

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2023-553983

(P2023-553983A)

(43)公表日 令和5年12月26日(2023.12.26)

(51)国際特許分類	F I	テーマコード (参考)
C 1 2 N 15/11 (2006.01)	C 1 2 N 15/11	Z Z N A 4 B 0 6 3
C 1 2 Q 1/6869(2018.01)	C 1 2 Q 1/6869	Z

審査請求 未請求 予備審査請求 未請求 (全103頁)

(21)出願番号	特願2023-535673(P2023-535673)	(71)出願人	515025398
(86)(22)出願日	令和3年12月10日(2021.12.10)		ザ ブロード インスティテュート, イン
(85)翻訳文提出日	令和5年8月14日(2023.8.14)		コーポレーテッド
(86)国際出願番号	PCT/US2021/062966		THE BROAD INSTITUTE
(87)国際公開番号	WO2022/125997		, INC.
(87)国際公開日	令和4年6月16日(2022.6.16)		アメリカ合衆国 マサチューセッツ州 0
(31)優先権主張番号	63/208,951		2142、ケンブリッジ、メイン スト
(32)優先日	令和3年6月9日(2021.6.9)		リート 415、ルーム 7003
(33)優先権主張国・地域又は機関	米国(US)		415 Main Street, Roo
			m 7003, Cambridge, M
(31)優先権主張番号	63/124,696		assachusetts 02142
(32)優先日	令和2年12月11日(2020.12.11)		United States of Am
(33)優先権主張国・地域又は機関	米国(US)		erica
(31)優先権主張番号	63/239,920	(71)出願人	511071555
			ダナーファーバー キャンサー インステ
	最終頁に続く		最終頁に続く

(54)【発明の名称】 二重鎖シーケンシングのための方法

(57)【要約】

本開示は、二重鎖シーケンシングを改善するデュアル鎖ハイスループット次世代シーケンシングに対する強力な新しいアプローチを提供する。方法は、各DNA二重鎖の両方の鎖を直鎖の配列に連結するライブラリー構築方法でシーケンシングされるDNAフラグメント（例として、ゲノムDNAフラグメント）にライゲーションされる新規なマルチオリゴヌクレオチドアダプター構築物を提供する。両方の鎖を物理的に繋ぎ合わせることで、産物は、二重鎖コンセンサスを形成するのにそれ自体で十分である。この戦略は、最低限の追加コストにて1000倍正確なシーケンシングを提供することへの潜在的な可能性を有し、そしてGenomics Platformで提供されている既存の製品（WGS、WES、標的化されたパネル）を直接的に強化することができる。

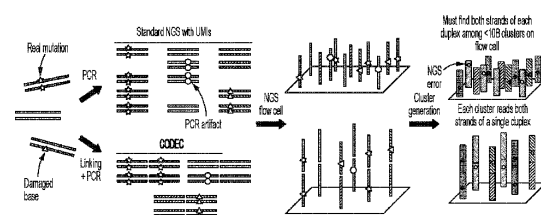


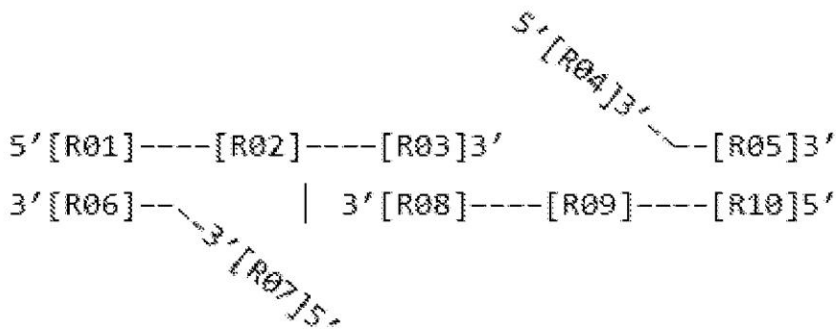
FIG. 1A

【特許請求の範囲】

【請求項 1】

以下の構成にある、少なくとも 10 個の領域 (R01~R10) を含む単離された核酸複合体 (複合体) であって：

【化 1】



10

ここで、「----」は、結合を表し、

ここで、R01、R02、およびR03は、第1のオリゴヌクレオチドを含み、

ここで、R04およびR05は、第2のオリゴヌクレオチドを含み、

ここで、R06およびR07は、第3のオリゴヌクレオチドを含み、

ここで、R08、R09、R10は、第4のオリゴヌクレオチドを含み、

ここで、R01とR06とは、相互にアニーリングされている、

ここで、R03とR08とは、相互にアニーリングされている、

ここで、R05とR10とは、相互にアニーリングされている、

ここで、R02とR07とは、相互にアニーリングされていない、および

ここで、R04とR09とは、相互にアニーリングされていない；

ここで、R02は、一本鎖リンカー、第1のユニークな分子識別子 (UMI)、および第1のリードプライマー部位を含み、および

ここで、R09は、一本鎖リンカー、第2のUMI、および第2のリードプライマー部位を含む、前記複合体。

20

【請求項 2】

30

(1) R01が、第1のアダプターを含む；

(2) R02が、一本鎖リンカー、第1のユニークな分子識別子 (UMI)、および第1のリードプライマー部位を含む；

(3) R03が、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端にあるかまたはその付近にある第1の配列を含む；

(4) R04が、第1の次世代シーケンシング (NGS) アダプター配列を含むフリーの5'末端を含む；

(5) R05が、第3のアダプターおよび第1のサンプルインデックスを含む；

(6) R06が、第2のアダプターおよび第2のサンプルインデックスを含む；

(7) R07が、第2の次世代シーケンシング (NGS) アダプター配列を含むフリーの5'末端を含む；

40

(8) R08が、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端にあるかまたはその付近にある第2の配列を含む；

(9) R09が、一本鎖リンカー、第2のUMI、および第2のリードプライマー部位を含む；および／または

(10) R10が、第4のアダプターを含む、

請求項1に記載の複合体。

【請求項 3】

第1の配列および第2の配列が、同じかまたは異なるプライマー結合部位をさらに含む、請求項2に記載の複合体。

50

【請求項 4】

第 1 および第 2 のプライマー部位が、反対方向の付加によってシーケンシングを開始するように配向されている、請求項 2 または請求項 2～3 のいずれか一項に記載の複合体。

【請求項 5】

第 1 および第 2 の UMI が、はっきり区別できるものである、請求項 1～4 のいずれか一項に記載の複合体。

【請求項 6】

R01 が、少なくとも 12 個のヌクレオチドを含む、
R02 が、少なくとも 14 個のヌクレオチドを含む、
R03 が、少なくとも 12 個のヌクレオチドを含む、
R04 が、少なくとも 20 個のヌクレオチドを含む、
R05 が、少なくとも 12 個のヌクレオチドを含む、
R06 が、少なくとも 12 個のヌクレオチドを含む、
R07 が、少なくとも 20 個のヌクレオチドを含む、
R08 が、少なくとも 12 個のヌクレオチドを含む、
R09 が、少なくとも 14 個のヌクレオチドを含む、および／または
R10 が、少なくとも 12 個のヌクレオチドを含む、

請求項 1～5 のいずれか一項に記載の複合体。

【請求項 7】

R01 が、30 個よりも少ないヌクレオチドを含む、
R02 が、75 個よりも少ないヌクレオチドを含む、
R03 が、99 個よりも少ないヌクレオチドを含む、
R04 が、49 個よりも少ないヌクレオチドを含む、
R05 が、30 個よりも少ないヌクレオチドを含む、
R06 が、30 個よりも少ないヌクレオチドを含む、
R07 が、49 個よりも少ないヌクレオチドを含む、
R08 が、99 個よりも少ないヌクレオチドを含む、
R09 が、75 個よりも少ないヌクレオチドを含む、および／または
R10 が、30 個よりも少ないヌクレオチドを含む、

請求項 1～6 のいずれか一項に記載の複合体。

【請求項 8】

R01 が、12～30 個の間のヌクレオチドを含む、
R02 が、14～75 個の間のヌクレオチドを含む、
R03 が、12～99 個の間のヌクレオチドを含む、
R04 が、20～49 個の間のヌクレオチドを含む、
R05 が、12～30 個の間のヌクレオチドを含む、
R06 が、12～30 個の間のヌクレオチドを含む、
R07 が、20～49 個の間のヌクレオチドを含む、
R08 が、12～99 個の間のヌクレオチドを含む、
R09 が、14～75 個の間のヌクレオチドを含む、および／または
R10 が、12～30 個の間のヌクレオチドを含む、

請求項 1～7 のいずれか一項に記載の複合体。

【請求項 9】

(a) R01 および R06 が、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/mol のハイブリダイゼーション自由エネルギーを含む；

(b) R03 および R08 が、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、約-35kcal/mol、約-40kcal/mol、約-45kcal/mol、約-50kcal/mol、約-55kcal/mol、約-60 のハイブリダイゼーション自由エネルギーを含む；および／または

10

20

30

40

50

(c) R05およびR10が、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギーを含む、
請求項1～8のいずれか一項に記載の複合体。

【請求項10】

(a) R01およびR06が、同じ数のヌクレオチドを各々含み、任意にここで、R06は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを有する；
(b) R03およびR08が、同じ数のヌクレオチドを各々含む；および／または
(c) R05およびR10が、同じ数のヌクレオチドを各々含み、任意にここで、R05は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを有する、
請求項1～9のいずれか一項に記載の複合体。

10

【請求項11】

(a) R01とR06とが、少なくとも90%の相補性がある配列を含む；
(b) R03とR08とが、少なくとも90%の相補性がある配列を含む；および／または
(c) R05とR10とが、少なくとも90%の相補性がある配列を含む、
請求項1～10のいずれか一項に記載の複合体。

【請求項12】

各々R01、R06、R05、およびR10が同じ数のヌクレオチドを含み、任意にここで、R06およびR05は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを各々有する、請求項1～11のいずれか一項に記載の複合体。

20

【請求項13】

請求項2に記載の少なくとも2つの要素を含む、請求項2～12のいずれか一項に記載の複合体。

【請求項14】

請求項2に記載の少なくとも3つの要素を含む、請求項2～13のいずれか一項に記載の複合体。

【請求項15】

請求項2に記載の少なくとも4つの要素を含む、請求項2～14のいずれか一項に記載の複合体。

【請求項16】

請求項2に記載の少なくとも5つの要素を含む、請求項2～15のいずれか一項に記載の複合体。

30

【請求項17】

請求項2に記載の少なくとも6つの要素を含む、請求項2～16のいずれか一項に記載の複合体。

【請求項18】

請求項2に記載の少なくとも7つの要素を含む、請求項2～17のいずれか一項に記載の複合体。

【請求項19】

請求項2に記載の少なくとも8つの要素を含む、請求項2～18のいずれか一項に記載の複合体。

40

【請求項20】

請求項2に記載の少なくとも9つの要素を含む、請求項2～19のいずれか一項に記載の複合体。

【請求項21】

(1) R01が、第1の連結二重鎖シーケンシング(CDS)アダプターを含み；
(2) R02が、一本鎖リンカーを含み；
(3) R03が、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端を含み；
(4) R04が、第1のユニークな分子識別子(UMI)を含み；

50

- (5) R05が、第3のCDSアダプターを含み；
 - (6) R06が、第2のCDSアダプターを含み；
 - (7) R07が、第2のUMIを含み；
 - (8) R08が、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端を含み；
 - (9) R09が、一本鎖リンカーを含み；および
 - (10) R10が、第4のCDSアダプターを含む、
- 請求項1～20のいずれか一項に記載の複合体。

【請求項22】

R01の5'末端が、標的DNA二重鎖の第1の鎖の3'末端ヘライゲーションされ；
 R05の3'末端が、標的DNA二重鎖の第1の鎖の5'末端ヘライゲーションされ；
 R10の5'末端が、標的DNA二重鎖の第2の鎖の3'末端ヘライゲーションされ；
 R06の3'末端が、標的DNA二重鎖の第2の鎖の5'末端ヘライゲーションされ；
 環状化されたDNA二重鎖または任意には部分的に二本鎖の環状DNAを形成している、

請求項1～21のいずれか一項に記載の複合体。

10

【請求項23】

DNA試料の次世代シーケンスにおける使用のための、請求項1～22のいずれか一項に記載の単離された核酸複合体。

【請求項24】

DNA試料の配列を得る次世代シーケンシングワークフローにおける二重鎖アダプターの代わりに使用するための、請求項1～22のいずれか一項に記載の単離された核酸複合体。

20

【請求項25】

第1の末端、第2の末端、ならびに第1および第2の末端の間に位置づけられた中央部分を有するシーケンシングアダプターであって、ここで、第1の末端は、第2のオリゴヌクレオチドにアニーリングされた第1のオリゴヌクレオチドを含む第1の二重鎖を含み、ここで、第2の末端は、第4のオリゴヌクレオチドにアニーリングされた第3のオリゴヌクレオチドを含む第2の二重鎖を含み、およびここで、第2および第4のオリゴヌクレオチドは、領域相補性にわたって相互にアニーリングされることで、中央部分に位置づけられた第3の二重鎖を形成しており、ここで、シーケンシングアダプターはさらに、一本鎖領域において第3の二重鎖の両側に1対のリードプライマー結合部位を含む、前記シーケンシングアダプター。

30

【請求項26】

第1の二重鎖が、長さ20bp、21bp、22bp、23bp、24bp、25bp、26bp、27bp、28bp、29bp、30bp、31bp、32bp、33bp、34bp、35bp、36bp、37bp、38bp、39bp、または40bpである、請求項25に記載のシーケンシングアダプター。

【請求項27】

第1の二重鎖が、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギー自由を有する、請求項25に記載のシーケンシングアダプター。

40

【請求項28】

第2の二重鎖が、長さ10bp、11bp、12bp、13bp、14bp、15bp、16bp、17bp、18bp、19bp、20bp、21bp、22bp、23bp、24bp、または25bpである、請求項25に記載のシーケンシングアダプター。

【請求項29】

第1の二重鎖が、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギー自由を有する、請求項25に記載のシーケンシングアダプター。

【請求項30】

第3の二重鎖が、10bp、11bp、12bp、13bp、14bp、15bp、16bp、17bp、1

50

8bp、19bp、20bp、21bp、22bp、23bp、24bp、または25bpである、請求項25に記載のシーケンシングアダプター。

【請求項31】

第3の二重鎖が、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギー自由を有する、請求項25に記載のシーケンシングアダプター。

【請求項32】

一本鎖領域が、長さ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、または100ヌクレオチドである、請求項25に記載のシーケンシングアダプター。

10

【請求項33】

第1のオリゴヌクレオチドが、第1の次世代シーケンシング（NGS）フローセル結合領域を含むフリーの5'末端を含む、請求項25に記載のシーケンシングアダプター。

【請求項34】

第3のオリゴヌクレオチドが、第2の次世代シーケンシング（NGS）フローセル結合領域を含むフリーの5'末端を含む、請求項25に記載のシーケンシングアダプター。

20

【請求項35】

第1の二重鎖が、第1のフリーの5'末端を有し、および第2の二重鎖が、第2のフリーの5'末端を有する、請求項25に記載のシーケンシングアダプター。

【請求項36】

第3の二重鎖が、二重鎖の各鎖のフリーの5'末端を含み、ここで、第1および第2の3'末端が、DNA依存性のDNAポリメラーゼによるDNA合成をプライムすることができる、請求項25に記載のシーケンシングアダプター。

【請求項37】

DNA試料の次世代シーケンスにおける使用のための、請求項23～36のいずれか一項に記載のシーケンシングアダプター。

30

【請求項38】

DNA試料の配列を得る次世代シーケンシングワークフローにおける二重鎖アダプターの代わりに使用するための、請求項23～36のいずれか一項に記載のシーケンシングアダプター。

【請求項39】

シーケンシングライブラリーを調製する方法であって、

(a) 請求項1～22のいずれか一項に記載の複合体を以下のとおりにdsDNA二重鎖ヘライゼーションすること：R01の5'末端をdsDNA二重鎖の第1の鎖の3'末端ヘライゼーションする；R05の3'末端をdsDNA二重鎖の第1の鎖の5'末端ヘライゼーションする；R10の5'末端をdsDNA二重鎖の第2の鎖の3'末端ヘライゼーションする；およびR06の3'末端をdsDNA二重鎖の第2の鎖の5'末端ヘライゼーションする；それによって、標的DNA分子および複合体を含む環状二本鎖DNA中間体が形成される；

40

(b) R03の3'末端から第1のDNA鎖を伸長させること；

(c) R08の3'末端から第2のDNA鎖を伸長させること；および

(d) 任意に、標的DNA分子の次世代シーケンシング（NGS）における使用のための二本鎖DNA分子を形成するために第1および第2のDNA鎖をアニーリングすることを含む、前記方法。

【請求項40】

二本鎖DNA分子が、標的DNA分子の2つのコピーを含む、請求項39に記載の方法。

【請求項41】

50

ステップ (a) のライゲーションすることが、リガーゼを添加することを含む、請求項 39 または 40 に記載の方法。

【請求項 42】

ステップ (b) および (c) の合成することが、環状二本鎖 DNA 中間体をポリメラーゼと接触させることを含む、請求項 39 ~ 41 のいずれか一項に記載の方法。

【請求項 43】

ポリメラーゼが、DNA 依存性の DNA ポリメラーゼである、請求項 42 に記載の方法。

【請求項 44】

ポリメラーゼが、鎖置換活性を有する、請求項 42 または 43 に記載の方法。

【請求項 45】

次世代シーケンシング (NGS) が、ショートリード戦略である、請求項 39 ~ 44 のいずれか一項に記載の方法。

【請求項 46】

次世代シーケンシングによって二本鎖 DNA 分子をシーケンシングすることをさらに含む、請求項 39 ~ 45 のいずれか一項に記載の方法。

【請求項 47】

シーケンシングされる複数の DNA 二重鎖を含むシーケンシングライブラリーを調製する方法であって、ライブラリーの各メンバーについて、以下のことを含む：

(a) 請求項 25 ~ 36 のいずれか一項に記載のシーケンシングアダプターの第 1 および第 2 の末端を、対向する top および bottom 鎖を有する試料 DNA フラグメントヘライゲーションすること；それによって、DNA フラグメントおよびシーケンシングアダプターを含む部分的に環状化された DNA 分子が形成される；および

(b) 部分的に環状化された DNA 分子の反対の鎖を鋳型として各々使用してシーケンシングアダプターのフリーの 3' 末端を伸長させることにより第 1 および第 2 の一本鎖 DNA 分子を合成すること、それによって、次世代シーケンシングのために構成された直鎖化された二本鎖 DNA 分子が形成され、該直鎖化された二本鎖 DNA 分子は、コピーされた bottom 鎖と対をなす元の top 鎖を含む第 1 の二本鎖領域および元の bottom 鎖と対をなすコピーされた top 鎖を含む第 2 の二本鎖領域を含んでいる、

ここで、各々異なる DNA フラグメントから調製された複数の直鎖化された二本鎖 DNA 分子が、次世代シーケンシングライブラリーを構成する、前記方法。

【請求項 48】

次世代シーケンシングのために構成され、および、第 1 および第 2 の末端を有する、直鎖化された二本鎖 DNA 分子が、以下の構造：

第 1 の末端 - [第 1 の次世代フローセルアダプター] - [元の bottom 鎖のコピーと対をなす元の top 鎖を含む第 1 の二重鎖領域] - [次世代シーケンシングアダプターの中央部分を含む第 2 の二重鎖領域] - [元の bottom 鎖と対をなす元の top 鎖のコピーを含む第 3 の二重鎖領域] - [第 2 の次世代フローセルアダプター] - 第 2 の末端を含む、請求項 47 に記載の方法。

【請求項 49】

第 1 の次世代フローセルアダプターが、Illumina P5 または P7 アダプター配列である、請求項 48 に記載の方法。

【請求項 50】

第 2 の次世代フローセルアダプターが、Illumina P5 または P7 アダプター配列である、請求項 48 に記載の方法。

【請求項 51】

第 2 の二重鎖領域が、第 1 および第 2 のリードプライマー結合部位を含み、ここで、各々第 1 および第 2 のリードプライマー部位は、ユニークな分子識別子 (UMI) およびサンプリングインデックス配列とさらに会合する、請求項 48 に記載の方法。

【請求項 52】

第 1 および第 2 のリードプライマー結合部位が、直鎖化された二本鎖 DNA 分子の末端

10

20

30

40

50

へ向かい外向きに配向されている、請求項 5 1 に記載の方法。

【請求項 5 3】

第 1 のリードプライマーを、UMI、サンプルインデックス、およびシーケンシングされる試料 DNA フラグメントの元の top 鎖またはその部分を含む配列リードを得るために使用することができる、請求項 5 2 に記載の方法。

【請求項 5 4】

第 2 のリードプライマーを、UMI、サンプルインデックス、およびシーケンシングされる試料 DNA フラグメントの元の bottom 鎖またはその部分を含む配列リードを得るために使用することができる、請求項 5 2 に記載の方法。

【請求項 5 5】

市販の次世代ライブラリー構築キットの代わりに使用される、請求項 4 7 に記載の方法。

10

【請求項 5 6】

ステップ (a) のライゲーションすることが、リガーゼを添加することを含む、請求項 4 7 に記載の方法。

【請求項 5 7】

ステップ (b) の合成することが、DNA ポリメラーゼ（任意には鎖置換活性を有するもの）を添加することを含む、請求項 4 7 に記載の方法。

【請求項 5 8】

第 1 および第 2 のリードプライマーを用いて次世代シーケンシングを実施することにより元の上のおよび元の bottom 鎖の配列を得るステップをさらに含む、請求項 5 1 に記載の方法。

20

【請求項 5 9】

請求項 4 7 に記載の方法によって得られる次世代シーケンシングのために構成された直鎖化された二本鎖 DNA 分子であって、第 1 および第 2 の末端を含み、および以下の構造：

第 1 の末端－[第 1 の次世代フローセルアダプター]－[元の bottom 鎖のコピーと対をなす元の top 鎖を含む第 1 の二重鎖領域]－[次世代シーケンシングアダプターの中央部分を含む第 2 の二重鎖領域]－[元の bottom 鎖と対をなす元の top 鎖のコピーを含む第 3 の二重鎖領域]－[第 2 の次世代フローセルアダプター]－第 2 の末端

を有する、前記直鎖化された二本鎖 DNA 分子。

30

【請求項 6 0】

第 1 の次世代フローセルアダプターが、Illumina P5 または P7 アダプター配列である、請求項 5 9 に記載の直鎖化された二本鎖 DNA 分子。

【請求項 6 1】

第 2 の次世代フローセルアダプターが、Illumina P5 または P7 アダプター配列である、請求項 5 9 に記載の直鎖化された二本鎖 DNA 分子。

【請求項 6 2】

第 2 の二重鎖領域が、第 1 および第 2 のリードプライマー結合部位を含み、ここで、各々第 1 および第 2 のリードプライマー部位は、ユニークな分子識別子 (UMI) およびサンプルインデックス配列とさらに会合する、請求項 5 9 に記載の直鎖化された二本鎖 DNA 分子。

40

【請求項 6 3】

第 1 および第 2 のリードプライマー結合部位が、直鎖化された二本鎖 DNA 分子の末端へ向かい外側へ向けられている、請求項 6 2 に記載の直鎖化された二本鎖 DNA 分子。

【請求項 6 4】

第 1 のリードプライマーを、UMI、サンプルインデックス、およびシーケンシングされる試料 DNA フラグメントの元の top 鎖またはその部分を含む配列リードを得るために使用することができる、請求項 6 2 に記載の直鎖化された二本鎖 DNA 分子。

【請求項 6 5】

第 2 のリードプライマーを、UMI、サンプルインデックス、およびシーケンシングさ

50

れる試料DNAフラグメントの元のbottom鎖またはその部分を含む配列リードを得るために使用することができる、請求項62に記載の直鎖化された二本鎖DNA分子。

【請求項66】

DNA試料の次世代シーケンシングのための方法であって、

(a) 生体源からDNA試料を得ること；

(b) 複数のDNAフラグメントを得るためにDNA試料をフラグメント化すること；

(c) 各鎖がDNAフラグメントのtopおよびbottom鎖のコンカテマーを含んでいる複数の直鎖化された二本鎖DNA分子を生成するために請求項47～57のいずれか一項に記載の方法によりDNAフラグメントの次世代シーケンシングライブラリーを構築すること；および

(d) 直鎖化された二本鎖DNA分子へ結合するリードプライマーを用いた次世代シーケンシングを使用してDNAフラグメントのtopおよびbottom鎖の配列を決定すること、それによって、DNA分子の配列が得られる、を含む、前記方法。

10

【請求項67】

生体試料が、血液である、請求項66に記載の方法。

【請求項68】

生体試料が、肝臓、腎臓、脳、心臓、皮膚、肺、結腸、または脾臓からの組織の試料である、請求項66に記載の方法。

【請求項69】

生体試料が、肝臓、腎臓、脳、心臓、皮膚、肺、結腸、または脾臓からの罹患組織の試料である、請求項66に記載の方法。

20

【請求項70】

罹患組織が、増殖性疾患である、請求項69に記載の方法。

【請求項71】

罹患組織が、腫瘍である、請求項69に記載の方法。

【請求項72】

シーケンシングエラー率が、Duplex Sequencingに基づく対照と同様であるが、しかしここで、必要とされるリードの数が、少なくとも100倍減少している、請求項66に記載の方法。

30

【請求項73】

少なくとも1つのオリゴヌクレオチドが、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための請求項1～22のいずれか一項に記載の単離された核酸複合体。

【請求項74】

第1、第2、第3、および第4のオリゴヌクレオチドの各々が、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための請求項1～22のいずれか一項に記載の単離された核酸複合体。

【請求項75】

少なくとも1つのオリゴヌクレオチドが、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための請求項25～38のいずれか一項に記載のシーケンシングアダプター。

40

【請求項76】

第1、第2、第3、および第4のオリゴヌクレオチドの各々が、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための請求項25～38のいずれか一項に記載のシーケンシングアダプター。

【請求項77】

50

(a) 請求項 25～38 のいずれか一項に記載のシーケンシングアダプターの第 1 および第 2 の末端を、対向する top および bottom 鎖を有する試料 DNA フラグメントヘライゲーションすること、それによって、DNA フラグメントおよびシーケンシングアダプターを含む部分的に環状化された DNA 分子が形成される、ここでシーケンシングアダプターは、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている；および

(b) 部分的に環状化された DNA 分子の反対の鎖を鋳型として各々使用してシーケンシングアダプターのフリーの 3' 末端を伸長させることにより第 1 および第 2 の一本鎖 DNA 分子を合成すること、

それによって、各鎖が DNA フラグメントの top および bottom 鎖のコンカテマーを含んでいる直鎖化された二本鎖 DNA 分子が形成され、ここで、合成するステップは、フリーの 3' 末端を、DNA ポリメラーゼ、および、標準の dATP、dGTP および dTTP デオキシヌクレオチドとともに用いられているメチル化された dCTP と、接触させることを含む；

(c) DNA フラグメントの、元の top 鎖において、メチル化されていないシトシンをウラシルに脱アミノ化させること；

(d) 次世代シーケンシングにより top および bottom 鎖の配列を決定すること、

(e) 元の DNA フラグメントにおけるメチル化位置を推測するために配列を比較することを含む、DNA 試料のメチル化シーケンシングの方法。

【請求項 78】

DNA 試料が、生体試料から得られる、請求項 77 に記載の方法。

【請求項 79】

生体試料が、肝臓、腎臓、脳、心臓、皮膚、肺、結腸、または脾臓組織から得られ、任意にここで、組織は、罹患している、請求項 78 に記載の方法。

【請求項 80】

疾患が、増殖性疾患である、請求項 79 に記載の方法。

【請求項 81】

疾患が、腫瘍である、請求項 79 に記載の方法。

【請求項 82】

dsDNA 二重鎖が、ステップ (a) に先立ちプレ増幅される、請求項 39 に記載の方法であって、

(a) dsDNA 二重鎖を第 1 および第 2 のプレ増幅分子と接触させること、ここで、2 つのプレ増幅分子の各々は、UMI、サンプルインデックス、ローリングサークル増幅 (RCA) プライマー、およびトランケーション部位を含む；

(b) プレ増幅 dsDNA 二重鎖を作るために、第 1 のプレ増幅分子を、dsDNA 二重鎖の、1 つの第 1 の末端ヘライゲーションし、および第 2 のプレ増幅分子を、dsDNA 二重鎖の第 2 の末端ヘライゲーションすること；

(c) プレ増幅 dsDNA 二重鎖を DNA ポリメラーゼ酵素に晒すこと；

(d) RCA を完了するのに十分な時間の間、プレ増幅 dsDNA 二重鎖および DNA ポリメラーゼ酵素をインキュベートすること；および

(e) プレ増幅 dsDNA 二重鎖をトランケーション部位で切断することにより RCA プライマーを除去することを含む、前記方法。

【請求項 83】

シーケンシングされる DNA 二重鎖が、ステップ (a) に先立ちプレ増幅される、請求項 47 に記載の方法であって、

(a) シーケンシングされる DNA 二重鎖の各々を第 1 および第 2 のプレ増幅分子と接触させること、ここで、2 つのプレ増幅分子の各々は、UMI、サンプルインデックス、ローリングサークル増幅 (RCA) プライマー、およびトランケーション部位を含む；

(b) 複数のプレ増幅 DNA 二重鎖を作るために、第 1 のプレ増幅分子を、シーケンシ

10

20

30

40

50

ングされるDNA二重鎖の各々の1つの第1の末端ヘライゲーションし、および第2のプレ増幅分子を、シーケンシングされるDNA二重鎖の各々の第2の末端ヘライゲーションすること；

(c) プレ増幅DNA二重鎖の各々をDNAポリメラーゼ酵素に晒すこと；

(d) RCAを完了するのに十分な時間の間、プレ増幅DNA二重鎖の各々およびDNAポリメラーゼ酵素をインキュベートすること；および

(e) プレ増幅DNA二重鎖の各々をトランケーション部位で切断することによりRCAプライマーを除去すること

を含む、前記方法。

【請求項84】

10

次世代シーケンシングライブラリーを調製する方法であって、

(a) R06の3'末端およびR05の3'末端を、ライゲーションを受けることからブロックすること；

(b) 請求項1～22のいずれか一項に記載の複合体を以下のとおりにdsDNA二重鎖ヘライゲーションすること：R01の5'末端をdsDNA二重鎖の第1の鎖の3'末端ヘライゲーションする；およびR10の5'末端をdsDNA二重鎖の第2の鎖の3'末端ヘライゲーションする；それによって、標的DNA分子および複合体を含む環状二本鎖DNA中間体が形成される；

(c) R03の3'末端から第1のDNA鎖を伸長させること；

(d) R08の3'末端から第2のDNA鎖を伸長させること；および

20

(e) 環状の一本鎖シーケンシング分子を形成するために第1および第2のDNA鎖の各々を環状化させること；

(f) 直鎖の一本鎖シーケンシング分子を形成するためにR03とR08との間の領域中へとニックを導入すること

を含む、前記方法。

【請求項85】

ステップ(a)のブロックすることが、ブロッキング溶液を添加することを含む、請求項84に記載の方法。

【請求項86】

ステップ(b)のライゲーションすることが、リガーゼを添加することを含む、請求項84または85に記載の方法。

30

【請求項87】

ステップ(c)および(d)の合成することが、環状二本鎖DNA中間体をポリメラーゼと接触させることを含む、請求項84～86のいずれか一項に記載の方法。

【請求項88】

ポリメラーゼが、DNA依存性のDNAポリメラーゼである、請求項84～87のいずれか一項に記載の方法。

【請求項89】

ポリメラーゼが、鎖置換活性を有する、請求項84～88のいずれか一項に記載の方法。

40

【請求項90】

次世代シーケンシング(NGS)が、ショートリード戦略である、請求項84～89のいずれか一項に記載の方法。

【請求項91】

シーケンシングライブラリーを調製する方法であって、

(a) dsDNA二重鎖を得ること；

(b) dsDNA二重鎖を二重鎖修復により処理すること；

(c) 請求項1～22のいずれか一項に記載の複合体を以下のとおりにdsDNA二重鎖ヘライゲーションすること：R01の5'末端をdsDNA二重鎖の第1の鎖の3'末端ヘライゲーションする；R05の3'末端をdsDNA二重鎖の第1の鎖の5'末端ヘライゲーション

50

する；R10の5'末端をdsDNA二重鎖の第2の鎖の3'末端ヘライゲーションする；およびR06の3'末端をdsDNA二重鎖の第2の鎖の5'末端ヘライゲーションする；それによって、標的DNA分子および複合体を含む環状二本鎖DNA中間体が形成される；

(d) R03の3'末端から第1のDNA鎖を伸長させること；

(e) R08の3'末端から第2のDNA鎖を伸長させること；および

(f) 任意に、標的DNA分子の次世代シーケンシング(NGS)における使用のための二本鎖DNA分子を形成するために第1および第2のDNA鎖をアニーリングすることを含む、前記方法。

【請求項92】

シーケンシングされる複数のDNA二重鎖を含むシーケンシングライブラリーを調製する方法であって、

(a) 二本鎖修復により複数のDNA二本鎖を処理し；

(b) 請求項25～36のいずれか一項に記載のシーケンシングアダプターの第1および第2の末端を、対向するtopおよびbottom鎖を有する試料DNAフラグメントヘライゲーションし、それによって、DNAフラグメントおよびシーケンシングアダプターを含む部分的に環状化されたDNA分子が形成され；および

(c) 部分的に環状化されたDNA分子の反対の鎖を鋳型として各々使用してシーケンシングアダプターのフリーの3'末端を伸長させることにより第1および第2の一本鎖DNA分子を合成し、それによって、次世代シーケンシングのために構成された直鎖化された二本鎖DNA分子が形成され、該直鎖化された二本鎖DNA分子は、コピーされたbottom鎖と対をなす元のtop鎖を含む第1の二本鎖領域および元のbottom鎖と対をなすコピーされたtop鎖を含む第1の二本鎖領域を含み、

ここで、各々異なるDNAフラグメントから調製された複数の直鎖化された二本鎖DNA分子が、次世代シーケンシングライブラリーを構成する、前記方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

この出願は、2020年12月11日出願の「METHOD FOR DUPLEX SEQUENCING」という表題の米国仮出願第63/124,696号、2021年1月29日に提出された「METHOD FOR DUPLEX SEQUENCING」という表題の米国仮出願第63/143,334号、2021年6月9日に提出された「METHOD FOR DUPLEX SEQUENCING」という表題の米国仮出願第63/208,951号、2021年6月30日に提出された「METHOD FOR DUPLEX SEQUENCING」という表題の米国仮出願第63/217,232号、および2021年9月1日に提出された「METHOD FOR DUPLEX SEQUENCING」という表題の米国仮出願第63/239,920号の、米国特許法第119(e)条の下の特許を主張するものであり、その各々の全開示は、それらの全体において参照により本明細書に組み込まれる。

【0002】

配列表

本出願には、EFS-Webを介してASCIIフォーマットで提出されておりその全体において参照により本明細書に組み込まれる配列表が含まれる。2021年12月10日に作り出された該ASCIIコピーは、B119570111WO00-SEQ-GJM.txtと名付けられておりおよびサイズ7,934バイトである。

【背景技術】

【0003】

背景

DNAは、生命の形成基盤である。DNAにおける突然変異は、遺伝的多様性を進め、遺伝子機能を変化させ、細胞の表現型に影響を及ぼし、細胞集団をマークし、進化の軌跡を定義し、疾患および状態を浮き彫りにし、ならびに精密医療および診断の標的を提供する。突然変異は、単一細胞から出現し、およびクローンの存在量において拡大または縮小す

10

20

30

40

50

る子孫に受け継がれる。よって、広範な存在量にわたって突然変異を検出することが可能であることは不可欠である。低い存在量の突然変異（例として、 $<0.1\sim 1\%$ VAF、「単一の二重鎖」の分解能まで下方）を検出することは、がんの進化および薬物耐性を研究すること、体細胞モザイクおよびクローン造血を理解すること、CRISPRなどの塩基編集技術の特徴付けすること、化学化合物の変異原性を評価すること、病原性のバリエーションを発見すること、ヒト胚発生を研究すること、微生物感染またはウイルス感染およびがん、および臨床的に対処可能な（clinically actionable）ゲノム変化を、組織または液体生検などの標本から検出すること、およびその他多くのことのために重要である。

【0004】

原理上は、第3世代の「単一分子」シーケンシング戦略（例として、PacBio、Oxford Nanopore Technologies）の使用は、各単一DNA二重鎖を全体としてシーケンシングすることで偽の突然変異とは別に真の突然変異を明らかにできるが、しかし実用上は、必要とされる正確性およびスループットに欠ける。次世代シーケンシング（NGS）は、他方、秀でたリードの正確性およびスループットを引き続き提供するが、しかし、少なくともそのスループットまたは実用性を損なうことなしに、単一の二重鎖をシーケンシングするようには構成されていない。

【0005】

NGSは、大規模並列蛍光分析において短いクローン的に増幅されたDNAフラグメントを読み取ることにより、高いスループットを提供する。その正確さは、しかしながら、各DNA二重鎖のワトソンとクリックの鎖を解離する必要性により、限定される。比較のための相補鎖がないと、塩基損傷、PCR、およびシーケンシングに起因していずれかの鎖へ導入されたエラー（すなわち、「偽の突然変異」）は、真の突然変異として見せかけられることがある（例として、図1Aを参照）。ユニークな分子識別子（UMI）を使用することで各DNA分子の両方の鎖を別々に追跡し、およびそれらの配列を比較することで真の突然変異（各二重鎖の両方の鎖に存在する）を偽の突然変異（二重鎖の鎖の片方にのみ存在する）から判別することが可能になる一方、それはNGSの根本的な限界、つまり二重鎖解離を解決しない。

【0006】

「二重鎖シーケンシング」と呼ばれている改変されたNGSワークフローは、Schmitt et al., 「Detection of ultra-rare mutations by next-generation sequencing」PNAS, Sept. 4, 2012, Vol. 109, No. 36, pp. 14508-14513（その全内容は参照により本明細書に組み込まれる）に初めて記載され、および、一本鎖DNAのシーケンシングに関連するNGSの限界を克服するように設計されたものである。方法は、Schmitt et al.において「Duplex Tag」と称している特殊アダプターに頼り、これは、NGSワークフロー（例として、フローセルに対するクラスター増幅、配列リードを生成するためのシーケンシング、およびアラインメント/データ分析）に進められることに先立ちDNAフラグメントとNGSフローセルアダプターとの間に挟まれるDNAフラグメントの末端に付けられる二本鎖のランダム化された配列である。分析ステージの間、配列リード（DNAフラグメントの両方の鎖の配列を包含する）は、適切なDuplex Tagをマッチングさせることにより同じDNAフラグメントのtopおよびbottom鎖の配列（top and bottom strand sequences）のセットへとグループ化される。これらのセットは、配列アラインメントされ、そして、シーケンシングされる二重鎖の各々の上および下の一本鎖についてのコンセンサス配列を表す一本鎖コンセンサス配列（SSCS）を生成するため比較される。このステージでは、SSCSは依然として真の突然変異および偽の突然変異を包含する。Duplex Tagが、次いで、topおよびbottom鎖のSSCSを対合させてそれによってコンセンサス二重鎖配列を確立するために使用され、それは次いで真の突然変異を偽の突然変異から仕分けるために分析される。各DNA二重鎖のtopおよびbottom鎖における内在の情動的対照を考えると、真の突然変異はtopおよびbottom鎖の配列の両方に現れるものであり、他方、偽の突然変異は鎖の配列の片方にのみ現れる。

【0007】

各々の元の二重鎖のワトソンとクリックのtopおよびbottom鎖に割り当てられたリード間で二重鎖コンセンサスを形成することにより、二重鎖シーケンシングは、最大1000倍までまたはそれよりも高い正確性を達成し、および単一のDNA二重鎖内で偽の突然変異から真の突然変異を明らかにできる。しかしながら、NGSフローセル（例として、Illumina、NovaSeq）上の最大100億もの他の鎖の中から両方の鎖の配列を回収することは、標準的なNGSワークフローと比較して100倍を超えるシーケンシングリードを必要とし、これは常にNGSのスループットを減退させ、そして過剰なコストを一因として二重鎖シーケンシングの適用可能性を著しく限定する。この二重鎖シーケンシングの高い非効率性はまた、アダプターライゲーションの後で両方の鎖が分離されおよびNGSワークフローの間に独立して増幅させられるということにも端を発している。これは、鎖の表現を歪め、そして両方の鎖を少なくとも1回読み取るのに膨大な数のリードを必要とすることにつながる。

10

【0008】

したがって、突然変異検出を損なうことなく、かつ、高コストを必要とすることなく、二重鎖シーケンシングなどのデュアル鎖シーケンシング方法の正確性およびスループットを改善するために、新しい方法が必要である。

【発明の概要】

【0009】

発明の概要

本開示は、本明細書中において「エラー訂正のための元の二重鎖の連結（Concatenating Original Duplex for Error Correction）」シーケンシングまたは「CODEC」シーケンシングと称する、伝統的な二重鎖シーケンシングの短所を改善した新規の二重鎖もしくは「デュアル鎖の」シーケンシング方法を提供する。方法は、低コストでありながら、希少な突然変異を検出することが可能な高品質のDNAシーケンシングリードを生じる。

20

【0010】

様々な側面において、本開示は、CODECシーケンシングのための方法、ならびに、アダプター（本明細書中様々な態様において「CODECアダプター」と称する）、シーケンシングされるDNAフラグメントの両末端ヘライゲーションされたCODECアダプターを各々含む環状化された中間体（本明細書中様々な態様において「CODEC環状化中間体」と称する）、および、シーケンシングされる単一DNAフラグメントの連結されたtopおよびbottom鎖を含む直鎖化された二本鎖産物（本明細書中様々な態様において「CODECライブラリー」または個々に「CODECライブラリーメンバー」と称する）を包含する、CODECシーケンシングに必要とされるおよび／またはCODECシーケンシングによって作られる組成物を提供する。様々な態様において、CODECアダプターは、NGSワークフロー（例として、NGSフローセルに対するクラスター増幅）のためのNGSアダプター、DNAフラグメントの両方の鎖を読み取るためのシーケンシングリードプライマー部位、および任意には、1つ以上のサンプルインデックスおよび1つ以上のユニークな分子識別子（UMI）を包含する。

30

【0011】

シーケンシングされる各DNAフラグメントについてtopおよびbottom鎖の配列を別々に得る（および、よって、topおよびbottom鎖の配列を同定、マッチング、および比較するためにコンピューターのアプローチを必要とする）伝統的な二重鎖シーケンシングと違って、CODECアダプターを使用したライブラリー形成ではシーケンシングされる元の各DNAフラグメント（すなわち、同じDNA分子内）の上下の配列のコンカテマーを各鎖が含む二本鎖ライブラリーメンバーを結果としてもたらすため、CODECライブラリーメンバーの各々は、同じリードにおいて二重鎖コンセンサスを形成するのにそれ自体で十分である。よって、CODECアダプターのシーケンシングは、top鎖、bottom鎖、および任意には、1つ以上のサンプルインデックスおよび1つ以上のUMIを含む、シーケンシング産物を結果としてもたらす。標準的な二重鎖シーケンシングと比較して、このアプロ

40

50

ーチの技術的利点は、2つの別々のシーケンシング産物（すなわち、上の配列に対して1つおよび下の配列に対して1つ）を生成する標準的な二重鎖シーケンシングと違って、CODECシーケンシングは、上の配列および下の配列の両方を含む単一のシーケンシング産物を結果としてもたらし、それによって、ユーザーが簡単に真の突然変異（シーケンシングリードの上および下の両方に現れる突然変異）を偽の突然変異（シーケンシングリードの上のみまたは下のみに現れる突然変異）から判別することを可能にするということである。

【0012】

他の側面において、本開示は、シーケンシングを実施するためのリードプライマー、ならびに、CODECライブラリーをシーケンシングする方法（例として、NGSシーケンシングによる）を記載する。本開示は、結果として得られる配列情報を分析（連結されたtopおよびbottom鎖の配列リードを含むビルトイン（built-in）二重鎖コンセンサスを分析することを包含するがこれに限定されない）するためのコンピューターベースの方法をさらに提供する。単一のリードのtopおよびbottomの配列を比較することにより、真の突然変異（シーケンシングリードのtopおよびbottomの両方に現れる突然変異）を偽の突然変異（シーケンシングリードのtopのみまたはbottomのみに現れる突然変異）から判別することが可能である。

【0013】

なおも別の側面において、本開示は、DNAをシーケンシングするための方法、DNA中の突然変異を検出するための方法、DNA中の希少もしくは低い存在量の突然変異を検出するための方法、DNA中の1つ以上の突然変異の検出に基づき疾患を診断および／または予測する方法、DNA中の1つ以上の突然変異の検出に基づき遺伝的状态を診断および／または予測する方法、ならびに、1つ以上の遺伝子をシーケンシングして1つ以上の疾患関連配列（例として、希少な突然変異）を検出することにより疾患または状態を診断および／または予測する方法を包含するがこれに限定されない、CODECシーケンシングのための方法および用途を提供する。他の側面において、本開示は、本明細書に記載のとおり主題の方法を実践するための組成物（例として、CODECアダプター）およびキットを提供する。

【0014】

もう1つの側面において、本明細書の例2および図16に例示されたとおり、本開示はまた、DNA試料の改善された突然変異およびメチル化シーケンシングを行うために使用することができるメチル化特異的CODECシーケンシングのための方法も記載する。一態様において、本開示は、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されたCODECアダプターを調製することを含み、ここで、メチル化されたシトシンには後続する脱アミノ化が無効であり、および、CODECワークフローに参与する増幅を受けることができる、DNAフラグメントのメチル化シーケンシング（または「メチル-seq」）のための方法を提供する。次に、改変されたCODECアダプターが、DNAフラグメントの両末端ヘライゲーションされ、それによって、CODECアダプター（CODECアダプターの中央の二重鎖における利用可能な5'末端を有する）およびDNAフラグメントを含む部分的に環状化された部分的に二本鎖の中間体コンストラクトが作られる。次に、利用可能な5'末端は、標準のdATP、dGTPおよびdTTPデオキシヌクレオチドとともに用いられているメチル化されたdCTPの存在下で、DNAポリメラーゼにより伸長させられ、ここで、DNAポリメラーゼは、中間体コンストラクトの反対の鎖を鋳型として使用する。両方の利用可能な5'末端からのこのようなDNA伸長は、bottom鎖（元のtop鎖をコピーした産物である）と連結されたDNAフラグメントの元のtop鎖を含む第1の鎖、ならびにtop鎖の逆相補鎖でありtop鎖（元のbottom鎖をコピーした産物である）と連結されたDNAフラグメントの元のbottom鎖を含む第2の鎖を包含する、図1Dのコンカテマーを含む二本鎖産物を生じる。図1Dを参照。しかしながら、この態様において、コピーされた領域は、後続する脱アミノ化が無効であるシトシン位置でメチル化される。次に、元のDNA鎖におけるメチル化されていないシトシンをウラシルに

10

20

30

40

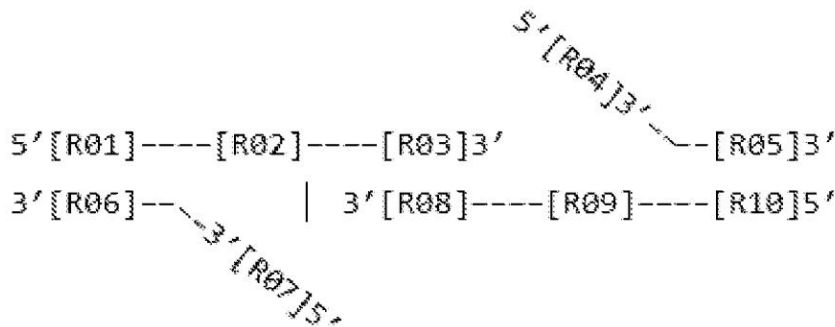
50

変換するために脱アミノ化ステップが行われる。様々な態様において、シトシンの脱アミノ化は、重亜硫酸での脱アミノ化によるもの²、メチル化されたシトシンとメチル化されていないシトシンとを差別化するためにTET2およびAPOBEC2酵素による酵素的ステップを使用する酵素的メチル-seq (EM-seq) 技法を使用した酵素的脱アミノ化によるもの³、またはTET補助Pic-ボランシーケンシング (TET Assisted Pic-borane Sequencing) (TAPS) 法によるもの⁴、などのあらゆる好適な方法によって行われることができる。脱アミノ化ステップに続いて、CODECアダプタープライマーを使用した増幅が適用され、本明細書に別途記載されているとおりの二重鎖シーケンシングがこれに続く。

【0015】

本開示の一側面は、以下の構成にある、少なくとも10個の領域 (R01~R10) を含む単離された核酸複合体 (複合体) であって：

【化1】



ここで、「----」は、結合を表し、ここで、R01、R02およびR03は、第1のオリゴヌクレオチドを含み、ここで、R04およびR05は、第2のオリゴヌクレオチドを含み、ここで、R06およびR07は、第3のオリゴヌクレオチドを含み、ここで、R08、R09、R10は、第4のオリゴヌクレオチドを含み、ここで、R01とR06とは、相互にアニーリングされており、ここで、R03とR08とは、相互にアニーリングされており、ここで、R05とR10とは、相互にアニーリングされており、ここで、R02とR07とは、相互にアニーリングされていない、およびここで、R04とR09とは、相互にアニーリングされていない；ここで、R02は、一本鎖リンカー、第1のユニークな分子識別子 (UMI)、および第1のリードプライマー部位を含み、およびここで、R09は、一本鎖リンカー、第2のUMI、および第2のリードプライマー部位を含む、複合体に関する。

【0016】

いくつかの態様において、R01は、第1のアダプターを含む；R02は、一本鎖リンカー、第1のユニークな分子識別子 (UMI)、および第1のリードプライマー部位を含む；R03は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端にあるかまたはその付近にある第1の配列を含む；R04は、第1の次世代シーケンシング (NGS) アダプター配列を含むフリーの5'末端を含む；R05は、第3のアダプターおよび第1のサンプルインデックスを含む；R06は、第2のアダプターおよび第2のサンプルインデックスを含む；R07は、第2の次世代シーケンシング (NGS) アダプター配列を含むフリーの5'末端を含む；R08は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端にあるかまたはその付近にある第2の配列を含む；R09は、一本鎖リンカー、第2のUMI、および第2のリードプライマー部位を含む；および／またはR10は、第4のアダプターを含む。

【0017】

いくつかの態様において、4つのオリゴヌクレオチドの各々は、ライブラリー調製の前に組み合わせられてもよく、それによって、ライブラリー調製に先立ち複合体が形成される。他の態様において、4つのオリゴヌクレオチドは、ライブラリー調製の間に別々に添加されてもよく、それによって、ライブラリー調製に相応してまたはその最中に、ハイブリダイズした複合体が形成される。

【0018】

いくつかの態様において、第1の配列および第2の配列は、同じかまたは異なるプライマー結合部位をさらに含む。いくつかの態様において、第1および第2のプライマー部位が、反対方向の付加によってシーケンシングを開始するように配向されている。いくつかの態様において、第1および第2のUMIは、はっきり区別できるものである。

【0019】

いくつかの態様において、R01は、少なくとも12個のヌクレオチドを含む、R02は、少なくとも14個のヌクレオチドを含む、R03は、少なくとも12個のヌクレオチドを含む、R04は、少なくとも20個のヌクレオチドを含む、R05は、少なくとも12個のヌクレオチドを含む、R06は、少なくとも12個のヌクレオチドを含む、R07は、少なくとも20個のヌクレオチドを含む、R08は、少なくとも12個のヌクレオチドを含む、R09は、少なくとも14個のヌクレオチドを含む、および／またはR10は、少なくとも12個のヌクレオチドを含む。いくつかの態様において、R01は、30個よりも少ないヌクレオチドを含む、R02は、75個よりも少ないヌクレオチドを含む、R03は、99個よりも少ないヌクレオチドを含む、R04は、49個よりも少ないヌクレオチドを含む、R05は、30個よりも少ないヌクレオチドを含む、R06は、30個よりも少ないヌクレオチドを含む、R07は、49個よりも少ないヌクレオチドを含む、R08は、99個よりも少ないヌクレオチドを含む、R09は、75個よりも少ないヌクレオチドを含む、および／またはR10は、30個よりも少ないヌクレオチドを含む。いくつかの態様において、R01は、12～30個の間のヌクレオチドを含む、R02は、14～75個の間のヌクレオチドを含む、R03は、12～99個の間のヌクレオチドを含む、R04は、20～49個の間のヌクレオチドを含む、R05は、12～30個の間のヌクレオチドを含む、R06は、12～30個の間のヌクレオチドを含む、R07は、20～49個の間のヌクレオチドを含む、R08は、12～99個の間のヌクレオチドを含む、R09は、14～75個の間のヌクレオチドを含む、および／またはR10は、12～30個の間のヌクレオチドを含む。

【0020】

いくつかの態様において、R01およびR06は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギーを含む；R03およびR08は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、約-35kcal/mol、約-40kcal/mol、約-45kcal/mol、約-50kcal/mol、約-55kcal/mol、約-60のハイブリダイゼーション自由エネルギーを含む；および／またはR05およびR10は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギーを含む。

【0021】

いくつかの態様において、R01およびR06は、同じ数のヌクレオチドを各々含み、任意にここで、R06は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを有する；R03およびR08は、同じ数のヌクレオチドを各々含む；および／または、R05およびR10は、同じ数のヌクレオチドを各々含み、任意にここで、R05は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを有する。

【0022】

いくつかの態様において、R01とR06とは、少なくとも90%の相補性がある配列を含む；R03とR08とは、少なくとも90%の相補性がある配列を含む；および／またはR05とR10とは、少なくとも90%の相補性がある配列を含む。いくつかの態様において、各々R01、R06、R05、およびR10は、同じ数のヌクレオチドを含み、任意にここで、R06およびR05は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを各々有する。

【0023】

いくつかの態様において、複合体は、上記のとおり少なくとも2つの要素を含む。い

10

20

30

40

50

いくつかの態様において、複合体は、上記のとおり少なくとも3つの要素を含む。いくつかの態様において、複合体は、上記のとおり少なくとも4つの要素を含む。いくつかの態様において、複合体は、上記のとおり少なくとも5つの要素を含む。いくつかの態様において、複合体は、上記のとおり少なくとも6つの要素を含む。いくつかの態様において、複合体は、上記のとおり少なくとも7つの要素を含む。いくつかの態様において、複合体は、上記のとおり少なくとも8つの要素を含む。いくつかの態様において、複合体は、上記のとおり少なくとも9つの要素を含む。

【0024】

いくつかの態様において、R01は、第1のアダプターを含み；R02は、一本鎖リンカーを含み；R03は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端を含み；R04は、第1のユニークな分子識別子(UMI)を含み；R05は、第3のアダプターを含み；R06は、第2のアダプターを含み；R07は、第2のUMIを含み；R08は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端を含み；R09は、一本鎖リンカーを含み；およびR10は、第4のアダプターを含む。

10

【0025】

いくつかの態様において、R01の5'末端は、標的DNA二重鎖の第1の鎖の3'末端ヘライゲーションされ；R05の3'末端は、標的DNA二重鎖の第1の鎖の5'末端ヘライゲーションされ；R10の5'末端は、標的DNA二重鎖の第2の鎖の3'末端ヘライゲーションされ；R06の3'末端は、標的DNA二重鎖の第2の鎖の5'末端ヘライゲーションされ；環状化されたDNA二重鎖または任意には部分的に二本鎖の環状DNAを形成している。

20

【0026】

本開示の別の側面は、DNA試料の次世代シーケンスにおける使用のための、本明細書に記載のとおり単離された核酸複合体に関する。

【0027】

本開示の別の側面は、DNA試料の配列を得る次世代シーケンシングワークフローにおける二重鎖アダプターの代わりに使用するための、本明細書に記載のとおり単離された核酸複合体に関する。

【0028】

本開示の別の側面は、第1の末端、第2の末端、ならびに第1および第2の末端の間に位置付けられた中央部分を有するシーケンシングアダプターであって、ここで、第1の末端は、第2のオリゴヌクレオチドにアニーリングされた第1のオリゴヌクレオチドを含む第1の二重鎖を含み、ここで、第2の末端は、第4のオリゴヌクレオチドにアニーリングされた第3のオリゴヌクレオチドを含む第2の二重鎖を含み、およびここで、第2および第4のオリゴヌクレオチドは、領域相補性(region complementarity)にわたって相互にアニーリングされることで、中央部分に位置付けられた第3の二重鎖を形成しており、ここで、シーケンシングアダプターがさらに、一本鎖領域において第3の二重鎖の両側に1対のリードプライマー結合部位を含む、前記シーケンシングアダプターに関する。

30

【0029】

いくつかの態様において、第1の二重鎖は、長さ20bp、21bp、22bp、23bp、24bp、25bp、26bp、27bp、28bp、29bp、30bp、31bp、32bp、33bp、34bp、35bp、36bp、37bp、38bp、39bp、または40bpである。いくつかの態様において、第1の二重鎖は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギーを有する。いくつかの態様において、第2の二重鎖は、長さ10bp、11bp、12bp、13bp、14bp、15bp、16bp、17bp、18bp、19bp、20bp、21bp、22bp、23bp、24bp、または25bpである。いくつかの態様において、第1の二重鎖は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギーを有する。いくつかの

40

50

態様において、第3の二重鎖は、長さ10bp、11bp、12bp、13bp、14bp、15bp、16bp、17bp、18bp、19bp、20bp、21bp、22bp、23bp、24bp、または25bpである。いくつかの態様において、第3の二重鎖は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギーを有する。

【0030】

いくつかの態様において、一本鎖領域は、長さ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、または100ヌクレオチドである。

10

【0031】

いくつかの態様において、第1のオリゴヌクレオチドは、第1の次世代シーケンシング（NGS）フローセル結合領域を含むフリーの5'末端を含む。いくつかの態様において、第3のオリゴヌクレオチドは、第2の次世代シーケンシング（NGS）フローセル結合領域を含むフリーの5'末端を含む。いくつかの態様において、第1の二重鎖は、第1のフリーの5'末端を有し、および、第2の二重鎖は、第2のフリーの5'末端を有する。いくつかの態様において、第3の二重鎖は、二重鎖の各鎖のフリーの5'末端を含み、ここで、第1および第2の3'末端は、DNA依存性のDNAポリメラーゼによるDNA合成をプライムすることができる。

20

【0032】

本開示の別の側面は、DNA試料の次世代シーケンスにおける使用のための、本明細書に記載のとおりシーケンシングアダプターに関する。

【0033】

本開示の別の側面は、DNA試料の配列を得る次世代シーケンシングワークフローにおける二重鎖アダプターの代わりに使用するための、本明細書に記載のとおりシーケンシングアダプターに関する。

【0034】

本開示の別の側面は、シーケンシングライブラリーを調製する方法であって、本明細書に記載の複合体を以下のとおりにdsDNA二重鎖ヘライゲーションすること：R01の5'末端をdsDNA二重鎖の第1の鎖の3'末端ヘライゲーションする；R05の3'末端をdsDNA二重鎖の第1の鎖の5'末端ヘライゲーションする；R10の5'末端をdsDNA二重鎖の第2の鎖の3'末端ヘライゲーションする；およびR06の3'末端をdsDNA二重鎖の第2の鎖の5'末端ヘライゲーションする；それによって、標的DNA分子および複合体を含む環状二本鎖DNA中間体が形成される；R03の3'末端から第1のDNA鎖を伸長させること；R08の3'末端から第2のDNA鎖を伸長させること；および任意に、標的DNA分子の次世代シーケンシング（NGS）における使用のための二本鎖DNA分子を形成するために第1および第2のDNA鎖をアニーリングすること、を含む方法に関する。

30

【0035】

いくつかの態様において、二本鎖DNA分子は、標的DNA分子の2つのコピーを含む。いくつかの態様において、上記の第1のステップのライゲーションすることは、リガーゼを添加することを含む。いくつかの態様において、上記の第2および第3のステップの合成することは、環状二本鎖DNA中間体をポリメラーゼと接触させることを含む。いくつかの態様において、ポリメラーゼは、DNA依存性のDNAポリメラーゼである。

40

いくつかの態様において、ポリメラーゼは、鎖置換活性を有する。

いくつかの態様において、次世代シーケンシング（NGS）は、ショートリード戦略である。いくつかの態様において、方法は、次世代シーケンシングによって二本鎖DNA分子をシーケンシングすることをさらに含む。

50

【0036】

本開示の別の側面は、シーケンシングされる複数のDNA二重鎖を含むシーケンシングライブラリーを調製する方法であって、ライブラリーの各メンバーについて、以下のことを含む方法に関する：本明細書に記載のシーケンシングアダプターの第1および第2の末端を、対向するtopおよびbottom鎖を有する試料DNAフラグメントヘライゲーションすること、それによって、DNAフラグメントおよびシーケンシングアダプターを含む部分的に環状化されたDNA分子が形成される；および、部分的に環状化されたDNA分子の反対の鎖を鋳型として各々使用してシーケンシングアダプターのフリーの3'末端を伸長させることにより第1および第2の一本鎖DNA分子を合成すること、それによって、次世代シーケンシングのために構成された直鎖化された二本鎖DNA分子が形成され、該直鎖化された二本鎖DNA分子は、コピーされたbottom鎖と対をなす元のtop鎖を含む第1の二本鎖領域および元のbottom鎖と対をなすコピーされたtop鎖を含む第2の二本鎖領域を含んでおり、ここで、各々異なるDNAフラグメントから調製された複数の直鎖化された二本鎖DNA分子が、次世代シーケンシングライブラリーを構成する。

10

【0037】

いくつかの態様において、次世代シーケンシングのために構成され、および、第1および第2の末端を有する、直鎖化された二本鎖DNA分子は、以下の構造：

第1の末端－[第1の次世代フローセルアダプター]－[元のbottom鎖のコピーと対をなす元のtop鎖を含む第1の二重鎖領域]－[次世代シーケンシングアダプターの中央部分を含む第2の二重鎖領域]－[元のbottom鎖と対をなす元のtop鎖のコピーを含む第3の二重鎖領域]－[第2の次世代フローセルアダプター]－第2の末端を含む。

20

【0038】

いくつかの態様において、第1の次世代フローセルアダプターは、Illumina P5またはP7アダプター配列である。いくつかの態様において、第2の次世代フローセルアダプターが、Illumina P5またはP7アダプター配列である。いくつかの態様において、第2の二重鎖領域は、第1および第2のリードプライマー結合部位を含み、ここで、各々第1および第2のリードプライマー部位は、ユニークな分子識別子（UMI）およびサンプルインデックス配列とさらに会合する。

【0039】

いくつかの態様において、第1および第2のリードプライマー結合部位は、直鎖化された二本鎖DNA分子の末端へ向かい外向きに配向されている。いくつかの態様において、第1のリードプライマーは、UMI、サンプルインデックス、およびシーケンシングされる試料DNAフラグメントの元のtop鎖またはその部分を含む配列リードを得るために使用することができる。いくつかの態様において、第2のリードプライマーは、UMI、サンプルインデックス、およびシーケンシングされる試料DNAフラグメントの元のbottom鎖またはその部分を含む配列リードを得るために使用することができる。いくつかの態様において、方法は、市販の次世代ライブラリー構築キットの代わりに使用される。いくつかの態様において、上記の第1のステップのライゲーションすることは、リガーゼを添加することを含む。いくつかの態様において、上記の第2のステップの合成することは、DNAポリメラーゼを添加することを含む。いくつかの態様において、ポリメラーゼは、鎖置換活性を有する。いくつかの態様において、方法は、第1および第2のリードプライマーを用いて次世代シーケンシングを実施することにより元の上のおよび元のbottom鎖の配列を得るステップをさらに含む。

30

40

【0040】

本開示の別の側面は、本明細書に記載の方法によって得られる次世代シーケンシングのために構成された直鎖化された二本鎖DNA分子に関し、ここで、直鎖化された二本鎖DNA分子は、第1および第2の末端を含み、および以下の構造を有する：

【0041】

第1の末端－[第1の次世代フローセルアダプター]－[元のbottom鎖のコピーと対

50

をなす元のtop鎖を含む第1の二重鎖領域] - [次世代シーケンシングアダプターの中央部分を含む第2の二重鎖領域] - [元のbottom鎖と対をなす元のtop鎖のコピーを含む第3の二重鎖領域] - [第2の次世代フローセルアダプター] - 第2の末端。

【0042】

いくつかの態様において、第1の次世代フローセルアダプターは、Illumina P5またはP7アダプター配列である。いくつかの態様において、第2の次世代フローセルアダプターは、Illumina P5またはP7アダプター配列である。いくつかの態様において、第2の二重鎖領域は、第1および第2のリードプライマー結合部位を含み、ここで、各々第1および第2のリードプライマー部位は、ユニークな分子識別子(UMI)およびサンプルインデックス配列とさらに会合する。いくつかの態様において、第1および第2のリードプライマー結合部位は、直鎖化された二本鎖DNA分子の末端へ向かい外向きに配向されている。いくつかの態様において、第1のリードプライマーは、UMI、サンプルインデックス、およびシーケンシングされる試料DNAフラグメントの元のtop鎖またはその部分を含む配列リードを得るために使用することができる。いくつかの態様において、第2のリードプライマーは、UMI、サンプルインデックス、およびシーケンシングされる試料DNAフラグメントの元のbottom鎖またはその部分を含む配列リードを得るために使用することができる。

10

【0043】

本開示の別の側面は、DNA試料の次世代シーケンシングのための方法であって、生体源からDNA試料を得ること；複数のDNAフラグメントを得るためにDNA試料をフラグメント化すること；各鎖がDNAフラグメントのtopおよびbottom鎖のコンカテマーを含んでいる複数の直鎖化された二本鎖DNA分子を生成するために本明細書に記載の方法によりDNAフラグメントの次世代シーケンシングライブラリーを構築すること；および直鎖化された二本鎖DNA分子へ結合するリードプライマーを用いた次世代シーケンシングを使用してDNAフラグメントのtopおよびbottom鎖の配列を決定すること（それによって、DNA分子の配列が得られる）を含む、方法に関する。

20

【0044】

いくつかの態様において、生体試料は、血液である。いくつかの態様において、生体試料は、肝臓、腎臓、脳、心臓、皮膚、肺、結腸、または脾臓からの組織の試料である。いくつかの態様において、生体試料は、肝臓、腎臓、脳、心臓、皮膚、肺、結腸、または脾臓からの罹患組織の試料である。いくつかの態様において、罹患組織は、増殖性疾患である。いくつかの態様において、罹患組織は、腫瘍である。いくつかの態様において、シーケンシングエラー率は、Duplex Sequencingに基づく対照と同様であるが、しかしここで必要とされるリードの数は、少なくとも100倍減少している。

30

【0045】

本開示の別の側面は、少なくとも1つのオリゴヌクレオチドが、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための本明細書に記載のとおり単離された核酸複合体に関する。

【0046】

本開示の別の側面は、第1、第2、第3、および第4のオリゴヌクレオチドの各々が、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための本明細書に記載のとおり単離された核酸複合体に関する。

40

【0047】

本開示の別の側面は、少なくとも1つのオリゴヌクレオチドが、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための本明細書に記載のとおりシーケンシングアダプターに関する。

【0048】

50

本開示の別の側面は、第1、第2、第3、および第4のオリゴヌクレオチドの各々が、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている、メチル化シーケンシングの方法における使用のための本明細書に記載のとおり

のシーケンシングアダプターに関する。

【0049】

本開示の別の側面は、本明細書に記載のシーケンシングアダプターの第1および第2の末端を、対向するtopおよびbottom鎖を有する試料DNAフラグメントヘライゲーションすること、それによって、DNAフラグメントおよびシーケンシングアダプターを含む部分的に環状化されたDNA分子が形成され、ここで、シーケンシングアダプターは、メチル化されていないシトシンの代わりにメチル化されたシトシンを含有するように改変されている；および、部分的に環状化されたDNA分子の反対の鎖を鋳型として各々使用してシーケンシングアダプターのフリーの3'末端を伸長させることにより第1および第2の一本鎖DNA分子を合成すること、それによって、各鎖がDNAフラグメントのtopおよびbottom鎖のコンカテマーを含んでいる直鎖化された二本鎖DNA分子が形成され、ここで、合成するステップは、フリーの3'末端を、DNAポリメラーゼ、および、標準のdATP、dGTPおよびdTTPデオキシヌクレオチドとともに用いられているメチル化されたdCTPと、接触させることを含む；DNAフラグメントの、元のtop鎖において、メチル化されていないシトシンをウラシルに脱アミノ化させること；次世代シーケンシングによりtopおよびbottom鎖の配列を決定すること、元のDNAフラグメントにおけるメチル化位置を推測するために配列を比較すること、を含む、DNA試料のメチル化シーケンシングの方法に関する。いくつかの態様において、DNA試料は、生体試料から得られる。いくつかの態様において、生体試料は、肝臓、腎臓、脳、心臓、皮膚、肺、結腸、または脾臓組織から得られ、任意にここで、組織は罹患している。いくつかの態様において、疾患は、増殖性疾患である。いくつかの態様において、疾患は、腫瘍である。

【0050】

いくつかの態様において、dsDNA二重鎖は、上記の第1のステップに先立ちプレ増幅され、方法は、dsDNA二重鎖を第1および第2のプレ増幅分子と接触させること、ここで、2つのプレ増幅分子の各々は、UMI、サンプルインデックス、ローリングサークル増幅（RCA）プライマー、およびトランケーション部位を含む；プレ増幅dsDNA二重鎖を作るために、第1のプレ増幅分子を、dsDNA二重鎖の、1つの第1の末端ヘライゲーションし、および第2のプレ増幅分子を、dsDNA二重鎖の第2の末端ヘライゲーションすること；プレ増幅dsDNA二重鎖をDNAポリメラーゼ酵素に晒すこと；RCAを完了するのに十分な時間の間、プレ増幅dsDNA二重鎖およびDNAポリメラーゼ酵素をインキュベートすること；および、プレ増幅dsDNA二重鎖をトランケーション部位で切断することによりRCAプライマーを除去することを含む。

【0051】

いくつかの態様において、シーケンシングされるDNA二重鎖は、上記の第1のステップに先立ちプレ増幅され、方法は、シーケンシングされるDNA二重鎖の各々を第1および第2のプレ増幅分子と接触させること、ここで、2つのプレ増幅分子の各々は、UMI、サンプルインデックス、ローリングサークル増幅（RCA）プライマー、およびトランケーション部位を含む；複数のプレ増幅DNA二重鎖を作るために、第1のプレ増幅分子を、シーケンシングされるDNA二重鎖の各々の1つの第1の末端ヘライゲーションし、および第2のプレ増幅分子を、シーケンシングされるDNA二重鎖の各々の第2の末端ヘライゲーションすること；プレ増幅DNA二重鎖の各々をDNAポリメラーゼ酵素に晒すこと；RCAを完了するのに十分な時間の間、プレ増幅DNA二重鎖の各々およびDNAポリメラーゼ酵素をインキュベートすること；および、プレ増幅DNA二重鎖の各々をトランケーション部位で切断することによりRCAプライマーを除去することを含む。

【0052】

本開示の別の側面は、次世代シーケンシングライブラリーを調製する方法であって、R06の3'末端およびR05の3'末端を、ライゲーションを受けることからブロックするこ

と；本明細書に記載の複合体を以下のとおりにdsDNA二重鎖ヘライゲーションすること：R01の5'末端をdsDNA二重鎖の第1の鎖の3'末端ヘライゲーションする；およびR10の5'末端をdsDNA二重鎖の第2の鎖の3'末端ヘライゲーションする；それによって、標的DNA分子および複合体を含む環状二本鎖DNA中間体が形成される；R03の3'末端から第1のDNA鎖を伸長させること；R08の3'末端から第2のDNA鎖を伸長させること；および環状の一本鎖シーケンシング分子を形成するために第1および第2のDNA鎖の各々を環状化させること；直鎖の一本鎖シーケンシング分子を形成するためにR03とR08との間の領域中へとニックを導入することを含む、方法に関する。

【0053】

いくつかの態様において、上記の第1のステップのブロックすることは、ブロッキング溶液を添加することを含む。いくつかの態様において、上記の第2のステップのライゲーションすることは、リガーゼを添加することを含む。いくつかの態様において、上記の第3および第4のステップの合成することは、環状二本鎖DNA中間体をポリメラーゼと接触させることを含む。いくつかの態様において、ポリメラーゼは、DNA依存性のDNAポリメラーゼである。いくつかの態様において、ポリメラーゼは、鎖置換活性を有する。いくつかの態様において、次世代シーケンシング（NGS）は、ショートリード戦略である。

【0054】

様々な態様において、CODECライブラリー調製および／またはシーケンシングに先立ち、シーケンシングの標的とされるDNAフラグメントは、従来型のER/AT修復により処理されてもよい。他の態様において、CODECライブラリー調製および／またはシーケンシングに先立ち、シーケンシングの標的とされるDNAフラグメントは、二重鎖修復により処理されてもよい。

【0055】

前述の概念、および以下で論じられる追加の概念は、本開示がこの点で限定されるものではなく、あらゆる好適な組み合わせでアレンジしてもよいということが当然理解されるべきである。さらに、付随する図面と関係づけて考えると、本開示の他の利点および新規な特色は、以下の様々な非限定的な態様の詳細な説明から明らかになる。

【0056】

図面の簡単な説明

以下の図面は本明細書の一部を形成し、これらは本開示の一定の側面をさらに実証するために含まれており、これら図面の1つ以上を本明細書に提示される特定の態様の詳細な説明と組み合わせて参照することにより、さらによく理解することができる。

【図面の簡単な説明】

【0057】

【図1A】図1A～1ARは、エラー訂正のための元の二重鎖の連結（Concatenating Original Duplex for Error Correction）（CODEC）の概観およびCODECの検証を示す。図1Aは、標準的なNGSワークフローには（例として、伝統的な二重鎖シーケンシングのように）、遺伝情報を2回コードするDNAの固有の特性が失われるDNA二重鎖の解離が関与するということを示す。塩基損傷、PCR、およびNGSエラーにより引き起こされる偽の突然変異を同定するためのユニークな分子識別子（UMI）を通じて二重鎖の両方の鎖を追跡することができる一方、他の数十億の鎖の中からそれらを見つけるにはスループットがかかり、それはクラスターによって強調されている。CODECワークフローは、各シーケンシングリードを得る前に各二重鎖を物理的に繋ぎ合わせて、各配列リードがライブラリー中の各DNAフラグメントについての連結されたtopおよびbottom鎖の配列を提供することから、各ライブラリー分子が配列リード毎に両方の鎖の情報を保持することを確実にする。

【図1B～1C】図1Bは、CODECが元の二重鎖の配列情報を一本鎖へと繋ぎ合わせることを、すなわち、各一本鎖配列リードがライブラリー中の各DNAフラグメントについての連結されたtopおよびbottom鎖の配列を提供することを示す。結果として、NGSリードの各対は、二重鎖コンセンサスを形成するのにそれ自体で十分となる（囲み枠）。それは

、ライゲーションのために二重鎖アダプターに代えてアダプター複合体を利用し、鎖置換伸長がこれに続く。図1Cは、CODECではライゲーションベースのNGSワークフローのアダプターライゲーションステップを改変することを示す。

【0058】

【図1D】図1Dは、CODECアダプター複合体が、Illumina NGSのために必要とされる全ての構成要素（リードプライマー結合部位、フローセルバインダー領域（すなわち、NGSアダプター）、UMIおよびインデックス領域、ならびにDNAフラグメントへのライゲーションを容易にするためのdTテールを包含する）とともに予めパッケージ化されたものを示す。標準的なNGSライブラリーと違って、CODECは、UMI、インデックス、およびインサートを一緒にシーケンシングするために外側へ向かって読み取る。インデックス付きのプライマーはインデックスとして必要とされず、および、フローセル結合領域（P5およびP7）がライゲーションによって付加される。図1Bに示されているとおり、CODECアダプター（左のコンストラクト）は、DNAフラグメント（topおよびbottom鎖を有する）の各末端へライゲーションされ、それによって、CODECアダプターの各末端へ配列結合されたDNAフラグメントを包含する部分的に環状化された部分的に二本鎖のDNA中間体（図1Bを参照）が作られる。部分的に環状化された中間体、部分的に二本鎖の中間体は、次いで、環状化された中間体のアダプター部分に位置する中央二重鎖領域のフリーの5'末端から伸長するDNAポリメラーゼでの鎖置換伸長を受ける。DNAポリメラーゼは、5'末端の各々から伸長することで一本鎖DNAを合成する。

10

【図1E-1F】図1Eは、アダプターの二本鎖領域が20℃にて500nMのオリゴヌクレオチド濃度および10mMのNaで安定を保つと予測されることを示す。図1Fは、標的二重鎖の曲げ剛性を軽減するように一本鎖リンカーの長さが決定されたことを示す。

20

【0059】

【図1G-1I】図1Gは、一方が鎖1からおおよばもう一方が鎖2から作り出された2つの二重鎖を含有し、その間のリンカーおよび両末端のNGSアダプターを伴う、CDS産物を示す。図1Hは、CDSが、インサートとアダプター複合体との間の環状ライゲーションで始まるということを示す。伸長は、次いで、開放3'末端から始まって鎖置換活性を持つポリメラーゼにより行われる。図1Iは、アダプターライゲーションステップを置き換えることにより、CDSを、全ゲノムシーケンシング（WGS）、全エクソームシーケンシング（WES）、または標的化シーケンシングの従来型のワークフロー中へと統合することができることを示す。

30

【図1J-1K】図1Jは、Illumina MiSeq（2×300bp）でのWGSで、全リードの56.7%が正しい構造を有しており、予想どおり、そのコンセンサスエラー比は未加工の比率の二乗に類似していたことが確認されたということを示す。図1Kは、鎖2を同調した様式でシーケンシングする追加のCDSリードプライマーを用いると、リード1の間の各サイクルでデュアル蛍光が生成されるということを示す。2つの鎖の間のあらゆる不一致は、低いスコアによってマークされる。

【0060】

【図1L-1M】図1Lは、バリエーション連結二重鎖シーケンシング（CDS）を示す概略図である。図1Mは、CDSとIlluminaワークフローとの統合を示す概略図である。

40

【図1N-1P】図1Nは、ミスマッチバブルバリエーションを伴う長い二重鎖を示す。図1Oは、ミスマッチバブルバリエーションを伴うモジュラー二重鎖を示す。図1Pは、ハーフアダプター複合体バリエーションを示す。

【図1Q-1S】図1Qは、UMIバリエーションを示す。図1Rは、部分的なリードプライマー結合部位としての領域2および3を伴うバリエーションを示す。図1Sは、完全リードプライマー結合部位としての領域2および3を伴うバリエーションを示す。

【0061】

【図1T-1U】図1Tは、領域1をインデックスとするバリエーションの形成を示す概略図である。図1Uは、連結された構造をCDSアダプター複合体がそれによって作り出すメカニズムを示す概略図である。

50

【図1V】図1Vは、CDSを示す概略図である。

【図1W】図1Wは、CDS構造が、NGSの質に影響を及ぼす単一インサート副産物を無視するということを示す概略図である。

【0062】

【図1X】図1Xは、混在クラスターをもたらす、ブリッジ増幅の間の単一インサート副産物形成のメカニズムを示す。

【図1Y】図1Yは、図1Wに描かれた単純な連結アプローチと同様にNGSリードプライマー結合部位が外側の端にあるときの混在クラスター形成の証拠を示す。Qスコアが、リードにおける位置に対してプロットされ、インサートの末端は垂直な線によりマークされており、CDSリンカー配列とSIアダプター配列との間で共通している塩基は赤いドットでアノテーションされている。共通塩基でのより高い塩基品質スコアは、CDSおよびSI副産物からの混在した蛍光を意味する。

10

【図1Z】図1Aは、図1Wに示されているとおりの「単純な連結」対CDSについてのインサートの後で読み取られた領域におけるメジアンQスコアを示す。単純な連結を用いると、ユニークな塩基についての対CDSリンカーとp7アダプターとの間の共通の塩基と比べて低いメジアンQスコアは、単一インサート副産物も読み取られるということの意味する。一方、CDSを用いると、インサートが読み取られた後の領域における高いメジアンQスコアは、単一インサート副産物が今度は混在クラスターから「無視」されるようになっているということの意味する。

【0063】

20

【図1AA】図1AAは、CDSがインデックスをインサートのすぐ隣に、および試料調製においてより早期に付着させるということを示す概略図である。

【図1AB-1AC】図1ABは、新規なCDSアダプター複合体に向けられた請求項の標的二本鎖DNAの次世代シーケンシングのためのCDSアダプター複合体を示す概略図であり、図1ACは、標的二本鎖DNAをシーケンシングするための二重鎖を示す概略図である。

【図1AD】図1ADは、標的二本鎖DNAをシーケンシングするための二重鎖を形成する方法を示す概略図である。

【0064】

【図1AE1】図1AEは、標的二本鎖DNAの次世代シーケンシングの方法を示す概略図である。

30

【図1AE2】図1AEは、標的二本鎖DNAの次世代シーケンシングの方法を示す概略図である。

【図1AF1】図1AFは、CDSの方法および組成物がDuplex-Repairと組み合わせられてもよいということを示す。

【図1AF2】図1AFは、CDSの方法および組成物がDuplex-Repairと組み合わせられてもよいということを示す。

【0065】

【図1AG】図1AGは、二重鎖シーケンシングを示す概略図である。二重鎖シーケンシングは、伝統的なシーケンシングよりも1000倍正確であり得、および、真の突然変異は同じDNA二重鎖の両方の鎖に存在するという前提に基づいて機能する。それはNGSリードを各々の元のDNA分子に繋ぎ戻せるようにする分子バーコードを伴う標準的なアダプターを使用することで、「二重鎖コンセンサス」を形成する。しかしながら、それは各二重鎖の両方の鎖を「見つける」ために約100倍多いNGSリードを必要とし、これはエクソーム／ゲノムについては不可能であり；遺伝子パネルにとって著しく限定される。

40

【図1AH-1AJ】図1AH~1AJは、CDSのメカニズムを示す。

【図1AK-1AL】図1AKは、連結された二重鎖シーケンシング（CDS）が、各二重鎖の両方の鎖を、それらが単一のリード対内で一緒にシーケンシングされることができるよう繋ぎ合わせることを示す。図1ALは、大半のNGSワークフローに参与する主要なステップを示す。

【図1AM】図1AMは、Duplex-Repairはアダプターライゲーションに先立ち鎖再合

50

成を限定し、およびよって、市販のER/AT方法で起こるような両方の鎖でコピーされる塩基損傷エラーの潜在的可能性を限定するということを示す。dsDNAの長さは、それが一本鎖であるときと比較してその軸に沿って短い。最大174bpの二重鎖は、全く曲がることなしに収容されることができる。

【図1AN】図1ANは、Duplex-Repair、Duplex-Repair v2対従来型ER/AT方法の概観を示す。

【0066】

【図1AO】図1AOは、キャピラリー電気泳動によって決定した、Duplex-Repairの各ステップおよび従来型ER/ATに供された様々な合成二重鎖の主な産物の概略を示す。合成分子のフルオロフォアタグなしの末端が示されており、フラグメントのサイズは一定の縮尺で示される。アスタリスク（*）で区切られた二重鎖はフルオロフォアを含有しないため、キャピラリー電気泳動で直接観察されなかった；ただしそれらの存在は、UDGおよびFPGの特徴的な活性により予測されている。鎖再合成の領域は薄い青で示される。

10

【図1AP】図1APは、ddPCRアッセイを使用することによるDNAインプットの関数としてのDuplex-Repair対KAPA HyperPrepキットの測定されたライブラリー変換効率を示す。

【図1AQ-1AR】図1AQは、二重鎖プレ増幅が、鎖識別子、ユニークな分子識別子（UMI）、およびサンプルインデックスを包含する各々の元の二重鎖の複数コピーを作り出すということを示す。エンドヌクレアーゼ消化を使用して、各々の元の二重鎖のコピーが各アンプリコンから解放され、およびCODEC鎖の繋ぎ合わせに使用できる状態になる。図1ARは、CODEC v2が今度はアダプターオリゴヌクレオチドを別々にライゲーションしてアダプター複合体を後から組み立てるということを示す。この新しいライゲーション作戦を使用することにより、各二重鎖の片方の鎖のみをライゲーションするために3'末端がブロックされたオリゴヌクレオチドを利用して第1の2つのアダプターオリゴヌクレオチドをライゲーションし、ブロックされたオリゴヌクレオチドを第2のライゲーションのための残りのアダプターオリゴヌクレオチドで置換することがこれに続く。アダプターブロッカーを除去することは、アダプター複合体を組み立てることを可能にし、これは鎖置換伸長のための鋳型として使用することができる。

20

【0067】

【図2】図2は、CODECアダプター複合体設計の背後にある理論、ならびに標準的なNGSおよびCODECのリードプライマー結合部位を示す。

30

【図3A】図3A~3B。図3Aは、NGSフローセルに対するクラスター生成サイクルの間に、インサート領域の途中における早期停止は、1つのみのインサートおよびリードプライマー結合領域を伴うより短いフラグメントへと変わる副産物を作り出し得るということを示す。これらのサブクローンフラグメントは、共通領域が終わるまで正しいフラグメントと同じ配列を有する。シーケンシングサイクルが共通領域を過ぎると、短いフラグメントは混在した蛍光を引き起こし、および、帰結としては低い品質スコアを生じる。

【図3B】図3Bは、ランダムに選択された100リード対からの共通領域中の最後の150bpおよび共通領域の後の最初の50bpを取ることに伴う各シーケンシングサイクルの平均品質スコアを示す。アダプター構造を設計し直す前は、品質スコアは共通領域の後で急に落ち、このことが、リードがCODEC構造を有するか否かを確認することを難しくしていた。この問題は、リンカーなしの全ての副産物を「サイレンシング」するためにリードプライマー結合領域をリンカーへ移動させたことによって解決した。

40

【0068】

【図4】図4は、UMIと4つのインデックスの各セットとが、集合的に各位置で4つ全ての塩基を包含しつつ、Illuminaシーケンサーの高品質な画像分析のために、同様のハイブリダイゼーション6-G（図1AL）を保つように設計されていることを示す。例えば、Illuminaソフトウェアは、クラスター同定、位相補正、および純度フィルター（chastity filter）などの様々な目的のために最初の25bpまでを使用する。配列は、上から下へ、配列番号19~26に対応する。

50

【図5A】図5A～5B。図5Aは、正しいCODEC産物およびそれらがどのようにして作り出されたかによって名付けられた副産物の比率を示す。

【図5B】図5Bは、副産物形成の予想されるメカニズムを示す。「ダブルライゲーション」は、2つのアダプター複合体がインサートの各末端ヘライゲーションされ、および、A/Tライゲーションとは反対に互いのT/Tミスマッチライゲーションを経たときに引き起こされ得る。「ブランクライゲーション」は、2つのアダプター複合体がインサートなしで両方の末端で互いのT/Tミスマッチライゲーションを経たときに引き起こされ得る。「分子間」は、鎖置換伸長が他のライゲーション産物を反対鎖リンカーに代えて鋳型として使用したときに引き起こされ得る。

【0069】

【図6A～6B】図6A～6Bは、概念実証を示す。図6Aは、CODEC、Duplex Sequencing、ならびに、典型的なペアエンドリード（R1+R2）および一本鎖コンセンサス（SSC）を包含する他のコンセンサス方法のエラー率を示す。汎がん遺伝子パネルを用いた標的富化を、2名の個人のセルフリーDNA（cfDNA）に対して行った。エラーバーは、95%二項信頼区間を意味する。図6Bは、各ファミリーサイズでの、同じUMIおよび開始・停止位置での未加工リードの数であるエラー率を示す。

【図6C～6D】図6Cは、汎がんパネルを用いると、2名の個人のセルフリーDNA（cfDNA）に適用したとき、CDSがDuplex Sequencingに匹敵する一塩基変異（SNV）エラー率を示し、それが典型的なペアエンドリード（R1+R2）または一本鎖コンセンサス（SSC）よりもはるかに低かったということを示す。エラーバーは、二項信頼（95%）区間を意味する。図6Dは、より少ない未加工のリードを伴うときさえも、CDSはより高いユニーク深度（3.96）を有しており、他方、Duplex Sequencingはほぼゼロのユニーク深度（0.025）を有していたということを示す。線は、累積的なフラクション（cumulative fractions）を意味する。

【図6E～6F】図6Eは、CDSのSNVエラー率が依然としてDuplex Sequencingのそれに匹敵するということを示す。図6Fは、最小アレル閾値が1であるとき、CDSが、リコールを維持しながらペアエンドリードよりも秀でた精度を示したということを示す。

【0070】

【図7】図7は、インプット試料のオーバーハング上の、ウラシルである脱アミノ化されたシトシンは、末端修復および鎖置換伸長を受けたということを示す。慎重に使用されたPhi29 DNAポリメラーゼは、HiFiポリメラーゼと違ってウラシルを認識することができ、および、後続するPCRにおいて増幅することができる鎖（クリック鎖）を作り出し得たものである。試料が高いレベルの脱アミノ化を有していた場合には、ウラシルからの偽陽性を抑制するためにCODECワークフローへUSER酵素ステップが加えられた。

【図8A】図8A～8Cは、標的とされたパネルのシーケンシングにおけるCDSの特徴付けを示す。図8Aは、配列コンテキストの関数として、汎がんパネルについてのエラー率を示す。健常ドナーからのCODECのC>Tエラー率は、Duplex Sequencingのそれよりも高かった。

【図8B～8C】図8Bは、健常ドナーのcfDNAにおいて、CDSはユニークな元の二重鎖を汎がんパネルシーケンシングにおいてDuplex Sequencingの350倍早く回収し始めたことを示す。実践は、移動平均を示し、および、影は、標準偏差を意味する。図8Cは、各ファミリーサイズでの、同じUMIおよび開始・停止位置での未加工リードの数であるCODECエラー率を示す。CDSは、低いSNVエラー率を達成するのに単一のペアリードのみを必要とし（すなわち、ファミリーサイズ=1）、一方、同じ元のDNA分子から複数CDSリードのコンセンサスを形成することは（すなわち、ファミリーサイズ>1）、エラー率への影響をほとんど有しなかった。

【0071】

【図9A】図9A～9Fは、Duplex Sequencingと比較した二重鎖コンセンサスデータを示す。図9Aは、二重鎖コンセンサスデータにおいて、両方のフラグメント末端からの、中心領域のそれらよりも高い12bpの平均エラー率は、Duplex Sequencingを使用し

10

20

30

40

50

た他の研究において以前に観察されたものである末端修復の前の $5'$ オーバーハングでの塩基損傷を示唆することを示す。これは、末端修復が $5'$ オーバーハングをフィルインし、および1つの鎖上の損傷した塩基を両方の鎖へコピーし、そして偽の二重鎖コンセンサスを作り出すからである。対照的に、SSCは、オーバーハングでも二重鎖領域でも塩基損傷を訂正せず、およびよって、最後の12bpと中心領域との間でより少ないエラー率の差を示す。

【図9B】図9Bは、CDSが、単一のライブラリー分子内で両方の鎖を繋ぎ合わせることで、両方を単一のリード対で読み取れることができるようにするということを示す。

【図9C-9D】図9Cは、271のセルフフリーDNA試料からの、標準的なNGS（コンセンサス＝なし）、一本鎖コンセンサスシーケンシング、および二本鎖（または二重鎖）コンセンサスシーケンシングにおけるエラー率の比較を示す。図9Dは、エラー率対配列あたりのリード数を示す。

【図9E-9F】図9Eは、シミュレーションされた20ngのDNAの二重鎖シーケンシングについてのリード深度に対する二重鎖回収、対、各リード対がユニークなDNA二重鎖を反映したとすると理論的に何を達成可能であるかを示す。図9Fは、271のcfDNA試料対2個のホルマリン固定パラフィン包埋（FFPE）腫瘍生検についての総二重鎖エラー率を示す。

【0072】

【図10A】図10A~10Lは、CODECベースのシーケンシングの有効性を示す。図10Aは、FFPE試料に対するWESの、全体としてのエラー率およびそれらの塩基コンテキストを示す。

【図10B-10D】図10B~10Dは、CDSがDuplex-Repairにおける原因不明のエラーを解消できるということを示す。

【図10E-10I】図10E~10Iは、CDSがDuplex-Repairにおける原因不明のエラーを解消できるということを示す。

【図10J-10L】図10Jは、4つのシーケンシング技法：CODEC、Duplex Sequencing、標準的なNGS、およびPacBio HiFiにおける、全ゲノムシーケンシング（WGS）コスト対エラー率を示す。PacBio HiFiの精度中央値は、製品パンフレットに基づいてQ30（99.9%）とした。残りのデータはBroadで生成し、および、シーケンシングしたコストは、Broad Genomic Platform価格に基づいてIllumina NovaSeq S4およびPacBio Sequel IIeに対して算出した。標準的なNGSの正確性は、最小Q30塩基品質でのR1+R2のコンセンサス正確性に基づくものとした。図10Kは、WGSデータ内の少なくとも所与のカバレッジでカバーされた全塩基の割合を意味する累積分布を示す。CODECと標準的なWGSとは12xでの平均カバレッジでマッチした。図10Lは、血液正常試料の全エクソームシーケンシング（約40Mb）におけるCODEC対標準的なNGSのエラー率を示す。左側は全体のエラー率であり、および、右側は、モノヌクレオチド配列のコンテキストによって分割されたものである。エラーバーは、二項信頼（95%）区間を意味する。

【0073】

【図11A-11C】図11A~11Cは、Genome in a Bottle ConsortiumのパイロットゲノムNA12878の全ゲノムシーケンシング（WGS）を示す。図11Aは、種々の方法のエラー率およびシーケンシングコストを示す。PacBio HiFiデータは、技術的仕様を使用した。図11Bは、CODECおよびDuplex Sequencingの各ユニーク二重鎖深度のフラクションを示す。図11Cは、より低い深度へダウンサンプリングされたときのCODECおよびR1+R2の偽陽性および偽陰性を示す。

【0074】

【図12A-12B】図12A~12Bは、患者データをシーケンシングするためのCODECの使用からのデータを示す。図12A~12Bは、NA12878試料へのWGSの、全体のエラー率およびそれらの塩基コンテキストを示す。

【図12C-12E】図12Cは、ペアエンドリード（R1+R2）で同じ鎖を2回読み

取ることがエラー率を4倍しか改善しなかったが、他方、元のtopおよびbottom鎖を両方読み取ること（二重鎖シーケンシングおよびCDS）はそれを1100倍改善したということを示す。エラーバーは、二項信頼（95%）区間を意味する。図12Dは、CDSが二重鎖シーケンシングよりも少ないリードでより効率的に元の二重鎖を回収したことを示し、および、そのより低いプラトーは、CDS+ハイブリッドキャプチャー富化ワークフローのためにさらなる最適化が必要であることを示唆する。点線は、図9Eにおけるシミュレーションされた曲線を意味する。図12Eは、Parsons et al.のとおり76部位を標的とした患者特異的アッセイに適用したとき、CDSは完全な特異性でがん患者のcfDNAから首尾よく体細胞突然変異を検出したということを示す。水平な線、囲み枠、およびひげは、メジアン、25~75%範囲、および5%~95%範囲をそれぞれ意味する。

10

【0075】

【図13A-13B】図13A~13Cは、モノヌクレオチドマイクロサテライトでのインデルを示す。図13Aは、NA12878のモノヌクレオチドマイクロサテライトでのインデルエラー頻度のまとめを示す。図13Bは、8~18ヌクレオチドの異なる長さのモノヌクレオチドマイクロサテライトでのインデルエラー頻度を示す。

【図13C】図13Cは、マイクロサテライト不安定性（MSI）検出限界を示す。MSIを伴う結腸がん患者の腫瘍および正常試料をシーケンシングし、およびin silicoで希釈した。

【0076】

【図14A】図14A~14Iは、MSI試料に対するWGSからのトリヌクレオチドコンテキストおよびCatalogue Of Somatic Mutations In Cancer（COSMIC）シグネチャーを示す。図14Aは、標準的なNGSは複数分子から高い存在量の突然変異のみ検出できるがバックグラウンドノイズによって不明瞭になった低い存在量の突然変異を検出できないということを示す。CODECは、その単一の二重鎖の分解能に起因して、高いおよび低い存在量の突然変異の両方をコールすることができる。

20

【図14B】図14Bは、バリエーションコーラーであるMutect2を用いてまたはそれなしで突然変異を閾値化した後の突然変異コンテキストを示す。高い存在量の突然変異のみを選択することが、標準的なNGSのためのゴールドスタンダードとなっていた。各バーは、トリヌクレオチドコンテキストを表す。

【図14C-14D】図14Cは、12×のカバレッジ（破線の囲み枠）での標準的なNGSからMutect2により選択された高い存在量の突然変異に対するコサイン類似性を示す。各方法は、そのコサイン類似性の有意な低下が観察されるまでより低い深度へダウンサンプリングされた。図14Dは、CODECにより検出されたがMutect2により選択されなかった突然変異率を示す。

30

【図14E】図14Eは、突然変異の種々の群から抽出されたCOSMIC一塩基置換（SBS）シグネチャーを示す。「Mutect2によりコールされていない」の下のグループは、「全ての突然変異」の下の対応するグループのサブセットである。

【0077】

【図14F】図14Fは、血漿cfDNAからの腫瘍全エクソームの適格性確認およびシーケンシングについての以前に記載されているワークフローを示す（Adalsteinsson et al. Nat Comms 2017）。

40

【図14G】図14Gは、ステージIV乳がんおよび前立腺がんの520患者における腫瘍由来cfDNAの推定フラクションを示し、血漿全エクソームシーケンシングに十分な腫瘍含量を有するのは33~45%のみであるということが示されている。

【図14H1】図14Hは、cfDNAの全エクソームシーケンシングと、cfDNAにおける腫瘍フラクションが>0.1である患者からのマッチした腫瘍生検との間での、クローンおよびサブクロンの腫瘍突然変異におけるオーバーラップを示す。

【図14H2】図14Hは、cfDNAの全エクソームシーケンシングと、cfDNAにおける腫瘍フラクションが>0.1である患者からのマッチした腫瘍生検との間での、クローンおよびサブクロンの腫瘍突然変異におけるオーバーラップを示す。

50

【図 1 4 I】図 14I は、転移性乳がんの患者におけるがん進行および進化をモニタリングするために使用される連続全エクソームシーケンシングの実証を示し、これは、選択的エストロゲン受容体分解剤による処理に反応した薬剤耐性の収束的進化（例として、多重 ESR1 突然変異）であり得るものの出現を特定する。

【0 0 7 8】

【図 1 5】図 15 は、低い存在量の突然変異（低いバリエーションアレル頻度（variant allele fraction））を検出する能力を CODEC と種々のカバレッジでの標準的な WGS（30×、60×、および 80×）との間で比較した二項モデルを示す。標準的な WGS は、エラー訂正のために少なくとも 2 つのユニークなフラグメントを必要とした。よって、このモデルは、シーケンシングエラーを無視した。0.3% 未満の VAF では、CODEC は、30x の標準的な WGS よりも良好な検出倍率を示した。0.03% 未満の VAF では、CODEC は、より高い深度の標準的な WGS のいずれのものよりも秀でた感受性を示した。

10

【図 1 6】図 16 は、CDS が DNA メチル化情報を保持および報告できるようにするために開発されたプロトコルを示す概略図である。

【0 0 7 9】

【図 1 7 A】図 17A～17D は、Kapa HyperPrep キットを使用した、ER/AT の間の鎖再合成の定量化を示す。図 17A は、ER/AT の間のフィルインされた塩基を定量化するための方法の概略図を示す。

【図 1 7 B】図 17B は、測定されたパルス間持続時間（IPD；フレーム単位）を、5 つの合成オリゴヌクレオチドの塩基位置の関数として示す。より長い IPD（60 フレームを超える場合は灰色）は、修飾塩基によって生じる。破線は、ER/AT の間にフィルインの開始が予想される場所を示す。

20

【図 1 7 C】図 17C は、健常ドナー cfDNA 試料について、測定された IPD を、塩基位置の関数として示す。

【図 1 7 D】図 17D は、広範な鎖再合成を受けた 4 つの強調された二重鎖を示す。

【0 0 8 0】

【図 1 8 A】図 18A～18I は、Duplex Repair の、従来型 ER/AT との比較を示す。図 18A は、Duplex-Repair アプローチの性能を、複数の異なる合成オリゴヌクレオチドについてキャピラリー電気泳動により決定し、従来型 ER/AT と比較して示す（i～v i i）。

30

【図 1 8 B－1 8 C】図 18B は、様々な量の DNase I（ニックの誘導のため）および CuCl₂/H₂O₂（酸化損傷の誘導のため）で処理した健常ドナーの cfDNA に適用した、Duplex-Repair 対市販の ER/AT および IDT xGEN「汎がん」パネルを使用して測定した、二重鎖シーケンシングのエラー率を示す。図 18C は、ホルマリン固定腫瘍 DNA の修復のために Duplex-Repair 対従来型 ER/AT を使用した後の、二重鎖シーケンシングのエラー率を示す（なお、Duplex-Repair 試料のエラー率が広いのは、シーケンシングされた合計の二重鎖が少ないことに起因している）。

【図 1 8 D－1 8 E】図 18D は、カスタムの単分子シーケンシングアッセイを使用して測定された、従来型 ER/AT および Duplex-Repair のいくつかのバリエーションを使用して再合成された内部塩基対（元の二重鎖フラグメントの両端から >12bp）の推定フラクションを示す。図 18E は、3 つの試料タイプにわたり従来型 ER/AT および Duplex-Repair の両方について、再合成された内部塩基対の推定フラクションを示す。

40

【0 0 8 1】

【図 1 8 F－1 8 G】図 18F は、従来型 ER/AT または Duplex-Repair で処理した、4 つの健常 cfDNA 試料（条件当たり 3 つの複製）、3 つのがん患者 cfDNA 試料（条件当たり 1 つの複製）、および 5 つのがん患者 FFPE 腫瘍生検（条件当たり 3 つの複製）の、二重鎖シーケンシングエラー率を示す。図 18G は、集計の変異塩基および元の二重鎖フラグメントの末端に対するそれらの位置を示す。破線はフラグメント内部の閾値（12bp）を表す。

【図 1 8 H－1 8 I】図 18H は、様々な濃度の DNase I（ニックを誘発する）および Cu

50

Cl₂/H₂O₂（酸化損傷を誘発する）で損傷され、その後、Duplex-Repairまたは従来型ER/ATを使用して修復したHD_78 cfDNAの、測定された二重鎖シーケンシングエラー率を示す（条件当たり3つの複製）。図18Iは、cfDNAおよびFFPE試料タイプに対する従来型ER/ATおよびDuplex-Repairの比較を示し、リードのin silicoダウンサンプリングを介した分析により、リード対の数の関数としての同等の二重鎖回収を示す。

【0082】

【図19A】図19A～19Bは、CODECシーケンシングにより対処され得る非限定的な課題を示す。図19Aは、シーケンシングがより安価になってきたがより正確にはなっていないことを示す。これは、生物医学研究および診断における全てのタイプのDNAシーケンシングに対して深刻な意義を有する。

【図19B】図19Bは、CDSが全てのタイプのDNAシーケンシングを「クリーンアップ」する潜在的可能性を示す。

【図20】図20は、CODECアダプターライゲーションおよびCODECシーケンシングに先立ち、核酸試料（例として、DNA試料）に対して二重鎖プレ増幅が実施されてもよいということを図解する。

【図21】図21は、改変されたCODECシーケンシングアダプターを使用したシーケンシングの態様を図解する。

【発明を実施するための形態】

【0083】

詳細な説明

本開示は、二重鎖シーケンシングを改善する本明細書中において「エラー訂正のための元の二重鎖の連結（Concatenating Original Duplex for Error Correction）」または「CODEC」と称する新規なDNAシーケンシング方法、ならびに、該新規なシーケンシング方法を実施するための組成物（例として、ライブラリー作成のためのマルチオリゴヌクレオチドアダプター、アダプターコンストラクト、およびシーケンシングライブラリー）、アダプターを作製するための方法、ライブラリー構築のための方法、および、より低いコストで二重鎖シーケンシングの正確性を改善する二重鎖シーケンシング方法を提供する。様々な側面において、CODECアダプターを使用したライブラリー調製は、各DNA分子が二重鎖コンセンサスを形成するのにそれ自体で十分となり、真の突然変異の同定を容易にし、および偽の突然変異を回避することを結果としてもたらす。

【0084】

様々な側面において、本開示は、各DNA二重鎖の両方の鎖を直鎖の配列に連結する強力な新しいライブラリー構築方法を提供する。両方の鎖を物理的に繋ぎ合わせることで、産物は、二重鎖コンセンサスを形成するのにそれ自体で十分である。この戦略は、最低限の追加コストにて1000倍正確なシーケンシングを提供することへの潜在的な可能性を有し、そしてGenomics Platformで提供されている既存の製品（WGS、WES、標的化されたパネル）を直接的に強化することができる。

【0085】

様々な側面において、本開示は、CODECシーケンシングのための方法、ならびに、アダプター（本明細書中様々な態様において「CODECアダプター」と称する）、シーケンシングされるDNAフラグメントの両末端へライゲーションされたCODECアダプターを各々含む環状化された中間体（本明細書中様々な態様において「CODEC環状化中間体」と称する）、および、シーケンシングされる単一DNAフラグメントの連結されたtopおよびbottom鎖を含む直鎖化された二本鎖産物（本明細書中様々な態様において「CODECライブラリー」または個々に「CODECライブラリーメンバー」と称する）を包含する、CODECシーケンシングに必要とされるおよび／またはCODECシーケンシングによって作られる組成物を提供する。様々な態様において、CODECアダプターは、NGSワークフロー（例として、NGSフローセルに対するクラスター増幅）のためのNGSアダプター、DNAフラグメントの両方の鎖を読み取るためのシーケンシングリードプライマー部位、および任意には、1つ以上のサンプルインデックスおよび1つ以上のユニークな分子識別子

10

20

30

40

50

(UMI) を包含する。

【0086】

シーケンシングされる各DNAフラグメントについてtopおよびbottom鎖の配列を別々に得る（および、よって、topおよびbottom鎖の配列を同定、マッチング、および比較するためにコンピューター的アプローチを必要とする）伝統的な二重鎖シーケンシングと違って、CODECアダプターを使用したライブラリー形成ではシーケンシングされる元の各DNAフラグメント（すなわち、同じDNA分子内）の上下の配列のコンカテマーを各鎖が含む二本鎖ライブラリーメンバーを結果としてもたらすため、CODECライブラリーメンバーの各々は、同じリードにおいて二重鎖コンセンサスを形成するのにそれ自体で十分である。よって、CODECアダプターのシーケンシングは、top鎖、bottom鎖、および任意には、1つ以上のサンプルインデックスおよび1つ以上のUMIを含む、シーケンシング産物を結果としてもたらす。標準的な二重鎖シーケンシングと比較して、このアプローチの技術的利点は、2つの別々のシーケンシング産物（すなわち、上の配列に対して1つおよび下の配列に対して1つ）を生成する標準的な二重鎖シーケンシングと違って、CODECシーケンシングは、上の配列および下の配列の両方を含む単一のシーケンシング産物を結果としてもたらし、それによって、ユーザーが簡単に真の突然変異（シーケンシングリードの上および下の両方に現れる突然変異）を偽の突然変異（シーケンシングリードの上のみまたは下のみに現れる突然変異）から判別することを可能にするということである。

10

【0087】

本明細書において別段の定義がない限り、本明細書で使用される全ての技術用語および科学用語は、本発明が属する技術分野の当業者によって一般的に理解されるのと同じ意味を有する。本明細書に記載のものと類似または等価のあらゆる方法および材料を、本発明の実施または試験に使用することができるが、好ましい方法および材料について説明する。

20

【0088】

本明細書で参照される全ての特許および刊行物（かかる特許および刊行物内に開示される全ての配列を含む）は、参照により明示的に組み込まれる。

【0089】

数値範囲は、範囲を定義する数値を含む。別段の指示がない限り、核酸は左から右に5'から3'の方向で記載される；アミノ酸配列はそれぞれ、左から右にアミノからカルボキシの方向で記載される。

30

【0090】

本明細書で提供される見出しは、本発明の様々な側面または態様を限定するものではない。したがって、すぐ下で定義される用語は、明細書全体を参照することによってより完全に定義される。

【0091】

別段の定義がない限り、本明細書で使用される全ての技術用語および科学用語は、本発明が属する技術分野の当業者によって一般に理解されるのと同じ意味を有する。Singleton, et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 3RD ED., John Wiley and Sons, New York (2006)、およびHale & Markham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, N.Y. (1991)は、本明細書で使用される多くの用語の一般的な意味を、当業者に提供する。ただし特定の用語については、明確さおよび参照の容易さのために、以下に定義する。用語の意味および範囲は明らかであるが、潜在的な曖昧さがある場合には、本明細書で提供される定義が、あらゆる辞書または外部の定義より優先される。さらに、文脈により別段の要求がない限り、単数形の用語には複数形が含まれ、複数形の用語には単数形が含まれるものとする。本開示において、別段の記載がない限り、「または」の使用は「および／または」を意味する。さらに、「含むこと」という用語、ならびに「含む」および「含まれる」などの他の形式の使用は、限定的ではない。また、具体的に別段の記載がない限り、

40

50

「要素」または「構成要素」などの用語は、1つのユニットを含む要素および構成要素と、2つ以上のサブユニットを含む要素および構成要素の両方を包含する。

【0092】

一般に、本明細書に記載の細胞および組織培養、分子生物学、免疫学、微生物学、遺伝学、ならびにタンパク質および核酸の化学およびハイブリダイゼーションに関連して使用される命名法、およびそれらの技法は、当技術分野でよく知られており、一般に使用されているものである。本開示の方法および技法は、一般に、当技術分野で周知の従来の方法に従って、および、別段の指示がない限り、本開示全体にわたって引用および議論される様々な一般的およびより具体的な参考文献に記載されているように実施される。酵素反応および精製技法は、製造業者の仕様書に従って、当技術分野で一般的に達成されるように、または本明細書に記載のように実施される。本明細書に記載の分析化学、有機合成化学、および医薬品および創薬化学に関連して使用される命名法、ならびにそれらの実験手順および技法は、当技術分野でよく知られ一般的に使用されているものである。標準的な技法は、化学合成、化学分析、医薬品の調製、製剤化、および送達、ならびに対象の処置に使用される。

10

【0093】

用語「およそ」または「約」は本明細書で互換的に使用され、1つ以上の興味ある値に適用される場合、記載された基準値に類似する値を指す。ある態様において、用語「およそ」または「約」は、記載された基準値のいずれかの方向の15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、またはそれ未満内（すなわち、これより大きいパーセンテージまたは小さいパーセンテージ）に入る値の範囲を指すが、ただし、別段の記載がない限り、または文脈から明らかでない限りにおいてである（例えば、かかる数値が可能な値の100%を超える場合）。

20

【0094】

本明細書で使用され得る用語「dAテーリング」は、非鋳型アデノシン（A）（例として、アデノシンーリン酸）を含む「テール」を有する核酸（例として、DNA、RNA）の状態または特徴を指す。「テール」とは、核酸（例として、DNA、RNA）の3'末端のアデノシン（例として、A A A A A）が、相補鎖の5'末端ヌクレオチドを越えるオーバーハングを含むことを意味する。用語（例として、dAテール）は、アデノシンが核酸の3'末端に付加されるプロセスを説明する動詞（例として、dAテーリング）として使用される場合があるいくつかの態様において、dAテーリングは、3'→5'エキソヌクレアーゼ活性を欠くKlenowフラグメントを用いて実施される。いくつかの態様において、dAテーリングは、Taqポリメラーゼを用いて実施される。

30

【0095】

本明細書で使用され得る用語「オーバーハング」は、反対側の鎖（例として、相補鎖）の末端（例として、末端ヌクレオチド）を越えて伸びる（例として、突き出る）二本鎖核酸の部分の指すと当業者に知られている技術用語を指す。例えば、これに限定されないが、5'オーバーハングは、それと結合して二本鎖核酸二重鎖を形成する反対側の鎖（例として、相補鎖）の3'末端（3'末端ヌクレオチド）を越えて伸びる核酸の鎖の部分の指すであろう。さらなる例として、限定されないが、3'オーバーハングは、それと結合して二本鎖核酸二重鎖を形成する反対側の鎖（例として、相補鎖）の5'末端（5'末端ヌクレオチド）を越えて伸びる核酸の鎖の部分の指すであろう。当業者に理解されるように、二本鎖二重鎖（double-stranded duplex）は、5'および3'オーバーハングの両方、単一の5'オーバーハング、2つの5'オーバーハング、単一の3'オーバーハング、2つの3'オーバーハング、1つのオーバーハング（例として、5'または3'）と1つの平滑末端、または2つの平滑末端を含み得る。本明細書で使用される用語「平滑末端」は、二本鎖二重鎖の性質を指し、ここで二重鎖を形成する2つの鎖は同じヌクレオチド対で終結し、したがって二重鎖のその末端にオーバーハングを有さない（例として、末端は平滑である）。

40

【0096】

50

本明細書で使用され得る用語「エキソヌクレアーゼ」は、核酸（例として、ポリヌクレオチド、オリゴヌクレオチド）の末端からヌクレオチドを切断する活性を少なくとも有する酵素を指すことが当業者に一般に知られている技術用語を指す。いくつかの態様において、エキソヌクレアーゼは、ヌクレオチドを一度に1つずつ切断する。エキソヌクレアーゼは、核酸のいずれかの方向（例として、5'末端からまたは3'末端からのいずれか）でヌクレオチドを切断することができる。かかる活性の説明において、ヌクレオチドを核酸の5'末端から開始して（例として、3'末端の遠位にある5'ヌクレオチド）切断するエキソヌクレアーゼを指す場合、多くの場合表記は5'→3'エキソヌクレアーゼ活性と示され、または、ヌクレオチドを核酸の3'末端から開始して（例として、5'末端の遠位にある3'ヌクレオチド）切断するエキソヌクレアーゼを指す場合、3'→5'エキソヌクレアーゼ活性と示される。いくつかの態様において、エキソヌクレアーゼは5'→3'エキソヌクレアーゼ活性を有する。いくつかの態様において、エキソヌクレアーゼは、Exo VIIであり得る。

【0097】

用語「相補的」および「相補性」は、本明細書において互換的に使用され得るように、鎖（例として、オリゴヌクレオチド）内の核酸（例として、RNA、DNA）におけるヌクレオチド（例として、A、C、G、T、U）の特性であって、反対方向の核酸鎖（例として、平行に走っているが逆方向（すなわち、5'-3'が3'-5'と整列する、および3'-5'が5'-3'と整列する））内の別の特定のヌクレオチドと対合する（すなわち、ワトソンとクリックの塩基対合ルール）ところの、前記特性を指す。デオキシリボ核酸（DNA）に関して、相補的塩基対合は、アデニン（A）とチミン（T）（例として、AとT、TとA）、グアニン（G）とシトシン（C）（例として、GとC、CとG）であり、リボ核酸（RNA）に関して、相補的塩基対合は、Aとウラシル（U）（例として、AとU、UとA）、およびGとC（例として、GとC、CとG）である。これは、各塩基対がその相補的な塩基（例として、A-T/U、T/U-A、C-G、G-C）と同数の水素結合を形成する能力によって生じ、例えばグアニンとシトシンの間の結合は、常に2つの水素結合を共有するA-T/U結合と比べて、3つの水素結合を共有する。

【0098】

核酸の対の少なくとも一方の鎖の全ての塩基が、その相補的塩基対の反対側にある場合、かかる鎖はもう一方の鎖の配列に対して完全に相補的であるとみなされる。かかる鎖の1つ以上の塩基が、その相補的な塩基対を除くあらゆる他の塩基の反対側の位置にある場合、その塩基は「ミスマッチ」とみなされ、鎖は部分的に相補的であるとみなされる。したがって、鎖は、整列する塩基がなくなるまで、様々な程度の部分相補性を示すことができ、整列した時点でそれらは非相補的となる。

【0099】

CODECアダプター、ライブラリー調製、およびシーケンシング

様々な側面において、本開示は、CODECシーケンシングのための方法、ならびに、アダプター（本明細書中様々な態様において「CODECアダプター」と称する）、シーケンシングされるDNAフラグメントの両末端ヘライゲーションされたCODECアダプターを各々含む環状化された中間体（本明細書中様々な態様において「CODEC環状化中間体」と称する）、および、シーケンシングされる単一DNAフラグメントの連結されたtopおよびbottom鎖を含む直鎖化された二本鎖産物（本明細書中様々な態様において「CODECライブラリー」または個々に「CODECライブラリーメンバー」と称する）を包含する、CODECシーケンシングに必要とされるおよび／またはCODECシーケンシングによって作られる組成物を提供する。様々な態様において、CODECアダプターは、NGSワークフロー（例として、NGSフローセルに対するクラスター増幅）のためのNGSアダプター、DNAフラグメントの両方の鎖を読み取るためのシーケンシングリードプライマー部位、および任意には、1つ以上のサンプルインデックスおよび1つ以上のユニークな分子識別子（UMI）を包含する。

【0100】

10

20

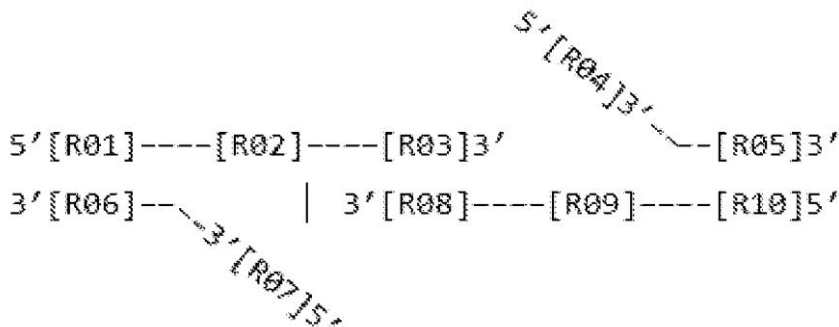
30

40

50

いくつかの態様において、CODECアダプター複合体は、連結とアダプター付着との両方に要求される各要素全てを包含する4つのハイブリダイゼーションされたオリゴヌクレオチドからなる。いくつかの態様において、CODECアダプター複合体は、以下の構成にある、少なくとも10個の領域(R01~R10)を含む：

【化2】



10

【0101】

いくつかの態様において、「----」は、結合を表す。いくつかの態様において、R01、R02、およびR03は、第1のオリゴヌクレオチドを含み、R04およびR05は、第2のオリゴヌクレオチドを含み、R06およびR07は、第3のオリゴヌクレオチドを含み、R08、R09、R10は、第4のオリゴヌクレオチドを含む。いくつかの態様において、R01とR06とは、相互にアニーリングされており、R03とR08とは、相互にアニーリングされており、R05とR10とは、相互にアニーリングされており、R02とR07とは、相互にアニーリングされておらず、および、R04とR09とは、相互にアニーリングされていない。

20

【0102】

いくつかの態様において、CODECアダプター複合体は、標的ニ重鎖（標的DNA分子）の一方の末端とライゲーションされ（アダプターライゲーション）、環状化された産物を作るための他方の末端間のライゲーションがこれに続く。本明細書で使用され得る用語「アダプターライゲーション」は、ヌクレオチド（例として、核酸、オリゴヌクレオチド、例えばアダプター）の既知の配列を、1つ以上の核酸（例として、DNAフラグメント、DNAの相補鎖）の1つ以上の末端に付着（例として、ライゲーション）するプロセスを一般に指すことが当業者に知られている用語を指す。多くの場合、アダプターは、それらが結合することが意図されている核酸フラグメントに相補的な特定の配列を含有するが、例えば限定されないが、核酸がdAテール付きの場合、アダプターは「T」オーバーハングを有し得、ここで「T」は、チミン核酸塩基を含むヌクレオチドを指す。TオーバーハングはdAテールに相補的であるため、ライゲーションが容易になる。用語「相補的」および「相補性」は、本明細書において互換的に使用され得るように、鎖（例として、オリゴヌクレオチド）内の核酸（例として、RNA、DNA）におけるヌクレオチド（例として、A、C、G、T、U）の特性であって、反対方向の核酸鎖（例として、平行に走っているが逆方向（すなわち、5' - 3' が3' - 5' と整列する、および3' - 5' が5' - 3' と整列する））内の別の特定のヌクレオチドと対合する（すなわち、ワトソンとクリックの塩基対合ルール）ところの、前記特性を指す。デオキシリボ核酸（DNA）に関して、相補的塩基対合は、アデニン（A）とチミン（T）（例として、AとT、TとA）、グアニン（G）とシトシン（C）（例として、GとC、CとG）であり、リボ核酸（RNA）に関して、相補的塩基対合は、Aとウラシル（U）（例として、AとU、UとA）、およびGとC（例として、GとC、CとG）である。これは、各塩基対がその相補的な塩基（例として、A - T / U、T / U - A、C - G、G - C）と同数の水素結合を形成する能力によって生じ、例えばグアニンとシトシンの間の結合は、常に2つの水素結合を共有するA - T / U結合と比べて、3つの水素結合を共有する。核酸の対の少なくとも一方の鎖の全ての塩基が、その相補的塩基対の反対側にある場合、かかる鎖はもう一方の鎖の配列に対して完全に相補的であるとみなされる。かかる鎖の1つ以上の塩基が、その相補的な塩基対を

30

40

50

除くあらゆる他の塩基の反対側の位置にある場合、その塩基は「ミスマッチ」とみなされ、鎖は部分的に相補的であるとみなされる。したがって、鎖は、整列する塩基がなくなるまで、様々な程度の部分相補性を示すことができ、整列した時点でそれらは非相補的となる。他の非標準ヌクレオチド（例として、5-メチルシトシン、5-ヒドロキシメチルシトシン）は当技術分野で知られており、それらの特性および相補性は当業者には容易に明らかであろう。

【0103】

いくつかの態様において、R01は、第1の連結二重鎖シーケンシング（CDS）アダプターを含み；R02は、一本鎖リンカー、第1のユニークな分子識別子（UMI）、および第1のリードプライマー部位を含み；R03は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端にあるかまたはその付近にある第1の配列を含み；R04は、第1の次世代シーケンシング（NGS）アダプター配列を含むフリーの5'末端を含み；R05は、第3のCDSアダプターおよび第1のサンプルインデックスを含み；R06は、第2のCDSアダプターおよび第2のサンプルインデックスを含み；R07は、第2の次世代シーケンシング（NGS）アダプター配列を含むフリーの5'末端を含み；R08は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端にあるかまたはその付近にある第2の配列を含み；R09は、一本鎖リンカー、第2のUMI、および第2のリードプライマー部位を含む；および／またはR10は、第4のアダプターを含む。

10

【0104】

本明細書で使用され得る「ポリメラーゼ」という用語は、核酸（例として、DNAポリメラーゼ、RNAポリメラーゼ）およびポリマーの合成を助けるかまたはそれらを合成する酵素を一般に指すことが当業者に知られている技術用語である。多数のポリメラーゼが知られており、例えば、限定することなく、これらは全て本明細書で企図される；DNAポリメラーゼI（Polガンマ、Polシータ、Polニュー）、DNAポリメラーゼII（Polアルファ、Polデルタ、Polイプシロン、Polゼータ）、DNAポリメラーゼIIIホロ酵素、DNAポリメラーゼIV（DinB）（SOS修復ポリメラーゼ、Polベータ、Polラムダ、Polミュー）、DNAポリメラーゼV（SOSポリメラーゼ、Polエータ、Polイオータ、Polカップ）、逆転写酵素、およびRNAポリメラーゼ（RNA Pol I、RNA Pol II、RNA Pol III、T7 RNA Pol、RNAレプリカーゼ、プライマーゼ）。さらに企図されるのは、細菌（例として、*Thermus aquaticus*）由来のポリメラーゼである。例えば、*Thermus aquaticus*からのTaqは、ポリメラーゼ連鎖反応（PCR）で使用される一般的なDNAポリメラーゼである。いくつかの態様において、ポリメラーゼは、Taqポリメラーゼである。いくつかの態様において、ポリメラーゼは、3'→5'エキソヌクレアーゼ活性を欠いている。いくつかの態様において、ポリメラーゼは、Klenowフラグメントである。いくつかの態様において、ポリメラーゼは、3'→5'エキソヌクレアーゼ活性を欠くKlenowフラグメントである。いくつかの態様において、ポリメラーゼは、本明細書に記載されるポリメラーゼのいずれかのヒトバリエーションである。

20

30

【0105】

様々な態様において、例示のCODECアダプターオリゴヌクレオチド配列が、例1の表2に提供される。

40

【0106】

用語「ユニークな分子識別子（UMI）」は、シーケンシングの間、エラー訂正および増大した正確さを提供する、短いオリゴヌクレオチド分子バーコードを指す。

【0107】

用語「核酸」、「ヌクレオチド配列」、「ポリヌクレオチド」、「オリゴヌクレオチド」、および「ヌクレオチドのポリマー」という用語は、本明細書において互換的に使用され得るように、少なくとも2つの、核酸塩基-糖-リン酸塩の組み合わせ（例えば、ヌクレオチド）のストリングを指し、とりわけ、一本鎖DNAおよび二本鎖DNA、一本鎖領域と二本鎖領域の混合物であるDNA、一本鎖RNAおよび二本鎖RNA、および一本鎖領域と

50

二本鎖領域の混合物であるRNA、ハイブリッド分子であって、一本鎖、より典型的には二本鎖、または一本鎖領域と二本鎖領域の混合物であり得るDNAおよびRNAを含む前記ハイブリッド分子である。さらに、本明細書で使用される用語（例として、核酸など）は、RNAまたはDNA、またはRNAとDNAの両方を含む、三本鎖領域を指し得る。かかる領域の鎖は、同じ分子に由来するものであっても、異なる分子に由来するものであってもよい。領域は、1つ以上の分子の全てを含み得るが、より一般的には、分子の一部の領域のみを含む。三重らせん領域の分子の1つは、しばしばオリゴヌクレオチドと呼ばれる。

【0108】

用語（例として、核酸など）はまた、化学的、酵素的、または代謝的に改変された核酸の形態、ならびに単純および複雑な細胞を含む、ウイルスおよび細胞に特徴的なDNAおよびRNAの化学的形態も包含する。例えば、本明細書で使用される用語（例として、核酸など）は、1つ以上の修飾塩基を含む本明細書に記載のDNAまたはRNAを含むことができる。核酸はまた、以下も含み得る：天然ヌクレオシド（すなわち、アデノシン、チミジン、グアノシン、シチジン、ウリジン、デオキシアデノシン、デオキシチミジン、デオキシグアノシン、およびデオキシシチジン）、ヌクレオシド類似体（例として、2-アミノアデノシン、2-チオチミジン、イノシン、ピロロピリミジン、3-メチルアデノシン、5-メチルシチジン、C5プロモウリジン、C5フルオロウリジン、C5ヨードウリジン、C5プロピニルウリジン、C5プロピニルシチジン、C5メチルシチジン、7デアザアデノシン、7デアザグアノシン、8オキソアデノシン、8オキソグアノシン、O（6）メチルグアニン、4-アセチルシチジン、5-（カルボキシヒドロキシメチル）ウリジン、ジヒドロウリジン、メチルシュドウリジン、1-メチルアデノシン、1-メチルグアノシン、N6-メチルアデノシン、および2-チオシチジン）、化学的修飾塩基、生物学的修飾塩基（例として、メチル化塩基）、インターカレート塩基、修飾糖（例として、2'-フルオロリボース、リボース、2'-デオキシリボース、2'-O-メチルシチジン、アラビノース、およびヘキソース）、または修飾リン酸基（例として、ホスホロチオアートおよび5'-Nホスホロアミダイト結合）。したがって、2つの例のみを挙げると、イノシンなどの異常な塩基を含むDNAまたはRNA、またはトリチル化塩基などの修飾塩基は、本明細書で使用される用語としての核酸である。用語（例として、核酸など）はまた、ペプチド核酸（PNA）、ホスホロチオアート、およびネイティブな核酸のリン酸主鎖の別のバリエーションも含む。天然の核酸はリン酸主鎖を有し、人工核酸は他の種類の主鎖を含有することができるが、含有される塩基は同じである。したがって、安定性または他の理由で主鎖が改変されたDNAまたはRNAは、その用語が本明細書で意図されているように、核酸である。

【0109】

本明細書で使用され得る「核酸塩基」という用語は、窒素含有塩基として当業者に知られている技術用語であり、ヌクレオシドの構成要素を形成する窒素含有生物学的化合物であり、それ自体はヌクレオチドの構成要素である。核酸塩基（本明細書では単に塩基とも呼ばれる）は、塩基対を形成しかつ互いに積み重なって長鎖らせん構造を形成する能力を有するため、核酸（例えば、DNA、RNA）の基本構成ブロックの1つである。5つの標準的な核酸塩基が存在する：アデニン（A）、シトシン（C）、グアニン（G）、チミン（T）、およびウラシル（U）であり、A、C、G、およびTはDNAに見出され、A、C、G、およびUはRNAに見出される。

【0110】

本明細書で使用され得る用語「ヌクレオシド」は、リン酸基のないヌクレオチドであることが一般に知られているグリコシルアミン（例として、N-グリコシド）を指す。ヌクレオシドは、核酸塩基（例として、窒素塩基）と五炭糖（例として、ペントース）から構成される。五炭糖は、リボースまたはデオキシリボースのいずれかであり得る。ヌクレオシドは、RNAおよびDNAの構成成分であるヌクレオチドの生化学的前駆体である。ヌクレオシドの例には、シチジン（C）、ウリジン（U）、アデノシン（A）、グアノシン（G）、チミジン（T）、およびイノシン（I）が含まれるが、バリエーション（例として、改

変または合成ヌクレオシド、改変または合成核酸塩基を含有するヌクレオシド) も含まれる。

【0111】

本明細書で使用され得る「ヌクレオチド」という用語は、一般に、核酸塩基、糖、およびリン酸塩(例として、ヌクレオシドおよびリン酸塩)を含む組成物を指すことが当業者に知られている技術用語である(これらの組成物(例として、ヌクレオチド)は、プリンとピリミジンに分離される)。ヌクレオチドは、ポリメラーゼを用いてコピーすることができる核酸の構成要素である。ヌクレオシドであるシチジン(C)、ウリジン(U)、アデノシン(A)、グアノシン(G)、チミジン(T)、およびイノシン(I)は、リン酸基と共に標準ヌクレオチドを表し、合成反応で使用される個々のヌクレオチド(例として、3つのリン酸基を有するヌクレオチド(例として、「三リン酸」))を指す場合、DNA形態(例として、デオキシリボースを有する)において、dATP、dGTP、dCTP、およびdTTPと呼ばれ得る。リン酸基のうちの2つを加水分解すると、核酸の重合に使用する一リン酸ヌクレオチドが得られる。一般に、dATP、dGTP、dCTP、およびdTTPは、dNTPと呼ばれることがあり、ここで「N」は、ヌクレオシドの性質に関する曖昧さを表す。したがってdNTPの混合物は、それぞれの全部または一部の濃度を含み得る。ヌクレオチドは、既知のプリン塩基およびピリミジン塩基のみでなく、損傷を受けた他の複素環塩基(例として、酸化、メチル化、アシル化、脱アデニル化された塩基など)も含有する。この用語は当技術分野ではよく知られており、当業者には容易に理解されるであろう。

10

20

【0112】

様々な態様において、4つのCODECアダプターオリゴヌクレオチドは、シーケンシングされるDNAフラグメントとのライゲーションの前にアニーリングされ(すなわち、プレアニーリングされ)てもよい。様々な他の態様において、4つのCODECアダプターオリゴヌクレオチドは、ライゲーションステップの間にまたはこれと同時にアニーリングされてもよい。

【0113】

4つのオリゴヌクレオチドをライゲーションの前にプレアニーリングすることの利点は、両末端が必ず異なるアダプターを得ることであり、他方、ハイブリダイゼーションなしでのライゲーションは、標的の50%が、環状化されることができない、両側での同じアダプターとのライゲーションをすることを結果としてもたらす。いくつかの態様において、収率を改善するために、単一A/Tオーバーハングがライゲーション部位で付加される。いくつかの態様において、DNA平滑末端またはDNA付着末端が付加される。いくつかの態様において、一本鎖DNA領域は、環状化のための柔軟性を加えるためにCODEC複合体中へと組み込まれている。

30

【0114】

いくつかの態様において、第1の配列および第2の配列は、同じかまたは異なるプライマー結合部位をさらに含む。いくつかの態様において、第1および第2のプライマー部位は、反対方向の付加によってシーケンシングを開始するように配向されている。いくつかの態様において、第1および第2のUMIは、はっきり区別できるものである。

【0115】

40

いくつかの態様において、R01は、12~30個の間のヌクレオチドを含む、R02は、14~75個の間のヌクレオチドを含む、R03は、12~99個の間のヌクレオチドを含む、R04は、20~49個の間のヌクレオチドを含む、R05は、12~30個の間のヌクレオチドを含む、R06は、12~30個の間のヌクレオチドを含む、R07は、20~49個の間のヌクレオチドを含む、R08は、12~99個の間のヌクレオチドを含む、R09は、14~75個の間のヌクレオチドを含む、および/またはR10は、12~30個の間のヌクレオチドを含む。

【0116】

いくつかの態様において、R01およびR06は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイ

50

ブリダイゼーション自由エネルギーを含み；R03およびR08は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、約-35kcal/mol、約-40kcal/mol、約-45kcal/mol、約-50kcal/mol、約-55kcal/mol、約-60のハイブリダイゼーション自由エネルギーを含み；および／またはR05およびR10は、約-10kcal/mol、約-15kcal/mol、約-20kcal/mol、約-25kcal/mol、約-30kcal/mol、または約-35kcal/molのハイブリダイゼーション自由エネルギーを含む。

【0117】

いくつかの態様において、R01およびR06は、同じ数のヌクレオチドを各々含み、任意にここで、R06は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを有する；R03およびR08は、同じ数のヌクレオチドを各々含む；および／または、R05およびR10は、同じ数のヌクレオチドを各々含み、任意にここで、R05は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを有する。

10

【0118】

いくつかの態様において、R01とR06とは、少なくとも90%の相補性がある配列を含み；R03とR08とは、少なくとも90%の相補性がある配列を含み；および／またはR05とR10とは、少なくとも90%の相補性がある配列を含む。

【0119】

いくつかの態様において、各々R01、R06、R05、およびR10は、同じ数のヌクレオチドを含み、任意にここで、R06およびR05は、ライゲーションを容易にするための1ヌクレオチドオーバーハングを各々有する。

20

【0120】

いくつかの態様において、R01は、第1の連結二重鎖シーケンシング（CDS）アダプターを含み；R02は、一本鎖リンカーを含み；R03は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端を含み；R04は、第1のUMIを含み；R05は、第3のCDSアダプターを含み；R06は、第2のCDSアダプターを含み；R07は、第2のUMIを含み；R08は、DNA依存性のDNAポリメラーゼによるDNA合成をプライミングすることが可能な3'末端を含み；R09は、一本鎖リンカーを含み；およびR10は、第4のCDSアダプターを含む。

【0121】

いくつかの態様において、R01の5'末端は、標的DNA二重鎖の第1の鎖の3'末端ヘライゲーションされ；R05の3'末端は、標的DNA二重鎖の第1の鎖の5'末端ヘライゲーションされ；R10の5'末端は、標的DNA二重鎖の第2の鎖の3'末端ヘライゲーションされ；R06の3'末端は、標的DNA二重鎖の第2の鎖の5'末端ヘライゲーションされ；環状化されたDNA二重鎖または任意には部分的に二本鎖の環状DNAを形成している。

30

【0122】

いくつかの態様において、CODECアダプター複合体は、NGSのために調製されおよび研究または臨床目的（例として、対象における突然変異の同定、疾患の診断）のために使用されてもよい。本明細書で使用される用語「対象」は、本明細書の主題を用いた処置または診断を必要とするあらゆる生物を指す。例えば、限定されないが、対象は哺乳動物および非哺乳動物を含み得る。いくつかの態様において、対象は哺乳動物である。いくつかの態様において、対象は非哺乳動物である。本明細書で使用される場合、「哺乳動物」とは、哺乳綱を構成するあらゆる動物（例として、ヒト、マウス、ラット、ネコ、イヌ、ヒツジ、ウサギ、ウマ、ウシ、ヤギ、ブタ、モルモット、ハムスター、ニワトリ、シチメンチョウ、または非ヒト霊長類（例として、マーモセット、マカク））である。いくつかの態様において、哺乳動物はヒトである。

40

【0123】

本明細書で使用され得る用語「突然変異」は、野生型配列と比較した場合の、核酸中のヌクレオチドに対する変化、変更、または改変を指す。例えば、限定されないが、突然変異は、置換、挿入、欠失、またはそれらのあらゆる組み合わせを含み得る。いくつかの態様において、少なくとも1つの突然変異が存在する。いくつかの態様において、複数の突

50

然変異が存在する。いくつかの態様において、複数の突然変異が存在する場合、突然変異は別個である（例えば、同じ種類ではない（例として、置換、挿入、欠失））。いくつかの態様において、複数の突然変異が存在する場合、それらの突然変異は同一である（例えば、同じ種類ではない（例として、置換、挿入、欠失））。さらに、いくつかの態様において、突然変異はフレームシフトを引き起こす。

【0124】

突然変異は、上記のように、野生型核酸とは異なる、所定の核酸（例として、DNA、RNA）の領域（例として、セクション、部分、核酸塩基、ヌクレオシド、ヌクレオチド）であり、ほとんどの場合、核酸の各々の鎖に反映される。すなわち、試料に突然変異が存在する場合、その突然変異とその相補体が、シーケンシング時に核酸の各鎖で観察されるだろう。しかし、試料が一本鎖部分（例として、ギャップ、オーバーハング）または鎖再合成を引き起こす可能性のある領域（例として、ニック）を含有し得ることを考慮すると、これは問題である。この問題は、損傷塩基がそのような一本鎖領域または再合成される他の領域に存在する場合、損傷塩基がその相補鎖の合成に、それから試料が生成された核酸には元々存在しなかった塩基を含めるよう指示し得るために、発生する（損傷塩基は非標準塩基対合に影響を与える可能性があるため）。1本の鎖がミスマッチ塩基を含有する場合にも、同じことが起こり得る。かかる場合、ミスマッチは、そのネイティブのミスマッチ塩基ではなく、再合成された補体中で対合したマッチ（paired match）を示す。これが起こると、両方の鎖のシーケンシングはそれぞれの鎖において突然変異を読み取り、突然変異を示し、しかしこの突然変異は、元の核酸を正確に反映していない可能性がある。かかる突然変異は、本明細書では「偽の突然変異」と呼ばれる。偽の突然変異は、核酸の相補鎖の再合成から生じる突然変異であり、試料が得られた元の（例として、ネイティブ、野生型）核酸の相補鎖を表さない。

【0125】

いくつかの態様において、CODECアダプター複合体の調製または方法は、標的DNA分子の次世代シーケンシング（NGS）における使用のための標的二本鎖DNA分子（dsDNA二重鎖）を調製する方法であって、請求項1～21のいずれか一項に記載の複合体を以下のとおりにdsDNA二重鎖ヘライゲーションすること：R01の5'末端をdsDNA二重鎖の第1の鎖の3'末端ヘライゲーションする；R05の3'末端をdsDNA二重鎖の第1の鎖の5'末端ヘライゲーションする；R10の5'末端をdsDNA二重鎖の第2の鎖の3'末端ヘライゲーションする；およびR06の3'末端をdsDNA二重鎖の第2の鎖の5'末端ヘライゲーションする；それによって、標的DNA分子および複合体を含む環状二本鎖DNA中間体が形成される；R03の3'末端から第1のDNA鎖を伸長させること；R08の3'末端から第2のDNA鎖を伸長させること；および任意に、標的DNA分子の次世代シーケンシング（NGS）における使用のための二本鎖DNA分子を形成するために第1および第2のDNA鎖をアニーリングすることを含む、方法であってもよい。いくつかの態様において、二本鎖DNA分子は、標的DNA分子の2つのコピーを含む。いくつかの態様において、ライゲーションするステップは、リガーゼを添加することを含む。いくつかの態様において、合成するステップは、環状二本鎖DNA中間体をポリメラーゼと接触させることを含む。本明細書で使用され得る用語「接触された」は、1つの物質（例として、酵素、試薬、dNTP）の別の物質（例として、試料、混合物）への暴露であって、ある量および意図で、すなわち、2つの物質が相互作用して、一方の物質の活性が他方の物質（例として、試料に作用する酵素）に影響を与えるようにすること、または2つの物質が相互作用することを意図しての、前記曝露を記述するために使用される。この用語は、2つの物質間の物理的接触を必要とするものと解釈されるべきではないが、さらに物理的接触を禁止するものでもない。例えば、物質間の相互作用および／または活性に影響を与えるのに、近接性は十分であり得る。いくつかの態様において、接触は、物質を同じ容器（例として、反応容器）に導入することによって達成される。いくつかの態様において、接触は、物質を同じ反応容器に導入することによって達成される。いくつかの態様において、接触は、物質A（例として、試薬、dNTP、酵素など）を、物質B（例として、試料）を含有するか、物

質 B が同時に導入されるか、または物質 B が後で導入される反応容器に導入することによって達成される。いくつかの態様において、接触は、物質が互いに物理的に接触する（例として、物理的に相互作用する）ときに達成される。いくつかの態様において、接触は、物質が互いに化学的に相互作用するときに達成される。いくつかの態様において、接触は、物質が互いに酵素的に相互作用するときに達成される。いくつかの態様において、接触は、物質が互いに近接しているときに達成される。

【0126】

いくつかの態様において、ポリメラーゼは、DNA依存性のDNAポリメラーゼである。いくつかの態様において、ここで、ポリメラーゼは、鎖置換活性を有する。いくつかの態様において、次世代シーケンシング（NGS）は、ショートリード戦略である。いくつかの態様において、方法は、次世代シーケンシングにより二本鎖DNA分子をシーケンシングすることを含む。

10

【0127】

いくつかの態様において、CODECアダプター配列は、R05およびR06 Illuminaアダプター（図1K）を作製することにより、Illumina NGSライブラリー構築ワークフローへ統合されることができる。インデックスは、NGSのためにプールされているデマルチプレックス試料へ付着する。

【0128】

他の態様において、本明細書に記載のCODECアダプターは、1以上の改変を包含してもよい。限定なしに、以下は、本明細書に記載のCODECシーケンシング方法に関連して使用され得る改変を表す：

20

【0129】

1. ミスマッチバブルを伴う長い二重鎖

図1Lに示されるこのバリエーションは、それがライゲーション後に領域4、5および6へと切断される必要があることを除き、基本のバージョンと同じように働く。最初にオリゴが2つのみであることで、より簡単に全てのコンポーネントと一緒に持つようになる。

【0130】

2. ミスマッチバブルを伴うモジュラー二重鎖

図1Mに示されるこのバリエーションは、無傷のアダプターを組み立てるために最初にライゲーションされる必要があることを除き、バリエーション4と同じように働く。

30

【0131】

3. ハーフアダプター複合体

4つ全てのオリゴをプレアニーリングすることは、CDSのために必然ではない。それらを2つのハーフアダプター複合体へとアニーリングし、続いてライゲーションをすることは、理論的には50%が領域4および4'を持つことを結果としてもたらす。ひとたびかかる構造が形成されると、領域4および4'は、最終的に、ライゲーションまたは鎖置換伸長の間のどこかの時点で互いにハイブリダイズする（図1N）。

【0132】

4. UMI

ユニークな分子識別子（UMI）は、領域1の一部としてライゲーション部位に導入されることができる（図10）。

40

【0133】

5 A. 部分的なリードプライマー結合部位としての領域2および3

領域2および3の主な目的は環状化のための柔軟性を加えることであるが、これらは、他の機能を有するものとして別目的で用いることもできる。図1Pは、領域2、3および4で正しい産物のみを読み取るためにそれらを部分的なリードプライマー結合部位として使用することを示す。

【0134】

これは、一部の副産物が従来型のNGS試料とちょうど同じように単一インサートのみを有し、そして領域2および3を使用することはそれらがリードプライマーとハイブリダ

50

イズすることを防ぐからである（図1P、「単一インサート（副産物）」）。

【0135】

しかしながら、通常のCDSアダプターおよびこのバリエーション5Aは両方とも、リードプライマーの3'末端がハイブリダイズできる部位を鎖内に2つ有することに悩まされる可能性がある（図1P、「デュアル蛍光」）。これは、2つの異なるプライマーが、データ分析を複雑化させるデュアル蛍光を生成することを引き起こし得る。図1Qに示されているバリエーションは、この問題を解決する。

【0136】

5B. 完全リードプライマー結合部位としての領域2および3

このバリエーションは、リードプライマー結合領域を完全に領域2および3内へと移すことによって、デュアル蛍光の問題に対処する（図1Q）。リードプライマーが今度は領域1とハイブリダイズしないので、それらの3'末端配列はユニークなものである。

【0137】

このバージョンのもう1つの利点は、UMIを導入することへの低いコストである。通常のNGSアダプターおよびCDSアダプターの両方で、バリエーション1には、標的フラグメントとのライゲーションの前に二本鎖アダプター領域の末端でそれが必要とする。UMIが長さ3bpである場合、 $4^3 = 64$ 対のアダプターオリゴが合成されおよびあらゆるUMIミスマッチを回避するため別々にアニーリングされなければならない、これは金額および時間の面で高価である。このバリエーションは、この要件を回避するためにUMIを一本鎖領域2および3内に配置することができる。UMI位置に混在した塩基があることで、あらゆる長さのUMIが単一バッチ中で合成され得る。

【0138】

新しいリードプライマー結合領域が領域1とオーバーラップしないため、全てのリードが同じ時間に領域1に入る場合に各シーケンシングサイクルでの塩基多様性は低くなる。これは、4つのオリゴを異なる長さのUMIと混合することまたは次のバリエーションを使用することにより、解決することができる。

【0139】

6. インデックスとしての領域1

アダプター複合体は必ずしも両側に同じ領域1を有するものではなく、独立した領域1aおよび1bがあり得る。（図1R）。バリエーション5Bと組み合わせられて、このバリエーションは領域1aおよび1bをサンプルインデックスとして使用することができ、インデックス付きのプライマーの必要性がなくなる。この例は、標的配列の隣にインデックスを直接付着させることで、インデックスホッピングとして知られている試料間のクロストークを低減させる。

【0140】

領域1をインデックスとして使用することはまた、先に言及した塩基多様性の問題にも対処することができる。複数インデックスが集合的に位置毎に4つ全ての塩基を有するとき、プールされたNGSライブラリーは領域1全体を通して完全な塩基多様性を得ることになる。

【0141】

（1）正確性の高い、直接リピートのシーケンシングを達成するために「混在クラスター」を克服すること

2本の鎖の連結が成功することは正確性が高く手ごろなNGSのために十分に見えるものの、いずれかの端に同じアダプター配列を持つ1本の鎖を含む副産物（本明細書中、単一インサート、SIと称する）が形成される可能性がある（図1U）。この危険性は、2要素から成る：（1）シーケンシングリードプライマーが末端アダプター領域に対して向けられた場合、SI対CDS分子からのフォワードおよびリバースリードを判別することが困難になる、および（2）SIライブラリー分子からの高いエラー率（0.1~1%）を考慮すると、SIリードのごく一部でもCDSリードとして誤分類されることが有害になる可能性がある。

10

20

30

40

50

【0142】

SI副産物は3つの主要なメカニズムによって形成され得るということが、ここで見出されている：(A)アダプターライゲーションが不完全である場合、すなわち、4つ全てのホスホジエステル結合が形成されない場合は、Phi29伸長（例として、図1S）、(B)CDS産物の直接リピート配列間の相同性を考慮すると、ライブラリー増幅におけるPCRジャンピング、および(C)フローセルに対するブリッジ増幅におけるPCRジャンピング（図1V）。(A)および(B)は、シーケンシングに先立つサイズ選択により、および、リンカー配列の「証拠」を必要とし、例として、インサート後にそれを検出するのに十分長いリードを使用することによって、幾分軽減することができる。しかしながら、いずれも(C)に対処するには十分でない。実際に、CDSフラグメントのブリッジ増幅では、クラスターを播種した元のCDSライブラリー分子および直接リピート配列の一方または両方から生成されたSI副産物を含む混合クラスターが形成されることが、ここで発見されている（図1V）。単一の「播種（seeding）」ライブラリー分子からのブリッジ増幅の対数線形的な性質を考慮すると、(i)CDS分子の、(ii)「top」の鎖のSI副産物に対する割合、(iii)「bottom」の鎖のSI副産物に対する割合は、数桁以上歪んでいる可能性がある。NGSリードプライマーが末端アダプター領域に対して向けられたとき、混在した蛍光が(i)～(iii)に対して起こるが、CDSライブラリー分子の由来となった元のDNA二本鎖のtop鎖対bottom鎖にどの塩基が実際に存在していたのかを判別することは困難になる（図1W）。

10

【0143】

20

ここでの解決策は、CDSフラグメントのみがシーケンシングされるように、リンカー領域にリードプライマー結合部位を配置することである。しかしなお、繋ぎ合わせプロセスの性質により、CDSアダプターのセグメント1/1'および1b/1b'（図1S）は、CDSおよびSI副産物の両方に存在する（図1Uを参照）。よって、SI副産物が読み取られないことをさらに確実にするために、NGSリードプライマー結合部位を、アダプターのセグメント2および3に由来する図1Uに示されている位置に置いた。これはまた、各シーケンシングリードの初期サイクルが茶色および薄緑のセグメントで始まるという意味でもあり、および、これらのサイクルが無駄にならないことを確実にするように、それらは、各DNAフラグメントについてサンプルインデックスおよびユニークな分子識別子をコードするために使用される。これは、次のセクションに記載のとおりインデックスホッピングを軽減すること、および塩基多様性をコードすることでシーケンサー上のクラスター認識および純度フィルターを改善することなどの、その他ユニークな利点を有する。一本鎖セグメントはまた、そうでなければ二本鎖DNAの硬さによって限定される環状化プロセスに柔軟性を導入することで、産物の収率も増大させる。最も重要なことに、これは、各フォワードおよびリバースリード対においてCDS分子のみがシーケンシングされることを確実にすることにより「混在したクラスター」の課題を解決する（図1X）。

30

【0144】

(2) 誤った試料へのCDSリードの誤割り当てを防ぐこと

このCDSのもう1つの重要な特色は、試料の誤割り当てを防ぐためのインデックスホッピングの抑制である。リードのほんの一部が誤った試料に不適切に割り当てられただけでも、多数のエラーを導入する可能性があるため、これは二重鎖シーケンシングの正確性を達成するのに単一CDSリードに頼ろうとするときに特に重要である。従来型のインデックス付けの限界は、インサートを避けてインデックスをタグ付けし、試料調製の最終ステップであるPCRまでタグ付けしないことである。インデックスは、アダプターの相同領域を標的とするプライマーの5'末端に向かってよく配置されるため、残余のプライマーは新しいライブラリー分子に簡単に「交換」され、そしてそれらが割り当てられる試料を変更される可能性がある。（PCRジャンピングにより、部分的に伸長したライブラリー分子でも同じことが起こる可能性がある。）これに対処するために、CDSインデックスを、アダプター複合体それ自体の内に配置し、これはアダプターライゲーションがあるとすぐにインデックスをインサートのすぐ隣に付着させることができるようにする（図1

40

50

Y)。インデックスおよびインサートを読み取ることが今度は単一のリードプライマーを用いてシームレスになるため、シーケンシングの間に分子間のクロストークが起きる可能性が大幅に低くなる。また、CDSではインサート1がインサート2とマッチすることを必要とするため、インサートまたはリンカー領域で起こるあらゆるPCRジャンピングは、異なるインサート1およびインサート2の配列を持つ分子間副産物を作り出すので明らかとなる。なお、この構成では、インデックスがシーケンシングの初期のサイクルで読み取られると仮定して、適切なクラスター生成および純度フィルタリングを確実にするようにCDSインデックス間に十分な多様性を導入した。インデックスリードサイクルはまた、各リードの始めにインデックスを読み取ることによってサイクルを「無駄に」しないように、リード1およびリード2に向けて「別目的で用いられ」もした。インデックスをインラインで読み取るとはまた、インデックスシーケンスがインサートとは分けて別々に読み取られるときに起こることが示されているクラスタークロストークを最小限にするという利点も有する。

10

【0145】

本明細書において別段に定義されない限り、本開示に関連して使用される科学用語および技術用語は、当業者 (those of ordinary skill in the art) (例として、当業者 (the skilled artisan)) によって一般に理解される意味を有するものとする。用語の意味および範囲は明らかであるが、潜在的な曖昧さがある場合には、本明細書で提供される定義が、あらゆる辞書または外部の定義より優先される。さらに、文脈により別段の要求がない限り、単数形の使用には複数形が含まれ、複数形の使用には単数形が含まれるものとする。本開示において、別段の記載がない限り、「または」の使用は「および／または」を意味する。さらに、「含むこと」という用語、ならびに「含む」および「含まれる」などの他の形式の使用は、限定的ではない。また、具体的に別段の記載がない限り、「要素」または「構成要素」などの用語は、1つのユニットを含む要素および構成要素と、2つ以上のサブユニットを含む要素および構成要素の両方を包含する。

20

【0146】

一般に、本明細書に記載の細胞および組織培養、分子生物学、免疫学、微生物学、遺伝学、ならびにタンパク質および核酸の化学およびハイブリダイゼーションに関連して使用される命名法、およびそれらの技法は、当技術分野でよく知られており、一般に使用されているものである。本開示の方法および技法は、一般に、当技術分野で周知の従来の方法に従って、および、別段の指示がない限り、本開示全体にわたって引用および議論される様々な一般的小およびより具体的な参考文献に記載されているように実施される。酵素反応および精製技法は、製造業者の仕様書に従って、当技術分野で一般的に達成されるように、または本明細書に記載のように実施される。本明細書に記載の分析化学、有機合成化学、および医薬品および創薬化学に関連して使用される命名法、ならびにそれらの実験手順および技法は、当技術分野でよく知られ一般的に使用されているものである。標準的な技法は、化学合成、化学分析、医薬品の調製、製剤化、および送達、ならびに対象の処置に使用される。

30

【0147】

本明細書で使用され得る用語「下流」は、複数ヌクレオチド (例として、核酸) の所与の配列におけるランドマークに対する、あるヌクレオチドの位置を指し、下流とは、ランドマークよりも「さらに3'側」を意味するものとする (核酸の場合)。例えばヌクレオチドは、それがランドマークよりも核酸の3'末端に近い (したがって5'末端から遠い) 場合、ランドマークの下流にある。逆に、本明細書で使用され得る用語「上流」は、複数ヌクレオチド (例として、核酸) の所与の配列のランドマークに対する、あるヌクレオチドの位置を指し、上流とは、ランドマークよりも「さらに5'側」を意味するものとする (核酸の場合)。例えばヌクレオチドは、それがランドマークよりも核酸の5'末端に近い (したがって3'末端から遠い) 場合、ランドマークの上流にある。

40

【0148】

用語「およそ」または「約」は本明細書で互換的に使用され、1つ以上の興味ある値に

50

適用される場合、記載された基準値に類似する値を指す。ある態様において、用語「およそ」または「約」は、記載された基準値のいずれかの方向の15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、またはそれ未満内（すなわち、これより大きいパーセンテージまたは小さいパーセンテージ）に入る値の範囲を指すが、ただし、別段の記載がない限り、または文脈から明らかでない限りにおいてである（例えば、かかる数値が可能な値の100%を超える場合）。パーセント同一性

【0149】

用語「パーセント同一性」、「配列同一性」、「%同一性」、「%配列同一性」、および「%同一である」は、本明細書では互換的に使用され得て、2つの配列間（例として、核酸またはアミノ酸）の類似性の定量的測度を指す。ヒトと他の種との間のゲノムDNA配列、イントロンおよびエクソン配列、およびアミノ酸配列のパーセント同一性は、種の種類によって異なり、チンパンジーは全ての種の中で各カテゴリーにおいて、最も高いヒトとのパーセント同一性を有する。

【0150】

2つの核酸配列のパーセント同一性の算出は、例えば、最適な比較目的のために2つの配列を整列させることによって行うことができる（例として、最適アライメントのために第1および第2の核酸配列の一方または両方にギャップを導入することができ、同一でない配列は比較目的では無視することができる）。ある態様において、比較目的で整列された配列の長さは、参照配列の長さの少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも80%、少なくとも90%、少なくとも95%、または100%である。次いで、対応するヌクレオチド位置のヌクレオチドを比較する。第1の配列内の位置が、第2の配列内の対応する位置と同じヌクレオチドによって占められている場合、分子はその位置で同一である。2つの配列間のパーセント同一性は、配列が共有する同一位置の数の関数であり、2つの配列の最適なアラインメントのために導入すべきギャップの数および各ギャップの長さを考慮する。

【0151】

配列の比較および2つの配列間のパーセント同一性の決定は、数学的アルゴリズムを用いて達成することができる。例えば、2つのヌクレオチド配列間のパーセント同一性は、以下に記載されているような方法を用いて決定することができる：Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; および Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; これらの各々は、参照により本明細書に組み込まれる。例えば、2つのヌクレオチド配列間のパーセント同一性は、PAM120重み付き残基表 (weight residue table)、12のギャップ長ペナルティおよび4のギャップペナルティを使用するALIGNプログラム (バージョン2.0) に組み込まれている、Meyers and Miller (CABIOS, 1989, 4:11-17)のアルゴリズムを使用して決定することができる。2つのヌクレオチド配列間のパーセント同一性は、代替的に、NWSgapdna.CMPマトリックスを用いるGCGソフトウェアパッケージのGAPプログラムを使用して決定することもできる。配列間のパーセント同一性を決定するために一般に使用される方法としては、参照により本明細書に組み込まれるCarillo, H., and Lipman, D., SIAM J Applied Math., 48:1073 (1988)に開示されたものが挙げられるが、これらに限定されない。同一性を特定するための技法は、公的に入手可能なコンピュータプログラムに体系化されている。2つの配列間の相同性を決定するための例示的なコンピュータソフトウェアとしては、GCGプログラムパッケージ、Devereux, J., et al., Nucleic Acids Research, 12(1), 387 (1984)、BLASTP、BLASTN、およびFASTA、Atschul, S. F. et al., J. Molec. Biol., 215, 403 (1990) が挙げら

れるが、これらに限定されない。

【0152】

パーセント同一性、またはその範囲（例として、少なくとも、それより多く、等）が記載されている場合、特に指定がない限りエンドポイントは含まれるものとし、その範囲（例として、少なくとも70%の同一性）には引用された範囲内の全ての範囲（例として、少なくとも71%、少なくとも72%、少なくとも73%、少なくとも74%、少なくとも75%、少なくとも76%、少なくとも77%、少なくとも78%、少なくとも79%、少なくとも80%、少なくとも81%、少なくとも82%、少なくとも83%、少なくとも84%、少なくとも85%、少なくとも86%、少なくとも87%、少なくとも88%、少なくとも89%、少なくとも90%、少なくとも91%、少なくとも92%、少なくとも93%、少なくとも94%、少なくとも95%、少なくとも95.5%、少なくとも96%、少なくとも96.5%、少なくとも97%、少なくとも97.5%、少なくとも98%、少なくとも98.5%、少なくとも99%、少なくとも99.5%、少なくとも99.6%、少なくとも99.7%、少なくとも99.8%、少なくとも99.9%の同一性）およびその全ての増分（例として、1パーセントの10分の1（例として、0.1%）、1パーセントの100分の1（例として、0.01%）など）が含まれるものとする。

10

【0153】

本明細書で使用され得る用語「実質的に」は、活性の程度または豊富さを記載するために使用する場合、一般に、過度の努力なしで達成可能な量としての、活性の値を指す。理解できるように、この量は実施される活動に応じて変化し、単純な活動ではより高い閾値が必要となり、より複雑な活動ではより低い閾値が必要となる。例えば、限定されないが、試薬、dNTP、または酵素を混合物から実質的に排除または除去することを指す場合、実質的な量は、50%以上の除去を指し得る。いくつかの態様において、実質的にとは、少なくとも50%（例として、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%、99.5%、99.9%、99.95%、99.99%、またはそれ以上）、および実験誤差内にある変数の全ての値（例として、平均について95%信頼区間）または示された値の $\pm 10\%$ 以内の、いずれか大きい方を指す。いくつかの態様において、実質的にとは、標的の少なくとも75%が除去されることを指す。いくつかの態様において、実質的にとは、標的の少なくとも80%が除去されることを指す。いくつかの態様において、実質的にとは、標的の少なくとも85%が除去されることを指す。いくつかの態様において、実質的にとは、標的の少なくとも90%が除去されることを指す。いくつかの態様において、実質的にとは、標的の少なくとも95%が除去されることを指す。

20

30

【0154】

本明細書で互換的に使用される「野生型」および「ネイティブな」という用語は当業者に理解される専門用語であり、自然界に存在する物品、生物、株、遺伝子、または特徴の典型的な形態であって、操作された、変異体、またはバリエーション形態から区別されるものを意味する。

40

【0155】

二重鎖修復とCODECシーケンシングとの組み合わせ

図1AFに描写されている改変されたNGSワークフローなどの、ある態様において、本開示は、(a)二重鎖修復と(b)CODECシーケンシングとを組み合わせるシーケンシング方法を提供する。

【0156】

核酸の調製に使用される既存の方法は、多くの作業およびステップを実施する。「末端修復」(ER)および「dAテーリング」(AT)(ER/AT)として知られる既存の方法はそれぞれ、dTMPテール付き(dTMP-tailed)シーケンシングアダプターのライゲーションの準備として、DNAフラグメントを平滑化およびリン酸化し、デオキシアデノシン

50

ーリン酸（「dAMP」）の3'末端への非鋳型付加を実施するために使用される（図1）。ERおよびATは、逐次的に、または「ワンポット」反応内（例えば、プロセスおよび方法の全体が、ステップを分離することなく1つの反応容器内で同時に行われる）のいずれかで実施され、3'オーバーハングを消化して5'オーバーハングをフィルインし、かつ単一のdAMPを二重鎖の鎖の各3'末端に残すことが意図されている、DNAポリメラーゼ（単数または複数）を使用する。しかし、ER/AT（そのまま、またはNEB PreCR（登録商標）もしくはExoVIIなどの前処理と組み合わせて；例として、図34および図35A～35Cを参照）は、伝統的に、5'エキソヌクレアーゼおよび／または鎖置換活性を有する、1つ以上のDNAポリメラーゼの使用を伴う。したがって、広範な鎖再合成が、二重鎖内の内部のニックおよびギャップ、および長い5'オーバーハングから生じる可能性があるという仮説が立てられた。再合成が、もともと一方の鎖に限定されていた増幅可能な病変または変化の存在下で生じると、エラーが両方の鎖にコピーされ、両方の鎖上の真の突然変異と区別できなくなり得るか、またはその可能性がある。二重鎖シーケンシングにおけるこの偽の発見の源は、短い5'オーバーハングがしばしばフィルインされるフラグメント末端で最も明確に見られるが（図2C）、本明細書では、かかるエラーは、以下が与えられている場合にフラグメントのさらに深い部分にまで及び得ることが示されている：（i）ER/ATで一般的に使用されるTaqおよびKlenowなどのポリメラーゼの、5'エキソヌクレアーゼおよび鎖置換活性、および（ii）複数の内因性または外因性因子によって誘導される様々な程度の主鎖損傷、これは、鎖再合成のプライミング部位（例として、ニック、ギャップ）として機能する。これは、271個のセルフリーDNA（cfDNA）試料よりも約100倍高いエラー率を示した重度の損傷を受けたFFPE腫瘍DNA試料において、3'フラグメント末端からの距離に応じて減少する長いテールのエラーが観察された理由を、説明できる可能性がある（図2C）。このメカニズムは、従来型のER/ATキットを用いた、ニック、ギャップ、およびオーバーハングを有する合成オリゴヌクレオチドの処理を含む実験によっても確認された（図2Bおよび図3A）。フラグメント末端でのエラーは、フラグメント末端のin silicoトリミングによって軽減できるが、各フラグメントの内部（またはフラグメント末端から事前に指定された距離を超える、例えば12bpを超える部分）で発生するエラーは、この方法では、DNAシーケンシングデータの収量を大幅に妥協することなく解決することはできない。これは、二重鎖シーケンシングが理論的には、一方の鎖の塩基損傷エラーを識別できるが、実際にはその能力が発効物質の品質に依存しており、これには多くの理由から大きな問題があることを意味する。例えば、ER/ATの前に、試料をフラグメント化してライブラリーを調製する。このフラグメント化は、核酸を小さなフラグメントに分解する。これは、物理的（例えば、超音波処理または物理的力によって）、酵素的、または化学的に達成することができる。しかし、あらゆる形態のフラグメント化は、本質的に鎖に損傷を与えて破断し、オフターゲット損傷（例として、オーバーハング、ニック、ギャップ、損傷塩基）を誘導する可能性がある。

【0157】

本明細書で開示されるのは、Duplex-Repair（DR）と呼ばれる新しいER/AT方法であり、これは、既存の方法に固有の問題の多くを最小化および／または除去する。例えば限定することなく、DRは、NGSアダプターのライゲーション前に、鎖再合成を最小限に抑え、これは偽の突然変異の発見を大幅に制限する。本明細書に示すように、この再合成を最小限に抑えることで、DRは、各二重鎖の両方の鎖からの配列のコンセンサスに依存する、二重鎖シーケンシングおよび他の関連方法の主要なアキレス腱に対処して、最大の精度および堅牢性を提供する。

【0158】

図1AFに示されている態様において、典型的なNGSワークフローは、左下の概略図に示されており、および、（i）シーケンシングされるDNA試料の末端修復、（ii）NGSアダプターライゲーション、（iii）PCR増幅（例として、フローセルクラスター増幅）、（iv）富化、（v）PCR、および（vi）NGSによるシーケンシングを含む。この図は、本開示の文脈においてNGSワークフローをある具体的なワークフローに限定

することは意図しない。あらゆるNGSワークフロー（例として、Illumina NGSワークフロー）が利用されてもよい。示された態様において、末端修復ステップは、二重鎖修復により置き換えられ、およびアダプターライゲーションステップはCODECにより置き換えられる。本開示は、改変された二重鎖修復が、CODECを実施することに先立って使用されてもよく、もともと天然では1本の鎖のみにあったヌクレオチド損傷または変化の増幅に起因する偽の突然変異などの、偽の突然変異の伝播を最小限にするために、核酸試料は二重鎖修復（DR）の方法により処理されてもよいということを提供する。

【0159】

したがって、いくつかの側面において、本開示は、もともと天然では一本鎖のみにあったヌクレオチド損傷または変化の増幅による偽の突然変異の伝播を最小限に抑える、シーケンシング用の核酸試料（試料；およびかかる用語は本明細書でさらに詳しく説明する）を調製する方法に関し、ここで試料の少なくとも一部は二本鎖であり、試料を反応容器に添加すること、ならびに以下を含む：（a）試料を、以下が可能な1つ以上の酵素と接触させること：（i）1つ以上の損傷塩基を、試料から切除すること；（ii）1つ以上の脱塩基部位を切断すること、および得られた末端を、DNAポリメラーゼによる伸長および／またはDNAリガーゼによるライゲーションに適合するように処理すること；（iii）および5'オーバーハングを消化すること；（b）試料を、以下の1つ以上と接触させること：（i）鎖置換および5'エキソヌクレアーゼ活性の両方を欠くが、試料の一本鎖セグメントをフィルインし、かつ／または試料の3'オーバーハングを消化することができる、DNA依存性DNAポリメラーゼ；および（ii）試料の鎖の5'末端をリン酸化することができる酵素；および（c）試料を、ニックを封止可能なDNAリガーゼと接触させること。いくつかの態様において、本開示の方法はさらに以下を含む：（d）アダプターライゲーション用の試料を調製すること、ここで調製は：（i）デオキシアデノシンリン酸（dAMP）を試料の鎖の3'末端に付加すること（dAテーリング）；または（ii）任意に試料の末端をさらに平滑化すること、を含む。

【0160】

いくつかの側面において、方法は、試料の少なくとも一部が二本鎖である核酸試料（試料）を調製することを含み、これは以下を含む：（a）試料を、以下が可能な1つ以上の酵素と接触させること：（i）試料の鎖の5'末端をリン酸化すること；3'ヒドロキシル部分を、試料の鎖の3'末端に付加すること；および（ii）ニックを封止すること；（b）試料を、5'および3'オーバーハングを除去すると共にギャップ領域を消化して平滑化二重鎖を生成することができる1つ以上の酵素と接触させること；および（c）デオキシアデノシンリン酸（dAMP）を、試料の鎖の3'末端に付加すること（dAテーリング）。かかる方法において、損傷塩基を切除する必要性、ExoVIIで処理する必要性、またはExoVII処理後に残されたギャップおよび短い5'オーバーハングを埋める必要性は、酵素（例として、エンドヌクレアーゼ（例として、ヌクレアーゼS1））を使用して、一本鎖ギャップ領域を切断し、オーバーハング領域に存在するヌクレオチドを切断することにより、軽減され得る。いくつかの態様において、ステップ（a）（1）で使用される酵素は、T4ポリヌクレオチドキナーゼ、HiFi Taqリガーゼ、またはそれらの組み合わせを含む。いくつかの態様において、ステップ（b）で使用される酵素は、ヌクレアーゼS1である。

【0161】

本明細書で使用され得る用語「エンドヌクレアーゼ」および「ヌクレアーゼ」は、一般にポリヌクレオチド鎖（例として、オリゴヌクレオチド、核酸）内のホスホジエステル結合（単数または複数）を切断する酵素を指すことが当業者に知られている技術用語である。ヌクレアーゼは天然に存在する場合もあれば、遺伝子操作された場合もある。いくつかの態様において、エンドヌクレアーゼは、エンドヌクレアーゼIV（EndoIV）である。いくつかの態様において、エンドヌクレアーゼは、エンドヌクレアーゼVIII（EndoVIII）である。いくつかの態様において、ヌクレアーゼはヌクレアーゼS1を含む（例えば、限定はされないが以下を参照されたい：thermofisher.com/order/catalog/product

t/EN0321#/EN0321;promega.com/products/cloning-and-dna-markers/molecular-biology-enzymes-and-reagents/s1-nuclease/?catNum=M5761;takarabio.com/products/cloning/modifying-enzymes/nucleases/s1-nuclease;およびsigmaaldrich.com/US/en/product/SIGMA/N5661)。ヌクレアーゼS1は一本鎖核酸を分解し、5'-ホスホリルモノヌクレオチドまたはオリゴヌクレオチドを放出し、および、二本鎖DNA(dsDNA)を、ニック、ギャップ、ミスマッチ、またはループによって生じた一本鎖領域で切断することもある。

【0162】

本明細書に記載の方法を実施することにより、偽の突然変異が導入される可能性は実質的に軽減される。例えば、最初に損傷塩基を切除し、脱塩基部位を切断し、かつ得られた末端を、DNAポリメラーゼによる伸長と試料からのDNAリガーゼによるライゲーションに適合するように処理する酵素を使用することにより、一方の鎖で塩基が切除されてギャップが生成されるか(ここで相補鎖は切除点にまだ存在し、二重鎖が無傷のままであるための主鎖を形成する)、または、二重鎖/鎖の破断が発生し、2つの「娘」二重鎖が作成される(ここで、相補鎖は切除点には存在せず、二重鎖は2つの小さな核酸に分解される)。このステップの利点は、限定されないが、損傷塩基が存在するギャップ領域において鎖破断を誘導することであり、なぜならば、本明細書に開示される方法のステップ(b)はDNAポリメラーゼを使用してギャップをフィルインすることを含み得るが、一方、アダプターライゲーション前に再合成されなかった完全二重鎖領域の1つの鎖上のあらゆる損傷塩基またはミスマッチ塩基は、未修正のまま残っていれば、コンピューターで二重鎖シーケンシングを用いて解決される可能性があるからである。さらに、これらの得られた二重鎖(無傷であるか、または分解されている(例として、鎖破断が起こっている))は、その後、5'オーバーハングを消化できる酵素に曝露される(例として、接触される)と、あらゆる5'オーバーハングの長さが実質的に低減され、その後のステップ(b)での、フラグメントの最末端までのフィルインを制限するであろう。次に得られた二重鎖を、鎖置換および5'エキソヌクレアーゼ活性の両方を欠くが、試料の一本鎖セグメントをフィルインすることができかつ3'オーバーハングの消化が可能なDNA依存性DNAポリメラーゼ、およびポリヌクレオチドキナーゼに曝露(例として、接触)させると、前のステップで完全に消化されなかったあらゆる残りの短い5'オーバーハングはフィルインされて、平滑末端が得られ;あらゆる残りの3'オーバーハングは消化されて平滑末端が生成され;および、あらゆる内部ギャップ(例として、損傷塩基の切除および脱塩基部位の切断によって生じる小さなギャップ、およびDNAフラグメントにも存在し得るあらゆるより長いギャップ)は、下流のDNAセグメントの5'末端まで埋められる。次に、得られた二重鎖を、ニックを(好ましくは、キメラ形成を避けるために最小限の末端結合活性で)封止可能なDNAリガーゼに曝露(例として、接触)させると、あらゆる残りのニック(例として、試料中に本質的に存在する他のもののうち、ギャップを埋めた後に残ったニック)は封止され、連続した平滑二重鎖を形成する。次に、得られた二重鎖を、それぞれ5'エキソヌクレアーゼ活性および鎖置換活性を有するTaqまたはKlenowフラグメントなどのDNAポリメラーゼを使用して、DNA二重鎖の3'末端への、dAMPの非鋳型伸長(例として、付加)(例として、dAテーリング)を実行できるDNAポリメラーゼに曝露する(例として、接触させる)と、鎖再合成に利用できる「プライミング部位」が実質的に少なくなる。さらに、ステップ(d)が、dAMP以外のヌクレオチドの付加を制限する条件下で実施される場合(例として、このステップの前にdNTPを実質的に除去することによって、または極端に過剰なdATPを提供することによって)、このステップでの鎖再合成の可能性は、大幅に軽減することができる。この保存された情報は、突然変異の精度および解像度の大幅な向上を可能にする。

【0163】

本明細書で使用され得る用語「接触された」は、1つの物質(例として、酵素、試薬、dNTP)の別の物質(例として、試料、混合物)への暴露であって、ある量および意図で、すなわち、2つの物質が相互作用して、一方の物質の活性が他方の物質(例として、試

10

20

30

40

50

料に作用する酵素)に影響を与えるようにすること、または2つの物質が相互作用することを意図しての、前記曝露を記述するために使用される。この用語は、2つの物質間の物理的接触を必要とするものと解釈されるべきではないが、さらに物理的接触を禁止するものでもない。例えば、物質間の相互作用および／または活性に影響を与えるのに、近接性は十分であり得る。いくつかの態様において、接触は、物質を同じ容器(例として、反応容器)に導入することによって達成される。いくつかの態様において、接触は、物質を同じ反応容器に導入することによって達成される。いくつかの態様において、接触は、物質A(例として、試薬、dNTP、酵素など)を、物質B(例として、試料)を含有するか、物質Bが同時に導入されるか、または物質Bが後で導入される反応容器に導入することによって達成される。いくつかの態様において、接触は、物質が互いに物理的に接触する(例として、物理的に相互作用する)ときに達成される。いくつかの態様において、接触は、物質が互いに化学的に相互作用するときに達成される。いくつかの態様において、接触は、物質が互いに酵素的に相互作用するときに達成される。いくつかの態様において、接触は、物質が互いに近接しているときに達成される。

10

20

30

40

50

【0164】

いくつかの態様において、本開示の方法はさらに(d)アダプターライゲーション用の試料を調製することを含み、ここで調製することは:(i)デオキシアデノシンリン酸(dAMP)を試料の鎖の3'末端に付加すること(dAテーリング);または(ii)試料の末端を平滑化すること、を含む。いくつかの態様において、dAテーリングは、試料を、デオキシアデノシンリン酸(dAMP)を試料の鎖の3'末端に組み込むことができる酵素と接触させること、および、試料を、dNTPと接触させることを含む。いくつかの態様において、本開示の方法のステップ(a)~(c)で使用される酵素および／またはdNTPは、dAテーリングの前に、反応容器から実質的に除去される。いくつかの態様において、dNTPは実質的にdATPを含む。いくつかの態様において、本明細書に開示される方法の1つ以上(例として、ステップ(a)、(b)、(c)、(d)等の代表として、1、2、3、4、5、またはそれ以上)は、「ワンポット」反応で行われ、ここでこれらのステップは、酵素および緩衝液を同じ反応容器に順次添加し、反応条件(例として、温度)を調整することによって、実施される。いくつかの態様において、ステップは連続して実施される。いくつかの態様において、前のステップからの試薬および酵素は、次のステップに進む前に混合物から除去されない。いくつかの態様において、前のステップからの試薬および酵素は、次のステップに進む前に混合物から除去される。いくつかの態様において、1つ以上のステップが1つの反応容器内で実施される。いくつかの態様において、1つ以上のステップが、2つ以上の反応容器内で実施される(例として、方法全体を通じて少なくとも1つの時点で移される)。

【0165】

二重鎖プレ増幅とCODECとの組み合わせ

様々な態様において、二重鎖プレ増幅が、核酸試料(例として、DNA試料)に対して、CODECアダプターライゲーションおよびCODECシーケンシングに先立ち実施されてもよい。CODECシーケンシングへのインプットとしての本明細書に記載の核酸試料は、低い存在量の核酸を含有してもよい。そのため、低い存在量の核酸はCODECアダプターライゲーションおよびCODECシーケンシングに先立ち増幅される必要があり得る。加えて、CODECアダプターライゲーションおよびCODECシーケンシングに先立ち核酸を増幅させることにより、CODECアダプターライゲーションおよびCODECシーケンシングの間の核酸材料の損失が許容可能になり、よって高い変換率および高い効率がもたらされる(図20)。

【0166】

いくつかの態様において、核酸試料内の核酸は、UMI、サンプルインデックス、ローリングサークル増幅プライマー、およびランケーション部位を各々含む2つのプレ増幅分子と接触させられる。本明細書において使用される用語「ローリングサークル増幅」は、核酸の複数コピーを迅速に合成することができる一方向的核酸複製のプロセスを指す。

本明細書において使用される用語「トランケーション部位」は、切断されやすい核酸部位を指す。いくつかの態様において、プレ増幅分子は、核酸の各末端ヘライゲーションされて核酸のローリングサークル増幅を可能にさせ、そのようにして核酸の複数コピーを合成させる。いくつかの態様において、ローリングサークル増幅および核酸の複数コピーの合成の後、ローリングサークル増幅プライマーを含むローリングサークル増幅アダプターがトランケーション部位で切断され、同じ核酸分子の複数コピーを結果としてもたらす。いくつかの態様において、ローリングサークル増幅の後、結果としてもたらされる複数の核酸分子は、サンプルインデックスおよびUMIを各々含む。いくつかの態様において、結果としてもたらされる複数の核酸分子は、CODECアダプターヘライゲーションされ、およびCODECライブラリー調製プロトコルおよび後続するシーケンシングを通じて続いていく。

10

【0167】

CODECシーケンシングの改変された方法

いくつかの態様において、CODECシーケンシングは、改変されたCODECシーケンシングアダプターを用いて実施されてもよい（図21）。いくつかの態様において、本明細書に記載のとおり、標準的なCODECシーケンシングアダプターは、CODECシーケンシングアダプターの途中のリンカー配列に隣接するリードプライマーを含む。いくつかの態様において、改変されたCODECシーケンシングアダプターは、CODECシーケンシングアダプターの末端にリードプライマーを含み、改変されたCODECシーケンシングアダプターの途中にリンカー配列を含まない。いくつかの態様において、改変されたCODECシーケンシングアダプターは、標準的なCODECシーケンシングアダプターを作るために使用される方法に類似のかつ本明細書に記載のとおり方法に従って作られる。いくつかの態様において、改変されたCODECシーケンシングアダプターがインプットのdsDNA二重鎖ヘライゲーションされる前に、改変されたCODECシーケンシングアダプターの3'末端がライゲーションからブロックされる。いくつかの態様において、改変されたCODECシーケンシングアダプターの3'末端のブロッキングの後で、改変されたCODECシーケンシングアダプターがインプットのdsDNA二重鎖ヘライゲーションされて部分的に環状のDNA分子を形成する。いくつかの態様において、部分的に環状のDNA分子は、鎖置換伸長を受け、そのようにしてdsDNA二重鎖を含む直鎖状の改変されたCODECシーケンシング分子が作られる。代替手段において、標準的なCODECシーケンシングアダプターを作りおよび標準的なCODECシーケンシングアダプターをインプットdsDNA二重鎖ヘライゲーションするために使用される方法に従って、改変されたCODECシーケンシングアダプターが作られおよびインプットdsDNA二重鎖ヘライゲーションされるが、しかし鎖置換伸長の後で、直鎖状の標準的なシーケンシングアダプターの末端がトランケーションされ、そのようにしてdsDNA二重鎖を含む改変されたCODECシーケンシング分子が作られる。いくつかの態様において、dsDNA二重鎖を含む改変されたCODECシーケンシング分子は、一本鎖DNA環状化を受ける。いくつかの態様において、CODECシーケンシング分子の途中のリンカーには、ニックが入れられており、そのようにしてdsDNA二重鎖の両方の鎖およびCODECシーケンシング分子の両末端のリードプライマーを含む直鎖のCODECシーケンシング分子が作られる。いくつかの態様において、CODECシーケンシング分子は、長さ1ヌクレオチド以下であるCODECシーケンシング分子の途中のリンカーを含む。いくつかの態様において、改変されたCODECシーケンシング分子は、標準的なCODECシーケンシング分子について使用されるものと同じシーケンシングプロトコルに従ってシーケンシングされることができる。

20

30

40

【0168】

核酸試料

様々な側面において、DNAをシーケンシングするためのCODECシーケンシング方法には、シーケンスのための核酸分子の試料を得ることが関与する。核酸は一般に、試料または対象から取得される。本発明の方法に従う標識化および／または検出のための標的分子は、限定されないが、DNA、ゲノムDNA、RNA、発現RNAおよび／または染色体（単数

50

または複数)などの遺伝的およびプロテオミクス材料を包含する。本発明の方法は、全細胞からのDNAに、または1つ以上の細胞から得られる遺伝的またはプロテオミクス材料の部分に、適用可能である。本発明の方法は、ウイルスなどの非細胞源からDNAまたはRNAが得られることを許容する。対象については、試料はあらゆる臨床的に許容し得る様式で得られてもよく、および、核酸鑄型は、当技術分野において知られている方法により試料から抽出される。一般に、核酸は、その内容がそれらの全体において参照により本明細書に組み込まれるManiatis, et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp. 280-281, 1982)に記載されているものなどの様々な技法により、生体試料から抽出されることができる。

核酸鑄型は、デオキシリボ核酸(DNA)および／またはリボ核酸(RNA)を包含する。核酸鑄型は、合成の、または天然源に由来するものとして行うことができる。核酸は、生物学的、環境的、物理的、または合成のいずれかを問わず、あらゆる入手源または試料から得られてもよい。一態様において、核酸鑄型は、タンパク質、脂質および非鑄型核酸などの、様々な他の構成要素を含有する試料から単離される。核酸鑄型は、動物、植物、細菌、真菌、またはあらゆるその他の細胞生物から得られる、あらゆる細胞性材料から得ることができる。本発明における使用のための試料は、ウイルス、ウイルス粒子または調製物を包含する。核酸はまた、環境試料などの試料からの、細菌または真菌などの微生物から取得されてもよい。

【0169】

本発明において、標的材料は、DNA、RNA、cDNA、PNA、LNAおよびその他の、試料中に含有されるものを包含する、あらゆる核酸である。核酸分子は、デオキシリボ核酸(DNA)および／またはリボ核酸(RNA)を包含する。核酸は、合成の、または天然源に由来するものとして行うことができる。一態様において、核酸分子は、タンパク質、脂質および非鑄型核酸などの、様々な他の構成要素を含有する生体試料から単離される。核酸鑄型分子は、動物、植物、細菌、真菌、またはあらゆるその他の細胞生物から得られる、あらゆる細胞性材料から得ることができる。ある態様において、核酸分子は、単一の細胞から得られる。本発明における使用のための生体試料は、ウイルス粒子または調製物を包含する。核酸分子は、生物から直接、または生物から得られた生体試料、例として、血液、尿、脳脊髄液、精液、唾液、痰、糞便および組織から、得ることができる。あらゆる組織または体液標本が、本発明における使用のための核酸のための入手源として使用されてもよい。核酸分子はまた、初代培養細胞または細胞株などの培養された細胞から単離されたものとして行うこともできる。鑄型核酸がそれから得られる細胞または組織は、ウイルスまたは他の細胞内病原体に感染したものとして行うことができる。加えて、核酸は、ウイルス試料または環境試料などの、非細胞もしくは非組織試料から得られるものとして行うこともできる。

【0170】

試料はまた、生体標本から抽出されたトータルRNA、cDNAライブラリー、ウイルスの、またはゲノムDNAとすることもできる。ある態様において、核酸分子は、タンパク質、酵素、基質、抗体、結合剤、ピーズ、小分子、ペプチド、またはあらゆるその他の分子などの、他の標的分子と結合し、および、標的分子を定量化および／または検出するための代用物として機能する。一般に、核酸は、Sambrook and Russell, Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor, N.Y. (2001)に記載されているものなどの様々な技法により、生体試料から抽出されることができる。核酸分子は、一本鎖、二本鎖、または一本鎖領域を伴う二本鎖であってもよい(例えば、ステムおよびループ構造)。抗体またはアプタマーなどの、高親和性結合部分へ結合することができるタンパク質またはタンパク質の部分(アミノ酸重合体)は、例えば液滴における、オリゴヌクレオチド標識化のための標的分子である。

【0171】

核酸鑄型は、生物から直接、または生物から得られた生体試料、例として、血液、尿、脳脊髄液、精液、唾液、痰、糞便および組織から、得ることができる。具体的な態様にお

いて、核酸は、新鮮凍結血漿（FFP）から得られる。具体的な態様において、核酸は、ホルマリン固定パラフィン包埋（FFPE）組織から得られる。あらゆる組織または体液標本が、本発明における使用のための核酸のための入手源として使用されてもよい。核酸鑄型はまた、初代培養細胞または細胞株などの培養された細胞から単離されたものとする 것도できる。鑄型核酸がそれから得られる細胞または組織は、ウイルスまたは他の細胞内病原体に感染したものとすることができる。試料はまた、生体標本から抽出されたトータルRNA、cDNAライブラリー、ウイルスの、またはゲノムDNAとすることもできる。

【0172】

生体試料は、界面活性剤（detergent）または表面活性剤（surfactant）の存在下で均質化または分画されてもよい。緩衝液中の界面活性剤の濃度は、約0.05%～約10.0%であってもよい。界面活性剤の濃度は、界面活性剤が溶液中で可溶性を維持できる量までとすることができる。好ましい態様において、界面活性剤の濃度は、0.1%～約2%の間である。界面活性剤、特に非変性の、穏やかなものは、試料を可溶化させるように作用することができる。界面活性剤は、イオン性であるかまたは非イオン性であってもよい。非イオン性の界面活性剤の例は、Triton Xシリーズ（Triton X-100 t-Oct-C₆H₄-(OC₂H₅-CH₂)_xOH、x=9-10、Triton X-100R、Triton X-114 x=7-8）などのトリトン、オクチルグルコシド、ポリオキシエチレン（9）ドデシルエーテル、ジギトニン、IGE PAL CA630オクチルフェニルポリエチレングリコール、n-オクチル-ベータ-D-グルコピラノシド（ベータOG）、n-ドデシル-ベータ、Tween 20ポリエチレングリコールソルビタンモノラウレート、Tween 80ポリエチレングリコールソルビタンモノオレアート、ポリドカノール、n-ドデシルベータ-D-マルトシド（DDM）、NP-40ノニルフェニルポリエチレングリコール、C12E8（オクタエチレングリコールn-ドデシルモノエーテル）、ヘキサエチレングリコールモノ-n-テトラデシルエーテル（C14E06）、オクチル-ベータ-チオグルコピラノシド（オクチルチオグルコシド、OTG）、Emulgen、およびポリオキシエチレン10ラウリルエーテル（C12E10）を包含する。イオン性の界面活性剤（アニオン性またはカチオン性）の例は、デオキシコラート、ドデシル硫酸ナトリウム（SDS）、N-ラウロイルサルコシン、およびセチルトリメチルアンモニウム臭化物（CTAB）を包含する。Chaps、両性イオン3-14、および3-[(3-コラミドプロピル)ジメチルアンモニオ]-1-ホスホパンスルホナートなどの、両性イオン性試薬もまた、本発明の精製スキームにおいて使用されてもよい。

【0173】

溶解または均質化溶液は、還元剤などの、他の剤をさらに含有してもよい。かかる還元剤の例は、ジチオスレイトール（DTT）、ベーターメルカプトエタノール、DTE、GSH、システイン、システアミン、トリカルボキシエチルホスフィン（TCEP）、または、亜硫酸の塩を包含する。ひとたび得られると、核酸は、一本鎖核酸鑄型を作るために当技術分野において知られているいずれかの方法により変性させられ、および、第1および第2のオリゴヌクレオチドの対は、一本鎖核酸鑄型に、第1および第2のオリゴヌクレオチドが鑄型上の標的領域に隣接するようにハイブリダイズされる。

【0174】

いくつかの態様において、核酸は、フラグメント化されまたはより小さな核酸フラグメントへと分解されてもよい。ゲノム核酸を包含する、核酸は、機械的なフラグメント化、化学的なフラグメント化、および酵素的なフラグメント化などの、様々な方法のいずれかを使用してフラグメント化されることができる。核酸のフラグメント化の方法は、当技術分野において知られており、そして、限定されないが、DNase消化、超音波処理、機械的せん断、および同類のものを包含する（J. Sambrook et al., "Molecular Cloning: A Laboratory Manual", 1989, 2.sup.nd Ed., Cold Spring Harbour Laboratory Press: New York, N.Y.; P. Tijssen, "Hybridization with Nucleic Acid Probes- Laboratory Techniques in Biochemistry and Molecular Biology (Parts I and II)", 1993, Elsevier; C. P. Ordahl et al., Nucleic Acids Res., 1976, 3: 2985-2999; P. J. Oefner et al., Nucleic Acids Res., 1996, 2

4: 3879-3889; Y. R. Thorstenson et al., *Genome Res.*, 1998, 8: 848-855) 。米国特許公開公報第2005/0112590号は、当技術分野で知られているフラグメント化の様々な方法の一般的な概観を提供する。

【0175】

ゲノム核酸は、均一なフラグメントへとフラグメント化されるかまたはランダムにフラグメント化されることができる。ある側面において、核酸は、約5キロベースまたは100キロベースのフラグメント長さを有するフラグメントを形成するようにフラグメント化される。好ましい態様において、ゲノム核酸フラグメントは、1キロベース～20キロベースの範囲とすることができる。好ましいフラグメントは、サイズが様々であり、および約10キロベースの平均フラグメント長さを有するものとしてすることができる。しかしながら、所望のフラグメント長さおよびフラグメント長さの範囲は、捕捉しようとする核酸標的のタイプに応じて調整することができる。フラグメントの具体的な方法は、所望のフラグメント長さを達成するように選択される。いくつかの非限定例が、以下に提供される。

【0176】

ゲノム核酸の化学的フラグメント化は、多種多様な方法を使用して達成することができる。例えば、塩基および酸加水分解を包含する加水分解反応は、核酸をフラグメント化するために使用されるよくある技法である。加水分解は、所望の加水分解の程度に応じて、温度上昇によって容易になる。フラグメント化は、下記のとおり温度およびpHを変更することにより成し遂げることができる。せん断のためのpHベースの加水分解の利益は、それが一本鎖の産物を結果としてもたらすことができるということである。加えて、温度は、加水分解を成し遂げるために一時的にpHを中性から上または下へシフトさせ、次いで長期保管等のために中性に戻すために、ある緩衝系（例として、Tris）とともに使用することができる。pHおよび温度の両方は、種々のせん断量（および、ゆえに様々な長さ分布）に影響を与えるように調節することができる。

【0177】

化学的切断はまた、特異的であることもできる。例えば、選択された核酸分子はアルキル化を介して切断することができ、特にホスホロチオアート修飾核酸分子がそうである（例として、K. A. Browne, "Metal ion-catalyzed nucleic Acid alkylation and fragmentation," *J. Am. Chem. Soc.* 124(27): 7950-7962 (2002)を参照）。ホスホロチオアート修飾でのアルキル化は、核酸分子を修飾部位で切断されやすくする。I. G. GutおよびS. Beck, "A procedure for selective DNA alkylation and detection by mass spectrometry," *Nucl. Acids Res.* 23(8): 1367-1373 (1995)を参照。

【0178】

本発明の方法はまた、Maxam-Gilbert Sequencing Method (Chemical or Cleavage Method), *Proc. Natl. Acad. Sci. USA.* 74:560-564に開示された技法を使用して化学的に核酸をフラグメント化することも企図する。そのプロトコルにおいては、グアニンでの、アデニンでの、シトシンおよびチミンでの、およびシトシンのみでの優先的切断などの特定の塩基で核酸をフラグメント化することのために設計された化学物質への曝露により、ゲノム核酸を化学的に切断することができる。

【0179】

核酸のフラグメントへの機械的せん断は、当技術分野において知られているあらゆる方法を使用して起こることができる。例えば、核酸をフラグメント化することは、流体せん断 (hydroshearing)、針を通した粉碎、および超音波処理によって成し遂げることができる。例えば、Quail, et al. (Nov 2010) *DNA: Mechanical Breakage*. In: eLS. John Wiley & Sons, Chichesterを参照。

【0180】

核酸はまた、霧状化を介してせん断することもできる (Roe, BA, Crabtree. JS and Khan, AS 1996) ; Sambrook & Russell, *Cold Spring Harb Protoc* 2006を参照。霧状化には、ネブライザーにおける小さな孔に核酸溶液を強制的に通すことによ

て作り出された霧からフラグメント化されたDNAを収集することが関与する。霧状化により得られるフラグメントのサイズは、主としてDNA溶液が孔を通過する速度によって決定され、ネブライザーを吹き抜ける気体の圧力、溶液の粘度、および温度が変更される。結果としてもたらされるDNAフラグメントは、狭い範囲のサイズにわたって分布する(700~1330bp)。核酸のせん断は、得られた核酸を狭いキャピラリーまたはオリフィスを通じて通過させることによって成し遂げることができる(Oefner et al., *Nucleic Acids Res.* 1996; Thorstenson et al., *Genome Res.* 1995)。この技法は、核酸試料がシリンジポンプによって小さな孔に強制的に通されたときに結果としてもたらされるポイントシンク流体力学に基づく。

【0181】

HydroShearing (Genomic Solutions, Ann Arbor, Mich., USA) では、急な狭窄部を伴うチューブに、溶液中のDNAが通過させられる。狭窄部に近づくにつれて、流体は、狭窄部のより小さな面積を通じて体積流量を維持するように加速する。この加速の間、抗力はDNAをそれが切れるまで引き伸ばす。DNAは、せん断力が化学結合を壊すには欠片が短すぎる状態となるまでフラグメント化する。流体の流速および狭窄部のサイズが、最終的なDNAフラグメントサイズを決定づける。

【0182】

超音波処理もまた、核酸を短時間の超音波処理、すなわち超音波エネルギーに供することにより核酸をフラグメント化するために使用される。超音波処理により核酸をフラグメントへとせん断する方法は、米国特許公開公報第2009/0233814号に記載されている。方法においては、精製された核酸が、内部に粒子が配置された懸濁液中に入れられて得られる。試料および粒子の懸濁液は、次いで核酸フラグメントへと超音波処理される。

【0183】

酵素的フラグメント化は、酵素的切断としても知られており、エンドヌクレアーゼ、エキソヌクレアーゼ、リボザイム、およびDNAザイムなどの酵素を使用して核酸をフラグメントへと切り分ける。かかる酵素は、広く知られており、および市販で利用可能である、Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 3rd (2001)およびRoberts RJ (January 1980). "Restriction and modification enzymes and their recognition sequences," *Nucleic Acids Res.* 8 (1): r63-r80参照。様々な酵素的フラグメント化技法は当技術分野において周知であり、およびかかる技法はシーケンシングのための核酸をフラグメント化するために頻繁に使用され、例えば、Alazard et al, 2002; Bentzley et al, 1998; Bentzley et al, 1996; Faulstich et al, 1997; Glover et al, 1995; Kirpekar et al, 1994; Owens et al, 1998; Pieles et al, 1993; Schuette et al, 1995; Smirnov et al, 1996; Wu & Aboleneen, 2001; Wu et al, 1998aである。

【0184】

核酸をフラグメント化するために使用される最もよくある酵素はエンドヌクレアーゼである。エンドヌクレアーゼは、二本鎖または一本鎖のいずれかの核酸に特異的であることができる。核酸分子の切断は、核酸分子内でランダムに生じることができるか、または核酸分子の特定の配列で切断することもできる。核酸分子の特異的なフラグメント化は、逐次反応または同時反応にて1つ以上の酵素を使用して成し遂げることができる。

【0185】

上記の側面および態様のいずれかは、概要、図面、および／または詳細な説明のセクション(下記の例／態様を包含する)において開示されたとおりの他のいずれかの側面または態様と組み合わせることができる。

【0186】

例

例1-CODECシーケンシング

ごく低いレベルの突然変異を単一の二本鎖DNA分子(「単一の二重鎖」)内で発見することは、診断[1]、予測[2]、および予後[3]バイオマーカーを見出すこと、がんの進化

10

20

30

40

50

[4]および体細胞モザイク[5]を理解すること、ならびに感染症[6]および老化[7]の研究に不可欠である。第3世代シーケンシング技術（例として、PacBio、Oxford Nanopore Technologies）は、原理上は、各単一DNA二重鎖を全体としてシーケンシングすることでいずれかの鎖上の偽の突然変異とは別に真の突然変異を明らかにできるが、しかし、実用上は、必要とされる正確性およびスループットに欠ける[8、9]。次世代シーケンシング（NGS）は、他方、秀でたリードの正確性およびスループットを引き続き提供するが[10]、しかし、少なくともそのスループットまたは実用性を著しく損なうことなしに、単一の二重鎖をシーケンシングするようには構成されていない。

【0187】

NGSは、大規模並列蛍光分析において短いクローン的に増幅されたDNAフラグメントを読み取ることにより、高いスループットを与える。それでも、その正確さは、各DNA二重鎖のワトソンとクリックの鎖を解離する必要性により、限定される。比較のための相補鎖がないと、塩基損傷、PCR、およびシーケンシングに起因していずれかの鎖へ導入されたエラー[11]は、真の突然変異として見せかけられることがある（図1A）。ユニークな分子識別子（UMI）を使用することで各DNA分子の両方の鎖を別々に追跡し、およびそれらの配列を比較することで各二重鎖の両方の鎖上の真の突然変異を検出することが可能[12]になる一方、それはNGSの根本的な限界を解決しない：二重鎖解離である。例えば、Duplex Sequencing[13]は、高い正確性のシーケンシングのゴールドスタンダードになっていて他の最近の方法で利用されており[14、15]、これはPCRおよびNGSの後でそれらを追跡し返すために各々の元の二重鎖上に二本鎖のUMIをタグ付けする。各々の元の二重鎖のワトソンとクリックの鎖に割り当てられたリード間で二重鎖コンセンサスを形成することにより、Duplex Sequencingは、1000倍またはそれよりも高い正確性を達成し、およびよって単一のDNA二重鎖内で真の突然変異を明らかにできる。しかしながら、NGSフローセル（例として、Illumina NovaSeq）上の最大100億もの他の鎖の中から両方の鎖を回収することは、100倍超のリードを必要とし[16]、これは常にNGSのスループットを減退させ、およびその適用可能性を著しく限定する。

【0188】

今日まで、Duplex Sequencingの高い非効率性を克服するために、いくつかの方法が試みられている。Duplex Proximity Sequencing（Pro-Seq）[17]は、二重鎖の元の鎖の5'末端を繋ぎ合わせるためにポリマーリンカーを使用するが、同じ反応において標的当たり複数のPCRプライマーを必要とすることは、Pro-Seqを、小規模の、標的化されたパネルに限定する。Pro-Seqの著者らは問題に対処するためのアイデアを提示したものの、彼らの提案はPCRと適合性がなく、それは非実用的となる。同じように、SaferSeqsもまたマルチプレックス化PCRを使用し、その適用は、小規模の、標的化されたパネルに限定される[18]。BotSeqs[14]およびNanoSeq[14、15]は、Duplex Sequencingができるように両方の鎖を回収する機会を増大させるために、繋ぎ合わせに代えて希釈を使用するが、しかし、そうすることによって、それはインプットDNAの0.001%のみしかシーケンシングしない。CypherSeq[19]は、環状化された二重鎖を生成し、ローリングサークル増幅がこれに続くが、しかし2本の鎖の間の対称性の欠如は、両方の鎖が実際にシーケンシングされたかどうかを、不明瞭にする。o2n-seq[20]およびCircle Sequencing[21]などのいくつかの技法は、二重鎖の単一の鎖のみを繋ぎ、およびよって、二重鎖コンセンサスを作り出す能力に欠ける。高い正確性およびスループットで二重鎖をシーケンシングすることの必要性にもかかわらず、ニッチな用途のための方法しか存在していない。よって、解離の前に両方の鎖の情報を繋ぎ合わせることが、NGSに高い正確性およびスループットで単一のDNA二重鎖を読み取ることを可能にするということが推論された。

【0189】

本開示は、単一リード対にて各DNA二重鎖の両方の鎖をシーケンシングするために、NGSの大規模並列の性質と第3世代シーケンシングの単一分子能とを組み合わせる方法に関する。エラー訂正のための元の二重鎖の連結（CODEC）と呼ばれているこのハイブ

リッドなアプローチにおいては、各分子がNGSを介して二重鎖コンセンサスを形成するのにそれ自体で十分になる（図1A）。直接それらを繋ぎ合わせることに代えて伸長のための鋳型として反対鎖を使用することにより、CODECは、強いヘアピン構造を形成することなしに、ワトソンとクリックの鎖の配列情報を物理的に一本鎖の形に連結する（図1B）。連結された配列間のあらゆる相違は、核酸塩基損傷によりまたは元のDNA二重鎖の1本の鎖に限った変化により作り出された非標準塩基対合、あるいはPCR増幅またはシーケンシングの間に導入されたエラーのいずれかを意味する。エラー率はシーケンシング技法そのものの以外の複数の要因に影響を受けるため、Duplex Sequencingを公平な比較のためにCODECと並行して行った。CODECを、標的化および全ゲノムNGSワークフローの両方で試験したことで、それがDuplex Sequencingと同じくらい正確にエラーを抑制したこと、および、突然変異シグネチャーを夫々100倍および280倍少ないリードで分析し、それによってNGSに「単一の二重鎖」の分解能が付与されたということを確認した。

10

【0190】

CODC構造は、市販のライゲーションをベースとしたNGS調製キットおよびCODCアダプター複合体を使用した合理化されたワークフローにより構築されることができる。最初に、典型的な二重鎖アダプターを、NGSのために必要とされる全ての要素を含有する4つのオリゴヌクレオチドからなるアダプター複合体に置き換えた。アダプターの二重鎖セグメントを、DNAハイブリダイゼーション熱力学に基づいて複合体全体を保持するように合理的に設計し（図1E）、および、硬い二重らせんの曲げ剛性を軽減するために一本鎖セグメントを導入した（図1F）。アダプターライゲーションがインプット分子の両末端を閉じた後、残りの3'末端で、反対鎖を鋳型として使用することにより各鎖を伸ばす鎖置換伸長が開始する。結果としてもたらされる構造は、2つの元の鎖が、中心にあるCODCリンカーおよび両側にあるNGSアダプターと連結されたものである。図1Bに描かれている分子プロセスは、市販のNGSライブラリー構築キットのアダプターライゲーションステップへと統合される（図1C）。

20

【0191】

連結構造を十分に利用するために、NGSライブラリーコンポーネントもまた、再配置した（図1D）。外側にNGSリードプライマー結合部位がある従来型のIllumina構造とは対照的に、リンカーなしの分子が読み取られることを防ぐために、リードプライマー結合部位を、中心にあるCODCリンカーへ移動させ、および外側へ向かってシーケンシングした（図2）。従来の位置にリードプライマー結合部位を有することは、悪い品質スコアを結果としてもたらしており、これはクラスター増幅における鋳型ホッピングに起因したと考えられるが、一方、リードプライマー結合部位をリンカーへ移動させたことは、この問題を克服した（図3B）。典型的にはリードプライマー結合部位の外側に配置されおよびインサートとは別々で読み取られるサンプルインデックスを、インサートのすぐ隣へ移動させた。インデックスをアダプターライゲーションの間に加え、およびそれらをインサートと共に単一のステップで読み取ることにより、CODECは、ユニークなデュアルインデックスを使用するというゴールドスタンダードよりもさらに良好にインデックスホッピングを抑制した[22]（0.056%対0.16%）。適切なクラスター同定、位相補正、および純度フィルターのための高い塩基多様性を確保するために、集合的に位置毎に4つ全ての塩基を有するような4つのサンプルインデックスのセットを設計した（図4）。インデックス付きプライマーはもはや必要でなくなったため、Illumina P5およびP7セグメントをアダプター複合体中に包含させて普遍的なプライマー結合領域として使用することが可能であった。

30

40

【0192】

記載されているアプローチの実現可能性を確認するために、最初に、CODECワークフローが、末梢血単核球からのフラグメント化されたヒトゲノムDNA（gDNA）をCODEC-NGSライブラリーへと変換することおよびそれをシーケンシングすることによって、意図するNGSライブラリー構造を作り出すことができたということを確認した。CODECリ

50

ードの新規な構造に起因して、「CODEC suite」と呼ばれているユーザーフレンドリーな分析パイプラインが、データを処理するために作り出された（「例1に関連する方法」を参照）。リードの半分よりも多くが、正しい構造を示し、および副産物のほぼ90%が、依然として、標準的なNGSとちょうど同じように二重鎖の片側に情報を保持しており、副産物であっても依然として有用なデータが得られる可能性があるということが提案される（図5A～5B）。

【0193】

次に、正しいCODEC構造を持つフラグメントが、有意により少ないリードを使用して Duplex Sequencingに匹敵するエラー率を提供することができるかどうかを探った。これを評価するために、直接比較（head-to-head comparison）を行った。Duplex Sequencingは遺伝子座当たりの高いシーケンシング深度を必要とするため、各方法で調製したNGSライブラリー上の、がん患者および健常ドナーからの20ngのセルフリーDNA（cfDNA）から構築された汎がんパネルを用いて標的富化を実施した。2名の個人の平均CODECエラー率（ 1.9×10^{-6} ）は、二重鎖シーケンシングのもの（ 5.9×10^{-7} ）と同様であり（図6A）、エラーの配列コンテキストに統計的に有意な差はなかったが、ただし健常ドナーにおけるC:G>T:Aは除き（図7）、これは改善された末端修復方法[15、23]を使用して解決することができると考えられた（図8A）。加えて、エラー率をフラグメントのいずれかの末端からの距離の関数としてプロットしたとき、二重鎖コンセンサスのフラグメントの末端に向かって上昇したエラー率がCODECおよびDuplex Sequencingデータから見られ、これは末端修復におけるエラー伝播の以前の報告と一貫している[15、23]（図9A）。この観察は、単一のCODECフラグメントを読み取ることが各鎖からの2つのDuplex Sequencingフラグメントを読み取ることと等価であるということを確認し、および、in silicoで元のDNA二重鎖の両末端から12塩基対（bp）をトリミングする必要があること[16]を肯定している。

【0194】

単純にフォワードおよびリバースリードのコンセンサスを形成することとは対照的に、元のDNA二重鎖を一緒に読み取ることによってCODECの潜在的なエラー抑制が独自に可能にできるということ、さらに確認するために、エラー率を次いで同じNGSデータからの3つの追加の方法：コンセンサスなし、ペアエンドリードコンセンサス（R1+R2、リード1およびリード2をコラプスしている）、および一本鎖コンセンサス（SSC、同じ元の鎖からのリードをコラプスしている）と比較した。興味深いことに、コンセンサスなしとR1+R2との間のエラー率のギャップは無視できるほど小さく（図6A）、NGSライブラリー分子には多数のエラーが物理的に存在しており、ライブラリー増幅の間に、または各ライブラリー分子がクラスター生成を受けるときに導入されていた可能性があったということが提案される（図1A）。SSCは、R1+R2およびコンセンサスなしのリードよりも正確であったが、ワトソンとクリックの鎖のコンセンサスがないことで、そのエラー率はCODECのそれよりも23倍高かった。

【0195】

同数のユニークなDNA二重鎖を発見するのに必要とされるリードの数を、次に探った。UMI、ならびに全てのリードをユニークな元の二重鎖へコラプスさせるための各分子の開始および停止のマッピング位置を使用したとき、Duplex Sequencingは700リードを受け取るまで二重鎖の再組み立てを始められなかったことが見出された（図8B）。対照的に、CODECは、350倍早くに二重鎖の再組み立てを始めた。要求されるリードの間のギャップは、少数の二重鎖を回収したときに最大化されており、CODECが浅い深度で幅広い遺伝子領域をシーケンシングすることを独自に可能とした可能性があったということが提案された。なお、各CODECリードは二重鎖コンセンサスを形成するのにそれ自体で十分であるので、CODECの単一のペアエンドリードさえも正確性が高かった（図6A）。これらの結果は、CODECが単一のペアエンドリードからの二重鎖シーケンシングの正確性を付与し、およびよって、有意により少ないリードを使用してより多くのDNA二重鎖をシーケンシングするということを提案する。

【0196】

次に、そうでなければ高コストに起因して非実用的となるところ、CODECがヒト全エクソームおよび全ゲノム「二重鎖」シーケンシングを可能にできるかどうかを決定しようと試みた。これを評価するために、CODEC全ゲノムシーケンシング（WES）を、その試料が以前に試験されたものであるヒトgDNA[16]に対して適用した。CODECは両方の試料のシーケンシングエラー率を低減させ、gDNAに対して100倍の改善があったということが見出された（図6A）。エラーの配列コンテキストを分析したことは、CODECが全てのタイプのSNVにわたって正確性を改善させたことを明らかにし、CODECがエラーを抑制する能力は特定のコンテキストに限られないということが提案された。なお、FFPE試料にはより多くのC>Tエラーが存在し（図10A）、脱アミノ化アーティファクトに起因したが[24]、これは改善された末端修復方法で解決できたものである[15、23]。

10

【0197】

次に、CODECおよびDuplex Sequencingを、Genome in a Bottle Consortium（GIAB）のパイロットゲノムNA12878のWGSに適用した[25]。Duplex Sequencingでは多数のユニークな二重鎖を回収することができなかったものの、公平な比較のために、各方法に同じ量のシーケンシングを割り当てた。費用対効果分析において、Duplex Sequencing（ 2.38×10^6 ）およびCODEC（ 3.37×10^6 ）の両方のエラー率は、標準的なNGS（ 2.2×10^4 ）のそれよりもはるかに低く（図11A）、これはR1+R2に類似する結果を示した（図12A）。これは、同じ条件下でCODECがDuplex Sequencingと同じくらい正確であるということを確認する。加えて、各配列コンテキストのエラー比は、CODECがDuplex Sequencingに類似するエラープロファイルを有するということを示した（図15）。シーケンシングコストの面で、Duplex Sequencingは他の方法よりも100～1000倍高価であり、その適用可能性は、標的化されたパネルに限定される。

20

【0198】

WGSについてのカバレッジ深度分析は、CODECがDuplex Sequencingよりも160倍大きいユニーク二重鎖深度を達成したということを実証した。GIAB v3.3.2 hg19高信頼性遺伝子領域（2.6B塩基）では、CODECは4.0の平均ユニーク二重鎖深度を有したが、一方、Duplex Sequencingは35%のより多いリードアウトプットを用いても0.025の平均深度しか有しておらず、それはほとんどのリードがそれらの元の二重鎖のマッチを見つけないためである（図11B）。よって、Duplex SequencingはWGSには実用的でないと結論付けられ、そしてDuplex SequencingのWGSデータは、この時点の後で二重鎖コンセンサスを生成することなく標準的なWGSデータとして扱った。他方、CODECは、単一の二重鎖を明らかにするという強みのおかげで、既存の方法のジレンマであった正確性とコストとの間の伝統的なトレードオフを打破する。

30

【0199】

CODECは、二次分析アプリケーションの最前線を押し広げる。WGS/WESにおけるDuplex Sequencingのエラー率を達成することは、CODECに多数の二次分析アプリケーションの限界を押し広げる能力を与える。1つのかかるアプリケーションは、全ゲノム生殖細胞系列小バリエーションコーリングをベンチマークすることである（SNV+インデル）。最先端の生殖細胞系列コーリングが通常30×の深度を必要とするということを確認しつつ、図8Bに示唆されるとおり低カバレッジでのCODECの潜在的可能性を試験するために、前述のNA12878試料のCODECデータを1×～5×の範囲にわたるカバレッジでR1+R2に対して比較した。GATK4[26]をバリエーションコーリングに使用し、および生殖細胞系列小バリエーションをベンチマークするためのGIABベストプラクティスがこれに続いた。全てのダウンサンプリングされた深度にわたって、CODECは、R1+R2を用いた標準的なWGSよりも90%少ない偽陽性（FP）を示し、コストは5%高い偽陰性（FN）であった（図11C、表1）。

40

表1．CODEC WGSと標準的なWGSとの間のSNP+小インデルコールの評価。表は、VcfEvalにより生成した。

50

【表 1】

方法	真陽性 ペー ス ライン	真陽性 コール	偽陽性	偽陰性	精度	感度	F値	タイプ	ds_ frac	FPPM	FNR	深度
CODEC	1149572	1149620	11542	2541289	0.9901	0.3115	0.4739	na1287 8_cds	0.1	4.482217885	0.68852 6593	1
	1786822	1786884	17480	1904039	0.9903	0.4841	0.6503	na1287 8_cds	0.2	6.788179573	0.51587 0691	2
	2222868	2222945	21154	1467993	0.9906	0.6023	0.7491	na1287 8_cds	0.3	8.21493997	0.39772 8978	3
	2544500	2544593	23624	1146361	0.9908	0.6894	0.8131	na1287 8_cds	0.4	9.174139258	0.31058 664	4
	2782660	2782755	25191	908201	0.991	0.7539	0.8564	na1287 8_cds	0.5	9.782667713	0.24606 1183	5
	2959035	2959132	25577	731826	0.9914	0.8017	0.8865	na1287 8_cds	0.6	9.932566873	0.19827 5353	6
	3089997	3090099	25548	600864	0.9918	0.8372	0.908	na1287 8_cds	0.7	9.921305019	0.16279 3287	7
	3186894	3186995	25231	503967	0.9921	0.8635	0.9233	na1287 8_cds	0.8	9.798201304	0.13654 0826	8
	3259747	3259845	24540	431114	0.9925	0.8832	0.9347	na1287 8_cds	0.9	9.529858508	0.11680 2706	9
	3314661	3314761	23829	376200	0.9929	0.8981	0.9431	na1287 8_cds	1	9.253748915	0.10192 4675	10
標準 NGS	1362228	1362280	263618	2328633	0.8379	0.3691	0.5124	na1287 8_r1r2	0.07 7	102.3733594	0.63090 9751	1
	2001169	2001242	332255	1689692	0.8576	0.5422	0.6644	na1287 8_r1r2	0.15	129.0278378	0.45779 5236	2
	2453841	2453932	366744	1237020	0.87	0.6648	0.7537	na1287 8_r1r2	0.23	142.4212889	0.33514 9306	3
	2778411	2778517	387976	912450	0.8775	0.7528	0.8104	na1287 8_r1r2	0.31	150.6665193	0.24721 1639	4
	2990581	2990709	392938	700280	0.8839	0.8103	0.8455	na1287 8_r1r2	0.38	152.5934614	0.18972 6927	5
	3169483	3169605	380937	521378	0.8927	0.8587	0.8754	na1287 8_r1r2	0.46	147.9329955	0.14125 7221	6
	3296709	3296823	354625	394152	0.9029	0.8932	0.898	na1287 8_r1r2	0.54	137.7149989	0.10678 8044	7
	3386823	3386934	320037	304038	0.9137	0.9176	0.9156	na1287 8_r1r2	0.62	124.2831022	0.08237 3424	8
	3443475	3443589	289759	247386	0.9224	0.933	0.9276	na1287 8_r1r2	0.69	112.5249499	0.06702 4567	9
	3491018	3491139	258701	199843	0.931	0.9459	0.9384	na1287 8_r1r2	0.77	100.4638927	0.05414 3586	10

ds_frac: ダウンサンプルフラクション。FPPM: 100万塩基当たりの偽陽性。FNR: 偽陰性率

【0200】

NGSデータをダウンサンプリングすることにより、FPおよびFNがどのように深度によって影響を受けるかもまた観察された。CODECにおけるより低いレベルのFPは、そのより低いエラー率を考慮すると、予想された結果であった。そのFNレベルは、標準的なWGSのそれよりもわずかに高く、それはおそらくより低いライブラリー変換効率がより高い二重化率を結果としてもたらしたからであるが、しかしCODECと標準的なWGSとのFN率間の差は、カバレッジが減少するにつれて小さくなった。それと同時に、低いFPを有するという利点はより低いカバレッジでより顕著になっており、浅い深度での適用はCODECを使用することから恩恵をより受け得るということを示唆する。

【0201】

低いカバレッジでのインデル検出についてのCODECの性能を考慮すると、CODECは、周知の突然変異ホットスポットであるマイクロサテライト (MS) のシーケンシングの正確性を改善することができると考えられた。実際に、NA12878におけるモノヌクレオチドMSの参照配列をCODECと標準的なNGSとの結果の間で比較したとき、CODECは、挿入および欠失の両方のエラーのより低い頻度を示した (図13A)。正しくないMS長さ

を伴うCODECリードの比率は0.45%であり、これは標準的なWGSのそれよりも12倍低かった。かかるより低い頻度は、8~18ヌクレオチドの様々な長さにわたって一貫して観察され（図13B）、とりわけ、より長いMSでの欠失についてそうであった。これらの知見は、がん免疫療法への応答の予測的なマーカーであることが示されているが液体生検試料からなどの低頻度で検出することは困難なままであったマイクロサテライト不安定（MSI）を検出するために、MS部位のリピート数／コピー数を読み取るのにCODECを使用できた可能性がある[27]ということを示唆する。よって、CODECが既存のMSI検出限界（0.1%）を改善するかどうか試験するために、MSI試料およびそれに対応する正常試料をシーケンシングした。in silico希釈シリーズからMSIを検出したとき、MSMuTect分析[28]は、標準的なNGSの検出限界を0.1%まで低減させたが、一方、CODECデータのそれは0.01%であった（図13C）。二次的アプリケーションにおける改善は、各NGSクラスター内で単一の二本鎖をシーケンシングすることによってCODECが何を可能にできるかを浮き彫りにする。

10

【0202】

CODECは、単一分子の突然変異シグネチャーを与える。低深度CODEC WGSで体細胞突然変異を検出することの潜在的可能性を探るために、CODEC（1×カバレッジ）により、およびバリエーションコーラーMutect2と対にして用いた標準的なNGS（12×カバレッジ）により検出されたMSI試料において突然変異のトリヌクレオチドコンテキストを比較した[29]。主な相違は、バリエーションコーラーが高いバックグラウンドノイズに起因して低い存在量の突然変異を捨てるのに対し、CODECは、高い存在量および低い存在量の突然変異の両方を受け入れられるということである（図14A）。例えば、統計的な閾値化なしで標準的なNGSの全ての一塩基置換（SBS）を受け入れることは、それらのコンテキストを、高い存在量の変異のそれと比較して有意に変化させたが（コサイン類似性=0.61）（図14B）、一方、CODECから全ての突然変異を受け入れることは、同じコンテキストを結果としてもたらした（コサイン類似性=0.98）。これは、標準的なNGSがランダムなエラーから低い存在量の突然変異を検出できない一方で、CODECのより低いエラー率は、複数回のリードなしに低い存在量の突然変異をコールすることを可能にできるということを示唆する（図15）。同じ傾向は、より低いCODECシーケンシング深度でさえも一貫して観察された。標準的なNGSおよびCODECのデータをダウンサンプリングしたとき、7×未満の深度の高い存在量の突然変異のコンテキストは、コサイン類似性を算出するのに参照として使用された12×の深度のそれから逸脱し始めた（図14C）。標準的なNGSの全ての突然変異を受け入れることは、一貫して、全てのシーケンシング深度において、同じ突然変異コンテキストを取得することができなかった。対照的に、CODECでは、0.025×でさえも同じ突然変異コンテキストが首尾よく得られ（コサイン類似性=0.95）、低い存在量の突然変異をコールするために必要とされるシーケンシング深度が280倍低減された。

20

30

【0203】

希少な突然変異を検出するCODECの能力を確認した後、CODECによってのみ専属的に検出される突然変異が真の体細胞突然変異であるかどうかを決定することを次に試みた。サブクローン性の体細胞突然変異を有する腫瘍試料は、正常試料よりもCODECに専属的な低い存在量の突然変異を多く示すであろうと仮定した。実に、かかる突然変異率は腫瘍試料では2.7倍高く（図14D）、および、MMR欠損症に関連するいくつかの突然変異シグネチャーにおいて富化されるT>C置換では、差は6.4倍にまで上った[30]。専属的な突然変異をさらに分析するために、Catalogue Of Somatic Mutations In Cancer（COSMIC）突然変異シグネチャーを、次いで種々の突然変異のセットから抽出した（図14E）。CODECは、Mutect2データ中のシグネチャーのみならず、1つ以上のMSIシグネチャー（SBS21）もまた検出しており、および、標準的なNGSからの全ての突然変異を利用したことはMSIシグネチャーの大半をキャンセルした。なお、SBS1シグネチャーは、腫瘍および正常細胞の両方において観察される5-メチルシトシンからチミンへの脱アミノ化から来ている。CODECにより検出されたがMutect2により捨てられた突然変

40

50

異のシグネチャーは、CODECからの全ての突然変異のそれに似ており、それらもまた低い存在量の体細胞突然変異であったということが提案された。興味深いことに、Mutect 2により見逃された2つの新しいシグネチャーの1つであるSBS29は、腫瘍および正常組織の両方に影響を与えていた可能性があるタバコ喫みに関連し、両方とも患者の結腸からのものである。正常組織はCODECデータにおいてMSIシグネチャーのいずれをも示さなかったということ、およびMutect2により捨てられた標準的なNGSからの突然変異は依然として散乱したシグネチャーを示したということもまた確認された。よって、CODECの単一の二重鎖の分解能は、バリエーションコーラーMutect2と対にして用いた標準的なNGSよりも良好に、有意により少ないシーケンシングで、突然変異シグネチャーを検出することを可能にできた。

10

【0204】

各DNA二重鎖の両方の鎖を物理的に繋ぎ合わせるにより、CODECは、各NGSクラスターが第3世代シーケンシングのように単一の二重鎖の分解能を有するようにすることを可能にできる。二重鎖を解離させることおよび二重鎖コンセンサスを形成するためにそれらを再び回収することを必要とするDuplex Sequencingと違って、CODECは、似たように高い正確性で、しかしそれよりも100倍少ないリードで、真の突然変異をエラーから判別する。このアプローチでは、最初に、汎がんパネルにより富化されたcfDNAを使用すること、続いて他の主なNGSワークフロー（例として、WESおよびWGS）にわたるその一貫性を試験することが示される。CODECのさらなる用途を提示するために、それがFPをとりわけ浅いシーケンシング深度で抑制したこと、MS部位でインデルエラーを低減させたこと、および極めて低いシーケンシング深度でがん患者からの突然変異シグネチャーを検出したこともまた示された。

20

【0205】

直接比較において、CODECはDuplex Sequencingと同じくらい正確であるが、Duplex Sequencingの主な限界となっていたシーケンシングに必要な要件はるかに低いということが示された。エラー率はシーケンシング技法そのものの複数の要因に影響を受けるため、あらゆる直接的な比較は、他の全てが同じであることを必要とする。同じ実験的およびコンピューターのプロトコルが、適用可能なときいつでも使用されており、インプット試料および質量、試薬、標的領域、エラーの定義、および精密比較のための分析パイプラインを包含した。

30

【0206】

CODECアダプター複合体は、2つの連続するライゲーションを通じて付着させられる：二分子ライゲーションとそれに続く単分子ライゲーションである。アダプター濃度を増大させることが変換効率をもまた増大させる典型的な二分子ライゲーションと違って、単分子ライゲーションは、アダプター濃度が高すぎると望ましくないものとなり得る。帰結として、CODECアダプター複合体の現在のバージョンは、2つのライゲーションの間のバランスを取る必要がある。

【0207】

この研究の全体を通して市販のキットの従来型の末端修復／dAテーリングを使用した、CODECの前に新しい末端修復方法を採用すれば、正確性をさらに改善することができる。最近の研究[15、23]は、オーバーハングでの塩基損傷および元の二重鎖の一本鎖破断が、一方の鎖上のエラーを両方の鎖にコピーされるようにする可能性があるということを報告している。この研究ではまた、DNAフラグメントの末端に向かうにつれてエラー率が高くなったということも間接的に観察された（図9A）。かかるエラーが二重鎖コンセンサス状に現れて偽の突然変異を結果としてもたらす一方で、新しい末端修復方法は、エラー伝播を防ぎ、そして、CODECが新しい末端修復方法と組み合わせられるとき、なおさらに高い正確性が達成可能になるであろうと考えられる。

40

【0208】

単一のCODECフラグメントを読み取ることは、元の二重鎖の両方の鎖を読み取ることと等価であり、これにより同じ遺伝子座を複数回読み取る必要性がなくなる。1×リード

50

深度でのCODECの低いエラー率は、診断からバイオインフォマティクスまでの分野にわたる様々な適用への可能性を開く。1つの例は、希少な体細胞突然変異に限られた数のリードで発見することであり、これはエラー率が低くなると真の突然変異を見つける可能性がより高くなる[32]。もう一つの例は、マイクロバイーム解析のためのショットガンメタゲノムシーケンスであり、そこでは偽SNVをCODECで抑制することが、正しくない分類学的分類、および微生物の多様性の不適切な評価を防ぐ[33]。de novoアセンブリにおいては、より低いエラー率が、de Bruijnグラフパラダイムにおけるより連続的なアセンブリ、およびオーバーラップレイアウトコンセンサスパラダイムにおけるより速い処理に貢献する[34]。

【0209】

要約すると、CODECは、各々の元のDNA二重鎖の両方の鎖を連結することにより、標準的なNGSの設備を大規模に並列的な単一の二重鎖のシーケンサーへと転換する。この戦略は、Duplex Sequencingと同じくらい正確であるが有意により少ないリードでのSNVおよびインデルの検出を、および0.025×までの低いシーケンシング深度でのがんシグネチャー検出を可能にできる。その上、標的化シーケンシングからWGSまでの範囲にわたるCODECの適用可能性は、他の高精度NGS方法とは一線を画す。よって、CODECは、液体生検から早期がんまたは微小残存病変を、液体または腫瘍生検から臨床的に対処可能な突然変異を、血液試料から未確定の潜在性のクローン造血（CHIP）を、正常組織試料において体細胞モザイクを、検出することなどの、多数の重要な生物医学用途に、およびその先へ、広く実施可能であると考えられる。

【0210】

例1に関連する方法

DNA試料およびオリゴヌクレオチド

コホート05-246からの患者315のセルフフリーDNA、および、コホート05-055からの患者95のFFPEおよびgDNAの両方は、他の研究からのものであった[16]。患者19のMSI DNAもまた、他の研究からのものであった[27]。NA12878はCoriellから購入した。全ての試料は、低TE緩衝液（10mM Tris-HCl、0.1mM EDTA、pH 8）中に保管し、および、cfDNAを除いて150bpの平均サイズを有するようにCovaris 超音波装置でフラグメント化した。CODECのための全てのオリゴヌクレオチドは、Integrated DNA Technologies（IDT）により合成され、およびPAGE精製を経た（表2）。Duplex Sequencingのためのアダプターは、IDTによりBroad Instituteのためにカスタムで注文された。

表2．オリゴヌクレオチドの配列。

10

20

30

40

50

【表 2】

CODEC アダプター		配列番号
LD4-adap5-1	AATGATACGGCGACCACCGAGATCTACACCTTGAACGGACTGTCCAC*T	1
LD4-adap5-2	AATGATACGGCGACCACCGAGATCTACACGAGCCTACTCAGTCAACG*T	2
LD4-adap5-3	AATGATACGGCGACCACCGAGATCTACACGCTTGTAAAGCAGGTAG*T	3
LD4-adap5-4	AATGATACGGCGACCACCGAGATCTACACCAAGCGTCTTACATGGTC*T	4
LD4-adap7-1	CAAGCAGAAGACGGCATACGAGATCACCAGAGCGTTAGACTAC*T	5
LD4-adap7-2	CAAGCAGAAGACGGCATACGAGATGTGTGCGAACACTTGACGG*T	6
LD4-adap7-3	CAAGCAGAAGACGGCATACGAGATCTGATCTTCAGCTGACTG*T	7
LD4-adap7-4	CAAGCAGAAGACGGCATACGAGATGAATCTGAGGCACTGTAC*T	8
LD4-brid5-1	P- GTGGACAGTCCGTTCAAGNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTT TACATAGTTATCCGCTAGACTCTGACGTGTTGATCCTCGAAGC	9
LD4-brid5-2	P- CGTTGACTGAGTAGGCTCNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTT TACATAGTTATCCGCTAGACTCTGACGTGTTGATCCTCGAAGC	10
LD4-brid5-3	P- CTAACCTGCCTTACAAGCTNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTT TACATAGTTATCCGCTAGACTCTGACGTGTTGATCCTCGAAGC	11
LD4-brid5-4	P- GACCATGTAAGACGCTTGANNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTT TACATAGTTATCCGCTAGACTCTGACGTGTTGATCCTCGAAGC	12
LD4-brid7-1	P- GTAGTCTAACGCTCGGTGNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC CAATCTATAAGTTGCTTCGAGGATCAACACGTCAGAGTCTAGC	13
LD4-brid7-2	P- CCGTCAAGTGTTTCGACACNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC CAATCTATAAGTTGCTTCGAGGATCAACACGTCAGAGTCTAGC	14
LD4-brid7-3	P- CAGTCAGCTGAAGATCAGTNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC CAATCTATAAGTTGCTTCGAGGATCAACACGTCAGAGTCTAGC	15
LD4-brid7-4	P- GTACAGTGCCTCAGATTCAANNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC CAATCTATAAGTTGCTTCGAGGATCAACACGTCAGAGTCTAGC	16
ハイブリダイゼーションキャプチャーのためのCODECブロック		
LD4-HybBlk1	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTTACATAGTTATCCGCTAGACTC TGACGT-3C	17
LD4-HybBlk2	AGATCGGAAGAGCACACGTCTGAACTCCAGTCACCAATCTATAAGTTGCTTCGAGG ATCAAC-3C	18
凡例	Illumina P5 Illumina P7 サンプルインデックス UMI リードプライマー結合領域 インサート	
ヌクレオチド間の"*"はホスホロチオアートの修飾を表す。 "P-"は5'-リン酸化を表す。 "-3C"はC3スベーターを表す。		

【0211】

CODECアダプターの調製

4つの100 μMオリゴヌクレオチドを低TE緩衝液および100mM NaClで5 μMに希釈し、続いて85℃にて3分間加熱し、-1℃/分で20℃まで冷却し、および室温で12時間インキュベートすることによって、CODECアダプター複合体を調製した。Mastercycler X50 (Eppendorf) およびMAXYMum Recovery PCRチューブ (Axygen) をアニーリングのために使用した。アニーリングされたアダプター複合体を、将来的な使用のために-20℃に保った。IlluminaのためのNEBNext Ultra II DNA Library Prep Kit (New England Biolabs, NEB) を使用し、および、製造者のマニュアルに

従ったがいくつかの例外があった：

1. ライゲーション時間を1時間へ増大させ、5 μ Mアダプターコンプレックスを使用の前にアダプター希釈緩衝液 (10mM Tris-HCl、1mM EDTA、10mM NaCl、pH 8) で500nMへ希釈し、およびNEBアダプターを置き換えた、
2. 3 μ Lの50-デアデニラーゼ (NEB) をライゲーション反応に添加した、
3. 鎖置換伸長 (試料 40 μ L、10 $\times$ 緩衝液 10 μ L、0.2mM dNTP、ポリメラーゼ 1 μ L、100 μ LまでのH₂O) を、phi29 DNAポリメラーゼ (New England Biolabs) で30°Cにて20分間行い、0.75 $\times$ 体積比での標準的なAMPure XP (Beckman Coulter) クリーンアップがこれに続いた、
4. 2分間の延長を伴って製造者のマニュアルに従うことによって、KAPA HiFi HotStart ReadyMixおよびxGen Library Amplification Primer Mix (IDT) をPCRのために使用した。
5. そして、0.75 $\times$ 体積比でのAMPure XPクリーンアップを、PCRの後で2回行った。

標準的なNGSおよびDuplex Sequencingのためのライブラリーを、他の箇所で記載されたとおりに調製した[16]。全てのライブラリー調製は、twin.tec PCR Plates Lo Bind 250 μ L (Eppendorf) 上で行った。ライブラリー定量化は、Bioanalyzer DNA High Sensitivityチップ (Agilent) と対にして用いたQubit dsDNA HSキット (Invitrogen) で行った。

【0212】

富化。

汎がんおよびWES富化は両方ともxGen HybridizationおよびWashキットならびにxGen Blocking Oligos (IDT) を用いて製造者のマニュアルに従って行った。キャプチャープローブについては、xGen Pan-cancer Panel (IDT、800kb)、およびTwist BioscienceによるBroad InstituteのためのカスタムWESパネルを使用した。

【0213】

シーケンシング。

標準的なNGSおよびDuplex Sequencingを、汎がんパネルおよびWGSについて、Illumina HiSeq 2500 Rapid Run (300サイクル) を用いて行った。CODECを、汎がんパネルおよびWGSについて、Illumina HiSeq 2500 Rapid Run (500サイクル) を用いて、およびWGSおよびWESについてNovaSeq SP (500サイクル) を用いて行った。余剰のサイクルは、CODEC構造を確認するために使用した。

【0214】

CODECデータ処理。

ユニークなCODECリード構造に起因して、CODECデータを処理するためにCODECsuite (github.com/broadinstitute/CODECsuiteで利用可能) (その全内容は参照により本明細書に組み込まれる) が開発された。CODECsuiteは、C++14およびpython 3.7で記述されており、およびsnakemake6.0.3がワークフローマネジメントシステムとして使用された。CODECsuiteは、4つの主要なステップからなる：デマルチプレックス化、アダプタートリミング、コンセンサスコーリング、および正確性の計算。最初の3ステップは、CODECデータに特異的である。ワークフローにはまた、BWA、FgbioおよびGATK Illumina bcl2fastqなどの標準的なツールが関与し、fastqファイルを生成するために使用されたが (-R -oを伴い、CODECsuiteがデマルチプレックス化するためサンプルシートなしで)、しかしこれはsuiteには包含されていない。データ処理を加速するために、fastqファイルをバッチに分割しおよびそれらを並列で処理することが推奨される。この例においては、40バッチを使用して、800M NovaSeqリードの前処理 (デマルチプレックス化およびアダプタートリミング) には、各バッチを単一のCPUおよび8GのRAMを使用して実行したHPC環境においてほんの数時間だけかかった。デマルチプレックス化およびアダプター除去の後、BWA(0.7.17-r1188)を使用して、未加工リードをヒト参照hg19に対してマッピングした。PCR二重鎖をコラプスし、および本質的に一本鎖コンセンサス (SSC) のリードを形成するために、Fgbio (github.com/fu

10

20

30

40

50

lcrumgenomics/fgbio) を次いで使用した。これらのSSCリードを、次いで再度BWAを使用して参照ゲノムに対してマッピングした。次に、R1とR2との間の二重鎖コンセンサスリードをSSCアラインメントから生成した。コンセンサス塩基は、R1またはR2からの塩基のいずれかが30未満の塩基品質を有する場合にはフィルタリングされた。二重鎖コンセンサスリードはBWAを使用して参照ゲノムにアライメントされ、および後続するアラインメントはGATK3 (hub.docker.com/r/broadinstitute/gatk3) を使用してインデル再アライメントされた。

デマルチプレックス化。

CODECシーケンシングリードは、ユニークな分子識別子(UMI)配列: NNNまたはNNAまたはNNNT (NNNはランダムな3-mer) で始まり、および、18bpの試料バーコード、および次いでT塩基がこれに続く(図11A~11C)。デマルチプレックス化するために、CODECSuiteは、バーコード(50末端からの4番目~21番目の塩基)を抽出し、および、サンプルインデックス(SD)割り当てのためにsmith-waterman(SW)アルゴリズムを使用する[1]。抽出されたバーコードが、唯一のサンプルインデックスから離れているのが編集距離x(デフォルトでは3)以内であれば、それはマッチとして宣言される。次いで、2つの抽出されたバーコード(リードペアの各末端から1つずつ)が両方とも、予想されたSID(P5およびP7)にマッチした場合にのみ、リード対は首尾よくデマルチプレックス化される。首尾よくデマルチプレックス化されたリードのみが、後続するステップに使用され、および、予想されたSIDが、後続するアダプタートリミングステップのためのリード名にて保存される。さらには、リード対からの2つのバーコードがキメラのサンプルインデックスの組み合わせにマッチしたときは、CODECSuiteはまた、2つのインサートをアラインメントすることによりインデックスホッピングもチェックし、それらがオーバーラップする場合には、それらをホッピングリードとしてフラグ付けする。そうでなければ、混在したインデックスは分子間副産物の結果である可能性が最も高い。

【0215】

アダプターのトリミングおよび副産物クリーニング。

デマルチプレックス化ステップは、リード名にSIDを加えるが、しかしリードシーケンシングは変化させない。アダプタートリミングステップは、アダプター配列をリードおよびuBAM(マッピングされていないBAMフォーマット)としてのアウトプットから除去する。R1およびR2の最初の3個の塩基が切り取られ、およびハイフンで区切られてbamレコードにおける「RX」タグに加えられる。各正しいCODECリードは、50アダプターおよび見込み得る30アダプターを(シーケンシング配向にて)含有する。R1のSIDはR1の50アダプターをトリミングするための鋳型として使用され、およびR2のSIDの逆相補鎖はR1の30アダプターをトリミングするために使用され、R2をトリミングするための逆も成り立つ。再び、SWアルゴリズムが、マッチを見つけるために使用される。リードは、50アダプターがR1およびR2の両方に見出された場合に基づいてグループ化される。換言すると、両方において50アダプターが見出されるリード対のみが、潜在的に正しいリードとみなされる。しかしながら、少数の副産物もまた、この基準を満足することがあり得る。ゆえに、30アダプターが存在するのであればそれをチェックすることが重要である。30アダプターが見出され、かつインサート部分が小さすぎる場合(例として、<15bp)、リードは捨てられる。R1およびR2の両方が捨てられる場合、この鋳型は空のライゲーションであるものとみなされる。リード末端の一方のみが捨てられる場合、それはダブルライゲーションとして分類される。副産物形成および定量化の概要は、これもまたCODECSuite githubサイトで利用可能であるカスタムpythonスクリプトにより作成される。

【0216】

ReadPair/二重鎖のコンセンサス。

CODECSuiteは、de novoまたは参照ベースのコンセンサスを生成できる。参照ベースのコンセンサスは、より良好な正確性を有し、およびこの例を通してずっと使用される

。2個のアラインメントされた塩基（または挿入もしくは欠失に関してはギャップ）が一致すればコンセンサス塩基が、およびそうでなければNが形成される。CODECsuiteはペアエンドリードを保つが、R1およびR2の両方に対してリード配列をコンセンサス配列に置き換える。配列の品質およびUMIなどのその他の補助タグは無傷に保たれる。コンセンサスは、uBAMフォーマットで生成される。

【0217】

アラインメントの正確性。

CODECsuiteは、アラインメント後の塩基レベルの正確性を評価するための手軽で素早いツールを提供する。それは、bedファイル領域（GIAB高信頼性領域など）内の塩基を評価し、および、通例、生殖細胞系列バリエーションおよび体細胞バリエーションのために夫々、VCFおよび／またはMAFファイルにおけるバリエーションに対してマスキングする。それはリードレベルで（例として、mapqまたは編集距離）および塩基レベルで（塩基品質により）フィルタリングする。それはまた、両方のフラグメント末端をトリミングする能力も提供し、および、対になったリードのオーバーラップした部分のみを評価する。それはフラグメント、サイクルおよび試料レベルに対し正確性を計算する。全ての非参照塩基について、それは塩基置換、品質スコア、リード上の位置および参照などの詳細をアウトプットすることができ、そうすることで、ポストプロセッシングスクリプトが一重鎖のコンテキストでエラー率を生成することができる。

【0218】

Duplex Sequencingのデータ処理。

この例において使用されたDuplex Sequencingデータ処理は、以前に記載されている[16、31]。手短には、二重鎖コンセンサスを生成するためおよびコンセンサスリードをフィルタリングするためFgbioを使用した。全ワークフローおよびさらなる詳細は、CODECsuite githubで利用可能である。見込み得る最良の二重鎖回収を得るための要求が各鎖の1コピーに緩和されたDuplex Sequencing WGSを除いて、各鎖の2つのコピーを伴うリードファミリーが、二重鎖コンセンサスを生成するために必要とされた。

【0219】

二重鎖のリカバリーおよびあるファミリーサイズへのダウンサンプリング。

2つのカスタムPythonスクリプトを使用して、図8Cおよび図6Bを夫々生成した。二重鎖のリカバリーのために、標的ごとのプレコンセンサスファミリー割り当て済みリード（Fgbio GroupReadsByUmiの後）を、 10^{-4} から始まる対数間隔分数（log spaced fractions）でサブサンプリングし（np.logspace(-4, 0, 30)）、および、各ダウンサンプリングフラクションにて形成された二重鎖の数を算出した。これは、限られたシーケンシングのみが与えられた場合（例として、<100リード対）の状況の理解を可能にした。エラー率へのファミリーサイズの影響を理解するために、ダウンサンプリングのための別のpythonスクリプトを記述した。この試料においては、正確なファミリーサイズを有する二重鎖コンセンサスの数（前もってコラプスされた（pre-collapsed）未加工リードの数）は限られており、およびよって、より信頼性の低い結果が得られた。よって、厳密により大きいファミリーサイズのファミリーを使用し、そして標的ファミリーサイズへダウンサンプリングした。各鎖からのリードの数の間で等しいかまたは近い比率を維持することもまた試みた。

【0220】

キャプチャーシーケンシングにおけるエラー率。

この全体を通して、エラー率は、参照ゲノム（hg19）へのマッピングの後の塩基レベルでの置換エラー率として定義した。Illuminaシーケンサーは通常100倍少ないインデルエラーを生成するため、一般的なエラー率の算出には置換エラー率を使用しており、およびこの定義は他の研究が報告してきたものに準拠した[15]。マッチ正常（match normal）を用いたパネルシーケンシングについては、以前の研究に従いエラー率を算出するためにMiredasを使用した[16]。cfDNAおよびマッチした正常試料の両方からの二重鎖BAMを同じやり方で生成し、および同じフィルターのセットに適用した：1．二次的お

よび補助的アライメントなし；2. $\text{Mapq} \geq 60$ ；3. ソフトクリッピングを除外したリードと参照ゲノムとの間のLevenshtein距離（L距離） ≤ 5 および非N塩基の数のL距離 ≤ 2 ；4. 両方のフラグメント末端からの距離が12bp以内の塩基を除外。エラーと真の突然変異との混乱が起きないように、生殖細胞系列SNVは事前に計算し、およびGATK4（HaplotypeCaller）をDuplex Sequencing標準試料から使用したが、これはそれらがより高いオンターゲット比率およびひいてはより高いカバレッジを有するためである（89%対CODECの40%）。患者試料については、3つの体細胞SNV（メジアンVAF=0.26、範囲0.24～0.28）が、キャプチャーされた領域（表3）においてMuTectを使用して見出された[32]。

表3. IDT汎がんパネルにおいて見出された患者315の体細胞SNV（800kb）。 10

【表3】

Hugoシンボル	染色体	開始位置	参照アレル	腫瘍配列 アレル2	腫瘍 フラクシオン	腫瘍参照 カウント	腫瘍代替 カウント	正常参照 カウント	正常代替 カウント
PIK3CA	3	178936091	G	C	0.245614	43	14	178	0
CDKN2A	9	21971000	C	A	0.264151	39	14	158	1
ARHGAP35	19	47507737	C	T	0.28125	46	18	140	0

【0221】

それらの体細胞突然変異（患者試料のみ）および生殖細胞系列突然変異は、エラー率を計算するときにマスキングされる。エラー率は、cfDNA試料についてのみ報告され、およびマッチ正常を、見込み得る生殖細胞系列（コールされなかったかまたはHaplotype Callerによる品質フィルターを通過しなかった）およびCHIPをフィルタリングするために使用した。それにより、あらゆるSNV位置もまた、少なくとも1つの二重鎖リードの支持がマッチ正常サンプルにあった場合にはマスキングしたところ、それはCHIPが極めて低い突然変異頻度で生じることができるためである。最後に、アラインメントエラーから生まれる可能性がある置換を除去するために、特異性チェック[16]をcfDNA試料に対して行った。 20

【0222】

全ゲノムシーケンシングにおけるエラー率。

WGSエラー率を、数点の違いを除いてキャプチャーデータと似たように計算した。1、C++プログラム「codec accuracy」を、そのスピード改善のため、Miredasに置き換わるものとして使用した。2、v3.3.2 GIAB NA12878高信頼性VCFおよびBEDファイルを、生殖細胞系列マスクおよび評価領域として使用した。3、マッチ正常はなかった。4、特異性チェックも大きなゲノムについては極めて遅いため省略した。ダウンサンプリングされたWGSにおける生殖細胞系列SNVおよび小インデルコーリング。生殖細胞系列バリエーションコーリングを評価するために、HiSeq 2500 Rapid RunおよびNovaSeq SP CODECのデータをマージさせた。マージされたCODECおよび標準のWGSのNA12878試料を、高信頼性領域において1～10×（ステップサイズは1×）のカバレッジ中央値まで、GATK DownsampleSamを使用してダウンサンプリングした。次に、CromwellおよびTerraワークフロー（webリソースで利用可能）を介してGATK4.1.4.1ベストプラクティスパイプラインを実行し、およびGoogle Cloud Platform上で計算をした。v3.3.2高信頼性VCFおよびBEDファイルをインプットとして使用して、ジェノタイプリングエラー（ヘテロ接合型バリエーションがホモ接合型としてコールされる場合、またはその逆の場合）にペナルティを課すことなしに、偽陽性（FP）および偽陰性（FN）をSNVおよびインデル（<50bp）に対して算出するためにRTG vcfevalを使用した。次いで、100万塩基当たりのFPを、高信頼領域のサイズに対する正規化により算出し、およびFN比率を、FNを真のバリエーションの総数で割ることにより算出した。 30 40

【0223】

マイクロサテライト不安定性検出。

PCRスタッターエラーを訂正しおよびよってMSI検出のためにバックグラウンドノイ 50

ズを低減させるCODECの能力を実証するために、NA12878についてフルカバレッジCODECコンセンサスBAMとフルカバレッジの標準的なNGS R1R2コンセンサスBAMとを互いに比較した。サイズ8~18ntのホモポリマーについてhg19をスキャンするために、MSIsensor-proを使用した。MSIsensor-proがマッピング品質または二次的アラインメントのフィルターを有しないので、BAMは、SAMtoolsを使用して、 $\text{mapq} \geq 60$ を要求し、二次的または補助的アラインメントなしとすることにより、予めフィルタリングした。そして、それを再度、それらの予め選択された部位で異なる長さのホモポリマーを支持するリードの数を計数するために使用した。あらゆる生殖細胞系列バリエーションとオーバーラップするかまたはごく近接している($\pm 5 \text{ bp}$)あらゆるホモポリマー部位を除去した。その後、ホモポリマー部位の参照長さが、真の長さとなされた。そして、リードからの観察された長さ分布を、真に対して比較した。結果は、染色体1のみから生成した。

10

例1に関連する参考文献

【表4-1】

- [1] Lennon, A. M. et al. Feasibility of blood testing combined with PET- CT to screen for cancer and guide intervention. *Science* 369, eabb9601 (2020).
- [2] Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature* 575, 299–309 (2019).
- [3] Griffith, O. L. et al. The prognostic effects of somatic mutations in ER-positive breast cancer. *Nat. Commun.* 9, 3476 (2018).
- [4] Gerlinger, M. et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* 366, 883–892 (2012).
- [5] D’Gama, A. M. & Walsh, C. A. Somatic mosaicism and neurodevelopmental disease. *Nature Neuroscience* 21, 1504–1514 (2018).
- [6] Blauwkamp, T. A. et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat. Microbiol.* 4, 663–674 (2019).
- [7] Brazhnik, K. et al. Single-cell analysis reveals different age-related somatic mutation profiles between stem and differentiated cells in human liver. *Sci. Adv.* 6, (2020).
- [8] Wenger, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 1155–1162 (2019).
- [9] Karst, S. M. et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat. Methods* 18, 165–169 (2021).
- [10] Shendure, J. et al. DNA sequencing at 40: past, present and future. *Nature* 550, 345–353 (2017).

20

30

40

【表 4 - 2】

- [11] Arbeithuber, B., Makova, K. D. & Tiemann-Boege, I. Artifactual mutations resulting from DNA lesions limit detection levels in ultrasensitive sequencing applications. *DNA Res.* 23, 547–559 (2016).
- [12] Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9530–9535 (2011).
- [13] Schmitt, M. W. et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14508–14513 (2012). 10
- [14] Hoang, M. L. et al. Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 113, 9846–9851 (2016).
- [15] Abascal, F. et al. Somatic mutation landscapes at single-molecule resolution. *Nature* 593, 405–410 (2021).
- [16] Parsons, H. A. et al. Sensitive Detection of Minimal Residual Disease in Patients Treated for Early-Stage Breast Cancer. *Clin. cancer Res.* 26, 2556–2564 (2020).
- [17] Pel, J. et al. Duplex Proximity Sequencing (Pro-Seq): A method to improve DNA sequencing accuracy without the cost of molecular bar-coding redundancy. *PLoS One* 13, 1–19 (2018). 20
- [18] Cohen, J. D. et al. Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat. Biotechnol.* (2021). doi:10.1038/s41587-021-00900-z
- [19] Gregory, M. T. et al. Targeted single molecule mutation detection with massively parallel sequencing. *Nucleic Acids Res.* 44, e22 (2016).
- [20] Wang, K. et al. Ultrasensitive and high-efficiency screen of de novo low-frequency mutations by o2n-seq. *Nat. Commun.* 8, 15335 (2017).
- [21] Lou, D. I. et al. High-Throughput DNA sequencing errors are reduced by orders of magnitude using Circle Sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19872–19877 (2013). 30
- [22] Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40, e3–e3 (2012).
- [23] Xiong, K. et al. Duplex-Repair enables highly accurate sequencing, despite DNA damage. *bioRxiv* (2021) doi:10.1101/2021.05.21.445162.
- [24] Kim, S. et al. Deamination Effects in Formalin-Fixed, Paraffin-Embedded Tissue Samples in the Era of Precision Medicine. *J. Mol. Diagnostics* 19, 137–146 (2017). 40

【表 4 - 3】

- [25] Zook, J. M. et al. An open resource for accurately benchmarking small variant and reference calls. *Nat. Biotechnol.* 37, 561–566 (2019).
- [26] DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498 (2011).
- [27] Yu, F. et al. NGS-based identification and tracing of microsatellite instability from minute amounts DNA using inter-Alu-PCR. *Nucleic Acids Res.* 49, e24–e24 (2021). 10
- [28] Maruvka, Y. E. et al. Analysis of somatic microsatellite indels identifies driver events in human tumors. *Nat. Biotechnol.* 35, 951–959 (2017).
- [29] Benjamin, D. et al. Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 861054 (2019). doi:10.1101/861054
- [30] Tate, J. G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 47, D941–D947 (2019).
- [31] Gydush, G. et al. MAESTRO affords ‘breadth and depth’ for mutation testing. *bioRxiv* (2021) doi:10.1101/2021.01.22.427323. 20
- [32] Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219 (2013).
- [33] May, A., Abeln, S., Crielaard, W., Heringa, J. & Brandt, B. W. Un- raveling the outcome of 16S rDNA-based taxonomy analysis through mock data and simulations. *Bioinformatics* 30, 1530–1538 (2014).
- [34] Limasset, A., Flot, J. F. & Peterlongo, P. Toward perfect reads: Self-correction of short reads via mapping on de Bruijn graphs. *Bioinform-atics* 36, 1374–1381 (2020). 30
- [35] Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197 (1981).

【 0 2 2 4】

CODEC (CDS) アダプター複合体は、ハイブリダイズされた4つのオリゴヌクレオチド (オリゴ) からなり、連結とアダプター付着との両方に要求される各要素全てを包含するように設計されている。ある態様において、全体として維持されるために、二本鎖領域 (1 および 4) の長さおよびハイブリダイゼーション ΔG° が十分に強いことは、極めて重要である。DNAハイブリダイゼーション熱力学に基づいて、 $>15\text{ bp}$ および $<-20\text{ kcal/mol}$ を有するように領域 1 を設計したところ、それは良好に機能した。領域 4 は、2つのオリゴを持つ必要があるため、余剰の長さ (30 bp) を与えられた。 40

【 0 2 2 5】

例 2 - メチル化特異的 CODEC シーケンシング

この例は、DNA試料の改善された突然変異およびメチル化シーケンシングのために使用することができる、「メチル化特異的 CDS」 (または同等に、「メチル化特異的 CODEC」) と称される態様を説明する。

【 0 2 2 6】

この例は、問われた DNA 試料からの、DNAメチル化ならびに突然変異についての情報 50

の抽出ができるようにしている。がんを包含する、いくつかの分野において、臨床試料からDNAメチル化情報を抽出することへの関心が高まっている。例えば、液体生検からメチル化したDNAのがん特異的フィンガープリントを抽出することが、最近、多発性がんの早期検出のためのアプローチにつながっている¹。

【0227】

DNA試料からのメチル化情報の抽出ができるようにするために、およびメチル化感受性のシーケンシングを行うために、ほとんどの場合、試料の増幅を行うことに先立ち、化学的または酵素的な脱アミノ化ステップが試料に適用される。このステップは、メチル化されていないシトシンのウラシルへの選択的変換ができるようにするが、一方、メチル化されたシトシンは、変わらないままである。このステップに続いて、標準のデオキシヌクレオチド（dNTP）を用いた試料の増幅は、メチル化されていないシトシンのチミジンへの変換という結果をもたらし、一方、メチル化されたシトシンはシトシンになる。後続するシーケンシングは、元の試料におけるどのシトシンがメチル化されたか、またはメチル化されていなかったかを、推察することができるようにする。

【0228】

CODECがDNAメチル化情報を保持および報告できるようにするために、図16に表されているとおり、以下のプロトコルが開発されている。

【0229】

プロトコルには、以下のステップが関与する：

（a）CODECアダプター複合体を、通常のメチル化されていないシトシンに代えてメチル化されたシトシンを含有するように合成した。このようなやり方で、それらには後続する脱アミノ化が無効となり、および、記載されたプライマーで増幅させられることができる。

（b）改変されたCODECアダプターのライゲーションに続いて、反対DNA鎖のコピーを、標準のdATP、dGTPおよびdTTPヌクレオチドとともに用いられているメチル化されたdCTPを使用して生成した。このようなやり方で、元の鎖のコピーが常にシトシン位置でメチル化され、および、後続する脱アミノ化が無効となる。

（c）元の上のDNA鎖においてメチル化されていないシトシンをウラシルに変換するための脱アミノ化ステップを実施する。シトシンの脱アミノ化は、標準的な重亜硫酸脱アミノ化²；メチル化されたシトシンとメチル化されていないシトシンとを差別化するためにTET2およびAPOBEC2酵素による酵素的ステップを使用する酵素的メチル-seq（EM-seq）技法を使用した酵素的脱アミノ化³、などのいくつかのアプローチの1つによって行われることができる。あるいは、最近報告されたTET補助Pic-ボランシーケンシング、TAPS法⁴。

（d）脱アミノ化ステップに続いて、CODECアダプタープライマーを使用した増幅が適用される。

（e）二重鎖シーケンシングを実施する。

【0230】

元の鎖におけるメチル化／非メチル化情報を保ちながら、脱アミノ化に対して非感受性である元のDNA鎖のコピーを生成することによって、突然変異情報と同様にメチル化シーケンシング情報も、2つの鎖から得られるシーケンシング結果の比較により推察することが今や可能である。例えば、コピーされた鎖にCが存在し、かつ、Cが元の鎖にもまた存在するのであれば、この配列位置は元の試料中でメチル化されていたと推察できる。一方、元の配列にTがあれば、この配列位置はおそらく元の試料ではメチル化されていなかったものである。（シーケンシングエラーが原因でTが現れているという可能性を除外するためには、追加の分析をすることが必要であり得る：例えば、このTが元の鎖で現れているヌクレオチドコンテキストを観察することができる。追加のTもまた近くに現れていれば、Tはおそらくメチル化されていないCを表し；それが孤立したTであれば、Tがシーケンシングエラーの結果である確率が高い。）

【0231】

10

20

30

40

50

元のメチル化感受性のDNA鎖とともにCODECプロトコルにおいてメチル化非感受性の第2のDNA鎖コピーを作り出すことには、可能ないくつかの実用的な用途がある。

【0232】

例えば、全位置でシトシンを保存することによりコピーされたDNA鎖は「シトシン不足」ではないため、それはシーケンシングの間の明確なアライメントのために使用することができ、よって、シーケンスリードのマッピングを向上させることができるようにする。また、複数のメチル化されていない部位を持つDNA鎖がハイブリッド捕捉にとってはしばしば問題となるため、メチル化非感受性の鎖は、改善されたハイブリッド捕捉のために使用することができる。また、それは配列コールのブルーフリーディングに、および他の塩基への一般的な二重鎖シーケンシング訂正のために、使用することができる。最後に、それは、単一のDNA試料を（一方が突然変異のためおよびもう一方がメチル化分析のためである2つの別々の試料を使用することに代えて）使用した後続する組み合わせられた「メチル-突然変異」シーケンシングのための、突然変異およびメチル化の情報の両方を保存するライブラリーを作り出すために使用することができる。

10

【0233】

メチル化されたdCTPを使用して反対鎖を合成することおよび続いてメチル化されていないシトシンの脱アミノ化をすることは、以下のことなどの利点を有する：1）4つ全ての塩基が存在するため、明確なアラインメント、これは配列多様性を保存しおよび配列をアラインメントする能力を向上させる、2）しばしば課題であるメチル化されていない部位についてさえも、改善されたハイブリッド捕捉、3）メチル化感受性の部分への配列コールについての、および他の塩基への一般的な二重鎖シーケンシング訂正のための、改善されたブルーフリーディング、ならびに、4）単一のDNA試料（2つの別々の試料に代えて）を使用して、後続する組み合わせられたメチル-突然変異シーケンシングのためのライブラリーを作り出すこと。

20

例2についての参考文献

30

40

50

【表 5】

- [1] Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, Douville C, Javed AA, Wong F, Mattox A, Hruban RH, Wolfgang CL, Goggins MG, Dal Molin M, Wang TL, Roden R, Klein AP, Ptak J, Dobbys L, Schaefer J, Silliman N, Popoli M, Vogelstein JT, Browne JD, Schoen RE, Brand RE, Tie J, Gibbs P, Wong HL, Mansfield AS, Jen J, Hanash SM, Falconi M, Allen PJ, Zhou S, Bettegowda C, Diaz LA, Jr., Tomasetti C, Kinzler KW, Vogelstein B, Lennon AM, Papadopoulos N: Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018, 359:926-30. 10
- [2] Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL: A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 1992, 89:1827-31.
- [3] Vaisvila R, Ponnaluri VKC, Sun Z, Langhorst BW, Saleh L, Guan S, Dai N, Campbell MA, Sexton B, Marks K, Samaranayake M, Samuelson JC, Church HE, Tamanaha E, Corrêa IR, Pradhan S, Dimalanta ET, Evans TC, Williams L, Davis TB: EM-seq: Detection of DNA Methylation at Single Base Resolution from Picograms of DNA. *bioRxiv* 2019:2019.12.20.884692. 20
- [4] Liu Y, Siejka-Zielińska P, Velikova G, Bi Y, Yuan F, Tomkova M, Bai C, Chen L, Schuster-Böckler B, Song CX: Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat Biotechnol* 2019, 37:424-9.

【0234】

例 3 - Duplex-RepairとCODECシーケンシングとの組み合わせ 30

本開示の図1AFに図解されたとおり、CODECシーケンシングは、偽の突然変異の存在を最小限にすることを狙いとして、CODECシーケンシングに先立ちDuplex-Repairと組み合わせられてもよいということが企図されている。Duplex-Repairは、当技術分野で知られている末端修復／dAテーリング（ER/AT）方法の代わりに使用されてもよい。この例は、Duplex-Repairを記載する。

【0235】

本開示はまた、鎖再合成（およびよって、NGSアダプターライゲーションに先立ち、塩基損傷エラーを両方の鎖にコピーする潜在的可能性）を最小限にするための「末端修復／dAテーリング」（ER/AT）のための新しいアプローチにも関する。この技術のための前提は、市販で利用可能なER/AT方法を使用すると大量の鎖再合成が起こり得るという観察から来ている（図17A～17D）。アッセイは、最初に、延長されたパルス間持続時間（IPD、図17A）に基づいてd6mATPおよびd4mCTP（標準のdATPおよびdCTPに置換された）の組み込みを検出するための単分子リアルタイムシーケンシングを使用して鎖再合成を測定するために開発された。それは、ニック、ギャップ、およびオーバーハングを抱えた合成オリゴに市販のER/ATを適用したときに広範な鎖再合成が起こるという仮説を検証するのに使用された。鎖内深くにニックまたはギャップがあるときには、全ての下流の塩基が再合成されるということが示されている（図17B）。（T4ポリメラーゼは、鎖を再合成するために前進する前に数塩基「後退する」ということ、および、それは、おそらくDNA呼吸に起因して、各オリゴのbottom鎖をその平滑末端で部分的に再合成することができるということもまた示されている。これは、再合成はまた、主鎖損傷に 50

関係なくとも幾分起こり、および、さらに特定の反応条件に依存し得るということを提案する。)このアッセイを、次いで健常ドナーcfDNAに適用したところ、ほとんどの再合成がフラグメント末端で起こりながら(図17C)、多数の二重鎖がほぼ完全に再合成された(図17D)ということが見出された。これは、明らかに二重鎖シーケンシングに与っての主な課題である;それでもなお、単分子シーケンシングアッセイは、この問題を特徴付けおよび解決するのに必要な分解能を提供できる。

【0236】

Duplex-Repairと呼ばれているこの新しい方法は、アダプターライゲーションに先立ち鎖再合成を制限するためにER/ATを注意深くかつ段階的な様式で行う。Duplex-Repairは、4つの主なステップからなる:(1)損傷した塩基の切除およびオーバーハング除去、(2)平滑化および制限されたフィルイン、(3)ニック封止、および(4)dAテーリング(図18A)。DNAは、最初に、エンドヌクレアーゼIV、ホルムアミドピリミジン[fapy]-DNAグリコシラーゼ、ウラシル-DNAグリコシラーゼおよびT4ピリミジンDNAグリコシラーゼおよびエンドヌクレアーゼVIIIを含んでなる酵素カクテルで処理され、これは、ウラシル、8-オキソG、酸化ピリミジン、シクロブタンピリミジン二量体および脱塩基部位などの損傷した塩基を認識して切除し、1ntのギャップ(二本鎖セグメント内である場合)または鎖破断(一本鎖領域内である場合)のいずれかを生じる。エキソヌクレアーゼVII(ExoVII)もまた存在し、および、3'および5'一本鎖オーバーハングを分解する。第2のステップにおいて、T4ポリヌクレオチドキナーゼは、DNA末端を(脱)リン酸化し、および、T4 DNAポリメラーゼ(これは3'エキソヌクレアーゼを有するが、5'エキソヌクレアーゼまたは鎖置換活性を有しない)は、3'オーバーハングを平滑化し、および、ギャップと、ExoVIIに取り残された短い(<7nt)残りの5'一本鎖オーバーハングとを埋める。次いで、HiFi Taq DNAリガーゼによりニックが封止される。最後に、Klenowフラグメント(exo-)およびTaq DNAポリメラーゼを使用して、しかし鎖再合成を防ぐためにdATPのみの存在下で、dAテーリングが行われる。Duplex-Repairの性能は、複数の合成オリゴヌクレオチドを使用して検証されており、それは現実のDNA試料における予想された主鎖損傷のよくあるタイプを反映していた(図18A)。各二重鎖について、topおよびbottom鎖を、それらの5'および3'末端で夫々、はっきり区別できる色素を用いて標識化し、および、キャピラリー電気泳動を使用したことで、様々な処理条件下で各々に追加されたかまたはこれから除去された塩基を測定した。(i)5'オーバーハング、(ii)3'オーバーハング、(iii)ニック、(iv~v)塩基損傷なしの様々な長さのギャップ、および(vi~vii)塩基損傷を伴うギャップ、を抱えた二重鎖オリゴヌクレオチドを評価した。3'オーバーハングを除いた全てのケースについて、市販のER/AT(Kapa HyperPrepキット)を用いるとかなりの鎖再合成が起こるということが示されたが、他方、Duplex-Repairは、再合成された塩基の数を有意に低減させた。Duplex-Repairは、二重鎖シーケンシングの正確性への、様々な塩基および主鎖損傷の影響を制限するということがさらに確認されている。大量のcfDNAを最初に1名の健常ドナーから収集し、および、その複数のアリコートを経験的な量のCuCl₂/H₂O₂(酸化的損傷を誘導するため)およびDNase I(ニックを作り出すため)で処理した。市販のER/ATキット(Kapa HyperPrepキット)を次いで適用し、そして、最も重く損傷した状態ではエラーが1桁増大することを示した(図18B)。(注目すべきことに、DNase I濃度に伴うエラー率の増大は、市販のER/ATキットを使用した液体生検試験について、ゲノム解析の信頼性は、患者の血流中において、諸々の細胞外ヌクレアーゼの中でも、DNase Iのレベルに幾分依存し得るということを提案する。)次いで、Duplex-Repairを、最も重く損傷された条件に対して適用したところ、それが塩基および主鎖損傷の影響を「レスキュー」し、および、市販のER/ATを使用して調製された損傷されていないcfDNA試料よりもなおさらに低いエラー率を提供したということが見出された(図18B)。類似した結果が、ホルマリン固定腫瘍生検について観察された(図18C)。塩基および主鎖損傷は、自然発生的に(例として、シトシン脱アミノ化)、および、環境および化学的曝露(例として、UV照射、活性酸素種、ホルマリン固定

、凍結・解凍、加熱、音響せん断、等）への応答で生じる可能性があることを考慮すると、Duplex-Repairは、広範な試料について二重鎖シーケンシングの信頼性を確保する必要があった。

【0237】

本開示の一側面は、最少の鎖再合成および最大のライブラリー変換効率（すなわち、アダプターでライゲーションされたライブラリー分子へと変換されたDNA二重鎖の割合）で、二重鎖DNAにおける主鎖損傷を訂正するために、Duplex-Repairを最適化させることに関する。Duplex-Repairは、鎖再合成を最小限にし、およびER/ATにおいて病変乗り越え合成に対して保護することが示されているが、しかし、現在のプロトコルには、複数回の緩衝液交換が関与し、これは図18B~18CにおけるDuplex-Repair試料についてより少ない合計の二重鎖を生じおよびより広いエラーバーを説明づける。ここでは、Duplex-Repairは可能な限り最少のステップへと練られ（例として、ステップ間の「クリーンアップ」をなくし）、および、複数の酵素が共に機能することができるよう、緩衝液組成および実験条件（例として、時間、温度、濃度、および代替の酵素）を最適化する。性能は、（i）単分子シーケンシングアッセイ（図17A）、（ii）合成オリゴヌクレオチド基質およびキャピラリー電気泳動（図18A）、および、ER/ATに続いてシーケンシングされた実際のDNA試料（図18B~18C）を使用して検証された。qPCRライブラリープライマーでのqPCRと、NGSとの両方が、種々の方法（例として、超音波処理、酵素的消化）で種々のインサートサイズ中央値（例として、50~250bp）にせん断して生殖系列配列が決定されている健常ドナーからの様々なインプット（例として、1~1000ng）のバッフィーコート由来ゲノムDNAを使用した変換効率の定量化のために使用される。ワークフローは、様々な程度の塩基および主鎖損傷に供されたDNA（図18B）およびホルマリン固定腫瘍生検（図18C）などの困難な試料に対して試験されている。

【0238】

Duplex-Repairは、試料中の塩基および主鎖の損傷の程度に関係なく、二重鎖シーケンシングにおいて一貫して高い正確性を提供する。これは、NGS結果が全ての臨床試料についてロバストであることを確実にするのに役立つ。Duplex-Repairは、ギャップをおよび事例としてExoVII処理の後に残る短いオーバーハングを埋めるためにいくつかの量のDNA重合を依然として必要とする。これは、in silicoでフラグメント末端を最大約8~12塩基までトリミングする必要性が依然としてあることを意味し、これはデータのアウトプットを低減させるが、誤発見に対する予防のためには必然である。各ポリメラーゼは、病変乗り越え合成について様々な傾向を有するが、一方、生じる可能性のある塩基損傷には多数のタイプがある。塩基損傷が二本鎖シーケンシングにおいてエラーを生成するには、ER/ATおよびライブラリー増幅の両方においてポリメラーゼによりコピーされることができなければならない。各ポリメラーゼの、よくある塩基損傷（例として、8-オキシグアニン、ウラシル、脱塩基部位等）を回避しおよび「誤った」塩基を挿入する傾向が、試験されるであろう。しかしながら、DNAには見込み得る多数の塩基損傷が生じる可能性があり、および、かかる病変を全て試験することは不可能である。さらに、各酵素は、100%効率的とはならず、ゆえに、DNA産物のいくつかの損失を被ることが予想される。合成オリゴおよびキャピラリー電気泳動を使用して、各ステップで最高の効率を提供する酵素および反応条件が同定される（図18A）。DNAフラグメント化のための将来的な戦略は、事例として、オーバーハング長さを限定することができれば、または、アダプターが二本鎖破断部でタグ付けまたはライゲーションを介して直接富化されれば、ER/ATの必要性を限定することができる。これらの方法は探索されるであろう。しかしながら、DNAにおいて自然に生じる主鎖損傷を訂正して二重鎖回収を最大にするためにER/ATの必要性があることが、依然として予想される。事例として、既にフラグメント化されたものであるcfDNAなどの多数の臨床標本があり、これにはER/ATが必ず必要になる。

【0239】

均等物および範囲

冠詞「a」、「an」、および「the」などは、逆の指示がない限り、または文脈から明らかでない限り、1またはそれ以上を意味し得る。グループの1つ以上のメンバー間に「または」を含む態様または説明は、逆の指示がない限り、または文脈から明らかでない限り、グループメンバーの1つ、複数、または全てが、所与の製品またはプロセスに存在する、使用される、またはその他で関連する場合に、満たされるとみなされる。本発明は、グループの正確に1つのメンバーが、所与の製品またはプロセスに存在する、使用される、またはその他で関連する態様を含む。本発明は、グループメンバーの1つより多くまたは全てが、所与の製品またはプロセスに存在する、使用される、またはその他で関連する態様を含む。

【0240】

さらに本開示は、列挙された請求項の1つ以上からの1つ以上の限定、要素、条項、および説明用語が別の請求項に導入される全ての変形、組み合わせ、および順列を包含する。例えば、別の請求項に従属するあらゆる請求項は、同じ基本請求項に従属するあらゆる他の請求項に見られる1つ以上の限定を含むように、改変することができる。要素がリストとして提示される場合、例えばマーカッシュグループ形式では、要素の各サブグループも開示され、あらゆる要素（単数または複数）をグループから削除できる。一般に、本発明または本発明の側面が特定の要素および／または特色を含むとされる場合、開示の特定の態様または開示の側面は、かかる要素および／または特色からなる、または本質的にそれらからなることを理解されたい。簡単にするために、これらの態様は、本明細書では直接的具体的に説明されてはいない。用語「含む (comprising)」および「含有する (containing)」は、オープンであることを意図しており、追加の要素またはステップを含めることを許容することにも留意されたい。範囲が指定されている場合は、エンドポイントも含まれる。さらに、別段の指示がない限り、または文脈および当業者の理解から明らかでない限り、範囲として表わされる値は、本発明の異なる態様において記載された範囲内のあらゆる特定の値または部分範囲を、文脈上明らかに別段の指示がない限り、範囲の下限の単位の10分の1までとることができる。

【0241】

本出願は、様々な発行された特許、公開された特許出願、学術文献、および他の刊行物を参照し、その全てが参照により本明細書に組み込まれる。組み込まれた参考文献のいずれかと本明細書との間に矛盾がある場合には、本明細書が優先するものとする。さらに、従来技術に該当する本発明の特定の態様は、あらゆる1つ以上の態様から明示的に除外することができる。かかる態様は当業者には既知であるとみなされるため、それらは、本明細書に明示的に除外が記載されていなくても、除外することができる。本発明のいかなる特定の態様も、先行技術の存在に関連するか否かにかかわらず、いかなる態様からいかなる理由によっても、除外することができる。

【0242】

当業者は、日常的な実験のみを使用して、本明細書に記載される特定の態様の多くの均等物を認識するか、確認することができるであろう。本明細書に記載される本態様の範囲は、上記の説明に限定されることを意図するものではなく、むしろ添付の態様に記載される通りである。当業者であれば、以下の態様で定義される本発明の精神または範囲から逸脱することなく、この説明に対する様々な変更および改変を行うことができることを、理解するであろう。

10

20

30

40

50

【図面】

【図 1 A】

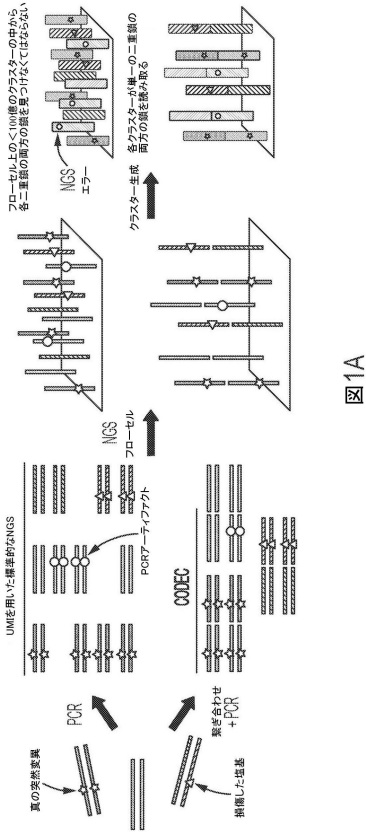


図 1A

【図 1 D】

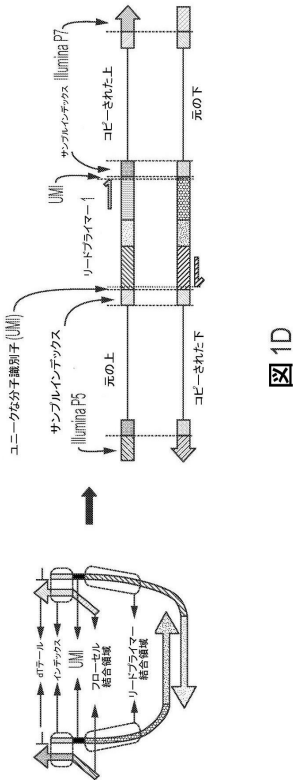


図 1D

【図 1 B - 1 C】

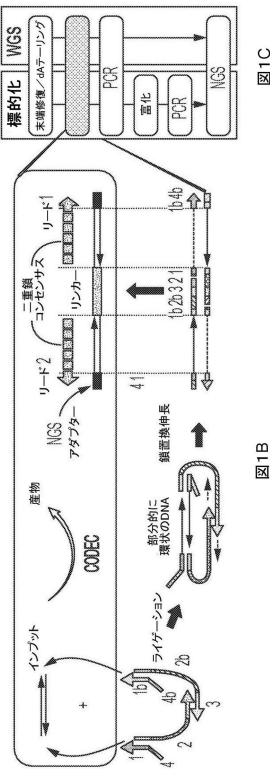


図 1B

図 1C

【図 1 E - 1 F】

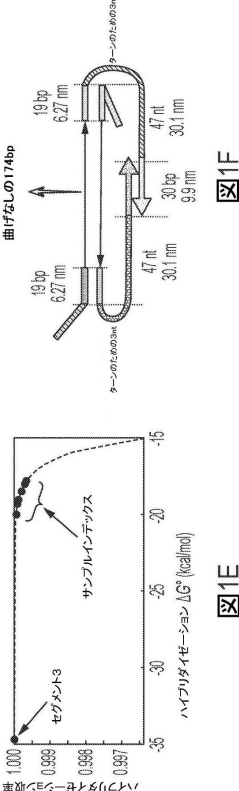
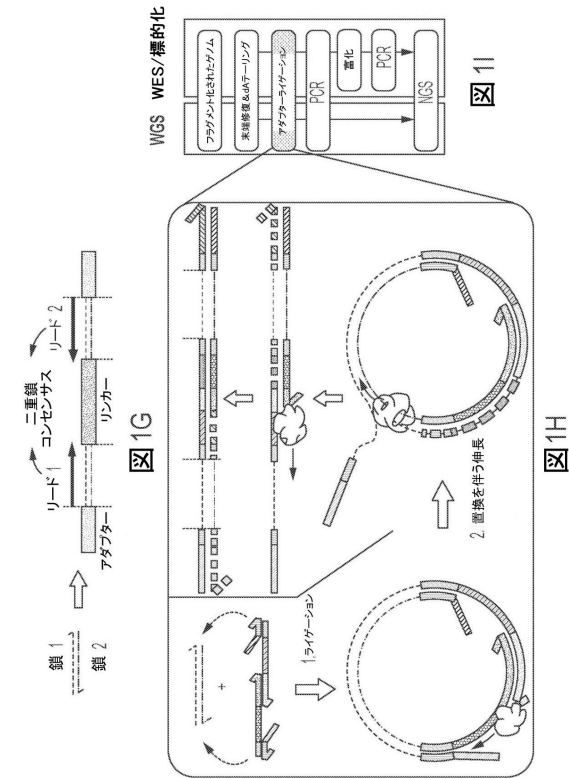


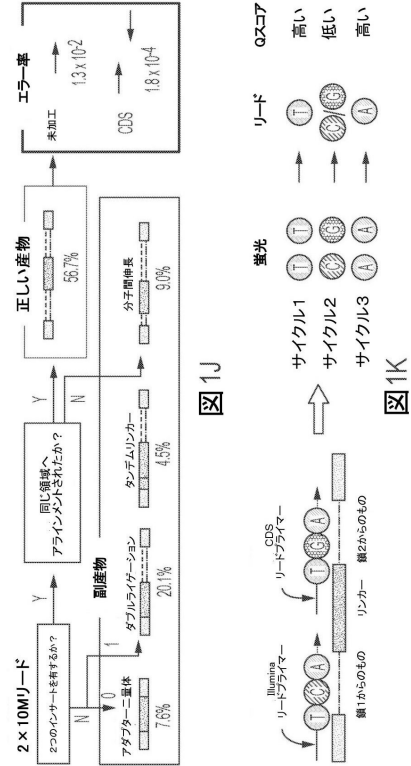
図 1E

図 1F

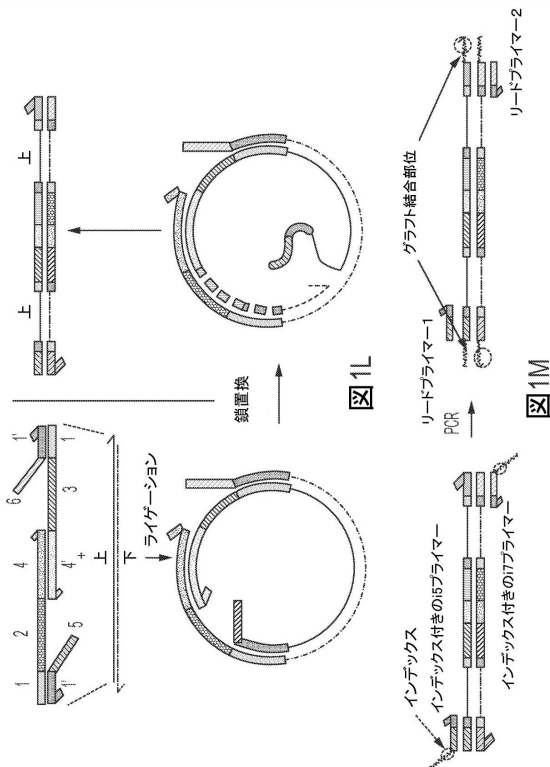
【図1G-1I】



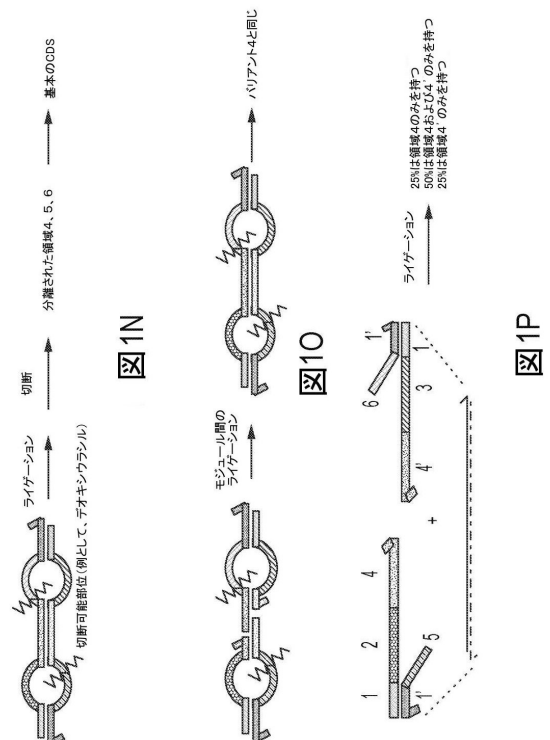
【図1J-1K】



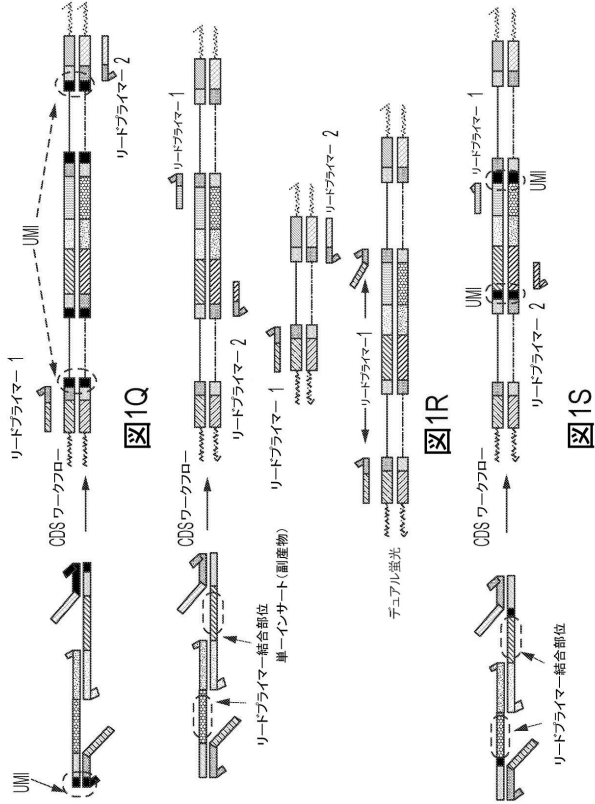
【図1L-1M】



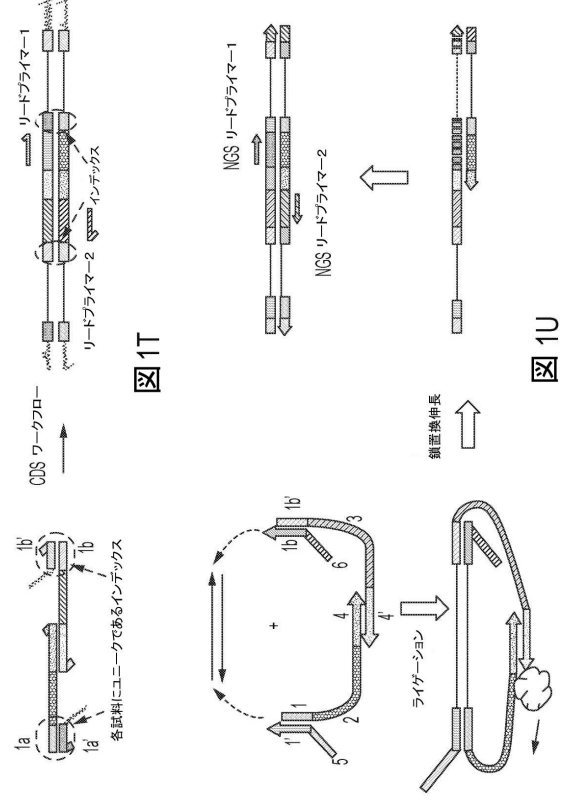
【図1N-1P】



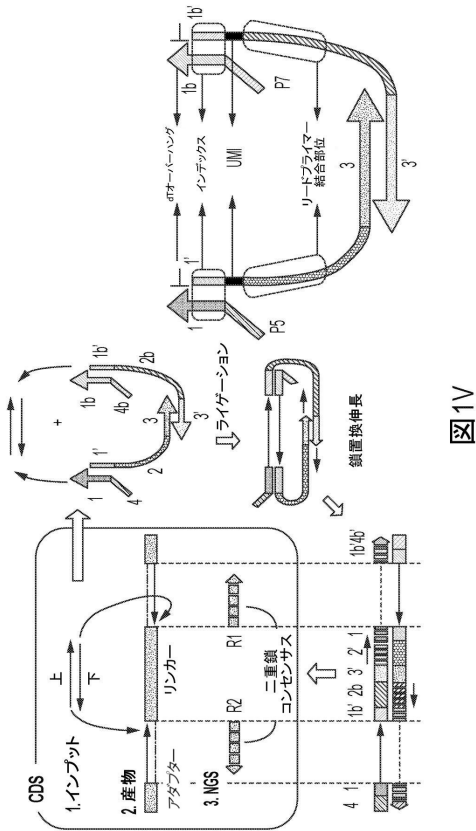
【図 1 Q - 1 S】



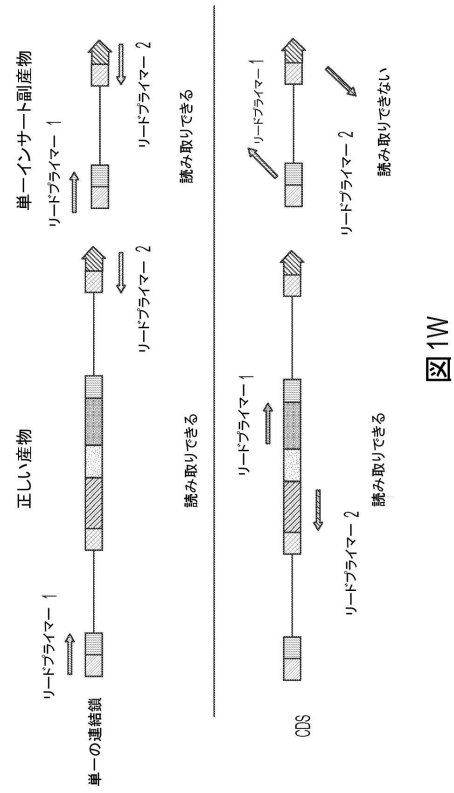
【図 1 T - 1 U】



【図 1 V】



【図 1 W】



10

20

30

40

50

【図 1 A B - 1 A C】

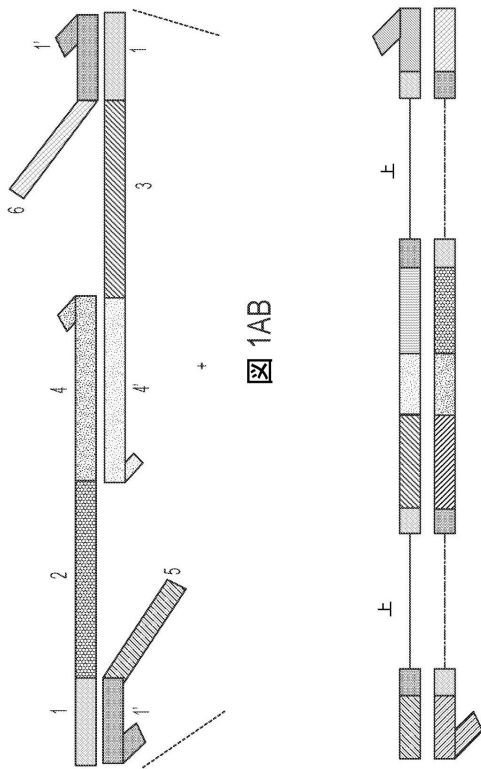


図 1AB

図 1AC

【図 1 A D】

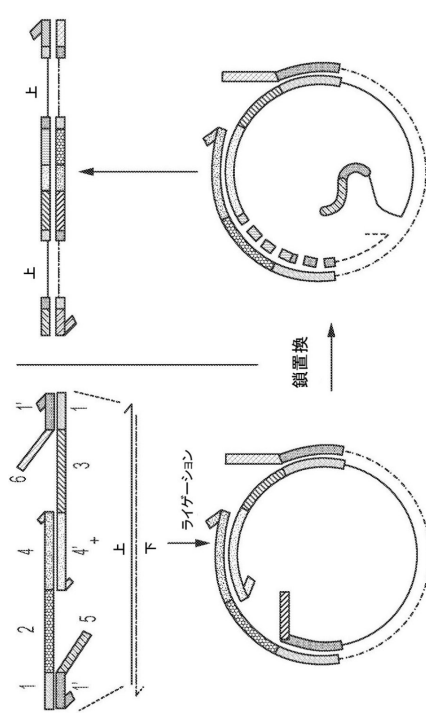


図 1AD

【図 1 A E 1】

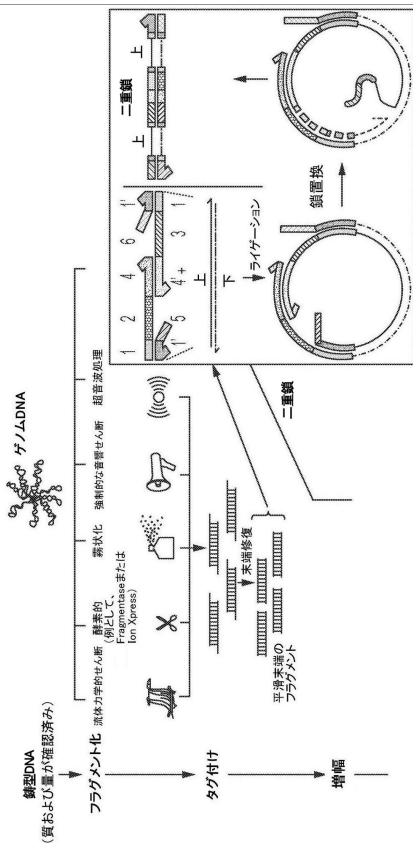


図 1AE

【図 1 A E 2】

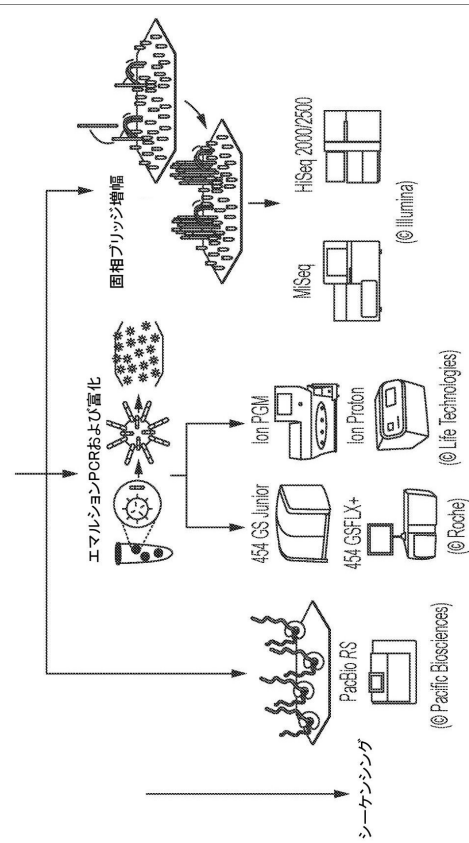


図 1AE 続き

10

20

30

40

50

【図 1 A F 1】

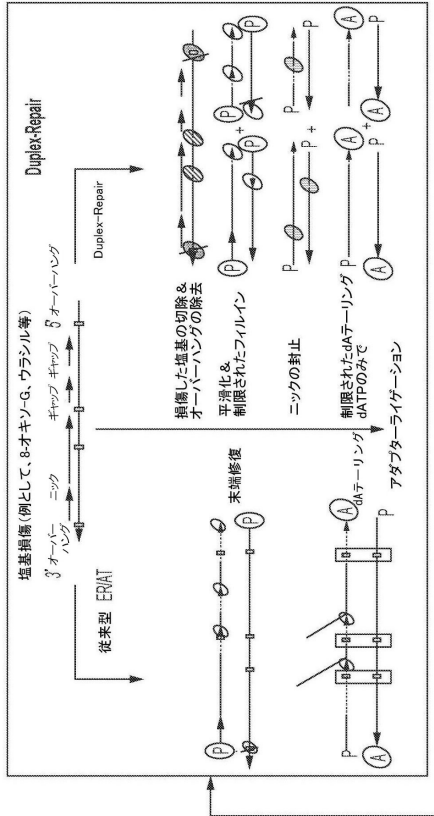


図1AF

【図 1 A F 2】

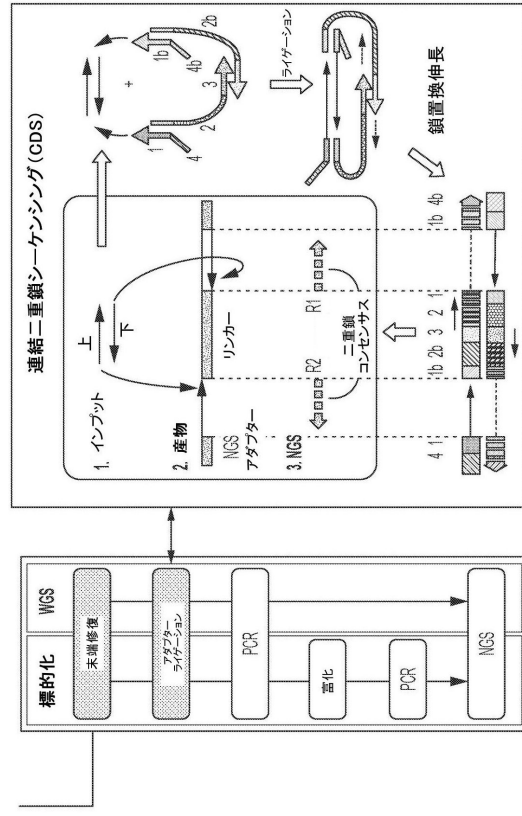


図1AF2
続き

【図 1 A G】

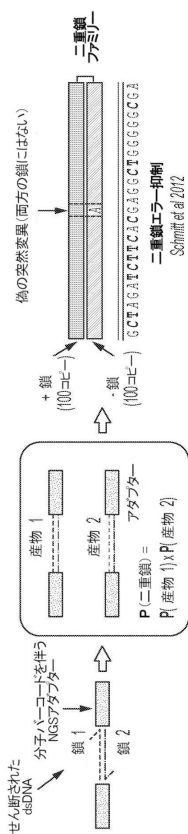


図1AG

【図 1 A H - 1 A J】

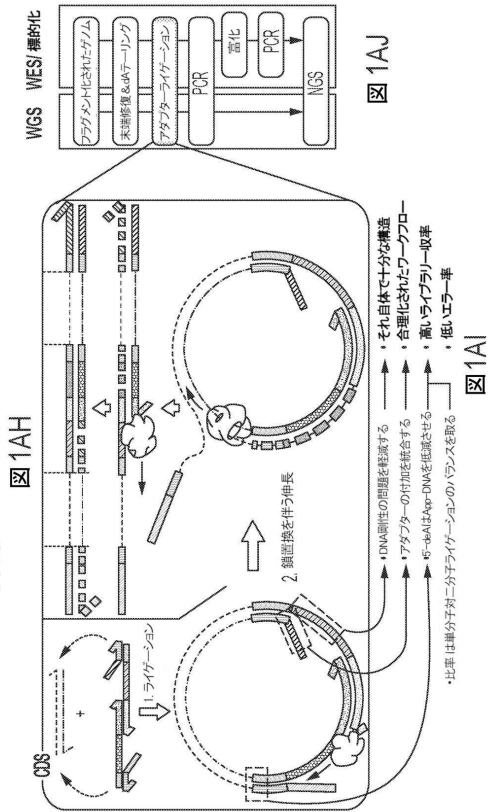


図 1AJ

図 1AI

【図 1AK-1AL】

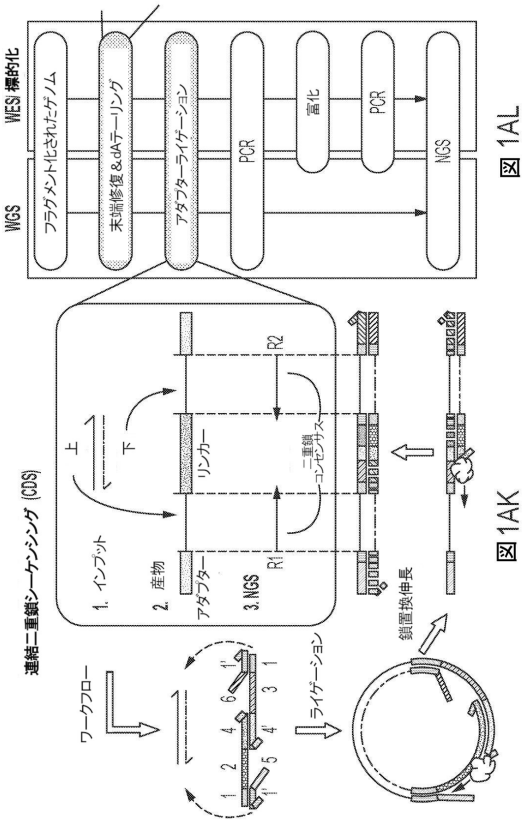


図 1AK

図 1AL

【図 1AM】

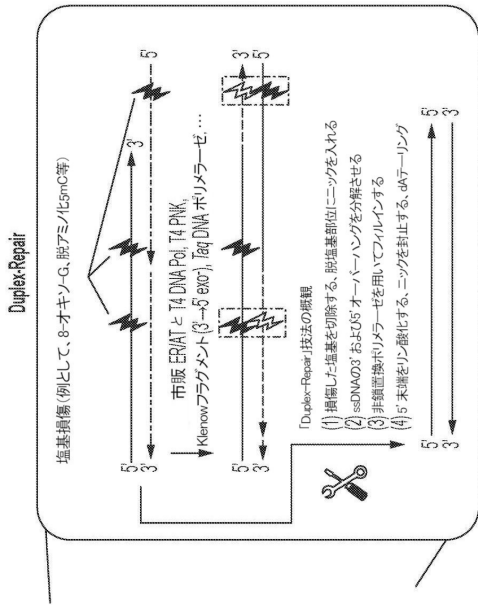


図 1AM

【図 1AN】

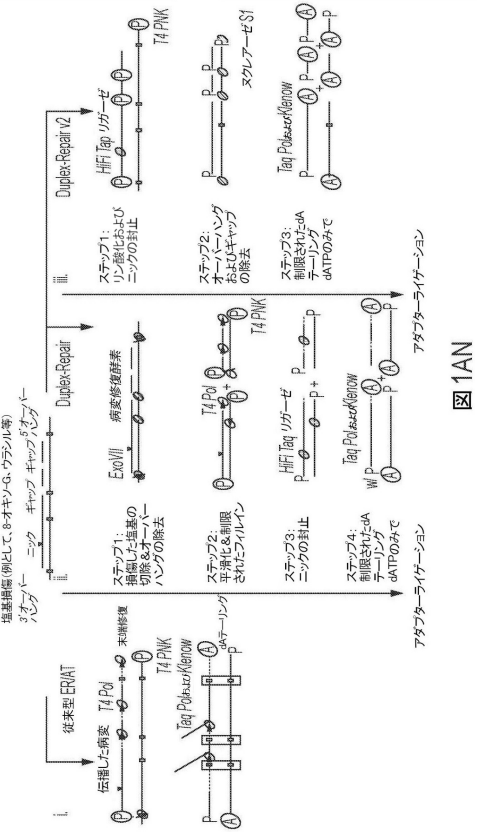


図 1AN

【図 1AO】

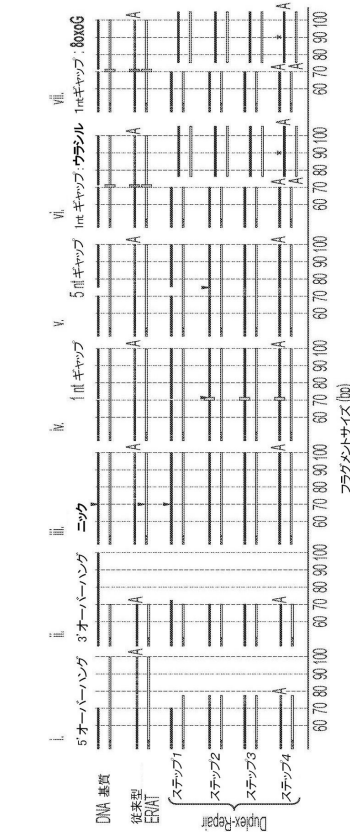


図 1AO

【図 1 A P】

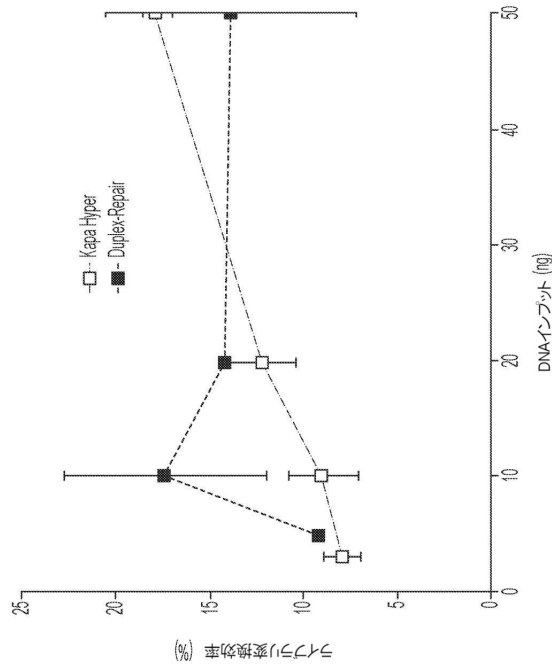


図 1AP

【図 1 A Q - 1 A R】

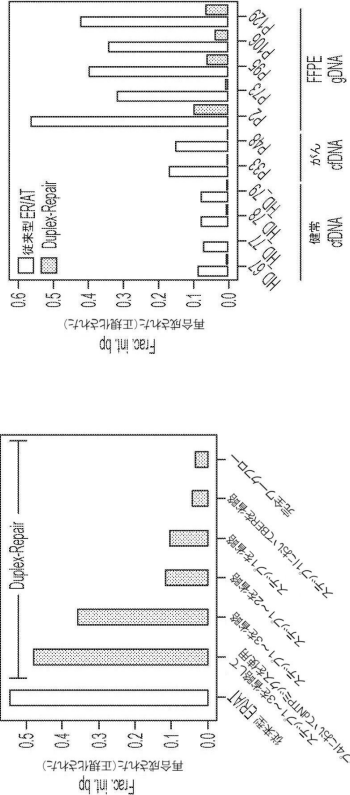


図 1AR

図 1AQ

【図 2】

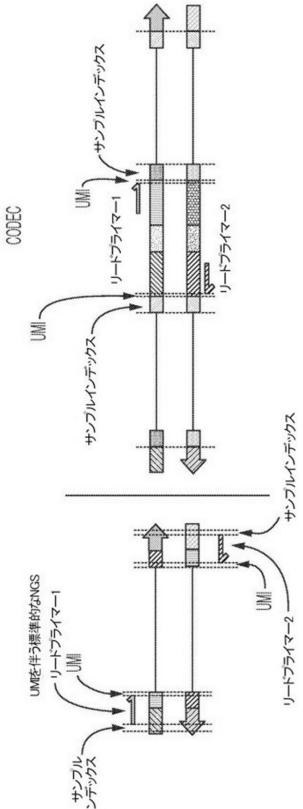


図 2

【図 3 A】

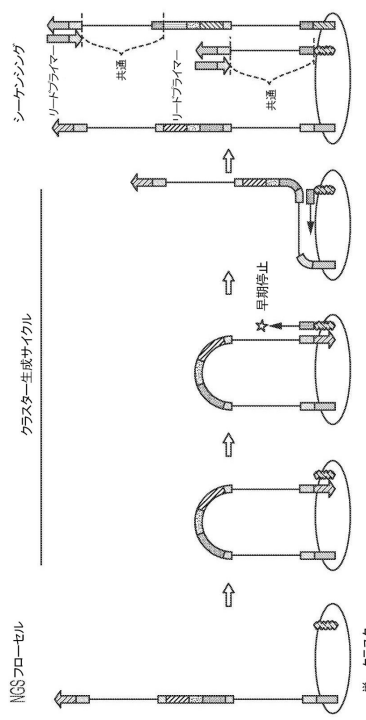


図 3A

10

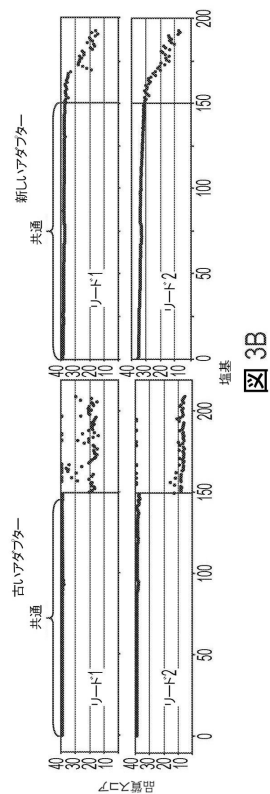
20

30

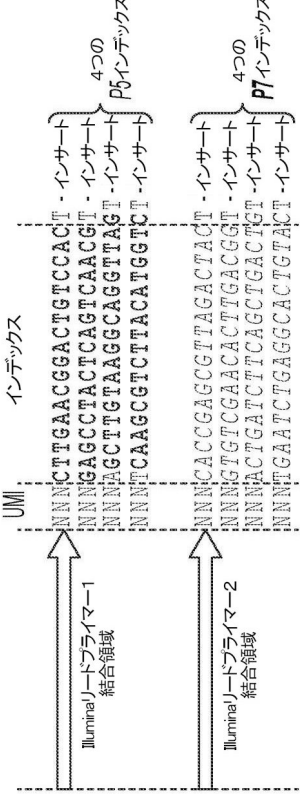
40

50

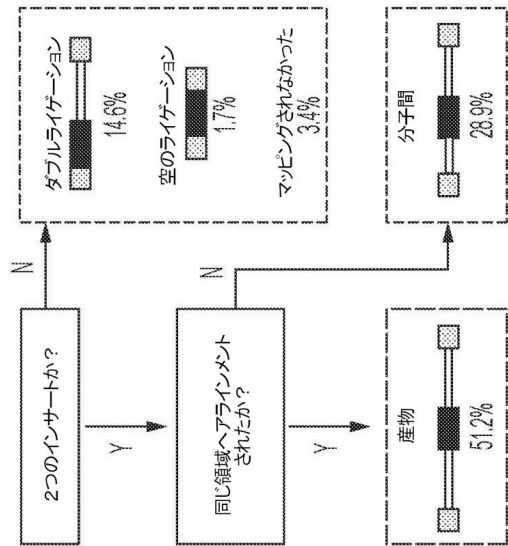
【図 3 B】



【図 4】



【図 5 A】



【図 5 B】

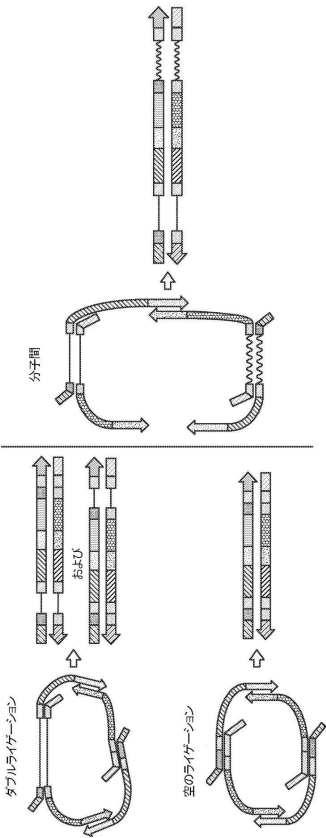


図 4

図 5B

図 5A

10

20

30

40

50

【図 6 A - 6 B】

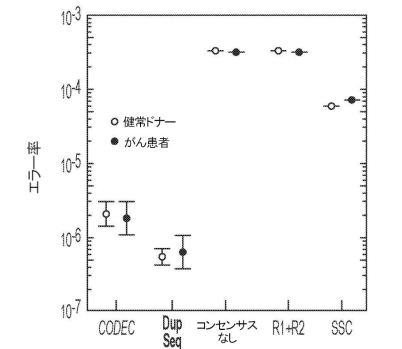


図 6A

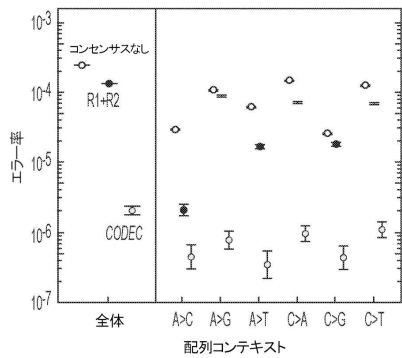


図 6B

【図 6 C - 6 D】

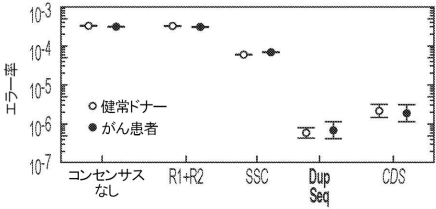


図 6C

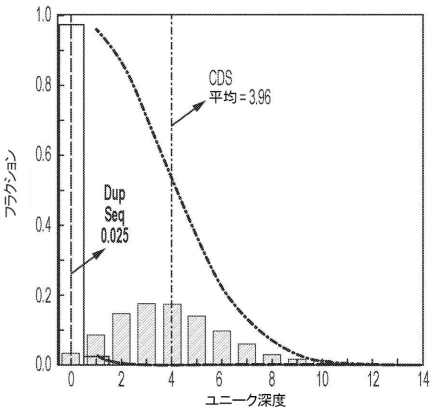


図 6D

【図 6 E - 6 F】

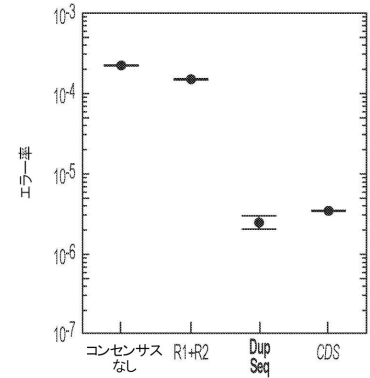


図 6E

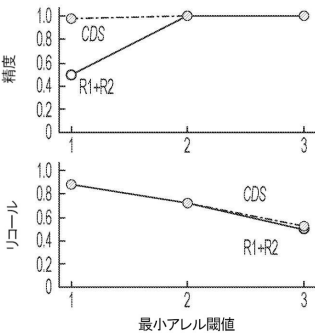


図 6F

【図 7】

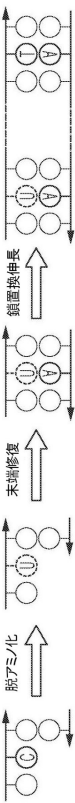


図 7

10

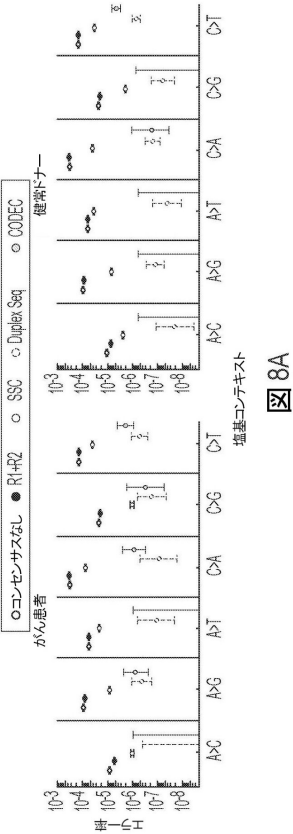
20

30

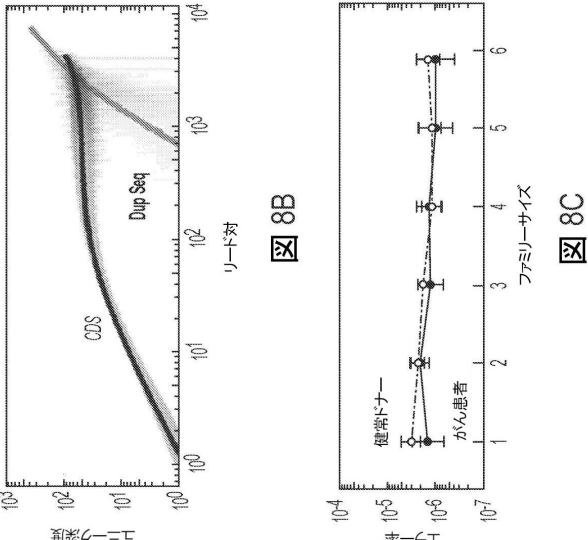
40

50

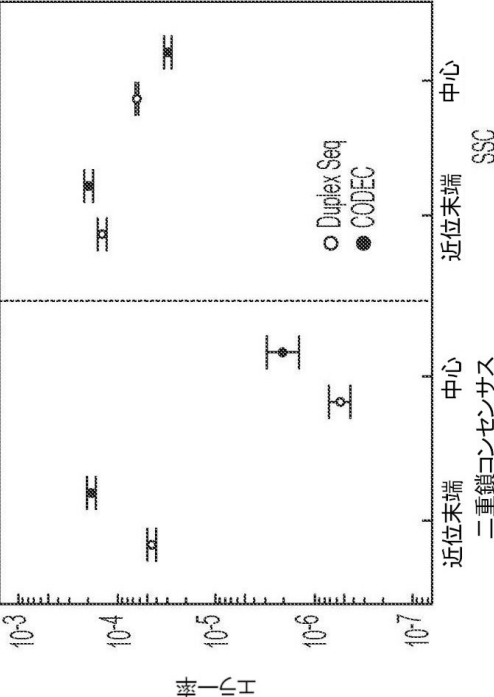
【図 8 A】



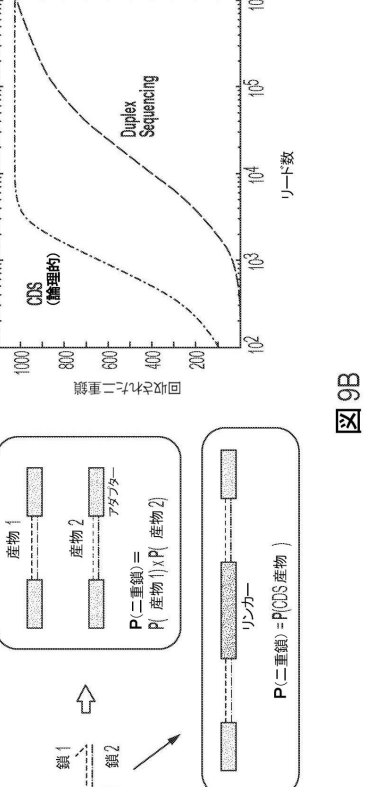
【図 8 B - 8 C】



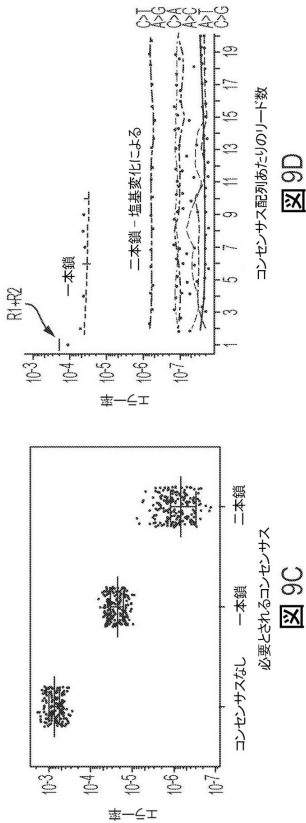
【図 9 A】



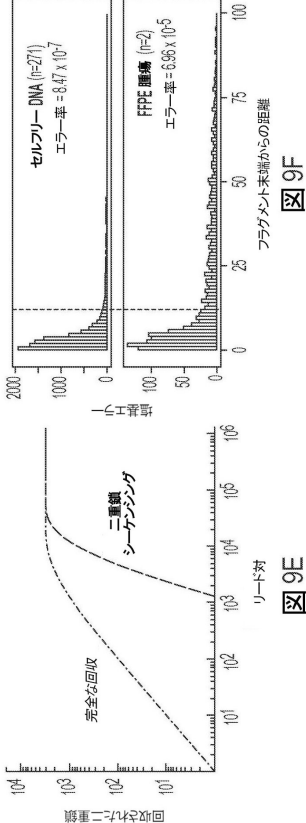
【図 9 B】



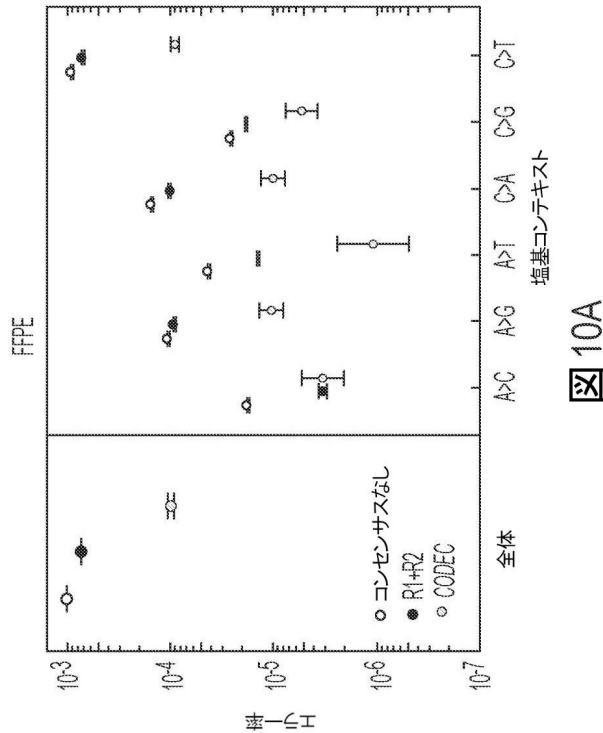
【図 9 C - 9 D】



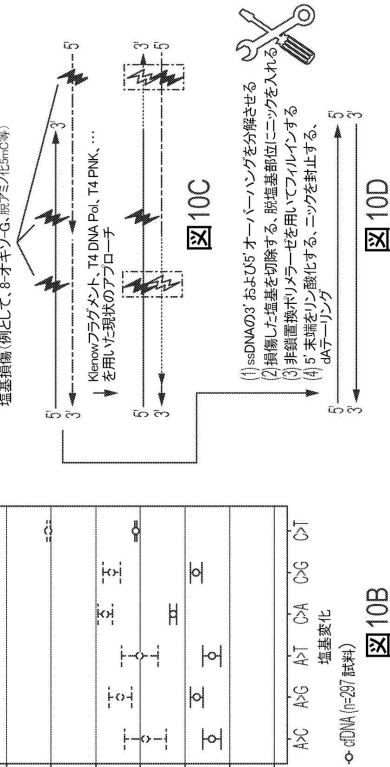
【図 9 E - 9 F】



【図 1 0 A】



【図 1 0 B - 1 0 D】



10

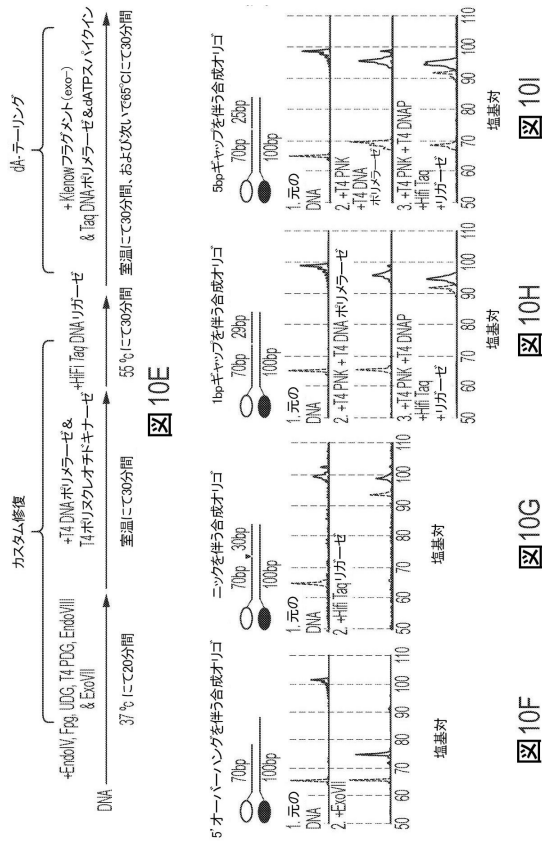
20

30

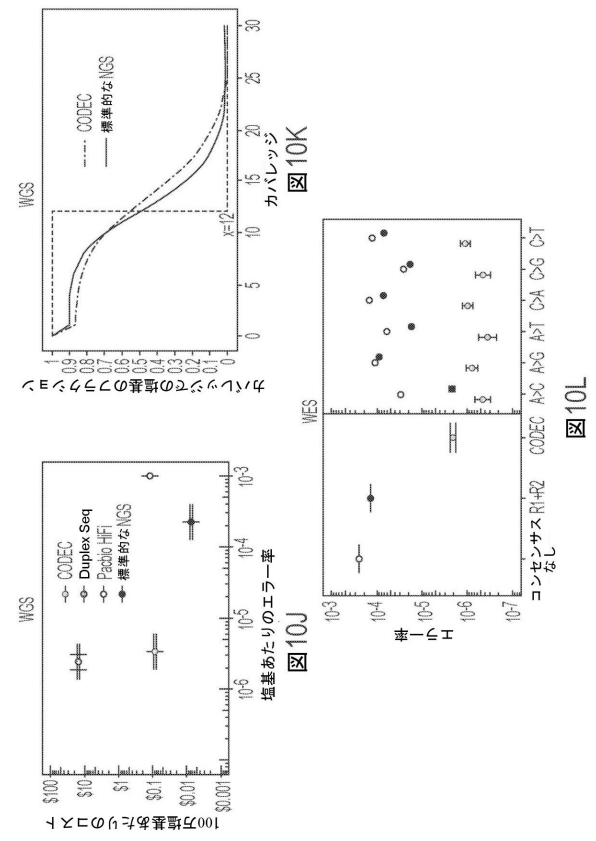
40

50

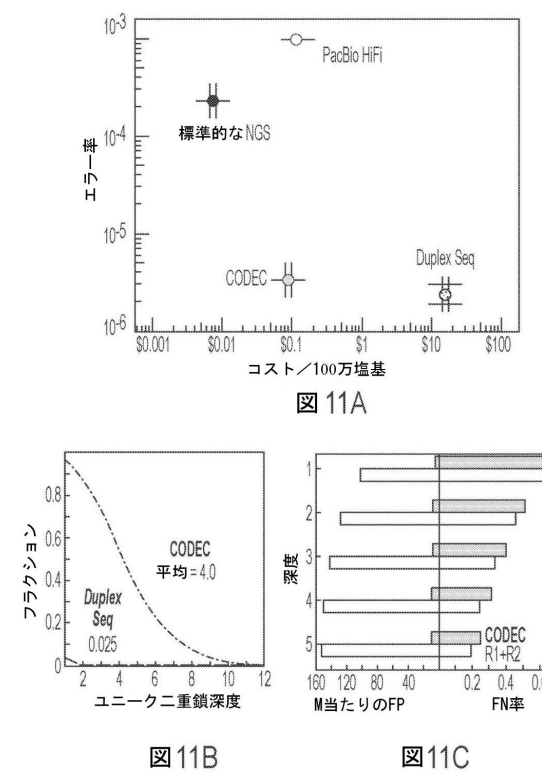
【図10E-10I】



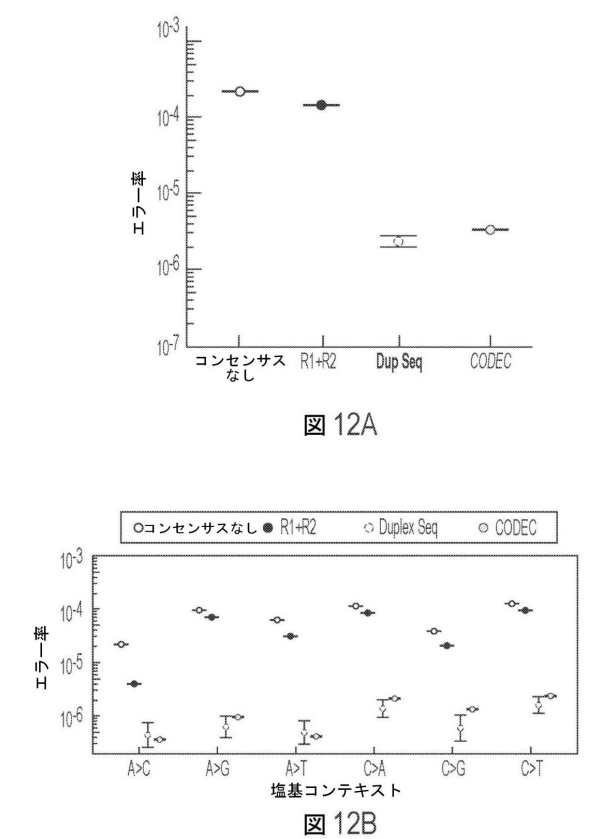
【図10J-10L】



【図11A-11C】



【図12A-12B】



10

20

30

40

50

【図 1 2 C - 1 2 E】

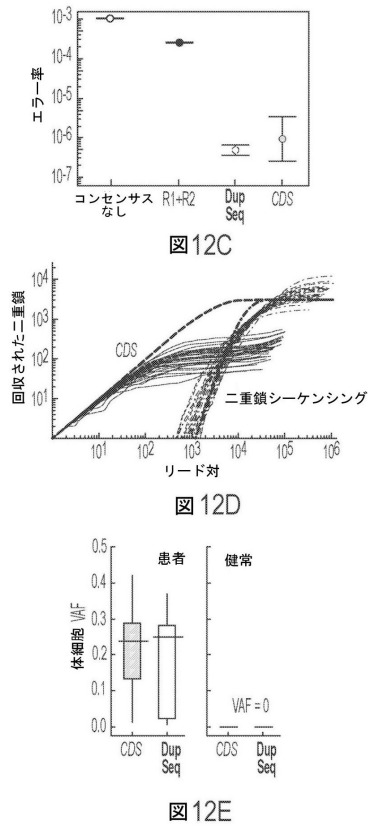


図 12C

図 12D

図 12E

【図 1 3 A - 1 3 B】

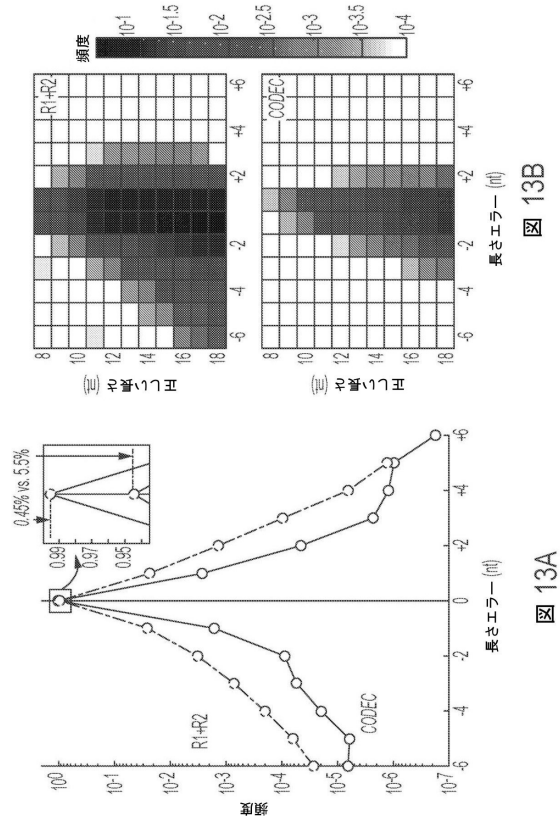


図 13B

図 13A

【図 1 3 C】

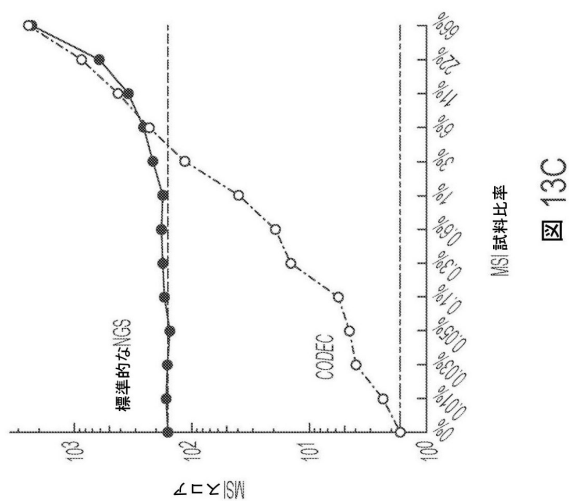


図 13C

【図 1 4 A】

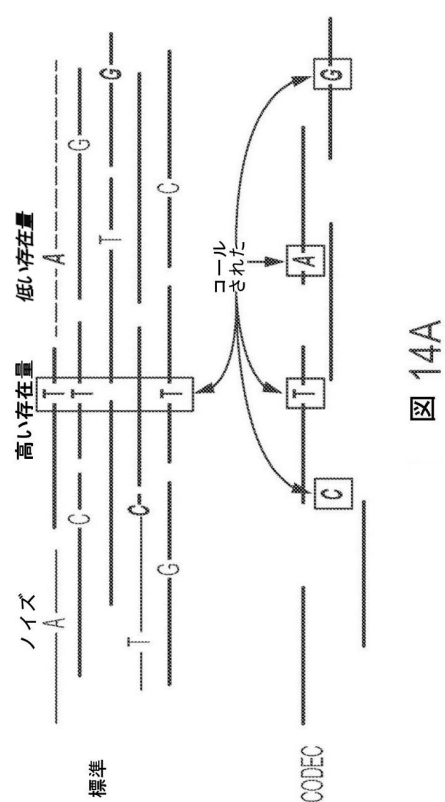


図 14A

10

20

30

40

50

【図 14B】

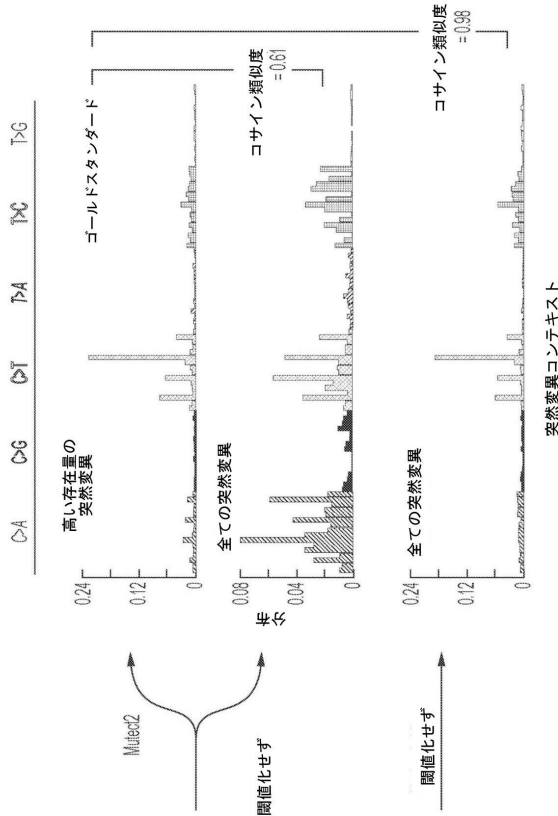


図 14B

【図 14C - 14D】

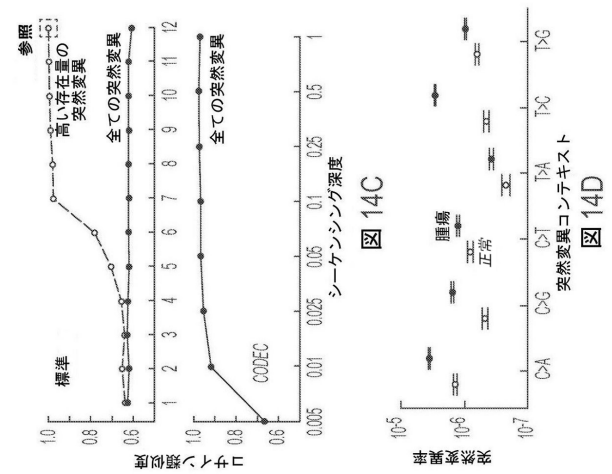


図 14C

図 14D

【図 14E】

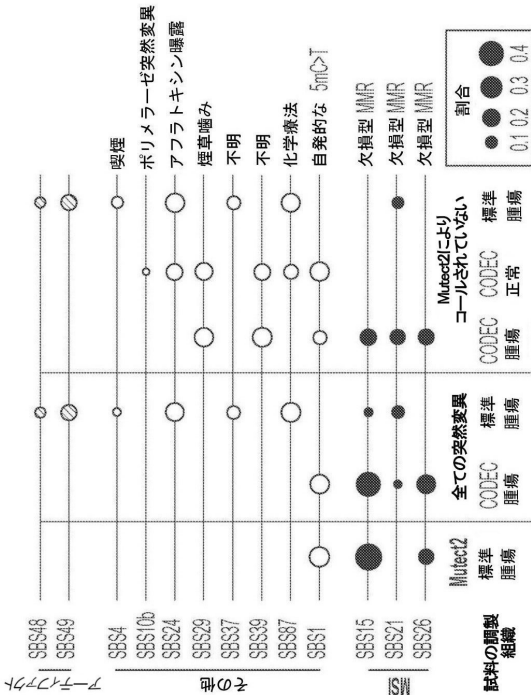


図 14E

【図 14F】

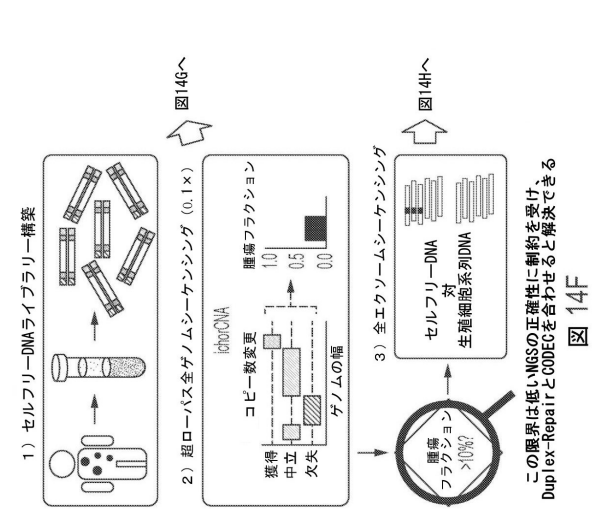


図 14F

【図 1 4 G】

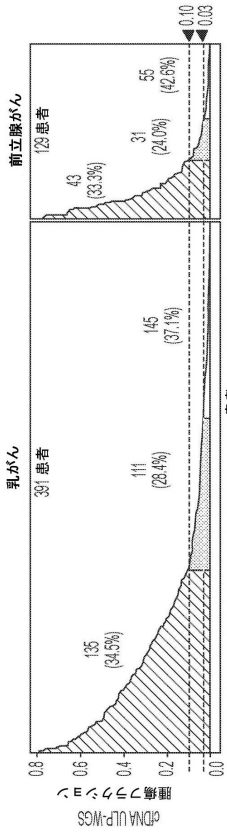


図 14G

【図 1 4 H 1】

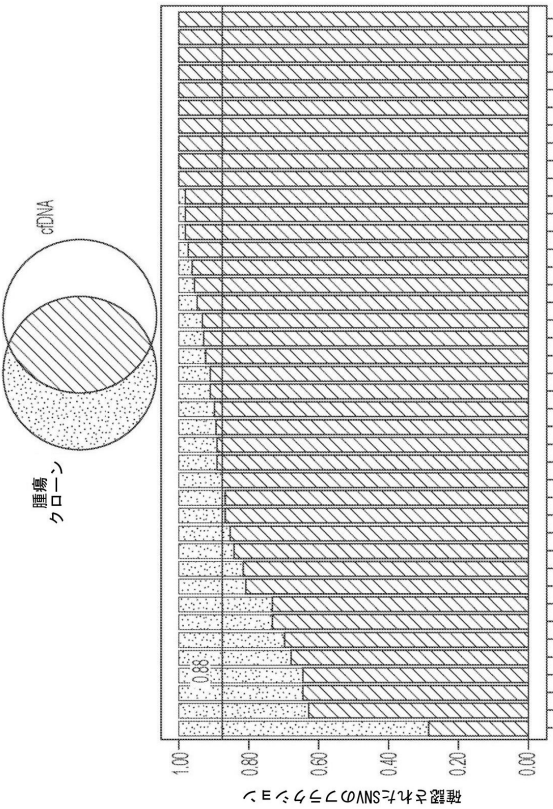


図 14H

【図 1 4 H 2】

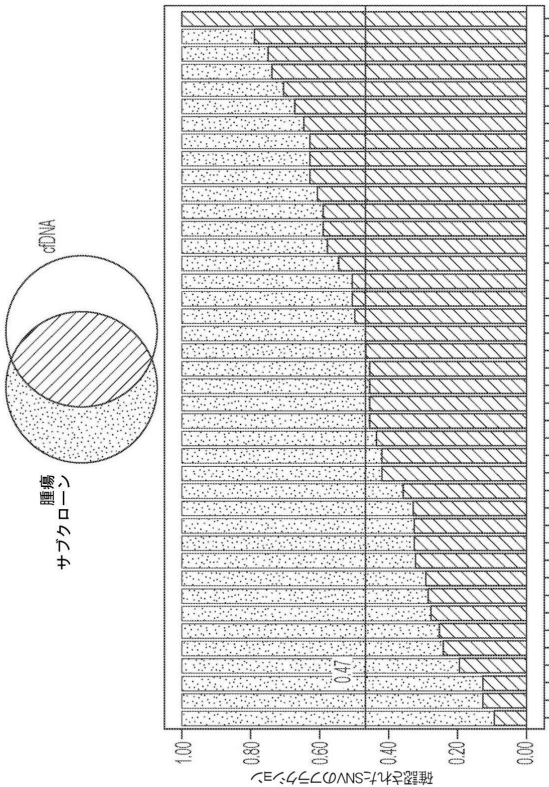


図 14H

続き

【図 1 4 I】

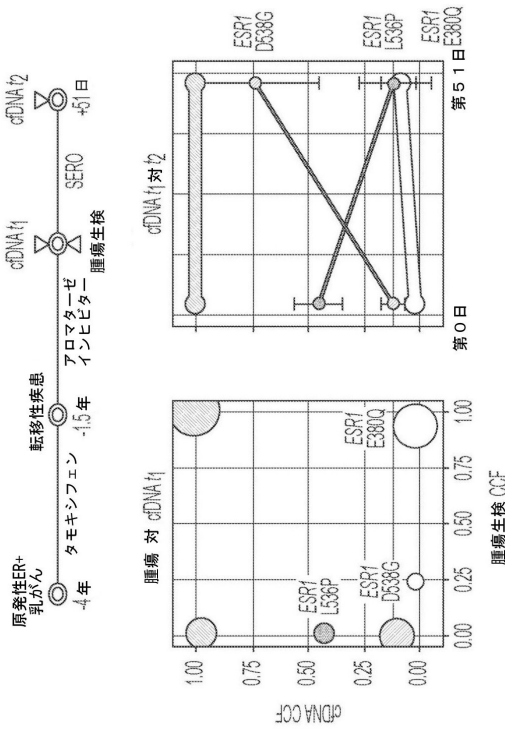


図 14I

【図 15】

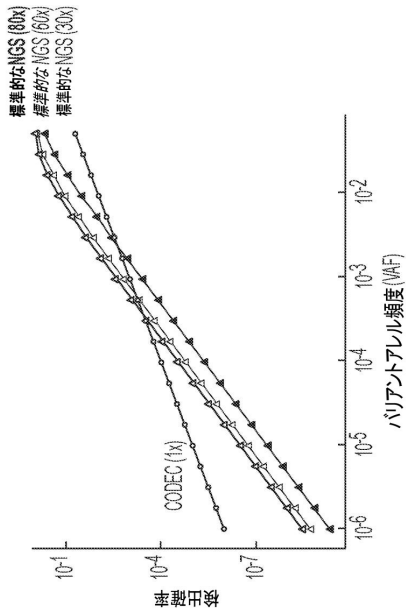


図15

【図 16】

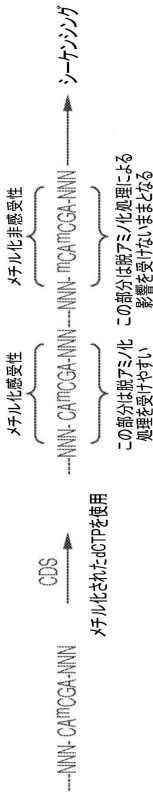


図16

【図 17 A】

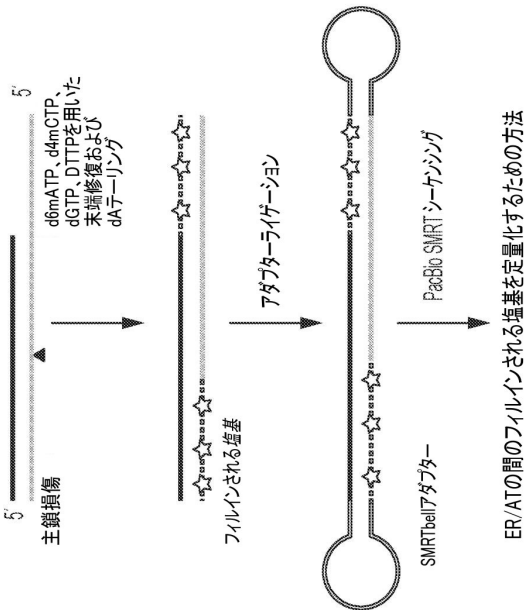


図17A

【図 17 B】

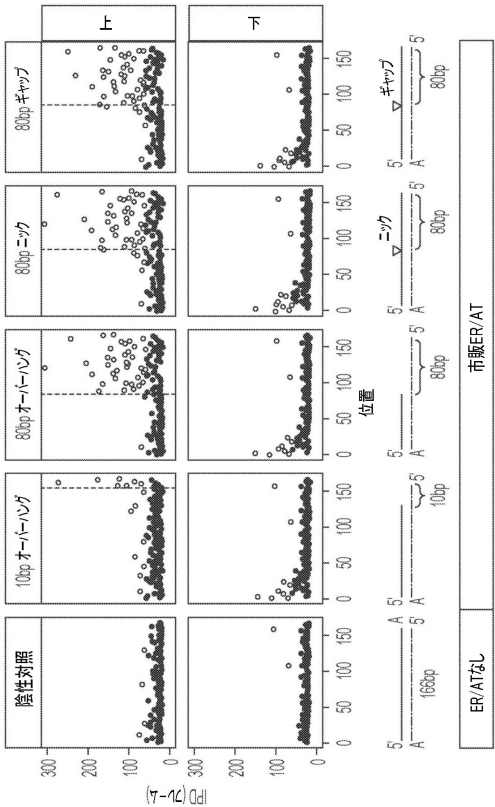


図17B

10

20

30

40

50

【図 17 C】

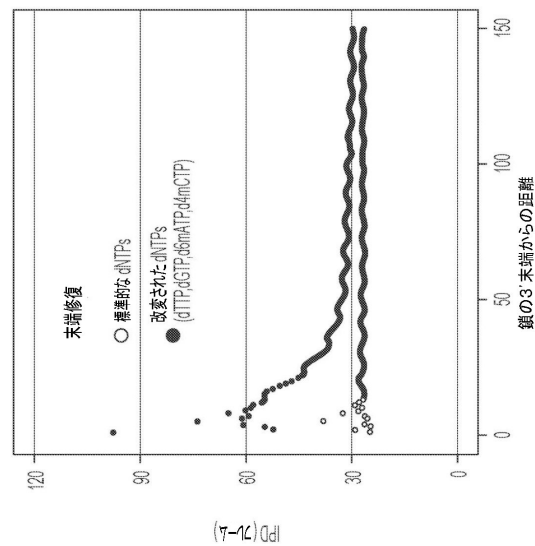


図 17C

【図 17 D】

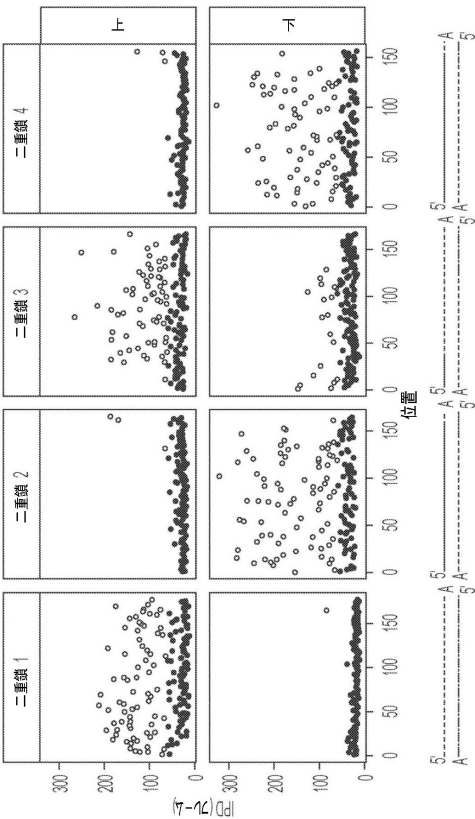


図 17D

【図 18 A】

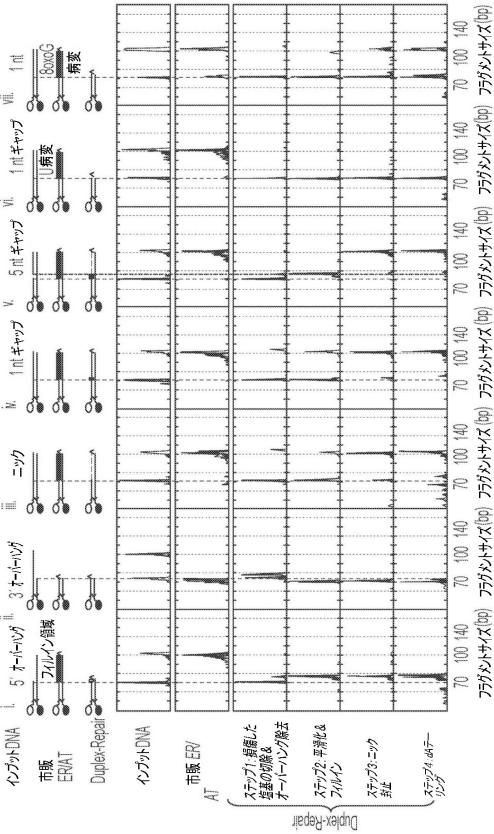


図 18A

【図 18 B - 18 C】

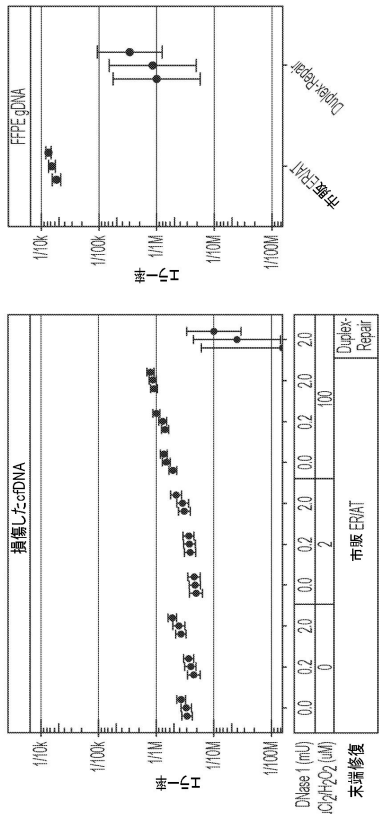


FIG. 18C

FIG. 18B

10

20

30

40

50

【図 18 D - 18 E】

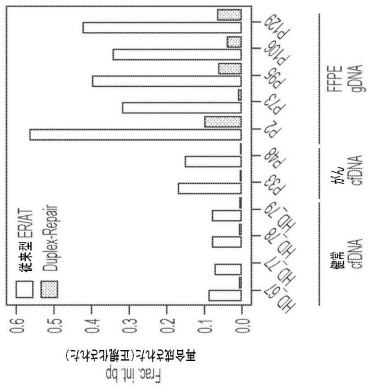


図 18E

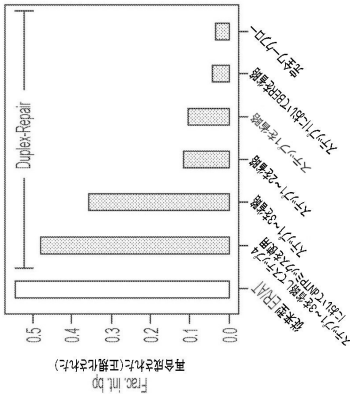


図 18D

【図 18 F - 18 G】

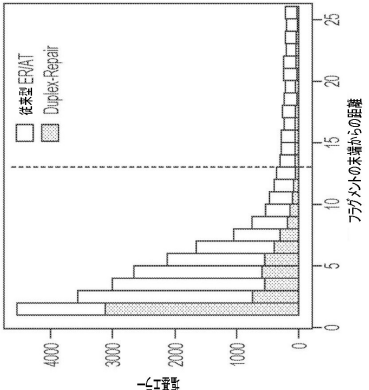


図 18F

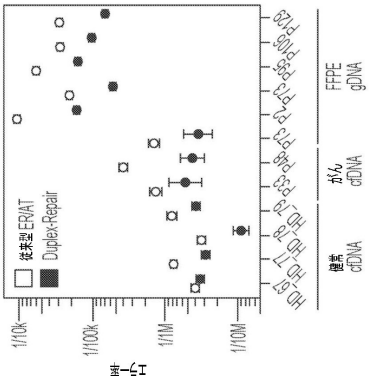


図 18G

【図 18 H - 18 I】

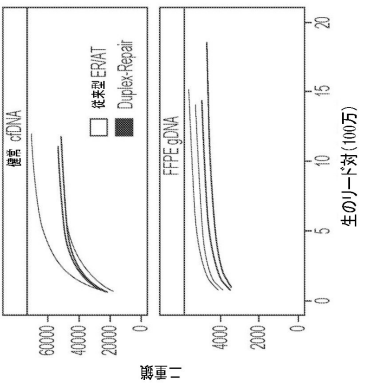


図 18I

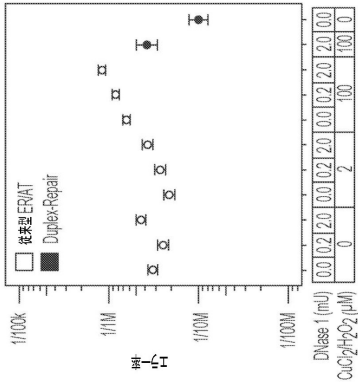


図 18H

【図 19 A】

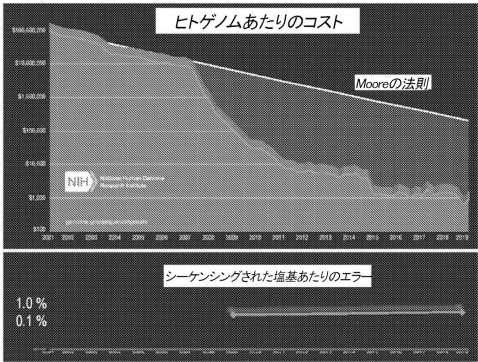


図 19A

10

20

30

40

50

【図 19 B】



図 19B

【図 20】

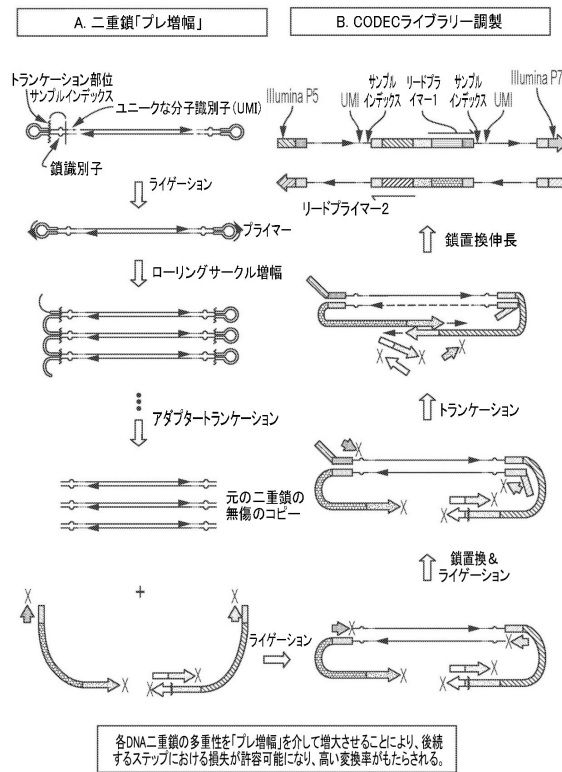


図 20

【図 21】

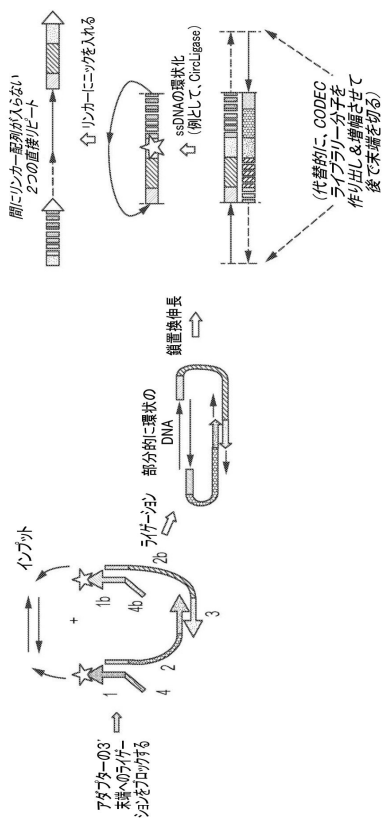


図 21

10

20

30

40

50

【配列表】

2023553983000001.app

10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/62966

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12Q 1/68, C12Q 1/6869, G16B 30/20 (2022.01)

CPC - C12Q 1/6806, C12Q 1/6855, C12Q 1/6869, C12Q 2525/191, C12Q 2535/119, C12Q 2535/122

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2020/0115752 A1 (LIFE TECHNOLOGIES CORPORATION) 16 April 2020 (16.04.2020). para [0064], [0069]-[0070], [0081], [0087], [0097], [0111], [0118], [0122], [0180], [0191], [0195], [0198], [0202], [0228], [0244], [0305]-[0306], Claim 1, Claim 4, Figure 1A, Figure 1B, Table 12	1-4, 25-36, 59-65
Y	US 2019/0241953 A1 (ROCHE SEQUENCING SOLUTIONS, INC.) 08 August 2019 (08.08.2019). para [0026], [0042], [0047]-[0048], [0061], [0083], [0065], [0068], [0070], Claim 1, Claim 12, Figure 1, Figure 2, Figure 8, Abstract	1-4, 25-36, 59-65
Y	US 2018/0121599 A1 (VITO NV (VLAAMSE INSTELLING VOOR TECHNOLOGISCH ONDERZOEK)) 03 May 2018 (03.05.2018). para [0007], Claim 4, Abstract	27, 29, 31

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

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

23 March 2022

Date of mailing of the international search report

APR 15 2022

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Kari Rodriguez

Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/62968

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 5-24, 37-46, 66-82, 84-91
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I: claims 1-4, 25-36, 59-65, drawn to an isolated nucleic acid complex.

Group II: claims 47-58, 83, 92, drawn to a method of preparing a sequencing library.

-----Please see continuation in first extra sheet-----

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-4, 25-36, 59-65

Remark on Protest

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/62966

Continuation of Box No. III. Observations where unity of invention is lacking

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features

Group I includes the special technical feature of a composition which differs from the special technical feature of a method, as disclosed by Group II.

Common Technical Features

The inventions of Groups I and II share the technical feature of sequencing adapter having a first end, a second end and a central portion positioned between the first and second ends.

However, these shared technical features do not represent a contribution over prior art in view of US 2020/0115752 A1 to Life Technologies Corporation (hereinafter "LTC") and US 2019/0241953 A1 to Roche Sequencing Solutions, Inc. (hereinafter "Roche").

LTC teaches (instant claim 25) an adapter having a first end, a second end and a central portion positioned between the first and second ends, wherein the first end comprises a first duplex comprising a first oligonucleotide annealed to a second oligonucleotide, wherein the second end comprises a second duplex comprising a third oligonucleotide annealed to a fourth oligonucleotide, and wherein the second and the fourth oligonucleotides are annealed to one another over a region complementarity to form a third duplex that is positioned in the central portion (abstract). Blocking oligonucleotide adaptors can be used to reduce the formation of adaptor dimers or trimers (or higher-order concatamers) which can improve the yield of desirable polynucleotide-adaptor products in any recombinant nucleic acid workflow. Blocking oligonucleotide adaptors can comprise a double-stranded oligonucleotide adaptor (duplex) having an overhang cohesive portion that anneals with a blocking oligonucleotide which can be a separate single-stranded oligonucleotide.; para [0048], FIG. 1A is a schematic showing a non-limiting example of a left and a right blocking oligonucleotide adaptor used to circularize a polynucleotide of interest.....(Bottom) A third single-stranded oligonucleotide (blocking oligonucleotide) can be removed and the overhang portions from the left and right blocking oligonucleotide adaptors anneal with each other to circularize the polynucleotide of interest. For the sake of clarity, the diagram shows only a portion of the polynucleotide of interest.), but does not specifically teach wherein the sequencing adapter further comprises a pair of read primer binding sites on either side of the third duplex in single stranded regions.

Roche teaches a sequencing adaptor suitable for sequencing circular nucleic acid (abstract, constructing libraries for single-molecule sequencing and a composition therefore.; para [0061], FIG. 1 depicts a single-stranded (denatured) library molecule.....Ligation of a single double-stranded adaptor to a double-stranded target molecule resulted in a double stranded circular molecule that can be denatured to yield structures depicted in FIG. 1. BC are barcodes present in the adaptor. Each adaptor contains two identical barcodes. In some embodiments, different barcodes may be used.....R and F are binding sites for a reverse and forward sequencing primers respectively. One of skill in the art will immediately appreciate that a single strand of nucleic acid (such as depicted in FIG. 1) contains a binding site (complementary sequence) for one primer (the F primer in FIG. 1), and the identical sequence to the opposite facing primer (R primer in FIG. 1), while the complementary strand (not shown in FIG. 1) would have identical sequence to the F primer and the binding site (complementary sequence) to the R primer. STOP is the strand synthesis terminator.; [0063], FIG. 2 depicts initiation of the first strand synthesis wherein the F primer binds to the primer binding site in the library molecule. The primer has an additional non-complementary sequence at its 5'-end. The additional sequence can contain a functional element, e.g., a sequencing primer binding site (P5).). Given that LTC also teaches sequencing circular nucleic acid (para [0069], Provided herein are blocking oligonucleotide adaptors that can be used as part of a workflow for joining together the two ends of a polynucleotide to form a circular molecule. In some embodiments, workflows can be used to prepare mate pair libraries for any next generation sequencing platform.), it would have been obvious to one of ordinary skill in the art to have provided read primer binding sites (Roche, Fig. 2, "P5-F") to the blocking oligonucleotide adaptor of LTC, facilitating the process of nucleic acid sequencing.

As said technical features were known in the art at the time of the invention, these cannot be considered special technical features that would otherwise unify the groups.

Groups I and II therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.

Item 4 (continued):

Claims 5-24, 37-46, 66-82, 84-91 are improper multiple dependent claims because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/62966

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

フロントページの続き

(32)優先日 令和3年9月1日(2021.9.1)

(33)優先権主張国・地域又は機関
米国(US)

(31)優先権主張番号 63/143,334

(32)優先日 令和3年1月29日(2021.1.29)

(33)優先権主張国・地域又は機関
米国(US)

(31)優先権主張番号 63/217,232

(32)優先日 令和3年6月30日(2021.6.30)

(33)優先権主張国・地域又は機関
米国(US)

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

(特許庁注：以下のものは登録商標)

1. TRITON

2. TWEEN

3. PYTHON

4. IGEPAL

イテュート， インコーポレイテッド

アメリカ合衆国 マサチューセッツ 02115-5450， ボストン， ブルックライン アヴェニュー 450

(74)代理人 110003971

弁理士法人葛和国際特許事務所

(72)発明者 アダルステインソン， ヴィクター， エー．

アメリカ合衆国 マサチューセッツ州 02142、ケンブリッジ、メイン ストリート 415、
ケア オブ ザ ブロード インスティテュート， インコーポレーテッド

(72)発明者 ベイ， ジン

アメリカ合衆国 マサチューセッツ州 02142、ケンブリッジ、メイン ストリート 415、
ケア オブ ザ ブロード インスティテュート， インコーポレーテッド

(72)発明者 リウ， ルオリン

アメリカ合衆国 マサチューセッツ州 02142、ケンブリッジ、メイン ストリート 415、
ケア オブ ザ ブロード インスティテュート， インコーポレーテッド

(72)発明者 マクリジョーゴス， ジェラッシモス

アメリカ合衆国 マサチューセッツ州 02215、ボストン、ブルックライン アヴェニュー 450

Fターム（参考） 4B063 QA13 QQ03 QQ42 QR08 QR62 QS36 QX01