

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6975507号  
(P6975507)

(45) 発行日 令和3年12月1日(2021.12.1)

(24) 登録日 令和3年11月10日(2021.11.10)

(51) Int.Cl.	F I
<b>C 1 2 Q 1/6876 (2018.01)</b>	C 1 2 Q 1/6876 Z N A Z
<b>C 1 2 Q 1/6869 (2018.01)</b>	C 1 2 Q 1/6869 Z
<b>C 1 2 Q 1/686 (2018.01)</b>	C 1 2 Q 1/686 Z
<b>C 1 2 Q 1/6844 (2018.01)</b>	C 1 2 Q 1/6844 Z
<b>C 1 2 N 15/11 (2006.01)</b>	C 1 2 N 15/11 Z

請求項の数 22 (全 56 頁)

(21) 出願番号	特願2018-530698 (P2018-530698)	(73) 特許権者	518201670
(86) (22) 出願日	平成28年12月8日 (2016.12.8)		ツインストランド・バイオサイエンシズ・ インコーポレイテッド
(65) 公表番号	特表2019-509015 (P2019-509015A)		TWINSTRAND BIOSCIEN CES, INC.
(43) 公表日	平成31年4月4日 (2019.4.4)		アメリカ合衆国、ワシントン 98121 、シアトル、エリオット・アベニュー 3 131、スイート 750
(86) 国際出願番号	PCT/US2016/065605	(74) 代理人	110001508
(87) 国際公開番号	W02017/100441		特許業務法人 津国
(87) 国際公開日	平成29年6月15日 (2017.6.15)	(72) 発明者	ローブ, ローレンス・エー
審査請求日	令和1年12月6日 (2019.12.6)		アメリカ合衆国、ワシントン 98005 、ベルビュー、ノースイースト・36ティ ーエイチ・プレイス 13808
(31) 優先権主張番号	62/264,822		
(32) 優先日	平成27年12月8日 (2015.12.8)		
(33) 優先権主張国・地域又は機関	米国 (US)		
(31) 優先権主張番号	62/281,917		
(32) 優先日	平成28年1月22日 (2016.1.22)		
(33) 優先権主張国・地域又は機関	米国 (US)		

最終頁に続く

(54) 【発明の名称】 二本鎖配列決定のための改善されたアダプター、方法、及び組成物

(57) 【特許請求の範囲】

【請求項1】

二本鎖標的核酸分子を配列決定する方法であって、

(1) アダプター核酸配列の対を、二本鎖標的核酸分子の少なくとも1つの終端に連結し、それにより第1の鎖アダプター - 標的核酸配列及び第2の鎖アダプター - 標的核酸配列を含む二本鎖アダプター - 標的核酸分子を形成するステップであって、各前記第1及び第2の鎖アダプター - 標的核酸配列が、単一分子識別子 (SMI) ドメインを有する、前記ステップと、

(2) 前記第2の鎖アダプター - 標的核酸配列から前記第1の鎖アダプター - 標的核酸配列を物理的に分離するステップと、

(3) 前記第1の鎖アダプター - 標的核酸配列を増幅し、それにより複数の第1の鎖アダプター - 標的核酸分子及び複数の第1鎖相補的分子を含む、増幅された生成物の第1の組を生成すること、及び

前記第2の鎖アダプター - 標的核酸配列を増幅し、それにより複数の第2の鎖アダプター - 標的核酸分子及び複数の第2鎖相補的分子を含む増幅された生成物の第2の組を生成するステップであって、

増幅された生成物の前記第2の組が、

SMIドメインによって増幅された生成物の前記第1の組に関連し、及び

増幅前の前記第2の鎖アダプター - 標的核酸配列から前記第1の鎖アダプター - 標的核酸配列の物理的分離によって増幅された生成物の前記第1の組から区別可能である、前記

ステップと、

(4) 増幅された生成物の前記第1の組を配列決定するステップと、

(5) 増幅された生成物の前記第2の組を配列決定するステップと、

(6) 増幅された生成物の前記第1の組から得られた少なくとも1つの配列を増幅された生成物の前記第2の組から得られた少なくとも1つの配列と比較し、二本鎖標的核酸分子のコンセンサス配列を作製するステップとを含む、前記方法。

【請求項2】

ステップ(2)が、前記二本鎖アダプター標的核酸分子を試料内に第1及び第2の鎖アダプター標的核酸配列を含む一本鎖分子に変性させ、前記二本の元々対合した第1及び第2の鎖アダプター標的核酸配列が同一の反応チャンバを共有する可能性が小さいように、複数の物理的に分離された反応チャンバ内に希釈することを含む、請求項1に記載の方法。

10

【請求項3】

前記物理的に分離された反応チャンバが、容器、管、ウェル、及び非連通液滴から選択される、請求項2に記載の方法。

【請求項4】

ステップ(3)が、タグ配列を有する少なくとも1つのプライマーの使用を通して各物理的に分離された反応チャンバのために実施され、

前記タグ配列は、タグ配列が鎖定義要素(SDE)ドメインとして動作するように各反応チャンバ内で実質的異なり、及び

20

分離された試料が、ステップ(4)及び(5)の前に組み換えられる、請求項2又は3に記載の方法。

【請求項5】

ステップ(3)が、前記標的核酸分子の配列の一部に特異的なプライマーの使用を通して第1及び第2の鎖アダプター-標的核酸配列を増幅することを含む、請求項1~4のいずれか1項記載の方法。

【請求項6】

アダプター核酸配列の対が、少なくとも部分的に相補的であるプライマー結合ドメインを有し、ステップ(3)が前記プライマー結合ドメインの特異的なプライマーの使用を通して前記第1及び第2の鎖アダプター-標的核酸配列を増幅することを含む、請求項1~5のいずれか1項記載の方法。

30

【請求項7】

前記二本鎖アダプター標的核酸分子が、非ヌクレオチド分子又は親和性パートナーによって結合されることができる親和性標識を含み、前記非ヌクレオチド分子又は親和性標識は、前記二本鎖アダプター-標的核酸分子の1つの鎖上に存在し、前記方法が、前記親和性パートナーを用いてステップ(2)において前記第2の鎖アダプター核酸配列から前記第1のアダプター核酸配列を分離し、前記非ヌクレオチド分子又は親和性標識を含む鎖を補足することをさらに含む、請求項1に記載の方法。

【請求項8】

前記非ヌクレオチド分子又は親和性標識が、コリシンE2、Im2、グルタチオン、グルタチオン-s-トランスフェラーゼ(GST)、ニッケル、ポリ-ヒスチジン、FLAG-タグ、myc-タグ、またはビオチンである含む群より選択される、請求項7に記載の方法。

40

【請求項9】

前記非ヌクレオチド分子又は親和性標識が、ビオチンであり、前記ビオチンが、ビオチン-16-アミノアリル-2'-デオキシウリジン-5'-トリホスフェート、ビオチン-16-アミノアリル-2'-デオキシシチジン-5'-トリホスフェート、ビオチン-16-アミノアリルシチジン-5'-トリホスフェート、N4-ビオチン-OBEA-2'-デオキシシチジン-5'-トリホスフェート、ビオチン-16-アミノアリルウリジ

50

ン - 5' - トリホスフェート、ピオチン - 16 - 7 - デアザ - 7 - アミノアリル - 2' - デオキシグアノシン - 5' - トリホスフェート、デスチオピオチン - 6 - アミノアリル - 2' - デオキシシチジン - 5' - トリホスフェート、5' - ピオチン - G - モノホスフェート、5' - ピオチン - A - モノホスフェート、5' - ピオチン - dG - モノホスフェート、または 5' - ピオチン - dA - モノホスフェートである、請求項 8 に記載の方法。

【請求項 10】

前記非ヌクレオチド分子又は親和性標識が、ピオチンであり、前記親和性パートナーが、基質に結合したストレプトアビジンである、請求項 8 に記載の方法。

【請求項 11】

前記固体基質が、固体表面、ビーズ、または物理的容器である、請求項 10 に記載の方法。 10

【請求項 12】

前記非ヌクレオチド分子又は親和性標識が、前記第 1 のアダプター - 標的核酸配列又は前記第 2 の鎖アダプター - アダプター核酸配列の終端に位置する、請求項 7 に記載の方法。

【請求項 13】

前記非ヌクレオチド分子又は親和性標識が、小分子、核酸、ペプチド、及び前記親和性パートナーによって結合されることができると特有に結合可能な部分から選択される、請求項 8 に記載の方法。

【請求項 14】

前記親和性標識が、核酸を含み、前記核酸が、DNA、RNA、又はそれらの組み合わせを含み、場合により、ペプチド核酸又はロックド核酸を含む、請求項 13 に記載の方法。 20

【請求項 15】

前記第 1 のアダプター - 標的核酸配列又は前記第 2 の鎖アダプター - 標的核酸配列のいずれかが、磁気特性、電荷特性、または不溶性特性を有する物理的基を含む、請求項 1 に記載の方法。

【請求項 16】

前記物理的基が、磁気特性を有し、ステップ (2) が、前記第 1 及び第 2 の鎖アダプター - 標的核酸配列に磁場を適用し、他のアダプター - 標的核酸配列から前記磁気特性を有する前記アダプター - 標的核酸配列を分離することを含む、請求項 15 に記載の方法。 30

【請求項 17】

前記物理的基が、電荷特性を有し、ステップ (2) が、前記第 1 及び第 2 の鎖アダプター - 標的核酸配列に電場を適用し、他のアダプター - 標的核酸配列から前記電荷特性を有する前記アダプター - 標的核酸配列を分離することを含む、請求項 15 に記載の方法。

【請求項 18】

前記物理的基が、不溶性特性を有し、ステップ (2) が、前記物理的基を含む前記アダプター - 標的核酸配列を沈殿させ、前記アダプター - 標的核酸配列を分離することを含む、請求項 15 に記載の方法。

【請求項 19】

増幅された生成物の第 1 の組を配列決定することが、第 1 の鎖コンセンサス配列を作製することを含み、増幅された生成物の第 2 の組を配列決定することが、第 2 の鎖コンセンサス配列を作製することを含み、前記比較するステップが、前記第 1 の鎖コンセンサス配列を前記第 2 の鎖コンセンサス配列と比較し、前記二本鎖標的核酸分子のコンセンサス配列を作製することを含み、前記第 1 のコンセンサス配列と前記第 2 のコンセンサス配列の間の差異が、人為産物とみなされ得る、請求項 1 ~ 18 のいずれか 1 項記載の方法。 40

【請求項 20】

ステップ (6) が、相補的ではない前記二本鎖標的核酸分子の両鎖における特定の位置を同定し除去することによって前記二本鎖標的核酸分子のエラーが訂正されたコンセンサス配列を提供することを含む、請求項 1 ~ 19 のいずれか 1 項記載の方法。 50

## 【請求項 2 1】

S M Iドメインが、少なくとも1つの縮重または半縮重核酸配列を含む、請求項 1 ~ 20のいずれか1項記載の方法。

## 【請求項 2 2】

S M Iドメインが、アダプター核酸配列及び二本鎖標的核酸分子のランダムまたは半ランダムに切断された末端に対応する配列の対に存在する核酸配列を含む、請求項 1 ~ 21のいずれか1項記載の方法。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

10

関連出願の相互参照

本出願は、2015年12月8日出願の米国特許仮出願第62/264,822号及び2016年1月22日出願の米国特許仮出願第62/281,917号に対する優先権及びそれらの利益を主張する。上述の出願の各々は、その全体が参照により本明細書に組み込まれる。

## 【0002】

配列表

本出願は、E F S - W e bを介してA S C I I形式で提出されている配列表を含み、その全体がこれにより参照により組み込まれる。2016年12月8日作製の該A S C I Iのコピーは、T W I N - 0 0 1 \_ S T 2 5 . t x tと称され、サイズは11,778バイトである。

20

## 【背景技術】

## 【0003】

二本鎖配列決定は、二本鎖DNAの二本の鎖を別個に増幅し、配列決定することによって、ハイスループットDNA配列決定の正確性の著しい改善を可能にし、故に、増幅及び配列決定エラーは典型的には二本の鎖のうち的一方のみに生じるため、それらを排除することができる。二本鎖配列決定は当初、DNA断片の末端に連結されたY形または「ループ」アダプター内に導入される非対称(すなわち、非相補的)PCRプライマー結合部位について記載された。アダプター内に存在する非対称プライマー結合部位自体は、二本のDNA鎖とは別個の生成物をもたらす、これが、二本のDNA鎖の各々からのエラー訂正

30

## 【0004】

したがって、非対称プライマー結合部位の使用を伴わない、二本鎖配列決定に対するアプローチへの満たされていない必要性が存在する。

40

## 【発明の概要】

## 【0005】

非対称プライマー結合部位の使用を必要としない、二本鎖配列決定に対する代替的かつ優れたアプローチが、本明細書に記載される。代わりに、配列決定されるアダプター内もしくはDNA分子内の別の場所の二本の鎖間のDNA配列内に少なくとも1つのヌクレオチドの差異を作製することによって、または他の方法(二本の鎖の物理的分離を可能にする、鎖の少なくとも一本への分子の結合など)で二本の鎖を差次的に標識することによって、二本の鎖間の非対称が導入され得る。

## 【0006】

第1の態様において、本発明は、第1のアダプター核酸配列及び第2のアダプター核酸

50

配列を含む、二本鎖標的核酸分子の配列決定において使用するためのアダプター核酸配列の対に関し、各アダプター核酸配列は、プライマー結合ドメイン、鎖定義要素 (SDE)、単一分子識別子 (SMI) ドメイン、及び連結ドメインを含む。第1のアダプター核酸配列のSDEは、第2のアダプター核酸配列のSDEと少なくとも部分的に非相補的であり得る。

【0007】

第1の態様の実施形態において、2つのアダプター配列は、少なくとも部分的にとともにアニールされている2つの別個のDNA分子を含み得る。第1のアダプター核酸配列及び第2のアダプター核酸配列は、リンカードメインを介して結合され得る。リンカードメインは、ヌクレオチドで構成され得る。リンカードメインは、1つ以上の修飾ヌクレオチドまたは非ヌクレオチド分子を含み得る。1つ以上の修飾ヌクレオチドまたは非ヌクレオチド分子は、脱塩基部位、ウラシル、テトラヒドロフラン、8-オキソ-7,8-ジヒドロ-2'-デオキシアデノシン (8-オキソ-A)、8-オキソ-7,8-ジヒドロ-2'-デオキシグアノシン (8-オキソ-G)、デオキシイノシン、5'-ニトロインドール、5-ヒドロキシメチル-2'-デオキシシチジン、イソ-シトシン、5'-メチル-イソシトシン、またはイソ-グアノシンであり得る。リンカードメインは、ループを形成し得る。第1のアダプター核酸配列のSDEは、第2のアダプター核酸配列のSDEと少なくとも部分的に非相補的であり得る。第1のアダプター核酸配列のプライマー結合ドメインは、第2のアダプター核酸配列のプライマー結合ドメインと少なくとも部分的に相補的であり得る。実施形態において、第1のアダプター核酸配列のプライマー結合ドメインは、第2のアダプター核酸配列のプライマー結合ドメインと相補的であり得る。第1のアダプター核酸配列のプライマー結合ドメインは、第2のアダプター核酸配列のプライマー結合ドメインと少なくとも部分的に非相補的であり得る。実施形態において、少なくとも1つのSMIドメインが、内在性SMIであり得、例えば、剪断点に関連付けられる (例えば、剪断点自体を使用して、剪断点の実際のマッピング位置 (例えば、染色体3、1、234、567位) を使用して、剪断点にすぐ隣接したDNA内の定義された数のヌクレオチドを使用して (例えば、剪断点から10ヌクレオチド、剪断点から7ヌクレオチド離れて開始する8ヌクレオチド、及び剪断点の後の「C」の最初の出現後に開始する6ヌクレオチド))。実施形態において、SMIドメインは、少なくとも1つの縮重または半縮重核酸を含む。実施形態において、SMIドメインは、非縮重であり得る。実施形態において、単一DNA分子を互いに区別することができるSMI配列を得るために、SMIドメインの配列は、連結されたDNAのランダムまたは半ランダムに剪断された末端に対応する配列と組み合わせ考慮され得る。第1のアダプター核酸配列のSMIドメインは、第2のアダプター核酸配列のSMIドメインと少なくとも部分的に相補的であり得る。第1のアダプター核酸配列のSMIドメインは、第2のアダプター核酸配列のSMIドメインと相補的であり得る。第1のアダプター核酸配列のSMIドメインは、第2のアダプター核酸配列のSMIドメインと少なくとも部分的に非相補的であり得る。実施形態において、各SMIドメインは、プライマー結合部位を含む。実施形態において、各SMIドメインは、その連結ドメインに対して遠位に位置し得る。第1のアダプター核酸配列のSMIドメインは、第2のアダプター核酸配列のSMIドメインと非相補的であり得る。実施形態において、各SMIドメインは、約1~約30個の縮重または半縮重核酸を含む。第1のアダプター核酸配列の連結ドメインは、第2のアダプター核酸配列の連結ドメインと少なくとも部分的に相補的であり得る。実施形態において、各連結ドメインは、二本鎖標的核酸配列の一方の鎖に連結されることが可能であり得る。実施形態において、連結ドメインの一方は、T-オーバーハング、A-オーバーハング、CG-オーバーハング、平滑末端、または別の連結可能な核酸配列を含む。実施形態において、両方の連結ドメインは、平滑末端を含む。実施形態において、連結ドメインの少なくとも一方は、修飾核酸を含む。修飾ヌクレオチドは、脱塩基部位、ウラシル、テトラヒドロフラン、8-オキソ-7,8-ジヒドロ-2'-デオキシアデノシン (8-オキソ-A)、8-オキソ-7,8-ジヒドロ-2'-デオキシグアノシン (8-オキソ-G)、デオキシイノシン、5'-ニト

10

20

30

40

50

ロインドール、5 - ヒドロキシメチル - 2' - デオキシシチジン、イソ - シトシン、5' - メチル - イソシトシン、またはイソ - グアノシンであり得る。実施形態において、連結ドメインの少なくとも一方は、脱リン酸化塩基を含む。実施形態において、連結ドメインの少なくとも一方は、脱ヒドロキシル化塩基を含む。実施形態において、連結ドメインの少なくとも一方は、それを連結不能にするように化学的に修飾されている。第1のアダプター核酸配列のSDEは、第2のアダプター核酸配列のSDEと少なくとも1つのヌクレオチドだけ異なる、かつ/またはそこで非相補的であり得る。実施形態において、少なくとも1つのヌクレオチドは、第1のアダプター核酸配列のSDEまたは第2のアダプター核酸のSDEのいずれかから酵素反応によって取り除かれ得る。酵素反応は、ポリメラーゼ、エンドヌクレアーゼ、グリコシラーゼ、またはリアーゼを含む。少なくとも1つのヌクレオチドは、修飾ヌクレオチドまたは標識を含むヌクレオチドであり得る。修飾ヌクレオチドまたは標識を含むヌクレオチドは、脱塩基部位、ウラシル、テトラヒドロフラン、8 - オキソ - 7, 8 - ジヒドロ - 2' - デオキシアデノシン (8 - オキソ - A)、8 - オキソ - 7, 8 - ジヒドロ - 2' - デオキシグアノシン (8 - オキソ - G)、デオキシイノシン、5' - ニトロインドール、5 - ヒドロキシメチル - 2' - デオキシシチジン、イソ - シトシン、5' - メチル - イソシトシン、またはイソ - グアノシンであり得る。第1のアダプター核酸配列のSDEは、ヘアピンループを形成することが可能であり得る自己相補的ドメインを含む。その連結ドメインに対して遠位の第1のアダプター核酸配列の末端は、その連結ドメインに対して遠位であり得る第2のアダプター核酸配列の末端に連結され、それによりループを形成し得る。ループは、制限酵素認識部位を含む。実施形態において、少なくとも第1のアダプター核酸配列は、第2のSDEを更に含む。第2のSDEは、第1のアダプター核酸配列の終端に位置し得る。第2のアダプター核酸配列は、第2のSDEを更に含む。第2のSDEは、第2のアダプター核酸配列の終端に位置し得る。第1のアダプター核酸配列の第2のSDEは、第2のアダプター核酸配列の第2のSDEと少なくとも部分的に非相補的であり得る。第1のアダプター核酸配列の第2のSDEは、第2のアダプター核酸配列の第2のSDEと少なくとも1つのヌクレオチドだけ異なる、かつ/またはそこで非相補的であり得る。実施形態において、少なくとも1つのヌクレオチドは、第1のアダプター核酸配列の第2のSDEまたは第2のアダプター核酸の第2のSDEのいずれかから酵素反応によって取り除かれ得る。酵素反応は、ポリメラーゼ、エンドヌクレアーゼ、グリコシラーゼ、またはリアーゼを含む。第1のアダプター核酸配列の第2のSDEは、第2のアダプター核酸配列の第2のSDEと非相補的であり得る。第1のアダプター核酸配列のSDEは、第2のアダプター核酸配列の第2のSDEに直接結合し得る。第1のアダプター核酸配列のプライマー結合ドメインは、第1のSDEに対して5'に位置し得る。第1のアダプター核酸配列の第1のSDEは、SMIDメインに対して5'に位置し得る。第1のアダプター核酸配列の第1のSDEは、SMIDメインに対して3'に位置し得る。第1のアダプター核酸配列の第1のSDEは、SMIDメインに対して5'に位置し得、プライマー結合ドメインに対して3'に位置し得る。第1のアダプター核酸配列の第1のSDEは、プライマー結合ドメインに対して3'に位置し得るSMIDメインに対して3'に位置し得る。第1のアダプター核酸配列のSMIDメインは、連結ドメインに対して5'に位置し得る。第1のアダプター核酸配列の3'終端は、連結ドメインを含む。第1のアダプター核酸配列は、5'から3'まで、プライマー結合ドメイン、第1のSDE、SMIDメイン、及び連結ドメインを含む。第1のアダプター核酸配列は、5'から3'まで、プライマー結合ドメイン、SMIDメイン、第1のSDE、及び連結ドメインを含む。実施形態において、第1のアダプター核酸配列または第2のアダプター核酸配列のいずれかは、修飾ヌクレオチドまたは非ヌクレオチド分子を含む。修飾ヌクレオチドまたは非ヌクレオチド分子は、コリシンE2、Im2、グルチチオン (グルチチオン (Glutithione))、グルタチオン - s - トランスフェラーゼ (GST)、ニッケル、ポリ - ヒスチジン、FLAG - タグ、myc - タグ、またはビオチンであり得る。ビオチンは、ビオチン - 16 - アミノアリル - 2' - デオキシウリジン - 5' - トリホスフェート、ビオチン - 16 - アミノアリル - 2' - デオキシシチジン

10

20

30

40

50

- 5' - トリホスフェート、ピオチン - 16 - アミノアリルシチジン - 5' - トリホスフェート、N4 - ピオチン - OBEA - 2' - デオキシシチジン - 5' - トリホスフェート、ピオチン - 16 - アミノアリルウリジン - 5' - トリホスフェート、ピオチン - 16 - 7 - デアザ - 7 - アミノアリル - 2' - デオキシグアノシン - 5' - トリホスフェート、デスチオピオチン - 6 - アミノアリル - 2' - デオキシシチジン - 5' - トリホスフェート、5' - ピオチン - G - モノホスフェート、5' - ピオチン - A - モノホスフェート、5' - ピオチン - dG - モノホスフェート、または5' - ピオチン - dA - モノホスフェートであり得る。ピオチンは、基質に結合したストレプトアビジンに結合することが可能であり得る。実施形態において、ピオチンが、基質に結合したストレプトアビジンに結合するとき、第1のアダプター核酸配列は、第2のアダプター核酸配列から分離することができる。実施形態において、第1のアダプター核酸配列または第2のアダプター核酸配列のいずれかは、小分子、核酸、ペプチド、及び親和性パートナーによって結合されることが可能であり得る特有に結合可能な部分から選択される、親和性標識を含む。実施形態において、親和性パートナーが、固体基質に結合し、親和性標識に結合するとき、親和性標識を含むアダプター核酸配列は、親和性標識を含まないアダプター核酸配列から分離されることができる。固体基質は、固体表面、ビーズ、または別の固定された構造であり得る。核酸は、DNA、RNA、またはこれらの組み合わせであってもよく、任意で、ペプチド - 核酸またはロックド核酸を含む。親和性標識は、第2のアダプター核酸配列内の対向するドメインと完全には相補的ではなくてもよい第1のアダプター核酸配列内の、アダプターの末端またはドメイン内に位置し得る。実施形態において、第1のアダプター核酸配列または第2のアダプター核酸配列のいずれかは、磁気特性、電荷特性、または不溶性特性を有する物理的基を含む。実施形態において、物理的基が磁気特性を有し、磁場が適用されるとき、物理的基を含むアダプター核酸配列は、物理的基を含まないアダプター核酸配列から分離される。実施形態において、物理的基が電荷特性を有し、電場が適用されるとき、物理的基を含むアダプター核酸配列は、物理的基を含まないアダプター核酸配列から分離される。実施形態において、物理的基が不溶性特性を有し、アダプター核酸配列の対が物理的基が不溶性である溶液中含有されるとき、物理的基を含むアダプター核酸配列は、溶液中に残る物理的基を含まないアダプター核酸配列から離れて沈殿する。物理的基は、第2のアダプター核酸配列内の対向するドメインと完全には相補的ではなくてもよい第1のアダプター核酸配列内の、アダプターの末端またはドメイン内に位置し得る。第2のアダプター核酸配列は、少なくとも1つのホスホロチオエート結合を含む。二本鎖標的核酸配列は、DNAまたはRNAであり得る。実施形態において、各アダプター核酸配列は、その末端の各々に連結ドメインを含む。第1のアダプター核酸配列または第2のアダプター核酸配列は、少なくとも部分的に一本鎖であり得る。第1のアダプター核酸配列または第2のアダプター核酸配列は、一本鎖であり得る。第1のアダプター核酸配列及び第2のアダプター核酸配列は、一本鎖であり得る。

【0008】

第2の態様において、本発明は、第1の態様のアダプター核酸配列の少なくとも1つの対及びアダプター核酸配列の第2の対を含む組成物に関し、アダプター核酸配列の第2の対の各鎖は、少なくともプライマー結合部位及び連結ドメインを含む。

【0009】

第2の態様は、第1の態様の少なくとも2対のアダプター核酸配列を含む組成物に更に関し、アダプター核酸配列の第1の対からの第1のアダプター核酸配列のSDEは、アダプター核酸配列の少なくとも第2の対からの第1のアダプター核酸配列のSDEとは異なる。

【0010】

第2の態様はまた、第1の態様の少なくとも2対のアダプター核酸分子を含む組成物にも関し、アダプター核酸分子の第1の対からの第1のアダプター核酸分子のSMIDメインは、アダプター核酸分子の少なくとも第2の対からの第1のアダプター核酸分子のSM

10

20

30

40

50

Iドメインとは異なる。

【0011】

第2の態様の実施形態において、本組成物は、アダプター核酸配列の第2の対の各鎖内にSMIドメインを更に含む。本組成物は、アダプター核酸配列の第2の対の各鎖内にプライマー結合部位を更に含む。一本鎖アダプター核酸分子の第1の対からの第1のアダプター核酸分子のSMIドメインは、一本鎖アダプター核酸分子の少なくとも第2の対からの第1の一本鎖アダプター核酸分子のSMIドメインと同一の長さであってもよい。一本鎖アダプター核酸分子第1の対からの第1のアダプター核酸分子のSMIドメインは、一本鎖アダプター核酸分子の少なくとも第2の対からの第1の一本鎖アダプター核酸分子のSMIドメインとは異なる長さを有してもよい。実施形態において、各SMIドメインは、SMI内またはそれに隣接する部位に1つ以上の固定された塩基を含む。実施形態において、第1の態様のアダプター核酸分子の第1の対を含む少なくとも第1の二本鎖複合化核酸は、二本鎖標的核酸分子の第1の終端に連結され、第1の態様のアダプター核酸分子の第2の対は、二本鎖標的核酸分子の第2の終端に連結される。アダプター核酸分子の第1の対は、アダプター核酸分子の第2の対とは異なってもよい。アダプター核酸分子の第1の対の第1の鎖アダプター-標的核酸分子は、第1のSMIドメインを含み、アダプター核酸分子の第2の対の第1の鎖アダプター-標的核酸分子は、第2のSMIドメインを含む。実施形態において、本組成物は、少なくとも第2の二本鎖複合化核酸を含む。

10

【0012】

第3の態様において、本発明は、第1のアダプター核酸配列及び第2のアダプター核酸配列を含む、二本鎖標的核酸分子の配列決定において使用するためのアダプター核酸配列の対に関する。第3の態様において、各アダプター核酸配列は、プライマー結合ドメイン及び単一分子識別子(SMI)ドメインを含む。

20

【0013】

第3の態様の実施形態において、第1のアダプター核酸配列または第2のアダプター核酸配列の少なくとも1つは、少なくとも1つの修飾ヌクレオチドを含むドメインを更に含む。第1のアダプター核酸配列及び第2のアダプター核酸配列は、少なくとも1つの修飾ヌクレオチドを含むドメインを更に含む。実施形態において、第1のアダプター核酸配列または第2のアダプター核酸配列の少なくとも1つは、連結ドメインを更に含む。第1のアダプター核酸配列及び第2のアダプター核酸配列は、連結ドメインを含み得る。少なくとも1つの修飾ヌクレオチドは、脱塩基部位、ウラシル、テトラヒドロフラン、8-オキソ-7,8-ジヒドロ-2'-デオキシアデノシン(8-オキソ-A)、8-オキソ-7,8-ジヒドロ-2'-デオキシグアノシン(8-オキソ-G)、デオキシイノシン、5'-ニトロインドール、5'-ヒドロキシメチル-2'-デオキシシチジン、イソ-シトシン、5'-メチル-イソシトシン、またはイソ-グアノシンであり得る。2つのアダプター配列は、少なくとも部分的にともにアニールされている2つの別個のDNA分子を含み得る。第1のアダプター核酸配列及び第2のアダプター核酸配列は、リンカードメインを介して結合され得る。リンカードメインは、ヌクレオチドで構成され得る。リンカードメインは、1つ以上の修飾ヌクレオチドまたは非ヌクレオチド分子を含み得る。実施形態において、少なくとも1つの修飾ヌクレオチドまたは非ヌクレオチド分子は、脱塩基部位、ウラシル、テトラヒドロフラン、8-オキソ-7,8-ジヒドロ-2'-デオキシアデノシン(8-オキソ-A)、8-オキソ-7,8-ジヒドロ-2'-デオキシグアノシン(8-オキソ-G)、デオキシイノシン、5'-ニトロインドール、5'-ヒドロキシメチル-2'-デオキシシチジン、イソ-シトシン、5'-メチル-イソシトシン、またはイソ-グアノシンであり得る。リンカードメインは、ループを形成し得る。第1のアダプター核酸配列のプライマー結合ドメインは、第2のアダプター核酸配列のプライマー結合ドメインと少なくとも部分的に相補的であり得る。第1のアダプター核酸配列のプライマー結合ドメインは、第2のアダプター核酸配列のプライマー結合ドメインと相補的であり得る。第1のアダプター核酸配列のプライマー結合ドメインは、第2のアダプター核酸配列のプライマー結合ドメインと非相補的であり得る。実施形態において、少なくとも1つのS

30

40

50

M Iドメインが、内在性S M Iであり、例えば、剪断点に関連付けられる（例えば、剪断点自体を使用して、剪断点の実際のマッピング位置（例えば、染色体3、1、234、567位）を使用して、剪断点にすぐ隣接したDNA内の定義された数のヌクレオチドを使用して（例えば、剪断点から10ヌクレオチド、剪断点から7ヌクレオチド離れて開始する8ヌクレオチド、及び剪断点の後の「C」の最初の出現後に開始する6ヌクレオチド））。S M Iドメインは、少なくとも1つの縮重または半縮重核酸を含む。S M Iドメインは、非縮重であり得る。単一DNA分子を互いに区別することができるS M I配列を得るために、S M Iドメインの配列は、連結されたDNAのランダムまたは半ランダムに剪断された末端に対応する配列と組み合わせて考慮され得る。第1のアダプター核酸配列のS M Iドメインは、第2のアダプター核酸配列のS M Iドメインと少なくとも部分的に相補的であり得る。第1のアダプター核酸配列のS M Iドメインは、第2のアダプター核酸配列のS M Iドメインと相補的であり得る。第1のアダプター核酸配列のS M Iドメインは、第2のアダプター核酸配列のS M Iドメインと少なくとも部分的に非相補的であり得る。第1のアダプター核酸配列のS M Iドメインは、第2のアダプター核酸配列のS M Iドメインと非相補的であり得る。実施形態において、各S M Iドメインは、約1～約30個の縮重または半縮重核酸を含む。第1のアダプター核酸配列の連結ドメインは、第2のアダプター核酸配列の連結ドメインと少なくとも部分的に相補的であり得る。実施形態において、各連結ドメインは、二本鎖標的核酸配列の一方の鎖に連結されることが可能であり得る。実施形態において、連結ドメインの一方は、T-オーバーハング、A-オーバーハング、CG-オーバーハング、平滑末端、または別の連結可能な核酸配列を含む。実施形態において、両方の連結ドメインは、平滑末端を含む。実施形態において、各S M Iドメインは、プライマー結合部位を含む。実施形態において、少なくとも第1のアダプター核酸配列は、S D Eを更に含む。S D Eは、第1のアダプター核酸配列の終端に位置し得る。第2のアダプター核酸配列は、S D Eを更に含む。S D Eは、第2のアダプター核酸配列の終端に位置し得る。第1のアダプター核酸配列のS D Eは、第2のアダプター核酸配列のS D Eと少なくとも部分的に非相補的であり得る。第1のアダプター核酸配列のS D Eは、第2のアダプター核酸配列のS D Eと少なくとも部分的に非相補的であり得る。第1のアダプター核酸配列のS D Eは、第2のアダプター核酸配列のS D Eに直接結合し得る。第1のアダプター核酸配列のS D Eは、第2のアダプター核酸配列のS D Eと少なくとも1つのヌクレオチドだけ異なる、かつ/またはそこで非相補的であり得る。少なくとも1つのヌクレオチドは、第1のアダプター核酸配列のS D Eまたは第2のアダプター核酸配列のS D Eのいずれかから酵素反応によって取り除かれ得る。酵素反応は、ポリメラーゼまたはエンドヌクレアーゼを含み得る。少なくとも1つのヌクレオチドは、修飾ヌクレオチドまたは標識を含むヌクレオチドであり得る。修飾ヌクレオチドまたは標識を含むヌクレオチドは、脱塩基部位、ウラシル、テトラヒドロフラン、8-オキソ-7,8-ジヒドロ-2'-デオキシアデノシン(8-オキソ-A)、8-オキソ-7,8-ジヒドロ-2'-デオキシグアノシン(8-オキソ-G)、デオキシイノシン、5'-ニトロインドール、5-ヒドロキシメチル-2'-デオキシシチジン、イソ-シトシン、5'-メチル-イソシトシン、またはイソ-グアノシンであり得る。第1のアダプター核酸配列のS D Eは、ヘアピンループを形成することが可能であり得る自己相補的ドメインを含み得る。その連結ドメインに対して遠位の第1のアダプター核酸配列の末端は、その連結ドメインに対して遠位である第2のアダプター核酸配列の末端に連結され、それによりループを形成し得る。ループは、制限酵素認識部位を含み得る。第1のアダプター核酸配列のプライマー結合ドメインは、S M Iドメインに対して5'に位置し得る。第1のアダプター核酸配列の少なくとも1つの修飾ヌクレオチドを含むドメインは、S M Iドメインに対して5'に位置し得る。第1のアダプター核酸配列の少なくとも1つの修飾ヌクレオチドを含むドメインは、S M Iドメインに対して3'に位置し得る。第1のアダプター核酸配列の少なくとも1つの修飾ヌクレオチドを含むドメインは、S M Iドメインに対して5'に位置し得る。第1のアダプター核酸配列の少なくとも1つの修飾ヌクレオチドを含むドメインは、プライマー結合ドメインに対して3

10

20

30

40

50



## 【 0 0 1 4 】

第4の態様において、本発明は、第3の態様の少なくとも2対のアダプター核酸分子を含む組成物に関し、アダプター核酸分子の第1の対からの第1のアダプター核酸分子のSMIドメインは、アダプター核酸分子の少なくとも第2の対からの第1のアダプター核酸分子のSMIドメインとは異なる。

## 【 0 0 1 5 】

第4の態様の実施形態において、一本鎖アダプター核酸分子の第1の対からの第1のアダプター核酸分子のSMIドメインは、一本鎖アダプター核酸分子の少なくとも第2の対からの第1の本鎖アダプター核酸分子のSMIドメインと同一の長さであってもよい。一本鎖アダプター核酸分子第1の対からの第1のアダプター核酸分子のSMIドメインは、一本鎖アダプター核酸分子の少なくとも第2の対からの第1の本鎖アダプター核酸分子のSMIドメインとは異なる長さを有してもよい。実施形態において、各SMIドメインは、SMI内またはそれに隣接する部位に1つ以上の固定された塩基を含む。

10

## 【 0 0 1 6 】

第5の態様において、本発明は、二本鎖標的核酸分子の第1の終端に連結された第3の態様のアダプター核酸分子の第1の対と、二本鎖標的核酸分子の第2の終端に連結された第3の態様のアダプター核酸分子の第2の対とを含む、少なくとも第1の二本鎖複合化核酸を含む組成物に関する。

## 【 0 0 1 7 】

第5の態様の実施形態において、アダプター核酸分子の第1の対は、アダプター核酸分子の第2の対とは異なってもよい。アダプター核酸分子の第1の対の第1の鎖アダプター-標的核酸分子は、第1のSMIドメインを含み得、アダプター核酸分子の第2の対の第1の鎖アダプター-標的核酸分子は、第2のSMIドメインを含み得る。アダプター核酸分子の第1の対の第1の鎖アダプター-標的核酸分子は、第1のSMIドメインを含み得、アダプター核酸分子の第2の対の第1の鎖アダプター-標的核酸分子は、第2のSMIドメインを含む。実施形態において、本組成物は、少なくとも第2の二本鎖複合化核酸を含む。

20

## 【 0 0 1 8 】

第6の態様において、本発明は、第1の態様のアダプター核酸分子の少なくとも1つの対と、第3の態様のアダプター核酸分子の少なくとも1つの対とを含む組成物に関する。

30

## 【 0 0 1 9 】

第7の態様において、本発明は、二本鎖標的核酸分子の第1の終端に連結された第1の態様のアダプター核酸分子の第1の対と、二本鎖標的核酸分子の第2の終端に連結された第3の態様のアダプター核酸分子の第2の対とを含む、少なくとも第1の二本鎖複合化核酸を含む組成物に関する。

## 【 0 0 2 0 】

第8の態様において、本発明は、二本鎖標的核酸を配列決定する方法であって、(1)第1の態様のアダプター核酸配列の対を、二本鎖標的核酸分子の少なくとも1つの終端に連結し、それにより第1の鎖アダプター-標的核酸配列及び第2の鎖アダプター-標的核酸配列を含む二本鎖核酸分子を形成するステップと、(2)第1の鎖アダプター-標的核酸配列を増幅し、それにより複数の第1の鎖アダプター-標的核酸配列及び複数のその相補的分子を含む、増幅された生成物の第1の組を生成するステップと、(3)第2の鎖アダプター-標的核酸配列を増幅し、それにより複数の第2の鎖アダプター-標的核酸配列及び複数のその相補的分子を含む増幅された生成物の第2の組を生成するステップであって、増幅された生成物の第2の組が、増幅された生成物の第1の組から区別可能であり得る、生成するステップと、(4)増幅された生成物の第1の組を配列決定するステップと、(5)増幅された生成物の第2の組を配列決定するステップとを含む、方法に関する。

40

## 【 0 0 2 1 】

第8の態様の実施形態において、少なくとも1つの終端は、2つの終端であり得る。増幅は、PCRによって、多置換増幅によって、または等温増幅によって実行され得る。二

50

本鎖標的核酸配列の第1の終端に連結されたアダプター核酸配列の対は、二本鎖標的核酸配列の第2の終端に連結されたアダプター核酸配列の対と同一の構造を有する。第8の態様の実施形態において、第1の鎖アダプター-標的核酸配列は、5'から3'の順で、(a)第1のアダプター核酸配列、(b)二本鎖標的核酸の第1の鎖、及び(c)第2のアダプター核酸配列を含む。第8の態様の実施形態において、第2の鎖アダプター-標的核酸配列は、3'から5'の順で、(a)第1のアダプター核酸配列、(b)二本鎖標的核酸の第2の鎖、及び(c)第2のアダプター核酸配列を含む。二本鎖標的核酸配列の第1の終端に連結されたアダプター核酸配列の対は、二本鎖標的核酸配列の第2の終端に連結されたアダプター核酸配列の対とは異なってもよい。二本鎖標的核酸配列の第1の終端に連結されたアダプター核酸配列の対は、第1のSMIDメインを有し、二本鎖標的核酸配列の第2の終端に連結されたアダプター核酸配列の対は、第2のSMIDメインを有し、第1のSMIDメインは、第2のSMIDメインとは異なってもよい。第8の態様の実施形態において、第1の鎖アダプター-標的核酸配列は、5'から3'の順で、(a)第1のSDEを含む第1のアダプター核酸配列、(b)第1のSMIDメイン、(c)二本鎖標的核酸の第1の鎖、及び(d)第2のアダプター核酸配列を含み得る。第8の態様の実施形態において、第2の鎖アダプター-標的核酸配列は、5'から3'の順で、(a)第1のSDEを含む第1のアダプター核酸配列、(b)第2のSMIDメイン、(c)二本鎖標的核酸の第2の鎖、及び(d)第2のアダプター核酸配列を含み得る。実施形態において、増幅された生成物の第1の組のコンセンサス配列は、増幅された生成物の第2の組のコンセンサス配列と比較され得、2つのコンセンサス配列間の差異は、人為産物であると見なされ得る。

10

20

**【0022】**

第9の態様において、本発明は、二本鎖標的核酸を配列決定する方法であって、(1)第3の態様のアダプター核酸配列の対を、二本鎖標的核酸分子の少なくとも1つの終端に連結し、それにより第1の鎖アダプター-標的核酸配列及び第2の鎖アダプター-標的核酸配列を含む二本鎖核酸分子を形成するステップと、(2)第1の鎖アダプター-標的核酸分子を増幅し、それにより複数の第1の鎖アダプター-標的核酸分子及び複数のその相補的分子を含む、増幅された生成物の第1の組を生成するステップと、(3)第2の鎖アダプター-標的核酸分子を増幅し、それにより複数の第2の鎖アダプター-標的核酸分子及び複数のその相補的分子を含む、増幅された生成物の第2の組を生成するステップと、(4)増幅された生成物の第1の組を配列決定し、それにより増幅された生成物の第1の組のコンセンサス配列を得るステップと、(5)増幅された生成物の第2の組を配列決定し、それにより増幅された生成物の第2の組のコンセンサス配列を得るステップとを含む、方法に関する。

30

**【0023】**

第9の態様の実施形態において、増幅された生成物の第2の組は、増幅された生成物の第1の組から区別可能であり得る。増幅は、PCRによって、多置換増幅によって、または等温増幅によって実行され得る。第9の態様の実施形態において、本方法は、ステップ(1)の後に、二本鎖核酸分子を、少なくとも1つの修飾ヌクレオチドを別の化学構造に変化させる少なくとも1つの酵素(例えば、グリコシラーゼ)と接触させるステップを更に含む。二本鎖標的核酸分子の第1の終端に連結されたアダプター核酸配列の対は、二本鎖標的核酸分子の第2の終端に連結されたアダプター核酸配列の対と同一であってもよい。二本鎖標的核酸分子の第1の終端に連結されたアダプター核酸配列の対は、二本鎖標的核酸分子の第2の終端に連結されたアダプター核酸配列の対とは異なってもよい。実施形態において、アダプター核酸配列の対は、二本鎖標的核酸分子の第1の終端に連結され得、標的DNA分子のDNA配列の一部に対応するプライマーは、DNA分子を増幅するために利用され得る。第9の態様の実施形態において、第1の鎖アダプター-標的核酸配列は、5'から3'の順で、(a)少なくとも1つの修飾ヌクレオチドまたは少なくとも1つの脱塩基部位を含む第1のアダプター核酸配列、(b)二本鎖標的核酸の第1の鎖、及び(c)第2のアダプター核酸配列を含む。第9の態様の実施形態において、第2の鎖

40

50

アダプター - 標的核酸配列は、3' から 5' の順で、( a ) 第 1 のアダプター核酸配列、( b ) 二本鎖標的核酸の第 2 の鎖、及び ( c ) 第 2 のアダプター核酸配列を含む。二本鎖標的核酸分子の第 1 の終端に連結されたアダプター核酸配列の対は、二本鎖標的核酸分子の第 2 の終端に連結されたアダプター核酸配列の対とは異なってもよい。二本鎖標的核酸分子の第 1 の終端に連結されたアダプター核酸配列の対は、第 1 の S M I ドメインを有し、二本鎖標的核酸配列の第 2 の終端に連結されたアダプター核酸配列の対は、第 2 の S M I ドメインを有し、第 1 の S M I ドメインは、第 2 の S M I ドメインとは異なってもよい。第 9 の態様の実施形態において、第 1 の鎖アダプター - 標的核酸配列は、5' から 3' の順で、( a ) 少なくとも 1 つの修飾ヌクレオチドまたは少なくとも 1 つの脱塩基部位、及び第 1 の S M I ドメインを含む第 1 のアダプター核酸配列、( b ) 二本鎖標的核酸の第 1 の鎖、ならびに ( c ) 第 2 の S M I ドメインを含む第 2 のアダプター核酸配列を含む。実施形態において、少なくとも 1 つの修飾ヌクレオチドが、8 - オキソ - G であり得るとき、第 2 のアダプター核酸配列は、8 - オキソ - G に対応する位置にシトシンを含む。第 9 の態様の実施形態において、第 2 の鎖アダプター - 標的核酸配列は、3' から 5' の順で、( a ) 第 1 の S M I ドメインを含む第 1 のアダプター核酸配列、( b ) 二本鎖標的核酸の第 2 の鎖、及び ( c ) 第 2 の S M I ドメインを含む第 2 のアダプター核酸配列を含む。実施形態において、少なくとも 1 つの修飾ヌクレオチドは、8 - オキソ - G であり得、第 2 のアダプター核酸配列は、8 - オキソ - G に対応する位置にシチジンを含む。実施形態において、ステップ ( 2 ) またはステップ ( 3 ) の増幅中に、少なくとも 1 つの脱塩基部位が、対応する増幅された生成物中のチミジンへの増幅時に変換され得、S D E の導入をもたらし。第 9 の態様の実施形態において、ステップ ( 2 ) またはステップ ( 3 ) の増幅中に、少なくとも 1 つの修飾ヌクレオチド部位が、対応する増幅された生成物中のアデノシンをコードする。

#### 【 0 0 2 4 】

第 1 0 の態様において、本発明は、区別可能な増幅生成物が、個々の D N A 分子の二本の鎖の各々から得られ得、増幅された生成物の第 1 の組のコンセンサス配列が、増幅された生成物の第 2 の組のコンセンサス配列と比較され得、2 つのコンセンサス配列間の差異が、人為産物と見なされ得る、方法に関する。

#### 【 0 0 2 5 】

第 1 0 の態様の実施形態において、増幅された生成物は、同一の S M I 配列を共有しているために、同一の初期 D N A 分子から生じていることが決定され得る。実施形態において、増幅された生成物は、S M I アダプターライブラリ合成の時点で、かつそれと組み合わせ生成されたデータベースに基づいて、互いに対応することが既知であり得る、別個の S M I 配列を担持しているために、同一の初期 D N A 分子から生じていることが決定され得る。実施形態において、増幅された生成物は、S D E によって導入された配列差異の少なくとも 1 つのヌクレオチドを介して、同一の初期二本鎖 D N A 配列の別個の鎖から生じていることが決定され得る。

#### 【 0 0 2 6 】

第 1 1 の態様において、本発明は、区別可能な増幅生成物が、個々の D N A 分子の二本の鎖の各々から得られ得、単一 D N A 分子の二本の初期 D N A 鎖のうち的一本に対応する増幅された生成物から得られる配列が、二本の初期 D N A 鎖のうちの本目に対応する増幅された生成物と比較され、2 つの配列間の差異が、人為産物と見なされ得る、方法に関する。

#### 【 0 0 2 7 】

第 1 2 の態様において、本発明は、単一 D N A 分子の二本の初期 D N A 鎖のうちの本目に対応する増幅された生成物から得られる配列が、二本の初期 D N A 鎖のうちの本目に対応する増幅された生成物と比較され、2 つの配列間にいかなる差異も特定されないとき、区別不可能な増幅生成物が、個々の D N A 分子の二本の鎖から得られ得る、方法に関する。

#### 【 0 0 2 8 】

10

20

30

40

50

第12の態様の実施形態において、増幅された生成物は、SMIアダプターライブラリ合成の時点で、かつそれと組み合わせて生成されたデータベースに基づいて、同一のSMI配列を共有しているために、同一の初期二本鎖DNA分子から生じていることが決定され得る。実施形態において、増幅された生成物は、SDEによって導入された配列差異の少なくとも1つのヌクレオチドを介して、同一の初期二本鎖DNA配列の別個の鎖から生じていることが決定され得る。実施形態において、本方法は、DNA二本鎖のそれらの構成要素の一本鎖への熱融解または化学融解後に、単一分子希釈のステップを更に含む。二本の元々対合した鎖が同一の容器を共有する可能性が小さく有り得るように、一本鎖は、複数の物理的に分離された反応チャンバ内へと希釈され得る。物理的に分離された反応チャンバは、容器、管、ウェル、及び少なくとも1対の非連通液滴から選択され得る。実施形態において、PCR増幅は、好ましくは異なるタグ配列を担持する各チャンバのためのプライマーを使用して、物理的に分離された各反応チャンバに実行され得る。実施形態において、各タグ配列は、SDEとして動作する。実施形態において、同一の初期DNAの二本の鎖に対応する一連の対合した配列は、互いに比較され得、一連の生成物からの少なくとも1つの配列が、初期DNA分子の正確な配列を表す可能性が最も高いものとして選択され得る。初期DNA分子の正確な配列を表す可能性が最も高いものとして選択される生成物は、少なくとも一部には、二本のDNA鎖から得られる生成物間に最小数のミスマッチを有するために選択され得る。初期DNA分子の正確な配列を表す可能性が最も高いものとして選択される生成物は、少なくとも一部には、参照配列と比較して最小数のミスマッチを有するために選択され得る。

10

20

## 【0029】

第13の態様において、本発明は、少なくとも2対のアダプター核酸配列を含む組成物に関し、アダプター核酸配列の第1の対は、プライマー結合ドメイン、鎖定義要素(SDE)、及び連結ドメインを含み、アダプター核酸配列の第2の対は、プライマー結合ドメイン、単一分子識別子(SMI)ドメイン、及び連結ドメインを含む。

## 【0030】

第14の態様において、本発明は、(1)プライマー結合ドメインと、SDEとを含む、アダプター核酸配列の第1の対と、(2)二本鎖標的核酸と、(3)プライマー結合ドメインと、単一分子識別子(SMI)ドメインとを含む、アダプター核酸配列の第2の対とを含む、二本鎖複合化核酸に関し、アダプター核酸分子の第1の対は、二本鎖標的核酸分子の第1の終端に連結され得、アダプター核酸分子の第2の対は、二本鎖標的核酸分子の第2の終端に連結され得る。第14の態様の実施形態において、アダプター核酸配列の第1の対及び/またはアダプター核酸分子の第2の対は、連結ドメインを更に含む得る。

30

## 【0031】

第15の態様において、本発明は、第1のアダプター核酸配列及び第2のアダプター核酸配列を含む、二本鎖標的核酸分子の配列決定において使用するためのアダプター核酸配列の対に関し、各アダプター核酸配列は、プライマー結合ドメイン、SDE、連結ドメインを含み、第1のアダプター核酸配列のSDEは、第2のアダプター核酸配列のSDEと少なくとも部分的に非相補的であり得る。

## 【0032】

第16の態様において、本発明は、二本鎖標的核酸分子の第1の終端に連結され、二本鎖標的核酸分子の第2の第2の終端に連結された、第1の態様のアダプター核酸分子の対を含む、二本鎖環状核酸に関する。

40

## 【0033】

第17の態様において、本発明は、二本鎖標的核酸分子の第1の終端に連結され、二本鎖標的核酸分子の第2の第2の終端に連結された、第3の態様のアダプター核酸分子の対を含む、二本鎖環状核酸に関する。

## 【0034】

第18の態様において、本発明は、二本鎖標的核酸分子の第1の終端に連結された第1の態様のアダプター核酸分子の対と、二本鎖標的核酸分子の第2の終端に連結されたブラ

50

イマー結合ドメインのアニールされた対とを含む、二本鎖環状核酸に関し、プライマー結合ドメインのアニールされた対は、アダプター核酸分子の対に連結され得る。

【0035】

第19の態様において、本発明は、二本鎖標的核酸分子の第1の終端に連結された第3の態様のアダプター核酸分子の対と、二本鎖標的核酸分子の第2の終端に連結されたプライマー結合ドメインのアニールされた対とを含む、二本鎖環状核酸に関し、プライマー結合ドメインのアニールされた対は、アダプター核酸分子の対に連結され得る。

【0036】

第20の態様において、本発明は、(1)プライマー結合ドメインと、鎖定義要素(SDE)と、単一分子識別子(SMI)ドメインとを含む、アダプター核酸配列の対と、(2)二本鎖標的核酸と、(3)アニールされた対プライマー結合ドメインとを含む、二本鎖複合化核酸に関し、アダプター核酸配列の対は、二本鎖標的核酸分子の第1の終端に連結され得、アニールされた対プライマー結合ドメインは、二本鎖標的核酸分子の第2の終端に連結され得る。第20の態様の実施形態において、アダプター核酸配列の対及び/またはアニールされた対プライマー結合ドメインは、連結ドメインを更に含む。

【0037】

二本鎖配列決定は、WO2013/142389A1及びSchmitt et al, PNAS 2012に更に記載され、これらの各々は、その全体が参照により本明細書に組み込まれる。

【0038】

上記の態様及び実施形態のいずれも、以下の本発明の具体的、非限定的な実施例/実施形態を含む、本明細書において要約、図面、及び/または発明を実施するための形態に開示されるいかなる他の態様または実施形態とも組み合わせることができる。

【0039】

本発明の他の特徴、利点、修正は、図面、発明を実施するための形態、及び特許請求の範囲から明らかとなるだろう。上述の記述は、説明が意図され、本開示の範囲の限定は意図されない。

【図面の簡単な説明】

【0040】

上記及び更なる特徴は、添付の図面と組み合わせると考慮されるとき、以下の発明を実施するための形態からより明確に理解されるだろう。

【0041】

【図1A-1I】元来記載されているY形アダプターを使用する、二本鎖配列決定を説明する。一例示的なY形アダプター(図1A)、そのようなアダプターに連結された二本鎖DNA分子(図1B)、それに由来するPCR生成物(図1C及び図1D)、ならびにこのようにして生成された配列決定読み取り(図1E~図1I)が示される。

【図2A-2K】非相補的「バブル」アダプターを使用する、本発明の二本鎖配列決定を説明する。一例示的な「バブル」アダプター(図2A及び図2H~図2K)、図2Aのアダプターに連結された二本鎖DNA分子(図2B)、それに由来するPCR生成物(図2C及び図2D)、ならびにこのようにして生成された配列決定読み取り(図2E~図2G)が示される。

【図3A-3G】共同で分子識別子及び非対称導入鎖定義要素(SDE)としての役割を果たす非相補的「バブル」形単一分子識別子(SMI)を有するアダプターを使用する、本発明の二本鎖配列決定を説明する。一例示的な「バブル」アダプター(図3A)、図3Aのアダプターに連結された二本鎖DNA分子(図3B)、それに由来するPCR生成物(図3C及び図3D)、ならびにこのようにして生成された配列決定読み取り(図3E及び図3F)が示される。図3Gは、特定のSMI配列及びそれらの対応する非相補的パートナーによってグルーブ化される図3E及び図3Fの配列決定読み取りを示す。

【図4A-4H】最初是对合した鎖DNAを形成するが、その後、後続する生化学反応後にDNAミスマッチにされる、ヌクレオチドまたはヌクレオチド類似体を有するアダプタ

10

20

30

40

50

ーを使用する、本発明の二本鎖配列決定を説明する。8 - オキソ - 7', 8 - ジヒドロ - 2' - デオキシグアノシン (8 - オキソ - G) を含む一例示的なアダプター (図 4 A)、図 4 A のアダプターに連結された二本鎖 DNA 分子 (図 4 B) が示され、図 4 C は、8 - オキソ - G 塩基を置き換える脱塩基部位を作製し、それによりアダプター内にミスマッチを作製する、グリコシラーゼでの処置の後の図 4 B の二本鎖 DNA 分子、それに由来する PCR 生成物 (図 4 D 及び図 4 E)、ならびにこのようにして生成された配列決定読み取り (図 4 F ~ 図 4 H) を示す。

【図 5 A - 5 H】DNA 分子の反対の末端上に異なるプライマー部位を導入するための二本鎖配列決定アダプター設計の組み合わせを使用する、本発明の二本鎖配列決定を説明する。一例示的な二本鎖配列決定アダプター (図 5 A) 及び「標準」アダプター (図 5 B)、図 5 A 及び図 5 B のアダプターが DNA 分子に連結されるとき、3 種類の二本鎖 DNA 分子が生成され (図 5 C ~ 図 5 E)、それに由来する PCR 生成物 (図 5 F 及び図 5 G)、ならびにこのようにして生成された配列決定読み取り (図 5 H) が示される。

【図 6 A - 6 I】非ペアードエンドプラットフォーム上での 2 つの読み取りを可能にする二本鎖配列決定アダプター設計の組み合わせを使用する、本発明の二本鎖配列決定を説明する。「標準」アダプター (図 6 A) 及び一例示的な二本鎖配列決定アダプター (図 6 B)、図 6 A 及び図 6 B のアダプターが DNA 分子に連結されるときに生成される好ましい二本鎖 DNA 分子 (図 6 C)、それに由来する PCR 生成物 (図 6 D 及び図 6 E、「上部」鎖 (図 6 F) 及び「下部」鎖 (図 6 G) に由来する配列決定鑄型鎖の配置、ならびにこのようにして生成された配列決定読み取り (図 6 H 及び図 6 I) が示される。

【図 7 A - 7 I】非ペアードエンドプラットフォーム上での 2 つの読み取りを可能にする二本鎖配列決定アダプター設計の組み合わせを使用する、本発明の二本鎖配列決定を説明する。縮重または半縮重 SMI 配列を更に含むアダプター (図 7 A) 及び一例示的な二本鎖配列決定アダプター (図 7 B)、図 7 A 及び図 7 B のアダプターが DNA 分子に連結されるときに生成される好ましい二本鎖 DNA 分子 (図 7 C)、それに由来する PCR 生成物 (図 7 D 及び図 7 E、「上部」鎖 (図 7 F) 及び「下部」鎖 (図 7 G) から得られる配列決定鑄型鎖の配置、ならびにこのようにして生成された配列決定読み取り (図 7 H 及び図 7 I) が示される。

【図 8 A - 8 J】非対称 SMI を有する Y 形二本鎖配列決定アダプターを使用する、本発明の二本鎖配列決定を説明する。一例示的な二本鎖配列決定アダプター (図 8 A)、図 8 A のアダプターが DNA 分子に連結されるときに生成される二本鎖 DNA 分子 (図 8 B)、それに由来する PCR 生成物 (図 8 C 及び図 8 D)、ならびにこのようにして生成された配列決定読み取り (図 8 E 及び図 8 F) が示される。図 8 G は、特定の SMI 配列及びそれらの対応する非相補的パートナーによってグループ化される図 8 E 及び図 8 F 配列決定読み取りを示す。図 8 H ~ 図 8 J は、この実施形態において有用である代替的なアダプター設計を示す。

【図 9 A - 9 G】遊離一本鎖尾部領域内に位置する非対称 SMI を有する Y 形またはループ形二本鎖配列決定アダプターを使用する、本発明の二本鎖配列決定を説明する。一例示的な二本鎖配列決定アダプター (図 9 A)、図 9 A のアダプターが DNA 分子に連結されるときに生成される好ましい二本鎖 DNA 分子 (図 9 B)、それに由来する PCR 生成物 (図 9 C 及び図 9 D) が示され、配列決定プライマー部位及びインデックスプライマー部位の配向は、図 9 E 及び図 9 F に示される。図 9 G は、図 9 E 及び図 9 F に示される方法で得られるグループ化配列決定読み取りを示す。

【図 10 A - 10 E】二本鎖配列決定に必要な全ての要素が 2 つの対合したアダプター内ではなく単一分子内に含まれる、本発明の二本鎖配列決定を説明する。図 10 A は、二本鎖 DNA 分子の連結前のそのような構成を示し、図 10 B は、二本鎖 DNA 分子の連結後の図 10 A の構成を示す。図 10 C ~ 図 10 E は、この実施形態のいくつかの代替形態を示す。

【図 11 A - 11 D】非対称化学標識及び鎖単離を通じた二本鎖配列決定を説明する。化学タグ (ここではビオチン) を有する一例示的な二本鎖配列決定アダプター (図 11 A)

10

20

30

40

50

及び第2のアダプター(図11B)、図11Aのアダプター及び図11BのアダプターがDNA分子に連結されるときに生成される好ましい二本鎖DNA分子(図11C)、ならびに化学タグを含む鎖が他方の鎖から分離され、各々が独立して増幅され、配列決定される、この方法の更なるステップ(図11D)が示される。

【図12A-12M】ニック翻訳によってSDEが導入される、本発明の二本鎖配列決定を説明する。図12A~図12Dは、ニック翻訳後にSDEが失われる、アダプター設計を示す。この実施形態において有用であるIon Torrent(商標)適合アダプター(図12E及び図12F)、図12E及び図12FのアダプターがDNA分子に連結されるときに生成される好ましい二本鎖DNA分子(図12G)、末端ヌクレオチドの誤組み込み(図12H)、それに由来し、ミスマッチを示す延長生成物(図12I)、図12Iの分子に由来するPCR生成物(図12J及び図12K)、ならびにこのようにして生成された配列決定読み取り(図12L及び図12M)が示される。

10

【図13A-13G】ニック翻訳後にSDEが導入される、本発明の二本鎖配列決定を説明する。脱リン酸化5'末端を含む二本鎖配列決定アダプター(図13A)、図13AのアダプターがDNA分子に連結されるときに生成される二本鎖DNA分子(図13B)、鎖置換合成が生じた後の構造(図13C)、いかなるミスマッチも示さない図13Cの構造の延長生成物(図13D)、ウラシルDNAグリコシラーゼ及び適切なAPエンドヌクレアーゼでの処置後にギャップを含む構造(図13E)、ギャップがミスマッチヌクレオチドで充填され、連結閉鎖された後の図13Eの構造(図13F)、ならびにこのようにして生成された配列決定読み取り(図13G)が示される。

20

【図14A-14I】配列決定されるDNA分子内に、ポリメラーゼ延長によってミスマッチが導入される、本発明の二本鎖配列決定を説明する。配列決定される二本鎖DNA分子(図14A)、5'オーバーハングを残すエンドヌクレアーゼで処理されている図14Aの二本鎖DNA分子(図14B)が示され、図14Bの部分的二本鎖DNA分子は、2つのミスマッチを導入するように処理され(図14C)、今では各ミスマッチに「バブル」を含む図14Cの構造の延長生成物(図14D)、アダプターの対が図14Eに示され、図14Fの構造は、図14Eのアダプターが図14DのDNA分子に連結されるときに生成され、図14Fの分子に由来するPCR生成物(図14G及び図14H)、ならびにこのようにして生成された配列決定読み取り(図14I)が示される。

【発明を実施するための形態】

30

【0042】

二本鎖配列決定は当初、二本のDNA鎖の別個の増幅のための非対称プライマー結合部位の使用について記載された。非対称プライマー結合部位の使用を必要としない、二本鎖配列決定に対する代替的かつ優れたアプローチが、本明細書に記載される。代わりに、配列決定されるアダプター内もしくはDNA分子内の別の場所の二本の鎖間のDNA配列内への少なくとも1つのヌクレオチドの差異(例えば、ミスマッチ、追加のヌクレオチド、及びヌクレオチドの取り除き)の作製、少なくとも1つのヌクレオチドの修飾ヌクレオチド(例えば、塩基を欠如するか、もしくは非定型塩基を有するヌクレオチド)での置き換え、ならびに/または二本の鎖を物理的に分離することができる少なくとも1つの標識ヌクレオチド(例えば、ピオチン化ヌクレオチド)の包含によって、二本の鎖間の非対称が導入され得る。表1は、本発明に開示される二本鎖配列決定のためのアダプターを組み立てるための、例示的な選択肢を説明する。

40

## 【表 1】

表 1 :

鎖定義要素 (SDE)	単一分子識別子 (SMI)
少なくとも1つのヌクレオチドのミスマッチは、アダプター (すなわち、「バブルアダプター」) 内の内部に存在する	バブルに隣接する SMI バブル自体内の SMI SMI は、第2のアダプター「内在性 SMI」 (剪断点) 内にある
「マッチした」配列は、後続するステップによってミスマッチに変換される -- ミスマッチは、酵素処理 (例: 8-オキソ-G) によって作製される -- ミスマッチは、ポリメラーゼによって導入される (例: ニック翻訳)	バブルに隣接する SMI バブル自体内の SMI SMI は、第2のアダプター「内在性 SMI」 (剪断点) 内にある
異なる配列が、アダプター内に存在する尾部 (すなわち、「Yアダプター」)	アダプター尾部に隣接する SMI 尾部自体内の SMI SMI は、第2のアダプター「内在性 SMI」 (剪断点) 内にある
二本の鎖は、(例えば、他方ではなく、一方の鎖上のビオチンで) 物理的に分離される	アダプター自体内の SMI SMI は、第2のアダプター「内在性 SMI」 (剪断点) 内にある
二本の鎖は、異なる長さである -- 一方のアダプター鎖内の内部「ループ」 -- 追加のヌクレオチドが、他方ではなく、一方の鎖に添加される	アダプター自体内の SMI SMI は、第2のアダプター「内在性 SMI」 (剪断点) 内にある

注記 :

(i) これらのアダプター設計の全てには、追加で任意の要素が追加されてもよい (例えば、二本のアダプター鎖がともに結合され、様々な構成で PCR プライマー部位を利用する)

(ii) SMI が使用される時は常に、それは、ランダム/縮重、半ランダム/半縮重、または事前定義されたものであってもよい。また、SMI が二本の鎖を含む場合、二本の鎖は、相補的、非相補的、または部分的に相補的のいずれであってもよい。

(iii) 少なくとも1つの SDE 及び少なくとも1つの SMI を含有する、完全に適応した分子複合体は、結合前にアダプター及び/または連結される DNA 内に存在しても、連結後に生成されても、これらの組み合わせであってもよい。

## 【0043】

本明細書に記載される、二本鎖配列決定のためのアダプター設計及びアプローチは、相補的 SMI 配列を有する Y-アダプターの使用には依存しない。

## 【0044】

いくつかの設計は、シングルエンド配列決定に直接適用することができる。本明細書に開示されるアプローチは、2つの一般的な特徴を共有し、つまり、(1) 個々の二本鎖 DNA 分子の各単鎖の半分は、究極的には二本の鎖の各々から得られる配列が同一の DNA 二本鎖に関連すると認識され得るような方法で標識され、かつ (2) 個々の二本鎖 DNA 分子の各単鎖は、究極的には二本の鎖の各々から得られる配列が反対の鎖から得られるものとは別個のものであると認識され得るような方法で標識される。これらのそれぞれの機能を果たす分子特徴は、本明細書では単一分子識別子 (SMI) 及び鎖定義要素 (SDE) と呼ばれる。

## 【0045】

これが、第1に開示される、異なるバージョンの内部非相補的「バブル」配列を介した鎖定義非対称の導入である。1つのそのような実施形態は、増幅プライマー部位内に位置

10

20

30

40

50

しない非相補的「バブル」配列の導入を伴い、その後、「バブル」の二本の鎖から得られる別個の配列は、二本の鎖の別個の標識をもたらすだろう。

【0046】

SDEとしての修飾DNA塩基の使用を通して、鎖定義非対称が、適応したDNA分子内にいかに同様に導入され得るかが、本明細書に開示される。実施例において、非対称は最初は、相補的配列をもたらす1つ以上のヌクレオチド類似体を含めることによって導入されるが、これは後続して非相補的配列に変換され得る。

【0047】

非Y形非対称アダプター設計が、各DNA分子の反対の末端上に異なるプライマー配列を必要とする配列決定プラットフォームに適用され得る方法もまた、開示される。

10

【0048】

異なる種類のSMIタグ及びSDEが、読み取り長さ及びSMIタグ付け多様性を最大化するために、アダプターを含有する2つの異なるプライマー-部位内で分布され得る代替的な方法が、本明細書に開示される。

【0049】

ペアードエンド配列決定に容易に受け入れられるY形またはループ形尾部を含むが、SMIタグは相補的配列ではないため、著しい設計柔軟性を可能にする、二本鎖配列決定アダプターのための追加の設計もまた、本明細書に開示される。

【0050】

そのような非対称のそのような導入が、二本鎖配列決定によるエラー訂正のために、二本のDNA鎖に由来する生成物の区別をいかに可能にするかが、本明細書に実証される。いくつかの実施形態が、シングルエンド読み取りプラットフォーム上での二本鎖配列決定の実行をいかに促進するかについての記述が、本明細書に更に実証される。

20

【0051】

プライマー部位ならびに単一アダプターによる二本鎖配列決定のためのSMI部位及びSDE部位を導入して、環状アダプター-DNA分子複合体を形成するための方法が、更に開示される。

【0052】

二本の鎖の差次的な配列に基づく分子タグ付けではなく、独立した分析のための、別個の反応区画への対合した鎖の物理的/機械的分離を可能にする非対称化学タグ付けに依存する、SDEの導入に対する全く異なるアプローチが、更に開示される。

30

【0053】

特にIon Torrent(商標)(Life Technologies(登録商標))配列決定プラットフォームのためのアダプター設計の例が、本明細書に開示される。

【0054】

二本鎖分子の各末端の両方の単鎖に連結され得るアダプターの変異形、及び最終調製分子内に必要なSMI及びSDE要素の両方を保持する「ニック翻訳」後の一本鎖連結を可能にする設計が、本明細書に開示される。

【0055】

SDEが、アダプター連結とは独立した方法で、DNA分子自体内にいかに組み込まれ得るかが、本明細書に開示される。

40

【0056】

最後に、前述の一本鎖コンセンサス配列(SSCS)生成に対する必要性を排除する、いかなる二本鎖アダプター設計とも使用することができる、二本鎖配列決定のための合理化された代替的なアルゴリズムのアプローチが、本明細書に開示される。

【0057】

いくつかの実施形態において、ヌクレオチド配列の一部分は、「縮重」であり得る。縮重配列において、各位置は、任意のヌクレオチドであり得、すなわち、「X」、「N」、または「M」によって表される各位置は、アデニン(A)、シトシン(C)、グアニン(G)

50

G)、チミン(T)、またはウラシル(U)、あるいは任意の他の天然もしくは非天然DNAもしくはRNAヌクレオチドまたはヌクレオチド様物質または塩基対合特性を有する類似体(例えば、キサントシン、イノシン、ヒポキサンチン、キサンチン、7-メチルグアニン、7-メチルグアノシン、5,6-ジヒドロウラシル、5-メチルシトシン、ジヒドゥリジン、イソシトシン、イソグアニン、デオキシヌクレオシド、ヌクレオシド、ペプチド核酸、ロックド核酸、グリコール核酸、及びトレオース核酸)であり得る。あるいは、ヌクレオチド配列の一部分は、配列が、少なくとも1つの事前定義されたヌクレオチド、または少なくとも1つの事前定義されたポリヌクレオチドと、任意のヌクレオチドであり得る位置、または可能性のあるヌクレオチドのサブセットの組み合わせのみを含む1つ以上の位置とを含むように、完全に縮重でなくてもよい。可能性のあるヌクレオチドのサブセットの組み合わせとしては、以下、A、C、G、及びTのうちのいずれか3つ;以下、A、C、G、T、及びUのうちのいずれか2つ;またはUプラス、以下、A、C、G、及びTのうちのいずれか3つを挙げることができる。そのようなサブセットの組み合わせは、任意の他の天然もしくは非天然DNAもしくはRNAヌクレオチドまたはヌクレオチド様物質または塩基対合特性を有する類似体を更に含んでも、それで置換されてもよい。分子の集団におけるこれらのヌクレオチドのうちのいずれかの間化学量論比は、およそ1:1または任意の他の比率であり得、本明細書において、そのような配列は、「半縮重」と呼ばれる。特定の実施形態において、「半縮重」配列とは、2つ以上の配列の組を指し、2つ以上の配列は、少なくとも1つのヌクレオチド位置で異なる。実施形態において、半縮重配列は、全てのヌクレオチドがその隣接するヌクレオチド(すぐ隣接するか、または2つ以上のヌクレオチド以内)に関してランダムではない配列である。実施形態において、本明細書で使用される場合、縮重及び半縮重という用語は、本出願が属する当業者によって一般に理解されるもの、及び本出願が属する当該技術分野において一般に使用されるもの同一の意味を有し得、そのような当該技術分野は、その全体が参照により組み込まれる。

#### 【0058】

実施形態において、配列は、各位置に全ての可能性のある塩基を含有する必要はない。縮重または半縮重n量体配列は、ポリメラーゼ媒介法によって生成されても、既知の配列の個々のオリゴヌクレオチドのライブラリを調製し、アニールすることによって生成されてもよい。あるいは、いかなる縮重または半縮重n量体配列も、標的DNA源とは異なる任意の代替的な供給源に由来する、ランダムまたは非ランダムに断片化された二本鎖DNA分子であってもよい。いくつかの実施形態において、代替的な供給源は、細菌に由来するゲノムもしくはプラスミド、標的DNAの生物以外の生物、またはそのような代替的な生物もしくは供給源の組み合わせである。ランダムまたは非ランダムに断片化されたDNAは、SMIアダプター内に導入されて、可変タグとしての役割を果たし得る。これは、酵素連結または当該技術分野において既知である任意の他の方法を通して達成することができる。

#### 【0059】

本明細書及び添付の特許請求の範囲で使用される場合、別段文脈が明確に指示しない限り、単数形の「1つの(a)」、「1つの(an)」、及び「その」は、複数形の参照対象を含む。

#### 【0060】

具体的に述べられるか、または文脈から明らかでない限り、本明細書で使用される場合、「または」という用語は、包括的であることが理解され、「または」及び「及び」の両方を網羅する。

#### 【0061】

「1つ以上の」、「少なくとも1つの」、及び「2つ以上の」などの用語は、少なくとも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、31、32、33、34、35、36、37、38、39、40、41、42、43

10

20

30

40

50

、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、もしくは150、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、またはそれ以上、及びその間の任意の数を含むが、これらに限定されないことが理解される。

10

【0062】

逆に、「以下」という用語は、述べられる値未満の各値を含む。例えば、「100以下のヌクレオチド」は、100、99、98、97、96、95、94、93、92、91、90、89、88、87、86、85、84、83、82、81、80、79、78、77、76、75、74、73、72、71、70、69、68、67、66、65、64、63、62、61、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3、2、1、及び0のヌクレオチドを含む。

20

【0063】

「複数の」、「少なくとも2つの」、「2つ以上の」、及び「少なくとも第2の」などの用語は、少なくとも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、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、もしくは150、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、またはそれ以上、及びその間の任意の数を含むが、これらに限定されないことが理解される。

30

40

【0064】

本明細書を通して、「を含む (comprising)」という用語、または「を含む (comprises)」もしくは「(comprising)」などの変異形は、述べられる要素、整数、もしくはステップ、または要素、整数、もしくはステップの群の包含を暗示するが、いかなる他の要素、整数、もしくはステップ、または要素、整数、もしくはステップの群の排除も暗示しないことが理解されるだろう。

【0065】

具体的に述べられるか、または文脈から明らかでない限り、本明細書で使用される場合、「約」という用語は、当該技術分野の通常の許容範囲内、例えば、平均値の2標準偏差以内にあることが理解される。約は、述べられる値の10%、9%、8%、7%、6%、

50

5%、4%、3%、2%、1%、0.5%、0.1%、0.05%、0.01%、または0.001%以内にあることが理解され得る。別段文脈から明確でない限り、本明細書に提供される全ての数値は、「約」という用語によって修飾される。

【0066】

本発明の実施または試験においては、本明細書に記載される方法及び材料と類似または同等の方法及び材料を使用することができるものの、好適な方法及び材料が後述される。本明細書に言及される全ての出版物、特許出願、特許、及び他の参考文献は、それらの全体が参照により組み込まれる。本明細書に引用される参考文献は、主張される本発明に対する先行技術であるとは認められない。矛盾が生じた場合、定義を含む本明細書が制御するものとする。加えて、材料、方法、及び実施例は説明的なものに過ぎず、限定的であることは意図されない。

10

【0067】

別段定義されない限り、本明細書で使用される全ての技術及び科学用語は、本出願が属する当業者によって一般に理解されるもの、及び本出願が属する当該技術分野において一般に使用されるものと同じの意味を有し、そのような当該技術分野は、その全体が参照により組み込まれる。

【0068】

上記の態様及び実施形態のうちのいずれも、以下の実施例/実施形態を含む、要約、図面、及び/または発明を実施するための形態の節に開示されるいかなる他の態様または実施形態とも組み合わせることができる。

20

【0069】

本発明の具体的、非限定的な実施例/実施形態

二本鎖配列決定にY形アダプターを使用することによる不利益

Y形アダプターによる二本鎖配列決定は、元来記載されているように(WO2013/142389A1及びSchmitt et al, PNAS2012、これらの各々は、その全体が参照により本明細書に組み込まれる)、ペアードエンド配列決定読み取りによって最も容易に実行される。しかしながら、全ての配列決定プラットフォームが、ペアードエンド配列決定読み取りと適合するわけではない。以前に記載されているY形またはループ形アダプター(非対称プライマー部位がアダプターの連結可能な末端の反対の一本鎖領域内に位置する)を使用する場合、シングルエンド配列決定読み取りによる二本鎖配列決定は、DNA分子を通じた配列決定読み取りの完全な延長を必要とする。これは、配列決定読み取りを二本の誘導体鎖から区別するために必要とされる、分子の両末端のSMIタグ配列を捕捉するために必要である。この必要性は、以下の通り説明される。

30

【0070】

以前に記載されているY形二本鎖配列決定アダプターを、図1Aに示す。図1Aにおいて、特徴A及びBは、異なるプライマー結合部位を表し、及び'は、縮重または半縮重配列及びその逆相補体を表し、は、異なる縮重または半縮重配列を表し、及びは、縮重または半縮重配列のプール内の2つの任意配列である。合わせて、これらは、単一分子識別子(SMI)としての役割を果たす。

【0071】

40

元来記載されているように(例えば、WO2013/142389A1)、SMIは、より大きなプール内の個々の分子を区別するために使用される。任意の2つのDNA分子が同一のSMI配列で標識される可能性が統計的に低いように、アダプターライブラリにおいてコードされるこれらの十分に大きな集団を有することが必要である。また、以前に記載されているように、ライブラリ生成中に導入される断片化部位は、特定の状況において、独立して、またはアダプター配列内にコードされる外来性SMIとの組み合わせでのいずれかで、内在性SMIとして機能し得る。本開示において、異なるアダプター設計の実施例において外来性SMIドメインのみが示されるが、外来性SMIドメインは、内在性SMIとして作用するDNA剪断点で置換されても、またはそれによって増強されてもよいことが理解され(かつ本明細書に含まれる)。

50

## 【 0 0 7 2 】

アダプターが、あるライブラリからの二本鎖 DNA 断片の各末端に連結された後、図 1 B に示される構造が現れるだろう。後続する図において誘導体を追跡する明確さのために、後続する図中の誘導体、特定の DNA 挿入物の「左」末端及び「右」末端、ならびに「上部」鎖及び「下部」鎖が留意される。

## 【 0 0 7 3 】

PCR の後、「上部」鎖に由来する二本鎖生成物が、図 1 C に示される。( L ) 及び ( R ) は、出発 DNA 分子のそれぞれ「左」末端及び「右」末端を示す。

「上部」鎖に由来する二本鎖 PCR 生成物が、図 1 D に示される。

## 【 0 0 7 4 】

「上部」鎖生成物及び「下部」鎖生成物中、A 及び B に対する 及び の異なる配置が留意されるべきである。ペアードエンド配列決定読み取り(すなわち、各 PCR 生成物のプライマー部位 A 及び B の両方からの読み取り)について、 タグは一方の鎖の A 読み取りに現れ、 タグは B 読み取りに現れ、相反の事例が他方の鎖で生じるために、各鎖に由来する生成物を区別することが可能である。図 1 E を参照されたい。

## 【 0 0 7 5 】

上述のペアードエンド読み取りの使用は、二本鎖配列の訂正を可能にする。しかしながら、シングルエンド配列決定読み取りのみ(すなわち、特定の分子の両方の部位からではなく、プライマー部位 A またはプライマー部位 B からの読み取りのみ)の使用では、配列決定読み取りが両末端の SMI 配列を捕捉するのに十分に長い場合にのみ二本鎖配列を得ることが可能である。配列決定プライマー A を使用する場合、異なる鎖から得られる完全長の配列決定読み取り(すなわち、両方の SMI 配列を含めるのに十分に長い)は、図 1 F に示される 2 つの配列をもたらすだろう。同様に、完全長の配列決定読み取りを有する配列決定プライマー B の配列決定の使用は、図 1 G に示される以下の 2 つの配列を生成するだろう。上記の事例の両方において、「上部」鎖及び下部鎖由来の生成物は、反対の配向に SMI を有する(一方では - 及び他方では - )のために、互いに区別することができる。しかしながら、両方の SMI 配列を捕捉するのに十分に長い配列決定読み取りがないと、二本鎖配列決定は、シングルエンド配列決定によっては容易には実行されない。これは、2 つの配列決定読み取りが各々、 タグ及び タグの両方を含有しないためである。この問題の別の見方は、DNA 分子の末端の部分について、相補体が配列決定され得ないため、比較を行うための第 2 の鎖についての情報が存在しないことである。

## 【 0 0 7 6 】

これを説明するために、プライマー A からの非完全長のシングルエンド配列決定読み取りを使用する場合に生成される 2 種類の配列を、図 1 H に示す。同様に、プライマー B からの非完全長のシングルエンド配列決定読み取りを使用する場合に生成される対応する配列を、図 1 I に示す。図 1 H 及び図 1 I に示される配列決定読み取りの両方について、各 DNA 断片の「左」末端及び「右」末端は、所与のプライマーで一度だけ配列決定されるため、二本鎖配列決定は達成することができないことに留意されたい。これは、それを比較するいかなる反対の鎖配列も存在しないためである。故に、増幅された分子の集団が 2 つの異なるプライマーの各々で配列決定された場合ですら、読み取り A 配列及び B 配列の特定の組が同一の誘導体分子に起源を持つことを明らかにする第 2 の鎖についてのいかなる情報も存在しないだろう。

## 【 0 0 7 7 】

シングルエンド配列決定の使用時、完全 DNA 分子の「リードスルー」に対する必要性は、読み取り長さが限定されているいくつかの配列決定プラットフォームに技術的な課題をもたらし得る。

## 【 0 0 7 8 】

二本鎖配列決定が、シングルエンド配列決定による亜完全長の配列決定読み取りと適合するためには、代替的なアダプター設計が必要である。ペアードエンド配列決定読み取り及び Y 形アダプター上の非対称プライマー部位の不在下では、鎖の区別を可能にするため

10

20

30

40

50

に、いくつかの他の形態の非対称が適応したDNA分子内に導入されなくてはならない。そのような設計の例を、以下に開示する。

【0079】

非相補的「バブル」による鎖定義非対称の導入

非ペアードエンド配列決定（すなわち、「バブルアダプター」）による二本鎖配列決定を可能にする（本発明の）非Y形アダプターの一例示的な設計が、図2Aに開示される。2つのプライマー部位を有する、以前に記載されているY形アダプターとは異なり、逆相補体（P'）を有する単一プライマー部位（P）のみが存在する。及び相補体'は、縮重または半縮重単一分子識別子（SMI）配列を表し、X及びYは、アダプター内の隣接する相補的配列の中間に非対合の「バブル」を形成する非相補的配列のセグメントである、鎖定義要素（SDE）の2つの半分を表す。最後に、アダプターは、連結可能な配列を有する。SDEによってこのアダプター設計内に導入される非対称は、図2B～図2Gに説明されるような各鎖から得られる配列決定読み取りを区別する。

10

【0080】

図2Aに示されるアダプターに類似するアダプターがDNA断片の各末端に連結した後、図2Bに示される構造が生成される。第2のアダプターは、第2の連結されたアダプターのSMI配列が一般に第1のアダプターのSMI配列とは異なることを説明するために、SMI配列'及び'とともに示される。あるいは、同一のアダプターが、DNA分子の両末端に連結されてもよい。

【0081】

PCR増幅後、「上部」鎖に由来する二本鎖生成物を図2Cに示し、「下部」鎖に由来する二本鎖生成物を図2Dに示す。

20

【0082】

この実施例において、プライマー部位配列は分子の両末端で同一であるため、どちらの一本鎖の半分が配列決定されるかによって、各鎖のPCR生成物のシングルエンド配列決定読み取りから、2つの異なる種類の配列の配列決定読み取りが得られるだろう。「上部」鎖のPCR生成物から得られる読み取りを図2Eに示し、「下部」鎖のPCR生成物から得られる読み取りを図2Fに示す。

【0083】

分析のために、図2Gに示されるように、配列決定読み取りは、特定のSMI（この場合、または'のいずれか）を含有するものによってグループ化される。DNAの所与の単一分子から生じた配列は、同一のSMI配列を有するために、ともにグループ化される。各SMI群において、2種類の配列（1つはSDE Xによってマークされ、1つはSDE Yによってマークされる）が見られることは明らかである。これらが、反対の鎖（すなわち、「上部」及び「下部」）から得られる配列決定読み取りを定義する。例えば、SMIタグ'を有する配列がともにグループ化される場合、得られる配列は、X - - DNA（図2E）及びY' - - DNA（図2F）である。元のDNA分子の「上部」鎖から生じる配列からなるコンセンサスは、X - - DNA配列をともにグループ化することによって作製することができる。同様に、「下部」鎖のコンセンサスは、Y' - - DNA配列をともにグループ化することによって作製することができる。最後に、二本の鎖のコンセンサスは、二本の鎖から生じる配列をともに比較すること（すなわち、配列Xで標識されたものは、配列Y'で標識されたものと比較されるだろう）によって作製することができる。合わせて、これらは、二本鎖配列決定分析の一部としての比較を可能にする。

30

40

【0084】

SMI配列及びSDE配列の順序を転換することによって、類似する結果を達成することができる。そのようなアダプターの一例を、図2Hに示す。

【0085】

上記及びWO2013/142389Aに明確化されるように、いくつかの実施形態において、アダプター配列内に含有されるSMIが、DNA分子自体の剪断点配列を含む内

50

在性 S M I 配列の代わりに取り除かれてもよい。1つのそのようなアダプター設計の構造は図 2 A に示されるものを伴うが、及び ' は排除されるだろう。

【 0 0 8 6 】

いくつかの用途において、図 2 H に示される配向が好ましい。例えば、現在 I l l u m i n a (登録商標)によって製造されているものなどのいくつかの配列決定プラットフォームにおいて、配列決定実行初期の特定の数の塩基をクラスター特定及び「不変異塩基」に使用することができ、つまり、配列決定される全ての分子内または実質的に複数の分子内で同一のものとして読み取られる塩基は、このプロセスの効率に影響を与える可能性がある。したがって、この状況において、配列決定実行の最初期の縮重または半縮重 S M I 配列がより望ましくあり得る。

10

【 0 0 8 7 】

他の用途において、図 2 A に示される配向が好ましい。二本鎖配列決定についての初期の記述(すなわち、W O 2 0 1 3 / 1 4 2 3 8 9 A 1)に記載されるように、相補的二本鎖 S M I 配列は、一本鎖縮重もしくは半縮重配列にわたってポリメラーゼによってプライマー延長すること、または異なる S M I 配列を含有するオリゴヌクレオチドを個々に合成及びアニールし、その後それらとともにプールして、多様なアダプターライブラリを作製することのいずれかによって、最も好都合に生成することができる。ポリメラーゼ延長法が選択される場合、アダプターの連結 - ドメイン末端上に S M I 配列を有することが、延長反応を促進するのに有利であり得る。I o n T o r r e n t (商標)によって製造されるものなどの特定の配列決定プラットフォーム上では、アダプターの連結不可能な末端に修飾塩基を有する 3 ' オーバーハングは、ポリメラーゼによる合成と容易には適合しない可能性があるため、ポリメラーゼ延長アプローチによるアダプターの合成は、図 2 A に示されるように、アダプターの連結可能な末端に向かって位置する S M I 配列で最も容易に実行される。

20

【 0 0 8 8 】

このアプローチがいかに実施されるかについての一具体例として、以下のアダプターの対を使用し得る I o n T o r r e n t (商標)配列決定プラットフォームを考慮する。

アダプター P 1

【化 1】

5' CCACTACGCCTCCGCTTTCTCTCTATGGGCAGTCGGTGAT 3' (配列番号 1)

30

3' T\*T\*GGTGATGCGGAGGCCGAAAGGAGAGATACCCGTCAGCCACTA 5' (配列番号 2)

アダプター A

【化 2】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG 3' (配列番号 3)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC 5' (配列番号 4)

アスタリスク「\*」は、ホスホロチオエート結合を表す。

40

【 0 0 8 9 】

配列決定プライマーはアダプター A にアニールし、故にアダプターの 3 ' 末端から開始する D N A 断片から配列情報が読み取られる。アダプター A は、以下の配列を使用して、図 2 に図示されるアプローチに適用可能な形態に変換され得る。

【化 3】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG GCGC NNNN G 3' (配列番号 5)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC ATAT MMMM C 5' (配列番号 6)

【 0 0 9 0 】

N N N N とは、縮重または半縮重 4 ヌクレオチド配列を指し、M M M M とは、その相補

50

体を指し、G - C塩基対は、連結を促進するように縮重配列の下流に含まれるが、他の形態の連結ドメインが使用されてもよい。

【0091】

この説明において、アダプターP1及びアダプターAはともに、配列決定される標的DNA分子に連結される。単純化のため、DNA分子の両末端に連結される同一のアダプターは無視され得る。しかしながら、Ion Torrent（商標）アダプターは、分子の各末端上の異なるアダプターを利用する。初期連結時、個々のDNA分子は、様々な構成、例えば、A - DNA - P1、A - DNA - A、またはP1 - DNA - P1で、アダプターによって連結され得る。部位A及び部位P1に対して指向されたプライマーによるエマルジョンPCRで増幅されるために、配列決定反応には、A - DNA - P1の正確な構成が利用され得る。あるいは、当該技術分野において既知である、2つの異なるアダプターに連結された分子のみを選択するための他の方法が、使用され得る。

10

【0092】

増幅及び配列決定時、以下の生成物が得られるだろう。

【化4】

GCGC NNNN [DNA配列]

TATA NNNN [DNA配列]

【0093】

これらは、図2Gに示される生成物X - - DNA及びY' - - DNAに対応することに留意されたい。

20

【0094】

その後、元来記載されているように（例えば、WO2013/142389A1を参照されたい）、二本の鎖に由来する生成物は、二本鎖配列決定を介したデータ処理のためにもマッチされ得る。具体的には、配列GCGC NNNNで始まる読み取りからコンセンサスを作製して、「上部」鎖のコンセンサスを得ることができる。配列TATA NNNNで始まる読み取りから別個のコンセンサスを作製して、「下部」鎖のコンセンサスを得ることができる。その後、2つの一本鎖コンセンサス配列を比較して、出発DNA分子の二本鎖コンセンサス配列を得ることができる。代替的なデータ処理アプローチを、以下に開示する。「二本鎖配列決定のための代替的なデータ処理スキーム」を参照されたい

30

【0095】

この実施形態において、DNA配列情報はDNA断片の2つの末端のうち的一方のみから必要とされるため、上記のアプローチは、ペアードエンド読み取りが不可能な短い読み取りを利用するプラットフォーム上での二本鎖配列決定を可能にする。

【0096】

このアプローチの代替的な一実施形態は、二本鎖で非相補的または部分的に非相補的なSMIの使用を介して、SMI配列自体内に非対称を導入することであるだろう。SMI配列自体は相補的ではない一方で、非相補的SMI配列から生じる生成物は、対を形成することが事前決定されているために、同一の出発二本鎖DNA分子から生じていることが決定され得る。

40

【0097】

この実施形態の一具体例として、以下の配列を有する一連のIon Torrent（商標）「アダプターA」分子を考慮する。

アダプター1：

【化5】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG AAAT GCAGC 3' (配列番号7)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC GGGC CGTCG 5' (配列番号8)

50

アダプター 2 :

【化 6】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG ATAT GCAGC 3' (配列番号 9)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC GCGC CGTCG 5' (配列番号 10)

アダプター 3 :

【化 7】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG TATT GCAGC 3' (配列番号 11)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC GGCC CGTCG 5' (配列番号 12)

10

アダプター 4 :

【化 8】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG ATTT GCAGC 3' (配列番号 13)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC CGGG CGTCG 5' (配列番号 14)

【0098】

単純化のため、4つのアダプターのみが上記に列挙されるが、実際には、そのようなアダプターのより大きなプールを有することが望ましくあり得る。この実施例において、相補的配列は非相補的配列の下流に含まれて、DNA分子への連結を促進する二本鎖領域を形成することに留意されたい。

20

【0099】

個々のDNA断片は、二本のDNA鎖の非対称標識をもたらす個々のアダプターに連結される。具体的には、配列決定時、出発DNA分子の「上部鎖」の配列は、アダプターの「上部鎖」内の配列で標識されるだろう。出発DNA分子の「下部鎖」の配列は、アダプターの「下部鎖」の配列の逆相補体で標識されるだろう。

【0100】

具体的な一例として、アダプター 1 に連結された二本のDNA鎖は、AAAT (上部鎖) 及びCCCG (下部鎖) と標識される。繰り返すと、配列決定時、下部鎖は、最初にはアダプターの下部鎖内に存在する配列の逆相補体をもたらすことに留意されたい。同様に、他のアダプターに連結された配列について、分子識別子は、それらの対合したタグのために、ともに対合され得る。その後、コンピュータプログラムは、そのアダプターから得られる既知のタグ配列の表を使用して、それらを単一DNA分子の相補的鎖から生じる読み取りへと組み立てることができる。表 2 は、結果として得られる配列読み取りが、上記の実施例に示される特定の非相補的識別子配列に基づいていかに標識されるかを示す。

30

【表 2】

表 2.

配列決定読み取りの最初 4 つのヌクレオチド

	上部鎖	下部鎖
アダプター 1	AAAT	CCCG
アダプター 2	ATAT	CGCG
アダプター 3	TATT	CCGG
アダプター 4	ATTT	GCCC

40

【0101】

これらは、特定の実施形態の具体例に過ぎない。当業者にとって、SMIタグはいかなる任意の長さであってもよいこと、SMIは完全にランダムであってもよいこと、全体が事前定義された配列からなってもよいことは明らかだろう。SMI配列が二本鎖分子の両方の鎖内にあるとき、(上記の実施例に言及される第 1 の例に記載されるように) 2つの

50

S M I 配列は完全に相補的であっても、部分的に非相補的であっても、完全に非相補的であってもよい。いくつかの実施形態において、いかなる外来性分子識別子タグも全く必要とされない。場合によっては、二本鎖DNAの所与の単一分子の二本の独立した鎖から生じるものとして生成物を区別することを可能にする、何らかの非対称（S D Eを含む）が存在する限り、特有な識別子として、DNA分子のランダムに剪断された末端が使用されてもよい。

#### 【0102】

本明細書に開示されるいかなる本発明の態様または実施形態においても（かつ現在記載されている実施形態に限定されない）、一本鎖S M I及び二本鎖S M Iの両方において、S M Iタグの組は、S M I配列の合成、増幅、または配列決定のエラーが、1つのS M I配列の別のS M I配列への変換をもたらさないような別個のタグ間の編集距離をもって設計され得る（例えば、Shiroguchi et al, Proc Nat Acad Sci USA, 109(4):1347-1352）を参照されたい。S M I配列間の編集距離の組み込みは、例えば、ハミング距離、ハミングコード、または当該技術分野において既知である別のエラー訂正法を使用することによる、S M Iエラーの特定及び除去を可能にする。ある組に由来する全てのS M Iは同一の長さであってもよく、あるいは2つ以上の異なる長さのS M Iの混合物があるS M Iの組において用いられてもよい。混合したS M I長さの使用は、S M I配列を使用し、かつS M I内またはそれに隣接する部位に1つ以上の固定された塩基を更に有するアダプター設計に有利であり得るが、これは、ある組における2つ以上の長さのS M Iの利用が、配列決定中に不変異塩基（複数可）を全て同一の読み取り位置で生じさせないためである（例えば、Hummelen et al, PLoS One, 5(8):e12078(2010)）を参照されたい。このアプローチは、特定の読み取り位置に不変異塩基が存在する状況において最適以下の性能に遭遇し得る、配列決定装置プラットフォーム上で生じ得る問題（例えば、クラスター特定に伴う困難）を回避することができる。

#### 【0103】

当業者にとって、非対称を導入する配列が、例えば、上記に示される内部「バブル」配列として、S M I配列の前後、またはそのような配列を有するアダプター設計において一本鎖「尾部」配列内を含む、配列決定アダプター内のどこに導入されてもよいこともまた明らかだろう。これらの配列及び任意の関連するS M I配列は、配列決定読み取りの一部として直接読み取ることができるか、またはあるいは独立した配列決定反応（例えば、インデックス読み取りにおいて）から決定することができる。更に、これらの配列は、Y形アダプター、「ループ」アダプター、または当該技術分野において既知である任意の他のアダプター設計と組み合わせて使用することができる。

#### 【0104】

実際には、S M I配列、S D E配列、及びプライマー結合部位の異なる相対的配向を有するアダプターが、本発明に想定され、含まれる。

#### 【0105】

図2A及び図2Hに示されるアダプター設計は、非連結末端を平滑末端として示す。しかしながら、この末端は、オーバーハングであっても、陥凹であっても、分解または所望されない連結を防止するための修飾塩基または化学基を有してもよい。

#### 【0106】

更に、アダプターの二本の鎖は接続されて、いくつかの用途において分解または所望されない連結を防止するのに望ましくあり得る、閉鎖「ループ」を形成してもよい。例えば、図2Iを参照されたい。図2Iの閉鎖「ループ」結合（「S」位にマークされる）は、従来のホスホジエステル結合によって、または任意の他の天然もしくは非天然化学リンカー基によって達成することができる。この結合は、連結の実行前、実行中、または実行後に「開放」末端を達成するために、化学的または酵素的に切断され得、ループの切断は、ローリングサークル型増幅産物を防止するために、PCR増幅前に望ましくあり得る。ここでウラシルなどの非標準塩基を使用してもよく、アダプター連結前、連結中、または連

10

20

30

40

50

結後に、酵素的ステップの組を使用して、ホスホジエステル骨格を切断することができる。例えば、ウラシルの場合、脱塩基部位を形成するためのウラシルDNAグリコシラーゼと、骨格を切断するためのエンドヌクレアーゼVIIIIとの組み合わせの使用が十分だろう。あるいは、この結合部位の嵩高い化学基または他の塩基転換不可能な修飾塩基を使用して、ポリメラーゼがループの末端を超えて横断するのを防止し、同一の目的を果たすことができる。

#### 【0107】

二本鎖SMI配列を使用するアダプター設計のための、本明細書に開示されるいかなる本発明の態様または実施形態においても（かつ現在記載されている実施形態に限定されない）、それが相補的、部分的に非相補的、または完全に非相補的であるかに関わらず、「ループ」形態へとアニールされている直鎖状分子としてアダプターを合成する具体的な利点は、「上部」及び「下部」鎖SMI配列が分子自体内に1：1の比率で存在することである。このアプローチは、個々の「上部」及び「下部」オリゴヌクレオチド対をアニールして、二本鎖SMIを形成することと比較して有利であり得るが、これは、そのようなアプローチにおいて、「上部」鎖及び「下部」鎖に使用されるオリゴヌクレオチドの濃度が完全な1：1の比率ではない場合、一方のアダプター鎖または他方のアダプター鎖の過剰な分子が存在し得、下流のステップにとって問題となり得る（例えば、追加の一本鎖オリゴヌクレオチドは、PCR増幅中に不適切なプライミングを引き起こし得るか、または存在し得る他の一本鎖オリゴヌクレオチドとアニールし得、二本のSMI鎖が適切に都合されていないアダプター分子を作製し得る）ためである。

#### 【0108】

いくつかの例において、完全なループ配列自体の複製を防止することが望ましくあり得、修飾配列位置が複製ブロックとして任意で含まれ得る。これは、酵素的に除去することができる塩基（例えば、ウラシルDNAグリコシラーゼによって除去され得るウラシル）であってもよく、例えば、DNA複製を部分的または完全に阻害する領域（例えば、脱塩基部位）であってもよい。

#### 【0109】

あるいはまたは更に、「開放」立体構造を達成するために使用され得る、制限エンドヌクレアーゼ部位が導入され（図2Iにおいて「T」位にマークされる）、小ヘアピン断片が結果として放出されてもよい。

#### 【0110】

二本のアダプター鎖間の塩基非対称の異なる配置が鎖定義要素として等しく役割を果たすことは、容易に明らかだろう。さもなければ相補的な鎖に対する1つのヌクレオチドまたは2つ以上のヌクレオチドの挿入が存在するとき、バブルがアダプター鎖内に形成され得るが、図2Jのアダプターに示される。図2Kは、2つ以上のヌクレオチド挿入が自己相補的である一部分を含むアダプターを示し、この後者のアダプターは、1つ以上のヌクレオチド位置に関与する二本の鎖間の単純な差異として類似の機能性を提供する。

#### 【0111】

非相補的SMI配列を使用する鎖定義非対称の導入

図2A～2Kに示されるアダプター設計は、タグに基づく二本鎖配列決定を可能にする2つの重要な特徴を含有する。一方は、特有な分子識別子（すなわち、SMI）であり、他方は、二本のDNA鎖内への非対称の導入手段（すなわち、SDE）である。二本鎖配列決定の初期の記述において、Y形アダプター及びペアードエンド配列決定読み取りが利用された。二本のDNA鎖内への非対称の導入は、非対称尾部自体のために達成された。図3Aに示されるように、別個かつ優れた二本鎖配列決定アダプター設計は、共同で分子識別子及び非対称導入SDEとしての役割を果たす非相補的「バブル」形SMIを含む。

#### 【0112】

この設計において、P及びP'はそれぞれ、プライマー部位及びその相補体を表し、i及びiiは、それらの長さの全てまたは一部分について非相補的である2つの縮重または半縮重配列を表す。この形態のアダプターの合成は、異なる縮重または半縮重配列を

10

20

30

40

50

有するオリゴヌクレオチドの対のうちの2つ以上をともにプールして、多様なプールを形成する前に、これらを個々に合成し、ハイブリダイズすることによって、最も容易に達成される。オリゴヌクレオチドが個々に合成され、アニールされるため、所与の  $i$  配列と  $i i$  配列との間の関係が既知となり、データベースに記録され、これが、配列決定後の分析中、対応するパートナー S M I 配列について検索され得る。

#### 【0113】

二本鎖 DNA 断片へのアダプター連結後、図 3 B に示される構造が生成される。この構造において、同一のアダプター構造ではあるが、 $i$  及び  $i i$  とは一般に別個である非相補的 S M I 配列の対である、 $i$  及び  $i i$  が、両末端に連結され得る。

#### 【0114】

P C R 増幅後、「上部」鎖に由来する二本鎖生成物を図 3 C に示し、「下部」鎖に由来する二本鎖生成物を図 3 D に示す。

#### 【0115】

(この実施例において)プライマー部位配列は分子の両末端で同一であるため、各鎖の P C R 生成物のシングルエンド配列決定読み取りから、及びどちらの一本鎖が配列決定されるかによって、2つの異なる種類の配列読み取りが得られるだろう。「上部」鎖 P C R 生成物からのシングルエンド配列決定読み取りを図 3 E に示し、「下部」鎖からのシングルエンド配列決定読み取りを図 3 F に示す。

#### 【0116】

その後、分析中、読み取りは、S M I アダプターライブラリ合成の時点で、かつそれと組み合わせて生成されたデータベースから既知である関係に基づいて、特定の S M I 配列及びそれらの対応する非相補的パートナーによってグループ化され得る。図 3 G に示されるように、元の分子の対合した「上部」及び「下部」鎖配列は、分子の一端に起源を持つ読み取りについて  $i$  及び  $i i$  で、ならびに反対の末端に起源を持つ読み取りについて  $i$  及び  $i i$  でタグ付けされる。

#### 【0117】

修飾または非標準ヌクレオチドを使用する鎖定義非対称の導入

二本鎖配列決定アダプターに鎖非対称が導入され得る別の方法は、最初是对合した鎖 DNA を形成するが、その後、更なる生化学的ステップ後に DNA ミスマッチをもたらす、ヌクレオチドまたはヌクレオチド類似体による。この一例が、DNA ポリメラーゼ誤組み込みである。誤組み込みは、増幅中に本質的に、または化学的もしくは酵素的ステップを介したミスマッチ領域への変換後のいずれかで生じ得る。

#### 【0118】

いくつかの用途では、上記に開示される「バブル型」配列にとってこの形態の S D E が好ましくあり得、これは、それらが、遊離一本鎖領域から生じ得る問題、例えば、他の DNA オリゴヌクレオチドへの誤アニール及びエキソヌクレアーゼ/エンドヌクレアーゼ分解を回避するためである。

#### 【0119】

当該技術分野において既知である多くの非標準ヌクレオチドが、この目的を果たすことができる。そのような修飾ヌクレオチドの非限定的な例としては、テトラヒドロフラン；8 - オキソ - 7 , 8 - ジヒドロ - 2 ' - デオキシアデノシン ( 8 - オキソ - A ) ; 8 - オキソ - 7 , 8 - ジヒドロ - 2 ' - デオキシグアノシン ( 8 - オキソ - G ) ; デオキシイノシン、5 ' - ニトロインドール；5 - ヒドロキシメチル - 2 ' - デオキシシチジン；イソ - シトシン；5 ' - メチル - イソシトシン；及びイソ - グアノシン、ならびに当該技術分野において既知である他のものが挙げられる。

#### 【0120】

8 - オキソ - G 含有二本鎖配列決定アダプターを、図 4 A に示す。8 - オキソ塩基は相補的シトシン塩基の反対に対合され、バブルは形成されない。上記及び以下の実施例のように、S M I 配列 ( この場合、 ) 及び S D E 部位 ( この場合、8 - オキソ - G 部位 ) の相対的順序は、必要に応じて転換されてもよい。P 及び P ' は、プライマー部位及びその

10

20

30

40

50

相補体を表す。

【0121】

二本鎖DNA断片へのアダプター連結後、図4Bに示される構造が生成される。

【0122】

その後、オキソグアニングリコシラーゼ(OGG1)などのグリコシラーゼによる図4Bの二本鎖DNAの処理が、(潜在的には、リアーゼ活性を有するグリコシラーゼによって生じ得る、結果として得られるニックを修復するために、DNAリガーゼと組み合わせ)実行され得る。この処理は、図4Cに示されるように、脱塩基部位の導入を有する、無傷のホスホジエステルDNA骨格をもたらすだろう。その後、二本の鎖の各々が、例えば、ポリメラーゼによってコピーされ得る。適切な反応条件下で、特定の熱安定性ポリメ  
10  
ラーゼは、Aの反対の脱塩基部位を優先的に挿入し(Belousova EA et al, Biochim Biophys Acta 2006)、GからTへの変異をもたらす。対照的に、相反の鎖は、連結時にアダプター内に存在していたCヌクレオチドを保持する。この処理は、二本の鎖の生成物の区別を可能にする鎖非対称をもたらす。

【0123】

PCRまたは他の形態のDNA増幅中、適切な条件下で、特定のポリメラーゼによって、アデニンは、鎖のコピー時、脱塩基の反対に優先的に挿入されるだろう。コピーに引き続いて、このアデニンはチミンに対合され、究極的には元の8-オキソ-G部位のTでの置き換えをもたらすだろう。更に、グリコシラーゼによる処理は必須ではない。適切な反応条件下で、ポリメラーゼは、示される脱塩基中間体なしで、Aの反対の8-オキソ-G  
20  
を挿入することができる(Sikorsky JA et al Biochem Biophys Res Commun 2007)。いずれの場合も、PCR増幅後、「上部」鎖に由来する二本鎖生成物を図4Dに示し、「下部」鎖に由来する二本鎖生成物を図4Eに示す。

【0124】

この(非限定的な)実施例において、プライマー部位配列は分子の両末端で同一であるため、どちらの一本鎖が配列決定されるかによって、各鎖のPCR生成物のシングルエンド配列決定読み取りから、2つの異なる種類の配列読み取りが得られるだろう。「上部」鎖PCR生成物に由来するそれらのPCR生成物は、図4Fに示される通りであり、「下部」鎖に由来するそれらのPCR生成物は、図4Gに示される通りである。  
30

【0125】

分析中、配列決定読み取りは、特定のSMI(この場合、または )を含有するものによってグループ化され得る。図4Hを参照されたい。各SMIグループ化内のT及びGでマークされた生成物は、起源の鎖を定義し、二本鎖配列比較を可能にする。

【0126】

当業者にとって、修飾ヌクレオチドまたは他の類似体から得られる配列がDNA配列決定時に回収され得る限り、上述のように、修飾ヌクレオチドまたは別の類似体が、配列決定アダプター内のどこに置かれてもよいこともまた明らかだろう。

【0127】

当業者にとって、同一の目的を果たすために、多くの他のヌクレオチド類似体が利用され得ることは明らかだろう。他の例としては、テトラヒドロフラン及び8-オキソ-7,8-ジヒドロ-2'-デオキシアデノシン(8-オキソ-A)が挙げられる。DNAポリメラーゼによる異なるヌクレオチドの誤組み込みを本質的にもたらし得るか、あるいは酵素的もしくは化学的ステップによって、または経時的に自然に、誤コード傷害またはミス  
40  
マッチ塩基へと変換され得る、任意のヌクレオチド修飾が、この実施形態のアダプター内で使用され得る。

【0128】

更に、二本の鎖を非対称的に標識するために、非ヌクレオチド分子を組み込んでよい。例えば、二本のアダプター鎖のうち的一方にビオチンを組み込んでよく、これは、ストレプトアビジンを利用して、ビオチンを欠如する鎖からビオチン含有鎖を物理的に分離  
50

することによって、二本の鎖の別個の分析を促進するだろう。この実施形態を、以下に詳細に開示する。

【0129】

DNA分子の反対の末端上に異なるプライマー部位を導入するための、二本鎖配列決定アダプター設計の組み合わせの使用

非Y形アダプターの前述の実施例は、DNA分子の両末端への、同一の種類のアダプターの対称連結を示す。現在、ほとんどの配列決定プラットフォームは、例えば、表面またはビーズのいずれかの上でのクラスター増幅を可能にするために、適応したDNA分子がいずれかの末端上に異なるプライマー部位を有することを必要とする。これらの異なるプライマー部位を作製するためにY形アダプターを通常使用しない配列決定プラットフォーム（例えば、Ion Torrent（商標）（Thermo（登録商標）Inc）、SOLID（Applied Biosystems（登録商標）Inc）、及び454（Roche（登録商標）Inc））では、2つの異なるアダプターの混合物が連結され、その後、最も一般的にはビーズに基づくエマルジョンPCRプロセスを通して、各プライマー部位のうちの1つを含有する分子が選択される。

10

【0130】

非Y形二本鎖配列決定アダプターを使用して非対称プライマー部位を生成するための1つの類似するアプローチを、以下に説明する。

【0131】

このために、1つの二本鎖アダプターと1つの標準アダプターとの混合物を生成し、各アダプターは異なるPCRプライマー部位を含有する。二本鎖アダプターは、本明細書に上述もしくは後述されるか、または当該技術分野において既知である、いかなる設計であってもよい。

20

【0132】

相補体P'を有するプライマー部位Pを有し、その後に mismatch 配列X及びY（各々1つ以上のヌクレオチドを含む）で構成されるSDEが続き、その後に縮重または半縮重SMI配列が続く、一例示的な二本鎖アダプターを図5Aに示す。図5Bに示される他のアダプターは、相補体O'を有する異なるプライマー部位Oを含有する、「標準」アダプターである。

【0133】

DNAライブラリへのこのアダプター混合物の連結後、図5C~図5Eに示される3つの異なる種類の生成物が生成される。平均で、適応に成功した分子の半分が各末端上に異なるアダプター配列を担持し（図5C）、4分の1が2つの二本鎖アダプターを有し（図5D）、4分の1が2つの標準アダプターを有する（図5E）だろう。適切な選択条件下で、1つのプライマー部位Pを有する分子及びプライマー部位Oのみが、クラスター増幅するだろう。故に、後者の2（有用ではない）種類の生成物は今後無視され得、後続する記述においては示されない。

30

【0134】

PCR増幅後、「上部」鎖に由来する二本鎖生成物を図5Fに示し、「下部」鎖に由来する二本鎖生成物を図5Gに示す。

40

【0135】

プライマー部位Pからの配列決定は、「上部」鎖及び「下部」鎖から得られる以下の配列をもたらすだろう。これらは、SDE XまたはY標識のいずれかを担持しているために、区別することができる。図5Hを参照されたい。

【0136】

本明細書に記載されるか、または当該技術分野において既知である、任意の他の形態の非Y形二本鎖アダプターが、この実施形態において使用されるものと同じの目的を果たし得ることは、容易に明らかである。例えば、1つの二本鎖アダプター及び1つの標準アダプターの代わりに、異なるプライマー部位を担持する2つの二本鎖アダプターを使用することが可能である。連結及びPCR後、増幅された生成物は分けることができ、一方の部

50

分はプライマーPで配列決定され、他方はプライマーOで配列決定される。これは、各適応した分子の両末端の二本鎖配列決定を可能にするだろう。異なるプライマー部位からの読み取りは、実際にはペアードエンドではないため、それらは、いかなる特定の分子についても、容易にはともに関連付けることができない。しかしながら、配列決定されるDNAの量が非常に限定されている用途では、分子の両末端の二本鎖配列決定から得られる追加の配列情報が依然として有利であり得る。

【0137】

非ペアードエンドプラットフォーム上での2つの読み取りの使用は、二本鎖配列決定中の読み取り長さを最大化することができる

Illumina (登録商標) 機器で実行されるものなどのペアードエンド配列決定は一般に、配列決定プラットフォームが、アダプターDNA分子の一方の末端上のプライマー部位からの一方の鎖を配列決定することができ、その後、異なるプライマー部位からのその分子の他方の末端を配列決定する前に逆相補体鎖を生成することを必要とする。この技術的な課題としては、相補的鎖生成のプロセスが挙げられ、これが、全てのプラットフォームがこのペアードエンド配列決定とは容易には適合しない理由である。

10

【0138】

しかしながら、適応したDNA分子の2つの異なる部分を配列する能力は、限定的な程度までは、相補的鎖を生成する必要性なく達成することができる。これは、配列決定読み取りがDNA分子及び第1のアダプターから離れて進行し、それにより第2のアダプター自体の配列決定読み取りを生成するように、第1のアダプターに対してDNA分子の反対の末端に結合した、第2のアダプター内に含有される第2のプライマー部位を使用することによって達成することができる。いくつかの状況において、そのような能力は、望ましくあり得る。例えば、二本鎖配列決定に必要とされるSMI配列及びSDE配列は、達成され得る本質的に限定された読み取り長さの一部分を消費するため、これらの要素を、第2のより短い読み取り中に読み取られる反対のアダプターに移動させることができると、最大読み取り長さが必要とされる場合に役立ち得る。類似の利益は、試料多重化にしばしば使用されるインデックスバーコード配列を再配置することによって実現することができる。

20

【0139】

このプロセスを可能にするために、2つの異なるアダプターが使用され得る。図6Aに示される第1のものは、その相補体P'の反対に単純なプライマー部位Pを含有する。

30

【0140】

図6Bに示される他のアダプター配列は、Y形尾部を有さず、二本鎖配列決定に必要な特徴SMI及びSDEを含有する。この二本鎖アダプターは、本明細書に記載され、かつSMI及びSDEが別個の配列要素であるか、非対合のSMIと同一の配列要素へと組み合わせられるか、またはSDEが修飾塩基で構成される、設計のうちのいずれであってもよい。

【0141】

図6Bに示される実施例において、SDEは、縮重または半縮重SMI配列に隣接するミスマッチ配列X及びYを伴う。相補体O'を有するPCRプライマー部位Oは、アダプターの非連結末端上にある。連結可能な末端に隣接するが、アニールされたプライマーがDNA断片の向かってではなくアダプター分子自体内へと延長するように配向される、相補体P2'を有する第2のプライマー部位P2が、このアダプター設計に特有である。

40

【0142】

DNAライブラリへのこのアダプター混合物の連結後、3つの異なる生成物が生成される。各アダプターのうちの1つを有する(図6Cに示される両方のプライマー部位P及びOを含有する)生成物のみがクラスター増幅及び配列決定に成功するため、反対の末端上に同一のアダプターの種類のうちの2つを有するものは無視され得る。

【0143】

PCR増幅後、「上部」鎖に由来する二本鎖生成物を図6Dに示し、「下部」鎖に由来

50

する二本鎖生成物を図 6 E に示す。

【 0 1 4 4 】

アニールされた配列決定プライマー P 1 及び P 2、ならびに各々によって配列決定され得る領域の配向を、以下に示す。これらの読み取りは、一方よりも配列決定で最も好都合に配列決定されるだろう。これは、1つの配列決定プライマーを導入し、第1の配列決定読み取りを受け、その後、第1の配列決定読み取りの完了後、第2の配列決定プライマーを導入することによって達成されるだろう。「読み取り番号2」(図6F及び図6Gに示される)が最初に実行される場合、配列決定は、分子の末端に到達し、配列決定を自己終結するまで実行され得る。「読み取り番号1」が最初に実行される場合、プライマー P 2 を添加して「読み取り番号2」を開始する前に、この配列決定反応を中断させることが必要であるだろう。これは、組み込み後に更なる延長が不可能な修飾 d N T P を導入することによって、または初期配列決定反応中に合成された鎖を、鋳型鎖から離して(熱的もしくは化学的にのいずれか)融解し、それを洗い流してから、次の配列決定プライマーを添加することによってのいずれかによって達成することができる。

10

【 0 1 4 5 】

「上部」鎖に由来する鋳型鎖を配列決定するための配置を図6Fに示し、「下部」鎖に由来する鋳型鎖を配列決定するための配置を図6Gに示す。

【 0 1 4 6 】

「上部」鎖由来鋳型からの配列決定読み取りは、図6Hに示される通りであり、「下部」鎖由来鋳型からの配列決定読み取りは、図6Iに示される通りである。

20

【 0 1 4 7 】

異なる元の鎖分子からの配列決定読み取りの対が、S D E X 標識または S D E Y 標識のいずれかを担持しているために、区別可能であることは、容易に明らかである。

【 0 1 4 8 】

非ペアードエンドプラットフォーム上での2つの読み取りの使用は、二本鎖配列決定のタグ多様性を最大化する。

上記に開示される形態の二重読み取りの使用から生じる潜在的な利点は、読み取り長さの単純な保存を超えて拡張する。Y形アダプターによるタグに基づく二本鎖配列決定についての初期の記述において、1つの S M I 配列が、適応した D N A 分子の各末端に付加された。この設計は、多様な S M I 含有アダプターの十分に大きな集団を効率的に生成して、全ての D N A 分子が特有に標識され得ることを確実にするために、特定の状況において、実用的利点を有する。

30

【 0 1 4 9 】

一説明として、完全縮重4ヌクレオチド S M I 配列が元の Y 形アダプター設計に導入され、(図1Bに示される) D N A 断片ライブラリに連結され、ペアードエンド読み取りで配列決定される場合、分子が標識され得る可能性のある方法の総数は、 $4^4 * 4^4 = 65,536$  である。完全縮重8塩基対 S M I 配列が二本鎖アダプターに組み込まれ、(図5Cに示される) シングルエンド読み取りのために D N A ライブラリに連結された場合、同一の  $65,536$  の標識の組み合わせが達成され得る。ポリメラーゼ延長法で相補的 S M I タグを生成するとき、 $65,536$  の標識を達成するこれら2つの手段は等しく実行可能であるが、個々に合成されたオリゴヌクレオチドを有するアダプタープールを生成する場合は、この限りではない。第1のシナリオにおいて、合計  $4^4 * 2 = 512$  のオリゴヌクレオチドが生成される必要があるだろう。後者のシナリオにおいて、 $4^8 * 2 = 131,072$  が生成され、個々にアニールされる必要があり、これは必要とされる財務費用及び労力を大いに増加させるだろう。

40

【 0 1 5 0 】

二本鎖配列決定のいくつかの実施形態について、S M I アダプター生成のオリゴヌクレオチド合成法が好ましく、上記に開示されるように、十分に多様な S M I 含有アダプター集団は、分子の一方の末端上の単一 S M I のみでは実用的には達成できない可能性がある。

50

## 【0151】

同一の反応の2つのステップで配列決定するために、両方のアダプター内にSMI配列が含まれることを可能にすることによって、非ペアードエンド適合プラットフォーム上での二重読み取りの上述の方法を使用して、この限界を克服することができる。これを以下に説明する。

## 【0152】

このために、各々が異なる増幅プライマー部位を有する2種類のアダプターが必要とされる。少なくとも1つはSDEを含有する必要がある。実施例において、両方が縮重または半縮重SMI配列を含有するだろう。図7Aに示されるように、第1のアダプターは、それがSMI配列(ここでは「 $\square$ 」と特定される)を更に含むことを除いて、図6Aのアダプターに類似している。図7Bに示される第2のアダプターは、図6Bに示されるアダプターに類似し、SMI配列(ここでは「 $\square$ 」と特定される)を含有する。

10

## 【0153】

当業者にとって、2つのアダプターのSMI特徴及びSDE特徴の相対的配置を相互交換して、同一の結果を達成することができることは明らかだろう。後者のアダプター内に上記に示されるSDEは、代わりに前者のアダプター内に置かれてもよい。以前に記載されているいかなる形態のSDEまたはSMIも、同等の効果をもって、この実施例において使用されるものと置換することができる。

## 【0154】

DNAライブラリへのこのアダプター混合物の連結後、各アダプターの種類のうちの1つへの結合に成功した生成物は、図7Cに示される通りである。

20

## 【0155】

PCR増幅後、「上部」鎖に由来する二本鎖生成物を図7Dに示し、「下部」鎖に由来する二本鎖生成物を図7Eに示す。

## 【0156】

以前の実施形態において記載されるように、配列決定プライマー部位P1及びP2、ならびに各々によって配列決定される領域の配向を、「上部」鎖については図7Fに示し、下部鎖については図7Gに示す。

## 【0157】

「上部」鎖由来鑄型からの配列決定読み取りは、図7Hに示される通りであり、「下部」鎖由来鑄型から得られる読み取りは、図7Iに示される通りである。

30

## 【0158】

繰り返すと、二本の鎖の生成物は、それらの異なるX及びY SDE標識のために、容易に区別可能である。二本鎖配列決定分析のために、SMI<sub>1</sub>及びSMI<sub>2</sub>の配列は、単一特定タグ配列へと組み合わせられてもよい。

## 【0159】

Y形二本鎖配列決定アダプター内の非対称SMI

いくつかの現在入手可能な配列決定プラットフォームは、クラスター増幅及び配列決定を可能にするために、DNA分子の反対の末端上に異なるプライマー部位を必要とする。これは、非対称プライマー結合部位を有するY形またはバブル形アダプターによって、または直前の3つの実施形態において説明される2つのアダプターの連結法を通して達成することができる。Illumina(登録商標)によって製造されているものなどのY形アダプターが、ペアードエンド配列決定適合プラットフォーム上では最も一般的に使用されているが、それらは、他のプラットフォーム上でも使用することができる。

40

## 【0160】

ライブラリ調製のためのY形または「バブル形」アダプターの1つの一般的利点は、理論上、全ての二重適応したDNA分子が配列決定可能であることである。しかしながら、2つの異なるアダプターを使用する方法では、生成される分子の半分のみが配列決定可能であり、それらは各アダプターの種類のうちの1つを有する一方で、生成される分子の他方の半分は同一のアダプターの2つのコピーを有するだろう。特定の状況において、例え

50

ば、入力DNAが限定的である場合、Y形アダプターのより高い変換が望ましくあり得る。

【0161】

しかしながら、上述の第1の実施形態（元来記載されている二本鎖配列決定法）において説明されるように、ペアードエンド読み取りまたは完全なリードスルーを行う能力がないと、元来記載されているY形二本鎖アダプターは、シングルエンド読み取りによる二本鎖配列決定を容易には許容しない。

【0162】

しかしながら、Y形アダプターの相補的「ステム」配列における配列決定プライマー部位の使用は、二本鎖配列決定のためのシングルエンド読み取りを可能にするが、これはアダプター配列内の別の場所に少なくとも1つのSDEによって非対称が導入される場合のみである。簡潔な説明を続ける。

【0163】

図8Aにおいて、 $i$ 及び $i i$ 配列を含む非対合のSMIを含有するY形アダプターが示される。この設計におけるこの配列はまた、SDEとしての役割も果たすだろう。3つのプライマー部位、つまり、遊離尾部上のPCRプライマーであるA及びB、ならびに配列決定プライマー部位（及びその相補体）を含むC（及びC'）が存在する。

【0164】

DNA断片へのアダプター連結後、図8Bに示される構造が生成され、2つの別個の非相補的SMIを有する2つのアダプターがいずれかの末端に添付される。

【0165】

部位A及びBと相補的なプライマーを使用するPCR増幅後、「上部」鎖に由来する二本鎖生成物は、図8Cに示される通りであり、「下部」鎖に由来する二本鎖PCR生成物は、図8Dに示される通りである。

【0166】

プライマー部位Cからの配列決定後、どちらの一本鎖の半分が配列決定されるかによって、各鎖のPCR生成物のシングルエンド読み取りから、2つの異なる種類の配列決定読み取りが得られるだろう。「上部」鎖のPCR生成物からの配列決定読み取りは、図8Eに示される通りであり、「下部」鎖のPCR生成物から得られる配列決定読み取りは、図8Fに示される通りである。

【0167】

分析中、配列決定読み取りは、SMIアダプターライブラリ合成の時点で、かつそれと組み合わせ生成されたデータベースから既知である関係に基づいて、特定のSMI配列及びそれらの対応する非相補的パートナーによってグループ化され得る。ここで、図8Gに示されるように、元の分子の対合した「上部」及び「下部」鎖配列は、分子の一末端に起源を持つ読み取りについて $i$ 及び $i i$ で、ならびに反対の末端に起源を持つ読み取りについて $i$ 及び $i i$ でタグ付けされる。

【0168】

したがって、二本鎖配列分析が実行され得る。分析は、「非相補的SMI配列を使用する鎖定義非対称の導入」と題する実施形態において上述される分析と類似している。

【0169】

図8Hに例証されるように、この種類のY形アダプターの一代替的な設計は、遊離腕の「デージェンチェーション化」だけでなく、エキソヌクレアーゼ消化または潜在的にはYの遊離腕への非特異的連結を防止するのにも有利である、閉鎖ループを含む。閉鎖「ループ」結合（矢印によってマークされる）は、従来のホスホジエステル結合によって、または任意の他の天然もしくは非天然化学リンカー基によって達成することができる。ローリングサークル型増幅産物を防止するために、PCR増幅前にしばしば望ましくあるように、この結合は、連結の実行後に「開放」末端を達成するために、化学的または酵素的に切断可能であり得る。あるいは、この結合部位の嵩高い化学基または修飾ヌクレオチドを使用して、ポリメラーゼがループの末端を超えて横断するのを防止し、同一の目的を果たすことが

10

20

30

40

50

できる。あるいは、図 8 I に例証されるように、ループ内のヘアピン相補性領域（矢印によってマークされる）に制限エンドヌクレアーゼ認識部位が導入され、これを使用して「開放」立体構造が達成され、小ヘアピン断片が結果として放出されてもよい。

【0170】

いくつかの状況において、アダプター連結後、PCR前に追加の酵素的ステップの実行が必要とされないことが好ましい。図 8 J に例証されるように、アダプターの尾部が相補的であるが、共有結合はしていないアダプター設計は、追加のステップの必要性の不在下でも、遊離非対合の DNA 尾部によって引き起こされる問題を依然として克服することができる。

【0171】

Y形二本鎖配列決定アダプター内の非対称 SMI

Y形またはループ形アダプターにおける非対合の SMI の概念の別の変異形は、PCRプライマー部位と相補的ステムとの間の遊離一本鎖尾部領域内に位置するこれらの非対合の SMI を含む。この設計の1つの利点は、それが、選ばれた Illumina（登録商標）配列決定システム上で利用可能であるように、「二重インデックス」読み取りの一部として SMI が完全に配列決定されることを可能にすることである（Kircher et al (2012) Nucleic Acid Res. Vol. 40, No. 1, e3）。長い読み取りが特に好ましい用途では、主要配列決定読み取りに SMI が含まれないことによって、DNA 挿入物の読み取り長さが最大化されるだろう。一例を続ける。

【0172】

図 9 A は、非対合の PCR プライマー部位 A 及び B を含有する Y 形二本鎖 - 配列決定アダプターを示す。i 及び ii は、少なくとも部分的に非相補的な縮重または半縮重 SMI の対を表す。P 及び P' は、配列決定プライマー部位及びその相補体である。

【0173】

DNA 断片へのアダプター連結後、図 9 B に示される構造が生成され、それにより、2つの少なくとも部分的に非相補的な SMI を有する2つのアダプターがいずれかの末端に添付される。

【0174】

部位 A 及び B と相補的なプライマーを使用する PCR 増幅後、「上部」鎖に由来する二本鎖生成物は、図 9 C に示される通りであり、「下部」鎖に由来する二本鎖生成物は、図 9 D に示される通りである。

【0175】

一例として、二重インデックスによるペアードエンド配列決定を使用する場合、Illumina（登録商標）プラットフォーム上では、1つの配列決定読み取り及び1つのインデックス読み取りの完了後、相補的鎖が生成され得、他方の鎖の対応する配列決定及びインデックス読み取りが実行され得る。

【0176】

しかしながら、ペアードエンド配列決定または二重インデックスのいずれも、それ自体、技術としては二本鎖配列決定を許容しないことに留意されたい。所与の PCR 生成物の両方の一本鎖がともに効果的に配列決定される一方で、各 PCR 生成物は、元の DNA 二本鎖の二本の鎖のうち一方のみに由来し、故に PCR 生成物の両方の鎖の配列決定は、元の DNA 二本鎖の両方の鎖の配列決定と同等ではない。

【0177】

配列決定プライマー及びインデックスプライマー、ならびにそれらが配列決定する領域の可能性のある相対的配向を、「上部」鎖に由来する PCR 生成物からの両方向の読み取りについては図 9 E に示し、「下部」鎖に由来する PCR 生成物からの両方向の読み取りについては図 9 F に示す。

【0178】

2つの別個の読み取りではなく、単一の配列決定読み取りで、SMI 及び配列自体の両方を配列決定することもまた十分だろう。多くの異なる構成及びプライマーの数を利用し

10

20

30

40

50

て、S M I 及び読み取り配列を配列決定することができることは、明らかである。ナノポア配列決定などのいくつかの実施形態において、S M I 及び/またはD N A 配列の配列決定は、いかなる特定のプライマー部位も全く必要としない可能性がある。更に、この実施例はP C Rの使用を記載する一方で、この及び他の実施形態は、ローリングサークル型増幅及び他のアプローチを含む、当該技術分野において既知である任意の他の方法によって増幅することができる。K i r c h e r e t a l ( 2 0 1 2 ) を参照されたい。

【 0 1 7 9 】

( 図 9 G に示される ) 「上部」鎖及び「下部」鎖から得られるものに関して、4つ全ての読み取りの配列の異なるパターンを比較するとき、一方はS M I タグ  $i'$  及び  $i$  を担持し、他方はタグ及び  $i i$  及び  $i i'$  を担持するため、それらを互いに区別することができることは明らかである。この非限定的な実施例において、二本の鎖はいかなるタグも共有しないものの、 $i$  と  $i i$  との関係及び  $i'$  と  $i i'$  との関係がアダプター調製時から既知であり、故に分析の構成要素としてデータベースから検索され得るため、それらは依然として互いに関連付けられ得る。

10

【 0 1 8 0 】

二本鎖配列決定のためのプライマー部位、S M I、及びS D Eを導入するための、単一環状ベクターの使用

二本鎖配列決定に必要な全ての要素を、2つの対合したアダプターではなく、単一分子内に導入する代替的な構造が、図10に説明される。

【 0 1 8 1 】

20

この実施形態において、環状構造は、( 二本鎖配列決定に必要な要素を含む ) 直鎖状二本鎖分子の2つの末端を、適合連結部位を有するD N A断片の2つの末端と結合させることによって形成される。

【 0 1 8 2 】

図10Aにおいて、A / A' 及びB / B' は、2つの異なるプライマー部位及びそれらの逆相補体を表し、 $i$  及び  $i'$  は、縮重または半縮重S M I配列を伴い、X 及びYは、それぞれS D Eの非相補的な半分である。

【 0 1 8 3 】

図10Aの二本鎖分子への二本鎖D N A断片の連結後、図10Bに示される閉鎖ループが生成される。

30

【 0 1 8 4 】

図10Bの連結された生成物の生成後、P C Rを使用して、プライマー部位から増幅が実行される。あるいは、ローリングサークル型増幅が最初に実行されてもよい。未連結のライブラリ及びアダプターの選択的破壊が有利であり、5' - 3' または3' - 5' エキソヌクレアーゼによって達成され得る。環状設計はこれらの機会を特有に提供し、これは多くの他の設計では容易には可能ではない。

【 0 1 8 5 】

上述及び後述のS M I 及びS D Eの形態のうちのいずれも、示される形態と置換されても、それらの順序が再配置されてもよいことは、容易に明らかだろう。

【 0 1 8 6 】

40

別の実施形態の一例として、図10Cに示されるように、「非相補的S M I配列を使用する鎖定義非対称の導入」と題する実施形態において論じられるものなどの、S M I 及びS D Eの両方としての役割を果たす1つの連結部位付近の単一の要素が使用されてもよい。

【 0 1 8 7 】

あるいは、図10D及び図10Eに示されるように、S D E 及びS M Iをアダプター連結部位の各々付近の配列内に設計して、ペアードエンド配列決定を促進してもよい。

【 0 1 8 8 】

この設計において、対応する配列( すなわち、 $i$  及び  $i i$  ) 間の関係が既知であり、分析中にデータベースで検索され得る限り、反対の鎖上のS M I配列が( 図10Eに示

50

されるように)相補的であることは必須ではないことに留意されたい。

【0189】

非対称化学標識及び鎖単離を通した二本鎖配列決定

上記に論じられるように、二本鎖配列決定は基本的に、DNA二本鎖の両方の鎖の区別が可能であるような、それらの配列決定に依存する。二本鎖配列決定の元来記載されている一実施形態(WO2013/142389A1)において、両方の鎖をヘアピン配列とともに結合して、対合した鎖とともに配列決定することができる。開示に開示される実施形態においてだけでなく、WO2013/142389A1も、特有なDNA二本鎖の二本の鎖が、DNAタグ付けを使用して区別され得る方法を記載する。この後者のアプローチは、特有なDNA配列(DNA断片の一末端もしくは両末端の座標を含む内在性SMI、または縮重もしくは半縮重配列を含む外来性SMI)で各DNA分子を標識付けし、少なくとも1つの形態のSDE(例えば、ペアードエンド読み取りによる非対称プライマー部位、「バブル」配列、非相補的SMI配列、及び天然または化学的のいずれかでミスマッチへと変換される非標準ヌクレオチド)を通して鎖定義非対称を導入することを伴う。

10

【0190】

二本の鎖が、独立した反応において配列決定のために物理的に分離され得るように、二本鎖内にそれらの非対称化学標識を含む、二本鎖配列決定を実行するための別のアプローチが、以下に開示される。この一例を続ける。

【0191】

図11Aに示されるように、2つの異なるアダプターが使用される。第1のアダプターは、相補体P'を有するプライマー部位P、及び相補体'を有するSMI配列を含有する。第1のアダプターの一方の鎖は、他方のDNA鎖にはできないように、既知の物質(例えば、固体表面、ビーズ、固定された構造、及び結合パートナー)に結合するか、またはそれによって結合されることが出来る化学タグを更に担持する。図11Aに示されるように、化学タグは、ストレプトアビジンの結合パートナー及びそれに対する親和性を有するビオチンである。

20

【0192】

好ましくは小分子、ペプチド、または任意の他の特有に結合可能な部分の形態の、当該技術分野において既知である他の結合パートナー対が使用されてもよい。この標識はまた、核酸配列(例えば、DNA、RNA、またはこれらの組み合わせ、及びペプチド-核酸もしくはロックド核酸などの修飾核酸)の形態、好ましくは一本鎖の形態であってもよく、固体基質(例えば、固体表面、ビーズ、または類似する固定された構造)に添付される実質的に相補的な「ベイト」配列を使用して、アダプター-連結分子の一方の鎖に結合し、それを選択的に捕捉し、他方の鎖から単離してもよい。

30

【0193】

この非限定的な実施例において、第2のアダプターは化学タグを担持しない。図11Bに示されるように、第2のアダプターは、相補体O'を有する異なるプライマー部位Oを有する。

【0194】

図11A及び図11BのアダプターがDNA断片に連結した後、図11Cに示される(好ましい)構造が生成される。

40

【0195】

加えて、2つの他の種類の構造が生成され、1つの構造はアダプターを含有する2つのプライマー部位Pを有し、別の構造はアダプターを含有する2つのプライマー部位Oに連結される。「DNA分子の反対の末端上に異なるプライマー部位を導入するための、二本鎖配列決定アダプター設計の組み合わせの使用」と題する実施形態において上記に論じられるように、他の2種類の構造に対する好ましい構造の富化は、他の2種類の構造が無視され得るような配列決定前の特定の増幅条件をもって、通例達成することができる。

【0196】

図11Dに示されるように、連結後、DNA鎖は、熱融解または化学融解で分けること

50

ができ、その後、特定の結合パートナーに対する選択的親和性を有する化学タグを有する鎖（この場合、例えば、常磁性ビーズに結合するストレプトアビジン）を他方の鎖から分離することができる。今では分離された2つの鎖は、任意で2つの分離された鎖が独立して増幅される前述のステップをもって、独立して配列決定され得る（配列決定は、物理的に異なる反応において、または例えば、標識PCRプライマーで各々に異なるインデックスを適用し、組み換えた後、同一の反応において生じ得る）。

## 【0197】

あるいは、両方の鎖は、2つの異なる種類のベイトに対する親和性を有する異なる化学タグで標識されてもよい。その後、一方の配列決定反応またはインデックス群に見出されるタグを、他方の集団の対応するタグと比較し、二本鎖配列決定を実行することができる。この実施例において、SDEが依然として使用されるが、それは、鎖を物理的に分離するために使用され得る非対称的に添付された化学タグを伴う。それらの物理的に異なる区画化は、二本の鎖が個々に配列決定されるか、またはプール前に後続する差次的標識ステップ（例えば、それらの尾部に異なるインデックス配列を担持するプライマーによるPCR）、及び後に情報学的にデコンヴォリューションすることができる組み合わせ配列決定を受けかのいずれかを可能にする。

10

## 【0198】

この概念の別の実施形態は、化学的親和性以外の手段による鎖分離を可能にする、他の特性を有する標識（すなわち、物理的基）を使用することだろう。例として、強い正電荷を有する分子（例えば、電荷特性を有する物理的基）を含む核酸鎖は、（例えば、電気泳動法による）電場の適用を通して、その対合した未標識の対合した鎖から優先的に分離され得、または強い磁気能力を有する分子（例えば、磁気特性を有する物理的基）を含む核酸鎖は、磁場の適用を通して、その対合した未標識の対合した鎖から優先的に分離され得る。沈殿に対して感受性である化学基（例えば、不溶性特性を有する物理的基）を含む核酸鎖は、DNA自体は可溶性であるが、物理的基を含むDNAが不溶性であるような、特定の適用される条件下で溶液中にあるときに、その対合した未標識の対合した鎖から優先的に分離され得る。

20

## 【0199】

（連結されたアダプター配列決定装置内の外来性タグとして、またはDNA断片の特有な剪断点を含む内在性SMIとしてのいずれかで）SMIが適用された後の、対合した鎖の物理的分離の概念についての更に別の変異形は、DNA二本鎖をそれらの構成要素の一本鎖へと熱融解または化学融解した後に、希釈を使用することである。精製可能な化学標識を一方の鎖に適用して、それを他方の鎖から分離する代わりに、二本の元々対合した鎖が同一の容器を共有する可能性が小さいように、一本鎖を複数（すなわち、2つ以上）の物理的に分離された反応チャンバ内に希釈する。例えば、混合物が100個の容器に分けられる場合、偶然に、約1%のパートナー鎖のみが同一の容器内に置かれるだろう。容器は、容器、試験管、もしくはマイクロウェルプレート内のウェル、または物理的に分離された非連通液滴（例えば、疎水性相内水相エマルジョン）などの物理的容器の組を伴う。核酸分子を含有する、2つ以上の空間的に別個の体積の液体または固体の内容物において、核酸分子な実質的の混合が防止される、任意の他の方法が使用され得る。各容器内で、好ましくは各々内に異なるタグ配列を担持するプライマーを使用して、PCR増幅が実行されてもよい。異なるプライマーによって全ての容器内に添加されるこの特有なタグ配列は、それが配列決定インデックス読み取り中に記録され得る場所に最も好都合に位置するだろう（例えば、図9Eを参照されたい）。これらの標識は、SDEとしての役割を果たす。この実施例において、同一のSMI標識を担持する約99%のパートナー鎖に、それらのパートナー鎖とは異なるSDE標識が割り当てられるだろう。約1%のみに、同一の標識が割り当てられるだろう。二本鎖配列決定分析及びコンセンサス作製は、SMI及びこれらのSDEを使用して、通常通り進行し得る。パートナー鎖が偶然に同一のSDEを取得する少数の場合において、これらの分子は、二本鎖分析中には本質的に無視され、偽変異には寄与しないだろう。

30

40

50

## 【0200】

## ニック翻訳中のSDEの導入

Ion Torrent (商標) プラットフォームのアダプター連結に使用される市販のキットにおけるものなどのいくつかの状況において、二本鎖アダプターは、配列決定される二本鎖標的DNA分子に連結される。しかしながら、ここでは標的DNA分子の二本の鎖のうち的一方のみが、アダプターに連結される。この一般的な実施形態は、連結ドメインの5'鎖がリン酸化されていない場合である。その後、鎖置換活性を有するポリメラーゼを使用して、一般的に「ニック翻訳」として知られるプロセスにおいて、連結された鎖からの配列を未連結の鎖にコピーする。本明細書に開示されるアダプター設計が、この方法で、かつ修飾なしで使用されるとき、多くの場合、SDEはニック翻訳ステップ中に失われ、それにより二本鎖配列決定を防止するだろう。これを以下に例証する。

10

## 【0201】

1種類の二本鎖配列決定アダプターが、図12Aに示される。N'は、縮重または半縮重SMI配列を表し、GGの反対のTTは、非相補的SDE領域であり、アスタリスクは、連結不可能な脱リン酸化5'塩基を表す。

## 【0202】

二本鎖DNA分子への図12Aのアダプターの連結後、図12Bに示されるように、1つの未連結のニックが残る。

## 【0203】

標準的な「ニック翻訳」アプローチでは、鎖置換ポリメラーゼを使用してライブラリDNA分子の3'末端を延長し、アダプターの未連結の鎖を置換する。これを図12Cに示す。延長後、図12Dに示されるように、非相補的SDEは失われる。SDEが失われると、鎖が区別不可能となるため、二本鎖配列決定は生じることができない。

20

## 【0204】

アダプター連結のニック翻訳法の使用を可能にし、SDEを保持する1つのアプローチは、以下の通りである。

## 【0205】

縮重または半縮重SMI配列を含むように修飾されているIon Torrent (商標) アダプター「A」の一例が、図12Eに示される。いかなるSDEも存在しないことに留意されたい。「A」は、プライマー部位である。アスタリスクは、非脱リン酸化5'塩基を表す。Ion Torrent (商標) P1プライマーの一例が、図12Fに示される。P1は、プライマー部位を表す。アスタリスクは、脱リン酸化5'塩基を示す。

30

## 【0206】

二本鎖DNAへの図12E及び図12Fの各アダプターの連結後、図12Gの構造が作製される。2つのP1または2つのAプライマー部位を有する生成物はクラスター増幅しないため、それらは示されない。明確さのために、非連結アダプター鎖も示されない。

## 【0207】

次に、典型的なニック翻訳プロトコルに従って、鎖置換ポリメラーゼ(例えば、その強い鎖置換活性のためにいくつかの市販のキットにおいて使用される、Bstポリメラーゼ)が添加される。しかしながら、図12Hに示されるように、4つのdNTPのうちのみ(この実施例ではdGTP)が最初に添加され、故にT-dGTP誤組み込みが生じるだろう(注目すべきことに、この誤組み込み事象は、適切な反応条件下でいくつかのDNAポリメラーゼについて生じさせられ得る。例えば、McCulloch and Kunkel, Cell Research 18: 148-161 (2008)及びその中で引用される参考文献を参照されたい)。

40

## 【0208】

ミスマッチ組み込みは、特定の条件下でかなり効率的であり得る一方で、ミスマッチ延長及び第2のミスマッチの作製はかなり非効率的である(McCulloch and Kunkel, 2008)。故に、適切な条件をもって、ヌクレオチド組み込みはミスマッチが生じた後に停止するだろう。この時点で、ポリメラーゼが4つ全てのdNTPに対

50

するアクセスを有するように、残りの3つのdNTPが添加され得る。「上部」鎖の増幅生成物が「下部」鎖の増幅生成物から区別可能であるような非相補的位置を有する、図12Iに示される構造を作製するために、残りのアダプター配列がコピーされるだろう。

【0209】

PCR後、元の「上部」鎖から生じる生成物は、図12Jに示される通りであり、元の「下部」鎖から生じるPCR生成物は、図12Kに示される通りである。

【0210】

「上部」鎖生成物の配列決定は、図12Lに示される構造をもたらし、「下部」鎖生成物の配列決定は、図12Mに示される構造をもたらし得るだろう。

【0211】

配列決定生成物は、導入されるミスマッチに基づいて、互いに区別することができることに留意されたい。

【0212】

この概念をIon Torrent (商標) アダプターで実施する具体的な一例を、以下に示す。

【0213】

Ion Torrent (商標) アダプターは、以下の配列を使用し得る。

アダプターP1

【化9】

5' CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT 3' (配列番号15)

3' T\*T\*GGTGATGCGGAGGCCGAAAGGAGAGATACCCGTCAGCCACTA 5' (配列番号16)

アダプターA

【化10】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG 3' (配列番号17)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC 5' (配列番号18)

アスタリスク「\*」は、ホスホロチオエート結合を表す。

【0214】

アダプターAの配列は、以下のように修飾され得る。NNNNは、縮重または半縮重SMI配列(4つのヌクレオチドが示されるが、この配列の長さは任意である)を示し、MMMMは、NNNNの相補体を示す。以前に記載されているように、二本鎖配列決定は、SMI配列なしで実行することができるが、ここでは、SMIは、この概念を二本鎖分子タグ付けによって適用する具体的な一例として示される。

【0215】

修飾アダプターA

【化11】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG NNNN AAC 3' (配列番号19)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC MMMM TTG 5' (配列番号20)

【0216】

アダプターA及びP1は、配列決定されるDNA分子の反対の末端に結合する。単純化のため、分子のアダプターA末端のみが示され、同様に単純化のため、二本の鎖はそれぞれX'及びY'として示される。配列決定される断片の長さが使用される配列決定プロセスと適合する限り、任意の長さの任意のDNA配列を使用することができる。

【0217】

「上部」鎖は連結されるが、「下部」鎖は連結されず、ニック(|として示される)を残す

10

20

30

40

50

## 【化 1 2】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3' (配列番号 2 1)  
 3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTCMMMMTTG|YYYYYYYYY 5' (配列番号 2 2)

## 【0 2 1 8】

鎖置換ポリメラーゼが d G T P とともに添加される。G は、5' - 3' 方向において遭遇される第 1 の位置 (C の反対の G の正確な組み込み)、及び遭遇される第 2 の位置 (A の反対の G の不正確な組み込み) に組み込まれる。ミスマッチ後の不正確な塩基の延長は非効率的であるため、ポリメラーゼ濃度、反応時間、及び緩衝剤条件の適切な条件下で、ポリメラーゼは失速し、更なる組み込みは生じない。「下部」アダプター鎖の最初 2 つのヌクレオチドはこの反応中に置換され、以下の略図にアダプター - DNA 構築物に以下に示されることに留意されたい。新たに組み込まれた塩基を太字で示す。

## 【化 1 3】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3'  
 3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTCMMMMT**GG**YYYYYYYYY 5'  
 \ TG

(上：配列番号 2 3 及び下：配列番号 2 4)

## 【0 2 1 9】

ここで、4 つ全てのヌクレオチドがポリメラーゼにとって利用可能であるように、d C T P、d A T P、及び d T T P が反応に添加される。鎖置換合成が進行し得、中間体生成物が説明目的で以下に示される。

## 【化 1 4】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3'  
 3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCT**GAGTCMMMMTGG**YYYYYYYYY 5'  
 \ GCTGAGTCMMMMTTG

(上：配列番号 2 5、中間：配列番号 2 6、及び下：配列番号 2 7)

## 【0 2 2 0】

鋳型の末端に到達した後、アダプターの元の「下部」鎖は完全に置換され (示さず)、相補的ではない単一の塩基対 (A : G 塩基対、下線) を有する、完全に合成された「下部」鎖が存在する。

## 【化 1 5】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNAACXXXXXXXXX 3' (配列番号 2 8)  
 3' **GGTAGAGTAGGGACGCACAGAGGCTGAGTCMMMMTGG**YYYYYYYYY 5' (配列番号 2 9)

## 【0 2 2 1】

その後、この構築物を、典型的な Ion Torrent (商標) プロトコルに従って、PCR 増幅及び配列決定に使用することができる。注目すべきことに、PCR 増幅は、「上部」鎖及び「下部」鎖の両方に由来する生成物をもたらす、これらの生成物は、ニック翻訳中に導入される非相補的塩基対のために、互いに区別することができる。

## 【0 2 2 2】

「上部」鎖から生じる生成物は、以下の形態のものであるだろう (塩基ミスマッチの位置に下線を引く)。

## 【化 1 6】

TCAGNNNNAACXXXXXXXXX (配列番号 3 0)

10

20

30

40

50

## 【0223】

対照的に、「下部」鎖から生じる生成物は、以下の形態のものであるだろう（塩基ミスマッチの位置に下線を引く）。

## 【化17】

TCAGNNNNACCXXXXXXXXXX (配列番号31)

## 【0224】

「下部」鎖生成物は、アダプター - 連結されたDNAの「下部」鎖内に最初に存在した配列の逆相補体である（故にニック翻訳中に導入される塩基誤挿入であるGヌクレオチドは、配列決定中にCヌクレオチドとして読み取られる）ことに留意されたい。

10

## 【0225】

ここで、エラー訂正のために、二本の鎖の各々から生じる増幅重複を互いに比較することができる。二本鎖DNAの所与の分子から生じる「上部鎖」生成物は、タグ配列NNNNAAACを有するだろう。対照的に、「下部」鎖生成物は、タグ配列NNNNAAACを有するだろう。故に、以前に記載されているように、二本の鎖からの重複は、エラー訂正の目的で分解され得る（Schmitt et al, PNAS 2012）。

## 【0226】

ニック翻訳後のミスマッチの導入

上記に対する代替的なアプローチは、存在する4つ全てのヌクレオチドについてニック翻訳を完了し、その後、鋳型鎖内の塩基を異なる塩基に変化させることだろう。

20

## 【0227】

プライマー配列及びその相補体（P/P'）、U-A塩基対（U=ウラシル）、ならびに一本鎖SMI配列及びその相補体（ / '）を含有するアダプターを図13Aに示し、アスタリスクは、脱リン酸化5'末端を表す。

## 【0228】

図13Aのアダプターが、配列決定される二本鎖DNA分子に連結された後、図13Bに示されるように、単一鎖ニックが脱リン酸化部位に残る。ここでは、「上部」鎖は、標的DNA分子内の5'リン酸のために連結するが、「下部」鎖は、アダプター内の5'リン酸の欠如のために標的DNAには連結せず、ニックを残す。

## 【0229】

鎖置換合成は、ポリメラーゼ（例えば、Bstポリメラーゼ）及び4つ全てのdNTPで実行することができ、図13Cに示される構造をもたらす。

30

## 【0230】

結果として得られる延長された生成物は、元のアダプター内で現れたように、ここで再度現れる。図13Dに示されるように、非対称のいかなる部位もまだ存在しない。

## 【0231】

ポリメラーゼ及びdNTPを除去するために、精製ステップが実行され得る。その後、ウラシルDNAグリコシラーゼ及び適切なAPエンドヌクレアーゼを添加することによって、（図13Dに示される構造の）「上部」鎖からウラシルを除去し、図13Eに示される単一ヌクレオチドギャップをもたらすことができる。

40

## 【0232】

次に、非鎖置換ポリメラーゼ（例えば、高度にエラーを受け易く、塩基誤組み込みを促進するスルホロプスDNAポリメラーゼIV）が、単一ヌクレオチド（例えば、他のヌクレオチドではなくdGTP）とともに添加される。この実施例において、これは、Aの反対のGの誤組み込みをもたらすだろう。結果として得られるニックはDNAリガーゼで封鎖され、図13Fに示されるアダプター内にミスマッチを有する生成物をもたらす得る。

## 【0233】

図13Gに示されるように、増幅及び配列決定後、「上部」鎖から生じる生成物は、同一のSMI配列を担持する配列決定読み取り上にGまたはT塩基のいずれかを有しているために、「下部」鎖から生じる生成物から区別可能である。

50

## 【0234】

この実施例は、G - A ミスマッチの作製によって説明されるが、分子内のいかなる位置の1つ以上の塩基のいかなる他のミスマッチも同一の効果を有することは、明らかだろう

## 【0235】

この概念を Ion Torrent (商標) プラットフォームに適用する具体的な一例を、以下に示す。

## 【0236】

太字の標準配列への付加 (U = ウラシル) を有する以下の「修飾アダプター A」を考慮する。

## 【化18】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG U NNNN C 3' (配列番号32)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTC A MMM G 5' (配列番号33)

## 【0237】

アダプターは上記のように標的 DNA 分子に連結され、ニックの位置は「|」として示される。

## 【化19】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGUNNNNCXXXXXXXXX 3' (配列番号34)

3' T\*T\*GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMG|YYYYYYYYY 5' (配列番号35)

## 【0238】

ここで、4つ全ての dNTP の存在下で鎖置換ポリメラーゼを使用して、アダプターの「下部鎖」の完全な鎖置換を可能にする (新たに組み込まれた塩基は太字、元の下部アダプター鎖は置換されており、示されない)。

## 【化20】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGUNNNNCXXXXXXXXX 3' (配列番号36)

3' **GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMG**YYYYYYYYY 5' (配列番号37)

## 【0239】

生成物を精製して、dNTP を除去し、その後、ウラシル DNA グリコシラーゼ及び A P エンドヌクレアーゼを添加して、「上部」鎖からウラシルを除去し、単一ヌクレオチドギャップを残す。

## 【化21】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAG NNNNCXXXXXXXXX 3' (配列番号38)

3' **GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMG**YYYYYYYYY 5' (配列番号39)

## 【0240】

次に、非鎖置換エラープロンポリメラーゼ (例えば、スルホロブス DNA ポリメラーゼ IV) を dGTP とともに添加し、これは、単一ヌクレオチドギャップの A の反対の G の組み込みをもたらす。その後、リガーゼを添加して、「上部」鎖に無傷アダプター - DNA 生成物をもたらす得る。これは、非相補的塩基対 (下線を引いた位置) をもたらす。

## 【化22】

5' CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNCXXXXXXXXX 3' (配列番号40)

3' **GGTAGAGTAGGGACGCACAGAGGCTGAGTCAMMMG**YYYYYYYYY 5' (配列番号41)

## 【0241】

この生成物は、直前の実施形態に記載される方法に類似する方法で、エラー訂正に使用することができる。

## 【0242】

ニック翻訳後のミスマッチの導入

10

20

30

40

50

「ニック翻訳中のSDEの導入」と題する実施形態は、アダプター配列内でニック翻訳中に非対称SDEがいかにかに導入され得るかを示した。可能性としてはアダプターが添加される前にすら、非対称部位(SDE)がライブラリ分子内に組み込まれるように、同一の原理がDNA分子ライブラリ自体に適用され得る。これは、様々な方法で達成することができる。以下は単なる一例に過ぎない。

【0243】

「上部」鎖及び「下部」鎖を有する二本鎖DNA分子を、図14Aに示す。DNA分子は、ライブラリ調製のために、様々な方法で断片化され得る。血漿中の無細胞DNAなどのいくつかのDNA源は既に小片であり、いかなる別個の断片化ステップも必要とされない。音響剪断は、しばしば使用される方法である。半ランダム酵素的剪断法が使用されてもよい。定義された認識部位で切断する非ランダムエンドヌクレアーゼは、別の方法である。この実施例において、5'オーバーハングを残すエンドヌクレアーゼを使用して、図14Bに示されるように、同様に5'オーバーハングした断片のライブラリを作製する。

10

【0244】

この非対称状態は、ポリメラーゼによってコピーされる第1のヌクレオチドと相補的ではない単一ヌクレオチドのみの存在下でポリメラーゼを使用することによって、配列非対称へと変換され得る。この実施例において、dGTPが使用され、これはT-dGTP誤組み込みをもたらすだろう(そのような誤組み込みは、適切な反応条件下でいくつかのDNAポリメラーゼについて生じさせられ得る。McCulloch and Kunkel, Cell Research 18:148-161(2008)及びその中で引用される参考文献を参照されたい)。2つのミスマッチを有する部分的二本鎖DNA分子を、図14Cに示す。

20

【0245】

次に、4つ全てのヌクレオチドが反応に添加され、コピーはDNA分子が二本鎖になるまでDNA分子の末端を延長し続ける。各断片末端上にミスマッチバブルが生成され、図14Dに示される2つのSDEを形成する。

【0246】

その後、二本鎖配列決定アダプターがDNA分子に連結され得る。図14Eに示される例示的なアダプターは、相補体P'を有するプライマー部位P、相補体O'を有する異なるプライマー部位O、及び相補体'を有する縮重または半縮重SMIを有する。

30

【0247】

図14Dの二本鎖DNA分子と図14Eのアダプターとの間に連結が実行されて、図14Fの構造が生成される。以前の実施形態において論じられるように、同一のアダプター配列のうちの2つに連結される生成物は無視され得るが、これは、それらが適切な条件下では増幅しないためである。

【0248】

PCR後、「上部」鎖に由来する生成物は、図14Gに示される通りであり、「下部」鎖に由来する生成物は、図14Hに示される通りである。

【0249】

プライマーPを使用する配列決定は、図14Iに示されるそれぞれの鎖から以下の配列をもたらすだろう。

40

【0250】

SMI配列後のC対Tの存在は、「上部」鎖の読み取りと、「下部」鎖から得られる読み取りとの区別を可能にすることに留意されたい。

【0251】

類似するSDE標識は、変異原性ヌクレオチド類似体を使用して3'陥凹末端ギャップを充填すること、または他の方法によって、同様に達成することができる。

【0252】

他の剪断方法が使用され、SDEを作製するように、充填前にエキソヌクレアーゼによって3'陥凹末端が作製されてもよい。

50

## 【 0 2 5 3 】

広義には、この実施例は、アダプター自体とは独立した方法で、S D E が導入され得ることを説明する。二本鎖配列決定が生じるためには、各最終適応した分子内のいくつかの形態のみの S M I 及び S D E が、二本鎖の各鎖から得られた配列が互いに遡って関連付けられ、しかし同時に互いに決定的に区別されることを可能にする。これらの要素は、上記に考慮されるように様々な形態を取り、アダプター連結前、連結中、または連結後に導入され得る。

## 【 0 2 5 4 】

二本鎖配列決定に適切な分子の組み立ての変異形

上記に開示される実施形態は、二本鎖配列決定のための改善された方法を説明し、組み立てられる最終分子は、少なくとも1つの鎖定義要素 ( S D E ) 及び少なくとも1つの単一分子識別子 ( S M I ) 配列を含み、S D E 及び S M I の両方は、配列決定される DNA の二本鎖または部分二本鎖分子に結合する。しかしながら、S M I 及び S D E は単一アダプター内に含まれている必要はなく、それらは単純に、理想的には任意の増幅及び/もしくは配列決定ステップ前またはステップ中に、最終分子内に存在する必要がある。

10

## 【 0 2 5 5 】

例えば、S D E は、図 4 D に示されるように、酵素反応を介した連結後にアダプター内に作製されてもよい。同様に、( W O 2 0 1 3 / 1 4 2 3 8 9 A 1 に ) 元来記載されているように、いくつかの実施形態において、個々の DNA ライブラリ断片の剪断点の特定の配列が、アダプター内に含まれる外来性 S M I の添加の必要性なく、内在性 S M I 配列としての役割を果たし得る。断片が参照ゲノムに整列されるとき、「剪断点」は、DNA 断片のいずれかの末端のマッピング座標と見なされ得る。いずれか一方の末端または両末端の座標を「内在性 S M I」として使用して、単独、または1つ以上の外来性 S M I 配列の配列との組み合わせのいずれかで、別個の DNA 分子を互いに区別することができる。

20

## 【 0 2 5 6 】

以下の一覧表は、そのようなアダプターの非限定的な変異形を含む。

## 【 0 2 5 7 】

- - S D E は両方の鎖内に存在するが、S M I 及びプライマー結合部位は一方のアダプター鎖内のみ存在する。その後、これらの要素は、ポリメラーゼによって他の鎖にコピーされる。

30

## 【 0 2 5 8 】

- - いかなる S D E も存在しない。S M I 及びプライマー結合部位は、一本の鎖内のみにある。存在する1つの不正確な d N T P のみについてポリメラーゼを使用して、S D E を作製し、その後、残りの d N T P を添加して、ポリメラーゼによる S M I 及びプライマー結合部位の二本鎖化を可能にする。

## 【 0 2 5 9 】

- - 連結ドメインは、( 第 2 のアダプター鎖が結合しないように ) 一方のアダプター鎖内にしか存在しない。その後、新たな第 2 のアダプター鎖が、ポリメラーゼによって第 1 のアダプター鎖からコピーされる。これが、S M I 及びプライマー結合ドメインを作製する。上記の通り、1つの不正確な d N T P のみを最初に添加して、S D E を作製し、その後、残りの d N T P を添加する。このアプローチを、上記に開示される実施形態に示す。

40

## 【 0 2 6 0 】

- - 連結ドメインは、( 第 2 のアダプター鎖が結合しないように ) 一方のアダプター鎖内のみ存在し、このアダプター鎖は、ウラシルを含む。その後、新たな第 2 のアダプター鎖が、4つ全てのヌクレオチドについて、ポリメラーゼによって第 1 のアダプター鎖からコピーされる。その後、元のアダプター鎖内のウラシル塩基を、ウラシル DNA グリコシラーゼ及び適切な A P エンドヌクレアーゼによって酵素的に除去する。その後、存在する不正確な単一ヌクレオチドとともに DNA ポリメラーゼを使用して、DNA 内のギャップにミスマッチを挿入し、その後、ギャップを DNA リガーゼによって連結する。図 4 に関連する上記に開示される実施形態において、このアプローチをより詳細に示す。

50

## 【0261】

- - 第1の結合したアダプターは、両方の鎖内に単独でSMIドメインを有する。その後、同様に両方の鎖内にプライマー結合ドメイン及びSDEを有する第2のアダプターが、これに結合する。

## 【0262】

- - 第1の結合したアダプターは、両方の鎖内にSMI及びSDEドメインを有する。両方の鎖内にプライマー結合ドメインを有する第2のアダプターが、結合する。

## 【0263】

- - 第1の結合したアダプターは、両方の鎖内にSMIドメインを有する。その後、2つの非相補的または部分的非相補的プライマー結合ドメインを有する第2の「Yアダプター」が、結合する。

10

## 【0264】

- - 第1の結合したアダプターは、両方の鎖及び一本鎖領域内に、連結ドメインとともにSMIを有する。オリゴヌクレオチドは、一本鎖領域内にアニールされ、連結されており、ミスマッチは、SDEドメインを作製するオリゴヌクレオチド内に含まれる。

## 【0265】

- - 他の実施形態において、バブルの位置は変更されてもよく、n量体の長さは変化してもよく、n量体は完全に排除され、DNA分子の末端の切断点から、各鎖からの重複が代わりに特定されてもよい。変異形ヌクレオチドまたはヌクレオチド様分子が、DNA（例えば、ロックド核酸（LNA）及びペプチド核酸（PNA）ならびにRNA）内で使用

20

## 【0266】

本明細書に開示される変異形の各々が、本発明に含まれる。

## 【0267】

これらの変異形の各々において、同一の一般的概念が適用され、二本鎖配列決定のための最終分子は、配列決定されるDNAのセグメントに接続されたSDE及びSMIのコア要素を含む。この場合SDEとしての役割を果たす、（例えば、「Y」構成の）2つの非対称プライマー結合部位と、二本鎖DNA分子に結合したSMI配列とを含むアダプターによって二本鎖配列決定が実行される、（WO2013/142389A1における）二本鎖配列決定の初期の記述にも、同一の一般的概念が適用されることにもまた留意されたい。これらの構成要素は、必要な構成要素が、理想的には任意の増幅もしくは配列決定ステップ前またはステップ中に最終分子内に存在する限り、様々な方法で標的DNA分子上に組み立てることができる。

30

## 【0268】

二本鎖配列決定のための代替的なデータ処理スキーム

二本鎖配列決定は、2つの個々のDNA鎖の各々から生じる増幅された重複の「コンセンサス」を得て、2つの一本鎖コンセンサス配列を得、その後、結果として得られる一本鎖コンセンサスを比較して、二本鎖コンセンサス配列を得ることによって実行することができる。単一分子の増幅された重複の配列を位置毎に「平均化」するこのアプローチは、いくつかの設定（例えば、重度に損傷したDNA内の所与の位置で反復増幅エラーが生じ得る場合）においては望ましくない可能性があり、故に異なるデータ処理スキームを有するいくつかの設定において、より信頼できる結果を得ることができる。

40

## 【0269】

代替的なアプローチとしては、以下のものが挙げられる。

## 【0270】

- - 「上部」鎖及び「下部」鎖に対応する所与のタグ配列を有する分子内、1つの「上部」鎖及び1つの「下部」鎖を任意で選び、二本の鎖の配列を比較する。両方の鎖が一致する位置は維持し、一致しない位置を未定義としてマークする。結果として得られる配列読み取りを、二本鎖読み取りと呼ぶ。

## 【0271】

50

- - 同一のタグ配列を共有する、任意で選択された「上部」鎖及び「下部」鎖のためのこのプロセスを繰り返して、一連の「二本鎖読み取り」を得る。

【0272】

- - 所与のタグ配列を有する結果として得られる「二本鎖読み取り」内、例えば、参照配列と比較して、読み取り内に最少の配列変化及び/または最少の未定義位置を有する二本鎖読み取りを選択する。その後、この読み取りを、出発DNA二本鎖の正しい配列を表す可能性が最も高い読み取りと見なすことができる。

【0273】

- 実施形態において、そのようなアプローチは、後述のアルゴリズムによって特に可能となり得る。これは、説明目的の単一の例に過ぎず、多くの他のアルゴリズムを使用して、二本鎖コンセンサス読み取りを形成することができることが理解される。更に、二本鎖配列決定の具体的な一実施形態のための実施例を示すが、二本鎖配列決定の多くの他の実施形態のための類似の例を適切に準備することができる。

10

【0274】

「バブル」配列を使用して、各二本鎖の「上部」鎖がGCGCと標識されること、及び「下部」鎖がTATAと標識されること（両方の鎖は同一の単一分子識別子(SMI)配列を共有する)をもたらす、本明細書に開示される一実施形態において、以下のステップを使用することができる。

1. 実験からの全ての配列決定読み取りを含有するファイルを準備する。

2. このファイルを2つのファイル(GCGCと標識される読み取りを含有する1つのファイルは「GCGC」と呼ばれ、TATAと標識される読み取りを含有する第2のファイルは「TATA」と呼ばれる)に分ける。

20

3. 「GCGC」ファイル内の任意の読み取りを選び、そのSMIタグを読み取り、「TATA」ファイル内でマッチするSMIタグを検索する。

4. マッチが見出される場合、これらの2つの配列から新たな配列を作製する。新たな配列内では、一致する読み取り内の全ての配列位置は維持し、2つの読み取り内の全ての一致しない位置を未定義としてマークする。この新たな配列を、「二本鎖」と呼ばれるファイルに書き込み、「GCGC」ファイル及び「TATA」ファイルからの2つの配列を除去する。

マッチが見出されない場合、「GCGC」ファイルからの配列を除去し、それを「非対応」と呼ばれるファイルに書き込む。

30

5. 「GCGC」ファイルから別の任意の読み取りを選び、ステップ3~4を再度実行する。

6. 「GCGC」ファイル内にいかなる読み取りも残らなくなるまで続ける。

【0275】

結果として得られる「二本鎖」ファイル内で、マッチするSMIタグ配列を有する全ての読み取りを考慮する。場合によっては、同一のSMIタグを有する複数の「二本鎖」読み取りが存在し得る（これは、例えば、単一出発DNA分子の複数のPCR重複が原因であり得る）。これらは、以下のアプローチのうちのいずれかによって単一の二本鎖読み取りに変換することができる。

40

【0276】

- - これらの読み取り内で、参照ゲノム配列と比較して、最少のミスマッチを有する読み取りを選択し、残りの読み取りを処分する。

【0277】

- - あるいは、参照ゲノム配列と比較して、最少の未定義位置を有する読み取りを選択し、残りの読み取りを処分する。

【0278】

- - あるいは、共有されたSMIタグ配列を有する読み取り内でコンセンサスを作製して、二本鎖コンセンサス読み取りを作製する。

【0279】

50

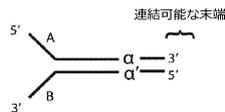
当業者にとって、上記の選択肢の組み合わせを使用して、二本鎖コンセンサ読み取りが開発され得ること、及び上述されていないいくつかの他の方法が使用され得ることは明らかだろう。

【0280】

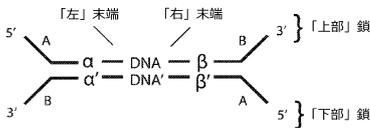
他の実施形態

本発明が、その発明を実施するための形態と組み合わせで記載されている一方で、上述の記述は、説明が意図され、本発明の範囲の限定は意図されず、本発明の範囲は、添付の特許請求の範囲によって定義される。他の態様、利点、及び修飾が、以下の特許請求の範囲内である。

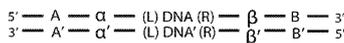
【図1A】



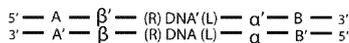
【図1B】



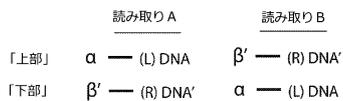
【図1C】



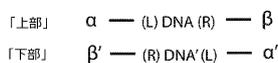
【図1D】



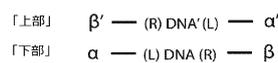
【図1E】



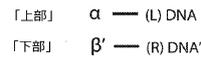
【図1F】



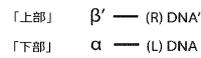
【図1G】



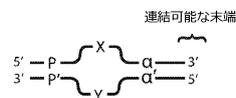
【図1H】



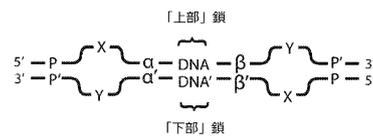
【図1I】



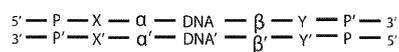
【図2A】



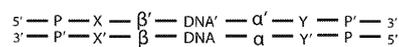
【図2B】



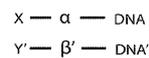
【図2C】



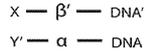
【図2D】



【図2E】



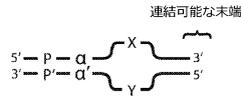
【図 2 F】



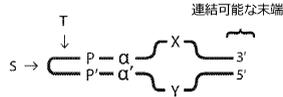
【図 2 G】



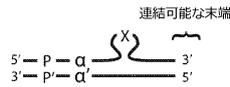
【図 2 H】



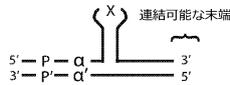
【図 2 I】



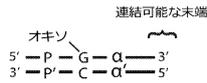
【図 2 J】



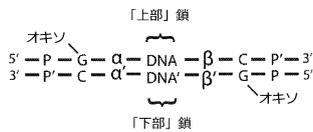
【図 2 K】



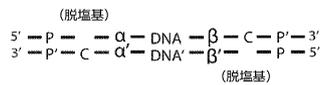
【図 4 A】



【図 4 B】



【図 4 C】



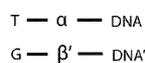
【図 4 D】



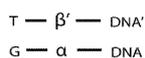
【図 4 E】



【図 4 F】



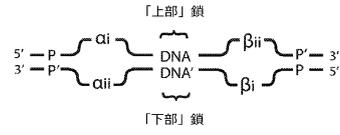
【図 4 G】



【図 3 A】



【図 3 B】



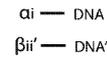
【図 3 C】



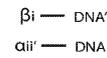
【図 3 D】



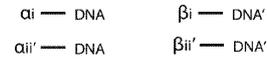
【図 3 E】



【図 3 F】



【図 3 G】



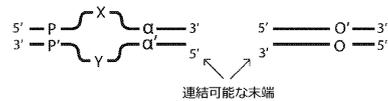
【図 4 H】



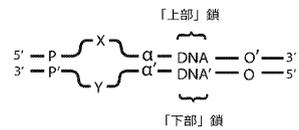
【図 5 A - 5 B】

図 5 A

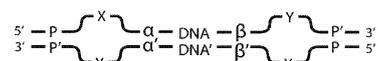
図 5 B



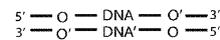
【図 5 C】



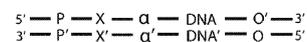
【図 5 D】



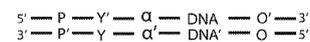
【図 5 E】



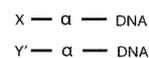
【図 5 F】



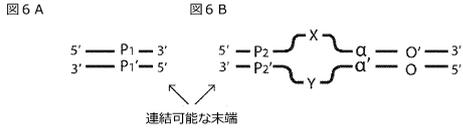
【図 5 G】



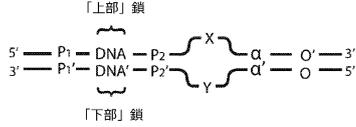
【図 5 H】



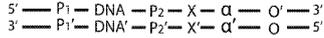
【図 6 A - 6 B】



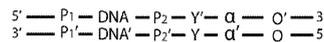
【図 6 C】



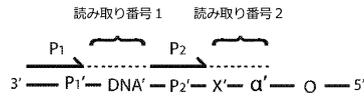
【図 6 D】



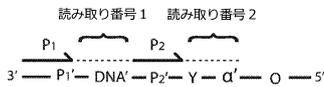
【図 6 E】



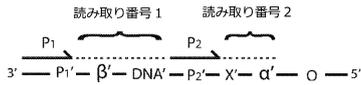
【図 6 F】



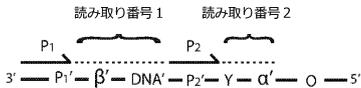
【図 6 G】



【図 7 F】



【図 7 G】



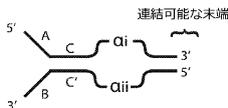
【図 7 H】

読み取り番号 1   β — DNA  
 読み取り番号 2   X — α

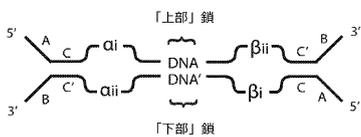
【図 7 I】

読み取り番号 1   β — DNA  
 読み取り番号 2   Y' — α

【図 8 A】



【図 8 B】



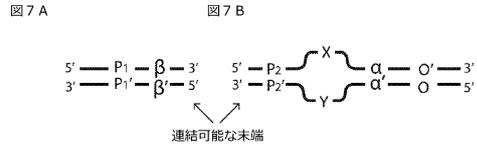
【図 6 H】

読み取り番号 1   DNA  
 読み取り番号 2   X — α

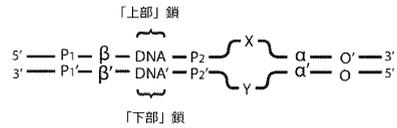
【図 6 I】

読み取り番号 1   DNA  
 読み取り番号 2   Y' — α

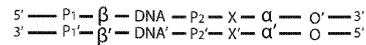
【図 7 A - 7 B】



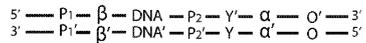
【図 7 C】



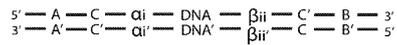
【図 7 D】



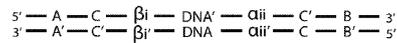
【図 7 E】



【図 8 C】



【図 8 D】



【図 8 E】

α<sub>i</sub> — DNA  
 β<sub>ii</sub>' — DNA'

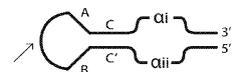
【図 8 F】

β<sub>i</sub> — DNA'  
 α<sub>ii</sub>' — DNA

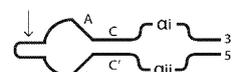
【図 8 G】

α<sub>i</sub> — DNA                      β<sub>i</sub> — DNA'  
 α<sub>ii</sub>' — DNA                      β<sub>ii</sub>' — DNA'

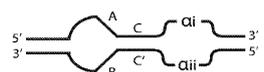
【図 8 H】



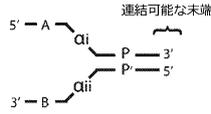
【図 8 I】



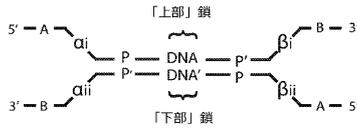
【図 8 J】



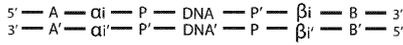
【図9A】



【図9B】



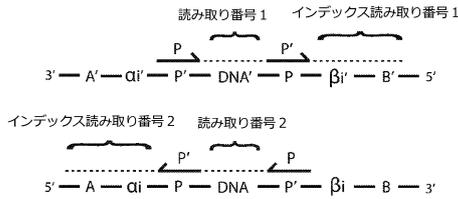
【図9C】



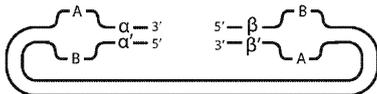
【図9D】



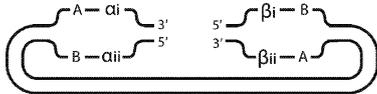
【図9E】



【図10D】



【図10E】



【図11A - 11B】

図11A

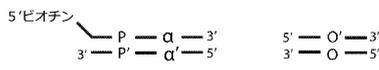
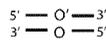
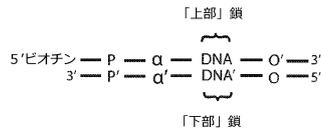


図11B

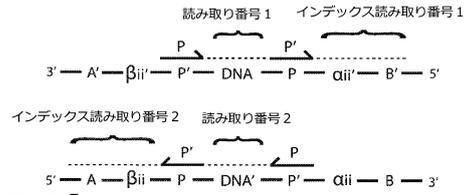


連結可能な末端

【図11C】



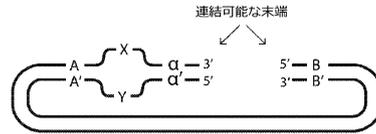
【図9F】



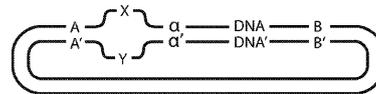
【図9G】

「上部」鎖に由来		「下部」鎖に由来	
読み取り番号1	DNA	読み取り番号1	DNA'
読み取り番号2	DNA'	読み取り番号2	DNA
インデックス番号1	βi	インデックス番号1	αii
インデックス番号2	αi'	インデックス番号2	βii'

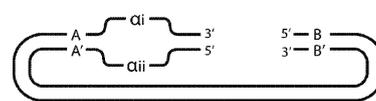
【図10A】



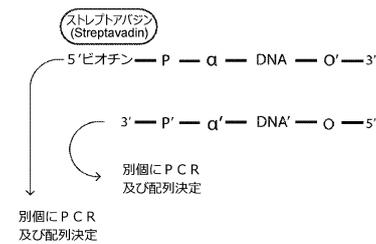
【図10B】



【図10C】



【図11D】



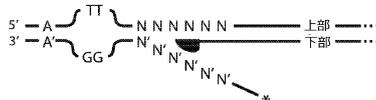
【図12A】



【図12B】



【図12C】



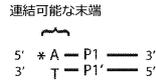
【図12D】



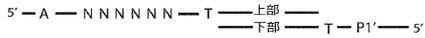
【図 1 2 E】



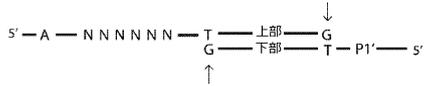
【図 1 2 F】



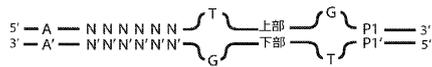
【図 1 2 G】



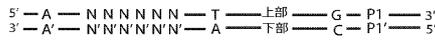
【図 1 2 H】



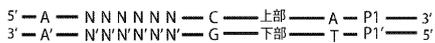
【図 1 2 I】



【図 1 2 J】



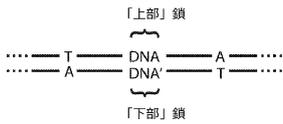
【図 1 2 K】



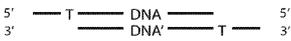
【図 1 2 L】



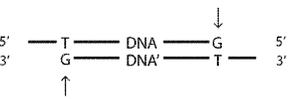
【図 1 4 A】



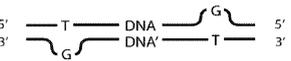
【図 1 4 B】



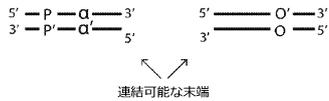
【図 1 4 C】



【図 1 4 D】



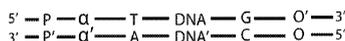
【図 1 4 E】



【図 1 4 F】



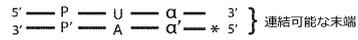
【図 1 4 G】



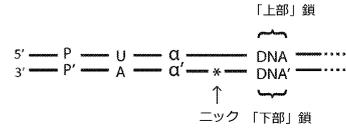
【図 1 2 M】



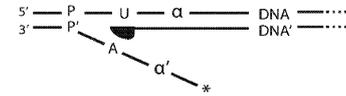
【図 1 3 A】



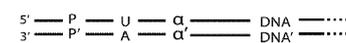
【図 1 3 B】



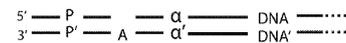
【図 1 3 C】



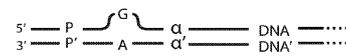
【図 1 3 D】



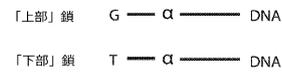
【図 1 3 E】



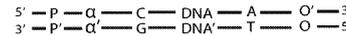
【図 1 3 F】



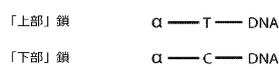
【図 1 3 G】



【図 1 4 H】



【図 1 4 I】



**【配列表】**

0006975507000001.app

---

フロントページの続き

(72)発明者 シュミット, マイケル・ダブリュー  
アメリカ合衆国、ワシントン 98112、シアトル、26ティーエイチ・アベニュー・イースト  
2027

(72)発明者 ソーク, ジェシー・ジェイ  
アメリカ合衆国、ワシントン 98125、シアトル、ダーランド・アベニュー・ノースイースト  
10704

審査官 竹内 祐樹

(56)参考文献 国際公開第2015/100427(WO, A1)  
国際公開第2013/142389(WO, A1)

(58)調査した分野(Int.Cl., DB名)  
C12Q 1/68 - 1/6897  
C12N 15/00 - 15/90  
CAplus/MEDLINE/EMBASE/BIOSIS(STN)