

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2023-519782  
(P2023-519782A)

(43)公表日 令和5年5月15日(2023.5.15)

(51)国際特許分類	F I	テーマコード(参考)
C 4 0 B 40/06 (2006.01)	C 4 0 B 40/06	4 B 0 6 3
C 1 2 Q 1/6844(2018.01)	C 1 2 Q 1/6844	Z
C 1 2 Q 1/6853(2018.01)	C 1 2 Q 1/6853	Z
C 1 2 Q 1/48 (2006.01)	C 1 2 Q 1/48	
C 1 2 Q 1/34 (2006.01)	C 1 2 Q 1/34	

審査請求 未請求 予備審査請求 未請求 (全46頁) 最終頁に続く

(21)出願番号 特願2022-542997(P2022-542997)  
 (86)(22)出願日 令和3年1月15日(2021.1.15)  
 (85)翻訳文提出日 令和4年9月12日(2022.9.12)  
 (86)国際出願番号 PCT/US2021/013598  
 (87)国際公開番号 WO2021/146534  
 (87)国際公開日 令和3年7月22日(2021.7.22)  
 (31)優先権主張番号 62/962,777  
 (32)優先日 令和2年1月17日(2020.1.17)  
 (33)優先権主張国・地域又は機関 米国(US)  
 (31)優先権主張番号 63/009,603  
 (32)優先日 令和2年4月14日(2020.4.14)  
 (33)優先権主張国・地域又は機関 米国(US)  
 (81)指定国・地域 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA)  
 最終頁に続く

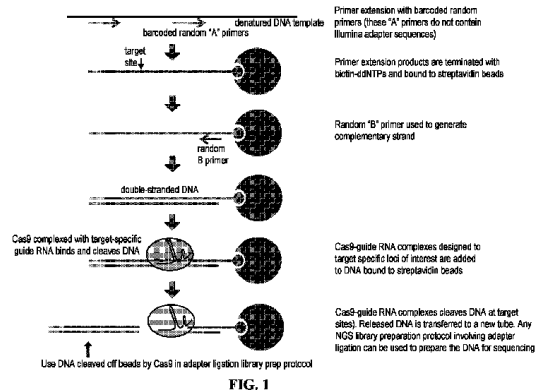
(71)出願人 518274618  
 ジャンプコード ゲノミクス, インク .  
 アメリカ合衆国 9 2 1 2 1 カリフォル  
 ニア州 サンディエゴ ネクサス・センタ  
 ー・ドライブ 4 7 5 5 スイート 1 0 0  
 (74)復代理人 110003797  
 弁理士法人清原国際特許事務所  
 (74)代理人 100082072  
 弁理士 清原 義博  
 (72)発明者 ブラウン, キース  
 アメリカ合衆国 9 2 0 0 8 カリフォル  
 ニア州 カールスバッド ドライデン・ブ  
 レイス 5 8 5 8 スイート 2 0 4  
 F ターム(参考) 4B063 QA13 QQ02 QQ06 QQ07  
 QQ08 QQ10 QQ42 QQ52  
 最終頁に続く

(54)【発明の名称】 標的化された配列決定の方法

(57)【要約】

本明細書において、標的核酸を含む配列決定ライブラリーを作製するための方法と組成物が提供される。本明細書の方法は：核酸サンプルをプライマーの第1の集団、ポリメラーゼ、dNTP、および標識されたddNTPに接触させる工程；伸長反応を実行する工程であって、それによって標識された伸長生成物を作製する、工程；標的核酸を含む二本鎖伸長生成物を作製するために、伸長生成物をプライマーの第2の集団に接触させる工程；二本鎖伸長生成物の少なくともサブセットの切断を可能にする条件下で、前記二本鎖伸長生成物を標的的特異的酵素に接触させる工程であって、それにより、切断された標的核酸を作製する、工程；および、切断された標的核酸を単離する工程を含み得る。

【選択図】図1



## 【特許請求の範囲】

## 【請求項 1】

標的核酸を含む配列決定ライブラリーを作製する方法であって、前記方法は、

(a) 核酸サンプルをプライマーの第 1 の集団、ポリメラーゼ、d N T P、および標識された d d N T P に接触させる工程、

(b) 伸長反応を実行する工程であって、それによって標識された伸長生成物を作製する、工程、

(c) 標的核酸を含む二本鎖伸長生成物を作製するために、前記伸長生成物をプライマーの第 2 の集団に接触させる工程

(d) 前記二本鎖伸長生成物の少なくともサブセットの切断を可能にする条件で、前記二本鎖伸長生成物を標的的特異的酵素に接触させる工程であって、それにより、切断された標的核酸を作製する、工程；および、

(e) 前記切断された標的核酸を単離する工程を含む、方法。

## 【請求項 2】

前記核酸サンプルは変性される、請求項 1 に記載の方法。

## 【請求項 3】

プライマーの前記第 1 の集団または第 2 の集団は、ランダムプライマーを含む、請求項 1 または 2 に記載の方法。

## 【請求項 4】

プライマーの前記第 1 の集団または第 2 の集団は、バーコードを含む、請求項 1 または 2 に記載の方法。

## 【請求項 5】

プライマーの前記第 1 の集団または第 2 の集団は、オリゴ-d T プライマーを含む、請求項 1 または 2 に記載の方法。

## 【請求項 6】

標識された d d N T P は、ビオチンを含む、請求項 1 から 5 のいずれか 1 つに記載の方法。

## 【請求項 7】

標識された伸長生成物は、ビーズ上で単離される、請求項 1 から 6 のいずれか 1 つに記載の方法。

## 【請求項 8】

前記ビーズはストレプトアビジン・ビーズである、請求項 7 に記載の方法。

## 【請求項 9】

前記ビーズは磁気ビーズである、前記請求項 7 に記載の方法。

## 【請求項 10】

標識された伸長生成物は、ポリプロピレンまたはポリカーボネート表面上で単離される、請求項 1 から 6 のいずれか 1 つに記載の方法。

## 【請求項 11】

標的的特異的酵素は、標的的特異的なガイド RNA と複合体化される、請求項 1 から 10 のいずれか 1 つに記載の方法。

## 【請求項 12】

標的的特異的な酵素は、C R I S P R 酵素である、請求項 1 から 11 のいずれか 1 つに記載の方法。

## 【請求項 13】

標的的特異的な酵素は、C a s ヌクレアーゼである、請求項 1 から 12 のいずれか 1 つに記載の方法。

## 【請求項 14】

標的的特異的な酵素は、C a s 9 ヌクレアーゼである、請求項 1 から 13 のいずれか 1 つに記載の方法。

## 【請求項 15】

10

20

30

40

50

標的特異的な酵素は、ヌクレアーゼである、請求項 1 から 10 のいずれか 1 つに記載の方法。

【請求項 16】

前記ヌクレアーゼは、制限エンドヌクレアーゼである、請求項 1 から 15 のいずれか 1 つに記載の方法。

【請求項 17】

前記標的核酸は、デオキシリボ核酸である、請求項 1 から 16 のいずれか 1 つに記載の方法。

【請求項 18】

前記標的核酸は、リボ核酸である、請求項 1 から 16 のいずれか 1 つに記載の方法。

10

【請求項 19】

前記ポリメラーゼは、DNAポリメラーゼ、RNAポリメラーゼ、逆転写酵素、またはそれらの組み合わせである、請求項 1 から 18 のいずれか 1 つに記載の方法。

【請求項 20】

前記標的核酸は、ウイルスの核酸、バクテリアの核酸、病原体の核酸、一塩基多型、疾病関連遺伝子、癌関連遺伝子、ヒト白血球抗原遺伝子、またはそれらの組み合わせを含む、請求項 1 から 19 のいずれか 1 つに記載の方法。

【請求項 21】

ウイルスは、インフルエンザウイルス、コロナウイルス、ライノウイルス、ヘルペスウイルス、またはヒト免疫不全ウイルスである、請求項 20 に記載の方法。

20

【請求項 22】

前記コロナウイルスは、COVID-19ウイルスである、請求項 21 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

相互参照

本出願は、2020年1月17日に本願の米国仮特許出願第62/962,777号、及び2020年4月14日に本願の米国仮特許出願第63/009,603号の利益を主張するものであり、これらは各々、引用によりその全体が本明細書に組み込まれる。

【背景技術】

30

【0002】

核酸配列決定の進歩は、ますます手頃な価格で大量のサンプルが配列決定されることを可能にする。しかしながら、ただ1つまたは少量の標的配列を配列決定することが、データ処理の時間とコスト、配列決定のコストを削減するために、または、稀な事象に対する検出感度を増加させるために、望ましい場合がある。

【発明の概要】

【0003】

標的化配列決定パネルは、焦点化された配列決定、ならびに既知の遺伝子配列内の既知および未知の分子事象の様々な発見を可能にする、重要な研究および診断ツールである。応用は、癌ホットスポット配列決定、遺伝病、薬理ゲノム学、感染症、ならびに、ヒトおよび農業に関する生物種の両方における他の応用を含む。標的化または遺伝子座特異的デオキシリボ核酸(DNA)(およびリボ核酸(RNA))配列決定の主な正当化理由は、コストを下げ、かつ感度を増大させることである。

40

【0004】

これらの欠陥を排除することができる標的化配列決定の技術は、従って、非常に魅力的である。そのような方法は、以下の特性を有するに違いない：鑄型分子の断片化がないこと；RNAとDNAの両方に使えること；各標的セットのための広範囲な最適化を必要としないこと；ポリメラーゼ連鎖反応(PCR)の感度および速度を有すること；ハイブリッド捕捉の規模とコストを有すること；前もって定義された「対照(reference)」配列に向けて結果を片寄せないこと；すべての標的遺伝子座の均等なカバレッジを

50

提供すること；および、微小のサンプル供給源と多様なサンプル品質（すなわち、非常に断片化されている材料か、劣化されている材料か）から作業可能であること。

【0005】

いくつかの実施形態では、本明細書における方法は、実績のある R i p t i d e R a p i d ライブラリー調製の性能と、規則的に間が置かれてクラスター化された短い回文反復 ( c l u s t e r e d r e g u l a r l y i n t e r s p a c e d s h o r t p a l i n d r o m i c r e p e a t s ) ( C R I S P R ) に基づく標的化の特異性と速度とを組み合わせる。サンプルインプットは、RNA または DNA であり得る。鋳型分子は、熱変性され、5' サンプル・バーコードテール ( b a r c o d e t a i l ) を備えたランダムプライマーに接触され得る。高忠実度のポリメラーゼは、テンプレート配列の正確なコピーを作りながら、プライマーを 3' 端から伸長するために使用することができる。伸長された生成物は、少量のピオチン化デオキシヌクレオチド三リン酸 ( d d N T P ) を使用して終止され得る。この第 1 の工程は、同時に配列決定に適用可能な断片サイズを生成し、および親和性分子を付加しながら、分子にバーコード付けするために、使用することができる。精製は、末端のピオチン分子を、ストレプトアビジンでコーティングされた磁気ビーズに結合することにより生じる場合がある。この方法は、バーコード付きの生成物を単離することができる一方、その分子がクロスハイブリダイゼーションや他のテンプレート分子のさらなる 3' 伸長ができないようにすることを可能にし得る。このことは、また、可能な限り早期の段階における多重化サンプル処理を可能にし得る。複数のサンプルは、この最初のバーコード付けする工程の後に、単一の反応管の中へとプールされ得る。

10

20

【0006】

生成物がビーズ上で（またはコーティングされたマイクロウェル上でさえ）固定された後、それらは、ランダムプライミングの工程によって、二本鎖分子に変換され得る。場合によっては、この工程において、高忠実度の鎖置換ポリメラーゼが使用され、つまり、ビーズに最も近く結合したプライマーが、下流のすべての生成物を置換することができる。得られた分子組成物は、二本鎖のバーコード（平滑末端）を片側に、およびテンプレートの二本鎖コピーをビーズに結合させて、含み得る。場合によっては、ランダムプライマーの代わりに、プライマーゼと鎖置換ポリメラーゼを組み合わせ、この生成物を産生する。複数のサンプルは、こうして、このフォーマットにおいて存在することができ、全てが、反応における第 1 のプライミング工程から生成された固有のサンプルバーコードを備える。すべての過剰な反応物質は、洗い流され得る。標的化された遺伝子座について富化するために、エクソンに隣接した配列を標的とするガイド RNA 分子のプール、単一のヌクレオチド変異体 ( S N V )、または他の対象とする遺伝子標的は、直接または DNA テンプレートのインビトロの転写を通じて、設計および合成され得る。ガイド RNA 分子のプールは、リボヌクレオプロテイン ( R N P ) 複合体の多重化プールを形成するために C A S 9 または C A S 酵素誘導体と共にインキュベートされ得る。その後、これらの R N P 複合体は、先に記載された捕捉された生成物と共にインキュベートすることができ、標的分子の二本鎖切断を生じ、これがビーズから溶出される。非標的分子は、ビーズに結合されたままであり得る。生じたプール（溶出）は、一端にサンプルバーコード、そして他端に C R I S P R RNA 標的に対応する平滑末端標的部位を包含する、多重サンプルライブラリー由来の平滑末端標的配列を包含し得る。その後、これらの平滑末端生成物は、汎用アダプターのライゲーションによって標準ライブラリー構築に供され得る。場合によっては、プレート特異的なバーコードは、サンプルのより大きな多重化とさらなるコスト削減を可能にするために（指標読取位置に）追加される。

30

40

【0007】

この方法は：癌ホットスポット / 液体生検；遺伝子特異的 N I P T；遺伝病 / 常染色体上の劣性遺伝子標的 ( T a y S a c h s, C y s t i c F i b r o s i s など)；微生物の 16s 同定；ゲノム創薬 / 吸収、分布、代謝、排泄 ( A D M E ) 遺伝子標的；感染症標的；法医学マーカーなどの応用に使用することができ、および、任意の標的部位にカ

50

スタマイズされ得る。

【0008】

場合によっては、このアプローチは、遺伝子標的の多連試験または遺伝子型の確認を可能にする。例えば、ローパス配列決定およびインピュテーションは、ゲノムワイド関連解析 (GWAS) を行なうために使用される場合があり、および、確認試験は、ライブラリーの第2のセットを調製することを必要とせず、候補SNVまたは候補遺伝子配列のより深いカバレッジのためのサンプルライブラリー上で行なわれ得る。同じバーコードを付されたサンプルセットは、新しい発見がなされ、低コストの確認を必要とするとき、様々な遺伝子標的の連続テストのために使用され得る。

【0009】

本明細書において、標的核酸を含む配列決定ライブラリーを作製する方法が提供される。いくつかの場合では、方法は：(a) 核酸サンプルをプライマーの第1の集団、ポリメラーゼ、デオキシリボヌクレオチド三リン酸 (dNTPs)、および標識された ddNTP に接触させる工程；(b) 伸長反応を実行する工程であって、それによって標識された伸長生成物を作製する、工程；(c) 二本鎖伸長生成物を作製するために前記伸長生成物をプライマーの第2の集団に接触させる工程；(d) 二本鎖伸長生成物の少なくともサブセットの切断を可能にする条件下で、前記二本鎖伸長生成物を標的特異的酵素に接触させる工程であって、それにより、切断された標的核酸を作製する、工程；および、(e) 切断された標的核酸を単離する工程を含む。場合によっては、核酸サンプルライブラリーは、変性される。場合によっては、プライマーの第1の集団または第2の集団は、ランダムプライマーを含む。場合によっては、プライマーの第1の集団または第2の集団は、バーコードを含む。場合によっては、プライマーの第1の集団または第2の集団は、オリゴ-dTプライマーを含む。場合によっては、標識された ddNTP はビオチンを含む。場合によっては、標識された伸長生成物は、ビーズ上で単離される。場合によっては、ビーズはストレプトアビジン・ビーズである。場合によっては、ビーズは磁気ビーズである。場合によっては、標識された伸長生成物は、ポリカーボネート、ポリプロピレン、またはポリカーボネート表面上で単離される。場合によっては、標的特異的酵素は、標的特異的なガイドRNAと複合体化される。場合によっては、標的特異的な酵素はCRISPR酵素である。場合によっては、標的特異的な酵素は、Casヌクレアーゼである。場合によっては、標的の特異的な酵素は、Cas9ヌクレアーゼである。場合によっては、標的特異的な酵素は、ヌクレアーゼである。場合によっては、標的核酸は、デオキシリボ核酸である。場合によっては、標的核酸は、リボ核酸である。場合によっては、ポリメラーゼは、DNAポリメラーゼ、RNAポリメラーゼ、逆転写酵素、またはそれらの組み合わせである。場合によっては、標的核酸は、ウイルスの核酸、バクテリアの核酸、一塩基多型、疾病関連遺伝子、癌関連遺伝子、ヒト白血球抗原遺伝子、またはそれらの組み合わせを含む。場合によっては、ウイルスは、インフルエンザウイルス、コロナウイルス、ライノウイルス、ヘルペスウイルス、またはヒト免疫不全ウイルスである。場合によっては、コロナウイルスは、COVID-19ウイルスである。場合によっては、標的核酸は、約5kb~約20kbである。場合によっては、標的核酸は、約10kb~約15kbである。

【0010】

参照による組み込み

本明細書で言及される刊行物、特許、および特許出願はすべて、あたかも個々の刊行物、特許、または特許出願がそれぞれ参照により引用されるように具体的かつ個々に指示されるのと同じ程度にまで、参照により本明細書に引用される。

【図面の簡単な説明】

【0011】

特許または出願ファイルは、カラーで実行された少なくとも1つの図面を包含している。カラーの図面を有するこの特許または特許出願公開のコピーが、必要な料金の請求および支払い後に当該事務局によって提供される。

【0012】

10

20

30

40

50

本発明の特徴と利点を良く理解するには、本発明の原理が用いられる例示的实施形態を説明する以下の詳細な説明と添付図面とを参照されたい。

【0013】

【図1】 標的化配列決定の方法を図示する。

【図2】 単管のHCoV2を標的とした配列決定法を図示する。

【図3】 大腸菌ゲノムの標的化配列決定を図示する。

【図4】 標的化配列決定のロングリードおよびショートリードの応用を図示する。

【図5】 標的化配列決定のアダプターおよびバーコードの応用を図示する。

【図6】 多重標的分子の生成を図示する。

【図7】 配列に対応したライブラリー生成のための小さな断片を図示する。

10

【発明を実施するための形態】

【0014】

最も一般に使用される標的化配列決定アプローチとして、PCRまたはPCR類似の派生、およびハイブリダイゼーションに基づくキャプチャープロトコルが挙げられる。両方の技術は、最初の非常に並列化されたシーケンサーが商業化されるかなり前に開発された。過去20年間にわたり、これらのアプローチの速度およびコストにおいて劇的な進歩が得られたが、それらは、恐らく革新の限界に達した。従って、両方の技術は高価であり且つ時間を消費する最適化を必要とするため、標的化配列決定へのこれらの時代遅れのアプローチを置き換える新技術が必要とされる。

【0015】

20

PCRとPCRの派生は、優れた特異性と感度に加えて、比較的速いサンプル処理時間の利点を有する。しかしながら、これらのアプローチは、多重化するとき、高価であり、大規模な最適化を必要とし、および、煩わしいポスト増幅正規化を必要とする。加えて、これらのアプローチは、2つの反対鎖のプライマーを使用する時、高い中断率を有し、同じ遺伝子座の冗長なサンプリングが可能でなく、および新規な再配列を検出するか発見する能力を欠く。ポリメラーゼエラーのクローン増幅もまた、大きな問題である。これらの制限のうちいくつかを克服する戦略として、クローンのポリメラーゼエラーを検出するための、固有分子指標(Unique Molecular Indices)(UMI)またはバーコードの使用、ならびに再配列を検出するための、半特異的(hemi-specific)もしくは「ネストされた」アプローチが挙げられる。後者は、半特異的アプローチに先立って、フラグメンテーションを使用するライブラリーを作製することを必要とする。これらの方法はまた、多くの最適化を必要とし、ロングリード配列決定システムには向かない。さらに、プライマリー部位の下の一塩基多型(SNP)は、しばしば脱落を起こす。加えて、対立遺伝子のバイアスは、カバレッジ(各アンプリコンの表現)の均一性の不足と共に、問題である。

30

【0016】

ハイブリッド捕捉と呼ばれるハイブリダイゼーションに基づくアプローチは、新規な構造の再配列を発見する能力を与え、および、同時に配列決定することができる標的の数の点で規模の増加を与える。しかしながら、ハイブリッド捕捉アプローチには、まだ多数の欠陥がある。現在まで、すべてのハイブリッド捕捉アプローチは、断片化工程で始まる。この工程は、鋳型分子が化学的または物理的な断片化によって引き起こされる損傷のために失われることになるため、情報消失を生じる。最も効率的な断片化プロトコルの後、最もよい場合で、個々の核酸分子の60%が保持される。加えて、ハイブリッド捕捉アプローチについては、ハイブリダイゼーションプローブが最初に設計されている配列に対する対照対立遺伝子バイアスが報告されている。キャプチャー効率は、配列の内容の違いのために幅広くばらつきがあり、そのため、臨床グレード配列決定に必要とされる最小のカバレッジに達するためには、サンプルの過剰配列決定(over sequencing)を必要とする。

40

【0017】

実際、University of Texas Southwesternの研究者

50

による2020年1月の刊行物は、主要な商用エクソーム配列決定提供者のうちの3つが重要な遺伝子をカバーしないことを明かした。「3つの主な米国の研究所の臨床試験の調査は、全エクソーム配列決定が、DNAの大きなセグメントを十分に解析することに常習的に失敗すること、医師が癩癩から癌におよぶ潜在的な遺伝障害を正確に診断することを妨害し得る潜在的に重要な欠陥を示した。」UT Southwesternによる再分析は、各研究室が、平均的で、遺伝子の4分の3未満 - 34、66、および69パーセントのカバレッジ - を適切に検査し、および特定の障害を検出するそれらの能力において、驚くほどに広いギャップを有していた、と説明する。研究者は、試験品質における広範な差異は臨床遺伝学の配列決定において特有であるが、十分報告されてこなかった、あるいは臨床医の間で共有されてこなかったと信じ、当該研究を行なったと述べる。(出典: <https://www.sciencedaily.com/releases/2020/01/2020106103451.htm>)

10

## 【0018】

ハイブリッド・キャプチャー戦略は、多重化の増加(多数の標的)を可能にし、クローンの増幅エラーを回避する。しかしながら、ワークフローは長く、煩わしく、および、キャプチャーはあまり特異的ではない。この方法は、また対立遺伝子バイアスと脱落の影響を受け、ロングリード配列決定のために理想的ではない。

## 【0019】

理想的な標的化配列決定戦略は無制限の多重化、サンプル中のすべての標的カバレッジ/表現の均一性を可能にし、およびロングリードとショートリードの配列決定技術において機能するだろう。また、それは初期設計において高転換率を有するだろう。

20

## 【0020】

現在、標的化配列決定技術は、主としてアンプリコンベース、またはハイブリダイゼーションベースのいずれかである。アンプリコンベースの戦略は、一般にPCRまたはPCRの派生を使用する。これらは高感度で、相対的に速いプロトコルであり得る。しかし、プライマー設計の制約(すなわち、アンプリコンの重複がない)のため、多重化能力に限界がある場合が多い。それはまた、一般に多くの最適化を必要とし、ロングリード配列決定システムには向かない。さらに、プライマー部位の下でSNPは脱落を起こす場合がある。加えて、クローンのポリメラーゼエラーと対立遺伝子のバイアスは問題となり得る。増幅ベースの標的化配列決定における別の問題は、カバレッジ(すなわち、各アンプリコンの表現)の均一性である。

30

## 【0021】

ハイブリダイゼーション捕捉戦略は、増加された多重化(多数の標的)を可能にし、クローンの増幅エラーを回避する。しかしながら、ワークフローは長く煩わしいものになり得る。具体的に、この方法では、特異性、対立遺伝子バイアス、および脱落、さらにポーズが問題となる。ハイブリダイゼーション法も、ロングリード配列決定のために理想ではない。

## 【0022】

理想的な標的化配列決定戦略は、無制限の多重化を可能にし、サンプル中のすべての標的のカバレッジと表現の均一性を提供し、およびロングリードとショートリードの配列決定技術において機能し、および初期設計上で高転換率を有するだろう。

40

## 【0023】

## 定義

関連する定義の部分的なリストは以下の通りである。

## 【0024】

本明細書で使用される「増幅された核酸」または「増幅されたポリヌクレオチド」は、その量が、その開始量と比較して、インピットロで実行される任意の核酸増幅または複製方法の少なくとも2倍増加された、任意の核酸またはポリヌクレオチド分子である。例えば、増幅された核酸は、ポリメラーゼ連鎖反応(PCR)から得られ、いくつかの事例では、DNAを指数関数的に増幅することができる(例えば、nサイクルで $2^n$ コピーに増幅

50

)。増幅された核酸は、線形の増幅から得ることもできる。

【0025】

本明細書に使用される「増幅産物」は、ポリメラーゼ連鎖反応などの増幅反応によって生じる生成物を指し得る。

【0026】

本明細書で使用される「アンプリコン」は、天然または人工の増幅または複製事象の源および/または産物であるポリヌクレオチドまたは核酸である。

【0027】

本明細書で使用される用語「生体サンプル」または「サンプル」は、一般に、生物学的存在から分離されたサンプルまたは部分を指す。生体試料は、全体の性質を示す場合があり、例としては、体液、解離した腫瘍標本、培養細胞、およびそれらの任意の組み合わせが挙げられるが、これらに限定されない。生体試料は1つ以上の個体由来であり得る。1つ以上の生体試料が同じ個体由来であり得る。1つの限定しない例において、第1のサンプルは個体の血液由来であり、第2のサンプルは個体の腫瘍生検標本由来であり得る。生体サンプルの例としては、限定されないが、血液、血清、血漿、鼻スワブ(nasal swab)または鼻咽頭洗浄液、唾液、尿、胃液、髄液、涙、便、粘液、汗、耳垢、油、腺分泌物、脳脊髄液、組織、精液、腔液、腫瘍組織由来の間質液を含む間質液、眼液、髄液、咽頭スワブ、息、毛、指の爪、皮膚、生検標本、胎盤液、羊水、臍帯血、強調液(emphatic fluids)、腔液、痰、膿、細菌叢、胎便、母乳および/または他の分泌物が挙げられる。サンプルは、鼻咽頭洗浄液を含み得る。被験体の組織サンプルの例としては、限定されないが、結合組織、筋組織、神経組織、上皮組織、軟骨、癌または腫瘍のサンプル、あるいは骨が挙げられる。サンプルはヒトまたは動物から提供され得る。サンプルは、ネズミ、サル、ヒト、家畜、スポーツ動物、またはペットなどの脊椎動物を含む哺乳動物から提供され得る。サンプルは、生きているか又は死んでいる被験体から収集され得る。サンプルは、被験体から新鮮に得られるか、あるいは前処理、保存、または輸送の何らかの形態を経ている場合がある。

10

20

【0028】

本明細書で使用される「体液」は、一般に被験体の身体から生じる流体または分泌物を記載し得る。いくつかの事例では、体液は、一緒に混合された1つを超えるタイプの体液の混合物である。体液の幾つかの限定しない例としては、血液、尿、骨髄、髄液、胸膜液、リンパ液、羊水、腹水、痰、またはそれらの組み合わせが挙げられる。

30

【0029】

「相補的」または本明細書に使用されるような「相補性」は、塩基対合によって関連づけられる核酸分子を指すことができる。相補的なヌクレオチドは、一般に、AおよびT(またはAおよびU)、あるいはCおよびG(またはGおよびU)である。最適にアラインメントされた、および適切なヌクレオチドの挿入または欠失が伴う、1つの鎖のヌクレオチドが、少なくとも約90%から約95%またはそれ以上の相補性、およびより好ましくは約98%から約100%の相補性、およびさらにより好ましくは100%の相補性を有して対になるときに、2つの一本鎖RNAまたはDNA分子は、実質的に相補的であると言われる。代替的に、RNAまたはDNAの鎖が選択的なハイブリダイゼーション条件下でその補体にハイブリダイズするときに、相当な相補性が存在する。選択的なハイブリダイゼーション条件は、限定されないが、ストリンジェントなハイブリダイゼーション条件を含む。ハイブリダイゼーション温度は、一般に、少なくとも約2 から約6 融解温度( $T_m$ )より低い。

40

【0030】

「バーコード」または「分子バーコード」は、本明細書では、標識化のための物質である。バーコードは、核酸またはポリペプチドなどの分子を標識することができる。標識化のための材料は、情報に関連づけられる。バーコードは、配列識別子(例えば、配列ベースのバーコードまたは配列指標)と呼ばれる。バーコードは、特定のヌクレオチド配列であり得る。バーコードは、識別子として使用することができる。バーコードは、異なる

50

サイズの分子、または同じ分子の異なる終了点であり得る。バーコードは、分子内の特異的な配列および異なる終了配列を含み得る。例えば、同じプライマーから増幅され且つ25ヌクレオチド位置を有する分子は、増幅され且つ27ヌクレオチド位置を有する分子とは異なる。27mer配列の付加位置はバーコードとみなし得る。バーコードは、ポリヌクレオチドに組み込まれ得る。バーコードは、多くの方法によってポリヌクレオチドに組み込まれる。バーコードを組み込むためのいくつかの非限定的な法は、分子生物学的方法を含み得る。バーコードを組み込むための分子生物学的方法のいくつかの非限定的な例は、プライマー（例えば、テールのプライマー伸長）、プローブ（即ち、ライゲーションによるプローブへの伸長）、またはライゲーション（即ち、分子への既知の配列のライゲーション）による方法である。

10

## 【0031】

バーコードは、ポリヌクレオチドの任意の領域の中に組み込まれ得る。前記領域は、既知であり得る。あるいは、前記領域は、未知であり得る。バーコードは、ポリヌクレオチドにわたって任意の位置に付加され得る。場合によっては、バーコードは、ポリヌクレオチドの5'端に付加され得る。あるいは、バーコードは、ポリヌクレオチドの3'端に付加され得る。バーコードは、ポリヌクレオチドの5'端と3'端との間に付加され得る。場合によっては、バーコードは、1つ以上の他の既知の配列を用いて付加され得る。1つの非限定的な例として、配列アダプターを用いたバーコードの付加がある。

## 【0032】

バーコードは、情報に関連づけられ得る。バーコードが情報に関連付けられる情報のタイプのいくつかの限定しない例は、次のものを含む：サンプルの供給源；サンプルの配向；サンプルが処理された領域または容器；隣接したポリヌクレオチド；またはそれらの任意の組み合わせ。

20

## 【0033】

場合によって、バーコードは、（コンビナトリアルバーコードとは異なる）配列の組み合わせから作られ、サンプルまたはゲノム座標および鎖の分子標識およびコピーが得られる異なる鋳型分子または一本鎖を同定するために使用される。場合によって、各生体分子に対するサンプル識別子、ゲノム座標、および特異的標識が、一緒に増幅され得る。バーコード、合成コード、または標識情報はまた、（エラーまたはエラー訂正を可能にする）コードの配列構成、コードの長さ、コードの方位、分子内のコードの位置から、および他の天然または合成のコードと組み合わせ、入手され得る。

30

## 【0034】

バーコードは、サンプルのプーリング（pooling）前に付加され得る。プールされたサンプルの配列が決定される時、バーコードは、ポリヌクレオチドの残りと共に配列決定され得る。場合によっては、バーコードは、配列決定された断片をサンプルの供給源に関連づけるために使用される。

## 【0035】

バーコードは、代替的に、撚り（strandedness）のサンプルを同定するために使用され得る。1つ以上のバーコードが一緒に使用され得る。2つ以上のバーコードは、互いに隣接しているか、隣接していないか、またはその組み合わせであり得る。

40

## 【0036】

場合によっては、バーコードはコンビナトリアル標識化のために使用される。

## 【0037】

本明細書で使用される「コンビナトリアル標識化（Combinatorial labeling）」は、2以上のバーコードが使用される方法である。2つ以上のバーコードがポリヌクレオチドを標識し得る。バーコードは、それぞれ、単独で情報に関連づけられる。バーコードの組み合わせは、一緒に情報に関連づけられ得る。場合により、バーコードの組み合わせは、増幅が、サンプル鋳型の合成コピーからではなく、元のサンプル鋳型から生じたことをランダムに増幅された分子において判定するために、一緒に使用される。場合により、別のバーコードの配列と組み合わせた1つのバーコードの長さは、ポリ

50

ヌクレオチドを標識するために使用される。場合により、別のバーコードの配向と組み合わせた1つのバーコードの長さは、ポリヌクレオチドを標識するために使用される。他の場合では、1つのバーコードの配列は、ポリヌクレオチドを標識するために別のバーコードの配向とともに使用される。場合により、第1および第2のバーコードの配列は、それらの間のヌクレオチドの距離との組み合わせにおいて、ポリヌクレオチドを標識または同定するために使用される。

**【0038】**

本明細書に使用される「二本鎖」は、逆相補的な配向などで、相補的塩基対合によってアニールされた2つのポリヌクレオチド鎖を指す。

**【0039】**

本明細書に使用される「既知のオリゴヌクレオチド配列」または「既知のオリゴヌクレオチド」あるいは「既知の配列」は、知られているポリヌクレオチド配列を指す。既知のオリゴヌクレオチド配列は、設計されたオリゴヌクレオチド、例えば、次世代型配列決定プラットフォームのためのユニバーサルプライマー（例えば、Illumina、454）、プローブ、アダプター、タグ、プライマー、分子バーコード配列、識別子に対応し得る。既知の配列は、プライマーの部分を含み得る。既知のオリゴヌクレオチド配列は、特定のユーザーに実際には知られていないが、例えば、コンピューターによって利用可能なデータとして保存されることによって構造的に知られている。既知の配列は、また、実際に知られていない通商上の秘密または1人以上のユーザーにとって秘密であるが、ユーザーが使用している実験の特定のコンポーネント、キット、装置、またはソフトウェアを設計した実体に知られている場合がある。

**【0040】**

本明細書で使用される「ライブラリー」は、核酸のコレクションを指し得る。ライブラリーは1つ以上の標的断片を含み得る。いくつかの例では、標的断片は増幅された核酸である。他の例では、標的断片は増幅されていない核酸である。ライブラリーは、3'末端、5'末端、または3'末端と5'末端の両方に加えられた1つ以上の既知のオリゴヌクレオチド配列を有する核酸を含有し得る。ライブラリーは、断片が、ライブラリーの供給源を特定する既知のオリゴヌクレオチド配列を含有するように調製され得る（例えば、患者またはDNA源を同定する分子同定バーコード）。いくつかの例では、2つ以上のライブラリーが、ライブラリープール（library pool）を作製するためにプールされる。ライブラリーはまた、当該技術分野で知られているような、トランスポゾン媒介性の標識化、または「タグ付け（tagmentation）」などの、他のキットおよび技術を用いて生成され得る。Illumina NEXTERAキット（Illumina, San Diego, CA）などのキットは、市販で入手可能である。

**【0041】**

本明細書で使用される「遺伝子座特異的な（Locus specific）」または「遺伝子座特異的な（loci specific）」は、核酸分子中の位置（例えば、染色体またはゲノム内の位置）に対応する1つ以上の遺伝子座を指し得る。いくつかの例では、遺伝子座は遺伝子型に関連づけられる。いくつかの例では、遺伝子座は、例えば、ハイブリダイゼーションおよび/または他の配列ベースの技術に基づいて、サンプルから直接単離され、富化されるか、あるいは、配列の検出前に鋳型としてサンプルを使用して選択的に増幅される。いくつかの例では、遺伝子座は個体間のDNAレベルの変異に基づいて、特定の染色体に対する特異性に基づいて、選択された遺伝子座のCG内容および/または必要とされる増幅条件に基づいて、または当業者にとって本開示を読むことで明白となる他の特徴に基づいて、選択される場合がある。遺伝子座はまた、特異的なゲノム座標またはそのゲノムの対照配列によって明示されるようなゲノムにおける位置を指す場合がある。

**【0042】**

本明細書で使用される「長い核酸」は、1、2、3、4、5、6、7、8、9、または10キロベースより長いポリヌクレオチドを指す。

10

20

30

40

50

## 【0043】

本明細書で使用される用語「融解温度」または「 $T_m$ 」は、通常、二本鎖核酸分子の集団が一本鎖へと半分に解離される温度を指す。核酸の $T_m$ を計算する等式は、当該技術分野において周知である。 $T_m$ 値の単純な概算を与える1つの方程式は次のとおりであり： $T_m = 81.5 + 16.6 (\log_{10} [Na^+]) - 0.41 (\% [G + C]) - 67.5 / n - 1.0m$ 、この時、核酸は0.5 M未満のカチオン濃度を有する水溶液中にあり、(G + C)の内容は、30% ~ 70%の間にあり、nは塩基の数であり、および、mは不適正塩基対のパーセンテージである（例えば、Sambrook J et al., *Molecular Cloning, A Laboratory Manual*, 3rd Ed., Cold Spring Harbor Laboratory Press (2001)を参照されたい）。他の参考文献は、より精巧な計算を含んでもよく、それらは構造ならびに配列の特徴を $T_m$ の計算のために考慮に入れる。

10

## 【0044】

本明細書で使用される「ヌクレオチド」は、塩基 - 糖 - リン酸の組み合わせを指し得る。ヌクレオチドは、核酸配列（例えば、DNAとRNA）の単量体単位である。用語ヌクレオチドは、例えば、自然発生および非自然発生のリボヌクレオシド三リン酸ATP、TTP、UTP、CTG、GTP、およびITP、およびdATP、dCTP、dITP、dUTP、dGTP、dTTP、またはそれらの誘導体などのデオキシリボヌクレオシド三リン酸を含む。そのような誘導体は、例えば、[aS]dATP、7-デアザ-dGTPおよび7-デアザ-dATP、および例えば、それらを含む核酸分子に対するヌクレアーゼ耐性を与えるヌクレオチド誘導体を含む。本明細書で使用される用語ヌクレオチドはまた、ジデオキシリボヌクレオシド三リン酸（ddNTP）およびそれらの誘導体を指す。ジデオキシリボヌクレオシド三リン酸の実例は、例えば、ddATP、ddCTP、ddGTP、ddITP、ddUTP、およびddTTPを含む。dd（2-6ジアミノ）プリンなど、他のddNTPは企図され、および、本明細書の開示に矛盾しない。

20

## 【0045】

本明細書で使用される「ポリメラーゼ」は、別の鎖を鋳型として使用して、個別のヌクレオチドを一緒に鎖へと連結する酵素を指す。

## 【0046】

「ポリメラーゼ連鎖反応」または「PCR」は、過剰な非特異的DNAの存在下でさえ、選択されたDNAの特定の断片をインビトロで複製するための技術を指す。選択されたDNAにプライマーが加えられ、そこでプライマーは、ヌクレオチドおよび典型的にTaqポリメラーゼなどを使用して、選択されたDNAのコピーを開始する。温度を周期させることによって、選択されたDNAは、反復して変性させられ、およびコピーされる。選択されたDNAの単一のコピーは、他のランダムなDNAと混合されたとしても、何千、何百万、または何十億もの複製を得るために増幅される。ポリメラーゼ連鎖反応は、非常に少量のDNAを検出および測定するために、ならびにDNAのカスタマイズされた断片を作製するために、使用される。

30

## 【0047】

本明細書で使用される用語「ポリヌクレオチド」または「核酸」は、限定されないが、様々なDNA、RNA分子、それらの誘導体または組み合わせを含み得る。これらは、dNTPs、ddNTP、2メチルNTP、DNA、RNA、ペプチド核酸、cDNA、dsDNA、ssDNA、プラスミドDNA、コスミドDNA、染色体DNA、ゲノムDNA、ウイルスDNA、細菌DNA、mtDNA（ミトコンドリアDNA）、mRNA、rRNA、tRNA、nRNA、siRNA、snRNA、snoRNA、scarNA、マイクロRNA、dsRNA、リボザイム、リボスイッチ、およびウイルスRNAなどの種を含み得る。「オリゴヌクレオチド」は、一般に、プライマーとして使用するために適している長さのポリヌクレオチドであり、一般に約6 ~ 50塩基であるが、特により長い例外も稀ではない。

40

50

## 【 0 0 4 8 】

本明細書で使用される「プライマー」は、ポリメラーゼ連鎖反応の合成工程、または一定の配列決定反応で使用されるプライマー伸長技術などにおいて、例えば、ヌクレオチドの伸長、ライゲーション、および/または合成をプライミングするために使用されるオリゴヌクレオチドを一般的に指す。プライマーは、代替的に、特異的な核酸領域の検出のための捕捉オリゴヌクレオチド (capture oligonucleotide) に対する遺伝子座の相補性を提供する手段としてハイブリダイゼーション技術に使用される。

## 【 0 0 4 9 】

本明細書で使用される「プライマー伸長産物」は、鋳型として連続したポリヌクレオチドを使用するプライマー伸長反応から生じる産物、および連続した配列に対する相補的な又は部分的に相補的なプライマーを一般的に指す。

## 【 0 0 5 0 】

本明細書で使用される「配列決定 (Sequencing)」、「配列決定 (sequence determination)」などは、核酸中のヌクレオチド塩基の順序を決定するために使用され得る全ての生化学的方法を一般的に指す。

## 【 0 0 5 1 】

本明細書に使用される「配列」は、核酸分子における隣接した核酸塩基の相対順序を反映し、および、必ずしもその核酸分子に固有にではないが、容易に特異的に同定され得る、一連の順序付けられた核酸塩基を指す。一般に、すべての場合ではないが、1つの配列は、情報を与えるために、5以上などの、複数の核酸塩基を必要とするが、この数はコンテキストによって変化し得る。従って、制限エンドヌクレアーゼは、たとえその配列がわずか4つの塩基でも、それが同定し、特異的に切断するならば「配列」を有すると言及される場合がある。配列は、サンプルの断片に対して「固有にマッピングする」必要はない。しかしながら、ほとんどの場合、配列は、その分子の出所に関して情報を与えるために十分な情報を含んでいると考えてよい。

## 【 0 0 5 2 】

本明細書に使用されるように、あるサンプルの全配列においてその配列が隣接して現れない場合、配列は「生じない」。あるサンプルにおいて生じない配列は、そのサンプルにおける自然発生の配列ではない。

## 【 0 0 5 3 】

本明細書に使用されるように、ライブラリーがサンプルの情報を与える配列を含む場合、ライブラリーは「サンプルの代表」として説明される。場合によっては、情報を与える配列は、サンプル配列の約5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または100%を含む。場合によっては、情報を与える配列は、サンプル配列の約90%、90%、または90%以上を含む。

## 【 0 0 5 4 】

本明細書に使用されるように、配列または配列長は、配列または配列長が第2の配列または配列長により、決定されるのではない、または第2の配列または配列長の関数ではない場合、「独立して判定される」と説明される。終止 ddNTP 塩基の取り込み、またはオリゴの鋳型に対する非特異的または正確でないアニーリングなどのランダムな事象は、一般に、そのような事象から生じる分子のライブラリーが配列または配列長に本質的な変異を含むように独立して判定される事象である。

## 【 0 0 5 5 】

本明細書に使用されるように、配列は、鋳型で媒介される合成によって判定されない場合、「不確定の」ものとして説明される。従って、ランダムなオリゴマーの鋳型へのアニーリングによってプライミングされた鋳型からの合成を起源とする核酸分子は、鋳型で駆動された核酸伸長から生じる、鋳型に方向づけられた配列の領域と、鋳型に駆動された伸長反応が構築される場所である3'OH基を提供するオリゴマー配列に対応する「不確定

10

20

30

40

50

の配列」とを含む場合がある。場合によっては、オリゴヌクレオチド・アニーリングは、当該オリゴマー配列が結合する分子の正確な逆相補ではないように、不完全である。

【0056】

本明細書に使用される「細分化」は、サンプル配列の文脈において、部分配列へと配列を分解することであり、ここで分解後の各々が依然として本明細書に定義される配列であるように、分解することを指す。

【0057】

本明細書で使用される「コンティグ」は、配列相同性の共通の又は重複する領域を共有する2つ以上の構成ヌクレオチド配列から構築されるヌクレオチド配列を指す。例えば、2つ以上の核酸断片のヌクレオチド配列が比較され、共通の又は重複する配列を同定するために整列される。共通の又は重複する配列が、2つ以上の核酸断片間に存在する場合、配列（およびそれ故それらの対応する核酸断片）は、単一の連続したヌクレオチド配列へと構築される。

10

【0058】

用語「ピオチン」は、本明細書に使用されるように、ピオチン（5 - [（3 a S , 4 S , 6 a R） - 2 - オキソヘキサヒドロ - 1 H - チエノ [ 3 , 4 - d ] イミダゾール - 4 - イル] ペンタン酸）およびあらゆるピオチン誘導体およびアナログを指すように意図されている。そのような誘導体およびアナログは、ネイティブな又は修飾されたストレプトアビジンまたはアビジンのピオチン結合ポケットとの複合体を形成する物質である。そのような化合物は、例えば、イミノピオチン、デスチオピオチンおよびストレプトアビジンの親和性ペプチドを含み、ピオチン - . イプシロン . - N - リシン、ピオシチンヒドラジド、2 - イミノピオチンおよびピオチニル - . アミノカブロン酸 - N - ヒドロキシスクシンイミドエステルのアミノまたはスルフヒドリルの誘導体、スルフォ - スクシンイミド - イミノピオチン、ピオチンプロモアセチルヒドラジド、p - ジアゾベンゾイルピオシチン、3 - ( N - マレイミドプロピオニル ) ピオシチンも含む。「ストレプトアビジン」は、ピオチンに結合することができるタンパク質またはペプチドを指し：天然の卵白アビジン、組み換えのアビジン、非糖鎖形成された形態のアビジン、バクテリアのストレプトアビジン、組み換えのストレプトアビジン、切り縮められたストレプトアビジン、および/またはそれらの任意の誘導体を含む。

20

【0059】

本明細書に使用される「被験体」は、複製することができるゲノムを有する、現在生存している生物またはかつて生きていた又は実体であった生物を一般的に指す。本開示の方法、キット、および/または組成物は、限定されないが、細菌および酵母などの微生物；ハエ、甲虫、およびハチを含み、これらに限定されない昆虫；トウモロコシ、コムギ、海藻、または藻類を含み、これらに限定されない植物；および、限定されないが、ヒト；マウス、ラット、サル、およびチンパンジーなどの、実験動物；イヌおよびネコなどの飼育哺乳動物；ウシ、ウマ、ブタ、ヒツジ、ヤギなどの農業動物；および、パンダ、ライオン、トラ、クマ、ヒョウ、ゾウ、シマウマ、キリン、ゴリラ、イルカ、およびクジラなどの野生動物を含む、動物を含む、1以上の単細胞または多細胞の被験体に適用される。本開示の方法は、ウイルスまたはウイルス粒子、あるいは1つ以上のウイルスによって感染された1つ以上の細胞などの、病原菌または感染病原体にも適用される。

30

40

【0060】

本明細書で使用される「支持体 ( support )」は、固形物、半固形物、ビーズ、表面である。支持体は、溶液中で可動であるか又は非可動である。

【0061】

本明細書で使用される用語「一意の識別子 ( unique identifier )」は、限定されないが、分子バーコード、または d U T P などの、混合物中の核酸のパーセンテージを含み得る。

【0062】

本明細書に使用される「反復配列」は、核酸配列データセットにおいて単一の位置に一

50

意にマッピングされない配列を指す。一部の反復配列は、所与のサイズの、および正確な又はおよその配列の、反復単位の整数または分数倍数として概念化される。

【 0 0 6 3 】

本明細書に使用される「プライマー」は、鋳型分子にアニール化されて、その場所で鋳型に方向づけられた核酸合成が生じることが出来る 3' OH基を提供する、オリゴヌクレオチドを指す。プライマーは、多くの場合、修飾されないデオキシ核酸を含むが、場合により、リボ核酸または 2' メチルリボ核酸などの修飾された核酸などの、代替の核酸を含む。

【 0 0 6 4 】

本明細書に使用されるように、核酸は、水素結合した塩基対合を含む場合、二本鎖である。二本鎖と呼ぶには、分子内のすべての塩基が対になっている必要はない。

【 0 0 6 5 】

本明細書で数字に関連して使用される用語「約」は、10%までの数を指す。範囲に関して使用される用語は、明示された下限の10%程度下の下限を有し、および明示された限界の10%以内で上の上限数を有する範囲を指す。

【 0 0 6 6 】

標的化配列決定法が本明細書に提供される。簡潔には、ある場合において、二本鎖 DNA または cDNA の生成物は固体基質に結合される。CRISPR-CAS システムは、標的分子を「切る」ために使用することができ、その後、物理的に固体基質に対して結合されたままであるバックグラウンド分子から分離することができる。その後、標的分子は、配列決定ライブラリーに変換されて配列決定され得る。

【 0 0 6 7 】

いくつかの RNA に標的化配列決定法において、RNA の合計はサンプルから得ることができる。場合によっては、ランダムプライマーがサンプル RNA 鋳型に結合される。逆転写酵素は RNA 鋳型の cDNA (コピー DNA) を作製するために使用され得る。場合により、dNTP とピオチン ddNTP の混合物が伸長反応に使用され、場合により、cDNA 生成物をピオチン ddNTP で終止させる。場合によっては、その後、生成物はストレプトアビジンでコーティングされた磁気ビーズ上でキャプチャーされ、および、過剰反応物質が洗い流される。場合によっては、第 2 鎖は、ランダムプライマーと鎖置換ポリメラーゼを使用して、捕捉された分子上で合成される。ビーズに最も近く結合したプライマーは、下流の生成物を置換することができ、二本鎖 cDNA 生成物は、ビーズに結合したまま残る。CRISPR ガイド RNA は、関心の標的配列を認識するように設計することができ、および、ビーズに結合された生成物は、RNA でガイドされる切断に供される。いくつかの場合では、CRISPR 切断生成物はビーズに結合されたバックグラウンドから磁気分離および溶出まで分離される。二本鎖 cDNA 生成物は、ビーズから溶出されて配列決定ライブラリーに変換され得る。

【 0 0 6 8 】

いくつかの DNA 標的化配列決定法では、上記されたものと同じ方法が使用される。場合によっては、二本鎖 DNA 生成物は変性され、および DNA ポリメラーゼは鋳型を調製するために使用される。

【 0 0 6 9 】

場合によっては、RNA と DNA は、RNA と DNA の鋳型を使用することができるポリメラーゼ酵素と同時に調製され得る。

【 0 0 7 0 】

標的化配列決定法

付加的な態様では、標準の RIPTIDE A および B の反応を使用する標的化配列決定法が提供される。この方法では、DNA サンプル (または cDNA) は、バーコード付きランダムプライマーを使用してポリメラーゼと共にコピーされ、ピオチン-ddNTP で終止される。場合によっては、サンプルは、ストレプトアビジンでコーティングされた磁気ビーズ上にプールされ (96 以上) およびキャプチャーされる。場合によっては、第

10

20

30

40

50

2鎖は、ポリメラーゼを置換する鎖を介して形成される。場合によっては、活性のある C R I S P R - C A S は、s g R N A によって標的化されて、標的化分子をビーズから切り離す。場合によっては、分子は、溶出され、そして、ライブラリーは、標準ライゲーションに基づくプロトコルを用いて生成される。

#### 【0071】

図1に図示されるのは、本明細書に提供される、別の標的化配列決定法である。図1では、プライマー伸長が、バーコード付きランダムプライマーを用いて行なわれる（これらの「A」プライマーは、Illuminaアダプター配列などのアダプター配列を包含していない）。プライマー伸長生成物は、ビオチン d d N T P で終止され、ストレプトアビジン・ビーズに結合される。その後、ランダムプライマーである「B」プライマーは、二本鎖 DNA を形成する相補鎖を生成するために使用される。関心の特異的な遺伝子座を標的とするように設計された Cas 9 - ガイド RNA 複合体は、ストレプトアビジン・ビーズに結合された DNA に付加される。Cas 9 ガイド RNA 複合体は、標的部位にて DNA を切断する。その後、放出された DNA は、新しい管に移される。その後、アダプターライゲーションを伴う任意の適切な N G S ライブラリー調製プロトコルが、配列決定のための DNA を調製するために使用され得る。

10

#### 【0072】

本明細書で提供される標的化配列決定のいくつかの方法では、ニックアーゼなどの酵素は、最初の工程で分子の 3' 端を露出させるために使用することができ、次に、鎖置換ポリメラーゼなどのポリメラーゼは、酵素に切断された部位およびビオチンなどの組み込みアフィニティー分子から伸長し得る。場合によっては、RNA 分子は、ビオチン d A T P s を使用してポリアデニル化され、および、同様にストレプトアビジン・ビーズ上でキャプチャーされ得るか、または、オリゴ d T キャプチャー・オリゴが第2鎖合成のためのプライマーとして使用され得る場所で、またはオリゴ d T キャプチャーされる得る。二本鎖分子を（好適には末端にて）キャプチャーするあらゆる適切な方法が、中間構成物を作るために使用されることが企図される。鍵となるのは、C R I S P R を使用して標的分子を切断すること、および物理的に固体基質に結合した分子からそれらを分離することである。

20

#### 【0073】

場合によっては、d s D N A は断片化され、末端修復され、A テーリングされ、およびビオチン化されたアダプターにライゲーションされる。場合によっては、ビオチン化されたプライマーは、RNA 分子の 3' 端にライゲーションされる。場合によっては、d s D N A は、ビオチン化された DNA でタグメント化される (tagmented)。場合によっては、単鎖または、d s D N A は、ターミナルトランスフェラーゼとビオチン標識された d N T P s または d d N T P でビオチン化される。場合によっては、オリゴヌクレオチド合成の間における 5' 端組み込みには、例えば、プライマー伸長のために使用されるプライマー、Nextera トランスポゾン端、T S O、その他のものがある。場合によっては、キナーゼ処理によってリン酸にコンジュゲートされたビオチンの 5' 端組み込みが使用される。場合によっては、3' 端組み込みは、クレノー (Klenow) などの、3' - 5' エキソヌクレアーゼ活性を備えたポリメラーゼによって行われ、ここで酵素は、d s D N A 分子の最後のいくらかの 3' ヌクレオチドを除去し、およびビオチン・コンジュゲートされたヌクレオチドを足し込むことになる。場合によっては、3' 端組み込みは、Terminal Transferase を用いた処理によってビオチン・コンジュゲートされたデオキシヌクレオチドを使用する。場合によっては、3' 端組み込みは、プライマー伸長する間にビオチン・コンジュゲートされたデオキシヌクレオチドヌクレオチドを使用する。場合によっては、5' 端または 3' 端は、ビオチン・ヌクレオチドを備えたアダプターのライゲーションによる組み込みを使用する。場合によっては、5' 端または 3' 端組み込みは、C L I C K 化学などの、コンジュゲーション化学によってビオチン・オリゴを使用する。

30

40

#### 【0074】

本明細書で提供される標的化配列決定のいくつかの方法では、分子の中間の中へのビオ

50

チン組み込みは、以下の方法の1つで行なわれ得る：プライマー伸長する間における、アニールされたプライマーからの、ヌクレオチドの組み込み；または、ニックング、次に *Sequenase* などの鎖置換ポリメラーゼを用いたニックの3'端からのプライマー伸長による、既存の dsDNA 鋳型の中へのヌクレオチドの組み込み。

【0075】

本明細書で提供される標的化配列決定のいくつかの方法では、バーコード付きランダムプライマーおよびアダプターがテールされたバーコード付きランダムプライマーは、サンプルを多重化するまたはライブラリー調製ワークフローを効率化するために、使用され得る。

【0076】

本明細書で提供される標的化配列決定のいくつかの方法では、CRISPRで切断された生成物からのライブラリー調製は、任意の適切な調製を使用し、限定されないが、タグ付け、断片化、Aテーリング、およびライゲーション、または他の適切な方法が含まれる。例えば、二本鎖ゲノムDNAは、ビオチン化されたトランスポゾンでタグ付けされ得る。このことは、捕捉されたとき、すべての分子が比較的等しい長さであることを確認し得る。

10

【0077】

本明細書における標的化配列決定法の例示的な実施形態では、プライマー5' - Illuminaアダプター - 8bpサンプルバーコード - ランダムプライマー - 3' (5' - Illumina adapter - 8bp sample barcode - random primer - 3') を含む、核酸分子が使用される。複数のサンプル(微小滴定プレートの個別のウェルなど)が第1の工程(伸長およびビオチン ddNTP 伸長停止)で処理され、次に、キャプチャーのためにまとめてプールされ、そして洗浄される場合がある。第2鎖合成とCRISPR標的化が、すべてのサンプルに同時に生じる場合がある。この場合、アダプターは、Aテーリングやライゲーションなどによって、CRISPRに切断された生成物のもう1つの端部に付加されなければならない。両端部におけるライゲーションは、CRISPR切断に先立ってビーズ/固体基質上の生成物をブロックするなどの任意の適切な方法を使用して防止される。生成物がビーズから切り離されるとき、アダプターはCRISPRに切断された位置の平滑末端だけにライゲーションすることができる。PCR濃縮は、配列決定のために2つのアダプターを有する生成物が富化されることを確認するために使用される。

20

30

【0078】

本明細書でされる標的化配列決定法の付加的な実施形態では、プライミングは、ランダムプライミングまたはプライマーゼに基づくイニシエーション、そしてポリメラーゼ伸長によって行うことができる。ランダムプライマーは、サンプルバーコードテール( )またはアダプターテール、またはバーコードを付されたアダプターテールを有し得る。アフィニティー分子は、ポリアデニル化などのビオチンまたは他の適切なアフィニティータグであり得る。固体基質は、磁気ビーズ、ストレプトアビジンでコーティングされたプレート、他の適切な基質であり得る。固体基質は、カラムの中、マイクロ流体デバイスの中、ピペットチップの中、または他の適切な基質面上にあり得る。

40

【0079】

本明細書における標的化配列決定法のさらなる実施形態では、サンプル中のすべての分子の二本鎖の、基質に結合された変換を作製することができ、そして、標的配列の、RNAでガイドされるエンドヌクレアーゼ切断によって物理的に分離することができる、あらゆる適切な方法が企図される。

【0080】

本明細書でされる標的化配列決定法の付加的な実施形態では、RNAでガイドされるエンドヌクレアーゼは、RNA/DNA二重螺旋、DNA/DNA二重螺旋を標的とするために使用され、PAM配列があるもの、またはPAM配列がないものを含む(PAM配列を必要としないCRISPER/Cas9酵素利用する)。

50

## 【0081】

本明細書における標的化配列決定法のさらなる実施形態では、一本鎖DNAまたはRNAはビーズ上にキャプチャーされ、特異性を改良するために標的とされるオリゴヌクレオチド（標的部点を唯一の二本鎖にする）にハイブリダイズされる。

## 【0082】

本明細書における標的化配列決定法の応用は、限定されないが、コロナウイルス、呼吸系ウイルスなどの既知のウイルスゲノムの全ゲノム配列決定を含む。RNA全体は、Riptide（登録商標）ライブラリーを用いて（例えば、i）第1のランダムプライマーとバーコードを付されたアダプターを含むオリゴヌクレオチドを核酸にアニーリングすること、ii）第1のランダムプライマーを伸長し、そして伸長生成物を生成するために伸長を終止させること、iii）アダプターを有する第2のランダムプライマーを伸長生成物にアニーリングし、そして第2のランダムプライマーを使用して、二本鎖伸長生成物を生成することを含む、ライブラリー構築技術）で準備し、キャプチャーすることができる。ガイドRNAは、500bp毎、1kb毎、3kb毎、またはそれ以上毎に設計され得、分離され配列決定されるウイルスcDNA配列を切り離し得る。場合によっては、このことは、サンプルからのRNA全体の複合的なバックグラウンドからウイルスの配列についてRNAを富化する。このことは、ハイスループット設定またはフィールド展開可能なOxfordナノポア配列決定において行うことができる。

## 【0083】

標的化配列決定の付加的な応用は、BRCA遺伝子などのACMG遺伝子リストの配列決定を含む。付加的な実施形態では、標的化配列決定は、キャリアスクリーニングパネル（carrier screening panel）（すべての既知の劣性の障害）を含む。

## 【0084】

場合によっては、CRISPR切断は対立遺伝子特異的である。例えば、特異的対立遺伝子が切り取られ、および、HLAなど、標的化遺伝子座について層化された情報が得られる。あるいは、低頻度癌変異が標的とされ得る。本明細書の方法は、低存在量核転写を標的とするために使用され得る。

## 【0085】

本明細書における標的化配列決定法の別の例では、標準Riptide（登録商標）ライブラリー（例えば、i）第1のランダムプライマーとバーコード付きアダプターを含むオリゴヌクレオチドを核酸にアニーリングすること、ii）第1のランダムプライマーを伸長し、そして伸長生成物を生成するために伸長を終止させること、iii）アダプターを有する第2のランダムプライマーを伸長生成物にアニーリングし、そして第2のランダムプライマーを使用して、二本鎖伸長生成物を生成することを含む、ライブラリー構築技術）は、ローパス配列決定と遺伝子型判定のために生成される。オリジナルのライブラリーは、研究における関心のSNPを切り取るために、および各サンプルにおけるそれらの変異体の高カバレッジの確認を得るために、使用することができる。

## 【0086】

本明細書の方法のワークフローの1つの例は、以下のとおりである。アダプターライゲーションを伴う現在の方法。Illuminaアダプターを含まないバーコード付きランダムプライマーを用いて標準A反応（プライマー伸長反応）を行なう。ストレプトアビジン磁気ビーズでプライマー伸長分子をキャプチャーする。非特異的に結合されたDNA分子を除去するために、ビーズを3回洗浄する。Illuminaアダプターを含まないバーコード付きランダムプライマーを用いてプライマー伸長反応（標準B反応）を行なって、相補鎖を作製する。非特異的に結合されたDNA分子を除去するために、ビーズを3回洗浄する。ビーズを1X Cas9緩衝液に再懸濁する。ストレプトアビジン・ビーズにCas9:sgRNAを添加する（sgRNAプールは、関心の遺伝子座すべてを標的とするガイドRNAを含む）。Cas9に媒介された部位特異的切断が起こるまで、37で30~60分インキュベートする。切り取られた分子は、もはやビーズに結合されな

10

20

30

40

50

いであろう。管を磁気スタンドに置く。上澄みを新しい管に移す。管を磁気スタンドに置く。上澄みを新しい管に移す（これは、残余のビーズを取り除くためである）。NEB Next Ultra II DNA Library Prep Kit の試薬を用いて、標準的なエンドリペア、リン酸化、Aテーリング反応を行う（NEB のプロトコルを使用）。NEB Next Ultra II DNA Library Prep Kit の試薬を用いて、Illuminaアダプターライゲーションを行う（NEBのプロトコルを使用）。0.9X Ampureビーズクリーンアップ、またはその他のビーズによるサイズセレクションを行う（NEBのプロトコルを使用）。全長Illumina P5およびP7プライマーを用いてPCR増幅を行う（NEBのプロトコルを使用）。Ampureビーズの最終的なクリーンアップを行う（NEBのプロトコルを使用）。

10

## 【0087】

標的外事象を減少させ、およびAプライマー・バーコードを含んでいる分子について富化することにより、プロトコルを改善するためのワークフローの別の例は、以下のとおりである。部分的なP5 Illuminaアダプターと5'アミン基（DNAの5'リン酸化をブロックするため）を持つバーコード付きランダムプライマーで標準A反応（プライマー伸長反応）を行う。ストレプトアビジン磁気ビーズでプライマー伸長分子を捕捉する。ビーズを3回洗浄し、非特異的に結合したDNA分子を除去する。Illuminaアダプターを含まないが5'アミンを含むバーコード付きランダムプライマーを用いてプライマー伸長反応（標準B反応）を行い、相補鎖を作製する。相補鎖は、5'末端でリン酸化がブロックされることになる。非特異的に結合されたDNA分子を除去するために、ビーズを2回洗浄する。Terminal Transferase酵素とデオキシヌクレオチドで処置することにより、DNA分子の利用可能な3'端をブロックする。ビーズをさらに2回洗浄する。ビーズを1X Cas9緩衝液に再懸濁する。ストレプトアビジン・ビーズにCas9:sgRNAを添加する（sgRNAプールは、関心の遺伝子座すべてを標的とするガイドRNAを含む）。Cas9に媒介された部位特異的切断が起こるまで、37°Cで30~60分インキュベートする。管を磁気スタンドに置く。上澄みを新しい管に移す。管を磁気スタンドに置く。上澄みを再度ピペットで吸い上げ、新しい管に移す（これは、残留ビーズを取り除くためである）。NEB Next Ultra II DNA Library Prep Kit の試薬を用いて、標準的なエンドリペア、リン酸化、Aテーリング反応を行う（NEBまたは別のプロトコルを使用）。P7アダプターのみでIlluminaアダプターライゲーションを行なう。アダプターは、分子のAプライマー端に、または存在するあらゆるBにプライマー端にライゲーションしないだろう。0.9X Ampureビーズクリーンアップ、またはその他のビーズによるサイズセレクションを行う（NEBのプロトコルを使用）。全長Illumina P5およびP7プライマーでPCR増幅を行なう。最終のAmpureビーズクリーンアップを行なう。

20

30

## 【0088】

ワークフローのさらなる例は、Cas9切断（先のプロトコルにおける工程1~5、部分的なP5 Illuminaアダプターを含むAプライマーと、いかなるIlluminaアダプター配列も含まないBプライマーを除く）まで正常なRipTideライブラリー調製の工程を行なうことを含み、Cas9切断を行ない、次に、P7アダプター置換可能な末端を有する切断された生成物の転移を実行し、そして、全長Illumina P5およびP7プライマーを使用して、P5-P7生成物をPCR増幅する。このプロトコルでは、アダプターライゲーションの必要はない。転移に基づくプロトコルは、以下のように詳述される：部分的なイルミナP5アダプターを持つバーコード付きランダムプライマーで標準A反応（プライマー伸長反応）を行う。ストレプトアビジン磁気ビーズでプライマー伸長分子を捕捉する。非特異的に結合されたDNA分子を除去するために、ビーズを3回洗浄する。Illuminaアダプターを含まないバーコード付きランダムプライマーを用いてプライマー伸長反応（標準B反応）を行なって、相補鎖を作製する。非特異的に結合されたDNA分子を除去するために、ビーズを3回洗浄する。ビーズを1X

40

50

Cas9 緩衝液に再懸濁する。Streptavidin・ビーズにCas9:sgRNA を添加する (sgRNA プールは、関心の遺伝子座すべてを標的とするガイドRNAを含む)。Cas9 に媒介された部位特異的切断が起こるまで、37 °C で30~60分インキュベートする。切り取られた分子は、もはやビーズに結合されないであろう。管を磁気スタンドに置く。上澄みを新しい管に移す。管を磁気スタンドに置く。上澄みを新しい管に移す (これは、残余のビーズを取り除くためである)。Nextera のようなシステムで転移を行なう。Illumina P7 アダプター (P5 アダプターは必要ない) は、CRISPR に切断された分子に挿入されることになる。Ampure ビーズ・サイズ選択を行なう (必要な場合も、必要でない場合もある)。P5 - P7 分子を増幅するために、全長 Illumina P5 および P7 プライマーを用いて PCR 増幅を行なう。最終の Ampure ビーズクリーンアップを行なう。

10

## 【0089】

代替のワークフローは、トランスポサゼへの dCas9 の融合を利用することにより、標的特異的なトランスポジションを行なうことを含み、前記トランスポサゼは、関心の標的部位にて Illumina アダプターでトランスポゾン末端を挿入することになる。トランスポジション後、アダプターライゲーションの必要も、Cas9 消化とトランスポジション事象とを分ける必要も、ないだろう。

## 【0090】

ワークフローの別の例は、Cas12a (Cpf1) を含み、5' オーバーハングを有するスタガードカットを作製するために、以下のように、dsDNA を切断する。部分的な P5 Illumina アダプター (ライゲーションをブロックするための 5' アミン基を含み得る) を含むバーコード付きランダムプライマーを用いて標準 A 反応 (プライマー伸長反応) を行なう。Streptavidin 磁気ビーズでプライマー伸長分子を捕捉する。非特異的に結合された DNA 分子を除去するために、ビーズを3回洗浄する。Illumina アダプターを含まないバーコード付きランダムプライマーを用いてプライマー伸長反応 (標準 B 反応) を行い、相補鎖を作製する。 (プライマーは、ライゲーションをブロックするための 5' アミン基を含む場合がある)。非特異的に結合された DNA 分子を除去するために、前記ビーズを3回洗浄する。ビーズを 1X Cas12a 緩衝液に再懸濁する。Streptavidin・ビーズに Cas12a:gRNA を添加する (gRNA プールは、関心の遺伝子座すべてを標的とするガイドRNAを含む)。Cas12a で媒介された部位特異的切断が起こるまで、37 °C でインキュベートする。切り取られた分子は、もはやビーズに結合されないであろう。管を磁気スタンドに置く。上澄みを新しい管に移す。管を磁気スタンドに置く。上澄みを新しい管に移す (これは、残余のビーズを取り除くためである)。3' 単一ヌクレオチド・オーバーハング (A、C、G、および T オーバーハングの混合物) がある Illumina P7 アダプターを加える。ギャップを埋めるために T4 DNA ポリメラーゼで処置し、次に、ニックをふさぐために T4 DNA リガーゼを加える。0.9X Ampure ビーズクリーンアップ、またはその他のビーズによるサイズセレクションを行う (NEB のプロトコルを使用)。全長 Illumina P5 および P7 プライマーで PCR 増幅を行なう。最終の Ampure ビーズクリーンアップを行なう。

20

30

40

## 【0091】

## さらなる実施形態

本開示の態様は、標的とされる、同一ではない、標的核酸配列由来の配列のサブセットを各々含む、タグ付けされた核酸分子の集団を生成するための、方法と組成物とを記載する。標的核酸サンプルは、植物、動物 (ヒトを含む)、細菌、菌類、または藻類を含む、任意の生物学的もしくは環境的供給源から得られ得る。任意の適切な生体試料が標的核酸として使用される。便利な適切なサンプルは、全血、組織、精液、唾液、涙、尿、糞便物質、汗、バツカル、皮膚、および毛髪を含む。いくつかの実施形態では、標的核酸は、50~500 の細胞から得られる。いくつかの実施形態では、標的核酸は、50~400、50~350 の、50~300、100~300、150~300、200~300、ま

50

たは200～250の細胞から得られる。

【0092】

一実施形態では、標的化配列決定法は、第1の分子のタグ配列を含む第1の核酸分子と、標的核酸サンプルから第1の長さを有している第1の標的配列とを得る工程を含む場合がある。第1の核酸分子は、様々な長さのものであり得る。いくつかの実施形態では、第1の核酸分子の長さは、特定の配列決定プラットフォームのための最適の長さに相当する。特定の配列決定プラットフォームのための最適の長さは、イオン半導体のための400以内のヌクレオチド塩基（例えば、ION TORRENT, Life Technologies, Carlsbad, CA）、パイロシーケンシング（pyrosequencing）のための700のヌクレオチド塩基（例えば、GS JUNIOR+, 454 Life Sciences, Branford, CT）、および合成による配列決定のための50～300のヌクレオチド塩基（SBS）（e.g., MISEQ, Illumina, San Diego, CA）を含む場合がある。いくつかの実施形態では、第1の核酸分子は、50～1000、100～1000、200～1000、300～1000、300～900、300～800、300～700、300～600、300～500または400～500のヌクレオチド塩基であり得る。いくつかの実施形態では、第1の核酸分子は、50、62.5、125、250、500、または1000ヌクレオチド塩基であり得る。

10

【0093】

いくつかの実施形態では、第1の核酸分子は分子リガンドを含む。いくつかの実施形態では、この分子のリガンドは、ビオチンまたは任意のビオチン誘導体またはアナログを含む。

20

【0094】

いくつかの実施形態では、分子タグ配列は、6、7、8、9、または10ヌクレオチド塩基長であり得る。いくつかの実施形態では、分子タグは、8ヌクレオチド塩基長である。一実施形態では、分子タグは、ランダムなヌクレオチド配列を含む。いくつかの実施形態では、ランダムな塩基配列は、標的核酸サンプルの含有量の変化を考慮して半ランダム的に合成される。ランダムなヌクレオチド配列は、代表的な「ランダムさ（randomness）」を反映するように選択され得、ゲノム中1%GCから100%GCグアニン-シトシン（GC）含量率のウィンドウに対して順序づけられ得、および、各GC%におけるゲノムの含有率に相対する比率で合成されてプールされ得る。

30

【0095】

いくつかの実施形態では、第1の核酸分子は、第1のランダムなオリゴヌクレオチド配列を含む第1のプライマーを標的核酸サンプルに接触させることによって得られる場合がある。いくつかの実施形態では、第1のプライマーを接触させることは、第1のプライマーを前記標的核酸サンプルの核酸にアニーリングすることを含む。アニーリングは、完全なハイブリダイゼーションまたは不完全なハイブリダイゼーションを生じ得る。さらなる実施形態では、第2の核酸は、第2のランダム・オリゴヌクレオチド配列を含む第2のプライマーを第1の核酸分子に接触させることによって生成される。この方法は、第1の核酸分子に第2の分子のタグ配列を含むオリゴヌクレオチドをアニールする工程、および、第1の分子のタグ配列、第1の長さを有する第1の標的配列、および第2の分子のタグ配列を含む第1の二本鎖核酸分子を得るために、前記オリゴヌクレオチドを伸長する工程を含む場合がある。

40

【0096】

本明細書に記載の標的化配列決定法は、第3の分子のタグ配列、第2の長さを有している第2の標的配列、および第4の分子のタグ配列を含む、第2の二本鎖核酸分子を得る工程、および、第3の分子のタグ配列が第1の分子のタグ配列と同一の場合に、第2の二本鎖核酸分子を切り捨てる工程をさらに含み、第4の分子のタグ配列は第2の分子のタグ配列と同一であり、第2の標的配列は第1の標的配列と同一であり、および、第2の標的配列の長さは第1の標的配列の長さと同じである。いくつかの実施形態では、第3の分子の

50

タグ配列が第1の分子のタグ配列とは異なる場合、第2の二本鎖分子が保持されてもよく、第4の分子のタグ配列は第2の分子のタグ配列とは異なり、第2の標的配列は第1の標的配列とは異なるか；または、第2の標的配列長さは第1の標的配列長さとは異なり、結果は、非同ーで、標的核酸サンプル由来の配列のサブセットを各々含む、タグ付き核酸分子の集団を生成する。

【0097】

いくつかの実施形態では、第1の核酸は、前記第1のランダムなオリゴヌクレオチド配列に対して5'側に位置するアダプター配列を含む。いくつかの実施形態では、このアダプター配列は、特定の配列決定プラットフォームのために増幅および/または配列決定を容易にするために付加される。配列決定プラットフォームは、イオン半導体（例えば、ION TORRENT, Life Technologies, Carlsbad, CA）、パイロシーケンシング（例えば、GS JUNIOR+, 454 Life Sciences, Branford, CT）、および、合成による配列決定（SBS）（例えば、MISEQ, Illumina, San Diego, CA）を含む。例示的なアダプター配列は、配列番号1と2を含む。

10

【0098】

場合によっては、標的化配列決定ライブラリー分子は、配列決定に先立って環状化される。ライブラリー分子の環状化は、例えば、アダプター配列配列番号1および配列番号2に逆の相補的な配列、または他のアダプター配列を含む「架橋オリゴ」または「スプリントオリゴ」の提供により達成され、その結果、単鎖のライブラリー生成物分子の5'末端と3'末端が、架橋オリゴによって同時に結合される。場合によっては、架橋オリゴは、塩基対水素結合相互作用により、一本鎖ライブラリー分子の5'末端と3'末端を近接させ、その結果、リガーゼの添加により、分子の5'末端と3'末端が結合し、環状化されたライブラリー分子が形成され得る。分子は、ライゲーション、cre-loxベースの融合、ニック-リペアベース(nick-repair-based)の技術、さもなければ単一の環状分子を形成することなどの、数ある分子の技術のいずれかによって環状化され得る。場合によっては、その後、ライブラリーは、架橋オリゴを除去するためにエキソヌクレアーゼで処置される。

20

【0099】

その後、環状化された分子は、配列情報を得るために、ローリングサークル増幅/配列決定などの、当該技術分野で既知の、多くの配列決定技術のうちの一つによって配列決定される。

30

【0100】

場合によっては、第1の核酸および第1のプライマーは、核酸ポリメラーゼとヌクレオチド三リン酸に接触され得る。核酸ポリメラーゼは、ファミリーA、B、C、D、X、Y、およびRT由来のDNAポリメラーゼを含む。いくつかの実施形態では、核酸ポリメラーゼは、鎖転移活性を有する。いくつかの実施形態では、核酸ポリメラーゼは、鎖転移活性を欠く。ヌクレオチド三リン酸塩は、dATP、dCTP、dITP、dUTP、dGTP、およびdTTPなどのデオキシリボヌクレオチド三リン酸、ならびに、ddATP、ddCTP、ddGTP、ddITP、およびddTTPなどの、ジデオキシリボヌクレオチド三リン酸(ddNTP)を含み得る。いくつかの実施形態では、ヌクレオチド三リン酸は、デオキシリボヌクレオチド三リン酸塩およびデオキシヌクレオチド三リン酸塩を含むプールから核酸ポリメラーゼによって選択される。いくつかの実施形態では、このプールは、0.01%~5.0%、0.01%~4.0%、0.01%~3.0%、0.01%~2.0%、0.02%~2.0%、0.03%~2.0%、0.04%~2.0%、0.05%~2.0%、0.06%~2.0%、0.07%~2.0%、0.08%~2.0%、0.09%~2.0%、または0.1%~2.0%に及ぶ量でデオキシヌクレオチド三リン酸塩を含み得る。いくつかの実施形態では、プールは、0.05、0.1%、0.2%、0.4%、0.8%、または1.0%の量でデオキシヌクレオチド三リン酸塩を含み得る。いくつかの実施形態では、ヌクレオチド三リン酸は、dATP、dCTP

40

50

、 d G T P、および d T T Pを含むプールから核酸ポリメラーゼによって選択され、ここで4つのデオキシリボヌクレオチド三リン酸のうち1つは他の3つよりも著しく低い濃度であり、または4つのデオキシリボヌクレオチド三リン酸うちの2つが他の2つよりも著しく濃度が低い。場合によっては、ヌクレオチド三リン酸は、デオキシリボヌクレオチド三リン酸、ならびに、2, 6ジアミノプリン(2, 6 Diaminopurine)および2-チオチミジン(2-thiothymidine)(または5位置にメチル基がないウラシル)などの修飾ヌクレオチドのプールから核酸ポリメラーゼによって選択される。場合によっては、修飾ヌクレオチドは「半適合性の(semi-compatible)」ヌクレオチド塩基対を含む。場合によっては、半適合性のヌクレオチド塩基対は、自然発生のヌクレオチド塩基、またはそれらの自然発生の類縁と対になる塩基と塩基対になることができるが、それらの自然発生の塩基対相手のアナログとは塩基対にならないように、選択された、修飾ヌクレオチドを含む。例えば、アデニン(Adenine)・アナログである2, 6-ジアミノプリンは、チミジン(Thymidine)と塩基対になることができ、および、チミジンアナログである2-チオチミジンはアデニンと塩基対になることができるが、しかし半適合性の対の2, 6-ジアミノプリンと2-チオチミジンは互いに塩基対になることができない。このように、アデニンアナログである2, 6-ジアミノプリンとチミジンアナログである2-チオチミジンは、半適合性の塩基対を構成する。ヌクレオチド三リン酸 d G T Pと d C T P(相補的か自然な対)を含む組成物、および半相補的な対であるデオキシ-2, 6-ジアミノプリンTPとデオキシ-2-チオチミジンTPは、従って、鋳型に方向づけられた核酸合成の3'OH位置からの伸長を支持する。

#### 【0101】

代替A:T対および代替G:C対などの、他の修飾塩基対合が企図される。

#### 【0102】

そのような半適合性の修飾塩基の利点は、核酸が取り出されるdNTPプールが十分な濃度のこれらの塩基を含んでいる場合、これらの修飾塩基を組み込む核酸鋳型は、合成のための鋳型として働かないことである。従って、これらの塩基を組み込む核酸は、他の合成された核酸によって鋳型とされるのではなく、むしろ元の核酸サンプルによって確実に鋳型とされる。この特徴は、核酸合成反応の初期における塩基組み込みミスマッチエラーが後の鋳型において広められるというリスクを伴わずに、サンプル核酸の複数コピーの合成を可能にする。しかしながら、dNTPプールを、アナログが置換である塩基の型の自然発生のdNTPから成る、または含む、プールに置換することによって、4つの自然発生の塩基をすべて含む核酸が、塩基対アナログを組み込む鋳型から生成される。

#### 【0103】

場合によっては、修飾ヌクレオチドの少なくとも1つは、標識されている。場合によっては、修飾ヌクレオチドの少なくとも1つは、ジゴキシゲニン(DIG)、ビオチン、フルオレセイン、またはテトラメチルローダミン(tetramethylrhodamine)で標識されている。場合によっては、鋳型は、第1の核酸と第1のプライマーを接触させる前に、特定の長さの断片へと断片化される。場合によっては、鎖伸長終止を達成するために、エンドヌクレアーゼ処置に敏感なヌクレオチドなどの、1つ以上のヌクレオチドアナログがエンドヌクレアーゼとの組み合わせで使用される。場合によっては、鎖伸長終止は、dNTP濃度の操作によって達成される。

#### 【0104】

一実施形態では、デオキシリボヌクレオチド三リン酸とデオキシヌクレオチド三リン酸を含むプールは、分子のリガンドに結合された少なくとも1つのデオキシヌクレオチド三リン酸を含む。いくつかの実施形態では、この分子のリガンドは、ビオチンを含む。いくつかの実施形態では、方法は、前記第1の核酸分子にアニールされた第2の分子のタグ配列を含むオリゴヌクレオチドを含む分子を、リガンド結合剤に接触させる工程を含む。いくつかの実施形態では、このリガンド結合剤は、アビジンまたはストレプトアビジンである。場合によっては、リガンド結合剤は、DIG、ビオチン、フルオレセイン、またはテ

トラメチルローダミンに対する高度親和性抗体である。

【0105】

いくつかの実施形態では、本明細書に記載の核酸の少なくとも1つは、デオキシリボ核酸である。実施形態では、デオキシリボ核酸は、10キロベースを超える断片へと断片化される。断片化は、機械的分断または酵素の消化を含む、多くの方法で遂行され得る。いくつかの実施形態では、本明細書に記載された核酸の少なくとも1つは、リボ核酸である。いくつかの実施形態では、標的核酸サンプルは、リボ核酸である。実施形態では、第1の核酸分子は、リボ核酸から生成された相補的デオキシリボ核酸(cDNA)分子である。いくつかの実施形態では、cDNAを生成した核酸ポリメラーゼは、RNA依存性DNAポリメラーゼである。いくつかの実施形態では、cDNAは、標的核酸サンプルにオリゴ(dT)配列を含む第1のプライマーを接触させることによって生成される。

【0106】

さらなる実施形態では、同じ分子タグを有する所与のコンテイング由来の配列は、すべて、特定の相同染色体に割り当てられる。

【0107】

また、本明細書において、第1の分子のタグ配列と第1の長さを有する第1の標的配列とを含む第1の核酸分子、および第2の分子タグ配列を含むオリゴヌクレオチドを含む、標的化配列決定組成物が記載される。いくつかの実施形態では、第1の核酸分子は、3'デオキシリボヌクレオチドを含む。いくつかの実施形態では、3'デオキシリボヌクレオチドは、デオキシヌクレオチドである。いくつかの実施形態では、第1の核酸は、第1の分子タグ配列に対して5'側に位置するアダプター配列を含む。このアダプター配列は、イオン半導体(例えば、ION TORRENT, Life Technologies, Carlsbad, CA)、パイロシーケンシング(例えば、GS JUNIOR+, 454 Life Sciences, Branford, CT)、または、合成による配列決定(SBS)(例えば、MISEQ, Illumina, San Diego, CA)などの特定の配列決定プラットフォームのために増幅および/または配列決定を容易にするために付加され得る。例示的なアダプター配列は、5' AAT GAT ACG GCG ACC ACC GA 3' (SEQ ID NO: 1)、および5' CAA GCA GAA GAC GGC ATA CGA GAT 3' (SEQ ID NO: 2)を含む。Illumina, 454, Ion Torrentおよび他の既知の配列決定技術と適合性のあるアダプターは、本明細書において企図される。

【0108】

いくつかの実施形態では、第1の核酸分子を含む標的化配列決定組成物は、分子リガンドを含む。いくつかの実施形態では、この分子のリガンドは、ビオチンを含む。いくつかの実施形態では、組成物はリガンド結合剤を含む。いくつかの実施形態では、このリガンド結合剤は、アビジンまたはストレプトアビジンである。本明細書に記載の組成物はまた、リガンド-リガンド結合剤洗浄緩衝剤を含む場合がある。いくつかの実施形態では、本明細書に記載の組成物は、ビオチン洗浄緩衝剤を含む。

【0109】

本明細書に記載の標的化配列決定組成物はまた、組み込まれていないヌクレオチドを含む場合がある。いくつかの実施形態では、組み込まれていないヌクレオチドは組み込まれていないデオキシリボヌクレオチドである。いくつかの実施形態では、組み込まれていないヌクレオチドはデオキシヌクレオチドである。

【0110】

いくつかの実施形態では、本明細書に記載の組成物は、第2の分子のタグ配列を含むオリゴヌクレオチドにハイブリダイズされた第1の核酸分子を含む。第1の核酸分子は、オリゴヌクレオチドの第2分子のタグ配列に完全にハイブリダイズされるか、または、第1の核酸分子は、オリゴヌクレオチドの第2の分子タグ配列に不完全にハイブリダイズされる場合がある。

【0111】

さらに本明細書に記載されるのは、核酸分子の集団を含む標的化配列決定組成物であり、ここで分子は各々、第1のアダプター配列を含む第1鎖、分子タグ配列、および独立した標的配列を個別に含み、および、ここで独立した標的配列は各々、サンプル核酸配列のサブセットを含み、および、ここで集団の少なくとも第1の分子は、サンプル核酸配列の第1のサブセットを含む独立した標的配列を含み、および、ここで集団の少なくとも第2の分子は、サンプル核酸配列の第2のサブセットを含む独立した標的配列を含む。いくつかの実施形態では、集団の各第1鎖のアダプターは、同一である。いくつかの実施形態では、集団の各分子の分子タグ配列は、少なくとも6つのヌクレオチド塩基を含む。いくつかの実施形態では、集団の第1のメンバおよび集団の第2のメンバは、異非同一の分子タグ配列を含む。いくつかの実施形態では、各第1鎖は、その3'末端に3'-デオキシヌクレオチドを含む。いくつかの実施形態では、各第1鎖は、その5'末端に分子リガンドを含み得るか、または、各第1鎖は、非終止位置にて付着された分子リガンドを含み得る。付加的には、各第1鎖は、その3'末端にて分子リガンドを含み得る。いくつかの実施形態では、前記分子リガンドは、ピオチンである。

#### 【0112】

いくつかの実施形態では、本明細書に記載の組成物は、核酸分子の集団を含み、ここで集団の各分子は、第2のアダプター配列および第2の分子タグ配列を含む第2鎖を含む。さらなる実施形態において、集団の少なくとも1つの分子の第2鎖は、第2鎖の第2の分子タグ配列の、第1鎖の独立した標的配列に対する、少なくとも部分的な塩基対合を介して、第1鎖にアニールされる場合がある。いくつかの実施形態では、集団の各第2鎖のアダプターは、同一であり得る。いくつかの実施形態では、集団の少なくとも1つの分子は、分子リガンド結合剤に結合される。いくつかの実施形態では、分子リガンド結合剤は、アビジンまたはストレプトアビジンを含む。

#### 【0113】

本明細書に記載の標的化配列決定組成物はまた、組み込まれていない核酸三リン酸を含む場合がある。いくつかの実施形態では、本明細書に記載の組成物は、分子リガンド結合剤洗浄緩衝剤、および/またはポリメラーゼ伸長緩衝剤、および/または核酸ポリメラーゼを含み得る。いくつかの実施形態では、核酸ポリメラーゼは、核酸ヘリカーゼ活性を持つ。いくつかの実施形態では、本明細書に記載された組成物は、核酸鎖転移活性を持つ核酸ポリメラーゼを含む。いくつかの実施形態では、本明細書に記載の組成物は、Illumina、Ion torrentまたは454配列決定技術に適合する配列を含む。いくつかの実施形態では、本明細書に記載の組成物は、配列番号1および配列番号2で詳述される配列を含む。

#### 【0114】

ここで得られた標的化配列情報は、場合により、核酸蓄積レベルを定量するために使用される。ライブラリーは、本明細書に開示されるとおり、生成され、および配列決定される。固有にタグ付けされたリードだけが含まれるように、複製のリード(read)は除外される。固有のリード配列は、ゲノムの配列またはcDNAライブラリー、または、所与の細胞型または処理についてのトランスクリプトーム、または生物体についての、トランスクリプトームセット全体までの又はトランスクリプトームセット全体を含むより大きなトランスクリプトームセットなどの、トランスクリプトームの配列にマッピングされる。標的領域にマッピングする固有のライブラリー配列リードの数は、数えられ、サンプルにおけるその配列の存在量を表わすために使用される。いくつかの実施形態では、固有にタグ付けされた配列リードはそれぞれ、サンプル配列における単一の部位に対してマッピングする。場合によっては、固有にタグ付けされた配列リードは、トランスポゾン挿入部位または循環要素部位などのゲノムの全体にわたる複数の部位に対してマッピングする。従って、場合により、トランスクリプトーム「遺伝子座」または転写にマッピングするライブラリー分子の数は、ライブラリーが生成されるもとなるサンプルにおけるその転写の蓄積のレベルに対応する。循環要素にマッピングするライブラリー分子の数は、ゲノムの所与の固有の領域にマッピングするライブラリー分子の数に比べて、サンプルにおける

循環要素の相対存在量をよく示す。このように、本明細書で開示されるのは、ライブラリーを生成するもととなったサンプル中の核酸分子配列の存在量に対応する核酸分子配列の出現頻度などの、サンプル中の核酸分子配列の相対存在量を定量する方法であり、前記方法は、固有にタグ付けされたライブラリー断片を有する配列ライブラリーを生成する工程、およびライブラリー上に核酸分子配列をマッピングする工程を含む。場合によっては、ライブラリーにおける核酸分子配列の出現頻度は、ライブラリーにおける第2の核酸分子配列の出現頻度、トランスクリプトームにおける既知の存在量の遺伝子座または転写に対応する前記第2の核酸配列、またはゲノムサンプルの1つのゲノムごとの既知のコピー数との比較で評価される。

#### 【0115】

10

前記組成物のいずれかを使用する標的化配列決定のためにサンプル中の核酸を調製する方法が、本明細書に記載される。いくつかの実施形態では、サンプルは、細胞、組織、または生物体の部分から得られる。生物体の非限定的な例は、ヒト、植物、細菌、ウイルス、原生動物、真核生物、および原核生物を含み得る。一例として、サンプルは、ヒトゲノム核酸を含むヒトゲノムである。サンプルは、核酸ライブラリーを調製するために使用される。ライブラリーは、配列決定される。

#### 【0116】

標的化配列決定の核酸ライブラリーの調製は、本明細書に記載されている、あるいは当該技術分野において周知の方法を用いて行われる。いくつかの実施形態において、核酸はヒトゲノムから得られる。ヒトゲノムの核酸を反応混合物Xにおいて増幅させる。いくつかの実施形態において、反応混合物XはDNA、少なくとも1つのプライマー、バッファー、デオキシヌクレオチド混合物、酵素、ヌクレアーゼフリー水を含み得る。反応混合物Xをエッペンドルフ管の中で調製する。反応混合物Xは、Eppendorf DNA LoBind微量遠心分離管の中で調整することが好ましい。場合によっては、DNAはヒトのDNAである。反応混合物Xにおける最終的なDNA濃度は約0.1 ng、0.2 ng、0.3 ng、0.4 ng、0.5 ng、0.6 ng、0.7 ng、0.8 ng、0.9 ng、1.0 ng、1.2 ng、1.4 ng、1.5 ng、1.8 ng、2.0 ngあるいはそれ以上である。反応混合物Xにおける最終的なDNA濃度は約0.1 ng、0.2 ng、0.3 ng、0.4 ng、0.5 ng、0.6 ng、0.7 ng、0.8 ng、0.9 ng、1.0 ng、1.2 ng、1.4 ng、1.5 ng、1.8 ng、2.0 ngあるいはそれ以下である。反応混合物Xにおける最終的なDNA濃度は0.1から約2.0 ngの間、約0.2 ngから約1.2 ngの間、約0.5 ngから約0.8 ngの間、あるいは約1.0 ngから約1.5 ngの間である。

20

30

#### 【0117】

場合によっては、反応混合物Xは1つのプライマーのみ、例えば、プライマーAを含む。全反応混合物におけるプライマーAの最終濃度は、約10 μM、20 μM、30 μM、40 μM、約50 μM、約100 μM、約150 μM、約200 μM、あるいはそれ以上である。全反応混合物XにおけるプライマーAの最終濃度は、約10 μM、20 μM、30 μM、40 μM、約50 μM、約100 μM、約150 μM、約200 μM、あるいはそれ以下である。全反応混合物XにおけるプライマーAの最終濃度は、約10 μMから約200 μMの間、約30 μMから約80 μMの間、約50 μMから約100 μMの間、または約40 μMから約150 μMの間である。

40

#### 【0118】

場合によっては、反応混合物XはThermo Sequenase Bufferのようなバッファーを含む。典型的に、反応混合物Xにおけるバッファーの最終濃度は元のバッファー濃度の約10%である。例えば、反応混合物Xの最終体積に応じて、加えるバッファーの量は、約1 μl、約2 μl、約2.5 μl、約3 μl、約4 μl、約5 μl、約10 μlか、あるいはそれらの量以下、それらの量以上である。

#### 【0119】

場合によっては、反応混合物Xは複数のデオキシリボヌクレオチドを含む。デオキシリ

50

ボヌクレオチドは、 $dATP$ 、 $dTTP$ 、 $dGTP$ 、 $dCTP$ 、 $ddATP$ 、 $ddTTP$ 、 $ddGTP$ 、および $ddCTP$ のうち1つ以上である。反応混合物Xにおけるデオキシリボヌクレオチドの最終濃度は約 $0.1\ \mu M$ 、約 $0.2\ \mu M$ 、約 $0.3\ \mu M$ 、約 $0.4\ \mu M$ 、約 $0.5\ \mu M$ 、約 $0.6\ \mu M$ 、約 $0.7\ \mu M$ 、約 $0.8\ \mu M$ 、約 $0.9\ \mu M$ 、約 $1.0\ \mu M$ 、約 $1.2\ \mu M$ 、約 $1.5\ \mu M$ 、約 $1.8\ \mu M$ 、約 $2.0\ \mu M$ 、あるいはそれ以上である。反応混合物Xにおけるデオキシリボヌクレオチドの最終濃度は約 $0.1\ \mu M$ 、約 $0.2\ \mu M$ 、約 $0.3\ \mu M$ 、約 $0.4\ \mu M$ 、約 $0.5\ \mu M$ 、約 $0.6\ \mu M$ 、約 $0.7\ \mu M$ 、約 $0.8\ \mu M$ 、約 $0.9\ \mu M$ 、約 $1.0\ \mu M$ 、約 $1.2\ \mu M$ 、約 $1.5\ \mu M$ 、約 $1.8\ \mu M$ 、約 $2.0\ \mu M$ 、あるいはそれ以下である。

【0120】

10

場合によっては、反応混合物Xはポリメラーゼなどの酵素を含む。例えば、場合によっては、酵素はThermo Sequenaseである。ポリメラーゼの最終濃度は約 $0.01\ \mu M$ 、約 $0.1\ \mu M$ 、約 $0.2\ \mu M$ 、約 $0.3\ \mu M$ 、約 $0.4\ \mu M$ 、約 $0.5\ \mu M$ 、約 $0.6\ \mu M$ 、約 $0.7\ \mu M$ 、約 $0.8\ \mu M$ 、約 $0.9\ \mu M$ 、約 $1.0\ \mu M$ 、約 $1.2\ \mu M$ 、約 $1.5\ \mu M$ 、約 $1.8\ \mu M$ 、約 $2.0\ \mu M$ 、あるいはそれ以上である。ポリメラーゼの最終濃度は約 $0.01\ \mu M$ 、約 $0.1\ \mu M$ 、約 $0.2\ \mu M$ 、約 $0.3\ \mu M$ 、約 $0.4\ \mu M$ 、約 $0.5\ \mu M$ 、約 $0.6\ \mu M$ 、約 $0.7\ \mu M$ 、約 $0.8\ \mu M$ 、約 $0.9\ \mu M$ 、約 $1.0\ \mu M$ 、約 $1.2\ \mu M$ 、約 $1.5\ \mu M$ 、約 $1.8\ \mu M$ 、約 $2.0\ \mu M$ 、あるいはそれ以下である。ポリメラーゼの最終濃度は、約 $2.0\ \mu M$ までの間、約 $0.1\ \mu M$ から約 $1.0\ \mu M$ の間、約 $0.5\ \mu M$ から約 $1.5\ \mu M$ の間、あるいは約 $0.8\ \mu M$ から約 $1.8\ \mu M$ の間である。

20

【0121】

典型的に、所望の最終体積にするために反応混合物Xに多量のヌクレアーゼフリー水を追加する。反応混合物の最終体積は約 $10\ \mu l$ 、約 $20\ \mu l$ 、約 $25\ \mu l$ 、約 $30\ \mu l$ 、約 $40\ \mu l$ 、約 $50\ \mu l$ 、あるいは $100\ \mu l$ である。反応混合物Xの最終体積に応じて、ヌクレアーゼフリー水の量は約 $0.1\ \mu l$ 、約 $0.5\ \mu l$ 、約 $0.8\ \mu l$ 、約 $1.0\ \mu l$ 、約 $2\ \mu l$ 、約 $5\ \mu l$ 、約 $10\ \mu l$ 、約 $15\ \mu l$ 、約 $20\ \mu l$ 、約 $25\ \mu l$ 、約 $30\ \mu l$ 、約 $40\ \mu l$ 、約 $50\ \mu l$ 、約 $80\ \mu l$ 、約 $90\ \mu l$ 、約 $95\ \mu l$ 、あるいはそれ以上である。ヌクレアーゼフリー水の量は約 $0.1\ \mu l$ 、約 $0.5\ \mu l$ 、約 $0.8\ \mu l$ 、約 $1.0\ \mu l$ 、約 $2\ \mu l$ 、約 $5\ \mu l$ 、約 $10\ \mu l$ 、約 $15\ \mu l$ 、約 $20\ \mu l$ 、約 $25\ \mu l$ 、約 $30\ \mu l$ 、約 $40\ \mu l$ 、約 $50\ \mu l$ 、約 $80\ \mu l$ 、約 $90\ \mu l$ 、約 $95\ \mu l$ 、あるいはそれ以下である。ヌクレアーゼフリー水の量は、約 $0.1\ \mu l$ から約 $95\ \mu l$ の間、約 $1.0\ \mu l$ から約 $10\ \mu l$ の間、約 $5\ \mu l$ から約 $50\ \mu l$ の間、あるいは約 $20\ \mu l$ から約 $80\ \mu l$ の間である。

30

【0122】

一般に、反応混合物Xは、DNAが変性するまで十分長くある温度( $T_m$ )でインキュベートされる。 $T_m$ は約 $80$ 、約 $85$ 、約 $90$ 、約 $91$ 、約 $92$ 、約 $93$ 、約 $94$ 、約 $95$ 、約 $96$ 、約 $97$ 、約 $98$ 、約 $99$ 、あるいはそれ以上である。反応混合物Xは約5秒間、約10秒間、約15秒間、約20秒間、約30秒間、約1分間、約2分間、約3分間、約4分間、約5分間、約6分間、約7分間、約8分間、約9分間、約10分間、あるいはそれら以上、それら以下の時間 $T_m$ でインキュベートされる。例えば、反応混合物Xは $95$ で約3分間インキュベートされる。変性後、管を氷の上に置くことにより反応混合物Xの温度は低下する。例えば、管は約5秒間、約10秒間、約15秒間、約20秒間、約30秒間、約5分間、約10分間、約15分間、約20分間、約30分間、約1分間、約2分間、約3分間、約4分間、約5分間、約6分間、約7分間、約8分間、約9分間、約10分間、あるいはそれら以上、それら以下の時間氷の上に配置される。ポリメラーゼ、例えばThermo Sequenase、を反応に追加し、丁寧に混ぜることが好ましい。一般的に、反応混合物Xはサーマルサイクラーへ移され、本明細書に記載のとおり、機器上での処理に進む。

40

【0123】

50

サーマルサイクラーは、(1)ある期間低温度を維持すること、(2)温度をDNAアニリング温度へと上昇させること、(3)ある期間アニリング温度を維持すること、(4)ある期間温度を変性温度へと上昇させること、(1)から(4)を少なくとも9回繰り返し行うことを含むプログラムの実行、そして保管のため8、4、それ以下、あるいは-20での冷凍を保つ。(1)における低温は約10、約12、約14、約16、約18、あるいは約20で維持される。(1)において低温は約5秒間、約10秒間、約15秒間、約20秒間、約30秒間、約1分間、約2分間、約3分間、約4分間、約5分間、約6分間、約7分間、約8分間、約9分間、約10分間、約15分間、あるいは約20分間維持される。代替として、サーマルサイクラーは約3分間約16で温度を維持することができる。いくつかの実施形態では、温度は(1)から(2)まで徐序に上昇され、温度は約0.1/秒の小さな増分で上昇される。(2)における温度は約45、約50、約55、約60、約65、約68、約70、あるいはそれ以上である。場合によっては、(2)の温度は0.1/秒で約60まで徐々に上昇される。場合によっては、(2)の温度は(3)の温度に等しい。場合によっては、(2)の温度は(3)の温度に達するためさらに大きく上昇される。(3)の温度は、約5秒間、約10秒間、約15秒間、約20秒間、約30秒間、約1分間、約2分間、約3分間、約4分間、約5分間、約6分間、約7分間、約8分間、約9分間、約10分間、約15分間、あるいは約20分間維持される。いくつかの実施形態では、(3)の温度は約10分間維持される。実施例として、(4)の温度は約95で、約10秒間、20秒間、30秒間、45秒間、60秒間、1分間、2分間、あるいはそれより長く維持される。

#### 【0124】

いくつかの実施形態では、反応混合物X中の反応成分はすべて、プライマーを除き、適切な分離装置上に組み合わせられロードされる。反応物が分離され、およびバーコード付きのプライマーと結合された後、反応混合物をサーマルサイクラーへと移し、95で2分間熱変性させ、その後本明細書に記載されたプログラムに従ってサーモサイクルされる。いくつかの実施形態において、生成物は、長期間保管のため、4、あるいは氷の上で保存、あるいは-20で冷凍される。いくつかの実施形態において、次の工程へと進む直前に、保存された生成物は約98で約3分間加熱され、その後氷の上で一時的に保存するため移される。

#### 【0125】

いくつかの実施形態において、上記記載の反応混合物XのDNA生成物は磁気ビーズを用いて捕捉される。このことは、上記記載の生成物を加える前にCapture Beadsを調製することにより達成される。まず、ビーズを再懸濁させるために、Capture Beadsの管を徹底的に振動させ約40 $\mu$ lのビーズを0.5mLの新しいEppendorf DNA LoBind管へと移す。場合によっては、ビーズの体積は約10 $\mu$ l、20 $\mu$ l、約30 $\mu$ l、約50 $\mu$ l、約100 $\mu$ l、あるいはそれ以上である。溶液が透明になるまで、管を約0.5~1分間磁気スタンドに配置する。上澄みはピペットで移し廃棄する。管を磁気スタンドから取り外す。約200 $\mu$ lの量のHS Bufferをビーズに加える。磁気スタンドへ戻す前に、サンプルをピペットで上下に移すことで構成要素を丁寧に混合する。溶液が透明になるまで、サンプルを約0.5~1分間磁気スタンドに設置した状態にする。上澄みはピペットで管の外へ丁寧に移すことで取り除き、および廃棄する。その後、管を磁気スタンドから取り外し、ビーズは40 $\mu$ lのHS Buffer中で再懸濁する。管を一時的に常温で実験台上に置いた状態にする。上記記載の反応混合物のDNA生成物を、本明細書に記載されているように調製したCapture Beadsに加え、約20分間常温でインキュベートした。場合によっては、DNAおよびCapture Beadsを含むサンプルを常温で約10分間、約15分間、約20分間、約30分間、あるいはそれより長くインキュベートする。DNA生成物およびCapture Beadsを約5分間、約10分間、約15分間、約20分間、約30分間、あるいはそれより長くピペットで上下に移すことで混合する。DNA生成物の混合物を含んだ管およびCapture Beadsを磁気スタンドに設置し、溶液が透明に

なるまで待つ。上澄みを管からピペットで慎重に移すことにより取り除く。その後、管は磁気スタンドから取り外されビーズを200 $\mu$ lのBead Wash Bufferに再懸濁し、溶液が透明になるまでの一定時間磁気スタンドへ戻され得る。上澄みは廃棄する。洗浄を少なくともさらに2回繰り返し、最終洗浄後に残った液体を慎重に取り除く。

【0126】

洗浄後のCapture Beadsおよび上記記載のDNA生成物を、反応混合物Yを生成するために試薬の混合物に加える。試薬はSequenase Buffer、複数のデオキシリボヌクレオチド、少なくとも1つプライマー、酵素およびヌクレアーゼフリー水を含み得る。

【0127】

場合によっては、反応混合物Yは1つのプライマーのみ、例えば、プライマーBを含む。全反応混合物YにおけるプライマーAの最終濃度は、約10 $\mu$ M、20 $\mu$ M、30 $\mu$ M、40 $\mu$ M、50 $\mu$ M、約100 $\mu$ M、約150 $\mu$ M、約200 $\mu$ M、あるいはそれ以上である。全反応混合物YにおけるプライマーBの最終濃度は、約10 $\mu$ M、20 $\mu$ M、30 $\mu$ M、40 $\mu$ M、50 $\mu$ M、約100 $\mu$ M、約150 $\mu$ M、約200 $\mu$ M、あるいはそれ以下である。全反応混合物YにおけるプライマーBの最終濃度は、約10 $\mu$ Mから約200 $\mu$ Mの間、約30 $\mu$ Mから約80 $\mu$ Mの間、約50 $\mu$ Mから約100 $\mu$ Mの間、あるいは約40 $\mu$ Mから約150 $\mu$ Mの間である。

【0128】

場合によっては、反応混合物YはSequenase Bufferを含む。典型的に、反応混合物Yにおけるバッファの最終濃度は元のバッファの濃度の約10%である。場合によっては、反応混合物Yにおけるバッファの最終濃度は元のバッファの濃度の約5%、約10%、約15%、約20%、約30%、あるいはそれ以下である。例えば、反応混合物Yの最終体積に応じて、加えるバッファの量は、約1 $\mu$ l、約2 $\mu$ l、約2.5 $\mu$ l、約3 $\mu$ l、約4 $\mu$ l、約5 $\mu$ l、約10 $\mu$ l、あるいはそれらの量以下、それらの量以上である。

【0129】

場合によっては、反応混合物Yは複数のデオキシリボヌクレオチドを含む。デオキシリボヌクレオチドはdATP、dTTP、dGTP、dCTP、ddATP、ddTTP、ddGTP、およびddCTPである。反応混合物におけるデオキシリボヌクレオチドの最終濃度は約0.1 $\mu$ M、約0.2 $\mu$ M、約0.3 $\mu$ M、約0.4 $\mu$ M、約0.5 $\mu$ M、約0.6 $\mu$ M、約0.7 $\mu$ M、約0.8 $\mu$ M、約0.9 $\mu$ M、約1.0 $\mu$ M、約1.2 $\mu$ M、約1.5 $\mu$ M、約1.8 $\mu$ M、約2.0 $\mu$ M、あるいはそれ以上である。反応混合物におけるデオキシリボヌクレオチドの最終濃度は約0.1 $\mu$ M、約0.2 $\mu$ M、約0.3 $\mu$ M、約0.4 $\mu$ M、約0.5 $\mu$ M、約0.6 $\mu$ M、約0.7 $\mu$ M、約0.8 $\mu$ M、約0.9 $\mu$ M、約1.0 $\mu$ M、約1.2 $\mu$ M、約1.5 $\mu$ M、約1.8 $\mu$ M、約2.0 $\mu$ M、あるいはそれ以下である。

【0130】

場合によっては、反応混合物Yは酵素を含む。酵素はポリメラーゼである。例えば、酵素はSequenaseである。場合によっては、Sequenasesは、Sequenaseおよび無機ピロホスファターゼを1:1の割合で含む。ポリメラーゼの最終濃度は約0.01 $\mu$ M、約0.1 $\mu$ M、約0.2 $\mu$ M、約0.3 $\mu$ M、約0.4 $\mu$ M、約0.5 $\mu$ M、約0.6 $\mu$ M、約0.7 $\mu$ M、約0.8 $\mu$ M、約0.9 $\mu$ M、約1.0 $\mu$ M、約1.2 $\mu$ M、約1.5 $\mu$ M、約1.8 $\mu$ M、約2.0 $\mu$ M、あるいはそれ以上である。ポリメラーゼの最終濃度は約0.01 $\mu$ M、約0.1 $\mu$ M、約0.2 $\mu$ M、約0.3 $\mu$ M、約0.4 $\mu$ M、約0.5 $\mu$ M、約0.6 $\mu$ M、約0.7 $\mu$ M、約0.8 $\mu$ M、約0.9 $\mu$ M、約1.0 $\mu$ M、約1.2 $\mu$ M、約1.5 $\mu$ M、約1.8 $\mu$ M、約2.0 $\mu$ M、あるいはそれ以下である。ポリメラーゼの最終濃度は約2.0 $\mu$ Mまでの間、約0.1 $\mu$ Mから約1.0 $\mu$ Mの間、約0.5 $\mu$ Mから約1.5 $\mu$ Mの間、あるいは約0.8 $\mu$ Mから約1.8 $\mu$ Mの間である。

10

20

30

40

50

## 【0131】

典型的に、所望の最終体積にするために反応混合物に多量のヌクレアーゼフリー水を加える。反応混合物 Y の最終体積は約 10  $\mu$ l、約 20  $\mu$ l、約 25  $\mu$ l、約 30  $\mu$ l、約 40  $\mu$ l、約 50  $\mu$ l、あるいは約 100  $\mu$ l である。最終反応液体積に応じて、ヌクレアーゼフリー水の量は約 0.1  $\mu$ l、約 0.5  $\mu$ l、約 0.8  $\mu$ l、約 1.0  $\mu$ l、約 2  $\mu$ l、約 5  $\mu$ l、約 10  $\mu$ l、約 15  $\mu$ l、約 20  $\mu$ l、約 25  $\mu$ l、約 30  $\mu$ l、約 40  $\mu$ l、約 50  $\mu$ l、約 80  $\mu$ l、約 90  $\mu$ l、約 95  $\mu$ l、あるいはそれ以上である。ヌクレアーゼフリー水の量は約 0.1  $\mu$ l、約 0.5  $\mu$ l、約 0.8  $\mu$ l、約 1.0  $\mu$ l、約 2  $\mu$ l、約 5  $\mu$ l、約 10  $\mu$ l、約 15  $\mu$ l、約 20  $\mu$ l、約 25  $\mu$ l、約 30  $\mu$ l、約 40  $\mu$ l、約 50  $\mu$ l、約 80  $\mu$ l、約 90  $\mu$ l、約 95  $\mu$ l、あるいはそれ以下である。ヌクレアーゼフリー水の量は約 0.1  $\mu$ l から約 95  $\mu$ l の間、約 1.0  $\mu$ l から約 10  $\mu$ l の間、約 5  $\mu$ l から約 50  $\mu$ l の間、あるいは約 20  $\mu$ l から約 80  $\mu$ l の間である。

10

## 【0132】

いくつかの実施形態において、反応混合物 Y は 24 時間で約 20 分間インキュベートされる。混合物はより長い時間、あるいはより短い時間インキュベートされる。例えば、反応混合物 Y は約 10 分間、約 15 分間、約 20 分間、約 30 分間、あるいはそれより長くインキュベートされる。温度は約 18  $^{\circ}$ C、約 20  $^{\circ}$ C、約 25  $^{\circ}$ C、約 28  $^{\circ}$ C、あるいはそれら以上、それら以下である。サーマルサイクラーあるいはヒートブロックにおいてインキュベーションを行なうことが好ましい。溶液が透明になるまで、その後管は一定時間磁気スタンドに設置され得る。上澄みを取り除き、および廃棄する。その後、管を磁気スタンドから取り外しビーズを約 200  $\mu$ l の Bead Wash Buffer 中で再懸濁して、磁気スタンドへ戻す前に、溶液が透明になるまでそのままにしておく。上澄みを慎重に取り除く。典型的に、洗浄の工程を少なくともさらに 2 回繰り返す。最終洗浄後に残った液体を慎重に取り除く。

20

## 【0133】

いくつかの実施形態において、反応混合物 Z を生成するために、反応混合物 Y を反応混合物に加える。一般に、反応混合物 Y は、PCR ユニバーサルプライマー I、バーコード付きの PCR プライマー II、KAPA HiFi PCR Amplification Mix およびヌクレアーゼフリー水を含む PCR 管における反応混合物 Z に加えられる。

30

## 【0134】

場合によっては、全反応混合物 Z' における PCR ユニバーサルプライマー I の最終濃度は約 10  $\mu$ M、20  $\mu$ M、30  $\mu$ M、40  $\mu$ M、約 50  $\mu$ M、約 100  $\mu$ M、約 150  $\mu$ M、約 200  $\mu$ M、あるいはそれ以上である。全反応混合物 Z' における PCR ユニバーサルプライマー I の最終濃度は約 10  $\mu$ M、20  $\mu$ M、30  $\mu$ M、40  $\mu$ M、約 50  $\mu$ M、約 100  $\mu$ M、約 150  $\mu$ M、約 200  $\mu$ M、あるいはそれ以下である。全反応混合物 Z' における PCR ユニバーサルプライマー I の最終濃度は約 10  $\mu$ M から約 200  $\mu$ M の間、約 30  $\mu$ M から約 80  $\mu$ M の間、約 50  $\mu$ M から約 100  $\mu$ M の間、あるいは約 40  $\mu$ M から約 150  $\mu$ M の間である。

40

## 【0135】

場合によっては、全反応混合物 Z' における PCR プライマー II の最終濃度は約 10  $\mu$ M、20  $\mu$ M、30  $\mu$ M、40  $\mu$ M、約 50  $\mu$ M、約 100  $\mu$ M、約 150  $\mu$ M、約 200  $\mu$ M、あるいはそれ以上である。全反応混合物 Z' における PCR プライマー II の最終濃度は約 10  $\mu$ M、20  $\mu$ M、30  $\mu$ M、40  $\mu$ M、約 50  $\mu$ M、約 100  $\mu$ M、約 150  $\mu$ M、約 200  $\mu$ M、あるいはそれ以下である。全反応混合物 Z' における PCR プライマー II の最終濃度は約 10  $\mu$ M から約 200  $\mu$ M の間、約 30  $\mu$ M から約 80  $\mu$ M の間、約 50  $\mu$ M から約 100  $\mu$ M の間、あるいは約 40  $\mu$ M から約 150  $\mu$ M の間である。

## 【0136】

50

場合によっては、反応混合物は K A P A H i F i P C R A m p l i f i c a t i o n M i x を含む。典型的に、反応混合物 Z ' における K A P A H i F i P C R A m p l i f i c a t i o n M i x の最終濃度は元のミックスの濃度の約 10 % である。場合によっては、反応混合物 Z ' における K A P A H i F i P C R A m p l i f i c a t i o n M i x の最終濃度は元のミックスの約 5 %、約 10 %、約 15 %、約 20 %、約 30 %、あるいはそれ以下である。例えば、反応混合物 Z ' の最終体積に応じて、加える K A P A H i F i P C R A m p l i f i c a t i o n M i x の量は約 1  $\mu$ l、約 2  $\mu$ l、約 2.5  $\mu$ l、約 3  $\mu$ l、約 4  $\mu$ l、約 5  $\mu$ l、約 10  $\mu$ l、あるいはそれら以下、それら以上である。

【0137】

典型的に、所望の最終体積にするために反応混合物 Z ' に多量のヌクレアーゼフリー水を加える。反応混合物 Z ' の最終体積は約 10  $\mu$ l、約 20  $\mu$ l、約 25  $\mu$ l、約 30  $\mu$ l、約 40  $\mu$ l、約 50  $\mu$ l、あるいは約 100  $\mu$ l である。最終反応液体積に応じて、ヌクレアーゼフリー水の量は約 0.1  $\mu$ l、約 0.5  $\mu$ l、約 0.8  $\mu$ l、約 1.0  $\mu$ l、約 2  $\mu$ l、約 5  $\mu$ l、約 10  $\mu$ l、約 15  $\mu$ l、約 20  $\mu$ l、約 25  $\mu$ l、約 30  $\mu$ l、約 40  $\mu$ l、約 50  $\mu$ l、約 80  $\mu$ l、約 90  $\mu$ l、約 95  $\mu$ l、あるいはそれ以上である。ヌクレアーゼフリー水の量は約 0.1  $\mu$ l、約 0.5  $\mu$ l、約 0.8  $\mu$ l、約 1.0  $\mu$ l、約 2  $\mu$ l、約 5  $\mu$ l、約 10  $\mu$ l、約 15  $\mu$ l、約 20  $\mu$ l、約 25  $\mu$ l、約 30  $\mu$ l、約 40  $\mu$ l、約 50  $\mu$ l、約 80  $\mu$ l、約 90  $\mu$ l、約 95  $\mu$ l、あるいはそれ以下である。ヌクレアーゼフリー水の量は約 0.1  $\mu$ l から約 95  $\mu$ l の間、約 1.0  $\mu$ l から約 10  $\mu$ l の間、約 5  $\mu$ l から約 50  $\mu$ l の間、あるいは約 20  $\mu$ l から約 80  $\mu$ l の間である。

【0138】

ポリメラーゼ連鎖反応 ( P C R ) を行ない、かつ生成物 X X を生成するために、反応混合物 Z をサーマルサイクラーに置く。P C R プログラムは D N A を変性させるために約 98 で 2 分間の少なくとも 1 サイクル、変性させるために約 98 で 20 秒間の少なくとも 15 サイクル、プライマーのアニーリングのために約 60 まで 30 秒間温度を低下させること、伸長のために約 70 まで 30 秒間温度を上昇させること、最終伸長のために 72 で 5 分間の少なくとも 1 サイクル、4 のまま保つことを含んでいる。場合によっては、D N A 変性温度は約 92、約 95、約 97、あるいは約 99 である。場合によっては、プライマーアニーリング温度は約 45、約 50、約 55、約 60、約 65、あるいは約 70 である。場合によっては、伸長温度は約 65、約 70、約 72、あるいは約 75 である。

【0139】

生成物 X X を A m p u r e X P B e a d s で浄化する。一般に、生成物 X X を含む P C R 管は、磁気スタンドに置かれ、および溶液が透明になるように静止状態に維持され、その後、上澄みがピペット操作によって除去される。上澄みを、新しい 0.5 mL E p p e n d o r f D N A L o B i n d 管に移す。C a p t u r e B e a d s を包含している P C R 管を廃棄する。典型的には、A m p u r e X P B e a d s の約 100  $\mu$ l を上澄みに添加し、そして、混合物を上下ピペット操作により混合し、その後、約 10 分間、室温でインキュベートする。場合によっては、インキュベーションの時間は、約 5 分、約 15 分、約 20 分、約 30 分、またはそれ以上など、10 分より長いまたは短い。溶液が透明になるまで、チューブを約 0.5 ~ 1 分間磁気スタンドに配置する。上澄みを廃棄する。約 200  $\mu$ l の 80 % エタノールを管に加え、および約 30 秒間おき、その後エタノールを除去および廃棄する。この手順の間、磁気スタンドから管を取り除くことが必要ではない場合がある。管は、少なくともさらに 1 回、200  $\mu$ l の 80 % エタノールで洗浄される。管のキャップを開き、そして約 10 ~ 15 分間、ビーズを大気乾燥にさらす。約 20  $\mu$ l ~ 約 30  $\mu$ l の 10 mM T r i c - H C l ( p H 7 . 8 ) をビーズに添加する。生じた混合物を上下ピペット操作で混合し、その後、約 2 分間、室温に置く。溶液が透明になるまで、チューブを約 0.5 ~ 1 分間磁気スタンドに配置する。溶出された

10

20

30

40

50

DNAを包含している上澄みを、新しいEppendorf DNA LoBind管に移す。その後、生成物は、ライブラリーを生成するために使用することができ、および、配列決定に先立って高感度DNAチップを使用して、Agilent Bioanalyzer上で量を計られる。

#### 【0140】

いくつかの実施形態では、この時点までのライブラリー調製のすべての工程が、単一ボリューム中で行なわれることが見られる。場合によっては、単一ボリュームは単一チューブである。ある場合には、単一ボリュームは、プレート中の単一ウェルである。任意選択的に、ライブラリー生成の後に、DNAは、ビーズに基づく方法またはアガロースゲルに基づく方法を使用して、選択されたサイズであり、および、ライブラリーの量は、配列決定に先立ち、高感度DNAチップを使用して、Agilent Bioanalyzer上で定量される。

10

#### 【0141】

核酸の標的化切断のためのエンドヌクラーゼ

本明細書に開示される標的化配列決定法は、部位特異的な、標的化可能な、および/または操作されたヌクラーゼもしくはヌクラーゼ・システムを使用して核酸の切断を標的化する工程を含む。そのようなヌクラーゼは、ゲノムの、cDNAまたは他の核酸分子の中の所望の位置に二本鎖切断(DSB)を生じ得る。他の例では、ヌクラーゼは、一本鎖切断を生じ得る。場合によっては、2つのヌクラーゼが使用され、その各々は一本鎖切断を生成する。本明細書の開示と調和する多くの切断酵素は、一本鎖か二本鎖のエキソヌクラーゼ活性のためのアクセス可能な末端を有する分子を産出する形質を共有する。

20

#### 【0142】

本明細書の開示と調和するエンドヌクラーゼは、クラスター化された短い回文反復(CRISPR)/Casシステム・タンパク質-gRNA複合体、ジンクフィンガー・ヌクラーゼ(ZFN)、および転写活性化因子様エフェクターヌクラーゼ(Transcription activator like effector nucleases)から選択される少なくとも1つを多様性を含む。いくつかの実施形態では、gRNAは、標的核酸における少なくとも1つの部位に相補的である。また、他のプログラム可能な、核酸配列特異的エンドヌクラーゼもまた、本明細書の開示に調和する。

30

#### 【0143】

ジンクフィンガー・ヌクラーゼ(ZFN)、転写活性化因子様エフェクターヌクラーゼ(TALEN)、操作されたホーミングエンドヌクラーゼなどの、操作されたヌクラーゼ、および、CRISPR/Cas9またはCPF1などの、RNAでガイドされる、またはDNAでガイドされるエンドヌクラーゼ、および/またはアルゴノートシステム(Argonate systems)は、本開示の方法のうちのいくつかを実行するのに特に適切である。付加的に、または代替的に、c2c2ヌクラーゼを含むCRISPR/Casシステムなどの、RNAを標的とするシステムが使用されてもよい。

#### 【0144】

本明細書で開示される方法は、タイプI、タイプII、タイプIII、タイプIV、タイプV、またはタイプVI CRISPRシステムなどの、CRISPRシステムを使用して、標的核酸を切断する工程を含む場合がある。CRISPR/Casシステムは、多タンパク質のシステムまたは単一のエフェクタータンパク質システムであり得る。多タンパク質の、またはClass 1の、CRISPRシステムは、タイプI、タイプIII、およびタイプIVシステムを含む。代替的に、クラス2システムは、単一のエフェクター分子を含み、およびタイプII、タイプV、およびタイプVIを含む。

40

#### 【0145】

本明細書に開示されるいくつかの標的化配列決定法において使用されるCRISPRシステムは、単一または複数のエフェクタータンパク質を含む場合がある。エフェクタータンパク質は、1つまたは複数のヌクラーゼドメインを含む場合がある。エフェクタータ

50

ンパク質は、DNAまたはRNAを標的とする場合があり、およびDNAまたはRNAは、単鎖の場合も二本鎖場合もある。CRISPRシステムは、単一または複数のガイドRNAを含む場合がある。gRNAはcrRNAを含み得る。gRNAは、crRNAとtracrRNAの配列を有するキメラRNAを含む場合がある。gRNAは、別個のcrRNAおよびtracrRNAを含む場合がある。ターゲット核酸配列は、プロトスペーサー隣接モチーフ(proto spacer adjacent motif)(PAM)またはプロトスペーサー側方部位(proto spacer flanking site)(PFS)を含む場合がある。PAMまたはPFSは、標的またはプロトスペーサー部位の3'または5'であり得る。標的配列の切断は、平滑末端、3'オーバーハング、または5'オーバーハングを生成する場合がある。

10

## 【0146】

gRNAは、スペーサー配列を含む場合がある。スペーサー配列は、標的配列またはプロトスペーサー配列に相補的な場合がある。スペーサー配列は、長さが10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、または36ヌクレオチドであり得る。いくつかの例では、スペーサー配列は、長さが10ヌクレオチド未満または36ヌクレオチド超であり得る。

## 【0147】

gRNAは、反復配列を含む場合がある。場合によっては、反復配列は、gRNAの二本鎖部分の一部である。反復配列は、長さが10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、または50ヌクレオチドであり得る。いくつかの例では、スペーサー配列は、長さ10ヌクレオチド未満、または50ヌクレオチド超であり得る。

20

## 【0148】

gRNAは、1つ以上の合成ヌクレオチド、非自然発生のヌクレオチド、修飾があるヌクレオチド、デオキシリボヌクレオチド、またはそれらの任意の組み合わせを含み得る。付加的に、または代替的に、gRNAは、ヘアピン、リンカー領域、一本鎖領域、二本鎖領域、またはそれらの任意の組み合わせを含み得る。付加的に、または代替的に、gRNAは、シグナル分子またはリポーター分子を含み得る。

30

## 【0149】

gRNAsは遺伝学的かエピゾームのDNAによってコードされる場合がある。gRNAsには、CRISPRヌクレアーゼに付随して、または連続して、提供あるいは送達され得る。ガイドRNAは、化学的に合成されるか、あるいはインビトロで転写されるか、さもなければ当該技術分野において既知の標準的なRNA生成技術を使用して生成され得る。

## 【0150】

CRISPRシステムは、タイプII CRISPRシステム、例えば、Cas9システムであり得る。タイプIIヌクレアーゼは、場合によりRuvCおよびHNHヌクレアーゼドメインを含む単一のエフェクタータンパク質を含む場合がある。場合によっては、機能的タイプIIヌクレアーゼは2つ以上のポリペプチドを含む場合があり、それぞれはヌクレアーゼドメインまたはその断片を含む。ターゲット核酸配列は、3'プロトスペーサー隣接モチーフ(PAM)を含む場合がある。いくつかの例では、PAMは、標的核酸の5'であり得る。ガイドRNA(gRNA)は単一のキメラgRNAを含む場合があり、それはcrRNAとtracrRNA配列の両方を含んでいる。あるいは、gRNAは、一組の2つのRNA、例えばcrRNAとtracrRNAを含む場合がある。タイプIIヌクレアーゼは、二本鎖切断を生成する場合があり、それは場合により2つの平滑末端を作る。場合によっては、タイプII CRISPRヌクレアーゼは、ヌクレアーゼが一本鎖切断だけを生成するように、ニックアーゼであるように操作される。そのような場合

40

50

、2つの一本鎖切断がニッカーゼによって生成されるように、2つの異なる核酸配列がgRNAによって標的とされ得る。いくつかの例では、2つの一本鎖切断は、効果的に二本鎖切断を生じる。場合によっては、タイプIIニッカーゼが2つの一本鎖切断を生成するために使用される場合、生じる核酸遊離末端は平滑であるか、3'オーバーハングを有するか、または5'オーバーハングを有するかのいずれかであり得る。いくつかの例では、タイプIIヌクレアーゼは、標的配列に結合するが切断しないように、触媒現象的に無効な場合がある。例えば、タイプIIヌクレアーゼは、RuvCとHNHの両方のドメインにおいて変異を有している場合があり、それによって、両方のヌクレアーゼドメインが機能を果たさなくなる。タイプII CRISPRシステムは、3つのサブタイプ、具体的には、タイプII-A、タイプII-B、タイプII-Cのうちの一つであり得る。

10

## 【0151】

CRISPRシステムは、タイプV CRISPRシステム、例えば、Cpf1、C2c1、C2c3システムであり得る。タイプVヌクレアーゼは、単一のエフェクタータンパク質を含む場合があり、それは場合により単一のRuvCヌクレアーゼドメインを含む。他の場合では、機能タイプVヌクレアーゼは、2つ以上のポリペプチド間で分割されたRuvCドメインを含む。そのような場合、標的核酸配列は、5'PAMまたは3'PAMを含み得る。ガイドRNA(gRNA)は、Cpf1である場合などの、単一のgRNAまたは単一のcrRNAを含み得る。場合によっては、tracrRNAは必要とされない。他の例では、C2c1が使用される場合など、gRNAは単一のキメラgRNAを含む場合があり、そのときcrRNAとtracrRNAの両方の配列を含むか、またはgRNAは一組の2つのRNA、例えば、crRNAとtracrRNAを含み得る。タイプV CRISPRヌクレアーゼは、二本鎖切断を生成する場合があり、それは場合により5'オーバーハングを生成する。場合によっては、タイプV CRISPRヌクレアーゼは、ヌクレアーゼが一本鎖切断だけを生成するように、ニッカーゼであるように操作される。そのような場合、2つの一本鎖切断がニッカーゼによって生成されるように、2つの異なる核酸配列がgRNAによって標的とされ得る。いくつかの例では、2つの一本鎖切断は効果的に二本鎖切断を生じる。場合によっては、タイプVニッカーゼが2つの一本鎖切断を生成するために使用される場合、生じる核酸遊離末端は平滑であるか、3'オーバーハングを有するか、または5'オーバーハングを有するかのいずれかであり得る。いくつかの例では、タイプVヌクレアーゼは、標的配列に結合するが切断しないように、触媒現象的に無効な場合がある。例えば、タイプVヌクレアーゼは、RuvCドメインにおいて変異を有している場合があり、それによって、ヌクレアーゼドメインが機能を果たさなくなる。

20

30

## 【0152】

CRISPRシステムはタイプVI CRISPRシステム、例えば、C2c2システムであり得る。タイプVIヌクレアーゼは、HEPNドメインを含む場合がある。いくつかの例では、タイプVIヌクレアーゼは2つ以上のポリペプチドを含む場合があり、それぞれはHEPNヌクレアーゼドメインまたはその断片を含む。そのような場合、ターゲット核酸配列は一本鎖RNAなどのRNAであり得る。タイプVI CRISPRシステムを使用する時、標的核酸はプロトスペーサー側方部位(PFS)を含み得る。PFSは、3'または5'、あるいは標的またはプロトスペーサー配列であり得る。ガイドRNA(gRNA)は、単一のgRNAまたは単一のcrRNAを含む場合がある。場合によっては、tracrRNAは必要とされない。他の例では、gRNAは単一のキメラgRNAを含む場合があり、そのときcrRNAとtracrRNAの両方の配列を含むか、またはgRNAは一組の2つのRNA、例えば、crRNAとtracrRNAを含み得る。いくつかの例では、タイプVIヌクレアーゼは、標的配列に結合するが切断しないように、触媒現象的に無効な場合がある。例えば、タイプVIヌクレアーゼは、HEPNドメインにおいて変異を有している場合があり、それによって、ヌクレアーゼドメインが機能を果たさなくなる。

40

## 【0153】

50

核酸にガイドされるヌクレアーゼを含む、本開示における使用のための適切なヌクレアーゼの非限定的な例は、C2c1、C2c2、C2c3、Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9 (Csn1およびCsx12としても知られる)、Cas10、Cpf1、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx100、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、それらの相同体、それらのオーソログ、またはそれらの修飾版を含む。

#### 【0154】

本明細書に開示されたいくつかの方法では、アルゴノート (Argonote) (Ago) システムが、ターゲット核酸配列を切断するために使用されてもよい。Agoタンパク質は、原核生物、真核生物、または古細菌に由来する場合がある。標的核酸は、RNAまたはDNAであり得る。DNA標的は、単鎖または二本鎖であり得る。いくつかの例では、標的核酸は、プロトスペーサー隣接モチーフまたはプロトスペーサー側方配列に等価の配列などの、特定の標的側方配列を必要としない。Agoタンパク質は、二本鎖切断または一本鎖切断を生じる場合がある。いくつかの例では、Agoタンパク質が一本鎖切断を形成する時、2つのAgoタンパク質は、二重鎖切断を生成するために組み合わせて使用されてもよい。いくつかの例では、Agoタンパク質は1、2、またはそれ以上のヌクレアーゼドメインを含む。いくつかの例では、Agoタンパク質は1、2、またはそれ以上の触媒ドメインを含む。1つ以上のヌクレアーゼまたは触媒ドメインはAgoタンパク質において変異させられる場合があり、それによって、一本鎖切断を生成することができるニックターゼタンパク質が生成される。他の例では、1つ以上のヌクレアーゼまたはAgoタンパク質の触媒ドメインにおける変異は、結合するが標的核酸を切断しない場合があるような、触媒現象的に無効なAgoタンパク質を生成する。

#### 【0155】

Agoタンパク質は、ガイドする核酸によって標的核酸配列に向けられる。多くの例では、ガイドする核酸は、ガイドDNA (gDNA) である。gDNAは、5'にリン酸化された末端を有し得る。gDNAは、単鎖または二本鎖であり得る。単鎖gDNAは、長さが10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、または50ヌクレオチドであり得る。いくつかの例では、gDNAは、長さが10ヌクレオチド未満であり得る。いくつかの例では、gDNAは、長さが50ヌクレオチド超であり得る。

#### 【0156】

アルゴノートに媒介された切断は、平滑末端、5'オーバーハング、または3'オーバーハングを生成する場合がある。いくつかの例では、1つ以上のヌクレオチドが、切断の間または後に、標的部位から除去される。

#### 【0157】

アルゴノートタンパク質は、内発的にまたは組み換え的に発現される場合がある。アルゴノートは、染色体上に、染色体外に、またはプラスミド、合成染色体、または人工染色体上に、コードされる場合がある。付加的に、または代替的に、アルゴノートタンパク質は、ポリペプチドをコードするポリペプチドまたはmRNAとして提供される場合がある。そのような例では、ポリペプチドまたはmRNAは、ペプチド、ナノ粒子、またはウイルス粒子の使用によるなどの、当該技術分野において既知の標準的な手段によって送達される場合がある。

#### 【0158】

ガイドDNAは、遺伝子の、またはエピゾームのDNAによって提供される場合がある。いくつかの例では、gDNAは、RNAまたはmRNAから逆転写される。いくつかの

10

20

30

40

50

例では、ガイドDNAは、Agotaタンパク質に付随して、または連続して、提供あるいは送達され得る。ガイドDNAは、化学的に合成、構築される場合があり、さもなければ当該技術分野において既知の標準的なDNA生成技術を使用して、生成される。ガイドDNAは、ゲノムDNA、エピゾームのDNA分子、単離された核酸分子、または核酸分子の他の任意の供給源から、切断されるか、放出されるか、さもなければ取り出され得る。

【0159】

ヌクレアーゼ融合蛋白質は、組み換え的に発現される場合がある。ヌクレアーゼ融合蛋白質は、染色体上に、染色体外に、またはプラスミド、合成染色体、または人工染色体上に、コードされる場合がある。ヌクレアーゼとクロマチンリモデリング酵素は、別々に操作され、その後共有結合でリンクされる場合がある。ヌクレアーゼ融合蛋白質は、ポリペプチドをコードするポリペプチドまたはmRNAとして提供される場合がある。そのような例では、ポリペプチドまたはmRNAは、ペプチド、ナノ粒子、またはウイルス粒子の使用によるなどの、当該技術分野において既知の標準的な手段によって送達される場合がある。

10

【0160】

ガイド核酸は、適合する核酸でガイドされるヌクレアーゼと複合体化し、および標的配列とハイブリダイズするする場合があり、それによって、標的配列にヌクレアーゼを向ける。ガイド核酸と複合体化することができる、対象の核酸でガイドされるヌクレアーゼは、ガイド核酸と適合する核酸でガイドされるヌクレアーゼとして言及され得る。同様に、核酸でガイドされるヌクレアーゼと複合体化することができるガイド核酸は、核酸でガイドされるヌクレアーゼと適合するガイド核酸として言及され得る。

20

【0161】

ガイド核酸はDNAであり得る。ガイド核酸はRNAであり得る。ガイド核酸は、DNAとRNAの両方を含む場合がある。ガイド核酸は、修飾された非自然発生のヌクレオチドを含んでもよい。ガイド核酸がRNAを含む場合、RNAガイド核酸は、本明細書に開示されるように、プラスミド、線状構築物、または編集カセットなどのポリヌクレオチド分子におけるDNA塩基配列によってコードされてもよい。

【0162】

ガイド核酸はガイド配列を含む場合がある。ガイド配列は、標的配列とハイブリダイズして、複合体化された核酸でガイドされるヌクレアーゼの配列特異的結合を標的配列に向けるために、標的ポリヌクレオチド配列と十分な相補性があるポリヌクレオチド配列である。ガイド配列とその対応する標的配列の間の相補性の程度は、適切なアラインメント・アルゴリズムを使用して、最適にアラインメントしたとき、約50%、60%、75%、80%、85%、90%、95%、97.5%、99%、またはそれらの値以上である。最適なアラインメントは、配列をアラインメントするための任意の適切なアルゴリズムを使用して求められ得る。いくつかの態様において、ガイド配列は、長さが約5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、75ヌクレオチド、またはそれ以上である。いくつかの態様において、ガイド配列は、長さが約75ヌクレオチド未満、50ヌクレオチド未満、45ヌクレオチド未満、40ヌクレオチド未満、35ヌクレオチド未満、30ヌクレオチド未満、25ヌクレオチド未満、20ヌクレオチド未満である。好適には、ガイド配列は、長さが10~30ヌクレオチドである。ガイド配列は、長さが10~25ヌクレオチドであり得る。ガイド配列は、長さが10~20ヌクレオチドであり得る。ガイド配列は、長さが15~30ヌクレオチドであり得る。ガイド配列は、長さが20~30ヌクレオチドであり得る。ガイド配列は、長さが15~25ヌクレオチドであり得る。ガイド配列は、長さが15~20ヌクレオチドであり得る。ガイド配列は、長さが20~25ヌクレオチドであり得る。ガイド配列は、長さが22~25ヌクレオチドであり得る。ガイド配列は、長さが15ヌクレオチドであり得る。ガイド配列は、長さが16ヌクレオチドであり得る。ガイド配列は、長さが17ヌクレオチドであり得る。ガイド配列は、長さが18ヌクレオチドであり得る。ガイド配列は、長さが19ヌ

30

40

50

クレオチドであり得る。ガイド配列は、長さが20ヌクレオチドであり得る。ガイド配列は、長さが21ヌクレオチドであり得る。ガイド配列は、長さが22ヌクレオチドであり得る。ガイド配列は、長さが23ヌクレオチドであり得る。ガイド配列は、長さが24ヌクレオチドであり得る。ガイド配列は、長さが25ヌクレオチドであり得る。

### 【0163】

ガイド核酸は、スキヤフォールド配列を含む場合がある。一般に、「スキヤフォールド配列」は、標的化可能なヌクレアーゼ複合体の生成を促進するために十分な配列を有している任意の配列を含み、ここで標的化可能なヌクレアーゼ複合体は核酸でガイドされるヌクレアーゼとガイド核酸を含み、ガイド核酸はスキヤフォールド配列とガイド配列を含む。標的化可能なヌクレアーゼ複合体の生成を促進するためのスキヤフォールド配列内の十分な配列は、2次構造の形成するに關与する1つまたは2つの配列領域などの、スキヤフォールド配列内の2つの配列領域の長さに沿った相補性の程度を含む場合がある。場合によっては、前記1つまたは2つの配列領域は、同じポリヌクレオチド上に含まれるか、またはコードされる。場合によっては、前記1つまたは2つの配列領域は、別個のポリヌクレオチド上に含まれるか、またはコードされる。最適なアラインメントは、任意の適切なアラインメント・アルゴリズムによって求められる場合があり、1つか2つの配列領域内の自己相補性などの、2次構造をさらに説明し得る。いくつかの態様では、最適にアラインメントされたとき2つのうちより短いものの長さに沿った、1つまたは2つの配列領域間の、相補性の程度は、約25%、30%、40%、50%、60%、70%、80%、90%、95%、97.5%、99%、またはそれら以上である。いくつかの態様では、2つの配列領域の少なくとも1つは、長さが約5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、40、50ヌクレオチド、またはそれらの値以上である。いくつかの態様では、2つの配列領域の少なくとも1つは、長さが約10~30ヌクレオチドである。2つの配列領域の少なくとも1つは、長さが10~25ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが10~20ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが15~30ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが20~30ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが15~25ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが15~20ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが20~25ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが22~25ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが15ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが16ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが17ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが18ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが19ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが20ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが21ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが22ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが23ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが24ヌクレオチドであり得る。2つの配列領域の少なくとも1つは、長さが25ヌクレオチドであり得る。

### 【0164】

対象ガイド核酸のスキヤフォールド配列は、2次構造を含む場合がある。2次構造はシュドノット領域を含む場合がある。いくつかの例では、ガイド核酸と核酸でガイドされるヌクレアーゼの適合性は、ガイドRNAのシュドノット領域に内在または隣接する配列によって、少なくとも部分的に求められる。場合によっては、核酸でガイドされるヌクレアーゼに対する、ガイド核酸の結合キネティクスは、ひとつには、スキヤフォールド配列内の2次構造によって求められる。場合によっては、核酸でガイドされるヌクレアーゼに対する、ガイド核酸の結合キネティクスは、ひとつには、スキヤフォールド配列を有する核酸配列によって求められる。

10

20

30

40

50

## 【0165】

本開示の態様において、用語「ガイド核酸」は、1) 標的配列にハイブリダイズすることができるガイド配列、および、2) 本明細書に記載されるような核酸でガイドされるヌクレアーゼと相互作用することができるか、または複合体化することができる、スキャフォールド配列を含む、ポリヌクレオチドを指す。

## 【0166】

ガイド核酸は、2つの要素が、1つの、標的配列を切断することができる機能的な標的化可能なヌクレアーゼ複合体を形成し得るとき、核酸でガイドされるヌクレアーゼと適合し得る。多くの場合、適合するガイド核酸に適合するスキャフォールド配列は、天然の核酸でガイドされるヌクレアーゼ遺伝子座に隣接している配列を走査することによって見つ

10

## 【0167】

核酸でガイドされるヌクレアーゼは、ヌクレアーゼの内在性ホスト内に存在しないガイド核酸に適合する場合がある。このような直交するガイド核酸は、経験的なテストによって判定することができる。直交するガイド核酸は、異なる細菌種に由来するものであっても、合成されたものであっても、あるいは、自然界に存在しないように操作されたものであってもよい。

## 【0168】

共通の核酸ガイドヌクレアーゼに適合する直交ガイド核酸は、1つまたは複数の共通の特徴を含んでいてもよい。共通の特徴は、シュドノット領域の外側の配列を含む場合がある。共通の特徴は、シュドノット領域を含む場合がある。共通の特徴は、一次配列または2次構造を含む場合がある。

20

## 【0169】

ガイド核酸は、ガイド配列を、ガイド配列が標的配列に相補的であるように変更することにより、所望の標的配列を標的にするように操作される場合があり、それによってガイド配列と標的配列の間のハイブリダイゼーションが可能になる。操作されたガイド核酸を有するガイド配列は、操作されたガイド核酸と呼ばれることがある。操作されたガイド核酸は、多くの場合、非自然発生で、自然界では見出されない。

## 【0170】

ガイドRNA分子は、配列決定するために単離するべき標的配列と塩基対となる配列を含む。いくつかの実施形態では、塩基対合は完全であり、一方、いくつかの実施形態では、塩基対合は部分的か、または対がない塩基を、非標的配列に対になる塩基と共に、含む。

30

## 【0171】

ガイドRNAは、RNA「ヘアピン」構造を形成する領域または複数の領域を含む場合がある。そのような領域または複数の領域は、回文配列を部分的にまたは完全に含み、その結果、領域の5'末端と3'末端は、二本鎖「ステム」構造を形成するように互いにハイブリダイズする場合があり、それは、いくつかの実施形態では、互いに二重鎖環の中の本鎖の各々をテザーする非回文の環によってキャップされる。

40

## 【0172】

いくつかの実施形態では、ガイドRNAは、tracrRNAステムループなどの、ステムループを含む。tracrRNAステムループなどのステムループは、Cas9 DNAエンドヌクレアーゼなどの核酸エンドヌクレアーゼと複合体化または結合し得る。代替的に、ステムループは、Cas9以外のエンドヌクレアーゼ、または塩基除去酵素、メチルトランスフェラーゼ、または1つ以上のDNAポリメラーゼ酵素を妨害する他の核酸修飾活性を有する酵素など、エンドヌクレアーゼ以外の核酸修飾酵素と複合体化してもよい。

## 【0173】

tracrRNA / CRISPR / Endonuclease システムは、真正細菌や

50

古細菌の原核生物が、既知の配列のウイルスに繰り返し感染することで細胞が耐性を獲得する適応免疫システムとして特定された。例えば、Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA et al. (2011) "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III" *Nature* 471 (7340): 602-7. doi:10.1038/nature09886. PMC 3070239. PMID 21455174; Terns MP, Terns RM (2011) "CRISPR-based adaptive immune systems" *Curr Opin Microbiol* 14 (3): 321-7. doi:10.1016/j.mib.2011.03.005. PMC 3119747. PMID 21531607; Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity" *Science* 337 (6096): 816-21. doi:10.1126/science.1225829. PMID 22745249; および、Brouns SJ (2012) "A swiss army knife of immunity" *Science* 337 (6096): 808-9. doi:10.1126/science.1227253. PMID 22904002を参照されたい。当該システムは、真核細胞における標的化遺伝子突然変異誘発を導くために、順応してきた。Wenzhi Jiang, Huanbin Zhou, Honghao Bi, Michael Fromm, Bing Yang, and Donald P. Weeks (2013) "Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice" *Nucleic Acids Res.* Nov 2013; 41(20): e188, Published online (Aug 31, 2013). doi: 10.1093/nar/gkt780、およびその中の参考文献を参照されたい。

#### 【0174】

本明細書に企図されたように、ガイドRNAは、Cas9エンドヌクラーゼなどのDNAエンドヌクラーゼに配列特異性を提供するために、いくつかの実施形態において、使用される。これらの実施形態では、ガイドRNAは、Cas9などのエンドヌクラーゼに結合するか、またはエンドヌクラーゼによって結合されるヘアピン構造を含み（他のエンドヌクラーゼは、いくつかの実施形態において、代替または追加として企図される）、および、ガイドRNAは、配列決定ライブラリーまたは配列決定反応から取り除かれることになっている配列に結合するか、特異的に結合するか、排他的に結合する、認識配列をさらに含む。ガイドRNAの中の認識配列の長さは、配列排除プロセスにおける所望の特異性の程度によって変化し得る。サンプル中で頻繁に生じる配列を含むか、または差動的に豊富な配列（ATリッチなゲノムサンプル中でのATの存在量またはGCリッチなゲノムサンプル中でGCの存在量）を含む、短い認識配列は、相対的に大きな数の部位を同定する可能性が高く、従って、エンドヌクラーゼ活性、塩基切除、メチル化、または少なくとも1つのDNAポリメラーゼ活性に干渉する他の活性などの、頻繁な核酸修飾を導く可能性が高い。サンプル中で希に生じる配列を含むか、または不十分に代表された塩基の組み合わせ（ATリッチなゲノムサンプル中でのGCの存在量またはGCリッチなゲノムサンプル中でATの存在量）を含む、長い認識配列は、相対的に小さな数の部位を同定する可能性が高く、従って、エンドヌクラーゼ活性、塩基切除、メチル化、または少なくとも1つのDNAポリメラーゼ活性に干渉する他の活性などの、希な核酸修飾を導く可能性が高い。従って、本明細書に開示されるように、いくつかの実施形態では、配列反応から認識配列の長さまたは内容への修飾を介して、配列除去の頻度を加減してもよい

## 【0175】

ガイドRNAは、本明細書に記載される開示と一致する多くの方法によって合成される。標準的な合成技術は、大きな分量のガイドRNAを産生するために、および/または非常に反復的な標的領域のために、使用されてもよく、そのことは、多くの不要な遺伝子座を標的とするほんの少数のガイドRNA分子だけを必要とし得る。二本鎖DNA分子は、RNA部位特異的結合配列、Cas9タンパク質のガイドRNA配列、およびT7プロモーター部位を含み得る。場合によっては、二本鎖DNA分子は、約100bp以下の長さであり得る。T7ポリメラーゼは、Cas9タンパク質のための標的RNA配列とガイドRNA配列を含み得る、一本鎖RNA分子を作製するために使用することができる。

10

## 【0176】

ガイドRNA配列は、数多くの方法によって設計され得る。例えば、いくつかの実施形態では、ヒトゲノムの非遺伝的な繰り返し配列は、例えば、100bpのスライディングウィンドウに分割される。二本鎖DNA分子は、フォトリソグラフィを使用して、マイクロアレイ上で並列に合成することができる。

## 【0177】

ウィンドウは、サイズにおいて変化する場合がある。30量体の標的配列は、標的設計配列の5'末端側方N-G-Gの短いトリヌクレオチド・プロトスペーサー隣接モチーフ(PAM)配列を用いて設計することができ、それは場合により、切断を容易にする。とりわけ、全体が参照によって本明細書に組み込まれる、Giedrius Gasionas et al., (2012) "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria" *Proc. Natl. Acad. Sci. USA*, Sep 25, 109(39): E2579-E2586、を参照されたい。冗長な配列は除去することができ、および、残りの配列は、refseq、ENSEMBLに対するハイブリダイゼーションを回避するために、ヒトゲノムに対してサーチエンジン(例えば、BLAST)を使用して、およびこれらの部位でヌクレアーゼ活性を回避するために、他の遺伝子データベースを使用して、解析することができる。ガイドRNA標的配列にユニバーサルCas9トレーサーRNA配列を付加し、T7プロモーターでフランキングすることができる。T7プロモーター部位上流の配列は、合成することができる。ヒトゲノムの標的領域は繰り返しが多いため、多くの実施形態では、相対的に少数のガイドRNA分子が、NGSライブラリー分子のより大きな割合を消化することになる。

20

30

## 【0178】

タンパク質のコードする遺伝子のわずか約50%がNGG PAM(フォトスペーサー隣接モチーフ)配列を含むエクソンを有すると推定されるが、Cas9切断システムで標的とすることができるゲノムの割合を増加させるために、複数の戦略が本明細書に提供される。例えば、DNA領域でPAM配列が利用可能でない場合、PAM配列からなるヘルパーDNAと結合したガイドRNAを用いた組み合わせ戦略により、PAM配列を導入することができる。ヘルパーDNAは、合成のもの、および/または単鎖のものであり得る。ヘルパーDNAのPAM配列は、NGSライブラリーのgDNAノックアウトターゲットと相補的ではないと考えられ、従って、標的NGSライブラリーテンプレートには結合しないかもしれないが、ガイドRNAには結合し得る。ガイドRNAは、ターゲット配列とPAM配列からなるヘルパーDNAの両方にハイブリダイズして、Cas9システムによって認識されるハイブリッドDNA:RNA:DNA複合体を形成するように設計することができる。

40

## 【0179】

PAM配列は、単鎖のオーバーハングまたはヘアピンとして表現される場合がある。ヘアピンは、場合により、任意選択的に劣化される場合がある、修飾されたヌクレオチドを含み得る。例えば、ヘアピンはウラシルを含むことができ、これはウラシルDNAグリコ

50

シラーゼによって分解され得る。

【0180】

PAM配列を含むDNAの使用の代替として、PAM配列の必要のない修飾されたCas9タンパク質またはPAM配列へのより低い感度を有する修飾されたCas9は、ヘルパーDNA塩基配列を必要とせずに、使用されてもよい。

【0181】

さらなる場合では、Cas9の認識に用いるガイドRNAの配列は、複数の部位を緊密に切断するためのデュアルカッティングシステムとして機能させるために、1つの末端で長くされ得、および反転され得る。ガイドRNA配列は、NGS DNAライブラリー標的上で2つの切断を作り出すことができる。このことは、制限された距離内の鎖を交互にするための単一のガイドRNAを設計することにより、達成され得る。ガイドRNAの一端は二本鎖DNAライブラリーの順鎖に結合し得、および他端は逆鎖に結合し得る。ガイドRNAの各末端は、PAM配列とCas9結合ドメインを含み得る。このことは、所定の距離だけ離れた同じDNA塩基配列由来のNGSライブラリー分子の2重の二本鎖切断をもたらし得る。いくつかの実施形態は、ガイドRNA分子の生成に関する。ガイドRNA分子は、場合によっては、DNA鋳型から転写される。T7ポリメラーゼ、RNA Pol I、RNA Pol II、RNA Pol III、オルガネラRNAポリメラーゼ、ウイルスRNAポリメラーゼ、真正細菌、または古細菌ポリメラーゼなど、多くのRNAポリメラーゼが使用され得る。場合によっては、ポリメラーゼはT7である。

10

【0182】

ガイドRNA生成テンプレートは、T7ポリメラーゼ、RNA Pol I、RNA Pol II、RNA Pol III、オルガネラRNAポリメラーゼ、ウイルス性RNAポリメラーゼ、真正細菌、または古細菌ポリメラーゼによって方向づけられる転写に適合したプロモーターなどの、プロモーターを含む。場合によっては、プロモーターはT7プロモーターである。

20

【0183】

ガイドRNA鋳型は、場合により、タグ配列をコードする。タグ配列は、メチラーゼ、塩基除去酵素、エンドヌクレアーゼなどの核酸修飾酵素に結合する。非標的部位に結合されたより大きなガイドRNA分子のコンテキストにおいて、タグ配列は、核酸の非標的領域に酵素を結びつけ(tether)、非標的部位に活性を向ける。例示的な、結びつけられる酵素(tethered enzyme)は、Cas9などのエンドヌクレアーゼである。

30

【0184】

ガイドRNAテンプレートは、リボソームRNA配列、グロビン蛋白質をコードする配列、トランスポゾンコードする配列、レトロウイルス配列をコードする配列、テロメア配列を含む配列、サブテロメア反復を含む配列、セントロメア配列を含む配列、イントロン配列を含む配列、Alu反復を含む配列、SINE反復を含む配列、LINE反復を含む配列、二核酸反復を含む配列、三核酸反復を含む配列、四核酸反復を含む配列、ポリA反復を含む配列、ポリT反復を含む配列、ポリC反復を含む配列、ポリG反復を含む配列、ATリッチな配列を含む配列、またはGCリッチな配列を含む配列に対応する核酸と相補的である。

40

【0185】

多くの場合において、タグ配列は、部分的または全体のステムループ構造などの、ステムループを含む。ステムループ構造の「ステム」は、場合によっては回文配列でコードされ、完全なものか、ステムに少なくとも1つの「キंक(kink)」またはターンを導入するために中断されたものかのいずれかである。ステムループ構造の「ループ」は、ほとんどの場合、ステム塩基のペアリングには関与しない。場合によっては、ステムループは、本明細書に組み込まれた参考文献に開示されたtracr配列などの、tracr配列によってコードされる。いくつかのステムループは、例えば、Cas9または他のエンドヌクレアーゼを結合する。

50

## 【0186】

ガイドRNA分子は、認識配列をさらに含む。認識配列は、核酸ライブラリー配列セットから除去されるべき非標的配列に対して、完全または不完全に逆相補的である。RNAは、DNA-DNAハイブリッドでは生じない塩基対の組み合わせ（例えば、G:Uの塩基対）でハイブリッドすることができるため、結合するために、認識配列が非標的配列の正確な逆相補配列である必要はない。加えて、完全な塩基対合からの微小摂動は、場合により、許容される。

## 【実施例】

## 【0187】

以下の例は、本発明の様々な実施形態を例証する目的で与えられ、いかなるやり方でも本発明を制限することを意図していない。本実施例は、本明細書に記載される方法と共に、好ましい実施形態を現時点において代表するものであり、典型的なものであるが、本発明の範囲を制限する意図はない。それらにおける変更、および特許請求の範囲により定義されるような本発明の趣旨に含まれる他の用途を、当業者は見出すだろう。

## 【0188】

実施例1：B反応におけるCRISPR Cas9を用いた標的化配列決定

アダプター配列を持たないバーコード付きランダムプライマーを用いて、プライマー伸長反応を行なう。プライマー伸長生成物は、ビオチンddNTPで終止され、ストレプトアビジン・ビーズに結合される。ランダムBプライマーは、相補鎖を生成するために使用される。関心の特異的な遺伝子座を標的とするように設計されたCas9-ガイドRNA複合体を、前記ストレプトアビジン・ビーズに結合されたDNAに付加する。Cas9-ガイドRNA複合体は、標的部位にてDNAを切断する。放出されたDNAを、新しい管に移す。アダプターライゲーションを伴う任意のNGSライブラリー調製プロトコルを、配列決定のためのDNAを調製するために使用する。この例は、図1に例示される。

## 【0189】

ロングリード配列決定が使用される場合、アダプターは、精製された生成物上に直接ライゲーションされる。ショートリード配列決定が使用される場合、サンプルは、断片化され、ポリアダニル化され、ライゲーションされ、または、タグ付けが使用される。これらの3つの選択肢は、図4に示される。

## 【0190】

まだビーズに結合されている間に（つまり、CRISPR反応が起こる前）、サンプルにアダプター/バーコードをライゲーションする。タグ付けやライゲーションによる切断の後、2つ目のアダプター/バーコードを付加する。このことは、図5に例示される。

## 【0191】

CRISPR切断部位を使用して、複数の標的分子サイズを得る。このことは、図6に例示される。より短い断片が所望される場合、より高い密度のCRISPR切断を設計する（図7）。

## 【0192】

実施例2：COVID-19の標的化配列決定

患者サンプルから全RNAを取得し、逆転写酵素を用いて逆転写し、ビオチン化ジデオキシヌクレオチド三リン酸（ddNTP）で終止されたcDNAを作り、ストレプトアビジン磁気ビーズで捕捉する。第2鎖はビーズ上に合成される。HCoV-2ゲノムに特異的なCRISPR複合体は、サンプル中でCOVID-19核酸を標的とする。バックグラウンド核酸は、磁石を使ってサンプルから除去され、精製されたHCoV2核酸分子は、Nexterraを使って配列決定のために調製される。この例は、図2に例示される。

## 【0193】

実施例3：大腸菌ゲノムの標的化配列決定

大腸菌ゲノムDNAを含む生体試料を得た。Illuminaアダプターを含まないバーコード付きランダムプライマーを用いてAプライマー伸長反応を行なった。ストレプト

10

20

30

40

50

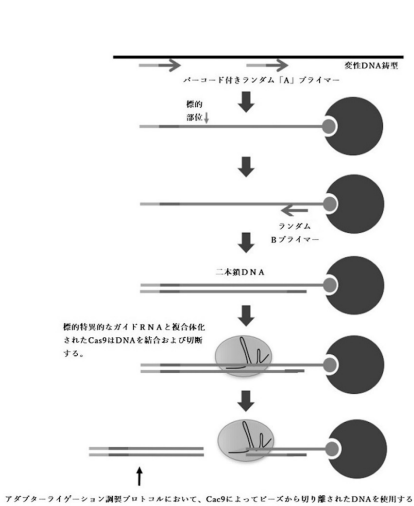
アビジン磁気ビーズでプライマー伸長分子を捕捉した。非特異的に結合されたDNA分子を除去するために、ビーズを3回洗浄した。相補鎖を作製するために、Illuminaアダプターを含まないバーコード付きランダムプライマーを用いてプライマー伸長反応を行なった。非特異的に結合されたDNA分子を除去するために、前記ビーズを3回洗浄した。ビーズを、1X Cas9緩衝剤において再懸濁した。事前に複合体化されたCas9 : sgRNAをストレプトアビジン・ビーズに加えた (sgRNAプールは、関心の大腸菌遺伝子座すべてを標的とするガイドRNAを含む)。Cas9に媒介された部位特異的切断が起こるまで、反応物を37℃で30~60分インキュベートした。切り取られた分子は、もはやビーズに結合されない。チューブを磁気スタンドに置き、上澄みを新しいチューブに移した。新しいチューブを磁気スタンドに戻し、上澄みを新しいチューブに移し、残留ビーズを取り除いた。NEBNext Ultra II DNA Library Prep Kitの試薬を用いて、標準的なエンドリペア、リン酸化、Aテーリング反応を行なった (NEBのプロトコルを使用)。NEBNext Ultra II DNA Library Prep Kitの試薬を用いて、Illuminaアダプターライゲーションを行なった (NEBのプロトコルを使用)。0.9X Ampureビーズのクリーンアップを行った (NEBのプロトコルを使用)。PCR増幅は全長のIllumina P5およびP7プライマーを用いて行った (NEBのプロトコルを使用)。最終的なAmpureビーズのクリーンアップを行った (NEBのプロトコルを使用)。精製した増幅産物の塩基配列を決定した。この実験からのデータを図3に示す。

【0194】

本発明の好ましい実施形態が本明細書において示され、かつ記載されているが、こうした実施形態はほんの一例として提供されているに過ぎないということが当業者にとって明白である。多くの変更、変化、および置換が、本発明から逸脱することなく、当業者の心に思い浮かぶであろう。本明細書に記載されるの実施形態の様々な代案が、利用され得ることを理解されたい。以下の請求項は本発明の範囲を定義するものであり、この請求項とその均等物の範囲内の方法、および構造体がそれによって包含されるものであるということが意図されている。

【図面】

【図1】



バーコード付きランダムプライマーを用いたプライマー伸長 (これら「A」プライマーは、Illuminaアダプター配列及そのアダプター配列を含まない)

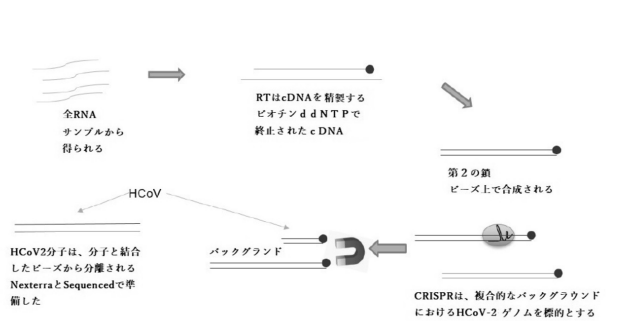
プライマー伸長生成物は、ピオチン dNTPs で終止され、ストレプトアビジン・ビーズに結合される

相補鎖を生成するために使用される、ランダム「B」プライマー

関心の特異的な遺伝子座を標的とするように設計された Cas9-ガイドRNA 複合体は、ストレプトアビジン・ビーズに結合された DNA に付加される。

Cas9 ガイドRNA 複合体は、標的的部位にて DNA を切断する。放出された DNA は、新しい管に移される。アダプターライゲーションを伴う任意の NGS フライアウプ 実験プロトコルが、配列決定のための DNA を調製するために使用される。

【図2】



10

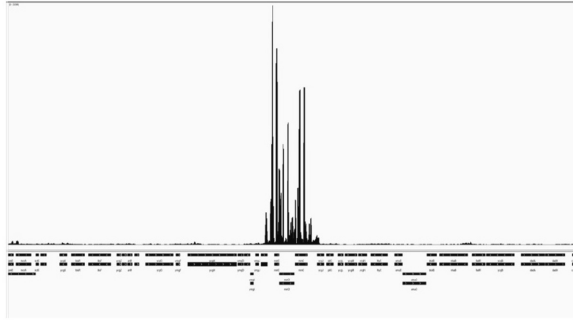
20

30

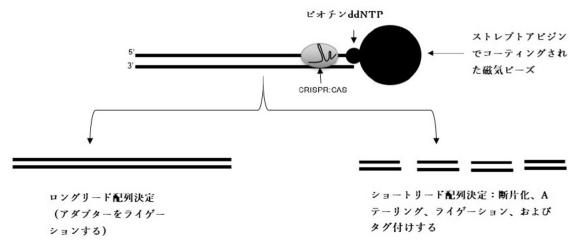
40

50

【 図 3 】

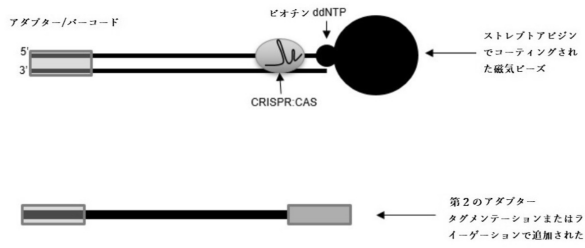


【 図 4 】

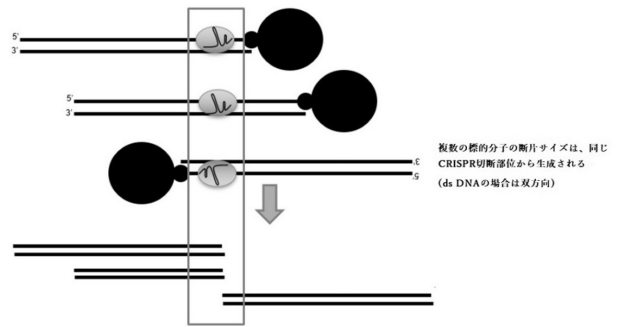


10

【 図 5 】

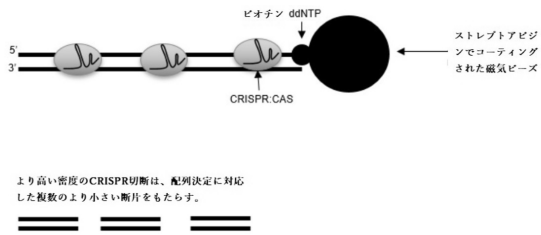


【 図 6 】



20

【 図 7 】



30

40

50

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2021/013598

A. CLASSIFICATION OF SUBJECT MATTER  
 IPC(B) - C12Q 1/6806; C12Q 1/6853; C12Q 1/6876 (2021.01)  
 CPC - C12N 15/1093; C12Q 1/6806; C12Q 1/6844; C12Q 1/686; C12Q 1/6874 (2021.02)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
 see Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2017/0101674 A1 (TOMA BIOSCIENCES, INC.) 13 April 2017 (13.04.2017) entire document	1-5
Y	US 2018/0010180 A1 (CENTRILLION TECHNOLOGY HOLDINGS CORPORATION) 11 January 2018 (11.01.2018) entire document	1-5
Y	US 2018/0016630 A1 (ROCHE SEQUENCING SOLUTIONS, INC.) 18 January 2018 (18.01.2018) entire document	1-5
Y	US 2019/0360044 A1 (THE BROAD INSTITUTE, INC. et al) 28 November 2019 (28.11.2019) entire document	5
A	US 2017/0247689 A1 (IGENOMX INTERNATIONAL GENOMICS CORPORATION) 31 August 2017 (31.08.2017) entire document	1-5

Further documents are listed in the continuation of Box C.  See patent family annex.

\* Special categories of cited documents:  
 "A" document defining the general state of the art which is not considered to be of particular relevance  
 "D" document cited by the applicant in the international application  
 "E" earlier application or patent but published on or after the international filing date  
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 "O" document referring to an oral disclosure, use, exhibition or other means  
 "P" document published prior to the international filing date but later than the priority date claimed  
 "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art  
 "&" document member of the same patent family

Date of the actual completion of the international search  
17 March 2021

Date of mailing of the international search report  
**APR 02 2021**

Name and mailing address of the ISA/US  
 Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
 P.O. Box 1450, Alexandria, VA 22313-1450  
 Facsimile No. 571-273-8300

Authorized officer  
 Blaine R. Copenheaver  
 Telephone No. PCT Helpdesk: 571-272-4300

10

20

30

40

50

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2021/013598

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

- 1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
- 2.  Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
- 3.  Claims Nos.: 6-22  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

10

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

- 1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
- 2.  As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
- 3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
- 4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

20

30

- Remark on Protest
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
  - The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
  - No protest accompanied the payment of additional search fees.

40

## フロントページの続き

## (51)国際特許分類

F I

テーマコード (参考)

<b>C 1 2 Q</b>	<b>1/6869(2018.01)</b>	C 1 2 Q	1/6869	Z
C 1 2 N	15/09 (2006.01)	C 1 2 N	15/09	Z

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

F ターム (参考)

QR08 QR14 QR62 QS25 QS36 QX02