

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6803327号  
(P6803327)

(45) 発行日 令和2年12月23日 (2020. 12. 23)

(24) 登録日 令和2年12月2日 (2020. 12. 2)

(51) Int. Cl.	F I
C 1 2 Q 1/6851 (2018. 01)	C 1 2 Q 1/6851 Z
C 1 2 Q 1/6855 (2018. 01)	C 1 2 Q 1/6855 Z
C 1 2 N 15/00 (2006. 01)	C 1 2 N 15/00

請求項の数 19 (全 57 頁)

(21) 出願番号	特願2017-506386 (P2017-506386)	(73) 特許権者	501430467
(86) (22) 出願日	平成27年8月6日 (2015. 8. 6)		ニューゲン テクノロジーズ, インコーポレイテッド
(65) 公表番号	特表2017-526353 (P2017-526353A)		アメリカ合衆国 カリフォルニア 940
(43) 公表日	平成29年9月14日 (2017. 9. 14)		70, サン カルロス, インダストリアル
(86) 国際出願番号	PCT/US2015/044065		ロード 201, スイート 310
(87) 国際公開番号	W02016/022833	(74) 代理人	100078282
(87) 国際公開日	平成28年2月11日 (2016. 2. 11)		弁理士 山本 秀策
審査請求日	平成30年7月17日 (2018. 7. 17)	(74) 代理人	100113413
(31) 優先権主張番号	62/034, 043		弁理士 森下 夏樹
(32) 優先日	平成26年8月6日 (2014. 8. 6)	(72) 発明者	スコルニック, ジョナサン
(33) 優先権主張国・地域又は機関	米国 (US)		アメリカ合衆国 カリフォルニア 941
			07, サンフランシスコ, カンザス
			ストリート 451, ナンバー518
			最終頁に続く

(54) 【発明の名称】 標的化されたシークエンシングからのデジタル測定値

(57) 【特許請求の範囲】

【請求項 1】

複数の特定核酸分子を定量化するための方法であって、

a . 各プローブが前記複数の特定核酸分子内の一特定核酸分子内のプローブ標的領域とハイブリダイズし、かつ各プローブがその 5' 末端において第 1 のアダプターを有する複数のプローブを伸長させて、複数の伸長産物を生成するステップと；

b . 第 2 のアダプターを、前記複数のプローブ伸長産物の二本鎖末端へと追加するステップと；

c . 前記複数のプローブ伸長産物をシークエンシングして、前記プローブ伸長産物の各々の配列データを生成するステップと；

d . プローブ標的領域とハイブリダイズした各プローブの数を決定するステップであって、前記数が、前記プローブ標的領域を含む、前記特定核酸分子の各々の量を指し示す、ステップと

を含む、方法。

【請求項 2】

組成物中の複数の特定核酸分子を定量化するための方法であって、

a . 複数のプローブを、特定核酸分子内のプローブ標的領域とハイブリダイズさせるステップであって、各プローブが、その 5' 末端に、第 1 のアダプターを有する、ステップと；

b . 各プローブを伸長させて、前記第 1 のアダプター配列を含む複数のプローブ伸長産

10

20

物を生成するステップと；

c．第2のアダプター配列を、前記複数のプローブ伸長産物の二本鎖末端へと追加するステップと；

d．前記複数のプローブ伸長産物をシーケンシングして、前記複数のプローブ伸長産物の各々の配列を生成するステップと；

e．前記複数のプローブ伸長産物の各々の前記配列を、プローブデータベース内の所定の配列に対してアラインさせるステップであって、前記プローブデータベースが、複数の所定の配列を含み、各所定の配列が、プローブに特異的である、ステップと；

f．各プローブ伸長産物の前記配列についての、前記シーケンシングデータベース内の所定の配列に対するアラインメントの数を決定するステップであって、前記アラインメントの数が、前記プローブがハイブリダイズする前記特定核酸分子の各々の量を指し示す、ステップと

を含む、方法。

【請求項3】

前記複数のプローブ伸長産物の前記配列が、フォワードリード配列、インデックスリード配列、およびリバーシブルリード配列のうちの少なくとも1つを含む、請求項1または2に記載の方法であって、

前記フォワードリード配列は、フォワードリードプライミング部位において開始される複数のプローブ伸長産物の一配列であって、フォワードプライマーを用いて生成され；

前記インデックスリード配列は、インデックスプライミング部位において開始される複数のプローブ伸長産物の一配列であって、インデックスプライマーを用いて生成され；

前記リバーシブルリード配列は、リバーシブルプライミング部位において開始される複数のプローブ伸長産物の一配列であって、リバーシブルプライマーを用いて生成される、方法。

【請求項4】

前記複数のプローブ伸長産物の前記配列を、ゲノムまたはトランスクリプトームの座標へとマッピングして、意図されるプローブのアニーリングおよび伸長を検証する、請求項3に記載の方法。

【請求項5】

前記フォワードリードおよび前記リバーシブルリードの前記複数のプローブ伸長産物の前記配列を、前記複数の特定核酸についてマッピングする、請求項4に記載の方法。

【請求項6】

前記第2のアダプターが、試料バーコードオリゴヌクレオチドとn-ランダムオリゴヌクレオチドのうち少なくとも1つを含み、前記複数のプローブ伸長産物の前記配列が、前記試料バーコードオリゴヌクレオチドと前記n-ランダムオリゴヌクレオチドのうちの少なくとも1つからの配列を含む、請求項3に記載の方法。

【請求項7】

前記アダプターの追加に続き、前記プローブ伸長産物を、増幅する、請求項1または2に記載の方法。

【請求項8】

前記第2のアダプターを付加する前に、前記プローブ伸長産物を、制限エンドヌクレアーゼで処理するか、またはこれに末端修復を施す、請求項1または2に記載の方法。

【請求項9】

前記末端修復が、平滑末端修復である、請求項8に記載の方法。

【請求項10】

前記制限エンドヌクレアーゼで処理されたプローブ伸長産物により、共通の末端を伴うフォワードリードをもたらす、請求項9に記載の方法。

【請求項11】

前記複数のプローブ伸長産物の前記配列を、ゲノムまたはトランスクリプトームの座標へとマッピングして、意図されるプローブのアニーリングおよび伸長を検証する、請求項10に記載の方法。

10

20

30

40

50

**【請求項 1 2】**

前記複数のプローブ伸長産物の前記配列を、プローブデータベースの基準コピーに対してアラインさせて、意図されるプローブのアニーリングを検証する、請求項 1 1 に記載の方法。

**【請求項 1 3】**

前記配列が、出発試料中に存在する特定核酸分子のコピー数の尺度として計数される、請求項 1 1 に記載の方法。

**【請求項 1 4】**

前記フォワードリード配列が、出発試料中に存在する特定核酸分子のコピー数の尺度として計数される、請求項 1 2 に記載の方法。

10

**【請求項 1 5】**

前記第 2 のアダプター配列が、インデックスプライミング部位、インデックスヌクレオチド配列、n-ランダムヌクレオチド配列、フォワードリードプライミング部位、リバースリードプライミング部位、ならびにこれらの組み合わせのうちの少なくとも 1 つを含む、請求項 1 から 2 および 9 のいずれか一項に記載の方法。

**【請求項 1 6】**

前記第 1 のアダプター配列が、フォワードリードプライミング部位、リバースリードプライミング部位、およびリンカー配列、ならびにこれらの組み合わせのうちの少なくとも 1 つを含む、請求項 1 から 2 および 9 のいずれか一項に記載の方法。

**【請求項 1 7】**

前記 5' の第 1 のアダプターが、各プローブ伸長産物に共通である、請求項 1 から 2 および 9 のいずれか一項に記載の方法。

20

**【請求項 1 8】**

前記インデックスヌクレオチド配列が、バーコード配列をさらに含む、請求項 1 5 のいずれか一項に記載の方法。

**【請求項 1 9】**

前記複数の核酸が、組織、臓器、単一の細胞、腫瘍、患者から採取された有機流体の検体、遊離循環核酸、真菌、原核生物、およびウイルスからなる群より選択される試料に由来する、請求項 1 または 2 に記載の方法。

**【発明の詳細な説明】**

30

**【技術分野】****【0001】**

発明の分野

本教示は、遺伝子発現およびコピー数変異についてのデジタル測定値をもたらす標的化核酸シーケンシングの使用に関する。

**【0002】**

関連出願への相互参照

この出願は、2014年8月6日に提出された米国仮出願第62/034,043号（これは、全ての目的のためにその全体が参考として本明細書に援用される）の利益を主張する。

40

**【背景技術】****【0003】**

特定核酸についてのデジタル計数をもたらす分子的方法は、研究および臨床のコミュニティにとって関心の対象である。これらの方法を使用して、遺伝子発現（デジタル遺伝子発現もしくはDGE）またはコピー数変異（CNV）を直接測定することができる。デジタルリードアウトにより得ることができる、高精度の測定値は、マイクロアレイ技術と比較して、データに高度な信頼性をもたらし、研究者が、試料の間のわずかな差を同定したり、同様に、腫瘍生検材料中など、細胞のサブセット内の差を同定したりすることを可能とするほか、細胞間の変動を決定することも可能とする。

**【発明の概要】**

50

## 【発明が解決しようとする課題】

## 【0004】

しかし、特化した装置を伴わずに、目的のトランスクリプトーム領域およびゲノム領域についてのハイスループットの解析を可能とする、選択的な標的の定量化のための異なる方法が、依然として必要とされている。本明細書で開示される方法、組成物、およびキットは、これらの必要を満たし、関連する利点をもたらす。

## 【課題を解決するための手段】

## 【0005】

一態様では、複数の核酸内の、複数の特定核酸を定量化するための方法であって、a. 複数のプローブ伸長産物のシーケンシングライブラリーを生成するステップであって、特定核酸配列内のプローブ標的領域と相補的であり、これとハイブリダイズするプローブを伸長させることにより、各プローブ伸長産物を導出するステップと；b. 複数のプローブ伸長産物を含むライブラリーをシーケンシングして、複数のプローブ伸長産物の配列データを生成するステップと；c. アラインさせた配列の各々を、計数するステップであって、アラインメントの数が、複数の核酸内の、対応する特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

10

## 【0006】

一態様では、組成物中の複数の特定核酸分子を定量化するための方法であって、a. 複数のプローブ伸長産物を生成するステップであって、各プローブ伸長産物が、特定核酸分子内のプローブ標的領域と相補的なプローブ配列を含む、ステップと；b. 複数のプローブ伸長産物をシーケンシングして、複数のプローブ伸長産物の各々の配列を生成するステップと；c. 複数のプローブ伸長産物の各々の配列を、基準配列データベースに対してアラインさせるステップであって、基準配列データベースが、プローブ配列を含む、ステップと；d. 基準配列データベース内の配列との、各プローブ伸長産物の配列についてのアラインメントの数を決定するステップであって、アラインメントの数が、プローブ伸長産物のプローブが相補的である特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

20

## 【0007】

一態様では、複数の核酸内の、複数の特定核酸を定量化するための方法であって、a. 複数のプローブ伸長産物のシーケンシングライブラリーを生成するステップであって、各プローブ伸長産物が、各プローブ伸長産物の5'末端に付着した、第1のアダプターを含み、特定核酸配列内のプローブ標的領域と相補的であり、これとハイブリダイズするプローブを伸長させることにより、各プローブ伸長産物を導出するステップと；b. ライブラリーをシーケンシングして、複数のプローブ伸長産物の配列データを生成するステップと；c. 配列データ内のプローブ配列の存在を同定し、複数のプローブ伸長産物内の各プローブ配列を計数するステップであって、計数されたプローブの数が、複数の核酸内の、複数の特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

30

## 【0008】

一態様では、複数の特定核酸分子を定量化するための方法であって、a. 複数のプローブ伸長産物を生成するステップであって、各プローブ伸長産物が、(i)第1のアダプターと、(ii)特定核酸分子内のプローブ標的領域と相補的なプローブ配列とを含む、ステップと；b. 複数のプローブ伸長産物をシーケンシングして、複数のプローブ伸長産物の各々についての配列を含む配列データを生成するステップと；c. 配列データ内の、各プローブ伸長産物のプローブ配列の存在を同定するステップと；d. 複数のプローブ伸長産物内の、プローブ配列の各々の数を決定するステップであって、プローブ配列の各々の数が、プローブ配列の各々が相補的である複数の特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

40

## 【0009】

一態様では、複数の核酸内の、複数の特定核酸を定量化するための方法であって、a. 複数の核酸の5'末端に、第1のアダプター配列を追加するステップと；b. 複数のプロ

50

ープをハイブリダイズさせるステップであって、各プローブが、複数の特定核酸内の一特定核酸内のプローブ標的領域と相補的である、ステップと； c . 各プローブを、追加された第 1 のアダプター配列へと伸長させて、第 1 のアダプター配列および第 2 のアダプター配列を有する、複数のプローブ伸長産物を生成するステップと； d . 複数のプローブ伸長産物を含むシーケンシングライブラリーを生成するステップと； e . ライブラリーをシーケンシングするステップであって、複数のプローブ伸長産物の各々についての配列データを得るステップと； f . 複数のプローブ伸長産物の各々についての配列データを、プローブデータベースの基準コピー内の所定の配列に対してアラインさせるステップであって、前記所定の配列が、各プローブに特異的である、ステップと； g . その所定の配列に対してアラインさせた各プローブ配列を計数するステップであって、その特定核酸に特異的な各プローブについての計数の数が、複数の核酸内の、複数の特定核酸内の、特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

10

**【 0 0 1 0 】**

一態様では、複数の核酸分子内の、複数の特定核酸分子を定量化するための方法であって、 a . 複数の核酸分子の各々の 5 ' 末端へと、第 1 のアダプター配列を追加するステップと； b . 複数のプローブを、複数の特定核酸分子とハイブリダイズさせるステップであって、各プローブが、特定核酸分子内のプローブ標的領域と相補的である、ステップと； c . 各プローブを、追加された第 1 のアダプター配列へと伸長させて、第 1 のアダプター配列および第 2 のアダプター配列を有する複数のプローブ伸長産物を生成して、複数のプローブ伸長産物を産生するステップと； d . 複数のプローブ伸長産物をシーケンシングして、複数のプローブ伸長産物の各々についての配列データを生成するステップと； e . 複数のプローブ伸長産物の各々の配列を、プローブデータベースの基準コピー内の所定の配列に対してアラインさせるステップであって、前記所定の配列が、各プローブに特異的である、ステップと； f . その所定の配列に対してアラインさせた各プローブ配列の数を決定するステップであって、数が、プローブが相補的である特定核酸分子の量を指し示すステップとを含む方法が開示される。

20

**【 0 0 1 1 】**

一態様では、複数の核酸内の、複数の特定核酸を定量化するための方法であって、 a . 複数のハイブリダイズさせたプローブを伸長させるステップであって、各プローブが複数の特定核酸内の一特定核酸内のプローブ標的領域と相補的であり、各プローブが 5 ' の第 1 のアダプターを有する、ステップと； b . 第 2 のアダプター配列を、複数のプローブ伸長産物の二本鎖末端へと追加して、シーケンシングライブラリーを生成するステップと； c . ライブラリーをシーケンシングするステップであって、複数のプローブ伸長産物の各々についての配列データを得ることができるステップと； d . 各プローブ標的領域に対応する各プローブ配列を計数するステップであって、その特定核酸に特異的な各プローブについての計数の数が、複数の核酸内の、複数の特定核酸内の、特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

30

**【 0 0 1 2 】**

一態様では、複数の特定核酸分子を定量化するための方法であって、 a . 各プローブが複数の特定核酸分子内の一特定核酸分子内のプローブ標的領域とハイブリダイズし、各プローブがその 5 ' 末端において第 1 のアダプターを有する複数のプローブを伸長させて、複数の伸長産物を生成するステップと； b . 第 2 のアダプターを、複数のプローブ伸長産物の二本鎖末端へと追加するステップと； c . 複数のプローブ伸長産物をシーケンシングして、プローブ伸長産物の各々の配列データを生成するステップと； d . プローブ標的領域とハイブリダイズした各プローブの数を決定するステップであって、数が、プローブ標的領域を含む、特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

40

**【 0 0 1 3 】**

一態様では、複数の核酸内の、複数の特定核酸を定量化するための方法であって、 a . 複数のプローブをハイブリダイズさせるステップであって、各プローブが、複数の特定核

50

酸内の一特定核酸内のプローブ標的領域と相補的であり、各プローブが、5'の第1のアダプターを有する、ステップと；b．各プローブを伸長させて、第1のアダプター配列を有する、複数のプローブ伸長産物を生成するステップと；c．第2のアダプター配列を、複数のプローブ伸長産物の二本鎖末端へと追加するステップと；d．複数のプローブ伸長産物を含むシーケンシングライブラリーを生成するステップと；e．ライブラリーをシーケンシングするステップであって、複数のプローブ伸長産物の各々についての配列データを得ることができるステップと；f．複数のプローブ伸長産物の各々についての配列データを、プローブデータベース内の所定の配列に対してアラインさせるステップであって、前記所定の配列が、各プローブに特異的である、ステップと；g．プローブ標的領域に対してアラインさせた各プローブ配列を計数するステップであって、その特定核酸に特異的な各プローブについての計数の数が、複数の核酸内の、複数の特定核酸内の、特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

10

## 【0014】

一態様では、組成物中の複数の特定核酸分子を定量化するための方法であって、a．複数のプローブを、特定核酸分子内のプローブ標的領域とハイブリダイズさせるステップであって、各プローブが、その5'末端に、第1のアダプターを有する、ステップと；b．各プローブを伸長させて、第1のアダプター配列を含む複数のプローブ伸長産物を生成するステップと；c．第2のアダプター配列を、複数のプローブ伸長産物の二本鎖末端へと追加するステップと；d．複数のプローブ伸長産物をシーケンシングして、複数のプローブ伸長産物の各々の配列を生成するステップと；e．複数のプローブ伸長産物の各々の配列を、プローブデータベース内の所定の配列に対してアラインさせるステップであって、前記プローブデータベースが、複数の所定の配列を含み、各所定の配列が、プローブに特異的である、ステップと；f．各プローブ伸長産物の配列について、シーケンシングデータベース内の所定の配列に対してアラインメントの数を決定するステップであって、アラインメントの数が、プローブがハイブリダイズする、特定核酸分子の各々の量を指し示すステップとを含む方法が開示される。

20

## 【0015】

一部の実施形態では、配列データ、またはシーケンシングされた複数のプローブ伸長産物は、フォワードリード、インデックスリード、およびリバースリードのうちの少なくとも1つを含む。一部の実施形態では、リバースリードは、プローブ標的領域を含む。一部の実施形態では、各プローブがそのそれぞれの特定核酸内のそのそれぞれのプローブ標的領域配列へとアニールした特異性を検証することができる。一部の実施形態では、配列データ、もしくはシーケンシングされた複数のプローブ伸長産物を、ゲノムデータベースもしくはトランスクリプトームデータベースの座標へとマッピングすることができ、かつ/または配列データ、もしくはシーケンシングされた複数のプローブ伸長産物を、プローブデータベースの基準コピーに対してアラインさせて、意図されるプローブのアニーリングおよび伸長を検証することができる。一部の実施形態では、配列データ、またはシーケンシングされた複数のプローブ伸長産物を、ゲノムデータベースまたはトランスクリプトームデータベースの座標へとマッピングすることができる。一部の実施形態では、リバースリードまたはフォワードリードは、プローブ標的領域を含む。一部の実施形態では、フォワードリードおよびリバースリードの配列データ、またはシーケンシングされた複数のプローブ伸長産物を、複数の特定核酸についてマッピングすることができ、インデックスリードの配列データ、またはシーケンシングされた複数のプローブ伸長産物により、バーコード配列およびn-ランダム配列のうちの少なくとも1つを同定することができる。一部の実施形態では、フォワードリードのマッピング座標と、インデックスリードのn-ランダム塩基との組み合わせにより、各プローブ伸長産物についてのPCR重複を決定し、同じフォワードリード座標および同じn-ランダム塩基配列を有する配列を、重複として同定し、単一の特定核酸分子として一括し、計数することができ、この場合、フォワードリード座標は同じであるが、n-ランダム塩基配列は異なる配列の各々を、別々の特定核酸分子として計数することができる。

30

40

50

## 【0016】

一部の実施形態では、フォワードリードと、対応するリバースリードとを、ペアエンドでアラインさせることができる。一部の実施形態では、重複の一括に続き、各プローブ配列について計数されるリバースリードまたはフォワードリードの数により、複数の特定核酸内の、各出発特定核酸分子についての分子の数を表す値を生成する。一部の実施形態では、ゲノムは、哺乳動物、細菌、ウイルス、リケッチア、または植物のゲノムまたはトランスクリプトームからなる群より選択される。一部の実施形態では、第1のアダプターを追加する前に、複数の特定核酸に末端修復を施す。一部の実施形態では、末端修復は、平滑末端修復である。一部の実施形態では、プローブは、DNAポリメラーゼ、RNAポリメラーゼ、または逆転写酵素からなる群より選択されるポリメラーゼにより伸長させることができる。

10

## 【0017】

一部の実施形態では、シーケンシングライブラリーを生成する前に、複数のプローブ伸長産物を増幅しうるか、または任意選択でこれを増幅する。一部の実施形態では、第2のアダプターを付加する前に、プローブ伸長産物を、制限エンドヌクレアーゼで処理するか、またはこれに平滑末端修復/末端修復を施すことができる。一部の実施形態では、プローブ伸長産物の伸長は、第1のアダプターの付加をさらに含む。一部の実施形態では、プローブ伸長産物の増幅は、増幅産物の各末端へのフローセル配列の付着をさらに含む。一部の実施形態では、制限エンドヌクレアーゼで処理されたプローブ伸長産物により、共通の末端を伴うフォワードリードをもたらす。一部の実施形態では、配列データ、またはシーケンシングされた複数のプローブ伸長産物を、ゲノムまたはトランスクリプトームの座標へとマッピングして、意図されるプローブのアニールリングおよび伸長を検証することができる。一部の実施形態では、配列データ、またはシーケンシングされた複数のプローブ伸長産物を、プローブデータベースの基準コピーに対してアラインさせて、意図されるプローブのアニールリングを検証することができる。一部の実施形態では、リバースリード配列またはフォワードリード配列を、それらが、どのプローブ配列を表すのかに従いビニングおよび計数することができ、この場合、各プローブが表される回数は、出発特定核酸分子が、元の試料中に存在する回数についての尺度でありうる。一部の実施形態では、フォワードリードは、特定核酸配列のうちの、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、または少なくとも25塩基を含みうる、特定核酸配列の少なくとも一部を含む。

20

30

## 【0018】

一部の実施形態では、第1のアダプター配列または第2のアダプター配列は、インデックス配列のプライミング部位、インデックスヌクレオチド配列、n-ランダムヌクレオチド配列、フォワードリードプライミング部位、およびリバースリードプライミング部位、ならびにこれらの組み合わせのうちの少なくとも1つを含む。一部の実施形態では、第2のアダプター配列または第1のアダプター配列は、フォワードリードプライミング部位、リバースリードプライミング部位、およびリンカー配列、ならびにこれらの組み合わせのうちの少なくとも1つを含む。一部の実施形態では、5'の第1のアダプターは、各プローブ伸長産物に共通でありうる。一部の実施形態では、5'テール配列は、第2のアダプター配列を含みうる。一部の実施形態では、プローブ伸長産物の増幅は、増幅産物の各末端へのフローセル配列の付着をもたらす。

40

## 【0019】

一部の実施形態では、インデックスリードは、インデックスヌクレオチド配列およびn-ランダム塩基配列のうちの、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、または少なくとも15塩基を含む。一部の実施形態では、インデックスリードは、n-ランダム塩基およびインデックスヌクレオチド配列のうちの、少なくと

50

も4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、または少なくとも10塩基を含む。一部の実施形態では、インデックスリードは、n-ランダム塩基および任意選択でインデックスヌクレオチド配列のうちの、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、または少なくとも10塩基を含む。一部の実施形態では、n-ランダム塩基のヌクレオチド配列は、少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、または少なくとも10ヌクレオチドを含む。一部の実施形態では、インデックスヌクレオチド配列は、バーコード配列をさらに含む。

#### 【0020】

一部の実施形態では、リバースリードは、プローブ配列および特定核酸配列の一部、ならびにこれらの組み合わせのうちの少なくとも1つを含む。一部の実施形態では、リバースリードは、プローブ配列のうちの、少なくとも15、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50、少なくとも55、または少なくとも60塩基を含む。一部の実施形態では、リバースリードは、プローブ配列に対して3'の特定核酸配列のうちの、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、または少なくとも20塩基を含む。

#### 【0021】

さらなる態様では、開示される方法により生成および/または増幅されたプローブ伸長産物の組成物が開示される。

#### 【0022】

なおさらなる態様では、複数の核酸は、組織、臓器、単一の細胞、腫瘍、患者から採取された有機流体 (organic fluid) の検体、遊離循環核酸、真菌、原核生物、およびウイルスからなる群より選択される試料に由来する。一部の実施形態では、患者は、腫瘍を有することが知られている場合もあり、これを有することが疑われる場合もある。一部の実施形態では、有機流体は、少なくとも1つの循環腫瘍細胞 (CTC) または播種性腫瘍細胞 (CTD) を含有する。一部の実施形態では、患者は、伝染性感染または伝染性疾患でありうる、ウイルス感染を有することが知られている場合もあり、これを有することが疑われる場合もある。

#### 【0023】

一部の実施形態では、本開示の組成物は、複数の核酸分子を含む。一部の実施形態では、各プローブ伸長産物は、特定核酸分子内のプローブ標的領域と相補的なプローブの伸長産物である。

#### 【0024】

なおさらなる態様では、オリゴヌクレオチドアダプター；プローブ標的領域配列の一部と相補的なプローブ；前記アダプター配列と相補的なプライマー；プローブ配列の一部と相補的なプライマー；リガーゼ；ポリメラーゼ；およびキットの使用のための指示のうちの少なくとも1または複数を含む核酸分子のデジタル測定のためのキットが開示される。なおさらなる態様では、核酸分子のデジタル測定のためのキットであって、本開示の1種または複数の態様を含むキットが開示される。

#### 【0025】

一部の実施形態では、本開示の方法、組成物、およびキットは、Liら、2012年、Bioinformatics、28巻(10号)：1307～1313頁；Bellowsら、2014年、Nucleic Acids Res.、42巻(20号)：e158頁；Jiangら、2015年、Nucleic Acids Res.、43巻(6号)：e39頁；Xiら、2011年、Proc. Natl. Acad. Sci.、108巻(46号)：1128～1136頁；FromerおよびPurcell、2014年、Curr. Protoc. Hum. Genet.、81巻：7.21.1～7.23.21節；Sathirapongsasutiら、2011年、Bioi

10

20

30

40

50



n f o r m a t i c s、31巻(15号):1~8頁;K r u m m ら、2012年、G e n o m e R e s .、22巻(8号):1525~1532頁;P l a g n o l ら、2012年、B i o i n f o r m a t i c s、28巻(21号):2747~2754頁において開示されている、1種または複数の態様を含む。

本発明は、例えば、以下の項目を提供する。

(項目1)

組成物中の複数の特定核酸分子を定量化するための方法であって、

a. 複数のプローブ伸長産物を生成するステップであって、各プローブ伸長産物が、特定核酸分子内のプローブ標的領域と相補的なプローブ配列を含む、ステップと;

b. 前記複数のプローブ伸長産物をシーケンシングして、前記複数のプローブ伸長産物の各々の配列を生成するステップと;

c. 前記複数のプローブ伸長産物の各々の前記配列を、基準配列データベースに対してアラインさせるステップであって、前記基準配列データベースが、プローブ配列を含む、ステップと;

d. 前記基準配列データベース内の配列との、各プローブ伸長産物の前記配列についてのアラインメントの数を決定するステップであって、前記アラインメントの数が、前記プローブ伸長産物の前記プローブが相補的である前記特定核酸分子の各々の量を指し示す、ステップと

を含む、方法。

(項目2)

複数の特定核酸分子を定量化するための方法であって、

a. 複数のプローブ伸長産物を生成するステップであって、各プローブ伸長産物が、(i)第1のアダプターと、(ii)特定核酸分子内のプローブ標的領域と相補的なプローブ配列とを含む、ステップと;

b. 前記複数のプローブ伸長産物をシーケンシングして、前記複数のプローブ伸長産物の各々の配列を含む配列データを生成するステップと;

c. 前記配列データ内の、各プローブ伸長産物の前記プローブ配列の存在を同定するステップと;

d. 前記複数のプローブ伸長産物内の、前記プローブ配列の各々の数を決定するステップであって、前記プローブ配列の各々の数が、前記プローブ配列の各々が相補的である前記複数の特定核酸分子の各々の量を指し示す、ステップと

を含む、方法。

(項目3)

複数の核酸分子内の、複数の特定核酸分子を定量化するための方法であって、

a. 複数の核酸分子の各々の5'末端へと、第1のアダプター配列を追加するステップと;

b. 複数のプローブを、前記複数の特定核酸分子とハイブリダイズさせるステップであって、各プローブが、特定核酸分子内のプローブ標的領域と相補的である、ステップと;

c. 各プローブを、追加された前記第1のアダプター配列へと伸長させて、前記第1のアダプター配列および第2のアダプター配列を有する複数のプローブ伸長産物を生成して、複数のプローブ伸長産物を産生するステップと;

d. 前記複数のプローブ伸長産物をシーケンシングして、前記複数のプローブ伸長産物の各々についての配列データを生成するステップと;

e. 前記複数のプローブ伸長産物の各々の配列を、プローブデータベースの基準コピー内の所定の配列に対してアラインさせるステップであって、前記所定の配列が、各プローブに特異的である、ステップと;

f. その所定の配列に対してアラインさせた各プローブ配列の数を決定するステップであって、前記数が、前記プローブが相補的である前記特定核酸分子の量を指し示す、ステップと

を含む、方法。

10

20

30

40

50

( 項目 4 )

複数の特定核酸分子を定量化するための方法であって、

a . 各プローブが前記複数の特定核酸分子内の一特定核酸分子内のプローブ標的領域とハイブリダイズし、かつ各プローブがその 5 ' 末端において第 1 のアダプターを有する複数のプローブを伸長させて、複数の伸長産物を生成するステップと；

b . 第 2 のアダプターを、前記複数のプローブ伸長産物の二本鎖末端へと追加するステップと；

c . 前記複数のプローブ伸長産物をシーケンシングして、前記プローブ伸長産物の各々の配列データを生成するステップと；

d . プローブ標的領域とハイブリダイズした各プローブの数を決定するステップであって、前記数が、前記プローブ標的領域を含む、前記特定核酸分子の各々の量を指し示す、

ステップと

を含む、方法。

( 項目 5 )

組成物中の複数の特定核酸分子を定量化するための方法であって、

a . 複数のプローブを、特定核酸分子内のプローブ標的領域とハイブリダイズさせるステップであって、各プローブが、その 5 ' 末端に、第 1 のアダプターを有する、ステップと；

b . 各プローブを伸長させて、前記第 1 のアダプター配列を含む複数のプローブ伸長産物を生成するステップと；

c . 第 2 のアダプター配列を、前記複数のプローブ伸長産物の二本鎖末端へと追加するステップと；

d . 前記複数のプローブ伸長産物をシーケンシングして、前記複数のプローブ伸長産物の各々の配列を生成するステップと；

e . 前記複数のプローブ伸長産物の各々の前記配列を、プローブデータベース内の所定の配列に対してアラインさせるステップであって、前記プローブデータベースが、複数の所定の配列を含み、各所定の配列が、プローブに特異的である、ステップと；

f . 各プローブ伸長産物の前記配列についての、前記シーケンシングデータベース内の所定の配列に対するアラインメントの数を決定するステップであって、前記アラインメントの数が、前記プローブがハイブリダイズする前記特定核酸分子の各々の量を指し示す、

ステップと

を含む、方法。

( 項目 6 )

前記複数のプローブ伸長産物の前記配列が、フォワードリード、インデックスリード、およびリバースリードのうちの少なくとも 1 つを含む、項目 1 から 5 のいずれか一項に記載の方法。

( 項目 7 )

各プローブがそのそれぞれの特定核酸内のそのそれぞれのプローブ標的領域配列へとアニールした特異性を検証する、項目 6 に記載の方法。

( 項目 8 )

前記複数のプローブ伸長産物の前記配列を、ゲノムまたはトランスクリプトームの座標へとマッピングして、意図されるプローブのアニーリングおよび伸長を検証する、項目 6 に記載の方法。

( 項目 9 )

前記複数のプローブ伸長産物の前記配列を、ゲノムまたはトランスクリプトームの座標へとマッピングする、項目 6 に記載の方法。

( 項目 10 )

前記複数のプローブ伸長産物の前記配列を、プローブデータベースの基準コピーに対してアラインさせて、意図されるプローブのアニーリングおよび伸長を検証する、項目 7 に記載の方法。

10

20

30

40

50

(項目 1 1)

前記リバースリードが、前記プローブ標的領域を含む、項目 6 に記載の方法。

(項目 1 2)

前記フォワードリードが、前記プローブ標的領域を含む、項目 6 に記載の方法。

(項目 1 3)

前記フォワードリードおよび前記リバースリードの前記複数のプローブ伸長産物の前記配列を、前記複数の特定核酸についてマッピングする、項目 8 および 9 のいずれか一項に記載の方法。

(項目 1 4)

前記インデックスリードの前記複数のプローブ伸長産物の前記配列により、試料バーコード配列および n - ランダム配列のうちの少なくとも 1 つを同定する、項目 6 に記載の方法。

10

(項目 1 5)

前記フォワードリードのマッピング座標と、前記インデックスリードの n - ランダム塩基との組み合わせにより、各プローブ伸長産物についての P C R 重複を決定する、項目 8 に記載の方法。

(項目 1 6)

同じフォワードリード座標および同じ n - ランダム塩基配列を有する配列を、重複として同定し、単一の特定核酸分子として一括し、計数する、項目 1 5 に記載の方法。

(項目 1 7)

フォワードリード座標は同じであるが、n - ランダム塩基配列は異なる配列の各々を、別々の特定核酸分子として計数する、項目 1 5 に記載の方法。

20

(項目 1 8)

前記フォワードリードのマッピング座標と、前記インデックスリードの n - ランダム塩基との組み合わせにより、特定核酸分子を同定する、項目 9 に記載の方法。

(項目 1 9)

同じフォワードリード座標および同じ n - ランダム塩基配列を有する配列を、重複として同定し、単一の特定核酸分子として一括し、計数する、項目 1 8 に記載の方法。

(項目 2 0)

フォワードリード座標は同じであるが、n - ランダム塩基配列は異なる配列の各々を、別々の特定核酸分子として計数する、項目 1 8 に記載の方法。

30

(項目 2 1)

前記フォワードリードと、対応するリバースリードとを、ペアエンドでアラインさせる、項目 1 8 に記載の方法。

(項目 2 2)

重複の一括に続き、各プローブ配列について計数されるリバースリードの数により、前記複数の特定核酸内の、各出発特定核酸分子についての分子の数を表す値を生成する、項目 1 6 および 1 9 のいずれか一項に記載の方法。

(項目 2 3)

重複の一括に続き、各プローブ配列について計数されるフォワードリードの数により、前記複数の特定核酸内の、各出発特定核酸分子についての分子の数を表す値を生成する、項目 1 6 および 1 9 のいずれか一項に記載の方法。

40

(項目 2 4)

各プローブ配列について計数されるリバースリードの数により、前記複数の特定核酸内の、各出発特定核酸分子についての分子の数を表す値を生成する、項目 1 7 および 1 9 のいずれか一項に記載の方法。

(項目 2 5)

前記ゲノムが、哺乳動物、細菌、ウイルス、リケッチア、または植物のゲノムまたはトランスクリプトームからなる群より選択される、項目 8 および 9 のいずれか一項に記載の方法。

50

( 項目 2 6 )

前記第 1 のアダプターを追加する前に、前記複数の特定核酸に末端修復を施す、項目 2 および 3 のいずれか一項に記載の方法。

( 項目 2 7 )

前記プローブを、ポリメラーゼにより伸長させる、項目 1 から 5 のいずれか一項に記載の方法。

( 項目 2 8 )

前記ポリメラーゼが、DNA ポリメラーゼ、RNA ポリメラーゼ、または逆転写酵素からなる群より選択される、項目 2 7 に記載の方法。

( 項目 2 9 )

前記アダプターの追加に続き、前記プローブ伸長産物を、任意選択で増幅する、項目 4 から 5 のいずれか一項に記載の方法。

( 項目 3 0 )

前記第 2 のアダプターを付加する前に、前記プローブ伸長産物を、制限エンドヌクレアーゼで処理するか、またはこれに末端修復を施す、項目 1 および 4 から 5 のいずれか一項に記載の方法。

( 項目 3 1 )

前記末端修復が、平滑末端修復である、項目 3 0 に記載の方法。

( 項目 3 2 )

前記プローブ伸長産物の伸長が、第 1 のアダプターの付加をさらに含む、項目 1 および 4 から 5 のいずれか一項に記載の方法。

( 項目 3 3 )

前記プローブ伸長産物の増幅が、前記増幅産物の各末端へのフローセル配列の付着をさらに含む、項目 2 9 に記載の方法。

( 項目 3 4 )

前記制限エンドヌクレアーゼで処理されたプローブ伸長産物により、共通の末端を伴うフォワードリードをもたらす、項目 3 1 に記載の方法。

( 項目 3 5 )

前記複数のプローブ伸長産物の前記配列を、ゲノムまたはトランスクリプトームの座標へとマッピングして、意図されるプローブのアニーリングおよび伸長を検証する、項目 3 4 に記載の方法。

( 項目 3 6 )

前記複数のプローブ伸長産物の前記配列を、プローブデータベースの基準コピーに対してアラインさせて、意図されるプローブのアニーリングを検証する、項目 3 5 に記載の方法。

( 項目 3 7 )

前記フォワードリードのマッピング座標と、前記インデックスリードの n - ランダム塩基との組み合わせにより、各プローブ伸長産物についての PCR 重複を決定する、項目 3 6 に記載の方法。

( 項目 3 8 )

同じフォワードリード座標および同じ n - ランダム塩基配列を有する配列を、重複として同定し、単一の特定核酸分子として一括し、計数する、項目 3 7 に記載の方法。

( 項目 3 9 )

前記複数のプローブ伸長産物の前記配列を、プローブデータベースの基準コピーに対してアラインさせて、意図されるプローブのアニーリングを検証する、項目 3 4 に記載の方法。

( 項目 4 0 )

リバーシブル配列を、それらが、どのプローブ配列を表すのかに従いビンニングおよび計数し、各プローブが表される回数が、前記出発特定核酸分子が、元の試料中に存在する回数についての尺度である、項目 3 5 に記載の方法。

10

20

30

40

50

(項目 4 1)

フォワードリード配列を、それらが、どのプローブ配列を表すのかに従いピニングおよび計数し、各プローブが表された回数が、前記出発特定核酸分子が、元の試料中に存在した回数についての尺度であった、項目 3 6 に記載の方法。

(項目 4 2)

前記フォワードリードが、前記特定核酸配列の少なくとも一部を含む、項目 6 に記載の方法。

(項目 4 3)

前記フォワードリードが、前記特定核酸配列のうちの、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 10、少なくとも 11、少なくとも 12、少なくとも 13、少なくとも 14、少なくとも 15、少なくとも 16、少なくとも 17、少なくとも 18、少なくとも 19、少なくとも 20、または少なくとも 25 塩基を含む、項目 6 に記載の方法。

10

(項目 4 4)

前記第 1 のアダプター配列が、インデックス配列のプライミング部位、インデックスヌクレオチド配列、n - ランダムヌクレオチド配列、フォワードリードプライミング部位、およびリバースリードプライミング部位、ならびにこれらの組み合わせのうちの少なくとも 1 つを含む、項目 2、3、および 3 1 のいずれか一項に記載の方法。

(項目 4 5)

前記第 2 のアダプター配列が、フォワードリードプライミング部位、リバースリードプライミング部位、およびリンカー配列、ならびにこれらの組み合わせのうちの少なくとも 1 つを含む、項目 2、3、および 3 1 のいずれか一項に記載の方法。

20

(項目 4 6)

前記第 2 のアダプター配列が、インデックス配列のプライミング部位、インデックスヌクレオチド配列、n - ランダムヌクレオチド配列、フォワードリードプライミング部位、リバースリードプライミング部位、ならびにこれらの組み合わせのうちの少なくとも 1 つを含む、項目 1、4 から 5、および 3 1 のいずれか一項に記載の方法。

(項目 4 7)

前記第 1 のアダプター配列が、フォワードリードプライミング部位、リバースリードプライミング部位、およびリンカー配列、ならびにこれらの組み合わせのうちの少なくとも 1 つを含む、項目 4 から 5、および 3 1 のいずれか一項に記載の方法。

30

(項目 4 8)

前記 5' の第 1 のアダプターが、各プローブ伸長産物に共通である、項目 4 から 5、および 3 1 のいずれか一項に記載の方法。

(項目 4 9)

前記インデックスリードが、インデックスヌクレオチド配列および n - ランダム塩基配列のうちの、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 10、少なくとも 11、少なくとも 12、少なくとも 13、少なくとも 14、または少なくとも 15 塩基を含む、項目 6 に記載の方法。

(項目 5 0)

前記インデックスリードが、n - ランダム塩基およびインデックスヌクレオチド配列のうちの、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、または少なくとも 10 塩基を含む、項目 6 に記載の方法。

40

(項目 5 1)

前記インデックスリードが、前記 n - ランダム塩基および任意選択で前記インデックスヌクレオチド配列のうちの、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、または少なくとも 10 塩基を含む、項目 6 に記載の方法。

。

(項目 5 2)

前記 n - ランダム塩基のヌクレオチド配列が、少なくとも 1、少なくとも 2、少なくと

50

も 3、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、または少なくとも 10ヌクレオチドを含む、項目 44 および 46 のいずれか一項に記載の方法。

(項目 53)

前記インデックスヌクレオチド配列が、バーコード配列をさらに含む、項目 44 および 46 のいずれか一項に記載の方法。

(項目 54)

前記リバースリードが、プローブ標的領域配列および前記特定核酸配列の一部、ならびにこれらの組み合わせのうちの少なくとも 1 つを含む、項目 6 に記載の方法。

(項目 55)

前記リバースリードが、プローブ配列のうちの、少なくとも少なくとも 20、少なくとも 25、少なくとも 30、少なくとも 35、少なくとも 40、少なくとも 45、少なくとも 50、少なくとも 55、または少なくとも 60 塩基を含む、項目 54 に記載の方法。

(項目 56)

前記リバースリードが、前記プローブ配列に対して 3' の特定核酸配列のうちの、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 10、少なくとも 11、少なくとも 12、少なくとも 13、少なくとも 14、少なくとも 15、または少なくとも 20 塩基を含む、項目 54 に記載の方法。

(項目 57)

項目 29 に記載の方法により増幅されるプローブ伸長産物を含む組成物。

(項目 58)

項目 1 から 5 のいずれか一項に記載の方法により産生されるプローブ伸長産物を含む組成物。

(項目 59)

前記複数の核酸が、組織、臓器、単一の細胞、腫瘍、患者から採取された有機流体の検体、遊離循環核酸、真菌、原核生物、およびウイルスからなる群より選択される試料に由来する、項目 1 から 5 のいずれか一項に記載の方法。

(項目 60)

前記患者が、腫瘍を有することが知られているか、またはこれを有することが疑われる、項目 59 に記載の方法。

(項目 61)

前記有機流体が、少なくとも 1 つの循環腫瘍細胞 (CTC) または播種性腫瘍細胞 (CTD) を含有する、項目 59 に記載の方法。

(項目 62)

前記患者が、ウイルス感染を有することが知られているか、またはこれを有することが疑われる、項目 59 に記載の方法。

(項目 63)

前記ウイルス感染が、伝染性感染または伝染性疾患である、項目 62 に記載の方法。

(項目 64)

前記組成物が、複数の核酸分子をさらに含む、項目 1 から 5 に記載の方法。

(項目 65)

各プローブ伸長産物が、特定核酸分子内のプローブ標的領域と相補的なプローブの伸長産物である、項目 1 から 5 に記載の方法。

【0026】

参照による組み込み

本明細書で言及される、全ての刊行物、特許、および特許出願は、各個別の刊行物、特許、または特許出願を、参照により組み込むことが、具体的かつ個別に指し示された場合と同じ程度に、参照により本明細書に組み込まれる。

【0027】

係属中の出願である、USN 13 / 750 , 768、USN 14 / 030 , 761

10

20

30

40

50

、U S S N 6 1 / 9 0 3 , 8 2 6、およびU S S N 6 1 / 9 8 9 , 1 1 3 は、参照によりそれらの全体が本明細書に組み込まれる。

【 0 0 2 8 】

開示される発明の新規の特徴および利点についてのよりよい理解は、開示される発明の原理を利用する、例示的な実施形態について明記する以下の記載と、付属の図面とを参照することにより得ることができる。

【図面の簡単な説明】

【 0 0 2 9 】

【図 1】図 1 は、g D N A を使用する本明細書で開示されるライブラリーの生成についての実施形態を例示するフローチャートである。

10

【 0 0 3 0 】

【図 2】図 2 は、c D N A を使用する本明細書で開示されるライブラリーの生成についての実施形態を例示するフローチャートである。

【 0 0 3 1 】

【図 3】図 3 は、二本鎖 g D N A を使用する本明細書で開示されるライブラリーの生成についての実施形態を例示するフローチャートである。

【 0 0 3 2 】

【図 4】図 4 は、二本鎖 g D N A を使用する本明細書で開示されるライブラリーの生成についての実施形態を例示するフローチャートである。

【 0 0 3 3 】

【図 5】図 5 は、シーケンシングライブラリーと、シーケンシングリードの領域とを構築するための本明細書で開示される実施形態を例示する。

20

【 0 0 3 4 】

【図 6】図 6 は、シーケンシングデータから、重複リードを除去するための本明細書で開示される実施形態を例示する。図 6 A : フォワードリード、図 6 B : インデックスリード、図 6 C : リバースリード。

【 0 0 3 5 】

【図 7】図 7 は、シーケンシングされた領域を同定して、配列データを得るための本明細書で開示される実施形態を例示する：図 7 A : フォワードリード、図 7 B : インデックスリード、図 7 C : リバースリード。

30

【 0 0 3 6 】

【図 8】図 8 は、シーケンシングされた領域を同定して、配列データを得るための本明細書で開示される実施形態を例示する：図 8 A : 配列リードを含有するプローブ、図 8 B : 特定核酸についてのシーケンシングリード、図 8 C : インデックス塩基リードおよび n - ランダム塩基リード、またはこれらの組み合わせのうちの少なくとも 1 つを含む、インデックス化シーケンシングリード。

【 0 0 3 7 】

【図 9】図 9 は、シーケンシングライブラリーの生成および後続のデジタル定量化のための本明細書で開示される実施形態を、図により例示する。

【 0 0 3 8 】

【図 1 0】図 1 0 は、デジタル定量化のための配列データをシーケンシングおよび解析する N G S のための、シーケンシングライブラリーを使用するための本明細書で開示される実施形態を、図により例示する。

40

【 0 0 3 9 】

【図 1 1 - 1】図 1 1 は、染色体の順序による、9 5 の遺伝子のパネル内の、R N A レベルにおける遺伝子の存在度についてのプロットを、グラフにより例示する。赤で染色された遺伝子は、顕著に下方調節され、緑で染色された遺伝子は、顕著に上方調節されている。エラーバーは、D N A データおよび R N A データの両方における標準偏差を反映する。

【図 1 1 - 2】図 1 1 は、染色体の順序による、9 5 の遺伝子のパネル内の、R N A レベルにおける遺伝子の存在度についてのプロットを、グラフにより例示する。赤で染色され

50

た遺伝子は、顕著に下方調節され、緑で染色された遺伝子は、顕著に上方調節されている。エラーバーは、DNAデータおよびRNAデータの両方における標準偏差を反映する。

【図11-3】図11は、染色体の順序による、95の遺伝子のパネル内の、RNAレベルにおける遺伝子の存在度についてのプロットを、グラフにより例示する。赤で染色された遺伝子は、顕著に下方調節され、緑で染色された遺伝子は、顕著に上方調節されている。エラーバーは、DNAデータおよびRNAデータの両方における標準偏差を反映する。

【0040】

【図12-1】図12は、染色体の順序で分取された509の遺伝子パネル内の全ての遺伝子について測定されたレベルのプロットを、グラフにより例示する。コピー数の変化した遺伝子は、緑で染色される。エラーバーは、試料データセットのプロープ計数と、対照データセットのプロープ計数とを組み合わせた変動を反映する。

10

【図12-2】図12は、染色体の順序で分取された509の遺伝子パネル内の全ての遺伝子について測定されたレベルのプロットを、グラフにより例示する。コピー数の変化した遺伝子は、緑で染色される。エラーバーは、試料データセットのプロープ計数と、対照データセットのプロープ計数とを組み合わせた変動を反映する。

【図12-3】図12は、染色体の順序で分取された509の遺伝子パネル内の全ての遺伝子について測定されたレベルのプロットを、グラフにより例示する。コピー数の変化した遺伝子は、緑で染色される。エラーバーは、試料データセットのプロープ計数と、対照データセットのプロープ計数とを組み合わせた変動を反映する。

【発明を実施するための形態】

20

【0041】

本開示は、デジタル測定値を結果としてもたらす標的化された核酸シーケンシングのための方法について記載する。これらのデジタル測定が有用である場合の例は、デジタル遺伝子発現およびコピー数変異においてである。出発材料は、核酸、DNA、RNA、cDNA、または二本鎖cDNAでありうる。開示される方法、組成物、およびキットは、そのプロープ標的領域とハイブリダイズさせた相補的プロープを利用して、プロープ標的領域に由来するプロープ伸長産物を生成することを記載する。プロープ伸長産物は、ハイスループットシーケンシングを進める、標的の濃縮およびライブラリーの生成のために使用する。シーケンシングデータの解析は、トランスクリプトームの遺伝子発現またはゲノムDNAのコピー数変異のデジタル測定値をもたらす。

30

【0042】

標的化プロープを、特定核酸とハイブリダイズさせ、NuGEN Ovation Target Enrichment Kitを使用して、ポリメラーゼにより伸長させる。ペアドエンドシーケンシングを、結果として得られる、濃縮されたライブラリーに対して実施することができる。リードを、ゲノムまたはトランスクリプトームへとマッピングし、PCR重複リードを同定する（特許出願であるUS 2016/0190382において記載されている）。次いで、プロープ配列を、重複を除外した（deduplicate）配列データセット内でそれらがどのくらいの回数出現するのかを、出発試料中に存在した元の核酸のコピーの数の尺度として計数する。ランダム配列ではなく、プロープ配列の計数を使用することにより、各デジタル測定値について、異なる試料にわたり、正確に同じ配列を評価するため、コピー数解析が簡素化される。これは、代替的エクソン使用に起因して、試料間で変化する、遺伝子の長さなどの因子について正規化するのに役立つほか、ゲノムまたはトランスクリプトームへのシーケンシングリードのマッピングに伴う公知の問題も低減する。

40

【0043】

開示される発明の方法は、RNA-Seq解析、デジタル遺伝子発現、遺伝子型決定、コピー数変異の決定、および全ゲノム増幅を含むがこれらに限定されない、遺伝子試料解析のための多様な適用と共に使用することができる。

【0044】

そうでないことが指定されない限りにおいて、生化学、核酸化学、分子生物学、および

50



分子遺伝学の用語および記号は、その分野内の標準的な取り決めおよび教科書、例えば、Sambrookら、Molecular Cloning: A Laboratory Manual、増補2版(Cold Spring Harbor Laboratory、1989年)；KornbergおよびBaker、DNA Replication、2版(W. H. Freeman、New York、1992年)；Gaits編、Oligonucleotide Synthesis: A Practical Approach(IRL Press、Oxford、1984年)；Lehninger、Biochemistry、2版(Worth Publishers、New York、1975年)；Eckstein編、Oligonucleotides and Analogs: A Practical Approach(Oxford University Press、New York、1991年)などによる用語および記号に従う。

10

**【0045】**

本明細書および付属の特許請求の範囲で使用される、単数形の「1つの(a)」、「1つの(an)」、および「その(the)」とは、文脈によりそうでないことが明らかに指示されるのでない限りにおいて、複数の指示対象を含む。こうして、例えば、「ポリメラーゼ」に対する言及は、1つの薬剤を指す場合もあり、このような薬剤の混合物を指す場合もあり、「方法」に対する言及は、当業者に公知の同等なステップおよび/または方法などに対する言及を含む。

**【0046】**

20

加えて、理解を容易とするために、本明細書で規定される、多数の用語についても開示する。

**【0047】**

本明細書で使用される「アダプター」という用語は、配列が公知のオリゴヌクレオチドであって、目的の特定核酸配列または標的ポリヌクレオチド鎖へのその付着が、目的の特定核酸または標的ポリヌクレオチド鎖の増幅準備のできた産物の生成を可能とするオリゴヌクレオチドを指す場合がある。少なくとも1つのアダプターの付加の前に、特定核酸試料を断片化する場合もあり、断片化しない場合もある。

**【0048】**

目的の特定配列領域/鎖の、増幅準備のできた産物の生成に適する、多様なアダプターデザインが想定される。例えば、二本鎖アダプターを使用する場合、アダプターの2つの鎖は、自己相補的な場合もあり、非相補的な場合もあり、部分的に相補的な場合もある。アダプターは、少なくとも部分的なフォワード配列プライミング部位と、ランダム配列とを含有しうる。

30

**【0049】**

一部の実施形態では、アダプターは、さらなる識別子配列、例えば、バーコード配列を含む。本明細書で使用する「バーコード」という用語は、バーコードを会合させたポリヌクレオチドのいくつかの特徴の同定を可能とする、公知の核酸配列を指す場合がある。一部の実施形態では、同定されるポリヌクレオチドの特徴は、ポリヌクレオチドが由来する試料でありうる。バーコードは、例えば、標的ポリヌクレオチドへと接続されると、標的ポリヌクレオチドが由来した試料の識別子として用いられうる核酸配列を含みうる。一部の実施形態では、バーコードは、少なくとも3、4、5、6、7、8、9、10、11、12、13、14、15、またはこれを超えるヌクレオチドの長さである。一部の実施形態では、バーコードは、10、9、8、7、6、5、または4ヌクレオチドの長さより短い。一部の実施形態では、複数のバーコード中の各バーコードは、少なくとも3つ、4つ、5つ、6つ、7つ、8つ、9つ、10、またはこれを超える位置など、少なくとも3つのヌクレオチド位置において、複数のバーコード中の他のあらゆるバーコードと異なる。一部の実施形態では、いくつかのポリヌクレオチドと会合させたバーコードは、他のポリヌクレオチドと会合させたバーコードと長さが異なる。バーコードは、それらが会合するバーコードに基づき、試料の同定を可能とするのに十分な長さであり、これを可能とす

40

50

るのに十分に異なる配列を含みうる。一部の実施形態では、フォワードアダプターおよびリバースアダプターのいずれも、複数のバーコード配列のうちの少なくとも1つを含みうる。一部の実施形態では、第1のアダプターおよび第2のアダプターは、複数のバーコード配列のうちの少なくとも1つを含む。一部の実施形態では、各リバースアダプターは、複数のバーコード配列のうちの少なくとも1つを含み、この場合、複数のバーコード配列のうちの、各バーコード配列は、複数のバーコード配列中の他のあらゆるバーコード配列と異なる。一部の実施形態では、第1のアダプターおよび第2のアダプターのいずれも、複数のバーコード配列のうちの少なくとも1つを含む。一部の実施形態では、第2のアダプターのオリゴヌクレオチドについてのバーコードを、第1のアダプターのオリゴヌクレオチドについてのバーコードと独立に選択する。一部の実施形態では、対をなすアダプターが、同じであるか、または異なる1種または複数のバーコードを含むように、バーコードを有する、第1のアダプターのオリゴヌクレオチドと、第2のアダプターのオリゴヌクレオチドとを対合させる。一部の実施形態では、本発明の方法は、標的ポリヌクレオチドが接続されるバーコード配列に基づき、標的ポリヌクレオチドが由来しうる試料を同定するステップをさらに含む。バーコードは、例えば、標的ポリヌクレオチドへと接続されると、標的ポリヌクレオチドが由来した試料の識別子として役立つ核酸配列を含みうる。

#### 【0050】

ライゲーションを利用して、目的の配列領域の所望の末端に、アダプターを追加することは、開示される方法を実行するのに適しうる。核酸、核酸修飾酵素、および結果として得られる、核酸のライゲーション可能な末端の選択に応じて、多様なライゲーションモダリティが想定される。例えば、目的の標的領域/配列を含む平滑末端産物を生成しうる場合は、平滑末端ライゲーションが適しうる。代替的に、公知の配列特異性を有する制限酵素を使用して、切断を実行し、公知の配列突出を伴う切断部位の生成をもたらす場合は、目的の配列領域の切断部位との、アダプターのハイブリダイゼーションと、後続のライゲーションとを可能とするように、アダプターの適切な末端を設計することができる。ライゲーションはまた、問題の核酸の配列を得るように、さらに修飾されうる、単一の核酸配列を結果としてもたらず2つの核酸分子の任意の接続も指す場合がある。アダプターの効率的かつ迅速なライゲーションのための試薬および方法は、市販されており、当技術分野でも公知である。

#### 【0051】

本明細書で使用される「増幅すること」、「増幅」、および特定核酸を「増幅する」という用語は、目的の核酸試料の複数のコピーを、例えば、DNAコピーの形態で生成する手順を指す場合がある。当技術分野では、例えば、PCRおよびqPCRなど、核酸を増幅する多くの方法およびプロトコルが公知である。

#### 【0052】

本明細書で使用される「cDNA」という用語は、相補的DNAを指す場合がある。相補的DNAは、酵素である逆転写酵素およびDNAポリメラーゼにより触媒される反応において、メッセンジャーRNA(mRNA)鋳型から合成することができる。

#### 【0053】

本明細書で使用される「相補的」という用語は、配列の全部に対する相補性を指す場合もあり、配列のうちの一部だけに対する相補性を指す場合もある。特異的オリゴヌクレオチドプライマーまたはプローブの、ハイブリダイズ可能な配列内のヌクレオチドの数は、オリゴヌクレオチドプライマーまたはプローブをハイブリダイズさせるのに使用されるストリンジェンシーの条件により、過剰でランダム非特異的ハイブリダイゼーションが防止されるような数でありうる。オリゴヌクレオチドプライマーまたはプローブのうちの、ハイブリダイズする部分内のヌクレオチドの数は、標的ポリヌクレオチド上の規定された配列であって、オリゴヌクレオチドプライマーまたはプローブがハイブリダイズする配列と少なくとも同じくらいの大きさの数、すなわち、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも約20であることが可

能であり、12～約200ヌクレオチド、通例、約20～約50ヌクレオチドのうちの、約6～約10、または6～約12ヌクレオチドでありうる。標的ポリヌクレオチド/オリゴヌクレオチドは、オリゴヌクレオチドプライマー、プライマー、またはプローブより大型でありうる。

【0054】

本明細書で使用される「～を変性させること」という用語は、二本鎖核酸の、一本鎖への分離を指す場合がある。変性は、当技術分野で公知の方法であって、物理的変性、熱的変性、および/または化学的変性を含むがこれらに限定されない方法のうちのいずれかを使用して達成することができる。

【0055】

本明細書で使用される「FFPE」という頭字語は、ホルマリン固定パラフィン包埋を明示する。FFPEとは、組織試料の保存において使用される方法であって、試料を、パラフィンと称する蠟を適用してカップリングさせたホルマリン溶液中で固定しうる方法である。

【0056】

本明細書で使用される「ゲノムDNA」という語句は、ゲノムデオキシリボ核酸を、gDNAと略記した、染色体のDNAを指す場合がある。gDNAは、生物の遺伝物質を含む。

【0057】

本明細書で使用される「ゲノム」という用語は、患者、組織、臓器、単一の細胞、腫瘍、患者から採取された有機流体の検体、遊離循環核酸、真菌、原核生物、およびウイルスに由来する、DNA、RNA、またはcDNAである配列を指す場合がある。本明細書で使用される「トランスクリプトーム」とは、発現される、生物の部分的ゲノムまたは全ゲノムを反映しうる、全てのRNA配列でありうる。

【0058】

本明細書で使用される「キット」という用語は、材料を送達するための任意のシステムを指す場合がある。反応アッセイの文脈では、このような送達システムは、適切な容器内の、オリゴヌクレオチド、緩衝成分、添加剤、反応増強剤、酵素などの反応成分の、1つの場所から別の場所への保存、輸送、または送達を可能とするエレメントであって、一般に、アッセイを実施するための書面での指示と共に提供されるエレメントを含みうる。キットは、関連する反応試薬と、支持材料とを含有する、1つまたは複数の筐体または箱を含みうる。キットは、2つまたはこれを超える個別の容器を含む場合があり、これらの容器の各々は、全キット成分のうちの一部を含む。容器は、意図されるレシピエントへと、一緒にまたは別個に送達される場合もある。

【0059】

本明細書で使用される「核酸(NA)修飾酵素」という語句は、DNA特異的修飾酵素を指す場合がある。NA修飾酵素は、二本鎖DNAに対する特異性について選択することができる。酵素は、二重鎖特異的エンドヌクレアーゼ、平滑末端頻発カッター(b l u n t - e n d f r e q u e n t c u t t e r)制限酵素、または他の制限酵素であり得る。平滑末端カッターの例は、D r a IまたはS m a Iを含みうる。NA修飾酵素は、N e w E n g l a n d B i o l a b sにより提供されている酵素でありうる。NA修飾酵素は、ホーミングエンドヌクレアーゼ(ホーミングエンドヌクレアーゼとは、厳密に規定された認識配列を有さないエンドヌクレアーゼでありうる)でありうる。NA修飾酵素は、ニッキ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とは、二本鎖DNA基質内のDNAの1つの鎖だけを切断しうる、エンドヌクレアーゼでありうる)でありうる。NA修飾酵素は、高忠実度エンドヌクレアーゼ(高忠実度エンドヌクレアーゼとは、エンドヌクレアーゼの野生型バージョンより「スター活性」が小さくなるように操作された、エンドヌクレアーゼでありうる)でありうる。一部の実施形態では、NA修飾酵素は、配列特異的であり、かつ、二重鎖特異的な、DNA修飾酵素でありうる。

【0060】

本明細書で使用される「核酸断片」および「特定核酸」という語句は、互換的に使用され、本明細書で使用される通り、核酸試料の一部を指す場合がある。インプット試料中の核酸は、断片化された核酸分子の集団へと断片化することもでき、1種または複数の特異的なサイズ範囲のポリヌクレオチドへと断片化することもできる。断片の平均の長さは、約10～約10,000ヌクレオチド、約50～約2,000ヌクレオチド、約100～2,500、10～1,000、10～800、10～500、50～500、50～250、または50～150ヌクレオチドの長さでありうる。断片の平均の長さは、10,000ヌクレオチド未満、5,000ヌクレオチド未満、2,500ヌクレオチド未満、2,000ヌクレオチド未満、1,000ヌクレオチド未満、400ヌクレオチド未満、300ヌクレオチド未満、200ヌクレオチド未満、または150ヌクレオチド未満など、500ヌクレオチド未満でありうる。

10

#### 【0061】

本明細書で使用される「特定核酸配列」または「特定配列」という語句は、デジタル測定および/またはデジタル定量化が所望される、目的のポリヌクレオチド配列であって、核酸断片を含むがこれらに限定されない配列でありうる。特定配列は、その実際の配列に関して、公知の場合もあり、公知でない場合もある。本明細書で使用される「鋳型」とは、特定核酸配列を含有するポリヌクレオチドでありうる。「特定配列」、「特定核酸配列」、「特定ヌクレオチド配列」、「目的の領域」、または「目的の配列」という用語、およびこれらの変形は、互換的に使用される。

#### 【0062】

20

本明細書で使用される「適格化核酸」および「標的核酸断片を適格化する」という語句は、i) DNAポリメラーゼに許容可能な鋳型、すなわち、鋳型がDNAポリメラーゼに対する架橋もしくは阻害剤を含有する可能性がない、またはii) 定量化、増幅、検出を目的として、断片を修飾しうるように、ポリヌクレオチド配列の5'末端および/もしくは3'末端における、バーコード、アダプター、プライマーと相補的な配列などのうちの少なくとも1つの付着を含むがこれらに限定されない修飾を有する鋳型である、gDNA配列またはRNA配列の断片を指す場合もあり、当業者に公知の他の方法であって、gDNA配列解析およびcDNA配列解析の方法を指す場合もある。阻害剤の存在は、FFPE調製中の固定を経た組織試料から得られたgDNAを使用する結果でありうる。

#### 【0063】

30

本明細書で使用される「オリゴヌクレオチド」という用語は、200残基長未満、例えば、15～100ヌクレオチド長の間のポリヌクレオチド鎖を指す場合があるが、また、より長いポリヌクレオチド鎖も包含しうる。オリゴヌクレオチドは、一本鎖の場合もあり、二本鎖の場合もある。本発明で使用される「オリゴヌクレオチド」という用語は、「プライマー」、「プローブ」、および「アダプター」という用語と互換的に使用することができる。

#### 【0064】

「PCR」とは、本発明の全ての方法において使用される核酸増幅技術であり、元は、Mullis K. B.ら、米国特許第4,683,195号およびMullis K. B.、米国特許第4,683,202号により発見され、記載された、「ポリメラーゼ連鎖反応」という用語の略号である。一部の実施形態では、PCRは、1つのプライマーの伸長が、次のPCRサイクルにおける別のプライマーのための鋳型をもたらすなど、設計される各鎖について、2つのオリゴヌクレオチドプライマーを用いる。本明細書では、考察されるオリゴヌクレオチドプライマーを識別することを目的として、オリゴヌクレオチドプライマーの対のうちのいずれか1つを、「フォワード」プライマーまたは「リバース」プライマーと名づけることができる。PCRは、(i) 二本鎖核酸の鎖を分離する変性ステップに続く、(ii) プライマーを、目的の配列を挟む位置にアニールさせるアニリングステップと、次いで、(iii) プライマーを、5'から3'の方向で伸長させ、これにより、標的配列と相補的な核酸断片を形成する伸長ステップとの反復(またはサイクル)からなりうる。上記のステップの各々は、自動式サーモサイクラーを使用して、異

40

50

なる温度で行うことができる。PCRサイクルを、所望される頻度で反復する結果として、その末端が、通例、使用されるプライマーの5'末端により規定される、標的DNA断片の指数関数的蓄積をもたらすことができる。この規則には、本明細書で記載される例外を含む、ある特定の例外も適用されうる。各ステップにおける、特定の温度、インキュベーション時間、およびステップの間の変化の速度は、当業者に周知の多くの因子に依存し、例は、例えば、McPherson M. J.ら(1991年および1995年)など、公表されている多数のプロトコール中に見出すことができる。PCR条件は、広い範囲で変動させうるが、二本鎖標的核酸は、>90の温度で変性させることができ、プライマーは、50~75の範囲の温度でアニールさせることができ、伸長は、72~78の範囲で実施することができる。

10

**【0065】**

本明細書で使用される「定量的PCR」または「qPCR」という語句は、試料中の、1種または複数の特定標的配列の存在度を測定するように設計されたPCRを指す場合がある。定量的測定値は、個別にアッセイすることもできるか、または標的核酸と併せてアッセイすることもできる、1種または複数の基準核酸配列を使用して作成することができる。当技術分野では、定量的PCRのための技法が周知であり、参照により本明細書に組み込まれる、以下の文書：Gu Z.ら(2003年)、J. Clin. Microbiol., 41巻：4636~4641頁；Becker-Andre M.およびHahlbrock K.(1989年)Nucleic Acids Res., 17巻：9437~9446頁；Freeman W. M.ら(1999年)Biotechniques, 26巻：112~122頁、124~125頁；Lutfalla G.およびUze G.(2006年)、Methods Enzymol., 410巻：386~400頁；Clementi M.ら(1993年)PCR Methods Appl., 2巻：191~196頁；Diviacco S.ら(1992年)Gene, 122巻：313~320頁において例示されている。

20

**【0066】**

本明細書で使用される「部分」という語句は、全長に満たない核酸配列、核酸配列断片、特定核酸配列、特定核酸断片、プローブ、プライマーなどを指す場合がある。部分は、約50~約2,000ヌクレオチド、約100~2,500、10~1,000、10~800、10~500、20~250、または20~150ヌクレオチド未満の長さでありうる。

30

**【0067】**

本明細書で使用される「プライマー」という用語は、鋳型(特異的なポリヌクレオチド、標的DNA、標的RNA、プライマー伸長産物、またはプローブ伸長産物など)とハイブリダイズまたはアニールすることが可能であり、また、鋳型と相補的なポリヌクレオチドの重合を促進することも可能でありうる、遊離3'ヒドロキシル基を一般に伴う、オリゴヌクレオチドを指す場合がある。プライマーは、プライマーのテールを構成する、ハイブリダイズしない配列を含有しうる。プライマーは、その配列が、標的と完全に相補的でなくてもやはり、標的とハイブリダイズしうる。

**【0068】**

本明細書で利用されるプライマーは、PCR、qPCR、伸長反応など、ポリヌクレオチド鋳型に沿った、ポリメラーゼによる伸長反応において用いられるオリゴヌクレオチドでありうる。オリゴヌクレオチドプライマーは、一本鎖でありうる合成ポリヌクレオチドであって、その3'末端において、標的ポリヌクレオチドの配列とハイブリダイズすることが可能でありうる配列を含有する合成ポリヌクレオチドでありうる。

40

**【0069】**

特定核酸とハイブリダイズするプライマーの3'領域は、配列またはプライマー結合性部位に対する、少なくとも80%、好ましくは90%、より好ましくは95%、最も好ましくは100%の相補性を含みうる。

**【0070】**

50

「テール配列」という用語は、プライマー配列またはプローブ配列に隣接し、かつその5'にあるハイブリダイズしない配列を指す場合がある。「プローブ伸長産物」という用語は、例えば、特定核酸配列内で、プローブのハイブリダイゼーションと、プローブから開始される鋳型指向性合成とから生じる、DNA断片を指す場合がある。プローブは、存在し、特定核酸へと追加される場合、ポリメラーゼにより、アダプター配列へと伸長させることができる。結果として得られるプローブ伸長産物は、第1のアダプター、例えば、特定核酸配列へと追加されるアダプターと、例えば、プライマーまたはプローブのテール配列内に見出される、第2のアダプターとの両方を有しうる。

【0071】

本明細書で使用される「ランダムプライマー」とは、必ずしも、試料中の特定の ( p a r t i c u l a r ) 配列または特定配列に基づき設計されうるわけではなく、ランダムプライマーの配列は、試料中の1種または複数の配列とハイブリダイズ可能でありうる ( 所与のセットの条件下で ) という統計学的予測 ( または経験的観察 ) に基づきうる配列を含むプライマーでありうる。ランダムプライマーは、オリゴヌクレオチド上の所与の位置におけるヌクレオチドが、4つのヌクレオチドのうちのいずれか、または4つのヌクレオチドのうちの選択される群のうちのいずれか ( 例えば、4つのヌクレオチドのうちの3つだけ、または4つのヌクレオチドのうちの2つだけ ) でありうる、ランダム配列を含むオリゴヌクレオチドまたはオリゴヌクレオチドの集団でありうる。本明細書で使用された、「n - ランダムオリゴヌクレオチド」という表記とは、アダプター内またはプライミング部位内の、少なくとも0、少なくとも1つ、少なくとも2つ、少なくとも3つ、少なくとも4つ、少なくとも6つ、少なくとも8つ、少なくとも9つ、少なくとも10などの塩基を指す場合がある。

【0072】

本明細書で使用される「ランダムヌクレオチド」および「n - ランダムヌクレオチド配列」とは、アダプター内またはプライマー内に、必ずしも、試料中の特定の配列または特定配列に基づき設計されうるわけではなく、ランダムヌクレオチドを有するアダプターまたはプライマーは、プライマー内、アダプター内、または試料中の、1種または複数の配列とハイブリダイズ可能でありうる ( 所与のセットの条件下で ) という統計学的予測 ( または経験的観察 ) に基づきうる配列を含みうる、ヌクレオチドでありうる。ランダムオリゴヌクレオチドは、オリゴヌクレオチド上の所与の位置におけるヌクレオチドが、4つのヌクレオチドのうちのいずれか、または4つのヌクレオチドのうちの選択される群のうちのいずれか ( 例えば、4つのヌクレオチドのうちの3つだけ、または4つのヌクレオチドのうちの2つだけ、またはヌクレオチドのうちの1つだけ ) でありうる、ランダム配列を含むオリゴヌクレオチドまたはオリゴヌクレオチドの集団でありうる。本明細書で使用された、「n - ランダムオリゴヌクレオチド」という表記とは、アダプター内またはプライマー内の、少なくとも0、少なくとも1つ、少なくとも2つ、少なくとも3つ、少なくとも4つ、少なくとも6つ、少なくとも7つ、少なくとも8つ、少なくとも9つ、少なくとも10などの塩基を指す場合がある。

【0073】

本明細書で使用される「試料」という用語は、目的の核酸を含有するか、または目的の核酸を含有することが仮定される任意の物質を指す場合があり、こうして、核酸、細胞、生物、組織、流体 ( 例えば、脊髄液またはリンパ液 ) 、患者から採取された有機流体の試料と、血液、血漿、血清、尿、涙液、糞便、気道および尿生殖路、唾液、異なる臓器の断片、組織、血液細胞、循環腫瘍細胞 ( C T C ) または播種性腫瘍細胞 ( C T D ) 、骨、核酸分子を含有することが疑われている *i n v i t r o* における細胞培養物または検体の試料を含むがこれらに限定されない試料とを含む。

【0074】

「伝染性感染」および「伝染性疾患」という語句は、人から人へ；動物から動物への、動物からヒトへの、またはヒトから動物への、直接的な接触または近接による偶発的な接触により感染可能な感染症および疾患を指す場合がある。

## 【 0 0 7 5 】

本明細書で使用される「PCR重複」という用語は、元の同じ核酸分子に由来するので、別のシーケンシングリードと同じプライマー/プローブ伸長産物配列であり、したがって、固有の核酸分子を表さない、任意のシーケンシングリードを指す場合がある。

## 【 0 0 7 6 】

本明細書で使用される「プローブ」という用語は、オリゴヌクレオチド配列を指す場合がある。プローブは、プローブ標的領域と相補的でありうる。プローブ標的領域と相補的なプローブ配列は、約200残基長未満、例えば、約15～100ヌクレオチド長の間でありうるが、また、より長いポリヌクレオチド鎖を包含することも意図されうる。プローブ標的領域は、一本鎖の場合もあり、二本鎖の場合もある。プローブ標的領域は、ポリメラーゼを使用して伸長を施される、相補的プローブのためのハイブリダイゼーション部位をもたらす。

10

## 【 0 0 7 7 】

本明細書で使用される「プローブ標的領域」という用語は、ゲノムデータベース内もしくはトランスクリプトームデータベース内の領域、またはゲノム配列内もしくはトランスクリプトーム配列内の領域であって、それに対してプローブが設計される領域を指す場合がある。領域は、特異的な相補的領域を越えて伸長させることが可能であり、ゲノムまたはトランスクリプトームの隣接領域を含みうる。そのプローブ標的領域に対してアラインさせたプローブ配列は、プローブのアニーリングの特異性の検証をもたらすので、プローブ伸長産物の特異性の検証ももたらすことが可能であり、こうして、計数される特定核酸分子の特異性の検証ももたらしうる。

20

## 【 0 0 7 8 】

プローブ標的領域は、特定核酸配列内にある。プローブ標的領域は、約500残基長であることが可能であり、また、約80～1000残基の間でもありうる。本明細書で使用される「プローブ標的領域」という用語は、「プローブのハイブリダイゼーション部位」という用語および「プローブのアニーリング部位」と互換的に使用することができる。

## 【 0 0 7 9 】

本明細書で使用される「検証されたプローブ」または「検証されたプローブ配列」という用語は、存在することが検証され、結果として得られるシーケンシングデータに由来する、意図される特定標的核酸とハイブリダイズすることが検証された、プローブの配列を指す場合がある。

30

## 【 0 0 8 0 】

ここで、開示される発明の例示的な実施形態について、詳細に言及する。開示される方法および組成物について、例示的な実施形態と共に記載するが、これらの例示的な実施形態は、開示される発明を限定することを意図するものではないことが理解されるであろう。逆に、開示される発明は、開示される発明の精神および範囲に含まれうる、代替物、改変、および等価物を包含することを意図する。

## 【 0 0 8 1 】

一部の実施形態では、本明細書では、複数の核酸を含む試料に由来する、目的の特定核酸配列を定量化するための方法および組成物が開示される。本明細書で記載される方法は、従来のアダプター、配列特異的プローブ標的領域のプローブ、ポリメラーゼ酵素およびライゲーション酵素、ならびにライゲーションを使用して、特定核酸配列を増幅しうる。方法はさらに、トランスクリプトームまたはゲノムDNAに由来する、少なくとも第1の特定核酸配列のデジタル測定も可能としうる。

40

## 【 0 0 8 2 】

デジタル遺伝子発現は、複数の方式で実施されているが、各々が重大な欠点を有するので、新たな方法論が、核酸分子の適正なデジタル計数を実施するために重要となっている。デジタル核酸計数のための現行の方法は、デジタルPCR、ハイスループットシーケンシング、およびNanosttring n-counter systemにより実施される、ハイブリダイゼーションベースの計数を含みうる。

50

## 【0083】

デジタルPCRは、PCR容器、すなわち1プレート内の1ウェル内で、またはエマルジョン1液滴のいずれか当たり1コピーを得る点まで、出発核酸材料を希釈することにより実施することができる。エンドポイントPCRは、所与の標的プライマーのセットについて実施することができ、増幅事象に対して陽性であるウェルまたは液滴の数を計数することができる。この方法の主要な欠点は、ポアソン分布に基づき、容器1つ当たりの標的核酸1コピーを正確に得ることの問題であり、また、反応も、探査されうる核酸試料1例当たり少数の標的に大きく限定されうる（低マルチプレックス能）。

## 【0084】

Nanosttringのn-counter systemでは、蛍光シグナルを測定することにより、インプット核酸を計数する、単分子分解能を伴う、プローブハイブリダイゼーションスキームを利用する。この技術の大きな欠点は、使用しなければならない蛍光タグと、同じ分子上の異なる領域を標的化することができないことに起因する、低い多重化である。例えば、使用される蛍光タグのサイズに起因して、n-counter systemは、同じRNA転写物内の、2つのエクソンの存在について探査することが不可能でありうる。

## 【0085】

ハイスループットシーケンシングは、核酸分子のデジタル計数のための優れた方法であると考えうるが、これもまた、大きな欠点を抱えている。ゲノムDNAならびにRNA計数のいずれのためにも、シーケンシングの前に、核酸をランダムにせん断する場合がある。このランダムなせん断が、標的の塩基組成へとバイアスを導入する結果として、所与の目的の標的の不均質な増幅またはシーケンシングがもたらされうる。核酸断片の計数における両義性の大きな供給源は、計数するのに現在使用されている方法に基づきうる。すなわち、所与の目的の遺伝子（またはゲノム標的領域）について、異なるサイズであり、したがって、必然的に異なる数のシーケンシングリードを生成する標的を、互いと比較しうるように、得られるシーケンシングリードの数を、標的領域のサイズにより正規化しなければならない。同じ遺伝子の、異なる長さのアイソフォームが、様々な存在度で存在するので、標的領域のサイズは、試料間で必ずしも固定されないため、両義性が生じる。これは、RNAシーケンシングの場合に、最も容易に見ることができるが、ゲノムDNAにも同等に当てはまる。

## 【0086】

RNAシーケンシングでは、遺伝子計数は、生成されるデータの種類に応じて、RPKMまたはFPKM（計数10億当たりのリード/断片、または計数10億当たりの断片）として表すことができる。シーケンシングデータの計数は、リード（またはペアドエンドシーケンシングの場合は断片）の数、標的RNAのサイズ（キロベース単位）、およびシーケンシングリードの総数（百万単位）により決定することができる。問題は、標的RNAのサイズを測定することであり、1つのサイズが、全ての試料にわたり仮定されている。しかし、代替的エクソン使用により、RNAのサイズは、異なる試料間で、最大で何kbもの配列にわたり異なりうるため、2つの試料の間のRPKM/FPKM測定値では、サイズ変数が潜在的に変更されることが周知である。加えて、一定数のシーケンシングリードについて、代替的エクソン使用により、1つの遺伝子のサイズを変更することは、他の遺伝子に由来するリードの数も変化させることになるので、1つの遺伝子についてのサイズ測定値の変化は、試料中の全ての遺伝子についてのRPKM/FPKM測定値にも影響を及ぼす。RNAシーケンシングに関して記載したのとまったく同様に、試料間で目的の標的領域のサイズを変更する部分的な重複および欠失を考慮する場合、ゲノムDNA計数も、同様の問題を抱える場合がある。

## 【0087】

一部の実施形態では、本明細書では、複数の核酸を有する試料に由来する特定核酸配列のデジタル測定のための方法および組成物が開示される。核酸は、DNAの場合もあり、RNAの場合もある。核酸は、一本鎖の場合もあり、二本鎖の場合もある。DNAは、ゲ

10

20

30

40

50



ノムDNA、cDNA、DNA/RNAハイブリッド体、またはこれらの任意の組み合わせでありうる。一部の実施形態では、インプット試料中の核酸は、二本鎖DNAでありうる。一部の実施形態では、方法は、インプット試料中の核酸を断片化して、核酸断片を生成するステップを含む。一部の実施形態では、試料を断片化しない。一部の実施形態では、核酸の断片化は、核酸を断片化するための、当技術分野で公知であるか、または本明細書で記載される方法であって、物理的（すなわち、超音波処理）断片化反応、および/または酵素的（すなわち、制限酵素処理）断片化反応を含みうるがこれらに限定されない方法により達成することができる。

#### 【0088】

物理的断片化法は、噴霧化、超音波処理、および/または流体力学的せん断を含みうる。一部の実施形態では、断片化は、機械的に達することができ、インプット試料中の核酸を、音響的超音波処理にかけることを含む。一部の実施形態では、断片化は、インプット試料中の核酸を、1種または複数の酵素が二本鎖核酸の切断を生成するのに適する条件下で、1種または複数の酵素により処理することを含む。核酸断片またはポリヌクレオチド断片の生成に有用な酵素の例は、配列特異的ヌクレアーゼおよび配列非特異的ヌクレアーゼを含みうる。ヌクレアーゼの非限定的な例は、DNアーゼI、Fragmentase、制限エンドヌクレアーゼ、これらの改変体、およびこれらの組み合わせを含みうる。酵素的断片化反応を実行するための試薬は、市販されている（例えば、New England Biolabsから）。例えば、DNアーゼIによる消化は、 $Mg^{++}$ の非存在下および $Mn^{++}$ の存在下で、DNA内のランダム二本鎖切断を誘導しうる。一部の実施形態では、断片化は、インプット試料中の核酸を、1種または複数の制限エンドヌクレアーゼで処理することを含む。断片化は、5'突出、3'突出、平滑末端、またはこれらの組み合わせを有する断片をもたらす。断片化が1種または複数の制限エンドヌクレアーゼの使用を含む場合など、一部の実施形態では、試料ポリヌクレオチドの切断は、予測可能な配列を有する突出を後に残す。

#### 【0089】

一部の実施形態では、インプット試料中の核酸は、断片化された核酸分子の集団へと断片化することもでき、1種または複数の特異的なサイズ範囲のポリヌクレオチドへと断片化することもできる。一部の実施形態では、断片の平均の長さは、約10~約10,000ヌクレオチドでありうる。一部の実施形態では、断片の平均の長さは、約50~約2,000ヌクレオチドでありうる。一部の実施形態では、断片の平均の長さは、約100~2,500、10~1,000、10~800、10~500、50~500、50~250、または50~150ヌクレオチドでありうる。一部の実施形態では、断片の平均の長さは、5,000ヌクレオチド未満、2,500ヌクレオチド未満、2,500ヌクレオチド未満、1,000ヌクレオチド未満など、10,000ヌクレオチド未満、400ヌクレオチド未満、300ヌクレオチド未満、200ヌクレオチド未満、または150ヌクレオチド未満など、500ヌクレオチド未満でありうる。

#### 【0090】

一部の実施形態では、核酸の断片化に続いて、核酸断片の末端修復を施すことができる。一部の実施形態では、断片化されていない試料に、末端修復を施すことができる。末端修復は、平滑末端、非平滑末端（すなわち、粘着末端または付着末端）、または3'-エクソヌクレアーゼ活性を欠くポリメラーゼによる、核酸断片の3'末端への、単一のdAヌクレオチドの付加など、一塩基突出の生成を含みうる。末端修復は、任意の数の酵素および/またはOvation（商標）UltraLow NGS Library System（NuGEN）など、市販のキットを含むがこれらに限定されない、当技術分野で公知の方法を使用して実施することができる。一部の実施形態では、末端修復を、二本鎖DNA断片上で実施して、平滑末端を生成することができ、この場合、二本鎖DNA断片は、5'ホスフェートおよび3'ヒドロキシルを含有する。一部の実施形態では、二本鎖DNA断片を、平滑末端ポリッシング（または「末端修復」）して、平滑末端を有するDNA断片を生成した後、アダプターへと接続することができる。二本鎖断片上の平滑末

端の生成は、二本鎖産物の突出する一本鎖末端を分解するための、例えば、エクソヌクレアーゼ1、エクソヌクレアーゼ7、またはこれらの組み合わせなど、一本鎖特異的DNAエクソヌクレアーゼの使用により行うことができる。代替的に、二本鎖DNA断片は、例えば、マングベーンエンドヌクレアーゼまたはS1エンドヌクレアーゼであるがこれらに限定されない、一本鎖特異的DNAエンドヌクレアーゼの使用により、平滑末端化することができる。代替的に、二本鎖産物は、二本鎖産物の突出する一本鎖末端を分解するための、例えば、T4 DNAポリメラーゼなど、一本鎖エクソヌクレアーゼ活性を含むポリメラーゼ、もしくは一本鎖エクソヌクレアーゼ活性を含む他の任意のポリメラーゼ、またはこれらの組み合わせの使用により平滑末端化することができる。場合によって、一本鎖エクソヌクレアーゼ活性を含むポリメラーゼは、1種または複数のdNTPを含むか、またはこれらを含まない、反応混合物中でインキュベートすることができる。他の場合には、一本鎖核酸特異的エクソヌクレアーゼと、1種または複数のポリメラーゼとの組み合わせを使用して、核酸を含む試料を断片化することにより生成される二本鎖断片を平滑末端化することができる。さらに他の場合には、二本鎖断片の突出する一本鎖末端に充填することにより、核酸断片を平滑末端化することもできる。例えば、断片を、1種または複数のdNTPの存在下で、T4 DNAポリメラーゼもしくはKlenowポリメラーゼ、またはこれらの組み合わせなどのポリメラーゼと共にインキュベートして、二本鎖断片の一本鎖部分を充填し得る。代替的に、二本鎖DNA断片は、エクソヌクレアーゼおよび/またはポリメラーゼを使用する一本鎖突出分解反応と、1種または複数のdNTPの存在下で、1種または複数のポリメラーゼを使用する充填反応との組み合わせにより、平滑にすることができる。平滑末端修復または末端ポリッシングのための市販のキットはまた、NEB Quick Blunting (商標) キットまたはNEB Next (登録商標) End Repair キット (New England Biolabs) も含む。

【0091】

一部の実施形態では、断片化された特定核酸を、一本鎖核酸断片へと変性させることができる。一部の実施形態では、非断片化試料を、一本鎖核酸鎖へと変性させることができる。当業者には、二本鎖核酸を一本鎖核酸へと変性させるための方法が周知である。方法は、熱変性、化学的変性などを含むがこれらに限定されない。

【0092】

特定核酸断片配列または非断片化核酸試料配列を定量化するための、本明細書で記載される方法は、少なくとも第1のアダプターを、本明細書で記載される方法により生成される、核酸断片または非断片化核酸試料配列へと追加するステップをさらに含む。一部の実施形態では、少なくとも第1のアダプターは、フォワードアダプターでありうる。少なくとも第1のアダプターを、本明細書で記載される方法により生成される、核酸断片または非断片化核酸試料配列へと追加するステップは、ライゲーション反応またはプライミング反応を使用して達成することができる。一部の実施形態では、少なくとも第1のアダプターの、核酸断片または非断片化核酸試料配列への追加は、ライゲーションを含む。一部の実施形態では、少なくとも第1のアダプターの、核酸断片または非断片化核酸試料配列へのライゲーションは、核酸断片または非断片化核酸試料配列の末端修復に続いて施すことができる。一部の実施形態では、少なくとも第1のアダプターの、核酸断片または非断片化核酸試料配列へのライゲーションは、核酸断片または非断片化核酸試料配列の末端修復を伴わない、核酸断片または非断片化核酸試料配列の生成に続いて施すことができる。

【0093】

少なくとも第1のアダプターは、当技術分野で公知の任意の種類のアダプターであって、アダプターが2つの相補鎖を含む、従来の二重鎖アダプターまたは二本鎖アダプターを含むがこれらに限定されないアダプターでありうる。一部の実施形態では、第1のアダプターは、二本鎖DNAアダプターでありうる。一部の実施形態では、第1のアダプターは、配列が公知のオリゴヌクレオチドであることが可能であり、こうして、少なくとも第1のアダプターを追加しうるか、または付着させうる、任意のポリヌクレオチドを増幅およ

び／またはシーケンシングするための、配列特異的プライマーの生成および／または使用を可能とする。一部の実施形態では、第1のアダプターは、従来の二重鎖アダプターであることが可能であり、この場合、第1のアダプターは、当技術分野で周知の配列を含む。一部の実施形態では、本明細書で記載される方法は、配列が公知の二本鎖DNAであって、平滑末端化することが可能であり、本明細書で記載される方法により、1つの配向性で生成される二本鎖核酸断片へとカップリングさせうる二本鎖DNAを含む、第1の二重鎖アダプターの使用を伴いうる。一部の実施形態では、核酸断片のライブラリー内の各核酸断片、または非断片化核酸のライブラリー内の非断片化核酸試料が、一方の末端へとライゲーションされた第1のアダプターを含むように、第1のアダプターを、本明細書で記載される方法により生成される核酸断片のライブラリーへと追加またはライゲーションすることができる。一部の実施形態では、少なくとも第1のアダプターを、一本鎖核酸断片または非断片化核酸試料配列へと、追加またはライゲーションすることができ、プローブ伸長産物へと組み込むことができる。

10

#### 【0094】

少なくとも第1のアダプターの、核酸断片または非断片化核酸試料配列へのライゲーションにより、ライゲーション産物である、第1のアダプターの特定核酸断片の複合体、または第1のアダプターの非断片化核酸試料の配列が生成される。一部の実施形態では、第1のアダプターの特定核酸断片の複合体を変性させることができる。一部の実施形態では、第1のアダプターの非断片化核酸試料の配列を変性させることができる。変性は、当技術分野で公知の方法であって、物理的変性、熱的変性、および／または化学的変性を含むがこれらに限定されない方法のうちのいずれかを使用して達成することができる。一部の実施形態では、変性は、熱的変性または熱変性を使用して達成することができる。一部の実施形態では、例えば、図1で描示される通り、少なくとも第1のアダプターの特定核酸断片の複合体、または少なくとも第1のアダプターの非断片化核酸試料の配列の変性により、少なくとも第1のアダプター配列を、核酸断片または非断片化核酸試料配列の5'末端だけにおいて含む、一本鎖核酸断片または非断片化核酸試料配列が生成される。

20

#### 【0095】

一部の実施形態では、5'末端または5'末端および3'末端の両方へと追加された第1のアダプター配列を含む、核酸断片または非断片化核酸試料配列を変性させて、5'末端または5'末端および3'末端の両方へと追加された第1のアダプター配列を含む、一本鎖核酸断片または非断片化核酸試料配列を生成する。一部の実施形態では、本明細書で記載される本発明の方法を使用して、5'末端または5'末端および3'末端の両方へと追加された第1のアダプター配列を含む、複数の一本鎖核酸断片または非断片化核酸試料配列を生成することができる。一部の実施形態では、第1の末端において、一本鎖特定核酸内に存在する、目的のプローブ標的領域配列と相補的な配列を含み、第2の末端において、第2のアダプターに由来する配列を含み、第2のアダプター配列が、プローブ標的領域と相補的でないオリゴヌクレオチドプローブを、一本鎖特定核酸断片または非断片化核酸試料配列とアニールさせることができる。一部の実施形態では、第2のアダプター配列は、リバースアダプターに由来する配列でありうる。

30

#### 【0096】

一部の実施形態では、目的のプローブ標的領域配列は、一本鎖特定核酸断片または非断片化核酸試料配列のうちの1種または複数の中に存在しうる。一部の実施形態では、目的とする異なるまたは別々のプローブ標的領域配列は、一本鎖核酸断片または非断片化核酸試料配列のうちの1種または複数の中に存在しうる。一部の実施形態では、1種または複数のオリゴヌクレオチドは、1種または複数の一本鎖核酸断片内または非断片化核酸試料配列内に存在する、目的とする同じ配列と相補的な配列を含みうる。この実施形態では、1種または複数のオリゴヌクレオチドは、目的とする同じ配列の異なる部分または領域と相補的でありうる配列を含みうる。一部の実施形態では、異なる領域は、互いと隣接する場合がある。一部の実施形態では、異なる領域は、互いと隣接しない場合がある。一部の実施形態では、目的とする同じ標的核酸配列と相補的な配列を含む、1種または複数のオ

40

50

リゴヌクレオチドは、同じ第2のアダプター配列をさらに含みうる。一部の実施形態では、1種または複数のプローブオリゴヌクレオチドは、1種または複数の一本鎖核酸断片内または非断片化核酸試料配列内に存在しうる、目的とする異なるまたは別々の配列と相補的な配列を含みうる。一部の実施形態では、目的とする異なるまたは別々の標的核酸配列と相補的な配列を含む、1種または複数のオリゴヌクレオチドプローブは、同じ第2のアダプター配列をさらに含みうる。一部の実施形態では、目的の標的配列と相補的な配列は、オリゴヌクレオチドプローブの3'末端にあることが可能であり、第2のアダプター配列は、オリゴヌクレオチドの5'末端にあることが可能である。一部の実施形態では、第2のアダプター配列は、目的の標的核酸配列と相補的でない場合がある。このようにして、第2のアダプター配列は、テールとして役立つ。第2のアダプター配列は、従来のアダプター配列でありうる。一部の実施形態では、第2のアダプター配列は、上記で記載した通り、一本鎖核酸断片または非断片化核酸試料配列へと追加される、第1のアダプターの配列と異なりうるか、または別々でありうる、従来のアダプター配列でありうる。一部の実施形態では、第2のアダプター配列は、公知の配列であることが可能であり、こうして、第2のアダプター配列を、追加しうるか、または付着させうる、任意のポリヌクレオチドの増幅および/またはシーケンシングのための、配列特異的プライマーの生成および/または使用を可能とする。別個の実施形態では、オリゴヌクレオチドプローブは、事前の変性を伴わずに、5'末端または5'末端および3'末端の両方へと追加された第1のアダプター配列を含む、特定核酸断片または非断片化核酸試料配列へとアニールさせることができる。この実施形態では、オリゴヌクレオチドのアニールリングは、オリゴヌクレオチドと、二本鎖核酸断片または非断片化核酸試料配列の5'末端または5'末端および3'末端の両方へと追加された第1のアダプター配列を含む、二本鎖核酸断片または非断片化核酸試料配列との間の、三重螺旋または三重鎖の形成を介するアニールリングでありうる。この実施形態では、二本鎖核酸断片または非断片化核酸試料配列は、目的の配列を含み、5'末端または5'末端および3'末端の両方へと追加された第1のアダプター配列を含む、複数の二本鎖核酸断片または非断片化核酸試料配列の中に存在しうる。この実施形態に付け加えて、オリゴヌクレオチドプローブは、二本鎖特定核酸断片内または非断片化核酸試料配列内の、プローブ標的領域と相補的な配列も含む。総じて、1種または複数の (more or a plurality of) 特定核酸断片または非断片化核酸試料配列の中の核酸断片内または非断片化核酸試料配列内に存在する、目的のプローブ標的領域配列と相補的な配列を含むオリゴヌクレオチドプローブの使用は、本明細書で記載される方法を使用する、前記核酸断片または非断片化核酸試料配列の選択的結合および後続の濃縮を可能とする。

【0097】

上記で記載したオリゴヌクレオチドプローブのアニールリングに続き、ポリメラーゼを使用して、オリゴヌクレオチドプローブを伸長させることができる。一部の実施形態では、ポリメラーゼは、DNA依存性DNAポリメラーゼでありうる。一部の実施形態では、DNA依存性DNAポリメラーゼは、本明細書で記載されるDNA依存性DNAポリメラーゼのうちのいずれかであることが可能であり、オリゴヌクレオチドの伸長は、当技術分野で公知の方法のうちのいずれかを介する伸長でありうる。一部の実施形態では、プローブ標的領域核酸と相補的でない第2のアダプター配列と、特定核酸断片内に存在する、目的のプローブ標的領域配列と相補的な配列であって、一方および/または両方の末端へと追加された第1のアダプターを含む配列とを含むオリゴヌクレオチドプローブを、核酸断片へとアニールさせ、ポリメラーゼにより伸長させて、第1の末端において、第1のアダプター配列を含み、第2の末端において、第2のアダプター配列を含む、プローブ伸長産物を生成することができる。一部の実施形態では、特定核酸断片は、一方および/または両方の末端へと追加された第1のアダプターを含む、複数の核酸断片の中に存在しうる。この実施形態では、プローブ伸長産物を、目的のプローブ標的領域配列を含有する核酸断片だけについて生成することができる。

【0098】

10

20

30

40

50

一部の実施形態では、本明細書で記載される方法により生成されるプローブ伸長産物を、増幅反応にかけることができる。一部の実施形態では、増幅反応は、指数関数的であることが可能であり、多様な温度サイクルで実行することができる。増幅反応は、等温反応でありうる。一部の実施形態では、増幅は、定量的ポリメラーゼ連鎖反応（qPCR）でありうる。一部の実施形態では、増幅反応は、等温でありうる。一部の実施形態では、プローブ伸長産物は、本明細書で記載される方法により生成される通り、一方の末端上に、少なくとも第1のアダプター配列を含み、他方の末端上に、第2のアダプター配列を含む。一部の実施形態では、プローブ伸長産物は、第1のアダプターと相補的な配列を含む第1のプライマーと、特定核酸鎖内の、プローブ標的領域と相補的な鎖内の、5'テール配列と相補的な配列を有する第2のプライマーとを使用して増幅することができる。このようにして、第1のアダプター配列およびプローブ標的領域の両方を含むプローブ伸長産物を増幅し、そのようにして濃縮することができる。少なくとも第1のアダプター配列およびプローブ標的領域配列の両方を有するプローブ伸長産物を増幅し、増幅されたプローブ伸長産物であって、前記ライゲーションされた特定核酸断片または非断片化核酸試料配列から生成されたプローブ伸長産物を定量化することができる。一部の実施形態では、少なくとも第1のアダプター配列および/または第2のアダプター配列は、識別子配列を含みうる。一部の実施形態では、識別子配列は、バーコード配列でありうる。一部の実施形態では、バーコード配列は、少なくとも第1のアダプターに固有でありうる。一部の実施形態では、少なくとも第1のアダプターおよび/または第2のアダプター配列は、例えば、シーケンシングおよびシーケンシング反応の後における特定核酸の同定などであるがこれらに限定されない、下流における適用のために使用しうる配列を含みうる。一部の実施形態では、少なくとも第1のアダプターおよび/または第2のアダプター配列は、Illuminaにより開発され、本明細書で記載されるシーケンシング法によるシーケンシングのために使用しうるフローセル配列33および35（図5）を含みうる。

#### 【0099】

目的の特定核酸配列断片を定量化するための、本明細書で記載される方法の、開示される実施形態についての概略図を、図1および図2に例示する。図中で使用される番号付けスキームは、例示的なものであるにすぎない。1つを超える図中に現れる同じ番号は、全部または一部において同一なオリゴヌクレオチド配列を指し示すことを意図するものではなく、開示される方法を実施するための基準となる構成要素、部位、または領域を指し示すことを意図するものである。

#### 【0100】

図1および図2の方法は、核酸断片、非断片化核酸試料、またはインサートの、ライゲーションされたライブラリーの生成を例示するが、この場合、ライゲーションされたライブラリーの各核酸配列は、フォワードリードプライミング部位と相補的なプライマーと、プローブ標的領域を含むプローブ伸長産物内の、リバースリードプライミング部位と相補的なプライマーとを使用するPCR増幅により、特異的なプローブ標的領域配列を有する特定核酸分子の定量化を可能とするシーケンシングカバレッジ（sequencing coverage）をもたらすように、アダプター内の、共通のフォワードリードプライミング部位と、特異的なプローブ標的領域配列とを含む。

#### 【0101】

図1は、せん断されたgDNAの使用について例示する。せん断されたDNA 8は、特定核酸断片10を有するgDNAの5'末端へとライゲーションされたアダプター11を有する。断片10は、プローブ標的領域50を含む。アダプターは、シーケンシングリード1のフォワードオリゴヌクレオチドプライミング部位12、6Nオリゴヌクレオチド配列14などのn-ランダムオリゴヌクレオチド塩基、インデックス塩基オリゴヌクレオチド配列16、および、使用されるハイスループットシーケンシング法に応じて、インデックスプライミング部位18のうちの少なくとも1つを含みうる。アダプター11をライゲーションすると、特定核酸断片10は、固有の識別子配列標識である、インデックスリードに加えた、n-ランダムオリゴヌクレオチドを有しうる。インデックス配列16

を使用して、特定核酸試料を同定し、6 N オリゴヌクレオチド配列 14 を、重複シーケンシングリードのマーキングに使用する。5' テールオリゴヌクレオチド配列 20 を有するプローブオリゴヌクレオチド配列 19 は、プローブ標的領域 50 と相補的であり、かつ、これとハイブリダイズすることが可能であり、アダプター 11 を介する、d N T P および D N A ポリメラーゼの存在下における、単一プライマー伸長反応により伸長させることができる。結果として得られるプローブ伸長産物 22 は、インデックスプライミング部位 18 と部分的に相補的でありうるフォワードプライマー 24 と、5' テール配列 20 の逆相補体と部分的に相補的でありうるリバースプライマー 26 とを使用して増幅することができる。増幅反応により、プローブ標的領域 50 を有する特定核酸 10 の存在を濃縮して、特定核酸配列のライブラリーを生成する。

10

#### 【0102】

図2に例示される通り、同様の単一プライマー伸長反応は、c D N A にも適用可能でありうる。c D N A 7 は、特定核酸断片9の5' 末端へとライゲーションされたアダプター 11 を有する。断片9は、プローブ標的領域60を含む。アダプターは、フォワードシーケンシングリードのオリゴヌクレオチドプライミング部位12、6 N オリゴヌクレオチド配列14など、公知のランダムオリゴヌクレオチド塩基、インデックス塩基オリゴヌクレオチド配列16、および、使用されるハイスループットシーケンシング法に応じて、インデックスシーケンシングリードのプライミング部位18のうちの少なくとも1つを含みうる。インデックス配列16を使用して、特定核酸試料を同定し、6 N オリゴヌクレオチド配列14を、重複シーケンシングリードの同定に使用する。5' テールオリゴヌクレオチド配列20を有するプローブオリゴヌクレオチド配列19は、プローブ標的領域配列60と相補的であり、かつ、これとハイブリダイズすることが可能であり、アダプター 15 を介する、d N T P および D N A ポリメラーゼの存在下における、単一プライマー伸長反応により伸長させることができる。結果として得られるプローブ伸長産物 21 は、18 と部分的に相補的でありうるフォワードプライマー 24 と、5' テール配列 20 の逆相補体と部分的に相補的でありうるリバースプライマー 26 とを使用して増幅することができる。増幅反応により、プローブ標的領域 60 を有する特定核酸 9 の存在を濃縮して、特定核酸配列のライブラリーを生成する。

20

#### 【0103】

目的の特定核酸配列断片を定量化するための、本明細書で記載される方法の、開示される実施形態についての概略図を、二本鎖 g D N A について図3および図4に例示する。図中で使用される番号付けスキームは、例示的なものであるにすぎない。1つを超える図中に現れる同じ番号は、全部または一部において同一なオリゴヌクレオチド配列を指し示すことを意図するものではなく、開示される方法を実施するための基準となる構成要素、部位、または領域を指し示すことを意図するものである。

30

#### 【0104】

図3および図4の方法は、核酸断片、非断片化核酸試料、またはインサートの、シーケンシングライブラリーの生成を例示するが、この場合、シーケンシングライブラリーの各核酸配列は、特異的なプローブ標的領域配列を有する特定核酸分子の定量化を可能とするシーケンシングカバレッジが存在し得るように、1つのアダプター内の、共通のフォワードプライミング部位と、特異的なプローブ標的領域配列とを含む。シーケンシングは、共通のフォワードプライミング部位と相補的なプライマーと、特定核酸配列内の特異的なプローブ標的領域配列と相補的なプライマーとを使用する、P C R 増幅を伴うか、またはこれを伴わない、ライゲーションされたプローブ伸長産物から作製されるシーケンシングライブラリーを使用して行うことができる。

40

#### 【0105】

図3は、せん断された g D N A の使用を例示する。特定核酸 10 を有する、せん断された g D N A は、プローブ標的領域 50 を含む。5' テールオリゴヌクレオチド配列 20 を有する、プローブオリゴヌクレオチド配列 19 は、プローブ標的領域配列 50 と相補的であり、かつ、これとハイブリダイズすることが可能であり、特定核酸 10 の末端を介する

50

、dNTPおよびDNAポリメラーゼの存在下における、単一プローブ伸長反応であって、二本鎖DNAを創出する伸長反応において伸長させることができる。結果として得られるプローブ伸長産物は、特定核酸断片10の3'末端へとライゲーションされたアダプターを有しうる。アダプターは、フォワードシークエンシングリード1のオリゴヌクレオチドプライミング部位12、6Nオリゴヌクレオチド配列14などのn-ランダムオリゴヌクレオチド塩基、インデックス塩基オリゴヌクレオチド配列16、および、使用されるハイスループットシークエンシング法に応じて、インデックスプライミング部位18のうちの少なくとも1つを含みうる。インデックス配列16を使用して、特定核酸試料を同定し、6Nオリゴヌクレオチド配列14を、重複シークエンシングリードのマーキングに使用する。ライゲーション産物22は、インデックスプライミング部位18と部分的に相補的でありうるフォワードプライマー24と、5'テール配列21の逆相補体と部分的に相補的でありうるリバースプライマー26とを使用して増幅することができる。増幅反応により、プローブ標的領域50を有する特定核酸10の存在を濃縮して、特定核酸配列のライブラリーを生成することができる。

#### 【0106】

図4は、せん断されたgDNAの使用を例示する。特定核酸10を有する、せん断されたgDNAは、プローブ標的領域50を含む。5'テールオリゴヌクレオチド配列20を有する、プローブオリゴヌクレオチド配列19は、プローブ標的領域配列50と相補的であり、かつ、これとハイブリダイズすることが可能であり、gDNA10の末端を介する、dNTPおよびDNAポリメラーゼの存在下における、単一プローブ伸長反応であって、二本鎖DNAを創出する伸長反応において伸長させることができる。結果として得られるプローブ伸長産物は、制限酵素70により消化することができる。例示的な制限酵素は、XbaI、EcoRI、EcoRV、およびBamHIを含むがこれらに限定されない。制限酵素による消化に続き、アダプターを、特定核酸断片10を有する二本鎖gDNAの末端へとライゲーションすることができる。アダプターは、リード1のフォワードオリゴヌクレオチドプライミング部位12、6Nオリゴヌクレオチド配列14などのn-ランダムオリゴヌクレオチド塩基、インデックス塩基オリゴヌクレオチド配列16、および、使用されるハイスループットシークエンシング法に応じて、インデックスプライミング部位18のうちの少なくとも1つを含みうる。インデックス配列16を使用して、特定核酸試料を同定し、6Nオリゴヌクレオチド配列14を、重複シークエンシングリードのマーキングに使用する。ライゲーション産物22は、図3に例示される通り、インデックスプライミング部位18と部分的に相補的でありうるフォワードプライマー24と、5'テール配列21と部分的に相補的でありうるリバースプライマー26とを使用して増幅することができる。増幅反応により、プローブ標的領域50を有する特定核酸10の存在を濃縮して、特定核酸配列のライブラリーを生成することができる。

#### 【0107】

図5（番号付けは、図1または図2で使用される番号付けを指す）に例示される通り、同様の単一プライマー伸長反応は、様々なシークエンシングプラットフォームのためのシークエンシングライブラリーを創出するのに、gDNAまたはcDNAへと適用可能でありうる。gDNAまたはcDNA（せん断されているか、またはせん断されていない）10または9は、特定核酸断片10または9の5'末端へとライゲーションされたアダプター11を有する。断片10または9は、それぞれ、プローブ標的領域50または60を含む。アダプターは、フォワードオリゴヌクレオチドプライミング部位12、6Nオリゴヌクレオチド配列14など、公知のランダムオリゴヌクレオチド塩基、インデックス塩基オリゴヌクレオチド配列16、および、使用されるハイスループットシークエンシング法に応じて、インデックスプライミング部位18のうちの少なくとも1つを含みうる。5'テールオリゴヌクレオチド配列20を有するプローブオリゴヌクレオチド配列19は、プローブ標的領域配列50または60と相補的であり、かつ、これとハイブリダイズすることが可能であり、アダプター11を介する、dNTPおよびDNAポリメラーゼの存在下における、単一プライマー伸長反応により伸長させることができる。結果として得られるブ

プローブ伸長産物 21 または 22 は、18 と部分的に相補的でありうるフォワードプライマー 24 と、5' テール配列 20 と部分的に相補的でありうるリバースプライマー 26 とを使用して増幅することができる。増幅反応により、プローブ標的領域 50 または 60 を有する特定核酸 10 または 9 の存在を濃縮して、特定核酸配列のライブラリーを生成することができる。

#### 【0108】

ライブラリーは、NuGEN Ovation (登録商標) Target Enrichment System (NuGEN) を使用して、PCR を介して、選択された目的のプローブ標的領域配列を有する、プローブ伸長産物配列を選択的に増幅することにより調製することができる。図 5 は、Illumina ハイスループットシーケンシングプラットフォームを使用する場合に、ハイスループットシーケンシングで使用される核酸ライブラリーの例を例示する。例えば、遺伝子発現またはコピー数変異の定量化のデジタル測定のために、各配列ライブラリーの特定配列リード領域を解析することができる。

#### 【0109】

一部の実施形態では、特定核酸は、ビオチンを含むがこれらに限定されない、インジケーター分子によりタグ付けすることができる。次いで、タグ付けされた特定核酸分子を、元の試料分子に由来するものとして識別することができる。一部の実施形態では、インジケーター分子の付着は、標識されたヌクレオチド、例えば、ビオチニル化されたヌクレオチドの、ライゲーションまたはポリメラーゼによる付加を介して達することができる。次いで、ポリメラーゼによるプローブの伸長を伴うか、またはこれを伴わずに、プローブ標的領域と相補的でありうるプローブを、タグ付けされた核酸とハイブリダイズさせることができる。一部の実施形態では、ハイブリダイズしないプローブは、例えば、ビオチン/ストレプトアビジン間相互作用を介して、タグ付けされた核酸を捕捉することにより除去する。一部の実施形態では、標的とハイブリダイズしたプローブを、標的と共に捕捉する。ハイブリダイズしないプローブの除去に続き、捕捉されたプローブを、標的核酸から溶出させ、計数する。一部の実施形態では、計数は、Illumina プラットフォームを介するシーケンシング、およびこれらのタグの計数により行うことができる。一部の実施形態では、プローブは、当業者に公知の通り、ナノ小孔または蛍光タグ付けによりタグ付けすることができる。

#### 【0110】

##### インプット核酸

インプットは、ヒト核酸でありうる。一部の実施形態では、インプットは、DNA でありうる。一部の実施形態では、インプットヒト核酸は、二本鎖 DNA、ゲノム DNA、または 1 つを超える生物に由来する混合型 DNA など、複合体 DNA でありうる。一部の実施形態では、インプットは、RNA でありうる。一部の実施形態では、RNA は、当技術分野で標準的な技法を使用して得、精製することができ、mRNA、tRNA、snRNA、rRNA、低分子非コード RNA、マイクロ RNA、ポリソーム RNA、プレ mRNA、イントロン RNA、無細胞 RNA、およびこれらの断片を含みうるがこれらに限定されない、精製形態または非精製形態にある RNA を含みうる。非コード RNA または ncRNA は、snRNA、マイクロ RNA、siRNA、piRNA、および長鎖 ncRNA を含みうる。一部の実施形態では、DNA 断片は、RNA を、プライマー（すなわち、ランダムプライマー）と組み合わせること、および、RNA 鋳型を、RNA 依存性 DNA ポリメラーゼを用いて逆転写することを含みうるがこれらに限定されない、RNA 鋳型から、cDNA を生成するための、当技術分野で周知の方法のうちのいずれかを使用する、第 1 の鎖合成反応を介して、cDNA へと転換された RNA に由来しうる。一部の実施形態では、DNA 断片は、当技術分野で周知の方法のうちのいずれかを使用する、第 1 の鎖合成反応および第 2 の鎖合成反応を介して、二本鎖 cDNA へと転換された RNA に由来しうる。

#### 【0111】

一部の実施形態では、インプット DNA は、異なる種のゲノムの混合物から作製される



cDNAでありうる。インプット複合体はまた、異なるヒトのゲノムの混合物にも由来しうる。インプットDNAは、異なるヒトのゲノムの混合物から作製されるcDNAでありうる。インプットDNAは、特定の種、例えば、ヒト、ラット、マウス、他の動物、特定の植物、細菌、藻類、ウイルスなどのDNAでありうる。インプット複合体はまた、宿主病原体、細菌集団など、異なる種のゲノムの混合物にも由来しうる。代替的に、インプット核酸は、合成供給源に由来しうる。インプットDNAは、ミトコンドリアDNAでありうる。インプットDNAは、無細胞DNAでありうる。無細胞DNAは、例えば、血清試料または血漿試料から得ることができる。インプットDNAは、1種または複数の染色体を含みうる。例えば、インプットDNAが、ヒトに由来する場合、DNAは、第1染色体、第2染色体、第3染色体、第4染色体、第5染色体、第6染色体、第7染色体、第8染色体、第9染色体、第10染色体、第11染色体、第12染色体、第13染色体、第14染色体、第15染色体、第16染色体、第17染色体、第18染色体、第19染色体、第20染色体、第21染色体、第22染色体、X染色体、またはY染色体のうちの1または複数を含みうる。DNAは、直鎖状ゲノムに由来する場合もあり、環状ゲノムに由来する場合もある。DNAは、プラスミドDNA、コスミドDNA、細菌の人工染色体(BAC)、または酵母人工染色体(YAC)でありうる。インプットDNAは、1例を超える個別のヒトに由来しうる。インプットDNAは、二本鎖の場合もあり、一本鎖の場合もある。インプットDNAは、クロマチンの部分でありうる。インプットDNAは、ヒストンと会合している場合がある。

10

#### 【0112】

20

一部の実施形態では、プローブオリゴヌクレオチドを、目的の特定核酸配列へと方向づけることができ、特定核酸内のプローブ標的領域を有する一本鎖特定核酸標的とハイブリダイズするように設計することができる。一部の実施形態では、選択された目的の配列領域を標的化するプローブを、一本鎖DNAまたはcDNAプローブ標的領域とハイブリダイズするように設計することができる。インプット核酸試料が、ゲノムDNAまたは他の二本鎖DNAを含む場合、インプット核酸試料をまず変性させて、標的を一本鎖とし、オリゴヌクレオチドプローブの、所望される目的のプローブ標的領域である配列領域とのハイブリダイゼーションを可能とすることができる。一部の実施形態では、他の二本鎖DNAは、1種または複数の標的RNAの第1の鎖合成および第2の鎖合成により生成される二本鎖cDNAでありうる。これらの実施形態では、本明細書で記載される方法および組成物により、複数のプローブ標的領域を含有する、複数の目的の特定核酸配列領域の、領域特異的濃縮および増幅を可能とすることができる。一部の実施形態では、本明細書で記載される方法および組成物は、各々が、対応する別々のプローブ標的領域を含有する目的の別々の領域を有する、少なくとも2つまたはこれを超える別々の特定核酸配列断片または非断片化核酸試料配列の、マルチプレックス増幅、濃縮、および定量化を可能とする。

30

#### 【0113】

他の実施形態では、選択された目的の配列領域を標的化するプローブを、二本鎖核酸断片または非断片化核酸試料配列の変性を伴わずに、二本鎖核酸標的断片または非断片化核酸試料配列とハイブリダイズするように設計することができる。他の実施形態では、選択された目的の配列領域を標的化するプローブを、dsDNAの変性を伴わずに、二本鎖DNA標的とハイブリダイズするように設計することができる。これらの実施形態では、選択された目的の配列領域を標的化するプローブを、選択された目的の配列領域において、三重螺旋(三重鎖)を形成するように設計することができる。プローブの、目的の二本鎖DNA配列領域とのハイブリダイゼーションは、二本鎖核酸試料の事前の変性を伴わずに実行することができる。このような実施形態では、本明細書で記載される方法および組成物により、目的の配列領域の領域特異的定量化のほか、鎖特異的増幅および定量化も可能とすることができる。この方法は、dsDNAであるインプットDNAを変性させる必要なしに、複合体核酸からの、目的の鎖特異的配列領域のコピーを生成し、こうして、天然の複合体核酸試料中の、多数の目的の配列領域の定量化および解析を可能とするために有用でありうる。方法は、*in situ*において実行される研究および解析のために使用

40

50

することが可能であり、単一の細胞、または極めて小規模な、十分に規定された細胞集団のコレクション中の複合体ゲノムDNAについての研究および解析を可能としうるほか、クロマチン構造の破壊を伴わない、複合体ゲノムDNAの解析も許容しうる。

#### 【0114】

一部の実施形態では、本明細書では、さらなる識別子配列、例えば、バーコード配列を含むアダプターが開示される。一部の実施形態では、少なくとも第1のアダプターは、複数のバーコード配列のうちの少なくとも1つを含む。一部の実施形態では、各リバースアダプターは、複数のバーコード配列のうちの少なくとも1つを含み、この場合、複数のバーコード配列のうちの、各バーコード配列は、複数のバーコード配列中の他のあらゆるバーコード配列と異なる。一部の実施形態では、第2のアダプターのオリゴヌクレオチドについてのバーコードを、少なくとも第1のアダプターのオリゴヌクレオチドについてのバーコードと独立に選択することができる。一部の実施形態では、対をなすアダプターが、同じであるか、または異なる1種または複数のバーコードを含むように、バーコードを有する、第1のアダプターのオリゴヌクレオチドと、第2のアダプターのオリゴヌクレオチドとを対合させることができる。一部の実施形態では、本発明の方法は、標的ポリヌクレオチドが由来する試料を、標的ポリヌクレオチドが接続されるバーコード配列に基づき同定するステップをさらに含みうる。バーコードは、例えば、標的ポリヌクレオチドへと接続されると、標的ポリヌクレオチドが由来した試料の識別子として役立つ核酸配列を含みうる。

#### 【0115】

目的のプロブ標的領域である配列領域/鎖の増幅準備のできた産物の生成に適しうる、多様なアダプター設計を想定することができる。一部の実施形態では、少なくとも第1のアダプターは、一本鎖の場合もあり、二本鎖の場合もある。例えば、二本鎖の場合、アダプターの2つの鎖は、自己相補的な場合もあり、非相補的な場合もあり、部分的に相補的な場合もある。近年、アダプター設計において、多くの改善がなされており、これにより、アダプター二量体の発生が低減されている。これらの改善は、ヌクレオチド類似体および構造化されたオリゴヌクレオチドの使用を含む場合があり、ライゲーション反応における高濃度のオリゴヌクレオチドの使用を可能としている。ライゲーション反応における高濃度のアダプターにより、研究者は、150コピーという少数のゲノムコピーから高品質のライブラリーを生成することが可能となっている。アダプターの、DNA断片、特に、目的の領域を含有する断片の末端へのライゲーションは、本発明の方法を実行するのに適しうる。核酸修飾酵素および結果として得られる二本鎖DNAの切断の選択に応じて、多様なライゲーションモダリティを想定することができる。例えば、目的の標的領域/配列を含む平滑末端産物を生成する場合は、平滑末端ライゲーションが適しうる。代替的に、公知の配列特異性の制限酵素を使用して、切断を実行し、公知の配列突出を伴う切断部位の生成をもたらしうる場合は、目的の配列領域の切断部位との、アダプターのハイブリダイゼーションと、後続のライゲーションとを可能とするように、アダプターの適切な末端を設計することができる。アダプターの効率的かつ迅速なライゲーションのための試薬および方法は、市販されており、当技術分野でも公知である。

#### 【0116】

##### 核酸修飾酵素

核酸(NA)修飾酵素は、DNA特異的修飾酵素でありうる。NA修飾酵素は、二本鎖DNAに対する特異性について選択することができる。酵素は、二重鎖特異的エンドヌクレアーゼ、平滑末端頻発カッター制限酵素または他の制限酵素であり得る。平滑末端カッターの例は、DraIまたはSmaIを含みうる。NA修飾酵素は、New England Biolabsにより提供されている酵素でありうる。NA修飾酵素は、ホーミングエンドヌクレアーゼ(ホーミングエンドヌクレアーゼとは、厳密に規定された認識配列を有さないエンドヌクレアーゼでありうる)でありうる。NA修飾酵素は、ニッキ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とは、二本鎖DNA基質内のDNAの1つの鎖だけを切断しうる、エンドヌクレアーゼでありうる)でありうる。NA修飾酵素

は、高忠実度エンドヌクレアーゼ（高忠実度エンドヌクレアーゼとは、エンドヌクレアーゼの野生型バージョンより「スター活性」が小さくなるように操作された、エンドヌクレアーゼでありうる）でありうる。

【0117】

DNA依存性DNAポリメラーゼ

本発明の方法および組成物における使用のためのDNA依存性DNAポリメラーゼは、本発明の方法に従うプローブ標的領域またはプライマーの伸長をもたらすことが可能でありうる。一部の実施形態では、DNA依存性DNAポリメラーゼは、DNA鋳型および/またはcDNA鋳型の存在下で、プローブ標的領域、核酸プライマーなどを伸長させることが可能なDNAポリメラーゼでありうる。本発明の方法に適する、例示的なDNA依存性DNAポリメラーゼは、3'-エクソヌクレアーゼを伴うか、またはこれを伴わない、Klenowポリメラーゼ、Bst DNAポリメラーゼ、Bsuポリメラーゼ、phi 29 DNAポリメラーゼ、Ventポリメラーゼ、Deep Ventポリメラーゼ、Taqポリメラーゼ、T4ポリメラーゼ、およびE. coli DNAポリメラーゼ1、これらの誘導体、またはポリメラーゼの混合物を含むがこれらに限定されない。場合によって、ポリメラーゼは、5'-エクソヌクレアーゼ活性を含まない。他の場合において、ポリメラーゼは、5'エクソヌクレアーゼ活性を含む。場合によって、本発明のプライマー伸長産物またはオリゴヌクレオチド伸長産物は、例えば、Bstポリメラーゼなど、強い鎖置換活性を含むポリメラーゼを使用して実施することができる。他の場合において、本発明のプライマー伸長は、弱い鎖置換活性を含むか、または鎖置換活性を含まないポリメラーゼを使用して実施することができる。当業者は、プライマー伸長ステップにおける、鎖置換活性の使用の利点および欠点、ならびに、どのポリメラーゼが、鎖置換活性をもたらすと予測されうるのか（例えば、New England Biolabs、「Polymerses」を参照されたい）を認識することができる。

【0118】

増幅法

本明細書で記載される方法、組成物、およびキットは、大規模並行シーケンシング（すなわち、次世代シーケンシング法）、目的の配列領域集団を濃縮したライブラリーの生成、またはハイブリダイゼーションプラットフォームなど、下流における適用のための、増幅準備のできた産物を生成するのに有用でありうる。当技術分野では、増幅法が周知である。適切な増幅反応は、指数関数的増幅の場合もあり、等温増幅の場合もあり、ポリメラーゼ連鎖反応（PCR）、鎖置換増幅（SDA）、線形増幅、多重置換増幅（MDA）、ローリングサークル増幅（RCA）、単一プライマー等温増幅（SPIA；例えば、米国特許第6,251,639号を参照されたい）、Ribo-SPIA、またはこれらの組み合わせを含むがこれらに限定されない、任意のDNA増幅反応を含みうる。場合によって、鋳型核酸をもたらすための増幅法は、例えば、cDNAの生成のために一般になされうる増幅など、少数ラウンド（例えば、1、2、3、4、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、30ラウンドなど）の増幅だけが実施されうるように、限定的な条件下で実施することができる。増幅のラウンド数は、約1~30、1~20、1~15、1~10、5~30、10~30、15~30、20~30、10~30、15~30、20~30、または25~30でありうる。

【0119】

PCRとは、好熱性の鋳型依存性ポリヌクレオチドポリメラーゼによる、変性、オリゴヌクレオチドプライマーのアニール、およびプライマーの伸長によるサイクルの反復に基づく、in vitroにおける増幅手順であって、プライマーに挟まれたポリヌクレオチド分析物の所望の配列のコピーの指数関数的増大を結果としてもたらす増幅手順である。DNAの対向する鎖へとアニールする、2つの異なるPCRプライマーは、ポリメラーゼに触媒される、一方のプライマーの伸長産物が、他方のプライマーのための鋳型鎖として役立ち得、その長さが、オリゴヌクレオチドプライマーの5'末端の間の距離によ

り規定されうる、個別の二本鎖断片の蓄積をもたらすように配置することができる。さらなる増幅法については、参照によりその全体において本明細書に組み込まれる、2013年1月25日に出願されたUS 2013/0175076において、さらに記載されている。

#### 【0120】

一部の実施形態では、増幅は、例えば、ポリメラーゼ連鎖反応(PCR)による、DNAの特定二本鎖配列の酵素的増幅における、指数関数的増幅でありうる。他の実施形態では、増幅法は、線形増幅でありうる。他の実施形態では、増幅法は、等温増幅でありうる。

#### 【0121】

下流における適用

本発明の一態様は、本明細書で開示される方法および組成物を、次世代シーケンシングまたはハイブリダイゼーションプラットフォームなど、下流における解析のために、目的の生体材料の喪失を最小限として、効率的かつ費用効果的に活用しうることである。本明細書で開示される方法はまた、目的の選択的ゲノム領域についての遺伝子情報の解析(例えば、SNP、コピー数変異、または他の疾患マーカーの解析)においても使用しうるほか、トランスクリプトーム解析、および目的の選択的領域と相互作用しうるゲノム領域による、デジタル遺伝子発現においても使用することができる。

#### 【0122】

シーケンシング

例えば、本発明の方法は、米国特許第5,750,341号;同第6,306,597号;および同第5,969,119号において記載されている通り、Illuminaにより市販されている方法によるシーケンシングに有用でありうる。一般に、二本鎖断片ポリヌクレオチドは、一方の末端(例えば、(A)/(A'))または両方の末端(例えば、(A)/(A')および(C)/(C'))においてタグ付けされた、増幅された核酸配列を生成するように、本発明の方法により調製することができる。場合によって、本発明の方法により(例えば、SPIAまたは線形PCRにより)、一方の末端または両方の末端においてタグ付けされた一本鎖核酸を増幅することもできる。次いで、結果として得られる核酸を変性させることができ、増幅された一本鎖ポリヌクレオチドを、フローセルチャネルの内部表面へとランダムに付着させることができる。標識されていないヌクレオチドを添加して、二本鎖DNAの稠密なクラスターを生成する、固相架橋増幅を開始することができる。第1塩基のシーケンシングサイクルを開始するために、4つの、標識された可逆性ターミネーター、プライマー、およびDNAポリメラーゼを添加することができる。レーザーによる励起の後、フローセル上の各クラスターからの蛍光をイメージングすることができる。次いで、各クラスターについて、第1塩基の正体を記録することができる。シーケンシングのサイクルは、断片の配列を一度に1塩基ずつ決定するように実施することができる。

#### 【0123】

一部の実施形態では、本発明の方法は、Applied Biosystemsにより市販されているライゲーションによるシーケンシング法(例えば、SOLiDシーケンシング)を介するシーケンシングのための標的ポリヌクレオチドを調製するのに有用でありうる。他の実施形態では、方法は、454/Roche Life Sciencesにより市販されている方法であって、Marguliesら、Nature(2005年)、437巻:376~380頁(2005年);ならびに米国特許第7,244,559号;同第7,335,762号;同第7,211,390号;同第7,244,567号;同第7,264,929号;および同第7,323,305号において記載されている方法および装置を含むがこれらに限定されない方法を使用する合成を介するシーケンシングのための標的ポリヌクレオチドを調製するのに有用でありうる。他の実施形態では、方法は、Helicos BioSciences Corporation(Cambridge, Mass.)により市販されている方法であって、米国出願第11

10

20

30

40

50

/ 1 6 7 , 0 4 6 号、および米国特許第 7 , 5 0 1 , 2 4 5 号；同第 7 , 4 9 1 , 4 9 8 号；同第 7 , 2 7 6 , 7 2 0 号；ならびに米国特許出願公開第 US 2 0 0 9 0 0 6 1 4 3 9 号；同第 US 2 0 0 8 0 0 8 7 8 2 6 号；同第 US 2 0 0 6 0 2 8 6 5 6 6 号；同第 US 2 0 0 6 0 0 2 4 7 1 1 号；同第 US 2 0 0 6 0 0 2 4 6 7 8 号；同第 US 2 0 0 8 0 2 1 3 7 7 0 号；および同第 US 2 0 0 8 0 1 0 3 0 5 8 号において記載されている方法を介するシークエンシングのための標的ポリヌクレオチドを調製するのに有用でありうる。他の実施形態では、方法は、Pacific Biosciences により市販されている方法であって、米国特許第 7 , 4 6 2 , 4 5 2 号；同第 7 , 4 7 6 , 5 0 4 号；同第 7 , 4 0 5 , 2 8 1 号；同第 7 , 1 7 0 , 0 5 0 号；同第 7 , 4 6 2 , 4 6 8 号；同第 7 , 4 7 6 , 5 0 3 号；同第 7 , 3 1 5 , 0 1 9 号；同第 7 , 3 0 2 , 1 4 6 号；同第 7 , 3 1 3 , 3 0 8 号；および米国出願公開第 US 2 0 0 9 0 0 2 9 3 8 5 号；同第 US 2 0 0 9 0 0 6 8 6 5 5 号；同第 US 2 0 0 9 0 0 2 4 3 3 1 号；および同第 US 2 0 0 8 0 2 0 6 7 6 4 号において記載されている方法を介するシークエンシングのための標的ポリヌクレオチドを調製するのに有用でありうる。

#### 【 0 1 2 4 】

本明細書で提示される発明の方法において使用されうるシークエンシング法の別の例は、Ion Torrent により提供されている半導体シークエンシング（例えば、Ion Personal Genome Machine (PGM) を使用する）である。Ion Torrent による技術では、複数の層、例えば、マイクロ加工されたウェルを伴う層、イオン感受性層、およびイオンセンサー層を伴う半導体チップを使用しうる。核酸を、ウェルへと導入することができる、例えば、単一の核酸のクローン集団を、単一のビーズへと付着させ、ビーズを、ウェルへと導入することができる。ビーズ上の核酸のシークエンシングを開始するために、1 種類のデオキシリボヌクレオチド（例えば、dATP、dCTP、dGTP、または dTTP）を、ウェルへと導入することができる。DNA ポリメラーゼにより、1 種または複数のヌクレオチドを組み込む場合、ウェル内では、プロトン（水素イオン）が放出される可能性があり、これを、イオンセンサーにより検出することができる。次いで、半導体チップを洗浄し、プロセスを、異なるデオキシリボヌクレオチドで反復することができる。半導体チップのウェル内では、複数の核酸を、シークエンシングすることができる。半導体チップは、DNA をシークエンシングする化学感受性電界効果トランジスター（chemFET）アレイ（例えば、米国特許出願公開第 2 0 0 9 0 0 2 6 0 8 2 号において記載されている）を含みうる。シークエンシングブライマーの 3' 末端における新たな核酸鎖への、1 種または複数のトリホスフェートの組み込みは、電流の変化を介して、chemFET により検出することができる。アレイは、複数の chemFET センサーを有しうる。

#### 【 0 1 2 5 】

本明細書で提示される発明の方法において使用されうるシークエンシング法の別の例は、ナノ小孔シークエンシング（例えば、Soni G V および M e l l e r A . ( 2 0 0 7 年 ) 、 C l i n C h e m 、 5 3 巻 : 1 9 9 6 ~ 2 0 0 1 頁を参照されたい）である。ナノ小孔は、直径 1 ナノメートルのオーダーの小型の穴でありうる。ナノ小孔を、伝導性流体中に浸漬し、これを横切る電位を印加する結果として、ナノ小孔を通るイオンの伝導に起因する、微弱な電流がもたらされる。流れる電流の量は、ナノ小孔のサイズに感受性である。DNA 分子が、ナノ小孔を通過するとき、DNA 分子上の各ヌクレオチドは、ナノ小孔を、異なる程度に塞ぐ。こうして、DNA 分子がナノ小孔を通過するときの、ナノ小孔を通過する電流の変化は、DNA 配列のリーディングを表しうる。

#### 【 0 1 2 6 】

##### 遺伝子解析

本発明の方法は、目的の選択的ゲノム領域についての遺伝子情報の解析のほか、目的の選択的領域と相互作用しうるゲノム領域内でも使用することができる。本明細書で開示される増幅法は、遺伝子解析のための、当技術分野で公知のデバイス、キット、および方法であって、米国特許第 6 , 4 4 9 , 5 6 2 号、同第 6 , 2 8 7 , 7 6 6 号、同第 7 , 3 6

10

20

30

40

50

1, 468号、同第7, 414, 117号、同第6, 225, 109号、および同第6, 110, 709号において見出されるデバイス、キット、および方法などであるがこれらに限定されない、デバイス、キット、および方法において使用することができる。場合によって、本発明の増幅法を使用して、多型の存在または非存在を決定するDNAハイブリダイゼーション研究のために、目的の標的核酸を増幅することができる。多型または対立遺伝子は、遺伝子疾患などの疾患または状態と関連しうる。他の場合、多型、例えば、嗜好、変性および加齢関連状態、がんなどと関連する多型は、疾患または状態への易罹患性に関連する。他の場合、多型は、冠動脈の健康の増大など、有益な形質と関連する場合もあり、HIVまたはマラリアなどの疾患に対する抵抗性と関連する場合もあり、骨粗鬆症、アルツハイマー病、または認知症などの変性疾患に対する抵抗性と関連する場合もある。

10

#### 【0127】

##### デジタル測定値

本発明の方法は、遺伝子発現、疾患と関連する遺伝子発現パターンについてのデジタル解析であって、診断、予後診断、および検出のほか、遺伝障害、例えば、染色体または遺伝子の転座、欠失、重複、および欠損の同定、ならびに目的の選択的ゲノム領域および目的の選択的領域と相互作用しうるゲノム領域についての研究も含むデジタル解析において使用することができる。一部の実施形態では、デジタル遺伝子発現(DGE)の決定またはコピー数変異(CNV)のデジタル測定値を、リード総数中の遺伝子リード数を定量化することにより達成することができる。一部の実施形態では、ペアドエンドシーケンシングを実施することができる。シーケンシングは、当業者に公知の様々なプラットフォーム上で、ハイスループットシーケンシングを介して実施することができる。一部の実施形態では、シーケンシングデータ/リードを、ゲノム/トランスクリプトーム(cDNAの場合)へとマッピングする。一部の実施形態では、米国特許出願公開第61/989, 113号において記載されている通り、配列データを評価して、重複リードを除去することができる。一部の実施形態では、プローブ配列を、重複を除外した配列データセット内でそれらが出現する回数であって、出発試料中に存在する元の核酸分子のコピーの数の尺度としての回数について計数する。

20

#### 【0128】

一部の実施形態では、特定核酸内の、その相補的なプローブ標的領域へと適正にアニールするプローブについての検証を評価することができる。一実施形態では、適正にアニールするプローブについての評価は、ペアドエンドアラインメントにより行うことができ、両方の末端である、フォワードリードおよびリバースリードが、予測される通りにアラインされれば、プローブを計数する。一部の実施形態では、適正にアニールするプローブについての評価は、プローブ配列の3'の特定核酸についてシーケンシングされるプローブ配列+20塩基を検討することにより行うことができ、重複解析のためには、フォワードリードだけを使用することができる。プローブ+20がアラインされれば、プローブは、所望の場所に存在した。

30

#### 【0129】

ランダム配列ではなく、プローブ配列計数を使用する利点は、各測定について、同じ配列を、異なる試料にわたり使用するため、コピー数解析の簡素化である。プローブ計数は、マルチプレックスシーケンシングを介する、高い試料スループット(例えば、シーケンシングラン1つ当たりの試料少なくとも96例)を可能とする。標的化されたRNAシーケンシングは、コード遺伝子または非コード遺伝子、特定のエクソン、UTR、RNAアイソフォーム、および遺伝子融合体を標的化する能力を拡張しながら、90%を超えるリードが、標的化された遺伝子に由来するので、RNAシーケンシング解析についての高レベルの焦点をもたらす。プローブ計数はまた、エクソンの使用、転写物のサイズ、および配列依存性増幅/シーケンシングのバイアスを低減することも可能であり、PCR重複の除去を可能とする。

40

#### 【0130】

50

デジタル解析は、定量化の前にPCR重複を決定することにより実施することができる。Illuminaシーケンシング技術を使用するこのような解析を、図6A~6Cに例示するが、図1も参照する。略述すると、gDNAについて例示されるフォワードリード(図1)は、図6Aにおいて示される通り、フォワード配列32のうちの少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも15、少なくとも20、少なくとも25、少なくとも30、少なくとも35塩基などの配列をシーケンシングするフォワードプライマー30であって、特定核酸10の配列へと伸長し、フォワードリード配列32を、ゲノム(またはcDNAの場合のトランスクリプトーム)領域へとマッピングするのに使用しうるフォワードプライマー30により活用されるフォワードプライミング部位12を含む。インデックスリードは、図6Bにおいて示される通り、試料の起源(例えば、ライブラリーと共通のライブラリーバーコード)を指し示しうる。インデックスリードは、インデックスプライマー34と共に、インデックスプライミング部位18において始まり、インデックス塩基(例えば、バーコード配列)16およびn-ランダム塩基14のうち、シーケンシングされる、少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10塩基などを含み、インデックスリード36をもたらす。一部の実施形態では、インデックス化リード塩基配列およびN-ランダム塩基36と組み合わせたフォワードリード配列32は、各特定核酸配列のためのライゲーション事象に固有である。一部の実施形態では、フォワードリード配列32の開始部位のゲノム(cDNAの場合のトランスクリプトーム)座標に、インデックスリード配列36のN-ランダム塩基14を加えた組み合わせを使用して、各プローブ伸長産物21または22についてのPCR重複を決定し、こうして、プローブ標的領域50または60を有する、対応する特定核酸配列10または9を決定することができる。リバースリード44は、図6Cに例示される通り、プローブが、適正なゲノム/トランスクリプトーム位置へとアニールされ、こうして、その相補的なプローブ標的領域へとアニールされたことを検証する。フローセル配列33および35は、濃縮時に、プローブ伸長産物の末端に追加する。

#### 【0131】

デジタル解析は、定量化の前にPCR重複を決定することにより実施することができる。このような解析を、図8A~8Cに例示するが、図1および図5も参照する。プローブ配列を有するリード44(図5)は、図8Aに例示される通り、プローブが、適正なゲノム/トランスクリプトーム位置へとアニールされ、こうして、その相補的なプローブ標的領域へとアニールされたことを検証する。プローブ配列を有するリードは、15塩基のリンカー38、40塩基のオリゴヌクレオチドの遺伝子特異的配列50または60(プローブ標的領域)、およびゲノム(またはトランスクリプトーム)データベースで表される40塩基のオリゴヌクレオチドの遺伝子特異的配列50または60に対して、約10塩基3'側にある、X塩基(例えば、10塩基)の領域10を含む。gDNAについて例示される、特定核酸配列10を有するリード(図1)は、図8Bにおいて示される通り、配列32のうちの少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも15、少なくとも20、少なくとも25、少なくとも30、少なくとも35塩基などの配列をシーケンシングするプライマー30(図5)であって、特定核酸9または10の配列へと伸長し、特定核酸配列を有するリード32を、ゲノム(またはcDNAの場合のトランスクリプトーム)領域へとマッピングするのに使用しうるプライマー30により活用されるプライミング部位12を含む。インデックス配列およびN6配列を含む配列リードは、図8Cにおいて示される通り、試料の起源(例えば、ライブラリーと共通のライブラリーバーコード)を指し示しうる。インデックスリードプライマー34は、インデックスプライミング部位18へとアニールし、インデックス塩基(例えば、バーコード配列)16およびn-ランダム塩基14の、シーケンシングされる、少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少な

10

20

30

40

50

くとも 8、少なくとも 9、少なくとも 10 塩基などのうちの少なくとも 1 つを含むリード配列 36 を産出し、シーケンシングリード 36 をもたらす。一部の実施形態では、特定核酸配列を有するリード (図 8 A) により、プローブの、プローブ標的領域へのアニーリングの特異性を検証する。一部の実施形態では、特定核酸配列を有するリードを、プローブ標的配列データベースへとマッピングするが、例えば、特定核酸配列 9 または 10 のうちの 10 塩基は、固有のプローブピン内にアラインされる必要があり、そのようにしてマッチングさせることにより、プローブの、そのプローブ標的領域へのアニーリングの特異性が検証される。一部の実施形態では、特定核酸配列を有するリードを、約 10 塩基のオリゴヌクレオチドマッチと比較し、配列がピン内で固有であるのかどうかを決定する。次いで、共通リードを、対応する N 6 と比較し、同一な N 6 リードを、単一のエントリーとして、併合し、1 回だけ計数する。一部の実施形態では、インデックスリードおよび N 6 リードのうちの少なくとも 1 つを有するリードは、約 14 塩基の長さでありうる。一部の実施形態では、特定核酸配列を有するリードは、約 10 塩基でありうる。一部の実施形態では、プローブ配列を有するリードは、約 65 塩基 (リンカー配列が約 15 塩基、プローブ標的領域 (ゲノム/トランスクリプトーム内で表される遺伝子特異的配列) が約 40 塩基、およびプローブ標的領域に対して 3' 側にある約 10 塩基) でありうる。一部の実施形態では、参照表 (look up table) を使用することができる。一部の実施形態では、プローブ配列を計数する。一部の実施形態では、N 6 配列は、排除する重複を指定する。

10

#### 【0132】

20

一部の実施形態では、リード配列 36 と組み合わせたリード配列 32 は、各特定核酸配列のためのライゲーション事象に固有である。一部の実施形態では、リード配列 32 の開始部位のゲノム (cDNA の場合のトランスクリプトーム) 座標に、リード配列 36 の N-ランダム塩基 14 を加えた組み合わせを使用して、各プローブ伸長産物 22 または 21 についての PCR 重複を決定し、プローブ標的領域 50 または 60 を有する、対応する特定核酸配列 10 または 9 を決定することができる。一部の実施形態では、リード配列 44 は、図 6 C に例示される通り、プローブが、適正なゲノム/トランスクリプトーム位置へとアニールされ、また、その相補的なプローブ標的領域 50 または 60 へもアニールされたことを検証する。

#### 【0133】

30

一部の実施形態では、上記で開示した通り、DGE または CNV の定量化の前に、重複リードを除去する。次いで、ゲノム/トランスクリプトーム内で適正にマッピングされたプローブ配列を計数する。一部の実施形態では、DGE または CNV を、各プローブ配列の計数により決定することができる。一部の実施形態では、プローブ計数を、例えば、遺伝子の全長において、プローブにわたり計数を平均することにより組み合わせることができる。一部の実施形態では、リード計数は、試料間で正規化することができ、例えば、リード計数を、全リードに対する百分率として正規化することができる。一部の実施形態では、リード計数は、例えば、各プローブ配列を計数する前に、全リード計数を正規化することにより正規化することができる。一部の実施形態では、リード計数は、ゲノムへとアラインさせたリードまたはプローブ標的領域に由来するリードの数により正規化することができる。

40

#### 【0134】

図 5 に例示される通り、シーケンシングライブラリーの構造およびシーケンシングリードの同定は、ハイスループットシーケンシング法を使用するマルチプレックスの定量化もたらす。図 6 を参照しながら、図 7 A に例示される通り、フォワードプライマー 30 は、フォワードリード 1 のプライミング部位 12 と相補的である可能性があり、リードは、リード 32 を、ゲノムまたはトランスクリプトームへとマッピングするのに十分な程度に特定核酸 10 (gDNA) または 13 (cDNA) へと伸長させることができる。加えて、図 7 B に例示されるインデックスリード配列は、相補的プライマー 34 と、「インデックス塩基」16 および「n-ランダム塩基」14 へのリーディング 36 とを使用して

50



、「インデックスプライミング部位」18から読み取ることができる。加えて、「リバー  
スリード配列」も決定することができる(図5、図6および図7C)。リバー  
ス配列プライマー42は、「リバー  
スリードプライミング部位」38とハイブリダイズし、プローブ  
50または60(それぞれ、図1、図2のgDNAまたはcDNAであり、図6に例示さ  
れる)(プローブ標的領域部位)、および少なくとも1、少なくとも2、少なくとも3、  
少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なく  
とも9、少なくとも10などの隣接塩基10(gDNA)または13(cDNA)を読み通し  
て44、プローブ伸長産物22または21(gDNAまたはcDNA、それぞれ、図1、  
図2)が、適正なゲノムまたはトランスクリプトームとハイブリダイズする  
プローブの産物であったのかどうかを検証する。ゲノム内またはトランスクリ  
プトーム内に適正にマッピングされたプローブ配列を計数する。DGEまたは  
CNVを、各プローブ配列の計数により決定することもでき、かつ/または  
プローブ計数を、特定核酸配列に対応する遺伝子の長さにおいて、  
プローブにわたり計数を平均することを含むがこれらに限定されないこと  
と組み合わせることもできる。リード計数は、試料間で正規化することが  
できる。一部の実施形態では、リード計数は、試料間で正規化することが  
でき、例えば、リード計数を、全リードに対する百分率として正規化す  
ることができる。一部の実施形態では、リード計数は、例えば、各  
プローブ配列を計数する前に、全リード計数を正規化することにより  
正規化することができる。一部の実施形態では、リード計数は、ゲノム  
へとアラインさせたリードまたはプローブ標的領域に由来するリードの  
数により正規化することができる。NGS配列解析の当業者には、正規化  
のための他の方法も周知である。

10

20

#### 【0135】

図9は、デジタル解析のために生成されるシーケンシングライブラリーの構築につ  
いてのグラフによる例示を提示する。

#### 【0136】

図10は、図9に例示される通りに構築されたシーケンシングライブラリーから生成  
されるシーケンシングデータの解析についてのグラフによる例示を提示する。

#### 【0137】

一部の実施形態では、開示される発明の方法は、例えば、組織、腫瘍、循環細胞につ  
いての遺伝子発現特徴および遺伝子発現特性であるがこれらに限定されない、  
遺伝子発現特徴および遺伝子発現特性を解析するほか、罹患患者を、非罹患患者と対  
比させて比較し、患者の正常組織を、患者の罹患組織と対比させて比較するた  
めのデジタル測定値のために使用することができる。一部の実施形態では、開  
示される発明の方法は、コピー数変異(CNV)のデジタル定量化のために使用  
することができる。CNVは、ゲノム内のDNAの変更であって、DNA区画の  
コピー数の異常な変異または正常な変異を有する細胞を結果としてもたら  
す変異を指し示しうる。CNVにより、大きなゲノム領域の欠失であって、  
染色体内で正常より少ない数を結果としてもたらす欠失、または大きなゲノ  
ム領域の重複であって、染色体内で通常を超える数を有する重複を同定す  
ることができる。CNVと、疾患に対する易罹患性または抵抗性との間には、  
関連が存在する。このような測定値は、診断、疾患病期分類、予後診断、  
疾患の進行、ウイルス負荷の決定のほか、遺伝子発現またはCNVの、治療  
剤の有効性または効能に対する影響であって、当業者に公知の影響などの  
決定にも有用でありうる。

30

40

#### 【0138】

別の態様では、開示される方法により増幅される、第1の核酸断片の配列を含む組成物  
が開示される。一部の実施形態では、第1の核酸断片または非断片化核酸試料は、  
同じヒトから選択されるヒト試料：単一の細胞、非罹患組織、罹患組織、FFPE  
試料または新鮮凍結試料、組織、臓器、腫瘍、患者から採取された有機流体の検  
体、遊離循環核酸、真菌、原核生物、およびウイルスに由来しうる。一部の  
実施形態では、第2の核酸断片または非断片化核酸試料は、罹患組織の場合  
もあり、非疾患組織の場合もあり、同じ日に回収される場合もあり、別の  
日に回収される場合もあり、異なる試料から回収される場合もあり、異  
なる方法により調製された試料から回収される場合もあり異なる精製法によ  
って試

50

料から回収される場合もあり、これらの組み合わせの場合もある組織を有する、同じヒトから選択される試料に由来しうる。一部の実施形態では、第1のアダプター配列を含む、第1の核酸断片または非断片化核酸試料をさらに濃縮し、大規模並行シーケンシングのために調製することができる。一部の実施形態では、第1の核酸断片または非断片化核酸試料は、二本鎖でありうる。一部の実施形態では、第1のアダプター配列を、前記第1の核酸断片または非断片化核酸試料の5'末端へと追加することができる。一部の実施形態では、第1のアダプター配列は、核酸修飾酵素のための制限部位および/または切断部位を含む。

#### 【0139】

さらに別の態様では、開示される方法は、アダプターを伴う、第2のヒト核酸断片または第2の非断片化核酸試料を有しうる。一部の実施形態では、第2のヒト試料は、第1の核酸試料が由来するヒトと異なるヒトに由来しうる。一部の実施形態では、第2の核酸断片または第2の非断片化核酸試料は、罹患組織の場合もあり、非疾患組織の場合もあり、同じ日に回収される場合もあり、別の日に回収される場合もあり、異なる試料から回収される場合もあり、異なる方法により調製された試料から回収される場合もあり異なる精製法によって試料から回収される場合もあり、これらの組み合わせの場合もある組織を有する、同じヒトから選択される試料でありうる。一部の実施形態では、第2の核酸断片または第2の非断片化核酸試料は、罹患組織の場合もあり、非疾患組織の場合もあり、同じ日に回収される場合もあり、別の日に回収される場合もあり、異なる試料から回収される場合もあり、異なる方法により調製された試料から回収される場合もあり異なる精製法によ

10

20

#### 【0140】

さらなる態様では、既の開示された方法に従い、第2のヒト核酸を定量化するための方法が開示される。

#### 【0141】

##### キット

本明細書で記載される組成物のいずれも、キット中に組み入れることができる。非限定的な例では、適切な容器手段内のキットは、公知の配列を伴う、1つのアダプターと、配列特異的部分と、配列が公知である共通部分とを有する、1つのプローブと、少なくともアダプターまたはプローブの共通部分に対する、直接的部分的な相補体を有する、1つのフォワードプライマーと、アダプターまたはプローブの共通部分に対する、直接的部分的な相補体を有する、1つのリバースプライマーとを含む。キットは、ライゲーション、標的の濃縮、およびライブラリーの調製に有用な、さらなるアダプター、プライマー、および/または試薬もさらに含有しうる。キットは任意選択で、DNAポリメラーゼもさらに含有しうる。キットは任意選択で、増幅のための試薬、例えば、PCR増幅法に有用な試薬もさらに含有しうる。キットは任意選択で、シーケンシングのための試薬、例えば、次世代大規模並行シーケンシング法に有用な試薬もさらに含有しうる。

30

#### 【0142】

キットの容器は、構成要素を入れることができ、好ましくは、適切にアリコート分割しうる、少なくとも1つのバイアル、試験管、フラスコ、ボトル、シリンジ、または他の容器を含みうる。キット中に、1つを超える構成要素が存在する場合、キットはまた、さらなる構成要素を個別に入れることができる、第2の容器、第3の容器、または他のさらなる容器も含有しうる。しかし、容器内には、構成要素の多様な組み合わせを含有させることもできる。

40

#### 【0143】

キットの構成要素を、1種または複数の溶液により提供しうる場合、溶液は、水溶液でありうる。しかし、キットの構成要素は、乾燥粉末として提供することもできる。試薬および/または成分を乾燥粉末として提供する場合、粉末は、適切な溶媒を添加することにより再構成することができる。

50

## 【0144】

キットは、キットの構成要素を用いるための指示のほか、キット中に含有されていない、他の任意の試薬の使用のための指示も含みうる。指示は、実装されうる変更を含みうる。

## 【実施例】

## 【0145】

(実施例 I：2つの試料の間の、特定転写物の示差的発現レベル)

100 ng の全 RNA から出発して、製造元の推奨に従い、Ovation (登録商標) Target Enrichment cDNA Module (NuGEN 型番 9301-32) を使用して、二本鎖 cDNA を作った。製造元の指示に従い、cDNA 試料を、NuGEN Ovation (登録商標) Target Enrichment Kit (NuGEN 型番 0400-32) へと直接添加した。ハイブリダイゼーションにおいて使用されるプローブは、270 の遺伝子を標的化するプローブのプール (NuGEN Ovation (登録商標) Target Enrichment System; 注文情報については、NuGEN に照会されたい) であった。

10

## 【0146】

結果として得られるライブラリーを、2 nM へと希釈し、ペアドエンドシーケンシングを、Illumina (登録商標) MiSeq (登録商標) DNA Sequencer 上で、濃縮されたライブラリーに対して実施した。以下のペアドエンドシリーズを、75 塩基のフォワードリード (R1)、75 塩基リバーズリード (R2)、14 塩基のインデックスリード (I1) で試行した。

20

## 【0147】

## データ解析

ペアドエンドアラインメントを、フォワードシーケンシングリードおよびリバーズシーケンシングリードについて実施し、デフォルトの設定で TopHat Alignment ソフトウェア (v. 2.0.10) を使用して、各々を、ヒトゲノム version hg19 へとマッピングした。標的化される領域へとマッピングされなかったリードのペアは、排除した。次いで、開始座標が同じフォワードリードを、インデックス配列の N6 配列 (n-ランダム配列) について評価した。N6 配列が同一な場合は、リードに、重複とマークし、群のうちの1つのリードだけを、単一の別々の核酸分子に由来するリードとして保持した。特許出願である US N 61/989,113 において記載されている通りに、同定された重複リードにマークし、次いで、除去した (重複のマーキング)。ねらい通りで (on target) 重複を除外したリードペアについてフィルタリングした後で、Trimalore (v. 0.3.1) を使用して、フィルタリングされたリバーズリード配列をトリミングし、アダプターおよびリンカー配列を除去し、FASTX Trimmer ソフトウェアを使用して、最初の35塩基へと短縮した。次いで、デフォルトパラメータおよび逆相補体マッチングを防止する「-norc」による Bowtie Alignment ソフトウェア (v. 1.0.0) を使用して、トリミングされたリバーズリード配列を、標的化オリゴヌクレオチドの配列を含有する、プローブ配列ファイル (使用されるプローブにより提示される) へとマッピングした。アラインさせたりリバーズリードを、それらの元のプライマーと関連させ、計数した。各プローブが検出される回数は、特定転写物が、元の試料中に存在する回数についての尺度であった。表1は、リード計数を試料間で正規化した DGE データを例示する。

30

40

## 【表 1】

表 1:2 つのがん細胞系の間、3 つの遺伝子についての、遺伝子発現の表示

遺伝子	正規化されたプロ ーブリード UHR	正規化されたプロ ーブリード H2228	UHR/H2228 の比
CCND3	499	494	1.01
TAF15	541	1074	0.52
PBX1	118	23	5.13

## 【0148】

表 1 に描示される通り、混合したがん細胞系の RNA 試料 (UHR、ユニバーサルヒト基準 RNA) における、TAF15 遺伝子の発現レベルは、H2228 細胞 (腺がん; 非小細胞肺癌) と比較して、比較的低レベルである。いずれの細胞型も、CCND3 の発現レベルは、ほぼ同等である。逆に、UHR による PBX1 の発現は、H2228 と比較して高度である。

## 【0149】

(実施例 II: ゲノムアラインメントを伴わない 2 つの試料の間、特定転写物の示差的発現レベル)

100 ng の全 RNA から出発して、製造元の推奨に従い、Ovation (登録商標) Target Enrichment cDNA Module (NuGEN 型番 9301-32) を使用して、二本鎖 cDNA を作る。製造元の指示に従い、cDNA 試料を、NuGEN Ovation (登録商標) Target Enrichment Kit (NuGEN 型番 0400-32) へと直接添加する。ハイブリダイゼーションにおいて使用されるプローブは、270 の遺伝子を標的化するプローブのプール (NuGEN Ovation (登録商標) Target Enrichment System; 注文情報については、NuGEN に照会されたい) である。

## 【0150】

結果として得られるライブラリーを、2 nM へと希釈し、ペアドエンドシーケンシングを、Illumina (登録商標) MiSeq (登録商標) DNA Sequencer 上で、濃縮されたライブラリーに対して実施する。以下のペアドエンドシリーズを、75 塩基のフォワードリード (R1)、75 塩基リバースリード (R2)、14 塩基のインデックスリード (I1) で試行する。

## 【0151】

## データ解析

5' 末端において、15 bp のリンカー配列および 0 ~ 3 塩基の多様性配列についてのパターンマッチを伴うリバースリード配列をトリミングする。リンカーのトリミングの後で、BEETL ソフトウェア (version 1.1.0; <https://github.com/BEETL/BEETL>) を使用して、リバースリードのうちの最初の 40 bp を、各リード内の 12 マーのタイル状 (tiled) のプローブにマッチするように、Burrows-Wheeler 変換 (BWT) へと構築する。次いで、各リードペアに、リバースリードへの 12 マーのマッチが最も多いプローブに由来すると標識する。次いで、インデックス配列の N6 配列 (n-ランダム配列) を、フォワードリードの最初の 10 塩基と共に解析することにより、プローブごとに、標識されたリードペアの重複を除外した。リードが同じプローブに由来し、N6 配列が同一であり、フォワードリードの最初の 10 塩基が同じである場合は、リードに、重複とマークし、群のうちの 1 つのリードだけを、単一の別々の核酸分子に由来するリードとして保持した。重複を除外したリードペアについてフィルタリングした後で、各プローブについて、重複を除外した総リード計数を得た。各プローブが検出される回数は、特異的なプローブが、元の試料中に存在する回数についての尺度であった。次いで、特定の試料中のその遺伝子についての相対存在度の尺度としての、プローブ注釈ファイルに基づき、プローブ 1 つ当たりの計数を平均して、遺伝子 1 つ当たりの計数を得る。

10

20

30

40

50

## 【0152】

(実施例III：フォワードリードをマッピングする、特定転写物の示差的発現)

100 ngの全RNAから出発して、製造元の推奨に従い、Ovation (登録商標) Target Enrichment cDNA Module (NuGEN型番9301-32)を使用して、ユニバーサルヒト基準試料(UHR)の二本鎖cDNAに由来するインプットを作った。製造元の指示に従い、cDNA試料を、NuGEN Ovation (登録商標) Target Enrichment Kit (NuGEN型番0400-32)へと直接添加した。また、Promega Male Reference Sampleによる100 ngのDNAインプットから出発する対照ライブラリーも、製造元の推奨に従い、Ovation (登録商標) Target Enrichment Kit (NuGEN型番0400-32)を使用して加工した。ハイブリダイゼーションにおいて使用されるプローブは、95の遺伝子を標的化するプローブのプールであった。

10

## 【0153】

結果として得られるライブラリーを、2 nMへと希釈し、ペアドエンドシーケンシングを、Illumina (登録商標) MiSeq (登録商標) DNA Sequencer上で、濃縮されたライブラリーに対して実施した。以下のペアドエンドシリーズを、70塩基のフォワードリード(R1)、88塩基のリバースリード(R2)、14塩基のインデックスリード(I1)で試行した。

## 【0154】

データ解析

RNA由来のデータおよびDNA由来のデータのいずれについても、フォワードリードを品質でトリミングし、かつ、フォワードリードから、リンカーおよびアダプター配列をトリミングした。DNA由来のデータについては、-m2パラメータによるBowtie Alignmentソフトウェアを使用して、フォワードリードを、ヒトゲノムversion hg19へとマッピングした。RNA由来のデータについては、STAR Alignmentソフトウェアを使用して、フォワードリードを、リボソームRNA基準配列へとまずマッピングし、次いで、リボソームRNAへとマッピングされなかったリードは、ここでもまた、STAR Alignmentソフトウェアを使用して、ヒトversion hg19へとマッピングした。アラインメントの後で、次いで、RNAデータおよびDNAデータの両方について、開始座標が同じフォワードリードを、インデックス配列のN6配列(n-ランダム配列)について評価した。N6配列が同一な場合は、リードに、重複とマークし、群のうちの1つのリードだけを、単一の別々の核酸分子に由来するリードとして保持した。特許出願であるUS 2016/0198913において記載されている通りに、同定された重複リードにマークし、次いで、除去した(重複のマーキング)。次いで、CoverageBedソフトウェアを、デフォルトの設定でを使用して、各データセットについて、各標的領域(エクソン)の任意の部分で重なり合う、重複を除外したフォワードリードを計数した。各標的領域についての計数を、データセットの全ての標的領域内の全リードについて正規化し、次いで、遺伝子内の各エクソンに対応する標的領域を、DNAデータおよびRNAデータの正規化された遺伝子計数について平均した。発現レベルが、リードを生成するプローブの能力に影響を及ぼさないため、DNA計数は、まったく均一であることが予測される。RNA計数は、発現レベルの変化に起因して、変動性を有することが予測される。この考えに基づき、次いで、正規化計数のRNA/DNAのlog2比を、RNA内の遺伝子存在度の尺度として計算した。次いで、スチューデントのT検定を使用して、各遺伝子測定値についてのp値を計算した。p値を<0.05とし、log比を>0とする遺伝子は、上方調節される遺伝子として注記し、p値を<0.05とし、log比を<0とする遺伝子は、下方調節される遺伝子として注記した。

20

30

40

## 【0155】

表2は、染色体の順序で、95の遺伝子によるパネル内のRNAレベルにおける全ての

50

遺伝子の存在度のプロットからの、顕著に上方調節された5つの遺伝子、および顕著に下方調節された5つの遺伝子を提示する(図11)。

【表2】

表2:有意に変化した遺伝子の相対存在度

方向	遺伝子	染色体	存在度	P 値
上方	GUSB	7	6.291068	3.18E-04
上方	ANXA1	9	8.132735	3.37E-07
上方	ITGB7	12	13.091	5.98E-05
上方	GAS6	13	5.678613	2.64E-04
上方	TSC2	16	3.440623	3.22E-04
下方	AMPD1	1	0.011096	2.95E-06
下方	CR2	1	0.133867	3.35E-04
下方	ITGAX	16	0.194365	2.39E-06
下方	NOS2	17	0.037126	9.74E-07
下方	ITGA2B	17	0.15729	2.31E-13

10

# 【0156】

(実施例IV:フォワードリードを使用する、特定転写物の相対発現レベル)

100ngの全RNAから出発して、製造元の推奨に従い、Ovation(登録商標) Target Enrichment cDNA Module (NuGEN型番9301-32)を使用して、二本鎖cDNAを作る。ILMNリバースフローセル配列に対応するアダプターを、各cDNA断片の5'末端へとライゲーションする。配列特異的領域に続いて、ILMNフォワードフローセル配列に対応する、15塩基のリンカーおよびXX塩基配列を含有するプローブを、標的へとアニールさせ、DNAポリメラーゼにより伸長させる。フォワードフローセル配列およびリバースフローセル配列の両方を含有するDNA断片を、NuGENにより推奨され、そして提供される条件下で、かつ試薬と共にPCRにより増幅する。

20

# 【0157】

結果として得られるライブラリーを、2nMへと希釈し、ペアドエンドシーケンシングを、Illumina(登録商標) MiSeq(登録商標) DNA Sequence r上で、濃縮されたライブラリーに対して実施する。以下のペアドエンドシリーズを、70塩基のフォワードリード(R1)、および14塩基のインデックスリード(I1)で試行する。

30

# 【0158】

データ解析

Trimaloreソフトウェア(v.0.3.1)を使用して、フォワードリード配列をトリミングして、リンカー配列を除去し、FASTX Trimmerソフトウェアを使用して、最後の55塩基へと短縮する。-m2パラメータによるBowtie Alignmentソフトウェアを使用して、トリミングされたフォワードリードを、ヒトゲノムversion hg19へとマッピングする。標的化される領域へとマッピングされないリードは、排除する。ゲノム内の同じ開始座標へとマッピングされるリードを同定する。次いで、開始座標が同じリードを、インデックス配列のN6配列(n-ランダム配列)について評価する。N6配列が同一な場合は、リードのペアに、重複とマークし、単一の別々の核酸分子に由来するペアとしてだけ計数する。特許出願であるUS61/989,113において記載されている通りに、同定された重複リードにマークし、次いで、除去する(重複のマーキング)。デフォルトパラメータおよび逆相補体マッチングを防止する「-norc」によるBowtie Alignmentソフトウェア(v.1.0.0)を使用して、残りのリードを、標的化オリゴヌクレオチドの配列を含有する、プローブ配列ファイル(使用されるプローブにより提示される)へとマッピングする。各プローブが検出される回数は、特定転写物が、元の試料中に存在する回数についての尺

40

50

度である。標的領域（エクソン）の任意の部分で重なり合うリードを計数する。遺伝子内の各エクソンに対応する計数を平均する。あるエクソンの計数が平均を2標準偏差下回る場合は、そのエクソンを棄却し、平均を計算し直す。

#### 【0159】

（実施例V：DNAシーケンシングによるコピー数変異（CNV）の決定）

2例のヒトgDNA試料（1例は、男性のトリソミーの第13染色体に由来し、もう1例は、女性のダイソミーの第13染色体に由来する）を、Covarisシステムによる超音波処理を介して、約500bpの長さへと断片化した。各試料に由来するgDNAの500bpの断片100ngを、製造元の指示に従い、NuGEN Ovation（登録商標）Target Enrichment Kit（NuGEN、型番0400-32）へと添加した。ハイブリダイゼーションにおいて使用されるプローブは、344の遺伝子を標的化するプローブのプール（NuGEN Ovation（登録商標）Cancer Panel Target Enrichment System）であった。

10

#### 【0160】

結果として得られるライブラリーを、2nMへと希釈し、ペアドエンドシーケンシングを、Illumina（登録商標）MiSeq（登録商標）DNA Sequence r上で、濃縮されたライブラリーに対して実施した。以下のペアドエンドシリーズを、75塩基のフォワードリード（R1）、88塩基のリバースリード（R2）、14塩基のインデックスリード（I1）で試行した。

#### 【0161】

20

データ解析

データは、2つの独立の方法：重複を除去する方法および重複を除去しない方法により解析した。略述すると、デフォルトの設定で、Bowtie Alignmentソフトウェア（v.1.0.0）を使用して、フォワードリードを、ヒトゲノムversion hg19に対してアラインさせた。あるフォワードリードが、同じゲノム開始座標に対してアラインすることが決定されたら、対応するインデックスリードを検討した。ゲノムの開始座標が同じであるフォワードリードに対応するインデックスリードの配列が同一である場合は、リードに、重複とマークし、単一の別々の核酸分子としてだけ計数する（特許出願であるUS 2016/0198913において記載されている通り）。残りの別々のフォワードリードに対応するリバースリードは、Bowtieを使用して、プローブデータベース内の配列に対してアラインさせた。アラインさせたリバースリードは、それらが、どのプローブ配列を表すのかに従いピニングおよび計数した。各プローブが表される回数は、出発特定核酸分子が、元の試料中に存在する回数についての尺度であった。

30

#### 【0162】

代替的に、重複リードを除去せず、プローブデータベース内に存在する配列に従い、40塩基のリバースリードを分類する（cataloging）ことにより、表示を確立した。プローブ基準データベース内の各代表に対してアラインさせたリードの数を決定した。データベース内の配列にマッチしないリードは、無視した。各プローブが検出される回数は、特定配列が、元の試料中に存在する回数についての尺度であった。表3は、上記で記載したいずれかの方法を使用するCNVデータであって、リード計数を、総シーケンシングリード数に対して正規化し、10を下回るあらゆる計数を解析から除去したCNVデータを描示する。トリソミー男性の試料中の所与のプローブについてのプローブ計数の、女性の野生型試料中の同じプローブの計数に対する比を、所与の染色体における全てのプローブについて平均した。

40

## 【表 3】

表 3:リード計数を、総シーケンシングリード数に対して正規化した、13 トリソミーの男性に由来するコピー数変異データ

染色体	重複を除去しない場合の平均プローブ計数比	重複を除去した場合の平均プローブ計数比
第 1 染色体	1.002485	0.990606365
第 2 染色体	1.025382	1.010290049
第 3 染色体	1.028736	1.016439439
第 4 染色体	1.045166	1.032544903
第 5 染色体	1.002378	0.998957554
第 6 染色体	1.015266	0.997262904
第 7 染色体	1.022412	1.021639631
第 8 染色体	1.046251	1.028980962
第 9 染色体	1.009415	0.991277289
第 10 染色体	1.035216	0.993768193
第 11 染色体	1.01177	1.00377304
第 12 染色体	1.027063	1.004790487
第 13 染色体	1.485411	1.471641235
第 14 染色体	0.996186	0.986919321
第 15 染色体	0.986867	0.981480187
第 16 染色体	0.967682	0.964463441
第 17 染色体	0.999821	0.992014077
第 18 染色体	1.035764	1.016860381
第 19 染色体	0.967125	0.958202381
第 20 染色体	1.012836	1.010031227
第 21 染色体	1.00104	1.013150115
第 22 染色体	0.975676	0.972111601
X 染色体	0.548004	0.54329808

## 【 0 1 6 3 】

表 3 に描示される通り、二倍体の男性は、2 つの X 染色体を有する、二倍体の正常（野生型、WT）な女性と対比して、単一の X 染色体を有し、これは、X 染色体についてのプローブ計数の 0.54 の比（または重複を除去しない場合の 0.55 の比）により同定される。同様に、男性および女性のいずれも、第 13 染色体を例外とする他の全ての染色体について、同等の正規化された計数を有しうる。13 トリソミーの男性は、1.47 のプローブ計数比（または重複を除去しない場合の 1.49）により解釈される通り、余分な第 13 染色体を有することから、WT の女性に対する比較と対比して、第 13 染色体のコピー数変異が確立される。

## 【 0 1 6 4 】

（実施例 V I : DNA シーケンシングによる、がん細胞系内のコピー数変異（CNV）の決定）

2 例のヒト gDNA 試料（1 例は、正常男性のプール（Promega）に由来し、もう 1 例は、正常男性の同じプールに由来するが、2 つの余分なコピーの EGF R 遺伝子および K I T 遺伝子が、各々合計 4 コピーずつスパイクされた（qPCR により既に検証されている））を、Covaris システムによる超音波処理を介して、約 500 bp の長さへと断片化した。各試料に由来する gDNA の 500 bp の断片 100 ng を、製造元の指示に従い、NuGEN Ovation（登録商標）Target Enrichment Kit（NuGEN、型番 0400-32）へと添加した。ハイブリダイゼーションにおいて使用されるプローブは、509 の遺伝子を標的化するプローブのプール（NuGEN Ovation（登録商標）Cancer Panel 2.0 Target



t Enrichment System)であった。

#### 【0165】

結果として得られるライブラリーを、2 nMへと希釈し、ペアドエンドシーケンシングを、Illumina(登録商標)MiSeq(登録商標)DNA Sequence r上で、濃縮されたライブラリーに対して実施した。以下のペアドエンドシリーズを、70塩基のフォワードリード(R1)、88塩基のリバースリード(R2)、14塩基のインデックスリード(I1)で試行した。

#### 【0166】

##### データ解析

いずれのデータセットについても、fastqフォーマットのフォワードリードから、リンカー配列および低品質の塩基を、Trim Galoreソフトウェアでトリミングする。デフォルトの設定で、Bowtie Alignmentソフトウェア(v1.0.0)を使用して、リードを、ヒトゲノム基準配列version hg19に対してアラインさせて、リードが、最大2つの地点へとマッピングされるようにし、単一の最良のアラインメントだけを取り出した(-m2 -best)。続いて、NuGEN社特製のNuDup deduplicationソフトウェア(<https://github.com/nugentechnologies/nudup>)を使用して、アラインさせたリードの重複を除外した。重複を除外するために、あるリードが、同じゲノム開始座標に対してアラインすることが決定されたら、対応するインデックスリードを検討する。ゲノムの開始位置が同じであるフォワードリードに対応するインデックスリードの配列が同一である場合は、リードに、重複とマークし、セットに由来する、品質が最良である、単一のリードだけを維持する。

#### 【0167】

濃縮実験で使用されるプローブは、プローブの開始座標～プローブの約300bp下流の範囲内に着地するリードを産出すると予測される。濃縮における全てのプローブについて、プローブ着地帯域は、bedファイル内で、「probePlus300」として規定される。各probePlus300領域内の、重複を除外したリードの数は、BEDtoolsのcoverageBedを使用して計数する。各probePlus300領域について、絶対計数を、全てのprobePlus300領域内に収まる、重複を除外した全リード(全probePlus300領域計数の合計)により、実験にわたり計数を比較するために正規化する。次に、各遺伝子またはゲノム領域について、probePlus300計数を平均する。細胞系試料に由来する各遺伝子についての、正規化された平均probePlus300計数を、正常ブレンドの男性試料計数と、比として比較する。さらに、スチューデントのt検定を使用して、多重仮説で補正されたp値を<0.005として、平均されたprobePlus300計数が、所与の遺伝子について、2つの試料間で顕著に異なる遺伝子またはゲノム領域を計算することができる。

#### 【0168】

表4は、試料中のスパイクにおける、有意なコピー数の変化およびp値について描述する。具体的には、EGFR遺伝子およびKIT遺伝子のコピー数だけが有意に増大している(2つの遺伝子を約4コピーでスパイクした場合)。

##### 【表4】

表4:試料中のスパイクにおける、有意なコピー数の変化およびp値

CNV	遺伝子	染色体	コピー	P値
増大	KIT	4	3.790214	5.05E-13
増大	EGFR	7	4.059194	1.62E-16

#### 【0169】

(実施例VII:DNAシーケンシングによりコピー数変異を決定するための、迅速ライブラリー生成)

2例のヒトgDNA試料(1例は、男性のトリソミーの第13染色体に由来し、もう1

10

20

30

40

50

例は、女性のダイソミーの第13染色体に由来する)を、Covarisシステムによる超音波処理を介して、約500bpの長さへと断片化することができる。各試料に由来するgDNAの500bpの断片1ugを、プローブおよびプローブのアニール溶液(NuGEN Ovation(登録商標)Target Enrichment Kit、型番0400-32)の存在下、95℃で5分間にわたり熱変性させ、1分間当たり0.1℃/sの速度で、60℃まで冷却し、この温度で少なくとも30分間にわたり保つことができる。アニールステップに続き、DNAポリメラーゼと、デオキシヌクレオチドとを、溶液へと添加して、それらの鋳型核酸と特異的にアニールしたプローブを伸長させることができる。この溶液を、室温まで冷却し、製造元の推奨に沿って、組み込まれなかったプローブを、示差的なビーズの結合と、SPRIビーズからの溶出とにより除去することができる。NuGEN Ovation(登録商標)Target Enrichment Kitにより提供されている溶液により、回収された二本鎖DNAに、末端修復およびライゲーションを施すことができる。

10

#### 【0170】

結果として得られるライブラリーを、2nMへと希釈し、ペアドエンドシーケンシングを、Illumina(登録商標)MiSeq(登録商標)DNA Sequencer上で、濃縮されたライブラリーに対して実施することができる。以下のペアドエンドシリーズを、75塩基のフォワードリード(リード1)、75塩基のリバースリード(リード2)、14塩基のインデックスリード(リード3)で試行することができる。

20

#### 【0171】

##### データ解析

プローブデータベース内に存在する配列により、75塩基のリバースリードを分類することを介して、データを解析することができる。プローブ基準データベース内の各代表に対してアラインするリードの数を決定することができる。データベース内の配列に対してアラインしなかったリードは、無視することができる。各プローブが検出される回数は、特定配列が、元の試料中に存在した回数についての尺度でありうる。リード計数を、総シーケンシングリード数に対して正規化し、10を下回るあらゆる計数を解析から除去することができる。トリソミーの男性の試料中の所与のプローブについてのプローブ計数の、野生型の女性の試料中の同じプローブの計数に対する比は、所与の染色体における全てのプローブについて平均することができる。

30

#### 【0172】

この試験によるデータは、男性試料を、2つのX染色体を有する、二倍体の正常(野生型、WT)な女性と対比して、単一のX染色体を有し、したがって、X染色体についてのプローブ計数が0.5の比である試料として明らかにするであろう。同様に、男性および女性のいずれも、第13染色体を例外とする他の全ての染色体について、同等の正規化された計数を有しうるであろう。13トリソミーの男性は、余分な第13染色体を有し、この結果としてとして、WTの女性と比べて、約1.5のプローブ計数比がもたらされるであろう。

#### 【0173】

当業者は、多くの改変、代替物、および等価物が可能であることを理解するであろう。全てのこのような改変、代替物、および等価物は、本明細書に包含されることを意図する。

40

#### 【0174】

本発明の原理について、具体的な実施形態との関連で記載してきたが、これらの記載は、例だけを目的としてなされるものであり、本発明の範囲を限定することを意図するものではないことを明らかに理解することができる。本明細書で開示されたことは、例示および記載のために提示されている。網羅的であること、または開示されることを、記載される正確な形態へと限定することは、意図されていない。当業者には、多くの改変および変更が明らかであろう。開示されていることは、記載される技術についての、開示される実施形態の、原理および実質的な適用を最もよく説明し、これにより、他の当業者が、想定

50

される特定の使用に適する、多様な実施形態および多様な改変を理解することを可能とするために、選出され、記載されている。開示されることの範囲は、以下の特許請求の範囲およびそれらの等価物により規定されることが意図される。

【図 1】

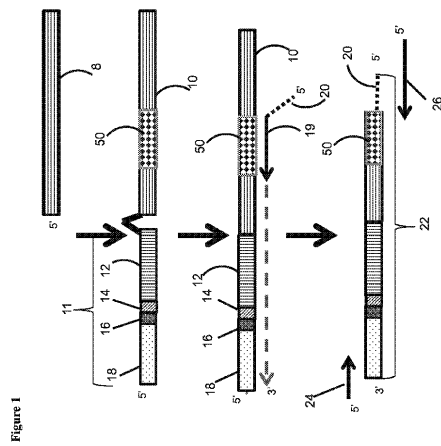


Figure 1

【図 2】

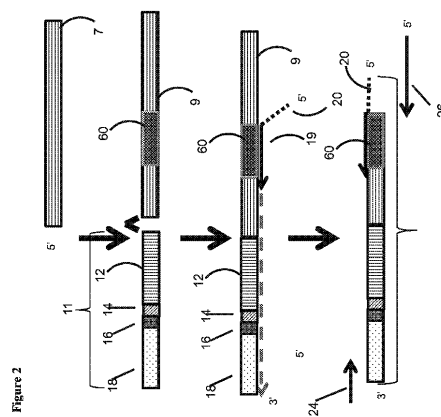


Figure 2

【図 3】

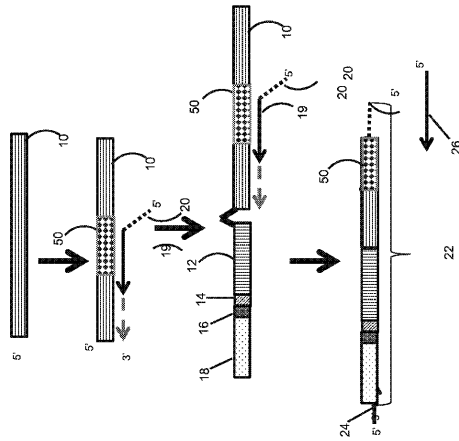


Figure 3

【図 4】

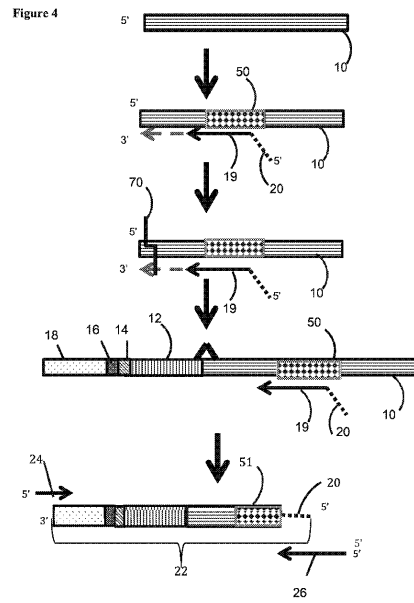
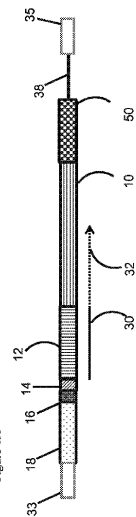


Figure 6A

【図 6 A】



【図 5】

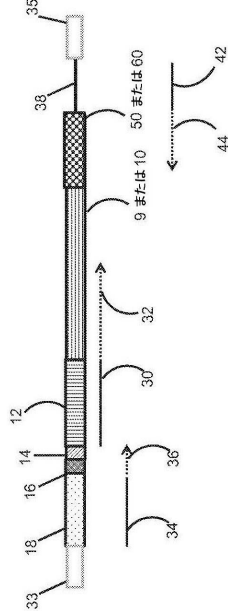
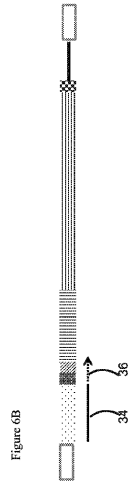
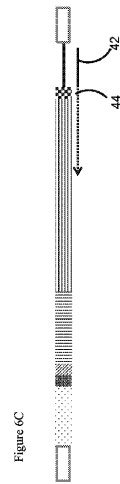


Figure 5

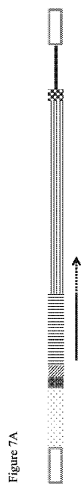
【図 6 B】



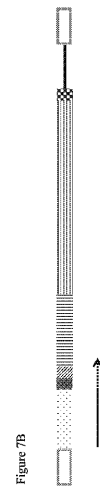
【図 6 C】



【図 7 A】



【図 7 B】



【 7 C 】

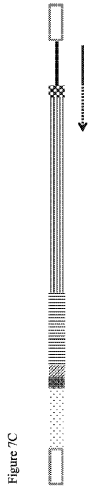


Figure 7C

【 8 A 】

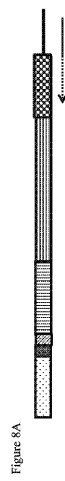


Figure 8A

【 8 B 】

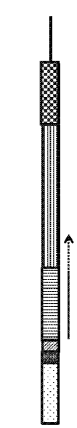


Figure 8B

【 8 C 】

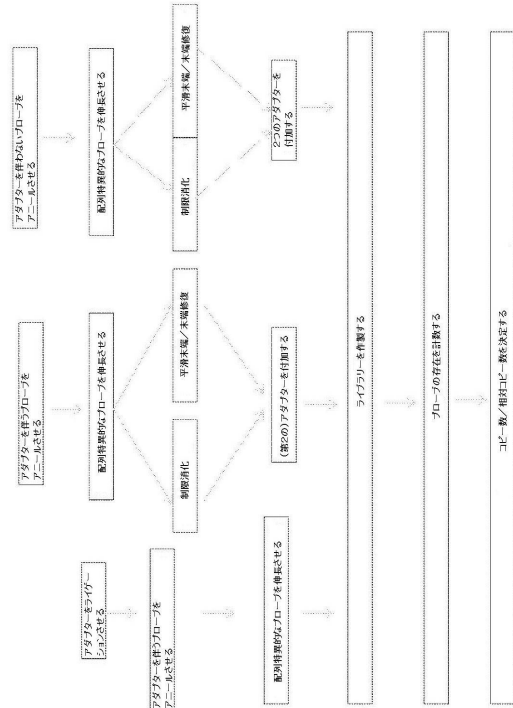


Figure 8C

【図 9】

ライブラリーの生成

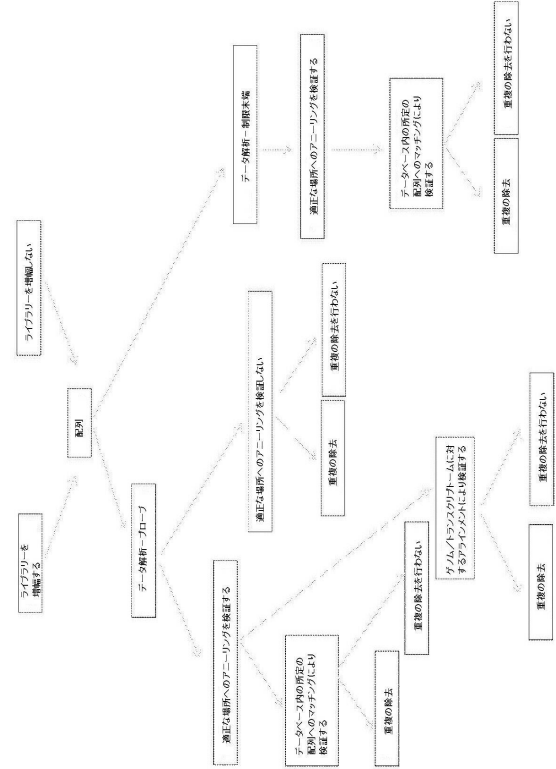
Figure 9



【図 10】

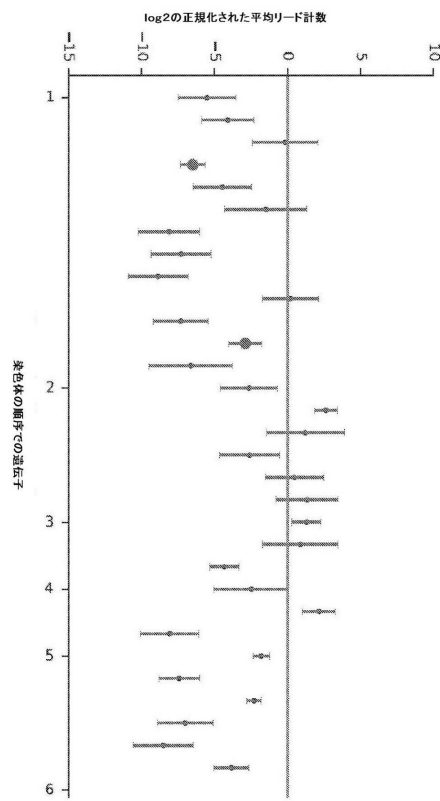
解析

Figure 10



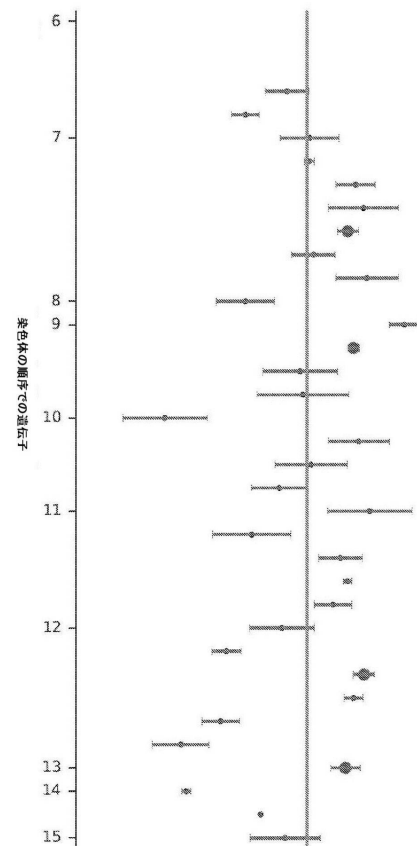
【図 11 - 1】

Figure 11



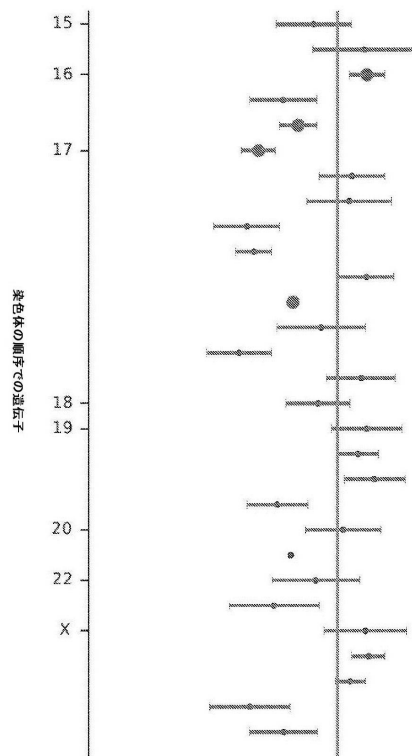
【図 11 - 2】

Figure 11 続き



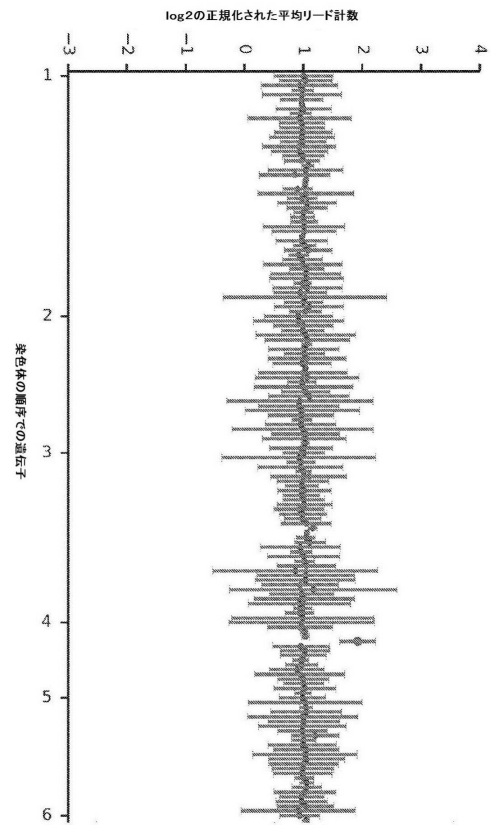
## 【図 1 1 - 3】

Figure 11 続き



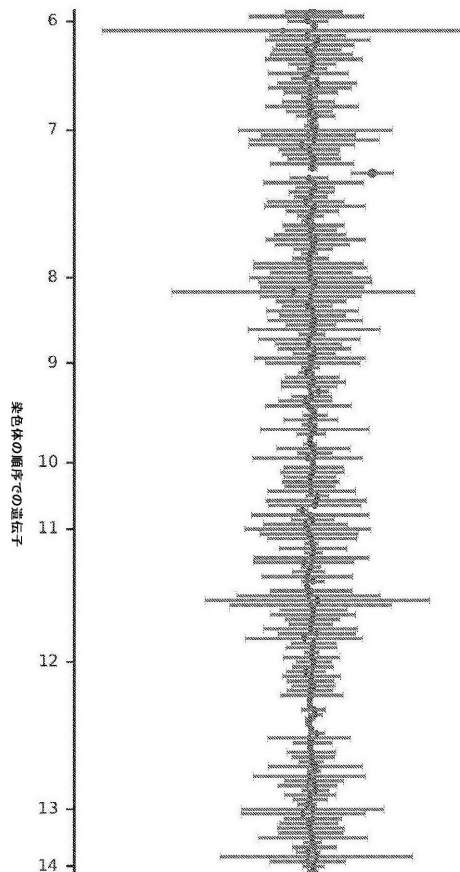
## 【図 1 2 - 1】

Figure 12



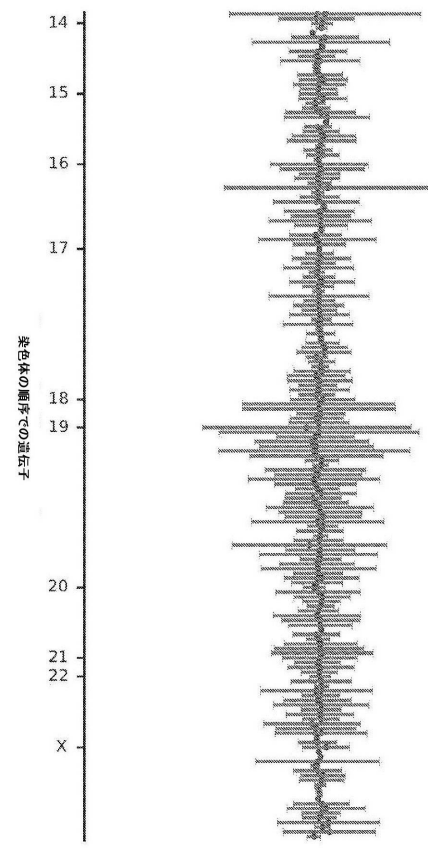
## 【図 1 2 - 2】

Figure 12 続き



## 【図 1 2 - 3】

Figure 12 続き





---

フロントページの続き

- (72)発明者 シュローダー, ベンジャミン  
アメリカ合衆国 カリフォルニア 94401, サン マテオ, ショアビュー アベニュー  
1964
- (72)発明者 アモアズ, ダグラス  
アメリカ合衆国 カリフォルニア 94022, ロス アルトス, エス. エル モンテ ア  
ベニュー 357
- (72)発明者 ウェルガ, ステファニー シー.  
アメリカ合衆国 カリフォルニア 94002, ベルモント, ハスティングス ドライブ 2  
203, アpartment 24

審査官 北村 悠美子

- (56)参考文献 国際公開第2014/039556 (WO, A1)  
Seda Eminaga et al., Curr Protoc Mol Biol., 2013年 7月, Vol.103, 4.17.1-4.17.14

- (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 )