

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2020-504709

(P2020-504709A)

(43) 公表日 令和2年2月13日 (2020.2.13)

(51) Int.Cl.	F I	テーマコード (参考)
C 4 O B 40/08 (2006.01)	C 4 O B 40/08 Z N A	4 B O 6 3
C 1 2 Q 1/686 (2018.01)	C 1 2 Q 1/686 Z	
C 1 2 Q 1/6874 (2018.01)	C 1 2 Q 1/6874 Z	
C 4 O B 50/14 (2006.01)	C 4 O B 50/14	

審査請求 未請求 予備審査請求 未請求 (全 97 頁)

(21) 出願番号	特願2019-525959 (P2019-525959)	(71) 出願人	516040017
(86) (22) 出願日	平成29年11月17日 (2017.11.17)		ツイスト バイオサイエンス コーポレー ション
(85) 翻訳文提出日	令和1年7月13日 (2019.7.13)		アメリカ合衆国 94158 カリフォル ニア州 サンフランシスコ ミッション・ ベイ・ブールバード サウス 455 ス イート 545
(86) 国際出願番号	PCT/US2017/062391	(74) 代理人	100082072
(87) 国際公開番号	W02018/094263		弁理士 清原 義博
(87) 国際公開日	平成30年5月24日 (2018.5.24)	(72) 発明者	ゼイトン, ラムゼイ イブライム
(31) 優先権主張番号	62/424,302		アメリカ合衆国 94107 カリフォル ニア州 サンフランシスコ 16番ストリ ート 1010 アpartment 608
(32) 優先日	平成28年11月18日 (2016.11.18)		
(33) 優先権主張国・地域又は機関	米国 (US)		
(31) 優先権主張番号	62/548,307		
(32) 優先日	平成29年8月21日 (2017.8.21)		
(33) 優先権主張国・地域又は機関	米国 (US)		

最終頁に続く

(54) 【発明の名称】 制御された化学量論を有するポリヌクレオチドライブラリおよびその合成

(57) 【要約】

ライブラリがあらかじめ決定された適用結果、例えば増幅後の制御された表示、および標的配列への結合後の均一な濃縮を可能にするように、ポリヌクレオチドの種に関してあらかじめ選択された化学量論を有するポリヌクレオチドのライブラリに関する組成物、方法、およびシステムが本明細書で提供される。さらに、均一かつ正確な次世代シーケンシングのための、ポリヌクレオチドプローブとその適用が提供される。

【選択図】 図 1 A

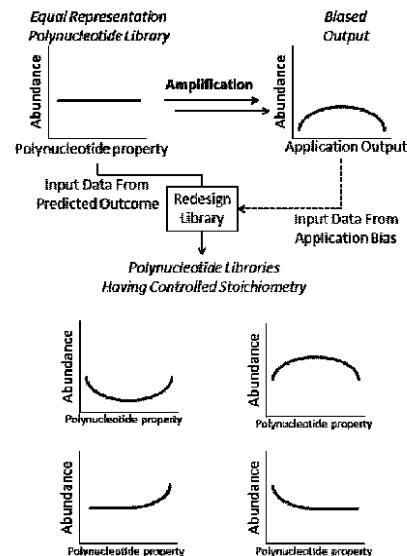


FIG. 1A

【特許請求の範囲】**【請求項 1】**

ポリヌクレオチドライブラリであって、ポリヌクレオチドライブラリは少なくとも 5 0 0 0 のポリヌクレオチドを含み、ゲノムフラグメントとのハイブリダイゼーションおよびハイブリダイズされたゲノムフラグメントのシーケンシングに続いて、ポリヌクレオチドライブラリが、ゲノムフラグメントの塩基に関して最大 5 5 倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも 9 0 パーセントの少なくとも 3 0 倍の読み取り深度を提供するような量で、少なくとも 5 0 0 0 のポリヌクレオチドの各々は存在する、ポリヌクレオチドライブラリ。

【請求項 2】

10

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大 5 5 倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも 9 5 パーセントの少なくとも 3 0 倍の読み取り深度を提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 3】

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大 5 5 倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも 9 8 パーセントの少なくとも 3 0 倍の読み取り深度を提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 4】

20

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも 9 0 パーセントの固有の読み取りを提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 5】

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも 9 5 パーセントの固有の読み取りを提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 6】

ポリヌクレオチドライブラリは、平均読み取り深度の約 1 . 5 倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも 9 0 パーセントを提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 7】

30

ポリヌクレオチドライブラリは、平均読み取り深度の約 1 . 5 倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも 9 5 パーセントを提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 8】

ポリヌクレオチドライブラリは、平均読み取り深度の約 1 . 5 倍以内の読み取り深度を有する、1 0 パーセントから 3 0 パーセントまたは 7 0 パーセントから 9 0 パーセントの G C の割合を有しているゲノムフラグメントの少なくとも 9 0 パーセントを提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 9】

40

ポリヌクレオチドライブラリは、平均読み取り深度の約 1 . 5 倍以内の読み取り深度を有する、1 0 パーセントから 3 0 パーセントまたは 7 0 パーセントから 9 0 パーセントの反復配列または二次構造配列の割合を有しているゲノムフラグメントの少なくとも約 8 0 パーセントを提供する、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 1 0】

ゲノムフラグメントの各々は、約 1 0 0 塩基から約 5 0 0 塩基の長さである、請求項 1 に記載のポリヌクレオチドライブラリ。

【請求項 1 1】

少なくとも 5 0 0 0 のポリヌクレオチドの少なくとも約 8 0 パーセントは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約 1 . 5 倍以内の量で表示される、請求項 1 に記載のポリヌクレオチドライブラリ。

50

【請求項 12】

少なくとも5000のポリヌクレオチドの少なくとも30パーセントは、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているポリヌクレオチドを含む、請求項1に記載のポリヌクレオチドライブラリ。

【請求項 13】

少なくとも5000のポリヌクレオチドの少なくとも約15パーセントは、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む、請求項1に記載のポリヌクレオチドライブラリ。

【請求項 14】

少なくとも5000のポリヌクレオチドは、少なくとも1000の遺伝子をコードする、請求項1に記載のポリヌクレオチドライブラリ。

【請求項 15】

ポリヌクレオチドライブラリは、少なくとも100,000のポリヌクレオチドを含む、請求項1に記載のポリヌクレオチドライブラリ。

【請求項 16】

ポリヌクレオチドライブラリは、少なくとも700,000のポリヌクレオチドを含む、請求項1に記載のポリヌクレオチドライブラリ。

【請求項 17】

少なくとも5,000のポリヌクレオチドは少なくとも1つのエクソン配列を含む、請求項1に記載のポリヌクレオチドライブラリ。

【請求項 18】

少なくとも700,000のポリヌクレオチドは、総体として単一のエクソン配列を含むポリヌクレオチドの少なくとも1つのセットを含む、請求項16に記載のポリヌクレオチドライブラリ。

【請求項 19】

少なくとも700,000のポリヌクレオチドは少なくとも150,000セットを含む、請求項18に記載のポリヌクレオチドライブラリ。

【請求項 20】

ポリヌクレオチドライブラリであって、ポリヌクレオチドライブラリは少なくとも5000のポリヌクレオチドを含み、ポリヌクレオチドの各々は約20~200塩基の長さであり、複数のポリヌクレオチドは、少なくとも1000のあらかじめ選択された遺伝子に関して各エクソンからの配列をコードし、各ポリヌクレオチドは分子タグを含み、少なくとも5000のポリヌクレオチドの各々は、ゲノムフラグメントとのハイブリダイゼーションとハイブリダイズされたゲノムフラグメントのシーケンシングに続いて、ポリヌクレオチドライブラリが、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも90パーセントの少なくとも30倍の読み取り深度を提供するような量で存在する、ポリヌクレオチドライブラリ。

【請求項 21】

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも95パーセントの少なくとも30倍の読み取り深度を提供する、請求項20に記載のポリヌクレオチドライブラリ。

【請求項 22】

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、少なくとも98パーセントのゲノムフラグメントの塩基の少なくとも30倍の読み取り深度を提供する、請求項20に記載のポリヌクレオチドライブラリ。

【請求項 23】

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも90パーセントの固有の読み取りを提供する、請求項20に記載のポリヌクレオチドライブラリ。

【請求項24】

ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも95パーセントの固有の読み取りを提供する、請求項20に記載のポリヌクレオチドライブラリ。

【請求項25】

ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも90パーセントを提供する、請求項20に記載のポリヌクレオチドライブラリ。

10

【請求項26】

ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも95パーセントを提供する、請求項20に記載のポリヌクレオチドライブラリ。

【請求項27】

ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているゲノムフラグメントの90パーセントより多くを提供する、請求項20に記載のポリヌクレオチドライブラリ。

20

【請求項28】

ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているゲノムフラグメントの約80パーセントより多くを提供する、請求項20に記載のポリヌクレオチドライブラリ。

【請求項29】

ゲノムフラグメントの各々は、約100塩基から約500塩基の長さである、請求項20に記載のポリヌクレオチドライブラリ。

【請求項30】

少なくとも5000のポリヌクレオチドの約80パーセントより多くが、ポリヌクレオチドライブラリに関する平均表示の少なくとも約1.5倍以内の量で表示される、請求項20に記載のポリヌクレオチドライブラリ。

30

【請求項31】

少なくとも5000のポリヌクレオチドの30パーセントより多くが、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているポリヌクレオチドを含む、請求項20に記載のポリヌクレオチドライブラリ。

【請求項32】

少なくとも5000のポリヌクレオチドの約15パーセントより多くが、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む、請求項20に記載のポリヌクレオチドライブラリ。

40

【請求項33】

ポリヌクレオチドライブラリは、少なくとも100,000のポリヌクレオチドを含む、請求項20に記載のポリヌクレオチドライブラリ。

【請求項34】

ポリヌクレオチドライブラリは、少なくとも700,000のポリヌクレオチドを含む、請求項20に記載のポリヌクレオチドライブラリ。

【請求項35】

少なくとも700,000のポリヌクレオチドは、総体として単一のエクソン配列を含むポリヌクレオチドの少なくとも1つのセットを含む、請求項34に記載のポリヌクレオ

50

チドライブラリ。

【請求項 36】

少なくとも 700,000 のポリヌクレオチドは少なくとも 150,000 セットを含む、請求項 35 に記載のポリヌクレオチドライブラリ。

【請求項 37】

ポリヌクレオチドライブラリを生成するための方法であって、該方法は：

a. 少なくとも 5000 のポリヌクレオチドをコードするあらかじめ決定された配列を提供する工程；

b. 少なくとも 5000 のポリヌクレオチドを合成する工程；および

c. ポリヌクレオチドライブラリを形成するためにポリメラーゼで少なくとも 5000 のポリヌクレオチドを増幅する工程であって、ここで、少なくとも 5000 のポリヌクレオチドの約 80 パーセントより多くは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約 2 倍以内の量で表示される、工程、を含む、方法。

10

【請求項 38】

少なくとも 5000 のポリヌクレオチドの約 80 パーセントより多くが、ポリヌクレオチドライブラリに関する平均表示の少なくとも約 1.5 倍以内の量で表示される、請求項 37 に記載の方法。

【請求項 39】

少なくとも 5000 のポリヌクレオチドの 30 パーセントより多くが、10 パーセントから 30 パーセントまたは 70 パーセントから 90 パーセントの GC の割合を有しているポリヌクレオチドを含む、請求項 37 に記載の方法。

20

【請求項 40】

少なくとも 5000 のポリヌクレオチドの約 15 パーセントより多くが、10 パーセントから 30 パーセントまたは 70 パーセントから 90 パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む、請求項 37 に記載の方法。

【請求項 41】

ポリヌクレオチドライブラリは、エラーの補正なしで、あらかじめ決定された配列と比較して、800 塩基において 1 未満の合計エラー率を有する、請求項 37 に記載の方法。

【請求項 42】

あらかじめ決定された配列は、少なくとも 700,000 のポリヌクレオチドをコードする、請求項 37 に記載の方法。

30

【請求項 43】

少なくとも 5000 のポリヌクレオチドの合成は、面を有する構造において行われ、面は複数のクラスタを含み、各クラスタは複数の遺伝子座を含み；および少なくとも 5000 のポリヌクレオチドの各々は、複数の遺伝子座のうちの異なる遺伝子座から伸長する、請求項 37 に記載の方法。

【請求項 44】

複数の遺伝子座はクラスタあたり最大 1000 の遺伝子座を含む、請求項 43 に記載の方法。

40

【請求項 45】

複数の遺伝子座はクラスタあたり最大 200 の遺伝子座を含む、請求項 43 に記載の方法。

【請求項 46】

ポリヌクレオチドライブラリの増幅のための方法であって、該方法は：

a. 少なくとも 5000 のポリヌクレオチドのための増幅分布を得る工程；

b. 少なくとも 1 つの配列特徴に基づいて 2 つ以上のピンへと増幅分布の少なくとも 5000 のポリヌクレオチドをクラスタリングする工程であって、配列特徴はパーセント GC 含量、パーセント反復配列含量、またはパーセント二次構造含量である、工程；

c. あらかじめ選択された表示を有するポリヌクレオチドライブラリを生成するために

50

少なくとも１つのピン中のポリヌクレオチドの相対度数を適合させる工程；

d．あらかじめ選択された表示を有するポリヌクレオチドライブラリを合成する工程；
および、

e．あらかじめ選択された表示を有するポリヌクレオチドライブラリを増幅する工程、
を含む、方法。

【請求項４７】

少なくとも１つの配列特徴はパーセントGC含量である、請求項４６に記載の方法。

【請求項４８】

少なくとも１つの配列特徴はパーセント二次構造含量である、請求項４６に記載の方法
。

10

【請求項４９】

少なくとも１つの配列特徴はパーセント反復配列含量である、請求項４６に記載の方法
。

【請求項５０】

反復配列含量は３つ以上のアデニンを有する配列を含む、請求項４９に記載の方法。

【請求項５１】

反復配列含量は、ポリヌクレオチドの少なくとも１つの末端に反復配列を含む、請求項
４９に記載の方法。

【請求項５２】

前記ポリヌクレオチドは、標的配列を結合するために１つ以上のポリヌクレオチド配列
の親和性に基づいてピンへとクラスタリングされる、請求項４６に記載の方法。

20

【請求項５３】

ピンの下３０パーセントにある配列の数は、調整前のピンの下３０パーセントにある配
列の数と比較して、調整後の下流での適用に少なくとも５０％より多くの表示を有する、
請求項４６に記載の方法。

【請求項５４】

ピンの上３０パーセントにある配列の数は、調整前のピンの上３０パーセントにある配
列の数と比較して、調整後の下流での適用に少なくとも５０％より多くの表示を有する、
請求項４６に記載の方法。

【請求項５５】

ゲノムDNAのシーケンシングのための方法であって、該方法は：

(a)複数のゲノムのフラグメントに、請求項１－３６のいずれか１つに記載のライブ
ラリを接触させる工程；

(b)少なくとも１つの濃縮された標的ポリヌクレオチドを生成するために、ライブラ
リに結合する少なくとも１つのゲノムフラグメントを濃縮する工程；および、

(c)少なくとも１つの濃縮された標的ポリヌクレオチドを配列決定する工程、
を含む、方法。

【請求項５６】

複数の濃縮された標的ポリヌクレオチドはcDNAライブラリを含む、請求項５５に記
載の方法。

40

【請求項５７】

少なくとも５０００のポリヌクレオチドは約８０～約２００の塩基の長さである、請求
項５５に記載の方法。

【請求項５８】

ゲノムフラグメントの各々は、約１００塩基から約５００塩基の長さである、請求項５
５に記載の方法。

【請求項５９】

接触させる工程は溶液中で行なわれる、請求項５５に記載の方法。

【請求項６０】

少なくとも５０００のポリヌクレオチドはゲノムフラグメントに対して少なくとも部分

50

的に相補的である、請求項 55 に記載の方法。

【請求項 61】

単離は (i) 個体担体上のポリヌクレオチド/ゲノムフラグメントのハイブリダイゼーション対を捕捉する工程；および (ii) 濃縮された標的ポリヌクレオチドを生成するために複数のゲノムフラグメントを放出する工程、を含む、請求項 55 に記載の方法。

【請求項 62】

シーケンシングは結果として、ゲノムフラグメントの塩基に関して最大 55 倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも 95 パーセントの少なくとも 30 倍の読み取り深度をもたらす、請求項 55 に記載の方法。

【請求項 63】

シーケンシングは結果として、ゲノムフラグメントの塩基に関して最大 55 倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも 98 パーセントの少なくとも 30 倍の読み取り深度をもたらす、請求項 55 に記載の方法。

【請求項 64】

シーケンシングは結果として、ゲノムフラグメントの塩基に関して少なくとも 90 パーセントの固有の読み取りをもたらす、請求項 55 に記載の方法。

【請求項 65】

シーケンシングは結果として、ゲノムフラグメントの塩基に関して少なくとも 95 パーセントの固有の読み取りをもたらす、請求項 55 に記載の方法。

【請求項 66】

シーケンシングは結果として、平均読み取り深度の約 1.5 倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも 90 パーセントをもたらす、請求項 55 に記載の方法。

【請求項 67】

シーケンシングは結果として、平均読み取り深度の約 1.5 倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも 95 パーセントをもたらす、請求項 55 に記載の方法。

【請求項 68】

シーケンシングは結果として、平均読み取り深度の約 1.5 倍以内の読み取り深度を有する、10 パーセントから 30 パーセントまたは 70 パーセントから 90 パーセントの GC の割合を有しているゲノムフラグメントの少なくとも 90 パーセントをもたらす、請求項 55 に記載の方法。

【請求項 69】

シーケンシングは結果として、平均読み取り深度の約 1.5 倍以内の読み取り深度を有する、10 パーセントから 30 パーセントまたは 70 パーセントから 90 パーセントの反復配列または二次構造配列の割合を有しているゲノムフラグメントの少なくとも約 80 パーセントをもたらす、請求項 55 に記載の方法。

【請求項 70】

少なくとも 5000 のポリヌクレオチドの少なくとも約 80 パーセントが、ポリヌクレオチドライブラリに関する平均表示の少なくとも約 1.5 倍以内の量で表示される、請求項 55 に記載の方法。

【請求項 71】

少なくとも 5000 のポリヌクレオチドの少なくとも 30 パーセントが、10 パーセントから 30 パーセントまたは 70 パーセントから 90 パーセントの GC の割合を有しているポリヌクレオチドを含む、請求項 55 に記載の方法。

【請求項 72】

少なくとも 5000 のポリヌクレオチドの少なくとも約 15 パーセントが、10 パーセントから 30 パーセントまたは 70 パーセントから 90 パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む、請求項 55 に記載の方法。

【請求項 73】

10

20

30

40

50

少なくとも5000のポリヌクレオチドは、少なくとも1000の遺伝子をコードする、請求項55に記載の方法。

【請求項74】

ポリヌクレオチドライブラリは、少なくとも100,000のポリヌクレオチドを含む、請求項55に記載の方法。

【請求項75】

ポリヌクレオチドライブラリは、少なくとも700,000のポリヌクレオチドを含む、請求項55に記載の方法。

【請求項76】

少なくとも5,000のポリヌクレオチドは少なくとも1つのエクソン配列を含む、請求項55に記載の方法。

【請求項77】

少なくとも700,000のポリヌクレオチドは、総体として単一のエクソン配列を含むポリヌクレオチドの少なくとも1つのセットを含む、請求項75に記載の方法。

【請求項78】

少なくとも700,000のポリヌクレオチドは少なくとも150,000セットを含む、請求項77に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

<相互参照>

本出願は、2016年11月18日に出願された米国仮特許出願第62/424,302号、2017年8月21日に出願された米国仮特許出願第62/548,307号、および2017年9月14日に出願された米国仮特許出願第62/558,666号の利益を主張し、これらの各々は参照によりその全体が本明細書に組み込まれる。

【0002】

<配列表>

本出願は、ASCIIフォーマットで電子的に提出され、参照によりその全体が本明細書に組み込まれる配列表を含んでいる。2017年11月13日に作成された前記ASCIIのコピーは、44854-730__601__SL.txtというファイル名であり、5,304バイトのサイズである。

【背景技術】

【0003】

高い忠実性と低コストでの非常に効率的な化学的遺伝子合成は、バイオテクノロジーと医療において、および基礎的な生物学的研究において中心的な役割を有している。デノボの遺伝子合成は、基礎的な生物学的研究およびバイオテクノロジーでの応用にとって強力なツールである。小規模での比較的短いフラグメントの合成のための様々な方法が知られているが、これらの技術にはしばしば、スケーラビリティ、自動化、速度、精度、およびコストの点で問題がある。

【0004】

<引用による組み込み>

本明細書で言及される全ての刊行物、特許、および特許出願は、個々の刊行物、特許、または特許出願が参照により組み込まれるように具体的かつ個々に示されているかのような程度で、参照により本明細書に組み込まれる。

【発明の概要】

【0005】

本明細書にはにおいてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは少なくとも5000のポリヌクレオチドを含み、ゲノムフラグメントとのハイブリダイゼーションおよびハイブリダイズされたゲノムフラグメントのシーケンシングに続いて、ポリヌクレオチドライブラリがゲノムフラグメントの塩基に関して最大55倍の理論

10

20

30

40

50

上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも90パーセントの少なくとも30倍の読み取り深度を提供するような量で、少なくとも5000のポリヌクレオチドの各々は存在する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも95パーセントの少なくとも30倍の読み取り深度を提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも98パーセントの少なくとも30倍の読み取り深度を提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも90パーセントの固有の読み取りを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも95パーセントの固有の読み取りを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも90パーセントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも95パーセントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているゲノムフラグメントの少なくとも90パーセントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているゲノムフラグメントの少なくとも約80パーセントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここでゲノムフラグメントの各々は、約100塩基から約500塩基の長さである。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドの少なくとも約80パーセントは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約1.5倍以内の量で表示される。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドの少なくとも30パーセントは、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドの少なくとも約15パーセントは、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドは少なくとも1000の遺伝子をコードする。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、少なくとも100,000のポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、少なくとも700,000のポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドは少なくとも1つのエクソン配列を含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも700,000のポリヌクレオチドは、総体として単一のエクソン配列を含むポリヌクレオチドの少なくとも1つのセットを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも700,000のポリヌクレオチドは少なくとも150,000のセットを含む。

【0006】

10

20

30

40

50

本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは少なくとも5000のポリヌクレオチドを含み、ポリヌクレオチドの各々は約20から200塩基の長さであり、複数のポリヌクレオチドは、少なくとも1000のあらかじめ選択された遺伝子に関して各エクソンからの配列をコードし、各ポリヌクレオチドは分子タグを含み、ゲノムフラグメントとのハイブリダイゼーションとハイブリダイズされたゲノムフラグメントのシーケンシングに続いて、ポリヌクレオチドライブラリがゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下でゲノムフラグメントの塩基の少なくとも90パーセントの少なくとも30倍の読み取り深度を提供するような量で、少なくとも5000のポリヌクレオチドの各々は存在する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも95パーセントの少なくとも30倍の読み取り深度を提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも98パーセントの少なくとも30倍の読み取り深度を提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも90パーセントの固有の読み取りを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、ゲノムフラグメントの塩基に関して少なくとも95パーセントの固有の読み取りを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも90パーセントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも95パーセントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有している90パーセントより多くのゲノムフラグメントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有している約80パーセントより多くのゲノムフラグメントを提供する。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここでゲノムフラグメントの各々は、約100塩基から約500塩基の長さである。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドの約80パーセントより多くは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約1.5倍以内の量で表示される。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドの30パーセントより多くは、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも5000のポリヌクレオチドの15パーセントより多くは、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、少なくとも100,000のポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ポリヌクレオチドライブラリは、少なくとも700,000のポリヌクレオチドを含む。さらに本明細書においてポリヌクレオチドライブラリが提供され、ここで少なくとも700,000のポリヌクレオチドは、総体として単一のエクソン配列を含むポリヌクレオチドの少なくとも1つのセットを含む。さらに本明細書においてポリヌクレオチドラ

イブラリが提供され、ここで少なくとも700,000のポリヌクレオチドは少なくとも150,000のセットを含む。

【0007】

本明細書で提供されるのは、ポリヌクレオチドライブラリを生成するための方法であって、該方法は：少なくとも5000のポリヌクレオチドをコードするあらかじめ決定された配列を提供する工程；少なくとも5000のポリヌクレオチドを合成する工程；および、ポリヌクレオチドライブラリを形成するためにポリメラーゼで少なくとも5000のポリヌクレオチドを増幅する工程を含み、ここで、少なくとも5000のポリヌクレオチドの約80パーセントより多くは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約2倍以内の量で表示される。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドの約80パーセントより多くは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約1.5倍以内の量で表示される。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドの30パーセントより多くは、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドの15パーセントより多くは、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、エラーの補正なしで、あらかじめ決定された配列と比較して、800塩基において1未満の合計エラー率を有する。さらに本明細書で提供される方法において、あらかじめ決定された配列は少なくとも700,000のポリヌクレオチドをコードする。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドの合成は、面を有する構造において行われ、面は複数のクラスタを含む。各クラスタは複数の遺伝子座を含み、および少なくとも5000のポリヌクレオチドの各々、複数の遺伝子座の異なる遺伝子座から伸長する。さらに本明細書で提供される方法において、複数の遺伝子座はクラスタあたり最大1000の遺伝子座を含む。さらに本明細書で提供される方法において、複数の遺伝子座はクラスタあたり最大200の遺伝子座を含む。

【0008】

本明細書で提供されるのは、ポリヌクレオチドライブラリの増幅のための方法であって、該方法は：少なくとも5000のポリヌクレオチドのための増幅分布を得る工程；少なくとも1つの配列特徴に基づいて2つ以上のピンへと、増幅分布の少なくとも5000のポリヌクレオチドをクラスタリングする工程であって、配列特徴はパーセントGC含量、パーセント反復配列含量、またはパーセント二次構造含量である、工程；あらかじめ選択された表示を有するポリヌクレオチドライブラリを生成するために少なくとも1つのピン中のポリヌクレオチドの相対度数を適合させる工程；あらかじめ選択された表示を有するポリヌクレオチドライブラリを合成する工程；および、あらかじめ選択された表示を有するポリヌクレオチドライブラリを増幅する工程、を含む。さらに本明細書で提供される方法において、少なくとも1つの配列特徴はパーセントGC含量である。さらに本明細書で提供される方法において、少なくとも1つの配列特徴はパーセント反復配列含量である。さらに本明細書で提供される方法において、反復配列含量は3つ以上のアデニンを有する配列を含む。さらに本明細書で提供される方法において、反復配列含量は、ポリヌクレオチドの少なくとも1つの末端に反復配列を含む。さらに本明細書で提供される方法において、前記ポリヌクレオチドは、標的配列を結合するために1つ以上のポリヌクレオチド配列の親和性に基づいてピンへとクラスタリングされる。さらに本明細書で提供される方法において、ピンの下30パーセントにある配列の数は、調整前のピンの下30パーセントにある配列の数と比較して、調整後の下流での適用に少なくとも50%より多くの表示を有する。さらに本明細書で提供される方法において、ピンの上30パーセントにある配列の数は、調整前のピンの上30パーセントにある配列の数と比較して、

調整後の下流での適用に少なくとも50%より多くの表示を有する。

【0009】

本明細書で提供されるのは、ゲノムDNAのシーケンシングのための方法であって、該方法は：複数のゲノムフラグメントに、本明細書に記載されるポリヌクレオチドライブラリのいずれかを接触させる工程；少なくとも1つの濃縮された標的ポリヌクレオチドを生成するためにライブラリに結合する少なくとも1つのゲノムフラグメントを濃縮する工程；および少なくとも1つの濃縮された標的ポリヌクレオチドを配列決定（シーケンシング）する工程、を含む。さらに本明細書で提供される方法において、複数の濃縮された標的ポリヌクレオチドはcDNAライブラリを含む。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドは約80～約200の塩基の長さである。さらに本明細書で提供される方法において、ゲノムフラグメントの各々は、約100塩基から約500塩基の長さである。さらに本明細書で提供される方法において、接触させる工程は溶液中で行なわれる。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドはゲノムフラグメントに対して少なくとも部分的に相補的である。さらに本明細書で提供される方法において、単離は（i）個体担体上のポリヌクレオチド/ゲノムフラグメントのハイブリダイゼーション対を捕捉する工程；および（ii）濃縮された標的ポリヌクレオチドを生成するために複数のゲノムフラグメントを放出する工程、を含む。さらに本明細書で提供される方法において、シーケンシングは結果として、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも95パーセントの少なくとも30倍の読み取り深度をもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、ゲノムフラグメントの塩基に関して最大55倍の理論上の読み取り深度のための条件下で、ゲノムフラグメントの塩基の少なくとも98パーセントの少なくとも30倍の読み取り深度をもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、ゲノムフラグメントの塩基に関して少なくとも90パーセントの固有の読み取りをもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、ゲノムフラグメントの塩基に関して少なくとも95パーセントの固有の読み取りをもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも90パーセントをもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも95パーセントをもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、平均読み取り深度の約1.5倍以内の読み取り深度を有するゲノムフラグメントの塩基の少なくとも95パーセントをもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているゲノムフラグメントの少なくとも90パーセントをもたらし。さらに本明細書で提供される方法において、シーケンシングは結果として、平均読み取り深度の約1.5倍以内の読み取り深度を有する、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているゲノムフラグメントの少なくとも約80パーセントをもたらし。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドの少なくとも約80パーセントは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約1.5倍以内の量で表示される。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドの少なくとも30パーセントは、10パーセントから30パーセントまたは70パーセントから90パーセントのGCの割合を有しているポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドの少なくとも15パーセントは、10パーセントから30パーセントまたは70パーセントから90パーセントの反復配列または二次構造配列の割合を有しているポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドは少なくとも1000の遺伝子をコードする。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、少なくとも100,000のポリヌクレオチドを含

10

20

30

40

50

む。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、少なくとも700,000のポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも5000のポリヌクレオチドは少なくとも1つのエクソン配列を含む。さらに本明細書で提供される方法において、少なくとも700,000のポリヌクレオチドは、総体として単一のエクソン配列を含むポリヌクレオチドの少なくとも1つのセットを含む。さらに本明細書で提供される方法において、少なくとも700,000のポリヌクレオチドは少なくとも150,000のセットを含む。

【図面の簡単な説明】

【0010】

【図1A】最初のポリヌクレオチドライブラリの適用、適用出力からのバイアスの測定、第2の制御された化学量論ポリヌクレオチドライブラリの設計と合成、および望ましい表示出力を生成するための第2のポリヌクレオチドライブラリの適用を含む、概略的なワークフローを示す。

10

【図1B】標的結合ポリヌクレオチドライブラリで標的ポリヌクレオチドを濃縮するための概要を示す。

【図1C】核酸サンプルの濃縮とシーケンシングに関する典型的なワークフローを表す。

【図2】クラスタ増幅からのポリヌクレオチドライブラリの生成の概略を表す。

【図3A】標的化と濃縮のための1対のポリヌクレオチドを表す。ポリヌクレオチドは、相補的な標的結合（挿入）配列、およびプライマー結合部位を含む。

【図3B】標的化と濃縮のための1対のポリヌクレオチドを表す。ポリヌクレオチドは、相補的な標的配列結合（挿入）配列、プライマー結合部位、および非標的配列を含む。

20

【図4A】より大きなポリヌクレオチドの標的配列へのポリヌクレオチド結合の形態を表す。標的配列はポリヌクレオチド結合領域より短く、およびポリヌクレオチド結合領域（または挿入配列）は標的配列に対してオフセットされ、さらに隣接する配列の一部に結合する。

【図4B】より大きなポリヌクレオチドの標的配列へのポリヌクレオチド結合の形態を表す。標的配列の長さは、ポリヌクレオチド結合領域以下であり、およびポリヌクレオチド結合領域は標的配列で中心におかれ、隣接する配列の一部にさらに結合する。

【図4C】より大きなポリヌクレオチドの標的配列へのポリヌクレオチド結合の形態を表す。標的配列はポリヌクレオチド結合領域よりわずかに長く、およびポリヌクレオチド結合領域は、各側の緩衝領域で標的配列上の中心におかれる。

30

【図4D】より大きなポリヌクレオチドの標的配列へのポリヌクレオチド結合の形態を表す。標的配列はポリヌクレオチド結合領域より長く、および2つのポリヌクレオチドの結合領域は、標的配列に及ぶようにオーバーラップする。

【図4E】より大きなポリヌクレオチドの標的配列へのポリヌクレオチド結合の形態を表す。標的配列はポリヌクレオチド結合領域より長く、および2つのポリヌクレオチドの結合領域は、標的配列に及ぶようにオーバーラップする。

【図4F】より大きなポリヌクレオチドの標的配列へのポリヌクレオチド結合の形態を表す。標的配列はポリヌクレオチド結合領域より長く、および2つのポリヌクレオチドの結合領域は、標的配列に及ぶようにオーバーラップせず、隙間（405）を残す。

40

【図4G】より大きなポリヌクレオチドの標的配列へのポリヌクレオチド結合の形態を表す。標的配列はポリヌクレオチド結合領域より長く、および3つのポリヌクレオチドの結合領域は、標的配列に及ぶようにオーバーラップする。

【図5】本明細書に開示される遺伝子合成のための典型的な処理ワークフローを明示する工程の図を示す。

【図6】コンピュータシステムを例示する。

【図7】コンピュータシステムのアーキテクチャを例示するブロック図である。

【図8】複数のコンピュータシステム、複数の携帯電話と個人用携帯情報端末、およびネットワークアタッチトストレージ（NAS）を組み込むように構成されたネットワークを明示する図表である。

50

【図 9】共有の仮想アドレスメモリ容量を使用するマルチプロセッサコンピュータシステムのブロック図である。

【図 10】256のクラスタを有するプレートの画像であり、各クラスタは121の遺伝子座を有し、そこからポリヌクレオチドが伸長する。

【図 11 A】240のクラスタからの29,040の特有のポリヌクレオチドの合成からの、プレートにわたるポリヌクレオチド表示（測定される吸収度等の、ポリヌクレオチド頻度 対 豊富量）のプロットであり、各クラスタは121のポリヌクレオチドを有する。

【図 11 B】個々の各クラスタにわたるポリヌクレオチド頻度 対 豊富量の吸収度（測定される吸収度）の測定のプロットであり、対照のクラスタはボックスで識別される。

【図 12】4つの個別のクラスタにわたるポリヌクレオチド頻度 対 豊富量（測定される吸収度）の測定のプロットである。

【図 13 A】240のクラスタからの29,040の特有のポリヌクレオチドの合成からの、プレートにわたる頻度 対 エラー率のプロットであり、各クラスタは121のポリヌクレオチドを有する。

【図 13 B】各個別のクラスタにわたるポリヌクレオチドのエラー率 対 頻度の測定のプロットであり、対照のクラスタはボックスで識別される。

【図 14】4つのクラスタにわたるポリヌクレオチド頻度 対 エラー率の測定のプロットである。

【図 15】ポリヌクレオチドの数 対 1つのポリヌクレオチドあたりのパーセントの測定数としてのGC含量のプロットである。

【図 16】2つの異なるポリメラーゼでのPCRからの結果を提供する。各チャートは、ポリヌクレオチドの数（0～2,000）対 観察された頻度（100,000あたりの数で計測した「0～35」）を表す。

【図 17】記録された増幅後のポリヌクレオチド集団の均一性の定量化でのチャートを提供する。

【図 18】配列ドロップアウトへの過剰増幅の影響を明示するプロットを表す。

【図 19】ポリヌクレオチドの未増幅集団と増幅集団における、ポリヌクレオチド頻度あたりの（100,000読み取りあたりの）GC含量の割合のプロットを表す。

【図 20】クラスタ増幅後の2つの別個の実行に関する、ポリヌクレオチド頻度あたりの（100,000読み取りあたりの）GC含量の割合のプロットである。

【図 21 A】ポリヌクレオチドのGC平衡ライブラリに関する、ポリヌクレオチド頻度あたりのGC含量の割合のプロットである。

【図 21 B】ポリヌクレオチドの極度に高い、および低いGCバイアスライブラリに関する、ポリヌクレオチド頻度あたりのGC含量の割合のプロットである。

【図 21 C】ポリヌクレオチドのやや高い、および低いGCバイアスライブラリに関する、ポリヌクレオチド頻度あたりのGC含量の割合のプロットである。

【図 21 D】ポリヌクレオチドの低GCバイアスライブラリに関する、ポリヌクレオチド頻度あたりのGC含量の割合のプロットである。

【図 21 E】ポリヌクレオチドの高GCバイアスライブラリに関する、ポリヌクレオチド頻度あたりのGC含量の割合のプロットである。

【図 22】15%～85%のGC含量を含む配列を有する理論上13,000plexポリヌクレオチドライブラリに関する、ポリヌクレオチド頻度あたりのGC含量の割合のプロットである。

【図 23】GCの平衡なポリヌクレオチドライブラリに関する、ポリヌクレオチドの数 対 ポリヌクレオチドの頻度（100,000読み取りあたり）のプロットである。

【図 24 A】単分散ライブラリの理論上の最大値と比較して、GCの平衡なポリヌクレオチドライブラリに関して80%のシーケンシングカバー率を得るのに必要なサンプリングの量を示すプロットである。

【図 24 B】単分散ライブラリの理論上の最大値と比較して、GCの平衡なポリヌクレオチドライブラリに関して90%のシーケンシングカバー率を得るのに必要なサンプリング

10

20

30

40

50

の量を示すプロットである。

【図 2 5】80ヌクレオチドの長さのポリヌクレオチドを含むライブラリに関する、ポリヌクレオチドの数 対 ポリヌクレオチドの頻度 (1, 000, 000 読み取りあたりの数) のプロットである。

【図 2 6】120ヌクレオチドの長さのポリヌクレオチドを含むライブラリに関する、ポリヌクレオチドの数 対 ポリヌクレオチドの頻度 (1, 000, 000 読み取りあたりの数) のプロットである。

【図 2 7】80および120のヌクレオチド長さのGCの平衡なポリヌクレオチドライブラリの両方に関して、(1, 000, 000の読み取りにつき)ポリヌクレオチドの平均頻度を示すプロットである。

【図 2 8】ポリヌクレオチド配列表示上の、PCR増幅サイクル数、GC含量、およびDNAポリメラーゼの選択の効果を表すプロットを表す。

【図 2 9】2つの異なる高忠実度のポリメラーゼに関する増幅サイクルの関数としての配列ドロップアウトのプロットである。

【図 3 0】配列表示に対する、異なるDNAポリメラーゼの効果を表すプロットを表す。同じポリヌクレオチドライブラリは、DNAポリメラーゼ1またはDNAポリメラーゼ2を用いて15サイクル、増幅された。

【図 3 1 A】制御された化学量論なしに、エクソームブロープライブラリを使用して、標的配列に関する任意の読み取り深度を達成するのに求められるシーケンシングに係る量を表す。

【図 3 1 B】制御された化学量論のないエクソームブロープライブラリと比較した、制御された化学量論を有するエクソームブロープライブラリを使用して標的配列に関する所与の読み取り深度を達成するために求められる配列全体に及ぶ減少を示す。

【図 3 2 A】コンパレータエクソームブローキットAと制御された化学量論ブロープライブラリ1の両方に関する、1x、20xまたは30xシーケンシング読み取り深度(Xカバー率)を有する塩基のパーセントのプロットである。

【図 3 2 B】コンパレータエクソームブローキットと制御された化学量論ブロープライブラリ1のパネルに関する、4.5Gbのシーケンシングで標準化された、1xまたは10xシーケンシング読み取り深度(Xカバー率)を有する塩基のパーセントのプロットである。

【図 3 3】ライブラリ内のポリヌクレオチドの数の関数として、異なるスケールのポリヌクレオチドブロープライブラリの合成を表す。

【図 3 4】コンパレータアレイベースのブロープライブラリの読み取り深度 対 制御された化学量論ブロープライブラリ2の関数としてのカバー率(塩基数)の比較を表す。

【図 3 5 A】10~30%、および30~50%のGC含量を有する標的に関する、コンパレータアレイベースのブロープライブラリ 対 制御された化学量論ブロープライブラリ2の読み取り深度の関数としてのカバー率間の比較を示す。

【図 3 5 B】50~70%より大きい、および30~50%より大きいGC含量を有する標的に関する、コンパレータアレイベースのブロープライブラリ 対 制御された化学量論ブロープライブラリ2の読み取り深度の関数としてのカバー率間の比較を示す。

【図 3 6 A】コンパレータアレイベースのブロープライブラリの標的比率のパーセント(%)と、制御された化学量論ブロープライブラリ3の0.1x、1x、および3xの濃度との間の比較を示す

【図 3 6 B】コンパレータアレイベースのブロープライブラリの読み取り深度と、制御された化学量論ブロープライブラリ4の0.1x、1x、および3xの濃度との間の比較を示す。

【図 3 7】コンパレータエクソームキット 対 制御された化学量論ブロープライブラリ4の、特有の読み取りの割合と、1X、20X、および30Xの読み取り深度での標的塩基との間の比較を示す。

【図 3 8】コンパレータエクソームキット 対 制御された化学量論ブロープライブラリ4

10

20

30

40

50

の、カバーされた塩基の割合と、1 X、20 X、および30 Xの読み取り深度での標的塩基との間の比較を示す。

【発明を実施するための形態】

【0011】

本開示は、別記されない限り、当技術分野の技術内である、従来の分子生物学技術を使用する。他に定義されない限り、本明細書で用いられる全ての技術的用語と科学的用語は、当業者によって一般に理解されるものと同じ意味を有する。

【0012】

本明細書において、大きなポリヌクレオチドライブラリの化学量論を設計し、合成し、制御する方法が提供される。ポリヌクレオチドの第1の集団が、濃縮のための補足プローブとして、例えば増幅のための、予備的な適用工程、および遺伝子合成にかけられる場合、続くポリヌクレオチドの集団の増幅反応は結果として、ポリヌクレオチド配列における変異によるバイアスのかかった表示出力をもたらし得、特定のポリヌクレオチドは他のものよりもより豊富に表示される。図11A。この予備的な適用出力から観察される、結果としてもたらされるバイアスが測定され、およびあらかじめ選択された化学量論、例えばGC含量、反復配列、後端(トレーリング)アデニン、二次構造、標的配列結合への親和性、または修飾されたヌクレオチド等の任意の数の配列特徴を考慮した、集団におけるポリヌクレオチドの相対頻度を用いて、ポリヌクレオチドの第1の集団を制御するために使用される。ポリヌクレオチドの化学量論を修正した後に、ポリヌクレオチドの第2の集団は、適用工程に関連する望ましくないバイアス効果を修正するためにあらかじめ選択された化学量論で、設計され、かつ合成される。いくつかの例では、ポリヌクレオチドの第2の制御された化学量論集団を、PCR増幅等の適用工程にかけることで、結果として、高度に均一な表示を伴う増幅されたポリヌクレオチド集団等の平衡な出力、または表示にあらかじめ選択された転移を有する不均一な表示等がもたらされる。図1Aの下のチャートを参照されたい。いくつかの例では、本明細書で記載される方法は、ポリヌクレオチド集団が高度に均一な標的配列捕捉頻度を提供するように、ポリヌクレオチドプローブの配列表示を制御する工程を含む(図1B)。例えば、ポリヌクレオチド(100)のサンプルは標的ポリヌクレオチド(101)を含む。適切な条件(102)下での標的結合ポリヌクレオチド(103)とのサンプル(100)の接触は、結果として、ハイブリダイゼーション対(104)の形成をもたらし、これはサンプル(100)において非標的ポリヌクレオチドから分離される。対(104)の変性と分離は、シーケンシング等の下流での適用のために濃縮された標的ポリヌクレオチド(107)を放出する。さらに本明細書において提供されるのは、シーケンシングプロセスの文脈における例として、ゲノムDNAへのハイブリダイゼーションで使用されるデノボ合成ポリヌクレオチドである。典型的なシーケンシングワークフロー(図1C)の第1の工程において、標的ポリヌクレオチドを含む核酸サンプル(108)が、フラグメント(109)のライブラリを形成するために、機械的切断または酵素による切断によってフラグメント化される。随意にプライマー配列および/またはバーコードを含むアダプター(115)は、アダプター標識ライブラリ(110)を形成するために結合される。このライブラリはその後、随意に増幅され、標的結合ポリヌクレオチド(117)とアダプター(115)のハイブリダイゼーションを防ぐブロッキング(遮断)ポリヌクレオチド(116)と共に、標的ポリヌクレオチドにハイブリダイズする標的結合ポリヌクレオチド(117)とハイブリダイズする。標的ポリヌクレオチド・標的結合ポリヌクレオチドのハイブリダイゼーション対(112)の補足、および標的結合ポリヌクレオチド(117)の除去は、標的ポリヌクレオチド(113)の単離/濃縮を可能にし、これはその後、随意に増幅され、配列される(114)。

【0013】

定義

【0014】

本開示の全体にわたって、数的な特徴は範囲のフォーマットで提示される。範囲のフォーマットでの記載が、単に利便性と簡潔性のためのものであることを理解すべきであり、

10

20

30

40

50

実施形態の範囲に対する柔軟性のない限定として解釈されるべきではない。したがって、範囲の記載は、文脈が他で明確に指示しない限り、下限の単位の10分の1までの範囲内の個々の数値と共に、すべての可能性のある部分範囲を具体的に開示していると考えべきである。例えば、1～6などの範囲の記載は、1～3、1～4、1～5、2～4、2～6、3～6などの部分範囲と共に、例えば1、1、2、2、3、5、および5、9などの、その範囲内の個々の値を具体的に開示していると考えべきである。これは範囲の幅にかかわらず適用される。これらの介在範囲の上限および下限は、より小さな範囲に独立して含まれてもよく、また、定められた範囲の具体的に除外された限度に従って、本発明内に包含される。定められた範囲が上限および下限の1つまたはその両方を含む場合、これらの含まれた上限および下限のいずれかまたは両方を除く範囲もまた、文脈が他で明確に指示しない限り、本発明に包含される。

10

【0015】

本明細書で使用される用語は、特定の実施形態のみを記載するためのものであり、実施形態を限定するようには意図されない。本明細書で使用されるように、単数形の「a」、「an」および「the」は、文脈が他で明確に指示しない限り、複数形も同様に含むことが意図される。用語「含む」および/または「含んでいる」は、本明細書で使用される場合、述べられた特徴、整数、工程、動作、要素、および/または構成要素の存在を規定するが、1つ以上の他の特徴、整数、工程、動作、要素、構成要素、および/またはそれらのグループの存在または追加を妨げないことが、さらに理解される。本明細書において使用されるように、用語「および/または」は、表記される関連項目の1つ以上の任意の、および全ての組み合わせを含む。

20

【0016】

具体的に規定されない限り、または文脈から明白でない限り、本明細書で使用されるように、数または数の範囲に関する用語「約」は、定められた数の+/-10%、または範囲に関して表記される値の表記された下限より10%下、および表記された上限の10%上を意味すると理解される。

【0017】

本明細書において使用されるように、用語「あらかじめ選択された配列」、「あらかじめ規定された配列」または「あらかじめ決定された配列」は、区別なく使用される。該用語は、ポリマーの配列が、ポリマーの合成または組立ての前に既知であり、かつ選択されることを意味する。特に、本発明の様々な態様は主として核酸分子の調製に関連して本明細書に記載され、オリゴヌクレオチドまたはポリヌクレオチドの配列は、核酸分子の合成または組立て前に既知であり、かつ選択されている。

30

【0018】

用語「核酸」は、単鎖の分子と共に、二重鎖または三重鎖の核酸を包含する。二重鎖または三重鎖の核酸では、核酸の鎖は同一の広がりを持つ必要はない（すなわち、二重らせん構造の核酸は、両方の鎖の全長に沿って二重らせん構造である必要はない）。核酸配列は、別記されない限り、提供される場合には5'から3'の方向に表記される。本明細書に記載される方法は、単離された核酸の生成を提供する。本明細書に記載される方法は、付加的に、単離され精製された核酸の生成を提供する。提供される場合、ポリヌクレオチドの長さは塩基の数で記載され、nt（ヌクレオチド）、bp（塩基）、kb（キロベース）、またはGb（ギガベース）等で略記される。

40

【0019】

本明細書において提供されるのは、合成の（すなわち、デノボ合成された、または化学的に合成された）ポリヌクレオチドの生成のための方法と組成物である。オリゴ核酸、オリゴヌクレオチド、オリゴ、およびポリヌクレオチドの用語は、全体にわたって同意語として規定される。本明細書で記載される合成されたポリヌクレオチドのライブラリは、総体として1つ以上の遺伝子または遺伝子フラグメントをコードする複数のポリヌクレオチドを含み得る。いくつかの例では、ポリヌクレオチドライブラリはコード配列または非コード配列を含む。いくつかの例では、ポリヌクレオチドライブラリは複数のcDNA配列

50

をコードする。cDNA配列の基盤となる参照遺伝子配列はイントロンを含んでもよく、しかしcDNA配列はイントロンを除外する。本明細書に記載されるポリヌクレオチドは、有機体からの遺伝子または遺伝子フラグメントをコードし得る。典型的な有機体として、限定されないが、原核生物（例えば細菌）と真核生物（例えばマウス、ウサギ、ヒト、およびヒト以外の霊長類）があげられる。いくつかの例では、ポリヌクレオチドライブラリは1つ以上のポリヌクレオチドを含み、1つ以上のポリヌクレオチドの各々は多数のエクソンに関する配列をコードする。本明細書に記載されるライブラリ内の各ポリヌクレオチドは、異なる配列、すなわち非同一の配列をコードし得る。いくつかの例では、本明細書に記載されるライブラリ内の各ポリヌクレオチドは、ライブラリ内の他のポリヌクレオチドの配列に相補的な少なくとも1つの部分を含む。本明細書に記載されるポリヌクレオチド配列は、別段の定めのない限り、DNAまたはRNAを含み得る。本明細書に記載されるポリヌクレオチドライブラリは、少なくとも10、20、50、100、200、500、1,000、2,000、5,000、10,000、20,000、30,000、50,000、100,000、200,000、500,000、1,000,000、または1,000,000より多くのポリヌクレオチドを含み得る。本明細書に記載されるポリヌクレオチドライブラリは、10、20、50、100、200、500、1,000、2,000、5,000、10,000、20,000、30,000、50,000、100,000、200,000、500,000、または1,000,000以下のポリヌクレオチドを有し得る。本明細書に記載されるポリヌクレオチドライブラリは、10～500、20～1000、50～2000、100～5000、500～10,000、1,000～5,000、10,000～50,000、100,000～500,000、または50,000～1,000,000のポリヌクレオチドを含み得る。本明細書に記載されるポリヌクレオチドライブラリは約370,000；400,000；500,000、またはより多くの異なるポリヌクレオチドを含み得る。

【0020】

本明細書において提供されるのは、合成の（すなわちデノボ合成された）遺伝子のための方法と組成物である。合成遺伝子を含むライブラリは、PCA/非PCA遺伝子組立法または階層的遺伝子組立法等の、本明細書でさらに詳細に別記される様々な方法によって構築されてもよく、より大きなDNAユニット（すなわちシャシー（chassis））を生成するために、2つ以上の二本鎖ポリヌクレオチドを組み合わせる（縫い合わせる）。大きな構築物のライブラリは、少なくとも1、1.5、2、3、4、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90、100、125、150、175、200、250、300、400、500kbの長さ、またはより長いポリヌクレオチドを含み得る。大きな構築物は、約5000、10000、20000、または50000塩基対の独立して選択された上限によって、制限することができる。非リボソームのペプチド（NRP）をコードする配列、非リボソームのペプチドシンターゼ（NRPS）モジュールをコードする配列、および合成変異体を含む、ヌクレオチド配列をコードするポリペプチドセグメント、抗体等の他の調節タンパク質のポリペプチドセグメント、例えばプロモーター、転写因子、エンハンサー、siRNA、shRNA、RNAi、miRNA、microRNA由来の小核小体RNA、または対象となる機能的または構造的DNAまたはRNAユニット等の、調節配列等の、非コードDNAまたはRNAを含む、他のタンパク質ファミリーからのポリペプチドセグメント、の任意の数の合成。下記はポリヌクレオチドの非限定な例である：遺伝子または遺伝子フラグメントのコード領域または非コード領域、遺伝子間DNA、連鎖解析により規定された遺伝子座、エクソン、イントロン、メッセンジャーRNA（mRNA）、転移RNA、リボソームRNA、siRNA（short interfering RNA）、小ヘアピンRNA（shRNA）、マイクロRNA（miRNA）、核小体低分子RNA、リボザイム、相補的DNA（cDNA）、これは、通常はメッセンジャーRNA（mRNA）の逆転写、または増幅によって得られるmRNAのDNA表示である；合成または増幅によって生成されたDNA分子、ゲノムDNA、組換えポリヌクレオチド、分岐ポリヌクレオチド、プ

10

20

30

40

50

ラスミド、ベクター、任意の配列の単離されたDNA、任意の配列の単離されたRNA、核酸プローブ、およびプライマー。本明細書で言及される遺伝子または遺伝子フラグメントをコードするcDNAは、対応するゲノム配列に見られる介在イントロン配列なしに、エクソン配列をコードする少なくとも1つの領域を含み得る。代替的に、cDNAに対応するゲノム配列は、そもそもイントロン配列を欠いている場合もある。

【0021】

増幅反応のための小さなポリヌクレオチド集団のデノボ合成

【0022】

面、例えばプレートからのポリヌクレオチドの合成の方法が、本明細書に記載される。いくつかの例では、ポリヌクレオチドは、ポリヌクレオチド拡張のための遺伝子座のクラスタ上で合成され、放出され、次に増幅反応、例えばPCRにさらされる。クラスタからのポリヌクレオチドの合成の典型的なワークフローが図2に示される。シリコンプレート(201)は多数のクラスタ(203)を含む。各クラスタ内には多数の遺伝子座(221)がある。ポリヌクレオチドは、クラスタ(203)からのプレート(201)上でデノボ合成される(207)。ポリヌクレオチドは、開裂され(211)、プレートから取り除かれて(213)、放出されたポリヌクレオチド(215)の集団を形成する。放出されたポリヌクレオチド(215)の集団はその後、増幅されて(217)、増幅されたポリヌクレオチド(219)のライブラリを形成する。

10

【0023】

本明細書で提供される方法において、クラスタ上で増幅されたポリヌクレオチドの合成は、そのようなクラスタ構造のない構造の全表面にわたるポリヌクレオチドの増幅と比較して、ポリヌクレオチド表示に対する制御の向上を提供する。いくつかの例では、ポリヌクレオチド拡張のための遺伝子座のクラスタ構造を有する面から合成されたポリヌクレオチドの増幅は、大きなポリヌクレオチド集団の繰り返しの合成によって、表示に対する負の効果に打ち勝つようにする。大きなポリヌクレオチド集団の繰り返しの合成による表示に対する典型的な負の効果は、限定されないが、高/低GC含量に起因する増幅バイアス、反復配列、後端アデニン、二次構造、標的配列結合への親和性、またはポリヌクレオチド配列中の修飾ヌクレオチドを含む。

20

【0024】

クラスタ構造のないプレート全体にわたるポリヌクレオチドの増幅とは対照的に、クラスタ増幅は、平均値付近のより緊密な分布を結果としてもたらすことができる。例えば、100,000の読み取りが無作為にサンプリングされると、1つの配列につき平均8の読み取りが、平均値から約1.5Xの分布を有するライブラリを生み出すだろう。場合によっては、単一のクラスタ増幅は、平均値から多くとも約1.5X、1.6X、1.7X、1.8X、1.9X、または2.0Xに帰結する。場合によっては、単一のクラスタ増幅は、平均値から少なくとも約1.0X、1.2X、1.3X、1.5X、1.6X、1.7X、1.8X、1.9X、または2.0Xに帰結する。

30

【0025】

プレート全体に及ぶ増幅と比較した場合、本明細書に記載されるクラスタ増幅法は、同等の配列表示のためのより少ないシーケンシングを必要とするポリヌクレオチドライブラリをもたらしすることができる。いくつかの例では、少なくとも10%、少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも80%、少なくとも90%、または少なくとも95%より少ないシーケンシングが必要である。いくつかの例では最大10%、最大20%、最大30%、最大40%、最大50%、最大60%、最大70%、最大80%、最大90%、または最大95%より少ないシーケンシングが必要である。時には30%より少ないシーケンシングが、プレート全体に及ぶ増幅と比較して、クラスタ増幅後に必要である。いくつかの例でのポリヌクレオチドのシーケンシングは、次世代シーケンシング等による高性能配列解読装置によって検証される。シーケンシングライブラリのシーケンシングは、限定されないが、単一分子のリアルタイム(SMRT)シーケンシング、P o l o n y シーケンシング

40

50

、結紮によるシーケンシング、可逆的ターミネーターシーケンシング、プロトン検出シーケンシング、イオン半導体シーケンシング、ナノポアシーケンシング、電子シーケンシング、ピロシーケンス、Maxam-Gilbertシーケンシング、連鎖停止反応（例えばサンガー法）シーケンシング、+Sシーケンシング、または合成によるシーケンシングを含む、任意の適切なシーケンシング技術で行なうことができる。単一のヌクレオチドまたはポリヌクレオチドが特定され、または「読み取られる」回数は、シーケンシング深度または読み取り深度として定義される。場合によっては、読み取り深度は、倍単位でのカバー率、例えば55倍（または55X）のカバー率として言及され、随意に塩基の割合を記載する。

【0026】

いくつかの例では、プレート全体に及び増幅と比較して、クラスタ構造からの増幅は、より少ないドロップアウト、または増幅産物のシーケンシング後に検出されない配列をもたらす。ドロップアウトはATおよび/またはGCのものであり得る。いくつかの例では、ドロップアウトの数は、多くともポリヌクレオチド集団の約1%、2%、3%、4%、または5%である。場合によっては、ドロップアウトの数は0である。

【0027】

本明細書に記載されるクラスタは、ポリヌクレオチド合成のための分離した重複しない遺伝子座の集合を含む。クラスタは、約50~1000、75~900、100~800、125~700、150~600、200~500、または300~400の遺伝子座を含むことができる。いくつかの例では、各クラスタは121の遺伝子座を含む。いくつかの例では、各クラスタは、約50~500、50~200、100~150の遺伝子座を含む。いくつかの例では、各クラスタは、少なくとも約50、100、150、200、500、1000、またはより多くの遺伝子座を含む。いくつかの例では、単一のプレートは、100、500、10000、20000、30000、50000、100000、500000、700000、1000000、またはより多くの遺伝子座を含む。遺伝子座は、スポット、ウェル、マイクロウェル、チャネル、またはポストであり得る。いくつかの例では、各クラスタは、少なくとも1X、2X、3X、4X、5X、6X、7X、8X、9X、10X、またはより多くの、同一の配列を有するポリヌクレオチドの拡張を支持する余剰な別個の特徴を有する。

【0028】

制御された化学量論を有するポリヌクレオチドライブラリの設計

【0029】

本明細書において、ポリヌクレオチドライブラリの設計と合成のための方法が提供され、各ポリヌクレオチド種（すなわち、ライブラリの他のポリヌクレオチドとは異なる配列を有する）の量（または化学量論）は、望ましい結果が下流での適用に対して制御されるように、あらかじめ決定された量に調節される。そのため、本明細書において提供されるのは、ポリヌクレオチド種化学量論の制御され、あらかじめ決定された修飾のための方法である。例えば、増幅反応後のポリヌクレオチド種分布は、本明細書に記載される方法を使用するために制御され得る。ポリヌクレオチド種分布は、例えば、シーケンシング解析等のためのハイブリダイゼーションベースのアッセイのためのポリヌクレオチドのパネルを使用して、標的配列の高度に均一な捕捉を提供するために、あらかじめ選択される。さらに、本明細書に記載される方法は、典型的には特定の「問題のある」ポリヌクレオチド配列に起因する不均一な増幅産物または捕捉産物をもたらす1つ以上の配列特徴を伴う配列のポリヌクレオチドライブラリを設計する工程を提供し、ここで「問題のある」ポリヌクレオチド配列は、ポリヌクレオチドライブラリの適用におけるバイアスの生成に係る1つ以上の特性を含む。本明細書に記載される方法を使用する化学量論の制御のための典型的な「問題のある」ポリヌクレオチド配列特性は、限定されないが、高いまたは低いGCあるいはAT含量、反復配列、後端アデニン、二次構造、（増幅、濃縮または検出に関する）標的配列結合への親和性、安定性、融解温度、生物学的活性、より大きなフラグメントに組立てるための性能、修飾されるヌクレオチドまたはヌクレオチドアナログを含

10

20

30

40

50

む配列、または予測データまたは実験データに基づいて配列の第2のポリヌクレオチドライブラリを生成するための他のポリヌクレオチド特性、を含む。いくつかの例では、配列のライブラリは、化学量論の制御のために取得され、1つ以上の配列特徴に基づいて2つ以上のあらかじめ決定されたグループ（ビン）に整理され、またはクラスタリング（ビン入れ）される。いくつかの例では、2つ以上のビンは個々のユニーク配列を表す。いくつかの例では、ビンは、各々が全配列の少なくとも10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、または少なくとも99%を含む、1つ以上の規定された配列特徴に基づいて、値の範囲を表示する。いくつかの例では、ビンは、各々が全配列の多くとも10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、または100%未満を含む、数値の範囲を表す。一例では、ビンは%GCによって規定されてもよく、多数のビンが5%のインクリメント（例えば25 - 29%、30 - 34%、35 - 39%等）で25 - 75%を表し、あるビンは25%未満を表し、またあるビンは75%より多くのGC含量を表す。各ビンにおける全ての配列に対する分子の化学量論を表す、各ビンに関する存在値が割り当てられる。いくつかの例では、存在値は最初に100に設定され、ビンあたりの配列の等しい表示をもたらす。いくつかの例では、化学量論の制御には、各ビンに対して存在値を増加させる、減少させる、または維持するために、得られた適用バイアスデータを使用することによって行われる。規格に一致するように配列の豊富量を適合させる他の方法も使用される。いくつかの例では、事前に取得された分布は、最初の存在値を判定するために使用される。

10

20

30

40

50

【0030】

いくつかの例では、適用バイアスデータは、予測的アルゴリズムによって得られる。適用バイアスデータは経験的に得られてもよく、または制御されていない、または事前に制御された化学量論ライブラリから得られてもよい。例えば、適用バイアスデータは、ポリヌクレオチドライブラリの増幅から得られる；増幅後のビンあたりのポリヌクレオチドの頻度は、%GC含量の関数として増幅バイアスを設定するために、%GCビンに対してプロットされる。他の例において、適用バイアスデータは、ポリヌクレオチドプローブラリを用いて標的配列の濃縮後に次世代シーケンシング（NGS）データから得られる；標的遺伝子あたりの読み取りは、ビンへとプローブ配列を選別するために使用される；標的遺伝子ごとの読み取りは、ポリヌクレオチドプローブ配列の関数としてNGSシーケンシングバイアスを設定するために、NGS読み取りビンの数に対してプロットされる。他の例において、適用バイアスデータは、ポリヌクレオチドライブラリを含むベクターでの細胞処理後に、蛍光等の細胞アッセイ出力から得られるだろう；蛍光細胞で識別された配列ごとの読み取りは、ビンへとプローブ配列を選別するために使用される；配列ごとの読み取りは、ポリヌクレオチドプローブ配列の関数としてバイアスを設定するために、読み取りビンの数に対してプロットされる。

【0031】

化学量論を制御した後に、修飾された配列ライブラリが、ポリヌクレオチドの制御された化学量論ライブラリを生成するために合成される。いくつかの例では、制御された化学量論ライブラリは、下流での適用に使用される。いくつかの例では、制御された化学量論ポリヌクレオチドライブラリを用いた下流での適用からのデータは、ライブラリの化学量論的修飾の追加ラウンドを実施するために使用される。

【0032】

GC含量の制御された化学量論を用いたポリヌクレオチドライブラリの生成

【0033】

本明細書で提供されるのは、GC含量、反復配列、後端アデニン、二次構造、標的配列結合への親和性、または予測データあるいは実験データに基づいて、ポリヌクレオチドの第2の集団を生成するための修飾されたヌクレオチド等の、規定された特性を有するポリヌクレオチドライブラリを合成するための方法である。例えば、ポリヌクレオチドライブラリが、増幅後に規定されたGC含量に帰結するように合成用を選択される場合、GC含量に依存する合成工程におけるライブラリ内のポリヌクレオチドに関する種の表示の調整

は、増幅後のポリヌクレオチド表示の改善をもたらす。ポリヌクレオチドライブラリの G C 含量は、少なくとも 10 %、20 %、30 %、40 %、50 %、60 %、70 %、80 %、90 %、95 %、または 95 % 以上であり得る。いくつかの例では、ポリヌクレオチドライブラリの G C 含量は、多くとも 10 %、20 %、30 %、40 %、50 %、60 %、70 %、80 %、90 %、95 %、または 100 % 未満である。場合によっては、G C 含量は、約 5 ~ 95 %、10 ~ 90 %、30 ~ 80 %、40 ~ 75 %、または 50 ~ 70 % の範囲である。

【0034】

本明細書に記載されるポリヌクレオチドライブラリは、それらの G C 含量に適合させてもよい。いくつかの例では、ポリヌクレオチドライブラリは高い G C 含量を好む。例えば、増加したポリヌクレオチド頻度が約 40 % から約 90 % の範囲の G C 含量を有するライブラリが設計される。いくつかの例では、ポリヌクレオチドライブラリは低い G C 含量を含む。例えば、増加したポリヌクレオチド頻度が約 10 % から約 60 % の範囲の G C 含量を有するライブラリが設計される。ライブラリは、高いおよび低い G C 含量を好むように設計することができる。例えば、増加したポリヌクレオチド頻度が、主として約 10 % から約 30 % の範囲、および約 70 % から約 90 % の G C 含量を有するライブラリを設計することができる。いくつかの例では、ライブラリは均一の G C 含量を好む。例えば、ポリヌクレオチド頻度は、約 10 % から約 90 % の範囲の G C 含量有し、均一である。いくつかの例では、ライブラリは、約 10 % から約 95 % の G C の割合でポリヌクレオチドを含む。いくつかの例では、本明細書に記載されるライブラリは、10 % から 30 % または 70 % から 90 % の G C の割合を有する、30 % より多くの異なるポリヌクレオチドを持つポリヌクレオチドを含む。いくつかの例では、本明細書に記載されるライブラリは、10 % から 30 % または 60 % から 90 % の G C の割合を有する、約 15 % 未満のポリヌクレオチドを持つポリヌクレオチドを含む。

【0035】

いくつかの場合において指定された G C 含量を有するポリヌクレオチドライブラリの生成は、異なる G C 含量を有する少なくとも 2 つのポリヌクレオチドライブラリを組み合わせることによって起こる。いくつかの例では、少なくとも 2、3、4、5、6、7、10、または 10 より多くのポリヌクレオチドライブラリが、指定された G C 含量を有するポリヌクレオチドの集団を生成するために組み合わせられる。場合によっては、2、3、4、5、6、7、または 10 以下のポリヌクレオチドライブラリが、指定された G C 含量を有する非同一のポリヌクレオチドの集団を生成するために組み合わせられる。

【0036】

いくつかの例では、G C 含量は、クラスタあたりより少数、またはより多くのポリヌクレオチドの合成によって適合される。例えば、少なくとも 25、50、100、200、300、400、500、600、700、800、900、1000、または 1000 より多くの非同一のポリヌクレオチドは、単一のクラスタ上で合成される。場合によっては、約 50、100、200、300、400、500、600 の、700、800、900、1000 以下の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。いくつかの例では、50 ~ 500 の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、100 ~ 200 の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、約 100、約 120、約 125、約 130、約 150、約 175、または約 200 の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。

【0037】

場合によっては、G C 含量は、可変長の非同一のポリヌクレオチドの合成によって適合される。例えば、合成された非同一のポリヌクレオチドの各々の長さは、少なくとも、または少なくとも約 10、15、20、25、30、35、40、45、50、100、150、200、300、400、500、2000 のヌクレオチド、またはそれ以上であり得る。合成された非同一のポリヌクレオチドの長さは、多くとも、または多くとも約 2

000、500、400、300、200、150、100、50、45、35、30、25、20、19、18、17、16、15、14、13、12、11、10のヌクレオチド、またはそれ以下でもよい。合成された非同一のポリヌクレオチドの各々の長さは、10～2000、10～500、9～400、11～300、12～200、13～150、14～100、15～50、16～45、17～40、18～35、および19～25であり得る。

【0038】

反復配列含量の制御された化学量論を用いたポリヌクレオチドライブラリの生成

【0039】

本明細書に記載されるポリヌクレオチドライブラリは、指定された反復配列分布で合成され得る。いくつかの例では、反復配列含量に関してポリヌクレオチドライブラリを適合させることで、結果としてポリヌクレオチド表示が改善する。

【0040】

反復配列は、単一のヌクレオチドの繰り返し、または2つ以上のヌクレオチドのブロックの繰り返しであり得る。いくつかの例では、反復配列は、少なくとも2、3、4、5、6、7、8、9、または少なくとも10のヌクレオチドである。いくつかの例では、反復配列は、多くとも2、3、4、5、6、7、8、9、または多くとも10のヌクレオチドである。いくつかの例では、ヌクレオチドのブロックは、少なくとも2、3、4、5、10、15、25、50、100、200、500、または少なくとも1000のヌクレオチドを含む。いくつかの例では、ヌクレオチドのブロックは、多くとも2、3、4、5、10、15、25、50、100、200、500、または多くとも1000のヌクレオチドを含む。反復配列は、より大きな合成ポリヌクレオチドの内部または末端位置に位置し得る。末端位置は、ポリヌクレオチドの5'、3'、または5'と3'末端の両方に近い場合もある。いくつかの例では、反復配列は、少なくとも1、2、3、4、5、6、7、8、9、または少なくとも10の末端のヌクレオチド以内にある。いくつかの例では、反復配列は、多くとも1、2、3、4、5、6、7、8、9、または多くとも10の末端のヌクレオチド以内にある。いくつかの例では、反復するヌクレオチドはアデニンである。いくつかの例では、反復配列は、ポリヌクレオチド末端、例えばポリアデニン末端に位置する。

【0041】

ポリヌクレオチドライブラリの反復配列含量は、少なくとも10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、または95%以上であり得る。いくつかの例では、ポリヌクレオチドライブラリの反復配列含量は、多くとも10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、または100%以下である。場合によっては、反復配列含量は、約5～95%、10～90%、30～80%、40～75%、または50～70%の範囲である。

【0042】

ポリヌクレオチドライブラリは、それらの反復配列含量に関して適合させることができる。いくつかの例では、ポリヌクレオチドライブラリは高い反復配列含量を好む。例えば、増加したポリヌクレオチド頻度が約40%から約90%の範囲の反復配列含量を有するライブラリが設計される。いくつかの例では、ポリヌクレオチドライブラリは低い反復配列含量を含む。例えば、増加したポリヌクレオチド頻度が約10%から約60%の範囲の反復配列含量を有するライブラリが設計される。ライブラリは、高いおよび低い反復配列含量を好むように設計することができる。例えば、増加したポリヌクレオチド頻度が、主として約10%と約30%の範囲、および約70%から約90%の反復配列含量を有するライブラリを設計することができる。いくつかの例では、ライブラリは均一の反復配列含量を好む。例えば、ポリヌクレオチド頻度は、約10%から約90%の範囲の反復配列含量を有し、均一である。いくつかの例では、ライブラリは、約10%から約95%の反復配列の割合でポリヌクレオチドを含む。いくつかの例では、本明細書に記載されるライブラリは、10%から30%または70%から90%の反復配列の割合を有する、30%よ

り多くの異なるポリヌクレオチドを持つポリヌクレオチドを含む。いくつかの例では、本明細書に記載されるライブラリは、10%から30%または60%から90%の反復配列の割合を有する、約15%未満のポリヌクレオチドを持つポリヌクレオチドを含む。

【0043】

いくつかの場合において指定された反復配列含量を有するポリヌクレオチドライブラリの生成は、異なる反復配列含量を有する少なくとも2つのポリヌクレオチドライブラリを組み合わせることによって起こる。いくつかの例では、少なくとも2、3、4、5、6、7、10、または10より多くのポリヌクレオチドライブラリが、指定された反復配列含量を有するポリヌクレオチドの集団を生成するために組み合わせられる。場合によっては、2、3、4、5、6、7、または10以下のポリヌクレオチドライブラリが、指定された反復配列含量を有する非同一のポリヌクレオチドの集団を生成するために組み合わせられる。

10

【0044】

いくつかの例では、反復配列含量は、クラスタあたりより少数、またはより多くのポリヌクレオチドの合成によって適合される。例えば、少なくとも25、50、100、200、300、400、500、600、700、800、900、1000、または1000より多くの非同一のポリヌクレオチドが、単一のクラスタ上で合成される。場合によっては、約50、100、200、300、400、500、600の、700、800、900、1000以下の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。いくつかの例では、50～500の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、100～200の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、約100、約120、約125、約130、約150、約175、または約200の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。

20

【0045】

場合によっては、反復配列含量は、可変長の非同一のポリヌクレオチドの合成によって適合される。例えば、合成された非同一のポリヌクレオチドの各々の長さは、少なくとも、または少なくとも約10、15、20、25、30、35、40、45、50、100、150、200、300、400、500、2000のヌクレオチド、またはそれ以上であり得る。合成された非同一のポリヌクレオチドの長さは、多くとも、または多くとも約2000、500、400、300、200、150、100、50、45、35、30、25、20、19、18、17、16、15、14、13、12、11、10のヌクレオチド、またはそれ以下でもよい。合成された非同一のポリヌクレオチドの各々の長さは、10～2000、10～500、9～400、11～300、12～200、13～150、14～100、15～50、16～45、17～40、18～35、および19～25であり得る。

30

【0046】

二次構造含量の制御された化学量論を用いたポリヌクレオチドライブラリの生成

【0047】

本明細書に記載されるポリヌクレオチドライブラリは、指定された二次構造含量で合成され得る。いくつかの例では、二次構造含量に関してポリヌクレオチドライブラリを適合させることで、結果としてポリヌクレオチド表示が改善する。

40

【0048】

二次構造は、ヘリックス（例えばアルファヘリックス）、ベータシート、ステムループ、シュードノット、ホモダイマー、またはヘテロダイマー等の構造を形成する1つ以上のポリヌクレオチド鎖中の3つ以上のヌクレオチドを含むことができる。ステムループはヘアピンループ、内部ループ、バルジ、またはマルチループであってもよい。二次構造の種類およびそれらの形成能力は、配列データから予測することができる。線形配列の二次構造への折り重ねまたはハイブリダイゼーションは、ポリヌクレオチドが個体担体に付着している間に、または溶液への分裂後に生じ得る。

50

【 0 0 4 9 】

ポリヌクレオチドライブラリの二次構造含量は、少なくとも 1 0 %、2 0 %、3 0 %、4 0 %、5 0 %、6 0 %、7 0 %、8 0 %、9 0 %、9 5 %、または 9 5 % 以上であり得る。いくつかの例では、ポリヌクレオチドライブラリの二次構造含量は、多くとも 1 0 %、2 0 %、3 0 %、4 0 %、5 0 %、6 0 %、7 0 %、8 0 %、9 0 %、9 5 %、または 1 0 0 % 以下である。場合によっては、二次構造含量は、約 5 ~ 9 5 %、1 0 ~ 9 0 %、3 0 ~ 8 0 %、4 0 ~ 7 5 %、または 5 0 ~ 7 0 % の範囲である。

【 0 0 5 0 】

ポリヌクレオチドライブラリは、それらの二次構造含量に適合させることができる。いくつかの例では、ポリヌクレオチドライブラリは高い二次構造含量を好む。例えば、増加したポリヌクレオチド頻度が約 4 0 % から約 9 0 % の範囲の二次構造含量を有するライブラリが設計される。いくつかの例では、ポリヌクレオチドライブラリは低い二次構造含量を含む。例えば、増加したポリヌクレオチド頻度が約 1 0 % から約 6 0 % の範囲の二次構造含量を有するライブラリが設計される。ライブラリは、高いおよび低い二次構造含量を好むように設計することができる。例えば、増加したポリヌクレオチド頻度が、主として約 1 0 % と約 3 0 % の範囲、および約 7 0 % から約 9 0 % の二次構造含量を有するライブラリを設計することができる。いくつかの例では、ライブラリは均一の二次構造含量を好む。例えば、ポリヌクレオチド頻度は、約 1 0 % から約 9 0 % の範囲の二次構造含量を有し、均一である。いくつかの例では、ライブラリは、約 1 0 % から約 9 5 % の二次構造の割合でポリヌクレオチドを含む。いくつかの例では、本明細書に記載されるライブラリは、1 0 % から 3 0 % または 7 0 % から 9 0 % の二次構造の割合を有する、3 0 % より多くの異なるポリヌクレオチドを持つポリヌクレオチドを含む。いくつかの例では、本明細書に記載されるライブラリは、1 0 % から 3 0 % または 6 0 % から 9 0 % の二次構造の割合を有する、約 1 5 % 未満のポリヌクレオチドを持つポリヌクレオチドを含む。

【 0 0 5 1 】

いくつかの場合において指定された二次構造含量を有するポリヌクレオチドライブラリの生成は、異なる反復配列含量を有する少なくとも 2 つのポリヌクレオチドライブラリを組み合わせることによって起こる。いくつかの例では、少なくとも 2、3、4、5、6、7、1 0、または 1 0 より多くのポリヌクレオチドライブラリが、指定された二次構造含量を有するポリヌクレオチドの集団を生成するために組み合わせられる。場合によっては、2、3、4、5、6、7、または 1 0 以下のポリヌクレオチドライブラリが、指定された二次構造含量を有する非同一のポリヌクレオチドの集団を生成するために組み合わせられる。

【 0 0 5 2 】

いくつかの例では、二次構造含量は、クラスタあたりより少数、またはより多くのポリヌクレオチドの合成によって適合される。例えば、少なくとも 2 5、5 0、1 0 0、2 0 0、3 0 0、4 0 0、5 0 0、6 0 0、7 0 0、8 0 0、9 0 0、1 0 0 0、または 1 0 0 0 以上の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。場合によっては、約 5 0、1 0 0、2 0 0、3 0 0、4 0 0、5 0 0、6 0 0 の、7 0 0、8 0 0、9 0 0、1 0 0 0 以下の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。いくつかの例では、5 0 ~ 5 0 0 の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、1 0 0 ~ 2 0 0 の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、約 1 0 0、約 1 2 0、約 1 2 5、約 1 3 0、約 1 5 0、約 1 7 5、または約 2 0 0 の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。

【 0 0 5 3 】

場合によっては、二次構造含量は、可変長の非同一のポリヌクレオチドの合成によって適合される。例えば、合成された非同一のポリヌクレオチドの各々の長さは、少なくとも、または少なくとも約 1 0、1 5、2 0、2 5、3 0、3 5、4 0、4 5、5 0、1 0 0、1 5 0、2 0 0、3 0 0、4 0 0、5 0 0、2 0 0 0 のヌクレオチド、またはそれ以上

であり得る。合成された非同一のポリヌクレオチドの長さは、多くとも、または多くとも約 2000、500、400、300、200、150、100、50、45、35、30、25、20、19、18、17、16、15、14、13、12、11、10 のヌクレオチド、またはそれ以下でもよい。合成された非同一のポリヌクレオチドの各々の長さは、10 ~ 2000、10 ~ 500、9 ~ 400、11 ~ 300、12 ~ 200、13 ~ 150、14 ~ 100、15 ~ 50、16 ~ 45、17 ~ 40、18 ~ 35、および 19 ~ 25 であり得る。

【0054】

配列含量の制御された化学量論を用いたポリヌクレオチドライブラリの生成

【0055】

いくつかの例では、ポリヌクレオチドライブラリは、望ましいポリヌクレオチド配列の指定された分布で合成される。いくつかの例では、特定の望ましい配列の濃縮のためにポリヌクレオチドライブラリを適合させることは、下流での適用の結果の改善をもたらす。

【0056】

1 つ以上の特異的配列は、下流での適用におけるそれらの評価に基づいて選択することができる。いくつかの例では、評価は、増幅、濃縮、または検出のための標的配列への結合親和性、安定性、融解温度、生物学的活性、より大きなフラグメントへの組立て能力、またはポリヌクレオチドの他の特性である。いくつかの例では、評価は実験による、または先の実験および / またはコンピュータアルゴリズムから予測される。

【0057】

ポリヌクレオチドライブラリの選択された配列は、配列の少なくとも 10 %、20 %、30 %、40 %、50 %、60 %、70 %、80 %、90 %、95 %、または 95 % 以上であり得る。いくつかの例では、ポリヌクレオチドライブラリの選択された配列は、配列の多くとも 10 %、20 %、30 %、40 %、50 %、60 %、70 %、80 %、90 %、95 %、または 100 % 未満である。場合によっては、選択された配列は、配列の約 5 ~ 95 %、10 ~ 90 %、30 ~ 80 %、40 ~ 75 %、または 50 ~ 70 % の範囲である。

【0058】

ポリヌクレオチドライブラリは、各選択された配列の頻度に適合される。いくつかの例では、ポリヌクレオチドライブラリは、より多くの選択された配列を好む。例えば、選択された配列の増加したポリヌクレオチド頻度が約 40 % から約 90 % の範囲であるライブラリが設計される。いくつかの例では、ポリヌクレオチドライブラリは、低い数の選択された配列を含む。例えば、選択された配列の増加したポリヌクレオチド頻度が約 10 % から約 60 % の範囲であるライブラリが設計される。ライブラリは、選択された配列のより高い、およびより低い頻度を好むように設計することができる。いくつかの例では、ライブラリは均一の配列表示を好む。例えば、ポリヌクレオチド頻度は、約 10 % から約 90 % の範囲の選択された配列の頻度に関して均一である。いくつかの例では、ライブラリは、配列の約 10 % から約 95 % の選択された配列頻度を伴うポリヌクレオチドを含む。

【0059】

いくつかの場合において指定された選択された配列の頻度を有するポリヌクレオチドライブラリの生成は、異なる選択された配列の頻度を有する少なくとも 2 つのポリヌクレオチドライブラリを組み合わせることによって起こる。いくつかの例では、少なくとも 2、3、4、5、6、7、10、または 10 より多くのポリヌクレオチドライブラリが、指定された選択された配列の頻度を有するポリヌクレオチドの集団を生成するために組み合わせられる。場合によっては、2、3、4、5、6、7、または 10 以下のポリヌクレオチドライブラリが、指定された選択された配列の頻度を有する非同一のポリヌクレオチドの集団を生成するために組み合わせられる。

【0060】

いくつかの例では、選択された配列の頻度は、クラスターあたりより少数、またはより多くのポリヌクレオチドの合成によって適合される。例えば、少なくとも 25、50、10

10

20

30

40

50

0、200、300、400、500、600、700、800、900、1000、または1000より多くの非同一のポリヌクレオチドが、単一のクラスタ上で合成される。場合によっては、約50、100、200、300、400、500、600の、700、800、900、1000以下の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。いくつかの例では、50～500の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、100～200の非同一のポリヌクレオチドが単一のクラスタ上で合成される。いくつかの例では、約100、約120、約125、約130、約150、約175、または約200の非同一のポリヌクレオチドが、単一のクラスタ上で合成される。

【0061】

場合によっては、選択された配列の頻度は、可変長の非同一のポリヌクレオチドの合成によって適合される。例えば、合成された非同一のポリヌクレオチドの各々の長さは、少なくとも、または少なくとも約10、15、20、25、30、35、40、45、50、100、150、200、300、400、500、2000のヌクレオチド、またはそれ以上であり得る。合成された非同一のポリヌクレオチドの長さは、多くとも、または多くとも約2000、500、400、300、200、150、100、50、45、35、30、25、20、19、18、17、16、15、14、13、12、11、10のヌクレオチド、またはそれ以下でもよい。合成された非同一のポリヌクレオチドの各々の長さは、10～2000、10～500、9～400、11～300、12～200、13～150、14～100、15～50、16～45、17～40、18～35、および19～25であり得る。

【0062】

ポリヌクレオチドプローブ構造

【0063】

ポリヌクレオチドプローブライブラリは、サンプルポリヌクレオチドのより大きな集団において特定の標的配列を濃縮するために使用することができる。いくつかの例では、ポリヌクレオチドプローブの各々は、1つ以上の標的配列に相補的な標的結合配列、1つ以上の非標的結合配列、およびユニバーサルプライマー結合部位等の1つ以上のプライマー結合部位を含む。いくつかの例では、相補的あるいは少なくとも部分的に相補的な標的結合配列は、標的配列に結合する（ハイブリダイズする）。ユニバーサルプライマー結合部位等のプライマー結合部位は、プローブライブラリの全ての構成員、または構成員の部分母集団の同時増幅を促進する。いくつかの例では、プローブはさらにバーコードまたはインデックス配列を含む。バーコードは核酸配列であり、バーコードに関連するポリヌクレオチドのいくつかの特徴を識別できるようにする。シーケンシングの後、バーコード領域は、コード領域に関係する特徴またはサンプル源を識別するための指標を提供する。バーコードは、十分な度合の識別を可能にするために、例えば少なくとも約3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、またはそれ以上の塩基の長さの、適切な長さで設計することができる。約2、3、4、5、6、7、8、9、10、またはそれ以上のバーコード等の多数のバーコードは、非バーコード配列によって随意に分離された同じ分子上で使用されてもよい。いくつかの実施形態では、複数のバーコードにある各バーコードは、少なくとも3つの塩基位置、少なくとも約3、4、5、6、7、8、9、10、またはそれ以上の位置等において、複数のうちの他のバーコードごとに異なっている。いくつかの例では、対象の標的配列の続く捕捉のためのプローブを形成するために、ポリヌクレオチドは、小分子、ペプチド、抗原、金属またはタンパク質等の1つ以上の分子（または親和性）タグに結合される。いくつかの例では、ハイブリダイゼーション可能な相補的な標的結合配列を有する2つのプローブが、二本鎖プローブ対を形成する。

【0064】

本明細書に記載のプローブは、ゲノム内の配列である標的配列に相補的であり得る。本明細書に記載のプローブは、ゲノム内のエクソーム配列である標的配列に相補的であり得る。本明細書に記載のプローブは、ゲノム内のイントロン配列である標的配列に相補的であり得る。いくつかの例では、プローブは、標的配列に相補的な標的結合配列、および標的に相補的ではない少なくとも1つの非標的結合配列を含む。いくつかの例では、プローブの標的結合配列は、約120ヌクレオチドの長さ、または少なくとも10、15、20、25、50、75、100、110、120、125、140、150、160、175、200、300、400、500、500以上のヌクレオチドの長さである。標的結合配列は、いくつかの例では、10、15、20、25、50、75、100、125、150、175、200以下、または500以下のヌクレオチドの長さである。プローブの標的結合配列は、いくつかの例では、約120ヌクレオチドの長さであり、または約10、15、20、25、40、50、60、70、80、85、87、90、95、97、100、105、110、115、117、118、119、120、121、122、123、124、125、126、127、128、129、130、135、140、145、150、155、157、158、159、160、161、162、163、164、165、166、167、168、169、170、175、180、190、200、210、220、230、240、250、300、400、または約500のヌクレオチドの長さである。標的結合配列は、いくつかの例では、約20～約400のヌクレオチドの長さであり、または約30～約175、約40～約160、約50～約150、約75～約130、約90～約120、または約100～約140のヌクレオチドの長さである。プローブの非標的結合配列は、いくつかの例では、少なくとも約20のヌクレオチドの長さであり、または少なくとも約1、5、10、15、17、20、23、25、50、75、100、110、120、125、140、150、160、175、または約175より多いヌクレオチドの長さである。非標的結合配列は、約5、10、15、20、25、50、75、100、125、150、175以下、または約200以下のヌクレオチドの長さである。プローブの非標的結合配列は、約20のヌクレオチドの長さ、または約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、25、40、50、60、70、80、90、100、110、120、130、140、150、または約200のヌクレオチドの長さである。いくつかの例では非標的結合配列は、約1～約250のヌクレオチドの長さであり、または約20～約200、約10～約100、約10～約50、約30～約100、約5～約40、または約15～約35のヌクレオチドの長さである。非標的結合配列はしばしば、標的配列に相補的でない配列を含み、および/またはプライマーを結合するのに使用されない配列を含む。いくつかの例では、非標的結合配列は、単一のヌクレオチド、例えばポリアデニン、ポリチミジンの反復を含む。プローブはしばしば、非標的結合配列を全く含まない、または少なくとも1つ含む。いくつかの例では、プローブは1つまたは2つの非標的結合配列を含む。非標的結合配列は、プローブ内の1つ以上の標的結合配列に隣接し得る。例えば、非標的結合配列は、プローブの5'または3'末端にある。いくつかの例では、非標的結合配列は、分子タグまたはスペーサーに付けられている。

【0065】

いくつかの例では、非標的結合配列はプライマー結合部位であり得る。プライマー結合部位はしばしば、各々が少なくとも約20のヌクレオチドの長さであり、または約10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、または少なくとも約40のヌクレオチドの長さである。いくつかの例では各プライマー結合部位は、約10、12、14、16、18、20、22、24、26、28、30、32、34、36、38以下、または約40以下のヌクレオチドの長さである。いくつかの例では各プライマー結合部位は、約10～約50のヌクレオチドの長さであり、または約15～約40である、約20～約30、約10～約40、約10～約30、約30～約50、または約20～約60のヌクレオチドの長さである。いくつかの例では、ポリヌ

クレオチドプローブは少なくとも2つのプライマー結合部位を含む。いくつかの例では、プライマー結合部位はユニバーサルプライマー結合部位である場合もあり、ここで全てのプローブは、これらの位置に同一のプライマー結合配列を含む。いくつかの例では、特定の配列およびその逆の補体（例えばゲノムDNAの領域）を標的とする1対のポリヌクレオチドは、図3Aに(300)で表され、第1の標的結合配列(301)、第2の標的結合配列(302)、第1の非標的結合配列(303)、および第2の非標的結合配列(304)を含む。例えば、特定の配列（例えばゲノムDNAの領域）に相補的な1対のポリヌクレオチドプローブ。

【0066】

いくつかの例では、標的結合配列(301)は、第1、第2の標的結合配列(302)の逆の補体である。いくつかの例では、両方の標的結合配列が、増幅に先立って化学的に合成される。代替的な構成では、特定の配列とその逆の補体（例えばゲノムDNAの領域）を標的とする1対のポリヌクレオチドプローブは、図3Bに(305)で表され、第1の標的結合配列(301)、第2の標的結合配列(302)、第1の非標的結合配列(303)、第2の非標的結合配列(304)、第3の非標的結合配列(306)、および第4の非標的結合配列(307)を含む。いくつかの例では、標的結合配列(301)は第1、第2の標的結合配列(302)の逆の補体である。いくつかの例では、1つ以上の非標的結合配列は、ポリアデニンまたはポリチミジンを含む。

【0067】

いくつかの例では、対に含まれる両方のプローブは、少なくとも1つの分子タグで標識される。いくつかの例では、PCRは、増幅中にプローブ上に(分子タグを含むプライマーを介して)分子タグを導入するために使用される。いくつかの例では、分子タグは、1つ以上のビオチン、葉酸、ポリヒスチジン、FLAGタグ、グルタチオン、または規格と一致する他の分子タグを含む。いくつかの例では、プローブは5'末端で標識される。いくつかの例では、プローブは3'末端で標識される。いくつかの例では、5'と3'末端の両方が、分子タグで標識される。いくつかの例では、対にある第1のプローブの5'末端は、少なくとも1つの分子タグで標識され、および対にある第2のプローブの3'末端は、少なくとも1つの分子タグで標識される。いくつかの例では、スペーサーは、プローブの1つ以上の分子タグと核酸の間に存在する。いくつかの例では、スペーサーは、アルキル、ポリオールまたはポリアミノ鎖、ペプチド、またはポリヌクレオチドを含み得る。いくつかの例では、プローブ標的核酸を捕捉するのに使用される個体担体は、ビーズまたは面である。いくつかの例における個体担体は、ガラス、プラスチック、または分子タグを結合する捕捉部分を含むことができる他の物質を含む。いくつかの例では、ビーズは磁気ビーズである。例えば、ビオチンで標識されたプローブは、ストレプトアビジンを含む磁気ビーズで捕捉される。標的配列へのプローブの結合を可能にするために、プローブは核酸のライブラリと接触させられる。いくつかの例では、ブロッキングポリ核酸が、標的核酸に付けられた1つ以上のアダプター配列へのプローブの結合を防ぐために加えられる。いくつかの例では、ブロッキングポリ核酸は、1つ以上の核酸アナログを含む。いくつかの例では、ブロッキングポリ核酸は、1つ以上の位置でチミンの代わりにウラシルを有する。

【0068】

本明細書に記載されるプローブは、1つ以上の標的核酸配列に結合する相補的な標的結合配列を含み得る。いくつかの例では、標的配列は任意のDNAまたはRNAの核酸配列である。いくつかの例では、標的配列はプローブ挿入より長い場合もある。いくつかの例において、標的配列はプローブ挿入より短い場合もある。いくつかの例において、標的配列はプローブ挿入と同じ長さである場合もある。例えば、標的配列の長さは、少なくとも、または少なくとも約2、10、15、20、25、30、35、40、45、50、100、150、200、300、400、500、1000、2000、5,000、12,000、20,000ヌクレオチド、またはそれ以上である。標的配列の長さは、多くとも、または多くとも約20,000、12,000、5,000、2,000、1,

10

20

30

40

50

【 0 0 6 9 】

【 0 0 7 0 】

本明細書に記載されるプローブライブラリは、少なくとも10、20、50、100、200、500、1,000、2,000、5,000、10,000、20,000、50,000、100,000、200,000、500,000、1,000,000、または1,000,000より多くのプローブを含み得る。プローブライブラリは、10、20、50、100、200、500、1,000、2,000、5,000、10,000、20,000、50,000、100,000、200,000、500,000、1,000,000、

00以下、または1, 000, 000以下のプローブを有し得る。プローブライブラリは、10~500、20~1000、50~2000、100~5000、500~10,000、1,000~5,000、10,000~50,000、100,000~500,000、または50,000~1,000,000のプローブを含み得る。プローブライブラリは約370,000; 400,000; 500,000、またはより多くの異なるプローブを含み得る。

【0071】

次世代シーケンシングの応用

【0072】

ポリヌクレオチドライブラリの下流での適用は次世代シーケンシングを含み得る。例えば、制御された化学量論ポリヌクレオチドプローブライブラリでの標的配列の濃縮は、より効果的なシーケンシングを結果としてもたらす。標的を捕捉する、または標的にハイブリダイズするためのポリヌクレオチドライブラリの性能は、効率、精度、および正確さを記す多くの異なるメトリックによって定義され得る。例えば、Picardのメトリックは、HSライブラリ（読み取り対から算出される、標的部位に対応するライブラリ内の特有の分子の数）、平均標的カバー率（特定のカバーレベルに達する塩基の割合）、カバー率の深度（任意のヌクレオチドを含む読み取りの数）、Fold enrichment（標的に固有にマッピングする配列読み取り/全サンプルにマッピングする読み取りを、全サンプルの長さ/標的長さに掛ける）、パーセントオフベイト塩基（プローブ/ベイトの塩基に対応しない塩基のパーセント）、標的の利用可能な塩基、ATまたはGCのドロップアウト率、「fold 80 base penalty」（平均カバー率レベルまで0でない標的の80パーセントを上昇させるために必要なfold over-coverage）、パーセント0カバー標的、PF読み取り（質フィルタを通過する読み取りの数）、パーセント選択塩基（合計の並んだ塩基で割った、ベイト上の塩基とベイト付近の塩基の合計）、パーセント複製、または規格と一致する他の変数等を含む変数を含む。

【0073】

読み取り深度（シーケンシング深度、またはサンプリング）は、第2の核酸フラグメント（「読み取り」）が配列のために取得される合計回数を表す。理論上の読み取り深度は、同じヌクレオチドが読み取られる予測回数として定義され、読み取りが理想化されたゲノム全体にわたって完全に分散していると仮定する。読み取り深度は%カバー率（またはカバー幅）の関数として表わされる。例えば、完全に分散した100万の塩基ゲノムの1000万の読み取りは、理論上、配列の100%の10x読み取り深度をもたらす。実験的に、より多くの読み取り（より高度な理論上の読み取り深度、またはオーバーサンプリング）が、標的配列の割合に関する望ましい読み取り深度を得るために必要である場合もある。より少ない合計読み取りが、標的配列の望ましい%を超える読み取りの許容可能な数を伴う実験結果を得るために求められるため、制御された化学量論プローブライブラリでの標的配列の濃縮は、下流のシーケンシングの効率を高める。例えば、いくつかの例では、標的配列の55xの理論上の読み取り深度は、配列の少なくとも90%の少なくとも30xのカバー率をもたらす。いくつかの例では、標的配列の55x以下の理論上の読み取り深度は、配列の少なくとも80%の少なくとも30xの読み取り深度をもたらす。いくつかの例では、標的配列の55x以下の理論上の読み取り深度は、配列の少なくとも95%の少なくとも30xの読み取り深度をもたらす。いくつかの例では、標的配列の55x以下の理論上の読み取り深度は、配列の少なくとも98%の少なくとも30xの読み取り深度をもたらす。いくつかの例では、標的配列の55x以下の理論上の読み取り深度は、配列の少なくとも98%の少なくとも20xの読み取り深度をもたらす。いくつかの例では、標的配列の55x以下の理論上の読み取り深度は、配列の少なくとも98%の少なくとも5xの読み取り深度をもたらす。標的とのハイブリダイゼーション中にプローブの濃度を高めることで、読み取り深度の増加をもたらしことができる。いくつかの例では、プローブの濃度は、少なくとも1.5x、2.0x、2.5x、3x、3.5x、4x、5x、または5x以上に増加する。いくつかの例では、プローブ濃度を高めることで、結

10

20

30

40

50

果として1000%の増加、または読み取り深度における20%、30%、40%、50%、60%、70%、80%、90%、100%、200%、300%、500%、750%、1000%、または1000%以上の増加がもたらされる。いくつかの例では、プローブ濃度を3Xまで高めることで、読み取り深度における1000%の増加がもたらされる。

【0074】

オンターゲット比率は、望ましい標的配列に対応するシーケンシング読み取りの割合を表す。いくつかの例では、制御された化学量論ポリヌクレオチドプローブライブラリは、少なくとも30%、または少なくとも35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、または少なくとも90%のオンターゲット比率をもたらし、標的核酸との接触中にポリヌクレオチドプローブの濃度を高めることで、オンターゲット比率の増加がもたらされる。いくつかの例では、プローブの濃度は、少なくとも1.5x、2.0x、2.5x、3x、3.5x、4x、5x、または5x以上に増加する。いくつかの例では、プローブ濃度を高めることで、少なくとも20%の増加、またはオンターゲット結合における10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、200%、300%、または少なくとも500%の増加がもたらされる。いくつかの例では、プローブ濃度を3Xまで高めることで、オンターゲット比率における20%の増加がもたらされる。

10

【0075】

カバー率の均一性は、場合によっては、標的配列の同定（アイデンティティ）の関数としての読み取り深度として算出される。より高いカバー率の均一性は、望ましい読み取り深度を得るために必要とされる、より少数のシーケンシング読み取りをもたらし、例えば、標的配列の特性は、読み取り深度、例えば高いまたは低いGCまたはAT含量、反復配列、後端アデニン、二次構造、標的配列結合（増幅、濃縮、または検出）に対する親和性、安定性、融解温度、生物学的活性、より大きなフラグメントへの組立て能力、修飾ヌクレオチドまたはヌクレオチドアナログを含む配列、またはポリヌクレオチドの他の特性に影響し得る。制御された化学量論ポリヌクレオチドプローブライブラリでの標的配列の濃縮は、シーケンシング後に、より高いカバー率の均一性をもたらし、いくつかの例では、配列の95%は、平均ライブラリ読み取り深度の1x以内、または平均ライブラリ読み取り深度の0.1、0.2、0.5、0.7、1、1.2、1.5、1.7、または約2x以内である読み取り深度を有する。いくつかの例では、配列の80%、85%、90%、95%、97%または99%は、平均値の1x以内である読み取り深度を有する。

20

30

【0076】

ポリヌクレオチドプローブライブラリでの標的核酸の濃縮

【0077】

本明細書に記載されるプローブライブラリは、サンプルポリヌクレオチドの集団にある標的ポリヌクレオチドを濃縮するために、様々な下流での適用に使用されてもよい。いくつかの例では、サンプルは1つ以上のソースから得られ、およびサンプルポリヌクレオチドの集団は、当技術分野で既知の従来技術を使用して単離される。サンプルは、（非限定な例として）唾液、血液、組織、皮膚、または完全に合成のソース等の生物学的ソースから得られる。サンプルから得られる複数のポリヌクレオチドはフラグメント化され、末端修復され、そしてアデニル化されて、二本鎖のサンプル核酸フラグメントを形成する。いくつかの例では、末端修復は、適切な緩衝液中で、T4 DNAポリメラーゼ、クレノウ酵素、およびT4ポリヌクレオチドキナーゼ等の1つ以上の酵素を用いた処理によって行われる。アダプターへの連結反応を促進するヌクレオチドオーバーハングが、いくつかの例では3'から5'のExo-MinusクレノウフラグメントとdATPを用いて、付加される。

40

【0078】

アダプターは、アダプター標識ポリヌクレオチド鎖のライブラリを生成するために、T4リガーゼ等のリガーゼでサンプルポリヌクレオチドフラグメントの両端に連結されても

50

よく、アダプター標識ポリヌクレオチドライブラリは、ユニバーサルプライマー等のプライマーで増幅される。いくつかの例では、アダプターは、1つ以上のプライマー結合部位、1つ以上の接ぎ木領域、および1つ以上のインデックス領域を含むY形のアダプターである。いくつかの例では、1つ以上のインデックス領域はアダプターの各々の鎖に存在する。いくつかの例では、接ぎ木領域は、フローセル面に相補的であり、およびサンプルライブラリの次世代シーケンシングを促進する。いくつかの例では、Y形のアダプターは部分的に相補的な配列を含む。いくつかの例では、Y形のアダプターは、二本鎖アダプター標識ポリヌクレオチド鎖のオーバーハングしたアデニンにハイブリダイズする単一のチミジンオーバーハングを含む。Y形のアダプターは、分裂に抵抗力のある修飾核酸を含んでもよい。例えば、ホスホロチオエート骨格は、アダプターの3'末端にオーバーハングしたチミジンを付けるために使用される。次に、二本鎖サンプル核酸フラグメントのライブラリは、アダプター遮断薬の存在下で変性される。アダプター遮断薬は、アダプター標識ポリヌクレオチド鎖上に存在するアダプター配列（標的配列の代わり）へのプローブのオフターゲットハイブリダイゼーションを最小化する。変性は、いくつかの例では96℃で、または約85、87、90、92、95、97、98、または約99℃で実行される。ポリヌクレオチド標的ライブラリ（プローブライブラリ）は、いくつかの例では96℃、85、87、90、92、95、97、98、または99℃で、ハイブリダイゼーション溶液で変性される。プローブが相補的な標的配列とハイブリダイズすることを可能にするために、変性アダプター標識ポリヌクレオチドライブラリおよびハイブリダイゼーション溶液は、適切な期間と適切な温度でインキュベートされる。いくつかの例では、適切なハイブリダイゼーション温度は、約45～80℃、または少なくとも45、50、55、60、65、70、75、80、85、または90℃である。いくつかの例では、ハイブリダイゼーション温度は70℃である。いくつかの例では、適切なハイブリダイゼーション時間は16時間、または少なくとも4、6、8、10、12、14、16、18、20、22時間、またはそれ以上、または約12～20時間である。次に結合緩衝液がハイブリダイズされたアダプター標識ポリヌクレオチドプローブに添加され、捕捉部分を含む個体担体が、ハイブリダイズされたアダプター標識ポリヌクレオチドプローブを選択的に結合するために使用される。溶出緩衝液が、個体担体からの濃縮された標識ポリヌクレオチドフラグメントに添加される前に、未結合のポリヌクレオチドを除去するために、個体担体を緩衝液で洗浄する。いくつかの例では、個体担体は、2回、または1、2、3、4、5、または6回、洗浄される。アダプター標識ポリヌクレオチドフラグメントの濃縮されたライブラリは増幅され、および濃縮されたライブラリが配列決定される。

【0079】

複数の核酸（すなわちゲノム配列）は、サンプルから得られてもよく、およびフラグメント化され、随意に末端修復され、アデニル化されてもよい。アダプターは、アダプター標識ポリヌクレオチド鎖のライブラリを生成するために、ポリヌクレオチドフラグメントの両端に連結され、およびアダプター標識ポリヌクレオチドライブラリが増幅される。次に、アダプター標識ポリヌクレオチドライブラリは、アダプター遮断薬の存在下において、高温、好ましくは96℃で変性される。ポリヌクレオチド標的ライブラリ（プローブライブラリ）は、高温で、好ましくは約90～99℃で、ハイブリダイゼーション溶液中で変性され、約10～24時間、約45～80℃で、ハイブリダイゼーション溶液中で、変性した標識ポリヌクレオチドライブラリと結合される。次に結合緩衝液がハイブリダイズされたアダプター標識ポリヌクレオチドプローブに添加され、捕捉部分を含む個体担体が、ハイブリダイズされたアダプター標識ポリヌクレオチドプローブを選択的に結合するために使用される。個体担体から濃縮されたアダプター標識ポリヌクレオチドフラグメントを放出するために溶出緩衝液を添加する前に、個体担体は、未結合のポリヌクレオチドを除去するために、緩衝液で1回以上、好ましくは約2～5回、洗浄される。アダプター標識ポリヌクレオチドフラグメントの濃縮されたライブラリが増幅され、その後ライブラリが配列決定される。インキュベーション時間、温度、反応容積/濃度、洗浄回数、または規格と一致する他の変数等の代替的な実験変数もまた、該方法で使用される。

10

20

30

40

50

【0080】

ポリヌクレオチドの集団は、アダプター連結反応に先立って濃縮されてもよい。一例では、複数のポリヌクレオチドはサンプルから得られ、フラグメント化され、随意に末端修復され、高温で、好ましくは90～99 で変性される。ポリヌクレオチド標的ライブラリ（プローブライブラリ）は、高温で、好ましくは約90～99 で、ハイブリダイゼーション溶液中で変性され、約10～24時間、約45～80 で、ハイブリダイゼーション溶液中で、変性した標識ポリヌクレオチドライブラリと結合される。次に結合緩衝液がハイブリダイズされたアダプター標識ポリヌクレオチドプローブに添加され、捕捉部分を含む個体担体が、ハイブリダイズされたアダプター標識ポリヌクレオチドプローブを選択的に結合するために使用される。個体担体から濃縮されたアダプター標識ポリヌクレオチドフラグメントを放出するために溶出緩衝液を添加する前に、個体担体は、未結合のポリヌクレオチドを除去するために、緩衝液で1回以上、好ましくは約2～5回、洗浄される。濃縮されたポリヌクレオチドフラグメントがその後、ポリアデニル化され、アダプターは、アダプター標識ポリヌクレオチド鎖のライブラリを生成するために、ポリヌクレオチドフラグメントの両端に連結され、およびアダプター標識ポリヌクレオチドライブラリが増幅される。次に、アダプター標識ポリヌクレオチドライブラリが配列決定される。

10

【0081】

ポリヌクレオチド標的ライブラリはまた、望ましくないフラグメントにハイブリダイズすることにより、望ましくない配列を複数のポリヌクレオチドからフィルタリングするために使用されてもよい。例えば、複数のポリヌクレオチドはサンプルから得られ、フラグメント化され、随意に末端修復され、アデニル化される。アダプターは、アダプター標識ポリヌクレオチド鎖のライブラリを生成するために、ポリヌクレオチドフラグメントの両端に連結され、およびアダプター標識ポリヌクレオチドライブラリが増幅される。代替的に、アデニル化とアダプター連結反応の工程は代わりに、サンプルポリヌクレオチドの濃縮の後に行なわれる。次に、アダプター標識ポリヌクレオチドライブラリは、アダプター遮断薬の存在下において、高温、好ましくは90～99 で変性される。望ましくない非標的配列を除去するために設計されたポリヌクレオチドフィルタリングライブラリ（プローブライブラリ）は、高温で、好ましくは約90～99 で、ハイブリダイゼーション溶液中で変性され、約10～24時間、約45～80 で、ハイブリダイゼーション溶液中で、変性した標識ポリヌクレオチドライブラリと結合される。次に結合緩衝液がハイブリダイズされたアダプター標識ポリヌクレオチドプローブに添加され、捕捉部分を含む個体担体が、ハイブリダイズされたアダプター標識ポリヌクレオチドプローブを選択的に結合するために使用される。個体担体は、未結合のアダプター標識ポリヌクレオチドフラグメントを溶出するために、緩衝液で1回以上、好ましくは約1～5時間、洗浄される。未結合のアダプター標識ポリヌクレオチドフラグメントの濃縮されたライブラリが増幅され、その後に増幅されたライブラリが配列決定される。

20

30

【0082】

高度に平行なデノボ核酸合成

【0083】

本明細書には、革新的な合成プラットフォームを作るために、シリコン上のナノウェル内のポリヌクレオチド合成から遺伝子組立てへの末端間プロセスの小型化、並列化、および垂直的統合を利用するプラットフォームアプローチが記載される。本明細書に記載されるデバイスは、96ウェルプレートと同じフットプリントを提供し、シリコン合成プラットフォームは、従来の合成方法と比較して、100～1,000倍スループットを増加させることができ、単一の高度に並列された実行（run）で、最大およそ1,000,000のポリヌクレオチドを産生することができる。いくつかの例では、本明細書に記載される単一のシリコンプレートは、約6,100の非同一のポリヌクレオチドの合成を提供する。いくつかの例では、非同一のポリヌクレオチドの各々はクラスタ内にある。クラスタは50～500の非同一のポリヌクレオチドを含み得る。

40

【0084】

50

20

基質

30

50

00, 000; 2, 000, 000; 2, 500, 000; 3, 000, 000; 3, 500, 000; 4, 000, 000; 4, 500, 000; 5, 000, 000; 10, 000, 000、またはより多くの、別個の配列をコードするポリヌクレオチドの合成のための支持を提供する。いくつかの例では、ポリヌクレオチドの少なくとも一部は、同一の配列を有し、または同一の配列で合成されるように構成される。

【0088】

本明細書には、約5、10、20、30、40、50、60、70、80、90、100、125、150、175、200、225、250、275、300、325、350、375、400、425、450、475、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、または2000の塩基の長さのポリヌクレオチドの製造および成長のための方法とデバイスが提供される。いくつかの例では、形成されるポリヌクレオチドの長さは、約5、10、20、30、40、50、60、70、80、90、100、125、150、175、200、または225の塩基の長さである。ポリヌクレオチドは、少なくとも5、10、20、30、40、50、60、70、80、90、または100の塩基の長さであり得る。ポリヌクレオチドは、10~225の塩基の長さ、12~100の塩基の長さ、20~150の塩基の長さ、20~130の塩基の長さ、または30~100の塩基の長さであり得る。

10

【0089】

いくつかの例では、ポリヌクレオチドは、基質の別個の遺伝子座において合成され、各遺伝子座はポリヌクレオチドの集団の合成を支持する。いくつかの例では、各遺伝子座は、別の遺伝子座で成長させポリヌクレオチドの集団とは異なる配列を有するポリヌクレオチドの集団の合成を支持する。いくつかの例では、デバイスの遺伝子座は複数のクラスタ内に位置する。いくつかの例では、デバイスは、少なくとも10、500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、11000、12000、13000、14000、15000、20000、30000、40000、50000、またはより多くのクラスタを含む。いくつかの例では、デバイスは、2, 000; 5, 000; 10, 000; 100, 000; 200, 000; 300, 000; 400, 000; 500, 000; 600, 000; 700, 000; 800, 000; 900, 000; 1, 000, 000; 1, 100, 000; 1, 200, 000; 1, 300, 000; 1, 400, 000; 1, 500, 000; 1, 600, 000; 1, 700, 000; 1, 800, 000; 1, 900, 000; 2, 000, 000; 300, 000; 400, 000; 500, 000; 600, 000; 700, 000; 800, 000; 900, 000; 1, 000, 000; 1, 200, 000; 1, 400, 000; 1, 600, 000; 1, 800, 000; 2, 000, 000; 2, 500, 000; 3, 000, 000; 3, 500, 000; 4, 000, 000; 4, 500, 000; 5, 000, 000; または10, 000, 000、またはより多くの別個の遺伝子座を含むいくつかの例では、デバイスは、約10, 000の別個の遺伝子座を含む。単一のクラスタ内の遺伝子座の量は、異なる事例では変更される。いくつかの例では、各クラスタは、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、120、130、150、200、300、400、500、1000、またはより多くの遺伝子座を含む。いくつかの例では、各クラスタは、約50~500の遺伝子座を含む。いくつかの例では、各クラスタは、約100~200の遺伝子座を含む。いくつかの例では、各クラスタは、約109、121、130、または137の遺伝子座を含む。いくつかの例では、各クラスタは、約19、20、61、64、またはそれ以上の遺伝子座を含む。

20

30

40

【0090】

デバイス上で合成された別々のポリヌクレオチドの数は、基質における利用可能な別々の遺伝子座の数に左右され得る。いくつかの事例では、デバイスのクラスタ内の遺伝子座

50

の密度は、 1 mm^2 あたり少なくともまたは約 1 の遺伝子座、 1 mm^2 あたり 10 の遺伝子座、 1 mm^2 あたり 25 の遺伝子座、 1 mm^2 あたり 50 の遺伝子座、 1 mm^2 あたり 65 の遺伝子座、 1 mm^2 あたり 75 の遺伝子座、 1 mm^2 あたり 100 の遺伝子座、 1 mm^2 あたり 130 の遺伝子座、 1 mm^2 あたり 150 の遺伝子座、 1 mm^2 あたり 175 の遺伝子座、 1 mm^2 あたり 200 の遺伝子座、 1 mm^2 あたり 300 の遺伝子座、 1 mm^2 あたり 400 の遺伝子座、 1 mm^2 あたり 500 の遺伝子座、 1 mm^2 あたり 1,000 の遺伝子座、またはそれ以上である。いくつかの例では、デバイスは、 1 mm^2 から約 500 mm^2 あたり約 10 の遺伝子座、 1 mm^2 から約 400 mm^2 あたり約 25 の遺伝子座、 1 mm^2 から約 500 mm^2 あたり約 50 の遺伝子座、 1 mm^2 から約 500 mm^2 あたり約 100 の遺伝子座、 1 mm^2 から約 500 mm^2 あたり約 150 の遺伝子座、 1 mm^2 から約 250 mm^2 あたり約 10 の遺伝子座、 1 mm^2 から約 250 mm^2 あたり約 50 の遺伝子座、 1 mm^2 から約 200 mm^2 あたり約 10 の遺伝子座、 1 mm^2 から約 200 mm^2 あたり約 50 の遺伝子座を含む。いくつかの例では、クラスタ内の 2 つの隣接した遺伝子座の中心からの距離は、約 $10\text{ }\mu\text{m}$ から約 $500\text{ }\mu\text{m}$ 、約 $10\text{ }\mu\text{m}$ から約 $200\text{ }\mu\text{m}$ 、または約 $10\text{ }\mu\text{m}$ から約 $100\text{ }\mu\text{m}$ である。いくつかの例では、隣接した遺伝子座の 2 つの中心からの距離は、約 $10\text{ }\mu\text{m}$ 、 $20\text{ }\mu\text{m}$ 、 $30\text{ }\mu\text{m}$ 、 $40\text{ }\mu\text{m}$ 、 $50\text{ }\mu\text{m}$ 、 $60\text{ }\mu\text{m}$ 、 $70\text{ }\mu\text{m}$ 、 $80\text{ }\mu\text{m}$ 、 $90\text{ }\mu\text{m}$ 、または $100\text{ }\mu\text{m}$ より長い。いくつかの例では、2 つの隣接した遺伝子座の中心からの距離は、約 $200\text{ }\mu\text{m}$ 、 $150\text{ }\mu\text{m}$ 、 $100\text{ }\mu\text{m}$ 、 $80\text{ }\mu\text{m}$ 、 $70\text{ }\mu\text{m}$ 、 $60\text{ }\mu\text{m}$ 、 $50\text{ }\mu\text{m}$ 、 $40\text{ }\mu\text{m}$ 、 $30\text{ }\mu\text{m}$ 、 $20\text{ }\mu\text{m}$ 、または $10\text{ }\mu\text{m}$ 未満である。いくつかの例では、遺伝子座それぞれの幅は、約 $0.5\text{ }\mu\text{m}$ 、 $1\text{ }\mu\text{m}$ 、 $2\text{ }\mu\text{m}$ 、 $3\text{ }\mu\text{m}$ 、 $4\text{ }\mu\text{m}$ 、 $5\text{ }\mu\text{m}$ 、 $6\text{ }\mu\text{m}$ 、 $7\text{ }\mu\text{m}$ 、 $8\text{ }\mu\text{m}$ 、 $9\text{ }\mu\text{m}$ 、 $10\text{ }\mu\text{m}$ 、 $20\text{ }\mu\text{m}$ 、 $30\text{ }\mu\text{m}$ 、 $40\text{ }\mu\text{m}$ 、 $50\text{ }\mu\text{m}$ 、 $60\text{ }\mu\text{m}$ 、 $70\text{ }\mu\text{m}$ 、 $80\text{ }\mu\text{m}$ 、 $90\text{ }\mu\text{m}$ 、または $100\text{ }\mu\text{m}$ である。いくつかの例では、遺伝子座それぞれの幅は、約 $0.5\text{ }\mu\text{m}$ から $100\text{ }\mu\text{m}$ 、約 $0.5\text{ }\mu\text{m}$ から $50\text{ }\mu\text{m}$ 、約 $10\text{ }\mu\text{m}$ から $75\text{ }\mu\text{m}$ 、または約 $0.5\text{ }\mu\text{m}$ から $50\text{ }\mu\text{m}$ である。

【0091】

いくつかの例では、デバイス内のクラスタの密度は、 100 mm^2 あたり少なくともまたは約 1 のクラスタ、 10 mm^2 あたり 1 のクラスタ、 5 mm^2 あたり 1 のクラスタ、 4 mm^2 あたり 1 のクラスタ、 3 mm^2 あたり 1 のクラスタ、 2 mm^2 あたり 1 のクラスタ、 1 mm^2 あたり 1 のクラスタ、 1 mm^2 あたり 2 のクラスタ、 1 mm^2 あたり 3 のクラスタ、 1 mm^2 あたり 4 のクラスタ、 1 mm^2 あたり 5 のクラスタ、 1 mm^2 あたり 10 のクラスタ、 1 mm^2 あたり 50 のクラスタ、またはそれ以上である。いくつかの例では、デバイスは、 10 mm^2 あたり約 1 のクラスタから 1 mm^2 あたり約 10 のクラスタを含む。いくつかの例では、2 つの隣接したクラスタの中心からの距離は、約 $50\text{ }\mu\text{m}$ 、 $100\text{ }\mu\text{m}$ 、 $200\text{ }\mu\text{m}$ 、 $500\text{ }\mu\text{m}$ 、 $1000\text{ }\mu\text{m}$ 、 $2000\text{ }\mu\text{m}$ 、または $5000\text{ }\mu\text{m}$ 未満である。いくつかの例では、2 つの隣接したクラスタの中心からの距離は、約 $50\text{ }\mu\text{m}$ から約 $100\text{ }\mu\text{m}$ 、約 $50\text{ }\mu\text{m}$ から約 $200\text{ }\mu\text{m}$ から、約 $50\text{ }\mu\text{m}$ から約 $300\text{ }\mu\text{m}$ 、約 $50\text{ }\mu\text{m}$ から約 $500\text{ }\mu\text{m}$ 、および約 $100\text{ }\mu\text{m}$ から約 $2000\text{ }\mu\text{m}$ である。いくつかの例では、2 つの隣接したクラスタの中心からの距離は、約 0.05 mm から約 50 mm 、約 0.05 mm から約 10 mm 、約 0.05 mm から約 5 mm 、約 0.05 mm から約 4 mm 、約 0.05 mm から約 3 mm 、約 0.05 mm から約 2 mm 、約 0.1 mm から 10 mm 、約 0.2 mm から 10 mm 、約 0.3 mm から約 10 mm 、約 0.4 mm から約 10 mm 、約 0.5 mm から 10 mm 、約 0.5 mm から約 5 mm 、または約 0.5 mm から約 2 mm である。いくつかの例では、各クラスタは、約 0.5 から 2 mm 、約 0.5 から 1 mm 、または約 1 から 2 mm の 1 寸法に沿った直径または幅を有している。いくつかの例では、各クラスタは、約 0.5 、 0.6 、 0.7 、 0.8 、 0.9 、 1 、 1.1 、 1.2 、 1.3 、 1.4 、 1.5 、 1.6 、 1.7 、 1.8 、 1.9 、または 2 mm の 1 寸法に沿った直径または幅を有している。いくつかの例では、各クラスタは、約 0.5 、 0.6 、 0.7 、 0.8 、 0.9 、 1 、 1.1 、 1.15 、 1.2 、 1.3 、 1.4 、 1.5 、 1.6 、 1.7 、 1.8 、 1.9 、または 2 mm の 1 寸法に沿った内径または幅を

有している。

【0092】

デバイスは、標準的な96ウェルプレートのおよそのサイズであってもよく、例えば、約100~200mm×約50~150mmであってもよい。いくつかの例では、デバイスは、約1000mm、500mm、450mm、400mm、300mm、250mm、200mm、150mm、100mm、または50mm未満の直径を有する。いくつかの例では、デバイスの直径は、約25mmから1000mm、約25mmから約800mm、約25mmから約600mm、約25mmから約500mm、約25mmから約400mm、約25mmから約300mm、または約25mmから約200である。デバイスサイズの非限定的な例は、約300mm、200mm、150mm、130mm、100mm、76mm、51mm、および25mmを含む。いくつかの例では、デバイスは、少なくとも約100mm²；200mm²；500mm²；1,000mm²；2,000mm²；5,000mm²；10,000mm²；12,000mm²；15,000mm²；20,000mm²；30,000mm²；40,000mm²；50,000mm²、またはより大きな平面面積を有している。いくつかの例では、デバイスの厚さは、約50mm~約2000mm、約50mm~約1000mm、約100mm~約1000mm、約200mm~約1000mm、または約250mm~約1000mmである。デバイスの厚さの非限定的な例として、約275mm、375mm、525mm、625mm、675mm、725mm、775mm、および925mmがあげられる。いくつかの例では、デバイスの厚さは、直径によって変わり、および基質の組成物に左右される。例えば、シリコン以外の物質を含むデバイスは、同じ直径のシリコンデバイスとは異なる厚さを有する。デバイスの厚さは、使用される物質の機械強度によって判定され、取り扱い中に割れることなく、それ自体の重量を支えるのに十分でなければならない。いくつかの例では、構造は、本明細書に記載される複数のデバイスを含む。

10

20

【0093】

面材料

【0094】

本明細書において、面を含むデバイスが提供され、面は、あらかじめ決定された位置でのポリヌクレオチド合成を支持するように修飾され、および結果として生じる低エラー率、低いドロップアウト率、高い収率、および高いオリゴ表示を伴う。いくつかの実施形態では、本明細書で提供されるポリヌクレオチド合成用のデバイスの面は、新たなポリヌクレオチド合成反応を支持するために修飾可能な様々な材料から製造される。いくつかの場合には、デバイスは十分に導電性であり、例えば、デバイスのすべてまたは一部にわたって均一な電場を形成することができる。本明細書に記載されるデバイスは、可撓性材料を含み得る。典型的な可撓性材料として、限定されないが、改質ナイロン、非改質ナイロン、ニトロセルロース、およびポリプロピレンがあげられる。本明細書に記載されるデバイスは、剛性材料を含み得る。典型的な剛性材料として、限定されないが、ガラス、石英ガラス(fuse silica)、シリコン、二酸化ケイ素、窒化ケイ素、プラスチック(例えば、ポリテトラフルオロエチレン、ポリプロピレン、ポリスチレン、ポリカーボネート、およびそれらの混合物など)、および金属(例えば、金、白金など)があげられる。本明細書に開示されるデバイスは、シリコン、ポリスチレン、アガロース、デキストラン、セルロース系ポリマー、ポリアクリルアミド、ポリジメチルシロキサン(PDMS)、ガラス、またはそれらの任意の組み合わせを含む材料から作られてもよい。いくつかの場合には、本明細書に開示されるデバイスは、本明細書にリストされる材料または当技術分野において既知の他の適切な材料を組み合わせで製造される。

30

40

【0095】

本明細書に記載される典型的な物質に関する引張強度のリストは、以下に提供される：ナイロン(70MPa)、ニトロセルロース(1.5MPa)、ポリプロピレン(40MPa)、シリコン(268MPa)、ポリスチレン(40MPa)、アガロース(1-10MPa)、ポリアクリルアミド(1-10MPa)、ポリジメチルシロキサン(PDM

50

S) (3.9 - 10.8 MPa)。本明細書に記載される個体担体は、1 ~ 300、1 ~ 40、1 ~ 10、1 ~ 5、または3 ~ 11 MPaの引張強度を有し得る。本明細書に記載される個体担体は、約1、1.5、2、3、4、5、6、7、8、9、10、11、20、25、40、50、60、70、80、90、100、150、200、250、270、またはより多くのMPaの引張強度を有し得る。いくつかの例では、本明細書に記載されるデバイスは、テープまたは軟質シート等の、連続したループまたはリールに保存することができる可撓性材料の形態の、ポリヌクレオチド合成用の個体担体を含む。

【0096】

ヤング率は、弾性の（復元可能な）荷重変形に対する物質の抵抗を測定する。本明細書に記載される典型的な物質の剛性に関するヤング率のリストは、以下に提供される：ナイロン（3 GPa）、ニトロセルロース（1.5 GPa）、ポリプロピレン（2 GPa）、シリコン（150 GPa）、ポリスチレン（3 GPa）、アガロース（1 - 10 GPa）、ポリアクリルアミド（1 - 10 GPa）、ポリジメチルシロキサン（PDMS）（1 - 10 GPa）。本明細書に記載される個体担体は、1 ~ 500、1 ~ 40、1 ~ 10、1 ~ 5、または3 ~ 11 GPaのヤング率を有し得る。本明細書に記載される個体担体は、約1、1.5、2、3、4、5、6、7、8、9、10、11、20、25、40、50、60、70、80、90、100、150、200、250、400、500 GPa、またはより多くのGPaのヤング率を有し得る。可撓性と剛性は真逆の関係にあるため、可撓性材料は低いヤング率を有し、その形状は負荷を受けてかなり変化する。

10

【0097】

いくつかの場合には、本明細書に開示されるデバイスは、二酸化ケイ素の基部と酸化シリコンの表層面を含む。代替的に、デバイスは、シリコン酸化物の基部を含んでもよい。本明細書に提供されるデバイスの面は織り込まれてもよく、結果としてポリヌクレオチド合成のための面の総面積を増加させる。本明細書に開示されるデバイスは、少なくとも5%、10%、25%、50%、80%、90%、95%、または99%のシリコンを含み得る。本明細書に開示されるデバイスは、SOI (silicon on insulator) ウェーハから製造されてもよい。

20

【0098】

面アーキテクチャ

【0099】

本明細書には、隆起した、および/または沈降した機構を含むデバイスが提供される。そのような機構を有する1つの利点は、ポリヌクレオチド合成を支持する表面積の増大である。いくつかの例では、隆起した、および/または沈降した機構を有するデバイスは、三次元基質と呼ばれる。いくつかの例では、三次元デバイスは1つ以上のチャネルを含む。いくつかの例では、1つ以上の遺伝子座はチャネルを含む。いくつかの例では、チャネルは、ポリヌクレオチドシンセサイザーなどの沈着デバイスによる試薬の沈着に利用可能である。いくつかの例では、試薬および/または流体は、1つ以上のチャネルと流体連通してより大きなウェルに集まる。例えば、デバイスは、クラスタを有する複数の遺伝子座に対応する複数のチャネルを含み、複数のチャネルは、クラスタの1つのウェルと流体連通している。いくつかの方法では、ポリヌクレオチドのライブラリは、クラスタの複数の遺伝子座において合成される。

30

40

【0100】

いくつかの例では、構造は、面上のポリヌクレオチド合成に関する流れの制御および物質移動経路の制御を可能にするように構成されている。いくつかの例では、デバイスの構成は、ポリヌクレオチド合成中の物質移動経路、化学暴露時間、および/または洗浄効果の制御およびその分布さえも可能にする。いくつかの例では、デバイスの構成は、例えば、成長しているポリヌクレオチドによる排除体積が、ポリヌクレオチドの成長に利用可能または適した最初の利用可能な体積の50、45、40、35、30、25、20、15、14、13、12、11、10、9、8、7、6、5、4、3、2、1%を超えないように、ポリヌクレオチドの成長に対する十分な体積を提供することによって、奏効率の増

50

大を可能にする。いくつかの例では、三次元構造は、化学暴露の急速な交換を可能にするために流体の流れの管理を可能にする。

【0101】

本明細書には、1 f M、5 f M、10 f M、25 f M、50 f M、75 f M、100 f M、200 f M、300 f M、400 f M、500 f M、600 f M、700 f M、800 f M、900 f M、1 p M、5 p M、10 p M、25 p M、50 p M、75 p M、100 p M、200 p M、300 p M、400 p M、500 p M、600 p M、700 p M、800 p M、900 p M、またはそれ以上の量のDNAを合成する方法が提供される。いくつかの例では、ポリヌクレオチドライブラリは、遺伝子の約1%、2%、3%、4%、5%、10%、15%、20%、30%、40%、50%、60%、70%、80%、90%、95%、または100%の長さに及び得る。遺伝子は、約1%、2%、3%、4%、5%、10%、15%、20%、30%、40%、50%、60%、70%、80%、85%、90%、95%、または100%まで変えられ得る。

10

【0102】

非同一のポリヌクレオチドは総体として、遺伝子の少なくとも1%、2%、3%、4%、5%、10%、15%、20%、30%、40%、50%、60%、70%、80%、85%、90%、95%、または100%に対する配列をコードし得る。いくつかの例では、ポリヌクレオチドは、遺伝子の50%、60%、70%、80%、85%、90%、95%、またはそれ以上の配列をコードし得る。いくつかの例では、ポリヌクレオチドは、遺伝子の80%、85%、90%、95%、またはそれ以上の配列をコードし得る。

20

【0103】

いくつかの例では、物理構造によって隔離が達成される。いくつかの例では、ポリヌクレオチド合成に対する能動領域および受動領域を生成する面の差次的な官能化によって、隔離が達成される。差次的な官能化はまた、デバイス面にわたる疎水性を交互にし、それによって沈着された試薬の水滴 (beading) または湿りを引き起こす水接触角の効果を作り出すことにより達成される。より大きな構造を利用することで、飛び散り (splashing) および隣接するスポットの試薬による別々のポリヌクレオチド合成位置の相互汚染を減らすことができる。いくつかの例では、ポリヌクレオチドシンセサイザーなどの装置は、別々のポリヌクレオチド合成位置に試薬を沈着するために使用される。三次元の機構を有する基質は、低いエラー率 (例えば、約1:500、1:1000、1:1500、1:2,000、1:3,000、1:5,000、または1:10,000未満) でポリヌクレオチド (例えば約10,000を超える) の合成を可能にする方法で構成される。いくつかの例では、デバイスは、1 mm² あたり約1、5、10、20、30、40、50、60、70、80、100、110、120、130、140、150、160、170、180、190、200、300、400、または500、またはより多くの特徴の密度で機構を含む。

30

【0104】

デバイスのチャンネルは、基質の別のウェルと同じ、または異なる幅、高さ、および/または体積を有し得る。デバイスのチャンネルは、基質の別のチャンネルと同じまたは異なる幅、高さ、および/または体積を有し得る。いくつかの例では、クラスタの幅は、約0.05 mmから約50 mm、約0.05 mmから約10 mm、約0.05 mmから約5 mm、約0.05 mmから約4 mm、約0.05 mmから約3 mm、約0.05 mmから約2 mm、約0.05 mmから約1 mm、約0.05 mmから約0.5 mm、約0.05 mmから約0.1 mm、約0.1 mmから10 mm、約0.2 mmから10 mm、約0.3 mmから約10 mm、約0.4 mmから約10 mm、約0.5 mmから10 mm、約0.5 mmから約5 mm、または約0.5 mmから約2 mmである。いくつかの例では、クラスタを含むウェルの幅は、約0.05 mmから約50 mm、約0.05 mmから約10 mm、約0.05 mmから約5 mm、約0.05 mmから約4 mm、約0.05 mmから約3 mm、約0.05 mmから約2 mm、約0.05 mmから約1 mm、約0.05 mmから約0.5 mm、約0.05 mmから約0.1 mm、約0.1 mmから10 mm、約0.2 m

40

50

mから10 mm、約0.3 mmから約10 mm、約0.4 mmから約10 mm、約0.5 mmから10 mm、約0.5 mmから約5 mm、約0.5 mmから約2 mmである。いくつかの例では、クラスタの幅は、約5 mm、4 mm、3 mm、2 mm、1 mm、0.5 mm、0.1 mm、0.09 mm、0.08 mm、0.07 mm、0.06 mm、または0.05 mmであり、またはこれらの数値未満である。いくつかの例では、クラスタの幅は、約1.0から1.3 mmである。いくつかの例では、クラスタの幅は、約1.150 mmである。いくつかの例では、ウェルの幅は、約5 mm、4 mm、3 mm、2 mm、1 mm、0.5 mm、0.1 mm、0.09 mm、0.08 mm、0.07 mm、0.06 mm、または0.05 mmであり、またはこれらの数値未満である。いくつかの例では、ウェルの幅は、約1.0から1.3 mmである。いくつかの例では、ウェルの幅は、約1.150 mmである。いくつかの例では、クラスタの幅は約0.08 mmである。いくつかの例では、ウェルの幅は約0.08 mmである。クラスタの幅は、二次元または三次元の基質内のクラスタを指す場合もある。

10

20

30

40

50

【0105】

いくつかの例では、ウェルの高さは、約20 μm から約1000 μm 、約50 μm から約1000 μm 、約100 μm から約1000 μm 、約200 μm から約1000 μm 、約300 μm から約1000 μm 、約400 μm から約1000 μm 、または約500 μm から約1000 μm である。いくつかの例では、ウェルの高さは、約1000 μm 未満、約900 μm 未満、約800 μm 未満、約700 μm 未満、または約600 μm 未満である。

【0106】

いくつかの例では、デバイスは、クラスタ内の複数の遺伝子座に対応する複数のチャネルを含み、チャネルの高さまたは深さは、約5 μm から約500 μm 、約5 μm から約400 μm 、約5 μm から約300 μm 、約5 μm から約200 μm 、約5 μm から約100 μm 、約5 μm から約50 μm 、または約10 μm から約50 μm である。いくつかの例では、チャネルの高さは、100 μm 未満、80 μm 未満、60 μm 未満、40 μm 未満、または20 μm 未満である。

【0107】

いくつかの例では、チャネル、遺伝子座（例えば、実質的に平面の基質における）またはチャネルと遺伝子座の両方の直径（例えば、遺伝子座がチャネルに対応する三次元構造デバイスにおける）は、約1 μm から約1000 μm 、約1 μm から約500 μm 、約1 μm から約200 μm 、約1 μm から約100 μm 、約5 μm から約100 μm 、または約10 μm から約100 μm 、例えば、約90 μm 、80 μm 、70 μm 、60 μm 、50 μm 、40 μm 、30 μm 、20 μm または10 μm である。いくつかの例では、チャネル、遺伝子座、またはチャネルと遺伝子座の両方の直径は、約100 μm 、90 μm 、80 μm 、70 μm 、60 μm 、50 μm 、40 μm 、30 μm 、20 μm 、または10 μm 未満である。いくつかの例では、2つの隣接チャネル、遺伝子座、またはチャネルと遺伝子座の中心からの距離は、約1 μm から約500 μm 、約1 μm から約200 μm 、約1 μm から約100 μm 、約5 μm から約200 μm 、約5 μm から約100 μm 、約5 μm から約50 μm 、または約5 μm から約30 μm 、例えば、約20 μm である。

【0108】

面修飾

【0109】

様々な例では、デバイス面またはデバイス面の選択された部位あるいは領域の1つ以上の化学的および/または物理的な特性を変更するために付加または減算のプロセスによって、面の化学的および/または物理的な変更のために、面修飾が利用される。例えば、面修飾は、限定されないが、（1）面の湿潤性の変更、（2）面の官能化、つまり面官能基の提供、修飾、または置換、（3）面の脱官能化、つまり面官能基の除去、（4）そうでなければ、例えばエッチングによって、面の化学組成を変更すること、（5）面の粗さを増大または減少させること、（6）面へのコーティング、例えば、面の湿潤性とは異なる湿

潤性を示すコーティングの提供、および／または(7)面への粒子の沈着、を含む。

【0110】

いくつかの例では、面の最上部の化学層(接着促進剤と呼ばれる)の付加は、基質の面上の遺伝子座の構造化されたパターン化を促進する。接着促進の適用のための典型的な面は、限定されないが、ガラス、シリコン、二酸化ケイ素、および窒化ケイ素を含む。いくつかの例では、接着促進剤は、高い表面エネルギーを有する化学物質である。いくつかの例では、基質の面上に第2の化学層が堆積される。いくつかの例では、第2の化学層は、低い表面エネルギーを有している。いくつかの例では、面上にコーティングされた化学層の表面エネルギーは、面への液滴の局在化を支持する。選択されるパターン配置に応じて、遺伝子座の接近および／または遺伝子座での流体接触の領域は変更可能である。

10

【0111】

いくつかの例では、例えばポリヌクレオチド合成のために、核酸または他の部分が堆積されるデバイス面または分解された遺伝子座は、滑らかであり、または実質的に平面であり(例えば、二次元)、あるいは隆起した、または沈降した機構(例えば三次元の機構)などの不規則性を有している。いくつかの例では、デバイス面は、化合物の1つ以上の異なる層で修飾される。対象のそのような修飾層は、限定されないが、金属、金属酸化物、ポリマー、小さな有機分子等の、無機層および有機層を含む。非限定的なポリマー層は、ペプチド、タンパク質、核酸またはそれらの模倣物(例えば、ペプチド核酸等)、多糖類、リン脂質、ポリウレタン、ポリエステル、ポリカーボネート、ポリ尿素、ポリアミド、ポリエチレンアミン、ポリアリーレンスルフィド、ポリシロキサン、ポリイミド、ポリアセテート、および本明細書に記載される、またはそうでなければ当該技術分野で既知の他の適切な化合物を含む。いくつかの例では、ポリマーはヘテロポリマーである。いくつかの例では、ポリマーはホモポリマーである。いくつかの例では、ポリマーは官能性部分を含む、またはそれに結合される。

20

【0112】

いくつかの例では、デバイスの分解された遺伝子座は、表面エネルギーを増大および／または減少させる1つ以上の部分で官能化される。いくつかの例では、部分は化学的に不活性である。いくつかの例では、部分は、ポリヌクレオチド合成反応における望ましい化学反応、例えば、1つ以上のプロセスを支持するように構成されている。面の表面エネルギー、すなわち疎水性は、面上へと付着するヌクレオチドの親和性を測定するための因子である。いくつかの例では、デバイスの官能化のための方法は、(a)二酸化ケイ素を含む面を有するデバイスを提供する工程；および(b)本明細書に記載される、またはそうでなければ当該技術分野で既知の適切なシラン化剤、例えば、有機官能性アルコキシシラン分子を使用して、面をシラン処理する工程を含む。

30

【0113】

いくつかの例では、有機官能性アルコキシシラン分子は、ジメチルクロロ-オクトデシル-シラン、メチルジクロロ-オクトデシル-シラン、トリクロロ-オクトデシル-シラン、トリメチル-オクトデシル-シラン、トリエチル-オクトデシル-シラン、またはそれらの任意の組み合わせを含む。いくつかの例では、デバイスの面は、ポリエチレン/ポリプロピレン(ガンマ線照射またはクロム酸酸化、およびヒドロキシルアルキル表面への還元によって官能化される)、高度に架橋されたポリスチレン-ジビニルベンゼン(クロロメチル化によって誘導体化され、ベンジルアミン官能面にアミノ化される)、ナイロン(末端のアミノヘキシル基は直接反応性である)で官能化され、または還元ポリテトラフルオロエチレンでエッチングされる。他の方法および官能化剤は、米国特許第5,474,796号に記載され、これは引用によってその全体が本明細書に組み込まれる。

40

【0114】

いくつかの例では、デバイスの面は、典型的にデバイス面上に存在する反応性の親水性部分を介して、デバイス面にシランをカップリングするのに有効な反応条件下で、シランの混合物を含む誘導体化組成物との接触によって官能化される。シラン処理は一般に、自己組織化により有機官能性アルコキシシラン分子で表面を覆う。

50

【0115】

当該技術分野において現在知られているように、例えば、表面エネルギーを低下または増大させるための、様々なシロキサン官能化試薬がさらに使用され得る。有機官能性アルコキシシランは、それらの有機官能性に従って分類され得る。

【0116】

本明細書には、ヌクレオシドに結合することができる薬剤のパターン化を含み得るデバイスが提供される。いくつかの例では、デバイスは、活性薬剤でコーティングされてもよい。いくつかの例では、デバイスは、受動剤 (passive agent) でコーティングされてもよい。本明細書に記載されるコーティング材に包含するための典型的な活性因子として、限定されないが、N - (3 - トリエトキシシリルプロピル) - 4 - ヒドロキシブチルアミド (HAPS)、11 - アセトキシウンデシルトリエトキシシラン、n - デシルトリエトキシシラン、(3 - アミノプロピル)トリメトキシシラン、(3 - アミノプロピル)トリエトキシシラン、3 - グリシドキシプロピルトリメトキシシラン (GOPS)、3 - ヨード - プロピルトリメトキシシラン、ブチル - アルデヒド - トリメトキシシラン、二量体二次アミノアルキルシロキサン、(3 - アミノプロピル) - ジエトキシ - メチルシラン、(3 - アミノプロピル) - ジメチル - エトキシシラン、および (3 - アミノプロピル) - トリメトキシシラン、(3 - グリシドキシプロピル) - ジメチル - エトキシシラン、グリシドキシ - トリメトキシシラン、(3 - メルカプトプロピル) - トリメトキシシラン、3 - 4 エポキシシクロヘキシル - エチルトリメトキシシラン、および (3 - メルカプトプロピル) - メチル - ジメトキシシラン、アリルトリクロクロクロシラン、7 - オクタ - 1 - エニルトリクロクロクロシラン、またはビス (3 - トリメトキシシリルプロピル) アミンがあげられる。

10

20

【0117】

本明細書に記載されるコーティング材に含有するための典型的な受動剤は、限定されないが、ペルフロオロオクチルトリクロクロシラン；トリデカフルオロ - 1, 1, 2, 2 - テトラヒドロオクチルトリクロクロシラン；1H, 1H, 2H, 2H - フルオロオクチルトリエトキシシラン (FOS)；トリクロロ (1H, 1H, 2H, 2H - ペルフロオロオクチル)シラン；tert - ブチル - [5 - フルオロ - 4 - (4, 4, 5, 5 - テトラメチル - 1, 3, 2 - ジオキサボロラン - 2 - イル)インドール - 1 - イル] - ジメチル - シラン；CYTOP (商標)；フロリナート (商標)；ペルフロオロオクチルトリクロクロシラン (PFOTCS)；ペルフロオロオクチルジメチルクロクロシラン (PFODCS)；ペルフロオロデシルトリエトキシシラン (PFDETS)；ペンタフルオロフェニル - ジメチルプロピルクロクロシラン (PFPTES)；ペルフロオロオクチルトリエトキシシラン；ペルフロオロオクチルトリメトキシシラン；オクチルクロクロシラン；ジメチルクロクロ - オクトデシル - シラン；メチルジクロクロ - オクトデシル - シラン；トリクロクロ - オクトデシル - シラン；トリメチル - オクトデシル - シラン；トリエチル - オクトデシル - シラン；またはオクタデシルトリクロクロシランを含む。

30

【0118】

いくつかの例では、官能化剤は、オクタデシルトリクロクロシランなどの炭化水素シランを含む。いくつかの例では、官能化剤は、11 - アセトキシウンデシルトリエトキシシラン、n - デシルトリエトキシシラン、(3 - アミノプロピル)トリメトキシシラン、(3 - アミノプロピル)トリエトキシシラン、グリシジルオキシプロピル / トリメトキシシランおよび N - (3 - トリエトキシシリルプロピル) - 4 - ヒドロキシブチルアミドを含む。

40

【0119】

ポリヌクレオチド合成

【0120】

ポリヌクレオチド合成のための本開示の方法は、ホスホラミダイトの化学作用を含むプロセスを含み得る。いくつかの例では、ポリヌクレオチド合成は、塩基をホスホラミダイトとカップリングする工程を含む。ポリヌクレオチドは、カップリング条件下でホスホラ

50

ミダイトの堆積によって塩基をカップリングする工程を含んでもよく、ここで同じ塩基が随意に、1回より多く、つまり二重のカップリングでホスホラミダイトと共に堆積される。ポリヌクレオチド合成は、未反応部位のキャッピングを含んでもよい。いくつかの例では、キャッピングは随意である。ポリヌクレオチド合成はまた、酸化または酸化工程を含んでもよい。ポリヌクレオチド合成は、分解、脱トリチル化、および硫化を含んでもよい。いくつかの例では、ポリヌクレオチド合成は、酸化または硫化のいずれかを含む。いくつかの例では、ポリヌクレオチド合成反応中の1工程または各工程間で、デバイスは、例えばテトラゾールまたはアセトニトリルを使用して洗浄される。ホスホラミダイト合成法における任意の1工程にかかる時間枠は、約2分、1分、50秒、40秒、30秒、20秒、および10秒未満であり得る。

10

【0121】

ホスホラミダイト法を使用するポリヌクレオチド合成は、亜リン酸塩トリエステル結合の形成のために成長しているポリヌクレオチド鎖へのホスホラミダイト構築ブロック（例えば、ヌクレオシドホスホラミダイト）の続く付加を含んでもよい。ホスホラミダイトポリヌクレオチド合成は、3'から5'の方向に進む。ホスホラミダイトポリヌクレオチド合成は、1合成サイクルにつき成長している核酸鎖への1つのヌクレオチドの制御された付加を可能にする。いくつかの例では、各合成サイクルはカップリング工程を含む。ホスホラミダイト結合は、活性化されたヌクレオシドホスホラミダイトと、例えばリンカーによって基質に結合されたヌクレオシドとの間の亜リン酸塩トリエステル結合の形成を伴う。いくつかの例では、ヌクレオシドホスホラミダイトは、起動されたデバイスに提供される。いくつかの例では、ヌクレオシドホスホラミダイトは、アクチベーター（*activator*）でデバイスに提供される。いくつかの例では、ヌクレオシドホスホラミダイトは、基質に結合されたヌクレオシドよりも1、5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、50、60、70、80、90、100倍、またはそれ以上の過剰量で、デバイスに提供される。いくつかの例では、ヌクレオシドホスホラミダイトの付加は、無水環境において、例えば無水アセトニトリルにおいて行われる。ヌクレオシドホスホラミダイトの付加に続いて、デバイスは随意に洗浄される。いくつかの例では、カップリング工程は、随意に、基質へのヌクレオシドホスホラミダイトの付加の間の洗浄工程と共に、追加で1回以上繰り返される。いくつかの例では、本明細書で使用されるポリヌクレオチド合成法は、1、2、3、またはより多くの連続したカップリング工程を含む。カップリング前に、多くの場合において、デバイスに結合されたヌクレオシドは、保護基の除去によって脱保護され、保護基は重合を防ぐように機能する。一般的な保護基は、4,4'-ジメトキシトリチル（DMT）である。

20

30

【0122】

カップリングに続いて、ホスホラミダイトポリヌクレオチド合成法は、随意にキャッピング工程を含む。キャッピング工程では、成長しているポリヌクレオチドがキャッピング剤で処理される。キャッピング工程は、カップリング後にさらなる鎖伸長から未反応の基質に結合した5'-OH基を遮断するのに有用であり、これによって、内部塩基欠失（*internal base deletions*）を伴うポリヌクレオチドの形成を防ぐ。さらに、1H-テトラゾールで活性化されたホスホラミダイトは、少しの程度、グアノシンのO6位置と反応し得る。理論に縛られることなく、I₂/水での酸化に際して、この副産物は、恐らくO6-N7遊走を介して脱プリン化を受けるであろう。脱プリン化部位は、最終的に、ポリヌクレオチドの最終的な脱保護の過程で切断され、したがって全長の生成物の収率が低下する。O6修飾は、I₂/水での酸化前にキャッピング試薬による処理によって除去され得る。いくつかの例では、ポリヌクレオチド合成中のキャッピング工程を含めることで、キャッピングなしでの合成と比較して、エラー率は低下する。一例として、キャッピング工程は、無水酢酸と1-メチルイミダゾールとの混合物で基質に結合したポリヌクレオチドを処理することを含む。キャッピング工程に続いて、デバイスは随意に洗浄される。

40

50

【0123】

いくつかの例では、ヌクレオシドホスホラミダイトの添加に続いて、および随意にカップリングと1回以上の洗浄工程後に、デバイスに結合した成長している核酸が酸化される。酸化工程は、亜リン酸塩トリエステルが、自然発生のリン酸ジエステルのヌクレオシド間結合の保護された前駆体である、四配位リン酸塩トリエステルへと酸化される工程を含む。いくつかの例では、成長しているポリヌクレオチドの酸化は、随意に弱塩基（例えば、ピリジン、ルチジン、コリジン）の存在下で、ヨウ素および水による処理によって達成される。酸化は、例えば、tert-ブチルヒドロペルオキシドまたは(1S)-(+)-(10-カンファースルホニル)-オキサジリジン(CSO)を使用して、無水条件下で実行され得る。いくつかの方法では、カップリング工程は、酸化に続いて行われる。持続し得る酸化からの残留水が続くカップリングを阻害することができるため、第2のカップリング工程はデバイスの乾燥を認める。酸化の後に、デバイスおよび成長しているポリヌクレオチドは、随意に洗浄される。いくつかの例では、酸化の工程は、ポリヌクレオチドホスホロチオエートを得る硫化工程と置き換えられ、ここでカップリング工程は硫化後に実行され得る。限定されないが、3-(ジメチルアミノメチリデン)アミノ-3H-1,2,4-ジチアゾール-3-チオン、DDTT、Beaucage試薬としても知られている3H-1,2-ベンゾジチオール-3-オン1,1-ジオキシド、およびN,N',N'-テトラエチルチウラムジスルフィド(TETD)を含む、多くの試薬が、効率的な硫黄移動を可能にする。

10

【0124】

20

ヌクレオシド取り込みの続くサイクルを、カップリングを介して生じさせるために、デバイスに結合した成長しているポリヌクレオチドの保護された5'末端は除去され、その結果、一次ヒドロキシル基が次のヌクレオシドホスホラミダイトと反応する。いくつかの例では、保護基はDMTであり、分解はジクロロメタン中でトリクロロ酢酸により生じる。時間を延長して、または推奨されるよりも強い酸の溶液を用いて脱トリチル化を行うことで、個体担体に結合したポリヌクレオチドの脱プリン化の増大がもたらされる場合もあり、したがって望ましい完全長の生成物の収率は低下する。本明細書に記載される開示の方法および組成物は、望ましくない脱プリン化反応を制限する制御された分解条件を提供する。いくつかの例では、デバイスに結合したポリヌクレオチドは、分解後に洗浄される。いくつかの例では、分解後の効率的な洗浄は、低いエラー率を有する合成ポリヌクレオチドに寄与する。

30

【0125】

ポリヌクレオチド合成のための方法は、典型的に、以下の工程の反復配列を含む：活性化された面、リンカーまたは事前に脱保護されたモノマーのいずれかと連結するための、保護されたモノマーの活発に官能化された面（例えば遺伝子座）への適用；続いて適用される保護されたモノマーと反応性であるようにする、適用されたモノマーの脱保護；および、連結のための別の保護モノマーの適用。1つ以上の中間工程は、酸化または硫化を含む。いくつかの例では、1つ以上の洗浄工程は、工程の1つまたはすべての前または後にある。

【0126】

40

ホスホラミダイトベースのポリヌクレオチド合成のための方法は、一連の化学工程を含む。いくつかの例では、合成方法の一つ以上の工程は、試薬のサイクリングを伴い、方法の一つ以上の工程は、工程に有用な試薬のデバイスへの適用を含む。例えば、試薬は、一連の液体沈着および真空乾燥の工程によって循環される。ウェル、マイクロウェル、チャネル等の三次元機構を含む基質に関して、試薬は、随意にウェルおよび/またはチャネルを介してデバイスの1つ以上の領域に通される。

【0127】

本明細書に記載される方法およびシステムは、ポリヌクレオチドの合成のためのポリヌクレオチド合成デバイスに関する。合成は平行して行われてもよい。例えば、少なくとも、または少なくとも約2、3、4、5、6、7、8、9、10、11、12、13、14

50

、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、10000、50000、75000、100000、またはそれ以上のポリヌクレオチドが、平行して合成され得る。平行して合成され得るポリヌクレオチドの総数は、2～100000、3～50000、4～10000、5～1000、6～900、7～850、8～800、9～750、10～700、11～650、12～600、13～550、14～500、15～450、16～400、17～350、18～300、19～250、20～200、21～150、22～100、23～50、24～45、25～40、30～35であり得る。当業者は、平行して合成されたポリヌクレオチドの総数が、これらの値のいずれかに制約される任意の範囲内、例えば25～100にあり得ることを認識する。平行して合成されたポリヌクレオチドの総数は、範囲のエンドポイントとして機能する値のいずれかによって定義された任意の範囲内にあり得る。デバイス内で合成されたポリヌクレオの総モル質量またはポリヌクレオの各々のモル質量は、少なくとも、または少なくとも約10、20、30、40、50、100、250、500、750、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000の、25000、50000、75000、100000ピコモル、またはそれ以上であり得る。デバイス内のポリヌクレオチドの各々の長さ、またはポリヌクレオチドの平均長は、少なくとも、または少なくとも約10、15、20、25、30、35、40、45、50、100、150、200、300、400、500のヌクレオチド、またはそれ以上であり得る。デバイス内のポリヌクレオチドの各々の長さ、またはその平均長は、多くとも、または多くとも約500、400、300、200、150、100、50、45、35、30、25、20、19、18、17、16、15、14、13、12、11、10のヌクレオチド、またはそれ以下であり得る。デバイス内のポリヌクレオチドの各々の長さ、またはその平均長は、10～500、9～400、11～300、12～200、13～150、14～100、15～50、16～45、17～40、18～35、19～25の間であり得る。当業者は、デバイス内のポリヌクレオチドの各々の長さ、またはその平均長が、これらの値のいずれかに制約される任意の範囲内、例えば100～300にあり得ることを認識する。デバイス内のポリヌクレオチドの各々の長さ、またはその平均長は、範囲のエンドポイントとして機能する値のいずれかによって定義された任意の範囲内にあり得る。

【0128】

本明細書において提供される面上でのポリヌクレオチド合成の方法は、速い速度での合成を可能にする。例えば、少なくとも1時間あたり3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、70、80、90、100、125、150、175、200ヌクレオチド、またはより多くが合成される。ヌクレオチドは、アデニン、グアニン、チミン、シトシン、ウリジン構築ブロック、またはそれらのアナログ/修飾されたバージョンを含む。いくつかの例では、ポリヌクレオチドのライブラリは基質上で平行に合成される。例えば、約、または少なくとも約100；1，000；10，000；30，000；75，000；100，000；1，000，000；2，000，000；3，000，000；4，000，000；または5，000，000の分解された遺伝子座を含むデバイスは、別個のポリヌクレオチドの少なくとも同じ数の合成を支持することができ、別個の配列をコードするポリヌクレオチドは分解された遺伝子座において合成される。いくつかの例では、ポリヌクレオチドライブラリは、約3ヶ月、2ヶ月、1ヶ月、3週間、15、14、13、12、11、10、9、8、7、6、5、4、3、2日、24時間未満、あるいはより短い期間で、本明細書に記載の低エラー率で、デバイス上で合成される。いくつかの例では、基質と本明細書に記載の方法を使用して低エラー率で合成されたポリヌクレオチドライブラリから組み立てられたより大きな核酸は、約3ヶ月、2ヶ月、1ヶ月、3週間、15、14、

13、12、11、10、9、8、7、6、5、4、3、2日、24時間未満、あるいはより短い期間で調製される。

【0129】

いくつかの例では、本明細書に記載の方法は、複数のコドン部位で異なっている変異ポリヌクレオチドを含むポリヌクレオチドのライブラリの生成を提供する。いくつかの例では、ポリヌクレオチドは、1つの部位、2つの部位、3つの部位、4つの部位、5つの部位、6つの部位、7つの部位、8つの部位、9つの部位、10の部位、11の部位、12の部位、13の部位、14の部位、15の部位、16の部位、17の部位、18の部位、19の部位、20の部位、30の部位、40の部位、50の部位、またはより多くの変異コドン部位を有し得る。

10

【0130】

いくつかの例では、変異コドン部位の1つ以上の部位は隣接し得る。いくつかの例では、変異コドン部位の1つ以上の部位は隣接しておらず、1、2、3、4、5、6、7、8、9、10、またはより多くのコドンによって分離され得る。

【0131】

いくつかの例では、ポリヌクレオチドは変異コドン部位の多数の部位を含んでもよく、すべての変異コドン部位は互いに隣接し、変異コドン部位の広がり形成する。いくつかの例では、ポリヌクレオチドは変異コドン部位の多数の部位を含んでもよく、変異コドン部位のいずれも互いに隣接していない。いくつかの例では、ポリヌクレオチドは変異コドン部位の多数の部位を含んでもよく、いくつかの変異コドン部位は互いに隣接し、変異コドン部位の広がり形成し、および変異コドン部位のいくつかは、互いに隣接していない。

20

【0132】

図5を参照すると、図5は、より短いポリヌクレオチドからの核酸（例えば遺伝子）の合成のための典型的な処理ワークフローを例示する。ワークフローは概して、以下のフェーズに分けられる：（1）一本鎖ポリヌクレオチドライブラリのデノボ合成、（2）より大きなフラグメントを形成するためのポリヌクレオチドの接合、（3）エラー補正、（4）品質管理、および（5）発送。デノボ合成に先立って、意図した核酸配列または一群の核酸配列が、あらかじめ選択される。例えば、一群の遺伝子は生成のためにあらかじめ選択される。

30

【0133】

一旦、生成のための大きなポリヌクレオチドが選択されると、ポリヌクレオチドのあらかじめ決定されたライブラリは、デノボ合成のために設計される。様々な適切な方法が、高密度ポリヌクレオチドアレイの生成に関して知られている。ワークフロー例において、デバイス面レイヤ（501）が提供される。例において、面の化学作用は、ポリヌクレオチド合成プロセスを改善するために修正される。低い界面エネルギーの領域は、液体をはね返すように生成され、他方で高い界面エネルギーの領域は、液体を引きつけるように生成される。面自体が平面形状であってもよく、または表面エリアで増加する突部またはマイクロウェル等の、形状の変化を有してもよい。ワークフローの例において、国際公開第WO/2015/021080号で開示されるように、選択された高い界面エネルギー分子は、DNAの化学作用を支持する二元機能を果たし、該出願はその全体が参照により本明細書に組み込まれる。

40

【0134】

ポリヌクレオチドアレイのインサイチュ調製は個体担体上で行われ、および平行して多数のオリゴマーを伸長させるために単一のヌクレオチド拡張プロセスを利用する。ポリヌクレオチドシンセサイザー等の材料沈着デバイスは、多数のポリヌクレオチドが並行して伸長するように段階的手法で試薬を放出するよう設計されており、あらかじめ決定された核酸配列を有するオリゴマーを生成するために一度に1試薬が放出される（502）。いくつかの例では、ポリヌクレオチドはこの段階で面から切断される。切断は、例えばアンモニアまたはメチルアミンを用いたガスによる切断を含む。

50

【0135】

生成されたポリヌクレオチドライブラリは、反応チャンバに配置される。この典型的なワークフローでは、反応チャンバ（「ナノリアクタ」とも呼ばれる）は、シリコンコーティングされたウェルであり、PCR試薬を含み、およびポリヌクレオチドライブラリ上に下げられる（503）。ポリヌクレオチドの密封（504）の前または後に、基質からポリヌクレオチドを放出するために試薬が加えられる。典型的なワークフローでは、ポリヌクレオチドはナノリアクタの密封後に放出される（505）。一旦、放出されると、一本鎖ポリヌクレオチドのフラグメントは、DNAの全長範囲全体に及ぶようにハイブリダイズする。部分的なハイブリダイゼーション（505）が可能であり、なぜなら合成ポリヌクレオチドの各々は、集団内の少なくとも1つの他のポリヌクレオチドをオーバーラップする少しの部分を含むように設計されるからである。

10

【0136】

ハイブリダイゼーション後に、PCR反応が開始される。ポリメラーゼサイクル中に、ポリヌクレオチドは相補的なフラグメントにアニール化され、隙間がポリメラーゼによって満たされる。各サイクルは、ポリヌクレオチドが互いを発見するかどうかは無作為に依存して、様々なフラグメントの長さを増加させる。フラグメント間の相補性は、二本鎖DNAの完全で大きなスパンを形成することを可能にする（506）。

【0137】

PCRが完了した後、ナノリアクタはデバイスから分離され（507）、PCRのためのプライマーを有するデバイスとのインタラクションのために配置される（508）。密封の後、ナノリアクタはPCRにさらされ（309）、およびより大きな核酸が増幅される。PCRの後（510）、ナノチャンバが開けられ（511）、エラー補正試薬が添加され（512）、チャンバが密封され（513）、そして、二本鎖PCR増幅産物からの相補性の乏しいミスマッチ塩基対および/または鎖を削除するためにエラー補正反応が生じる（514）。ナノリアクタが開かれ、分割される（515）。エラー補正された生成物は次に、PCRおよび分子バーコーディング等の追加の処理工程にさらされ、そして出荷（523）のために包装される（522）。

20

【0138】

いくつかの例では、品質管理測定が行われる。エラー補正の後に、品質管理工程は、例えば、エラー補正された生成物の増幅のためのシーケンシングプライマーを有するウェーハとのインタラクション（516）、エラー補正された増幅産物を含むチャンバへのウェーハの密封（517）、および増幅の追加サイクルを行う工程（518）を含む。ナノリアクタが開けられ（519）、そして生成物がプールされ（520）、配列される（521）。合格品質であるという管理決定がなされた後、包装された生成物（522）が出荷（523）のために承認される。

30

【0139】

いくつかの例では、図5等のワークフローによって生成される核酸は、本明細書に開示の重複するプライマーを使用して、突然変異誘発にさらされる。いくつかの例では、プライマーのライブラリは、個体担体上でのインサイチュ調製によって生成され、平行に多数のオリゴマーを伸長させるために単一のヌクレオチド拡張プロセスを利用する。ポリヌクレオチドシンセサイザー等の材料沈着デバイスは、多数のポリヌクレオチドが並行して伸長するように段階的手法で試薬を放出するよう設計されており、あらかじめ決定された核酸配列を有するオリゴマーを生成するために一度に1試薬が放出される（502）。

40

【0140】

大きなポリヌクレオチドライブラリは低エラー率を有する。

【0141】

提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオチドに関する平均エラー率は、1000分の1未満、1250分の1未満、1500分の1未満、2000分の1未満で、3000分の1未満、またはより低い頻度であり得る。いくつかの例では、提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオ

50

チドに関する平均エラー率は、 $1/500$ 、 $1/600$ 、 $1/700$ 、 $1/800$ 、 $1/900$ 、 $1/1000$ 、 $1/1100$ 、 $1/1200$ 、 $1/1250$ 、 $1/1300$ 、 $1/1400$ 、 $1/1500$ 、 $1/1600$ 、 $1/1700$ 、 $1/1800$ 、 $1/1900$ 、 $1/2000$ 、 $1/3000$ 未満、またはより低い。いくつかの例では、提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオチドに関する平均エラー率は、 $1/1000$ 未満である。

【0142】

いくつかの例では、提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオチドに関する合計エラー率は、あらかじめ決定された配列と比較して、 $1/500$ 、 $1/600$ 、 $1/700$ 、 $1/800$ 、 $1/900$ 、 $1/1000$ 、 $1/1100$ 、 $1/1200$ 、 $1/1250$ 、 $1/1300$ 、 $1/1400$ 、 $1/1500$ 、 $1/1600$ 、 $1/1700$ 、 $1/1800$ 、 $1/1900$ 、 $1/2000$ 、 $1/3000$ 未満、またはより低い。いくつかの例では、提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオチドに関する合計エラー率は、 $1/500$ 、 $1/600$ 、 $1/700$ 、 $1/800$ 、 $1/900$ 、または $1/1000$ 未満である。いくつかの例では、提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオチドに関する合計エラー率は、 $1/1000$ 未満である。

10

【0143】

いくつかの例では、エラー補正酵素が、使用可能な提供される方法とシステムを使用してライブラリ内で合成されたポリヌクレオチドに使用され得る。いくつかの例では、エラー補正によるポリヌクレオチドに関する合計エラー率は、あらかじめ決定された配列と比較して、 $1/500$ 、 $1/600$ 、 $1/700$ 、 $1/800$ 、 $1/900$ 、 $1/1000$ 、 $1/1100$ 、 $1/1200$ 、 $1/1300$ 、 $1/1400$ 、 $1/1500$ 、 $1/1600$ 、 $1/1700$ 、 $1/1800$ 、 $1/1900$ 、 $1/2000$ 、 $1/3000$ 未満、またはそれ以下であり得る。いくつかの例では、提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオチドに関するエラー補正による合計エラー率は、 $1/500$ 、 $1/600$ 、 $1/700$ 、 $1/800$ 、 $1/900$ 、または $1/1000$ 未満であり得る。いくつかの例では、提供されるシステムと方法を使用してライブラリ内で合成されたポリヌクレオチドに関するエラー補正での合計エラー率は、 $1/1000$ 未満であり得る。

20

30

【0144】

エラー率は、遺伝子変異体のライブラリの生成のための遺伝子合成の値を制限し得る。 $1/300$ のエラー率では、 1500 の塩基対遺伝子におけるクローンの約 0.7% が正しくなる。ポリヌクレオチド合成からのエラーのほとんどは、結果としてフレームシフト突然変異をもたらすため、そのようなライブラリにおけるクローンの 99% 以上が、全長タンパク質を生成しない。エラー率を 75% 低下させることによって、正しいクローンのフラクションは 40 倍増加するだろう。本開示の方法および組成物は、超並列および時間効率の良い方法で可能になる合成の質の改善およびエラー補正方法の適用性の両方によって、一般に観察される遺伝子合成方法よりも低いエラー率での大規模なオリゴヌクレオチドと、遺伝子ライブラリの速いデノボ合成を可能にする。したがって、ライブラリは、ライブラリ全体にわたって、またはライブラリの 80% 、 85% 、 90% 、 93% 、 95% 、 96% 、 97% 、 98% 、 99% 、 99.5% 、 99.8% 、 99.9% 、 99.95% 、 99.98% 、 99.99% 、またはそれ以上にわたって、塩基の挿入、欠失、置換、あるいは $1/300$ 、 $1/400$ 、 $1/500$ 、 $1/600$ 、 $1/700$ 、 $1/800$ 、 $1/900$ 、 $1/1000$ 、 $1/1250$ 、 $1/1500$ 、 $1/2000$ 、 $1/2500$ 、 $1/3000$ 、 $1/4000$ 、 $1/5000$ 、 $1/6000$ 、 $1/7000$ 、 $1/8000$ 、 $1/9000$ 、 $1/10000$ 、 $1/12000$ 、 $1/15000$ 、 $1/20000$ 、 $1/25000$ 、 $1/30000$ 、 $1/40000$ 、 $1/50000$ 、 $1/60000$ 、 $1/70000$ 、 $1/80000$ 、 $1/90000$ 、 $1/100000$ 、 $1/125000$ 、 $1/150000$ 、 $1/200000$ 、 $1/300000$ 、 $1/400000$

40

50

、1/500000、1/600000、1/700000、1/800000、1/900000、1/1000000未満、またはそれ以下である合計のエラー率を伴って、合成され得る。本開示の方法および組成物はさらに、あらかじめ決定された／あらかじめ選択された配列と比較して、エラーのない配列に関連するライブラリの少なくともサブセットにおいて、ポリヌクレオチドまたは遺伝子の少なくとも30%、40%、50%、60%、70%、75%、80%、85%、90%、93%、95%、96%、97%、98%、99%、99.5%、99.8%、99.9%、99.95%、99.98%、99.99%、またはそれ以上に関連付けられる低いエラー率での大規模な合成ポリヌクレオチドと遺伝子ライブラリに関する。いくつかの例では、ライブラリ内の単離した量でのポリヌクレオチドまたは遺伝子の少なくとも30%、40%、50%、60%、70%、75%、80%、85%、90%、93%、95%、96%、97%、98%、99%、99.5%、99.8%、99.9%、99.95%、99.98%、99.99%、またはそれ以上は、同じ配列を有している。いくつかの例では、95%、96%、97%、98%、99%、99.5%、99.6%、99.7%、99.8%、99.9%、またはそれ以上の類似性または同一性に関連する、ポリヌクレオチドまたは遺伝子の少なくとも30%、40%、50%、60%、70%、75%、80%、85%、90%、93%、95%、96%、97%、98%、99%、99.5%、99.8%、99.9%、99.95%、99.98%、99.99%、またはそれ以上は、同じ配列を有している。いくつかの例では、ポリヌクレオチドまたは遺伝子上の特定の遺伝子座に関連するエラー率は、最適化される。したがって、大規模なライブラリの一部としての1つ以上のポリヌクレオチドまたは遺伝子の複数の選択された遺伝子座の任意の遺伝子座は各々、1/300、1/400、1/500、1/600、1/700、1/800、1/900、1/1000、1/1250、1/1500、1/2000、1/2500、1/3000、1/4000、1/5000、1/6000、1/7000、1/8000、1/9000、1/10000、1/12000、1/15000、1/20000、1/25000、1/30000、1/40000、1/50000、1/60000、1/70000、1/80000、1/90000、1/100000、1/125000、1/150000、1/200000、1/300000、1/400000、1/500000、1/600000、1/700000、1/800000、1/900000、1/1000000未満、またはそれより低いエラー率を有し得る。様々な例では、そのようなエラーを最適化した遺伝子座は、少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、4000、5000、6000、7000、8000、9000、10000、30000、50000、75000、100000、500000、1000000、2000000、3000000、またはそれ以上の遺伝子座を含み得る。エラーを最適化した遺伝子座は、少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、4000、5000、6000、7000、8000、9000、10000、30000、75000、100000、500000の、1000000、2000000、3000000、またはそれ以上のポリヌクレオチドまたは遺伝子に分布され得る。

【0145】

エラー率は、エラー補正の有無に関わらず達成され得る。エラー率は、ライブラリ全体にわたって、またはライブラリの80%、85%、90%、93%、95%、96%、97%、98%、99%、99.5%、99.8%、99.9%、99.95%、99.98%、99.99%、またはそれ以上にわたって達成され得る。

【0146】

コンピュータシステム

【0147】

本明細書に記載されるシステムのいずれかは、コンピュータに動作可能に接続されてもよく、およびローカルまたは遠隔のいずれかでコンピュータを通じて自動化されてもよい。様々な例において、本開示の方法とシステムはさらに、コンピュータシステムのソフトウェアプログラムとその使用を含む。したがって、材料沈着デバイスの動き、分配動作、および真空発動の組織化と同期等の、分配/真空/再補充機能の同期のためのコンピュータ制御は、本開示の範囲内にある。コンピュータシステムは、ユーザーに特異的な塩基配列と、基質の特定領域に適正な試薬を送達するための材料沈着デバイスの位置をインターフェースで接続するようにプログラムされてもよい。

10

【0148】

図6に例示されるコンピュータシステム(600)は、固定媒体(612)を有するサーバー(609)に随意に接続され得る、媒体(611)および/またはネットワークポート(605)からの命令を読み取ることができる論理装置として理解してもよい。図6に示されるシステム等のシステムは、CPU(601)、ディスクドライブ(603)、キーボード(615)および/またはマウス(616)等の随意の入力デバイス、および随意のモニタ(607)を含むことができる。データ通信は、示された通信媒体を通して、ローカル位置または遠隔位置のサーバーへと到達され得る。通信媒体は、データを送信および/または受信する任意の手段を含むことができる。例えば、通信媒体は、ネットワーク接続、無線接続またはインターネット接続であり得る。そのような接続は、ワールド

20

【0149】

図7は、本開示の例としての事例に関連して使用することができるコンピュータシステム(700)の第1の例であるアーキテクチャを例示するブロック図である。図7で示されるように、例としてのコンピュータシステムは、命令を処理するためのプロセッサ(702)を含むことができる。プロセッサの非限定的な例として、以下があげられる: Intel Xeon(商標)プロセッサ、AMD Opteron(商標)プロセッサ、Samsung 32-bit RISC ARM 1176JZ(F)-Sv1.0(商標)プロセッサ、ARM Cortex-A8 Samsung S5PC100(商標)プロセッサ、ARM Cortex-A8 Apple A4(商標)プロセッサ、Marvell PXA 930(商標)プロセッサ、または機能的に同等なプロセッサ。実行の多数のスレッドが、並列処理に使用され得る。いくつかの例では、多数のプロセッサ、または多数のコアを備えたプロセッサはまた、単一のコンピュータシステムにおいて、またはクラスタにおいて使用することができ、または複数のコンピュータ、携帯電話、および/またはパーソナルデータアシスタントデバイスを含むネットワークを通じてシステム全体に分配され得る。

30

【0150】

図7に例示されるように、プロセッサ(702)によって最近使用された、または頻繁に使用される命令またはデータのための高速メモリを提供するために、高速キャッシュ(704)がプロセッサ(702)に接続され、または組み込まれ得る。プロセッサ(702)は、プロセッサバス(708)によってノースブリッジ(706)に接続される。ノースブリッジ(706)は、メモリバス(712)によってランダムアクセスメモリ(RAM)(710)に接続され、プロセッサ(702)によってRAM(710)に対するアクセスを管理する。ノースブリッジ(706)はまた、チップセットバス(716)によってサウスブリッジ(714)に接続される。サウスブリッジ(714)は、順に周辺バス(718)に接続される。周辺バスは、例えば、PCI、PCI-X、PCI Express、または他の周辺バスであり得る。ノースブリッジとサウスブリッジは、しばしばプロセッサチップセットと呼ばれ、プロセッサ、RAM、および周辺バス(718)

40

50

にある周辺コンポーネントの間のデータ転送を管理する。いくつかの代替的なアーキテクチャでは、ノースブリッジの機能性は、別個のノースブリッジチップを使用する代わりに、プロセッサに組み込むことができる。いくつかの例では、システム(700)は、周辺バス(718)に付けられたアクセラレータカード(722)を含むことができる。アクセラレータは、フィールドプログラマブルゲートアレイ(FPGA)、または特定の処理を加速するための他のハードウェアを含むことができる。例えば、アクセラレータは、適応データの再構成のために、または拡張された設定処理に使用される代数式を評価するために使用され得る。

【0151】

ソフトウェアとデータは外部ストレージ(724)に保存され、プロセッサによる使用のためにRAM(710)および/またはキャッシュ(704)にロードされ得る。システム(700)は、システム資源を管理するためのオペレーティングシステムを含み、オペレーティングシステムの非限定的な例として、以下があげられる: Linux、Windows(商標)、MACOS(商標)、BlackBerry OS(商標)、iOS(商標)、および他の機能的に同等なオペレーティングシステムと共に、本開示の例としての事例に係るデータの記憶および最適化を管理するためのオペレーティングシステム上で実行されるアプリケーションソフトウェア。この例において、システム(700)はまた、ネットワーク接続ストレージ(NAS)、および分散並列処理に使用可能な他のコンピュータシステム等の、外部ストレージにネットワークインターフェースを提供するための周辺バスに接続されたネットワークインターフェースカード(NIC)(720)と(721)を含むことができる。

【0152】

図8は、複数のコンピュータシステム(802a)と(802b)、複数の携帯電話とパーソナルデータアシスタント(802c)、およびネットワーク接続ストレージ(NAS)(804a)と(804b)を備えるネットワーク(800)を示す図である。例としての事例において、システム(802a)、(802b)および(802c)は、データ記憶を管理し、ネットワーク接続ストレージ(NAS)(804a)および(804b)に保存されたデータのためのデータアクセスを最適化することができる。数学的モデルをデータに使用することができ、およびコンピュータシステム(802a)と(802b)、および携帯電話とパーソナルデータアシスタントシステム(802c)にわたる分散並列処理を使用して評価することができる。コンピュータシステム(802a)と(802b)、および携帯電話とパーソナルデータアシスタントシステム(802c)はまた、ネットワークアタッチトストレージ(NAS)(804a)と(804b)に保存されたデータの適応データ再構成に並列処理を提供することができる。図8は、単に一例を例示するものであり、種々様々な他のコンピュータアーキテクチャおよびシステムが、本開示の様々な例と共に使用され得る。例えば、並列処理を提供するためにブレードサーバーを使用することができる。並列処理を提供するために、プロセッサブレードをバックプレーンを介して接続することができる。ストレージもまたバックプレーンに接続することができ、または別のネットワークインターフェースを介してネットワークアタッチトストレージ(NAS)として接続することができる。いくつかの例としての事例において、プロセッサは、別個のメモリ空間を維持することができ、ネットワークインターフェース、バックプレーン、または他のプロセッサによる並列処理のための他のコネクタを介してデータを送信することができる。他の例では、プロセッサのいくつかまたはすべては、共有の仮想アドレスメモリ空間を使用することができる。

【0153】

図9は、例としての事例に係る共有の仮想アドレスメモリ空間を使用するマルチプロセッサコンピュータシステム(900)のブロック図である。システムは、共有メモリサブシステム(904)にアクセス可能な複数のプロセッサ(902a-f)を含む。システムは、複数のプログラム可能なハードウェアメモリアルゴリズムプロセッサ(MAP)(906a-f)をメモリサブシステム(904)に組み込む。各MAP(906a-f)

は、メモリ(908a-f)と、1つ以上のフィールドプログラマブルゲートアレイ(FPGA)(910a-f)を含むことができる。MAPは、設定可能な機能ユニットを提供し、特定のアルゴリズムまたはアルゴリズムの部分が、それぞれのプロセッサと密に協働した処理のためにFPGA(910a-f)に提供され得る。例えば、MAPは、データモデルに関する代数式を評価するために、および例としての事例において適応データ再構成を実行するために、使用することができる。この例において、各MAPは、これらの目的のためにすべてのプロセッサによって地球規模でアクセス可能である。1つの構成では、各MAPは、関連するメモリ(908a-f)にアクセスするためにダイレクトメモリアクセス(DMA)を使用することができ、これによって、それぞれのマイクロプロセッサ(902a-f)から独立して、およびそれらとは非同期的にタスクを実行することができる。この構成では、MAPは、アルゴリズムのパイプライン処理および並列実行のために別のMAPに結果を直接供給することができる。

10

20

30

40

50

【0154】

上記のコンピュータアーキテクチャおよびシステムは単なる例であり、一般的なプロセッサ、コプロセッサ、FPGAおよび他のプログラマブルロジックデバイス、システムオンチップ(SOC)、特定用途向け集積回路(ASIC)、および他の処理素子と論理素子の任意の組み合わせを使用するシステムを含む、広範な他のコンピュータ、携帯電話、パーソナルデータアシスタントのアーキテクチャおよびシステムが、例としての事例に関連して使用され得る。いくつかの例では、コンピュータシステムのすべてまたは一部は、ソフトウェアまたはハードウェアに実装され得る。ランダムアクセスメモリ、ハードドライブ、フラッシュメモリ、テープドライブ、ディスクアレイ、ネットワークアタッチトストレージ(NAS)および他のローカルまたは分散データストレージデバイスとシステムを含む、あらゆる種類のデータストレージメディアが、例としての事例に関連して使用され得る。

【0155】

例としての事例において、コンピュータシステムは、上記または他のコンピュータアーキテクチャおよびシステムのいずれかにおいて実行されるソフトウェアモジュールを使用して実装され得る。他の例において、システムの機能は、ファームウェア、図9で参照されるようなフィールドプログラマブルゲートアレイ(FPGA)等のプログラマブルロジックデバイス、システムオンチップ(SOC)、特定用途向け集積回路(ASIC)、または他の処理素子および論理素子等において部分的または完全に実装され得る。例えば、セットプロセッサ(Set Processor)およびオプティマイザ(Optimizer)は、図7に例示されるアクセラレータカード(722)などのハードウェアアクセラレータカードの使用によって、ハードウェアアクセラレーションで実装され得る。

【0156】

追加の方法と組成物

【0157】

本明細書で提供されるのは、ポリヌクレオチドライブラリを生成するための方法であって、該方法は：少なくとも約5000の非同一のポリヌクレオチドをコードするあらかじめ決定された配列を提供する工程；面を有する構造を提供する工程であって、面は複数のクラスタを含む、工程；少なくとも約5000の非同一のポリヌクレオチドを合成する工程であって、少なくとも5000の非同一のポリヌクレオチドの各々は異なる遺伝子座を伸長させる、工程；および、ポリヌクレオチドライブラリを形成するために少なくとも5000の非同一のポリヌクレオチドを増幅する工程を含み、ここで、少なくとも5000の非同一のポリヌクレオチドの約80パーセントより多くは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約2倍以内の量で表示される。さらに本明細書で提供される方法において、少なくとも5000の非同一のポリヌクレオチドの約80パーセントより多くは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約1.5倍以内の量で表示される。さらに本明細書で提供される方法において、少なくとも5000の非同一のポリヌクレオチドの約90パーセントより多くは、ポリヌクレオチドライブラリに

関する平均表示の少なくとも約 2 倍以内の量で表示される。さらに本明細書で提供される方法において、少なくとも 5 0 0 0 の非同一のポリヌクレオチドの約 9 0 パーセントより多くは、ポリヌクレオチドライブラリに関する平均表示の少なくとも約 1 . 5 倍以内の量で表示される。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、クラスタ化されていない遺伝子座の面を有する構造を使用する方法からの増幅産物と比較して、より少ないドロップアウトを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは少なくとも約 1 0 % の G C の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは、多くとも 9 5 % の G C の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは、約 1 0 % から約 9 5 % の G C の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドの約 3 0 % より多くが、1 0 % から 3 0 % または 7 0 % から 9 0 % の G C の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドの約 1 5 % 未満が、1 0 % から 3 0 % または 6 0 % から 9 0 % の G C の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは少なくとも約 1 0 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは、多くとも 9 5 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは、約 1 0 % から約 9 5 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドの約 3 0 % より多くが、1 0 % から 3 0 % または 7 0 % から 9 0 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドの約 1 5 % 未満が、1 0 % から 3 0 % または 6 0 % から 9 0 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは少なくとも約 1 0 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは、多くとも 9 5 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドは、約 1 0 % から約 9 5 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドの約 3 0 % より多くが、1 0 % から 3 0 % または 7 0 % から 9 0 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約 5 0 0 0 の非同一のポリヌクレオチドの約 1 5 % 未満が、1 0 % から 3 0 % または 6 0 % から 9 0 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは変形ライブラリをコードする。さらに本明細書で提供される方法において、少なくとも 5 0 0 0 の非同一のポリヌクレオチドライブラリは少なくとも 1 つの遺伝子をコードする。さらに本明細書で提供される方法において、少なくとも 5 0 0 0 の非同一のポリヌクレオチドライブラリは少なくとも 5 0 の遺伝子をコードする。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは少なくとも 1 つの遺伝子をコードする。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、抗体、酵素またはペプチドの少なくとも一部をコードする。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、エラーを補正せずに、あらかじめ決定された配列と比較して、5 0 0 塩基に 1 未満の合計エラー比率を有する。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、エラーを補正せずに、あらかじめ決定された配列と比較して、1 0 0 0 塩基に 1 未満の合計エラー比率を有する。さらに本明細書で提供される方法において、あらかじめ決定された配列は少なくとも 7 0 0 , 0 0 0 の非同一のポリ

ヌクレオチドをコードする。さらに本明細書で提供される方法において、各クラスタは、ポリヌクレオチド合成のための50～約500の遺伝子座を含む。さらに本明細書で提供される方法において、各クラスタは、ポリヌクレオチド合成のための最大約500の遺伝子座を含む。

【0158】

本明細書で提供されるのは、ポリヌクレオチドライブラリを生成するための方法であって、該方法は：少なくとも約5000の非同一のポリヌクレオチドをコードするあらかじめ決定された配列を提供する工程；面を有する構造を提供する工程であって、面は複数のクラスタを含む、工程；少なくとも約5000の非同一のポリヌクレオチドを合成する工程であって、少なくとも5000の非同一のポリヌクレオチドの各々は異なる遺伝子座を伸長させる、工程；および、ポリヌクレオチドライブラリを形成するために少なくとも5000の非同一のポリヌクレオチドを増幅する工程であって、ポリヌクレオチドライブラリは、増幅反応後に75%より大きい正確な配列比率を有する工程、を含む。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、増幅反応後に85%より大きい正確な配列比率を有する。さらに本明細書で提供される方法において、少なくとも約5000の非同一のポリヌクレオチドは少なくとも約10%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約5000の非同一のポリヌクレオチドは、多くとも95%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約5000の非同一のポリヌクレオチドは、約10%から約95%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約5000の非同一のポリヌクレオチドの約30%より多くが、10%から30%または70%から90%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、少なくとも約5000の非同一のポリヌクレオチドの約15%未満が、10%から30%または60%から90%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、クラスタ化されていない遺伝子座の面を有する構造を使用する方法からの増幅産物と比較して、より少ないドロップアウトを含む。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは変形ライブラリをコードする。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは少なくとも1つの遺伝子をコードする。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、抗体、酵素またはペプチドの少なくとも一部をコードする。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、エラーを補正せずに、あらかじめ決定された配列と比較して、500塩基に1未満の合計エラー比率を有する。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリは、エラーを補正せずに、あらかじめ決定された配列と比較して、1000塩基に1未満の合計エラー比率を有する。さらに本明細書で提供される方法において、あらかじめ決定された配列は少なくとも700,000の非同一のポリヌクレオチドをコードする。さらに本明細書で提供される方法において、少なくとも5000の非同一のポリヌクレオチドは少なくとも1つの遺伝子をコードする。さらに本明細書で提供される方法において、少なくとも5000の非同一のポリヌクレオチドは少なくとも50の遺伝子をコードする。さらに本明細書で提供される方法において、各クラスタは、ポリヌクレオチド合成のための50～約500の遺伝子座を含む。さらに本明細書で提供される方法において、各クラスタは、ポリヌクレオチド合成のための最大約500の遺伝子座を含む。

【0159】

本明細書で提供されるのは、少なくとも5000の非同一のポリヌクレオチドを含む核酸ライブラリであって、少なくとも5000の非同一のポリヌクレオチドは合成されたポリヌクレオチドの増幅産物であり、および少なくとも5000の非同一のポリヌクレオチドの約80%より多くが、核酸ライブラリに関する平均表示の少なくとも約2倍以内の量で表示される。さらに本明細書において核酸ライブラリが提供され、ここで少なくとも5000の非同一のポリヌクレオチドの約80パーセントより多くは、核酸ライブラリに関

する平均表示の少なくとも約 1.5 倍以内の量で表示される。さらに本明細書において核酸ライブラリが提供され、ここで少なくとも 5000 の非同一のポリヌクレオチドの約 90 パーセントより多くは、核酸ライブラリに関する平均表示の少なくとも約 2 倍以内の量で表示される。さらに本明細書において核酸ライブラリが提供され、ここで少なくとも 5000 の非同一のポリヌクレオチドの約 90 パーセントより多くは、核酸ライブラリに関する平均表示の少なくとも約 1.5 倍以内の量で表示される。さらに本明細書において核酸ライブラリが提供され、核酸ライブラリは、クラスタ化されていない遺伝子座の面を有する構造を使用する方法からの増幅産物と比較して、より少ないドロップアウトを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは少なくとも約 10 % の GC の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは、多くとも 95 % の GC の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは、約 10 % から約 95 % の GC の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドの約 30 % より多くが、10 % から 30 % または 70 % から 90 % の GC の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドの約 15 % 未満が、10 % から 30 % または 60 % から 90 % の GC の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは少なくとも約 10 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは、多くとも 95 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは、約 10 % から約 95 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドの約 30 % より多くが、10 % から 30 % または 70 % から 90 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドの約 15 % 未満が、10 % から 30 % または 60 % から 90 % の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは少なくとも約 10 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは、多くとも 95 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドは、約 10 % から約 95 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドの約 30 % より多くが、10 % から 30 % または 70 % から 90 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約 5000 の非同一のポリヌクレオチドの約 15 % 未満が、10 % から 30 % または 60 % から 90 % の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、ポリヌクレオチドライブラリは変形ライブラリをコードする。さらに本明細書において核酸ライブラリが提供され、少なくとも 5000 の非同一のポリヌクレオチドは少なくとも 1 つの遺伝子をコードする。さらに本明細書において核酸ライブラリが提供され、少なくとも 5000 の非同一のポリヌクレオチドは少なくとも 50 の遺伝子をコードする。さらに本明細書において核酸ライブラリが提供され、ポリヌクレオチドライブラリは、抗体、酵素またはペプチドの少なくとも一部をコードする。さらに本明細書において核酸ライブラリが提供され、あらかじめ決定された配列は少なくとも 700,000 の非同一のポリヌクレオチドをコードする。

【0160】

10

20

30

40

50

本明細書において、少なくとも5000の非同一のポリヌクレオチドを含む核酸ライブラリが提供され、ここでGC含量が制御され、およびライブラリは、増幅反応後に75%より大きい正確な配列比率を提供する。さらに本明細書において、増幅反応後に85%より大きい正確な配列比率を有する核酸ライブラリが提供される。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは少なくとも約10%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは、多くとも95%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは、約10%から約95%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドの約30%より多くが、10%から30%または70%から90%のGCの割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドの約15%未満が、10%から30%または60%から90%のGCの割合を有するポリヌクレオチドを含む。

10

20

30

40

50

【0161】

本明細書において、少なくとも5000の非同一のポリヌクレオチドを含む核酸ライブラリが提供され、ここで反復配列含量が制御され、およびライブラリは、増幅反応後に75%より大きい正確な配列比率を提供する。さらに本明細書において核酸ライブラリが提供され、ポリヌクレオチドライブラリは、増幅反応後に85%より大きい正確な配列比率を有する。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは少なくとも約10%の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは、多くとも95%の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは、約10%から約95%の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドの約30%より多くが、10%から30%または70%から90%の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドの約15%未満が、10%から30%または60%から90%の反復配列の割合を有するポリヌクレオチドを含む。

【0162】

本明細書において、少なくとも5000の非同一のポリヌクレオチドを含む核酸ライブラリが提供され、ここで少なくとも5000の非同一のポリヌクレオチドによってコードされる二次構造含量があらかじめ選択され、およびライブラリは、増幅反応後に75%より大きい正確な配列比率を提供する。さらに本明細書において核酸ライブラリが提供され、核酸ライブラリは、増幅反応後に85%より大きい正確な配列比率を有する。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは少なくとも約10%の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは、多くとも95%の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドは、約10%から約95%の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドの約30%より多くが、10%から30%または70%から90%の二次構造の割合を有するポリヌクレオチドを含む。さらに本明細書において核酸ライブラリが提供され、少なくとも約5000の非同一のポリヌクレオチドの約15%未満が、10%から30%または60%から90%の反復配列の割合を有するポリヌクレオチドを含む。さらに本明細書において、変形ライブラリをコードする核酸ライブラリが提供される。さ

らに本明細書において、少なくとも1つの遺伝子をコードする核酸ライブラリが提供される。さらに本明細書において、抗体、酵素またはペプチドの少なくとも一部をコードする核酸ライブラリが提供される。さらに本明細書において、エラーを補正せずに、あらかじめ決定された配列と比較して、500塩基に1未満の合計エラー比率を有する核酸ライブラリが提供される。さらに本明細書において、エラーを補正せずに、あらかじめ決定された配列と比較して、1000塩基に1未満の合計エラー比率を有する核酸ライブラリが提供される。さらに本明細書において核酸ライブラリが提供され、あらかじめ決定された配列は少なくとも700,000の非同一のポリヌクレオチドをコードする。さらに本明細書において核酸ライブラリが提供され、少なくとも5000の非同一のポリヌクレオチドは少なくとも1つの遺伝子をコードする。さらに本明細書において核酸ライブラリが提供され、少なくとも5000の非同一のポリヌクレオチドは少なくとも50の遺伝子をコードする。

10

【0163】

本明細書で提供されるのは、ポリヌクレオチドライブラリの増幅のための方法であって、該方法は：少なくとも5000の非同一のポリヌクレオチドのための増幅分布を得る工程；少なくとも1つの配列特徴に基づいて、2つ以上のピンへと、増幅分布の少なくとも5000の非同一のポリヌクレオチドをクラスタリングする工程；および、あらかじめ選択された表示を有するポリヌクレオチドライブラリを生成するために、2つ以上のピンの各々において、少なくとも5000の非同一のポリヌクレオチドの頻度数に基づいて、非同一のポリヌクレオチドの各々の合成のための表示を調整する工程；あらかじめ選択された表示を有するポリヌクレオチドライブラリを合成する工程；および、あらかじめ選択された表示を有するポリヌクレオチドライブラリを増幅する工程、を含む。さらに本明細書で提供される方法において、少なくとも1つの配列特徴はパーセントGC含量である。さらに本明細書で提供される方法において、少なくとも1つの配列特徴はパーセント反復配列含量である。さらに本明細書で提供される方法において、少なくとも1つの配列特徴はパーセント二次構造含量である。さらに本明細書で提供される方法において、反復配列は3つ以上のアデニンを含む。さらに本明細書で提供される方法において、反復配列はポリヌクレオチドの片方または両方の末端にある。さらに本明細書で提供される方法において、前記ポリヌクレオチドは、標的配列を結合するために1つ以上のポリヌクレオチド配列の親和性に基づいてピンへとクラスタリングされる。さらに本明細書で提供される方法において、ピンの下30%にある配列の数は、調整前のピンの下30%にある配列の数と比較して、調整後の下流での適用に少なくとも50%多くの表示を有する。さらに本明細書で提供される方法において、ピンの上30%にある配列の数は、調整前のピンの上30%にある配列の数と比較して、調整後の下流での適用に少なくとも50%多くの表示を有する。さらに本明細書で提供される方法において、前記増幅分布は経験的に得られる。さらに本明細書で提供される方法において、前記増幅分布は、予測的アルゴリズムを通じて得られる。いくつかの例における調整は、ライブラリのポリヌクレオチドの化学量論を制御する工程を含む。

20

30

【0164】

本明細書で提供されるのは、少なくとも100,000の非同一のポリヌクレオチドを含む核酸ライブラリであって、各非同一のポリヌクレオチドは少なくとも1つの異なるエクソーム配列をコードし、および少なくとも100,000の非同一のポリヌクレオチドの少なくとも約80%は各々、ライブラリの非同一のポリヌクレオチドの各々に関して平均頻度の2x以内の量でポリヌクレオチドライブラリに存在する。さらに本明細書で提供されるのは、核酸ライブラリであって、核酸ライブラリはアンブリコンライブラリであり、および複数の非同一のポリヌクレオチドの少なくとも約80%は各々、ライブラリの非同一のポリヌクレオチドの各々に関して平均頻度の2x以内の量でアンブリコンライブラリに存在する。さらに本明細書で核酸ライブラリが提供され、ライブラリを最大55倍の理論上の読み取り深度で配列決定することで、結果として、少なくとも30倍の読み取り深度を有する塩基の少なくとも90%がもたらされる。さらに本明細書で核酸ライブラリ

40

50

が提供され、ライブラリを最大 5 5 倍の理論上の読み取り深度で配列決定することで、結果として、少なくとも 1 0 倍の読み取り深度を有する塩基の少なくとも 9 8 % がもたらされる。

【 0 1 6 5 】

本明細書で提供されるのは、ポリヌクレオチドライブラリの合成のための方法であって、該方法は：(a) 少なくとも 1 0 0 , 0 0 0 の非同一のポリヌクレオチドにあらかじめ決定された配列を提供する工程であって、各非同一のポリヌクレオチドは、ゲノム DNA の 1 つ以上の部分をコードする、工程；(b) 少なくとも 1 0 0 , 0 0 0 の非同一のポリヌクレオチドを合成する工程；および(c) ポリヌクレオチドのライブラリを生成するために、少なくとも 1 0 0 , 0 0 0 の非同一のポリヌクレオチドを増幅する工程であって、ライブラリのポリヌクレオチドの少なくとも約 7 5 % は、少なくとも 1 0 0 , 0 0 0 の非同一のポリヌクレオチドに関してあらかじめ決定された配列と比較してエラーがない工程、を含む。さらに本明細書で提供される方法において、ポリヌクレオチドライブラリはアンプリコンライブラリであり、複数の非同一のポリヌクレオチドの少なくとも約 8 0 % は各々、ライブラリの非同一のポリヌクレオチドに関して平均頻度の 2 x 以内の量でアンプリコンライブラリに存在する。さらに本明細書で提供される方法において、各非同一のポリヌクレオチドは 1 つ以上のエクソンをコードする。さらに本明細書で提供される方法において、増幅された非同一のポリヌクレオチドの各々は、少なくとも 1 つの分子タグを含む。

10

【 0 1 6 6 】

本明細書で提供されるのは、ポリヌクレオチドライブラリの合成のための方法であって、該方法は：(a) 少なくとも 2 , 0 0 0 の非同一のポリヌクレオチドの第 1 のライブラリを増幅する工程；(b) 1 つ以上の配列特徴の関数として第 1 のライブラリ内の配列の分布を特定する工程；および(c) 第 2 のライブラリを生成するために、配列の分布に基づいて第 1 のライブラリにおいて配列の相対比率を変更する工程であって、それによって第 2 のライブラリの 2 . 5 x 以下のサンプリングは少なくとも 8 0 % のカバー率をもたらす工程、を含む。さらに本明細書で提供される方法において、1 つ以上の配列特徴はパーセント GC 含量を含む。さらに本明細書で提供される方法において、1 つ以上の配列特徴はパーセント反復配列含量を含む。さらに本明細書で提供される方法において、1 つ以上の配列特徴はパーセント二次構造含量を含む。さらに本明細書で提供される方法において、1 つ以上の配列特徴はシーケンシングカバー率を含む。さらに本明細書で提供される方法において、1 . 7 x 以下のサンプリングは、少なくとも 8 0 % のシーケンシングカバー率をもたらす。さらに本明細書で提供される方法において、2 . 5 x 以下のサンプリングは、少なくとも 9 0 % のシーケンシングカバー率をもたらす。さらに本明細書で提供される方法において、該方法は第 2 のライブラリを合成する工程をさらに含む。さらに本明細書で提供される方法において、該方法は第 2 のライブラリを増幅する工程をさらに含む。さらに本明細書で提供される方法において、ライブラリは少なくとも 5 , 0 0 0 のポリヌクレオチドを含む。さらに本明細書で提供される方法において、ライブラリは少なくとも 1 0 , 0 0 0 のポリヌクレオチドを含む。さらに本明細書で提供される方法において、ライブラリは少なくとも 3 0 , 0 0 0 のポリヌクレオチドを含む。

20

30

40

【 0 1 6 7 】

本明細書で提供されるのは、標的濃縮のための方法であって、該方法は：少なくとも 2 , 0 0 0 の非同一の二本鎖ポリヌクレオチドのライブラリを、標的核酸を含むサンプルポリヌクレオチドの集団に接触させる工程であって、少なくとも 2 , 0 0 0 の非同一のポリヌクレオチドの各々は、(5 ' ~ 3 ' まで) 第 1 の非標的配列と第 2 の非標的配列、および 1 つ以上の標的核酸配列に相補的な挿入配列を含む、工程；個体担体上の少なくとも 2 , 0 0 0 の非同一のポリヌクレオチドの 1 つ以上にハイブリダイズする標的核酸配列を捕捉する工程；および、標的ポリヌクレオチドライブラリを生成するために、捕捉された標的核酸を放出する工程、を含む。さらに本明細書で提供される方法において、各ポリヌクレオチドは、少なくとも 1 つの分子タグをさらに含む。さらに本明細書で提供される方法

50

において、各非同一の配列はプライマー結合部位をさらに含む。さらに本明細書で提供される方法において、第1の非標的配列はポリヌクレオチドの5'末端に位置し、および第2の非標的配列はポリヌクレオチドの3'末端に位置する。さらに本明細書で提供される方法において、1つ以上の分子タグはポリヌクレオチドの5'末端に付けられている。さらに本明細書で提供される方法において、1つ以上の分子タグはポリヌクレオチドの3'末端に付けられている。さらに本明細書で提供される方法において、1つ以上の分子タグとポリヌクレオチドはスパーサーによって結合される。さらに本明細書で提供される方法において、挿入配列は少なくとも1つのエクソンに相補的である。さらに本明細書で提供される方法において、1つ以上の分子タグはビオチン、葉酸、ポリヒスチジン、FLAGタグ、またはグルタチオンである。さらに本明細書で提供される方法において、1つ以上の分子タグは2つのビオチン分子である。さらに本明細書で提供される方法において、個体担体は磁気ビーズである。さらに本明細書で提供される方法において、第1の非標的配列と第2の非標的配列は、20~40の塩基の長さである。さらに本明細書で提供される方法において、挿入配列は90~200の塩基の長さである。さらに本明細書で提供される方法において、ライブラリは少なくとも5,000のポリヌクレオチドを含む。さらに本明細書で提供される方法において、ライブラリは少なくとも10,000のポリヌクレオチドを含む。さらに本明細書で提供される方法において、ライブラリは少なくとも30,000のポリヌクレオチドを含む。

10

20

30

40

50

【0168】

本明細書において、複数の部分的に相補的な二本鎖ポリヌクレオチドを含むプローブライブラリが提供され、各々は：第1の非標的配列と第2の非標的配列、および1つ以上の標的核酸配列に相補的な第1の挿入配列を含む、第1のポリヌクレオチド：第1の非標的配列と第2の非標的配列、および第1の挿入配列に相補的な第2の挿入配列を含む、第2のポリヌクレオチド：を含み、第1のポリヌクレオチドと第2のポリヌクレオチドは部分的にハイブリダイズされる。さらに本明細書で提供されるライブラリにおいて、二本鎖ポリヌクレオチドの各鎖は、少なくとも2つの分子タグをさらに含む。さらに本明細書で提供されるライブラリにおいて、第1の非標的配列と第2の非標的配列は相補的ではない。さらに本明細書で提供されるライブラリにおいて、第1の非標的配列はポリヌクレオチドの5'末端に位置し、および第2の非標的配列はポリヌクレオチドの3'末端に位置する。さらに本明細書で提供されるライブラリにおいて、1つ以上の分子タグはポリヌクレオチドの5'末端に付けられている。さらに本明細書で提供されるライブラリにおいて、1つ以上の分子タグはポリヌクレオチドの3'末端に付けられている。さらに本明細書で提供されるライブラリにおいて、1つ以上の分子タグとポリヌクレオチドはスパーサーによって結合される。さらに本明細書で提供されるライブラリにおいて、挿入配列は少なくとも1つのエクソンに相補的である。さらに本明細書で提供されるライブラリにおいて、1つ以上の分子タグはビオチン、葉酸、ポリヒスチジン、FLAGタグ、またはグルタチオンである。さらに本明細書で提供されるライブラリにおいて、1つ以上の分子タグは2つのビオチン分子である。さらに本明細書で提供されるライブラリにおいて、個体担体は磁気ビーズである。さらに本明細書で提供されるライブラリにおいて、第1の非標的配列と第2の非標的配列は、20~40の塩基の長さである。さらに本明細書で提供されるライブラリにおいて、挿入配列は90~200の塩基の長さである。

【0169】

本明細書で提供されるのは、プローブライブラリを設計するための方法であって、該方法は：標的配列のライブラリを得る工程；および、標的配列に相補的な挿入配列のライブラリを設計する工程を含み、設計する工程は：標的配列が挿入配列より短い長さであれば、標的配列に相補的な挿入配列を生成する工程；標的配列が挿入配列+Xより短い長さであれば、標的配列に少なくとも部分的に相補的な挿入配列を生成する工程；または、標的配列が挿入配列+Xより長ければ、一般的な標的配列に少なくとも部分的に相補的な1セットの挿入配列を生成工程であって、ここでXは挿入配列によって標的とされない連続塩基の数である、工程；挿入配列のライブラリを生成するためにライブラリにおいて各標的

配列に対して工程 (b) を繰り返す工程、を含む。さらに本明細書で提供される方法において、X は 30 未満のヌクレオチドである。さらに本明細書で提供される方法において、X は 10 未満のヌクレオチドである。さらに本明細書で提供される方法において、X は約 6 のヌクレオチドである。

【0170】

本明細書において、次世代シーケンシングのための方法が提供され、該方法は、複数の標的ポリヌクレオチドを含むサンプルに本明細書に記載されるライブラリを接触させる工程；ライブラリに結合する少なくとも 1 つの標的ポリヌクレオチドを濃縮する工程；および少なくとも 1 つの濃縮された標的ポリヌクレオチドを配列決定する工程、を含む。

【0171】

本明細書で提供されるのは、次世代シーケンシングのための方法であって、該方法は：複数のポリヌクレオチドを含むサンプルに、本明細書に記載のライブラリを接触させる工程；ライブラリに結合しない少なくとも 1 つのポリヌクレオチドから、ライブラリに結合するサンプル中の少なくとも 1 つのポリヌクレオチドを分離する工程；および、ライブラリに結合しない少なくとも 1 つのポリヌクレオチドを配列決定する工程、を含む。

【実施例】

【0172】

以下の実施例は、本発明の様々な実施形態を例示する目的で与えられ、いかなる方法でも本発明を制限するようには意図されていない。実施例は、本明細書に記載される方法と共に、好ましい実施形態を現時点で代表するものであり、例示的なものであり、および本発明の範囲を限定するものとしては意図されていない。そこでの変更、および請求項の範囲により定義される本発明の精神内に包含される他の使用が、当業者に想到されるであろう。

【0173】

実施例 1：基質面の官能化

【0174】

ポリヌクレオチドのライブラリのアタッチメントと合成を支持するために、基質を官能化した。基質面をまず、90%の H_2SO_4 と10%の H_2O_2 を含むピラニア溶液を使用して20分間、湿式洗浄（ウェットクリーニング）した。DI水を用いて基質をいくつかのビーカー内ですすぎ、5分間、DI水のゲースネック型の栓下で保持し、そして N_2 で乾燥させた。基質をその後5分間、 NH_4OH （1：100；3mL：300mL）に浸し、ピストルを使用してDI水ですすぎ、DI水を用いてこれらの3つの連続したビーカーにそれぞれ1分間浸し、次に再びピストルを使用してDI水ですすいだ。次に、基質面を O_2 にさらすことによってデバイスをプラズマ洗浄した。SAMCO PC-300器具を使用して、下流モードで1分間、250ワットで O_2 のプラズマエッチングを行った。

【0175】

浄化した基質面を、以下のパラメータで、YES-1224P蒸着オープンシステムを使用して、N-（3-トリエトキシシリルプロピル）-4-ヒドロキシブチルアミドを含む溶液で能動的に官能化した：0.5～1トール、60分間、70℃、135の気化器。基質面を、Brewer Science 200Xスピンコーターを使用してレジスト塗布した。SPR（商標）3612フォトレジストを、40秒間、2500rpmで基質上にスピンコーティングした。Brewerホットプレート上で30分間、90℃で基質をあらかじめ焼いた。基質を、Karl Suss MA6マスクアライナー器を使用して、フォトリソグラフィにさらした。基質を2.2秒間、露出させ、MSF 26Aで1分間、現像した。残りの現像液をピストルですすぎ、基質を5分間、水に浸した。基質をオープンで30分間、100℃で焼き、続いてNikon L200を使用してリソグラフィの欠損を目視検査した。250ワットで1分間、 O_2 プラズマエッチングを行うためのSAMCO PC-300器具を使用して残りのレジストを取り出すためにデスカム処理を使用した。

【0176】

基質面を、10 μ Lの軽油と混合した100 μ Lのペルフルオロオクチルトリクロロシラン溶液を用いて受動的に官能基化した。基質をチャンバに置き、10分間、ポンプでくみ上げ、その後にバルブでポンプを止め、10分間立たせたままにした。チャンバを空気に通気させた。最大出力（クレストシステム上で9）の超音波処理を用いて、基質を70で5分間、500 mLのNMPに2回漬けることによって、レジストストリップを行った。次に、最大出力の超音波処理を用いて、基質を室温で5分間、500 mLのイソプロパノールに漬けた。基質を300 mLの200プルーフエタノールに漬け、N₂で吹き付けて乾かした。官能化された面を、ポリヌクレオチド合成の支持の役割を果たすように活性化させた。

10

【0177】

実施例2：ポリヌクレオチド合成デバイスにおける50量体配列の合成

【0178】

2次元のポリヌクレオチド合成デバイスをフローセルに組み立て、それをフローセル（Applied Biosystems（ABI394 DNA Synthesizer））に接続した。2次元のポリヌクレオチド合成デバイスを、N-（3-トリエトキシシリルプロピル）-4-ヒドロキシブチルアミド（Gelest）で均一に官能化し、本明細書に記載されるポリヌクレオチド合成方法を使用して50bp（「50量体ポリヌクレオチド」）の典型的なポリヌクレオチドを合成するために使用した。

【0179】

20

50量体配列は、SEQ ID NO. : 1、5' AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTTCGGCAT##TTTTTTT3'（SEQ ID NO. : 1）として記される通りであり、ここで#は、チミジン-スクシニルヘキサナミドCEDホスホラミダイト（ChemGenesからのCLP-2244）を意味し、これは脱保護中に面からのポリヌクレオチドの放出を可能にする切断リンカーである。

【0180】

表1のプロトコルに従い、およびABIABIシンセサイザで、標準的なDNA合成化学（カップリング、キャッピング、酸化および分解）を使用して、合成を行った。

【0181】

30

【表 1 - 1】

表 1		
一般的な DNA 合成プロセス名	処理工程	時間 (秒)
洗浄 (アセトニトリル洗浄フロー)	アセトニトリルシステムフラッシュ	4
	アセトニトリルからフローセル	23
	N2 システムフラッシュ	4
	アセトニトリルシステムフラッシュ	4
DNA 塩基の付加 (ホスホラミダイト+ 活性化因子フロー)	活性化因子マニホールドフラッシュ	2
	活性化因子からフローセル	6
	活性化因子+ホスホラミダイトからフローセル	6
	アセトニトリルからフローセル	0.5
	活性化因子+ホスホラミダイトからフローセル	5
	活性化因子からフローセル	0.5
	活性化因子+ホスホラミダイトからフローセル	5
	活性化因子からフローセル	0.5
	活性化因子+ホスホラミダイトからフローセル	5
	25 秒間インキュベート	25
洗浄 (アセトニトリル洗浄フロー)	アセトニトリルシステムフラッシュ	4
	アセトニトリルからフローセル	15
	N2 システムフラッシュ	4
	アセトニトリルシステムフラッシュ	4
DNA 塩基の付加 (ホスホラミダイト+ 活性化因子フロー)	活性化因子マニホールドフラッシュ	2
	活性化因子からフローセル	5
	活性化因子+ホスホラミダイトからフローセル	18
	25 秒間インキュベート	25
洗浄 (アセトニトリル洗浄フロー)	アセトニトリルシステムフラッシュ	4
	アセトニトリルからフローセル	15
	N2 システムフラッシュ	4
	アセトニトリルシステムフラッシュ	4

【表 1 - 2】

表 1		
一般的な DNA 合成プロセス名	処理工程	時間 (秒)
キャッピング (CapA+B, 1:1, フロー)	CapA+B からフローセル	15
洗浄 (アセトニトリル洗浄 フロー)	アセトニトリルシステムフ ラッシュ	4
	アセトニトリルからフロー セル	15
	アセトニトリルシステムフ ラッシュ	4
酸化 (酸化剤フロー)	酸化剤からフローセル	18
洗浄 (アセトニトリル洗浄 フロー)	アセトニトリルシステムフ ラッシュ	4
	N2 システムフラッシュ	4
	アセトニトリルシステムフ ラッシュ	4
	アセトニトリルからフロー セル	15
	アセトニトリルシステムフ ラッシュ	4
	アセトニトリルからフロー セル	15
	N2 システムフラッシュ	4
	アセトニトリルシステムフ ラッシュ	4
	アセトニトリルからフロー セル	23
	N2 システムフラッシュ	4
	アセトニトリルシステムフ ラッシュ	4
	デブロッキング (デブロッ クフロー)	36
洗浄 (アセトニトリル洗浄 フロー)	アセトニトリルシステムフ ラッシュ	4
	N2 システムフラッシュ	4
	アセトニトリルシステムフ ラッシュ	4
	アセトニトリルからフロー セル	18
	N2 システムフラッシュ	4.13
	アセトニトリルシステムフ ラッシュ	4.13
	アセトニトリルからフロー セル	15

ホスホラミダイト / 活性化因子の組み合わせを、フローセルを通る大量の試薬の送達と同様に送達した。環境が試薬でずっと「湿った」ままであるため、乾燥工程は行わなかった。

【0184】

より速い流れを可能にするために、流れレストリクタをABI 394シンセサイザから取り除いた。流れレストリクタなしで、アミダイト (ACN中に0.1M)、活性剤 (ACN中に0.25Mのベンゾイルチオテトラゾール (「BT T」; Glen Researchからの30-3070-xx))、およびOX (20%のピリジン10%の水、および70%のTHF中に0.02M I2)に関する流速は、およそ~100 uL / 秒であり、アセトニトリル (「ACN」) およびキャッピング試薬 (1:1のCap AとCap Bの混合物、Cap AはTHF / ピリジン中の無水酢酸であり、Cap BはTHF中の16%の1-メチルイミジゾール (methylimidazole) である) に関してはおよそ~200 uL / 秒であり、およびDeblock (トルエン中の3%のジクロロ酢酸) に関してはおよそ~300 uL / 秒であった (流れレストリクタがある場合のすべての試薬の~50 uL / 秒と比較)。酸化剤が完全に押し出されるまでの時間を観察し、化学物質の流れ時間のタイミングを適宜、調整し、および余分なACN洗浄を異なる化学物質間に導入した。ポリヌクレオチド合成後、75 psiで一晩、チップをガス状アンモニアで脱保護した。ポリヌクレオチドを回復させるために水を5滴、面に適用した。次に、回復させたポリヌクレオチドを、BioAnalyzer小型RNAチップで分析した (データは示さず)。

10

20

【0185】

実施例3: ポリヌクレオチド合成デバイスにおける100量体配列の合成

【0186】

50量体配列の合成に関して実施例2に記載されたのと同じプロセスを使用して、100量体のオリゴヌクレオチド ("100-mer polynucleotide" ; 5' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTGTGCTGTCCACCAAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAAACCCCGCAT##TTTTTTTTTTT3'、#はチミン-スクシニルヘキサナミドCEDホスホラミダイト (ChemGenesからのCLP-2244) を意味する; SEQ ID NO.: 2) を、2つの異なるシリコンチップ上で合成し、1つ目は、N-(3-トリエトキシシリルプロピル)-4-ヒドロキシブチルアミドで均一に官能化し、2つ目は11-アセトキシウンデシルトリエトキシシランとn-デシルトリエトキシシランの5/95の混合物で官能化し、そして面から抽出されたポリヌクレオチドをBioAnalyzer器具で分析した (データは示さず)。

30

【0187】

2つのチップからの10のサンプルはすべて、以下の熱サイクルプログラムを用いて、順方向プライマー (5' ATGCGGGGTTCTCATCATC3' ; SEQ ID NO.: 3) と逆方向プライマー (5' CGGGATCCTTATCGTCATCG3' ; SEQ ID NO.: 4) を使用し、50 uLのPCR混合物中 (25 uLのNEB Q5マスターミックス、2.5 uL 10 uMの順方向プライマー、2.5 uL 10 uMの逆方向プライマー、面から抽出された1 uLのポリヌクレオチド、および最大50 uLの水) で、さらにPCR増幅された:

40

98 C、30秒

98 C、10秒; 63 C、10秒; 72 C、10秒; 12サイクルを繰り返す

72 C、2分

【0188】

PCR生成物もまた、BioAnalyzerで分析し (データは示さず)、100量体位置で鋭いピークを実証した。次に、PCR増幅サンプルをクローニングし、およびサンガーシーケンシングを行った。表2は、チップ1からのスポット1-5から採取されたサンプル、およびチップ2からのスポット6-10から採取されたサンプルに関する、サ

50

ンガーシーケンシングの結果を概説する。

【 0 1 8 9 】

【 表 2 】

スポット	エラー率	サイクル効率
1	1/763 bp	99.87%
2	1/824 bp	99.88%
3	1/780 bp	99.87%
4	1/429 bp	99.77%
5	1/1525 bp	99.93%
6	1/1615 bp	99.94%
7	1/531 bp	99.81%
8	1/1769 bp	99.94%
9	1/854 bp	99.88%
10	1/1451 bp	99.93%

10

20

【 0 1 9 0 】

したがって、合成ポリヌクレオチドの高い品質と均一性は、異なる面化学作用を用いて2つのチップ上で繰り返された。全体として、配列された262の100量体のうち233に対応する89%が、エラーのない完全な配列であった。

【 0 1 9 1 】

最後に表3は、スポット1 - 10からのポリヌクレオチドサンプルから得られた配列に関するエラー特性を概説する。

【 0 1 9 2 】

【表 3】

表3

サンプル ID/スポット no.	OSA_0 046/1	OSA_0 047/2	OSA_0 048/3	OSA_0 049/4	OSA_0 050/5	OSA_0 051/6	OSA_0 052/7	OSA_0 053/8	OSA_0 054/9	OSA_00 55/10
全配列数	32	32	32	32	32	32	32	32	32	32
シーケンシ ングの質	25/28	27/27	26/30	21/23	25/26	29/30	27/31	29/31	28/29	25/28
オリゴの質	23/25	25/27	22/26	18/21	24/25	25/29	22/27	28/29	26/28	20/25
ROI マッチ数	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI 突然変異	2	2	1	3	1	0	2	1	2	1
ROI複数 塩基欠失	0	0	0	0	0	0	0	0	0	0
ROI小 挿入	1	0	0	0	0	0	0	0	0	0
ROI一塩 基欠失	0	0	0	0	0	0	0	0	0	0
大きな欠失 の数	0	0	1	0	0	1	1	0	0	0
突然変異: G>A	2	2	1	2	1	0	2	1	2	1
突然変異: T>C	0	0	0	1	0	0	0	0	0	0
ROIエラ ー数	3	2	2	3	1	1	3	1	2	1
ROIエラ ー率	Err: ~1 (834 の 内)	Err: ~1 (1350 の 内)	Err: ~1 (1282 の 内)	Err: ~1 (708 の 内)	Err: ~1 (2500 の 内)	Err: ~1 (2667 の 内)	Err: ~1 (876 の 内)	Err: ~1 (2900 の 内)	Err: ~1 (1400 の 内)	Err: ~1 (2349 の 内)
ROIマイ ナス プラ イマー エ ラー率	MP Err: ~1(763 の内)	MP Err: ~1(824 の内)	MP Err: ~1(780 の内)	MP Err: ~1(429 の内)	MP Err: ~1(1525 の内)	MP Err: ~1(1615 の内)	MP Err: ~1(531 の内)	MP Err: ~1(1769 の内)	MP Err: ~1(854 の内)	MP Err: ~1(1451 の内)

10

20

30

【 0 1 9 3 】

実施例 4 : 2 9 , 0 4 0 の特有のポリヌクレオチドの平行組立て

【 0 1 9 4 】

図 1 0 に示されるように、平らなシリコンプレート (1 0 0 1) 上の 1 2 1 の遺伝子座を各々含む 2 5 6 のクラスタ (1 0 0 5) を有する構造を製造した。クラスタの拡大図は、1 2 1 の遺伝子座を伴い (1 0 1 0) で示される。2 5 6 のクラスタのうち 2 4 0 個からの遺伝子座は、別個の配列を有するポリヌクレオチドの合成のためのアタッチメントと支持を提供した。ポリヌクレオチド合成は、実施例 3 からの一般的方法を使用して、ホスホラミダイトの化学作用によって行なわれた。2 5 6 のクラスタのうち 1 6 個からの遺伝子座は、対照クラスタであった。合成された 2 9 , 0 4 0 の特有のポリヌクレオチドの全体的な分布 (2 4 0 × 1 2 1) は、図 1 1 A で示される。ポリヌクレオチドライブラリは高い均一性で合成された。配列の 9 0 % が、平均の 4 X 以内のシグナルに存在し、1 0 0 % の表示を可能にした。図 1 1 B で示されるように、分布はクラスタごとに測定された。4 つの代表的なクラスタで合成された特有のポリヌクレオチドの分布は図 1 2 に示される。全体的なレベルに、実行された全てのポリヌクレオチドが存在し、およびポリヌクレオチドの 9 9 % は、合成の均一性を示す平均の 2 x 以内である豊富量を有していた。この同じ所見は、クラスタごとのレベルで一貫していた。

40

50

【 0 1 9 5 】

各ポリヌクレオチドのエラー率を、I l l u m i n a M i S e q 遺伝子シーケンサーを使用して判定した。29,040の特有のポリヌクレオチドに関するエラー率分布が図13Aに示され、平均は500分の1塩基あたりであり、800分の1塩基と同等に低いいくつかのエラー率が存在する。図13Bで示されるように、分布はクラスタごとに測定された。4つの代表的なクラスタにおける特有のポリヌクレオチドのエラー率分布は、図14に示される。29,040の特有のポリヌクレオチドのライブラリは20時間以内に合成された。

【 0 1 9 6 】

G C の割合 対 29,040の特有のポリヌクレオチドのすべてにわたるポリヌクレオチド表示の分析は、合成がG C 含量に関わらず均一であったことを示した。図15。

【 0 1 9 7 】

実施例5：合成されたポリヌクレオチドライブラリのP C R 増幅

【 0 1 9 8 】

9,996のポリヌクレオチド、G C 含量の変動する無作為化された配列の各100塩基の長さ、20～80%のG C 含量を設計し、実施例3に記載の類似する構成を有する構造上で合成した。G C 表示に対するP C R 増幅の効果を判定するために、ポリヌクレオチド集団を、高忠実度のD N A ポリメラーゼ（D N A ポリメラーゼ1）を用いて6または20のサイクルのいずれかで増幅した。代替的に、ポリメラーゼ選択が増幅後の全体的な配列表示に効果を有するかどうかを判定するために、2つの他の高忠実度のP C R 酵素を使用し、6、8、10、または15サイクルの間、ポリヌクレオチド集団を増幅した。P C R 増幅後、サンプルは、次世代シーケンシングのために準備され、I l l u m i n a M i S e q プラットフォーム上で配列決定された。150のb p S E の読み取りが、100Xのおよその読み取り深度へと生成された。生のF A S T Q ファイルが分析された。6、10、または15サイクルのいずれかのポリメラーゼによるポリヌクレオチド表示は、図16に表される。ポリヌクレオチド表示の均一性が、様々な疾病に関して評価され、表4に概説される。

【 0 1 9 9 】

【 表 4 】

表4

	サイクル	1.5x 以内の%	2x 以内の%
ポリメラーゼ 1	6	72.1%	92.6%
	8	76.1%	90.3%
	10	70.9%	86.6%
	15	64.1%	82.7%
ポリメラーゼ 2	6	91.9%	98.9%
	8	89.9%	98.1%
	10	90.1%	98.4%
	15	89.2%	97.9%

【 0 2 0 0 】

増幅された各ポリヌクレオチド集団に関するドロップアウトの数を、図15に示すよう

に定量化し、増幅サイクル 対 集団のフラクションは平均閾値の 10 %を下回った。ポリメラーゼ 1 ドロップアウトが素早く成長したのに対し、ポリメラーゼ 2 ドロップアウトは比較的一定のままであった。

【0201】

GC 分布に対する過剰増幅の影響を評価した。図 18。一般的には、GC 含量 30 % ~ 70 % のポリヌクレオチドは、近似曲線、 $Y = X$ に従い、およびより多くのサイクルで頻度が増加した。70 % より多くの GC 含量を有するポリヌクレオチドが、概して 20 サイクル後にわずかに頻度が多かったのに対し、30 % 未満の GC 含量のポリヌクレオチドは、概して、6 サイクル後にわずかに頻度が多かった。

【0202】

実施例 6：全プレート増幅からのポリヌクレオチド表示と、平行なポリヌクレオチドクラスタ増幅との比較

【0203】

図 10 に示されるように、平らなシリコンプレート上の 121 の遺伝子座を各々含む 256 のクラスタを有する構造上で、ポリヌクレオチドが合成された。ポリヌクレオチド合成は、実施例 3 からの一般的方法を使用して、ホスホラミダイトの化学作用によって行なわれた。構造上のポリヌクレオチドを切断して組み合わせた。

【0204】

プレートにわたってポリヌクレオチドを組み合わせて増幅した。増幅後に、図 19 に見られるような線で表される平均値からの顕著な GC のバイアスと分散があった。その結果、より多くのシーケンシングが必要であり、およびより多くのドロップアウトがあった。

【0205】

クラスタの増幅からのポリヌクレオチドの分布は図 20 に見られる。実行 1 と実行 2 では、平均値（線）からの度数分布は約 8 であり、および平均値からの分散は約 $1.7 \times$ であった。GC の割合は 17 % から 94 % の範囲であった。図 23 と図 20 は、再現性があることを例証し、およびポリヌクレオチド集団は、GC バイアスの劇的な減少を示す（図 20）。加えて、0 のドロップアウトがあり、および 30 % 少ないシーケンシングが求められた。

【0206】

実施例 7：種々の GC 含量で合成されたポリヌクレオチドライブラリ

【0207】

約 15 % から約 85 % の GC 含量を含む 13,000 のポリヌクレオチド配列のライブラリを、合成のためにあらかじめ選択した（図 22）。第 1 のポリヌクレオチドライブラリは構造上で合成され、および合成は、実施例 3 からの一般的方法を使用してホスホラミダイト化学作用によって行なわれた。構造上のポリヌクレオチドは、ポリヌクレオチドの PCR バイアスのかかったライブラリを生成するために、増幅後に切断されて組み合わされた。ライブラリ内のポリヌクレオチド配列は、GC 含量に従ってビンに入れられ、各ビンの化学量論は、PCR 増幅によって生成された、観察された GC バイアスを考慮して適合された。例えば、より高い、またはより低い GC 含量を含むポリヌクレオチドは、増幅後に化学量論の表示を均一にするより高い初期濃度を有する。これは、増幅工程からの PCR GC のバイアスを効果的に減らし、または除去する。第 2 のポリヌクレオチドライブラリは構造上で合成され、および合成は、実施例 3 からの一般的方法を使用してホスホラミダイト化学作用によって行なわれた。構造上のポリヌクレオチドは、増幅後に均一の GC の表示で（図 21A）、ポリヌクレオチドの高度に均一なライブラリを生成するために（図 23）、切断されて組み合わされ、その後に増幅された。GC の平衡なライブラリの 1 つの利点は、望ましいサンプリングカバー率のためにより少ないサンプリングが求められることである。例えば、ライブラリの 80 % と 90 % のカバー率のためのサンプリングは、単分散ライブラリのための理論上の最小に近づいた（図 24A と図 24B）。度合の変動する高いおよび低い GC 含量の両方を好むポリヌクレオチドライブラリをさらに合成した（それぞれ図 21B と図 21C）。低い GC 含量（図 21D）または高い GC 含量

10

20

30

40

50

(図21E)を好むポリヌクレオチドライブラリをさらに合成した。

【0208】

実施例8：80量体および120量体のポリヌクレオチド長さで合成された、GCの平衡なポリヌクレオチドライブラリ

【0209】

各々が80の核酸長さである、およそ20,000の特有のポリヌクレオチドを含むライブラリを設計し、実施例7の一般的方法を使用してGCを平衡にし、そして構造上で合成した；合成は、実施例3からの一般的方法を使用して、ホスホラミダイトの化学作用によって行なわれた。各々が120の核酸の長さであり、ポリヌクレオチドを含む類似のライブラリをさらに合成した(図26)。両ライブラリは高度に均一な分布を示し、>99%のユニーク配列が識別された。ライブラリはまた、GC含量の分散を通じて均一性を示し、複製を通じて高い一致を有し(図27)、雑音にさらされる末尾に少数のポリヌクレオチドを有した。

10

【0210】

実施例9：ポリヌクレオチドライブラリの繰り返しの増幅後のGC含量評価

【0211】

20-80%のGC含量を含む9,996の特有のポリヌクレオチドから成るポリヌクレオチドライブラリは、各々100塩基の長さであり、実施例3からの一般的方法を使用して、ホスホラミダイト化学作用によって構造上で合成された。ライブラリを、2つの異なる高忠実度のDNAポリメラーゼを用いて8または15PCRサイクルのいずれかで増幅し、集団におけるポリヌクレオチドの頻度を、これらの2つの条件間で比較した(図28)。同定線(identity line)(黒色破線)は、8または15PCRサイクルのいずれか後の、集団内の同じ頻度のポリヌクレオチドを示す。8サイクルの増幅と比較して、同定線上の配列は、15サイクル後の集団での過剰表示であり、線の下に配列は、15サイクル後の集団での提示不足である。この場合、ポリメラーゼ1は、増加するPCRサイクルと共にGCバイアスを示す。8サイクルの増幅と比較して、15サイクルの増幅後、高GCの配列(70%を超えるGC、中くらいの灰色)については過剰表示が観察され、低GCの配列(30%未満のGC、最も暗い灰色)は提示不足であった。加えて、同様のGC割合内での濃縮における大きな変動は、GC含量以外の因子、ヘアピン形成またはホモポリマー伸張等が、増幅バイアスに影響を及ぼし得ることを示唆する。ポリメラーゼ2は同じ配列表示バイアスを示さなかった。

20

30

【0212】

実施例10：ポリヌクレオチドライブラリの繰り返しの増幅後のドロップアウトと表示評価

【0213】

種々のDNAポリメラーゼ酵素での増幅によって導入されたバイアスを調査した。実施例9のポリヌクレオチドライブラリを、6、8、10、または15サイクル間、DNAポリメラーゼ1(図29、暗い線)またはDNAポリメラーゼ2(図29、明るい線)のいずれかで増幅した。PCRサイクルの増加は、ポリヌクレオチド配列のドロップアウト頻度と相関していた。ドロップアウト頻度は、平均の10%未満の豊富量を有する配列として定義される。この効果の範囲は、増幅に使用されたDNAポリメラーゼに依存した。配列のより大きな部分は、DNAポリメラーゼ2での増幅と比較して、DNAポリメラーゼ1での増幅後にドロップアウトした(15サイクルでのドロップアウトのおよそ20倍多い)。種々のポリメラーゼは、GC含量、長さ、および配列の複雑性に依拠して、異なるライブラリ配列の増幅のために最適である場合もある。

40

【0214】

各DNAポリメラーゼを用いて15PCRサイクル増幅されたライブラリを、ポリヌクレオチド配列の表示を評価するためにより詳細に研究した(図30)。DNAポリメラーゼ1を用いたポリヌクレオチドライブラリの増幅は、15PCRサイクル度に20倍より多くの配列ドロップアウトをもたらした(図29)。

50

DNAポリメラーゼ1を用いて増幅されたポリヌクレオチドの分布は、DNAポリメラーゼ2で増幅されたポリヌクレオチドの分布より大きかった。

DNAポリメラーゼ1で増幅されたライブラリのポリヌクレオチド分布は、平均値の1.5倍以内に存在する配列の64%を有した。

同じライブラリをDNAポリメラーゼ2で増幅した場合、>89%の配列が平均の1.5倍以内に存在し、DNAポリメラーゼ2で増幅されたライブラリがDNAポリメラーゼ1よりもはるかに低いバイアスを有することを示した。

DNAポリメラーゼ1を用いて増幅されたライブラリに導入されたバイアスは、ポリヌクレオチドライブラリをカバーするのに必要なスクリーニング成果を高める。

【0215】

10

実施例11：次世代シーケンシング(NGS)を用いた、エクソームを標的とするための制御された化学量論ポリヌクレオチドライブラリの使用

【0216】

1つ以上の遺伝子エクソンとオーバーラップする最大370,000以上の非同一のポリヌクレオチドを含むライブラリ(プローブライブラリ)を標的とする第1のポリヌクレオチドcDNAは、実施例3からの一般的方法を使用して、ホスホラミダイト化学作用によって構造上で設計され、合成される。

ポリヌクレオチドは、対象の標的エクソンの続く捕捉のためのプローブを形成するために、PCRを使用して(または個体フェーズ合成中に直接)、ビオチン等の分子タグに結合される。

20

プローブをゲノム核酸のライブラリの配列にハイブリダイズし、未結合の配列から分離する。

未結合のプローブを洗い流して、cDNA配列で濃縮された標的ライブラリを残した。

次に、濃縮されたライブラリを、NGSを使用して配列し、および各予測される遺伝子に関する読み取りを、遺伝子を標的とするのに使用されるcDNAプローブの関数として測定する。

【0217】

いくつかの例では、標的配列の読み取り頻度は、標的配列の豊富量、プローブ結合、二次構造、または濃縮に関わらず標的配列のシーケンシング後に表示を低下させる他の因子によって影響を受ける。ポリヌクレオチドライブラリの化学量論的制御は、第2のポリヌクレオチドcDNA標的ライブラリを得るために第1のポリヌクレオチドcDNA標的ライブラリの化学量論の修正によって行われ、より少ない読み取りに帰結するポリヌクレオチドプローブ配列に関する化学量論を増加させる。この第2のcDNA標的ライブラリは、実施例3からの一般的方法を使用し、ホスホラミダイト化学作用によって構造上で設計されて合成され、そして前記のように標的ゲノムDNAライブラリの配列エクソンを濃縮するために使用される。

30

【0218】

実施例12：エクソームプローブライブラリを用いた化学量論制御の多数の反復

【0219】

エクソームプローブライブラリは、実施例11の一般的方法を使用して合成され、試験された。化学量論的な修飾を多数、繰り返し、結果として制御された化学量論プローブライブラリ、ライブラリ1を得た。いくつかのコンパレータエクソーム濃縮キットと比較して、これは、標的の望ましいカバー率を得るために、有意に少ないシーケンシング読み取りをもたらす。正確なシーケンシングのために、標的エクソーム塩基の少なくとも90%の30xの読み取り深度が望ましく、および過剰シーケンシング(理論上の読み取り深度、30xより大きな読み取り深度)がしばしば、均一性の問題を補うために必要とされる。制御された化学量論エクソームプローブライブラリは、55xの理論上の読み取り深度で、標的塩基の90%の30xの読み取り深度を達成することができ(図32A)、これは、他のコンパレータエクソーム濃縮キットによって必要とされるよりも有意に少ないシーケンシングカバー率であり、およびより速い配列スループットであった(1実行あたり

40

50

のサンプル、表 5)。4 . 5 G b のシーケンシングに標準化される場合、制御された化学量論プローブライブラリは 1 0 x の読み取り深度の全標的塩基の > 9 5 % を提供し、これは他のコンパレータエクソームプローブキットより有意に高かった (図 3 2 B)。

【 0 2 2 0 】

【表 5】

表5

必要な平均シーケンシングカバー率	実行あたりのサンプル
コンパレータエクソーム濃縮キットA	4
制御された化学量論エクソームプローブ (ライブラリ 1)	17

10

【 0 2 2 1 】

実施例 1 3 : ハイブリダイゼーションパネルの生成

【 0 2 2 2 】

ポリヌクレオチド標的ライブラリは、特定遺伝子、疾患、パネルの組み合わせ、またはカスタムエクソームを標的とする実施例 1 1 の一般的方法を使用して調製された。反応サイズは 1 0 ³ のスケールに及び、およびプローブパネルサイズは、約 8 0 ~ 約 9 0 0 , 0 0 0 のプローブに及んだ (図 3 3)。

20

【 0 2 2 3 】

実施例 1 4 : 7 0 , 0 0 0 のプローブパネルの生成

【 0 2 2 4 】

7 0 , 0 0 0 の非同一のポリヌクレオチドを含むポリヌクレオチド標的ライブラリ (プローブライブラリ) を、実施例 3 から的一般的方法を使用してホスホラミダイト化学作用によって、およびライブラリ 2 を生成するために実施例 1 1 の一般的方法を使用して制御された G C によって、構造上で設計して合成した。シーケンシング後の読み取り分布は図 3 4 に示され、および G C ビン標的カバー率は、図 3 5 A と図 3 5 B に示される。

30

【 0 2 2 5 】

実施例 1 5 : 2 , 5 4 4 のプローブパネルの生成

【 0 2 2 6 】

2 , 5 4 4 の非同一のポリヌクレオチドを含むポリヌクレオチド標的ライブラリ (プローブライブラリ) を、実施例 3 から的一般的方法を使用してホスホラミダイト化学作用によって、およびライブラリ 3 を生成するために実施例 1 1 の一般的方法を使用して制御された化学量論によって、構造上で設計して合成した。正確な比率は図 3 6 A に示され、およびカバー率は図 3 6 B に示される。ライブラリ 3 での標的濃縮は、コンパレータアレイベースのキット # 2 よりも高い、正確な比率とカバー率の両方をもたらした。

【 0 2 2 7 】

実施例 1 6 : ポリヌクレオチド標的ライブラリを用いたサンプルの調製と濃縮

40

【 0 2 2 8 】

ゲノム DNA (g DNA) をサンプルから得て、切断緩衝液中で酵素によりフラグメント化し、末端修復し、そして 3 ' アデニル化を行う。アダプターは、アダプター標識 g DNA 鎖のライブラリを生成するためにゲノム DNA フラグメントの両端に連結され、およびアダプター標識 DNA ライブラリが高忠実度のポリメラーゼで増幅される。次に g DNA ライブラリを、アダプター遮断薬の存在下において、9 6 で一本鎖に変性させる。ポリヌクレオチド標的ライブラリ (プローブライブラリ) を 9 6 で、ハイブリダイゼーション溶液中で変性させ、ハイブリダイゼーション溶液中で 7 0 で 1 6 時間、変性した標識 g DNA ライブラリと組み合わせる。次に、結合緩衝液をハイブリダイズされた標識 g DNA プローブに添加し、およびストレプトアビジンを含む磁気ビーズを使用してピオチ

50

ン化プローブを捕捉する。磁石を使用してビーズを分離し、そしてビーズを緩衝液で3回洗浄して、溶出緩衝液を加える前に未結合のアダプター、gDNA、およびアダプター遮断薬を除去し、ビーズから濃縮された標識gDNAフラグメントを放出する。クラスタ生成に十分な産出を得るために、標識gDNAフラグメントの濃縮されたライブラリを高忠実度のポリメラーゼで増幅し、次にライブラリを、NGS器具を使用して配列決定する。

【0229】

実施例17：ポリヌクレオチド標的ライブラリを用いた一般的なサンプルの調製と濃縮

【0230】

複数のポリヌクレオチドをサンプルから得て、フラグメント化し、随意に末端修復し、アデニル化する。アダプターは、アダプター標識ポリヌクレオチド鎖のライブラリを生成するために、ポリヌクレオチドフラグメントの両端に連結され、およびアダプター標識ポリヌクレオチドライブラリが増幅される。次に、アダプター標識ポリヌクレオチドライブラリを、アダプター遮断薬の存在下において、高温、好ましくは96℃で変性させる。ポリヌクレオチド標的ライブラリ（プローブライブラリ）を高温で、好ましくは約90～99℃で、ハイブリダイゼーション溶液中で変性させ、約45～80℃で、約10～24時間、ハイブリダイゼーション溶液中で、変性した標識ポリヌクレオチドライブラリと結合させる。次に結合緩衝液をハイブリダイズされたアダプター標識ポリヌクレオチドプローブに添加し、捕捉部分を含む個体担体を、ハイブリダイズされたアダプター標識ポリヌクレオチドプローブを選択的に結合するために使用する。個体担体を1回以上、好ましくは約2～5回、緩衝液で洗浄して、溶出緩衝液を添加する前に未結合のポリヌクレオチドを除去し、個体担体から濃縮されたアダプター標識ポリヌクレオチドフラグメントを放出する。アダプター標識ポリヌクレオチドフラグメントの濃縮されたライブラリが増幅され、その後ライブラリが配列される。

【0231】

実施例18：ポリヌクレオチド標的ライブラリを用いた、標識前の一般的な濃縮

【0232】

複数のポリヌクレオチドをサンプルから得て、フラグメント化し、随意に末端修復する。次に、フラグメント化ポリヌクレオチドサンプルを、高温、好ましくは96℃で変性させる。ポリヌクレオチド標的ライブラリ（プローブライブラリ）を高温で、好ましくは約90～99℃で、ハイブリダイゼーション溶液中で変性させ、約10～24時間、約45～80℃で、ハイブリダイゼーション溶液中で、変性した標識ポリヌクレオチドライブラリと結合させる。次に結合緩衝液をハイブリダイズされたポリヌクレオチドプローブに添加し、捕捉部分を含む個体担体を使用して、ハイブリダイズされたフラグメント化ポリヌクレオチドプローブを選択的に結合する。個体担体を1回以上、好ましくは約2～5回、緩衝液で洗浄して、溶出緩衝液を添加する前に未結合のポリヌクレオチドを除去し、個体担体から濃縮されたポリヌクレオチドフラグメントを放出する。濃縮されたポリヌクレオチドをアデニル化し、アダプター標識ポリヌクレオチド鎖の濃縮されたライブラリを生成するためにアダプターをポリヌクレオチドの両端に連結し、およびアダプター標識ポリヌクレオチドライブラリを増幅する。次に、アダプター標識ポリヌクレオチドフラグメントの濃縮されたライブラリを配列決定する。

【0233】

実施例19：ポリヌクレオチド標的ライブラリを用いた一般的なサンプル調製とフィルタリング

【0234】

複数のポリヌクレオチドをサンプルから得て、フラグメント化し、随意に末端修復し、アデニル化する。アダプター標識ポリヌクレオチド鎖のライブラリを生成するために、アダプターをポリヌクレオチドフラグメントの両端に連結し、およびアダプター標識ポリヌクレオチドライブラリを増幅する。次に、アダプター標識ポリヌクレオチドライブラリを、アダプター遮断薬の存在下において、高温、好ましくは96℃で変性させる。望ましくない非標的配列を除去するために設計されたポリヌクレオチドフィルタリングライブラリ

(プローブライブラリ)を、高温で、好ましくは約90~99 で、ハイブリダイゼーション溶液中で変性させ、約10~24時間、約45~80 で、ハイブリダイゼーション溶液中で、変性した標識ポリヌクレオチドライブラリと結合させる。次に結合緩衝液をハイブリダイズさせたアダプター標識ポリヌクレオチドプローブに添加し、捕捉部分を含む個体担体を、ハイブリダイズさせたアダプター標識ポリヌクレオチドプローブを選択的に結合するために使用する。個体担体は、標的アダプター標識ポリヌクレオチドフラグメントを溶出するために、緩衝液で1回以上、好ましくは約1~5回、洗浄される。標的アダプター標識ポリヌクレオチドフラグメントの濃縮されたライブラリを増幅し、その後ライブラリを配列決定する。

【0235】

実施例20：160量体のプローブライブラリの調製

【0236】

少なくとも1000のプローブを含むライブラリを構造上で合成し、および合成は、実施例3からの一般的方法を使用してホスホラミダイト化学作用によって行なう。各プローブは二本鎖であり、プローブの各鎖は、標的に相補的な120ヌクレオチドの標的結合配列を含む。各プローブは、20ヌクレオチドの順方向プライミング部位と、20ヌクレオチドの逆方向プライミング部位を含む。プローブの各鎖は、2つのビオチン分子で5'位置に標識される。

【0237】

実施例21：非標的結合配列を含む210量体のプローブライブラリの調製

【0238】

少なくとも1000のプローブを含むライブラリを構造上で合成し、および合成は、実施例3からの一般的方法を使用してホスホラミダイト化学作用によって行なう。各プローブは二本鎖であり、プローブの各鎖は、標的に相補的な120のヌクレオチド標的結合配列を含む。各プローブはさらに、20ヌクレオチドの順方向プライマー結合部位、20ヌクレオチドの逆方向プライマー結合部位、25ヌクレオチドの5'非標的結合配列、およびポリアデニンを含む25ヌクレオチドの3'非標的結合配列を含む。プローブの各鎖は、2つのビオチン分子で5'位置に標識される。

【0239】

実施例22：ヒトHLAのエクソン1を標的とする210量体のプローブ

【0240】

少なくとも1000のプローブを含むライブラリを構造上で合成し、および合成は、実施例3からの一般的方法を使用してホスホラミダイト化学作用によって行なう。各プローブは二本鎖であり、プローブの各鎖は、ヒトHLAのエクソン1の領域に相補的な120ヌクレオチドの標的結合配列を含む。各プローブはさらに、20ヌクレオチドの順方向プライマー結合部位、20ヌクレオチドの逆方向プライマー結合部位、25ヌクレオチドの5'非標的結合配列、およびポリアデニンを含む25ヌクレオチドの3'非標的結合配列を含む。プローブの各鎖は、2つのビオチン分子で5'位置に標識される。

【0241】

実施例22：オーバーラップしないプローブライブラリの設計方法

【0242】

少なくとも100の標的配列が提供され、相補的なプローブ標的結合配列の所望の長さと比較し、長さに基づいて別個のカテゴリに選別される。例えば、カテゴリは、限定されないが、(a)挿入長さよりも短い標的、(b)挿入長さ+Xよりも短い、または同等の標的、および(c)挿入長さ+Xより長い標的を含み、ここでXは、プローブによって標的とされない望ましい隙間の長さである。カテゴリ(a)の標的配列は、挿入もまた相補的になる非標的領域(反復、高/低GC、回文配列等)の複雑性に応じて、標的配列の左または右側に集中あるいは整列する標的結合配列で、標的とされる。カテゴリ(b)の標的配列は、カテゴリ(a)と同じ様式で標的とされ、ここでXは、標的結合配列が標的としない望ましい隙間の長さである。カテゴリ(c)の標的に関しては、標的

10

20

30

40

50

の全長は、標的結合配列の長さで割られ、および最も近い整数値に切り上げられ、標的配列のための挿入セットを生成するために標的配列のすべてを完全に標的とするのに必要とされる標的結合配列の数を表す。随意に、標的結合配列の数は、減数後に減らされてもよく、および標的結合配列間の隙間は望ましい隙間の長さY未満である。その後、この全過程を標的配列ごとに繰り返し、挿入ライブラリを形成する。次に、ライブラリ内の標的結合配列を、1つ以上のプライミング配列を含む1つ以上の非標的配列を加えることによって修飾し、ライブラリを構造上で合成し、合成は実施例3からの一般的方法を使用して、ホスホラミダイト化学作用によって行なわれ、そしてプローブを分子タグで標識する。

【0243】

実施例23：オーバーラップするプローブライブラリのための設計方法

10

【0244】

少なくとも100の標的配列が提供され、相補的なプローブ標的結合配列の所望の長さと比較し、長さに基づいて別個のカテゴリに選別される。例えば、カテゴリは、限定されないが、(a)挿入長さよりも短い標的、(b)挿入長さ+Xよりも短い、または同等の標的、および(c)挿入長さ+Xより長い標的を含み、ここでXは、プローブによって標的とされない望ましい隙間の長さである。カテゴリ(a)の標的配列は、挿入もまた相補的になる非標的領域(反復、高/低GC、回文配列等)の複雑性に依りて、標的配列の左または右側に集中あるいは整列する標的結合配列で、標的とされる。カテゴリ(b)の標的配列は、カテゴリ(a)と同じ様式で標的とされ、ここでXは、標的結合配列が標的としない望ましい隙間の長さである。カテゴリ(c)の標的に関して、標的の全長は、標的結合配列の長さで割られ、および最も近い整数値に切り上げられ、標的配列のすべてを完全に標的とするのに必要とされる標的結合配列の数を表す。相補的な標的結合配列は、その後、標的配列にわたって(随意に均一に)一定間隔で配置され、標的配列のための挿入セットを生成するために、オーバーラップを可能にする。その後、この全過程を標的配列ごとに繰り返して、挿入ライブラリを形成する。次に、ライブラリ内の標的結合配列を、1つ以上のプライミング配列を含む1つ以上の非標的配列を加えることによって修飾し、ライブラリを構造上で合成し、合成は実施例3からの一般的方法を使用して、ホスホラミダイト化学作用によって行なわれ、そしてプローブを分子タグで標識する。

20

【0245】

実施例24：混合されたプローブライブラリのための設計方法

30

【0246】

プローブライブラリは、修飾を伴う実施例22と23の一般的方法に従って合成される。オーバーラップしない挿入、オーバーラップする挿入、または混合された(オーバーラップし、かつオーバーラップしない)挿入を含むセットが、各標的配列のために生成される。

【0247】

実施例25：エクソン標的のためのポリヌクレオチドプローブ

【0248】

ポリヌクレオチドプローブは、ゲノム内のエクソンを標的とする場合もあり、および遺伝子は複数のエクソンを含み得る。例えば、ヒト白血球抗原(HLA)遺伝子は7つのエクソンを含み、それらの3つが表6に表記される。

40

【0249】

【表 6】

表6

Seq ID	種類	エクソン 長さ (bp)	<u>エクソン</u> (下線)を含むゲノム配列
5	ヒト HLA 遺伝子エクソン 1	63	CCCACCGGGACTCAGATTCTCCCCAGACGCCGAGGAT GGTGCTCATGGCGCCCCGAACCCTCCTCCTGCTGCTC <u>TCAGGGGGCCCTGGCCCTGACCCAGACCTGGGCGCGT</u> GAGTGCAGGGTCTGCAGGGAAATGGGCGCGTGAGTG CAGGGTCTGC
6	ヒト HLA 遺伝子エクソン 2	270	GCTCCCAGGTTCCCCTCCATGAGGTATTCTACACC <u>ACCATGTCCCGGCCCGCGCCGGGGAGCCCCGCTTCA</u> <u>TCTCCGTCGGCTACGTGGACGATACGCAGTTCGTGCG</u> <u>GTTCGACAGCGACGACGCGAGTCCGAGAGAGGAGCC</u> <u>GCGGGCGCCGTGGATGGAGCGGGAGGGGCCAAAGTA</u> <u>TTGGGACCGGAACACACAGATCTGCAAGGCCCAGGC</u> <u>ACAGACTGAACGAGAGAACCTGCGGATCGCGCTCCG</u> <u>CTACTACAACCAGAGCGAGGGCGGTGAGTTGACCCC</u> GG
7	ヒト HLA 遺伝子エクソン 5	117	TAGCAGGGTTCAGGGTTCCTCACCTTCCCCCTTTTCC CAGCCATCTTCCCAGCCCACCGTCCCCATCGTGGGCA <u>TCGTTGCTGGCTTGGTTCTACTTGTAGCTGTGGTCACT</u> <u>GGAGCTGTGGTCGCTGCTGTAATGTGGAGGAAGAAG</u> <u>AGCTCAGGTAAGGAAGGGGT</u>

10

20

任意のエクソンに関して、標的結合配列と非標的結合配列の様々な組み合わせを、様々な構成と長さのプローブを設計するために使用してもよい。HLAエクソン1を標的とする非限定なプローブ設計は、表7に例として示され、および1本鎖のみのプローブの配列が示される。非標的結合配列と標的結合配列のサイズ、および全プローブ長さは、表8に表記される。

30

【0250】

【表 7】

表7

Seq ID	標的	5' 非標的結合配列	標的結合配列	3' 非標的結合配列
8	ヒト HLA エクソン 1	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTGATTC</u> TCCCCAGA CGCCGAGG ATGGTGCT C	ATGGCGCCCCGAACCCTCCTCCTGC TGCTCTCAGGGGCCCTGGCCCTGAC CCAGACCTGGGCG	CGTGAGTGCAG GGTCTGCAGGG AAATGGTAGTG <u>TCGGAGGTCGT</u> <u>TCCT</u>
9	ヒト HLA エクソン 2	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTG</u>	GGTTCCTCCACTCCATGAGGTATTTCT ACACCACCATGTCCCGGCCCGGCG CCGGGGAGCCCCGCTTCATCTCCGT CGGCTACGTGGACGATACGCAGTT CGTGCGGTTTCGACAGCGACGAC	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
10	ヒト HLA エクソン 2	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTG</u>	TGGATGGAGCGGGAGGGGCCAAAG TATTGGGACCGGAACACACAGATC TGCAAGGCCAGGCACAGACTGAA CGAGAGAACCTGCGGATCGCGCTC CGCTACTACAACCAGAGCGAGGGC	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
11	ヒト HLA エクソン 2	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTG</u>	GGCTACGTGGACGATACGCAGTTC GTGCGGTTTCGACAGCGACGACGCG AGTCCGAGAGAGGAGCCGCGGGCG CCGTGGATGGAGCGGGAGGGGCCA AAGTATTGGGACCGGAACACACAG	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
12	ヒト HLA エクソン 1	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTG</u>	ATGGCGCCCCGAACCCTCCTCCTGC TGCTCTCAGGGGCCCTGGCCCTGAC CCAGACCTGGGCG	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
13	ヒト HLA エクソン 2	AAAAAAAAA AAAAAAAAA AAAAAAAAA <u>AGTTACCC</u> <u>AAGAACGC</u> <u>AGCTG</u>	GGTTCCTCCACTCCATGAGGTATTTCT ACACCACCATGTCCCGGCCCGGCG CCGGGGAGCCCCGCTTCATCTCCGT CGGCTACGTGGACGATACGCAGTT CGTGCGGTTTCGACAGCGACGAC	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u> AAA AAAAAAAAAAAA AAAAAAAAAAAA
14	ヒト HLA エクソン 2	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTG</u> AAAA AAAA AAAA	GGTTCCTCCACTCCATGAGGTATTTCT ACACCACCATGTCCCGGCCCGGCG CCGGGGAGCCCCGCTTCATCTCCGT CGGCTACGTGGACGATACGCAGTT CGTGCGGTTTCGACAGCGACGAC	AAAAAAAAAAAA AAAAAAAAAAAA AAATAGTGTCG <u>GAGGTCGTTTCCT</u>
15	ヒト HLA エクソン 5	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTG</u>	AGCCATCTTCCCAGCCCACCGTCCC CATCGTGGGCATCGTTGCTGGCTTG GTTCTACTTGTA	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
16	ヒト HLA エクソン 5	<u>GTTACCCA</u> <u>AGAACGCA</u> <u>GCTG</u>	GCTGTGGTCACTGGAGCTGTGGTCG CTGCTGTAATGTGGAGGAAGAAGA GCTCAG	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>

非標的結合配列のプライミング配列に下線を付す。

【表 8】

表8

Seq ID	標的	5' 非標的結合配列の長さ (bp)	標的結合配列の長さ (bp)	3' 非標的結合配列の長さ (bp)	全プローブ長さ (bp)
8	ヒト HLA エクソン 1	49	63	48	160
9	ヒト HLA エクソン 2	20	120	20	160
10	ヒト HLA エクソン 2	20	120	20	160
11	ヒト HLA エクソン 2	20	120	20	160
12	ヒト HLA エクソン 1	20	63	20	103
13	ヒト HLA エクソン 2	45	120	45	210
14	ヒト HLA エクソン 2	45	120	45	210
15	ヒト HLA エクソン 5	20	62	20	102
16	ヒト HLA エクソン 5	20	55	20	95

10

20

複数のプローブの様々な構成を、任意のエクソン、例えばヒト H L A エクソン 2 をカバーするために使用してもよい。SEQ ID : 9 または 10 を含むプローブは、1 つのセットと標的ヒト H L A エクソン 2 を含むが、共に標的エクソンに隙間を残し、および重複配列を含まない (図 4 F 参照)。SEQ ID : 11 を含むプローブは H L A エクソン 2 を標的とし、および SEQ ID : 9 および 10 とオーバーラップする標的結合配列を含む (図 4 G 参照)。各々が SEQ ID : 15 または 16 を含むプローブは、セット、標的ヒト H L A エクソン 5 を含み、および標的または非標的領域の重複する領域を標的としない。SEQ ID : 8 - 16 に対応するプローブは構造上で合成され、および合成は、実施例 3 からの一般的方法を使用し、ホスホラミダイト化学作用によって行なわれ、ピオチン等の少なくとも 1 つの分子タグで随意に標識される。

30

40

50

【0252】

実施例 26 : ポリヌクレオチドプローブライブラリを用いたゲノム DNA 捕捉

【0253】

ヒトエクソームを標的とする少なくとも 500,000 の非同一のポリヌクレオチドを含むポリヌクレオチド標的ライブラリは、実施例 3 からの一般的方法を使用してホスホラミダイト化学作用によって、およびライブラリ 4 を生成するために実施例 11 の一般的方法を使用して制御された化学量論によって、構造上で設計され、合成された。次に、ポリヌクレオチドをピオチンで標識し、溶解して、エクソームプローブライブラリ溶液を作った。乾燥させたインデックス付けライブラリプールを、実施例 16 の一般的方法を使用してゲノム DNA (gDNA) サンプルから得た。

【0254】

エクソームプローブライブラリ溶液、ハイブリダイゼーション溶液、遮断薬混合物 A、および遮断薬混合物 B を 2 秒間、パルス渦で混ぜた。ハイブリダイゼーション溶液を 10 分間、またはすべての沈殿物が溶解するまで、65 °C で加熱し、その後さらに 5 分間、ベンチトップ上で室温に戻した。20 µL のハイブリダイゼーション溶液と 4 µL のエクソームプローブライブラリ溶液を、薄壁の PCR 0.2 mL ストリップチューブに加え、ピペットで優しく混ぜ合わせた。混ぜ合わせたハイブリダイゼーション溶液 / エクソ

ムプローブ溶液を 95 °C で 2 分間、105 °C のふたの温度で、サーマルサイクラーにおいて加熱し、そして少なくとも 10 分間、氷の上で素早く冷却した。溶液をその後 5 分間、ベンチトップ上で室温に冷ました。ハイブリダイゼーション溶液 / エクソームプローブライブラリ溶液を冷却しながら、各ゲノム DNA サンプルにつき 9 µL に水を加え、および 5 µL の遮断薬ミックス A と 2 µL の遮断薬ミックス B を、薄壁の PCR 0.2 mL のストリップチューブ内の乾燥させたインデックス付けライブラリプールに加えた。溶液をその後、ピペットで優しく混ぜ合わせた。プールしたライブラリ / 遮断薬チューブを、105 °C のふた温度でサーマルサイクラーにおいて 95 °C で 5 分間、加熱し、その後、次の工程に進む前に 5 分間以下で、ベンチトップ上で室温に戻した。ハイブリダイゼーション混合物 / プローブ溶液をピペットで混ぜ合わせ、全 24 µL のプールされたライブラリ / 遮断薬チューブに添加した。全捕捉反応ウェルをピペットで優しく混ぜ合わせて、泡が立たないようにした。サンプルチューブが確実にしっかりと密封されるように、チューブをパルススピンした。捕捉 / ハイブリダイゼーション反応物を 70 °C で 16 時間、PCR サーマルサイクラーで加熱し、ふたの温度は 85 °C にした。

10

【0255】

結合緩衝液、洗浄緩衝液 1、および洗浄緩衝液 2 を、すべての沈殿物が溶液に溶解するまで 48 °C で加熱した。700 µL の洗浄緩衝液 2 を、捕捉ごとに等分して、48 °C にあらかじめ加熱した。ストレプトアビジン結合ビーズと DNA 精製ビーズを、少なくとも 30 分間、室温で平衡に保った。KAPA HiFi HotStart Ready Mix 等のポリメラーゼと増幅プライマーを氷の上で解凍した。一旦、試薬を解凍し、パルス渦で 2 秒間、攪拌した。捕捉反応につき 500 µL の 80 パーセントのエタノールを調製した。ストレプトアビジン結合ビーズをあらかじめ室温と等しくし、均質化されるまで攪拌した。100 µL のストレプトアビジン結合ビーズを加えて、捕捉反応ごとに 1.5 mL の微量遠心機チューブを浄化した。各チューブに 200 µL の結合緩衝液を加えて、均質化されるまで各チューブをピペットで混ぜ合わせた。チューブを磁気スタンドに置いた。ストレプトアビジン結合ビーズを 1 分以内にペレットにした。チューブを取り除き、そして透明な上清を捨て、確実にビーズペレットを妨害しないようにした。チューブを磁気スタンドから取り出して、追加で 2 回の洗浄を繰り返した。第 3 の洗浄後、チューブを取り除き、そして透明な上清を捨てた。最終 200 µL の結合緩衝液を添加し、およびビーズを均質になるまで攪拌して再懸濁した。

20

30

【0256】

ハイブリダイゼーション反応を完了した後に、サーマルサイクラーのふたを開き、および捕捉反応の全量を洗浄したストレプトアビジン洗浄ビーズへと素早く移した (36 - 40 µL)。混合物を室温で 30 分間、振とう機、ロッカー、または回転器上で、捕捉反応 / ストレプトアビジン結合ビーズ溶液を均質に保つのに十分な速度で混ぜた。捕捉反応 / ストレプトアビジン結合ビーズ溶液を混合器から取り出し、パルススピンによって確実にすべての溶液がチューブの底にあるようにした。サンプルを磁気スタンドに置き、ストレプトアビジン結合ビーズをペレット化し、1 分以内に透明な上清を残した。透明な上清を除去して捨てた。チューブを磁気スタンドから取り出し、200 µL の洗浄緩衝液を室温に加え、その後、均質化されるまでピペットで混ぜた。すべての溶液が確実にチューブの底にあるように、チューブをパルススピンした。サーマルサイクラーは以下の疾病でプログラムされた (表 9)。

40

【0257】

加熱されたふたの温度は 105 °C に設定された。

【0258】

【表 9】

表9

工程	温度	時間	サイクル数
1	98°C	45 秒	1
2	98°C	15 秒	9
	60°C	30 秒	
	72°C	30 秒	
3	72°C	1 分	1
4	4°C	保持	

10

【0259】

増幅プライマー（2.5 μ L）と、KAPA HiFi HotStart Ready Mix（25 μ L）等のポリメラーゼを、水/ストレプトアビジン結合ビーズスラリーを含むチューブに加えて、ピペットでチューブを混ぜた。次にチューブを2つの反応に分けた。チューブをパルススピンしてサーマルサイクラーに移し、そして表9のサイクリングプログラムを開始した。サーマルサイクラープログラムが完了すると、サンプルをブロックから取り出し、直ちに精製にかけた。あらかじめ室温と等しくしたDNA精製ビーズを、均質化されるまで攪拌した。90 μ L（1.8x）の均質化したDNA精製ビーズをチューブに加えて、攪拌してよく混ぜた。チューブを室温で5分間インキュベートし、そして磁気スタンドに置いた。DNA精製ビーズをペレット化し、1分以内に透明な上清を残した。透明な上清を捨て、およびチューブを磁気スタンド上に残した。DNA精製ビーズペレットを200 μ Lの新たに準備した80パーセントのエタノールで洗浄し、1分間インキュベートし、その度に取り除いてエタノールを捨てた。磁気スタンドにチューブを保持しながら、洗浄を1回繰り返し、つまり合計2回の洗浄を行った。残ったすべてのエタノールを10 μ Lのピペットで取り除いて捨て、DNA精製ビーズペレットを確実に妨害しないようにした。DNA精製ビーズペレットを5～10分間、またはペレットが乾燥するまで、磁気スタンドで風乾した。チューブを磁気スタンドから取り出し、32 μ Lの水を加え、均質化するまでピペットで移し、そして室温で2分間、インキュベートした。チューブを3分間、またはビーズが完全にペレット化するまで、磁気スタンドに置いた。30 μ Lの透明な上清を回収して、清潔な薄壁のPCR 0.2 mLのストリップチューブに移し、DNA精製ビーズペレットを確実に妨害しないようにした。平均のフラグメント長さは、分析器具において150 bp～1000 bpに設定した範囲を使用して、約375 bp～約425 bpであった。理想的には、最終濃度値は少なくとも約15 ng/ μ Lである。各捕捉を定量化し、次世代（NGS）を使用して検証した。

20

30

40

【0260】

コンパレータエクソーム捕捉キット（コンパレータキットD）と比較した、NGSメトリックの概要は表10、表11、および図37に示される。ライブラリ4は、コンパレータキットDよりも高いエクソン標的率に対応するプローブ（ベイト）を有する。これは、ライブラリ4を使用した標的配列の同等の質とカバー率を得るためのより少ないシーケンシングをもたらす。

【0261】

【表 10】

表10

NGS メトリック	コンパレータキット D	ライブラリ 4
標的領域	38.8 Mb	33.2 Mb
ベイト領域	50.8 Mb	36.7 Mb
ベイト設計の効率	76.5%	90.3%
捕捉 Plex	8-plex	8-plex
PF 読み取り	57.7M	49.3M
標準化カバー率	150X	150X
HS ライブラリサイズ	30.3 M	404.0 M
複製のパーセント	32.5%	2.5%
Fold Enrichment	43.2	48.6
Fold 80 Base Penalty	1.84	1.40

10

20

【0262】

【表 11】

表11

NGS メトリック	コンパレータキット D	ライブラリ 4
フィルタを通過した固有の読み取りパーセント (PCT_PF_UQ_READS)	67.6%	97.5%
1X での標的塩基のパーセント	99.8%	99.8%
20X での標的塩基のパーセント	90.3%	99.3%
30X での標的塩基のパーセント	72.4%	96.2%

30

【0263】

キット D とライブラリ 4 の両方に関する、オーバーラップする標的部位の比較（全読み取りは 96X のカバー率に標準化）は、表 12 と図 37 に示される。ライブラリ 4 は、ハイブリダイゼーションにつき 8 つのサンプルとして処理され、およびキット D は、ハイブリダイゼーションにつき 2 つのサンプルで処理された。付加的に、両方のライブラリに関して、オーバーラップ領域からの一塩基多型とインフレーム欠失呼出しが、“Genome in a Bottle” NA12878 参照データから特定された高信頼度の領域に対して比較された（表 13）。ライブラリ 4 は、SNP とインデルを特定する際に、キット D と同様に、またはより良好に（より高いインデル精度）機能した。

40

【0264】

【表 1 2】

表12

NGS メトリック	コンパレータキット D	ライブラリ 4
フィルタを通過した固有の読み取りパーセント (PCT_PF_UQ_READS)	94.60%	97.7%
選択塩基のパーセント	79%	80%
1X での標的塩基のパーセント	100%	100%
20X での標的塩基のパーセント	90%	96%
30X での標的塩基のパーセント	71%	77%
Fold Enrichment	44.9	49.9
Fold 80 Base Penalty	1.76	1.4
HS ライブラリサイズ	122 M	267 M

10

20

【 0 2 6 5】

【表 1 3】

表13

変異	コンパレータキット D		ライブラリ 4	
	精度	感度	精度	感度
一塩基多型(SNP)	98.59%	99.23%	99.05%	99.27%
インフレーム欠失(インデル)	76.42%	94.12%	87.76%	94.85%
合計	98.14%	99.15%	98.85%	99.20%

30

【 0 2 6 6】

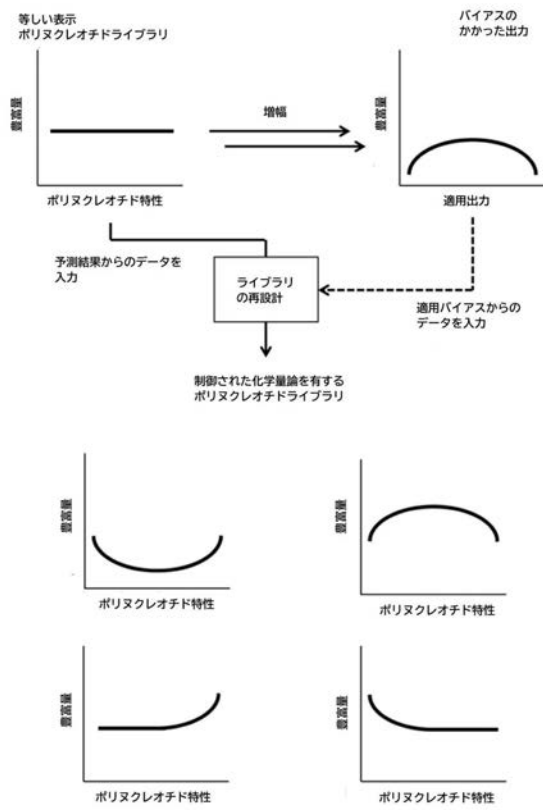
精度は、真の呼出し 対 合計の（真または偽の）陽性の呼出しの比率を表す。感度は、真の陽性の呼出し 対 合計の真値（真の陽性と偽の陰性）の比率を表す。

【 0 2 6 7】

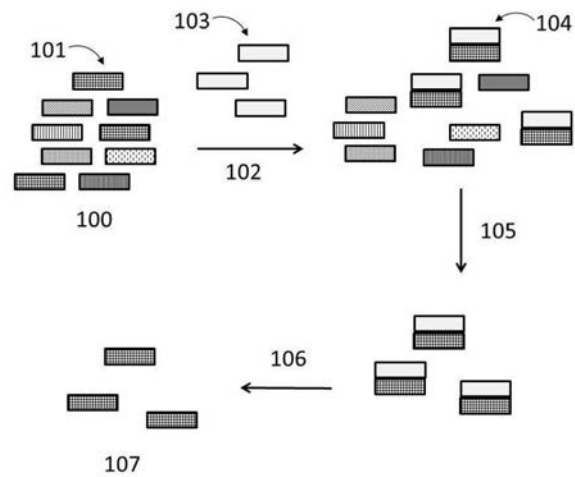
本発明の好ましい実施形態が本明細書に示され、記載されたが、そのような実施形態が単なる例として提供されていることは、当業者にとって明白だろう。多数の変形、変更、および置き換えが、本発明から逸脱することなく、当業者によって想到されるだろう。本明細書に記載される本発明の実施形態の様々な代案が、本発明の実施において利用され得ることを理解されたい。以下の特許請求の範囲は本発明の範囲を定義するものであり、この特許請求の範囲内の方法と構造、およびそれらの同等物を包含するように意図されている。

40

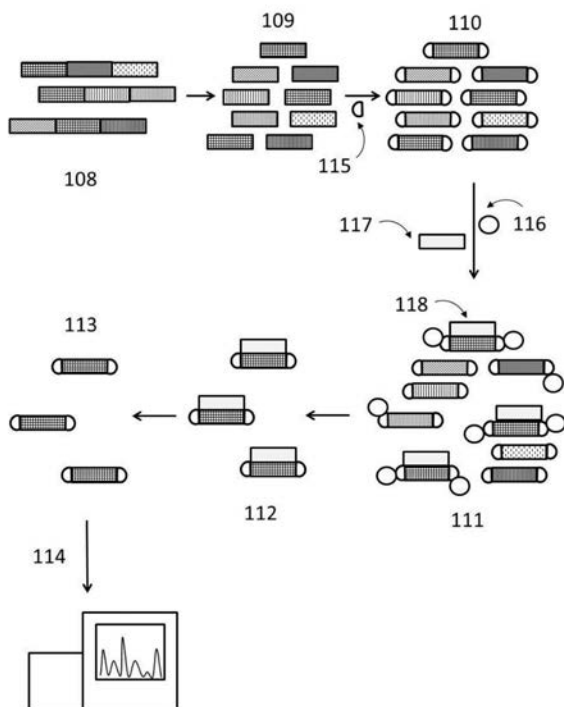
【図 1 A】



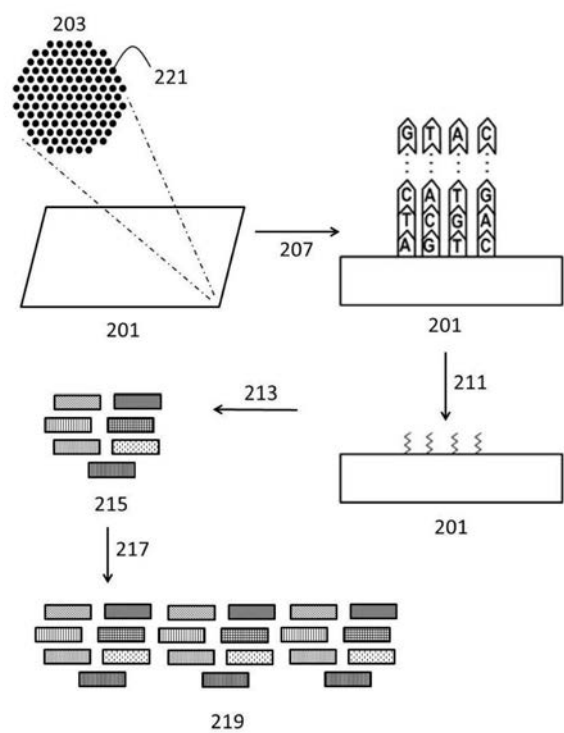
【図 1 B】



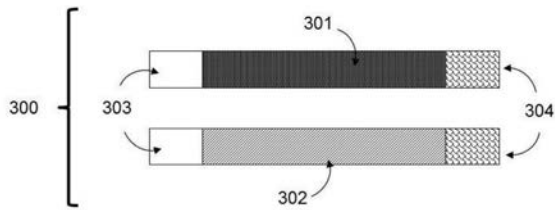
【図 1 C】



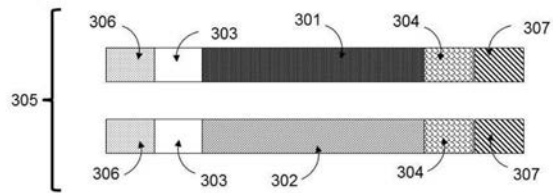
【図 2】



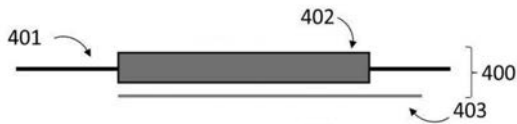
【図 3 A】



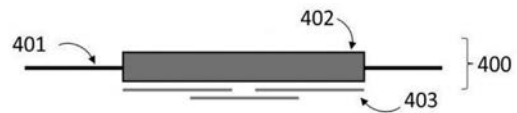
【図 3 B】



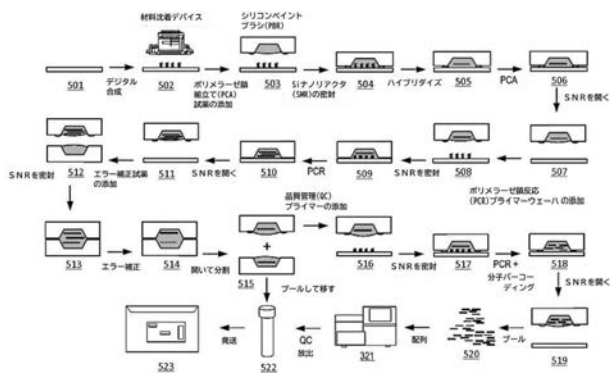
【図 4 A】



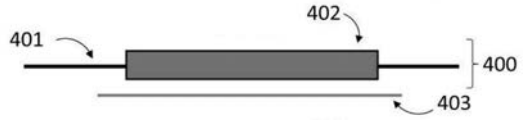
【図 4 G】



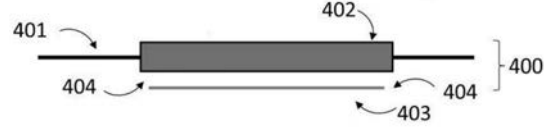
【図 5】



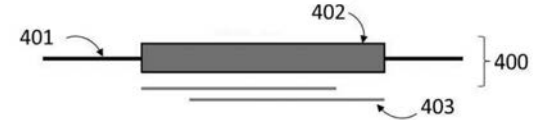
【図 4 B】



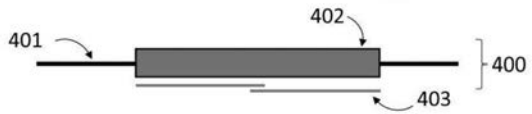
【図 4 C】



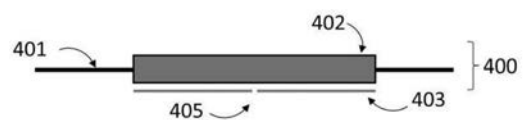
【図 4 D】



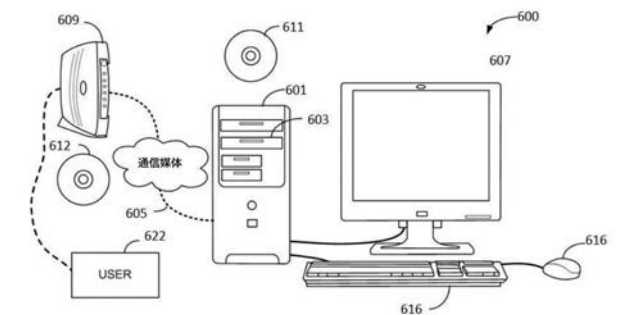
【図 4 E】



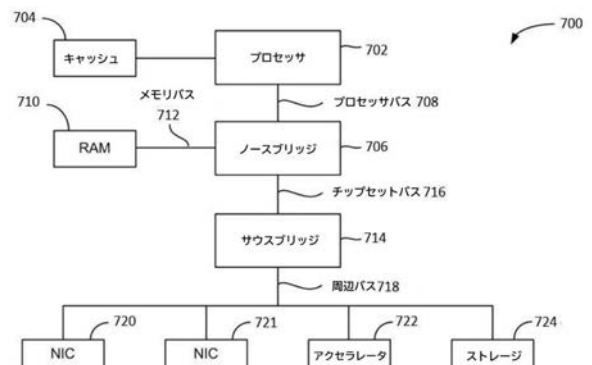
【図 4 F】



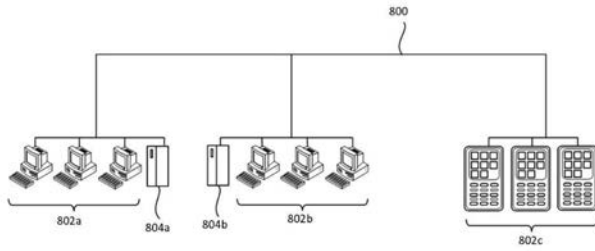
【図 6】



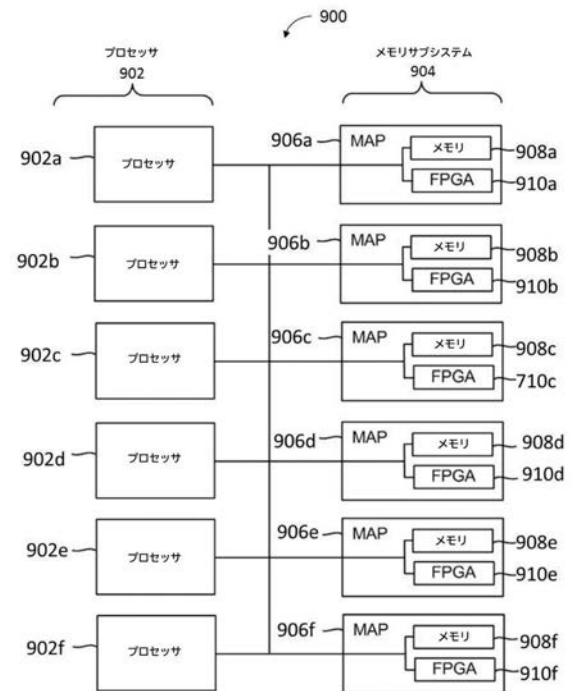
【図 7】



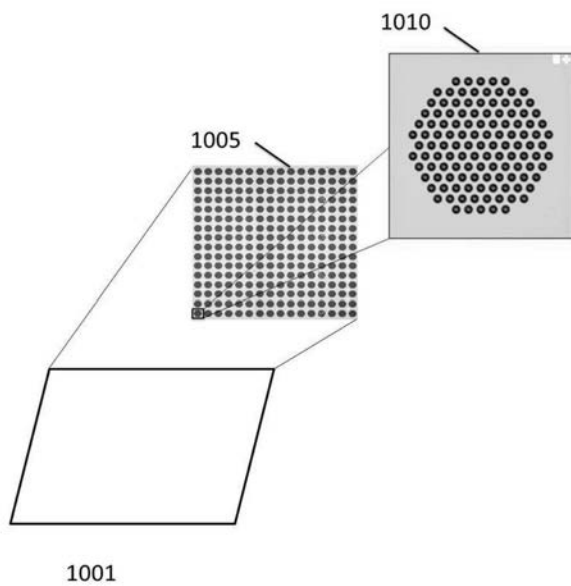
【図 8】



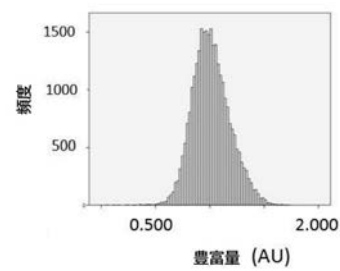
【図 9】



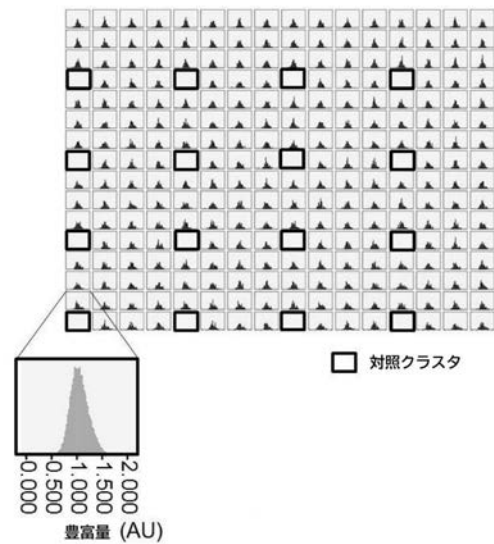
【図 10】



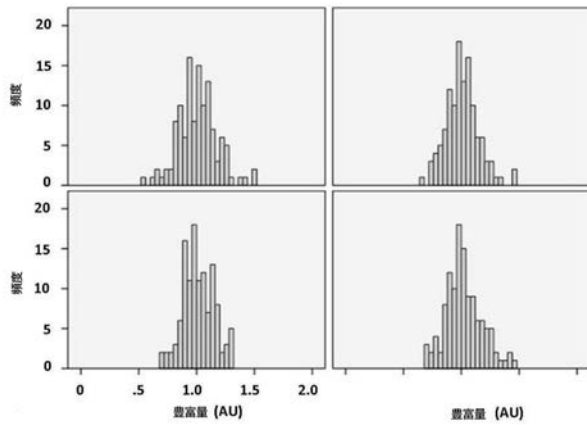
【図 11 A】



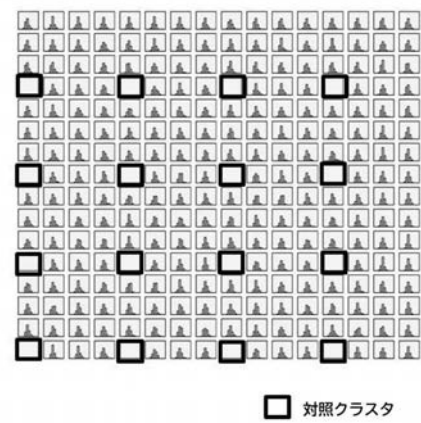
【図 11 B】



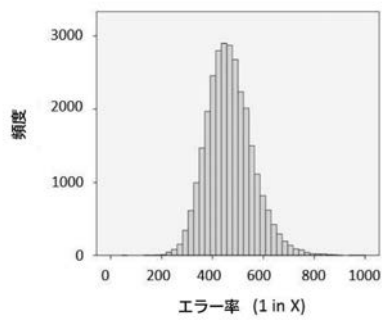
【図 1 2】



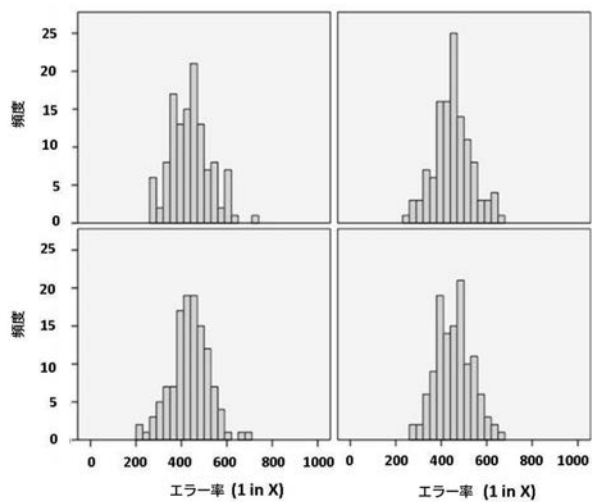
【図 1 3 B】



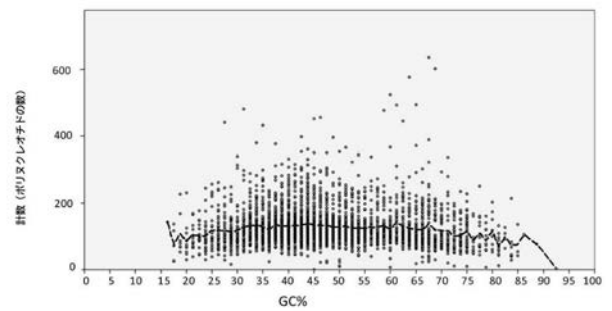
【図 1 3 A】



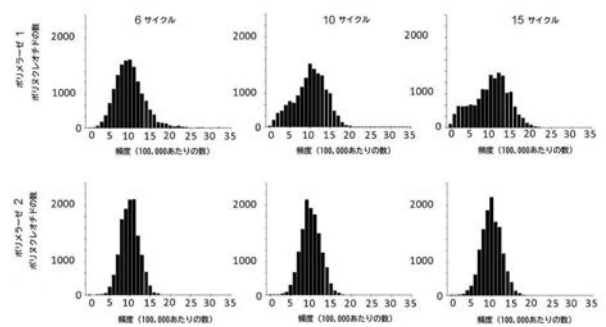
【図 1 4】



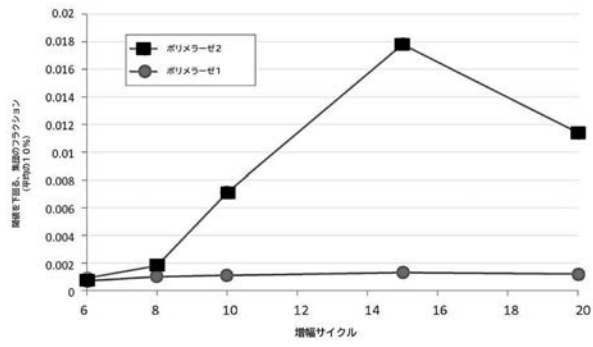
【図 1 5】



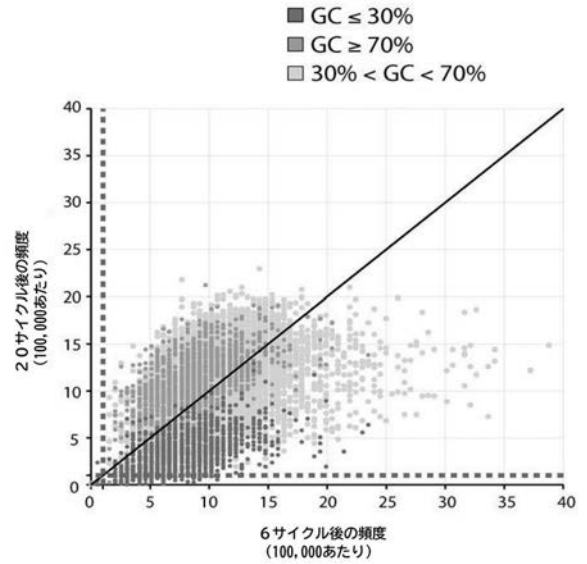
【図 1 6】



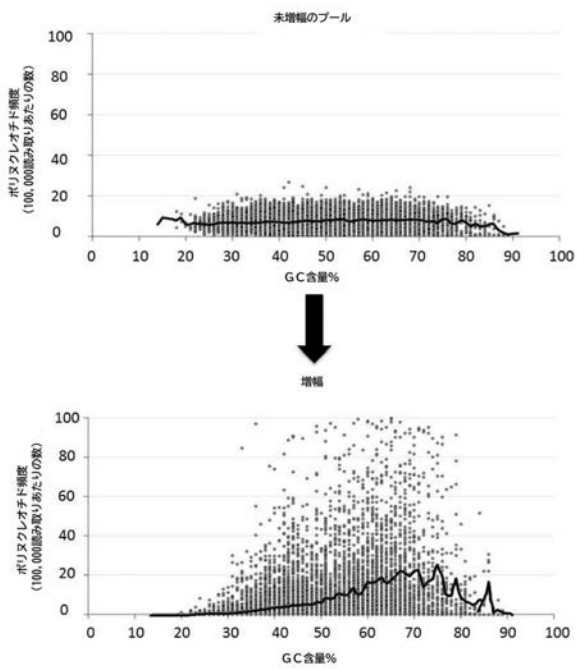
【図 17】



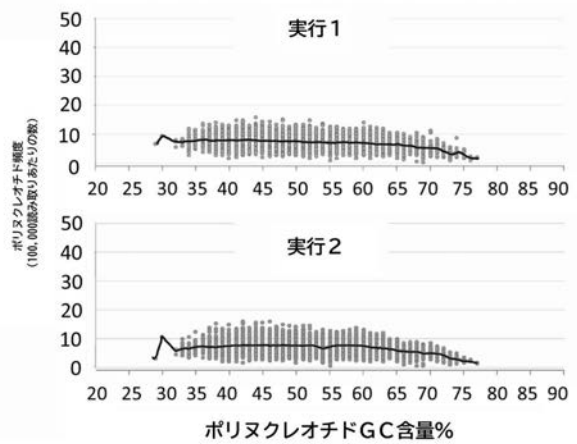
【図 18】



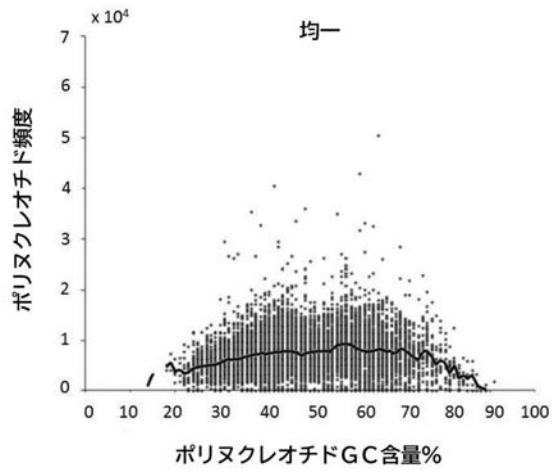
【図 19】



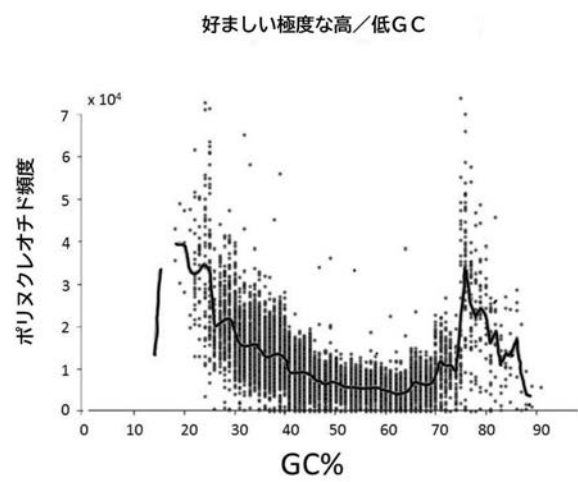
【図 20】



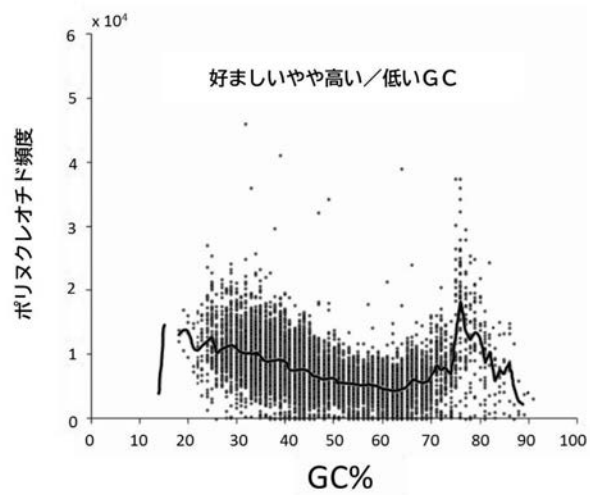
【図 2 1 A】



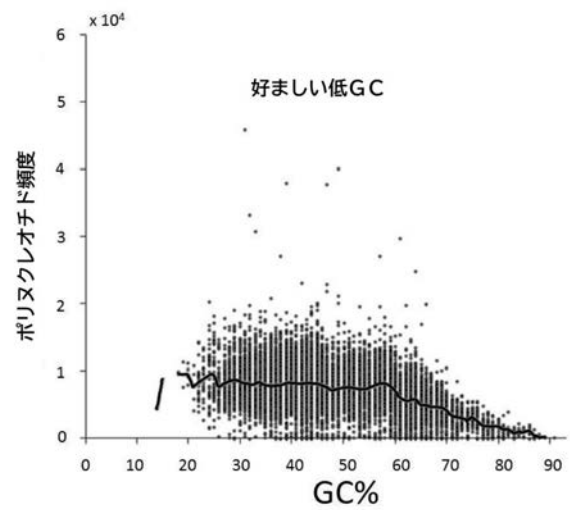
【図 2 1 B】



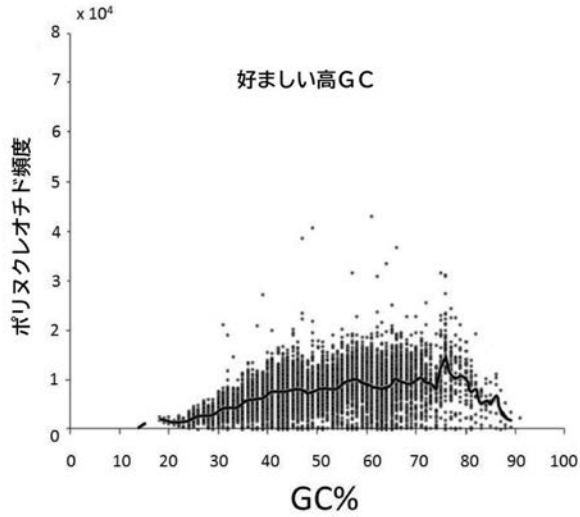
【図 2 1 C】



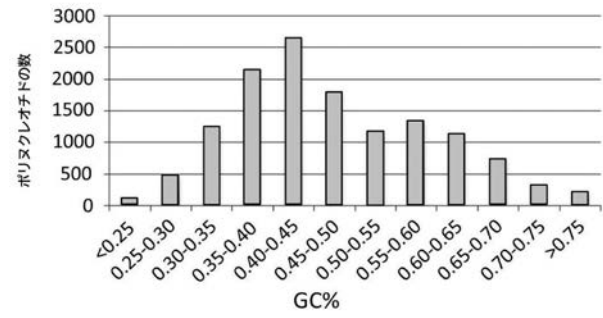
【図 2 1 D】



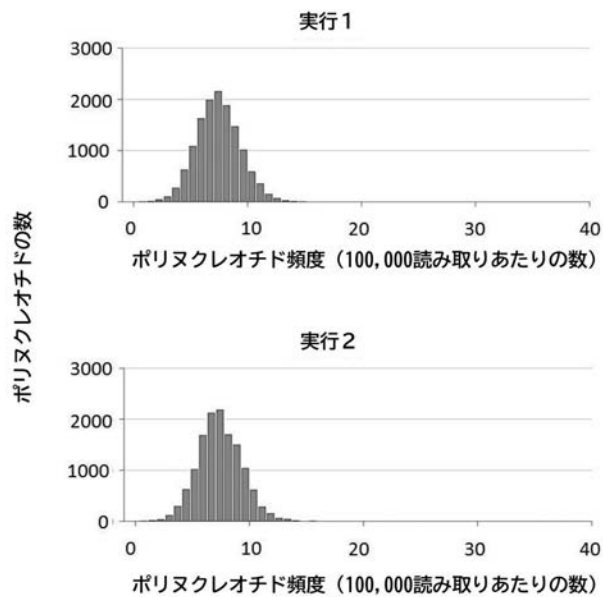
【図 2 1 E】



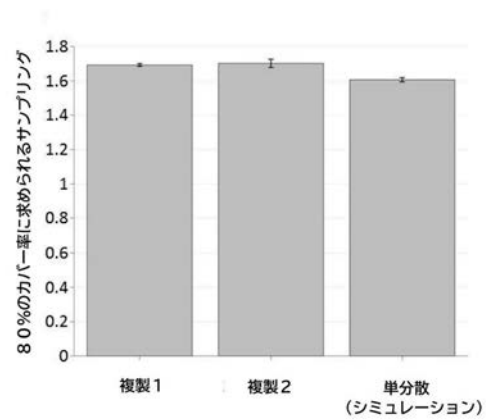
【図 2 2】



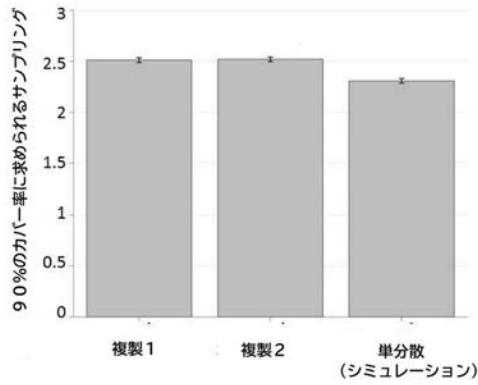
【図 2 3】



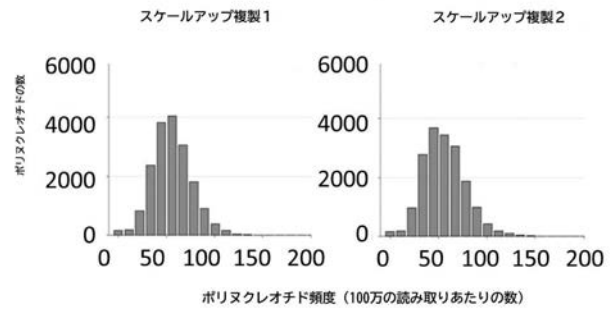
【図 2 4 A】



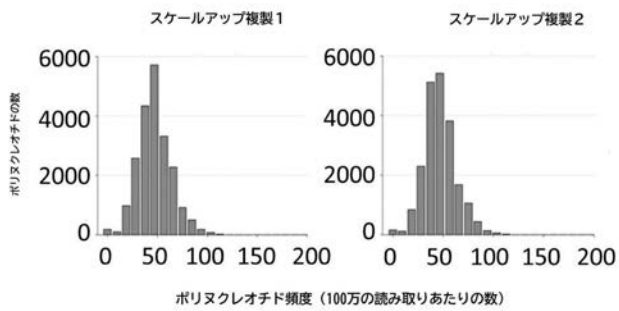
【図 2 4 B】



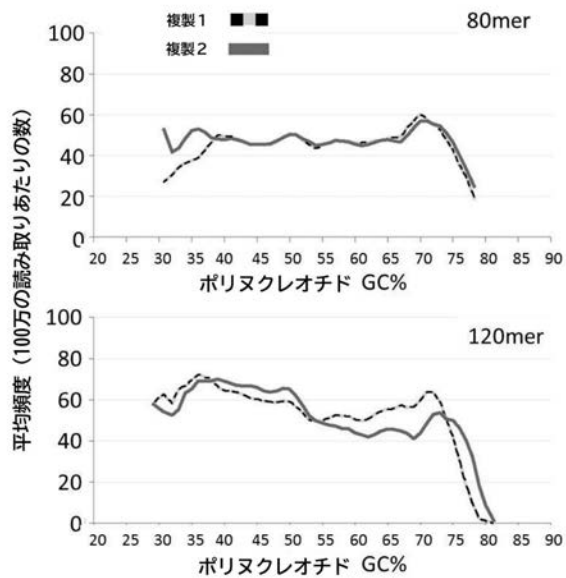
【図 2 6】



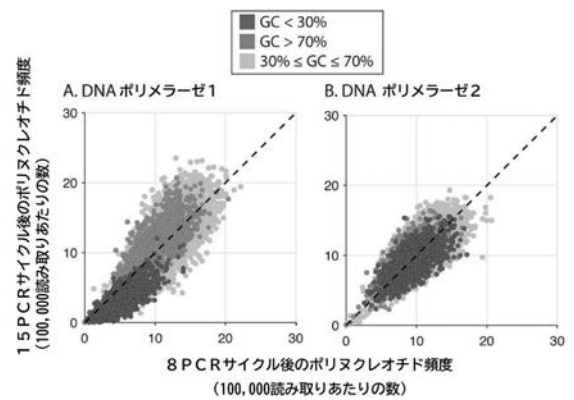
【図 2 5】



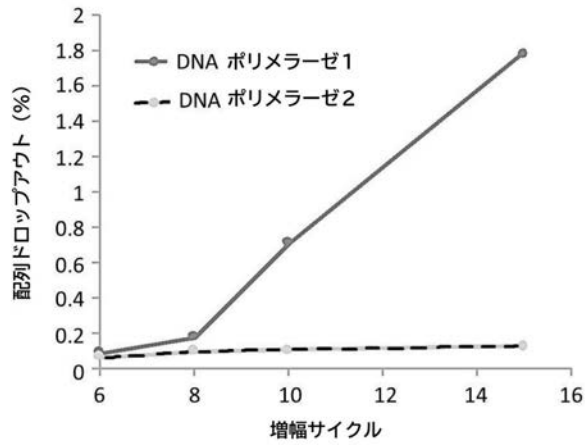
【図 2 7】



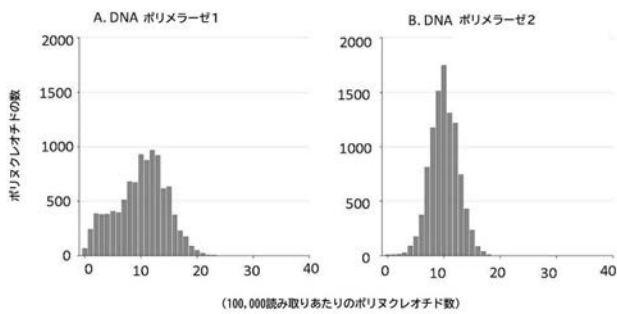
【図 2 8】



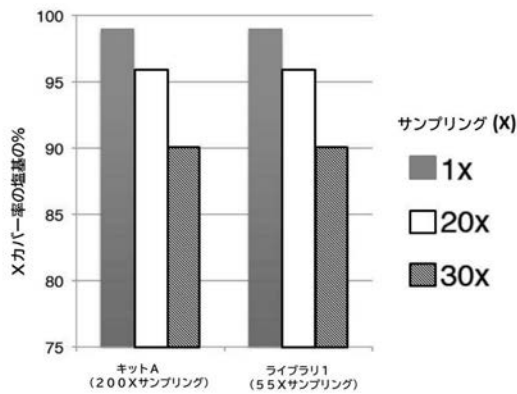
【図 29】



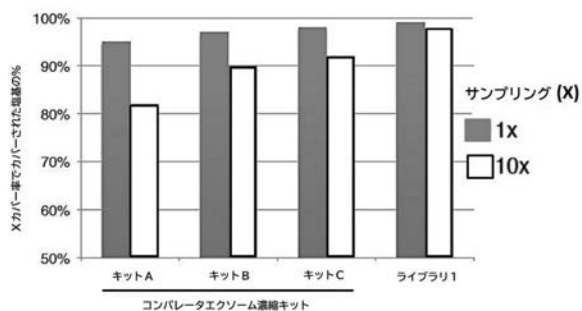
【図 30】



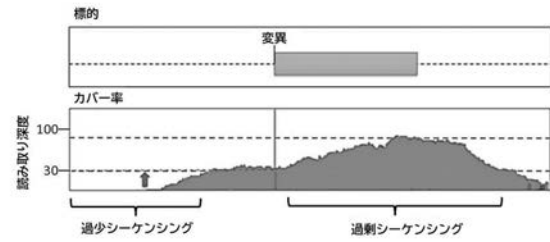
【図 32 A】



【図 32 B】



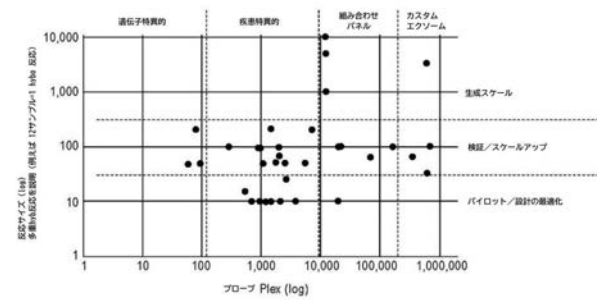
【図 31 A】



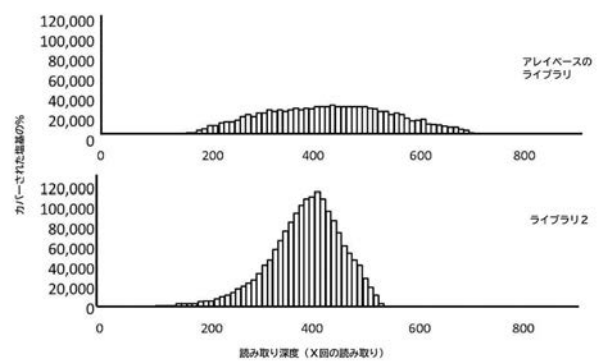
【図 31 B】



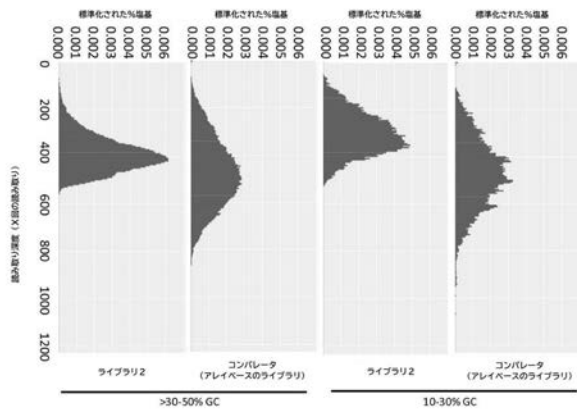
【図 33】



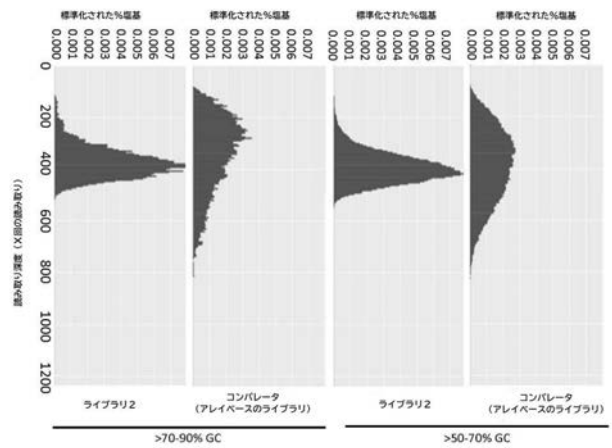
【図 34】



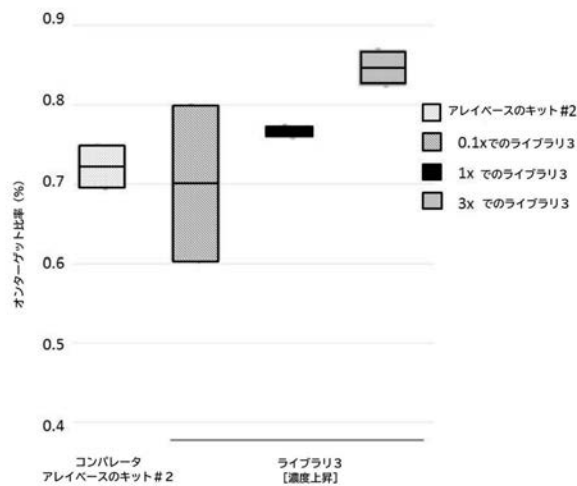
【図 3 5 A】



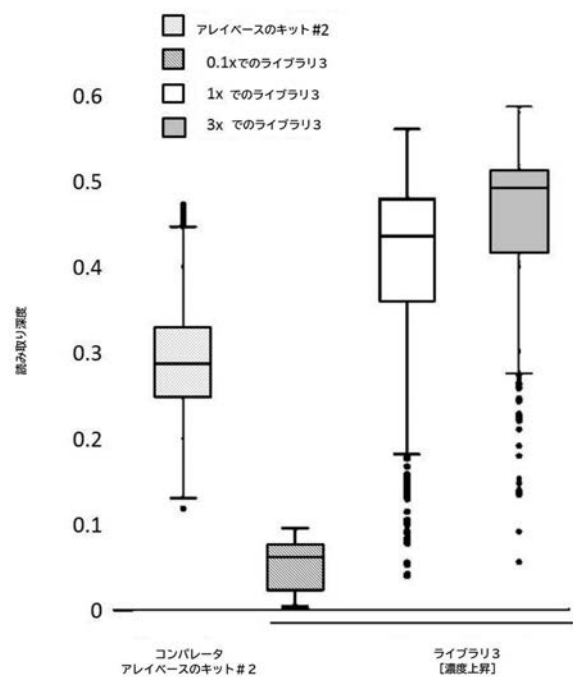
【図 3 5 B】



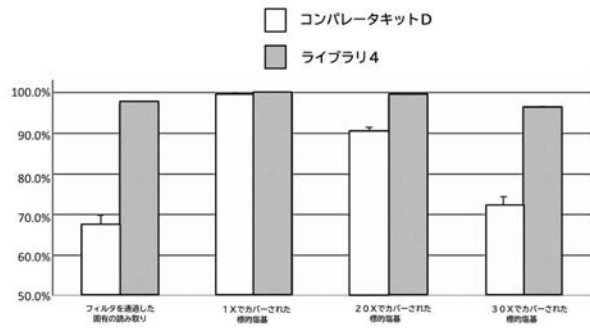
【図 3 6 A】



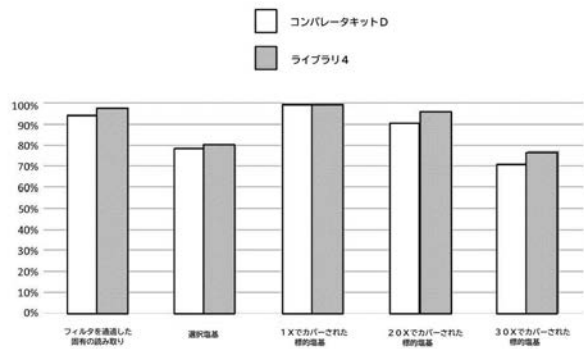
【図 3 6 B】



【図 3 7】



【図 3 8】



【配列表】

2020504709000001.app

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 17/62391									
A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - C40B 50/14, C12Q 1/68, C12N 15/10 (2018.01) CPC - C40B 50/14, C12Q 1/6869, C12Q 1/6876, C40B 50/00, C40B 40/08											
According to International Patent Classification (IPC) or to both national classification and IPC											
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) See Search History Document Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched See Search History Document Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) See Search History Document											
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%; padding: 5px;">Category*</th> <th style="width: 60%; padding: 5px;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="width: 30%; padding: 5px;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td style="vertical-align: top; padding: 5px;"> X — Y </td> <td style="vertical-align: top; padding: 5px;"> US 2016/0032396 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 4 February 2016 (04.02.2016) para [0013], [0043], [0089], [0127], [0140], [0268], [0293], [0303], [0325], [0418], [0419], [0466], [0489], [0493], [0513], [0523], [0576], [0590], [0601], [0684], [0711], [0731], [0918], [0925], [0938] </td> <td style="vertical-align: top; padding: 5px;"> 1-7, 9-11, 13-26, 28-30, 32-36, (55-67, 69-70, 72-78)/(1-7, 9-11, 13-26, 28-30, 32-36) 8, 12, 27, 31, (55-67)/(8, 12, 27, 31), 68, (69-70)/(8, 12, 27, 31), 71, (72-82)/(8, 12, 27, 31) </td> </tr> <tr> <td style="vertical-align: top; padding: 5px;">Y</td> <td style="vertical-align: top; padding: 5px;"> US 2015/0056609 A1 (LONGHORN VACCINES AND DIAGNOSTICS, LLC) 26 February 2015 (26.02.2015) para [0017] </td> <td style="vertical-align: top; padding: 5px;"> 8, 12, 27, 31, (55-67)/(8, 12, 27, 31), 68, (69-70)/(8, 12, 27, 31), 71, (72-82)/(8, 12, 27, 31) </td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X — Y	US 2016/0032396 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 4 February 2016 (04.02.2016) para [0013], [0043], [0089], [0127], [0140], [0268], [0293], [0303], [0325], [0418], [0419], [0466], [0489], [0493], [0513], [0523], [0576], [0590], [0601], [0684], [0711], [0731], [0918], [0925], [0938]	1-7, 9-11, 13-26, 28-30, 32-36, (55-67, 69-70, 72-78)/(1-7, 9-11, 13-26, 28-30, 32-36) 8, 12, 27, 31, (55-67)/(8, 12, 27, 31), 68, (69-70)/(8, 12, 27, 31), 71, (72-82)/(8, 12, 27, 31)	Y	US 2015/0056609 A1 (LONGHORN VACCINES AND DIAGNOSTICS, LLC) 26 February 2015 (26.02.2015) para [0017]	8, 12, 27, 31, (55-67)/(8, 12, 27, 31), 68, (69-70)/(8, 12, 27, 31), 71, (72-82)/(8, 12, 27, 31)
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.									
X — Y	US 2016/0032396 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 4 February 2016 (04.02.2016) para [0013], [0043], [0089], [0127], [0140], [0268], [0293], [0303], [0325], [0418], [0419], [0466], [0489], [0493], [0513], [0523], [0576], [0590], [0601], [0684], [0711], [0731], [0918], [0925], [0938]	1-7, 9-11, 13-26, 28-30, 32-36, (55-67, 69-70, 72-78)/(1-7, 9-11, 13-26, 28-30, 32-36) 8, 12, 27, 31, (55-67)/(8, 12, 27, 31), 68, (69-70)/(8, 12, 27, 31), 71, (72-82)/(8, 12, 27, 31)									
Y	US 2015/0056609 A1 (LONGHORN VACCINES AND DIAGNOSTICS, LLC) 26 February 2015 (26.02.2015) para [0017]	8, 12, 27, 31, (55-67)/(8, 12, 27, 31), 68, (69-70)/(8, 12, 27, 31), 71, (72-82)/(8, 12, 27, 31)									
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.											
<table style="width: 100%;"> <tr> <td style="width: 50%; vertical-align: top;"> * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="width: 50%; vertical-align: top;"> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family							
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family										
Date of the actual completion of the international search 15 March 2018	Date of mailing of the international search report 28 MAR 2018										
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young <small>PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774</small>										

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/62391

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

—— please see extra sheet ——

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-36, 55-78

Remark on Protest

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/82391

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

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-36, 55-78, drawn to compositions comprising a polynucleotide library and method for using said polynucleotide library to sequence genomic DNA

Group II: Claims 37-45, drawn to a method for generating a polynucleotide library

Group III: Claims 46-54, drawn to a method for amplifying a polynucleotide library having a preselected representation of a sequence feature

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

Special Technical Features

Group I requires compositions of matter comprising a polynucleotide library, and method steps for using said polynucleotide library for sequencing genomic DNA, not required by Groups II, III.

Group II requires method steps for generating a polynucleotide library by providing predetermined sequences encoding a set of polynucleotides, not required by Groups I, III.

Group III requires method steps for amplifying a polynucleotide library having a preselected representation by obtaining an amplification distribution for a set of polynucleotides, clustering polynucleotides into two or more bins based on a sequence feature, and adjusting the relative frequency of polynucleotides in at least one bin, not required by Groups I, II.

Common Technical Features

The feature shared by Groups I, II, III is a polynucleotide library comprising at least 5000 polynucleotides.

The feature shared by Groups II and III is synthesizing a polynucleotide library and amplifying the polynucleotide library.

However, these shared technical features do not represent a contribution over prior art, because the shared technical features are taught by US 2015/0038373 A1 to Twist Bioscience Corp. (hereinafter 'Twist').

Twist discloses a polynucleotide library comprising at least 5000 polynucleotides (para [0010] - "In practicing any of the methods of constructing a gene library as provided herein, in some embodiments, the list of genes comprises at least 500 genes. In some embodiments, the list comprises at least 5000 genes").

Twist discloses a method comprising synthesizing a polynucleotide library and amplifying the polynucleotide library (para [0204] - "Various methods are contemplated for the de novo synthesis of gene libraries with low error rates. . . . For example, oligonucleotide synthesis typically in situ on a DNA synthesis wafer, may be followed by a gene assembly reaction, such as polymerase cycling assembly (PCA), of the synthesized oligonucleotides into longer sequences. The assembled sequences may be amplified, e.g. through PCR").

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

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

フロントページの続き

(31)優先権主張番号 62/558,666

(32)優先日 平成29年9月14日(2017.9.14)

(33)優先権主張国・地域又は機関
米国(US)

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

(特許庁注：以下のものは登録商標)

1 . L i n u x

(72)発明者 チェン , スーユアン

アメリカ合衆国 9 4 4 0 3 カリフォルニア州 サン・マテオ トヨン・コート 1 6 4 5

Fターム(参考) 4B063 QA13 QQ42 QR08 QR32 QR42 QR56 QR62 QS25 QS34 QS39
QX02