

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7436493号
(P7436493)

(45)発行日 令和6年2月21日(2024.2.21)

(24)登録日 令和6年2月13日(2024.2.13)

(51)国際特許分類			F I		
C 1 2 Q	1/6806(2018.01)		C 1 2 Q	1/6806	Z Z N A
G 1 6 B	30/00 (2019.01)		G 1 6 B	30/00	
C 1 2 N	15/09 (2006.01)		C 1 2 N	15/09	Z
請求項の数 24 (全165頁)					
(21)出願番号	特願2021-544100(P2021-544100)	(73)特許権者	521151027	最終頁に続く	
(86)(22)出願日	令和1年10月4日(2019.10.4)		エイアールシー バイオ リミテッド ラ		
(65)公表番号	特表2022-513343(P2022-513343 A)		イアビリティ カンパニー		
(43)公表日	令和4年2月7日(2022.2.7)		アメリカ合衆国 マサチューセッツ州 0		
(86)国際出願番号	PCT/US2019/054843		2 1 4 1 ケンブリッジ カナル パーク		
(87)国際公開番号	WO2020/072990	(74)代理人	100094569	最終頁に続く	
(87)国際公開日	令和2年4月9日(2020.4.9)		弁理士 田中 伸一郎		
審査請求日	令和4年10月3日(2022.10.3)	(74)代理人	100103610		
(31)優先権主張番号	62/741,466		弁理士 ▲吉▼田 和彦		
(32)優先日	平成30年10月4日(2018.10.4)	(74)代理人	100109070		
(33)優先権主張国・地域又は機関	米国(US)		弁理士 須田 洋之	最終頁に続く	
(31)優先権主張番号	62/832,560	(74)代理人	100119013		
(32)優先日	平成31年4月11日(2019.4.11)		弁理士 山崎 一夫		
最終頁に続く		(74)代理人	100123777		

(54)【発明の名称】 次世代シーケンスにおいて低サンプルインプットを扱うための正規化対照

(57)【特許請求の範囲】
【請求項1】

サンプル中の標的生物の力価を定量する方法であって、

a. 人間を手術又は治療する方法を除く、前記標的生物を含むサンプルを提供することであって、前記標的生物が、少なくとも1つの標的配列を含む、提供することと、

b. 少なくとも3つの生物の種の既知の力価を含む、多検体対照を提供することであって、前記生物が不活化されている、提供することと、

c. 正規化対照の既知の量を前記サンプルおよび前記多検体対照と混合することと、

d. ハイスループットシーケンスライブラリを前記サンプルおよび前記多検体対照から調製することと、

e. 前記ライブラリをシーケンスしてサンプルリードの集合および多検体対照リードの集合を産生することと、

f. (e)からの前記サンプルリードの集合および前記多検体対照リードの集合を、前記正規化対照を使用して正規化することと、

g. 正規化されたリードと前記多検体対照中の前記少なくとも3つの生物の種の既知の力価との間の関係性を判定することであって、前記関係性が線形関係又は対数関係である、判定することと、

h. (g)で判定された前記関係性を使用して前記サンプル中の前記標的生物の力価を計算することと、を含む、方法。

【請求項2】

前記正規化対照が少なくとも3つのポリヌクレオチドのグループを含み、各グループ内の前記ポリヌクレオチドは同じ長さのものである、請求項1に記載の方法。

【請求項3】

前記正規化対照中のすべての前記ポリヌクレオチドの前記長さは同じである、請求項2に記載の方法。

【請求項4】

前記正規化対照の各グループ内の前記ポリヌクレオチドは、前記正規化対照の他のグループ内の前記ポリヌクレオチドと比較した場合、異なる長さのものである、請求項2に記載の方法。

【請求項5】

前記正規化対照中の前記ポリヌクレオチドのグループの前記長さは、線形配列または幾何学配列で分布している、請求項4に記載の方法。

【請求項6】

前記正規化対照中の前記少なくとも3つのグループ中の前記ポリヌクレオチドの長さは、約15bpおよび約50,000bpの間である、または

前記正規化対照中の前記少なくとも3つのグループ中の前記ポリヌクレオチドの長さは、約500bpおよび約1500bpの間、約100および1200bpの間、または約150および約600bpの間である、請求項2～5のいずれか一項に記載の方法。

【請求項7】

前記正規化対照中の各グループ内の前記ポリヌクレオチドが、同じ配列を含む、請求項2～6のいずれか一項に記載の方法。

【請求項8】

前記正規化対照中の各グループ内の前記ポリヌクレオチドが、同じ配列を含まない、請求項2～6のいずれか一項に記載の方法。

【請求項9】

前記正規化対照中の各ポリヌクレオチドのグループが、少なくとも3つのポリヌクレオチドのサブグループを含み、

各サブグループ内の前記ポリヌクレオチドは、同じ配列を含み、かつ、各サブグループ内の前記ポリヌクレオチドは、他のサブグループとは同じ配列を含まない、請求項8に記載の方法。

【請求項10】

前記正規化対照中の少なくとも1つのポリヌクレオチドのグループ中の少なくとも1つの配列が、コンポーネント配列を含む、請求項7～9のいずれか一項に記載の方法。

【請求項11】

前記正規化対照中のすべてのポリヌクレオチドのグループ中のすべての配列が、コンポーネント配列を含む、請求項7～9のいずれか一項に記載の方法。

【請求項12】

前記正規化対照中の少なくとも1つのポリヌクレオチドのグループの前記配列が、単離配列を含む、請求項7～10のいずれか一項に記載の方法。

【請求項13】

前記正規化対照中のすべてのポリヌクレオチドのグループの前記配列が、単離配列を含む、請求項7～9のいずれか一項に記載の方法。

【請求項14】

前記正規化対照中の少なくとも1つのポリヌクレオチドのグループの少なくとも1つのサブグループの前記配列が、単離配列を含む、請求項9または10に記載の方法。

【請求項15】

前記正規化対照中の少なくとも1つのグループのすべてのサブグループの前記配列が、単離配列を含む、請求項9または10に記載の方法。

【請求項16】

前記正規化対照中のすべてのグループのすべてのサブグループの前記配列が、単離配列

10

20

30

40

50

を含む、請求項 9 に記載の方法。

【請求項 17】

前記正規化対照中の各サブグループの前記単離配列が、他のサブグループの前記単離配列と同じではない、請求項 15 または 16 に記載の方法。

【請求項 18】

前記正規化対照中の前記単離配列が、約 6 b p および約 200,000 b p の間、約 15 b p および約 50,000 b p の間、約 500 b p および約 1500 b p の間、約 100 b p および約 1200 b p の間、または約 150 b p および約 600 b p の間の配列を含む、請求項 12 ～ 17 のいずれか一項に記載の方法。

【請求項 19】

前記正規化対照中の前記単離配列が、ウイルス、細菌、真菌、または真核性寄生虫から単離される、または由来する、請求項 12 ～ 18 のいずれか一項に記載の方法。

【請求項 20】

前記正規化対照中の前記単離配列が、前記サンプル中の少なくとも 1 つの標的配列と同じではない、請求項 12 ～ 19 のいずれか一項に記載の方法。

【請求項 21】

前記正規化対照中の少なくとも 1 つのポリヌクレオチドのサブグループまたはグループの少なくとも 1 つの配列が、前記サンプル中の前記少なくとも 1 つの標的配列と少なくとも 1 つの配列特性を共有し、

前記少なくとも 1 つの配列特性が、GC 含量パーセント、エントロピー、複雑度、長さ、電子イオン間相互作用ポテンシャル (EIP)、転位性要素配列、ウイルス配列、細菌配列、真菌配列、真核性寄生虫配列、もしくは 1 つまたは複数のヒト遺伝子配列の配列特性、またはこれらの組み合わせを含む、請求項 9 ～ 20 のいずれか一項に記載の方法。

【請求項 22】

前記サンプルが、宿主および非宿主核酸の混合物を含む、請求項 21 に記載の方法。

【請求項 23】

前記少なくとも 1 つの標的配列が、前記サンプル中の非宿主配列である、請求項 20 ～ 22 のいずれか一項に記載の方法。

【請求項 24】

前記サンプルが、癌と診断された対象からのサンプルを含む、請求項 20 ～ 23 のいずれか一項に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本出願は、2018年10月4日に提出された米国仮特許出願第62/741,466号、および2019年4月11日に提出された米国仮特許出願第62/832,560号の優先権および利益を主張するものであり、各出願の内容はその全体で参照により本明細書に援用される。

【0002】

本発明は、ハイスループットシーケンスを使用したサンプル中の核酸の定量的な方法および組成物に関する。

【0003】

援用による配列表の取り込み

本出願は、電子フォーマットの配列表とともに提出される。配列表は、2019年10月3日に作成された、ARCB-006-02WO_SeqList.txtという表題のファイルとして提供され、そのサイズは5943バイトである。電子フォーマットでの配列表の情報は、その全体で参照により援用される。

【背景技術】

【0004】

次世代シーケンス (NGS) は、ハイスループットシーケンスとも呼ばれており、研究および診断の強力なツールである。NGSの技術により、核酸の大規模なハイスループットシ

10

20

30

40

50

ーケンスが可能となった。NGS技術を使用することで、一つのサンプルに由来する数百万のDNA分子またはRNA分子（ポリヌクレオチド）を安価で効率的にシーケンスすることができる。NGSシーケンス中にポリヌクレオチドがサンプリングされる度合いは、サンプル中のポリヌクレオチドの初期濃度に比例する。つまりNGSを使用して、シーケンスサンプル中の個々のポリヌクレオチドの量、またはポリヌクレオチドの集団の量を判定することができる。しかし、NGSにより生成されたリードの数と、サンプル中のポリヌクレオチドの開始濃度は、必ずしも良好に相関するわけではない。NGSのリード数と開始濃度の相関は、低サンプルインプットで開始する場合に特に問題となる。その理由は、開始サンプルインプットが低い場合にはサンプル調製とシーケンスを行う間の確率過程によって、NGS法でシーケンスされる核酸分子の数に対して不釣り合いな効果がもたらされる可能性があるためである。様々な臨床応用において、例えば、低力価でサンプル中に存在する病原体の検出を試みる際に、開始サンプルインプットが低くなる可能性がある。

10

【0005】

当分野において、NGSを使用するときにサンプル中のポリヌクレオチドの存在量を正確に定量するための組成物および方法に対するニーズが存在する。本明細書において、NGSを使用した、かかるサンプル中のポリヌクレオチドの定量のための方法および組成物が提供される。

【発明の概要】

【0006】

本明細書において、次世代シーケンスのための正規化対照、当該対照を作製する方法、および当該対照を使用する方法が提供される。

20

【0007】

1つの態様において、本明細書において、少なくとも3つのポリヌクレオチドのグループを含む正規化対照が提供され、当該各グループ内のポリヌクレオチドは、同一の長さのポリヌクレオチドである。一部の実施形態では、正規化対照中のすべてのポリヌクレオチドの長さは、同一である。一部の実施形態では、各グループ内のポリヌクレオチドは、任意の他のグループ内のポリヌクレオチドと比較したとき、異なる長さのポリヌクレオチドである。

【0008】

本開示の正規化対照の一部の実施形態では、各グループ内のポリヌクレオチドは、同一の配列を含む。一部の実施形態では、各グループ内のポリヌクレオチドは、同一の配列を含まない。一部の実施形態では、ポリヌクレオチドグループの各々は、少なくとも3つのポリヌクレオチドのサブグループを含み、当該各サブグループ内のポリヌクレオチドは同一の配列を含み、および当該各サブグループ内のポリヌクレオチドは任意の他のサブグループと同一の配列は含まない。

30

【0009】

本開示の正規化対照の一部の実施形態では、少なくとも1つのポリヌクレオチドのグループの配列は、コンポーネント配列（component sequence）を含む。一部の実施形態では、すべてのポリヌクレオチドのグループの配列が、コンポーネント配列を含む。一部の実施形態では、コンポーネント配列は、ランダム配列を含み、またはランダム配列からなる。

40

【0010】

本開示の正規化対照の一部の実施形態では、少なくとも1つのポリヌクレオチドのグループの配列は、単離配列（isolated sequence）を含む。一部の実施形態では、すべてのポリヌクレオチドのグループの配列が、単離配列を含む。一部の実施形態では、単離配列は、ウイルス、細菌、真菌または真核生物から単離され、または誘導される。一部の実施形態では、単離配列は、シーケンスサンプル中の標的配列と同一ではない。

【0011】

本開示の正規化対照の一部の実施形態では、正規化対照中の少なくとも1つのポリヌクレオチドのサブグループ、または少なくとも1つのポリヌクレオチドのグループの配列が

50

、シーケンスサンプル中の標的配列と、少なくとも1つの配列プロパティを共有する。当該少なくとも1つの配列プロパティは、例えば、転位因子配列、ウイルス配列、細菌配列、真菌配列、真核寄生体配列、または一つ以上のヒト遺伝子配列の配列プロパティである。

【0012】

本開示の正規化対照の一部の実施形態では、シーケンスサンプルは、宿主核酸と非宿主核酸の混合物を含む。一部の実施形態では、宿主は、例えば昆虫、植物または動物などの真核生物である。一部の実施形態では、宿主は、ヒトである。一部の実施形態では、非宿主は、共生生物 (symbiote)、共生生物 (commensal organism)、寄生体、または病原体を含む。非宿主は、複数の種を含み得る。一部の実施形態では、標的配列は、シーケンスサンプル中の非宿主配列である。例示的な実施形態では、非宿主は、ウイルス、細菌または真菌であり、標的配列は、当該非宿主の毒性因子の配列を含む。

10

【0013】

別の態様では、本明細書において、正規化対照を作製する方法が提供され、当該方法は、(a)少なくとも3つのDNA分子のグループを合成することであって、当該グループ内のDNA分子は同一の配列を有し、当該各DNA分子の配列は、5'から3'に向かって、第一のコンポーネント配列、第一の制限酵素部位、第二のコンポーネント配列、第二の制限部位、および第三のコンポーネント配列を含むこと、(b)溶液中の当該DNA分子のグループの各々の量を定量すること、(c)当該DNA分子のグループの各々を、第一の制限部位と第二の制限部位を切断する制限酵素を用いて消化し、少なくとも3つの消化されたDNA分子のグループを生成すること、および(d)当該消化されたDNA分子のグループの各々の所定量を混合し、それにより正規化対照を生成すること、を含む。

20

【0014】

別の態様では、本明細書において、RNA分子を含む正規化対照を作製する方法が提供され、当該方法は、(a)少なくとも3つのDNA分子のグループを合成することであって、当該少なくとも3つのDNA分子の各々は、5'から3'に向かって、第一のコンポーネント配列、プロモーター配列、第一の制限酵素部位、第二のコンポーネント配列、第二のプロモーター配列、第二の制限部位、第三のコンポーネント配列、および第三のプロモーター配列を含むこと、(b)当該DNA分子のグループの各々を、第一の制限酵素部位と第二の制限酵素部位を切断する制限酵素を用いて消化し、少なくとも3つの消化されたDNA分子のグループを生成すること、(c)当該少なくとも3つの消化されたDNA分子のグループの各々をインビトロ転写し、少なくとも3つのRNA分子の集団を生成すること、(d)当該RNA分子の集団の各々で生成されたRNAの量を定量すること、および(e)当該RNA分子の集団の各々の所定量を混合し、それによりRNA分子を含む正規化対照を生成すること、を含む。

30

【0015】

別の態様では、本明細書において、少なくとも3つのポリヌクレオチドのグループを含む正規化対照を作製する方法が提供され、この場合において各グループ内のポリヌクレオチドは、同一の長さのポリヌクレオチドであり、当該方法は、(a)サンプルからDNAを抽出すること、(b)当該DNAを、制限酵素を用いて消化し、DNA断片のコレクションを生成すること、(c)当該DNA断片のコレクションを分離させること、(d)当該DNA断片を精製して、少なくとも3つのポリヌクレオチドのグループを生成することであって、当該各グループ内のポリヌクレオチドは、同一の長さのポリヌクレオチドであること、および(e)ポリヌクレオチドのグループの各々の所定量を混合し、正規化対照を生成すること、を含む。

40

【0016】

別の態様では、本明細書において、サンプル中の少なくとも1つの標的核酸分子の発現レベルを定量する方法が提供され、当該方法は、(a)既知の量の本開示の正規化対照と、当該サンプルを混合すること、(b)ハイスループットシーケンスライブラリを調製すること、(c)当該ライブラリをシーケンスして、リードのコレクションを生成すること、(d)当該サンプルまたは当該正規化対照に対して各リードをマッピングすること、(e)当

50

該正規化対照中の各グループまたは各サブグループのポリヌクレオチドにより生成されたリードの数を決定すること、(f) (a)においてサンプルと混合された正規化対照中の複数の核酸分子の各々の開始濃度と、(c)において生成されたリードの数との間の関連性を計算すること、(g) リードと、サンプル中の濃度の間の関連性を決定すること、および(h) (g)のモデルを使用して、標的核酸分子により生成されたリードの数から、サンプル中の少なくとも1つの標的核酸分子の初期濃度を計算すること、を含む。

【0017】

本明細書において、次世代シーケンスのための多検体対照、当該対照を作製する方法、および当該対照を使用する方法が提供される。一部の実施形態では、本明細書に記載される正規化対照を使用して、多検体対照からのリードおよびサンプルからのリードを正規化し、正規化されたリードを使用して、サンプル中の一つ以上の標的生物の力価が決定される。

10

【0018】

1つの態様では、本明細書において、少なくとも3つの異なる種の生物の混合物を含む多検体対照が提供され、この場合において当該生物は不活化されている。一部の実施形態では、多検体対照は、少なくとも10の異なる種の生物を含む。

【0019】

別の態様では、本明細書において、サンプル中の少なくとも1つの標的配列を定量する方法が提供され、この場合において当該サンプルは、少なくとも3つの異なる種の生物の混合物を含む多検体対照を含み、当該生物は不活化されている。

20

【0020】

別の態様では、本明細書において、サンプル中の少なくとも1つの標的配列を定量する方法が提供され、当該方法は、既知の量の多検体対照と、当該サンプルを混合することを含み、この場合において当該多検体対照は、少なくとも3つの異なる種の生物の混合物を含み、当該生物は不活化されている。

【0021】

別の態様では、本明細書において、サンプル中の標的生物の力価を定量する方法が提供され、当該方法は、(a) 当該標的生物を含むサンプルを提供することであって、当該標的生物は、少なくとも1つの標的配列を含むこと、(b) 力価が既知の少なくとも3種の生物を含む多検体対照を提供することであって、当該生物は不活化されていること、(c) 既知の量の本開示の正規化対照と当該サンプルおよび当該多検体対照を混合すること、(d) 当該サンプルおよび当該多検体対照からハイスループットシーケンスライブラリを調製すること、(e) 当該ライブラリをシーケンスし、サンプルリードのコレクションおよび多検体対照リードのコレクションを生成すること、(f) 当該正規化対照を使用して、(e)からのサンプルリードのコレクションおよび多検体対照リードのコレクションを正規化すること、(g) 当該多検体対照中の少なくとも3種の生物からの生物の正規化リードと、既知の力価の間の関連性を決定すること、および(h) (g)で決定された関連性を使用して、当該サンプル中の標的生物の力価を計算すること、を含む。

30

【0022】

一部の実施形態では、(f)の工程での正規化は、(i) サンプルリードのコレクションからのリードを、当該サンプルまたは正規化対照に対してマッピングすること、(ii) 多検体対照リードのコレクションからのリードを、当該多検体対照または正規化対照に対してマッピングすること、(iii) 当該サンプルリードのコレクションまたは当該多検体対照リードのコレクションに関し、正規化対照中のポリヌクレオチドのグループまたはサブグループの各々により生成されたリードの数を決定すること、(iv) 正規化対照中のポリヌクレオチドのグループまたはサブグループの各々の開始濃度と、当該サンプルおよび多検体対照の両方に関し、(e)の工程で生成されたリードの数との間の関連性を計算すること、および(v) 当該サンプルおよび当該多検体対照におけるリードと濃度の間の関連性を決定すること、を含む。

40

【0023】

50

本開示は、複数の正規化対照オリゴヌクレオチドを作製する方法を提供し、当該方法は、(a) 少なくとも1つの参照配列から、複数の参照配列断片を生成すること、(b) 参照配列断片の数の関数として、少なくとも1つのパラメータの分布を生成すること、(c) 当該分布を少なくとも5つのビン (bins) に分割すること、(d) 当該少なくとも5つのビンのうちの少なくとも3つから、少なくとも1つの参照配列断片を選択すること、(e) 当該少なくとも3つの参照配列をシャッフルして、シャッフル化配列を生成すること、および (f) 当該シャッフル化配列を含むオリゴヌクレオチドを合成し、それにより複数の正規化対照オリゴヌクレオチドを生成すること、を含む。

【0024】

本開示は、複数の正規化対照オリゴヌクレオチドを作製する方法を提供し、当該方法は、(a) スライディングウィンドウを使用して、少なくとも1つの参照配列から、複数の参照配列断片を生成すること、(b) 参照配列断片の数の関数として、少なくとも1つのパラメータの分布を生成すること、(c) 当該分布を少なくとも5つのビン (bins) に分割すること、(d) 当該少なくとも5つのビンのうちの少なくとも3つから、少なくとも2つの参照配列断片を選択することであって、当該少なくとも2つの参照配列断片は、参照配列中で非連続であるか、または異なる参照配列であること、(e) 当該少なくとも3つのビンの各々に由来する当該少なくとも2つの参照配列断片を連結させること、および (f) 当該連結された参照配列断片を含むオリゴヌクレオチドを合成し、それにより複数の正規化対照オリゴヌクレオチドを生成すること、を含む。一部の実施形態では、当該スライディングウィンドウは、1bp、2bp、3bp、4bpまたは5bpのスライディングウィンドウを含む、請求項203~206のいずれか1項に記載の方法。一部の実施形態では、参照配列断片は、対応する参照配列の約15~60、約20~40、約20~30、約15~32、20~32、または約25~35の連続ヌクレオチドである。一部の実施形態では、参照配列断片は、対応する参照配列の29、30、31、32、33または34の連続ヌクレオチドである。

【0025】

本開示の方法の一部の実施形態では、パラメータは、(1) GC含量パーセント、(2) エントロピー、(3) 複雑度、(4) EIIP、またはそれらの組み合わせのうちの少なくとも1つを含む。

【0026】

本開示の方法の一部の実施形態では、当該少なくとも1つの参照配列は、少なくとも2個、少なくとも10個、少なくとも20個、少なくとも50個、少なくとも100個、少なくとも200個、少なくとも300個、少なくとも400個、少なくとも500個、少なくとも600個、少なくとも700個、少なくとも800個、少なくとも900個、少なくとも1,000個、少なくとも1,200個、少なくとも1,300個、少なくとも1,400個、少なくとも1,500個、少なくとも1,600個、少なくとも1,700個、少なくとも1,800個、少なくとも1,900個、少なくとも2,000個、少なくとも2,200個、少なくとも2,400個、少なくとも2,600個、少なくとも2,800個、少なくとも3,000個、少なくとも4,000個、少なくとも5,000個、少なくとも6,000個、少なくとも7,000個、少なくとも8,000個、少なくとも9,000個、または少なくとも10,000個の参照配列を含む。一部の実施形態では、当該少なくとも1つの参照配列は、約2,000個の参照配列を含む。

【0027】

本開示の方法の一部の実施形態では、各正規化対照オリゴは、参照配列分布中の対応するビンからの参照配列断片と、(1) 同様のGC含量パーセント、(2) 同様のエントロピー、(3) 同様のEIIP、(4) 同様の長さ、または(5) 同様の複雑度、のうちの少なくとも3つを有する。

【0028】

本開示は、本明細書に提供される正規化対照と多検体対照とを組み合わせた使用のための、対象の標的核酸に対するサンプルを富化する方法を提供する。一部の実施形態では、当該方法は、ヌクレオチド修飾を基にした富化方法を含む。一部の実施形態では、当該方法は、核酸誘導ヌクレアーゼ (nucleic acid-guide nuclease) を基にした富化方法を含

む。一部の実施形態では、当該方法は、ヌクレオチド修飾を基にした富化方法と、核酸誘導ヌクレアーゼを基にした富化方法の両方を含む。

【0029】

本開示の方法の一部の実施形態では、対象の核酸に対するサンプルを富化する方法は、(a) 対象の核酸と、枯渇の標的とされる核酸を含むサンプルを提供することであって、少なくとも当該対象の核酸のサブセット、または枯渇の標的とされる核酸のサブセットが、第一の修飾感受性制限酵素に対する複数の第一の認識部位を含むこと、(b) 当該サンプル中の複数の核酸を末端脱リン酸化すること、(c) (b) からのサンプルと、当該第一の修飾感受性制限酵素とを、当該サンプル中の核酸中の当該第一の修飾感受性制限部位の少なくとも一部の切断が可能となる条件下で接触させること、および (d) (c) からのサンプルと、アダプターとを、対象の複数の核酸の5'末端と3'末端に当該アダプターがライゲーションすることが可能となる条件下で接触させ、それにより自身の5'末端と3'末端でアダプターがライゲーションされた対象核酸が富化されたサンプルが生成されること、を含む。

10

【0030】

本開示の方法の一部の実施形態では、当該方法は、サンプルと、複数の核酸誘導ヌクレアーゼ-ガイド核酸 (gNA) 複合体とを接触させることを含み、当該gNAは、枯渇の標的とされる核酸中の標的部位に対して相補的であり、それにより一方の末端でアダプターがライゲーションされた枯渇の対象とされる切断核酸と、5'末端と3'末端の両方でアダプターがライゲーションされた対象の標的核酸が生成される。一部の実施形態では、当該方法は、サンプルを、少なくとも 10^2 個の固有核酸-誘導ヌクレアーゼ-gNA複合体、少なくとも 10^3 個の固有核酸-誘導ヌクレアーゼ-gNA複合体、少なくとも 10^4 個の固有核酸-誘導ヌクレアーゼ-gNA複合体、または少なくとも 10^5 個の固有核酸-誘導ヌクレアーゼ-gNA複合体とを接触させることを含む。一部の実施形態では、核酸誘導ヌクレアーゼは、Cas9、Cpf1、またはそれらの組み合わせである。

20

【0031】

他の態様では、本明細書において、本開示の多検体対照を含むキット、本開示の正規化対照を含むキット、および本開示の正規化対照と多検体対照の両方を含むキットが提供される。一部の実施形態では、当該キットはさらに、ヌクレオチド修飾を基にした対象標的核酸の富化のための試薬、核酸誘導ヌクレアーゼを基にした対象標的核酸の富化のための試薬、またはその両方を含む。

30

【0032】

本開示は、本開示の正規化対照を含むキットを提供する。

【0033】

本開示は、本開示の多検体対照を含むキットを提供する。

【0034】

本開示は、正規化対照、多検体対照、試薬、および使用説明書を含むキットを提供する。一部の実施形態では、当該キットはさらに、枯渇の対象とされる配列を枯渇させるための試薬、および使用説明書を含む。

【0035】

本開示は、複数の正規化対照ポリヌクレオチド配列を設計するためのシステムを提供する。一部の実施形態では、当該システムは、コンピュータ実行可能命令を保存するコンピュータ可読記憶媒体を含み、当該命令は、(i) 少なくとも1つの参照配列をインポートする命令、(ii) 少なくとも1つの参照配列から複数の参照配列断片を生成する命令、(iii) 参照配列断片の数の関数として少なくとも1つのパラメータの分布を生成する命令、(iv) 当該分布をビンに分割する命令、(v) 少なくとも当該ビンのサブセットから複数の参照配列断片を選択する命令、および (vi) 当該複数の参照配列断片をシャッフルしてシャッフル化配列を生成し、それにより複数の正規化ポリヌクレオチド配列を生成する命令、を含む。一部の実施形態では、当該システムは、コンピュータ実行可能命令を保存するコンピュータ可読記憶媒体を含み、当該命令は、(i) 少なくとも1つの参照配列をインポート

40

50

する命令、(ii) 少なくとも1つの参照配列から複数の参照配列断片を生成する命令、(iii) 参照配列断片の数の関数として少なくとも1つのパラメータの分布を生成する命令、(iv) 当該分布をビンに分割する命令、(v) 少なくとも当該ビンの各サブセットから少なくとも2つの参照配列断片を選択する命令であって、当該少なくとも2つの参照配列断片は、当該参照配列中で非連続であるか、または異なる参照配列に由来し、および(vi) 各ビンからの少なくとも2つの参照配列断片を連結し、それにより複数の正規化ポリヌクレオチド配列を生成する命令、を含む。一部の実施形態では、当該システムはさらに、(a) 当該少なくとも1つの参照配列を含む入力ファイルのセットを受信する工程、および(b) 当該コンピュータ可読記憶媒体中に保存されたコンピュータ実行可能命令を実行する工程、を含む工程を実施するよう構成されたプロセッサを含む。一部の実施形態では、パラメータは、(1) GC含量パーセント、(2) エントロピー、(3) 複雑度、(4) EIIP、(5) 長さ、またはそれらの組み合わせのうちの少なくとも1つを含む。

【図面の簡単な説明】

【0036】

【図1】図1は、正規化対照(NC: normalization control)リードのクオリティコントロール分析を示す表である。

【図2】図2は、サイトメガロウイルス(CMV)の存在量(x軸)と力価(y軸)のグラフである。正規化を行わなければ、 R^2 値は、0.89である。

【図3】図3は、正規化対照(NC)で正規化されたCMV存在量(x軸)と力価(y軸)のグラフである。正規化を行うと、 R^2 値は、0.98である。

【図4】図4は、3種の仮想次世代シーケンスライブラリを用いて分析されたサンプル中のウイルスの存在量パーセント(y軸、任意の単位)を示す棒グラフであり、異なる転換を使用して計算されたウイルスの存在量パーセントを示す。各ライブラリに対し、左から右に向かって、棒グラフは、判明しているウイルス量、次世代シーケンス後の生リードエビデンスを使用して計算されたウイルス量、および $f(r)$ リードエビデンスと $f(r)+\Delta r$ リードエビデンスを使用して計算されたウイルス量を表す。

【図5】図5は、 $f(r)$ を使用してウイルス量を計算したときと、生リードを使用してウイルス量を計算したときの変動性における予測される減少を示すプロットである。

【図6】図6は、RNA抽出プロセスの間に正規化対照が添加されたサンプルにおける、標的配列レベルの定量に関するプロトコルの概要である。

【図7】図7は、Exact Diagnostics Negative Plasmaサンプルのライブラリにおける、正規化対照リードのパーセントを示すグラフであり、NCは、サンプル抽出の間に少量、中量、および多量にサンプルに添加された。

【図8】図8は、二重インデックス化プライマー(dual indexing primers)を正規化対照にアニーリングするプロトコルの表である。

【図9】図9は、正規化対照において、添加されたライブラリ中の正規化対照リードのパーセントを示すグラフであり、DNA NCは、多量(左側の2つの棒)、中量(右側の2つの棒)、および少量(真ん中の2つの棒)で添加されている。

【図10】図10は、例示的な正規化対照のGC含量、総カウントおよび長さを示す表である。

【図11】図11は、本開示の例示的な正規化対照の濃度と比率を示す表である。左から右に、列は、正規化対照組成物中のポリヌクレオチドのグループ(図11において「オリゴ」と呼称され、対応するオリゴ番号とともに示されている)、ポリヌクレオチドの各グループ、および組成物中の正規化対照ポリヌクレオチド集団の最終濃度($\text{ng}/\mu\text{L}$)、正規化対照組成物中のポリヌクレオチドの指定グループのDNAの量と、正規化対照組成物中のすべてのポリヌクレオチドの全DNAの総量の比率、を示している。

【図12】図12は、本開示の例示的なRNA正規化対照のサイズ、インプット分子の数、および濃度を示す表である。左から右に向かって、列は、正規化対照中の個々のRNA断片、正規化対照中のRNAグループ、RNA分子の長さ(単位はkbとbp)、インプット分子の数、オリジナルのgBlock中の各RNA分子のアトモル数、および結果として得られたgBlock

10

20

30

40

50

断片化後のアトム／マイクロリットル単位での各RNA分子の濃度を示す。

【図13】図13は、較正および実験サンプルに対して観察されたシグナルの関数としてのlog単位での力価を示すグラフである。暗灰色の円は、較正に使用された多検体対照サンプル中の種を示し、一方で明灰色の円は実験サンプルである。

【図14A】図14Aは、本開示の正規化対照を設計するための例示的プロセスを示すフローチャートである。

【図14B】図14Bは、本開示の正規化対照の一つの設計でのシグナル（y軸）とlog濃度（x軸）を示すグラフである。

【図14C】図14Cは、本開示の正規化対照の一つのセットの濃度を示す表である。各オリゴは、異なる配列である。

【図14D】図14Dは、本開示の正規化対照の一つの設計でのシグナル（y軸）とlog濃度（x軸）を示すグラフである。濃度における各桁は、1log当たり少なくとも1つの異なるオリゴを含む6つの濃度により表されており、7～8桁が対象となっている。

【図15】図15は、およそ1200個の参照ゲノムに対する、エントロピー、電子-イオン相互作用ポテンシャル（EIIP：electron-ion interaction potential）およびGC含量の間の相互作用を示す三元プロットである。

【図16】図16は、シャッフリング法を使用して生成された参照配列からのKmerの分布を示す一連の3つのプロットである。左から右に向かって、電子EIIP、エントロピー、およびGC含量。EIIP、エントロピー、およびGC含量の値は、それぞれ左から右にx軸上に示されている。y軸上にはKmer密度が示されている。線は分布を表し、ヒストグラム棒は、所与のピン中のKmer数の実際のカウント数を表す。

【図17A】図17Aは、シャッフリング法を使用して生成された1000万個のKmersの分布を示す一連の3つのプロットである。左から右に向かって、電子EIIP、エントロピー、およびGC含量。EIIP、エントロピー、およびGC含量の値は、それぞれ左から右にx軸上に示されている。y軸上にはKmer密度が示されている。

【図17B】図17Bは、シャッフリング法を使用して生成された130万個のKmersの分布を示す一連の3つのプロットである。左から右に向かって、電子EIIP、エントロピー、およびGC含量。EIIP、エントロピー、およびGC含量の値は、それぞれ左から右にx軸上に示されている。y軸上にはKmer密度が示されている。

【図18】図18は、シャッフリング法を行って正規化対照配列を生成する前にKmerを連結するためのプロセスを示すダイアグラムである。

【図19A】図19Aは、EIIP（左上）、GC含量（右上）および複雑度（下の行）に対し、Cantorシャッフリング法とBKポリオーマウイルス単離株CH-1を使用して生成された正規化対照の間の対応を示す一連の3つのグラフである。非シャッフル化Kmerはx軸上にプロットされ、Cantorシャッフル化されたKmerはy軸上にプロットされている。線は、1：1の対応を示している。

【図19B】図19Bは、EIIP（左上）、GC含量（右上）および複雑度（下の行）に対し、Cantorシャッフリング法とJCポリオーマウイルス株NIID 12-31を使用して生成された正規化対照の間の対応を示す一連の3つのグラフである。非シャッフル化Kmerはx軸上にプロットされ、Cantorシャッフル化されたKmerはy軸上にプロットされている。線は、1：1の対応を示している。

【図19C】図19Cは、EIIP（左上）、GC含量（右上）および複雑度（下の行）に対し、Cantorシャッフリング法とヒトポリオーマウイルス1のBK2株を使用して生成された正規化対照の間の対応を示す一連の3つのグラフである。非シャッフル化Kmerはx軸上にプロットされ、Cantorシャッフル化されたKmerはy軸上にプロットされている。線は、1：1の対応を示している。

【図19D】図19Dは、EIIP（左上）、GC含量（右上）および複雑度（下の行）に対し、Cantorシャッフリング法とヒトアデノウイルスCを使用して生成された正規化対照の間の対応を示す一連の3つのグラフである。非シャッフル化Kmerはx軸上にプロットされ、Cantorシャッフル化されたKmerはy軸上にプロットされている。線は、1：1の対応を示し

10

20

30

40

50

ている。

【図20A】図20Aは、Basic Local Alignment Search Tool (BLAST)アライメントであり、Cantorシャッフル化された正規化対照配列は、BKポリオーマウイルス単離株CH-1とはNCBI BLASTヌクレオチド配列を使用してもアラインしないことを示す。

【図20B】図20Bは、blastアライメントであり、Cantorシャッフル化された正規化対照配列は、JCポリオーマウイルスNIID 12-31株とはNCBI BLASTヌクレオチド配列を使用してもアラインしないことを示す。

【図20C】図20Cは、blastアライメントであり、Cantorシャッフル化された正規化対照配列は、ヒトポリオーマウイルス1のBK2株とはNCBI BLASTヌクレオチド配列を使用してもアラインしないことを示す。

【図20D】図20Dは、BLASTアライメントであり、Cantorシャッフル化された正規化対照配列は、ヒトアデノウイルスCとはNCBI BLASTヌクレオチド配列を使用してもアラインしないことを示す。

【図21A】図21Aは、選択されたGC含量パーセントのKmerに対するEIIP、エントロピー、およびGC含量を示す一連の3つのプロットである。

【図21B】図21Bは、選択されたGC含量パーセントのKmerに対するEIIP、エントロピー、およびGC含量を示す一連の3つのプロットである。

【図21C】図21Cは、選択されたGC含量パーセントのKmerに対するEIIP、エントロピー、およびGC含量を示す一連の3つのプロットである。

【図22】図22は、EIIP、GC含量およびエントロピーに対する第一および第二の主要コンポーネント（それぞれx軸とy軸）を示す一連のプロットである。PC1およびPC2は、参照ゲノム配列（左）およびスライディングウィンドウ／包括的Kmerアプローチを使用して生成された正規化対照配列（右）に対して計算された。

【図23】図23は、参照ゲノム配列（x軸）とスライディングウィンドウ／包括的Kmerアプローチを使用して生成された正規化対照オリゴ（y軸）のGC含量（上）、EIIP（左下）およびエントロピー（右下）を比較する一連の3つのQ-Qプロットである。

【図24】図24は、参照ゲノム配列と、スライディングウィンドウ／包括的Kmerアプローチを使用して生成された正規化対照配列のランダムサブサンプリング集団におけるエントロピーを比較するコルモゴロフ-スミルノフ（KS）検定のプロットである。ECDF＝経験的累積分布関数。矢印は正規化対照のエントロピーを示す。点はK-S統計を示す。

【図25A】図25AはBLASTアライメントであり、スライディングウィンドウ／包括的Kmerアプローチを使用して生成された785,000個の正規化対照配列の先頭部分の配列は、約750塩基対（bp）位置で約50～80bp、シュードモナスに対してアラインすることを示す。

【図25B】図25BはBLASTアライメントであり、スライディングウィンドウ／包括的Kmerアプローチを使用して生成された785,000個の正規化対照配列の後部の配列は、BLASTを使用しても任意の配列に対してアラインしないことを示す。

【図26】図26は、シャッフルリング法および包括的Kmer／スライディングウィンドウアプローチを使用して設計された正規化対照を比較する表である。

【図27】図27は、本発明の例示的な方法を示すダイアグラムである。サンプル中の核酸は脱リン酸化され、次いで制限酵素で消化される。当該制限酵素は、制限酵素認識部位に修飾が存在すると阻害される。次いで得られた消化物から露出されたリン酸塩を使用して、アダプターを対象核酸にライゲートする。

【図28】図28は、本発明の例示的な方法を示すダイアグラムである。サンプル中の核酸は脱リン酸化され、次いで制限酵素で消化される。当該制限酵素は、一つ以上の修飾ヌクレオチドを含む制限酵素部位を認識する。次いで切断された核酸を、露出末端リン酸塩を使用するエキソヌクレアーゼを用いて消化し、アダプターを残りの対象核酸にライゲートする。

【図29】図29は、本発明の例示的な方法を示すダイアグラムである。サンプル中の核酸はアダプターをライゲートされ、次いで制限酵素で消化される。当該制限酵素は、一つ以

10

20

30

40

50

上の修飾ヌクレオチドを含む制限酵素部位を認識し、両末端にアダプターがライゲートされた対象核酸を生成する。

【図30】図30は、本開示の例示的な方法を示すダイアグラムである。サンプル中の核酸はアダプターをライゲートされ、次いで核酸誘導ヌクレアーゼで切断される。当該核酸誘導ヌクレアーゼは枯渴の標的とされる核酸を切断し、両末端にアダプターがライゲートされた対象核酸を生成する。本方法は、本開示のヌクレオチド修飾を基にした方法と併せて使用してもよい。

【発明を実施するための形態】

【0037】

本明細書において、次世代シーケンス (NGS) におけるサンプル定量のための正規化対照として有用な組成物が提供される。一部の実施形態では、正規化対照は、低サンプルインプットを用いてNGSを実施するとき使用される。一部の実施形態では、例えば本明細書に記載の多検体対照を使用したサンプル中の生物の力価の計算のために、正規化対照が使用される。一部の実施形態では、正規化対照は、本明細書に記載されるように、標的配列のサンプルを富化する方法と組み合わせられる。

【0038】

本明細書において、正規化対照を作製する方法、およびNGS中に正規化対照を使用する方法が提供される。

【0039】

本明細書において別段の規定がない限り、本明細書で使用される全ての技術用語および科学用語は、本開示が属する分野の当業者によって普遍的に理解される意味と同じ意味を有する。本明細書に記載されるものと類似または均等の方法および材料を本開示の実施または検証に使用することができるが、例示的な方法および材料を記載する。

【0040】

「次世代シーケンス」 (NGS) という用語は、例えばIllumina社、Life Technologies社、およびRoche社などにより現在行われている、いわゆる並行処理化sequencing-by-synthesis法、またはsequencing-by-ligationプラットフォームを指す。次世代シーケンス法はさらに、ナノポアシーケンス法、または例えばLife Technologies社により販売されるIon Torrent技術などの電子検出を基にした方法を含む場合もある。NGSは本明細書において、「ハイスループットシーケンス」とも呼称される。

【0041】

「シーケンスサンプル」は、NGSを使用してシーケンスされる核酸を含有するサンプルである。核酸は、デオキシリボ核酸 (DNA)、リボ核酸 (RNA) またはその両方であり得る。

【0042】

「標的配列」とは、シーケンスサンプル中の個々の配列または配列セットを指し、サンプル中のその存在量は、NGS反応において、本開示のNCを使用することにより決定される。標的配列の存在量は、非ゼロ (すなわち標的配列が、いくつかの存在量でサンプル中に存在する)、またはゼロ (すなわち標的配列が存在しない、または検出値未満である) であってもよい。

【0043】

本明細書において使用される場合、「参照配列」とは、任意の公知の標的配列と同一または類似した一つ以上の配列を指す。標的配列は、一つの参照配列、または複数の参照配列を含み得、またはそれからなり得る。

【0044】

「標的生物」とは、標的配列を含む、サンプル中の生物を指す。

【0045】

本明細書において使用される場合、「コンポーネント配列」という用語は、正規化対照中に存在するポリヌクレオチドの一部または全体を指す。正規化対照ポリヌクレオチドは、コンポーネント配列を含み、またはコンポーネント配列からなる。コンポーネント配列

10

20

30

40

50

は、天然配列から単離されてもよく、遺伝子操作プラスミドから単離されてもよく、または合成であってもよい。「単離配列」という用語は、生物から、もしくは別手段により既存配列から単離または誘導されたある種のコンポーネント配列を指す。単離配列は、当該単離配列が単離または誘導された生物の配列に類似していてもよく、または例えばシャッフルリングもしくは連結などの一つ以上の転換が行われてもよい。

【0046】

「低サンプル」または「低インプットサンプル」という用語は、従来のライブラリ調製プロトコルで使用するよりも少量の核酸を含むサンプルを指す。低サンプルのインプットは、サンプル中の総核酸のうち、100ngもの核酸を含んでもよく、わずか1ピコグラムの核酸を含んでもよい。ある実施形態では、「低サンプル」または「低サンプルのインプット」という用語は、サンプル中の一つ以上の標的配列をコードする核酸の量を指す。低サンプルのインプットは、100ngもの標的配列を含む核酸を含んでもよく、または1フェムトグラム未満の標的配列を含む核酸を含んでもよい。特に稀な標的配列に関しては、フェムトモル単位およびアトモル範囲の標的配列濃度が、本開示の範囲内にあるとみなされる。

【0047】

本開示全体を通じて使用される場合、「配列同一性」または「配列類似性」は、独立型で実行可能なBLASTエンジンプログラムを使用して2配列をブラストすることにより決定されてもよい (bl2seq)。このプログラムは、National Center for Biotechnology Information (NCBI)のftpサイトからデフォルトパラメータを使用して検索可能である (Tatusova and Madden, FEMS Microbiol Lett., 1999, 174, 247-250。当該論文はその全体で参照により本明細書に組み込まれる)。2つ以上の核酸配列またはポリペプチド配列に関する文脈において使用される場合、「同一」または「同一性」という用語は、各配列の指定される領域にわたり同じである残基の指定される百分率を指す。当該百分率は、2つの配列を最適にアラインし、指定される領域にわたり当該2つの配列を比較して、同一の残基が両方の配列中に存在する位置の数を決定して、合致位置数を出し、その合致位置数を指定領域中の総位置数で割り、その結果に100を掛け、配列同一性の百分率を出すことにより計算され得る。2つの配列が異なる長さの配列である場合、またはアライメントによって一つ以上ずれた末端が生成され、比較される指定領域が一つの配列しか含まない場合、当該一つの配列の残基は計算の分母には含まれるが、分子には含まれない。DNAとRNAを比較する場合、チミン (T)とウラシル (U)は同等とみなされ得る。同一性は手計算で算出することができ、または例えばBLASTやBLAST2.0などのコンピュータ配列アルゴリズムを使用して算出することができる。

【0048】

配列は、アライメントにより少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、少なくとも99.5%、または少なくとも99.9%同一である場合に、「類似」であり得る。

【0049】

本開示は、NGSを目的としたサンプルに添加される正規化対照を提供する。任意選択的に、当該サンプルは低インプットサンプルである。

【0050】

本開示は、正規化対照を作製する方法、および当該対照を使用する方法を提供する。

【0051】

一部の実施形態では、正規化対照は、同じ量で、各サンプルに添加される。

【0052】

一部の実施形態では、正規化対照は、同じ量で、各サンプルに添加されない。

【0053】

一部の実施形態では、正規化対照は、核酸抽出の前にサンプルに添加される。他の実施

10

20

30

40

50

形態では、正規化対照は、核酸抽出の後、およびライブラリ調製の間に添加される。他の代替的实施形態では、正規化対照は、ライブラリ調製の後、およびシーケンスの前にサンプルに添加される。

【0054】

本開示は、少なくとも3つのポリヌクレオチドのグループを含む正規化対照を提供し、当該少なくとも3グループの各々の内のポリヌクレオチドは、同一の長さのポリヌクレオチドである。一部の实施形態では、正規化対照ポリヌクレオチドは、コンポーネント配列を含み、一部の实施形態では、シーケンスされる標的サンプルのなんらかの性質または特徴に似せて、または再現するように設計される。一部の实施形態では、コンポーネント配列は、ランダム配列である。一部の实施形態では、正規化対照ポリヌクレオチドは、単離配列を含む。一部の实施形態では、正規化対照ポリヌクレオチドは、単離配列とコンポーネント配列の両方を含む。一部の实施形態では、正規化対照ポリヌクレオチドの配列は、シーケンスサンプル中の標的配列と、少なくとも1つの配列プロパティを共有する。

【0055】

本開示は、正規化対照を作製する方法を提供するものであり、当該方法は、(a)少なくとも3つのDNA分子の集団を合成することであって、当該集団内の各DNA分子は同一の配列を有し、当該各DNA分子の配列は、5'から3'に向かって、第一のコンポーネント配列、第一の制限酵素部位、第二のコンポーネント配列、第二の制限部位、および第三のコンポーネント配列を含むこと、(b)溶液中の当該DNA分子の集団の各々の量を定量すること、(c)当該DNA分子の集団の各々を、第一の制限部位と第二の制限部位を切断する制限酵素を用いて消化し、少なくとも3つの消化されたDNA分子の集団を生成すること、および(d)当該消化されたDNA分子の集団の各々の所定量を一つの正規化対照組成物中で混合すること、を含む。

【0056】

例えばサンプルがRNAサンプルであり、正規化対照がRNA正規化対照である実施形態など、一部の实施形態では、本開示は、正規化対照を作製する方法を提供し、当該方法は、(a)少なくとも3つのDNA分子の集団を合成することであって、当該少なくとも3つの合成DNA分子の各々は、5'から3'に向かって、第一のコンポーネント配列、プロモーター配列に相補的な配列、第一の制限酵素部位、第二のコンポーネント配列、プロモーター配列に相補的な第二の配列、第二の制限部位、第三のコンポーネント配列、およびプロモーター配列に相補的な第三の配列を含むこと、(b)当該DNA分子の集団の各々を、第一の制限酵素部位と第二の制限酵素部位を切断する制限酵素を用いて消化し、少なくとも3つの消化されたDNA分子の集団を生成すること、(c)当該少なくとも3つの消化DNA分子の集団の各々をインビトロ転写し、少なくとも3つのRNA分子の集団を生成すること、(d)当該RNA分子の集団の各々で生成されたRNAの量を定量すること、および(e)当該RNA分子の集団の各々の所定量を混合し、一つの正規化対照組成物を生成すること、を含む。

【0057】

本開示は、正規化対照を作製する方法を提供するものであり、当該方法は、(a)サンプルからDNAを抽出すること、(b)当該DNAを、制限酵素を用いて消化し、DNA断片のコレクションを生成すること、および(c)当該DNA断片のコレクションを分離させること、(d)少なくとも3つのDNA断片を精製して、少なくとも3つのポリヌクレオチドのグループを生成することであって、当該少なくとも3つのポリヌクレオチドのグループの各々は、複数のポリヌクレオチドを含むこと、および(e)当該少なくとも3つのポリヌクレオチドのグループの各々の所定量を混合し、正規化対照を生成すること、を含む。

【0058】

本開示はさらに、正規化対照組成物を作製する方法を提供するものであり、この場合において当該正規化対照は、本開示方法を使用して合成されるポリヌクレオチドと、本開示方法を使用して単離および／または誘導される単離配列を含むポリヌクレオチドの両方を含む。

【0059】

10

20

30

40

50

本開示は、正規化対照を使用してサンプル中の少なくとも1つの標的核酸分子の発現レベルを定量する方法を提供するものであり、当該方法は、(a) 既知の量の本開示の正規化対照と、当該サンプルを混合すること、(b) ハイスループットシーケンスライブラリを調製すること、(c) 当該ライブラリをシーケンスして、リードのコレクションを生成すること、(d) 当該サンプルまたは当該正規化対照に対して各リードをマッピングすること、(e) 当該正規化対照中の複数の核酸分子の各々により生成されたリードの数を決定すること、(f) (a) においてサンプルと混合された正規化対照中の複数の核酸分子の各々の開始濃度と、(c) において生成されたリードの数との間の関連性を計算すること、(g) リードと、サンプル中の濃度の間の関連性をモデル化すること、および(h) (g) のモデルを使用して、標的核酸分子により生成されたリードの数から、サンプル中の少なくとも1つの標的核酸分子の初期濃度を計算すること、を含む。

10

【0060】

正規化対照組成物

本開示は、NGSアプリケーションにおけるサンプルインプットの定量に有用な正規化対照組成物（本明細書において正規化対照または（NC）と相互交換可能に呼称される）を提供する。本明細書に提供されるように、本開示のNCは、少なくとも2つのポリヌクレオチドのグループを含んでもよく、この場合において当該少なくとも2つのグループの各々の内のポリヌクレオチドは、同一の長さのポリヌクレオチドである。一部の実施形態では、NCは、ポリヌクレオチドとして新たに（de novo）合成される。一部の実施形態では、NCは、例えばDNAサンプルもしくはRNAサンプルなどのサンプル、生物、またはベクターから単離され、または誘導される。一部の実施形態では、NCは、新たに合成されたポリヌクレオチド、およびサンプルから単離または誘導されるポリヌクレオチドの両方を含む。一部の実施形態では、NCは、少なくとも3つのポリヌクレオチドのグループを含み、この場合において当該少なくとも3つのグループの各グループの内のポリヌクレオチドは、同一の長さのポリヌクレオチドである。

20

【0061】

一部の実施形態では、NCは、少なくとも3つの異なる長さのポリヌクレオチド配列のグループを含み、およびポリヌクレオチドの各グループは、異なる濃度で正規化対照中に存在する。

【0062】

一部の実施形態では、NCは、少なくとも15の異なる長さのポリヌクレオチド配列のグループを含み、およびポリヌクレオチドの各グループは、異なる濃度で正規化対照中に存在する。

30

【0063】

一部の実施形態では、NCは、15の異なる長さのポリヌクレオチド配列のグループを含み、およびポリヌクレオチドの各グループは、異なる濃度で正規化対照中に存在する。

【0064】

NC中のポリヌクレオチドの長さ

正規化対照内のポリヌクレオチドの長さは、濃度範囲全体で変化する。その結果、断片化比率におけるバイアスまたはクラスター化が発生し得る（Illumina社系のNGS法の場合に時折認められる）。例えば、正規化対照は、複数のポリヌクレオチドのグループを含み、各ポリヌクレオチドのグループは異なる長さであり、正規化対照内で異なる濃度で存在する。

40

【0065】

異なる長さ、および長さの組み合わせのポリヌクレオチドのグループは、本開示の正規化対照の範囲内にあるとみなされる。所与のグループ内の個々のポリヌクレオチドのすべてが、同じ長さのポリヌクレオチドである。しかし少なくとも3つのグループを含む正規化対照組成物内では、グループ内のポリヌクレオチドの長さは、当該正規化対照組成物中の任意の他のグループ内のポリヌクレオチドの長さと同じであっても、異なってもよい。

50

【0066】

したがって、一部の実施形態では、正規化対照中のすべてのポリヌクレオチドの長さは、同一である。一部の実施形態では、当該少なくとも3つのグループの各々の内のポリヌクレオチドは、任意の他のグループ内のポリヌクレオチドとは異なる長さのポリヌクレオチドである。

【0067】

一部の実施形態では、当該少なくとも3つのポリヌクレオチドのグループの長さは、直線状数列または等比数列で分布される。

【0068】

直線状数列は、数列中の次の数が、前の数と比較して都度同じ量で増加または減少する数列である。例示的な直線状数列は、「1、2、3、4、5、6」の数列を含む。概して、直線状数列は、以下の式により表される：

$$u_n = d \times n + c$$

式中、 d は、数列中の連続する項の間の第一の差異であり、 n は、数列中の項であり、 c は定数である。

【0069】

等比数列は、最初の項の後の各項が、公比と呼ばれる固定されたゼロではない数値で前の項を乗じることにより見出される数列である。例示的な等比数列は、「2、4、8、16」（各数値は、前の数値に2を乗じている）を含む。概して、等比数列は、以下の式により表される：

$$u_n = u_1 r^{n-1}$$

式中、 r は、公比である。

【0070】

例えば、ポリヌクレオチドがコンポーネント配列（本明細書において詳述される）を含む実施形態などの一部の実施形態では、各グループのポリヌクレオチドの長さは、約15bp～約3000bp、約50bp～約3000bp、約100bp～約3000bp、約1000bp～約3000bp、約1200～約3000bp、約1500bp～約3000bp、約15bp～約50bp、約15bp～約100bp、約15bp～約150bp、約15bp～約200bp、約15bp～約300bp、約15bp～約400bp、約15～約500bp、約50bp～約1200bp、約100bp～約1200bp、約150bp～約1200bp、または約150～約1100bpである。例示的な実施形態では、ポリヌクレオチドのグループのポリヌクレオチドの長さは、約15bp～約3000bpである。例示的な実施形態では、ポリヌクレオチドのグループのポリヌクレオチドの長さは、約500bp～約1500bpである。例示的な実施形態では、ポリヌクレオチドのグループのポリヌクレオチドの長さは、約100bp～約1200bpである。例示的な実施形態では、ポリヌクレオチドのグループのポリヌクレオチドの長さは、約150bp～約600bpである。

【0071】

一部の実施形態では、当該少なくとも3つのポリヌクレオチドのグループは、175bp、250bpおよび450bp；192bp、250bpおよび450bp；200bp、300bpおよび500bp；217bp、300bpおよび517bp；436bp、552bpおよび974bp；450bp、612bpおよび1034bp；510bp、626bpおよび1048bp；ならびに450bp、612bpおよび1034bpからなる長さから選択される長さを有するポリヌクレオチドを含む。

【0072】

本明細書において提供されるように、ポリヌクレオチドのグループは、インビトロで合成されることができ、または既に合成された核酸から単離もしくは誘導されることができ、または天然核酸から単離もしくは誘導されることもできる。インビトロで合成されるポリヌクレオチドは、およそ3kbの最大長を有し得ることが予期されるが、当業者であれば、かかる長さは、現在のオリゴヌクレオチド合成技術により影響を受け得ることを理解するであろう。既に合成された核酸もしくは天然核酸から単離または誘導された正規化対照ポリヌクレオチドは、合成方法を使用して作製される正規化対照よりも広範な長さを包含し得るが、これも現在のオリゴヌクレオチド合成技術により影響を受けることが予期され

10

20

30

40

50

る。

【0073】

例えば正規化対照ポリヌクレオチドが単離配列（本明細書において詳述される）を含む一部の実施形態では、当該少なくとも3つの異なるポリヌクレオチドのグループの長さは、約15bp～約500kbである。一部の実施形態では、当該少なくとも3つの異なるポリヌクレオチドのグループの長さは、約15bp～約50kbである。一部の実施形態では、当該少なくとも3つの異なるポリヌクレオチドのグループの長さは、約50bp～約50kbである。一部の実施形態では、当該少なくとも3つの異なるポリヌクレオチドのグループの長さは、約500bp～約50kbである。一部の実施形態では、当該少なくとも3つの異なるポリヌクレオチドのグループの長さは、約1000bp～約50kbである。一部の実施形態では、当該少なくとも3つの異なるポリヌクレオチドのグループの各々のうちで最も長いものは、両端を含む、少なくとも15kbの長さ、少なくとも20kbの長さ、少なくとも25kbの長さ、少なくとも30kbの長さ、少なくとも35kbの長さ、または少なくとも40kbの長さである。

10

NC中のポリヌクレオチド配列

【0074】

本開示のポリヌクレオチドのグループは、同じ配列のポリヌクレオチドを含むか、またはグループ中の全ポリヌクレオチド配列がすべて同じではないポリヌクレオチドを含むと予測される。

【0075】

一部の実施形態では、各グループ内のポリヌクレオチドは、同一の配列からなる。

20

【0076】

一部の実施形態では、各グループ内のポリヌクレオチドは、同一の配列を含む。各グループ内の全ポリヌクレオチドが同じであり得る配列の非限定的な例としては、アダプター配列、プライマー配列、インデックス化配列、制限部位、プロモーター配列、コンポーネント配列、単離配列、またはそれらの組み合わせが挙げられる。

【0077】

一部の実施形態では、各グループ内のポリヌクレオチドは、全ポリヌクレオチドで同じ配列である配列、および全ポリヌクレオチドで同じではない配列の両方を含む。例えば、各グループ内の全ポリヌクレオチドは、同じアダプター配列および／またはインデックス化配列を含み、コンポーネント配列は異なってもよい。他の実施形態では、各グループ内の全ポリヌクレオチドは、異なるアダプター配列および／またはインデックス化配列を含み、同じコンポーネント配列または単離配列を共有してもよい。

30

【0078】

一部の実施形態では、各グループ内のポリヌクレオチドの配列は、少なくとも5%同一、少なくとも10%同一、少なくとも20%同一、少なくとも30%同一、少なくとも40%同一、少なくとも50%同一、少なくとも60%同一、少なくとも70%同一、少なくとも80%同一、少なくとも90%同一、少なくとも95%同一、少なくとも97%同一、少なくとも98%同一、または少なくとも99%同一である。

【0079】

一部の実施形態では、各グループ内のポリヌクレオチドは、同一の配列を含まない。

40

【0080】

一部の実施形態では、一つ以上、またはすべてのポリヌクレオチドのグループは、少なくとも3つのポリヌクレオチドのサブグループを含み、当該各サブグループ内のポリヌクレオチドは同一の配列を含み、および当該各サブグループ内のポリヌクレオチドは任意の他のサブグループと同一の配列は含まない。

【0081】

本開示の正規化対照の一部の実施形態では、当該少なくとも3つのグループの各々の内の各ポリヌクレオチドは、同一の配列を含む。

【0082】

一部の実施形態では、当該少なくとも3つのグループの各1つのグループ内の複数のポリ

50

ヌクレオチドの配列は、任意の他のグループ内の複数のポリヌクレオチドの配列と同一の配列を含まない。

【0083】

一部の実施形態では、当該少なくとも3つのグループの各々のグループ内の複数のポリヌクレオチドの各々は、同一の配列を有さないポリヌクレオチドを含む。

【0084】

一部の実施形態では、当該少なくとも3つのポリヌクレオチドのグループの各々は、少なくとも3つのポリヌクレオチドのサブグループを含み、当該ポリヌクレオチドの各サブグループは、複数の同一配列のポリヌクレオチドを含み、および当該各サブグループの複数のポリヌクレオチドは、当該グループの他のサブグループと同一ではない。

【0085】

コンポーネント配列

本開示の正規化対照の一部の実施形態では、少なくとも1つのポリヌクレオチドのグループの配列は、例えば長さ、配列またはGC含量などの特定のプロパティを有するように設計され、選択され、または有することが知られているコンポーネント配列を含む。これによって、設計され、選択され、または公知のそれらプロパティ範囲が現れたポリヌクレオチドのグループが生成され得る。コンポーネント配列のプロパティは、シーケンシングサンプル中の標的配列のプロパティと合致するように設計または選択され得る。

【0086】

本明細書において使用される場合、「コンポーネント配列」という用語は、正規化対照中に存在するポリヌクレオチドの一部または全体を指す。正規化対照ポリヌクレオチドは、コンポーネント配列を含み、またはコンポーネント配列からなる。一部の実施形態では、コンポーネント配列はインシリコで設計され、いずれか生物から単離され、または誘導されない。

【0087】

一部の実施形態では、コンポーネント配列は、ランダム配列を含む。一部の実施形態では、コンポーネント配列は、シーケンシングサンプル中の標的配列の一つ以上の配列プロパティを模倣または共有するように設計される。

【0088】

一部の実施形態では、すべてのポリヌクレオチドのグループの配列が、コンポーネント配列を含む。一部の実施形態では、各ポリヌクレオチドのグループのコンポーネント配列は、任意の他のグループのコンポーネント配列と同じではない。

【0089】

一部の実施形態では、少なくとも1つのポリヌクレオチドのグループの少なくとも1つのサブグループの配列が、コンポーネント配列を含む。一部の実施形態では、少なくとも1つのグループの少なくとも3つの各サブグループの配列が、コンポーネント配列を含む。一部の実施形態では、すべてのグループの少なくとも3つのサブグループの各々の配列が、コンポーネント配列を含む。一部の実施形態では、各サブグループのコンポーネント配列は、任意の他のサブグループのコンポーネント配列と同じではない。

【0090】

例えば、3つのグループを有する正規化対照組成物において、それら各々が3つのサブグループを有し、各サブグループは異なるコンポーネント配列を有し、それらコンポーネント配列は同じではない。この例では、正規化対照組成物中のポリヌクレオチドの集団は、9つの異なるコンポーネント配列とともにポリヌクレオチドを含む。

【0091】

一部の実施形態では、コンポーネント配列は、約15bp～約3000bp、約50bp～約3000bp、約100bp～約3000bp、約1000bp～約3000bp、約1200～約3000bp、約1500bp～約3000bp、約15bp～約50bp、約15bp～約100bp、約15bp～約150bp、約15bp～約200bp、約15bp～約300bp、約15bp～約400bp、約15～約500bp、約50bp～約1200bp、約100bp～約1200bp、約150bp～約1200bp、または約150～約1100bp

10

20

30

40

50

の配列を含む。一部の実施形態では、コンポーネント配列は、約6bp～約3000bpの配列を含む。一部の実施形態では、コンポーネント配列は、約150bp～約500bpの配列を含む。

【0092】

本開示方法の一部の実施形態では、コンポーネント配列は、シーケンスサンプル中の標的配列と同一ではない。一部の実施形態では、コンポーネント配列は、シーケンシングサンプル中の標的配列と、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、30%、40%、50%、60%、70%、80%、85%、90%、95%、96%、97%、98%、99%または99.5%、99.8%または99.9%以下の同一性を有する。一部の実施形態では、コンポーネント配列は、標的配列と、20bpごとに約1bp、50bpごとに約1bp、150bpごとに約1bp、250bpごとに約1bp、または300bpごとに約1bp、異なってもよい。

【0093】

一部の実施形態では、コンポーネント配列は、ランダム配列を含み、またはランダム配列からなる。一部の実施形態では、ランダム配列は、ランダム配列ジェネレーターを使用して生成される。一部の実施形態では、ランダム配列は、参照配列にマッピングされない。一部の実施形態では、ランダム配列は、対象の任意の生物のゲノム、例えばNCBIのヌクレオチドデータベース中の任意の生物のゲノム、または正規化対照を使用して検出、定量もしくは別手段により分析する任意の生物のゲノムにマッピングされない。一部の実施形態では、ランダム配列は、ヒトDNAを切断するよう設計された90k gRNAアレイ中の任意のCRISPR関連（Cas）ガイドRNA（gRNA）により切断されない。一部の実施形態では、ランダム配列は、特定の配列性能が除外されるよう設計され得る。例えば、ランダム配列は、CRISPR/Cas gRNA認識部位、制限部位、転写因子結合部位、または反復配列などの性能が除外され得る。一部の実施形態では、ランダム配列は、標的配列のプロパティを模倣するよう設計され得る。例えば、ランダム配列は、標的配列と同じGC含量、または同じ頻度の特定の配列モチーフ（例えば、トリヌクレオチドリピート）を有して設計され得る。

【0094】

単離配列

本開示の正規化対照の一部の実施形態では、正規化対照ポリヌクレオチドは、単離配列を含む。本明細書において使用される場合、「単離配列」は、ある種のコンポーネント配列である。「単離配列」という用語は、生物から、もしくは別手段により既存配列から単離または誘導された配列を指す。単離配列は、ゲノム、ミトコンドリア、または葉緑体のDNA配列を含み得る。単離配列は、例えば、メッセンジャーRNA（mRNA）、トランスファーRNA（tRNA）、非コードRNA（ncRNA）、またはmicroRNAなどのRNA配列を含み得る。単離配列は、生物または既存配列から単離され得、および既存配列の断片を含み得る。あるいは、またはさらに、例えば配列シャッフリング、配列連結、ヌクレオチドの順序を反転させる、または他の配列内で入れ子にするなどの一つ以上の転換を受けて、既存配列から誘導されてもよい。

【0095】

一部の実施形態では、少なくとも1つのポリヌクレオチドのグループの配列が、単離配列を含む。一部の実施形態では、すべてのポリヌクレオチドのグループの配列が、単離配列を含む。

【0096】

一部の実施形態では、少なくとも1つのポリヌクレオチドのグループの少なくとも1つのサブグループの配列が、単離配列を含む。一部の実施形態では、少なくとも1つのグループのすべてのサブグループの配列が、単離配列を含む。一部の実施形態では、すべてのグループのすべてのサブグループの配列が、単離配列を含む。一部の実施形態では、各サブグループの単離配列は、任意の他のサブグループの単離配列と同じではない。

【0097】

一部の実施形態では、単離配列の長さは、約15bp～約500kbである。一部の実施形態

では、単離配列の長さは、約15bp～約100kbである。一部の実施形態では、単離配列の長さは、約15bp～約50kbである。一部の実施形態では、単離配列の長さは、約50bp～約50kbである。一部の実施形態では、単離配列の長さは、約500bp～約50kbである。一部の実施形態では、単離配列の長さは、約1000bp～約50kbである。

【0098】

一部の実施形態では、単離配列は、約6bp～約200,000bp、約15bp～約50,000bp、約500bp～約1500bp、約100bp～約1200bp、または約150bp～約600bpである。

【0099】

一部の実施形態では、単離配列は、少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、少なくとも約300bp、少なくとも約350bp、少なくとも約400bp、少なくとも約450bp、少なくとも約500bp、少なくとも約550bp、少なくとも約600bp、少なくとも約650bp、少なくとも約600bp、少なくとも約750bp、少なくとも約800bp、少なくとも約850bp、少なくとも約900bp、少なくとも約950bp、少なくとも約1000bp、少なくとも約1250bp、少なくとも約1500bp、少なくとも約2000bp、少なくとも約2500bp、または少なくとも約3000bpである。

【0100】

一部の実施形態では、単離配列は、約100bp、約200bp、約300bp、約400bp、約500bp、約600bp、約700bp、約800bp、約900bp、約1000bp、約1100bp、約1200bp、約1300bp、約1400bp、または約1500bpである。

【0101】

一部の実施形態では、単離配列は、ウイルス、細菌、真菌または真核生物から単離され、または誘導される。一部の実施形態では、ウイルスは、T4バクテリオファージ (T4) またはサイトメガロウイルス (CMV) である。一部の実施形態では、単離配列は、例えば、プラスミドなどのクローニングベクターから単離される。

【0102】

一部の実施形態では、正規化対照の単離配列が精製される核酸サンプルは、クローニングベクターである。一部の実施形態では、クローニングベクターは、細菌人工染色体 (BAC)、酵母人工染色体 (YAC)、コスミド、フォスミド、またはプラスミドである。一部の実施形態では、単離正規化対照が精製される核酸サンプルは、プラスミドである。一部の実施形態では、正規化対照配列は、例えば、ウイルス、細菌または真核寄生体などの対象の種からクローニングベクターへとクローニングされ、そして正規化対照は、本開示方法を使用して当該クローニングベクターから精製される。

【0103】

一部の実施形態では、単離配列は、シーケンスサンプル中の標的配列と同一ではない。一部の実施形態では、正規化対照中のポリヌクレオチドのいずれかの配列は、シーケンスサンプル中の標的配列に対し、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、30%、40%、50%、60%、70%、80%、85%、90%、95%、96%、97%、98%、または99%以下の同一性を有する。

【0104】

一部の実施形態では、単離配列は、複数の参照配列から単離され、または誘導される。「参照配列」とは、当分野に公知の標的配列と同一または類似した一つ以上の配列を指す。参照配列とは、例えば標的生物の遺伝子またはゲノムなどの一つの標的配列を指す場合もある。

【0105】

さらなる例として、「参照配列」とは、複数の異なる生物に由来する複数の配列を指す場合もある。生物は、ウイルス、細菌、真菌、または単細胞真核生物、例えば寄生性真核生物もしくは病原性真核生物、またはそれらの組み合わせであり得る。参照配列は、ゲノムDNA配列、cDNA配列、またはそれらの組み合わせであり得る。

【0106】

一部の実施形態では、正規化対照は、一つの参照配列から単離された、または誘導された複数の単離配列を含む。

【0107】

一部の実施形態では、正規化対照は、例えば複数の異なる生物など、複数の参照配列から単離された、または誘導された複数の単離配列を含む。一部の実施形態では、複数の参照配列は、例えばウイルス、細菌または真菌など（例えば、宿主が哺乳動物、植物または多細胞真核生物などの場合）、非宿主生物であると一般的にみなされる生物に由来する複数の配列を含む。複数の参照配列を使用することで、本明細書に記載される正規化対照において標的生物の集団レベルの特性パネルをモデル化させることができる。標的生物のパネルのうちの任意のものがシーケンスサンプル中に存在してもよく、その存在と力価は、本明細書に記載の方法を使用して測定することができる。シーケンスサンプル中の標的配列または標的生物が、シーケンスを行う前に判明していない実施形態において、複数の生物をモデル化する正規化対照を使用することで、識別と力価測定の正確度を高めることができる。

【0108】

一部の実施形態では、複数の単離配列は、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20個、少なくとも50個、少なくとも100個、少なくとも200個、少なくとも300個、少なくとも400個、少なくとも500個、少なくとも600個、少なくとも700個、少なくとも800個、少なくとも900個、少なくとも1,000個、少なくとも1,200個、少なくとも1,300個、少なくとも1,400個、少なくとも1,500個、少なくとも1,600個、少なくとも1,700個、少なくとも1,800個、少なくとも1,900個、少なくとも2,000個、少なくとも2,200個、少なくとも2,400個、少なくとも2,600個、少なくとも2,800個、少なくとも3,000個、少なくとも4,000個、少なくとも5,000個、少なくとも6,000個、少なくとも7,000個、少なくとも8,000個、少なくとも9,000個、少なくとも10,000個の参照配列、少なくとも11,000個の参照配列、もしくは少なくとも12,000個の参照配列から単離され、または誘導される。一部の実施形態では、複数の単離配列は、少なくとも10個の参照配列から単離され、または誘導される。一部の実施形態では、複数の単離配列は、少なくとも100個の参照配列から単離され、または誘導される。一部の実施形態では、複数の単離配列は、少なくとも500個の参照配列から単離され、または誘導される。一部の実施形態では、複数の単離配列は、少なくとも1000個の参照配列から単離され、または誘導される。一部の実施形態では、複数の単離配列は、少なくとも1,200個の参照配列から単離され、または誘導される。一部の実施形態では、複数の単離配列は、少なくとも1,500個の参照配列から単離され、または誘導される。一部の実施形態では、複数の単離配列は、少なくとも2,000個の参照配列から単離され、または誘導される。一部の実施形態では、各参照配列は、異なる生物に由来し、または同じ生物の異なる株もしくは亜種に由来する。一部の実施形態では、複数の参照配列は、異なる生物、または同じ生物の異なる株もしくは亜種に由来するゲノムDNA配列である。

【0109】

一部の実施形態では、複数の参照配列は、複数の生物に由来する参照配列を含む。一部の実施形態では、複数の参照配列は、少なくとも2個、少なくとも10個、少なくとも20個、少なくとも50個、少なくとも100個、少なくとも200個、少なくとも300個、少なくとも400個、少なくとも500個、少なくとも600個、少なくとも700個、少なくとも800個、少なくとも900個、少なくとも1,000個、少なくとも1,200個、少なくとも1,300個、少なくとも1,400個、少なくとも1,500個、少なくとも1,600個、少なくとも1,700個、少なくとも1,800個、少なくとも1,900個、少なくとも2,000個、少なくとも2,200個、少なくとも2,400個、少なくとも2,600個、少なくとも2,800個、少なくとも3,000個、少なくとも4,000個、少なくとも5,000個、少なくとも6,000個、少なくとも7,000個、少なくとも8,000個、少なくとも9,000個、または少なくとも10,000個の生物に由来す

る配列を含む。一部の実施形態では、複数の参照配列は、約1,000個の生物、約1,200個の生物、約1,400個の生物、約1,500個の生物、約1,600個の生物、約1,700個の生物、約1,800個の生物、約2,000個の生物、約2,200個の生物、約2,400個の生物、約2,500個の生物、約3,000個の生物、約5,000個の生物、約7,000個の生物、または約10,000個の生物に由来する配列を含む。

【0110】

一部の実施形態では、複数の単離配列は、参照配列の断片から単離され、または誘導される配列を含む。一部の実施形態では、参照配列断片は、参照配列の約10-100、約10-80、約10-70、約10-60、約10-50、約10-40、約10-30、約10-20、約20-100、約20-80、約20-60、約20-50、約20-40、約20-30、約10-35、約20-35、または約20-25個の連続するヌクレオチドである。一部の実施形態では、参照配列断片は、参照配列の約15~60、約20~40、約20~30、約15~32、20~32、または約25~35個の連続するヌクレオチドである。一部の実施形態では、参照配列断片は、参照配列の20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、または65個の連続する塩基対を含む。一部の実施形態では、参照配列断片は、参照配列の25、26、27、28、29、30、31、32、33、34、または35個の連続するbpを含む。一部の実施形態では、参照配列断片は、参照配列の31個または32個の連続するbpを含む。

【0111】

一部の実施形態では、参照配列断片は、シーケンスリードの平均長の半分未満である。例えば、NGSシーケンス法によって平均でそれぞれ120bpのリードが生成される場合、参照配列断片は60bp未満である。異なるNGSプラットフォームは、異なる長さのシーケンスリードを生成する。例えば、HiSeqおよびHiSeq Xは、約300bpの平均リード長であり、一方でMiSeqプラットフォームは、約300~600bpの平均リード長である。一部の実施形態では、平均NGSリード長は、約100bp、約150bp、約200bp、約250bp、約300bp、約350bp、約400bp、約450bp、約500bp、約550bp、約600bp、約650bp、約700bp、約750bp、約800bp、約850bp、約900bp、約950bp、約1000bp、約1100bp、約1200bp、約1300、about 1400bp、約1500bp、約1600bp、約1700bp、約1800bp、約1900bp、または約2000bpである。

【0112】

一部の実施形態では、複数の単離配列は、シャッフルされた参照配列断片を含む。配列をシャッフルする方法は当分野に公知であり、ランダムシャッフリング法およびCantorシャッフリング法が挙げられる。様々なプログラムを使用してヌクレオチド配列をシャッフルすることができ、そうした方法は当業者に公知である（例えば、www.bioinformatics.org/sms2/shuffle_dna.htmlを参照のこと）。本明細書において予期されるように、シャッフル化された参照配列断片は、参照配列の特性を一つ以上保持しており、と同時に、サンプル中の標的配列とは同一ではない。ゆえに例えばblastなどの方法によって標的配列と識別可能であることが予測される。

【0113】

一部の実施形態では、複数の単離配列は、連結された参照配列断片を含む。一部の実施形態では、参照配列断片はシャッフルおよび連結されて、例えば、所望の長さの単離配列が生成される。一部の実施形態では、単離配列は、シャッフルおよび連結された少なくとも第一の参照配列と、第二の参照配列の少なくとも二つの断片を含む。一部の実施形態では、当該第一および第二の参照配列は、生物ゲノム中で隣接していない。一部の実施形態では、当該第一および第二の参照配列は、異なる生物に由来する。単離配列は、シャッフルおよび連結された少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20個の参照配列断片を含むことができ、そのいずれの断片も、同じ参照配列もしくは参照生物、または異なる参照配列もしくは参照生物に由来する非隣接断片であってもよい。

【0114】

一部の実施形態では、単離配列は、(a)参照配列断片の分布により、少なくとも1つの参照配列を提示すること、(b)ビンに分割すること、(c)当該ビンの少なくともサブセットから複数の参照配列断片を選択すること、および(d)参照配列断片をシャッフル化し、および任意で連結して、単離配列を生成すること、のプロセスにより参照配列から誘導される。

【0115】

一部の実施形態では、正規化対照の単離配列は、(a)少なくとも1つの参照配列上でスライディングウィンドウを使用して生成された重複する参照配列断片の分布を生成すること、(b)当該分布をビンに分割すること、(c)当該ビンの少なくともサブセットから複数の参照配列断片を選択すること、および(d)当該参照配列断片を連結して、単離配列を生成すること、により生成される。

10

【0116】

一部の実施形態では、スライディングウィンドウは、1塩基対(bp)、2bp、3bp、4bp、5bp、6bp、7bp、8bp、9bp、10bp、11bp、12bp、13bp、14bp、15bp、16bp、17bp、18bp、19bp、または20bpのスライディングウィンドウを含む。スライディングウィンドウは長さNのウィンドウであり、参照配列に沿ってスライドして、例えば、1、2、3、4、5bpまたはそれ以上など、ウィンドウの長さによりオフセットされる参照配列断片を生成する。

【0117】

分布は、(1)GC含量パーセント、(2)エントロピー、(3)複雑度、(4)電子-イオン相互作用ポテンシャル(EIIP)、(5)長さ、またはそれらの組み合わせのうちの任意の一つの分布であり得る。分布を生成するために、参照配列は指定されたサイズ(本明細書においてKmerと称される場合もある)の断片に断断され、参照配列断片の数は、例えばGC含量パーセント、EIIP、長さまたはエントロピーなどの所与のパラメータにわたりビン化される。各ビンは、パラメータ分布のパーセントを表しており、例えば、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%または25%となる。

20

【0118】

分布は任意の数のビンに分割され得る。例えば、分布は、2、3、4、5、6、7、8、9、10、20、50、100、または1000のビンに分割され得る。一部の実施形態では、各ビンは、例えばGC含量パーセント、エントロピー、複雑度、長さ、またはEIIPなどのパラメータの0.1%、0.5%、1%、2%、5%、10%、15%または20%の分布を表している。

30

【0119】

一部の実施形態では、正規化対照の各グループ内のポリヌクレオチドは、5つのビンに分割された分布のうち、少なくとも3つのビンのうちの1つから選択された参照配列断片を使用して生成される。一部の実施形態では、正規化対照の各グループ内のポリヌクレオチドは、分布中の少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19または20のビンから選択された参照配列断片を使用して生成される。一部の実施形態では、分布は10のビンに分割され、参照配列断片は、GC含量パーセント分布からの10、30、50、70、80および90パーセンタイルのビンから選択される。一部の実施形態では、正規化対照の各グループ内のポリヌクレオチドは、分布中の各ビンから選択された参照配列断片を使用して生成される。

40

【0120】

一部の実施形態では、一つ以上の参照配列断片が、例えばGC含量パーセント分布などの分布のビンから選択される。一部の実施形態では、一つ以上の参照配列断片が、分布中の各ビンから選択される。一部の実施形態では、一つ以上の参照配列断片が、分布のビンのサブセットから選択される。一部の実施形態では、同じビンから選択された参照配列断片は連結されて、所望の長さの単離配列を形成する。これによって、選択されたビンサブサンプリングすることにより全分布にわたって正規化対照中の単離配列が、例えばGC含量、EIIP、長さおよびエントロピーなどの様々なパラメータの分布をモデル化することが可

50

能となり、参照配列と正規化対照との間で1:1の対応をとって分布全体の参照配列をモデル化するための非常に大量の正規化対照オリゴは必要ではなくなる。

【0121】

一部の実施形態では、各グループ内のポリヌクレオチドは、参照配列分布中の対応するピンからの参照配列断片と、(1) 同様のGC含量パーセント、(2) 同様のエントロピー、(3) 同様のEIIP、(4) 同様の長さ、または(5) 同様の複雑度、のうちの少なくとも3つを有する。

【0122】

同様のGC含量パーセント、EIIP、長さ、複雑度またはエントロピーとは、参照配列と正規化対照の平均GC含量、EIIP、長さ、複雑度またはエントロピーの比較を指す場合もある。例えば、正規化対照と参照配列は、参照配列の平均GC含量、EIIP、長さ、複雑度またはエントロピーが、正規化対照の平均GC含量、EIIP、長さ、複雑度またはエントロピーの0.1、0.5、1、2、3、4、5、6、7、8、9、10、11、12、13、14または15パーセント率以内である場合に、同様のGC含量パーセント、EIIP、長さ、複雑度またはエントロピーであり得る。あるいは、またはさらに、同様のGC含量、EIIP、長さ、複雑度またはエントロピーとは、参照配列中の一部またはすべての配列にわたる、および正規化対照の配列にわたる、GC含量、EIIP、長さ、複雑度またはエントロピーの分布を指す場合もある。類似性の度合いの決定方法は当分野に公知であり、限定されないが、コルモゴロフ-スミルノフ検定、Z検定、Q-Qプロット、および分散分析(ANOVA)が挙げられる。当業者であれば、対象の分布の特性(例えば、正規分布、ポアソン分布、ガウス分布または傾斜分布など)に基づき、適切な検定を選択することができるであろう。

【0123】

一部の実施形態では、複数の単離配列は、連結された参照配列断片を含む。一部の実施形態では、参照配列断片は、シャッフルされることなく連結される。単離配列を形成するために連結される参照配列断片が、参照配列に隣接していない場合、または異なる参照配列に由来する場合、非隣接配列または異なる源由来の配列との間を「結合」させることによって、NGS後に単離配列がblastアライメントを介して特定されることが可能となるはずである。結合があることによって、正規化対照中の単離配列が、サンプル中の標的配列から識別されることが可能となる。学説に拘束されることは望まないが、このアプローチに対しては、約15~約50bp、好ましくは約20~約35bpの配列断片が好ましい場合があることが予測される。その理由は、小さすぎる配列、例えばアライナー(aligner)のシード長より小さい配列などは堅固なアライメントを形成させるには固有性が不十分である可能性が高く、一方で大きすぎる配列は、結合を有さず、個々のNGSリードに捕捉されない。

【0124】

一部の実施形態では、単離配列は連結される少なくとも1つの参照配列の少なくとも2つの断片を含み、当該少なくとも2つの断片は、当該少なくとも1つの参照配列中で隣接していない。すなわち、当該少なくとも2つの断片は、少なくとも1bp、少なくとも2bp、少なくとも3bp、少なくとも4bp、少なくとも5bp、または任意のbp数で分離されている。

【0125】

一部の実施形態では、単離配列は、少なくとも第一の参照配列の少なくとも第一の断片、および少なくとも第二の参照配列の少なくとも第二の断片を含み、この場合において当該少なくとも第一および第二の断片は連結され、および当該第一および第二の参照配列は、同じ参照配列ではない。例えば、第一の断片はヒトアデノウイルス由来であり、第二の断片はカンジダ・アルビカンスに由来する。さらなる例としては、単離配列は、入れ子型参照配列断片：例えば、第二の参照生物由来の参照配列断片内に、第一の参照生物由来の参照配列断片など、を含むことができ、当該入れ子型参照配列断片は、さらに別の異なる生物由来の参照配列断片に任意で連結される。単離配列は、連結された少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20個の参照配列断片を含むことができ、そのいずれも、同じ参照配列もしくは参照生物、また

は異なる参照配列もしくは参照生物に由来する非隣接断片であってもよい。

【0126】

一部の実施形態では、単離配列（例えば、シャッフル化単離配列または連結単離配列）は、ヒトDNAを切断するよう設計された90k gRNAアレイ中の任意のCRISPR関連（Cas）ガイドRNA（gRNA）により切断されない。一部の実施形態では、単離配列は、特定の配列性能が除外されるよう設計され得る。例えば、CRISPR/Cas gRNA認識部位、制限部位、転写因子結合部位、または反復配列などの性能を排除したシャッフル化単離配列または連結単離配列を選択することができる。

【0127】

配列プロパティ

一部の実施形態では、ポリヌクレオチドの少なくとも3つのグループの各々の配列は、シーケンスサンプル中の標的配列と共通したプロパティを少なくとも1つ有する。一部の実施形態では、コンポーネント配列は、シーケンスサンプル中の標的配列と共有したプロパティを少なくとも1つ有するように設計される。一部の実施形態では、単離配列は、シーケンスサンプル中の標的配列と共通したプロパティを少なくとも1つ有することが知られているために選択された生物、またはクローニングベクター中の配列から単離される。

【0128】

一部の実施形態では、当該少なくとも1つのプロパティは、転位因子配列、ウイルス配列、細菌配列、真菌配列、真核寄生体配列、または一つ以上のヒト遺伝子配列の配列プロパティを含む。

【0129】

一部の実施形態では、転位因子のプロパティは、転位因子配列を含む。転位因子は、ゲノム中の位置を変えることができるDNA配列である。転位因子（トランスポゾン）は概して、トランスポザゼ酵素をコードするオープンリーディングフレームに隣接する逆位末端配列（ITR:inverted terminal repeats）の対からなる。したがって転位因子の例示的なプロパティは、ITR配列、トランスポザゼ配列、またはトランスポゾンの任意の配列を含む。トランスポゾンの任意の配列、またはトランスポゾンの配列プロパティを、本開示のコンポーネント配列に組み込むことができる。本開示の単離配列は、トランスポゾンの任意の配列、またはトランスポゾンの配列プロパティを含むことができる。

【0130】

一部の実施形態では、ウイルス配列のプロパティは、GC含量パーセント、反復配列因子、逆位末端配列（ITR）、配列内リボソーム進入部位（IRES:internal ribosome entry site）、タンパク質コード配列、転写後制御因子、転写制御因子、プロモーター配列、シス作用RNA因子、RNA構造因子、ゲノムパッケージングシグナル、5'非翻訳領域（5'UTR）配列、3'非翻訳領域（3'UTR）配列、またはそれらの組み合わせを含む。ウイルスの任意の配列、またはウイルスの配列プロパティを、本開示のコンポーネント配列に組み込むことができる。

【0131】

一部の実施形態では、細菌配列のプロパティは、GC含量パーセント、反復配列、micro RNA結合部位、配列内リボソーム進入部位（IRES）、タンパク質コード配列、転写制御因子、プロモーター配列、5'UTR配列、3'UTR配列、またはそれらの組み合わせを含む。一部の実施形態では、反復配列因子は、ポリAモチーフ、ポリTモチーフ、ポリGモチーフ、ポリCモチーフ、ジヌクレオチドモチーフ、トリヌクレオチドモチーフ、テトラヌクレオチドモチーフ、ペンタヌクレオチドモチーフ、ヘキサヌクレオチドモチーフ、ヘプタヌクレオチドモチーフ、オクタヌクレオチドモチーフ、ノナヌクレオチドモチーフ、散在反復配列因子、リボソームRNA配列、トランスファーRNA（tRNA）配列、末端逆位反復（TIR:terminal inverted repeat）、非自律性小型逆位反復転位因子（MITE:non-autonomous miniature inverted repeat transposable element）、Clustered Regularly Interspaced Short Palindromic Repeat（CRISPR）またはそれらの組み合わせを含む。ポリAモチーフ、ポリTモチーフ、ポリGモチーフ、およびポリCモチーフは、A、T、G

10

20

30

40

50

およびCの連続である（例えば、4～50個のA、T、GまたはC）。細菌の任意の配列、または細菌の配列プロパティを、本開示のコンポーネント配列に組み込むことができる。本開示の単離配列は、細菌の配列、または細菌の配列プロパティを含むことができる。

【0132】

一部の実施形態では、真菌配列のプロパティは、GC含量パーセント、反復配列因子、microRNA結合部位、配列内リボソーム進入部位（IRES）、タンパク質コード配列、転写制御因子、プロモーター配列、5' UTR配列、3' UTR配列、セントロメア配列、テロメア配列、テロメア周辺配列、ミトコンドリア配列またはそれらの組み合わせを含む。真菌の任意の配列、または真菌の配列プロパティを、本開示のコンポーネント配列に組み込むことができる。本開示の単離配列は、真菌の配列、または真菌の配列プロパティを含むことができる。

10

【0133】

一部の実施形態では、1つ以上のヒト遺伝子のプロパティは、GC含量パーセント、反復配列、タンパク質コード配列、イントロン配列、5' UTR配列、3' UTR配列、転写制御因子配列、プロモーター配列、microRNA結合配列、またはそれらの組み合わせを含む。例示的な実施形態では、当該1つ以上のヒト遺伝子のプロパティは、ファミリー中の複数のタンパク質で共通する保存タンパク質ドメインをコードする配列である。タンパク質ドメインの非限定的な例は、例えばジンクフィンガーDNA結合ドメインまたはホメオドメインなどのDNA結合ドメインである。例示的な実施形態では、当該1つ以上のヒト遺伝子のプロパティは、同族microRNAにより制御される全遺伝子に共通するmicroRNA結合部位である。ヒト遺伝子の任意の配列、またはヒト遺伝子の配列プロパティを、本開示のコンポーネント配列に組み込むことができる。本開示の単離配列は、ヒト遺伝子の配列、またはヒト遺伝子の配列プロパティを含むことができる。

20

【0134】

一部の実施形態では、真核寄生体配列のプロパティは、GC含量パーセント、反復配列、microRNA結合部位、タンパク質コード配列、転写制御因子、プロモーター配列、5' UTR配列、3' UTR配列、セントロメア配列、テロメア配列、テロメア周辺配列、ミトコンドリア配列またはそれらの組み合わせを含む。真核寄生体の任意の配列、または配列プロパティを、本開示のコンポーネント配列に組み込むことができる。本開示の単離配列は、真核寄生体の配列、または真核寄生体の配列プロパティを含むことができる。

30

【0135】

一部の実施形態では、正規化対照配列と標的配列に共通する少なくとも1つのプロパティは、エントロピーを含む。エントロピーは、ヌクレオチド配列の情報量および複雑度の測定尺度である。エントロピーの計算方法は当業者に公知である（例えば、Bioinformatics, Vol. 27 2011, Pages 1061-1067を参照のこと）。

【0136】

一部の実施形態では、正規化対照配列と標的配列に共通する少なくとも1つのプロパティは、複雑度を含む。複雑度（Nにより示される）は、所与のポリヌクレオチドセグメント中、もしくはゲノムの構成要素中の固有塩基対、または非反復性のヌクレオチドの数である。ポリヌクレオチドの一部が反復している場合、Nは、配列の長さ（L）とは異なってくる。

40

【0137】

一部の実施形態では、正規化対照配列と標的配列に共通する少なくとも1つのプロパティは、GC含量パーセントを含む。

【0138】

本開示の正規化対照の一部の実施形態では、ポリヌクレオチドの1つ以上のグループまたはサブグループのGC含量は、シーケンスサンプル中の標的配列と同じである。一部の実施形態では、当該グループまたはサブグループの全ポリヌクレオチドのGC含量が、標的配列と同じである。一部の実施形態では、コンポーネント配列および／または単離配列のGC含量は、標的配列と同じである。GC含量は、NGS中のリードカウントに影響を及ぼすこと

50

が知られている（GCバイアス）。低GC含量の配列または高GC含量の配列は、NGS中に生成するリード数が実際よりも少なく表される傾向がある。そのため、NC組成物において標的配列のGC含量を模倣することで、NGS中の標的配列の振る舞いをモデル化するNCの能力が増強される。

【0139】

一部の実施形態では、グループ内のポリヌクレオチドは、同一のGC含量パーセントを有する。一部の実施形態では、グループ内のポリヌクレオチドは、すべての他のグループ内のポリヌクレオチドと同一のGC含量パーセントを有する。一部の実施形態では、グループ内のポリヌクレオチドは、任意の他のグループ内のポリヌクレオチドと同一のGC含量パーセントを有さない。

10

【0140】

一部の実施形態では、各グループ内のポリヌクレオチドのGC含量パーセントは、40%以上、60%以下である。一部の実施形態では、各グループ内のポリヌクレオチドのGC含量パーセントは、43%以上56%以下である。

【0141】

一部の実施形態では、正規化対照配列と標的配列に共通する少なくとも1つのプロパティは、電子-イオン相互作用ポテンシャル（EIIP）を含む。EIIPは、ヌクレオチド配列中のヌクレオチドの非局在化電子の平均エネルギーと規定することができる。ヌクレオチド配列は、各ヌクレオチドと、対応するEIIP値を置き換えることにより、数値で表されるEIIP配列へと転換されることができる。例えば、 $A=0.1260$ 、 $C=0.1340$ 、 $T=0.1335$ 、 $G=0.0806$ である。一部の実施形態では、EIIP値は、例えば単離配列または参照配列断片などの配列のすべて、または一部で合計されることができる。

20

【0142】

一部の実施形態では、正規化対照配列と標的配列に共通する少なくとも1つのプロパティは、GC含量、EIIP、およびエントロピーを含む。GC含量、EIIPおよびエントロピーはすべて配列依存性であり、したがって一緒に変化する。一部の実施形態では、正規化対照は、これらパラメータの3つすべてが標的配列と共通する。

【0143】

例えばGC含量パーセント、EIIP、エントロピーまたは複雑度などのパラメータが正規化対照配列と標的配列もしくは参照配列と類似している度合いは、当分野に公知の様々な統計的方法により決定することができる。そうした方法としては限定されないが、分散分析（ANOVA）、例えばスチューデントt検定などのt検定、コルモゴロフ-スミルノフ（KS）検定およびQ-Qプロットなどが挙げられる。一部の実施形態では、正規化対照は、修飾ヌクレオチドを含む。任意のタイプのヌクレオチド修飾が、本開示の範囲内に予測される。本開示のヌクレオチド修飾の非限定的な例は、以下である。

30

【0144】

本開示方法に使用されるヌクレオチド修飾は、任意のヌクレオチド（アデニン、シトシン、グアニン、チミンまたはウラシルなど）に存在することができる。これらヌクレオチド修飾は、デオキシリボ核酸（DNA）またはリボ核酸（RNA）に存在することができる。これらヌクレオチド修飾は、二本鎖DNA分子もしくは一本鎖DNA分子、または二本鎖RNA分子もしくは一本鎖RNA分子で存在することができる。

40

【0145】

一部の実施形態では、ヌクレオチド修飾は、アデニン修飾またはシトシン修飾を含む。

【0146】

一部の実施形態では、アデニン修飾は、アデニンメチル化を含む。一部の実施形態では、アデニンメチル化は、 N^6 -メチルアデニン（6mA）を含む。 N^6 -メチルアデニン（6mA）は、原核生物ゲノムと真核生物ゲノムの両方に存在している。

【0147】

一部の実施形態では、アデニンメチル化は、EcoKIメチル化を含む。EcoKIメチル化は、EcoKIメチラーゼにより行われるタイプのDNAヌクレオチド修飾である。EcoKIメチラ

50

ーゼは、AAC(N₆)GTGC配列とGCAC(N₆)GTT配列のアデニン残基を修飾する。EcoKIメチラーゼとEcoKIメチル化は原核生物に存在する。

【0148】

一部の実施形態では、アデニン修飾は、グリシンによりN⁶で修飾されるアデニン（モミレーション（momylation））を含む。モミレーションは、N⁶-(1-アセトアミド)-アデニンとアデニンを交換する。モミレーションは例えばバクテリオファージなどのウイルス中に存在する。

【0149】

一部の実施形態では、修飾は、シトシン修飾を含む。一部の実施形態では、ゲノム中のシトシン修飾の存在量とタイプは、種によって変化する。一部の実施形態では、ゲノム中のシトシン修飾の位置（特に制限酵素認識部位内での位置）は、種によって変化する。

【0150】

一部の実施形態では、シトシン修飾は、5-メチルシトシン（5mC）、5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）、5-カルボキシルシトシン（5caC）、5-グルコシルヒドロキシメチルシトシン（5ghmC）、または3-メチルシトシン（3mC）を含む。

【0151】

一部の実施形態では、シトシン修飾は、シトシンメチル化を含む。一部の実施形態では、シトシンメチル化は、5-メチルシトシン（5mC）またはN⁴-メチルシトシン（4mC）を含む。一部の実施形態では、4mCシトシンメチル化は、細菌中に存在する。一部の実施形態では、シトシンメチル化は、Dcmメチル化を含む。一部の実施形態では、シトシンメチル化は、DNMT1メチル化、DNMT3Aメチル化、またはDNMT3Bメチル化を含む。

【0152】

一部の実施形態では、シトシンメチル化は、CpGメチル化、CpAメチル化、CpTメチル化、CpCメチル化、またはそれらの組み合わせを含む。一部の実施形態では、シトシンメチル化は、CpGメチル化を含む。

【0153】

一部の実施形態では、シトシン修飾は、5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）、5-カルボキシルシトシン（5caC）、5-グルコシルヒドロキシメチルシトシン、または3-メチルシトシンを含む。

【0154】

標的生物

本開示は、サンプル中の一つ以上の標的生物の力価を決定する組成物および方法を提供する。標的生物は、本明細書に記載される標的配列を一つ以上含む生物である。

【0155】

一部の実施形態では、サンプルは、宿主生物と、一つ以上の非宿主生物の混合物を含み、そして標的生物は、非宿主生物のうちの一つ以上である。一部の実施形態では、宿主生物が、標的生物である。

【0156】

一部の実施形態では、非宿主標的生物は、病原体である。一部の実施形態では、標的生物は、ウイルス、細菌、真菌または真核病原体である。例えば一部の実施形態では、標的生物は、ヒト臨床サンプル中のウイルス病原体であり、標的配列は、当該ウイルス病原体の配列である。

【0157】

一部の実施形態では、標的生物は共生生物（symbiote）または共生生物（commensal organism）である。

【0158】

一部の実施形態では、標的生物は、メタゲノムサンプル中の種混合物中の種である。

【0159】

標的配列

10

20

30

40

50

一部の実施形態では、NCは、NGS後にシーケンスサンプル中の標的配列がシーケンスリードへと転換されることに影響を及ぼす可能性がある変数の要因となるよう設計される。例えば、NCの配列は、本開示の正規化対照と方法を使用して開始サンプル中の存在量が決定される標的配列の核酸のGC含量と合致するように設計されることができる。

【0160】

本開示の正規化対照の一部の実施形態では、コンポーネント配列および／または単離配列は、シーケンスサンプル中の標的配列と共通したプロパティを少なくとも1つ有する。標的配列とコンポーネント配列の間で共有される共通プロパティによって、サンプル調製および／またはライブラリ調製ならびにシーケンスの間に、NCのコンポーネント配列が標的配列に擬態することが可能となる。これによって、サンプル中の標的配列の開始濃度からNGS後の標的配列のリード数への転換を、NCがモデル化することが可能となる。

10

【0161】

一部の実施形態では、正規化対照中の少なくとも1つのポリヌクレオチドのサブグループの配列が、シーケンスサンプル中の標的配列と、少なくとも1つの配列プロパティを共有する。一部の実施形態では、当該少なくとも1つの配列プロパティは、転位因子配列、ウイルス配列、細菌配列、真菌配列、真核寄生体配列、または一つ以上のヒト遺伝子配列の配列プロパティを含む。

【0162】

標的配列の全ての配列プロパティが、本開示の正規化対照組成物の範囲内にあるとみなされる。一部の実施形態では、少なくとも1つのプロパティは、転位因子配列、ウイルス配列、細菌配列、真菌配列、真核寄生体配列、または一つ以上のヒト遺伝子の配列の配列プロパティを含む。一部の実施形態では、当該少なくとも1つのプロパティは、GC含量、エントロピー、複雑度、長さ、EIIP、またはそれらの組み合わせを含む。

20

【0163】

一部の実施形態では、シーケンスサンプルは、宿主核酸と非宿主核酸の混合物を含む。一部の実施形態では、非宿主は、標的配列を含む。

【0164】

一部の実施形態では、宿主は、真核生物である。一部の実施形態では、宿主は、昆虫、動物または植物である。一部の実施形態では、動物は、ヒトである。

【0165】

一部の実施形態では、宿主は、任意の哺乳動物の生物である。特定の実施形態では、哺乳動物は、ヒトである。他の実施形態では、哺乳動物は、家畜動物、例えばウマ、ヒツジ、ウシ、ブタまたはロバである。他の実施形態では、哺乳動物の生物は、ヒトに慣れたペット、例えばネコ、イヌ、スナネズミ、マウスまたはラットである。他の実施形態では、哺乳動物は、ある種のサルである。

30

【0166】

一部の実施形態では、宿主は、任意のトリ、または鳥類の生物である。鳥類の生物としては限定されないが、ニワトリ、七面鳥、アヒルおよびガチョウが挙げられる。

【0167】

一部の実施形態では、宿主は、昆虫である。昆虫としては限定されないが、ミツバチ、孤立性のハチ (solitary bee)、アリ、ハエ、カリバチ、および蚊が挙げられる。

40

【0168】

一部の実施形態では、宿主は、植物である。一部の特定の実施形態では、植物は、イネ、トウモロコシ、コムギ、バラ、ブドウ、コーヒー、果実、トマト、ジャガイモまたは綿花である。

【0169】

一部の実施形態では、非宿主は、共生生物 (symbiote)、共生生物 (commensal organism)、寄生体、または病原体を含む。一部の実施形態では、非宿主は、共生生物 (symbiote)、共生生物 (commensal organisms)、寄生体、または病原体のうちの一つ以上の種を含む。

50

【0170】

一部の実施形態では、非宿主は、病原体である。一部の実施形態では、非宿主は、ある種のウイルス、ある種の細菌、ある種の真菌、またはある種の真核寄生体である。一部の実施形態では、非宿主は、ある種の藻類である。

【0171】

一部の実施形態では、真核寄生体は、哺乳動物の寄生体である。一部の実施形態では、寄生体は、蠕虫である。他の実施形態では、寄生体は、マラリアの原因となる寄生体である。他の実施形態では、寄生体は、リーシュマニア病の原因となる寄生体である。他の実施形態では、寄生体は、アメーバである。

【0172】

一部の実施形態では、非宿主は、ある種の細菌である。特定の実施形態では、寄生体は、結核の原因となる細菌である。

【0173】

一部の実施形態では、標的配列は、シーケンスサンプル中の非宿主配列である。

【0174】

一部の実施形態では、標的配列は、参照配列を含む。一部の実施形態では、標的配列は、複数の参照配列を含む。一部の実施形態では、標的配列は、少なくとも2個、少なくとも10個、少なくとも20個、少なくとも50個、少なくとも100個、少なくとも200個、少なくとも300個、少なくとも400個、少なくとも500個、少なくとも600個、少なくとも700個、少なくとも800個、少なくとも900個、少なくとも1,000個、少なくとも1,200個、少なくとも1,300個、少なくとも1,400個、少なくとも1,500個、少なくとも1,600個、少なくとも1,700個、少なくとも1,800個、少なくとも1,900個、少なくとも2,000個、少なくとも2,200個、少なくとも2,400個、少なくとも2,600個、少なくとも2,800個、少なくとも3,000個、少なくとも4,000個、少なくとも5,000個、少なくとも6,000個、少なくとも7,000個、少なくとも8,000個、少なくとも9,000個、または少なくとも10,000個の参照配列を含む。一部の実施形態では、各参照配列は、異なる生物に由来する。一部の実施形態では、当該複数の参照配列は、異なる生物に由来するゲノムDNA配列またはcDNA配列である。一部の実施形態では、標的配列は、少なくとも2個、少なくとも10個、少なくとも20個、少なくとも50個、少なくとも100個、少なくとも200個、少なくとも300個、少なくとも400個、少なくとも500個、少なくとも600個、少なくとも700個、少なくとも800個、少なくとも900個、少なくとも1,000個、少なくとも1,200個、少なくとも1,300個、少なくとも1,400個、少なくとも1,500個、少なくとも1,600個、少なくとも1,700個、少なくとも1,800個、少なくとも1,900個、少なくとも2,000個、少なくとも2,200個、少なくとも2,400個、少なくとも2,600個、少なくとも2,800個、少なくとも3,000個、少なくとも4,000個、少なくとも5,000個、少なくとも6,000個、少なくとも7,000個、少なくとも8,000個、少なくとも9,000個、または少なくとも10,000個の生物に由来する複数の参照配列を含む。

【0175】

一部の実施形態では、標的配列は非宿主配列を含み、当該非宿主は、ある種のウイルス、ある種の細菌、ある種の真菌、またはある種の真核寄生体である。

【0176】

一部の実施形態では、非宿主は、共生生物を含む。共生生物とは、相互的または相利共生的な生物と呼ばれる場合もあり、別の生物と、両者にとって相互的な利益をもたらす関係性にある生物である。

【0177】

一部の実施形態では、非宿主は、共生生物 (commensal organism) を含む。共生生物の相互関係においては、一方の生物 (例えば、非宿主) が利益を得て、他方の生物 (例えば、宿主) は影響を受けない。共生生物または相互生物は、食物と住居、サポート、移動またはそれらの組み合わせを宿主に依存する場合がある。宿主は限定されないが、感染からの防御、消化改善および免疫調節をはじめとする様々な利益をこの相

10

20

30

40

50

互関係から受け取ることができる。相互的關係および／または片利共生的關係の非限定的な例は、腸管に生息する微生物集団（腸内微生物叢または微生物叢）と宿主の關係である。腸内微生物叢は腸管の粘膜バリアの維持を補助し、例えばビタミンなどの栄養素を提供し、病原体から防御し、そして適切な免疫機能の維持を補助する。

【0178】

一部の実施形態では、非宿主は複数の種、または同種由来の異なる生物を含む。一部の実施形態では、非宿主は、少なくとも5、少なくとも10、少なくとも50、少なくとも100、少なくとも200、少なくとも500、少なくとも1000、少なくとも1,500、または少なくとも2,000の種、同じ種由来の異なる生物、またはそれらの組み合わせを含む。一部の実施形態では、非宿主は、例えば2,000の種、例えば少なくとも2,000の腸管細菌叢の種または微生物叢の生物を含む。

10

【0179】

一部の実施形態では、標的配列は、病原性と関連する非宿主配列である。病原性と関連する例示的な非宿主配列としては限定されないが、抗生物質耐性遺伝子、毒性因子、または真核寄生体と関連する病原性マーカーが挙げられる。

【0180】

毒性因子は、例えば宿主の感染において細菌、ウイルス、または真菌などの微生物の病原性の一因となる遺伝子または遺伝子グループである。毒性とは、微生物により宿主に与えられるダメージの程度を指し、毒性が大きくなると、ダメージも大きくなる。

【0181】

限定ではないが、例として毒性因子としては宿主においてニッチのコロニー形成（例えば、宿主細胞への非宿主の付加）、免疫回避（宿主の免疫反応からの回避）、免疫抑制（宿主免疫反応の非宿主による抑制）、（細胞内非宿主の）細胞進入および細胞脱出、ならびに宿主からの栄養分の取得、の一因となる遺伝子が挙げられる。宿主のコロニー形成を補助し、促進するために使用される例示的な毒性因子としては限定されないが、接着、侵入および抗貪食の因子が挙げられる。

20

【0182】

宿主へのダメージを増加させる例示的な毒性因子としては限定されないが、毒素、破壊性酵素、および宿主のGTPaseを制御するタンパク質が挙げられる。例示的な毒素としては、内毒素と外毒素が挙げられる。内毒素（リポ多糖またはLPS）は、細菌の細胞壁成分であり、宿主において炎症を誘発させる能力を有する。外毒素は、一部の細菌から能動的に分泌される毒素であり、宿主に対し、限定されないが特定の生化学経路の阻害をはじめとする広範な効果を有する。例示的な外毒素としては限定されないが、破傷風菌により分泌される破傷風毒素、ボツリヌス菌により分泌されるボツリヌス毒素、ならびに大腸菌、ウェルシュ菌、炭そ菌およびクロストリジウム・ディフィシレ菌により分泌される毒素が挙げられる。一部の実施形態では、内毒素は、真菌により分泌される（マイコトキシンと呼ばれる）。マイコトキシンの例には、アスペルギルスにより産生されるアフラトキシンが含まれる。

30

【0183】

一部の実施形態では、毒性因子は、破壊性酵素である。破壊性酵素は、宿主の細胞と組織にダメージを与えることにより毒性をもたらす酵素である。例示的な破壊性酵素としては限定されないが、プロテアーゼ、リパーゼおよびDNaseが挙げられる。例えば大腸菌および黄色ブドウ球菌などの細菌により産生される溶血素は、赤血球を含む宿主細胞を破壊する。

40

【0184】

一部の実施形態では、毒性因子は、宿主のGTPase活性を制御するタンパク質である。毒性因子はグアニンヌクレオチド交換因子（GEF：guanine nucleotide exchange factor）またはGTPase活性化タンパク質（GAP）とし作用して、宿主のGTPase活性を改変し、あるいは宿主のGTPase自体を共有結合的に改変することができる。

【0185】

50

例示的なウイルス毒性遺伝子としては、ウイルスの潜伏をコントロールするウイルス遺伝子が挙げられる。さらなるウイルス毒性遺伝子の例としては、宿主免疫反応のウイルスの回避の一因となるウイルス遺伝子、ならびに例えば宿主細胞への結合と進入、またはビリオンの産生および放出などのウイルスのライフサイクルの態様の一因となるウイルス遺伝子が挙げられる。

【0186】

一部の実施形態では、毒性因子は染色体にあり、内在性である。例えば、リポ多糖（内毒素）などの細菌毒性因子は染色体にある。

【0187】

一部の実施形態では、毒性因子は、遺伝子の水平伝達を介して拡散される可動性の遺伝因子によりコードされる。一部の実施形態では、可動性の遺伝因子は、プラスミド、転位遺伝因子、バクテリオファージ因子またはリボザイムである。一部の実施形態では、毒性因子の水平遺伝子伝達は、当該毒性因子の獲得前には片利共生または相利共生（共生）微生物であった微生物を、病原性微生物へと転換させる。

【0188】

一部の実施形態では、非宿主は、ウイルス、細菌または真菌であり、標的配列は、毒性因子の配列を含む。

【0189】

一部の実施形態では、非宿主は、細菌であり、標的配列は、抗生物質耐性遺伝子の配列を含む。一部の実施形態では、抗生物質耐性遺伝子は、水平遺伝子伝達を介して拡散される可動性の遺伝因子によりコードされる。一部の実施形態では、可動性の遺伝因子は、プラスミド、転位遺伝因子、バクテリオファージ因子またはリボザイムである。

【0190】

一部の実施形態では、抗生物質耐性遺伝子は、キノロン系、 β -ラクタム系、マクロライド系、テトラサイクリン系、スルホンアミド系、アミノグリコシド系、またはバンコマイシン系に対する耐性を伝達する。

【0191】

一部の実施形態では、シーケンスサンプルは、メタゲノムサンプルを含む。本明細書において使用される場合、「メタゲノム」という用語は、本開示の組成物および方法を使用した、例えば生物混合サンプルの直接的な分析を指す。

【0192】

一部の実施形態では、メタゲノムサンプルは、環境サンプルである。環境サンプルの例としては限定されないが、土壌、水および空気のサンプルが挙げられる。一部の実施形態では、環境サンプルは、例えば病院の表面など、人工物の表面から採取されたサンプルを含む。

【0193】

一部の実施形態では、サンプルは環境サンプルを含み、標的配列は毒性因子の配列、抗生物質耐性遺伝子、または病原性と関連する配列を含む。例えば、水平遺伝子伝達を介した抗生物質耐性の発現は、公衆衛生上の大きな問題となっている。ゆえに例えば病院または農園など、抗生物質が使用されている環境から採取されたメタゲノムサンプルにおける抗生物質耐性遺伝子の頻度を本開示の組成物および方法を使用して定量することは有益である。

【0194】

一部の実施形態では、本開示の正規化対照は、サンプル抽出および／またはライブラリ調製の間、シーケンスサンプル中の標的配列、すなわち非宿主配列のプロパティを模倣するように設計される。一部の実施形態では、正規化対照のポリヌクレオチドの集団は、核酸サンプルから精製され、当該核酸サンプルは、ウイルス、細菌、真菌もしくは真核寄生体から単離または誘導される。一部の実施形態では、NCが精製されるウイルス、細菌、真菌または真核生物は、非宿主のウイルス、細菌、真菌または真核寄生体に類似しているが、同一ではない。

【0195】

一部の実施形態では、非宿主はウイルスであり、正規化対照のポリヌクレオチドは関連ウイルス、または非宿主に似たプロパティを有するウイルスから精製される。例えば非宿主がレンチウイルスである場合、正規化対照のポリヌクレオチドは、関連種のレンチウイルスから精製される。一部の実施形態では、ウイルスは、T4バクテリオファージまたはサイトメガロウイルス（CMV）である。あるいは非宿主が細菌である場合、正規化対照のポリヌクレオチドは、細菌の関連種、または非宿主と類似したプロパティを有する細菌の種から精製される。

【0196】

NCは、一つ以上の生物に由来してもよい。例えば、NCは、宿主／非宿主混合サンプル中の非宿主配列の一つ以上のプロパティを模倣する、例えばウイルス、細菌、真菌または真核生物などの一つ以上の生物または生物種から抽出されたポリヌクレオチドを含んでもよい。

10

【0197】

ある実施形態では、NCを使用して、サンプル中の異なる核酸の相対量が決定される。例えば、正規化対照を使用して、感染宿主（例えばウイルスまたは細菌に感染したヒト宿主）に由来するサンプル中の病原体（非宿主）の存在量が決定されてもよい。病原体の存在量は、病原体にマッピングされるNGSリード数を、NCにマッピングされるNGSリード数で割り、様々なバックグラウンド宿主含量を補正することにより決定されてもよい。次いでこの値を、同様に作成され、正規化された標準曲線に位置づけてもよく、病原体の力価は当該標準曲線から推測されてもよい。そうした方法を使用することで、病原体の絶対量の決定は必須ではなくなる。

20

【0198】

ある実施形態では、NCを使用して、サンプル中の異なる核酸の絶対量が決定される。例えば、NCを使用して、感染宿主に由来するサンプル（例えばウイルスまたは細菌に感染したヒト宿主に由来するサンプル）中の病原体（非宿主）の絶対量が決定されてもよい。病原体の存在量は、病原体にマッピングされるNGSリード数を、NCにマッピングされるNGSリード数で割り、様々なバックグラウンド宿主含量を補正することにより決定されてもよい。この比率と、NCのインプット量（例えばバクテリオファージまたはポリヌクレオチドのコピー量）に基づき、相対的な病原体ゲノムコピー量が決定されてもよい。次いでこの値は、病原体のゲノムサイズに対して補正され、開始材料中の「絶対的な」病原体の存在量が導かれてもよい。

30

【0199】

一部の実施形態では、正規化対照ポリヌクレオチドが精製される対象種は、非宿主の種に類似しているが、同一ではない。

【0200】

一部の実施形態では、シーケンスサンプルは、あるタイプの癌であると診断された対象に由来する癌関連サンプルを含む。癌の進行は複数工程のプロセスであり、経時的な遺伝子変化の蓄積を必要とする。一部の実施形態では、癌をもたらす遺伝子変化には、促進性癌遺伝子を活性化し、腫瘍抑制遺伝子とDNA修復遺伝子を脱制御する遺伝子変化が含まれる。癌をもたらす遺伝子変化によって、それまでは正常であった細胞が細胞分割の制御から逸脱し、チェックを受けずに分割するようになる。ゆえに癌は、様々な変異を有する細胞の不均一な集団を含む。したがって一部の実施形態では、癌関連サンプルは、対象の非癌性細胞中、または癌を有していない対象に由来する細胞中、または異なるタイプの癌を有する対象に由来する細胞中の核酸と比較して、少なくとも1つの遺伝子変化を有する核酸（例えば、遺伝子、DNA、またはRNA）を含む細胞の集団を含む。

40

【0201】

一部の実施形態では、対象の変異は、癌サンプル中の不均一な集団において、稀にしか存在しない。しかし癌サンプル中の特定の変異は、治療に対する臨床反応を予測する。ゆえに不均一な癌サンプル中の変異の存在を本開示の組成物および方法を使用して特定およ

50

び定量することは有益である。本開示の組成物および方法は特に、細胞型および／または変異が稀である場合に情報として有益であり、有用である。

【0202】

一部の実施形態では、癌細胞中の少なくとも1つの遺伝子変化は、一塩基変異多型 (SNP)、挿入、欠失、逆位または染色体再構成を含む。

【0203】

一部の実施形態では、少なくとも1つの遺伝子変化は、促進性癌遺伝子の活性化または腫瘍抑制遺伝子の不活化と関連している。促進性癌遺伝子の例としては限定されないが、ヒト上皮成長因子受容体2 (HER2)、RAS、MYC促進性癌遺伝子 (MYC)、SRC促進性癌遺伝子 (SRC)、テロメラーゼ、BCL-2、アポトーシス調節因子 (BCL-2) および表皮成長因子受容体 (EGFR) が挙げられる。腫瘍抑制遺伝子の例としては限定されないが、腫瘍タンパク質p53 (p53)、網膜芽細胞腫 (Rb)、大腸腺腫様ポリポーシス (APC)、乳癌感受性遺伝子1 (BRCA1)、および乳癌感受性遺伝子2 (BRCA2) が挙げられる。

【0204】

一部の実施形態では、標的配列は、癌中の少なくとも1つの遺伝子変化を伴う配列を含む。

【0205】

一部の実施形態では、正規化対照ポリヌクレオチドが精製される核酸サンプル中の核酸の配列は、シーケンシングサンプル中の標的配列と同じではない。一部の実施形態では、核酸サンプル中の核酸の配列は、シーケンスサンプル中の標的配列に対し、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、30%、40%、50%、60%、70%、80%、85%、90%、95%、96%、97%、98%、または99%以下の同一性を有する。

【0206】

濃度

一部の実施形態では、NCは、非宿主 (例えば、病原体) 特定の流れの一部としてシーケンスリードの正規化のためにDNAを抽出する間に宿主 (例えば、患者) サンプルに添加されるよう設計される。サンプルから取得される全シーケンスリードのごく一部となるように、正規化対照断片は集合的に設計される。正規化対照断片は集合的に、サンプルから取得されるシーケンスリードのごく一部 (<5%) となるものでなければならない。一部の

【0207】

ポリヌクレオチドのグループ間の濃度で可能性のある構成はすべて、本開示の範囲内にあるとみなされる。

【0208】

一部の実施形態では、各グループのポリヌクレオチドは、同一の濃度である。例えば、一部の実施形態では、3つのグループの各々は1ピコグラム (pg) / マイクロリットル (μL) の同じ濃度であり、正規化対照組成物中の総濃度は3pg/μLである。

【0209】

一部の実施形態では、各グループのポリヌクレオチドは、同一の濃度ではない。一部の実施形態では、ポリヌクレオチドのグループの各々は、異なる濃度で存在してもよい。例えば一部の実施形態では、3つのグループは1:2:4の濃度比率で存在し、第一のグループは1pg/μLの濃度、第二のグループは2pg/μLの濃度、そして第三のグループは、4pg/μLの濃度であり、正規化対照組成物中の総濃度は7pg/μLとなる。

【0210】

郡内のサブグループ間の濃度、および異なるグループのサブグループ間の濃度で可能性のある構成はすべて、本開示の範囲内にあるとみなされる。

【0211】

一部の実施形態では、当該少なくとも3つのグループは、グループ中の少なくとも3つのサブグループのポリヌクレオチドの濃度が等しいグループを少なくとも1つと、グループ

中のサブグループの濃度が等しくないグループを少なくとも1つ、含む。

【0212】

一部の実施形態では、当該少なくとも3つのグループの各々における、少なくとも3つのサブグループの各々のポリヌクレオチドの濃度は、当該グループの他のサブグループの濃度と等しくない。一部の実施形態では、グループ中の当該少なくとも3つのサブグループの各々の濃度は、直線状数列または等比数列で分布される。直線状数列は、数列中の次の数が、前の数と比較して都度同じ量で増加または減少する数列である。概して、直線状数列は、以下の式により表される：

$$u_n = d \times n + c$$

【0213】

式中、 d は、数列中の連続する項の間の第一の差異であり、 n は、数列中の項であり、 c は定数である。濃度が直線状数列であるグループに3つのサブグループが存在する非限定的な実施形態の例において、当該3つのサブグループの濃度は、例えば、 $1\text{ pg}/\mu\text{L}$ 、 $2\text{ pg}/\mu\text{L}$ 、および $3\text{ pg}/\mu\text{L}$ である。

【0214】

等比数列は、最初の項の後の各項が、公比と呼ばれる固定されたゼロではない数値で前の項を乗じることにより見出される数列である。概して、等比数列は、以下の式により表される：

$$u_n = u_1 r^{n-1}$$

式中、 r は、公比である。濃度が等比数列であるグループに3つのサブグループが存在する非限定的な実施形態の組み合わせの例において、当該3つのサブグループの濃度は、例えば、 $2\text{ pg}/\mu\text{L}$ 、 $4\text{ pg}/\mu\text{L}$ 、および $8\text{ pg}/\mu\text{L}$ である。各ポリヌクレオチドのグループ内のサブグループの相対的濃度は、本開示の正規化対照の有利な特性である。NGS後のポリヌクレオチドのサブグループおよび／またはグループの各々に戻しマッピングされるシーケンスリードの相対的存在量は、正規化対照組成物中のサブグループおよび／またはグループの相対的濃度を反映する。すなわちNGS前のサンプルを反映する。したがってグループまたはサブグループ内のリードの相対的存在量を計算することで、内部クオリティコントロールの比較尺度が提供され、シーケンスサンプル中の標的配列の初期濃度のより正確なモデル化が可能となる。

【0215】

一部の実施形態では、グループ中の1つのサブグループは、基準濃度であり、グループ中の非基準サブグループの濃度は、基準の倍数の整数値である。一部の実施形態では、グループ中の1つのサブグループは基準濃度であり、グループ中の各サブグループの濃度は、基準を除き、グループ中の別のサブグループの濃度の2倍である。一部の実施形態では、グループ中の少なくとも3つのサブグループは、1：2：4の濃度比率で存在する。

【0216】

一部の実施形態では、複数のポリヌクレオチドが、個々のNC配列の濃度においてワイドダイナミックレンジが可能となる方法で混合される。濃度は、 $10^1 \sim 10^7$ の濃度範囲に及んでもよく、一部の断片は中間レベルの濃度を有する。正規化対照は、QPCRを含む様々な方法を使用して認証される。

【0217】

一部の実施形態では、ポリヌクレオチドの集団は、正規化対照組成物中、約 $0.005 \sim$ 約 $5000\text{ pg}/\mu\text{L}$ 、約 $0.005 \sim$ 約 $1000\text{ pg}/\mu\text{L}$ 、約 $0.005 \sim$ 約 $700\text{ pg}/\mu\text{L}$ 、約 $0.005 \sim$ 約 $500\text{ pg}/\mu\text{L}$ 、約 $0.005 \sim$ 約 $100\text{ pg}/\mu\text{L}$ 、約 $0.005 \sim$ 約 $50\text{ pg}/\mu\text{L}$ 、約 $0.05 \sim$ 約 $5000\text{ pg}/\mu\text{L}$ 、約 $0.05 \sim$ 約 $1000\text{ pg}/\mu\text{L}$ 、約 $0.05 \sim$ 約 $700\text{ pg}/\mu\text{L}$ 、約 $0.05 \sim$ 約 $500\text{ pg}/\mu\text{L}$ 、約 $0.05 \sim$ 約 $100\text{ pg}/\mu\text{L}$ 、約 $0.05 \sim$ 約 $50\text{ pg}/\mu\text{L}$ 、約 $0.5 \sim$ 約 $1000\text{ pg}/\mu\text{L}$ 、約 $0.5 \sim$ 約 $700\text{ pg}/\mu\text{L}$ 、約 $0.5 \sim$ 約 $500\text{ pg}/\mu\text{L}$ 、約 $0.5 \sim$ 約 $100\text{ pg}/\mu\text{L}$ 、約 $0.05 \sim$ 約 $50\text{ pg}/\mu\text{L}$ 、約 $5 \sim$ 約 $1000\text{ pg}/\mu\text{L}$ 、約 $5 \sim$ 約 $700\text{ pg}/\mu\text{L}$ 、約 $5 \sim$ 約 $500\text{ pg}/\mu\text{L}$ 、約 $5 \sim$ 約 $100\text{ pg}/\mu\text{L}$ 、または約 $5 \sim$ 約 $50\text{ pg}/\mu\text{L}$ の濃度である。一部の実施形態では、ポリヌクレオチドの集団は、正規化対照組成物中、約 $0.005 \sim 500\text{ pg}/\mu\text{L}$ の濃度である。

イクロリットル (μL) の濃度である。一部の実施形態では、ポリヌクレオチドの集団は、正規化対照組成物中、約0.5～約50pg/ μL の濃度である。一部の実施形態では、ポリヌクレオチドの集団は、正規化対照組成物中、約10～約20pg/ μL の濃度である。一部の実施形態では、ポリヌクレオチドの集団は、正規化対照組成物中、約1～約2pg/ μL の濃度である。一部の実施形態では、ポリヌクレオチドの集団は、正規化対照組成物中、約0.1～約0.2pg/ μL の濃度である。一部の実施形態では、ポリヌクレオチドの集団は、正規化対照組成物中、末端の値を含み、1.4pg/ μL の濃度である。

【0218】

インデックス化

一部の実施形態では、正規化対照ポリヌクレオチドは、固有の分子識別子をコードする配列を含む。一部の実施形態では、固有の分子識別子をコードする配列は、第一のインデックス化配列または第二のインデックス化配列を含む。

【0219】

本開示の非限定的なインデックス化戦略の例は、第一のインデックスと第二のインデックスの2つのインデックスを使用する。一部の実施形態では、固有の分子識別子をコードする配列は、共通プライマー配列、インデックス配列、およびアダプター配列を含む。一部の実施形態では、共通プライマー配列は、NGS用のフローセルアタッチメント部位を含む。

【0220】

一部の実施形態では、インデックス配列は、固有の4～20bpの配列である。一部の実施形態では、インデックス配列は、固有の6bp、8bp、または10bpの配列である。一部の実施形態では、インデックス配列は、固有の6bpの配列である。インデックスの長さは、ライブラリの複雑さに依存する。適切なインデックス配列の長さは、当業者には明白である。一部の実施形態では、第一のプライマー配列、第一のインデックス配列、および第一のアダプター配列を含む配列は、ポリヌクレオチドの5'末端に付加され、第二のプライマー配列、第二のインデックス配列、および第二のアダプター配列を含む配列は、同ポリヌクレオチドの3'末端に付加される。当該ポリヌクレオチドは例えば、本開示の正規化対照組物のポリヌクレオチドである。一部の実施形態では、ポリヌクレオチドのすべてのグループが、当該グループ中の複数のポリヌクレオチドの5'末端と3'末端に固有の二重インデックスを含む。この実施形態では、当該二重インデックスは、当該グループ内の複数のポリヌクレオチドのすべてで同じであり、グループ間のポリヌクレオチドでは異なる。一部の実施形態では、ポリヌクレオチドのすべてのサブグループが、当該サブグループ中の複数のポリヌクレオチドの5'末端と3'末端に固有の二重インデックスを含む。この実施形態では、当該二重インデックスは、当該サブグループ内の複数のポリヌクレオチドのすべてで同じであるが、異なるサブグループのポリヌクレオチドでは異なる。一部の実施形態では、各グループ内の複数のポリヌクレオチド、および各サブグループ内の複数のポリヌクレオチドは、当該グループおよび／またはサブグループに特異的な固有二重インデックスを有する。

【0221】

本開示の非限定的な二重インデックス化システムの例は、Illumina i5とi7の二重インデックス化システムである。

【0222】

一部の実施形態では、正規化対照は、シーケンスサンプル中のDNAまたはRNAの標的配列の、シーケンス可能ライブラリ分子への転換を測定するよう設計される。一部の実施形態では、ライブラリ調製中の開始時材料と終了時材料の転換比率は、すべてのサンプルで信頼性をもって測定するには不十分で、困難である。NCポリヌクレオチドの末端を二重インデックス化することによって、ライブラリ調製中の任意の時点で1サンプル当たり、どの分子が転換され、どのように転換されたかを追跡することが可能となる。これにより、サンプルのタイプに依存しないライブラリ転換因子の計算が可能となる。また、サンプル調製にクオリティコントロールを組み込むことも可能となる。この技術を使用して、試薬

10

20

30

40

50

ロット、酵素反応の効率、およびNGSに影響を及ぼす他の多くの基準の比較が可能となる。これらの比較は、通常であれば、問題が生じているかを判定するためにはライブラリ調製プロセスの最後まで待たなければならないシーケンス反応の工程に沿って行うことができる。

タンパク質

【0223】

正規化対照ポリヌクレオチドは、デオキシリボ核酸（DNA）分子、リボ核酸（RNA）分子、またはDNA-RNAハイブリッド分子を含む。

【0224】

タンパク質との会合は、NGSリードカウントからの標的配列濃度を算定する際の変動性および誤差の重大な原因である。例えば、いくつかの実施態様では、標的配列はウイルス配列であり、このウイルス配列を含むポリヌクレオチドとキャプシドとの会合は、サンプルおよびライブラリ調製の間の標的配列の精製および／または断片化を防ぐ。したがって、シーケンシングサンプル内の標的配列の核酸／タンパク質相互作用を模倣する本開示の正規化対照組成物へのタンパク質の添加は、NGS間の標的配列の挙動をモデル化するためのNCの能力を増大させる。

【0225】

本開示の正規化対照のいくつかの実施態様では、正規化対照はさらに、少なくとも一つのタンパク質を含む。いくつかの実施態様では、少なくとも一つのタンパク質は、細菌、ウイルスまたは真核生物の寄生体から単離または誘導される。いくつかの実施態様では、少なくとも一つのタンパク質は、ウイルスキャプシドタンパク質または細菌細胞壁タンパク質である。例示的であるが非限定的なウイルスキャプシドタンパク質は、アデノウイルス、アデノ関連ウイルス（AAV）、レンチウイルスまたはレトロウイルスから単離または誘導化されたウイルスキャプシドタンパク質を含む。

【0226】

いくつかの実施態様では、ポリヌクレオチドの集団とタンパク質とは操作可能に連結されている。例示的な実施態様では、タンパク質はウイルスキャプシドタンパク質であり、ポリヌクレオチドはウイルスキャプシドタンパク質内にカプセル化されている。他の実施態様では、タンパク質とポリヌクレオチドは結合している。

付加的配列

【0227】

いくつかの実施態様では、正規化対照ポリヌクレオチドは、プロモーター配列を含む。いくつかの実施態様では、プロモーター配列は、T7プロモーター配列、SP6プロモーター配列またはT3プロモーター配列からなるグループから選択されたプロモーター配列を含む。いくつかの実施態様では、特にプロモーターがT7プロモーターである実施態様では、T7プロモーターは、5'-TAATACGACTCACTATAGGG-3'の配列（配列番号14）を含む。いくつかの実施態様では、T7プロモーターは、5'-TAATACGACTCACTATAGGG-3'の配列（配列番号15）を含む。いくつかの実施態様では、T7プロモーターは、5'-GCCCTCGAGCTAATACGACTCACTATAGAG-3'の配列（配列番号16）を含む。いくつかの実施態様では、SP6プロモーターは、5'-ATTTAGGTGACACTATAG-3'の配列（配列番号17）を含む。いくつかの実施態様では、SP6プロモーターは、5'-CATACGATTTAGGTGACACTATAG-3'の配列（配列番号18）を含む。いくつかの実施態様では、T3プロモーターは、5'-AATTAACCCCTCACTAAAG-3'の配列（配列番号19）を含む。

【0228】

いくつかの実施態様では、正規化対照ポリヌクレオチドは、付加的プライマー配列を含む。プライマー配列は、コンポーネント配列および／または単離された配列の5'および3'に位置することができる。プライマー配列は、コンポーネント配列のインシリコデザインの部分として付加するか、またはコンポーネント配列および／または単離された配列

10

20

30

40

50

に二次的に付加することができる。プライマー配列は、例えば、アダプターライゲーションまたはランダムプライミングおよびコンポーネント配列および／または単離された配列の伸長により、コンポーネント配列および／または単離された配列に付加することができる。

【0229】

混合組成物

いくつかの実施態様では、本開示の正規化対照は、合成ポリヌクレオチドと、一つまたは複数の生物から単離または誘導されたポリヌクレオチドとの混合物を含む。例えば、サンプル内の標的配列のセットは、合成ポリヌクレオチドの現在の製造方法の長さ（例えば10 kbの長さ）を越えるように伸長された長さの範囲をカバーしてよい。この例では、正規化対照を用いて標的配列の長さの範囲を完全にモデル化するために、正規化対照は、コンポーネントと単離されたポリヌクレオチドとの混合物を含む。

10

【0230】

いくつかの実施態様では、正規化対照内のポリヌクレオチドの少なくとも一つのグループは、本開示の方法を使用して生物から精製される。いくつかの実施態様では、生物は、ウイルス、細菌、真菌または真核生物の寄生体である。いくつかの実施態様では、生物は、混合宿主／非宿主配列サンプル内の非宿主と一つまたは複数の特徴を共有する。いくつかの実施態様では、ウイルスはバクテリオファージT4である。

【0231】

いくつかの実施態様では、バクテリオファージT4はまた、既知の濃度で、付加的にまたは代替的に、合成ポリヌクレオチドを基礎とするNC（オリゴベースのNC）に添加される。

20

【0232】

正規化対照へのバクテリオファージT4の添加は、正規化対照が、抽出変動を考慮に入れることを可能にする。バクテリオファージT4は、オリゴベースのNCよりも、特に長い配列にとって、製造が安価でかつ容易であるという付加的利点を有する。

正規化対照の製造方法

デノボ合成

【0233】

本開示は、ポリヌクレオチドの少なくとも三つのグループを含む正規化対照の製造方法を提供し、各グループ内のポリヌクレオチドは、同じ長さである。いくつかの実施態様では、ポリヌクレオチドの少なくとも三つのグループの各々はさらに、ポリヌクレオチドの少なくとも三つのサブグループを含む。

30

【0234】

いくつかの実施態様では、少なくとも三つのグループは、同じ配列のポリヌクレオチドを含む。いくつかの実施態様では、少なくとも三つのグループは、同じ配列ではないポリヌクレオチドを含む。

【0235】

本開示の方法のいくつかの実施態様では、方法は、各固有のグループおよび／またはサブグループのポリヌクレオチドのオリゴヌクレオチドとしての合成、各固有のグループおよび／またはサブグループの濃度の定量化、およびNC組成物の製造のための各固有のグループおよび／またはサブグループの予定量の混合を含む。

40

【0236】

いくつかの実施態様では、ポリヌクレオチドは、DNAポリヌクレオチドである。いくつかの実施態様では、ポリヌクレオチドは、約10 bp～約3000 bpである。いくつかの実施態様では、ポリヌクレオチドは、約10 bp～約1200 bpである。いくつかの実施態様では、ポリヌクレオチドは、10 bp～250 bpである。いくつかの実施態様では、ポリヌクレオチドは、約150 bp～約1500 bpである。いくつかの実施態様では、ポリヌクレオチドは、約400 bp～約1100 bpである。いくつかの実施態様では、ポリヌクレオチドは、約500 bp～約1500 bpである。

50

【0237】

いくつかの実施態様では、ポリヌクレオチドは、シーケンシングサンプル内の標的配列と共通する少なくとも一つの特性を有する。標的配列の全ての配列特性は、本開示の正規化対照の範囲内にあると認識される。いくつかの実施態様では、標的配列は、転移因子配列、ウイルス配列、細菌配列、真菌配列、真核生物の寄生体配列または一つもしくは複数のヒト遺伝子の配列を含む。いくつかの実施態様では、少なくとも一つの特性は、転移因子配列、ウイルス配列、細菌配列、真菌配列、真核生物の寄生体配列または一つもしくは複数のヒト遺伝子の配列の配列特性を含む。

【0238】

いくつかの実施態様では、正規化対照のポリヌクレオチドは、RNAを含む。いくつかの実施態様では、DNAポリヌクレオチドはさらに、プロモーターをコード化する配列を含み、この方法はさらに、複数のRNAの製造のための各DNAポリヌクレオチドのインビトロ転写、複数のRNAの各々の濃度の定量化、および正規化対照を製造するための複数のRNAの予定量での混合を含む。例示的プロモーターは、T3、SP6およびT7を含むが、これらに限定されるものではない。

【0239】

ポリヌクレオチドのグループおよびサブグループ間での濃度のあらゆる可能な手配は、本開示の範囲内にあると認識される。

【0240】

一つのグループ内でのサブグループ間での、および異なるグループのサブグループ間での濃度のあらゆる可能な手配は、本開示の範囲内にあると認識される。

【0241】

濃度は、体積当たりの質量（例えばマイクロリットル（ μL ）当たりのピコグラム（ pg ）または μL 当たりのナノグラム（ ng ）として、あるいは体積当たりのポリヌクレオチドの分子の数（モル濃度）として表すことができる。濃度の測定をどのように変換するかは、当業者には容易に理解される。

【0242】

いくつかの実施態様では、ポリヌクレオチドの各々は、DNAオリゴヌクレオチドとして合成され、かつ任意に、インビトロ転写して複数のRNAポリヌクレオチドを製造する。各グループまたはサブグループのポリヌクレオチドの複数の各々は、定量化され、この複数のものは本開示の濃度比で混合される。最終的な正規化対照組成物は、次いで、マイクロリットル（ μL ）当たり約0.001～約5000ピコグラム（ pg ）、約0.005～約1000 $\text{pg}/\mu\text{L}$ 、約0.005～約700 $\text{pg}/\mu\text{L}$ 、約0.005～約50 $\text{pg}/\mu\text{L}$ 、約0.005～約100 $\text{pg}/\mu\text{L}$ 、約0.005～約50 $\text{pg}/\mu\text{L}$ 、約0.05～5000 $\text{pg}/\mu\text{L}$ 、約0.05～約1000 $\text{pg}/\mu\text{L}$ 、約0.05～約700 $\text{pg}/\mu\text{L}$ 、約0.05～約500 $\text{pg}/\mu\text{L}$ 、約0.05～約00 $\text{pg}/\mu\text{L}$ 、約0.05～約50 $\text{pg}/\mu\text{L}$ 、約0.5～約1000 $\text{pg}/\mu\text{L}$ 、約0.5～約700 $\text{pg}/\mu\text{L}$ 、約0.5～約500 $\text{pg}/\mu\text{L}$ 、約0.5～約100 $\text{pg}/\mu\text{L}$ 、約0.05～約50 $\text{pg}/\mu\text{L}$ 、約5～約1000 $\text{pg}/\mu\text{L}$ 、約5～約700 $\text{pg}/\mu\text{L}$ 、約5～約500 $\text{pg}/\mu\text{L}$ 、約5～約100 $\text{pg}/\mu\text{L}$ 、または約5～約50 $\text{pg}/\mu\text{L}$ の濃度に希釈される。いくつかの実施態様では、正規化対照は、約0.005～約500 $\text{pg}/\mu\text{L}$ の濃度である。いくつかの実施態様では、正規化対照は、約0.5～約50 $\text{pg}/\mu\text{L}$ の濃度である。いくつかの実施態様では、正規化対照は、約10～約20 $\text{pg}/\mu\text{L}$ の濃度である。いくつかの実施態様では、正規化対照は、約1～約2 $\text{pg}/\mu\text{L}$ の濃度である。いくつかの実施態様では、正規化対照は、約0.1～約0.2 $\text{pg}/\mu\text{L}$ の濃度である。いくつかの実施態様では、正規化対照は、約1.4 $\text{pg}/\mu\text{L}$ の濃度である。

【0243】

本発明の開示は、（a）少なくとも一つの参照配列から複数の参照配列フラグメントを生成すること；（b）参照配列フラグメントの数の関数として少なくとも一つのパラメーターの分布を生成すること；（c）分布をビンに分割すること；（d）ビンの少なくとも

10

20

30

40

50

サブセットから少なくとも一つの参照配列フラグメントを選択すること；（e）参照配列をシャッフルして、シャッフル配列を生成すること；および（f）シャッフル配列を含むオリゴを合成すること；それにより複数の正規化対照オリゴを生成することを含む、複数の正規化対照オリゴの製造方法を提供する。

【0244】

本開示は、（a）スライディングウィンドウを使用して少なくとも一つの参照配列から複数の参照配列フラグメントを生成すること；（b）参照配列フラグメントの数の関数として少なくとも一つのパラメーターの分布を生成すること；（c）分布をビンに分割すること；（d）ビンの少なくともサブセットから少なくとも二つの参照配列フラグメントを選択すること、ここで、この少なくとも二つの参照配列フラグメントは、参照配列内で非連続であるか、または異なる参照配列に由来する；（e）少なくとも3つのビンの各々から少なくとも二つの参照配列フラグメントを連結すること；および（f）連結された参照配列フラグメントを含むオリゴを合成すること；それにより複数の正規化対照オリゴを生成することを含む、複数の正規化対照オリゴの製造を提供する。いくつかの実施態様では、スライディングウィンドウは、1塩基対（bp）、2bp、3bp、4bpまたは5bpのスライディングウィンドウを含む。いくつかの実施態様では、参照配列フラグメントは、対応する参照配列の約15～60、約20～40、約20～30、約15～32、20～32または約25～35の連続ヌクレオチドである。いくつかの実施態様では、参照配列フラグメントは、対応する参照配列の29、30、31、32、33または34の塩基対である。いくつかの実施態様では、参照配列フラグメントは、シーケンシングリードの平均長さの半分未満である。

【0245】

いくつかの実施態様では、パラメーターは、（1）パーセントGC含有率、（2）エントロピー、（3）複雑性、（4）EIP、またはそれらの組合せを含む。いくつかの実施態様では、パラメーターは、パーセントGC含有率を含む。

【0246】

分布、例えばGC含有率の分布は、正規化対照の所望の解像度に応じて、任意の数の適切なビンに分割することができる。例えば、各ビンは、分布の1パーセント、2パーセント、5パーセント、10パーセント、15パーセント、または20パーセントを表すことができる。参照配列フラグメントは、各ビンから、またはビンのサブセットから選択することができる。例えば、10のビンを含むGC含有率分布では、参照配列フラグメントは、パーセントGC含有率分布の10、30、50、70、80および90パーセンタイルビンから選択される。

【0247】

いくつかの実施態様では、ステップ（d）は、各ビンから少なくとも2つの参照配列フラグメントを選択することを含む。例えば、少なくとも2、3、4、5、6、7、8、9、または10の参照配列フラグメントが各ビンから選択される。いくつかの実施態様では、ステップ（e）はさらに、各ビンからのシャッフル参照配列を連結することを含む。いくつかの実施態様では、所定のビンからの配列だけが連結されて、対応するビン内の参照配列フラグメントの特性を模倣し、したがって分布に沿った複数の点をモデル化する。

【0248】

いくつかの実施態様では、少なくとも一つの参照配列は、少なくとも2、少なくとも10、少なくとも20、少なくとも50、少なくとも100、少なくとも200、少なくとも300、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1,000、少なくとも1,200、少なくとも1,300、少なくとも1,400、少なくとも1,500、少なくとも1,600、少なくとも1,700、少なくとも1,800、少なくとも1,900、少なくとも2,000、少なくとも2,200、少なくとも2,400、少なくとも2,600、少なくとも2,800、少なくとも3,000、少なくとも4,000、少なくとも5,000、少なくとも6,000、少なくとも7,000、少なくとも8,000、少

なくとも9,000、少なくとも10,000の参照配列を含む。いくつかの実施態様では、少なくとも一つの参照配列は、約2,000の参照配列を含む。いくつかの実施態様では、少なくとも一つの参照配列は、ゲノム配列もしくはcDNA配列、またはそれらの組合せを含む。例えば、第一の生物に対応する参照配列は、ゲノム配列であってよく、第二の生物に対応する参照配列は、cDNA配列であってよい。あるいは、多数の生物（全ての生物）のための参照配列は、ゲノム配列またはcDNA配列であってよい。

【0249】

いくつかの実施態様では、参照配列フラグメントは、参照配列の約10~100、約10~80、約10~70、約10~60、約10~50、約10~40、約10~30、約10~20、約20~100、約20~80、約20~60、約20~50、約20~40、約20~30、約10~35、約20~35、約20~25の連続ヌクレオチドである。いくつかの実施態様では、参照配列フラグメントは、参照配列の約15~60、約20~40、約20~30、約15~32、20~32または約25~35の連続ヌクレオチドである。いくつかの実施態様では、参照配列フラグメントは、参照配列の20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64または65の連続塩基対を含む。

【0250】

いくつかの実施態様では、各正規化対照オリゴは、参照配列分布内の対応するピンからの一つまたは複数の参照配列フラグメントと、（1）同様のパーセントGC含有率、（2）同様のエントロピー、（3）同様のEIP、（4）同様の複雑性、または（5）同様の長さの少なくとも三つを有する。

多重合成

【0251】

いくつかの実施態様では、正規化対照は、合成DNAポリヌクレオチドの混合物を含む。

【0252】

本開示は、（a）DNA分子の少なくとも三つの集団を合成すること；ここで、少なくとも三つの集団の各々の中のDNA分子の各々は同じ配列を有し、DNA分子の各々の配列は、5'から3'に向かって、第一のコンポーネント配列、第一の制限酵素部位、第二のコンポーネント配列、第二の制限部位、および第三のコンポーネント配列を含む；（b）溶液内の集団DNA分子の各々の量を定量化すること；（c）DNA分子の少なくとも三つの集団のそれぞれを第一および第二の制限部位を切断する制限酵素を用いて消化して、消化されたDNA分子の少なくとも三つの集団を製造すること；および（d）消化された合成DNA分子の少なくとも三つの集団の各々の予定量を単一の組成物内で混合することを含む合成ポリヌクレオチドの集団を製造する方法を提供する。いくつかの実施態様では、少なくとも三つの合成DNA分子の各々は、少なくとも三つの合成DNA分子のいずれか別の第一、第二および第三のコンポーネント配列と同じ長さの第一、第二および第三のコンポーネント配列を含む。

【0253】

本開示の方法のいくつかの実施態様では、第一、第二および第三のコンポーネント配列はそれぞれ、約15bp~約3000bpである。いくつかの実施態様では、第一、第二および第三のコンポーネント配列はそれぞれ、約15bp~約1200bpである。いくつかの実施態様では、第一、第二および第三のコンポーネント配列はそれぞれ、約100bp~約600bpである。いくつかの実施態様では、第一、第二および第三のコンポーネント配列はそれぞれ、約150bp~約500bpである。いくつかの実施態様では、第一、第二および第三のコンポーネント配列は、それぞれ175bp、250bpおよび450bpである。

【0254】

いくつかの実施態様では、少なくとも三つのDNA分子のいずれの第一、第二および第

三のコンポーネント配列のいずれも同じ配列ではない。いくつかの実施態様では、第一、第二および第三のコンポーネント配列は、ランダム配列である。いくつかの実施態様では、ランダム配列は、ランダム配列ジェネレーターを用いて生成される。例示的ではあるが非限定的なランダム配列ジェネレーターは、www.bioinformatics.org/sms2/random_dna.htmlに見ることができ、参照により本明細書に組み込まれる。いくつかの実施態様では、ランダム配列は、NCBIヌクレオチドデータベース内の全ての生物のゲノムにもマッピングされない。いくつかの実施態様では、ランダム配列は、GC含有率のような特定の配列特性を含み、かつその他（例えば共通の制限部位）が除外されるように設計されていることができる。

【0255】

いくつかの実施態様では、少なくとも三つのDNA分子の各々の第一、第二および第三のコンポーネント配列は、シーケンシングサンプル内の標的配列と共通する少なくとも一つの特性を有する。

【0256】

いくつかの実施態様では、第一および第二の制限部位の配列は同じである。いくつかの実施態様では、制限部位は、II型制限エンドヌクレアーゼにより認識および切断される制限部位である。適切な制限部位の選択は、第一、第二および第三のコンポーネント配列に依存し、かつ当業者に容易に理解される。いくつかの例示的実施態様では、第一、第二および第三のコンポーネント配列のいずれにも存在しない制限部位が選択される。

【0257】

いくつかの実施態様では、第一および第二の制限部位の配列は、EcoRI部位、BamHI部位、HindIII部位、NotI部位、PvuII部位、SmaI部位、EcoRV部位、KpnI部位、PstI部位、SalI部位、ScaI部位、SpeI部位、SphI部位、StuI部位およびXbaI部位の配列からなるグループから選択される。いくつかの実施態様では、少なくとも三つの合成DNA分子の少なくとも三つの集団のそれぞれを消化するために使用する制限酵素は、EcoRI、BamHI、HindIII、NotI、PvuII、SmaI、EcoRV、KpnI、PstI、SalI、ScaI、SpeI、SphI、StuIまたはXbaIを含む。いくつかの実施態様では、第一および第二の制限部位の配列は、XhoI部位を含み、かつ制限酵素はXhoIを含む。

【0258】

いくつかの実施態様では、少なくとも三つのDNA分子の各々の配列はさらに、第一のプライマー配列と第二のプライマー配列に相補的な配列とを含み、第一のプライマー配列は、第一のランダム配列の5'に位置し、第二のプライマー配列に相補的な配列は、第三のランダム配列の3'に位置する。いくつかの実施態様では、第一のプライマー配列は、5'-TGAAGAACTGCGGCAGG-3'の配列（配列番号12）（MitOアレイプライマー）を含む。いくつかの実施態様では、第二のプライマー配列は、5'-AGCTGGAAGTGCAAGACC-3'の配列（配列番号20）（RibOアレイプライマー配列）を含み、第二のプライマー配列に相補的な配列は、5'-GGTCTGCACCTTCCAGCT-3'の配列（配列番号13）を含む。

【0259】

いくつかの実施態様では、本開示の方法は、溶液内の少なくとも三つのDNA分子の各々の量を定量化することを含む。いくつかの実施態様では、本開示の方法はさらに、溶液内の消化されたDNA分子の少なくとも三つの集団の各々の量を定量化することを含む。溶液内のDNA分子の集団の各々の量を定量化する方法および／または溶液内の消化されたDNA分子の少なくとも三つの集団の各々の量を定量化する方法は、当業者に容易に理解される。DNA定量化法の非限定的例は、リアルタイムPCR、ドロップレットデジタルPCR（ddPCR）および分光光度法を含む。

【0260】

いくつかの実施態様では、本開示の方法は、消化されたDNA分子の少なくとも三つの

10

20

30

40

50

集団の各々の予定量を単一組成物中で混合することを含む。

【0261】

いくつかの実施態様では、消化されたDNA分子の少なくとも三つの集団の予定量は同じである。いくつかの実施態様では、消化されたDNA分子の少なくとも三つの集団の予定量は同じではない。いくつかの実施態様では、消化されたDNA分子の少なくとも三つの集団の予定量は、線形配列であるか、または幾何学的配列内にある。

【0262】

いくつかの実施態様では、消化された合成DNA分子の一つの集団の予定量は、ベースラインであり、消化されたDNA分子の少なくとも二つの付加的集団の予定量は、ベースラインの整数倍である。いくつかの実施態様では、消化されたDNA分子の一つの集団の予定量は、ベースラインであり、消化されたDNA分子の各々の付加的集団の予定量は、ベースラインを除く消化されたDNA分子の別の集団の予定量の二倍である。

【0263】

いくつかの実施態様では、消化された合成DNA分子の少なくとも三つの集団の予定量は、1：2：4の比率である。

【0264】

例示的な実施態様では、消化されたDNA分子の三つの集団は、A：B：Cの比率で混合され、「A」、「B」および「C」は、濃度のあらゆる予定比率（例えば1：2：4、または1：3：5、またはあらゆる線形配列または幾何学的配列）を表す。この実施態様では、三つのDNA分子の各々の第一のコンポーネント配列は同じ長さであり、三つのDNA分子の各々の第二のコンポーネント配列は同じ長さであり、三つの合成DNA分子の各々の第三のコンポーネント配列は同じ長さである。得られるNC組成物は、それぞれが三つのサブグループからなる三つのグループからなる。第一のグループの各サブグループは、DNA分子の各々の第一のコンポーネント配列を含むポリヌクレオチドからなり、第二のグループの各サブグループは、三つのDNA分子の各々の第二のコンポーネント配列を含むポリヌクレオチドからなり、第三のグループの各サブグループは、DNA分子の各々の第三のコンポーネント配列を含むポリヌクレオチドからなる。第一のグループの三つのサブグループの三つの複数のポリヌクレオチドは、消化されたDNA分子の三つの集団がA：B：Cの比率で混合されたため、A：B：Cの比率の濃度で存在する。消化されたDNA分子の集団の各々が各グループに等しく寄与するので、各グループは、同じ濃度で組成物内に存在する。

インビトロ転写

【0265】

いくつかの実施態様では、正規化対照は、RNAポリヌクレオチドを含む。本開示は、このようなRNA正規化対照を製造する方法を提供する。

【0266】

本開示は、ポリメラーゼ（例えば、T3、SP6またはT7ポリメラーゼ）のプロモーター配列の3'である配列を転写するためのインビトロ転写方法を提供する。5'プロモーター配列を有するあらゆる配列は、本開示のインビトロ転写方法に適したテンプレートであると認識される。

【0267】

本開示は、少なくとも三つのDNAポリヌクレオチドを合成し、各ポリヌクレオチドはプロモーターの配列とコンポーネント配列とを含み、各ポリヌクレオチドをインビトロ転写して、複数のRNAを製造し、複数のRNAの各々の濃度を定量化し、複数のRNAを予定量で混合して、NC組成物を製造することを含む正規化対照の製造方法を提供する。

【0268】

本開示は、(a) DNA分子の少なくとも三つのグループを合成し；ここで、DNA分子の各々は、5'から3'に向かって、第一のコンポーネント配列、プロモーター配列、第一の制限酵素部位、第二のコンポーネント配列、第二のプロモーター配列、第二の制限部位、第三のコンポーネント配列、および第三のプロモーター配列を含む；(b) 第一お

よび第二の制限酵素部位を切断する制限酵素でDNA分子のグループの各々を消化して、消化された合成DNA分子の少なくとも三つの集団を製造し；(c) DNA分子のグループの各々をインビトロ転写して、RNA分子の少なくとも三つの集団を製造し；(d) RNA分子の集団の各々内に製造されたRNAの量を定量化し；および(e) RNA分子の集団の各々の予定量を混合して、単一の正規化組成物を製造することを含む合成核酸分子の集団の製造方法を提供する。

【0269】

いくつかの実施態様では、DNA分子の各々は、DNA分子のいずれか別の第一、第二および第三のコンポーネント配列と同じ長さの第一、第二および第三のコンポーネント配列を含む。

【0270】

いくつかの実施態様では、第一、第二および第三のコンポーネント配列はそれぞれ、約15bp～約1200bpである。いくつかの実施態様では、第一、第二および第三のコンポーネント配列はそれぞれ、約150bp～約500bpである。いくつかの実施態様では、第一、第二および第三のコンポーネント配列はそれぞれ、約100bp～約600bpである。

【0271】

いくつかの実施態様では、DNA分子のいずれの第一、第二および第三のコンポーネント配列のいずれも同じ配列ではない。いくつかの実施態様では、DNA分子の各々の第一、第二および第三のコンポーネント配列は、ランダム配列である。いくつかの実施態様では、DNA分子の各々の第一、第二および第三のコンポーネント配列は、サンプル内の標的配列と共通する少なくとも一つの特徴を有する。

【0272】

本開示は、少なくとも三つの合成DNA分子の各々を、第一および第二の制限酵素部位を切断するII型制限エンドヌクレアーゼのような制限酵素で消化して、消化されたDNA分子の少なくとも三つの集団を製造する方法を提供する。

【0273】

いくつかの実施態様では、第一および第二の制限部位の配列は同じである。いくつかの実施態様では、制限部位は、II型制限エンドヌクレアーゼにより認識および切断される制限部位である。適切な制限部位の選択は、第一、第二および第三のコンポーネント配列に依存し、かつ当業者に容易に理解される。いくつかの例示的実施態様では、第一、第二および第三のコンポーネント配列のいずれにも存在しない制限部位が選択される。

【0274】

いくつかの実施態様では、第一および第二の制限部位の配列は、EcoRI部位、BamHI部位、HindIII部位、NotI部位、PvuII部位、SmaI部位、EcoRV部位、KpnI部位、PstI部位、SalI部位、ScaI部位、SpeI部位、SphI部位、StuI部位およびXbaI部位の配列からなるグループから選択される。いくつかの実施態様では、少なくとも三つの合成DNA分子の少なくとも三つの集団の各々を消化するために使用する制限酵素は、EcoRI、BamHI、HindIII、NotI、PvuII、SmaI、EcoRV、KpnI、PstI、SalI、ScaI、SpeI、SphI、StuIまたはXbaIを含む。いくつかの実施態様では、第一および第二の制限部位の配列は、XhoI部位を含み、かつ制限酵素はXhoIを含む。

【0275】

いくつかの実施態様では、消化されたDNA分子の得られる集団は、インビトロ転写のための5'プロモーターを有するDNA分子、およびコンポーネント配列を含む。例えば、少なくとも三つの合成DNA分子の各集団は、消化の際に、三つのDNAフラグメントを製造し、各々のフラグメントは、第一、第二または第三のコンポーネント配列のいずれかに操作可能に連結された5'プロモーターを含む。これらのDNAフラグメントは、次いで、本開示のRNAポリヌクレオチドを製造するために、インビトロ転写のためのテン

10

20

30

40

50

プレートとして使用することができる。いくつかの実施態様では、これらのフラグメントは、インビトロ転写の前に別々に精製および単離される。いくつかの実施態様では、フラグメントのプールは、インビトロ転写のためのテンプレートとして使用される。

【0276】

本開示は、DNA分子の集団の各々をインビトロ転写して、RNA分子の少なくとも三つの集団を製造する方法を提供する。いくつかの実施態様では、第一のプロモーター配列、第二のプロモーター配列および第三のプロモーター配列は、同じプロモーター配列を含む。いくつかの実施態様では、プロモーター配列は、T7プロモーター配列、SP6プロモーター配列またはT3プロモーター配列からなるグループから選択されるプロモーター配列を含む。いくつかの実施態様では、プロモーターがT7プロモーターである特別な実施態様では、T7プロモーターは、5'-TAATACGACTCACTATAGG-3'の配列（配列番号14）を含む。いくつかの実施態様では、T7プロモーターは、5'-TAATACGACTCACTATAGGG-3'の配列（配列番号15）を含む。いくつかの実施態様では、T7プロモーターは、5'-GCCCTCGAGCTAATACGACTCACTATAGAG-3'の配列（配列番号16）を含む。いくつかの実施態様では、SP6プロモーターは、5'-ATTTAGGTGACACTATAG-3'の配列（配列番号17）を含む。いくつかの実施態様では、SP6プロモーターは、5'-CATACGATTTAGGTGACACTATAG-3'の配列（配列番号18）を含む。いくつかの実施態様では、T3プロモーターは、5'-AATTAACCCCTCACTAAAG-3'の配列（配列番号19）を含む。

【0277】

本開示のポリメラーゼは、RNAポリメラーゼIIまたはRNAポリメラーゼIIIポリメラーゼであることができる。いくつかの実施態様では、ポリメラーゼは、T7ポリメラーゼ、SP6ポリメラーゼまたはT3ポリメラーゼである。本開示のRNAポリメラーゼは、野生型ポリメラーゼ、コンポーネントポリメラーゼ、または（インビトロ転写のために）最適化または操作されたポリメラーゼであってよい。本開示のポリメラーゼの活性は、所定のプロモーター配列に対して高度に特異的であってよい（例えば、T7プロモーター用のT7ポリメラーゼ、SP6プロモーター用のSP6ポリメラーゼ、またはT3プロモーター用のT3ポリメラーゼ）。

【0278】

T7プロモーターは、T7バクテリオファージRNAポリメラーゼによって認識されかつ転写がサポートされる。本開示のT7ポリメラーゼは、野生型T7ポリメラーゼ、コンポーネントT7ポリメラーゼ、または（例えば、インビトロ転写のために）最適化または操作されたT7ポリメラーゼであってよい。T7ポリメラーゼは、5'から3'の方向にDNAテンプレートからRNAの形成を触媒するDNA依存性RNAポリメラーゼである。DNAテンプレートは、二本鎖または一本鎖であってよい。T7ポリメラーゼは、T7プロモーターに対して高い特異性を示し、インビトロでロバストな転写を製造することができ、修飾ヌクレオチド（例えば標識ヌクレオチド）の新生RNA転写物内への組み込みを可能にする。T7ポリメラーゼのこれらの特徴は、本開示のRNAを合成するための優れたポリメラーゼである。

【0279】

本開示は、合成RNA分子の少なくとも三つの集団の各々において製造されるRNAの量を定量化する方法を提供する。RNAを定量化する方法は、当業者に容易に理解される。RNAを定量化する非限定的な例示的方法是、分光光度法または臭化エチジウムのような色素による蛍光標識付けを含む。

【0280】

本開示は、RNA分子の集団の各々の予定量を混合して単一組成物を製造する方法を提供する。

【0281】

いくつかの実施態様では、RNA分子の集団の各々の予定量は同じである。いくつかの

実施態様では、RNA分子の集団の各々の予定量は同じではない。いくつかの実施態様では、RNA分子の集団の予定量は、線形配列内にあるか、または幾何学的配列内にある。

【0282】

いくつかの実施態様では、RNA分子の一つの集団の予定量は、ベースラインであり、RNA分子の付加的集団の予定量は、ベースラインの整数倍である。いくつかの実施態様では、RNA分子の一つの集団の予定量は、ベースラインであり、RNA分子の各付加的集団の予定量は、ベースラインを除くRNA分子の別の集団の濃度の二倍である。いくつかの実施態様では、RNA分子の少なくとも三つの集団の予定量は、1：2：4の比率である。

【0283】

いくつかの実施態様では、この方法はさらに、正規化対照を約0.001～約500 pg/μLの濃度に希釈することを含む。いくつかの実施態様では、この方法はさらに、正規化対照を約0.5～約500 pg/μLの濃度に希釈することを含む。いくつかの実施態様では、この方法はさらに、正規化対照を約0.5～約50 pg/μLの濃度に希釈することを含む。いくつかの実施態様では、この方法はさらに、正規化対照を約10～約20 pg/μLの濃度に希釈することを含む。いくつかの実施態様では、この方法はさらに、正規化対照を約1.0～約2.0 pg/μLの濃度に希釈することを含む。いくつかの実施態様では、この方法はさらに、正規化対照を約0.1～約0.2 pg/μLの濃度に希釈することを含む。いくつかの実施態様では、正規化対照の濃度は、1.4 pg/μLである。

単離された配列

【0284】

本開示は、正規化対照が単離された配列を含む正規化対照組成物を製造する方法を提供する。いくつかの実施態様では、単離された配列は、配列のシャッフルまたは付加的な単離された配列もしくはコンポーネント配列との連結のような付加的ステップなしで、天然に存在する配列から単離される。

【0285】

したがって、本開示は、(a) DNAをサンプルから抽出すること；(b) DNAを制限酵素で消化して、DNAフラグメントのコレクションを製造すること；(c) DNAフラグメントのコレクションを分離すること；(d) 少なくとも三つのDNAフラグメントをDNAフラグメントのコレクションから精製して、ポリヌクレオチドの少なくとも三つのグループを製造すること、ここでポリヌクレオチドの少なくとも三つのグループの各々は、複数のポリヌクレオチドを含む；および(e) ポリヌクレオチドの少なくとも三つのグループの各々の予定量を混合して組成物を製造することを含む正規化対照の製造方法を提供する。

【0286】

本開示は、DNAをサンプルから抽出する方法を提供する。いくつかの実施態様では、サンプルは、ウイルス、細菌、真菌または真核生物の寄生体から単離または誘導される。いくつかの実施態様では、サンプルはクローニングベクターを含む。いくつかの実施態様では、クローニングベクターは、細菌人工染色体(BAC)、酵母人工染色体(YAC)、コスミド、フォスミドまたはプラスミドである。いくつかの実施態様では、サンプルはプラスミドである。いくつかの実施態様では、正規化対照配列は、目的の種、例えばウイルス、細菌または真核生物の寄生体からクローニングベクター内にクローニングされ、DNAは、クローニングベクターを産生する細胞の培養から抽出される。いくつかの実施態様では、細胞は細菌(例えばエシェリキア コリ)である。DNA抽出の例示的な方法は、キット(例えば、Qiagen MiniPrep)またはフェノールに基づく抽出方法を含む。DNA抽出の適切な方法は、当業者に容易に理解される。

【0287】

本開示は、DNAを制限酵素で消化して、DNAフラグメントのコレクションを製造する方法を提供する。いくつかの実施態様では、制限酵素は、EcoRI, BamHI、H

i n d I I I、N o t I、P v u I I、S m a I、E c o R V、K p n I、P s t I、S a l I、S c a I、S p e I、S p h I、S t u IおよびX b a Iからなるグループから選択される。いくつかの実施態様では、制限酵素は、E c o R I、B a m H I、H i n d I I I、P v u I I、S m a I、E c o R V、K p n I、P s t I、S a l I、S c a I、S p e I、S p h I、S t u I、X b a I、N o t I、A s c I、F s e I、P a c I、P m e I、B g l I I、B s t B I、H i n c I I IおよびS g f Iからなるグループから選択される。

【0288】

いくつかの実施態様では、制限酵素でDNAを消化することにより製造されたフラグメントは、約15bp～約100kbである。いくつかの実施態様では、フラグメントは、約15bp～約50kbである。いくつかの実施態様では、フラグメントは、約50bp～約50kbである。いくつかの実施態様では、フラグメントは、約500bp～約50kbである。いくつかの実施態様では、フラグメントは、約1000bp～約50kbである。いくつかの実施態様では、ポリヌクレオチドの少なくとも三つの異なるグループの各々の最長のものは、少なくとも15kbの長さ、少なくとも20kbの長さ、少なくとも25kbの長さ、30kbの長さ、少なくとも35kbの長さまたは少なくとも40kbの長さである。

【0289】

本開示は、DNAフラグメントのコレクションから少なくとも三つのDNAフラグメントを精製して、ポリヌクレオチドの少なくとも三つのグループを製造する方法を提供し、ここで、ポリヌクレオチドの少なくとも三つのグループの各々は、複数のポリヌクレオチドを含む。いくつかの実施態様では、DNAフラグメントのコレクションを分離することは、フラグメントサイズ、フラグメント電荷、またはその組合せに基づく分離を含む。いくつかの実施態様では、DNAフラグメントのコレクションを分離することは、フラグメントサイズに基づく分離を含む。いくつかの実施態様では、サイズに基づくDNAフラグメントの分離は、ゲル電気泳動、クロマトグラフィーまたはタンジェンシャルフローフィルトレーション(TFF)を含む。

【0290】

いくつかの実施態様では、サイズに基づくDNAフラグメントの分離は、ゲル電気泳動を含む。ゲル電気泳動は、そのサイズに従ってポリヌクレオチドを分離する。ポリヌクレオチドサンプルは、ゲルの一端のウェル内に置かれ、電流を供給してそれがゲルを通過するように引き込む。ポリヌクレオチドは負に帯電しているため、正極に向かって移動する。ポリヌクレオチドのサイズは、それがゲルの細孔を通過して移行する速度を決定する。多様なサイズのポリヌクレオチドを、ゲルのアガロース濃度(例えば、0.8%～3%のアガロース)の変化により分離することができる。あるいは、より小さなポリヌクレオチドを分離するためにアクリルアミドゲルを使用することもできる。いくつかの実施態様、特に、大きなポリヌクレオチドを分離する実施態様では、電磁場の角度、強度および持続時間を変更することで、分離を支援することができる。いくつかの実施態様では、ゲル電気泳動は、パルスフィールドゲル電気泳動である。

【0291】

いくつかの実施態様では、サイズに基づくDNAフラグメントの分離は、クロマトグラフィーを含む。例示的なクロマトグラフィー法は、アニオン交換クロマトグラフィーを含む。アニオン交換クロマトグラフィーは、負に帯電した分子に結合する正に帯電した基を含むイオン交換樹脂を使用して、その電荷に基づいてポリヌクレオチドを分離する。ポリヌクレオチドがカラムまたはマトリックスに結合する強さは、ポリヌクレオチドの負電荷の全体的な強度に、したがってサイズに基づく。

【0292】

いくつかの実施態様では、サイズに基づくDNAフラグメントの分離は、TFFを含む。TFFは、膜に基づく濾過プロセスである。膜多孔性にに基づき、これは、精密濾過または限外濾過プロセスとして分類することができる。限外濾過膜は、一般的に、0.001

10

20

30

40

50

〜0.1 μmの細孔サイズを有する。クロスフローフィルトレーションとしても公知のタンジェンシャルフローフィルトレーション(TFF)は、膜面に平行に供給流を適用する。この流の一部は、膜を通過する(透過液)が、残り(濃縮液)は、再循環させて供給リザーバーに戻される。ポリヌクレオチドは、膜の細孔を通過することにより分離される。

【0293】

いくつかの実施態様では、この方法はさらに、ポリヌクレオチドの少なくとも三つの異なるグループの各々の少なくとも末端にアダプターをライゲーションすることを含む。

【0294】

いくつかの実施態様では、この方法はさらに、ポリヌクレオチドの少なくとも三つの異なるグループの各末端にアダプターをライゲーションすることを含む。

【0295】

いくつかの実施態様では、正規化対照内のポリヌクレオチドの少なくとも三つの異なるグループの配列はさらに、合成ポリヌクレオチド配列を含む。例示的であるが非限定的な合成配列は、アダプターの配列、PCR増幅用のプライマー、インビトロ転写のためのプロモーター配列、およびNGSの間のフローセルアタッチメント用の結合部位を含む。

【0296】

いくつかの実施態様では、アダプター配列は、固有の分子識別子をコードする配列を含む。いくつかの実施態様では、固有の分子識別子をコードする配列は、第一のインデキシング配列または第二のインデキシング配列を含む。本開示の例示的で非限定的なインデキシング戦略は、二つのインデックス、第一のインデックスと第二のインデックスを使用する。いくつかの実施態様では、固有の分子識別子をコードする配列は、共通のプライマー配列、インデックス配列およびアダプター配列を含む。いくつかの実施態様では、共通のプライマー配列は、NGS用のフローセルアタッチメントを含む。

【0297】

いくつかの実施態様では、インデックス配列は、固有の4〜20bp配列である。いくつかの実施態様では、インデックス配列は、固有の6bp、8bpまたは10bp配列である。いくつかの実施態様では、インデックス配列は、固有の6bp配列である。インデックスの長さは、ライブラリーの複雑性に依存する。適切なインデックス配列長さは、当業者に容易に理解される。いくつかの実施態様では、第一のプライマー配列、第一のインデックス配列および第一のアダプター配列を含む配列は、ポリヌクレオチドの5'末端に取り付けられ、第二のプライマー配列、第二のインデックス配列および第二のアダプター配列を含む配列は、同じポリヌクレオチドの3'末端に取り付けられる。ポリヌクレオチドは、例えば本開示の正規化対照組成物のポリヌクレオチドである。いくつかの実施態様では、ポリヌクレオチドのあらゆるグループは、グループ内の複数のポリヌクレオチドの5'および3'末端に固有のデュアルインデックスを含む。この実施態様では、デュアルインデックスは、グループ内の複数のポリヌクレオチドの全てについて同じであり、グループ間のポリヌクレオチドについては異なる。

【0298】

いくつかの実施態様では、合成配列は、プロモーター配列を含む。いくつかの実施態様では、プロモーター配列は、T7プロモーター配列、SP6プロモーター配列またはT3プロモーター配列からなるグループから選択されるプロモーター配列を含む。

単離とデノボ合成との連結

【0299】

本開示は、本開示の方法を使用して、正規化対照組成物の製造方法を提供し、ここで、正規化対照は、合成された配列を含むポリヌクレオチドと、サンプルから単離された配列を含むポリヌクレオチドの両方を含む。

【0300】

本開示は、(a) DNA分子の少なくとも三つのグループを合成すること、ここでグループ内のDNA分子は同じ配列を有し、DNA分子の各々の配列は、5'から3'に向かって、第一のコンポーネント配列、第一の制限酵素部位、第二のコンポーネント配列、第

10

20

30

40

50

二の制限部位、および第三のコンポーネント配列を含む；（b）溶液内のDNA分子のグループの各々の量を定量化すること；（c）第一および第二の制限部位を切断する制限酵素でDNA分子のグループの各々を消化して、消化されたDNA分子の少なくとも三つのグループを製造すること；および（d）消化されたDNA分子のグループの各々の予定量を混合し、それにより単一の正規化対照組成物を生成することを含む正規化対照の製造方法を提供する。

【0301】

いくつかの実施態様では、（d）の混合ステップはさらに、DNA分子の少なくとも一つの付加的グループの予定量を、ステップ（a）～（c）の消化されたDNA分子の少なくとも三つのグループと混合して、単一の正規化対照内にDNA分子の少なくとも四つのグループを製造することを含み、ここで、DNA分子の少なくとも一つの付加的グループは、サンプルから単離または誘導される。特定の実施態様では、DNA分子の少なくとも一つの付加的グループは、本開示の方法を使用してサンプルから単離または誘導される。

10

【0302】

いくつかの実施態様では、DNA分子の少なくとも四つのグループの予定量は等しい。

【0303】

いくつかの実施態様では、DNA分子の少なくとも四つのグループの予定量は等しくない。特定の実施態様では、DNA分子の少なくとも四つのグループの予定量は、線形配列または幾何学的配列内にある。

【0304】

本開示は、（a）DNA分子の少なくとも一つのグループを合成すること、ここでグループ内のDNA分子は同じ配列を有し、DNA分子の各々の配列は、5'から3'に向かって、第一のコンポーネント配列、第一の制限酵素部位、第二のコンポーネント配列、第二の制限部位、および第三のコンポーネント配列を含む；（b）溶液内のDNA分子のグループの量を定量化すること；（c）第一および第二の制限部位を切断する制限酵素でDNA分子の少なくとも一つのグループを消化して、消化されたDNA分子の少なくとも一つのグループを製造すること；および（d）消化されたDNA分子のグループの予定量をDNA分子の少なくとも二つの付加的グループの予定量と混合して、単一の正規化対照組成物内でDNA分子の少なくとも三つのグループを製造し、ここで、DNA分子の少なくとも二つの付加的グループは、サンプルから単離または誘導されることを含む正規化対照の製造方法を提供する。

20

【0305】

いくつかの実施態様では、DNA分子の少なくとも三つのグループの予定量は等しい。

【0306】

いくつかの実施態様では、DNA分子の少なくとも三つのグループの予定量は等しくない。いくつかの実施態様では、DNA分子の少なくとも三つのグループの予定量は、線形配列または幾何学的配列内にある。

30

【0307】

いくつかの実施態様では、DNA分子の一つのグループの予定量は、ベースラインであり、DNA分子の付加的グループの予定量は、ベースラインの整数倍である。いくつかの実施態様では、消化されたDNA分子の一つのグループの予定量は、ベースラインであり、消化されたDNA分子の各々の付加的グループの予定量は、ベースラインを除く消化されたDNA分子の別のグループの予定量の二倍である。いくつかの実施態様では、消化されたDNA分子の少なくとも四つのグループの予定量は、1：2：4、任意に1：2：4：8、または1：2：4：8：16等の比率である。

40

【0308】

いくつかの実施態様では、サンプルは、プラスミドDNA、ミトコンドリアDNA、葉緑体DNAまたはゲノムDNAを含む。特定の実施態様では、サンプルは、クローニングベクターから単離されたDNAを含む。いくつかの実施態様では、クローニングベクターは、細菌成分染色体（BAC）、酵母人工染色体（YAC）、コスミド、フォスミドまた

50

はプラスミドである。

サンプル調製

【0309】

本開示は、(a) 本開示の正規化対照組成物の既知量とサンプルとを混合すること、(b) ハイスループットシーケンシングライブラリーを調製すること、(c) リードのコレクションを製造するために前記ライブラリーをシーケンシングすること、(d) サンプルまたは組成物に各リードをマッピングすること、(e) 組成物内の複数の核酸分子の各々によって製造されたリードの数を決定すること、(f) (a) においてサンプルと混合した組成物内の複数の核酸分子の各々の開始濃度と、(c) において製造されたリードの数との間の関係を計算すること、(g) リードとサンプル内の濃度との間の関係をモデリング(決定)すること、およびh. (g) におけるモデルを使用して、標的核酸分子により製造されたリードの数から、サンプル内の少なくとも一つの標的核酸分子の初期濃度を計算することを含む、サンプル内の少なくとも一つの標的核酸分子の発現のレベルを定量化する方法を提供する。

10

【0310】

いくつかの実施態様では、サンプルは核酸を含む。核酸は、DNA、RNAまたはDNAとRNAとの混合物であることができる。核酸は、一本鎖(例えば、一本鎖RNA、一本鎖DNA)、二本鎖(例えば、二本鎖DNA、DNA:RNAハイブリッド)、またはこれらの混合物であることができる。

【0311】

いくつかの実施態様では、サンプルは、哺乳動物から採取された全血、血漿、血清、涙、唾液、粘液、脳脊髄液、歯、骨、指爪、便、尿、組織またはバイオブシーを含む。いくつかの実施態様では、サンプルは、ヒトから採取された全血、血漿、血清、涙、唾液、粘液、脳脊髄液、歯、骨、指爪、便、尿、組織またはバイオブシーを含む。いくつかの実施態様では、サンプルは、植物からの葉、花またはその一部、茎、根、堅果または種子を含む。いくつかの実施態様では、サンプルは、昆虫からの体の部分、組織、便または血リンパを含む。

20

【0312】

いくつかの実施態様では、この方法はさらに、サンプルから核酸を抽出することを含む。いくつかの実施態様では、核酸は、DNA、RNAまたはこれらの混合物を含む。サンプルから核酸を抽出する方法は、当業者に周知である。例示的な抽出方法は、フェノール/クロロホルムに基づく抽出方法、およびQiagen Miniprepキット(カタログ番号27104)およびQiagen RNeasyミニキット(カタログ番号74104)のようなキットを含むが、これらに限定されるものではない。

30

【0313】

いくつかの実施態様では、サンプルは哺乳動物サンプルである。いくつかの実施態様では、標的核酸分子は、ウイルス、細菌、真菌、または真核生物の寄生体(例えば、哺乳動物サンプル内に見られるウイルス、細菌、真菌、または真核生物の寄生体)の配列を含む。いくつかの実施態様では、標的核酸分子は、一つまたは複数の哺乳動物遺伝子を含む。

【0314】

いくつかの実施態様では、サンプルはヒトサンプルである。いくつかの実施態様では、標的核酸分子は、ウイルス、細菌、真菌、または真核生物の寄生体(例えば、ヒトサンプル内に見られるウイルス、細菌、真菌、または真核生物の寄生体)の配列を含む。いくつかの実施態様では、標的核酸分子は、一つまたは複数のヒト遺伝子を含む。

40

【0315】

本開示は、本開示の正規化対照の既知量をサンプルと混合する方法を提供する。いくつかの実施態様では、正規化対照組成物は、正規化対照が、サンプルから得られた全配列リードの小さな部分を表すようにサンプルと混合される。いくつかの実施態様では、正規化対照は、集合的に、NGSを使用してサンプルから得られた配列リードの小さな部分(<5%)を表すべきである。

50

【0316】

いくつかの実施態様では、正規化対照組成物内のポリヌクレオチドのグループまたはサブグループの各々は、正規化対照内のポリヌクレオチドの全体量に対して約0.001～約0.500:1の比率である。いくつかの実施態様では、正規化対照組成物内のポリヌクレオチドのグループまたはサブグループの各々は、正規化対照内のポリヌクレオチドの全体量に対して約0.001～約0.420:1の比率である。いくつかの実施態様では、正規化対照組成物内のポリヌクレオチドのグループまたはサブグループの各々は、約0.001～約0.450:1、約0.001～約0.400:1、約0.001～約0.400:1、約0.001～約0.350:1、約0.001～約0.300:1、約0.001～約0.250:1、約0.001～約0.200:1、約0.001～約0.150:1、約0.001～約0.100:1、約0.001～約0.05:1、約0.005～約0.500:1、約0.010～約0.500:1、約0.050～約0.500:1、約0.100～約0.500:1、約0.010～約0.400:1、約0.010～約0.300:1、約0.010～約0.200:1、約0.010～約0.100:1、約0.050～約0.400:1、約0.050～約0.300:1、約0.050～約0.200:1または約0.050～約0.100:1の比率である。

10

【0317】

いくつかの実施態様では、組成物内のポリヌクレオチドのグループまたはサブグループの各々は、約0.0001 pg/μL～約45 pg/μL、約0.001 pg/μL～約45 pg/μL、約0.001 pg/μL～約10 pg/μL、約0.005 pg/μL～約5 pg/μLまたは約0.009 pg/μL～約5 pg/μLの最終濃度でサンプルに添加される。いくつかの実施態様では、組成物内のポリヌクレオチドのグループまたはサブグループの各々は、約0.009 pg/μL～約5 pg/μLの最終濃度でサンプルに添加される。

20

【0318】

いくつかの実施態様では、組成物内のポリヌクレオチドのグループまたはサブグループの各々は、約0.001アトモル/μL～約50アトモル/μLの最終濃度でサンプルに添加される。いくつかの実施態様では、組成物内のポリヌクレオチドのグループまたはサブグループの各々は、約0.01アトモル/μL～50アトモル/μLの最終濃度でサンプルに添加される。いくつかの実施態様では、組成物内のポリヌクレオチドのグループまたはサブグループの各々は、約0.05アトモル/μL～25アトモル/μLの最終濃度でサンプルに添加される。

30

【0319】

いくつかの実施態様では、この方法はさらに、サンプルから核酸を抽出すること、およびサンプルから核酸を抽出する前に正規化対照をサンプルと混合することを含む。いくつかの実施態様では、核酸抽出前の正規化対照とサンプルとの混合は、正規化対照がサンプル調製の間の核酸抽出内のバイアスを測定することを可能にする。例えば、タンパク質と複合体形成した標的核酸は、核酸抽出に対して耐性であってよく、したがって、NGS後のリード存在度に不十分に表される。このプロセスは、適切な正規化対照、例えばウィルスキャプシドタンパク質によりカプセル化された正規化対照の添加によって模倣することができる。

40

【0320】

いくつかの実施態様では、正規化対照とサンプルとの混合は、ハイスループットシーケンシングライブラリーの調製の前に行われる。いくつかの実施態様、特にライブラリー調製の前に混合ステップを行う実施態様では、NCは、シーケンシングサンプル内のDNAまたはRNA標的配列からシーケンシング可能なライブラリー分子への変換を測定する。サンプルポリヌクレオチドと同時のNCポリヌクレオチドの末端のデュアルインデキシングは、ライブラリー調製の間にいつでも、どの分子が変換されるかおよびサンプル毎にどのように変換されるかの追跡を可能にする。それにより、サンプルタイプに依存しないライブラリー変換係数の算出を可能にする。これは、サンプル調製に品質管理を組み

50

込むことを可能にする。この技術を使用して、試薬ロット、酵素反応の効率およびシーケンシングに関する多様な指標を比較することを可能にする。これらの比較は、通常では、問題が発生したかを決定するために、ライブラリー調製プロセスの終了を待たなければならないシーケンシング反応のステップに沿って行うことができる。

【0321】

いくつかの実施態様では、正規化対照とサンプルとの混合は、ライブラリー調製ステップと同時に進行。必要に応じて、適切なアダプターを有する正規化対照は、ライブラリー調製プロセスの任意の時点で添加されてよい。

【0322】

いくつかの実施態様では、正規化対照とサンプルとの混合は、ライブラリー調製ステップの後に行う。適切なインデキシングおよび／またはシーケンシングアダプターを有する正規化対照は、ライブラリー調製の後であるがNGSの前にライブラリーに添加することができる。いくつかの実施態様では、プロセス内でこの時点で添加された正規化対照は、シーケンシング反応自体におけるバイアスをコントロールする。例えば、高いまたは低いGC含有率を有する正規化対照は、パイロシーケンシング法における既知のGCバイアスをコントロールすることができる。

ライブラリー

【0323】

本開示のシーケンシング方法は、シーケンシングライブラリーの調製を含む。シーケンシングライブラリー調製は、次に、シーケンシングの準備ができたアダプター修飾DNAフラグメントのコレクションの製造を含む。ポリヌクレオチドのシーケンシングライブラリーは、等価物、DNAの類似体またはcDNA（これは例えば、逆転写酵素の作用によりRNAテンプレートから製造された相補的または複製DNAである）を含めたDNAまたはRNAから調製することができる。ポリヌクレオチドは、二本鎖DNA（dsDNA）型（例えばゲノムDNAフラグメント、PCRおよび増幅産物）に起源があってもよく、またはDNAもしくはRNAとして一本鎖型に起源があり、かつdsDNA型に変換されたポリヌクレオチドであってよい。例えば、mRNA分子は、シーケンシングライブラリーの調製で使用するために適した二本鎖cDNA内へ複製されてよい。

【0324】

いくつかのNGSシーケンシングプラットフォームのためのシーケンシングライブラリーの調製は、ポリヌクレオチドが特定の範囲、例えば0～1200bpのフラグメントサイズであることを必要とする。したがって、大きなポリヌクレオチドの断片化を必要としてもよい。ポリヌクレオチドが強制的に断片化されるかまたは自然にフラグメントとして存在するかにかかわらず、これらは、5-リン酸塩および3'-ヒドロキシルを有する平滑末端DNAに変換される。ライブラリー調製における断片化のプロセスは、NGSを使用して標的配列を定量化する際の誤差の潜在的な一つの原因であり、ライブラリー調製における断片化ステップの前に正規化対照の包含によりモデル化することができる。

【0325】

典型的には、フラグメント末端は、末端修復され、例えば当業者に公知の方法またはキットを用いて平滑末端にされる。平滑末端フラグメントは、例えばポリヌクレオチドキナーゼを使用して、酵素処理によりリン酸化することができる。いくつかの実施態様では、単一のデオキシヌクレオチド、例えばデオキシアデノシン（A）は、例えば、Taqポリメラーゼまたはクレノウエキソマイナスポリメラーゼのような特定の型のDNAポリメラーゼの活性により、ポリヌクレオチドの3'-末端に付加される。dAテール産物は、後続のステップでライゲーションされるアダプターの各二重領域の3'末端に存在する「T」オーバーハングと適合性がある。dAテリングは、アダプターライゲーション配列の形成に偏りを持つような平滑末端ポリヌクレオチドの両方のセルフライゲーションを防ぐ。dAテールポリヌクレオチドは、二本鎖アダプターポリヌクレオチド配列にライゲーションされる。同じアダプターを、ポリヌクレオチドの両末端に使用することができるか、またはアダプターの二つのセットを使用することができる。ライゲーション方法は、当業

10

20

30

40

50

者に公知であり、DNAリガーゼのようなリガーゼ酵素を利用して、アダプターをd-Aテールポリヌクレオチドに共有結合させる。アダプターは、標的3'-OHへのライゲーションを容易にするために、5'-リン酸塩基を含んでいてよい。d-Aテールポリヌクレオチドは、切断プロセスから残ったかまたは酵素処理ステップを用いて付加された5'-リン酸塩基を含み、末端修復され、任意に、単数または複数のオーバーハングする塩基により伸長され、ライゲーションのために適した3'-OHが付与される。ライゲーション反応の産物を精製して、ライゲーションしていないアダプター、互いにライゲーションしたアダプターを除去し、クラスター生成のためのテンプレートのサイズ範囲を選択し、この選択の前に増幅、例えばPCR増幅を行うことができる。ライゲーション産物の精製は、ゲル電気泳動および固相可逆固定(SPRI)を含む方法により行うことができる。

10

【0326】

標準プロトコル、例えばIlluminaプラットフォームを使用したシーケンシング用プロトコルは、d-Aテールリングの前に末端修復された産物を精製し、ライブラリー調製のアダプターライゲーションステップの前にd-Aテールリング産物を精製することをユーザーに教示してよい。末端修復産物およびd-Aテール産物の精製は、酵素、緩衝液、塩などを除去して、引き続き酵素によるステップのために好ましい反応条件を提供する。いくつかの実施態様では、末端修復、d-Aテールリングおよびアダプターライゲーションのステップは、精製ステップを除外する。したがって、いくつかの実施態様では、本開示の方法は、末端修復、d-Aテールリングおよびアダプターライゲーションの連続したステップを含むシーケンシングライブラリーの調製を含む。d-Aテールリングステップを必要としないシーケンシングライブラリーを調製するための実施態様、例えばRoche 454およびSOLID(商標)3プラットフォームを使用するシーケンシング用プロトコルでは、末端修復およびアダプターライゲーションのステップは、アダプターライゲーションの前の末端修復産物の精製ステップを除外する。

20

【0327】

いくつかの実施態様では、増幅反応は調製される。増幅ステップは、アダプターライゲーションテンプレート分子に、フローセルへのハイブリダイゼーションに必要なポリヌクレオチド配列を導入する。増幅反応の内容は、当業者に公知であり、適切な基質(例えばdNTP)、酵素(例えばDNAポリメラーゼ)および増幅反応に必要な緩衝液成分を含む。場合により、アダプターライゲーションポリヌクレオチドの増幅は省くことができる。一般的に、増幅反応は、少なくとも二つの増幅プライマー、すなわち同じであってもよいプライマーオリゴヌクレオチドを必要とし、アニーリングステップの間に、増幅されるポリヌクレオチド分子(または、テンプレートが一本鎖として表される場合にその相補体)内のプライマー結合配列にアニーリングすることができる「アダプター特異的部分」を含む。一端形成されると、上述の方法により調製されたテンプレートのライブラリーは、固相核酸増幅のために使用することができる。ここで使用されるような「固相増幅」の用語は、増幅産物の全部または一部が形成される際に固体支持体上に固定されるように、固体支持体上でまたは固体支持体と会合して実施されるあらゆる核酸増幅反応を言う。特に、この用語は、順方向および逆方向増幅プライマーの一方または両方が固体支持体上に固定されていることを除いて、標準溶液相増幅に類似する反応である固相ポリメラーゼ連鎖反応(固相PCR)および固相等温増幅を含む。固相PCRは、一方のプライマーがビーズに固定され、他方が遊離溶液内にあるエマルジョンのような系、および一方のプライマーが表面に固定され、他方が遊離溶液内にある固相ゲルマトリックス中でのコロニー形成をカバーする。増幅の後に、シーケンシングライブラリーは、このライブラリーがアダプター二量体または一本鎖DNAを含まないことを確認するために、マイクロ流体キャピラリー電気泳動により分析することができる。テンプレートポリヌクレオチド分子のライブラリーは、固相シーケンシング法で使用するために特に適している。固相シーケンシングおよび固相PCRのためのテンプレートを提供することに加えて、ライブラリーテンプレートは、全ゲノム増幅のためのテンプレートを提供する。

30

40

【0328】

50

いくつかの実施態様では、アダプターライゲーションポリヌクレオチドのライブラリーは、例えば、数百万のクラスターを有する高密度シーケンシングフローセルを作製するために、ランダムに断片化されたゲノムDNAの、平坦な光学的に透明な表面へのアタッチメントを用いてかつ固相増幅する、数百万の核酸のフラグメントをシーケンシングするための技術を含む超並列シーケンシングに供される。Solexa/Illumina法は、ランダムに断片化されたゲノムDNAの、平坦な光学的に透明な表面へのアタッチメントに頼る。付着したDNAフラグメントは伸長され、ブリッジ増幅されて、同じテンプレートの数千の複製をそれぞれ含む数百万のクラスターを有する超高密度シーケンシングフローセルを作製する。クラスターテンプレートは、除去可能な蛍光色素を有する可逆ターミネーターを使用するロバストな四色DNAシーケンシングバイーシンテーシス技術を使用して配列決定される。あるいは、このライブラリーは、各ビーズが順方向および逆方向の増幅プライマーを含むビーズ上で増幅されてよい。

10

【0329】

増幅されたライブラリーのシーケンシングは、ここに記載されたようなあらゆる適切なシーケンシング技術を使用して実施することができる。

シーケンシング

【0330】

本開示は、次世代シーケンシングを使用して、サンプル、例えば正規化対照を含むサンプルをシーケンシングする方法を提供する。

【0331】

20

いくつかの実施態様では、ここに記載された方法は、クローン増幅DNAテンプレートまたは単一DNA分子がフローセル内で超並列方式で配列決定される次世代シーケンシング技術(NGS)を使用する(例えばVolkerding et al. Clin Chem 55:641-658 [2009]; Metzker M Nature Rev 11:31-46 [2010]に記載されている)。ハイスループット配列情報に加えて、NGSは、各配列リードが個々のクローンDNAテンプレートまたは単一DNA分子を表す可算の「配列タグ」であるデジタル定量情報を提供する。NGSのシーケンシング技術は、パイロシーケンシング、可逆色素ターミネーターを用いるシーケンシングバイーシンテーシス、オリゴヌクレオチドプロプライゲーションによるシーケンシングおよびリアルタイムシーケンシングを含む。

【0332】

30

シーケンシング技術のいくつかは、Affymetric Inc. (Sunnyvale, Calif.) からのシーケンシングバイーハイブリダイゼーションプラットフォームおよび454 Life Sciences (Bradford, Conn.)、Illumina/Solexa (Hayward, Calif.) およびHelicos Biosciences (Cambridge, Mass.) からのシーケンシングバイーシンテーシスプラットフォームならびにApplied Biosystems (Foster City, Calif.) からのシーケンシングバイーライゲーションプラットフォームのように、商業的に入手可能である。Helicos Biosciencesのシーケンシングバイーシンテーシスを用いて行われる単一分子シーケンシングプラットフォームに加えて、他の単一分子シーケンシング技術は、本開示の方法に包含され、Pacific BiosciencesのSMRT (商標) 技術、Ion Torrent (商標) 技術、および例えばOxford Nanopore Technologiesにより開発されたナノポアシーケンシングを含む。しかしながら、本開示は、本来的に低入力サンプルの定量化の問題を抱える、開発されたあらゆる将来のハイスループットシーケンシング技術と適合すると考えられる。

40

【0333】

自動化サンガー法は、「第一世代」技術と見なされているが、自動化サンガーシーケンシングを含むサンガーシーケンシングを、本開示の方法により使用することもできる。

【0334】

本開示の方法のいくつかの実施態様では、使用されたDNAシーケンシング技術は、(例えば、Margulies, M. et al. Nature 437:376-380 [2005]に記載されたような) 454シーケンシング(Roche)である。約300~800塩基対の454シーケン

50

シングDNAフラグメントは、平滑末端にされ、アダプターが、このフラグメントの末端にライゲーションされる。このアダプターは、フラグメントの増幅およびシークエンシングのためのプライマーとして機能する。このフラグメントを、DNA捕捉ビーズ、例えば5'-ビオチンタグを含むアダプターBを用いて、例えばストレプトアビジン被覆ビーズに取り付けることができる。ビーズに取り付けられたフラグメントは、油水エマルションの液滴内でPCR増幅される。その結果、各ビーズ上にクローン増幅されたDNAフラグメントの多数の複製が生じる。第二のステップでは、これらのビーズをウェル（ピコリットルサイズ）内に取り込む。パイロシークエンシングは、各DNAフラグメントに関して並行して行われる。一つまたは複数のヌクレオチドの添加は、光信号を生成し、これは、シークエンシング機器内でCCDカメラにより記録される。信号強度は、組み込まれたヌクレオチドの数に比例する。パイロシークエンシングは、ヌクレオチド添加の際に放出されるピロリン酸塩（PPi）を利用する。PPiは、アデノシン5'ホスホ硫酸の存在でATPスルフィラーゼによりATPに変換される。ルシフェラーゼは、ATPを使用してルシフェリンをオキシルシフェリンに変換し、この反応は、光を生成し、この光を識別し、かつ分析する。454シークエンシングは、長さが100~1000bpであるリードを製造することができる。

【0335】

本開示の方法のいくつかの実施態様では、使用されるDNAシークエンシング技術は、SOLiD（商標）技術（Applied Biosystems）である。SOLiD（商標）シークエンシングバイーライゲーションにおいて、5'および3'は、DNAフラグメントに取り付けられて、フラグメントライブラリーを生成する。次に、クローンビーズ集団は、ビーズ、プライマー、テンプレート、およびPCRコンポーネントを含むマイクロリアクター内で調製される。PCRの後、テンプレートは変性され、ビーズは、伸長されたテンプレートを有するビーズを分離するために濃縮される。選択されたビーズ上のテンプレートは、ガラススライドへの結合を可能にする3'修飾が行われる。配列は、部分的なランダムオリゴヌクレオチドと、特定のフルオロフォアにより同定される中央の決定された塩基（または塩基の対）との連続的ハイブリダイゼーションおよびライゲーションにより決定することができる。色が記録された後、ライゲーションされたオリゴヌクレオチドは切断され、除去され、次いでこのプロセスが繰り返される。SOLiD技術は、約50bpのリードを製造することができる。

【0336】

本開示の方法のいくつかの実施態様では、使用されたDNAシークエンシング技術は、Pacific Biosciencesの単一分子リアルタイム（SMRT（商標））シークエンシング技術である。SMRT技術において、色素標識ヌクレオチドの連続的組み込みは、DNA合成の間に画像化される。単一DNAポリメラーゼ分子は、配列情報を取得する個々のゼロモード波長識別器（ZMW識別器）の底面に取り付けられていて、リン酸連結されたヌクレオチドは成長するプライマー鎖内に取り込まれる。ZMWは閉じ込め構造であり、これは、ZMWの外に（マイクロ秒で）急速に拡散する蛍光ヌクレオチドのバックグラウンドに対して、DNAポリメラーゼによる単一ヌクレオチドの組み込みの観察を可能にする。ヌクレオチドを成長する鎖内に組み込むために数マイクロ秒かかる。この時間の間、蛍光ラベルは励起され、蛍光信号が製造され、蛍光タグは切断される。色素の対応する蛍光の同定は、どの塩基が組み込まれたかを示す。このプロセスは繰り返される。RS機器は、時には2500または2900bpまでのリードで、約1100bpの平均リード長さを有する。RS IIは、8,500~60,000bpのリード長さを製造することができる。

【0337】

本開示の方法のいくつかの実施態様では、使用されたDNAシークエンシング技術は、（例えば、Soni G VおよびMeller A. Clin Chem 53: 1996-2001 [2007]）に記載されたような）ナノポアシークエンシングである。ナノポアシークエンシングDNA分析技術は、Oxford Nanopore Technologies（Oxford, United Kingdom）を含む多くの企業により工業的に開発されている。ナノポアシークエンシングは、DNAの単一分子がナ

ノポアを通過する際に直接配列決定する単一分子シーケンシング技術である。ナノポアは、直径1ナノメートルのオーダーの小さな穴であり、ナノポアを導電性液体に浸漬し、横切るように電位（電圧）を印加することで、ナノポアを通したイオンの伝導により僅かな電流が生じる。流れる電流の量は、ナノポアのサイズおよび形状に敏感である。DNA分子がナノポアを通過する際に、DNA分子の各ヌクレオチドはナノポアを様々な程度で妨害し、ナノポアを通過する電流の大きさを様々な程度で変化させる。したがって、DNA分子がナノポアを通過する際の電流についての変化は、DNA配列の読取りを示す。NanoNanopore MinIONは、10,000bpを越えるリード長さを製造することができる。

【0338】

いくつかの実施態様では、DNAシーケンシング技術は、Ion Torrent単一分子シーケンシングであり、これは半導体技術と単純なシーケンシング化学とを組み合わせ、化学的にコードされた情報（A、C、G、T）を半導体チップ上でデジタル信号（0、1）に直接変換する。実際に、ヌクレオチドがポリメラーゼによりDNAの鎖内に組み込まれる際に、水素イオンが副産物として放出される。Ion Torrentは、微細加工されたウェルの高密度アレーを使用して、この生化学的プロセスを超並列法で実施する。各ウェルは、異なるDNA分子を保持する。ウェルの下に、イオン感受性層があり、その下に、イオンセンサーがある。ヌクレオチド、例えばCが、DNAテンプレートに付加され、次いでDNAの鎖内に組み込まれる際に、水素イオンが放出される。そのイオンからの電荷は、溶液のpHを変化させ、これをIon Torrentのイオンセンサーにより識別することができる。本質的に世界最小の固体pHメータであるシーケンサーは、化学的情報からデジタル情報に直結する塩基を呼び出す。次に、Ion personal Genome Machine（PGM（商標））シーケンサーは、一つのヌクレオチドを次々にチップに流し込む。チップに流し込まれた次のヌクレオチドはマッチしない場合、電圧の変化は記録されず、塩基は呼び出されない。DNA鎖に二つの同一の塩基が存在する場合、電圧は二倍になり、チップは、呼び出される二つの同一の塩基を記録する。直接同定は、数秒でのヌクレオチドの組み込みの記録を可能にする。Ion Torrentは、約400bpのリード長さを製造することができる。

【0339】

いくつかの実施態様では、この方法は、Illuminaのシーケンシングバイーシンテシスおよび（例えば、Bentley et al., Nature 6:53-59 [2009]に記載されたような）可逆ターミネーターに基づくシーケンシング化学を用いる数百万のDNAフラグメントの超並列シーケンシングを使用する。Illuminaのシーケンシング技術は、断片化されたゲノムDNAの、オリゴヌクレオチドアンカーが結合している平坦な光学的に透明な表面への取り付けに頼る。テンプレートDNAは、末端修復されて、5'-リン酸化平滑末端が生成され、クレノウフラグメントのポリメラーゼ活性を使用して、平滑リン酸化DNAフラグメントの3'末端に単一のA塩基を付加する。この付加は、ライゲーション効率を高めるためにその3'末端に単一のT塩基のオーバーハングを有するオリゴヌクレオチドアダプターへのライゲーション用のDNAフラグメントを調製する。このアダプターオリゴヌクレオチドは、フローセルアンカーに対して相補的である。限界希釈条件下で、アダプター修飾一本鎖テンプレートDNAはフローセルに添加され、アンカーへのハイブリダイゼーションにより固定される。取り付けられたDNAフラグメントは伸長され、ブリッジ増幅され、同じテンプレートの約1,000の複製をそれぞれ含む数億のクラスターを有する超高密度シーケンシングフローセルを作製する。このテンプレートは、除去可能な蛍光色素を有する可逆ターミネーターを使用するロバストな四色DNAシーケンシングバイーシンテシス技術を使用して配列決定される。高感度蛍光同定は、レーザー励起および全反射光学系を使用して達成される。短い配列リードは、リピートマスクされた参照ゲノムに対してアライメントされ、特別に開発されたデータ分析パイプラインソフトウェアを用いて遺伝的相違が呼び出される。第一のリードの完了後に、テンプレートをインサイトウで再生成して、フラグメントの反対側の末端から第二のリードを可能にするこ

10

20

30

40

50

とができる。したがって、DNAフラグメントのシングルエンドかペアードエンドのシーケンシングのいずれかが、この方法により使用される。Illuminaリード長さは、特定のプラットフォームに依存するが、一般的に50～400bpである。

定量化

【0340】

「式」、「アルゴリズム」、または「モデル」は、一つまたは複数の連続的またはカテゴリー的入力を取りかつ出力値を計算するあらゆる数学的方程式、アルゴリズム的な、分析的なもしくはプログラムされたプロセス、または統計的手法であり、時には「インデックス」または「インデックス値」とも言われる。「式」の非限定的な例は、和、比、および係数または指数のような回帰演算子、バイオマーカー値の変換および正規化、統計的分類モデル、および履歴的母集団で学習されたニューラルネットワークを含む。サンプル内の開始濃度とNGS後の出力リードとの間の関係を決定する際に特に使用するの、線形および非線形方程式ならびに統計的分類分析であり、サンプル内の開始濃度とNGS後の配列リードとの間の関係が決定される。

【0341】

特に興味深いのは、相互相関、主成分分析(PCA)、因子回転、ロジスティック回帰(LogReg)、線形判別分析(LDA)、サポートベクターマシン(SVM)、ランダムフォレスト(RF)、回帰的分割木(RPART)、ならびに他の関連する決定木分類技術、縮小セントロイド(SC)、StepAIC、Kth-Nearest Neighbor、Boosting、決定木、ニューラルネットワーク、ベイジアンネットワーク、および隠れマルコフモデル等のような確立された技術を含むパターン認識機能を利用した構造的および構文的な統計分類アルゴリズムである。これらは、付加的配列データとモデルの改良との間のトレードオフを定量化し、オーバーフィットを最小化するために、赤池情報量基準(AIC)またはベイズ情報量基準(BIC)のような情報量基準と組み合わせてよい。得られた予測モデルは、他の研究において検証するか、またはブートストラップ、Leave-One-Out(LOO)および10分割交差検証(10-Fold CV)のような技術を用いて、最初に学習した研究において交差検証してよい。多様なステップで、偽発見率は、当分野で公知の技術による値の順列により推測されてよい。

線形相関

【0342】

本開示の方法のいくつかの実施態様では、正規化対照のグループおよび／またはサブグループの濃度は、線形配列を含む。いくつかの実施態様では、正規化対照の設計の線形性は、ライブラリー内のリードの間の相関を可能にする。サンプルに添加したときのまたはライブラリーを調製する間のNC分子の濃度は完全に相関し、線形モデルに従う。この関係は、これらのNCポリヌクレオチドにより共有される出力リードについても当てはまると考えられる。NCの開始濃度を対応する正規化されたリードカウントに対してプロットし、回帰の R^2 値を決定する場合に、0.8より高い R^2 は、抽出および下流ライブラリー調製プロセスが誤りなく行われたことを示す。

【0343】

このNCは、一定の濃度および異なるサイズでライブラリーに添加される。異なるポリヌクレオチドのサイズの範囲は、ライブラリー内で同定される標的DNAまたはRNAのインサートサイズの表示であることを意味する。調製の間にサンプルまたはライブラリーに添加されるNCの濃度は、予め決定されているため、サンプル内に存在する標的配列の濃度は、標的のリードカウントをNCによって決定される線形モデルにフィッティングすることにより決定することができる。このモデルでは、標的配列、サンプル内のバックグラウンド核酸、またはNCが、ライブラリー調製プロセスの間に有するバイアスは、最小である。各々のNC線形性は、その独自の線形方程式、 $y = mx + c$ を有する。Yは、所定の問題における標的配列／NCのlog(百万当たりのリード、RPM)であり、Mは直線の傾きであり、Cは切片である。Xは、標的配列に対して未決定である。理論的には、NCは、アtomol／ μ Lで示すlog(濃度)として計算することができる線形性に従

う。この濃度は、 μL 当たりのライブラリー内に存在する標的配列のフラグメントの数に変換される。

【0344】

このモデルの成功は、標的力価の推定を開始するために、0.8を超える R^2 を有するNCの線形性に依存する。0.8より低い R^2 は、サンプル調製および／またはハイスループットシーケンシングの間のNCの品質管理の失敗を示唆する。このモデルでは、NCの線形性は、絶対濃度に基づいて確立される。これはさらに、インサートサイズが種類だけであり、100%のゲノムカバレッジであり、かつ最小のサンプル変動であることを前提とする。サンプル変動は、バックグラウンドサンプル配列（「標的」以外のサンプル配列）にマッピングされたリードのパーセント、NCにマッピングされたリードのパーセント、および理論的には枯渇の割合を計算することにより決定することができる。標的配列の単一挿入サイズが望ましいが、これはまれにしか達成可能でない。理想的には、ヒストグラム内で、標的挿入サイズの3～5の最も代表的なビンが選択され、各々のビン内のリードのパーセントが計算される。これにより、本質的にゲノムカバレッジの問題を処理する。挿入分子の数は、これらのビンの各々における数に依存する。

10

【0345】

標的挿入分子の数または濃度は、線形関係 $Y = mx + C$ を使用して計算することができる。標的分子の数

【数1】

×

20

力価（の関数である）；および力価

【数2】

×

f（インサートサイズビン、ゲノムカバレッジ、%ヒトマッピング、%NCマッピング）。

【0346】

例えば、品質管理を可能にするため、正規化対照は、相互の比率で添加することができる。これは、観察された比率の計算、および全体のプロセス／アッセイを通して核酸が期待通りに変換されたことの確認を可能にする。この比率が期待通りになった場合（例えば、NC1、NC2、およびNC3は、1：2：4の濃度比でサンプルに添加され、次いで、リードカウント／エビデンスは、その比率の100リード：200リード：400リードを回復する）。図1は、例示的な正規化対照品質管理チェックを示す。合成DNAポリヌクレオチドではなくゲノムDNAを含む正規化対照、例えばバクテリオファージT4正規化NCの品質管理のためにそのゲノム（例えばバクテリオファージT4ゲノム）から回復されたリードの最小数が探される。

30

【0347】

いくつかの場合には、NCは、標準曲線を正規化するために使用される。単純な場合では、存在量が決定される観察されたリード（ここでは「標的」リードとも言われる）は、観察されたNCリード（組合せまたは個別のいずれかで、より良好な標準曲線の生成が可能であるかに依存する）で除算される。この場合、正規化標的リード存在量は、次のように計算される：

40

正規化されたリード存在量 = [（観察された標的）／（観察されたNC）] * C

（式中、C = 定数、必要な場合）

図2は、CMV DNAが添加された血漿からDNA（標的）が抽出されたDNAライブラリーの次世代シーケンシングにおけるサイトメガロウイルスCMV（標的）ウイルス力価対リード存在量のプロットを示す。点線は、CMV力価対CMVリード存在量の線形回帰を示す。正規化なしでは、 R^2 値は0.89である。対照的に、正規化した場合、R

50

²値は0.98である(図3)。したがって、正規化対照は、初期サンプル内の標的(ここではCMV)の存在量と、ライブラリー調製および次世代シーケンシングの後の標的にマッピングされたリードのパーセンテージの間の相関を改善する。

【0348】

ライブラリー調製の前にサンプルに添加された正規化対照は、異なる実験条件から生じる誤差およびサンプル内の核酸に影響を及ぼす統計的／ランダム誤差を説明することができる。可能な限り厳密に実際の生物学を模倣するように設計されている場合、調製、変換、シーケンシング等における多くの相違を説明しかつ正規化することができる。

【0349】

しかしながら、いくつかの場合では、模倣できない、したがって正規化対照によって説明することができないいくつかの生物学的要因がある。このような場合では、生のリードから標的濃度への変換は、次の等式により表すことができる：

【数3】

$$\tilde{r} = f(r) + \Delta r$$

式中、 r は、生の標的リードエビデンス(またはカウント)であり、 $f(r)$ は、説明された全ての変数についてリードを正規化することができる変換関数であり、 Δ 関数は、模倣または説明できないプロセスによる残留誤差である。理想的には、 $\Delta(r)$ はゼロになり、この変換は、生の標的リードエビデンスを初期開始濃度(例えばミリリットル当たりのウイルス複製)に変換する。単純な関数 $f(r)$ の例は、単純な乗法的因子である。この場合では、 $f(r)$ は、実際の入力パーセンテージ／正規化対照の存在量および正規化対照の(リードカウント／エビデンスを経て)観察された存在量から計算される：

$$f(r) = f \cdot r$$

式中、 f は次のように計算することができる

$$f = \text{実際の存在量} / \text{観察された存在量}$$

【0350】

$f(r)$ の他の型は、標的GC含有率、標的サイズ、フラグメンテーション条件、標的がタンパク質(例えばキャプシド内のウイルスゲノム)と会合しているかまたは遊離DNA分子であるかなどの標的特性を説明しかつ模倣することができる。この正規化対照が全てのこれらの変数を説明するように設計されている場合、 Δ はますます小さくなる。この視覚的表現は、図4における仮想的な例に記されている。図4では、ウイルスに感染した宿主から単離した核酸サンプルが、次世代シーケンシングを使用して三つに分けて分析されている。図4では、各々の次世代シーケンシングライブラリーの左端のバーは、開始サンプル内の既知の非宿主濃度(ウイルス負荷)である(つまり、サンプル内の「標的」の開始存在量)。ライブラリーの調製、次世代シーケンシングおよびリードのアライメントの後に、リードカウントは、ウイルス負荷を回復するように変換される。各ライブラリー内の左からの二番目のバーは、得られたウイルスエビデンスだけを使用してウイルスのパーセント存在量を計算する例である。図4は、得られたウイルスエビデンスだけから計算された開始サンプル内のウイルスのパーセント存在量(ウイルス負荷)を計算する場合、各ライブラリー間に大きな変動性があることを示す。また、この計算は、オリジナルサンプル内の実際のウイルス負荷を反映していない。 $f(r)$ 関数を使用して、生のリードエビデンスを、ライブラリー間でより比較可能な量に変換することができる(図4、各ライブラリーについて左から三番目のバー)。 $f(r)$ 関数を使用して、ウイルスの計算されたパーセント存在量の変動性は、ライブラリー間で減少し、得られた値は、実際のウイルス量に近づく。ウイルス負荷に直接つながる唯一の残りの変換は、 $\Delta(r)$ である。 $\Delta(r)$ の正確な計算は、ライブラリー間の変動性をさらに減少させ、ウイルス負荷が(誤差内で)得られる。

【0351】

正規化対照を注意深く構築することで、各ライブラリーおよび各標的についての $f(r)$ の推定が可能になる。ランダム配列を使用することができるが、より多くの生物学的情報を追加することもできる。このアプローチの特徴は、 $f(r)$ を計算できる場合には、生のリードを初期サンプル内の標的存在量に直接変換するために、多数のサンプルの実行を回避することができることである。 $f(r)$ は、リードカウントと標的存在量との間のギャップを埋め、変動性を低下させる。 $\Delta(r)$ は、正規化対照内に含めることができなかった全ての変数を説明する実験を設計することにより推定することができる。例えば、 $\Delta(r)$ は、細胞壁、例えば細菌細胞壁、またはウイルスについてのキャプシド環境に起因することができ、この効果は、実験的に決定することができる。初期サンプル内の標的存在量と出力リードカウントの間のギャップを $f(r)$ で埋めることにより、 Δ は最小化される。 Δ 、おそらくは乗法係数 ($y = ax + b$) は、多数のサンプルから計算することができる。特定の実施態様では、 Δ についての高次モデルが適切であってもよい。図 5 は、ウイルス負荷を計算するために、生のリードとは反対に、 $f(r)$ を使用する場合、変動性における予想される低減が示される。多くのサンプルを実行して、 Δr を計算して図 5 に示された線の等式を取得する前に、変動性を、 $f(r)$ によって可能な限り低減させる。

10

【0352】

$f(r)$ および Δr の後計算における変動性の一つの潜在的原因は、アッセイの実効の部位である。いくつかの場合では、アッセイが二次サイトで実行される（例えば、別の病院が患者サンプル内のウイルス力価を計算する）場合、アッセイの実効は、標的と正規化対照とのセットに対して、最初に生成された線に完全には当てはまらないデータを提供することがある。この場合では、アッセイは二次サイトで確認される。例えば、二次サイトは、ウイルス負荷の変換を計算するために十分なサンプルを実行し、その等式を、実際のサンプルからウイルス負荷を計算する際に使用されるウェブアプリに入力することができる。

20

【0353】

サンプルからシーケンシングリードへの変換に影響を与える可能性のある変数のいくつかの表を、以下に表 1 として示す。表 1 は、NC 対照設計において考えられる可能な変数の概要を示す。表 1 は、標的がウイルスである場合に NC の設計において考えられる可能な変数の概要を示す。しかしながら、これらの可能な変数のいくつかは、他の標的に適用されてよい。

30

40

50

【表 1 - 1】

表 1 NC設計における変数

変数	カテゴリー／ノイズ	NC設計戦略	寄与	試験の説明
GC含有率	生物学	全てのTPxウイルス内のGC含有率を計算し、NCにその範囲を模倣させる	$f(r)$	これらのNCでライブラリーを実行し、リードエビデンスを計算し、NCカウントの平均が $f(r)$ のために使用できるか、またはウイルスに近いGC含有率を有するNCのコンポーネントが、より良好な $f(r)$ 乗算係数を生成するか（または $f(r)$ がより複雑である場合に観察し、生のリード存在量をウイルス負荷に僅かな変動性で近づくか）どうかを見る。
場所バイアス	ランダム	なし	$\Delta(r)$	場所の変動性の影響を推定するために、多様な場所で十分なサンプルを送る。
ウイルス断片化	生物学	フラグメントに対して十分な大きさのNCを得る	$f(r)$	これらのNCでライブラリーを実行し、リードエビデンスを計算し、NCカウントの平均が $f(r)$ のために使用できるか、またはウイルスに近い断片化パターンを有するNCのコンポーネントが、より良好な $f(r)$ 乗算係数を生成するか（または $f(r)$ がより複雑である場合に観察し、生のリード存在量をウイルス負荷に僅かな変動性で近づく）どうかを見る。

10

20

30

40

50

【表 1 - 2】

変数	カテゴリー／ノイズ	NC設計戦略	寄与	試験の説明
キャプシド環境からの抽出（抽出バイアス）	生物学	なし／ウイルスと同等のキャプシドにNCを添加	$\Delta(r)$	十分なサンプルを実行し、線形範囲が観察される場合、高いR ² で決定し、線の等式を取得する。
NCの存在量対ウイルス	生物学	ウイルスにとって臨床的に適切なレベルでNCをスパイクイン	$f(r)$	これらのNCでライブラリーを実行し、リードエビデンスを計算し、NCカウントの平均が $f(r)$ のために使用できるか、またはウイルスに近い存在量を有するNCのコンポーネントが、より良好な $f(r)$ 乗算係数を生成するか（または $f(r)$ がより複雑である場合に観察し、生のリード存在量をウイルス負荷に僅かな変動性で近づく）どうかを見る。

【0354】

したがって、本開示は、サンプルと混合された正規化対照内の複数の核酸分子の各々の開始濃度と、NGSにより製造されたリードの数との間の関係を計算し、リードとサンプル内の濃度との間の関係をモデル化し、このモデルを使用して標的核酸分子により製造されたリードの数から、サンプル内の少なくとも一つの標的核酸分子の初期濃度を計算する方法を提供する。

【0355】

いくつかの実施態様では、モデルは線形モデルである。いくつかの実施態様では、サンプル内の少なくとも一つの核酸分子の初期濃度対標的核酸分子により製造されたリードの数の線形回帰は、正規化後に0.95より大、0.96より大、0.97より大、0.98より大または0.99より大のR²値を有する。いくつかの実施態様では、線形回帰のR²値は、引き続き正規化の間に少なくとも0.01、0.03、0.05、0.07、0.09、0.1、0.13、0.15、0.17、0.19、0.2、0.23、0.25、0.27、0.29、0.3、0.33、0.35、0.37、0.39、0.4、0.43、0.45、0.47、0.49、0.5、0.53、0.55、0.57、0.59、0.6、0.63、0.65、0.67、0.69、0.7、0.8、0.9、1.0または任意の値だけ改善する。

機械学習を使用した力価の決定

【0356】

サポートベクターマシン（SVM）は、薬物毒性および抗菌物質耐性を予測するための教師付学習アプローチの公知のメカニズムである。これらは、分類、回帰および異常値検出のために使用される方法のセットである。SVMがデータを予測するためには、データにフィットしていなければならない。トレーニングデータのセットが与えられると、各デ

ータポイントは二つのカテゴリーの一方に属するとマークされ、SVMトレーニングアルゴリズムは、新しいデータをカテゴリーの一つに割り当てるモデルを構築する。SVMは、分類装置をトレーニングするために教師付機械学習アプローチを使用する非確率的バイナリ線形分類装置である。

【0357】

いくつかの実施態様では、正規化対照データは、SVMモデルをフィッティングするために使用される。いくつかの実施態様では、このデータの80%は、SVMをトレーニングするために使用され、このデータの残りの20%は、SVMをテストするために使用される。モデルへのフィッティングは、線型モデルだけが考慮されていないフィードバックプロセスとして機能する。NGSライブラリーに入った正規化対照の実際の濃度に最も近い予測を有するモデルは、サンプル内の標的配列の力価の予測のために選択されたモデルを決定する。

10

【0358】

したがって、本開示は、サンプルと混合された正規化対照内の複数の核酸分子の各々の開始濃度と、NGSにより生成されたリードの数との間の関係を計算し、リードとサンプル内の濃度との間の関係をモデル化し、このモデルを使用して標的核酸分子により製造されたリードの数から、サンプル内の少なくとも一つの標的核酸分子の初期濃度を計算する方法を提供する。

【0359】

いくつかの実施態様では、モデルは、機械学習分類装置を用いて作成される。いくつかの実施態様では、機械学習分類装置は、教師付である。いくつかの実施態様では、機械学習分類装置は、ベクターサポートマシンである。

20

【0360】

全てのパーセンテージおよび比率は、特に明記されていない限り、質量または分子の数（例えばモル比）により計算される。適切な測定単位は、当業者に容易に理解される。

【0361】

全てのパーセンテージおよび比率は、特に明記されていない限り、全ての正規化対照組成物に基づいて計算される。

マルチアナライト対照

【0362】

ここに提供されるのは、マルチアナライト対照、マルチアナライト対照を含む組成物、マルチアナライト対照の製造方法、およびマルチアナライト対照の使用方法である。ここに記載されたように、マルチアナライト対照は、不活性化された生物種の混合物を含む。マルチアナライト対照に含めることができる生物種を制限することなく、ウイルス、細菌、真菌および真核生物の寄生体（単細胞および多細胞の両方）は全て、混合物中に含めることができる生物種またはマルチアナライト対照内の種として想定される。いくつかの実施態様では、マルチアナライト対照はさらに、適切な担体、希釈剤または付形剤を含む。マルチアナライト対照、およびマルチアナライト対照を含む組成物は、ここでは互換的に言及される。

30

【0363】

特定の用途に拘束されることを望むものではないが、本開示のマルチアナライト対照は、ハイスループットシーケンシングおよびサンプル分析において多くの用途を有する。いくつかの実施態様では、マルチアナライト対照は、自律的に、すなわちここに記載された正規化対照の使用なしで使用される。例えば、マルチアナライト対照は、ライブラリー調製およびシーケンシングのようなサンプル処理ステップの間に、陽性対照として使用することができる。いくつかの実施態様では、マルチアナライト対照は、ここに記載された正規化対照と協調して使用される。例えば、正規化対照は、マルチアナライト対照内の一つまたは複数の種からのリードの正規化のために使用することができる。したがって、いくつかの実施態様では、正規化対照は、ここに記載されたように処理および配列決定されるマルチアナライト対照に添加される。マルチアナライト対照内の種の濃度（力価）は

40

50

、予め決定されているため、マルチアナライト対照は、ここに記載された正規化対照について陽性対照として使用することができる。あるいは、または付加的に、正規化対照は、マルチアナライト対照内の一つまたは複数の種からのリードを正規化するために使用することができ、正規化されたリードカウントと力価と間の関係は、検量線の生成のために使用される。この検量線は、ここに記載された正規化対照を使用して正規化されたリードカウントからサンプル内の標的生物の力価を決定するために使用することができる。いくつかの実施態様では、正規化対照は、マルチアナライト対照に添加され、マルチアナライト対照は処理され、配列決定され、リードはサンプルと共に並行して正規化されて、検量線が生成される。いくつかの実施態様では、マルチアナライト対照および正規化対照は、参照検量線を生成するために使用される。いくつかの実施態様では、正規化対照およびマルチアナライト対照は両方ともサンプルに添加される（つまりスパイクイン）。いくつかの実施態様では、検量線は、サンプル内の標的生物の力価を計算するために使用される。いくつかの実施態様では、本開示の正規化対照は、マルチアナライト対照からのリードを正規化するために使用される。いくつかの実施態様では、マルチアナライト対照からの正規化されたリードは、サンプル内の一つまたは複数の標的生物の力価を決定するために使用される。

【0364】

本開示は、少なくとも三つの異なる生物種の混合物を含むマルチアナライト対照を提供する。いくつかの実施態様では、マルチアナライト対照は、少なくとも十の生物種を含む。いくつかの実施態様では、マルチアナライト対照は、少なくとも4つの生物種、少なくとも5つの生物種、少なくとも6つの生物種、少なくとも7つの生物種、少なくとも8つの生物種、少なくとも9の生物種、少なくとも10の生物種、少なくとも11の生物種、少なくとも12の生物種、少なくとも13の生物種、少なくとも14の生物種、少なくとも15の生物種、少なくとも16の生物種、少なくとも18の生物種、少なくとも19の生物種、少なくとも20の生物種、少なくとも25の生物種、少なくとも30の生物種、少なくとも35の生物種、少なくとも40の生物種、少なくとも45の生物種、少なくとも50の生物種、少なくとも75の生物種、少なくとも100の生物種、少なくとも150の生物種または少なくとも200の生物種を含む。いくつかの実施態様では、マルチアナライト対照は、本質的に、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39または40の生物種からなる。いくつかの実施態様では、マルチアナライト対照は、本質的に、10の生物種からなる。いくつかの実施態様では、マルチアナライト対照は、本質的に、11の生物種からなる。いくつかの実施態様では、マルチアナライト対照内の生物は、不活性化されている。いくつかの実施態様では、マルチアナライト対照はさらに、許容可能な担体、希釈剤または付形剤を含む。

【0365】

いくつかの実施態様では、マルチアナライト対照は、少なくとも4つのウイルス、少なくとも5つのウイルス、少なくとも6つのウイルス、少なくとも7つのウイルス、少なくとも8つのウイルス、少なくとも9のウイルス、少なくとも10のウイルス、少なくとも11のウイルス、少なくとも12のウイルス、少なくとも13のウイルス、少なくとも14のウイルス、少なくとも15のウイルス、少なくとも16のウイルス、少なくとも18のウイルス、少なくとも19のウイルス、少なくとも20のウイルス、少なくとも25のウイルス、少なくとも30のウイルス、少なくとも35のウイルス、少なくとも40のウイルス、少なくとも45のウイルス、少なくとも50のウイルス、少なくとも75のウイルス、少なくとも100のウイルス、少なくとも150のウイルスまたは少なくとも200のウイルスを含む。いくつかの実施態様では、マルチアナライト対照は、本質的に、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39または40のウイルスからなる。いくつかの実施態様では、マルチアナ

ライト対照は、本質的に、10のウイルスからなる。いくつかの実施態様では、マルチアナライト対照は、本質的に、11のウイルスからなる。いくつかの実施態様では、マルチアナライト対照内の生物は、不活性化されている。いくつかの実施態様では、マルチアナライト対照はさらに、許容可能な担体、希釈剤または付形剤を含む。

【0366】

本開示のマルチアナライト対照のいくつかの実施態様では、マルチアナライト対照内の種の混合物内の各々の種は、同じ力価で存在する。

【0367】

いくつかの実施態様では、各々の種は、マルチアナライト対照内に異なる力価で存在する。いくつかの実施態様では、マルチアナライト対照内の生物種の力価は、予定の所望の比率で存在する。予定の所望の比率が1:1である場合、各々の種は、同じ力価で存在する。別の例では、三つの種を含むマルチアナライト対照内の各々の生物は、1:5:35または1:20:90の比率で存在することができる。力価のあらゆる予定の所望の比率は、本開示の範囲内にあると認識される。いくつかの実施態様では、マルチアナライト対照内の生物種の力価は、線形配列、幾何学配列または対数配列である。例えば、三つの種を含むマルチアナライト対照内の種は、1:2:3または1:10:100の比率で存在することができる。当業者は、ここに記載されたマルチアナライト対照の特別な使用のために適した予定の所望の比率を選択することができる。

【0368】

いくつかの実施態様では、少なくとも二つの種は、マルチアナライト対照内で同じ力価で存在する。いくつかの実施態様、例えば、マルチアナライト対照が少なくとも三つの種を含む実施態様では、少なくとも二つの種は、マルチアナライト対照内で同じ力価で存在し、かつ少なくとも二つの種は、異なる力価で存在する。

【0369】

病原性の種を含む本開示のマルチアナライト対照のいくつかの実施態様では、混合物内の病原体の各々の異なる種は、mL当たり約500~1000万感染単位(IU/mL)の濃度で存在する。いくつかの実施態様では、混合物内の各々の異なる種は、約50~500万IU/mL、約500~500万IU/mL、約500~100万IU/mL、約500~100,000IU/mL、約500~10,000IU/mL、または約1,000~10,000IU/mLの濃度で存在する。いくつかの実施態様では、マルチアナライト対照内の各々の種は、0IU/mL、100IU/mL、1,000IU/mL、5,000IU/mL、10,000IU/mL、100,000IU/mLまたは1,000,000IU/mLで存在する。

【0370】

いくつかの実施態様では、マルチアナライト対照は、サンプルに、100IU/mL~1,000,000IU/mL、100IU/mL~100,000IU/mL、1,000IU/mL~1,000,000IU/mL、1,000IU/mL~100,000IU/mL、100IU/mL~10,000IU/mLまたは1,000IU/mL~10,000IU/mLの最終濃度で添加される。いくつかの実施態様では、マルチアナライト対照は、サンプルに、100IU/mL、150IU/mL、1,000IU/mL、5,000IU/mL、10,000IU/mL、100,000IU/mLの最終濃度で添加される。いくつかの実施態様では、マルチアナライト対照は、多数のサンプルに、例えば0IU/mL、100IU/mL、1,000IU/mL、5,000IU/mL、10,000IU/mL、100,000IU/mLの濃度範囲で添加することができる。更なる実施例として、マルチアナライト対照は、サンプルに、1IU/mL、100IU/mL、150IU/mL、1,000IU/mL、5,000IU/mL、10,000IU/mL、100,000IU/mLの濃度で添加することができる。いくつかの実施態様では、この濃度範囲は、サンプル内の標的生物の力価を測定するための検量線を計算するために使用される。

【0371】

ここで使用されるような「力価」または「物理的力価」は、溶液内の生物の濃度を言う。感染性生物について、「力価」または「感染力価」は、溶液の体積当たりの濃度感染単位を言う（例えば、mLまたは μ L当たりの感染単位、IU）。力価を測定する方法は、当業者に容易に理解される。例えば、物理的力価は、溶液内の生物の濃度を測定することにより測定することができる。これは、計数のような直接的な方法、または例えば問題の生物に特有のタンパク質のレベルを測定することによる間接的な方法を使用して行うことができる。物理的力価は、溶液の体積当たりの細胞、生物、粒子の数または単位（例えば単位/mL）として表すことができる。感染力価を測定する方法は、当業者に公知である。例えば、ウイルスについての感染力価は、プラーク形成アッセイのようなこの分野で通常のアッセイを使用して決定することができる。プラーク形成アッセイでは、ウイルスストックの希釈液を調製し、敏感な細胞単一層上にプレーティングし、ウイルスプラークを形成する感染細胞の数をカウントする。細菌力価は、プレーティングおよび形成されたコロニー数の測定により、カウティング法により、分光光度法により、または当分野で公知の他のあらゆる方法により決定することができる。

【0372】

本開示のマルチアナライト対照のいくつかの実施態様では、マルチアナライト対照内の生物種は、ウイルス、細菌、真菌、真核生物の寄生体またはそれらの組合せの種を含む。

【0373】

いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、ウイルス種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも20、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50、少なくとも75、少なくとも100、少なくとも150または少なくとも200のウイルス種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19または20のウイルス種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、11のウイルス種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも20、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50、少なくとも75、少なくとも100、少なくとも150または少なくとも200のウイルス種を含む。いくつかの実施態様では、マルチアナライト対照内の生物種は、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19または20のウイルス種を含む。いくつかの実施態様では、マルチアナライト対照内の生物種は、11のウイルス種を含む。

【0374】

いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、細菌種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも20、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50、少なくとも75、少なくとも100、少なくとも150または少なくとも200の細菌種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19または20の細菌種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少な

【 0 3 7 5 】

【 0 3 7 6 】

【 0 3 7 7 】

いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、ウイルス種および真菌種からなる。いくつかの実施態様では、マルチアナライト対照内の生物種は、本質的に、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、

10

20

30

40

50

少なくとも１４、少なくとも１５、少なくとも２０、少なくとも２０、少なくとも２５、少なくとも３０、少なくとも３５、少なくとも４０、少なくとも４５、少なくとも５０、少なくとも７５、少なくとも１００、少なくとも１５０または少なくとも２００のウイルス種、細菌種および真菌種を含む。いくつかの実施態様では、マルチアナライト対照内の生物種は、３、４、５、６、７、８、９、１０、１１、１２、１３、１４、１５、１６、１７、１８、１９または２０のウイルス種、細菌種および真菌種を含む。

【０３８０】

いくつかの実施態様では、マルチアナライト対照内の生物種は、一つまたは複数のウイルス種を含む。全てのウイルス種は、本開示のマルチアナライト対照内に含めるのに好適であると認識される。ウイルスは、ＤＮＡウイルス、ＲＮＡウイルスまたはこれらの組合せであることができる。いくつかの実施態様では、ウイルスは、ＤＮＡウイルスまたはＲＮＡウイルスである。いくつかの実施態様では、ＤＮＡウイルスは、二本鎖ＤＮＡウイルスゲノムを含む。いくつかの実施態様では、ＤＮＡウイルスは、一本鎖ＤＮＡウイルスゲノムを含む。いくつかの実施態様では、ＲＮＡウイルスは、二本鎖ＲＮＡウイルスゲノムを含む。いくつかの実施態様では、ＲＮＡウイルスは、一本鎖ＲＮＡウイルスゲノムを含む。いくつかの実施態様では、一本鎖ＲＮＡウイルスゲノムはセンス鎖を含む。いくつかの実施態様では、一本鎖ＲＮＡウイルスゲノムはアンチセンス鎖を含む。いくつかの実施態様では、ウイルスは、ボルティモア分類系により決定してⅠ型、Ⅱ型、Ⅲ型、Ⅳ型、Ⅴ型、Ⅵ型またはⅦ型ウイルスである。

【０３８１】

いくつかの実施態様では、マルチアナライト対照は、ヒト病原体であるウイルス種を含む。例示的であるが非限定的な、ヒト病原体であるウイルスを、以下の表２に示す：

10

20

30

40

50

【表 2 - 1】

表 2 例示的なヒトウイルス病原体

ウイルス	属および科	疾患例
アデノ関連ウイルス	デペンドウイルス、パルボウイルス科	なし
アイチウイルス	コブウイルス、ピコルナウイルス科	胃腸炎
アリューシャン病ウイルス	パルボウイルス	形質細胞増加症
オーストラリアコウモリリッサウイルス	リッサウイルス、ラブドウイルス科	致死性脳炎
BKポリオマウイルス	ポリオマウイルス、ポリマウイルス科	なし
バンナウイルス	シードルナウイルス、レオウイルス科	脳炎
バルマー森林ウイルス	アルファウイルス、トガウイルス科	発熱、関節痛
ブニヤンウェラウイルス	オルソブニヤウイルス、ブニヤウイルス科	脳炎
ブニヤウイルス ラクロス	オルソブニヤウイルス、ブニヤウイルス科	脳炎
ブニヤウイルス カンジキウサギ	オルソブニヤウイルス、ブニヤウイルス科	脳炎
ケルコピテクス ヘルペスウイルス	リンフォクリプトウイルス、ヘルペスウイルス科	脳炎
チャンディプラウイルス	ベシクロウイルス、ラブドウイルス科	脳炎
チクングニアウイルス	アルファウイルス、トガウイルス科	発熱、関節痛
コサウイルス A	コサウイルス、ピコルナウイルス科	—
牛痘ウイルス	オルソポックスウイルス、ポックスウイルス科	なし
コクサッキーウイルス	エンテロウイルス、ピコルナウイルス科	髄膜炎、心筋炎、麻痺
クリミア・コンゴ出血熱ウイルス	ナイロウイルス、ブニヤウイルス科	出血性熱
デングウイルス	フラビウイルス、フラビウイルス科	出血性熱
ドリウイルス	トゴトウイルス、オルソミクソウイルス科	熱、脳炎

10

20

30

40

50

【表 2 - 2】

ウイルス	属および科	疾患例
ダグベウイルス	ナイロウイルス、ブニヤウイルス科	血小板減少症
デュベーンハゲウイルス	リッサウイルス、ラブドウイルス科	致死性脳炎
東部アフリカ脳炎ウイルス	アルファウイルス、トガウイルス科	脳炎
エボラウイルス	エボラウイルス、フィロウイルス科	出血性熱
エコーウイルス	エンテロウイルス、ピコルナウイルス科	風邪
脳心筋炎ウイルス	カーディオウイルス、ピコルナウイルス科	脳炎
エプスタインバーウイルス	リンフォクリプトウイルス、ヘルペスウイルス科	単核球症
ヨーロッパコウモリリッサウイルス	リッサウイルス、ラブドウイルス	致死性脳炎
GBウイルスC/G型肝炎ウイルス	ペギウイルス、フラビウイルス科	なし
ハンタアンウイルス	ハンタウイルス、ブニヤウイルス科	腎または呼吸器症候群
ヘンドラウイルス	ヘニパウイルス、パラミクソウイルス科	脳炎
A型肝炎ウイルス	ヘパトウイルス、ピコルナウイルス科	肝炎
B型肝炎ウイルス	オルソヘパドナウイルス、ヘパドナウイルス科	肝炎
C型肝炎ウイルス	ヘパシーウイルス、フラビウイルス科	肝炎
E型肝炎ウイルス	ヘペウイルス、未分類	肝炎
デルタ型肝炎ウイルス	デルタウイルス、未分類	肝炎
単純ヘルペスウイルス 1	ヘルペスウイルス科	ヘルペス
単純ヘルペスウイルス 2	ヘルペスウイルス科	ヘルペス
ホースボックスウイルス	オルソボックスウイルス、ボックスウイルス科	なし
ヒトアデノウイルス	マスタデノウイルス、アデノウイルス科	呼吸器系、消化器系
ヒトアストロウイルス	マムアストロウイルス、アストロウイルス科	胃腸炎
ヒトコロナウイルス	アルファコロナウイルス、コロナウイルス科	呼吸器

10

20

30

40

50

【表 2 - 3】

ウイルス	属および科	疾患例
ヒトサイトメガロウイルス	サイトメガロウイルス、ヘルペスウイルス科	単核球症、肺炎
ヒトエンテロウイルス 68、70	エンテロウイルス、ピコルナウイルス科	下痢、神経系障害
ヒトヘルペスウイルス 1	シンプレクスウイルス、ヘルペスウイルス科	皮膚病変
ヒトヘルペスウイルス 2	シンプレクスウイルス、ヘルペスウイルス科	皮膚病変
ヒトヘルペスウイルス 6 (AおよびB)	ロゼオロウイルス、ヘルペスウイルス科	皮膚病変
ヒトヘルペスウイルス 7	ロゼオロウイルス、ヘルペスウイルス科	皮膚病変
ヒトヘルペスウイルス 8	ラディノウイルス、ヘルペスウイルス科	皮膚リンパ腫
ヒト免疫不全ウイルス	レンチウイルス、レトロウイルス科	A I D S
ヒトパピローマウイルス 1	ムパピローマウイルス、パピローマウイルス科	皮膚疣贅
ヒトパピローマウイルス 2	アルファパピローマウイルス、パピローマウイルス科	皮膚疣贅
ヒトパピローマウイルス 16、18	アルファパピローマウイルス、パピローマウイルス科	性器疣贅
ヒトパラインフルエンザ	レスピロウイルス、パラミクソウイルス科	呼吸器
ヒトパルボウイルス B 19	エリスロウイルス、パルボウイルス科	皮膚病変
ヒト呼吸器シンシチアルウイルス	ニューモウイルス、パラミクソウイルス科	呼吸器
ヒトライノウイルス	エンテロウイルス、ピコルナウイルス科	呼吸器
ヒト S A R S コロナウイルス	ベータコロナウイルス、コロナウイルス科	呼吸器
ヒトスパーマレトロウイルス	スプーマウイルス、レトロウイルス科	なし
ヒト T リンパ球向性ウイルス	デルタレトロウイルス、レトロウイルス科	白血病
ヒトトロウイルス	トロウイルス、コロナウイルス科	胃腸炎
インフルエンザ A 型ウイルス	インフルエンザウイルス A、オルソミクソウイルス科	インフルエンザ

10

20

30

40

50

【表 2 - 4】

ウイルス	属および科	疾患例
インフルエンザB型ウイルス	インフルエンザウイルス B、オルソミクソウイルス 科	インフルエンザ
インフルエンザC型ウイルス	インフルエンザウイルス C、オルソミクソウイルス 科	インフルエンザ
イスファハンウイルス	ベシクロウイルス、ラブド ウイルス科	脳炎
J Cポリオマウイルス (J Cウイルス、ヒトポリオマ ウイルス 2)	ポリオマウイルス、ポリオ マウイルス科	脳炎
日本脳炎ウイルス	フラビウイルス、フラビウ イルス科	脳炎
ジュニンアレナウイルス	アレナウイルス、アレナウ イルス科	出血性熱
K I ポリオマウイルス	ポリオマウイルス、ポリオ マウイルス科	脳炎
クンジンウイルス	フラビウイルス、フラビウ イルス科	脳炎
ラゴスコウモリウイルス	リッサウイルス、ラブドウ イルス科	致死性脳炎
ビクトリア湖マールブルグ ウイルス	マールブルグウイルス、フ ィロウイルス科	出血性熱
ランガットウイルス	フラビウイルス、フラビウ イルス科	脳炎
ラッサウイルス	アレナウイルス、アレナウ イルス科	出血性熱
ロルドスデールウイルス	ノロウイルス、カリシウイ ルス科	胃腸炎
ルーピングイルウイルス	フラビウイルス、フラビウ イルス科	脳炎
リンパ球性絨毛膜炎ウイル ス	アレナウイルス、アレナウ イルス科	脳炎
マチュポウイルス	アレナウイルス、アレナウ イルス科	脳炎
マヤロウイルス	アルファウイルス、トガウ イルス科	発熱、関節痛
MERS コロナウイルス	ベータコロナウイルス、コ ロナウイルス科	呼吸器
麻疹ウイルス	モルビリウイルス、パラミ クソウイルス科	発熱、発疹

10

20

30

40

50

【表 2 - 5】

ウイルス	属および科	疾患例
メング脳筋心筋炎ウイルス	カーディオウイルス、ピコルナウイルス科	脳炎
メルケル細胞ポリオマウイルス	ポリオマウイルス、ポリオマウイルス科	メルケル細胞癌
モコラウイルス	リッサウイルス、ラブドウイルス科	脳炎
伝染性軟属腫ウイルス	モルシボックスウイルス、ボックスウイルス科	皮膚病変
モンキーボックスウイルス	オルソボックスウイルス、ボックスウイルス科	皮膚病変
ムンプスウイルス	ルブラウイルス、パラミクソウイルス科	流行性耳下腺炎
マレーバレー脳炎ウイルス	フラビウイルス、フラビウイルス科	脳炎
ニューヨークウイルス	ハンタウイルス、ブニヤウイルス	出血性熱
ニパウイルス	ヘニパウイルス、パラミクソウイルス科	脳炎
ノーウォークウイルス	ノロウイルス、カリシウイルス科	胃腸炎
オニオンオニオンウイルス	アルファウイルス、トガウイルス科	発熱、関節痛
オルフウイルス	パラボックスウイルス、ボックスウイルス科	皮膚病変
オロポーシェウイルス	オルソブニヤウイルス、ブニヤウイルス科	発熱、関節痛
ピシンデウイルス	アレナウイルス、アレナウイルス科	出血性熱
ポリオウイルス	エンテロウイルス、ピコルナウイルス科	灰白髄炎
プンタトロフレボウイルス	フレボウイルス、ブニヤウイルス科	出血性熱
プーマラウイルス	ハンタウイルス、ブニヤウイルス	出血性熱
狂犬病ウイルス	リッサウイルス、ラブドウイルス科	致死性脳炎
リフトバレー熱ウイルス	フレボウイルス、ブニヤウイルス科	出血性熱
ロザウイルスA	ロザウイルス、ピコルナウイルス科	

10

20

30

40

50

【表 2 - 6】

ウイルス	属および科	疾患例
ロスリバーウイルス	アルファウイルス、トガウイルス科	発熱、関節痛
ロタウイルス A	ロタウイルス、レオウイルス科	胃腸炎
ロタウイルス B	ロタウイルス、レオウイルス科	胃腸炎
ロタウイルス C	ロタウイルス、レオウイルス科	胃腸炎
風疹ウイルス	ルビウイルス、トガウイルス科	風疹
サギヤマウイルス	アルファウイルス、トガウイルス科	発熱、関節痛
サリウイルス A	サリウイルス、ピコルナウイルス科	胃腸炎
サシチョウバエ熱シチリアウイルス	フレボウイルス、ブニヤウイルス科	出血性熱
サッポロウイルス	サポウイルス、カリシウイルス科	胃腸炎
セムリキ森林ウイルス	アルファウイルス、トガウイルス科	発熱、関節痛
ソウルウイルス	ハンタウイルス、ブニヤウイルス	出血性熱
シミアンフォアミーウイルス	スプーマウイルス、レトロウイルス科	なし
シミアンウィルス 5	ルブラウイルス、パラミクソウイルス科	未証明
シンドビスウイルス	アルファウイルス、トガウイルス科	発熱、関節痛
サザンプトンウイルス	ノロウイルス、カリシウイルス科	胃腸炎
セントルイス脳炎ウイルス	フラビウイルス、フラビウイルス科	脳炎
ダニ媒介性ポワッサンウイルス	フラビウイルス、フラビウイルス科	脳炎
トルクテノウイルス	アルファトルケウイルス、アネロウイルス科	なし
トスカーナウイルス	フレボウイルス、ブニヤウイルス科	出血性熱
ウクニエミウイルス	フレボウイルス、ブニヤウイルス科	出血性熱

10

20

30

40

50

【表 2 - 7】

ウイルス	属および科	疾患例
ワクシニアウイルス	オルソポックスウイルス、 ポックスウイルス科	なし
水痘帯状疱疹ウイルス	バリセロウイルス、ヘルペ スウイルス科	水痘
天然痘ウイルス	オルソポックスウイルス、 ポックスウイルス科	痘瘡
ベネズエラ馬脳炎ウイルス	アルファウイルス、トガウ イルス科	発熱、関節痛
水疱性口内炎ウイルス	ベシクロウイルス、ラブド ウイルス科	脳炎
西部ウマ脳炎ウイルス	アルファウイルス、トガウ イルス科	発熱、関節痛
WUポリオマウイルス	ポリオマウイルス、ポリオ マウイルス科	なし
西ナイルウイルス	フラビウイルス、フラビウ イルス科	出血性熱
ヤバサル腫瘍ウイルス	オルソポックスウイルス、 ポックスウイルス科	なし
ヤバ様病ウイルス	オルソポックスウイルス、 ポックスウイルス科	なし
黄熱病ウイルス	フラビウイルス、フラビウ イルス科	出血性熱
ジカウイルス	フラビウイルス、フラビウ イルス科	発熱、関節痛、発疹

【0382】

いくつかの実施態様では、マルチアナライト対照内の生物種は、一つまたは複数の細菌種を含む。全ての細菌種は、本開示のマルチアナライト対照内に含めるのに好適であると認識される。いくつかの実施態様では、細菌は、グラム陽性細菌またはグラム陰性細菌である。

【0383】

いくつかの実施態様では、マルチアナライト対照は、ヒト病原体である細菌種を含む。例示的であるが非限定的な、ヒト病原体である細菌を、以下の表 3 に示す：

10

20

30

40

50

【表 3 - 1】

表 3 例示的なヒト細菌病原体

疾患	細菌種
アクネ	プロピオニバクテリウム アクネ
炭疽	バシラス アントラシス
細菌性血管腫／肝紫斑病	バルトネラ ヘンセラ、バルトネラ・キンタナ
細菌性敗血症	ストレプトコッカス アガラクチエ（グループ b strep.）、エシェリキア コリ、ストレプトコッカス ニューモニエ
細菌性膣炎	ガードネレラ バジナリス、アトポビウム バギナエ、マイコプラズマ ホミニス、モビルンカス属を含むさまざまな嫌気性細菌、プレボテラ属
バランチジウム症	バランティディウム コリ
腫れ	スタフィロコッカス属
ボツリヌス中毒	クロストリジウム ボツリヌム
気管支炎	マイコプラズマ ニューモニア、クラミドフィア ニューモニエ、S. ピオゲネス
ブルセラ症	ブルセラ属、Y. ペステリス
腺ペスト	エルシニア ペステリス
ブルーリ潰瘍	マイコバクテリウム ウルセランス
ネコひっかき病	バルトネラ ヘンセラ
蜂巣炎	ストレプトコッカスおよびストレプトコッカス属
軟性下疳	ヘモフィルス デュクレイ
クラミジア	クラミジア トラコマチス
コレラ	ビブリオ コレラ
クロストリジウム ソルデリイ感染症	クロストリジウム ソルデリイ
下痢性疾患	カンピロバクター ジェジュニ
エールリヒア症	エールリヒア カフェエンシス、E. エウインギイ、E. ムリス エウクレイレンシス
疫病性再発性熱ー	ボレリア属（15の異なる種；B. ヘルムシイ、B. パルケリを含む）
フラビモナス オリジハビタンスの感染	フラビモナス オリジハビタンス
毛嚢炎	スタフィロコッカス アウレウス；スタフィロコッカス属
食中毒	エシェリキア コリ、スタフィロコッカス アウレウス、バチルス セレウス、クロストリジウム

10

20

30

40

50

【表 3 - 2】

疾患	細菌種
	ムパープリンジェス、クロストリジウム ボツリヌム
鼻疽	バークホルデリア マレイ
淋病	ナイセリア ゴノロエエ
鼠径部肉芽腫	クレブシエラ グラヌロマチス
B群連鎖球菌症	グループBストレプトコッカス (GBS)、またはストレプトコッカス アガラクチェ
ヘモフィルス インフルエンゼ	ヘモフィルス インフルエンゼ
ハンセン病	マイコバクテリウム レブレ
医療関連感染症	アシネトバクター バウマンニイ、バクテロイデス フラジリス、バークホルデリア セパシア、クロストリジウム ディフィシル、クロストリジウム ソルデリイ、カルバペネム耐性腸内細菌科、エンテロコッカス フェカリス、クレブシエラ ニューモニエ、メチシリン耐性スタフィロコッカス アウレウス、モルガネラ モルガニー、マイコバクテリウム アブセッサス、シュードモナス アエルギノーザ、スタフィロコッカス アウレウス、ステノトロホモナス マルトフィリア、マイコバクレイウム ツベルクローシス、バンコマイシン耐性スタフィロコッカス アウレウス、バンコマイシン耐性腸球菌、ストレプトコッカス ニューモニエ
ヒト顆粒球エーリキア症	アナプラズマ ファゴサイトフィラム
ヒト単球性エーリキア症	エールリヒア シャフェンシス
膿痂疹	スタフィロコッカス属
腸管感染症	カンピロバクター ジェジュニ、エシェリキア コリ (EIEC)、エシェリキア コリ (EHEC)、サルモネラ タイフィウム、サルモネラ タイピイ、シゲラ ディセンテリエタイプ1、シゲラ ソンネイ/フレキシネリ、エルシニア エンテロコリチカ、クロストリジウム ディフィシル、ビブリオ コレラ
クレブシエラ感染症	クレブシエラ属
レジオネラ症	レジオネラ属、レジオネラ ニューモフィラ
レミエール症候群	フソバクテリウム ネクロフォルム
リステリア感染症	リステリア モノサイトゲネス
ライム病	ボレリア ブルグドルフェリ
性病性リンパ肉芽腫	クラミジア トラコマチス
類鼻疽	バークホルデリア シュードマレイ

10

20

30

40

50

【表 3 - 3】

疾患	細菌種
髄膜炎	マイコバクテリウム ツベルクローシス、ストレプトコッカス アガラクチエ（グループb ストレプトコッカス）、エシェリキア コリ、リステリア モノサイトゲネス、ストレプトコッカス アガラクチエ、エシェリキア コリ、ストレプトコッカス ニューモニエ、ヘモフィルス インフルエンゼ
菌腫	アクチノミセス属、アクチノマドゥラ ペレティエル、アクチノマドゥラ マデュラエ
マイコバクテリア感染症	マイコバクテリウム アブセッサス、M. アビウム、M. イントラセルラーレ、M. シェロネア、M. フォルチュイタム、M. ゴルドナエ、M. カンサシイ、M. マルモエンセ、M. マリナム、M. ツベルクローシス、M. キセノピ
壊死性筋膜炎	グループAストレプトコッカスおよび他の属；S. アウレウス、S. ピオゲネス、クロストリジウム属
非結核性マイコバクテリア肺疾患	マイコバクテリア属
骨髄炎	スタフィロコッカス アウレウス、ストレプトコッカス属、腸内細菌科
パスツレラ マルトシダ感染症	パスツレラ マルトシダ
心膜炎	スタフィロコッカス アウレウス、ストレプトコッカス ニューモニエおよび他の連鎖球菌、マイコバクテリウム ツベルクローシス
熱帯白斑性皮膚病	トレボネマ カラテウム
筋炎	スタフィロコッカス アウレウス
Q熱	コクシエラ バーネッティ
リウマチ熱	グループAストレプトコッカス細菌
ロッキー山紅斑熱	リケッチア
サルモネラ	サルモネラ タイフィムリウム
セネツ熱	ノエリケッチア センネツ
細菌性赤痢	シゲラ属
ブドウ球菌食中毒	スタフィロコッカス属
ステノトロホモナス マルトフィリア感染症	ステノトロホモナス マルトフィリア
ストレプトコッカス グループA侵襲性疾患	ストレプトコッカス グループA
ストレプトコッカス グループB侵襲性疾患	ストレプトコッカス グループB

10

20

30

40

50

【表 3 - 4】

疾患	細菌種
梅毒	トレポネマ パリ ドウム
破傷風	クロストリジウム テタニ、クロストリジウム テタナス
結核	マイコバクテリウム ツベルクローシス
野兔病	フランシセラ チュラレンシス
チフス	リケッチア プロワゼキイ、オリエンティア ツ ツガムシ、リケッチア タイピイ
潰瘍	ヘリコバクター ピロリ
尿道炎、子宮頸管炎	クラミジア トラコマチス、ナイセリア ゴノロ エエ
ビブリオ感染症	ビブリオ属、V. パラエモリディクス、V. ブルニフィクス、V. アルギノリディクス
ウォーターハウス フリデリクセン 症候群	髄膜炎菌属
ウィップル病	トロフェリマ ウィップレイ
百日咳	ボルデテラ ペルツッシス
イチゴ腫	トレポネマ ペルテヌエ

10

20

【 0 3 8 4 】

マルチアナライト対照内に含まれることができる、代表的な片利共生種、共生種、病原性種および薬剤耐性種、ならびにメタゲノミックサンプル内に見られる種を含めた付加的細菌種を、以下の表 4 に示す：

30

40

50

【表 4－1】

表 4 例示的な付加的細菌種

細菌の属および種
アシネトバクター バウマンニイ
アシネトバクター ジョンソニイ
アシネトバクター属
アクチノミセス オドントリテिकास
アロバシラス ハロトレランス
バシラス セレウス
バシラス スブチリス
バクテリオデス フラジリス
バクテロイデス バルガタス
ビフィドバクテリウム アドレッセンチス
クロモバクター ビオラセウム
クロストリジオイデス ディフィシレ
クロストリジウム ベイジェリンキ
クロストリジウム ペルフリンゲンズ
コアグラ－ゼ陰性ブドウ球菌
コリネバクテリウム ストリアタム
キューティバクテリウム アクネス
デイノコッカス ラジオデュランス
エンテロコッカス フェカーリス
エンテロバクター クロアカエ
エンテロバクター属
エシェリキア コリ

細菌の属および種
ラクトバチルス ガセリ
ラクトバチルス ジェンセニイ
ラクトバチルス プランタラム
リステリア モノサイトゲネス
ミクロコッカス ルテウス
マイコプラズマ ホミニス
ナイセリア メニンギティディス
ポルフィロモナス ジンジバリス
プレボテラ ビビア
プレボテラ メラニノゲニカ
プロピオニバクテリウム アクネス
プロテウス ミラビリス
プロビデンシア レットゲリ
シュードアルテロモナス ハロプラנקティス
シュードモナス アエルギノーサ
シュードモナス フルオレッセン
シュードモナス属
ロドバクター スフェロイデス
サルモネラ エンテリカ
セラチア リクイファシエン
セラチア マルセッセンス
セラチア属

10

20

30

40

50

【表 4－2】

フソバクテリウム ヌクレアタム	スタフィロコッカス アウレウス
ガードネレラ バジナリス	スタフィロコッカス アウレウス (MRSA、MSSA)
ヘモフィルス パラインフルエンザ	スタフィロコッカス エピデルミディス
ハロバシルス ハロフィルス	ストレプトコッカス アガラクチエ
ハロフェラックス ボルカニイ	ストレプトコッカス ミディス
ヘリコバクター ピロリ	ストレプトコッカス ミュータンス
イムテセラ ハロトレランス	ストレプトコッカス ニューモニエ
クレブシエラ ニューモニアエ	ストレプトコッカス ピオゲネス
クレブシエラ属	ストレプトコッカス属
ラクトバチルス フェルメンタム	ベイルロネラ パルブラ
	エルシニア エンテロコリチカ

【0385】

いくつかの実施態様では、マルチアナライト対照は、一つまたは複数の真核生物の寄生体種を含む。いくつかの実施態様では、寄生体はヒト寄生体である。いくつかの実施態様では、真核生物の寄生体は、多細胞寄生体（例えば、線形動物の種）である。いくつかの実施態様では、真核生物の寄生体は、単細胞生物である。いくつかの実施態様では、真核生物の寄生体は、酵母または真菌である。いくつかの実施態様では、真核生物の寄生体は、酵母または真菌に分類されない種（例えばアメーバ）である。

【0386】

いくつかの実施態様では、マルチアナライト対照は、ヒト病原体である一つまたは複数の真菌種を含む。例示的であるが非限定的な、ヒト病原体である真菌を、以下の表 5 に示す：

10

20

30

40

50

【表 5】

表 5 例示的ヒト真菌性病原体

真菌種	疾患
アスペルギルス属、A. フミガタス、A. フラバス、A. クラバタス他	アスペルギルス症
トリコフィトン、エピデルモフィトン、マイクロスポルム、他の属	水虫
ブラストミセス属	ブラストミセス症
クリプトコッカス ガッティ	C. ガッティ感染症
クリプトコッカス ネオフォルマンس	C. ネオフォルマンス感染症
クリプトコッカス ネオフォルマンス	C. ネオフォルマンス感染症
カンジダアルビカンス、C. アウリス、および他のカンジダ属	カンジダ症
カンジダおよびアスペルギルス属	心内膜炎
カンジダ属、他の属	真菌性眼感染症
多様な属	真菌性爪感染症
ヒストプラズマ属、ヒストプラズマ カプスラタム	ヒストプラズマ症
クリプトコッカス ネオフォルマンス、C. ローレンティ、C. アルビドゥス	髄膜炎
ムコルミセテス	ムーコル症
アクレモニウム ストリクタム、アスペルギルス ニドゥランス、ノエテストディナ ロザチイ、ファエオアクレモニウム カラジデニイ、シュードアレスケリア ボイジイ、アスペルギルス テレウス、クルブラリアルナータ、クラドフィアロフォラ バンチアナ、エキソフィアラ ジェンセルメイ、レプトスフェリア セネガレンシス、レプトスファエリア トンピキンシイ、マズレラ グリセア、マズレラ ミセトマチス、ピレノカエタ ロメロイ、他の種	菌腫
ニューモシスチス ジロベシイ	ニューモシスチス肺炎
ニューモシスチス ジロベシイ	肺炎
スタキボトリス カルタルム	呼吸器系の問題
トリコフィトン ルブルム、トリコフィトン トンズランス、トリコフィトン インタージギターレ、および/またはトリコフィトン メンタグロフィテス、マイクロスポルム カニス、エピデルモフィトン フロッコスム	白癬
スポロトリクス	スポロトリコーシス
タラロマイシス マルネファイ	タラロマイコシス
コクシジオイデス属	バレー熱

【0387】

本開示のマルチアナライト対照内に含めることができる付加的真菌種は、クリプトコッカス ネオフォルマンスおよびサッカロミセス セレビスエを含むが、これらに限定されるものではない。

【0388】

いくつかの実施態様では、マルチアナライト対照は、真核生物の寄生体種を含む。例示的であるが非限定的な真核生物の寄生体を、以下の表 6 に示す：

10

20

30

40

50

【表 6】

表 6 例示的なヒトの真核生物の寄生体

真核生物の寄生体	疾患
マンソン住血吸虫、ビルハルツ住血吸虫、日本住血吸虫、メコン住血吸虫、インターカラタム住血吸虫	住血吸虫症
フィラリア属、バンクロフト糸状虫、マレー糸状虫、チモール糸状虫	フィリアリア症
トリパノソーマ ブルセイ	トリパノソーマ症
リーシュマニア マコール	リーシュマニア症
エントアメーバ ヒストリチカ	アメーバ症
バベシア ボビス、B. ダイバージェンス	バベシア症
トリパノソーマ クルジ	シャーガス病
ランブル鞭毛虫	ジアルジア症
アカントアメーバ カルベルツォーニ	肉芽腫性アメーバ状脳炎
アメリカ鉤虫	鉤虫感染症
マレー糸状虫	リンパ系フィラリア症
ファルシパラム原虫、ピバックス原虫	マラリア
トキソプラズマ原虫	トキソプラズマ症
トリコモナス ヴァギナリス	トリコモナス症

【 0 3 8 9 】

例示的な実施態様では、表 7（2016 年 2 月 18 日に wwwn.cdc.gov/nndss/conditions/notifiable/2019/ でアクセスされる CDC の「Summary of Notifiable Diseases - United States, 2019」から適応）は、ここで提供された方法および組成物を使用して定量化することができる、更なる例示的な疾患を引き起こす病原体を提供する。

【表 7 - 1】

表 7 例示的な疾患の原因となる病原体

炭疽菌病原体	リステリア症
アルボウイルス病原体、神経侵襲性と非神経侵襲性	ルジョウイルス
バベシア症	ライム病
ボツリヌス中毒病原体	マラリア
ブルセラ症病原体	マールブルグウイルス
カリフォルニア血清型ウイルス	麻疹
カリフォルニア血清型ウイルス病	髄膜炎菌性疾患
カンピロベリア症	おたふく風邪
カンジダアウリス	新世界アレナウイルス
軟性下痢	新型インフルエンザ A ウイルス感染症
チクングニアウイルス病	百日咳
クラミジアトラコマチス感染症	ペスト
コレラ	灰白髄炎、麻痺
コレラ	ポリオウイルス感染、非麻痺性
コクシジオイデス症	ポワッサンウイルス
クリミアーコンゴ出血熱ウイルス	オウム病
クリプトスポリジウム症	Q 熱
サイクロスポラ症	狂犬病
デング熱	風疹
デング出血熱	風疹、先天性症候群
デングショック症候群	サルモネラ症
デングウイルス感染症	重症急性呼吸器症候群関連コロナウイルス (SARS-CoV) 疾患
ジフテリア	志賀毒素産生性大腸菌 (STEC)
東部馬脳炎ウイルス	細菌性赤痢
エボラウイルス	天然痘
エールリヒア症/アナプラズマ病	紅斑熱リケッチア症
食中毒病原体	セントルイス脳炎ウイルス
ジアルジア症	連鎖球菌毒素性ショック症候群
淋病	ストレプトコッカス ニューモニエ、侵襲性疾患
ヘモフィルス インフルエンザ菌、侵襲性疾患	梅毒
ハンセン病 (らい病)	梅毒、先天性
ハンタウイルス肺症候群	破傷風
溶血性尿毒症症候群、下痢後遺症	毒素性ショック症候群 (連鎖球菌以外)

10

20

30

40

50

【表 7 - 2】

A型肝炎、急性	旋毛虫症
B型肝炎ウイルス、周産期感染	結核
B型肝炎、急性	野兔病
B型肝炎、慢性	腸チフス
C型肝炎、急性	バンコマイシン中程度（VISA）感染症
C型肝炎、慢性	バンコマイシン耐性（VRSA）感染症
肝炎、ウイルス	水痘
ヒト免疫不全ウイルス（HIV）	ビブリオ症
インフルエンザ関連の小児死亡率	ウイルス性出血熱
ラッサウイルス	西ナイルウイルス
レジオネラ症	西部ウマ脳炎ウイルス
レプトスピラ症	黄熱
	ジカウイルス

10

20

【0390】

本開示のマルチアナライト対照のいくつかの実施態様では、マルチアナライト対照内の異なる生物種は、ヒト病原体である。ヒト病原体の例示的な種は、表2～3および5～7に列挙されたあらゆる種が含まれる。マルチアナライト対照内に既知のヒト病原体を使用することで、マルチアナライト対照は、本開示の方法を使用して力価が決定されるサンプル内の標的種の挙動をモデル化することを可能にする。核酸抽出の有効性、病原体ゲノムサイズ、ライブラリー調製の効率、配列リードの数、およびサンプルの組成の全ては、ハイスループットシーケンシングライブラリーからのリードカウントがサンプル内の標的生物の初期力価をどの程度反映するかに影響を及ぼすことができる。サンプル内の標的種に類似する種（例えば同属）または同一の種を含むマルチアナライト対照は、マルチアナライト対照が、正規化されたリードがサンプル内の力価を反映する精度に影響を及ぼすパラメーターについての効果的な制御を可能にする。いくつかの実施態様では、マルチアナライト対照は、サンプル内の標的生物の一つまたは複数の種を含む。標的生物は、サンプル内にあることが既知であるか、またはサンプル内にある疑いがあることができる（例えば、マルチアナライト対照は、一般的なヒト病原体のパネルを含む）。

30

【0391】

いくつかの実施態様では、ヒト病原体は、血液または組織移植において使用される組織内に一般的に見られるヒト病原体を含む。移植後の後天的感染症は、移植レシピエントにおいて罹患率および死亡率の大きな原因である。したがって、本開示のマルチアナライト対照は、一つまたは複数のヒト病原体の存在について移植組織および血液をスクリーニングする際に陽性対照として使用することができ、それにより、移植組織または血液からの感染のリスクを軽減する。ヒトの血液または組織移植において一般的に見られる例示的な病原体は、サイトメガロウイルス（CMV）、エプスタインバーウイルス（EBV）、アデノウイルス（ADV）、BKウイルス（BKV）、JCポリマウイルス（JCV）、ヒトヘルペスウイルス6A（HHV6A）、ヒトヘルペスウイルス6B（HHV6B）、単純ヘルペスウイルス1型（HSV1）、単純ヘルペスウイルス2型（HSV2）、水痘-帯状疱疹ウイルス（VZV）およびヒトパルボウイルスB19（B19）、ヒト免疫不全ウイルス（HIV）、B型肝炎、C型肝炎、ヒトTリンパ球向性ウイルスI型およびII型（HTLV）、西ナイルウイルス、ジカウイルス、ストレプトコッカス種、スタフィロコッカス アウレウス、エンテロコッカス フェカーリス、バシラス セレウス、プロピオニ

40

50

バクテリウム アクネス、セラチア リキファシエンシス、セラチア マルセスセン、エルシニア エンテロコリチカ、エンテロバクター属、アシネトバクター属、シュードモナス属、E. コリ、クレブシエラ ニューモニエ、プロテウス ミラビリス、A. バウマニ、バシラス セレウス、コアグラールゼ陰性ブドウ球菌、ストレプトコッカス属、クレブシエラ属、セラチア属、プロビデンシア レットゲリ、レボネーマ パリズム、トリパノソーマ クルージおよびバベシア ミクロチを含むが、これらに限定されるものではない。

【0392】

いくつかの実施態様では、マルチアナライト対照は、11のウイルス種の混合物を含むかまたは本質的にこの混合物からなる。例示的なマルチアナライト対照は、サイトメガロウイルス (CMV)、エプスタインバーウイルス (EBV)、アデノウイルス (ADV)、BKウイルス (BKV)、JCポリマウイルス (JCV)、ヒトヘルペスウイルス6A (HHV6A)、ヒトヘルペスウイルス6B (HHV6B)、単純ヘルペスウイルス1型 (HSV1)、単純ヘルペスウイルス2型 (HSV2)、水痘-帯状疱疹ウイルス (VZV) およびヒトパルボウイルスB19 (B19) を含むかまたは本質的にこれらからなる。いくつかの実施態様では、マルチアナライト対照は、10のウイルス種の混合物を含むかまたは本質的にこの混合物からなる。例示的なマルチアナライト対照は、CMV (例えばAD169株)、EBV、ADV (例えば1型)、BKV (例えばサブタイプ1b-2)、JCV (例えば1a型)、HHV6A (例えばGS株) HHV6B (例えばZ-29株)、HSV1 (例えば95株)、HSV2 (例えば09株) およびVZV (例えば9/84株) を含むか、または本質的にこれらからなる。いくつかの実施態様では、マルチアナライト対照は、CMV、EBV、ADV、BKV、JCV、HHV6A、HHV6B、HSV1、HSV2 およびVZVを含むか、または本質的にこれらからなる。更なる例示的なマルチアナライト対照は、麻疹ウイルス、西ナイルウイルス、ジカウイルス、黄熱ウイルス、シンドビスウイルス、痘瘡ウイルス、ノーウォークウイルス、狂犬病ウイルスおよびヒトライノウイルスを含むか、または本質的にこれらからなる。また更なる例示的なマルチアナライト対照は、エプスタインバーウイルス、ヒトサイトメガロウイルス、ヒト呼吸器シンシチアルウイルス、バリオラウイルス、インフルエンザウイルス、風疹ウイルス、ムンプスウイルスおよびヒトSARSコロナウイルスを含むか、または本質的にこれらからなる。

【0393】

本開示のマルチアナライト対照を製造することができる生物の例示的な混合物を、以下の表8に示す：

10

20

30

40

50

【表 8】

表 8 例示的なマルチアナライト対照

例示的なマルチ アナライト対照	種
ウイルスパネル	サイトメガロウイルス (CMV)、エプスタインバーウイルス (EBV)、アデノウイルス (ADV)、BKウイルス (BKV)、JCポリオマウイルス (JCV)、ヒトヘルペスウイルス 6A (HHV 6A)、ヒトヘルペスウイルス 6B (HHV 6B)、単純ヘルペスウイルス 1 型 (HSV 1)、単純ヘルペスウイルス 2 型 (HSV 2)、水痘帯状疱疹ウイルス (VZV)、ヒトパルボウイルス B 19 (B 19)
ウイルスパネル	CMV、B 型肝炎、C 型肝炎、ワクシニアウイルス、風疹ウイルス、ロタウイルス A、ロタウイルス B、ムンプスウイルス、麻疹ウイルス、インフルエンザ A、インフルエンザ B、インフルエンザ C、ヒトヘルペスウイルス 6 型
ウイルスパネル	エボラウイルス、バリオラ、黄熱病ウイルス、狂犬病ウイルス、マチュポウイルス、ラッサウイルス、HIV、デングウイルス、チクングニアウイルス
ウイルスパネル	CMV、EBV、ADV、BKV、JCV、HHV 6A、HHV 6B、HSV 1、HSV 2 および VZV
ウイルス+細菌パネル	CMV、EBV、ADV、BKV、JCV、HHV 6A、HHV 6B、HSV 1、HSV 2、VZV、B 19、スタフィロコッカス Aureus、エンテロコッカス フェカリス、バシラス セレウス、プロピオニバクテリウム アクネス、セラチア リクイ ファシエンス、セラチア マルセッセンス、エルシニア エンテロ コリチカ
ウイルス、細菌および真菌パネル	CMV、EBV、ADV、BKV、JCV、HHV 6A、HHV 6B、HSV 1、HSV 2、VZV、B 19、スタフィロコッカス Aureus、エンテロコッカス フェカリス、バシラス セレウス、プロピオニバクテリウム アクネス、セラチア リクイ ファシエンス、セラチア マルセッセンス、エルシニア エンテロ コリチカ、カンジダ アルビカンス、アスペルギルス フミガタス、クリプトコッカス ネオフォルマンس
多生物パネル	CMV、EBV、ADV、BKV、JCV、HHV 6A、HHV 6B、HSV 1、HSV 2、VZV、B 19、スタフィロコッカス Aureus、エンテロコッカス フェカリス、バシラス セレウス、プロピオニバクテリウム アクネス、セラチア リクイ ファシエンス、セラチア マルセッセンス、エルシニア エンテロ コリチカ、カンジダ アルビカンス、アスペルギルス フミガタス、クリプトコッカス ネオフォルマンス、トリパノソーマ ブル チェイ、リーシュマニア マコール、ジアルジア ランブリア

【0394】

本開示のマルチアナライト対照のいくつかの実施態様では、マルチアナライト対照内の少なくとも三つの異なる生物種は、ヒト病原体ではない。いくつかの実施態様では、マルチアナライト対照内の少なくとも三つの異なる生物種は、ヒト病原体ではないが、ヒト病原体である種と同じ属内にある。同じ属内の種は、しばしば同様の物理特性を有し、例えば、同じ属内の細菌は、同様の細胞壁を有してよく、ウイルスは、同様のウイルスタンパク質を有してよい。同じ属内の種は、ここに記載されたサンプル処理方法に供する場合、同様に挙動することを予想することができる。マルチアナライト対照内の種を同じ属のヒト病原体、例えば、移植組織または血液内に一般的に見られるヒト病原体とマッ

チングすることにより、マルチアナライト対照は、サンプル処理の間にサンプル内のそれらのヒト病原性種の挙動を模倣してよい。これは、本開示のマルチアナライト対照および正規化対照を使用して、サンプル内の一つまたは複数の標的生物、例えばヒト臨床サンプル内の病原体の力価を計算するために、本開示の方法の優れた能力を得ることができる。

【0395】

本開示のマルチアナライト対照のいくつかの実施態様では、マルチアナライト対照内の種は、ヒト病原体である種と、ヒト病原体ではない種との混合物を含む。

【0396】

いくつかの実施態様では、マルチアナライト対照はさらに、血漿を含む。血漿は、あらゆる脊椎動物に由来することができる。いくつかの実施態様では、血漿は、ヒト、非ヒト霊長類、マウス、ラット、ウサギ、イヌ、ネコ、アレチネズミ、ヒツジ、ブタ、ウマ、ウシまたは口バに由来する。いくつかの実施態様では、血漿はヒト血漿である。いくつかの実施態様では、血漿は、合成により製造される。

【0397】

ここで提供されるのは、本開示のマルチアナライト対照のいずれか一つから抽出された核酸のコレクションである。ここに記載されたマルチアナライト対照から抽出することができる核酸の抽出方法は、当業者に容易に理解される。例示的方法は、フェノール／クロロホルム抽出、引き続きエタノールまたは同様の適切な溶媒を用いた沈殿、およびQiagenおよびZymo mini-prepキットのような市場で入手可能なキットを含むが、これらに限定されるものではない。マルチアナライト対照から抽出される核酸は、ここに記載された実施態様のいずれかで使用することができる。例えば、マルチアナライト対照から抽出される核酸は、ライブラリー調製ステップでサンプルに添加するか、またはサンプルに並行して処理および配列決定して、陽性対照としてもしくは検量線を生成するために使用することができる。

マルチアナライト対照の製造方法

【0398】

本開示は、ここに記載されたマルチアナライト対照の製造方法を提供する。いくつかの実施態様では、この方法は、(a) マルチアナライト対照内に含まれる生物のパネルを選択すること；(b) 適切な培養条件下で各生物を培養すること；(c) 各生物を不活性化することおよび(d) 不活性化された各生物の予定量を、適切な担体、希釈剤または付形剤と混合して、マルチアナライト対照を製造することを含む。

【0399】

マルチアナライト対照内の生物を不活性化する適切な方法は、当業者に公知である。例えば、多くのウイルスは、乾燥、熱、蒸気、または4.0未満のpHのような低いpHで不活性化することができる。エンベロープを備えたウイルスは、ウイルスエンベロープを破壊する溶媒または界面活性剤の添加により不活性化することができる。ウイルスは、ホルムアルデヒドのような化学的架橋剤に曝すことにより不活性化することもできる。細菌を不活性化する方法は、紫外線、放射線、熱、低いPH、酸化エチレン、ホルムアルデヒドおよびチオシアン酸グアニジウムのような化学薬品による処理を含む。アメーバまたは酵母のような単細胞真核生物は、熱、またはホルムアルデヒドのような化学的架橋剤の適用により不活性化することができる。

【0400】

いくつかの実施態様では、生物は、化学的処理、熱処理、pH処理または紫外線照射により不活性化される。いくつかの実施態様では、化学的処理は、ホルムアルデヒドを含む。

【0401】

いくつかの実施態様では、マルチアナライト対照内の不活性化された生物の既知量が血漿に添加される。血漿は、例えば、遠心分離および／または濾過により血球、血小板および他の細胞成分を除去することにより、血液から単離することができる。血漿は、あらゆる脊椎動物に由来することができる。いくつかの実施態様では、血漿は、ヒト、非ヒト霊長類、マウス、ラット、ウサギ、アレチネズミ、イヌ、ネコ、ウシ、ヒツジ、ブタ、ウマ

または口バに由来する。いくつかの実施態様では、血漿はヒト由来である。不活性化された生物を血漿に添加することで、マルチアナライト対照は、サンプル処理の間に、ヒト血漿を含むヒト臨床サンプルのような実験サンプルの挙動を厳密に模倣することができる。

マルチアナライト対照の使用方法

【0402】

本開示は、ここに記載されたマルチアナライト対照を使用する方法を提供する。いくつかの実施態様では、本開示のマルチアナライト対照は、サンプル、例えば臨床サンプルまたは実験サンプルと並行して処理することができる。サンプルと並行して処理されるマルチアナライト対照は、生物の混合物を含むサンプル内の一つまたは複数の生物の力価を決定するために当業者が普通に使用する検量線を生成するために使用することができる。例えば、サンプルは、ヒトホストからの細胞、ウイルス、細菌または真菌性病原体のような非ホストからの細胞または粒子を含む臨床サンプルであることができる。別の実施態様では、マルチアナライト対照は、臨床サンプルまたは実験サンプルのようなサンプルに直接添加され、生物の混合物を含むサンプル内の一つまたは複数の生物の力価を計算するための内部標準として使用される。いくつかの実施態様では、マルチアナライト対照は、本開示の方法のための陽性対照として使用される。

【0403】

いくつかの実施態様では、マルチアナライト対照は、ハイスループットシーケンシングライブラリーおよびそこからリードのコレクションの生成のために、サンプルと並行してであるが、それとは別個に処理される。いくつかの実施態様では、マルチアナライト対照およびサンプルに添加される正規化対照は、マルチアナライト対照およびサンプルからのリードの正規化のために使用される。マルチアナライト対照およびサンプルが並行して（例えば、マルチアナライト対照はサンプル内に「スパイクイン」されずに）処理される実施態様では、マルチアナライト対照は、ヒト病原体である種、ヒト病原体ではない種、またはこれらの組合せを含む。いくつかの実施態様では、マルチアナライト対照から正規化されたリードは、検量線の生成のために使用され、この検量線は、サンプル内の標的生物の力価を計算するために使用される。

【0404】

本開示の方法のいくつかの実施態様では、方法は、マルチアナライト対照内の各々の生物種についての少なくとも一つの標的配列を発現するレベルを定量化するために使用される。いくつかの実施態様では、方法は、（a）正規化対照の既知量をマルチアナライト対照と混合すること、（b）ハイスループットシーケンシングライブラリーを調製すること、（c）前記ライブラリーを配列決定して、リードのコレクションを製造すること、（d）リードのコレクションからのリードをマルチアナライト対照または正規化対照にマッピングすること、（e）正規化対照内のポリヌクレオチドのグループまたはサブグループの各々により製造されたリードの数を決定すること、（f）（a）においてマルチアナライト対照と混合された正規化対照内のポリヌクレオチドのグループまたはサブグループの各々の開始濃度と、（c）において製造されたリードの数との間の関係を計算すること、（g）リードと、マルチアナライト対照内の各々の種についての少なくとも一つの標的配列の濃度との間の関係をモデル化すること、および（h）（g）におけるモデルを使用して標的配列により製造されたリードの数からマルチアナライト対照内の各々の種についての少なくとも一つの標的配列の初期濃度を計算することを含む。いくつかの実施態様では、方法はさらに、マルチアナライト対照から核酸を抽出することを含む。いくつかの実施態様では、（a）の混合ステップは、（b）のライブラリー調製ステップの前に、（b）のライブラリー調製ステップと同時にまたは（b）のライブラリー調製ステップの後に行われる。いくつかの実施態様では、（a）の混合ステップは、サンプルからの核酸の抽出の前に行われる。

【0405】

いくつかの実施態様では、マルチアナライト対照内の異なる生物種の各々は、少なくとも一つの標的配列を含む。いくつかの実施態様では、少なくとも一つの標的配列は、マル

チアナライト対照内の各々の生物種にとって異なる標的配列である。いくつかの実施態様では、マルチアナライト対照内の各々の種は、一つまたは複数の異なる標的配列を含む。いくつかの実施態様では、マルチアナライト対照内の各々の生物種からの少なくとも一つの標的配列は、マルチアナライト対照内の他の生物種の全てまたは複数からの標的配列に対して99%未満の同一性、95%未満の同一性、90%未満の同一性、85%未満の同一性、80%未満の同一性、75%未満の同一性、70%未満の同一性、60%未満の同一性、50%未満の同一性、40%未満の同一性、30%未満の同一性または20%未満の同一性を有する。

【0406】

いくつかの実施態様では、マルチアナライト対照内の各々の生物種からの少なくとも一つの標的配列は、サンプルからの標的配列に対して同一であるか、99%未満の同一性、95%未満の同一性、90%未満の同一性、85%未満の同一性、80%未満の同一性、75%未満の同一性、70%未満の同一性、60%未満の同一性、50%未満の同一性、40%未満の同一性、30%未満の同一性、20%未満の同一性、10%未満の同一性または5%未満の同一性を有する。いくつかの実施態様では、マルチアナライト対照内の生物種からの少なくとも一つの標的配列は、サンプルからの標的配列と同一である。いくつかの実施態様では、マルチアナライト対照は、サンプルからの標的配列と同一の標的配列を含む少なくとも一つの種と、サンプルからの標的配列と同一でない標的配列を含む少なくとも一つの種とを含む。

【0407】

いくつかの実施態様では、マルチアナライト対照内の各々の生物種についての少なくとも一つの標的配列は、ゲノム配列を含む。いくつかの実施態様では、マルチアナライト対照内の各々の生物種からの少なくとも一つの標的配列は、その種の参照ゲノムを含むかまたはその参照ゲノムからなる。いくつかの実施態様では、マルチアナライト対照内の各々の生物種からの少なくとも一つの標的配列は、その種のトランスクリプトームを含むかまたはそのトランスクリプトームからなる。いくつかの実施態様では、マルチアナライト対照内の各々の生物種からの少なくとも一つの標的配列は、その種の参照ゲノムの一部を含む。少なくとも一つの標的配列は、タンパク質をコードする配列または非コードの配列であることができる。

【0408】

いくつかの実施態様では、ステップ(d)でのリードのコレクションからのリードのマルチアナライト対照または正規化対照へのマッピングはさらに、マルチアナライト対照(d)内の各々の生物種へのリードのマッピングを含む。ハイスループットシーケンシングリードのコレクションからのリードのマッピング方法は、当業者が容易に理解される。例えば、マルチアナライト対照内の各々の生物が参照ゲノムを有する場合、ハイスループット配列により生成されたリードのコレクションからのリードは、BWA、Novoalign、Bowtie、SOAP2、BFAST)、SSAHA2、MPscan、GASSST)またはPerMアルゴリズムを介して参照ゲノムにマッピングすることができる。いくつかの実施態様では、マルチアナライト対照内の各々の生物についての参照ゲノムへのリードのマッピングは、マルチアナライト対照内の各々の生物種についての少なくとも一つの標的配列へのリードのマッピングを含む。

【0409】

いくつかの実施態様では、(h)での少なくとも一つの標的配列の初期濃度の計算は、マルチアナライト対照内の各々の種の少なくとも一つの標的配列の初期濃度の計算を含む。マルチアナライト対照内の各々の種についての少なくとも一つの標的配列の初期濃度は、リードカウントと正規化対照組成物および本開示の方法を用いて決定される正規化対照内のグループまたはサブグループの各々の初期濃度との間の関係から計算することができる。

【0410】

本開示は、サンプル内の標的生物の力価を定量化する方法を提供する。いくつかの実施

10

20

30

40

50

態様では、方法は、（a）標的生物を含むサンプルを準備すること、ここで標的生物は少なくとも一つの標的配列を含む；（b）少なくとも三つの生物種の既知の力価を含むマルチアナライト対照を準備すること、ここで生物は不活性化されている；（c）本開示の正規化対照の既知量をサンプルおよびマルチアナライト対照と混合すること；（d）サンプルと少なくとも一つのマルチアナライト対照とからハイスループットシーケンシングライブラリーを調製すること；（e）前記ライブラリーをシーケンシングして、サンプルリードのコレクションとマルチアナライト対照リードのコレクションとを製造すること；（f）正規化対照を使用して（e）からのサンプルリードのコレクションとマルチアナライト対照リードのコレクションを正規化すること；（g）正規化されたリードとマルチアナライト対照内の生物種の既知の力価との間の関係を決定すること；および（h）（g）で決定した関係を使用して、サンプル内の標的生物の力価を計算することを含む。

10

【0411】

いくつかの実施態様では、生物種はそれぞれ、少なくとも一つの種特異的標的配列を含む。いくつかの実施態様では、少なくとも一つの種特異的標的配列は、マルチアナライト対照内の各々の種において異なる。

【0412】

いくつかの実施態様では、マルチアナライト対照からのリードの正規化は、（i）サンプルリードのコレクションからのリードをサンプルまたは正規化対照にマッピングすること；（ii）マルチアナライト対照リードのコレクションからのリードをマルチアナライト対照または正規化対照にマッピングすること；（iii）サンプルリードのコレクションおよびマルチアナライト対照リードのコレクションについて正規化対照内のポリヌクレオチドのグループまたはサブグループの各々により製造されたリードの数を決定すること；（iv）正規化対照内のポリヌクレオチドのグループまたはサブグループの各々の開始濃度と、サンプルおよびマルチアナライト対照の両方についてステップ（e）で製造されたリードの数との間の関係を計算すること；および（v）リードと、サンプル内およびマルチアナライト対照内の濃度との間の関係を決定することを含む。

20

【0413】

いくつかの実施態様では、マルチアナライト対照およびサンプル内のリードと力価との間の関係の決定は、（i）マルチアナライト対照内の各々の種からの少なくとも一つの種特異的標的配列の初期濃度を計算すること；および（ii）サンプル内の少なくとも一つの標的配列の初期濃度を計算することを含む。いくつかの実施態様では、ステップ（i）でのマッピングはさらに、サンプルリードのコレクションの参照ゲノムでのマッピングを含む。いくつかの実施態様では、参照ゲノムは標的配列を含まない。いくつかの実施態様では、参照ゲノムはヒトゲノムである。例えば、ヒト宿主細胞と病原体とを含むヒト臨床サンプル内で、病原体は標的配列を含み、サンプルからのハイスループットシーケンシングライブラリーから生成されたリードは、参照ゲノムとしてヒトゲノムにマッピングされる。ヒトゲノムにマッピングされかつヒトゲノムとアライメントされたリードは、ヒト細胞から単離された核酸に由来し、標的配列は含まない。ヒト参照ゲノムにマッピングされていないリードは、標的配列に由来し、本開示の方法を使用して正規化される。この方法の利点は、標的配列を産生する標的生物の同定を必ずしも事前に知る必要がないことである。本開示の方法は、ヒト臨床サンプル内の未知の病原体のような、サンプル内の未知の標的生物の存在および力価の両方を効率的に決定するために使用することができる。

30

40

【0414】

いくつかの実施態様では、ステップでのマッピングはさらに、一つより多くの参照ゲノムにサンプルリードのコレクションをマッピングすることを含む。いくつかの実施態様では、第一の参照ゲノムは標的配列を含まず、第二の参照ゲノムは標的配列を含む。例えば、ヒト宿主細胞と病原体とを含むヒト臨床サンプル内で、病原体は標的配列を含み、サンプルからのハイスループットシーケンシングライブラリーから製造されたリードは、標的配列を含んでいないヒトゲノムと、標的配列を含む病原体に対応する第二の参照ゲノムとにマッピングされる。いくつかの実施態様では、マッピングはさらに、マルチアナラ

50

イト対照内の生物種の各々について種特異的参照ゲノムへのマッピングを含む。

【0415】

いくつかの実施態様では、正規化されたリードと、マルチアナライト対照内の生物種の既知の力価との間の関係は線形関係である。いくつかの実施態様では、この関係は対数的である。正規化されたリードと、マルチアナライト対照内の生物種の既知の力価との間の関係を決定する方法は、当業者に容易に理解される。

【0416】

いくつかの実施態様では、マルチアナライト対照は、正規化されたリードカウントからサンプル内の生物の力価を決定するために使用される検量線を生成するために使用される。例示的な検量線が図13に示されており、検量線を生成するために使用されるマルチアナライト対照内の生物の対数力価に対するシグナル（正規化されたリード）として示される。いくつかの実施態様では、検量線は、一つまたは複数のサンプル（例えば臨床サンプルまたは実験サンプル）と並行して生成される。いくつかの実施態様では、検量線は、多数の異なる濃度で分析されるマルチアナライト対照から生成される。例えば、例示的な検量線は、1:10、1:100、1:1000などに希釈され、次いで配列決定されるマルチアナライト対照から生成することができる。いくつかの実施態様では、検量線は、2つの濃度で、3つの濃度で、4つの濃度で、5つの濃度で、6つの濃度で、7つの濃度で、8つの濃度で、9つの濃度でまたは10の濃度で提供されるマルチアナライト対照から生成される。いくつかの実施態様では、検量線は、参照検量線を含む。例えば、サンプル内の標的配列からの正規化されたリードは、標的配列に対応するサンプル内の生物の力価を決定するために、参照検量線と比較することができる。いくつかの実施態様、例えば参照検量線を使用する実施態様では、一つまたは複数のマルチアナライト対照は、検量線を検証するためおよび／または陽性対照として機能させるために、サンプルと並行して処理される。例えば、高濃度と低濃度との二つのマルチアナライト対照は、サンプルと並行して処理される。

【0417】

本開示の方法のいくつかの実施態様では、方法はさらに、マルチアナライト対照の既知量とサンプルとの混合を含む（スパイクイン法）。マルチアナライト対照は、生物の混合物を含むサンプル内の一つまたは複数の生物の力価を計算するための内部標準として使用される。例えば、サンプルは、宿主細胞（ヒト細胞）とウイルス粒子の一つまたは複数、または細菌性もしくは真菌性病原体からの細胞（非宿主）を含むヒト臨床サンプルであることができる。マルチアナライト対照は、ヒト病原体でない生物を含み、したがって、配列の相違を使用してサンプル内のヒト病原体と識別化能である。

【0418】

いくつかの実施態様では、マルチアナライト対照は、ヒト病原体と同じ属にある生物を含む。ヒト病原体に関連する生物は、処理の間にサンプル内のヒト病原体と同様に挙動し、サンプル内のヒト病原体の力価を計算するための優れた計量を提供する。いくつかの実施態様では、マルチアナライト対照内の生物は、ヒト病原体の配列に対して99%未満の同一性、95%未満の同一性、90%未満の同一性、85%未満の同一性、80%未満の同一性、75%未満の同一性、70%未満の同一性、60%未満の同一性、50%未満の同一性、40%未満の同一性、30%未満の同一性または20%未満の同一性を有する標的配列を含む。

製剤

【0419】

本開示の正規化対照およびマルチアナライト対照組成物は、当分野で公知の許容可能なあらゆる担体、希釈剤または付形剤内に調剤することができる。

【0420】

正規化対照は、許容可能なあらゆる濃度で、溶液内で、沈殿物で、または凍結乾燥組成物として提供することができる。正規化対照のような核酸は、水、トリス緩衝液、トリス-E D T A緩衝液および最大で0.5 MまでNaClを有するトリス-E D T Aを含む溶

液内に懸濁されてもよいが、これらに限定されるものではない。

【0421】

本開示の製剤内に使用される例示的な緩衝液は、トリスー（ヒドロキシメチル）アミノメタン（TRIS）の酢酸塩、硫酸塩、塩酸塩、リン酸塩または遊離酸の形態を含むが、TRISとほぼ同じイオン強度およびpKaの代替緩衝液を使用しても同等の結果が得られる。緩衝塩に加えて、カリウム（例えば、塩化カリウムまたは酢酸カリウム）およびマグネシウム（例えば、塩化マグネシウムまたは酢酸マグネシウム）のような補因子塩が、組成物中に含まれる。

【0422】

マルチアナライト対照内の不活性化された生物は、溶液、フリーズドライまたは凍結乾燥で供給することができる。フリーズドライまたは凍結乾燥する場合、マルチアナライト対照は、許容可能なあらゆる担体の適切な体積内に再懸濁することができる。許容可能な担体は、水、リン酸塩で緩衝された食塩水（PBS）またはリンゲル溶液を含むが、これらに限定されるものではない。

枯渇または濃縮方法

【0423】

正規化対照、マルチアナライト対照、およびこれらを使用する方法は、目的の標的配列についてサンプルを濃縮するか、または枯渇の標的となる配列のサンプルを枯渇させる方法と組み合わせることができる。いくつかの実施態様では、濃縮または枯渇方法は、核酸誘導ヌクレアーゼに基づく方法、ヌクレオチド修飾に基づく方法、またはこれらの組合せを含む。

【0424】

したがって、本開示は、枯渇の標的となる配列のサンプルを枯渇させ、ここに記載された正規化対照を添加し、NGSを使用してサンプルをシーケンシングすることを含む、サンプル内の標的配列を定量化または同定する方法を提供する。正規化対照は、目的の標的配列の濃縮または枯渇の標的となる核酸の枯渇の前、その間またはその後に添加することができる。

【0425】

いくつかの実施態様では、目的の標的配列の濃縮または枯渇の標的となる配列の枯渇の前に、正規化対照をシーケンシングサンプルに添加する。いくつかの実施態様では、正規化対照は、枯渇の標的となる配列からこのような配列を識別するために使用される、目的の標的配列のそれらの特性を模倣するように設計することができる。例えば、目的の標的配列が、核酸誘導ヌクレアーゼと、枯渇の標的となる配列を標的とする複数のガイド核酸（gNA）を使用して濃縮する場合、正規化対照は、複数のgNAにより標的とされる配列を含んでいない。更なる例として、枯渇の標的となる配列が、ヌクレオチド修飾に基づく方法を用いて枯渇される場合、正規化対照は、濃縮された目的の標的配列の修飾と同じヌクレオチド修飾を含んでよい。

ヌクレオチド修飾に基づく方法

【0426】

ここで提供されるのは、目標の標的核酸と枯渇の標的となる核酸との間のヌクレオチド修飾における相違を使用することを含む、枯渇の標的となる核酸に関する目的の標的核酸についてサンプルを濃縮する方法である。いくつかの実施態様では、標的核酸は、シーケンシングサンプル内に標的配列を含み、その存在および／または力価は、正規化対照および任意に本開示の目的のマルチアナライト対照を使用して決定することができる。目的の標的配列と同じヌクレオチド修飾を含む正規化対照は、修飾に基づく枯渇方法の前にサンプルに添加することができる。あるいは、正規化対照は、修飾に基づく枯渇の後にサンプルに添加することができる。

【0427】

あらゆるタイプのヌクレオチド修飾は、本開示の範囲内にあることが想定される。本開示のヌクレオチド修飾の例示的であるが非限定的な例を以下に記載する。

【0428】

本開示の方法により使用されるヌクレオチド修飾は、あらゆるヌクレオチド（例えば、アデニン、シトシン、グアニン、チミンまたはウラシル）に行うことができる。これらのヌクレオチド修飾は、デオキシリボ核酸（DNA）またはリボ核酸（RNA）に行うことができる。これらのヌクレオチド修飾は、二本鎖もしくは一本鎖DNA分子、または二本鎖もしくは一本鎖RNA分子に行うことができる。

【0429】

いくつかの実施態様では、ヌクレオチド修飾は、アデニン修飾またはシトシン修飾を含む。

【0430】

いくつかの実施態様では、アデニン修飾は、アデニンメチル化を含む。いくつかの実施態様では、アデニンメチル化は、N6-メチルアデニン（6mA）を含む。いくつかの実施態様では、アデニンメチル化は、デオキシアデノシンメチラーゼにより実施されるDamメチル化を含む。いくつかの実施態様では、アデニンメチル化は、EcoKIメチル化を含む。いくつかの実施態様では、アデニン修飾は、グリシンによるN⁶で修飾されたアデニンを含む（mom化）。

【0431】

いくつかの実施態様では、修飾は、シトシン修飾を含む。いくつかの実施態様では、シトシン修飾は、5-メチルシトシン（5mC）、5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）、5-カルボキシルシトシン（5caC）、5-グルコシルヒドロキシメチルシトシン（5ghmC）または3-メチルシトシン（3mC）を含む。いくつかの実施態様では、シトシンメチル化は、5-メチルシトシン（5mC）またはN4-メチルシトシン（4mC）を含む。いくつかの実施態様では、シトシンメチル化は、Dcmメチル化、DNMT1メチル化、DNMT3Aメチル化またはDNMT3Bメチル化を含む。いくつかの実施態様では、シトシンメチル化は、CpGメチル化、CpAメチル化、CpTメチル化、CpCメチル化またはこれらの組合せを含む。いくつかの実施態様では、シトシンメチル化は、CpGメチル化を含む。例えば、CpGメチル化は、本開示の方法を使用して枯渇のための哺乳類ゲノム内の活性領域を選択的に標的にするために使用することができる。

【0432】

いくつかの実施態様では、シトシンメチル化は、5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）を含む。5-ホルミルシトシンは、5mC、5-カルボキシルシトシン（5caC）、5-グルコシルヒドロキシメチルシトシンまたは3-メチルシトシンの酸化誘導体である。

【0433】

本開示の方法のいくつかの実施態様では、方法は、少なくとも第一の修飾感受性制限酵素と第二の修飾感受性制限酵素とを使用する。いくつかの実施態様では、第一と第二の修飾感受性制限酵素は同じである。いくつかの実施態様では、第一と第二の修飾感受性制限酵素は同じでない。いくつかの実施態様では、第一または第二の修飾感受性制限酵素は、単一種の制限酵素である（例えば、AluI、またはMcrBCであるが、両方ではない）。いくつかの実施態様では、第一または第二の修飾感受性制限酵素は、2つまたはそれ以上の種の修飾感受性制限酵素の混合物（例えばFspEIおよびAbaSIの混合物）である。本開示の方法のいくつかの実施態様では、二つより多くの異なる方法が組み合わせられ、各々は、異なる修飾感受性制限酵素または修飾感受性制限酵素のカクテルを使用する。

【0434】

ここで使用されるような「修飾感受性制限酵素」の用語は、制限酵素についての認識部位内またはその認識部位に隣接する修飾ヌクレオチドの存在に敏感である制限酵素を言う。あるいは、または付加的に、修飾感受性制限酵素は、認識部位自体内の修飾ヌクレオチドに敏感であることができる。修飾感受性制限酵素は、認識部位に隣接する、例えば認識

10

20

30

40

50

部位の5'または3'の1~50のヌクレオチド内にある修飾ヌクレオチドに敏感であることができる。本開示のヌクレオチド修飾は、認識部位自体内にあるか、または認識部位に隣接するヌクレオチド（例えば、認識部位の5'または3'の1~50のヌクレオチド内、またはその両方）を含むことができる。

【0435】

修飾感受性制限酵素の活性を遮断または低減することができる例示的な修飾は、N⁶-メチルアデニン、5-メチルシトシン（5mC）、5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）、5-カルボキシルシトシン（5caC）、5-グルコシルヒドロキシメチルシトシン、3-メチルシトシン（3mC）、N⁴-メチルシトシン（4mC）またはこれらの組合せを含むが、これらに限定されるものではない。修飾感受性制限酵素を遮断することができる例示的な修飾は、Dam、Dcm、EcoKI、DNMT1、DNMT3A、DNMT3BおよびTET酵素により媒介される修飾を含む。

【0436】

いくつかの実施態様では、修飾は、Damメチル化を含む。Damメチル化により遮断される制限酵素は、AlwI、BcgI、BclI、BsaBI、BspDI、BspEI、BspHI、ClaI、DpnII、HphI、Hpy188I、Hpy188II、MboI、MboII、NruI、Nt. AlwI、TaqαIおよびXbaIを含むが、これらに限定されるものではない。

【0437】

いくつかの実施態様では、修飾は、Dcmメチル化を含む。Dcmメチル化により遮断される制限酵素は、Acc65I、AlwNI、ApaI、AvaI、AvaII、BanI、BsaI、BsaHI、BslI、BsmFI、BssKI、BstXI、EaeI、Esp3I、EcoO109I、MscI、NlaIV、PflMI、PspGI、PspOMI、Sau96I、ScrFI、SexAI、SfiI、SfoIおよびStuIを含むが、これらに限定されるものではない。いくつかの実施態様では、修飾は、CpGメチル化を含む。CpGメチル化により遮断される制限酵素は、AatII、AccII、AciI、AclI、AfeI、AgeI、Aor13HI、Aor51HI、AscI、AsiSI、AluI、AvaI、BceAI、BmgBI、BsaI、BsaHI、BsiEI、BsiWI、BsmBI、BspDI、BspT104I、BsrFαI、BssHII、BstBI、BstUI、Cfr10I、ClaI、CpoI、EagI、Esp3I、Eco52I、FauI、FseI、FspI、HaeII、HgaI、HhaI、HpaII、HpyCH4IV、Hpy99I、KasI、MluI、NaeI、NgoMIV、NotI、NruI、Nt. BsmAI、Nt. CviPII、NsbI、PmaCI、Psp1406I、PluTI、PmlI、PvuI、RsrII、SacII、SalI、SmaI、SnaBI、SfoI、SgrAI、SmaI、SrfI、Sau3AI、TspMIおよびZraIを含むが、これらに限定されるものではない。

【0438】

いくつかの実施態様では、修飾感受性制限酵素は、少なくとも一つの修飾ヌクレオチドを含む認識部位では活性であり、少なくとも一つの修飾ヌクレオチドを含まない認識部位では活性ではない。一つまたは複数の修飾ヌクレオチドを含む認識部位で切断する修飾感受性制限酵素により認識される例示的な修飾は、N⁶-メチルアデニン、5-メチルシトシン（5mC）、5-ヒドロキシメチルシトシン（5hmC）、5-ホルミルシトシン（5fC）、5-カルボキシルシトシン（5caC）、5-グルコシルヒドロキシメチルシトシン、3-メチルシトシン（3mC）、N⁴-メチルシトシン（4mC）またはこれらの組合せを含むが、これらに限定されるものではない。一つまたは複数の修飾ヌクレオチドを含む認識部位を特異的に切断する修飾感受性制限酵素が認識する例示的な修飾は、Dam、Dcm、EcoKI、DNMT1、DNMT3A、DNMT3BおよびTET酵素により媒介される修飾を含む。

【0439】

認識部位内または認識部位に隣接する一つまたは複数の修飾ヌクレオチドを含む認識部位で切断する例示的であるが制限のない修飾感受性制限酵素は、AbaSI、DpnI、FspEI、LpnPI、MspJIおよびMcrBCを含むが、これらに限定されるものではない。

【0440】

いくつかの実施態様では、修飾は、5-グルコシルヒドロキシメチルシトシンを含み、修飾感受性制限酵素は、AbaSIを含む。AbaSIは、グルコシルヒドロキシメチルシトシンを含むAbaSI制限部位を切断するが、グルコシルヒドロキシメチルシトシンを含まないAbaSI認識部位は切断しない。

10

【0441】

いくつかの実施態様では、ヌクレオチド修飾は、5-ヒドロキシメチルシトシンを含み、修飾感受性制限酵素は、AbaSIおよびT4ファージβ-グルコシルトランスフェラーゼを含む。T4ファージβ-グルコシルトランスフェラーゼは、ウリジンジホスホグルコース(UDP-Glc)のグルコース基を、二本鎖DNA内で、例えばAbaSI認識部位内で、5-ヒドロキシメチルシトシン(5-hmC)残基に特異的に変換し、グルコシルヒドロキシメチルシトシン修飾AbaSI認識部位を製造する。AbaSIは、グルコシルヒドロキシメチルシトシンを含むAbaSI制限部位を切断するが、グルコシルヒドロキシメチルシトシンを含まないAbaSI認識部位は切断しない。

【0442】

20

いくつかの実施態様では、ヌクレオチド修飾は、メチルシトシンを含み、修飾感受性制限酵素は、McrBCを含む。McrBCは、メチルシトシンを含むMcrBC部位を切断するが、メチルシトシンを含まないMcrBC部位は切断しない。McrBC部位は、一方または両方のDNA鎖でメチルシトシンにより修飾することができる。いくつかの実施態様では、McrBCはまた、一方または両方のDNA鎖でヒドロキシメチルシトシンを含むMcrBC部位を切断する。いくつかの実施態様では、McrBCハーフ部位は、3,000までのヌクレオチドにより分離される。いくつかの実施態様では、McrBCハーフ部位は、55~103のヌクレオチドにより分離される。

【0443】

いくつかの実施態様では、修飾は、アデニンメチル化を含み、方法は、DpnIを用いた消化を含む。DpnIは、GATC認識の両方の鎖のアデニンがメチル化されている場合、GATC認識部位を切断する。いくつかの実施態様では、アデニンメチル化とシトシン修飾の両方を含むDpnI GATC認識部位は、細菌DNAでは生じるが、哺乳動物DNAでは生じない。メチル化アデニンと修飾シトシンの両方を含むこれらの認識部位は、(例えば、細菌と哺乳動物との混合したDNAの)サンプル内でDpnIにより選択的に切断することができ、次いでT4ポリメラーゼで処理して、切断末端でのメチル化アデニンと修飾シトシンを、非修飾アデニンおよびシトシンに置き換える。T4ポリメラーゼは、テンプレート、プライマーおよびヌクレオチドの存在で、5'から3'方向でDNAの合成を触媒する。T4ポリメラーゼは、非修飾のヌクレオチドを新たに合成されたDNA内に組み込む。これは、目的の核酸内で非修飾のシトシンと、枯渇の標的となる核酸内の修飾シトシンとを含むサンプルが製造する。修飾シトシンにおけるこの相違は、本開示の方法を使用して目的の核酸を濃縮するために使用することができる。

30

【0444】

本開示の方法のいくつかの実施態様では、サンプル内の核酸は、末端脱リン酸されるので、サンプル内の核酸は、修飾感受性制限酵素と接触して、目的の核酸か、または目的の核酸についてのサンプルを濃縮する本開示の方法において使用することができる露出した末端リン酸塩を有する枯渇の標的となる核酸のいずれかが製造される。例えば、この露出した末端リン酸塩は、エキソヌクレアーゼによる分解のための枯渇について核酸を標的とするため(図28)、またはアダプターライゲーションについて目的の核酸を標的にするため(図27)に使用することができる。

40

50

【0445】

ここで使用されるような「末端脱リン酸」の用語は、核酸分子の5'および3'末端から末端リン酸基が除去された核酸を言う。いくつかの実施態様では、サンプル内の核酸は、アルカリホスファターゼのようなホスファターゼを使用して末端脱リン酸される。本開示の例示的なホスファターゼは、エビアルカリホスファターゼ(SAP)、組換えエビアルカリホスファターゼ(rSAP)、子ウシ腸アルカリホスファターゼ(CIP)および南極ホスファターゼを含むが、これらの限定されるものではない。

【0446】

ここで使用されるような「エキソヌクレアーゼ」の用語は、核酸分子の3'または5'末端からヌクレオチドを連続的に除去する酵素のクラスを言う。核酸分子は、DNAまたはRNAであることができる。DNAまたはRNAは、一本鎖または二本鎖であることができる。例示的なエキソヌクレアーゼは、ラムダヌクレアーゼ、エキソヌクレアーゼI、エキソヌクレアーゼIIIおよびBAL-31を含むが、これらに限定されるものではない。エキソヌクレアーゼは、本開示の方法を使用して、枯渇の標的となる核酸を選択的に分解するために使用することができる(例えば図28)。

【0447】

本開示は、サンプル、および任意に正規化対照内の核酸または目的の核酸の5'および3'末端にライゲーションされたアダプターを提供する。いくつかの実施態様では、アダプターは、正規化対照配列にライゲーションされる。他の実施態様では、アダプター配列を含む正規化対照は、デノボ合成される。

【0448】

本開示の方法のいくつかの実施態様では、アダプターは、サンプル内の全ての核酸にライゲーションされ、次いで、ヌクレオチド修飾における相違を使用して、枯渇の標的となる核酸を選択的に切断し、両端にアダプターがライゲーションされた目的の核酸と、一端にアダプターがライゲーションされた枯渇の標的となる核酸とを生成する(図29、図30)。いくつかの実施態様では、ヌクレオチド修飾の相違を使用して、枯渇の標的となる核酸を選択的に枯渇させ、次いでアダプターを目的の標的核酸にライゲーションする(図28)。いくつかの実施態様では、ヌクレオチド修飾の相違を使用して、露出した末端リン酸塩を有する目的の核酸を製造し、アダプターを目的の標的核酸にライゲーションするために使用する(図27)。

【0449】

本開示の方法のいくつかの実施態様では、アダプターは、サンプル内の核酸の5'および3'にライゲーションされる。いくつかの実施態様では、アダプターはさらに、5'末端および/または3'末端の間の介在配列を含む。

【0450】

いくつかの実施態様では、アダプターは、二本鎖DNA分子の両方の鎖にライゲーション可能な核酸である。

【0451】

いくつかの実施態様では、アダプターは、枯渇/濃縮の前にライゲーションされる。いくつかの実施態様では、アダプターは、後のステップでライゲーションされる。

【0452】

アダプターの非限定的な例は、線形、線型Y型、またはヘアピン型のアダプターを含む。いくつかの実施態様では、アダプターは、ポリG配列を含む。

【0453】

多様な実施態様では、アダプターは、ヘアピン型アダプター、すなわち二本鎖のステムとループとを有する構造を形成するために塩基がそれ自体と対になる分子であってよく、この分子の3'および5'末端は、それぞれフラグメントの二本鎖DNA分子の5'および3'末端にライゲーションする。

【0454】

あるいは、アダプターは、フラグメントの一端または両端とライゲーションするY型ア

10

20

30

40

50

アダプターであってよく、これはユニバーサルアダプターとも呼ばれる。あるいは、アダプター自体は、互いに対になる塩基である二つの別個のオリゴヌクレオチド分子からなっていてよい。付加的に、アダプターのライゲーション可能な末端は、制限酵素により切断することによって製造されたオーバーハングに適合性となるように設計されているか、または平滑末端または5' Tオーバーハングを有していてよい。いくつかの実施態様では、制限酵素は、修飾感受性制限酵素である。

【0455】

アダプターは、二本鎖および一本鎖の分子を含んでよい。したがって、アダプターは、DNAまたはRNA、またはその二つの混合物であることができる。RNAを含むアダプターは、RNase処理またはアルカリ加水分解によって切断されてよい。

10

【0456】

アダプターは、10～100bpの長さであることができるが、この範囲外のアダプターも、本開示から逸脱することなく使用可能である。

【0457】

アダプターは、次世代シーケンシングプラットフォームのため、例えばHiSeqもしくはMiSeqのようなイルミナシーケンシングプラットフォームで使用するため、またはIonTorrentプラットフォームで使用するため、またはナノポア技術を用いて使用するために構成されていてよい。いくつかの実施態様では、アダプターは、シーケンシングアダプター（例えばイルミナシーケンシングアダプター）を含む。いくつかの実施態様では、アダプターは、時にはバーコードと言われる固有分子識別子（UMI）配列を含む。いくつかの実施態様では、UMI配列は、各々の元の核酸分子に固有の配列（例えばランダム配列）を含む。いくつかの実施態様では、アダプターは、各々の核酸分子に固有のUMI、特別な供給源からの核酸分子の間で共通のバーコード、およびシーケンシングアダプターのような多数の個別の配列を含む。

20

【0458】

枯渇の標的となる核酸は、差動的アダプター取付により枯渇させることができる。いくつかの実施態様では、アダプターは、サンプルの核酸に取り付けられ、引き続き一つまたは複数のアダプターは、それらの修飾状態に基づいて、枯渇の標的となる核酸から除去される。例えば、両端に取り付けられたアダプターを有する枯渇の標的となる核酸は、修飾感受性制限酵素により切断することができ、それにより、一端だけに取り付けられたアダプターを有する枯渇の標的となる核酸が製造される。引き続くステップ（例えば、増幅）は、両端に取り付けられたアダプターを有する核酸だけを標的として使用することができ、それにより、枯渇の標的となる核酸は枯渇される。他の例では、サンプルの核酸は、切断された核酸だけが取り付けられたアダプターを有することができるように（例えば脱リン酸により）処理され；引き続き、目的の核酸は、修飾感受性制限酵素により切断することができ（例えば、それによりリン酸塩基を露出させる）、アダプターを取り付けることができる。引き続くステップ（例えば、増幅）は、取り付けられたアダプターを有する核酸だけを標的とするために使用することができ、それにより枯渇の標的となる核酸を枯渇させる。

30

【0459】

枯渇の標的となる核酸は、エキソヌクレアーゼによる消化のように、消化により枯渇させることができる。

40

【0460】

枯渇の標的となる核酸は、サイズ選択により枯渇させることができる。例えば、修飾感受性制限酵素は、目的の核酸か、枯渇の標的となる核酸のいずれかを切断するために使用することができ、引き続き、目的の核酸は、切断によるサイズの相違に基づき枯渇の標的となる核酸から分離することができる。

【0461】

いくつかの場合には、枯渇の標的となる核酸は、サイズ選択を使用することなく枯渇される。

50

【0462】

プロトコル1：ここに記載された適用の例示的な方法は、図27に示されている。目的の標的核酸(2701)と枯渇の標的となる核酸(2702)と、任意に正規化対照とを含む核酸のサンプルを、最終的に脱リン酸(2705)させて、目的のリン酸化されていない核酸(2706)と、枯渇の標的となる核酸(2707)とを製造する。いくつかの実施態様では、核酸は、脱リン酸の前に断片化される。いくつかの実施態様では、サンプル内の核酸は、最終的にホスファターゼ、例えば組換えエビアルカリホスファターゼ(rSAP)により脱リン酸される。いくつかの実施態様では、目的の核酸と枯渇の標的となる核酸との両方は、修飾感受性制限酵素のための一つまたは複数の認識部位(それぞれ2703、2704)を含む。目的の核酸では、修飾感受性制限酵素のための認識部位は、修飾ヌクレオチド(2703)を含まないか、あるいは、枯渇の標的となる核酸の対応する認識部位よりも低い頻度の修飾ヌクレオチドを含む。枯渇の標的となる核酸では、修飾感受性制限酵素のための認識部位は、制限部位内にまたは制限部位に隣接する修飾ヌクレオチド(2704)を含むか、あるいは、目的の核酸の対応する認識部位よりも高い頻度の修飾ヌクレオチドを含む。修飾感受性制限酵素(2709)の活性は、同種の認識部位内にまたはその認識部位に隣接する修飾ヌクレオチドの存在により遮断され(2708)、それにより修飾感受性制限酵素の活性は目的の核酸を標的にする(2710と2711とを比較)。いくつかの実施態様では、修飾感受性制限酵素(2709)は、AatII、AccII、Aor13HI、Aor51HI、BspT104I、BssHII、Cfr10I、ClaI、CpoI、Eco52I、HaeII、HapII、HhaI、MluI、NaeI、NotI、NruI、NsbI、PmaCI、Psp1406I、PvuI、SacII、SalI、SmaI、SnaBI、AluIまたはSau3AIを含む。いくつかの実施態様では、修飾感受性制限酵素(2709)は、AluIまたはSau3AIを含む。サンプルを修飾感受性制限酵素で消化すると(2713)、末端リン酸塩の5'および3'末端で末端リン酸塩を有する目的の核酸が製造される(2714)。この末端リン酸塩は、目的の核酸の末端にアダプターをライゲーション(2715、ライゲーションステップ; 116、アダプター)するために使用され、両端にアダプターがライゲーションされた目的の核酸(2717)が製造される。対照的に、枯渇の標的とされる核酸は、アダプターがライゲーションされていない(2711)。これらのアダプターは、下流の適用、例えばアダプター媒介PCR増幅、シーケンシング(例えばハイスループットシーケンシング)、およびサンプル内の目的の核酸の定量化および／またはクローニングのために使用することができる。これは、目的の核酸にアダプターを選択的にライゲーションすることにより、枯渇の標的となる核酸を枯渇させる。この枯渇は、サイズ選択を使用せずに達成することができる。あるいは、アダプターがライゲーションした目的の核酸は、ここに記載された付加的な濃縮方法の一つまたはそれ以上に供される。例えば、アダプターがライゲーションした核酸は、本開示の付加的な修飾依存性濃縮方法(例えば、図29に示された方法)に供される。あるいは、または付加的に、アダプターがライゲーションした核酸は、本開示の核酸誘導ヌクレアーゼに基づく濃縮方法(例えば、図30に示された方法)に供される。

【0463】

プロトコル2：ここに記載された適用の例示的な方法は、図28に示されている。目的の標的核酸(2801)と枯渇の標的となる核酸(2802)と、任意に正規化対照とを含む核酸のサンプルを、最終的に脱リン酸(2805)させて、目的のリン酸化されていない核酸(2806)と、枯渇の標的となる核酸(2807)とを製造する。いくつかの実施態様では、核酸は、脱リン酸の前に断片化される。いくつかの実施態様では、サンプル内の核酸は、最終的にホスファターゼ、例えば組換えエビアルカリホスファターゼ(rSAP)により脱リン酸される。いくつかの実施態様では、目的の核酸と枯渇の標的となる核酸との両方は、修飾感受性制限酵素のための一つまたは複数の認識部位(それぞれ2803および2804)を含む。目的の核酸では、修飾感受性制限酵素のための認識部位は、修飾ヌクレオチド(2803)を含まないか、あるいは、枯渇の標的となる核酸の対応

する認識部位よりも低い頻度の修飾ヌクレオチドを含む。枯渇の標的となる核酸では、修飾感受性制限酵素のための認識部位は、制限部位内にまたは制限部位に隣接する修飾ヌクレオチド（2804）を含むか、あるいは、目的の核酸の対応する認識部位よりも高い頻度の修飾ヌクレオチドを含む。修飾感受性制限酵素（2809）は、認識部位内にまたは認識部位に隣接する一つまたは複数の修飾ヌクレオチドが存在する場合、その同種の認識部位を切断し（2808）、この認識部位が一つまたは複数の修飾ヌクレオチドを含まない場合、その同種の認識部位を切断せず（2808）、それにより、修飾感受性制限酵素の活性は枯渇の標的となる核酸を標的とする（2810と2811とを比較）。いくつかの実施態様では、修飾感受性制限酵素は、AbaSI、FspEI、LpnPI、MspJIまたはMcrBCを含む。いくつかの実施態様では、修飾感受性制限酵素は、FspEIである。いくつかの実施態様では、修飾感受性制限酵素は、MspJIである。サンプルを修飾感受性制限酵素で消化すると（2812）、核酸の5'および3'末端の一端（2813）または両方（2814）に末端リン酸塩を有する枯渇の標的となる核酸が製造される。対照的に、修飾感受性制限酵素により切断されなかった目的の核酸は、核酸の5'およびまたは3'末端に露出した末端リン酸塩を有しない（2810と2813-2814とを比較）。次いで、枯渇の標的となる核酸内の末端リン酸塩を使用して、核酸分子の末端からヌクレオチドを連続して除去するエキソヌクレアーゼを用いてサンプルを消化（2815、消化ステップ；2816 エキソヌクレアーゼ）して、サンプルから枯渇の標的となる核酸を枯渇させる。この枯渇は、サイズ選択を使用せずに達成することができる。エキソヌクレアーゼ消化の後で、アダプターは、末端リン酸塩を欠き、エキソヌクレアーゼにより消化されなかった目的の核酸にライゲーションされる（2817）。これは、両端にアダプターがライゲーションされた目的の核酸（2818）を製造する。これらのアダプターは、下流の適用、例えばアダプター媒介PCR増幅、シーケンシング（例えばハイスループットシーケンシング）、およびサンプル内の目的の核酸の定量化および／またはクローニングのために使用することができる。あるいは、アダプターがライゲーションした目的の核酸は、ここに記載された付加的な濃縮方法の一つまたはそれ以上に供される。例えば、アダプターがライゲーションした核酸は、本開示の付加的な修飾依存性濃縮方法（例えば、図29に示された方法）に供される。あるいは、または付加的に、アダプターがライゲーションした核酸は、本開示の核酸誘導ヌクレアーゼに基づく濃縮方法（例えば、図30に示された方法）に供される。

【0464】

プロトコル3：ここに記載された適用の例示的な方法は、図29に示されている。目的の核酸（2901）と、枯渇の標的となる核酸（2902）と、任意に正規化対照を含む核酸のサンプルを、アダプターライゲーションする（2905）か、または本開示の濃縮方法（2906）（例えば、図27または図28に示された方法）に供し、アダプターライゲーションされた目的の核酸（2907）と、アダプターライゲーションされた枯渇の標的となる核酸（2908）とを製造する。いくつかの実施態様では、目的の核酸と枯渇の標的となる核酸との両方は、修飾感受性制限酵素のための一つまたは複数の認識部位（それぞれ2903、2904）を含む。目的の核酸では、修飾感受性制限酵素のための認識部位は、修飾ヌクレオチド（2903）を含まないか、あるいは、枯渇の標的となる核酸では、修飾感受性制限酵素のための認識部位は、制限部位内にまたは制限部位に隣接する修飾ヌクレオチド（2904）を含むか、あるいは、目的の核酸の対応する認識部位よりも高い頻度の修飾ヌクレオチドを含む。修飾感受性制限酵素（2909）は、認識部位内にまたは認識部位に隣接する一つまたは複数の修飾ヌクレオチドが存在する場合、その同種の認識部位を切断し（2908）、この認識部位が一つまたは複数の修飾ヌクレオチドを含まない場合、その同種の認識部位を切断せず（2908）、それにより、修飾感受性制限酵素の活性は枯渇の標的となる核酸を標的とする（2910と2911とを比較）。いくつかの実施態様では、修飾感受性制限酵素は、AbaSI、FspEI、LpnPI、MspJIまたはMcrBCを含む。いくつかの実施態様では、修飾感受性制限酵素

は、FspEIである。いくつかの実施態様では、修飾感受性制限酵素は、MspJIである。サンプルは、修飾感受性制限酵素で消化され（2911）、アダプターがライゲーションされていない（2912）か、または一端にだけアダプターがライゲーションされている（2913）枯渇の標的となる核酸を製造する。これは、枯渇の標的となる核酸からアダプターを選択的に除去することにより、枯渇の標的となる核酸を枯渇させる。この枯渇は、サイズ選択を使用せずに達成することができる。対照的に、修飾感受性制限酵素により切断されなかった目的の核酸は、両端にアダプターがライゲーションされる（2910と2912-2913とを比較）。これらのアダプターは、下流の適用、例えばアダプター媒介PCR増幅、シーケンシング（例えばハイスループットシーケンシング）、およびサンプル内の目的の核酸の定量化および／またはクローニングのために使用することができる。

10

【0465】

プロトコル4：ここに記載された適用の例示的な方法は、図30に示されている。複数のgNA（3001）を使用して、核酸誘導ヌクレアーゼ（3002）は、アダプターライゲーションされた核酸のサンプル内で枯渇の標的となる核酸（3003）を標的とする。アダプターがライゲーションされた核酸は、ここに記載された濃縮方法のいずれかにより生成され、最初のアダプターライゲーションの前または後で、修飾感受性制限酵素を使用してサンプルからの枯渇の標的となる核酸を枯渇させるために使用される。この方法では、gNAは、枯渇の標的となる核酸（3003）を特異的に標的にし、目的の核酸（3004）は標的にしないため、この目的の核酸は、核酸誘導ヌクレアーゼ（3002）により切断されない。核酸誘導ヌクレアーゼによる切断は、一端にアダプターがライゲーションされた枯渇の標的となる核酸（3005）と、両端にアダプターがライゲーションされた目的の核酸（3003）とを生じさせる。これらのアダプターは、下流の適用、例えばアダプター媒介PCR増幅、シーケンシング（例えばハイスループットシーケンシング）、サンプル内の目的の核酸の定量化およびクローニングのために使用することができる。

20

【0466】

ここに記載された方法のいずれも、サンプルから枯渇の標的となる核酸を枯渇させ、それにより目的の核酸を濃縮させる独立した方法として使用することができる。あるいは、ここに記載された方法を、独立したあらゆる個々の方法よりも高い程度の濃縮を達成するために組み合わせることができる。

30

【0467】

方法の特別な組合せ、および方法の組合せの順序が、ここに記載されているが、本開示の方法と組み合わせることができる経路を制限することを意図するものではない。方法の産物として目的のアダプターがライゲーションされた核酸を製造する本開示の目的の核酸についてサンプルを濃縮するあらゆる方法は、その出発基質としてのアダプターがライゲーションされた核酸を使用する本開示のあらゆる付加的な方法と組み合わせることができる。

核酸誘導ヌクレアーゼを基礎とする濃縮方法

【0468】

本開示は、正規化対照およびマルチアナライト対照、および任意に、ここに記載された他の濃縮または枯渇の方法と組み合わせることができる、標的配列について濃縮するための核酸誘導ヌクレアーゼを基礎とする方法を提供する。核酸誘導ヌクレアーゼを基礎とする濃縮方法は、目的の配列についてサンプルを濃縮するために核酸誘導ヌクレアーゼを使用する方法である。核酸誘導ヌクレアーゼを基礎とする濃縮方法は、国際公開第2016/100955号、国際公開第2017/031360号、国際公開2017/100343号、国際公開2017/147345号および国際公開2018/227025号に記載されており、これらの各々の内容は、参照により、全体として本明細書に組み込まれる。

40

【0469】

50

正規化および任意に、マルチアナライト対照は、核酸誘導ヌクレアーゼを基礎とする方法にサンプルを供する前に、または後に、サンプルに添加することができる。いくつかの実施態様では、正規化対照は、ここに記載された g N A の標的配列の配列と同じではない配列を含む。

【0470】

「核酸誘導ヌクレアーゼー g N A 複合体」の用語は、核酸誘導ヌクレアーゼタンパク質とガイド核酸 (g N A、例えば g R N A または g D N A) とを含む複合体を言う。例えば、「C a s 9ー g R N A 複合体」は、C a s 9 タンパク質とガイド R N A (g R N A) を含む複合体を言う。核酸誘導ヌクレアーゼは、限定されるものではないが、野生型核酸誘導ヌクレアーゼ、触媒的に死んだ核酸誘導ヌクレアーゼ、または核酸誘導ヌクレアーゼニックアーゼを含むあらゆるタイプの核酸誘導ヌクレアーゼであってよい。核酸誘導ヌクレアーゼが C R I S P R / C a s 核酸誘導ヌクレアーゼである場合、この複合体を、「C R I S P R / C a s システムタンパク質 g N A 複合体」と言うことができる。

【0471】

本開示の方法は、核酸誘導ヌクレアーゼを利用することができる。ここで使用されるような「核酸誘導ヌクレアーゼ」は、DNA、RNA または DNA / RNA ハイブリッドを切断するあらゆるヌクレアーゼであり、特異性を付与するために一つまたは複数のガイドヌクレアーゼ (g N A) を使用する。核酸誘導ヌクレアーゼは、C R I S P R / C a s システムタンパク質ならびに非 C R I S P R / C a s システムタンパク質を含む。ここで提供される核酸誘導ヌクレアーゼは、DNA 誘導 DNA ヌクレアーゼ；DNA 誘導 RNA ヌクレアーゼ、RNA 誘導 DNA ヌクレアーゼ；または RNA 誘導 RNA ヌクレアーゼであることができる。ヌクレアーゼはエンドヌクレアーゼであることができる。ヌクレアーゼはエキソヌクレアーゼであることができる。一実施態様では、核酸誘導ヌクレアーゼは、核酸誘導 DNA エンドヌクレアーゼである。一実施態様では、核酸誘導ヌクレアーゼは、核酸誘導 RNA エンドヌクレアーゼである。

【0472】

いくつかの実施態様では、本開示の修飾を基礎とする濃縮方法および核酸誘導ヌクレアーゼを基礎とする濃縮方法は、サンプル内の異なる核酸を枯渇させ、それにより単独のいずれかのアプローチよりも目的の核酸についてより高い程度の濃縮を達成する。

【0473】

ここで提供されるのは、複数のガイド核酸 (g N A) (交換可能にライブラリー、またはコレクションとも言う) である。いくつかの実施態様では、正規化対照は、複数の g N A 内のあらゆる g N A の配列と同一または高度に類似する配列を含まない。

【0474】

「ガイド核酸」の用語は、核酸誘導ヌクレアーゼ、および任意に付加的な核酸と複合体を形成することができるガイド核酸 (g N A) を言う。g N A は、単離された核酸として、または核酸誘導ヌクレアーゼー g N A 複合体、例えば C a s 9ー g R N A 複合体の一部として存在してよい。

【0475】

ここで使用されるように、複数の g N A は、少なくとも 10^2 の固有の g N A を含む g N A の混合物を示す。いくつかの実施態様では、複数の g N A は、少なくとも 10^2 の固有の g N A、少なくとも 10^3 の固有の g N A、少なくとも 10^4 の固有の g N A、少なくとも 10^5 の固有の g N A、少なくとも 10^6 の固有の g N A、少なくとも 10^7 の固有の g N A、少なくとも 10^8 の固有の g N A、少なくとも 10^9 の固有の g N A または少なくとも 10^{10} の固有の g N A を含む。いくつかの実施態様では、g N A のコレクションは、合計で少なくとも 10^2 の固有の g N A、少なくとも 10^3 の固有の g N A、少なくとも 10^4 の固有の g N A または少なくとも 10^5 の固有の g N A を含む。

【0476】

いくつかの実施態様では、g N A のコレクションは、標的配列を含む第一の N A セグメント；および核酸誘導ヌクレアーゼシステム (例えば、C R I S P R / C a s システム)

タンパク質結合配列を含む第二のNAセグメントを含む。いくつかの実施態様では、第一および第二のセグメントは、5'から3'までの順序である。いくつかの実施態様では、第一および第二のセグメントは、3'から5'までの順序である。

【0477】

いくつかの実施態様では、第一のセグメントのサイズは、複数のgNAにわたって、12~100bp、12~75bp、12~50bp、12~30bp、12~25bp、12~22bp、12~20bp、12~18bp、12~16bp、14~250bp、14~100bp、14~75bp、14~50bp、14~30bp、14~25bp、14~22bp、14~20bp、14~18bp、14~17bp、14~16bp、15~250bp、15~100bp、15~75bp、15~50bp、15~30bp、15~25bp、15~22bp、15~20bp、15~18bp、15~17bp、15~16bp、16~250bp、16~100bp、16~75bp、16~50bp、16~30bp、16~25bp、16~22bp、16~20bp、16~18bp、16~17bp、17~250bp、17~100bp、17~75bp、または17~50bp、17~30bp、17~25bp、17~22bp、17~20bp、17~18bp、18~250bp、18~100bp、18~75bp、18~50bp、18~30bp、18~25bp、18~22bp、18~20bp、19~250bp、19~100bp、19~75bp、または19~50bp、19~30bp、19~25bp、または19~22bpで変化する。いくつかの特別な実施態様では、第一のセグメントのサイズは、15bp、16bp、17bp、18bp、19bp、または20bpである。

【0478】

いくつかの実施態様では、複数の第一のセグメントの少なくとも10%、または少なくとも15%、または少なくとも20%、または少なくとも25%、または少なくとも30%、または少なくとも35%、または少なくとも40%、または少なくとも45%、または少なくとも50%、または少なくとも55%、または少なくとも60%、または少なくとも65%、または少なくとも70%、または少なくとも75%、または少なくとも80%、または少なくとも85%、または少なくとも90%、または少なくとも95%、または100%は、15~50bpである。

【0479】

いくつかの実施態様では、複数のgNAは、枯渇の標的となる核酸内の標的配列と塩基対になることができる標的配列を含み、枯渇の標的となる核酸内の標的配列は、サンプル内の枯渇の標的となるゲノムまたはトランスクリプトームにわたって、少なくとも1bpごとに、少なくとも2bpごとに、少なくとも3bpごとに、少なくとも4bpごとに、少なくとも5bpごとに、少なくとも6bpごとに、少なくとも7bpごとに、少なくとも8bpごとに、少なくとも9bpごとに、少なくとも10bpごとに、少なくとも11bpごとに、少なくとも12bpごとに、少なくとも13bpごとに、少なくとも14bpごとに、少なくとも15bpごとに、少なくとも16bpごとに、少なくとも17bpごとに、少なくとも18bpごとに、少なくとも19bpごとに、20bp、少なくとも25bpごとに、少なくとも30bpごとに、少なくとも40bpごとに、少なくとも50bpごとに、少なくとも100bpごとに、少なくとも200bpごとに、少なくとも300bpごとに、少なくとも400bpごとに、少なくとも500bpごとに、少なくとも600bpごとに、少なくとも700bpごとに、少なくとも800bpごとに、少なくとも900bpごとに、少なくとも1000bpごとに、少なくとも2500bpごとに、少なくとも5000bpごとに、少なくとも10,000bpごとに、少なくとも15,000bpごとに、少なくとも20,000bpごとに、少なくとも25,000bpごとに、少なくとも50,000bpごとに、少なくとも100,000bpごとに、少なくとも250,000bpごとに、少なくとも500,000bpごとに、少なくとも750,000bpごとにまたは少なくとも1,000,000bpごとに離間されている。

【0480】

いくつかの実施態様では、複数の gNA は、標的配列を含む第一の NA セグメントと、核酸誘導ヌクレアーゼシステム（例えば、CRISPR/Cas システム）タンパク質結合配列を含む第二の NA セグメントとを含み、複数の gNA は、核酸誘導ヌクレアーゼシステム（例えば、CRISPR/Cas システム）のタンパク質メンバーに対する多様な特異性を有する多様な第二の NA セグメントを有することができる。例えば、ここに提供されるような gNA のコレクションは、第二のセグメントが、第一の核酸誘導ヌクレアーゼシステム（例えば、CRISPR/Cas システム）タンパク質に特異的な核酸誘導ヌクレアーゼシステム（例えば、CRISPR/Cas システム）タンパク質結合配列を含むメンバーを含み；また、第二のセグメントが、第二の核酸誘導ヌクレアーゼシステム（例えば、CRISPR/Cas システム）タンパク質に特異的な核酸誘導ヌクレアーゼシステム（例えば、CRISPR/Cas システム）タンパク質結合配列を含むメンバーを含むことができ、ここで、第一および第二の核酸誘導ヌクレアーゼシステム（例えば、CRISPR/Cas システム）タンパク質は同じではない。いくつかの実施態様では、ここに提供されるような複数の gNA は、Cas9 タンパク質および Cpf1、Cas3、Cas8a-c、Cas10、CasX、CasY、Cas13、Cas14、Cse1、Csy1、Csn2、Cas4、Csm2、および Cm5 からなるグループから選択される他のタンパク質に特異性を示すメンバーを含む。標的配列を含む第一の NA セグメントと核酸誘導ヌクレアーゼシステムタンパク質結合配列を含む第二の NA セグメントとの順序は、核酸誘導ヌクレアーゼシステムタンパク質に依存する。第一と第二の NA セグメントとの適切な 5' から 3' までの配置および核酸誘導ヌクレアーゼシステムタンパク質の選択は、当業者に理解される。

【0481】

いくつかの実施態様では、gNA は、DNA および RNA を含む。いくつかの実施態様では、gNA は、DNA (gDNA) からなる。いくつかの実施態様では、gNA は、RNA (gRNA) からなる。

【0482】

いくつかの実施態様では、gNA は、gRNA を含み、gRNA は、crRNA と tracrRNA とをコードする二つのサブセグメントを含む。いくつかの実施態様では、crRNA は、標的配列に加えて、tracrRNA とハイブリダイズすることができる余剰の配列を含まない。いくつかの実施態様では、crRNA は、tracrRNA とハイブリダイズすることができる余剰の配列を含む。いくつかの実施態様では、二つのサブセグメントは、独立して転写される。いくつかの実施態様では、二つのサブセグメントは、単一ユニットとして転写される。いくつかの実施態様では、crRNA をコードする DNA は、配列 G T T T T A G A G C T A T G C T G T T T T G (配列番号：1) の 5' の標的配列を含む。いくつかの実施態様では、tracrRNA をコードする DNA は、配列 G G A A C C A T T C A A A A C A G C A T A G C A A G T T A A A A T A A G G C T A G T C C G T T A T C A A C T T G A A A A A G T G G C A C C G A G T C G G T G C T T T T T T T (配列番号：2) を含む。

【0483】

ここで使用されるように、標的配列は、サンプル内の枯渇の標的となる核酸内の標的配列に gNA が向かう配列である。例えば、標的配列は、ここに記載された非ホスト配列のいずれかを標的とする。

【0484】

ここに提供されるのは、標的配列を含むセグメントを含む gNA および複数の gNA である。

【0485】

いくつかの実施態様では、標的配列は、DNA を含むかまたは DNA からなる。いくつかの実施態様では、標的配列は、RNA を含むかまたは RNA からなる。

【0486】

いくつかの実施態様では、標的配列は、RNAを含み、かつ目的の配列のPAM配列の5'の配列に対して少なくとも70%の配列同一性、少なくとも75%の配列同一性、少なくとも80%の配列同一性、少なくとも85%の配列同一性、少なくとも90%の配列同一性、少なくとも95%の配列同一性、または100%の配列同一性を共有するが、ただし、RNAは、チミンの代わりにウラシルを含む。いくつかの実施態様では、標的配列は、RNAを含み、かつ目的の配列のPAM配列の3'の配列に対して少なくとも70%の配列同一性、少なくとも75%の配列同一性、少なくとも80%の配列同一性、少なくとも85%の配列同一性、少なくとも90%の配列同一性、少なくとも95%の配列同一性、または100%の配列同一性を共有するが、ただし、RNAは、チミンの代わりにウラシルを含む。いくつかの実施態様では、PAM配列は、AGG、CGG、TGG、GGGまたはNAGである。いくつかの実施態様では、PAM配列は、TTN、TCNまたはTGNである。

【0487】

いくつかの実施態様では、標的配列は、RNAを含み、かつPAM配列の5'のヌクレオチドの配列の反対側の鎖に相補的である。いくつかの実施態様では、標的配列は、PAM配列の5'のヌクレオチドの配列の反対側の鎖に、少なくとも70%相補的、少なくとも75%相補的、少なくとも80%相補的、少なくとも85%相補的、少なくとも90%相補的、少なくとも95%相補的、または100%相補的である。いくつかの実施態様では、標的配列は、RNAを含み、かつPAM配列の3'のヌクレオチドの配列の反対側の鎖に相補的である。いくつかの実施態様では、標的配列は、PAM配列の3'のヌクレオチドの配列の反対側の鎖に、少なくとも70%相補的、少なくとも75%相補的、少なくとも80%相補的、少なくとも85%相補的、少なくとも90%相補的、少なくとも95%相補的、または100%相補的である。いくつかの実施態様では、PAM配列は、AGG、CGG、TGG、GGGまたはNAGである。いくつかの実施態様では、PAM配列は、TTN、TCNまたはTGNである。

【0488】

異なるCRISPR/Casシステムタンパク質は、異なるPAM配列を認識する。PAM配列は、標的配列の5'または3'に配置することができる。例えば、Cas9は、標的配列の隣接する3'末端に位置するNGG PAMを認識することができる。Cpf1は、標的配列の隣接する5'末端に位置するTTN PAMを認識することができる。全てのCRISPR/Casシステムタンパク質により認識される全てのPAM配列は、本開示の範囲内にあると認識される。どのPAM配列が、特別のCRISPR/Casシステムタンパク質と適合するかは、当業者に容易に理解される。

【0489】

ここに提供されるのは、核酸誘導ヌクレアーゼタンパク質結合配列を含むセグメントを含むgNAおよび複数のgNAである。核酸誘導ヌクレアーゼは、核酸誘導ヌクレアーゼシステムタンパク質（例えば、CRISPR/Casシステム）であることができる。核酸誘導ヌクレアーゼシステムは、RNA誘導ヌクレアーゼシステムであることができる。核酸誘導ヌクレアーゼシステムは、DNA誘導ヌクレアーゼシステムであることができる。

【0490】

核酸誘導ヌクレアーゼタンパク質結合配列は、核酸誘導ヌクレアーゼシステムのあらゆるタンパク質メンバーに結合する核酸配列である。例えば、CRISPR/Casタンパク質結合配列は、CRISPR/Casシステムのあらゆるタンパク質メンバーに結合する核酸配列である。

【0491】

いくつかの実施態様では、CRISPR/Casシステムタンパク質は、あらゆる細菌種または古細菌種に由来することができる。いくつかの実施態様では、CRISPR/Casシステムタンパク質は、単離され、組換えにより製造、または合成されている。いくつかの実施態様では、CRISPR/Casシステムタンパク質の例は、天然に存在するかまたは操作された変種である。いくつかの実施態様では、核酸誘導ヌクレアーゼシステム

タンパク質（例えば、C R I S P R / C a s システムタンパク質）は、あらゆる細菌種または古細菌種に由来することができる。いくつかの実施態様では、天然に存在する C R I S P R / C a s システムタンパク質は、C A S クラス I タイプ I、I I I、もしくは I V、または C A S クラス I I タイプ I I もしくは V に属することができ、C a s 9、C a s 3、C a s 8 a - c、C a s 1 0、C a s X、C a s Y、C a s 1 3、C a s 1 4、C s e 1、C s y 1、C s n 2、C a s 4、C s m 2、C m r 5、C s f 1、C 2 c 2、および C p f 1 を含むことができる。例示的な実施態様では、C R I S P R / C a s システムタンパク質は、C a s 9 を含む。例示的な実施態様では、C R I S P R / C a s システムタンパク質は、C p f 1 を含む。

【0492】

いくつかの実施態様では、核酸誘導ヌクレアーゼシステムタンパク質（例えば、C R I S P R / C a s システムタンパク質）は、ストレプトコッカス パイオゲネス、スタフィロコッカス アウレウス、ナイセリア メニンギチジス、ストレプトコッカス サーモフィルス、トレポネーマ デンティコラ、フランシセラ ツラレンシス、パスツレラ ムルトシダ、カンピロバクター ジェジュニ、カンピロバクター ラリ、マイコプラズマ ガリナルム、ニトラティフラクター サルスギニス、パルビバクルム ラバメンチボランス、ロゼブリア インテスチナリス、ナイセリア シネレア、グルコンアセトバクター ジアゾトロフィクス、アゾスピリルム、スフェロチャエタ グロブス、フラボバクテリウム コルムナレ、フルビコラタフェンシス、バクテロイデス コプロフィルス、マイコプラズマ モビレ、ラクトバシルス ファルシミニス、ストレプトコッカス パストゥリアヌス、ラクトバシルス ジョンソニ、スタフィロコッカス シュードインターメディウス、フィリファクタ アロシス、レジオネラ ニューモフィラ、ステレラ ワズボルテンシス、コリネバクター ジフテリア、アシドアミノコックス、ラクノスピラセア細菌またはプレボテラからの核酸誘導ヌクレアーゼシステムタンパク質（例えば C R I S P R / C a s システムタンパク質）に由来するかまたはそれらから誘導される。

【0493】

いくつかの実施態様では、核酸誘導ヌクレアーゼシステムタンパク質結合配列は、g N A（例えば、g R N A）ステムループ配列を含む。異なる C R I S P R / C a s システムタンパク質は、異なる核酸誘導ヌクレアーゼシステムタンパク質結合配列と適合可能である。どの P C R I S P R / C a s システムタンパク質が、どの核酸誘導ヌクレアーゼシステムタンパク質結合配列と適合可能であるかは、当業者に容易に容易に理解される。

【0494】

いくつかの実施態様では、g N A（例えば、g R N A）ステムループ配列をコードする二本鎖 DNA 配列は、一方の鎖に以下の DNA 配列（5'>3'、G T T T T T A G A G C T A G A A A T A G C A A G T T A A A A T A A G G C T A G T C C G T T A T C A A C T T G A A A A A G T G G C A C C G A G T C G G T G C T T T T T T T T）（配列番号：3）と、他方の鎖にその逆相補的 DNA（5'>3'、A A A A A A A A G C A C C G A C T C G G T G C C A C T T T T T C A A G T T G A T A A C G G A C T A G C C T T A T T T T A A C T T G C T A T T T C T A G C T C T A A A A C）（配列番号：4）を含む。

【0495】

いくつかの実施態様では、g N A（例えば、g R N A）ステムループ配列をコードする一本鎖 DNA 配列は、以下の DNA 配列：（5'>3'、A A A A A A A A G C A C C G A C T C G G T G C C A C T T T T T C A A G T T G A T A A C G G A C T A G C C T T A T T T T A A C T T G C T A T T T C T A G C T C T A A A A C）（配列番号：4）を含み、この一本鎖 DNA は、転写テンプレートとして機能する。

【0496】

いくつかの実施態様では、g N A（例えば、g R N A）ステムループ配列は、以下の R N A 配列：（5'>3'、G U U U U A G A G C U A G A A A U A G C A A G U U A A A A U A A G G C U A G U C C G U U A U C A A C U U G A A A A A G U G G C A C C G A G U C G G U G C U U U U U U U U）（配列番号：5）を含む。

【0497】

いくつかの実施態様では、gNA（例えば、gRNA）ステムループ配列をコードする二本鎖DNA配列は、一方の鎖に以下のDNA配列（5'>3'、GTTTTTAGAGCTATGCTGGAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCAGGTCGGTGTCTTTT TTTC）（配列番号：6）と、他方の鎖にその逆相補的DNA（5'>3'、GAAAAA AAGCACCCGACTCGGTGCCACTTTTTTCAAGTTGATAACGGG ACTAGCCTTATTTTAACTTGCTATGCTGTTTCCAGCATAGC TCTAAAAC）（配列番号：7）を含む。

【0498】

いくつかの実施態様では、gNA（例えば、gRNA）ステムループ配列をコードする一本鎖DNA配列は、以下のDNA配列：（5'>3'、GAAAAAAAGCACCCG A CTCGGTGCCACTTTTTTCAAGTTGATAACGGGACTAGCCTTA TTTTAACTTGCTATGCTGTTTCCAGCATAGCTCTAAAAC）（配列番号：7）を含み、この一本鎖DNAは、転写テンプレートとして機能する。

【0499】

いくつかの実施態様では、gNA（例えば、gRNA）ステムループ配列は、以下のRNA配列：（5'>3'、GUUUUAGAGCUAUGCUGGAAACAGCAUAG CAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAG UGGCACCCGAGUCGGUGCUUUUUUUUC）（配列番号：8）を含む。

【0500】

いくつかの実施態様では、CRISPR/Casシステムタンパク質は、Cpf1タンパク質である。いくつかの実施態様では、Cpf1タンパク質は、フランシセラ種またはアシドアミノコックス種から単離または誘導される。いくつかの実施態様では、gNA（例えば、gRNA）CRISPR/Casシステムタンパク質結合配列は、以下のRNA配列：（5'>3'、AAUUUCUACUGUUGUAGAU）（配列番号：9）を含む。

【0501】

いくつかの実施態様では、CRISPR/Casシステムタンパク質は、Cpf1タンパク質である。いくつかの実施態様では、Cpf1タンパク質は、フランシセラ種またはアシドアミノコックス種から単離または誘導される。いくつかの実施態様では、gNA（例えば、gRNA）CRISPR/Casシステムタンパク質結合配列をコードするDNA配列は、以下のDNA配列：（5'>3'、AATTTCTACTGTTGTAGAT）（配列番号：10）を含む。いくつかの実施態様では、DNAは一本鎖である。いくつかの実施態様では、DNAは二本鎖である。

【0502】

いくつかの実施態様では、ここで提供されるのは、標的配列を含む第一のNAセグメントと、核酸誘導ヌクレアーゼ（例えば、CRISPR/Cas）システムタンパク質結合配列を含む第二のNAセグメントとを含むgNA（例えばgRNA）である。いくつかの実施態様では、第一のセグメントのサイズは、15bp、16bp、17bp、18bp、19bpまたは20bpである。いくつかの実施態様では、第二のセグメントは、gRNAステムループ配列を含む単一セグメントを含む。いくつかの実施態様では、gRNAステムループ配列は、以下のRNA配列：（5'>3'、GUUUUAGAGCUAGAA AUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCCGAGUCGGUGCUUUUUUUU）（配列番号：5）を含む。いくつかの実施態様では、gRNAステムループ配列は、以下のRNA配列：（5'>3'、GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAA UAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCCGAG UCGGUGCUUUUUUUUC）（配列番号：8）を含む。いくつかの実施態様では、第二のセグメントは、二つのサブセグメント：第二のRNAサブセグメント（tracr

10

20

30

40

50

RNA)とハイブリッドを形成する第一のRNAサブセグメント(c r RNA)を含み、これらは一緒になって、核酸誘導ヌクレアーゼ(例えば、CRISPR/Cas)システムタンパク質結合に向かうように作用する。いくつかの実施態様では、第二のサブセグメントの配列は、GUUUUAGAGCUAUGCUGUUUUUを含む。いくつかの実施態様では、第一のRNAセグメントと第二のRNAセグメントとは、一緒にc r RNA配列を形成する。いくつかの実施態様では、第二のRNAセグメントとハイブリッドを形成する他のRNAは、tracrRNAである。いくつかの実施態様では、tracrRNAは、: 5'>3', GGAACCAUUC AAAACAGCAUAGCAAGUUAAA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCAACCGA GUCGGUGCUUUUUUU (配列番号: 11)の配列を含む。

10

【0503】

いくつかの実施態様では、ここに提供されるのは、標的配列を含む第一のNAセグメントと、核酸誘導ヌクレアーゼ(例えば、CRISPR/Cas)システムタンパク質結合配列を含む第二のNAセグメントを含むgNA(例えば、gRNA)である。いくつかの実施態様では、例えば、CRISPR/Casシステムタンパク質がCpf1システムタンパク質である実施態様では、第二のセグメントは、第一のセグメントの5'である。いくつかの実施態様では、第一のセグメントのサイズは、20bpである。いくつかの実施態様では、第一のセグメントのサイズは、20bpより大きい。いくつかの実施態様では、第一のセグメントのサイズは、30bpより大きい。いくつかの実施態様では、第二のセグメントは、gRNAステムループ配列を含む単一セグメントを含む。いくつかの実施態様では、gRNAステムループ配列は、以下のRNA配列:(5'>3', AAUUUC UACUGUUGUAGA U)(配列番号9)を含む。

20

【0504】

CRISPR/Casシステムタンパク質は、野生型CRISPR/Casシステムタンパク質に対して少なくとも60%同一(例えば、少なくとも70%、少なくとも80%、または90%同一、少なくとも95%同一または少なくとも98%同一または少なくとも99%同一)であってよい。CRISPR/Casシステムタンパク質は、野生型CRISPR/Casシステムタンパク質の全ての機能を有するか、または結合活性およびヌクレアーゼ活性を含む機能の一つだけまたはいくつかを有してよい。

【0505】

「CRISPR/Casシステムタンパク質会合ガイドNA」の用語は、ガイドNA(gNA)を言う。CRISPR/Casシステムタンパク質会合ガイドNAは、単離されたNAとして、またはCRISPR/Casシステムタンパク質-gNA複合体の一部として存在してよい。

30

【0506】

いくつかの実施態様では、CRISPR/Casシステムタンパク質は、RNA誘導DNAヌクレアーゼである。いくつかの実施態様では、CRISPR/Casシステムタンパク質により切断されるDNAは二本鎖である。二本鎖DNAを切断する例示的なRNA誘導DNAヌクレアーゼは、Cas9、Cpf1、CasXおよびCasYを含むが、これらの限定されるものではない。更なる例示的なRNA誘導DNAヌクレアーゼは、Cas10、Csm2、Csm3、Csm4、およびCsm5を含む。いくつかの実施態様では、Cas10、Csm2、Csm3、Csm4、およびCsm5は、gRNAとリボ核タンパク質複合体を形成する。いくつかの実施態様では、CRISPR/Casシステムタンパク質核酸誘導ヌクレアーゼは、Cas9であるかまたはCas9を含む。本開示のCas9は、単離、組換えによる製造、または合成することができる。いくつかの実施態様では、Cas9タンパク質は、熱安定性である。この実施態様で 사용할 ことができるCas9タンパク質の例は、F.A. Ran, L. Cong, W.X. Yan, D. A. Scott, J.S. Gootenberg, A.J. Kriz, B. Zetsche, O. Shalem, X. Wu, K.S. Makarova, E.V. Koonin, P.A. Sharp,およびF. Zhang; "In vivo genome editing using Staphylococcus aureus Cas9," Nature 520, 186-191 (09 April 2015) doi:10.1038/nature14299に見るこ

40

50

とができ、これは、参照により本明細書に組み込まれる。いくつかの実施態様では、C a s 9 は、ストレプトコッカス パイオゲネス、スタフィロコッカス アウレウス、ナイセリア メニンギチジス、ストレプトコッカス サーモフィルス、トレポネーマ デンティコラ、フランシセラ ツラレンシス、パスツレラ ムルトシダ、カンピロバクター ジェジュニ、カンピロバクター ラリ、マイコプラズマ ガリナルム、ニトラティフラクター サルスギニス、パルビバクulum ラバメンチボランス、ロゼブリア インテスチナリス、ナイセリア シネレア、グルコンアセトバクター ジアゾトロフィクス、アゾスピリウム、スフェロチャエタ グロブス、フラボバクテリウム コルムナレ、フルビコラ タフェンシス、バクテロイデスコプロフィルス、マイコプラスマ モビレ、ラクトバシルス ファルシミニス、ストレプトコッカス パストゥリアヌス、ラクトバシルス ジョンソニ、スタフィロコッカス シュードインターメディウス、フィリファクタ アロシス、レジオネラ ニューモフィラ、ステレラ ワズボルテンシス、またはコリネバクター ジフテリアから誘導されるタイプII C R I S P R システムである。

【0507】

いくつかの実施態様では、C a s 9 は、S. パイオゲネスから誘導されたタイプII C R I S P R システムであり、P A M 配列は、標的特異的ガイド配列の隣接する3'末端に位置するN G Gである。例示的な細菌種由来のタイプII C R I S P R システムのP A M 配列は、ストレプトコッカス パイオゲネス(N G G)、スタフィロコッカス アウレウス(N N G R R T)、ナイセリア メニンギチジス(N N N G A T T)、ストレプトコッカス サーモフィルス(N N A G A A) およびトレポネーマ デンティコラ(N A A A A C) も含み、これら全ては、本開示から逸脱することなく使用可能である。

【0508】

例示的な一実施態様では、C a s 9 配列は、例えば、p X 3 3 0 プラスミド(Addgene から入手可能)から得られ、P C R により再増幅され、次いでp E T 3 0 (EMD biosciencesから)内ヘクローニングして、細菌内で発現させ、組換え6 H i s タグ付きタンパク質を精製することができる。

【0509】

「C a s 9 - g N A 複合体」は、C a s 9 タンパク質とガイドN A とを含む複合体を言う。C a s 9 タンパク質は、野生型C a s 9 タンパク質、例えばストレプトコッカス パイオゲネスC a s 9 タンパク質に対して少なくとも60%同一(例えば、少なくとも70%、少なくとも80%、または90%同一、少なくとも95%同一または少なくとも98%同一または少なくとも99%同一)であってよい。C a s 9 タンパク質は、野生型C a s 9 タンパク質の全ての機能を有するか、または結合活性、ヌクレアーゼ活性、およびヌクレアーゼ活性を含む機能の一つだけまたはいくつかを有してよい。

【0510】

「C a s 9 会合ガイドN A」の用語は、上述のようなガイドN A を言う。C a s 9 会合ガイドN A は、単離されて、またはC a s 9 - g N A 複合体の一部として存在してよい。

【0511】

いくつかの実施態様では、C R I S P R / C a s システムタンパク質核酸誘導ヌクレアーゼは、C p f 1 システムタンパク質であるかまたはC p f 1 システムタンパク質を含む。本開示のC p f 1 システムタンパク質は、単離、組換えによる製造、または合成することができる。いくつかの実施態様では、C p f 1 タンパク質は、熱安定性である。

【0512】

C p f 1 システムタンパク質は、クラスII、タイプV C R I S P R システムタンパク質である。いくつかの実施態様では、C p f 1 タンパク質は、フランシセラ ツラレンシスから単離または誘導される。いくつかの実施態様では、C p f 1 タンパク質は、アシドアミノコックス、ラクノスピラセア細菌またはプレボテラから単離または誘導される。

【0513】

C p f 1 システムタンパク質は、核酸誘導ヌクレアーゼシステムタンパク質結合配列(例えば、ステムループ)と標的配列とを含む単一のガイドR N A に結合する。C p f 1 標

10

20

30

40

50

的配列は、標的核酸内のCpf1 PAM配列の3'に隣接するように位置する配列を含む。Cas9とは異なり、Cpf1核酸誘導ヌクレアーゼシステムタンパク質結合配列は、Cpf1 gRNA内で標的配列の5'に位置する。Cpf1は、標的核酸内で平滑末端切断ではなく、付着末端を製造することができる。Cpf1タンパク質-gRNAタンパク質複合体を標的核酸に標的とした後、フランシセラ由来のCpf1は、例えば、標的核酸を付着末端に切断し、標的配列の3'末端でPAMから18~23塩基離れて約5ヌクレオチドの5'オーバーハングを生成する。対照的に、野生型Cas9による切断は、Cas9 PAMの3ヌクレオチド上流に平滑末端を製造する。

【0514】

例示的Cpf1 gRNAステムループ配列は、以下のRNA配列：(5'>3', AAU UUCUACUGUUGUAGA U) (配列番号：9)を含む。

【0515】

「Cpf1タンパク質-gNA複合体」は、Cpf1タンパク質とガイドNA (例えば、gRNA)とを含む複合体を言う。gNAが、gRNAである場合、gRNAは、単一分子、すなわち標的とハイブリダイズし、配列特異性を提供する一つのRNA (「crRNA」) からなっていてよい。

【0516】

Cpf1タンパク質は、野生型Cpf1タンパク質に対して少なくとも60%同一 (例えば、少なくとも70%、少なくとも80%、または90%同一、少なくとも95%同一または少なくとも98%同一または少なくとも99%同一) であってよい。Cpf1タンパク質は、野生型Cpf1タンパク質の全ての機能を有するか、または結合活性およびヌクレアーゼ活性を含む機能の一つだけまたはいくつかを有してよい。

【0517】

Cpf1システムタンパク質は、多様なPAM配列を認識する。Cpf1システムタンパク質により認識される例示的PAMは、TTN、TCNおよびTGNを含むが、これらに限定されるものではない。付加的Cpf1 PAM配列は、TTTNを含むが、これに限定されるものではない。Cpf1 PAM配列の一つの特徴は、Cas9タンパク質により使用されるNGGまたはNAG PAM配列よりも高いA/T含有率を有する。

コンピューターシステムおよびソフトウェア

【0518】

ここに記載された方法は、コンピューターシステムの文脈において、またはコンピューター可読記憶媒体内に記憶されたソフトウェアもしくはコンピューター実行可能な命令の一部として使用されてよい。

【0519】

いくつかの実施態様では、システム (例えば、コンピューターシステム) は、本発明の実施態様のいくつかの特定の特徴を実行するために使用されてよい。

【0520】

特定の実施態様では、システムは、一つ以上のメモリおよび/または記憶装置を含んでよい。メモリおよび記憶装置は、本発明の多様な実施態様の少なくとも一部を実行するコンピューター実行可能な命令を記憶してよい一つまたは複数のコンピューター可読記憶媒体であってよい。

【0521】

ここに提供されているのは、複数の正規化対照ポリヌクレオチド配列を設計するためのシステムである。いくつかの実施態様では、システムは、(i) 少なくとも一つの参照配列をインポートする命令；(ii) 少なくとも一つの参照配列から複数の参照配列フラグメントを生成する命令；(iii) 参照配列フラグメントの数の関数として、少なくとも一つのパラメーターの分布を生成する命令；(iv) 分布をビンに分割する命令；(v) ビンの少なくともサブセットから複数の参照配列フラグメントを選択する命令；および(vi) 複数の参照配列フラグメントをシャッフルして、シャッフル配列を生成する命令を含み、それにより複数の正規化ポリヌクレオチド配列を生成するコンピューター実行可能

10

20

30

40

50

な命令を記憶するコンピューター可読記憶媒体を含む。いくつかの実施態様では、システムは、(i) 少なくとも一つの参照配列をインポートする命令；(ii) 少なくとも一つの参照配列から複数の参照配列フラグメントを生成する命令；(iii) 参照配列フラグメントの数の関数として、少なくとも一つのパラメーターの分布を生成する命令；(iv) 分布をビンに分割する命令；(v) ビンの少なくとも一つのサブセットの各々から少なくとも二つの参照配列フラグメントを選択する命令、ここで、少なくとも二つの参照配列フラグメントは、参照配列内で非連続的であるか、または異なる参照配列に由来する；および(vi) 各ビンからの少なくとも二つの参照配列フラグメントを連結する命令を含み、それにより複数の正規化ポリヌクレオチド配列を生成するコンピューター実行可能な命令を記憶するコンピューター可読記憶媒体を含む。いくつかの実施態様では、システムはさらに、(a) 少なくとも一つ一つの参照配列を含む入力ファイルのセットを受信すること；および(b) コンピューター可読記憶媒体内に記憶されたコンピューター実行可能な命令を実行することを含むステップを実行するように構成されたプロセッサを含む。いくつかの実施態様では、パラメーターは、(1) パーセントGC含有率、(2) エントロピー、(3) 複雑性、(4) EIEP、(5) 長さ、またはそれらの組合せの少なくとも一つを含む。

【0522】

コンピューティングシステムは、一つまたは複数の中央演算ユニット（「プロセッサ」）、メモリ、入力／出力デバイス、例えば、キーボードおよびポインティングデバイス、タッチデバイス、ディスプレイデバイス、ストレージデバイス、例えばディスクドライブ、およびネットワークアダプター、例えばネットワークインターフェースを含んでよく、これらは相互接続するように接続されてよい。相互接続は、適切なブリッジ、アダプター、またはコントローラーにより接続された、一つまたは複数の個別の物理バス、ポイントツーポイント接続、またはその両方を表す抽象概念である。したがって、相互接続は、例えば、システムバス、ペリフェラルコンポーネントインターコネクト（PCI）バスまたはPCI-エクスプレッスバス、ハイパートランスポートもしくはインダストリアルスタンダードアーキテクチャ（ISA）バス、スモールコンピューターシステムインターフェース（SCSI）バス、ユニバーサルシリアルバス（USB）、IIC（12C）バス、またはファイヤワイヤとも言われる米国電気電子技術者協会（IEEE）標準1394バスを含んでよい。

【0523】

さらに、データ構造およびメッセージ構造は、データ伝送媒体、例えば通信リンク上の信号を介して記憶または伝送されてよい。多様な通信リンクは、例えば、インターネット、ローカルエリアネットワーク、ワイドエリアネットワーク、またはポイントツーポイントダイアルアップ接続を使用してよい。したがって、コンピューター可読媒体は、コンピューター可読記憶媒体、例えば一時的でない媒体、およびコンピューター可読伝送媒体を含むことができる。

【0524】

メモリ内に記憶された命令は、上述の動作を実行するように一つまたは複数のプロセッサをプログラムするために、ソフトウェアおよび／またはファームウェアとして実施することができる。本発明のいくつかの実施態様では、このようなソフトウェアまたはファームウェアは、最初に、コンピューティングシステムを通して、例えばネットワークアダプターを介してリモートシステムからダウンロードすることにより処理システムに提供されてよい。

【0525】

ここに紹介された本発明の多様な実施態様は、例えば、プログラマブルな回路、例えばソフトウェアおよび／またはファームウェアでプログラムされた一つまたは複数のマイクロプロセッサ、完全に特殊な目的のハードワイヤード、つまり非プログラマブルな回路、またはこれらの形式の組合せにより実施することができる。特殊な目的のハードワイヤード回路は、例えば、一つまたは複数のASIC、PLD、FPGA等の形式であってよ

い。

【0526】

詳細な説明のいくつかの部分は、アルゴリズムの観点で表されていてよく、このアルゴリズムは、コンピューターメモリ内のデータビットに関する操作の記号表現であってよい。このアルゴリズムの記載および表現は、データ処理技術の当業者が、自分の仕事の内容を他の当業者に最も効果的に伝えるために使用する方法である。アルゴリズムは、ここで、および一般的に、所望の結果を導くための自己矛盾のない一連の操作であると考えられる。操作は、物理量の物理的な処理を必要とする操作である。必ずしもそうとは限らないが、通常では、これらの量は、記憶、伝送、結合、比較、および他の処理が可能である電氣的または磁氣的信号の形式を取る。時には、主に、一般的な慣例の理由から、これらの信号をビット、値、要素、記号、文字、用語、数字等と言うことが便利であることが判明した。

10

【0527】

ここに提示されるアルゴリズムおよび表示は、あらゆる特定のコンピューターまたは他の装置に本質的に関連していない。多様な汎用システムは、ここでの教示に従ったプログラムで使用されてよいが、またはいくつかの実施態様の方法を実施するより特殊な装置を構築することが便利であると判明してよい。

【0528】

さらに、実施態様は、完全に機能するコンピューターおよびコンピューターシステムの文脈で説明されたが、当業者は、多様な実施態様が、多様な形式のプログラム製品として配布することが可能であり、かつ本開示は、実際に配布を行うために使用される特定の形式の機械またはコンピューター可読媒体に関係なく同様に適用されることを理解する。

20

【0529】

機械可読記憶媒体、機械可読媒体、またはコンピューター可読（記憶）媒体の更なる例は、揮発性または不揮発性のメモリデバイス、フロッピーおよび他のリムーバブルディスク、ハードディスクドライブ、光ディスク（例えば、コンパクトディスクリードオンリーメモリ（CD ROM）、デジタル多用途ディスク（DVD）等）などのような記録可能型の媒体、ならびにデジタルおよびアナログの通信リンクのような伝送型の媒体を含むが、これらの限定されるものではない。

【0530】

キットおよび製品

本明細書に記載される正規化対照組成物のうちのいずれか1つ以上、および／または本明細書に記載される正規化対照組成物のうちのいずれかを製造もしくは使用するために使用される試薬を含むキットおよび製品が本明細書に提供される。いくつかの実施形態では、例示的なキットは、本明細書に記載されるDNA正規化対照組成物を含む。いくつかの実施形態では、例示的なキットは、本明細書に記載されるRNA正規化対照組成物を含む。いくつかの実施形態では、例示的なキットは、本明細書に記載されるDNAおよびRNA正規化対照組成物の組み合わせを含む。いくつかの実施形態では、キットは、正規化対照組成物を使用するための説明書、例えば、希釈、濃縮、および正規化対照をサンプルに添加するための説明書をさらに含む。

30

40

【0531】

本明細書に記載されるキットのいくつかの実施形態では、キットは、本開示の多検体対照のうちのいずれか1つ以上、およびそれらの使用のための試薬を含む。いくつかの実施形態では、キットは、多検体対照を使用するための説明書、例えば、希釈、濃縮、およびサンプルに多検体対照を添加するための説明書、またはサンプルと並行して多検体対照を処理するための説明書をさらに含む。本開示のキットと共に提供される例示的な多検体対照には、不活性化されている少なくとも3つの種の生物を含む多検体対照が含まれる。いくつかの実施形態では、キットは、多検体対照を希釈するための試薬をさらに含む。いくつかの実施形態では、キットは、多検体対照の使用のための説明書をさらに含む。

【0532】

50

いくつかの実施形態では、キットは、正規化対照、多検体対照、ならびに試薬およびそれらの使用のための説明書を含む。いくつかの実施形態では、キットは、陽性対照をさらに含む。例示的な陽性対照には、本明細書に記載される正規化対照および多検体対照について陽性対照または校正するために使用することができる、既知の濃度（例えば、 1 IU/mL 、 10 IU/mL 、 100 IU/mL 、 $1,000\text{ IU/mL}$ 、 $10,000\text{ IU/mL}$ 、 $100,000\text{ IU/mL}$ 、または $1,000,000\text{ IU/mL}$ ）の不活性化生物が挙げられる。いくつかの実施形態では、陽性対照は、ウイルス、細菌、真菌、単細胞真核生物、またはそれらの任意の組み合わせを含む。

【0533】

本明細書に記載されるように、NGSシーケンスの前に、枯渇を標的とする配列のサンプルを枯渇させる、または目的の配列のサンプルを濃縮するために使用される試薬のうちのいずれか1つ以上を含むキットおよび製品が本明細書に提供される。いくつかの実施形態では、キットは、アダプター、gNA（例えば、gRNAまたはgDNA）、gNA収集物（例えば、gRNAまたはgDNA多重性）、修飾感受性制限酵素、対照などに限定されない、本明細書に記載される組成物のうちのいずれか1つ以上を含む。

【0534】

例示的な実施形態では、キットは、gRNAを含み、gRNAは、本明細書に記載される任意の宿主ゲノムまたはcDNA配列を標的とする。さらなる例示的な実施形態では、キットは、gRNAを含み、gRNAは、ヒトゲノムまたはDNA配列の他の源を標的とする。

【0535】

本開示はまた、本明細書に記載されるように、ヌクレオチド修飾の違いを使用して、目的の核酸のサンプルを濃縮する方法を実施するためのすべての必須試薬および説明書を提供する。

【0536】

本開示はまた、正規化対照、多検体対照、枯渇を標的とする配列のサンプルを枯渇させるために使用される試薬、またはNGSシーケンスの前に目的の配列についてサンプルを濃縮するために使用される試薬、ならびに試薬およびそれらの使用のための説明書を含むキットを提供する。

【0537】

本明細書に提供される方法を使用してサンプルを濃縮する前後の情報を監視するコンピュータソフトウェアもまた本明細書に提供される。一例示的な実施形態では、ソフトウェアは、本明細書に記載される方法を適用する前後のサンプル中の枯渇を標的とする核酸の配列の豊富さを計算および報告し、オフターゲット枯渇のレベルを評価することができ、ソフトウェアは、本明細書に提供される濃縮方法を使用してサンプルを処理する前後の目的の配列の豊富さを比較することによって、標的化枯渇／濃縮／捕捉／分画／標識／調節／編集の有効性を確認することができる。

【0538】

本開示のキットのいくつかの実施形態では、キットは、正規化対照、多検体対照、サンプル濃縮および／または枯渇のための試薬、それらの使用のための説明書、ならびに許容される希釈剤などのそれらの使用のための試薬の両方を含む。

【0539】

本開示のキットのいくつかの実施形態では、多検体対照は、11種のウイルスの混合物を含むか、またはそれからなる。いくつかの実施形態では、ウイルスは、サイトメガロウイルス（CMV）、エプスタインバーウイルス（EBV）、アデノウイルス（ADV）、BKウイルス（BKV）、JCポリオマウイルス（JCV）、ヒトヘルペスウイルス6A（HHV6A）、ヒトヘルペスウイルス6B（HHV6B）、単純ヘルペスウイルス1型（HSV1）、単純ヘルペスウイルス2型（HSV2）、水痘帯状疱疹ウイルス（VZV）、およびヒトパルボウイルスB19（B19）からなる。

【0540】

本開示のキットのいくつかの実施形態では、多検体対照は、10種のウイルスの混合物を含むか、またはそれからなる。いくつかの実施形態では、ウイルスは、CMV、EBV、ADV、BKV、JCV、HHV6A、HHV6B、HSV1、HSV2、およびVZVからなる。

【0541】

本開示を通じて与えられるすべての最大数値制限は、かかるより低い数値制限が本明細書に明示的に書かれたかのように、すべてのより低い数値制限を含む。本開示を通じて与えられるすべての最小数値制限は、かかるより高い数値制限が本明細書に明示的に書かれたかのように、より高い数値制限すべてを含むであろう。本開示を通じて与えられるすべての数値範囲は、かかるより狭い数値範囲がすべて本明細書に明示的に書かれたかのように、かかるより広い数値範囲内に入るすべてのより狭い数値範囲を含むであろう。

10

【0542】

本明細書に開示される値は、列挙される正確な数値に厳密に限定されるものとして理解されるものではない。代わりに、別段の指定がない限り、そのような各値は、列挙された値およびその値を取り囲む機能的に等価な範囲の両方を意味することが意図される。例えば、「20 μm 」として開示される値は、「約20 μm 」を意味することが意図される。

【0543】

明示的に除外されないまたは別様に限定されない限り、任意の相互参照または関連する特許もしくは出願を含む本明細書に引用されるすべての文書は、参照によりその全体が本明細書に組み込まれる。任意の文書の引用は、本明細書に開示または特許請求される任意の発明に関して先行技術であること、またはそれが単独で、もしくは任意の他の参考文献（複数可）と任意の組み合わせで、任意のそのような発明を教示、示唆、または開示することを認めるものではない。さらに、本文書における用語の任意の意味または定義が、参照により組み込まれた文書における同一用語の任意の意味または定義と矛盾する場合には、本文書においてその用語に割り当てられた意味または定義が適用されるものとする。

20

【0544】

本開示の特定の実施形態は例示および記載されているが、様々な他の変更および修正は、本開示の主旨および範囲から逸脱することなく行われ得る。添付の特許請求の範囲は、本開示の範囲内のすべてのかかる変更および修正を含む。

【0545】

列挙される実施形態

本発明は、以下の列挙される例示的な実施形態を参照することによって定義され得る。

30

【0546】

1. ポリヌクレオチドの少なくとも3つの基を含む正規化対照であって、各グループ内のポリヌクレオチドが、同じ長さのものである、正規化対照。

【0547】

2. 正規化対照におけるポリヌクレオチドのすべての長さが、同じである、実施形態1に記載の正規化対照。

【0548】

3. 各グループ内のポリヌクレオチドが、任意の他のグループ内のポリヌクレオチドと比較した場合、異なる長さのものである、実施形態1に記載の正規化対照。

40

【0549】

4. ポリヌクレオチドのグループの長さが、線形配列または幾何学配列で配置される、実施形態3に記載の正規化対照。

【0550】

5. 少なくとも3つのグループ中のポリヌクレオチドの長さが、約15 bp～約50,000 bpの間である、実施形態1～4のいずれか1つに記載の正規化対照。

【0551】

6. 少なくとも3つのグループ中のポリヌクレオチドの長さが、約500 bp～約1500 bpの間である、実施形態1～4のいずれか1つに記載の正規化対照。

50

【0552】

7. 少なくとも3つのグループ中のポリヌクレオチドの長さが、約100bp～約1200bpの間である、実施形態1～4のいずれか1つに記載の正規化対照。

【0553】

8. 少なくとも3つのグループ中のポリヌクレオチドの長さが、約150bp～約600bpの間である、実施形態1～4のいずれか1つに記載の正規化対照。

【0554】

9. 少なくとも3つのグループが、175bp、250bp、および450bp、192bp、250bp、および450bp、200bp、300bp、および500bp、217bp、300bp、および517bp、436bp、552bp、および974bp、450bp、612bp、および1034bp、510bp、626bp、および1048bp、ならびに450bp、612bp、および1034bpから選択される長さを有するポリヌクレオチドを含む、実施形態1～4のいずれか1つに記載の正規化対照。

10

【0555】

10. 各グループ内のポリヌクレオチドが、同じ配列を含む、実施形態1～9のいずれか1つに記載の正規化対照。

【0556】

11. 各グループ内のポリヌクレオチドが、同じ配列を含まない、実施形態1～9のいずれか1つに記載の正規化対照。

【0557】

12. ポリヌクレオチドの各グループが、ポリヌクレオチドの少なくとも3つのサブグループを含み、

20

【0558】

各サブグループ内のポリヌクレオチドが、同じ配列を含み、各サブグループ内のポリヌクレオチドが、任意の他のサブグループと同じ配列を含まない、実施形態11に記載の正規化対照。

【0559】

13. ポリヌクレオチドの少なくとも1つグループ中の少なくとも1つの配列が、コンポーネント配列を含む、実施形態10～12のいずれか1つに記載の正規化対照。

【0560】

14. すべてのポリヌクレオチドのグループ中のすべての配列が、コンポーネント配列を含む、実施形態10～12のいずれか1つに記載の正規化対照。

30

【0561】

15. ポリヌクレオチドの各グループのコンポーネント配列が、任意の他の基のコンポーネント配列と同じではない、実施形態13または14に記載の正規化対照。

【0562】

16. ポリヌクレオチドの少なくとも1つの基の少なくとも1つのサブグループの配列が、コンポーネント配列を含む、実施形態12に記載の正規化対照。

【0563】

17. 少なくとも1つの基の少なくとも3つのサブグループの各々の配列が、コンポーネント配列を含む、実施形態12に記載の正規化対照。

40

【0564】

18. すべてのグループの各サブグループの配列が、コンポーネント配列を含む、実施形態12に記載の正規化対照。

【0565】

19. 各サブグループのコンポーネント配列が、任意の他のサブグループのコンポーネント配列と同じではない、実施形態17または18に記載の正規化対照。

【0566】

20. コンポーネント配列が、約6bp～約3000bpの間の配列を含む、実施形態13～19のいずれか1つに記載の正規化対照。

50

【0567】

21. コンポーネント配列が、ランダム配列を含む、実施形態13～20に記載の正規化対照。

【0568】

22. ポリヌクレオチドの少なくとも1つのグループの配列が、単離配列を含む、実施形態10～13に記載の正規化対照。

【0569】

23. すべてのポリヌクレオチドのグループの配列が、単離配列を含む、実施形態10～12のいずれか1つに記載の正規化対照。

【0570】

24. ポリヌクレオチドの少なくとも1つのグループの少なくとも1つのサブグループの配列が、単離配列を含む、実施形態12、13、または15～17のいずれか1つに記載の正規化対照。

【0571】

25. 少なくとも1つのグループのすべてのサブグループの配列が、単離配列を含む、実施形態12、13、または15～17のいずれか1つに記載の正規化対照。

【0572】

26. すべてのグループの各サブグループの配列が、単離配列を含む、実施形態12に記載の正規化対照。

【0573】

27. 各サブグループの単離された配列が、任意の他のサブグループの単離配列と同じではない、実施形態25または26のいずれか1つに記載の正規化対照。

【0574】

28. 単離配列が、約6bp～約200,000bpの間、約15bp～約50,000bpの間、約500bp～約1500bpの間、約100bp～約1200bpの間、または約150bp～約600bpの間の配列を含む、実施形態22～26のいずれか1つに記載の正規化対照。

【0575】

29. 単離配列が、ウイルス、細菌、真菌、または真核性寄生虫から単離される、または由来する、実施形態22～28のいずれか1つに記載の正規化対照。

【0576】

30. ウイルスが、T4バクテリオファージ(T4)またはサイトメガロウイルス(CMV)である、実施形態29に記載の正規化対照。

【0577】

31. 単離配列が、プラスミドから単離される、実施形態22～30のいずれか1つに記載の正規化対照。

【0578】

32. 単離配列が、シーケンスサンプル中の少なくとも1つの標的配列と同じではない、実施形態22～31のいずれか1つに記載の正規化対照。

【0579】

33. 正規化対照におけるポリヌクレオチドのうちのいずれかの配列が、シーケンスサンプル中の少なくとも1つの標的配列と、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、30%、40%、50%、60%、70%、80%、85%、90%、95%、96%、97%、98%、または99%以下の同一性を有する、実施形態1～32のいずれか1つに記載の正規化対照。

【0580】

34. 正規化対照におけるポリヌクレオチドの少なくとも1つのサブグループの少なくとも1つの配列が、シーケンスサンプル中の少なくとも1つの標的配列と少なくとも1つの配列特性を共有する、実施形態12～33のいずれか1つに記載の正規化対照。

10

20

30

40

50

【0581】

35. 少なくとも1つの配列特性が、GC含量パーセント、エントロピー、複雑度、長さ、電子イオン間相互作用ポテンシャル（E I I P）、転位性要素配列の配列特性、ウイルス配列の配列特性、細菌配列の配列特性、真菌配列の配列特性、真核性寄生虫配列、1つ以上のヒト遺伝子配列（複数可）の配列特性、またはそれらの任意の組み合わせを含む、実施形態34に記載の正規化対照。

【0582】

36. ウイルス配列の特性が、GC含量パーセント、反復配列、逆位末端反復（I T R）配列、内部リボソーム侵入部位（I R E S）、タンパク質コード配列、転写後調節エレメント配列、転写調節エレメント、プロモーター配列、シス作用性RNAエレメント、RNA構造エレメントの配列、ゲノムパッケージシグナル、5'非翻訳領域（5' U T R）配列、3'非翻訳領域（3' U T R）配列、またはそれらの組み合わせを含む、実施形態35に記載の正規化対照。

10

【0583】

37. 細菌配列の特性が、GC含量パーセント、反復配列、マイクロRNA結合部位、内部リボソーム侵入部位（I R E S）、タンパク質コード配列、転写調節エレメント配列、プロモーター配列、5' U T R配列、3' U T R配列、またはそれらの組み合わせを含む、実施形態35に記載の正規化対照。

【0584】

38. 反復配列要素が、ポリAモチーフ、ポリTモチーフ、ポリGモチーフ、ポリCモチーフ、ジヌクレオチドモチーフ、トリヌクレオチドモチーフ、テトラヌクレオチドモチーフ、ペンタヌクレオチドモチーフ、ヘキサヌクレオチドモチーフ、ヘプタヌクレオチドモチーフ、オクタヌクレオチドモチーフ、ノナヌクレオチドモチーフ、散在反復配列要素、リボソームRNA配列、tRNA配列、またはそれらの組み合わせを含む、実施形態37に記載の正規化対照。

20

【0585】

39. 真菌配列の特性が、GC含量パーセント、反復配列、マイクロRNA結合部位、内部リボソーム侵入部位（I R E S）配列、タンパク質コード配列、転写調節エレメント、プロモーター配列、5' U T R配列、3' U T R配列、セントロメア配列、テロメア配列、サブテロメア配列、ミトコンドリア配列、またはそれらの組み合わせを含む、実施形態35に記載の正規化対照。

30

【0586】

40. 1つ以上のヒト遺伝子の特性が、GC含量パーセント、反復配列、タンパク質コード配列、イントロン配列、5' U T R配列、3' U T R配列、転写調節エレメント、プロモーター配列、マイクロRNA結合部位、またはそれらの組み合わせを含む、実施形態35に記載の正規化対照。

【0587】

41. シーケンスサンプルが、宿主および非宿主核酸の混合物を含む、実施形態34～40のいずれか1つに記載の正規化対照。

【0588】

42. 宿主が、真核生物である、実施形態41に記載の正規化対照。

40

【0589】

43. 宿主が、昆虫、植物、または動物である、実施形態41に記載の正規化対照。

【0590】

44. 動物が、ヒトである、実施形態43に記載の正規化対照。

【0591】

45. 非宿主が、共生生物（s y m b i o t e）、共生生物（c o m m e n s a l o r g a n i s m）、寄生虫、または病原体を含む、実施形態41～44のいずれか1つに記載の正規化対照。

【0592】

50

46. 非宿主が、複数の生物種を含む、実施形態45に記載の正規化対照。

【0593】

47. 非宿主が、ウイルス種、細菌種、真菌種、真核性寄生虫種、またはそれらの組み合わせを含む、実施形態41～44のいずれか1つに記載の正規化対照。

【0594】

48. 少なくとも1つの標的配列が、シーケンスサンプル中の非宿主配列である、実施形態32～47のいずれか1つに記載の正規化対照。

【0595】

49. 非宿主が、ウイルス、細菌、または真菌であり、少なくとも1つの標的配列が、病原性因子の配列を含む、実施形態48に記載の正規化対照。

10

【0596】

50. 病原性因子が、エンドトキシン、エクソトキシン、溶血素、プロテアーゼ、リパーゼ、DNase、アドヘシン、侵襲、抗食作用、破壊酵素、またはGTPase活性の調節因子をコードする1つ以上の遺伝子を含む、実施形態49に記載の正規化対照。

【0597】

51. 非宿主が、細菌であり、少なくとも1つの標的配列が、抗生物質耐性遺伝子の配列を含む、実施形態48に記載の正規化対照。

【0598】

52. 少なくとも1つの標的配列が、複数の参照配列を含む、実施形態32～51のいずれか1つに記載の正規化対照。

20

【0599】

53. 複数の参照配列が、少なくとも2、少なくとも10、少なくとも50、少なくとも100、少なくとも200、少なくとも300、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1,000、少なくとも1,200、少なくとも1,300、少なくとも1,400、少なくとも1,500、少なくとも1,600、少なくとも1,700、少なくとも1,800、少なくとも1,900、少なくとも2,000、少なくとも2,200、少なくとも2,400、少なくとも2,600、少なくとも2,800、少なくとも3,000、少なくとも4,000、少なくとも5,000、少なくとも6,000、少なくとも7,000、少なくとも8,000、少なくとも9,000、または少なくとも10,000の参照配列を含む、実施形態52に記載の正規化対照。

30

【0600】

54. 参照配列が、非宿主配列を含む、実施形態53に記載の正規化対照。

【0601】

55. 参照配列が、ウイルス配列、細菌配列、真菌配列、またはそれらの組み合わせを含む、実施形態53または54に記載の正規化対照。

【0602】

56. 複数の参照配列が、複数の生物からの参照配列を含む、実施形態52～55のいずれか1つに記載の方法。

【0603】

40

57. 参照配列が、ゲノムDNA配列もしくはcDNA配列、またはそれらの組み合わせを含む、実施形態56に記載の正規化対照。

【0604】

58. 単離配列が、シャッフルされた少なくとも1つの参照配列の少なくとも1つの断片を含む、実施形態52～57のいずれか1つに記載の正規化対照。

【0605】

59. 単離配列が、シャッフルおよび連結された少なくとも第1の参照配列および第2の参照配列の少なくとも2つの断片を含む、実施形態58に記載の正規化対照。

【0606】

60. 第1および第2の参照配列が、生物のゲノムにおいて隣接していない、実施形態

50

58に記載の正規化対照。

【0607】

61. 第1および第2の参照配列が、異なる生物由来である、実施形態58に記載の正規化対照。

【0608】

62. 参照配列の少なくとも1つの断片が、参照配列の約10～100、約10～80、約20～60、約20～40、約20～30、または約20～25個の連続するヌクレオチドを含む、実施形態58～61のいずれか1つに記載の正規化対照。

【0609】

63.

a. 少なくとも1つの参照配列が、参照配列断片の分布によって表され、
b. 分布が、少なくとも5つのピンに分割され、
c. 複数の参照配列断片が、少なくとも5つのうちの少なくとも3つから選択され、シャッフルされ、任意に連結され、それによって単離配列を生成する、実施形態52～62のいずれか1つに記載の正規化対照。

【0610】

64. 参照配列断片が、参照配列長、および(1)GC含量パーセント、(2)エントロピー、(3)複雑度、(4)EIIIP、(5)長さ、またはそれらの組み合わせのうちの少なくとも1つに関して分布される、実施形態63に記載の正規化対照。

【0611】

65. 各ピンが、分布の1パーセント、2パーセント、5パーセント、10パーセント、15パーセント、または20パーセントを表す、実施形態63または64に記載の正規化対照。

【0612】

66. 各ピンが、GC含量の分布の5パーセント、10パーセント、15パーセント、または20パーセントを表す、実施形態63または64に記載の正規化対照。

【0613】

67. 正規化対照の各グループ内のポリヌクレオチドが、シャッフルすること、任意に連結することにより生成され、参照配列断片が、少なくとも3つのピンのうちの1つから選択される、実施形態63～66のいずれか1つに記載の正規化対照。

【0614】

68. 各グループ内のポリヌクレオチドが、参照配列分布における対応するピンからの参照配列断片と(1)同様のGC含量パーセント、(2)同様のエントロピー、(3)同様のEIIIP、(4)同様の長さ、または(5)同様の複雑度のうちの少なくとも3つを有する、実施形態65～67のいずれか1つに記載の正規化対照。

【0615】

69. 参照配列断片が、GC含量パーセント分布から第10番目、第30番目、第50番目、第70番目、第80番目、および第90番目の百分位数のピンから選ばれる、実施形態63～68のいずれか1つに記載の正規化対照。

【0616】

70. 単離配列が、連結された少なくとも1つの参照配列の少なくとも2つの断片を含み、少なくとも2つの断片が、少なくとも1つの参照配列中で隣接していない、実施形態52～57のいずれか1つに記載の正規化対照。

【0617】

71. 単離配列が、少なくとも第1の参照配列の少なくとも第1の断片、および少なくとも第2の参照配列の少なくとも第2の断片を含み、少なくとも第1および第2の断片が連結され、第1および第2の参照配列が、同じ参照配列ではない、実施形態52～57のいずれか1つに記載の正規化対照。

【0618】

72. 少なくとも第1および第2の断片が、対応する参照配列の約15～60、約20

10

20

30

40

50

～40、約20～30、約15～32、20～32、または約25～35個の連続するヌクレオチドである、実施形態70または71に記載の正規化対照。

【0619】

73. 少なくとも第1および第2の断片が、対応する参照配列の29、30、31、32、33、または34個の連続するヌクレオチドである、実施形態70または71に記載の正規化対照。

【0620】

74. 少なくとも第1および第2の断片が、シーケンスリードの平均長の半分未満である、実施形態70～73のいずれか1つに記載の正規化対照。

【0621】

75.

a. 少なくとも1つの参照配列が、スライディングウィンドウを使用して生成された重複する参照配列断片の分布によって表され、

b. 分布が、少なくとも5つのビンに分割され、

c. 複数の参照配列断片が、少なくとも5つビンのうちの少なくとも3つから選択され、連結され、それによって単離配列を生成する、実施形態70～74のいずれか1つに記載の正規化対照。

【0622】

76. 参照配列断片が、参照配列長、および(1)GC含量パーセント、(2)エントロピー、(3)複雑度、(4)EIIIP、またはそれらの組み合わせのうちの少なくとも1つに関して分布される、実施形態75に記載の正規化対照。

【0623】

77. 各ビンが、分布の1パーセント、2パーセント、5パーセント、10パーセント、15パーセント、または20パーセントを表す、実施形態75または76に記載の正規化対照。

【0624】

78. 各ビンが、GC含量の分布の5パーセント、10パーセント、15パーセント、または20パーセントを表す、実施形態75または76に記載の正規化対照。

【0625】

79. 各グループ内のポリヌクレオチドが、少なくとも3つのビンのうちの1つから選択される参照配列断片を連結することによって生成される、実施形態75～78のいずれか1つに記載の正規化対照。

【0626】

80. 各グループ内のポリヌクレオチドが、参照配列分布における対応するビンからの参照配列断片と(1)同様のGC含量パーセント、(2)同様のエントロピー、(3)同様のEIIIP、(4)同様の長さ、または(5)同様の複雑度のうちの少なくとも3つを有する、実施形態75～79のいずれか1つに記載の正規化対照。

【0627】

81. スライディングウィンドウが、1bp、2bp、3bp、4bp、または5bpの摺動ウィンドウを含む、実施形態75～80のいずれか1つに記載の正規化対照。

【0628】

82. 参照配列断片が、GC含量パーセント分布から第10番目、第30番目、第50番目、第70番目、第80番目、および第90番目の百分位数のビンから選ばれる、実施形態75～81のいずれか1つに記載の正規化対照。

【0629】

83. シーケンスサンプルは、癌と診断された対象からのサンプルを含む、実施形態32～80のいずれか1つに記載の正規化対照。

【0630】

84. 癌が、対象の非癌性細胞または健康な対象からの非癌性細胞に対して、少なくとも1つの遺伝子変化を有する配列を含む細胞集団を含む、実施形態83に記載の正規化対

10

20

30

40

50

照。

【0631】

85. 少なくとも1つの遺伝子変化が、一塩基多型 (SNP)、挿入、欠失、反転、または染色体再配置を含む、実施形態84に記載の正規化対照。

【0632】

86. 標的配列が、少なくとも1つの遺伝子変化を含む、実施形態83または84に記載の正規化。

【0633】

87. シーケンスサンプルが、メタゲノムサンプルを含む、実施形態32～82のいずれか1つに記載の正規化対照。

【0634】

88. メタゲノムサンプルが、環境サンプルである、実施形態87に記載の正規化対照。

【0635】

89. 少なくとも1つの標的配列が、メタゲノムサンプル内の種の配列を含む、実施形態88に記載の正規化対照。

【0636】

90. 少なくとも1つの標的配列が、病原性因子の配列、抗生物質耐性遺伝子、または病原性に関連する配列を含む、実施形態88に記載の正規化対照。

【0637】

91. すべてのグループのポリヌクレオチドが、同じ濃度である、実施形態1～90のいずれか1つに記載の正規化対照。

【0638】

92. すべてのグループのポリヌクレオチドが、同じ濃度ではない、実施形態1～90のいずれか1つに記載の正規化対照。

【0639】

93. すべてのグループのポリヌクレオチドが、線形配列または幾何学配列における濃度である、実施形態92に記載の正規化対照。

【0640】

94. ポリヌクレオチドの1つのグループが、ベースライン濃度を有し、残りの非ベースライン群の濃度が、ベースライン濃度の整数倍である、実施形態93に記載の正規化対照。

【0641】

95. ポリヌクレオチドの1つのグループが、ベースライン濃度を有し、ポリヌクレオチドの各群の濃度が、ベースライン濃度を除く、正規化対照における別の群の濃度の2倍である、実施形態93に記載の正規化対照。

【0642】

96. ポリヌクレオチドの少なくとも3つのグループが、1:2:4の濃度比率で存在する、実施形態93に記載の正規化対照。

【0643】

97. グループ内のポリヌクレオチドが、同じGC含量パーセントを有する、実施形態1～96のいずれか1つに記載の正規化対照。

【0644】

98. グループ内のポリヌクレオチドが、他のすべてのグループ内のポリヌクレオチドと同じGC含量パーセントを有する、実施形態97に記載の正規化対照。

【0645】

99. グループ内のポリヌクレオチドが、任意の他のグループ内のポリヌクレオチドと同じGC含量パーセントを有さない、実施形態97に記載の正規化対照。

【0646】

100. 各グループのポリヌクレオチドのGC含量パーセントが、エンドポイントを含む40%～60%の間である、実施形態97～99のいずれか1つに記載の正規化対照。

10

20

30

40

50

【0647】

101. 各グループのポリヌクレオチドのGC含量パーセントが、エンドポイントを含む43%～56%の間である、実施形態97～99のいずれか1つに記載の正規化対照。

【0648】

102. ポリヌクレオチドが、固有の分子識別子をコードする配列を含む、実施形態1～101のいずれか1つに記載の正規化対照。

【0649】

103. 固有の分子識別子をコードする配列が、第1のインデックス配列または第2のインデックス配列を含む、実施形態102に記載の正規化対照。

【0650】

104. ポリヌクレオチドが、デオキシリボ核酸（DNA）分子、リボ核酸（RNA）分子、またはDNA-RNAハイブリッド分子を含む、実施形態1～103のいずれか1つに記載の正規化対照。

【0651】

105. ポリヌクレオチドが、正規化対照において、1マイクロリットル（ μL ）あたり約0.005～約500ピコグラム（pg）の間の濃度にある、実施形態104に記載の正規化対照。

【0652】

106. ポリヌクレオチドが、正規化対照において約0.5～約50 pg/ μL の間の濃度にある、実施形態105に記載の正規化対照。

【0653】

107. ポリヌクレオチドが、正規化対照において1.4 pg/ μL の濃度である、実施形態105に記載の正規化対照。

【0654】

108. 少なくとも1つのタンパク質をさらに含む、実施形態1～107のいずれか1つに記載の正規化対照。

【0655】

109. 少なくとも1つのタンパク質が、細菌、ウイルス、または真核性寄生虫から単離される、または由来する、実施形態108に記載の正規化対照。

【0656】

110. 少なくとも1つのタンパク質が、ウイルスカプシドタンパク質または細菌細胞壁タンパク質である、実施形態108に記載の正規化対照。

【0657】

111. ポリヌクレオチドおよびタンパク質の集団が、動作可能に連結される、実施形態108～110のいずれか1つに記載の正規化対照。

【0658】

112. 正規化対照を製造する方法であって、

【0659】

a. DNA分子の少なくとも3つのグループを合成することであって、グループ内のDNA分子が、同じ配列を有し、DNA分子の各々の配列が、5'から3'に、第1のコンポーネント配列、第1の制限酵素部位、第2のコンポーネント配列、第2の制限部位、および第3のコンポーネント配列を含む、合成することと、

b. 溶液中のDNA分子のグループの各々の量を定量することと、

c. 第1および第2の制限部位を切断する制限酵素でDNA分子の群の各々を消化して、消化されたDNA分子の少なくとも3つのグループを産生することと、

d. 所定の消化されたDNA分子のグループのそれぞれの量を混合することで、正規化対照を産生することと、を含む、方法。

【0660】

113. DNA分子の各々が、すべての他のDNA分子の第1、第2、および第3のコンポーネント配列と同じ長さの第1、第2、および第3のコンポーネント配列を含む、実

10

20

30

40

50

施形態 1 1 2 に記載の方法。

【0 6 6 1】

1 1 4. 第 1、第 2、および第 3 のコンポーネント配列が各々、約 1 5 b p ～約 1 2 0 0 b p の間である、実施形態 1 1 2 に記載の方法。

【0 6 6 2】

1 1 5. 第 1、第 2、および第 3 のコンポーネント配列が各々、約 1 0 0 b p ～約 6 0 0 b p の間である、実施形態 1 1 2 に記載の方法。

【0 6 6 3】

1 1 6. DNA 分子の少なくとも 3 つのグループのうちのいずれかの第 1、第 2、および第 3 のコンポーネント配列が、同じ配列ではない、実施形態 1 1 2 ～ 1 1 5 のいずれか 1 つに記載の方法。

10

【0 6 6 4】

1 1 7. 第 1 および第 2 の制限部位の配列が、E c o R I 部位、B a m H I 部位、H i n d I I I 部位、N o t I 部位、P v u I I 部位、S m a I 部位、E c o R V 部位、K p n I 部位、P s t I 部位、S a l I 部位、S c a I 部位、S p e I 部位、S p h I 部位、S t u I 部位、および X b a I 部位の配列からなる群から選択される、実施形態 1 1 2 に記載の方法。

【0 6 6 5】

1 1 8. 第 1 および第 2 の制限部位の配列が、X h o I 部位を含み、制限酵素が、X h o I を含む、実施形態 1 1 2 に記載の方法。

20

【0 6 6 6】

1 1 9. 少なくとも 3 つの DNA 分子の各々の配列が、第 1 のプライマー配列と、第 2 のプライマー配列に対して相補的な配列とをさらに含み、第 1 のプライマー配列が、第 1 のコンポーネント配列の 5 ' に位置し、第 2 のプライマー配列に対して相補的な配列が、第 3 のコンポーネント配列の 3 ' に位置する、実施形態 1 1 2 ～ 1 1 8 のいずれか 1 つに記載の方法。

【0 6 6 7】

1 2 0. 第 1 のプライマー配列が、5 ' - T G A A G A A C T G C G G C A G G - 3 ' (配列番号 1 2) の配列を含む、実施形態 1 1 9 に記載の方法。

【0 6 6 8】

1 2 1. 第 2 のプライマー配列に対して相補的な配列が、5 ' - G G T C T G C A C T T C C A G C T - 3 ' (配列番号 1 3) の配列を含む、実施形態 1 1 9 または 1 2 0 に記載の方法。

30

【0 6 6 9】

1 2 2. 所定の消化された DNA 分子の少なくとも 3 つのグループの量が、等しい、実施形態 1 1 2 ～ 1 2 1 のいずれか 1 つに記載の方法。

【0 6 7 0】

1 2 3. 所定の消化された DNA 分子の少なくとも 3 つのグループの量が、等しくない、実施形態 1 1 2 ～ 1 2 2 のいずれか 1 つに記載の方法。

【0 6 7 1】

1 2 4. 所定の消化された DNA 分子の少なくとも 3 つのグループの量が、線形配列または幾何学配列にある、実施形態 1 2 3 に記載の方法。

40

【0 6 7 2】

1 2 5. 所定の消化された DNA 分子の 1 つのグループの量が、ベースラインであり、所定の消化された DNA 分子の少なくとも 2 つの追加のグループの量が、ベースラインの整数倍である、実施形態 1 2 3 に記載の方法。

【0 6 7 3】

1 2 6. 所定の DNA 分子の 1 つのグループの量が、ベースラインであり、所定の消化された DNA 分子の各追加のグループの量が、ベースラインを除いて、所定の消化された DNA 分子の別のグループの量の 2 倍である、実施形態 1 2 3 に記載の方法。

50

【0674】

127. 所定の消化されたDNA分子の少なくとも3つのグループの量が、1:2:4の比率である、実施形態123に記載の方法。

【0675】

128. (d)の混合工程が、所定のDNA分子の少なくとも1つの追加のグループの量を、工程(a)～(c)の消化されたDNA分子の少なくとも3つのグループと混合して、単一の正規化対照においてDNA分子の少なくとも4つのグループを産生することをさらに含み、DNA分子の少なくとも1つの追加のグループが、サンプルから単離される、または由来する、実施形態112～127のいずれか1つに記載の方法。

【0676】

129. サンプルが、プラスミドDNA、ミトコンドリアDNA、葉緑体DNA、またはゲノムDNAを含む、実施形態128に記載の方法。

【0677】

130. 所定のDNA分子の少なくとも4つのグループの量が、等しい、実施形態128または129に記載の方法。

【0678】

131. 所定のDNA分子の少なくとも4つのグループの量が、等しくない、実施形態128または129に記載の方法。

【0679】

132. 所定のDNA分子の少なくとも4つのグループの量が、線形配列または幾何学配列にある、実施形態131に記載の方法。

【0680】

133. 所定のDNA分子の1つのグループの量が、ベースラインであり、所定のDNA分子の追加のグループの量が、ベースラインの整数倍である、実施形態131に記載の方法。

【0681】

134. 所定の消化されたDNA分子の1つのグループの量が、ベースラインであり、所定の消化されたDNA分子の各追加のグループの量が、ベースラインを除いて、所定の消化されたDNA分子の別のグループの量の2倍である、実施形態131に記載の方法。

【0682】

135. 所定の消化されたDNA分子の少なくとも4つのグループの量が、1:2:4:8の比率である、実施形態131に記載の方法。

【0683】

136. RNA分子を含む正規化対照を製造する方法であって、
a. DNA分子の少なくとも3つのグループを合成することであって、少なくとも3つのDNA分子の各々が、5'から3'に、第1のコンポーネント配列、プロモーター配列に対して相補的な配列、第1の制限酵素部位、第2のコンポーネント配列、プロモーター配列に対して相補的な第2の配列、第2の制限部位、第3のコンポーネント配列、およびプロモーター配列に対して相補的な第3の配列を含む、合成することと、
b. 第1および第2の制限酵素部位を切断する制限酵素でDNA分子の各グループを消化して、消化されたDNA分子の少なくとも3つのグループを産生することと、
c. 消化されたDNA分子の少なくとも3つのグループの各々をインビトロで転写して、RNA分子の少なくとも3つの集団を産生することと、
d. RNA分子の集団のそれぞれに産生されたRNAの量を定量することと、
e. 所定のRNA分子の集団の各々の量を混合し、それによってRNA分子を含む正規化対照を産生することと、を含む、方法。

【0684】

137. 少なくとも3つのDNA分子の各々が、少なくとも3つのDNA分子のいずれかの他の第1、第2、および第3のコンポーネント配列と同じ長さの第1、第2、および第3のコンポーネント配列を含む、実施形態136に記載の方法。

10

20

30

40

50

【0685】

138. 第1のプロモーター配列、第2のプロモーター配列、および第3のプロモーター配列が、同じプロモーター配列を含む、実施形態136に記載の方法。

【0686】

139. プロモーター配列が、T7プロモーター配列、SP6プロモーター配列、またはT3プロモーター配列からなる群から選択されるプロモーター配列を含む、実施形態138に記載の方法。

【0687】

140. T7プロモーター配列が、5' - T A T A C G A C T C A C T A T A G G G - 3' (配列番号15) の配列を含む、実施形態139に記載の方法。

10

【0688】

141. 第1および第2の制限部位の配列が、同じである、実施形態136～140のいずれか1つに記載の方法。

【0689】

142. 第1、第2、および第3のコンポーネント配列が各々、約15bp～約120bpの間である、実施形態136～141のいずれか1つに記載の方法。

【0690】

143. 第1、第2、および第3のコンポーネント配列が各々、約100bp～約600bpの間である、実施形態136～141のいずれか1つに記載の方法。

【0691】

144. DNA分子の少なくとも3つのグループのうちのいずれかの第1、第2、および第3のコンポーネント配列が、同じ配列ではない、実施形態136～141のいずれか1つに記載の方法。

20

【0692】

145. 第1および第2の制限部位の配列が、EcoRI部位、BamHI部位、HindIII部位、NotI部位、PvuII部位、SmaI部位、EcoRV部位、KpnI部位、PstI部位、SalI部位、ScaI部位、SpeI部位、SphI部位、StuI部位、およびXbaI部位の配列からなる群から選択される、実施形態136～144のいずれか1つに記載の方法。

【0693】

146. 第1および第2の制限部位の配列が、XhoI部位を含み、制限酵素が、XhoIを含む、実施形態145に記載の方法。

30

【0694】

147. 所定のRNA分子の少なくとも3つの集団の各々の量が、等しい、実施形態136～146のいずれか1つに記載の方法。

【0695】

148. 所定のRNA分子の少なくとも3つの集団の各々の量が、等しくない、実施形態136～146のいずれか1つに記載の方法。

【0696】

149. 所定のRNA分子の少なくとも3つの集団の量が、線形配列または幾何学配列にある、実施形態148に記載の方法。

40

【0697】

150. 所定の合成RNA分子の1つの集団の量が、ベースラインであり、所定のRNA分子の少なくとも2つの追加の集団の量が、ベースラインの整数倍である、実施形態148に記載の方法。

【0698】

151. 所定のRNA分子の1つの集団の量が、ベースラインであり、所定のRNA分子の各追加の集団の量が、ベースラインを除いて、RNA分子の別の集団の濃度の2倍である、実施形態148に記載の方法。

【0699】

50

152. 所定のRNA分子の少なくとも3つの集団の量が、1:2:4の比率である、実施形態148に記載の方法。

【0700】

153. 正規化対照を約0.005~約500 pg/μLの濃度に希釈することをさらに含む、実施形態136~152のいずれか1つに記載の方法。

【0701】

154. 正規化対照を約0.5~約50 pg/μLの濃度に希釈することをさらに含む、実施形態136~152のいずれか1つに記載の方法。

【0702】

155. 正規化対照濃度が、1.4 pg/μLである、実施形態136~152に記載の方法。

【0703】

156. ポリヌクレオチドの少なくとも3つの基を含む正規化対照を製造する方法であって、各グループ内のポリヌクレオチドが、同じ長さのものであり、方法が、

i. サンプルからDNAを抽出することと、

ii. 制限酵素でDNAを消化してDNA断片の集合を産生することと、

iii. DNA断片の集合を分離することと、

iv. 少なくとも3つのポリヌクレオチドのグループを産生するために、DNA断片を精製することであって、各グループ内のポリヌクレオチドが、同じ長さである、精製することと、

v. 所定のポリヌクレオチドのグループの各々の量を混合して正規化対照を産生することと、を含む、方法。

【0704】

157. 制限酵素が、EcoRI、BamHI、HindIII、PvuII、SmaI、EcoRV、KpnI、PstI、SalI、ScaI、SpeI、SphI、StuI、XbaI、NotI、AscI、FseI、PacI、PmeI、BglII、BstBI、HincII、およびSgfIからなる群から選択される、実施形態156に記載の方法。

【0705】

158. DNA断片の集合を分離することが、断片サイズに基づく分離を含む、実施形態156または157に記載の方法。

【0706】

159. サイズに基づいてDNA断片を分離することが、ゲル電気泳動、クロマトグラフィー、またはタンジェンシャルフロー過(TFF)を含む、実施形態158に記載の方法。

【0707】

160. 少なくとも3つの異なるグループのポリヌクレオチドの各々にアダプターをライゲーションすることをさらに含む、実施形態156~159のいずれか1つに記載の方法。

【0708】

161. 実施形態112~160のいずれか1つに記載の方法によって生成される正規化対照。

【0709】

162. 実施形態1~111のいずれか1つに記載の正規化対照を含むキット。

【0710】

163. 少なくとも3つの異なる種の生物の混合物を含む多検体対照であって、生物が、不活性化されている、多検体対照。

【0711】

164. 各種が、同じ力価で多検体対照中に存在する、実施形態163に記載の多検体対照。

10

20

30

40

50

【0712】

165. 各種が、異なる力価で多検体対照中に存在する、実施形態163に記載の多検体対照。

【0713】

166. 力価が、線形配列または幾何配列にある、実施形態163に記載の多検体対照。

【0714】

167. 少なくとも2つの種が、同じ力価で多検体対照中に存在し、少なくとも2つの種が、異なる力価で混合物中に存在する、実施形態163に記載の多検体対照。

【0715】

168. 多検体対照中の各異なる種が、1mLあたり約500万～約1000万単位（U/mL）の力価で存在する、実施形態163～167のいずれか1つに記載の多検体対照。

10

【0716】

169. 生物の種が、ウイルス、細菌、真菌、または真核性寄生虫の種を含む、実施形態168に記載の多検体対照。

【0717】

170. ウイルスが、DNAウイルスである、実施形態169に記載の多検体対照。

【0718】

171. ウイルスが、RNAウイルスである、実施形態169に記載の多検体対照。

【0719】

172. 細菌が、グラム陽性細菌またはグラム陰性細菌である、実施形態169に記載の多検体対照。

20

【0720】

173. 生物の種が、ヒト病原体である、実施形態163～172のいずれか1つに記載の多検体対照。

【0721】

174. 多検体対照中の各異なる種は、1mLあたり約500万～約1000万の感染単位（IU/mL）の力価で存在する、実施形態173に記載の多検体対照。

【0722】

175. 多検体対照が、10または11種のウイルスを含むか、または本質的にそれからなる、実施形態163に記載の多検体対照。

30

【0723】

176. 多検体対照中の種は、表8に列挙される種の集合から選択される、実施形態175に記載の多検体対照。

【0724】

177. ヒト病原体が、血液または組織移植に見出されるヒト病原体を含む、実施形態173または175に記載の多検体対照。

【0725】

178. 血液または組織移植において見出されるヒト病原体が、サイトメガロウイルス（CMV）、エプスタインバーウイルス（EBV）、アデノウイルス（ADV）、BKウイルス（BKV）、JCポリオマウイルス（JCV）、ヒトヘルペスウイルス6A（HHV6A）、ヒトヘルペスウイルス6B（HHV6B）、単純ヘルペスウイルス1型（HSV1）、単純ヘルペスウイルス2型（HSV2）、水痘帯状疱疹ウイルス（VZV）、およびヒトパルボウイルスB19（B19）、ヒト免疫不全ウイルス（HIV）、B型肝炎、C型肝炎、ヒトTリンパ性ウイルスI型およびII型（HTLV）、ウエストナイルウイルス、ジカウイルス、Streptococcus種、Staphylococcus aureus、Enterococcus faecalis、Bacillus cereus、Propionibacterium acnes、Serratia liquifaciens、Serratia marcescens、Yersinia enterocolitica、Enterobacter spp.、Acinetobac

40

50

ter spp.、Pseudomonas spp.、E. coli、Klebsiella pneumoniae、Proteus mirabilis、A. baumannii、Bacillus cereus、コアグラゼ陰性staphylococci、Streptococcus spp.、Klebsiella spp.、Serratia spp. Providencia rettgeri、Treponema pallidum、Trypanosoma cruzi、およびBabesia microtiのうちの1つ以上を含む、実施形態177に記載の多検体対照。

【0726】

179. 多検体対照中のヒト病原体が、表2～3または5～7から選択される、実施形態173または174に記載の多検体対照。

【0727】

180. 種が、ヒト病原体ではない、実施形態163～172のいずれか1つに記載の多検体対照。

【0728】

181. 種が、ヒト病原体である種と同じ属にある、実施形態180に記載の多検体対照。

【0729】

182. 種が、ヒト病原体である種およびヒト病原体ではない種の混合物を含む、実施形態163～172のいずれか1つに記載の多検体対照。

【0730】

183. 生物が、化学処理、熱処理、pH処理、または紫外線照射によって不活性化されている、実施形態163～182のいずれか1つに記載の多検体対照。

【0731】

184. 化学処理が、ホルムアルデヒドを含む、実施形態183に記載の多検体対照。

【0732】

185. 血漿をさらに含む、実施形態163～184のいずれか1つに記載の多検体対照。

【0733】

186. 血漿が、ヒト血漿である、実施形態183に記載の多検体対照。

【0734】

187. 実施形態163～186のいずれか1つに記載の多検体対照から抽出される核酸を含む集合。

【0735】

188. 実施形態1～111のいずれか1つに記載の正規化対照をさらに含む、実施形態163～186のいずれか1つに記載の多検体対照または実施形態185に記載の集合。

【0736】

189. 実施形態163～186のいずれか1つに記載の多検体対照または実施形態187に記載の集合を含むキット。

【0737】

190. 実施形態1～111のいずれか1つの正規化対照をさらに含む、実施形態189に記載のキット。

【0738】

191. サンプル中の少なくとも1つの標的配列の発現レベルを定量化する方法であって、

a. 実施形態1～111のいずれか1つに記載の正規化対照の既知の量をサンプルと混合することと、

b. ハイスループットシーケンスライブラリを調製することと、

c. ライブラリをシーケンスしてリードの集合を産生することと、

d. リードの集合からのリードをサンプルまたは正規化対照にマップすることと、

e. 正規化対照中のポリヌクレオチドのグループまたはサブグループの各々によって産

10

20

30

40

50

生されたリードの数を判定することと、

f. (a) でサンプルと混合された正規化対照中のポリヌクレオチドのグループまたはサブグループの各々の開始濃度と、(c) で產生されたリードの数との間の関係性を計算することと、

g. リードとサンプル中の少なくとも1つの標的配列の濃度の間の関係性をモデリングすることと、

h. サンプル中の少なくとも1つの標的ポリヌクレオチドの初期濃度を、(g) のモデルを使用して標的配列によって產生されたリードの数から計算することと、を含む、方法。

【0739】

192. サンプルから核酸を抽出することをさらに含む、実施形態191に記載の方法。

【0740】

193. 核酸が、DNA、RNA、またはそれらの混合物を含む、実施形態191に記載の方法。

【0741】

194. (a) の混合工程が、(b) のライブラリ調製工程の前に、(b) のライブラリ調製工程と同時に、または(b) のライブラリ調製工程の後に生じる、実施形態191～193のいずれか1つに記載の方法。

【0742】

195. サンプルから核酸を抽出することをさらに含み、(a) の混合工程が、サンプルから核酸を抽出する前に生じる、実施形態191～194のいずれか1つに記載の方法。

【0743】

196. 正規化対照中のポリヌクレオチドの少なくとも3つグループの各々が、約0.001対0.420対1の比率でサンプルに添加される、実施形態191～195のいずれか1つに記載の方法。

【0744】

197. 正規化対照中のポリヌクレオチドの少なくとも3つグループの各々が、約0.001 pg/μL～約5 pg/μLの間の最終濃度でサンプルに添加される、実施形態191～196のいずれか1つに記載の方法。

【0745】

198. サンプルが、ヒトサンプルである、実施形態191～197のいずれか1つに記載の方法。

【0746】

199. 標的配列が、ウイルス、細菌、真菌、真核性寄生虫、または1つ以上のヒト遺伝子の配列を含む、実施形態191～198のいずれか1つに記載の方法。

【0747】

200. 標的配列が、病原性因子の配列、抗生物質耐性遺伝子、または病原性に関連する配列を含む、実施形態191～198のいずれか1つに記載の方法。

【0748】

201. サンプルは、癌と診断された対象からのサンプルを含む、実施形態191～198のいずれか1つに記載の方法。

【0749】

202. 癌が、対象の非癌性細胞に対して、少なくとも1つの遺伝子変化を有する配列を含む細胞集団を含む、実施形態201に記載の方法。

【0750】

203. 少なくとも1つの遺伝子変化が、一塩基多型(SNP)、挿入、欠失、反転、または再配置を含む、実施形態202に記載の方法。

【0751】

204. 標的配列が、少なくとも1つの遺伝子変化を有する配列を含む、実施形態201～203のいずれか1つに記載の方法。

【0752】

10

20

30

40

50

205. サンプルが、メタゲノムサンプルを含む、実施形態191～197のいずれか1つに記載の方法。

【0753】

206. メタゲノムサンプルが、環境サンプルである、実施形態205に記載の方法。

【0754】

207. 標的ポリヌクレオチドが、メタゲノムサンプル内の種の配列を含む、実施形態205または206に記載の方法。

【0755】

208. 標的ポリヌクレオチドが、病原性因子の配列、抗生物質耐性遺伝子、または病原性に関連する配列を含む、実施形態205～207のいずれか1つに記載の方法。

10

【0756】

209. サンプルが、少なくとも3つの異なる種の生物の混合物を含む多検体対照を含み、生物が、不活性化されている、実施形態191～193のいずれか1つに記載の方法。

【0757】

210. 少なくとも3つの異なる種の生物の各々が、少なくとも1つの標的配列を含む、実施形態209に記載の方法。

【0758】

211. 少なくとも1つの標的配列が、生物の各種において異なる標的配列である、実施形態210に記載の方法。

【0759】

20

212.

(i) 工程(d)において、リードの集合からサンプルヘリッドをマップすることが、多検体対照(d)中の少なくとも3つの種の生物の各々にリードをマッピングすることをさらに含み、

(ii) (h)における少なくとも1つの標的配列の初期濃度を計算することが、各種についての少なくとも1つの標的配列の初期濃度を計算することを含む、実施形態209に記載の方法。

【0760】

213. 各種が、同じ力価で多検体対照中に存在する、実施形態209～212のいずれか1つに記載の方法。

30

【0761】

214. 各種が、異なる力価で多検体対照中に存在する、実施形態209～212のいずれか1つに記載の方法。

【0762】

215. 力価が、線形配列または幾何配列にある、実施形態214に記載の方法。

【0763】

216. 少なくとも2つの種が、同じ力価で多検体対照中に存在し、少なくとも2つの種が、異なる力価で存在する、実施形態209～212のいずれか1つに記載の方法。

【0764】

217. 多検体対照中の各種が、1mLあたり500万～1000万単位(U/mL)の力価で存在する、実施形態209～216のいずれか1つに記載の方法。

40

【0765】

218. 生物が、ウイルス、細菌、真菌、真核性寄生虫、またはそれらの組み合わせである、実施形態209～217のいずれか1つに記載の方法。

【0766】

219. ウイルスが、DNAウイルスである、実施形態218に記載の方法。

【0767】

220. ウイルスが、RNAウイルスである、実施形態218に記載の方法。

【0768】

221. 細菌が、グラム陽性細菌またはグラム陰性細菌である、実施形態218に記載

50

の方法。

【0769】

222. 生物の種が、ヒト病原体である、実施形態209～221のいずれか1つに記載の方法。

【0770】

223. 多検体対照中の各異なる種は、1mLあたり約500万～約1000万の感染単位(IU/mL)の力価で存在する、実施形態222に記載の方法。

【0771】

224. 多検体対照が、10または11種のウイルスを含むか、または本質的にそれらなる、実施形態222または223に記載の方法。

【0772】

225. 多検体対照中の種は、表8に列挙される種の集合から選択される、実施形態224に記載の方法。

【0773】

226. ヒト病原体が、血液または組織移植に見出されるヒト病原体を含む、実施形態222または223に記載の方法。

【0774】

227. 血液または組織移植において見出されるヒト病原体が、サイトメガロウイルス(CMV)、エプスタインバーウイルス(EBV)、アデノウイルス(ADV)、BKウイルス(BKV)、JCポリオマウイルス(JCV)、ヒトヘルペスウイルス6A(HHV6A)、ヒトヘルペスウイルス6B(HHV6B)、単純ヘルペスウイルス1型(HSV1)、単純ヘルペスウイルス2型(HSV2)、水痘帯状疱疹ウイルス(VZV)、およびヒトパルボウイルスB19(B19)、ヒト免疫不全ウイルス(HIV)、B型肝炎、C型肝炎、ヒトTリンパ性ウイルスI型およびII型(HTLV)、ウエストナイルウイルス、ジカウイルス、*Streptococcus* 種、*Staphylococcus aureus*、*Enterococcus faecalis*、*Bacillus cereus*、*Propionibacterium acnes*、*Serratia liquefaciens*、*Serratia marcescens*、*Yersinia enterocolitica*、*Enterobacter* spp.、*Acinetobacter* spp.、*Pseudomonas* spp.、*E. coli*、*Klebsiella pneumonia*、*Proteus mirabilis*、*A. baumannii*、*Bacillus cereus*、コアグラゼ陰性staphylococci、*Streptococcus* spp.、*Klebsiella* spp.、*Serratia* spp.、*Providencia rettgeri*、*Treponema pallidum*、*Trypanosoma cruzi*、および*Babesia microti*のうちの1つ以上を含む、実施形態226に記載の方法。

【0775】

228. 多検体対照中のヒト病原体が、表2～3または5～7から選択される、実施形態222または223に記載の方法。

【0776】

229. 生物が、化学処理、熱処理、pH処理、または紫外線照射によって不活性化されている、実施形態209～224のいずれか1つに記載の方法。

【0777】

230. 化学処理が、ホルムアルデヒドを含む、実施形態229に記載の方法。

【0778】

231. 多検体対照が、血漿をさらに含む、実施形態209～230のいずれか1つに記載の方法。

【0779】

232. 血漿が、ヒト血漿である、実施形態231に記載の方法。

【0780】

10

20

30

40

50

233. 方法が、既知の量の多検体対照をサンプルと混合することを含み、多検体対照が、少なくとも3つの異なる種の生物の混合物を含み、生物が、不活性化されている、実施形態191～232のいずれか1つに記載の方法。

【0781】

234. 生物が、ウイルス、細菌、真菌、真核性寄生虫、またはそれらの組み合わせである、実施形態233に記載の方法。

【0782】

235. ウイルスが、DNAウイルスである、実施形態234に記載の方法。

【0783】

236. ウイルスが、RNAウイルスである、実施形態234に記載の方法。

10

【0784】

237. 細菌が、グラム陽性細菌またはグラム陰性細菌である、実施形態234に記載の方法。

【0785】

238. 生物が、ヒト病原体ではない、実施形態234～237のいずれか1つに記載の方法。

【0786】

239. 生物が、ヒト病原体である種と同じ属にある種である、実施形態238に記載の方法。

【0787】

20

240. ヒト病原体が、血液または組織移植に見出されるヒト病原体を含む、実施形態239に記載の方法。

【0788】

241. 血液または組織移植において見出されるヒト病原体が、サイトメガロウイルス (CMV)、エプスタインバーウイルス (EBV)、アデノウイルス (ADV)、BKウイルス (BKV)、JCポリオマウイルス (JCV)、ヒトヘルペスウイルス6A (HHV6A)、ヒトヘルペスウイルス6B (HHV6B)、単純ヘルペスウイルス1型 (HSV1)、単純ヘルペスウイルス2型 (HSV2)、水痘帯状疱疹ウイルス (VZV)、およびヒトパルボウイルスB19 (B19)、ヒト免疫不全ウイルス (HIV)、B型肝炎、C型肝炎、ヒトTリンパ性ウイルスI型およびII型 (HTLV)、ウエストナイルウイルス、ジカウイルス、*Streptococcus* 種、*Staphylococcus aureus*、*Enterococcus faecalis*、*Bacillus cereus*、*Propionibacterium acnes*、*Serratia liquefaciens*、*Serratia marcescens*、*Yersinia enterocolitica*、*Enterobacter* spp.、*Acinetobacter* spp.、*Pseudomonas* spp.、*E. coli*、*Klebsiella pneumoniae*、*Proteus mirabilis*、*A. baumannii*、*Bacillus cereus*、コアグラセ陰性 *staphylococci*、*Streptococcus* spp.、*Klebsiella* spp.、*Serratia* spp.、*Providencia rettgeri*、*Treponema pallidum*、*Trypanosoma cruzi*、および *Babesia microti* のうちの1つ以上を含む、実施形態240に記載の方法。

30

【0789】

242. 生物が、化学処理、熱処理、pH処理、または紫外線照射によって不活性化されている、実施形態234～240のいずれか1つに記載の方法。

【0790】

243. 既知の量の多検体対照をサンプルと混合することが、工程(b)におけるライブラリ調製の前に生じる、実施形態234～242のいずれか1つに記載の方法。

【0791】

244. 多検体対照が、多検体対照中の各種の約500万～1000万U/mLの濃度

50

でサンプルに添加される、実施形態 2 3 4 ~ 2 4 3 のいずれか 1 つの実施形態に記載の方法。

【0 7 9 2】

2 4 5 . モデルが、線形モデルである、実施形態 1 9 1 ~ 2 4 4 のいずれか 1 つに記載の方法。

【0 7 9 3】

2 4 6 . モデルが、機械学習分類子によって作成される、実施形態 1 9 1 ~ 2 4 4 のいずれか 1 つに記載の方法。

【0 7 9 4】

2 4 7 . 機械学習分類子が、監督される、実施形態 2 4 6 に記載の方法。

【0 7 9 5】

2 4 8 . 機械学習分類子が、サポートベクターマシンである、実施形態 2 4 7 に記載の方法。

【0 7 9 6】

2 4 9 . サンプル中の少なくとも 1 つの標的核酸分子の初期濃度対標的核酸分子によって產生されるリードの数の線形回帰が、正規化後、0 . 9 5 超、0 . 9 6 超、0 . 9 7 超、0 . 9 8 超、または 0 . 9 9 超の R^2 値を有する、実施形態 2 4 6 に記載の方法。

【0 7 9 7】

2 5 0 . 線形回帰の R^2 値が、少なくとも 0 . 0 1、0 . 0 3、0 . 0 5、0 . 0 7、0 . 0 9、0 . 1、0 . 1 3、0 . 1 5、0 . 1 7、0 . 1 9、0 . 2、0 . 2 3、0 . 2 5、0 . 2 7、0 . 2 9、0 . 3、0 . 3 3、0 . 3 5、0 . 3 7、0 . 3 9、0 . 4、0 . 4 3、0 . 4 5、0 . 4 7、0 . 4 9、0 . 5、0 . 5 3、0 . 5 5、0 . 5 7、0 . 5 9、0 . 6、0 . 6 3、0 . 6 5、0 . 6 7、0 . 6 9、0 . 7、または以下の正規化の間の任意の値で改善する、実施形態 2 4 9 に記載の方法。

【0 7 9 8】

2 5 1 . サンプル中の標的生物の力価を定量する方法であって、

a . 標的生物を含むサンプルを提供することであって、標的生物が、少なくとも 1 つの標的配列を含む、提供することと、

b . 少なくとも 3 つの種の生物の既知の力価を含む多検体対照を提供することであって、生物が不活性化されている、提供することと、

c . 実施形態 1 ~ 1 1 1 のいずれか 1 つに記載の正規化対照の既知の量を、サンプルおよび多検体対照と混合することと、

d . サンプルおよび多検体対照からハイスループットシーケンスライブラリを調製することと、

e . 当該ライブラリをシーケンスして、サンプルリードの集合および多検体対照リードの集合を產生することと、

f . 正規化対照を使用してサンプルリードの集合および (e) からの多検体対照リードの集合を正規化することと、

g . 正規化されたリードと多検体対照中の少なくとも 3 つの種の生物の既知の力価との間の関係性を判定することと、

h . (g) で判定された関係性を使用して、サンプル中の標的生物の a を計算することと、を含む、方法。

【0 7 9 9】

2 5 2 . 少なくとも 3 つの種の生物が各々、少なくとも 1 つの種特異的標的配列を含む、実施形態 2 5 1 に記載の方法。

【0 8 0 0】

2 5 3 . 少なくとも 1 つの種特異的標的配列が、多検体対照中の各種において異なる、実施形態 2 5 2 に記載の方法。

【0 8 0 1】

2 5 4 . 工程 (f) で正規化することが、

10

20

30

40

50

i. サンプルリードの集合からサンプルまたは正規化対照にリードをマップすることと、
ii. 多検体対照リードの集合から多検体対照または正規化対照にリードをマップすることと、

iii. サンプルリードの収集および多検体対照リードの収集のために、正規化対照中のポリヌクレオチドのグループまたはサブグループの各々によって產生されるリードの数を決定することと、

iv. 正規化対照中のポリヌクレオチドのグループまたはサブグループの各々の開始濃度と、サンプルおよび多検体対照の両方について工程 (e) で產生されるリードの数との間の関係性を計算することと、

v. サンプルおよび多検体対照におけるリードと濃度との関係性を決定することと、を含む、実施形態 251 に記載の方法。

10

【0802】

255. (g) でリードと力価との間の関係性を判定することが、

vi. 多検体対照中の少なくとも3つの種の各々から、少なくとも1つの種特異的標的配列の初期濃度を計算することと、

vii. サンプル中の少なくとも1つの標的配列の初期濃度を計算することと、を含む、実施形態 251 ~ 254 のいずれか1つに記載の方法。

【0803】

256. 工程 (i) でのマップが、サンプルリードの集合を参照ゲノムにマップすることを含む、実施形態 251 または 255 に記載の方法。

20

【0804】

257. 参照ゲノムが、標的配列を含まない、実施形態 256 に記載の方法。

【0805】

258. 参照ゲノムが、ヒトゲノムである、実施形態 256 または 257 に記載の方法。

【0806】

259. 工程 (ii) でのマップが、多検体対照中の少なくとも3つの種の生物の各々についての種特異的参照ゲノムへのマップをさらに含む、実施形態 254 ~ 258 のいずれか1つに記載の方法。

【0807】

260. 工程 (b) において、2濃度で、3濃度で、4濃度で、5濃度で、6濃度で、7濃度で、8濃度で、9濃度で、または10濃度で多検体対照を提供することと、工程 (b) で提供される多検体対照の各濃度に工程 (c) ~ (g) を適用することと、をさらに含む、実施形態 251 ~ 259 のいずれか1つに記載の方法。

30

【0808】

261. 工程 (g) での正規化されたリードと多検体対照中の少なくとも3つの種の生物の既知の力価との間の関係性が、線形関係である、実施形態 251 ~ 260 のいずれか1つに記載の方法。

【0809】

262. 各種が、同じ力価で多検体対照中に存在する、実施形態 251 ~ 261 のいずれか1つに記載の方法。

40

【0810】

263. 各種が、異なる力価で多検体対照中に存在する、実施形態 251 ~ 261 のいずれか1つに記載の方法。

【0811】

264. 力価が、線形配列または幾何配列にある、実施形態 263 に記載の方法。

【0812】

265. 少なくとも2つの種が、同じ力価で多検体対照中に存在し、少なくとも2つの種が、異なる力価で存在する、実施形態 251 ~ 261 のいずれか1つに記載の方法。

【0813】

266. 各種が、1 mL あたり 500 万 ~ 1000 万単位 (U/mL) の濃度で多検体

50

対照中に存在する、実施形態 251～265 のいずれか 1 つに記載の方法。

【0814】

267. 生物が、ウイルス、細菌、真菌、またはそれらの組み合わせである、実施形態 251～266 のいずれか 1 つに記載の方法。

【0815】

268. ウイルスが、DNA ウイルスである、実施形態 267 に記載の方法。

【0816】

269. ウイルスが、RNA ウイルスである、実施形態 267 に記載の方法。

【0817】

270. 細菌が、グラム陽性細菌またはグラム陰性細菌である、実施形態 267 に記載の方法。

10

【0818】

271. 生物が、ヒト病原体である、実施形態 251～270 のいずれか 1 つに記載の方法。

【0819】

272. 多検体対照中の各異なる種は、1 mL あたり約 500 万～約 1000 万の感染単位 (IU/mL) の力価で存在する、実施形態 271 に記載の方法。

【0820】

273. 多検体対照が、10 種のウイルスまたは 11 種のウイルスを含むか、または本質的にそれからなる、実施形態 271 または 272 に記載の方法。

20

【0821】

274. 多検体対照中の種は、表 8 に列挙される種の集合から選択される、実施形態 273 に記載の方法。

【0822】

275. ヒト病原体が、血液または組織移植に見出されるヒト病原体を含む、実施形態 271 または 272 に記載の方法。

【0823】

276. 血液または組織移植において見出されるヒト病原体が、サイトメガロウイルス (CMV)、エプスタインバーウイルス (EBV)、アデノウイルス (ADV)、BK ウイルス (BKV)、JC ポリオマウイルス (JCV)、ヒトヘルペスウイルス 6A (HHV 6A)、ヒトヘルペスウイルス 6B (HHV 6B)、単純ヘルペスウイルス 1 型 (HSV 1)、単純ヘルペスウイルス 2 型 (HSV 2)、水痘帯状疱疹ウイルス (VZV)、およびヒトパルボウイルス B19 (B19)、ヒト免疫不全ウイルス (HIV)、B 型肝炎、C 型肝炎、ヒト T リンパ性ウイルス I 型および II 型 (HTLV)、ウエストナイルウイルス、ジカウイルス、*Streptococcus* 種、*Staphylococcus aureus*、*Enterococcus faecalis*、*Bacillus cereus*、*Propionibacterium acnes*、*Serratia liquefaciens*、*Serratia marcescens*、*Yersinia enterocolitica*、*Enterobacter* spp.、*Acinetobacter* spp.、*Pseudomonas* spp.、*E. coli*、*Klebsiella pneumoniae*、*Proteus mirabilis*、*A. baumannii*、*Bacillus cereus*、コアグラゼ陰性 *staphylococci*、*Streptococcus* spp.、*Klebsiella* spp.、*Serratia* spp.、*Providencia rettgeri*、*Treponema pallidum*、*Trypanosoma cruzi*、および *Babesia microti* のうちの 1 つ以上を含む、実施形態 275 に記載の方法。

30

【0824】

277. 多検体対照中のヒト病原体が、表 2～3 または 5～7 から選択される、実施形態 271 または 272 に記載の方法。

【0825】

40

50

278. 生物が、ヒト病原体である種およびヒト病原体ではない種の混合物を含む、実施形態251～270のいずれか1つに記載の方法。

【0826】

279. 生物が、化学処理、熱処理、pH処理、または紫外線照射によって不活性化されている、実施形態251～278のいずれか1つに記載の方法。

【0827】

280. 化学処理が、ホルムアルデヒドを含む、実施形態279に記載の方法。

【0828】

281. 多検体対照が、血漿をさらに含む、実施形態251～280のいずれか1つに記載の方法。

【0829】

282. 血漿が、ヒト血漿である、実施形態281に記載の方法。

【0830】

283. 複数の正規化対照オリゴヌクレオチドを製造する方法であって、

a. 少なくとも1つの参照配列から複数の参照配列断片を生成することと、

b. 参照配列断片の数の関数として少なくとも1つのパラメータの分布を生成することと、

c. 分布を少なくとも5つのビンに分割することと、

d. 少なくとも5ビンのうちの少なくとも3つから、少なくとも1つの参照配列断片を選択することと、

e. 少なくとも3つの参照配列をシャッフルして、シャッフルされた配列を生成することと、

f. シャッフル配列を含むオリゴヌクレオチドを合成することと、を含み、

それにより複数の正規化対照オリゴヌクレオチドを生成する、方法。

【0831】

284. パラメータが、(1) GC含量パーセント、(2) エントロピー、(3) 複雑度、(4) EIIIP、またはそれらの組み合わせのうちの少なくとも1つを含む、実施形態283に記載の方法。

【0832】

285. パラメータが、GC含量パーセントを含む、実施形態283に記載の方法。

【0833】

286. 各ビンが、分布の1パーセント、2パーセント、5パーセント、10パーセント、15パーセント、または20パーセントを表す、実施形態283～285のいずれか1つに記載の方法。

【0834】

287. 参照配列断片が、GC含量パーセント分布の第10番目、第30番目、第50番目、第70番目、第80番目、および第90番目の百分位数のビンから選ばれる、実施形態283～285のいずれか1つに記載の方法。

【0835】

288. 工程(d)が、少なくとも3つのビンの各々から少なくとも2つの参照を選択することを含み、工程(e)が、各ビンからシャッフルされた参照配列を連結することをさらに含む、実施形態283～287のいずれか1つに記載の方法。

【0836】

289. 少なくとも1つの参照配列が、少なくとも2、少なくとも10、少なくとも20、少なくとも50、少なくとも100、少なくとも200、少なくとも300、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1,000、少なくとも1,200、少なくとも1,300、少なくとも1,400、少なくとも1,500、少なくとも1,600、少なくとも1,700、少なくとも1,800、少なくとも1,900、少なくとも2,000、少なくとも2,200、少なくとも2,400、少なくとも2,600、少なくとも2

10

20

30

40

50

、800、少なくとも3,000、少なくとも4,000、少なくとも5,000、少なくとも6,000、少なくとも7,000、少なくとも8,000、少なくとも9,000、または少なくとも10,000の参照配列を含む、実施形態283～288のいずれか1つに記載の方法。

【0837】

290. 少なくとも1つの参照配列が、約2,000個の参照配列を含む、実施形態283～288のいずれか1つに記載の方法。

【0838】

291. 少なくとも1つの参照配列が、ゲノム配列もしくはcDNA配列、またはそれらの組み合わせを含む、実施形態283～290のいずれか1つに記載の方法。

10

【0839】

292. 複数の参照配列断片が、参照配列の約10～100、約10～80、約20～60、約20～40、約20～30、または約20～25個の連続するヌクレオチドである断片を含む、実施形態283～291のいずれか1つに記載の方法。

【0840】

293. 各正規化対照オリゴが、参照配列分布における対応するビンからの参照配列断片（複数可）と（1）同様のGC含量パーセント、（2）同様のエントロピー、（3）同様のEIIIP、（4）同様の長さ、または（5）同様の複雑度のうちの少なくとも3つを有する、実施形態283～292のいずれか1つに記載の方法。

【0841】

20

294. 複数の正規化対照オリゴヌクレオチドを製造する方法であって、

a. スライディングウィンドウを使用して、少なくとも1つの参照配列から複数の参照配列断片を生成することと、

b. 参照配列断片の数の関数として少なくとも1つのパラメータの分布を生成することと、

c. 分布を少なくとも5つのビンに分割することと、

d. 少なくとも5つのビンのうちの少なくとも3つから、少なくとも2つの参照配列断片を選択することであって、少なくとも2つの参照配列断片が、参照配列中で連続していないか、または異なる参照配列からのものであるかのいずれかである、選択することと、

e. 少なくとも3つのビンのそれぞれからの少なくとも2つの参照配列断片を連結させることと、

30

f. 連結された参照配列断片を含むオリゴヌクレオチドを合成することと、を含み、それにより複数の正規化対照オリゴヌクレオチドを生成する、方法。

【0842】

295. パラメータが、（1）GC含量パーセント、（2）エントロピー、（3）複雑度、（4）EIIIP、またはそれらの組み合わせのうちの少なくとも1つを含む、実施形態294に記載の方法。

【0843】

296. パラメータが、GC含量パーセントを含む、実施形態294に記載の方法。

【0844】

40

297. 各ビンが、分布の1パーセント、2パーセント、5パーセント、10パーセント、15パーセント、または20パーセントを表す、実施形態294～296のいずれか1つに記載の方法。

【0845】

298. 参照配列断片が、GC含量パーセント分布の第10番目、第30番目、第50番目、第70番目、第80番目、および第90番目の百分位数のビンから選ばれる、実施形態294～296のいずれか1つに記載の方法。

【0846】

299. スライディングウィンドウが、1bp、2bp、3bp、4bp、または5bpの摺動ウィンドウを含む、実施形態294～298のいずれか1つに記載の方法。

50

【0847】

300. 参照配列断片が、対応する参照配列の約15～60、約20～40、約20～30、約15～32、20～32、または約25～35個の連続するヌクレオチドである、実施形態294～299のいずれか1つに記載の方法。

【0848】

301. 参照配列断片が、対応する参照配列の29、30、31、32、33、または34個の連続するヌクレオチドである、実施形態294～299のいずれか1つに記載の方法。

【0849】

302. 参照配列断片が、シーケンスリードの平均長の半分未満である、実施形態294～301のいずれか1つに記載の方法。

10

【0850】

303. 少なくとも1つの参照配列が、少なくとも2、少なくとも10、少なくとも20、少なくとも50、少なくとも100、少なくとも200、少なくとも300、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1,000、少なくとも1,200、少なくとも1,300、少なくとも1,400、少なくとも1,500、少なくとも1,600、少なくとも1,700、少なくとも1,800、少なくとも1,900、少なくとも2,000、少なくとも2,200、少なくとも2,400、少なくとも2,600、少なくとも2,800、少なくとも3,000、少なくとも4,000、少なくとも5,000、少なくとも6,000、少なくとも7,000、少なくとも8,000、少なくとも9,000、または少なくとも10,000の参照配列を含む、実施形態294～302のいずれか1つに記載の方法。

20

【0851】

304. 少なくとも1つの参照配列が、約2,000個の参照配列を含む、実施形態294～302のいずれか1つに記載の方法。

【0852】

305. 参照配列が、ゲノム配列またはcDNA配列を含む、実施形態294～304のいずれか1つに記載の方法。

【0853】

306. 各正規化対照オリゴが、参照配列分布における対応するビンからの参照配列断片（複数可）と（1）同様のGC含量パーセント、（2）同様のエントロピー、（3）同様のEIP、（4）同様の長さ、または（5）同様の複雑度のうちの少なくとも3つを有する、実施形態294～305のいずれか1つに記載の方法。

30

【0854】

307. 実施形態283～306のいずれか1つに記載の方法によって產生される正規化対照。

【0855】

308. 実施形態307に記載の正規化対照を含むキット。

【0856】

309. 複数の正規化対照ポリヌクレオチド配列をデザインするためのシステムであって、

40

- a. コンピュータ実行可能命令を記憶するコンピュータ可読記憶媒体であって、
 - i. 少なくとも1つの参照配列をインポートするための命令と、
 - ii. 少なくとも1つの参照配列から複数の参照配列断片を生成するための命令と、
 - iii. 配列断片の数の関数として少なくとも1つのパラメータの分布を生成するための命令と、
 - iv. 分布をビンに分割するための命令と、
 - v. 少なくともビンのサブセットから複数の参照配列断片を選択するための命令と、
 - vi. 複数の参照配列断片をシャッフルして、シャッフルされた配列を生成するため

50

の命令と、を含み、

それにより複数の正規化ポリヌクレオチド配列を生成する、コンピュータ可読記憶媒体と、

b. 以下の工程を実行するように構成されたプロセッサであって、

i. 少なくとも少なくとも1つの参照配列を含むインプットファイルのセットを受信することと、

ii. コンピュータ可読記憶媒体に記憶されたコンピュータ実行可能命令を実行することと、を含む、プロセッサと、を備える、システム。

【0857】

310. 複数の正規化対照ポリヌクレオチド配列をデザインするためのシステムであって、

a. コンピュータ実行可能命令を記憶するコンピュータ可読記憶媒体であって、

i. 少なくとも1つの参照配列をインポートするための命令と、

ii. 少なくとも1つの参照配列から複数の参照配列断片を生成するための命令と、

iii. 配列断片の数の関数として少なくとも1つのパラメータの分布を生成するための命令と、

iv. 分布をビンに分割するための命令と、

v. ビンの少なくともサブセットの各々から、少なくとも2つの参照配列断片を選択するための命令であって、少なくとも2つの参照配列断片が、参照配列中で連続していないか、または異なる参照配列からのものであるかのいずれかである、命令と、

vi. 各ビンからの少なくとも2つの参照配列断片を連結するための命令と、を含み、それにより複数の正規化ポリヌクレオチド配列を生成する、コンピュータ可読記憶媒体と、

b. 以下の工程を実行するように構成されたプロセッサであって、

i. 少なくとも少なくとも1つの参照配列を含むインプットファイルのセットを受信することと、

ii. コンピュータ可読記憶媒体に記憶されたコンピュータ実行可能命令を実行することと、を含む、プロセッサと、を備える、システム。

【0858】

311. パラメータが、(1) GC含量パーセント、(2) エントロピー、(3) 複雑度、(4) EII P、またはそれらの組み合わせのうちの少なくとも1つを含む、実施形態309 j または310に記載のシステム。

【実施例】

【0859】

本明細書に開示される本発明をより効率的に理解できるようにするため、実施例を以下に提供する。これらの実施例は、例示目的のみのためであり、いかなる方法でも本発明を限定するものとして解釈されるべきではないことが理解されるべきである。

【0860】

実施例1：RNA正規化対照

RNA正規化対照は、ランダムに生成されたDNA配列を転写することによって行われ得る。ランダムに生成された配列を、転写のための末端T7プロモーター配列、およびXhoI制限部位を用いてgBlockとして合成して、ランダムDNAテンプレートを線形化した。RNA__NC__1、RNA__NC__2、およびRNA__NC__3と称される、以下に示されるデザインの3つのバージョンを作成した。各バージョンを、以下のデザインで作成したが、異なるランダムに生成された配列で作成した。

5'-TGAGGAAC TGCGGCAGG (配列番号12) (Mit oアレイプライマー)+175bpのランダム配列(ランダムDNAジェネレーターを使用してデザインされる)+T7 RNAポリメラーゼプロモーター+XhoI部位+250bpのランダム配列(ランダムDNAジェネレーターを使用してデザインされる)+T7 RNAポリメラーゼプロモーター+450bpのランダム配列(ランダムDNAジェネレーターを使用

10

20

30

40

50

してデザインされる) + T7 RNAポリメラーゼプロモーター + (リボアレイプライマー) GGTCTGCACCTTCCAGCT-3' (配列番号13)。

【0861】

Riboアレイプライマーの配列は、配列5'-AGCTGGAAGTGCAGACC-3' (配列番号20) を含む。

【0862】

T7プロモーターの配列は、5'-TAATACGACTCACTATAGGG-3' (配列番号15) の配列を含む。

【0863】

デザインの全長は981bpであった。異なるランダム配列を有するデザインの3つのバージョンを、gBlock (遺伝子ブロック) として指示し、RNA_{NC}1、RNA_{NC}2、およびRNA_{NC}3と称した。RNA_{NC}1、RNA_{NC}2、RNA_{NC}3のgBlockをXhoIで消化して、3つの線形DNA断片を産生し、得られた3つの断片の混合物をT7 RNAポリメラーゼで転写して、3つのRNA_{NC}を産生した。gBlock RNA_{NC}1は、3つのRNA断片: RNA_{NC}1.1、RNA_{NC}1.2、RNA_{NC}1.3をもたらした。gBlock RNA_{NC}2は、3つのRNA断片: RNA_{NC}2.1、RNA_{NC}2.2、RNA_{NC}2.3をもたらした。gBlock RNA_{NC}3は、3つのRNA断片: RNA_{NC}3.1、RNA_{NC}3.2、RNA_{NC}3.3をもたらした。RNA_{NC}1.1、RNA_{NC}2.1、およびRNA_{NC}3.1 NC RNAに対応するDNA断片のT7転写は、Mittoアレイプライマーに対して相補的な配列、175bpのランダムDNA配列に対して相補的な配列、および転写中にT7ポリメラーゼによって導入される末端GGG配列を含んだNC RNAをもたらした。NC RNA RNA_{NC}1.2、RNA_{NC}2.2、およびRNA_{NC}3.2に対応するDNA断片のT7転写は、250bpのランダム配列に対して相補的な配列および転写中に導入された末端GGG配列を含むNC RNAをもたらした。NC RNA RNA_{NC}1.3、RNA_{NC}2.3、およびRNA_{NC}3.3に対応するDNA断片のT7転写は、450bpのランダム配列に対して相補的な配列および転写中に導入された末端GGG配列を含むNC RNAをもたらした。

【0864】

サンプル抽出中にNC対照を添加するためのプロトコルを図6に示す。3つのNC転写反応 (それぞれが上述の3つの転写産物を含む、NC1、NC2、およびNC3、またはRNA_{NC}1、RNA_{NC}2、およびRNA_{NC}3とも呼ばれるRNA正規化対照1、2、および3のセット) の各々の出力を、20,000倍で0.1ng/μLまで希釈した。次いで、25μLのRNA_{NC}1、50μLのRNA_{NC}2、および50μLのRNA_{NC}3を混合することによって、NC転写産物の3つセットを1:2:4の比率で混合し、最終体積を175μLにした。このNC混合物は、合計9個の異なる転写産物を含み、3個は3個の異なる175bpのランダム配列を含み、3個は3個の異なる250bpのランダム配列を含み、3個は3個の異なる450bpランダム配列を含む。次いで、NC混合物をさらに10倍希釈し、30μLの混合物を270μLのAVE緩衝液に添加して、10pg/μLのNC混合物を産生した。

【0865】

EZ1ウイルスMini Kit (Qiagen) を使用して、核酸を陰性血漿サンプル (Exact Diagnostics) から抽出した。8μLの10pg/μLのNC混合物および52μLのAVE緩衝液を、EZ1キット3列目の1.5mLチューブに添加した。3列目のチューブから50μLのアリコートを採取し、400μLのサンプルに添加し、抽出中に70pgのNC混合物のサンプルへの添加をもたらした。

【0866】

抽出中に陰性血漿サンプルに3つの異なる量のNC混合物を添加した結果をアッセイした。「高度」の量のNCについて、サンプルあたり38.10pgのNC混合物を添加し

10

20

30

40

50

た（NC1：5.4 pg、NC2：10.8 pg、NC3：21.8 pg）。「中等度」の量のNCについて、サンプルあたり19.04 pgのNC混合物を添加した（NC1：2.7 pg、NC2：5.4 pg、NC3：10.8 pg）。「低度」の量のNCについて、サンプルあたり9.52 pgのNC混合物を添加した（NC1：1.4 pg、NC2：2.7 pg、NC3：5.4 pg）。「低度」の量は、NextSeq 500次世代シーケンス後に4%のリードをもたらした。重複排除されたリードに基づいて、ライブラリに変換された「低度」のNC1の質量は、1.076 pg、または入力76%であった。

【0867】

図7は、核酸抽出中に添加されたNCを有する正確な診断の陰性血漿から調製されたRNA-SeqライブラリにおけるRNA正規化対照の表現を要約する。ライブラリあたりの平均リード出力は23,000,000リードであり、平均ヒト表現は80%であった。RNA正規化対照の「高度」のレベルの添加（38.10 pg）は、これらのライブラリにおいて総リードの平均16%を表すリードをもたらした。RNA正規化対照の「中等度」のレベルの添加（19.04 pg）は、これらのライブラリにおいて総リードの平均6%を表すリードをもたらした。RNA正規化対照の「低度」のレベルの添加（9.52 pg）は、これらのライブラリにおいて総リードの平均4%を表すリードをもたらした。「低度」の入力時のRNA__NC1断片の変換率は76%（1.067 pg）であった。

【0868】

実施例2：DNA正規化対照

DNA NCは、ランダムに生成された配列を使用して行われ得る。以下に、DNA NCのデザインの一例を示す。このデザインの3つのバージョンは、異なるランダムに生成された配列で作成された：

5'-TGAAGAACTGCGGCAGG（配列番号12）（Mittoアレイプライマー）+ランダムDNAジェネレーターを使用するランダム配列（200 bp）+XhoI部位+ランダムDNAジェネレーターを使用するランダム配列（300 bp）+XhoI部位+ランダムDNAジェネレーターを使用するランダム配列（500 bp）+T7 RNAポリメラーゼプロモーター+（リボアレイプライマー）GGTCTGCACCTTCCAGCT-3'（配列番号13）。

【0869】

5'から3'まで、リボアレイプライマーの配列は、5'-AGCTGGAAGTGCAAGACC-3'である（配列番号20）。デザインの全長は1048 bpであった。デザインの3つのバージョンを、gBlock DNA__NC__1、gBlock DNA__NC__2、およびgBlock DNA__NC__3と称されるgBlock（遺伝子ブロック）として指示した。gBlock DNA__NC__1のXhoI消化は、3つの断片：DNA__NC__1.1、DNA__NC__1.2、およびDNA__NC__1.3をもたらした。gBlock DNA__NC__2の消化は、3つの断片：DNA__NC__2.1、DNA__NC__2.2、およびDNA__NC__2.3をもたらした。gBlock DNA__NC__3の消化は、3つの断片：DNA__NC__3.1、DNA__NC__3.2、およびDNA__NC__3.3をもたらした。断片DNA__NC__1.1、DNA__NC__2.1、およびDNA__NC__3.1はそれぞれ、MittoアレイプライマーのDNA配列、200 bpのランダム配列（それぞれ異なる）、およびXhoI半部位を含んでいた。断片DNA__NC__1.2、DNA__NC__2.2、およびDNA__NC__3.2は、300 bpのランダム配列を含み、2つのXhoI半部位に隣接した。断片DNA__NC__1.3、DNA__NC__2.3、およびDNA__NC__3.3は、XhoI半部位、500 bpのランダム配列、T7プロモーター配列、およびリボアレイプライマー配列を含んでいた。

【0870】

ゲノムDNAの抽出の前に、正規化対照をサンプルに添加した。次いで、正規化対照をシーケンスライブラリの残りとは並行してインデックス化した。

【0871】

正規化対照を有するDNAライブラリは、i7およびi5アダプターを有する二重インデックス戦略を使用して作成された。下記のi7およびi5配列を有するプライマーを、200 μMの0.5 μLの濃度で再懸濁し、i5およびi7プライマーをNEB緩衝液2中の総体積20 μLで混合した（図8）。次いで、i5およびi7プライマー混合物をサーモサイクラーに入れ、95℃に1分間ランプし、次いで、20℃にゆっくりと冷却した。次いで、i5/i7混合物を1200 μLの総体積（1反応用）で15 μMの濃度に希釈した。i7/i5アダプターを1:20にさらに希釈して、750 nMの総濃度を得た。デュアルインデックスプライマー

i701-UM16プライマー: 5' phos/GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC TAC AGG TCN NNN NNA TCT CGT (配列番号21)

i702-UM16プライマー: 5' phos/GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC AGT TAC ATN NNN NNA TCT CGT (配列番号22)

i501-UM16プライマー: AAT GAT ACG GCG ACC GAG ATC TAC ACN NNN NNC CAG TAC AAC ACT CTT TCC CTA (配列番号23)

i502-UM16プライマー: 5' AAT GAT ACG GCG ACC ACC GAG ATG TAC CAN NNN NNG TGC TTA TAC ACT CTT TGG CTA N (配列番号24)

【0872】

図9は、含まれるNCでシーケンスされた例示的なDNAライブラリにおけるNCリードのパーセントを示す。NC対照を、図12に示される濃度でサンプルに添加した。

【0873】

実施例3: サンプル中の非宿主の定量レベル

この実施例では、非宿主対NCの比率を決定することを使用して、混合宿主および非宿主核酸のサンプル中の非宿主定量化に対する異なる宿主バックグラウンド量の効果を減算する。

【0874】

NCは、別個のDNAもしくはRNAのいずれか、またはDNAおよびRNAの両方の混合物としての核酸のセットから構成される。様々なサイズのNCを混合して、異なる長さおよび塩基組成物の線形分布を提供し、NGSライブラリ内の異なる変動性をシミュレートする。NCに含まれるこれらの核酸は、合成的に製造され得るか、または天然核酸の培養混合物から製造され得る。断片は、非常に特異的なサイズ、核酸組成物を表すことができるか、または様々な濃度で非常に広範囲のサイズの断片をカバーすることができる。配列モチーフを含めて、ライブラリプロセスが配列反復または反復配列に対して偏っているかどうかを理解することができる

【0875】

NCは1つ以上の生物である。例えば、NCは、混合宿主/非宿主サンプル中の非宿主配列の1つ以上の特性をシミュレートする、ウイルス、細菌、真菌、または真核生物などの1つ以上の生物または生物種から抽出されるポリヌクレオチドを含む。

【0876】

NCを使用して、サンプル中の異なる核酸の相対量を判定する。例えば、正規化対照を使用して、感染した宿主（例えば、ウイルスまたは細菌に感染したヒト宿主）からのサンプル中の病原体（非宿主）の存在量値を決定する。病原体の存在量値は、病原体にマップするNGSリードを、異なるバックグラウンド宿主含量を補償するためにNCにマップするNGSリードの数で分割ことによって決定される。この値は、同じように生成され正規化された標準曲線にマップされ、病原体力価は標準曲線から推測される。かかる方法を使用して、必ずしも絶対病原体負荷を決定する必要はない。

10

20

30

40

50

【0877】

NCを使用して、サンプル中の異なる核酸の絶対量を判定する。例えば、NCを使用して、感染した宿主からのサンプル（例えば、ウイルスまたは細菌に感染したヒト宿主からのサンプル）中の病原体（非宿主）の絶対量を決定する。病原体の存在量値は、病原体にマップするNGSリードを、異なるバックグラウンド宿主含量を補償するためにNCにマップするNGSリードの数で分割ことによって決定される。この比率およびNCの入力量（例えば、バクテリオファージまたはポリヌクレオチドのコピー）に基づいて、相対病原体ゲノムコピーを決定する。次いで、この値を病原体ゲノムサイズについて補正して、出発物質中の「絶対」病原体存在量を導出する。

【0878】

いくつかの場合では、NCは、T4ファージ核酸配列を含み、病原体は、サイトメガロウイルスを含む。T4ファージおよびCMVゲノムはともに約200kbであるため、2つはほぼ同じサイズであり、プラーク形成単位/mL（pfu/mL、T4）およびコピー/mL（cp/mL、CMV）は同等であると仮定される。次世代シーケンス後、T4ファージNCの1リードごとに2CMVリードがあり、NCを100pfu/mLの濃度で添加した場合、CMVは200cp/mLの濃度であった必要がある。CMVがT4のゲノムサイズの1/2であった場合、CMVおよびT4は、初期サンプル中でほぼ同じ濃度（100pfu/mLまたはcp/mL）で存在するであろう。

【0879】

実施例4：ウイルス力価を決定するための多検体対照の使用

本開示の正規化対照を使用して、実験サンプル中のウイルスの力価を計算するために使用される、多検体対照におけるウイルスのパネルの力価を決定することができる。

【0880】

この実施例では、多検体対照は、以下の一般に見出されるヒト移植サンプルである、11個のウイルスの混合物である：サイトメガロウイルス（CMV）、エプスタインバーウイルス（EBV）、アデノウイルス（ADV）、BKウイルス（BKV）、JCポリオマウイルス（JCV）、ヒトヘルペスウイルス6A（HHV6A）、ヒトヘルペスウイルス6B（HHV6B）、単純ヘルペスウイルス1型（HSV1）、単純ヘルペスウイルス2型（HSV2）、水痘帯状疱疹ウイルス（VZV）、およびヒトパルボウイルスB19（B19）。11個のウイルスの混合物を使用して、異なる力価、例えば、0、100、1,000、および10,000個の感染単位（IU）/mLの直線性パネルを構築し、血漿中にスパイクする。多検体対照を異なるレベルで血漿中にスパイクして、多検体対照のキャリブレーションパネルを生成することができる。正規化対照をサンプルおよびキャリブレーションパネル内の多検体対照に添加し、核酸を抽出し、次いで、実験サンプルおよび多検体対照からハイスループットシーケンスライブラリ（複数可）を生成する。リードは、多検体対照パネルまたは正規化対照内の異なるウイルスにマップされ、正規化対照を使用して、多検体対照内のウイルスの各々からのリードカウントを正規化することができる。次いで、標準曲線を生成するために、多検体対照の力価に対して正規化されたリードカウントをプロットし（図13）、正規化されたリードと対数（IU/mL）との関係性の方程式を計算する。この実施例（図13）では、この関係性は比例である。実験サンプルからのウイルスシグナルは、正規化対照と同様に正規化され、多検体対照から計算された関係性を使用して、実験サンプル中のウイルス力価を計算する。

【0881】

多検体対照は、実験サンプルとは別に、完全なサンプル調製プロセスを経る陽性外部対照としても使用される。高濃度および低濃度の多検体対照を実験サンプルに含めることもでき、以前のキャリブレーションパネル、例えば、基準パネルが現在の条件に依然として適用可能であることを確実にする。

【0882】

実施例5：シーケンスシャッフルを使用する正規化対照シーケンスの生成

正規化対照をデザインするための1つのアプローチは、ベースシャッフルアプローチ（

10

20

30

40

50

Cantor Shuffling) によるものである。このデザインプロセスの図を図 14 A に示す。シャッフルアプローチでは、標的配列（複数可）のすべての同じ塩基が同じ比率で保持され、標的配列（複数可）のパラメータが保持されるが、正規化対照の配列はシャッフルされ、標的配列（複数可）にマップされない。

【0883】

潜在的病原体を表す約 1200~2000 の参照ゲノムを使用して、ゲノム長に関して GC 含量、エントロピー、複雑度、および電子イオン間相互作用ポテンシャル (E I I P) の分布を生成した。参照ゲノムの GC 含量、エントロピー、および E I I P の間の相互作用を示す 3 元系プロットを図 15 に示す。

【0884】

異なる分布 (GC 含量、エントロピー、複雑度、E I I P) に関して x 軸 (長さ) の第 10 番目、第 30 番目、第 50 番目、第 70 番目、第 80 番目、および第 90 番目の百分位数にあった長さピンから配列を得た。各配列を、それ自体内で 20~25mer ウィンドウによってシャッフルし、潜在的なアレイの配列のリポジトリに配置された。各配列はまた、それ自体内で 100mer ウィンドウによってシャッフルされ、「マップ対照」として使用するために配列のリポジトリに配置された。

【0885】

潜在的なアレイおよび「マップ対照」を参照ゲノムに対して分画し、NCBI 非冗長配列データベース (NR) およびガイド配列に対してブラストした。マップされた潜在的なアレイからの配列を排除した。正規化対照を、55 の千鳥状の濃度でデザインした。

【0886】

正規化対照を、プロビット (1000 cp/ml)、臨床サンプル、オリゴの同定および定量化、病原体の同定および定量化に情報学的にスパイクインした。「マップ対照」配列を参照に対してマップして、シャッフルが機能することを証明した。

【0887】

対数濃度あたり 1~3 オリゴでデザインされた正規化対照の例を図 14 B および 14 C に示す。対数濃度あたり少なくとも 1 つの異なるオリゴを有する 6 個の濃度で表される濃度の各規模の順序、かつカバーされる 7~8 の規模の順序でデザインされた例の正規化対照を図 14 D に示す。

【0888】

図 16 および 17 は、参照ゲノムのパネルを使用して生成された Kmer における E I I P、エントロピー、および GC 含量の分布を示す。32 億 Kmers を生成し、20 億 Kmers (図 16)、1000 万 Kmers (図 17 A)、および 130 万 Kmers (図 17 B) を使用して初期プロファイルを構築した。Kmer をシャッフルして、参照ゲノムに見られる E I I P、エントロピー、および GC 含量の分布を模倣した。

【0889】

Python におけるマルチ処理を使用して、Kmer データセットで処理を実行することができる (例えば、Process および Pool)。Pandas および Process を使用して、大きなファイルに対して同様の操作を行う。Python におけるデータマッシュを 12 億行の Kmer (7 つのファイルに分割) に使用して、特徴のパーセントイル値を取得した。AWK を使用して選択を実行し、選択した Kmer をファイルに書き込み、これを使用してオリゴを構築した。

【0890】

Kmer を処理し、出力オリゴを生成するための例示的なパイプラインを図 18 に示す。Kmer は、正規化対照オリゴ構築で何回使用されか、およびグローバル変数 DIST_RANGE = {1KB...10KB} が開始されたことを示すためにインデックス化される。出力データフレーム (正規化対照オリゴの) は、長さおよび位置が DIST_RANGE 内であるかどうか、および生成されるオリゴの数を確認する。「いいえ」の場合、カバーされていない長さピンが選択され、次いで、ステップ 1 で作成されたインデックスから最も使用されていない Kmer が選択される。これらは、デザインされた正規化対照長

10

20

30

40

50

が達成されるまで連結され、所望の数のオリゴが生成されるまでプロセスが繰り返される。次いで、配列をシャッフル関数に進める。

【0891】

図19A～19Dは、Cantorシャッフルによって生成された正規化対照が、BKポリオーマウイルス分離株CH-1、JCポリオーマウイルス株NIID 12-31、ヒトポリオーマウイルス1株BK2およびヒトアデノウイルスCについて、EIP、複雑度、およびGC含量を模倣することを示す。図20A～20Dは、正規化対照が、国立バイオテクノロジーインフォメーションセンター(NCBI)BLASTヌクレオチド配列サーバ(blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE__TYPE=BlastSearch)を使用して、BKポリオーマウイルス分離株CH-1、JCポリオーマウイルスNIID 12-31、ヒトポリオーマウイルス1株BK2、およびヒトアデノウイルスCにBlastしないことを示す。

10

【0892】

実施例6：スライディングウィンドウを使用する正規化対照シーケンスの生成（徹底的なKmerアプローチ）

実施例5に記載されるCantorシャッフルに対する代替アプローチは、標的配列からの天然配列を使用することである。このアプローチでは、天然配列にわたってスライディングウィンドウを取り、千鳥状の重複したKmersを生成する。次いで、非隣接Kmerを連結して、正規化対照配列を産生した。約31または32bpのKmerサイズを使用する場合、約150bpのシーケンスリードは複数のKmerを包含し、標的配列（複数可）に隣接しないKmerの並列配置を通じて正規化対照配列を同定することができる。

20

【0893】

このアプローチでは、単一のゲノムは、そのゲノムがどのように断片化するかを表す複数の配列をゲノム全体に広げることによって表すことができる。さらに、個々の正規化対照は、複数の生物からのKmerを組み込むことができ、複数の生物のパラメータと一致し得る。例えば、正規化対照オリゴは、長いKmerを含むことができ、長いKmer内では、他の生物からの配列を含むより短いKmerであり得る。

【0894】

約1200～2000の参照ゲノムを使用して、千鳥状のKmerを生成し、Kmer分布を実施例5について上述のように計算した。

30

【0895】

Kmersは、最初にGC含量パーセントに基づいて選択され、すなわち、GC含量のパーセンテージの設定された数が正規化対照に含まれるように選択された。次に、選択されたGC含量パーセントのKmerを連結して、GCピン上に1キロ塩基(KB)オリゴを生成した。連結された配列のGC含量を確認し、エントロピーも確認した。次いで、配列をホモポリマーフィルターに通し、ブラスト(Blastnおよびガイドブラスト)して、それぞれ1Kbの長さの785,000オリゴを産生した。

【0896】

図21A～21Cは、選択されたGC含量パーセンテージで生成されたKmerのEIP、エントロピー、およびGC含量の分布を示す。これらの方法を使用して生成された参照ゲノム配列および正規化対照配列の両方に対して主成分分析(PCA)を行った場合、両方ともエントロピー、EIP、およびGC含量の間で同様の相関を示した(図22)。さらに、この方法を使用して生成された正規化対照のエントロピー、EIP、およびGC含量は、参照ゲノム配列のエントロピー、EIP、およびGC含量と相関していた(図23)。加えて、参照ゲノム配列のエントロピーの確率分布を正規化対照と比較するKolmogorov-Smirnov(KS)試験を実行した(図24)。

40

【0897】

785,000個の正規化対照オリゴ配列が参照ゲノム配列に整列しなかったことを、NCBI BLASTヌクレオチド配列サーバを使用してBLASTによって検証した（

50

図25A～25B)。正規化対照オリゴの頭部は、50～80塩基について、常に約750bpでPseudomonasにBLASTした(図25A)。ファイルの尾部にはBLASTヒットはなかった(図25)。すべての785,000個の正規化対照オリゴ配列は、質量BLASTされ得、BLASTヒットがない配列のみ、正規化対照として使用するために選択され得る。

本発明のまた別の態様は、以下のとおりであってもよい。

〔1〕 正規化対照であって、少なくとも3つのポリヌクレオチドのグループを含み、各グループ内の前記ポリヌクレオチドは同じ長さのものである、正規化対照。

〔2〕 前記正規化対照中のすべての前記ポリヌクレオチドの前記長さは同じである、前記〔1〕に記載の正規化対照。

〔3〕 各グループ内の前記ポリヌクレオチドは、他のグループ内の前記ポリヌクレオチドと比較した場合、異なる長さのものである、前記〔1〕に記載の正規化対照。

〔4〕 前記ポリヌクレオチドのグループの前記長さは、線形配列または幾何学配列で配置される、前記〔3〕に記載の正規化対照。

〔5〕 前記少なくとも3つのグループ中の前記ポリヌクレオチドの長さは、約15bpおよび約50,000bpの間である、前記〔1〕～〔4〕のいずれか一項に記載の正規化対照。

〔6〕 前記少なくとも3つのグループ中の前記ポリヌクレオチドの長さは、約500bpおよび約1500bpの間、約100および1200bpの間、または約150および約600bpの間である、前記〔1〕～〔4〕のいずれか一項に記載の正規化対照。

〔7〕 各グループ内の前記ポリヌクレオチドが、同じ配列を含む、前記〔1〕～〔6〕のいずれか一項に記載の正規化対照。

〔8〕 各グループ内の前記ポリヌクレオチドが、同じ配列を含まない、前記〔1〕～〔6〕のいずれか一項に記載の正規化対照。

〔9〕 各ポリヌクレオチドのグループが、少なくとも3つのポリヌクレオチドのサブグループを含み、

各サブグループ内の前記ポリヌクレオチドは、同じ配列を含み、かつ、各サブグループ内の前記ポリヌクレオチドは、他のサブグループとは同じ配列を含まない、前記〔8〕に記載の正規化対照。

〔10〕 少なくとも1つのポリヌクレオチドのグループ中の少なくとも1つの配列が、コンポーネント配列を含む、前記〔7〕～〔9〕のいずれか一項に記載の正規化対照。

〔11〕 すべてのポリヌクレオチドのグループ中のすべての配列が、コンポーネント配列を含む、前記〔7〕～〔9〕のいずれか一項に記載の正規化対照。

〔12〕 各ポリヌクレオチドのグループの前記コンポーネント配列が、他のグループの前記コンポーネント配列と同じではない、前記〔10〕または〔11〕に記載の正規化対照。

〔13〕 少なくとも1つのポリヌクレオチドのグループの少なくとも1つのサブグループの前記配列が、コンポーネント配列を含む、前記〔9〕に記載の正規化対照。

〔14〕 少なくとも1つのグループの前記少なくとも3つのサブグループのそれぞれの前記配列が、コンポーネント配列を含む、前記〔9〕に記載の正規化対照。

〔15〕 すべてのグループの各サブグループの前記配列が、コンポーネント配列を含む、前記〔9〕に記載の正規化対照。

〔16〕 各サブグループの前記コンポーネント配列が、他のグループの前記コンポーネント配列と同じではない、前記〔14〕または〔15〕に記載の正規化対照。

〔17〕 前記コンポーネント配列が、約6bpおよび約3000bpの間の配列を含む、前記〔10〕～〔16〕のいずれか一項に記載の正規化対照。

〔18〕 前記コンポーネント配列が、ランダム配列を含む、前記〔10〕～〔17〕のいずれか一項に記載の正規化対照。

〔19〕 少なくとも1つのポリヌクレオチドのグループの前記配列が、単離配列を含む、前記〔7〕～〔10〕のいずれか一項に記載の正規化対照。

〔20〕 すべてのポリヌクレオチドのグループの前記配列が、単離配列を含む、前記〔7〕

10

20

30

40

50

〕～〔 9 〕のいずれか一項に記載の正規化対照。

〔 2 1 〕少なくとも1つのポリヌクレオチドのグループの少なくとも1つのサブグループの前記配列が、単離配列を含む、前記〔 9 〕、〔 1 0 〕または〔 1 2 〕～〔 1 4 〕に記載の正規化対照。

〔 2 2 〕少なくとも1つのグループのすべてのサブグループの前記配列が、単離配列を含む、前記〔 9 〕、〔 1 0 〕または〔 1 2 〕～〔 1 4 〕に記載の正規化対照。

〔 2 3 〕すべてのグループのすべてのサブグループの前記配列が、単離配列を含む、前記〔 9 〕に記載の正規化対照。

〔 2 4 〕各サブグループの前記単離配列が、他のサブグループの前記単離配列と同じではない、前記〔 2 2 〕または〔 2 3 〕に記載の正規化対照。

〔 2 5 〕前記単離配列が、約 6 b p および約 2 0 0 , 0 0 0 b p の間、約 1 5 b p および約 5 0 , 0 0 0 b p の間、約 5 0 0 b p および約 1 5 0 0 b p の間、約 1 0 0 b p および約 1 2 0 0 b p の間、または約 1 5 0 b p および約 6 0 0 b p の間の配列を含む、前記〔 1 9 〕～〔 2 3 〕のいずれか一項に記載の正規化対照。

〔 2 6 〕前記単離配列が、ウイルス、細菌、真菌、真核性寄生虫から単離される、または由来する、前記〔 1 9 〕～〔 2 5 〕のいずれか一項に記載の正規化対照。

〔 2 7 〕前記単離配列が、配列サンプル中の少なくとも1つの標的配列と同じではない、前記〔 1 9 〕～〔 2 6 〕のいずれか一項に記載の正規化対照。

〔 2 8 〕前記正規化対照中の少なくとも1つのポリヌクレオチドのサブグループまたはグループの少なくとも1つの配列が、前記配列サンプル中の前記少なくとも1つの標的配列と少なくとも1つの配列特性を共有する、前記〔 9 〕～〔 2 7 〕のいずれか一項に記載の正規化対照。

〔 2 9 〕前記少なくとも1つの配列特性が、G C 含量パーセント、エントロピー、複雑度、長さ、電子イオン間相互作用ポテンシャル (E I I P) 、転位性要素配列の配列特性、ウイルス配列、細菌配列、真菌配列、寄生虫配列、1つまたは複数のヒト遺伝子配列、またはこれらの組み合わせを含む、前記〔 2 8 〕に記載の正規化対照。

〔 3 0 〕前記配列サンプルが、宿主および非宿主核酸の混合を含む、前記〔 2 8 〕または〔 2 9 〕に記載の正規化対照。

〔 3 1 〕前記宿主が、真核生物である、前記〔 3 0 〕に記載の正規化対照。

〔 3 2 〕前記宿主が、昆虫、植物、または動物である、前記〔 3 0 〕に記載の正規化対照。

〔 3 3 〕前記動物が、ヒトである、前記〔 3 2 〕に記載の正規化対照。

〔 3 4 〕前記非宿主が、共生生物 (s y m b i o t e) 、共生生物 (c o m m e n s a l o r g a n i s m) 、寄生虫または病原体を含む、前記〔 3 0 〕～〔 3 2 〕のいずれか一項に記載の正規化対照。

〔 3 5 〕前記非宿主が、複数の生物種を含む、前記〔 3 4 〕に記載の正規化対照。

〔 3 6 〕前記非宿主が、ある種のウイルス、ある種の細菌、ある種の真菌もしくはある種の寄生虫、またはこれらの組み合わせを含む、前記〔 3 0 〕～〔 3 3 〕のいずれか一項に記載の正規化対照。

〔 3 7 〕前記少なくとも1つの標的配列が、前記配列サンプル中の非宿主配列である、前記〔 2 7 〕～〔 3 6 〕のいずれか一項に記載の正規化対照。

〔 3 8 〕前記少なくとも1つの標的配列が、複数の参照配列を含む、前記〔 2 7 〕～〔 3 6 〕のいずれか一項に記載の正規化対照。

〔 3 9 〕前記複数の参照配列が、少なくとも2、少なくとも10、少なくとも50、少なくとも100、少なくとも200、少なくとも300、少なくとも400、少なくとも500、少なくとも600、少なくとも700、少なくとも800、少なくとも900、少なくとも1,000、少なくとも1,200、少なくとも1,300、少なくとも1,400、少なくとも1,500、少なくとも1,600、少なくとも1,700、少なくとも1,800、少なくとも1,900、少なくとも2,000、少なくとも2,200、少なくとも2,400、少なくとも2,600、少なくとも2,800、少なくとも3,000、少なくとも4,000、少なくとも5,000、少なくとも6,000、少なく

10

20

30

40

50

とも7,000、少なくとも8,000、少なくとも9,000または少なくとも10,000の参照配列を含む、前記〔38〕に記載の正規化対照。

〔40〕前記参照配列が、非宿主配列を含む、前記〔39〕に記載の正規化対照。

〔41〕前記参照配列が、ウイルス配列、細菌配列、真菌配列、またはこれらの組み合わせを含む、前記〔39〕または〔40〕に記載の正規化対照。

〔42〕前記複数の参照配列が、複数の生物からの参照配列を含む、前記〔38〕～〔41〕のいずれか一項に記載の正規化対照。

〔43〕前記参照配列が、ゲノムDNA配列もしくはcDNA配列、またはこれらの組み合わせを含む、前記〔42〕に記載の正規化対照。

〔44〕前記単離配列が、シャッフルされた少なくとも1つの参照配列の少なくとも1つの断片を含む、前記〔38〕～〔43〕のいずれか一項に記載の正規化対照。

〔45〕前記単離配列が、シャッフルされ、連結された少なくとも第1の参照および第2の参照配列の少なくとも2つの断片を含む、前記〔44〕に記載の正規化対照。

〔46〕前記第1および第2の参照配列が、生物のゲノム中で隣接してない、列前記〔44〕に記載の正規化対照。

〔47〕前記第1のおよび第2の参照配列が、異なる生物からのものである、列前記〔44〕に記載の正規化対照。

〔48〕前記参照配列の前記少なくとも1つの断片が、約10～100、約10～80、約20～60、約20～40、約20～30、または約20～25の連続した前記参照配列のヌクレオチドを含む、前記〔44〕～〔47〕のいずれか一項に記載の正規化対照。

〔49〕 a. 前記少なくとも1つの参照配列が参照配列断片の分布により表され、
b. 前記分布が少なくとも5つのビンに分割され、

c. 複数の参照配列断片が、前記少なくとも5つのビンのうちの少なくとも3つから選択され、シャッフルされ、任意に連結され、それにより単離配列を生成する、前記〔34〕～〔38〕のいずれか一項に記載の正規化対照。

〔50〕前記参照配列断片が、参照配列長さ、および（1）GC含量パーセント、（2）エントロピー、（3）複雑度、（4）EIIIP、またはこれらの組み合わせのうちの少なくとも1つに対して分布される、前記〔49〕に記載の正規化対照。

〔51〕各ビンが、前記分布の1パーセント、2パーセント、5パーセント、10パーセント、15パーセント、または20パーセントを表す、前記〔49〕または〔50〕に記載の正規化対照。

〔52〕各ビンが、GC含量の分布の5パーセント、10パーセント、15パーセント、または20パーセントを表す、前記〔49〕または〔50〕に記載の正規化対照。

〔53〕前記正規化対照の各グループ内の前記ポリヌクレオチドが、シャッフルすること、任意に連結することにより生成され、参照配列断片は、前記少なくとも3つのビンのうちの1つから選択される、前記〔50〕～〔52〕のいずれか一項に記載の正規化対照。

〔54〕各グループ内の前記ポリヌクレオチドが、前記参照配列分布中の対応するビンからの前記参照配列断片と、（1）同様のGC含量パーセント、（2）同様のエントロピー、（3）同様のEIIIP、（4）同様の長さ、または（5）同様の複雑度のうちの少なくとも3つを有する、前記〔51〕～〔53〕のいずれか一項に記載の正規化対照。

〔55〕参照配列断片が、前記GC含量パーセント分布からの第10番目、第30番目、第50番目、第70番目、第80番目の百分位数のビンから選ばれる、前記〔49〕～〔54〕のいずれか一項に記載の正規化対照。

〔56〕前記単離配列が、連結された少なくとも1つの参照配列の少なくとも2つの断片を含み、前記少なくとも2つの断片は、前記少なくとも1つの参照配列中で隣接していない、前記〔38〕～〔43〕のいずれか一項に記載の正規化対照。

〔57〕前記単離配列が、少なくとも第1の参照配列の少なくとも第1の断片、および少なくとも第2の参照配列の少なくとも第2の断片を含み、前記少なくとも第1のおよび第2の断片は、連結しており、前記第1のおよび第2の参照配列は、同じ参照配列ではない、前記〔38〕～〔43〕のいずれか一項に記載の正規化対照。

10

20

30

40

50

〔５８〕前記少なくとも第１のおよび第２の断片が、対応する参照配列の約１５～６０、約２０～４０、約２０～３０、約１５～３２、２０～３２または約２５～３５の連続したヌクレオチドである、前記〔５６〕または〔５７〕に記載の正規化対照。

〔５９〕前記少なくとも第１のおよび第２の断片が、対応する参照配列の２９、３０、３１、３２、３３または３４の連続したヌクレオチドである、前記〔５６〕または〔５７〕に記載の正規化対照。

〔６０〕前記少なくとも第１のおよび第２の断片が、シーケンスリードの平均長さの半分より少ない、前記〔５６〕～〔５９〕のいずれか一項に記載の正規化対照。

〔６１〕a. スライディングウィンドウを使用して生成されたオーバーラップする参照配列断片の分布により表され、

b. 前記分布は少なくとも５つのビンに分割され、

c. 複数の参照配列断片が、前記少なくとも５つのビンのうちの少なくとも３つから選択され、それにより単離配列を生成する、前記〔５６〕～〔６０〕のいずれか一項に記載の正規化対照。

〔６２〕前記参照配列断片が、参照配列長さ、および少なくとも（１）ＧＣ含量パーセント、（２）エントロピー、（３）複雑度、（４）ＥＩＩＰ、またはこれらの組み合わせの少なくとも１つに対して分布される、前記〔６１〕に記載の正規化対照。

〔６３〕各ビンが、前記分布の１パーセント、２パーセント、５パーセント、１０パーセント、１５パーセント、または２０パーセントを表す、前記〔６１〕または〔６２〕に記載の正規化対照。

〔６４〕各ビンが、ＧＣ含量の分布の５パーセント、１０パーセント、１５パーセント、または２０パーセントを表す、前記〔６１〕または〔６２〕に記載の正規化対照。

〔６５〕各グループ内の前記ポリヌクレオチドが、前記少なくとも３つのビンのうちの１つから選択される参照配列断片を連結することにより生成される、前記〔６１〕～〔６４〕のいずれか一項に記載の正規化対照。

〔６６〕各グループ内の前記ポリヌクレオチドが、前記参照配列分布中の対応するビンからの前記参照配列断片と、（１）同様のＧＣ含量パーセント、（２）同様のエントロピー、（３）同様のＥＩＩＰ、（４）同様の長さ、または（５）同様の複雑度のうちの少なくとも３つを有する、前記〔６１〕～〔６５〕のいずれか一項に記載の正規化対照。

〔６７〕前記スライディングウィンドウが、１bp、２bp、３bp、４bpまたは５bpスライディングウィンドウを含む、前記〔６１〕～〔６６〕のいずれか一項に記載の正規化対照。

〔６８〕参照配列断片が、前記ＧＣ含量パーセント分布からの第１０番目、第３０番目、第５０番目、第７０番目、第８０番目の百分位数のビンから選ばれる、前記〔６１〕～〔６７〕のいずれか一項に記載の正規化対照。

〔６９〕前記配列サンプルが、癌と診断された対象からのサンプルを含む、前記〔２７〕～〔６８〕のいずれか一項に記載の正規化対照。

〔７０〕前記配列サンプルが、メタゲノムのサンプルを含む、前記〔２７〕～〔６９〕のいずれか一項に記載の正規化対照。

〔７１〕前記少なくとも１つの標的配列が、前記メタゲノムのサンプル内のある種の配列を含む、前記〔７０〕に記載の正規化対照。

〔７２〕すべてのグループの前記ポリヌクレオチドが、同じ濃度にある、前記〔１〕～〔７１〕のいずれか一項に記載の正規化対照。

〔７３〕すべてのグループの前記ポリヌクレオチドが、同じ濃度でない、前記〔１〕～〔７１〕のいずれか一項に記載の正規化対照。

〔７４〕グループ内の前記ポリヌクレオチドが、同じＧＣ含量パーセントを有する、前記〔１〕～〔７３〕のいずれか一項に記載の正規化対照。

〔７５〕グループ内の前記ポリヌクレオチドが、他のグループ内の前記ポリヌクレオチドとは同じＧＣ含量パーセントを有さない、前記〔７４〕に記載の正規化対照。

〔７６〕前記ポリヌクレオチドが、一意的分子識別子をコードする配列を含む、前記〔１

10

20

30

40

50

〕～〔 7 5 〕のいずれか一項に記載の正規化対照。

〔 7 7 〕前記一意的分子識別子をコードする前記配列が、第 1 の索引配列または第 2 の索引配列を含む、前記〔 7 6 〕に記載の正規化対照。

〔 7 8 〕前記ポリヌクレオチドが、デオキシリボ核酸（DNA）分子、リボ核酸（RNA）分子、またはDNA-RNAハイブリッド分子を含む、前記〔 1 〕～〔 7 7 〕のいずれか一項に記載の正規化対照。

〔 7 9 〕前記ポリヌクレオチドが、前記正規化対照中、1 マイクロリットル（ μL ）当たり、約 0.005 および約 500 ピコグラム（ pg ）の間の濃度にある、前記〔 7 8 〕に記載の正規化対照。

〔 8 0 〕少なくとも 1 つのタンパク質をさらに含む、前記〔 1 〕～〔 7 9 〕のいずれか一項に記載の正規化対照。

〔 8 1 〕前記少なくとも 1 つのタンパク質が、細菌、ウイルスまたは寄生虫から単離される、または由来する、前記〔 8 0 〕に記載の正規化対照。

〔 8 2 〕正規化対照を製造する方法であって、

- a. 少なくとも 3 つの DNA 分子のグループを合成することであって、
グループ内の前記 DNA 分子は同じ配列を有し、それぞれの前記 DNA 分子の前記配列が、5'～3' の、第 1 のコンポーネント配列、第 1 の制限酵素部位、第 2 のコンポーネント配列、第 2 の制限部位、および第 3 のコンポーネント配列を含む、合成することと、
- b. 溶液中のそれぞれの前記 DNA 分子のグループの量を定量することと、
- c. それぞれの前記 DNA 分子のグループを、制限酵素で消化し、前記第 1 のおよび第 2 の制限部位を切断して少なくとも 3 つの消化された DNA 分子のグループを生成することと、
- d. 所定のそれぞれの前記 DNA 分子のグループの量を混合することで、正規化対照を生成することと、を含む、方法。

〔 8 3 〕 RNA 分子を含む正規化対照を製造する方法であって、

- a. 少なくとも 3 つの DNA 分子のグループを合成することであって、前記少なくとも 3 つの DNA 分子のそれぞれが、5'～3' の、第 1 のコンポーネント配列、プロモーター配列に相補的な配列、第 1 の制限酵素部位、第 2 のコンポーネント配列、プロモーター配列に相補的な第 2 の配列、第 2 の制限部位、第 3 のコンポーネント配列およびプロモーター配列に相補的な第 3 の配列を含む、合成することと、
- b. 各 DNA 分子のグループを、制限酵素で消化し、前記第 1 のおよび第 2 の制限酵素部位を切断して少なくとも 3 つの消化された DNA 分子のグループを生成することと、
- c. 前記少なくとも 3 つの DNA 分子のグループのそれぞれを、インビトロで複製して、少なくとも 3 つの RNA 分子の集団を生成することと、
- d. 前記 RNA 分子の集団のそれぞれに生成された RNA の量を定量することと、
- e. 所定の前記 RNA 分子の集団のそれぞれの量を混合することで、RNA 分子を含む正規化対照を生成することと、を含む、方法。

〔 8 4 〕少なくとも 3 つのポリヌクレオチドのグループを含む正規化対照を製造する方法であって、各グループ内の前記ポリヌクレオチドは同じ長さのものであり、前記方法は、

- i. サンプルから DNA を抽出することと、
- i i. 前記 DNA を制限酵素で消化して、DNA 断片の集合を生成することと、
- i i i. 前記 DNA 断片の集合を分離させることと、
- i v. 少なくとも 3 つのポリヌクレオチドのグループを生成するために、前記 DNA 断片を精製することであって、各グループ内の前記ポリヌクレオチドは同じ長さのものである、精製することと、
- v. 所定のポリヌクレオチドのグループのそれぞれの量を混合して、前記正規化対照を生成することと、を含む、方法。

〔 8 5 〕前記〔 8 3 〕～〔 8 5 〕のいずれか一項に記載の方法によって生成された正規化対照。

〔 8 6 〕前記〔 1 〕～〔 8 2 〕のいずれか一項に記載の正規化対照を含む、キット。

10

20

30

40

50

〔 8 7 〕 少なくとも 3 つの異なる生物の種の混合物を含む多検体対照であって、前記生物が不活化されている、多検体対照。

〔 8 8 〕 前記多検体対照が、10 種のウイルス種もしくは 11 種のウイルス種を含むか、またはそれから本質になる、前記〔 8 7 〕に記載の多検体対照。

〔 8 9 〕 少なくとも 1 つの標的配列の発現レベルを定量する方法であって、前記方法は、

a. 前記〔 1 〕～〔 8 2 〕のいずれか一項に記載の正規化対照の既知の量を前記サンプルと混合することと、

b. ハイスループットシーケンスライブラリを調製することと、

c. 前記ライブラリをシーケンスしてリードの集合を産生することと、

d. 前記リードの集合からのリードを前記サンプルまたは前記正規化対照にマップすることと、

e. 前記正規化対照中にそれぞれのポリヌクレオチドの前記グループまたはサブグループによって産生されたリードの数を判定することと、

f. (a) にて前記サンプルと混合された前記正規化対照中の各ポリヌクレオチドのグループまたはサブグループの開始濃度と、(c) にて産生されたリードの数との間の関係性を計算することと、

g. リードと前記サンプル中の前記少なくとも 1 つの標的配列の濃度の間の関係性をモデリングすることと、

h. 前記サンプル中の前記少なくとも 1 つの標的ポリヌクレオチドの初期濃度を、(g) のモデルを使用して前記標的配列によって産生されたリードの数から計算することと、を含む、方法。

〔 9 0 〕 サンプル中の標的生物の力価を定量する方法であって、

a. 前記標的生物を含むサンプルを提供することであって、前記標的生物が、少なくとも 1 つの標的配列を含む、提供することと、

b. 少なくとも 3 つの生物の種の既知の力価を含む、多検体対照を提供することであって、前記生物が不活化されている、提供することと、

c. 前記〔 1 〕～〔 8 0 〕のいずれか一項に記載の正規化対照の既知の量を前記サンプルおよび前記多検体対照と混合することと、

d. ハイスループットシーケンスライブラリを前記サンプルおよび前記多検体対照から調製することと、

e. 前記ライブラリをシーケンスしてサンプルリードの集合および多検体対照リードの集合を産生することと、

f. (e) からの前記サンプルリードの集合および前記多検体対照リードの集合を、前記正規化対照を使用して正規化することと、

g. 正規化されたリードと前記多検体対照中の前記少なくとも 3 つの生物の種の既知の力価との間の関係性を判定することと、

h. (g) で判定された前記関係性を使用して前記サンプル中の前記標的生物を計算することと、を含む、方法。

〔 9 1 〕 複数の正規化対照オリゴヌクレオチドを製造する方法であって、

a. 少なくとも 1 つの参照配列から複数の参照配列断片を生成することと、

b. 参照配列断片の数の関数として少なくとも 1 つのパラメータの分布を生成することと、

c. 前記分布を少なくとも 5 つのビンに分割することと、

d. 前記少なくとも 5 つのビンのうちの少なくとも 3 つから、少なくとも 1 つの参照配列断片を選択することと、

e. 前記少なくとも 3 つの参照配列をシャッフルして、シャッフルされた配列を生成することと、

f. 前記シャッフルされた配列を含むオリゴヌクレオチドを合成することと、を含み、それにより複数の正規化対照オリゴヌクレオチドを生成する、方法。

〔 9 2 〕 複数の正規化対照オリゴヌクレオチドを製造する方法であって、

10

20

30

40

50

a. スライディングウィンドウを使用して、少なくとも1つの参照配列から複数の参照配列断片を生成することと、

b. 参照配列断片の数の関数として少なくとも1つのパラメータの分布を生成することと、

c. 前記分布を少なくとも5つのビンに分割することと、

d. 前記少なくとも5つのビンのうちの少なくとも3つから、少なくとも2つの参照配列断片を選択することであって、前記少なくとも2つの参照配列断片が、前記参照配列中で連続していないか、または異なる参照配列からのものであるかのいずれかである、選択することと、

e. 前記少なくとも3つのビンのそれぞれからの前記少なくとも2つの参照配列断片を連結させることと、

f. 前記連結された参照配列断片を含むオリゴヌクレオチドを合成することと、を含み、それにより複数の正規化対照オリゴヌクレオチドを生成する、方法。

〔93〕複数の正規化対照ポリヌクレオチド配列をデザインするためのシステムであって、

a. コンピュータ実行可能命令を記憶するコンピュータ可読記憶媒体であって、前記命令が、

i. 少なくとも1つの参照配列をインポートするための命令と、

ii. 少なくとも1つの参照配列から複数の参照配列断片を生成するための命令と、

iii. 配列断片の数の関数として少なくとも1つのパラメータの分布を生成するための命令と、

iv. 前記分布をビンに分割するための命令と、

v. 少なくとも前記ビンのサブセットから複数の参照配列断片を選択するための命令と、

vi. 前記複数の参照配列断片をシャッフルして、シャッフルされた配列を生成するための命令と、を含み、

それにより複数の正規化ポリヌクレオチド配列を生成する、コンピュータ可読記憶媒体と、

b. 以下の工程を実行するように構成されたプロセッサであって、前記工程が、

i. 前記少なくとも1つの参照配列を含むインプットファイルのセットを受信することと、

ii. 前記コンピュータ可読記憶媒体に記憶された前記コンピュータ実行可能命令を実行することと、を含む、プロセッサと、を備える、システム。

〔94〕複数の正規化対照ポリヌクレオチド配列をデザインするためのシステムであって、

a. コンピュータ実行可能命令を記憶するコンピュータ可読記憶媒体であって、前記命令が、

i. 少なくとも1つの参照配列をインポートするための命令と、

ii. 少なくとも1つの参照配列から複数の参照配列断片を生成するための命令と、

iii. 参照配列断片の数の関数として少なくとも1つのパラメータの分布を生成するための命令と、

iv. 前記分布をビンに分割するための命令と、

v. 少なくとも前記ビンのサブセットのそれぞれから少なくとも2つの参照配列断片を選択するための命令であって、前記少なくとも2つの参照配列断片が、前記参照配列中で連続していないか、または異なる参照配列からのものであるかのいずれかである、命令と、

vi. 各ビンからの前記少なくとも2つの参照配列断片を連結するための命令と、を含み、

それにより複数の正規化ポリヌクレオチド配列を生成する、コンピュータ可読記憶媒体と、

b. 以下の工程を実行するように構成されたプロセッサであって、前記工程が、

i. 前記少なくとも1つの参照配列を含むインプットファイルのセットを受信することと、

ii. 前記コンピュータ可読記憶媒体に記憶された前記コンピュータ実行可能命令を実行

10

20

30

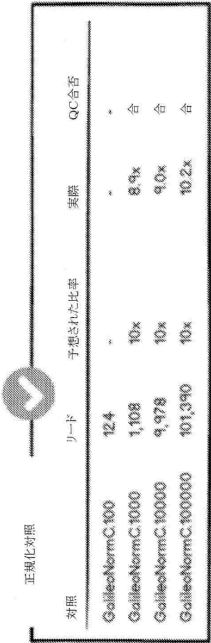
40

50

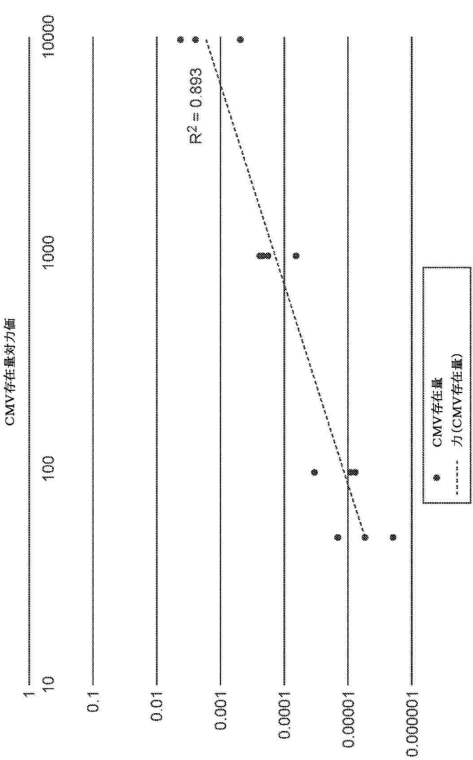
することと、を含む、プロセッサと、を備える、システム。

【図面】

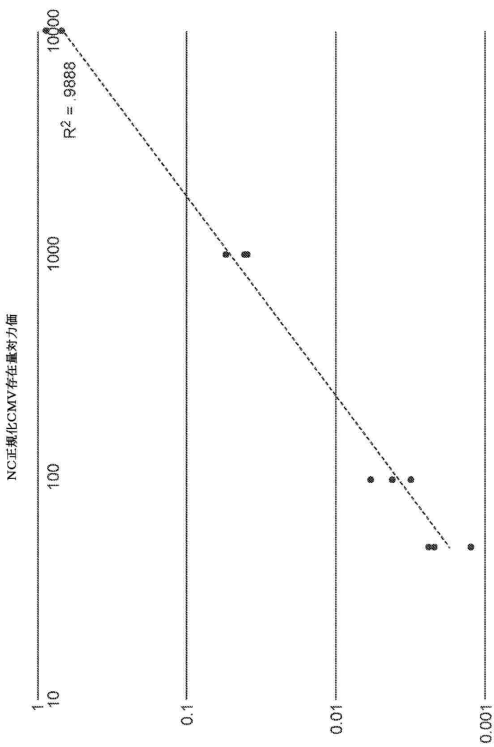
【図 1】



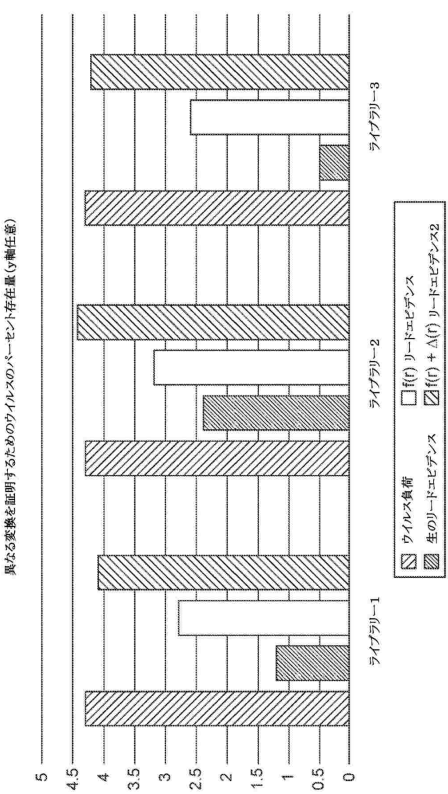
【図 2】



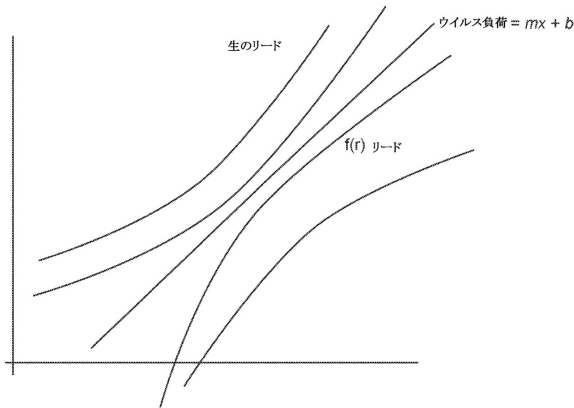
【図 3】



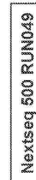
【図 4】



【図 5】



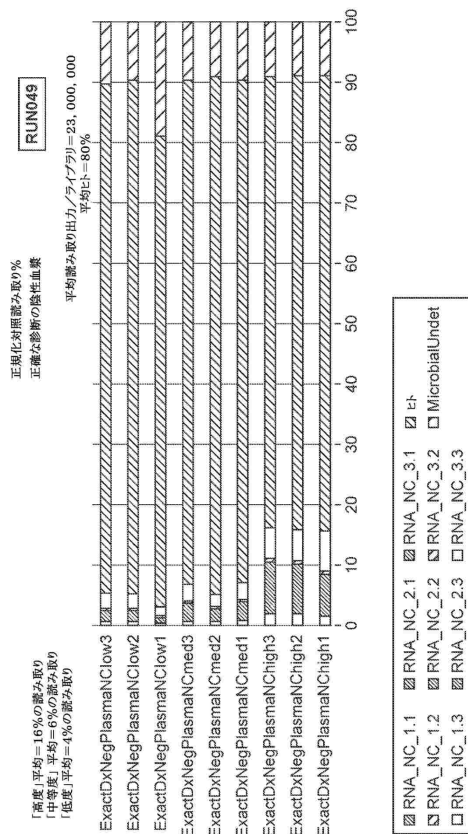
【図 6】



抽出中の正規化対照の添加

- 0.1mg/ μ L に対して細胞の定着物の各セプト20,000倍希釈
- 正規化処理混合液、
 - RNA正規化対照1： 25 μ L
 - RNA正規化対照2： 50 μ L
 - RNA正規化対照3： 100 μ L
 - 175 μ L
- 少なくとも10倍希釈：30 μ L以上の上記のプール+270 μ LのAVE緩衝液を添加した/飲料(NC1:1.5)
 - EZ1は3列目のチューブに1.5mLチューブ：8 μ Lの10mg/ μ L NC濃度
 - EZ1は、3列目のチューブから50 μ L L7アプコートを取り、400 μ L
 - 「高濃度」=98.10mg/NCが添加された/飲料(NC1:1.5)
 - 「中濃度」=19.04mg/NCが添加された/飲料(NC1:7.5)
 - 「低濃度」=9.52mg/NCが添加された/飲料(NC1:15)
- 「低濃度」の量は、4%の錠量取りをならした、RUNの「低濃度」(NC1 (1.4mg)のライブラリー)の変換 (重複排除)
- 「低濃度」(NC1 (1.4mg)のライブラリー)の変換 (重複排除)

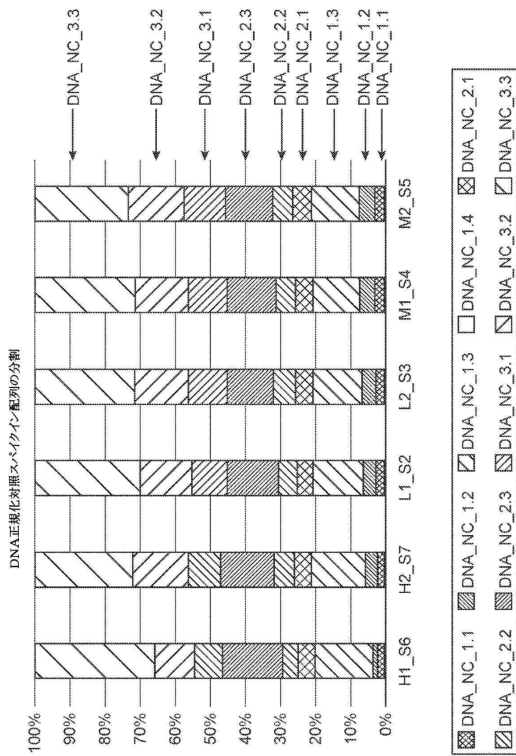
【圖 7】



【図 8】

Aグループ-混合									反応	2		
	供給業者	PIN#	MW	[開始液体] または グラム固体	UOM	RXN 1倍体積	UOM	生成相模 RXN	最終濃度 1倍RXN	UOM		
	Sigma	793566-25KG		1	x	8	uL	16	0.4	uL		
	Mo Bio lalae			200	uM	5	uL	10	50	uM		
	NEB 2			10	x	2	uL	4	1	uL		
					合計	20	uL	40				
20℃までゆーくひと1分間95からランプする												
15mMに希釈する												
正屈化対照								反応	2			
	供給業者	PIN#	MW	[開始液体] または グラム固体	UOM	RXN 1倍体積	UOM	生成相模 RXN	最終濃度 1倍RXN	UOM		
	Sigma	793566-25KG		1	x	42	uL	84	0.7	uL		
	Mo Bio lalae			50	uM	18	uL	36	15	uM		
					合計	60	uL	120				
15mMに希釈する												
正屈化対照								反応	2			
	供給業者	PIN#	MW	[開始液体] または グラム固体	UOM	RXN 1倍体積	UOM	生成相模 RXN	最終濃度 1倍RXN	UOM		
	Sigma	793566-25KG		1	x	1140	uL	2280	19	uL		
	Mo Bio lalae			15	uM	60	uL	120	0.75	uM		
					合計	1200	uL	2400				

【図 9】



【図 10】

Fasta ヘッダ	GC 含量	合計カウント	長さ
>Oligo177_RNA_正酸化_対照_1_および_TT_RNA_ポリマーゼ_プロモーター_IVT用	45.41284404	218	436
>Oligo178_RNA_正酸化_対照_2_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	46.78898083	218	436
>Oligo179_RNA_正酸化_対照_3_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	43.57798165	218	436
>Oligo180_RNA_正酸化_対照_4_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	46.73913043	276	552
>Oligo181_RNA_正酸化_対照_5_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	43.84057971	276	552
>Oligo182_RNA_正酸化_対照_6_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	40.57971014	276	552
>Oligo183_RNA_正酸化_対照_7_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	44.35318275	487	974
>Oligo184_RNA_正酸化_対照_8_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	43.32648871	487	974
>Oligo185_RNA_正酸化_対照_9_with_TT_RNA_ポリマーゼ_プロモーター_IVT用	43.53182752	487	974
>Oligo186_RNA_正酸化_対照_1	49.77777778	225	450
>Oligo187_RNA_正酸化_対照_2	52	225	450
>Oligo188_RNA_正酸化_対照_3	49.77777778	225	450
>Oligo189_RNA_正酸化_対照_4	42.15686275	306	612
>Oligo190_RNA_正酸化_対照_5	43.48405229	306	612
>Oligo191_RNA_正酸化_対照_6	41.50326797	306	612
>Oligo192_RNA_正酸化_対照_7	45.64796905	517	1034
>Oligo193_RNA_正酸化_対照_8	46.61508704	517	1034
>Oligo194_RNA_正酸化_対照_9	44.87427466	517	1034
>Oligo241_manorm1seg1	47.05882353	255	510
>Oligo242_manorm2seg1	48.62745098	255	510
>Oligo243_manorm3seg1	53.7254902	255	510
>Oligo244_manorm1seg2	51.7571885	313	626
>Oligo245_manorm2seg2	51.7571885	313	626
>Oligo246_manorm3seg2	44.7284345	313	626
>Oligo247_manorm2seg3	50.57251908	524	1048
>Oligo248_manorm3seg3	49.42748092	524	1048
>Oligo249_manorm4seg3	48.85496183	524	1048
>Oligo250_dnanorm1seg1	46.66666667	225	450
>Oligo251_dnanorm2seg1	55.55555556	225	450
>Oligo252_dnanorm3seg1	54.22222222	225	450
>Oligo253_dnanorm1seg2	49.34640523	306	612
>Oligo254_dnanorm2seg2	50.65359477	306	612
>Oligo255_dnanorm3seg2	49.01960784	306	612
>Oligo256_dnanorm1seg3	49.9032882	517	1034
>Oligo257_dnanorm2seg3	48.54932302	517	1034
>Oligo258_dnanorm3seg3	48.54932302	517	1034

【図 11】

オリゴ番号	最終濃度 ng/ul	DNA の総量に対する比率
Oligo186	4.20E-03	0.420
Oligo187	4.21E-04	0.042
Oligo188	2.11E-04	0.021
Oligo189	2.86E-03	0.286
Oligo190	4.29E-04	0.043
Oligo191	1.43E-04	0.014
Oligo192	9.68E-04	0.097
Oligo193	4.84E-04	0.048
Oligo194	9.68E-05	0.010
Oligo250	4.22E-05	0.004
Oligo251	2.11E-05	0.002
Oligo253	4.31E-05	0.004
Oligo254	1.43E-05	0.001
Oligo256	4.84E-05	0.005
Oligo257	9.71E-06	0.001
ライブラリサイズ		

【図 12】

群	長さ (kb)	長さ (bp)	分子の数	合計 アトムール/断片	濃度 (アトムール/μL)
RNA_NC_1.1 A	0.175	175	263158	0.44	0.07
RNA_NC_1.2 B	0.25	250	1144737	1.90	0.29
RNA_NC_1.3 C	0.45	450	5131579	8.52	1.28
RNA_NC_2.1 A	0.175	175	2785474	4.63	0.69
RNA_NC_2.2 B	0.25	250	4513157	7.49	1.12
RNA_NC_2.3 C	0.45	450	44473884	73.84	11.08
RNA_NC_3.1 A	0.175	175	4605263	7.65	1.15
RNA_NC_3.2 B	0.25	250	17500000	29.06	4.36
RNA_NC_3.3 C	0.45	450	77894737	129.33	19.40

10

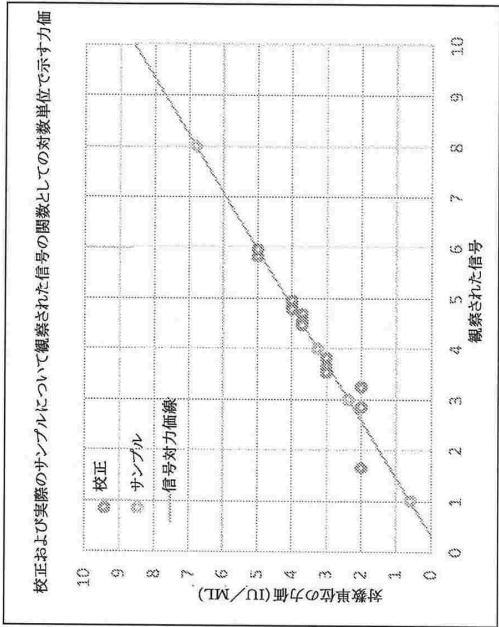
20

30

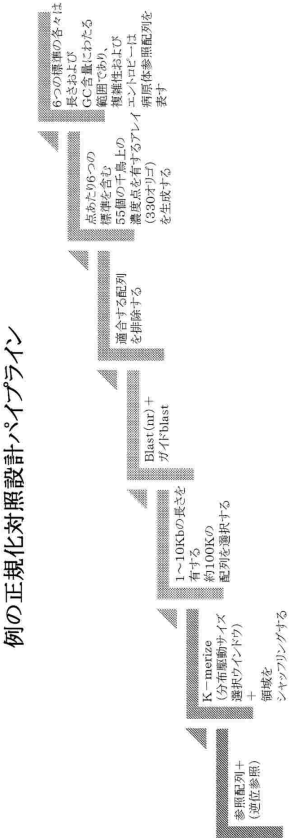
40

50

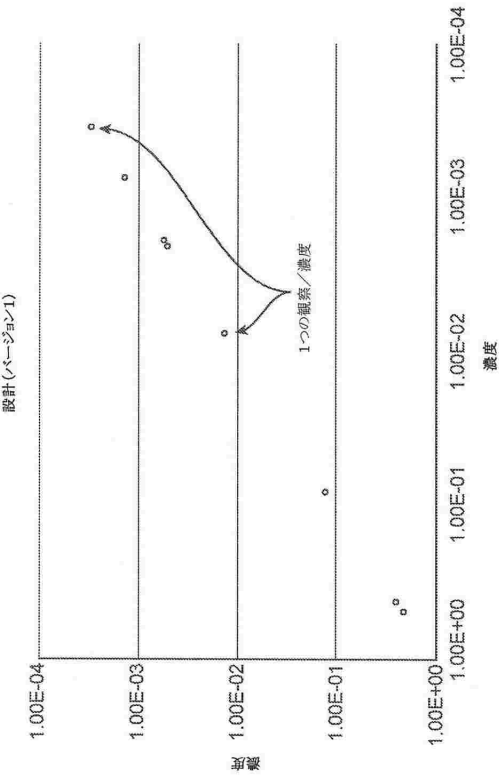
【図 1 3】



【図 1 4 A】



【図 1 4 B】



【図 1 4 C】

名称	ライブラリ	観測入力	最低	観測出力
Oligo298.dne	298	4.88E-01	なし	4.88E-01
Oligo299.dne	299	4.19E-01	なし	4.19E-01
Oligo301.dne	301	8.13E-02	なし	8.13E-02
Oligo302.dne	302	2.04E-03	なし	2.04E-03
Oligo303.dne	303	7.63E-03	なし	7.63E-03
Oligo304.dne	304	1.87E-03	なし	1.87E-03
Oligo305.dne	305	7.39E-04	最低	7.39E-04
Oligo306.dne	306	3.45E-04	最低	3.45E-04

10

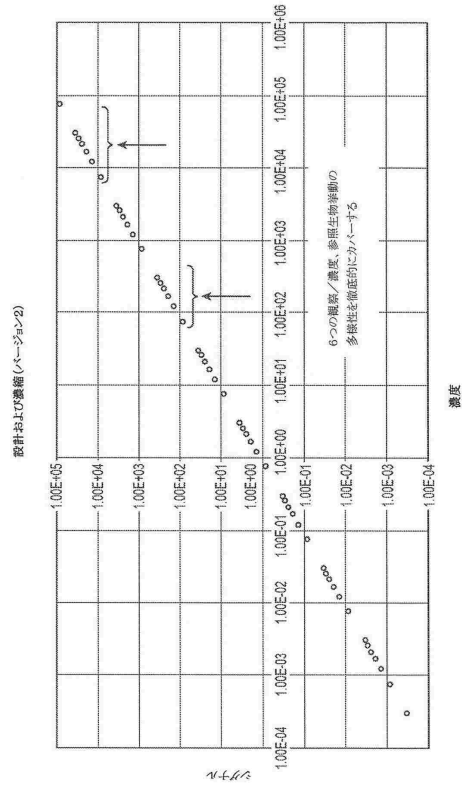
20

30

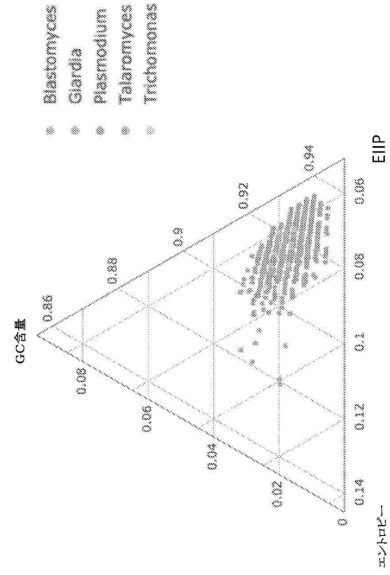
40

50

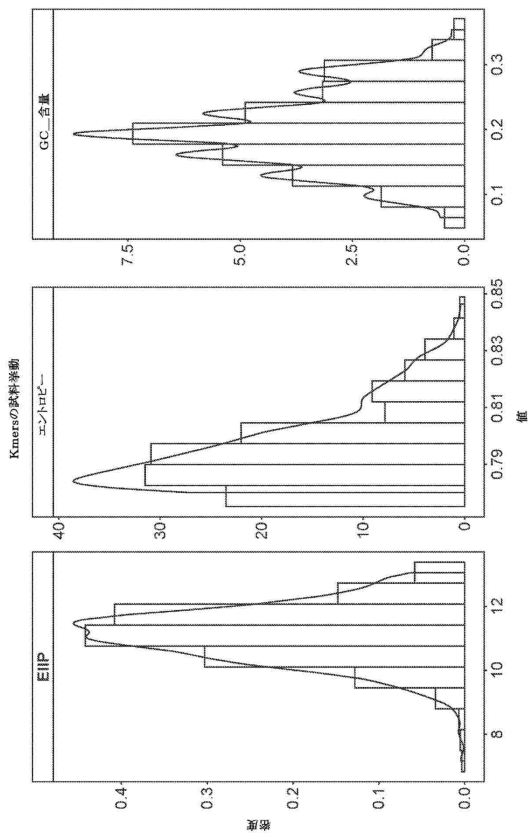
【図 1 4 D】



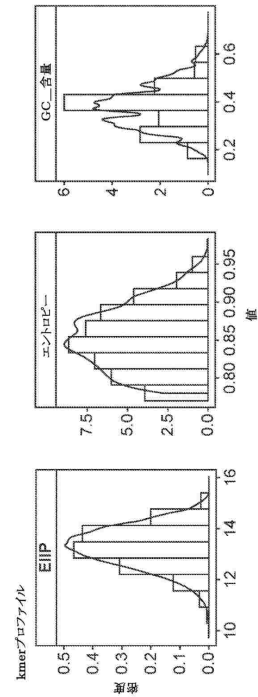
【図 1 5】



【図 1 6】



【図 1 7 A】



10

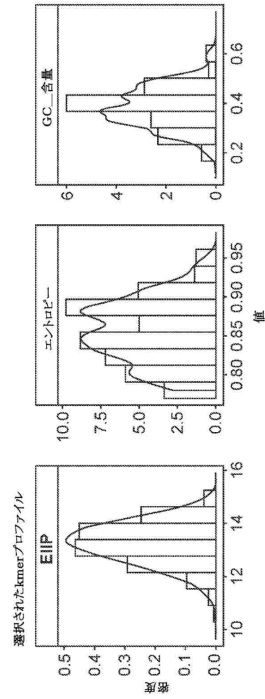
20

30

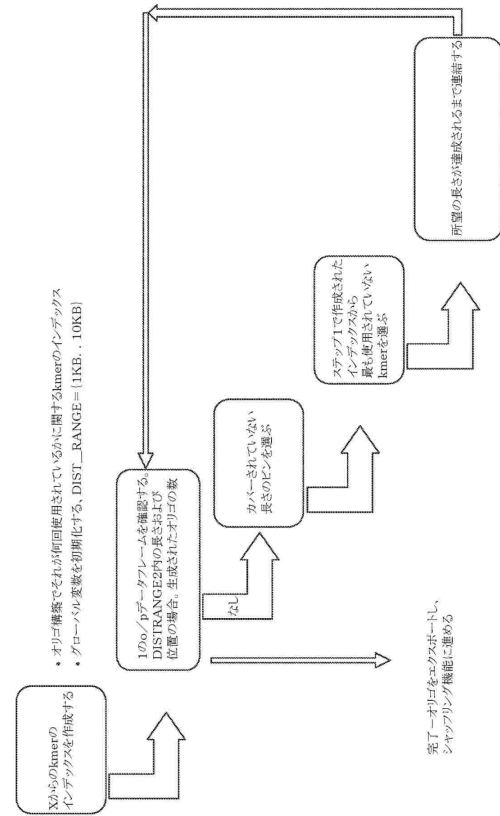
40

50

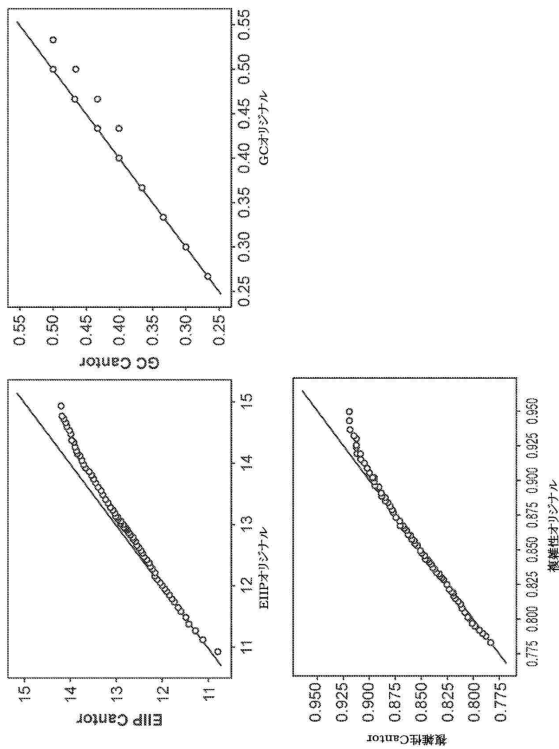
【図 17 B】



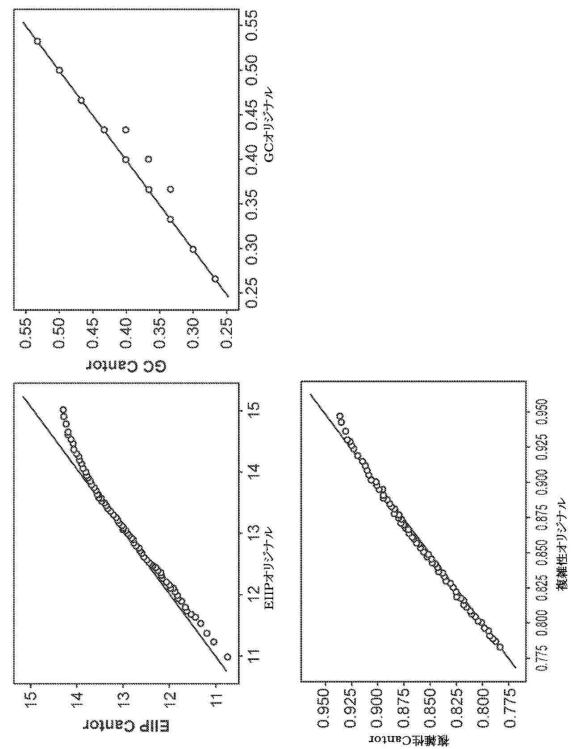
【図 18】



【図 19 A】



【図 19 B】



10

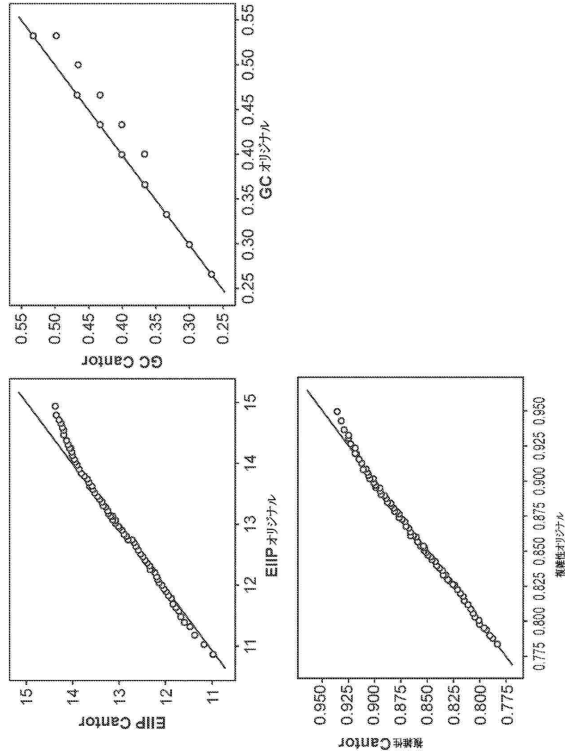
20

30

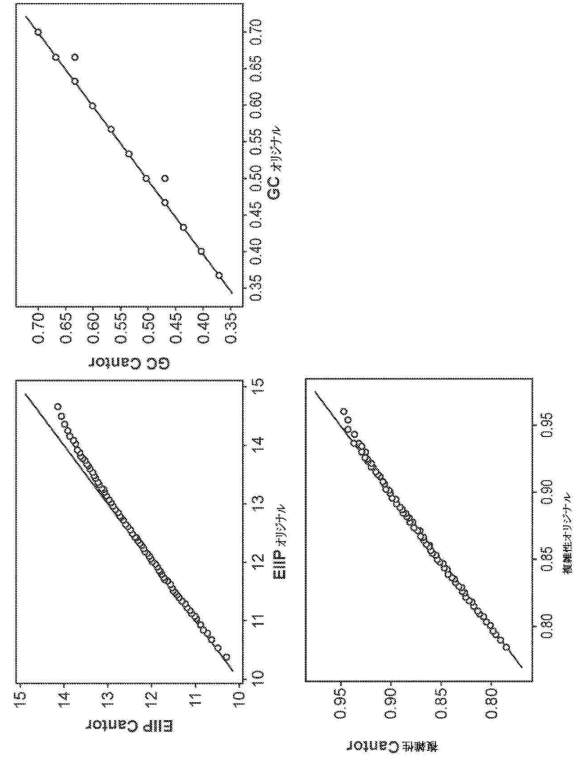
40

50

【図 19 C】



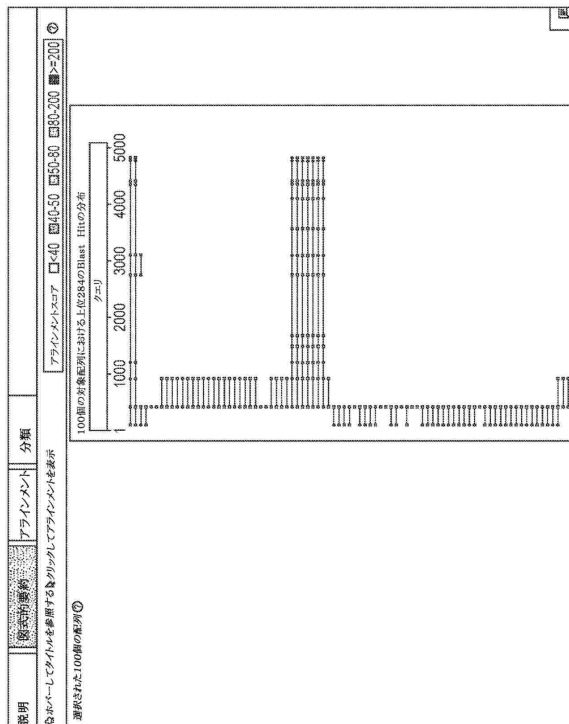
【図 19 D】



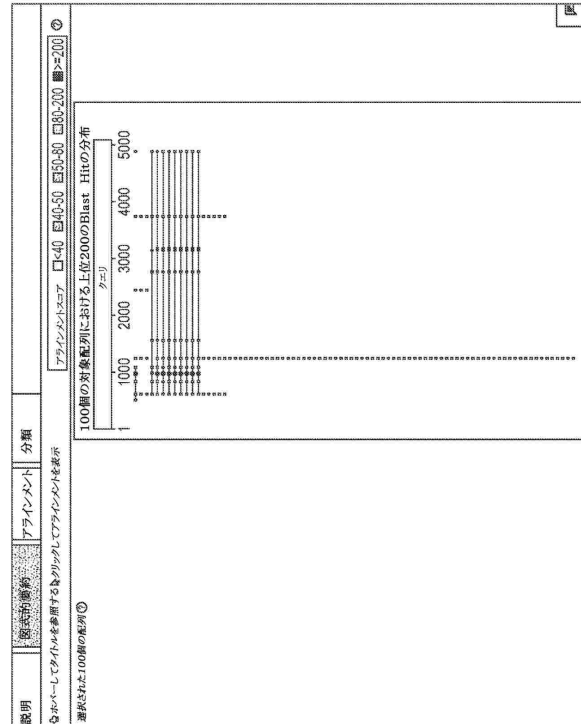
10

20

【図 20 A】



【図 20 B】

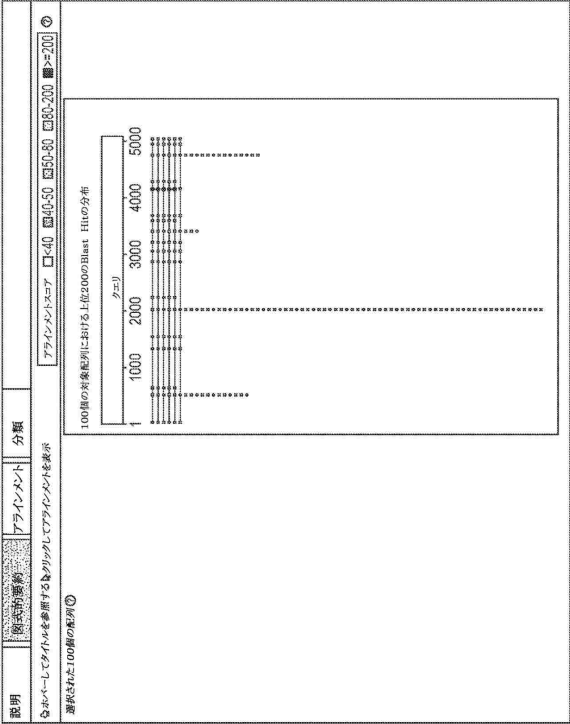


30

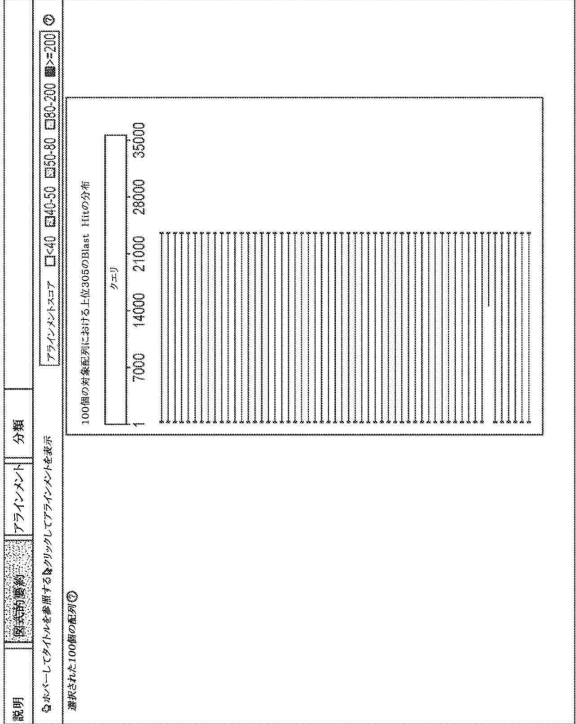
40

50

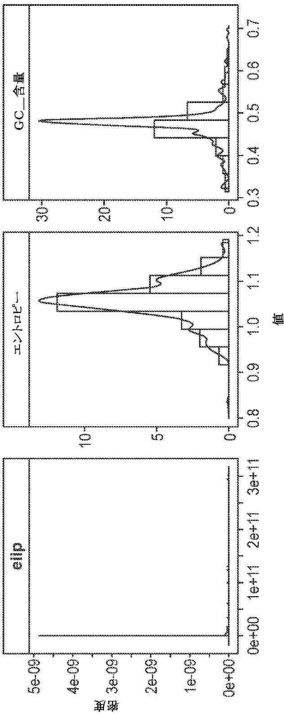
【図 20 C】



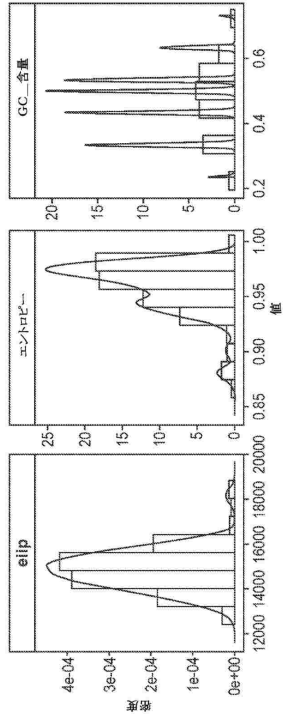
【図 20 D】



【図 21 A】



【図 21 B】



10

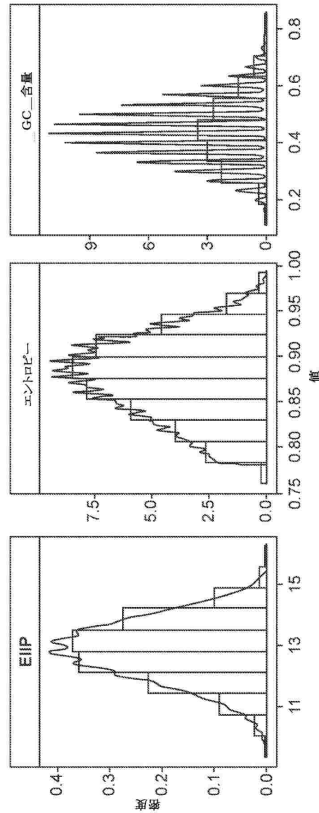
20

30

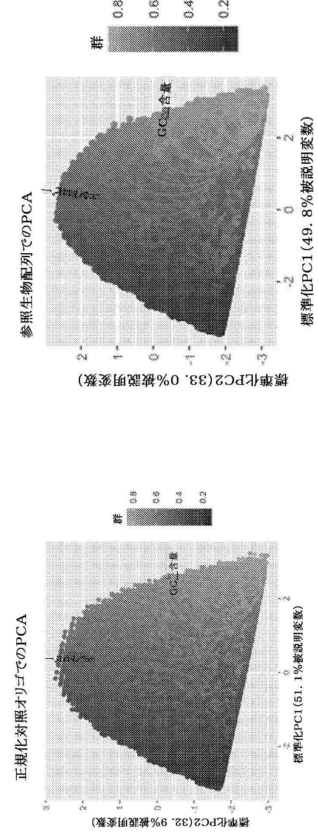
40

50

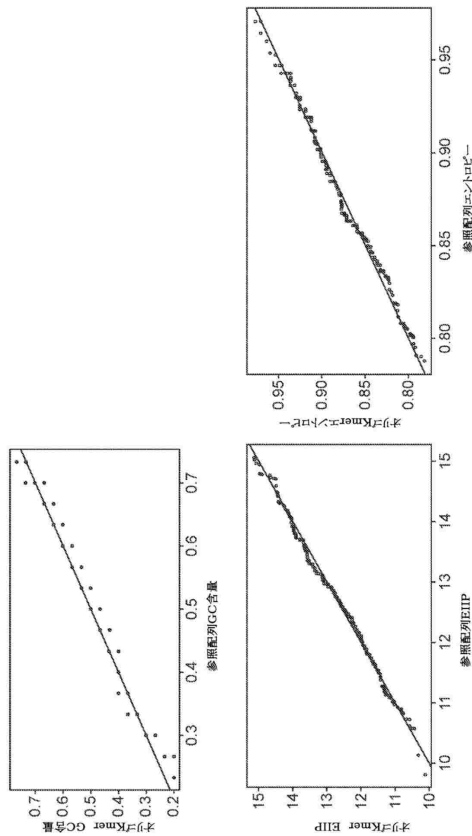
【図 2 1 C】



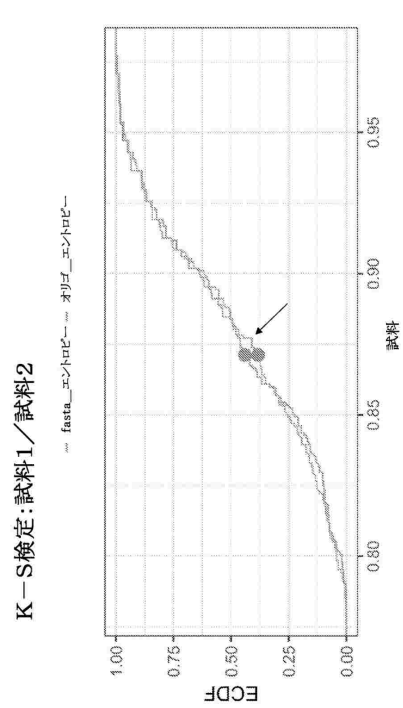
【図 2 2】



【図 2 3】



【図 2 4】



K-S検定: 試料1/試料2

10

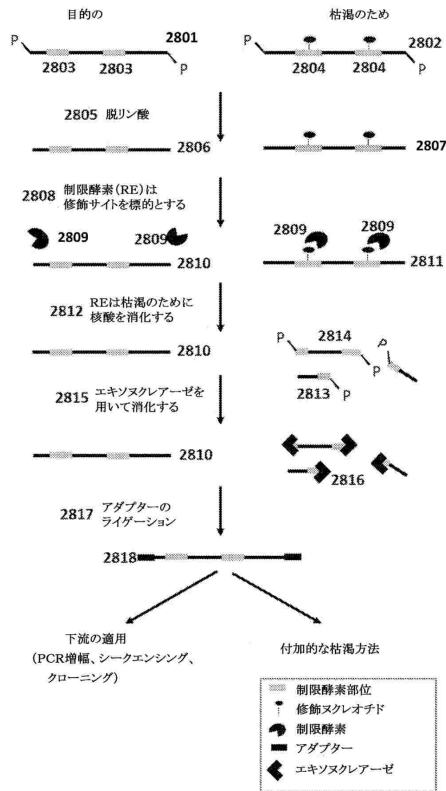
20

30

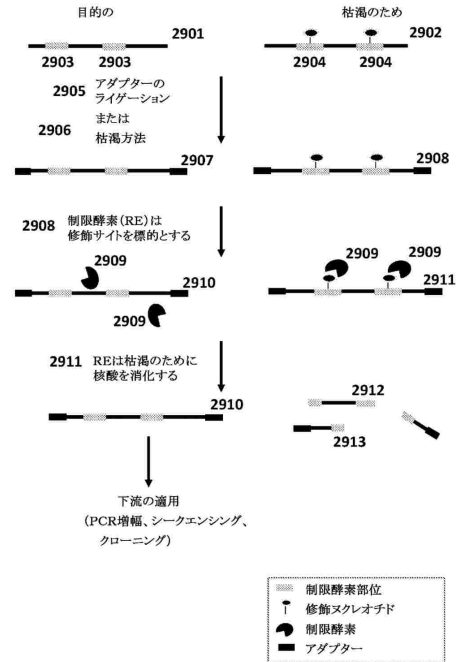
40

50

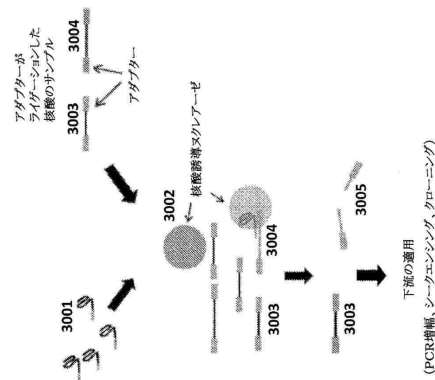
【図 28】



【図 29】



【図 30】



【配列表】

0007436493000001.app

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

弁理士 市川 さつき

(74)代理人 100111796

弁理士 服部 博信

(74)代理人 100136249

弁理士 星野 貴光

(72)発明者 ハーネス エリック

アメリカ合衆国 カリフォルニア州 94086 サニーベール タマラック レーン 920 アパートメント 7

(72)発明者 ナゲシュ ヴァイシュナヴィ

アメリカ合衆国 マサチューセッツ州 02141 ケンブリッジ カナル パーク 10 スイート 201 エイアールシー バイオ リミテッド ライアビリティ カンパニー内

(72)発明者 ベントレー ロイド ゴードン

アメリカ合衆国 マサチューセッツ州 02141 ケンブリッジ カナル パーク 10 スイート 201 エイアールシー バイオ リミテッド ライアビリティ カンパニー内

(72)発明者 カーペンター メレディス エル

アメリカ合衆国 カリフォルニア州 94401 サンマテオ テンス アベニュー 170

審査官 福間 信子

(56)参考文献 国際公開第2017/165864 (WO, A1)

国際公開第2016/001736 (WO, A1)

国際公開第2015/073080 (WO, A1)

国際公開第2017/024138 (WO, A1)

(58)調査した分野 (Int.Cl., DB名)

C12Q

C12N

JSTPlus/JMEDPlus/JST7580 (JDreamIII)

Caplus/MEDLINE/EMBASE/BIOSIS (STN)