

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7525402号
(P7525402)

(45)発行日 令和6年7月30日(2024.7.30)

(24)登録日 令和6年7月22日(2024.7.22)

(51)国際特許分類		F I		
C 1 2 Q	1/6813(2018.01)	C 1 2 Q	1/6813	Z Z N A
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z
G 0 1 N	33/53 (2006.01)	G 0 1 N	33/53	M
G 0 1 N	33/536(2006.01)	G 0 1 N	33/536	D
請求項の数 22 (全146頁)				
(21)出願番号	特願2020-564214(P2020-564214)	(73)特許権者	513322707	
(86)(22)出願日	令和1年5月14日(2019.5.14)		ナノストリング テクノロジーズ, イン	
(65)公表番号	特表2021-523723(P2021-523723		コーポレイティド	
	A)		アメリカ合衆国, ワシントン 9 8 1 0	
(43)公表日	令和3年9月9日(2021.9.9)		9, シアトル, フェアビュー アベニュー	
(86)国際出願番号	PCT/US2019/032168		ノース 5 3 0	
(87)国際公開番号	WO2019/222178	(74)代理人	100099759	
(87)国際公開日	令和1年11月21日(2019.11.21)		弁理士 青木 篤	
審査請求日	令和4年5月13日(2022.5.13)	(74)代理人	100123582	
(31)優先権主張番号	62/671,091		弁理士 三橋 真二	
(32)優先日	平成30年5月14日(2018.5.14)	(74)代理人	100117019	
(33)優先権主張国・地域又は機関			弁理士 渡辺 陽一	
	米国(US)	(74)代理人	100141977	
(31)優先権主張番号	62/836,327		弁理士 中島 勝	
(32)優先日	平成31年4月19日(2019.4.19)	(74)代理人	100150810	
最終頁に続く		最終頁に続く		

(54)【発明の名称】 化学的組成物とそれを利用する方法

(57)【特許請求の範囲】

【請求項1】

標的核酸の存在を同定するための方法であって、以下：

- (1) 第1のプローブの標的結合ドメインを、標的核酸の第1の領域にハイブリダイズさせ、前記プローブは、標的結合ドメインとバーコードドメインを含み、
前記標的結合ドメインは、長さが少なくとも12個のヌクレオチドであり；
前記バーコードドメインは、L-DNAからなり、前記バーコードドメインは、少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子にハイブリダイズする少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含んでおり、
前記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、並びに
前記バーコードドメインの各付着位置は、長さが6〜20個のヌクレオチドであり；
— (2) 少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識とを含む第1の相補的核酸分子を、前記バーコードドメインの前記少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズさせ；
(3) 前記第1の付着位置にハイブリダイズした前記第1の相補的核酸分子の前記少なくとも1つの第1の検出可能な標識と前記少なくとも1つの第2の検出可能な標識を同定し；
(4) 前記第1の付着位置にハイブリダイズした前記少なくとも1つの第1の検出可能な標識と前記少なくとも1つの第2の検出可能な標識を除去し；
(5) 少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識とを含む第2の相補的核酸分子を、前記バーコードドメインの前記少なくとも3つの付着位置

のうちの第2の付着位置にハイブリダイズさせ；

(6) 前記第2の付着位置にハイブリダイズした前記第2の相補的核酸分子の前記少なくとも1つの第3の検出可能な標識と前記少なくとも1つの第4の検出可能な標識を同定し；

(7) 前記第2の付着位置にハイブリダイズした前記少なくとも1つの第3の検出可能な標識と前記少なくとも1つの第4の検出可能な標識を除去し；

(8) 少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を含む第3の相補的核酸分子を、前記バーコードドメインの前記少なくとも3つの付着位置のうちの第3の付着位置にハイブリダイズさせ；並びに

(9) 前記第3の付着位置にハイブリダイズした前記第3の相補的核酸分子の前記少なくとも1つの第5の検出可能な標識と前記少なくとも1つの第6の検出可能な標識を同定し、それによって、前記少なくとも1つの第1の検出可能な標識、前記少なくとも1つの第2の検出可能な標識、前記少なくとも1つの第3の検出可能な標識、前記少なくとも1つの第4の検出可能な標識、前記少なくとも1つの第5の検出可能な標識、及び前記少なくとも1つの第6の検出可能な標識の属性に基づき、前記標的核酸の存在を決定すること、を含む、方法。

【請求項2】

前記プローブが、前記標的結合ドメインと前記バーコードドメインとの間に二本鎖DNAスパーサを含む、請求項1に記載の方法。

【請求項3】

前記標的結合ドメイン内のヌクレオチドの数が、前記バーコード内の付着位置の数よりも多い、請求項1に記載の方法。

【請求項4】

前記バーコードドメインが、少なくとも4つの付着位置を含む、請求項3に記載の方法。

【請求項5】

前記バーコードドメイン内の各付着位置が、1つの付着領域を含む、請求項1に記載の方法。

【請求項6】

前記バーコードドメイン内の各付着位置が、

- a) 長さが9個のヌクレオチド；
 - b) 長さが12個のヌクレオチド；
 - c) 長さが14個のヌクレオチド；又は
 - d) 長さが16個のヌクレオチド、
- である、請求項1に記載の方法。

【請求項7】

各付着位置の前記少なくとも1つの核酸配列が、3'末端グアノシンヌクレオチドを含む、請求項1に記載の方法。

【請求項8】

各付着位置の前記少なくとも1つの核酸配列が、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、3'末端グアノシンヌクレオチドとを含む、請求項1に記載の方法。

【請求項9】

前記標的結合ドメインの前記少なくとも12個のヌクレオチドの各ヌクレオチドが、D-DNAである、請求項1に記載の方法。

【請求項10】

工程(4)と(5)が逐次的に、または同時に起こる、請求項1に記載の方法。

【請求項11】

工程(7)と(8)が逐次的に、または同時に起こる、請求項1に記載の方法。

【請求項12】

前記第1の検出可能な標識と前記第2の検出可能な標識が、同じ発光スペクトルを有する

10

20

30

40

50

か、異なる発光スペクトルを有する、請求項1に記載の方法。

【請求項13】

前記第3の検出可能な標識と前記第4の検出可能な標識が、同じ発光スペクトルを有するか、異なる発光スペクトルを有する、請求項1に記載の方法。

【請求項14】

前記第5の検出可能な標識と前記第6の検出可能な標識が、同じ発光スペクトルを有するか、異なる発光スペクトルを有する、請求項1に記載の方法。

【請求項15】

前記第1の相補的核酸分子、前記第2の相補的核酸分子、および前記第3の相補的核酸分子のそれぞれが、切断可能なリンカーを含む、請求項1に記載の方法。

【請求項16】

前記第1の相補的核酸分子が、1個の一次核酸分子を含むレポータプローブを含み、
前記一次核酸分子が、少なくとも2つのドメインを含み、
第1のドメインは、前記バーコードドメインの前記第1の付着位置にハイブリダイズし、及び

第2のドメインは、6個の二次核酸分子にハイブリダイズし、
これら二次核酸分子のそれぞれが、5個の三次核酸分子にハイブリダイズし、前記三次核酸分子のそれぞれが、検出可能な標識を含む、
請求項1に記載の方法。

【請求項17】

前記一次核酸分子が、前記第1のドメインと前記第2のドメインとの間に位置する切断可能なリンカーを含む、請求項16に記載の方法。

【請求項18】

前記切断可能なリンカーが、光で切断可能なリンカーである、請求項17に記載の方法。

【請求項19】

前記二次核酸分子のそれぞれが少なくとも2つのドメインを含み、
第1のドメインは、前記一次核酸分子の第2のドメインにハイブリダイズし、
第2のドメインは、前記5個の三次核酸分子にハイブリダイズする、請求項16に記載の方法。

【請求項20】

前記二次核酸分子のそれぞれが、前記第1のドメインと前記第2のドメインとの間に位置する切断可能なリンカーを含む、請求項19に記載の方法。

【請求項21】

前記切断可能なリンカーが、光で切断可能なリンカーである、請求項20に記載の方法。

【請求項22】

前記第1の付着位置にハイブリダイズした前記少なくとも1つの第1の検出可能な標識と前記少なくとも1つの第2の検出可能な標識の除去が、
前記一次核酸の第1のドメインと第2のドメインの間の切断可能なリンカーの切断、それぞれの二次核酸の第1のドメインと第2のドメインの間の切断可能なリンカーの切断、またはこれらの任意の組み合わせ、
を含む、請求項17、18、20及び21のいずれか1項に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連する出願の相互参照

本出願は、2018年5月14日に米国合衆国仮出願第62/671,091号と、2019年4月19日に米国合衆国仮出願第62/836,327号の優先権と利益を主張する。これら特許出願それぞれの内容は、その全体が参照によって本明細書に組み込まれている。

【0002】

配列リスト

本出願は、EFS-Webを通じてASCII形式で提出された配列リストを含有しており、その全体が参照によって本明細書に組み込まれている。このASCIIのコピーは2019年5月13日に作成されて「NATE-039_001WO_SeqList.txt」と名づけられており、サイズは25,129バイトである。

【背景技術】

【0003】

現在、核酸のシークエンシング（すなわち1個の核酸分子内のヌクレオチドの正確な順序を求めるプロセス）のための多彩な方法が存在している。現在の方法は、核酸を酵素で増幅すること（例えばPCR）、および／またはクローニングによって増幅することを必要とする。光検出手段によって検出できるシグナルを発生させるには、酵素によるさらなる重合が必要とされる。このような増幅工程と重合工程はコストおよび／または時間がかかる。したがって本分野では、迅速で増幅と酵素が不要な核酸シークエンシングの方法が必要とされている。本開示はこうした要求に応えるものである。

【発明の概要】

【0004】

本開示により、読み出し長が長く、エラー率が小さく、迅速で、酵素と増幅とライブラリが不要な核酸シークエンシングを提供するシークエンシング用プローブと、方法と、キットと、装置が提供される。本明細書に記載したシークエンシング用プローブはバーコードドメインを含んでいて、そのバーコードドメイン内の各位置は、標的結合ドメイン内の少なくとも2個のヌクレオチドに対応している。さらに、この方法、キット、装置は、サンプルから結果を迅速に取得する能力を有する。これらの特徴は、臨床状況におけるシークエンシングで特に有用である。本開示は、米国特許出願公開第2016/0194701号に開示されている開示内容（その内容全体が参照によって本明細書に組み込まれている）の改良である。

【0005】

本開示により、標的結合ドメインとバーコードドメインとを含むプローブが提供され、このプローブでは、標的結合ドメインは、少なくとも8個のヌクレオチドを含んでいて標的核酸にハイブリダイズし、標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定し、標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子にハイブリダイズする少なくとも1つの核酸配列を含む少なくとも1つの付着位置を含んでおり、合成骨格はL-DNAを含み、上記少なくとも3つの付着位置の各付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置の各付着位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドの対応する2個のヌクレオチドの位置と属性（identity）を決めており；第1の相補的二次核酸分子が上記少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズし、第1の相補的二次核酸分子は少なくとも2つのドメインと切断可能なリンカーを含み、第1のドメインはバーコードドメインの第1の付着位置にハイブリダイズし、第2のドメインは少なくとも1つの相補的二次核酸分子にハイブリダイズすることができ、リンカーの修飾は、

10

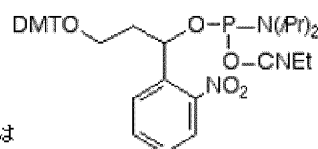
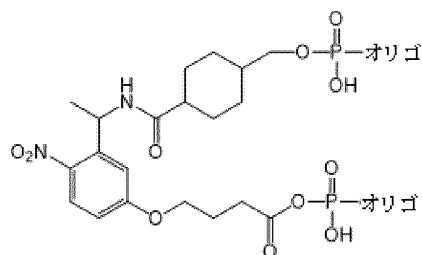
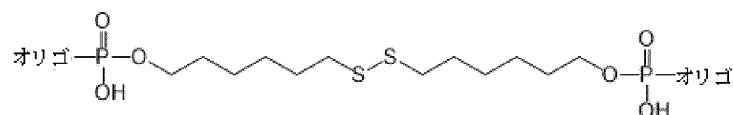
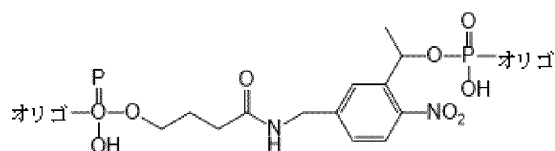
20

30

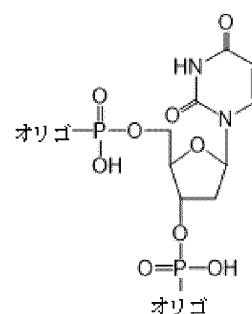
40

50

【化1】



または



10

20

のいずれかであり、上記リンカーの修飾は、第1のドメインと第2のドメインの間に位置している。

【0006】

プローブは約60個のヌクレオチドを含むことができる。プローブは、一本鎖DNA合成骨格と、標的化結合ドメインとバーコードドメインの間の二本鎖DNAスペーサを含むことができる。一本鎖DNA合成骨格はL-DNAを含むことができる。一本鎖DNA合成骨格は約27個のヌクレオチドを含むことができる。二本鎖DNAスペーサはL-DNAを含むことができる。二本鎖DNAスペーサは長さが約25個のヌクレオチドを含むことができる。

30

【0007】

プローブの標的結合ドメイン内のヌクレオチドの数は、このプローブのバーコードドメイン内の付着位置の数よりも多くすることができる。標的結合ドメインは8個のヌクレオチドを含むことができ、バーコードドメインは3つの付着位置を含むことができる。標的結合ドメインの中であって標的核酸分子内の対応するヌクレオチドを同定しないヌクレオチドのうちの少なくとも1個は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドよりも前に位置することができ、標的核酸分子内の対応するヌクレオチドを同定しない標的結合ドメイン内のヌクレオチドのうちの少なくとも1個は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドの後に位置することができる。

40

【0008】

バーコードドメイン内の1つの付着位置は、1つの付着領域を含むことができる。バーコードドメイン内の各付着位置の少なくとも1つの核酸配列は、約9個のヌクレオチドを含むことができる。付着位置の少なくとも1つの核酸配列は、3'末端グアノシンヌクレオチドを含むことができる。各付着位置の少なくとも1つの核酸配列は、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、3'末端グアノシンヌクレオチドとを含むことができる。付着位置の少なくとも1つの核酸配列の各ヌクレオチドとして、L-DNAが可能である。標的結合ドメインの少なくとも8個のヌクレオチドの各ヌクレオチドとして、D-DNAが可能である。

50

【0009】

相補的核酸分子として一次核酸分子が可能であり、この一次核酸分子は、バーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域に直接結合することができる。一次核酸分子は少なくとも2つのドメインを含むことができ、第1のドメインは、バーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域に結合することができる、第2のドメインは、少なくとも1つの相補的二次核酸分子に結合することができる。一次核酸分子の第1のドメインはL-DNAを含むことができる。一次核酸分子の第2のドメインはD-DNAを含むことができる。一次核酸分子の第1のドメインは5'末端シトシンヌクレオチドを含むことができる。一次核酸分子の第1のドメインは、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のグアニンヌクレオチド、またはこれらの任意の組み合わせと、5'末端シトシンヌクレオチドを含むことができる。切断可能なリンカーは、一次核酸分子の第1のドメインと一次核酸分子の第2のドメインとの間に位置することができる。切断可能なリンカーは、少なくとも1つの切断可能な部分を含むことができる。切断可能な部分として、光で切断可能な部分 (photocleavable moiety) が可能である。

10

【0010】

一次核酸分子は、バーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域にハイブリダイズするとともに、少なくとも1個の二次核酸分子にハイブリダイズすることができる。一次核酸分子は4個の二次核酸分子にハイブリダイズすることができる。

20

【0011】

二次核酸分子は少なくとも2つのドメインを含むことができ、第1のドメインは、少なくとも1個の一次核酸分子内の相補的配列に結合することができる、第2のドメインは、(a) 第1の検出可能な標識と少なくとも第2の検出可能な標識に、または (b) 少なくとも1個の相補的三次核酸分子に、または (c) これらの組み合わせに結合することができる。二次核酸分子は切断可能なリンカーを含むことができる。切断可能なリンカーは、第1のドメインと第2のドメインとの間に位置することができる。切断可能な部分は、光で切断することができる。二次核酸分子は、少なくとも1個の一次核酸分子にハイブリダイズするとともに、少なくとも1個の三次核酸分子にハイブリダイズすることができる。二次核酸分子は、(a) 少なくとも1個の一次核酸分子と、(b) 少なくとも1個の三次核酸分子と、(c) 第1の検出可能な標識と少なくとも第2の検出可能な標識にハイブリダイズすることができる。それぞれの二次核酸分子は1個の三次核酸分子にハイブリダイズすることができる。第1の検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

30

【0012】

三次核酸分子は少なくとも2つのドメインを含むことができ、第1のドメインは、二次核酸分子内の相補的配列に結合することができる、第2のドメインは、第1の検出可能な標識と少なくとも第2の検出可能な標識に結合することができる。三次核酸分子は切断可能なリンカーを含んでいる。切断可能なリンカーは、第1のドメインと第2のドメインとの間に位置することができる。切断可能なリンカーは、光で切断することができる。三次核酸分子は少なくとも1個の二次核酸分子にハイブリダイズさせることができ、第1の検出可能な標識と少なくとも第2の検出可能な標識を含むことができる。第1の検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

40

【0013】

二次核酸分子上に位置する少なくとも第1および第2の検出可能な標識は同じ発光スペクトルを有することができ、三次核酸分子上に位置する少なくとも第1および第2の検出可能な標識は同じ発光スペクトルを有することができ、二次核酸分子上の検出可能な標識の発光スペクトルは、三次核酸分子上の検出可能な標識の発光スペクトルと異なってもよい。

50

【0014】

一次核酸分子は4個の二次核酸分子にハイブリダイズさせることができ、これら4個の二次核酸分子のそれぞれは4つの第1の検出可能な標識を含んでいて、1個の三次核酸分子にハイブリダイズし、この三次核酸分子は5つの検出可能な標識を含んでいる。二次核酸分子の第1の検出可能な標識の発光スペクトルは、三次核酸分子上の第2の検出可能な標識の発光スペクトルと異なっている。10

【0015】

本開示により、核酸のヌクレオチド配列を決定する方法として、(1) 請求項1の少なくとも1つの第1のプロープの標的結合ドメインを、標的核酸の第1の領域にハイブリダイズさせ、標的核酸は、任意により基板に1つ以上の位置で固定化され；(2) 少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズさせ；(3) 第1の付着位置にハイブリダイズした第1の相補的核酸分子の少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を同定し；(4) 第1の付着位置にハイブリダイズした少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を除去し；(5) 少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を含む第2の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置にハイブリダイズさせ；(6) 第2の付着位置にハイブリダイズした第2の相補的核酸分子の少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を同定し；(7) 第2の付着位置にハイブリダイズした少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を除去し；(8) 少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を含む第3の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第3の付着位置にハイブリダイズさせ；(9) 第3の付着位置にハイブリダイズした第3の相補的核酸分子の少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を同定し；並びに(10) 上記少なくとも1つの第1の検出可能な標識と、上記少なくとも1つの第2の検出可能な標識と、上記少なくとも1つの第3の検出可能な標識と、上記少なくとも1つの第4の検出可能な標識と、上記少なくとも1つの第5の検出可能な標識と、上記少なくとも1つの第6の検出可能な標識の属性に基づき、上記少なくとも1つの第1のプロープの標的結合ドメインの上記少なくとも6個のヌクレオチドにハイブリダイズした任意により固定化された標的核酸の少なくとも6個のヌクレオチドのヌクレオチド配列を決定することを含む方法が提供される。20

【0016】

前段の方法はさらに、(11) 任意により固定化された標的核酸の第1の領域から少なくとも1つの第1のプロープを除去し；(12) 請求項1に記載の少なくとも1つの第2のプロープの標的結合ドメインを、任意により固定化された標的核酸の第2の領域にハイブリダイズさせ（ただし第1のプロープと少なくとも第2のプロープの標的結合ドメインは異なる）；(13) 少なくとも1つの第7の検出可能な標識と少なくとも1つの第8の検出可能な標識を含む第4の相補的核酸分子を、少なくとも1つの第2のプロープのバーコードドメインの少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズさせ；(14) 第1の付着位置にハイブリダイズした第4の相補的核酸分子の少なくとも1つの第7の検出可能な標識と少なくとも1つの第8の検出可能な標識を同定し；(15) 第1の付着位置にハイブリダイズした第4の相補的核酸分子の少なくとも1つの第7の検出可能な標識と少なくとも1つの第8の検出可能な標識を除去し；(16) 少なくとも1つの第9の検出可能な標識と少なくとも1つの第10の検出可能な標識を含む第5の相補的核酸分子を、少なくとも1つの第2のプロープのバーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置にハイブリダイズさせ；(17) 第2の付着位置にハイブリダイズした第5の相補的核酸分子の少なくとも1つの第9の検出可能な標識と少なくとも1つの第10の検出可能な標識を同定し；(18) 第2の付着位置にハイブリダイズした少なくとも1つの第9の検出可能な標識と少なくとも1つの第10の検出可能な標識を除去し；(19) 少なくとも1つの第11の検出40

可能な標識と少なくとも1つの第12の検出可能な標識を含む第6の相補的核酸分子を、少なくとも1つの第2のプロープのバーコードドメインの上記少なくとも3つの付着位置のうちの第3の付着位置にハイブリダイズさせ；(20) 第3の付着位置にハイブリダイズした第6の相補的核酸分子の少なくとも1つの第11の検出可能な標識と少なくとも1つの第12の検出可能な標識を同定し；(21) 上記少なくとも1つの第7の検出可能な標識と、上記少なくとも1つの第8の検出可能な標識と、上記少なくとも1つの第9の検出可能な標識と、上記少なくとも1つの第10の検出可能な標識と、上記少なくとも1つの第11の検出可能な標識と、上記少なくとも1つの第12の検出可能な標識の属性に基づき、上記少なくとも1つの第2のプロープの標的結合ドメインの上記少なくとも6個のヌクレオチドにハイブリダイズした任意により固定化された上記標的核酸の少なくとも6個のヌクレオチドのヌクレオチド配列を決定することを含むことができる。

10

【0017】

この方法はさらに、任意により固定化された標的核酸の少なくとも第1の領域と少なくとも第2の領域の中のそれぞれ同定された線形順序 (linear order) のヌクレオチドをアセンブルし、それによって、任意により固定化されたその標的核酸の配列を同定することをさらに含むことができる。

【0018】

工程(4)と(5)は、逐次的に、または同時に起こることができる。工程(7)と(8)は、逐次的に、または同時に起こることができる。

【0019】

20

第1の検出可能な標識と第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。第3の検出可能な標識と第4の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。第5の検出可能な標識と第6の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

【0020】

第1の相補的核酸分子、第2の相補的核酸分子、第3の相補的核酸分子は、切断可能なリンカーを含むことができる。切断可能なリンカーは、光で切断することができる。

【0021】

第1の相補的核酸分子は、1個の一次核酸分子と、4個の二次核酸分子と、4個の三次核酸分子を含むことができ、一次核酸分子は4個の二次核酸分子にハイブリダイズし、4個の二次核酸分子のそれぞれは、4つの第1の検出可能な標識を含むとともに、1個の三次核酸分子にハイブリダイズし、4個の三次核酸分子のそれぞれは、5つの第2の検出可能な標識を含んでいる。

30

【0022】

一次核酸分子は、少なくとも2つのドメインを含むこと、すなわちバーコードドメインの第1の付着位置にハイブリダイズする第1のドメインと、4個の二次核酸分子にハイブリダイズする第2のドメインを含むことができる。一次核酸分子は、第1のドメインと第2のドメインとの間に位置する切断可能なリンカーを含むことができる。

【0023】

40

二次核酸分子は、少なくとも2つのドメインを含むこと、すなわち一次核酸分子の第2のドメインにハイブリダイズする第1のドメインと、4つの第1の検出可能な標識を含んでいて1個の三次核酸分子にハイブリダイズする第2のドメインを含むことができる。二次核酸分子は、第1のドメインと第2のドメインとの間に位置する切断可能なリンカーを含むことができる。

【0024】

第1の付着位置にハイブリダイズした少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識の除去は、一次核酸の第1のドメインと第2のドメインの間の切断可能なリンカーを切断すること、または各二次核酸の第1のドメインと第2のドメインの間の切断可能なリンカーを切断すること、またはこれらの任意の組み合わせを含むこ

50

とができる。

【0025】

本開示により、少なくとも1つの分子複合体を含む組成物として、この少なくとも1つの分子複合体が、(A) 生物サンプルから得られた標的核酸分子と、(B) 少なくとも2つの核酸分子複合体を含んでいて、これら2つの核酸分子複合体のうちの第1の複合体は、部分的に二本鎖の第1の核酸分子を含み、この部分的に二本鎖の第1の核酸分子の1本の鎖は、標的核酸分子の第1の部分にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールする二本鎖ドメインと、少なくとも1つの第1の親和性部分を含み、部分的に二本鎖の第1の核酸分子の他方の鎖は、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールする二本鎖ドメインと、基板に付着した相補的核酸にハイブリダイズする基板特異的ドメインと、少なくとも1つの第2の親和性部分を含み、第2の複合体は、部分的に二本鎖の第2の核酸分子を含み、この部分的に二本鎖の第2の核酸分子の1本の鎖は、標的核酸分子の第2の部分にハイブリダイズする標的特異的ドメイン（第1の部分と第2の部分は重複していない）と、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールする二本鎖ドメインを含み、部分的に二本鎖の第2の核酸分子の他方の鎖は、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールする二本鎖ドメインと、標的核酸の供給源となった生物サンプルを同定するサンプル特異的ドメインと、第1の一本鎖精製配列と、二本鎖ドメインとサンプル特異的ドメインの間に位置する第1の切断可能な部分と、サンプル特異的ドメインと第1の一本鎖精製配列の間に位置する第2の切断可能な部分を含む組成物が提供される。

【0026】

本開示により、少なくとも1つの分子複合体を含む組成物として、この少なくとも1つの分子複合体が、(A) 生物サンプルから得られた標的核酸分子と、(B) 少なくとも2つの核酸分子複合体を含んでいて、これら2つの核酸分子複合体のうちの第1の複合体は、部分的に二本鎖の第1の核酸分子を含み、この部分的に二本鎖の第1の核酸分子の1本の鎖は、標的核酸分子の第1の部分にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールする二本鎖ドメインと、少なくとも1つの第1の親和性部分を含み、部分的に二本鎖の第1の核酸分子の他方の鎖は、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールするとともに標的核酸分子の3'末端に機能可能に連結された二本鎖ドメインと、基板に付着した相補的核酸にハイブリダイズする基板特異的ドメインと、少なくとも1つの第2の親和性部分を含み、第2の複合体は、部分的に二本鎖の第2の核酸分子を含み、この部分的に二本鎖の第2の核酸分子の1本の鎖は、標的核酸分子の第2の部分にハイブリダイズする標的特異的ドメイン（第1の部分と第2の部分は重複していない）と、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールする二本鎖ドメインを含み、部分的に二本鎖の第2の核酸分子の他方の鎖は、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールするとともに標的核酸分子の5'末端に機能可能に連結された二本鎖ドメインと、標的核酸の供給源となった生物サンプルを同定するサンプル特異的ドメインと、第1の一本鎖精製配列と、二本鎖ドメインとサンプル特異的ドメインの間に位置する第1の切断可能な部分を含む組成物が提供される。

【0027】

本開示により、少なくとも1つの分子複合体を含む組成物として、この少なくとも1つの分子複合体が、(A) 生物サンプルから得られた標的核酸分子と、(B) 少なくとも2つの核酸分子複合体を含んでいて、これら2つの核酸分子複合体のうちの第1の複合体は、部分的に二本鎖の第1の核酸分子を含み、この部分的に二本鎖の第1の核酸分子の1本の鎖は、標的核酸分子の第1の部分にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールする二本鎖ドメインと、少なくとも1つの第1の親和性部分を含み、部分的に二本鎖の第1の核酸分子の他方の鎖は、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールするとともに標的核酸分子の3'末端に機能可能に連結された二本鎖ドメインと、基板に付着した相補的核酸にハイブリダイズする基板特異的ドメインと、少なくとも1つの第2の親和性部分を含み、第2の複合体は、部分的に二

本鎖の第2の核酸分子を含み、この部分的に二本鎖の第2の核酸分子の1本の鎖は、標的核酸分子の第2の部分にハイブリダイズする標的特異的ドメイン（第1の部分と第2の部分は重複していない）と、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールする二本鎖ドメインを含み、部分的に二本鎖の第2の核酸分子の他方の鎖は、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールするとともに標的核酸分子の5'末端に機能可能に連結された二本鎖ドメインを含む組成物が提供される。

【0028】

本開示により、平坦な固体支持基板と；この平坦な固体支持基板上の第1の層と；この第1の層の上の第2の層を含んでいて、第2の層が複数のナノウェル（nanowell）を含み、各ナノウェルが第1の層の露出部分へのアクセスを提供し、各ナノウェルが、第1の層の露出部分に共有結合した複数の第1のオリゴヌクレオチドを含む組成物も提供される。

10

【0029】

本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインは少なくとも8個のヌクレオチドを含んでいて標的核酸にハイブリダイズし、この標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定し、この標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子にハイブリダイズする少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含んでおり、上記少なくとも3つの付着位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドの位置と属性を決めていて、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有するシーケンシング用プローブが提供される。

20

【0030】

本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインは少なくとも8個のヌクレオチドを含んでいて標的核酸にハイブリダイズし、この標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定し、この標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子にハイブリダイズする少なくとも1つの核酸配列を含む少なくとも1つの付着位置を含んでおり、上記少なくとも3つの付着位置の各付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置の各付着位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドのうちの対応する2個のヌクレオチドの位置と属性を決めているシーケンシング用プローブも提供される。

30

【0031】

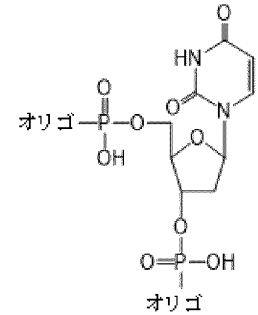
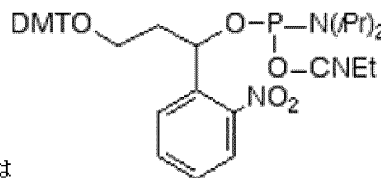
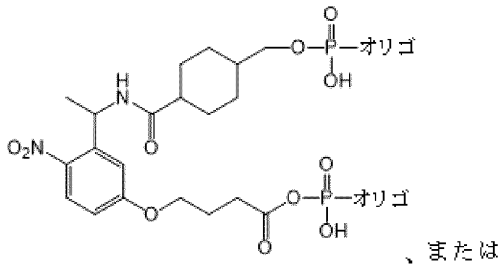
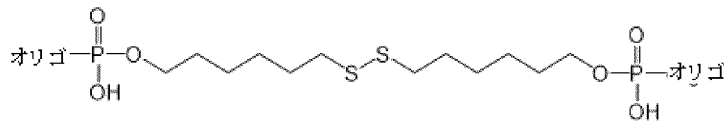
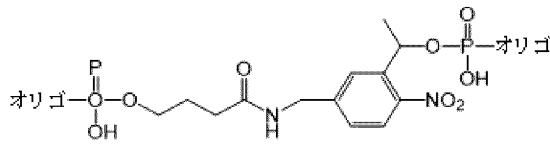
本開示により、a) 標的結合ドメインとバーコードドメインを含む組成物を含む複合体が提供され、この複合体では、標的結合ドメインは、少なくとも8個のヌクレオチドを含んでいて標的核酸にハイブリダイズし、標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定し、標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子にハイブリダイズする少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含んでおり、上記少なくとも3つの付着位置の各付着位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドの対応する2個のヌクレオチドの位置と属性を決めており、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し；第1の相補的一次核酸分子が上記少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズし、第1の相補的一次核酸分子は少なくとも2つの

40

50

ドメインと切断可能なリンカーを含み、第1のドメインはバーコードドメインの第1の付着位置にハイブリダイズし、第2のドメインは少なくとも1つの相補的二次核酸分子にハイブリダイズすることができ、切断可能なリンカーは、

【化2】



のいずれかであり、第1のドメインと第2のドメインとの間に位置する。

【0032】

本開示により、核酸のヌクレオチド配列を決定する方法として、(1) 本開示の第1のシーケンシング用プローブの標的結合ドメインを、標的核酸の第1の領域にハイブリダイズさせ、標的核酸は、任意により基板に1つ以上の位置で固定化され；(2) 少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズさせ；(3) 第1の付着位置にハイブリダイズした第1の相補的核酸分子の少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を同定し；(4) 第1の付着位置にハイブリダイズした少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を除去し；(5) 少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を含む第2の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置にハイブリダイズさせ；(6) 第2の付着位置にハイブリダイズした第2の相補的核酸分子の少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を同定し；(7) 第2の付着位置にハイブリダイズした少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を除去し；(8) 少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を含む第3の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第3の付着位置にハイブリダイズさせ；(9) 第3の付着位置にハイブリダイズした第3の相補的核酸分子の少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を同定し；(10) 上記少なくとも1つの第1の検出可能な標識と、上記少なくとも1つの第2の検出可能な標識と、上記少なくとも1つの第3の検出可能な標識と、上記少なくとも1つの第4の検出可能な標識と、上記少なくとも1つの第5の検出可能な標識と、上記少なくとも1つの第6の検出可能な標識の属性に基づき、第1のシ

ークエンシング用プローブの標的結合ドメインの上記少なくとも6個のヌクレオチドにハイブリダイズした任意により固定化された標的核酸の少なくとも6個のヌクレオチドのヌクレオチド配列を決定することを含む方法が提供される。

【0033】

本開示により、核酸のヌクレオチド配列を決定する方法として、(1) 請求項113または114の第1のシークエンシング用プローブの標的結合ドメインを、標的核酸の第1の領域にハイブリダイズさせ、標的核酸は、任意により基板に1つ以上の位置で固定化され；(2) 少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズさせ；(3) 第1の付着位置にハイブリダイズした第1の相補的核酸分子の少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を同定し；(4) 上記少なくとも1つの第1の検出可能な標識と上記少なくとも1つの第2の検出可能な標識の属性に基づき、標的結合ドメインの上記少なくとも6個のヌクレオチドのうちの2個にハイブリダイズする任意により固定化された標的核酸内の第1のヌクレオチドと第2のヌクレオチドの位置と属性を同定し；(5) 第1の付着位置にハイブリダイズした少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を除去し；(6) 少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を含む第2の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置にハイブリダイズさせ；(7) 第2の付着位置にハイブリダイズした第2の相補的核酸分子の少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を同定し；(8) 上記少なくとも1つの第3の検出可能な標識と上記少なくとも1つの第4の検出可能な標識の属性に基づき、標的結合ドメインの上記少なくとも6個のヌクレオチドのうちの2個にハイブリダイズする任意により固定化された標的核酸内の第3のヌクレオチドと第4のヌクレオチドの位置と属性を同定し；(9) 第2の付着位置にハイブリダイズした少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を除去し；(10) 少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を含む第3の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第3の付着位置にハイブリダイズさせ；(11) 第3の付着位置にハイブリダイズした第3の相補的核酸分子の少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を同定し；(12) 上記少なくとも1つの第5の検出可能な標識と上記少なくとも1つの第6の検出可能な標識の属性に基づき、標的結合ドメインの上記少なくとも6個のヌクレオチドのうちの2個にハイブリダイズする任意により固定化された標的核酸内の第5のヌクレオチドと第6のヌクレオチドの位置と属性を同定することにより；第1のシークエンシング用プローブの標的結合ドメインの上記少なくとも6個のヌクレオチドにハイブリダイズした任意により固定化された標的核酸の少なくとも6個のヌクレオチドのヌクレオチド配列を決定することを含む方法が提供される。

【0034】

本開示により、標的核酸内の所定のヌクレオチド配列の存在を特定する方法として、(1) 本開示の第1のシークエンシング用プローブの標的結合ドメインを、標的核酸の第1の領域にハイブリダイズさせ、標的核酸は、任意により基板に1つ以上の位置で固定化され；(2) 少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置にハイブリダイズさせ；(3) 第1の付着位置にハイブリダイズした第1の相補的核酸分子の少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を同定し；(4) 第1の付着位置にハイブリダイズした少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を除去し；(5) 少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を含む第2の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置にハイブリダイズさせ；(6) 第2の付着位置にハイブリダイズした第2の相補的核酸分子の少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を同定し；

(7) 第2の付着位置にハイブリダイズした少なくとも1つの第3の検出可能な標識と少なくとも1つの第4の検出可能な標識を除去し；(8) 少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を含む第3の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第3の付着位置にハイブリダイズさせ；(9) 第3の付着位置にハイブリダイズした第3の相補的核酸分子の少なくとも1つの第5の検出可能な標識と少なくとも1つの第6の検出可能な標識を同定することにより、上記少なくとも1つの第1の検出可能な標識と、上記少なくとも1つの第2の検出可能な標識と、上記少なくとも1つの第3の検出可能な標識と、上記少なくとも1つの第4の検出可能な標識と、上記少なくとも1つの第5の検出可能な標識と、上記少なくとも1つの第6の検出可能な標識の属性に基づき、所定のヌクレオチド配列の存在を明らかにすることを含む方法が提供される。

10

【0035】

本開示により、キットとして、(A) 部分的に二本鎖の第1の核酸分子を含む第1の核酸分子複合体（この部分的に二本鎖の第1の核酸分子の1本の鎖は、標的核酸分子の第1の部分にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールする二本鎖ドメインと、少なくとも1つの第1の親和性部分を含み、部分的に二本鎖の第1の核酸分子の他方の鎖は、この部分的に二本鎖の第1の核酸分子の他方の鎖にアニールする二本鎖ドメインと、基板に付着した相補的核酸にハイブリダイズする基板特異的ドメインと、少なくとも1つの第2の親和性部分を含む）と；(B) 部分的に二本鎖の第2の核酸分子を含む第2の核酸分子複合体（この部分的に二本鎖の第2の核酸分子の1本の鎖は、標的核酸分子の第2の部分にハイブリダイズする標的特異的ドメイン（第1の部分と第2の部分は重複していない）と、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールする二本鎖ドメインを含み、部分的に二本鎖の第2の核酸分子の他方の鎖は、この部分的に二本鎖の第2の核酸分子の他方の鎖にアニールする二本鎖ドメインと、標的核酸を取得する元になった生物サンプルを同定するサンプル特異的ドメインと、基板に付着した相補的核酸にハイブリダイズする基板特異的ドメインと、第1の一本鎖精製配列と、二本鎖ドメインとサンプル特異的ドメインの間に位置する第1の切断可能な部分と、サンプル特異的ドメインと第1の一本鎖精製配列の間に位置する第2の切断可能な部分を含む）を含むキットが提供される。

20

【0036】

本開示により、キットとして、(A) 標的核酸分子の第1の部分にハイブリダイズする標的特異的ドメインと、第2の核酸分子の二本鎖ドメインにアニールする二本鎖ドメインと、少なくとも1つの第1の親和性部分を含む第1の一本鎖核酸分子と、(B) 第1の一本鎖核酸分子の二本鎖ドメインにアニールする二本鎖ドメインと、基板に付着した相補的核酸にハイブリダイズする基板特異的ドメインと、少なくとも1つの第2の親和性部分を含む第2の一本鎖核酸分子と、(C) 標的核酸の第2の部分にハイブリダイズする標的特異的ドメイン（ただし第1の部分と第2の部分は重複していない）と、第4の一本鎖核酸分子の二本鎖ドメインにアニールする二本鎖ドメインを含む第3の一本鎖核酸分子と、(D) 第3の一本鎖核酸分子の二本鎖ドメインにアニールする二本鎖ドメインと、標的核酸を取得する元になった生物サンプルを同定するサンプル特異的ドメインと、第1の一本鎖精製配列と、二本鎖ドメインとサンプル特異的ドメインの間に位置する第1の切断可能な部分と、サンプル特異的ドメインと第1の一本鎖精製配列の間に位置する第2の切断可能な部分を含む第4の一本鎖核酸分子を含むキットも提供される。

30

40

【0037】

上記の態様の任意のものを他の任意の態様と組み合わせることができる。

【0038】

特に断わらない限り、あらゆる科学技術用語は、本開示が属する分野の当業者が一般に理解しているのと同じ意味を有する。本明細書では、単数形は、文脈から明らかに異なることがわかる場合を除いて複数も含んでおり、例えば「1つの」、「その」は、単数または複数であると理解され、「または」という用語は、包括的であると理解される。例えば

50

「1つの要素」は、1つ以上の要素を意味する。本明細書全体を通じ、「含んでいる」またはそのバリエーションである「含む」は、記載されている1つの要素、1つの整数、1つの工程、または一群の要素、一群の整数、一群の工程が含まれると理解するが、他の任意の要素、整数、工程や、一群の要素、一群の整数、一群の工程が除外されることはない。「約」は、記載されている数値の10%以内、9%以内、8%以内、7%以内、6%以内、5%以内、4%以内、3%以内、2%以内、1%以内、0.5%以内、0.1%以内、0.05%以内、0.01%以内のいずれかであると理解することができる。文脈から明確である場合を除き、本明細書に提示されているあらゆる数値は「約」という用語によって変更される。

【0039】

本開示を実施または試験する際に本明細書に記載の方法および材料と類似するか同等な方法および材料を使用できるが、以下には適切な方法および材料が記載されている。本明細書で言及されているあらゆる刊行物、特許出願、特許と、それ以外の参考文献は、その全体が参照によって本明細書に組み込まれている。本明細書で引用されている参考文献は、権利を請求する本発明の先行技術であることを認めるものではない。矛盾がある場合には、本明細書が、定義を含め、優先する。それに加え、材料、方法、実施例は説明だけ目的としており、制限する意図はない。本開示の他の特徴と利点は、以下の詳細な説明と請求項から明らかになる。

【0040】

本特許または本出願のファイルは、少なくとも1枚のカラー図面を含有している。カラー図面を伴うこの特許または出願の公開物のコピーは、請求して必要な手数料を支払うと特許商標庁から提供される。

【0041】

上記の特徴とさらなる特徴は、添付の図面と組み合わせたとき、以下の詳細な説明からより明確になる。

【図面の簡単な説明】

【0042】

【図1】図1は、本開示の代表的な1つのシーケンシング用プローブの図である。

【図2】図2は、本開示の標準プローブと、3部構成シーケンシング用プローブと、1部構成リンカープローブの設計を示す。

【図3】図3は、本開示の代表的なシーケンシング用プローブにハイブリダイズした本開示の代表的なレポータ複合体の図である。

【図4】図4は、本開示の代表的なレポータプローブの模式図を示している。

【図5】図5は、さまざまな構成の三次核酸を含む本開示の代表的ないくつかのレポータプローブの模式図である。

【図6】図6は、分岐している三次核酸を含む本開示の代表的ないくつかのレポータプローブの模式図である。

【図7】図7は、本開示の代表的なレポータプローブの中で切断可能なリンカーの修飾が可能な位置を示している。

【図8】図8は、本開示の2捕捉用プローブ系を用いて標的核酸を捕捉する模式図である。

【図9】図9は、FFPEサンプルを用いた100個の標的からなる多重がんパネルを捕捉して検出するために本発明の方法を利用する実験からの結果を示している。

【図10】図10は、本開示のシーケンシング法の1つのサイクルの模式図である。

【図11-1】図11は、本開示のシーケンシング法の1つのサイクルの模式図と、このサイクルの間に回収された対応するイメージングデータである。

【図11-2】図11は、本開示のシーケンシング法の1つのサイクルの模式図と、このサイクルの間に回収された対応するイメージングデータである。

【図12】図12は、本開示の代表的なシーケンシング用プローブのプールの構成を示しており、8通りの色組み合わせを用いて異なる8通りのシーケンシング用プローブのプールを設計している。

【図13】図13は、米国特許出願公開第2016/0194701号に開示されているバーコード

10

20

30

40

50

ドメインの設計を本開示のバーコードドメインの設計と比較している。

【図14】図14は、本開示のシーケンシングサイクルの模式図であり、切断可能なリンカーの修飾を利用してバーコード内の1つの位置を暗くしている。

【図15】図15は、本開示の代表的なシーケンシングサイクルの模式図であり、バーコードドメイン内の1つの位置が一次核酸の置換によって暗くされている。

【図16】図16は、本開示のシーケンシング法を利用すると、異なるシーケンシング用プローブを用いた標的核酸の同じ塩基のシーケンシングがどのようにして可能になるかの模式図である。

【図17】図17は、標的核酸上の1つの具体的なヌクレオチド位置における複数の塩基読み取り (base call) が1つ以上のシーケンシング用プローブから記録され、どのように組み合わせられてコンセンサス配列を作り出し、そのことによって最終的な塩基読み取りの精度を上昇させているかを示している。

10

【図18-1】図18は、本開示のシーケンシング法を用いて取得してAssembly Algorithmを利用して分析した、シーケンシング実験からの結果を示している。グラフに関しては、示されている配列は、左上のグラフから始まって時計回りに配列番号3、4、6、8、7、5に対応している。右側の表に関しては、配列は、上から下へと配列番号3、4、7、8、6、5に対応している。

【図18-2】図18は、本開示のシーケンシング法を用いて取得してAssembly Algorithmを利用して分析した、シーケンシング実験からの結果を示している。グラフに関しては、示されている配列は、左上のグラフから始まって時計回りに配列番号3、4、6、8、7、5に対応している。右側の表に関しては、配列は、上から下へと配列番号3、4、7、8、6、5に対応している。

20

【図19】図19は、FFPEサンプルからがん遺伝子標的を多重捕捉してシーケンシングするための実験設計の模式図を示している。

【図20】図20は、直接的なRNAシーケンシングの模式図と、本開示のシーケンシング法を利用してRNA分子の適合性を調べる実験からの結果を示している。

【図21】図21は、本開示のシーケンシング法を利用して同じヌクレオチド配列を有するRNA分子とDNA分子をシーケンシングした結果を示している。

【図22】図22は、本開示の標準シーケンシング用プローブと3部構成シーケンシング用プローブの性能の比較を示している。

30

【図23】図23は、個別のプローブを用いた本開示の代表的な標的結合ドメイン内のLNA置換の効果を示している。

【図24】図24は、9個のプローブからなるプールを用いたときの本開示の代表的な標的結合ドメインの中でのLNA置換の効果を示している。

【図25】図25は、本開示の代表的な標的結合ドメインの中での修飾されたヌクレオチドと核酸類似体による置換の効果を示している。

【図26】図26は、本開示のシーケンシング法の生の精度を定量する実験からの結果を示している。

【図27】図27は、標的核酸内のヌクレオチドを2個以上のシーケンシング用プローブによってシーケンシングするときの本開示のシーケンシング法の精度を求める実験からの結果を示している。

40

【図28】図28は、ポケットオリゴを含む本開示のシーケンシング用プローブの模式図である。

【図29】図29は、各付着位置の間にPEGリンカー領域を含む本開示のシーケンシング用プローブの模式図である。

【図30】図30は、各付着位置の間に非塩基性領域を含む本開示のシーケンシング用プローブの模式図である。

【図31】図31は、本開示の代表的なシーケンシング用プローブにコネクタオリゴを通じて間接的にハイブリダイズした本開示の代表的なレポータ複合体の図である。

【図32】図32は、本開示の方法で用いられるパリティスキームの説明図である。

50

【図 3 3】図 33 は、捕捉用プローブとアダプタオリゴヌクレオチドとローン（lawn）オリゴヌクレオチドからなる本発明の複合体の模式図である。

【図 3 4】図 34 は、標的核酸にハイブリダイズする本開示の c5 プローブ複合体と c3 プローブ複合体の模式図である。

【図 3 5】図 35 は、FEN1 を用いて消化させた後の本開示の標的核酸-c3 プローブ-c5 プローブ複合体の模式図である。

【図 3 6】図 36 は、連結後の本開示の標的核酸-c3 プローブ-c5 プローブ複合体の模式図である。

【図 3 7】図 37 は、本開示の標的核酸-c3 プローブ-c5 プローブ複合体を、利用者を媒介として切断する模式図である。

10

【図 3 8】図 38 は、USER によって切断した後の本開示の標的核酸-c3 プローブ-c5 プローブ複合体の模式図である。

【図 3 9】図 39 は、本開示の標的核酸-c3 プローブ-c5 プローブ複合体を UV によって切断する模式図である。

【図 4 0】図 40 は、相補的核酸を介して基板に付着させた本開示の標的核酸-c3 プローブ-c5 プローブ複合体を UV によって切断した後の模式図である。

【図 4 1】図 41 は、標的核酸にハイブリダイズした本開示の c3.2 プローブ複合体と c5.2 プローブ複合体の模式図である。

【図 4 2】図 42 は、c3.2 プローブ複合体と c5.2 プローブ複合体を連結させた後の本開示の標的核酸複合体の模式図である。

20

【図 4 3】図 43 は、本開示の核酸複体内の一本鎖精製配列の切断と放出の模式図である。

【図 4 4】図 44 は、本開示の基板の上に固定化された本開示の標的核酸複合体の模式図である。

【図 4 5】図 45 は、本開示の標的核酸複合体を本開示の基板に固定化した後の基板特異的ドメインの切断と放出の模式図である。

【図 4 6】図 46 は、本開示の基板に固定化された基板特異的ドメインを放出した後の本開示の標的核酸複合体の模式図である。

【図 4 7】図 47 は、本発明の代表的なアレイの断面の模式図である。

【図 4 8】図 48 は、ピラミッド形のナノウェルを含む本発明の代表的なアレイの断面の模式図である。

30

【図 4 9】図 49 は、ランダムなパターンに配置された複数の円筒形ナノウェルを含む本開示の代表的なアレイの模式図である。

【図 5 0】図 50 は、一定ピッチの規則正しいグリッドの中に配置された円筒形ナノウェルを含む本開示の代表的なアレイの模式図である。

【図 5 1】図 51 は、本開示の代表的なアレイの模式図であり、単一の標的核酸複合体が各ナノウェルの中に固定化されている。

【図 5 2】図 52 は、本開示の代表的なアレイの模式図であり、単一の標的核酸複合体が各ナノウェルの中に固定化されていることで、他の標的核酸複合体が固定化されるのを阻止している。

40

【図 5 3】図 53 は、全面的に L-DNA からなり、3' 末端 L-dG ヌクレオチドを有する付着領域を含む本開示のシーケンシング用プローブの模式図である。

【図 5 4】図 54 は、全面的に D-DNA からなり、付着領域 1（スポット 1）と付着領域 2（スポット 2）の間と、付着領域 2（スポット 2）と付着領域 3（スポット 3）の間にポケットオリゴを含む本開示のシーケンシング用プローブの模式図である。

【図 5 5】図 55 は、捕捉用プローブとローンオリゴヌクレオチドをタンパク質ロックと組み合わせ用いて固体基板の表面に固定化された合成標的核酸の模式図である。

【図 5 6】図 56 は、本開示の LG 介在シーケンシング用プローブと D-ポケットシーケンシング用プローブを用いたシーケンシング実験の結果を示す一連のチャートである。x 軸は、シーケンシングしている標的核酸の具体的なヌクレオチドを表わす。最上部のチ

50

ャートは、LG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブに関する理論的なシークエンシング多様性と、観察されたシークエンシング多様性と、観察されたシークエンシングカバレッジを示している。赤色の箱は、シークエンシングに関して予想される問題領域を示す。

【図57】図57は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。x軸は、シークエンシングしている標的核酸の具体的なヌクレオチドを表わす。最上部のチャートは、LG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブに関する理論的なシークエンシング多様性と、観察されたシークエンシング多様性と、観察されたシークエンシングカバレッジを示している。赤色の箱は、シークエンシングに関して予想される問題領域を示す。

10

【図58】図58は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。x軸は、シークエンシングしている標的核酸の具体的なヌクレオチドを表わす。最上部のチャートは、LG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブに関する理論的なシークエンシング多様性と、観察されたシークエンシング多様性と、観察されたシークエンシングカバレッジを示している。赤色の箱は、シークエンシングに関して予想される問題領域を示す。

【図59】図59は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。x軸は、シークエンシングしている標的核酸の具体的なヌクレオチドを表わす。最上部のチャートは、LG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブに関する観察されたシークエンシング多様性と、観察されたシークエンシングカバレッジを示している。

20

【図60】図60は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。x軸は、シークエンシングしている標的核酸の具体的なヌクレオチドを表わす。最上部のチャートは、LG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブに関する観察されたシークエンシング多様性と、観察されたシークエンシングカバレッジを示している。

30

【図61】図61は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。x軸は、シークエンシングしている標的核酸の具体的なヌクレオチドを表わす。最上部のチャートは、LG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブに関する観察されたシークエンシング多様性と、観察されたシークエンシングカバレッジを示している。

【図62】図62は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験におけるバーコードイベントの総数と有効な3スポット読み出しの数を示す一連のヒストグラムである。

【図63】図63は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験におけるオンターゲットイベント、無効なイベント、オフターゲットイベント、b₁-b₆で1エラーのイベント、b₁-b₆で2エラーのイベント、b₁-b₆で3エラーのイベント、b₁-b₆で4エラーのイベント、b₁-b₆で5エラーのイベント、b₁-b₆で6エラーのイベントそれぞれの総数を示す一連のグラフである。

40

【図64-1】図64は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験におけるオンターゲットイベント、無効なイベント、オフターゲットイベント、b₁-b₆で1エラーのイベント、b₁-b₆で2エラーのイベント、b₁-b₆で3エラーのイベント、b₁-b₆で4エラーのイベント、b₁-b₆で5エラーのイベント、b₁-b₆で6エラーのイベントそれぞれの総数を示す一連のグラフである。

【図64-2】図64は、本開示のLG介在シークエンシング用プローブとD-ポケットシー

50

クエンシング用プローブを用いたシークエンシング実験におけるオンターゲットイベント、無効なイベント、オフターゲットイベント、 b_1 - b_6 で1エラーのイベント、 b_1 - b_6 で2エラーのイベント、 b_1 - b_6 で3エラーのイベント、 b_1 - b_6 で4エラーのイベント、 b_1 - b_6 で5エラーのイベント、 b_1 - b_6 で6エラーのイベントそれぞれの総数を示す一連のグラフである。

【図64-3】図64は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験におけるオンターゲットイベント、無効なイベント、オフターゲットイベント、 b_1 - b_6 で1エラーのイベント、 b_1 - b_6 で2エラーのイベント、 b_1 - b_6 で3エラーのイベント、 b_1 - b_6 で4エラーのイベント、 b_1 - b_6 で5エラーのイベント、 b_1 - b_6 で6エラーのイベントそれぞれの総数を示す一連のグラフである。

【図65】図65は、本開示のD-ポケットシークエンシング用プローブ（サイクル1～50）とLG介在シークエンシング用プローブ（サイクル51～100）を用いたシークエンシング実験の各サイクルにおける（可能な3つのレポータプローブのうちの1つだけがうまく記録された）1スポットイベント、（可能な3つのレポータプローブのうちの2つだけがうまく記録された）2スポットイベント、（可能な3つのレポータプローブのすべてがうまく記録された）3スポットイベントを示すチャートである。

【図66-1】図66は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。最も左側の図は、シークエンシング実験の各サイクルにおけるオンターゲットイベント、新規6量体イベント、冗長6量体イベント、オフターゲットイベント、無効なイベントの数を示している。サイクル1～50は本開示のD-ポケットシークエンシング用プローブを用いて実施し、サイクル51～100は本開示のLG介在シークエンシング用プローブを用いて実施した。

【図66-2】図66は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。最も左側の図は、シークエンシング実験の各サイクルにおけるオンターゲットイベント、新規6量体イベント、冗長6量体イベント、オフターゲットイベント、無効なイベントの数を示している。サイクル1～50は本開示のD-ポケットシークエンシング用プローブを用いて実施し、サイクル51～100は本開示のLG介在シークエンシング用プローブを用いて実施した。

【図66-3】図66は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。最も左側の図は、シークエンシング実験の各サイクルにおけるオンターゲットイベント、新規6量体イベント、冗長6量体イベント、オフターゲットイベント、無効なイベントの数を示している。サイクル1～50は本開示のD-ポケットシークエンシング用プローブを用いて実施し、サイクル51～100は本開示のLG介在シークエンシング用プローブを用いて実施した。

【図66-4】図66は、本開示のLG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブを用いたシークエンシング実験の結果を示す一連のチャートである。最も左側の図は、シークエンシング実験の各サイクルにおけるオンターゲットイベント、新規6量体イベント、冗長6量体イベント、オフターゲットイベント、無効なイベントの数を示している。サイクル1～50は本開示のD-ポケットシークエンシング用プローブを用いて実施し、サイクル51～100は本開示のLG介在シークエンシング用プローブを用いて実施した。

【図67】図67は、本開示の方法と組成物を用いて固体基板に固定化された標的核酸の模式図である。標的核酸は、捕捉用プローブ上に位置するビオチン部分とローンオリゴヌクレオチドとニュートラアビジン部分の間のタンパク質ロックを利用して固定化される。

【発明を実施するための形態】

【0043】

本開示により、酵素なし、増幅なし、ライブラリなしで迅速に実施され、読み出し長が長くてエラー率が低い核酸シークエンシングを提供するシークエンシング用プローブと、

10

20

30

40

50

レポータープローブと、方法と、キットと、装置が提供される。

【0044】

本開示の組成物

【0045】

本開示により、標的結合ドメインとバーコードドメインを含むシークエンシング用プローブとして、標的結合ドメインが表1に記載されている任意のコンストラクトを含むシークエンシング用プローブが提供される。代表的な標的結合ドメインは少なくとも8個のヌクレオチドを含んでいて標的核酸に結合することができ、標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたはヌクレオチド類似体であることが可能であり、標的結合ドメイン内にあって標的核酸分子内の対応するヌクレオチドを同定しない上記少なくとも2個のヌクレオチドは、標的結合ドメイン内の上記少なくとも6個のヌクレオチドによって指定される標的に特異的でない4種類の正規塩基、またはユニバーサル塩基、または縮重塩基のうちの任意のものが可能である。代表的なバーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子が結合することのできる少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含み、上記少なくとも3つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置のそれぞれの位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドのうちの対応する2個のヌクレオチドの位置と属性を決めている。

【0046】

別の態様では、代表的な標的結合ドメインは、標的核酸にハイブリダイズすることのできる少なくとも6個のヌクレオチドを含むことができ、標的結合ドメイン内のこれら少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ；標的結合ドメイン内のこれら少なくとも6個のヌクレオチドの任意のヌクレオチドは、改変されたヌクレオチドまたはヌクレオチド類似体であることが可能である。

【0047】

本開示により、標的結合ドメインとバーコードドメインを含むシークエンシング用プローブとして；標的結合ドメインは少なくとも10個のヌクレオチドを含んでいて標的核酸に結合することができ、この標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも4個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子が結合することのできる少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含み、上記少なくとも3つの付着位置の各付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置の各付着位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドの対応する2個のヌクレオチドの位置と属性を決めているプローブも提供される。

【0048】

本開示により、本明細書に開示されている任意のシークエンシング用プローブを複数含むシークエンシング用プローブの集団も提供される。

【0049】

開示されているシークエンシング用プローブの標的結合ドメイン、バーコードドメイン

、骨格のほか、相補的核酸分子（例えばレポータ分子またはレポータ複合体）は、下により詳細に記載してある。

【0050】

本開示のシーケンシング用プローブは、標的ドメインとバーコードドメインを含んでいる。図1は、本開示の代表的な1つのシーケンシング用プローブの模式図である。図1は、標的ドメインが標的核酸に結合できることを示している。標的核酸として、本開示のシーケンシング用プローブがハイブリダイズすることのできる任意の核酸が可能である。標的核酸として、DNAまたはRNAが可能である。標的核酸は、対象に由来する生物サンプルから得ることができる。「標的結合ドメイン」と「シーケンシングドメイン」という用語は、本明細書では交換可能に用いられる。

10

【0051】

標的結合ドメインは、一連のヌクレオチドを含むことができる（例えばポリヌクレオチドである）。標的結合ドメインは、DNA、またはRNA、またはこれらの組み合わせを含むことができる。標的結合ドメインがポリヌクレオチドである場合には、標的結合ドメインは、図1に示してあるように、標的核酸のうちでシーケンシング用プローブの標的結合ドメインと相補的な部分にハイブリダイズすることにより、その標的核酸に結合する。

【0052】

シーケンシング用プローブの標的結合ドメインは、シーケンシング用プローブのハイブリダイゼーションおよび／または脱ハイブリダイゼーションの可能性と、それが起こる速度を制御できるように設計することができる。一般に、プローブの T_m が低くなるほど、プローブが標的核酸からより速く、より大きな確率で脱ハイブリダイズする。したがって T_m がより低いプローブを用いると、標的核酸に結合するプローブの数が少なくなろう。

20

【0053】

標的結合ドメインの長さが、プローブが標的核酸にハイブリダイズする可能性と標的核酸にハイブリダイズしたままになる可能性に部分的に影響を与える。一般に、標的結合ドメインが長くなるほど（ヌクレオチドの数が増えるほど）、相補的配列が標的ヌクレオチドの中に存在する可能性が小さくなる。逆に、標的結合ドメインが短くなるほど、相補的配列が標的ヌクレオチドの中に存在する可能性が大きくなる。例えば4量体配列が標的核酸の中に位置する確率は $1/256$ であるのに対し、6量体配列が標的核酸の中に位置する確率は $1/4096$ である。その帰結として、より短いプローブの集合は、より長いプローブの集合と比べたとき、所与の長さの核酸に対してより多くの位置に結合する可能性がある。

30

【0054】

複数の状況において、例えば変異またはSNPアレルを検出するとき、プローブがより短い標的結合ドメインを有するようにして所与の長さの核酸内のリードの数を増やすことにより、標的核酸、または標的核酸の一部（特に目的の部分）のカバレッジを大きくすることが好ましい。

【0055】

標的結合ドメインとして、長さが任意の量または数のヌクレオチドが可能である。標的結合ドメインとして、長さが少なくとも12個のヌクレオチド、長さが少なくとも10個のヌクレオチド、長さが少なくとも8個のヌクレオチド、長さが少なくとも6個のヌクレオチド、長さが少なくとも3個のヌクレオチドのいずれかが可能である。

40

【0056】

標的結合ドメイン内の各ヌクレオチドは、標的分子の相補的なヌクレオチドを同定（またはコード）することができる。あるいは標的結合ドメイン内のいくつかのヌクレオチドは、標的分子の相補的なヌクレオチドを同定（またはコード）することができ、標的結合ドメイン内のいくつかのヌクレオチドは、標的分子の相補的なヌクレオチドを同定（またはコード）しない。

【0057】

標的結合ドメインは、少なくとも1個の天然塩基を含むことができる。標的結合ドメイ

50

ンは、天然塩基を含まないことが可能である。標的結合ドメインは、少なくとも1個の修飾されたヌクレオチドまたは核酸類似体を含むことができる。標的結合ドメインは、修飾されたヌクレオチドまたは核酸類似体を含まないことが可能である。標的結合ドメインは、少なくとも1個のユニバーサル塩基を含むことができる。標的結合ドメインは、ユニバーサル塩基を含まないことが可能である。標的結合ドメインは、少なくとも1個の縮重塩基を含むことができる。標的結合ドメインは、縮重塩基を含まないことが可能である。

【0058】

標的結合ドメインは、天然塩基（例えば0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはそれよりも多い個数の天然塩基）、修飾されたヌクレオチドまたは核酸類似体（例えば0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはそれよりも多い個数の修飾されたヌクレオチドまたは核酸類似体）、ユニバーサル塩基（例えば0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはより多くのユニバーサル塩基）、縮重塩基（例えば0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはそれよりも多い個数の縮重塩基）の任意の組み合わせを含むことができる。個々の標的結合ドメインの天然塩基、修飾されたヌクレオチドまたは核酸類似体、ユニバーサル塩基、縮重塩基は、組み合わせで存在するとき、任意の順番で配置することができる。

【0059】

「修飾されたヌクレオチド」または「核酸類似体」という用語の非限定的な例に含まれるのは、ロックド核酸（locked nucleic acids）（LNA）、架橋型核酸（bridged nucleic acids）（BNA）、プロピンで修飾された核酸、ジップ核酸（ZNA（登録商標））、イソグアニン、イソシトシン、6-アミノ-1-(4-ヒドロキシ-5-ヒドロキシメチル-テトラヒドロフラン-2-イル)-1,5-ジヒドロ-ピラゾロ[3,4-d]ピリミジン-4-オン（PPG）、2'が修飾された核酸（2'-O-メチル核酸など）である。標的結合ドメインは、0個～6個（例えば0個、1個、2個、3個、4個、5個、6個）の修飾されたヌクレオチドまたは核酸類似体を含むことができる。修飾されたヌクレオチドまたは核酸類似体は、ロックド核酸（LNA）であることが好ましい。

【0060】

本明細書では、「ロックド核酸（locked nucleic acids）（LNA）」という用語の非限定的な例に含まれるのは、2'酸素と4'炭素を接続するメチレン架橋をリボース部分が含まれている修飾されたRNAヌクレオチドである。このメチレン架橋が、リボースをA型RNA二本鎖で見られる3'エンドコンホメーション（ノースコンホメーションとしても知られる）の中にロックする。アクセス不能なRNAという用語は、LNAと交換可能に用いることができる。本明細書では、「架橋化核酸（BNA）」という表現の非限定的な例に含まれるのは、固定された3'エンドコンホメーション（ノースコンホメーションとしても知られる）を有する5員または6員の架橋構造を含む修飾されたRNA分子である。架橋構造は、リボースの2'酸素をリボースの4'炭素に接続している。炭素原子、窒素原子、水素原子を含む異なるさまざまな架橋構造が可能である。本明細書では、「プロピンで修飾された核酸」という用語の非限定的な例に含まれるのは、核酸塩基のC5位にプロピン修飾を含むピリミジン類、すなわちシトシンとチミン／ウラシルである。本明細書では、「ジップ核酸（ZNA（登録商標））」という用語の非限定的な例に含まれるのは、カチオン性スベルミン部分に共役したオリゴヌクレオチドである。

【0061】

本明細書では、「ユニバーサル塩基」という用語の非限定的な例に含まれるのは、ワトソン-クリック塩基対合規則に従わないが、標的核酸の上に位置する4種類の正規塩基（A、T/U、C、G）のうちの任意のものに結合することのできるヌクレオチド塩基である。本明細書では、「縮重塩基」という用語の非限定的な例に含まれるのは、ワトソン-クリック塩基対合規則に従わないが、標的核酸の上に位置する4種類の正規塩基（A、T/U、C、G）の4種類全部ではなくてそのうちの少なくとも2種類に結合することのできるヌクレオチド塩基である。縮重塩基は、ゆらぎ塩基と呼ぶこともでき、これらの用語は本明細書では

交換可能に用いられる。

【0062】

図1に示した代表的なシーケンシング用プローブは、シーケンシングすべき標的核酸の相補的なヌクレオチド1～6に特異的にハイブリダイズする長さ6個のヌクレオチド（6量体）配列（ b_1 - b_2 - b_3 - b_4 - b_5 - b_6 ）を含む標的結合ドメインを示している。標的結合ドメインのこの6量体部分（ b_1 - b_2 - b_3 - b_4 - b_5 - b_6 ）が、標的配列内の相補的なヌクレオチド（1-2-3-4-5-6）を同定（またはコード）する。この6量体配列は、それぞれの側に塩基（N）が隣接している。（N）によって示される塩基として、独立に、ユニバーサル塩基または縮重塩基が可能である。典型的には、（N）によって示される塩基は、独立に、正規塩基のうちの1つである。（N）によって示される塩基は、標的配列の中でその塩基が結合する相補的なヌクレオチドを同定（またはコード）することがなく、（6量体）配列（ b_1 - b_2 - b_3 - b_4 - b_5 - b_6 ）の核酸配列とは独立である。

10

【0063】

図1に示したシーケンシング用プローブは、本開示のシーケンシング法と組み合わせて用いて標的核酸をシーケンシングすることができ、ハイブリダイゼーション反応だけを利用して、共有結合化学、酵素、増幅を必要としない。標的核酸分子の中の可能なすべての6量体配列をシーケンシングするには合計で4096通りのシーケンシング用プローブが必要とされる（ $4^6=4096$ ）。

【0064】

図1は、本開示によるシーケンシング用プローブの標的結合ドメインの構成の一例である。表1に、本開示の標的結合ドメインの別のいくつかの構成を示してある。好ましい1つの標的結合ドメインは「6 LNA」標的結合ドメインと呼ばれ、標的結合ドメインの位置 b_1 - b_6 に6個のLNAを含んでいる。これら6個のLNAはそれぞれの側に塩基（N）が隣接している。本明細書では、塩基（N）として、（6量体）配列（ b_1 - b_2 - b_3 - b_4 - b_5 - b_6 ）の核酸配列とは独立なユニバーサル塩基／縮重塩基、または正規塩基が可能である。言い換えると、塩基 b_1 - b_2 - b_3 - b_4 - b_5 - b_6 は所与の任意の標的配列に対して特異的である可能性があるのに対し、（N）塩基は、ユニバーサル塩基／縮重塩基であること、または4種類の正規塩基のうちで、 b_1 - b_2 - b_3 - b_4 - b_5 - b_6 によって指定される標的に対して特異的ではない任意のもので構成することができる。例えば調べる標的配列がCAGGCATAである場合には、標的結合ドメインの塩基 b_1 - b_2 - b_3 - b_4 - b_5 - b_6 はTCCGTAになると考えられるのに対し、標的結合ドメインの（N）塩基のそれぞれは、独立に、A、C、T、Gのいずれかになることができると考えられるため、得られる標的結合ドメインは、配列ATCCGTAG、TTCCGTAC、GTCCGTAGのいずれか、または可能な16通りの配列のうちの他のどれかになることができる。あるいは2個の（N）塩基が6 LNAの前に位置することができると考えられる。あるいはさらに、2個の（N）塩基が6 LNAの後ろに位置することができると考えられる。

20

30

【0065】

40

50

【表 1】

表1

標的結合ドメインの塩基			B1	B2	B3	B4	B5	B6		
「6 量体」			b	b	b	b	b	b		
「8 量体」		b	b	b	b	b	b	b	b	
「10 量体」	b	b	b	b	b	b	b	b	b	b
「Natural I」	N	N	b	b	b	b	b	b	N	N
「Natural II」		N	b	b	b	b	b	b	N	
「2LNA」		N	b	b	+	+	b	b	N	
		N	b	+	b	b	+	b	N	
		N	+	b	b	b	b	+	N	
「4LNA」		N	+	+	b	b	+	+	N	
		N	+	b	+	+	b	+	N	
		N	b	+	+	+	+	b	N	
「6LNA」			+	+	+	+	+	+		
		N	+	+	+	+	+	+	N	
「LNA を含む 8 量体」		N	b/+	b/+	b/+	b/+	b/+	b/+	N	
「MGB」		Q	b	b	b	b	b	b		
	Q	b	b	b	b	b	b	b	b	

b=天然塩基；+=修飾されたヌクレオチド、またはヌクレオチド類似体（例えばLNA、2'-O'-メチル-修飾塩基、6-アミノ-1-(4-ヒドロキシ-5-ヒドロキシメチル-テトラヒドロフラン-2-イル)-1,5-ジヒドロピラゾロ[3,4-d]ピリミジン-4-オン（PPG））；N=天然塩基、ユニバーサル塩基、縮重塩基；Qは副溝バインダ（例えば振じれたインターカレーティング核酸、MGB-BP3、プロスタリシン）

【0066】

表1には、10個の標的的特異的な天然塩基を含む「10量体」標的結合ドメインも記載されている。表1には、8個の標的的特異的な天然塩基を含む「8量体」標的結合ドメインも記載されている。

【0067】

表1にはさらに、位置b₁-b₆に6個の天然塩基を含む「Natural I」標的結合ドメインが記載されている。これら6個の天然塩基のそれぞれの側に2個の（N）塩基が隣接している。あるいは4個の（N）塩基すべてがそれら6個の天然塩基の前に位置することができよう。あるいはさらに、4個の（N）塩基すべてがそれら6個の天然塩基の後ろに位置することができよう。4個の（N）塩基のうちの任意の個数（すなわち1個、2個、3個、4個）をそれら6個の天然塩基の前に位置させることができると、残りの（N）塩基はそれら6個の天然塩基の後ろに位置することになる。

【0068】

表1にはさらに、位置b₁-b₆に6個の天然塩基を含む「Natural II」標的結合ドメインが記載されている。これら6個の天然塩基のそれぞれの側に1個の（N）塩基が隣接している。あるいは両方の（N）塩基を6個の天然塩基の前に位置させることができよう。あるいはさらに、両方の（N）塩基を6個の天然塩基の後ろに位置させることができよう。典型的には、Natural II標的結合ドメインの（N）塩基は縮重塩基である。

【0069】

表1には、標的結合ドメインの位置b₁-b₆に2個のLNAと4個の天然塩基の組み合わせを含む「2LNA」標的結合ドメインも記載されている。2個のLNAと4個の天然塩基は任意の

順番で登場することができる。例えば位置 b_3 と b_4 をLNAにすることができる一方で、位置 b_1 、 b_2 、 b_5 、 b_6 は天然塩基である。塩基 b_1 - b_6 のそれぞれの側に1個の(N)塩基が隣接している。あるいは塩基 b_1 - b_6 の前に2個の(N)塩基が位置することができる。あるいはさらに、塩基 b_1 - b_6 の後に2個の(N)塩基が続くことができる。

【0070】

表1にはさらに、標的結合ドメインの位置 b_1 - b_6 に4個のLNAと2個の天然塩基の組み合わせを含む「4LNA」標的結合ドメインが記載されている。4個のLNAと2個の天然塩基は任意の順番で登場することができる。例えば位置 b_2 - b_5 をLNAにすることができる一方で、位置 b_1 と b_6 は天然塩基である。塩基 b_1 - b_6 のそれぞれの側に1個の(N)塩基が隣接している。あるいは塩基 b_1 - b_6 の前に2個の(N)塩基が位置することができる。あるいはさらに、塩基 b_1 - b_6 の後に2個の(N)塩基が続くことができる。

10

【0071】

表1にはさらに、標的結合ドメインの位置 b_1 - b_6 に6個のLNAを含む「6LNA」標的結合ドメインが記載されている。塩基 b_1 - b_6 のそれぞれの側に1個の(N)塩基が隣接することができる。

【0072】

表1にはさらに、標的結合ドメインの位置 b_1 - b_6 の任意の位置に天然塩基またはLNAを個別に含む「8量体」標的結合ドメインが記載されている。塩基 b_1 - b_6 のそれぞれの側に1個の(N)塩基が隣接することができる。

【0073】

標的結合ドメインは、副溝バインダ(minor-groove binder)部分も含むことができる。副溝バインダ部分はオリゴヌクレオチドの化学修飾であり、オリゴヌクレオチドがハイブリダイズする標的ヌクレオチドの副溝に結合することのできる化学的部分を付加する。理論に囚われなければ、副溝バインダ部分を含めることで、標的核酸に対する標的結合ドメインの親和性が増大し、標的結合ドメイン-標的核酸二本鎖の融解温度が上昇する。結合親和性が大きいほど、より小さな標的結合ドメインを用いることが可能になる。

20

【0074】

標的結合ドメインは、1つ以上の捩じれたインターカレーティング核酸(twisted intercalating nucleic acid)(TINA)も含むことができる。TINAは、二本鎖オリゴヌクレオチドと三本鎖形成オリゴヌクレオチドからのHoogsteen三本鎖DNAの形成を安定化させる核酸分子である。TINAを用いて二本鎖オリゴヌクレオチドを安定化させることで、標的核酸に対するオリゴヌクレオチドプローブの特異度と感度を改善することができる。

30

【0075】

標的結合ドメインは、2'-O-メチル修飾塩基を含む核酸分子も含むことができる。2'-O-メチル修飾塩基はRNAのヌクレオシド修飾であり、メチル基がリボースの2'ヒドロキシル基に付加されて2'メトキシ基を生成させる。2'-O-メチル修飾塩基は、塩基加水分解と、ヌクレアーゼによる消化に対する優れた保護を提供する。理論に囚われないと、2'-O-メチル修飾塩基の付加により、核酸二本鎖の融解温度も上昇する。

【0076】

標的結合ドメインは、共有結合したスチルベン修飾も含むことができる。スチルベン修飾は核酸二本鎖の安定性を増大させることができる。

40

【0077】

本開示のシーケンシング用プローブは合成骨格を含んでいる。標的結合ドメイン(本明細書ではシーケンシングドメインとしても記載されている)とバーコードドメインは、機能可能に連結されている。標的結合ドメインとバーコードドメインを共有結合させて1つの合成骨格の一部にすることができる。標的結合ドメインとバーコードドメインは、リンカー(例えば核酸リンカー、化学的リンカー)を介して結合させることができる。合成骨格は、任意の材料(例えば多糖、ポリヌクレオチド、ポリマー、プラスチック、繊維、ペプチド、ペプチド核酸、ポリペプチド)を含むことができる。合成骨格は堅固であることが好ましい。合成骨格は一本鎖DNA分子を含むことができる。骨格は、6つのDNA二

50

重螺旋からなる「DNA折り紙」を含むことができる（例えばLin他、「DNAから自己組織化したミクロン未満の幾何学的にコードされた蛍光バーコード」Nature Chemistry；2012年10月；第4巻(10)：832～839ページを参照されたい）。バーコードは、DNA折り紙タイルから作製することができる（Jungmann他、「DNA-PAINTとExchange-PAINTを用いた多重化3D細胞超解像イメージング」、Nature Methods、第11巻、第3号、2014年）。

【0078】

本開示のシーケンシング用プローブは、部分的に二本鎖の合成骨格を含むことができる。シーケンシング用プローブは、一本鎖DNA合成骨格と、標的結合ドメインとバーコードドメインの間の二本鎖DNAスペーサを含むことができる。二本鎖DNAスペーサは、少なくとも1つの修飾されたヌクレオチドまたは核酸類似体を含むことができる。二本鎖DNAスペーサで有用な典型的な修飾されたヌクレオチドまたは核酸類似体は、イソグアニンとイソシトシンである。あるいはさらに、二本鎖DNAスペーサを含む核酸のそれぞれとして、独立に、L-DNAが可能である。いくつかの態様では、二本鎖DNAスペーサはL-DNAを含むことができる。二本鎖DNAスペーサはL-DNAからなることができる。二本鎖DNAスペーサは、主にL-DNAからなることができる。

【0079】

二本鎖DNAスペーサは、長さが約1個～約100個のヌクレオチドを含むことができる。二本鎖DNAスペーサは、長さが約25個のヌクレオチドを含むことができる。

【0080】

合成骨格はL-DNAを含むことができる。合成骨格は、L-DNAからなることができる。合成骨格は、主にL-DNAからなることができる。一本鎖DNA合成骨格は、長さが約10個～約100個のヌクレオチドを含むことができる。一本鎖DNA合成骨格は、長さが約52個のヌクレオチドを含むことができる。一本鎖DNA合成骨格は、長さが約27個のヌクレオチドを含むことができる。

【0081】

バーコードドメインはL-DNAを含むことができる。バーコードドメインは、L-DNAからなることができる。バーコードドメインは、主にL-DNAからなることができる。バーコードドメインは、約27個のヌクレオチド、または約52個のヌクレオチド、または約99個のヌクレオチド、または約74個のヌクレオチドを含むことができる。バーコードドメインは、長さが約27個のヌクレオチド、または約52個のヌクレオチド、または約99個のヌクレオチド、または約74個のヌクレオチドを含むことができる。

【0082】

シーケンシング用プローブは、一本鎖DNA合成骨格と、標的結合ドメインとバーコードドメインの間にあって二本鎖DNAと同様の機械的特性を有するポリマー系スペーサを含むことができる。典型的なポリマー系スペーサは、ポリエチレングリコール（PEG）型ポリマーを含んでいる。

【0083】

二本鎖DNAスペーサとして、長さが約1個～約100個のヌクレオチド、または長さが約2個～約50個のヌクレオチド、または長さが約20個～約40個のヌクレオチドが可能である。二本鎖DNAスペーサは、長さが約36個のヌクレオチドであることが好ましい。

【0084】

本開示の1つのシーケンシング用プローブを「標準プローブ」と名づけ、図2の左図に示してある。図2の標準プローブは、標的結合ドメインに共有結合したバーコードドメインを含んでいるため、標的結合ドメインとバーコードドメインは、同じ一本鎖オリゴヌクレオチドの中に存在する。図2の左図では、一本鎖オリゴヌクレオチドはステムオリゴヌクレオチドに結合してステム配列と呼ばれる長さ36個のヌクレオチドの二本鎖スペーサ領域を生み出している。プローブのプールの中の各シーケンシング用プローブは、このアーキテクチャを利用して同じステム配列にハイブリダイズすることができる。

【0085】

別の態様では、バーコードドメインと、標準プローブのステムオリゴヌクレオチドに結合する領域に含まれるそれぞれの核酸として、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。バーコードドメインと、標準プローブのステムオリゴヌクレオチドに結合する領域で有用な典型的な修飾されたヌクレオチドまたは核酸類似体は、イソグアニンとイソシトシンである。あるいはさらに、バーコードドメインと、標準プローブのステムオリゴヌクレオチドに結合する領域に含まれるそれぞれの核酸として、L-DNAが可能である。例えばバーコードドメインと、標準プローブのステムオリゴヌクレオチドに結合する領域は、全面的にL-DNAで構成することができる。別の例では、バーコードドメインと、標準プローブのステムオリゴヌクレオチドに結合する領域は、非塩基性である一本鎖核酸の区画、または二本鎖DNAと機械的特性が似たポリマー（下により詳しく記載するPEGなど）の区画によって隔てられたL-DNAの複数の区画で構成することができる。

10

【0086】

本開示の別のシーケンシング用プローブを「3部構成プローブ」と名づけ、図2の中央の図に示してある。図2の3部構成プローブは、リンカーを介して標的結合ドメインに付着したバーコードドメインを含んでいる。この例では、リンカーは、標的結合ドメインを含有する一本鎖オリゴヌクレオチドと、バーコードドメインを含有する一本鎖オリゴヌクレオチドとにハイブリダイズする一本鎖ステムオリゴヌクレオチドであり、バーコードドメイン（18個のヌクレオチド）と標的結合ドメイン（18個のヌクレオチド）を架橋する長さ36個のヌクレオチドの二本鎖スペーサ領域を作り出している。この代表的なプローブの構成を利用し、バーコードドメインの交換を阻止することを目的として、各バーコードを、1つだけのステム配列にハイブリダイズするように設計することができる。さらに、各バーコードドメインを対応するステムオリゴヌクレオチドにハイブリダイズさせた後に、異なるシーケンシング用プローブをまとめてプールすることもできる。

20

【0087】

別の態様では、一本鎖ステムオリゴヌクレオチドに含まれるそれぞれの核酸として、正規核酸、または修飾されたヌクレオチド、または核酸類似体が可能である。一本鎖ステムオリゴヌクレオチドで有用な典型的な修飾されたヌクレオチドまたは核酸類似体は、イソグアニンとイソシトシンである。あるいはさらに、一本鎖ステムオリゴヌクレオチドに含まれるそれぞれの核酸として、独立に、L-DNAが可能である。

30

【0088】

別の態様では、一本鎖ステムオリゴヌクレオチドがハイブリダイズするバーコードドメイン上の領域に含まれるそれぞれの核酸として、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。一本鎖ステムオリゴヌクレオチドで有用な典型的な修飾されたヌクレオチドまたは核酸類似体は、イソグアニンとイソシトシンである。あるいはさらに、一本鎖ステムオリゴヌクレオチドがハイブリダイズするバーコードドメイン上の領域に含まれるそれぞれの核酸として、独立に、L-DNAが可能である。

【0089】

別の態様では、一本鎖ステムオリゴヌクレオチドがハイブリダイズする標的結合ドメインを含有する一本鎖オリゴヌクレオチド上の領域に含まれるそれぞれの核酸として、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。一本鎖ステムオリゴヌクレオチドで有用な典型的な修飾されたヌクレオチドまたは核酸類似体は、イソグアニンとイソシトシンである。あるいはさらに、一本鎖ステムオリゴヌクレオチドがハイブリダイズする標的結合ドメインを含有する一本鎖オリゴヌクレオチド上の領域に含まれるそれぞれの核酸として、独立に、L-DNAが可能である。

40

【0090】

本開示の別のシーケンシング用プローブを「1部構成リンカープローブ（1-Part Linker Probe）」と名づけ、図2の右図に示してある。図2の1部構成リンカープローブは、リンカーを介して標的結合ドメインに付着したバーコードドメインを含んでいる。この例では、リンカーはPEG分子である。あるいはリンカーはトランス-スチルベンにすることがで

50

きよう。あるいはさらに、リンカーとして、二本鎖DNAと機械的特性が似た任意のポリマーが可能である。典型的なポリマー系スペーサに含まれるのは、ポリエチレングリコール(PEG)型ポリマーである。

【0091】

本開示のシーケンシング用プローブは、約60個のヌクレオチドを含むことができる。本開示のシーケンシング用プローブは、約107個のヌクレオチドを含むことができる。本開示のシーケンシング用プローブとして、長さが約60個のヌクレオチド、または長さが約107個のヌクレオチドが可能である。シーケンシング用プローブを含むヌクレオチドとして、個別に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能であり、その中にL-DNAとD-DNAが含まれる。

10

【0092】

バーコードドメインは、複数の付着位置(例えば1つ、2つ、3つ、4つ、5つ、6つ、7つ、8つ、9つ、10、またはそれよりも多い数の付着位置)を含んでいる。付着位置の数は、標的結合ドメインに含まれるヌクレオチドの数よりも少なくすること、またはその数と同じにすること、またはその数よりも多くすることができる。標的結合ドメインは、骨格ドメインに含まれる付着位置の数よりも多い個数のヌクレオチド(例えば1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはそれよりも多い個数のヌクレオチド)を含むことができる。標的結合ドメインは8個のヌクレオチドを含むことができ、バーコードドメインは3つの付着位置を含んでいる。標的結合ドメインは10個のヌクレオチドを含むことができ、バーコードドメインは3つの付着位置を含んでいる。

20

【0093】

バーコードドメインは、下に説明するように、少なくとも3つの付着位置のための十分なスペースが存在する限り、長さに制限がない。「付着位置」、「位置」、「スポット」という用語は、本明細書では交換可能に用いられる。「バーコードドメイン」と「レポータリングドメイン」という用語は、本明細書では交換可能に用いられる。

【0094】

バーコードドメイン内の各付着位置は、標的結合ドメイン内の2個のヌクレオチド(ジヌクレオチド)に対応しているため、標的結合ドメイン内のそのジヌクレオチドにハイブリダイズする標的核酸内の相補的なジヌクレオチドに対応する。非限定的な一例として、バーコードドメイン内の第1の付着位置は、標的結合ドメインの中の第1と第2のヌクレオチドに対応する(例えば図1では、 R_1 がバーコードドメイン内の第1の付着位置であり、この R_1 は標的結合ドメイン内のジヌクレオチド b_1 と b_2 に対応し、このジヌクレオチドが今度は標的核酸のジヌクレオチド1と2を同定する)。バーコードドメイン内の第2の付着位置は、標的結合ドメイン内の第3と第4のヌクレオチドに対応する(例えば図1では、 R_2 がバーコードドメイン内の第2の付着位置であり、この R_2 は標的結合ドメイン内のジヌクレオチド b_3 と b_4 に対応し、このジヌクレオチドが今度は標的核酸のジヌクレオチド3と4を同定する)。バーコードドメイン内の第3の付着位置は、標的結合ドメイン内の第5と第6のヌクレオチドに対応する(例えば図1では、 R_3 がバーコードドメイン内の第3の付着位置であり、この R_3 は標的結合ドメイン内のジヌクレオチド b_5 と b_6 に対応し、このジヌクレオチドが今度は標的核酸のジヌクレオチド5と6を同定する)。さらに別の非限定的な一例では、バーコードドメイン内の第1の付着位置、バーコードドメイン内の第2の付着位置、バーコードドメイン内の第3の付着位置は、合わせて、標的結合ドメインの中の第1~第6のヌクレオチドに対応する(例えば図1では標的結合ドメイン内のヌクレオチド b_1 - b_6 であり、これが今度は標的核酸の6個のヌクレオチドを同定する)。

30

40

【0095】

バーコードドメイン内の各付着位置は、少なくとも1つの付着領域(例えば1~50箇所、またはそれよりも多い数の付着領域)を含んでいる。バーコードドメイン内のいくつかの位置は、他の位置よりも多くの付着領域を有することができる(例えば第1の付着位置は3つの付着領域を有することができるのに対し、第2の付着位置は2つの付着領域を有することができる)。あるいはバーコードドメイン内の各付着位置は、同じ数の付着領域を

50

有する。バーコードドメイン内の各付着位置は、1つの付着領域を含むことができる。バーコードドメイン内の各付着位置は、2つ以上の付着領域を含むことができる。バーコードドメイン内の少なくとも3つの付着位置のうちの少なくとも1つは、バーコードドメイン内の他の2つの付着位置とは異なる数の付着領域を含むことができる。いくつかの態様では、バーコードドメイン内の各付着位置は、1つの付着領域を含むことができる。

【0096】

各付着領域は、相補的核酸分子（例えばDNAまたはRNA）が可逆的に結合することのできる核酸配列のコピーを少なくとも1つ（すなわち1〜50、例えば10〜30）含んでいる。1つの付着位置の複数の付着領域の核酸配列は互いに同じにすることができる。したがってこれら付着領域に結合する相補的核酸分子は互いに同じである。あるいは1つの位置の複数の付着領域の核酸配列は互いに同じでない。したがってこれら付着領域に結合する相補的核酸分子は互いに同じでない。

10

【0097】

バーコードドメイン内の各付着領域を含む核酸配列として、長さが約6個のヌクレオチド〜約20個のヌクレオチドが可能である。バーコードドメイン内の各付着領域を含む核酸配列として、長さが約12個のヌクレオチドが可能である。バーコードドメイン内の各付着領域を含む核酸配列として、長さが約16個のヌクレオチドが可能である。バーコードドメイン内の各付着領域を含む核酸配列として、長さが約14個のヌクレオチドが可能である。バーコードドメイン内の各付着領域を含む核酸配列として、長さが約8個のヌクレオチドが可能である。バーコードドメイン内の各付着領域を含む核酸配列として、長さが約9個

20

【0098】

付着位置は、または付着領域は、または付着領域の少なくとも1つの核酸配列は、少なくとも1個のスーパーT塩基（5-ヒドロキシブチル-2'-デオキシウリジン）を含むことができる。付着位置は、または付着領域は、または付着領域の少なくとも1つの核酸配列は、少なくとも1個の3'末端スーパーT塩基（5-ヒドロキシブチル-2'-デオキシウリジン）を含むことができる。付着位置は、または付着領域は、または付着領域の少なくとも1つの核酸配列は、少なくとも1つの5'末端スーパーT塩基（5-ヒドロキシブチル-2'-デオキシウリジン）を含むことができる。

【0099】

バーコードドメイン内の各付着領域に含まれるそれぞれの核酸として、独立に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。バーコードドメイン内の付着領域の中の少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個のヌクレオチドとして、修飾されたヌクレオチドまたはヌクレオチド類似体が可能である。バーコードドメイン内の修飾されたヌクレオチドまたはヌクレオチド類似体と正規塩基の典型的な比率は、1：2〜1：8である。バーコードドメイン内の付着領域で有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンまたはイソシトシンである。修飾されたヌクレオチドまたはヌクレオチド類似体（イソグアニン、イソシトシンなど）を用いると、例えばバーコードドメイン内の適切な付着領域に対するレポータの結合の効率と精度を向上させる一方で、他の場所（標的が含まれる）への結合を最少にすることができる。

30

40

【0100】

バーコードドメイン内の1つ以上の付着領域がL-DNAを含むことができる。L-DNAは左巻きであり、天然に生じる右巻きのD-DNAの鏡像バージョンである。L-DNAは、酵素消化に対してより安定かつ抵抗性である。L-DNAはD-LNAにハイブリダイズすることができないため、L-DNAは、バーコードドメイン内の適切な付着領域へのレポータの結合の効率と精度を向上させるとともに、シーケンシング用プローブ上の他の場所にレポータが結合するのを阻止することができる。いくつかの態様では、付着位置の少なくとも1つの核酸配列の各ヌクレオチドとして、L-DNAが可能である。

【0101】

50

バーコードドメイン内の各付着領域に含まれるそれぞれの核酸は、独立に、アデニン塩基、シトシン塩基、グアニン塩基、チミン塩基のいずれかを含むことができる。あるいはバーコードドメイン内の各付着領域に含まれるそれぞれの核酸は、独立に、アデニン塩基、グアニン塩基、チミン塩基のいずれかを含むことができる。

【0102】

バーコードドメイン内の各付着領域に含まれるそれぞれの核酸配列は、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、3'末端グアノシンヌクレオチドとを含むことができる。バーコードドメイン内の各付着領域に含まれるそれぞれの核酸配列は、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、3'末端グアノシンヌクレオチドからなることができる。バーコードドメイン内の各付着領域に含まれるそれぞれの核酸配列は、主に、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、3'末端グアノシンヌクレオチドからなることができる。

10

【0103】

バーコードドメイン内の各付着領域に含まれるそれぞれの核酸配列は、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、5'末端グアノシンヌクレオチドを含むことができる。バーコードドメイン内の各付着領域に含まれるそれぞれの核酸配列は、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、5'末端グアノシンヌクレオチドからなることができる。バーコードドメイン内の各付着領域に含まれるそれぞれの核酸配列は、主に、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のシトシンヌクレオチド、またはこれらの任意の組み合わせと、5'末端グアノシンヌクレオチドからなることができる。

20

【0104】

いくつかの態様では、バーコードドメインの少なくとも1つの付着位置の中の少なくとも1つの付着領域は、3'末端グアノシンヌクレオチドを含むことができる。いくつかの態様では、バーコードドメインの少なくとも2つの付着位置の中の少なくとも1つの付着領域は、3'末端グアノシンヌクレオチドを含むことができる。いくつかの態様では、バーコードドメインの少なくとも3つの付着位置の中の少なくとも1つの付着領域は、3'末端グアノシンヌクレオチドを含むことができる。3'末端グアノシンヌクレオチドとして、L-DNAが可能である。

30

【0105】

いくつかの態様では、バーコードドメインの少なくとも1つの付着位置の中の少なくとも1つの付着領域は、5'末端グアノシンヌクレオチドを含むことができる。いくつかの態様では、バーコードドメインの少なくとも2つの付着位置の中の少なくとも1つの付着領域は、5'末端グアノシンヌクレオチドを含むことができる。いくつかの態様では、バーコードドメインの少なくとも3つの付着位置の中の少なくとも1つの付着領域は、5'末端グアノシンヌクレオチドを含むことができる。5'末端グアノシンヌクレオチドとして、L-DNAが可能である（例えばL-デオキシグアノシン（L-dG））。末端L-dGヌクレオチドは、付着領域および／または付着位置の間のクロスジャンクショナルハイブリダイゼーションを減らすとともに、塩基積層相互作用を提供することによって安定性を維持する。

40

【0106】

1つ以上の付着領域をポリヌクレオチド骨格と一体化することができる。すなわち骨格は単一のポリヌクレオチドであり、付着領域はその単一のポリヌクレオチドの配列の一部である。1つ以上の付着領域を合成骨格内の修飾された単量体（例えば修飾されたヌクレオチド）に連結させ、その付着領域を合成骨格から分岐させることができる。1つの付着位置は2つ以上の付着領域を含むことができ、そのうちのいくつかの付着領域は合成骨格

50

から分岐する一方で、いくつかの付着領域は合成骨格と一体化している。少なくとも1つの付着位置の中の少なくとも1つの付着領域を合成骨格と一体化することができる。上記少なくとも3つの付着位置それぞれの中の各付着領域を合成骨格と一体化することができる。少なくとも1つの付着位置の中の少なくとも1つの付着領域を合成骨格から分岐させることができる。上記少なくとも3つの付着位置それぞれの中の各付着領域を合成骨格から分岐させることができる。

【0107】

バーコードドメイン内の各付着位置は、16通りのジヌクレオチドのうちの1つ、すなわちアデニン-アデニン、アデニン-チミン/ウラシル、アデニン-シトシン、アデニン-グアニン、チミン/ウラシル-アデニン、チミン/ウラシル-チミン/ウラシル、チミン/ウラシル-シトシン、チミン/ウラシル-グアニン、シトシン-アデニン、シトシン-チミン/ウラシル、シトシン-シトシン、シトシン-グアニン、グアニン-アデニン、グアニン-チミン/ウラシル、グアニン-シトシン、グアニン-グアニンのいずれかに対応する。したがってバーコードドメインの単一の付着位置内に位置する1つ以上の付着領域は、16通りのジヌクレオチドのうちの1つに対応していて、その付着領域が対応しているジヌクレオチドに対して特異的な核酸配列を含んでいる。バーコードドメインの異なる付着位置内に位置する付着領域は、バーコード領域内のこれら付着位置が同じジヌクレオチドに対応する場合でさえ、独自の核酸配列を含有している。例えば配列A-G-A-G-A-Cをコードする6量体を有する標的結合ドメインを含有する本開示のシーケンシング用プローブが与えられると、このシーケンシング用プローブのバーコードドメインは3つの付着位置を含有することになり、第1の付着位置はアデニン-グアニンジヌクレオチドに対応し、第2の付着位置はアデニン-グアニンジヌクレオチドに対応し、第3の付着位置はアデニン-シトシンジヌクレオチドに対応することになると考えられる。この例のプローブの位置1に位置する付着領域は、付着位置1と付着位置2の両方がアデニン-グアニンジヌクレオチドに対応する場合でさえ、位置2に位置する付着領域の核酸配列とは異なる独自の核酸配列を含むと考えられる。具体的な付着位置の配列を設計し、個々の付着位置の相補的核酸が異なる付着位置と相互作用しないことを調べる。それに加え、相補的核酸のヌクレオチド配列には制約がない。このヌクレオチド配列は、既知のヌクレオチド配列との実質的な相同性（例えば50%~99.9%）がないことが好ましい。そうすることで、相補的核酸と標的核酸の望ましくないハイブリダイゼーションが制限される。

【0108】

図1は、代表的なバーコードドメインを含む本開示の代表的な1つのシーケンシング用プローブの図を示している。図1に例示したバーコードドメインは、3つの付着位置 R_1 、 R_2 、 R_3 を含んでいる。それぞれの付着位置は、標的結合ドメインの6量体配列(b_1 - b_6)の中に存在する特定のジヌクレオチドに対応する。この例では、 R_1 は位置 b_1 と b_2 に対応し、 R_2 は位置 b_3 と b_4 に対応し、 R_3 は位置 b_5 と b_6 に対応する。したがって各位置は、標的結合ドメインの6量体配列の中に存在する特定のジヌクレオチドをデコードし、そのことによってそれぞれのジヌクレオチドの中に存在する特定の2個の塩基(A、C、G、Tのいずれか)を同定することが可能になる。

【0109】

図1に示されている代表的なバーコードドメインでは、各付着位置は、合成骨格に一体化された単一の付着領域を含んでいる。上記の3つの付着位置の各付着領域は、各付着位置によってコードされた個々のジヌクレオチドに対応する特定のヌクレオチド配列を含有している。例えば付着位置 R_1 は、ジヌクレオチド b_1 - b_2 の属性に対応する特定の配列を有する付着領域を含んでいる。

【0110】

バーコードドメインはさらに、1つ以上の結合領域を含むことができる。バーコードドメインは、少なくとも1つの付着位置に近接するか隣接した少なくとも1つの一本鎖核酸配列を含むことができる。バーコードドメインは、少なくとも2つの付着位置に近接するか隣接した少なくとも2つの一本鎖核酸配列を含むことができる。バーコードドメインは、

少なくとも3つの付着位置に近接するか隣接した少なくとも3つの一本鎖核酸配列を含むことができる。これら隣接部分は「トー-ホールド」として知られており、一本鎖オリゴヌクレオチドのための追加結合部位を提供することによってトー-ホールドに近接してハイブリダイズしたオリゴヌクレオチドの交換速度を大きくするのに用いることができる（例えば「トー-ホールド」プローブ；例えばSeeling他、「準安定DNA燃料の触媒による緩和」；J. Am. Chem. Soc. 2006年、第128巻(37)、12211～12220ページを参照されたい）。

【0111】

バーコードドメイン内の少なくとも1つの付着領域を二本鎖核酸配列によって少なくとも1つの側に隣接させることができる。バーコードドメイン内の少なくとも2つの付着領域を二本鎖核酸配列によって少なくとも1つの側に隣接させることができる。バーコードドメイン内の少なくとも3つの付着領域を二本鎖核酸配列によって少なくとも1つの側に隣接させることができる。

【0112】

バーコードドメイン内の任意の付着領域を、「ポケットオリゴ (pocket oligo)」と呼ばれる二本鎖核酸配列によって隣の任意の付着位置から分離することができる。図28は、3つの付着位置を含むバーコードドメインを有するシーケンシング用プローブの一例を示している。付着位置1はポケットオリゴによって隣の付着位置2から分離されている。付着位置2はさらに、別のポケットオリゴによって隣の付着位置3から分離されている。

【0113】

ポケットオリゴに含まれるそれぞれの核酸として、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。ポケットオリゴで有用な典型的な修飾されたヌクレオチドまたは核酸類似体は、イソグアニンとイソシトシンである。あるいはさらに、ポケットオリゴに含まれるそれぞれの核酸として、独立に、L-DNAが可能である。ポケットオリゴは、少なくとも1つのスーパーT塩基（5-ヒドロキシブチル-2'-デオキシウリジン）を含むことができる。ポケットオリゴとして、長さが約25個のヌクレオチドが可能である。

【0114】

いくつかの態様では、バーコードドメイン内の少なくとも1つ、または少なくとも2つ、または少なくとも3つの付着位置を少なくとも1つのフランキング二本鎖ポリヌクレオチドに隣接させることができる。少なくとも1つのフランキング二本鎖ポリヌクレオチドは、少なくとも1つの修飾されたヌクレオチドまたは核酸類似体を含むことができる。少なくとも1つのフランキング二本鎖ポリヌクレオチドは、L-DNAを含むことができる。少なくとも1つのフランキング二本鎖ポリヌクレオチドは、少なくとも1つのスーパーT塩基（5-ヒドロキシブチル-2'-デオキシウリジン）を含むことができる。少なくとも1つの隣接二本鎖ポリヌクレオチドとして、長さが約25個のヌクレオチドが可能である。

【0115】

バーコードドメイン内の少なくとも1つの付着領域を、二本鎖DNAと機械的特性が似た任意のポリマーによって少なくとも1つの側に隣接させることができる。典型的なポリマー系スペーサに含まれるのはポリエチレングリコール（PEG）型ポリマーである。バーコードドメイン内の少なくとも2つの付着領域を、二本鎖DNAと機械的特性が似た任意のポリマーによって少なくとも1つの側に隣接させることができる。バーコードドメイン内の少なくとも3つの付着領域を、二本鎖DNAと機械的特性が似た任意のポリマーによって少なくとも1つの側に隣接させることができる。

【0116】

バーコードドメイン内の任意の付着領域を、二本鎖DNAと機械的特性が似た任意のポリマーによって隣接した任意の付着位置から分離することができる。典型的なポリマー系スペーサに含まれるのはポリエチレングリコール（PEG）型ポリマーである。図29は、3つの付着位置を含むシーケンシング用プローブの一例を示している。付着位置1はPEGリンカーによって隣の付着位置2から分離されている。付着位置2はさらに、別のPEGリンカーによって隣の付着位置3から分離されている。

【0117】

バーコードドメイン内の少なくとも1つの付着領域を、非塩基性である一本鎖核酸分子によって少なくとも1つの側に隣接させることができる。非塩基性の核酸分子は、プリン塩基もピリミジン塩基も持たない核酸分子である。バーコードドメイン内の少なくとも2つの付着領域を、非塩基性である一本鎖核酸分子によって少なくとも1つの側に隣接させることができる。バーコードドメイン内の少なくとも3つの付着領域を、非塩基性である一本鎖核酸分子によって少なくとも1つの側に隣接させることができる。

【0118】

バーコードドメイン内の任意の付着領域を、非塩基性である一本鎖核酸分子によって任意の隣の付着位置から分離することができる。図30は、3つの付着位置を含むバーコードドメインを有するシーケンシング用プローブの一例を示している。付着位置1は非塩基性である一本鎖核酸分子によって隣の付着位置2から分離されている。付着位置2はさらに、非塩基性である一本鎖核酸分子によって隣の付着位置3から分離されている。

10

【0119】

バーコードドメイン内の任意の付着領域を、3'末端グアノシンヌクレオチドによって任意の隣の付着位置から分離することができる。いくつかの態様では、バーコードドメインの少なくとも2つの付着位置の中の少なくとも1つの付着領域は、3'末端グアノシンヌクレオチドを含むことができる。図53は、3つの付着位置を含むバーコードドメインを有するシーケンシング用プローブの一例を示しており、それぞれの付着位置が末端L-Gヌクレオチドによって分離されている。付着位置1は、L-Gヌクレオチドによって隣の付着位置2から分離されている。付着位置2はさらに、L-Gヌクレオチドによって隣の付着位置3から分離されている。付着位置3は、3'末端がL-Gヌクレオチドで終わっている。

20

【0120】

本開示のシーケンシング用プローブは、約20ナノメートル～約50ナノメートルの全長を有することができる（その中に、標的結合ドメイン、バーコードドメインと、オプションの任意のドメインが含まれる）。シーケンシング用プローブの骨格として、約120個のヌクレオチド、または約60個のヌクレオチド、または約52個のヌクレオチド、または約27個のヌクレオチドを含むポリヌクレオチド分子が可能である。

【0121】

シーケンシング用プローブは、切断可能なリンカーの修飾を含むことができる。切断可能なリンカーの修飾は、少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個、または少なくとも7個、または少なくとも8個、または少なくとも9個、または少なくとも10個、または任意の個数の切断可能な部分を含むことができる。当業者に知られている任意の切断可能なリンカーの修飾、または切断可能な部分を使用することができる。切断可能なリンカーの修飾と切断可能な部分の非限定的な例に含まれるのは、UV光で切断可能なリンカー、還元剤で切断可能なリンカー、酵素で切断可能なリンカーである。酵素で切断可能なリンカーの一例は、USER（商標）酵素によって切断するためのデオキシウラシルの挿入である。切断可能なリンカーの修飾は、シーケンシング用プローブの長さに沿った任意の位置に配置することができ、その位置の非限定的な例に含まれるのは、標的結合ドメインとバーコードドメインの間の領域である。図7の右図は、本開示のプローブに組み込むことのできる代表的な切断可能なリンカーの修飾を示している。

30

40

【0122】

レポータプローブ

【0123】

本開示によるシーケンシング用プローブのバーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域内の相補的核酸配列に結合（例えばハイブリダイズ）する核酸分子は、（直接的または間接的に）検出可能な標識を含んでいる。この検出可能な標識を本明細書では「レポータプローブ」または「レポータプローブ複合体」と呼び、本明細書ではこれらの用語は交換可能に用いられる。レポータプローブとして、DNA、

50

RNA、PNAのいずれかが可能である。レポータプローブはDNAであることが好ましい。

【0124】

レポータプローブは少なくとも2つのドメインを含むことができ、そのうちの第1のドメインは少なくとも1個の第1の相補的核酸分子に結合することができ、第2のドメインは、第1の検出可能な標識と少なくとも第2の検出可能な標識に結合することができる。図3は、代表的なシーケンシング用プローブのバーコードドメインの第1の付着位置に結合した本開示の代表的なレポータプローブの模式図を示している。図3では、レポータプローブの第1のドメイン（栗色の市松模様で表示）がバーコードドメインの付着位置 R_1 内の相補的核酸配列に結合し、このレポータプローブの第2のドメイン（灰色で表示）が2つの検出可能な標識（1つは緑色の標識、1つは赤色の標識）に結合している。

10

【0125】

あるいはレポータプローブは少なくとも2つのドメインを含むことができ、そのうちの第1のドメインは、少なくとも1個の第1の相補的核酸分子に結合することができ、第2のドメインは、少なくとも1個の第2の相補的核酸分子に結合することができる。これらの少なくとも1個の第1と第2の相補的核酸分子は、異なっている（異なる核酸配列を有する）ことが可能である。

【0126】

「一次核酸分子（primary nucleic acid molecule）」は、少なくとも2つのドメインを含むレポータプローブであり、そのうちの第1のドメインは、シーケンシング用プローブのバーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域内の相補的核酸配列に結合（例えばハイブリダイズ）することができ、第2のドメインは、少なくとも1つの追加の相補的核酸に結合（例えばハイブリダイズ）することができる。一次核酸分子は、シーケンシング用プローブのバーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域内の相補的核酸配列に直接的に結合することができる。一次核酸分子は、シーケンシング用プローブのバーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域内の相補的核酸配列に核酸リンカーを介して間接的に結合することができる。核酸リンカーは「コネクタオリゴ」と呼ばれる。

20

【0127】

コネクタオリゴは少なくとも2つのドメインを含むことができ、そのうちの第1のドメインは、バーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域内の少なくとも1つの第1の相補的核酸配列に結合（例えばハイブリダイズ）することができ、第2のドメインは、一次核酸分子の第1のドメインに結合（例えばハイブリダイズ）することができる。図31は、レポータプローブにコネクタオリゴを介して結合したシーケンシング用プローブを示している。

30

【0128】

コネクタオリゴの第1のドメインまたは第2のドメインに含まれるそれぞれの核酸として、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。コネクタオリゴの第1のドメインまたは第2のドメインで有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンまたはイソシトシンである。修飾されたヌクレオチドまたはヌクレオチド類似体（イソグアニン、イソシトシンなど）を用いると、例えばシーケンシング用プローブのバーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域内の適切な相補的核酸配列に対するコネクタオリゴの第1のドメインの結合の効率と精度を向上させる一方で、他の場所（標的が含まれる）への結合を最少にすることができる。修飾されたヌクレオチドまたはヌクレオチド類似体（イソグアニン、イソシトシンなど）を用いると、例えばレポータプローブの適切な第1のドメインに対するコネクタオリゴの第2のドメインの結合の効率と精度を向上させる一方で、他の場所（標的が含まれる）への結合を最少にすることができる。あるいはコネクタオリゴの第1のドメインまたは第2のドメインに含まれるそれぞれの核酸として、独立に、L-DNAが可能である。コネクタオリゴの一例では、第1のドメインはD-DNAを含み、第2のドメインはL-DNAを含んでいる。コネクタオリゴの別の一例では、第1のドメインはD-DNAを含み

40

50

、第2のドメインはイソグアニンおよび／またはイソシトシンを含んでいる。

【0129】

コネクタオリゴの第1のドメインとして、長さが約8個～約16個のヌクレオチドが可能である。コネクタオリゴの第1のドメインは、長さが14個のヌクレオチドであることが好ましい。コネクタオリゴの第2のドメインとして、長さが約4～12個のヌクレオチドが可能である。コネクタオリゴの第2のドメインは、長さが約8個のヌクレオチドであることが好ましい。

【0130】

コネクタオリゴを含むいくつかの態様では、付着領域を部分的に二本鎖の付着領域と呼ぶことができる。部分的に二本鎖の付着領域は、二本鎖領域と一本鎖領域を含むことができる。部分的に二本鎖の付着領域の一本鎖領域は、少なくとも1つの相補的核酸配列に結合（例えばハイブリダイズ）する少なくとも1つの核酸配列を含むことができる。部分的に二本鎖の付着領域の一本鎖領域に結合（例えばハイブリダイズ）するその少なくとも1つの相補的核酸配列として、一次核酸分子が可能である。

【0131】

部分的に二本鎖の付着領域の二本鎖領域に含まれるそれぞれの核酸として、独立に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。部分的に二本鎖の付着領域の二本鎖領域内の少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個、または少なくとも7個、または少なくとも8個のヌクレオチドとして、修飾されたヌクレオチドまたは核酸類似体が可能である。二本鎖領域内の修飾されたヌクレオチドまたはヌクレオチド類似体と正規塩基の典型的な比率は、1：2～1：8である。部分的に二本鎖の付着領域の二本鎖領域で有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンとイソシトシンである。あるいは部分的に二本鎖の付着領域の二本鎖領域に含まれるそれぞれの核酸として、独立に、L-DNAが可能である。

【0132】

部分的に二本鎖の付着領域の一本鎖領域に含まれるそれぞれの核酸として、独立に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。部分的に二本鎖の付着領域の一本鎖領域の中の少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個、または少なくとも7個、または少なくとも8個のヌクレオチドとして、修飾されたヌクレオチドまたは核酸類似体が可能である。一本鎖領域内の修飾されたヌクレオチドまたはヌクレオチド類似体と正規塩基の典型的な比率は、1：2～1：8である。部分的に二本鎖の付着領域の一本鎖領域で有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンとイソシトシンである。修飾されたヌクレオチドまたはヌクレオチド類似体（イソグアニン、イソシトシンなど）を用いると、例えば一次核酸分子の適切な相補的核酸配列に対する部分的に二本鎖の付着領域の一本鎖領域の結合の効率と精度を向上させる一方で、他の場所（標的が含まれる）への結合を最少にすることができる。あるいは部分的に二本鎖の付着領域の一本鎖領域に含まれるそれぞれの核酸として、独立に、L-DNAが可能である。

【0133】

一次核酸分子は、切断可能なリンカーを含むことができる。切断可能なリンカーは、第1のドメインと第2のドメインとの間に位置することができる。切断可能なリンカーは、光で切断可能であることが好ましい。切断可能なリンカーは、少なくとも1つ、または少なくとも2つの切断可能な部分を含むことができる。その少なくとも1つ、または少なくとも2つの切断可能な部分は、光で切断することができる。

【0134】

一次核酸分子の第1のドメインとして、長さが約6～16個のヌクレオチドが可能である。一次核酸分子の第1のドメインは、長さが約8個のヌクレオチドであることが好ましい。

【0135】

一次核酸分子の第1のドメインに含まれるそれぞれの核酸として、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。一次核酸分子の第1のドメイン内の少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個、または少なくとも7個、または少なくとも8個のヌクレオチドとして、修飾されたヌクレオチドまたは核酸類似体が可能である。第1のドメイン内の修飾されたヌクレオチドまたはヌクレオチド類似体と正規塩基の典型的な比率は、1:2~1:8である。一次核酸分子の第1のドメインで有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンとイソシトシンである。修飾されたヌクレオチドまたはヌクレオチド類似体（イソグアニン、イソシトシンなど）を用いると、例えばシークエンシング用プローブのバーコードドメインの少なくとも1つの付着位置内の少なくとも1つの付着領域内の適切な相補的核酸配列に対する一次核酸分子の第1のドメインの結合の効率と精度を向上させる一方で、他の場所（標的が含まれる）への結合を最少にすることができる。あるいは一次核酸分子の第1のドメインに含まれるそれぞれの核酸として、独立に、L-DNAが可能である。

【0136】

いくつかの態様では、一次核酸分子の第1のドメインは、全面的にL-DNAで構成することができ、一次核酸分子の第2のドメインは、全面的にD-DNAで構成することができる。

【0137】

いくつかの態様では、一次核酸分子の第1のドメインは、3'末端シトシンヌクレオチドを含むことができる。いくつかの態様では、一次核酸分子の第1のドメインは、3'末端シトシンヌクレオチドを含むことができ、この3'末端シトシンヌクレオチドはL-DNAである。

【0138】

いくつかの態様では、一次核酸分子の第1のドメインは、5'末端シトシンヌクレオチドを含むことができる。いくつかの態様では、一次核酸分子の第1のドメインは、5'末端シトシンヌクレオチドを含むことができ、この5'末端シトシンヌクレオチドはL-DNAである。

【0139】

いくつかの態様では、一次核酸分子の第1のドメインは、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のグアニンヌクレオチド、またはこれらの任意の組み合わせと、3'末端シトシンヌクレオチドを含むことができる。いくつかの態様では、一次核酸分子の第1のドメインは、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のグアニンヌクレオチド、またはこれらの任意の組み合わせと、3'末端シトシンヌクレオチドからなることができる。いくつかの態様では、一次核酸分子の第1のドメインは、主に、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のグアニンヌクレオチド、またはこれらの任意の組み合わせと、3'末端シトシンヌクレオチドからなることができる。

【0140】

いくつかの態様では、一次核酸分子の第1のドメインは、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のグアニンヌクレオチド、またはこれらの任意の組み合わせと、5'末端シトシンヌクレオチドを含むことができる。いくつかの態様では、一次核酸分子の第1のドメインは、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のグアニンヌクレオチド、またはこれらの任意の組み合わせと、5'末端シトシンヌクレオチドからなることができる。いくつかの態様では、一次核酸分子の第1のドメインは、主に、少なくとも1個のアデニンヌクレオチド、少なくとも1個のチミンヌクレオチド、少なくとも1個のグアニンヌクレオチド、またはこれらの任意の組み合わせと、5'末端シトシンヌクレオチドからなることができる。

【0141】

一次核酸分子に結合する少なくとも1個の追加の相補的核酸を本明細書では「二次核酸分子 (secondary nucleic molecule)」と呼ぶ。一次核酸分子は、少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、またはそれよりも多い個数の二次核酸分子に結合（例えばハイブリダイズ）することができる。一次核酸分子は、4個の二次核酸分子に結合（例えばハイブリダイズ）することが好ましい。

【0142】

二次核酸分子は少なくとも2つのドメインを含むことができ、そのうちの第1のドメインは、少なくとも1個の一次核酸分子の中の少なくとも1つの相補的配列に結合（例えばハイブリダイズ）することができ、第2のドメインは、(a) 第1の検出可能な標識と、少なくとも第2の検出可能な標識；または (b) 少なくとも1つの追加の相補的核酸；または (c) これらの組み合わせに結合（例えばハイブリダイズ）することができる。いくつかの態様では、二次核酸分子の第1のドメインを全面的にL-DNAで構成し、二次核酸分子の第2のドメインを全面的にD-DNAで構成することができる。いくつかの態様では、二次核酸分子の第1のドメインと第2のドメインの両方を全面的にD-DNAで構成することができる。

【0143】

二次核酸分子は、切断可能なリンカーを含むことができる。切断可能なリンカーは、第1のドメインと第2のドメインとの間に位置することができる。切断可能なリンカーは、光で切断できることが好ましい。

【0144】

二次核酸分子の第1のドメインに含まれるそれぞれの核酸として、独立に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。二次核酸分子の第1のドメイン内の少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個のヌクレオチドとして、修飾されたヌクレオチドまたは核酸類似体が可能である。二次核酸分子の第1のドメイン内の修飾されたヌクレオチドまたはヌクレオチド類似体と正規塩基の典型的な比率は、1:2~1:8である。二次核酸分子の第1のドメインで有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンとイソシトシンである。修飾されたヌクレオチドまたはヌクレオチド類似体（イソグアニン、イソシトシンなど）を用いると、例えば二次核酸分子の第2のドメイン内の適切な相補的核酸配列に対する二次核酸分子の第1のドメインの結合の効率と精度を向上させる一方で、他の場所（標的が含まれる）への結合を最少にすることができる。

【0145】

二次核酸分子に結合する少なくとも1個の追加の相補的核酸を本明細書では「三次核酸分子 (tertiary nucleic molecule)」と呼ぶ。二次核酸分子は、少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個、または少なくとも7個、またはそれよりも多い個数の三次核酸分子に結合（例えばハイブリダイズ）することができる。その少なくとも1個の二次核酸分子は、1個の三次核酸分子に結合（例えばハイブリダイズ）することが好ましい。

【0146】

三次核酸分子は少なくとも2つのドメインを含んでいて、そのうちの第1のドメインは、少なくとも1個の二次核酸分子内の少なくとも1つの相補的配列に結合（例えばハイブリダイズ）することができ、第2のドメインは、第1の検出可能な標識と、少なくとも第2の検出可能な標識に結合（例えばハイブリダイズ）することができる。あるいは第2のドメインは、第1の検出可能な標識と、少なくとも第2の検出可能な標識を含むことが、例えばホスホロアミダイト化学またはNHS化学を利用したオリゴヌクレオチド合成の間にこれらの標識を直接的または間接的に付着させることを通じて可能になる。いくつかの態様では、三次核酸分子の第1のドメインを全面的にL-DNAで構成し、三次核酸分子の第2のドメインを全面的にD-DNAで構成することができる。いくつかの態様では、三次核酸分子の第1のドメインと第2のドメインの両方を全面的にD-DNAで構成することができる。三次核酸分

子は、切断可能なリンカーを含むことができる。切断可能なリンカーは、第1のドメインと第2のドメインとの間に位置することができる。切断可能なリンカーは、光で切断できることが好ましい。

【0147】

三次核酸分子の第1のドメインに含まれるそれぞれの核酸として、独立に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。三次核酸分子の第1のドメイン内の少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個のヌクレオチドとして、修飾されたヌクレオチドまたは核酸類似体が可能である。三次核酸分子の第1のドメイン内の修飾されたヌクレオチドまたはヌクレオチド類似体と正規塩基の典型的な比率は、1：2～1：8である。三次核酸分子の第1のドメインで有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンとイソシトシンである。修飾されたヌクレオチドまたはヌクレオチド類似体（イソグアニン、イソシトシンなど）を用いると、例えば二次核酸分子の第2のドメイン内の適切な相補的核酸配列に対する三次核酸分子の第1のドメインの結合の効率と精度を向上させる一方で、他の場所（標的が含まれる）への結合を最少にすることができる。

【0148】

レポータプローブは、第1の検出可能な標識と少なくとも第2の検出可能な標識に結合して2色の組み合わせを生み出す。蛍光染料のこの二重の組み合わせには、1つの色の重複（例えば青色-青色）を含めることができる。本明細書では、「標識」という用語に、検出可能なシグナルを発生させることのできる単一の部分、または互いに同じか実質的に同じ検出可能なシグナルを発生させることのできる複数の部分が含まれる。例えば標識には、単一の黄色蛍光染料（ALEXA FLUOR（商標）532など）、または複数の黄色蛍光染料（ALEXA FLUOR（商標）532など）が含まれる。

【0149】

レポータプローブは、第1の検出可能な標識と少なくとも第2の検出可能な標識に結合することができ、それぞれの検出可能な標識は、青（B）、緑（G）、黄（Y）、赤（R）という4種類の蛍光染料のうちの1つである。これら4種類の染料を用いると、可能な2色の組み合わせが10通り（BB；BG；BR；BY；GG；GR；GY；RR；RY；YY）生じる。いくつかの態様では、図3に示したように、本開示のレポータプローブに、8通りの可能な色組み合わせ、すなわちBB、BG、BR、BY、GG、GR、GY、YYのうちの1つで標識する。検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

【0150】

シーケンシング用プローブと一次核酸分子を含む態様では、本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインが表1に記載されている任意のコンストラクトを含むシーケンシング用プローブが提供される。代表的な標的結合ドメインは少なくとも8個のヌクレオチドを含んでいて標的核酸にハイブリダイズすることができ、標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたはヌクレオチド類似体であることが可能であり、標的結合ドメイン内にあって標的核酸分子内の対応するヌクレオチドを同定しない上記少なくとも2個のヌクレオチドは、標的結合ドメイン内の上記少なくとも6個のヌクレオチドによって指定される標的に特異的ではない4つの正規塩基、またはユニバーサル塩基、または縮重塩基のうちの任意のものが可能である。代表的なバーコードドメインは、合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、少なくとも1個の相補的な一次核酸分子が結合する少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含み、相補的な一次核酸分子は、第1の検出可能な標識と少なくとも1つの

第 2 の検出可能な標識を含み、上記少なくとも 3 つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも 6 個のヌクレオチドのうちの 2 個のヌクレオチドに対応し、上記少なくとも 3 つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも 3 つの付着位置のそれぞれの位置に結合するそれぞれの相補的な一次核酸分子の上記少なくとも 1 つの第 1 の検出可能な標識と上記少なくとも 1 つの第 2 の検出可能な標識が、標的結合ドメインが結合する標的核酸内の上記少なくとも 6 個のヌクレオチドのうちの対応する 2 個のヌクレオチドの位置と属性を決めている。標的結合ドメイン内にあつて標的核酸分子内の対応するヌクレオチドを同定しない上記少なくとも 2 個のヌクレオチドとして、標的結合ドメイン内の上記少なくとも 6 個のヌクレオチドによって指定される標的に特異的ではない 4 つの正規塩基、またはユニバーサル塩基、または縮重塩基のうちの任意のものが可能である。

10

【0151】

いくつかの態様では、標的結合ドメイン内にあつて標的核酸分子内の対応するヌクレオチドを同定しない少なくとも 1 個のヌクレオチドは、標的結合ドメイン内にあつて標的核酸分子内の対応するヌクレオチドを同定するヌクレオチドよりも前に来ることができる。いくつかの態様では、標的結合ドメイン内にあつて標的核酸分子内の対応するヌクレオチドを同定しない少なくとも 1 個のヌクレオチドは、標的結合ドメイン内にあつて標的核酸分子内の対応するヌクレオチドを同定するヌクレオチドの後に続くことができる。

【0152】

別の態様では、代表的な標的結合ドメインは、標的核酸にハイブリダイズすることのできる少なくとも 6 個のヌクレオチドを含むことができ、標的結合ドメイン内のこれら少なくとも 6 個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内のこれら少なくとも 6 個のヌクレオチドはどれも、修飾されたヌクレオチドまたは核酸類似体でないか、標的結合ドメイン内のその少なくとも 6 個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたは核酸類似体であることが可能である。

20

【0153】

シーケンシング用プローブと一次核酸分子を含む態様では、本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインは少なくとも 10 個のヌクレオチドを含んでいて標的核酸にハイブリダイズすることができ、標的結合ドメイン内の少なくとも 6 個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも 4 個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも 3 つの付着位置を含み、それぞれの付着位置は、少なくとも 1 個の相補的な一次核酸分子が結合する少なくとも 1 つの核酸配列を含む少なくとも 1 つの付着領域を含み、その相補的な一次核酸分子は、少なくとも 1 つの第 1 の検出可能な標識と少なくとも 1 つの第 2 の検出可能な標識を含み、上記少なくとも 3 つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも 6 個のヌクレオチドのうちの 2 個のヌクレオチドに対応するとともに、上記少なくとも 3 つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも 3 つの付着位置の各位置に結合したそれぞれの相補的な一次核酸分子の上記少なくとも 1 つの第 1 の検出可能な標識と上記少なくとも 1 つの第 2 の検出可能な標識が、標的結合ドメインが結合した標的核酸内の上記少なくとも 6 個のヌクレオチドの対応する 2 個のヌクレオチドの位置と属性を決めているシーケンシング用プローブも提供される。

30

40

【0154】

シーケンシング用プローブと一次核酸分子と二次核酸分子を含む態様では、本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインが表 1 に記載されている任意のコンストラクトを含むシーケンシング用プローブが提供される。代表的な標的結合ドメインは少なくとも 8 個のヌクレオチドを含んでいて、標的核酸にハイブリダイズすることができ、標的結合ドメイン内の少なく

50

とも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたはヌクレオチド類似体であることが可能であり、標的結合ドメイン内にあって標的核酸分子内の対応するヌクレオチドを同定しない上記少なくとも2個のヌクレオチドは、標的結合ドメイン内の上記少なくとも6個のヌクレオチドによって指定される標的に特異的ではない4つの正規塩基、またはユニバーサル塩基、または縮重塩基のうちの任意のものが可能である。代表的なバーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、少なくとも1個の相補的な一次核酸分子が結合する少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含み、相補的な一次核酸分子には、少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む少なくとも1つの相補的な一次核酸分子がさらに結合し、上記少なくとも3つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置のそれぞれの位置に結合するそれぞれの相補的な二次核酸分子の上記少なくとも1つの第1の検出可能な標識と上記少なくとも1つの第2の検出可能な標識が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドのうちの対応する2個のヌクレオチドの位置と属性を決めている。

【0155】

別の態様では、代表的な標的結合ドメインは、標的核酸にハイブリダイズすることのできる少なくとも6個のヌクレオチドを含むことができ、標的結合ドメイン内のこれら少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ；これら少なくとも6個のヌクレオチドのどれも、修飾されたヌクレオチドまたはヌクレオチド類似体ではないか、これら少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたはヌクレオチド類似体である。

【0156】

シーケンシング用プローブと一次核酸分子と二次核酸分子を含む態様では、本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインは少なくとも10個のヌクレオチドを含んでいて標的核酸にハイブリダイズすることができ、標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも4個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、少なくとも1つの相補的な一次核酸分子が結合する少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含んでおり、相補的な一次核酸分子には、少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む少なくとも1個の相補的な二次核酸分子がさらに結合し、上記少なくとも3つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置のそれぞれの位置に結合するそれぞれの相補的な二次核酸分子の上記少なくとも1つの第1の検出可能な標識と上記少なくとも1つの第2の検出可能な標識が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドの対応する2個のヌクレオチドの位置と属性を決めているシーケンシング用プローブも提供される。

【0157】

シーケンシング用プローブと一次核酸分子と二次核酸分子と三次核酸分子を含む態様では、本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインが表1に記載されている任意のコンストラクトを含むシーケンシング用プローブが提供される。代表的な標的結合ドメインは少なくとも8個

のヌクレオチドを含んでいて標的核酸にハイブリダイズすることができ、標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも2個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたはヌクレオチド類似体であることが可能であり、標的結合ドメイン内にあって標的核酸分子内の対応するヌクレオチドを同定しない上記少なくとも2個のヌクレオチドは、標的結合ドメイン内の上記少なくとも6個のヌクレオチドによって指定される標的に特異的ではない4つの正規塩基、またはユニバーサル塩基、または縮重塩基のうちの任意のものが可能である。代表的なバーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、少なくとも1つの相補的な一次核酸分子が結合する少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含み、相補的な一次核酸分子には、少なくとも1個の相補的な二次核酸分子がさらに結合し、この少なくとも1個の相補的な二次核酸分子にはさらに、少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む少なくとも1個の相補的な三次核酸分子が結合し、上記少なくとも3つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置のそれぞれの位置に結合するそれぞれの相補的な三次核酸分子の上記少なくとも1つの第1の検出可能な標識と上記少なくとも1つの第2の検出可能な標識が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドのうちの対応する2個のヌクレオチドの位置と属性を決めている。

【0158】

別の態様では、代表的な標的結合ドメインは、標的核酸にハイブリダイズすることのできる少なくとも6個のヌクレオチドを含むことができ、標的結合ドメイン内のこれら少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ；これら少なくとも6個のヌクレオチドのどれも、修飾されたヌクレオチドまたはヌクレオチド類似体ではないか、これら少なくとも6個のヌクレオチドのいずれかが、修飾されたヌクレオチドまたはヌクレオチド類似体である。

【0159】

シーケンシング用プローブと一次核酸分子と二次核酸分子と三次核酸分子を含む態様では、本開示により、標的結合ドメインとバーコードドメインを含むシーケンシング用プローブとして、標的結合ドメインは少なくとも10個のヌクレオチドを含んでいて標的核酸にハイブリダイズすることができ、標的結合ドメイン内の少なくとも6個のヌクレオチドは、標的核酸分子内の対応する（相補的な）ヌクレオチドを同定することができ、標的結合ドメイン内の少なくとも4個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定せず；バーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、少なくとも1個の相補的な一次核酸分子が結合する少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含んでおり、相補的な一次核酸分子には、少なくとも1個の相補的な二次核酸分子がさらに結合し、この少なくとも1個の相補的な二次核酸分子には、少なくとも1つの第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を含む少なくとも1個の相補的な三次核酸分子がさらに結合し、上記少なくとも3つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置のそれぞれの位置に結合するそれぞれの相補的な三次核酸分子の上記少なくとも1つの第1の検出可能な標識と上記少なくとも1つの第2の検出可能な標識が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドの対応する2個のヌクレオチドの位置と属性を決めているシーケンシング用プローブも提供される。

【0160】

本開示により、シーケンシング用プローブと、検出可能な標識を二次核酸分子と三次

核酸分子の両方に有するレポータプローブも提供される。例えば二次核酸分子は一次核酸分子に結合することができ、二次核酸分子は、第1の検出可能な標識と少なくとも第2の検出可能な標識の両方を含むとともに、第1の検出可能な標識と少なくとも第2の検出可能な標識を含む少なくとも1個の三次核酸分子にも結合することができる。二次核酸分子上に位置する第1の検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。三次核酸分子上に位置する第1の検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。二次核酸分子上の検出可能な標識の発光スペクトルは、三次核酸分子上の検出可能な標識の発光スペクトルと同じ発光スペクトルにすること、または異なった発光スペクトルにすることができる。

10

【0161】

図4は、代表的な一次核酸分子と二次核酸分子と三次核酸分子を含む本開示の代表的なレポータプローブの模式図である。一次核酸分子は3'末端に第1のドメインを含み、この第1のドメインは、シーケンシング用プローブのバーコードドメインの付着位置内の相補的な付着領域にハイブリダイズする12個のヌクレオチドからなる配列を含んでいる。5'末端には、6個の二次核酸分子にハイブリダイズする第2のドメインがある。示されている代表的な二次核酸分子のほうは、一次核酸分子にハイブリダイズする5'末端内の第1のドメインと、5個の三次核酸分子にハイブリダイズする3'末端内のドメインを含んでいる。

【0162】

三次核酸分子は少なくとも2つのドメインを含んでいる。第1のドメインは、二次核酸分子に結合することができる。三次核酸分子の第2のドメインは、第1の検出可能な標識と、少なくとも第2の検出可能な標識に結合することができる。三次核酸分子の第2のドメインは、蛍光標識した1つ以上のヌクレオチド単量体を三次核酸分子の第2のドメインの配列に直接組み込むことにより、第1の検出可能な標識と少なくとも第2の検出可能な標識に結合させることができる。三次核酸分子の第2のドメインには、この三次核酸分子の第2のドメインに対する標識となる短いポリヌクレオチドをハイブリダイズさせることにより、第1の検出可能な標識と少なくとも第2の検出可能な標識を結合させることができる。これらの短いポリヌクレオチドは「標識されたオリゴ」と呼ばれ、蛍光標識されたヌクレオチド単量体の直接的な組み込みによって、または当業者に知られている他の核酸標識法によって標識することができる。図4に示されている代表的な三次核酸分子は「標識されたオリゴ」と見なすことができ、二次核酸分子にハイブリダイズする第1のドメインを含むとともに、例えばNHS化学を利用してオリゴヌクレオチド合成の間に標識を間接的に付着させることによって、または三次核酸分子の合成中に1つ以上の蛍光標識されたヌクレオチド単量体を組み込むことによって蛍光標識された第2のドメインを含んでいる。標識されたオリゴとして、DNA、RNA、PNAのいずれかが可能である。

20

30

【0163】

標識されたオリゴは、蛍光部分とポリヌクレオチド分子の間に切断可能なリンカーを含むことができる。切断可能なリンカーは、光で切断できることが好ましい。切断可能なリンカーは、化学的に切断すること、または酵素で切断することもできる。

【0164】

40

別の態様では、二次核酸の第2のドメインは、第1の検出可能な標識と少なくとも第2の検出可能な標識に結合することができる。二次核酸の第2のドメインは、蛍光標識した1つ以上のヌクレオチド単量体を二次核酸の第2のドメインに直接組み込むことにより、第1の検出可能な標識と少なくとも第2の検出可能な標識に結合することができる。二次核酸分子の第2のドメインには、この二次核酸の第2のドメインに対する標識となる短いポリヌクレオチドをハイブリダイズさせることにより、第1の検出可能な標識と少なくとも第2の検出可能な標識を結合させることができる。これらの短いポリヌクレオチドは標識されたオリゴと呼ばれ、蛍光標識したヌクレオチド単量体の直接的な組み込みによって、または当業者に知られている他の核酸標識法によって標識することができる。

【0165】

50

一次核酸分子は、約100個、または約95個、または約90個、または約85個、または約80個、または約75個のヌクレオチドを含むことができる。一次核酸分子は約100個～約80個のヌクレオチドを含むことができる。一次核酸分子は約90個のヌクレオチドを含むことができる。二次核酸分子は、約90個、または約85個、または約80個、または約75個、または約70個のヌクレオチドを含むことができる。二次核酸分子は約90個～約80個のヌクレオチドを含むことができる。二次核酸分子は約87個のヌクレオチドを含むことができる。二次核酸分子は、約25個、または約20個、または約15個、または約10個のヌクレオチドを含むことができる。三次核酸分子は、約20個～約10個のヌクレオチドを含むことができる。三次核酸分子は約15個のヌクレオチドを含むことができる。

【0166】

本開示のレポータプローブとしてさまざまな設計のものが可能である。例えば一次核酸分子は、少なくとも1個（例えば1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはより多い個数）の二次核酸分子にハイブリダイズさせることができる。それぞれの二次核酸分子は、少なくとも1個（例えば1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはより多い個数）の三次核酸分子にハイブリダイズさせることができる。特定の2色組み合わせで標識したレポータプローブを作製するため、その特定の2色の組み合わせの各色で標識した二次核酸分子、または三次核酸分子、または標識したオリゴ、または二次核酸分子と三次核酸分子と標識したオリゴの任意の組み合わせを含むレポータプローブを設計する。例えば図4は、合計で30個の染料（15個の染料が色1、15個の染料が色2）を含む本開示のレポータプローブを示している。異なる蛍光染料の間の色の交換または交差ハイブリダイゼーションを阻止するため、特定の標識または蛍光染料に結合したそれぞれの三次核酸または標識されたオリゴは、独自のヌクレオチド配列を含んでいる。

【0167】

いくつかの態様では、本開示により、5×5レポータプローブが提供される。5×5レポータプローブは一次核酸分子を含んでいて、この一次核酸分子は、12個のヌクレオチドからなる第1のドメインを含んでいる。一次核酸分子は第2のドメインも含んでいて、この第2のドメインは、5個の二次核酸分子にハイブリダイズすることのできるヌクレオチド配列を含んでいる。それぞれの二次核酸分子は、検出可能な標識が結合した5個の三次核酸がそれぞれの二次核酸にハイブリダイズできるようなヌクレオチド配列を含んでいる。

【0168】

いくつかの態様では、本開示により、4×3レポータプローブが提供される。4×3レポータプローブは一次核酸分子を含んでいて、この一次核酸分子は、12個のヌクレオチドからなる第1のドメインを含んでいる。一次核酸分子は第2のドメインも含んでいて、この第2のドメインは、4個の二次核酸分子にハイブリダイズすることのできるヌクレオチド配列を含んでいる。それぞれの二次核酸分子は、検出可能な標識が結合した3個の三次核酸がそれぞれの二次核酸にハイブリダイズできるようなヌクレオチド配列を含んでいる。

【0169】

いくつかの態様では、本開示により、3×4レポータプローブが提供される。3×4レポータプローブは一次核酸分子を含んでいて、この一次核酸分子は、12個のヌクレオチドからなる第1のドメインを含んでいる。一次核酸分子は第2のドメインも含んでいて、この第2のドメインは、3個の二次核酸分子にハイブリダイズすることのできるヌクレオチド配列を含んでいる。それぞれの二次核酸分子は、検出可能な標識が結合した4個の三次核酸がそれぞれの二次核酸にハイブリダイズできるようなヌクレオチド配列を含んでいる。

【0170】

いくつかの態様では、本開示により、スパーサ3×4レポータプローブが提供される。スパーサ3×4レポータプローブは一次核酸分子を含んでいて、この一次核酸分子は、12個のヌクレオチドからなる第1のドメインを含んでいる。一次核酸分子の第1のドメインと第2のドメインの間には、20～40個のヌクレオチドからなるスパーサ領域が位置する。スパーサは、長さが20～40個のヌクレオチドとして同定されるが、スパーサの長さには限定は

なく、20個よりも短いヌクレオチド、または40個よりも長いヌクレオチドが可能である。一次核酸分子の第2のドメインは、3個の二次核酸分子にハイブリダイズすることのできるヌクレオチド配列を含んでいる。それぞれの二次核酸は、検出可能な標識が結合した4個の三次核酸がそれぞれの二次核酸にハイブリダイズできるようなヌクレオチド配列を含んでいる。

【0171】

いくつかの態様では、一次核酸は、長さが12個のヌクレオチドである第1のドメインを含むことができる。しかし一次核酸の第1のドメインの長さに制限はなく、12個未満のヌクレオチド、または13個以上のヌクレオチドが可能である。一例では、一次核酸の第1のドメインは14個のヌクレオチドである。別の一例では、一次核酸の第1のドメインは9個のヌクレオチドである。さらに別の一例では、一次核酸の第1のドメインは8個のヌクレオチドである。レポータプローブの一次核酸の9個のヌクレオチドからなる第1のドメインには、表15に示されているものが含まれる。

【0172】

【表2】

表15

レポータ位置	9 量体配列	色	レポータ位置	9 量体配列	色
1	CATTGGGT	BB	2	CGGGGTTTA	GR
1	CTGGTATGT	BG	2	CAAATTGGT	GY
1	CAGTGAGTG	BR	2	CGAAGTGGT	RR
1	CAGGAAGGT	BY	2	CTGTTAGGG	YR
1	CGATGGATG	GG	2	CGTGTGTGTG	YY
1	CGGTGGAAT	GR	3	CTTTGGTTT	BB
1	CAAAAGAGG	GY	3	CGAGTGGGA	BG
1	CAGGAGAAA	RR	3	CTAGTAGGG	BR
1	CAAGGGTAG	YR	3	CTTTGTGTT	BY
1	CGAGATGAG	YY	3	CATGGGGTG	GG
2	CTTGTGATG	BB	3	CGAAGTTGA	GR
2	CGGGTTAGA	BG	3	CGGTGATTT	GY
2	CGTATGGTT	BR	3	CTATTGTGG	RR
2	CGATTGGTA	BY	3	CTTAGGGAG	YR
2	CATGGTGTA	GG	3	CGGTGGAGG	YY

【0173】

本開示による特定の1つのレポータプローブの設計の任意の特徴を、本開示による別のレポータプローブの設計の任意の特徴と組み合わせることができる。例えば5×5レポータプローブを改変して相補的核酸と一次核酸の間に約20～40個のヌクレオチドのスペーサ領域が含まれるようにすることができる。別の一例では、4×3レポータプローブを改変して、4個の二次核酸が、検出可能な標識に結合した5個の三次核酸がそれぞれの二次核酸にハイブリダイズすることを可能にするヌクレオチド配列を含むように4×5レポータプローブを作り出すことができる。

【0174】

理論に囚われなければ、5×5レポータプローブが含有する蛍光標識（25）は、4×3レポータプローブ（12）よりも数が多いため、5×5レポータプローブの蛍光強度はより大きくなる。所与の任意の視野FOVで検出される蛍光は、多彩な変数（所与のレポータプローブの蛍光強度、そのFOV内の任意により結合している標的分子の数が含まれる）の関数である。視野（FOV）当たりの任意により結合している標的分子の数は、FOV当たり100万～250万個の標的になる可能性がある。FOV当たりの結合している標的分子の数の典型値は、20,000～40,000個、または220,000～440,000個、または100万～200万個の標的分子である。典型的なFOVは0.05 mm²～1 mm²である。典型的なFOVのさらなる例

は $0.05\text{ mm}^2 \sim 0.65\text{ mm}^2$ である。

【0175】

いくつかの態様では、本開示により、二次核酸分子が、三次核酸分子にハイブリダイズしていなくて一次核酸分子の遠位にある「追加ハンドル (extra-handle)」を含むレポータプローブの設計が提供される。いくつかの態様では、「追加ハンドル」は長さが12個のヌクレオチド (「12量体」) が可能だが、その長さ限定されず、12個よりも少ないヌクレオチド、または12個よりも多いヌクレオチドが可能である。それぞれの「追加ハンドル」は、一次核酸分子のうちで二次核酸分子がハイブリダイズする第1のドメインのヌクレオチド配列を含むことができる。したがってレポータプローブが「追加ハンドル」を含んでいるとき、レポータプローブは、一次核酸分子の第1のドメインを介して、または「追加ハンドル」を介してシーケンシング用プローブにハイブリダイズすることができる。したがってレポータプローブがシーケンシング用プローブに結合する確率が大きくなる。「追加ハンドル」のこの設計は、ハイブリダイゼーションの動力学を改善することもできる。理論に囚われなければ、「追加ハンドル」はレポータプローブの相補的核酸の有効濃度を大きくすることができる。 5×4 「追加ハンドル」レポータプローブは、標準的なFOV当たり約4750蛍光カウントを生じさせることが予想される。 5×3 「追加ハンドル」レポータプローブ、 4×4 「追加ハンドル」レポータプローブ、 4×3 「追加ハンドル」レポータプローブ、 3×4 「追加ハンドル」レポータプローブはすべて、標準的なFOV当たり約6000蛍光カウントを生じさせることが予想される。本開示の任意のレポータプローブの設計を改変して「追加ハンドル」が含まれるようにすることができる。

【0176】

レポータプローブの個々の二次核酸分子は、すべて同じ検出可能な標識で標識された三次核酸分子にハイブリダイズすることができる。例えば図5の左図は「 5×6 」レポータプローブを示している。 5×6 レポータプローブは、第2のドメインを含む1個の一次核酸を含んでいて、この第2のドメインは、6個の二次核酸分子にハイブリダイズするヌクレオチド配列を含んでいる。それぞれの二次核酸は、検出可能な標識に結合した5個の三次核酸分子がそれぞれの二次核酸にハイブリダイズするようにするヌクレオチド配列を含んでいる。特定の1個の二次核酸分子に結合した5個の三次核酸分子のそれぞれは、同じ検出可能な標識で標識されている。例えば3個の二次核酸分子が、黄色蛍光染料で標識された三次核酸分子に結合し、他の3個の二次核酸分子が、赤色蛍光染料で標識された三次核酸分子に結合している。

【0177】

レポータプローブの個々の二次核酸分子は、異なる検出可能な標識で標識された三次核酸分子にハイブリダイズすることができる。例えば図5の中央の図は、「 $3 \times 2 \times 6$ 」レポータプローブの設計を示している。 $3 \times 2 \times 6$ レポータプローブは、第2のドメインを含む1個の一次核酸を含んでいて、この第2のドメインは、6個の二次核酸分子にハイブリダイズするヌクレオチド配列を含んでいる。それぞれの二次核酸は、検出可能な標識に結合した5個の三次核酸分子がそれぞれの二次核酸にハイブリダイズするようにするヌクレオチド配列を含んでいる。それぞれの二次核酸は、黄色蛍光染料で標識された三次核酸分子と赤色蛍光染料で標識された三次核酸分子の両方に結合する。この具体例では、3個の二次核酸分子が2個の赤色三次核酸分子と3個の黄色三次核酸分子に結合する一方で、他の3個の二次核酸分子は2個の赤色三次核酸分子と3個の黄色三次核酸分子に結合している。それぞれの二次核酸分子は、異なる検出可能な標識が結合した任意の数の三次核酸分子に結合することができる。図5の中央の図では、個々の二次核酸分子に結合した三次核酸分子は、標識の色が交互になるように配置されている (すなわち赤-黄-赤-黄-赤、または黄-赤-黄-赤-黄)。

【0178】

説明したレポータプローブのどの設計でも、異なる検出可能な標識で標識した三次核酸を二次核酸に沿って任意の順番で配置することができる。例えば図5の右図は、「FRET抵抗性 $3 \times 2 \times 6$ 」レポータプローブを示している。これは $3 \times 2 \times 6$ レポータプローブの設計

に似ているが、それぞれの二次核酸分子に沿った赤色三次核酸分子と黄色三次核酸分子の配置（例えば線形順序またはグループ化）が異なっている。

【0179】

図6は本開示のより代表的なレポータプローブの設計を示しており、ここにはさまざまな三次核酸分子に結合する個々の二次核酸分子が含まれている。左図は、1個の一次核酸分子を含む「 $6 \times 1 \times 4.5$ 」レポータプローブを示している。この一次核酸分子は第2のドメインを含んでいて、この第2のドメインは、6個の二次核酸分子にハイブリダイズするヌクレオチド配列を含んでいる。それぞれの二次核酸分子は5個の三次核酸分子にハイブリダイズする。それぞれの二次核酸分子にハイブリダイズする5個の三次核酸分子のうちの4個は、同じ色の検出可能な標識で直接標識されている。5番目の三次核酸（分岐三次核酸と表記）は、2色組み合わせの他方の色にされた5個の標識されたオリゴに結合している。6個の二次核酸のうちの3個が、2色組み合わせの一方の色（この例では赤色）の標識された分岐三次核酸に結合する一方で、残る3個の二次核酸は、2色組み合わせの他方の色（この例では黄色）の標識された分岐三次核酸に結合している。まとめると、 $6 \times 1 \times 4.5$ レポータプローブは、合計で54個の染料（それぞれの色に27個の染料）で標識されている。図6の中央の図は、「 $4 \times 1 \times 4.5$ 」レポータプローブを示している。この $4 \times 1 \times 4.5$ レポータプローブは、 $6 \times 1 \times 4.5$ レポータプローブと同じ全体的アーキテクチャを共有しているが、一次核酸は4個の二次核酸にだけ結合するため、合計すると36個の染料で、それぞれの色については18個の染料となる点が異なっている。

【0180】

レポータプローブは、2色組み合わせのそれぞれの色について同数の染料を含むことができる。レポータプローブは、2色組み合わせのそれぞれの色について異なる数の染料を含むことができる。レポータプローブ内でどちらの色の染料の数をより多くするかを選択は、2つの染料が吸収する光のエネルギーレベルに基づいてなされる。例えば図6の右図は、「 5×5 エネルギー最適化」レポータプローブの設計を示している。このレポータプローブの設計は、（エネルギーがより高い）15個の黄色染料と（エネルギーがより低い）10個の赤色染料を含んでいる。この例では、15個の黄色染料が第1の標識を構成し、10個の赤色染料が第2の標識を構成することができる。

【0181】

検出可能な部分、または標識、またはレポータは、二次核酸分子、三次核酸分子、標識されたオリゴのいずれかに、さまざまなやり方（検出可能な部分（蛍光部分、測色部分など）の直接的または間接的な付着）で結合することができる。当業者は、核酸の標識に関する参考文献を参照することができる。例えば蛍光部分の非限定的な例に含まれるのは、黄色蛍光タンパク質（YFP）、緑色蛍光タンパク質（GFP）、シアン色蛍光タンパク質（CFP）、赤色蛍光タンパク質（RFP）、ウンベリフェロン、フルオレセイン、フルオレセインイソチオシアネート、ローダミン、ジクロロトリアジニルアミンフルオレセイン、シアニン、ダンシルクロリド、フィコシアニン、フィコエリトリンなどである。

【0182】

蛍光標識と、ヌクレオチドおよび／またはオリゴヌクレオチドへの蛍光標識の付着は、多くの文献に記載されており、その中に含まれるのは、Haugland、『Handbook of Fluorescent Probes and Research Chemicals』、第9版（Molecular Probes, Inc.社、ユージン2002年）；KellerとManak、『DNA Probes』、第2版（Stockton Press社、ニューヨーク、1993年）；Eckstein編、『Oligonucleotides and Analogues: A Practical Approach』（IRL Press社、オックスフォード、1991年）；Wetmur、『Critical Reviews in Biochemistry and Molecular Biology』、第26巻：227～259ページ（1991年）である。本開示に適用できる具体的な方法は、参考文献のサンプルである米国特許第4,757,141号、第5,151,507号、第5,091,519号に記載されている。標識された標的配列のための標識として1つ以上の蛍光染料を用いることができ、その例が開示されているのは、例えば米国特許第5,188,934号（4,7-ジクロロフルオレセイン染料）；第5,366,860号（スペクトル分解可能なローダミン染料）；第5,847,162号（4,7-ジクロロ

ローダミン染料) ; 第4,318,846号(エーテルで置換されたフルオレセイン染料) ; 第5,800,996号(エネルギー移動染料) ; Lee他、第5,066,580号(キサンチン染料) ; 第5,688,648号(エネルギー移動染料) などである。標識操作は、量子ドットを用いて実施することもでき、それが開示されているのは、以下の特許と特許公開である：米国特許第6,322,901号 ; 第6,576,291号 ; 第6,423,551号 ; 第6,251,303号 ; 第6,319,426号 ; 第6,426,513号 ; 第6,444,143号 ; 第5,990,479号 ; 第6,207,392号 ; 米国特許出願公開第2002/0045045号 ; 第2003/0017264号。本明細書では、「蛍光標識」という用語に、1個以上の分子の蛍光吸収および／または発光特性を通じて情報を伝達するシグナル伝達部分が含まれる。このような蛍光特性に含まれるのは、蛍光強度、蛍光の寿命、発光スペクトルの特徴、エネルギー移動などである。

10

【0183】

ヌクレオチド配列および／またはオリゴヌクレオチド配列に組み込むのが容易な市販の蛍光ヌクレオチド類似体の非限定的な例に含まれるのは、Cy3-dCTP、Cy3-dUTP、Cy5-dCTP、Cy5-dUTP (Amersham Biosciences社、ピスカタウェイ、ニュージャージー州)、フルオレセイン-12-dUTP、テトラメチルローダミン-6-dUTP、TEXAS RED (商標) -5-dUTP、CASCADE BLUE (商標) -7-dUTP、BODIPY TMFL-14-dUTP、BODIPY TMR-14-dUTP、BODIPY TMTR-14-dUTP、RHODAMINE GREEN (商標) -5-dUTP、OREGON GREENR (商標) 488-5-dUTP、TEXAS RED (商標) -12-dUTP、BODIPY TM 630/650-14-dUTP、BODIPY TM 650/665-14-dUTP、ALEXA FLUOR (商標) 488-5-dUTP、ALEXA FLUOR (商標) 532-5-dUTP、ALEXA FLUOR (商標) 568-5-dUTP、ALEXA FLUOR (商標) 594-5-dUTP、ALEXA FLUOR (商標) 546-14-dUTP、フルオレセイン-12-UTP、テトラメチルローダミン-6-UTP、TEXAS RED (商標) -5-UTP、mCherry、CASCADE BLUE (商標) -7-UTP、BODIPY TM FL-14-UTP、BODIPY TMR-14-UTP、BODIPY TM TR-14-UTP、RHODAMINE GREEN (商標) -5-UTP、ALEXA FLUOR (商標) 488-5-UTP、ALEXA FLUOR (商標) 546-14-UTP (Molecular Probes, Inc.社、ユージン、オレゴン州) などである。あるいは上記の蛍光体と本明細書で言及している蛍光体を、例えばホスホロアミダイト化学またはNHS化学を利用してオリゴヌクレオチドの合成中に添加することができる。他の蛍光体を有するヌクレオチドをカスタム合成するプロトコルは本分野で知られている (Henegariu他 (2000年) Nature Biotechnol. 第18巻 : 345ページを参照されたい)。2-アミノプリンは、オリゴヌクレオチド配列の合成中にその中に直接組み込むことのできる蛍光塩基である。核酸は、事前にインターカレーティング染料 (DAPI、YOYO-1、臭化エチジウム、シアニン染料 (例えばSYBR Green) など) で染色することもできよう。

20

30

【0184】

合成後に付着させるのに使用できる他の蛍光体の非限定的な例に含まれるのは、ALEXA FLUOR (商標) 350、ALEXA FLUOR (商標) 405、ALEXA FLUOR (商標) 430、ALEXA FLUOR (商標) 532、ALEXA FLUOR (商標) 546、ALEXA FLUOR (商標) 568、ALEXA FLUOR (商標) 594、ALEXA FLUOR (商標) 647、BODIPY 493/503、BODIPY FL、BODIPY R6G、BODIPY 530/550、BODIPY TMR、BODIPY 558/568、BODIPY 558/568、BODIPY 564/570、BODIPY 576/589、BODIPY 581/591、BODIPY TR、BODIPY 630/650、BODIPY 650/665、Cascade Blue、Cascade Yellow、Dansyl、リッサミンローダミンB、Marina Blue、Oregon Green 488、Oregon Green 514、Pacific Blue、Pacific Orange、ローダミン6G、ローダミン・グリーン、ローダミン・レッド、テトラメチルローダミン、Texas Red (Molecular Probes, Inc. 社、ユージン、オレゴン州から入手可能)、Cy2、Cy3、Cy3.5、Cy5、Cy5.5、Cy7 (Amersham Biosciences社、ピスカタウェイ、ニュージャージー州) などである。FRETタンデム蛍光体も使用することができ、その非限定的な例に含まれるのは、PerCP-Cy5.5、PE-Cy5、PE-Cy5.5、PE-Cy7、PE-Texas Red、APC-Cy7、PE-Alexa染料 (610、647、680)、APC-Alexa染料などである。

40

【0185】

50

金属銀粒子または金属金粒子を使用して、蛍光標識したヌクレオチド配列および／またはオリゴヌクレオチド配列からのシグナルを増強することができる（Lakowicz他（2003年）BioTechniques 第34巻：62ページ）。

【0186】

オリゴヌクレオチド配列にとって適切な他の標識に含めることができるのは、フルオレセイン（FAM、FITC）、ジゴキシゲニン、ジニトロフェノール（DNP）、ダンシル、ビオチン、プロモデオキシウリジン（BrdU）、ヘキサヒスチジン（6×His）、リン-アミノ酸（例えばP-tyr、P-ser、P-thr）などである。以下のハブテン／抗体のペア、すなわちビオチン／a-ビオチン、ジゴキシゲニン／a-ジゴキシゲニン、ジニトロフェノール（DNP）／a-DNP、5-カルボキシフルオレセイン（FAM）／a-FAMを検出に使用することができ、このペアの中では、それぞれの抗体が検出可能な標識を用いて誘導体化されている。

【0187】

本明細書に記載の検出可能な標識は、スペクトル分解可能である。複数の蛍光標識に関する「スペクトル分解可能」は、複数の標識の蛍光発光バンドが互いに十分に異なっていて（すなわち非重複が十分であり）、それぞれの標識が発生させる蛍光シグナルに基づいてそれぞれの標識が付着している分子タグを標準的な光検出システムによって（例えば帯域通過フィルタと光増倍管のシステムなどを用いて）識別できることを意味する。そのようなシステムの例は、米国特許第4,230,558号、第4,811,218号など、または『Flow Cytometry: Instrumentation and Data Analysis』（Academic Press社、ニューヨーク、1985年）の中のWheeless他、21～76ページに記載されているシステムである。スペクトル分解可能な有機染料（フルオレセイン、ローダミンなど）とは、発光波長の最大値が互いに少なくとも20 nm離れていること、別の1つの態様では、互いに少なくとも40 nm離れていることを意味する。キレート化ランタニド化合物、量子ドットなどに関するスペクトル分解可能とは、最大発光波長が互いに少なくとも10 nm、または互いに少なくとも15 nm離れていることを意味する。

【0188】

バーコードドメイン内に3つの付着位置が存在していることで、それぞれの付着位置には10通りまでの2色組み合わせが可能であることに基づき、1000通りまでの色組み合わせが存在できる。レポータプローブをプール1つ当たり1000プローブ未満でプールする場合には、エラーを克服するためのパリティチェックを用いる能力を利用することができる。パリティチェックを可能にする多くの潜在的なパリティスキームが存在でき、スキームの一例が図32に示されている。この例では、存在する実際の色はパリティチェックに使用されず、バーコードドメイン内のそれぞれの付着位置における単一色（S）レポータプローブ（例えば赤色）と複数色（M）レポータプローブ（例えば赤色／黄色）の存在が使用される。パリティの設計に見られるように、任意の2つのレポータ位置の状態に関する知識（SまたはM）により、第3の位置の予測が可能になる。示されている例では、任意の2つの位置でSが観察されると、観察されていない位置はMでなければならない、任意の2つの位置でSとMが観察されると、他の位置はSでなければならないことを意味し、2つのMレポータプローブが観察されると、他の位置はMでなければならない。これは、検出されたレポータの色が正確でない3つのレポータプローブのコードを得るには、2回の正しくない読み出しがなされなければならないことを意味する。図32は、レポータプローブのエラーが5%であるときのシミュレーションの結果を示しており、パリティチェックを適用したときにエラーの除外が増加することがわかる。適用可能な多数のパリティシステムが存在しており、これはほんの一例である。

【0189】

別のエラー訂正ルーチンは、レポータプローブの各プールのための色パレットを交換するというものである。色パレットは、プールを測定するのに実際に用いられるレポータプローブのセットである。どのプールでも使用されないレポータプローブが多数あり、500個のレポータプローブが1つのプールに存在している場合には、可能な色組み合わせの半分だけが必要とされる。これを実現する最も簡単なやり方は、パレットを2つにすること

であり、パレットAは500個のレポータプローブを含有し、パレットBは別の500個のレポータプローブを含有する。したがってシーケンシングプール1、3、5、7がパレットAを有し、プール2、4、6、8がパレットBを有する場合には、ランニングプールの順番が1、2、3、4、5、6、7、8であるというのは、連続したそれぞれのシーケンシングプールが別々のパレットを有することを意味する。したがってプール2からのバーコードは、直前のプールと直後のプール（例えばプール1と3）には存在しないことになる。こうすることで、トラブル対応を簡単に自動化することと、エラー検出を制限することが可能になる。

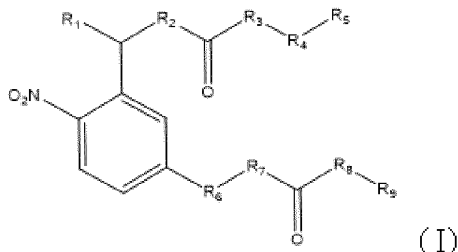
【0190】

レポータプローブは、切断可能なリンカーの修飾を1つ以上含むことができる。切断可能なリンカーのこの1つ以上の修飾は、レポータプローブ内の任意の場所に位置させることができる。切断可能なリンカーの修飾は、レポータプローブの一次核酸分子の第1のドメインと第2のドメインの間に位置させることができる。切断可能なリンカーの修飾は、レポータプローブの二次核酸分子の第1のドメインと第2のドメインの間に位置させることができる。切断可能なリンカーの修飾は、レポータプローブの一次核酸分子と二次核酸分子の第1のドメインと第2のドメインの間に位置させることができる。図7の左図は本開示の代表的なレポータプローブを示しており、一次核酸の第1のドメインと第2のドメインの間の切断可能なリンカーの修飾と、二次核酸の第1のドメインと第2のドメインの間の切断可能なリンカーの修飾を含んでいる。図7の左図に例示されているようなケースでは、切断可能なリンカーの修飾は、1つ以上の切断可能な部分（図7の左図に例示されているようなもの）を含むことができる。

【0191】

切断可能なリンカーの修飾として、式 (I) の化合物：

【化3】



またはその立体異性体、またはその塩が可能である。ただしこの式において、 R_1 は、水素、ハロゲン、 C_{1-6} アルキル、 C_{2-6} アルケニル、 C_{2-6} アルキニルのいずれかであり、その中の C_{1-6} アルキル、 C_{2-6} アルケニル、 C_{2-6} アルキニルはそれぞれ独立に、任意により少なくとも1個の置換基 R_{10} で置換されており； R_2 は、O、NH、N(C_{1-6} アルキル)のいずれかであり； R_3 は、シクロアルキル、ヘテロシクロアルキル、アリール、ヘテロアリエールのいずれかであり、そのそれぞれが任意により少なくとも1個の置換基 R_{10} で置換されており；それぞれの R_4 と R_7 は、独立に、 C_{1-6} アルキル、 C_{2-6} アルケニル、 C_{2-6} アルキニルであり、その中の C_{1-6} アルキル、 C_{2-6} アルケニル、 C_{2-6} アルキニルは、それぞれ独立に、任意により少なくとも1個の置換基 R_{10} で置換されており； R_5 と R_9 は、それぞれ独立に、シクロアルキル、ヘテロシクロアルキル、アリール、ヘテロアリエールのいずれかであり、そのそれぞれは、任意により少なくとも1個の置換基 R_{10} で置換されており； R_6 は、O、NH、N(C_{1-6} アルキル)のいずれかであり； R_8 は、O、NH、N(C_{1-6} アルキル)のいずれかであり；それぞれの R_{10} は、独立に、水素、ハロゲン、 $-C_{1-6}$ アルキル、 $-C_{2-6}$ アルケニル、 $-C_{2-6}$ アルキニル、ハロ C_{1-6} アルキル、ハロ C_{2-6} アルケニル、ハロ C_{2-6} アルキニル、シクロアルキル、ヘテロシクリル、アリール、ヘテロアリエール、 $-CN$ 、 $-NO_2$ 、オキソ、 $-OR_{11}$ 、 $-SO_2R_{11}$ 、 $-SO_3^-$ 、 $-COR_{11}$ 、 $-CO_2R_{11}$ 、 $-CONR_{11}R_{12}$ 、 $-C(=NR_{11})NR_{12}R_{13}$ 、 $-NR_{11}R_{12}$ 、 $-NR_{12}COR_{12}$ 、 $-NR_{11}CONR_{12}R_{13}$ 、 $-NR_{11}CO_2R_{12}$ 、 $-NR_{11}SONR_{12}R_{13}$ 、 $-NR_{11}$

$\text{SO}_2\text{NR}_{12}\text{R}_{13}$ 、 $-\text{NR}_{11}\text{SO}_2\text{R}_{12}$ のいずれかであり； R_{11} 、 R_{12} 、 R_{13} は同じでも異なってもよく、それぞれ独立に、水素、 $-\text{C}_{1-6}$ アルキル、 $-\text{C}_{2-6}$ アルケニル、 $-\text{C}_{2-6}$ アルキニル、ハロ C_{1-6} アルキル、ハロ C_{2-6} アルケニル、ハロ C_{2-6} アルキニル、 C_{1-6} アルキルオキシ C_{1-6} アルキル、シクロアルキル、ヘテロシクリル、アリール、ヘテロアリールのいずれかである。

【0192】

1つの態様では、 R_1 は C_{1-6} アルキル、好ましくは C_{1-3} アルキル（メチル、エチル、プロピル、イソプロピルなど）であり； R_2 は NH または $\text{N}(\text{C}_{1-6}\text{アルキル})$ であり； R_3 は5員～6員のシクロアルキル、好ましくはシクロヘキシルであり； R_4 は C_{1-6} アルキレン、好ましくは C_{1-3} アルキレン（メチレン、エチレン、プロピレン、イソプロピレンなど）であり； R_5 は、1個の窒素原子を含むとともに N 、 O 、 S から選択された0個または1個の追加のヘテロ原子を含む5員～6員のヘテロシクリルであり、このヘテロシクリルは、任意により1個または2個の R_{10} で置換されており； R_6 は O であり； R_7 は C_{1-6} アルキレン、好ましくは C_{1-3} アルキレン（メチレン、エチレン、プロピレン、イソプロピレンなど）であり； R_8 は O であり； R_9 は、1個の窒素原子を含むとともに N 、 O 、 S から選択された0個または1個の追加のヘテロ原子を含む5員～6員のヘテロシクリルであり、このヘテロシクリルは、任意により1個または2個の R_{10} で置換されており；それぞれの R_{10} は、独立に、ハロゲン、 C_{1-6} アルキル、ハロ C_{1-6} アルキル、オキソ、 $-\text{SO}_2\text{H}$ 、 $-\text{SO}_3^-$ のいずれかである。

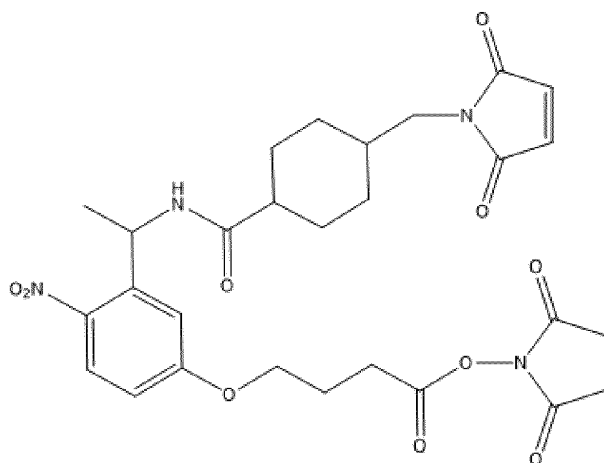
【0193】

1つの態様では、 R_3 はシクロヘキシルであり、 R_4 はメチレンであり、 R_5 は1H-ピロール-2,5-ジオンであり、 R_9 は、任意により SO_3^- で置換されたピロリジン-2,5-ジオンである。

【0194】

リンカー化合物として、

【化4】



またはその立体異性体、またはその塩が可能である。

【0195】

リンカー化合物として、

10

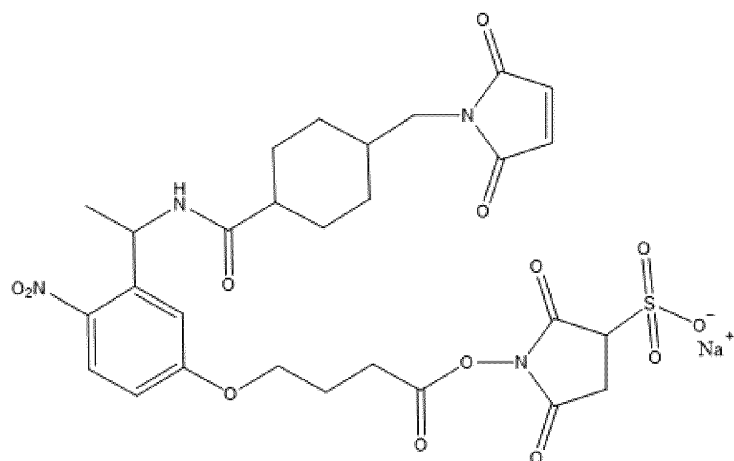
20

30

40

50

【化 5】



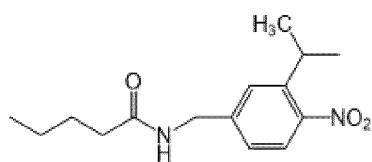
10

またはその立体異性体、またはその塩が可能である。

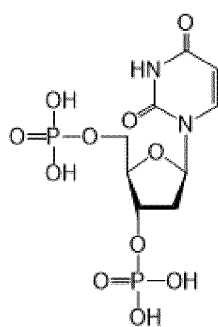
【 0 1 9 6】

リンカー化合物またはリンカーの修飾として、

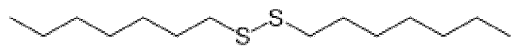
【化 6】



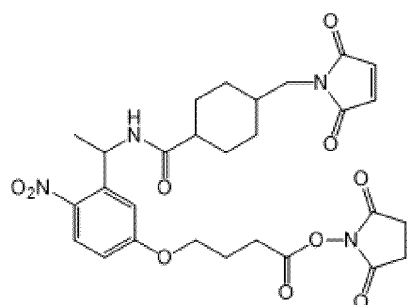
20



30



または



40

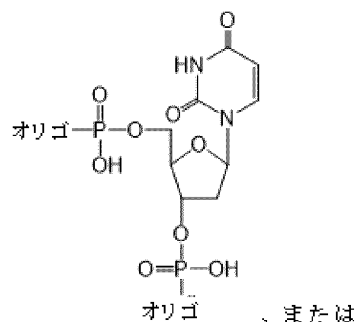
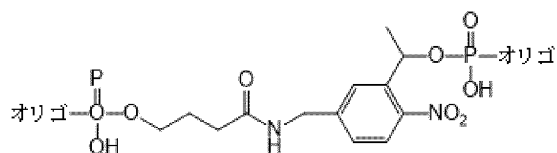
のいずれかが可能である。

【 0 1 9 7】

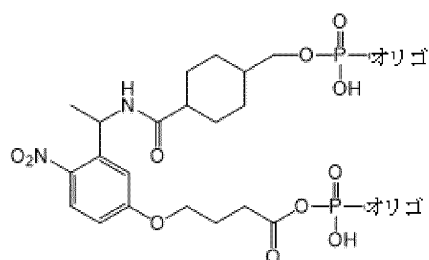
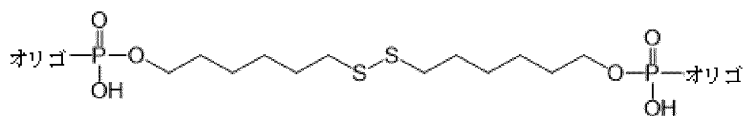
リンカー化合物またはリンカーの修飾として、

50

【化 7】



10



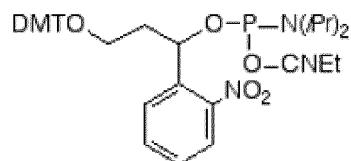
20

のいずれかが可能である。

【0198】

切断可能なリンカーの修飾または切断可能な部分として、

【化 8】



30

が可能である。

【0199】

レポータープローブは、3種類の貯蔵溶液を水と混合することによって構成することができる。1つの貯蔵溶液は一次核酸分子を含有し、1つの貯蔵溶液は二次核酸分子を含有し、最後の貯蔵溶液は三次核酸分子を含有している。表2は、個々の設計のレポータープローブを構成するために混合することのできる各貯蔵溶液の代表的な量を示している。

【0200】

40

【表 3】

表 2

レポータープロ ーブの設計	一次核酸分子（ 10 μ M の貯蔵溶 液）の体積（ μ l）	二次核酸分子（ 10 μ M の貯蔵溶 液）の体積（ μ l）	三次核酸分子（ 10 μ M の貯蔵溶 液）の体積（ μ l）	水の体積（ μ l ）
5 $\times$ 4	1	4.5	2.25	92.25
5 $\times$ 3	1	4.5	1.8	92.7
4 $\times$ 4	1.28	4.5	2.25	91.97
4 $\times$ 3	1.28	4.5	1.8	92.42
3 $\times$ 4	1.8	4.5	2.25	91.45

【0201】

標的核酸

【0202】

本開示により、本明細書に開示されているシーケンシング用プローブを用いて核酸をシーケンシングする方法が提供される。本開示の方法を用いてシーケンシングする核酸を、本明細書では「標的核酸」と呼ぶ。「標的核酸」という用語は、本開示のプローブ、方法、装置によって配列を決定する核酸分子（DNA、RNA、PNA）を意味する。一般に、「標的核酸」、「標的核酸分子」、「標的核酸配列」、「標的核酸断片」、「標的オリゴヌクレオチド」、「標的ポリヌクレオチド」という用語は交換可能に用いられ、その非限定的な例に含まれるのは、さまざまな長さが可能なポリマー形態のヌクレオチド（デオキシリボヌクレオチドまたはリボヌクレオチド）、またはその類似体である。核酸の非限定的な例に含まれるのは、遺伝子、遺伝子断片、エキソン、イントロン、遺伝子間DNA（その非限定的な例にヘテロクロマチンDNAが含まれる）、メッセンジャーRNA（mRNA）、トランスファーRNA、リボソームRNA、リボザイム、小さな干渉性RNA（siRNA）、非コードRNA（ncRNA）、cDNA、組み換えポリヌクレオチド、分岐したポリヌクレオチド、プラスミド、ベクター、1つの配列から単離されたDNA、1つの配列から単離されたRNA、核酸プローブ、プライマーである。本開示の方法を利用してシーケンシングする前に、標的核酸の属性および／または配列は既知である。あるいは属性および／または配列は未知である。本開示の方法を利用してシーケンシングする前に標的核酸の配列の一部が既知であることも可能である。例えばこの方法を適用して、既知の標的核酸分子の中の点変異を明らかにすることができる。

【0203】

本開示の方法は、サンプル（例えば生物からのサンプル）から得られた核酸分子を直接シーケンシングしており、好ましいことに変換（または増幅）工程は存在しない。一例として、RNAに基づく直接的シーケンシングでは、本開示の方法は、配列を取得できるようになる前にRNA分子からDNA分子への（すなわちcDNAの合成を通じた）変換を必要としない。増幅または変換が必要ないため、本開示でシーケンシングされる核酸は、核酸がサンプル中に存在するときや核酸がサンプルから得られたときには、その核酸の中に存在しているあらゆる独自の塩基および／またはエピジェネティックなマーカーを保持していることになる。このような独自の塩基および／またはエピジェネティックなマーカーは、本分野で知られているシーケンシング法では失われる。

【0204】

本開示の方法を利用すると、1分子の分解能でシーケンシングすることができる。言い換えると、本開示の方法により、利用者は、異なる標的核酸分子からのデータを組み合わせる必要がなく、単一の標的核酸分子から回収したデータに基づいて最終的な配列を生成させることが可能になるため、その特定の標的の独自の特徴はすべて保持されている。

【0205】

標的核酸は、核酸の任意のサンプルまたは供給源（例えば任意の細胞、組織、生物、インビトロ、化学的合成装置など）から取得することができる。標的核酸は、本分野で知られている任意の方法によって取得することができる。核酸は、臨床対象の血液サンプルから取得することができる。核酸は、本分野で周知の方法とキットを用いて供給源またはサンプルから抽出、または単離、または精製することができる。

【0206】

標的核酸は、本分野で知られている任意の手段によって断片化することができる。断片化は、酵素的手段または機械的手段によって実施することが好ましい。機械的手段として、超音波処理または物理的剪断が可能である。酵素的手段は、ヌクレアーゼ（例えばデオキシリボヌクレアーゼI（DNアーゼI））または1つ以上の制限エンドヌクレアーゼを用いた消化によって実施することができる。

10

【0207】

標的核酸を含む核酸分子が1つの完全な染色体であるときには、その染色体の断片化を避ける工程を実施する必要がある。

【0208】

本分野で周知のように、標的核酸は、天然のヌクレオチドまたは非天然のヌクレオチドを含むことができ、その中には修飾されたヌクレオチドまたは核酸類似体が含まれる。

【0209】

標的核酸分子に含めることができるのは、長さが数百キロベース（例えば1キロベース、2キロベース、3キロベース、4キロベース、5キロベース、10キロベース、20キロベース、30キロベース、40キロベース、50キロベース、100キロベース、200キロベース、500キロベース、またはそれよりも大きな数のキロベース）までのDNA分子、RNA分子、PNA分子である。標的核酸分子は、約50個～約400個のヌクレオチド、または約90個～約350個のヌクレオチドを含むことができる。

20

【0210】

捕捉用プローブ

【0211】

標的核酸分子は、基板に（例えば1つ、2つ、3つ、4つ、5つ、6つ、7つ、8つ、9つ、10、またはそれよりも多くの位置で）固定化することができる。

【0212】

有用な基板の代表例に含まれるのは結合部分を含む基板であり、その結合部分の選択は、リガンド、抗原、炭水化物、核酸、受容体、レクチン、抗体からなるグループからなされる。捕捉用プローブは、基板の結合部分に結合することのできる基板結合部分を含んでいる。反応性部分を含む有用な代表的な基板の非限定的な例に含まれるのは、エポキシ、アルデヒド、金、ヒドラジド、スルフヒドリル、NHS-エステル、アミン、アルキン、アジド、チオール、カルボン酸塩、マレイミド、ヒドロキシメチルホスフィン、イミドエステル、イソシアネート、ヒドロキシル、ペンタフルオロフェニル-エステル、ソラーレン、ピリジルジスルフィドまたはビニルスルホン、ポリエチレングリコール（PEG）、ヒドロゲルのいずれか、またはこれらの混合物を含む表面である。このような表面は、市販品提供者から入手すること、または標準的な技術に従って用意することができる。反応性部分を含む有用な代表的な基板の非限定的な例に含まれるのは、OptArray-DNA NHS基（Accler 8社）、Nexterion Slide AL（Schott社）、Nexterion Slide E（Schott社）である。

30

【0213】

基板として、標的核酸を固定化することのできる本分野で知られている任意の固体支持体（例えば被覆されたスライド、マイクロ流体装置）が可能である。基板として、表面、膜、ビーズ、多孔性材料、電極、アレイが可能である。基板として、例えばポリマー材料、金属、ケイ素、ガラス、石英が可能である。標的核酸は、当業者にとって明らかな任意の基板の表面に固定化することができる。

40

【0214】

基板がアレイであるときには、基板は複数のウェルを含むことができる。ウェルのサイ

50

【0219】

ローンオリゴヌクレオチドまたはアダプタオリゴヌクレオチドに含まれるそれぞれの核酸として、独立に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。ローンオリゴヌクレオチドまたはアダプタオリゴヌクレオチドで有用な典型的な修飾されたヌクレオチドまたは核酸類似体は、イソグアニンとイソシトシンである。あるいはさらに、ローンオリゴヌクレオチドに含まれるそれぞれの核酸として、独立に、L-DNAが可能である。いくつかの態様では、ローンオリゴヌクレオチド (lawn oligonucleotide) はL-DNAを含むことができる。ローンオリゴヌクレオチドはL-DNAからなることができる。ローンオリゴヌクレオチドは主にL-DNAからなることができる。修飾されたヌクレオチドまたはヌクレオチド類似体 (イソグアニン、イソシトシン、L-DNAなど) を用いると、例えばローンオリゴヌクレオチド内の適切な相補的核酸配列に対するアダプタオリゴヌクレオチドの結合の効率と精度を改善しつつ、他の場所への結合を最少にすることができる。

10

【0220】

ローンオリゴヌクレオチドはさらに、6炭素リンカーを有する5'アミンを含んでいる (本明細書では5AmMC6と呼ぶ)。5AmMC6を用いてローンオリゴヌクレオチドを基板に付着させることができる。

【0221】

捕捉用プローブとアダプタオリゴヌクレオチドとローンオリゴヌクレオチドの複合体の一例が図33に示されている。この図では、標的核酸 (この例では遺伝子TP53.1) にハイブリダイズする代表的なアダプタの配列と代表的な捕捉用プローブの配列は緑色であり、代表的なローンオリゴの逆相補体である配列は青色であり、標的核酸にハイブリダイズする捕捉用プローブ上の代表的な配列は赤色である。代表的な捕捉用プローブの配列は、3' - CCGGTCAACCGTTTTGTAGAACAACCTCCCGTCCCTCACTCACTAGCCTCCAGTACCGA AAGC-5' (配列番号111) である。代表的なアダプタの配列は、5' -GAGTGATCGGAG GTCATGGCTTTCGAC/iMe-イソdC/CTA/iMe-イソdC/AAA/iMe-イソdC/TCA/iMe-イソdC/TA/iMe-イソdC/TA/iMe-イソdC/CAA/iMe-イソdC/AAC/iMe-イソdC/TCA/iMe-イソdC/CA-3' (配列番号110) である。代表的なローンオリゴヌクレオチドの配列は、TG/iイソdG/GAT/iイソdG/TTT/iイソdG/AGT/iイソdG/AT/iイソdG/AT/iイソdG/GT T/iイソdG/TTG/iイソdG/AGT/iイソdG/GT/5AmMC6 (配列番号108) である。

20

30

【0222】

いくつかの態様では、ローンオリゴヌクレオチドは、少なくとも1個の親和性部分、または少なくとも2個の親和性部分、または少なくとも3個の親和性部分、または少なくとも4個の親和性部分、または少なくとも5個の親和性部分、または少なくとも6個の親和性部分、または少なくとも7個の親和性部分、または少なくとも8個の親和性部分、または少なくとも9個の親和性部分、または少なくとも10個の親和性部分を含むことができる。親和性部分としてビオチンが可能である。したがってローンオリゴヌクレオチドは、少なくとも1個のビオチン部分、または少なくとも2個のビオチン部分、または少なくとも3個のビオチン部分、または少なくとも4個のビオチン部分、または少なくとも5個のビオチン部分、または少なくとも6個のビオチン部分、または少なくとも7個のビオチン部分、または少なくとも8個のビオチン部分、または少なくとも9個のビオチン部分、または少なくとも10個のビオチン部分を含むことができる。

40

【0223】

いくつかの態様では、標的核酸にハイブリダイズする本開示の捕捉用プローブは、少なくとも1つの親和性部分 (その非限定的な例はビオチン部分) を含むことができる。標的核酸にハイブリダイズする捕捉用プローブは次いで基板上の少なくとも1つのローンオリゴヌクレオチドに直接的または間接的にハイブリダイズすることができる。この少なくとも1つのローンオリゴヌクレオチドは、少なくとも1つの親和性部分 (その非限定的な例はビオチン部分) を含んでいる。捕捉用プローブがローンオリゴヌクレオチドにハイブリダイズした後、得られた捕捉用プローブ-標的核酸-ローンオリゴヌクレオチド複合体は第2の

50

親和性部分とともにインキュベートすることができ、第2の親和性部分は、捕捉用プローブ上に位置する第1の親和性部分と、ローンオリゴヌクレオチド上に位置する第1の親和性部分に結合することができる。非限定的な一例では、捕捉用プローブ上に位置する第1の親和性部分と、ローンオリゴヌクレオチド上に位置する第1の親和性部分の両方がビオチンである場合には、ニュートラアビジンを第2の親和性部分として用いることができる。この第2の親和性部分は、捕捉用プローブ上に位置する第1の親和性部分と、ローンオリゴヌクレオチド上に位置する第1の親和性部分に結合し、本明細書で「タンパク質ロック」と呼ぶタンパク質架橋を作り出す。タンパク質ロックを用いて基板上に標的核酸をより安定に固定化することができる。図67は、ビオチニル化された捕捉用プローブとローンオリゴヌクレオチドとニュートラアビジンを用いたタンパク質ロックの模式図を示している。

10

【0224】

標的核酸には1つ以上の捕捉用プローブ（すなわち2個、3個、4個、5個、6個、7個、8個、9個、10個、またはそれよりも多い個数の捕捉用プローブ）を結合させることができる。捕捉用プローブは、標的核酸の一部と相補的なドメインと、基板結合部分を含むドメインを含んでいる。標的核酸のうちで捕捉用プローブと相補的な部分として、標的核酸の一端、または末端近くでない部分が可能である。捕捉用プローブは、標的核酸の一部と相補的なドメインと、基板結合部分を含むドメインの間に切断可能な部分を含むことができる。

【0225】

あるいは捕捉用プローブは、標的核酸の一部と相補的な第1のドメインと、基板結合部分を含む第2のドメインと、異なる基板結合部分を含む第3のドメインを含むことができる。捕捉用プローブは、任意のドメイン間に切断可能な部分を含むことができる。

20

【0226】

捕捉用プローブは、5'末端をリン酸化することができる。あるいは捕捉用プローブは、少なくとも1つのホスホロチオエート結合を含むことができる。捕捉用プローブは、少なくとも2つのホスホロチオエート結合を含むことができる。これら少なくとも2つのホスホロチオエート結合は、捕捉用プローブの5'末端に位置することが好ましい。

【0227】

捕捉用プローブの基板結合部分としてビオチンが可能であり、基板としてアビジン（例えばストレプトアビジン）が可能である。アビジンを含む有用な基板は市販されており、その中に含まれるのは、TB0200（Accelr8社）、SAD6、SAD20、SAD100、SAD500、SAD2000（Xantec社）、SuperAvidin（Array-It社）、ストレプトアビジン・スライド（カタログ番号MPC 000、Xenopore社）、STREPTAVIDINnslide（カタログ番号439003、Greiner Bio-one社）である。捕捉用プローブの基板結合部分としてアビジン（例えばストレプトアビジン）が可能であり、基板としてビオチンが可能である。ビオチンを含む市販の有用な基板の非限定的な例に含まれるのは、Optiarray-ビオチン（Accler8社）、BD6、BD20、BD100、BD500、BD2000（Xantec社）である。

30

【0228】

捕捉用プローブの基板結合部分として、光活性化によって基板に結合することのできる反応性部分が可能である。基板が光活性化部分を含むこと、またはナノレポータの第1の部分が光活性化部分を含むことができる。光活性化部分のいくつかの例に含まれるのは、アジ化アリール（4-アジド-2,3,5,6-テトラフルオロ安息香酸など）；ベンゾフェノン系試薬（4-ベンゾイル安息香酸スクシンイミジルエステルなど）；5-プロモ-デオキシウリジンである。

40

【0229】

捕捉用プローブの基板結合部分として、基板の相補的な結合部分にハイブリダイズすることができる核酸が可能である。捕捉用プローブの基板結合部分に含まれるそれぞれの核酸として、独立に、正規塩基、または修飾されたヌクレオチド、または核酸類似体が可能である。捕捉用プローブの基板結合部分内の少なくとも1個、または少なくとも2個、または少なくとも3個、または少なくとも4個、または少なくとも5個、または少なくとも6個

50

のヌクレオチドとして、修飾されたヌクレオチドまたは核酸類似体が可能である。捕捉用プローブの基板結合部分内の修飾されたヌクレオチドまたはヌクレオチド類似体と正規塩基の典型的な比率は、1：2～1：8である。捕捉用プローブの基板結合部分で有用な典型的な修飾されたヌクレオチドまたはヌクレオチド類似体は、イソグアニンとイソシトシンである。

【0230】

捕捉用プローブの基板結合部分は、当業者にとって明らかな他の結合ペアを介して基板に固定化することができる。標的核酸は、基板に結合させた後、標的核酸を伸長させるのに十分な力（例えば重力、流体力、電磁力「エレクトロストレッチング」、フロー-ストレッチング、後退メニスカス（receding meniscus）技術、これらの組み合わせ）を加えることによって伸長させることができる。捕捉用プローブは、標識可能な標識（すなわち基準スポット）を含むこと、または標識可能な標識と会合させることができる。

【0231】

標的核酸には、この標的核酸の第2の部分と相補的なドメインを含む第2の捕捉用プローブが結合することができる。第2の捕捉用プローブが結合する標的核酸の第2の部分は、第1の捕捉用プローブが結合する標的核酸の第1の部分とは異なっている。この部分として、標的核酸の一端、または末端近くではない部分が可能である。第2の捕捉用プローブの結合は、標的核酸の伸長後に起こること、または伸長中に起こること、または伸長しなかった標的核酸に対して起こることができる。第2の捕捉用プローブは、上記のような結合を有することができる。

【0232】

標的核酸には、標的核酸の第3の部分、または第4の部分、または第5の部分、または第6の部分、または第7の部分、または第8の部分、または第9の部分、または第10の部分と相補的なドメインを含む第3の捕捉用プローブ、または第4の捕捉用プローブ、または第5の捕捉用プローブ、または第6の捕捉用プローブ、または第7の捕捉用プローブ、または第8の捕捉用プローブ、または第9の捕捉用プローブ、または第10の捕捉用プローブが結合することができる。この部分として、標的核酸の一端、または末端ではない部分が可能である。第3の捕捉用プローブ、または第4の捕捉用プローブ、または第5の捕捉用プローブ、または第6の捕捉用プローブ、または第7の捕捉用プローブ、または第8の捕捉用プローブ、または第9の捕捉用プローブ、または第10の捕捉用プローブの結合は、標的核酸の伸長後に起こること、または伸長中に起こること、または伸長しなかった標的核酸に対して起こることができる。第3の捕捉用プローブ、または第4の捕捉用プローブ、または第5の捕捉用プローブ、または第6の捕捉用プローブ、または第7の捕捉用プローブ、または第8の捕捉用プローブ、または第9の捕捉用プローブ、または第10の捕捉用プローブは、上記のような結合を有することができる。

【0233】

捕捉用プローブは、サンプルから標的核酸を単離することができる。ここでは、標的核酸を含むサンプルに捕捉用プローブを添加する。捕捉用プローブは、この捕捉用プローブのうちで標的核酸の1つの領域と相補的な領域を介して標的核酸に結合する。標的核酸が、捕捉用プローブの基板結合部分に結合する部分を含む基板と接触するとき、核酸は基板の表面に固定化される。

【0234】

図8は、本開示による2捕捉用プローブシステムを用いた標的核酸の捕捉を示している。ゲノムDNAを95℃で変性させ、捕捉用試薬のプールにハイブリダイズさせる。捕捉用試薬のこのプールは、オリゴヌクレオチドであるプローブA、プローブB、アンチセンスブロックプローブを含んでいる。プローブAは、このプローブの3'末端にあるビオチン部分と、標的核酸の5'末端と相補的な配列を含んでいる。プローブBは、このプローブの5'末端に常磁性ビーズによって結合させることのできる精製結合配列と、標的核酸の3'末端と相補的なヌクレオチド配列を含んでいる。アンチセンスブロックプローブは、標的核酸のうちのシーケンシングする部分のアンチセンス鎖と相補的なヌクレオチド配列を含んでいる

。捕捉用試薬にハイブリダイズさせた後、標的核酸上でハイブリダイズしたプローブAとプローブBに挟まれた位置にシーケンシングウインドウを作る。プローブBの5'配列に結合する常磁性ビーズを用いて標的核酸を精製する。過剰なあらゆる捕捉用試薬または相補的アンチセンスDNA鎖を洗い流すと、求める標的核酸が精製される。次に、精製された標的核酸に、ハイブリダイズしたプローブA上のビオチン部分（ストレプトアビジンなど）に結合することのできる表面を含むフローセルを通過させる。すると標的核酸の一端がフローセルの表面とつながる。他端を捕捉するには、標的核酸をフロー-ストレッチングさせ、プローブBの精製結合配列と相補的なビオチニル化プローブを添加する。ビオチニル化プローブは、プローブBの精製結合配列とハイブリダイズするとフローセルの表面に結合することができ、その結果として、伸長していてフローセルの表面に両端が結合した捕捉された標的核酸になる。

【0235】

利用者が高度に断片化されたサンプルからできるだけ多くの標的核酸分子を確実に「捕捉する」ためには、複数の捕捉用プローブが含まれていて、それぞれの捕捉用プローブが標的核酸の異なる領域と相補的になるようにすることが有用である。例えば捕捉用プローブの3つのプールが存在することが可能であり、そのうちの第1のプールは標的核酸のうちに5'末端近くの領域と相補的であり、第2のプールは標的核酸のうちに中央部の領域と相補的であり、第3のプールは標的核酸のうちに3'末端近くの領域と相補的である。これは、1つの標的核酸につき「目的のn個の領域」に一般化することができる。この例では、断片化された標的核酸の個々のプールが、ビオチンタグを含むかビオチンタグに結合した捕捉用プローブに結合する。入力サンプルの1/n（nは標的核酸内の異なる領域の数）を各プール室のために単離する。捕捉用プローブは目的の標的核酸に結合する。次に標的核酸を、捕捉用プローブのビオチンを介し、基板に付着させたアビジン分子に固定化する。任意により、標的核酸を例えば流体力または静電力を通じて伸長させる。n個のプールすべてを同時に伸長させて結合させることや、完全に伸長した分子の数を最大にするため、（最も5'側の領域を捕捉する）プール1を伸長させて最初に結合させた後、（標的の中央領域を捕捉する）プール2を伸長させて結合させ、最後にプール3を伸長させて結合させることができる。

【0236】

標的核酸は、本開示の「ビーズに基づく2工程精製」システムを用いて捕捉することができる。4つの捕捉用プローブ、すなわちプローブA、プローブB、プローブC、プローブDが存在している。プローブAは、標的核酸の5'末端と相補的な核酸配列であるOA配列と、ビオチン部分に付着した核酸配列を含んでいる。OA配列はヌクレオチド配列CGAAAGCCATGACCTCCGATCACTC（配列番号109）を含むことができ、ローンオリゴヌクレオチドに結合することができる。ビオチン部分に付着した核酸配列は、切断可能なリンカーを介して標的核酸の5'末端と相補的な核酸配列に接続される。プローブBとプローブCは、標的核酸と相補的な核酸配列と、ビオチン部分に付着した核酸配列を含んでいる。ビオチン部分に付着した核酸配列は、切断可能なリンカーを介して標的核酸と相補的な核酸配列に接続される。プローブDは、標的核酸の3'末端と相補的な核酸配列（G配列と名づける精製結合配列）と、ビオチン部分を含んでいる。ビオチン部分は切断可能なリンカーを介してG配列に接続される。上記の4つの捕捉用プローブは最初に標的核酸にハイブリダイズする。すべてのプローブが、標的核酸に沿って重複していない位置にハイブリダイズし、プローブBとプローブCは、プローブAとプローブDの間にハイブリダイズする。その後、標的核酸は、捕捉用プローブ上のビオチン部分に結合したストレプトアビジン常磁性ビーズを用いて精製される。標的でない過剰なゲノムDNAはビーズから洗い流される。その後、標的核酸-捕捉用プローブ複合体は、各捕捉用プローブ内の切断可能なリンカーが切断されることによってストレプトアビジン磁性ビーズから放出される。標的核酸-捕捉用プローブ複合体はさらに、プローブD上の精製G配列に結合した常磁性ビーズを用いて精製される。過剰な捕捉用プローブは洗い流され、標的核酸-捕捉用プローブ複合体が常磁性ビーズから分離される。

【0237】

標的核酸は、本開示の「λエキソヌクレアーゼを用いたビーズに基づく1工程精製」システムを用いて捕捉することができる。4つの捕捉用プローブ、すなわちプローブA、プローブB、プローブC、プローブDが存在している。プローブAは、標的核酸の5'末端と相補的な配列を含んでいる。プローブAの5'末端は、2つのホスホロチオエート結合を含んでいる。プローブB、プローブC、プローブDは、ビオチン部分に付着した核酸配列をプローブの3'末端に、標的核酸と相補的な核酸配列をプローブの5'末端に含んでいる。プローブB、プローブC、プローブDの5'末端はリン酸化されている。プローブA、プローブB、プローブC、プローブDは、標的核酸に沿って重複しない位置にハイブリダイズする。プローブが標的核酸にハイブリダイズした後、ストレプトアビジン常磁性ビーズを用いて標的核酸を精製する。過剰なgDNAと捕捉用プローブを洗い流す。標的核酸-捕捉用プローブ複合体をビーズから溶離させる。その後、5'末端がリン酸化された二本鎖DNAを選択的に分解するλエキソヌクレアーゼを用いてプローブB、プローブC、プローブDを消化させる。

【0238】

標的核酸は、本開示の「FEN1を用いたビーズに基づく1工程精製」システムを用いて捕捉することができる。4つの捕捉用プローブ、すなわちプローブA、プローブB、プローブC、プローブDが存在している。プローブAは、標的核酸にハイブリダイズしない3'核酸配列と、核酸配列の5'末端と相補的な核酸配列と、ビオチン部分を含んでいて標的核酸にハイブリダイズしない5'核酸配列を含んでいる。プローブBとプローブCは、標的核酸にハイブリダイズしない3'核酸配列と、核酸配列と相補的な核酸配列と、ビオチン部分を含んでいて標的核酸にハイブリダイズしない5'核酸配列を含んでいる。プローブDは、標的核酸にハイブリダイズしない3'配列と、標的核酸と相補的な5'配列を含んでいる。標的核酸へのプローブA、プローブB、プローブC、プローブDがハイブリダイズするとき、プローブAがプローブBの隣に来て、プローブA上のビオチン部分を含んでいて標的核酸配列にハイブリダイズしない5'核酸配列と、プローブB上の標的核酸にハイブリダイズしない3'核酸配列が、5' DNAフラップを有する分岐した二本鎖DNA基板を形成し、プローブBがプローブCの隣に来て、プローブB上のビオチン部分を含んでいて標的核酸配列にハイブリダイズしない5'核酸配列と、プローブC上の標的核酸にハイブリダイズしない3'核酸配列が、5' DNAフラップを有する分岐した二本鎖DNA基板を形成し、プローブCがプローブDの隣に来て、プローブC上のビオチン部分を含んでいて標的核酸配列にハイブリダイズしない5'核酸配列と、プローブD上の標的核酸にハイブリダイズしない3'核酸配列が、5' DNAフラップを有する分岐した二本鎖DNA基板を形成する。プローブが標的核酸配列にハイブリダイズした後、標的核酸はストレプトアビジン常磁性ビーズを用いて精製される。過剰なゲノムDNAと過剰なプローブはビーズから洗い流される。標的核酸は、熱安定性フラップエンドヌクレアーゼ1（FEN1）とともにインキュベートすることによってビーズから溶離する。FEN1は5' DNAフラップを切断し、そのことによってビオチン部分をハイブリダイズした捕捉用プローブから分離し、標的核酸-捕捉用プローブ複合体を放出する。

【0239】

本開示により、利用者は、複数の標的核酸を捕捉して同時にシーケンシングすることも可能になり、複数の捕捉用プローブを複数の標的核酸が混合されたサンプルにハイブリダイズさせることができる。複数の標的核酸には、それぞれの核酸が同じ配列を含有する2個以上の核酸からなるグループ、またはそれぞれの核酸が必ずしも同じ配列を含有しない2個以上の核酸からなるグループを含めることができる。同様に、複数の捕捉用プローブには、配列が同じ2つ以上の捕捉用プローブからなるグループ、または配列が必ずしも同じでない2つ以上の捕捉用プローブからなるグループを含めることができる。例えばすべてが同じ配列を含有する複数の捕捉用プローブを用いると、利用者は、すべてが同じ配列を含有する複数の標的核酸を捕捉することができる。同じ配列を含有するこの複数の標的核酸をシーケンシングすることにより、データの冗長性を理由として高レベルのシーケンシング精度を実現することができる。別の一例では、目的のそれぞれの遺伝子と相補的な捕捉用プローブを含む一群の捕捉用プローブを用いて目的の2つ以上の特定の遺伝

子を捕捉して同時にシークエンシングすることができる。こうすることで、利用者は、特定の遺伝子の多重化されたシークエンシングを実施することが可能になる。図9は、FFPEサンプルを使用して100個の標的からなる多重がんパネルを捕捉して検出する本開示の方法を用いた実験からの結果を示している。

【0240】

捕捉用プローブは、「多重化オリゴ (multiplexing oligo)」に結合 (ハイブリダイズ) するドメインも含むことができる。多重化オリゴは少なくとも3つのドメインを含むことができる。第1のドメインは、捕捉用プローブにハイブリダイズする核酸配列を含むことができる。第2のドメインは、サンプルを同定する独自の核酸配列を含むことができる。第3のドメインは、基板結合部分を含むことができる。複数の多重化オリゴを本開示の捕捉用プローブと組み合わせて用いて少なくとも2つのサンプルから複数の標的核酸を同時にシークエンシングすることができる。多重化オリゴを用いて少なくとも3つのサンプル、または少なくとも4つのサンプル、または少なくとも5つのサンプル、または少なくとも6つのサンプル、または少なくとも7つのサンプル、または少なくとも8つのサンプル、または少なくとも9つのサンプル、または少なくとも10つのサンプル、または少なくとも100個のサンプル、または少なくとも1000個のサンプルから複数の標的核酸をシークエンシングすることができる。

【0241】

多重化オリゴを用いて3つのサンプルから3つの標的核酸分子を同時にシークエンシングする一例は以下の通りである：3つのサンプル (サンプル1、サンプル2、サンプル3) それぞれからの標的核酸を2つの捕捉用プローブ (プローブAとプローブB) にハイブリダイズさせる。プローブAは2つのドメインを含んでいる。第1のドメインは基板結合部分を含んでいる。第2のドメインは、標的核酸の5'末端と相補的な配列を含んでいる。プローブBは2つのドメインを含んでいる。第1のドメインは標的核酸の3'末端と相補的な配列を含んでいる。第2のドメインは、多重化オリゴと相補的な配列を含んでいる。これら2つの捕捉用プローブを標的核酸にハイブリダイズさせた後、プローブBの第2のドメインを多重化オリゴにハイブリダイズさせる。多重化オリゴは3つのドメインを含んでいる。第1のドメインは、プローブBの第2のドメインと相補的な配列を含んでいる。第2のドメインは、サンプルを同定する独自の核酸配列を含んでいる。第3のドメインは、基板結合部分を含んでいる。多重化オリゴをハイブリダイズさせた後、エンドヌクレアーゼを用いた切断工程を実施して標的核酸上のあらゆるオーバーハングDNAを除去し、プローブAが標的核酸の5'末端にハイブリダイズし、プローブBが標的核酸の3'末端にハイブリダイズするようにする。エンドヌクレアーゼ処理の後、多重化オリゴを標的核酸の3'末端に連結させ、次いでプローブBを除去する。標的核酸-プローブA複合体をさらに精製した後、シークエンシングする。各サンプルからの各標的核酸は多重化オリゴに連結されているため、標的核酸の出所であるサンプルを、多重化オリゴのシークエンシングによって同定することができる。

【0242】

全範囲をシークエンシングすることが望ましい場合には、必要とされる独立な捕捉用プローブの数は、標的核酸の断片のサイズと逆になる関係がある。言い換えると、高度に断片化された標的核酸ではより多くの捕捉用プローブが必要になろう。高度に断片化され分解されている標的核酸を含むタイプのサンプル (例えばホルマリン固定パラフィン包埋組織) では、捕捉用プローブのプールを多数含めると有用である可能性がある。逆に、標的核酸の断片が長いサンプル (例えばインビトロで得られた単離された核酸) では、5'末端の位置の1つの捕捉用プローブで十分である可能性がある。

【0243】

標的核酸のうちで2つの捕捉用プローブに挟まれた領域、または1つの捕捉用プローブの後かつ標的核酸の末端の前の領域を本明細書では「シークエンシングウインドウ」と呼ぶ。図8では、2つの捕捉用プローブを用いて1個の標的核酸を捕捉するときに生じるシークエンシングウインドウに標識している。シークエンシングウインドウは、標的核酸のうちでシークエンシング用プローブの結合に利用できる部分である。最小シークエンシングウ

10

20

30

40

50

インドウは標的結合ドメインの長さ（例えば4～10個のヌクレオチド）であり、最大シーケンシングウインドウは1つの染色体全体の大半である。

【0244】

本開示の方法を用いて大きな標的核酸分子をシーケンシングするとき、1つの「ブロッカーオリゴ (blocker oligo)」または複数のブロッカーオリゴを標的核酸の長さに沿ってハイブリダイズさせ、シーケンシングウインドウのサイズを制御することができる。ブロッカーオリゴは標的核酸に特定の位置でハイブリダイズし、そのことによってシーケンシング用プローブがこれらの位置に結合するのを阻止し、目的のより小さなシーケンシングウインドウを作り出す。より小さなシーケンシングウインドウを作り出すことにより、シーケンシング反応は標的DNA分子上の目的の特定の領域に限定されるため、シーケンシングの速度と精度が増大する。ブロッカーオリゴの使用は、標的核酸内の既知の位置にある特定の変異をシーケンシングするときに特に有用である。なぜなら標的核酸全体をシーケンシングする必要がないからである。非限定的な一例では、本開示の方法を2つのヘテロ接合部位を標的としたシーケンシングに用いて2つの異なるハプロタイプを識別することができる。

10

【0245】

捕捉用プローブは核酸分子複合体を含むことができる。核酸分子複合体は、部分的に二本鎖の核酸分子を含むことができる。いくつかの態様では、部分的に二本鎖の核酸分子は、標的特異的ドメイン、二本鎖ドメイン、一本鎖精製配列、切断可能な部分、一本鎖オーバーハングドメイン、サンプル特異的ドメイン、基板特異的ドメインのいずれか、またはこれらの任意の組み合わせを含むことができる。

20

【0246】

いくつかの態様では、部分的に二本鎖の核酸分子の任意の1つの鎖は、約40～150個のヌクレオチド、または約60～135個のヌクレオチド、または約10～90個のヌクレオチド、または約25～75個のヌクレオチド、または約60個のヌクレオチド、または約50～100個のヌクレオチドを含むことができる。

【0247】

いくつかの態様では、部分的に二本鎖の核酸分子の任意の1つの鎖は、少なくとも1個の親和性部分、または少なくとも2個の親和性部分、または少なくとも3個の親和性部分、または少なくとも4個の親和性部分、または少なくとも5個の親和性部分、または少なくとも6個の親和性部分、または少なくとも7個の親和性部分、または少なくとも8個の親和性部分、または少なくとも9個の親和性部分、または少なくとも10個の親和性部分を含むことができる。

30

【0248】

いくつかの態様では、部分的に二本鎖の核酸分子の任意の1つの鎖は、少なくとも1個の架橋部分を含むことができる。架橋部分として、化学的架橋部分、または光活性化架橋部分が可能である。

【0249】

捕捉用プローブは一本鎖核酸分子を含むことができる。いくつかの態様では、一本鎖核酸分子は、標的特異的ドメイン、二本鎖ドメイン、一本鎖精製配列、切断可能な部分、一本鎖オーバーハングドメイン、サンプル特異的ドメイン、基板特異的ドメインのいずれか、またはこれらの任意の組み合わせを含むことができる。

40

【0250】

標的特異的ドメイン、二本鎖ドメイン、一本鎖精製配列、切断可能な部分、一本鎖オーバーハングドメイン、サンプル特異的ドメイン、基板特異的ドメインは、少なくとも1つの天然塩基を含むこと、または天然塩基を含まないことが可能である。いくつかの態様では、標的特異的ドメイン、二本鎖ドメイン、一本鎖精製配列、切断可能な部分、一本鎖オーバーハングドメイン、サンプル特異的ドメイン、基板特異的ドメインは、少なくとも1つの修飾されたヌクレオチドまたは核酸類似体を含むこと、または修飾されたヌクレオチドを含まないことが可能である。

50

【0251】

標的特異的ドメイン、二本鎖ドメイン、一本鎖精製配列、切断可能な部分、一本鎖オーバーハングドメイン、サンプル特異的ドメイン、基板特異的ドメインは、天然塩基（例えば0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはそれよりも多い個数の天然塩基）と、修飾されたヌクレオチドまたは核酸類似体（例えば0個、1個、2個、3個、4個、5個、6個、7個、8個、9個、10個、またはそれよりも多い個数の修飾されたヌクレオチドまたは核酸類似体）の任意の組み合わせを含むことができる。天然塩基と、修飾されたヌクレオチドまたは核酸類似体は、組み合わせで存在するときには、任意の順番で配置することができる。

【0252】

標的特異的ドメインは、標的核酸分子の一部と相補的で標的核酸分子にハイブリダイズする核酸配列を含むことができる。いくつかの態様では、標的特異的ドメインは、約10～約150個のヌクレオチド、または約25～約100個のヌクレオチド、または約35～約100個のヌクレオチド、または約25～約125個のヌクレオチド、または約15～約100個のヌクレオチドを含むことができる。

【0253】

いくつかの態様では、標的特異的ドメインは、標的核酸分子の3'末端から少なくとも約100塩基対以内にハイブリダイズすることができる。いくつかの態様では、標的特異的ドメインは、標的核酸分子の5'末端から少なくとも約100塩基対以内にハイブリダイズすることができる。

【0254】

二本鎖ドメインは、別の核酸鎖にアニールして部分的に、または完全に二本鎖の核酸分子を形成することができる核酸配列を含むことができる。いくつかの態様では、二本鎖ドメインは、約14～約45個のヌクレオチド、または約25～約35個のヌクレオチド、または約30個のヌクレオチド、または約10～約60個のヌクレオチド、または約30～約50個のヌクレオチドを含むことができる。

【0255】

一本鎖精製配列は、精製で用いるのに適した核酸配列を含むことができる。一本鎖精製配列はF-タグを含むことができる。一本鎖精製配列はF様タグを含むことができる。一本鎖精製配列はヌクレオチド配列AACATCACACAGACC（配列番号112）を含むことができる。一本鎖精製配列はヌクレオチド配列GTCTATCATCACAGC（配列番号113）を含むことができる。

【0256】

一本鎖精製配列は、少なくとも1個の親和性部分、または少なくとも2個の親和性部分、または少なくとも3個の親和性部分、または少なくとも4個の親和性部分、または少なくとも5個の親和性部分、または少なくとも6個の親和性部分、または少なくとも7個の親和性部分、または少なくとも8個の親和性部分、または少なくとも9個の親和性部分、または少なくとも10個の親和性部分を含むことができる。親和性部分としてビオチンが可能である。したがっていくつかの態様では、一本鎖精製配列は、少なくとも1個のビオチン部分、または少なくとも2個のビオチン部分、または少なくとも3個のビオチン部分、または少なくとも4個のビオチン部分、または少なくとも5個のビオチン部分、または少なくとも6個のビオチン部分、または少なくとも7個のビオチン部分、または少なくとも8個のビオチン部分、または少なくとも9個のビオチン部分、または少なくとも10個のビオチン部分を含むことができる。

【0257】

一本鎖精製配列は、少なくとも50個のヌクレオチド、または約15～約50個のヌクレオチドを含むことができる。

【0258】

切断可能な部分は、酵素で切断可能な部分を含むことができる。酵素で切断可能な部分は、USER酵素によって切断するためのUSER配列を含むことができる。あるいは切断可能

10

20

30

40

50

な部分は、光で切断可能な部分を含むことができる。

【0259】

一本鎖オーバーハングドメインは、標的核酸分子と合わさって5'オーバーハングフラップ構造を形成することのできる一本鎖核酸配列を含むことができる。

【0260】

サンプル特異的ドメインは、標的核酸分子の供給源となった生物サンプルを同定する核酸配列を含むことができる。サンプル特異的ドメインはL-DNAを含むことができる。サンプル特異的ドメインはD-DNAを含むことができる。サンプル特異的ドメインはL-DNAとD-DNAの組み合わせを含むことができる。サンプル特異的ドメインは本開示の任意のプロープにハイブリダイズすることができる。サンプル特異的ドメインは約28個のヌクレオチドを含むことができる。

10

【0261】

いくつかの態様では、サンプル特異的ドメインは、少なくとも1つの付着位置、または少なくとも2つの付着位置を含むことができる。サンプル特異的ドメインが少なくとも1つの付着位置、または少なくとも2つの付着位置を含む態様では、付着位置は、約14個、または約10個、または約8個のヌクレオチドを含むことができる。

【0262】

基板特異的ドメインは、基板に付着した相補的核酸分子にハイブリダイズする核酸配列を含むことができる。基板としてアレイが可能である。基板特異的ドメインは、ローンオリゴヌクレオチドにハイブリダイズする核酸配列を含むことができる。

20

【0263】

基板特異的ドメインはポリ-A配列を含むことができる。基板特異的ドメインはポリ-T配列を含むことができる。基板特異的ドメインはL-ポリ-A配列を含むことができ、ポリ-A配列のヌクレオチドはL-DNAである。基板特異的ドメインはL-ポリT配列を含むことができ、ポリT配列のヌクレオチドはL-DNAである。基板特異的ドメインはL-DNAを含むことができる。基板特異的ドメインは約30個のヌクレオチドを含むことができる。

【0264】

図34は、標的核酸に結合する核酸分子複合体を含む「c5プロープ複合体」と名づけた代表的な捕捉用プロープの模式図を示している。このc5プロープ複合体は、部分的に二本鎖の核酸分子を含んでいる。この部分的に二本鎖の核酸分子の1本の鎖は、標的核酸にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の核酸分子の他方の鎖にアニールする二本鎖ドメインと、第1の一本鎖精製配列と、標的特異的ドメインと二本鎖ドメインの間に位置する切断可能な部分を含んでいる。この非限定的な例では、一本鎖精製配列はF様タグを含み、切断可能な部分は、酵素で切断可能なUSER配列を含んでいる。部分的に二本鎖の核酸分子の他方の鎖は、部分的に二本鎖の核酸分子のこの他方の鎖にアニールする二本鎖ドメインと、一本鎖オーバーハングドメインを含んでいる。この非限定的な例では、一本鎖オーバーハングドメインと標的核酸分子が5'オーバーハングフラップ構造を形成する。

30

【0265】

図34は、標的核酸に結合する核酸分子複合体を含む「c3プロープ複合体」と名づけた代表的な捕捉用プロープの模式図も示している。このc3プロープ複合体は、部分的に二本鎖の核酸分子を含んでいる。この部分的に二本鎖の核酸分子の1本の鎖は、標的核酸にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の核酸分子の他方の鎖にアニールする二本鎖ドメインと、標的特異的ドメインと二本鎖ドメインの間に位置する切断可能な部分を含んでいる。この非限定的な例では、切断可能な部分は、酵素で切断可能なUSER配列を含んでいる。部分的に二本鎖の核酸分子の他方の鎖は、部分的に二本鎖の核酸分子のこの他方の鎖にアニールする二本鎖ドメインと、サンプル特異的ドメインと、基板特異的ドメインと、一本鎖精製配列と、一本鎖精製配列と基板特異的ドメインの間に位置する切断可能な部分を含んでいる。この非限定的な例では、サンプル特異的ドメインはL-DNAを含み、基板特異的ドメインはL-DNAを含み、一本鎖精製配列はF-タグを含み、切断可

40

50

能な部分は光で切断可能な部分を含んでいる。

【0266】

図41は、標的核酸に結合する核酸分子複合体を含む「c3.2プローブ複合体」と名づけた代表的な捕捉用プローブの模式図を示している。このc3.2プローブ複合体は、部分的に二本鎖の核酸分子を含んでいる。この部分的に二本鎖の核酸分子の1本の鎖は、標的核酸にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の核酸分子の他方の鎖にアニールする二本鎖ドメインを含んでいる。いくつかの態様では、この鎖は、任意により少なくとも1つの第1の親和性部分を含んでいる。いくつかの態様では、この鎖は、任意により、標的特異的ドメインと二本鎖ドメインの間に位置する切断可能な部分を含んでいる。部分的に二本鎖の核酸分子の他方の鎖は、部分的に二本鎖の核酸分子のこの他方の鎖にアニールする二本鎖ドメインと、基板特異的ドメインを含んでいる。いくつかの態様では、この鎖は、任意により、少なくとも1つ、または少なくとも2つ、または少なくとも3つの第2の親和性部分を含むことができる。

10

【0267】

図41は、標的核酸に結合する核酸分子複合体を含む「c5.2プローブ複合体」と名づけた代表的な捕捉用プローブの模式図も示している。このc5.2プローブ複合体は、部分的に二本鎖の核酸分子を含んでいる。この部分的に二本鎖の核酸分子の1本の鎖は、標的核酸にハイブリダイズする標的特異的ドメインと、この部分的に二本鎖の核酸分子の他方の鎖にアニールする二本鎖ドメインを含んでいる。いくつかの態様では、この鎖は、任意により、標的特異的ドメインと二本鎖ドメインの間に位置する切断可能な部分を含んでいる。部分的に二本鎖の核酸分子の他方の鎖は、部分的に二本鎖の核酸分子のこの他方の鎖にアニールする二本鎖ドメインと、サンプル特異的ドメインと、第1の一本鎖精製配列と、二本鎖ドメインとサンプル特異的ドメインの間に位置する第1の切断可能な部分と、サンプル特異的ドメインと第1の一本鎖精製配列の間に位置する第2の切断可能な部分を含んでいる。いくつかの態様では、第1の一本鎖精製配列は、少なくとも1つの親和性部分、例えば少なくとも1つのビオチン部分を含むことができる。いくつかの態様では、第1の一本鎖精製配列は、少なくとも1つのビオチン部分で置き換えることができる。そのため部分的に二本鎖の核酸分子の他方の鎖は、部分的に二本鎖の核酸分子のこの他方の鎖にアニールする二本鎖ドメインと、サンプル特異的ドメインと、少なくとも1つのビオチン部分と、二本鎖ドメインとサンプル特異的ドメインの間に位置する第1の切断可能な部分と、サンプル特異的ドメインと少なくとも1つのビオチン部分の間に位置する第2の切断可能な部分を含んでいる。

20

30

【0268】

本開示のサンプル調製法

【0269】

本開示により、標的核酸分子を基板に固定化することを含むサンプル調製法が提供される。

【0270】

本発明のサンプル調製法は、CRISPRに基づく断片化工程を含むことができる（例えばBakerとMueller、「CRISPRを媒介とした、ゲノムDNAの特定のメガ塩基区画の単離」、Nucleic Acids Research 2017年、第45巻(19)、e165ページ；Tsai他、「増幅なしの、CRISPR-Cas9による、リピート伸長疾患を引き起こすゲノム領域の富化とSMRTシーケンシング」、bioRxiv 203919；doi: <https://doi.org/10.1101/203919>；Nachmanson他、「CRISPR/Cas9を利用した標的化ゲノム断片化によって小さな標的のハイブリダイゼーション捕捉が改善され、PCRバイアスが減少し、効率的な高精度シーケンシングが可能になる」、bioRxiv 207027；doi: <https://doi.org/10.1101/207027>を参照されたい）。CRIPPRによる断片化は、生物サンプルから得られたゲノムDNA (gDNA) を、gDNAの中に位置するプロトスペーサ隣接モチーフ (PAM) 部位の近位で切断することによってインビトロで断片化することができる。PAM部位はヌクレオチド配列NGGを含むことができる (Nは任意の核塩基である)。あるいはPAM部位はヌクレオチド配列

40

50

NGAを含むことができる（Nは任意の核塩基である）。CRISPRに基づく断片化によって生成した断片は、ビオチニル化CRISPR複合体、または抗CAS9抗体を用いて精製することができる。

【0271】

標的核酸を捕捉する方法は、（1）CRISPRに基づく断片化工程を利用してgDNAを断片化し；（2）断片化されたgDNAを少なくとも2つの捕捉用プローブと接触させ、これら少なくとも2つの捕捉用プローブの少なくとも1つは上記のc5プローブ複合体であり、これら少なくとも2つの捕捉用プローブの少なくとも1つは上記のc3プローブ複合体であり、c3プローブ複合体とc5プローブ複合体が標的核酸にハイブリダイズして図34に示されている複合体を形成し；（3）この複合体をFEN1と接触させることによって5'オーバーハングフラップ構造を除去し；（4）標的核酸の3'末端を、基板特異的ドメインを含むc3プローブ複合体の鎖の5'末端に連結させ；（5）c5プローブ複合体の一本鎖精製配列を第1の基板に結合させ；（6）c3プローブ複合体とc5プローブ複合体それぞれの二本鎖ドメインと標的特異的ドメインの間に位置する切断可能な部分を切断し；（7）c3プローブ複合体の一本鎖精製配列を第2の基板に結合させ；（8）連結されたc3プローブ複合体の一本鎖精製配列と基板特異的ドメインの間に位置する切断可能な部分を切断し；（9）第3の基板に付着した相補的核酸分子に基板特異的ドメインをハイブリダイズさせることを含むことができる。

【0272】

上記の方法のいくつかの態様では、工程（9）を工程（8）の前に実施することができる。

【0273】

上記の方法のいくつかの態様では、工程（3）と（4）を併せて実施することができる。上記の方法のいくつかの態様では、工程（3）と（4）を同時に実施することができる。

【0274】

いくつかの態様では、上記の方法は、任意により工程（6）と（7）に1つの工程を含むことができ、その工程では、異なる生物サンプルに由来する標的核酸-捕捉用プローブ複合体がまとめてプールされる。この態様では、異なる生物サンプルに由来する標的核酸-捕捉用プローブ複合体は、独自のサンプル特異的ドメインを含むc3プローブ複合体を含むことになるため、標的特異的ドメインが、取得したそれぞれの標的核酸の出所である生物サンプルを同定する。

【0275】

本開示のサンプル調製法の一例が図34～図40に示されている。この非限定的な例では、生物サンプルから得られたgDNAが最初にCRISPRに基づく断片化を利用して断片化される。断片化の後、図34に示されているように、標的核酸が2つの捕捉用プローブにハイブリダイズする。この非限定的な例では、2つの捕捉用プローブは、上に説明したc3プローブ複合体とc5プローブ複合体である。c3プローブ複合体とc5プローブ複合体は、標的核酸に沿って重複しない位置で標的核酸にハイブリダイズする。c3プローブ複合体は、標的核酸の3'末端から8個以内のヌクレオチドの位置で標的特異的ドメインを介して標的核酸にハイブリダイズし、c5プローブ複合体は、標的特異的ドメインを介して標的核酸にハイブリダイズする結果、c5プローブ複合体は、c3プローブ複合体の5'側にハイブリダイズする。c5プローブ複合体の一本鎖オーバーハングドメインと標的核酸分子は5'オーバーハングフラップ構造を形成する。これら2つの捕捉用プローブがハイブリダイズした後、標的核酸-捕捉用プローブ複合体をFEN1およびリガーゼとともにインキュベートする。図35に示されているように、FEN1は5'オーバーハングフラップ構造を除去し、リガーゼは、標的核酸の3'末端を、基板特異的ドメインを含むc3プローブ複合体の鎖に連結する。図36に示されている得られた複合体は、c5プローブ複合体の中に存在するF様タグにハイブリダイズするF様ビーズに結合する。ビーズを洗浄し、USER酵素を添加する。図37に示されているように、USER酵素が、c3プローブ複合体とc5プローブ複合体両方の標的特異的ドメインと二本鎖ドメインの間に位置する切断可能な部分を切断することで、F様ビーズから標的核酸を放出する。図38に示されているように、溶離した複合体は、SPRIビーズを用い

10

20

30

40

50

てさらに精製される。その後、精製された複合体は、c3プローブ複合体の中に存在するF-タグにハイブリダイズするF-ビーズに結合する。洗浄後、図39に示されているように、F-ビーズをUV光に曝露してc3プローブ複合体の中で基板特異的ドメインとF-タグの間に位置する光で切断可能な部分を切断することにより、標的核酸をF-ビーズから溶離させる。その後、図40に示されているように、連結されたc3プローブ複合体の基板特異的ドメインを、基板に付着した相補的核酸にハイブリダイズさせることにより、得られた複合体を基板に結合させる。

【0276】

本開示の別のサンプル調製法の一例が図41～図46に示されている。この非限定的な例では、生物サンプルから得られたgDNAが最初にCRISPRに基づく断片化を用いて断片化される。断片化の後、図41に示されているように、標的核酸が2つの捕捉用プローブにハイブリダイズする。この非限定的な例では、2つの捕捉用プローブは、上に説明したc3.2プローブ複合体とc5.2プローブ複合体である。c3.2プローブ複合体とc5.2プローブ複合体は、標的核酸に沿って重複しない位置で標的核酸にハイブリダイズする。c5.2プローブ複合体は標的特異的ドメインを介して標的核酸にハイブリダイズする結果、c5.2プローブ複合体はc3.2プローブ複合体の5'側にハイブリダイズする。これら2つの捕捉用プローブがハイブリダイズした後、図42に示されているように、標的核酸はc3.2プローブ複合体の1本の鎖とc5.2プローブ複合体の1本の鎖に連結される。連結に含めることができるのは、酵素による連結、自己連結、化学的連結のいずれか、またはこれらの任意の組み合わせである。酵素による連結を含む態様では、酵素による連結は、高忠実度の鋳型依存性ニックリガーゼを用いて実施できる。その後、図42に示されている得られた複合体は、一本鎖精製配列にハイブリダイズする少なくとも1つのオリゴヌクレオチドを含むビーズに結合することができる。図43に示されているように、ビーズを洗浄し、サンプル特異的ドメインと一本鎖精製配列の間に位置する切断可能な部分を切断することにより、ビーズから標的核酸を放出させることができる。その後、得られた複合体は、図44に示されているように、基板に付着したオリゴヌクレオチドに基板特異的ドメインをハイブリダイズさせることにより、基板の表面に固定化することができる。基板／オリゴヌクレオチド複合体として、本開示の任意のアレイが可能である。

【0277】

上記の方法はさらに、第1の検出可能な標識と第2の検出可能な標識を含む少なくとも1つのレポータープローブをサンプル特異的ドメインにハイブリダイズさせることを含むことができる。その後、第1と第2の検出可能な標識を同定することにより、第1の検出可能な標識と第2の検出可能な標識の属性に基づいて標的核酸の出所であるサンプルを同定することができる。

【0278】

あるいは上記の方法は、第1の検出可能な標識と第2の検出可能な標識を含む第1のレポータープローブをサンプル特異的ドメインにハイブリダイズさせることをさらに含むことができる。その後、第1の検出可能な標識と第2の検出可能な標識を同定することができる。その後、第1の検出可能な標識と第2の検出可能な標識を除去し、第3の検出可能な標識と第4の検出可能な標識を含む第2のレポータープローブをサンプル特異的ドメインにハイブリダイズさせることができる。その後、第3の検出可能な標識と第4の検出可能な標識を同定することにより、第1の検出可能な標識、第2の検出可能な標識、第3の検出可能な標識、第4の検出可能な標識の属性に基づいて標的核酸の出所であるサンプルを同定することができる。

【0279】

標的核酸の出所であるサンプルを同定した後、図45に示されているようにサンプル特異的ドメインと二本鎖ドメインの間に位置する切断可能な部分を切断することにより、サンプル特異的ドメインを放出させることができる。

【0280】

本開示の方法

【0281】

本開示のシーケンシング法は、本明細書に開示されている少なくとも1つのシーケンシング用プローブを標的核酸に可逆的にハイブリダイズさせることを含んでいる。

【0282】

核酸をシーケンシングする方法は、(1) 本明細書に記載のシーケンシング用プローブを標的核酸にハイブリダイズさせることを含むことができる。標的核酸は、任意により基板に1つ以上の位置で固定化することができる。代表的なシーケンシング用プローブは、標的結合ドメインとバーコードドメインを含むことができ；標的結合ドメインは、表1に記載されている任意のコンストラクトを含んでいる。代表的な標的結合ドメインは標的核酸にハイブリダイズする少なくとも8個のヌクレオチドを含んでいて、標的結合ドメイン内の少なくとも6個のヌクレオチドは標的核酸分子内の対応するヌクレオチドを同定することができ（例えばこれら6個のヌクレオチドが、標的結合ドメインがハイブリダイズする標的分子と相補的な6個のヌクレオチドを同定する）、標的結合ドメイン内の少なくとも2個のヌクレオチドは標的核酸分子内の対応するヌクレオチドを同定せず（例えばこれら少なくとも2個のヌクレオチドは、標的結合ドメインがハイブリダイズする標的分子と相補的な2個のヌクレオチドを同定しない）；標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたはヌクレオチド類似体であることが可能であり；標的結合ドメイン内にあって標的核酸分子内の対応するヌクレオチドを同定しない上記少なくとも2個のヌクレオチドは、標的結合ドメイン内の上記少なくとも6個のヌクレオチドによって指定される標的に特異的でない4種類の正規塩基、またはユニバーサル塩基、または縮重塩基のうちの任意のものが可能である。代表的なバーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的な核酸分子が結合することのできる少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含み、上記少なくとも3つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置のそれぞれの位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドのうちの対応する2個のヌクレオチドの位置と属性を決めている。

【0283】

別の態様では、代表的な標的結合ドメインは、標的核酸にハイブリダイズする少なくとも6個のヌクレオチドを含むことができ、標的結合ドメイン内のこれら少なくとも6個のヌクレオチドは標的核酸分子内の対応するヌクレオチドを同定することができ（例えば標的結合ドメイン配列が正確に6個のヌクレオチドであるとき、これら6個のヌクレオチドは、ハイブリダイズする標的分子と相補的な6個のヌクレオチドを同定する）；標的結合ドメイン内のこれら少なくとも6個のヌクレオチドはどれも、修飾されたヌクレオチドまたは核酸類似体でないか、標的結合ドメイン内のその少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたは核酸類似体であることが可能である。

【0284】

この方法は、シーケンシング用プローブを標的核酸にハイブリダイズさせた後、(2) 第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的な核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；(3) 結合した第1の相補的な核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出し；(4) 固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定することを含んでいる。例えば第1の相補的な核酸分子が2つの検出可能な標識を含んでいるときには、これら2つの検出可能な標識が、固定化された核酸分子内の上記少なくとも2個のヌクレオチドを同定する。

【0285】

上記少なくとも2つの検出可能な標識を検出した後、第1の相補的な核酸分子からこれら少なくとも2つの検出可能な標識を除去する。したがってこの方法はさらに、(5) 検出可能

な標識が欠けた第1のハイブリダイジング核酸分子 (hybridizing nucleic acid molecule) を第1の付着位置に結合させることによって検出可能な標識を含む第1の相補的核酸分子を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子を、第1の検出可能な標識と少なくとも第2の検出可能な標識を放出させるのに十分な力と接触させることを含んでいる。したがって工程 (5) の後には、どの検出可能な標識も第1の付着位置に結合していない。この方法はさらに、(6) 第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核酸分子をバーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；(7) 結合した第2の相補的核酸分子の第3の検出可能な標識と少なくとも第4の検出可能な標識を検出し；(8) 任意により固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定し；(9) バーコードドメイン内の上記少なくとも3つの付着位置のそれぞれに、2つの検出可能な標識を含む相補的核酸分子が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されるまで、工程 (5) ~ (8) を繰り返すことにより、シーケンシング用プローブの標的結合ドメインにハイブリダイズした固定化された標的核酸の少なくとも第1の領域のための少なくとも6個のヌクレオチドの線形順序を同定し；(10) 任意により固定化された標的核酸からシーケンシング用プローブを除去することを含んでいる。

【0286】

この方法はさらに、(11) 第2のシーケンシング用プローブを、任意により基板に1つ以上の位置で固定化された標的核酸にハイブリダイズさせ（ただし第1のシーケンシング用プローブと第2のシーケンシング用プローブの標的結合ドメインは異なる）；(12) 第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；(13) 結合した第1の相補的核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出し；(14) 任意により固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定し；(15) 検出可能な標識が欠けた第1のハイブリダイジング核酸分子を第1の付着位置に結合させることによって検出可能な標識を含む第1の相補的核酸分子または複合体を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子または複合体を、第1の検出可能な標識と少なくとも第2の検出可能な標識を放出させるのに十分な力と接触させ；(16) 第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核酸分子をバーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；(17) 結合した第2の相補的核酸分子の第3の検出可能な標識と少なくとも第4の検出可能な標識を検出し；(18) 任意により固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定し；(19) バーコードドメイン内の上記少なくとも3つの付着位置のそれぞれに、2つの検出可能な標識を含む相補的核酸分子が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されるまで、工程 (15) ~ (18) を繰り返すことにより、第2のシーケンシング用プローブの標的結合ドメインにハイブリダイズした固定化された標的核酸の少なくとも第2の領域のための少なくとも6個のヌクレオチドの線形順序を同定し；(20) 任意により固定化された標的核酸から第2のシーケンシング用プローブを除去することを含むことができる。

【0287】

この方法はさらに、固定化された標的核酸の少なくとも第1の領域と少なくとも第2の領域の中のヌクレオチドの同定されたそれぞれの線形順序をアセンブルすることにより、固定化された標的核酸の配列を同定することを含むことができる。

【0288】

工程 (5) と (6) は逐次的に、または同時に起こることができる。第1の検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。第3の検出可能な標識と少なくとも第4の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

【0289】

第1の相補的核酸分子は、切断可能なリンカーを含むことができる。第2の相補的核酸分子は、切断可能なリンカーを含むことができる。第1の相補的核酸分子と第2の相補的核酸分子は、それぞれが切断可能なリンカーを含むことができる。切断可能なリンカーは、光で切断できることが好ましい。放出力として光が可能である。UV光が好ましい。光は、アークランプ、レーザー、集束UV光源、発光ダイオードからなるグループから選択された光源によって提供することができる。

【0290】

第1の相補的核酸分子と、検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。例えば検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、第1の相補的核酸分子のうちでバーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合する部分と同じ核酸配列を含むことができる。検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、バーコードドメイン内の第1の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0291】

第2の相補的核酸分子と、検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、バーコードドメイン内の第2の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0292】

本開示により、核酸をシーケンシングする方法として、(1) 本明細書に記載されているシーケンシング用プローブを標的核酸にハイブリダイズさせることを含む方法も提供される。標的核酸は、任意により基板に1つ以上の位置で固定化することができる。代表的なシーケンシング用プローブは標的結合ドメインとバーコードドメインを含むことができ；標的結合ドメインは、表1に記載されている任意のコンストラクトを含んでいる。代表的な標的結合ドメインは、標的核酸にハイブリダイズする少なくとも8個のヌクレオチドを含んでいて、標的結合ドメイン内の少なくとも6個のヌクレオチドは標的核酸分子内の対応するヌクレオチドを同定することができ（例えばこれら6個のヌクレオチドは、ハイブリダイズする標的分子と相補的な6個のヌクレオチドを同定する）、標的結合ドメイン内の少なくとも2個のヌクレオチドは標的核酸分子内の対応するヌクレオチドを同定せず（例えばこれら少なくとも2個のヌクレオチドは、ハイブリダイズする標的分子と相補的な2個のヌクレオチドを同定しない）；標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの任意のものが、修飾されたヌクレオチドまたはヌクレオチド類似体であることが可能であり、標的結合ドメイン内にあって標的核酸分子内の対応するヌクレオチドを同定しない上記少なくとも2個のヌクレオチドは、標的結合ドメイン内の上記少なくとも6個のヌクレオチドによって指定される標的に特異的でない4つの正規塩基、またはユニバーサル塩基、または縮重塩基のうちの任意のものが可能である。代表的なバーコードドメインは合成骨格を含むとともに少なくとも3つの付着位置を含み、それぞれの付着位置は、相補的核酸分子が結合することのできる少なくとも1つの核酸配列を含む少なくとも1つの付着領域を含み、上記少なくとも3つの付着位置のそれぞれの付着位置は、標的結合ドメイン内の上記少なくとも6個のヌクレオチドのうちの2個のヌクレオチドに対応し、上記少なくとも3つの付着位置のそれぞれは異なる核酸配列を有し、上記少なくとも3つの付着位置のそれぞれの位置の核酸配列が、標的結合ドメインが結合する標的核酸内の上記少なくとも6個のヌクレオチドのうちの対応する2個のヌクレオチドの位置と属性を決めている。

【0293】

別の態様では、代表的な標的結合ドメインは、標的核酸にハイブリダイズする少なくとも6個のヌクレオチドを含むことができ、標的結合ドメイン内のこれら少なくとも6個のヌクレオチドは、標的核酸分子内の対応するヌクレオチドを同定することができ（例えば標的結合ドメイン配列が正確に6個のヌクレオチドであるとき、これら6個のヌクレオチドは、ハイブリダイズする標的分子と相補的な6個のヌクレオチドを同定する）；標的結合ド

メイン内の上記少なくとも6個のヌクレオチドのどれも、修飾されたヌクレオチドまたはヌクレオチド類似体でないか、上記少なくとも6個のヌクレオチドのうちのいずれかが、修飾されたヌクレオチドまたはヌクレオチド類似体であることが可能である。

【0294】

この方法は、標的核酸にシーケンシング用プローブをハイブリダイズさせた後、(2) 第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；(3) 結合した第1の相補的核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出して記録することを含んでいる。

【0295】

上記少なくとも2つの検出可能な標識を検出して記録した後、これら少なくとも2つの検出可能な標識を第1の相補的核酸分子から除去する。したがってこの方法はさらに、(4) 第1の付着位置に、検出可能な標識が欠けた第1のハイブリダイジング核酸分子を結合させることにより、検出可能な標識を含む第1の相補的核酸分子を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子を、第1の検出可能な標識と少なくとも第2の検出可能な標識を放出させるのに十分な力と接触させることを含んでいる。したがって工程(4)の後には、どの検出可能な標識も第1の付着位置に結合していない。この方法はさらに、(5) 第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；(6) 結合した第2の相補的核酸分子の第3の検出可能な標識と少なくとも第4の検出可能な標識を検出して記録し；(7) バーコードドメイン内の上記少なくとも3つの付着位置のそれぞれの付着位置に2つの検出可能な標識を含む相補的核酸が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されて記録されるまで、工程(4)～(6)を繰り返し；(8) 工程(3)、工程(6)、工程(7)で記録された検出可能な標識を用いて、シーケンシング用プローブの標的結合ドメインにハイブリダイズした固定化された標的核酸の少なくとも第1の領域の上記少なくとも6個のヌクレオチドの位置と属性を同定し；(9) 任意により固定化された標的核酸からシーケンシング用プローブを除去することを含んでいる。

【0296】

この方法はさらに、(10) 第2のシーケンシング用プローブを、基板に任意により1つ以上の位置で固定化された標的核酸にハイブリダイズさせ(ただし第1のシーケンシング用プローブと第2のシーケンシング用プローブの標的結合ドメインは異なる)；(11) 第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；(12) 結合した第1の相補的核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出して記録し；(13) 検出可能な標識が欠けた第1のハイブリダイジング核酸分子を第1の付着位置に結合させることにより、検出可能な標識を含む第1の相補的核酸分子または複合体を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子または複合体を、第1の検出可能な標識と少なくとも第2の検出可能な標識を放出させるのに十分な力と接触させ；(14) 第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；(15) 結合した第2の相補的核酸分子の第3の検出可能な標識と少なくとも第4の検出可能な標識を検出して記録し；(16) バーコードドメイン内の上記少なくとも3つの付着位置のそれぞれの付着位置に2つの検出可能な標識を含む相補的核酸分子が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されて記録されるまで、工程(13)～(15)を繰り返し；(17) 工程(12)、工程(15)、工程(16)で記録された検出可能な標識を用い、第2のシーケンシング用プローブの標的結合ドメインがハイブリダイズした固定化された標的核酸の少なくとも第2の領域の上記少なくとも6個のヌクレオチドの位置と属性を同定し；(18) 任意により固定化された標的核酸から第2のシーケンシング用プローブを除去することを含むことができる。

10

20

30

40

50

【0297】

この方法はさらに、固定化された標的核酸の少なくとも第1の領域と少なくとも第2の領域の中のヌクレオチドの同定されたそれぞれの線形順序をアセンブルすることにより、固定化された標的核酸の配列を同定することを含むことができる。

【0298】

工程(4)と(5)は逐次的に、または同時に起こることができる。第1の検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。第3の検出可能な標識と少なくとも第4の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

【0299】

第1の相補的核酸分子は、切断可能なリンカーを含むことができる。第2の相補的核酸分子は、切断可能なリンカーを含むことができる。第1の相補的核酸分子と第2の相補的核酸分子は、それぞれが切断可能なリンカーを含むことができる。切断可能なリンカーは、光で切断できることが好ましい。放出力として光が可能である。UV光が好ましい。光は、アークランプ、レーザー、集束UV光源、発光ダイオードからなるグループから選択された光源によって提供することができる。

【0300】

第1の相補的核酸分子と、検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。例えば検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、第1の相補的核酸分子のうちでバーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合する部分と同じ核酸配列を含むことができる。検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、バーコードドメイン内の第1の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0301】

第2の相補的核酸分子と、検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、バーコードドメイン内の第2の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0302】

上記の方法はさらに、検出可能な標識を記録するのに適した媒体を含むことができる。この媒体として、適切なコンピュータ可読媒体が可能である。

【0303】

本開示によりさらに、本明細書に開示されている複数のシークエンシング用プローブを用いて核酸をシークエンシングする方法が提供される。例えば標的核酸を2つ以上のシークエンシング用プローブにハイブリダイズさせると、各プローブは、核酸のうちでそのプローブがハイブリダイズした部分をシークエンシングすることができる。

【0304】

本開示により、核酸をシークエンシングする方法として、(1)本明細書に記載されている複数のシークエンシング用プローブを含む第1のシークエンシング用プローブの少なくとも1つの第1の集団を、基板の1つ以上の位置に任意により固定化された標的核酸にハイブリダイズさせ；(2)第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的核酸分子をバーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；(3)結合した第1の相補的核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出し；(4)固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定し；(5)検出可能な標識が欠けた第1のハイブリダイジング核酸分子を第1の付着位置に結合させることにより、検出可能な標識を含む第1の相補的核酸分子を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子を、第1の検出可能な標識と少なくとも第2の検出可能な標識を放出させるのに十分な力と接触させ；(6)第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核

10

20

30

40

50

酸分子をバーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；（7）結合した第2の相補的核酸分子の第3の検出可能な標識と少なくとも第4の検出可能な標識を検出し；（8）任意により固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定し；（9）バーコードドメイン内の上記少なくとも3つの付着位置の各付着位置に2つの検出可能な標識を含む相補的核酸が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されるまで工程（5）～（8）を繰り返すことにより、シーケンシング用プローブの標的結合ドメインがハイブリダイズした固定化された標的核酸の少なくとも1つのヌクレオチドの線形順序を同定し；（10）任意により固定化された標的核酸から第1のシーケンシング用プローブの少なくとも1つの第1の集団を除去することを含む方法も提供される。

10

【0305】

この方法はさらに、（11）本明細書に開示されている複数のシーケンシング用プローブを含む第2のシーケンシング用プローブの少なくとも1つの第2の集団を、基板の1つ以上の位置に任意により固定化された標的核酸にハイブリダイズさせ（ただし第1のシーケンシング用プローブと第2のシーケンシング用プローブの標的結合ドメインは異なる）；（12）第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；（13）結合した第1の相補的核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出し；（14）任意により固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定し；（15）検出可能な標識が欠けた第1のハイブリダイジング核酸分子を第1の付着位置に結合させることにより、検出可能な標識を含む第1の相補的核酸分子または複合体を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子または複合体を、第1の検出可能な標識と少なくとも第2の検出可能な標識を放出させるのに十分な力と接触させ；（16）第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核酸分子をバーコードドメイン内の上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；（17）結合した第2の相補的核酸分子の第3の検出可能な標識と少なくとも第4の検出可能な標識を検出し；（18）固定化された標的核酸内の少なくとも2個のヌクレオチドの位置と属性を同定し；（19）バーコードドメイン内の上記少なくとも3つの付着位置のそれぞれの付着位置に2つの検出可能な標識を含む相補的核酸が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されるまで工程（15）～（18）を繰り返すことにより、シーケンシング用プローブの標的結合ドメインがハイブリダイズした固定化された標的核酸の少なくとも1つの第2の領域のための少なくとも6個のヌクレオチドの線形順序を同定し；（20）任意により固定化された標的核酸からシーケンシング用プローブの少なくとも1つの第2の集団を除去することを含むことができる。

20

30

【0306】

この方法はさらに、固定化された標的核酸の少なくとも第1の領域と少なくとも第2の領域の中のヌクレオチドの同定されたそれぞれの線形順序をアSEMBLすることにより、固定化された標的核酸の配列を同定することを含むことができる。

【0307】

工程（5）と（6）は逐次的に、または同時に起こることができる。第1の検出可能なと少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。第3の検出可能なと少なくとも第4の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

40

【0308】

第1の相補的核酸分子は、切断可能なリンカーを含むことができる。第2の相補的核酸分子は、切断可能なリンカーを含むことができる。第1の相補的核酸分子と第2の相補的核酸分子は、それぞれが切断可能なリンカーを含むことができる。切断可能なリンカーは、光で切断できることが好ましい。放出力として光が可能である。UV光が好ましい。光は、ア

50

ークランプ、レーザー、集束UV光源、発光ダイオードからなるグループから選択された光源によって提供することができる。

【0309】

第1の相補的核酸分子と、検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、バーコードドメイン内の第1の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0310】

第2の相補的核酸分子と、検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、バーコードドメイン内の第2の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0311】

本開示により、核酸をシークエンシングする方法として、(1) 本明細書に記載されている複数のシークエンシング用プローブを含む第1のシークエンシング用プローブの少なくとも1つの第1の集団を、基板の1つ以上の位置に任意により固定化された標的核酸にハイブリダイズさせ；(2) 第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的核酸分子をバーコードドメイン内の上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；(3) 結合した第1の相補的核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出して記録し；(4) 検出可能な標識が欠けた第1のハイブリダイジング核酸分子を第1の付着位置に結合させることにより、検出可能な標識を含む第1の相補的核酸分子を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子を、第1の検出可能な標識と少なくとも1つの第2の検出可能な標識を放出させるのに十分な力と接触させ；(5) 第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核酸分子をバーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；(6) 結合した第2の相補的核酸分子の第3の検出可能な標識と少なくとも第4の検出可能な標識を検出して記録し；(7) バーコードドメイン内の上記少なくとも3つの付着位置のそれぞれの付着位置に2つの検出可能な標識を含む相補的核酸分子が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されて記録されるまで、工程(4)～(6)を繰り返し；(8) 工程(3)、工程(6)、工程(7)で記録された検出可能な標識を用い、シークエンシング用プローブの標的結合ドメインがハイブリダイズした固定化された標的核酸の少なくとも1つの第1の領域の上記少なくとも6個のヌクレオチドの位置と属性を同定し；(9) 任意により固定化された標的核酸から第1のシークエンシング用プローブの少なくとも1つの第1の集団を除去することを含む方法も提供される。

【0312】

この方法はさらに、(10) 本明細書に記載されている複数のシークエンシング用プローブを含む第2のシークエンシング用プローブの少なくとも1つの第2の集団を、基板の1つ以上の位置に任意により固定化された標的核酸にハイブリダイズさせ（ただし第1のシークエンシング用プローブと第2のシークエンシング用プローブの標的結合ドメインは異なる）；(11) 第1の検出可能な標識と少なくとも第2の検出可能な標識を含む第1の相補的核酸分子をバーコードドメインの上記少なくとも3つの付着位置のうちの第1の付着位置に結合させ；(12) 結合した第1の相補的核酸分子の第1の検出可能な標識と少なくとも第2の検出可能な標識を検出して記録し；(13) 検出可能な標識が欠けた第1のハイブリダイジング核酸分子を第1の付着位置に結合させることにより、検出可能な標識を含む第1の相補的核酸分子または複合体を自由な状態にするか、検出可能な標識を含む第1の相補的核酸分子または複合体を、第1の検出可能な標識と少なくとも第2の検出可能な標識を放出させるのに十分な力と接触させ；(14) 第3の検出可能な標識と少なくとも第4の検出可能な標識を含む第2の相補的核酸分子を、バーコードドメインの上記少なくとも3つの付着位置のうちの第2の付着位置に結合させ；(15) 結合した第2の相補的核酸分子の第3の検出

可能な標識と少なくとも第4の検出可能な標識を検出して記録し；(16) バーコードドメイン内の上記少なくとも3つの付着位置のそれぞれの付着位置に2つの検出可能な標識を含む相補的核酸分子が結合し、結合した相補的核酸分子のこれら2つの検出可能な標識が検出されて記録されるまで、工程(13)～(15)を繰り返し；(17) 工程(12)、工程(15)、工程(16)で記録された検出可能な標識を用い、第2のシーケンシング用プローブの標的結合ドメインがハイブリダイズした固定化された標的核酸の少なくとも1つの第2の領域の上記少なくとも6個のヌクレオチドの位置と属性を同定し；(18) 任意により固定化された標的核酸から第2のシーケンシング用プローブの少なくとも1つの第2の集団を除去することを含むことができる。

【0313】

この方法はさらに、固定化された標的核酸の少なくとも第1の領域と少なくとも第2の領域の中のヌクレオチドの同定されたそれぞれの線形順序をアセンブルすることにより、固定化された標的核酸の配列を同定することを含むことができる。

【0314】

工程(4)と(5)は逐次的に、または同時に起こることができる。第1の検出可能な標識と少なくとも第2の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。第3の検出可能な標識と少なくとも第4の検出可能な標識は、同じ発光スペクトルを有すること、または異なる発光スペクトルを有することができる。

【0315】

第1の相補的核酸分子は、切断可能なリンカーを含むことができる。第2の相補的核酸分子は、切断可能なリンカーを含むことができる。第1の相補的核酸分子と第2の相補的核酸分子は、それぞれが切断可能なリンカーを含むことができる。切断可能なリンカーは、光で切断できることが好ましい。放出力として光が可能である。UV光が好ましい。光は、アークランプ、レーザー、集束UV光源、発光ダイオードからなるグループから選択された光源によって提供することができる。

【0316】

第1の相補的核酸分子と、検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。検出可能な標識が欠けた第1のハイブリダイジング核酸分子は、バーコードドメイン内の第1の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0317】

第2の相補的核酸分子と、検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、同じ核酸配列を含むことができる。検出可能な標識が欠けた第2のハイブリダイジング核酸分子は、バーコードドメイン内の第2の付着位置に隣接する一本鎖ヌクレオチドと相補的な核酸配列を含むことができる。

【0318】

上記の方法はさらに、検出可能な標識を記録するのに適した媒体を含むことができる。この媒体として、適切なコンピュータ可読媒体が可能である。

【0319】

このシーケンシング法をここでさらに説明する。

【0320】

図10は、本開示の代表的な1つのシーケンシングサイクル全体の模式図を示している。本開示の方法ではシーケンシングの前に標的核酸を固定化する必要はないが、この例では、この方法は、左最上部の図に示してあるように、標的核酸が捕捉用プローブを用いて捕捉されてフローセルの表面に結合している状態から始まる。次にシーケンシング用プローブのプールをフローセルの中に流入させ、シーケンシング用プローブが標的核酸にハイブリダイズできるようにする。この例では、シーケンシング用プローブは図1に示したものである。これらシーケンシング用プローブは、標的核酸にハイブリダイズする標的結合ドメインの中に6量体配列を含んでいる。6量体は、それぞれの側に(N)塩基

10

20

30

40

50

が隣接している。その(N)塩基は、ユニバーサル塩基／縮重塩基であること、または塩基b₁-b₂-b₃-b₄-b₅-b₆によって指定されない標的に対して特異的ではない4種類の正規塩基のうちの任意のもので構成することが可能である。6量体配列を利用すると、4096(4⁶)通りのシーケンシング用プローブのセットにより、任意の標的核酸をシーケンシングすることが可能になる。この例に関しては、4096通りのシーケンシング用プローブのセットが、それぞれ512のシーケンシング用プローブからなる8つのプールの中の標的核酸にハイブリダイズする。シーケンシング用プローブの標的結合ドメイン内の6量体配列は、図10の中央上方の図に示してあるように、標的核酸の長さに沿って、その6量体と標的核酸の間で完全に相補性が一致している位置にハイブリダイズすることになる。この例では、単一のシーケンシング用プローブが標的核酸にハイブリダイズする。結合しないあらゆるシーケンシング用プローブがフローセルから洗い流される。

10

【0321】

これらシーケンシング用プローブは、上述のように、3つの付着位置R₁、R₂、R₃を有するバーコードドメインも含んでいる。付着位置R₁内の付着領域は、シーケンシング用プローブの6量体の第1のジヌクレオチドに対応する1つ以上のヌクレオチド配列を含んでいる。したがってシーケンシング用プローブの標的結合ドメインの中に存在する第1のジヌクレオチドの属性に対応する相補的核酸を含むレポータプローブだけが付着位置R₁にハイブリダイズすることになる。同様に、シーケンシング用プローブの付着位置R₂内の付着領域は、標的結合ドメインの中に存在する第2のジヌクレオチドに対応し、シーケンシング用プローブの付着位置R₃内の付着領域は、標的結合ドメインの中に存在する第3のジヌクレオチドに対応する。

20

【0322】

この方法は、図10の右最上部の図に続く。レポータプローブのプールをフローセルの中に流入させる。レポータプローブのプールに含まれる各レポータプローブは、2色組み合わせの形態になった検出可能な標識と、シーケンシング用プローブの付着位置R₁の中の対応する付着領域にハイブリダイズすることのできる相補的核酸を含んでいる。2色組み合わせと特定のレポータプローブの相補的核酸は、上に説明したように、可能な16通りのジヌクレオチドのうちの1つに対応する。レポータプローブの各プールは、特定のジヌクレオチドに対応する2色組み合わせがシーケンシング前に確立されるように設計する。例えば図10に示したシーケンシング実験では、付着位置R₁にハイブリダイズするレポータプローブの第1のプールについて、黄-赤という2色組み合わせをアデニン-チミンというジヌクレオチドに対応させることができる。図10の右上の図に示してあるようにレポータプローブが付着位置R₁にハイブリダイズした後、結合しないあらゆるレポータプローブをフローセルから洗い流し、結合したレポータプローブの検出可能な標識を記録して6量体の第1のジヌクレオチドの属性を明らかにする。

30

【0323】

付着位置R₁にハイブリダイズしたレポータプローブに属する検出可能な標識を除去する。検出可能な標識を除去するため、レポータプローブに切断可能なリンカーを含めることと、適切な切断試薬の添加を追加することができる。あるいは検出可能な標識を欠く相補的核酸をシーケンシング用プローブの付着位置R₁にハイブリダイズさせ、検出可能な標識を有するレポータプローブと置き換える。検出可能な標識を除去する方法が何であれ、付着位置R₁はもはや検出可能なシグナルを発生しない。以前に検出可能なシグナルを発生していたバーコードドメインの付着位置がもはや検出可能なシグナルを発生できないようにする方法を、本明細書では「暗化(darkening)」と呼ぶ。

40

【0324】

レポータプローブの第2のプールをフローセルの中に流入させる。レポータプローブのプールに含まれる各レポータプローブは、2色組み合わせの形態になった検出可能な標識と、シーケンシング用プローブの付着位置R₂の中の対応する付着領域にハイブリダイズすることのできる相補的核酸を含んでいる。2色組み合わせと特定のレポータプローブの相補的核酸は、可能な16通りのジヌクレオチドのうちの1つに対応する。ある特定の2色

50

組み合わせが、レポータプローブの第1のプールの文脈では1つのジヌクレオチドに対応し、レポータプローブの第2のプールの文脈では別のジヌクレオチドに対応することが可能である。図10の右下の図に示してあるように、レポータプローブが付着位置 R_2 にハイブリダイズした後、結合しないあらゆるレポータプローブをフローセルから洗い流し、結合したレポータプローブの検出可能な標識を記録して、シーケンシング用プローブの中に存在する6量体の第2のジヌクレオチドの属性を明らかにする。

【0325】

位置 R_2 の検出可能な標識を除去するため、レポータプローブに切断可能なリンカーを含めることと、適切な切断試薬の添加を追加することができる。あるいは検出可能な標識を欠く相補的核酸をシーケンシング用プローブの付着位置 R_2 にハイブリダイズさせ、検出可能な標識を有するレポータプローブと置き換える。検出可能な標識を除去する方法が何であれ、付着位置 R_2 はもはや検出可能なシグナルを発生しない。

【0326】

次にレポータプローブの第3のプールのフローセルの中に流入させる。第3のレポータプローブのプールに含まれる各レポータプローブは、2色組み合わせの形態になった検出可能な標識と、シーケンシング用プローブの付着位置 R_3 の中の対応する付着領域にハイブリダイズすることのできる相補的核酸を含んでいる。2色組み合わせと特定のレポータプローブの相補的核酸は、可能な16通りのジヌクレオチドのうちの1つに対応する。図10の中央下方の図に示してあるようにレポータプローブが位置 R_3 にハイブリダイズした後、結合しないあらゆるレポータプローブをフローセルから洗い流し、結合したレポータプローブの検出可能な標識を記録して、シーケンシング用プローブの中に存在する6量体の第3のジヌクレオチドの属性を明らかにする。このようにして、標的結合ドメインの3つのジヌクレオチドがすべて同定されるため、それらをまとめてアセンブルして標的結合ドメインの配列を明らかにすること、したがって標的核酸の配列を明らかにすることができる。

【0327】

標的核酸のシーケンシングを続けるため、結合したあらゆるシーケンシング用プローブを標的核酸から除去することができる。レポータプローブがバーコードドメインの位置 R_3 にハイブリダイズしたままである場合でさえ、シーケンシング用プローブを標的核酸から除去することができる。あるいは位置 R_3 にハイブリダイズしたレポータプローブは、シーケンシング用プローブを標的結合ドメインから除去する前に、例えば位置 R_1 と R_2 のレポータに関して上に説明した暗化手続きを利用してバーコードドメインから除去することができる。

【0328】

図10に示したシーケンシングサイクルは何回でも繰り返すことができ、各シーケンシングサイクルを、シーケンシング用プローブの同じプールを標的核酸分子にハイブリダイズさせることから始めたり、シーケンシング用プローブの異なるプールを標的核酸分子にハイブリダイズさせることから始めたりする。第1のシーケンシング用プローブ、またはシーケンシング用プローブの第1のプールが第1のシーケンシングサイクルの間に標的核酸に結合した位置と重複する位置でシーケンシング用プローブの第2のプールが標的核酸に結合することが可能である。そうすることで、標的核酸の中のいくつかのヌクレオチドを2回以上シーケンシングすることと、2つ以上のシーケンシング用プローブを用いることができる。

【0329】

図11は、本開示のシーケンシング法の1つのサイクル全体の模式図と、このサイクルの間に回収された対応するイメージングデータを示している。この例では、使用するシーケンシング用プローブは図1に示したものであり、シーケンシング工程は、図10に示して上に説明したのと同じである。シーケンシング用プローブのシーケンシングドメインを標的核酸にハイブリダイズさせた後、レポータプローブをシーケンシング用プローブの第1の付着位置(R_1)にハイブリダイズさせる。次に第1のレポータプローブの画像を取得して色ドットを記録する。図11では、色ドットを点線の円で示してある。色ドット

は、1つのサイクル全体の間記録される単一のシーケンシング用プローブに対応する。この例では、7つのシーケンシング用プローブが記録される(1~7)。次にバーコードドメインの第1の付着位置を暗化し、二重蛍光レポータプローブをシーケンシング用プローブの第2の付着位置(R_2)にハイブリダイズさせる。次に第2のレポータプローブの画像を取得して色ドットを記録する。次にバーコードドメインの第2の付着位置を暗化し、二重蛍光レポータプローブをシーケンシング用プローブの第3の付着位置(R_3)にハイブリダイズさせる。次に第3のレポータプローブの画像を取得して色ドットを記録する。次に各シーケンシング用プローブ1~7からの3つの色ドットを順番に並べる。次にデコーディングマトリックスを用いてそれぞれの色ドットを特定のジヌクレオチドにマッピングしてシーケンシング用プローブ1~7の標的結合ドメインの配列を決定する。

10

【0330】

1つのシーケンシングサイクルの間に標的核酸に結合した任意のシーケンシング用プローブの標的結合ドメインの配列を決定するのに必要なレポータプローブのプールの数は、バーコードドメイン内の付着位置の数と同じである。したがって3つの位置を有するバーコードドメインでは、シーケンシング用プローブ上で3つのレポータプローブをサイクルさせることになる。

【0331】

シーケンシング用プローブの1つのプールは、配列がすべて同じ複数のシーケンシング用プローブを含むこと、または配列がすべて同じではない複数のシーケンシング用プローブを含むことができる。シーケンシング用プローブの1つのプールが、配列がすべて同じではない複数のシーケンシング用プローブを含んでいるとき、異なるそれぞれのシーケンシング用プローブが同数存在すること、または異なるシーケンシング用プローブが異なる数で存在することができる。

20

【0332】

図12は、本開示によるシーケンシング用プローブのプールの構成例を示しており、ここではシーケンシング用プローブが、(a) 6個のヌクレオチド(6量体)を含んでいて標的核酸に特異的に結合する標的結合ドメインと、(b) バーコードドメイン内の3つの付着位置(R_1 、 R_2 、 R_3)を含有しているため、上に具体的に示した8通りの色組み合わせを用いてシーケンシング用プローブの8つの異なるプールを設計する。可能な4096通りの異なる6量体配列($4 \times 4 \times 4 \times 4 \times 4 \times 4 = 4096$)が存在する。バーコードドメイン内の3つの付着位置のそれぞれは、異なる8通りの色組み合わせのうちの1つが結合した相補的核酸にハイブリダイズすることができるため、可能な3つの色組み合わせからなる異なる512のセットが存在する($8 \times 8 \times 8 = 512$)。例えば R_1 が色組み合わせGGに結合した相補的核酸にハイブリダイズし、 R_2 が色組み合わせBGに結合した相補的核酸にハイブリダイズし、 R_3 が色組み合わせYRに結合した相補的核酸にハイブリダイズするプローブでは、3つの色組み合わせのセットはGG-BG-YRである。シーケンシング用プローブの1つのプールの中では、3つの色組み合わせがそれぞれ異なるセットが、標的結合ドメイン内の異なる6量体に対応することになる。各プールは512個の異なる6量体を含有していて合計で4096通りの可能な6量体が存在するため、可能なすべての6量体をシーケンシングするには8つのプールが必要とされる($4096/512=8$)。8つのプールそれぞれの中に入る具体的なシーケンシング用プローブは、各シーケンシング用プローブが標的核酸に最適にハイブリダイズするように決められる。最適なハイブリダイゼーションを保証するにはいくつかの注意事項がある。その注意事項に含まれるのは、(a) 6量体の完全な相補体を異なるプールに分けること；(b) 高い T_m と低い T_m を有する6量体を異なるプールに分けること；(c) 経験的に学習したハイブリダイゼーションパターンに基づいて6量体を異なるプールに分けることである。

30

40

【0333】

図13は、米国特許出願公開第20160194701号に記載されているシーケンシング用プローブと、本開示のシーケンシング用プローブの違いを示している。図13の左図に示してあるように、米国特許出願公開第2016/0194701号には、相補的核酸にハイブリダ

50

イズする6つの付着位置を含むバーコードドメインを有するシーケンシング用プローブが記載されている。それぞれの相補的核酸は異なる4種類の蛍光染料のうちの1つに結合する。この構成では、それぞれの色（赤、青、緑、黄）が、標的結合ドメインの中の1つのヌクレオチド（A、T、C、G）に対応する。プローブをこのように設計することにより、4096通りの異なるプローブ（ 4^6 ）が生まれる。図13の右図に示してあるように、本開示の一例では、各シーケンシング用プローブのバーコードドメインは、相補的核酸にハイブリダイズする3つの付着位置を含んでいる。米国特許出願公開第2016/0194701号とは異なり、これら相補的核酸には、8通りの色組み合わせ（GG、RR、GY、RY、YY、RG、BB、RB）のうちの1つが結合する。それぞれの色組み合わせは、標的結合ドメイン内の特定の1つのジヌクレオチドに対応する。この構成では、512通り（ 8^3 ）の異なるプローブが生まれる。標的核酸ドメイン内の可能なあらゆる6量体の組み合わせ（4096）をカバーしようとする、1個の標的核酸全体をシーケンシングするのにこれら512の異なるプローブからなる8つの別々のプールが必要とされる。相補的核酸を標識するのに8通りの色組み合わせを用いているが可能なジヌクレオチドは16通り存在するため、シーケンシング用プローブのどのプールを用いるかに応じていくつかの色組み合わせは異なるジヌクレオチドに対応することになる。例えば図13では、シーケンシング用プローブの第1、第2、第3、第4のプールにおいて色組み合わせBBがジヌクレオチドAAに対応し、色組み合わせGGがジヌクレオチドATに対応する。シーケンシング用プローブの第5、第6、第7、第8のプールにおいては色組み合わせBBがジヌクレオチドCAに対応し、色組み合わせCTがジヌクレオチドATに対応する。

【0334】

複数のシーケンシング用プローブ（すなわち2つ以上のシーケンシング用プローブ）をシーケンシングウインドウの中でハイブリダイズさせることができる。シーケンシング中、ハイブリダイズした複数のシーケンシング用プローブの中の各シーケンシング用プローブに結合した検出可能な標識の属性と空間的位置を記録する。こうすることで、複数のジヌクレオチドの位置と属性の両方をその後同定することが可能になる。言い換えると、複数のシーケンシング用プローブを同時に単一の標的核酸分子にハイブリダイズさせることにより、この標的核酸に沿った複数の位置を同時にシーケンシングすることができるため、シーケンシングの速度が向上する。

【0335】

いくつかの態様では、単一のシーケンシング用プローブを捕捉された1個の標的核酸にハイブリダイズさせることができる。いくつかの態様では、複数のシーケンシング用プローブを捕捉された1つの標的核酸にハイブリダイズさせることができる。2つのハイブリダイズした5'捕捉用プローブと3'捕捉用プローブの間のシーケンシングウインドウにより、単一のシーケンシング用プローブまたは複数のシーケンシング用プローブを標的核酸分子の長さに沿ってハイブリダイズさせることが可能になる。複数のシーケンシング用プローブを標的核酸分子の長さに沿ってハイブリダイズさせることによって標的核酸分子上の2つ以上の位置を同時にシーケンシングすることができるため、シーケンシングの速度が大きくなる。標的核酸の長さに沿って結合した複数のプローブの個々のプローブからの蛍光シグナルは、空間的に分離することができる。

【0336】

いくつかの態様では、シーケンシング用プローブを標的核酸の長さに沿って均等な間隔で結合させることができる。いくつかの態様では、シーケンシング用プローブは標的核酸の長さに沿って均等な間隔で結合する必要がない。標的核酸の長さに沿って結合した複数のシーケンシング用プローブからのシグナルを空間的に分離することで、標的核酸の複数の位置でシーケンシング情報を同時に得ることができる。

【0337】

標的核酸の長さに沿ったプローブの分布が、検出可能なシグナルの解像度にとって極めて重要である。1つの領域内のプローブが多すぎると検出可能な標識の重複が起こる可能性があるため、2つの近接したプローブの分離が妨げられる。これは以下のように説明さ

10

20

30

40

50

れる。1個のヌクレオチドは長さが0.34 nmであり、シーケンシング装置の横方向（x-y）の空間解像度は約200 nmであるため、シーケンシング装置の解像度の限界は約588塩基対（すなわち1個のヌクレオチド／0.34 nm×200 nm）である。すなわち2つのプローブが互いに約588塩基対以内の位置にあるときには、上記のシーケンシング装置だと標的核酸にハイブリダイズした2つのプローブからのシグナルを分離することができないと考えられる。したがって検出可能な標識を別々の「スポット」として分離できるようにするには、シーケンシング装置の解像度に応じて2つのプローブを約600塩基対離す必要がある。そのため最適な間隔では、標的核酸600 bpごとに1つのプローブが存在すべきである。プローブの集団の中の各シーケンシング用プローブは互いに600ヌクレオチドよりも近い距離で結合しないことが好ましい。（例えば蛍光強度の値と、波長に依存した比を利用する）さまざまなソフトウェアのアプローチを利用して標的核酸の1つの分離可能な領域の中にハイブリダイズするプローブの数をモニタし、制限し、潜在的に解析することが可能であり、それに応じてプローブの集団を設計することができる。さらに、より多くの明確に異なるシグナルを提供する検出可能な標識（例えば蛍光標識）を選択することができる。さらに、文献に記載されている方法（SmallとParthasarthy：「超解像位置特定法」Annu. Rev. Phys. Chem., 2014年；第65巻：107～125ページ）には、構造化された照明と、シーケンシング用顕微鏡の解像限界を数10ナノメートルまで低下させるさまざまな超解像アプローチが記載されている。より大きな解像度のシーケンシング装置を用いると、より短い標的結合ドメインを有するプローブを用いることが可能になる。

【0338】

上述のように、プローブの T_m の設計が、標的核酸にハイブリダイズするプローブの数に影響を与える可能性がある。その代わりに、またはそれに加えて、1つの集団内のシーケンシング用プローブの濃度を上昇させて標的核酸の特定の領域内のプローブのカバレッジを大きくすることができる。例えばシーケンシング装置の解像限界よりも上にするため、シーケンシング用プローブの濃度を低下させて標的核酸の特定の領域内のプローブのカバレッジを小さくすることができる。

【0339】

2つの検出可能な標識の解像限界は約600ヌクレオチドだが、そのことが本開示の強力なシーケンシング法の障害となることはない。いくつかの態様では、任意の集団の中の複数のシーケンシング用プローブが標的核酸上で互いに600ヌクレオチド離れることはなかろう。しかし統計的には（ポアソン分布に従うと）、1つのシーケンシング用プローブが結合しているだけの標的核酸が存在するであろうゆえ、そのシーケンシング用プローブは光学的に分離できる。600ヌクレオチド以内に複数のプローブを有する（ため、光学的に分離できない）標的核酸については、これら分離不能なシーケンシング用プローブに関するデータを廃棄することができる。重要なことだが、本開示の方法では、複数のシーケンシング用プローブを結合させて検出する多数回のラウンドが提供される。したがって何回かのラウンドではあらゆるシーケンシング用プローブからシグナルを検出し、何回かのラウンドではシーケンシング用プローブの一部だけからシグナルを検出し、何回かのラウンドではどのシーケンシング用プローブからもシグナルを検出しないことが可能である。いくつかの態様では、標的核酸に結合したシーケンシング用プローブの分布を（例えば濃度または希釈度を制御することによって）管理し、1個の標的核酸に1つのシーケンシング用プローブだけが結合するようにできる。

【0340】

ランダムにだが、部分的には標的結合ドメインの長さ、と、プローブの T_m と、適用するプローブの濃度に依存して、1つの集団の中の2つの明確に異なるシーケンシング用プローブが互いに600ヌクレオチド以内で結合する可能性がある。

【0341】

その代わりに、またはそれに加えて、例えばシーケンシング装置の解像限界よりも上にするため、1つの集団の中のシーケンシング用プローブの濃度を低下させて標的核酸

の特定の領域の中のプローブのカバレッジを小さくし、そのことによって解像度が制限されたスポットから単一の読み出しを発生させることができる。

【0342】

本開示の方法を用いて標的核酸をシーケンシングする前に、標的核酸の配列、または配列の一部がわかっている場合には、どの2つのシーケンシング用プローブも互いにヌクレオチド600個以内で標的核酸に結合することのないようにシーケンシング用プローブを設計して選択することができる。

【0343】

シーケンシング用プローブを標的核酸にハイブリダイズさせる前に、1個以上の相補的核酸分子に第1の検出可能な標識を結合させ、少なくとも第2の検出可能な標識をシーケンシング用プローブのバーコードドメイン内の1つ以上の付着位置にハイブリダイズさせることができる。例えば標的核酸にハイブリダイズさせる前に、第1の検出可能な標識と少なくとも第2の検出可能な標識が結合した1個以上の相補的核酸分子を各シーケンシング用プローブの第1の付着位置にハイブリダイズさせることができる。したがってシーケンシング用プローブは、標的核酸に接触すると第1の付着位置から検出可能なシグナルを発生させることができるため、バーコードドメイン上の第1の位置に向かう相補的核酸またはレポータプローブの第1のプールを用意する必要はない。別の一例では、第1の検出可能な標識と少なくとも第2の検出可能な標識が結合した1個以上の相補的核酸分子をシーケンシング用プローブのバーコードドメイン内のすべての付着位置にハイブリダイズさせることができる。したがってこの例では、相補的核酸を順番に交換する必要なしに6個のヌクレオチドからなる配列を読み出すことができる。このあらかじめハイブリダイズしたシーケンシング用プローブ-レポータプローブ複合体を用いると説明した方法の多くの工程が省略されるため、配列情報を取得する時間が短縮されることが考えられる。しかしこのプローブは、検出可能な標識が重複していない場合に有益であると考えられる。例えば蛍光体は、重複していない波長の光によって励起される。あるいは蛍光体は、重複していない波長の光を出す。

【0344】

本開示の方法のいくつかの態様では、記録された色ドットからのシグナル強度を用いて標的核酸を正確にシーケンシングすることができる。いくつかの態様では、1つの色ドット内の特定の色のスポット強度を用いて特定の色ドットが1つの色の重複（すなわちBB、GG、YY、RR）である色組み合わせに対応する確率を求めることができる。

【0345】

バーコードドメイン内のある位置の暗化は、その位置にハイブリダイズしたレポータプローブの中に存在する切断可能なリンカーの修飾の位置で鎖を切断することによって実現できる。図14は、切断可能なリンカーの修飾を利用して1つのシーケンシングサイクルの間にバーコード位置を暗化することを示している。図14の最も左側の図に示した第1の工程は、レポータプローブの一次核酸をシーケンシング用プローブの第1の付着位置にハイブリダイズさせることを含んでいる。一次核酸は、バーコードドメインの第1の位置の付着領域内の特定の相補的配列にハイブリダイズする。一次核酸の第1のドメインと第2のドメインは、切断可能なリンカーの修飾によって共有結合している。第2の工程では、検出可能な標識を記録してシーケンシング用プローブの標的結合ドメイン内の特定のジヌクレオチドの属性と位置を求める。第3の工程では、切断可能なリンカーの修飾の位置でレポータプローブを切断することにより、バーコードドメインの第1の位置を暗化する。こうすることで一次核酸の第2のドメインが放出され、そのことによって検出可能な標識が放出される。一次核酸分子の第1のドメインは、今や検出可能な標識がまったく欠けた状態でバーコードドメインの第1の付着位置にハイブリダイズしたままになっているため、バーコードドメインの第1の位置はもはや検出可能なシグナルを発生させず、その後のシーケンシング工程において他のどのレポータプローブにもハイブリダイズすることはできないであろう。図14の最も右側の図に示した最終工程では、レポータプローブがバーコードドメインの第2の位置にハイブリダイズし、シーケンシングが続いていく。

【0346】

バーコードドメインの付着位置は、検出可能な標識が結合しているレポータプローブ内の任意の二次核酸または三次核酸を置換する一方で、レポータプローブの一次核酸はシーケンシング用プローブにハイブリダイズしたままにすることによって暗化することができる。この置換は、検出可能な標識が結合していない二次核酸または三次核酸を一次核酸にハイブリダイズさせることによって実現できる。図15は本開示の代表的なシーケンシングサイクルの一例であり、バーコードドメイン内の1つの位置が、標識を有する二次核酸の置換によって暗化されている。図15の最も左側の図はシーケンシングサイクルの開始を示しており、レポータプローブの一次核酸分子がシーケンシング用プローブのバーコードドメインの第1の付着位置にハイブリダイズしている。次に、検出可能な標識に結合している二次核酸分子が一次核酸分子にハイブリダイズし、検出可能な標識が記録される。バーコードドメインの第1の位置を暗化するため、検出可能な標識に結合している二次核酸分子を、検出可能な標識を欠く二次核酸分子で置換する。シーケンシングサイクルの次の工程では、検出可能な標識を含むレポータプローブをバーコードドメインの第2の位置にハイブリダイズさせる。バーコードドメインの付着位置の暗化は、検出可能な標識が結合していない核酸をシーケンシング用プローブにバーコードドメインの対応する付着位置でハイブリダイズさせてレポータプローブの任意の一次核酸分子を置換することにより可能である。バーコードドメインが少なくとも1つの付着位置に近接または隣接した少なくとも1つの一本鎖核酸配列を含んでいる場合には、検出可能な標識が結合していない核酸が一次核酸分子と置き換わることが、そのフランキング配列と、バーコードドメインのうちでその一次核酸分子によって占められている部分とにハイブリダイズすることによって可能になる。必要な場合には、検出可能な標識の交換速度を大きくする小さな一本鎖オリゴヌクレオチドを組み込むことによって検出可能な標識の交換速度を大きくすることができる（例えば「トールホールド」プローブ；例えばSeeling他、「準安定DNA燃料の触媒による緩和」；J. Am. Chem. Soc. 2006年、第128巻(37)、12211～12220ページを参照されたい）。

【0347】

検出可能な標識を含む相補的核酸すなわちレポータプローブは、付着領域から除去することができるが、検出可能な標識を欠くハイブリダイズする核酸で置き換えることはできない。これは、例えばカオトロピック剤を添加することによって、および／または温度を上昇させることによって、および／または塩の濃度を変えることによって、および／またはpHを調整することによって、および／または流体力を加えることによって起こすことができる。これらの例では、より少数の試薬（すなわち検出可能な標識を欠くハイブリダイズする核酸）が必要とされる。

【0348】

本開示の方法を利用して同じサンプルからRNA分子とDNA分子（mRNAとgDNAが含まれる）を同時に捕捉してシーケンシングすることができる。同じサンプルからRNA分子とDNA分子の両方を捕捉してシーケンシングすることを同じフローセルの中で実行できる。いくつかの態様では、本開示の方法を利用してFFPEサンプルからgDNAとmRNAの両方を同時に捕捉し、検出し、シーケンシングすることができる。

【0349】

本開示のシーケンシング法はさらに、固定化された標的核酸の各領域についてそれぞれ同定されたヌクレオチドの線形順序をアセンブルすることにより、固定化された標的核酸の配列を同定し、それによって、任意により固定化された前記標的核酸のための配列を同定する工程を含んでいる。アセンブルする工程では、実行可能なプログラムが記憶されている非一時的コンピュータ可読記憶媒体を利用する。そのプログラムが、標的核酸の各領域についてそれぞれ同定されたヌクレオチドの線形順序をアセンブルする指示をマイクロプロセッサに出すことにより、核酸の配列が得られる。アセンブルする作業は、「リアルタイム」で、すなわちすべてのデータが回収された後や完全なデータ取得後ではなく、データがシーケンシング用プローブから回収されている間に実行することができる。

【0350】

本開示のシーケンシング法の生の特異度は約94%である。本開示のシーケンシング法の精度は、2つ以上のシーケンシング用プローブを用いて標的核酸内の同じ塩基をシーケンシングすることによってほぼ99%まで上昇させることができる。図16は、本開示のシーケンシング法によりどのようにして異なるシーケンシング用プローブを用いて標的核酸の同じ塩基のシーケンシングが可能になるかを示している。この例における標的核酸は、NRASエキソン2の断片（配列番号1）である。目的の特定の塩基は、標的核酸の中で強調されているシトシン（C）である。目的のこの塩基は、標的核酸に対するハイブリダイゼーションのフットプリントが異なる2つのシーケンシング用プローブにハイブリダイズすることになる。この例では、シーケンシング用プローブ1~4（バーコード1~4）は、目的の塩基の左側の3個のヌクレオチドに結合するのに対し、シーケンシング用プローブ5~8（バーコード5~8）は、目的の塩基の左側の5個のヌクレオチドに結合する。すると目的の塩基が異なる2つのプローブによってシーケンシングされるため、その特定の位置での塩基読み取りの量が増加し、そのことによってその特定の位置における全体的な精度が上昇する。図17は、標的ヌクレオチド上の特定のヌクレオチド位置における複数の異なる塩基読み取りが1つ以上のシーケンシング用プローブから記録された後、どのように組み合わせられてコンセンサス配列（配列番号2）を作り出し、そのことによって最終的な塩基読み取りの精度を向上させているかを示している。

10

【0351】

「Hyb & Seq化学」、「Hyb & Seqシーケンシング」、「Hyb & Seq」という用語は、上に説明した本開示の方法を意味する。

20

【0352】

本開示のアレイと、このアレイを利用する方法

【0353】

本明細書に詳しく記載されているように、本開示により、核酸分子を固定化するための組成物と方法（アレイと、アレイを利用する方法が含まれる）が提供される。

【0354】

本開示により、平坦な固体支持基板と；この平坦な固体支持基板上の第1の層と；この第1の層の上の第2の層を含んでいて；第2の層が複数のナノウェルを含み、各ナノウェルが第1の層の露出部分へのアクセスを提供し、各ナノウェルは、第1の層の露出部に結合した複数の第1のオリゴヌクレオチドを含んでいる組成物が提供される。

30

【0355】

本開示により、平坦な固体支持基板と；この平坦な固体支持基板上にあってこの平坦な固体支持基板の第1の表面と接触している第1の層と；この第1の層の上にあるこの第1の層の第2の表面と接触している第2の層を含んでいて；第1の層の第2の表面が平坦な固体支持基板の表面と接触しておらず、第2の層が複数のナノウェルを含み、各ナノウェルが第1の層の露出部分へのアクセスを提供し、各ナノウェルは、第1の層の露出部に結合した複数の第1のオリゴヌクレオチドを含んでいる組成物が提供される。

【0356】

第1の層は、平坦な固体支持基板の表面と接触している第1の表面と、第2の層と接触しているが平坦な固体支持基板の表面とは接触していない第2の表面を含むことができる。

40

【0357】

第2の層は、第1の層の表面と接触している第1の表面と、環境に露出している第2の表面を含むことができる。

【0358】

図47は、本発明の代表的なアレイの断面模式図である。このアレイは、平坦な固体支持基板101と、この平坦な固体支持基板101上の第1の層102と、この第1の層102の上の第2の層103を含んでいる。第2の層103は、複数のナノウェル104を含んでいる。各ナノウェル104は2つの側が開放されているため、各ナノウェル内で第1の層の一部105が露出している。複数の第1のオリゴヌクレオチド106が各ナノウェルの中で露出した第1の層105

50

に共有結合している。

【0359】

いくつかの態様では、平坦な固体支持基板として、表面、膜、ビーズ、多孔性材料、電極のいずれかが可能である。平坦な固体支持基板の非限定的な例に含まれるのは、例えばポリマー材料、金属、ケイ素、ガラス、石英である。

【0360】

いくつかの態様では、第1の層102は、酸化物フィルムを含むことができ、その非限定的な例は二酸化ケイ素である。

【0361】

いくつかの態様では、第1の層102は厚みを約50～約150 nmにすることができる。第1の層102は厚みを約90 nmにすることができる。

10

【0362】

いくつかの態様では、第2の層103は、非限定的な例としてビス(トリメチルシリル)アミン(ヘキサメチルジシラザン(HMDSまたはHDMS))としても知られる)を含むことができる。

【0363】

いくつかの態様では、第2の層103は、化学的に反応しない材料を含むことができるため、第2の層は生体高分子に結合しない。

【0364】

いくつかの態様では、第2の層103は厚みを約1 nm～約10 nmにすることができる。第2の層103は厚みを約3 nm～約4 nmにすることができる。

20

【0365】

いくつかの態様では、平坦な固体支持基板はケイ素を含み、第1の層は二酸化ケイ素を含み、第2の層はHMDSを含んでいる。

【0366】

いくつかの態様では、平坦な固体支持基板はガラスを含み、第1の層は二酸化ケイ素を含み、第2の層はHMDSを含んでいる。

【0367】

いくつかの態様では、第2の層は、1平方ミリメートル当たり約 0.1×10^5 ～約 100×10^7 個のナノウエルを含むことができる。第2の層は、1平方ミリメートル当たり約 0.1×10^6 ～約 100×10^6 個のナノウエルを含むことができる。第2の層は、1平方ミリメートル当たり約 1×10^6 ～約 10×10^6 個のナノウエルを含むことができる。第2の層は、1平方ミリメートル当たり約 2×10^6 ～約 5×10^6 個のナノウエルを含むことができる。第2の層は、1平方ミリメートル当たり約 3×10^6 個のナノウエルを含むことができる。

30

【0368】

本明細書では、「ナノウエルの密度」は、特定の表面積の中に存在するナノウエルの数を意味する。例えば表面積が 1.0 mm^2 であり、 1.0×10^6 ナノウエルを含む第2の層は、ナノウエルの密度が 1.0×10^6 ナノウエル/ mm^2 であると言われる。

【0369】

いくつかの態様では、ナノウエルの密度は、約 0.1×10^5 と約 100×10^7 ナノウエル/ mm^2 の間にすることができる。ナノウエルの密度は、約 0.1×10^6 と約 100×10^6 ナノウエル/ mm^2 の間にすることができる。ナノウエルの密度は、約 1×10^6 と約 10×10^6 ナノウエル/ mm^2 の間にすることができる。ナノウエルの密度は、約 2×10^6 と約 5×10^6 ナノウエル/ mm^2 の間にすることができる。ナノウエルの密度は、約 3×10^6 ナノウエル/ mm^2 にすることができる。

40

【0370】

いくつかの態様では、ナノウエル内の第1の層の露出部分の表面積は、約200～約50,000 nm^2 が可能である。ナノウエル内の第1の層の露出部分の表面積は、約300～約40,000 nm^2 が可能である。ナノウエル内の第1の層の露出部分の表面積は、約700～約8,000 nm^2 が可能である。ナノウエル内の第1の層の露出部分の表面積は、約2,000～約3,000

50

nm²が可能である。

【0371】

いくつかの態様では、各ナノウエル内の第1の層の露出部分は円形である。いくつかの態様では、各ナノウエル内の第1の層の露出部分は楕円形である。いくつかの態様では、各ナノウエル内の第1の層の露出部分は長方形である。いくつかの態様では、各ナノウエル内の第1の層の露出部分は正方形である。いくつかの態様では、各ナノウエル内の第1の層の露出部分は六角形または八角形である。いくつかの態様では、各ナノウエル内の第1の層の露出部分は、正多角形の形状を有する。いくつかの態様では、各ナノウエル内の第1の層の露出部分は、不規則な多角形の形状を有する。

【0372】

ナノウエル内の第1の層の露出部分が円形であるいくつかの態様では、第1の層の露出部分は、約10 nm～約200 nmの直径を有することができる。第1の層の露出部分は、約20 nm～約200 nmの直径を有することができる。第1の層の露出部分は、約30 nm～約100 nmの直径を有することができる。第1の層の露出部分は、約50 nm～約60 nmの直径を有することができる。第1の層の露出部分は、約60 nmの直径を有することができる。

【0373】

いくつかの態様では、ナノウエルは円柱が可能である。ナノウエルは長方形が可能である。ナノウエルは直方体が可能である。ナノウエルは多面体が可能である。ナノウエルは、直円柱の形状を有することができる。ナノウエルは、楕円柱の形状を有することができる。ナノウエルは、四角柱の形状を有することができる。ナノウエルは、立方体の形状を有することができる。ナノウエルは、三角柱の形状を有することができる。ナノウエルは、円錐の形状を有することができる。ナノウエルは、ピラミッドの形状を有することができる。ナノウエルは、正四角錐の形状を有することができる。

【0374】

いくつかの態様では、ナノウエルの一部は、同じナノウエルの別の部分と異なるサイズを有することができる。図48に示されている非限定的な一例では、本開示のアレイは、平坦な固体支持基板201と、この平坦な固体支持基板201上の第1の層202と；この第1の層202の上の第2の層203を含んでいる。第2の層203は、複数のナノウエル204を含んでいる。各ナノウエル204は2つの側が開放されているため、各ナノウエル205内で第1の層の一部が露出している。ナノウエルで第1の層202に最も近い部分は、ナノウエルで第1の層から最も遠い部分の直径207よりも小さい直径206を有する。

【0375】

いくつかの態様では、複数のナノウエルをランダムなパターンに配置することができる。本明細書では、「ランダムなパターンに配置された」、または「ランダムにパターン化された」、または「ランダム」という表現は、意図的な設計（またはそのような設計を実現することのできるプログラム）や特定の配置を通じて実現されているのではない凹凸（features）（ナノウエルやパッドなど）が、規則正しくない非デカルト座標式分布をしていることを意味する（言い換えると、グリッドのx軸とy軸に沿ったあらかじめ決められた点に配置されていない、あるいは放射状パターンの中心から規定の「時計回りの位置」、または角度、または半径に配置されていない）。凹凸の「ランダムにパターン化された」セットは、どの凹凸も特定の部位に向かうことがないようなやり方で、溶液、乳液、エアロゾル、蒸気、乾燥調製物のいずれかを支持体の表面に落下させること、またはスプレーすること、または置くこと、または広げることによって実現できる。図49は本開示の代表的なアレイの模式図を示しており、複数の円柱形ナノウエル301がランダムなパターンに配置されている。

【0376】

いくつかの態様では、複数のナノウエルを規則正しいパターンに配置することができる。本明細書では、「規則正しいパターンに配置」または「規則正しいパターン」という表現は、グリッドのx軸とy軸に沿った規則的なあらかじめ決められた点に配置されるか、放射状パターンの中心から規定の「時計回りの位置」、または角度、または半径の位置に配

10

20

30

40

50

置された凹凸の分布を意味する。いくつかの態様では、複数のナノウェルを規則正しいグリッドに配置することができる。規則正しいグリッドは、約50 nm～約3500 nmのピッチを有することができる。規則正しいグリッドは、約100 nm～約3200 nmのピッチを有することができる。規則正しいグリッドは、約300 nm～約1000 nmのピッチを有することができる。規則正しいグリッドは、約440 nm～約710 nmのピッチを有することができる。規則正しいグリッドは、約575 nmのピッチを有することができる。規則正しいグリッドは、約577 nmのピッチを有することができる。本明細書では、「ピッチ」という用語は、規則正しいグリッド上に配置された隣り合った2つの凹凸の中心間の距離を意味する。図50は、一定のピッチ402で規則正しいグリッドに配置されたナノウェル401を含む本開示の代表的なアレイの模式図を示している。

10

【0377】

アレイおよび／または基板に付着したオリゴヌクレオチドを本明細書ではローンオリゴヌクレオチドと呼ぶこともできる。

【0378】

いくつかの態様では、第1のオリゴヌクレオチドの核酸の長さをヌクレオチド約10個～約100個にすることができる。第1のオリゴヌクレオチドの核酸の長さはヌクレオチド約20個～約40個にすることができる。

【0379】

いくつかの態様では、第1のオリゴヌクレオチドは、少なくとも1個の天然塩基を含むことができる。第1のオリゴヌクレオチドは、天然塩基を含まないことができる。第1のオリゴヌクレオチドは、少なくとも1個の修飾されたヌクレオチドまたは核酸類似体を含むことができる。第1のオリゴヌクレオチドは、修飾されたヌクレオチドまたは核酸類似体を含まないことができる。第1のオリゴヌクレオチドは、少なくとも1個のユニバーサル塩基を含むことができる。第1のオリゴヌクレオチドは、ユニバーサル塩基を含まないことができる。第1のオリゴヌクレオチドは、少なくとも1個の縮重塩基を含むことができる。第1のオリゴヌクレオチドは、縮重塩基を含まないことができる。

20

【0380】

いくつかの態様では、第1のオリゴヌクレオチドは、RNA、D-DNA、L-DNA、LNA、イソグアニン、イソシトシン、非塩基性ヌクレオチドのいずれか、またはこれらの任意の組み合わせを含むことができる。

30

【0381】

いくつかの態様では、複数の第1のオリゴヌクレオチドのそれぞれの第1のオリゴヌクレオチドは、同じ配列を含むことができる。複数の第1のオリゴヌクレオチドのそれぞれの第1のオリゴヌクレオチドは、異なる配列を含むことができる。

【0382】

いくつかの態様では、複数のナノウェルの各ナノウェルは、同じ複数の第1のオリゴヌクレオチドを含むことができる。別の態様では、複数のナノウェルの各ナノウェルは、異なる複数の第1のオリゴヌクレオチドを含むことができる。

【0383】

複数の第1のオリゴヌクレオチドは、2個、3個、4個、5個、6個、7個、8個、9個、10個、50個、100個、200個、300個、400個、500個、600個、700個、800個、900個、1,000個、10,000個、100,000個、1,000,000個いずれかの第1のオリゴヌクレオチドを含むことができる。

40

【0384】

本開示のアレイは、本分野で知られている方法（その非限定的な例に含まれるのはフォトリソグラフィなどである）を用いて作製することができる。アレイを作製する方法は、米国特許第7,250,371号と米国特許出願公開第2006/0134917号に記載されており、その内容の全体が参照によって本明細書に組み込まれている。

【0385】

いくつかの態様では、生体高分子（第1のオリゴヌクレオチド、標的核酸複合体など）

50

の表面に存在する特定の化学基を引き付けてその化学基に結合させるため、反応性部分を用いて各ナノウェル内の第1の層の露出部分を機能化することができる。これら官能基は、さまざまな共役化学を通じて生体高分子を特異的に引き付けてその生体高分子に結合できることがよく知られている。代表的な反応性部分の非限定的な例に含まれるのは、エポキシ、アルデヒド、金、ヒドラジド、スルフヒドリル、NHS-エステル、アミン、アルキン、アジド、チオール、カルボン酸塩、マレイミド、ヒドロキシメチルホスフィン、(3-アミノプロピル)トリメチルオキシシラン、イミドエステル、イソシアネート、ヒドロキシル、ペンタフルオロフェニルエステル、ソラーレン、ピリジルジスルフィドまたはビニルスルホン、ポリエチレングリコール (PEG)、ヒドロゲルのいずれか、またはこれらの混合物を含む表面である。

10

【0386】

いくつかの態様では、光反応性部分を用いて生体高分子（その非限定的な例は核酸分子である）を各ナノウェル内の第1の層の露出部に付着させることができる。光反応性部分のいくつかの例に含まれるのは、アジ化アリール（N((2-ピリジルジチオ)エチル)-4-アジドサリシルアミドなど）；フッ化アジ化アリール（4-アジド-2,3,5,6-テトラフルオロ安息香酸など）；ベンゾフェノン系試薬（4-ベンゾイル安息香酸のスクシンイミジルエステルなど）；5-ブromo-デオキシウリジンである。

【0387】

本開示により、サンプルからの少なくとも1個の標的核酸を固定化する方法として、a) 本開示の組成物を用意し；b) この組成物を少なくとも1個の標的核酸と接触させ、この少なくとも1個の標的核酸を第1のオリゴヌクレオチドにハイブリダイズさせることにより少なくとも1個の標的核酸を固定化することで、2個以上の標的核酸が単一のナノウェルの中に固定化されることがないようにすることを含む方法が提供される。

20

【0388】

図51は、図47に示した代表的なアレイの模式図において1つの標的核酸複合体501が各ナノウェルの中に固定化されることを示している。

【0389】

本開示の固定化法は、アレイ上のナノウェルのサイズ（例えば体積）が理由で、2つ以上の標的核酸複合体が単一のナノウェル内でハイブリダイズするのが阻止されるという利点を有する。したがって固定化後のアレイ上の各ナノウェルは、標的核酸複合体を1個または0個含有している。これは、一分子シーケンシングなどの用途で有利である。単一のナノウェル内で標的核酸複合体を1つだけ固定化することにより、個々の標的核酸複合体を個別に調べることが、例えば本分野で知られている蛍光プローブ（例えば米国特許第8,148,512号、第7,473,767号、第7,919,237号、第7,941,279号、第8,415,102号、第8,492,094号、第8,519,115号、米国特許出願公開第2009/0220978号、第2009/0299640号、第2010/0015607号、第2010/0261026号、第2011/0086774号、第2011/0145176号、第2011/0201515号、第2011/0229888号、第2013/0004482号、第2013/0017971号、第2013/0178372号、第2013/0230851号、第2013/0337444号、第2013/0345161号、第2014/0005067号、第2014/0017688号、第2014/0037620号、第2014/0087959号、第2014/0154681号、第2014/0162251号、第2016/0194701号を参照されたい。そのそれぞれは、その全体が参照によって本明細書に組み込まれている）を用いて可能になる。

30

40

【0390】

図52は、図47に示した代表的なアレイの模式図において1つの標的核酸複合体601が各ナノウェルの中に固定化されることを示している。標的核酸複合体602は最も左側のナノウェル603には結合することができない。というのもナノウェル603はすでに標的核酸複合体601で占められているからである。

【0391】

いくつかの態様では、ナノウェルは、1個の標的核酸の排除体積とほぼ等しい体積、または等しい体積、またはそれよりも少ない体積を有することができるため、1つの標的核

50

酸複合体だけが、所与の時刻に単一のナノウェルに物理的にフィットすることができる。

【0392】

いくつかの態様では、ナノウェル同士の間隔は、1つのナノウェルの中に固定化された生体高分子にハイブリダイズ／結合した蛍光プローブからのシグナルを、隣のナノウェルの中に固定化された生体高分子にハイブリダイズ／結合した蛍光プローブからのシグナルから光学的に識別できるようにされている。

【0393】

いくつかの態様では、標的核酸は、約10個～約100,000個のヌクレオチドを含んでいる。いくつかの態様では、標的核酸は、約100個、または約250個、または約500個、または約750個、または約1,000個、または約5,000個、または約10,000個、または約100,000個、または約1,000,000個、または約10,000,000個のヌクレオチドを含むことができる。

10

【0394】

いくつかの態様では、標的核酸は、非限定的な例として、さまざまな長さを有することができるポリマー形態のヌクレオチドを含むことができる（デオキシリボヌクレオチドまたはリボヌクレオチド、またはその類似体）。標的核酸の非限定的な例に含まれるのは、遺伝子、遺伝子断片、エキソン、イントロン、遺伝子間DNA（その非限定的な例にヘテロクロマチンDNAが含まれる）、メッセンジャーRNA（mRNA）、トランスファーRNA、リボソームRNA、リボザイム、小さな干渉性RNA（siRNA）、非コードRNA（ncRNA）、cDNA、組み換えポリヌクレオチド、分岐したポリヌクレオチド、プラスミド、ベクター、1つの配列から単離されたDNA、1つの配列から単離されたRNA、核酸プローブ、プライマーである。標的核酸の属性および／または配列は既知であることが可能である。あるいは属性および／または配列は未知である。標的核酸の配列の一部が既知であることも可能である。

20

【0395】

標的核酸は、任意の核酸サンプルまたは核酸供給源（例えば任意の細胞、組織、生物、インビトロ、化学的合成装置など）から取得することができる。標的核酸は、本分野で知られている任意の方法によって取得することができる。標的核酸は、臨床対象の血液サンプルから取得することができる。標的核酸は、生物サンプルから単離することができる。標的核酸は、本分野で周知の方法とキットを用いて供給源またはサンプルから抽出、または単離、または精製することができる。いくつかの態様では、組織サンプルは、生検された腫瘍またはその一部、すなわち臨床で重要な組織サンプルである。例えば腫瘍は乳がん由来するものが可能である。サンプルはリンパ節から切除することができる。

30

【0396】

標的核酸は、本分野で知られている任意の手段によって断片化することができる。断片化は酵素的手段または機械的手段によって実施することが好ましい。機械的手段として、超音波処理または物理的剪断が可能である。酵素的手段は、ヌクレアーゼ（例えばデオキシリボヌクレアーゼI（DNアーゼI））または1つ以上の制限エンドヌクレアーゼを用いる消化によって実施することができる。標的核酸は、CRISPRに基づく断片化工程を利用して断片化することができる。CRISPRに基づく断片化工程により、標的断片化と、望むサイズの標的核酸の生成が可能になる。非限定的な一例では、CRISPRに基づく断片化により、2つ以上の標的核酸複合体が単一のナノウェル内に固定化されることを阻止するサイズの標的核酸を生成させることができる。

40

【0397】

本開示の方法のいくつかの態様では、標的核酸は、第2のオリゴヌクレオチドを介して第1のオリゴヌクレオチドに間接的にハイブリダイズさせることができる。ただし第2のオリゴヌクレオチドは、第1のオリゴヌクレオチドにハイブリダイズする第1の領域と、上記少なくとも1個の標的核酸にハイブリダイズする第2のオリゴヌクレオチドを含んでいる。

【0398】

第2のオリゴヌクレオチドは、少なくとも1個の天然塩基を含むことができる。第2のオ

50

リゴヌクレオチドは、天然塩基を含まないことができる。第2のオリゴヌクレオチドは、少なくとも1個の修飾されたヌクレオチドまたは核酸類似体を含むことができる。第2のオリゴヌクレオチドは、修飾されたヌクレオチドまたは核酸類似体を含まないことができる。第2のオリゴヌクレオチドは、少なくとも1個のユニバーサル塩基を含むことができる。第2のオリゴヌクレオチドは、ユニバーサル塩基を含まないことができる。第2のオリゴヌクレオチドは、少なくとも1個の縮重塩基を含むことができる。第2のオリゴヌクレオチドは、縮重塩基を含まないことができる。

【0399】

第2のオリゴヌクレオチドは、RNA、D-DNA、L-DNA、LNA、イソグアニン、イソシトシン、非塩基性ヌクレオチドのいずれか、またはこれらの任意の組み合わせを含むことができる。第2のオリゴヌクレオチドはL-DNAを含むことができる。第2のオリゴヌクレオチドは、サンプルを同定する配列を含むことができる。

10

【0400】

本開示の方法のいくつかの態様では、標的核酸は第2のオリゴヌクレオチドと第3のオリゴヌクレオチドを介して第1のオリゴヌクレオチドに間接的にハイブリダイズすることができる。ただし第2のオリゴヌクレオチドは、第1のオリゴヌクレオチドにハイブリダイズする第1の領域と、第3のオリゴヌクレオチド上の第1の領域にハイブリダイズする第2の領域を含み、第3のオリゴヌクレオチドは、第2のオリゴヌクレオチドの第2の領域にハイブリダイズする第1の領域と、標的核酸にハイブリダイズする第2の領域を含んでいる。

【0401】

20

第3のオリゴヌクレオチドは、少なくとも1個の天然塩基を含むことができる。第3のオリゴヌクレオチドは、天然塩基を含まないことができる。第3のオリゴヌクレオチドは、少なくとも1個の修飾されたヌクレオチドまたは核酸類似体を含むことができる。第3のオリゴヌクレオチドは、修飾されたヌクレオチドまたは核酸類似体を含まないことができる。第3のオリゴヌクレオチドは、少なくとも1個のユニバーサル塩基を含むことができる。第3のオリゴヌクレオチドは、ユニバーサル塩基を含まないことができる。第3のオリゴヌクレオチドは、少なくとも1個の縮重塩基を含むことができる。第3のオリゴヌクレオチドは、縮重塩基を含まないことができる。

【0402】

第3のオリゴヌクレオチドは、RNA、D-DNA、L-DNA、LNA、イソグアニン、イソシトシン、非塩基性ヌクレオチドのいずれか、またはこれらの任意の組み合わせを含むことができる。第3のオリゴヌクレオチドはL-DNAを含むことができる。第3のオリゴヌクレオチドは、サンプルを同定する配列を含むことができる。

30

【0403】

本開示の方法のいくつかの態様では、標的核酸は第1のオリゴヌクレオチドに直接ハイブリダイズすることができる。標的核酸は、この標的核酸の3'末端に連結した核酸プローブを介して第1のオリゴヌクレオチドに直接ハイブリダイズすることができる。

【0404】

核酸プローブは、少なくとも1個の天然塩基を含むことができる。核酸プローブは、天然塩基を含まないことができる。核酸プローブは、少なくとも1個の修飾されたヌクレオチドまたは核酸類似体を含むことができる。核酸プローブは、修飾されたヌクレオチドまたは核酸類似体を含むことができる。核酸プローブは、少なくとも1個のユニバーサル塩基を含むことができる。核酸プローブは、ユニバーサル塩基を含まないことができる。核酸プローブは、少なくとも1個の縮重塩基を含むことができる。核酸プローブは、縮重塩基を含まないことができる。

40

【0405】

標的核酸複合体の核酸プローブは、RNA、D-DNA、L-DNA、LNA、イソグアニン、イソシトシン、非塩基性ヌクレオチドのいずれか、またはこれらの任意の組み合わせを含むことができる。

【0406】

50

核酸プローブは、切断可能なリンカーを含むことができる。切断可能なリンカーとして、光で切断可能なリンカー、または酵素で切断可能なリンカーが可能である。

【0407】

本開示の方法のいくつかの態様では、標的核酸を固定化した後、この標的核酸をシークエンシングすることができる。標的核酸は、本分野で知られている任意の方法を用いてシークエンシングすることができる。標的核酸は、直接的な検出法によってシークエンシングすることが好ましい（例えばWO 2016/081740、WO 2018/094385を参照されたい。それぞれの全体が、参照によって本明細書に組み込まれている）。

【0408】

本明細書では、「アレイ」という用語は最も広い意味で用いられ、複数の凹凸を含む基板を意味する。なお1つの凹凸は、固定化された1個以上の生体高分子を含んでいる、および／または1つの凹凸は、1個以上の生体高分子を捕捉して固定化することができる。したがっていくつかの態様では、本発明の組成物をアレイと見なすことができる。

10

【0409】

本明細書では、「凹凸 (feature)」、「パッド」、「スポット」、「ナノウェル」という用語は交換可能に用いられ、固定化された1個以上の生体高分子を含む、および／または1個以上の生体高分子を捕捉して固定化することのできる構造および／または領域を意味する。

【0410】

本明細書では、「生体高分子」という用語は最も広い意味で用いられ、有機分子を意味し、その非限定的な例は、炭水化物、脂質、タンパク質、ペプチド、核酸分子である。

20

【0411】

本明細書では、「固定化された」という用語は、核酸分子と表面の間の連結を意味する。連結として、非共有結合が可能である。例えば標的核酸は、表面に共有結合しているオリゴヌクレオチドにハイブリダイズさせることによって固体支持基板に固定化することができる。

【0412】

本明細書では、「排除体積」は、特定の1個の分子が占めていて他のそのような分子が排除される空間の体積を意味する。

【0413】

上記の任意の態様を本明細書に開示されている他の任意の態様と組み合わせることができる。

30

【0414】

定義

【0415】

本明細書では、「アニーリング」と「ハイブリダイゼーション」という用語は交換可能に用いられ、安定な二本鎖の形成を意味する。1つの態様では、安定な二本鎖とは、二本鎖構造が、温度が二本鎖のうちの1本の鎖のT_mよりも約5℃下または約5℃上、かつ1価の塩の濃度が低い（例えば0.2 M未満、または0.1 M未満、または当業者に知られている塩濃度）という条件でのストリンジェントな洗浄によって破壊されないことを意味する。「完全に一致した」という表現は、二本鎖に関して用いるときには、その二本鎖を構成するポリヌクレオチド鎖および／またはオリゴヌクレオチド鎖が互いに二本鎖構造を形成し、それぞれの鎖に含まれるすべてのヌクレオチドが、他方の鎖に含まれるヌクレオチドとワトソン-クリック塩基対合を形成することを意味する。「二本鎖」という用語には、使用可能なヌクレオシド類似体（デオキシイノシンなど）、2-アミノプリン塩基を有するヌクレオシド、PNAなどのペアが含まれるが、これらに限定されない。二本鎖における2個のオリゴヌクレオチドの間の「ミスマッチ」は、その二本鎖の中のヌクレオチドのペアがワトソン-クリック結合を形成できないことを意味する。

40

【0416】

本明細書では、「ハイブリダイゼーション条件」という用語は、典型的には、約1 M未

50

満、より一般には約500 mM未満、さらに一般には約200 mM未満の塩の濃度を含むことになる。ハイブリダイゼーションの温度として5℃という低温が可能だが、典型的には22℃超、より典型的には約30℃超であり、約37℃を超えることがしばしばある。ハイブリダイゼーションは、通常はストリンジェントな条件（例えばプローブが標的の対応する部分配列に特異的にハイブリダイズするであろう条件）下で実施される。ストリンジェントな条件は配列に依存しており、状況が異なれば異なる。より長い断片は、特異的なハイブリダイゼーションのためにはより高いハイブリダイゼーション温度を必要とする可能性がある。他の因子（相補的な鎖の塩基組成と長さ、有機溶媒の存在、塩基ミスマッチの程度が含まれる）がハイブリダイゼーションのストリンジェンシーに影響を与える可能性があるため、パラメータの組み合わせのほうが、どれか1つだけのパラメータの絶対的な指標よりも重要である。

10

【0417】

一般に、ストリンジェントな条件は、規定されたイオン強度とpHにおける特定の配列のT_mよりも約5℃低くなるように選択される。ストリンジェントな条件の例に含まれるのは、pH 7.0～8.3かつ少なくとも25℃の温度において少なくとも0.01 M～1 M以下のNaイオン濃度（または他の塩）である。例えば5×SSPE（750 mM NaCl、50 mMリン酸ナトリウム、5 mM EDTA、pH 7.4）かつ温度25～30℃という条件が、アレルト異質的プローブのハイブリダイゼーションに適している。ストリンジェントな条件に関しては、例えばSambrook、Fritzsche、Maniatis、『Molecular Cloning A Laboratory Manual』、第2版、Cold Spring Harbor Press（1989年）と、Anderson、『Nucleic Acid Hybridization』、初版、BIOS Scientific Publishers Limited社（1999年）を参照されたい。本明細書では、「～に特異的にハイブリダイズする」または「特異的に～にハイブリダイズする」や同様の表現は、ストリンジェントな条件下である分子が実質的に特定の1つまたは複数のヌクレオチド配列に結合すること、または二本鎖化すること、またはハイブリダイズすることを意味する。

20

【0418】

プローブの特定の位置に付随する検出可能な標識は、1回または複数回「読み出すこと」（例えばその蛍光の検出）が可能である。「読み出し」は、「塩基読み取り」という用語の同義語として使用できる。複数回の読み出しによって精度が向上する。標的核酸配列は、元の単一の標的分子に由来する配列情報の連続ストレッチが検出されるときに「読み出される」。典型的には、これは、（下に定義する）マルチパスコンセンサスを通じて生成する。本明細書では、「カバレッジ（coverage）」または「カバレッジの深度（depth of coverage）」という用語は、（別々の読み出しを通じて）標的の1つの領域をシーケンシングして参照配列とアラインメントする回数を意味する。読み出しカバレッジは、特定の参照標的配列にマッピングする読み出しの合計数である。塩基カバレッジは、特定のゲノム位置における塩基読み取りの合計数である。

30

【0419】

1つの「読み出し（read）」は、配列出力の1つの単位である。元の単一の標的分子に由来する配列情報の連続ストレッチ。各読み出しは、その読み出しの中の塩基読み取りの信頼レベルと関係した品質指標を有する。シーケンサ出力の単位。元の単一の標的分子に由来する配列情報の連続ストレッチ。Hyb & Seqでは、すべての読み出しがマルチパスコンセンサスを通じて生成する。

40

【0420】

「読み出し長（readlength）」は、各読み出しからの配列の長さを記述する指標である（単位はbp）。この指標はシーケンシング技術によって決まる。

【0421】

本明細書では、「Hyb & Seqサイクル」は、特定の1つのプローブ、またはプローブの集団の表面にある各付着領域を検出するのに必要な全工程を意味する。例えば標的核酸の表面にある6つの位置を検出することのできるプローブでは、1つの「Hyb & Seqサイクル」は、少なくとも、プローブを標的核酸にハイブリダイズさせ、相補的核酸／レポータ

50

プローブをプローブのバーコードドメインの表面にある6つの位置それぞれの付着領域にハイブリダイズさせ、それら6つの位置それぞれに付随する検出可能な標識を検出することを含むことになる。

【0422】

「k量体プローブ (k-mer probe)」という用語は、本開示によるシーケンシング用プローブの同義語である。k量体読み出しは、Hyb & Seqのデータの基本単位である。1つのk量体読み出しは、1つのHyb & Seqサイクルごとに1つの標的分子から得られる。多数回のHyb & Seqサイクルを実施して単一の標的分子から十分な数の別々のk量体読み出しを生成させると、別々のk量体があいまいさなくアラインメントされて配列の連続的なストレッチを得ることが可能になる。

10

【0423】

別々の読み出しからの2つ以上の配列をアラインメントするとき、重複している部分を組み合わせて単一のコンセンサス配列を作り出すことができる。重複している部分が同じ塩基を有する位置（アラインメントの1つの列）では、それらの塩基がコンセンサスになる。さまざまな規則を利用することで、重複している配列の間に不一致が存在する位置に関するコンセンサスを生成させることができる。単純過半数という規則では、コンセンサスとして、列の中に最もよく出現する塩基を用いる。「マルチパスコンセンサス」は、単一の標的分子からの別々のプローブ読み出しすべてのアラインメントである。適用するプローブの集団／調査の全サイクル数に応じ、単一の標的分子内の各塩基位置への問い合わせを異なるレベルの冗長性または重複度で実施することができる。一般に、冗長性は塩基読み取りの信頼レベルを向上させる。

20

【0424】

「コンセンサス」は、別々の読み出しからの2つ以上のDNA配列をアラインメントする場合であり、重複している部分を組み合わせて単一のコンセンサス配列を生成させることができる。重複している部分が同じ塩基を有する位置（アラインメントの1つの列）では、それらの塩基がコンセンサスになる。さまざまな規則を利用することで、重複している配列の間に不一致が存在する位置に関するコンセンサスを生成させることができる。単純過半数という規則では、コンセンサスとして、列の中に最もよく出現する塩基を用いる。

【0425】

「生の精度 (Raw Accuracy)」は、システムが正しく塩基を同定する固有の能力の1つの指標である。生の精度はシーケンシング技術に依存する。「コンセンサスの精度」は、追加の読み出しと統計学の力を用いてシステムが正しく塩基を同定する能力の1つの指標である。「特異度」は、1回の試行ごとの全読み出しのうちで意図する標的にマッピングされる読み出しの割合を意味する。「一様性」は、標的領域相互間の配列カバレッジの変動性を意味し、大きな一様性は小さな変動性と相関している。この特徴は、標的とするすべての領域のうちで、20%以上の平均カバー深度でカバーされている標的とする領域の割合として一般に報告されている。ランダムなエラー（すなわちシーケンシング化学に固有のエラー）は、同じ標的核酸を「マルチパス」シーケンシングすることによって容易に補正でき、パスの数が十分だと、実質的に「完全なコンセンサス」または「エラーなし」のシーケンシングを実現することができる。

30

【0426】

本明細書に記載した方法を実現すること、および／または結果を記録することは、その方法を実現すること、および／またはその結果を記録することのできる任意の装置を用いて可能である。使用できる装置の非限定的な例に含まれるのは、あらゆるタイプのコンピュータを含む電子計算装置である。本明細書に記載した方法をコンピュータの中に実現および／または記録するとき、そのコンピュータにその方法の工程を実行させるのに使用できるコンピュータプログラムは、そのコンピュータプログラムを収容できる任意のコンピュータ可読媒体の中に収容することができる。使用可能なコンピュータ可読媒体の非限定的な例に含まれるのは、ディスク、CD-ROM、DVD、ROM、RAM、非一時的コンピュータ可読媒体、それ以外のメモリ、それ以外のコンピュータ記憶装置である。コンピュ

40

50

ータにその方法の工程を実行させる、および／または配列情報をアセンブルする、および／または結果を記録するのに使用できるコンピュータプログラムは、電子的ネットワーク（例えばインターネット、イントラネット、それ以外のネットワーク）に提供することもできる。

【0427】

「使い捨てシーケンシングカード（Consumable Sequencing Card）」を本分野で知られている蛍光イメージング装置に組み込むことができる。異なる多くの特徴を備えた任意の蛍光顕微鏡が、このシーケンシング読み出しを実施することができる。例えば広視野ランプ、レーザー、LED、マルチフォトン、共焦点照明、全反射照明のいずれかを励起および／または検出に用いることができる。蛍光顕微鏡の発光検出チャンネルには、フィルタまたは格子に基づくスペクトル分解（スペクトル分解された1つ以上の発光波長）ができるカメラ（1台または複数台）および／または光増倍管（1個または複数個）を取り付けることが可能である。標準的なコンピュータを使用して、使い捨てシーケンシングカードと、そのカードを通過する試薬の両方の制御することと、蛍光顕微鏡による検出を制御することができる。

【0428】

シーケンシングのデータは、任意の数の標準的な次世代シーケンシングアセンブラによって分析することができる（例えばWajidとSerpedin、「次世代シーケンサ用ゲノムアセンブラのためのアルゴリズムの一般的な特徴のレビュー」Genomics, proteomics & bioinformatics、第10巻(2)、58～73ページ、2012年を参照されたい）。顕微鏡の1つの回折制限領域の中で得られたシーケンシングデータを「局所的にアセンブルする」ことにより、1つの回折スポット内の複数の読み出しからコンセンサ配列を生成させる。次に、多数の回折スポットをアセンブルした読み出しをまとめてマッピングすることで、標的とする遺伝子セット全体を表わす連続配列、またはゲノム全体の新規なアセンブリを生成させる。

【0429】

本開示に係る追加の教示は、米国特許第8,148,512号、第7,473,767号、第7,919,237号、第7,941,279、第8,415,102号、第8,492,094号、第8,519,115、米国特許出願公開第2009/0220978号、第2009/0299640号、第2010/0015607号、第2010/0261026号、第2011/0086774号、第2011/0145176号、第2011/0201515号、第2011/0229888号、第2013/0004482号、第2013/0017971号、第2013/0178372号、第2013/0230851号、第2013/0337444号、第2013/0345161号、第2014/0005067号、第2014/0017688号、第2014/0037620号、第2014/0087959号、第2014/0154681号、第2014/0162251号、第2016/0194701号のうちの1つ以上に記載されている（それぞれの全体が参照によって本明細書に組み込まれている）。

【実施例】

【0430】

実施例1：Hyb & Seq化学を利用した一分子長の読み出し

【0431】

本開示のシーケンシング用プローブと、これらシーケンシング用プローブを用いたシーケンシング法を便宜上Hyb & Seqと呼ぶ。開示されているシーケンシング用プローブとシーケンシング法を記述するのに明細書全体を通じてこの用語を使用する。Hyb & Seqは、ライブラリなし、増幅なしの、一分子シーケンシング技術であり、天然標的の表面にある蛍光分子バーコードの核酸ハイブリダイゼーションサイクルを利用している。

【0432】

Hyb & Seqを用いた長い読み出しを長さ33キロベース（kb）の一分子DNA標的について示す。重要な工程は以下の通りである：（1）長いDNA分子がシーケンシング用フローセルの表面に捕捉され、流体力によって伸長される；および／または（2）複数の完全に一致したシーケンシング用プローブが長い一分子標的に沿ってハイブリダイズする；および／または（3）蛍光レポータがシーケンシング用プローブ内のバーコード領域に

ハイブリダイズし、結合したすべての配列が同定される；および／または（４）空間的に分離された蛍光データを利用して一分子標的内の配列の相対的位置が決定される。

【０４３３】

Hyb & Seqを利用した長い読み出しの重要な利点の非限定的な例に含まれるのは、分子の長さによって決まる読み出し長が化学によって制限されないこと；および／または簡単に制限されたサンプル調製の結果として断片化がより少なくなること；および／またはシーケンシング用プローブと関係する位置情報がアセンブル操作を助けること；および／またはバリエーションのフェージングによって長い範囲のハプロタイプにする能力である。

【０４３４】

Hyb & Seq化学の設計：シーケンシング用プローブは、一分子標的と塩基対を形成する標的結合ドメインと、標的結合ドメインの中に存在する6量体配列に対応する少なくとも3つの位置（ R_1 、 R_2 、 R_3 ）を有するバーコードドメインとを含んでいる。4096通りのシーケンシング用プローブによって任意の標的配列のシーケンシングが可能になる。レポータープローブ：3つのレポータープローブがバーコードドメインの上記位置に順番に結合する。それぞれのレポーター複合体は特定の1つのジヌクレオチドに対応する。ハイブリダイゼーションによって機能が発揮される。

【０４３５】

本開示の長い読み出しシーケンシング法と短い読み出しシーケンシング法は、核酸を標的とした捕捉に同じ簡単なプローブハイブリダイゼーション作業フローを利用することができる。複数のシーケンシング用プローブが1個の標的核酸に同時にハイブリダイズすることができ、光学的分解によって長い標的1つにつきいくつかのスポットを個別に識別することが可能になる。複数のシーケンシング用プローブを同時にハイブリダイズさせて記録することにより、1回の読み出し情報の内容が増える。長い範囲のハプロタイプは一分子分析に固有であり、コンピュータによる再構成ではなくて実際の物理的位置によってアセンブルすることができる。本開示のシーケンシング法を利用すると、数百キロベースまでの長いシーケンシング読み出しが可能である。

【０４３６】

シーケンシング用プローブは、予想される配列特異的位置かつ相対的な物理的距離で、伸長した標的（流体力で伸長した標的が好ましい）にハイブリダイズすることができる。本開示のシーケンシング法は、短い読み出し技術と比べて情報の内容が増えているため、各サイクルでより多くの塩基を読み出すことができる。本開示のシーケンシング法は、シーケンシングの読み出し結果の相対位置も記録する。これは、長い読み出しのアセンブル操作を助ける。本開示のシーケンシング法を利用する場合には、読み出し長＝コンセンサス配列の長さ＝捕捉された標的分子の長さである。

【０４３７】

本開示の方法を用いた1つの実験において、33キロベースのDNA断片を捕捉し、伸長させ、シーケンシング用プローブとレポータープローブにハイブリダイズさせ、検出した。本開示のシーケンシング法は、33キロベースまでのDNA断片と33キロベース超のDNA断片に適合性がある。読み出し長は、標的核酸の断片の初期長によってだけ制限され、酵素またはシーケンシング化学によっては制限されない。

【０４３８】

本開示の方法は、標的となるフェージングされた長い読み出しに関して追加の能力を有する。フェージングされた長い範囲のハプロタイプがデータには固有であり、バリエーションのフェージングに関して容易に同定される。長い標的分子全体のシーケンシングは不要である。というのも「ブロッカーオリゴ」を用いて配列のサイクリングを目的のシーケンシングウィンドウに制限できるからである。

【０４３９】

実施例1の結果は、本開示のシーケンシング法で長い読み出し長の一分子シーケンシングが可能であることを示している。この結果は特に、15キロベースと33キロベースの一本鎖DNA分子の捕捉と流体力による伸長がうまくいくこと；空間的に分離された蛍光

10

20

30

40

50

データが、長い単一の分子に沿った実際の相対位置に正確に対応していること；シーケンシングサイクルごとに10塩基超の複数の配列が同時に読み出されることを示している。

【0440】

実施例2：Assembly Algorithm：短いヌクレオチドバリエーションとInDelを分離する標的シーケンシングのため参照基準にガイドされてHyb & Seq読み出しを正確にアセンブルする

【0441】

Assembly Algorithmは、Hyb & Seqの独自の6量体読み出し（6量体スペクトル）をアセンブルする設計のオープンソースアルゴリズムである。Assembly Algorithmは、ShortStackまたはHexSembler（商標）という分析ソフトウェアとしても知られる。このアルゴリズムは、一分子をベースとしていて、画像化されたそれぞれの特徴から6量体読み出しを用いて標的を同定し、それら6量体読み出しをアセンブルしてコンセンサス配列にするための統計的アプローチであり、エラー補正を伴っている。

【0442】

Hyb & Seq化学とAssembly Algorithmを利用した一分子シーケンシングを以下のように実施した：Hyb & Seq化学を利用して各ハイブリダイゼーションサイクルの後に一分子標的の6量体読み出しを生成させた；ハイブリダイゼーションサイクルが多数回終了した後、それぞれの一分子標的の領域をカバーする6量体スペクトルが生成した；6量体スペクトルを標的核酸分子それぞれの参照配列とともに用いて各一分子標的のコンセンサス配列を導出する。

【0443】

Hyb & Seq技術をAssembly Algorithmとともに用いた標的シーケンシングの結果は以下のことを示している：6量体のスペクトルを用いた一分子標的の同定アルゴリズムは成功率が100%であった；および／または参照基準によってガイドされるアセンブリアルゴリズムにより一分子コンセンサスの精度が5回のカバレッジで99%超（QVが約32）になった；および／または合致する体細胞バリエーションの検出（ R^2 約90%）が、あらかじめ特徴付けられたgDNAサンプルを用いて実証された；および／またはすべての6量体とAssembly Algorithmを用いたインシリコ実験から、より大きな標的パネルで平均QVが90超であることが確認された。Assembly AlgorithmはHyb & Seqデータを正確にアセンブルすることができる。図18は、本開示のシーケンシング法を用いて取得してAssembly Algorithmを用いて分析したシーケンシング実験からの結果を示している。この実験では、シーケンシングされた標的核酸に含まれているのは、遺伝子BRAF（配列番号3）の断片、遺伝子EGFRex18（配列番号4）の断片、遺伝子KRAS（配列番号5）の断片、遺伝子PIK3CA（配列番号6）の断片、遺伝子EGFRex20（配列番号7）の断片、遺伝子NRAS（配列番号8）の断片である。図18は、塩基カバレッジとバリエーション読み取りの両方を示している。カバレッジのプロットは、FFPE（ホルマリン固定パラフィン包埋）gDNAにおける塩基のカバレッジを示している。結果は、利用可能なシーケンシング用プローブによって多彩な標的の大半の塩基がカバーされることを示している。エラープロットは、多彩な標的に関してFFPE gDNAサンプルの中の調べた位置におけるエラー率とカバレッジの関係を示している。結果は、8回のカバレッジでエラー率が1%未満であることを示している。頻度プロットは、シーケンシングされたHorizon gDNAサンプルに含まれるバリエーションの予想頻度と既知の頻度の間の相関を示している。表にはシーケンシングされたHorizon Genomic ReferenceのgDNAが掲載されており、バリエーション分子の一部が参照サンプルの既知の頻度と一致していることがわかる。

【0444】

実施例2の結果は、Assembly Algorithmが、本開示のシーケンシング法を利用して得られた6量体スペクトルの部分的アセンブル操作のための正確なアルゴリズムであることを示している。結果から特にわかるのは以下のことである：シミュレーションデータを用いると、標的の同定が100%の精度、塩基ごとの平均品質値が30超である；Hyb & Seq実験データにおいて塩基読み取りが5回のカバレッジで99%超の精度である；ゲノムDNA

からのバリエーションの検出が既知の値に合致する頻度である (R^2 約90%) ; 計算性能が優れていてアセンブルされる6量体の数に正比例し、パーソナルコンピュータで約15分間に69 k個の分子をアセンブルすることができる。

【0445】

実施例3: Hyb & Seq (商標) 技術を利用してFFPEサンプルから天然gDNAをライブラリなしで標的シーケンシングする - ハイブリダイゼーションに基づく一分子シーケンシングシステム

【0446】

本開示のシーケンシング法 (Hyb & Seq) を利用してFFPEサンプルに由来する天然のgDNAに関してがんパネルを標的としたシーケンシングを実施して以下のことを実証した: がん遺伝子標的を標的とする一分子シーケンシングにおける正確な塩基読み取り ; および / または既知のがん遺伝子の一ヌクレオチドバリエーション (SNV) と挿入 / 欠失 (InDel) の正確な検出 ; および / またはFFPEから抽出されたgDNA (DNA断片のサイズの中央値は200塩基) からのがん遺伝子標的の多重捕捉 ; および / または改良型プロトタイプ装置での末端から末端までの自動化シーケンシングの実行。

【0447】

Hyb & Seq化学と作業フローは以下のものであった: 目的のゲノム標的をシーケンシング用フローセルの表面に直接捕捉する ; 数百の6量体シーケンシング用プローブを含むプールをそのシーケンシング室に流入させる ; 3回のレポーター交換サイクルで蛍光レポータープローブをシーケンシング用プローブのバーコード領域に順番にハイブリダイズさせて6量体の塩基を同定する ; 塩基が同定されると、シーケンシング用プローブを洗い流す ; 標的領域が十分な深度で読み出されるまで、シーケンシング用プローブの新しいプールを用いてサイクルを繰り返す。

【0448】

Hyb & Seqの重要な利点: 単純かつ迅速なFFPE作業フロー - 60分以内にシーケンシングを開始するための臨床試料 ; および / または酵素または増幅がない / ライブラリがない構成 ; および / または全作業時間が15分間 ; および / または高い精度 - 低い化学エラー率 + 固有エラー補正 ; および / または長い読み出しと短い読み出しの両方 - 読み出し長は入力サンプルによって規定され、化学によっては制限されない。

【0449】

Hyb & Seq化学の設計は実施例1に記載した通りである。FFPE組織を処理するためのHyb & Seqサンプルの調製は3つの簡単な工程からなる: (1) 単一の試験管での脱パラフィンと溶解 ; (2) シリンジフィルタを用いた粒子の除去 ; (3) オプションのDNA断片化と標的捕捉。このプロセスでは、サンプルごとに1~3個の10ミクロンFFPE切片を使用する必要がある。このプロセス全体は60分以内に完了し、一般的な実験室の装置だけ、すなわちヒートブロック、ピペット、フィルタ、試薬だけを必要とする。

【0450】

図19は、FFPEサンプルに由来するがん遺伝子標的の捕捉とシーケンシングを多重化するための実験設計の模式図を示している。11個のがん遺伝子標的の一部 (配列番号3~13) をシーケンシングするため、合計425のシーケンシング用プローブを設計して構成した。各遺伝子標的に関する既知のバリエーションの遺伝子座が多数のシーケンシング用プローブでカバーされた (完全な一致 + 単一のミスマッチ)。塩基カバレッジと塩基ごとの精度をこれらの領域で測定した。あらかじめ特徴付けた参照サンプルを用いてバリエーション検出の精度を取得した。図19の最も上の図は、シーケンシング用プローブ (青色) が、既知のバリエーションの位置 (赤色) を取り囲む標的配列 (灰色) にアラインメントされることを示している。各バリエーションの位置 (赤色) について、それぞれの (A、G、C、T) 塩基バリエーションを有する4つのプローブ配列が含まれるようにした。シーケンシング中に1つの標的DNA分子を800回のバーコード交換サイクルにわたって追跡すると、多数の6量体読み出しが得られ、それらを実施例2に記載したようにしてAssembly Algorithmによって再アセンブルした。

【0451】

図18はシーケンシングの結果を示しており、そこには、各標的の平均カバレッジ、一塩基エラー率、バリエーションの頻度の観測値と予想値が含まれている。実施例3の結果は、Hyb & Seqシーケンシングを利用すると、FFPEと参照gDNAサンプルの中の11の標的領域の多重化シーケンシングを実施して一ヌクレオチドバリエーションを小さなエラー率で検出できることを示している。

【0452】

実施例4：Hyb & Seq（商標）化学を利用してcDNAへの変換なしに一分子RNAシーケンシングを直接実施する

【0453】

Hyb & Seq化学を利用した直接的な一分子RNAシーケンシングを以下のように実施した：天然RNA分子をcDNAへの変換なしに直接捕捉し、シーケンシング用フローセルの表面に固定化した；数百の6量体シーケンシング用プローブを含有するプールをそのシーケンシング用フローセルに流入させた；完全に一致するシーケンシング用プローブを一分子RNA標的の表面にランダムにハイブリダイズさせた；蛍光レポータープローブをシーケンシング用プローブのバーコード領域に順番にハイブリダイズさせて6量体の塩基を同定した；塩基を同定した後、シーケンシング用プローブを洗い流した；標的が十分な深度で読み出されるまで、サイクルを繰り返した。

【0454】

重要な結果：標的とされる一分子RNAをシーケンシングすると、DNAと似たカバレッジプロファイルが示された；および／または200回超のHyb & SeqサイクルにわたってRNA分子がフローセルの表面に安定に維持された；および／または単一のFFPEスライスからmRNAとゲノムDNAが同時に捕捉されて定量された；および／またはわずか10 ngの全RNAを用いて8個の転写産物が多重捕捉されて定量された。

【0455】

Hyb & Seq化学の設計は実施例1に記載した通りである。図20の最上部の図は、直接的なRNAシーケンシングに関する実験工程を、cDNA変換を利用して実施する従来のRNAシーケンシングに関する工程と比較した模式図を示している。図20の中央の図と最下部の図は、RNA分子と本開示のシーケンシング法の適合性を調べる実験からの結果を示している。この実験では、4個の標的RNA分子をシーケンシングした（配列番号14～17）。結果は、少なくとも200回のシーケンシングサイクルにわたってRNA分子を捕捉して検出できることを示している。これは、本開示のシーケンシング法がRNA分子に適合していることを証明している。

【0456】

図21は、本開示のシーケンシング法を利用した直接的な一分子RNAシーケンシングを確認する実験からの結果を示している。NRASex2（配列番号18）の断片をコードしている天然RNA分子をcDNAへの変換なしで直接捕捉し、シーケンシング用フローセルの表面に固定化し、本開示の方法を利用してシーケンシングした。この実験は、RNAの代わりに捕捉されたDNA分子も用いて繰り返した。図21は、DNAとRNAのシーケンシングのカバレッジが同等であったことを示している。これは、本開示のシーケンシング法を利用してRNAをcDNAへの変換なしに直接シーケンシングできることを証明している。

【0457】

いくつかの態様では、本開示により、FFPEサンプルからRNAとDNAをまとめて捕捉することが提供される。サンプルは、実施例3に記載されているのと同じFFPE作業フローを利用して調製される。同じ捕捉プロトコルを利用するが、RNA特異的捕捉用プローブとDNA特異的捕捉用プローブが用いられる。DNA分子とRNA分子は同じフローセルの中で同じシーケンシング用プローブを用いて同時にシーケンシングされる。いくつかの態様では、RNAとDNAを同時に捕捉するのに特異的RNA捕捉用プローブと特異的DNA捕捉用プローブが必要とされる。

【0458】

全RNAをさまざまな入力量（0 ng、1 ng、10 ng、100 ng、1000 ng）にしたときのヒトユニバーサル参照RNA（Human Universal Reference RNA）に関して中程度から高度に発現した8個の転写産物の多重捕捉も実施した。多重捕捉されたRNA分子をフローセルの表面に固定化し、特異的シーケンシング用プローブとレポータプローブを固定化されたそのRNA分子にハイブリダイズさせて定量した。それぞれの特異的RNA標的のカウント数の定量結果は、入力RNAの総量が増えるにつれて増加することを示していた。

【0459】

実施例4の結果は、一分子RNAシーケンシングがHyb & Seq化学で実現されることを示している。結果から特にわかるのは、（1）cDNAへの変換なしの直接的なRNAシーケンシングが可能であること；（2）Hyb & Seqサイクリングプロセスを通じてRNA分子が安定であること；（3）1つのHyb & Seq作業フローでRNA分子とDNA分子の両方を捕捉してシーケンシングできること；（4）10 ngという少ない全RNA入力でmRNAパネルの標的捕捉を実施できることである。

【0460】

実施例5：Hyb & Seqシーケンシングプラットフォームから生成するハイスループットな分子レベルでの短い読み出しのための統合バイオインフォマティクスアルゴリズム

【0461】

Assembly Algorithmソフトウェアを設計して標準的なシーケンシングに基づくバイオインフォマティクス分析のタスク（アラインメント、エラー補正、変異読み取り、読み出しのアセンブル操作など）を実行した。Assembly Algorithmソフトウェアパイプラインの工程に含めることができるのは、6量体のアラインメントとカバレッジの推定；および／または変異した配列の同定；および／またはグラフデータ構造の構成；および／または分子レベルでの配列再構成とエラー補正である。

【0462】

すべてのアルゴリズムが、一分子から得られた情報の範囲内で厳密に実行されたことで、最終的な変異コール（mutation call）結果がサンプルの変異頻度に左右されないことが保証される。6量体をパネル結合位置に従って異なる分子にグループ化した。分子を標的に割り当てるため、6量体を分子ごとに異なるすべての標的領域にアラインメントさせ、最もよく一致した遺伝子標的を選択した。

【0463】

1つの統計的指標を測定して分子同定の品質を評価した。N個の異なる標的領域に対するアラインメントから、各標的についてのカバレッジ値の和N個の分布が生成する。カバレッジ値の和の一致スコアが最上位のものを正しい一致として選択した。異なるN個の標的すべてのスコア分布に対して選択された最上位の一致スコアのzスコア統計を測定した。信頼度の低い分子属性（zスコアが2.5σ未満）はフィルタで除外した。

【0464】

Assembly Algorithmの重要な利点に含まれるのは、階層的ハッシュインデックスの設計を実現することにより、生じる可能性のある配列のあいまいさを正確に取り扱えること；および／または進歩したアルゴリズム設計構造が優先順位付けによってマッピングの品質を保証し、変異の過大評価を阻止していることである。

【0465】

それに加え、変異グラフデータ構造によりさまざまなタイプの変異（置換、挿入、欠失）の計算モデルを作ることが可能になり、配列再構成とバリエーション読み取りのための出力が生成する（置換バリエーションは、元の配列と同じ長さのグラフにおいて追加ノードとして表わされ；挿入は、任意の長さの接続されたノードを追加することによってモデル化することができ；欠失は、グラフの中の人工ノードに空の塩基対ストリングを追加することとしてモデル化される）；および／またはブラインド変異検索（すなわち変異許容配列アラインメントの検索）において参照配列のすべての位置からのハミング距離が測定され、検索された変異を表わすグラフに新たなノードが追加される；および／または変異した6量体に関するカバレッジの推定が、階層的ハッシュ表を用いて実行される。

【0466】

構成されたグラフデータ構造により、分子レベルでの配列再構成と装置エラーの補正が可能になる。構成されたグラフにおいて、ダイナミックプログラミングアルゴリズムを適用し、スコアが規格化された塩基カバレッジとして定義されるスコアリング経路のうちで最良のものを見いだす。グラフの最良のスコアリング経路が、分子レベルでの配列再構成を表わしていた。正しい変異配列が含まれていた一方で、6量体における装置エラーは廃棄された。

【0467】

シミュレーションしたデータセットから、このソフトウェアが、分子レベルでの非常に正確な配列再構成と変異コーリング (mutation calling) 結果を供給できることを確認した。結果は、10個のランダムな変異についての変異コーリング精度を示している。装置エラーが中程度のデータセットでは、精度は平均で99.39% (標的を明確にした検索) と98.02% (ブラインド検索) を示した。装置エラーが多いシミュレーションのもとでは、性能は平均で97.19% (標的を明確にした検索) と93.53% (ブラインド検索) を示した。分子レベルでの塩基カバレッジの閾値を2倍にしたとき、結果は99.5% (2回のカバレッジ) と99.9% (3回のカバレッジ) に改善された。

【0468】

Assembly Algorithmソフトウェアは、広い範囲のさまざまな変異を処理することができる。変異には、1 bpと15 bpの間で選択された挿入と欠失が含まれ、変異コーリング精度は94.4% (1回のカバレッジ)、97.2% (2回のカバレッジ)、98.5% (3回のカバレッジ) である。

【0469】

実施例6: Hyb & SeqのためにFFPE組織を処理することを目的としたサンプル調製

【0470】

ホルマリン固定パラフィン包埋 (FFPE) 組織は、伝統的なシーケンシングプラットフォームにとって挑戦的なタイプのサンプル入力である。Hyb & Seqのサンプルを調製する方法によって下流シーケンシングのためのFFPE組織入力がうまく処理される。最初に、シーケンシングする核酸を1工程プロセスでホルマリン固定パラフィン包埋 (FFPE) 組織から抽出する。厚み10 μm の1つ以上のFFPE切片を核酸抽出緩衝水溶液の中で加熱してパラフィンワックスを同時に融解させ、組織を分解し、細胞から核酸を放出させる。適切な抽出緩衝液が本分野で知られており、その中に典型例に含まれているのは、プロテイナーゼ、洗浄剤 (Triton-100など)、キレート剤 (EDTAなど)、アンモニウムイオンである。FFPE切片と抽出緩衝液を56°Cで30分間インキュベートして組織からパラフィンを分離し、プロテイナーゼKに組織構造を消化させることで包埋された細胞を洗浄剤に対して露出させ、細胞の溶解が可能になるようにする。溶液を8分間の間隔で3回上下反転させ、組織の脱パラフィンと消化のプロセスの間に試薬が混ざりやすくする。この工程の後、溶液を98°Cまで加熱してホルムアルデヒド架橋の逆転を促進することで、核酸がさらに抽出されやすくする。

【0471】

核酸がFFPE組織から抽出されると、孔のサイズが2.7 μm のガラスファイバーフィルタ (Whatman社) を用いて溶液を濾過して組織デブリと凍結したパラフィンを除去する。得られた溶液は、核酸を含有する均一で半透明な溶液である。その核酸は、ホルマリン固定プロセスと保管条件が原因で高度に断片化されている。さらに断片化する必要がある場合には、Covaris社の集中照射式超音波処理装置を用いてDNAを物理的に剪断することができる。緩衝液の条件が理由で、核酸の剪断には延長された超音波処理が必要とされる。ピーク投入電力50 W、デューティ因子20%、200サイクル/バーストという標準的な設定を利用した超音波処理を600秒間実施して捕捉される標的をできるだけ増加させる。断片の長さをより短くするには、4°Cにて21,000 gで15分間にわたって遠心分離することにより、乳化したパラフィンを濾過された溶液から沈殿させることができる。こうすることで、DNAを剪断して約225 bpにすることができる。

【0472】

次に、迅速なハイブリダイゼーションプロセスの間に捕捉用プローブのペアを標的核酸分子に結合させることにより、標的の捕捉を実行する。5'捕捉用プローブは3'ビオチン部分を含んでいるため、標的堆積プロセスの間に、標的が、ストレプトアビジンで被覆したフローセルの表面に結合することができる。3'捕捉用プローブは、精製プロセスの間にビーズに結合できるようにする5'タグ配列（G配列）を含んでいる。反応速度は、反応速度を最大にするため低いナノモル範囲で添加される捕捉用プローブの濃度に支配される。捕捉用プローブは、目的の領域に隣接するようにして標的にハイブリダイズしてシーケンシングウインドウを生成させる。それぞれのDNA標的について、捕捉用プローブのセットは、標的のアンチセンス鎖にハイブリダイズして再アニーリングを阻止するため、シーケンシングウインドウと配列が同じ構成のオリゴも含んでいる。捕捉用プローブを含む溶液を3分間にわたって98℃に加熱してゲノムDNAを変性させた後、65℃で15分間インキュベートする。400 mM～600 mMの範囲の濃度のNaClをこのハイブリダイゼーション反応で用いる。実験で確認されている100個超の標的からなるパネルを表3に掲載してある。この表には、遺伝子と、標的となるDNA領域のエキソンが詳しく記載されている。

【0473】

10

20

30

40

50

【表 5】

表3：遺伝子と、標的となるDNA領域のエキソン

遺伝子	標的	遺伝子	標的	遺伝子	標的
ABL1	ABL1_ex4	EZH2	EZH2_ex8	NOTCH1	NOTCH1_ex26
	ABL1_ex6		NRAS	NRAS_ex2	
	ABL1_ex7			NRAS_ex3	
AKT1	AKT1_ex6	FBXW7		FBXW7_ex2	NRAS_ex3
ALK	ALK_ex26		FBXW7_ex5	NRAS_ex4	
APC	APC_ex5		FBXW7	FBXW7_ex7	PDGFRA
	APC_ex16	FBXW7_ex8		PDGFRA_ex4	
	APC_ex17	FBXW7_ex9		PDGFRA_ex7	
	APC_ex17	FBXW7_ex10		PDGFRA_ex10	
	APC_ex17	FGFR1	FGFR1_ex6	PDGFRA_ex11	
	APC_ex17	FGFR2	FGFR2_ex7	PDGFRA_ex14	
	APC_ex17	FLT3	FLT3_ex11	PDGFRA_ex15	
ATM_ex8	FLT3_ex12		PDGFRA_ex16		
ATM_ex9	FLT3_ex21		PDGFRA_ex18		
ATM	ATM_ex11	GNAQ	GNAQ_ex5	PIK3CA	PDGFRA_ex23
	ATM_ex26	IDH1	IDH1_ex4		PIK3CA_ex2
	ATM_ex34		IDH1_ex10		PIK3CA_ex3
	ATM_ex39	IDH2	IDH2_ex4		PIK3CA_ex7
	ATM_ex49	JAK2	JAK2_ex3		PIK3CA_ex10
	ATM_ex49		JAK2_ex7		PIK3CA_ex14
	ATM_ex55		JAK2_ex14		PIK3CA_ex21
	ATM_ex59		JAK2_ex20		PIK3CA_ex21
	BRAF		BRAF_ex8		KDR
BRAF_ex11		KDR_ex7	PTEN_ex7		
BRAF_ex13		KDR_ex9	PTENP1	PTEN_ex8	
BRAF_ex15		KDR_ex11		PTENP1_ex1	
CDH1	CDH1_ex9	KIT	KDR_ex27	RB1	RB1_ex10
CSF1R	CSF1R_ex3		KDR_ex30		RB1_ex17
	CSF1R_ex22		KIT_ex5		RB1_ex17
CTNNB1	CTNNB1_ex3	KIT	KIT_ex9	RET	RB1_ex20
	CTNNB1_ex6		KIT_ex14		RB1_ex22
	CTNNB1_ex16		KIT_ex17		RET_ex12
EGFR	EGFR_ex3	KRAS	KIT_ex18	SMAD4	RET_ex15
	EGFR_ex10		KRAS_ex2		SMAD4_ex3
	EGFR_ex15		KRAS_ex3		SMAD4_ex8
	EGFR_ex18		KRAS_ex4		SMAD4_ex9
	EGFR_ex20	MEK	MEK_ex3		SMAD4_ex10
ERBB2	ERBB2_ex7	MET	MET_ex2	SMARCB1	SMARCB1_ex5
ERBB4	ERBB4_ex4		MET_ex3	TP53	TP53_ex4
	ERBB4_ex5		MET_ex11		TP53_ex6
	ERBB4_ex7		MET_ex14		
	ERBB4_ex8		MET_ex16		
	ERBB4_ex23		MET_ex19		
	ERBB4_ex25		MLH1		MLH1_ex12
		MLH1_ex16			

【0474】

捕捉用プローブを標的となるDNA領域に結合させた後、ゲノムDNAの残部から精製すると、標的が豊富になった溶液が生じる。3'捕捉用プローブの結合配列に対するアンチセンスオリゴ（アンチG配列）で被覆したビーズを捕捉反応混合物とともに室温で15分間インキュベートする。この結合工程の後、ビーズを0.1×SSPEで3回洗浄して非標的DNAとピオチン含有5'捕捉用プローブを除去する。洗浄の後、ビーズを14 μlの0.1×SSPEに再懸濁させ、次いで10分間にわたって45℃に加熱して精製されたDNA標的をビーズから溶離させる。溶離後、1 μlの5 M NaClを添加して捕捉用プローブがDNA標的に結合したままになるようにする。

10

20

30

40

50

【0475】

サンプル調製プロセスの最終工程は、DNA標的をフローセルの表面に堆積させることであり、その場所で本明細書に開示されているようにして本開示のプロープを用いてDNA標的を分析することができる。シリンジポンプを用いて標的をフローセルの流体チャンネルに装填する速度を制御する。そのためすべての標的に、チャンネルの高さを横断して拡散してストレプトアビジンの表面に結合する時間がある。この装填法によって標的の密度勾配が生じ、単位面積当たりの分子数は流体チャンネルの入口において最大になり、出口に向かう流体流の方向でチャンネルの長さに沿って減少していく。0.35 μl /秒という流速だと、幅が1.6 mmで高さが40 μm のチャンネルでは約10 mmのチャンネルの長さの中に大量の捕捉が実現する。ビオチニル化された5'捕捉用プロープによって標的が表面に結合すると、3'捕捉用プロープの結合配列の逆相補体であるビオチニル化されたオリゴ（Gフック）の溶液が注入されて標的の自由端が垂れ下がり、架橋した構造を生み出す。その中央部の中のssDNA領域が、目的のシーケンシングウインドウである。次に、G配列オリゴの溶液を添加して表面にある過剰なGフックにハイブリダイズさせ、その表面上のssDNAの量を減らす。図8は、本開示の2捕捉用プロープシステムを用いた1個の標的核酸の捕捉を示している。

10

【0476】

実施例7：Hyb & Seqのための多色レポート画像処理

【0477】

画像処理パイプラインに含まれる工程は、バックグラウンドを差し引く工程、登録する工程、特徴を検出する工程、分類の工程である。バックグラウンドを差し引く工程では、所与の任意のチャンネルの平均バックグラウンドは、ショット雑音と露出の関数である。われわれのシステムでは、青色チャンネルが最大レベルのバックグラウンドを有し、それよりも大きな変動がカップルしている。半径が7画素の円形構造要素を有する単純なシルクハット形フィルタを適用して局在化したバックグラウンドを差し引く。登録に関しては、多色かつ多サイクルの特徴分析のため、目的の特徴を完全にアラインメントすることが必須である。このシステムは、2つの形態の登録を必要とする。第1の形態に関しては、単一取得スタックの中のあらゆる画像チャンネルに局所アフィン変換を適用する。この変換は光学系の関数であるため、所与の装置に適合している。この関数をすべての試行についてあらかじめ計算しておき、取得したすべての画像に適用する。第2の形態に関しては、規格化された交差相関を利用して固定シフトの形態の全体変換を計算し、試行中の機械的ガントリのドリフトを捕捉する。次の工程は特徴の検出である。

20

30

【0478】

すべての画像が登録されると、一致したフィルタを用い、LoG (Laplacian of Gaussian) フィルタを通じて特徴を検出する。フィルタは、スポットの応答の増強と一致させるため、（特徴の回折限界に一致した）固定されたカーネルサイズと（対応するチャンネルの波長に一致した）変化する標準偏差で適用する。極大を用いてレポートの潜在的な位置を同定する。同定されたそれぞれの特徴に関連する強度値を回収して分類する。最終工程は分類である。ガウス型ナイーブベイズモデルを用いて多色レポートの強度を分類する。このモデルは、レポートの強度が互いに独立であって正規分布に従うことを仮定している。次にこのモデルで、最大事後 (maximum a posteriori) 規則、すなわちMAP規則を用い、（すべてのチャンネルにおける強度によって指定される）特定の特徴

40

【数1】

 $\hat{y}$

が所定のクラス (C_k) に属する確率を計算する。

【0479】

【数2】

50

$$\hat{y} = \operatorname{argmax}_{\{k \in \{1, \dots, K\}\}} p(C_k) \prod_{i=1}^n p(x_i | C_k)$$

【0480】

2つの染料（青と赤）を用いたコーディングスキームの一例では、2色コーディングシナリオに可能な6つのクラス（バックグラウンドを含む）が存在する。実現したシステムでは、4色を選択することで14のクラスが可能になっている。単一の半染料と全染料の分布の間にはいくらか重複があることに注意されたい。その帰結として、これらクラスの間での分類はより大きなエラー率を示し、「xG」と「GG」の間で誤って分類される割合の最大値は11.8%である。10クラスモデルで誤って分類される割合は0.2%未満である。各レポートは最大で8つのクラスを必要とするため、分類エラーが最小のものを選択するのは簡単である。検出された色コードを、参照表に基づき、同定された塩基対に翻訳する。本明細書に開示されているように本開示のプロープを用いて複数のサイクルで1つの特徴を追跡する。

10

【0481】

実施例8：捕捉用プロープを用いた標的核酸の精製と堆積

【0482】

標的核酸分子を捕捉するのに2捕捉用プロープシステムを用い、極めて特殊な富化を実現する。捕捉用プロープを、標的とする目的の領域に隣接した位置で標的核酸に結合するように設計し、「シーケンシングウインドウ」を生み出す。5'捕捉用プロープ（CapBと呼ぶ）は3'ピオチン部分を含有している。3'捕捉用プロープ（CapAと呼ぶ）は、5'親和性タグ配列（G配列と呼ぶ）を含有している。平均して捕捉用プロープは長さが約40個のヌクレオチドであり、Tmと配列の文脈に基づいて設計される。シーケンシングウインドウは長さが約70個のヌクレオチドであり、容易に調節される。図8は、2捕捉用プロープシステムの模式図を示している。

20

【0483】

CapB上のピオチン部分が、標的核酸を、ストレプトアビジンで被覆したシーケンシング用フローセルの表面に接続している。CapB上の親和性タグにより、精製中に標的核酸分子を磁性ビーズに可逆的に結合させることが可能になる。CapAとCapBを用いることで、高度にストリンジェントな標的富化が可能になる。というのも両方のプロープが1個の標的核酸分子に結合したままになり、その標的が磁性ビーズ精製プロセスと表面堆積プロセスの両方を生き延びるからである。多重捕捉は、一度に100個までの標的で実証されている。短時間での効率的な捕捉を実現するため、捕捉用プロープは1 nM～10 nMの範囲の濃度で添加する。

30

【0484】

実験的な試験において、約10個の標的核酸分子からなるパネルを、Gビーズと2プロープ捕捉システムを用いて精製した。CapAプロープとCapBプロープを標的核酸に最初にハイブリダイズさせた。次に、結合したCapAプロープのG配列部分がGビーズ上のGフックにハイブリダイズすることで、標的核酸分子がGビーズに連結される。0.1×SSPEを用いた一連のストリンジェントな洗浄を実施して標的とされないDNAと結合しなかったCapBを除去した。標的核酸分子をGビーズから放出させるため、低い塩濃度で45℃にて溶離させてG配列を変性させる一方で、CapAとCapBは標的核酸にハイブリダイズしたままになるようにした。

40

【0485】

試験から、約100個の標的核酸分子からなるパネルを精製するとき、非特異的なシグナル／バックグラウンドのシグナルが有意に増加することがわかる。バックグラウンドのこの増加は、いくつかの因子に起因する可能性がある。すなわち、（1）CapAプロープ種とCapBプロープ種との相互作用が増加し、それが精製を通じて有し越される増加した量の

50

遊離したCapBプローブにつながることで、(2) CapBプローブとGフックまたはGビーズの間の相互作用が増加し、それが望まない標的核酸の精製につながることである。さらに、パネルのサイズが大きくなるにつれて、CapB種とCapA種とシーケンシング用プローブの間で可能な相互作用が指数関数的に大きくなる。これら相互作用は、標的を密に堆積させる能力に干渉する可能性があるため、シーケンシングの読み出しが無駄になる。

【0486】

遊離したプローブ種と望まない標的核酸分子の精製に起因する非特異的なシグナルとバックグラウンドシグナルを減らすため、精製手続きに対する改変をいくつか実行することができる。第1に、標的核酸分子をGビーズに結合させるときに用いる緩衝液の中にホルムアミドを30%v/vの濃度で含めると、おそらく、遊離した捕捉用プローブとGフックの不完全なハイブリダイゼーションに干渉することを通じ、(標的分子を欠く対照におけるカウント数によって測定したとき)バックグラウンドのカウント数が1/2に減少する。そのため過剰なプローブを洗い流すことができる。第2に、Gビーズ上のGフックの中に4個のイソdG塩基を含め(イソ-Gフック(iso-G-hooks))、CapA G配列の中に相補的イソdC塩基を含めると、(標的分子を欠く対照におけるカウント数によって測定したとき)バックグラウンドのカウント数が1/3に減少する。イソdCとイソdGは天然のdC塩基とdG塩基の異性体バリエーションである。イソ塩基は他のイソ塩基と塩基対を形成するが天然塩基とは塩基対を形成しないため、捕捉用プローブとイソGフックの間の不完全なハイブリダイゼーションだけが、G配列の非イソ塩基とイソGフックの間に存在することができる。これらの不完全な相互作用は、ストリンジェントな洗浄の間により簡単に破壊される。最後に、Ampure(登録商標)XP(Agencourt Biosciences Company社)ビーズを用いてイソGビーズをその後精製することにより、(標的分子を欠く対照におけるカウント数によって測定したとき)バックグラウンドのカウント数が少なくとも1/20に減少する。Ampure(登録商標)XPビーズを用いた精製の間、ポリエチレングリコール(PEG)とNaClの溶液中でDNAサンプルをカルボキシル化された磁性ビーズの懸濁液と混合する。PEGとNaClの濃度を滴定し、分子量の閾値を超えた分子だけが沈殿してビーズに結合するようにできる。捕捉用プローブにハイブリダイズしたHyb & Seq標的は81 kDaのオーダーであるのに対し、遊離したプローブは17 kDa以下のオーダーである。Ampure(登録商標)XPビーズの懸濁液をイソGビーズ溶離液と体積比1.8:1で混合することにより、ハイブリダイズした標的はビーズに結合するため、遊離したプローブの有意な部分を最終的な溶離の前に洗い流すことができる。

【0487】

したがってモデル精製作業フローは以下の工程からなる：(1) 5×SSPE/30%ホルムアミドの中で捕捉用プローブ-標的核酸複合体をイソGビーズにハイブリダイズさせる；(2) 0.1×SSPEを用いてイソGビーズを洗浄する；(3) 0.1×SSPEの中で45℃にて捕捉用プローブ-標的核酸複合体を溶離させる；(4) イソGビーズ溶離液を1.8倍の体積のAmpure(登録商標)XPビーズに結合させる；(5) 75%エタノールを用いてAmpure(登録商標)XPビーズを洗浄する；(6) 0.1×SSPEの中で捕捉用プローブ-標的核酸複合体を溶離させると標的が7.5 µlの中に溶離するため、その後0.5 µlの5 M NaClを添加する。

【0488】

精製の後、輸液シリンジポンプを用いて捕捉用プローブ-標的核酸複合体をシーケンシング表面に堆積させ、精製された標的をフローセルにゆっくりと注入する。堆積勾配を求めるため、フローセルをチャンネルの長さに沿ったさまざまな位置で撮影する。チャンネルの高さが20 µmだと、サンプルを0.167 µl/分の流速で装填すると標的が濃縮され、全標的の80%がチャンネルの長さに沿って5.1 mm以内に結合する。これは、FOVが0.0357 mm²、フローセルのチャンネル幅が1.7 mmのGen2イメージャでは約240 FOVであることを表わしている。勾配は、堆積中の流速を調節することによって変えることができる。

【0489】

上に記載した手続きを利用して、約300塩基対のサイズに切断されたゲノムDNAを有する100重の標的核酸パネルの精製と堆積を調べた。DNA入力範囲を25 ng~500 ngに

10

20

30

40

50

して一連の実験をトリプリケートで実施した。フローセル上の標的の合計数を堆積勾配のイメージングによって外挿し、平均カウント数を取得した。捕捉効率は6.6%であり、DNA質量入力範囲で一貫性があった。

【0490】

実施例9：シーケンシング用プローブの設計と特徴

【0491】

シーケンシング用プローブは標的結合ドメインを介して標的核酸分子にハイブリダイズする。この実施例では、標的結合ドメインは長さ8個のヌクレオチドであり、(N)塩基((N)塩基として、ユニバーサル塩基／縮重塩基または正規塩基が可能である)が隣接したロックド核酸(LNA)6量体を含む(N₁-B₁-B₂-B₃-B₄-B₅-B₆-N₂、ただしB₁~B₆はLNAであり、N₁とN₂は、(6量体)配列B₁-B₂-B₃-B₄-B₅-B₆とは独立なユニバーサル塩基／縮重塩基または正規塩基である)。4,096通りのシーケンシング用プローブの完全なセットは可能なすべての6量体をコードしているため、任意の標的核酸のシーケンシングが可能である。それぞれのシーケンシング用プローブは、標的結合ドメイン内に存在する6量体配列をコードするバーコードドメインも含んでいる。それぞれのバーコードドメインは、3つの位置(R₁、R₂、R₃)を含む。バーコードドメイン内の各位置は、標的結合ドメインの6量体の中の特定のジヌクレオチドに対応し、特定の標識されたレポータ複合体に結合することのできる独自の配列を含む。シーケンシング用プローブの全体模式図を図1に示してある。バーコードドメイン内の各位置は、4種類の蛍光染料、すなわち青(B)、緑(G)、黄(Y)、赤(R)を用いて作り出される8通りの「色組み合わせ」をコードしている。シーケンシングの各サイクル中にレポータ複合体がバーコードドメイン内の3つの位置のうちの1つに結合して、標的結合ドメインの6量体の中の対応するジヌクレオチドの属性を示す。3回の連続したシーケンシングサイクルの間に3つの「色組み合わせ」がバーコードドメイン内のそれぞれの位置について1つ記録されるため、標的結合ドメインの6量体全体の同定が可能になる。4,096通りのシーケンシング用プローブは8つのプールに分割され、それぞれは512の可能なバーコードのうちの1つと関係している。

【0492】

実施例10：レポータ複合体の設計、精製、結合の条件

【0493】

この実施例では、各レポータ複合体は37個のDNAオリゴマーからなる分岐構造であり、合計30個の蛍光染料(色組み合わせのそれぞれの色の染料が15個)を保持する設計にされている。レポータ複合体を構成する37個のDNAオリゴマーはそのサイズによって分類することができる。一次核酸と呼ばれる最大のオリゴマーは、長さが12個または14個のヌクレオチドである相補的核酸に共有結合する。一次核酸は、長さが96個のヌクレオチドである。相補的一次核酸は、シーケンシング用プローブのバーコードドメイン上の位置R₁、R₂、R₃のいずれかに結合する。次に大きいDNAオリゴマーは長さが89個のヌクレオチドであり、二次核酸と呼ばれる。レポータ複合体1つにつき6個の二次核酸が存在していて、色組み合わせのそれぞれの色につき3個の二次核酸である。それぞれの二次核酸はヌクレオチド14個の長さの配列を含んでいて、その配列により二次核酸が一次核酸にハイブリダイズすることが可能になる。最小のDNAオリゴマーは長さが15個のヌクレオチドであり、三次核酸と呼ばれる。2色プローブ1つにつき30個の三次核酸が存在していて、1色につき15個の三次核酸である。5個の三次核酸がそれぞれの二次核酸に結合する。37個のDNAオリゴマーからなる分岐構造の模式図を図4に示す。

【0494】

三次核酸は、蛍光染料の形態になった検出可能な標識を含んでいる。4種類の蛍光染料、すなわち青(B)、緑(G)、黄(Y)、赤(R)が存在している。1つのレポータ複合体の中で染料を組み合わせると、2色の組み合わせが10通り可能になる(BB、BG、BR、BY、GG、GR、GY、RR、YR、YY)。異なる蛍光染料の間の色交換または交差ハイブリダイゼーションを阻止するため、特定の蛍光染料に対応するそれぞれの二次核酸と三次核酸は

独自の配列を含有している。例えばAlexa 488蛍光体すなわち青色で標識したそれぞれの三次核酸は、青色二次核酸に対してだけ相補的な配列を含んでいる。青色二次核酸はさらに、青を含む色組み合わせに対応する一次核酸分子に対してだけ相補的な異なる配列を有する。

【0495】

それぞれの相補的核酸は、シークエンシング用プローブのバーコードドメインの位置 R_1 、 R_2 、 R_3 の間で互いに異なる配列を含有している。したがって同じバーコードドメインの位置 R_1 と R_2 が同じジヌクレオチドをコードしている場合でさえ、位置 R_1 のそのジヌクレオチドを同定する相補的核酸分子が位置 R_2 に結合することはない。同様に、位置 R_2 のそのジヌクレオチドを同定する相補的核酸分子が位置 R_1 に結合することはない。相補的核酸は、競合的トーホールド交換 (competitive toe-hold exchange) (長さが12個のヌクレオチドである相補的核酸の場合) を利用するか、UV切断 (長さが14個のヌクレオチドである相補的核酸の場合) を利用してシークエンシング用プローブから効率的に外れることができるように設計される。

【0496】

レポーター複合体の調製は、(1) 三次核酸を二次核酸にハイブリダイズさせた後、(2) 三次核酸+二次核酸を一次核酸にハイブリダイズさせるという2つの連続したハイブリダイゼーション工程でなされる。4.2×SSPE緩衝液の中で100 μ Mの二次核酸と600 μ Mの三次核酸を室温で30分間にわたって組み合わせることにより、二次核酸に対する三次核酸の反応物を別々に4つ調製する。次に、4.8×SSPEの中で2 μ Mの一次核酸、7.2 μ Mの二次核酸+染料#1三次核酸、7.2 μ Mの二次核酸+染料#2三次核酸を用いて24のレポータープローブを別々に調製する。これらの反応物を5分間45℃に加熱した後、30分間かけて室温に冷却する。次に、これら24の反応物を、バーコードドメイン (すなわち R_1 、 R_2 、 R_3) に対応する3つの異なるプールに分類する。例えば R_1 バーコードドメインに結合した8つの異なるレポータープローブ (それぞれ2 μ M) をまとめてプールし、10倍に希釈して各レポーター複合体の最終作業濃度を200 nMにする。レポーター複合体は、高圧液体クロマトグラフィ (HPLC) を利用して精製することができる。遊離したオリゴマーとうまく形成されなかったプローブをHPLC精製によって除去してレポータープローブを生成させることができる。

【0497】

レポーター複合体を調製した後は、品質保証のための標準的な試験である。3つの別々のフローセルの中で、レポータープローブの3つのプールのそれぞれを、対応するバーコード領域 (R_1 、 R_2 、 R_3 のいずれか) への結合に関して試験する。試験は、バーコードドメインが1つだけ存在していてフローセルの表面に固定化されている修飾されたシークエンシング用プローブコンストラクトで実施する。それぞれの色を表わす8個の12量体すべてを多重化すると、8つのレポータープローブすべてが大きな色カウント数で同定されることが予想される。

【0498】

レポータープローブとシークエンシング用プローブのバーコードドメインのハイブリダイゼーションの効率と精度を向上させるため、さまざまな緩衝液添加物を試験した。実験からの結果は、5%硫酸デキストラン (500K) と、15%ホルムアミドまたは15%炭酸エチレンとを含有する緩衝液により、レポータープローブとシークエンシング用プローブのハイブリダイゼーションが短いハイブリダイゼーション時間のうちに最高の効率と精度で可能になることを示している。しかし別の実験からの結果は、炭酸エチレンがシークエンシング用スライドの表面に対してマイナスの影響を有し、その結果として時間が経つと標的核酸が大きく失われることを示している。したがって5%硫酸デキストラン (500K) と15%ホルムアミドを含有する緩衝液が、レポータープローブとシークエンシング用プローブの効率的で正確なハイブリダイゼーションに関してより優れている。

【0499】

実施例11：相補的核酸配列の設計と確認

【0500】

レポータプローブは、シーケンシング用プローブのバーコードドメイン上の特定の位置（R₁、R₂、R₃のいずれか）に結合する相補的核酸を含有している。12個のヌクレオチド（12量体）または14個のヌクレオチド（14量体）を含有する相補的核酸を設計して試験し、ハイブリダイゼーションに最適な配列を決定した。スクリーニングのため、以下の基準を用いて最適な配列を決定した。すなわちその基準とは、レポータプローブとシーケンシング用プローブが10回のシーケンシングサイクルで80%超の効率で結合するという大きな結合効率を配列が示さねばならないことと；配列が15秒～30秒以内に起こる速いハイブリダイゼーション速度を示さねばならないことと；配列がレポータプールにおいて5%未満の交差ハイブリダイゼーションエラーという大きな特異度を示さねばならないことである。

【0501】

表4は、同定された24通りの12量体配列（配列番号19～42）を示している。各バーコードドメインは3つの位置を含有しているため、24通りの12量体配列を3つのグループに分割して8×8×8 12量体レポータセットを作り出すことができる。

【0502】

【表6】

レポータの位置	12 量体配列	レポータの名称	色	配列番号
1	AGGACAGATGAC	R1BB-07	BB	19
1	GTATCGGATGAC	R1BG-07d (R1RR-06)	BG	20
1	AGGAGTGATGAC	R1BR-07	BR	21
1	AGGGGTGAGGAG	R1GG-07c (R1YR-07)	GG	22
1	AGAGGGGATGAC	R1GR-07	GR	23
1	AGTGGGGAGGAG	R1GY-07c (R1BY-07)	GY	24
1	AGCCGAGATGAC	R1RR-07	RR	25
1	AGGGTGGATGAC	R1YY-07	YY	26
2	TGGATGAAAAAG	R2 BB (GRv5 用)	BB	27
2	GAAGGAGAAAAAG	R2 BG (GYv5 用)	BG	28
2	GGGGATGAAAAAG	R2 BR (GRv4 用)	BR	29
2	GTGAGGGAAAAAG	R2 BY (YYv5 用)	BY	30
2	AGCCGAGAAAAAG	R2 GG	GG	31
2	CGAGAGGAAAAAG	R2 GY (GGv5 用)	GY	32
2	GAGGGCGAAAAAG	R2 RR (GGv4 用)	RR	33
2	AGCGTGGAAAAAG	R2 YY	YY	34
3	TGAGAAGGGTAG	RPTR12-BG_Screen3_D2	BG	35
3	GTTGTTATTGTG	RPTR12-BR_RC_D4	BR	36
3	TTTGGGTTTAGG	RPTR12-BY_RC_D3	BY	37
3	GTTAGTGGGAAA	RPTR12-GR_RC_D7	GR	38
3	ATGGGAAAAAGT	RPTR12-GY_RC_D6	GY	39
3	GAGTTGGATGAG	RPTR12-RR_RC_D10	RR	40
3	ATGTTGTGGGTA	RPTR12-YR_RC_D9	YR	41
3	GAGGGTTTTAAG	RPTR12-YY_RC_D8	YY	42

【0503】

同様にして14量体配列を設計したが、3つの点で12量体配列とは異なっている。第1に、14量体配列はより長いハイブリダイゼーション配列を含有している。というのも14量体配列は、バーコードドメイン上の特定の位置に結合する14個の一本鎖ヌクレオチドを、12量体の中に存在する12個の一本鎖ヌクレオチドの代わりに含有しているからである。第2に、14量体配列は、配列の多様性がより大きい。なぜなら14量体配列はトーホールド

を介した除去に適した設計にできなかったからである。14量体配列はシーケンシング用プローブにより強くハイブリダイズするため、トーホールドを介した除去の効率が低下する。したがって14量体配列に関しては配列と独立な除去戦略を探し、スクリーニング中の配列の制約を緩くした。以下の一群の規則を含むアルゴリズムを利用してスクリーニングのための配列を設計した。その規則とは、「G」または「C」を欠くヌクレオチド組成（すなわち複雑度が低い配列）；GC含量が40%～60%；融点（Tm）が35℃～37℃；ヘアピン折り畳みエネルギー（dG）が2超；他のシーケンシング用プローブとの適合性（ハミング距離が7以上）である。標的核酸の中に存在する可能性のあるゲノム配列にハイブリダイズする14量体配列をできるだけ少なくするため、外部RNA管理コンソーシアム（External RNA Controls Consortium）配列をガイドとして用いて可能性のある配列をフィルタした。第3に、14量体相補的核酸がレポータ複合体の一次核酸に付着している位置で切断可能なリンカーの修飾を利用して鎖を切断することによってシーケンシング用プローブのバーコードドメインから除去されるように14量体配列を設計した。14量体配列を除去するとレポータ複合体のシグナルの「暗化」が起こり、次のスクリーニングサイクルとシグナルの検出が可能になる。UV光で切断可能なリンカー、還元剤（TCEPなど）で切断可能なリンカー、酵素で切断可能なリンカー（USER（商標）酵素によって切断されるウラシルなど）を含むさまざまな切断可能なリンカーの修飾を調べた。これらの切断可能なリンカーの修飾のすべてが、レポータ複合体の効率的な暗化を促進することが見いだされた。暗化は、切断可能なリンカーの修飾を二次核酸の中に導入することによってさらに増強された。これらの切断可能なリンカーの修飾は、一次核酸にハイブリダイズする配列と、三次核酸にハイブリダイズする配列の間に位置させた。図7は、切断可能なリンカーの修飾にとってレポータプローブの中で可能な位置を示している。

【0504】

可能性のある14量体配列のスクリーニングにより、受容可能な配列の2つのグループが同定された。表5は第1のグループを示しており、ここには24個の配列が含まれていた（配列番号43～66）。これら24個の配列を3つのグループに分割して8×8×8 14量体レポータセットを作り出すことができた。

【0505】

10

20

30

40

50

【表 7】

表5

レポータ	14 量体配列	レポータの名称	色	配列番号
A	ATCTTTTCCCCACT	R14-BG_RC-Sc3_B2	BG	43
A	CCCCACTATTTCTT	RPTR14-BY_Screen4_I2	BY	44
A	CTACCCACAACATA	RPTR14-YR_Screen3_D9	YR	45
A	CCATATAAACCCCA	R14-GG_RC-Sc3_B5	GG	46
A	AAACTCCAATCTCC	R14-GR_RC-Sc3_B7	GR	47
A	CTATTCTCAACCTA	RPTR14-YY_RS0255_H8	YY	48
A	CCCCCTCTTTTAAA	R14-BB_RC-Sc3_B1	BB	49
A	CCAATCTTACCTCA	RPTR14-RR_Screen3_B10	RR	50
B	CCCTCACATAACTT	RPTR14-BG_Screen4_I1	BG	51
B	CTCCTCTACTTTCC	RPTR14-BB_ERCC_00014.1_10	BB	52
B	CCCTAAACCCAAAA	RPTR14-BY_Screen3_D3	BY	53
B	CACTTTTTCCCATC	RPTR14-GY_Screen3_D6	GY	54
B	CATCTGATTCCTCC	R14-RR_ERCC_00042.1_150RC	RR	55
B	CTAAACCCCTACT	R14-BR_RC-Sc3_B4	BR	56
B	CCTTTACAAACACA	RPTR14-GR_RS0247_H7	GR	57
B	ATACCACCCTCTTT	RPTR14-YY_Screen3_B8	YY	58
C	TATTCTTCTACCCC	RPTR14-YR_Screen4_I5	YR	59
C	TCTACCCTTCTCAT	R14-BG_RC-Sc3_D2	BG	60
C	CCACAATAACAACC	RPTR14-BR_Screen3_D4	BR	61
C	ACCTTAACATTCCC	R14-GG_RC-Sc3_D5	GG	62
C	ATTTCCCACTAACC	RPTR14-GR_Screen3_D7	GR	63
C	ACTTAAACGCTCC	RPTR14-YY_Screen3_D8	YY	64
C	TACCTATTCCTCCA	RPTR14-BB_Screen3_D1	BB	65
C	CCCCTTTCTCTAAG	RPTR14-RR_ERCC_00051.1_220	RR	66

【0506】

表6は他方のグループを示しており、この中には30個の配列が含まれている（配列番号67～96）。これら30個の配列を3つのグループに分割して10×10×10 14量体レポータセットを作り出すことができた。

【0507】

10

20

30

40

50

【表 8】

表6

レポータ	14 量体配列	レポータの名称	色	配列番号
A	GATGATGGTAGGTG	R14_PC_J2_BB_v2	BB	67
A	ATGAGAAGGGTAGA	R14_PC_D2_BG_v2	BG	68
A	GTTTTGTTGGTGAG	R14_PC_K2_BY_v2	BY	69
A	TTAGTGTGTTGGAG	R14_PC_K5_BR_v2	BR	70
A	ATGTAGGAGAGAGA	R14_PC_L1_GG_v2	GG	71
A	GGGAATGTTAAGGT	R14_PC_D5_GY_v2	GY	72
A	GGTTAGTGGGAAAT	R14_PC_rcD7_GR_v2	GR	73
A	GGAGGGTTTTAAGT	R14_PC_rcD8_YY_v2	YY	74
A	GTAGTGTGGATGTT	R14_PC_J5_YR_v2	YR	75
A	CTTAGAGAAAGGGG	R14_PC_ERCC51_RR_v2	RR	76
B	GGAAGAGGATGAAA	R14_PC_K1_BB_v2	BB	77
B	AAGTTATGTGAGGG	R14_PC_spB_BG_v1	BG	78
B	GGAAAGTAGAGGAG	R14_PC_spB_BY_v1	BY	79
B	TTTTGGGTTTAGGG	R14_PC_spB_BR_v1	BR	80
B	AGATGTATGGGTGA	R14_PC_L2_GG_v2	GG	81
B	GATGGGAAAAAGTG	R14_PC_spB_GY_v1	GY	82
B	GGAGGAATCAGATG	R14_PC_spB_GR_v1	GR	83
B	AGAGGGATTGATGA	R14_PC_J4_YY_v2	YY	84
B	TGTGTTTGTAAGG	R14_PC_spB_YR_v1	YR	85
B	AAGGAGTGATAGGA	R14_PC_J1_RR_v2	RR	86
C	TGGTGATTTAGAGG	R14_J3_BB_v2	BB	87
C	GGGGTAGAAGAATA	R14_rcI5_BG_v2	BG	88
C	AAGAAATAGTGGGG	R14_PC_spA_BY_v1	BY	89
C	TATGTTGTGGGTAG	R14_PC_spA_BR_v1	BR	90
C	GTAAAGGGAGGTT	R14_K3_GG_v2	GG	91
C	TGGGGTTTATATGG	R14_PC_spA_GY_v1	GY	92
C	AGGGAATATGGAGA	R14_K6_GR_v2	GR	93
C	TAGGTTGAGAATAG	R14_PC_spA_YY_v1	YY	94
C	TTTAAAAGAGGGGG	R14_PC_spA_YR_v1	YR	95
C	TGAGGTAAGATTGG	R14_PC_spA_RR_v1	RR	96

【0508】

スクリーニングの後、 $8 \times 8 \times 8$ 12量体レポータセット、 $8 \times 8 \times 8$ 14量体レポータセット、 $10 \times 10 \times 10$ 14量体レポータセットを実験で確認した。 $8 \times 8 \times 8$ 12量体結合スキームに関しては、Hyb & Seqプロトタイプを用いてシーケンシングサイクルを10回記録する確認を実行した。レポータプローブの3つのプールを長い作業フロー法と短い作業フロー法の両方で使用した。512通りの可能なシーケンシング用プローブのバーコードドメインをすべて調べた。表7は、長い作業フロー法と短い作業フロー法の実験工程を示している。

【0509】

10

20

30

40

50

【表 9】

表7

1つのサイクル 内の工程	長い作業フロー： トーホールドなしでのレポータの ハイブリダイゼーション	短い作業フロー： トーホールドありでのレポータの ハイブリダイゼーション
1	レポータ 1 を 15 秒間、または 30 秒間、または 60 秒間	レポータ 1 を 30 秒間
2	画像取得	画像取得
3	暗化するためトーホールド 1 を 60 秒間	レポータ 2+トーホールド 1 を 15 秒間
4	画像取得	画像取得
5	レポータ 2 を 15 秒間、または 30 秒間、または 60 秒間	レポータ 3+トーホールド 2 を 15 秒間
6	画像取得	画像取得
7	暗化するためトーホールド 2 を 60 秒間	洗浄
8	画像取得	画像取得
9	レポータ 3 を 15 秒間、または 30 秒間、または 60 秒間	
10	画像取得	
11	洗浄	
12	画像取得	

【0510】

長い作業フロー実験によって暗化効率が97%超になった。短い作業フロー実験に関して
は暗化が同様に効率的であると推定したが、暗化されないレポータが各画像にわずかな頻
度で残って新しいレポータと間違っ呼ばれる可能性があることが予想された。実際、短
い作業フロー実験でバーコードカウント数が最も多かったのはYYYYYYであった。これは
、暗化していないアーチファクトとバックグラウンドである可能性が大きかった。8×8×
8 12量体レポータセットの性能は、短い作業フローでは長い作業フローと比べて一般に低
かった。（バーコードドメインの位置R₁に結合する）レポータ複合体1と（バーコードド
メインの位置R₃に結合する）レポータ複合体3は、短い作業フローでは長い作業フローと
比べて効率が低かった。レポータ複合体3ではそうなることが予想される。というのもレ
ポータ複合体3は、8個の追加トーホールドオリゴヌクレオチドをそれぞれ2.5 μMという
高い濃度で含んでいるため、レポータのハイブリダイゼーションに干渉する可能性がある
からである。レポータ複合体1は、2つの作業フローで同様に機能するはずである。とい
うのも短い作業フローと長い作業フローのどちらでも、第1のレポータ複合体を除去する
のにトーホールドを使用しなかったからである。全エラーも、3つのレポータプローブのす
べてで、短い作業フローでは長い作業フローよりも多かった（1.3～2倍）。

【0511】

8×8×8 14量体レポータセットは、可能な512通りのシーケンシング用プローブす
べてのバーコードドメインにハイブリダイズする効率、特異性、速度を調べることでよ
って確認した。シーケンシング用プローブのバーコードドメインは、Hyb & Seqシーク
エンシングカートリッジのガラスの表面に直接固定化した。8×8×8 14量体レポータプ
ローブは、わずか15秒の間に平均効率88%でハイブリダイズし、平均エラー率は5.1%であ
った。このエラーの大半は、正確でないハイブリダイゼーションが原因ではなく、レポー
タの正確でない同定が原因である。レポータの分類ミスのエラーが、レポータエラーの最
大の要素として残る。

【0512】

10×10×10 14量体レポータセットは、30通りの相補的な切断されたシーケンシン

グ用プローブバーコードドメインにハイブリダイズする効率、特異性、速度を調べることによって確認した。それぞれのバーコードドメインは、レポータ結合部位を1つだけ含んでいた。これらバーコードドメインは、Hyb & Seqシーケンシングカートリッジのガラスの表面に直接固定化した。10×10×10 14量体レポータセットは、わずか15秒の間に平均効率90%でハイブリダイズし、平均エラー率は5.0%であった。ここでもエラーの大半は、正確でないハイブリダイゼーションが原因ではなく、レポータの正確でない同定が原因であった。

【0513】

実施例12：標準シーケンシング用プローブと3部構成シーケンシング用プローブの設計と試験

【0514】

シーケンシング用プローブの標的結合ドメインとバーコードドメインは二本鎖「ステム」によって隔てられている。図2は、実験で調べた2つのシーケンシング用プローブアーキテクチャを示している。標準シーケンシング用プローブでは、標的結合ドメインとバーコードドメインは同じオリゴヌクレオチド上に存在し、このオリゴヌクレオチドがステムオリゴヌクレオチドに結合して長さが36個のヌクレオチドの二本鎖領域を生み出している。このアーキテクチャを用いると、プローブのプールの中の各シーケンシング用プローブは同じステム配列を利用する。3部構成プローブでは、標的結合ドメインとバーコードドメインは別々のDNAオリゴヌクレオチドであり、36個のヌクレオチドからなるステムオリゴヌクレオチドによって互いに結合している。起こる可能性のあるバーコードドメインの交換を阻止するため、各バーコードは独自のステム配列を持っていて、シーケンシング用プローブをプールする前に別々にハイブリダイズさせる。

【0515】

図22は、3部構成シーケンシング用プローブを標準シーケンシング用プローブと比べた一連の実験の結果を示している。これらの実験から、3部構成シーケンシング用プローブが、第3のレポータプローブの検出を含め、両方の構成に関してあらゆる読み出しの約80%で1つのシーケンシングサイクル全体を生き延びることが確認された。3部構成プローブは、標準シーケンシング用プローブと比べて約12%少ないカウント数を示す。バーコードドメインオリゴの交換の傾向を調べるため、同じステム配列を含有する高濃度の短い代わりのオリゴヌクレオチドを反応物に添加した。結果は、検出された3部構成シーケンシング用プローブの約13%がバーコードオリゴを交換したことを示していた。オリゴヌクレオチドの交換は、独自のステム配列を組み込むことによって減らす必要がある。3部構成プローブは、性能がわずかに低いにもかかわらず、設計の柔軟性、迅速なオリゴ合成、コスト低減という利点を提供する。

【0516】

実施例13：標的結合ドメインにおけるロックド核酸置換の効果

【0517】

シーケンシング用プローブの標的結合ドメインの中に置換によってロックド核酸（LNA）を入れることの効果を以下のようにして調べた。溶液の中でシーケンシング用プローブをレポータプローブにハイブリダイズさせ、適切に形成されたシーケンシング用プローブ-レポータプローブを精製した。次に、溶液の中でシーケンシング用プローブ-レポータプローブを合成標的核酸にハイブリダイズさせ、プロトタイプシーケンシングカートリッジの表面に堆積させた。合成標的核酸は50個のヌクレオチドからなり、ビオチニル化されていた。シーケンシング用プローブは、個別に試験するか、9つのシーケンシング用プローブからなるプールの中で試験した。9つのシーケンシング用プローブからなるプールに関しては、プローブを、標的核酸の長さに沿って結合する設計にした。分析のため、ストレプトアビジンで被覆したカバースライドの表面に反応物全体をブレッドボード装置によって堆積させた後、流体力で伸長させた。次にレポータプローブの画像を取得し、適切な装置とソフトウェア（例えばNanoString nCounter（登録商標）装置とソフトウェア）を利用してカウントした。

【0518】

それぞれのシークエンシング用プローブは、10個のヌクレオチド（配列番号97）からなる標的結合ドメインを含有していた。標的結合ドメイン内のLNA置換を実行し、2個、または3個、または4個のLNA塩基が図23に示した位置に含まれるようにした。図23は、LNA塩基の数が増えるにつれて標的核酸に対する個々のシークエンシング用プローブの結合親和性が大きくなったことを示している。重要なことだが、図23は、LNA塩基の組み込みによってシークエンシング用プローブの結合の特異度は低下しなかったことを示している。9つのシークエンシング用プローブからなるプールを試験し、プローブが標的への結合を完了できたときの塩基カバレッジを求めた。図24は、単一のLNAプローブをプールの中に導入したとき、影響を受ける塩基のカバレッジは増加するが、周囲のプローブの結合にはほとんど影響を与えなかったことを示している。これらの結果は、LNA塩基による置換が特異度の低下なしに塩基感受性を向上させうることを示していた。

10

【0519】

実施例14：標的結合ドメインにおける修飾されたヌクレオチドと核酸類似体による置換の効果

【0520】

さまざまな修飾されたヌクレオチドと核酸類似体（ロックド核酸（LNA）、架橋型核酸（BNA）、プロピンで修飾された核酸、ジップ核酸（ZNA（登録商標））、イソグアニン、イソシトシンが含まれる）による置換をシークエンシング用プローブの標的結合ドメインの中に導入することの効果以下のようにして調べた。ヌクレオチド50個の長さを有するビオチニル化された標的核酸を、プロトタイプシークエンシングカートリッジのストレプトアビジンカバースライドの表面に堆積させた。次にシークエンシング用プローブとレポータプローブをサンプル室の中に順番に導入し、Hyb & Seqプロトタイプ装置を用いて画像を取得した。画像を処理し、異なるそれぞれのシークエンシング用プローブについてカウント数を比較した。シークエンシング用プローブの10個のヌクレオチド（配列番号99）からなる標的結合ドメインで置換を実行し、図25に示した位置にLNA塩基、BNA塩基、プロピン塩基、ZNA（登録商標）塩基が含まれるようにした。図25は、LNAとBNAを含有するプローブが結合親和性に関して最大の増加を示した一方で、特異度は維持したことを示している。そのことは、一致した標的と一致しない標的で検出されたカウント数に示されている。これらの結果は、LNA塩基またはBNA塩基による置換が特異度の低下なしに塩基感受性を向上させうることを示していた。

20

30

【0521】

実施例15：本開示によるシークエンシング法の精度を求める

【0522】

図26は、本開示によるシークエンシング法の生の特異度を定量する実験からの結果を示している。この実験では、異なる4つのシークエンシング用プローブからなるプールを、N RASエキソン2（配列番号1）の断片を含む標的核酸にハイブリダイズさせるというシークエンシング反応を実施した。それぞれのシークエンシング用プローブ（バーコード1~4）は、図26の最も上の図に示されているように、6量体の位置b₅が異なる以外は同じである標的結合ドメインを持っていた。この実施例では、バーコード4は正しいシークエンシング用プローブである。シークエンシング用プローブをハイブリダイズさせた後、バーコードドメインの3つの位置（R₁、R₂、R₃）のそれぞれにレポータプローブを順番にハイブリダイズさせ、対応する蛍光データを記録した。図26の中央の図は、バーコードドメインの3つの位置について、それぞれの色組み合わせが記録された回数と、正しい組み合わせが記録された回数の割合を示している。R₁における色組み合わせは96%の回数で正しく同定され、R₂における色組み合わせは97%の回数で正しく同定され、R₃における色組み合わせは94%の回数で正しく同定された。図26の最も下の図に示されているように、こうすることで全体的な生の特異度が94%になる。バーコードドメインの位置が誤って読み取られたことを説明できる可能性のあるエラー発生源に含まれるのは、（a）フローセルの表面へのレポータプローブの非特異的結合と、（b）レポータプローブの正確でないハイブリ

40

50

ダイゼーションである。レポータハイブリダイゼーションエラーの推定値は約2~4%であった。

【0523】

図27は、標的核酸の中のヌクレオチドを2つ以上のシークエンシング用プローブによってシークエンシングするときの本開示によるシークエンシング法の精度を求める実験からの結果を示している。図27の最も上の図に示されているように、この実施例の標的核酸は、NRASエキソン2（配列番号1）の断片である。目的の特定の塩基は、標的核酸の中で強調してあるシトシン（C）である。目的のこの塩基は、2つの異なるシークエンシング用プローブにハイブリダイズし、標的核酸へのハイブリダイゼーションのフットプリントがそれぞれ異なることになる。この実施例では、シークエンシング用プローブ1~4（バーコード1~4）は、目的の塩基の左側に3個のヌクレオチドを結合させるのに対し、シークエンシング用プローブ5~8（バーコード5~8）は、目的の塩基の左側に5個のヌクレオチドを結合させる。図27の中央の図は、シークエンシング用プローブのバーコードドメインのそれぞれの位置で特定の色組み合わせが記録された回数を示している。画像の定量後、図26に示した塩基読み取り技術を利用すると、約98.98%という平均精度を記録することができる。

【0524】

実施例16：CRISPRに基づく断片化を利用して標的核酸を捕捉する

【0525】

以下に示すようにして、（上に説明した）c3プローブ複合体と（上に説明した）c5プローブ複合体を用いて標的核酸を捕捉し、基板の表面に固定化した。CRISPRに基づく断片化を利用してゲノムDNA（gDNA）を断片化した。c3プローブ複合体とc5プローブ複合体を含む溶液を断片化されたgDNAと混合してインキュベートし、c3プローブ複合体とc5プローブ複合体を標的核酸にハイブリダイズさせると、図34に示されている複合体が形成された。次に、この標的核酸-捕捉用プローブ複合体を、HiFi TaqリガーゼとFEN1の混合物とともに65℃でインキュベートした。5'オーバーハングフラップ構造をFEN1によって除去し、標的核酸分子の3'末端を、基板特異的ドメインを含むc5プローブ複合体の鎖に連結した。インキュベーションの後、標的核酸-捕捉用プローブ複合体を20 µlのF様ビーズとともに攪拌しながら10分間インキュベートした。F様ビーズを磁石で捕捉し、3回洗浄した。標的核酸-捕捉用プローブ複合体をF様ビーズから溶離させるため、F様ビーズを20 µlのUSER酵素混合物の中に懸濁させ、25℃で15分間インキュベートした。ビーズを磁石で除去し、溶離した標的核酸-捕捉用プローブ複合体を含有する得られた上清を、SPRIビーズを用いてさらに精製した。36 µlのSPRIビーズを上清と混合し、25℃で10分間インキュベートした。次いでビーズを磁石に結合させ、100 µlの80%エタノールで2回洗浄した。次いでエタノールを除去し、ビーズを25℃で5分間乾燥させた。乾燥後、ビーズを10 µlの0.1×SSPEの中に再懸濁させた。精製した核酸-捕捉用プローブ複合体をSPRIビーズから10分間かけて溶離させた。溶離後、異なる生物サンプルに由来する標的核酸-捕捉用プローブ複合体をまとめてプールし、F-ビーズに添加した。プールした複合体をF-ビーズとともに25℃で10分間インキュベートした。F-ビーズを3回洗浄し、8 µlの中に再懸濁させた。次いでF-ビーズを、c5プローブ複合体の基板特異的ドメインと相補的で基板に付着させた核酸分子を含むカードに移した。基板に付着したこれらの相補的核酸分子を基板特異的ドメインにハイブリダイズさせることにより、標的核酸を基板に固定化した。次いで基板をUV光に曝露してc3プローブ複合体上の光で切断可能なリンカーを切断することにより、標的核酸-捕捉用プローブ複合体をF-ビーズから放出させた。

【0526】

実施例17：CRISPRに基づくgDNAの断片化

【0527】

以下のプロトコルを用いると、CRISPRに基づく断片化工程を利用してゲノムDNAを断片化することができる。このプロトコルにより、二本鎖標的DNAをインビトロで切断し、本開示の方法で用いることのできる標的核酸分子を生成させることができる。

【0528】

以下のプロトコルでは、複数の一本鎖RNA (sgRNA) とV3-Cas9複合体の両方を含有するCas9 RNP複合体を使用する。この非限定的な実施例では、ガイドRNAは、切断することを望むdsDNAと正確に一致する分子の5'末端にある標的配列に対応する17～20個のヌクレオチドを含有していた。

【0529】

RNアーゼなしの環境で以下のプロトコルを実施した：(1) それぞれのsgRNA分子を含む試験管を遠心分離することにより、乾燥したRNAペレットが試験管の底に回収されるようにした。この実施例では、154個の別々のsgRNAを用いて154個の切片を作ると、77個の標的核酸が生成した；(2) それぞれのsgRNAペレットを、ヌクレアーゼを含まない1×TE緩衝液に溶かし、sgRNAの最終濃度を50 μMにした；(3) その後、工程(2)からの5 μlの各サンプルと混合することによってすべてのsgRNAをまとめてプールした。

【0530】

sgRNAをまとめてプールした後、プロトコルはさらに、(4) sgRNAとCas9酵素を表9と表10に記載されている同じモル量で組み合わせることを含んでいた。各切断部位について優れた切断効率となるようにするため、切断するdsDNA 1000 ngにつき1.0ピコモルのsgRNAとCas9を使用した。

【0531】

【表10】

表9

切断部位1つ当たりの量	sgRNA	1 ピコモル
	Cas9	1 ピコモル
	10×Cas9 緩衝液	0.3 μl
	100×BSA	0.1 μl
	DTT (1 M)	0.2 μl
	dH ₂ O (最終体積が10 μlになるまで添加)	X μl

【0532】

【表11】

表10

切断部位1つ当たりの量	sgRNA 1～154 (それぞれ50 μM)	3.1 μl
	Cas9	2.5 μl
	10×Cas9 緩衝液	0.3 μl
	100×BSA	0.1 μl
	DTT (1 M)	0.2 μl
	dH ₂ O	3.8 μl

【0533】

sgRNAとCas9酵素を混合した後、プロトコルはさらに、(5) このsgRNA-Cas9混合物を37℃で10分間インキュベートしてRNP複合体を形成することと；(6) その後、1000 ngの抽出されたヒトゲノムDNA (hgDNA) を、(5) で形成されたRNP複合体と、表11に示されている量に従って室温で混合することを含んでいた。

【0534】

10

20

30

40

50

【表 1 2】

表11

1 μ g の hgRNA (77 個の標的核酸)	NEB3 緩衝液	4 μ l
	DTT	2 μ l
	CRISPR154	10 μ l
	hgDNA	1000 ng
	ddH ₂ O (最終体積が 10 μ l になるまで添加)	X μ l

【0 5 3 5】

hgDNAをRNP複合体と混合した後、プロトコルはさらに、(7) 工程 (6) でアセンブルした反応物を37℃で60分間インキュベートし；(8) 工程 (7) からの反応物を70℃で10分間インキュベートすることによりCRISPR切断物を不活性化し；(9) 1 μ lのプロテイナーゼK (20 mg/mlの貯蔵溶液) を工程 (8) からの反応物に添加し、この混合物を56℃で10分間インキュベートしてCas9エンドヌクレアーゼからDNA基質を放出させることを含んでいた。あるいは2 μ lのQiagenプロテアーゼ (1.07 AU/ml) を工程 (8) からの反応物に添加した後、37℃で30分間インキュベートし、その後2回目のインキュベーションを70℃で15分間実施してQiagenプロテアーゼを加熱により不活性化することができる。

【0 5 3 6】

実施例18：CRISPRに基づく断片化工程とUSERが媒介する切断を利用して断片化したDNAからの標的核酸の精製

【0 5 3 7】

CRISPRに基づく断片化の後、以下のプロトコルを利用すると、本開示のプロープと方法を用いて特定の核酸を捕捉して精製することができる。

【0 5 3 8】

サンプルから抽出したDNAをCRISPRに基づいて断片化した後、断片化されたDNAを捕捉用プロープと混合する。この非限定的な実施例では、断片化されたDNAを2つの捕捉用プロープと混合する。第1の捕捉用プロープは、特定の標的核酸の5'末端と相補的なドメインと、親和性部分を含んでいる。この実施例では、親和性部分はF-タグである。切断可能な部分は、F-タグと、標的核酸の5'末端と相補的なドメインの間に位置する。この実施例では、切断可能な部分はUSER配列である。第2の捕捉用プロープは、標的核酸の3'末端と相補的なドメインを含んでいる。第2の捕捉用プロープは、任意により親和性部分を含むことができる。1 nMの捕捉用プロープを20 fMの標的核酸と5×SSPE緩衝液の中で混合し、最終反応体積を100 μ lにする。

【0 5 3 9】

捕捉用プロープを標的核酸にハイブリダイズさせるため、断片化されたDNAとプロープを含む溶液を最初に98℃で3分間インキュベートした後、65℃で15分間インキュベートし、最後に、プロトコルの次の工程まで22℃でインキュベートした。

【0 5 4 0】

ハイブリダイゼーションの後に標的核酸-捕捉用プロープ複合体を捕捉するため、CRISPRで断片化されたDNAと捕捉用プロープを含む溶液を（第1の捕捉用プロープの表面に位置するF-タグに結合することができる）、F-ビーズとともに、回転させながら、または手で混合しながら、室温で10分間インキュベートする。インキュベーションの後、ビーズを1 mlの0.1×SSPE緩衝液で3回洗浄し、ビーズに特異的に結合していないDNAを除去する。ビーズを100 μ lの1×SSPEの中に再懸濁させ、PCRチューブに移し、沈降させる。上清を除去し、1×Cutsmart緩衝液と1 μ lのUSER酵素を含有する10 μ lのUSERマスター混合物の中にビーズを再懸濁させる。ビーズをUSERマスター混合物の中で室温にて15分間インキュベートすることによって第1の捕捉用プロープの中のUSER配列を切断し、ビーズから

標的核酸を放出させる。ビーズを沈降させ、標的核酸を含有する上清を回収する。

【0541】

標的核酸をさらに精製するため、AMPure精製を2回実施する。AMPure精製では、標的核酸を含有する上清をAMPure XPビーズ（固相可逆的固定化ビーズ、SPRI）と、サンプル：ビーズ＝1：1.8の比率で混合する。この混合物を10分間インキュベートして標的核酸をビーズに結合させる。次いでビーズを200 μ lの75～80%エタノールで2回洗浄する。各回の洗浄中に磁石を用いてビーズをエタノールから2回引っ張る。次いでビーズを10 μ lの0.1×SSPE溶液に再懸濁させ、10分間インキュベートする。最後に、精製された標的核酸をビーズから溶離させる。

【0542】

実施例19：CRISPRに基づく断片化工程とUSERが媒介する切断を利用して断片化したDNAからの標的核酸の精製と堆積

【0543】

CRISPRに基づく断片化の後、以下のプロトコルを用いると、本開示のプロープと方法を用いて特定の核酸を捕捉して精製し、適切な基板上に堆積させることができる。

【0544】

サンプルから抽出したDNAをCRISPRに基づいて断片化した後、断片化されたDNAを捕捉用プロープと混合する。この非限定的な実施例では、図34に示されているように、断片化されたDNAをc3プロープ複合体およびc5プロープ複合体と混合するが、C5プロープ上に位置するF様タグがGタグで置き換えられている点が異なっている。1 nMの捕捉用プロープと、20 fMの標的核酸と、1 μ lのHiFi DNAリガーゼと、2 μ lのFEN1酵素を1×HiFi緩衝液の中で混合して最終反応体積を100 μ lにする。

【0545】

捕捉用プロープを標的核酸にハイブリダイズさせるため、標的核酸とc5プロープ複合体によって形成される5'オーバーハングフラップを切断し、標的分子の3'末端を、基板特異的ドメインを含むc3プロープの鎖に連結させ、断片化されたDNAとプロープを含む溶液を98℃で3分間インキュベートした後、65℃で60分間インキュベートし、最後に、プロトコルの次の工程まで22℃でインキュベートする。65℃でのインキュベーションの最後の10分間は10 μ lの5 M NaClを添加して塩の濃度を500 mMにし、捕捉用プロープのハイブリダイゼーションを安定化させるのを助ける。

【0546】

ハイブリダイゼーション、切断、連結の後に標的核酸-捕捉用プロープ複合体を捕捉するため、断片化されたDNAと捕捉用プロープを含む溶液を（G-タグに結合することのできる）、G-ビーズとともに、回転させながら、または手で混合しながら、室温で10分間インキュベートする。インキュベーションの後、ビーズを1 mlの0.1×SSPE緩衝液で3回洗浄し、ビーズに特異的に結合しないDNAを除去する。ビーズを100 μ lの1×SSPEの中に再懸濁させ、PCRチューブに移し、沈降させる。上清を除去し、1×Cutsmart緩衝液と1 μ lのUSER酵素を含有する10 μ lのUSERマスター混合物の中にビーズを再懸濁させる。ビーズをUSERマスター混合物の中で室温にて15分間インキュベートすることによって第1の捕捉用プロープの中のUSER配列を切断し、ビーズから標的核酸を放出させる。ビーズを沈降させ、標的核酸を含有する上清を回収する。

【0547】

標的核酸をさらに精製するため、AMPure精製を2回実施する。AMPure精製では、標的核酸を含有する上清をAMPure XPビーズ（固相可逆的固定化ビーズ、SPRI）と、サンプル：ビーズ＝1：1.8の比率で混合する。この混合物を10分間インキュベートして標的核酸をビーズに結合させる。次いでビーズを200 μ lの75～80%エタノールで2回洗浄する。各回の洗浄中に磁石を用いてビーズをエタノールから2回引っ張る。次いでビーズを磁石の上で5分間乾燥させる。次いでビーズを10 μ lの0.1×SSPE溶液に再懸濁させ、10分間インキュベートする。最後に、精製された標的核酸をビーズから溶離させる。次に、精製された標的核酸を、標的核酸に連結した基板特異的ドメインに結合することのできる

10

20

30

40

50

基板の上に堆積させることができる。

【0548】

実施例20：CRISPRに基づく断片化工程を利用して断片したDNAからの標的核酸の精製

【0549】

最初に、DEPCで処理した／無ヌクレアーゼ水を用い、あらかじめ断片化されたgDNAを捕捉用プローブと5×SSPE緩衝液の中で混合し、100 μlのハイブリッド-捕捉反応物を調製する。あらかじめ断片化されたgDNAをハイブリッド-捕捉反応物に添加する前に、Qubit dsDNA BRアッセイキット（Thermo Scientific社）を用いてそのあらかじめ断片化されたgDNAの濃度を測定する。Qubitアッセイの定量範囲は2～1000 ngである。好ましくは、Qubitアッセイを利用して10～100 ngのgDNAサンプルを測定すべきである。gDNAサンプルの濃度が1000 ng/μlよりも大きいと考えられる場合には、サンプルを最初に約10～1000 ng/μlに希釈すべきである。合計して100～1000 ngのgDNA、好ましくは1000 ngのgDNAをハイブリッド-捕捉反応物に添加すべきである。

【0550】

この非限定的な実施例では、精製するそれぞれの標的核酸に2つの捕捉用プローブを用いる。第1の捕捉用プローブは、特定の標的核酸の5'末端と相補的なドメインと、親和性部分を含んでいる。この実施例では、親和性部分はF-タグである。第2の捕捉用プローブは、標的核酸の3'末端と相補的なドメインを含んでいる。第2の捕捉用プローブは、任意により親和性部分を含むことができる。最初に1 μM/プローブの貯蔵溶液を調製する。この貯蔵溶液を希釈して50 nM/プローブの作業溶液にする。2 μlの50 nM/プローブの作業溶液を100 μlの最終ハイブリッド-捕捉反応物に添加する。

【0551】

捕捉用プローブを標的核酸にハイブリダイズさせるため、ハイブリッド-捕捉反応物を98℃で3分間インキュベートした後、65℃で15分間インキュベートする。この65℃でのインキュベーションの後、反応物はプロトコルの次の工程まで22℃で保管することができる。

【0552】

F-タグに結合することのできるF-ビーズを、1 mlの5×SSPEで2回洗浄することによって調製する。洗浄後、ビーズを沈降させ、上清を除去し、5×SSPEと60%ホルムアミドと0.1%トウイン-20を含む緩衝液の中にビーズを再懸濁させる。

【0553】

おおまかには、F-ビーズは、サンプルの濃度が10 mg/mlだと1.25ピコモル/μlの結合能力を有することが予想される。反応物の中に存在する第1の捕捉用プローブの全モル数によって判断すると、F-ビーズの結合能力の1.5倍を利用することが好ましい。例えば77個の異なる標的核酸を捕捉することを目的とした1つの反応では、（それぞれの標的核酸に対応する）それぞれの第1のプローブを0.1ピコモル使用する。これは、合計して7.7ピコモルの第1のプローブが存在することを意味する。7.7ピコモルの1.5倍は11.55ピコモルである。F-ビーズの結合能力は1.25ピコモル/μlであると計算されているため、ビーズは9.24 μl使用せねばならない。

【0554】

捕捉用プローブがハイブリダイズした後、100 μlのハイブリッド-捕捉反応物を100 μlの再懸濁させたF-ビーズとともに、回転させながら室温で10分間インキュベートする。捕捉用プローブ-標的核酸複合体がビーズに結合した後、ビーズを洗浄して捕捉されなかったgDNAを除去する。ビーズを洗浄するため、ビーズを最初に磁石で濃縮し、次いで1 mlの0.1×SSPEで3回洗浄する。最後の洗浄の後、900 μlの最終洗浄溶液を除去し、ビーズを残っている100 μlの中に再懸濁させる。再懸濁させたビーズをクリーンなPCRチューブに移し、磁石を用いて濃縮し、上清を除去する。

【0555】

捕捉用プローブ-標的核酸複合体をビーズから溶離させるため、ビーズを12 μlの0.1×SSPEの中に再懸濁させ、45℃で7.5分間インキュベートする。インキュベーションの後、

ビーズを磁石の上で迅速に濃縮し、12 μ lの溶離液を除去する。

【0556】

2回のSPRI (AMPure) 精製を利用して標的核酸をさらに精製する。12 μ lの溶離液をAMPureビーズと、溶離液：ビーズ=1：1.8の比率で混合し、ローテータの上で室温にて10分間インキュベートする。次にビーズを磁石で濃縮し、100 μ lの80%エタノールで2回洗浄する。1回の洗浄につき2回、磁石を通じてビーズを移動させ、エタノールによってビーズペレット全体が確実に洗浄されるようにする。2回目のエタノール洗浄の後、できるだけ多くのエタノールを除去し、蓋をしていないチューブの中でビーズを磁石の上で5分間乾燥させる。次いでビーズを12 μ lの0.1×SSPEの中に再懸濁させ、室温で10分間溶離させる。ビーズを濃縮し、溶離液を除去する。2回目のSPRI精製の後、1 μ lの20×SSPE緩衝液を最終溶液に添加して捕捉用プローブ-標的核酸複合体を安定化させる。

【0557】

実施例21：本開示のシーケンシング用プローブを用いた標的核酸のシーケンシング

【0558】

以下は、合成標的核酸をシーケンシングするために設計した2つの異なるシーケンシング用プローブの利用を説明する1つの実施例である。

【0559】

第1のシーケンシング用プローブの設計（ここではLG介在（LG-spaced）シーケンシング用プローブと呼ぶ）が図53に示されている。5'末端から開始すると、このシーケンシング用プローブは標的結合ドメインを含んでいる。標的結合ドメインは長さが6個のヌクレオチド（6量体）配列（b₁-b₂-b₃-b₄-b₅-b₆）を含んでいて、標的核酸内の6個の相補的なヌクレオチドに特異的にハイブリダイズする。この6量体配列はそれぞれの側に縮重塩基（N）が隣接している。このシーケンシング用プローブは、L-DNAステムと呼ぶ二本鎖DNAスペーサを介して標的結合ドメインに共有結合しているバーコードドメインも含んでいる。バーコードドメインと二本鎖DNAスペーサの両方とも全面的にL-DNAからなる一方で、標的結合ドメインは全面的にD-DNAからなる。L-DNAステムは長さが25個のヌクレオチドである。バーコードドメインは長さが27個のヌクレオチドであり、3つの付着領域（スポット1、スポット2、スポット3と呼ぶ）を含んでいる。それぞれの付着位置は9個のヌクレオチドを含んでいる。各付着領域の3'末端のヌクレオチドはL-dGヌクレオチドである。

【0560】

これらの実験のため、512通りの異なる種類のLG介在シーケンシング用プローブのプール（LG介在プール）を設計した。それぞれのシーケンシング用プローブは、スポット1核酸配列とスポット2核酸配列とスポット3核酸配列の独自の組み合わせを含んでいる。付着領域のために合計で30通りの複雑性が低い独自の核酸配列を設計した。そのうちの10通りの配列はスポット1で使用され、10通りの配列はスポット2で使用され、10通りの配列はスポット3で使用される。それぞれ異なる種類のLG介在シーケンシング用プローブは、標的結合ドメインに異なる6量体配列も含んでいる。

【0561】

第2のシーケンシング用プローブの設計（ここではD-ポケットシーケンシング用プローブと呼ぶ）が図54に示されている。5'末端から開始すると、このシーケンシング用プローブは標的結合ドメインを含んでいる。標的結合ドメインは長さが6個のヌクレオチド（6量体）配列（b₁-b₂-b₃-b₄-b₅-b₆）を含んでいて、標的核酸内の6個の相補的なヌクレオチドに特異的にハイブリダイズする。この6量体配列はそれぞれの側に縮重塩基（N）が隣接している。このシーケンシング用プローブは、D-DNAステムと呼ぶ二本鎖DNAスペーサを介して標的結合ドメインに共有結合しているバーコードドメインも含んでいる。D-DNAステムは長さが25個のヌクレオチドである。バーコードドメインは長さが74個のヌクレオチドであり、3つの付着領域（スポット1、スポット2、スポット3と呼ぶ）を含んでいる。それぞれの付着位置は8個のヌクレオチドを含んでいる。スポット1とスポット2は、ヌクレオチド25個の長さのポケットオリゴヌクレオチドにハイブリダイズしたバ

ーコードドメインからなる二本鎖DNA領域によって隔てられている。スポット2とスポット3も、ヌクレオチド25個の長さのポケットオリゴヌクレオチドにハイブリダイズしたバーコードドメインからなる二本鎖DNA領域によって隔てられている。D-ポケットシークエンシング用プローブは全面的にD-DNAからなる。

【0562】

これらの実験のため、512通りの異なる種類のD-ポケット (D-pocket) シークエンシング用プローブの2つのプール (D-ポケットプール1とD-ポケットプール3) を設計した。異なる種類のそれぞれのシークエンシング用プローブは、標的結合ドメイン内に異なる6量体配列を含んでいた。

【0563】

これらの実験のため、D-ポケットシークエンシング用プローブのプールの中の一連のシークエンシング用プローブの標的結合ドメインは、LG介在シークエンシング用プローブのプールの中の一連のシークエンシング用プローブの標的結合ドメインと同じであり、レポータ付着位置1、2、3それぞれに割り当てられる色は、D-ポケットシークエンシング用プローブとLG介在シークエンシング用プローブで同じであった。言い換えると、それぞれのD-ポケットシークエンシング用プローブについて、同じ標的結合ドメイン配列と対応する3色レポータ読み出しを有する1つの対応するLG介在シークエンシング用プローブが存在していた。

【0564】

LG介在シークエンシング用プローブとD-ポケットシークエンシング用プローブの効率と精度を調べるため、合成標的核酸を設計した。5'末端から開始すると、合成標的核酸は一本鎖シークエンシングウインドウを含んでおり、その長さは25~45個のヌクレオチドであった。これらシークエンシングウインドウは、合成ヌクレオチド配列と、既知のヒト遺伝子 (P53、ALX1、SMO1-1など) からのヌクレオチド配列の両方を含んでいた。シークエンシングウインドウの後には長さが35個のヌクレオチドのユニバーサル配列 (ここでは35-8配列と呼ぶ) が続く。35-8配列の後には3'ビオチン-TEG修飾が続く。この合成核酸は全面的にD-DNAからなる。合成標的核酸の模式図が図55に示されている。

【0565】

続く実験のため合成標的核酸を捕捉して固定化することを目的として、長さがヌクレオチド70個の捕捉用プローブを使用した。この捕捉用プローブは、35-8配列と相補的な配列を含んでいたため、合成標的核酸にハイブリダイズすることができた。35-8配列と相補的な配列は、全面的にD-DNAで構成されていた。この捕捉用プローブは、長さがヌクレオチド35個のポリ-L-dA配列も含んでいた。図55に示されているように、捕捉用プローブを合成標的核酸にハイブリダイズさせた後、捕捉用プローブ-標的核酸複合体を、ポリ-L-dA配列を含むローンオリゴヌクレオチドを含む表面に固定化することができる。それに加え、ローンオリゴヌクレオチドはビオチン分子を含むことができる。したがって図55に示されているように、捕捉用プローブのポリ-L-dA配列がローンオリゴヌクレオチドのポリ-L-dA配列にハイブリダイズした後、ニュートラアビジンを添加して合成標的核酸とローンオリゴヌクレオチドの間にタンパク質ロックを作り出すことができる。なぜならニュートラアビジンは、合成標的核酸の3'ビオチンと、ローンオリゴヌクレオチド上に位置するビオチンの両方に結合するからである。

【0566】

以下の実験では、合成標的核酸を捕捉用プローブにハイブリダイズさせ、規則正しいアレイに固定化した。規則正しいアレイは、ローンオリゴヌクレオチドで機能化した直径200 nmのパッドを含有していた。規則正しいアレイのピッチは800 nmであった。固定化された捕捉用プローブ-標的核酸複合体の上に0.2 mg/mlのニュートラアビジン溶液をピペットで添加することによって合成標的核酸とローンオリゴヌクレオチドの間のタンパク質ロックを実施した後、5分間インキュベートした。その後、固定化されてタンパク質ロックされた標的核酸を1 μ Mのビオチニル化されたPEG溶液の中で1分間インキュベートすることにより、空いているニュートラアビジン結合部位を1 μ Mの二重にビオチニル化された

10

20

30

40

50

PEGでブロックした。最後に、低塩（ $0.0033\times\text{SSPE}$ ）洗浄液を用い、タンパク質ロックされていない標的核酸をすべて、規則正しいアレイから除去した。

【0567】

標的核酸の固定化とタンパク質ロックの後、シーケンシング用プローブのプールを、シーケンシング用プローブ1つ当たり200 nMの濃度（ $3.75\times\text{SSPE}$ と15%ホルムアミド緩衝液の中）で8℃にて1分間アレイの上に流し、シーケンシング用プローブを標的核酸にハイブリダイズさせた。シーケンシング用プローブは、標的核酸にハイブリダイズした後、 $8.75\times\text{SSPE}$ という緩衝液条件下での連続した3ラウンドのハイブリダイゼーションによってレポータープローブにハイブリダイズさせた。これらの実験におけるレポータープローブは、レポータープローブのうちで付着領域に結合する部分と、検出可能な標識を含む部分の間にUVで切断可能なリンカーを含んでいた。

10

【0568】

第1ラウンドでは、第1のレポータープローブをシーケンシング用プローブのスポット1にハイブリダイズさせた。スポット1にハイブリダイズした第1のレポータープローブの検出可能な標識を同定した後、サンプルをUV光に曝露して第1のレポータープローブを切断することにより検出可能な標識を除去した。第2ラウンドでは、第2のレポータープローブをシーケンシング用プローブのスポット2にハイブリダイズさせた。スポット2にハイブリダイズした第2のレポータープローブの検出可能な標識を同定した後、サンプルをUV光に曝露して第2のレポータープローブを切断することにより検出可能な標識を除去した。最後の第3ラウンドでは、第3のレポータープローブをシーケンシング用プローブのスポット3にハイブリダイズさせた。スポット3にハイブリダイズした第3のレポータープローブの検出可能な標識を同定した後、低塩洗浄液（ $0.0033\times\text{SSPE}$ ）を用い、固定化された合成標的核酸にハイブリダイズしているシーケンシング用プローブを外した。脱ハイブリダイゼーションの後、シーケンシング用プローブの新たなプールを、固定化された合成標的核酸にハイブリダイズさせることができる。これらの工程が、シーケンシングの1つのサイクルを構成する。

20

【0569】

第1セットの実験では、LG介在プールを用い、固定化されたいくつかの異なる合成標的核酸を50サイクルのシーケンシング操作でシーケンシングした。第2セットの実験では、D-ポケットプールを用い、固定化されたいくつかの異なる合成標的核酸を50サイクルのシーケンシング操作でシーケンシングした。これら2セットの実験の結果を表12に示す。

30

【0570】

40

50

【表 1 3】

表12：シーケンシングの結果

プール	バーコード 効率 (%)	クリーン3 -スポット	クリーン2 -スポット	クリーン1 -スポット			
LG 介在 (2044)	41.1	12.3	10.4	13.9			
D-ポッケ ット (2073)	32.0	10.1	10.0	10.2			
プール	クリーンス ポット1 効 率	クリーンス ポット2 効 率	クリーンス ポット3 効 率	クリーンな 3-スポット 読み出し (%)	多色 暗化 (%)	有効率 (%)	無効率 (%)
LG 介在 (2044)	76.4	70.0	65.4	29.8	11.2	86.8	13.2
D-ポッケ ット (2073)	74.0	69.2	63.6	31.5	5.5	91.7	8.4

生存カットオフと票決

プール	バーコード 効率 (%)	クリーン3 -スポット	クリーン2 -スポット	クリーン1 -スポット			
LG 介在 (2044)	46.9	15.2	11.8	14.4			
D-ポッケ ット (2073)	29.3	10.5	8.7	8.7			
プール	クリーンス ポット1 の 効率	クリーンス ポット2 の 効率	クリーンス ポット3 の 効率	クリーンな 3-スポット 読み出し (%)	多色暗 化 (%)	有効率 (%)	無効率 (%)
LG 介在 (2044)	82.3	71.7	65.4	32.5	11.6	86.9	13.1
D-ポッケ ット (2073)	75.6	69.9	66.8	35.9	4.6	93.5	6.5

【0 5 7 1】

表12は、LG 介在プールがD-ポケットプールと比べて大きなバーコード効率を示したことを示している（47%対30%）。さらに、LG 介在プールはD-ポケットプールと比べて大きなクリーンスポット1効率も示した（82%対76%）。クリーンスポット2効率とクリーンスポット3効率は、LG 介在プールとD-ポケットプールの両方でほぼ同じであった。

【0 5 7 2】

さらに、図56～図58は、LG 介在プールとD-ポケットプールが予想される問題の標的領域でカバレッジの増大を示したことも示している。図56～図58では、最下部のx軸は、シーケンシングされつつある標的核酸の具体的なヌクレオチド／位置を示している。最上部のグラフは、LG 介在シーケンシング実験とD-ポケットシーケンシング実験での1塩基当たりの多様性の理論値と観測値を示している。1塩基のシーケンシング多様性は、

標的核酸内のその特定のヌクレオチドにハイブリダイズすることのできる標的結合ドメインを含む明確に異なるプローブ種の数である。したがって1塩基のシーケンシング多様性は、1つの位置を調べる（例えばシーケンシングする、または同定する）のに使用できる異なるシーケンシング用プローブがどれくらい多いかの1つの指標である。理論に囚われなければ、個々の塩基のシーケンシング多様性が大きいほど、多重シーケンシングサイクルの間の予想カバレッジが大きくなる。というのも、各サイクルにおいてプローブがその塩基に結合してシーケンシングされる機会がより多く存在するからである。図56～図58の最上部のグラフは、LG介在シーケンシング実験とD-ポケットシーケンシング実験における各塩基のカバレッジの観測値も示している。赤い箱は、予想される問題領域を示している。LG介在シーケンシング用プローブは、これらの問題領域で、D-ポケットシーケンシング用プローブと比べて大きなカバレッジを示し、いくつかのケースでは、LG介在シーケンシング用プローブは、D-ポケットシーケンシング用プローブがカバレッジをまったく示さない領域でカバレッジを示した。さらに、図59～図61は、LG介在シーケンシング用プローブとD-ポケットシーケンシング用プローブを用いたより多くの実験からのシーケンシング結果を示している。図59～図61に示されているように、LG介在シーケンシング用プローブと用いるときには、D-ポケットシーケンシング用プローブを用いるときと比べて、シーケンシングカバレッジの観測値とシーケンシング多様性の観測値の両方が増大している。

【0573】

LG介在シーケンシング用プローブとD-ポケットシーケンシング用プローブの性能を比較したより多くの実験結果を図62に示してある。図62の各ヒストグラムのx軸は、実験で記録された観察されたバーコードイベント（すなわち何らかの形態の検出可能なシグナル）の数を表わす。ヒストグラムの黒い区画は、有効な3-スポット読み出しを具体的に表わしている（すなわちシーケンシング用プローブの3つの付着領域のそれぞれが有効であると同定された）。y軸は、特定の1つの標的に関する対応する数のバーコードイベントを示したアレイ上の凹凸（すなわち個々の固定化された標的核酸）の数を表わす。図62は、LG介在プールがD-ポケットと比べてバーコードイベントの有効な3-スポット読み出しが多いことを示している。

【0574】

LG介在シーケンシング用プローブとD-ポケットシーケンシング用プローブの性能を比較したより多くの実験結果を図63に示してある。合計して20タイプの異なる合成標的核酸を、LG介在シーケンシング用プローブまたはD-ポケットシーケンシング用プローブを用いてシーケンシングした。図63のグラフは、標的核酸ドメインの任意の1つの塩基位置 b_1 および／または b_2 および／または b_3 および／または b_4 および／または b_5 および／または b_6 におけるオンターゲットイベント、無効イベント、オフターゲット、1/6ミスマッチ、2/6ミスマッチ、3/6ミスマッチ、4/6ミスマッチ、5/6ミスマッチ、6/6ミスマッチの合計数を示している。図63からわかるように、LG介在シーケンシング用プローブは、オフターゲットイベントに対するオンターゲットイベントの比率と、無効イベントに対するオンターゲットイベントの比率が、D-ポケットシーケンシング用プローブと比べて約2倍であった。図64からわかるように、LG介在シーケンシング用プローブのこの向上した性能は、異なる多彩な特異的合成標的核酸をシーケンシングするときと整合している。

【0575】

D-ポケットシーケンシング用プローブとLG介在シーケンシング用プローブを直接比較するため、D-ポケットシーケンシング用プローブとLG介在シーケンシング用プローブの組み合わせを用いて合成標的核酸をシーケンシングした。D-ポケットプール（D-ポケットプール1）を最初の50回のシーケンシングサイクルで使用した。次に、同じ固定化された合成標的核酸を、LG介在プールを用いたその後の50回のサイクルでシーケンシングした。次いで、D-ポケットシーケンシング用プローブを用いた最初の50回のサイクルからの結果と、LG介在シーケンシング用プローブを用いた最後の50回のサイクル

からの結果を比較した。これらの結果を表13に示す。

【0576】

【表14】

表13：シーケンシングの結果

プール	バーコード 効率 (%)	クリーン3 -スポッタ	クリーン2 -スポッタ	クリーン1 -スポッタ			
D-ポケット (サイクル 1~50)	26.9	6.8	8.4	10.6			
LG 介在 (サイクル 51~100)	37.8	12.8	11.6	11.0			
プール	クリーンス ポット1 効 率	クリーンス ポット2 効 率	クリーンス ポット3 効 率	クリーン な3-スポ ット読み 出し (%)	多色暗 化 (%)	有効率 (%)	無効率 (%)
D-ポケット (サイクル 1~50)	70.0	62.6	58.0	25.4	4.2	87.3	12.7
LG 介在 (サイクル 51~100)	73.3	74.5	66.7	33.9	6.4	86.2	13.8

【0577】

表13は、LG 介在プールが、D-プールと比べてバーコード効率、クリーンスポット1の効率、クリーンスポット2の効率、クリーンスポット3の効率、クリーンな3-スポット読み出しが大きかったことを示している。

【0578】

さらに、図65は、1サイクル当たりの1スポッタイベント（可能な3つのレポータプローブのうちの1つだけがうまく記録される）、2スポッタイベント（可能な3つのレポータプローブのうちの2つだけがうまく記録される）、3スポッタイベント（可能な3つのレポータプローブのすべてがうまく記録される）の数を示している。図65からわかるように、1スポッタイベント、2スポッタイベント、3スポッタイベントの数はすべて、（LG 介在シーケンシングプローブへの切り換えに対応する）サイクル51以降に増加している。特に、3スポッタイベントの数で最大の増加が起こる。

【0579】

図66は、上記の実験の各サイクルにおけるオンターゲットイベント、新たな6量体イベント、冗長な6量体イベント、オフターゲットイベント、無効イベントの数を示している。オンターゲットイベントは、シーケンシング用プローブが標的核酸に正しく結合し、3つのレポータプローブがすべてうまく正確に同定されたイベントである。新たな6量体イベントは、以前のサイクルでは同定されていなかった新たな6量体が1個の固定化された標的核酸の中で同定／シーケンシングされるイベントである。冗長な6量体イベントは、特定の6量体が1個の固定化された標的核酸の中で再度同定されるイベントである。オフターゲットイベントと無効イベントは、シーケンシング用プローブとレポータプローブのハイブリダイゼーションおよび／または検出でエラーが起こるイベントである。図66の最も左側の図に示されているように、オンターゲットイベント、新たな6量体イベント、冗長な6量体イベントの数はすべて、（LG 介在シーケンシングプローブへの切り換えに対応する）サイクル51以降に増加している。したがって新たな6量体イベントの数は、サイクル51~100の間は減少している。というのも、連続した各サイクルにおいて標的核酸の

中でまだ同定されていない6量体はより少なくなり、冗長な6量体イベントの数に対応して増加することが見られるからである。

【0580】

別のセットの実験では、1つのLG介在プールと2つのD-ポケットプール（D-ポケットプール1とD-ポケットプール3）を用い、固定化された合成標的核酸をシーケンシングした。LG介在プールは最初の50回のシーケンシングサイクルで使用した。次に、同じ固定化された合成標的核酸を、D-ポケットプール1とD-ポケットプール3を用いた100回のサイクルでシーケンシングした。これら2つのプールは交互に使用した。すなわちD-ポケットプール1をサイクル51で使用し、D-ポケットプール3をサイクル52で使用し、D-ポケットプール1をサイクル53で使用し、といった具合である。これら150回のシーケンシングサイクルからの結果を表14に示す。

【0581】

10

20

30

40

50

【表 15】

表14：シークエンシングの結果

プール	バーコード効率 (%)	クリーン 3-スポット	クリーン 2-スポット	クリーン 1-スポット			
LG 介在プール 1 (サイクル 1~50)	36.6	11.3	12.4	11.7			
D-ポケット プール 1 (51~150 の奇数 サイクル)	13.1	2.8	4.1	5.9			
D-ポケット プール 3 (51~150 の偶数 サイクル)	11.8	2.6	3.7	5.4			
プール	クリーン スポット 1 効率	クリーン スポット 2 効率	クリーン スポット 3 効率	クリーン な 3-スポット 読み出し (%)	多色暗化 (%)	有効率 (%)	無効率 (%)
LG 介在プール 1 (サイクル 1~50)	66.6	69.9	67.0	30.9	3.3	88.1	12.0
D-ポケット プール 1 (51~150 の奇数 サイクル)	64.2	58.4	55.0	21.4	1.8	89.5	10.6
D-ポケット プール 3 (51~150 の偶数 サイクル)	64.1	58.3	56.1	21.8	2.0	88.7	11.3

【0582】

表14に示した結果は、LG 介在プールが、D-ポケットプールと比べてバーコード効率、クリーンスポット1効率、クリーンスポット2効率、クリーンスポット3効率、クリーンな3-スポット読み出しが大きかったことを実証している。

【0583】

上記の結果を合わせて考えると、LG 介在シークエンシングプローブとD-ポケットシークエンシングプローブの両方を用いて1個の固定化された標的核酸を調べることができ、LG 介在シークエンシングプローブは、D-ポケットシークエンシングプローブと比べて大きなカバレッジ、効率、有効バーコード事象を示す。

【図面】

【図 1】

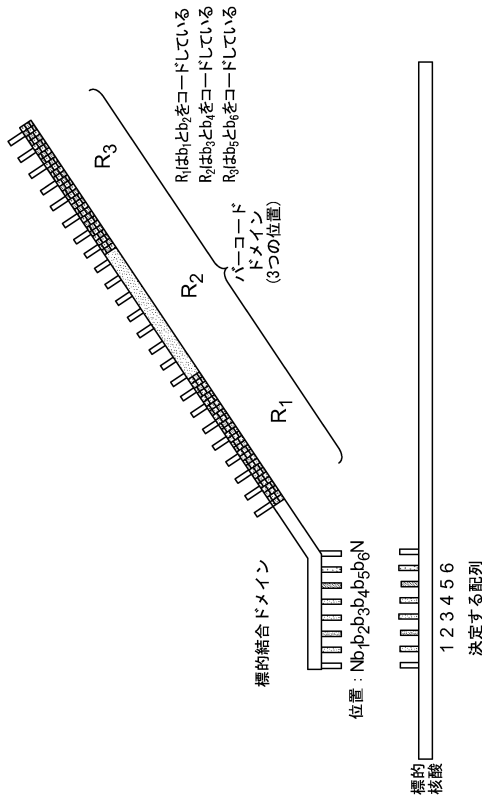


FIG. 1

【図 2】

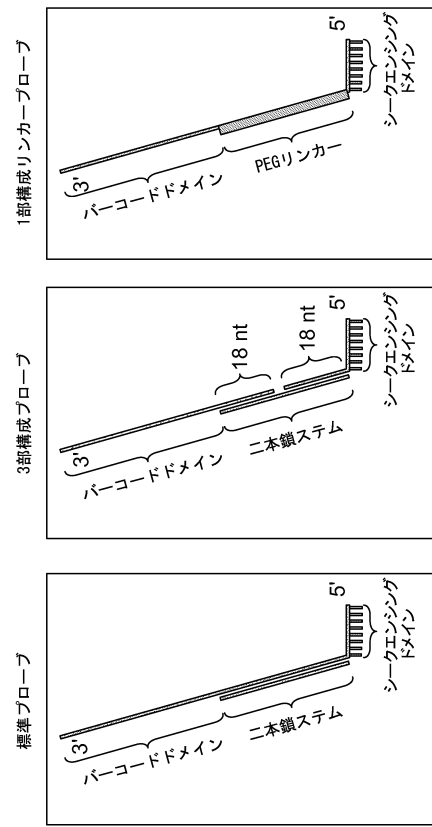


FIG. 2

【図 3】

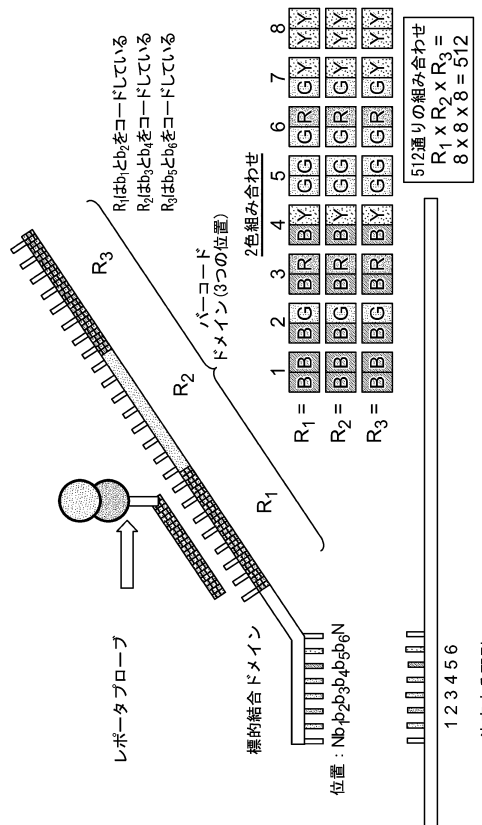


FIG. 3

【図 4】

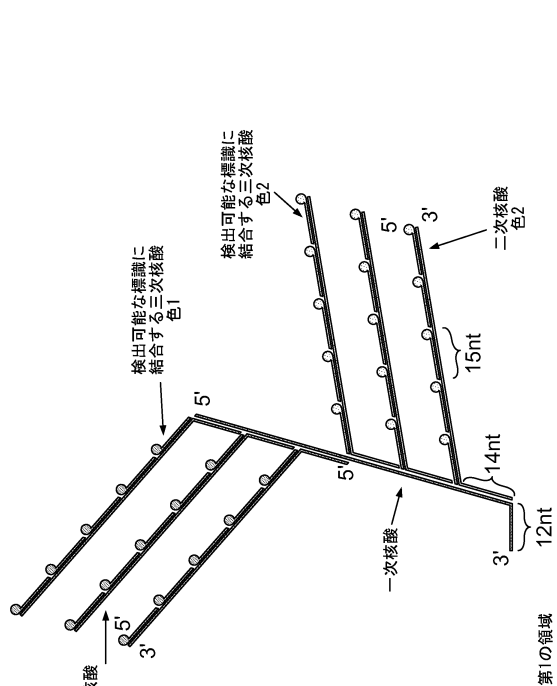


FIG. 4

10

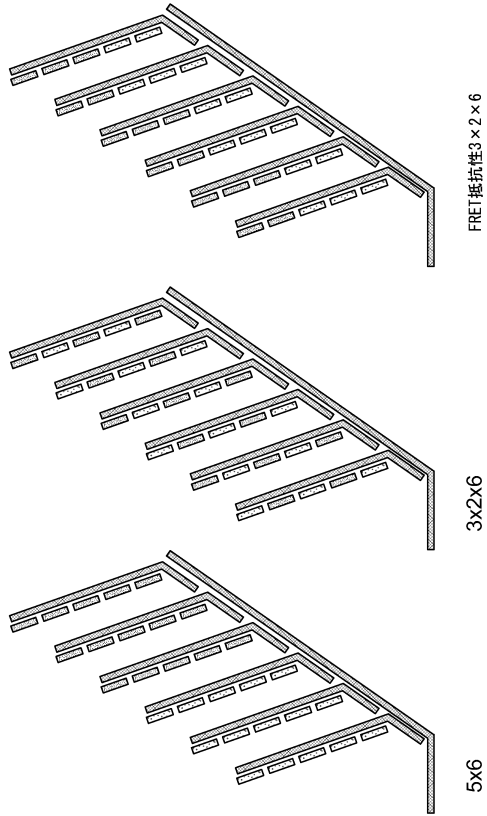
20

30

40

50

【図 5】



【図 6】

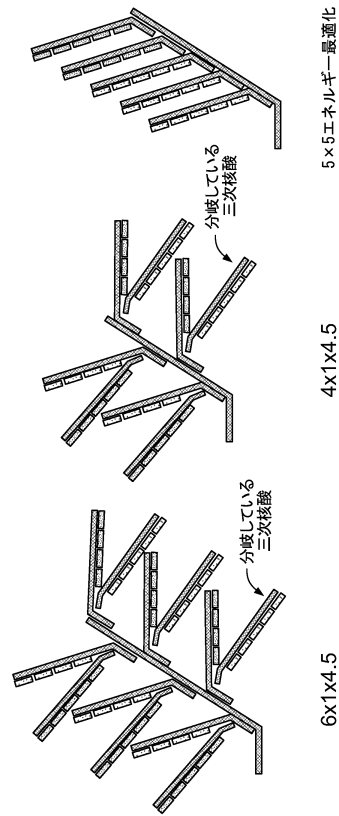


FIG. 6

10

20

【圖 7】

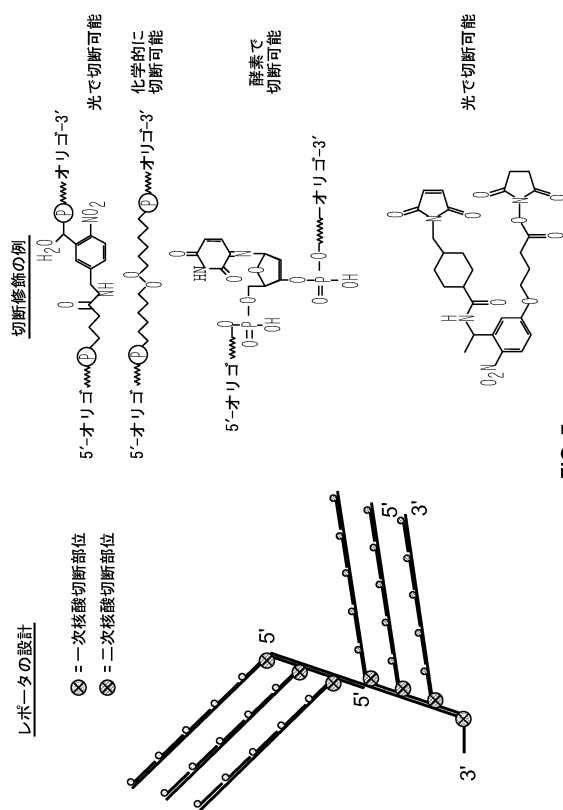


FIG. 7

【図 8】

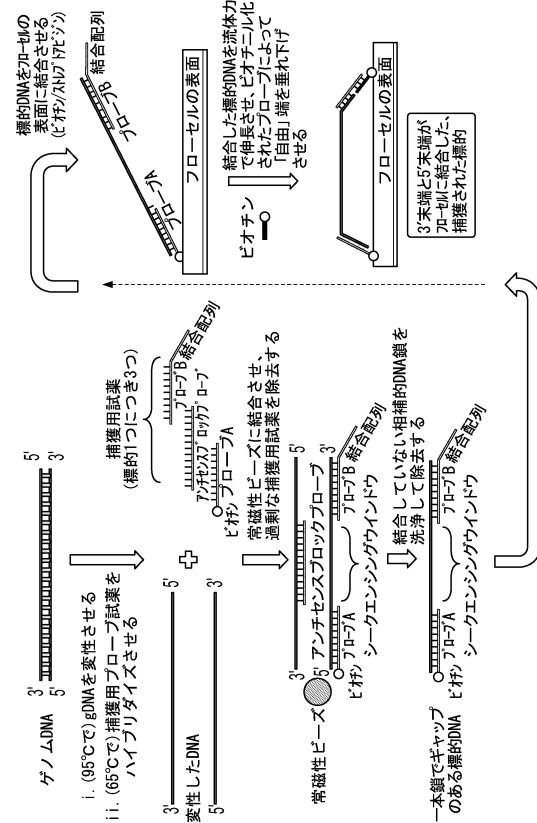


FIG. 8

30

40

50

【図 1 2】

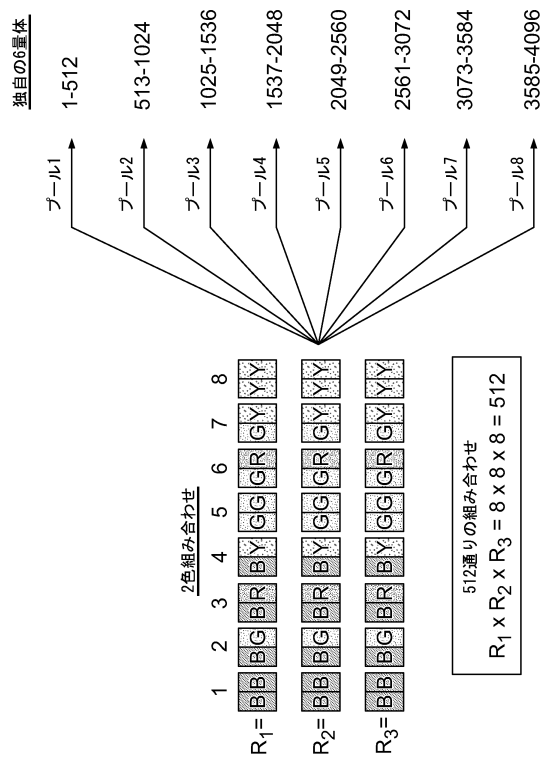


FIG. 12

【図 1 3】

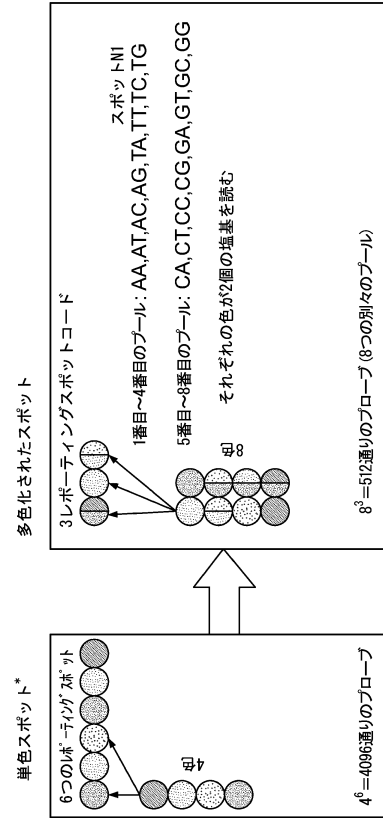


FIG. 13

*アメリカ合衆国特許出願公開第2016/0194701号に開示されている (その全体が参照によって本明細書に組み込まれている)

【図 1 4】

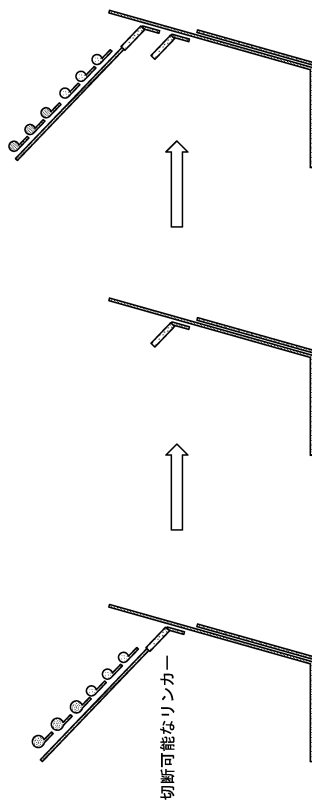


FIG. 14

【図 1 5】

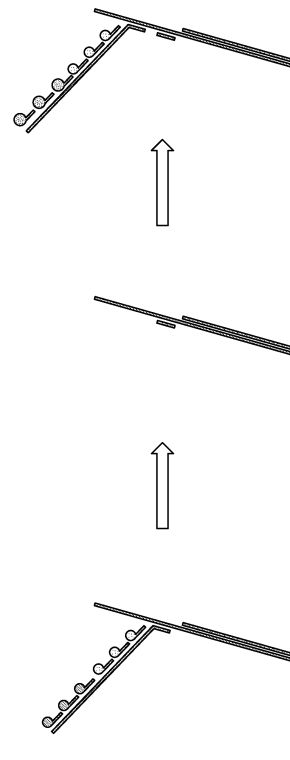


FIG. 15

10

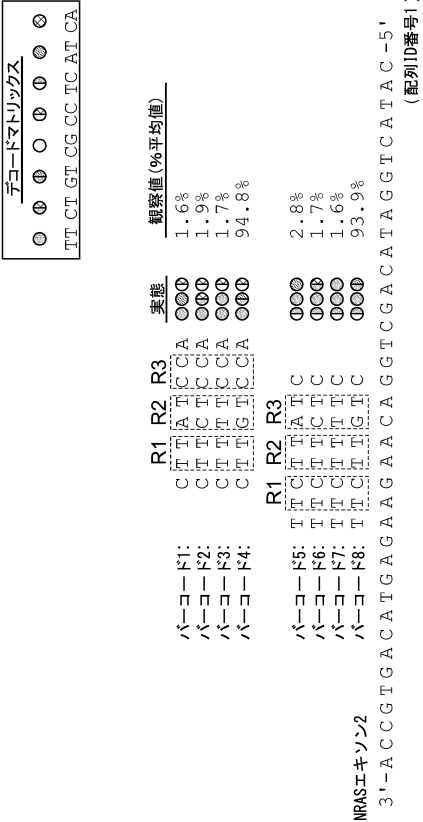
20

30

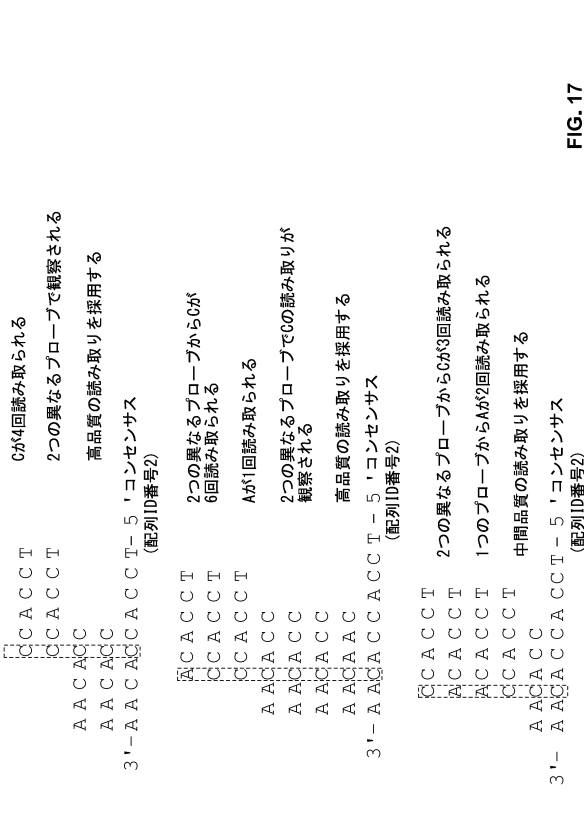
40

50

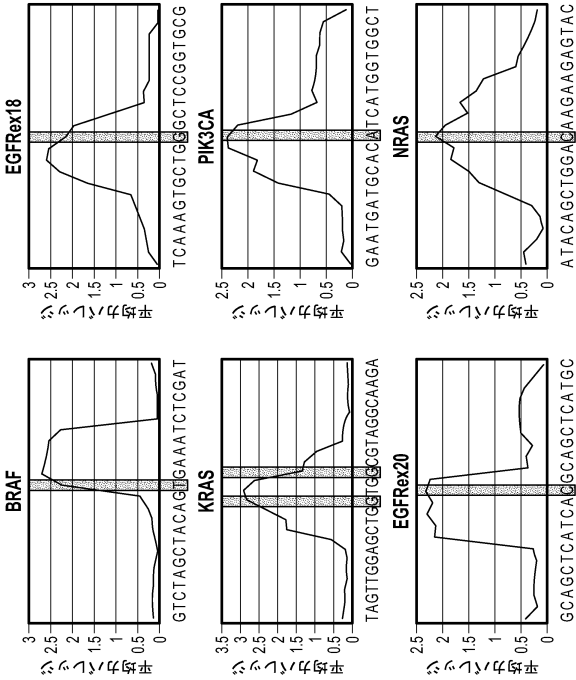
【図 16】



【図 17】



【図 18 - 1】



【図 18 - 2】

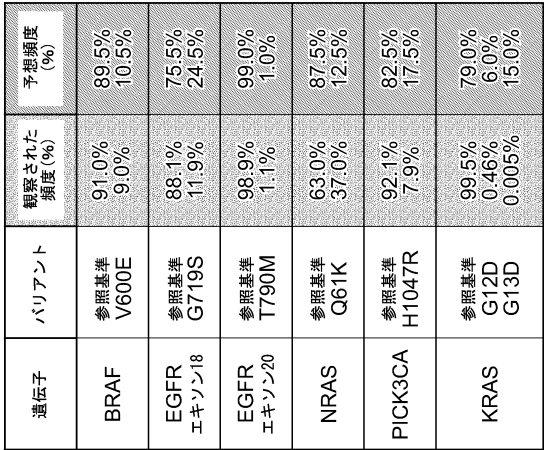


FIG. 18 (Cont.)

FIG. 17

【図 19】

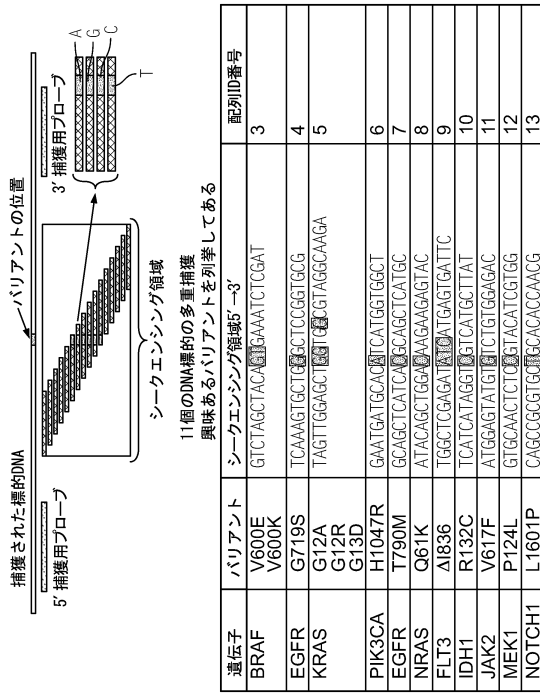
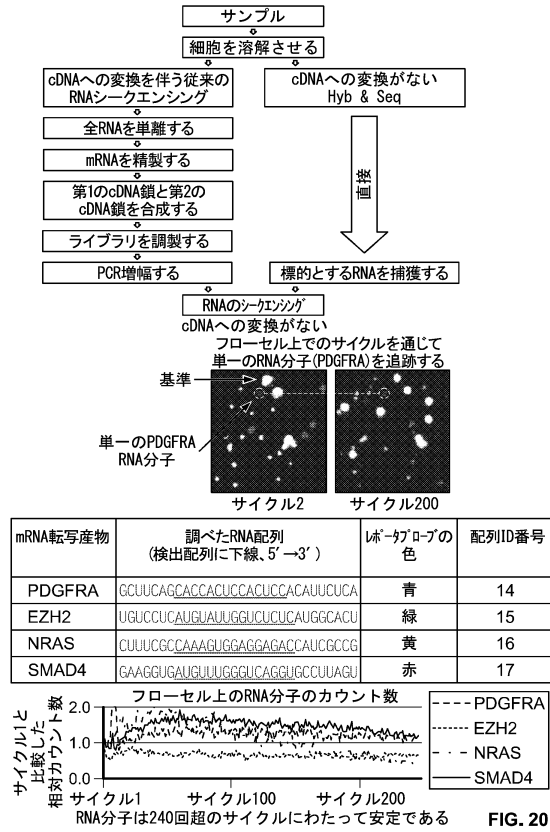


FIG. 19

【図 20】



【図 21】

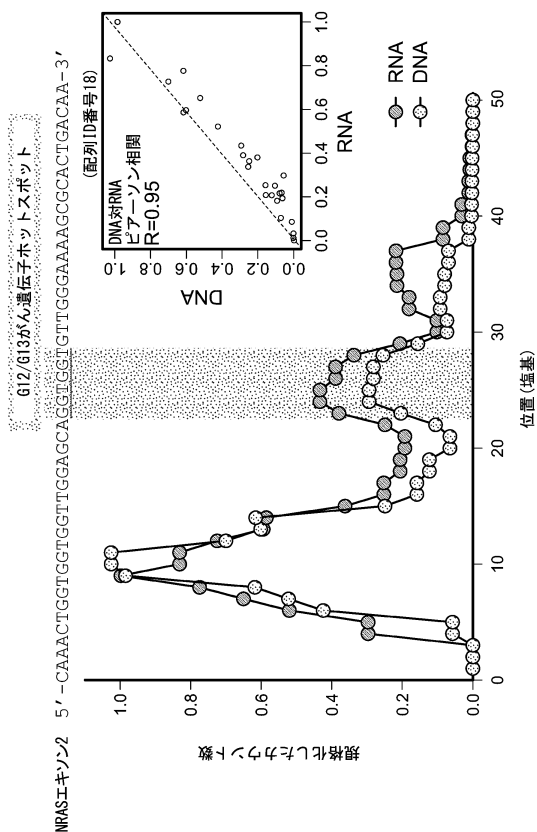
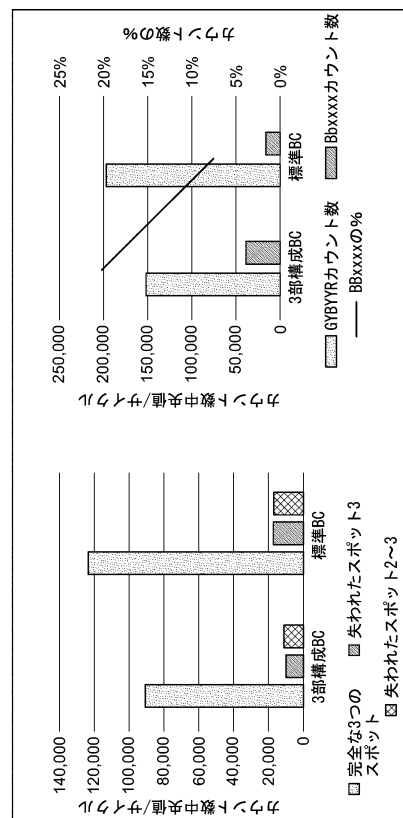


FIG. 21

【図 22】



10

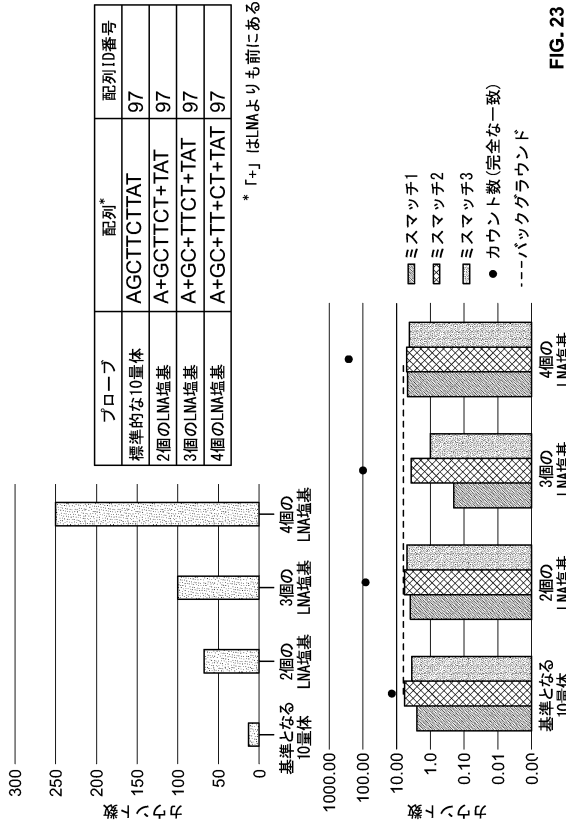
20

30

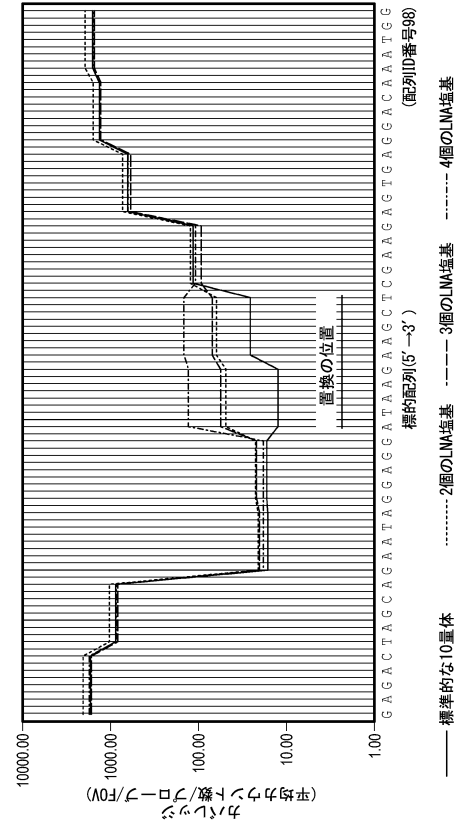
40

50

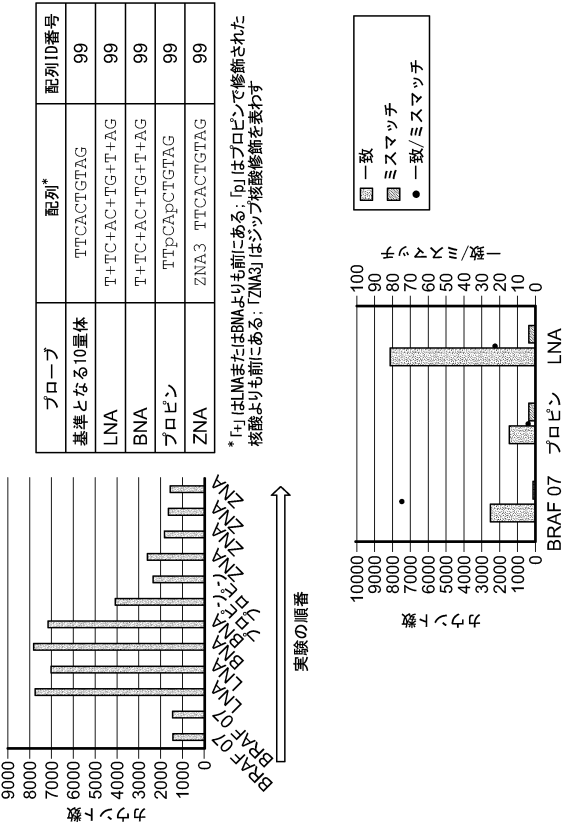
【図 2 3】



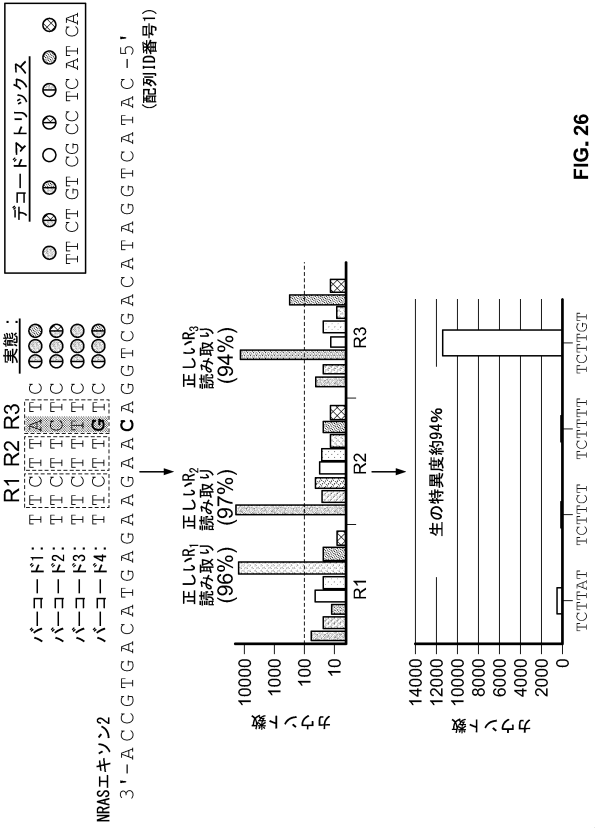
【図 2 4】



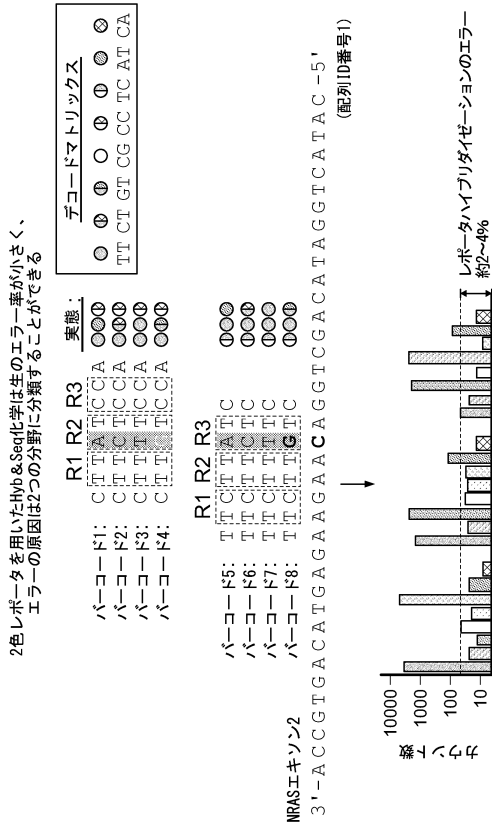
【図 2 5】



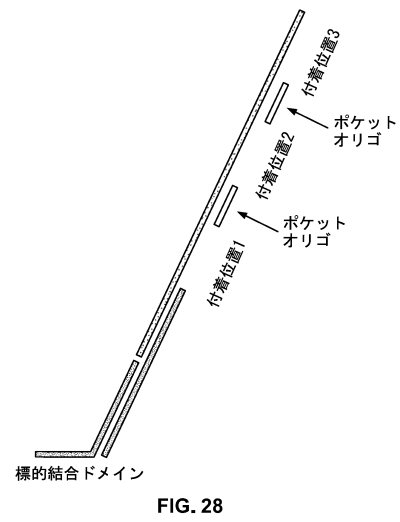
【図 2 6】



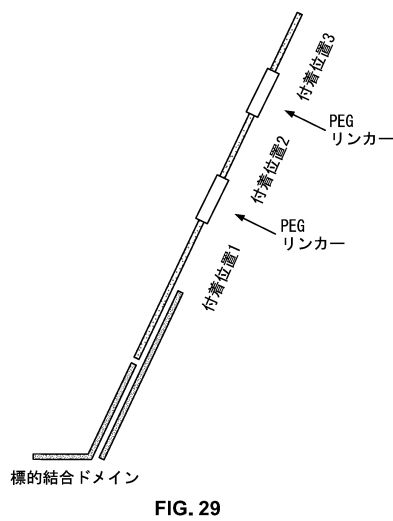
【図 27】



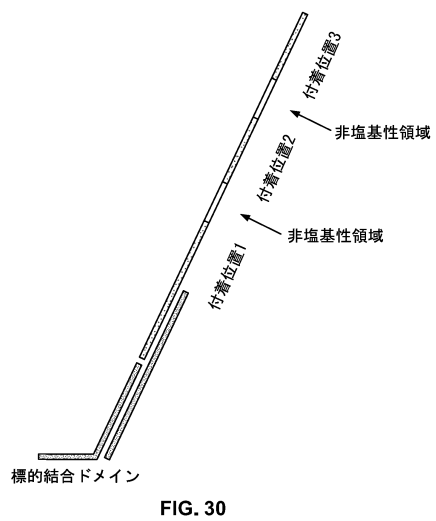
【図 28】



【図 29】



【図 30】



10

20

30

40

50

【図 39】

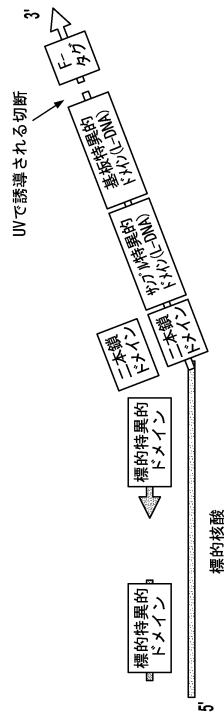


FIG. 39

【図 40】

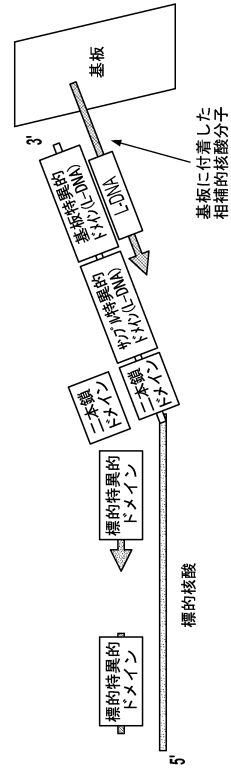


FIG. 40

【図 41】

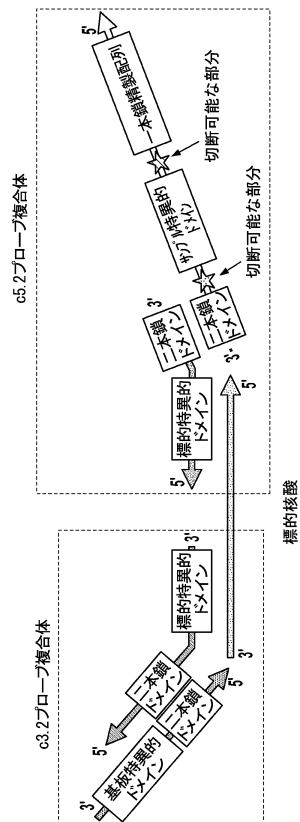


FIG. 41

【図 42】

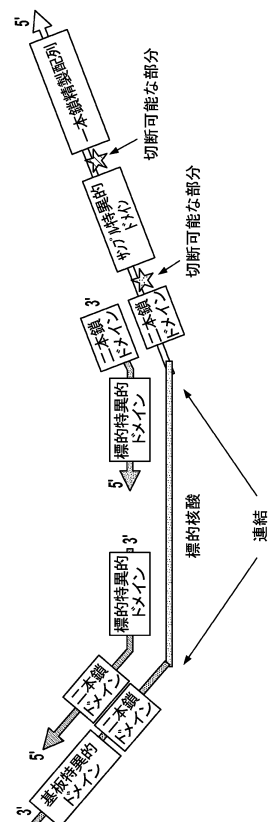


FIG. 42

10

20

30

40

50

【図 4 3】

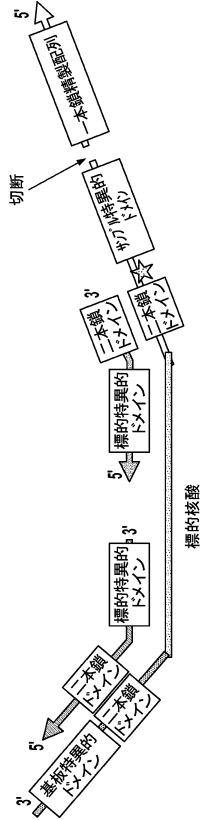


FIG. 43

【図 4 4】

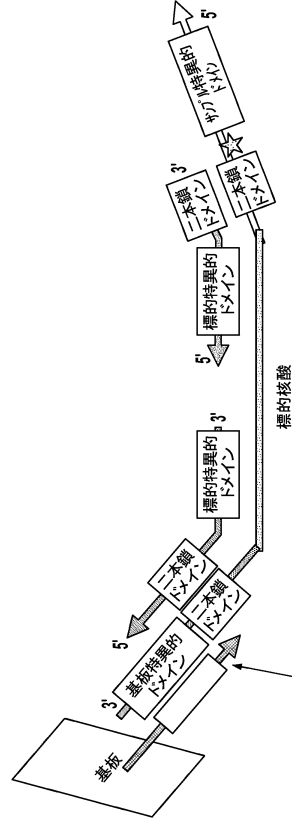


FIG. 44

【図 4 5】

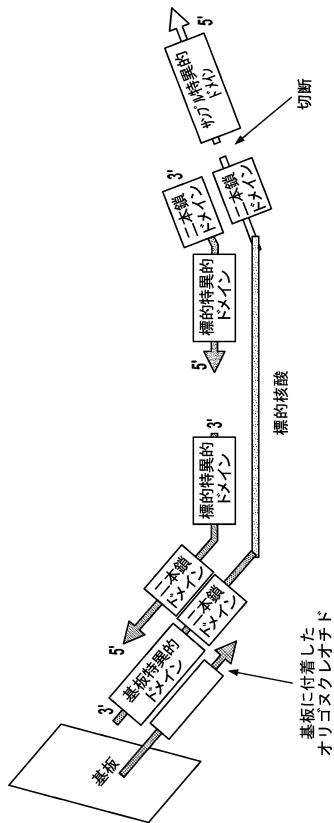


FIG. 45

【図 4 6】

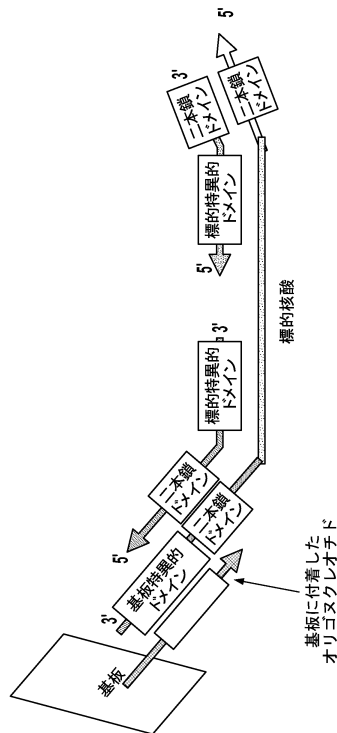


FIG. 46

10

20

30

40

50

【図 4 7】

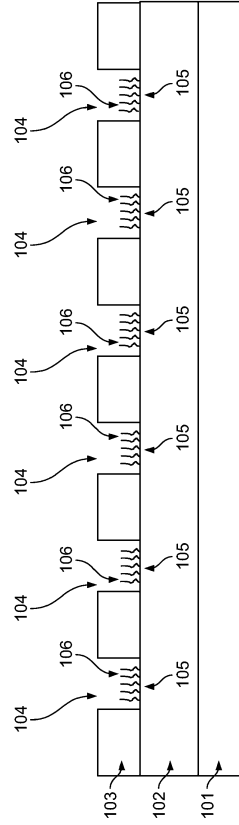


FIG. 47

【図 4 8】

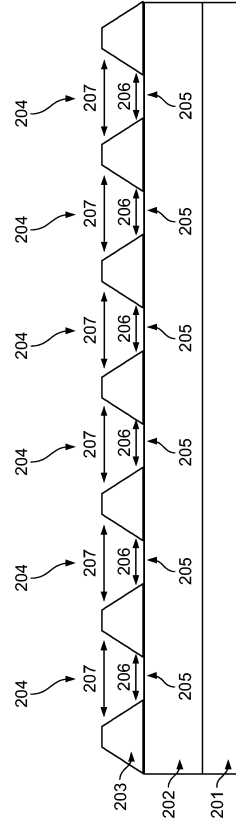


FIG. 48

【図 4 9】

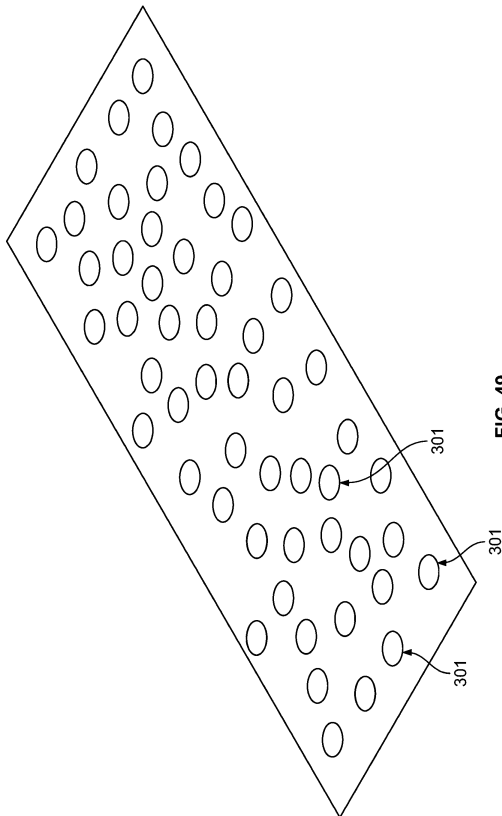


FIG. 49

【図 5 0】

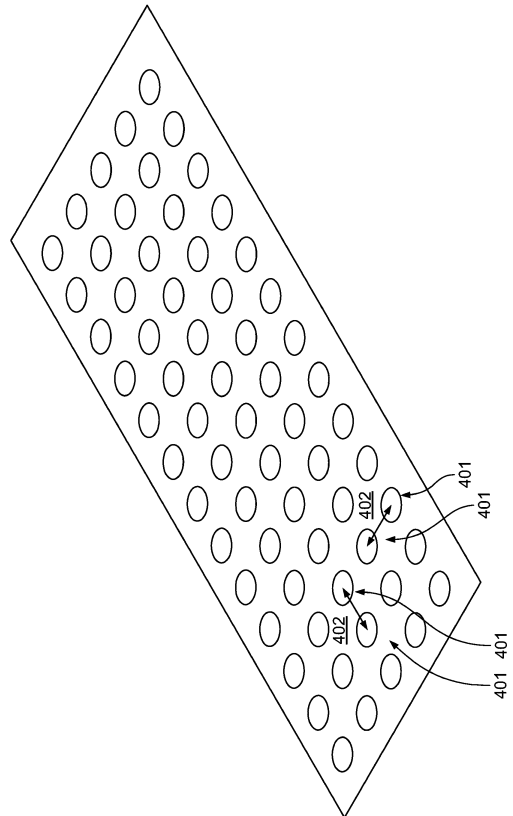


FIG. 50

10

20

30

40

50

【図 5 1】

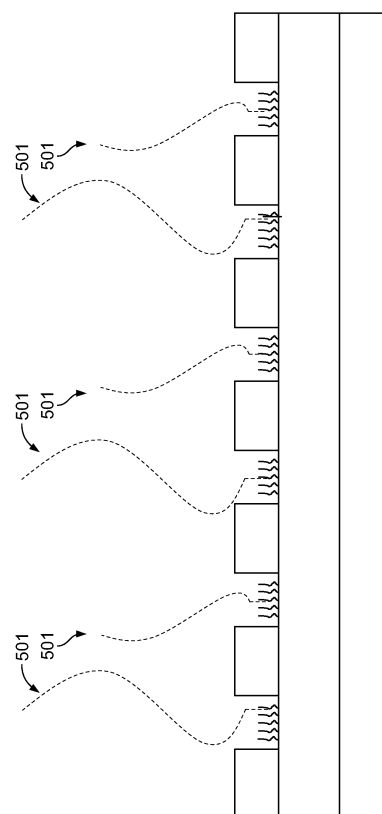


FIG. 51

【図 5 2】

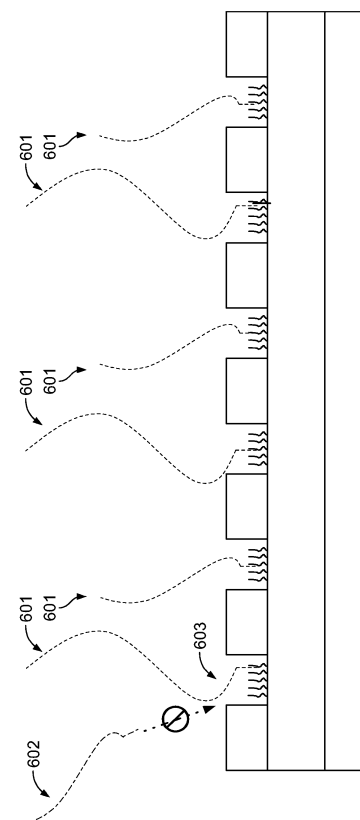


FIG. 52

【図 5 3】

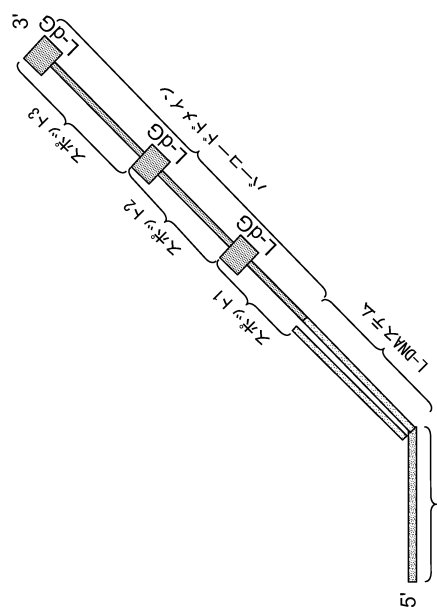


FIG. 53

【図 5 4】

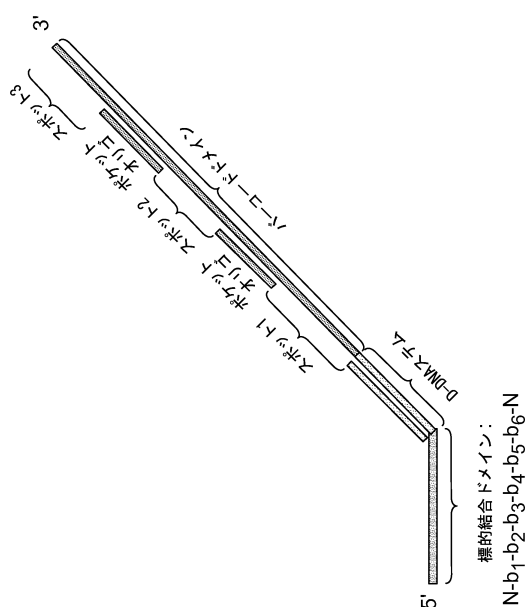


FIG. 54

【図 55】

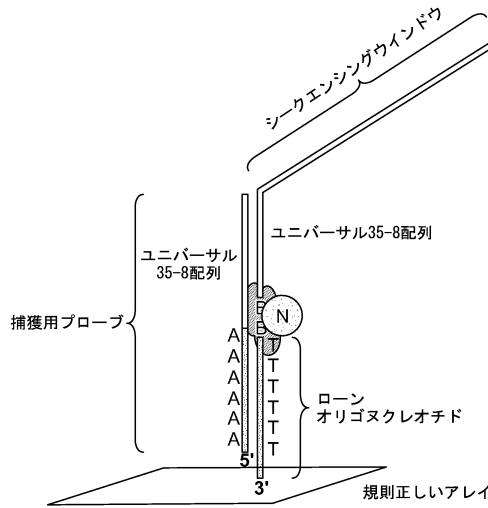


FIG. 55

【図 56】

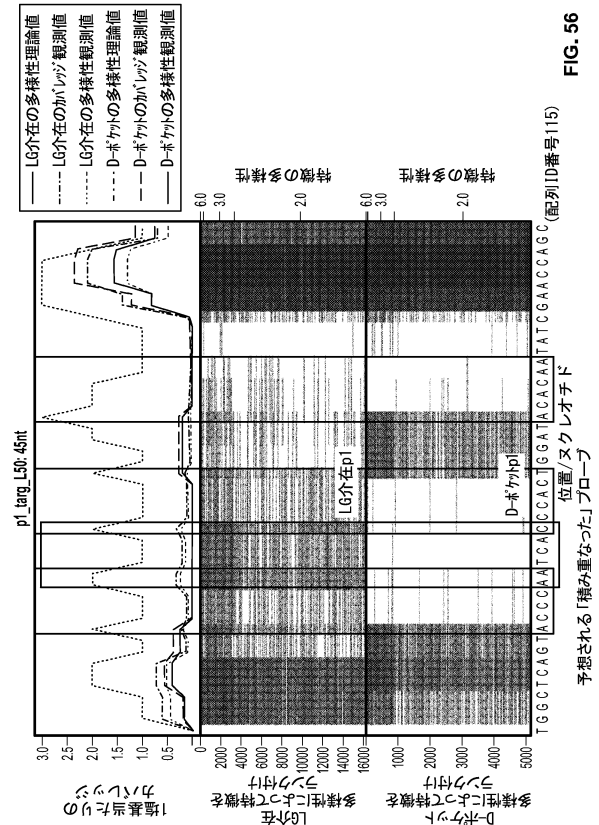


FIG. 56

【図 57】

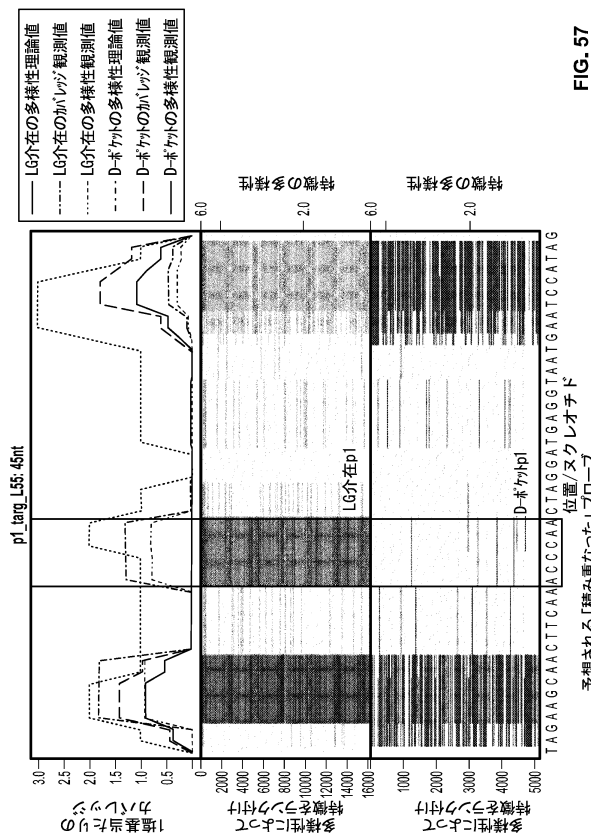


FIG. 57

【図 58】

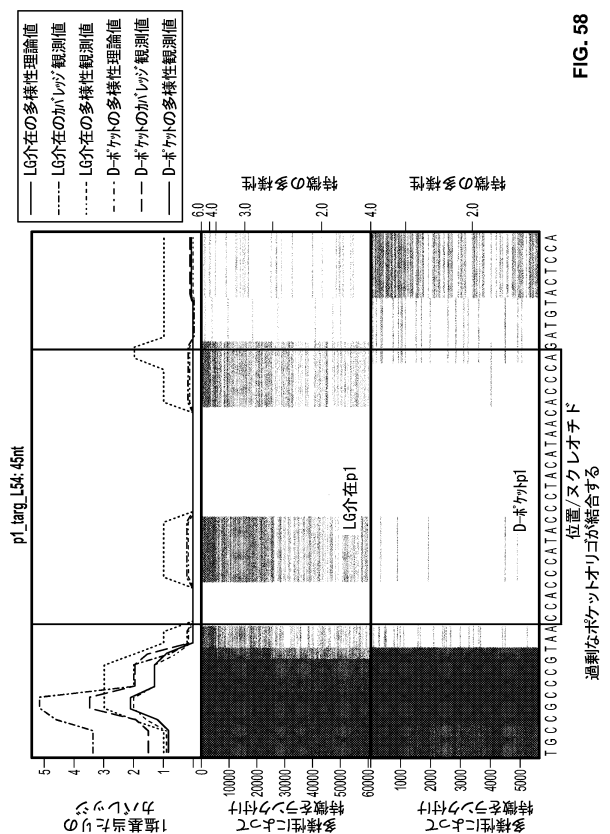


FIG. 58

10

20

30

40

50

【図 59】

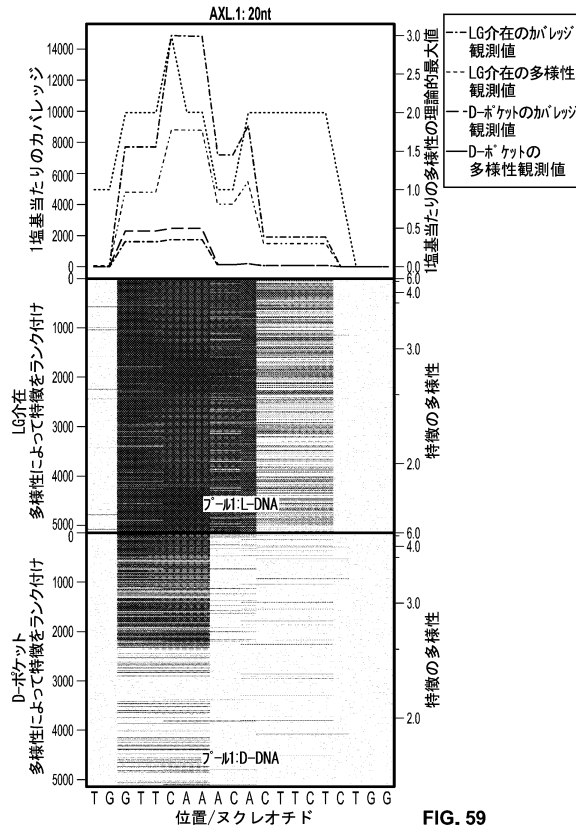


FIG. 59

【図 60】

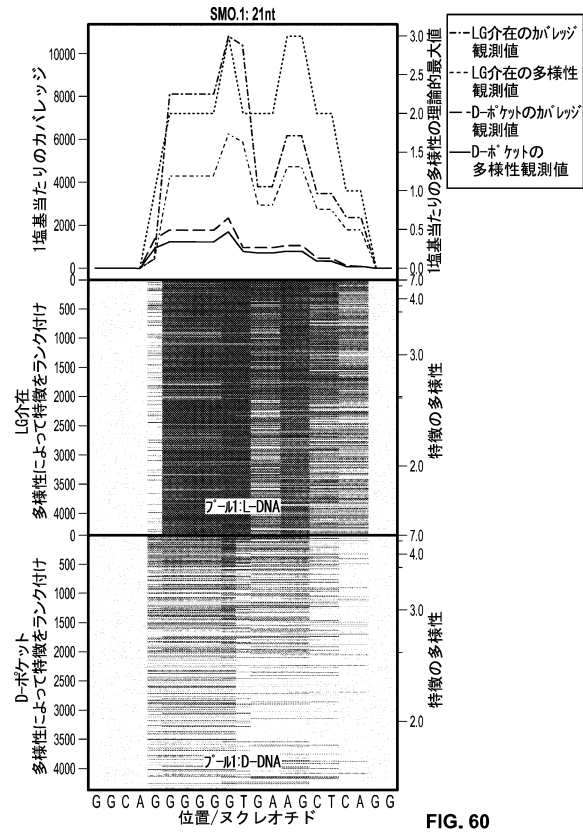


FIG. 60

【図 61】

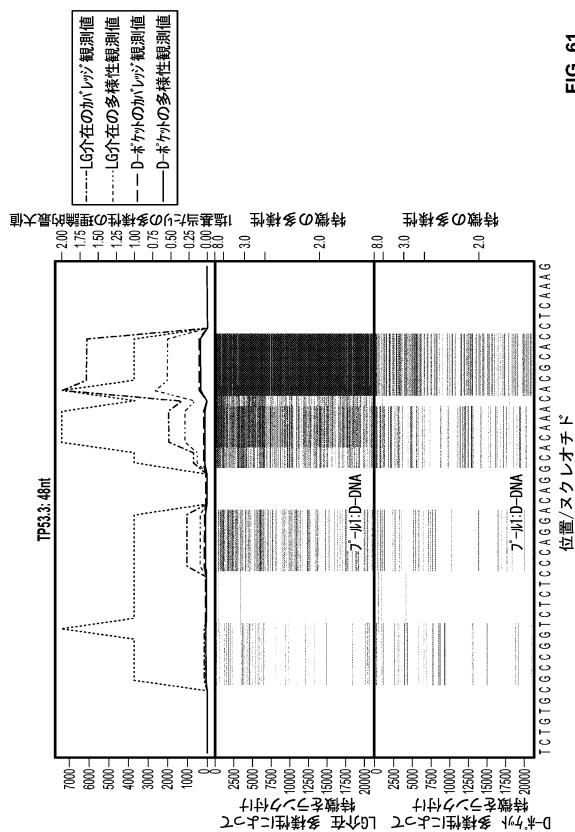


FIG. 61

【図 62】

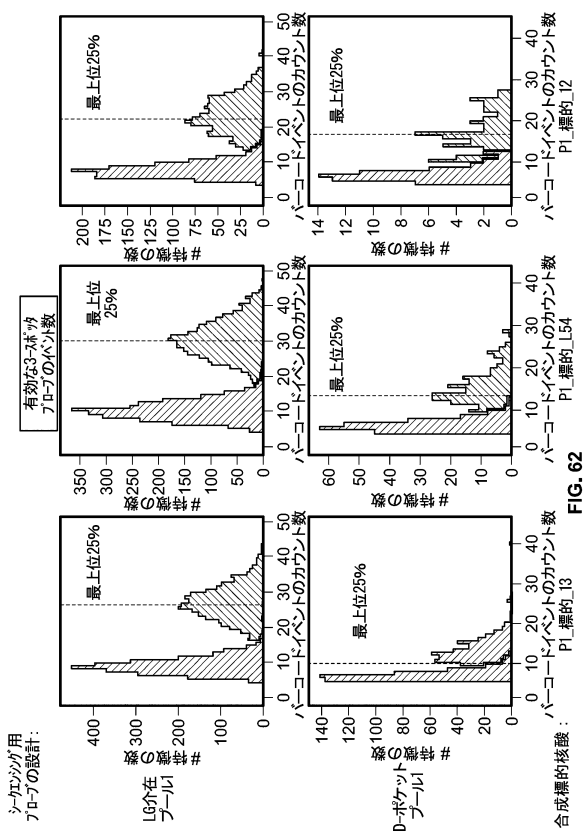


FIG. 62

10

20

30

40

50

【図 6 3】

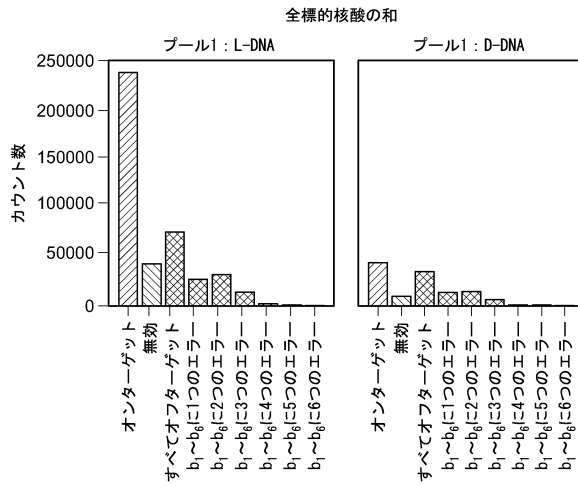


FIG. 63

【図 6 4 - 1】

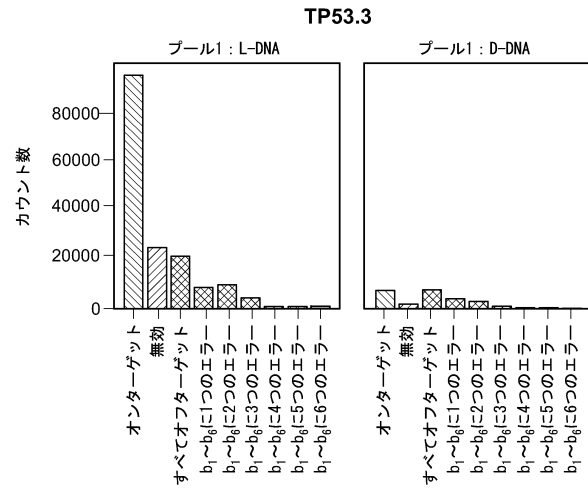


FIG. 64

【図 6 4 - 2】

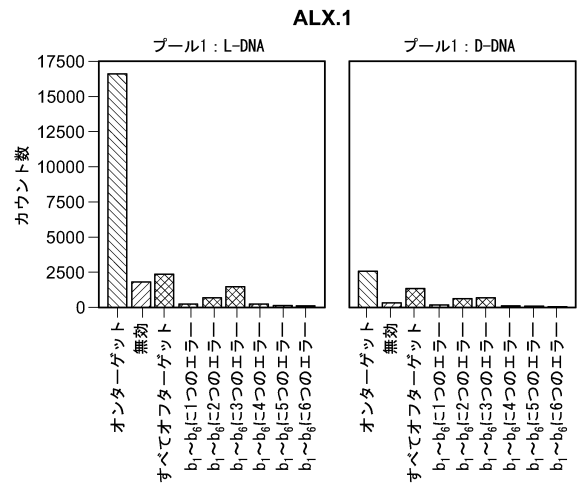


FIG. 64 (Cont.)

【図 6 4 - 3】

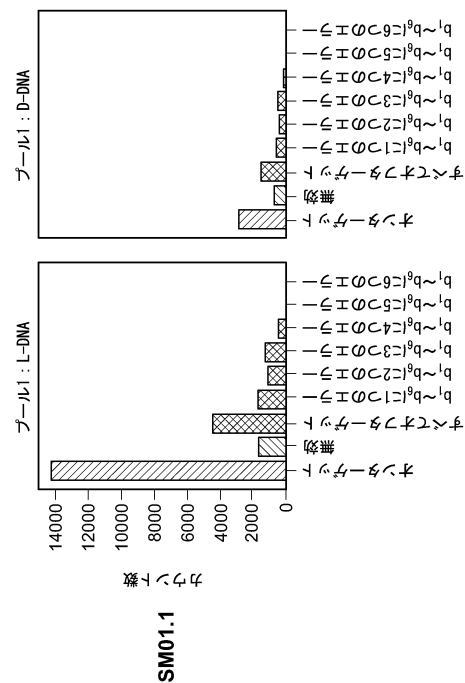


FIG. 64 (Cont.)

10

20

30

40

50

【図 6 5】

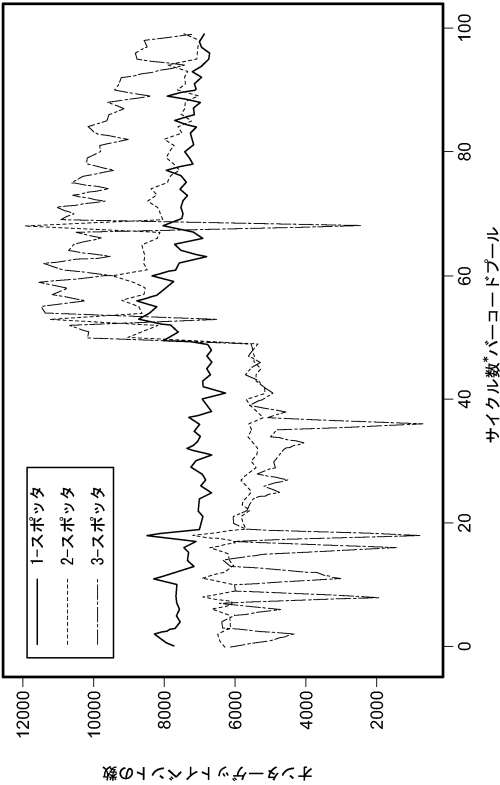


FIG. 65

【図 6 6 - 1】

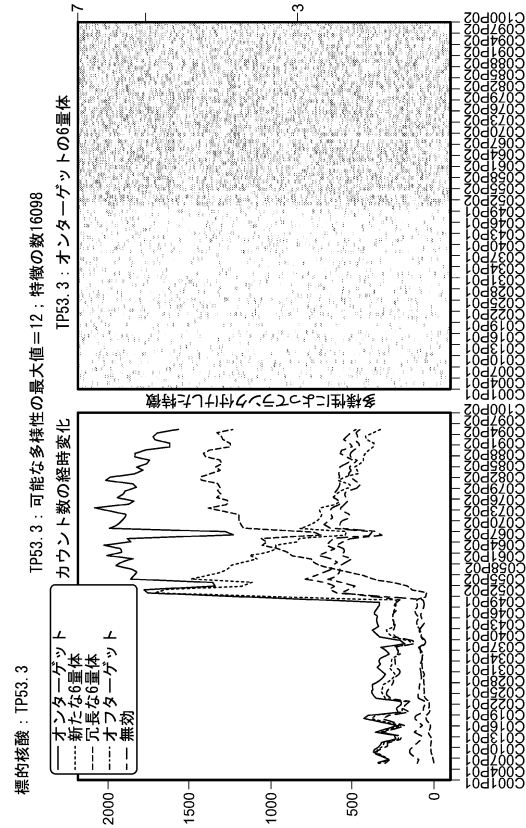


FIG. 66

【図 6 6 - 2】

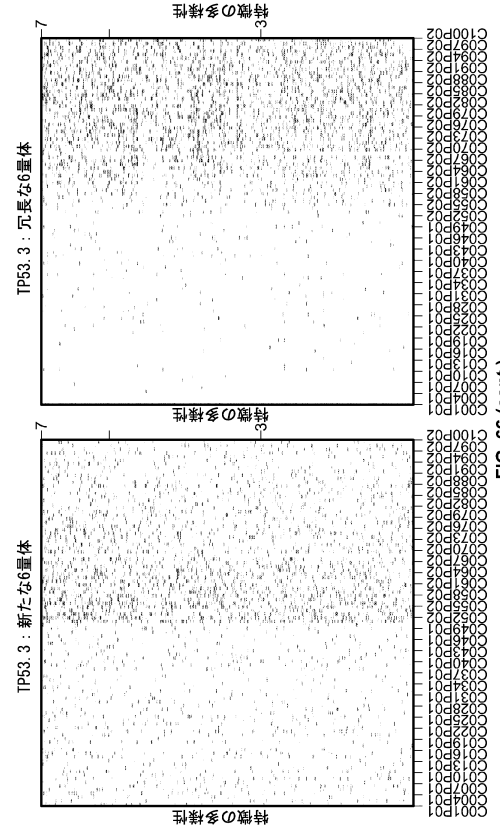


FIG. 66 (cont.)

【図 6 6 - 3】

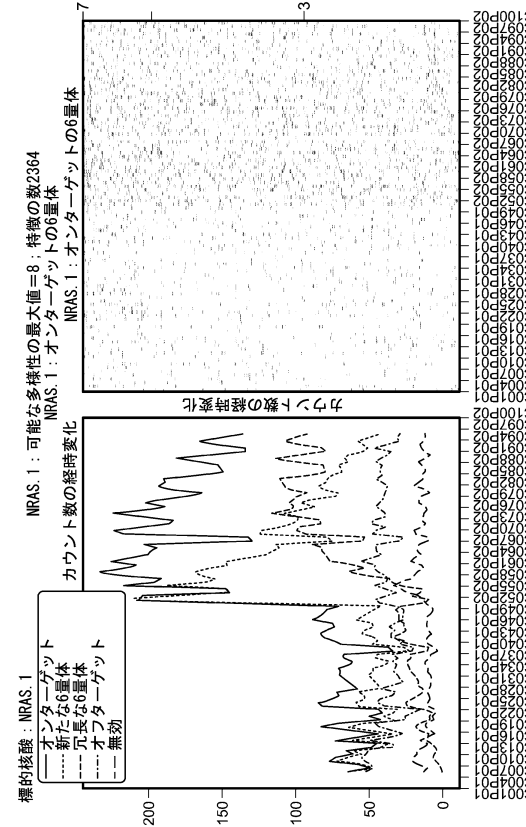
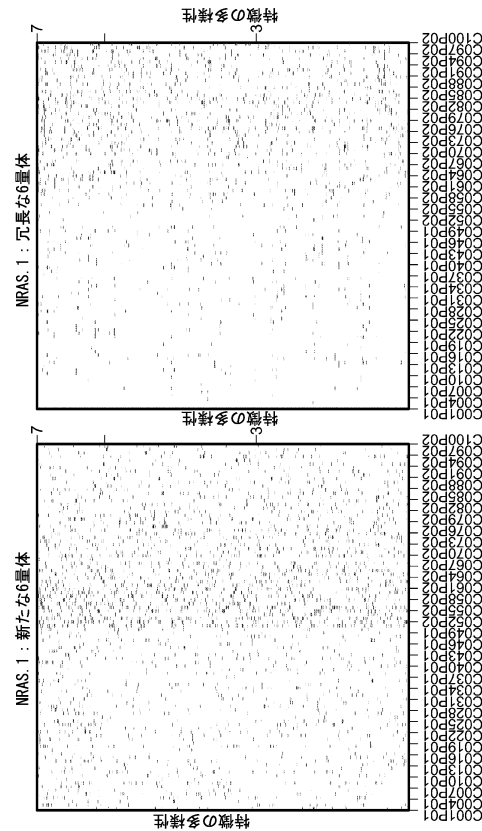
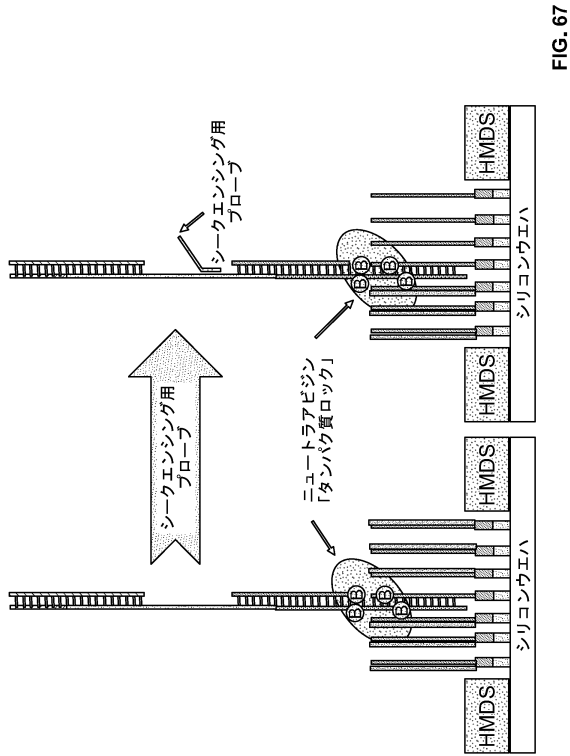


FIG. 66 (cont.)

【図 6 6 - 4】



【図 6 7】



【配列表】

0007525402000001.app

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

弁理士 武居 良太郎

(74)代理人 100203828

弁理士 喜多村 久美

(72)発明者 ドウェイン ダナウェイ

アメリカ合衆国, ワシントン 98115, シアトル, ノース イースト ナインティフォース ストリート 2751

(72)発明者 エリザベス マンラオ

アメリカ合衆国, ワシントン 98115, レイク フォレスト パーク, ノース イースト ワンハンドレッドナインティサード プレイス 5220

(72)発明者 ジョーセフ ビーチェム

アメリカ合衆国, オレゴン 97405, ユージーン, リッジライン ドライブ 46

(72)発明者 ルステム カフィゾフ

アメリカ合衆国, ワシントン 98109, シアトル, マーサー ストリート 1200, アpartment 212

(72)発明者 サンガミスラ コルコンダ

アメリカ合衆国, ワシントン 98122, シアトル, トウェンティフォース アベニュー 1716

(72)発明者 イ デン

アメリカ合衆国, ワシントン 98125, シアトル, サーティーセカンド アベニュー ノース イースト 14018, アpartment ディー

(72)発明者 テ キム

アメリカ合衆国, ワシントン 98006, ベルビュー, サウス イースト シックスティサード プレイス 16729

(72)発明者 マーク グレゴリー

アメリカ合衆国, ワシントン 98125, シアトル, トウェンティフォース アベニュー ノース イースト 14327

(72)発明者 マーガレット ホアン

アメリカ合衆国, ワシントン 98108, シアトル, サウス ネバダ ストリート 1111

(72)発明者 マシュー ウォルシュ

アメリカ合衆国, ワシントン 98102, シアトル, ハーバード アベニュー イースト 206, アpartment エー

(72)発明者 ギャビン メレディス

アメリカ合衆国, ワシントン 98199, シアトル, サーティファースト アベニュー ウェスト 3020ビー

(72)発明者 マーク マッケルウェイン

アメリカ合衆国, ワシントン 98109, シアトル, フェアビュー アベニュー ノース 530, スイート 2000

(72)発明者 ピーター スキーン

アメリカ合衆国, ワシントン 98027, イサクア, スカイ リッジ ロード ノース ウェスト 135

(72)発明者 カサンドラ バーク

アメリカ合衆国, ワシントン 98115, シアトル, ノース イースト ラッドフォード ドライブ 6101, #712

審査官 福岡 信子

(56)参考文献 国際公開第2017/201073 (WO, A1)

特表2017-535269 (JP, A)

米国特許出願公開第2006/0014191 (US, A1)

米国特許出願公開第2016/0054308 (US, A1)

(58)調査した分野 (Int.Cl., DB名)

C12Q

C12N

JSTPlus/JMEDPlus/JST7580 (JDreamIII)

CAplus/MEDLINE/EMBASE/BIOSIS (STN)