

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6743268号  
(P6743268)

(45) 発行日 令和2年8月19日(2020.8.19)

(24) 登録日 令和2年7月31日(2020.7.31)

(51) Int. Cl.			F I		
<b>C 1 2 Q</b>	<b>1/6809</b>	<b>(2018.01)</b>	C 1 2 Q	1/6809	Z N A Z
<b>C 1 2 Q</b>	<b>1/6888</b>	<b>(2018.01)</b>	C 1 2 Q	1/6888	Z
<b>C 4 O B</b>	<b>40/06</b>	<b>(2006.01)</b>	C 4 O B	40/06	
<b>C 1 2 N</b>	<b>15/11</b>	<b>(2006.01)</b>	C 1 2 N	15/11	Z

請求項の数 30 (全 118 頁)

(21) 出願番号	特願2019-500764 (P2019-500764)	(73) 特許権者	518335735
(86) (22) 出願日	平成29年3月24日 (2017.3.24)		カリウス・インコーポレイテッド
(65) 公表番号	特表2019-509068 (P2019-509068A)		アメリカ合衆国、カリフォルニア・940
(43) 公表日	平成31年4月4日 (2019.4.4)		25、メンロー・パーク、アダムス・ドラ
(86) 国際出願番号	PCT/US2017/024176		イブ・1505-エー
(87) 国際公開番号	W02017/165864	(74) 代理人	100114188
(87) 国際公開日	平成29年9月28日 (2017.9.28)		弁理士 小野 誠
審査請求日	令和1年10月30日 (2019.10.30)	(74) 代理人	100119253
(31) 優先権主張番号	62/397,873		弁理士 金山 賢教
(32) 優先日	平成28年9月21日 (2016.9.21)	(74) 代理人	100124855
(33) 優先権主張国・地域又は機関	米国 (US)		弁理士 坪倉 道明
(31) 優先権主張番号	62/451,363	(74) 代理人	100129713
(32) 優先日	平成29年1月27日 (2017.1.27)		弁理士 重森 一輝
(33) 優先権主張国・地域又は機関	米国 (US)	(74) 代理人	100137213
			弁理士 安藤 健司

最終頁に続く

(54) 【発明の名称】 合成核酸スパイクイン

(57) 【特許請求の範囲】

【請求項1】

(a) 少なくとも1,000個のユニーク合成核酸の出発量を初期サンプルに加え、ここで、前記の少なくとも1,000個のユニーク合成核酸のそれぞれは(i) 識別タグ、および(ii) 少なくとも5縮重塩基を含む可変領域を含み、

(b) 標的核酸の一部および前記の少なくとも1,000個のユニーク合成核酸の一部に関して配列決定アッセイを行い、それにより、標的および合成核酸配列リードを得、

(c) 前記の少なくとも1,000個のユニーク合成核酸の部分に関する多様性減少値を計算することを含み、

ここで、初期サンプルにおける標的核酸の存在量は、前記多様性減少値を用いて決定される、標的核酸を含む初期サンプルにおける核酸の存在量を決定するための方法。

【請求項2】

標的核酸が病原体核酸を含む、請求項1記載の方法。

【請求項3】

標的核酸が、少なくとも5つの異なる病原体からの病原体核酸を含む、請求項2に記載の方法。

【請求項4】

前記の少なくとも1,000個のユニーク合成核酸がDNAを含む、請求項1に記載の方法。

## 【請求項 5】

前記の少なくとも 1, 000 個のユニーク合成核酸のそれぞれが 500 塩基対またはヌクレオチド長未満である、請求項 1 に記載の方法。

## 【請求項 6】

初期サンプルが血液、血漿、血清、脳脊髄液、滑液、気管支肺胞洗浄液、尿、便、唾液または鼻サンプルである、請求項 1 に記載の方法。

## 【請求項 7】

初期サンプルが、単離された核酸のサンプルである、請求項 1 に記載の方法。

## 【請求項 8】

初期サンプルから配列決定ライブラリーを製造することを更に含み、ここで、配列決定ライブラリーを製造する前に前記の少なくとも 1, 000 個のユニーク合成核酸をサンプルに加える、請求項 1 に記載の方法。

10

## 【請求項 9】

前記の少なくとも 1, 000 個のユニーク合成核酸の部分に関する多様性減少値が初期サンプルのサンプル処理中の 1 以上の核酸の減少を示す、請求項 1 に記載の方法。

## 【請求項 10】

前記の少なくとも 1, 000 個のユニーク合成核酸のそれぞれの識別タグが、共通配列を含む、請求項 1 に記載の方法。

## 【請求項 11】

初期サンプルがヒト対象由来である、請求項 1 に記載の方法。

20

## 【請求項 12】

前記の少なくとも 1, 000 個のユニーク合成核酸が少なくとも  $10^4$  個のユニーク合成核酸を含む、請求項 1 に記載の方法。

## 【請求項 13】

初期サンプルに追加的合成核酸を加えることを更に含み、該追加的合成核酸が少なくとも 3 つの異なる長さを有する、請求項 1 に記載の方法。

## 【請求項 14】

第 1 の長さを有する追加的合成核酸の第 1 群、第 2 の長さを有する追加的合成核酸の第 2 群、および第 3 の長さを有する追加的合成核酸の第 3 群を初期サンプルに加えることを更に含み、ここで、追加的合成核酸の第 1 群、追加的合成核酸の第 2 群および追加的合成核酸の第 3 群のそれぞれは、少なくとも 3 つの異なる GC 含量を有する合成核酸を含む、請求項 1 に記載の方法。

30

## 【請求項 15】

該追加的合成核酸を使用して、初期サンプルにおける標的核酸の存在量を計算することを更に含み、請求項 13 に記載の方法。

## 【請求項 16】

該追加的合成核酸を使用して、該追加的合成核酸の長さ、GC 含量、または長さおよび GC 含量の両方に基づいて初期サンプルにおける標的核酸の存在量を計算することを更に含み、請求項 13 に記載の方法。

## 【請求項 17】

第 1 サンプル処理工程において、前記の少なくとも 1, 000 個のユニーク合成核酸を初期サンプルに加える、請求項 1 に記載の方法。

40

## 【請求項 18】

第 2 サンプル処理工程において、少なくとも 1, 000 個のユニーク合成核酸の追加的プールを初期サンプルに加えることを更に含み、ここで、第 2 サンプル処理工程は第 1 サンプル処理工程とは異なる、請求項 17 に記載の方法。

## 【請求項 19】

少なくとも 1, 000 個のユニーク合成核酸の追加的プールに関する多様性減少値を計算することを更に含み、請求項 18 に記載の方法。

## 【請求項 20】

50

前記の少なくとも1,000個のユニーク合成核酸に関する多様性減少値を少なくとも1,000個のユニーク合成核酸の追加的プールに関する多様性減少値と比較することにより、比較的高い多様性減少を示すサンプル処理工程を特定することを更に含む、請求項18記載の方法。

【請求項21】

少なくとも1,000個のユニーク合成核酸の追加的プールにおけるユニーク合成核酸のそれぞれが、少なくとも1,000個のユニーク合成核酸の追加的プールのメンバーとして該合成核酸を特定するドメインを含む、請求項18記載の方法。

【請求項22】

サンプル識別核酸を初期サンプルに加えることを更に含む、請求項1記載の方法。

10

【請求項23】

(a)が更に、非ユニーク合成核酸を初期サンプルに加えることを含む、請求項1記載の方法。

【請求項24】

方法の結果を、介護人、患者またはその他の人に報告することをさらに含む、請求項1に記載の方法。

【請求項25】

決定される存在量が相対的存在量である、請求項1記載の方法。

【請求項26】

決定される存在量が絶対的存在量である、請求項1記載の方法。

20

【請求項27】

異なる可変領域を含む合成核酸配列リードの数を測定することを更に含む、請求項1に記載の方法。

【請求項28】

前記の少なくとも1000個のユニーク合成核酸の出発量を、異なる可変領域を含む合成核酸配列リードの数と比較することによって前記多様性減少値を計算する、請求項27に記載の方法。

【請求項29】

異なる可変領域を含む異なる合成核酸配列リードを参照配列とアライメントさせることで、該合成核酸配列リードの数を決定する、請求項27に記載の方法。

30

【請求項30】

異なる可変領域を含む異なる合成核酸配列リードを互いにアライメントさせ、重複配列リードを除外することにより、該合成核酸配列リードの数を定量する、請求項28に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は合成核酸スパイクイン(spike-in)に関する。

【0002】

相互参照

40

本出願は2016年3月25日付け出願の米国仮特許出願第62/313,668号、2016年9月21日付け出願の米国仮特許出願第62/397,873号および2017年1月27日付け出願の米国仮特許出願第62/451,363号(それらの全体を参照により本明細書に組み入れることとする)の利益を主張するものである。

【背景技術】

【0003】

次世代シーケンシングは、サンプルの遺伝的内容に関する大量のデータを収集するために用いられる。それは臨床サンプルのような複雑なサンプルにおける核酸の分析および全ゲノムの配列決定に特に有用でありうる。しかし、核酸、特に低含量の核酸または患者のサンプルにおける核酸を検出および定量するための、より効率的かつ高精度の方法が当

50

技術分野で必要とされている。

【発明の概要】

【0004】

概要

スパイクイン合成核酸を使用する、次世代シーケンシングアッセイおよび他のアッセイにおける核酸の、改良された特定または定量のための方法および組成物を、本発明において提供する。幾つかの場合には、スパイクイン合成核酸は特定の配列、長さ、GC含量、縮重度、多様性の度合および/または既知の出発濃度のような特別な特徴を有する。本発明で提供する方法は、血漿のような臨床サンプルにおける病原体核酸の検出に特に有用であるが、他のタイプの標的を検出するためにも使用されうる。

10

【0005】

1つの態様においては、標的核酸を含む初期サンプルにおける核酸の存在量を決定するための方法を本発明で提供し、該方法は、(a)少なくとも1,000個の合成核酸の出発量をサンプルに加え(添加し)、ここで、前記の少なくとも1,000個の合成核酸のそれぞれはユニーク(unique;一意)可変領域を含み、(b)該サンプルにおける標的核酸の一部および前記の少なくとも1,000個の合成核酸の一部に関して配列決定アッセイを行い、それにより、標的および合成核酸配列リード(read;読取り)を得、ここで、該合成核酸配列リードはユニーク可変領域配列を含み、(c)(i)該合成核酸配列リード内の異なる可変領域配列の数を定量して、ユニーク配列決定値を得、(ii)前記の少なくとも1,000個の合成核酸の出発量を前記のユニーク配列決定値と比較して、前記の少なくとも1,000個の合成核酸の多様性減少を得ることにより、前記の少なくとも1,000個の合成核酸の多様性減少を検出し、(d)前記の少なくとも1,000個の合成核酸の多様性減少を用いて、初期サンプルにおける標的核酸の存在量を計算することを含む。幾つかの場合には、比較する出発量は出発濃度である。

20

【0006】

幾つかの実施形態においては、標的核酸は病原体核酸を含む。幾つかの場合には、標的核酸は、少なくとも5つの異なる病原体からの病原体核酸を含む。幾つかの場合には、標的核酸は、少なくとも2つの異なる病原体からの病原体核酸を含む。幾つかの場合には、標的核酸は、少なくとも10種の異なる病原体からの病原体核酸を含む。

【0007】

幾つかの場合には、前記の少なくとも1,000個の合成核酸はDNAを含む。幾つかの場合には、前記の少なくとも1,000個の合成核酸はRNA、ssRNA、dsDNA、ssDNAまたはそれらの幾つかの組合せを含む。幾つかの場合には、前記の少なくとも1,000個の合成核酸のそれぞれは500塩基対またはヌクレオチド長未満である。幾つかの場合には、前記の少なくとも1,000個の合成核酸のそれぞれは200塩基対またはヌクレオチド長未満である。幾つかの場合には、前記の少なくとも1,000個の合成核酸のそれぞれは100塩基対またはヌクレオチド長未満である。幾つかの場合において、サンプルは血液、血漿、血清、脳脊髄液、滑液、気管支肺胞洗浄液、尿、便、唾液または鼻サンプルである。幾つかの場合には、サンプルはヒト対象由来である。幾つかの場合には、サンプルは、単離された核酸のサンプルである。

30

40

【0008】

幾つかの場合には、該方法は更に、該サンプルから配列決定ライブラリーを製造(作製)することを含み、ここで、配列決定ライブラリーを製造する前に前記の少なくとも1,000個の合成核酸をサンプルに加える。幾つかの場合には、前記の少なくとも1,000個の合成核酸の多様性減少はサンプルのサンプル処理中の1以上の核酸の減少を示す。

【0009】

幾つかの場合には、前記の少なくとも1,000個の合成核酸のそれぞれは識別タグ配列を含む。幾つかの場合には、ユニーク可変領域配列の数の定量は、該タグ配列を含有する配列を検出することを含む。幾つかの場合には、第1配列リード内の少なくとも1,000個のユニーク配列の定量は、第1配列リード内のユニーク配列のリード数(リードカ

50

ウント)を決定することを含む。幾つかの場合には、少なくとも1,000個のユニーク合成核酸は少なくとも $10^4$ 個のユニーク合成核酸を含む。

【0010】

幾つかの場合には、該方法は更に、第1の長さを有する第1の追加的合成核酸群、第2の長さを有する第2の追加的合成核酸群、および第3の長さを有する第3の追加的合成核酸群を加えることを含み、ここで、第1、第2および第3の追加的合成核酸群のそれぞれは、少なくとも3つの異なるGC含量を有する合成核酸を含む。幾つかの場合には、該方法は更に、該追加的合成核酸を使用して、サンプルにおける標的核酸の絶対的存在量値を計算することを含む。幾つかの場合には、該方法は更に、該追加的合成核酸を使用して、該追加的合成核酸の長さ、GC含量または長さおよびGC含量の両方に基づいてサンプルにおける標的核酸の絶対的または相対的存在量を計算することを含む。

10

【0011】

幾つかの場合には、第1サンプル処理工程において、前記の少なくとも1,000個の合成核酸をサンプルに加える。幾つかの場合には、該方法は更に、第2サンプル処理工程において、少なくとも1,000個のユニーク合成核酸の追加的プールをサンプルに加えることを含み、ここで、第2サンプル処理工程は第1サンプル処理工程とは異なる。幾つかの場合には、該方法は更に、少なくとも1,000個の合成核酸の追加的プールに関する多様性減少を計算することを含む。幾つかの場合には、該方法は更に、少なくとも1,000個の合成核酸に関する多様性減少を少なくとも1,000個の合成核酸の追加的プールに関する多様性減少と比較することにより、比較的高い多様性減少を示すサンプル処理工程を特定することを含む。

20

【0012】

幾つかの場合には、少なくとも1,000個のユニーク合成核酸の追加的プールにおけるユニーク合成核酸のそれぞれは、少なくとも1,000個の合成核酸の追加的プールのメンバーとして該合成核酸を特定するドメインを含む。幾つかの場合には、該方法は更に、サンプル識別核酸をサンプルに加えることを含み。幾つかの場合には、前記の(a)は更に、非ユニーク合成核酸をサンプルに加えることを含み。

【0013】

幾つかの実施形態においては、計算される存在量は相対的存在量である。幾つかの実施形態においては、計算される存在量は絶対的存在量である。

30

【0014】

もう一つの態様においては、サンプルにおける病原体核酸の相対的存在量または初期存在量を決定する方法を本発明で提供し、該方法は、(a)病原体に感染している又は感染していると疑われる対象からサンプルを得、ここで、該サンプルは複数の病原体核酸を含み、(b)該サンプルが既知初期存在量の合成核酸を含むように、複数の合成核酸を該サンプルに加え、ここで、(i)該合成核酸は500塩基対長未満であり、(ii)該合成核酸は、第1の長さを有する合成核酸、第2の長さを有する合成核酸、および第3の長さを有する合成核酸を含み、ここで、第1、第2および第3の長さは異なり、(iii)第1の長さを有する合成核酸は、少なくとも3つの異なるGC含量を有する合成核酸を含み、(c)前記の複数の合成核酸を含むサンプルに関して配列決定アッセイを行い、それにより、該合成核酸の最終存在量および前記の複数の病原体核酸の最終存在量を決定し、(d)合成核酸の最終存在量および既知初期存在量を比較して、該合成核酸に関する回収プロファイルを得、(e)該合成核酸に関する回収プロファイルを使用して、該病原体核酸を、最も近いGC含量および長さを有する合成核酸と比較し、それにより、前記の複数の病原体核酸の相対的存在量または初期存在量を決定することにより、前記の複数の病原体核酸の最終存在量を正規化することを含む。

40

【0015】

幾つかの場合には、前記の少なくとも3つの異なるGC含量は、10%~40%である第1のGC含量、40%~60%である第2のGC含量、および60%~90%である第3のGC含量を含む。幾つかの場合には、前記の少なくとも3つの異なるGC含量はそれ

50

ぞれ10%~50%である。幾つかの場合には、前記の少なくとも3つの異なるGC含量はそれぞれ5%~40%である。幾つかの場合には、該合成核酸は200塩基対またはヌクレオチド長未満である。幾つかの場合には、該合成核酸は100塩基対またはヌクレオチド長未満である。幾つかの場合には、前記の少なくとも3つの異なるGC含量は、少なくとも4つ、少なくとも5つ、少なくとも6つ、少なくとも7つ、または少なくとも8つの異なるGC含量である。幾つかの場合には、該合成核酸は、少なくとも第4の長さ、少なくとも第5の長さ、少なくとも第6の長さ、少なくとも第7の長さ、少なくとも第9の長さ、少なくとも第10の長さ、少なくとも第12の長さ、または少なくとも15番目の長さを有する。幾つかの実施形態においては、各長さは、少なくとも3、4、5、6、7、8、9、10個の異なるGC含量、または50個以下の異なるGC含量を有する合成核酸を含む。

10

**【0016】**

幾つかの場合には、該合成核酸は二本鎖DNAを含む。幾つかの場合には、該方法は更に、該合成核酸を使用して、病原体核酸の変性をモニターすることを含む。幾つかの場合には、該方法は更に、重み係数を使用することにより、病原体核酸の相対的存在量または初期存在量を正規化することを含む。幾つかの場合には、第1合成核酸の既知濃度および第2合成核酸の既知濃度と比較して、前記の複数の合成核酸の第1合成核酸の生測定値および前記の複数の合成核酸の第2合成核酸の生測定値を分析することにより、重み付け係数を得る。

**【0017】**

もう一つの態様においては、病原体からの核酸を検出するための方法を本発明で提供し、該方法は、(a)第1病原体核酸を含む第1サンプルを得、ここで、第1サンプルは、第1病原体に感染している第1対象から得られ、(b)第2対象から第2サンプルを得、(c)第1病原体核酸にハイブリダイズし得ない異なる合成核酸をそれぞれが含む第1サンプル識別子および第2サンプル識別子を得、第1サンプル識別子を第1サンプルに割り当て、第2サンプル識別子を第2サンプルに割り当て、(d)第1サンプル識別子を第1サンプルに、そして第2サンプル識別子を第2サンプルに加え、(e)第1サンプル識別子を含む第1サンプルに関して、そして第2サンプル識別子を含む第2サンプルに関して配列決定アッセイを行い、それにより、第1サンプルおよび第2サンプルに関する配列結果を得、(f)第1サンプルに関する配列結果における第1サンプル識別子、第2サンプル識別子および第1病原体核酸の存在または非存在を検出し、(g)該配列決定アッセイが、第1サンプルにおいて、(i)第1サンプル識別子を検出し、(ii)第1病原体核酸を検出し、および(iii)第2サンプル識別子を検出せず、または閾値レベル未満の第2サンプル識別子を検出しない場合には、検出された第1病原体核酸が第1サンプルに元々存在すると決定することを含む。

20

30

**【0018】**

もう一つの態様においては、核酸を検出するための方法を本発明で提供し、該方法は、(a)第1核酸を含む第1核酸サンプルを得、(b)第1陽性対照核酸を含む第1対照核酸サンプルを得、(c)第1核酸にハイブリダイズし得ない合成核酸を含む第1サンプル識別子を第1対照核酸に加え、(d)第1サンプル識別子を含む第1対照核酸サンプルおよび第1核酸サンプルに関して配列決定アッセイを行い、それにより、第1および対照核酸サンプルの配列リードを得、(e)第1核酸サンプルに関する配列リードを参照配列とアライメント(整列)させて、第1核酸サンプルに関する配列リードにおける第1サンプル識別子の存在または非存在を検出し、(f)該配列リードのアライメントに基づいて、第1陽性対照核酸が第1核酸サンプルに存在するかどうかを決定することを含む。

40

**【0019】**

幾つかの場合には、第1サンプル識別子の合成核酸は150塩基対またはヌクレオチド長未満である。幾つかの場合には、第1陽性対照核酸は病原体核酸である。幾つかの場合には、第1サンプル識別子は修飾核酸を含む。幾つかの場合には、第1サンプル識別子はDNAを含む。幾つかの場合には、サンプルは無細胞体液を含む。幾つかの場合には、サ

50

ンプルは、病原体に感染している対象からのものである。

【0020】

もう一つの態様においては、サンプルにおいて試薬を検出するための方法を本発明で提供し、該方法は、(a)第1合成核酸を試薬に加え、ここで、第1合成核酸はユニーク配列を含み、(b)第1合成核酸を含む試薬を核酸サンプルに加え、(c)配列決定アッセイのための核酸サンプルを調製し、(d)核酸サンプルに関して配列決定アッセイを行い、それにより、核酸サンプルに関する配列結果を得、(e)核酸サンプルに関する配列結果に基づいて、該サンプルにおける第1合成核酸の存在または非存在を決定することにより、該サンプルにおいて試薬を検出することを含む。

【0021】

幾つかの場合には、第1合成核酸は150塩基対またはヌクレオチド長未満である。幾つかの場合には、第1合成核酸を第1試薬ロットに加え、更に、第2合成核酸を第2試薬ロットに加えることを含む。幾つかの場合には、サンプルにおいて試薬を検出することは、試薬の特定のロットを検出することを含む。幾つかの場合には、該合成核酸はヌクレアーゼにより分解可能でない。幾つかの場合には、試薬は水性バッファーを含む。幾つかの場合には、試薬は抽出試薬、酵素、リガーゼ、ポリメラーゼまたはdNTPを含む。

【0022】

もう一つの態様においては、配列決定ライブラリーの製造(作製)方法を本発明で提供し、該方法は、(a)(i)標的核酸、(ii)配列決定アダプター、および(iii)少なくとも一つの合成核酸を含むサンプルを得、ここで、前記の少なくとも一つの合成核酸はDNAを含み、核酸への連結に抵抗し、(b)該配列決定アダプターが前記の少なくとも一つの合成核酸よりも優先的に標的核酸に連結するように、該サンプルに関して連結反応を行うことを含む。

【0023】

もう一つの態様においては、配列決定ライブラリーの製造方法を本発明で提供し、該方法は、(a)標的核酸と少なくとも一つの合成核酸とを含むサンプルを得、(b)前記の少なくとも一つの合成核酸を該サンプルから除去し、それにより、該標的核酸を含み前記の少なくとも一つの合成核酸を含まない配列決定サンプルを得、(c)配列決定アダプターを配列決定サンプルにおける標的核酸に結合させることを含む。

【0024】

もう一つの態様においては、配列決定ライブラリーの製造方法を本発明で提供し、該方法は、(a)標的核酸と少なくとも一つの合成核酸とを含むサンプルを得、ここで、前記の少なくとも一つの合成核酸は、(i)一本鎖DNA、(ii)該合成核酸の増幅を抑制するヌクレオチド修飾、(iii)固定化タグ、(iv)DNA-RNAハイブリッド、(v)標的核酸の長さより長い長さを有する核酸、または(vi)それらの任意の組合せを含み、(b)配列決定反応のためのサンプルから配列決定ライブラリーを製造すること(ここで、前記の少なくとも一つの合成核酸の少なくとも一部は該配列決定反応において配列決定されない)を含む。

【0025】

もう一つの態様においては、配列決定ライブラリーの製造方法を本発明で提供し、該方法は、(a)(i)標的核酸、(ii)配列決定アダプター、および(iii)少なくとも一つの合成核酸を含むサンプルを得、ここで、前記の少なくとも一つの合成核酸はDNAを含み、末端修復に抵抗し、(b)標的核酸が前記の少なくとも一つの合成核酸よりも優先的に末端修復されるように、該サンプルに関して末端修復反応を行うことを含む。

【0026】

もう一つの態様においては、配列決定ライブラリーを製造するためのキットを本発明で提供し、該キットは、(a)配列決定アダプター、および(b)少なくとも一つの合成核酸を含み、ここで、前記の少なくとも一つの合成核酸はDNAを含み、核酸に対する末端修復に抵抗する。

【0027】

10

20

30

40

50

1つの態様においては、標的核酸を含む初期サンプルにおける核酸の絶対的または相対的存在量を決定するための方法を本発明で提供し、該方法は、(a)少なくとも1,000個のユニーク合成核酸の出発量をサンプルに加え、ここで、前記の少なくとも1,000個のユニーク合成核酸のそれぞれは、(i)識別タグ、および(ii)可変領域を含み、(b)該サンプルにおける標的核酸の一部および前記の少なくとも1,000個のユニーク合成核酸の一部に関して配列決定アッセイを行い、それにより、標的および合成核酸配列リードを得、ここで、該合成核酸配列リードは識別タグ配列および可変領域配列を含み、(c)(i)該識別タグ配列の少なくとも一部に対応する配列リードを検出して、第1配列リードのセットを得、(ii)第1配列リード内の異なる可変領域配列の数を定量して、ユニーク配列決定値を得、(iii)前記の少なくとも1,000個のユニーク合成核酸の出発量を該ユニーク配列決定値と比較して、前記の少なくとも1,000個のユニーク合成核酸の多様性減少を得ることにより、前記の少なくとも1,000個のユニーク合成核酸の多様性減少を検出し、(d)前記の少なくとも1,000個のユニーク合成核酸の多様性減少を用いて、初期サンプルにおける標的核酸の絶対的または相対的存在量を計算することを含む。幾つかの場合には、比較する出発量は出発濃度である。

10

## 【0028】

幾つかの場合には、標的核酸は病原体核酸を含む。幾つかの場合には、標的核酸は、少なくとも5つの異なる病原体からの病原体核酸を含む。幾つかの場合には、前記の少なくとも1,000個のユニーク合成核酸はDNAを含む。

## 【0029】

20

幾つかの場合には、前記の少なくとも1,000個のユニーク合成核酸のそれぞれは500塩基対またはヌクレオチド長未満である。幾つかの場合には、前記の少なくとも1,000個のユニーク合成核酸のそれぞれは200塩基対またはヌクレオチド長未満である。幾つかの場合には、前記の少なくとも1,000個のユニーク合成核酸のそれぞれは100塩基対またはヌクレオチド長未満である。

## 【0030】

幾つかの場合において、サンプルは血液、血漿、血清、脳脊髄液、滑液、気管支肺胞洗浄液、尿、便、唾液または鼻サンプルである。幾つかの場合には、サンプルは、単離された核酸のサンプルである。幾つかの場合には、サンプルはヒト対象由来である。

## 【0031】

30

幾つかの場合には、該方法は更に、該サンプルから配列決定ライブラリーを製造することを含み、ここで、配列決定ライブラリーを製造する前に前記の少なくとも1,000個のユニーク合成核酸をサンプルに加える。幾つかの場合には、前記の少なくとも1,000個のユニーク合成核酸の多様性減少はサンプルのサンプル処理中の1以上の核酸の減少を示す。幾つかの場合には、該識別タグは共通配列を含む。幾つかの場合には、第1配列リード内の少なくとも1,000個のユニーク配列の定量は、第1配列リード内のユニーク配列のリード数を決定することを含む。

## 【0032】

幾つかの場合には、前記の少なくとも1,000個のユニーク合成核酸は少なくとも $10^4$ 個のユニーク合成核酸を含む。幾つかの場合には、前記の少なくとも1,000個のユニーク合成核酸は少なくとも $10^5$ 個のユニーク合成核酸を含む。幾つかの場合には、該方法は更に、少なくとも3つの異なる長さを有する追加的合成核酸を加えることを含む。

40

## 【0033】

幾つかの場合には、該方法は更に、第1の長さを有する第1の追加的合成核酸群、第2の長さを有する第2の追加的合成核酸群、および第3の長さを有する第3の追加的合成核酸群を加えることを含み、ここで、第1、第2および第3の追加的合成核酸群のそれぞれは、少なくとも3つの異なるGC含量を有する合成核酸を含む。幾つかの場合には、該方法は更に、該追加的合成核酸を使用して、サンプルにおける標的核酸の絶対的または相対的存在量を計算することを含む。幾つかの場合には、該方法は更に、該追加的合成核酸

50



を使用して、該追加的合成核酸の長さ、GC含量または長さおよびGC含量の両方に基づいてサンプルにおける標的核酸の絶対的または相対的存在量を計算することを含む。

【0034】

幾つかの場合には、第1サンプル処理工程において、前記の少なくとも1,000個のユニーク合成核酸をサンプルに加える。幾つかの場合には、該方法は更に、第2サンプル処理工程において、少なくとも1,000個のユニーク合成核酸の追加的プールをサンプルに加えることを含み、ここで、第2サンプル処理工程は第1サンプル処理工程とは異なる。幾つかの場合には、該方法は更に、少なくとも1,000個のユニーク合成核酸の追加的プールに関する多様性減少を計算することを含む。幾つかの場合には、該方法は更に、前記の少なくとも1,000個のユニーク合成核酸に関する多様性減少を少なくとも1,000個のユニーク合成核酸の追加的プールに関する多様性減少と比較することにより、比較的高い多様性減少を示すサンプル処理工程を特定することを含む。

10

【0035】

幾つかの場合には、少なくとも1,000個のユニーク合成核酸の追加的プールにおけるユニーク合成核酸のそれぞれは、少なくとも1,000個の合成核酸の追加的プールのメンバーとして該合成核酸を特定するドメインを含む。幾つかの場合には、該方法は更に、サンプル識別核酸をサンプルに加えることを含み。幾つかの場合には、前記の(a)は更に、非ユニーク合成核酸をサンプルに加えることを含み。幾つかの場合には、参照配列とアライメント(整列)させることにより、可変配列リードを検出する。幾つかの場合には、可変配列リードを互いにアライメントさせ、重複配列リードを除外することにより、異なる可変配列リードの数を定量する。

20

【0036】

核酸のサンプルにおける病原体核酸の相対的存在量または濃度を決定する方法を本発明で提供する。幾つかの場合には、該方法は、病原体に感染している又は感染していると疑われる対象からサンプルを得(ここで、該サンプルは2以上の病原体核酸を含み、ここで、前記の2以上の病原体核酸は、異なる長さを有する第1病原体核酸および第2病原体核酸を含む)、既知濃度の2以上の合成核酸をサンプルに加え(ここで、前記の2以上の合成核酸は、第1病原体核酸の65%~135%、75%~125%、または85%~115%の長さを有する第1合成核酸、および第2病原体核酸の65%~135%、75%~125%、または85%~115%の長さを有する第2合成核酸を含み、ここで、前記の2以上の合成核酸は第1または第2病原体核酸にハイブリダイズしない)、サンプルに関して配列決定アッセイを行い、それにより、前記の2以上の合成核酸、第1病原体核酸および第2病原体核酸に関する生測定値を得、第1合成核酸の生測定値を第1合成核酸の既知濃度と比較して、第1合成核酸に関する回収プロファイルを得、第1合成核酸に関する回収プロファイルを使用して、第1病原体核酸に関する生測定値を正規化し、それにより、第1病原体核酸の相対的存在量または出発濃度を決定することを含む。

30

【0037】

幾つかの場合において、第1病原体核酸および第2の病原体核酸は、同じ病原体に由来する。幾つかの場合には、第1病原体核酸および第2の病原体核酸は、異なる病原体に由来する。幾つかの場合には、本明細書に記載されている方法は更に、重み係数を使用することにより、第1病原体核酸の相対的存在量または出発濃度を正規化することを含む。幾つかの場合には、第1合成核酸の既知濃度および第2合成核酸の既知濃度と比較して、前記の複数の合成核酸の第1合成核酸の生測定値および前記の複数の合成核酸の第2合成核酸の生測定値を分析することにより、重み付け係数を得る。

40

【0038】

核酸のサンプルにおける核酸の相対的存在量または出発濃度を決定する方法を本発明で提供し、該方法は、(a)対象から核酸サンプルを得(ここで、該核酸サンプルは、異なる長さを有する第1核酸および第2核酸を含む)、既知濃度の2以上の合成核酸をサンプルに加え(ここで、(i)前記の2以上の合成核酸は、第1核酸の65%~135%、75%~125%、または85%~115%の長さを有する第1合成核酸、および第2核酸

50

の長さの65%~135%、75%~125%、または85%~115%の長さを有する第2合成核酸を含み、(ii)第1合成核酸は特定の長さのロード(load)ドメインと、ロードドメインの特定の長さを識別するようにコードされたユニーク配列を有する識別ドメインとを含み、(iii)前記の2以上の合成核酸は第1または第2核酸にハイブリダイズし得ない)、(b)サンプルに関して配列決定アッセイを行い、それにより、前記の2以上の合成核酸、第1核酸および第2核酸に関する生測定値を得、(c)第1合成核酸の生測定値を第1合成核酸の既知濃度と比較して、回収プロファイルを得、(d)該回収プロファイルを使用して、第1核酸に関する生測定値を正規化し、それにより、第1核酸の相対的存在量または出発濃度を決定することを含む。

【0039】

幾つかの場合には、第1核酸は病原体核酸である。幾つかの場合には、前記の2以上の合成核酸の既知濃度は2以上、3以上、5以上、10以上、50以上、100以上、または1,000以上の異なる濃度を含む。幾つかの場合には、前記の2以上の合成核酸の既知濃度は等モル濃度である。幾つかの場合には、前記の2以上の合成核酸はDNAまたは修飾DNAを含む。幾つかの場合には、前記の2以上の合成核酸はRNAまたは修飾RNAを含む。幾つかの場合には、前記の2以上の合成核酸は、2以上、3以上、5以上、8以上、10以上、50以上、100以上、または1,000以上の異なる長さの核酸を含む。幾つかの場合には、前記の2以上の合成核酸は、2以上、3以上、5以上、8以上、10以上、50以上、100以上、または1,000以上の異なる配列の核酸を含む。幾つかの場合には、前記の2以上の合成核酸は50ヌクレオチド長以下、100ヌクレオチド長以下、200ヌクレオチド長以下、300ヌクレオチド長以下、350ヌクレオチド長以下、400ヌクレオチド長以下、450ヌクレオチド長以下、500ヌクレオチド長以下、750ヌクレオチド長以下、または1,000ヌクレオチド長以下である。幾つかの場合には、前記の2以上の合成核酸は少なくとも10ヌクレオチド長、少なくとも20ヌクレオチド長、または少なくとも30ヌクレオチド長、少なくとも50ヌクレオチド長、少なくとも100ヌクレオチド長、または少なくとも150ヌクレオチド長である。幾つかの場合には、前記の2以上の合成核酸は、前記の2以上の合成核酸を合成物として特定する核酸配列を含む。幾つかの場合には、前記の2以上の合成核酸を合成物として特定する核酸配列は10ヌクレオチド長以下、20ヌクレオチド長以下、30ヌクレオチド長以下、40ヌクレオチド長以下、50ヌクレオチド長以下、100ヌクレオチド長以下、200ヌクレオチド長以下、または500ヌクレオチド長以下である。幾つかの場合には、前記の2以上の合成核酸は、該合成核酸の長さを特定する核酸配列を含む。幾つかの場合には、該合成核酸の長さを特定する核酸配列は10ヌクレオチド長以下、20ヌクレオチド長以下、30ヌクレオチド長以下、40ヌクレオチド長以下、50ヌクレオチド長以下、100ヌクレオチド長以下、200ヌクレオチド長以下、または500ヌクレオチド長以下である。

【0040】

幾つかの場合において、サンプルは、血液、血漿、血清、脳脊髄液、滑液、気管支肺胞洗浄液、尿、便、唾液、鼻スワブおよびそれらの任意の組合せからなる群から選択される。幾つかの場合には、サンプルは無細胞核酸を含む。幾つかの場合には、サンプルは循環無細胞核酸を含む。幾つかの場合には、対象はヒトである。幾つかの場合には、病原体は細菌、ウイルス、真菌または寄生生物である。幾つかの場合には、対象は敗血症を有する又は有すると疑われる。幾つかの場合には、病原体は敗血症に関連している。幾つかの場合には、前記の2以上の病原体核酸は3以上、5以上、10以上、50以上、100以上、1,000以上、2,000以上、5,000以上、8,000以上、10,000以上、15,000以上、または20,000以上の病原体核酸配列を含む。

【0041】

幾つかの場合には、第1病原体核酸の相対的存在量の決定は、1以上のゲノムコピーを生成させることを含む。幾つかの場合には、1以上のゲノムコピーの生成は体積当たりのゲノムコピーとして表される。幾つかの場合には、本明細書に記載されている方法は更に

10

20

30

40

50

、サンプルから核酸を抽出することを含む。幾つかの場合には、サンプルからの核酸の抽出は、磁気ビーズを使用して行う。幾つかの場合には、本明細書に記載されている方法は更に、低品質の配列決定リードを除去することを含む。幾つかの場合には、本明細書に記載されている方法は更に、対象の種の参照配列に対してアライメント（整列）またはマッピングされた配列決定リードを除去することを含む。幾つかの場合には、本明細書に記載されている方法は更に、1以上の異なる長さの核酸を回収する相対的効率を決定することを含む。幾つかの場合には、本明細書に記載されている方法は更に、1以上の合成核酸の測定濃度を決定することを含む。幾つかの場合には、本明細書に記載されている方法は、前記の1以上の合成核酸の測定濃度を既知濃度と比較することを含む。幾つかの場合には、本明細書に記載されている方法は更に、配列決定アッセイにおいて、1以上、2以上、3以上、5以上、10以上、50以上、100以上、1,000以上、2,000以上、5,000以上、8,000以上、10,000以上、15,000以上、または20,000以上の病原体核酸を検出することを含む。幾つかの場合には、本明細書に記載されている方法は更に、配列決定アッセイにおいて、抗微生物、抗細菌、抗ウイルスまたは抗真菌耐性を示す1以上、2以上、3以上、5以上、10以上、50以上、100以上、1,000以上、2,000以上、5,000以上、8,000以上、10,000以上、15,000以上、または20,000以上の病原体核酸を検出することを含む。幾つかの場合には、本明細書に記載されている方法は更に、サンプルにおける2以上、3以上、5以上、10以上、50以上、100以上の病原体の同時存在を特定することを含む。

#### 【0042】

幾つかの場合には、サンプルからの核酸の抽出の前または途中で、前記の2以上の合成核酸をサンプルに加える。幾つかの場合には、サンプルからの核酸の抽出の後および核酸のライブラリー調製の前に、前記の2以上の合成核酸をサンプルに加える。幾つかの場合には、前記の2以上の合成核酸の長さは少なくとも約20塩基対異なる。幾つかの場合には、前記の2以上の合成核酸は3以上、5以上、8以上、10以上、20以上または50以上の合成核酸を含む。幾つかの場合には、前記の2以上の合成核酸は、配列番号111～配列番号118およびそれらの任意の組合せからなる群から選択される。幾つかの場合には、前記の2以上の合成核酸は共通のフォワード配列を共有する。幾つかの場合には、前記の共通のフォワード配列は約20塩基対長以下である。幾つかの場合には、前記の2以上の合成核酸は共通のリバース配列を共有する。幾つかの場合には、前記の共通のリバース配列は約20塩基対長以下である。

#### 【0043】

幾つかの場合には、本明細書に記載されている方法は更に、第2合成核酸の生測定値を第2合成核酸の既知濃度と比較して、第2合成核酸に関する回収プロファイルを得、第2合成核酸に関する回収プロファイルを使用して、第2病原体核酸に関する生測定値を正規化し、それにより、第2病原体核酸の相対的存在量または出発濃度を決定することを含む。

#### 【0044】

幾つかの場合には、前記の2以上の病原体核酸は、異なる長さを有する5以上の病原体核酸を含み、前記の2以上の合成核酸は、前記の5以上の病原体核酸のそれぞれの長さの65%～135%、75%～125%、または85%～115%の長さを有する1以上の合成核酸を含み、前記の2以上の核酸は前記の5以上の病原体核酸にハイブリダイズせず、サンプルに関して配列決定アッセイを行うことにより、前記の2以上の合成核酸および前記の5以上の病原体核酸に関する生測定値を得、該生測定値の比較は、該生測定値を各合成核酸の既知濃度と比較して、各合成核酸に関する回収プロファイルを得ることを含み、および/または該回収プロファイルの使用は、各合成核酸に関する回収プロファイルを使用して、前記の5以上の病原体核酸のそれぞれの生測定値を正規化し、それにより、前記の5以上の病原体核酸のそれぞれの相対的存在量または出発濃度を決定することを含む。幾つかの場合には、前記の5以上の病原体核酸は、10以上、50以上、100以上、1,000以上、2,000以上、5,000以上、8,000以上、10,000以上

、15,000以上、または20,000以上の病原体核酸を含む。幾つかの場合には、本明細書に記載されている方法は更に、前記の2以上の合成核酸および核酸のサンプルにおける核酸を抽出または精製することを含む。幾つかの場合には、前記の2以上の合成核酸および核酸のサンプルにおける核酸の抽出または精製は、前記の2以上の合成核酸および核酸のサンプルにおける核酸の相対的濃度を变化させる。幾つかの場合には、該生測定値はリード数である。

#### 【0045】

病原体からの核酸を検出するための方法を本発明で提供し、該方法は、(a)第1病原体核酸を含む第1核酸サンプルを得、ここで、第1病原体に感染している又は第1病原体に感染していると疑われる第1対象から第1核酸サンプルを得、(b)第2病原体核酸を含む第2核酸サンプルを得、ここで、第2病原体に感染している又は第2病原体に感染していると疑われる第2対象から第2核酸サンプルを得、(c)病原体核酸にハイブリダイズし得ない異なる合成核酸をそれぞれが含む第1サンプル識別子および第2サンプル識別子を得、第1サンプル識別子を第1核酸サンプルに、そして第2サンプル識別子を第2核酸サンプルに割り当て、(d)第1サンプル識別子を第1核酸サンプルに、そして第2サンプル識別子を第2核酸サンプルに加え、(e)第1サンプル識別子を含む第1核酸サンプルに関して、および第2サンプル識別子を含む第2核酸サンプルに関して配列決定アッセイを行い、それにより、第1サンプルおよび第2サンプルに関する配列結果を得、(f)該配列結果における第1サンプル識別子、第2サンプル識別子および病原体核酸の存在または非存在を検出し、(g)該配列決定アッセイが第1サンプル識別子および標的核酸を検出するが、第2サンプル識別子を検出しない場合には、第1サンプル中に標的核酸が元々存在すると決定することを含む。

#### 【0046】

幾つかの場合には、該合成核酸は約500塩基対長以下である。幾つかの場合には、該合成核酸は約100塩基対長以下である。幾つかの場合には、該合成核酸は少なくとも約50塩基対長である。幾つかの場合には、該合成核酸は少なくとも約100塩基対長である。幾つかの場合には、該合成核酸はDNAまたは修飾DNAを含む。幾つかの場合には、該合成核酸はRNAまたは修飾RNAを含む。幾つかの場合には、該合成核酸は修飾核酸である。幾つかの場合には、該合成核酸は、配列番号1~配列番号110およびそれらの任意の組合せからなる群から選択される配列を含む。幾つかの場合には、第1サンプルは無細胞体液を含む。

#### 【0047】

サンプルにおいて試薬を検出するための方法を本発明で提供し、該方法は、第1合成核酸を試薬に加え(ここで、第1合成核酸はユニーク配列を含む)、第1合成核酸を含む試薬を核酸サンプルに加え、配列決定アッセイのための核酸サンプルを調製し、核酸サンプルに関して配列決定アッセイを行い、それにより、核酸サンプルに関する配列結果を得、核酸サンプルに関する配列結果に基づいて、該サンプルにおける第1合成核酸の存在または非存在を決定することにより、該サンプルにおいて試薬を検出することを含む。

#### 【0048】

幾つかの場合には、工程aにおいて第1合成核酸を試薬に加えることは、試薬の特定のロットに第1合成核酸を加えることを含む。幾つかの場合には、本明細書に記載されている方法は更に、核酸サンプルに関する配列結果に基づいて、試薬の特定のロットを検出することを含む。幾つかの場合には、第1合成核酸は病原体からの核酸にハイブリダイズしない。幾つかの場合には、本明細書に記載されている方法は更に、試薬の異なるロットに第2合成核酸を加えることを含み、ここで、第2合成核酸は、試薬の異なるロットをユニーク(一意)に特定する。幾つかの場合には、本明細書に記載されている方法は更に、核酸サンプルの配列決定アッセイからの結果に基づいて標的核酸を検出することを含む。幾つかの場合には、本明細書に記載されている方法は更に、(i)標的核酸が正確に検出される場合には、将来の配列決定アッセイにおいて、試薬の特定のロットを使用し、または(ii)標的核酸が正確に検出されない場合には、将来の配列決定アッセイにおいて、試

10

20

30

40

50

薬の特定のロットを使用することを控えることを含む。幾つかの場合には、該試薬は水溶液を含む。幾つかの場合には、該合成核酸は約50～約500塩基対長である。幾つかの場合には、該合成核酸はDNAまたは修飾DNAを含む。幾つかの場合には、該合成核酸はRNAまたは修飾RNAを含む。幾つかの場合には、該合成核酸は、配列番号1～配列番号110およびそれらの任意の組合せからなる群から選択される配列を含む。幾つかの場合には、該合成核酸はDNアーゼにより分解され得ない。

**【0049】**

サンプルにおける核酸の多様性減少または存在量を決定するための方法を本発明で提供し、該方法は、標的核酸を含むサンプルに既知濃度の1,000個のユニーク合成核酸を加え、サンプルに関して配列決定アッセイを行い、それにより、標的核酸の配列リード数、および前記の1,000個のユニーク合成核酸の少なくとも一部の配列リード数を、前記の1,000個のユニーク合成核酸の少なくとも一部の配列リード数を、工程aにおいて標的核酸を含むサンプルに加えられた1,000個のユニーク核酸の配列とアライメント(整列)させ、アライメント配列リード数の多様性を前記の1,000個以上のユニーク合成核酸の多様性と比較することにより、前記の1,000個のユニーク合成核酸の多様性減少を検出し、前記の1,000個のユニーク合成核酸の多様性減少を使用して、該サンプルにおける標的核酸における多様性減少または該標的核酸の存在量を計算することを含む。

**【0050】**

幾つかの場合には、前記の1,000個のユニーク合成核酸は約500塩基対長以下または約100塩基対長以下である。幾つかの場合には、前記の1,000個のユニーク合成核酸を等モル濃度で加える。幾つかの場合には、前記の1,000個のユニーク合成核酸は少なくとも約 $1 \times 10^6$ の多様性を有する。幾つかの場合には、前記の1,000個のユニーク合成核酸は少なくとも約 $1 \times 10^6$ の多様性を有する。幾つかの場合には、前記の1,000個のユニーク合成核酸は少なくとも約 $1 \times 10^7$ の多様性を有する。幾つかの場合には、前記の1,000個のユニーク合成核酸は少なくとも約 $1 \times 10^8$ の多様性を有する。幾つかの場合には、前記の1,000個のユニーク合成核酸はランダム化部分を有する。幾つかの場合には、前記の1,000個のユニーク合成核酸はDNA、修飾DNA、RNAまたは修飾RNAを含む。幾つかの場合には、前記の1,000個のユニーク合成核酸は、配列番号119および配列番号120において特定されている配列を含む。幾つかの場合には、第1サンプル処理工程において、前記の1,000個のユニーク合成核酸をサンプルに加える。幾つかの場合には、該方法は更に、第2サンプル処理工程において、1,000個のユニーク合成核酸の追加的プールをサンプルに加えることを含み、ここで、第2サンプル処理工程は第1サンプル処理工程とは異なる。幾つかの場合には、1,000個のユニーク合成核酸の追加的プールに関する多様性減少を計算する。幾つかの場合には、本明細書に記載されている方法は、前記の1,000個のユニーク合成核酸に関する多様性減少を1,000個のユニーク合成核酸の追加的プールに関する多様性減少と比較することにより、比較的高い多様性減少を示すサンプル処理工程を特定することを含む。幾つかの場合には、前記の1,000個のユニーク合成核酸は、前記の1,000個のユニーク合成核酸のプールのメンバーとして該合成核酸を特定するドメインを含む。幾つかの場合には、1,000個のユニーク合成核酸の追加的プールは、1,000個のユニーク合成核酸の追加的プールのメンバーとして該合成核酸を特定するドメインを含む。幾つかの場合には、前記の1,000個のユニーク合成核酸を標的核酸の抽出の前にサンプルに加える。幾つかの場合には、前記の1,000個のユニーク合成核酸を標的核酸のライブラリー調製の前にサンプルに加える。幾つかの場合には、本明細書に記載されている方法は更に、標的核酸を含むサンプルに既知濃度の5,000個のユニーク合成核酸を加えることを含む。

**【0051】**

更に、分子を分析するための方法および組成物を本明細書において開示する。1つの態様においては、配列決定ライブラリーの製造方法を本明細書において開示し、該方法は、

10

20

30

40

50

a) (i) 標的核酸、(ii) 配列決定アダプター、および(iii) 少なくとも1つの合成核酸を含むサンプルを得、ここで、前記の少なくとも1つの合成核酸はDNAを含み、核酸への連結に抵抗し、b) 該配列決定アダプターが前記の少なくとも1つの合成核酸よりも優先的に標的核酸に連結するように、該サンプルに関して連結反応を行うことを含む。

【0052】

幾つかの場合には、前記の少なくとも1つの合成核酸はホスホジエステル結合を介した該核酸への連結に抵抗する。幾つかの場合には、前記の少なくとも1つの合成核酸は該配列決定アダプターへの連結に抵抗する。もう1つの態様においては、配列決定ライブラリーの製造方法を本明細書において開示し、該方法は、a) 標的核酸と少なくとも1つの合成核酸とを含むサンプルを得、b) 前記の少なくとも1つの合成核酸を該サンプルから除去し、それにより、該標的核酸を含み前記の少なくとも1つの合成核酸を含まない配列決定サンプルを得、c) 配列決定アダプターを配列決定サンプルにおける標的核酸に結合させることを含む。幾つかの場合には、前記の少なくとも1つの合成核酸の除去はエンドヌクレアーゼ消化によっては実施されない。幾つかの場合には、サンプルから除去された前記の少なくとも1つの合成核酸は別の合成核酸に結合していない。幾つかの場合には、前記の少なくとも1つの合成核酸は末端修復に抵抗する。

10

【0053】

もう1つの態様においては、配列決定ライブラリーの製造方法を本明細書において開示し、該方法は、a) 標的核酸と少なくとも1つの合成核酸とを含むサンプルを得、b) 配列決定アダプターを該サンプルにおける標的核酸に結合させ、それにより、配列決定サンプルを得、c) 前記の少なくとも1つの合成核酸を、アフィニティに基づく枯渇(除去)、RNA誘導性DNAアーゼ消化またはそれらの組合せにより、該配列決定サンプルから除去する(ここで、該配列決定サンプルからの前記の少なくとも1つの合成核酸の除去は、該配列決定アダプターよりも、および該配列決定アダプターの多量体よりも、前記の少なくとも1つの合成核酸を優先的に除去することを含む)ことを含む。

20

【0054】

幾つかの場合には、該方法は更に、エンドヌクレアーゼ消化、サイズに基づく除去またはそれらの組合せにより、前記の少なくとも1つの合成核酸を除去することを含む。幾つかの場合には、該配列決定アダプターは核酸である。幾つかの場合には、前記の少なくとも1つの合成核酸の除去を、アフィニティに基づく枯渇により行い、前記の少なくとも1つの合成核酸は固定化タグを含む。幾つかの場合には、前記の少なくとも1つの合成核酸の除去をRNA誘導性DNAアーゼ消化により行う。幾つかの場合には、RNA誘導性DNAアーゼはCRISPR関連タンパク質を含む。幾つかの場合には、前記の少なくとも1つの合成核酸の除去をエンドヌクレアーゼ消化により行う。幾つかの場合には、前記の少なくとも1つの合成核酸の除去を、サイズに基づく除去により行い、前記の少なくとも1つの合成核酸は、標的核酸の長さより長い長さを有する。幾つかの場合には、前記の少なくとも1つの合成核酸の除去をRNAアーゼで行い、前記の少なくとも1つの合成核酸の除去はDNA-RNAハイブリッドである。幾つかの場合には、配列決定アダプターを標的核酸に結合させることは、配列決定アダプターを標的核酸に連結することを含む。幾つかの場合には、配列決定アダプターを標的核酸に結合させることは、配列決定アダプターを標的核酸に連結することを含む。

30

40

【0055】

もう1つの態様においては、配列決定ライブラリーの製造方法を本明細書において開示し、該方法は、a) 標的核酸と少なくとも1つの合成核酸とを含むサンプルを得、ここで、前記の少なくとも1つの合成核酸は、(i) 一本鎖DNA、(ii) 該合成核酸の増幅を抑制するヌクレオチド修飾、(iii) 固定化タグ、(iv) DNA-RNAハイブリッド、(v) 標的核酸の長さより長い長さを有する核酸、または(vi) それらの任意の組合せを含み、b) 配列決定反応のためのサンプルから配列決定ライブラリーを製造すること(ここで、前記の少なくとも1つの合成核酸の少なくとも一部は該配列決定反応にお

50

いて配列決定されない)を含む。

【0056】

幾つかの場合には、前記の少なくとも1つの合成核酸は更に、エンドヌクレアーゼ認識部位を含む。幾つかの場合には、サンプルを得ることは、試験サンプルから標的核酸を抽出することを含み、更に、試験サンプルから標的核酸を抽出した後、前記の少なくとも1つの合成核酸を試験サンプルに加えることを含む。幾つかの場合には、サンプルを得ることは、試験サンプルから標的核酸を抽出することを含み、更に、試験サンプルから標的核酸を抽出する前に、前記の少なくとも1つの合成核酸を試験サンプルに加えることを含む。幾つかの場合には、前記の少なくとも1つの合成核酸が、連結反応を阻害するブロッキング基を含む場合には、該ブロッキング基は修飾ヌクレオチドを含む。幾つかの場合には、修飾ヌクレオチドは反転(inverted)デオキシ-糖を含む。幾つかの場合には、反転デオキシ-塩基は3'反転デオキシ-糖を含む。幾つかの場合には、修飾ヌクレオチドは反転チミジン、反転アデノシン、反転グアノシンまたは反転シチジンを含む。幾つかの場合には、修飾ヌクレオチドは反転ジデオキシ-糖を含む。幾つかの場合には、反転ジデオキシ-糖は5'反転ジデオキシ-糖を含む。幾つかの場合には、修飾ヌクレオチドは反転ジデオキシ-シチミジン、反転ジデオキシ-アデノシン、反転ジデオキシ-グアノシンまたは反転ジデオキシ-シチジンを含む。幾つかの場合には、修飾ヌクレオチドはジデオキシ-シチジンである。幾つかの場合には、前記の少なくとも1つの合成核酸は連結反応を阻害するブロッキング基を含み、ブロッキング基はスペーサーを含む。幾つかの場合には、スペーサーはC3スペーサーまたはスペーサー18を含む。幾つかの場合には、前記の少なくとも1つの合成核酸は、連結反応を阻害するブロッキング基を含み、ブロッキング基はヘアピン構造を含む。幾つかの場合には、該合成核酸は、前記の少なくとも1つの合成核酸の増幅を阻害するヌクレオチド修飾を含み、該ヌクレオチド修飾は少なくとも1つの脱塩基部位を含む。幾つかの場合には、前記の少なくとも1つの脱塩基部位は少なくとも1つの内部脱塩基部位である。幾つかの場合には、該ヌクレオチド修飾は8~10個の脱塩基部位を含む。幾つかの場合には、前記の少なくとも1つの脱塩基部位は単一脱塩基部位である。幾つかの場合には、前記の少なくとも1つの脱塩基部位は修飾リボース上に存在する。幾つかの場合には、前記の少なくとも1つの脱塩基部位は、1', 2'-ジデオキシリボース、ロック化(locked)核酸、架橋核酸、またはねじれたインターカレーティング核酸を含む。幾つかの場合には、前記の少なくとも1つの合成核酸は固定化タグを含み、固定化タグはビオチン、ジゴキシゲニン、ポリヒスチジンまたはNi-ニトリロ三酢酸を含む。幾つかの場合には、前記の少なくとも1つの合成核酸はDNAを含み、内部ウラシルで標識されている。幾つかの場合には、前記の少なくとも1つの合成核酸を、ウラシル特異的切除試薬酵素を使用して、配列決定サンプルから除去する。

【0057】

幾つかの場合には、試験サンプルは生物学的サンプルである。幾つかの場合には、生物学的サンプルは全血、血漿、血清または尿である。幾つかの場合には、標的核酸は無細胞核酸である。幾つかの場合には、無細胞核酸は無細胞DNAである。幾つかの場合には、無細胞核酸は病原体核酸である。幾つかの場合には、無細胞核酸は循環無細胞核酸である。幾つかの場合には、前記の少なくとも1つの合成核酸は二本鎖核酸を含む。幾つかの場合には、前記の少なくとも1つの合成核酸は一本鎖核酸を含む。幾つかの場合には、前記の少なくとも1つの合成核酸はDNA、RNA、DNA-RNAハイブリッドまたはそれらの任意の類似体を含む。

【0058】

幾つかの場合には、該方法は更に、(a)サンプルから標的核酸を抽出すること、(b)サンプルから標的核酸を精製すること、(c)標的核酸を末端修復すること、(d)標的核酸を断片化すること、(e)標的核酸を増幅すること、(f)配列決定アダプターを標的核酸に結合させること、および(g)標的核酸を配列決定することの1以上を含む。幾つかの場合には、該方法は、配列決定アダプターを標的核酸に結合させることを含み、更に、配列決定アダプターを標的核酸に結合させる前に、配列決定サンプルをエンドヌク

10

20

30

40

50

レアーゼで処理することを含む。幾つかの場合には、該方法は、配列決定アダプターを標的核酸に結合させることを含み、更に、配列決定アダプターを標的核酸に結合させた後、配列決定サンプルをエンドヌクレアーゼで処理することを含む。幾つかの場合には、該方法は、標的核酸を末端修復することを含み、ここで、標的核酸を末端修復する前に、前記の少なくとも1つの合成核酸をサンプルに加える。幾つかの場合には、該方法は、標的核酸を末端修復することを含み、ここで、標的核酸を末端修復した後、前記の少なくとも1つの合成核酸をサンプルに加える。幾つかの場合には、該方法は、配列決定アダプターを標的核酸に結合させることを含み、配列決定アダプターを標的核酸に結合させる前に、前記の少なくとも1つの合成核酸をサンプルに加える。幾つかの場合には、サンプルにおける前記の少なくとも1つの合成核酸の濃度とサンプルにおける標的核酸の濃度との比は1 : 1 ~ 1000 : 1である。

10

**【0059】**

幾つかの場合には、前記の少なくとも1つの合成核酸のサイズと標的核酸のサイズとの差異は、前記の少なくとも1つの合成核酸を標的核酸からサイズに基づいて分離することを可能にする。幾つかの場合には、合成核酸は、連結反応を阻害するブロッキング基、および増幅反応を阻害するヌクレオチド修飾を含む。幾つかの場合には、連結反応を阻害するブロッキング基は3'反転デオキシ-Tを含み、増幅反応を阻害するヌクレオチド修飾は内部脱塩基部位を含む。幾つかの場合には、ブロッキング基は更に、5'反転ジデオキシ-Tを含む。幾つかの場合には、該方法は更に、サンプルをエンドヌクレアーゼV I I Iと共にインキュベートすることを含む。幾つかの場合には、サンプルをエンドヌクレアーゼV I I Iと共に1時間以下、インキュベートする。幾つかの場合には、該方法は、サンプルから標的核酸を抽出することを含み、標的核酸を抽出することは、前記の少なくとも1つの合成核酸を含有しないサンプルから標的核酸を抽出することと比較して高い収率を示す。幾つかの場合には、該方法は、標的核酸を末端修復することを含み、標的核酸を末端修復することは、前記の少なくとも1つの合成核酸を含有しないサンプルにおける標的核酸を末端修復することと比較して高い効率を示す。幾つかの場合には、標的核酸は、天然に存在する核酸またはそのコピーを含む。幾つかの場合には、該方法は更に、コンピュータを使用して、標的核酸の少なくとも1つの配列情報を得ることを含む。

20

**【0060】**

もう1つの態様においては、配列決定ライブラリーの製造方法を本明細書において開示し、該方法は、(a)(i)標的核酸、(ii)配列決定アダプター、および(iii)少なくとも1つの合成核酸(ここで、前記の少なくとも1つの合成核酸はDNAを含み、末端修復に抵抗する)を含むサンプルを得、b)前記の少なくとも1つの合成核酸よりも標的核酸が優先的に末端修復されるように、該サンプルに関して末端修復反応を行うことを含む。

30

**【0061】**

幾つかの実施形態においては、前記方法のいずれかは、該方法の結果を患者、介護者または他の者に報告することを含む。

**【0062】**

もう1つの態様においては、配列決定ライブラリーを製造するためのキットを本明細書において開示し、該キットは、a)配列決定アダプター、およびb)少なくとも1つの合成核酸(ここで、前記の少なくとも1つの合成核酸はDNAを含み、核酸に対する末端修復に抵抗する)を含む。幾つかの場合には、前記の少なくとも1つの合成核酸の量と配列決定アダプターの量との比は1 : 1以下である。

40

**【0063】**

開示されている内容の新規特徴は添付の特許請求の範囲に詳細に記載されている。本開示の内容の特徴および利点のより深い理解は、開示されている内容の原理を利用する例示的な実施形態に記載する以下の詳細な説明および後記の添付図面を参照することによりもたらされる。

**【0064】**

50



## 文献の援用

本明細書に挙げられている全ての刊行物、特許および特許出願の全体を、各個の刊行物、特許または特許出願が参照により本明細書に組み入れられると具体的かつ個別に示されている場合と同様に、参照により本明細書に組み入れることとする。

### 【0065】

#### 詳細な説明

#### 概要

本開示は、次世代シーケンシングアッセイおよび他のアッセイにおける核酸の、改良された特定または定量のための複数の方法およびアプローチを提供する。一般に、本発明で提供する方法は、特定の配列、長さ、GC含量、縮重度、多様性の度合および/または既知の出発濃度のような特別な特徴を有するスパイクイン合成核酸の使用を含む。そのようなスパイクイン合成核酸の使用は、絶対的存在量の決定、相対的存在量の決定、存在量の正規化、汎用的な定量、バイアス制御、サンプルの特定(同定)、交差汚染の検出、情報伝達効率、試薬追跡、多様性減少の正規化、絶対的または相対的な損失の決定、品質管理および多数の他の用途を可能にし、改善しうる。本発明で提供するスパイクイン合成核酸は、特別に設計された担体(carrier)核酸をも含み、該担体核酸はサンプルにおける核酸の全濃度を増加させうるが、配列決定または他のアッセイによる検出を回避する能力を有する。

10

### 【0066】

好ましい実施形態においては、本開示は、スパイクイン合成核酸の種のセットを提供し、ここで、それぞれの種の長さおよび/またはGC含量は、分析されるべき標的核酸のセットの、予想されるまたは観察可能な長さおよび/またはGC含量に合致または近似するように設計される。例えば、スパイクイン合成核酸の長さは、そのような病原体に感染しているヒト患者から得られたサンプル(例えば、血漿)における疾患特異的または病原体特異的無細胞核酸の長さに近似可能である。他の好ましい実施形態においては、本開示は、サンプル、試薬または試薬ロットをユニークに特定するための配列を含むスパイクイン合成核酸を提供する。更に他の好ましい実施形態においては、本開示は、ユニーク配列を有する多数のスパイクイン合成核酸(例えば、 $10^4$ 、 $10^5$ 、 $10^6$ 、 $10^7$ 、 $10^8$ 、 $10^9$ または $10^{10}$ 個のユニークスパイクイン合成核酸)を含むプールを提供し、これは、ハイスループット配列決定アッセイの経過中、特にサンプル処理工程、例えば核酸抽出および/またはライブラリー製造におけるユニークスパイクイン配列の多様性の減少を介してサンプルにおける絶対的核酸減少を追跡するために使用されうる。

20

30

### 【0067】

絶対的核酸減少を追跡しうることは初期サンプルにおける標的核酸の絶対的存在量の決定を可能にしうる。例えば、臨床サンプルにおける病原体の絶対量は、その病原体に起因する配列決定リードの数に基づいて決定されうる。抗生物質または医薬組成物での治療の前、途中および後などに経時的に採取された臨床サンプルにおける病原体の絶対的存在量を決定することにより、医学的治療がモニターまたは調節されうる。特定の病原体が存在するかどうかを決定することに加えて、感染または疾患の度合または段階も決定されうる。

40

### 【0068】

該方法は、臨床サンプル、処理サンプル(例えば、抽出核酸、抽出無細胞DNA、抽出無細胞RNA、血漿、血清)、未処理サンプル(例えば、全血)および任意の他のタイプのサンプル、特に核酸を含むサンプル(これらに限定されるものではない)を含む多種多様なサンプルにスパイクイン合成核酸を加えることを含みうる。幾つかの場合には、該方法は、試薬、特に、配列決定(例えば、次世代シーケンシング)によるサンプルの分析の任意の段階で使用される検査用試薬(または特定の試薬ロット)に該スパイクイン合成核酸を加えることを含みうる。好ましい実施形態においては、該方法は、既知濃度の合成核酸を試薬およびサンプル内に導入することを含みうる。該方法は、臨床サンプルにおける病原体に由来する低存在量の病原体または核酸を検出、特定、モニターまたは定量するよ

50

うに設計されたアッセイの精度および効率を高めるのに特に有用でありうる。該方法はまた、サンプルトラッキング（サンプル追跡）のエラーで生じる、あるいはサンプル調製、核酸精製または配列決定ライブラリー製造中の核酸配列の不均一な減少から生じる、あるいは異なる標的核酸または異なるサンプルの分析を比較する際の内部正規化標準の欠如から生じる望ましくない影響を低減しうる。

#### 【0069】

図1は、特に存在量の正規化に関する本発明で提供する方法の多くの工程の一般的概要を示す。該方法は、対象110（例えば、ヒト患者）からサンプルを得ることを含む。幾つかの特定の実施形態においては、対象は感染症を有し、または病原体に感染していると疑われる。サンプルは、図示されているとおり、血液サンプル120または血漿サンプル130、あるいは任意の他のタイプの生物学的サンプル、特に、体液、組織および/または細胞を含有する生物学的サンプル、あるいは無細胞生物学的サンプルでありうる。

10

#### 【0070】

サンプル140からの核酸（例えば、無細胞核酸）を抽出し、アッセイ、例えば配列決定アッセイ（例えば、次世代シーケンシングアッセイ）において使用することが可能である。1以上のタイプの合成核酸150を、該方法における1以上の工程において、例えば血液サンプル120、血漿サンプル130またはサンプル核酸140に加える（またはスパイクイン（添加）する）。該合成核酸は、分析すべき標的核酸のセットの長さに近似するように設計された長さ、および/または分析すべき標的核酸のセットのGC含量に近似するように設計されたGC含量を有しうる。一般に、該合成核酸は既知の出発濃度をも有する。ついで該合成核酸を含むサンプルを配列決定アッセイ160、例えば次世代シーケンシングアッセイにより分析することが可能である。幾つかの場合には、配列決定アッセイにより特定された合成核酸の量を該合成核酸の既知出発濃度と比較して、リード数を既知出発濃度と相関させる。結果として、特に、検出された標的核酸の存在量を、長さおよび/またはGC含量においてそのような標的核酸170に最も近い合成核酸の存在量と比較することにより、サンプル核酸における標的核酸を特定または定量することが可能である。そのような方法および本発明で提供する他の方法の使用により、対象の状態を、より高い精度および確実性のレベルで特定することが可能である。幾つかの特定の実施形態においては、配列決定アッセイ（例えば、次世代シーケンシングアッセイ）はヒト患者由来の無細胞核酸（例えば、DNA）のサンプルにおける病原体核酸を検出する。

20

30

#### 【0071】

それらの工程は任意の順序および任意の組合せで行われうる。幾つかの場合には、ある工程は行われぬ。幾つかの場合には、図示されている工程に新たな工程を加え、または図示されている工程の間に介在させる。

#### 【0072】

図2は典型的な感染の概要図を示す。病原体感染源は、例えば肺におけるものでありうる。病原体に由来する無細胞核酸、例えば無細胞DNAなどの無細胞核酸は血流を通過して移動し、分析のために血漿サンプル中に収集されうる。ついでサンプルにおける核酸を、図1に示されているとおりに配列決定アッセイにより分析することが可能である。

#### 【0073】

図3は、本発明で提供する方法の幾つかの一般的スキームを示す。該方法は、宿主（例えば、ヒト）核酸および非宿主（例えば、病原体）核酸を含有するサンプルを得ることを含む。サンプルは対象、例えば患者から得られうる。幾つかの特定の実施形態においては、対象は感染症を有し、または病原体に感染していると疑われる。サンプルは血液サンプルまたは血漿サンプル、または任意の他のタイプの生物学的サンプル、特に、体液、組織および/または細胞を含有する生物学的サンプルでありうる。サンプルからの核酸（例えば、無細胞核酸）を既知量の合成核酸と一緒にすることが可能である。ついで、合成核酸を含有するサンプルを、配列決定アッセイ、例えば次世代シーケンシングアッセイにより分析することが可能である。配列決定結果を既知の宿主および非宿主参照配列に対してマッピングすることが可能である。幾つかの場合には、配列決定アッセイにより特定さ

40

50

れた合成核酸の量を該合成核酸の既知出発濃度と比較して、リード数を該既知出発濃度と相関させる。結果として、非宿主配列の相対的存在量が決定されうる。それらの工程は任意の順序および任意の組合せで行われうる。幾つかの場合には、ある工程は数回反復されうる。幾つかの場合には、ある工程は行われぬ。幾つかの場合には、図示されている工程に新たな工程を加え、または図示されている工程の間に介在させる。

**【0074】**

本発明で提供する方法は、標的核酸がサンプル中に低存在量で存在する場合または複数のサンプルもしくは複数の標的核酸を比較もしくは追跡する場合には特に、次世代シーケンシングによる標的核酸の改良された特定または定量を可能にしうる。例えば、次世代シーケンシングによる臨床サンプルにおける標的病原体、腫瘍細胞または腫瘍原性マーカーの正確な検出および定量は、該サンプルが不適切に追跡される場合または標的核酸が不正確に正規化もしくは定量される場合には、損なわれ、または負の影響を受けうる。したがって、本発明で提供する方法は、サンプルトラッキングもしくは特定における又は核酸の定量もしくは配列決定データのクラウド分析におけるエラーから生じる陥穽を回避することを助けうる。

10

**【0075】**

本発明で提供する方法および組成物は、出発サンプルが比較的少量の核酸を含有する場合には特に、配列決定ライブラリーの収率、品質または効率を向上させるために配列決定ライブラリーの製造中に合成核酸を加えるおよび/または除去するために使用されうる。一般に、幾つかの場合には、該合成核酸はこれらの用途において担体核酸として作用して、サンプル調製プロセス中に全核酸の濃度を上昇させうる。サンプルへの該合成核酸の添加は配列決定ライブラリーの製造(作製)の1以上の工程の収率および/または効率を増加させうる。前記の1以上の工程は核酸濃度に感受性でありうる。例えば、該工程の収率および/または効率はサンプルにおける核酸濃度に左右されうる。そのような工程は核酸抽出、精製、連結および末端修復を含みうる。幾つかの場合には、該合成核酸は該配列決定ライブラリーから除去されうる。該合成核酸は、該配列決定ライブラリーの製造における1以上の工程にそれらが関与することを妨げる或る特徴を含みうる。したがって、該合成核酸は配列決定工程において配列決定されない可能性がある。

20

**【0076】**

該方法および組成物は、複数の対象からのサンプルを分析するために(例えば、サンプルにおける標的核酸から配列決定ライブラリーを製造するために)使用されうる。これらのサンプルにおける標的核酸の濃度は対象によって異なりうる。この場合のこれらのサンプルへの該合成核酸の添加はサンプル間の濃度変動を低減して、分析の精度を改善しうる。

30

**【0077】**

該方法および組成物は、少なくとも1つの合成核酸を加えることにより、サンプルから配列決定ライブラリーを製造するために使用されうる。該合成核酸は、配列決定反応においてそれらが配列決定されないようにする1以上の特徴を有しうる。幾つかの場合には、該合成核酸は、配列決定ライブラリーの製造における1以上の反応、例えばアダプター連結および核酸増幅を抑制する修飾を含む。例えば、該核酸は一方または両方の末端に反転糖および/または1以上の脱塩基部位を含みうる。

40

**【0078】**

幾つかの場合には、該合成核酸は配列決定前に配列決定ライブラリーから除去されうる。幾つかの場合には、該合成核酸は酵素消化により除去されうる。例えば、該合成核酸は制限酵素認識部位を含むことが可能であり、制限酵素により分解されうる。幾つかの場合には、該合成核酸は、アフィニティに基づく枯渇により除去されうる。例えば、該合成核酸は1以上の固定化タグを含むことが可能であり、アフィニティに基づく枯渇により除去されうる。ある場合においては、該合成核酸は、サイズに基づく除去により除去されうる。該合成核酸は配列決定ライブラリーにおける他の分子とは異なるサイズを有することも可能であり、その結果、該合成核酸は、サイズに基づく除去により除去されうる。幾つか

50

の場合には、該合成核酸は本明細書における特徴および/または修飾の組合せを含むことが可能であり、その結果、それらは配列ライブラリーの製造の1以上の工程に關与せず、配列決定前に除去されうる。

【0079】

#### サンプル

本発明で提供する方法は多種多様なサンプルの改良された分析を可能にしうる。本発明で提供する合成核酸は、そのようなサンプルを分析するために使用可能であり、それは、サンプルに、またはサンプルの加工形態、例えば臨床血漿サンプルからの抽出無細胞核酸に該合成核酸を直接加えることを含みうる。

【0080】

本発明で提供する方法において分析されるサンプルは、好ましくは、任意のタイプの臨床サンプルである。幾つかの場合には、サンプルは細胞、組織または体液を含有する。好ましい実施形態においては、サンプルは液体または流体サンプルである。幾つかの場合には、サンプルは体液、例えば全血、血漿、血清、尿、便、唾液、リンパ液、髄液、滑液、気管支肺胞洗浄液、鼻腔スワブ、呼吸器分泌物、腔液、羊水、精液または月経物を含有する。幾つかの場合には、サンプルは、全体的または部分的に、細胞または組織から構成される。幾つかの場合には、細胞、細胞断片またはエキソソームは、例えば遠心分離または濾過によりサンプルから取り出される。本明細書におけるサンプルは生物学的サンプルでありうる。

【0081】

サンプルは任意の濃度の核酸を含みうる。本明細書における組成物および方法は、低濃度の全核酸を含有するサンプルに有用でありうる。幾つかの場合には、サンプルは、多くとも100 ng/μL、50 ng/μL、10 ng/μL、5 ng/μL、2 ng/μL、1.5 ng/μL、1.2 ng/μL、1 ng/μL、0.8 ng/μL、0.4 ng/μL、0.2 ng/μL、0.1 ng/μL、0.05 ng/μL、0.01 ng/μL、10 ng/mL、5 ng/mL、2 ng/mL、1 ng/mL、0.8 ng/mL、0.6 ng/mL、0.5 ng/mLまたは0.1 ng/mLの核酸の全濃度を有する。幾つかの場合には、サンプルは、少なくとも0.1 ng/mL、0.5 ng/mL、0.6 ng/mL、0.8 ng/mL、1 ng/mL、2 ng/mL、5 ng/mL、10 ng/mL、0.01 ng/μL、0.05 ng/μL、0.1 ng/μL、0.2 ng/μL、0.4 ng/μL、0.8 ng/μL、1 ng/μL、1.2 ng/μL、1.5 ng/μL、2 ng/μL、5 ng/μL、10 ng/μL、50 ng/μLまたは100 ng/μLの核酸の全濃度を含む。幾つかの場合には、サンプルは約0.1 ng/mL ~ 約10,000 ng/mL (すなわち、約0.1 ng/mL ~ 約10 ng/μL) の範囲内の核酸の全濃度を有する。

【0082】

サンプルは1以上の対照を含みうる。幾つかの場合には、サンプルは1以上の陰性対照を含む。典型的な陰性対照は、汚染物質を特定するために調製されたサンプル(例えば、血漿マイナスサンプル)、健康な対象からの血漿、および低多様性サンプル(例えば、見かけ上健康な対象から採取されたサンプル)を含む。幾つかの場合には、サンプルは1以上の陽性対照を含む。典型的な陽性対照は、既知病原体からのゲノムDNAを有する健康対象からのサンプル(例えば、血漿サンプル)を含む。既知病原体からのゲノムDNAは完全なゲノムDNAでありうる。幾つかの場合には、既知病原体からのゲノムDNAは、例えば種々の平均長までせん断されうる。せん断は機械的せん断(例えば、超音波、流体力学的せん断力)、酵素的せん断(例えば、エンドヌクレアーゼ)、熱分解(例えば、高温でのインキュベーション)、化学的断片化(例えば、アルカリ溶液、二価イオン)により行われうる。

【0083】

サンプルは標的核酸を含みうる。標的核酸は、サンプルにおいて分析される核酸を意味しうる。例えば、標的核酸はサンプル中に元々存在することが可能であり、例えば、天然

10

20

30

40

50

核酸でありうる。サンプルは更に、本明細書に開示されている 1 以上の合成核酸を含みうる。幾つかの場合には、標的核酸は、本明細書に記載されている無細胞核酸である。例えば、標的核酸は無細胞 DNA、無細胞 RNA（例えば、無細胞 mRNA、無細胞 miRNA、無細胞 siRNA）、またはそれらの任意の組合せでありうる。ある場合においては、無細胞核酸は病原体核酸、例えば、病原体からの核酸である。無細胞核酸は循環核酸、例えば、循環腫瘍 DNA または循環胎児 DNA でありうる。サンプルは、病原体、例えばウイルス、細菌、真菌および / または真核寄生生物からの核酸を含みうる。

**【 0 0 8 4 】**

ある場合においては、サンプルはアダプターをも含む。アダプターは、既知または未知配列を有する核酸でありうる。アダプターは核酸の 3' 末端、5' 末端または両方の末端に結合されうる。アダプターは既知配列および / または未知配列を含みうる。アダプターは二本鎖または一本鎖でありうる。幾つかの場合には、アダプターは配列決定アダプターである。配列決定アダプターは標的核酸に結合し、標的核酸の配列決定を助けうる。例えば、配列決定アダプターは、配列決定用プライマー結合部位、ユニーク識別子配列、非ユニーク識別子配列、および固体支持体上に標的核酸を固定化するための配列の 1 以上を含みうる。配列決定アダプターに結合される標的核酸はシーケンサー上の固体支持体上に固定化されうる。配列決定用プライマーはアダプターにハイブリダイズし、配列決定反応において標的核酸を鋳型として使用して伸長されうる。幾つかの場合には、アダプターにおける識別子を使用して、異なる標的配列の配列リードを標識し、それにより、複数の標的核酸のハイスループット配列決定が可能となる。

**【 0 0 8 5 】**

「結合」なる語およびその文法的等価体は、任意の結合形態を用いて 2 つの分子を連結することを意味しうる。例えば、結合は、2 つの分子を化学結合または他の方法により連結して新しい分子を生成させることを意味しうる。核酸にアダプターを結合させることは、アダプターと核酸との間に化学結合を形成させることを意味しうる。幾つかの場合には、結合は、例えばリガーゼを使用する連結により行われる。例えば、核酸アダプターは、リガーゼによって触媒されるホスホジエステル結合を形成させることにより、連結により標的核酸に結合されうる。

**【 0 0 8 6 】**

配列決定ライブラリーは、本発明で提供する方法および組成物を使用して、サンプルから製造されうる。配列決定ライブラリーは、使用される配列決定システムに適合した複数の核酸を含みうる。例えば、配列決定ライブラリーにおける核酸は、1 以上のアダプターに結合した標的核酸を含みうる。配列決定ライブラリーを製造するための工程は、サンプルから標的核酸を抽出すること、標的核酸を断片化すること、標的核酸にアダプターを結合させること、標的核酸 - アダプター複合体を増幅すること、および増幅された標的核酸 - アダプター複合体を配列決定することの 1 以上を含みうる。

**【 0 0 8 7 】**

サンプル（特に、細胞サンプルまたは組織生検）は身体の任意の部分または領域からのものでありうる。典型的なサンプルは、例えば血液、中枢神経系、脳、脊髄、骨髄、膵臓、甲状腺、胆嚢、肝臓、心臓、脾臓、結腸、直腸、肺、呼吸器系、咽喉、鼻腔、胃、食道、耳、眼、皮膚、四肢、子宮、前立腺、生殖器または身体の任意の他の器官または領域から得られうる。

**【 0 0 8 8 】**

一般に、サンプルはヒト対象、特にヒト患者からのものである。しかし、サンプルはまた、任意の他のタイプの対象、例えば任意の哺乳動物、非ヒト哺乳動物、非ヒト霊長類、飼育動物（例えば、実験動物、家庭用ペットまたは家畜）または非飼育動物（例えば、野生動物）からのものでありうる。幾つかの特定の実施形態においては、対象はイヌ、ネコ、げっ歯類、マウス、ハムスター、ウシ、トリ、ニワトリ、ブタ、ウマ、ヤギ、ヒツジ、ウサギ、類人猿、サルまたはチンパンジーである。

**【 0 0 8 9 】**

好ましい実施形態においては、対象は、病原体に感染している、または病原体による感染のリスクを有する、または病原体感染を有すると疑われる宿主生物（例えば、ヒト）である。幾つかの場合には、対象は、特定の感染を有することが疑われ、例えば、結核を有することが疑われる。他の場合においては、対象は未知起源の感染を有することが疑われる。幾つかの場合には、宿主または対象は（例えば、1以上の微生物、病原体、細菌、ウイルス、真菌または寄生生物に）感染している。幾つかの場合には、宿主または対象は、1以上の型の癌を有すると診断されており、または1以上の型の癌を発生するリスクを有する。幾つかの場合には、宿主または対象は（例えば、1以上の微生物、病原体、細菌、ウイルス、真菌または寄生生物に）感染していない。幾つかの場合には、宿主または対象は健康である。幾つかの場合には、宿主または対象は感染に感受性である、または感染の

10

**【0090】**

幾つかの場合には、対象は抗微生物剤、抗細菌剤、抗ウイルス剤または抗寄生生物剤で治療されていてもよく、または治療されうる。対象は（例えば、1以上の微生物、病原体、細菌、ウイルス、真菌または寄生生物による）実際の感染を有しうる。幾つかの場合には、対象は（例えば、1以上の微生物、病原体、細菌、ウイルス、真菌または寄生生物に）感染していない。幾つかの場合には、対象は健康である。幾つかの場合には、対象は感染に感受性であり、または感染のリスクを有する（例えば、患者は免疫無防備状態である）。対象は、別の疾患または障害を有する、またはそのリスクを有しうる。例えば、対象は、疾患、例えば癌（例えば、乳癌、肺癌、膵臓癌、血液癌など）を有しうる、またはその

20

**【0091】**

サンプルは核酸サンプルでありうる。幾つかの場合には、サンプルは或る量の核酸を含有する。サンプル中の核酸は二本鎖（ds）核酸、一本鎖（ss）核酸、DNA、RNA、cDNA、mRNA、cRNA、tRNA、リボソームRNA、dsDNA、ssDNA、miRNA、siRNA、循環核酸、循環無細胞核酸、循環DNA、循環RNA、無細胞核酸、無細胞DNA、無細胞RNA、循環無細胞DNA、無細胞dsDNA、無細胞ssDNA、循環無細胞RNA、ゲノムDNA、エキソソーム、無細胞病原体核酸、循環病原体核酸、ミトコンドリア核酸、非ミトコンドリア核酸、核DNA、核RNA、染色体DNA、循環腫瘍DNA、循環腫瘍RNA、環状核酸、環状DNA、環状RNA、環状一本鎖DNA、環状二本鎖DNA、プラスミドまたはそれらの任意の組合せを含みうる。幾つかの場合には、サンプル核酸は合成核酸を含みうる。ある場合においては、合成核酸は、本明細書に開示されている任意のタイプの核酸、例えばDNA、RNA、DNA-RNAハイブリッドを含む。例えば、合成核酸はDNAでありうる。

30

**【0092】**

幾つかの場合には、サンプル中には種々のタイプの核酸が存在しうる。例えば、サンプルは無細胞RNAおよび無細胞DNAを含みうる。同様に、本発明で提供する方法は、サンプル中に存在するRNAおよびDNAの両方を単独でまたは組合せて分析する方法を含みうる。

**【0093】**

本明細書中で用いる「無細胞」なる語は、体からサンプルが得られる前に体内に出現した核酸の状態を意味する。例えば、サンプルにおける循環無細胞核酸は、人体の血流中を循環する無細胞核酸に起源を有しうる。これとは対照的に、生検のような固体組織から抽出された核酸は一般に「無細胞」であるとはみなされない。

40

**【0094】**

幾つかの場合には、サンプルは、無細胞核酸または細胞結合核酸を含有する処理サンプル（例えば、血清、血漿）または未処理サンプル（例えば、全血）でありうる。幾つかの場合には、サンプルは、あるタイプの核酸、例えばDNA、RNA、無細胞DNA、無細胞RNA、無細胞循環DNA、無細胞循環RNAなどに関して富化されている。幾つかの場合には、サンプルは、核酸を単離するために、またはサンプルにおける他の成分から核

50

酸を分離するために、何らかの方法で処理されている。幾つかの場合には、サンプルは病原体特異的核酸に関して富化されている。

【0095】

しばしば、サンプルは新鮮なサンプルである。幾つかの場合には、サンプルは凍結サンプルである。幾つかの場合には、サンプルは、例えばホルマリン固定パラフィン包埋組織のように、例えば化学的固定剤で固定される。

【0096】

#### 標的核酸

本発明で提供する方法は、多数の標的核酸を検出するために使用されうる。標的核酸には、全ゲノムまたは部分ゲノム、エクソーム、遺伝子座、遺伝子、エキソン、イントロン、修飾核酸（例えば、メチル化核酸）、および/またはミトコンドリア核酸が含まれるが、これらに限定されるものではない。しばしば、本発明で提供する方法は、病原体標的核酸を検出するために使用されうる。幾つかの場合には、病原体標的核酸は、対象からの核酸を含有する複雑な臨床サンプル中に存在する。病原体標的核酸は、感染症、例えばインフルエンザ、結核または任意の他の公知の感染性の疾患または障害（本明細書に更に詳細に記載されているものを含む）に関連していることが可能である。幾つかの場合には、本明細書に記載されている標的核酸は標的核酸でありうる。

【0097】

幾つかの場合には、病原体標的核酸は、組織サンプル、例えば、感染部位からの組織サンプル中に存在する。他の場合においては、病原体標的核酸は感染部位から移動しており、例えば、それは、循環無細胞核酸（例えば、DNA）を含有するサンプルから得られうる。

【0098】

幾つかの場合には、標的核酸は癌組織に由来する。標的核酸は組織または腫瘍から直接得られうる。幾つかの場合には、標的癌核酸は循環無細胞核酸または循環腫瘍細胞（CTC）から得られる。

【0099】

幾つかの場合には、標的核酸は、サンプル全体の非常に小さな部分のみ、例えば、サンプルにおける全核酸の1%未満、0.5%未満、0.1%未満、0.01%未満、0.001%未満、0.0001%未満、0.00001%未満または0.0000001%未満を構成しうる。幾つかの場合には、標的核酸はサンプルにおける全核酸の約0.00001%~約0.5%を構成しうる。しばしば、元のサンプルにおける全核酸は変動しうる。例えば、全無細胞核酸（例えば、DNA、mRNA、RNA）は1~100ng/ml（例えば、約1、5、10、20、30、40、50、80、100ng/ml）の範囲でありうる。幾つかの場合には、サンプルにおける無細胞核酸の全濃度はこの範囲外である（例えば、1ng/ml未満；他の場合においては、全濃度は100ng/mlを超える）。これは、ヒトDNAおよび/またはRNAから主に構成される無細胞核酸（例えば、DNA）サンプルの場合に当てはまりうる。そのようなサンプルにおいては、病原体標的核酸または癌標的核酸は、例えば化学療法を受けている対象からのサンプルに関しては、ヒト核酸または健常核酸と比較して十分には存在しないかもしれない。例えば、病原体標的核酸はサンプルにおける全核酸の0.001%未満を構成することが可能であり、癌標的核酸はサンプルにおける全核酸の1%未満を構成することが可能である。

【0100】

標的核酸の長さは様々でありうる。幾つかの場合には、標的核酸は少なくとも約20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250、300、350、400、450、500、750、1000、1500、2000、3000、4000、5000、10000、15000、20000、25000または50000ヌクレオチド（または塩基対）長でありうる。幾つかの場合には、標的核酸は多くとも約20、30、40、50、60、70、80、90、100、110、120、130、140、15

10

20

30

40

50

0、160、170、180、190、200、250、300、350、400、450、500、750、1000、1500、2000、3000、4000、5000、10000、15000、20000、25000または50000ヌクレオチド（または塩基対）長でありうる。幾つかの特定の形態においては、標的核酸は比較的短く、例えば、500塩基対（またはヌクレオチド）長未満または1000塩基対（またはヌクレオチド）長未満である。幾つかの場合には、標的核酸は比較的長く、例えば、1000塩基対（またはヌクレオチド）長を超える、1500塩基対（またはヌクレオチド）長を超える、2000塩基対（またはヌクレオチド）長を超える、2500塩基対（またはヌクレオチド）長を超える、3000塩基対（またはヌクレオチド）長を超える、または5000塩基対（またはヌクレオチド）長を超える。幾つかの場合には、標的核酸は約20

10

#### 【0101】

サンプル核酸の場合と同様に、標的核酸は、二本鎖（ds）核酸、一本鎖（ss）核酸、DNA、RNA、cDNA、mRNA、cRNA、tRNA、リボソームRNA、dsDNA、ssDNA、miRNA、siRNA、循環核酸、循環無細胞核酸、循環DNA、循環RNA、無細胞核酸、無細胞DNA、無細胞RNA、循環無細胞DNA、無細胞dsDNA、無細胞ssDNA、循環無細胞RNA、ゲノムDNA、エキソソーム、無細胞病原体核酸、循環病原体核酸、ミトコンドリア核酸、非ミトコンドリア核酸、核DNA、核RNA、染色体DNA、循環腫瘍DNA、循環腫瘍RNA、環状核酸、環状DNA、環状RNA、環状一本鎖DNA、環状二本鎖DNA、プラスミドまたはそれらの任意の組合せを含む任意のタイプの核酸でありうる。標的核酸は、好ましくは、ウイルス、細菌、真菌、寄生生物および任意の他の微生物、特に感染性微生物（これらに限定されるものではない）を含む病原体に由来する核酸である。標的核酸は、特定の器官または組織に由来する核酸でありうる。幾つかの場合には、標的核酸は、病原体ではなく対象から直接的に誘導される。

20

#### 【0102】

##### スパイクイン（spike-in）合成核酸

本開示は、特にハイスループットまたは次世代シーケンシングアッセイに関連した種々の用途で使用される単一合成核酸および合成核酸のセットを記載する。幾つかの場合には、スパイクイン合成核酸は、記載されている方法で使用される場合、例えば、それが由来する個体、分析前サンプル処理条件、核酸抽出の方法、分子生物学的手段および方法による核酸操作、核酸精製の方法、測定自体の実施、保存条件および時間経過には無関係に、サンプルにわたる核酸（例えば、疾患特異的核酸、病原体核酸）の効率的な正規化を可能にしうる。幾つかの場合には、本開示は、例えば多数のユニーク配列のような特定の特徴を有する合成核酸のプールまたはセットを提供する。合成核酸のセットは、サンプル分析の経過中に多様性減少をモニターするために使用可能であり、そして該多様性減少は、出発核酸の存在量を決定するために使用されうる。本発明で提供する合成核酸はまた、サンプルを追跡するため、サンプル間の交差汚染をモニターするため、試薬を追跡するため、試薬ロットを追跡するため、および多数の他の用途に使用されうる。しばしば、合成核酸の設計、長さ、品質、濃度、多様性レベルおよび配列は個々の用途に適合化されうる。幾つかの場合には、スパイクイン合成核酸には、本明細書に記載されている担体合成核酸（例えば、担体合成核酸）が含まれる。

30

40

#### 【0103】

本発明で提供する合成核酸の集合体（コレクション）（またはセット）は幾つかの種の合成核酸を含みうる。幾つかの場合には、該種の長さ、濃度および/または配列は同じである、または類似していることが可能である。幾つかの場合には、該種の長さ、濃度および/または配列は異なってもよい。

#### 【0104】

好ましい実施形態においては、合成核酸の種は長さにおいて様々である。例えば、合成

50



核酸種の集合体は全体として、サンプルにおける或る標的核酸の長さの観察可能な範囲、またはそのような観察可能な範囲の少なくとも一部にわたりうる。例えば、該種は全体として、サンプル（特に、病原体に感染している又は感染していると疑われる対象から得られたサンプル）における疾患特異的または病原体特異的核酸の長さにわたりうる。幾つの場合には、サンプルにおける疾患特異的または病原体特異的核酸の長さは約40～約100塩基対の範囲でありうる。幾つの場合には、該種は全体として、サンプルにおける多種多様な疾患特異的または病原体特異的核酸の長さにわたりうる。幾つの場合には、該種は全体として、特定の病原体特異的核酸、例えば、特定の病原体ゲノム内の核酸の長さにわたりうる。幾つの場合には、該核酸は、病原体ゲノム内の特異的核酸、例えば、病原体のビルレンス領域内の核酸、病原体の抗生物質耐性領域、あるいは他の領域または特定の核酸もしくは遺伝子でありうる。幾つの場合には、長さまたは核酸は感染の個々のタイプ（例えば、急性、慢性、活動性または潜伏性）に特異的でありうる。他の例においては、該種は全体として、サンプル（例えば、感染対象由来のもの）における或る対象核酸および/または病原体核酸の長さにわたりうる。

10

## 【0105】

集合体内の合成核酸の種の長さは特定の標的核酸の長さ（例えば、サンプルにおける病原体特異的または疾患特異的核酸の観察可能な範囲）と厳密に一致しうる。他の場合においては、合成核酸の集合体内の合成核酸の種の長さは標的核酸の長さと同様に一致し、またはそのような長さに実質的に一致しうる。例えば、合成核酸の種の長さは標的核酸の長さの50%～150%、標的核酸の長さの55%～145%、標的核酸の長さの60%～140%、標的核酸の長さの65%～135%、標的核酸の長さの70%～130%、標的核酸の長さの75%～125%、標的核酸の長さの80%～120%、標的核酸の長さの85%～115%、標的核酸の長さの90%～110%、標的核酸の長さの95%～105%、標的核酸の長さの96%～104%、標的核酸の長さの99%～101%、または標的核酸の長さの99.5%～100.5%の範囲内である。幾つの場合には、合成核酸の種の長さは標的核酸の長さの50%～150%の範囲内でありうる。幾つの場合には、合成核酸の種の長さは標的核酸の長さの2倍、3倍、4倍または5倍まででありうる。幾つの場合には、合成核酸の種の長さは標的核酸の長さの1、2、3、4、5、10、20、30、40、50、60、70、80、90、100、125、150または200ヌクレオチド以内でありうる。幾つの場合には、集合体内の合成核酸の種は、最も厳密に一致した標的核酸の長さの65%、75%、80%、85%、90%、92%、95%、97%または99%より大きい。

20

30

## 【0106】

本明細書に開示されている合成核酸の集合体（またはプール）内のそれぞれ又はほとんどの核酸「種」は、関心のある1以上のドメインまたは領域を含みうる。幾つの場合には、関心のあるドメインまたは領域は長さ識別子配列である。長さ識別子配列は、特定の長さを示す又は表すと予め定められたコードを含有しうる。しばしば、そのような長さ識別子は短い配列であることが可能であり、例えば、10塩基対（bp）、9bp、8bp、7bp、6bp、5bp、4bpまたは3bp；9bp未満、8bp未満、7bp未満または6bp未満；あるいは6～15bp、5～10bp、4～8bp、または6～9bpでありうる。該種は1個、2個またはそれ以上の長さ識別子配列を含有しうる。幾つの場合には、長さ識別子はフォワードおよび/またはリバース配列として存在する。

40

## 【0107】

幾つの場合には、合成核酸の集合体内の核酸種におけるドメインは、存在する場合の該合成核酸における長さ識別配列によりコードされる長さに対応する特定の長さのロード（load）配列でありうる。スパイクイン核酸またはロードの長さは様々でありうる。幾つの場合には、スパイクイン核酸全体は少なくとも約20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250、300、350、400、450または500ヌクレオチド長でありうる。幾つの場合には、スパイクイン核酸は多くとも約20

50

、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250、300、350、400、450または500ヌクレオチド長でありうる。幾つかの場合には、スパイクイン核酸は約20～約200塩基対、例えば約20～約120塩基対の範囲でありうる。幾つかの場合には、スパイクイン核酸内のロード配列ドメインの長さは少なくとも約20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250、300、350、400、450または500ヌクレオチド長でありうる。幾つかの場合には、スパイクイン核酸内のロード配列ドメインの長さは多くとも約20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250、300、350、400、450または500ヌクレオチド長でありうる。幾つかの場合には、スパイクイン核酸内のロード配列ドメインの長さは0～約200bpの範囲内でありうる。

10

#### 【0108】

合成核酸の集合体内の核酸種におけるドメインは、該核酸が元のサンプルの一部ではなくスパイクインであることを示すユニークコードを含む合成核酸識別配列〔例えば、スパーク(Spark)識別配列、スパンク(Spank)識別配列〕でありうる。一般に、該ユニークコードは、元のサンプルに又は標的核酸のプールに存在しないコードである。該合成核酸識別配列は特定の数のbp、例えば25bp、20bp、19bp、18bp、16bp、15bp、12bp、10bpまたは他の長さを含みうる。該種は1個、2個またはそれ以上の合成核酸識別配列またはドメインを含有しうる。幾つかの場合には、該合成核酸識別配列はフォワードおよび/またはリバース配列として存在する。

20

#### 【0109】

幾つかの場合には、合成核酸の集合体内の核酸種におけるドメインは、合成核酸の全体的なプールまたは集合体に関連する「多様性コード」でありうる。多様性コードドメインは、合成核酸のプール内の多様性の量を示すユニークコードでありうる。そのような場合、該多様性プール内の各合成核酸は、該プールの多様性の度合(例えば、 $10^8$ 個のユニーク配列)を示す配列でコードされうる。幾つかの場合、例えば、2以上の多様性プールが同一サンプルに関して使用される場合には、該多様性コードは、それらの2以上のプールにおける多様性減少を特定するために使用されうる。

30

#### 【0110】

幾つかの場合には、合成核酸の集合体内の核酸種におけるドメインは、用途に応じて、サンプルまたは試薬の特徴の1以上に関連する特徴ドメインでありうる。例えば、該特徴ドメインは、特定の試薬、特定の試薬ロットまたは特定のサンプル(例えば、サンプル番号、患者番号、患者名、患者の年齢、患者の性別、患者の人種、サンプルが患者から得られた場所)を示すようにコードされた配列を含みうる。

#### 【0111】

関心のあるドメインまたは領域は任意の組合せおよび数で存在しうる。例えば、合成核酸は、1以上の長さ識別子配列、1以上のロード配列、1以上の合成核酸識別配列、1以上の多様性コードおよび/または1以上の特徴ドメインを任意の組合せまたは比で含みうる。例えば、幾つかの場合には、合成核酸は長さ識別子配列およびロード配列を含有する。幾つかの場合には、合成核酸は合成核酸識別子配列および特徴ドメイン配列を含む。幾つかの場合には、合成核酸は合成核酸識別子配列を含み、他の場合においては、それはそのような配列を含有しない。

40

#### 【0112】

幾つかの場合には、合成核酸は、重複する目的を有するドメインを含有しうる。例えば、幾つかの場合には、合成核酸は、ロード配列としても機能する1以上の長さ識別子配列を含有する。幾つかの場合には、長さ識別子配列および/またはロード配列は合成核酸識別子配列としても機能する。

#### 【0113】

50

合成またはスパイクイン核酸は、核酸ライブラリーに適合するように選択または設計される。幾つかの場合には、合成核酸またはスパイクインは、アダプター、共通配列、ランダム配列、ポリ(A)尾部、平滑末端もしくは不整(ragged)末端またはそれらの任意の組合せを含有しうる。幾つかの場合には、合成核酸またはスパイクインは、これらの又は他の特性の1以上において、サンプルにおける核酸を模倣するように設計される。

#### 【0114】

本発明で提供する合成核酸(例えば、スパイクイン合成核酸)は任意のタイプの核酸、または核酸タイプの組合せを含有しうる。好ましい実施形態においては、合成またはスパイクイン核酸はDNAである。幾つかの場合には、合成またはスパイクイン核酸は一本鎖DNAである。幾つかの場合には、合成またはスパイクイン核酸は二本鎖DNAである。幾つかの場合には、合成またはスパイクイン核酸はRNAである。幾つかの場合には、合成またはスパイクイン核酸は修飾塩基または人工塩基を含有しうる。二本鎖合成またはスパイクイン核酸は平滑末端または陥凹末端を有しうる。合成またはスパイクイン核酸はリン酸化または脱リン酸化末端を有しうる。幾つかの場合には、合成核酸は二本鎖(ds)核酸、一本鎖(ss)核酸、DNA、RNA、cDNA、mRNA、cRNA、tRNA、リボソームRNA、dsDNA、ssDNA、snRNA、ゲノムDNA、オリゴヌクレオチド、二本鎖オリゴヌクレオチド、より長い合体(assembled)二本鎖DNA(例えば、Integrated DNA TechnologiesのgBlocks)、プラスミド、PCR産物、インビトロ合成転写産物、ウイルス粒子、断片化または非断片化ゲノムDNA、環状核酸、環状DNA、環状RNA、環状一本鎖DNA、環状二本鎖DNA、プラスミドまたはそれらの任意の組合せを含有しうる。合成核酸は、しばしば、核酸塩基、例えばアデニン(A)、シトシン(C)、グアニン(G)、チミン(T)および/またはウラシル(U)を含みうる。

#### 【0115】

合成核酸は任意の合成核酸または核酸類似体であることが可能であり、あるいは任意の合成核酸または核酸類似体を含みうる。合成核酸は、修飾または改変リン酸骨格、修飾ペントース糖(例えば、修飾リボースまたはデオキシリボース)、あるいは修飾または改変核酸塩基(例えば、修飾アデニン(A)、シトシン(C)、グアニン(G)、チミン(T)、ウラシル(U))を含みうる。幾つかの場合には、合成核酸は1以上の修飾塩基、例えば5-メチルシトシン(m5C)、シュードウリジン( )、ジヒドロウリジン(D)、イノシン(I)および/または7-メチルグアノシン(m7G)を含みうる。幾つかの場合には、合成核酸はペプチド核酸(PNA)、架橋核酸(BNA)、類似核酸、グリセロール核酸(GNA)、トレオース核酸(TNA)、ロック核酸(LNA)、2'-O-メチル置換RNA、モルホリノ、またはヌクレオチド側鎖を有する他の合成ポリマーを含みうる。幾つかの場合には、合成核酸はDNA、RNA、PNA、LNA、BNAまたはそれらの任意の組合せを含みうる。幾つかの場合には、合成核酸は二重らせんまたは三重らせんまたは他の構造を含みうる。

#### 【0116】

合成核酸は任意のヌクレオチドの任意の組合せを含みうる。ヌクレオチドは天然物または合成物でありうる。幾つかの場合には、ヌクレオチドは酸化またはメチル化されていてもよい。ヌクレオチドには以下のものが含まれうる(それらに限定されるものではない): アデノシン一リン酸(アデノシンモノホスファート)(AMP)、アデノシン二リン酸(ADP)、アデノシン三リン酸(ATP)、グアノシン一リン酸(GMP)、グアノシン二リン酸(GDP)、グアノシン三リン酸(GTP)、チミジン一リン酸(UTP)、ウリジン二リン酸(UDP)、ウリジン三リン酸(UTP)、シチジン一リン酸(CMP)、シチジン二リン酸(CDP)、シチジン三リン酸(CTP)、5-メチルシチジン一リン酸、5-メチルシチジン二リン酸、5-メチルシチジン三リン酸、5-ヒドロキシメチルシチジン一リン酸、5-ヒドロキシメチルシチジン二リン酸、5-ヒドロキシメチルシチジン三リン酸、環状アデノシン一リン酸(cAMP)、環状グアノシン一リン酸(c

10

20

30

40

50

GMP)、デオキシアデノシンーリン酸(dAMP)、デオキシアデノシンニリン酸(dADP)、デオキシアデノシン三リン酸(dATP)、デオキシグアノシンーリン酸(dGMP)、デオキシグアノシンニリン酸(dGDP)、デオキシグアノシン三リン酸(dGTP)、デオキシチミジンーリン酸(DTMP)、デオキシチミジンニリン酸(dTDP)、デオキシチミジン三リン酸(dTTP)、デオキシウリジンーリン酸(dUMP)、デオキシウリジンニリン酸(dUDP)、デオキシウリジン三リン酸(dUTP)、デオキシシチジンーリン酸(dCMP)、デオキシシチジンニリン酸(dCDP)およびデオキシシチジン三リン酸(dCTP)、5-メチル-2'-デオキシシチジンーリン酸、5-メチル-2'-デオキシシチジンニリン酸、5-メチル-2'-デオキシシチジン三リン酸、5-ヒドロキシメチル-2'-デオキシシチジンーリン酸、5-ヒドロキシメチル-2'-デオキシシチジンニリン酸および5-ヒドロキシメチル-2'-デオキシシチジン三リン酸。

10

## 【0117】

合成またはスパイクイン核酸は、サンプルに添加される任意の分子を意味することが可能であり、例えばカラム上で化学合成された分子には限定されない。幾つかの場合には、合成またはスパイクイン核酸は、例えばPCR増幅、インビトロ転写、または鋳型に基づく他の複製により合成されうる。幾つかの場合には、合成またはスパイクイン核酸は、せん断または断片化された核酸であり、あるいは、せん断または断片化された核酸を含む。該せん断または断片化核酸はゲノム核酸、例えばヒトまたは病原体ゲノム核酸を含みうる。幾つかの場合には、合成核酸はヒト核酸を含有しない。幾つかの場合には、合成核酸は、天然で見出されうる核酸を含有しない。幾つかの場合には、合成核酸はサンプル核酸を含有しない。

20

## 【0118】

スパイクインまたは合成核酸のグアニン-シトシン含量(GC含量)は様々でありうる。幾つかの場合には、スパイクインまたは合成核酸のGC含量は少なくとも約0%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%または100%でありうる。幾つかの場合には、GC含量は多くとも約5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%または100%でありうる。幾つかの場合には、スパイクインまたは合成核酸のGC含量は約15%~約85%、例えば約20%~約80%の範囲内でありうる。集合体内の合成核酸の種のGC含量は特定の標的核酸(例えば、サンプルにおける病原体特異的または疾患特異的核酸の観察可能な範囲)のGC含量と厳密に一致しうる。他の場合においては、合成核酸の集合体内の合成核酸の種のGC含量は標的核酸のGC含量と厳密に一致し、またはそのようなGC含量に実質的に一致しうる。例えば、合成核酸の種のGC含量は標的核酸のGC含量の75%~125%、標的核酸のGC含量の80%~120%、標的核酸のGC含量の85%~115%、標的核酸のGC含量の90%~110%、標的核酸のGC含量の95%~105%、標的核酸のGC含有量の96%~104%、標的核酸のGC含量の99%~101%、または標的核酸のGC含有量の99.5%~100.5%の範囲内でありうる。

30

40

## 【0119】

スパイクイン核酸は、異なる分子、例えばビーズ、発蛍光団、ポリマーに結合され、連結され、またはコンジュゲート化されうる。発蛍光団の例には、蛍光タンパク質、緑色蛍光タンパク質(GFP)、Alexa色素、フルオレセイン、赤色蛍光タンパク質(RFP)および黄色蛍光タンパク質(YFP)が含まれるが、これらに限定されるものではない。スパイクイン核酸はタンパク質(例えば、ヒストン、核酸結合タンパク質、DNA結合タンパク質、RNA結合タンパク質)に結合されうる。他の場合においては、スパイクイン核酸はタンパク質に結合されない。スパイクイン核酸は粒子で保護されうる(例えば、ピリオン内の核酸に類似)。幾つかの場合には、スパイクイン核酸は粒子内に封入され、または粒子に結合している。幾つかの場合には、粒子はタンパク質、脂質、金属、金属

50

酸化物、プラスチック、ポリマー、生体高分子、セラミックまたは複合材料を含む。

【0120】

スパイクイン核酸は、サンプルまたは宿主において潜在的に見出される配列とは異なる配列を有しうる。幾つかの場合には、スパイクイン核酸配列は天然に存在する。幾つかの場合には、スパイクイン核酸配列は天然に存在しない。幾つかの場合には、スパイクイン核酸配列は宿主に由来する。幾つかの場合には、スパイクイン核酸配列は宿主に由来しない。幾つかの場合には、スパイクインまたは合成核酸は1以上の標的核酸（例えば、病原体核酸、疾患特異的核酸）および/または1以上のサンプル核酸にハイブリダイズし得ない（または相補的でない）。

【0121】

サンプルにおけるスパイクイン核酸の濃度は様々でありうる。スパイクインは、感度およびサンプル損失を決定するのに有用でありうる広範囲の濃度で添加されうる。幾つかの場合には、10万、50万、100万、200万、300万、400万、500万、600万、700万、800万、900万、1000万、2000万、3000万、4000万、5000万、6000万、7000万、8000万、9000万、1億、5億または10億個の、各スパイクイン核酸の分子が、血漿またはサンプルの1 mLあたりに添加される。幾つかの場合には、約1000万～約10億個の、各スパイクイン核酸の分子が、血漿またはサンプルの1 mLあたりに添加される。他の場合においては、合成核酸は、異なる濃度でサンプルにスパイクイン（添加）される。

【0122】

サンプルに添加される種々のスパイクイン核酸の数は様々でありうる。複数のスパイクイン核酸がサンプルまたは試薬に添加されうる。幾つかの場合には、少なくとも約1、2、3、4、5、6、7、8、9または10個のスパイクイン核酸がサンプルまたは試薬に添加される。幾つかの場合には、多くとも約1、2、3、4、5、6、7、8、9または10個のスパイクイン核酸がサンプルまたは試薬に添加される。幾つかの場合には、サンプルまたは試薬に添加されるスパイクイン核酸は、同じ長さである。幾つかの場合には、サンプルまたは試薬に添加されるスパイクイン核酸は、異なる長さである。幾つかの場合には、スパイクイン核酸は、配列番号1～120およびそれらの任意の組合せからなる群から選択される。

【0123】

スパイクイン核酸のユニーク性（一意性）のレベルは様々でありうる。本質的に無制限の数のスパイクイン（例えば、IDスパイク）が設計または使用されうる。

【0124】

スパイクイン核酸が添加されるプロセスにおける工程は様々でありうる。サンプルトラッキングのためには、スパイクイン核酸の、より早期の添加がより良好であり、その後のオペレータまたはシステムエラーの可能性を減少させうる。幾つかの場合には、サンプル（例えば、血液）が最初に添加されるチューブは既にスパイクイン核酸を含有していてもよい。これらのチューブの製造は、診療所または研究検査施設におけるサンプルへのスパイクイン核酸の添加と比較して、より体系的に制御および試験されることが可能であり、それにより、サンプルの混同の可能性が低減されうる。幾つかの場合には、IDスパイクが全ての外部ラベル（「ホワイトラベル」）に取って代わりうる。

【0125】

幾つかの場合には、各配列リードが識別マーカ含有するように、識別核酸マーカがサンプルにおける各核酸断片に加えられうる。この方法は初期のものと下流のものとの交差汚染の識別を可能にするであろう。断片のタグ付けが十分に完了している場合、それは、バーコードがサンプル断片に加えられるとすぐに、サンプルの意図的な多重化をも可能にしうるであろう。タグを組み込むための方法には、トランスポゾン、末端トランスフェラーゼ、メチル化部位での切断、および脱メチル化部位での切断が含まれるが、これらに限定されるものではない。

【0126】

10

20

30

40

50

プロセス品質管理または開発作業に関連する用途（これらに限定されるものではない）を含む他の用途の場合、スパイクイン核酸は該プロセスにおける種々の工程で添加される。例えば、RNA分析の場合には、異なる濃度、長さ、配列および/またはGC含量をそれぞれが有する複数のRNAスパイクインがサンプル調製の開始時に添加可能であり、RNAがDNAに変換された後、DNAスパイクインが添加可能である。DNAライブラリーの場合には、種々の形態のDNAがライブラリー作製プロセスの種々の工程で添加される。例えば、末端修復工程を試験するために、非平滑末端を有する、5'-リン酸を有する又は有さない（+/- 5'-リン酸）、および3'-アデニン伸長を有する又は有さない（+/- 3'-アデニン伸長）DNAスパイクインが使用される。アダプターを末端修復断片に連結する工程を試験するために、予め適合化された又はされていない（+/- pre-adapted）スパイクインが使用される。配列決定qPCRは個々の工程におけるサンプル損失を定量しうる。スパイクインのqPCRはまた、配列決定前の最終的なライブラリー評価のための他のライブラリー定量法と併用される。

10

#### 【0127】

「スパイクイン」、「スパイクイン合成核酸」、「スパイク」および「合成核酸」なる語は本明細書においては互換的に用いられ、異なる解釈を文脈が示す場合を除き、そのようなものとして解釈されるべきである。「IDスパイク」または「トレーサー」なる語は、例えばサンプル識別トラッキング、交差汚染検出、試薬トラッキングまたは試薬ロットトラッキングに使用される識別スパイクを意味するものとして本明細書において一般に用いられる。「スパーク（Spark）」なる語は、存在量の正規化、開発および/または分析の目的ならびに他の目的で使用される、サイズまたは長さマーカーである核酸を意味するものとして本明細書において一般に用いられる。「スパンク（Spank）」なる語は、縮重プール、または多様な配列を有する核酸のプールを意味するものとして一般に用いられ、しばしば、多様性評価および存在量の計算のために用いられる。

20

#### 【0128】

##### 核酸の測定結果の一般的な正規化

本開示は、記載されている方法において使用される場合、サンプルにおける疾患特異的核酸、病原体特異的核酸または他の標的核酸の量の効率的かつ改善された正規化を可能にしうる合成核酸のセットを記載する。添加（spiked）核酸のセットは、長さにおいて異なる核酸の幾つかの「種」を含有することが可能であり、その結果、添加核酸種の集合体は全体として、測定される病原体核酸、疾患特異的核酸または他の標的核酸の長さの観察可能な範囲にわたる。

30

#### 【0129】

スパイクイン合成核酸は、多種多様な方法でサンプルを正規化するために使用される。しばしば、正規化は、サンプルが由来する対象、分析前サンプル処理条件、核酸抽出の方法、分子生物学的手段および方法による核酸操作、核酸精製の方法、測定自体の実施、保存条件ならびに/または時間経過には無関係に、サンプル全体にわたりうる。

#### 【0130】

幾つかの好ましい実施形態では、スパイクイン核酸は、疾患特異的核酸、病原体特異的核酸または他の標的核酸を測定する全ての方法および全てのサンプルにわたって正規化する。幾つかの場合には、スパイクインは、サンプルにおける病原体核酸（または疾患特異的核酸または標的核酸）の、他の病原体核酸と比較した場合の相対的存在量を決定するために使用される。

40

#### 【0131】

一般に、本発明で提供する方法は、合成核酸のセットの1以上をサンプル内に添加（スパイクイン）または導入することを含む。この添加工程は、プロセスの初期、中期または終期を含む、該方法全体の任意の時点で行われうる。例えば、合成核酸は、対象からサンプルが採取された時点またはその直後、サンプルの保存の前または途中、サンプルの輸送の前、核酸抽出の前または途中、ライブラリー調製の前または途中、配列決定アッセイの直前、あるいは該方法の任意の他の工程において導入される。幾つかの場合には、該方

50

法は、同じ方法により測定されうるが病原体特異的もしくは疾患特異的核酸または他のサンプル核酸から容易に識別される既知量のユニーク核酸分子を該プロセスの初期に生物学的サンプルに添加することを含みうる。幾つかの場合には、該プロセスの単一工程（例えば、サンプルを対象から採取したとき、分析を行うためにサンプルを得たとき、サンプル保存中、核酸抽出の前または途中、ライブラリー調製の前または途中、あるいは配列決定アッセイの直前）において合成核酸を生物学的サンプルに添加する。他の場合においては、同じ又は異なるスパイクイン合成核酸を該プロセスにおける異なる工程において導入する。例えば、ユニーク合成核酸を該プロセスの初期（例えば、サンプル採取時）に導入することが可能であり、異なるセットのユニーク核酸を、該プロセスの、より後の時点（例えば、抽出、精製またはライブラリー調製の前または後）で導入することが可能である。また、スパイクイン核酸の同一集合体または何らかの態様で異なる集合体を使用して、該方法の異なる工程で添加工程を繰り返すことも可能である。

10

**【 0 1 3 2 】**

一般に、既知濃度（または複数濃度）の合成核酸の種を各サンプルに添加することが可能である。多くの場合においては、合成核酸の種を各種の等モル濃度で添加する。幾つかの場合には、合成核酸種の濃度は異なる。

**【 0 1 3 3 】**

サンプルが処理され、最終的に測定される場合、サンプルの取り扱い、調製および測定固有の偏りゆえに、核酸種の相対的存在量は変化しうる。測定後、添加核酸の各「種」の測定存在量を最初の添加量と比較することにより、各長さの核酸の回収効率が決定されうる。これは「長さに基づく回復プロファイル」を与えうる。

20

**【 0 1 3 4 】**

「長さに基づく回復プロファイル」を用いて、最も近い長さの添加分子に対して、または種々の長さの添加分子にフィットされた関数に対して、疾患特異的核酸の存在量（または病原体核酸もしくは他の標的核酸の存在量）を正規化することにより、疾患特異的核酸、病原体核酸または他の標的核酸の全て（または大部分または幾つか）の存在量を正規化することが可能である。このプロセスは疾患特異的核酸に適用可能であり、サンプル添加の時点の「全ての疾患特異的核酸の元の長さの分布」の推定をもたらしうる。同様に、このプロセスは他の標的核酸、例えば病原体特異的核酸にも適用可能であり、サンプル添加の時点の「全ての病原体特異的核酸の元の長さの分布」の推定をもたらしうる。「全ての標的核酸の元の長さの分布」はサンプル添加の時点の標的核酸（例えば、疾患特異的核酸、病原体特異的核酸）に関する長さの分布プロファイルを示しうる。完璧またはほぼ完璧な存在量の正規化を達成するために添加核酸が再現（*recapitulate*）しようとするのは、この長さの分布である。

30

**【 0 1 3 5 】**

特定のサンプルにおける疾患特異的核酸、病原体核酸または他の標的核酸の相対的存在量プロファイルを厳密に再現する既知核酸の混合物をサンプルに添加することは可能でない可能性があるため（これは1つには、該サンプルが使い果たされた、または時間が相対的存在量プロファイルを変化させた可能性があるからである）、スパイクインの各「種」は「全ての疾患特異的核酸の元の長さの分布」におけるその相対的存在量に比例して加重（*weight*）されうる。全ての「加重係数（*weighting factor*）」の和は1.0に等しいことが可能である。

40

**【 0 1 3 6 】**

正規化は単一工程または一連の工程を含みうる。幾つかの場合には、最も近いサイズの添加核酸の存在量の生測定値を用いて、病原体特異的核酸（または病原体核酸もしくは他の標的核酸）の存在量を正規化して、「正規化疾患特異的核酸（または病原体核酸もしくは他の標的核酸）存在量」を得ることが可能である。ついで「正規化疾患特異的核酸存在量」（または病原体核酸もしくは他の標的核酸存在量）に「加重係数」を掛け算して、その長さの回収の相対的重要性に関して補正して、「加重正規化疾患特異的（または病原体特異的もしくは他の標的）核酸存在量」を得ることが可能である。この正規化方法の利点

50

の1つは、それが、疾患特異的核酸の存在量を測定する全て（またはほとんど）の方法にわたって、方法には無関係に、標的核酸（例えば、疾患特異的核酸、病原体核酸）の存在量の同等の測定を可能にすることであろう。

【0137】

標的核酸存在量または相対的存在量の測定は、検出、予測、モニターおよび診断アッセイに特に有用でありうる。そのようなアッセイは、病原体の存在を検出するために、または病態を特定するために、生物学的サンプル（例えば、血漿）における標的核酸（例えば、疾患特異的核酸）の量を測定することを含みうる。本明細書に記載されている方法は、これらの測定を、サンプル、測定時間、核酸抽出方法、核酸操作方法、核酸測定方法および/または種々のサンプル処理条件の全体にわたって同等にしうる。

10

【0138】

添加分子の厳密な配列、「種」の厳密な数、「種」の長さの範囲、添加分子の濃度、各分子の相対量、各添加分子の実際の量、分子が添加される段階は、サンプルに基づいて最適化または調整されうる。長さはGC含量、核酸構造、DNA損傷またはDNA修飾状態で置換または分析されうる。

【0139】

幾つかの場合には、本発明で提供する方法は、（幾つかの方法における幾つかの短いランダム化部分を除き）大部分が固定された配列組成をしばしば伴う、核酸の単一の長さを含む添加核酸の使用を含みうる。この方法は、疾患特異的核酸、病原体特異的核酸または他の標的核酸が添加核酸とほぼ同じ長さのものである場合に、良好に機能しうる。

20

【0140】

単一の長さの核酸は単独で使用可能であり、あるいは該方法は、複数の長さの核酸の使用を伴う別の方法と組合せられうる。例えば、サンプルが得られた際または核酸の抽出の前に、複数の長さの核酸のプールをサンプルに添加することが可能であり、単一の長さの核酸のプールを該プロセスにおける異なる時点で（例えば、核酸の抽出の後およびライブラリー調製の前に）サンプルに添加することが可能である。単一の長さおよび/または複数の長さの核酸を使用する場合、疾患特異的核酸、病原体核酸または他の標的核酸の量を、該方法の終了時の測定された添加核酸の量に対して正規化することが可能である。

【0141】

多くの場合、本明細書に記載されているとおり、複数の長さを有する合成核酸の使用は、単一の長さの合成核酸の使用を含む方法より好ましいかもしれない。本発明で提供する核酸は、標的核酸が複数の長さを有する場合に特に有用である。例えば、疾患特異的（または病原体特異的）核酸は長さにおいて広範囲に変動しうる。したがって、疾患特異的核酸の観察可能な長さにわたるスパイクイン核酸の使用が特に役立つ。更に、測定された疾患特異的核酸の長さは、それが由来する個体の代謝、分析前的サンプル処理条件、核酸抽出の方法、分子生物学手段および方法による核酸操作、核酸精製の方法、測定自体の実施、保存条件および時間経過を含む多数の要因によっても劇的な影響を受けうる。これらの要因は、異なる長さの核酸に対して差動的な影響を及ぼし、したがって、単一の添加核酸は、混合した長さの核酸に関して行われるプロセスの全体的な効率を適切に反映しない可能性がある。

30

40

【0142】

「体積当たりのゲノムコピー数」の計算

本発明で提供する方法および合成核酸は、次世代シーケンシングの結果からサンプルにおける微生物または病原体の体積当たりのゲノムコピー数を決定することを含む或る計算を助けるために使用されうる。一般に、体積当たりのゲノムコピー数は流体（例えば、血漿、尿、バッファーなど）1ml当たりの標的核酸（例えば、特定の病原体に由来する標的核酸）の量の絶対的尺度を示すことが可能であり、個々の病原体の存在量または相対的存在量を示すための表現としてしばしば用いられうる。病原体の存在量のリード（read；読取り）の総数および/または大きさはサンプルごとに異なりうる。感染の生物学的レベルに対応し、サンプルとサンプルとの比較に有用でありうる値を報告することが望ま

50



しいことがある。

【0143】

特定の例においては、該方法は、サンプル（特に、病原体に感染している又は病原体に感染していると疑われる対象から得られたサンプル）における病原体核酸の体積当たりのゲノムコピー数を決定するために使用されうる。体積当たりのゲノムコピー数は、統計的枠組みを使用して決定または推定されうる。該統計的枠組みは、サンプルからの配列決定結果における非ヒトリード（例えば、病原体リード）の集合体（コレクション）を与える1以上のゲノムの相対的存在量を推定するために使用されうる。

【0144】

本発明で提供するスパイクイン合成核酸を使用して、サンプルにおける1以上の病原体/生物の「体積当たりのゲノムコピー数」の推定値が計算されうる。一般に、種々の長さの核酸が既知濃度でサンプルに添加されうる。幾つかの場合には、配列データにおいて実際に観察されるサンプルからの情報の割合が、（例えば、観察されたリード数を、添加核酸に関連するリード数と比較することにより、または観察されたリード数を添加リード数で割り算することにより）各スパイクイン長に関して観察されうる。各長さにおける非宿主または病原体分子の元の数を逆算することも可能である（例えば、各長さにおけるスパイクインリードの数から部分的に推測される）。各病原体のゲノムの長さが知られている場合、このロード（load）は「体積当たりのゲノムコピー数」の尺度に変換されうる。

【0145】

多くの場合、体積当たりのゲノムコピーを検出するための方法（および本発明で提供する他の方法）は低品質のリードの除去または隔離を含みうる。低品質のリードの除去は、本発明で提供する方法の精度および信頼性を改善しうる。幾つかの場合には、該方法は、マッピング不可能なリード、PCRデュプリケート（duplicate）から得られたリード、低品質リード、アダプター二量体リード、配列決定アダプターリード、非ユニークマッピングリードおよび/または情報価値のない配列に位置するリードの（任意の組合せでの）除去または隔離を含みうる。

【0146】

幾つかの場合には、配列リードは参照ゲノムに対してマッピングされ、そのような参照ゲノムにマッピングされていないリードが1以上の標的または病原体ゲノムにマッピングされる。幾つかの例においては、リードはヒト参照ゲノム（例えば、hg19）にマッピングされ、一方、残りのリードはウイルス、細菌、真菌および他の真核生物病原体（例えば、真菌、原虫、寄生生物）のキュレート参照データベースに対してマッピングされる。

【0147】

幾つかの特定の例においては、該方法は、DNA抽出（例えば、無細胞DNA抽出、無細胞RNA抽出）の前またはアッセイの異なる段階（例えば、抽出後、ライブラリー調製前、配列決定前、サンプルの保存中）において、既知濃度の合成核酸（例えば、DNA）をサンプル（例えば、血漿サンプル）に添加することを含みうる。合成核酸は陰性および/または陽性対照サンプルにも添加されうる。幾つかの場合には、対照サンプルはサンプルと並行して処理されうる。該方法は更に、サンプル（例えば、血漿サンプル、陽性対照、陰性対照）のための配列決定ライブラリーを製造することを含みうる。該ライブラリーは、当技術分野で公知の配列決定装置、特に、次世代シーケンシングの性能を有する装置において多重化され、配列決定されうる。該方法は更に、低品質のリードを破棄すること、およびヒト参照配列（例えば、hg19）に対するアライメントによりヒトリードを除去することを含みうる。ついで、残りのリードは病原体配列のデータベースにアライメントされうる。幾つかの場合には、関心のある標的配列（例えば、病原体配列）に対応するリードはNGSリードセットから定量される。この情報から、標的核酸（例えば、病原体核酸）の相対的存在量が体積当たりのゲノムコピー数として表されうる。体積当たりのゲノムコピー数は、例えば、サンプル（例えば、血漿）に添加された既知量のオリゴヌクレオチドに対して正規化された各生物（例えば、病原体）に関する存在配列の数を決定する

10

20

30

40

50

ことにより決定されうる。体積当たりのゲノム数の計算は個々の病原体ゲノムの相対的長さをも考慮しうる。幾つかの場合には、体積当たりのゲノムコピー数は、各生物（例えば、病原体）に関する存在配列の数を定量し、サンプルに添加された既知量の合成核酸に対して正規化することにより決定可能であり、ここで、病原体配列の正規化は、長さにおいて病原体配列に最も近い合成核酸を考慮する。同様に、正規化は、種々の長さ（例えば、2、3、4、5、6、10、15、20個またはそれ以上の異なる長さ）のスパイクイン合成核酸の集合体の使用を含むことが可能であり、ここで、病原体核酸は、スパイクインの集合体内の、長さにおいて最も近いそれぞれのスパイクイン核酸に対して正規化される。

#### 【0148】

##### サンプルの追跡（トラッキング）および/または分析のためのスパイクイン

分子をサンプルに添加して、ユニーク（unique；一意の、特有の）識別子およびトレーサーを得ることが可能である。これらの分子はサンプルの一部となることが可能であり、適切な測定装置により読取られうる（レーザースキャナーにより読取られるサンプルチューブの外表面上の1Dまたは2Dバーコードに類似した概念）。光学的、放射性および他のトレーサーが可能であるが、核酸サンプルの分析のためには、核酸トレーサーが最適な選択肢でありうる。なぜなら、スパイクインが何であるかが、サンプルの核酸を評価する同じプロセス（例えば、DNAまたはRNA配列決定）において示されうるからである。

#### 【0149】

外部由来の核酸には、オリゴヌクレオチド、二本鎖オリゴヌクレオチド、より長い合体（assembled）二本鎖DNA（例えば、Integrated DNA TechnologiesのgBlocks）、プラスミド、PCR産物、インビトロ合成転写産物、ウイルス粒子、および断片化または非断片化ゲノムDNA（これらに限定されるものではない）が含まれ、それらはサンプル、例えば対象からの体液に添加されうる。スパイクインを使用する利点には、核酸配列、長さ、多様性および濃度がサンプルまたは用途に適合化されうるということが含まれるが、これに限定されるものではない。

#### 【0150】

用途には以下のものが含まれるが、それらに限定されるものではない：サンプルの追跡（トラッキング）[例えば、通常のラベルバーコードに加えて、または潜在的にその代わりに、IDスパイク（Spike）が使用されうる]、サンプル交差汚染（例えば、サンプルのいずれにおいてもIDスパイクが天然で見出されない場合、およびサンプルによって異なるIDスパイクがサンプルに添加される場合には、サンプルの混合が判定されうる）、試薬の追跡[例えば、IDスパイクは試薬にも添加されうる。例えば、各試薬ロットは、それが使用される各サンプルに関して追跡可能であり、エラーのより少ない試薬追跡分子実験室情報管理システム（LIMS）がもたらされうる]、品質管理または開発作業[例えば、ライブラリーの複雑性（例えば、PCRデュープリケート（duplicate））、サンプル損失または感度をモニターするために、異なるスパイクインがサンプル処理プロセスにおける種々の時点で添加されうる]、正規化または収率[例えば、既知入力をスパイクインの測定出力と比較することは、（例えばサンプルにおける）未知入力、その測定出力による推測を可能にしうる。これらの測定および計算から、例えばサンプルの病原体量が判明しうる]、および核酸濃度の増加（例えば、バーコードが核酸である場合、それは、限定的である核酸濃度を有するサンプルに関して高濃度で使用可能であり、このことはサンプルの回収を改善しうる）。

#### 【0151】

幾つかの好ましい実施形態においては、スパイクインは、関心のある特定の核酸配列が、それが観察されたサンプルに由来する可能性、または観察されたサンプルにおけるその存在が、異なるサンプルからの交差汚染もしくは持ち越し汚染の結果でありうるかどうかを推定するために使用されうる。ユニークスパイクイン分子を、特定の病原体からの分子（または関心のある他の配列クラス）から合理的に予想される濃度より高い濃度で各サン

10

20

30

40

50

プル内に導入することにより、交差汚染または持ち越し汚染によって偶然に導入されたいずれかの病原体配列（または関心のある他の配列クラス）が、該汚染または持ち越し汚染配列の起源からの、より一層多数のスライクイン分子を伴う可能性が高い。したがって、交差汚染または持ち越し汚染スライクイン分子数に対する病原体配列数（または他のクラスの配列）の比率を用いて、サンプル間の交差汚染または持ち越し汚染の結果でありうる任意の病原体配列を特定することが可能である。幾つかの場合には、交差汚染または持ち越し汚染スライクイン分子の非存在、または閾値レベル未満のレベルでのその存在を利用して、サンプルが汚染されていないことを示す。

#### 【 0 1 5 2 】

幾つかの用途には、サンプルが由来する対象の遺伝子型を、特にサンプル追跡のために使用することが可能である。幾つかの場合には、遺伝子型は分析操作中に決定可能であり、あるいはアリコートを取り出し、別個の遺伝子型決定法を行うことにより決定可能である。幾つかの場合には、サンプルの遺伝子型は既に知られている。対象のDNAの配列決定出力を、独立して得られた遺伝子型と比較することが可能である。遺伝子型を使用する利点は、それが既にサンプルの一部であり、サンプルに固有のものであることである。典型的な直交性（orthogonal）遺伝子型決定法は、ショート・タンデム・リピート（STR）分析である。例えば、ATCCの試験サービスを参照されたい。

10

#### 【 0 1 5 3 】

幾つかの場合には、表現型の特徴がサンプルの特定を助けうる。例えば、対象の眼の色、血液型、性別、人種および他の形質が遺伝子型の手がかりをもたらさうであろう。

20

#### 【 0 1 5 4 】

##### IDスライク

ユニークサンプル識別子は完全にスクランブルされることが可能であり（例えば、DNAではA、C、GおよびTのランダム化、またはRNAではA、C、GおよびUのランダム化）、あるいはそれらは共通配列の幾つかの領域を有することが可能である。例えば、各末端の共有領域は連結事象における配列の偏りを減少させうる。幾つかの場合には、共有領域は少なくとも約1、2、3、4、5、6、7、8、9、10、15または20個の共通塩基対である。幾つかの場合には、共有領域は多くとも約1、2、3、4、5、6、7、8、9、10、15または20個の共通塩基対である。典型的な配列は、表1を参照されたい。

30

#### 【 0 1 5 5 】

IDスライクの組合せは、膨大な数のIDスライクを使用しなくても多様性を増大させるために添加されうる。IDスライクはマイクロタイタープレートにおけるウェル位置に関する識別子として使用可能であり（例えば、96ウェルプレートの場合には96個の異なるIDスライク）、別のIDスライクはプレート番号に関する識別子として使用可能であり（例えば、24個の異なるプレートの場合には24個の異なるIDスライク）、僅か $96 + 24 = 120$ の配列を使用して、 $96 \times 24 = 2,304$ の組合せが得られる。サンプル当たり3個以上のIDスライクの使用は、達成可能な多様性を、より一層劇的に増大させうる。

【表 1】

表1: 典型的なIDスパイク配列

配列番号	名称	配列
配列番号 1	ID_028_100bp	ACGTTGTCTGCGAGTCAGCTAATCCTATCCTGGTGCAT GCTTGACTTGTCACCGAGGTAATCATCGAATCCTGGAT GAGGACGCAAGAGATGTATGGTCA
配列番号 2	ID_029_100bp	ACGTCGAATCCTACGCGACTGCGCGTGACTAGGTGAA GGTGAACCTTAGAGGCTCTCAACCTCATCCACTCGGTAT CATCCTGTGTGTATCAAGAGAGTCA
配列番号 3	ID_030_100bp	ACGTGAGTGAGAGTACTCAATCAATCTTCTCCGCACCG TGAATGCGTGAGTGTGGCCAATGCCGCCATCAACGATT CTACTGAGCGAGTGCTCGCAGTCA
配列番号 4	ID_031_100bp	ACGTTGGTCTCAATGCCTGTGACGGACATAACGCATTA AGGACCGATATGGTAGAACTGTTCCCTCATGTGACTAGG AGGTAATCCTGGCGCATAACGTCA
配列番号 5	ID_032_100bp	ACGTACAACCTATGGTATGTCCACTGAGCGGCAACCAG GTTCTCATCATCGCTGCGGAGAAGTCACGTAATATTCT GAAGGTAGTGGCGTGTAGACGGTCA
配列番号 6	ID_033_100bp	ACGTCGTATGCAACGTATATGGAATTATCTGTGTACGT GCATACGTGACCAACAACCAGACGGCACCGATCATCT TAGTCGCCGAGAGATCTAATTGTCA
配列番号 7	ID_034_100bp	ACGTACGGCATTGTTCTCAGGAACGTGTGTTTCATACGA TCTTCGACTCTAGCATATCCAACGTCTGAAGTTATCATT

10

20

30

		ACCGAGCCGGAACAACGTCGGTCA	
配列番号 8	ID_035_100bp	ACGTACCGTTAAGTGCTGTTGAAGAATATGAGTCTATA GGTTCCGGACCTGTTGCGACGTGCGATGGCTACTTACG CATTAAACCAGTGTTGTATAGGTCA	
配列番号 9	ID_036_100bp	ACGTTGACCAAGAGGACGATACTTGGATAAGTTCCTTGC TATATGAGCGCGGTCAACGTGAAGGCCGATATCGCGC CAATCTTCACGATGGAGCTTCGTCA	
配列番号 10	ID_037_100bp	ACGTCCTGAATCATCGACTGAACCGGCTTACTAGGAAT TAGTCAGCGCATAATACTCCAATGAATGTTCTGAGC TCGACGCTCTTGCCGGAGTGGTCA	10
配列番号 11	ID_038_100bp	ACGTTCTGTCTATACGTATATGCTAAGCGCAATATGAT TCAATGGTAGTCACTTCGTCCGACGGCGAGTAACGCAC CACGTGTCCATTCTCTGGAGGTCA	
配列番号 12	ID_039_100bp	ACGTTACTGGTGACGGATGCCTCTCCTGACCGAGTATC TAGGCACCACACGGACGTTGACAGGCATGCTTAATCC GGTGATGAACGGATCGGTCCTGTCA	20
配列番号 13	ID_040_100bp	ACGTGAGGCGTCTCGCGATGTCAAGTGAATATGATGC AGTTCTGACTGCACACCACAGAGGTCGCTTGAAGTCG ATGACTCCGCATAGTCATTCAGTCA	
配列番号 14	ID_041_100bp	ACGTGCGACGTATGACGTGCGAGGTTAGAGTTGGCCT ACTACTTAGATAATCTTGCCGCATGCTTATTAAGTAGA ACGGTTGCCGAATTGCACCTGGTCA	
配列番号 15	ID_042_100bp	ACGTCACCTATGCACGACTCAACTTGAGGTAAGCGTGT AGAGAAGCCTTGATAGTCTTGTCGTCGCGGCCGGAGAT CTCCGTTACCTTCACACTTGGGTCA	30
配列番号 16	ID_043_100bp	ACGTTCCGATAGATAAGCACAAAGTCAGCGAAGCCTTG TCCGCTGTGTAACATATATCATCGAATGCGATAGTCGG TTGGAACGCGATCTGACGTACGTCA	
配列番号 17	ID_044_100bp	ACGTCCAGCATTCAATAAGCATCGACTCTTAGTGCGGT GTGGAGTGTACTTCCATGTGACAACCTCGAGTAGACCG ATTAGGCTGCTGACAAGTTAGGTCA	40
配列番号	ID_045_100bp	ACGTGGAATAGAGATGGTAAGCCACCGGCTTCGGTGC	

18		CTTGGCACGTGAGACCTATAGCTTAGTGCTCATCCATC GTTATCTACGACGAACATCGCGTCA	
配列番号 19	ID_046_100bp	ACGTGCCGCGTGTGTGTACCGTACAGTGTCTAGAAG TCTATTGGATTGGCTAGCTAGCGAGATTACCGACGCTA CTCAACAGAGTGGAGCTCATGTCA	
配列番号 20	ID_047_100bp	ACGTAAGTAGCGATCGCAGAGGCTAAGGCGTACAGTT CGTGGTTCGCGTCCACGTAGGCCGTTATCTGCTTACTT CCGTAGGTACTGAGAGATTCAGTCA	10
配列番号 21	ID_048_100bp	ACGTAGGCCGGATGTGACTGATGATTCATCTCAGCAGC ATAGCCTTCACGTGTAGCGGCTTATCACGCTGAGTTCG TTCCAACAGCTGGATAGTAGGTCA	
配列番号 22	ID_049_100bp	ACGTTATGTACACGTGAAGCTGGCGCGTGCGCGCTACT TATATTACGGAGGTTAGTTCATCAATATCAGAGAAGTT CCTCATGTATACATGAATTCGTC	
配列番号 23	ID_050_100bp	ACGTCTGGTACAGACGGATGCTGCTTGTCCAATTATGG TGACTAACTCCTAATCATCAGCAATCGCGAGTCCGGTA AGCGTCCGCGACTGATCGCCGTCA	20
配列番号 24	ID_051_100bp	ACGTGTTATACCGTACCGGTAGGAACACAAGTGTACCT GGTGAGATAGGTTCCATGCTTCTGGCGACCTGACAATC TACAGCAGACTAGTATGCGGGTCA	
配列番号 25	ID_052_100bp	ACGTTGCGGCCAGGACAATCATATCGCAATGATCATG ACTGTGATTCAACGGACTCGCGTCATGAGATGAATCTG CCAGAGTATGTCGGCTCTGGAGTCA	30
配列番号 26	ID_053_100bp	ACGTTGAGAGATAATTGAGCGATACTGTTATCGCGGCT ATACATAGCTCTCACCTCATGGACAGCGTAGGATTGTG AATAGCTGCTCCACTCGTCCGTCA	
配列番号 27	ID_054_100bp	ACGTACGGAAGACGGCAAGTTCTTGATCTTCACGCATC TGCCGCTATTGCAATATGTGGTATGACGTGATAACTAG CCGGCGTCCGACGTAACATGGTCA	
配列番号 28	ID_055_100bp	ACGTCATCGATCGTAGGCCGCTTCCGGCAAGGACCTAGT AGACTTCCAGATTGGACTTCTACGGCTCGGATATTATA CCGTCTCAAGGAACGGTGCTGTCA	40

配列番号 29	ID_056_100bp	ACGTCCTGCGCCTTAAGGCGCATCGCTTGTTACCAGGA TTAAGGATACGTCGTCGAATGCGCAGGTCTAGGATTGT TGTTGCTATAGAACTAATAACGTCA
配列番号 30	ID_057_100bp	ACGTCAGGCGTACATCGAACTGCTTACTCTCACGTCTG TCCGTATCAGGTTGACCTAACGCGTCTGGTGGCAAGCT AATACTGAGCATAGCAGTAGGTCA
配列番号 31	ID_058_100bp	ACGTTACGCAATGTGTCATACGCGCATCTGAGTTCAGA CTGCAATCACGCTGACGCACGACTATATGGTTGCCTCC GAGTACTTGGTACTAGGTCCGGTCA
配列番号 32	ID_059_100bp	ACGTTTGACAGCGCGCTAGCAATCTGATGCAGTTGACG TGTTCCGTATTCCTAGTTGACACATAACAACGTCCTCAC GTGTGCGAAGACCGTATTCGTGTCA
配列番号 33	ID_060_100bp	ACGTTTCTGCCTCTTAAGCTGTGAAGCAAGCCGCTTAG TCAGGACTGTATTCCAAGTGTCCAGCTTCCGGTTGAA CACTGTATCAGGCAACGAGCGTCA
配列番号 34	ID_061_100bp	ACGTACTIONACCAATTGGCCAATACAGTTAGTATCA GCTGGAGCCGTGAGACTTATGCTTCGGATGACTTCAGT TGGCTAATTGGAGGCGACCAGTCA
配列番号 35	ID_062_100bp	ACGTGATTAGATAATGAGTCCACACGCATCTTGTGTTG GACTGCAGCGTTAACAATGACGATGCCTTGCGTTACGG CGCTATCTTCATTAGGACTCGTCA
配列番号 36	ID_063_100bp	ACGTGAGATACGCGCCTGACGGATACCGTGACCGTAT ATGTCCAGGTTATGAGAGAACATGACTTGAGAGAGTC TGGCACTTCCTTGACGTGTCCGGTCA
配列番号 37	ID_064_100bp	ACGTCAGCGGCGTTGGTCACACGTCAACACCTCTCGTA GAATCTGCTGCAGCTCCTGTCTCGTTCAGTAGGCTAAGCA CTGAGGTAGTAAGGCTCAGTGTCA
配列番号 38	ID_065_100bp	ACGTACTIONCTACTCCTTCGTGTGAGGAAGGTGCCACG GTGATGTGTCACGTAGTCTCGGAATATATGGCCACACT ACTTCCATGCAACTGCGGATGTCA
配列番号 39	ID_066_100bp	ACGTGTGGCCTCTAGGAAGGAAGGTGCAACCAGCTAC TTGATAGACGTCGACTAGCATACTACTCCTCACGTGTG

10

20

30

40

		CTGTGCGAGTATGAGTTCCTAGTCA	
配列番号 40	ID_067_100bp	ACGTATTCGAATTGATGCGGACTACCACTAGCAACGCC GTTGATAATAGCAAGATGGAACGTACTTGTACCTTGCC TGAGGCGCTCGATTAACGCGGTCA	
配列番号 41	ID_068_100bp	ACGTAGCAACGTGTTAAGACTTGCGACAACGGTCCTG CGGAAGGTACTGATACTTACATTATTAGACTGCAGTTG ACTACCGTGTAAGTGGACGTCGTCA	
配列番号 42	ID_069_100bp	ACGTAATTCGAGCAGGCTACACTGGCTTAAGTAGGCT GTGTTCAGTAGCGTACACATTCTACAACGTCCGAGCCA CATCGATATGTGCCTAAGTCGGTCA	10
配列番号 43	ID_070_100bp	ACGTTAGGAATTGGCTGTGGTTCATTCGGCTGAATCTC GCCAACAACTCTCGATTGTTAAGGCACTTAAGAAGCA GAGTCGGCCGCCAATCTGGTAGTCA	
配列番号 44	ID_071_100bp	ACGTTCTTGAACCAGTGAGAAGTCAGCATAGTAACTCT CTGGTCAATTAACATAGACCATCGTCTCGGATTGCGTG GTCGACGCCTGCCAGAATGAGTCA	20
配列番号 45	ID_072_100bp	ACGTGGAATGAGGAATAGGATGCTAAGACAGGACAGA ACTCTGACCATAGGCTCCTCATTGTGGCTTGAGATCTT CTTCCACGAACCGTCCGGCACGTCA	
配列番号 46	ID_073_100bp	ACGTCTGACGACTACTGAGATCACCTAGTTCGGATGAA TGCGCCATTCATGCGGAGGTATACGAGCTTACGTCGGA TCCTAGCGCGTACTGACGTGGTCA	
配列番号 47	ID_074_100bp	ACGTAGGCACTATTCTAACTATTCTTAAGGCAGAGGCG AACGGTTAGTTATCCGCGCTCAGATAAGCCTCCTTACA GATCCGATATCAATGCTGGCGTCA	30
配列番号 48	ID_075_100bp	ACGTCGCTTCATGGACAATTAGTTACTGCTTACAGCCA GCACAGTGGTACTAACGATCGCCGTTAGCGCAACGCC TGAGATTATCGTAGTTGAATGGTCA	
配列番号 49	ID_076_100bp	ACGTGAGTCAGTAGTCTCCAGTCATCACGAGCGAATC GAGCTCGGTGAACAGTCGTGAACAATATACCTGGTTC AGATACGTATATAGTCAGTGCCGTCA	40
配列番号	ID_077_100bp	ACGTAGAATTCCACGTTACTGATGACCGGTAGATGAA	



50		GTTAGAGAGTAGCGCTCACTGTCTCGAACATCGACGCAA TTGTACTGTGTAACGTGCCATGGTCA	
配列番号 51	ID_078_100bp	ACGTCTCTTGCGCCGATATCCACTCTAAGATGTGCCAC GCGTAATCTAATAGAAGGAGCCGGAGCCGTAGTGGTA CCATCAACTTGACTGGTACTAGTCA	
配列番号 52	ID_079_100bp	ACGTCCTAAGTGC GGAAGTCCTAATTGTTGGTACGGTA TGCAAGCTTGTACGACCTTCTGCTACCTTATATAGAAG TACACAGTCGGAAGCGTCGGGTCA	10
配列番号 53	ID_080_100bp	ACGTGTGAACAAGATGATCCGGTATCCAGTAGGACCG TCATAAGATCACGAGTTACCACGTAGAGTCAATTGGAT AGTCCTGGCGGTTATGCTAGTGTCA	
配列番号 54	ID_081_100bp	ACGTCGGCCGTAATCGCTTGTATCTGCTACAGTGCATT GTGGCGCACTAGAGTAATACGGATATAGGTTCTCACA CATGCGCATCCAGGCGCATGGGTCA	
配列番号 55	ID_082_100bp	ACGTGGCAACTACCGACCTGTAATTGCTAGTCGACGCA TACGGTGTCCGTGCTGGTTCTTGGTGC GATCATATCAC CAGTTGAAGCAGTGATCTGAGTCA	20
配列番号 56	ID_083_100bp	ACGTCAATCGTACCGCTCTAAGTCCACGTGAACCTTGCT CTCCACAATGTAATAAGAAGTTCGCCGCTGCGGACGG AGAAGGTTGCTAGATTAGGCTGTCA	
配列番号 57	ID_084_100bp	ACGTCAGCCTTATTAGCAACATACACATCGCTCGCGAT ATGTAAGAATGGTACTCTTGCTCCAAGGTGGAGCGTGT AAGAACCGTTGGTTGCTGACGTCA	30
配列番号 58	ID_085_100bp	ACGTGCTCCTTGGAGGAGCATGATATAGTAACCTTCAG TTCTAACAGTCTTATACCGCTTCGAGCTAGCTCGCAAT GGCAACA ACTTCATCGGAAGGTCA	
配列番号 59	ID_086_100bp	ACGTAGCTGGTAACGTGGTTAGCCACCAGATCTCGCA GGAGCCATAGGTGTATTCGCATATGAATCTTCGGCGAC ATACGTCGGATATGCAGAGTCGTCA	
配列番号 60	ID_087_100bp	ACGTCATAGCTGCCGATCACCTATGGCATAACTCATT TACCAATTAGTCGGCAGGTGGCATGTACGCAGTGTGAT GCTCCAATGGCTTCTGAGTGGTCA	40

配列番号 61	ID_088_100bp	ACGTCAAGCGAGTGTTAACTGTCTACTCCACCGATGCA TGAGTCAACAAGCCATCCTGTCTGCGCTGGTGATTACT CTTAAGAGTCCATAGGCGAGGTCA
配列番号 62	ID_089_100bp	ACGTCCATATAGCGCGCTGCGAAGCTACACTACAGCG CATGATTGAATCAGGCCTTCTTAATTCAGGAGTCAGAT ATTCAGTGGCGCGGCGAGGACGTCA
配列番号 63	ID_090_100bp	ACGTGATCAGCGGTATCTCGTCATACGAATATCGTGAA CGTTACGCCTAATTCATGGTAGACTCGATGGCGCAAG AAGCGACCGAGTATCTGTTCGTCA
配列番号 64	ID_091_100bp	ACGTTCTTGTTAATGGTTGATAGCAACAAGGCGAGTAG TCCGAATAGCTCCGGAATCACACTGTTCTCCACGACGG TAGCCATCTTGCGTCTTAGCGTCA
配列番号 65	ID_092_100bp	ACGTAGCTGGAGAGACAACCTAGATCGGCGTCAATATC TGAGCGGTATAATATGCTTGGAATGCTAGCTGTATTGG CTCTCTCTCAATCTGGTAGTGGTCA
配列番号 66	ID_093_100bp	ACGTTTCATCAACAGAAGGAGAGACTGATACGATTAGC GCCATCCTGTTAGTGGCTCTTAACAGGCCGGTAGCGAT TCTGGCCATGGTATTCCGATGGTCA
配列番号 67	ID_094_100bp	ACGTGTGTGCTTAGCACTACGTGTCGTGAGTCCGTGAT AGTCCGTTGGTATGTCACCTCACTTGGCTAACGCCAAGA TACGTTTCAGCATTAAATCCAAGTCA
配列番号 68	ID_095_100bp	ACGTGAAGGCACTCTTATGGTAGTCGCAGCGTGATATT GTCCGACCATTGTGAGGATCGCACGGTACTCATCTTAC TGCCAATAGTGCTCCAGTAGGTCA
配列番号 69	ID_096_100bp	ACGTGAATCTCTGCAACGCGGAGACTGCCTCGGTTGTA AGCTACCACGTAGTCAGAACTCCGCGTCGCATATTGGC TATGATATTACGGAAGACCGGTCA
配列番号 70	ID_097_100bp	ACGTTGTAGCACCGTCCGTCCAACCTAGAGTTATACGAA GTGCTCGCCAGTTCATGATCGCTGCACTGCCAGGCTGT CACCTGTAGCTACAGTGCTTGTC
配列番号 71	ID_098_100bp	ACGTACTGTATCTGTAGCTGCGGTAGTGCTGACATTGT ACAAGAGCGTCCGTGCCACAGATACATCGACACGCAT

10

20

30

40

		GCGTTCGCCAGGTTAAGCGTGGTCA	
配列番号 72	ID_099_100bp	ACGTCGCGTCTCGTCGTAGGTAGTCAATCCTTGCAACC GCGCATAAGGACAGCATGAGGTGTCTATCTTATAAGA TAGACTAGAGGTTACGTGAGAGTCA	
配列番号 73	ID_100_100bp	ACGTTGCGGCAAGCAAGTAGCATTGAGACTGTTTACA GGACGACTACAGATGGCTCGTACACCTCCATACCTGGT TGCCGTTAGCTTGGACAGCGTGTCA	
配列番号 74	ID_101_100bp	ACGTGATGTGCGTCATCCGGAACCAACGCATTGTATGA TAACATACGGCTTGCAGCAGACGTTATCTCATAGACG TCGCGGCCTCCGCAAGGTAGGTCA	10
配列番号 75	ID_102_100bp	ACGTACCAGGTGTTGACCGGAGTACGGCAACTCGCGC AGGTGTGCACGTACCTATCACGGTAGTAATACTAGTAA CACGTCGATACTGTATAGATCGTCA	
配列番号 76	ID_103_100bp	ACGTATACACTCTGCGCGCACCACGATACGGCTAACGT GCGATATCAGTCCATGTCACAACCTCTGGCGTGGTAATG TAGCTTCTTGGAGTCGCTTAGTCA	20
配列番号 77	ID_104_100bp	ACGTAACATCGTGGACGTGTCTAAGCTCCTGGAGAAT AGACTCGTTATTGGCATCACGTCACCTTGCACGCGATAT TCCGTCTGCCGATATGGTCCTGTCA	
配列番号 78	ID_105_100bp	ACGTCTAGCGTGTAGTTGTCGGCTCCTCAAGTACTCAA GACCGCCTCTGGTGCCTCGAGAGCTCACTGCGTAGGA CATATGCTGACACGTTAGTTAGTCA	
配列番号 79	ID_106_100bp	ACGTAGATAAGTCCGTAAGCGTATTCTCACTGGAT TCATGCTGAACGTAGAACTGCTATAGTCCACGTGCTAC GCCGTTGACCGACGTACGAAGTCA	30
配列番号 80	ID_107_100bp	ACGTCTCTGTCGATCACACGCCGGACGCACTGTTACTT GACTGGAGGTGACCTTCGCACCAATAACGTTGAAGAG CTAGATTAGATGGTAGAACGAGTCA	
配列番号 81	ID_108_100bp	ACGTGGACCTGCTCGCCATAACGCGGATCGGTCCTGCG ATGTATCAACTGATGATTGACACACAGCGCAAGTATG ACCGTTAGATAATTCAGCTCGGTCA	40
配列番号	ID_109_100bp	ACGTACCAACCACGCGTCGGTGAGCCGTGATACTCGA	

82		TCTACCTAGGATATTGCACTTCCTGGCGGTATGCGACC GATCTTCGTGTTCAAGGAATCGTCA	
配列番号 83	ID_110_100bp	ACGTTACTAACGTGGCCGAGGCAACTTCATGGTGAATT GCTCTAGGCCACTATGTACCGCCAGTGACTCGGCAGAC CCGACAGAACCGCAATACGGGTCA	
配列番号 84	ID_111_100bp	ACGTAGACCAGCTGGCGCGACGTCTGCGAGCAGTCTC CTGAGAGGTGTGTGTCTCACTACTCAAGTACTCTGGCA CAGGCCGCATTAATTGGATTGGTCA	10
配列番号 85	ID_112_100bp	ACGTTACTATGCTTGCCGTACGGTTGTCGATGAGGTCC GTACAGCAGAATCGACTAGGTATAAGACCGACATGCT CAATTAGTCTGTCACCTCATAGTCA	
配列番号 86	ID_113_100bp	ACGTTACTAGGAGACCGTAATTAGTAGTGTAACCTAGC GACAGCCGCGTACCGCAACTTCACTTGGTATTGCTGCT ATCCATTAGGTGTCACGCGAGTCA	
配列番号 87	ID_114_100bp	ACGTCTAAGTCCGCCGAAGGCATATGTGAATTCACCTG GAGACTGCCATTCACTCCGTTGAAGTACTAGTTACGGA GCTGCTAGTGCACAGAGGAAGTCA	20
配列番号 88	ID_115_100bp	ACGTCACAGACCGGATTACGGTATTAGTTGCTGGAGG CGCACGCACGTCGCCGTATCAGTAAGGTGATTCCTGCG TATACTGGCCGGAAGCAGAATGTCA	
配列番号 89	ID_116_100bp	ACGTGCAGTCGCGATGACGGCTCGATCGCGAACTAAT TCTGAGCTGTAGACCGATGATCTGGTGGTAAGGCTATC ACTGTCAGTGCATCGAGCACTGTCA	30
配列番号 90	ID_117_100bp	ACGTATTGGATATAACAAGGATTATTGGCACGTAGTCAT ACGCACAGTGCCATGGCAAGTTGGCTTCTTACCACCTG GCCACCGTCGATCAGTGCAAGTCA	
配列番号 91	ID_118_100bp	ACGTGCGCGTCAAGGAATATGGCCATTCATGACAGAC GGAACCTCGTCGCATCTGTGTTCCATAAGGAGGACGAAT CTGCATCGACGTGGTCCTCGCAGTCA	
配列番号 92	ID_119_100bp	ACGTACAGCGCTGACAGGACTGGCGCAACTAGGAATG TCAGCACCTGGATCGGCTTAGAGGTGTGTCGATGTATC TGCTGCTAGTGTAACCTCCTCTGTCA	40

配列番号 93	ID_120_100bp	ACGTGCGGCCTGTGCGATTAGTCTTAAGCGGATCGAGT TATCCGCAAGCAGATTACAAGAAGCGTTCCTACATGA GTATTCCTGCATCGTATGGTAGTCA
配列番号 94	ID_121_100bp	ACGTTGTATTCACCTCGAGCTCCGCTGCACCTTCGGAAT TGAGAGAGTGGCAATCTCATGGACAGTCGTCGTCGTG CGCTAAGGTACAGTATTAGACGTCA
配列番号 95	ID_122_100bp	ACGTAGTGCCGAACATGCGTGAGCTCGCCTCGGAATT AGGCCTTGGCTAAGAATACTAAGGTATAGGCACTAA TACGCGGCTGTGGCAACATGATGTCA
配列番号 96	ID_123_100bp	ACGTAAGGCGGCATCATTCTCTGAGGTGACTCAAGTG CCAACCTCATTGTATAGTGGCTTAAGTTCATGCGGCAC TCTGAGAGGAACACGTATAAGTCA
配列番号 97	ID_124_100bp	ACGTGTCATATGACGAGGATTCGCACGGTCCGTACAC ATGCCGTCTCTCCAGGTTACATTATTCGAAGTGTGTCC ATGCGGAAGACGGATGTCTCCGTCA
配列番号 98	ID_125_100bp	ACGTTTGGACAGTGACGGAATAAGTTATGTCGCGT AACTCTCCGCTGATAATACATGCTAACGAGCCACGAGT TCTTGGCGAGGTCCACAAGGGTCA
配列番号 99	ID_126_100bp	ACGTAGAGCCACTCTCCATATCGATCTCCTCTATAGTC GTAGGTCGGCTACTGTGAGCAGGTGGAGTGACTGAAG CTTGAAGGACGAGATTCATAGTCA
配列番号 100	ID_127_100bp	ACGTAAGCGATAGGATAGCAGTATCCGACCAGCCTAG CGTGGCACATTCCGCATTCTCAGGCAGTTACCATGTAT ATCTCTCAAGAATGCGGTTGAGTCA
配列番号 101	ID_128_100bp	ACGTTGGTACATAGGAGGTTGACGCGGACCATCTGC AGTTACCACTCCGACTGTATGTTCTGTCGGCGGCAATG CGGAATGTACATGCGGATGATGTCA
配列番号 102	ID_129_100bp	ACGTAGTGTTGTTGTTCTGGCCTCTAGGAGAAGATTCA TACTTGCAAGACCGTACTACTAGACGAGTTACAGTCAT CCAGTAATCGGCGTGCCACAGTCA
配列番号 103	ID_130_100bp	ACGTCTTCCCTAGGAGTCGAGGTATGAGTTGTCCAGTTC GTTCTAGATCCTCAAGGTCCTATAAGGCTCCAACGACC

10

20

30

40

		GAGCAGCGGAATATGACCGTGTCA
配列番号 104	ID_131_100bp	ACGTGGCTCTATCGAGTTGTGCGACTACAAGAATGCGAC TTGGTATCGGCCTACGAGACTCAACGTGGAAGTAAGG AGCCACAAGGTATCACTCGTAGTCA
配列番号 105	ID_132_100bp	ACGTTATGAGATATAACGTCGCAGCGTGCTTCCGCATC GCGCAGACCATTAATGGTAATCCTAGACCTGGTACTTA GTCCGCAGGTTGTGTGAATAGTCA
配列番号 106	ID_133_100bp	ACGTCCAGGAGGACCTTATGTGTGCGAATATGGTCTGTA GGAGCATCCGTCCACATCTGATACTGAGTTAGCGCCAA CATCGGCCGCGATCTAGACGTCA
配列番号 107	ID_134_100bp	ACGTGTCTAACTGATCAGAGGAGGTGTAGTGCTCGTGT CCTAACCGCACACACGATTCGGTGCCAGCACGTAGA TCGGACGTGTTCGGTACATATAGTCA
配列番号 108	ID_135_100bp	ACGTGATTGCTGATAACGAGTTCTCACAAGGACAGTTA TATGGCGAACTGTCTGTCGTCAGTCGGCATTGAA CTCGCGCATCGGCGTAAGACGTCA
配列番号 109	ID_136_100bp	ACGTGAGGACACTAGTGTACGTGCTCGTACCATTCCCTG TTAGGTGATGCCTAATAACCTGTAATGCAGATAGAGTT ACAGTTCTACCGCCGACAAGTCA
配列番号 110	ID_137_100bp	ACGTCGTTCTGACCGTACATAGAGCGGCTACCGAGCT CACTGTTGGTAGCATAGTAGTCCAGTAATGTGCGAGCGG ATCGCATAACAAGGCTTGATGTCA

10

20

30

## 【 0 1 5 6 】

スパーク ( Spark ) パイアス制御スパイクイン

複数の長さにわたる核酸配列のセット ( 「スパーク ( Spark ) 」 ) はサイズマーカーとして働きうる。これらの配列はサンプル核酸と共にサンプルに添加され、処理 ( 例えば、抽出、精製、配列決定 ) されうる。あるプロセスは、異なる長さの核酸に差動的に影響を及ぼしうる。例えば、シリカ膜カラムを使用する核酸精製は、より長い長さの配列に対して偏向することが可能であり、または特定の長さの配列を保持するように最適化されうる。核酸配列決定は、典型的には、サンプルから核酸が抽出された後に行われるため、配列決定結果における長さの出現率または分布は元のサンプルを代表するものではない可能性がある。既知の量および長さのスパーク配列を添加することにより、種々の長さのサンプル核酸に対する処理および配列決定の効果をモニターし、定量化することが可能である。また、サンプル核酸およびスパークサイズセット ( Spark size set ) 核酸に関する配列決定リードの最終的な数を測定し、元のサンプルに添加された既知量のスパークサイズセット核酸に対して正規化することにより、元のサンプルにおける種々の長さのサンプル核酸の相対および / または絶対量を推定することが可能である。

40

## 【 0 1 5 7 】

幾つかの場合には、スパークサイズセットは、少なくとも約 3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、100、200、250、300、350、400、500、600、700、800、1000 個またはそれ以上の核酸を含みうる。幾つかの場合には、スパークサイズセッ

50

トは、多くとも約3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、100または200個の核酸を含みうる。幾つかの場合には、スパークサイズセットは約3～約50個の核酸、例えば約3～約30個の核酸を含む。幾つかの場合には、スパークサイズセットにおける核酸は、1以上の異なる特性、例えば、異なる長さ、異なるGC含量および/または異なる配列を有する。

#### 【0158】

スパーク核酸は、長さ識別配列、ロード (load) 配列、合成核酸識別配列 (それは、この場合、スパーク識別配列であろう) および特徴ドメインを含む、本明細書に記載されている合成スパイクイン核酸の特徴のいずれかを含みうる。幾つかの場合には、スパークサイズセットにおける核酸は、固定されたフォワード配列および/または固定されたりバース配列を含有する。固定されたフォワード配列および/または固定されたりバース配列はスパークサイズセットにおける全ての核酸に共通であることが可能であり、配列をスパークとして識別しうる。幾つかの場合には、固定されたフォワード配列および/または固定されたりバース配列は、少なくとも約5、10、11、12、13、14、15、16、17、18、19、20、25、30、32、40、50、60、70、80、90または100塩基対長である。幾つかの場合には、固定されたフォワード配列および/または固定されたりバース配列は、多くとも約5、10、11、12、13、14、15、16、17、18、19、20、25、30、32、40、50、60、70、80、90または100塩基対長である。幾つかの場合には、固定されたフォワード配列および/または固定されたりバース配列は、約8bp～約50bp、例えば、約8bp～約20bp、または約16bp～約40bpの範囲内である。幾つかの場合には、スパーク識別配列はサンプルにおいて見出されず、または天然で存在しない。幾つかの場合には、固定されたフォワード配列は、固定されたりバース配列とは異なる。

#### 【0159】

幾つかの場合には、スパークサイズセットにおける核酸はユニークフォワード配列および/またはユニークリバース配列を含有する。ユニークフォワード配列および/またはユニークリバース配列は該サイズセットにおけるスパークをお互いから識別しうる。幾つかの場合には、ユニークフォワード配列および/またはユニークリバース配列は、少なくとも約0、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、32、40、50、60、70、80、90または100塩基対長である。幾つかの場合には、ユニークフォワード配列および/またはユニークリバース配列は、多くとも約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、32、40、50、60、70、80、90、100、200、300、306、400または500塩基対長である。幾つかの場合には、ユニークフォワード配列および/またはユニークリバース配列は約4～約10塩基対長の範囲内である。幾つかの場合には、スパークサイズセットにおける各核酸は、異なるユニークフォワード配列および/またはユニークリバース配列を有する。幾つかの場合には、スパークサイズセットにおける各核酸は、同じ長さを有するユニークフォワード配列および/またはユニークリバース配列を有する。幾つかの場合には、スパークサイズセットにおける各核酸は、異なる長さを有するユニークフォワード配列および/またはユニークリバース配列を有する。

#### 【0160】

幾つかの場合には、スパークサイズセットにおける核酸は充填 (filler) 配列を含有する。幾つかの場合には、充填配列はサイズセットにおけるスパークをお互いから識別しうる。幾つかの場合には、充填配列は、少なくとも約0、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、32、40、50、60、70、80、90または100塩基対長である。幾つかの場合には、充填配列は、多くとも約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、32、40、50、60、70、80、90、100、200、300、306、400または5

10

20

30

40

50

00塩基対長である。幾つかの場合には、充填配列は0～約350bpの範囲内である。幾つかの場合には、スパークサイズセットにおける各核酸は、異なる長さを有する充填配列を有する。幾つかの場合には、充填配列の長さは、0、8、31、56、81、106、131および306bpからなる群から選択される。

【0161】

幾つかの場合には、スパークサイズセットにおける核酸は、少なくとも約10、20、30、32、40、50、60、70、80、90または100塩基対長である。幾つかの場合には、スパークサイズセットにおける核酸は、多くとも約100、200、300、350、400、500、600、700、800、900または1,000塩基対長である。幾つかの場合には、スパークサイズセットにおける核酸は約20～約500塩基対長の範囲内、約20～約400塩基対長の範囲内、または約20～約200塩基対長の範囲内である。

10

【0162】

例えば、以下の特徴を有する8個の二重鎖DNA配列(表2、図4における配列番号111～118)のセットが設計されうる: 32～350bpサイズ範囲(例えば、それぞれ0、8、31、56、81、106、131および306bpの充填配列長を有する32、52、75、100、125、150、175および350bpの断片)、固定された16bpのフォワード配列、フォワード配列とは異なる固定された16bpのリバース配列、ならびにユニーク6bpフォワードおよびリバース配列。

【表2】

20

表2: 典型的なスパークサイズセット配列

配列番号	名称	長さ(bp)	配列
配列番号	Spark-032	32	TAGTCAGTACAGCTGCTGCACGATACGTGTAC



111				
配列 番号 112	スパーク- 052	52	TAGTCAGTACAGCTGCGTCTACAGTCTGCCTGCT GTTGCACGATACGTGTAC	
配列 番号 113	スパーク- 075	75	TAGTCAGTACAGCTGCAGTGCAAGTCTGCCTTCA TCCTAAGTCACCGTCATTAGATGAGTGCACGATA CGTGTAC	10
配列 番号 114	スパーク- 100	100	TAGTCAGTACAGCTGCCGCATAAGTCTGCCTTCA TCCTAAGTCACCGTCATTAGGTTACGTGCACTT CAGGATCTGCACTCAGTGCACGATACGTGTAC	
配列 番号 115	スパーク- 125	125	TAGTCAGTACAGCTGCCTCAGTAGTCTGCCTTCA TCCTAAGTCACCGTCATTAGGTTACGTGCACTT CAGGATCTGCTATTACAGCGCGGAAGATGCCTG CGCTGTCATGCACGATACGTGTAC	20
配列 番号 116	スパーク- 150	150	TAGTCAGTACAGCTGCCATCTGAGTCTGCCTTCA TCCTAAGTCACCGTCATTAGGTTACGTGCACTT CAGGATCTGCTATTACAGCGCGGAAGATGCCTG CGACCAACGATCTAACACTTATATTGAGACATCT GCACGATACGTGTAC	
配列 番号 117	スパーク- 175	175	TAGTCAGTACAGCTGCTCGCTAAGTCTGCCTTCA TCCTAAGTCACCGTCATTAGGTTACGTGCACTT CAGGATCTGCTATTACAGCGCGGAAGATGCCTG CGACCAACGATCTAACACTTATATTGATCCGAAT CAATCATATGAGTGTTGATCTGCTTGCACGATAC GTGTAC	30
配列 番号 118	スパーク- 350	350	TAGTCAGTACAGCTGCGCTACAAGTCTGCCTTCA TCCTAAGTCACCGTCATTAGGTTACGTGCACTT CAGGATCTGCTATTACAGCGCGGAAGATGCCTG CGACCAACGATCTAACACTTATATTGATCCGAAT CAATCATATGAGTGTTGACCGGACATCGGCGTGT GGTGGCCGTGGAATCACTTATTCCATGGCTCCTG CCGCGATGTATATGTGACAGCGAGTTAGATAC GACAGATAAGTCGACGCGTGCCTTGTAGCCGTA	40

			CAACGCATATGTCTTCATTCCTGATAGAGTGTGCG GTACTCATAGGAGTGAACCTATACGGTATCGTGC ACGATACGTGTAC
配列 番号 121	スパーク 100.3a	100	ACGCGGAGTGAACGCTGTATAATCCAGTGTGCGT ATGATTCGTCTATCCTGTTTCGGATGAAGGCACCT GCGACGAAGGTATGAAGCATTGCCACGCACATT
配列 番号 122	スパーク 125.3a	125	GTCTCCAATTAGAGGTCTACCACGACACTTATT GGCGATCGATATAGACTGGTGACGCTGACCGGC ACCTCCAGTCCGGCTGATCAGGACTATCGTGAA GCGGTTCTAGTTCGTAACCTGTGTT
配列 番号 123	スパーク 150.3a	150	TGCTATACGGAACGGTCTCTAGGCGAACAAGTG AGGAATGTCAACAGAGACTAACATCGATATTCT CCTCGTCATTACTGTTGACGTAATTGCTCCGATG TCGCGCGCGGTCATGCCAGCTATACTGGCTAAG AGTACTATCCATATAC
配列 番号 124	スパーク 175.3a	175	AGCTGATGGACAATATCGCCACCTGCGACTGCT AGGCATGCTGCTATAAGCGAGGCTCCTACTAAG CGCTCGCTGTACTGGTGCGGAGGACTAGGAGTT CAATACGTGCGCCATTAACGGACGTATCGAGCA GACGGAACCTGCTTGGATCACCCTTCATGTTAGT TCTTGGAGA

10

20

## 【 0 1 6 3 】

G C 含量スパイクインパネル

既知濃度でサンプルに添加され、ついで、処理後に測定された核酸（例えば、DNA）は、収率およびプロセスに関する他の情報をもたらすことが可能であり、これらは、収率およびサンプル自体に関する追加的特性を推定するために使用されうる。例えば、あるサイズ範囲を含む核酸スパイクインセットはサンプル（例えば、血漿）に添加され、ついで抽出およびそれに続いて次世代シーケンシング（NGS）に付されうる。各サイズスパイクの収率は、意図的なサイズ選択、温度および他の変性要因ならびにPCRバイアスを含む、処理中の多数の要因に応じて変動しうる。この情報は、所望のサイズ範囲の回収を最大にすることを目的とした新規方法を開発するために、または既存プロセスをモニターする（例えば、品質管理）ために有用でありうる。

30

40

## 【 0 1 6 4 】

二本鎖DNAライブラリー調製物の場合、比較的低い融解温度（ $T_m$ ）のDNA二重鎖の変性は、 $T_m$ に反比例して、これらの二本鎖の収率を低下させる。与えられた条件（例えば、塩濃度、温度、pHなど）において、二重鎖の $T_m$ に影響を及ぼす寄与因子には、長さおよびGC含量が含まれる。単一GC含量を有する単一種で各サイズが代表される、二重鎖のサイズ範囲は、種々の条件に対する $T_m$ の応答に関する部分的な情報のみを提供しうる。

## 【 0 1 6 5 】

核酸の長さおよび/またはGC含量が核酸の $T_m$ および処理にどのように影響を及ぼすかに関する情報は、例えば、血中の種々の病原体からの短い無細胞断片の回収を推測する

50

ための代用物としてスパイクインを使用する場合に重要でありうる。病原体核酸はそれらのGC含量において劇的に変動する可能性があり、したがって、短い断片長において非常に様々な $T_m$ を有しうる。多数のcfDNA断片の短い長さ(例えば、30、40、50bp)を考慮すると、それらは、例えばNGSのための処理中の変性に感受性でありうる。広範な $T_m$ 範囲にわたって回収を追跡するための、より詳細なスパイクインセットは、未知サンプルの出発量のより良好な推測を可能にしうる。

#### 【0166】

ある範囲の $T_m$ 、GCおよび/または長さにわたるスパイクイン核酸のパネルは、絶対的存在量の決定のために、および/または変性の詳細なモニターを可能にするために使用されうる。例えば、表3に示されているとおり、4つの異なる長さ(例えば、32、42、52および75bp)および各長さについて7つの異なるGC含量(約20、30、40、50、60、70または80%のGC)を有する核酸を含む28個の異なる核酸(例えば、二重鎖)のパネルが使用されうる。全体として、該パネルは、各サイズごとに単一のGC含量を有するセットより高い細分性をもたらしうる。幾つかの場合には、合成核酸(dsDNA、ssDNA、dsRNA、ssRNA)のパネルは、少なくとも3個の異なる長さ、および各長さについて少なくとも2個の異なるGC含量、少なくとも3個のGC含量、少なくとも4個のGC含量、少なくとも5個のGC含量、少なくとも7個のGC含量または少なくとも10個のGC含量の核酸を含有しうる。幾つかの場合には、合成核酸(dsDNA、ssDNA、dsRNA、ssRNA)のパネルは、少なくとも5個の異なる長さ、および各長さについて少なくとも2個の異なるGC含量、少なくとも3個のGC含量、少なくとも4個のGC含量、少なくとも5個のGC含量、少なくとも7個のGC含量または少なくとも10個のGC含量の核酸を含有しうる。

#### 【0167】

幾つかの場合には、スパイクインパネルは、少なくとも3、5、10、15、20、25または30個のユニーク核酸を含む。幾つかの場合には、スパイクインパネルは、多くとも15、20、25、30、35、40、45、50または100個のユニーク核酸を含む。

#### 【0168】

種々のGC含量を有するスパイクイン核酸が使用されうる。幾つかの場合には、スパイクインパネルは、約40~60%のGC、約45~65%のGC、約30~70%のGC、約25~75%のGC、または約20~80%のGCの範囲にわたるGC含有量を有する核酸を含む。幾つかの場合には、スパイクインパネルは、少なくとも2、3、4、5、6、7、8、9または10個の異なるGC含量を有する核酸を含む。幾つかの場合には、スパイクインパネルは、多くとも3、4、5、6、7、8、9、10、15または20個の異なるGC含量を有する核酸を含む。幾つかの場合には、スパイクインパネルは、GCが少なくとも1、2、3、4、5、6、7、8、9、10、15または20%異なる種々のGC含有量を有する核酸を含む。GCの百分率は、GヌクレオチドおよびCヌクレオチドの数の和を配列内の総ヌクレオチド数で割り算することにより算出されうる。例えば、配列ACTGの場合、GCの百分率(%GC)は $(1+1)/4=50\%$  GCとして算出されるであろう。

#### 【0169】

種々の長さを有するスパイクイン核酸が使用されうる。幾つかの場合には、スパイクインパネルは、少なくとも3、4、5、6、7、8、9、10または15個の異なる長さを有する核酸を含む。幾つかの場合には、スパイクインパネルは、多くとも3、4、5、6、7、8、9、10、15、20、25、50または100個の異なる長さを有する核酸を含む。幾つかの場合には、スパイクインパネルは、約40~50bp、約35~55bp、約30~60bp、約35~60bp、約35~65bp、約35~70bp、約35~75bp、約30~70bp、約30~80bp、約30~90bp、約30~100bp、約25~150bp、約20~300bp、または約20~500bpの範囲にわたる長さを有する核酸を含む。幾つかの場合には、スパイクインパネルは、少なくとも

1、2、3、4、5、6、7、8、9、10、15または20bp異なる種々の長さを有する核酸を含む。幾つかの場合には、スパイクインパネルは、32、42、52および75bpの長さ、または27、37、47、57、62および67bpの長さを有する核酸を含む。

【0170】

一連の値から選択された長さおよびGC含量を有するスパイクイン核酸が使用されうる。例えば、合成核酸のセットは2以上の長さおよび2以上のGC含量から選択されうる。表3における28個の合成核酸のセット(配列番号125~配列番号152)は4つの異なる長さ(例えば、32、42、52および75塩基対)および7つの異なるGC含量(例えば、約20、30、40、50、60、70および80% GC)から構成される。異なる長さ(例えば、27、37、47、57、62および67bp)および異なるGC含量(例えば、約15、25、35、45、55、65および75% GC)を用いて、合成核酸の類似セットが得られうる。

10

【0171】

種々の融解温度( $T_m$ )を有するスパイクイン核酸が使用されうる。幾つかの場合には、スパイクインパネルは、約40~50、約35~55、約30~60、約35~60、約35~65、約35~70、約35~75、または約30~70の範囲にわたる融解温度( $T_m$ )を有する核酸を含む。幾つかの場合には、スパイクインパネルは、少なくとも1、2、3、4、5、6、7、8、9、10、15、20、25または30異なる種々の融解温度( $T_m$ )を有する核酸を含む。

20

【0172】

幾つかの場合には、 $T_m$ は、二本鎖の長さおよびGC含量に加えて、二本鎖濃度、ヌクレオチド配列の最近傍効果、高次DNA構造、1価および/または2価カチオン濃度ならびにヌクレオチド濃度に基づいて計算されうる。幾つかの場合には、 $T_m$ は、与えられた条件、例えば、二本鎖DNA特異的色素および温度の漸増および色素シグナルの検出に関して、実験的に計算されうる。

【0173】

種々の配列を有するスパイクイン核酸が使用されうる。好ましくは、非天然もしくは非自然配列、またはサンプル核酸にハイブリダイズし得ない配列が使用される。幾つかの場合には、スパイクインパネルは、少なくとも3、4、5、6、7、8、9、10または15個の異なる配列を有する核酸を含む。幾つかの場合には、スパイクインパネルは、多くとも3、4、5、6、7、8、9、10、15、20、25、50または100個の異なる配列を有する核酸を含む。

30

【0174】

種々の数のスパイクイン核酸が使用されうる。幾つかの場合には、約1、2、3、4、5、6、7、8、9、10、15、20、25、30、40または50個の核酸が使用される。例えば、表3に列挙された28個の配列のサブセット(例えば、32/42/52/75bp x 20/50/80% GC)が使用されうる。

【0175】

RNAの用途には、RNAパネルが使用されうる。本明細書に記載されているとおり、RNAパネルは、同一分子、あるいは長さ、GC含量および/または他の特性に関して異なる多様な分子を含みうる。

40

【0176】

8個のDNA配列のセット(それぞれ約50% GCである、表2における配列番号111~118)は、表3に列挙された28個のメンバーのGCパネルの部分的な適用範囲(カバレッジ)をもたらす。

## 【表 3】

表3: 典型的なGCスパイク配列

配列番号	名称	配列 (5'-3')	長さ (bp)	GC%
125 配列番号	スパーク32.20.1	AGTCATATTATTATATTAATTTAACTATCACG	32	19
126 配列番号	スパーク32.30.1	AGTCACGTTATTTCTTTTGAAAATACACACG	32	31
127 配列番号	スパーク32.40.1	AGTCTAAAGTCCTAACTCTTTGATCACACACG	32	41
128 配列番号	スパーク32.50.1	AGTCGGTTACTTTCTGGAGAATCCTTGGCACG	32	50
129 配列番号	スパーク32.60.1	AGTCCTTGCCAAGAGGCTCCGTAAGCACACG	32	59
130 配列番号	スパーク32.70.1	AGTCCGGCTGAGGCTCGGGACCTTGGCTCACG	32	69
配列番号	スパーク32.80.1	AGTCCCACCGGCTCGGGGCGGCAGCGGCCACG	32	81

10

20

131				
配列番号 132	スパーク42.20.1	AGTCTTATATATTAATACTATTTTCTTTTAAAGAT AATCACG	42	19
配列番号 133	スパーク42.30.1	AGTCACTATATATATTTAGAGACGAATATCAAGT AAAGCACG	42	31
配列番号 134	スパーク42.40.1	AGTCTACACTACTCGGGCTTTAAACGAAATTCAA CATTCACG	42	41
配列番号 135	スパーク42.50.1	AGTCTGGATACGATAACGGGAGCCCTTATTGAC GGATACACG	42	50
配列番号 136	スパーク42.60.1	AGTCTGTTTACGCGGGCGGCCCAAGAGACGTAG TCGTACACG	42	59
配列番号 137	スパーク42.70.1	AGTCCAGGGATCCACCCGTCAGGCTGCTAGCCG CCAGCCACG	42	69
配列番号 138	スパーク42.80.1	AGTCCGGTGGCTCCGCGTGCGGGCGCGGCACCG GCCACCACG	42	81
配列番号 139	スパーク52.20.1	AGTCAAATATCGATACAAATTAATAATTTTACT ATTTTAAAGATTATCACG	52	19
配列番号 140	スパーク52.30.1	AGTCATGTTTAATTACTGAGAACGTTATGTAATA TATGTCCTGTAAATCACG	52	31
配列番号 141	スパーク52.40.1	AGTCGGTGTAGTTGTGAGTTAATCTAAGGAATAC CTTTGTTCTATGTCACG	52	41
配列番号 142	スパーク52.50.1	AGTCTAGGCTGCTTGGCTTCTTCTAGCTCACTTG GTTATCCCGACATACACG	52	50
配列番号 143	スパーク52.60.1	AGTCTCCTAGCGGTACAGCTACTGTCATTCTCG GGCCCTCTAGTCGCCACG	52	59
配列番号 144	スパーク52.70.1	AGTCGGCGCGCCGTTATATGGCAGAGCGGCGGT CGCCCGTCTGCGGAACACG	52	69
配列番号 145	スパーク52.80.1	AGTCGCCAACGACTCGCGCCCCGGGGCGCGCC GCGCAGCCGGCTCGTCACG	52	81
配列番号	スパーク75.20.1	AGTCATTAATATATTATGGTTTATTATGAAACGA	75	19

10

20

30

40

146		TAAAAGTATCTTTGTATAAAAATTTAGAGTTTAAA TTACACG		
配列番号 147	スパーク75.30.1	AGTCGAAAGCTTATTTATAAGCATACTAAAATAT TAACTTCTTGATIGGCATCGAATATATACTTTCA CAACACG	75	31
配列番号 148	スパーク75.40.1	AGTCGAAATGAAAAGGTTTAGGATGAATCCTTA AATAAATCTACCTTAGGTCGTACTACCGAGTG GACTCACG	75	41
配列番号 149	スパーク75.50.1	AGTCAGGAGGAAGGACAATATAGAATGCGCGTT ATCTCCTTGTCTCCAACCGGCTCAATGCTTAGTT GGCCACG	75	50
配列番号 150 配列番号	スパーク75.60.1	AGTCTCGCACTTTTGCTTCCGCTGAGATACTCGC CGCCTGGACCGATTGAGGTCGGGAGCCCACCTCT TGACACG	75	59
配列番号 151	スパーク75.70.1	AGTCTGATAGCACGCCGCCACGGGCTCAGCCTT CCACCCGGCCGACGTTGCCGTACCTCTCCCTGCG GAGCACG	75	69
配列番号 152	スパーク75.80.1	AGTCGGGACTCACCCGTGGCCTGCTGAGGGCC CGCGCGGCGCCCCGGCCCTCTGGCGAAGGCGGC GAGCCCACG	75	81

10

20

## 【0177】

縮重スパイクイン：スパーク (Spank)

スパイクイン合成核酸は、核酸の縮重プール、または高い多様度を有する核酸のプールでありうる（本明細書においては「スパーク (Spank)」と称されることもある）。一般に、スパークは、配列決定反応につながるおよび/または配列決定反応を含むサンプル処理工程中に生じる絶対的もしくは相対的な核酸損失または多様性減少を決定するために使用される。スパーク配列のユニークプールの場合、プール内の配列多様性の減少は核酸存在量の減少に直接対応するはずであり、この場合、増幅またはPCRバイアスの影響を考慮する必要はない。例えば、 $10^8$  個のユニークスパーク配列をサンプルに添加し、配列決定後に  $10^4$  個のユニークスパーク配列しか回収されなかった場合、核酸の存在量および核酸の多様性は共に、 $10^4$  分の1に減少したことになる。幾つかの場合には、スパークは、重複分子の回収の度合を決定するために使用される。例えば、抽出およびライブラリー処理（これは種々の投入分子のPCRおよび潜在的不均一増幅を含む）の後、個々のスパークの配列決定およびアライメントは重複分子の回収の度合を示しうる。

30

40

## 【0178】

ついで、決定された多様性減少を用いて、1以上のサンプル処理または配列決定工程の前に、初期サンプルにおける核酸（例えば、標的核酸）の絶対的存在量を決定することが可能である。幾つかの場合には、決定された多様性減少を用いて、初期サンプルにおける核酸の相対的存在量を決定する。図5に示されているとおり、サンプル核酸 ( $S_1$ 、 $S_2$ 、 $\dots$ 、 $S_m$ ) は、1以上のサンプル処理工程の前に、スパーク (Spank) スパイクイン合成核酸 ( $SP_1$ 、 $SP_2$ 、 $\dots$ 、 $SP_n$ ) と一緒にされうる。例えば、約  $10^8$  個のユニークスパークがサンプルに添加される。サンプル処理（例えば、核酸抽出、

50

精製、連結および/または末端修復)中に、サンプル核酸の一部および合成核酸の一部が失われうる。サンプル処理後、最初の $10^8$ 個のユニーク配列のうちの約 $10^6$ 個のユニーク配列が残存しうる。ついで、これらの配列の一部、例えば $10^4$ 個のユニーク配列が配列決定されうる。絶対的多様性減少は、最初のユニーク配列の数を配列決定または回収されたユニーク配列の数で割り算したものとして算出されうる(例えば、 $10^8 / 10^4 = 10^4$ )。同様に、回復値は、配列決定または回収されたユニーク配列の数を最初のユニーク配列の数で割り算したものとして算出されうる(例えば、 $10^4 / 10^8 = 10^{-4}$ )。算出された多様性減少は、初期サンプルにおける核酸の絶対的存在量を決定するために用いられうる。例えば、スパイク配列に関する及びサンプル配列に関する配列決定リード数は配列決定分析から決定可能であり、サンプルに添加されるスパイク配列の初期濃度または量は既知である。決定された多様性減少を用いて、初期サンプルにおける核酸(例えば、特定の生物、病原体、腫瘍または器官からの核酸)の初期濃度または量が決定されうる。元のサンプルにおけるサンプル核酸の絶対量は、サンプル核酸およびスパイク核酸に関する配列決定リードの最終的な数ならびに/またはスパイク核酸の最終的な多様性を測定し、元のサンプルに添加されたスパイク核酸の既知量または多様性に対して正規化することにより推定されうる。

10

**【0179】**

ユニーク配列リードの数は種々の方法により決定されうる。例えば、識別タグを有する配列リードが特定(識別)されうる。ついで、識別タグを有する配列リード内のユニーク配列の数は、重複配列を重複排除(duplicate)(「デデュープ(deduplicate)」)または除去することにより決定されうる。例えば、それらの配列を、可能な配列の参照データベースに対して又はお互いに対してアライメントさせて、どれが重複しているのか、およびどれがユニークである又は異なるのかを決定することが可能である。識別タグは典型的には配列間で保存されているため、各添加分子内に埋め込まれたランダム化配列領域が分析されうる。幾つかの場合には、スパイク核酸は識別タグを含まず、そのような場合には、スパイクは、例えば、既知配列を含むデータベースに対する参照またはアライメントのような他の方法により特定されうる。

20

**【0180】**

スパイク配列は、相対的損失および/または絶対的損失をモニターするために使用されうる。幾つかの場合には、スパイク配列の多様性が十分に高い場合には、サンプルに添加されるスパイク配列は実質的に全て非ユニークであると仮定されうる。したがって、配列決定された重複スパイク配列が存在する場合、それはPCR増幅によるものであり、サンプルに添加された同じスパイク配列の複数コピーによるものではない可能性が高く、分析から除外されうる。また、各スパイク配列がユニークである場合、サンプルに最初に添加されたスパイク配列の総数は、サンプルに添加された核酸の濃度および体積に基づいて既知であり、配列決定後のユニークスパイク配列決定リードの総数は既知である。これらの値を一緒に使用して、多様性減少値または回収値が計算されうる。

30

**【0181】**

本発明で提供する方法は、ボトルネック効果(population bottleneck)または多様性の減少に関連するサンプル処理中の工程を特定する方法を含む。幾つかの場合には、ボトルネック効果が特定された場合、出発集団における最初は未知であったその他の分子に補正係数が適用されうる。例えば、投入スパイク分子は実質的に全てユニークであるが、回収されたスパイクは50%しかユニークでない場合、これは、サンプルからのその他の分子の多様性の解釈に関する情報をもたらしうるボトルネック効果および多様性の減少を示す。

40

**【0182】**

ボトルネック効果が生じる工程を特定するために、サンプル処理中の任意の工程において、スパイクの集合体(コレクション)がサンプルに添加されうる。例えば、サンプル(例えば、体液)を対象から採取した際に、スパイクの第1集合体を導入することが可能であり、採取サンプルの後続処理(例えば、残存細胞の除去、保存)の前または途中で、ス

50



パンクの第2集合体をサンプル内に導入することが可能であり、および/または、ライブラリーの製造の前に、スパングの第3集合体を導入することが可能である。幾つかの場合には、サンプル処理中の異なる工程においてサンプルに添加されたスパングの集合体は、同じまたは類似した組成を有する。幾つかの場合には、スパングの別の集合体をサンプル処理中の異なる工程においてサンプルに添加する。

**【0183】**

幾つかの場合には、スパング核酸はそれぞれ、ユニーク配列を有するランダム化部分を含有しうる。スパングは1以上の異なるドメインを含みうる。幾つかの場合には、スパングは1以上のプロセスコード、1以上の多様性コード、1以上の長さ識別配列、1以上のロード (load) 配列、1以上の合成核酸識別配列 (またはスパング識別配列) および/または1以上の特徴ドメインを含みうる。幾つかの場合には、スパングは識別タグおよびユニーク核酸配列を含みうる。

10

**【0184】**

スパングの種々の集合体を使用される場合、各集合体は、特定の工程 (例えば、サンプル採取、抽出、ライブラリー処理) においてサンプル内に導入されるスパング集合体を識別するための「プロセスコード」でコードされうる。そのような場合においては、同一プロセスコードを有するスパングは生物情報学的に分類され、多様性減少に関して分析されうる。ついで、特定の工程に関連する多様性減少の度合いが決定され、ついで各サンプル処理工程にわたって比較されうる。

**【0185】**

スパングは、合成核酸またはスパングの全プールまたは集合体に関連する「多様性コード」を含みうる。多様性コードドメインは、合成核酸のプール内の多様性の量を示すユニークコードでありうる。そのような場合においては、多様性プール内の各合成核酸は、プールの多様性の度合い (例えば、 $10^8$  個のユニーク配列) を示す配列でコードされうる。幾つかの場合、例えば、2以上の多様性プールが同一サンプルに関して使用される場合には、該多様性コードは、それらの2以上のプールにおける多様性減少を特定するために使用されうる。

20

**【0186】**

幾つかの場合には、スパングは、特定のスパングプールまたは集合体のメンバーとしてスパングを特定する1以上のコード (例えば、プロセスコード) を含みうる。幾つかの場合には、スパングは、サンプルに最初に存在した核酸ではなくスパングとしてスパングを特定する1以上のスパング識別ドメインを含みうる。本明細書に更に詳細に記載されているとおり、スパングは特徴ドメイン、長さ識別子ドメイン (長さ識別ドメイン) およびロード (load) ドメインをも含みうる。

30

**【0187】**

スパングは、核酸の存在量を計算するため、または他の用途のために、単独で、または他の合成核酸と組合せて使用されうる。幾つかの場合には、スパングは他の合成核酸と共に使用されうる。例えば、幾つかの場合には、スパングのパネルおよびスパーク (Spark) のパネルがサンプルに添加されうる。幾つかの場合には、サンプル識別核酸もサンプルに添加されうる。

40

**【0188】**

スパングプールは、好ましくは、核酸配列の多様な混合物を含む。したがって、スパングプールは、多様性を最大にするように設計されうる。幾つかの場合には、スパングプールは、より一層大きなスパングプールに由来する。例えば、幾つかの場合には、75 bp のオリゴヌクレオチドは、2つの8 bp スtring (文字列) のN (例えば、等しい比率のA/C/G/T) を使用して合成されうる。スパングは、(i) 1以上の識別タグと (ii) ユニーク核酸配列とを含む合成核酸でありうる。幾つかの場合には、ユニーク核酸配列は複数の縮重またはランダム位置であることが可能であり、例えば、図6に示されているとおり、1以上のヌクレオチドによって隔てられた8 bp スtringの縮重位置の2つのグループでありうる。2つの典型的な配列が表4に列挙されている。2つの8 bp スト

50

リングのNを有するオリゴヌクレオチド設計は、 $4^{16} = 4 \cdot 3 \times 10^9$  個の異なるオリゴヌクレオチドのプールで、合計16個のNを含有する。例えば、このプールの $1 \times 10^8$  個の分子が1 mLの血漿に添加され、IDスパイクおよびスパークに関して前記で記載されているとおりに処理された場合には、スパークのほぼ全てがユニークとなる。例えば、そのような場合には、スパークの90%超(すなわち、90%を超える)、95%超、99%超がユニークでありうる。

#### 【0189】

幾つかの場合には、スパーク核酸は、少なくとも約20、30、40、50、60、70、75、80、90、100、110、120、125、130、140、150、160、170、175、180、190、200、250、300、350、400、450、500、600、700、800、900または1000ヌクレオチド長でありうる。幾つかの場合には、スパーク核酸は、多くとも約20、30、40、50、60、70、75、80、90、100、110、120、125、130、140、150、160、170、175、180、190、200、250、300、350、400、450、500、600、700、800、900または1000ヌクレオチド長でありうる。幾つかの場合には、スパーク核酸は約20~約175塩基対の範囲内の長さを有しうる。幾つかの場合には、スパークセットにおける核酸は同じ長さを有する。幾つかの場合には、スパークセットにおける核酸は2以上の異なる長さ(例えば、2、3、4、5個またはそれ以上の長さ)を有する。

#### 【0190】

幾つかの場合には、スパーク核酸は、少なくとも約5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29または30個の縮重位置を有しうる。幾つかの場合には、スパーク核酸は、多くとも約10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29または30個の縮重位置を有しうる。スパーク核酸は約5個~約25個の範囲内の数の縮重位置を有しうる。幾つかの場合には、縮重位置は連続的であり、分離しており、または2以上のグループ、例えば2、3、4または5個のグループに分割されうる。縮重位置が幾つかのグループに分割される幾つかの場合には、縮重位置はグループ間で均等に分割されることが可能であり(例えば、8bpストリングの縮重位置の2つのグループで、合計16個の縮重位置)、あるいはグループ間で不均等に分割されることが可能である(例えば、1つのグループは10個の縮重位置で、別のグループは6個の縮重位置で、合計16個の縮重位置)。縮重位置が幾つかのグループに分割される幾つかの場合には、それらのグループは1以上のヌクレオチドによって隔てられうる。幾つかの場合には、それらのグループは少なくとも約1、2、3、4、5、6、7、8、9、10、15、20、30、40または50個のヌクレオチドによって隔てられる。幾つかの場合には、それらのグループは少なくとも約1、2、3、4、5、6、7、8、9、10、15、20、30、40または50ヌクレオチドによって隔てられる。

#### 【0191】

幾つかの場合には、スパーク核酸は、少なくとも $1 \times 10^3$ 、 $1 \times 10^4$ 、 $1 \times 10^5$ 、 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ または $1 \times 10^{11}$  個のユニーク配列の多様性を有しうる。幾つかの場合には、スパーク核酸は、多くとも $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、

$0^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ または $1 \times 10^{11}$ 個のユニーク配列の多様性を有しうる。幾つかの場合には、スパンク核酸は約 $1 \times 10^4$ ～約 $1 \times 10^{11}$ 個のユニーク配列の範囲内の多様性を有しうる。

【表4】

表4: 典型的なスパンク多様性配列

配列番号	名称	配列
配列番号 119	スパンク- 75A	CCTGTGCTCTAGAGTAGAGTANNNNNNNNAGCTNNNNNN NNGAGCGATCTGTATAGATAGCTACACGCTGAGTGA
配列番号 120	スパンク- 75B	CCTGACTCTCGATGATAGTACNNNNNNNNAGCTNNNNNN NNGTCTATAGCTCTAGCGACATACGTA CTGTGTCGT

10

【0192】

トレーサー配列

実験室（検査室）由来の核酸（例えば、病原体ゲノムDNA）は、感染症診断試験の開発、立証、検証、アッセイ対照などのための標準として有用である。しかし、これらの同じ生物は臨床サンプル（例えば、病原体感染サンプル）中に存在しうるため、実験室由来の物質が、試験中に、臨床サンプルを交差汚染して、偽陽性判定を与える危険性があり、これは患者および医師に不正確な情報を与えうるだけでなく、ある病原体種に関しては、保健当局への法定の届出をも生じさせうるであろう。実際の参照核酸（例えば、実際の病原体ゲノムDNA、癌核酸、腫瘍核酸または他の疾患関連核酸）は陽性対照として有用であり、または更には不可欠であるが、それを取り扱う際の通常の又は更には特別な注意は、次世代シーケンシング（NGS）のような高感度アッセイの場合には特に、交差汚染を予防するには不十分でありうる。

20

【0193】

天然に存在しない又はサンプル核酸にハイブリダイズし得ない合成トレーサー核酸は、陽性対照核酸と少なくとも同程度に高い有効濃度で、陽性対照核酸ストックに添加される。トレーサーおよび陽性対照核酸は、それらが同じ状態で処理され検出されるような形態で存在する。したがって、エンドポイント（例えば、NGSの場合のアライメントされた配列リード）はトレーサーおよび陽性対照核酸の両方に関して同じであり、トレーサーは、より高いその有効濃度ゆえに、陽性対照核酸と少なくとも同じくらい容易に検出される。幾つかの場合には、陽性対照核酸は病原体ゲノムDNAである。幾つかの場合には、陽性対照核酸は発癌遺伝子のような疾患関連核酸を含む。

30

【0194】

トレーサー配列は、例えば配列、長さ、濃度、GC含量などのような1以上の特性において変動可能である。表5に示され実施例6で使用される配列は約50%のGC含量を有するが、トレーサー配列は、それが組合される陽性対照またはゲノムの組成に合致するように変動可能であり、例えば、30%のGC含量、35%のGC含量、40%のGC含量、45%のGC含量、50%のGC含量、55%のGC含量、60%のGC含量、65%のGC含量、または70%のGC含量を有しうる。

40

【0195】

幾つかの場合には、トレーサー配列は、例えば実施例6に記載されているとおり、断片化後の陽性対照核酸またはゲノムDNAに添加されうる。幾つかの場合には、陽性対照核酸またはサンプル核酸に対して行われる完全な処理をより良好に表すために、トレーサー配列は断片化前の陽性対照核酸またはゲノムDNAに添加されうる。臨床サンプル（例えば、病原体DNA）において稀であり低濃度で見出される陽性対照核酸は、未標識核酸での交差汚染を最小にするために、可能な限り早くトレーサー配列で標識されうる。

50

## 【 0 1 9 6 】

幾つかの場合には、2以上のトレーサー配列が各陽性対照核酸に添加される。幾つかの場合には、2以上、3以上、4以上または5以上のトレーサー配列が、同じ濃度または異なる濃度で添加される。

## 【 0 1 9 7 】

用途によって異なる形態のトレーサー配列が使用されうる。例えば、トレーサー配列の長さは、対照配列の長さに、例えば平均または中央値の長さに一致させることが可能である。幾つかの場合には、トレーサー配列の長さは対照配列の平均または中央値の長さの5%、10%または20%以内でありうる。

## 【 0 1 9 8 】

RNA用途にはRNAトレーサー配列が使用されうる。

## 【表5】

表5: 典型的なトレーサー配列

配列番号	名前	配列
配列番号	トレーサー138	GCGTTGGCCGAGATGAAGACCTCGTGCTCACGTTACCAC
153		GCATGAGTTAATCAGTTGGCACGAAGGTCGGCATT
配列番号	トレーサー139	GCGCCACAGCCTGTGCTACTATGGCTAACAGCGTATGCCG
154		TCCGGATAGTGACCTGTCCGCGTCGGATACTTGGC
配列番号	トレーサー140	ACGTCGGTTGTTACATATGACGCAACGCTTGATTGAAGGC
155		GTTGTGAATCGCGCAGTACCGTGCTCGCTCAGAGC
配列番号	トレーサー141	GACAATTGAATTGTGGCCGCTTACTTCGCACTACCGCACG

10

20

156	CGGCAATGCTATGATGTGAGAACTTGATTCTTGGC	
配列番号 トレーサー142	GTAATGGAATCCTACAGACATGTGTAAGTATGCTGATTGA	
: 157	TCCGACTTACATCAGTCAGTCAGAGGCCGCCATAT	
配列番号 トレーサー143	CGCTAGTCCTCTCCTCGGTTCCGAACCTTCGTCGTCACGAG	
158	GTGGCTAACCATGGCATTAGTGCTCTAAGTCCTAT	
配列番号 トレーサー144	CTAGATGAAGGCAGGCGAGCAACCGGATAGATGGCTCCG	
159	TACCGCTGCGAGGTATTCTCCGATATATTGTACTCT	
配列番号 トレーサー145	ATGTCCTGAAGGAATAGAGAATCACTCTGCCAGCGACAA	10
160	GACGGCCTGGTTCAGTAGATCTAGAGATCAGAATCT	
配列番号 トレーサー146	TACGAGTATGACGATTGGCGACAACGTGGCATCTGCTTA	
161	ATTATTATGTCGCCTGCGACCACAGCACGCGAGACT	
配列番号 トレーサー147	AGTTGTGACTGCAGATGCCATAACAGCACCGAACCATAT	
: 162	GTGACTGCGGTGGCGAGCGAGCTAATGCTTGCGTGG	
配列番号 トレーサー148	AACTGTAAGAAGAATATTCCGGCTGCCAGCCTTGAATGTC	
163	TAGCGAAGCCGAACGCATAGAGGATGCATGTGCGG	20
配列番号 トレーサー149	GATTCCGAGGCCTAGTGCGACAGCAGCTCCGTCAACTGA	
164	TATTCCACTGGCAGTCCACGAATAGAGGTGGTGACA	
配列番号 トレーサー150	TTCGGTACCGAGCACGCATATGAACTCGTCGTAGATACTA	
165	GTAGATCACCGTAAGACCTTGCTGTGCGCGCCGTA	
配列番号 トレーサー151	AGAGCGCACTTAATGTCTCTGGAATGTTGCGTGAATCGCA	
166	GCGGCGTAAGTATGAGCAATCGTACCTCGGACCGT	
配列番号 トレーサー152	CTTAAGGTAGATCTTGTATCATGAATCACCAATTATGTAT	
167	GCATGCGGCGCGCACCTAAGAGCCTGTGAGATGTC	30
配列番号 トレーサー153	AGCGCTGTGGAACACGCAACGGTGTATGTCCACTCATTG	
168	TTCCGCAGTTGCGAAGTAGACAGGATCCTACTAAC	
配列番号 トレーサー154	CAGAAGCTCATGGCCTCAGTGCAGTTGCAAGGATGCGCC	
168	TGCATCTCGACTAGCAACTGTTCATACGTCATGGTC	
配列番号 トレーサー155	TCGCTCAATACACTTGGACCAGAATTATGTCCTATTCAGA	
169	ACCTTGCCGCGCGGCAGTCGCGCAGATGGTCCTGA	

## 【 0 1 9 9 】

## 分子 L I M S

実験室情報管理システム ( L I M S ) は、消耗品の消費および使用を追跡する方法であり、幾つかの場合には、与えられた実験に必要な化学物質または試薬、および与えられた実験に必要な化学物質または試薬のみがその実験に使用されたことを保証するための方法である。L I M S は、実験の各繰り返しに使用される化学物質のロット番号を追跡するのにも役立つ。これらの機能 (例えば、ロット番号の追跡) は全て、例えば単一の化学物質の品質が低下した場合または誤った試薬が実験で使用された場合、実験の失敗の問題解決を助けうる。

## 【 0 2 0 0 】

L I M S システムは、プロセスにおいて使用される各消耗品に関するカタログ番号およびロット番号を実験室員が入力する電子的またはウェブアプリケーションとして設計され

40

50

うる。典型的には、該プロセスを促進し、その精度を高めるために、バーコードが使用される。しかし、人為的な誤りが尚も、反応の所与反復に関する不完全な記録をもたらさうる。

#### 【0201】

試薬、特に試薬、試薬ロット、アリコート、または出荷品を分子的に標識する方法を本発明で提供する。幾つかの場合には、該方法は、種々の試薬の異なる容器を分子的にバーコード化するための、スパイクイン合成核酸の使用を含む。例えば、ユニーク配列（例えば、非ヒト、非病原体）を有するスパイクイン核酸または短い核酸オリゴマー（例えば、50～100bp）を各試薬、試薬ロット、試薬アリコートまたは試薬出荷品に添加することは、個々のライブラリーを製造するために使用されるために使用される試薬の在庫を追跡するのに役立つ。幾つかの場合には、1以上のIDスパイク（Spike）、スパーク（Spark）またはスパンク（Spank）配列が分子LIMSに使用されうる。ついで、各サンプルの処理において使用されるロット番号および試薬が配列決定によって自動的に検出されることが可能であり、例えば、成功した実施において使用されたロット番号と比較すること、またはそのサンプルの処理において使用される欠如している若しくは余分な試薬を特定することにより、問題のある実施の問題解決に使用されることが可能である。

10

#### 【0202】

同様に、特定の試薬、試薬ロット番号、アリコートまたは出荷品に関連するスパイクイン核酸の検出は、成功する配列決定実施において使用される試薬のロット番号、アリコートまたは出荷品を特定するために用いられうる。幾つかの場合には、核酸またはスパイクインは配列決定以外の方法により検出されることが可能であり、例えば、1以上の蛍光プローブで標識された一般的なポリマーが、蛍光を用いて検出されうる。

20

#### 【0203】

DNAオリゴマーは多数の水溶液に有効でありうるが、DNアーゼ作用を受けない核酸オリゴマー（例えば、RNA、修飾主鎖を有するDNAオリゴマー）がDNアーゼ含有溶液のために設計されうる。同様に、RNアーゼに抵抗性の合成核酸（例えば、DNA）が、RNアーゼ含有溶液を追跡するために使用されうる。

#### 【0204】

##### 核酸の富化およびライブラリーの製造

本発明で提供する方法においては、核酸は、当技術分野で公知の任意の手段を用いて、サンプルから単離されうる。例えば、核酸は、液体抽出（例えば、トリゾール（Trizol）、DNAzol）技術を用いて抽出されうる。核酸は、商業的に入手可能なキット [QIAamp循環核酸キット（Circulating Nucleic Acid Kit）、Qiagen DNeasyキット、QIAampキット、Qiagen Midiキット、QIAprep] を使用することによっても抽出されうる。

30

#### 【0205】

核酸は、単なる例示として遠心分離を含む公知方法により濃縮または沈殿されうる。核酸は、精製目的で、選択的膜（例えば、シリカ）に結合されうる。核酸は、所望の長さの断片、例えば、1000、500、400、300、200または100塩基対長未満の断片に関しても富化されうる。サイズに基づくそのような富化は、例えば、PEG誘導沈殿、電気泳動ゲルまたはクロマトグラフィー材（Huberら（1993）Nucleic Acids Res. 21:1061-6）、ゲル濾過クロマトグラフィーまたはTSKゲル（Katoら（1984）J. Biochem, 95:83-86）（それらの刊行物の全体をあらゆる目的で参照により本明細書に組み入れることとする）を使用して行われうる。

40

#### 【0206】

核酸サンプルは、標的ポリヌクレオチド、特に、炎症または感染に関連する標的核酸に関して富化されうる。幾つかの好ましい場合には、標的核酸は病原体核酸（例えば、無細胞病原体核酸）である。幾つかの好ましい場合には、標的核酸は、子宮、心臓、肺、腎臓

50

、胎児脳、肝臓または子宮頸組織（これらに限定されるものではない）を含む特定の器官または組織に関連する無細胞RNAである。

【0207】

標的富化は、当技術分野で公知の任意の手段によるものでありうる。例えば、核酸サンプルは、標的特異的プライマー（例えば、病原体核酸に特異的なプライマー）を使用して標的配列を増幅することにより富化されうる。標的増幅は、当技術分野で公知の任意の方法またはシステムを使用して、デジタルPCR形態で行われうる。核酸サンプルは、標的選択的オリゴヌクレオチドが固定化されたアレイ上に標的配列を捕捉することにより富化されうる。核酸サンプルは、溶液中に遊離した又は固体支持体上の標的選択的オリゴヌクレオチドにハイブリダイズさせることにより富化されうる。オリゴヌクレオチドは、捕捉試薬による捕捉を可能にする捕捉部分を含みうる。幾つかの実施形態においては、核酸サンプルは標的ポリヌクレオチドに関しては富化されず、例えば、全ゲノムを表す。

10

【0208】

幾つかの場合には、標的（例えば、病原体器官）核酸は、例えばプルダウン（pull-down）（例えば、ビオチンタグのような標識に結合した相補的オリゴヌクレオチドに標的核酸をハイブリダイズさせ、そして、例えば、固体支持体に結合したアビジンまたはストレプトアビジンを使用することにより、プルダウンアッセイにおいて標的核酸を優先的にプルダウンさせること）、標的化PCRまたは他の方法により、サンプル中のバックグラウンド（例えば、対象、健常組織）核酸に対して富化されうる。富化技術の例には、（a）核酸のサンプルにおける主要集団がサンプルにおける少数集団より迅速に自己ハイブリダイズする、自己ハイブリダイゼーション技術、（b）遊離DNAからのヌクレオソーム関連DNAの枯渇、（c）特定の長さ間隔のDNAを除去および/または単離すること、（d）エキソソーム枯渇または富化、ならびに（e）関心領域の戦略的捕捉が含まれるが、これらに限定されるものではない。

20

【0209】

幾つかの場合には、富化工程は、（a）宿主からの核酸のサンプルを準備すること（ここで、宿主からの核酸のサンプルは宿主からの一本鎖核酸のサンプルであり、宿主核酸および非宿主核酸を含む）、（b）宿主からの一本鎖核酸の少なくとも一部を再生させ、それにより、サンプルにおける二本鎖核酸の集団を生成させること、および（c）ヌクレアーゼを使用してサンプル内の二本鎖核酸の少なくとも一部を取り出し、それにより、宿主からの核酸のサンプルにおける非宿主配列を富化させることを含む。幾つかの場合には、富化工程は、（a）宿主からの核酸のサンプルを準備すること（ここで、宿主からの核酸のサンプルは、ヌクレオソームに関連する宿主核酸および非宿主核酸を含む）、および（b）ヌクレオソームに関連する宿主核酸の少なくとも一部を除去し、それにより、宿主からの核酸のサンプルにおける非宿主核酸を富化することを含む。幾つかの場合には、富化工程は、（a）宿主からの核酸のサンプルを準備すること（ここで、宿主からの核酸のサンプルは宿主核酸および非宿主核酸を含む）、および（b）1以上の長さ間隔のDNAを除去または単離し、それにより、宿主からの核酸のサンプルにおける非宿主核酸を富化することを含む。幾つかの場合には、富化工程は、（a）宿主からの核酸のサンプルを準備すること（ここで、宿主からの核酸のサンプルは宿主核酸、非宿主核酸およびエキソソームを含む）、および（b）エキソソームの少なくとも一部を除去または単離し、それにより、宿主からの核酸のサンプルにおける非宿主配列を富化することを含む。幾つかの場合には、富化工程は、約300塩基長以上の長さを有する核酸をサンプルから優先的に除去することを含む。幾つかの場合には、富化工程は、サンプルからの非宿主核酸を優先的に増幅または捕捉することを含む。

30

40

【0210】

富化工程は、約120、約150、約200または約250塩基長以上の核酸をサンプルから優先的に除去することを含みうる。幾つかの場合には、富化工程は、約10塩基～約60塩基長、約10塩基～約120塩基長、約10塩基～約150塩基長、約10塩基～約300塩基長、約30塩基～約60塩基長、約30塩基～約120塩基長、約30塩

50

基～約150塩基長、約30塩基～約200塩基長、または約30塩基～約300塩基長の、サンプルからの核酸を優先的に富化することを含む。幾つかの場合には、富化工程は、宿主（例えば、対象）からの核酸を優先的に消化させることを含む。幾つかの場合には、富化工程は、非宿主核酸を優先的に複製させることを含む。

【0211】

幾つかの場合には、富化工程は、宿主（例えば、対象）核酸に対する非宿主核酸の比率を、少なくとも2倍、少なくとも3倍、少なくとも4倍、少なくとも5倍、少なくとも6倍、少なくとも7倍、少なくとも8倍、少なくとも9倍、少なくとも10倍、少なくとも11倍、少なくとも12倍、少なくとも13倍、少なくとも14倍、少なくとも15倍、少なくとも16倍、少なくとも17倍、少なくとも18倍、少なくとも19倍、少なくとも20倍、少なくとも30倍、少なくとも40倍、少なくとも50倍、少なくとも60倍、少なくとも70倍、少なくとも80倍、少なくとも90倍、少なくとも100倍、少なくとも1000倍、少なくとも5000倍または少なくとも10,000倍増加させる。幾つかの場合には、富化工程は、宿主（例えば、対象）核酸に対する非宿主核酸の比率を少なくとも10倍増加させる。幾つかの場合には、富化工程は宿主（例えば、対象）核酸に対する非宿主核酸の比率を約10倍～約100倍増加させる。

10

【0212】

幾つかの場合には、核酸ライブラリーを製造する。核酸ライブラリーは一本鎖核酸ライブラリーまたは二本鎖核酸ライブラリーでありうる。幾つかの場合には、一本鎖核酸ライブラリーは一本鎖DNAライブラリー（ssDNAライブラリー）またはRNAライブラリーでありうる。幾つかの場合には、二本鎖核酸ライブラリーは二本鎖DNAライブラリー（dsDNAライブラリー）である。ssDNAライブラリーの製造方法は、二本鎖DNA断片をssDNA断片に変性させ、プライマードッキング配列をssDNA断片の一端に連結し、プライマーを該プライマードッキング配列にハイブリダイズさせることを含む。プライマーは、次世代シーケンシングプラットフォームと組合されるアダプター配列の少なくとも一部を含む。該方法は更に、ハイブリダイズしたプライマーを伸長させて二本鎖を生成させることを含むことが可能であり、ここで、該二本鎖は元のssDNA断片および伸長プライマー鎖を含む。該伸長プライマー鎖は元のssDNA断片から分離されうる。該伸長プライマー鎖は回収可能であり、ここで、該伸長プライマー鎖はssDNAライブラリーのメンバーである。RNAライブラリーの製造方法は、プライマードッキング配列をRNA断片の一端に連結し、プライマーをプライマードッキング配列にハイブリダイズさせることを含む。プライマーは、次世代シーケンシングプラットフォームと組合されるアダプター配列の少なくとも一部を含む。該方法は更に、ハイブリダイズしたプライマーを伸長させて二本鎖を生成させることを含むことが可能であり、ここで、該二本鎖は元のRNA断片および伸長プライマー鎖を含む。該伸長プライマー鎖は元のRNA断片から分離されうる。該伸長プライマー鎖は回収可能であり、ここで、該伸長プライマー鎖はRNAライブラリーのメンバーである。dsDNAライブラリーの製造方法は、アダプター配列をdsDNA断片の一端または両端に連結することを含む。

20

30

【0213】

種々の態様においては、dsDNAは、当技術分野で公知の又は本明細書に記載されている任意の手段により断片化されうる。幾つかの場合においては、dsDNAは物理的手段（例えば、機械的剪断、噴霧または超音波処理）、酵素的手段または化学的手段により断片化されうる。

40

【0214】

幾つかの実施形態においては、RNAからcDNAを生成させる。例えば、ランダムプライム逆転写（RNアーゼH<sup>+</sup>）を用いて、ランダムなサイズのcDNAを得ることにより、cDNAを生成させることが可能である。

【0215】

核酸の長さは様々でありうる。核酸または核酸断片（例えば、dsDNA断片、RNA

50



、またはランダムなサイズのcDNA)は、1000bp未満、800bp未満、700bp未満、600bp未満、500bp未満、400bp未満、300bp未満、200bp未満または100bp未満である。DNA断片は約40~約100bp、約50~約125bp、約100~約200bp、約150~約400bp、約300~約500bp、約100~約500、約400~約700bp、約500~約800bp、約700~約900bp、約800~約1000bp、または約100~約1000bpである。幾つかの場合には、核酸または核酸断片(例えば、dsDNA断片、RNA、またはランダムなサイズのcDNA)は約20~約200bpの範囲内、例えば約40~約100bpの範囲内でありうる。

#### 【0216】

dsDNA断片の末端はポリッシュ(polish)(例えば、平滑末端化)されうる。DNA断片の末端はポリメラーゼでの処理によりポリッシュされうる。ポリッシュは、3'オーバーハング(突出)の除去、5'オーバーハングの補充(フィルイン; fill-in)またはそれらの組合せを含みうる。ポリメラーゼはブルーフリーディングポリメラーゼ(例えば、3'から5'へのエキソヌクレアーゼ活性を含む)でありうる。ブルーフリーディングポリメラーゼは、例えば、T4 DNAポリメラーゼ、Pol Iクレノー断片またはPfuポリメラーゼでありうる。ポリッシュは、当技術分野で公知の任意の手段を用いて損傷ヌクレオチド(例えば、脱塩基部位)を除去することを含みうる。

#### 【0217】

核酸断片の3'末端へのアダプターの連結は該断片の3'OH基とアダプターの5'ホスファートとの間の結合の形成を含みうる。したがって、核酸断片からの5'ホスファートの除去は2つのライブラリーメンバーの異常連結を最小限に抑えうる。したがって、幾つかの実施形態においては、5'ホスファートが核酸断片から除去される。幾つかの実施形態においては、5'ホスファートはサンプルにおける核酸断片の少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%または95%超から除去される。幾つかの実施形態においては、実質的に全てのホスファート基(リン酸基)が核酸断片から除去される。幾つかの実施形態においては、実質的に全てのホスファートがサンプルにおける核酸断片の少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%または95%超から除去される。核酸サンプルからのホスファート基の除去は、当技術分野で公知の任意の手段によるものでありうる。ホスファート基の除去は、サンプルを熱不安定性ホスファターゼで処理することを含みうる。幾つかの実施形態においては、ホスファート基は核酸サンプルから除去されない。幾つかの実施形態においては、核酸断片の5'末端へのアダプターの連結が行われる。

#### 【0218】

##### 配列決定(シーケンシング)

本開示は核酸の分析方法を提供する。そのような分析方法は核酸の配列決定および配列決定結果の生物情報学的(バイオインフォマティクス)分析を含む。本方法により得られた核酸は、ゲノム、エピジェネティック(例えば、メチル化)およびRNA発現を含む種々のタイプの情報を得るために分析されうる。メチル化分析は、例えば、メチル化塩基の変換およびそれに続くDNA配列決定により行われうる。RNA発現分析は、例えば、ポリヌクレオチドアレイハイブリダイゼーション、RNA配列決定技術、またはRNAから生成されたcDNAの配列決定により行われうる。

#### 【0219】

好ましい実施形態においては、配列決定は、次世代シーケンシングアッセイを用いて行われる。本明細書中で用いる「次世代」なる語は当技術分野において十分に理解されており、一般に、限定的なものではないが以下の1以上を含む任意のハイスループット配列決定アプローチを意味する：大規模並列シグネチャー配列決定、ピロシーケンス(例えば、Roche 454配列決定装置を使用するもの)、イルミナ(Illumina)(ソレクサ(Solexa))配列決定、合成による配列決定(Illumina)、イオントレント(Ion torrent)配列決定、連結による配列決定(例えば、SOLID

10

20

30

40

50

配列決定)、単分子リアルタイム(SMRT)配列決定(例えば、Pacific Bioscience)、ポロニー(polony)配列決定、DNAナノボール(nanoball)配列決定、ヘリスコープ単分子(heliscopesingle molecule)配列決定(Helicos Biosciences)およびナノポア(nanopore)配列決定(例えば、Oxford Nanopore)。幾つかの場合には、配列決定アッセイはナノポア配列決定を用いる。幾つかの場合には、配列決定アッセイはサンガー配列決定の幾つかの形態を含む。幾つかの場合には、配列決定はショットガン配列決定を含み、幾つかの場合には、配列決定はブリッジ(bridge)PCRを含む。幾つかの場合には、配列決定は広域スペクトルである。幾つかの場合には、配列決定は標的化される。

10

**【0220】**

幾つかの場合には、配列決定アッセイはギルバートの配列決定法を含む。そのようなアプローチにおいては、核酸(例えば、DNA)を化学修飾し、ついで特定の塩基において切断する。幾つかの場合には、配列決定アッセイはジデオキシヌクレオチド鎖終結またはサンガー配列決定を含む。

**【0221】**

本発明で提供する方法においては、合成による配列決定アプローチが用いられうる。幾つかの場合においては、蛍光標識可逆的ターミネーターヌクレオチドを、ガラスフローセルの表面上に固定化されたクローン増幅DNA鑄型に導入する。各配列決定サイクル中に、単一の標識デオキシヌクレオチド三リン酸(dNTP)が核酸鎖に付加されうる。標識ターミネーターヌクレオチドは、塩基を特定するために、添加されたらイメージングされることが可能であり、ついで酵素的に切断されて、次のヌクレオチドの取り込みを可能にしうる。全4個の可逆的ターミネーター結合dNTP(A、C、T、G)は、一般に、単一の分離した分子として存在するため、自然競合が取り込みバイアスを最小にしうる。

20

**【0222】**

幾つかの場合には、単一分子リアルタイム(SMRT)と称される方法を用いる。そのようなアプローチにおいては、核酸(例えば、DNA)をゼロモード導波路(ZMW)において合成する。ZMWは、ウェルの底部に位置する捕捉手段を含有する小さなウェル様容器である。未修飾ポリメラーゼ(ZMW底部に結合している)および溶液中で自由流動する蛍光標識ヌクレオチドを使用して、配列決定を行う。蛍光標識は、DNA鎖内への取り込みに際してヌクレオチドから分離され、未修飾DNA鎖が残る。ついで、カメラのような検出器を使用して、発光を検出することが可能である。データを生物情報学的に分析して、配列情報を得ることが可能である。

30

**【0223】**

幾つかの場合においては、連結アプローチによる配列決定を用いて、サンプルにおける核酸を配列決定する。一例としては、SOLID[オリゴヌクレオチド連結および検出による配列決定(Sequencing by Oligonucleotide Ligation and Detection)]配列決定(Life Technologies)の次世代シーケンシング法が挙げられる。この次世代技術は数億個~数十億個の小さな配列リード(read)を同時に生成しうる。該配列決定方法は、配列決定されるべきサンプルからDNA断片のライブラリーを製造することを含みうる。幾つかの場合には、該ライブラリーを使用して、各ビーズ(例えば、磁気ビーズ)の表面上にただ1つの種の断片が存在するクローンビーズ集団を調製する。磁気ビーズに結合した断片は、各断片の出発配列が既知かつ同一となるように結合した汎用P1アダプター配列を有しうる。幾つかの場合には、該方法は更に、PCRまたはエマルジョンPCRを含みうる。例えば、エマルジョンPCRは、PCR用の試薬を含有するマイクロリアクターの使用を含みうる。ついで、ビーズに結合した得られたPCR産物をガラススライドに共有結合させることが可能である。配列決定アッセイ、例えばSOLID配列決定アッセイまたは連結アッセイによる他の配列決定は、プライマーの使用を含む工程を含みうる。プライマーはP1アダプター配列またはライブラリー鑄型内の他の配列にハイブリダイズしうる。該方法は

40

50

更に、配列決定プライマーへの連結に関して競合する4つの蛍光標識二塩基プローブを導入することを含みうる。該二塩基プローブの特異性は、各連結反応における各第1および第2塩基をイントロゲート ( i n t e r r o g a t e ) することにより達成されうる。連結、検出および切断の複数のサイクルが、最終的なリード長を決定するサイクル数で行われうる。幾つかの場合には、一連の連結サイクルの後、伸長産物を除去し、第2ラウンドの連結サイクルのために、n - 1位に相補的なプライマーを使用して、鋳型を再配置する。各配列タグのために、複数ラウンド (例えば、5ラウンド) のプライマー再配置が完了されうる。プライマー再配置プロセスを通じて、各塩基は、2つの異なるプライマーによる2つの独立した連結反応においてイントロゲートされうる。例えば、リード位置5における塩基は、連結サイクル2においてはプライマー番号2によって、そして連結サイクル1においてはプライマー番号3によってアッセイされる。

10

#### 【0224】

該実施形態のいずれかにおいては、オリゴヌクレオチドの検出または定量分析は配列決定により達成されうる。サブユニットまたは全合成オリゴヌクレオチドは、本明細書に記載されている配列決定方法を含む当技術分野で公知の任意の適切な方法 (例えば、Illumina HiSeq 2500) による全オリゴヌクレオチドの完全な配列決定により検出されうる。

#### 【0225】

配列決定は、当技術分野でよく知られた古典的なサンガー配列決定法により達成される。配列決定は、ハイスループット系を使用することによっても達成可能であり、それらの中の幾つかは、配列決定ヌクレオチドが、成長中の鎖内へのその取り込みの直後または該取り込みの際して検出されること (例えば、リアルタイムまたは実質的にリアルタイムでの配列の検出) を可能にする。幾つかの場合には、ハイスループット配列決定は、1時間当たり少なくとも1,000、少なくとも5,000、少なくとも10,000、少なくとも20,000、少なくとも30,000、少なくとも40,000、少なくとも50,000、少なくとも100,000または少なくとも500,000個の配列リード ( r e a d ) を生成する。幾つかの場合には、各リードはリード当たり少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも120または少なくとも150塩基である。幾つかの場合には、各リードはリード当たり多くとも2000、多くとも1000、多くとも900、多くとも800、多くとも700、多くとも600、多くとも500、多くとも400、多くとも300、多くとも200または多くとも100塩基である。長いリード配列決定は、例えば、500塩基より長い、800塩基より長い、1000塩基より長い、1500塩基より長い、2000塩基より長い、3000塩基より長い、または4500より長い連続配列リードを与える配列決定を含みうる。

20

30

#### 【0226】

幾つかの場合には、ハイスループット配列決定は、IlluminaのGenome Analyzer IIX、MiSeqパーソナルシーケンサー、またはHiSeqシステム、例えば、HiSeq 2500、HiSeq 1500、HiSeq 2000もしくはHiSeq 1000を使用するものにより利用可能な技術の使用を含む。これらの装置は、合成化学による可逆的ターミネーターベース配列決定を利用する。これらの装置は8日間で2,000億個以上のDNAの読取りを行いうる。より小さいシステムは、3日、2日もしくは1日以内またはそれより短い時間内の実施のために利用されうる。短い合成サイクルは、配列決定結果を得るために要する時間を最小するために用いられうる。

40

#### 【0227】

幾つかの場合には、ハイスループット配列決定は、ABI Solid Systemにより利用可能な技術の使用を含む。この遺伝子解析プラットフォームは、ビーズに結合したクローン増幅DNA断片の大規模並列配列決定を可能にしうる。該配列決定方法は色素標識オリゴヌクレオチドとの連続的連結に基づく。

50

## 【0228】

次世代シーケンシングはイオン半導体配列決定（例えば、Life Technologies (Ion Torrent) の技術を利用するもの) を含む。イオン半導体配列決定は、ヌクレオチドがDNA鎖内に取り込まれる際にイオンが放出されうるという事実を利用する。イオン半導体配列決定を行うために、微細加工ウェルの高密度アレイを形成させることが可能である。各ウェルは単一DNA鋳型を保持する。ウェルの下にはイオン感受性層が存在することが可能であり、イオン感受性層の下にはイオンセンサーが存在することが可能である。ヌクレオチドがDNAに加えられると、H<sup>+</sup>が放出される可能性があり、これはpHの変化として測定される。H<sup>+</sup>イオンは電圧に変換され、半導体センサにより記録される。アレイチップは1ヌクレオチドずつ連続的に水に浸され

10

## 【0229】

幾つかの場合には、ハイスループット配列決定は、Helicos Biosciences Corporation (Cambridge, Massachusetts) により利用可能な技術、例えば、合成による単一分子配列決定 (Single Molecule Sequencing by Synthesis) (SMS) 法の使用を含む。SMSは、ヒトゲノム全体を24時間以内に配列決定することを可能にする。SMSは、MIP技術と同様に、ハイブリダイゼーションの前に増幅工程を要しないであろう。SMSは増幅を要しないであろう。SMSは、米国特許出願公開第20060024711号、第20060024678号、第20060012793号、第20060012784号および第20050100932号に部分的に記載されている。

20

## 【0230】

幾つかの場合には、ハイスループット配列決定は、454 Lifesciences、Inc. (Branford, Connecticut) により利用可能な技術、例えばPico Titer Plate装置の使用を含み、該装置は、該装置内のCCDカメラにより記録される、配列決定反応により生成された化学発光シグナルを伝達する光ファイバプレートを含む。光ファイバーのこの使用は4.5時間で少なくとも2,000万塩基対の検出を可能にする。

30

## 【0231】

ビーズ増幅およびそれに続く光ファイバー検出を使用するための方法は、Margules, M.ら、"Genome sequencing in microfabricated high-density picolitre reactors", Nature, doi:10.1038/nature03959、ならびに米国特許出願公開第20020012930号、第20030058629号、第20030100102号、第20030148344号、第20040248161号、第20050079510号、第20050124022号および第20060078909号に記載されている。

40

## 【0232】

幾つかの場合には、ハイスループット配列決定は、Clonal Single Molecule Array (Solexa, Inc.) を用いて、または可逆的ターミネーター化学を利用する合成による配列決定 (SBS) を用いて行われる。これらの技術は、米国特許第6,969,488号、第6,897,023号、第6,833,246号、第6,787,308号、および米国特許出願公開第20040106110号、第20030064398号、第20030022207号、およびConstans, A., The Scientist 2003, 17(13):36に部分的に記載されている。

50



場合には、ナノポア配列決定技術はIBM/Rochéからのものである。電子ビームを用いて、マイクロチップ内にナノポアサイズの開口部を生成させることが可能である。電場を用いて、ナノポアを介してDNAを引っ張り、または通すことが可能である。ナノポア内のDNAトランジスタデバイスは金属および誘電体の交互のナノメートルサイズの層を含みうる。DNA骨格内の孤立電荷はDNAナノポア内の電場により捕捉されうる。ゲート電圧をオフおよびオンにすることにより、DNA配列を読取ることが可能となりうる。

#### 【0235】

次世代シーケンシングはDNAナノボール(nanoball)配列決定(例えば、Complete Genomicsにより行われるもの; 例えば、Drmanac et al (2010) Science 327:78-81を参照されたい)を含みうる。DNAを単離し、断片化し、サイズ選択することが可能である。例えば、DNAは(例えば、超音波処理により)約500bpの平均長に断片化されうる。アダプター(Ad1)が断片の末端に結合される。アダプターを使用して、配列決定反作用アンカーにハイブリダイズさせることが可能である。各末端に結合したアダプターを有するDNAをPCR増幅することが可能である。アダプター配列は、相補的な一本鎖末端が互いに結合して環状DNAを形成するように修飾されうる。DNAは、後続工程で使用されるIIS型制限酵素による切断から保護するためにメチル化されうる。アダプター(例えば、右アダプター)は制限認識部位を有することが可能であり、制限認識部位は非メチル化状態のままでありうる。アダプター内の非メチル化制限認識部位は制限酵素(例えば、AclI)により認識されることが可能であり、該DNAは、右アダプターの13bp右側で、AclIにより切断されて、線状二本鎖DNAを形成しうる。第2ラウンドの右および左アダプター(Ad2)を該線状DNAのいずれかの末端に連結することが可能であり、両方のアダプターが結合された全てのDNAを(例えばPCRにより)PCR増幅することが可能である。Ad2配列は、それらが互いに結合して環状DNAを形成するように修飾されうる。該DNAはメチル化されうるが、制限酵素認識部位は左Ad1アダプター上で非メチル化状態のままでありうる。制限酵素(例えば、AclI)が適用可能であり、該DNAはAd1の13bp左側で切断されて線状DNA断片を形成しうる。第3ラウンドの右および左アダプター(Ad3)を該線状DNAの右および左側に連結することが可能であり、得られた断片をPCR増幅することが可能である。該アダプターは、それらが互いに結合して環状DNAを形成しうるように修飾されうる。III型制限酵素(例えば、EcoPI5)が添加されうる。EcoPI5はAd3の26bp左側およびAd2の26bp右側でDNAを切断しうる。この切断は大きなDNAセグメントを除去し、DNAを再び線状化しうる。第4ラウンドの右および左アダプター(Ad4)をDNAに連結し、DNAを(例えばPCRにより)増幅し、修飾して、それらが互いに結合し、完全な環状DNA鋳型を形成するようにすることが可能である。

#### 【0236】

ローリングサークル複製(例えば、Phi29 DNAポリメラーゼを使用するもの)を用いて、DNAの小さな断片を増幅することが可能である。4つのアダプター配列は、ハイブリダイズ可能な回文配列を含むことが可能であり、単一鎖がそれ自体に対してフォールディングして、約200~300ナノメートルの平均直径を有しうるDNAナノボール(DNB(商標))を形成しうる。DNAナノボールはマイクロアレイ(配列決定フローセル)に(例えば吸着により)結合されうる。該フローセルは、二酸化ケイ素、チタンおよびヘキサメチルジシラザン(HMDS)でコーティングされたシリコンウェハならびにフォトレジスト材でありうる。配列決定は、蛍光プローブをDNAに連結することにより、非連鎖(unchained)配列決定により行われうる。イントロゲート(introgated)位置の蛍光の色は高分解能カメラにより可視化されうる。アダプター配列間のヌクレオチド配列の同一性が決定されうる。

#### 【0237】

本発明で提供する方法はシステムの使用を含むことが可能であり、例えば、DNAまた

10

20

30

40

50

はRNA配列情報を得るための核酸シーケンサー（例えば、DNAシーケンサー、RNAシーケンサー）を含むシステムの使用を含みうる。該システムは、DNAまたはRNA配列情報に関するバイオインフォマティクス（生物情報学的）分析を実行するソフトウェアを含むコンピュータを含みうる。バイオインフォマティクス分析には、限定的なものではないが、配列データの構築、サンプルにおける遺伝的変異体〔生殖系列変異体および体細胞変異体（例えば、癌または前癌状態に関連する遺伝的変異、感染に関連する遺伝的変異）を含む〕の検出および定量が含まれる。

#### 【0238】

配列データを使用して、遺伝子配列情報、倍数性状態、1以上の遺伝的変異体の同一性、および変異体の定量的尺度（相対的および絶対的相対尺度を含む）を決定することが可能である。

10

#### 【0239】

幾つかの場合には、ゲノムの配列決定は全ゲノム配列決定または部分ゲノム配列決定を含む。配列決定は不偏性（unbiased）であることが可能であり、サンプルにおける核酸の全てまたは実質的に全て（例えば、70%、80%、90%を超える）の配列決定を含みうる。ゲノムの配列決定は選択的であることが可能であり、例えば、関心のあるゲノムの部分に向けられたものでありうる。例えば、多数の遺伝子（およびこれらの遺伝子の変異形態）は種々の癌に関連していることが公知である。所望の分析のためには、選択された遺伝子または遺伝子の一部の配列決定で十分でありうる。関心のある対象である、ゲノム内の特定の遺伝子座に位置決定（マッピング）されたポリヌクレオチドは、例えば配列捕捉または部位特異的増幅により、配列決定のために単離されうる。

20

#### 【0240】

##### 用途

本発明で提供する方法は種々の目的に使用可能であり、例えば、状態（例えば、感染）の診断または検出、状態の発生または再発の予測、治療のモニター、治療レジメンの選択または修飾、あるいは療法の最適化に使用されうる。このアプローチにより、治療および/または診断レジメンを、治療経過にわたる種々の時点で得られたデータに従い個別化および調整し、それにより、個別に適切なレジメンを提供することが可能である。

#### 【0241】

##### 状態の検出 / 診断 / 予後判定

本発明で提供する方法は、患者のサンプル、例えばヒトの血液サンプルにおいて感染または疾患を検出、診断または予後判定するために使用されうる。該方法は、ヒト核酸から主に構成されるサンプルにおける希少微生物核酸断片を検出するために使用されうる。例えば、血液中の無細胞DNA（cfDNA）は、主に宿主由来のDNA断片からなるが、体内の微生物由来の少量の断片をも含有する。cfDNAの抽出およびそれに続く詳細な配列決定（例えば、次世代シーケンシングまたはNGS）は、宿主および非宿主ゲノムデータベースに対して位置決定（マッピング）されうる数百万個または数十億個の配列リード（read；読取り）を生成しうる。同様に、該方法は、特定の器官からの循環または無細胞RNAの希少集団を検出するためにも使用されうる。非宿主リードが全体のうちのごく少数であるサンプルの場合、本発明で提供する方法によらなければ、異なる標的核酸（例えば、異なる微生物または生物に由来するもの）と比較するための又は異なるサンプルもしくは試薬を追跡するための内部正規化標準の欠如により損なわれるアッセイの感度および特異性を、本発明で提供する方法は改善しうる。また、該方法は、標的核酸が核酸の全集団のより大きな部分を構成する場合に使用されうる。

30

40

#### 【0242】

本発明で提供する方法は、多種多様な疾患および障害を検出、モニター、診断、予後判定、治療または予防するために使用されうる。特に、該方法は、感染性疾患または障害に関連する病原体に由来する1以上の標的核酸を検出するために使用されうる。典型的な疾患および障害には、感染に関連する任意の疾患または障害、例えば、敗血症、肺炎、結核、HIV感染、肝炎感染（例えば、Hep A、BまたはC）、ヒトパピローマウイルス

50

(HPV)感染、クラミジア感染、梅毒感染、エボラ感染、スタフィロコッカス・アウレウス (*staphylococcus aureus*) 感染またはインフルエンザが含まれる。本発明で提供する方法は、薬剤耐性微生物、例えば、多剤耐性微生物、または容易には培養されず若しくは典型的には試験されない微生物による感染を検出するのに特に有用である。本方法で検出されうる疾患および障害の幾つかの非限定的な例には以下のものが含まれる：癌、拡張型心筋症、ギラン・バレー症候群、多発性硬化症、結核、炭疽病、睡眠病、赤痢、トキソプラズマ症、白癬、カンジダ症、ヒストプラズマ症、エボラ、アシネトバクター感染、放線菌症、アフリカ睡眠病 (アフリカトリパノソーマ症)、エイズ (後天性免疫不全症候群)、HIV感染、アメーバ症、アナプラズマ症、炭疽病、アルカンバクテリウム・ヘモリティカム (*Arcanobacterium haemolyticum*) 感染、アルゼンチン出血熱、アスカリアシス、アスペルギルス症、アストロウイルス感染、バベシア症、バシルス・セレウス (*Bacillus cereus*) 感染、細菌性肺炎、細菌性膣症 (BV)、バクテロイデス感染、バランチジウム症、ベイリサスカリス (*Baylisascaris*) 感染、BKウイルス感染、黒色砂毛、胚盤胞感染、胚盤葉感染、ボリビア出血熱、ボレリア感染、ボツリヌス中毒 (および幼児ボツリヌス中毒症)、ブラジル出血熱、ブルセラ症、腺ペスト、パークホルデリア感染、ブルリ潰瘍、カリシウイルス感染 (ノロウイルスおよびサポウイルス)、カンピロバクター症、カンジダ症 (モニリア症；膣炎)、猫引っかき病、蜂巣炎、シャガス病 (アメリカトリパノソーマ症)、軟性下疳、水痘、チクングニア熱、クラミジア、クラミドフィラ・ニューモニエ (*Chlamydia pneumoniae*) 感染 (台湾急性呼吸器症候群またはTWAR)、コレラ、クロモブラストミコーシス、肝吸虫症、クロストリジウム・ディフィシレ (*Clostridium difficile*) 感染、コクシジオイデス症、コロラドダニ熱 (CTF)、感冒 (急性ウイルス性鼻咽頭炎；急性コリーザ)、クロイツフェルト・ヤコブ病 (CJD)、クリミア・コンゴ出血熱 (CCHF)、クリプトコックス症、クリプトスポリジウム症、皮膚幼虫遊走 (CLM)、シクロスポラ症、嚢虫症、サイトメガロウイルス感染、デング熱、二核アメーバ症、ジフテリア、裂頭条虫症、メジナ虫症、エボラ出血熱、エキノコックス症、エーリキア症、蟻虫症 (蟻虫感染症)、腸球菌感染症、エンテロウイルス感染、流行チフス、伝染性紅斑 (第五病)、突発性発疹 (第六病)、肥大吸虫症、肝蛭症、フィラリア症、クロストリジウム・パーFRINGENS (*Clostridium perfringens*) による食中毒、自由生活アメーバ感染、フゾバクテリウム感染、ガス壊疽 (クロストリジウム筋壊死)、ゲオトリウム症、ゲルストマン-シュトロイスラー-シャインカー症候群 (GSS)、ジアルジア症、鼻疽、顎口虫症、淋病、鼠径肉芽腫 (ドノヴァン症)、A群連鎖球菌感染、B群レンサ球菌感染、ヘモフィルス・インフルエンザ (*Haemophilus influenzae*) 感染、手足口病 (HFMD)、ハンタウイルス肺症候群 (HPS)、ハートランド (*Heartland*) ウイルス疾患、ヘリコバクター・ピロリ (*Helicobacter pylori*) 感染、溶血性尿毒症症候群 (HUS)、腎症候性出血熱 (HFRS)、A型肝炎、B型肝炎、C型肝炎、D型肝炎、E型肝炎、単純ヘルペス、ヒストプラズマ症、鉤虫感染、ヒトボカウイルス感染、ヒト・エウイングイェールリキオシス (*ewingii ehrlichiosis*)、ヒト顆粒球アナプラズマ症 (HGA)、ヒトメタニューモウイルス感染、ヒト単球エールリヒア症、ヒトパピローマウイルス (HPV) 感染、ヒトパラインフルエンザウイルス感染、膜様条虫症、エプスタイン-バーウイルス感染性単核球症 (Mono)、インフルエンザ (flu)、イソスポラ症、川崎病、角膜炎、キングラ・キング (*Kingella kingae*) 感染、クーラー、ラッサ熱、レジオネラ症 (レジオネラ病)、レジオネラ症 (ポンティアック熱)、リーシュマニア症、ハンセン病、レプトスピラ症、リステリア症、ライム病 (ライムボレリア症)、リンパ管フィラリア症 (象皮病)、リンパ球性脈絡髄膜炎、マラリア、マールブルグ出血熱 (MHF)、麻疹、中東呼吸器症候群 (MERS)、類鼻疽 (ホイットモア病)、髄膜炎、髄膜炎菌性疾患、横川吸虫症、微孢子虫症、伝染性軟属腫 (MC)、サル痘、流行性耳下腺炎、マウスチフス (流行チフス)、マイコプラズマ肺炎、菌腫、ハエ幼虫症、新生児結膜炎 (新

10

20

30

40

50



生児眼炎)、(新)変種クロイツフェルト・ヤコブ病(vCJD、nvCJD)、ノカルジア症、オンコセルカ症(河川盲目症)、パラコクシジオイデス(南米プラストミセス症)、肺吸虫症、パスツレラ症、アタマジラミ寄生症(アタマジラミ)、コロモジラミ寄生症(着物虱)、ケジラミ症(恥毛シラミ、カニシラミ)、骨盤内炎症性疾患(PID)、百日咳(百日ぜき)、ベスト、肺炎球菌感染、ニューモシスチス肺炎(PCP)、肺炎、ポリオ、プレボテラ感染、原発性アメーバ性髄膜脳炎(PAM)、進行性多巣性白質脳症、オウム病、Q熱、狂犬病、呼吸器合胞体ウイルス感染、リスノポリジウム症、ライノウイルス感染、リケッチア感染、リケッチア痘、リフトバレー熱(RVF)、ロッキー山紅斑熱(RMSF)、ロタウイルス感染、風疹、サルモネラ症、SARS(重症急性呼吸器症候群)、疥癬、住血吸虫症、敗血症、細菌性赤痢(細菌赤痢)、帯状疱疹(帯状ヘルペス)、天然痘(痘瘡)、スポロトリクス症、ブドウ球菌食中毒、ブドウ球菌感染、糞線虫症、亜急性硬化性全脳炎、梅毒、テニア症、破傷風(咬痘)、白癬性毛瘡(床屋痒み症)、頭部白癬(しらくも)、体部白癬(ぜにたむし)、股部白癬(いんきんたむし)、手白癬(手の白癬)、黒癬、足白癬(水虫)、爪白癬(爪真菌症)、澱風(なまず)、トキソカラ症(眼幼虫移行症(OLM))、トキソカラ症(内臓幼虫移行症(VLM))、トラコーマ、トリノククリアシス(Trinocchiasis)、旋毛虫病(Trichinellosis)、鞭虫症(鞭虫感染)、結核、野兔病、腸チフス、ウレアプラズマ・ウレアリチカム(Ureaplasma urealyticum)感染、渓谷熱、ベネズエラウマ脳炎、ベネズエラ出血熱、ウイルス性肺炎、西ナイル熱、白色砂毛(チネア・ブランカ(Tinea blanca)、エルシニア・シュードツベルクローシス(Yersinia pseudotuberculosis)感染、エルシニア症、黄熱病、ジカウイルス、および接合菌症。

10

20

## 【0243】

幾つかの場合には、本明細書に記載されている方法は、感染が活動性であるか潜在性であるかを決定することを含む。幾つかの場合には、遺伝子発現の定量は、活動性感染を検出、予測、診断またはモニターするための方法を提供しうる。幾つかの場合には、本明細書に記載されている方法は、活動性感染を検出することを含む。幾つかの場合には、遺伝子発現は、関心のある1以上の標的核酸の検出または配列決定により定量されうる。幾つかの場合には、遺伝子発現の定量は、潜伏感染を検出、予測、診断またはモニターするための方法を提供しうる。幾つかの場合には、本明細書に記載されている方法は、潜伏感染を検出することを含む。

30

## 【0244】

本発明で提供する方法は、癌を検出するために使用可能であり、特に、そのような癌を有する対象、そのような癌を有するリスクのある対象、またはそのような癌を有すると疑われる対象において、癌を検出するために使用されうる。癌の例には、限定的なものではないが、脳腫瘍、頭頸部癌、喉頭癌、口腔癌、乳癌、骨癌、血液癌、白血病、リンパ腫、肺癌、腎癌、膵臓癌、胃癌、結腸癌、直腸癌、皮膚癌、生殖路の癌、前立腺癌などが含まれる。幾つかの場合には、本発明で提供する方法は、非血液癌、例えば、実質臓器の癌(例えば、肺癌、乳癌、膵臓癌など)を検出するのに特に有用である。

40

## 【0245】

該方法は、対象の任意の他のタイプの疾患または状態を検出するのにも有用でありうる。しばしば、それは、希少遺伝的変異を検出するのに有用であり、あるいは、サンプルにおける全核酸集団のうちの非常に小さな部分のみを構成する核酸配列を検出するのに有用である。

## 【0246】

病原体または器官核酸の検出は、病原体または器官核酸の存在もしくは非存在および/または病原体もしくは器官核酸の量を決定するために、病原体または器官核酸のレベルを対照または参照値と比較することを含みうる。レベルは定性的または定量的レベルでありうる。幾つかの場合には、対照または参照値は、無細胞病原体核酸または無細胞器官由来核酸の存在または非存在を示す所定の絶対的な値である。例えば、対照値を超える無細胞

50

病原体核酸のレベルの検出は病原体または感染の存在を示しうる。一方、対照値より低いレベルは病原体または感染の非存在を示しうる。該対照値は、感染を有さない対象の無細胞核酸レベルを分析することにより得られる値でありうる。幾つかの場合には、対照値は陽性対照値であることが可能であり、特定の感染を有する、または特定の器官の特定の感染を有する対象からの無細胞核酸を分析することにより得られうる。

【0247】

幾つかの場合には、感染が存在するか否かを決定するために、そしてしばしば、正確な結果を得るために、以下の方法の1以上が適用されうる：(i) WO 2015070086 A1の特許に記載されているとおり、配列決定により得られたリード(read)の全体はキュレート(curated)宿主ゲノム参照データベース(これはヒト、イヌ、ネコ、霊長類由来または任意の他の宿主由来であることが可能であり、例えば、GenBank hg19ヒト参照配列を含む)に対してアライメントされうる；(ii) 病原体関連配列を含む非宿主配列のみが更に分析されうるように、バイオインフォマティクス分析のためのデータプロセッサは宿主配列を除去または隔離しうる；(iii) データプロセッサは、例えばGenBankおよびRefSeqからの例示参照配列を含むキュレート微生物参照配列データベースに対して非宿主配列をアライメントさせることにより、1以上の病原体の存在を決定しうる；(iv) 1以上の病原体の存在が統計的に有意であるかどうかを判定するために、統計的解析フレームワークが適用されうる；および/または(v) 幾つかの場合には、データプロセッサは、配列決定前に既知濃度でサンプルに添加された対照分子により得られたリードの数と比較した場合の、病原体に関して得られたリードの数に基づいて、存在する病原体の量を定量しうる。

【0248】

対照値は、異なる時点、例えば、試験時点の前の時点で、対象(例えば、感染を有する対象または感染を有すると疑われる対象)から得られた無細胞病原体または器官特異的核酸のレベルでありうる。そのような場合、異なる時点でのレベルの比較は、感染の存在、特定の器官における感染の存在、感染の改善、または感染の悪化を示しうる。例えば、無細胞病原体核酸の経時的な一定量の増加は感染の存在または感染の悪化を示しうる。例えば、元の値と比較して、少なくとも5%、10%、20%、25%、30%、50%、75%、100%、200%、300%または400%の、病原体または器官特異的無細胞核酸の増加は、感染の存在または感染の悪化を示しうる。他の例においては、元の値と比較して、少なくとも5%、10%、20%、25%、30%、50%、75%、100%、200%、300%または400%の、病原体または器官特異的無細胞核酸の減少は、感染の非存在または感染の改善を示しうる。しばしば、そのような測定は、特定の期間にわたって、例えば毎日、隔日、毎週、隔週、毎月または隔月に行われうる。例えば、1週間わたる、少なくとも50%の、病原体または器官無細胞核酸の増加は、感染の存在を示しうる。

【0249】

対照または参照値は濃度として又は配列決定リードの数として測定されうる。対照または参照値は病原体依存的でありうる。例えば、大腸菌(*Escherichia coli*)の対照値はマイコプラズマ・ホミニス(*Mycoplasma hominis*)の対照値と異なりうる。レベルまたは対照値のデータベースは、1以上の病原体、1以上の器官および/または1以上の時点に関して、1以上の対象から得られたサンプルに基づいて作成されうる。そのようなデータベースはキュレート(curated)物または独占物でありうる。推奨治療選択肢は種々の閾値レベルに基づきうる。例えば、低レベルは感染を示しうるが、治療は不要でない可能性があり、中等度のレベルは抗生物質治療につながる可能性があり、高レベルは即座の又は重大な介入を要しうる。

【0250】

本発明で提供する方法は高効率、高精度および/または高感度での配列決定データの作成を可能にしうる。しばしば、そのような方法は、プレート培養またはポリメラーゼ連鎖反応(PCR)のような他の方法によっては検出されず又は検出可能である病原体または

感染を検出する。該方法は一般に、非常に高い感度、例えば、80%、85%、90%、95%、99%または99.5%を超える感度を有する。該方法は一般に、非常に低い偽陽性率、例えば、5%、4%、3%、2%、1%、0.1%、0.05%、0.01%未満の偽陽性率を有する。

【0251】

本発明で提供する方法は、高い特異性、高い感度、高い陽性的中度および/または低い陰性的中度を与えうる。本発明で提供する方法は、少なくとも70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%またはそれ以上である特異性（または陰性一致率）および/または感度（または陽性一致率）を示しうる。幾つかの場合には、名目上の特異性は70%以上である。名目上の陰性的中度（NPV）は95%以上である。幾つかの場合には、NPVは少なくとも95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%またはそれ以上である。

10

【0252】

感度、陽性一致率（PPA）または真陽性率（有病正診率）（TPR）は $TP / (TP + FN)$ または $TP / (\text{感染対象の総数})$ の式に関するものでありうる（ここで、TPは真陽性の数であり、FNは偽陰性の数である）。前記式の分母を計算する場合、該値は、特定の独立した感染検出方法（例えば、血液培養またはPCR）に基づく感染結果の総数を反映しうる。

【0253】

特異性、陰性一致率または真陰性率（無病正診率）は、 $TN / (TN + FP)$ または $TN / (\text{未感染対象の総数})$ のような式に関するものでありうる（ここで、TNは真陰性であり、FPは偽陽性である）。前記式の分母を計算する場合、該値は、独立した感染検出方法（例えば、血液培養またはPCR）により決定された実際の「未感染」の総数を反映しうる。

20

【0254】

幾つかの場合には、サンプルは、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.5%またはそれ以上の精度で、感染していると特定される。幾つかの場合には、サンプルは、95%を超える感度で、感染していると特定される。幾つかの場合には、サンプルは、95%を超える特異性で、感染していると特定される。幾つかの場合には、サンプルは、95%を超える感度および95%を超える特異性で、感染していると特定される。幾つかの場合には、精度は、学習したアルゴリズムを使用して計算される。本明細書中で用いる診断精度は特異性、感度、陽性的中度、陰性的中度および/または偽発見率を含む。幾つかの場合には、本明細書に記載されている方法は、70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、または99.5%を超える特異性または感度、あるいは少なくとも95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%またはそれ以上の陽性的中度または陰性的中度を有する。

30

【0255】

感染の診断のためにサンプルを分類する場合、典型的には、二項分類指標からの4つの可能な結果が存在する。予測からの結果がpであり、実際の値もpである場合、それは真陽性（TP）と称される。しかし、実際の値がnであれば、それは偽陽性（FP）であると称される。逆に、予測結果と実際の値の両方がnである場合には真陰性が生じ、実際の値がpである場合に予測結果がnである場合には偽陰性となる。そのような感染症のような疾患または障害を検出する試験では、対象（被験体）が陽性試験結果を示すが実際には感染を有さない場合には、この場合の偽陽性が生じうる。一方、対象が実際に感染しているが、そのような感染に関する陰性試験結果を示す場合には、偽陰性が生じうる。

40

【0256】

50

陽性的中度 ( P P V )、または精度 ( p r e c i s i o n r a t e )、または疾患の検査後確率は、正確に診断された陽性試験結果を有する患者の割合である。それは、以下の式： $PPV = TP / (TP + FP)$ を適用することにより計算されうる。PPVは、陽性試験結果が、試験されている基礎状態 ( 基礎疾患 ) を反映する確率を反映しうる。しかし、その割合は、変動しうる当該疾患の罹患率に左右されうる。陰性的中度 ( N P V ) は、以下の式： $TN / (TN + FN)$ により計算される。陰性的中度は、正確に診断された陰性試験結果を有する患者の割合でありうる。PPVおよびNPV測定値は、適切な疾患有病率推定値を用いて導出されうる。

【 0 2 5 7 】

幾つかの場合には、本明細書に記載されている方法の配列決定分析の結果は、与えられた診断が正しいという統計的信頼水準を示す。幾つかの場合には、そのような統計的信頼水準は85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99.5%を超える。

10

【 0 2 5 8 】

モニターおよび治療

該方法は、経時的に対象が感染を有するかどうかをモニター ( 監視 ) することを含みうる。例えば、感染の存在または非存在を決定するために、サンプルを種々の時点で連続的に採取されうる。他の例では、該方法は、経時的に感染経過をモニターすることを含みうる。そのような場合、サンプルは感染または疾患中の種々の時点で連続的に採取可能であり、幾つかの場合には、連続的に採取されたサンプルを互いに比較して、感染が改善しているか悪化しているかを決定する。

20

【 0 2 5 9 】

本発明で提供する方法は、対象、例えば、感染を有する対象または感染を有すると疑われる対象の治療方法を含む。治療は対象における感染を軽減、予防または排除しうる。幾つかの場合には、治療は感染および/または炎症を軽減、予防または排除しうる。

【 0 2 6 0 】

治療は、炎症および/または感染を軽減または排除するために薬物または他の療法を投与することを含みうる。幾つかの場合には、例えば、感染または炎症の発生を予防するために、対象を薬物で予防的に治療 ( 処置 ) する。

【 0 2 6 1 】

感染または炎症の症状を改善または軽減するための任意の療法 ( 薬物を含む ) を対象に投与することが可能である。典型的な薬物には、限定的なものではないが、抗生物質、抗ウイルス薬、アンピシリン、スルバクタム、ペニシリン、バンコマイシン、ゲンタマイシン、アミノグリコシド、クリンダマイシン、セファロsporin、メトロニダゾール、チメンチン、チカルシリン、クラブラン酸、セフォキシチン、抗レトロウイルス薬 ( 例えば、高活性抗レトロウイルス療法 ( H A A R T )、逆転写逆転インヒビター、ヌクレオシド/ヌクレオチド逆転写酵素インヒビター ( N R T I )、非ヌクレオシドRTインヒビターおよび/またはプロテアーゼインヒビター)、抗体 - 薬物コンジュゲート、および免疫グロブリンが含まれる。

30

【 0 2 6 2 】

該方法は治療レジメンを調節する方法を含みうる。例えば、対象は既知の感染を有する可能性があり、該感染を治療するための薬物が投与されている可能性がある。本発明で提供する方法は、薬物治療の有効性を追跡またはモニターするために使用されうる。幾つかの場合には、該治療レジメンは、そのようなモニターの結果に応じて調節されうる。例えば、本発明で提供する方法が、薬物治療の結果として感染が改善していないことを示している場合には、患者に投与する薬物または治療のタイプを変更すること、先行薬物の使用を中止すること、該薬物の使用を継続すること、薬物治療の用量を増加させること、あるいは新たな薬物または他の治療を対象の治療レジメンに加えることにより、治療レジメンが調節されうる。幾つかの場合には、治療レジメンは特定の処置を含みうる。同様に、該方法が、感染が改善している又は消散していることを示している場合には、調節は、薬物

40

50

治療を低減または中止することを含みうる。

【0263】

本明細書に記載されている方法は更に、RNA配列決定(RNA-Seq)を含むことが可能であり、またはRNA-Seqを含む方法と組合せられる。組織損傷または感染は特定の器官または組織からの無細胞核酸の放出をもたらす。例えば、組織におけるアポトーシス細胞によりRNAが放出される。無細胞RNAのRNA-Seqは体内の種々の組織の健康または状態を示す。

【0264】

RNA配列決定を含む方法は、感染した特定の器官または組織の検出を可能にし、器官の健康状態を検出またはモニターするために使用される。RNA-Seqは、器官の健康を調べるために独立して使用可能であり、または本明細書に記載されている方法により検出された感染が特定の器官の感染であるという信頼性を高めうる。RNA-Seq試験は、感染の検出方法と同時に、感染の検出方法の後で、または感染の検出方法の前に行われうる。

【0265】

本発明が提供する病原体検出方法が体液中の無細胞RNAのRNA配列決定による感染部位の検出方法と組合せられる多数の潜在的シナリオが存在する。例えば、本発明で提供する方法は、病原体からの循環無細胞核酸を検出するために使用される。該方法は更に、対象血液における器官特異的無細胞RNAの増加を検出するためのRNA-Seq試験を行うことを含みうる。試験結果の組合せは、病原体が器官に感染していることを示すことが可能であり、更には、どの器官組織が感染しているかをも決定することが可能でありうる。

【0266】

RNA-Seq試験(または一連のRNA-Seq試験)は、時には、本明細書に記載されている方法が陽性試験結果(例えば、病原体感染の検出)を示した後に行われうる。RNA-Seq試験は、感染を確認するため、または感染の位置を特定するために特に有用でありうる。例えば、該方法は、循環無細胞核酸を分析することにより、対象における病原体の存在を検出しうるが、感染部位は不明でありうる。そのような場合には、該方法は更に、(例えば、器官組織に由来する循環無細胞RNAのレベルの増加の検出により)感染が器官内に存在することを確認するために、対象からの無細胞RNAを配列決定することを含みうる。ついで、特定の器官または組織において感染が悪化または改善しているかどうか、あるいはそれが異なる器官または組織に広がっているかどうかを判定するために、RNA配列決定試験を経時的に反復することが可能である。同様に、病原体検出アッセイを経時的に反復することが可能である。

【0267】

幾つかの場合には、本明細書に記載されている病原体検出方法は、RNA-Seq試験の実施の後で行われる。例えば、器官に関連する無細胞RNAの血漿レベルの上昇は器官の感染のような障害を示す。そのような場合、該方法は更に、器官感染に関連する循環無細胞核酸のレベルを検出することを含みうる。

【0268】

本明細書に記載されている方法は、例えば、経時的に感染または治療をモニターするために反復される。本明細書に記載されている方法は、1、2、3、4、5、6、7、8、9または10日ごと、1、2、3、4、5または6週ごと、あるいは1、2、3、4、5、6、7、8または9ヶ月ごとに反復される。

【0269】

幾つかの場合には、本明細書に記載されている方法が陰性試験結果を示す(例えば、病原体が検出されない)場合、対象における病原体核酸をモニターするために、方法を経時的に連続的に反復することが可能である。また、幾つかの場合においては、陰性病原体試験結果または陰性RNA-Seq結果の後、RNA-Seqアッセイを経時的に連続的に反復する。

10

20

30

40

50

## 【0270】

幾つかの場合には、本明細書に記載されている方法が陽性試験結果（例えば、病原体の検出）を示す場合、治療レジメンを対象に投与することが可能である。治療レジメンには、薬物投与、抗生物質投与または抗ウイルス投与が含まれ得るが、これに限定されるものではない。

## 【0271】

幾つかの場合には、本明細書に記載されている方法が陽性試験結果を示す場合、感染の経過をモニターするために、方法または試験を経時的に連続的に反復することが可能である。例えば、感染の上方または下方の経過に応じて、治療レジメンを調節することが可能である。他の場合には、最初はいずれの治療レジメンをも行わない。例えば、追加的な医学的介入を行うことなく感染が消失するかどうかを見るために、「監視的待機」または「経過観察」の方法で感染がモニターされうる。幾つかの場合には、本明細書に記載されている方法が陽性試験結果を示す場合、薬物を投与することが可能であり、該薬物がどれほど良好に作用するか、または薬物治療をいつ終了すべきかを調べるために、感染の経過をモニターすることが可能である。幾つかの場合には、必要に応じて療法が改変されうる。

## 【0272】

コンピュータ制御システム

本開示は、本開示の方法を実施するようにプログラムされたコンピュータ制御システムを提供する。図7は、本開示の方法を実施するようにプログラムまたは他の方法で構成されたコンピュータシステム701を示す。

## 【0273】

コンピュータシステム701は中央処理装置（CPU；本明細書においては「プロセッサ」および「コンピュータプロセッサ」とも称される）705を含み、これは単一コアもしくはマルチコアプロセッサ、または並列処理のための複数のプロセッサでありうる。コンピュータシステム701はまた、メモリまたはメモリロケーション710（例えば、ランダムアクセスメモリ、読み出し専用メモリ、フラッシュメモリ）、電子記憶装置715（例えば、ハードディスク）、1以上の他のシステムと通信するための通信インターフェイス720（例えば、ネットワークアダプタ）、および周辺装置725、例えばキャッシュ、他のメモリ、データ記憶装置および/または電子ディスプレイアダプタを含む。メモリ710、記憶装置715、インタフェース720および周辺装置725は、マザーボードのような通信バス（実ライン）を介してCPU705につながっている。記憶装置715は、データを格納するためのデータ記憶装置（またはデータリポジトリ）でありうる。コンピュータシステム701は、通信インタフェース720を用いてコンピュータネットワーク（「ネットワーク」）730に機能的に接続されうる。ネットワーク730は、インターネット、インターネットおよび/またはエクストラネット、あるいはインターネットとつながったイントラネットおよび/またはエクストラネットでありうる。ネットワーク730は、幾つかの場合には、テレコミュニケーションおよび/またはデータネットワークである。ネットワーク730は1以上のコンピュータサーバを含み、これは分散コンピューティング、例えばクラウドコンピューティングを可能にしうる。ネットワーク730は、幾つかの場合には、コンピュータシステム701を用いて、ピアツーピアネットワークを実行することが可能であり、これは、コンピュータシステム701に接続した装置がクライアントまたはサーバとして動作することを可能にしうる。

## 【0274】

CPU705は、プログラムまたはソフトウェアにおいて表されうる一連の機械可読命令を実行しうる。命令はメモリ位置、例えばメモリ710に格納されうる。命令はCPU705に向けられ、これはついで、本開示の方法を実施するようにCPU705をプログラムまたは構成しうる。CPU705によって実行される動作の例には、フェッチ、デコード、実行およびライトバックが含まれうる。

## 【0275】

CPU705は集積回路のような回路の一部でありうる。システム701の1以上の他

10

20

30

40

50

の構成要素が回路に含まれうる。幾つかの場合には、回路は特定用途向け集積回路（ASIC）である。

【0276】

記憶装置715は、ファイル、例えばドライバ、ライブラリおよび保存プログラムを格納しうる。記憶装置715は、ユーザデータ、例えば、利用者選好およびユーザプログラムを格納しうる。コンピュータシステム701は、幾つかの場合には、コンピュータシステム701の外部にある1以上の追加的データ記憶装置を含むことが可能であり、該追加的データ記憶装置は、例えば、イントラネットまたはインターネットを介してコンピュータシステム701につながっているリモートサーバ上に位置する。

【0277】

コンピュータシステム701はネットワーク730を介して1以上のリモートコンピュータシステムと通信可能である。例えば、コンピュータシステム701はユーザ（例えば、医療提供者）のリモートコンピュータシステムと通信可能である。リモートコンピュータシステムの例には、パーソナルコンピュータ（例えば、ポータブルPC）、スレートまたはタブレットPC（例えば、Apple（登録商標）iPad、Samsung（登録商標）Galaxy Tab）、電話、スマートフォン（例えば、Apple（登録商標）iPhone、Android対応デバイス、BlackBerry（登録商標））またはパーソナルデジタルアシスタントが含まれる。ユーザはネットワーク730を介してコンピュータシステム701にアクセス可能である。

【0278】

本明細書に記載されている方法は、コンピュータシステム701の電子記憶場所、例えばメモリ710または電子記憶装置715上に記憶された機械（例えば、コンピュータプロセッサ）実行可能コードにより実行されうる。機械実行可能コードまたは機械可読コードはソフトウェアの形態で提供されうる。コードは、使用中に、プロセッサ705により実行されうる。幾つかの場合には、コードは記憶装置715から検索され、プロセッサ710によって容易にアクセスできるようにメモリ710に記憶されうる。幾つかの場合には、電子記憶装置715は除外可能であり、機械実行可能命令はメモリ710に格納される。

【0279】

コードは、コードを実行するように適合化されたプロセッサを有する機械と共に使用されるように、プリコンパイルされ、構成されることが可能であり、または実行中にコンパイルされうる。コードは、プリコンパイルまたはコンパイル形態でコードが実行可能となるように選択されうるプログラミング言語で供給されうる。

【0280】

コンピュータシステム701のような、本発明で提供するシステムおよび方法の態様は、プログラミングにおいて具体化されうる。該技術の種々の態様は、典型的には機械（またはプロセッサ）実行可能コードの形態および/または機械可読媒体の一種に搭載または組み込まれた関連データの形態の「製品」または「製造品」であると考えられうる。機械実行可能コードは、電子記憶装置、例えばメモリ（例えば、リードオンリメモリ、ランダムアクセスメモリ、フラッシュメモリ）またはハードディスクに格納されうる。「記憶（ストレージ）」タイプのメディアには、コンピュータ、プロセッサなどの有形メモリ、または関連モジュール、例えば種々の半導体メモリ、テープドライブ、ディスクドライブなどのありとあらゆるものが含まれることが可能であり、それらはソフトウェアプログラミングのために任意の時点で非一時的ストレージを提供しうる。ソフトウェアの全部または一部は、時には、インターネットまたは種々の他のテレコミュニケーションネットワークを介して通信されうる。そのような通信は、あるコンピュータまたはプロセッサから別のコンピュータまたはプロセッサへ、例えば、管理サーバまたはホストコンピュータからアプリケーションサーバのコンピュータプラットフォームへの、ソフトウェアのロードを可能にしうる。したがって、ソフトウェア要素を担いうる別のタイプの媒体には、有線および光陸線ネットワークならびに種々のエアリンクを介してローカルデバイス間の物理イン

10

20

30

40

50

タフェースを越えて使用されるような、光波、電波および電磁波が含まれる。そのような波を運ぶ物理的要素、例えば有線または無線リンク、光リンクなども、ソフトウェアを担持する媒体とみなされうる。本明細書中で用いる、コンピュータまたは機械「可読媒体」のような語は、非一時的な有形「記憶」媒体に限定されない限り、実行のためにプロセッサに命令を与えることに関与する任意の媒体を意味する。

#### 【0281】

したがって、機械可読媒体、例えばコンピュータ実行可能コードは、有形記憶媒体、搬送波媒体または物理的伝送媒体（これらに限定されるものではない）を含む多数の形態を取りうる。不揮発性記憶媒体には、例えば、光学または磁気ディスク、例えば、任意のコンピュータなどにおける記憶装置のいずれか、例えば、図面に示されているデータベースを実行するために使用されうるものが含まれる。揮発性記憶媒体には、動的メモリ、例えば、そのようなコンピュータプラットフォームのメインメモリが含まれる。有形伝送媒体には、同軸ケーブル、銅線および光ファイバ（コンピュータシステム内のバスを構成するワイヤを含む）が含まれる。搬送波伝送媒体は、電気または電磁信号、あるいは音響波または光波、例えば、無線周波数（RF）および赤外線（IR）データ通信中に生成されるものの形態を取りうる。したがって、コンピュータ可読媒体の一般的な形態には、例えば、フロッピーディスク、フレキシブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、CD-ROM、DVDまたはDVD-ROM、任意の他の光学媒体、パンチカード紙テープ、ホールのパターンを有する任意の他の物理的記憶媒体、RAM、ROM、PROMおよびEPROM、FLASH-EPROM、任意の他のメモリチップまたはカートリッジ、データまたは命令を搬送する搬送波、そのような搬送波を搬送するケーブルまたはリンク、あるいはコンピュータがプログラミングコードおよび/またはデータを読取りうる任意の他の媒体が含まれうる。コンピュータ可読媒体のこれらの形態の多くは、実行のために1以上の命令のシーケンスの1以上をプロセッサに搬送することに関与しうる。

#### 【0282】

コンピュータシステム701は、電子ディスプレイ735を含むことが可能であり、または電子ディスプレイ735と通信可能であり、電子ディスプレイ735は、対象の診断または対象に対する治療的介入を含みうる報告の出力を提供するユーザインターフェース（UI）740を含む。UIの例には、限定的なものではないが、グラフィカルユーザインターフェース（GUI）およびウェブベースユーザインターフェースが含まれる。分析は報告として提供されうる。報告は対象、医療従事者、研究者または他の個人に提供されうる。

#### 【0283】

本開示の方法およびシステムは1以上のアルゴリズムにより実行されうる。アルゴリズムは、中央処理ユニット705による実行に際してソフトウェアにより実行されうる。アルゴリズムは、例えば、病原体または他の標的核酸の富化、配列決定および/または検出を容易にしうる。

#### 【0284】

患者または対象に関する情報、例えば患者の背景、患者の病歴または医学的スキャンがコンピュータシステムに入力されうる。コンピュータシステムは、本明細書に記載されている方法からの結果を分析するため、あるいは結果を患者または医師に報告するため、あるいは治療計画を策定するために使用されうる。

#### 【0285】

##### 試薬およびキット

本明細書に記載されている方法の1以上を行うための試薬およびそのキットも提供する。該試薬およびそのキットは多種多様でありうる。関心のある試薬には、対象から得られたサンプルにおける1以上の病原体または他の標的核酸の特定、検出および/または定量における使用のために特別に設計された試薬が含まれる。該キットは、本明細書に記載されている方法、例えばPCRおよび配列決定を用いて核酸抽出および/または核酸検出を

10

20

30

40

50



行うのに必要な試薬を含みうる。該キットは更に、データ解析用のソフトウェアパッケージを含むことが可能であり、これは、試験プロファイルとの比較のための参照プロファイルを含むことが可能であり、特に、参照データベースを含みうる。該キットは、試薬、例えばバッファーおよび水を含みうる。

#### 【0286】

そのようなキットはまた、情報、例えば科学文献リファレンス、添付文書資料、臨床試験結果および/またはこれらの概要などを含むことが可能であり、これらは、該組成物の活性および/もしくは利点を示し、もしくは確定し、および/または用量、投与、副作用、薬物相互作用、もしくは医療提供者にとって有用な他の情報を記載している。そのようなキットは、データベースにアクセスするための説明をも含みうる。そのような情報は、種々の研究の結果、例えば、インビボモデルを含む実験動物を用いる研究およびヒト臨床試験に基づく研究に基づくものでありうる。本明細書に記載されているキットは、医師、看護師、薬剤師、処方職員などを含む医療提供者に提供、販売および/または宣伝されうる。キットはまた、幾つかの実施形態においては、消費者に直接販売されうる。

10

#### 【0287】

本開示は、配列決定ライブラリーを製造するためのキットをも提供する。該キットは、本明細書に記載されている少なくとも1つの合成核酸、および配列決定ライブラリー反応のための試薬を含みうる。幾つかの場合には、該キットは、1以上の配列決定アダプターおよび1以上の担体(carrier)核酸を含む。該キットにおける担体核酸は、i) 末端修復に抵抗性である1以上の担体核酸、ii) 連結に抵抗性である1以上の担体核酸、iii) 増幅に抵抗性である1以上の担体核酸、iv) 固定化タグを含む1以上の担体核酸、v) サイズに基づく枯渇を可能にするサイズを有する1以上の担体核酸、および/またはvi) それらの任意の組合せを含みうる。例えば、該キットは、1以上の配列決定アダプター、および末端修復に抵抗性である1以上の担体核酸を含みうる。

20

#### 【0288】

キットにおける配列決定ライブラリーアダプターの量および1以上の担体核酸の量は一定の比率でありうる。幾つかの場合には、配列決定ライブラリーアダプターの量と1以上の担体核酸の量との比は、1:10、1:5、1:1、5:1、10:1、20:1、50:1、100:1、500:1、または1000:1以下である。例えば、配列決定ライブラリーアダプターの量と1以上の担体核酸の量との比は1:1以下でありうる。

30

#### 【0289】

##### 担体核酸(CNA)

本開示は、担体(carrier)核酸(CNA)、特に、配列決定アッセイの工程の1以上からそれを除外するように設計された特徴を含有する密かな(surreptitious)CNAを提供する。本開示はまた、配列決定アッセイの工程の1以上を回避しうるCNAを使用する方法を提供する。本発明で提供するCNAは密かに作用しうるが、それらは、一般に、サンプルにおける全核酸量を増加させることが可能であり、それにより、典型的な「担体」核酸として作用しうる。担体核酸は、一般に、サンプルから配列決定ライブラリーを製造する際の収率および/または効率を改善するために核酸量を増加させ、最終的に配列決定アッセイの精度および/または感度を改善しうる。本発明で提供する修飾CNAを含む担体核酸の添加は、サンプルが少量(例えば、1ng未満)の標的核酸を含有する場合に特に有用でありうる。なぜなら、核酸の量が少ないと、ライブラリー製造の1以上の工程(例えば、核酸抽出、核酸精製、核酸末端修復、アダプター連結など)または配列決定アッセイにおける後の工程(例えば、増幅)の効率および/または収率が低下しうるからである。DNAおよび/またはRNAに基づく核酸は、それらの構造的形態がいずれであっても、および/または1以上の化学修飾を伴う場合も伴わない場合も、関心のある核酸サンプルにCNAとして添加されうる。典型的には、CNAは、例えば阻害によっても、あるいは、配列決定スループットの過度な部分を占めることによっても、核酸配列決定を妨げない。幾つかの場合には、DNAサンプルおよび/またはRNAサンプルにDNA CNAを添加する。幾つかの場合には、DNAサンプルおよび/または

40

50

RNAサンプルにRNA CNAを添加する。

【表6】

表6: 典型的な担体核酸配列

配列番号	配列 (5'-3')
配列番号 170	CTG TTG GGC CGC CAC TGC GTG AGC CTC GGC CC
配列番号 171	GAC CTA TTT TGA CGG CAC CGT TGG rCrGrG rArArG TTG CTG GGC CTG CGC ACC GCG G
配列番号 172	GCG TCC CGG CGC GCG TTT AGG GAT AAC AGG GTA ATG GCG CAA GGG TGC TGG C
配列番号 173	/5InvddT/GC GTC CCG GCG CGC GTT TAG GGA TAA CA/idSp/ /idSp//idSp//idSp/ GGG TAA TGG CGC AAG GGT GCT GGC /3InvdT/

10

【0290】

20

本発明で提供するCNAは、配列決定ライブラリー製造の工程の1以上、例えば、末端修復、断片化、増幅、連結および配列決定を回避するように設計または修飾されうる。CNAは配列決定ライブラリー製造における1以上の工程に添加されうる。例えば、図8に示されているとおり、CNAは、サンプル採取802の途中もしくは直後、サンプル調製、例えば血漿803の単離の途中もしくは後、核酸単離804もしくは抽出805の前、途中もしくは後、核酸精製の前、途中もしくは後に、核酸806の末端修復の前、途中もしくは後、連結807もしくは核酸にアダプターを結合させるための他の操作の前、途中もしくは後、および/または増幅808の前もしくは途中に添加されうる。幾つかの場合には、CNAは、例えば、酵素消化、アフィニティに基づく枯渇および/またはサイズに基づく枯渇により、配列決定アッセイにおける工程から除去されうる。例えば、本発明で提供するCNAは、それが配列決定ライブラリーに含まれないように、配列決定アッセイの工程から物理的に除去されうる。幾つかの場合には、CNAは配列決定ライブラリー自体から物理的に除去されうる。

30

【0291】

#### 結合に抵抗するCNA

本発明で提供するCNAは、1以上の配列決定アダプターおよび/または標的核酸のような他の分子に対する結合または連結に抵抗しうる(抵抗性でありうる)。幾つかの場合には、CNAは、アダプターがCNAよりも優先的に標的核酸に優先的に連結するように設計されうる。アダプターまたは標的核酸への連結または結合を回避することにより、CNAは配列決定されることも回避されうる。

40

【0292】

幾つかの場合、特に、サンプルにおける核酸にアダプターを結合させるために連結を用いる場合には、CNAは、連結反応に加わることに抵抗するように設計されうる。一般に、連結反応は、2つの核酸をホスホジエステル結合により連結することを含む。幾つかの場合には、CNAは、連結反応に抵抗する二次構造(例えば、一本鎖構造、ヘアピン構造)を有するように設計されうる。二次構造はRNA、DNA、ssDNA、dsDNA、DNA-RNAハイブリッドおよび/または他の特徴を含みうる。幾つかの場合には、CNAは、連結を妨げるように設計されたブロッキング基または他の構造を含有しうる。

【0293】

本発明で提供するCNAは、結合または連結に抵抗し又はそれらを低減するように設計

50

された一本鎖および/または二本鎖二次構造を含有しうる。CNAは1以上の一本鎖領域を含有することが可能であり、あるいは全体が一本鎖でありうる。一本鎖領域はCNAの任意の位置に存在しうるが、幾つかの好ましい場合には、CNAは、その末端付近または末端の一方もしくは両方に一本鎖領域を含有する。例えば、CNAは、一方または両方の末端から50ヌクレオチド以内に、例えば、一方または両方の末端から50nt、45nt、40nt、35nt、30nt、25nt、20nt、15nt、10ntまたは5nt以内に、一本鎖領域を含有しうる。幾つかの好ましい場合には、CNAは、その末端の一方または両方(例えば、5'末端、3'末端)に一本鎖領域を含有しうる。幾つかの場合には、CNAは全体が二本鎖であることが可能であり、あるいは、二本鎖である領域を単に含有することが可能である。二次構造(特にヘアピンループ)は、リガーゼによるCNAの結合および/または認識を妨げうる。幾つかの場合には、CNAはY字型二本鎖核酸を含有することが可能であり、その結果、CNAのY字型部分は別の核酸に連結または結合できない。

#### 【0294】

本発明で提供するCNAに存在しうるヘアピン構造は、一般に、ループおよびハイブリダイゼーション領域、例えばヘアピンステムを有する。例えば、ヘアピンは、二本鎖ハイブリダイゼーション領域を形成する2つの相補的領域と、それらの2つの相補的領域を連結するループとを含みうる。相補的領域は少なくとも5、10、15、20、30、40、50ヌクレオチドを含みうる。ループ領域は少なくとも3、4、5、10、15、20、30、40、50ヌクレオチドを含みうる。一般に、ヘアピン構造は、しばしば、結合を伴わない一本鎖核酸であるため、製造が比較的容易でありうる。ヘアピンはRNAまたはDNAを含有しうる。

#### 【0295】

本発明で提供するCNAは、結合または連結に抵抗しまたはそれを低減しうる環状構造を含有しうる。環状構造は環状DNA、環状RNAまたは環状DNA-RNAハイブリッドでありうる。幾つかの場合には、環状構造は環状DNAである。環状構造は二本鎖または一本鎖でありうる。環状構造は、特定の長さ、例えば少なくとも5nt、10nt、20nt、30nt、32nt、40nt、50nt、60nt、70nt、80nt、90nt、100nt、120nt、140nt、160nt、180nt、200nt、250nt、300nt、400nt、500ntまたは1000ntのものでありうる。幾つかの場合には、環状構造は約30~約100ヌクレオチドを含む。幾つかの場合には、環状構造は約10ヌクレオチド~約1,000ヌクレオチドの範囲内、例えば約100ヌクレオチド~約1,000ヌクレオチドの範囲内のサイズを有しうる。環状構造が二本鎖である場合、環状構造は少なくとも10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、250bp、300bp、400bp、500bpまたは1000bpのサイズを有しうる。幾つかの場合にて、二本鎖環状構造は約30bp~100bpを含む。幾つかの場合には、二本鎖環状構造は、約10塩基対~約1,000塩基対の範囲内、例えば、約100塩基対~約1,000塩基対の範囲内のサイズを有しうる。幾つかの場合には、環状構造は、ある酵素、例えばエンドヌクレアーゼからの消化に対してCNAが抵抗性となることを可能にしうる。例えば、CNAは二本鎖環状構造を含有することが可能であり、エンドヌクレアーゼ、例えば、二本鎖線状DNAを消化するが二本鎖環状DNAを消化しないエンドヌクレアーゼによる消化に抵抗しうる。幾つかの場合には、CNAは、大部分または全体が環状であり、例えば、環状二本鎖DNA、環状一本鎖DNAである。幾つかの場合には、CNAは、エンドヌクレアーゼ、例えば、CNAの二次構造に結合しないおよび/またはそれを認識しないエンドヌクレアーゼによる消化に抵抗する二次構造を含む。例えば、CNAは、一本鎖DNAを認識するが二本鎖DNAを認識しないエンドヌクレアーゼによる消化に抵抗する二本鎖DNAを含みうる。もう1つの例では、CNAは、二本鎖DNAを認識するが一本鎖DNAを認識しないエンドヌクレアーゼによる消化に抵抗する一本鎖DNAを含みうる。

10

20

30

40

50

## 【0296】

幾つかの場合には、CNAは、1以上のニック（切れ目）を有する二本鎖である。ニックは二本鎖核酸分子において不連続性であることが可能であり、この場合、該鎖の1つの隣接ヌクレオチド間にホスホジエステル結合が存在しない。ニックは酵素（例えば、ニックキングエンドヌクレアーゼ）により生成されうる。幾つかの場合には、ニックは酵素（例えば、リガーゼ）により連結されうる。ある場合には、ニックはエキソヌクレアーゼ消化および/または連結から保護される。

## 【0297】

CNAは、連結反応に抵抗する1以上の修飾（例えば、修飾ヌクレオチド）を含みうる。幾つかの場合には、修飾は、核酸にCNAが連結するのを妨げるブロッキング基でありうる。例えば、CNAは3'末端、5'末端または両末端にブロッキング基を有しうる。ブロッキング基は、反転（*inverted*）デオキシ糖を含みうる。反転デオキシ糖は反転デオキシ糖、反転ジデオキシ糖または他の反転デオキシ糖でありうる。反転デオキシ糖は3'反転デオキシ糖または5'反転ジデオキシ糖でありうる。例えば、ブロッキング基は3'反転チミジン（dT）、3'反転アデノシン（dA）、3'反転グアノシン（dG）、3'反転シチジン（dC）、3'反転デオキシウラシル（dU）、5'反転ジデオキシチミジン（ddT）、5'反転ジデオキシアデノシン（ddA）、5'反転ジデオキシグアノシン（ddG）、5'反転ジデオキシシチジン（ddC）、5'（反転）ジデオキシウラシル（ddU）またはそれらの任意の類似体でありうる。幾つかの場合には、CNAは3'反転チミジンを含む。幾つかの場合には、CNAは5'反転ジデオキシチミジンを含む。幾つかの場合には、CNAは3'反転チミジンおよび/または5'反転ジデオキシチミジンを含む。幾つかの場合には、ブロッキング基はジデオキシシチジンを含む。幾つかの場合には、修飾はウラシル（U）塩基、2'OMe修飾RNA、C3-18スパーサー（例えば、3~18個の連続炭素原子を有する構造）、ピオチン、ジ-デオキシヌクレオチドトリホスファート、エチレングリコール、アミンおよび/またはホスファート（リン酸）を含む。

## 【0298】

増幅に抵抗する担体核酸

CNAは、核酸増幅を阻害することにより配列決定反応においてCNAが増幅されることを妨げる1以上の核酸修飾を含みうる。幾つかの場合には、修飾は、例えば、ポリメラーゼを機能停止または阻害（例えば、機能低下）することにより、核酸ポリメラーゼの機能を妨げうる。幾つかの場合には、修飾は1以上の脱塩基（*abasic*）部位を含みうる。脱塩基部位は、塩基を有さない核酸内の位置を意味しうる。例えば、核酸内の脱塩基部位は、塩基を有さない1'末端に存在しうる。脱塩基部位はアプリンもしくはアピリミジン構造、塩基類似体またはリン酸骨格類似体を有しうる。幾つかの場合には、脱塩基部位は、アミド結合により連結されたN-（2-アミノエチル）-グリシン、テトラヒドロフランまたは1', 2'-ジデオキシリボース（*dSpacer*）を有する。幾つかの場合には、修飾は、脱塩基部位および修飾糖残基、例えば、3個の炭素原子を含有する糖残基、例えば、部分リボース構造（例えば、3'、4'、5'末端炭素原子のみが保持されている）を含むことが可能であり、それにより、骨格（バックボーン）に沿った接続性が維持されうる。

## 【0299】

脱塩基部位は、CNAをポリメラーゼが増幅するのを妨げうる。幾つかの場合には、CNAにおける脱塩基部位はポリメラーゼ（例えば、Taqポリメラーゼ）を脱塩基部位1個当たり1桁阻害しうる。

## 【0300】

本発明で提供するCNAは複数の脱塩基部位、例えば、複数の内部脱塩基部位および1以上の他の特徴を含みうる。CNAは、1以上のライブラリー生成反応に加わることを妨げる特徴を含みうる。例えば、CNAは1以上の内部脱塩基部位、3'反転dTおよび/または5'反転ddTを任意の組合せで含みうる。

## 【0301】

幾つかの場合には、CNAは、核酸増幅を阻害する他の修飾を含みうる。幾つかの場合には、核酸増幅を阻害する修飾には、ウラシル(U)塩基、2'OMe修飾RNA、C3-18スペーサー(例えば、C3スペーサーのような3~18個の連続炭素原子を有する構造)、エチレングリコール多量体スペーサー(例えば、スペーサー18(ヘキサ-エチレングリコールスペーサー)、ピオチン、ジ-デオキシヌクレオチドトリホスファート、エチレングリコール、アミンおよび/またはホスファート)が含まれる。

## 【0302】

修飾

CNAは少なくとも1、2、3、4、5、6、7、8、9、10個またはそれ以上の修飾(例えば、脱塩基部位)を含みうる。CNAが複数の修飾(例えば、核酸増幅を阻害する修飾)を含む場合、修飾はクラスター化されうる(例えば、修飾は互いに隣接して連続して位置する)。幾つかの場合には、1以上の修飾がCNAの5'末端に存在する。幾つかの場合には、1以上の修飾がCNAの3'末端に存在する。幾つかの場合には、前記の1以上の修飾はCNAの3'末端および5'末端の両方に存在する。幾つかの場合には、前記の1以上の修飾はCNAの内部位置に存在する。例えば、CNAは、1以上の内部 *d spacer (ids p)* を含みうる。

## 【0303】

本明細書に記載されている修飾には、2-アミノプリン、2,6-ジアミノプリン、5-プロモdU、デオキシウリジン、反転dT、反転ジデオキシ-T、ジデオキシ-C、5-メチルdC、デオキシノシン、汎用塩基、例えば5-ニトロインドール、2'-O-メチルRNA塩基、イソ-dC、イソ-dG、リボヌクレオチド、モルホリノ、タンパク質ヌクレオチド類似体、糖ヌクレオチド類似体、ロックド(Locked)ヌクレオチド類似体、トレオースヌクレオチド類似体、鎖終結ヌクレオチド類似体、チオウリジン、プソイドウリジン、ジヒドロウリジン、キューオシン、ワイオシンヌクレオチド、脱塩基部位、官能基、例えばアルキン官能基、アジド官能基、例えばアジド(NHSエステル、非天然結合、例えばホスホロチオアート結合、スペーサー、例えば2'-ジデオキシリボース(*d Spacer*)、ヘキサンジオール、光切断性スペーサー、種々の数の炭素原子を有する種々の長さのスペーサー、例えばC3スペーサーホスホラミダイト、C9スペーサー、例えばトリエチレングリコールスペーサー、CI8(18原子ヘキサエチレングリコールスペーサー)が含まれうる。そのようなスペーサーはCNAまたはアダプターの5'末端または3'末端または内部に組み込まれうる。更に、CNAの少なくとも1つの鎖は、例えば、5'ホスファートまたは3'ホスファート(例えば、相補鎖上)のいずれかまたは両方を含むように、リン酸化により修飾されうる。

## 【0304】

酵素認識部位

CNAは、配列決定ライブラリーからCNAが除去されることを可能にする特徴を含みうる。そのような特徴は酵素認識部位を含みうる。例えば、CNAは、合成核酸が酵素により分解されうるように、1以上の酵素認識部位を含みうる。幾つかの場合には、CNAは、標的核酸およびアダプターに存在しない1以上の酵素認識部位を含みうる。したがって、担体核酸は、標的核酸またはアダプターの酵素分解をもたらすことなく、認識部位を標的とする酵素により除去されうる。

## 【0305】

幾つかの場合には、CNAはヌクレアーゼ認識部位を含みうる。例えば、ヌクレアーゼ認識部位はエンドヌクレアーゼ認識部位でありうる。エンドヌクレアーゼはI型、II型(II S型、II G型を含む)、III型またはIV型エンドヌクレアーゼでありうる。幾つかの場合には、エンドヌクレアーゼ認識部位は制限ヌクレアーゼ認識部位である。例えば、エンドヌクレアーゼ認識部位は、AatII、Acc65I、AccI、AclI、AatII、Acc65I、AccI、AclI、AfeI、AflII、AgeI、ApaI、ApalI、ApoI、AscI、AseI、AsiSI、AvrII、Ba

10

20

30

40

50

mH I、B c l I、B g l I I、B m e 1 5 8 0 I、B m t I、B s a H I、B s i E I、B s i W I、B s p E I、B s p H I、B s r G I、B s s H I I、B s t B I、B s t Z 1 7 I、B t g I、C l a I、D r a I、E a e I、E a g I、E c o R I、E c o R V、F s e I、F s p I、H a e I I、H i n c I I、H i n d I I I、H p a I、K a s I、K p n I、M f e I、M l u I、M s c I、M s p A 1 I、M f e I、M l u I、M s c I、M s p A 1 I、N a e I、N a r I、N c o I、N d e I、N g o M I V、N h e I、N o t I、N r u I、N s i I、N s p I、P a c I、P c i I、P m e I、P m l I、P s i I、P s p O M I、P s t I、P v u I、P v u I I、S a c I、S a c I I、S a l I、S b f I、S c a I、S f c I、S f o I、S g r A I、S m a I、S m l I、S n a B I、S p e I、S p h I、S s p I、S t u I、S w a I、X b a I、X h o IまたはX m a Iの認識部位でありうる。酵素認識部位は、前記で挙げられていないDNアーゼ、例えばエキソデオキシリボヌクレアーゼの部位でありうる。酵素認識部位はウラシルDNAグリコシラーゼ(UDG)、DNAグリコシラーゼ-リアーゼ(エンドヌクレアーゼV I I I)またはそれらの混合物(例えば、ウラシル特異的切断試薬(USER)酵素)の部位でありうる。例えば、CNAは1以上のウラシル(例えば、内部ウラシル)を含みうる。酵素認識部位はRNA誘導DNアーゼの部位、例えばCRISPR関連タンパク質ヌクレアーゼ、例えばCas9の部位でありうる。ある場合には、ヌクレアーゼ認識部位は、RNアーゼ、例えばエンドリボヌクレアーゼ、例えばRNアーゼA、RNアーゼH、RNアーゼI I I、RNアーゼL、RNアーゼP、RNアーゼPhyM、RNアーゼT1、RNアーゼT2、RNアーゼU2、RNアーゼV、またはエキソリボヌクレアーゼ、例えばポリヌクレオチドホスホリラーゼ、RNアーゼPH、RNアーゼR、RNアーゼD、RNアーゼT、オリゴリボヌクレアーゼ、エキソリボヌクレアーゼIまたはエキソリボヌクレアーゼI Iの認識部位でありうる。幾つかの特定の例では、CNAは制限酵素認識部位を含むことが可能であり、本発明で提供する方法は、そのような部位を認識する制限酵素でCNAを消化することを含みうる。幾つかの場合には、CNAは、酵素(例えば、CNAに結合しおよび/またはCNAを分解する酵素)、リボザイム、アダプターおよびDNAに基づく触媒または結合ポリマーにより認識されうる二次または三次構造を含む。幾つかの場合には、CNAは、酵素により認識されうる1以上の特異的結合性核酸配列を含む。

#### 【0306】

幾つかの場合には、CNAは、DNアーゼまたはRNアーゼにより分解されうるDNA-RNAハイブリッドを含みうる。幾つかの場合には、CNAはDNA-RNA-DNAハイブリッドを含む。そのような分子は二本鎖でありうる。CNAの末端領域はデオキシリボヌクレオチドを含みうる。内部領域はリボヌクレオチドを含みうる。幾つかの場合には、DNA-RNAハイブリッドは標的核酸またはアダプターに連結することが可能であり、ついでDNA-RNAハイブリッドは配列決定の前(例えば、増幅工程の前)にRNアーゼにより消化されうる。幾つかの特定の例では、DNA-RNAハイブリッドは(例えば、RNアーゼにより)消化され、一方、標的核酸(例えばDNA、例えば無細胞DNA)はRNアーゼによっては消化されない。

#### 【0307】

CNAのDNA部分が、増幅に抵抗するのに十分な程度に長い場合には、配列決定の前にDNA-RNAハイブリッドを除去するためのRNアーゼ消化工程は必要ないかもしれない。あるいは、DNA-RNAハイブリッド分子が増幅前に酵素消化により分解される場合には、DNA-RNAハイブリッドは、増幅に抵抗するサイズまたは長さを有する必要がないかもしれない。

#### 【0308】

##### サイズに基づく枯渇のためのCNA

CNAは、サイズに基づく枯渇(depletion)により配列決定ライブラリーから分離されうるサイズを有しうる。幾つかの場合には、CNAは、標的核酸の長さよりも長い、または標的核酸の平均長よりも長い長さを有する。例えば、CNAは、標的核酸の

10

20

30

40

50

長さ又は標的核酸の平均長より少なくとも1.5、2、3、4、5、10、20または50倍長い長さを有しうる。CNAは少なくとも150bp、200bp、300bp、400bp、500bp、600bp、800bp、1kb、2kb、5kbまたは10kbの長さを有しうる。例えば、CNAは少なくとも500bpの長さを有しうる。幾つかの場合には、CNAは約150bp～約1000bpの範囲内のサイズを有しうる。幾つかの場合には、CNAは2kbまでのサイズを有しうる。幾つかの場合には、CNAの長さは標的核酸の長さまたは標的核酸の平均長より短い。例えば、CNAは標的核酸の長さまたは標的核酸の平均長の多くとも99%、95%、90%、80%、60%、50%、40%、20%または10%の長さを有しうる。幾つかの場合には、CNAは標的核酸のサイズまたは標的核酸の平均サイズの多くとも50%のサイズを有しうる。ある場合には、CNAは標的核酸または標的核酸の平均長と実質的に同じ長さを有する。

10

#### 【0309】

サイズに基づく枯渇を可能にするサイズまたは長さを有するCNAは、本開示に記載されている任意の修飾、例えば、連結、増幅、末端修復またはそれらの組合せを妨げるための修飾を含有しうる。幾つかの場合には、CNAの末端の一方または両方は該修飾の1以上を含有しうる。幾つかの場合には、修飾は内部修飾、例えば内部脱塩基部位、または末端修飾と内部修飾との組合せでありうる。

#### 【0310】

幾つかの特定の例においては、CNAは、サイズに基づく枯渇を可能にする、より長い長さ、および連結を妨げる反転塩基のような修飾（例えば、末端修飾など）を有しうる。連結を防止するまたは妨げる構造の他の組合せも可能である（例えば、ヘアピンループ、ヘアピンループと末端修飾との組合せ）。幾つかの場合には、CNAは1以上のヘアピン構造および1以上の脱塩基部位を含みうる。幾つかの特定の例においては、CNAは、500bpを超えるサイズまたは長さを有することが可能であり、3'反転dT、5'反転ddT、C3スペーサー、またはスペーサー18、またはヘアピン構造（一方の末端に存在する）を有しうる。幾つかの特定の例においては、CNAは、600bpを超えるサイズまたは長さを有することが可能であり、3'反転dT、5'反転ddT（一方の末端に存在する）および1以上の内部脱塩基部位を有しうる。

20

#### 【0311】

##### 固定化タグ

CNAは1以上の固定化タグを含みうる。固定化タグは、アフィニティに基づく枯渇による溶液（例えば、配列決定ライブラリーの溶液）からCNAを除去するために使用されうる。例えば、固定化タグは固体支持体、例えばビーズまたはプレートに結合しうる。溶液を固体支持体に接触させた際に、CNAは溶液から除去されうる。1以上の固定化タグを含むCNAは標的核酸より短いことが可能である。あるいは、CNA分子は、例えば、配列決定反応へのCNAの持ち越し（キャリアオーバー）を最小限に抑えるために、標的核酸より長いことが可能である。

30

#### 【0312】

固定化タグには、ビオチン、ジゴキシゲニン、Ni-ニトリロトリ酢酸、デスチオビオチン、ヒスチジン、ポリヒスチジン、myc、ヘマグルチニン（HA）、FLAG、蛍光タグ、タンデムアフィニティー精製（TAP）タグ、グルタチオンSトランスフェラーゼ（GST）、ポリヌクレオチド、アプタマー、ポリペプチド（例えば、抗原または抗体）またはそれらの誘導体が含まれうる。例えば、CNAはビオチン、例えば、内部または末端ビオチン化鎖を含みうる。幾つかの場合には、固定化タグは磁気感受性材料、例えば磁石、または磁気感受性金属を含みうる。幾つかの特定の例においては、ビオチン化CNAは、増幅工程の前に、サンプルまたは配列決定ライブラリーからの、CNAの、磁気ビーズに基づく枯渇（例えば、アビジン-磁気ビーズによるもの）を可能にしうる。幾つかの場合には、CNAは、固体支持体に結合しうる又は固定化タグに結合しうる二次または三次構造を含む。

40

#### 【0313】

50

幾つかの場合には、標的核酸および/または配列決定ライブラリー核酸は1以上の固定化タグを含む。これらの場合には、CNAは固定化タグを含まず、または標的核酸とは異なる固定化タグを含む。したがって、CNAは、異なる固定化タグを使用する、アフィニティに基づく枯渇により、標的核酸および/または配列決定ライブラリー核酸から分離される。例えば、標的核酸および/または配列決定ライブラリー核酸は固体支持体上に固定化されることが可能であり、一方、CNAは洗い流される。幾つかの場合には、CNAは直接的または間接的に固定化タグに連結される。幾つかの場合には、CNAは固定化タグから切断される。

#### 【0314】

CNAは、本明細書に開示されている特徴および構造の組合せを含みうる。幾つかの場合には、CNAは、核酸増幅を阻害する1以上の修飾と、連結反応に抵抗する1以上の修飾とを含む。例えば、CNAは1以上の脱塩基部位（例えば、内部d spacer）および反転デオキシ塩基（例えば、3'反転チミジン）を含みうる。修飾を含むCNAは更に、酵素認識部位および/または固定化タグを含みうる。ある場合には、CNAは、1以上の固定化タグを有するDNA-RNAハイブリッド、例えばビオチン化DNA-RNA-DNAハイブリッド分子を含む。CNAは、特定の酵素またはタンパク質、非アミノ酸に基づく任意の触媒またはアフィニティ単位、例えばリボザイム、DNAに基づく触媒高分子および分子刷込高分子に対して高いアフィニティを有する核酸の二次および/または三次構造をも有しうる。

#### 【0315】

##### サンプルにおける核酸に対する担体核酸の比

例えば、サンプルにおける核酸から配列決定ライブラリーを製造するために、核酸を含むサンプルに特定の量のCNAが添加される。幾つかの場合には、サンプルにおける全核酸の量とサンプルに添加されるCNAの量との比は少なくとも1:100、1:50、1:10、1:1、10:1、50:1、100:1、500:1、1000:1、2000:1、または5000:1である。幾つかの場合には、サンプルにおける標的核酸の量とサンプルに添加されるCNAの量との比は少なくとも1:100、1:50、1:10、1:1、10:1、50:1、100:1、500:1、1000:1、2000:1または5000:1である。幾つかの場合には、サンプルにおける全核酸の量とサンプルに添加されるCNAの量との比は多くとも10:1、1:1、1:10、1:50、1:100、1:500、1:1000、1:2000または1:5000である。幾つかの場合には、サンプルにおける標的核酸の量とサンプルに添加されるCNAの量との比は多くとも10:1、1:1、1:10、1:50、1:100、1:500、1:1000、1:2000または1:5000である。幾つかの場合には、サンプルにおける全核酸の量とサンプルに添加されるCNAの量との比は約1:1~約1:100の範囲内である。幾つかの場合には、サンプルにおける標的核酸の量とサンプルに添加されるCNAの量との比は約1:1~約1:100の範囲内である。幾つかの場合には、該比はモル比である。

#### 【0316】

##### 配列決定ライブラリーを製造する際のCNAの使用法

本明細書の開示は配列決定ライブラリーの製造方法を含む。該方法は、配列決定ライブラリーの製造の効率および/または収率を改善するために、本明細書に開示されているCNAを添加することを含みうる。配列決定ライブラリーは、配列決定に付される核酸分子の集団を意味しうる。該方法は、標的核酸および/またはアダプター（例えば、配列決定アダプター）を含むサンプル、ならびに1以上のCNAを得ることを含みうる。該方法は更に、配列決定ライブラリーを製造するための1以上を含みうる。該方法はまた、配列決定ライブラリーにおける1以上の核酸を配列決定することを含みうる。CNAは配列決定されないことが可能である。例えば、CNAはライブラリーから物理的に除去されることが可能であり、あるいは、配列決定ライブラリーの製造における1以上の工程にそれが関与しないように設計されることが可能である。

10

20

30

40

50



## 【0317】

該方法は、標的核酸および/またはアダプターを含むサンプルにCNAを添加することを含みうる。サンプルに添加されるCNAの量は少なくとも0.1ng、0.5ng、1ng、5ng、10ng、20ng、30ng、40ng、50ng、60ng、70ng、80ng、90ng、100ng、150ng、200ng、300ng、400ngまたは500ngである。幾つかの場合には、CNAの量は0.1ng~200ng、1ng~100ng、5ng~80ng、10~60ng、または20ng~50ngでありうる。サンプルにおけるCNAの濃度は少なくとも0.1ng/mL、0.5ng/mL、0.6ng/mL、0.8ng/mL、1ng/mL、2ng/mL、5ng/mL、10ng/mL、0.01ng/μL、0.05ng/μL、0.1ng/μL、0.2ng/μL、0.4ng/μL、0.8ng/μL、1ng/μL、1.2ng/μL、1.5ng/μL、2ng/μL、5ng/μLまたは10ng/μLでありうる。幾つかの場合には、サンプルに添加されるCNAの量は約1ng/15μL~約5ng/15μLの範囲内でありうる。幾つかの場合には、サンプルに添加されるCNAの量は約0.05ng/μL~約0.5ng/μLの範囲内でありうる。

10

## 【0318】

本明細書における方法は、本開示の全体にわたって記載されている任意のタイプの合成核酸を添加することを含みうる。例えば、該方法は、以下の合成核酸、すなわち、配列決定ライブラリーの製造のための合成核酸、標的核酸の相対的存在量を正規化するための合成核酸(例えば、既知濃度の合成核酸)、および/またはサンプルにおける核酸の多様性減少を決定するための合成核酸の1以上を添加することを含みうる。

20

## 【0319】

核酸抽出

該方法はサンプルから核酸(例えば、標的核酸、無細胞核酸)を抽出することを含みうる。抽出は、サンプル中に存在しうる他の細胞成分および汚染物、例えば生物学的流体または組織サンプルから核酸を分離することを含みうる。幾つかの場合には、フェノールクロロホルム抽出または有機溶媒(例えば、エタノールまたはイソプロパノール)による沈殿により抽出を行う。幾つかの場合には、核酸結合カラムを使用して抽出を行う。幾つかの場合には、商業的に入手可能なキット、例えばQiagen Qiaamp循環核酸キット(Circulating Nucleic Acid Kit) Qiagen Qubit dsDNA HSアッセイキット、Agilent(商標)DNA 1000キット、TruSeq(商標)配列決定ライブラリー製造(Sequencing Library Preparation)、または核酸結合スピンカラム(例えば、Qiagen DNAミニプレップキット)を使用して抽出を行う。幾つかの場合には、無細胞核酸の抽出は濾過または限外濾過を含みうる。

30

## 【0320】

CNAは抽出の前または途中でサンプルに添加されうる。例えば、担体核酸は、それが抽出試薬、例えば抽出バッファーと混合される前に、サンプルに添加されうる。あるいは、担体核酸は、抽出試薬、例えば抽出バッファーに添加され、ついでそれがサンプルと混合されうる。ある場合には、CNAはサンプルと抽出試薬、例えば抽出バッファーとの混合物にも添加されうる。これらの場合には、標的核酸およびCNAは同時に抽出されうる。

40

## 【0321】

サンプルへのCNAの添加は核酸抽出の収率を増加させうる。標的核酸をCNAと共に抽出する収率は、標的核酸をCNAの非存在下で抽出する収率よりも、例えば少なくとも10%、20%、40%、60%、80%、100%、2倍、4倍、6倍、8倍または10倍高くなりうる。幾つかの場合には、CNAは、核酸抽出後に標的核酸を含むサンプルに添加されうる。該抽出は少なくとも10ng、50ng、100ng、200ng、300ng、400ng、500ng、600ng、700ng、800ng、900ngまたは1000ngの核酸を与えうる。

50

## 【0322】

核酸の精製

該方法は、標的核酸を精製することを含みうる。典型的な精製方法には、エタノール沈殿、イソプロパノール沈殿、フェノールクロロホルム精製、およびカラム精製（例えば、アフィニティに基づくカラム精製）、透析、濾過または限外濾過が含まれる。

## 【0323】

CNAは精製の前または途中にサンプルに添加されうる。例えば、担体核酸は、それが精製試薬、例えば精製バッファーと混合される前に、サンプルに添加されうる。あるいは、担体核酸は、精製試薬、例えば精製バッファーに添加され、ついでそれがサンプルと混合されうる。ある場合には、CNAはサンプルと精製試薬、例えば精製バッファーとの混合物にも添加されうる。これらの場合には、標的核酸およびCNAは同時に抽出されうる。

10

## 【0324】

サンプルへのCNAの添加は核酸精製の収率を増加させうる。標的核酸をCNAと共に精製する収率は、標的核酸をCNAの非存在下で精製する収率よりも、例えば少なくとも10%、20%、40%、60%、80%、100%、2倍、4倍、6倍、8倍または10倍高くなりうる。幾つかの場合には、CNAは、核酸精製後に標的核酸を含むサンプルに添加されうる。幾つかの場合には、CNAが添加されたサンプルにおける核酸の精製はサンプルにおける全核酸の少なくとも1pg、10pg、50pg、100pg、500pg、1ng、5ng、10ng、50ng、100ng、200ng、300ng、400ng、500ng、600ng、700ng、800ng、900ngまたは1000ngを与える。幾つかの場合には、CNAが添加されたサンプルにおける核酸の精製はサンプルにおける標的核酸の少なくとも1pg、10pg、50pg、100pg、500pg、1ng、5ng、10ng、50ng、100ng、200ng、300ng、400ng、500ng、600ng、700ng、800ng、900ngまたは1000ngを与える。

20

## 【0325】

断片化

該方法は、標的核酸を断片化することを含みうる。標的核酸の断片化は、例えば機械的剪断、サンプルをシリンジに通すこと、超音波処理、加熱処理またはそれらの組合せにより行われうる。幾つかの場合には、標的核酸の断片化は、ヌクレアーゼまたはトランスポザゼを含む酵素を使用することにより行われる。断片化に使用されるヌクレアーゼは、制限エンドヌクレアーゼ、ホーミングエンドヌクレアーゼ、ニッキングエンドヌクレアーゼ、高忠実度制限酵素、または本明細書に開示されている任意の酵素を含みうる。該方法は、標的核酸を、ある長さ、例えば少なくとも50、60、80、100、120、140、160、180、200、300、400、500、1000、2000、4000、6000、8000または10000bpの長さの断片に断片化することを含みうる。CNAは標的核酸の断片化の前にサンプルに添加されうる。CNAは標的核酸の断片化の後でサンプルに添加されうる。

30

## 【0326】

A-テーリング

該方法は、標的核酸に対してA-テーリング(tailing)を行うことを含みうる。A-テーリング反応は、1以上のA-テーリング酵素を使用して行われうる。例えば、単一の3'アデニン残基を付加する非ブルーフリーディングDNAポリメラーゼおよびdATPと共にDNAをインキュベートすることにより、アデニン(A)残基を付加することが可能である。A-テーリングの前に、標的核酸を含むサンプルにCNAを添加することが可能である。あるいは、A-テーリングの後に、標的核酸を含むサンプルにCNAを添加することが可能である。

40

## 【0327】

末端修復

50

該方法は、標的核酸に対して末端修復を行うことを含みうる。例えば、標的核酸が配列決定ライブラリーの他の工程に適合しうるように、標的核酸に対して末端修復が行われうる。末端修復反応は、1以上の末端修復酵素を使用して行われうる。DNAを修復するための酵素はポリメラーゼおよびエキソヌクレアーゼを含みうる。例えば、ポリメラーゼは、5'から3'の方向に、DNA鎖の欠失塩基を埋めることが可能である。得られる二本鎖DNAは元の最長DNA鎖と実質的に同じ長さを有しうる。エキソヌクレアーゼは3'オーバーハングを除去しうる。得られる二本鎖DNAは元の最短DNA鎖と実質的に同じ長さを有しうる。

#### 【0328】

CNAは、末端修復の前に、標的核酸を含むサンプルに添加されうる。幾つかの場合には、CNAの添加は末端修復反応の効率を、例えば少なくとも10%、20%、40%、60%、80%または100%増加させる。幾つかの場合には、CNAは、末端修復の後で、標的核酸を含むサンプルに添加されうる。幾つかの場合には、CNAの添加は、酵素、例えば末端修復酵素の活性および/または機能を維持しうる。例えば、酵素は、核酸の量が少ないサンプルにおいては、低下した活性または異常な機能を有する可能性があり、CNAの添加はサンプルにおける全核酸量を増加させることが可能であり、その結果、酵素はサンプルにおいて正常に機能することが可能である。

#### 【0329】

##### アダプターの結合

該方法は、1以上のアダプターを標的核酸に結合させることを含みうる。アダプターは、プライマー伸長、逆転写またはハイブリダイゼーションにより標的核酸に結合されうる。幾つかの場合には、アダプターは連結により標的核酸に結合される。例えば、アダプターは、リガーゼにより標的核酸に結合されうる。例えば、アダプターは粘着末端連結または平滑末端連結により標的核酸に結合されうる。幾つかの場合には、アダプターはトランスポザーゼにより標的核酸に結合されうる。標的核酸は3'末端、5'末端またはそれらの両方の末端においてアダプターに結合されうる。幾つかの場合には、標的核酸は、両方の末端において、同じアダプターまたは異なるアダプターに結合される。幾つかの場合には、標的核酸は、一方の末端において、1以上のアダプターに結合されうる。

#### 【0330】

CNAは結合工程の前に添加されうる。あるいは、CNAは結合工程の後で添加されうる。CNAは連結反応に抵抗しうる。例えば、CNAは、標的核酸および/またはアダプターとの連結に抵抗しうる。これらの場合、CNAが結合工程の前に添加された場合、それらは標的核酸またはアダプターのいずれにも連結せず、配列決定工程において配列決定されない。他の場合には、CNAは結合工程の前にサンプルから除去されうる。あるいは、CNAはサンプル抽出の後および結合工程の前に除去されうる。

#### 【0331】

サンプルにおける標的核酸にアダプターを結合させる前に、サンプルを酵素で処理することが可能である。例えば、サンプルをエンドヌクレアーゼで処理して、連結部位、例えば粘着末端または平滑末端を生成させることが可能である。あるいは、アダプターが標的核酸に結合した後、サンプルを酵素で処理することが可能である。

#### 【0332】

##### 増幅

該方法は、標的核酸を増幅することを含みうる。増幅は、核酸配列のコピー数を増加させるための任意の方法を意味しうる。例えば、増幅は、例えば1以上のポリメラーゼ連鎖反応において、ポリメラーゼを使用して行われうる。増幅は、当技術分野で公知の方法を用いて行われうる。これらの方法は、しばしば、核酸またはその相補体の複数のコピーの産物触媒形成によるものである。そのような方法の1つとして、以下のものを含むポリメラーゼ連鎖反応(PCR)が挙げられる: AFLP(増幅断片長多型)PCR、対立遺伝子特異的PCR、Alu PCR、アセンブリ、非対称PCR、コロニーPCR、ヘリカーゼ依存性PCR、ホットスタートPCR、インバースPCR、in situ(インシ

10

20

30

40

50

トゥ)PCR、配列間特異的PCRまたはISSR PCR、デジタルPCR、ドロップレット(droplet)デジタルPCR、線形後指数関数的PCRまたはレート(Late)PCR、ロング(long)PCR、ネスティッドPCR、リアルタイムPCR、二重PCR、マルチプレックスPCR、定量的PCRまたはシングルセルPCR。リガーゼ連鎖反応(LCR)、核酸配列ベース増幅(NASBA)、線形増幅、等温線形増幅、Q-ベータ-レプリカーゼ法、3SR、転写媒介増幅(TMA)、鎖置換(Strand Displacement)増幅(SDA)またはローリングサークル増幅(RCA)を含む他の増幅方法も用いられうる。

#### 【0333】

CNAは増幅前に添加されうる。あるいは、CNAは増幅後に添加されうる。CNAは増幅されないことが可能である。例えば、CNAは、増幅を阻害する修飾を含みうる。これらの場合、CNAが増幅前に添加されると、それは増幅されない。したがって、CNAは配列決定ライブラリーに存在しない、または配列決定されないことが可能である。

10

#### 【0334】

##### CNAの除去

該方法は更に、CNAをサンプルから除去することを含むことが可能であり、これは、しばしば、CNAが配列決定されることを妨げる。幾つかの場合には、該方法は、サンプルからCNAの一部または全部を除去して、配列決定サンプルを調製することを含む。得られる配列決定サンプルはCNAを含有していないことが可能であり、配列決定にそのまま使用されうる。幾つかの場合には、該方法は、サンプルにおける他の核酸、例えば標的核酸、アダプターまたはアダプターの多量体よりも優先的に、少なくとも1つのCNAを除去することを含む。

20

#### 【0335】

CNAの除去は、酵素を使用して行われうる。例えば、CNAは酵素、例えば酵素消化により分解されうる。幾つかの場合には、該方法は、ヌクレアーゼを使用してCNAを除去することを含む。例えば、該方法は、エンドヌクレアーゼ、例えばI型、II型(II S型、II G型を含む)、III型またはIV型エンドヌクレアーゼを使用して、CNAを除去することを含みうる。該方法は、制限エンドヌクレアーゼ、例えばAatII、Acc65I、AccI、AclI、AatII、Acc65I、AccI、AclI、AfeI、AflII、AgeI、ApaI、ApaLI、ApoI、AscI、AseI、AsiSI、AvrII、BamHI、BclI、BglII、Bme1580I、BmtI、BsaHI、BsiEI、BsiWI、BspEI、BspHI、BsrGI、BssHII、BstBI、BstZ17I、BtgI、ClaI、DraI、EaeI、EagI、EcoRI、EcoRV、FseI、FspI、HaeII、HincII、HindIII、HpaI、KasI、KpnI、MfeI、MluI、MscI、MspA1I、MfeI、MluI、MscI、MspA1I、NaeI、NarI、NcoI、NdeI、NgoMIV、NheI、NotI、NruI、NsiI、NspI、PacI、PciI、PmeI、PmlI、PsiI、PspOMI、PstI、PvuI、PvuII、SacI、SacII、SalI、SbfI、ScaI、SfcI、SfoI、SgrAI、SmaI、SmlI、SnaBI、SpeI、SphI、SspI、StuI、SwaI、XbaI、XhoI、XmaIまたはそれらの任意の組合せを使用して、CNAを除去することを含みうる。該方法は、前記で挙げられていないDNアーゼ、例えばエキソデオキシリボヌクレアーゼを使用して、CNAを除去することを含みうる。該方法は、ウラシルDNAグリコシラーゼ(UDG)、DNAグリコシラーゼ-リアーゼ(エンドヌクレアーゼVII)またはそれらの混合物(例えば、ウラシル特異的切断試薬(USER)酵素)を使用して、CNAを除去することを含みうる。該方法は、RNA誘導DNアーゼの部位、例えばCRISPR関連タンパク質ヌクレアーゼ、例えばCas9、RNAアーゼを使用して、CNAを除去することを含みうる。該方法は、RNAアーゼ、例えばエンドリボヌクレアーゼ、例えばRNAアーゼA、RNAアーゼH、RNAアーゼII、RNAアーゼL、RNAアーゼP、RNAアーゼPhyM、RNAアーゼT1、RNAアーゼ

30

40

50

T 2、RNアーゼU 2、RNアーゼV、またはエキソリボヌクレアーゼ、例えばポリヌクレオチドホスホリラーゼ、RNアーゼPH、RNアーゼR、RNアーゼD、RNアーゼT、オリゴリボヌクレアーゼ、エキソリボヌクレアーゼIまたはエキソリボヌクレアーゼII、あるいはそれらの任意の組合せを使用して、担体合成核酸を除去することを含みうる。幾つかの場合には、該方法は、当技術分野で公知の任意の核酸分解試薬を使用して、CNAを除去することを含む。幾つかの場合には、該方法は、CNAを物理的処理、例えば加熱、冷却またはせん断に付すことにより、CNAを除去することを含みうる。幾つかの場合には、CNAの除去方法は、標的核酸、アダプター、または配列決定ライブラリーにおける他の任意の分子をサンプルから除去しない。幾つかの場合には、CNAの除去は酵素分解、例えばエンドヌクレアーゼ消化によっては行われない。

10

## 【0336】

CNAを除去するために、該方法は、酵素が機能する温度でCNAを酵素と共にインキュベートすることを含みうる。例えば、該方法は、10 ~ 80、例えば20 ~ 60、20 ~ 40、30 ~ 40、または20 ~ 25の温度でCNAを酵素と共にインキュベートすることを含みうる。該方法は、少なくとも10、20、25、30、31、32、33、34、35、36、37、38、39、40、50、60または70の温度でCNAを酵素と共にインキュベートすることを含みうる。幾つかの場合には、該方法は、約20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41または42の温度でC

20

## 【0337】

CNAを除去するために、該方法は、酵素が機能的である時間にわたってCNAを酵素と共にインキュベートすることを含みうる。幾つかの場合には、該方法は、少なくとも1分間、5分間、10分間、15分間、20分間、30分間、40分間、50分間、1時間、2時間、5時間、12時間、24時間、48時間または72時間、CNAを酵素と共にインキュベートすることを含みうる。

## 【0338】

該方法は、アフィニティに基づく枯渇 (depletion) により担体合成核酸を除去することを含みうる。アフィニティに基づく枯渇は、1以上の固定化タグを含みうる担体合成核酸上で行われうる。これらの場合には、該方法は、固体支持体に固定化タグを結合させることによりCNAを除去することを含みうる。そのような固体支持体は紙、ガラス(例えば、制御孔ガラス(CPG))、プラスチック(例えば、ポリメチルアクリル酸、ポリエチレン、ポリプロピレン、ポリアクリレート、ポリメチルメタクリレート、ポリ塩化ビニル、ポリテトラフルオロエチレン、ポリスチレン、ポリカーボネートマクロポロスピリスチレン(MPPS)またはナイロン)、ポリアクリルアミド、酢酸セルロース、硝酸セルロース、ニトロセルロース、シリコンまたは他の金属、または光ファイバーでありうる。

30

## 【0339】

アフィニティに基づく枯渇のための固体支持体は任意の形状および形態に成形されうる。幾つかの場合には、固体支持体は、ウェル、桶(トラフ)、台(ペDESTAL)、疎水性もしくは親水性パッチ、ダイカット接着剤リザーバ、または流体流動に対する他の物理的障壁の形態の、分離した隔離領域を有する平面装置の形態で製造されうる。そのような固体支持体の例には、スライド、マイクロプレート、シート、フィルム、ディップスティックなどが含まれる。

40

## 【0340】

他の場合には、固体支持体は、被覆カチオン性表面を含有するビーズまたはペレットの形態でありうる。ビーズは、被覆固体支持体上のプローブ密度を増加させるための手段を提供しうる。ビーズは、例えばアミノ化によりビーズをカチオン性にするのに適した種々の表面化学または官能基(例えば、アミン、カルボキシルまたはヒドロキシル)を提供し

50

うる。適切なビーズ組成物には、例えば、プラスチック、例えば、ポリスチレン、メチル  
 スチレン、アクリルポリマー、セラミック、ガラス、ポリマー材料、例えば、架橋デキス  
 トラン、セルロース、ナイロンおよびラテックス、常磁性材料、二酸化チタン、ラテック  
 スが含まれる。ビーズは、任意のタイプの中実もしくは中空の球、ボール、ベアリング、  
 シリンダーまたは他の固体形状を含みうる。ビーズは本質的に多孔質または非多孔質であ  
 りうる。多孔質ビーズの使用は、核酸検出に利用可能なビーズの表面積を増加させうる。  
 ビーズサイズは、100 nm ~ 5 mm、例えば0.2 μm ~ 200 μm、または0.5 μ  
 m ~ 5 μmの範囲でありうる。幾つかの場合には、固体支持体は磁性または磁気感受性で  
 ありうる。固体支持体は被覆されうる。被覆（コーティング）は固定化タグに結合しうる  
 。例えば、固体支持体は、固定化タグの結合相手、例えば、ストレプトアビジン、抗原、  
 抗体（例えば、抗ポリヒスチジン抗体）、グルタチオンSトランスフェラーゼ、またはそ  
 の類似体で被覆されうる。

10

#### 【0341】

該方法は、サイズに基づく枯渇によりCNAを除去することを含みうる。例えば、サイ  
 ズに基づく枯渇は、多孔性ビーズ（例えば、固相可逆的固定化（Solid Phase  
 Reversible Immobilization）（SPRI）磁気ビーズ、電  
 気泳動ゲル精製（例えば、アガロースゲル精製）および/またはゲル濾過を用いて行われ  
 うる。幾つかの場合には、該方法は、50 bp、100 bp、200 bp、300 bp、  
 400 bp、500 bp、600 bp、700 bp、800 bp、900 bp、1 kb、  
 2 kb、5 kbまたは10 kbの長さを有する合成核酸を除去することを含みうる。例え  
 ば、該方法は、少なくとも500 bpの長さを有する合成核酸を除去することを含みうる  
 。

20

#### 【0342】

##### 配列決定

該方法は、配列決定ライブラリーにおける標的核酸および/またはアダプターを配列決  
 定することを含みうる。配列決定は、マキサム-ギルバート（Maxam-Gilbert）  
 配列決定、鎖終結配列決定、ショットガン配列決定またはブリッジPCRを含む基本  
 的な配列決定方法により行われうる。配列決定は、大規模並列配列決定法（例えば、次世  
 代シーケンシング）、例えば、ハイスループット配列決定、パイロシーケンシング、合成  
 による配列決定、単一分子配列決定、ナノポア配列決定、半導体配列決定、連結による配  
 列決定、ハイブリダイゼーションによる配列決定、RNA-Seq（Illumina）  
 、デジタル遺伝子発現（Digital Gene Expression）（Heli  
 cos）、次世代シーケンシング、合成による単分子配列決定（Single Mole  
 cule Sequencing by Synthesis）（SMSS）（Heli  
 cos）、大規模並列配列決定、クローン単一分子アレイ（Clonal Single  
 Molecule Array）（Solexa）、ショットガン配列決定、マキサム  
 -ギルバート（Maxam-Gilbert）またはサンガー（Sanger）配列決定  
 、プライマーウォーキング、Illumina、PacBio、SOLiD、Ion T  
 orrent、454またはナノポアプラットフォームを使用する配列決定によっても行  
 われうる。配列決定を次世代シーケンシング法により行う場合、ここで作製される配列決  
 定ライブラリーは次世代シーケンシングライブラリーである。

30

40

#### 【0343】

本明細書の全体にわたって用いる、数字または数値範囲に関する「約」なる語は、言及  
 された数字または数値範囲が実験的変動内（または統計的実験誤差内）の近似であること  
 、あるいは数字または数値範囲が、例えば、示されている数字または数値範囲の1% ~ 1  
 5%で変動しうることを意味する。例えば、「約」なる語は、示されている数または値の  
 ±10%を意味する。

#### 【0344】

本明細書中で用いる「または」なる語は、特に示されていない限り、非排他的であり、  
 例えば、「AまたはB」は「BではなくA」、「AではなくB」および「AおよびB」を

50

含む。

【図面の簡単な説明】

【0345】

【図1】図1は本開示の基本的方法の概要図を示す。

【図2】図2は典型的な感染の概要図を示す。

【図3】図3は、本発明で提供する方法の幾つかの一般的スキームを示す。

【図4】図4は8つの典型的なスパーク(Spark)サイズセットスパイクインの設計を示す。

【図5】図5は、多様性減少を決定するための本発明で提供する方法の一般的スキームを示す。

【図6】図6は典型的なスパーク(Spark)スパイクインの設計を示す。

【図7】図7は、本発明で提供する方法を実行するようにプログラムまたは構成されたコンピュータ制御システムを示す。

【図8】図8は、担体(carrier)核酸が加えられうる配列決定ライブラリーの製造における工程を示す。

【図9】図9は110個の典型的なIDスパイクに関する正規化リード数を示す。

【図10】図10は110個の典型的なシグナル正規化IDスパイクに関する正規化リード数を示す。

【図11】図11は、病原体トレーサーを使用してシゲラ・フレックスネリ(*Shigella flexneri*)の陽性対照からの交差汚染を特定するための方法からの結果を示す。

【図12】図12は、病原体トレーサーを使用してサルモネラ・エンテリカ(*Salmonella enterica*)の陽性対照からの交差汚染を特定するための方法からの結果を示す。

【図13】図13は、病原体トレーサーを使用してスタフィロコッカス・アウレウス(*Staphylococcus aureus*)の陽性対照からの交差汚染を特定するための方法からの結果を示す。

【図14】図14は、病原体トレーサーを使用してシュードモナス・エルジノーサ(*Pseudomonas aeruginosa*)の陽性対照からの交差汚染を特定するための方法からの結果を示す。

【図15】図15は、病原体トレーサーを使用してクロストリジウム・スポロゲネス(*Clostridium sporogenes*)の陽性対照からの交差汚染を特定するための方法からの結果を示す。

【図16】図16は典型的なスパーク(Spark)サイズスパイクインの相対収率に対するサイズ選択ライブラリ処理方法の効果を示す。

【図17】図17は、種々のGC含量の核酸を用いて、酵素加熱失活工程を含む配列決定ライブラリーの製造方法からの結果を示す。

【図18】図18は、種々のGC含量の核酸を使用する、酵素加熱失活工程を欠く配列決定ライブラリーの製造方法からの結果を示す。

【図19】図19は、両端の連結を妨げる大きなサイズを有する担体合成核酸を使用する、配列決定ライブラリーの典型的な製造方法を示す。

【図20】図20Aは、脱塩基部位および修飾を有する担体合成核酸を使用する、配列決定ライブラリーの典型的な製造方法を示す。図20Bは、エンドヌクレアーゼVII消化の非存在下で担体合成核酸を使用する配列決定ライブラリーの製造からの結果を示す。

レーンA1: Tape Stationラダー。レーンB1: CNA無し、第1複製。レーンC1: CNA無し、第2複製。レーンD1: CNA無し、第3複製。レーンE1: 10ngのCNA、第1複製。レーンF1: 10ngのCNA、第2複製。レーンG1: 10ngのCNA、第3複製。図20Cは、エンドヌクレアーゼVII消化の存在下で担体合成核酸を使用する配列決定ライブラリーの製造からの結果を示す。レーンA1: Tape Stationラダー。レーンB1: CNA無し、第1複製。レーンC1: CNA無し

10

20

30

40

50

、第2複製。レーンD1：CNA無し、第3複製。レーンE1：10ngのCNA、第1複製。レーンF1：10ngのCNA、第2複製。レーンG1：10ngのCNA、第3複製。

【図21】図21Aは、エンドヌクレアーゼVII消化の非存在下で脱塩基部位を有する担体合成核酸を使用する配列決定ライブラリーの製造からの結果を示す。図21Bは、エンドヌクレアーゼVII消化の存在下で脱塩基部位を有する担体合成核酸を使用する配列決定ライブラリーの製造からの結果を示す。

【図22】図22は、DNA-RNAハイブリッドを有する担体合成核酸の典型的配列を示す。「rX」なる文字（例えば、rG、rC、rA）はRNA配列を示す。

【図23】図23Aは、DNA-RNAハイブリッドを有する担体合成核酸を使用する、配列決定ライブラリーの典型的な製造方法を示す。図23Bは、DNA-RNAハイブリッドを有する担体合成核酸を使用する、配列決定ライブラリーの製造からの結果を示す。

【0346】

#### 実施例

##### 実施例1：無細胞DNA配列決定アッセイによる診断

無細胞血漿サンプルを調製する。次世代シーケンシングのためのDNAライブラリーを、既に記載されているとおりに製造する(De Vlamincck I, Khush K, Strehl Cら, Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* 2013; 155(5): 1178-87.; De Vlamincck I, Martin L, Kertesz Mら, Noninvasive monitoring of infection and rejection after lung transplantation. *Proceedings of the National Academy of Sciences of the United States of America* 2015; 112(43): 13336-41; それらのそれぞれの全体を参照により本明細書に組み入れることとする)。配列決定をIllumina NextSeq装置で行い、分析する。簡潔に説明すると、低品質リードを除去した後、リードをヒト参照ゲノム(例えば、hg19)に対してマッピングする。残りのリードをウイルス性、細菌性、真菌性および他の真核性病原体のキュレート参照データベースに対してマッピングする。個々の病原体の存在量を、血漿1ml当たりの特定の病原体由来の核酸量の絶対的尺度である、体積当たりのゲノムコピー数として表す。耐性を付与することが知られている配列を特定するために、更なる分析を行うことが可能である。

【0347】

##### 患者の血漿の直接次世代シーケンシング(NGS)

DNA抽出の前に既知濃度の合成DNA分子を血漿サンプルに添加する。改変された磁気ビーズに基づく方法(Omega Biotek, Norcross, GA)を用いてDNAを抽出する。改変されたライブラリー調製キット(NuGEN, San Carlos, CA)を使用して、NGSライブラリーを構築する。陰性(バッファーを含有するが血漿を含有しない)および陽性(健常ドナーからの血漿および既知濃度の剪断された実験室由来の病原体DNAを含有する)対照サンプルを、サンプルと並行して処理する。75サイクルのシングルエンド(single-end)デュアルインデックス(dual-index)配列決定キットを使用するIllumina NextSeqで全3個のDNAライブラリータイプを多重化し、配列決定する。

【0348】

##### バイオインフォマティクス分析

病原体リードをNGSリードセットから定量する。簡潔に説明すると、低品質リードを破棄した後、ヒト参照配列(例えば、hg19)に対してアライメントすることによりヒトリードを除去する。完全スパイクイン配列のデータベースに対するアライメントにより、合成スパイクインリードを特定する。残りのリードを、ウイルス、原核生物および真菌

10

20

30

40

50



(真菌、原生動物および寄生生物を含む)の8000個を超える参照配列のキュレートデータベースに対してアライメントする。PCR重複(duplicate)または配列決定装置のエラーに由来すると推定される重複リードをアライメントに基づき特定し、除去する。生物の相対的存在量を、推定される重複排除(deduped)リード(EDR)または百万当たりのリード(RPM; サンプルに関する全リードに対して正規化されたもの)、またはサンプルの体積当たりのリード(MPM; 1マイクロリットル当たりの分子)として表す。MPMは、血漿1マイクロリットルにおける各生物に関して表される、核酸断片の推定数を見積もる正規化量である。この計算は、抽出の開始時に血漿に添加された既知量の合成DNAに対して正規化された各生物に関して存在する配列の数から導出される。

10

## 【0349】

2つの患者サンプルの処理の説明は以下の通りである。スパイク(Spank)-75B(配列番号120)、スパーク(Spark)-32/52/75/100/125/150/175/350(配列番号111~118)、およびIDスパイク(血漿1マイクロリットル当たりの各スパイクインの $3 \times 10^5$ 個の分子)の混合物を血漿に添加する。各サンプルには同じスパイク/スパーク混合物を加えるが、異なるIDスパイクを加える。添加された血漿を16,000gで10分間遠心分離し、無細胞血漿からなる上清を新鮮なチューブに移す。デュアルインデックスIlluminaアダプターを添加するための無細胞DNA抽出およびライブラリー調製の後、サンプルをプールし、それと並行して、陰性対照サンプルおよび陽性対照サンプルを処理し、ついでIllumina NextSeqで配列決定する。典型的には、約4億個のリードがバッチ内のサンプルに分布し、任意の個々のサンプルのリード数は、サンプルが含む全ライブラリープールの割合に比例し、そしてこの割合は無細胞血漿中のDNAの量に比例する。

20

## 【0350】

計算分析: 個々のサンプルのリードを、対応するアダプターバーコード配列に基づいて特定した(「脱多重化」)。アダプター二量体配列の除去、および品質に基づくリードのトリミングの後、ヒトゲノム、スパイクインおよび病原体ゲノム参照配列に対するアライメントにより、リード配列の推定起源を特定した。IDスパイクおよびSPANK(スパイク)-75Bリードの数を、スパイクインアライメントを用いてカウントし、各添加分子内に埋め込まれたランダム化配列タグを使用して、スパイク(SPANK)-75Bリードを重複排除(duplicate)(「デデュープ(dedupe)」)した。ゲノム位置に基づいて病原体アライメントを重複排除し、機械学習アプローチを用いて、各リードの最も可能性の高い分類学的起源を決定して、特定の病原体に起因する推定重複リードを得た。正規化病原体存在量を1マイクロリットル当たりの病原体分子(MPM-スパイク)としての濃度で表し、以下のとおりに計算する:  $MPM - \text{スパイク} = (\text{推定重複排除リード} / \text{スパイク} - 75B \text{ の数}) \times c$  (ここで、 $c$ は、サンプルに添加されたスパイク-75Bリードの濃度、すなわち、1マイクロリットル当たり $3 \times 10^5$ リードである)。

30

## 【0351】

IDスパイクは、配列決定バッチにおける各サンプルに関してユニークであるスパイクインの一種でありうる。スパイク分子は全ライブラリーにわたって一定濃度で添加されうる。したがって、個々のライブラリーにおいて検出される重複排除スパイク分子の数は、そのライブラリーにおいて検出可能な最小濃度の代用物でありうる。より一般的には、それは、そのライブラリーが元のサンプルにおける核酸(例えば、DNA)分子を核酸配列決定データにおけるリードに変換した効率に正比例しうる。スパイク分子の目的は、サンプルにおいて表される混合物における標的(例えば、病原体または疾患関連)分子の相対的存在量を確定するのを助けることでありうる。

40

【表 7】

計量	サンプル A	サンプル B, 病原体 1	サンプル B, 病原体 2
リードの数	58,078,895	35,045,795	35,045,795
ID スパイクリードの数	344,677	45,785	45,785
重複排除スパンク 75B リードの数	194,124	55,040	55,040
病原体に起因する推定重複排除リード	178,758.561	74,324.581	16,866.832
MPM スパンク	276,254.189	405,112.18	91,934.042

10

## 【0352】

実施例 2：ID スパイクの合成および処理

インテグレートド DNA テクノロジーズ (Integrated DNA Technologies) により合成された相補的な 100 マー (mer) オリゴヌクレオチドのペアをアニーリングさせることにより、110 個の典型的な ID スパイク (ID スパイク 28 ~ 137; 配列は表 1 に列挙されている) を構築した。ついで、これらの配列を 4 つのグループにサブプールし、ヒト血漿の 4 つのアリコートに加え、抽出し、血漿の無細胞 DNA と共に配列決定ライブラリーへと処理し、Illumina NextSeq 500 装置で配列決定した。110 個の ID スパイクのそれぞれにマッピングされたリードの数を決定し、ついで 100 万個の全リードごとに正規化した。100 万個の全リード当たり最小約 12,000 個および最大約 88,000 個を示すプロットを図 9 に示す。

20

## 【0353】

ID スパイクからのシグナルは調節可能である。ID スパイクは、例えば、図 9 に示されているシグナル中央値に基づいて各 ID スパイクの入力量を調節することにより、シグナル正規化されうる。シグナル正規化 ID スパイクの再試験は、図 10 に示されているとおり、より均一なリード数を与える。この場合、各 ID スパイクにマッピングされるリードの数は 100 万個の全リード当たり 10,000 ~ 25,000 の範囲内であり、ほとんどの ID スパイクは 100 万個の全リード当たり 15,000 ~ 20,000 の範囲内のリード数を有する。異なる ID スパイクのシグナルを等化または正規化する利点は、それが交差汚染検出の精度を向上させうることである。例えば、1 つの ID スパイクが典型的には 100 万当たり 70,000 個のリードを与え、別の ID スパイクが 19,000 個を与える場合、後者はより低感度の交差汚染トレーサでありうる。それらの 2 つの ID スパイク間のシグナルの正規化はより高い均一性をもたらさう。

30

## 【0354】

実施例 3：ID スパイクを使用するサンプルの交差汚染の測定

4 つのサンプルを調製する。ID スパイク 124 ~ 127 をサンプル 1 に添加する。ID スパイク 128 ~ 131 をサンプル 2 に添加する。ID スパイク 132 ~ 134 をサンプル 3 に添加する。ID スパイク 135 ~ 137 をサンプル 4 に添加する。それらの 4 つのサンプルを処理し、配列決定する。表 7 に ID スパイクのサブセットに関して示されているとおり、交差汚染 (例えば、ID スパイクが意図的には添加されなかったサンプルにおける ID スパイクに起因するリード) は 1:10,000 未満であることが示されている。

40

## 【表 8】

表 7:

	サンプル 1	サンプル 2	サンプル 3	サンプル 4
ID スパイク 124	70,896	1	1	2
ID スパイク 125	88,403	2	2	2
ID スパイク 126	81,767	2	1	2
ID スパイク 127	86,878	2	2	2
ID スパイク 128	1	58,131	1	1
ID スパイク 129	1	54,839	1	1
ID スパイク 130	1	49,834	1	1
ID スパイク 131	1	53,491	1	1
ID スパイク 132	0	0	27,537	1
ID スパイク 133	0	0	25,055	0
ID スパイク 134	1	0	32,739	0
ID スパイク 135	1	0	1	19,459
ID スパイク 136	0	0	0	17,030
ID スパイク 137	0	1	0	19,644

10

20

## 【 0 3 5 5 】

実施例 4 : ID スパイクを使用するサンプルの交差汚染の特定

4 つのサンプルを調製する。ID スパイク 1 2 4 をサンプル A に添加する。ID スパイク 1 2 3 をサンプル B に添加する。ID スパイク 1 2 2 をサンプル C に添加する。ID スパイク 1 1 9 をサンプル D に添加する。それらの 4 つのサンプルを処理し、配列決定する。サンプル A および B における ID スパイク 1 2 3 および ID スパイク 1 2 4 の両方からの相当数のリードは、表 8 に示されているとおり、それらの 2 つの ID スパイクストック間またはサンプル A および B の間の交差汚染を示している。

## 【表 9】

表 8:

	サンプル A	サンプル B	サンプル C	サンプル D
ID スパイク 119	0	0	0	20,804
ID スパイク 122	0	0	7,291	0
ID スパイク 123	21	23,451	0	0
ID スパイク 124	5,990	13	0	0

30

40

## 【 0 3 5 6 】

実施例 5 : ID スパイクを使用する交差汚染源の特定

幾つかのサンプルにおいては、実質的にだた 1 つの ID スパイクのみが特定されるが、他のサンプルにおいては、幾つかの汚染 ID スパイクが有意レベルで観察される。そのような状況においては、表 9 に示されているとおり、汚染 ID スパイクが何であるかが、例えば、マイクロタイタープレートの隣接ウェルからの交差汚染源を示しうる。

【表 10】

表 9:

	RD-1819- Lib144-1	RD-1819- Lib133-1	RD-1819- Lib137-1	EC-1991- Lib2-1	EC-1997- Lib2-1	RD-1819- Lib148-1
ID スパイク 052		38,541				
ID スパイク 053		65		5		
ID スパイク 054				156,813		
ID スパイク 055				40		
ID スパイク 056					2	
ID スパイク 057						
ID スパイク 058			57,163	0		
ID スパイク 059						0
ID スパイク 060	1	15				
ID スパイク 061						
ID スパイク 062					0	
ID スパイク 063				0		
ID スパイク 064					0	
ID スパイク 065						
ID スパイク 066			419			2
ID スパイク 067						1
ID スパイク 068	42,351	3				
ID スパイク 069		19				
ID スパイク 070						
ID スパイク 071						
ID スパイク 072			6		145,380	6
ID スパイク 073			11		1,516	3
ID スパイク 074			221		2,270	210
ID スパイク 075			46		388	58,921

## 【 0 3 5 7 】

実施例 6 : 病原体 DNA の陽性対照からの交差汚染を特定するためのトレーサー配列  
 実験室 (検査室) 病原体 DNA サンプルは陽性対照として使用されるが、臨床サンプル  
 を交差汚染し、偽陽性リードまたは診断をもたらすリスクを伴う。交差汚染検出を可能に  
 するために、トレーサーを実験室病原体 DNA サンプルに添加する。血液からの無細胞病

10

20

30

40

50

原体DNA断片のNGS検出の場合には、比較的短いトレーサーが使用されうる。なぜなら、無細胞病原体断片は比較的短く、例えば20～120bpであり、しばしば、平均約75bpであるからである。ここでは、病原体対照当たり1個のユニークトレーサー配列を含有する75bpの合成DNA二本鎖のセットをトレーサーとして使用する。実験室で得られたゲノムDNA(ATCCまたはNIST)を、DNase I(New England Biolabs)またはフラグメンターゼ(Fragmentase)ヌクレアーゼ混合物(New England Biolabs)を使用して剪断することにより、病原体断片を得た。

#### 【0358】

表10～12に列挙されている11種の異なる病原体由来のゲノムDNAを約75bpの平均断片長に個々に剪断し、精製し、定量した(Qubit, Thermo Fisher)。各断片プールに、別の75bpの合成DNA二本鎖(Integrated DNA Technologies)を約10倍高い質量で加えて(Qubit, Thermo Fisher)、各断片化病原体にユニークトレーサーを付与した。好ましくは、ゲノムDNAに対して少なくとも約5倍高いシグナルが該トレーサーに関して得られる。同時感染をシミュレートするために、病原体/トレーサーペアを3つの種々の組合せと一緒に混合し(表10～12に列挙されているとおり、混合物1は4つの病原体を含有し、混合物2は4つの病原体を含有し、混合物3は3つの病原体を含有していた)、ヒト血漿に加え、ヒト血漿において更に希釈して濃度系列を得、ついで無細胞DNA抽出、ライブラリー調製およびNGSに付した。データベースに対するアライメントは全サンプルにおける全11個のトレーサーおよび全11個の病原体の検出濃度を算出した。

#### 【0359】

表10～12および図11～図15に示されている結果はトレーサーと病原体との1:1のペア形成を示している。各場合において、トレーサーは病原体より高い濃度で検出され、濃度差は希釈系列にわたって一貫したままである。低レベルの交差汚染事象から予想されるような非常に低い濃度への外挿は、病原体の前にトレーサーが検出されることを強く示唆している。表10～12は全てのサンプルからのデータを示し、図11～15は、それぞれシゲラ・フレックスネリ(*Shigella flexneri*)、サルモネラ・エンテリカ(*Salmonella enterica*)、スタフィロコッカス・アウレウス(*Staphylococcus aureus*)、シュードモナス・エルジノーサ(*Pseudomonas aeruginosa*)およびクロストリジウム・スポロゲネス(*Clostridium sporogenes*)に関する代表的なトレーサー:病原体のペアをプロットしている。トレーサーは、それが意図的に添加されたサンプルでのみ観察された。未添加サンプルにおいては、1つの病原体、すなわち、大腸菌(*E. coli*)のみが観察された。それらのサンプルにおいては、大腸菌(*E. coli*)トレーサー#143は観察されなかったため、大腸菌(*E. coli*)はヒト血漿中に低レベルで存在していたと結論づけることが可能であり、これはこの共生生物ではよくあることである。

10

20

30

## 【表 1 1】

表 10: 病原体トレーサー混合物 1

病原体	混合物 1 -	混合物 1 -	混合物 1 -	混合物 1 -
	2	1	0.5	0.25
スパイク	pg/mL	pg/mL	pg/mL	pg/mL
	100 万当たりのリード数			
サルモネラ・エンテリカ ( <i>Salmonella enterica</i> )	348	163	85	60
トレーサー_138	1339	814	274	303
クロストリジウム・スポロゲネス ( <i>Clostridium sporogenes</i> )	387	170	75	68
トレーサー_141	2616	1318	541	463
シゲラ・フレックスネリ ( <i>Shigella flexneri</i> )	286	139	65	55
トレーサー_144	2865	1195	628	508
アスペルギルス・フミガーツス ( <i>Aspergillus fumigatus</i> )	150	73	31	25
トレーサー_147	1205	742	269	274
スタフィロコッカス・アウレウス ( <i>Staphylococcus aureus</i> )	0	0	0	0
トレーサー_139	0	0	0	0
マイコバクテリウム・ツベルクローシス ( <i>Mycobacterium tuberculosis</i> )	0	0	0	0
トレーサー_142	0	0	0	0
ブルデテラ・ペルツッシス ( <i>Bordetella pertussis</i> )	0	0	0	0
トレーサー_145	0	0	0	0
リーシュマニア・メジャー ( <i>Leishmania major</i> )	0	0	0	0
トレーサー_148	0	0	0	0
シュードモナス・エルジノーサ ( <i>Pseudomonas aeruginosa</i> )	0	0	0	0
トレーサー_140	0	0	0	0
大腸菌 ( <i>Escherichia coli</i> )	2	0	0	1
トレーサー_143	0	0	0	0
スタフィロコッカス・エピデルミデイス ( <i>Staphylococcus epidermidis</i> )	0	0	0	0
トレーサー_146	0	0	0	0

10

20

30

## 【表 1 2】

表 11: 病原体トレーサー混合物 2

病原体 スパイク	混合物 2 -	混合物 2 -	混合物 2 -	混合物 2 -	
	2 pg/mL	1 pg/mL	0.5 pg/mL	0.25 pg/mL	
	100 万当たりのリード数				
サルモネラ・エンテリカ ( <i>Salmonella enterica</i> )	0	0	0	0	
トレーサー_138	0	0	0	0	
クロストリジウム・スポロゲネス ( <i>Clostridium sporogenes</i> )	0	0	0	0	10
トレーサー_141	0	0	0	0	
シゲラ・フレックスネリ ( <i>Shigella flexneri</i> )	0	0	0	0	
トレーサー_144	0	0	0	0	
アスペルギルス・フミガーツス ( <i>Aspergillus fumigatus</i> )	0	0	0	0	
トレーサー_147	0	0	0	0	
スタフィロкокカス・アウレウス ( <i>Staphylococcus aureus</i> )	484	207	93	109	
トレーサー_139	1435	816	252	416	
マイコバクテリウム・ツベルクローシス ( <i>Mycobacterium tuberculosis</i> )	256	172	60	62	20
トレーサー_142	3942	1767	626	1327	
ブルデテラ・ペルツシス ( <i>Bordetella pertussis</i> )	207	152	52	45	
トレーサー_145	3149	1667	632	1093	
リーシュマニア・メジャー ( <i>Leishmania major</i> )	138	76	31	45	
トレーサー_148	1261	605	249	363	
シュードモナス・エルジノーサ ( <i>Pseudomonas aeruginosa</i> )	0	0	0	0	
トレーサー_140	0	0	0	0	
大腸菌 ( <i>Escherichia coli</i> )	1	7	1	0	30
トレーサー_143	0	0	0	0	
スタフィロкокカス・エピデルミデイス ( <i>Staphylococcus epidermidis</i> )	0	0	0	0	
トレーサー_146	0	0	0	0	

## 【表 13】

表 12: 病原体トレーサー混合物 3

病原体	混合物 3 -	混合物 3 -	混合物 3 -	混合物 3 -
	2	1	0.5	0.25
スパイク	pg/mL	pg/mL	pg/mL	pg/mL
	100 万当たりのリード数			
サルモネラ・エンテリカ ( <i>Salmonella enterica</i> )	0	0	0	0
トレーサー_138	0	0	0	0
クロストリジウム・スポロゲネス ( <i>Clostridium sporogenes</i> )	0	0	0	0
トレーサー_141	0	0	0	0
シゲラ・フレックスネリ ( <i>Shigella flexneri</i> )	0	0	0	0
トレーサー_144	0	0	0	0
アスペルギルス・フミガーツス ( <i>Aspergillus fumigatus</i> )	0	0	0	0
トレーサー_147	0	0	0	0
スタフィロコッカス・アウレウス ( <i>Staphylococcus aureus</i> )	0	0	0	0
トレーサー_139	0	0	0	0
マイコバクテリウム・ツベルクローシス ( <i>Mycobacterium tuberculosis</i> )	0	0	0	0
トレーサー_142	0	0	0	0
ブルデテラ・ペルツシス ( <i>Bordetella pertussis</i> )	0	0	0	0
トレーサー_145	0	0	0	0
リーシュマニア・メジャー ( <i>Leishmania major</i> )	0	0	0	0
トレーサー_148	0	0	0	0
シュードモナス・エルジノーサ ( <i>Pseudomonas aeruginosa</i> )	480	154	162	33
トレーサー_140	3544	1031	1132	227
大腸菌 ( <i>Escherichia coli</i> )	363	104	133	24
トレーサー_143	5020	1202	1359	295
スタフィロコッカス・エピデルミディス ( <i>Staphylococcus epidermidis</i> )	265	69	80	25
トレーサー_146	5399	1506	1820	361

10

20

30

## 【0360】

実施例 7 : スパーク配列を使用する、種々の長さの核酸の相対的収量の決定

8 個のスパーク ( Spark ) をヒト血漿に等モル量で加え、抽出し、血漿の無細胞 DNA と共に配列決定ライブラリーへと処理し、Illumina NextSeq500 装置で配列決定した。種々の処理方法 ( 例えば、図 16 に示されているとおり、種々のサイズ範囲を有するライブラリーのサブセットを選択するもの ) が、種々のスパークの相対的収率を決定することによりモニターされうる。更に、全てのサンプルに同じ量のスパークを添加した場合、例えば 8 個のスパークのそれぞれの 1 億個の分子を血漿 1 mL 当たりに加えた場合、与えられたスパークのリード数を用いて、サンプルにおける他の類似サイズの断片 ( 例えば、感染因子由来の無細胞 DNA ) の出発濃度を推測することが可能である。

40

## 【0361】

実施例 8 : 種々の GC 含量を有する合成核酸を使用する配列決定ライブラリーの製造

無細胞病原体核酸はそれらの GC 含量において様々であることが可能であり、短い断片長で多種多様な  $T_m$  を有しうる。無細胞病原体断片は比較的短い長さ ( 例えば、20 ~ 120 bp、そしてしばしば平均約 75 bp ) を有するため、より短い断片は、例えば NG

50



Sのための処理中に、より変性し易く、したがって、サンプル中に存在していたとしても配列決定も検出もされない可能性がある。低い $T_m$ の断片（例えば、低いGCゲノムからの短い断片）、特に、32～75bpの範囲の断片の回収を追跡するための方法を用いて、より高い比率の低い $T_m$ の断片が維持されるように、核酸処理を最適化することが可能である。

#### 【0362】

28個の二本鎖のそれぞれを、2個のオリゴヌクレオチドをアニーリングすることによって作製した。二本鎖DNAの濃度がQubit (Thermo Fisher)により測定された。それらの28個の等モル量を1つの混合物へと一緒にした。該混合物の8個の重複サンプルを標準的なライブラリー製造方法1（酵素加熱不活性化工程を含む）で処理し、別の8個を、改変されたライブラリー製造方法2（酵素加熱不活性化工程を欠く）で処理した。ライブラリー製造の後、それらの16個のサンプルを1回の配列決定の実施において組合せ、各サンプルに関する28個のスパイクのそれぞれの収率を計算し、各ライブラリーに関する100万当たりのリード数に対して正規化した。

10

#### 【0363】

該正規化リード数は、標準方法1（図17に示されている）と比較した場合の、改変方法2（図18に示されている）における低い $T_m$ の種の回収の増加を示しており、例えば、32bp長および20%GC含量、32bp長および30%GC含量、32bp長および40%GC含量、42bp長および20%GC含量、42bp長および30%GC含量、または52bp長および20%GC含量であるスパイクインに関して、それが認められる。GCパネルのこの粒度が無かったなら、そのような差は遥かに不明確であったであろう。例えば、50%GC含量においては、試験した4つの長さのうち、32bpの長さのみが回復レベルの差を示す。

20

#### 【0364】

実施例9：サイズに基づく枯渇を可能にする長い長さを有する担体合成核酸を使用する配列決定ライブラリーの製造

本実施例は、サンプルにおける標的核酸より長い合成DNA（例えば、PCRにより合成されたDNA）を使用して配列決定ライブラリーを製造するための典型的な方法を示す。合成DNAは、サンプルにおける標的核酸からの、サイズに基づく分離を可能にする長さを有しうる。合成DNAの一方または両方の末端は、連結に抵抗する修飾を有しうる。該修飾は、1以上の末端における1以上の内部脱塩基部位および/または反転（inverted）ヌクレオチドを含みうる。合成DNAは、配列決定ライブラリーに加えられた後、合成DNAの比較的長い長さを利用する、サイズに基づく枯渇法を用いて、任意の時点で、該ライブラリーから枯渇されうる。

30

#### 【0365】

無細胞DNAを含む血漿サンプルを対象から得る。合成DNAを、ライブラリー作製キットのための最小必要量のDNAの濃度で、ライブラリーDNA投入溶液（例えば、無細胞DNA抽出物）に添加する。末端修復工程の前、または末端修復工程の後かつアダプター連結工程の前に、合成DNAを血漿DNA抽出物に添加する。

#### 【0366】

ついで、DNA濃度感受性連結を、キット製造者の説明に従い行う。合成DNAはPCR増幅されない。むしろ、合成DNAをサイズ選択し、短い断片（例えば、110bp未満の断片）の富化中に配列決定ライブラリーから枯渇させる。また、合成DNAを、末端修復または連結に抵抗するように修飾した場合、またはそれが連結に抵抗するように末端修復後にそれを添加した場合には、それは両末端においてアダプターを欠き、したがって配列決定されない。

40

#### 【0367】

図19は配列決定ライブラリーの製造における工程を示す。サンプル（例えば、血漿）における無細胞DNA断片1901を工程1902において単離して、非常に低い濃度の無細胞DNA1903を得ることが可能である。工程1904において該断片を末端修復

50

に付すことが可能である。ついで、一方の末端において連結に抵抗する修飾を有する長い担体核酸1910を添加することが可能である。ついで、核酸をアダプター連結工程1905に付すことが可能であり、ここで、末端修復断片は、両端に連結されたアダプターを有するが、担体核酸は、一方の末端に連結されたアダプターのみを有する。増幅工程1906中に、連結断片は増幅されるが、担体核酸は増幅されない。ついで、サイズ選択工程1907を行うことが可能である。

#### 【0368】

合成DNAの一方の末端は、ライブラリー製造における反応に合成DNAが関与することを妨げる修飾または構造を含む。アダプターを合成DNAの3'末端に連結する場合、合成DNAを2つの制限酵素で二重消化して、2つの異なるオーバーハングを有する、または一方の末端にオーバーハングを有し他方の末端に平滑末端を有する500bpを超える分子を得る。次に、それぞれ相補的オーバーハングまたは平滑末端化ヘアピンを特異的に使用して、オーバーハングまたは平滑末端にヘアピンを連結する。アダプターが合成DNAの5'末端に連結することが予想される場合には、PCRプライマーのペア[そのうちの一方は不活性化5'末端(例えば、5'反転ジデオキシ-T、C3スペーサー、スペーサー18など)を有する]を使用して、合成DNAを合成する。

#### 【0369】

実施例10: 脱塩基部位および修飾を有する担体合成核酸を使用する配列決定ライブラリーの製造

cfDNA抽出工程中に担体核酸として機能し、ライブラリー製造中に最小ライブラリー投入量をもたらす担体合成核酸を設計した。担体合成核酸は中央脱塩基伸長を含有し、修飾を含む両端を有していた。担体合成核酸の配列を以下に示す(5Invddtは5'反転ddTを示し、3invdTは3'反転dTを示し、idSpは内部脱塩基部位を示す)。

#### 【0370】

5' - / 5 I n v d d T / G C G T C C C G G C G C G C G T T T A G G G A T A A C A  
/ i d S p / i d S p / i d S p / i d S p / G G G T A A T G G C G C A A G G G T G  
C T G G C / 3 I n v d T / - 3 ' ;  
3' - / 3 I n v d T / C G C A G G G C C G C G C G C A A A T C C C T A T T G T /  
i d S p / i d S p / i d S p / i d S p / C C C A T T A C C G C G T T C C C A C G  
A C C G / 5 I n v d d T / - 5 ' 。

#### 【0371】

該方法の工程を図20Aに示す。2つの並行実験を行った。2つの実験のうちの1つでエンドヌクレアーゼVII消化を行った。エンドヌクレアーゼVII消化を伴わない実験においては、該方法は末端修復(工程2001)、酢酸ナトリウムおよびエタノールでのMagBind精製(工程2002)、アダプター連結(工程2003)、アンブア(Ampure)精製(工程2004)およびライブラリー増幅(工程2005)を含む。エンドヌクレアーゼVII消化を伴う実験においては、該方法は末端修復(工程2006)、酢酸ナトリウムおよびエタノールでのMagBind精製(工程2007)、アダプター連結(工程2008)、アンブル(Ampure)精製(工程2009)、エンドヌクレアーゼVII消化(工程2010)およびライブラリー増幅(工程2011)を含む。

#### 【0372】

エンドヌクレアーゼVII消化を37で行い、1時間進行させた。アダプター連結後には担体合成核酸の枯渇は不要であった。なぜなら、脱塩基部位が鋳型の増幅を既に効率的に阻害したからである。また、該修飾はアダプター連結を妨げ、全ては、担体合成核酸が配列決定されることを妨げた。エンドヌクレアーゼVIIはライブラリーにおけるアダプター二量体を枯渇させるために使用されうる。

#### 【0373】

図20Bおよび20Cは、エンドヌクレアーゼVII消化を伴う又は伴わない配列決

10

20

30

40

50

定ライブラリーの作製を示す。担体合成核酸を含有するライブラリーにおけるアダプター連結後のエンドヌクレアーゼV I I Iによる消化は、担体合成核酸を含有しないライブラリーと比較して改善された再現性およびより高いスパイクインシグナルをもたらした。

【0374】

実施例11：脱塩基部位を有する合成核酸を使用する配列決定ライブラリーの製造

c f DNA抽出工程中に担体核酸として機能し、ライブラリー製造中に最小ライブラリー投入量をもたらす脱無塩基含有担体合成核酸を設計した。配列決定ライブラリーの製造方法は、実施例8で用いたものと実質的に同じであった。種々のタイプの脱塩基含有担体合成核酸を設計した。担体合成核酸分子の配列を以下に示す。

【0375】

連結に抵抗する修飾を含む二本鎖の一端を有する部分的に活性な脱塩基担体合成核酸(部分的 a b - C N A ) (連結用に二本鎖の一端を残したことは、末端修復およびアダプター連結反応に対するいずれかの濃度効果をもたらすことを助けた) ( 5 I n v d d t は 5 ' 反転 d d T を示し、 3 i n v d T は 3 ' 反転 d T を示し、 i d S p は内部脱塩基部位を示す) :

5 ' - G C G T C C C G G C G C G C G T T T A G G G A T A A C A / i d S p / i d S p / i d S p / i d S p / G G G T A A T G G C G C A A G G G T G C T G G C / 3 I n v d T / - 3 ' ;

3 ' - C G C A G G G C C G C G C G C A A A T C C C T A T T G T / i d S p / i d S p / i d S p / C C C A T T A C C G C G T T C C C A C G A C C G / 5 I n v d d T / - 5 ' .

【0376】

二本鎖の両端を含有する活性な脱塩基担体合成核酸(活性 a b - C N A ) は連結可能であった(連結用に両端を残したことは、濃度効果の効率的低減に末端が要求される場合に担体合成核酸投入量を減少させるのに有用であった)( i d S p は内部脱塩基部位を示す) :

5 ' - G C G T C C C G G C G C G C G T T T A G G G A T A A C A / i d S p / / i d S p / / i d S p / / i d S p / G G G T A A T G G C G C A A G G G T G C T G G C - 3 ' ;

3 ' - C G C A G G G C C G C G C G C A A A T C C C T A T T G T / i d S p / / i d S p / / i d S p / / i d S p / C C C A T T A C C G C G T T C C C A C G A C C G - 5 ' .

【0377】

鎖当たりだた1つの脱塩基部位を有する単一脱塩基担体合成核酸(単一 a b - C N A ) (単一脱塩基部位はエンドヌクレアーゼV I I Iで、より効率的に消化された)( i d S p は内部脱塩基部位を示す) :

5 ' - G C G T C C C G G C G C G C G T T T A G G G A T A A C A G T / i d S p / G G G T A A T G G C G C A A G G G T G C T G G C - 3 ' ;

3 ' - C G C A G G G C C G C G C G C A A A T C C C T A T T G T C A T C C C A T T / i d S p / C C G C G T T C C C A C G A C C G - 5 ' .

【0378】

全てのライブラリーを製造し、各変異を3回重複して施した。また、連結用の二本鎖末端の作製はスパイクイン分子(例えば、スパーク)のバンドを拡散させた。このことは、この実験条件下、多様性の減少が有意でありうることを示唆している。図21Aおよび21Bは配列決定ライブラリーの製造の結果を示す。エンドヌクレアーゼV I I I消化は該重複体の幾つかにおいてアダプター二量体バンドの消失を引き起こした。エンドヌクレアーゼV I I I消化物を用いた場合、より多数の非アダプター二量体鑄型が増幅に利用可能となった。

【0379】

実施例12：DNA-RNAハイブリッドを含有する合成核酸を使用する配列決定ライ

10

20

30

40

50

## ブラリーの製造

c f DNA抽出工程中に担体核酸として機能し、ライブラリー製造中に最小ライブラリー投入量をもたらす、DNA-RNAハイブリッドを含有する担体合成核酸(RnD-CNA)を設計した。図22に示されているとおり、配列決定可能な分子の最終プールにおけるRnD-CNA枯渇はアダプター連結後かつライブラリー増幅前のRNアーゼH消化により達成された。RNアーゼHに基づく枯渇をライブラリー増幅バッファー中、37で行い、1時間進行させた。

### 【0380】

該方法の工程を図23Aに示す。典型的な実験においては、該方法は末端修復(工程2301)、酢酸ナトリウムおよびエタノールでのMagBind精製(工程2302)、アダプター連結(工程2303)、アンプル(Ampure)精製(工程2304)、担体核酸枯渇(工程2305)およびライブラリー増幅(工程2306)を含む。これらの実験におけるRnD-CNAは、連結または増幅を妨げる末端を有さなかった。図23Bはライブラリーの製造の結果を示す。アダプター二量体バンドは、RnD-CNAがライブラリー投入物の一部として導入された場合には消失した。

10

### 【0381】

図23Aおよび23BはRNアーゼH消化後のRnD-CNAを示す。RNアーゼH消化により得られた断片は、両側をアダプターに連結した場合、2本のアダプターの全長が145bpであると仮定すると、175bpおよび166bpの断片を与えた。この連結は、最初に増幅バッファー中でポリメラーゼにより3'陥凹末端が埋められた場合に可能であった。これは、消化後の断片が共に3'オーバーハングを含有するように、そして増幅ポリメラーゼが3'エンドヌクレアーゼ活性を示さないように、RNA-CNA内のリボヌクレオチドの位置を設計することにより妨げられた。

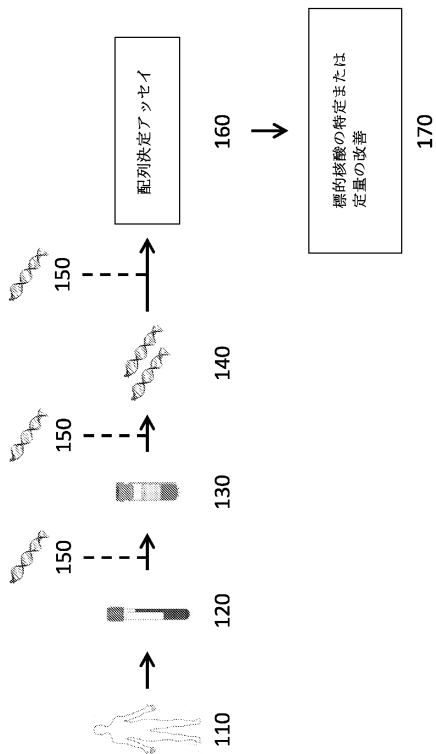
20

### 【0382】

本開示内容の好ましい実施形態が本明細書に示され、記載されているが、そのような実施形態は単なる例示として記載されていることが当業者に明らかであろう。本開示内容から逸脱することなく、多数の変形、変更および置換が当業者に今や見出されるであろう。本明細書に記載されている開示内容の実施形態の種々の代替物が本開示内容の実施において使用されうると理解されるべきである。以下の特許請求の範囲は本開示内容の範囲を定め、特許請求の範囲の範囲内の方法および構造ならびにそれらの均等物はそれに包含されると意図される。

30

【 図 1 】



【 図 3 】

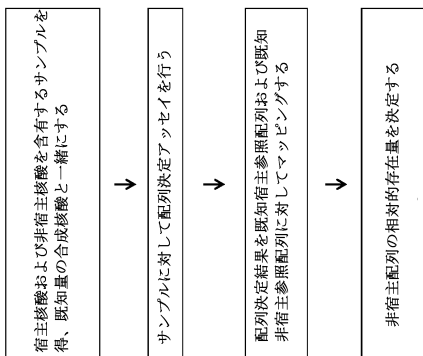


FIG. 3

【 図 2 】

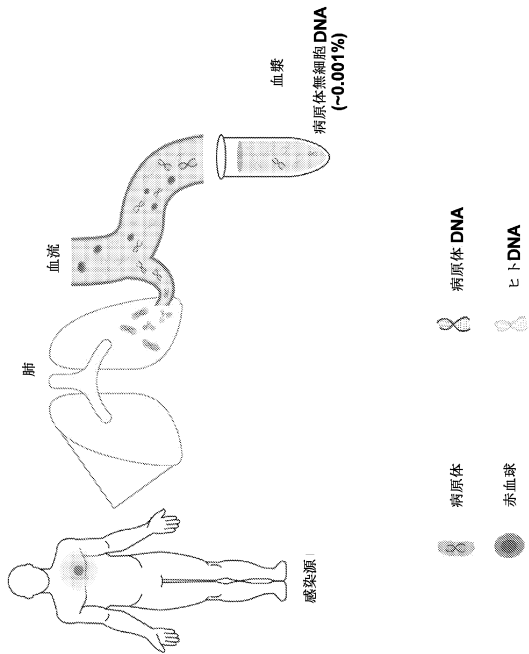


FIG. 1

FIG. 2

【 図 4 】

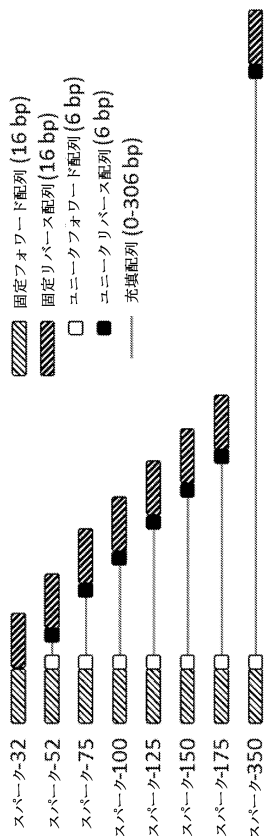


FIG. 4

【 図 5 】

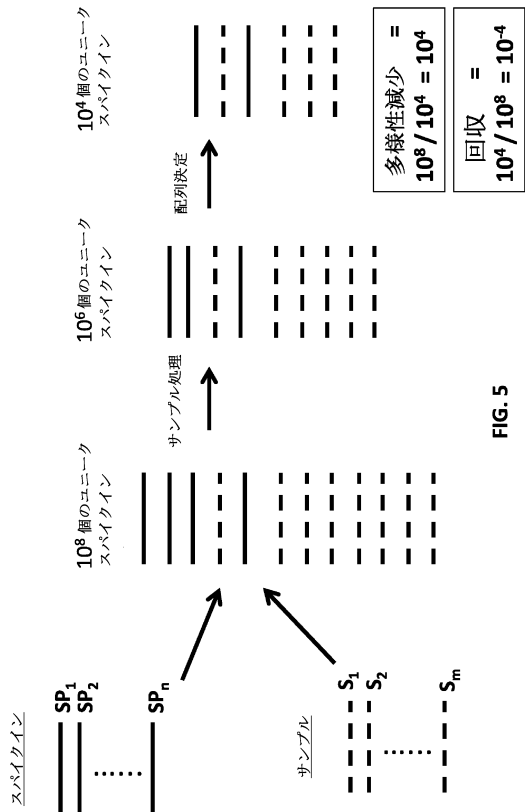


FIG. 5

【 図 6 】

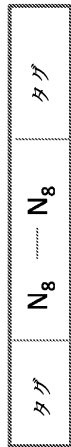


FIG. 6

【 図 7 】

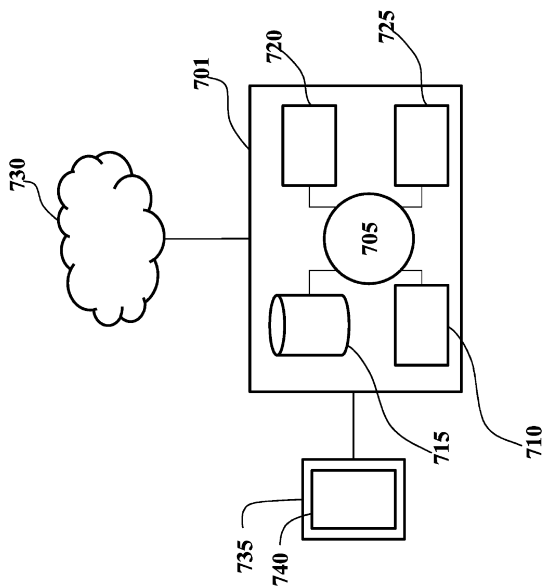


FIG. 7

【 図 8 】

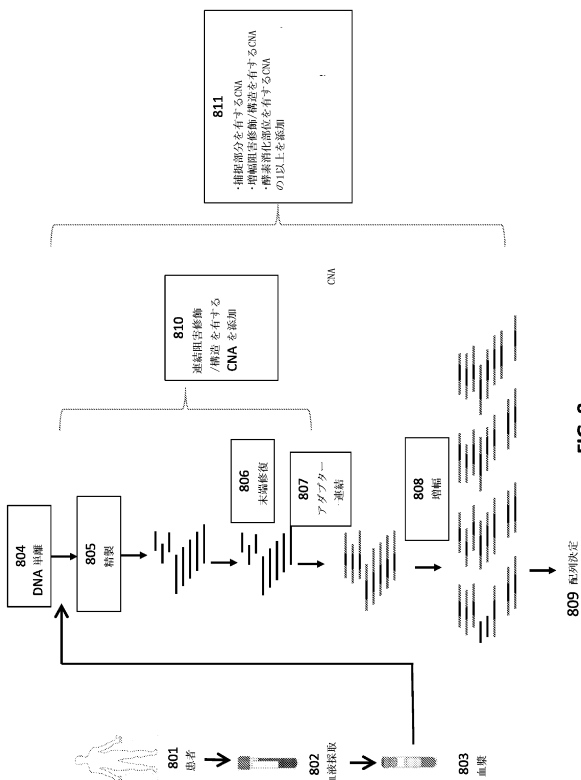


FIG. 8

【図 9】

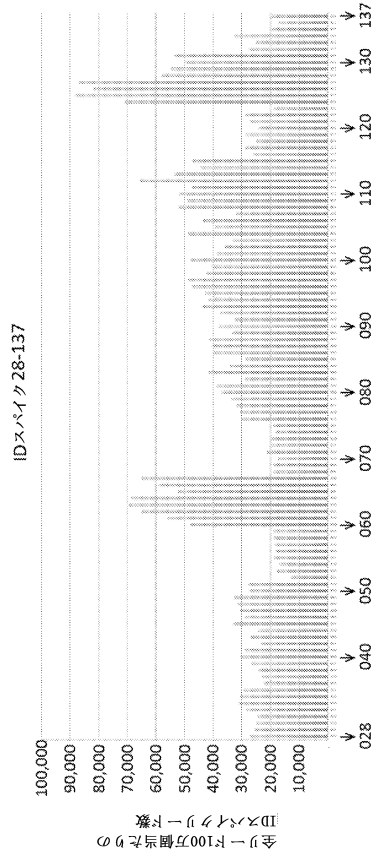


FIG. 9

【図 10】

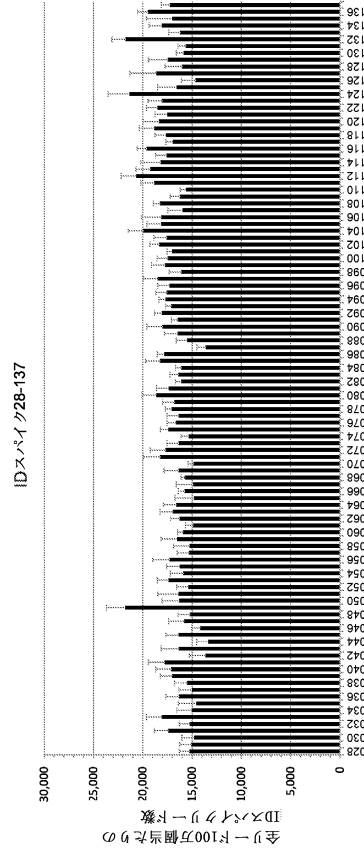


FIG. 10

【図 11】

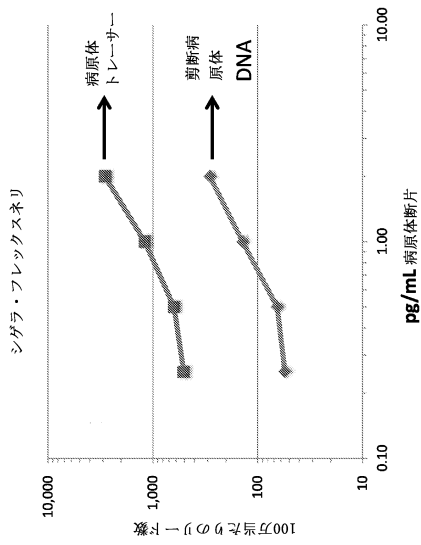


FIG. 11

【図 12】

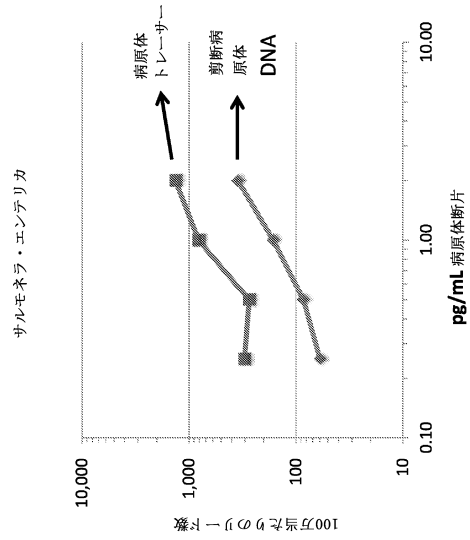


FIG. 12

【 図 1 3 】

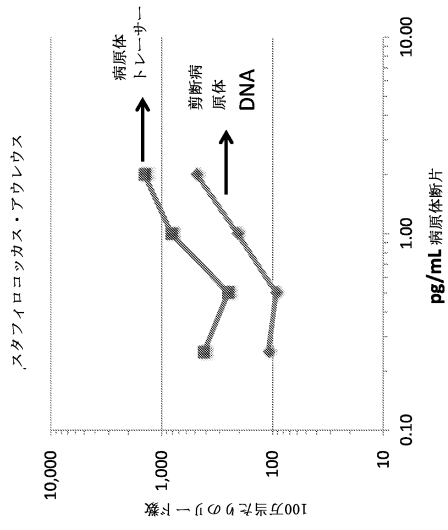


FIG. 13

【 図 1 4 】

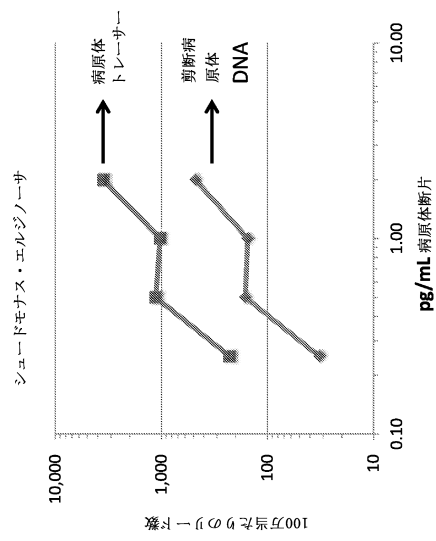


FIG. 14

【 図 1 5 】

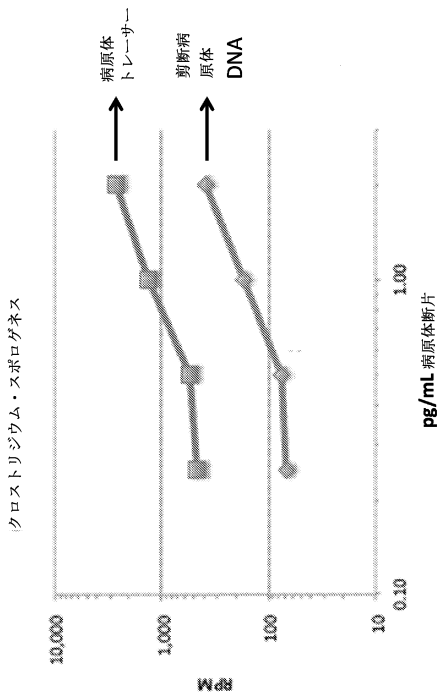


FIG. 15

【 図 1 6 】

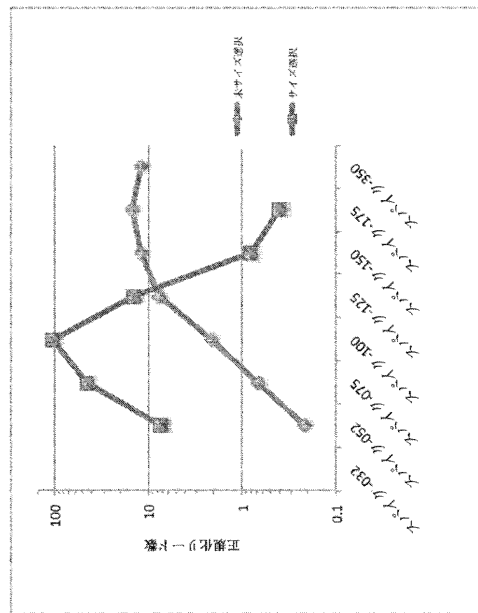


FIG. 16



【図 17】

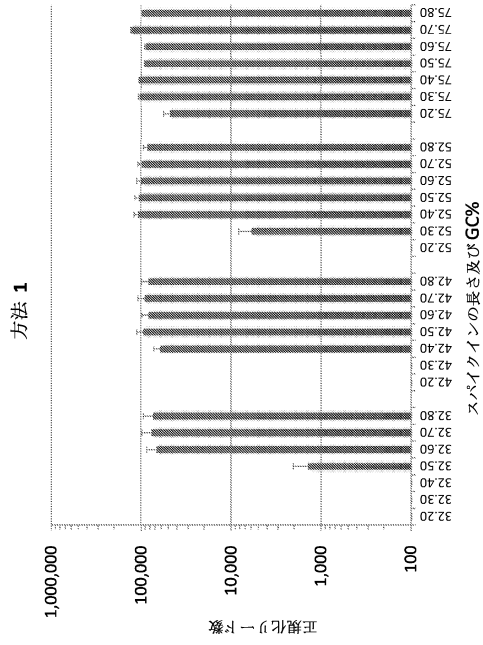


FIG. 17

【図 18】

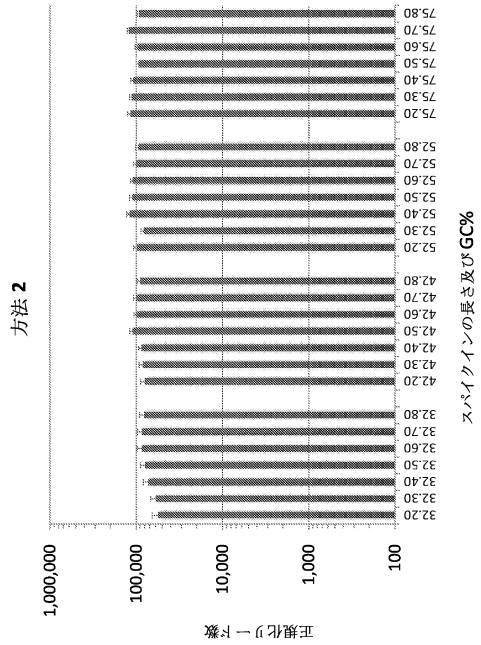


FIG. 18

【図 19】

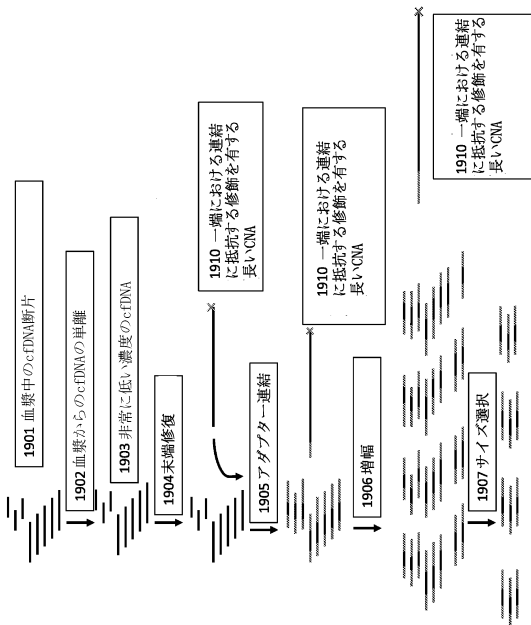


FIG. 19

【図 20 A】

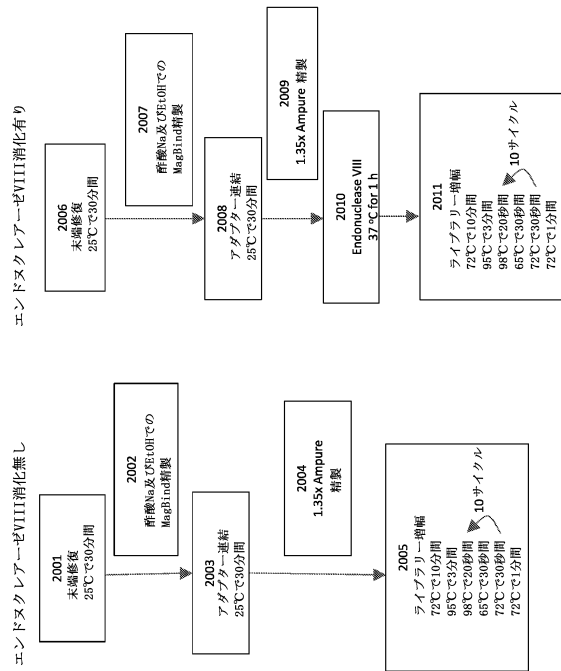


FIG. 20A

【 20B 】

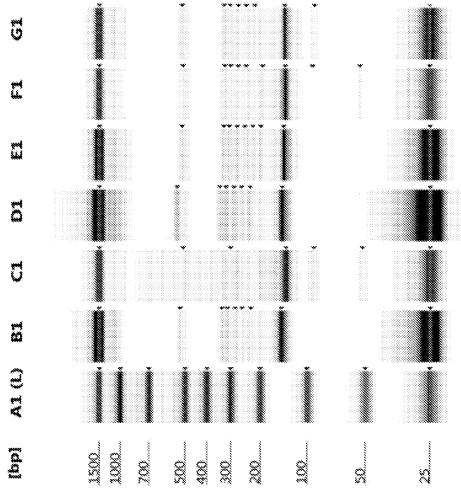


FIG. 20B

【 20C 】

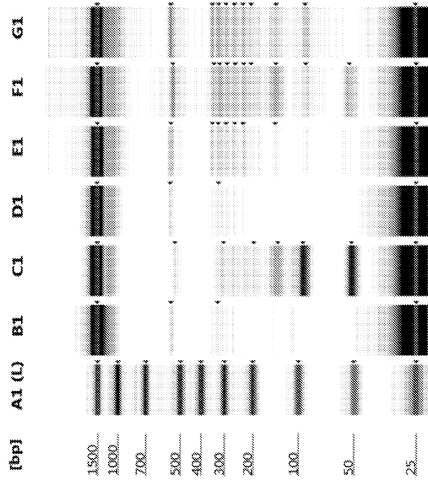


FIG. 20C

【 21A 】

ライブラリー製造における脱塩基CNA

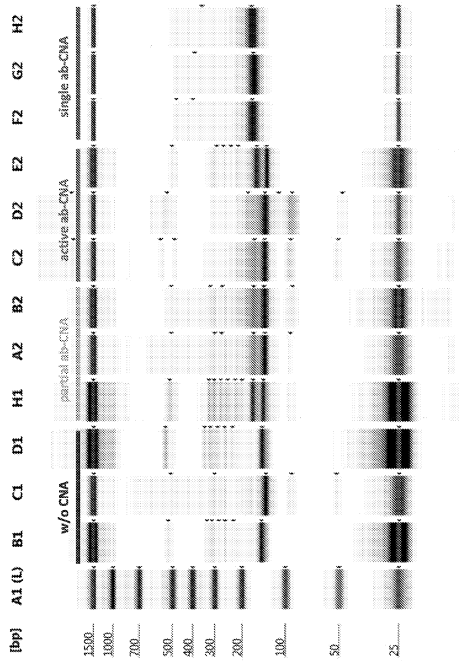


FIG. 21A

【 21B 】

アダプター連結後のエンドスクレアーゼVIII消化による脱塩基CNA

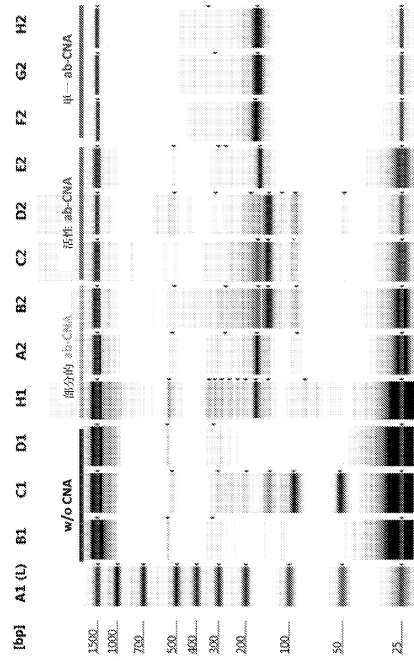


FIG. 21B

【 2 2 】

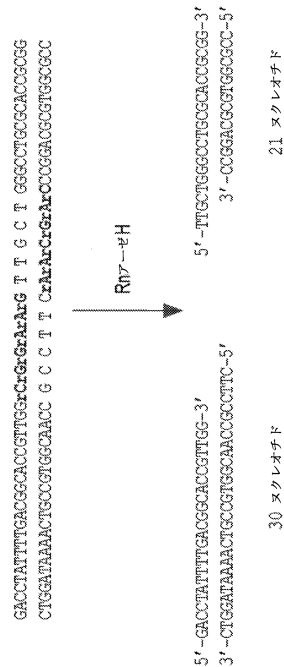


FIG. 22

【 2 3 A 】

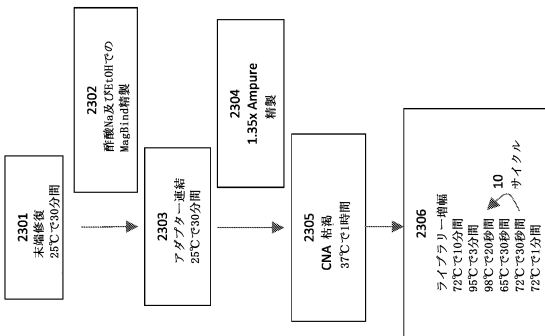


FIG. 23A

【 2 3 B 】

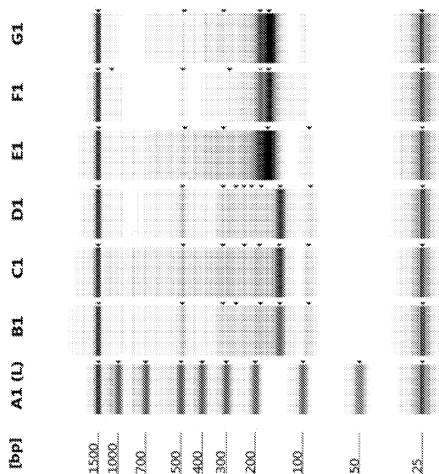


FIG. 23B

**【配列表】**

0006743268000001.app

## フロントページの続き

- (31)優先権主張番号 62/313,668  
(32)優先日 平成28年3月25日(2016.3.25)  
(33)優先権主張国・地域又は機関  
米国(US)

## 早期審査対象出願

- (74)代理人 100143823  
弁理士 市川 英彦
- (74)代理人 100151448  
弁理士 青木 孝博
- (74)代理人 100183519  
弁理士 櫻田 芳恵
- (74)代理人 100196483  
弁理士 川崎 洋祐
- (74)代理人 100203035  
弁理士 五味淵 琢也
- (74)代理人 100185959  
弁理士 今藤 敏和
- (74)代理人 100160749  
弁理士 飯野 陽一
- (74)代理人 100160255  
弁理士 市川 祐輔
- (74)代理人 100202267  
弁理士 森山 正浩
- (74)代理人 100146318  
弁理士 岩瀬 吉和
- (74)代理人 100127812  
弁理士 城山 康文
- (72)発明者 クリスチャンズ, フレッド・シー  
アメリカ合衆国、カリフォルニア・9 4 0 2 2、ロス・アルト・ヒルズ、アマースト・コート・1  
4 1 3 4
- (72)発明者 ヴィルフアン, イゴール・ディー  
アメリカ合衆国、カリフォルニア・9 4 3 0 3、イースト・パロ・アルト、ウッドランド・アベニ  
ュー・1 6 7 7
- (72)発明者 ケルテース, マイケル  
アメリカ合衆国、カリフォルニア・9 4 0 2 5、メンロー・パーク、トレントン・ウェイ・3 2 8
- (72)発明者 ブラウキャンブ, ティモシー・エー  
アメリカ合衆国、カリフォルニア・9 4 3 0 3、パロ・アルト、サザーランド・ドライブ・4 0 2  
0
- (72)発明者 ベンカタサブラーマニヤン, シヴクマール  
アメリカ合衆国、カリフォルニア・9 5 1 2 4、サン・ノゼ、カービー・ウェイ・1 9 5 8
- (72)発明者 ローゼン, マイケル  
アメリカ合衆国、カリフォルニア・9 4 0 3 6、パロ・アルト、カレッジ・アベニュー・3 2 2、  
アパートメント・イー
- (72)発明者 シット, レネ  
アメリカ合衆国、カリフォルニア・9 4 0 8 6、サニーベール、リード・アベニュー・1 0 8 5、  
ユニット・シー

審査官 北村 悠美子

(56)参考文献 特表2015-535431(JP,A)

(58)調査した分野(Int.Cl., DB名)

C12Q 1/00-3/00

C12N 15/00-15/90

C A p l u s / M E D L I N E / E M B A S E / B I O S I S ( S T N )

P u b M e d