可変標的領域のレベルの増加は、がんの存在（または対象の既往に応じて再発）の指標であり得る。

【0162】

さらに、がんは、メチル化などの非配列改変によって示され得る。がんにおけるメチル化変化の例としては、正常な成長の制御、DNA修復、細胞周期調節、および／または細胞分化に関係する遺伝子のTSSでCpGアイランドにおけるDNAメチル化の局所増加が挙げられる。この高メチル化は、関係する遺伝子の転写能の異常な喪失に関連し得、変更された遺伝子発現の原因としての点突然変異および欠失と少なくとも同程度の頻度で起こる。

【0163】

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

【0164】

一部の実施形態では、DNAメチル化は、CpG部位（シトシンーリン酸ーグアニン部位（すなわち、核酸配列の5'→3'方向にシトシンの後にグアニン）でのシトシン残基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N⁶-メチルアデニンの場合のように、アデニン残基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-炭素環の5番目の炭素の改変）である。一部の実施形態では、5-メチル化は、5-メチルシトシン（m5cまたは5-mCまたは5mC）を作製するためのシトシン残基の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、これらに限定されないが、5-ヒドロキシメチルシトシン（5-hmCまたは5hmC）、5-ホルミルシトシン（5-fC）、および5-カルボキシルシトシン（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（シトシン残基の6-炭素環の3番目の炭素の改変）である。一部の実施形態では、3Cメチル化は、3-メチルシトシン（3mC）を生成するためのシトシン残基の3C位へのメチル基の付加を含む。メチル化はまた、非CpG部位でも起こり得、例えばメチル化は、CpA、CpT、またはC

10

20

30

40

50

p C 部位で起こり得る。DNA メチル化は、メチル化 DNA 領域の活性を変化させることができる。例えば、プロモーター領域の DNA がメチル化される場合、遺伝子の転写は抑制され得る。DNA メチル化は、正常な発達にとって極めて重要であり、メチル化の異常はエピジェネティック調節を破壊し得る。エピジェネティック調節の破壊、例えば抑制は、がんなどの疾患を引き起こし得る。DNA におけるプロモーターのメチル化はがんを示し得る。

【0165】

メチル化プロファイリングは、ゲノムの異なる領域にわたるメチル化パターンを決定することを含み得る。例えば、メチル化の程度（例えば、分子あたりのメチル化ヌクレオチドの相対数）に基づいて分子を分配およびシーケンシングした後、異なる分配中の分子の配列を参照ゲノムにマッピングすることができる。これは、他の領域と比較してより高度にメチル化されるか、またはあまり高度にメチル化されないゲノムの領域を示すことができる。このようにして、ゲノム領域は、個々の分子と対比するとメチル化の程度が異なり得る。

10

【0166】

メチル化プロファイリングから得られたシグナルを、体細胞変異（例えば、SNV、インデル、CNV、および遺伝子融合）から得られたシグナルと組み合わせると、がんの検出が容易となる。

【0167】

試料中の核酸分子は、核酸分子のメチル化状態に基づいて、分画または分配され得る。試料中の核酸分子を分配することにより、希少なシグナルが増加し得る。例えば、高メチル化 DNA に存在するが、低メチル化 DNA にはあまり（またはまったく）存在しない、遺伝的変異は、試料を、高メチル化核酸分子と低メチル化核酸分子とに分配することによって、より容易に検出することができる。試料の複数の分画を解析することによって、単一分子の多次元解析を行うことができ、したがって、より高い感度を達成することができる。分配は、核酸分子を、1 つまたは複数のメチル化ヌクレオチド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分配することを含み得る。試料は、差次的遺伝子発現または疾患状態を示す特徴に基づいて、1 つまたは複数の分配セットに分画または分配され得る。試料は、核酸、例えば、無細胞 DNA（「cfDNA」）、非 cfDNA、腫瘍 DNA、循環腫瘍 DNA（「ctDNA」）、および無細胞核酸（「cfNA」）の解析中に、正常状態および罹患状態の間でシグナルに相違をもたらす特徴またはその組合せに基づいて、分画され得る。

20

30

【0168】

分配する手順は、部分試料間で、DNA 分子の不完全な選別をもたらす場合がある。例えば、第 2 の部分試料中の分子のうちの少数部分は、高度に改変されていてもよく（例えば、高メチル化）、および／または第 1 の部分試料中の分子のうちの少数部分は、改変されていないかもしくはほとんど改変されていなくてもよい（例えば、非メチル化であるかもしくはほとんど非メチル化である）。第 2 の部分試料中の高度に改変された分子および第 1 の部分試料中の改変されていないかほとんど改変されていない分子は、非特異的に分配されると考えられる。本明細書に記載される方法は、非特異的に分配された DNA に由来する技術的ノイズを、例えば、それを分解することによって、および／または非特異的に分配された DNA をシーケンシングの後に同定することができるようにある特定の塩基を変換することによって、低減させることができるステップを含む。したがって、本明細書に記載される方法は、改善された感度および／または合理化された解析を提供し得る。

40

【0169】

図 2 は、対象から得られた試料中の核酸分子のメチル化状態を決定するための方法 200 の例としての実施形態を図示する。202 において、ポリヌクレオチド試料を、対象から得る。一部の実施形態では、試料は、腫瘍組織生検から得られた DNA 試料である。一部の実施形態では、試料は、血液から得られた無細胞 DNA（cfDNA）試料である。204 において、ポリヌクレオチド試料を、少なくとも 2 つの分配セット（部分試料）に

50

分配する。一部の実施形態では、分配は、メチル化ヌクレオチドを含むポリヌクレオチドに優先的に結合する結合剤に対するポリヌクレオチドの差次的結合親和性に基づいて、核酸分子を分配することを含む。

【0170】

206において、少なくとも1つの分配セット中の核酸分子は、少なくとも1つのメチル化依存性ヌクレアーゼ、例えば、メチル化依存性制限酵素（MDRE）で消化される。一部の実施形態では、少なくとも1つの分配セット中の核酸は、少なくとも2つのMDREで消化される。一部の実施形態では、2つのMDREが、少なくとも1つの分配セット中の核酸分子を消化するために使用される。本明細書の他の箇所に記載されるMDREのうちのいずれかまたはその組合せを、使用することができる。

10

【0171】

一部の実施形態では、MDREでの制限消化の前に、少なくとも1つのアダプターを、核酸分子の少なくとも1つの末端（すなわち、DNA分子の5'および／または3'末端）に付着させる。他の実施形態では、消化の後であるが、208において富化する前に、少なくとも1つのアダプターを、核酸分子の少なくとも1つの末端に付着させる。一部の実施形態では、アダプターは、例えば、非メチル化であるヌクレオチドまたは適切なヌクレオチドアナログ（例えば、連結改変、例えば、ホスホロチオエートを有するヌクレオチドアナログ）の存在に起因して、メチル化依存性ヌクレアーゼまたは制限酵素による消化に対して抵抗性である。

【0172】

20

208において、MDRE消化の後に、1つまたは複数の分配セット中の核酸分子は、目的のゲノム領域に関して富化することができる。あるいは、富化ステップは、分配するステップの前に行うことができる。一部の実施形態では、目的のゲノム領域は、がんの検出のために差次的にメチル化された領域（例えば、高メチル化可変標的領域セットおよび／または低メチル化可変標的領域セット）を含み得る。210において、富化された分子の少なくともサブセットを、次世代シーケンサーによってシーケンシングする。212において、シーケンサーによって生成されたシーケンシング読み取りデータを、次いで、バイオインフォマティクツツール／アルゴリズムを使用して解析して、1つまたは複数の分配セット中の分子の数を決定し、次にこれを使用して、少なくとも1つの分配セット中の核酸分子の1つまたは複数の遺伝子座でのメチル化状態を決定する。一部の実施形態では、1つまたは複数の遺伝子座は、複数の遺伝子座を含み得る。一部の実施形態では、1つまたは複数の遺伝子座は、1つまたは複数のゲノム領域を含み得る。一部の実施形態では、ゲノム領域は、遺伝子のプロモーター領域であり得る。一部の実施形態では、シーケンシングの前に、核酸分子を、PCR増幅によって増幅することができる。一部の実施形態では、増幅において使用されるプライマーは、少なくとも1つの試料インデックスを含み得る。

30

【0173】

図3は、本開示の実施形態による、対象におけるがんの存在または非存在を検出するための方法300の例としての実施形態を図示する。302において、ポリヌクレオチド試料を、対象から得る。一部の実施形態では、ポリヌクレオチド試料は、腫瘍組織生検から得られたDNA試料である。一部の実施形態では、ポリヌクレオチド試料は、血液から得られた無細胞DNA（cfDNA）試料である。304において、ポリヌクレオチド試料を、少なくとも2つの分配セットに分配する。一部の実施形態では、分配は、メチル化ヌクレオチドを含むポリヌクレオチドに優先的に結合する結合剤に対するポリヌクレオチドの差次的結合親和性に基づいて、核酸分子を分配することを含む。結合剤の例としては、メチル結合ドメイン（MBD）およびメチル結合タンパク質（MBP）が挙げられるがこれらに限定されず、これらは、本明細書の他の箇所において詳細に考察されている。

40

【0174】

306において、1つまたは複数の分配セット中の核酸分子を、アダプターに付着させ、アダプターは、少なくとも1つのタグを含み、核酸分子の少なくとも一方の末端（すな

50

わち、DNA分子の5'および/3'末端)に付着させられる。一部の実施形態では、アダプターは、メチル化依存性制限酵素による消化に対して抵抗性である。一部の実施形態では、アダプターは、非メチル化ヌクレオチドを含む。一部の実施形態では、アダプターは、メチル化依存性制限酵素に対して抵抗性の1つまたは複数のヌクレオチドアナログを含む。一部の実施形態では、アダプターは、メチル化依存性制限酵素によって認識されないヌクレオチド配列を含む。一部の実施形態では、タグが、アダプターの構成要素として提供され得る。一部の実施形態では、タグは、分子バーコード(すなわち、分子識別子)を含む。一部の実施形態では、一方の分配セット中の核酸分子に付着させられるタグは、他方の分配セット中の核酸分子に付着させられるタグとは異なる。一部の実施形態では、一方の分配セットは、他方の分配セットとは差次的にタグ付けされる。分配セットの差次的タグ付けは、特定の分配セットに属する核酸分子を追跡し続けるのを補助する。異なる分配セット中の核酸分子は、一方の分配セットのメンバーを別のものと識別することが可能な異なるタグを受容する。同じ分配セットの核酸分子に連結されるタグは、互いに同じであっても、異なってもよい。しかし、互いに異なる場合、タグは、それらが付着した分子を、特定の分配セットのものであるとして同定するように、共通した配列の部分を持し得る。例えば、試料の分子が、P1およびP2の2つの分配セットに分配される場合には、P1内の分子は、A1、A2、A3などでタグ付けされ得、P2内の分子は、B1、B2、B3、などでタグ付けされ得る。そのようなタグ付けシステムにより、分配セット、および分配セット中の分子間で、識別することが可能となる。一部の実施形態では、タグは、分配タグ(すなわち、分配識別子)を含む。そのような実施形態では、分配セット中の核酸分子は、同じ分配タグを受容し、これは、他の分配セットの核酸分子に付着させられる分配タグとは異なる。

10

20

【0175】

308において、少なくとも1つの分配セット中の核酸分子は、少なくとも1つのメチル化依存性ヌクレアーゼまたはメチル化依存性制限酵素(MDRE)で消化される。一部の実施形態では、少なくとも1つの分配セット中の核酸は、少なくとも2つのMDREで消化される。一部の実施形態では、2つのMDREが、少なくとも1つの分配セット中の核酸分子を消化するために使用される。MDREは、本明細書の他の箇所に記載される任意のMDREまたはその組合せであり得る。

【0176】

310において、MDRE消化の後に、1つまたは複数の分配セット中の核酸分子は、目的のゲノム領域に関して富化させることができる。あるいは、富化ステップは、分配するステップの前に行うことができる。一部の実施形態では、目的のゲノム領域は、がん検出のために差次的にメチル化された領域を含み得る。312において、富化された分子の少なくともサブセットを、次世代シーケンサーによってシーケンシングする。314において、シーケンサーによって生成されたシーケンシング読み取りデータを、次いで、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つまたは複数の分配セット中の分子の数を決定し、次にこれを使用して、少なくとも1つの分配セット中の核酸分子の1つまたは複数の遺伝子座でのメチル化状態を決定する。一部の実施形態では、1つまたは複数の遺伝子座は、複数の遺伝子座を含み得る。一部の実施形態では、1つまたは複数の遺伝子座は、1つまたは複数のゲノム領域を含み得る。一部の実施形態では、ゲノム領域は、遺伝子のプロモーター領域であり得る。一部の実施形態では、シーケンシングの前に、核酸分子を、PCR増幅によって増幅することができる。一部の実施形態では、増幅において使用されるプライマーは、少なくとも1つの試料インデックスを含み得る。

30

40

【0177】

一部の実施形態では、方法は、例えば、少なくとも1つの分配セット中の核酸分子の1つまたは複数の遺伝子座でのメチル化状態に基づいて、対象におけるがんの存在または非存在を検出するステップをさらに含み得る。一部の実施形態では、方法は、ポリヌクレオチド試料における腫瘍細胞に由来するDNAのレベルを決定するステップをさらに含む。

50

【0178】

図4は、例えば、cfDNA試料で開始する、本開示のある特定の実施形態による、がんの存在または非存在を検出するための例示的なワークフローを図示し、ここで、cfDNAを、血液試料から単離し、cfDNA試料は、高メチル化可変標的領域（高DMR）に属するcfDNA分子および低メチル化可変標的領域（低DMR）に属するcfDNA分子、ならびに非メチル化対照領域に属するcfDNA分子を含む。cfDNAは、メチル結合ドメインタンパク質（MBD）を使用して、低メチル化部分試料および高メチル化部分試料に分配され、各分配セットを、分子バーコード化に供して、部分試料に由来するDNAを識別可能にタグ付けし、低分配セットを、1つまたは複数のMDREで消化させ、メチル化cfDNA分子を、RE認識部位で切断し、必要に応じて、高分配セットを、1つまたは複数のMSREで消化させ、非メチル化cfDNA分子を、RE認識部位で切断し、次いで、分配セット（MDREで消化させた低分配セットを含む）をプールし、捕捉し、増幅させ、シーケンシングする。

10

2. 試料を複数の部分試料に分配する；試料の態様

【0179】

本明細書に記載されるある特定の実施形態では、異なる形態の核酸の集団（例えば、試料中の高メチル化および低メチル化DNA、例えば、cfDNA）は、さらなる解析、例えば、ヌクレアーゼと接触させ、核酸塩基を差次的に改変もしくは単離し、タグ付けし、および／またはシーケンシングする前に、核酸の1つまたは複数の特徴に基づいて、物理的に分配され得る。このアプローチを使用して、例えば、ある特定の配列が、高メチル化であるかまたは低メチル化であるかを決定することができる。さらに、不均一な核酸集団を分配することによって、例えば、集団の1つの分画（または分配）により多く存在する希少な核酸分子を富化することにより、希少なシグナルを増加させることができる。例えば、高メチル化DNAに存在するが、低メチル化DNAにはあまり（またはまったく）存在しない、遺伝的変異は、試料を、高メチル化核酸分子と低メチル化核酸分子とに分配することによって、より容易に検出することができる。試料の複数の分画を解析することによって、核酸のゲノムまたは種の単一の座の多次元解析を行うことができ、したがって、より高い感度を達成することができる。

20

【0180】

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

30

【0181】

分配に使用することができる特徴の例としては、配列の長さ、メチル化レベル、ヌクレオソーム結合、配列ミスマッチ、免疫沈降、および／またはDNAに結合するタンパク質が挙げられる。結果として得られる分配は、以下の核酸形態のうちの1つまたは複数を含み得る：一本鎖DNA（ssDNA）、二本鎖DNA（dsDNA）、より短いDNA断片、およびより長いDNA断片。一部の実施形態では、シトシン改変（例えば、シトシンメチル化）またはメチル化に基づく分配が一般に行われ、必要に応じて、前述のDNAの特徴または形態のうちのいずれかに基づき得る少なくとも1つの追加の分配するステップと組み合わせられる。一部の実施形態では、不均一な核酸集団は、1つまたは複数のエピジェネティック改変を有する核酸および1つまたは複数のエピジェネティック改変を有さない核酸に分配される。エピジェネティック改変の例としては、メチル化の存在または非存在；メチル化のレベル；メチル化のタイプ（例えば、5-メチルシトシン対他のタイプのメチル化、例えば、アデニンメチル化および／またはシトシンヒドロキシメチル化）、な

40

50

らびに1つまたは複数のタンパク質、例えばヒストンとの会合および会合のレベルが挙げられる。あるいはまたはさらに、不均一な核酸集団は、ヌクレオソームと会合した核酸分子およびヌクレオソームが欠如した核酸分子に分配され得る。あるいはまたはさらに、不均一な核酸集団は、一本鎖DNA (ssDNA) および二本鎖DNA (dsDNA) に分配してもよい。あるいはまたはさらに、不均一な核酸集団は、核酸の長さ (例えば、最大で160bpの分子および160bpより長い長さを有する分子) に基づいて分配してもよい。

【0182】

一部の例では、各分配 (異なる核酸形態を表す) は、差次的に標識され、分配は、シーケンシングする前に一緒にプールされる。他の例では、異なる形態は、別個にシーケンシングされる。

10

【0183】

一部の実施形態では、異なる核酸の集団は、2つまたはそれよりも多くの異なる分配に分配される。各分配は、異なる核酸形態を表し、第1の分配 (部分試料とも称される) は、第2の部分試料よりも高い割合でシトシン改変を有するDNAを含む。各分配は、明確にタグ付けされる。第1の部分試料は、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供され、第1の核酸塩基は、改変または非改変核酸塩基であり、第2の核酸塩基は、第1の核酸塩基とは異なる改変または非改変核酸塩基であり、第1の核酸塩基および第2の核酸塩基は、同じ塩基対形成特異性を有する。タグ付けされた核酸は、シーケンシングの前に一緒にプールされる。配列読み取りデータを得、例えば、第1の部分試料のDNAにおいて第1の核酸塩基を第2の核酸塩基と識別するために、*in silico*で解析する。タグを使用して、異なる分配に由来する読み取りデータを選別する。遺伝的変異体を検出するための解析は、分配ごとのレベル、ならびに全核酸集団レベルで行われ得る。例えば、解析は、各分配中の核酸における遺伝的変異体、例えば、CNV、SNV、インデル、融合を決定するための*in silico*解析を含み得る。一部の例では、*in silico*解析は、クロマチン構造を決定することを含み得る。例えば、配列読み取りデータのカバレッジを使用して、クロマチンにおけるヌクレオソームの位置付けを決定することができる。より高いカバレッジは、ゲノム領域におけるより高いヌクレオソーム占有率と相関し、一方でより低いカバレッジは、より低いヌクレオソーム占有率またはヌクレオソーム枯渇領域 (NDR) と相関し得る。

20

30

【0184】

試料は、ヌクレオチドに対する複製後改変を含む改変、および通常は非共有結合による1つまたは複数のタンパク質への結合が異なる核酸を含み得る。

【0185】

一実施形態では、核酸の集団は、新生物、腫瘍、もしくはがんを有することが疑われる対象、または新生物、腫瘍、もしくはがんを有すると以前に診断されている対象に由来する血清、血漿、または血液試料から得られたものである。核酸の集団は、様々なレベルのメチル化を有する核酸を含む。メチル化は、任意の1つまたは複数の複製後または転写改変により生じ得る。複製後改変は、ヌクレオチドシトシンの改変、特に、核酸塩基の5位におけるもの、例えば、5-メチルシトシン、5-ヒドロキシメチルシトシン、5-ホルミルシトシン、および5-カルボキシルシトシンを含む。

40

【0186】

親和性剤は、所望の特異性を有する抗体、天然の結合パートナーもしくはその変異体 (Bock et al., Nat Biotech 28: 1106-1114 (2010)、Song et al., Nat Biotech 29: 68-72 (2011))、または例えばファージディスプレイによって所与の標的に対する特異性を有するように選択された人工ペプチドであり得る。

【0187】

本明細書において企図される捕捉部分の例としては、本明細書に記載されるメチル結合ドメイン (MBD) およびメチル結合タンパク質 (MBP) が挙げられ、これには、タン

50

パク質、例えば、MeCP2、MBD、例えば、MBD2、ならびに5-メチルシトシンに優先的に結合する抗体が挙げられる。抗体を使用してメチル化DNAを免疫沈降する場合、メチル化DNAは、一本鎖形態で回収され得る。そのような実施形態では、第2の鎖を合成することができる。高メチル化（および必要に応じて中間メチル化）部分試料は、次いで、半メチル化DNAを切断しないメチル化感受性ヌクレアーゼ、例えば、HpaI、BstUI、またはHin6Iと接触させられ得る。あるいはまたはさらに、低メチル化（および必要に応じて、中間メチル化）部分試料は、次いで、半メチル化DNAを切断するメチル化依存性ヌクレアーゼと接触させられ得る。

【0188】

同様に、異なる形態の核酸を分配するステップは、ヒストンに結合した核酸を遊離または非結合の核酸から分離することができるヒストン結合タンパク質を使用して行うことができる。本明細書に開示される方法において使用することができるヒストン結合タンパク質の例としては、RBBP4、RbAp48、およびSANTドメインペプチドが挙げられる。

【0189】

一部の親和性剤および改変に関して、作用剤への結合は、核酸が改変を有するかどうかに応じて、本質的に全か無かの様式で生じ得るが、分離は、程度に応じたものであってもよい。そのような例では、改変が過剰発現する核酸は、改変が過少発現する核酸よりも高い程度で、作用剤に結合する。あるいは、改変を有する核酸は、全か無かの様式で結合し得る。しかし、次いで、様々なレベルの改変は、結合剤から逐次的に溶出されてもよい。

【0190】

例えば、一部の実施形態では、分配は、バイナリであり得るか、または改変の程度／レベルに基づき得る。例えば、全てのメチル化断片は、メチル結合ドメインタンパク質（例えば、MethylMiner Methylated DNA Enrichment Kit (ThermoFisher Scientific)）を使用して、非メチル化断片から分配することができる。続いて、さらなる分配には、メチル結合ドメインおよび結合断片を有する溶液中の塩濃度を調整することによって、異なるレベルのメチル化を有する断片を溶出することが含まれ得る。塩濃度が増加するにつれ、より高いメチル化レベルを有する断片が溶出される。

【0191】

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

【0192】

MethylMiner Methylated DNA Enrichment Kit (ThermoFisher Scientific)を使用する場合、様々なレベルのメチル化を、逐次溶出を使用して分配することができる。例えば、低メチル化分配（メチル化なし）は、核酸集団を、磁気ビーズに付着した、キットからのMBDと接触させることによって、メチル化分配から分離することができる。ビーズを使用して、メチル化核酸を、非メチル化核酸から分離する。続いて、1つまたは複数の溶出ステップを逐次的に行って、異なるレベルのメチル化を有する核酸を溶出させる。例えば、メチル化核酸の第1のセットを、160mMまたはそれよりも高い、例えば、少なくとも150mM、少なくとも200mM、300mM、400mM、500mM、600mM、700mM、800mM、900mM、1000mM、または2000mMの塩濃度で溶出させることができる。そのようなメチル化核酸を溶出させた後、磁気分離を、もう一度使用して、よ

り高いレベルのメチル化核酸を、より低いレベルのメチル化を有するものから分離する。溶出および磁気分離ステップ自体を反復して、様々な分配、例えば、低メチル化分配（メチル化がないことを表す）、メチル化分配（低いレベルのメチル化を表す）、および高メチル化分配（高いレベルのメチル化を表す）を作成することができる。

【0193】

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

10

【0194】

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

【0195】

メチル化などの特徴に基づいて核酸試料を分配することに関するさらなる詳細については、参照により本明細書に組み込まれるWO 2018/119452を参照されたい。

20

【0196】

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

【0197】

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

30

【0198】

一部の実施形態では、核酸を分配するステップは、核酸を、メチル化結合タンパク質（「MBP」）のメチル化結合ドメイン（「MBD」）と接触させることによって行われる。MBDは、5-メチルシトシン（5mC）に結合する。MBDは、常磁性ビーズ、例えば、Dynabeads（登録商標）M-280ストレプトアビジンにビオチンリンカーを介してカップリングされる。異なる程度のメチル化を有する分画への分配は、NaCl濃度を増加させることによって分画を溶出させることにより行うことができる。

40

【0199】

本明細書において企図されるMBPの例としては、以下のものが挙げられるが、これらに限定されない：

（a）MeCP2およびMBD2は、非改変シトシンよりも、5-メチルシトシンに優先的に結合するタンパク質である。

（b）RPL26、PRP8、およびDNAミスマッチ修復タンパク質MHS6は、非改

50

変シトシンよりも、5-ヒドロキシメチルシトシンに優先的に結合する。

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

(d) 1つまたは複数のメチル化ヌクレオチド塩基に特異的な抗体。

【0200】

一般に、溶出は、分子あたりのメチル化部位の数の関数であり、高い塩濃度下では、より多くのメチル化を有する分子が溶出する。DNAを、メチル化の程度にも基づいて、別個の集団に溶出させるために、漸増NaCl濃度の一連の溶出緩衝液を使用することができ。塩濃度は、約100nM~約2500mMのNaClの範囲であり得る。一実施形態では、プロセスにより、3つの分配が得られる。分子を、第1の塩濃度で、メチル結合ドメインを含む分子を含む溶液と接触させ、分子が、捕捉部分、例えば、ストレプトアビジンに付着し得る。第1の塩濃度において、ある分子集団は、MBDに結合し、ある集団は、結合しないままとなる。非結合の集団を、「低メチル化」集団として分離することができる。例えば、DNAの低メチル化形態を表す第1の分配は、低い塩濃度、例えば、100mMまたは160mMにおいて、非結合のままであるものである。中間メチル化DNAを表す第2の分配は、中間の塩濃度、例えば、100mM~2000mMの間の濃度を使用して溶出される。これもまた、試料から分離される。DNAの高メチル化形態を表す第3の分配は、高い塩濃度、例えば、少なくとも約2000mMを使用して溶出される。

a. 分配のタグ付け

【0201】

一部の実施形態では、2つまたはそれよりも多くの分配、例えば、各分配は、差次的にタグ付けされる。タグまたはインデックスは、タグが会合する分子の特色を示す情報を含む、分子、例えば、核酸であり得る。タグは、配列読み取りデータの起源である分子を区別することを可能にし得る。例えば、分子は、試料タグまたは試料インデックス（これにより、1つの試料中の分子を、異なる試料中ものと識別する）、分配タグ（これにより、1つの分配中の分子を、異なる分配中ものと識別する）、または分子タグ／分子バーコード／バーコード（これにより、異なる分子を、互いに識別する（一意的および非一意的タグ付けシナリオの両方で））を有し得る。ある特定の実施形態では、タグは、1つのバーコードまたはバーコードの組合せを含み得る。本明細書で使用される場合、「バーコード」という用語は、文脈に応じて、特定のヌクレオチド配列を有する核酸分子、またはヌクレオチド配列自体を指す。バーコードは、例えば、10~100個の間のヌクレオチドを有し得る。バーコードのコレクションは、具体的な目的に所望されるように、縮重配列を有し得るか、またはある特定のハミング距離を有する配列を有し得る。そのため、例えば、分子バーコードは、1つのバーコード、または各々が分子の異なる末端に付着する2つのバーコードの組合せから構成され得る。さらにまたはあるいは、異なる分配および／または試料について、バーコードが、個々の配列を通じて分子タグとしての機能を果たし、それらがメンバーであるセットに基づいて対応する分配および／または試料を同定する機能も果たすように、異なるセットの分子バーコード、分子タグ、または分子インデックスが、使用され得る。バーコードを含むタグは、アダプターに組み込まれ得るか、またはそうでなければ、アダプターに接合され得る。タグは、他の方法の中でもとりわけ、ライゲーション、オーバーラップ伸長PCRによって、組み込まれ得る。

【0202】

タグ付け戦略は、一意的タグ付け戦略および非一意的タグ付け戦略に分けることができる。一意的タグ付けでは、試料中の全てまたは実質的に全ての分子が、異なるタグを有し、そのため、読み取りデータは、タグ情報単独に基づいて、もともとの分子に割り当てることができる。そのような方法において使用されるタグは、「一意的タグ」と称される場合がある。非一意的タグ付けでは、同じ試料中の異なる分子が同じタグを有し得、そのため、タグ情報に加えて他の情報を使用して、配列読み取りデータがもともとの分子に割り当てられる。そのような情報としては、開始および終止座標、分子がマッピングされる座

10

20

30

40

50

標、開始または終止座標単独などを挙げることができる。そのような方法において使用されるタグは、「非一意的タグ」と称される場合がある。したがって、必ずしも、試料中の全ての分子を一意的にタグ付けするわけではない。試料内の同定可能なクラス内に含まれる分子を一意的にタグ付けすることで十分である。したがって、異なる同定可能なファミリー内の分子は、タグ付けされた分子の同一性に関する情報を失うことなく、同じタグを有し得る。

【0203】

非一意的タグ付けのある特定の実施形態では、使用される異なるタグの数は、特定の群の全ての分子が異なるタグを有する非常に高い可能性（例えば、少なくとも99%、少なくとも99.9%、少なくとも99.99%、または少なくとも99.999%）であれば、十分であり得る。バーコードをタグとして使用する場合、およびバーコードが、例えば、ランダムに分子の両方の末端に付着している場合、バーコードの組合せが、一緒に、タグを構成し得ることに留意されたい。この数は、明確にコールに含まれる分子の数の関数である。例えば、クラスは、参照ゲノム上の同じ開始-終止位置にマッピングされる全ての分子であり得る。クラスは、特定の遺伝子座、例えば、特定の塩基または特定の領域（例えば、最大100個の塩基または遺伝子または遺伝子のエクソン）にわたってマッピングする全ての分子であり得る。ある特定の実施形態では、クラス内の分子の数 z を一意的に同定するために使用される異なるタグの数は、 $2 * z$ 、 $3 * z$ 、 $4 * z$ 、 $5 * z$ 、 $6 * z$ 、 $7 * z$ 、 $8 * z$ 、 $9 * z$ 、 $10 * z$ 、 $11 * z$ 、 $12 * z$ 、 $13 * z$ 、 $14 * z$ 、 $15 * z$ 、 $16 * z$ 、 $17 * z$ 、 $18 * z$ 、 $19 * z$ 、 $20 * z$ 、または $100 * z$ のうちのいずれか（例えば、下限）と、 100 、 $000 * z$ 、 10 、 $000 * z$ 、 $1000 * z$ 、または $100 * z$ のうちのいずれか（例えば、上限）との間であり得る。

【0204】

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

【0205】

一般に、反応における一意的または非一意的タグバーコードの割り当ては、米国特許出願第20010053519号、同第20030152490号、同第20110160078号、ならびに米国特許第6,582,908号および米国特許第7,537,898号および米国特許第9,598,731号によって記載される方法およびシステムに従う。

【0206】

タグは、ランダムまたは非ランダムに、試料核酸に連結され得る。

【0207】

一部の実施形態では、タグ付けされた核酸は、マイクロウェルプレートにロードした後シーケンシングされる。マイクロウェルプレートは、96、384、または1536個のマイクロウェルを有し得る。一部の場合には、それらは、予想される、一意的タグのマイクロウェルに対する比で、導入される。例えば、一意的タグは、ゲノム試料あたり約1個を上回る、2個を上回る、3個を上回る、4個を上回る、5個を上回る、6個を上回る、7個を上回る、8個を上回る、9個を上回る、10個を上回る、20個を上回る、50個を上回る、100個を上回る、500個を上回る、1000個を上回る、5000個を上回る、10000個を上回る、50,000個を上回る、100,000個を上回る、500,000個を上回る、1,000,000個を上回る、10,000,000個を上回る、50,000,000個を上回る、または1,000,000,000個を上回る一意的タグがロードされるように、ロードされ得る。一部の場合には、一意的タグは、

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

10

20

【0208】

好ましいフォーマットは、標的核酸の両方の末端にライゲーションされた20～50個の異なるタグ（例えば、バーコード）を使用する。例えば、標的分子の両方の末端に、35個の異なるタグ（例えば、バーコード）をライゲーションすると、 35×35 の順列を作製し、これは、35個のタグについて1225個に等しい。そのような数のタグは、同じ開始点および終止点を有する異なる分子が、高い確率（例えば、少なくとも94%、99.5%、99.99%、99.999%）で異なる組合せのタグを受容するのに十分である。他のバーコードの組合せとしては、10～500の間の任意の数、例えば、約 15×15 、約 35×35 、約 75×75 、約 100×100 、約 250×250 、約 500×500 が挙げられる。

30

【0209】

一部の場合には、一意的タグは、所定のまたはランダムなまたは半ランダムな配列のオリゴヌクレオチドであり得る。他の場合には、バーコードが複数の中で必ずしも互いに一意的とならないような、複数のバーコードが使用されてもよい。この例では、バーコードは、バーコードおよびそれがライゲーションされ得る配列の組合せが、個々に追跡され得る一意的配列を作製するように、個々の分子にライゲーションされ得る。本明細書に記載されるように、配列読み取りデータの始まり（開始）および終わり（終止）の部分の配列データと組み合わせた非一意的バーコードの検出により、特定の分子に対する一意的同一性の割り当てが可能となり得る。個々の配列読み取りデータの長さまたは塩基対の数もまた、そのような分子に一意的同一性を割り当てるために使用することができる。本明細書に記載されるように、一意的同一性が割り当てられた核酸の一本鎖に由来する断片は、それによって、後続の親鎖に由来する断片の同定を可能にし得る。

40

【0210】

タグ（単数または複数）を特定の分配と関連付けるために、タグを使用して、個々のポリヌクレオチド集団分配を標識することができる。あるいは、タグは、分配するステップを用いない本発明の実施形態において使用することができる。一部の実施形態では、単一のタグを使用して、特定の分配を標識することができる。一部の実施形態では、複数の異なるタグを使用して、特定の分配を標識することができる。複数のタグを用いて特定の分配を標識する実施形態では、1つの分配を標識するために使用されるタグのセットは、他の分配を標識するために使用されるタグのセットと容易に区別することができる。一部の

50

実施形態では、タグは、追加の機能を有してもよく、例えば、タグを、試料の起源に索引を付けるために使用するか、または一意的分子識別子（これは、例えば、Kinde et al., Proc Nat'l Acad Sci USA 108: 9530-9535 (2011)、Kou et al., PLoS ONE, 11: e0146638 (2016)にあるように、シーケンシングエラーを突然変異とは区別することによって、シーケンシングデータの品質を改善するために使用することができる）として使用するか、または例えば、米国特許第9,598,731号に記載されるように、非一意的分子識別子として使用することができる。同様に、一部の実施形態では、タグは、追加の機能を有し得、例えば、タグを、試料の起源に索引を付けるために使用するか、または非一意的分子識別子（これは、シーケンシングエラーを突然変異とは区別することによって、シーケンシングデータの品質を改善するために使用することができる）として使用することができる。

10

【0211】

一実施形態では、分配のタグ付けは、各分配における分子を、分配タグでタグ付けすることを含む。分配を再度組み合わせ（例えば、必要とされるシーケンシング実行数を低減させ、不必要な費用を回避するため）、分子をシーケンシングした後に、分配タグにより、起源の分配が同定される。別の実施形態では、異なる分配は、例えば、バーコード対から構成される、分子タグの異なるセットでタグ付けされる。このようにして、各分子バーコードは、起源の分配、ならびに分配内の分子を識別するのに有用であることを示す。例えば、35個のバーコードの第1のセットを使用して、第1の分配における分子をタグ付けすることができ、一方で、35個のバーコードの第2のセットを使用して、第2の分配中の分子をタグ付けすることができる。

20

【0212】

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

【0213】

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

30

【0214】

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

40

【0215】

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

【0216】

50

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

【0217】

一部の実施形態では、方法において使用されるヌクレアーゼによって認識される配列を含まない、および／または例えば、ヌクレオチド改変、例えば、連結改変（例えば、ホスホロチオエート）の存在に起因して、切断に対して抵抗性である、アダプターが使用される。一部の実施形態では、方法において使用されるヌクレアーゼによって認識される配列を含まない、および／または例えば、ヌクレオチド改変、例えば、連結改変（例えば、ホスホロチオエート）の存在に起因して、切断に対して抵抗性である、タグが使用される。1つまたは複数のメチル化依存性制限酵素および1つまたは複数のメチル化感受性制限酵素の両方が使用される場合、アダプターおよび／またはタグは、それらが、使用される制限酵素のうちのいずれかによる切断の基質とはならないように、メチル化が欠如し得、1つまたは複数のメチル化感受性制限酵素の認識配列が欠如し得る。

10

b. 改変された核酸解析の代替的な方法

20

【0218】

一部の実施形態では、アダプターは、核酸を分配した後に核酸に付加され、他の実施形態では、アダプターは、核酸を分配する前に核酸に付加され得る。一部のそのような方法では、異なる程度で改変（例えば、核酸分子あたり0個、1個、2個、3個、4個、5個、またはそれよりも多くのメチル基）を有する核酸の集団を、改変の程度に応じて、集団の分画の前に、アダプターと接触させる。アダプターは、集団内の核酸分子の一方の末端または両方の末端に付着する。好ましくは、アダプターは、タグの組合せ数が、同じ開始および終止点を有する2つの核酸が同じタグの組合せを受容する低い確率、例えば、95、99、または99.9%をもたらすのに十分な数の異なるタグを含む。アダプターは、同じタグを有するかまたは異なるタグを有するかに関係なく、同じかまたは異なるプライマー結合部位を含み得るが、好ましくは、アダプターは、同じプライマー結合部位を含む。アダプターの付着後に、核酸を、改変を有する核酸に優先的に結合する作用剤（例えば、以前に説明されたそのような作用剤）と接触させる。核酸を、作用剤への結合から、核酸が改変を有する程度が異なる少なくとも2つの部分試料に分配する。例えば、作用剤が、改変を有する核酸に対する親和性を有する場合、改変が過剰発現した核酸（集団内の中央値出現と比較して）は、作用剤に優先的に結合し、一方で、改変が過少発現する核酸は、結合しないか、または作用剤からより容易に溶出される。分配した後に、第1の部分試料は、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供され、第1の核酸塩基は、改変または非改変核酸塩基であり、第2の核酸塩基は、第1の核酸塩基とは異なる改変または非改変核酸塩基であり、第1の核酸塩基および第2の核酸塩基は、同じ塩基対形成特異性を有する。核酸を、次いで、アダプター内のプライマー結合部位に結合するプライマーから増幅させる。増幅の後に、異なる分配は、次いで、さらなるプロセッシングステップに供され得るが、これには、典型的に、並行であるが別個での、さらなる（例えば、クローン）増幅、および配列解析が含まれる。異なる分配からの配列データを、次いで、比較することができる。

30

40

【0219】

別の実施形態では、以下の例示的な手順を使用して、分配するスキームを行うことができる。核酸を、プライマー結合部位およびタグを含む、Y形状アダプターの両方の末端に連結させる。分子を増幅させる。増幅させた分子を、次いで、5-メチルシトシンに優先的に結合する抗体との接触によって分画して、2つの分配を産生する。一方の分配には

50

、メチル化が欠如したもともとの分子、およびメチル化が失われた増幅コピーが含まれる。他方の分配には、メチル化を有するもともとのDNA分子が含まれる。メチル化を有するもともとのDNA分子を含む分配を、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供し、ここで、第1の核酸塩基は、改変または非改変核酸塩基であり、第2の核酸塩基は、第1の核酸塩基とは異なる改変または非改変核酸塩基であり、第1の核酸塩基および第2の核酸塩基は、同じ塩基対形成特異性を有する。2つの分配を、次いで、メチル化分配のさらなる増幅とは別個に、プロセッシングおよびシーケンシングする。2つの分配の配列データを、次いで、比較することができる。この例では、タグは、メチル化DNAと非メチル化DNAとを識別するためではなく、同じ開始および終止点を有する読み取りデータが、同じ分子に基づくか異なる分子に基づくかを決定することができるように、これらの分配内の異なる分子間での識別するために使用される。

10

【0220】

本開示はさらに、核酸のうちの少なくとも一部が5-メチルシトシンなどの1つまたは複数の改変シトシン残基および前述の他の改変のうちのいずれかを含む、核酸の集団を解析するための方法を提供する。これらの方法において、分配した後に、核酸の部分試料を、5'位置において改変された1つまたは複数のシトシン残基、例えば、5-メチルシトシンを含むアダプターと接触させる。好ましくは、そのようなアダプター内の全てのシトシン残基も改変されているか、またはアダプターのプライマー結合領域内の全てのそのようなシトシンが改変されている。アダプターは、集団内の核酸分子の両方の末端に付着する。好ましくは、アダプターは、タグの組合せ数が、同じ開始および終止点を有する2つの核酸が同じタグの組合せを受容する低い確率、例えば、95、99、または99.9%をもたらすのに十分な数の異なるタグを含む。そのようなアダプター内のプライマー結合部位は、同じであっても異なってもよいが、好ましくは、同じである。アダプターの付着後に、核酸を、アダプターのプライマー結合部位に結合するプライマーから増幅させる。増幅させた核酸を、第1のアリコートおよび第2のアリコートに分ける。第1のアリコートを、さらなるプロセッシングありまたはなしで、配列データに関してアッセイする。第1のアリコート内の分子の配列データは、したがって、核酸分子の初期メチル化状態に関係なく、決定される。第2のアリコート内の核酸分子を、DNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供し、ここで、第1の核酸塩基は、5'位で改変されたシトシンを含み、第2の核酸塩基は、非改変シトシンを含む。この手順は、バイサルファイト処置または非改変シトシンをウラシルに変換する別の手順であり得る。手順に供した核酸を、次いで、核酸に連結したアダプターのもともとのプライマー結合部位に対するプライマーで増幅させる。これらの核酸は、アダプターのプライマー結合部位にシトシンを保持しているが、一方で増幅産物は、バイサルファイト処置においてウラシルへの変換を受けてこれらのシトシン残基のメチル化が失われているため、アダプターにもともと連結していた核酸分子（その増幅産物とは異なる）のみが、ここで増幅可能である。したがって、少なくとも一部がメチル化されている集団内のもともとの分子のみが、増幅を受ける。増幅後に、これらの核酸は、配列解析に供される。第1のアリコートおよび第2のアリコートから決定された配列の比較は、とりわけ、核酸集団内のどのシトシンがメチル化に供されたかを示し得る。

20

30

40

【0221】

そのような解析は、以下の例示的な手順を使用して行うことができる。分配した後に、メチル化DNAは、プライマー結合部位およびタグを含むY形状アダプターの両方の末端に連結される。アダプター内のシトシンは、5'位において改変されている（例えば、5-メチル化）。アダプターの改変は、後続の変換ステップ（例えば、バイサルファイト処置、TAP変換、または改変シトシンには影響を及ぼさないが、非改変シトシンに影響を及ぼす任意の他の変換）においてプライマー結合部位を保護するように機能する。アダプターの付着後に、DNA分子を増幅させる。増幅産物を、変換ありおよび変換なしでのシーケンシングのために2つのアリコートに分ける。変換に供さないアリコートは、さらなる

50

プロセッシングありまたはなしで、配列解析に供され得る。他方のアリコートは、DNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供され、ここで、第1の核酸塩基は、5位で改変されたシトシンを含み、第2の核酸塩基は、非改変シトシンを含む。この手順は、バイサルファイト処置または非改変シトシンをウラシルに変換する別の手順であり得る。もともとのプライマー結合部位に特異的なプライマーと接触させた場合に、シトシンの改変によって保護されたプライマー結合部位のみが、増幅を補助し得る。したがって、第1の増幅に由来するコピーではなく、もともとの分子のみが、さらなる増幅に供される。さらに増幅された分子を、次いで、配列解析に供する。配列を、次いで2つのアリコートから比較することができる。上記で考察される分離スキームにあるように、アダプター内の核酸タグは、メチル化DNAと非メチル化DNAとを識別するためではなく、同じ分配内の核酸分子を識別するために使用される。

10

3. 部分試料を、メチル化依存性またはメチル化感受性ヌクレアーゼと接触させる

【0222】

一部の実施形態では、部分試料（例えば、本明細書に記載されるように、例えば、シトシン改変、例えば、メチル化、例えば、5-メチル化のレベルに基づいて、試料を分配することによって調製された、第1、第2、または第3の部分試料）を、メチル化依存性ヌクレアーゼまたはメチル化感受性ヌクレアーゼと接触させる。別途示されない限り、分配が、シトシン改変に基づいて行われる場合、第1の部分試料は、より高いレベルの改変を有する部分試料であり、第2の部分試料は、より低いレベルの改変を有する部分試料であり、存在する場合、第3の部分試料は、第1の部分試料と第2の部分試料との間の中間のレベルの改変を有する。

20

【0223】

上記で考察されるように、分配する手順は、部分試料間で、DNA分子の不完全な選別をもたらす場合がある。メチル化依存性ヌクレアーゼまたはメチル化感受性ヌクレアーゼの選択は、非特異的に分配されたDNAを分解するように行われ得る。例えば、第2の部分試料を、メチル化依存性ヌクレアーゼ、例えば、メチル化依存性制限酵素と接触させてもよい。これにより、第2の部分試料中の非特異的に分配されたDNA（例えば、メチル化DNA）を分解して、処置された第2の部分試料を産生することができる。あるいはまたはさらに、第1の部分試料を、メチル化感受性エンドヌクレアーゼ、例えば、メチル化感受性制限酵素と接触させ、それによって、第1の部分試料中の非特異的に分配されたDNAを分解して、処置された第1の部分試料を産生することができる。第1または第2の部分試料の一方または両方における非特異的に分配されたDNAの分解は、例えば、試料中の異常に改変されたDNAの存在を検出するため、DNAの起源である組織を決定するため、および／または対象ががんを有するかどうかを決定するために、シトシン改変に基づくDNAの正確な分配に依存する方法の性能の改善として、提案される。例えば、そのような分解は、改善された感度をもたらし、かつ／または下流の解析を単純化し得る。一般に、低メチル化分配においてなど、非特異的に分配されたDNAが高メチル化である場合、メチル化依存性ヌクレアーゼ、例えば、メチル化依存性制限酵素を使用すべきである。対照的に、高メチル化分配においてなど、非特異的に分配されたDNAが低メチル化である場合、メチル化感受性ヌクレアーゼ、例えば、メチル化感受性制限酵素を使用すべきである。メチル化依存性ヌクレアーゼ、例えば、メチル化依存性制限酵素は、非メチル化DNAと比べて、メチル化DNAを優先的に切断するが、メチル化感受性ヌクレアーゼ、例えば、メチル化感受性制限酵素は、メチル化DNAと比べて、非メチル化DNAを優先的に切断する。

30

40

【0224】

部分試料をヌクレアーゼと接触させるとき、1つまたは複数のヌクレアーゼを使用することができる。一部の実施形態では、部分試料を、複数のヌクレアーゼと接触させる。部分試料は、逐次的または同時に、ヌクレアーゼと接触させてもよい。ヌクレアーゼの同時使用は、ヌクレアーゼが、類似の条件（例えば、緩衝液組成）下において活性である場合に、不必要な試料操作を回避するため有益であり得る。第2の部分試料を、1つを上回る

50

メチル化依存性制限酵素と接触させることにより、非特異的に分配された高メチル化DNAをより完全に分解することができる。同様に、第1の部分試料を、1つを上回るメチル化感受性制限酵素と接触させることにより、非特異的に分配された低メチル化DNAおよび／または非メチル化DNAをより完全に分解することができる。

【0225】

一部の実施形態では、メチル化依存性ヌクレアーゼは、MspJI、LpnPI、FspEI、またはMcrBCのちの1つまたは複数を含む。一部の実施形態では、少なくとも2つのメチル化依存性ヌクレアーゼが使用される。一部の実施形態では、少なくとも3つのメチル化依存性ヌクレアーゼが使用される。一部の実施形態では、メチル化依存性ヌクレアーゼは、FspEIを含む。一部の実施形態では、メチル化依存性ヌクレアーゼは、例えば、逐次的に使用される、FspEIおよびMspJIを含む。

10

【0226】

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

20

【0227】

一部の実施形態では、FspEIが、少なくとも1つの部分試料（例えば、低メチル化分配）中の核酸分子を消化するために使用される。一部の実施形態では、BstUI、HpaII、およびHin6Iが、少なくとも1つの部分試料（例えば、高メチル化分配）中の核酸分子を消化するために使用され、FspEIが、少なくとも1つの他の部分試料（例えば、低メチル化分配）中の核酸分子を消化するために使用される。中間メチル化分配を含む実施形態では、その中の核酸分子は、メチル化感受性ヌクレアーゼまたはメチル化依存性ヌクレアーゼで消化され得る。一部の実施形態では、中間メチル化分配中の核酸分子は、高メチル化分配と同じヌクレアーゼで消化される。例えば、中間メチル化分配は、高メチル化分配とともにプールされてもよく、次いで、プールされた分配を、消化に供してもよい。一部の実施形態では、中間メチル化分配中の核酸分子は、低メチル化分配と同じヌクレアーゼで消化される。例えば、中間メチル化分配は、低メチル化分配とともにプールされてもよく、次いで、プールされた分配を、消化に供してもよい。

30

【0228】

一部の実施形態では、部分試料を、DNAの両方の末端へタグ付けするまたはアダプターを付着するステップの後に、上述のようにヌクレアーゼと接触させる。タグまたはアダプターは、上述のアプローチのうちのいずれかを使用したヌクレアーゼによる切断に対して抵抗性であり得る。このアプローチにおいて、切断は、非特異的に分配された分子が解析まで持ち込まれることを防止し得るが、これは、切断産物が、両方の末端においてタグまたはアダプターが欠如しているためである。

40

【0229】

あるいは、タグ付けするまたはアダプターを付着するステップは、上述のようにヌクレアーゼによる切断の後に行うことができる。切断された分子を、次いで、ヌクレアーゼ認識部位に対応する末端（タグまたはアダプターへの付着点）を有することに基づいて、配列読み取りデータ内で同定することができる。このようにして分子をプロセッシングすることによっても、切断された分子、例えば、体細胞突然変異の観察からの情報の獲得が可能

50

となり得る。部分試料をヌクレアーゼと接触させた後にタグ付けするまたはアダプターを付着し、低分子量DNA、例えば、cfDNAを解析する場合、接触させるステップの前に、試料から高分子量DNA（例えば、混入しているゲノムDNA）を除去することが望ましい場合がある。また、変性は後続のライゲーションステップを妨害し得るという点で、比較的低い温度（例えば、65℃もしくはそれ未満、または60℃もしくはそれ未満）で熱不活化することができるヌクレアーゼを使用して、DNAの変性を回避することが望ましい場合もある。

【0230】

試料を、中間メチル化分子を含有する第3の部分試料を含む、3つの部分試料に分配する場合、第3の部分試料は、一部の実施形態では、メチル化感受性ヌクレアーゼと接触させられる。そのようなステップは、接触させるステップと関連する本明細書の他の箇所に記載される特色のうちのいずれかを有し得、上記で考察されるように、タグ付けするまたはアダプターを付着するステップの前または後に行われ得る。一部の実施形態では、第1および第3の部分試料は、メチル化感受性ヌクレアーゼと接触させる前に、組み合わせられる。そのようなステップは、接触させるステップと関連する本明細書の他の箇所に記載される特色のうちのいずれかを有し得、上記で考察されるように、タグ付けするまたはアダプターを付着するステップの前または後に行われ得る。一部の実施形態では、第1および第3の部分試料は、組み合わせる前に、差次的にタグ付けされる。

10

【0231】

あるいは、試料を、中間メチル化分子を含有する第3の部分試料を含む、3つの部分試料に分配する場合、第3の部分試料は、一部の実施形態では、メチル化依存性ヌクレアーゼと接触させられる。そのようなステップは、接触させるステップと関連する本明細書の他の箇所に記載される特色のうちのいずれかを有し得、上記で考察されるように、タグ付けするまたはアダプターを付着するステップの前または後に行われ得る。一部の実施形態では、第2および第3の部分試料は、メチル化依存性ヌクレアーゼと接触させる前に、組み合わせられる。そのようなステップは、接触させるステップと関連する本明細書の他の箇所に記載される特色のうちのいずれかを有し得、上記で考察されるように、タグ付けするまたはアダプターを付着するステップの前または後に行われ得る。一部の実施形態では、第2および第3の部分試料は、組み合わせる前に、差次的にタグ付けされる。

20

【0232】

一部の実施形態では、DNAは、ヌクレアーゼと接触させた後、例えば、SPRIビーズを使用して精製される。そのような精製は、ヌクレアーゼの熱不活化後に行われ得る。あるいは、精製は、省略されてもよく、したがって、例えば、増幅などの後続のステップが、熱不活化されたヌクレアーゼを含有する部分試料に行われ得る。別の実施形態では、接触ステップは、例えば、チューブ移送と関連する消失が最小限となるように、精製試薬、例えば、SPRIビーズの存在下で行われ得る。切断および熱不活化後に、SPRIビーズは、分子クラウディング試薬（例えば、PEG）および塩を添加することによるクリーンアップに再利用され得る。

30

4. 第1の部分試料を、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供する

40

【0233】

本明細書に開示される方法は、第1の部分試料を、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供するステップを含み得、ここで、第1の核酸塩基は、改変または非改変核酸塩基であり、第2の核酸塩基は、第1の核酸塩基とは異なる改変または非改変核酸塩基であり、第1の核酸塩基および第2の核酸塩基は、同じ塩基対形成特異性を有する（が、例えば、第2の部分試料は、本明細書の他の箇所に記載される実施形態のいずれかに従って、メチル化依存性ヌクレアーゼと接触させられる）。一部の実施形態では、第1の核酸塩基が、改変または非改変アデニンである場合には、第2の核酸塩基は、改変または非改変アデニンであり、第1の核酸塩基が、改変または非改変シトシンである場合には、第2の核酸塩基は、改変また

50

は非改変シトシンであり、第1の核酸塩基が、改変または非改変グアニンである場合には、第2の核酸塩基は、改変または非改変グアニンであり、第1の核酸塩基が、改変または非改変チミンである場合には、第2の核酸塩基は、改変または非改変チミンである（改変または非改変ウラシルは、このステップの目的で、改変チミン内に包含される）。そのような手順を使用して、ある特定の改変、例えば、メチル化を有するかまたはそれが欠如した、部分試料中のヌクレオチドを同定することができる。

【0234】

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

【0235】

一部の実施形態では、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順は、バイサルファイト変換を含む。バイサルファイトでの処置により、非改変シトシンおよびある特定の改変シトシンヌクレオチド（例えば、5-ホルミルシトシン（fC）または5-カルボキシルシトシン（caC））を、ウラシルに変換し、一方で他の改変シトシン（例えば、5-メチルシトシン、5-ヒドロキシメチルシトシン）は変換されない。したがって、バイサルファイト変換が使用される場合、第1の核酸塩基は、非改変シトシン、5-ホルミルシトシン、5-カルボキシルシトシン、またはバイサルファイトによって影響を受ける他のシトシン形態のうちの1つまたは複数を含み、第2の核酸塩基は、mCおよびhmCのうちの1つまたは複数、例えば、mCおよび必要に応じてhmCを含み得る。バイサルファイトで処置したDNAのシーケンシングは、シトシンとして読み取られる位置を、mCまたはhmC位置であるとして同定する。一方、Tとして読み取られる位置は、T、またはCのバイサルファイト感受性形態、例えば、非改変シトシン、5-ホルミルシトシン、または5-カルボキシルシトシンであるとして同定される。本明細書に記載されるように第1の部分試料にバイサルファイト変換を行うことは、したがって、第1の部分試料から得られた配列読み取りデータを使用して、mCまたはhmCを含有する位置を同定することを容易にする。バイサルファイト変換の例示的な説明については、例えば、Moss et al., Nat Commun. 2018; 9: 5068を参照されたい。

【0236】

一部の実施形態では、第1の部分試料DNA内の第1の核酸塩基に対してのDNA内の第2の核酸塩基とは異なって影響を及ぼす手順は、酸化的バイサルファイト（Ox-B S）変換を含む。この手順は、まず、hmCを、バイサルファイト感受性であるfCに変換し、続いて、バイサルファイト変換を行う。したがって、酸化的バイサルファイト変換が使用される場合、第1の核酸塩基は、非改変シトシン、fC、caC、hmC、またはバイサルファイトによって影響を受ける他のシトシン形態のうちの1つまたは複数を含み、第2の核酸塩基は、mCを含む。Ox-B Sで変換されたDNAのシーケンシングは、シトシンとして読み取られる位置を、mC位置であるとして同定する。一方、Tとして読み取られる位置は、T、hmC、またはCのバイサルファイト感受性形態、例えば、非改変シトシン、fC、またはhmCであるとして同定される。本明細書に記載されるように第1の部分試料にOx-B S変換を行うことは、したがって、第1の部分試料から得られた配列読み取りデータを使用して、mCを含有する位置を同定することを容易にする。酸化的バイサルファイト変換の例示的な説明については、例えば、Booth et al., Science 2012; 336: 934-937を参照されたい。

【0237】

一部の実施形態では、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順は、Tet補助バイサルファイト (TAB) 変換を含む。TAB変換において、hmCは、変換から保護され、mCは、バイサルファイト処置に先立って酸化され、その結果、もともとmCによって占有されていた位置が、Uに変換され、もともとhmCによって占有されていた位置が、シトシンの保護形態として残る。例えば、Yu et al., Cell 2012; 149: 1368-80に記載されるように、β-グルコシルトランスフェラーゼを使用して、hmCを保護することができ(5-グルコシルヒドロキシメチルシトシン (ghmC) を形成する)、次いで、TETタンパク質、例えば、mTet1を使用して、mCをcaCに変換することができ、次いで、バイサルファイト処置を使用して、CおよびcaCをUに変換することができるが、ghmCは影響を受けないままである。したがって、TAB変換が使用される場合、第1の核酸塩基は、非改変シトシン、fC、caC、mC、またはバイサルファイトによって影響を受ける他のシトシン形態のうちの1つまたは複数を含み、第2の核酸塩基は、hmCを含む。TABで変換されたDNAのシーケンシングは、シトシンとして読み取られる位置を、hmC位置であるとして同定する。一方、Tとして読み取られる位置は、T、mC、またはCのバイサルファイト感受性形態、例えば、非改変シトシン、fC、またはcaCであるとして同定される。本明細書に記載されるように第1の部分試料にTAB変換を行うことは、したがって、第1の部分試料から得られた配列読み取りデータを使用して、hmCを含む位置を同定することを容易にする。

10

20

30

40

50

【0238】

一部の実施形態では、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順は、置換ボラン還元剤を用いたTet補助変換を含み、必要に応じて、置換ボラン還元剤は、2-ピコリンボラン、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランである。置換ボラン還元剤変換を用いたTet補助pic-ボラン変換では、TETタンパク質を使用して、非改変Cに影響を及ぼすことなく、mCおよびhmCをcaCに変換する。caCおよび存在する場合にはfCは、次いで、2-ピコリンボラン (pic-ボラン) または他の置換ボラン還元剤、例えば、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランでの処置によって、これもまた非改変Cに影響を及ぼすことなく、ジヒドロウラシル (DHU) に変換される。例えば、Liu et al., Nature Biotechnology 2019; 37:424-429 (例えば、補足の図1および補足説明7) を参照されたい。DHUは、シーケンシングにおいてTとして読み取られる。したがって、このタイプの変換が使用される場合、第1の核酸塩基は、mC、fC、caC、またはhmCのうちの1つまたは複数を含み、第2の核酸塩基は、非改変シトシンを含む。変換されたDNAのシーケンシングは、シトシンとして読み取られる位置を、非改変C位置であるとして同定する。一方、Tとして読み取られる位置は、T、mC、fC、caC、またはhmCであるとして同定される。本明細書に記載されるように第1の部分試料にTAP変換を行うことは、したがって、第1の部分試料から得られた配列読み取りデータを使用して、非改変Cを含有する位置を同定することを容易にする。この手順は、Liu et al. 2019 (上記) にさらに詳細に記載されているTet補助ピリジンボランシーケンシング (TAPS) を包含する。

【0239】

あるいは、hmCの保護 (例えば、βGTを使用) は、置換ボラン還元剤を用いたTet補助変換と組み合わせることができる。hmCは、βGTを使用したグルコシル化を通じて上述のように保護して、ghmCを形成することができる。TETタンパク質、例えば、mTet1での処置は、次いで、mCをcaCに変換するが、CもghmCも変換しない。caCは、次いで、pic-ボランまたは別の置換ボラン還元剤、例えば、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランでの処置によって、これもまた非改変CまたはghmCに影響を及ぼすことなく、DHUに変換される。したがって、置換ボラン還元剤を用いたTet補助変換が使用される場合、第1の核酸塩基は、mCを含み、第2の核酸塩基は、非改変シトシンまたはhmCのうちの1つまたは複数

、例えば、非改変シトシン、ならびに必要なに応じて h m C、f C、および／または c a C を含む。変換された DNA のシーケンシングは、シトシンとして読み取られる位置を、h m C または非改変 C 位置であるとして同定する。一方、T として読み取られる位置は、T、f C、c a C、または m C であるとして同定される。本明細書に記載されるように第 1 の部分試料に T A P S β 変換を行うことは、したがって、第 1 の部分試料から得られた配列読み取りデータを使用して、一方で非改変の C または h m C を含有する位置を m C を含有する位置と識別することを容易にする。このタイプの変換の例示的な説明については、例えば、Liu et al., Nature Biotechnology 2019; 37:424-429 を参照されたい。
【0240】

一部の実施形態では、第 1 の部分試料の DNA 内の第 1 の核酸塩基に対して DNA 内の第 2 の核酸塩基とは異なって影響を及ぼす手順は、置換ボラン還元剤を用いた化学物質補助変換を含み、必要なに応じて、置換ボラン還元剤は、2-ピコリンボラン、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランである。置換ボラン還元剤を用いた化学物質補助変換において、酸化剤、例えば、過ルテニウム酸カリウム (K R u O₄) (ox-B S 変換における使用にも好適である) を使用して、特異的に h m C を f C に酸化させる。pic-ボランまたは他の置換ボラン還元剤、例えば、ボランピリジン、tert-ブチルアミンボラン、またはアンモニアボランでの処置は、f C および c a C を D H U に変換させるが、m C にも非改変 C にも影響を及ぼさない。したがって、このタイプの変換が使用される場合、第 1 の核酸塩基は、h m C、f C、および c a C のうちの 1 つまたは複数を含み、第 2 の核酸塩基は、非改変シトシンまたは m C のうちの 1 つまたは複数、例えば、非改変シトシンおよび必要なに応じて m C を含む。変換された DNA のシーケンシングは、シトシンとして読み取られる位置を、m C または非改変 C 位置であるとして同定する。一方、T として読み取られる位置は、T、f C、c a C、または h m C であるとして同定される。本明細書に記載されるように第 1 の部分試料にこのタイプの変換を行うことは、したがって、第 1 の部分試料から得られた配列読み取りデータを使用して、一方で非改変の C または m C を含有する位置を h m C を含有する位置と識別することを容易にする。このタイプの変換の例示的な説明については、例えば、Liu et al., Nature Biotechnology 2019; 37:424-429 を参照されたい。

【0241】

一部の実施形態では、第 1 の部分試料の DNA 内の第 1 の核酸塩基に対して DNA 内の第 2 の核酸塩基とは異なって影響を及ぼす手順は、A P O B E C カップリングエピジェネティックス (A C E) 変換を含む。A C E 変換において、A I D / A P O B E C ファミリー DNA デアミナーゼ酵素、例えば、A P O B E C 3 A (A 3 A) を使用して、h m C、f C、または c a C を脱アミノ化することなく、非改変シトシンおよび m C を脱アミノ化する。したがって、A C E 変換が使用される場合、第 1 の核酸塩基は、非改変 C および／または m C (例えば、非改変 C および必要なに応じて m C) を含み、第 2 の核酸塩基は、h m C を含む。A C E 変換 DNA のシーケンシングは、シトシンとして読み取られる位置を、h m C、f C、または c a C 位置であるとして同定する。一方、T として読み取られる位置は、T、非改変 C、または m C であるとして同定される。本明細書に記載されるように第 1 の部分試料に A C E 変換を行うことは、したがって、第 1 の部分試料から得られた配列読み取りデータを使用して、h m C を含有する位置を、m C または非改変 C を含有する位置と識別することを容易にする。A C E 変換の例示的な説明については、例えば、Schutsky et al., Nature Biotechnology 2018; 36: 1083-1090 を参照されたい。

【0242】

一部の実施形態では、第 1 の部分試料の DNA 内の第 1 の核酸塩基に対して DNA 内の第 2 の核酸塩基とは異なって影響を及ぼす手順は、例えば、E M - S e q にあるように、第 1 の核酸塩基の酵素的変換を含む。例えば www.biorxiv.org/content/10.1101/2019.12.20.884692v1 において利用可能な、Vaisvila R, et al. (2019) EM-seq: Detection of DNA methylation at single base resolution from picograms of DNA. bioRxiv; DOI: 10.1101/2019.12.20.884692 を参照

されたい。例えば、T E T 2 および T 4 - β G T を使用して、5 m C および 5 h m C を、デアミナーゼ（例えば、A P O B E C 3 A）によって脱アミノ化することができない基質に変換することができ、次いで、デアミナーゼ（例えば、A P O B E C 3 A）を使用して、非改変シトシンを脱アミノ化してウラシルに変換することができる。

【0243】

一部の実施形態では、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順は、もともと第1の核酸塩基を含むDNAを、もともと第1の核酸塩基を含んでいないDNAから分離することを含む。一部のそのような実施形態では、第1の核酸塩基は、h m C である。もともと第1の核酸塩基を含むDNAは、もともと第1の核酸塩基を含んでいた位置をビオチン化することを含む標識化手順を使用して、他のDNAから分離され得る。一部の実施形態では、第1の核酸塩基は、まず、アジド含有部分、例えば、グルコシルーアジド含有部分で誘導体化される。アジド含有部分は、次いで、例えば、ヒュスゲン環化付加化学反応を通じてビオチンを付着させるための試薬としての機能を果たし得る。次いで、ここでビオチン化されたもともと第1の核酸塩基を含むDNAは、ビオチン結合剤、例えば、アビジン、ニュートラアビジン（約6.3の等電点を有する脱グリコシル化アビジン）、またはストレプトアビジンを使用して、もともと第1の核酸塩基を含まないDNAから分離され得る。もともと第1の核酸塩基を含むDNAを、もともと第1の核酸塩基を含まないDNAから分離するための手順の例は、h m C - シールであり、これは、h m C を標識して、 β - 6 - アジド - グルコシルー5 - ヒドロキシメチルシトシンを形成し、次いで、ヒュスゲン環化付加を通じてビオチン部分を付着させ、続いて、ビオチン結合剤を使用して、ビオチン化DNAを他のDNAから分離する。h m C - シールの例示的な説明については、例えば、Han et al., Mol. Cell 2016; 63: 711-719を参照されたい。このアプローチは、1つまたは複数のh m C 核酸塩基を含む断片を特定するのに有用である。

【0244】

一部の実施形態では、そのような分離に続いて、方法は、もともと第1の核酸塩基を含むDNAの各々を、もともと第1の核酸塩基を含まないDNAおよび第2の部分試料のDNAとは差次的にタグ付けするステップをさらに含む。方法は、差次的タグ付けの後に、もともと第1の核酸塩基を含むDNA、もともと第1の核酸塩基を含まないDNA、および第2の部分試料のDNAをプールするステップをさらに含む。もともと第1の核酸塩基を含むDNA、もともと第1の核酸塩基を含まないDNA、および第2の部分試料のDNAは、次いで、同じシーケンシングセルにおいてシーケンシングされ得るが、差次的タグを使用して、所与の読み取りデータが、もともと第1の核酸塩基を含むDNA、もともと第1の核酸塩基を含まないDNA、または第2の部分試料のDNAの分子に由来するかを解明する能力は保持する。

【0245】

一部の実施形態では、第1の核酸塩基は、改変または非改変アデニンであり、第2の核酸塩基は、改変または非改変アデニンである。一部の実施形態では、改変アデニンは、N⁶ - メチルアデニン（mA）である。一部の実施形態では、改変アデニンは、N⁶ - メチルアデニン（mA）、N⁶ - ヒドロキシメチルアデニン（hmA）、またはN⁶ - ホルミルアデニン（fA）のうちの1つまたは複数である。

【0246】

メチル化DNA免疫沈降（MeDIP）を含む技術を使用して、改変塩基、例えば、mAを含有するDNAを、他のDNAから分離することができる。例えば、Kumar et al., Frontiers Genet. 2018; 9: 640; Greer et al., Cell 2015; 161: 868-878を参照されたい。mAに特異的な抗体は、Sun et al., Bioessays 2015; 37:1155-62に記載されている。様々な改変核酸塩基、例えば、ハロゲン化形態、例えば、5 - ブロモウラシルを含むチミン／ウラシルの形態に対する抗体が、市販されている。様々な改変塩基はまた、その塩基対形成特異性における変化に基づいて検出することができる。例えば、ヒポキサンチンは、脱アミノ化により生じ得るアデニンの改変形態で

あり、シーケンシングにおいてGとして読み取られる。例えば、米国特許第8,486,630号、Brown, Genomes, 2nd Ed., John Wiley & Sons, Inc., New York, N.Y., 2002, chapter 14, "Mutation, Repair, and Recombination"を参照されたい。

5. 富化／捕捉ステップ；増幅、アダプター、バーコード

【0247】

一部の実施形態では、本明細書に開示される方法は、DNA、例えば、cfDNAの標的領域の1つまたは複数のセットを捕捉するステップを含む。捕捉は、当技術分野において公知の任意の好適なアプローチを使用して行われ得る。

【0248】

一部の実施形態では、捕捉するステップは、捕捉しようとするDNAを、標的特異的プローブのセットと接触させることを含む。標的特異的プローブのセットは、上述の実施形態および下記のプロープに関連する節におけるものを含むがこれらに限定されない、標的特異的プローブのセットに関して本明細書に記載される特色のうちのいずれかを有し得る。捕捉するステップは、本明細書において開示される方法の間に調製される1つまたは複数の部分試料に行われ得る。一部の実施形態では、DNAは、少なくとも第1の部分試料または第2の部分試料、例えば、少なくとも第1の部分試料および第2の部分試料から、捕捉される。第1の部分試料が、分離ステップ（例えば、もともと第1の核酸塩基（例えば、hmc）を含むDNAを、もともと第1の核酸塩基を含まないDNAから分離すること、例えば、hmc-シール）を受ける場合、捕捉するステップは、もともと第1の核酸塩基（例えば、hmc）を含むDNA、もともと第1の核酸塩基を含まないDNA、および第2の部分試料のうちのいずれか、いずれか2つ、または全てに対して、行われ得る。一部の実施形態では、部分試料は、差次的にタグ付けされ（例えば、本明細書に記載されるように）、次いで、プールされた後、捕捉を受ける。

【0249】

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

【0250】

一部の実施形態では、本明細書に記載される方法は、複数の標的領域のセットに関して、試験対象から得られたcfDNAを捕捉するステップを含む。標的領域は、エピジェネティック標的領域を含み、これは、それらが腫瘍を起源とするか、健康な細胞を起源とするかに応じて、メチル化レベルおよび／または断片化パターンに相違を示し得る。標的領域はまた、配列可変標的領域を含み、これは、それらが腫瘍を起源とするか、健康な細胞を起源とするかに応じて、配列に相違を示し得る。捕捉するステップは、cfDNA分子の捕捉されたセットを産生し、配列可変標的領域セットに対応するcfDNA分子は、cfDNA分子の捕捉されたセットにおいて、エピジェネティック標的領域セットに対応するcfDNA分子よりも高い捕捉収率で捕捉される。捕捉するステップ、捕捉収率、および関連する態様に関するさらなる考察については、全ての目的に関して参照により本明細書に組み込まれる、WO2020/160414を参照されたい。

【0251】

一部の実施形態では、本明細書に記載される方法は、試験対象から得られたcfDNAを、標的特異的プローブのセットと接触させることを含み、標的特異的プローブのセットは、エピジェネティック標的領域セットに対応するcfDNAよりも高い捕捉収率で配列可変標的領域セットに対応するcfDNAを捕捉するように構成される。

【0252】

配列可変標的領域を十分な信頼度または精度で解析するためには、エピジェネティック標的領域を解析するために必要であり得るよりも大きい深度のシーケンシングが必要であ

10

20

30

40

50

り得ることから、配列可変標的領域セットに対応する c f D N A を、エピジェネティック標的領域セットに対応する c f D N A よりも高い捕捉収率で捕捉することは有益であり得る。断片化パターン（例えば、転写開始部位または C T C F 結合部位の乱れに関して試験するために）または断片の存在量（例えば、高メチル化および低メチル化分配における）を決定するために必要なデータ量は、一般的に、がん関連配列突然変異の存在または非存在を決定するために必要なデータ量よりも少ない。異なる収率で標的領域セットを捕捉することは、同じシーケンシングの実行においてシーケンシングの異なる深度まで標的領域をシーケンシングすることを容易にし得る（例えば、プールした混合物を使用しておよび／または同じシーケンシングセル中で）。

【0253】

10

様々な実施形態では、方法は、本明細書における考察と一貫して、捕捉された c f D N A を、例えば、エピジェネティックおよび配列可変標的領域セットに関して様々な程度のシーケンシング深度までシーケンシングするステップをさらに含む。

【0254】

一部の実施形態では、標的特異的プローブおよび D N A の複合体を、標的特異的プローブに結合していない D N A から分離する。例えば、標的特異的プローブが共有結合または非共有結合によって固体支持体に結合している場合、洗浄または吸引ステップを使用して非結合材料を分離することができる。あるいは、複合体が、非結合材料とは別個のクロマトグラフィー特性を有する場合（例えば、プローブが、クロマトグラフィー樹脂に結合するリガンドを含む場合）、クロマトグラフィーを使用することができる。

20

【0255】

本明細書において他所で詳細に考察するように、標的特異的プローブのセットは、配列可変標的領域セットのプローブおよびエピジェネティック標的領域セットのプローブなどの複数のセットを含み得る。一部のそのような実施形態では、捕捉するステップは、同じ容器で同時に、配列可変標的領域セットのプローブおよびエピジェネティック標的領域セットのプローブについて実施され、例えば配列可変標的領域セットおよびエピジェネティック標的領域セットのプローブは、同じ組成物中にある。このアプローチは、比較的合理化されたワークフローを提供する。一部の実施形態では、配列可変標的領域セットのプローブの濃度は、エピジェネティック標的領域セットのプローブの濃度よりも高い。

【0256】

30

あるいは、捕捉するステップは、第1の容器中で配列可変標的領域プローブセットについて、および第2の容器中でエピジェネティック標的領域プローブセットについて実施されるか、または接触させることは、第1の時間および第1の容器で配列可変標的領域プローブセットについて、ならびに第1の時間の前もしくは後の第2の時間でエピジェネティック標的領域プローブセットについて実施される。このアプローチは、配列可変標的領域セットに対応する捕捉された D N A およびエピジェネティック標的領域セットに対応する捕捉された D N A を含む個別の第1および第2の組成物の調製を可能にする。組成物を、所望のように個別にプロセッシングし（例えば、本明細書において他所で記載されるようにメチル化に基づいて分画するために）、適切な割合で組み換えて、さらなるプロセッシングおよびシーケンシングなどの解析のための材料を提供することができる。

40

【0257】

一部の実施形態では、D N A は増幅される。一部の実施形態では、増幅は、捕捉するステップの前に実施される。一部の実施形態では、増幅は、捕捉するステップの後に実施される。

【0258】

一部の実施形態では、アダプターは、D N A に含まれる。これは、例えば上記のように、例えばプライマーの 5' 部分にアダプターを提供することによって、増幅手順と同時に行われ得る。あるいは、アダプターは、ライゲーションなどの他のアプローチによって付加され得る。

【0259】

50

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

10

【0260】

増幅、タグ、およびバーコードに関するさらなる詳細は、以下の「方法の全般的特色」の節において考察されており、これは、実用的な程度で、前述の実施形態ならびに序説および概要の節に記載される実施形態のいずれかと組み合わせることができる。

6. 捕捉されたセット

【0261】

一部の実施形態では、DNA（例えば、cfDNA）の捕捉されたセットが提供される。開示される方法に関して、DNAの捕捉されたセットは、例えば、本明細書に記載されるように分配するステップの後に捕捉するステップを行うことによって、提供され得る。捕捉されたセットは、配列可変標的領域セット、エピジェネティック標的領域セット、またはこれらの組合せに対応するDNAを含み得る。

20

【0262】

一部の実施形態では、第1の標的領域セットは、少なくともエピジェネティック標的領域を含む、第1の部分試料から捕捉される。第1の部分試料から捕捉されるエピジェネティック標的領域は、高メチル化可変標的領域を含み得る。一部の実施形態では、高メチル化可変標的領域は、健康な対象に由来するcfDNAにおいて、非メチル化であるか、または低いメチル化（例えば、バルクcfDNAと比べて、平均を下回るメチル化）を有するCpGを含有領域である。一部の実施形態では、高メチル化可変標的領域は、健康なcfDNAにおいて、少なくとも1つの他の組織タイプにおけるものよりも低いメチル化を示す、領域である。いずれの特定の理論によっても束縛されることを望むものではないが、がん細胞は、同じ組織タイプの健康な細胞よりも多くのDNAを、血流中に脱落させ得る。そのため、cfDNAの起源の組織の分布は、発癌の際に変化し得る。したがって、第1の部分試料における高メチル化可変標的領域のレベルの増加は、がんの存在（または対象の既往に応じて、再発）を示し得る。

30

【0263】

一部の実施形態では、第2の標的領域セットは、少なくともエピジェネティック標的領域を含む、第2の部分試料から捕捉される。エピジェネティック標的領域は、低メチル化可変標的領域を含み得る。一部の実施形態では、低メチル化可変標的領域は、健康な対象に由来するcfDNAにおいて、メチル化されているか、または高いメチル化（例えば、バルクcfDNAと比べて、平均を上回るメチル化）を有するCpG含有領域である。一部の実施形態では、低メチル化可変標的領域は、健康なcfDNAにおいて、少なくとも1つの他の組織タイプにおけるものよりも高いメチル化を示す、領域である。いずれの特定の理論によっても束縛されることを望むものではないが、がん細胞は、同じ組織タイプの健康な細胞よりも多くのDNAを、血流中に脱落させ得る。そのため、cfDNAの起源の組織の分布は、発癌の際に変化し得る。したがって、第2の部分試料における低メチル化可変標的領域のレベルの増加は、がんの存在（または対象の既往に応じて、再発）を示し得る。

40

【0264】

一部の実施形態では、捕捉された配列可変標的領域DNAの数量は、標的とされる領域のサイズ（フットプリントサイズ）の差に関して正規化した場合、捕捉されたエピジェネ

50

ティック標的領域DNAの数量よりも多い。

【0265】

あるいは、それぞれ、配列可変標的領域セットに対応するDNAおよびエピジェネティック標的領域セットに対応するDNAを含む、第1および第2の捕捉されたセットが提供されてもよい。第1および第2の捕捉されたセットを組み合わせて、組合せの捕捉されたセットを提供してもよい。

【0266】

配列可変標的領域セットおよびエピジェネティック標的領域セットに対応するDNAを含む捕捉されたセットが、上記で考察したように組合せの捕捉されたセットを含む一部の実施形態では、配列可変標的領域セットに対応するDNAは、エピジェネティック標的領域セットに対応するDNAよりも高い濃度、例えば、1.1倍～1.2倍高い濃度、1.2倍～1.4倍高い濃度、1.4倍～1.6倍高い濃度、1.6倍～1.8倍高い濃度、1.8倍～2.0倍高い濃度、2.0倍～2.2倍高い濃度、2.2倍～2.4倍高い濃度、2.4倍～2.6倍高い濃度、2.6倍～2.8倍高い濃度、2.8倍～3.0倍高い濃度、3.0倍～3.5倍高い濃度、3.5倍～4.0倍、4.0倍～4.5倍高い濃度、4.5倍～5.0倍高い濃度、5.0倍～5.5倍高い濃度、5.5倍～6.0倍高い濃度、6.0倍～6.5倍高い濃度、6.5倍～7.0倍高い濃度、7.0倍～7.5倍高い濃度、7.5倍～8.0倍高い濃度、8.0倍～8.5倍高い濃度、8.5倍～9.0倍高い濃度、9.0倍～9.5倍高い濃度、9.5倍～10.0倍高い濃度、10倍～11倍高い濃度、11倍～12倍高い濃度、12倍～13倍高い濃度、13倍～14倍高い濃度、14倍～15倍高い濃度、15倍～16倍高い濃度、16倍～17倍高い濃度、17倍～18倍高い濃度、18倍～19倍高い濃度、19倍～20倍高い濃度、20倍～30倍高い濃度、30倍～40倍高い濃度、40倍～50倍高い濃度、50倍～60倍高い濃度、60倍～70倍高い濃度、70倍～80倍高い濃度、80倍～90倍高い濃度、または90倍～100倍高い濃度で存在し得る。濃度の差の程度は、定義の節において考察されるように、標的領域のフットプリントサイズに関する正規化を説明する。

a. エピジェネティック標的領域セット

【0267】

エピジェネティック標的領域セットは、新生物（例えば、腫瘍またはがん）細胞に由来するDNAと、健康な細胞、例えば、非新生物循環細胞に由来するものとを区別する可能性が高い1つまたは複数のタイプの標的領域を含み得る。そのような領域の例示的なタイプは、本明細書において詳細に考察されている。エピジェネティック標的領域セットはまた、例えば、本明細書に記載されるように、1つまたは複数の対照領域も含み得る。

【0268】

一部の実施形態では、エピジェネティック標的領域セットは、少なくとも100kbp、例えば、少なくとも200kbp、少なくとも300kbp、または少なくとも400kbpのフットプリントを有する。一部の実施形態では、エピジェネティック標的領域セットは、100～200Mbp、例えば、100～200kbp、200～300kbp、300～400kbp、400～500kbp、500～600kbp、600～700kbp、700～800kbp、800～900kbp、900～1,000kbp、1～1.5Mbp、1.5～2Mbp、2～3Mbp、3～4Mbp、4～5Mbp、5～6Mbp、6～7Mbp、7～8Mbp、8～9Mbp、9～10Mbp、または10～20Mbpの範囲のフットプリントを有する。一部の実施形態では、エピジェネティック標的領域セットは、少なくとも20Mbpのフットプリントを有する。

i. 高メチル化可変標的領域

【0269】

一部の実施形態では、エピジェネティック標的領域セットは、1つまたは複数の高メチル化可変標的領域を含む。一般に、高メチル化可変標的領域は、例えば、cfDNA試料において観察されるメチル化のレベルの増加が、（例えば、cfDNAの）試料が新生物細胞、例えば、腫瘍細胞またはがん細胞によって産生されるDNAを含有する可能性の増

加を示す、領域を指す。例えば、腫瘍抑制遺伝子のプロモーターの高メチル化は、反復的に観察されている。例えば、Kang et al., Genome Biol. 18:53 (2017)およびそこに引用されている参考文献を参照されたい。別の例では、上記で考察されるように、高メチル化可変標的領域は、がん性組織において、必ずしも、同じタイプの健康な組織に由来するDNAと比べてメチル化が異なるわけではないが、健康な対象において典型的なcfDNAと比べてメチル化が異なる（例えば、より多くのメチル化を有する）、領域を含み得る。例えば、がんの存在が、細胞死、例えば、がんに対応する組織タイプの細胞のアポトーシスの増加をもたらす場合、そのようながんは、少なくとも部分的に、そのような高メチル化可変標的領域を使用して検出することができる。

【0270】

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

【表1】

表1. CRC研究に基づく例示的な高メチル化標的領域

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

【0271】

一部の実施形態では、高メチル化可変標的領域は、表1に列挙される座のうちの複数、例えば、表1に列挙される座のうちの少なくとも10%、20%、30%、40%、50

%、60%、70%、80%、90%、または100%を含む。例えば、標的領域として含まれる各座に関して、遺伝子の転写開始部位と終止コドン（選択的にスプライシングされる遺伝子の最後の終止コドン）との間または遺伝子のプロモーター領域に結合するハイブリダイゼーション部位を有する1つまたは複数のプローブが存在し得る。一部の実施形態では、1つまたは複数のプローブは、表1における遺伝子の転写開始部位から300bp以内、例えば、200または100bp以内に結合する。

【0272】

様々なタイプの肺がんにおけるメチル化可変標的領域は、例えば、Ooki et al., Clin. Cancer Res. 23:7141-52 (2017)、Belinsky, Annu. Rev. Physiol. 77:453-74 (2015)、Hulbert et al., Clin. Cancer Res. 23:1998-2005 (2017)、Shi et al., BMC Genomics 18:901 (2017)、Schneider et al., BMC Cancer. 11:102 (2011)、Lissa et al., Transl Lung Cancer Res 5(5):492-504 (2016)、Skvortsova et al., Br. J. Cancer. 94(10):1492-1495 (2006)、Kim et al., Cancer Res. 61:3419-3424 (2001)、Furonaka et al., Pathology International 55:303-309 (2005)、Gomes et al., Rev. Port. Pneumol. 20:20-30 (2014)、Kim et al., Oncogene. 20:1765-70 (2001)、Hopkins-Donaldson et al., Cell Death Differ. 10:356-64 (2003)、Kikuchi et al., Clin. Cancer Res. 11:2954-61 (2005)、Heller et al., Oncogene 25:959-968 (2006)、Licchesi et al., Carcinogenesis. 29:895-904 (2008)、Guo et al., Clin. Cancer Res. 10:7917-24 (2004)、Palmisano et al., Cancer Res. 63:4620-4625 (2003)、およびToyooka et al., Cancer Res. 61:4556-4560, (2001)において詳細に考察されている。

【0273】

肺がん研究に基づく高メチル化可変標的領域の例示的なセットは、表2に提供されている。これらの遺伝子の多くは、肺がん以外のがんにも関連性を有する可能性が高く、例えば、Casp8（カスパーゼ8）は、プログラム細胞死における鍵となる酵素であり、この遺伝子の高メチル化に基づく不活化は、肺がんに限定されない一般的な発癌機序であり得る。さらに、いくつかの遺伝子は、表1および2の両方に出現し、一般性を示す。

10

20

30

40

50

【表 2 - 1】

表2. 肺がん研究に基づく例示的な高メチル化標的領域

遺伝子名	染色体
MARCH11	chr5
TAC1	chr7
TCF21	chr6
SHOX2	chr3
p16	chr3
Casp8	chr2
CDH13	chr16
MGMT	chr10
MLH1	chr3
MSH2	chr2
TSLC1	chr11
APC	chr5
DKK1	chr10
DKK3	chr11
LKB1	chr11
WIF1	chr12

10

20

【表 2 - 2】

RUNX3	chr1
GATA4	chr8
GATA5	chr20
PAX5	chr9
E-カドヘリン	chr16
H-カドヘリン	chr16

30

【0274】

表 2 において同定される標的領域に関する前述の実施形態のうちのいずれかを、表 1 において同定される標的領域に関して上述された実施形態のうちのいずれかと組み合わせてもよい。一部の実施形態では、高メチル化可変標的領域は、表 1 または表 2 に列挙される座のうちの複数、例えば、表 1 または表 2 に列挙される座のうちの少なくとも 10 %、20 %、30 %、40 %、50 %、60 %、70 %、80 %、90 %、または 100 % を含む。

40

【0275】

さらなる高メチル化標的領域は、例えば、Cancer Genome Atlas から得ることができる。Kang et al., Genome Biology 18:53 (2017) は、乳房、結腸、腎臓、肝臓、および肺に由来する高メチル化標的領域を使用した Cancer Locator と称される確率的方法の構築について記載している。一部の実施形態では、

50

高メチル化標的領域は、がんの1つまたは複数のタイプに特異的であり得る。したがって、一部の実施形態では、高メチル化標的領域は、乳がん、結腸がん、腎臓がん、肝臓がん、および肺がんのうちの1つ、2つ、3つ、4つ、または5つにおいて高メチル化を集合的に示す高メチル化標的領域の1つ、2つ、3つ、4つ、または5つのサブセットを含む。

【0276】

一部の実施形態では、異なるエピジェネティック標的領域が、第1および第2の部分試料から捕捉される場合、第1の部分試料から捕捉されるエピジェネティック標的領域は、高メチル化可変標的領域を含む。

i i . 低メチル化可変標的領域

【0277】

全体的な低メチル化は、様々ながんにおいて共通して観察される現象である。例えば、Hon et al., *Genome Res.* 22:246-258 (2012) (乳がん)、Ehrlich, *Epi genomics* 1:239-259 (2009) (結腸がん、卵巣がん、前立腺がん、白血病、肝細胞がん、および子宮頸がんにおける低メチル化の観察に言及した総説論文)を参照されたい。例えば、反復エレメント、例えば、LINE1エレメント、Aluエレメント、セントロメアタンデムリピート、ペリセントロメアタンデムリピート、およびサテライトDNAなどの領域、ならびに健康な細胞において通常メチル化されている遺伝子間領域は、腫瘍細胞において、低減されたメチル化を示し得る。したがって、一部の実施形態では、エピジェネティック標的領域セットは、観察されるメチル化のレベルの減少が、(例えば、cfDNAの)試料が新生物細胞、例えば、腫瘍細胞またはがん細胞によって産生されるDNAを含有する可能性の増加を示す、低メチル化可変標的領域を含む。別の例では、上記で考察されるように、低メチル化可変標的領域は、がん性組織において、必ずしも、同じタイプの健康な組織に由来するDNAと比べてメチル化が異なるわけではないが、健康な対象において典型的なcfDNAと比べてメチル化が異なる(例えば、あまり非メチル化である)、領域を含み得る。例えば、がんの存在が、細胞死、例えば、がんに対応する組織タイプの細胞のアポトーシスの増加をもたらす場合、そのようながんは、少なくとも部分的に、そのような低メチル化可変標的領域を使用して検出することができる。

【0278】

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

【0279】

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

【0280】

一部の実施形態では、異なるエピジェネティック標的領域が、第1および第2の部分試料から捕捉される場合、第2の部分試料から捕捉されるエピジェネティック標的領域は、低メチル化可変標的領域を含む。

i i i . CTCF 結合領域

【0281】

CTCFは、クロマチン構成に寄与し、しばしば、コヒーシと共局在化する、DNA結合タンパク質である。CTCF結合部位の乱れが、様々な異なるがんにおいて報告されている。例えば、Katainen et al., *Nature Genetics*, doi:10.1038/ng.3335, published online 8 June 2015、Guo et al., *Nat. Commun.* 9:1520 (2018)を参照されたい。CTCF結合は、シーケンシングによって、例えば、断片長解析を通じて検出することができる、cfDNAにおける認識可能なパターンをもたら

10

20

30

40

50

す。シーケンシングに基づく断片長解析に関する詳細は、Snyder et al., Cell 164:57-68 (2016)、WO2018/009723、およびUS20170211143A1において提供されており、これらのそれぞれは、参照により本明細書に組み込まれる。

【0282】

したがって、CTCF結合の乱れは、cfDNAの断片化パターンに変動をもたらす。そのため、CTCF結合部位は、断片化可変標的領域の1つのタイプを表す。

【0283】

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

【0284】

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

【0285】

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

i v . 転写開始部位

【0286】

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

【0287】

したがって、転写開始部位の乱れもまた、cfDNAの断片化パターンに変動をもたらす。そのため、転写開始部位もまた、断片化可変標的領域の1つのタイプを表す。

【0288】

ヒト転写開始部位は、btss.hgc.jpにおいてインターネット上で利用可能なDBTSS

(Data Base of Human Transcription Start Sites) から利用可能であり、Yamashita et al., Nucleic Acids Res. 34(Database issue): D86-D89 (2006)において記載されており、これは、参照により本明細書に組み込まれる。

【0289】

したがって、一部の実施形態では、エピジェネティック標的領域セットは、転写開始部位を含む。一部の実施形態では、転写開始部位は、少なくとも10個、20個、50個、100個、200個、もしくは500個の転写開始部位、または10～20個、20～50個、50～100個、100～200個、200～500個、もしくは500～1000個の転写開始部位、例えば、DBTSSにおいて列挙されている転写開始部位などを含む。一部の実施形態では、転写開始部位のうちの少なくともいくつかは、メチル化されていても、メチル化されていなくてもよく、ここで、メチル化状態は、細胞ががん細胞であるかそうでないかと相関する。一部の実施形態では、エピジェネティック標的領域セットは、転写開始部位の少なくとも100bp、少なくとも200bp、少なくとも300bp、少なくとも400bp、少なくとも500bp、少なくとも750bp、少なくとも1000bp上流および下流の領域を含む。

10

v. 局所増幅

【0290】

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

20

vi. メチル化対照領域

【0291】

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

30

b. 配列可変標的領域セット

【0292】

一部の実施形態では、配列可変標的領域セットは、がんにおいて体細胞突然変異を受けることが公知の複数の領域を含む。

40

【0293】

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

50

れる配列読み取りデータ深度または配列読み取りデータカバレッジを達成するように選択され得る。パネルは、試料中の1つまたは複数の遺伝的変異体を検出するための理論上の感度、理論上の特異度、および／または理論上の正確度を達成するように選択され得る。

【0294】

領域のパネルを検出するためのプローブとしては、目的のゲノム領域（ホットスポット領域）を検出するためのもの、ならびにヌクレオソーム認識プローブ（例えば、KRASコドン12および13）を挙げることができ、ヌクレオソーム結合パターンおよびGC配列組成によって影響を受けるcfDNAカバレッジおよび断片サイズ変動の解析に基づいて捕捉を最適化するように設計され得る。本明細書において使用される領域としては、ヌクレオソーム位置およびGCモデルに基づいて最適化された非ホットスポット領域も含まれ得る。

10

【0295】

目的のゲノムの場所の一覧の例は、表3および表4に見出すことができる。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表3の遺伝子のうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、または70個の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表3のSNVのうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、または70個を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表3の融合体のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、または6個を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表3のインデルのうちの少なくとも1個、少なくとも2個、または3個の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表4の遺伝子のうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、少なくとも70個、または73個の少なくとも一部分を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表4のSNVのうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、少なくとも70個、または73個を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表4の融合体のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、または6個を含む。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは表4のインデルのうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくとも16個、少なくとも17個、または18個の少なくとも一部分を含む。これらの目的のゲノムの場所の各々は、所与のパネルの骨格領域またはホットスポット領域として同定され得る。目的のホットスポットゲノムの場所の一覧の例は、表5に見出すことができる。一部の実施形態では、本開示の方法において使用される配列可変標的領域セットは、表5の遺伝子のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくと

20

30

40

50

も 16 個、少なくとも 17 個、少なくとも 18 個、少なくとも 19 個、または少なくとも 20 個の少なくとも一部分を含む。各ホットスポットゲノム領域は、関連する遺伝子、それが存在する染色体、遺伝子の座を表すゲノムの開始位置および終止位置、塩基対における遺伝子の座の長さ、遺伝子によってカバーされるエクソン、および目的の所与のゲノム領域が捕捉しようとし得る重要な特色（例えば、突然変異のタイプ）を含む、いくつかの特徴とともに列挙されている。

【表 3】

表3

点突然変異(SNV)						融合体
AKT1	ALK	APC	AR	ARAF	ARID1A	ALK
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	FGFR2
CCNE1	CDH1	CDK4	CDK6	CDKN2A	CDKN2B	FGFR3
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	NTRK1
FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	RET
GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	ROS1
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	
PTPN11	RAF1	RB1	RET	RHEB	RHOA	
RIT1	ROS1	SMAD4	SMO	SRC	STK11	
TERT	TP53	TSC1	VHL			

10

20

【表 4 - 1】

表4

点突然変異(SNV)						融合体
AKT1	ALK	APC	AR	ARAF	ARID1A	ALK
ATM	BRAF	BRCA1	BRCA2	CCND1	CCND2	FGFR2
CCNE1	CDH1	CDK4	CDK6	CDKN2A	DDR2	FGFR3
CTNNB1	EGFR	ERBB2	ESR1	EZH2	FBXW7	NTRK1

【表 4 - 2】

FGFR1	FGFR2	FGFR3	GATA3	GNA11	GNAQ	RET
GNAS	HNF1A	HRAS	IDH1	IDH2	JAK2	ROS1
JAK3	KIT	KRAS	MAP2K1	MAP2K2	MET	
MLH1	MPL	MYC	NF1	NFE2L2	NOTCH1	
NPM1	NRAS	NTRK1	PDGFRA	PIK3CA	PTEN	
PTPN11	RAF1	RB1	RET	RHEB	RHOA	
RIT1	ROS1	SMAD4	SMO	MAPK1	STK11	
TERT	TP53	TSC1	VHL	MAPK3	MTOR	
NTRK3						

30

40

50

【表 5 - 1】

表 5

遺伝子	染色体	開始位置	終止位置	長さ(bp)	カバーされるエクソン	重要な特色
ALK	chr2	29446405	29446655	250	イントロン 19	融合体
ALK	chr2	29446062	29446197	135	イントロン 20	融合体
ALK	chr2	29446198	29446404	206	20	融合体
ALK	chr2	29447353	29447473	120	イントロン 19	融合体
ALK	chr2	29447614	29448316	702	イントロン 19	融合体
ALK	chr2	29448317	29448441	124	19	融合体
ALK	chr2	29449366	29449777	411	イントロン 18	融合体
ALK	chr2	29449778	29449950	172	18	融合体
BRAF	chr7	140453064	140453203	139	15	BRAF V600
CTNNB1	chr3	41266007	41266254	247	3	S37
EGFR	chr7	55240528	55240827	299	18 および 19	G719 および欠失
EGFR	chr7	55241603	55241746	143	20	挿入/T790M
EGFR	chr7	55242404	55242523	119	21	L858R
ERBB2	chr17	37880952	37881174	222	20	挿入
ESR1	chr6	152419857	152420111	254	10	V534, P535, L536, Y537, D538
FGFR2	chr10	123279482	123279693	211	6	S252
GATA3	chr10	8111426	8111571	145	5	SS /インデル
GATA3	chr10	8115692	8116002	310	6	SS /インデル
GNAS	chr20	57484395	57484488	93	8	R844
IDH1	chr2	209113083	209113394	311	4	R132
IDH2	chr15	90631809	90631989	180	4	R140, R172
KIT	chr4	55524171	55524258	87	1	
KIT	chr4	55561667	55561957	290	2	
KIT	chr4	55564439	55564741	302	3	

10

20

30

40

【表 5 - 2】

KIT	chr4	55565785	55565942	157	4	
KIT	chr4	55569879	55570068	189	5	
KIT	chr4	55573253	55573463	210	6	
KIT	chr4	55575579	55575719	140	7	
KIT	chr4	55589739	55589874	135	8	
KIT	chr4	55592012	55592226	214	9	
KIT	chr4	55593373	55593718	345	10 および 11	557, 559, 560, 576
KIT	chr4	55593978	55594297	319	12 および 13	V654
KIT	chr4	55595490	55595661	171	14	T670, S709
KIT	chr4	55597483	55597595	112	15	D716
KIT	chr4	55598026	55598174	148	16	L783
KIT	chr4	55599225	55599368	143	17	C809, R815, D816, L818, D820, S821F, N822, Y823
KIT	chr4	55602653	55602785	132	18	A829P
KIT	chr4	55602876	55602996	120	19	
KIT	chr4	55603330	55603456	126	20	
KIT	chr4	55604584	55604733	149	21	
KRAS	chr12	25378537	25378717	180	4	A146
KRAS	chr12	25380157	25380356	199	3	Q61
KRAS	chr12	25398197	25398328	131	2	G12/G13
MET	chr7	116411535	116412255	720	13, 14, イ ントロン 13, イント ロン 14	MET エクソン 14 SS
NRAS	chr1	115256410	115256609	199	3	Q61
NRAS	chr1	115258660	115258791	131	2	G12/G13
PIK3CA	chr3	178935987	178936132	145	10	E545K
PIK3CA	chr3	178951871	178952162	291	21	H1047R
PTEN	chr10	89692759	89693018	259	5	R130
SMAD4	chr18	48604616	48604849	233	12	D537
TERT	chr5	1294841	1295512	671	プロモータ ー	chr5:1295228
TP53	chr17	7573916	7574043	127	11	Q331, R337, R342
TP53	chr17	7577008	7577165	157	8	R273
TP53	chr17	7577488	7577618	130	7	R248
TP53	chr17	7578127	7578299	172	6	R213/Y220

10

20

30

40

50

【表 5 - 3】

TP53	chr17	7578360	7578564	204	5	R175 / 欠失
TP53	chr17	7579301	7579600	299	4	
				12574 (全標的領域)		
				16330 (全プローブカバレッジ)		

10

【0296】

さらに、またはあるいは、好適な標的領域セットは、文献から入手可能である。例えば、参照により本明細書に組み込まれる、Gale et al., PLoS One 13: e0194630 (2018)は、配列可変標的領域セットの一部または全てとして使用することができる35個のがん関連遺伝子標的のパネルを記載する。これらの35個の標的は、AKT1、ALK、BRAF、CCND1、CDK2A、CTNNB1、EGFR、ERBB2、ESR1、FGFR1、FGFR2、FGFR3、FOXL2、GATA3、GNA11、GNAQ、GNAS、HRAS、IDH1、IDH2、KIT、KRAS、MED12、MET、MYC、NFE2L2、NRAS、PDGFRA、PIK3CA、PPP2R1A、PTEN、RET、STK11、TP53、およびU2AF1である。

20

【0297】

一部の実施形態では、配列可変標的領域セットは、上に列挙されたがん関連遺伝子などの少なくとも10、20、30、または35個のがん関連遺伝子からの標的領域を含む。

【0298】

一部の実施形態では、配列可変標的領域セットは、少なくとも50kbp、例えば、少なくとも100kbp、少なくとも200kbp、少なくとも300kbp、または少なくとも400kbpのフットプリントを有する。一部の実施形態では、配列可変標的領域セットは、100~2000kbp、例えば、100~200kbp、200~300kbp、300~400kbp、400~500kbp、500~600kbp、600~700kbp、700~800kbp、800~900kbp、900~1,000kbp、1~1.5Mbp、または1.5~2Mbpの範囲のフットプリントを有する。一部の実施形態では、配列可変標的領域セットは、少なくとも2Mbpのフットプリントを有する。

30

7. 対象

【0299】

一部の実施形態では、DNA（例えば、cfDNA）は、がんを有する対象から得られる。一部の実施形態では、DNA（例えば、cfDNA）は、がんを有することが疑われる対象から得られる。一部の実施形態では、DNA（例えば、cfDNA）は、腫瘍を有する対象から得られる。一部の実施形態では、DNA（例えば、cfDNA）は、腫瘍を有することが疑われる対象から得られる。一部の実施形態では、DNA（例えば、cfDNA）は、新生物を有する対象から得られる。一部の実施形態では、DNA（例えば、cfDNA）は、新生物を有することが疑われる対象から得られる。一部の実施形態では、DNA（例えば、cfDNA）は、（例えば、化学療法、外科的切除、放射線照射、またはこれらの組合せの後に）腫瘍、がん、または新生物から寛解した対象から得られる。前述の実施形態のいずれかでは、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、肺、結腸、直腸、腎臓、乳房、前立腺、または肝臓のものであり得る。一部の実施形態では、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、肺のものである。一部の実施形態では、がん、腫瘍、もしくは新生物、

40

50

または疑われるがん、腫瘍、もしくは新生物は、結腸または直腸のものである。一部の実施形態では、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、乳房のものである。一部の実施形態では、がん、腫瘍、もしくは新生物、または疑われるがん、腫瘍、もしくは新生物は、前立腺のものである。前述の実施形態のいずれかでは、対象は、ヒト対象であり得る。

8. 定量

【0300】

一部の実施形態では、第1の部分試料、処置された第1の部分試料、または処置された第2の部分試料のうちの1つまたは複数から捕捉されるエピジェネティック標的領域は、定量される。例えば、低メチル化可変標的領域が、処置された第2の部分試料において定10
量され得る、かつ／または高メチル化可変標的領域が、第1の部分試料もしくは処置された第1の部分試料において定量され得る。定量は、任意の適切な技術、例えば、定量的増幅、例えば、定量的PCRによるものであり得る。一部の実施形態では、定量は、シーケンシングデータ（例えば、シーケンシング読み取りデータの数またはシーケンシングされる一意的分子の数）に基づく。

【0301】

上記で考察されるようなエピジェネティック標的領域の定量は、対象におけるがんの存在、非存在、または可能性を決定するために使用することができる。例えば、がんの存在または非存在の決定は、少なくとも部分的に、第1の部分試料もしくは処置された第1の部分試料における高メチル化可変標的領域の量、および／または処置された第2の部分試10
料における低メチル化可変標的領域の量が、所定の閾値を超えるかどうかに基づき得る。一部の実施形態では、そのような量は、試料から収集された他のデータ、例えば、本明細書の他の箇所に記載される突然変異および／または他のエピジェネティック特色の存在、例えば、転写開始部位および／またはCTCF結合部位の乱れと一緒に使用することができる。

9. 第1および第2の部分試料またはその一部分に由来するDNAのプール

【0302】

一部の実施形態では、方法は、第2の部分試料のDNAの少なくとも一部分（低メチル化分配とも称される）および第1の部分試料のDNAの少なくとも一部分（高メチル化分配とも称される）を含むプールを調製するステップを含む。例えば、エピジェネティック10
標的領域および／または配列可変標的領域を含む、標的領域は、プールから捕捉され得る。本明細書の他の箇所に記載される部分試料の少なくとも一部分から標的領域セットを捕捉するステップは、第1および第2の部分試料に由来するDNAを含むプールに対して行われる捕捉ステップを包含する。プール内のDNAを増幅させるステップは、プールから標的領域を捕捉する前に行われ得る。捕捉するステップは、本明細書の他の箇所に記載される特色のうちのいずれかを有し得る。

【0303】

エピジェネティック標的領域は、本明細書の他の箇所で考察されるように、それらが、腫瘍を起源とするかもしくは健康な細胞を起源とするか、またはどの組織タイプを起源とするかに応じて、メチル化レベルおよび／または断片化パターンに相違を示し得る。配列10
可変標的領域は、それらが腫瘍を起源とするか、健康な細胞を起源とするかに応じて、配列に相違を示し得る。

【0304】

低メチル化分配に由来するエピジェネティック標的領域の解析は、一部の適用において、高メチル化および低メチル化分配に由来する配列可変標的領域ならびに高メチル化分配に由来するエピジェネティック標的領域の解析よりも、情報量が少ない場合がある。そのため、配列可変標的領域およびエピジェネティック標的領域を捕捉する方法では、後者は、高メチル化および低メチル化分配に由来する配列可変標的領域ならびに高メチル化分配10
に由来するエピジェネティック標的領域のうちの1つまたは複数よりも、低い程度で捕捉され得る。例えば、配列可変標的領域は、高メチル化分配とともにプールされていない低

10

20

30

40

50

メチル化分配の部分から捕捉され得、プールは、高メチル化分配に由来するDNAの一部（例えば、大部分、実質的に全て、または全て）および低メチル化分配に由来するDNAなしまたはその一部（例えば、少数）を用いて調製され得る。そのようなアプローチにより、低メチル化分配に由来するエピジェネティック標的領域のシーケンシングを低減または排除し、それによって、さらなる解析にとって十分であるシーケンシングデータの量を低減することができる。

【0305】

一部の実施形態では、プール内の低メチル化分配のDNAのうちの少数部分を含めることで、1つまたは複数のエピジェネティック特色（例えば、メチル化または本明細書の他の箇所において考察されている他のエピジェネティック特色）の定量が、例えば、相対的に、容易になる。

10

【0306】

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

20

【0307】

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

30

【0308】

一部の実施形態では、方法は、低メチル化分配のDNAのうちの少なくとも一部分を含む、第1のプールを調製するステップを含む。一部の実施形態では、方法は、高メチル化分配のDNAのうちの少なくとも一部分を含む、第2のプールを調製するステップを含む。一部の実施形態では、第1のプールは、高メチル化分配のDNAのうちの一部分をさらに含む。一部の実施形態では、第2のプールは、低メチル化分配のDNAのうちの一部分をさらに含む。一部の実施形態では、第1のプールは、低メチル化分配のDNAのうちの大部分を含み、必要に応じて、高メチル化分配のDNAのうちの少数部分を含む。一部の実施形態では、第2のプールは、高メチル化分配のDNAのうちの大部分を含み、低メチル化分配のDNAのうちの少数部分を含む。中間メチル化分配を含む一部の実施形態では、第2のプールは、中間メチル化分配のDNAのうちの少なくとも一部分、例えば、中間メチル化分配のDNAのうちの大部分を含む。一部の実施形態では、第1のプールは、低メチル化分配のDNAのうちの大部分を含み、第2のプールは、高メチル化分配のDNAのうちの大部分および中間メチル化分配のDNAのうちの大部分を含む。

40

【0309】

一部の実施形態では、方法は、第1のプールから少なくとも標的領域の第1のセットを捕捉するステップを含み、例えば、第1のプールは、上述の実施形態のうちのいずれかに記載される通りである。一部の実施形態では、第1のセットは、配列可変標的領域を含む

50

。一部の実施形態では、第1のセットは、低メチル化可変標的領域および／または断片化可変標的領域を含む。一部の実施形態では、第1のセットは、配列可変標的領域および断片化可変標的領域を含む。一部の実施形態では、第1のセットは、配列可変標的領域、低メチル化可変標的領域、および断片化可変標的領域を含む。第1のプール内のDNAを増幅させるステップは、この捕捉ステップの前に行われ得る。一部の実施形態では、第1のプールから標的領域の第1のセットを捕捉するステップは、第1のプールのDNAを、標的特異的プローブの第1のセットと接触させることを含む。一部の実施形態では、標的特異的プローブの第1のセットは、配列可変標的領域に特異的な標的結合プローブを含む。一部の実施形態では、標的特異的プローブの第1のセットは、配列可変標的領域、低メチル化可変標的領域、および／または断片化可変標的領域に特異的な標的結合プローブを含む。

10

【0310】

一部の実施形態では、方法は、第2のプールから、標的領域の第2のセットまたは標的領域の複数のセットを捕捉するステップを含み、例えば、第1のプールは、上述の実施形態のうちのいずれかに記載される通りである。一部の実施形態では、第2の複数物は、エピジェネティック標的領域、例えば、高メチル化可変標的領域および／または断片化可変標的領域を含む。一部の実施形態では、第2の複数物は、配列可変標的領域およびエピジェネティック標的領域、例えば、高メチル化可変標的領域および／または断片化可変標的領域を含む。第2のプール内のDNAを増幅させるステップは、この捕捉ステップの前に行われ得る。一部の実施形態では、第2のプールから標的領域の第2の複数のセットを捕捉するステップは、第1のプールのDNAを、標的特異的プローブの第2のセットと接触させることを含み、標的特異的プローブの第2のセットは、配列可変標的領域に特異的な標的結合プローブおよびエピジェネティック標的領域に特異的な標的結合プローブを含む。一部の実施形態では、標的領域の第1のセットおよび標的領域の第2のセットは、同一ではない。例えば、標的領域の第1のセットは、標的領域の第2のセットには存在しない1つまたは複数の標的領域を含み得る。あるいは、または追加として、標的領域の第2のセットは、標的領域の第1のセットには存在しない1つまたは複数の標的領域を含み得る。一部の実施形態では、少なくとも1つの高メチル化可変標的領域は、第1のプールからではなく、第2のプールから捕捉される。一部の実施形態では、複数の高メチル化可変標的領域は、第1のプールからではなく、第2のプールから捕捉される。一部の実施形態では、標的領域の第1のセットは、配列可変標的領域を含み、および／または標的領域の第2のセットは、エピジェネティック標的領域を含む。一部の実施形態では、標的領域の第1のセットは、配列可変標的領域および断片化可変標的領域を含み、標的領域の第2のセットは、エピジェネティック標的領域、例えば、高メチル化可変標的領域および断片化可変標的領域を含む。一部の実施形態では、標的領域の第1のセットは、配列可変標的領域、断片化可変標的領域を含み、および低メチル化可変標的領域を含み、標的領域の第2のセットは、エピジェネティック標的領域、例えば、高メチル化可変標的領域、および断片化可変標的領域を含む。

20

30

【0311】

一部の実施形態では、第1のプールは、低メチル化分配のDNAのうちの大部分、および高メチル化分配のDNAのうち的一部分（例えば、約半分）を含み、第2のプールは、高メチル化分配のDNAのうち的一部分（例えば、約半分）を含む。一部のそのような実施形態では、標的領域の第1のセットは、配列可変標的領域を含み、および／または標的領域の第2のセットは、エピジェネティック標的領域を含む。配列可変標的領域および／またはエピジェネティック標的領域は、本明細書の他の箇所に記載される実施形態のうちのいずれかに記載される通りであり得る。

40

10. シーケンシング

【0312】

一般に、以前の増幅の有無にかかわらず、アダプターが隣接している試料核酸は、シーケンシングに供することができる。シーケンシング方法としては、例えば、サンガーシー

50

ケンシング、ハイスルーブ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、一分子シーケンシング、ナノポアシーケンシング、半導体シーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシング、デジタル遺伝子発現 (Helicos)、次世代シーケンシング (NGS)、合成による一分子シーケンシング (SMSS) (Helicos)、大規模並列シーケンシング、クロン単一分子アレイ (Solexa)、ショットガンシーケンシング、Ion Torrent、Oxford Nanopore、Roche Genia、マキサムーギルバートシーケンシング、プライマーウォーキング、および PacBio、SOLID、Ion Torrent、または Nanopore プラットフォームを使用したシーケンシングが挙げられる。シーケンシング反応は、様々な試料プロセッシングユニットにおいて行うことができ、これは、複数のレーン、複数のチャンネル、複数のウェル、または複数の試料セットを実質的に同時にプロセッシングする他の手段を含み得る。試料プロセッシングユニットにはまた、複数の実行のプロセッシングを同時に可能にする複数の試料チャンバーが含まれ得る。

10

【0313】

一部の実施形態では、シーケンシングするステップは、捕捉された標的領域のセットを含むライブラリーに行われ、これは、本明細書に記載される標的領域セットのうちのいずれかを含み得る。一部の実施形態では、シーケンシングするステップは、捕捉／富化を受けていない部分試料（例えば、全ゲノム部分試料）を含むライブラリーに行われる。例えば、標的領域は、第1の部分試料および第2の試料から捕捉され、次いでシーケンシングされ得るか、または標的領域は、第1の部分試料から捕捉され、接触させるステップおよびタグ付けするステップなどのプロセッシングの後に、第2の部分試料と組み合わせられ得るか、または標的領域は、第2の部分試料から捕捉され、接触させるステップおよびタグ付けするステップなどのプロセッシングの後に、第1の部分試料と組み合わせられ得るか、または第1および第2の部分試料の両方が、捕捉／富化を受けることなく、プロセッシングされ組み合わせられ得る。

20

【0314】

シーケンシング反応は、少なくとも1つががんまたは他の疾患のマーカーを含有することが公知である、核酸の1つまたは複数の形態に行うことができる。シーケンシング反応はまた、試料中に存在する任意の核酸断片に行うこともできる。一部の実施形態では、ゲノムの配列カバレッジは、5%を下回る、10%を下回る、15%を下回る、20%を下回る、25%を下回る、30%を下回る、40%を下回る、50%を下回る、60%を下回る、70%を下回る、80%を下回る、90%を下回る、95%を下回る、99%を下回る、99.9%を下回る、または100%を下回り得る。一部の実施形態では、配列反応は、ゲノムの少なくとも5%、10%、15%、20%、25%、30%、40%、50%、60%、70%、または80%の配列カバレッジを提供し得る。配列カバレッジは、少なくとも5個、10個、20個、70個、100個、200個、もしくは500個の異なる遺伝子、または多くとも5000個、2500個、1000個、500個、もしくは100個の異なる遺伝子に行うことができる。

30

【0315】

多重シーケンシングを使用して、同時のシーケンシング反応を行ってもよい。一部の場合には、無細胞核酸は、少なくとも1000回、2000回、3000回、4000回、5000回、6000回、7000回、8000回、9000回、10000回、50000回、100,000回のシーケンシング反応でシーケンシングされ得る。他の場合には、無細胞核酸は、1000回を下回る、2000回を下回る、3000回を下回る、4000回を下回る、5000回を下回る、6000回を下回る、7000回を下回る、8000回を下回る、9000回を下回る、10000回を下回る、50000回を下回る、100,000回を下回るシーケンシング反応でシーケンシングされ得る。シーケンシング反応は、逐次的に、または同時に行われてもよい。後続のデータ解析は、シーケンシング反応の全てまたは一部で行われ得る。一部の場合には、データ解析は、少なくとも1

40

50

000回、2000回、3000回、4000回、5000回、6000回、7000回、8000回、9000回、10000回、50000回、100,000回のシーケンシング反応で行われ得る。他の場合には、データ解析は、1000回を下回る、2000回を下回る、3000回を下回る、4000回を下回る、5000回を下回る、6000回を下回る、7000回を下回る、8000回を下回る、9000回を下回る、10000回を下回る、50000回を下回る、100,000回を下回るシーケンシング反応で行われ得る。例示的な読み取りデータ深度は、座（塩基）1つあたり1000～5000読み取りデータである。

a. シーケンシングの差次的深度

【0316】

一部の実施形態では、配列可変標的領域セットに対応する核酸は、エピジェネティック標的領域セットに対応する核酸よりも大きいシーケンシング深度までシーケンシングされる。例えば、配列変異体標的領域セットに対応する核酸のシーケンシング深度は、エピジェネティック標的領域セットに対応する核酸のシーケンシング深度よりも、少なくとも1.25倍、1.5倍、1.75倍、2倍、2.25倍、2.5倍、2.75倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、もしくは15倍大きいか、または1.25倍～1.5倍、1.5倍～1.75倍、1.75倍～2倍、2倍～2.25倍、2.25倍～2.5倍、2.5倍～2.75倍、2.75倍～3倍、3倍～3.5倍、3.5倍～4倍、4倍～4.5倍、4.5倍～5倍、5倍～5.5倍、5.5倍～6倍、6倍～7倍、7倍～8倍、8倍～9倍、9倍～10倍、10倍～11倍、11倍～12倍、13倍～14倍、14倍～15倍、もしくは15倍～100倍大きくあり得る。一部の実施形態では、前記シーケンシング深度は、少なくとも2倍大きい。一部の実施形態では、前記シーケンシング深度は、少なくとも5倍大きい。一部の実施形態では、前記シーケンシング深度は、少なくとも10倍大きい。一部の実施形態では、前記シーケンシング深度は、4～10倍大きい。一部の実施形態では、前記シーケンシング深度は、4～100倍大きい。これらの実施形態の各々は、配列可変標的領域セットに対応する核酸が、エピジェネティック標的領域セットに対応する核酸よりも大きいシーケンシング深度までシーケンシングされる程度に言及している。

【0317】

一部の実施形態では、配列可変標的領域セットに対応する捕捉されたcfDNA、およびエピジェネティック標的領域セットに対応する捕捉されたcfDNAは、同時に、例えば、同じシーケンシングセル（例えば、Illuminaシーケンサーのフローセル）において、ならびに／または別個に捕捉されたセットを組み換えることに起因するプールされた組成物もしくは配列可変標的領域セットに対応するcfDNAおよびエピジェネティック標的領域セットに対応する捕捉されたcfDNAを同じ容器に捕捉することによって得られた組成物であり得る同じ組成物において、シーケンシングされる。

11. 解析

【0318】

一部の実施形態では、本明細書に記載される方法は、腫瘍（または新生物細胞もしくはがん細胞）によって産生されるDNAの存在を同定するステップを含む。

【0319】

本方法は、対象における状態、特に、がんの存在を診断するため、状態を特徴付ける（例えば、がんをステージ分類するまたはがんの不均一性を決定する）ため、状態の処置に対する応答をモニタリングするため、状態を発症するリスクまたは状態の後続の経過の予後判定を行うために、使用することができる。本開示はまた、特定の処置選択肢の有効性を決定するのににも有用であり得る。奏功した処置選択肢では、処置が成功している場合、より多くのがんが死滅し、DNAが脱落し得るため、対象の血液中で検出されるコピー数変異または希少な突然変異の量が増加し得る。他の例では、これは生じない場合がある。別の例では、おそらくはある特定の処置選択肢は、経時的に、がんの遺伝子プロファイルと相関付けられ得る。この相関性は、治療を選択するのに有用であり得る。

【0320】

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

【0321】

検出することができるがんのタイプおよび数としては、血液のがん、脳のがん、肺がん、皮膚がん、鼻のがん、喉のがん、肝臓がん、骨がん、リンパ腫、膵臓がん、皮膚がん、腸がん、直腸がん、甲状腺がん、膀胱がん、腎臓がん、口腔がん、胃がん、固形状態の腫瘍、不均一な腫瘍、均一な腫瘍などを挙げることができる。がんのタイプおよび／またはステージは、突然変異、希少な突然変異、インデル、コピー数変異、塩基転換、転座、逆位、欠失、異数性、部分異数性、倍数性、染色体不安定性、染色体構造の変更、遺伝子融合、染色体融合、遺伝子トランケーション、遺伝子増幅、遺伝子重複、染色体病変、DNA病変、核酸化学修飾の異常な変化、エピジェネティックパターンの異常な変化、および核酸5-メチルシトシンの異常な変化を含む、遺伝的変異から検出することができる。

10

【0322】

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

20

【0323】

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

30

【0324】

本方法は、不均一な疾患における異なる細胞に由来する遺伝子情報をまとめたものであるデータのプロファイル、フィンガープリント、またはセットを生成するために使用することができる。このデータのセットは、コピー数変異、エピジェネティック変異、および突然変異の解析を、単独または組合せで含み得る。

【0325】

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

40

【0326】

第1の部分試料を、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供するステップを含む、NGSを通じてMBDビーズにより分配されるライブラリーの分子タグ同定のための例示的な方法は、以下の通りである。

50

1. メチル結合ドメインタンパク質ービーズ精製キットを使用して、抽出されたDNA試料（例えば、ヒト試料に由来する抽出された血液血漿DNA、これは、必要に応じて、本明細書に記載される標的捕捉に供されている）を物理的に分配し、下流プロセッシングのためにプロセスからの全ての溶出物を保管する。
2. 差次的分子タグおよびNGSにより実行されるアダプター配列の各分配への並行した適用。例えば、高メチル化、残留メチル化（「ウォッシュ」）、および低メチル化分配を、分子タグを有するNGSーアダプターとライゲーションする。
3. 高メチル化分配を、本明細書に記載されるもののうちのいずれかなど、DNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供する。
4. 全ての分子タグ付けした分配を組み換え、続いて、アダプター特異的DNAプライマー配列を使用して増幅させる。
5. 目的のゲノム領域（例えば、がん特異的遺伝的変異体および差次的にメチル化された領域）を標的とする、組み換え、増幅させた全ライブラリーの捕捉／ハイブリダイゼーション。
6. 捕捉されたDNAライブラリーを再増幅させ、試料タグを付加する。異なる試料をプールし、NGS機器で多重にアッセイする。
7. 分子タグを使用して一意的分子を同定するNGSデータのバイオインフォマティクス解析、ならびに試料の差次的にMBD分配された分子へのデコンボリューション。この解析により、標準的な遺伝子シーケンシング／変異体検出と同時に、ゲノム領域に関して相対的5-メチルシトシンの情報を得ることができる。

10

20

【0327】

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

30

【0328】

一般に、DNA内の第1の核酸塩基に対してDNAの内の第2の核酸塩基とは異なって影響を及ぼす手順は、代わりに、差次的分子タグおよびNGSにより実行されるアダプター配列の各分配への並行した適用のステップの前に、行われてもよい。例えば、これは、DNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順が、分離、例えば、hmCーシールである場合に行われてもよく、そのような場合には、分離された集団は、それら自体が、互いに対して差次的にタグ付けされていてもよい。そのような例示的な方法は、以下の通りである。

40

1. メチル結合ドメインタンパク質ービーズ精製キットを使用して、抽出されたDNA試料（例えば、ヒト試料に由来する抽出された血液血漿DNA、これは、必要に応じて、本明細書に記載される標的捕捉に供されている）を物理的に分配し、下流プロセッシングのためにプロセスからの全ての溶出物を保管すること。
2. 高メチル化分配を、本明細書に記載されるもののうちのいずれかなど、DNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供する。

50

3. 差次的分子タグおよびNGSにより実行されるアダプター配列の各分配物への並行した適用。例えば、高メチル化分配（または該当する場合には、高メチル化分配の2つもしくはそれよりも多くの部分分配）、残留メチル化（「ウォッシュ」）分配、および低メチル化分配を、分子タグを有するNGS-アダプターとライゲーションする。

4. 全ての分子タグ付けした分配を組み換え、続いて、アダプター特異的DNAプライマー配列を使用して増幅させる。

5. 目的のゲノム領域（例えば、がん特異的遺伝的変異体および差次的にメチル化された領域）を標的とする、組み換え、増幅させた全ライブラリーの捕捉／ハイブリダイゼーション。

6. 捕捉されたDNAライブラリーを再増幅させ、試料タグを付加する。異なる試料をプールし、NGS機器で多重にアッセイする。

7. 分子タグを使用して一意的分子を同定する、NGSデータのバイオインフォマティクス解析、ならびに試料の差次的にMBD分配された分子へのデコンボリューション。この解析により、標準的な遺伝子シーケンシング／変異体検出と同時に、ゲノム領域に関して相対的5-メチルシトシンの情報を得ることができる。

12. 例示的なワークフロー

【0329】

分配およびライブラリー調製の例示的なワークフローが、本明細書に提供される。一部の実施形態では、分配およびライブラリー調製ワークフローの一部または全ての特色は、組み合わせて使用され得る。

a. 分配

【0330】

一部の実施形態では、試料DNA（例えば、5～200ngの間）を、メチル結合ドメイン（MBD）緩衝液と混合し、磁気ビーズをMBDタンパク質とコンジュゲートさせ、一晚インキュベートする。メチル化DNA（高メチル化DNA）は、このインキュベーション中に磁気ビーズ上のMBDタンパク質に結合する。非メチル化である（低メチル化DNA）またはあまり非メチル化であるDNA（中間メチル化）は、漸増濃度の塩を含有する緩衝液で、ビーズから洗い流される。例えば、非メチル化、低メチル化、および／または中間メチル化DNAを含有する、1つ、2つ、またはそれよりも多くの分画が、そのような洗浄から得られ得る。最終的に、高塩緩衝液を使用して、MBDタンパク質から、高度メチル化DNA（高メチル化DNA）を溶出させる。一部の実施形態では、これらの洗浄により、漸増メチル化レベルを有するDNAの3つの分配（低メチル化分配、中間メチル化分画、および高メチル化分配）が得られる。

【0331】

一部の実施形態では、DNAの3つの分配は、ライブラリー調製の酵素ステップの準備において脱塩および濃縮される。

b. ライブラリーの調製

【0332】

一部の実施形態では（例えば、分配中のDNAを濃縮した後）、分配されたDNAは、例えば、DNA分子の末端オーバーハングを伸長させ、断片の3'末端にアデノシン残基を付加し、各DNA断片の5'末端をリン酸化することによって、ライゲーション可能となる。DNAリガーゼおよびアダプターを添加して、各分配されたDNA分子を、各末端でアダプターとライゲーションさせる。これらのアダプターは、他の分配において使用されるアダプターの分配タグと識別可能な分配タグ（例えば、非ランダム、非一意的バーコード）を含有する。小分けにしたDNAをライゲーション可能にすることおよびライゲーションを行うことの前または後のいずれかで、少なくとも1つの部分試料（例えば、低メチル化分配、または該当する場合には低メチル化分配および中間メチル化分配）を、メチル化依存性ヌクレアーゼ（例えば、メチル化依存性制限酵素、例えば、FspEI）で消化させる。必要に応じて、高メチル化分配は、メチル化感受性ヌクレアーゼ、例えば、メチル化感受性制限酵素（例えば、HpaII、BstUI、およびHin6iのうちの1

10

20

30

40

50

つもしくは複数、または各々）で消化させてもよい。必要に応じて、高メチル化分配は、本明細書に記載されるもののうちのいずれかなど、DNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供され得る。DNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順により、さらに、高メチル化分配を分配する場合、アダプターのライゲーションは、高メチル化分配の部分分配が差示的にタグ付けされ得るように、この手順の後に行うべきである。次いで、3つ（またはそれよりも多くの）分配が、一緒にプールされ、増幅される（例えば、PCRによって、例えば、アダプターに特異的なプライマーを用いて）。

【0333】

PCR後、増幅されたDNAは、富化の前に、清浄され、濃縮され得る。増幅されたDNAは、目的の特定の領域を標的とする、本明細書に記載されるプローブのコレクション（これは、例えば、ビオチン化RNAプローブであり得る）と接触させられる。混合物を、例えば、塩緩衝液中で、例えば、一晚インキュベートする。プローブを捕捉し（例えば、ストレプトアビジン磁気ビーズを使用して）、捕捉されなかった増幅されたDNAから、例えば、一連の塩洗浄によって分離させ、それによって、試料を富化する。富化後に、富化させた試料を、PCRによって増幅させる。一部の実施形態では、PCRプライマーは、試料タグを含有し、それによって、試料タグをDNA分子に組み込む。一部の実施形態では、異なる試料に由来するDNAと一緒にプールされ、次いで、例えば、Illumina NovaSeqシーケンサーを使用して、多重シーケンシングされる。

C. ある特定の開示される方法のさらなる特色

1. 試料

【0334】

試料は、対象から単離された任意の生体試料であり得る。試料は、身体試料であり得る。試料としては、体組織、例えば、判明しているかまたは疑われる固形腫瘍、全血、血小板、血清、血漿、糞便、赤血球、白血球（white blood cell）または白血球（leucocyte）、内皮細胞、組織生検、脳脊髄液、滑液、リンパ液、腹水、間質液または細胞外液、細胞間の空間の流体が挙げられ得、歯肉溝滲出液、骨髓、胸水、脳脊髄液、唾液、粘液、喀痰、精液、汗、尿が挙げられる。試料は、好ましくは、体液、特に、血液およびその分画、ならびに尿である。試料は、対象からもともと単離されている形態であってもよく、または成分、例えば、細胞を除去もしくは追加するため、または1つの構成要素を別の構成要素と比べて富化するために、さらなるプロセッシングに供されていてもよい。したがって、解析のための好ましい体液は、無細胞核酸を含有する血漿または血清である。試料は、対象から単離または取得することができ、試料解析の現場へと輸送され得る。試料は、望ましい温度、例えば、室温、4℃、-20℃、および／または-80℃で保存または発送され得る。試料は、試料解析の現場で、対象から単離または取得することができる。対象は、ヒト、哺乳動物、動物、コンパニオン動物、介助動物、または愛玩動物であり得る。対象は、がんを有し得る。対象は、がんまたは検出可能ながんの症状を有さない場合がある。対象は、1つまたは複数のがん治療、例えば、化学療法、抗体、ワクチン、または生物学のうちのいずれか1つまたは複数で処置されていてもよい。対象は、寛解期にあってもよい。対象は、がんまたは任意のがん関連遺伝子突然変異／障害に罹患しやすいと診断されていてもよく、またはされていなくてもよい。

【0335】

血漿の体積は、シーケンシングされる領域の所望の読み取りデータ深度に依存し得る。例示的な体積は、0.4~40mL、5~20mL、10~20mLである。例えば、体積は、0.5mL、1mL、5mL、10mL、20mL、30mL、または40mLであり得る。試料採取される血漿の体積は、5~20mLであり得る。

【0336】

試料は、ゲノム等価物を含有する様々な量の核酸を含み得る。例えば、約30ngのDNAの試料は、約10,000（ 10^4 ）個の半数体ヒトゲノム等価物を含有し、cfDNAの場合、約2000億（ 2×10^{11} ）個の個々のポリヌクレオチド分子を含有し得

る。同様に、約100 ngのDNAの試料は、約30,000個の半数体ヒトゲノム等価物を含み、cfDNAの場合、約6000億個の個々の分子を含み得る。

【0337】

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

10

【0338】

増幅前の試料中の無細胞核酸の例示的な量は、約1 fg～約1 μg、例えば、1 pg～200 ng、1 ng～100 ng、10 ng～1000 ngの範囲である。例えば、量は、最大約600 ng、最大約500 ng、最大約400 ng、最大約300 ng、最大約200 ng、最大約100 ng、最大約50 ng、または最大約20 ngの無細胞核酸分子であり得る。量は、少なくとも1 fg、少なくとも10 fg、少なくとも100 fg、少なくとも1 pg、少なくとも10 pg、少なくとも100 pg、少なくとも1 ng、少なくとも10 ng、少なくとも100 ng、少なくとも150 ng、または少なくとも200 ngの無細胞核酸分子であり得る。量は、最大1フェムトグラム（fg）、10 fg、100 fg、1ピコグラム（pg）、10 pg、100 pg、1 ng、10 ng、100 ng、150 ng、または200 ngの無細胞核酸分子であり得る。方法は、1フェムトグラム（fg）～200 ngを得ることを含み得る。

20

【0339】

無細胞DNAは、対象からのその単離の時点において、細胞内に含有されていないDNAを指す。例えば、cfDNAは、細胞を溶解すること、それ以外では細胞内DNAを抽出することなしに、インタクトな細胞を除去した後に試料中に残留しているDNAとして、試料から単離され得る。無細胞核酸としては、DNA、RNA、およびそれらのハイブリッドが挙げられ、これには、ゲノムDNA、ミトコンドリアDNA、siRNA、miRNA、循環RNA（cRNA）、tRNA、rRNA、核小体低分子RNA（snRNA）、Piwi相互作用RNA（piRNA）、長鎖非コードRNA（長鎖ncRNA）、またはこれらのうちのいずれかの断片が含まれる。無細胞核酸は、二本鎖、一本鎖、またはそれらのハイブリッドであり得る。無細胞核酸は、分泌または細胞死プロセス、例えば、細胞壊死およびアポトーシスを通じて、体液中に放出され得る。一部の無細胞核酸は、がん細胞、例えば、循環腫瘍DNA（ctDNA）から、体液中に放出される。その他のものは、健康な細胞から放出される。一部の実施形態では、cfDNAは、無細胞胎児DNA（cffDNA）である。一部の実施形態では、無細胞核酸は、腫瘍細胞によって産生される。一部の実施形態では、無細胞核酸は、腫瘍細胞と非腫瘍細胞との混合物によって産生される。

30

40

【0340】

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

【0341】

無細胞核酸は、溶液に見出される無細胞核酸を、インタクトな細胞および体液の他の不溶性構成要素から分離する、分画または分配ステップによって、体液から単離することができる。分配するステップには、遠心分離または濾過などの技術が含まれ得る。あるいは、体液中の細胞を、溶解し、無細胞核酸および細胞核酸を、一緒にプロセッシングしても

50

よい。一般に、緩衝液の添加および洗浄ステップの後に、核酸は、アルコールで沈殿させることができる。シリカに基づくカラムなどのさらなる洗浄ステップを使用して、夾雑物または塩を除去してもよい。バイサルファイトシーケンシング、ハイブリダイゼーション、および／またはライゲーションのための非特異的バルク担体核酸、例えば、C 1 D N A、D N A、またはタンパク質を、手順のある特定の態様、例えば、収量を最適化するために、反応全体にわたって添加してもよい。

【0342】

そのようなプロセッシングの後、試料は、二本鎖D N A、一本鎖D N A、および一本鎖R N Aを含む、様々な形態の核酸を含み得る。一部の実施形態では、一本鎖D N AおよびR N Aは、後続のプロセッシングおよび解析ステップに含められるように、二本鎖形態に変換してもよい。

10

【0343】

試料中の二本鎖D N A分子および二本鎖D N A分子に変換される一本鎖核酸分子は、一方の末端または両方の末端において、アダプターに連結され得る。典型的には、二本鎖分子は、4つ全ての標準的なヌクレオチドの存在下において、5' - 3' ポリメラーゼおよび3' - 5' エクソヌクレアーゼ（またはブルーフリーディング機能）を有するポリメラーゼでの処置によって、平滑末端化される。クレノウラージ断片およびT 4ポリメラーゼが、好適なポリメラーゼの例である。平滑末端化されたD N A分子は、少なくとも部分的に二本鎖のアダプター（例えば、Y形状またはベル形状のアダプター）とライゲーションされ得る。あるいは、ライゲーションを容易にするために、相補的なヌクレオチドが、試料核酸およびアダプターの平滑末端に付加されてもよい。平滑末端ライゲーションおよび付着末端ライゲーションの両方が、本明細書において企図される。平滑末端ライゲーションでは、核酸分子およびアダプタータグの両方が、平滑末端を有する。付着末端ライゲーションでは、典型的には、核酸分子は、「A」オーバーハングを有し、アダプターは、「T」オーバーハングを有する。

20

2. 増幅

【0344】

アダプターが隣接する試料核酸は、P C Rおよび他の増幅方法によって増幅され得る。増幅は、典型的には、増幅させようとするD N A分子に隣接するアダプター内のプライマー結合部位にプライマーが結合することによって、プライミングされる。増幅方法は、サーモサイクリングの結果としての変性、アニーリング、および伸長のサイクルを含み得るか、または転写媒介増幅におけるような等温のものであってもよい。他の増幅方法としては、リガーゼ連鎖反応、鎖置換増幅、核酸配列に基づく増幅、および自家持続性配列に基づく複製法が挙げられる。

30

【0345】

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

40

3. ベイトセット；捕捉部分

【0346】

上記で考察されるように、試料中の核酸は、捕捉ステップに供することができ、ここで、標的配列を有する分子が、後続の解析のために捕捉される。標的の捕捉は、捕捉部分、例えば、ビオチンまたは以下に記載される他の例で標識されたオリゴヌクレオチドベイトを含むベイトセットの使用を伴い得る。プローブは、遺伝子など、領域のパネルにわたってタイリングするように選択された配列を有し得る。一部の実施形態では、ベイトセットは、本明細書の他の箇所において考察されるように、標的領域のセット、例えば、それぞれ、配列可変標的領域セットおよびエピジェネティック標的領域セットに関して、より高い捕捉収率およびより低い捕捉収率を有し得る。そのようなベイトセットは、標的分子の

50

ベイトとのハイブリダイゼーションを可能にする条件下において、試料と組み合わせられる。次いで、捕捉された分子は、捕捉部分を使用して単離される。例えば、ビーズに基づくストレプトアビジンによるビオチン捕捉部分。そのような方法は、例えば、2017年12月26日に発行された米国特許第9,850,523号において、さらに説明されており、これは、参照により本明細書に組み込まれる。

【0347】

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

10

D. 標的特異的プローブのコレクション

【0348】

一部の実施形態では、標的特異的プローブのコレクションを、本明細書に記載される方法において使用する。一部の実施形態では、標的特異的プローブのコレクションは、配列可変標的領域セットに対して特異的な標的結合プローブおよびエピジェネティック標的領域セットに対して特異的な標的結合プローブを含む。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの捕捉収率は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの捕捉収率よりも高い（例えば、少なくとも2倍高い）。一部の実施形態では、標的特異的プローブのコレクションは、エピジェネティック標的領域セットに対して特異的なその捕捉収率よりも高い（例えば、少なくとも2倍高い）配列可変標的領域セットに対して特異的な捕捉収率を有するように構成される。

20

【0349】

一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの捕捉収率は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの捕捉収率よりも少なくとも1.25倍、1.5倍、1.75倍、2倍、2.25倍、2.5倍、2.75倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、または15倍高い。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの捕捉収率は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの捕捉収率よりも1.25倍～1.5倍、1.5倍～1.75倍、1.75倍～2倍、2倍～2.25倍、2.25倍～2.5倍、2.5倍～2.75倍、2.75倍～3倍、3倍～3.5倍、3.5倍～4倍、4倍～4.5倍、4.5倍～5倍、5倍～5.5倍、5.5倍～6倍、6倍～7倍、7倍～8倍、8倍～9倍、9倍～10倍、10倍～11倍、11倍～12倍、13倍～14倍、または14倍～15倍高い。

30

40

【0350】

一部の実施形態では、標的特異的プローブのコレクションは、エピジェネティック標的領域セットに対するその捕捉収率よりも少なくとも1.25倍、1.5倍、1.75倍、2倍、2.25倍、2.5倍、2.75倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、または15倍高い、配列可変標的領域セットに対して特異的な捕捉収率を有するように構成される。一部の実施形態では、標的特異的プローブのコレクションは、エピジェネティック標的領域セットに対して特異的なその捕捉収率よりも1.25倍～1.5倍、1.5倍～1.75倍、1.75倍～2倍、2倍～2.25倍、2.25倍～2.5倍、2.5倍～2.75倍、2

50

． 7 5 倍～ 3 倍、 3 倍～ 3 . 5 倍、 3 . 5 倍～ 4 倍、 4 倍～ 4 . 5 倍、 4 . 5 倍～ 5 倍、 5 倍～ 5 . 5 倍、 5 . 5 倍～ 6 倍、 6 倍～ 7 倍、 7 倍～ 8 倍、 8 倍～ 9 倍、 9 倍～ 1 0 倍、 1 0 倍～ 1 1 倍、 1 1 倍～ 1 2 倍、 1 3 倍～ 1 4 倍、または 1 4 倍～ 1 5 倍高い、配列可変標的領域セットに対して特異的な捕捉収率を有するように構成される。

【 0 3 5 1 】

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

【 0 3 5 2 】

一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、エピジェネティック標的領域セットに対して特異的な標的特異的プローブよりも高い濃度で存在する。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの濃度は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの濃度よりも少なくとも 1 . 2 5 倍、1 . 5 倍、1 . 7 5 倍、2 倍、2 . 2 5 倍、2 . 5 倍、2 . 7 5 倍、3 倍、3 . 5 倍、4 倍、4 . 5 倍、5 倍、6 倍、7 倍、8 倍、9 倍、1 0 倍、1 1 倍、1 2 倍、1 3 倍、1 4 倍、または 1 5 倍高い。一部の実施形態では、配列可変標的領域セットに対して特異的な標的結合プローブの濃度は、エピジェネティック標的領域セットに対して特異的な標的結合プローブの濃度よりも 1 . 2 5 倍～ 1 . 5 倍、1 . 5 倍～ 1 . 7 5 倍、1 . 7 5 倍～ 2 倍、2 倍～ 2 . 2 5 倍、2 . 2 5 倍～ 2 . 5 倍、2 . 5 倍～ 2 . 7 5 倍、2 . 7 5 倍～ 3 倍、3 倍～ 3 . 5 倍、3 . 5 倍～ 4 倍、4 倍～ 4 . 5 倍、4 . 5 倍～ 5 倍、5 倍～ 5 . 5 倍、5 . 5 倍～ 6 倍、6 倍～ 7 倍、7 倍～ 8 倍、8 倍～ 9 倍、9 倍～ 1 0 倍、1 0 倍～ 1 1 倍、1 1 倍～ 1 2 倍、1 3 倍～ 1 4 倍、または 1 4 倍～ 1 5 倍高い。そのような実施形態では、濃度は、各セット中の個々のプローブの体積濃度あたりの平均質量を指し得る。

【 0 3 5 3 】

一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、エピジェネティック標的領域セットに対して特異的な標的特異的プローブよりもそれらの標的に対する高い親和性を有する。親和性は、異なるプローブ化学を使用することによってを含む、当業者に公知の任意の方法でモジュレートすることができる。ある特定のヌクレオチド改変、例えばシトシン 5 -メチル化（ある特定の配列の文脈において）、糖の 2 ' 位でヘテロ原子を提供する改変、および L N A ヌクレオチドは、二本鎖核酸の安定性を増加させることができ、そのような改変を有するオリゴヌクレオチドが、その相補的配列に関して比較的高い親和性を有することを示している。例えば、Severin et al., *Nucleic Acids Res.* 39: 8740-8751 (2011); Freier et al., *Nucleic Acids Res.* 25: 4429-4443 (1997); 米国特許第 9, 7 3 8, 8 9 4 号を参照されたい。同様に、より長い配列長は、一般的に親和性の増加を提供する。他のヌクレオチド改変、例えば、核酸塩基ヒポキサンチンをグアニンの代わりに置換すると、オリゴヌクレオチドとその相補配列の間の水素結合の量を低減させることによって親和性を低減する。一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、その標的に対するその親和性を増加させる改変を有する。一部の実施形態では、あるいはまたはさらに、エピジェネティック標的領域セットに対して特異的な標的特異的プローブは、その標的に対するその親和性を減少させる改変を有する。一部の実施形態では、配列可変標的領域セットに対して特異的な標的特異的プローブは、エピジェネティック標的領域セットに対して特異的な標的特異的プローブよりも長い平均長および／または高い平均融解温度を有する。これらの実施形態を互いに、および／または上記で考察した濃度の差と組み合わせて、上記の任意の倍率の差またはその範囲などの、捕捉収率の所望の倍率の差を達成してもよい。

【 0 3 5 4 】

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

【0355】

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

10

【0356】

一部の実施形態では、標的特異的プローブは、単一の組成物中で提供される。単一の組成物は溶液（液体または凍結）であり得る。あるいは、これは凍結乾燥物であり得る。

【0357】

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

20

1. エピジェネティック標的領域に対して特異的なプローブ

【0358】

エピジェネティック標的領域セットのプローブは、新生物（例えば、腫瘍またはがん）細胞からのDNAを、健康な細胞、例えば非新生物の循環細胞と区別する可能性が高い1つまたは複数のタイプの標的領域に対して特異的なプローブを含み得る。そのような領域の例示的なタイプは、本明細書において、例えば捕捉されたセットに関する上記の節に詳細に考察されている。エピジェネティック標的領域セットのプローブはまた、例えば本明細書に記載される1つまたは複数の対照領域のプローブも含み得る。

30

【0359】

一部の実施形態では、エピジェネティック標的領域セットのプローブは、少なくとも100 kbp、例えば、少なくとも200 kbp、少なくとも300 kbp、または少なくとも400 kbpのフットプリントを有する。一部の実施形態では、エピジェネティック標的領域セットは、100～20 Mbp、例えば、100～200 kbp、200～300 kbp、300～400 kbp、400～500 kbp、500～600 kbp、600～700 kbp、700～800 kbp、800～900 kbp、900～1,000 kbp、1～1.5 Mbp、1.5～2 Mbp、2～3 Mbp、3～4 Mbp、4～5 Mbp、5～6 Mbp、6～7 Mbp、7～8 Mbp、8～9 Mbp、9～10 Mbp、または10～20 Mbpの範囲のフットプリントを有する。一部の実施形態では、エピジェネティック標的領域セットは、少なくとも20 Mbpのフットプリントを有する。

40

a. 高メチル化可変標的領域

【0360】

一部の実施形態では、エピジェネティック標的領域セットのプローブは、1つまたは複数の高メチル化可変標的領域に対して特異的なプローブを含む。高メチル化可変標的領域はまた、本明細書において高メチル化DMR（差次的にメチル化された領域）とも呼ばれ得る。高メチル化可変標的領域は、上記のそれらのいずれかであり得る。例えば、一部の実施形態では、高メチル化可変標的領域に対して特異的なプローブは、表1に列挙される複数の座、例えば表1に列挙される座の少なくとも10%、20%、30%、40%、50%、60%、70%、80%、90%、または100%に対して特異的なプローブを含

50

む。一部の実施形態では、高メチル化可変標的領域に対して特異的なプローブは、表 2 に列挙される複数の座、例えば表 2 に列挙される座の少なくとも 10 %、20 %、30 %、40 %、50 %、60 %、70 %、80 %、90 %、または 100 % に対して特異的なプローブを含む。一部の実施形態では、高メチル化可変標的領域に対して特異的なプローブは、表 1 または表 2 に列挙される複数の座、例えば表 1 または表 2 に列挙される座の少なくとも 10 %、20 %、30 %、40 %、50 %、60 %、70 %、80 %、90 %、または 100 % に対して特異的なプローブを含む。一部の実施形態では、標的領域として含まれる各座に関して、遺伝子の転写開始部位と終止コドン（選択的にスプライシングされる遺伝子の場合には最後の終止コドン）の間に結合するハイブリダイゼーション部位を有する 1 つまたは複数のプローブが存在し得る。一部の実施形態では、1 つまたは複数のプローブは、列挙された位置の 300 bp 以内、例えば 200 または 100 bp 以内に結合する。一部の実施形態では、プローブは、上に列挙された位置と重複するハイブリダイゼーション部位を有する。一部の実施形態では、高メチル化標的領域に対して特異的なプローブは、乳がん、結腸がん、腎臓がん、肝臓がん、および肺がんのうちの 1、2、3、4、または 5 つにおいて集合的に高メチル化を示す高メチル化標的領域の 1、2、3、4、または 5 つのサブセットに対して特異的なプローブを含む。

10

b. 低メチル化可変標的領域

【0361】

一部の実施形態では、エピジェネティック標的領域セットのプローブは、1 つまたは複数の低メチル化可変標的領域に対して特異的なプローブを含む。低メチル化可変標的領域はまた、本明細書において低メチル化 DMR（差次的にメチル化された領域）とも呼ばれ得る。低メチル化可変標的領域は、上記のそれらのいずれかであり得る。例えば、1 つまたは複数の低メチル化可変標的領域に対して特異的なプローブは、反復エレメント、例えば、LINE1 エレメント、Alu エレメント、セントロメアタンデムリピート、ペリセントロメアタンデムリピート、およびサテライト DNA などの領域のプローブを含んでもよく、健康な細胞において通常メチル化される遺伝子間領域は、腫瘍細胞において低減されたメチル化を示し得る。

20

【0362】

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

30

【0363】

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

40

c. CTCF 結合領域

【0364】

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

50

なプローブを含む。一部の実施形態では、エピジェネティック標的領域セットのプローブは、CTCF結合部位の少なくとも100bp、少なくとも200bp、少なくとも300bp、少なくとも400bp、少なくとも500bp、少なくとも750bp、または少なくとも1000bp上流および下流の領域を含む。

d. 転写開始部位

【0365】

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

10

e. 局所増幅

【0366】

上記のように、局所増幅が体細胞突然変異であるが、それらは、メチル化の変化などのある特定のエピジェネティック変化を検出するためのアプローチと類似の方法で読み取りデータの頻度に基づいてシーケンシングすることによって検出することができる。そのため、がんにおいて局所増幅を示し得る領域を、上記で考察したようにエピジェネティック標的領域セットに含めることができる。一部の実施形態では、エピジェネティック標的領域セットに対して特異的なプローブは、局所増幅に対して特異的なプローブを含む。一部の実施形態では、局所増幅に対して特異的なプローブは、AR、BRAF、CCND1、CCND2、CCNE1、CDK4、CDK6、EGFR、ERBB2、FGFR1、FGFR2、KIT、KRAS、MET、MYC、PDGFRA、PIK3CA、およびRAF1のうちの1つまたは複数に対して特異的なプローブを含む。例えば、一部の実施形態では、局所増幅に対して特異的なプローブは、前述の標的の少なくとも2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、または18個のうちの1つまたは複数に対して特異的なプローブを含む。

20

30

f. 対照領域

【0367】

データ検証を容易にするために対照領域を含めることは有用であり得る。一部の実施形態では、エピジェネティック標的領域セットに対して特異的なプローブは、本質的に全ての試料においてメチル化されると予想される対照メチル化領域に対して特異的なプローブを含む。一部の実施形態では、エピジェネティック標的領域セットに対して特異的なプローブは、本質的に全ての試料において低メチル化であると予想される対照低メチル化領域に対して特異的なプローブを含む。

2. 配列可変標的領域に対して特異的なプローブ

【0368】

配列可変標的領域セットのプローブは、がんにおいて体細胞突然変異を受けることが公知である複数の領域に対して特異的なプローブを含み得る。プローブは、本明細書に記載される任意の配列可変標的領域セットに対して特異的であり得る。例示的な配列可変標的領域セットは、本明細書において、例えば捕捉されたセットに関する上記の節で詳細に考察される。

40

【0369】

一部の実施形態では、配列可変標的領域プローブセットは、少なくとも0.5kb、例えば、少なくとも1kb、少なくとも2kb、少なくとも5kb、少なくとも10kb、少なくとも20kb、少なくとも30kb、または少なくとも40kbのフットプリントを有する。一部の実施形態では、エピジェネティック標的領域プローブセットは、0.5

50

～100kb、例えば、0.5～2kb、2～10kb、10～20kb、20～30kb、30～40kb、40～50kb、50～60kb、60～70kb、70～80kb、80～90kb、および90～100kbの範囲のフットプリントを有する。一部の実施形態では、配列可変標的領域プローブセットは、少なくとも50kbp、例えば、少なくとも100kbp、少なくとも200kbp、少なくとも300kbp、または少なくとも400kbpのフットプリントを有する。一部の実施形態では、配列可変標的領域プローブセットは、100～2000kbp、例えば、100～200kbp、200～300kbp、300～400kbp、400～500kbp、500～600kbp、600～700kbp、700～800kbp、800～900kbp、900～1,000kbp、1～1.5Mbp、または1.5～2Mbpの範囲のフットプリントを有する。一部の実施形態では、配列可変標的領域セットは、少なくとも2Mbpのフットプリントを有する。

10

【0370】

一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表3の遺伝子のうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、または70個の少なくとも一部分に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表3のSNVのうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、または70個に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表3の融合体のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、または6個に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表3のインデルのうちの少なくとも1個、少なくとも2個、または3個の少なくとも一部分に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表4の遺伝子のうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、少なくとも70個、または73個の少なくとも一部分に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表4のSNVのうちの少なくとも5個、少なくとも10個、少なくとも15個、少なくとも20個、少なくとも25個、少なくとも30個、少なくとも35個、少なくとも40個、少なくとも45個、少なくとも50個、少なくとも55個、少なくとも60個、少なくとも65個、少なくとも70個、または73個に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表4の融合体のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくとも16個、少なくとも17個、または18個の少なくとも一部分に特異的なプローブを含む。一部の実施形態では、配列可変標的領域セットに特異的なプローブは、表5の遺伝子のうちの少なくとも1個、少なくとも2個、少なくとも3個、少なくとも4個、少なくとも5個、少なくとも6個、少なくとも7個、少なくとも8個、少なくとも9個、少なくとも10個、少なくとも11個、少なくとも12個、少なくとも13個、少なくとも14個、少なくとも15個、少なくとも16個、少なくとも17個、少なくとも18個、少なくとも19個、または少なくとも20個の少なくとも一部分に特異的なプローブを

20

30

40

50

含む。

【0371】

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

10

E. 捕捉されたDNAを含む組成物

【0372】

DNAの第1および第2の集団を含む組合せであって、第2の集団が、本明細書に記載されるメチル化依存性ヌクレアーゼのうちのいずれか1つまたは任意の組合せであり得る少なくとも1つのメチル化依存性ヌクレアーゼの認識部位において、末端、または付着したタグもしくはアダプターを有するDNAの断片を含む、組合せが、本明細書において提供される。一部の実施形態では、第1および第2の集団は、差次的にタグ付けされる。第1の集団は、第2の集団よりも高い割合でシトシン改変を有するDNAを含むか、またはそれに由来し得る。第1の集団は、変更された塩基対形成特異性を有するDNAにもともと存在していた第1の核酸塩基の形態および変更された塩基対形成特異性を有さない第2の核酸塩基を含み得、ここで、塩基対形成特異性の変更前にDNAにもともと存在する第1の核酸塩基の形態は、改変または非改変核酸塩基であり、第2の核酸塩基は、第1の核酸塩基とは異なる改変または非改変核酸塩基であり、塩基対形成特異性の変更前にDNAにもともと存在する第1の核酸塩基の形態および第2の核酸塩基は、同じ塩基対形成特異性を有する。一部の実施形態では、シトシン改変は、シトシンメチル化である。一部の実施形態では、第1の核酸塩基は、改変または非改変シトシンであり、第2の核酸塩基は、改変または非改変シトシンである。第1および第2の核酸塩基は、本明細書において発明の概要で、または第1の部分試料を、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供することに関して、考察されているもののうちのいずれかであり得る。一部の実施形態では、第1の集団は、本明細書に記載されるメチル化感受性ヌクレアーゼのうちのいずれか1つまたは任意の組合せであり得る少なくとも1つのメチル化感受性ヌクレアーゼの認識部位において、末端、または付着したタグもしくはアダプターを有するDNAの断片を含む。

20

30

【0373】

一部の実施形態では、第1の集団は、1つまたは複数の配列タグの第1のセットから選択される配列タグを含み、第2の集団は、1つまたは複数の配列タグの第2のセットから選択される配列タグを含み、配列タグの第2のセットは、配列タグの第1のセットとは異なる。配列タグは、バーコードを含み得る。

【0374】

一部の実施形態では、第1の集団は、保護されたhmC、例えばグルコシル化hmCを含む。

40

【0375】

一部の実施形態では、第1の集団は、本明細書で考察される変換手順、例えばバイサルファイト変換、Ox-Bs変換、TAB変換、ACE変換、TAP変換、TAPsβ変換、またはCAP変換のいずれかに供された。一部の実施形態では、第1の集団は、hmCの保護後にmCおよび／またはCの脱アミノ化に供された。

【0376】

組合せの一部の実施形態では、第1の集団は、第2の集団よりも大きい割合でシトシン改変を有するDNAを含むかまたはそれに由来し、第1の集団は、第1および第2の部分集団を含み、第1の核酸塩基は、改変または非改変核酸塩基であり、第2の核酸塩基は、

50

第 1 の核酸塩基とは異なる改変または非改変核酸塩基であり、第 1 の核酸塩基および第 2 の核酸塩基は同じ塩基対形成特異性を有する。一部の実施形態では、第 2 の集団は、第 1 の核酸塩基を含まない。一部の実施形態では、第 1 の核酸塩基は、改変または非改変シトシンであり、第 2 の核酸塩基は、改変または非改変シトシンであり、必要に応じて改変シトシンは mC または hmC である。一部の実施形態では、第 1 の核酸塩基は、改変または非改変アデニンであり、第 2 の核酸塩基は改変または非改変アデニンであり、必要に応じて改変アデニンは mA である。

【0377】

一部の実施形態では、第 1 の核酸塩基（例えば、改変シトシン）はビオチン化される。一部の実施形態では、第 1 の核酸塩基（例えば、改変シトシン）は親和性標識（例えば、ビオチン）を含む β -6-アジド-グルコシル-5-ヒドロキシメチルシトシンに対するヒュスゲン環化付加の生成物である。

【0378】

本明細書に記載される組合せのいずれかでは、捕捉された DNA は、cfDNA を含む得る。

【0379】

捕捉された DNA は、例えばエピジェネティック標的領域セットに対応する DNA の濃度よりも配列可変標的領域セット（上記で考察したようにフットプリントサイズに関して正規化）に対応する DNA の高い濃度を含む、捕捉セットに関する本明細書に記載される特色のいずれかを有し得る。一部の実施形態では、捕捉セットの DNA は、本明細書に記載される DNA に付加され得る配列タグを含む。一般的に、配列タグを含めることは、その天然に存在する非タグ付け形態とは異なる DNA 分子をもたらす。

【0380】

組合せはさらに、その各々が、天然に存在する核酸分子とは異なり得る、本明細書に記載されるプローブセットまたはシーケンシングプライマーを含む得る。例えば、本明細書に記載されるプローブセットは、捕捉部分を含み得、シーケンシングプライマーは、天然に存在しない標識を含み得る。

F. コンピュータシステム

【0381】

本開示の方法は、コンピュータシステムを使用して、またはその補助を用いて、実装することができる。例えば、そのような方法は、試料を、第 1 の部分試料および第 2 の部分試料を含む複数の部分試料に分配するステップであって、第 1 の部分試料が、第 2 の部分試料よりも高い割合でシトシン改変を有する DNA を含む、ステップと、第 1 の部分試料を、第 1 の部分試料の DNA 内の第 1 の核酸塩基に対して DNA 内の第 2 の核酸塩基とは異なって影響を及ぼす手順に供するステップであって、第 1 の核酸塩基が、改変または非改変核酸塩基であり、第 2 の核酸塩基が、第 1 の核酸塩基とは異なる改変または非改変核酸塩基であり、第 1 の核酸塩基および第 2 の核酸塩基が、同じ塩基対形成特異性を有する、ステップと、第 1 の部分試料中の DNA および第 2 の部分試料中の DNA を、第 1 の部分試料の DNA 内の第 1 の核酸塩基を第 2 の核酸塩基と識別する様式でシーケンシングするステップとを含み得る。

【0382】

図 5 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それ以外の方法で構成されたコンピュータシステム 501 を示す。コンピュータシステム 501 は、試料調製、シーケンシング、および／または解析の様々な態様を調節することができる。一部の例では、コンピュータシステム 501 は、試料調製および核酸シーケンシングを含む試料解析を実施するように構成される。

【0383】

コンピュータシステム 501 は、シングルコアもしくはマルチコアプロセッサ、または並列プロセッシングのための複数のプロセッサであり得る中央処理装置（CPU、または本明細書において「プロセッサ」および「コンピュータプロセッサ」）505 を含む

。コンピュータシステム 501 はまた、メモリーまたはメモリー場所 510（例えば、ランダムアクセスメモリー、読み取り専用メモリー、フラッシュメモリー）、電子ストレージユニット 515（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 520（例えば、ネットワークアダプター）、ならびに周辺デバイス 525、例えばキャッシュ、他のメモリー、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も含む。メモリー 510、ストレージユニット 515、インターフェース 520、および周辺デバイス 525 は、マザーボードなどの通信ネットワークまたはバス（実線）を通して CPU 505 と通信する。ストレージユニット 515 は、データを保存するためのデータストレージユニット（またはデータレポジトリ）であり得る。コンピュータシステム 501 は、通信インターフェース 520 の助けを借りてコンピュータネットワーク 530 に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 530 は、Internet、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 Internet 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り得る。コンピュータネットワーク 530 は、一部の例では、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 530 は、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し得る 1つまたは複数のコンピュータサーバーを含み得る。コンピュータネットワーク 530 は、一部の例では、コンピュータシステム 501 の助けを借りて、デバイスをコンピュータシステム 501 にカップリングさせて、クライアントまたはサーバーとして挙動することを可能にし得る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る。

10

20

【0384】

CPU 505 は、プログラムまたはソフトウェアにおいて具体化することができる機械可読命令のシーケンスを実行することができる。命令は、メモリー 510 などのメモリー場所に保存され得る。CPU 505 によって実施される操作の例としては、フェッチ、復号、実行、およびライトバックが挙げられ得る。

【0385】

ストレージユニット 515 は、ドライバー、ライブラリー、およびセーブしたプログラムなどのファイルを保存することができる。ストレージユニット 515 は、ユーザーが作成したプログラム、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を保存することができる。ストレージユニット 515 は、ユーザーデータ、例えばユーザーの好みおよびユーザープログラムを保存することができる。コンピュータシステム 501 は、一部の例では、コンピュータシステム 501 に外付けの 1つまたは複数の追加の、例えば、イントラネットまたは Internet を通してコンピュータシステム 501 と通信するリモートサーバー上に位置する、データストレージユニットを含み得る。データは、例えば通信ネットワークまたは物理的データ転送を使用して（例えば、ハードドライブ、サムドライブ、または他のデータストレージ機構を使用して）1つの場所から別の場所へと転送され得る。

30

【0386】

コンピュータシステム 501 は、ネットワーク 530 を通して 1つまたは複数のリモートコンピュータシステムと通信することができる。実施形態に関して、コンピュータシステム 501 は、ユーザー（例えば、オペレーター）のリモートコンピュータシステムと通信することができる。リモート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）、スレートもしくはタブレット PC（例えば、Apple（登録商標）iPad（登録商標）、Samsung（登録商標）Galaxy Tab）、電話、スマートフォン（例えば、Apple（登録商標）iPhone（登録商標）、Android（登録商標）可能デバイス、Blackberry（登録商標））、またはパーソナルデジタルアシスタントが挙げられる。ユーザーは、ネットワーク 530 を介してコンピュータシステム 501 にアクセスすることができる。

40

【0387】

本明細書に記載される方法は、コンピュータシステム 501 の電子ストレージ場所、例

50

例えばメモリー５１０または電子ストレージユニット５１５などに保存された機械（例えば、コンピュータプロセッサ）実行可能コードによってインプリメントすることができる。機械実行可能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５０５によって実行され得る。一部の例では、コードを、ストレージユニット５１５から読み出し、プロセッサ５０５が容易にアクセスするためにメモリー５１０に保存することができる。一部の状況では、電子ストレージユニット５１５を除外することができ、機械実行可能命令はメモリー５１０に保存される。

【０３８８】

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、DNAを含む試料を、第１の部分試料および第２の部分試料を含む複数の部分試料に分配するステップであって、第１の部分試料が、第２の部分試料よりも高い割合でシトシン改変を有するDNAを含む、ステップと、第２の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、第２の部分試料中の非特異的に分配されたDNAを分解して、処置された第２の部分試料を産生するステップ、必要に応じて、第１の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、第１の部分試料中の非特異的に分配されたDNAを分解して、処置された第１の部分試料を産生するステップと、第１の部分試料および処置された第１の部分試料から、エピジェネティック標的領域を含む第１の標的領域セットを捕捉し、処置された第２の部分試料から、エピジェネティック標的領域を含む第２の標的領域セットを捕捉するステップと、第１の標的領域セットおよび第２の標的領域セット中のDNAをシーケンシングするステップとを含む方法の少なくとも一部分を行う、コンピュータ実行可能命令を含む、非一時的コンピュータ可読媒体を提供する。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、試料を、第１の部分試料および第２の部分試料を含む複数の部分試料に分配するステップであって、第１の部分試料が、第２の部分試料よりも高い割合でシトシン改変を有するDNAを含む、ステップと、第２の部分試料を、メチル化依存性ヌクレアーゼと接触させ、それによって、第２の部分試料中の非特異的に分配されたDNAを分解して、処置された第２の部分試料を産生するステップ、必要に応じて、第１の部分試料を、メチル化感受性エンドヌクレアーゼと接触させ、それによって、第１の部分試料中の非特異的に分配されたDNAを分解して、処置された第１の部分試料を産生するステップと、第１の部分試料および処置された第１の部分試料から、エピジェネティック標的領域を含む第１の標的領域セットを捕捉するステップと、第１の標的領域セット中のDNAおよび第２の部分試料に由来するDNAをシーケンシングするステップとを含む方法の少なくとも一部分を行う、コンピュータ実行可能命令を含む、非一時的コンピュータ可読媒体を提供する。一部の実施形態では、方法は、核酸シーケンサーによってシーケンシングから生成された複数の配列読み取りデータを取得するステップと、複数の配列読み取りデータを、１つまたは複数の参照配列にマッピングして、マッピングされた配列読み取りデータを生成するステップと、マッピングされた配列読み取りデータをプロセッシングして、対象ががんを有する可能性を決定するステップとをさらに含む。

【０３８９】

コードは、事前にコンパイルされ、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と共に使用するように構成することができ、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、事前にコンパイルされた、またはコンパイルされた通りに実行することが可能と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することができる。

【０３９０】

本明細書に提供されるシステムおよび方法の態様、例えばコンピュータシステム５０１は、プログラミングにおいて具体化することができる。テクノロジーの様々な態様は、典型的に、機械可読媒体の１つのタイプにおいて実行または具体化される機械（またはプロセッサ）実行可能コードおよび／または関連データの形態の「生成物」または「製造品」であると考えられ得る。機械実行可能コードは、電子ストレージユニット、そのような

メモリー（例えば、読み取り専用メモリー、ランダムアクセスメモリー、フラッシュメモリー）またはハードディスクに保存することができる。「ストレージ」型の媒体は、ソフトウェアプログラミングのためにいついかなる時にも非一時的ストレージを提供し得るコンピュータの有形メモリー、プロセッサなど、またはその関連するモジュール、例えば様々な半導体メモリー、テープドライブ、ディスクドライブなどのいずれかまたは全てを含み得る。

【0391】

ソフトウェアの全てまたは一部は、時に、Internetまたは様々な他の電気通信ネットワークを通して通信され得る。そのような通信は、例えば、1つのコンピュータまたはプロセッサからのソフトウェアを別のコンピュータまたはプロセッサにロードすること、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ロードすることを可能にし得る。このように、ソフトウェア要素を有し得る別のタイプの媒体は、光波、電波、および電磁波、例えば有線および光学的固定電話ネットワーク、ならびに様々なエアリンクを通してローカルデバイス間の物理的インターフェースを越えて使用される媒体を含む。そのような波動を伝える物理的要素、例えば有線または無線リンク、光学的リンクなどもまた、ソフトウェアを有する媒体であると考えられ得る。本明細書で使用される場合、非一時的有形「ストレージ」媒体に限定されない限り、コンピュータまたは機械「可読媒体」などの用語は、実行のためにプロセッサに命令を提供することに関与する任意の媒体を指す。

【0392】

したがって、機械可読媒体、例えばコンピュータ実行可能コードは、これらに限定されないが有形ストレージ媒体、搬送波媒体、または物理的伝送媒体を含む多くの形態をとり得る。不揮発性ストレージ媒体は、図面に示されるように、例えば、光学または磁気ディスク、例えば任意のコンピュータなどにおけるストレージデバイスのいずれか、例えばデータベースをインプリメントするために使用され得るものなどを含む。揮発性ストレージ媒体は、ダイナミックメモリー、例えばそのようなコンピュータプラットフォームのメインメモリーを含む。有形伝送媒体は、同軸ケーブル；コンピュータシステム内のバスを含むワイヤーを含む銅線および光ファイバーを含む。搬送波伝送媒体は、電気もしくは電磁シグナル、または音波もしくは光波の形態、例えば高周波（RF）および赤外線（IR）データ通信の間に生成される形態をとり得る。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もしくはDVD-ROM、任意の他の光学媒体、パンチカード、紙テープ、穴パターンを有する任意の他の物理的ストレージ媒体、RAM、ROM、PROMおよびEPROM、FLASH-EPROM、任意の他のメモリーチップもしくはカートリッジ、搬送波輸送データもしくは命令、そのような搬送波を輸送するケーブルもしくはリンク、またはそこからコンピュータがプログラミングコードおよび／もしくはデータを読み取り得る任意の他の媒体を含む。コンピュータ可読媒体のこれらの形態の多くは、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伝えることに関係し得る。

【0393】

コンピュータシステム501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ーフェース（UI）540を含む電子ディスプレイ535を含むか、またはそれと通信することができる。UIの例としては、これらに限定されないが、グラフィカルユーザーインターフェース（GUI）およびウェブに基づくユーザーインターフェースが挙げられる。

【0394】

コンピュータシステムおよびネットワーク、データベース、ならびにコンピュータプログラム生成物に関する追加の詳細はまた、例えば、その各々の全体がこれにより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 Net

10

20

30

40

50

working: 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)にも提供されている。

G. 応用

1. がんおよび他の疾患

【0395】

本方法は、対象における状態、特に、がんの存在を診断するため、状態を特徴付ける（例えば、がんをステージ分類するかまたはがんの不均一性を決定する）ため、状態の処置に対する応答をモニタリングするため、状態を発症するリスクまたは状態の後続の経過の予後判定を達成するために、使用することができる。本開示はまた、特定の処置選択肢の有効性を決定するのにとも有用であり得る。奏功した処置選択肢では、処置が成功している場合、より多くのがんが死滅し、DNAが脱落し得るため、対象の血液中で検出されるコピー数変異または希少な突然変異の量が増加し得る。他の例では、これは生じない場合がある。別の例では、おそらくはある特定の処置選択肢は、経時的に、がんの遺伝子プロファイルと相関付けられ得る。この相関性は、治療を選択するのに有用であり得る。一部の実施形態では、高メチル化可変エピジェネティック標的領域は、それらが、腫瘍細胞もしくは通常はcfDNAに有意に寄与しない細胞の高メチル化特徴を示すかどうかを決定する、および／または低メチル化可変エピジェネティック標的領域は、それらが、腫瘍細胞もしくは通常はcfDNAに有意に寄与しない細胞の低メチル化特徴を示すかどうかを決定するために解析される。

【0396】

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

【0397】

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、核酸変異体が体細胞または生殖系列起源であるという分類に基づいて、患者における所定の疾患または状態を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同定してもよい。典型的に、検討中の疾患は、あるタイプのがんである。そのようながんの非限定的な例としては、胆管がん、膀胱がん、移行上皮癌、尿路上皮癌、脳がん、神経膠腫、星細胞腫、乳癌、化生癌、子宮頸がん、子宮頸部扁平上皮癌、直腸がん、結腸直腸癌、結腸がん、遺伝性非ポリポーシス大腸がん、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、明細胞腎細胞癌、移行上皮癌、尿路上皮癌、ウィルムス腫瘍、白血病、急性リンパ球性白血病（ALL）、急性骨髄性白血病（AML）、慢性リンパ球性白血病（CLL）、慢性骨髄性白血病（CML）、慢性骨髄単球性白血病（CMML）、肝臓がん、肝臓癌、肝腫、肝細胞癌、胆管癌、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細胞リンパ腫、マントル細胞リンパ腫、T細胞リンパ腫、非ホジキンリンパ腫、前駆Tリンパ芽球性リンパ腫／白血病、末梢性T細胞リンパ腫、多発性骨髄腫、鼻咽頭癌（NPC）、神経芽腫、中咽頭がん、口腔扁平上皮癌、骨肉腫、卵巣癌、膵臓がん、膵管腺癌、偽乳頭状腫瘍、膵腺房細胞癌、前立腺がん、前立腺腺癌、皮膚がん、黒色腫、悪性黒色腫、皮膚黒色腫、小腸癌、胃がん、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。がんのタイプおよび／またはステージは、突然変異、まれな突然変異、インデル、コピー数変異、塩基転換、転座、逆位、欠失、異数性、部分的異数性、倍数性、染色体不安定性、染色体構造の変更、遺伝子融合、染色体融合、遺伝子トランケーション、遺伝子増幅、遺伝子重複、染色体病変、DNA病変、核酸化学修飾の異常な変化、エ

10

20

30

40

50

ピジェネティックパターンの異常な変化、および核酸5-メチルシトシンの異常な変化を含む遺伝的変異から検出することができる。

【0398】

遺伝子データはまた、がんの特定の形態を特徴付けるためにも使用することができる。がんはしばしば、組成およびステージ分類の両方が不均一である。遺伝子プロファイルデータは、その特定のサブタイプの診断または処置において重要であり得るがんの特定のサブタイプの特徴付けを可能にし得る。この情報はまた、がんの特定のタイプの予後に関する手がかりを対象または従事者に提供し、対象または従事者は、疾患の進行に従って処置選択肢を適合させることを可能にし得る。一部のがんは、進行してより侵襲性となり、遺伝的に不安定となり得る。他のがんは、良性、不活性、または休眠中のままであり得る。本開示のシステムおよび方法は、疾患進行を決定するために有用であり得る。

10

【0399】

さらに、本開示の方法を使用して、対象における異常な状態の不均一性を特徴付けてもよい。そのような方法は、例えば、対象に由来する細胞外ポリヌクレオチドの遺伝的プロファイルを作成するステップであって、遺伝的プロファイルが、コピー数変異およびまれな突然変異の解析に起因する複数のデータを含む、ステップを含み得る。一部の実施形態では、異常な状態はがんである。一部の実施形態では、異常な状態は、不均一なゲノム集団をもたらす状態であり得る。がんの例では、一部の腫瘍は、がんの異なるステージの腫瘍細胞を含むことが公知である。他の例では、不均一性は、疾患の複数の病巣を構成し得る。この場合も、がんの例では、複数の腫瘍病巣が存在してもよく、おそらく1つまたは複数の病巣は、原発部位から広がった転移の結果である。

20

【0400】

本方法を使用して、不均一な疾患における異なる細胞に由来する遺伝情報の要約であるデータのフィンガープリントまたはセットを作成またはプロファイリングすることができる。このデータのセットは、コピー数変異、エピジェネティック変異、および突然変異解析を単独または組み合わせて含み得る。

【0401】

本方法を使用して、がんまたは他の疾患を診断、予後判定、モニタリング、または観察することができる。一部の実施形態では、本明細書の方法は、胎児を診断することも、予後判定することも、モニタリングすることも伴わず、そのため、非侵襲性の出生前検査を対象としない。他の実施形態では、これらの方法論を、妊娠中の対象に用いて、そのDNAおよび他のポリヌクレオチドが、母体の分子と共循環し得るまだ生まれていない対象におけるがんまたは他の疾患を診断、予後判定、モニタリング、または観察してもよい。

30

【0402】

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乏症、抗リン脂質抗体症候群、自閉症、常染色体優性多発性嚢胞腎、シャルコーマリーートゥース病（CMT）、クリデュチャット病、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、全前脳胞症、ハンチントン病、クライنفelter症候群、マルファン症候群、筋強直性ジストロフィー、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、ポーランド症候群、ポルフィリン症、早老症、網膜色素変性症、重度複合免疫不全（SCID）、鎌状赤血球症、脊髄性筋萎縮症、テイサックス病、サラセミア、トリメチルアミン尿症、ターナー症候群、口蓋心臓顔面症候群、WAGR症候群、ウィルソン病などが挙げられる。

40

【0403】

一部の実施形態では、本明細書に記載される方法は、本明細書に記載されるように得られた配列情報のセットを使用してがんを有することが以前に診断された対象の以前のがん処置後の予め選択した時点で、腫瘍細胞を起源とするまたは腫瘍細胞に由来するDNAの

50

存在または非存在を検出するステップを含む。方法は、試験対象に関して腫瘍細胞を起源とするまたは腫瘍細胞に由来するDNAの存在または非存在を示すがん再発スコアを決定するステップをさらに含み得る。

【0404】

がん再発スコアが決定される場合、これをさらに使用してがん再発ステータスを決定してもよい。例えば、がん再発スコアが既定の閾値より上である場合、がん再発ステータスは、がん再発のリスクがあり得る。例えば、がん再発スコアが既定の閾値より上である場合、がん再発ステータスは、がんの再発のリスクが低いまたはより低いリスクであり得る。特定の実施形態では、既定の閾値に等しいがん再発スコアは、がん再発のリスクがあるか、またはがん再発のリスクが低いもしくはより低いリスクであるいずれかのがん再発ステータスをもたらし得る。

10

【0405】

一部の実施形態では、がん再発スコアを既定のがん再発閾値と比較し、がん再発スコアががん再発閾値より上である場合、試験対象はその後のがん処置の候補であると分類され、がん再発スコアががん再発閾値より下である場合、試験対象は治療の候補ではないと分類される。特定の実施形態では、がん再発閾値に等しいがん再発スコアは、その後のがん処置の候補であるか、または治療の候補ではないといういずれかの分類をもたらし得る。

【0406】

上記で考察した方法は、試験対象におけるがん再発のリスクを決定するおよび／または試験対象がその後のがん処置の候補であると分類する方法に関する節を含む、本明細書の

20

他所で記載した任意の適合性の特色（単数または複数）をさらに含み得る。

2. 試験対象におけるがん再発のリスクを決定するおよび／または試験対象がその後のがん処置の候補であると分類する方法

【0407】

一部の実施形態では、本明細書に提供される方法は、試験対象におけるがん再発のリスクを決定する方法である。一部の実施形態では、本明細書に提供される方法は、試験対象をその後のがん処置の候補であると分類する方法である。

【0408】

そのような方法のいずれかは、試験対象に対する1つまたは複数の以前のがん処置の後の1つまたは複数の事前に選択された時点で、がんを有すると診断された試験対象からDNA（例えば、腫瘍細胞を起源とするまたは腫瘍細胞に由来する）を収集するステップを含み得る。対象は、本明細書に記載される対象のいずれかであり得る。DNAはcfDNAであり得る。DNAは、組織試料から得られ得る。

30

【0409】

そのような方法のいずれかは、対象のDNAから複数の標的領域のセットを捕捉するステップであって、複数の標的領域セットが、配列可変標的領域セットおよびエピジェネティック標的領域セットを含み、それによって捕捉されたDNA分子のセットが産生される、ステップを含み得る。捕捉するステップは、本明細書の他所で記載される実施形態のいずれかに従って実施され得る。

【0410】

そのような方法のいずれかにおいて、以前のがん処置は、手術、治療組成物の投与、および／または化学療法を含み得る。

40

【0411】

そのような方法のいずれかは、捕捉されたDNA分子をシーケンシングすることであって、それによって配列情報のセットが産生されることを含み得る。配列可変標的領域セットの捕捉されたDNA分子は、エピジェネティック標的領域セットの捕捉されたDNA分子よりもシーケンシングの大きい深度までシーケンシングされ得る。

【0412】

そのような方法のいずれかは、配列情報のセットを使用して事前に選択した時点で腫瘍細胞を起源とするまたは腫瘍細胞に由来するDNAの存在または非存在を検出するステッ

50

ブを含み得る。腫瘍細胞を起源とするまたは腫瘍細胞に由来するDNAの存在または非存在の検出は、本明細書の他所で記載されるその実施形態のいずれかに従って実施され得る。

【0413】

試験対象においてがん再発のリスクを決定する方法は、試験対象の腫瘍細胞を起源とするまたは腫瘍細胞に由来するDNAの存在もしくは非存在、または量を示すがん再発スコアを決定するステップを含み得る。がん再発スコアをさらに使用して、がん再発ステータスを決定してもよい。がん再発ステータスは、例えばがん再発スコアが既定の閾値より上である場合、がん再発のリスクがあり得る。がん再発ステータスは、例えばがん再発スコアが既定の閾値より上である場合、がん再発のリスクが低いまたはより低いリスクであり得る。特定の実施形態では、既定の閾値に等しいがん再発スコアは、がん再発のリスクがあるか、またはがん再発のリスクが低いもしくはより低いリスクであるいずれかのがん再発ステータスをもたらし得る。

10

【0414】

試験対象をその後のがん処置の候補であると分類する方法は、試験対象のがん再発スコアを既定のがん再発閾値と比較して、それによってがん再発スコアががん再発閾値より上である場合、試験対象をその後のがん処置の候補であると分類するステップ、またはがん再発スコアががん再発閾値より下である場合、治療の候補ではないと分類するステップを含み得る。特定の実施形態では、がん再発閾値に等しいがん再発スコアは、その後のがん処置の候補であるか、または治療の候補ではないといういずれかの分類をもたらし得る。一部の実施形態では、その後のがん処置は、化学療法または治療組成物の投与を含む。

20

【0415】

そのような方法のいずれかは、がん再発スコアに基づいて試験対象の無病生存（DFS）期間を決定するステップを含み得、例えば、DFS期間は、1年、2年、3年、4年、5年、または10年であり得る。

【0416】

一部の実施形態では、配列情報のセットは、配列可変標的領域配列を含み、がん再発スコアを決定するステップは、配列可変標的領域配列に存在するSNV、挿入／欠失、CNV、および／または融合の量を示す少なくとも第1のサブスコアを決定するステップを含み得る。

30

【0417】

一部の実施形態では、1個、2個、3個、4個、または5個から選択される配列可変標的領域における突然変異の数は、第1のサブスコアが、がん再発に関して陽性であると分類されるがん再発スコアをもたらし十分に十分である。一部の実施形態では、突然変異の数は、1個、2個、または3個から選択される。

【0418】

一部の実施形態では、配列情報のセットは、エピジェネティック標的領域配列を含み、がん再発スコアを決定することは、健康な対象に由来する対応する試料において見出されるDNA（例えば、健康な対象に由来する血液試料において見出されるcfDNA、または健康な対象に由来する組織試料において見出されるDNA、組織試料は、試験対象から得られたものと同じタイプの組織のものである）とは異なるエピジェネティック状態を表す分子（エピジェネティック標的領域配列から得られる）の量を示す、第2の部分スコアを決定することを含む。これらの異常な分子（すなわち、健康な対象に由来する対応する試料において見出されるDNAとは異なるエピジェネティック状態を有する分子）は、がんに関連するエピジェネティック変化、例えば、高メチル化可変標的領域のメチル化、および／または断片化可変標的領域の断片化の乱れと一致し得、ここで、「乱れ」は、健康な対象に由来する対応する試料において見出されるDNAとは異なることを意味する。

40

【0419】

一部の実施形態では、0.001%～10%の範囲の値よりも大きいまたはそれに等しい、高メチル化可変標的領域セットにおける高メチル化および／または断片化可変標的領域

50

セットにおける異常な断片化を示す高メチル化可変標的領域セットおよび／または断片化可変標的領域セットに対応する分子の割合は、第2の部分スコアががん再発に関して陽性として分類されるのに十分である。範囲は、0.001%～1%、0.005%～1%、0.01%～5%、0.01%～2%、または0.01%～1%であり得る。

【0420】

一部の実施形態では、そのような方法のいずれかは、腫瘍細胞からの起源を示す1つまたは複数の特色を示す配列情報のセット中の分子の分画から腫瘍DNAの分画を決定するステップを含み得る。これは、例えば、高メチル化可変標的領域および断片化可変標的領域（高メチル化可変標的領域の高メチル化および／または断片化可変標的領域の異常な断片化は、腫瘍細胞からの起源を示すと考えられ得る）の一方または両方を含む、エピジェネティック標的領域の一部または全てに対応する分子について行われ得る。これは、配列可変標的領域に対応する分子、例えばがんと一貫する変更、例えばSNV、インデル、CNV、および／または融合を含む分子について行われ得る。腫瘍DNAの分画は、エピジェネティック標的領域に対応する分子および配列可変標的領域に対応する分子の組合せに基づいて決定され得る。

【0421】

がん再発スコアの決定は、腫瘍DNAの分画に少なくとも部分的に基づき得、 10^{-1} ～1または 10^{-10} ～1の範囲の閾値よりも大きい腫瘍DNAの分画は、がん再発スコアががん再発に関して陽性であると分類されるために十分である。一部の実施形態では、 10^{-10} ～ 10^{-9} 、 10^{-9} ～ 10^{-8} 、 10^{-8} ～ 10^{-7} 、 10^{-7} ～ 10^{-6} 、 10^{-6} ～ 10^{-5} 、 10^{-5} ～ 10^{-4} 、 10^{-4} ～ 10^{-3} 、 10^{-3} ～ 10^{-2} 、または 10^{-2} ～ 10^{-1} の範囲の閾値よりも大きいまたはそれに等しい腫瘍DNAの分画は、がん再発スコアが、がん再発に関して陽性であると分類されるために十分である。一部の実施形態では、少なくとも 10^{-7} の閾値よりも大きい腫瘍DNAの分画は、がん再発スコアが、がん再発に関して陽性であると分類されるために十分である。腫瘍DNAの分画が閾値、例えば前述の実施形態のいずれかに対応する閾値よりも大きいという決定は、累積確率に基づいて行われ得る。例えば、試料は、腫瘍分画が前述の範囲のいずれかにおける閾値よりも大きい累積確率が、少なくとも0.5、0.75、0.9、0.95、0.98、0.99、0.995、または0.999の確率閾値を超える場合に陽性であるとみなされた。一部の実施形態では、確率閾値は、少なくとも0.95、例えば0.99である。

【0422】

一部の実施形態では、配列情報のセットは、配列可変標的領域配列およびエピジェネティック標的領域配列を含み、がん再発スコアを決定するステップは、配列可変標的領域配列に存在するSNV、挿入／欠失、CNVおよび／または融合の量を示す第1のサブスコア、ならびにエピジェネティック標的領域配列中の異常な分子の量を示す第2のサブスコアを決定するステップ、ならびに第1および第2のサブスコアを組み合わせるがん再発スコアを提供するステップを含む。第1および第2のサブスコアを組み合わせる場合、それらを、独立して各サブスコア（例えば、配列可変標的領域における既定の数の突然変異（例えば、>1）よりも大きく、エピジェネティック標的領域中の異常な分子（すなわち、健康な対象からの対応する試料中に見出されるDNAとは異なるエピジェネティック状態を有する分子、例えば腫瘍）の既定の分画よりも大きい）に閾値を適用するか、または機械学習分類器を訓練して、複数の陽性および陰性訓練試料に基づいて状態を決定することによって組み合わせてもよい。

【0423】

一部の実施形態では、-4～2または-3～1の範囲の組み合わせたスコアの値は、がん再発スコアががん再発に関して陽性であると分類されるために十分である。

【0424】

がん再発スコアががん再発に関して陽性であると分類される任意の実施形態では、対象のがん再発ステータスはがん再発のリスクがあり得る、および／または対象はその後のが

ん処置の候補であると分類され得る。

【0425】

一部の実施形態では、がんは、本明細書の他所で記載されるがんのタイプのいずれか1つ、例えば結腸直腸がんである。

3. 治療および関連する投与

【0426】

ある特定の実施形態では、本明細書に開示される方法は、核酸変異体の状態が体細胞または生殖系列起源であることを考慮して、カスタマイズされた治療を同定すること、およびそれを患者に投与することに関する。一部の実施形態では、本質的に任意のがん治療（例えば、外科治療、放射線治療、化学療法および／またはそれと同様のもの）を、これらの方法の一部として含めてもよい。典型的に、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。免疫療法は一般的に、所定のがんタイプに対する免疫応答を増強する方法を指す。ある特定の実施形態では、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。

10

【0427】

ある特定の実施形態では、体細胞または生殖系列起源である対象由来の試料からの核酸変異体の状態を、その対象のカスタマイズされたまたは標的化治療を同定するために参照集団からの比較器の結果のデータベースと比較してもよい。典型的に、参照集団は、試験対象と同じがんもしくは疾患タイプを有する患者、ならびに／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ことがある患者を含む。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）は、核酸変異体および比較器の結果がある特定の分類基準（例えば、実質的またはおおよそマッチする）を満たす場合に同定され得る。

20

【0428】

ある特定の実施形態では、本明細書に記載されるカスタマイズされた治療は、典型的に非経口（例えば、静脈内、または皮下）投与される。免疫療法剤を含有する医薬組成物は、典型的に静脈内投与される。ある特定の治療剤は経口投与される。しかし、カスタマイズされた治療（例えば、免疫療法剤など）はまた、例えば口腔内、舌下、直腸、膣、尿道内、局所、眼内、鼻腔内、および／または耳介内などの方法によって投与されてもよく、投与は、錠剤、カプセル剤、顆粒剤、水性懸濁剤、ゲル剤、スプレー剤、坐剤、塗擦剤、軟膏剤などを含み得る。

30

【0429】

本発明の好ましい実施形態が本明細書において示され、説明されているが、そのような実施形態が単なる例として提供されることは、当業者には明らかであろう。本発明が本明細書内に提供されている具体的な例によって制限されることは、意図されない。本発明は、前述の明細書を参照して記載されているが、本明細書における実施形態の説明および例示は、制限的な意味で解釈されることを意味するものではない。当業者であれば、本発明から逸脱することなく、多数の変化形、変更、および置換を想起するであろう。さらに、本発明の全ての態様は、本明細書において記載されている具体的な説明、構成、または相対的割合に制限されず、それらは、様々な条件および変数に依存することを理解すべきである。本明細書に記載される本開示の実施形態に対する様々な代替形態を、本開示の実施に用いることができることを理解すべきである。したがって、本開示はまた、任意のそのような代替形態、改変形態、変動形態、または等価物を網羅するものとするのが企図される。以下の特許請求の範囲が本発明の範囲を定めること、ならびにこれらの特許請求の範囲内の方法および構造、ならびにそれらの等価物が、それによって網羅されることが意図される。

40

【0430】

前述の開示は、明確さおよび理解の目的で、例証および例によって一部詳細に説明されているが、当業者には、本開示を読むことによって、形態および細部における様々な変化が、本開示の真の範囲から逸脱することなく行われ得、添付の特許請求の範囲内で実施され得ることは明白である。例えば、全ての方法、システム、コンピュータ可読媒体、およ

50

び／または構成要素の特色、ステップ、要素、もしくはこれらの他の態様を、様々な組合せで使用することができる。

H. キット

【0431】

また、本明細書に記載される組成物を含む、キットも提供される。キットは、本明細書に記載される方法を行うのに有用であり得る。一部の実施形態では、キットは、本明細書に記載されるように、試料を複数の部分試料に分配するための第1の試薬、例えば、本明細書の他の箇所に記載される分配試薬のいずれかを含む。一部の実施形態では、キットは、第1の部分試料を、第1の部分試料のDNA内の第1の核酸塩基に対してDNA内の第2の核酸塩基とは異なって影響を及ぼす手順に供するための第2の試薬を含み、ここで、第1の核酸塩基は、改変または非改変核酸塩基であり、第2の核酸塩基は、第1の核酸塩基とは異なる改変または非改変核酸塩基であり、第1の核酸塩基および第2の核酸塩基は、同じ塩基対形成特異性を有する（例えば、核酸塩基、例えば、シトシンまたはメチル化シトシンを異なる核酸塩基に変換するための本明細書の他の箇所に記載される試薬のうちのいずれか）。キットは、第1および第2の試薬、ならびに以下および／または本明細書の他の箇所において考察される追加の要素を含み得る。

10

【0432】

キットは、さらに、ALK、APC、BRAF、CDKN2A、EGFR、ERBB2、FBXW7、KRAS、MYC、NOTCH1、NRAS、PIK3CA、PTEN、RBI、TP53、MET、AR、ABLL、AKTL、ATM、CDHL、CSFIR、CTNNBL、ERBB4、EZH2、FGFRL、FGFR2、FGFR3、FLT3、GNA11、GNAQ、GNAS、HNF1A、HRAS、IDH1、IDH2、JAK2、JAK3、KDR、KIT、MLH1、MPL、NPM1、PDGFRA、PROG、PTPN11、RET、SMAD4、SMARCB1、SMO、SRC、STK11、VHL、TERT、CCND1、CDK4、CDKN2B、RAF1、BRCA1、CCND2、CDK6、NF1、TP53、ARID1A、BRCA2、CCNE1、ESR1、RIT1、GATA3、MAP2K1、RHEB、ROS1、ARAF、MAP2K2、NFE2L2、RHOA、およびNTRK1からなる群から選択される少なくとも5個、6個、7個、8個、9個、10個、20個、30個、40個または全ての遺伝子に選択的にハイブリダイズする複数のオリゴヌクレオチドプローブを含み得る。オリゴヌクレオチドプローブが選択的にハイブリダイズすることができる遺伝子の数は、異なり得る。例えば、遺伝子の数は、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、11個、12個、13個、14個、15個、16個、17個、18個、19個、20個、21個、22個、23個、24個、25個、26個、27個、28個、29個、30個、31個、32個、33個、34個、35個、36個、37個、38個、39個、40個、41個、42個、43個、44個、45個、46個、47個、48個、49個、50個、51個、52個、53個、または54個を含み得る。キットは、本明細書に記載される方法のいずれかを実施するための複数のオリゴヌクレオチドプローブを含む容器および使用説明書を含み得る。

20

30

【0433】

オリゴヌクレオチドプローブは、遺伝子、例えば少なくとも5個の遺伝子のエクソン領域に選択的にハイブリダイズすることができる。一部の例では、オリゴヌクレオチドプローブは、遺伝子、例えば少なくとも5個の遺伝子の少なくとも30個のエクソンに選択的にハイブリダイズすることができる。一部の例では、複数のプローブは、少なくとも30個のエクソンの各々に選択的にハイブリダイズすることができる。各エクソンにハイブリダイズするプローブは、少なくとも1つの他のプローブと重複する配列を有し得る。一部の実施形態では、オリゴプローブは、本明細書に開示される遺伝子の非コード領域、例えば遺伝子のイントロン領域に選択的にハイブリダイズすることができる。オリゴプローブはまた、本明細書に開示される遺伝子のエクソンおよびイントロン領域の両方を含む遺伝子の領域に選択的にハイブリダイズすることもできる。

40

50

【0434】

任意の数のエクソンを、オリゴヌクレオチドプローブによって標的化することができる。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、10個、11個、12個、13個、14個、15個、16個、17個、18個、19個、20個、21個、22個、23個、24個、25個、30個、35個、40個、45個、50個、55個、60個、65個、70個、75個、80個、85個、90個、95個、100個、105個、110個、115個、120個、125個、130個、135個、140個、145個、150個、155個、160個、165個、170個、175個、180個、185個、190個、195個、200個、205個、210個、215個、220個、225個、230個、235個、240個、245個、250個、255個、260個、265個、270個、275個、280個、285個、290個、295個、300個、400個、500個、600個、700個、800個、900個、1,000個、またはそれより多くのエクソンを標的化することができる。

10

【0435】

キットは、別個の分子バーコードおよび同一の試料バーコードを有する少なくとも4個、5個、6個、7個、または8個の異なるライブラリーアダプターを含み得る。ライブラリーアダプターは、シーケンシングアダプターでなくてもよい。例えば、ライブラリーアダプターは、フローセル配列またはシーケンシングのためのヘアピンループの形成を可能にする配列を含まない。異なる変異ならびに分子バーコードおよび試料バーコードの組合せが、本明細書全体を通して記載されており、キットに応用可能である。さらに、一部の例では、アダプターは、シーケンシングアダプターではない。さらに、キットに提供されるアダプターはまた、シーケンシングアダプターも含み得る。シーケンシングアダプターは、1つまたは複数のシーケンシングプライマーにハイブリダイズする配列を含み得る。シーケンシングアダプターはさらに、固体支持体にハイブリダイズする配列、例えばフローセル配列も含み得る。例えば、シーケンシングアダプターはフローセルアダプターであり得る。シーケンシングアダプターを、ポリヌクレオチド断片の一方または両方の末端に付着させることができる。一部の例では、キットは、別個の分子バーコードおよび同一の試料バーコードを有する少なくとも8個の異なるライブラリーアダプターを含み得る。ライブラリーアダプターは、シーケンシングアダプターでなくてもよい。キットはさらに、ライブラリーアダプターに選択的にハイブリダイズする第1の配列およびフローセル配列に選択的にハイブリダイズする第2の配列を有するシーケンシングアダプターも含み得る。別の例では、シーケンシングアダプターは、ヘアピン形状であり得る。例えば、ヘアピン形状のアダプターは、相補的二本鎖部分およびループ部分を含み得、二本鎖部分は、二本鎖ポリヌクレオチドに付着（例えば、ライゲーション）することができる。ヘアピン形状のシーケンシングアダプターを、ポリヌクレオチド断片の両方の末端に付着させて、環状分子を生成することができ、これを複数回シーケンシングすることができる。シーケンシングアダプターは、末端から末端まで最大10個、11個、12個、13個、14個、15個、16個、17個、18個、19個、20個、21個、22個、23個、24個、25個、26個、27個、28個、29個、30個、31個、32個、33個、34個、35個、36個、37個、38個、39個、40個、41個、42個、43個、44個、45個、46個、47個、48個、49個、50個、51個、52個、53個、54個、55個、56個、57個、58個、59個、60個、61個、62個、63個、64個、65個、66個、67個、68個、69個、70個、71個、72個、73個、74個、75個、76個、77個、78個、79個、80個、81個、82個、83個、84個、85個、86個、87個、88個、89個、90個、91個、92個、93個、94個、95個、96個、97個、98個、99個、100個、またはそれより多くの塩基であり得る。シーケンシングアダプターは、末端から末端まで20～30個、20～40個、30～50個、30～60個、40～60個、40～70個、50～60個、50～70個の塩基を含み得る。特定の例では、シーケンシングアダプターは、末端から末端まで20～30個の塩基を含み得る。別の例では、シーケンシングアダプターは、末端から末端ま

20

30

40

50

で50～60個の塩基を含み得る。シーケンシングアダプターは、1つまたは複数のバーコードを含み得る。例えば、シーケンシングアダプターは、試料バーコードを含み得る。試料バーコードは既定の配列を含み得る。試料バーコードを使用してポリヌクレオチドの起源を同定することができる。試料バーコードは、少なくとも1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、11個、12個、13個、14個、15個、16個、17個、18個、19個、20個、21個、22個、23個、24個、25個、またはそれより多くの（または本明細書全体を通して記載される任意の長さの）核酸塩基、例えば少なくとも8個の塩基であり得る。バーコードは、上記のように連続または不連続配列であり得る。

【0436】

ライブラリーアダプターは、平滑末端化され、Y字形状であり得、40核酸塩基未満またはそれに等しい長さであり得る。他の変形形態は、全体を通して見出すことができ、キットに応用可能である。

【実施例】

【0437】

III. 実施例

以下の実施例は、開示される方法のある特定の態様を例示するために提供される。実施例は、本開示を限定するものではない。

（実施例1）

腫瘍の存在／非存在を検出するためのcfDNAの解析

【0438】

患者試料のセットを、がんの存在／非存在を検出するために、Guardant Health (Redwood City, CA, USA)において血液に基づくNGSアッセイによって解析する。cfDNAを、これらの患者の血漿から抽出する。患者試料のcfDNAを、次いで、メチル結合ドメイン(MBD)緩衝液と組み合わせ、磁気ビーズをMBDタンパク質とコンジュゲートさせ、一晚インキュベートする。メチル化cfDNA（存在する場合、cfDNA試料中）は、このインキュベーション中に、MBDタンパク質に結合する。非メチル化であるか、またはあまり非メチル化であるDNAは、漸増濃度の塩を含有する緩衝液で、ビーズから洗い流される。最終的に、高塩緩衝液を使用して、MBDタンパク質から、高度メチル化DNAを洗い流す。これらの洗浄により、漸増メチル化cfDNAの3つの分配（低メチル化、残留メチル化、および高メチル化分配）が得られる。

【0439】

必要に応じて、高メチル化分配中のcfDNA分子を、酵素的改変(EM)に供し、それによって、非改変シトシンは、脱アミノ化を受けるが、mCおよびhmCはそれを受けず、それによって、非改変シトシンからウラシルへの変換により、第1の部分試料中の非特異的に分配された低メチル化分子に印を付ける。

【0440】

分配中のcfDNAを濃縮した後、分配されたcfDNAの末端オーバーハングを伸長させ、アデノシン残基を、伸長中にポリメラーゼによってcfDNA断片の3'末端に付加する。各断片の5'末端を、リン酸化する。これらの改変により、分配されたcfDNAがライゲーション可能となる。DNAリガーゼおよびアダプターを添加して、各分配されたcfDNA分子を、各末端でアダプターとライゲーションさせる。これらのアダプターは、非一意的分子バーコードを含有し、各分配を、他の分配において使用されるアダプター内のバーコードと識別可能な非一意的分子バーコードを有するアダプターとライゲーションさせる。

【0441】

低メチル化分配におけるcfDNAを、1つまたは複数のメチル化依存性ヌクレアーゼと接触させる。ヌクレアーゼは、低メチル化分配中の非特異的に分配されたDNAの少なくとも一部分を切断する。

10

20

30

40

50

【0442】

ライゲーションの後に、4つの分配を一緒にプールし、PCRによって増幅させる。1つまたは複数のメチル化依存性ヌクレアーゼによって切断された分子は、各末端にアダプターを有さないため、指数的増幅を受けない。

【0443】

PCR後、増幅したDNAを、富化の前に、洗浄し、濃縮する。濃縮した後、増幅したDNAを、塩緩衝液ならびに配列可変標的領域セットのプロープおよびエピジェネティック標的領域セットのプロープを含むビオチン化RNAプロープと組み合わせ、この混合物を、一晚インキュベートする。配列可変領域セットのプロープは、約50kbのフットプリントを有し、エピジェネティック標的領域セットのプロープは、約500kbのフットプリントを有する。配列可変標的領域セットのプロープは、表3～5において同定される遺伝子の少なくともサブセットを標的とするオリゴヌクレオチドを含み、エピジェネティック標的領域セットのプロープは、高メチル化可変標的領域、低メチル化可変標的領域、CTCF結合標的領域、転写開始部位標的領域、局所増幅標的領域、およびメチル化対照領域の選択物を標的とするオリゴヌクレオチドを含む。

10

【0444】

ビオチン化RNAプロープ（DNAにハイブリダイズしている）を、ストレプトアビジン磁気ビーズによって捕捉し、一連の塩に基づく洗浄によって捕捉されない増幅されたDNAから分離し、それによって、試料を富化する。富化後に、富化した試料のアリコート

を、Illumina NovaSeqシーケンサーを使用してシーケンシングする。シーケンサーによって生成された配列読み取りデータを、次いで、バイオインフォマティクツール／アルゴリズムを使用して解析する。分子バーコードは、一意的分子を同定するため、ならびに試料の差次的にMBD分配された分子へのデコンボリューションのために、使用する。この実施例に記載される方法はまた、その分配に基づく分子の全体的なメチル化レベル（すなわち、メチル化シトシン残基）に関する情報を提供することとは別に、低メチル化分配における非特異的に分配されたcfDNAの切断に起因する正確度および／または信頼度の増加を含め、高メチル化分配における非メチル化シトシンの変換に基づいてメチル化シトシンの位置に関するより高い分解能の情報も提供し得る。配列可変標的領域配列を、実際の腫瘍変異体を技術的エラー（例えば、PCRエラー、シーケンシングエラー）とは区別する十分な根拠でコールすることができるゲノム変更、例えば、SNV、挿入、欠失、および融合を検出することによって、解析する。エピジェネティック標的領域配列を独立して解析して、例えば、可能性としてがん性の組織において、健康なcfDNAと比較して、差次的にメチル化されることが示されている領域において、cfDNA分子のメチル化状態を検出する。最後に、両方の解析の結果を組み合わせ、最終的な腫瘍の存在／非存在のコールを生成する。

20

30

（実施例2）

健康な対象および初期結腸直腸がんを有する対象に由来するcfDNA試料における単一ヌクレオチド分解能におけるメチル化の解析

【0445】

健康な対象および初期結腸直腸がんを有する対象に由来するcfDNAの試料を、以下のように解析する。cfDNAを、MBDを使用して分配して、高メチル化分配、中間分配、および低メチル化分配を得た。各分配の分配されたDNAを、アダプターにライゲーションし、それによって、非改変シトシンは脱アミノ化を受けるが、mCおよびhmCは受けないEM-seq変換手順に供したが、代替的な手順において、低メチル化分配の分配されたDNAは、本明細書に記載されるメチル化依存性ヌクレアーゼと接触させてもよい。そのような脱アミノ化の後に、分配を、シーケンシングのために調製し、全ゲノムシーケンシングに供した。各分配を、別個にシーケンシングしたが、代替的な手順では、分配は、差次的にタグ付し（例えば、分配後でEM-seq変換前、または分配およびEM-seq変換後でシーケンシングのためのさらなる調製の前に）、プールし、プロセシングされ、並列してシーケンシングしてもよい。

40

50

【0446】

高メチル化可変標的領域からのた配列データを、バイオインフォマティクスによって単離したが、代替的な手順では、標的領域は、シーケンシングの前に *in vitro* で富化してもよい。高メチル化可変標的領域の塩基あたりのメチル化を、図7に示されるように定量し、これは、高メチル化分配に由来する高メチル化可変標的領域内の分子あたりのメチル化 CpG の数を示す。x 軸は、分子あたりの CpG の総数を示し、そのため、対角線に沿った点は、CpG ごとにメチル化を有する分子を表す。したがって、メチル化を、単一塩基分解能で解析し、MBD 分配材料の塩基あたりのメチル化および部分的な分子メチル化を定量することが可能であった。結腸直腸がんを有する対象に由来する試料は、これらの領域において、健康な対象に由来する試料よりもはるかに高い全体的なメチル化を示した。

10

(実施例3)

非特異的に分配された DNA の消化による技術的ノイズの低減

【0447】

2つの健康な正常試料に由来する cfDNA のプールを組み合わせ、そこから 18.6 ng を、本明細書に記載される MBD 分配アッセイへの入力として使用した。試料のサブセットに対して、結腸直腸がん試料 (CRC) に由来する cfDNA を、0.5% の MAF (変異型アレル分画) で添加し、0.16% の MAF を有する希釈 CRC 試料を得た。正常試料および希釈 CRC 試料の3つのセットを、アッセイにおいて使用した。3つのセットの試料を、次いで、MBD タンパク質を使用して、3つの分配 (高分配、残留分配、および低分配) に分配した。クリーンアップ後に、各分配中の cfDNA 分子を、分子バーコードを含む分配特異的アダプターとライゲーションさせた。高分配および残留分配に使用する分子バーコードは、MSRE 認識部位を有さず、そのため下流のプロセッシング (cfDNA メチル化状態とは無関係) において消化されないように、選択する。ライゲーション後に、ライゲーションクリーンアップを行った。ライゲーションクリーンアップの後に、高および残留分配を、MSRE 消化反応に供した。第1の試料セット (正常試料および希釈 CRC 試料) を、BstUI および HpaII で処置し、別の試料セットを、BstUI、HpaII、および Hin6I 酵素で処置した。第3の試料セットに、対照として、MBD 分配アッセイにおいて偽消化 (MSRE なし) を行った。MSRE 消化後に、酵素を、熱不活化させ (65°C、20 分間)、SPRI ビーズを使用してクリーンアップした。消化クリーンアップ後に、高、残留、および (非消化) 低分配 (アダプターライゲーションした cfDNA) を組み合わせ、PCR 増幅、目的のゲノム領域における分子の富化、試料をプールし、それによって、NovaSeq を使用したプールした試料の多重シーケンシングおよびシーケンシングを可能にすることを含む、NGS アッセイワークフローを通じてプロセッシングした。代替的な手順では、低分配を、1つまたは複数のメチル化依存性ヌクレアーゼ、例えば、本明細書に記載される MDRE のうちのいずれかと接触させて、低分配中の非特異的に分配された DNA を切断してもよい。

20

30

【0448】

図6は、MSRE 消化を適用した場合に、正常試料中の非メチル化分子からの技術的ノイズと比較して DMR でのがんメチル化シグナルの増加を明らかに示す。図6で示される陰性対照領域 (DNA 分子が疾患状態によらず、ほぼ全ての時間で非メチル化である) では、「a」は、MSRE 消化が高分配に誤って分配された非メチル化分子を除去することが明白であったことを明らかに示し、すなわち 90 個の分子が、偽消化では高分配に分配されたが、BstUI、HpaII、および Hin6I 消化では、分子数は 10 個に低減された。図6に示される分類 DMR では、cfDNA 分子は、MSRE による消化により正常試料 (b; 350 → 100) では希釈した CRC 試料 (c; 1500 → 1100) よりもかなり高い割合で除去された。

40

(実施例4)

MDRE 消化 cfDNA の解析

【0449】

50

2人の健康なドナーに由来するcfDNAの複数のアリコート単離し、メチル化cfDNAのMBDに基づく分配に供した。低メチル化cfDNA分配を、次いで、cfDNA分子へのNGSアダプターのライゲーシオンに供した。ライゲーシオンされた各ドナーに由来するcfDNAを、次いで、FspEI、LpnPI、MspJI、もしくはSgeIでのメチル化依存性制限酵素(MDRE)消化、または「偽」消化(酵素を消化に添加しない)、または対照反応としてMDRE反応を省略した非消化状態に供した。MDREステップの後に、低メチル化cfDNA分配を、MDREによって切断されているDNAが、アダプターが各末端に存在しないために指数的に増幅されない、ユニバーサルPCRにおいて増幅させた。PCR産物を、次いで、ハイブリッド捕捉パネルを使用して、標的とされるゲノム領域の富化に供し、第2のPCRにおいて増幅させ、NGSによってシーケンシングした。ハイブリッド捕捉パネル標的には、富化のためのゲノムの「陽性対照(c t r l)」および「陰性対照(c t r l)」領域が含まれる。陽性対照領域は、血液およびがん性組織を含む全てのヒト組織において、広範に高度にメチル化されている(バイサルファイト-seqによる>85%メチル化)ことが見出される、ゲノムのCpGが高密度な領域である。対照的に、陰性対照領域は、全てのヒト組織において広範に非メチル化である(<15%メチル化)。NGS解析により、全ての条件下においてシーケンシングされる陽性対照分子(すなわち、陽性対照領域内の分子)および陰性対照分子(すなわち、陰性対照領域内の分子)の数を、比較して、それぞれ、MDRE感度および特異度を推定する。図8A~Bは、FspEI酵素処置が、「偽」条件と比較して、陽性対照分子の数を>100倍低減させ、メチル化分子の消化に関して、約99%の感度を実証することを示す。図8C~Dは、FspEI処置が、陰性対照分子を有意義に低減させず、FspEI消化による高い特異度を示す(非メチル化分子を消化しない)ことを示す。MspJIは、いくらかの感度を示すが、FspEIと比較して特異度が不良である一方、LpnIおよびSgeIは、特異度をほとんど／まったく示さないことに留意されたい。

10

20

30

40

【0450】

MDRE消化効率を、異なる認識部位を有する分子および分子あたりの部位数を使用して計算した。消化効率は、 $1 - [\text{MDRE条件での陽性対照分子の数}] / [\text{偽条件での陽性対照分子の数}]$ として計算される。^{5m}CpGを含むFspEIの全般的な認識配列は、^{5m}C^{5m}CGH(H=A、C、またはT)であり、切断は12~16塩基下流で起こる。FspEIパリンドローム部位^{5m}CGGは、反対方向の上の鎖および下の鎖に2つのFspEI認識部位を含有する。全般的な^{5m}CpG含有コンセンサスは、^{5m}CpGNRであり、これはFspEIコンセンサスと重複し得る。図9A~Dは、消化効率が、分子あたり最少数の^{5m}C^{5m}CGHまたは^{5m}CGG部位によって増加し、パリンドローム部位(^{5m}CGG)でより効率的であることを示している。少なくとも1つの^{5m}CGGまたは少なくとも2つの^{5m}C^{5m}CGH部位を有する陽性対照分子は、95%の効率で切断された。

【0451】

さらに、FspEIおよびMspJIによる同時または逐次的消化を試験した。2つのMDRE(FspEIの後にMspJI)による逐次的消化は、最高の効率を有した。同時消化(FspEIおよびMspJI)では、MspJIは、時にDNAに結合するが、切断せず(より低い個々の効率)、このためFspEI活性を立体的に遮断することが起こり得る。ここでFspEIの後にMspJIは、FspEI単独より高い全効率を有するが、FspEI単独は、より良好な切断特異性を有する。このように、異なる状況では、FspEI単独による消化またはFspEIの後にMspJIによる消化が好ましくなり得る。最少部位の数が多い場合、より少ない陽性対照分子が観察されること(図9C~D)およびこのように消化効率の推定がよりノイズが多くなることに留意されたい。

(実施例5)

MDRE処置後の腫瘍DNAの検出

【0452】

4人の健康なドナーから単離したcfDNAを使用して、「正常」およびシミュレート

50

した「がん」cfDNA試料を作製した。ドナー試料を「正常」として未希釈のまま使用し、結腸直腸がん(CRC)患者のcfDNAでスパイクして、「がん」試料を作製した。CRC cfDNA試料の循環腫瘍DNA分画は、以前に測定されており、これを使用して正常なドナーcfDNAにCRC cfDNAの計算量をスパイクし、その結果、得られた「がん」試料は、0.5%の循環腫瘍DNAを含有した(図10A~Jにおける「0.5% CRC」)。全ての試料を、MBDに基づく分配に供し、cfDNAを高メチル化および低メチル化cfDNA分配に分割した。次いで、低メチル化cfDNA分配を、NGSアダプターにライゲーションした。次いで、各ドナーからのライゲーションしたcfDNAを、FspEI、MspJI、またはFspEI+MspJIのいずれかによるMDRE消化に供した。「偽消化」(消化反応に酵素を加えない)および「非消化」条件(MDRE反応を完全に省略)を、対照反応として役立てた。MDREステップ後、非消化低メチル化分配cfDNAを、ユニバーサルPCRにおいて増幅し、次いで、ハイブリッド捕捉パネルを使用して標的化ゲノム領域の富化に供し、次いで、2回目のPCRにおいて増幅し、NGSによってシーケンシングした。ハイブリッド捕捉パネル標的は、富化のためにゲノムの低メチル化可変標的領域および「陰性対照(ctrl)」領域を含む。陰性対照領域は、血液およびがん様組織を含む全てのヒト組織において広範に低度にメチル化される(バイサルファイト-seqによって<15%メチル化)ことが見出されたゲノムのCpGが高密度の領域である。低メチル化可変標的領域は、健康な結腸組織および血液と比較してCRC組織において低減されたメチル化パーセンテージを有すると文献で注釈されているゲノム領域である。NGS解析から、2つまたはそれより多くのCCGG部位を有する(MDREによって高い効率で消化されるはずである)低メチル化可変標的領域分子の数を、全ての消化条件にわたって「正常」と「がん」試料の間で比較する(図10A~E)。低メチル化可変標的領域分子数の比をまた、陰性対照分子数と比較し、これを、低メチル化可変標的領域分子数に影響を及ぼし得るcfDNA入力量を変化させるために正規化する(図10F~J)。低メチル化可変標的領域がんシグナルの分解可能な検出は、MDRE消化なしの条件(「非消化」および「偽消化」)では観察されなかった。すなわち、低メチル化可変標的領域分子および正規化された比のレベルは「がん」と「正常」試料の間で識別不能であった(有意差はなかった)(これは、図10C、E、H、およびJにおいて水平の矢印によって印される)。逆に、MDRE処置が存在する場合、低メチル化可変標的領域数および正規化した比のシフト(増加)が、「正常」試料と比較して「がん」試料において検出された(図10A、B、D、F、G、およびIにおいて上向きの右矢印によって印される)。このように、MDRE処置は、「がん」試料において0.5% CRC cfDNAでがん低メチル化可変標的領域シグナルの検出を可能にするが、これはMBD分配アッセイ単独では検出可能ではなかった。

【0453】

本発明の好ましい実施形態を本明細書に示し、記載してきたが、そのような実施形態が単なる例として提供されることは、当業者には明白である。本発明は、本明細書内に提供された具体的な例に限定されることは意図されない。本発明は、上記の明細書を参照して説明してきたが、本明細書の実施形態の説明および例証は、限定的な意味であると解釈されることを意味しない。多数の変形形態、変化、および置換が、本発明から逸脱することなく当業者に想起されるであろう。さらに、本発明の全ての態様は、多様な条件および変数に依存する、本明細書に記載される具体的な説明、構成、または相対的割合に限定されないと理解すべきである。本明細書に記載される本開示の実施形態に対する様々な代替を、本発明の実践において用いてもよいと理解すべきである。したがって、本開示はまた、任意のそのような代替、改変、変形形態、または等価物を網羅するものとするのが企図される。以下の特許請求の範囲は本発明の範囲を定義すること、これらの特許請求の範囲内に入る方法および構造、ならびにその等価物がそれによって網羅されることが意図される。

【0454】

前述の開示は、明快にするためおよび理解目的のために、例証および例によって一部詳

10

20

30

40

50

細に記載されているが、当業者には、本開示を読むことによって、形態および詳細の様々な変化が、本発明の真の範囲から逸脱することなく行われ得ること、ならびに添付の特許請求の範囲内で実践され得ることは明白である。例えば、全ての方法、システム、コンピュータ可読媒体、および／またはその構成要素の特色、ステップ、要素、もしくは他の態様は、様々な組合せで使用することができる。

【0455】

本明細書で引用される全ての特許、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、各々の個々の項目が具体的かつ個々に参照により組み込まれることが示されているのと同程度に全ての目的に関してその全体が参照により組み込まれる。配列の異なるバージョンが異なる時間で受託番号に関連している場合、本出願の有効な提出日で受託番号に関連するバージョンを意味する。有効な提出日は、実際の提出日、または該当する場合、受託番号を参照する優先権出願の提出日の早いほうを意味する。同様に、異なるバージョンの刊行物、ウェブサイトなどが異なる時期に公開されている場合、特に示していない限り、出願の有効な提出日で公開された最新のバージョンを意味する。

【図面】

【図1】

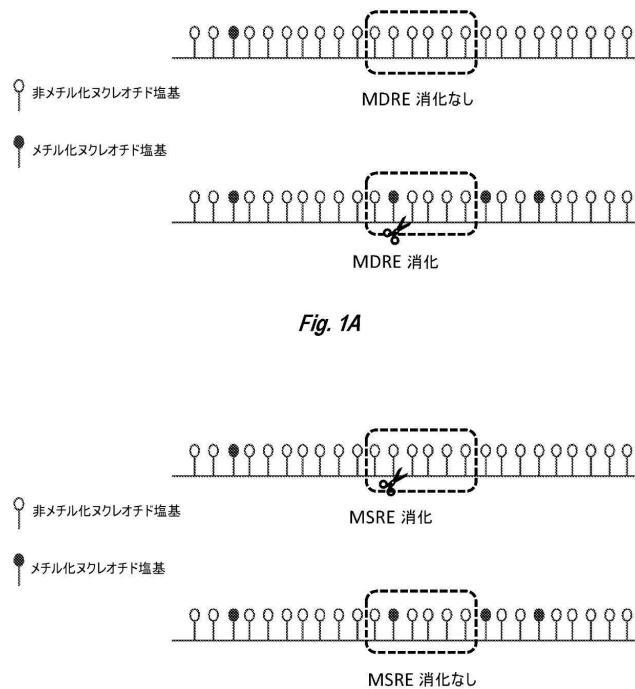


Fig. 1A

Fig. 1B

【図2】

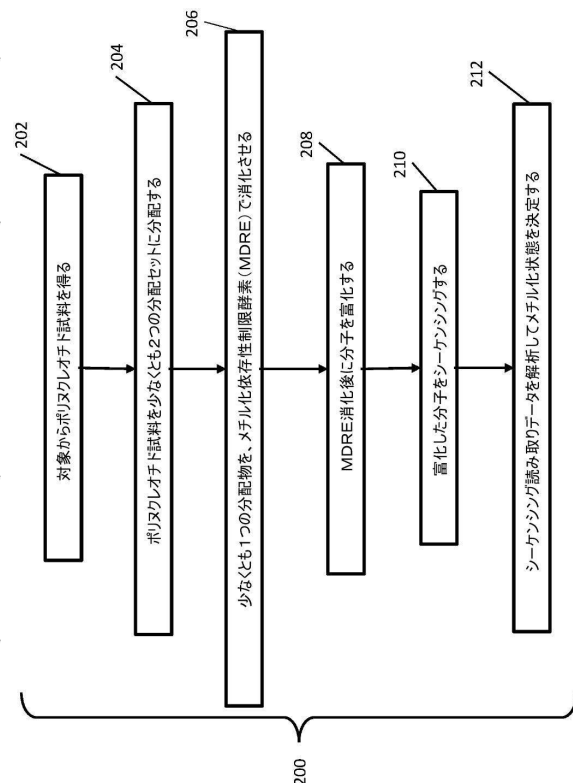


Fig. 2

【図 6】

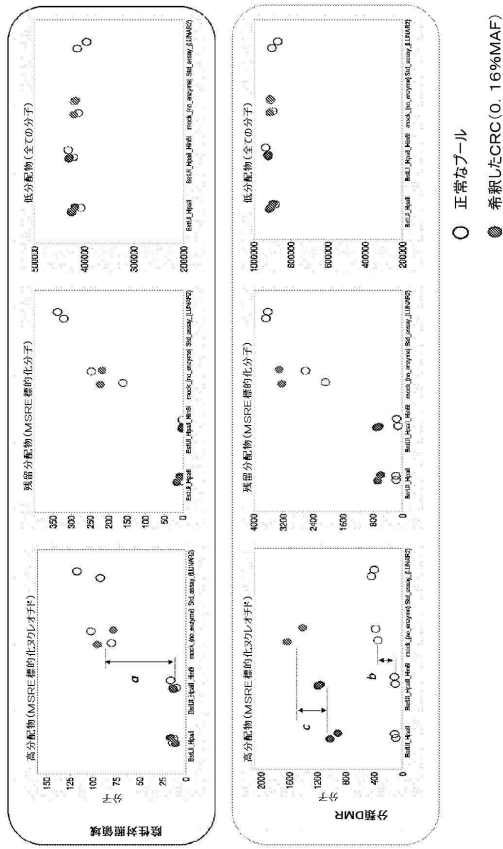


Fig. 6

【図 7】

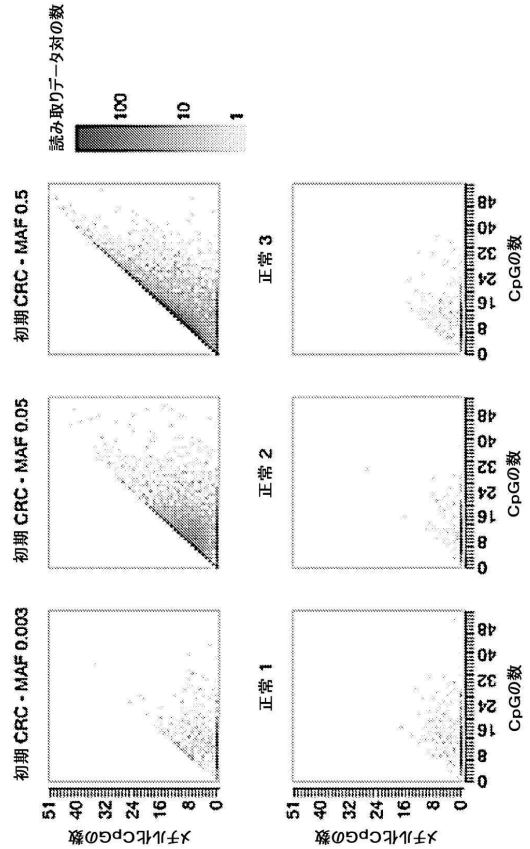


Fig. 7

【図 8 - 1】

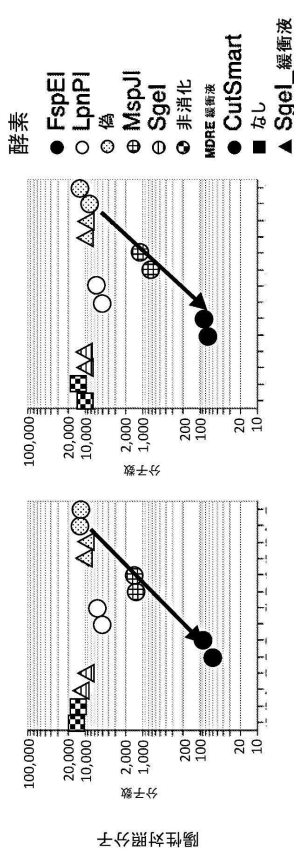


Fig. 8A

Fig. 8B

【図 8 - 2】

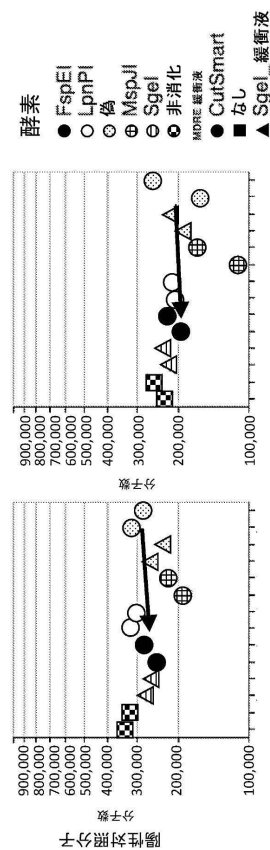


Fig. 8C

Fig. 8D

10

20

30

40

50

10

20

30

40

50

【図 9 - 1】

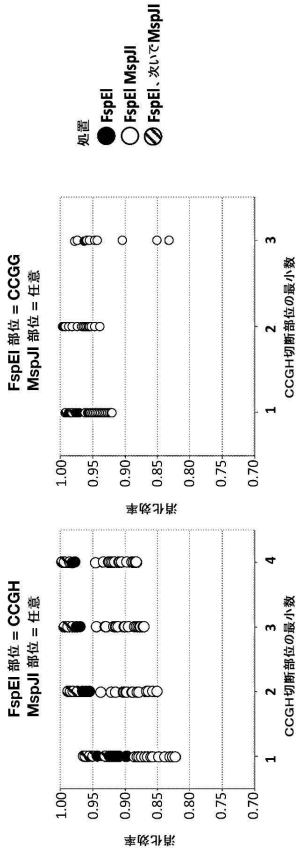


Fig. 9B

【図 9 - 2】

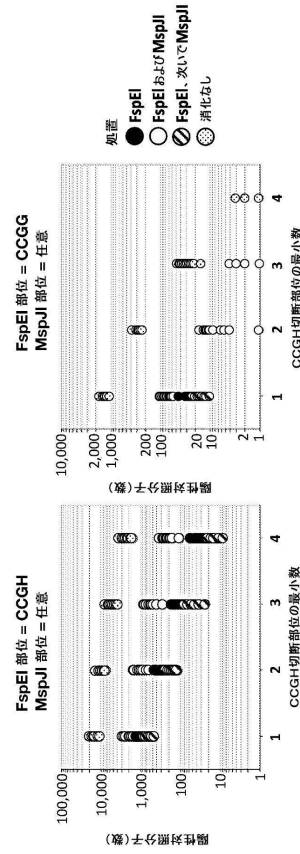


Fig. 9D

【図 10 - 1】

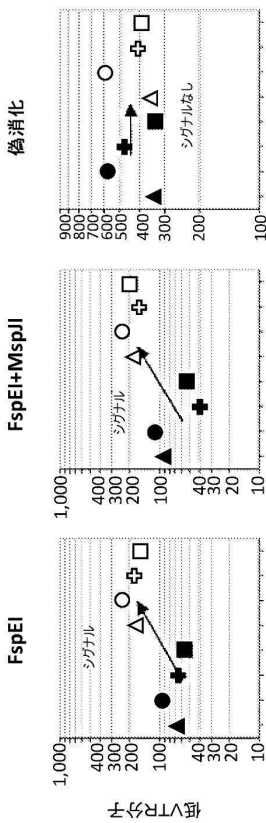


Fig. 10B

Fig. 10C

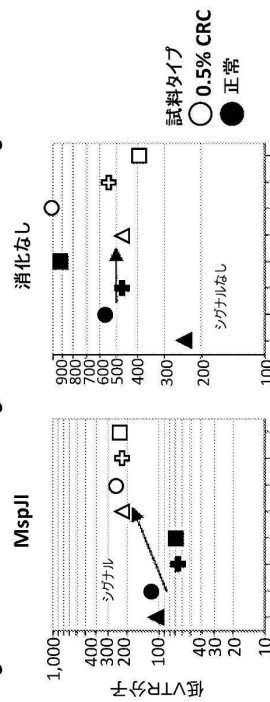


Fig. 10E

Fig. 10F

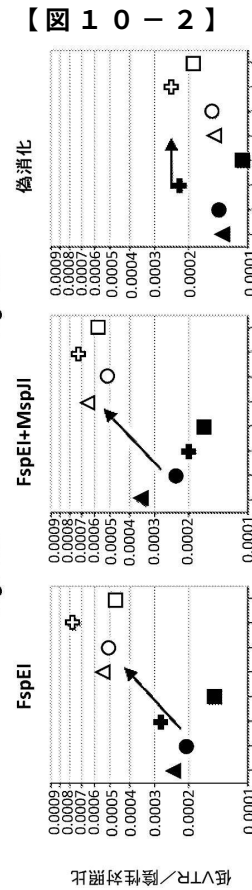


Fig. 10H

Fig. 10I

Fig. 10J

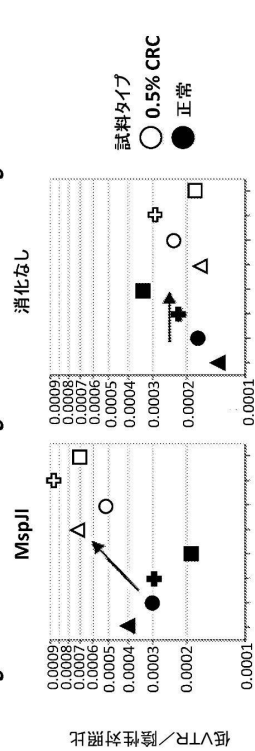


Fig. 10L

Fig. 10M

Fig. 10N



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/071649

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6804 C12Q1/6869
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2018/119452 A2 (GUARDANT HEALTH INC [US]) 28 June 2018 (2018-06-28) cited in the application abstract paragraphs [0156], [0187], [0212], [0238] examples 1,14,16 claims 1,14,20,29,37 -----	1-12, 23-26, 35-47, 54-62, 85-88
X	WO 2020/160414 A1 (GUARDANT HEALTH INC [US]) 6 August 2020 (2020-08-06) cited in the application abstract paragraphs [0237], [0409] - [0427] claims 1-2,9-10,19-21 figure 1 ----- -/-	1-90

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

11 January 2022

Date of mailing of the international search report

20/01/2022

Name and mailing address of the ISA/
European Patent Office, P.B. 5618 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Barz, Wolfgang

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/071649

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011/237444 A1 (CLANCY KEVIN [US] ET AL) 29 September 2011 (2011-09-29) abstract figure 1 claims 4-8, 15, 28-29, 33, 69 claims 1,10 -----	1-90
A	US 2018/305738 A1 (KENNEDY ANDREW [US] ET AL) 25 October 2018 (2018-10-25) abstract paragraph [0240] -----	1-90
A	WO 2017/190215 A1 (UNIV HEALTH NETWORK [CA]) 9 November 2017 (2017-11-09) abstract page 25 - page 26 -----	1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2021/071649

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2018119452 A2	28-06-2018	AU 2017382439 A1	20-06-2019
		BR 112019012958 A2	26-11-2019
		CA 3046007 A1	28-06-2018
		CN 110325650 A	11-10-2019
		EP 3559270 A2	30-10-2019
		JP 2020504606 A	13-02-2020
		KR 20190095410 A	14-08-2019
		US 2019390253 A1	26-12-2019
		WO 2018119452 A2	28-06-2018

WO 2020160414 A1	06-08-2020	AU 2020216438 A1	29-07-2021
		CA 3126428 A1	06-08-2020
		CN 113661249 A	16-11-2021
		EP 3918089 A1	08-12-2021
		US 2020248272 A1	06-08-2020
		WO 2020160414 A1	06-08-2020

US 2011237444 A1	29-09-2011	NONE	

US 2018305738 A1	25-10-2018	NONE	

WO 2017190215 A1	09-11-2017	BR 112018072536 A2	26-03-2019
		CA 3022606 A1	09-11-2017
		CN 109415763 A	01-03-2019
		EP 3452615 A1	13-03-2019
		JP 6979554 B2	15-12-2021
		JP 2019521315 A	25-07-2019
		JP 2021168657 A	28-10-2021
		KR 20190003987 A	10-01-2019
		KR 20210095962 A	03-08-2021
		US 2019144848 A1	16-05-2019
		WO 2017190215 A1	09-11-2017

フロントページの続き

(51)国際特許分類

C 1 2 N 15/11 (2006.01)

F I

C 1 2 N 15/11

Z

テーマコード (参考)

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

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 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

グリーンリーフ, ウィリアム ジェイ.

アメリカ合衆国 カリフォルニア 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

F ターム (参考)

4B063 QA01 QQ03 QQ42 QR08 QR14 QR42 QR62 QS25 QS34 QX02