

(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2017-532042

(P2017-532042A)

(43) 公表日 平成29年11月2日(2017.11.2)

(51) Int.Cl.		F I				テーマコード (参考)
<b>C 1 2 Q</b>	<b>1/68</b>	<b>(2006.01)</b>	C 1 2 Q	1/68	A	4 B 0 2 9
C 1 2 N	15/09	(2006.01)	C 1 2 N	15/00	A	4 B 0 6 3
C 1 2 M	1/00	(2006.01)	C 1 2 M	1/00	A	

審査請求 未請求 予備審査請求 未請求 (全 49 頁)

(21) 出願番号	特願2017-519628 (P2017-519628)	(71) 出願人	516380360
(86) (22) 出願日	平成27年10月29日 (2015.10.29)		1 O エックス ゲノミクス, インコーポレイテッド
(85) 翻訳文提出日	平成29年4月12日 (2017.4.12)		アメリカ合衆国, カリフォルニア州 9 4
(86) 国際出願番号	PCT/US2015/058142		5 6 6, プレザントン, コール センター
(87) 国際公開番号	W02016/069939		パークウェイ 7 0 6 8, スイート 4
(87) 国際公開日	平成28年5月6日 (2016.5.6)		O 1
(31) 優先権主張番号	62/072, 164	(74) 代理人	100079108
(32) 優先日	平成26年10月29日 (2014.10.29)		弁理士 稲葉 良幸
(33) 優先権主張国	米国 (US)	(74) 代理人	100109346
			弁理士 大貫 敏史
		(74) 代理人	100117189
			弁理士 江口 昭彦
		(74) 代理人	100134120
			弁理士 内藤 和彦

最終頁に続く

(54) 【発明の名称】 標的化核酸配列決定のための方法及び組成物

## (57) 【要約】

本発明は、ゲノムの標的化領域に含有されている配列情報を捕捉し、分析する方法、組成物及び系を対象とする。そのような標的化領域は、エクソーム、部分エクソーム、イントロン、エクソン及びイントロン領域の組み合わせ、遺伝子、遺伝子のパネル、ならびに目的でありうる全ゲノムの任意の他のサブセットを含むことができる。

【選択図】なし

**【特許請求の範囲】****【請求項 1】**

ゲノムの 1 つ以上の選択部分を配列決定する方法であって、

( a ) 出発ゲノム材料を準備することと、

( b ) 前記出発ゲノム材料の個別の核酸分子を、分離区分がそれぞれ第 1 の個別核酸分子を含有するように、分離区分の中に分配することと、

( c ) 前記分離区分中にある前記個別の核酸分子を断片化して、複数の断片を形成し、断片のそれぞれがバーコードを更に含み、所定の分離区分内の断片が、それぞれ共通のバーコードを含み、それによって、それぞれの断片を、それが誘導された前記個別の核酸分子と関連づけることと、

( d ) 前記ゲノムの前記 1 つ以上の選択部分の少なくとも一部を含む断片が豊富化された個体群を提供することと、

( e ) 前記個体群から配列情報を取得し、それによってゲノムの 1 つ以上の選択部分を配列決定することと

を含む、前記方法。

**【請求項 2】**

前記提供工程 ( d ) が、

( i ) 前記ゲノムの前記 1 つ以上の選択部分の中の、または近くの領域に相補的なプローブを、前記断片とハイブリッド形成させて、プローブ断片複合体を形成するステップと、

( i i ) プローブ断片複合体を固体支持体の表面に捕捉し、それによって、前記ゲノムの前記 1 つ以上の選択部分の少なくとも一部を含む断片を、前記個体群において豊富化するステップとを含む、請求項 1 に記載の方法。

**【請求項 3】**

前記固体支持体がビーズを含む、請求項 2 に記載の方法。

**【請求項 4】**

前記プローブが結合部分を含み、前記表面が捕捉部分を含み、前記プローブ断片複合体が、前記結合部分と前記捕捉部分との反応によって前記表面に捕捉される、請求項 2 ~ 3 に記載の方法。

**【請求項 5】**

前記捕捉部分がストレプトアビジンを含み、前記結合部分がビオチンを含む、請求項 4 に記載の方法。

**【請求項 6】**

前記捕捉部分がストレプトアビジン磁気ビーズを含み、前記結合部分がビオチン化 RNA ライブラリーベイトを含む、請求項 4 に記載の方法。

**【請求項 7】**

前記捕捉部分が、全体的もしくは部分的なエクソームの捕捉、パネルの捕捉、標的化エクソンの捕捉、アンカード ( a n c h o r e d ) エクソームの捕捉及びタイル状 ( t i l e d ) ゲノム領域の捕捉からなる群から選択されるメンバーに向けられている、請求項 4 に記載の方法。

**【請求項 8】**

前記取得ステップ ( e ) の前に、前記断片を増幅して、増幅産物を形成する、請求項 1 ~ 7 に記載の方法。

**【請求項 9】**

前記増幅産物が、部分的または完全なヘアピン構造を形成する能力がある、請求項 8 に記載の方法。

**【請求項 10】**

前記取得ステップ ( e ) が、短い読み込み長さ ( r e a d - l e n g t h ) 配列決定反応及び長い読み込み長さ配列決定反応からなる群から選択される配列決定反応を含む、請

10

20

30

40

50

求項 1 ~ 9 に記載の方法。

【請求項 1 1】

前記配列決定反応が、短い読み込みで高い精度の配列決定反応である、請求項 1 0 に記載の方法。

【請求項 1 2】

前記取得ステップ (e) が、前記出発ゲノム材料の 9 0 % 未満についての配列情報を提供する、請求項 1 ~ 1 1 に記載の方法。

【請求項 1 3】

前記取得ステップ (e) が、前記出発ゲノム材料の 7 5 % 未満についての配列情報を提供する、請求項 1 ~ 1 1 に記載の方法。

10

【請求項 1 4】

前記取得ステップ (e) が、前記出発ゲノム材料の 5 0 % 未満についての配列情報を提供する、請求項 1 ~ 1 1 に記載の方法。

【請求項 1 5】

前記方法が、単離された前記断片の重複配列に基づいて、推定コンティグ中において 2 つ以上の前記個別の核酸分子を連結することを更に含み、前記推定コンティグが、少なくとも 1 0 k b の N 5 0 長さを含む、請求項 1 ~ 1 4 に記載の方法。

【請求項 1 6】

前記推定コンティグが、少なくとも 2 0 k b の N 5 0 長さを含む、請求項 1 5 に記載の方法。

20

【請求項 1 7】

前記推定コンティグが、少なくとも 4 0 k b の N 5 0 長さを含む、請求項 1 5 に記載の方法。

【請求項 1 8】

前記推定コンティグが、少なくとも 5 0 k b の N 5 0 長さを含む、請求項 1 5 に記載の方法。

【請求項 1 9】

前記推定コンティグが、少なくとも 1 0 0 k b の N 5 0 長さを含む、請求項 1 5 に記載の方法。

【請求項 2 0】

前記推定コンティグが、少なくとも 2 0 0 k b の N 5 0 長さを含む、請求項 1 5 に記載の方法。

30

【請求項 2 1】

前記方法が、単離された前記断片の配列内の重複位相変異体に基づいて、位相ブロック中において 2 つ以上の前記個別の核酸分子を連結することを更に含み、前記位相ブロックが、少なくとも 1 0 k b の N 5 0 長さを含む、請求項 1 ~ 1 4 に記載の方法。

【請求項 2 2】

前記位相ブロックが、少なくとも 2 0 k b の N 5 0 長さを含む、請求項 2 1 に記載の方法。

【請求項 2 3】

前記位相ブロックが、少なくとも 4 0 k b の N 5 0 長さを含む、請求項 2 1 に記載の方法。

40

【請求項 2 4】

前記位相ブロックが、少なくとも 5 0 k b の N 5 0 長さを含む、請求項 1 8 に記載の方法。

【請求項 2 5】

前記位相ブロックが、少なくとも 1 0 0 k b の N 5 0 長さを含む、請求項 2 1 に記載の方法。

【請求項 2 6】

前記位相ブロックが、少なくとも 2 0 0 k b の N 5 0 長さを含む、請求項 2 1 に記載の

50

方法。

【請求項 27】

前記ゲノムの前記選択部分が、エクソームを含む、請求項 1 ~ 26 に記載の方法。

【請求項 28】

それぞれの分離区分中の前記個別の核酸分子が、単一細胞のゲノム DNA を含む、請求項 1 ~ 26 に記載の方法。

【請求項 29】

それぞれの分離区分が、異なる染色体のゲノム DNA を含む、請求項 1 ~ 28 に記載の方法。

【請求項 30】

前記分離区分が、エマルションの液滴を含む、請求項 1 ~ 29 に記載の方法。

【請求項 31】

前記断片に結合した前記バーコードが、少なくとも 700,000 個のバーコードのライブラリーからのものである、請求項 1 ~ 30 に記載の方法。

【請求項 32】

前記バーコードが、追加の配列セグメントを更に含む、請求項 1 ~ 31 に記載の方法。

【請求項 33】

前記追加の配列セグメントが、プライマー、結合配列、ランダム  $n$ -mer オリゴヌクレオチド、ウラシル核酸塩基を含むオリゴヌクレオチドからなる群から選択される 1 つ以上のメンバーを含む、請求項 32 に記載の方法。

【請求項 34】

ゲノム試料の 1 つ以上の標的化部分から配列情報を取得する方法であって、

(a) 前記ゲノム試料の個別の第 1 の核酸断片分子を分離区分の中に提供することと、

(b) 前記分離区分内の前記個別の第 1 核酸断片分子を断片化して、前記個別の第 1 の核酸断片分子のそれぞれから複数の第 2 の断片を作り出すことと、

(c) 共通のバーコード配列を、前記複数の第 2 の断片のそれぞれが、それらが含有されている前記分離区分に起因するように、前記分離区分内の前記複数の第 2 の断片に結合することと、

(d) 前記ゲノム試料の前記 1 つ以上の標的化部分に向けられているプローブのライブラリーを、前記第 2 の断片に適用することと、

(e) 配列決定反応を実施して、前記プローブのライブラリーとハイブリッド形成する前記複数の第 2 の断片の配列を特定し、それによって前記ゲノム試料の前記 1 つ以上の標的化部分から配列情報を取得することと

を含む、前記方法。

【請求項 35】

前記プローブのライブラリーを結合部分に結合し、前記実施ステップ (e) の前に、前記第 2 の断片を、捕捉部分を含んだ表面に、前記結合部分と前記捕捉部分の反応を介して捕捉する、請求項 34 に記載の方法。

【請求項 36】

前記実施ステップ (e) の前に、前記第 2 の断片を、前記第 2 の断片が前記表面に捕捉される前後に増幅する、請求項 35 に記載の方法。

【請求項 37】

前記結合部分がビオチンを含み、前記捕捉部分がストレプトアビジンを含む、請求項 35 ~ 36 に記載の方法。

【請求項 38】

前記配列決定反応が、短い読み込み長さ配列決定反応及び長い読み込み長さ配列決定反応からなる群から選択されるメンバーである、請求項 34 ~ 37 に記載の方法。

【請求項 39】

前記配列決定反応が、短い読み込みで高い精度の配列決定反応である、請求項 34 ~ 38 に記載の方法。

10

20

30

40

50

## 【請求項 40】

前記方法が、前記複数の第2断片の重複配列に基づいて、推定コンティグ中において2つ以上の前記個別の核酸分子を連結することを更に含み、前記推定コンティグが、少なくとも10 kbのN50長さを含む、請求項34～39に記載の方法。

## 【請求項 41】

前記推定コンティグが、少なくとも20 kbのN50長さを含む、請求項40に記載の方法。

## 【請求項 42】

前記推定コンティグが、少なくとも40 kbのN50長さを含む、請求項40に記載の方法。

10

## 【請求項 43】

前記推定コンティグが、少なくとも50 kbのN50長さを含む、請求項40に記載の方法。

## 【請求項 44】

前記推定コンティグが、少なくとも100 kbのN50長さを含む、請求項40に記載の方法。

## 【請求項 45】

前記推定コンティグが、少なくとも200 kbのN50長さを含む、請求項40に記載の方法。

## 【請求項 46】

前記方法が、前記複数の第2断片の配列内の重複位相変異体に基づいて、位相ブロック中において2つ以上の複数の前記個別の核酸分子を連結することを更に含み、前記位相ブロックが、少なくとも10 kbのN50長さを含む、請求項34～39に記載の方法。

20

## 【請求項 47】

前記位相ブロックが、少なくとも20 kbのN50長さを含む、請求項46に記載の方法。

## 【請求項 48】

前記位相ブロックが、少なくとも40 kbのN50長さを含む、請求項46に記載の方法。

## 【請求項 49】

前記位相ブロックが、少なくとも50 kbのN50長さを含む、請求項46に記載の方法。

30

## 【請求項 50】

前記位相ブロックが、少なくとも100 kbのN50長さを含む、請求項46に記載の方法。

## 【請求項 51】

前記位相ブロックが、少なくとも200 kbのN50長さを含む、請求項46に記載の方法。

## 【請求項 52】

前記ゲノムの前記標的化部分が、エクソームを含む、請求項34～51に記載の方法。

40

## 【請求項 53】

それぞれの分離区分中の前記ゲノム試料が、単一細胞のゲノムDNAを含む、請求項34～51に記載の方法。

## 【請求項 54】

それぞれの分離区分が、異なる染色体のゲノムDNAを含む、請求項34～51に記載の方法。

## 【請求項 55】

前記分離区分が、エマルジョンの液滴を含む、請求項34～54に記載の方法。

## 【請求項 56】

前記第2断片に結合した前記バーコード配列が、少なくとも700,000個のバー

50

コードのライブラリーからのものである、請求項 34 ~ 55 に記載の方法。

【請求項 57】

前記バーコードが、追加の配列セグメントを更に含む、請求項 34 ~ 56 に記載の方法。

【請求項 58】

前記追加の配列セグメントが、プライマー、結合配列、ランダム  $n$ -mer オリゴヌクレオチド、ウラシル核酸塩基を含むオリゴヌクレオチドからなる群から選択される 1 つ以上のメンバーを含む、請求項 57 に記載の方法。

【請求項 59】

前記第 2 の断片が、得られた増幅産物が部分的または完全なヘアピン構造を形成する能力があるように増幅される、請求項 36 ~ 58 に記載の方法。

【請求項 60】

分子構成を保持しながら、ゲノム試料の 1 つ以上の標的化部分から配列情報を取得する方法であって、

(a) 出発ゲノム材料を準備することと、

(b) 前記出発ゲノム材料の個別の核酸分子を、分離区分がそれぞれ第 1 の個別核酸分子を含有するように、分離区分の中に分配することと、

(c) 前記分離区分中における前記第 1 の個別核酸分子を断片化して、複数の断片を形成することと、

(d) 前記ゲノムの 1 つ以上の選択部分の少なくとも一部を含む断片が豊富化された個体群を提供することと、

(e) 前記個体群から配列情報を取得し、それによって、分子構成を保持しながら、前記ゲノム試料の 1 つ以上の標的化部分を配列決定することを含む、前記方法。

【請求項 61】

前記取得ステップ (e) の前に、前記複数の断片にバーコードをタグ付けして、それぞれの断片を、それらが形成された前記分離区分と関連づける、請求項 60 に記載の方法。

【請求項 62】

ステップ (b) の前記個別の核酸分子が、第 1 の個別の核酸分子のそれぞれの分子構成が維持されるように分配される、請求項 60 に記載の方法。

【請求項 63】

ゲノム試料の 1 つ以上の標的化部分から配列情報を取得する方法であって、

(a) 前記ゲノム試料の個別の核酸断分子を分離区分の中に提供することと、

(b) 前記分離区分中にある前記個別の核酸分子を断片化して、複数の断片を形成し、断片のそれぞれがバーコードを更に含み、所定の分離区分内の断片が、それぞれ共通のバーコードを含み、それによって、それぞれの断片を、それが誘導された前記個別の核酸分子と関連づけることと、

(c) 前記ゲノム試料の前記 1 つ以上の標的化部分に向けられているプローブのライブラリーを、前記複数の断片に適用し、前記プローブのライブラリーにおける少なくとも大部分のプローブが、有益な一塩基多形 (SNP) とハイブリッド形成するように設計されていることと、

(d) 配列決定反応を実施して、前記プローブのライブラリーとハイブリッド形成する前記複数の断片の配列を特定し、それによって前記ゲノム試料の前記 1 つ以上の標的化部分から配列情報を取得することを含む、前記方法。

【請求項 64】

前記プローブのライブラリーにおける約 80 % ~ 99 % のプローブが、有益な SNP とハイブリッド形成するように設計されている、請求項 63 に記載の方法。

【請求項 65】

前記プローブのライブラリーにおける約 65 % ~ 85 % のプローブが、有益な SNP と

10

20

30

40

50

ハイブリッド形成するように設計されている、請求項 63 に記載の方法。

【請求項 66】

前記プローブのライブラリーにおける約 70% ~ 80% のプローブが、有益な SNP とハイブリッド形成するように設計されている、請求項 63 に記載の方法。

【請求項 67】

前記プローブのライブラリーにおける少なくとも 65% のプローブが、有益な SNP とハイブリッド形成するように設計されている、請求項 63 に記載の方法。

【請求項 68】

前記プローブのライブラリーにおける少なくとも 75% のプローブが、有益な SNP とハイブリッド形成するように設計されている、請求項 63 に記載の方法。

10

【請求項 69】

前記プローブのライブラリーにおける少なくとも 85% のプローブが、有益な SNP とハイブリッド形成するように設計されている、請求項 63 に記載の方法。

【請求項 70】

前記プローブのライブラリーにおける少なくとも 90% のプローブが、有益な SNP とハイブリッド形成するように設計されている、請求項 63 に記載の方法。

【請求項 71】

前記有益な SNP が、前記ゲノム試料の前記標的化部分のエクソンとイントロンの両方に位置する、請求項 63 ~ 70 に記載の方法。

【請求項 72】

前記プローブのライブラリーにおける前記大部分のプローブが、約 1 キロベースから約 15 キロベース (kb) の間隔を置いた有益な SNP とハイブリッド形成するように更に設計されている、請求項 63 ~ 70 に記載の方法。

20

【請求項 73】

前記プローブのライブラリーにおける前記大部分のプローブが、約 5 kb ~ 約 10 kb の間隔を置いた有益な SNP とハイブリッド形成するように更に設計されている、請求項 63 ~ 70 に記載の方法。

【請求項 74】

前記プローブのライブラリーにおける前記大部分のプローブが、約 3 kb ~ 約 6 kb の間隔を置いた有益な SNP とハイブリッド形成するように更に設計されている、請求項 63 ~ 70 に記載の方法。

30

【請求項 75】

前記プローブのライブラリーにおける前記大部分のプローブが、約 1 kb の間隔を置いた有益な SNP とハイブリッド形成するように更に設計されている、請求項 63 ~ 70 に記載の方法。

【請求項 76】

前記プローブのライブラリーにおける前記大部分のプローブが、約 3 kb の間隔を置いた有益な SNP とハイブリッド形成するように更に設計されている、請求項 63 ~ 70 に記載の方法。

【請求項 77】

前記プローブのライブラリーにおける前記大部分のプローブが、約 10 kb の間隔を置いた有益な SNP とハイブリッド形成するように更に設計されている、請求項 63 ~ 70 に記載の方法。

40

【請求項 78】

前記プローブのライブラリー内の複数のプローブが、以下の条件の 1 つ以上の任意の組み合わせを満たすように設計され、

(i) エクソンとイントロンの境界の 10 ~ 50 kb の範囲内に有益な SNP がない前記ゲノム試料の標的化部分では、前記複数のプローブが、その境界からのイントロン内の有益な SNP とハイブリッド形成するように設計され、

(ii) エクソン内に第 1 の有益な SNP が存在し、その第 1 の有益な SNP が隣接イ

50

ントロンとの境界から 10 ~ 50 kb に位置し、前記隣接イントロン内に第 2 の有益な SNP が存在し、その第 2 の有益な SNP が前記境界から 10 ~ 50 kb に位置する前記ゲノム試料の標的化部分では、前記複数のプローブが、前記第 1 及び第 2 の有益な SNP の間の前記ゲノム試料の領域とハイブリッド形成するように設計され、

( i i i ) 少なくとも 10 ~ 50 kb に有益な SNP を含まない前記ゲノム試料の標的化部分では、前記複数のプローブが、前記ゲノム試料のこれらの標的化部分と 0 . 5 、 1 、 3 、または 5 kb 毎にハイブリッド形成するように設計され、

( i v ) エクソンとイントロンの境界の 10 ~ 50 kb の範囲内に有益な SNP が不在前記ゲノム試料の標的化部分では、前記複数のプローブが、前記エクソンとイントロンの境界に二番目に近い有益な SNP とハイブリッド形成するように設計される

請求項 63 ~ 77 に記載の方法。

【請求項 79】

前記プローブのライブラリーが、バーコードに連結情報を提供する密度でエクソンに隣接する、前記ゲノム試料の領域と、ハイブリッド形成するように設計されたプローブを含む、請求項 63 ~ 78 に記載の方法。

【請求項 80】

前記プローブのライブラリーにより表される適用範囲が、長い個別の核酸断片分子を高い割合で含有する方法が小さい適用範囲のプローブのライブラリーを使用するように、前記分離区分内の前記ゲノム試料の前記個別の核酸断片分子の長さの分配に反比例する、請求項 63 ~ 79 に記載の方法。

【請求項 81】

前記プローブのライブラリーが、前記ゲノム試料の前記標的化部分の適用範囲を最適化し、前記ゲノム試料の前記標的化部分が、高品質のマッピング範囲を含む、請求項 63 ~ 80 に記載の方法。

【請求項 82】

前記プローブのライブラリーが、ゲノム試料の前記 1 つ以上の標的化部分の特徴により情報提供される特徴を有し、

( i ) 高品質マッピングの標的化部分では、前記プローブのライブラリーが、エクソンとイントロンの境界の 1 kb ~ 1 メガベース ( Mb ) の範囲内の有益な SNP とハイブリッド形成するプローブを含み、

( i i ) バーコード付き前記断片の長さの前記分配が、高い割合で約 250 kb より長い断片を有する標的化部分では、前記プローブのライブラリーが、少なくとも 50 kb 離れている有益な SNP とハイブリッド形成するプローブを含み、

( i i i ) 低品質マッピングの標的化部分では、前記プローブのライブラリーが、エクソンとイントロンの境界の 1 kb 内の有益な SNP とハイブリッド形成するプローブを含み、かつエクソン内及びイントロン内の有益な SNP とハイブリッド形成するプローブを含み、

( i v ) 遺伝子間領域を含む標的化部分では、前記プローブのライブラリーが、少なくとも 2 kb の距離の間隔を置いた有益な SNP とハイブリッド形成するプローブを含む

請求項 63 ~ 80 に記載の方法。

【請求項 85】

ステップ ( d ) で取得した前記配列情報が、遺伝子融合、コピー数変動、挿入及び欠失からなる群の 1 つ以上のメンバーについての情報を含む、請求項 63 ~ 84 に記載の方法。

【請求項 86】

ゲノム試料の 1 つ以上の標的化部分から配列情報を取得する方法であって、

( a ) 前記ゲノム試料の個別の核酸断片分子を準備することと、

( b ) 前記個別の核酸分子を断片化して、複数の断片を形成し、前記断片のそれぞれがバーコードを更に含み、同じ前記個別の核酸分子の断片が共通のバーコードを含み、それによって、それぞれの断片を、それが誘導された前記個別の核酸分子と関連づけることと

10

20

30

40

50



、

(c) 前記複数の断片を、前記ゲノム試料の前記1つ以上の標的化部分を含有する断片で豊富化することと、

(d) 配列決定反応を実施して、豊富化された前記複数の断片の配列を特定し、それによって前記ゲノム試料の前記1つ以上の標的化部分から配列情報を取得することとを含む、前記方法。

【請求項87】

バーコードが、前記断片化ステップ(b)の前に、前記個別の核酸分子に付加される、請求項86に記載の方法。

【請求項88】

前記バーコードが、トランスポソンの使用により、前記個別の核酸分子に付加される、請求項87に記載の方法。

【請求項89】

バーコードが、前記断片化と同時に付加される、請求項86に記載の方法。

【請求項90】

前記断片化が増幅ステップを含む、請求項89に記載の方法。

【請求項91】

前記豊富化ステップ(c)が、前記ゲノム試料の前記1つ以上の標的化部分に向けられたプローブのライブラリーを適用することを含む、請求項86～90に記載の方法。

【請求項92】

前記プローブのライブラリーを結合部分に結合し、前記実施ステップ(d)の前に、前記断片を、前記結合部分と前記捕捉部分の反応を介して捕捉する、請求項91に記載の方法。

【請求項93】

前記結合部分と前記捕捉部分との前記反応が、前記断片を表面に固定化する、請求項92に記載の方法。

【請求項94】

前記結合部分がビオチンを含み、前記捕捉部分がストレプトアビジンを含む、請求項92～93に記載の方法。

【請求項95】

前記配列決定反応が、短い読み込み長さ配列決定反応及び長い読み込み長さ配列決定反応からなる群から選択されるメンバーである、請求項86～94に記載の方法。

【請求項96】

前記配列決定反応が、短い読み込みで高い精度の配列決定反応である、請求項86～94に記載の方法。

【請求項97】

前記断片が、得られた増幅産物が部分的または完全なヘアピン構造を形成する能力があるように増幅される、請求項86～96に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本出願は、米国特許仮出願第62/072,164号、2014年10月29日出願の利益を主張し、これは、全ての目的においてその全体が参照として本明細書に明確に組み込まれる。

【背景技術】

【0002】

ゲノムを正確及び迅速に配列決定することができる能力は、生物学及び医学に革命を起こしている。複雑なゲノムの研究、特にヒトの疾患の遺伝的根拠の探索は、大規模な遺伝子分析を伴う。全ゲノムレベルにおけるそのような遺伝子分析は、金銭的のみならず、時

10

20

30

40

50

間的及び労力的にも代償が大きい。これらの代償は、個別DNA試料の別々の分析を伴うプロトコールによって増加している。疾患発生に関連するゲノムの多形領域の配列決定（及び再配列決定）は、癌などの疾患を理解すること及び治療の開発に大きく寄与し、薬剤応答の多様性に関連する遺伝子及び機能的な多形を特定するという、薬理ゲノミクスの課題に対処する助けとなる。統計的に有意なデータを生じるために十分に大きな個体群において実施される多数の遺伝子マーカーの選別は、関連性が所定の遺伝子型と特定の疾患との間に確立できる前に必要なことである。

#### 【0003】

遺伝子配列決定に関連する代償を低減し、同時に大規模なゲノム分析の利益を得る一つの方法は、ゲノムの標的化領域において高い処理量で高い精度の配列決定を実施することである。広く使用されている手法は、ヒトゲノムの約1%までを構成する、ゲノムのタンパク質全体をコードする領域（エクソーム）の大部分を捕捉し、臨床及び基礎研究において日常的な技術になっている。エクソーム配列決定は、全ゲノム配列決定よりも利点をもたらす。これは、費用が著しく少なく、機能的な解釈において理解が容易であり、分析が著しく速く、非常に深い配列決定（very deep sequencing）が可能になり、管理の容易なデータベースをもたらす。高精度及び高処理量の配列決定、ならびに遺伝子分析のために目的の標的領域を豊富化する方法、系及び組成物の必要性が存在する。

10

#### 【発明の概要】

#### 【0004】

したがって、本発明は、遺伝子の標的化領域についての配列情報を得る方法、系及び組成物を提供する。

20

#### 【0005】

いくつかの態様において、本開示は、ゲノムの1つ以上の選択部分を配列決定する方法であって、一般に、（a）出発ゲノム材料を準備するステップ、（b）個別の核酸分子を、分離区分がそれぞれ第1の個別核酸分子を含有するように、出発ゲノム材料から分配するステップ、（c）分離区分中にある個別の核酸分子を断片化して、複数の断片を形成し、断片のそれぞれがバーコードを更に含み、所定の分離区分内の断片が、それぞれ共通のバーコードを含み、それによって、それぞれの断片を、それが誘導された個別の核酸分子と関連づけるステップ、（d）ゲノムの1つ以上の選択部分の少なくとも一部を含む断片を豊富化した個体群を提供するステップ、（e）個体群から配列情報を取得し、それによってゲノムの1つ以上の選択部分を配列決定するステップを含む方法を提供する。

30

#### 【0006】

更なる実施形態において、上記によると、ゲノムの1つ以上の選択部分の少なくとも一部を含む断片が豊富化された個体群を提供することは、（i）ゲノムの1つ以上の選択部分の中の、または近くの領域に相補的なプローブを、断片とハイブリッド形成させて、プローブ断片複合体を形成するステップ、及び（ii）プローブ断片複合体を固体支持体の表面に捕捉し、それによってゲノムの少なくとも1つ以上の選択部分の少なくとも一部を含む断片を個体群において豊富化するステップを含む。なお更なる実施形態において、固体支持体にはビーズが含まれる。なお更なる実施形態において、プローブは結合部分を含み、表面は捕捉部分を含み、プローブ断片複合体は、結合部分と捕捉部分との反応によって表面に捕捉される。更なる例において、捕捉部分はストレプトアビジンを含み、結合部分はビオチンを含む。なお更なる例において、捕捉部分はストレプトアビジン磁気ビーズを含み、結合部分はビオチン化RNAライブラリーベイトを含む。

40

#### 【0007】

いくつかの実施形態において、上記のうちのいずれかによると、本発明の方法は、全体的もしくは部分的なエクソームの捕捉、パネルの捕捉、標的化エクソンの捕捉、アンカー（anchored）エクソームの捕捉、またはタイル状（tiled）ゲノム領域の捕捉に向けられている、捕捉部分の使用を含む。

#### 【0008】

50

なお更なる実施形態において、上記のうちのいずれかによると、本明細書に開示されている方法は、配列決定反応を含む取得ステップを含む。更なる実施形態において、配列決定反応は、短い読み込み長さ (read-length) 配列決定反応または長い読み込み長さ配列決定反応である。なお更なる例において、配列決定反応は、出発ゲノム材料の 90% 未満、75% 未満、または 50% 未満についての配列情報を提供する。

#### 【0009】

なお更なる実施形態において、本明細書に記載されている方法は、単離断片の重複配列に基づいて、推定コンティグ中において 2 つ以上の個別の核酸分子を連結することを更に含み、推定コンティグは、少なくとも 10 kb、20 kb、40 kb、50 kb、100 kb、または 200 kb の N50 長さを含む。

10

#### 【0010】

なお更なる例において、上記のうちのいずれかによると、本明細書に開示されている方法は、単離断片の配列内の重複位相変異体に基づいて、位相ブロック中において 2 つ以上の個別の核酸分子を連結することを含み、位相ブロックは、少なくとも 10 kb、少なくとも 20 kb、少なくとも 40 kb、少なくとも 50 kb、少なくとも 100 kb、または少なくとも 200 kb の N50 長さを含む。

#### 【0011】

なお更なる実施形態において、上記のうちのいずれかによると、本明細書に開示されている方法は、一緒になってエクソームを網羅するゲノムの選択部分の配列情報を提供する。なお更なる実施形態において、分離区分中の個別の核酸分子は、単一細胞のゲノム DNA を含む。なお更なる実施形態において、分離区分は、それぞれ、異なる染色体のゲノム DNA を含む。

20

#### 【0012】

更なる態様において、本開示は、ゲノム試料の 1 つ以上の標的化部分から配列情報を取得する方法を提供する。そのような方法は、限定されることなく、(a) ゲノム試料の個別の第 1 の核酸断片分子を分離区分の中に提供するステップ、(b) 分離区分内の個別の第 1 核酸断片分子を断片化して、個別の第 1 の核酸断片分子のそれぞれから複数の第 2 の断片を作り出すステップ、(c) 共通のバーコード配列を、複数の第 2 の断片のそれぞれが、それらが含有されている分離区分に起因するように、分離区分内の複数の第 2 の断片に結合するステップ、(d) ゲノム試料の 1 つ以上の標的化部分に向けられているプローブのライブラリーを、第 2 の断片に適用するステップ、(e) 配列決定反応を実施して、プローブのライブラリーとハイブリッド形成する複数の第 2 の断片の配列を特定し、それによってゲノム試料の 1 つ以上の標的化部分から配列情報を取得するステップ、を含む。更なる実施形態において、プローブのライブラリーを結合部分に結合し、実施ステップ (e) の前に、第 2 の断片を、捕捉部分を含んだ表面に、結合部分と捕捉部分の反応を介して捕捉する。なお更なる実施形態において、実施ステップ (e) の前に、第 2 の断片を、第 2 の断片が表面に捕捉される前後に増幅する。なお更なる実施形態において、結合部分はビオチンを含み、捕捉部分はストレプトアビジンを含む。なお更なる実施形態において、配列決定反応は、短い読み込みで高い精度の配列決定反応である。なお更なる実施形態において、第 2 の断片は、得られた産物が部分的または完全なヘアピン構造を形成する能力があるように増幅される。

30

40

#### 【0013】

更なる態様において、上記のうちのいずれかによると、本開示は、分子構成を保持しながら、ゲノム試料の 1 つ以上の標的化部分から配列情報を取得する方法を提供する。そのような方法は、(a) 出発ゲノム材料を準備するステップ、(b) 出発ゲノム材料の個別の核酸分子を、分離区分がそれぞれ第 1 の個別核酸分子を含有するように、分離区分の中に分配するステップ、(c) 分離区分中における第 1 の個別核酸分子を断片化して、複数の断片を形成するステップ、(d) ゲノムの 1 つ以上の選択部分の少なくとも一部を含む断片が豊富化された個体群を提供するステップ、(e) 個体群から配列情報を取得し、それによって、分子構成を保持しながら、ゲノム試料の 1 つ以上の標的化部分を配列決定す

50

るステップ、を含む。更なる実施形態において、取得ステップ(e)の前に、複数の断片にバーコードをタグ付けして、それぞれの断片を、それらが形成された分離区分と関連づける。なお更なる実施形態において、ステップ(b)の個別の核酸分子は、第1の個別の核酸分子のそれぞれの分子構成が維持されるように分配される。

#### 【0014】

いくつかの態様において、本開示は、ゲノム試料の1つ以上の標的化部分から配列情報を取得する方法を提供する。そのような方法は、限定されることなく、(a)ゲノム試料の個別の核酸分子を準備するステップ、(b)個別の核酸分子を断片化して、複数の断片を形成し、断片が、それぞれバーコードを更に含み、同じ個別の核酸分子の断片が、共通のバーコードを有し、それによって、それぞれの断片を、それが誘導された個別の核酸分子と関連づけるステップ、(c)ゲノム試料の1つ以上の標的化部分を含有する断片を、複数の断片において豊富化するステップ、及び(d)配列決定反応を実施して、豊富化された複数の断片の配列を特定し、それによって、ゲノム試料の1つ以上の標的化部分から配列情報を取得するステップ、を含む。更なる実施形態では、ゲノム試料の1つ以上の標的化部分に向けられ得たプローブのライブラリーを適用することを含む、豊富化ステップである。なお更なる実施形態において、プローブのライブラリーを結合部分に結合し、実施ステップの前に、断片を、結合部分と捕捉部分の反応を介して捕捉する。例示的な実施形態において、結合部分と捕捉部分の反応は、断片を表面に固定化する。

#### 【図面の簡単な説明】

#### 【0015】

【図1】本明細書に記載されているプロセス及び系と比べた、従来のプロセスを使用する標的化ゲノム領域の特定及び分析の略図を提供する。

【図2A】本明細書に記載されているプロセス及び系を使用する、標的化ゲノム領域の特定及び分析の略図を提供する。

【図2B】本明細書に記載されているプロセス及び系を使用する、標的化ゲノム領域の特定及び分析の略図を提供する。

【図3】本明細書に記載されている方法及び組成物を使用して配列情報を検出するアッセイを実施する、典型的なワークフローを例示する。

【図4】核酸試料をビーズと組み合わせ、核酸及びビーズを分離液滴に区分するプロセスの略図を提供する。

【図5】染色体核酸断片のバーコード付け及び増幅のプロセスの略図を提供する。

【図6】配列データを個別の染色体に起因させる、染色体核酸断片のバーコード付けの使用についての略図を提供する。

【図7】本発明の方法の一般的実施形態を例示する。

【図8】本発明の方法の一般的実施形態を例示する。

#### 【発明を実施するための形態】

#### 【0016】

本発明の実施は、特に示されない限り、有機化学、ポリマー技術、分子生物学(組み換え技術を含む)、細胞生物学、生化学及び免疫学の従来の技術及び記載を用いることができ、これらは当該技術の技能の範囲内である。そのような従来技術には、ポリマーアレイ合成、ハイブリッド形成、連結、ファージ提示及び標識の使用によるハイブリッド形成の検出が含まれる。適切な技術の特定の例示は、本明細書下記の実施例を参照することによって入手することができる。しかし、他の同等の慣用的な手順も、当然のことながら使用することができる。そのような従来技術及び記載は、Genome Analysis: A Laboratory Manual Series (Vols. I-IV)、Using Antibodies: A Laboratory Manual、Cells: A Laboratory Manual、PCR Primer: A Laboratory Manual及びMolecular Cloning: A Laboratory Manual(全て、Cold Spring Harbor Laboratory Pressのもの)、Stryer, L. (1995) Biochemist

ry (4<sup>th</sup> Ed.) Freeman, New York、Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London、Nelson and Cox (2000), Lehninger, Principles of Biochemistry 3<sup>rd</sup> Ed., W.H. Freeman Pub., New York, N.Y., ならびに Berg et al. (2002) Biochemistry, 5<sup>th</sup> Ed., W.H. Freeman Pub., New York, N.Y. などの標準的な実験マニュアルにおいて見出すことができ、これらは全て、全ての目的において全体が参照として本明細書に組み込まれる。

【0017】

本明細書及び添付の特許請求の範囲で使用されるとき、単数形「a」、「an」及び「the」は、特に文脈により明確に指示されない限り、複数対象を含むことに留意すること。したがって、例えば、「ポリメラーゼ」への参照は、1つの作用物質またはそのような作用物質の混合物を指し、「方法」への参照は、当業者に既知の同等のステップ及び方法への参照を含む。

【0018】

特に定義されない限り、本明細書において使用される全ての技術及び科学用語は、本発明が属する当業者に一般的に理解されるものと同じ意味を有する。本明細書に記述される全ての出版物は、出版物に記載されており、かつ本記載の発明と関連して使用されうる、デバイス、組成物、製剤及び方法を記載及び開示する目的において、参照として本明細書に組み込まれる。

【0019】

ある範囲の値が提供される場合、その範囲の上下限と、その記述された範囲の他の任意の記述された、または中間にある値との、特に文脈により明確に指示されない限り下限の10分の1の単位までのそれぞれの中間値は、本発明の範囲内に包含されることが理解される。小さい範囲に独立して含まれうるこれらの小さい範囲の上下限も、本発明の範囲内に包含され、記述された範囲の任意の明確に除外された限界に制約される。記述された範囲が一方または両方の限界を含む場合、これらの両方の含まれた限界のいずれかを除外する範囲も、本発明に含まれる。

【0020】

以下の記載では、多数の特定の詳細が本発明のさらに十分な理解を提供するために記載されている。しかし、本発明は、これらの特定の詳細の1つ、またはそれ以上を用いることなく実施できることが、当業者には明白である。他の場合において、当業者に周知である周知の特徴及び手順は、本発明を不明瞭にすることを避けるために記載されていない。

【0021】

本明細書で使用されるとき、用語「comprising (含む)」は、組成物及び方法が、列挙された要素を含むが、他を除外しないことを意味することが意図される。「consisting essentially of (から実質的になる)」は、組成物及び方法を定義するために使用されるとき、組成物または方法に対する任意の本質的な有意性のある他の要素を除外することを意味する。「consisting of (からなる)」は、特許請求される組成物の他の成分及び実質的な方法ステップにおける微量を超える要素を除外することを意味する。これらの移行用語のそれぞれにより定義される実施形態は、本発明の範囲内である。したがって、方法及び組成物は、追加のステップ及び構成要素を含むことができる (comprising)、あるいは有意性のないステップ及び組成物を含む (consisting essentially of)、あるいは記述された方法ステップまたは組成物のみが意図される (consisting of) ことが意図される。

【0022】

全ての数値指定、例えば範囲を含む、pH、温度、時間、濃度及び分子量は、0.1の増分により (+) または (-) に変動する近似値である。全ての数値指定には用語「約」

10

20

30

40

50

が先行することが理解されるべきであるが、必ずしも明確に記述されているとは限らない。用語「約」は、「 $X + 0.1$ 」または「 $X - 0.1$ 」などの僅かな増分に加えて、正確な値「 $X$ 」も含む。本明細書に記載されている試薬は、単に例示的であること及びそのようなものの同等物は当該技術において知られていることも理解されるべきであるが、必ずしも明確に記述されているとは限らない。

#### 【0023】

##### I. 概観

本開示は、遺伝子材料の特徴決定に有用な方法、組成物及び系を提供する。特に、本明細書に記載されている方法、組成物及び系は、特定の染色体、染色体の領域、全てのエクソン（エクソーム）、エクソームの一部、特定の遺伝子、遺伝子のパネル（例えば、カイノームもしくは他の標的化遺伝子パネル）、イントロン領域、ゲノムのタイル状部分、またはゲノムの任意の他の選択部分が限定されることなく含まれる、ゲノムの標的化領域の遺伝子特徴決定を提供する。

10

#### 【0024】

一般に、本明細書に記載されている方法及び系は、個別の長い核酸分子の配列の決定、及び／または配列の長い伸展により離されている2つの配列セグメントの間の直接的な分子連結の特定を提供することによって達成され、このことは、長い範囲の配列情報の特定及び使用を可能にするが、この配列決定情報は、極めて低い配列決定誤差及び高処理量の短い読み込み配列決定技術の利点を有する方法の使用によって得られる。本明細書に記載されている方法及び系は、長い核酸分子を、高処理量で高い精度の短い読み込み配列決定技術の使用により配列決定されうる小さい断片に分割し、その分割は、配列情報が元の長い範囲の分子配列構成を保持する小さな断片から誘導されることを可能するように、すなわち、短い配列読み込みが個別の長い核酸分子に由来するように起因することを可能にするように達成される。配列読み込みを、元の長い核酸分子に起因させることによって、短い配列読み込みのみから一般に得ることができない、長い核酸配列についての有意な特徴決定情報を獲得することができる。長い範囲の分子構成は、配列決定プロセスによって保存されるのみならず、本明細書に記載されている標的化配列決定手法に使用される標的化豊富化プロセスによっても保存され、他の配列決定手法はこの能力を示していない。

20

#### 【0025】

一般に、小さい断片の配列情報は、本明細書に記載されており、当該技術に既知のバーコードの付加を含むタグ付け手順の使用を介して、元の長い範囲の分子配列構成を保持する。特定の例において、同じ元の長い個別の核酸分子に由来する断片に、これらの断片の後の任意の配列読み込みが長い個別の核酸分子に由来するように起因するように、共通のバーコードをタグ付けする。そのようなバーコードを、当該技術に既知の任意の方法を使用して付加することができ、個別の核酸分子のセグメントを増幅する増幅方法の際にバーコードを配列に付加すること、ならびにトランスポゾンの使用による元の個別の核酸分子へのバーコードの挿入が含まれ、Amini et al., Nature Genetics 46: 1343-1349 (2014) (2014年10月29日にオンライン先行公開)に記載されたものなどの方法が含まれ、これは、全ての目的において、特に、トランスポゾンの使用によるアダプター及び他のオリゴヌクレオチドの付加に関する全ての教示において、その全体が参照として本明細書に組み込まれる。いったん核酸がそのような方法によりタグ付けされると、得られたタグ付け断片を、本明細書に記載されている方法の使用により、断片の個体群がゲノムの標的化領域を表すように豊富化することができる。このように、その個体群からの配列読み込みは、ゲノムの選択領域の標的化配列決定を可能にし、これらの配列読み込みも、核酸分子に由来するように起因させ、したがって、元の長い範囲の分子配列構成を保存することができる。配列読み込みは、当該技術において既知であり、本明細書に記載されている任意の配列決定方法及びプラットフォームを使用して得ることができる。

30

40

#### 【0026】

ゲノムの標的化領域から配列情報を得る能力を提供することに加えて、本明細書に記載

50

されている方法及び系は、ハプロタイプ位相、構造変動の特定及びコピー数変動の特定が限定されることなく含まれる、ゲノム材料の他の特徴決定も提供することができ、両方とも2015年6月26日に出願された同時係属特許出願のUS N 14 / 752, 589及び14 / 752, 602に記載されており、これらは、全ての目的において、特に、ゲノム材料の特徴決定を対象とする全ての記載、図面及び実施例において、その全体が参照として本明細書に組み込まれる。

#### 【0027】

本出願に記載されている方法及び系により核酸を処理及び配列決定する方法は、US N 14 / 316, 383、14 / 316, 398、14 / 316, 416、14 / 316, 431、14 / 316, 447及び14 / 316, 463においても更に詳細に記載されており、これらは、全ての目的において、特に、核酸の処理及び配列決定、ならびにゲノム材料の他の特徴決定を対象とする全ての記載、図面及び実施例において、その全体が参照として本明細書に組み込まれる。

10

#### 【0028】

一般に、図1に示されているように、本明細書に記載されている方法及び系を使用して、核酸を特徴決定することができる。特に、示されているように、2つの分離された個別の核酸102及び104が例示されており、それぞれ、多数の目的領域、例えば、核酸102に領域106及び108、ならびに核酸104に領域110及び112を有する。それぞれの核酸における目的領域は、同じ核酸分子内に連結されているが、互いに比較的離されていることがあり、例えば、1kbを超える間隔、5kbを超える間隔、10kbを超える間隔、20kbを超える間隔、30kbを超える間隔、40kbを超える間隔、50kbを超える間隔、いくつかの場合では100kbまでの間隔でありうる。この領域は、個別の遺伝子、遺伝子群、エクソン、またはゲノムの単に分離された別個の部分を示すことがある。単に考察の容易さのため、図1に示されている領域は、エクソン106、108、110及び112と呼ばれる。示されているように、核酸102及び104は、それぞれ、それ自体の区分114及び116にそれぞれ分けられている。本明細書の他の部分に示されているように、これらの区分は、多くの場合において油中水エマルションの水性液滴である。それぞれの液滴内において、それぞれの断片の部分は、これらの断片の元の分子構成を保存するように、例えば同じ分子に由来するように、コピーされる。示されているように、これは、由来の断片が区分された液滴を表す、バーコード配列、例えば例示されているバーコード配列「1」または「2」のそれぞれのコピー断片に含めることによって達成される。全ゲノム配列分析用途では、全てのコピー断片及び関連バーコードを単にプールして、それぞれの由来の核酸102及び104から全範囲の配列情報を配列決定し、再会合させることができる。しかし、多くの場合において、ゲノム全体における特定の標的化部分のみ、例えばエクソーム、特定の遺伝子などを分析して、ゲノムの化学的に関連する部分により多く集中すること、ならびにゲノムの関連性の低い、または関連性のない部分の配列決定を実施する時間及び費用を最小限にすることが、より望ましい。

20

30

#### 【0029】

本明細書に記載されている方法によると、標的豊富化ステップをバーコード付き配列断片のライブラリーに適用して、所望の標的に関連する配列を「プルダウン (pull down)」することができる。これらには、エクソン標的化プルダウン、遺伝子パネル特異的標的化プルダウンなどが含まれてもよい。ゲノムの特定の標的化領域の豊富化分離を可能にする多数の標的化プルダウンキットが市販されており、例えば、Agilent SureSelectエクソームプルダウンキットなどである。図1に示されているように、標的化豊富化の適用は、豊富化されたバーコード付き配列ライブラリー118をもたらす。更に、ライブラリー118内のプルダウン断片が、元の分子構成を、例えばバーコード情報の保持によって保持しているので、これらを、埋め込まれた長い範囲の連結情報によって、例えば目的の会合領域それぞれの間の推測連結106:108及び110:112より、元の分子構成に再会合させることができる。例として、ゲノムの2つの異種標的化部分、例えば2つ以上のエクソンの間の直接的な分子連結を特定することができ、そ

40

50

の直接的な分子連結を使用して、構造的変動及び他のゲノムの特徴を特定すること、ならびに2つ以上のエクソンの位相情報を特定すること、例えば、潜在的に全位相エクソーム、またはゲノムの他の位相標的化部分を含む位相エクソンを提供することができる。

#### 【0030】

一般に、本発明の方法は、本明細書において更に詳細に考察される本発明の方法の概略図を提供する、図7に例示されているステップを含む。理解されるように、図9に概説されている方法は、必要に応じて、本明細書に記載されているように、変更及び修正される例示的な実施形態である。

#### 【0031】

図7に示されているように、本明細書に記載されている方法は、大部分の例において、目的の標的領域を含有する試料核酸が区分されるステップ(701)を含む。一般に、それぞれの区分は特定の遺伝子座からの単一の個別核酸分子を含み、これは、通常、含有されている区分に特異的な断片にバーコードを付けることによって(702)、断片の元の分子構成を保存するように断片化またはコピーされる。それぞれの区分は、いくつかの例において、1つを超える核酸を含み、いくつかの場合において、数百の核酸分子を含有し、複数の核酸が区分内にある状態では、ゲノムのあらゆる特定の遺伝子座は、一般に、バーコード付けの前には単一の個別核酸によりに表される。ステップ702のバーコード付き断片は、当該技術に既知の任意の方法を使用して生成することができ、いくつかの例では、オリゴヌクレオチドが分離区分内の試料である。そのようなオリゴヌクレオチドは、試料の異なる多数の領域を無作為に刺激することが意図されるランダム配列を含んでもよく、または試料の標的化領域の上流を刺激するように標的化された特定のプライマー配列を含んでもよい。更なる例において、これらのオリゴヌクレオチドは、複製プロセスも元の試料核酸の得られた複製断片にバーコードを付けるように、バーコード配列も含有する。試料を増幅し、バーコード付けにこれらのバーコードオリゴヌクレオチドを使用する特に洗練されたプロセスは、全て2014年6月26日に出願された米国特許出願第14/316,383号、同第14/316,398号、同第14/316,416号、同第14/316,431号、同第14/316,447号、同第14/316,463号に詳細に記載されており、これらは、それぞれ全ての目的においてその全体が参照として本明細書に組み込まれる。伸長反応試薬、例えば、DNAポリメラーゼ、ヌクレオシド三リン酸、補助因子(例えば、 $Mg^{2+}$ または $Mn^{2+}$ など)も区分中に含有されており、試料をテンプレートとして使用してプライマー配列を伸長し、プライマーがアニールされたテンプレートの鎖に相補的な断片を産生し、相補的な断片は、オリゴヌクレオチド及び関連するバーコード配列を含む。試料の異なる部分への複数のプライマーのアニール及び伸長は、バーコード配列が作り出された区分を示すそれ自体のバーコード配列をそれぞれ有する、試料の重複相補的断片の大きなプールをもたらすことができる。いくつかの場合において、これらの相補的断片は、それ自体を、区分に存在するオリゴヌクレオチドにより刺激されるテンプレートとして使用して、ここでもバーコード配列を含む補体の補体を産生することができる。更なる例において、この複製プロセスは、第1の補体が複製されたとき、2つの相補的配列を末端または末端の近くに産生して、分子が更に反復コピーを産生する基礎となる能力を低減する、ヘアピン構造または部分的ヘアピン構造の形成を可能にする。

#### 【0032】

図7に例示されている方法に戻ると、区分特異的バーコードをコピー断片に結合すると、次のバーコード付き断片をプールする(703)。次に標的豊富化技術を適用して(704)、目的の標的化領域を「プルダウン」する。次にこれらの目的の標的化領域を配列決定し(705)、目的の標的化領域が由来分子構成を特定し、また関連づけられるように、断片の配列を元の分子構成に起因させる(706)。本明細書に記載され、図7に例示されている方法及び系の独自の特徴は、バーコードを、標的豊富化ステップ(704)の前に断片に結合する(702)ことである。本明細書に記載されている方法及び系の利点は、区分または試料特異的バーコードを、標的化ゲノム領域の断片を豊富化する前にコ

10

20

30

40

50



ピー断片に結合することが、標的化領域の元の分子構成を保存し、元の区分、したがって由来試料核酸に起因させることができることである。

【0033】

一般に、標的化ゲノム領域は、チップに基づいた及び溶液に基づいた捕捉方法の両方を使用して、更なる分析、特に配列決定のために豊富化、単離、または分離、すなわち、「ブルダウン」される。そのような方法は、目的のゲノム領域または目的のゲノム領域の近くの、もしくは隣接した領域に相補的であるプローブを利用する。例えば、ハイブリッド（または、チップに基づいた）捕捉では、目的の領域と一緒に網羅する配列を有する捕捉プローブ（通常、一本鎖オリゴヌクレオチド）を含有するマイクロアレイが、表面に固定される。ゲノムDNAは断片化され、平滑末端を産生する末端修復及び/または汎発性刺激配列などの追加の特徴の付加、などの処理を更に受けてもよい。これらの断片を、マイクロアレイにおいてプローブとハイブリッド形成させる。非ハイブリッド形成断片を洗い流し、所望の断片を、配列決定または他の分析のために表面に溶出させ、そうでなければ表面において処理し、したがって表面に残留した断片の個体群は、目的の標的化領域（例えば、捕捉プローブに含有された配列に相補的な配列を含む領域）を含有する断片が豊富化される。断片が豊富化された個体群を、当該技術に既知の任意の増幅技術を使用して更に増幅してもよい。

【0034】

標的化ゲノム領域捕捉の追加的な方法は、ゲノムDNA断片がオリゴヌクレオチドプローブとハイブリッド形成する、溶液に基づいた方法を含む。オリゴヌクレオチドプローブは、多くの場合、「ベイト」と呼ばれる。これらのベイトは、一般に、ビオチン分子が限定されることなく含まれる、捕捉分子に結合される。ベイトは、ゲノムDNA断片に適用されると、ベイトが断片とハイブリッド形成するように、ゲノムの標的化領域（または目的の標的化領域の近く、または隣接した領域）に相補的であり、次に捕捉分子（例えば、ビオチン）を使用して、目的の標的化領域を選択的にブルダウンし（例えば、磁気ストレプトアビジンビーズを用いる）、それにより、得られた個体群を、目的の標的化領域を含有する断片で豊富化する。

【0035】

全エクソームを網羅する標的化領域が必要な例では、全エクソームと一緒に網羅するベイトのライブラリーを使用して、標的化配列を捕捉する。そのような例では、捕捉プロトコールは、Roche/NimbleGen、Illumina及びAgilentにより製造されたエクソーム捕捉プロトコール及びキットのいずれかが限定されることなく含まれる、当該技術に知られている任意のものを含むことができる。

【0036】

本明細書に記載されている方法及び系に使用される標的化ゲノム領域の捕捉は、全エクソームに限定されず、部分エクソーム、遺伝子、遺伝子のパネル、イントロン及びイントロンとエクソンの組み合わせのいずれか1つ、または組み合わせを含むことができる。これらの異なる種類の標的化領域の捕捉手順は、目的の標的化領域を含有する断片をブルダウンするのにベイトを使用する、一般的方法に従っている。ベイトは、特に、目的の標的化領域と、またはその近くとハイブリッド形成するベイトのオリゴヌクレオチドプローブ部分は、捕捉される標的化領域の種類によってある程度決まる。

【0037】

部分的なエクソームのみが更なる分析に必要な例では、ベイトを、エクソームの一部を補足するように設計することができる。特定の例において、必要とされるエクソームの部分の特定の素生は知られており、ベイトのライブラリーは、特定された部分、またはそれらの部分に近い、もしくは隣接した領域に相補的であるオリゴヌクレオチドを含む。そのような例は、障害または疾患などの特定の表現型と関連することが知られている特定の遺伝子及び/もしくは遺伝子のパネル、またはエクソームの特定された部分を捕捉することを、限定されることなく更に含むことができる。いくつかの例において、エクソームの特定の部分または全ゲノム（イントロンとエクソンの領域の両方を含む）が更なる分析に必

要でありうるが、捕捉されるゲノムの部分の特定の配列は、不明である。そのような実施形態において、使用されるベイトは、全ゲノムに向けられたライブラリーのサブセットであり、そのサブセットを、無作為に、またはベイトのライブラリーがゲノムもしくはエクソムの標的化小単位に相補的なプローブのために選択または豊富化されている任意の種類の知的設計を介して、選択することができる。

#### 【0038】

本明細書に記載されている任意の方法のため、標的化領域を、標的化領域の全体もしくは一部に相補的なオリゴヌクレオチドプローブを含むベイトを使用して捕捉することができる、またはオリゴヌクレオチドプローブは、標的領域に近い、もしくは標的領域に隣接する別の領域、例えば、イントロンに相補的でありうる。例えば、図2Aに概略的に例示されているように、遺伝子配列201は、エクソン領域202及び203を含む。これらのエクソン領域は、ベイトを近くの1つ以上のイントロン配列（例えば、エクソン領域202の捕捉にはイントロン領域204及び/または205、エクソン領域203の捕捉にはイントロン領域206）に向けることによって捕捉することができる。換言すると、エクソン領域202または203を含む断片の個体群は、イントロン領域204及び/または205及び206に相補的なベイトの使用によって捕捉することができる、図2Aに示されているように、近くのエクソン領域のイントロンベイトとして使用されるイントロン領域は、目的のエクソン領域に隣接することができ、すなわち、イントロン領域と標的化エクソン領域の間には間隙がない。他の例において、近くのエクソン領域の捕捉に使用されるイントロン領域は、両方の領域が同じ断片内にありうるが、エクソン領域とイントロン領域の間に1つ以上のヌクレオチドの間隙がある（例えば、図2Aの202及び205）ように十分に近くにあってもよい。

10

20

#### 【0039】

いくつかの例では、ベイトを、ゲノムの特定の領域を標的にするように設計するのではなく、タイリング (t i l i n g) 手法が使用される。そのような手法では、特定のエクソンまたはイントロン領域を標的にするのではなく、ベイトは、特定の範囲または距離でゲノムの部分に相補的であるように設計される。例えば、ベイトのライブラリーを、このベイトのライブラリーを断片化試料に適用することが、ゲノムの特定のサブセット、すなわち、ベイトに相補的な配列を含有する断片に含有されている領域のみを捕捉するように、ゲノムに沿って5キロベース (k b) 毎に配列を網羅するように設計することができる。理解されるように、ベイトを、ヒトゲノム基準配列などの基準配列に基づいて設計することができる。更なる例において、タイル状のベイトライブラリーは、ゲノムの1、2、5、10、15、20、25、50、100、200、250、500、750、1000、または10000キロベース毎に領域を捕捉するように設計される。なお更なる例において、タイル状のベイトライブラリーは、距離の混合物を捕捉するように設計され、その混合物は、距離の無作為混合物でありうる、またはゲノムの特定の部分もしくは割合が捕捉されるように知的に設計されうる。理解されるように、そのようなタイリング捕捉方法は、配列決定などの更なる分析のため、イントロンとエクソンの領域を両方とも捕捉する。本明細書に記載されているタイリングまたは他のイントロンベイト方法のいずれかは、長い介在イントロン領域により広く離されているエクソンの配列情報を関連づける方法を提供する。

30

40

#### 【0040】

更なる例において、本明細書に記載されているタイリングまたは他の捕捉方法は、全ゲノムの約5%、10%、15%、20%、30%、40%、50%、60%、70%、80%、90%、95%を捕捉する。なお更なる例において、本明細書に記載されている捕捉方法は、全ゲノムの約1~10%、5~20%、10~30%、15~40%、20~50%、25~60%、30~70%、35~80%、40~90%、または45~95%を捕捉する。

#### 【0041】

いくつかの例において、ゲノムDNAを断片化する、増幅する、区分化する、別の方法

50

で処理する方法を含む試料の調製方法は、ゲノムの特定の領域に偏った、または低い適用範囲をもたらすことができる。そのように偏った、または低い適用範囲は、ゲノムの標的化領域の捕捉に使用されるベイトの濃度またはゲノム位置を変更することによって、本明細書に開示されている方法及び系において補うことができる。いくつかの例において、高いGC含有量または他の構造的変動を含有するゲノムの特定の領域が低い適用範囲をもたらすことが知られている場合があり、そのような状況では、ベイトのライブラリーを変更して、低い適用範囲の領域に向けられるベイトの濃度を増加することができ、換言すると、使用されるベイトの個体群を「スパイク (spiked)」して、低い適用範囲の区域におけるゲノムの標的化領域を含有する十分な数の断片が、配列決定される断片の最終個体群において得られることを確実にすることができる。そのようなベイトのスパイクは、低い適用範囲の領域に向けられたベイトの特別仕様のライブラリーを既製のエクソーム捕捉キットに加えるようにして、市販の全エクソームキットによって実施することができる。加えて、ベイトを、目的の領域に非常に近いが、より多好ましい適用範囲を有する、ゲノム領域を標的にするように設計することができ、このことは、また本明細書において更に詳細に考察され、その実施形態は、図2に概略的に例示されている。

10

20

30

40

50

#### 【0042】

更なる例において、本発明の方法に使用されるベイトのライブラリーは、本明細書に更に記載される特徴の1つ以上を満たす情報価値のある設計 (informed design) の産物である。情報価値のある設計は、ベイトのライブラリーが有益な (informative) 単一ヌクレオチド多形 (SNP) に向けられている場合を含む。用語「有益な SNP」は、本明細書において使用されるとき、ヘテロ接合性である SNP を指す。ベイトのライブラリーは、いくつかの例において、有益な SNP を含有するゲノム試料の領域に向けられている複数のプローブを含有するように設計される。「向けられている」とは、本明細書において使用されるとき、プローブが、SNP を包含する配列に相補的である配列を含有することを意味する。更なる例において、ベイトのライブラリーは、エクソンとイントロンの境界から所定の距離で SNP に向けられているプローブを含有するように設計される。ゲノムの標的化領域が、SNP を欠いている、また非常に僅かな SNP しか含有しない領域を含む場合では、ベイトのライブラリーは、所定の距離でそのような領域にわたってタイル状になる、及び/または二番目に近いイントロンもしくはエクソン内の第1の有益な SNP とハイブリッド形成する、プローブを含む。

#### 【0043】

本明細書に記載されている方法及び系の利点は、捕捉された標的化領域が、標的化領域を捕捉する及び配列決定分析を実施するステップの後であっても、これらの標的化領域の元の分子構成が保持されるように、捕捉の前に処理されることである。本明細書において更に詳細に考察されるように、特定の標的化領域を元の分子構成 (誘導されている元の染色体もしくは染色体領域、及び/または完全ゲノム内の特定の標的化領域の互いの位置が含まれうる) に起因させる能力は、そうでなければ伝統的な配列決定技術の使用により不十分にマップされる、または不十分な適用範囲を有するゲノムの領域から、配列情報を得る方法を提供する。

#### 【0044】

例えば、いくつかの遺伝子は、一般に利用可能な配列決定技術を使用して、特に、長い読み込み技術と比較して優れた精度を有する短い読み込み技術を使用してスパンするには長すぎる長いイントロンを有する。しかし、本明細書に記載されている方法及び系では、一般に図1に例示されており、本明細書において更に詳細に記載されるタグ付け手順を介して、標的化領域の分子構成が保持される。このように、関連性がゲノムの伸長領域にわたって作られる。例えば、図2Bに概略的に例示されているように、核酸分子207は、長いイントロン領域 (208) により中断された2つのエクソン (影付バー) を含有する。一般に使用される配列決定技術は、イントロンを横断する距離をスパン (span) して2つのエクソンの関係に関する情報を提供することができない。本明細書に記載されている方法では、個別の核酸分子207は、それ自体の分離区分209に分配され、次に異

なる断片が異なる部分のエクソン及びイントロンを含有するように断片化される。これらの断片のそれぞれが、断片から得られた任意の配列情報が、断片が生成された分離区分に起因するようにタグ付けされるので、それぞれの断片も、断片が誘導された個別の核酸分子207に起因する。一般に、本明細書において更に記載されるように、断片化及びタグ付けの後、異なる区分からの断片を一緒に合わせる。次に標的化捕捉方法を使用して、配列決定などの更なる分析を受ける断片の個体群を、目的の標的化領域を含有する断片で豊富化することができる。図2Bに例示されている例では、使用されるベイトが断片の個体群を豊富化して、2つのエクソンのうちの1つの一部及び/または介在イントロンの部分を含有するもののみを捕捉するが、エクソン及びイントロンの外側にある領域(例えば、209及び210)は捕捉しない。したがって、配列決定を受ける断片の最終個体群は、目的の2つのエクソンの部分を含有する断片が豊富化される。次に短い読み込みで高い精度の配列決定技術を使用して、この断片豊富化個体群の配列を特定することができ、それぞれの断片がタグ付けされ、したがって、元の分子構成、すなわち、元の個別の核酸分子に起因するので、短い読み込み配列は、長い長さの介在イントロンを超えてスパンする情報を提供して、2つのエクソンの関係についての情報を提供することができる。

10

20

30

40

50

#### 【0045】

上記に示されたように、本明細書に記載されている方法及び系は、長い核酸の短い配列読み込むために、個別の分子構成を提供する。本明細書において使用されるとき、個別の分子構成は、特定の配列読み込みを超えた配列構成、例えば、配列読み込みそれ自体の範囲内に含まれない、隣接または近位配列との関係を指し、このように、典型的には短い配列読み込みの全体または一部に含まれないようなもの、例えば、対読み込みでは約150塩基または約300塩基の読み込みである。特に好ましい態様において、方法及び系は、短い配列読み込みのために長い範囲の配列構成を提供する。そのような長い範囲の構成は、所定の配列読み込みと、互いに1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長い、もしくはさらに100kbより長い、またはそれ以上長い距離内にある配列読み込みとの関係または連結を含む。理解されるように、長い範囲の個別の分子構成を提供することによって、その個別の分子構成内にある変異体の位相情報を導き出すこともでき、例えば、特定の長い分子における変異体は、定義によると一般的に位相されている。

#### 【0046】

長い範囲の個別の分子構成を提供することによって、本発明の方法及び系は、はるかに長い推定分子構成(本明細書において、「長い仮想単一分子読み込み」とも呼ばれる)も提供する。配列構成は、本明細書において記載されているように、異なる(一般にキロベース尺度に基づく)範囲の完全長ゲノム配列にわたる断片のマッピング、または連結の提供を含むことができる。これらの方法は、短い配列読み込みを、個別の長い分子または連鎖分子のコンティグにマッピングすること、ならびに、例えば、個別の分子の連続決定配列を有する長い個別の分子の大きな部分を長い範囲で配列決定することを含み、ここで、そのような決定配列は、1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長い、またはさらには100kbより長い。配列構成によって、短い配列を、長い核酸、例えば個別の長い核酸分子、または連結核酸分子のコレクション、またはコンティグに起因させることは、短い配列を長い核酸伸展にマッピングして、高レベルの配列構成を提供すること、ならびにこれらの長い核酸を介して短い配列から会合配列を提供することを含んでもよい。

#### 【0047】

更に、長い個別の分子と関連する長い範囲の配列構成を利用してもよいが、そのような長い範囲の配列構成を有することは、さらに長い範囲の配列構成を推測することも可能にする。一例として、上記に記載された長い範囲の分子構成を提供することによって、異なる由来分子からの長い配列の中に重複した変異部分、例えば、位相変異体、転位配列など

を特定することができ、これらの分子間の推定連結を可能にする。そのような推定連結または分子構成は、本明細書において「推定コンティグ」と呼ばれる。いくつかの場合において、位相配列の文脈で考察するとき、推定コンティグは、一般に位相された配列を表すことができ、例えば、重複位相変異体によって、個別の由来分子より実質的に大きな長さの位相コンティグを推定することができる。これらの位相コンティグを、本明細書において「位相ブロック」と呼ぶ。

#### 【0048】

長い単一分子読み込み（例えば、上記に考察された「長い仮想単一分子読み込み」）から出発することによって、短い読み込み配列決定技術または位相配列決定の他の手法を使用して得られるものより長い推定コンティグまたは位相ブロックを、導き出すことができる。例えば、米国特許出願公開第2013-0157870号を参照すること。特に、本明細書に記載されている方法及び系を使用して、少なくとも約10 kb、少なくとも約20 kb、少なくとも約50 kbのN50を有する推定コンティグまたは位相ブロック長さを得ることができる（ここで、記述されているN50数値より大きなブロック長さの合計は、全てのブロック長さの合計の50%である）。より好ましい態様において、少なくとも約100 kb、少なくとも約150 kb、少なくとも約200 kb、多くの場合において、少なくとも約250 kb、少なくとも約300 kb、少なくとも約350 kb、少なくとも約400 kb、いくつかの場合において、少なくとも約500 kb、またはそれ以上のN50を有する推定コンティグまたは位相ブロック長さが得られる。なお他の場合において、200 kbを超える、300 kbを超える、400 kbを超える、500 kbを超える、1 Mbを超える、または2 Mbさえもを超える最大位相ブロック長さを得ることもできる。

10

20

#### 【0049】

1つの態様において、本明細書上記及び下記に記載されている捕捉方法のいずれかと共に、本明細書に記載されている方法及び系は、試料核酸またはその断片を分離区画または区分（本明細書において、区分と交換可能に呼ばれる）に区画化する、沈着させる、または区分化することを提供し、ここで、それぞれの区画は、それ自体の内容物を他の区分の内容物から離れた状態に維持している。独自の識別子、例えばバーコードを、区画化または区分化した試料核酸を保持する区分に前もって、すぐ後に、または同時に送達して、特徴を、例えば核酸配列情報を、特定の区画内に含まれる試料核酸に、特に、区画に元々沈着されていてもよい連続試料核酸の比較的長い伸展に、後に起因させることを可能にすることができる。

30

#### 【0050】

本明細書に記載されている方法に利用される試料核酸は、典型的には、分析される試料全体の多数の重複部分、例えば染色体全体、エクソーム、または他の大きな遺伝子部分を表す。これらの試料核酸は、全ゲノム、個別の染色体、エクソーム、アンブリコン、または目的とする様々な異なる核酸のいずれかを含んでもよい。試料核酸は、典型的には、核酸が区分に比較的長い断片または伸展の連続核酸分子で存在するように区分される。典型的には、試料核酸のこれらの断片は、1 kbより長く、5 kbより長く、10 kbより長く、15 kbより長く、20 kbより長く、30 kbより長く、40 kbより長く、50 kbより長く、60 kbより長く、70 kbより長く、80 kbより長く、90 kbより長く、またはさらに100 kbより長くてもよく、上記に記載された長い範囲の分子構成が許容される。

40

#### 【0051】

試料核酸も、所定の区分が非常に低い確率で出発試料核酸の2つの重複断片を含むというレベルで、典型的に区分される。これは、典型的には、試料核酸を区分化プロセスの際に低い投入量及び/または濃度で提供することによって達成される。その結果、好ましい場合では、所定の区分は、出発試料核酸の長いが重複していない多数の断片を含むことができる。次に異なる区分中の試料核酸は、独自の識別子に関連づけられ、任意の所定の区分では、含有された核酸は、同じ独自の識別子を有するが、異なる区分は、異なる独自の

50

識別子を含むことができる。更に、区分化ステップは、試料構成成分を非常に小さい体積の区分または液滴の中に配分するので、上記に記載された所望の配分を達成するため、高体積プロセスに必要とされる試料の相当な希釈を、例えば管またはマルチウエルプレートのウエルにおいて実施する必要がないことが理解される。更に、本明細書に記載されている系が高レベルのバーコード多様性を用いるので、多様なバーコードを、上記に提示されている多数のゲノム等価物に配分することができる。特に、以前に記載されているマルチウエルプレート手法（例えば、米国特許出願公開第2013-0079231号及び同第2013-0157870号を参照すること）は、典型的には、百から数百個の異なるバーコード配列を操作し、試料の限定された希釈プロセスを用いることによってのみ、バーコードを異なる細胞／核酸に起因させることができる。このように、これらは一般に100個よりかなり少ない細胞を用いて操作され、典型的には、1:10、当然のことながら1:100をはるかに超えるオーダーでゲノム：（バーコード型）の比をもたらす。一方、本明細書に記載されている系は、高レベルのバーコード多様性、例えば、10,000、100,000、500,000超などの多様なバーコード型のため、1:50以下、1:100以下、1:1000以下のオーダー、またはさらに小さな比のゲノム：（バーコード型）の比で操作されうるが、ゲノム毎にはるかに改善されたバーコード多様性を依然として提供しながら、より多数のゲノム（例えば、アッセイ毎に100個を超えるゲノム、アッセイ毎に500個を超えるゲノム、アッセイ毎に1000個を超えるゲノム、または更にそれ以上のオーダー）を装填することも可能である。

10

20

30

40

50

#### 【0052】

多くの場合、試料は、区分化ステップの前に、ビーズに遊離可能に結合されるオリゴヌクレオチドタグのセットと組み合わせられる。次にその組み合わせは、当該技術に既知であり、かつ本明細書に記載されている方法の使用により、試料中の核酸のバーコード付けをもたらすことができる。いくつかの例において、増幅方法を使用して、得られた増幅産物にバーコードを付加し、いくつかの例では、増幅産物は、それが誘導される完全由来核酸分子の小さなセグメント（断片）を含有する。いくつかの例において、トランスポゾンを使用する方法は、Amini et al, Nature Genetics 46:1343-1349 (2014) (2014年10月29日にオンライン先行公開)に記載されているように利用されており、これは、全ての目的において、特に、バーコードまたは他のオリゴヌクレオチドタグを核酸に結合することに関する全ての教示において、その全体が参照として本明細書に組み込まれる。更なる例において、バーコードを結合する方法は、ニックング酵素もしくはポリメラーゼ、及び／またはrecAなどの侵襲性プロローブを使用して、二本鎖試料核酸に沿って間隙を生じることを含むことができ、次にバーコードをこれらの間隙に挿入することができる。

#### 【0053】

増幅を使用して核酸断片にタグ付けする例において、オリゴヌクレオチドタグは、少なくとも第1及び第2の領域を含んでもよい。第1の領域は、所定の区分内のオリゴヌクレオチドでは、実質的に同じバーコード配列であってもよいが、異なる区分間では、異なるバーコード配列であってもよく、大部分の場合において異なるバーコード領域である。第2の領域は、区分内の試料内の核酸を刺激するのに使用できるN-mer（ランダムN-merまたは特定の配列を標的にするように設計されたN-mer）でありうる。N-merが特定の配列を標的にするように設計されるいくつかの場合において、特定の染色体（例えば、染色体1、13、18、もしくは21）、または染色体の領域、例えば、エクソームもしくは他の標的化領域を標的にするように設計されうる。いくつかの場合において、N-merは、疾患または障害（例えば、癌）に関連する遺伝子または領域などの特定の遺伝子または遺伝子領域を標的にするように設計されうる。区分内では、増幅反応は、第2のN-merを使用して、核酸の長さに沿って異なる場所で核酸試料を刺激することによって実施してもよい。増幅の結果、それぞれの区分は、同一またはほぼ同一のバーコードに結合している核酸の増幅産物を含有することができ、これは、それぞれの区分中の核酸の重複している小さい断片を表すことがある。バーコードは、同じ区分に由来し、し

たがって潜在的に同じ核酸鎖にも由来する核酸のセットを示すマーカーとして、機能することができる。増幅の後、核酸をプールし、配列決定し、配列決定アルゴリズムの使用により整列させることができる。短い配列読み込みが、関連するバーコード配列によって整列され、試料核酸の単一の長い断片に起因しうるので、配列において特定された変異体は、全て、単一の由来断片及び単一の由来染色体に起因することができる。更に、複数の共同配置変異体を複数の長い断片にわたって配列させることによって、染色体の寄与について更に特徴決定することができる。したがって、長い範囲の遺伝子配列にわたって分析することができるので、例えば、ゲノムの不十分に特徴決定された領域の伸展にわたって配列情報を特定することができるので、特定の遺伝子変異体の位相に関する結論を導き出すことができる。そのような情報は、一般に同じ核酸鎖または異なる核酸鎖に存在する特定の遺伝子変異体セットである、ハロタイプの特定にも有用でありうる。コピー数変動も、この方法によって特定することができる。

10

#### 【0054】

記載されている方法及び系は、現行の核酸配列決定技術及び関連する試料調製方法よりも大きな利点を提供する。試料調製及び配列決定方法は、全体として、主に、試料中の大多数の構成要素を特定し、特徴決定することに傾いており、抽出試料中の総DNAにおいて少ない比率を構成する少数の構成要素、例えば、1つの染色体または1個もしくは数個の細胞が寄与する遺伝子材料、または血流中に循環する断片化腫瘍細胞DNA分子を特定及び特徴決定するように設計されていない。記載されている方法及び系は、大型試料内に存在する個体群を検出するのに大きな利点も提供する。このように、ハロタイプ及びコピー数変動の評価に特に有用であり、本明細書に開示されている方法は、不十分に特徴決定されている、または試料調製の際に導入された偏向に起因して、核酸標的の個体群によって不十分に表されているゲノムの領域に、配列情報を提供することにも有用である。

20

#### 【0055】

本明細書に開示されているバーコード付け技術の使用は、所定の遺伝子マーカーのセットに個別の分子構成を提供する、すなわち、所定の遺伝子マーカーnセットを（単一のマーカーとは対照的に）個別の試料核酸分子に起因させる、ならびに変異体座標会合を介して、より広範囲の、さらには長い範囲の推定個別分子構成を、複数の試料核酸分子及び/または特定の染色体に提供する、独自の能力を付与する。これらの遺伝子マーカーには、特定の遺伝子座、例えば、SNPなどの変異体を含んでもよく、または短い配列を含んでもよい。更に、バーコード付けの使用は、例えば、血流中の循環腫瘍DNAの検出及び特徴決定のため、試料から抽出された総核酸個体群において小数の構成成分と大多数の構成成分とを区別する能力を促進する追加的な利点を付与し、また任意の増幅ステップの際に増幅偏向を低減または排除する。加えて、マイクロ流体工学フォーマットの実施は、極めて小さい試料体積及び低投入量のDNAを扱う能力、ならびに多数の試料区分（液滴）を迅速に処理して、ゲノム規模のタグ付けを促進する能力を付与する。

30

#### 【0056】

前記載のように、本明細書に記載されている方法及び系の利点は、広範に入手可能な短い読み込み配列決定技術の使用を介して所望の結果を達成できることである。そのような技術は、十分に特徴づけられている極めて効果的なプロトコール及び試薬系によって、研究界において容易に入手可能であり、広く普及しているという利点を有する。これらの短い読み込み配列決定技術には、例えば、Illumina, inc. (GXII、NextSeq、MiSeq、HiSeq、X10)、Ion Torrent division of Thermo-Fisher (Ion Proton及びIon PGM)、パイロシーケンシング (pyrosequencing) の方法、ならびに他から入手可能なものが含まれる。

40

#### 【0057】

本明細書に記載されている方法及び系がこれらの短い読み込み配列決定技術を利用し、関連する誤差率が低いことが、特に利点である。特に、本明細書に記載されている方法及び系は、上記に記載された所望の個別の分子読み込み長さ、または構成を達成するが、メ

50

イトペア (mate pair) 伸展を除いて、1000bpより短い、500bpより短い、300bpより短い、200bpより短い、150bpより短い、またはさらに短い個別の配列決定読み込みを伴い、5%未満、1%未満、0.5%未満、0.1%未満、0.05%未満、0.01%未満、0.005%未満、またはさらには0.001%未満の個別の分子読み取り長さの配列決定誤差率を伴って達成する。

#### 【0058】

##### II. ワークフロー概観

1つの例示的な態様において、本開示に記載されている方法及び系は、個別の試料（例えば、核酸）を分離区分に沈着または区分化することを提供し、それぞれの区分は、それ自体の内容物を他の区分の内容物から離れた状態に維持している。本明細書において使用されるとき、区分とは、様々な異なる形態、例えば、ウェル、管、マイクロまたはナノウェル、中空などの容器または器を指す。しかし好ましい態様では、区分は流体流の中で流動可能である。これらの器は、例えば、内側流体中心またはコアを囲む外側隔壁を有するマイクロカプセルまたは微小胞から構成されていてもよく、あるいは、マトリックス内に材料を導入させる及び/または保持することができる多孔質マトリックスであってもよい。しかし好ましい態様では、これらの区分は、非水性連続相、例えば油相内の水性流体の液滴を構成しうる。様々な異なる器は、例えば、米国特許出願第13/966,150号、2013年8月13日出願に記載されている。同様に、非水性または油連続相中に安定した液滴を作り出すエマルジョン系は、例えば米国特許出願公開第2010-0105112号に詳細に記載されている。特定の場合において、マイクロ流体チャンネル網状組織は、本明細書に記載されている区分の生成に特に適している。そのようなマイクロ流体デバイスには、米国特許出願第14/682,952号、2015年4月9日出願において詳細に記載されたものが含まれ、この全開示は、全ての目的においてその全体が参照として本明細書に組み込まれる。細胞の水性混合物がその中を通して非水性流体中に抽出される多孔質膜を含む、代替的な機構を個別の細胞の区分化に用いることもできる。そのような系は、例えば、Nanomi, Inc から一般に入手可能である。

#### 【0059】

エマルジョンの液滴の場合、試料材料、例えば、核酸を分離区分に区分化することは、一般に、試料含有水性流を、区分化流体の非水性流、例えばフッ素化油も流れている接合部に、水性液滴が、流れている区分流体流の中に作り出されるように流すことによって達成することができ、ここで、そのような液滴は試料材料を含む。下記に記載されるように、区分、例えば液滴は、同時に区分化されたバーコードオリゴヌクレオチドも典型的に含む。任意の特定の区分内の試料材料の相対量は、例えば、水性流中の試料の濃度、水性流及び/または非水性流の流速などが含まれる系の様々な異なるパラメーターを制御することによって、調整することができる。本明細書に記載されている区分は、多くの場合、極めて小さい体積を有することによって特徴づけられる。例えば、液滴に基づいた区分の場合では、液滴は、1000pL未満、900pL未満、800pL未満、700pL未満、600pL未満、500pL未満、400pL未満、300pL未満、200pL未満、100pL未満、50pL未満、20pL未満、10pL未満、またはさらに1pL未満の総体積を有することがある。ビーズと同時区分化される場合、区分内の試料流体体積は、上記に記載された体積の90%未満、上記に記載された体積の80%未満、70%未満、60%未満、50%未満、40%未満、30%未満、20%未満、またはさらには10%未満でありうる。いくつかの場合では、低反応体積区分の使用は、非常に少量の出発試薬、例えば、投入核酸によって反応を実施するのに特に有利である。低投入量核酸により試料を分析する方法及び系は、米国特許出願第14/752,602号、2015年6月26日出願に表されており、その全開示は、その全体が参照として本明細書に組み込まれる。

#### 【0060】

試料が、本明細書に記載されている方法及び系によって、それぞれの区分に導入されると、区分内の試料核酸は、これらの核酸が特徴決定されたときに、それぞれの起源から誘



導されたように起因されうるように、一般に独自の識別子が提供される。したがって、試料核酸は、典型的には独自の識別子（例えば、バーコード配列）と共に同時区分化される。特に好ましい態様において、独自の識別子は、試料に結合していてもよい核酸バーコード配列を含むオリゴヌクレオチドの形態で提供される。オリゴヌクレオチドは、所定の区分内のオリゴヌクレオチドでは、含有された核酸バーコード配列は同じであるが、異なる区分では、オリゴヌクレオチドはバーコード配列が異なりうるように、好ましくは異なるバーコード配列を有するように、区分化される。好ましい態様において、1つの核酸バーコード配列のみが所定の区分と関連するが、いくつかの場合では、2つ以上の異なるバーコード配列が存在してもよい。

#### 【0061】

核酸バーコード配列は、典型的には6～約20個以上のヌクレオチドをオリゴヌクレオチドの配列内に含む。これらのヌクレオチドは、完全に連続的、すなわち、隣接ヌクレオチドの単一伸展であってもよく、または1個以上のヌクレオチドによって離されている、2つ以上の別々の部分配列に離されていてもよい。典型的には、離された部分配列は、典型的には約4～約16個のヌクレオチドの長さでありうる。

#### 【0062】

同時区分化されたオリゴヌクレオチドは、区分化された核酸の処理に有用な他の機能性配列も典型的に含む。これらの配列には、例えば、関連するバーコード配列を結合しながら、区分内の個別の核酸からゲノムDNAを増幅する標的化もしくはランダム/汎発性増幅プライマー配列、配列決定プライマー、例えば、配列の存在の特定のため、もしくはバーコード付き核酸のプルダウンのためのハイブリッド形成もしくはプローブ配列、または任意の多数の他の潜在的な機能性配列が含まれる。ここでも、オリゴヌクレオチド及び関連するバーコード、ならびに他の機能性配列の同時区分化は、試料材料と共に、例えば、全て2014年6月26日出願の米国特許出願第14/316,383号、同第14/316,398号、同第14/316,416号、同第14/316,431号、同第14/316,447号、同第14/316,463号、ならびに2014年2月7日出願の米国特許出願第14/175,935号に記載されており、これらの全開示は、その全体が参照として本明細書に組み込まれる。

#### 【0063】

簡潔には、1つの例示的なプロセスにおいて、それぞれ多数の、ビーズに遊離可能に結合し上記に記載されたオリゴヌクレオチドを含みうるビーズが提供され、特定のビーズに結合したオリゴヌクレオチドは、全て、同じ核酸バーコード配列を含んでもよいが、多数の多様なバーコード配列が、使用されるビーズの個体群にわたって表されていてもよい。典型的には、ビーズの個体群は、少なくとも1000個の異なるバーコード配列、少なくとも10,000個の異なるバーコード配列、少なくとも100,000個の異なるバーコード配列、またはいくつかの場合では、少なくとも1,000,000個の異なるバーコード配列を含みうる多様なバーコード配列ライブラリーを提供してもよい。加えて、それぞれのビーズは、典型的には、結合された多数のオリゴヌクレオチド部分を備えていることができる。特に、個別のビーズにおいてバーコード配列を含むオリゴヌクレオチド分子の数は、少なくとも約10,000個のオリゴヌクレオチド分子、少なくとも100,000個のオリゴヌクレオチド分子、少なくとも1,000,000個のオリゴヌクレオチド分子、少なくとも10億個のオリゴヌクレオチド分子でありうる。

#### 【0064】

オリゴヌクレオチドは、ビーズへの特定の刺激の適用によって、ビーズから遊離可能でありうる。いくつかの場合において、刺激は、例えば、オリゴヌクレオチドを遊離しうる光解離性連結の切断を介した、光刺激でありうる。いくつかの場合では、熱刺激を使用してもよく、ビーズ環境の温度の上昇は、連結の切断またはビーズからのオリゴヌクレオチドの他の遊離をもたらすことができる。いくつかの場合では、ビーズへのオリゴヌクレオチドの連結を切断する、または別の方法でビーズからのオリゴヌクレオチドの遊離をもた

10

20

30

40

50

らすことができる、化学刺激を使用してもよい。

【0065】

本明細書に記載されている方法及び系によると、結合されたオリゴヌクレオチドを含むビーズは、単一ビーズ及び単一試料が個別の区分内に含有されるように、個別の試料と共に同時区分化されうる。単一ビーズ区分が望ましい、いくつかの場合において、平均して、区分が1区分あたり1個未満のビーズを含有して、占有されたこれらの区分が主に単一占有されることを確実にするように、流体の相対流速を制御することが望ましいことがある。同様に、流速を制御して、区分が占有される高い率をもたらす、例えば、僅かな率の非占有区分のみを可能にすることが望ましい場合もある。好ましい態様において、流れ及びチャンネル構造は、望ましい数の単一占有区分が、非占有区分の特定のレベルより少なく、複数占有区分の特定のレベルより少ないことを確実にするように制御される。

10

【0066】

図3は、特にコピー数変動またはハロタイプアッセイに使用するため、試料核酸をバーコード付け及び続く配列決定する1つの特定例の方法を例示する。最初に、核酸を含む試料を供給源300から得ることができ、バーコード付きビーズのセットも、得ることができる(310)。ビーズは、好ましくは、1つ以上のバーコード配列を含有するオリゴヌクレオチドと、ならびにランダムN-merなどのプライマーまたは他のプライマーと連結している。好ましくは、バーコード配列は、例えば、バーコードとビーズとの連結の切断を介して、もしくは下にあるビーズの分解によるバーコードの遊離を介して、または2つの組み合わせを介して、バーコード付きビーズから遊離される。例えば、特定の好ましい態様において、バーコード付きビーズは、還元剤なその作用物質により分解または溶解されて、バーコード配列を遊離することができる。この例において、核酸を含む低量の試料(305)、バーコード付きビーズ(315)及び任意に他の試薬、例えば還元剤(320)を組み合わせ、区分化に付す。例として、そのような区分化は、構成成分をマイクロ流体デバイス(325)などの液滴生成系に導入することを伴うことができる。マイクロ流体デバイス(325)の助けを借りて、油中水エマルション(330)を形成することができ、エマルションは、試料核酸(305)、還元剤(320)及びバーコード付きビーズ(315)を含有する水性液滴を含有する。還元剤は、バーコード付きビーズを溶解または分解し、それによって、バーコード及びランダムN-merを有するオリゴヌクレオチドを、液滴(335)内でビーズから遊離させることができる。次にランダムN-merは、試料核酸の異なる領域を刺激して、増幅後に試料増幅コピーをもたらし、それぞれのコピーには、バーコード配列がタグ付けされる(340)。好ましくは、それぞれの液滴は、同一のバーコード配列及び異なるランダムN-mer配列を含有するオリゴヌクレオチドのセットを含有する。続いて、エマルションが壊され(345)、追加の配列(例えば、特定の配列決定方法を助ける配列、追加のバーコードなど)を、例えば、増幅法(350)(例えば、PCR)を介して付与することができる。次に配列決定を実施し(355)、アルゴリズムを適用して、配列決定データを解釈する(360)ことができる。配列決定アルゴリズムは、一般に、例えばバーコードの分析を実施して配列読み込みを整列すること、及び/または特定の配列読み込みが属する試料を特定することができる。加えて、本明細書に記載されているように、これらのアルゴリズムを更に使用して、コピーの配列を由来分子構成に起因させることもできる。

20

30

40

【0067】

上記に示されたように、単一ビーズ占有が最も望ましい状態でありうるが、複数占有区分または非占有区分が多くの場合に存在しうるということが理解される。試料及びバーコードオリゴヌクレオチドを含むミーズを同時区分化するマイクロ流体チャンネル構造の例が、図4に概略的に例示されている。示されているように、チャンネルセグメント402、404、406、408及び410は、チャンネル接合部412において流体連通するように提示されている。個別の試料414を含む水性流は、チャンネルセグメント402の中をチャンネル接合部412に向かって流れる。本明細書の他に記載されているように、これらの試料は、区分化プロセスの前には水性流体内に浮遊されていてもよい。

50

## 【0068】

同時に、バーコード担持ビーズ416を含む水性流は、チャンネルセグメント404の中をチャンネル接合部412に向かって流れる。非水性区分流体が、側面チャンネル406及び408のそれぞれからチャンネル接合部412に導入され、合わされた流は、出口チャンネル410に流れる。チャンネル接合部412内では、チャンネルセグメント402及び404からの2つの合わされた水性流が合わされ、同時区分化された試料414及びビーズ416を含む液滴418に区分化される。前記に示されているように、チャンネル接合部412において合わされるそれぞれの流体の流れ特徴を制御することによって、ならびにチャンネル接合部の形状を制御することによって、組み合わせ及び区分化を最適化して、生成される区分418内のビーズ、試料、またはその両方の所望の占有レベルを達成することができる。

10

## 【0069】

理解されるように、例えば、化学刺激、核酸伸長、転写、ならびに/またはポリメラーゼ、逆転写酵素、ヌクレオシ三リン酸もしくはNTP類縁体などの増幅試薬、このような反応に使用される二価金属イオンなどのプライマー配列及び追加の補助因子、リガーゼ酵素及び連結配列などの連結反応試薬、色素、標識、または他のタグ付け試薬を含む多数の他の試薬を、試料及びビーズと共に同時区分化することができる。

## 【0070】

いったん同時区分化されると、ビーズに配置されたオリゴヌクレオチドを使用して、区分化試料をバーコード付けし、増幅することができる。試料を増幅し、バーコード付けにこれらのバーコードオリゴヌクレオチドを使用する特に洗練されたプロセスは、全て2014年6月26日に出願された米国特許出願第14/316,383号、同第14/316,398号、同第14/316,416号、同第14/316,431号、同第14/316,447号、同第14/316,463号に詳細に記載されており、これらの全開示は、その全体が参照として本明細書に組み込まれる。簡潔には、1つの態様において、オリゴヌクレオチドは、試料と共に同時区分化されるビーズに提示され、ビーズから、試料を有する区分内に遊離される。オリゴヌクレオチドは、典型的には、バーコード配列と共に、5'末端にプライマー配列を含む。このプライマー配列は、ランダムでありうる、または構造化されうる。ランダムプライマー配列は、一般に、試料の多数字野異なる領域を無作為に刺激することが意図される。構造化プライマーには、試料の特定の標的化領域の上流を刺激するように標的化された確定配列を含む一連の異なる構造体、ならびにある率の特定の塩基(例えば、ある率のGC N-mer)を含有するプライマー、部分的もしくは全体的に縮重した配列を含有するプライマー、及び/または部分的にランダムであり、本明細書のいずれかの記載に従って部分的に構造化されている配列を含有するプライマーが限定されることなく含まれる、ある種の部分的に確定された構造を有するプライマー、が含まれうる。理解されるように、1つ以上の上記の種類のランダム及び構造化プライマーを、任意の組み合わせでオリゴヌクレオチドに含めることができる。

20

30

## 【0071】

いったん遊離されると、オリゴヌクレオチドのプライマー部分は、試料の相補的領域にアニールすることができる。伸長反応試薬、例えば、DNAポリメラーゼ、ヌクレオシ三リン酸、補助因子(例えば、Mg<sup>2+</sup>またはMn<sup>2+</sup>など)も試料及びビーズと共に同時区分化されており、試料をテンプレートとして使用してプライマー配列を伸長し、プライマーがアニールされたテンプレートの鎖に相補的な断片を産生し、相補的な断片は、オリゴヌクレオチド及び関連するバーコード配列を含む。試料の異なる部分への複数のプライマーのアニリング及び伸長は、バーコード配列が作り出された区分を示すそれ自体のバーコード配列をそれぞれ有する、試料の重複相補的断片の大きなプールをもたらすことができる。いくつかの場合において、これらの相補的断片は、それ自体を、区分に存在するオリゴヌクレオチドにより刺激されるテンプレートとして使用して、ここでもバーコード配列を含む補体の補体を産生することができる。いくつかの場合において、この複製プロセスは、第1の補体が複製されたとき、2つの相補的配列を末端または末端の近くに産

40

50

生して、分子が更に反復コピーを産生する基礎となる能力を低減する、ヘアピン構造または部分的ヘアピン構造の形成を可能にする。これの一例の概略的例示が、図5に示されている。

#### 【0072】

図に示されているように、バーコード配列を含むオリゴヌクレオチドは、例えば、エマルジョンの液滴502の中に、試料核酸504と共に同時区分化されている。本明細書の他に示されているように、オリゴヌクレオチド508は、試料核酸504と共に同時区分化されているビーズ506に提供されていてもよく、オリゴヌクレオチドは、パネルAに示されているように、ビーズ506から好ましくは遊離可能である。オリゴヌクレオチド508は、バーコード配列512、1つ以上の機能性配列に加えて、例えば、配列510、514及び516を含む。例えば、オリゴヌクレオチド508は、バーコード配列512、ならびに所定の配列決定系において結合または固定化配列として機能しうる配列510、例えば、Illumina HiSeqまたはMiSeq系のフローセルにおいて結合として使用されるP5配列を含むように示されている。示されているように、オリゴヌクレオチドはプライマー配列516を含み、これは、試料核酸504の部分の複製を刺激するランダムまたは標的化N-merを含んでもよい。配列決定系における合成反応によるポリメラーゼ媒介テンプレート指向配列決定の刺激に使用される、「リード1」またはR1刺激領域などの配列決定刺激領域を提供することができる配列514も、オリゴヌクレオチド508に含まれる。多くの場合において、バーコード配列512は、固定化配列510及びR1配列514は、所定のビーズに結合された全てのオリゴヌクレオチドに共通のものでありうる。プライマー配列516は、ランダムN-merプライマーに変わりうる、または特定の標的化用途において、所定のビーズのオリゴヌクレオチドでは共通のものでありうる。

#### 【0073】

プライマー配列516の存在に基づいて、オリゴヌクレオチドは、パネルBに示されているように試料核酸を刺激することができ、このことは、ビーズ506及び試料核酸504と共に同時分割されたポリメラーゼ酵素及び他の伸長試薬の使用によるオリゴヌクレオチド508及び508aの伸長を可能にする。パネルCに示されているように、ランダムN-merプライマーでは、試料核酸504の複数の異なる領域にアニールするオリゴヌクレオチドの伸長の後、核酸の複数の重複補体または断片が作り出され、例えば、断片518及び520である。試料核酸の部分に相補的な配列部分、例えば配列522及び524を含んでいるが、これらの構造体は、結合バーコード配列を有する試料核酸504の断片を含むものと、本明細書において一般に呼ばれる。理解されるように、上記に記載されたテンプレート配列の複製部分は、そのテンプレート配列の「断片」と本明細書において多くの場合に呼ばれる。しかし前述にもかかわらず、用語「断片」は、例えば、酵素的、化学的または機械的断片化を介した所定の分子配列の実際の断片化などの、テンプレート配列の部分を提供する他の機構によって作り出されたものを含む、由来核酸配列の任意の提示部分、例えばテンプレートまたは試料核酸を包含する。しかし好ましい態様において、テンプレートまたは試料核酸配列の断片は、基礎配列またはその補体の複製部分を示す。

#### 【0074】

次にバーコード付き核酸断片を、例えば、配列分析を介して特徴決定に付すことができる、またはパネルDに示されているように、プロセスによって更に増幅することができる。例えば、追加のオリゴヌクレオチド、例えば、オリゴヌクレオチド508bも、ビーズ306から遊離されており、断片518及び520を刺激することができる。特に、ここでも、オリゴヌクレオチド508bにおけるランダムN-merプライマー516bの存在に基づいて（多くの場合では、所定の区分内の他のランダムN-mer、例えばプライマー配列516と異なる）、オリゴヌクレオチドは断片518とアニールし、伸長して、同じ核酸配列の一部の複製を含む配列528を含む断片518の少なくとも一部まで、補体526作り出す。オリゴヌクレオチド508bの伸長は、断片518のオリゴヌクレオ

チド部分 508 を介して複製されるまで続く。本明細書の他に示されており、パネル D に例示されているように、オリゴヌクレオチドは、例えば断片 518 に含まれるオリゴヌクレオチド 508 の配列 516 及び 514 を介する複製の後、所望の地点でポリメラーゼの複製に終止を促すように形成されうる。本明細書に記載されているように、このことは、例えば、使用されるポリメラーゼ酵素により処理されない異なるヌクレオチド及び/またはヌクレオチド類縁体の組み込みを含む、異なる方法によって達成することができる。例えば、これには、ヌクレオチドを含有するウラシルを配列領域 512 内に含めて、非ウラシル耐性ポリメラーゼがその領域の複製を止めることを防止することが含まれうる。その結果、一方の末端に完全長オリゴヌクレオチド 508 b を含み、バーコード配列 512、結合配列 510、R1プライマー領域 514 及びランダム N-mer 配列 516 b を含む断片 526 が作り出される。配列の他方の末端には、第 1 のオリゴヌクレオチド 508 のランダム N-mer の補体 516'、ならびに R1 配列の全てまたは一部の補体が配列 514' と示されて含まれる。R1 配列 514 及びその補体 514' を一緒にハイブリッド形成させて、部分ヘアピン構造 528 を形成することができる。理解されるように、ランダム N-mer が、異なるオリゴヌクレオチドにおいて異なるので、これらの配列及びそれらの補体は、ヘアピン形成に参加しないことが予測され、例えば、ランダム N-mer 516 の補体である配列 516' は、ランダム N-mer 配列 516 b の補体であると予測されない。このことは、他の用途には、例えば、N-mer が所定の区分内のオリゴヌクレオチドに共通得している標的化プライマーには当てはまらない。

10

20

#### 【0075】

これらの部分ヘアピン構造を形成することによって、試料配列の第 1 レベルの複製を更なる複製から排除すること、例えば、コピーの反復複製の防止を可能にする。部分ヘアピン構造は、作り出された断片、例えば、断片 526 の続く処理にとって有用な構造も提供する。

#### 【0076】

次に複数の異なる区分から全ての断片を、本明細書に記載されている高処理量配列決定装置による配列決定のためにプールすることができる。それぞれの断片が、その起源区分に対してコードされているので、その断片の配列を、バーコードの存在に基づいてその起源に起因させることができる。これは、図 6 に概略的に例示されている。一例として示されているように、第 1 の供給源 600 (例えば、個別の染色体、核酸の鎖など) に結いライする核酸 604 及び異なる染色体 602 または核酸の鎖から誘導される核酸 606 は、それぞれ、上記に記載されたバーコードオリゴヌクレオチドのセットと共に区分化される。

30

#### 【0077】

それぞれの区分内において、それぞれの核酸 604 及び 606 は処理されて、第 1 の断片 (複数可) の第 2 の断片の重複セット、例えば、第 2 の断片セット 608 及び 610 をもたらす。またこの処理は、特定の第 1 の断片から誘導される第 2 の断片のそれぞれと同じであるバーコード配列を有する第 2 の断片をもたらす。示されているように、第 2 の断片セット 608 のバーコード配列は、「1」と示されており、断片セット 610 のバーコード配列は、「2」と示されている。多様なバーコードライブラリーを使用して、多数の異なる断片セットを異なってバーコード付けすることができる。しかし、異なる第 1 の断片からの全ての第 2 の断片セットを、異なるバーコード配列でバーコード付けする必要はない。事実、多くの場合において、複数の異なる第 1 の断片は、同じバーコード配列を含むように同時に処理されることがある。多様なバーコードライブラリーは、本明細書の他に詳細に記載されている。

40

#### 【0078】

次に、例えば、断片セット 608 及び 610 のバーコード付き断片を、例えば、Illumina または Ion Torrent division of Thermo Fisher, Inc. から入手可能な合成技術による配列を使用する配列決定のためにプールすることができる。いったん配列決定されると、配列読み込み 612 を、集合読み込

50

み 6 1 4 及び 6 1 6 に示されているような、それぞれの断片セットに、少なくとも部分的には含まれているバーコードに基づいて、場合により、かつ好ましくは、部分的には断片それ自体の配列に基づいて起因させることができる。次にそれぞれの断片セットの起因配列読み込みを会合して、それぞれの試料断片の会合配列、例えば、配列 6 1 8 及び 6 2 0 をもたらし、次にこれら、それぞれの元の染色体 (6 0 0 及び 6 0 2) に更に起因させることができる。ゲノム配列を会合する方法及び系は、例えば、米国特許出願第 1 4 / 7 5 2 , 7 7 3 号、2 0 1 5 年 6 月 2 6 日出願に表されており、その全開示は、その全体が参照として本明細書に組み込まれる。

#### 【 0 0 7 9 】

III . 標的化配列決定へする方法及び系の適用

10

1 つの態様において、本明細書に記載されている系及び方法は、ゲノムの標的化領域から配列情報を得るために使用される。

#### 【 0 0 8 0 】

ゲノムの「標的化」領域 (ならびに、任意の文法的に同等なもの) とは、全ゲノム、または目的のものとして特定されている、及び / もしくは本明細書に記載されている 1 つ以上の方法により選択されたゲノムの任意の 1 つ以上の領域を意味する。本明細書に記載されている方法及び系により配列決定されたゲノムの標的化領域には、イントロン、エクソン、遺伝子間領域、またはこれらの任意の組み合わせが限定されることなく含まれる。特定の例において、本明細書に記載されている方法及び系は、全エクソーム、エクソームの一部、1 つ以上の選択された遺伝子 (選択された遺伝子パネルを含む)、1 つ以上のイントロン及びイントロンとエクソン配列の組み合わせについての配列情報をもたらす。

20

#### 【 0 0 8 1 】

ゲノムの標的化領域には、配列により特定された領域ではなく、ゲノムある特定の部分または割合も含まれる。特定の実施形態において、本明細書に記載されている方法及び系により捕捉及び分析されるゲノム標的化部分には、ゲノムの 1 、 2 、 5 、 1 0 、 1 5 、 2 0 、 2 5 、 5 0 、 1 0 0 、 2 0 0 、 2 5 0 、 5 0 0 、 7 5 0 、 1 0 0 0 、または 1 0 0 0 0 キロベース毎に位置するゲノムの部分が含まれる。更なる実施形態において、ゲノムの標的化領域は、全ゲノムの 5 % 、 1 0 % 、 1 5 % 、 2 0 % 、 3 0 % 、 4 0 % 、 5 0 % 、 6 0 % 、 7 0 % 、 8 0 % 、 9 0 % 、 9 5 % を構成する。なお更なる例において、標的化領域は、全ゲノムの 1 ~ 1 0 % 、 5 ~ 2 0 % 、 1 0 ~ 3 0 % 、 1 5 ~ 4 0 % 、 2 0 ~ 5 0 % 、 2 5 ~ 6 0 % 、 3 0 ~ 7 0 % 、 3 5 ~ 8 0 % 、 4 0 ~ 9 0 % 、または 4 5 ~ 9 5 % を構成する。

30

#### 【 0 0 8 2 】

一般に、ゲノムの標的領域は、当該技術に既知であり、本明細書に記載されている任意の配列決定方法を使用して捕捉される。「捕捉される」とは、本明細書において使用されるとき、核酸及び / もしくは核酸談判の個体群を、得られた個体群が、目的のものではないゲノム領域と比較して、増加した割合で目的の標的化領域を含有するように豊富化する任意の方法または系を意味する。更なる実施形態において、豊富化された個体群は、標的化領域を含む核酸 / 核酸断片の少なくとも 5 0 % 、 5 5 % 、 6 0 % 、 7 0 % 、 7 5 % 、 8 0 % 、 8 5 % 、 9 0 % 、 9 5 % 、 9 7 % 、 9 8 % 、 9 9 % 、または 1 0 0 % を含有する。

40

#### 【 0 0 8 3 】

捕捉方法は、標的化領域がハイブリッド形成または他の会合を介して表面の捕捉分子により捕捉される、チップに基づいた方法、及び標的化領域 (または標的化領域に近い領域) に相補的なオリゴヌクレオチドプローブ (ベイト) が、ゲノム断片ライブラリーとハイブリッド形成する、溶液に基づいた方法、が一般に含まれる。本明細書に開示されている捕捉方法に使用されるプローブは、一般に、ハイブリッド形成するプローブ及び断片を「プルダウン」することに使用できるビオチンなどの捕捉分子に結合されており、これらのプルダウン方法には、目的の標的化領域を含有する核酸または核酸断片とハイブリッド形成するベイトが、目的の領域を含有しない断片から離されている任意の方法が含まれる。プローブがビオチン化されている実施形態では、磁気ストレプトアビジンビーズを使用し

50

て、結合した標的化領域を有するベイトを選択的にブルダウンし、豊富化する。

【0084】

更なる態様において、更なる研究に望ましい全ての標的化領域を網羅するベイトのライブラリーが、使用される。全エクソーム分析において、そのようなベイトライブラリーは、したがって、完全エクソームと一緒に網羅するオリゴヌクレオチドプロ-ブを含む。特定の実施形態において、エクソームの部分のみが、更なる分析に必要とされる。そのような実施形態において、ベイトはエクソームのサブセットを標的化するように設計される。この設計は、当該技術に既知の方法及びアルゴリズムを使用して達成することができ、一般に、ヒトゲノムなどの基準配列に基づいている。

【0085】

いくつかの例において、本明細書に記載されている方法及び系により処理及び配列決定された標的化ゲノム領域は、完全または部分エクソームである。これらの完全または部分エクソームを、NimbleGen 2.1M Human Exomeアレイ及びNimbleGen SeqCap EZ Exome Libraryを含む任意のRoche/NimbleGenエクソームプロトコル、任意のAgilent SureSelect製品、TruSeq及びNextera Exome製品を含む任意のIlluminaエクソーム捕捉製品、ならびに当該技術に既知の任意の他の製品、方法、系及びプロトコルが限定されることなく含まれる、当該技術に既知の任意の方法を使用して、配列決定のために捕捉することができる。

【0086】

更なる実施形態において、目的の標的化領域が、エクソームの全体または部分を含む場合、これらの標的化領域の捕捉に使用されるベイトは、これらのエクソン配列に相補的であるように設計することができる。他の実施形態において、ベイトはそれ自体エクソン配列に相補的ではないが、その代わりに、エクソン配列の近くの配列または2つのエクソン間のイントロン配列に相補的である。そのような設計は、本明細書において「アンカードエクソーム捕捉」または「イントロンベйтиング (baiting)」とも呼ばれ、これは、本明細書において考察されたように、エクソームの1つ以上の部分が、目的のエクソームの1つ以上の部分に近い、または隣接する1つ以上のイントロン配列に相補的なベイトの使用を介して捕捉されるプロセスを意味する。例えば、図2に概略的に例示されているように、遺伝子配列201は、エクソン領域202及び203を含む。これらのエクソン領域は、近くの1つ以上のイントロン配列（例えば、エクソン領域202の捕捉にはイントロン領域204及び/または205、エクソン領域203の捕捉にはイントロン領域206）に向けられたベイトを利用することによって、捕捉することができる。換言すると、エクソン領域202または203を含む断片の個体群は、イントロン領域204及び/または205及び206に相補的なベイトの使用によって捕捉される、いくつかの実施形態において、イントロンベйтиングを使用して、長いイントロンを低密度でベйтиングすることによって、長いイントロン領域により離されているエクソンを架橋する。そのような実施形態において、ベイトは、目的のエクソン領域に近いイントロン領域を標的にする必要がなく、代わりにベイトは特定の距離（もしくは、距離のセット）で離されている領域を標的にするように設計される、または特定の数の塩基もしくは多数の塩基の組み合わせによってイントロン領域にわたってタイル状になるように設計される。そのような実施形態は、下記に更に詳細に記載される。

【0087】

いくつかの実施形態において、本発明のアンカードエクソーム捕捉/イントロンベйтиング技術に使用されるイントロン領域は、捕捉されるエクソン領域に隣接している。更なる実施形態において、イントロン領域は、捕捉されるエクソン領域から、約1~50個、2~45個、3~40個、4~35個、5~30個、6~25個、7~20個、8~15個、9~10個、2~20個、3~15個、4~10個、5~30個、10~40個、15~50個、20~75個、25~100個のヌクレオチドによって離されている。更なる実施形態において、イントロン領域は、捕捉されるエクソン領域から、約5、10、

20、30、40、50、60、70、80、90、100、125、150、175、200、300、400、または500個のヌクレオチドによって離されている。更なる実施形態において、特に、イントロン領域の低密度ベითィングが使用される（位相変異体の検出する、または大きなイントロン距離のわたる連結エクソン領域の特定などの）状況では、イントロン領域は、捕捉されるエクソン領域から、キロベースのオーダーの距離、例えば、1～20、2～18、3～16、4～14、5～12、6～10キロベースによって離されている。オリゴヌクレオチドの豊富化個体群の元の分子構成が保持されているので、このイントロン領域の低密度ベითィングは、長いイントロンにより離されているエクソン領域間の配列情報の連結を可能にする。

#### 【0088】

更なる態様においてベイトを、ゲノムの特定の領域を標的にするように設計するのではなく、タイリング手法が使用される。そのような手法では、特定のエクソンまたはイントロン領域を標的にするのではなく、ベイトは、その代わりに、特定の範囲または距離でゲノムの部分に相補的であるように設計される。例えば、ベイトのライブラリーを、このベイトのライブラリーを断片化試料に適用することが、ゲノムの特定のサブセット、すなわち、ベイトに相補的な配列を含有する断片に含有されている領域のみを捕捉するように、ゲノムに沿って5キロベース（kb）毎に位置する配列とハイブリッド形成するように設計することができる。理解されるように、ベイトを、ヒトゲノム基準配列などの基準配列に基づいて設計することができる。更なる実施形態において、タイル状のベイトライブラリーは、ゲノムの1、2、5、10、15、20、25、50、100、200、250、500、750、1000、または10000キロベース毎に領域を捕捉するように設計される。いくつかの例において、このタイリング方法は、イントロン領域を低密度に捕捉する効果を有し、したがって、イントロン領域の低密度捕捉を介して捕捉されたこれらのエクソン領域の元の分子構成が保持されているので、長いイントロン領域により離されているエクソン領域の配列情報を連結する方法を提供する。

#### 【0089】

なお更なる実施形態において、ベイトはゲノムを無作為または組み合わせ様式でタイル状にするように設計され、例えば、タイル状ライブラリーの混合物を、ライブラリーのうちのいくつかは1kb毎に領域を捕捉し、混合物の他のライブラリーが100kb毎に領域を捕捉する場合に使用することができる。なお更なる実施形態において、タイル状ライブラリーは、ベイトがゲノム内の位置の範囲内を標的にするように設計され、例えば、ベイトは、ゲノムの1～10、2～5、5～200、10～175、15～150、20～125、30～100、40～75、50～60kb毎の領域を標的にすることができる。更なる例において、本明細書に記載されているタイル状または他の捕捉方法は、全ゲノムの約5%、10%、15%、20%、30%、40%、50%、60%、70%、80%、90%、95%を捕捉する。理解されるように、そのようなタイリング捕捉方法は、配列決定などの更なる分析のため、イントロンとエクソンの領域を両方とも捕捉する。

#### 【0090】

なお更なる実施形態において、本明細書に記載されている方法のいずれかによると、発明の方法に使用されるベイトのライブラリーは、本明細書に更に記載される特徴の1つ以上を満たす情報価値のある設計の産物である。情報価値のある設計（*informative*）単一ヌクレオチド多形（*SNP*）に向けられている場合を含む。上記に考察されたように、用語「有益なSNP」は、本明細書において使用されるとき、ヘテロ接合性であるSNPを指す。ベイトのライブラリーは、いくつかの例において、有益なSNPを含有するゲノム試料の領域に向けられている複数のプローブを含有するように設計される。「向けられている」とは、本明細書において使用されるとき、プローブが、ゲノム配列の領域に相補的である配列を含有することを意味する。有益なベイト設計は、完全な網羅を伴う標的化豊富化を可能にし、同時に必要とされるプローブの数を低減する（したがって費用を低減し、ワークフローを合理化する）ことによって、標的化配列決定方法を最適化する能力を提供する。

10

20

30

40

50



## 【 0 0 9 1 】

一般に、有益なベイト設計を利用する方法において、ベイトのライブラリーは、ゲノムの標的化領域における特定に配列に、これらの領域における有益な S N P の存在もしくは不在及び / またはこれらの有益な S N P の位置に基づいて向けられるように設計される。有益なベイト設計を一般的に考慮した例示的な図が、図 8 に提供されている。ゲノム 8 0 1 の領域は、エクソン ( 8 0 2 及び 8 0 3 ) を含むことができる。いくつかの例において、有益な S N P ( 8 0 4 ) は、エクソン ( 8 0 2 ) と隣接するイントロンの境界に位置する。そのような状況において、ベイトライブラリーは、境界から特定の距離で離れた 1 つ以上のヌクレオチド ( 8 0 5 ) に向けられたプローブを含むように設計することができる。エクソンと隣接イントロン ( 8 0 6 ) の境界に有益な S N P が存在しない更なる例において、ベイトライブラリーは、境界に近いイントロン ( 8 0 7 及び 8 0 8 ) の 1 つ以上の位置に向けたプローブを含むように設計することができる。これらの位置は、好ましくは有益な S N P を含むが、必要であれば他の S N P 及び / または他の配列を含むこともできる。エクソン 8 0 3 がエクソンの内部に有益な S N P 8 0 9 を含有するが、境界に有益な S N P が存在しない、なお更なる例において、ベイトライブラリーは、有益及び有益ではない S N P の混合物 ( ならびに、必要であれば任意の他の配列 ) を含む隣接イントロンのいくつかの位置 8 1 0 、 8 1 1 及び 8 1 2 に向けたプローブを含むように設計することができる。

10

## 【 0 0 9 2 】

いくつかの態様において、1 つ以上の投入特性を使用して、これらの投入特性、ならびに様々な領域のマップ品質に基づいてゲノムに沿って位置を移動するように向けられた、プローブベイトライブラリーを設計する。この設計は、一般に、イントロン及びエクソンの位置ではなく、有益な S N P の間に間隔を置くことに基づいている。しかし、理解されるように、イントロン及びエクソンの位置に基づいたベイト設計に関する本明細書に提示されている記載のいずれかを、有益な S N P に基づいて情報価値のあるベイト設計方法と組み合わせて使用することもできる。情報価値のあるベイト設計に使用される投入特性には、エクソン、イントロン、遺伝子間領域、有益な S N P 、ならびに反復配列の領域 ( 例えば、G C 豊富領域 ) 、セントロメア及び試料核酸長さが、限定されることなく、任意の組み合わせ位置で含まれる。

20

## 【 0 0 9 3 】

考察を容易にするため、プローブライブラリーの情報価値のある設計の異なる特性が、異なる潜在的实施形態の観点から下記に記載される。理解されるように、本明細書において考察されるプローブライブラリーのいずれかを、情報価値のある設計要素または上記に考察された設計の任意の他の種類を使用するかにかかわらず、単一または任意の組み合わせで使用することができる。利用される設計要素は、目的の標的化ゲノム領域、ならびに目的の領域への試料投入及びマッピングの品質に基づいて選択される。

30

## 【 0 0 9 4 】

いくつかの実施形態において、プローブベイトライブラリーは、所定の試料中に有益な S N P を含有する高い可能性を有する領域に向けられたプローブを含むように設計される。そのような標的には、個別の塩基 ( 有益な S N P それ自体 ) または有益な S N P に近位の、または隣接している 1 個以上の塩基が含まれる。なお更なる実施形態において、プローブベイトの標的は、有益な S N P と直接隣接していてもよく、または有益な S N P から約 1 ~ 2 0 0 、 1 0 ~ 1 9 0 、 2 0 ~ 1 8 0 、 3 0 ~ 1 7 0 、 4 0 ~ 1 6 0 、 5 0 ~ 1 5 0 、 6 0 ~ 1 4 0 、 7 0 ~ 1 3 0 、 8 0 ~ 1 2 0 、 9 0 ~ 1 0 0 の塩基の距離で離れていてもよい。

40

## 【 0 0 9 5 】

更なる実施形態において、プローブベイトライブラリーには、核酸分子の平均長さに関して特定の密度の領域に向けられたプローブが含まれる。例えば、プローブは、プローブがハイブリッド形成する核酸分子 / 断片の平均長さより  $\times$  倍密度がある標的配列の密度でプローブを含み、ここで  $\times$  は、限定されることなく、1、5、10、20、50、75、

50

100、125、150、または200でありうる。プローブ標的の密度を核酸の長さに対して増加させると、同じ物理的な分子における遺伝子座にわたってプローブを連結させる能力が増加する。そのような方法は、連結された領域が有益なSNPを含み、したがって、プローブベイトライブラリーがゲノムの標的化領域に結合する能力を更に改善する確率も改善されうる。

#### 【0096】

プローブ標的の密度は、(個体群のレベルにおいて)目的の所定の領域に有益なSNPがある高い確率のない状況においても増加させることができる。そのような領域において、本明細書に記載されているものなどのタイリング方法を使用して、領域に沿って周期的な間隔においてプローブを向けることができる。特定の実施形態において、間隔の密度は、有益なSNPを欠いている領域におけるプローブ間隔の密度が、有益なSNPを含有する領域におけるプローブ間隔より、距離が1、2、5、10、25、50倍短いように、基礎が異なりうる。

10

#### 【0097】

更なる実施形態において、プローブベイトライブラリーは、遺伝子(エクソン及びイントロンを含む)内の有益なSNPの分配のみを考慮するように設計される。この設計方法は、遺伝子の一方の末端から他方へと連結/位相する主要な位置においてヘテロ接合性SNPの十分な数を捕捉することに向けられている。そのような設計方法は、エクソンの有益なSNPを、遺伝子における有益なSNP間の距離が上記に記載された間隔密度より低いように、1つ以上の非エクソンSNPと組み合わせる、標的のセットに向けられている

20

#### 【0098】

そのような情報価値のある設計方法は、ゲノムの一般の標的化領域の検出を可能にするのみならず、転位及び遺伝子融合などのゲノム構造変動の検出及び位相も可能にする。任意の個別の遺伝子が位相されうることを確実にすることによって、続いて、圧倒的多数の遺伝子融合事象を、本明細書に記載されている方法の使用により検出及び位相することができる。

#### 【0099】

特定の実施形態において、上記のいずれかによって、ベイトライブラリーは、約1kb~約2Mbの距離でプローブを標的にするように設計される。更なる実施形態において、距離は、約1~50、5~45、10~40、15~35、20~30、10~50kbである。

30

#### 【0100】

更なる実施形態において、プローブベイトによって標的にされる核酸断片は、約2kb~約250Mbである。なお更なる実施形態において、断片は、約10~1000、20~900、30~800、40~700、50~600、60~500、70~400、80~300、90~200、100~150、50~500、25~300kbである。

#### 【0101】

いくつかの実施形態において、プローブベイトライブラリーは、約60~95%のプローブが、有益なSNPを含有する配列とハイブリッド形成するように設計される。更なる実施形態において、プローブベイトライブラリーは、プローブライブラリーにおける約65%~85%、70%~80%、60~90%、80~90%、90~95%、95%~99%のプローブが有益なSNPとハイブリッド形成するように設計されるように、設計される。なお更なる実施形態では、プローブライブラリーにおける少なくとも65%、75%、85%、90%、91%、92%、93%、94%、95%、96%、97%のプローブが、有益なSNPとハイブリッド形成するように設計される。理解されるように、有益なSNPと「ハイブリッド形成する」ように設計されるプローブとは、そのようなプローブが、有益なSNPを含む配列領域とハイブリッド形成することを意味する。

40

#### 【0102】

50

更なる実施形態において、プローブペイトライブラリーは、ゲノム試料の標的部分におけるエクソン及びイントロンの両方の範囲内に位置する有益なSNPに向けられた複数のプローブを含むように、設計される。

【0103】

なお更なる実施形態において、ライブラリーは、ライブラリーにおける大部分のプローブが、約1~15、5~10、3~6kbの間隔を置いた有益なSNPとハイブリッド形成するように設計される。なお更なる実施形態において、プローブのライブラリーにおける大部分のプローブは、約1、3、5、10、20、30、50kbの間隔を置いた有益なSNPとハイブリッド形成するように更に設計される。

【0104】

更なる実施形態において、プローブのライブラリー内の複数のプローブは、エクソンとイントロンの境界の5~300、10~50、20~100、30~150、または40~200kbの範囲内に有益なSNPがないゲノム試料の標的部分では、複数のプローブが、その境界からのイントロン内の有益なSNPとハイブリッド形成するように設計されるように、設計される。

【0105】

更なる実施形態において、プローブのライブラリー内の複数のプローブは、エクソン内に第1の有益なSNPが存在し、その第1の有益なSNPが隣接イントロンとの境界から5~300、10~50、20~100、30~150、または40~200kbに位置し、隣接イントロン内に第2の有益なSNPが存在し、その第2の有益なSNPが境界から10~50kbに位置するゲノム試料の標的部分では、複数のプローブが、第1及び第2の有益なSNPの間のゲノム試料の領域とハイブリッド形成するように設計されるように、設計される。

【0106】

更なる実施形態において、プローブのライブラリー内の複数のプローブは、少なくとも5~300、10~50、20~100、30~150、または40~200kbに有益なSNPを含まないゲノム試料の標的部分では、複数のプローブが、ゲノム試料のこれらの標的部分と0.5、1、3、または5kb毎にハイブリッド形成するように設計されるように、設計される。更なる実施形態において、プローブのライブラリー内の複数のプローブは、ゲノム試料の標的部分に沿っての0.1、0.5、1、1.5、3、5、10、15、20、30、35、40、45、50kb毎にハイブリッド形成するように設計される。

【0107】

更なる実施形態において、プローブのライブラリー内の複数のプローブは、エクソンとイントロンの境界の5~300、10~50、20~100、30~150、または40~200kbの範囲内に有益なSNPがないゲノム試料の標的部分では、複数のプローブが、エクソンとイントロンの境界に二番目に近い有益なSNPとハイブリッド形成するように設計されるように、設計される。

【0108】

更なる実施形態において、プローブのライブラリーは、バーコードに連結情報を提供する密度でエクソンに隣接するゲノム試料の領域とハイブリッド形成するように設計されたプローブを含む。

【0109】

なお更なる実施形態において、プローブのライブラリーにより表される適用範囲は、長い個別の核酸断片分子を高い割合で含有する方法が小さい適用範囲のプローブのライブラリーを使用するように、分離区分内のゲノム試料の個別の核酸断片分子の長さの分配に反比例する。

【0110】

なお更なる実施形態において、プローブのライブラリーは、ゲノム使用の標的部分の適用範囲を最適化する。なお更なる実施形態において、適用範囲の密度は、高品質マップ

10

20

30

40

50

の領域では、特に、有益な S N P を含有する領域では低い場合があり、密度は、連結情報が標的化領域にわたって提供されることを確実にするため、低品質マップの領域において更に高い場合がある。

【 0 1 1 1 】

なお更なる実施形態において、プローブのライブラリーは、高品質マップの標的化部分では、プローブのライブラリーが、エクソンとイントロンの境界の 1 k b ~ 1 M b の範囲内の有益な S N P とハイブリッド形成するプローブを含む、というような、ゲノム試料の 1 つ以上の標的化部分の特徴により情報提供される特徴を有する。プローブのライブラリーは、そのような状況において、エクソンとイントロンの境界の 1 0 ~ 5 0 0 、 2 0 ~ 4 5 0 、 3 0 ~ 4 0 0 、 4 0 ~ 3 5 0 、 5 0 ~ 3 0 0 、 6 0 ~ 2 5 0 、 7 0 ~ 2 0 0 、 8 0 ~ 1 5 0 、 9 0 ~ 1 0 0 k b の範囲内の有益な S N P とハイブリッド形成するプローブを更に含んでもよい。

10

【 0 1 1 2 】

なお更なる実施形態において、プローブのライブラリーは、バーコード付き前記断片の長さの前記分配が、高い割合で約 1 0 0 、 1 5 0 、 2 0 0 、 2 5 0 k b より長い断片を有する標的化部分では、前記プローブのライブラリーが、少なくとも 5 0 k b 離れている有益な S N P とハイブリッド形成するプローブを含む、というような、ゲノム試料の 1 つ以上の標的化部分の特徴により情報提供される特徴を有する。プローブのライブラリーは、そのような状況において、少なくとも 5 、 1 0 、 1 5 、 2 0 、 2 5 、 3 0 、 3 5 、 4 0 、 4 5 、 5 0 、 7 5 、 1 0 0 、 1 2 5 、 1 5 0 、 1 7 5 、 2 0 0 k b 離れている有益な S N P とハイブリッド形成するプローブを更に含んでもよい。

20

【 0 1 1 3 】

なお更なる実施形態において、プローブのライブラリーは、低品質マップの標的化部分では、プローブのライブラリーが、エクソンとイントロンの境界の 1 k b 内の有益な S N P とハイブリッド形成するプローブを含む、というような、ゲノム試料の 1 つ以上の標的化部分の特徴により情報提供される特徴を有する。プローブのライブラリーは、そのような状況において、エクソンとイントロンの境界の 2 、 5 、 1 0 、 1 5 、 2 0 、 2 5 、 3 0 、 3 5 、 4 0 、 4 5 、 5 0 、 7 5 、 1 0 0 、 1 2 5 、 1 5 0 、 1 7 5 、 2 0 0 k b 内の有益な S N P とハイブリッド形成するプローブを更に含んでもよい。そのような状況において、ライブラリーは、エクソン内、イントロン内、またはその両方における有益な S N P とハイブリッド形成するプローブを更に含む。

30

【 0 1 1 4 】

なお更なる実施形態において、プローブのライブラリーは、遺伝子間領域を含む標的化部分では、プローブのライブラリーが、少なくとも 1 、 2 、 5 、 1 0 、 1 5 、 2 0 、 2 5 、 3 0 、 3 5 、 4 0 、 4 5 、 5 0 、 7 5 、 1 0 0 k b の距離で間隔を置いた有益な S N P とハイブリッド形成するプローブを含む、というような、ゲノム試料の 1 つ以上の標的化部分の特徴により情報提供される特徴を有する。

【 0 1 1 5 】

本明細書に記載されている捕捉方法に使用されるベイトは、断片の個体群を、ゲノムの標的亜溶離機を含有する断片で豊富化するのに有用な任意のサイズ及び構造でありうる。上記において考察されたように、一般に本発明に使用されるベイトは、ビオチンなどの捕捉分子に結合されたオリゴヌクレオチドプローブを含む。オリゴヌクレオチドプローブは、目的の標的化領域内の配列と相補的でありうる、または標的化領域の外側であるが、ベイトが近くの領域（例えば、隣接イントロン）とハイブリッド形成することによって標的化領域をブルダウンすることができるように、「アンカーリング」領域と標的化領域の両方が同じ断片内にある標的化領域に十分近い領域と相補的でありうる。

40

【 0 1 1 6 】

ベイトに結合した捕捉分子は、ベイト及びそのハイブリッド形成パートナーを固体分の他の断片から単離することに使用できる任意の捕捉分子でありうる。一般に、本明細書に使用されるベイトは、ビオチンに結合しており、次にストレプトアビジン（磁気ストレブ

50

トアビジンビーズが限定されることなく含まれる)を含む固体支持体を使用して、ベイト及びベイトがハイブリッド形成する断片を捕捉することができる。他の捕捉分子対には、ビオチン/ニュートラアビジン、抗原/抗体、または相補的オリゴヌクレオチド配列が限定されることなく含まれてもよい。

#### 【0117】

更なる実施形態において、ベイトのオリゴヌクレオチドプローブ部分は、標的化領域または標的化領域に近い領域とのハイブリッド形成に適した任意の長さでありうる。いくつかの実施形態において、本明細書に記載されている方法により使用されるベイトのオリゴヌクレオチドプローブ部分、すなわち、ゲノムの標的化領域または標的化領域に近い領域とハイブリッド形成する部分は、一般に、約10～約150個の長さのヌクレオチド長さ(例えば、35個のヌクレオチド、50個のヌクレオチド、100個のヌクレオチド)を有し、目的の標的配列と特異的にハイブリッド形成するように選択される。更なる実施形態において、オリゴヌクレオチドプローブ部分は、約5～10個、10～50個、20～100個、30～90個、40～80個、50～70個の長さのヌクレオチド長さを備える。理解されるように、本明細書に記載されている任意のオリゴヌクレオチドプローブ部分は、RNA、DNA、PNA、LNAなどの非天然ヌクレオチド、またはこれらの任意の組み合わせを含むことができる。

#### 【0118】

本明細書に記載されている方法及び系の利点は、捕捉された標的化領域が、標的化領域を捕捉する及び配列決定分析を実施するステップの後であっても、これらの標的化領域の元の分子構成が保持されるように、捕捉の前に処理されることである。特定の標的化領域を元の分子構成(誘導されている元の染色体もしくは染色体領域、及び/または完全ゲノム内の特定の標的化領域の互いの位置が含まれる)に起因させる能力は、そうでなければ伝統的な配列決定技術の使用により不十分にマップされる、または不十分な適用範囲を有するゲノムの領域から、配列情報を得る方法を提供する。

#### 【0119】

例えば、いくつかの遺伝子は、一般に利用可能な配列決定技術を使用して、特に、短い読み込み技術を使用してスパンするには長すぎる長いイントロンを有する。短い読み込み技術は、長い読み込み技術と比較して優れた精度を有するので、多くの場合に好ましい配列決定技術である。しかし、一般に使用される短い読み込み技術は、ゲノムの長い領域にわたってスパンすることができず、したがって、情報は、長いタンDEM反復配列の長さ、高GC含有量及び長いイントロンを含有するエクソンなどの構造的特徴に起因して特徴決定することが困難であるゲノムの領域に、これらの従来の技術を使用して得ることができない。しかし、本明細書に記載されている方法及び系では、一般に図1に例示されており、本明細書において更に詳細に記載されるタグ付け手順を介して、標的化領域の分子構成が保持される。このように、関連性がゲノムの伸長領域にわたって作られる。例えば、図2Bに概略的に例示されているように、核酸分子207は、長いイントロン領域(208)を伴う2つのエクソン(影付バー)を含有する。本明細書に記載されている方法では、個別の核酸分子207は、それ自体の分離区分211に分配され、次に異なる断片が異なる部分のエクソン及びイントロンを含有するように断片化される。これらの断片のそれぞれが、断片から得られた任意の配列情報が、断片が生成された分離区分に起因するようにタグ付けされるので、それぞれの断片も、断片が誘導された個別の核酸分子207に起因する。

#### 【0120】

一般に、本明細書において更に記載されるように、断片化及びタグ付けの後、異なる区分からの断片を一緒に合わせる。次に標的化捕捉方法を使用して、配列決定などの更なる分析を受ける断片の個体群を、目的の標的化領域を含有する断片で豊富化することができる。図2Bに例示されている例では、使用されるベイトが断片の個体群を豊富化して、エクソンの一部を含有するもののみを捕捉するが、エクソン及びイントロンの外側にある領域(例えば、209及び210)は捕捉しない。したがって、配列決定を受ける断片の最

終個体群は、エクソンが長いイントロン領域により離されていたとしても、エクソンの部分を含有する断片が豊富化される。次に短い読み込みで高い精度の配列決定技術を使用して、この断片豊富化個体群の配列を特定することができ、それぞれの断片がタグ付けされ、したがって、元の分子構成、すなわち、元の個別の核酸分子に起因するので、短い読み込み配列は、つなぎ合わされて、エクソン間の関係についての情報を提供することができる。いくつかの実施形態において、1つ以上のエクソンの全て、または一部を含有する断片の捕捉に使用されるベイトは、1つ以上のエクソンそれ自体の1つ以上の部分に相補的である。他の実施形態において、ベイトは、介在イントロンの1つ以上の部分、またはエクソン領域の3'もしくは5'の側に隣接した、もしくは近い配列に相補的である(そのようなベイトは、本明細書において「イントロンベイト」とも呼ばれる)。更なる実施形態において、エクソンの全て、または一部を含有する断片の捕捉に使用されるベイトには、エクソンそれ自体及びイントロンバイトに相補的なベイトが含まれる。

10

#### 【0121】

捕捉された標的化領域の分子構成を配列決定のために保持する能力は、ゲノムの不十分に特徴決定された領域にわたって配列決定することができる利点ももたらす。理解されるように、有意な率(例えば、Alt et al., PLOS Computational Biology, May 15, 2014, Vol. 10, Issue 5によると、少なくとも5~10%)のヒトゲノムが、未会合であり、位置が定まらず、不十分に特徴決定されたままである。基準会合は、一般に、これらの欠損領域を、主にセントロメア及び末端動原体型染色体の短いアームの近くに見出される、マルチメガベース(multi-megabase)異質染色性間隙と注釈する。ゲノムの欠損部分には、一般に使用される配列決定技術を使用する正確な特徴決定に抵抗する状態のままである、構造的特徴が含まれる。ゲノムの伸長領域にわたって情報を連結する能力を提供することによって、本明細書に記載されている方法は、これらの不十分に特徴決定された領域にわたる配列決定を可能にする方法を提供する。

20

#### 【0122】

いくつかの例において、ゲノムDNAを断片化する、増幅する、区分化する、別の方法で処理する方法を含む試料の調製方法は、ゲノムの特定の領域に偏った、または低い適用範囲をもたすことができる。そのように偏った、または低い適用範囲は、ゲノムの標的化領域の捕捉に使用されるベイトの濃度を変更することによって、本明細書に開示されている方法及び系において補うことができる。例えば、いくつかの実施形態において、ゲノムの他の領域より特定の領域に偏向をもたす高いGC含有量または他の構造的変動を含有する領域など、ゲノムの特定の領域は、断片ライブラリーが処理された後に低い適用範囲を有することが知られている。そのような状況では、ベイトのライブラリーを変更して、低い適用範囲の領域に向けられるベイトの濃度を増加することができ、換言すると、使用されるベイトの個体群を「スパイク(spike)」して、低い適用範囲の区域におけるゲノムの標的化領域を含有する十分な数の断片が、配列決定される断片の最終個体群において得られることを確実にすることができる。そのようなベイトのスパイクは、いくつかの実施形態において、特別仕様のライブラリーを設計することによって実施してもよい。更なる実施形態において、そのようなベイトのスパイクは、低い適用範囲の領域に向けられたベイトの特別仕様のライブラリーを既製のエクソーム捕捉キットに加えるようにして、市販の全エクソームキットによって実施することができる。

30

40

#### 【0123】

本明細書に記載されている方法及び系の利点は、捕捉された標的化領域が、標的化領域を捕捉する及び配列決定分析を実施するステップの後であっても、これらの標的化領域の元の分子構成が保持されるように、捕捉の前に処理されることである。本明細書において更に詳細に考察されるように、特定の標的化領域を元の分子構成(誘導されている元の染色体もしくは染色体領域、及び/または完全ゲノム内の特定の標的化領域の互いの位置が含まれる)に起因させる能力は、そうでなければ伝統的な配列決定技術の使用により不十分にマップされる、または不十分な適用範囲を有するゲノムの領域から、配列情報を得

50

る方法を提供する。

【0124】

例えば、いくつかの遺伝子は、一般に利用可能な配列決定技術を使用して、特に、長い読み込み技術と比較して優れた精度を有する短い読み込み技術を使用してスパンするには長すぎる長いイントロンを有する。しかし、本明細書に記載されている方法及び系では、一般に図1に例示されており、本明細書において更に詳細に記載されるタグ付け手順を介して、標的化領域の分子構成が保持される。このように、関連性がゲノムの伸長領域にわたって作られる。例えば、図2Bに概略的に例示されているように、核酸分子207は、長いイントロン領域により中断されたエクソン（影付バー）を含有する。一般に使用される配列決定技術は、イントロンを横断する距離をスパン（span）して2つのエクソンの関係に関する情報を提供することができない。本明細書に記載されている方法では、個別の核酸分子207は、それ自体の分離区分209に分配され、次に異なる断片が異なる部分のエクソン及びイントロンを含有するように断片化される。これらの断片のそれぞれが、断片から得られた任意の配列情報が、断片が生成された分離区分に起因するようにタグ付けされるので、それぞれの断片も、断片が誘導された個別の核酸分子207に起因する。一般に、本明細書において更に記載されるように、断片化及びタグ付けの後、異なる区分からの断片を一緒に合わせる。次に標的化捕捉方法を使用して、配列決定などの更なる分析を受ける断片の個体群を、目的の標的化領域を含有する断片で豊富化することができる。図2Bに例示されている例では、使用されるベイトが断片の個体群を豊富化して、エクソンのうちの1つの一部を含有するもののみを捕捉するが、エクソンの外側にある領域（例えば、209及び210）は捕捉しない。したがって、配列決定を受ける断片の最終個体群は、目的のエクソンを含有する断片が豊富化される。次に短い読み込みで高精度の配列決定技術を使用して、この断片豊富化個体群の配列を特定することができ、それぞれの断片がタグ付けされ、したがって、元の分子構成、すなわち、元の個別の核酸分子に起因するので、短い読み込み配列は、つなぎ合わされて、介在イントロンの長さ（幾つの場合では、1、2、5、10、またはそれ以上のキロベースの長さのオーダーでありうる）を超えてスパンする情報を提供して、2つのエクソンの関係についての情報を提供することができる。

10

20

【0125】

上記に示されたように、本明細書に記載されている方法及び系は、長い核酸の短い配列読み込むために、個別の分子構成を提供する。本明細書において使用されるとき、個別の分子構成は、特定の配列読み込みを超えた配列構成、例えば、配列読み込みそれ自体の範囲内に含まれない、隣接または近位配列との関係を指し、このように、典型的には短い配列読み込みの全体または一部に含まれないようなもの、例えば、対読み込みでは約150塩基または約300塩基の読み込みである。特に好ましい態様において、方法及び系は、短い配列読み込みのために長い範囲の配列構成を提供する。そのような長い範囲の構成は、所定の配列読み込みと、互いに1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長い、もしくはさらに100kbより長い、またはそれ以上長い距離内にある配列読み込みとの関係または連結を含む。長い範囲の個別の分子構成を提供することによって、本発明の方法及び系は、はるかに長い推定分子構成も提供する。配列構成は、本明細書に記載されているように、例えば、短い配列読み込みを、個別の長い分子または連鎖分子のコンティグにマッピングすることから低い解像度構成、ならびに、例えば、個別の分子の連続決定配列を有する長い個別の分子の大きな部分を、長い範囲で配列決定することから高解像度構成を含むことができ、ここで、そのような決定配列は、1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長い、またはさらには100kbより長い。配列構成によって、短い配列を、長い核酸、例えば個別の長い核酸分子、または連結核酸分子のコレクション、ま

30

40

50

たはコンティグに起因させることは、短い配列を長い核酸伸展にマッピングして、高レベルの配列構成を提供すること、ならびにこれらの長い核酸を介して短い配列から会合配列を提供することを含んでもよい。

#### 【0126】

##### IV. 試料

理解されるように、本明細書において考察される方法及び系を使用して、任意の種類のゲノム材料から標的化配列情報を得ることができる。そのようなゲノム材料は、患者から採取した試料によって得てもよい。本明細書において考察される方法及び系に使用されるゲノム材料の例示的な試料及び種類には、ポリヌクレオチド、核酸、オリゴヌクレオチド、循環無細胞核酸、循環腫瘍細胞 (CTC)、核酸断片、ヌクレオチド、DNA、RNA、ペプチドポリヌクレオチド、相補的DNA (cDNA)、二本鎖DNA (dsDNA)、一本鎖DNA (ssDNA)、プラスミドDNA、コスミドDNA、染色体DNA、ゲノムDNA (gDNA)、ウイルスDNA、細菌DNA、mtDNA (ミトコンドリアDNA)、リボソームRNA、無細胞DNA、無細胞胎児DNA (cffDNA)、mRNA、rRNA、tRNA、nRNA、siRNA、snRNA、snoRNA、scarRNA、マイクロRNA、dsRNA、ウイルスRNAなどが限定されることなく含まれる。まとめると、使用される試料は、特定の処理に必要性に応じて変わりうる。

#### 【0127】

核酸を含む任意の物質が、試料の供給源でありうる。物質は、流体、例えば、生物学的流体でありうる。流体物質には、血液、臍帯血、唾液、尿、汗、血清、精液、膣液、胃及び消化液、髄液、胎盤液、体腔液、眼液、血清、母乳、リンパ液、またはこれらの組み合わせが含まれるが、これらに限定されない。物質は、固体、例えば、生物学的組織でありうる。物質は、正常な健常組織、疾患組織、または健常及び疾患組織の混合物から構成される。いくつかの場合において、物質は腫瘍から構成される。腫瘍は、良性 (非癌) または悪性 (癌) でありうる。腫瘍の非限定例には、線維肉腫、粘液肉腫、脂肪肉腫、軟骨肉腫、骨原性肉腫、脊索腫、血管肉腫、内皮肉腫 (endotheliosarcoma)、リンパ管肉腫、リンパ管内皮肉腫 (lymphangioendotheliosarcoma)、滑膜腫、中皮腫、ユーイング腫瘍、平滑筋肉腫、横紋筋肉腫、胃腸管系癌、結腸癌、膀胱癌、乳癌、泌尿生殖器系癌、卵巣癌、前立腺癌、扁平上皮癌、基底細胞癌、腺癌、汗腺癌、皮脂腺癌、乳頭癌、乳頭腺癌、嚢胞腺癌、髄様癌、気管支原性肺癌、腎細胞癌、肝細胞癌、胆管癌、絨毛癌、精上皮腫、胚性癌、ウィルムス腫瘍、子宮頸癌、内分泌系癌、精巣腫瘍、肺癌、小細胞肺癌、非小細胞肺癌、膀胱癌、上皮癌、神経膠腫、星状細胞腫、髄芽細胞腫、頭蓋咽頭腫、上衣腫、松果体腫、血管細胞芽腫、聴神経腫、乏突起神経膠腫、髄膜腫、黒色腫、神経芽細胞腫、網膜芽細胞腫、またはこれらの組み合わせが含まれる。物質は、様々な種類の臓器と関連していてもよい。臓器の非限定例には、脳、肝臓、肺、腎臓、前立腺、卵巣、脾臓、リンパ節 (扁桃腺を含む)、甲状腺、膵臓、心臓、骨格筋、腸、咽頭、食道、胃、またはこれらの組み合わせが含まれる。いくつかの場合において、物質は、真核細胞、原核細胞、真菌細胞、心臓細胞、肺細胞、腎細胞、肝細胞、膵臓細胞、生殖細胞、幹細胞、誘導多能性幹細胞、胃腸細胞、血球、癌細胞、癌細胞、細菌細胞、ヒトマイクロバイーム試料から単離された細菌細胞などが含まれるが、これらに限定されない様々な細胞から構成される。いくつかの場合において、物質は、例えば単一細胞の構成または複数の細胞の構成などの細胞の構成を含んでもよい。個別の細胞を分析する方法及び系は、例えば、米国特許出願第14/752,641号、2015年6月26日出願に提供されており、この全開示は、その全体が、特に個別の細胞から核酸を分析することの関する全ての教示が参照として本明細書に組み込まれる。

#### 【0128】

試料は、様々な対象から得ることができる。対象は、生存している対象または死亡している細胞でありうる。対象に例には、ヒト、哺乳動物、非ヒト哺乳動物、齧歯類、両生類、爬虫類、イヌ、ネコ、ウシ、ブタ、ヤギ、ヒツジ、メンドリ、鳥類、マウス、ウサギ、昆虫、ナメクジ、微生物、細菌、寄生虫、または魚が含まれるが、これらに限定されな



い。いくつかの場合において、対象は、疾患または障害を有する、有すると疑われる、または発生する危険性のある患者でありうる。いくつかの場合において、患者は妊婦でありうる。いくつかの場合において、患者は正常で健康な妊婦でありうる。いくつかの場合において、ある特定の先天性欠損を有する子どもを身ごもっている危険性のある、妊婦でありうる。

#### 【0129】

試料は、対象から当該技術に既知の任意の方法によって得ることができる。例えば試料は、循環系（例えば、シリンジもしくは他の器具による静脈内もしくは動脈内）からの入手、分泌された生物学的試料（例えば、唾液、痰、尿、糞便など）、生物学的試料の外科的な（例えば、生検による）取得（例えば、術中試料、術後試料など）、スワブ拭き取り（swabbing）（例えば、頬側スワブ、口腔咽頭スワブ）、またはピペット操作を介して、対象から得てもよい。

10

#### 【0130】

本発明の好ましい実施形態が本明細書において示され、記載されてきたが、そのような実施形態は単なる例として提供されていることが、当業者には明白である。多数の変更、改変及び置き換えが、当業者によって、本発明から逸脱することなく生じる。本明細書に記載されている本発明の実施形態に対する様々な代替案を、本発明の実施に用いてもよいことが、理解されるべきである。以下の特許請求の範囲が本発明の範囲を確定すること、ならびにこれらの特許請求の範囲内の方法及び構造、及びその等価物が網羅されることが、意図される。

20

#### 【実施例】

#### 【0131】

実施例1：全エクソームの捕捉及び配列決定：NA12878

NA12878ヒト細胞系のゲノムDNAを、Blue-Pippin DNAサイズ分類系の使用によりサイズに基づいた断片の分離に付して、およそ10kb以上の長さの断片を回収した。次に、サイズに基づいて選択された試料核酸を、マイクロ流体区分化系の使用により、フッ素化油連続相内において水性液滴中でバーコードビーズと共に同時区分化し（例えば、2015年4月9日に出願され、全ての目的においてその全体が参照として本明細書に組み込まれる、米国特許出願第14/682,952号を参照すること）、ここで水性液滴は、dNTP、熱安定性DNAポリメラーゼ及び液滴内で増幅を実施する他の試薬、ならびにビーズからバーコードオリゴヌクレオチドを遊離するDTTも含んだ。これを、1ngの総投入DNA及び2ngの総投入DNAの両方で繰り返した。バーコードビーズを、700,000個を超える異なるバーコード配列のバーコード密度を表す貯蔵ライブラリーのサブセットとして得た。バーコード含有オリゴヌクレオチドは、追加の配列構成成分を含み、一般構造：

30

ビーズ-P5-BC-R1-Nmer

を有した。

#### 【0132】

ここでP5及びR1は、それぞれIllumina結合及びRead1プライマー配列を指し、BCは、オリゴヌクレオチドのバーコード部分を示し、Nmerは、テンプレート核酸の刺激に使用されるランダム10塩基Nmer刺激配列を示す。例えば、全開示が、全ての目的において、その全体が参照として本明細書に組み込まれる、米国特許出願第14/316,383号、2014年6月26日出願を参照すること。

40

#### 【0133】

ビーズを溶解した後、液滴を熱循環させて、それぞれの液滴内の試料核酸のテンプレートに対するバーコードオリゴのプライマー伸長を可能にした。これは、上記に記載された他の含有配列に加えて、由来区分を表すバーコード配列を含む試料核酸の、増幅されたコピー断片をもたらした。

#### 【0134】

コピー断片をバーコード標識化した後、増幅コピー断片を含むエマルジョンの液滴を壊

50

し、配列決定装置に必要な追加の構成成分、例えば、read 2 プライマー配列及び P 7 結合配列を、追加の増幅ステップを介してコピー断片に付加し、これらの配列はコピー断片の他の末端に結合された。次にバーコード付き DNA を、Agilent SureSelect E x o m e 捕捉キットの使用によりハイブリッド捕捉に付した。

#### 【0135】

下記の表は、NA 1 2 8 7 8 ゲノムの標的化の統計を提示する。

【表 1】

試料	挿入物サイズの中央値	標的における断片%	標的に基づいた%
バージョン 1. A	258	81%	51%
バージョン 1. B	224	81%	55%
バージョン 1. C	165	81%	63%

10

#### 【0136】

上記に提示された 3 つの異なるバージョンは、第 2 のアダプター結合ステップの前のバーコード付き断片の 3 つの異なる剪断長さを表す。

#### 【0137】

実施例 2：全エクソームの捕捉及び配列決定：NA 1 9 7 0 1 及び NA 1 9 6 6 1

20

NA 1 9 7 0 1 及び NA 1 9 6 6 1 細胞系のゲノム DNA は、上記の実施例 1 に記載された方法に従って調製した。これらの 2 つの細胞系の、位相データを含むデータを、下記の表に提示する。

【表 2】

	NA19661	NA19701
N50 の位相ブロック	29,535	83,953
N90 の位相ブロック	8,595	25,684
平均位相ブロック	5,968	21,126
位相ブロック中央値	0	76.5
最長位相部録	209,323	504,140
位相した遺伝子の割合	0.719	0.841
完全に位相した遺伝子の割合	0.678	0.778
位相した s n p の割合	0.869	0.832
両方の対立遺伝子にバーコード付けした s n p の割合	0.328	0.351
遺伝子において修正される s n p の確率	0.906	0.927
遺伝子において位相される s n p の確率	0.807	0.889
短い s n p の切換誤差	0.013	0.013
長い s n p の切換誤差	0.012	0.013

30

表 3. 位相距離。図 1 に示されているように、NA 1 9 7 0 1 の断片長さは、NA 1 9 6 6 1 のよりもかなり長く、かなり良好な位相性能をもたらす。

#### 【0138】

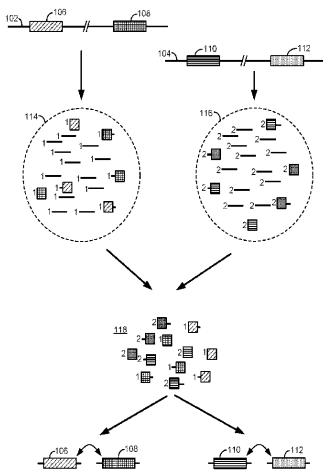
本明細書は、方法論、系及び / または構造の完全な記載、ならびに本明細書に記載されている技術の例示的な態様におけるその使用を提供する。本技術の様々な態様が、ある程度の特定性をもって、または 1 つ以上の独立した態様を参照しながら、上記に記載されてきたが、当業者は、本技術の精神または範囲から逸脱することなく開示されている態様に多数の変更を行うことができる。多くの態様を、本開示の技術の精神または範囲から逸脱することなく行うことができるので、適切な範囲は、本明細書以降に添付されている特許請求の範囲にある。したがって他の態様も考慮される。更に、任意の操作を、特に明確に主張されない限り、または特定の順番が請求項の語句により本質的に必要とならない限り、任意の順番で実施できることが理解されるべきである。上記の記載に含有され、添付の図面に示されている全ての事項は、特定の態様の例示のためのみであり、示されている実施形態を制限するものではないと解釈されることが意図される。文脈から明白ではない、または明確に記述されていない限り、本明細書に提供されている任意の濃度値は、一般に、混合物の特定の成分の添加によって、または添加後に生じる任意の変換を無視している

40

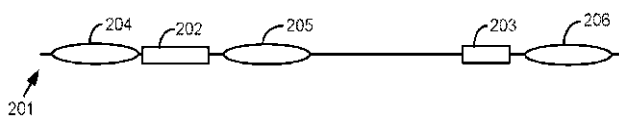
50

混合物の値または百分率として提示されている。既に明白に本明細書に組み込まれていない程度に、本開示において参照されている全ての公開参考文献及び特許文献は、全ての目的においてそれらの全体が参照として本明細書に組み込まれる。詳細または構造における変更は、以下の特許請求の範囲に定義されている本技術の基本的要素から逸脱することなく行うことができる。

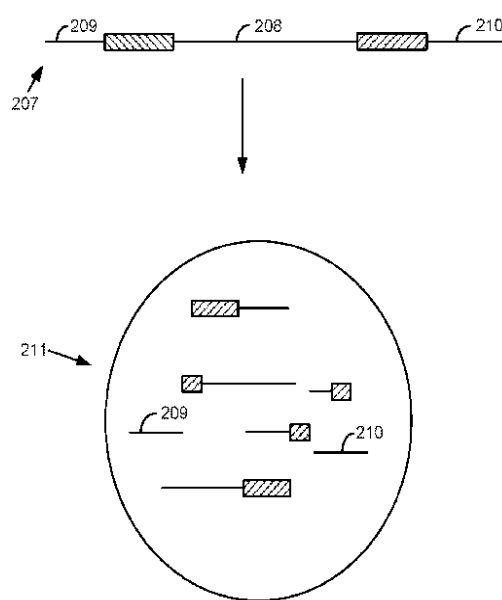
【図 1】



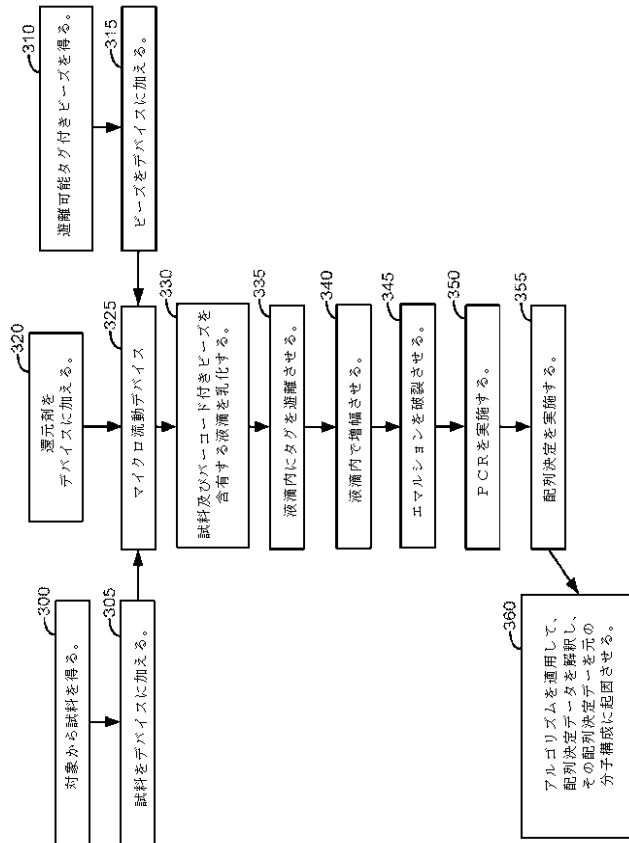
【図 2 A】



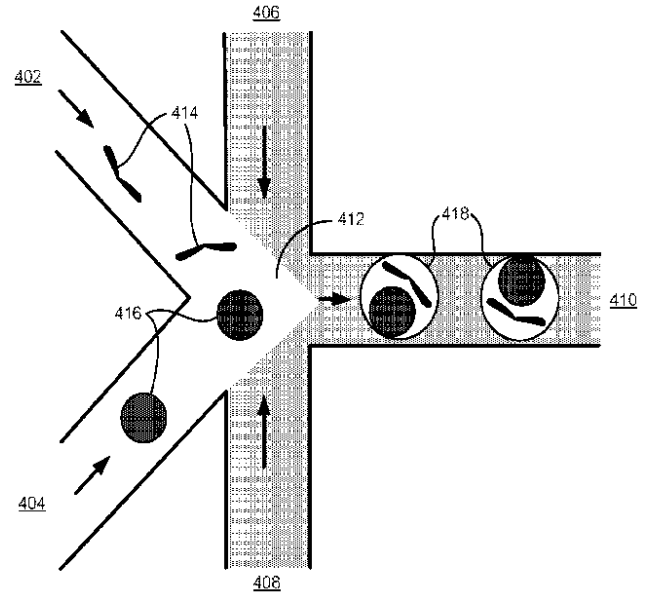
【図 2 B】



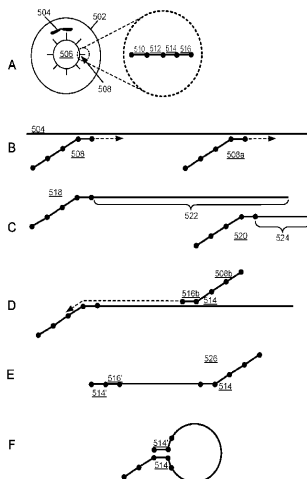
【図 3】



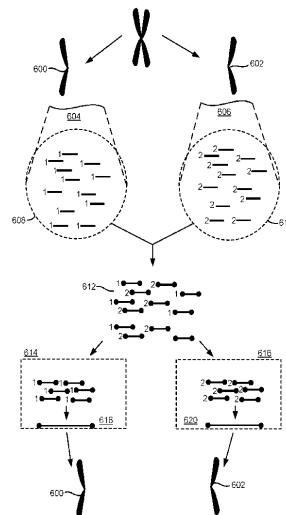
【図 4】



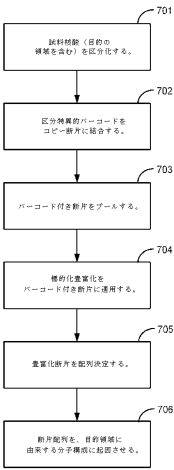
【図 5】



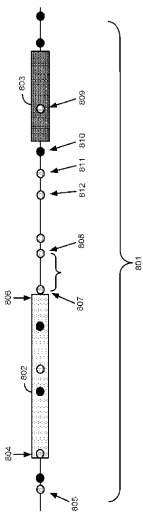
【図 6】



【 図 7 】



【 図 8 】



## 【 国際調査報告 】

## INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/058142

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. C12Q1/68 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) C12Q		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data, BIOSIS, CHEM ABS Data, Sequence Search, EMBASE		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012/106546 A2 (UNIV WASHINGTON CT COMMERCIALI [US]; SHENDURE JAY ASHOK [US]; SCHWARTZ) 9 August 2012 (2012-08-09) example 2, pages 37-45 -----	1-97
X	WO 2013/036929 A1 (UNIV LELAND STANFORD JUNIOR; PUSHKAREV DMITRY [US]; QUAKE STEPHEN R [U]) 14 March 2013 (2013-03-14) claims 1, 25 and 63 -----	1-97
Y	WO 2012/142611 A2 (COMPLETE GENOMICS INC [US]; DRMANAC RADOJE [US]; PETERS BROCK A [US];) 18 October 2012 (2012-10-18) Claim 44 -----	1-97
Y	WO 2014/093676 A1 (10X TECHNOLOGIES INC [US]) 19 June 2014 (2014-06-19) page 38; claims -----	1-97
-/-		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "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		Date of mailing of the international search report
27 January 2016		08/02/2016
Name and mailing address of the ISA/ European Patent Office, P.B. 6818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer
		Hennard, Christophe

1

## INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/058142

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>JIANGFENG HE ET AL:  "Genotyping-by-sequencing (GBS), an  ultimate marker-assisted selection (MAS)  tool to accelerate plant breeding",  FRONTIERS IN PLANT SCIENCE,  vol. 5, 30 September 2014 (2014-09-30),  XP055217380,  DOI: 10.3389/fpls.2014.00484  figure 1</p> <p>-----</p>	1-97
Y	<p>BROCK A. PETERS ET AL: "Accurate  whole-genome sequencing and haplotyping  from 10 to 20 human cells",  NATURE,  vol. 487, no. 7406,  11 July 2012 (2012-07-11), pages 190-195,  XP055209563,  ISSN: 0028-0836, DOI: 10.1038/nature11236  figure 1</p> <p>-----</p>	1-97

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2015/058142

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012106546 A2	09-08-2012	AU 2012212148 A1 CA 2826131 A1 CN 103443338 A EP 2670894 A2 JP 2014506788 A US 2013203605 A1 WO 2012106546 A2	15-08-2013 09-08-2012 11-12-2013 11-12-2013 20-03-2014 08-08-2013 09-08-2012
WO 2013036929 A1	14-03-2013	AU 2012304328 A1 CA 2848304 A1 EP 2753715 A1 GB 2496016 A HK 1199745 A1 US 2013079231 A1 US 2013157870 A1 WO 2013036929 A1	20-03-2014 14-03-2013 16-07-2014 01-05-2013 17-07-2015 28-03-2013 20-06-2013 14-03-2013
WO 2012142611 A2	18-10-2012	AU 2012242508 A1 CA 2832643 A1 CN 103987857 A HK 1201077 A1 JP 2014523233 A US 2013124100 A1 WO 2012142611 A2	02-05-2013 18-10-2012 13-08-2014 21-08-2015 11-09-2014 16-05-2013 18-10-2012
WO 2014093676 A1	19-06-2014	AU 2013359165 A1 CA 2894694 A1 EP 2931919 A1 US 2014206554 A1 US 2015218633 A1 WO 2014093676 A1	02-07-2015 19-06-2014 21-10-2015 24-07-2014 06-08-2015 19-06-2014



## フロントページの続き

(81)指定国 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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, TZ, UA, UG, US

(72)発明者 ジャローズ, ミルナ

アメリカ合衆国, カリフォルニア州 9 4 5 6 6, プレザントン, コール センター パークウェイ 7 0 6 8, スイート 4 0 1

(72)発明者 シュナル - レヴィン, マイケル

アメリカ合衆国, カリフォルニア州 9 4 5 6 6, プレザントン, コール センター パークウェイ 7 0 6 8, スイート 4 0 1

(72)発明者 サクソノブ, サージ

アメリカ合衆国, カリフォルニア州 9 4 5 6 6, プレザントン, コール センター パークウェイ 7 0 6 8, スイート 4 0 1

(72)発明者 ハインドソン, ベンジャミン

アメリカ合衆国, カリフォルニア州 9 4 5 6 6, プレザントン, コール センター パークウェイ 7 0 6 8, スイート 4 0 1

(72)発明者 ジョン, シンイン

アメリカ合衆国, カリフォルニア州 9 4 5 6 6, プレザントン, コール センター パークウェイ 7 0 6 8, スイート 4 0 1

F ターム(参考) 4B029 AA07 FA15 GB10

4B063 QA13 QQ42 QR08 QR48 QR55 QR82 QS13 QS15 QS25 QX02