

(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2018-505688

(P2018-505688A)

(43) 公表日 平成30年3月1日(2018.3.1)

(51) Int.Cl.	F I	テーマコード (参考)
<b>C 1 2 N 15/09 (2006.01)</b>	C 1 2 N 15/00 A	4 B 0 6 3
<b>C 1 2 Q 1/68 (2018.01)</b>	C 1 2 Q 1/68 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2017-543973 (P2017-543973) (86) (22) 出願日 平成28年2月24日 (2016.2.24) (85) 翻訳文提出日 平成29年9月4日 (2017.9.4) (86) 国際出願番号 PCT/US2016/019382 (87) 国際公開番号 W02016/138148 (87) 国際公開日 平成28年9月1日 (2016.9.1) (31) 優先権主張番号 62/146,834 (32) 優先日 平成27年4月13日 (2015.4.13) (33) 優先権主張国 米国 (US) (31) 優先権主張番号 62/119,996 (32) 優先日 平成27年2月24日 (2015.2.24) (33) 優先権主張国 米国 (US)	(71) 出願人 516380360 1 O エックス ゲノミクス, インコーポレイテッド アメリカ合衆国, カリフォルニア州 94566, プレザントン, コール センター パークウェイ 7068, スイート 401 (74) 代理人 100079108 弁理士 稲葉 良幸 (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
---	--

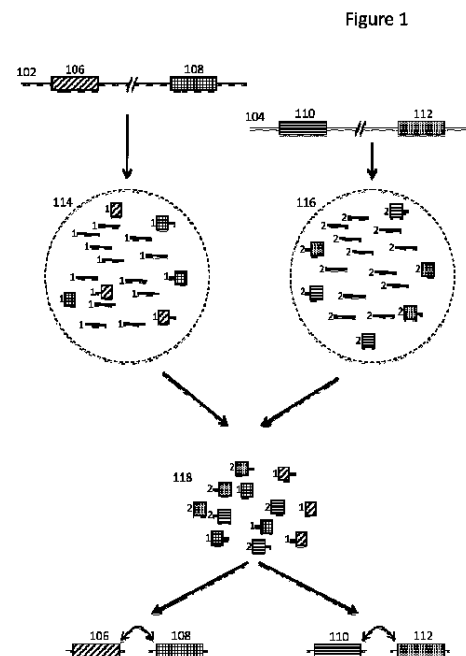
最終頁に続く

(54) 【発明の名称】 標的化核酸配列包括度 (COVERAGE) のための方法

## (57) 【要約】

本発明は、ゲノムの標的化領域の配列情報を分析する方法、組成物及び系を対象とする。そのような標的化領域には、不十分に特徴決定されている、高度に多形性である、または基準ゲノム配列から逸脱している、ゲノムの領域が含まれ得る。

【選択図】図 1



**【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、それぞれの分離区分が個別の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 ) 前記分離区分中の前記個別の核酸分子の少なくともいくつかの選択部分を増幅して、アンプリコンの個体群を形成することと、

( d ) 前記アンプリコンの個体群にバーコードを付けて、前記アンプリコンの複数のバーコード付き断片を形成し、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、

10

( e ) 前記複数の断片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】**

前記ゲノムの前記 1 つ以上の選択部分が、前記ゲノムの高度に多形性の領域を含む、請求項 1 に記載の方法。

**【請求項 3】**

前記ゲノムの前記 1 つ以上の選択部分の前記配列決定が、新規の配列決定である、請求項 1 ~ 2 に記載の方法。

20

**【請求項 4】**

前記増幅が、少なくとも 3 . 5 メガベース対 ( M b ) の範囲にわたる P C R 増幅を含む、請求項 1 ~ 3 に記載の方法。

**【請求項 5】**

前記増幅が、少なくとも 3 . 0 M b の領域にわたってねじれている複数のプライマー対を利用する P C R 増幅を含む、請求項 1 ~ 4 の記載の方法。

**【請求項 6】**

前記複数のプライマー対が、前記プライマー配列の増幅を防止するためにウラシルを含有する、請求項 5 に記載の方法。

30

**【請求項 7】**

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、請求項 1 ~ 6 に記載の方法。

**【請求項 8】**

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 7 に記載の方法。

**【請求項 9】**

取得ステップ ( e ) において生成される前記配列情報が、由来する個別の核酸の分子構成を保持する、請求項 1 ~ 8 に記載の方法。

40

**【請求項 10】**

前記取得ステップ ( e ) の前に、前記複数の断片は、

( 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こと、

( i i ) プローブ断片複合体を固体支持体の表面に捕捉すること

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がさらに豊富化される、請求項 1 ~ 9 に記載の方法。

**【請求項 11】**

前記固体支持体がビーズを含む、請求項 10 に記載の方法。

**【請求項 12】**

50

前記方法が、前記複数の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さらに含み、前記推定コンティグが、少なくとも10 kbのN50長さを含み、請求項1～11に記載の方法。

【請求項13】

前記推定コンティグが、少なくとも20 kbのN50長さを含み、請求項12に記載の方法。

【請求項14】

前記推定コンティグが、少なくとも40 kbのN50長さを含み、請求項12に記載の方法。

【請求項15】

前記推定コンティグが、少なくとも50 kbのN50長さを含み、請求項12に記載の方法。

【請求項16】

前記推定コンティグが、少なくとも100 kbのN50長さを含み、請求項12に記載の方法。

【請求項17】

前記推定コンティグが、少なくとも200 kbのN50長さを含み、請求項12に記載の方法。

【請求項18】

前記バーコード付き断片の前記バーコードが、追加の配列セグメントをさらに含み、請求項1～17に記載の方法。

【請求項19】

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18に記載の方法。

【請求項20】

前記バーコード付けが、少なくとも700,000個のバーコードのライブラリーから選択されるバーコードを結合することを含み、請求項1～19に記載の方法。

【請求項21】

前記分離区分内の前記アンプリコンの前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の約100x～5000xの包括度を表す、請求項1～20に記載の方法。

【請求項22】

前記分離区分内の前記アンプリコンの前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の約200x～1000xの包括度を表す、請求項1～20に記載の方法。

【請求項23】

前記分離区分内の前記アンプリコンの前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも1000xの包括度を表す、請求項1～20に記載の方法。

【請求項24】

前記分離区分内の前記アンプリコンの前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも2000xの包括度を表す、請求項1～20に記載の方法。

【請求項25】

前記分離区分内の前記アンプリコンの前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも5000xの包括度を表す、請求項1～20に記載の方法。

【請求項26】

ゲノム試料の1つ以上の不十分に特徴決定された部分から配列情報を取得する方法であっ

10

20

30

40

50

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) 不十分に特徴決定された前記複数の第2の断片の選択領域を増幅して、アンプリコンの個体群を形成することと、

(d) 共通のバーコード配列を、前記アンプリコンのそれぞれが、それらが含有されている前記分離区分に属するように、それぞれの分離区分内の前記アンプリコンに結合することと、

(e) 前記アンプリコンの配列を特定し、それによって、前記ゲノム試料の1つ以上の不十分に特徴決定された部分から配列情報を取得することとを含む、前記方法。 10

【請求項27】

前記増幅が、少なくとも3.5メガベース対(Mb)の範囲にわたるPCR増幅を含む、請求項26に記載の方法。

【請求項28】

前記増幅が、少なくとも3.0Mbの領域にわたってねじれている複数のプライマー対を利用するPCR増幅を含む、請求項26～27の記載の方法。

【請求項29】

前記複数のプライマー対が、前記プライマー配列の増幅を防止するためにウラシルを含有する、請求項28に記載の方法。 20

【請求項30】

前記特定ステップ(e)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配列決定反応を含む、請求項26～29に記載の方法。

【請求項31】

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30に記載の方法。

【請求項32】

前記特定ステップ(e)が、前記特定が、同じ個別の第1の核酸断片分子から誘導されるアンプリコンを特定することをさらに含むように、前記アンプリコン配列の前記分子構成を保存する、請求項26～31に記載の方法。 30

【請求項33】

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第1の断片分子を連結することをさらに含み、前記推定コンティグが、少なくとも10kbのN50長さを含む、請求項26～32に記載の方法。

【請求項34】

前記推定コンティグが、少なくとも20kbのN50長さを含む、請求項33に記載の方法。

【請求項35】

前記推定コンティグが、少なくとも40kbのN50長さを含む、請求項33に記載の方法。 40

【請求項36】

前記推定コンティグが、少なくとも50kbのN50長さを含む、請求項33に記載の方法。

【請求項37】

前記推定コンティグが、少なくとも100kbのN50長さを含む、請求項33に記載の方法。

【請求項38】

前記推定コンティグが、少なくとも200kbのN50長さを含む、請求項33に記載 50

の方法。

【請求項 39】

前記バーコード配列が、追加の配列セグメントをさらに含む、請求項 26 ~ 38 に記載の方法。

【請求項 40】

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 39 に記載の方法。

【請求項 41】

前記結合ステップ (d) が、少なくとも 700,000 個のバーコードのライブラリーから選択されるバーコードを結合することを含む、請求項 26 ~ 40 に記載の方法。

【請求項 42】

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 26 ~ 41 に記載の方法。

【請求項 43】

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 26 ~ 41 に記載の方法。

【請求項 44】

前記分離区分が、エマルションの液滴を含む、請求項 26 ~ 43 に記載の方法。

【請求項 45】

前記特定ステップ (e) の前に、前記アンプリコンが、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さらに増幅される、請求項 26 ~ 44 に記載の方法。

【請求項 46】

前記分離区分内の前記バーコード付きアンプリコンが、前記ゲノムの前記 1 つ以上の不十分に特徴決定された部分の約  $100 \times \sim 5000 \times$  の包括度を表す、請求項 26 ~ 45 に記載の方法。

【請求項 47】

前記分離区分内の前記バーコード付きアンプリコンが、前記ゲノムの前記 1 つ以上の不十分に特徴決定された部分の約  $200 \times \sim 1000 \times$  の包括度を表す、請求項 26 ~ 45 に記載の方法。

【請求項 48】

前記分離区分内の前記バーコード付きアンプリコンが、前記ゲノムの前記 1 つ以上の不十分に特徴決定された部分の少なくとも  $1000 \times$  の包括度を表す、請求項 26 ~ 45 に記載の方法。

【請求項 49】

前記分離区分内の前記バーコード付きアンプリコンが、前記ゲノムの前記 1 つ以上の不十分に特徴決定された部分の少なくとも  $2000 \times$  の包括度を表す、請求項 26 ~ 45 に記載の方法。

【請求項 50】

前記分離区分内の前記バーコード付きアンプリコンが、前記ゲノムの前記 1 つ以上の不十分に特徴決定された部分の少なくとも  $5000 \times$  の包括度を表す、請求項 26 ~ 45 に記載の方法。

【請求項 51】

分子構成を保持しながら、ゲノム試料の 1 つ以上の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

10

20

30

40

50

た個体群を提供することと、

(d) 共通のバーコード配列を、前記断片のそれぞれが、それらが含有されていた前記分離区分に属するように、それぞれの分離区分内の前記断片に結合することと、

(e) 前記断片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とを含む、前記方法。

【請求項52】

前記提供ステップ(c)が、前記ゲノムの前記1つ以上の選択部分の配列を含有する前記断片の少なくとも一部のPCR増幅を含む、請求項51に記載の方法。

【請求項53】

前記ゲノムの前記1つ以上の選択部分が、前記ゲノムの少なくとも3.0Mbの長さの近接領域を含む、請求項51～52に記載の方法。

【請求項54】

前記取得ステップ(e)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配列決定反応を含む、請求項51～53に記載の方法。

【請求項55】

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54に記載の方法。

【請求項56】

前記特定ステップ(e)が、前記特定が、同じ個別の第1の核酸断片分子から誘導される断片を特定することをさらに含むように、前記断片配列の前記分子構成を保存する、請求項51～55に記載の方法。

【請求項57】

前記方法が、前記複数の第2の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第1の断片分子を連結することをさらに含み、前記推定コンティグが、少なくとも10kbのN50長さを含む、請求項51～56に記載の方法。

【請求項58】

前記推定コンティグが、少なくとも20kbのN50長さを含む、請求項57に記載の方法。

【請求項59】

前記推定コンティグが、少なくとも40kbのN50長さを含む、請求項57に記載の方法。

【請求項60】

前記推定コンティグが、少なくとも50kbのN50長さを含む、請求項57に記載の方法。

【請求項61】

前記推定コンティグが、少なくとも100kbのN50長さを含む、請求項57に記載の方法。

【請求項62】

前記推定コンティグが、少なくとも200kbのN50長さを含む、請求項57に記載の方法。

【請求項63】

前記バーコード配列が、追加の配列セグメントをさらに含む、請求項51～62に記載の方法。

【請求項64】

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63に記載の方法。

【請求項65】

前記結合ステップ(d)が、少なくとも700,000個のバーコードのライブラリーから選択されるバーコードを結合することを含む、請求項51～64に記載の方法。

【請求項66】

それぞれの分離区分中の前記ゲノム材料が、単一細胞のゲノムDNAを含む、請求項51～65に記載の方法。

【請求項67】

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51～65に記載の方法。

【請求項68】

前記分離区分が、エマルションの液滴を含む、請求項51～67に記載の方法。

10

【請求項69】

前記取得ステップ(e)の前に、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さらに増幅される、請求項51～68に記載の方法。

【請求項70】

前記提供ステップ(c)が、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前記断片のPCR増幅を含む、請求項51～69に記載の方法。

【請求項71】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の約100x～5000xの包括度を表す、請求項51～70に記載の方法。

20

【請求項72】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の約200x～1000xの包括度を表す、請求項51～70に記載の方法。

【請求項73】

前記分離区分内の前記アンプリコンの前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも1xの包括度を表す、請求項51～70に記載の方法。

【請求項74】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも2000xの包括度を表す、請求項51～70に記載の方法。

【請求項75】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも5000xの包括度を表す、請求項51～70に記載の方法。

30

【請求項76】

分子構成を保持しながら、ゲノム試料の1つ以上の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の配列が豊富化されている少なくともいくつかの前記分離区分内に、個体群を提供することと、

40

(d)共通のバーコード配列を、前記断片のそれぞれが、それらが含有されていた前記分離区分に属するように、それぞれの分離区分内の前記断片に結合することと、

(e)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含有する分離区分を、前記ゲノムの前記1つ以上の選択部分を含む断片を含有しない分離区分から分別することと、

(f)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前記断片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と

を含む、前記方法。

【請求項77】

50

前記提供ステップ(c)が、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アンプリコンの個体群を産生する、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前記断片の定方向PCR増幅を含む、請求項76に記載の方法。

【請求項78】

前記提供ステップ(c)が、検出可能標識を前記アンプリコンに結合することをさらに含む、請求項77に記載の方法。

【請求項79】

前記分別ステップ(e)が、前記検出可能標識のシグナルを放出する前記区分を、そのような信号を有さない前記区分から選別することを含む、請求項78に記載の方法。

【請求項80】

前記検出可能標識が蛍光分子を含む、請求項78～79に記載の方法。

【請求項81】

前記取得ステップ(f)の前に、前記分離区分が合わされ、前記断片と一緒にプールされる、請求項76～80に記載の方法。

【請求項82】

前記特定ステップ(f)が、前記特定が、同じ個別の第1の核酸断片分子から誘導される断片を特定することをさらに含むように、前記断片配列の前記分子構成を保存する、請求項76～81に記載の方法。

【請求項83】

前記特定ステップ(f)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配列決定反応を含む、請求項76～82に記載の方法。

【請求項84】

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83に記載の方法。

【請求項85】

前記分離区分が、エマルションの液滴を含む、請求項76～84に記載の方法。

【請求項86】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の約100×～5000×の包括度を表す、請求項76～85に記載の方法。

【請求項87】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の約200×～1000×の包括度を表す、請求項76～85に記載の方法。

【請求項88】

前記分離区分内の前記アンプリコンの前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも1000×の包括度を表す、請求項76～85に記載の方法。

【請求項89】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも2000×の包括度を表す、請求項76～85に記載の方法。

【請求項90】

前記分離区分内の前記バーコード付き断片が、前記ゲノムの前記1つ以上の選択部分の少なくとも5000×の包括度を表す、請求項76～85に記載の方法。

【請求項91】

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

(a)ゲノム材料を準備することと、

(b)個別の核酸分子を前記ゲノム材料から分別して、分別された個別の核酸分子を形成することと、

(c)前記分別された個別の核酸分子からの前記ゲノムの前記1つ以上の選択部分の少

10

20

30

40

50



なくとも一部を含む断片が豊富化された個体群を提供し、少なくとも複数の前記断片が、それらが誘導される前記個別の核酸分子に属することと、

(d) 前記断片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とを含む、前記方法。

【請求項92】

前記分別ステップ(b)が、1つ以上の個別の核酸分子を分離区分に分配することを含む、請求項91に記載の方法。

【請求項93】

前記提供ステップ(c)が、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アンプリコンの個体群を産生する、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前記核酸分子の定方向増幅を含む、請求項91～92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、2015年4月13日出願の米国特許仮出願第62/146,834号及び2015年2月24日出願の米国特許仮出願第62/119,996号の利益を主張し、これらは、その全体が全ての目的において参照として本明細書に組み込まれる。

【背景技術】

【0002】

配列決定技術における顕著な進歩にもかかわらず、約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マルチメガベース(multi-megabase)異質染色性間隙と注釈する。この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。ゲノム全体の新規の配列決定は経済的に実現可能ではなく、したがって、ゲノムの配列決定に関連する費用を低減し、同時に大規模なゲノム分析の利益を保持する必要性が、依然として存在している。

【発明の概要】

【0003】

したがって、本開示は、選択された領域の新規配列会合を可能にする、いくつかの態様では、その新規包括度とゲノムの残りの領域の再配列決定との組み合わせを、高い処理量及び高い精度で可能にする、ゲノムの選択領域の標的化包括度を提供する方法、系及び組成物を提供する。

【0004】

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、(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、それぞれの分離区分が個別の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)分離区分内の少なくともいくつかの個別の核酸分子の選択された部分を増幅して、アンプリコンの個体群を形成するステップ、(d)アンプリコンの個体群にバーコード付けして、アンプリコンの複数のバーコード付き断片を形成し、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、それが誘導される個別の核酸分子と関連づけるステップ、(e)複数の断片から配列情報を取得して、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。

【0005】

更なる実施形態において、上記によると、ゲノムの1つ以上の選択部分は、ゲノムの高度に多形性の領域を含む。なお更なる実施形態において、ゲノムの1つ以上の選択部分の配列決定は、新規の配列決定である。

【0006】

10

20

30

40

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、増幅は、少なくとも3.5メガベース対(Mb)の領域にわたるPCR増幅を含む。なお更なる実施形態において、増幅は、少なくとも3.0Mbの領域にわたってねじれている複数のプライマー対を利用するPCR増幅を含む。

【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、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。更なる実施形態において、取得ステップにおいて生成される配列情報は、由来する個別の核酸の分子構成を保持する。

【0008】

特定の実施形態において、上記のうちのいずれかによると、取得ステップの前に、複数の断片は、(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するステップによって、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さらに豊富化される。

【0009】

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、分離区分内のアンプリコンのバーコード付き断片は、ゲノムの1つ以上の選択部分の約100x~5000xの包括度を表す。更なる実施形態において、分離区分内のアンプリコンのバーコード付き断片は、ゲノムの1つ以上の選択部分の約200x~1000xの包括度を表す。なお更なる実施形態において、分離区分内のアンプリコンのバーコード付き断片は、ゲノムの1つ以上の選択部分の少なくとも1000xの包括度を表す。なお更なる実施形態において、分離区分内のアンプリコンのバーコード付き断片は、ゲノムの1つ以上の選択部分の少なくとも2000xまたは5000xの包括度を表す。

【0010】

更なる態様において、本開示は、ゲノム試料の1つ以上の不十分に特徴決定された部分から配列情報を取得する方法であって、(a)分離区分内のゲノム試料の個別の第1の核酸断片分子を準備するステップ、(b)分離区分内の個別の第1の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)不十分に特徴決定されている複数の第2の断片の選択領域を増幅して、アンプリコンの個体群を形成するステップ、(d)共通バーコード配列を、それぞれのアンプリコンが、それが含有されている分離区分に属させるように、それぞれの分離区分内のアンプリコンに結合するステップ、(e)アンプリコンの配列を特定し、それによって、ゲノム試料の1つ以上の不十分に特徴決定された部分から配列情報を取得するステップを含む方法を提供する。

【0011】

特定の実施形態において、上記のうちのいずれかによると、増幅は、少なくとも3.5メガベース対(Mb)の領域にわたるPCR増幅を含む。更なる実施形態において、増幅は、少なくとも3.0Mbの領域にわたってねじれている複数のプライマー対を利用するPCR増幅を含む。なお更なる実施形態において、複数のプライマー対は、プライマー配列の増幅を防止するためにウラシルを含有する。

【0012】

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、特定ステップは、特定することが、同じ個別の第1の核酸断片分子から誘導されるアンプリコンを特定することをさらに含むように、アンプリコン配列の分子構成を保存する。更なる実施形態において、方法は、複数の第2の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つ以上の個別の第1の断片分子を連結することをさらに含み、ここで推定コンティグは、少なくとも10kbのN50長さを含む。

【0013】

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、バーコード配列は、追加の配列セグメントをさらに含む。更なる実施形態において、追加の配列セグメントは、

10

20

30

40

50

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。なお更なる実施形態において、バーコードは、少なくとも700,000個のバーコードのライブラリーから選択される。

【0014】

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、それぞれの分離区分内のゲノム試料は、単一細胞のゲノムDNAを含む。更なる実施形態において、それぞれの分離区分は、異なる染色体のゲノムDNAを含む。

【0015】

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、分離区分は、エマルジョンの液滴を含む。

【0016】

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、分離区分内のバーコード付きアンプリコンは、ゲノムの1つ以上の不十分に特徴決定された部分の約1000x~5000xの包括度を表す。

【0017】

更なる態様において、本出願は、分子構成を保持しながら、ゲノム試料の1つ以上の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ステップ、(b)出発ゲノム材料の個別の核酸分子を、それぞれの分離区分が第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)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、(d)共通バーコード配列を、それぞれの断片が、それが含有されていた分離区分に属させるように、それぞれの分離区分内の断片に結合するステップ、(e)断片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るステップを含む方法を提供する。

【0018】

なお更なる態様において、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ステップ、(b)出発ゲノム材料の個別の核酸分子を、それぞれの分離区分が第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)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の配列が豊富化されている少なくともいくつかの分離区分内に、個体群を提供するステップ、(d)共通バーコード配列を、それぞれの断片が、それが含有されていた分離区分に属させるように、それぞれの分離区分内の断片に結合するステップ、(e)ゲノムの1つ以上の選択部分少なくとも一部を含む断片を含有する分離区分を、ゲノムの1つ以上の選択部分を含む断片を含有しない分離区分から分別するステップ、(f)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るステップを含む方法を提供する。

【0019】

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、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の配列が豊富化されている個体群を提供することは、ゲノムの1つ以上の選択部分の少なくとも一部を含むアンプリコンの個体群を産生する、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の定方向PCR増幅を含む。なお更なる実施形態において、この提供ステップは、検出可能標識をアンプリコンに結合することをさらに含み、いくつかの実施形態では、これには蛍光分子が含まれ得る。なお更なる実施形態において、ゲノムの1つ以上の選択部分を含む断片を含有しない分離区分から、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含有する分離区分を分別するステップは、検出可能標識からのシグナルを発する区分を、そのようなシグナルを有さない区分から選別することを含む。

【0020】

10

20

30

40

50

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、断片から配列情報を取得する前に、分離区分は合わされ、断片は一緒にプールされる。更なる実施形態において、断片から配列情報を取得するステップは、確定することが同じ第1の個別の核酸分子から誘導される断片を確定することをさらに含むので、断片の配列の分子構成を維持するように実施される。なお更なる実施形態において、この配列情報の取得は、短い読み込み長さ (read-length) 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。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。

#### 【0021】

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、分離区分は、エマルジョンの液滴を含む。更なる実施形態において、分離区分内のバーコード付き断片は、ゲノムの1つ以上の選択部分の約  $100 \times \sim 5000 \times$  の包括度を表す。なお更なる実施形態において、分離区分内のバーコード付き断片は、ゲノムの1つ以上の選択部分の約  $200 \times \sim 1000 \times$  の包括度を表す。なお更なる実施形態において、分離区分内のアンブリコンのバーコード付き断片は、ゲノムの1つ以上の選択部分の少なくとも  $1000 \times$  の包括度を表す。なお更なる実施形態において、分離区分内のバーコード付き断片は、ゲノムの1つ以上の選択部分の少なくとも  $2000 \times$  または  $5000 \times$  の包括度を表す。

#### 【0022】

いくつかの態様において、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つ以上の部分から配列情報を取得する方法であって、(a)ゲノム材料を準備するステップ、(b)個別の核酸分子をゲノム材料から分別して、分別された個別の核酸分子を形成するステップ、(c)分離された個別の核酸分子からのゲノムの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を含む方法を提供する。特定の実施形態において、分別ステップは、1つ以上の核酸分子が他の1つ以上の核酸分子に対して相対的に単離されるように選別及び処理されることを可能にする任意の方法を使用して達成される。いくつかの実施形態において、分別とは、基質上の異なる区分または分離区分に物理的に分別することである。更なる実施形態において、少なくとも複数の断片は、それらが誘導される個別の核酸分子に属する。属性は、特定の個別の核酸分子に由来するように個別の断片を指定することを可能にする、任意の方法を使用して得られる。特定の例示的な実施形態において、この属性は、断片にバーコード付けすることによって得られる。更なる態様では、配列情報が断片から取得され、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る。

#### 【図面の簡単な説明】

#### 【0023】

【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2】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クセスを実施する、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6A】配列データを由来供給源の核酸分子に属させる、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6B】配列データを由来供給源の核酸分子に属させる、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1つの実施形態の略図を提供する。

【図8】本発明の1つの実施形態の略図を提供する。

10

20

30

40

50

【図 9】テンプレートを含有しないもの（NTC）と比較した、テンプレートを用いて実施した増幅反応の比較実験のデータを示す。

【図 10】一連のアニーリング温度にわたって実施した増幅反応のデータを示す。

【発明を実施するための形態】

【0024】

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) Biochemistry (4th Ed.) Freeman, New York、Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London、Nelson and Cox (2000), Lehninger, Principles of Biochemistry 3rd Ed., W.H. Freeman Pub., New York, N.Y., ならびに Berg et al. (2002) Biochemistry, 5th 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。

10

20

【0025】

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。

30

【0026】

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得る、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。

40

【0027】

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得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。

【0028】

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

50

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。

【0029】

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。

10

【0030】

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」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「X」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が、必ずしも明確に記述されているとは限らないことも理解されるべきである。

20

【0031】

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、追加の多重複（redundant）配列情報がゲノムのこれらの選択部分から取得され得るように、ゲノムの選択部分の増加した多重複包括度を提供する。特定の場合において、この追加の配列情報は、ゲノムのこれらの選択部分の新規配列決定を可能にするのに十分な情報を提供する。

30

【0032】

一般に、本明細書に記載されている方法、組成物及び系は、ゲノムの選択領域の遺伝子特徴決定を提供する。この遺伝子特徴決定は、ゲノムの選択領域の新規配列決定を可能にするのに十分な深さのものである。この新規配列決定は、不十分に特徴決定されている、高度に多形性がある、及び/または基準配列から逸脱している、ゲノム領域において特定の使用される。理解されるように、有意な率（例えば、Altmeose 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、マルチメガベース異質染色性間隙と注釈する。この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、構造的な特徴が含まれる。正確な特徴決定に抵抗性がある例示的な領域には、密接な相同性がある偽遺伝子（例えば、SMN1/2 CYP2D6）を有する区域、トランスポゾン（SINE、LINEなど）が限定されることなく含まれる、ゲノム全体にわたって実質的に反復している配列を有する区域、及び基準配列が不十分な先導としてしか機能しない、途方もない変動を有する特定の区域（ヒト白血球抗原（HLA）複合体の遺伝子をコードする領域など）が含まれる。本明細書に記載されている方法、組成物及び系は、目的の領域の選的増幅を、分子構成を維持する能力と組み合わせ、それによって、一般的に不十分に特徴決定されているゲノム領域の新規配列決定を可能にし、同様に、より大きなゲ

40

50

ノムにおけるこれらの領域の長い範囲の分子構成を最適に提供することを可能にする。

【0033】

特定の場合において、本明細書に記載されている方法は、ゲノムの選択領域が配列決定される前に選択的に増幅されるステップを含む。この増幅は、一般に当該技術に既知の方法（PCR増幅が限定されることなく含まれる）を使用して実施され、ゲノムの選択領域の少なくとも1x、10x、20x、50x、100x、200x、500x、1000x、1500x、2000x、5000x、または10000xの包括度を提供し、それによって、これらの選択領域の新規配列決定を可能にする量の核酸を提供する。更なる実施形態において、増幅は、ゲノムの選択領域の少なくとも1x~20x、50x~100x、200x~1000x、1500x~5000x、5000x~10,000x、10000x~100000x、1500x~9000x、2000x~8000x、2500x~7000x、3000x~6500x、3500x~6000x、4000x~5500xの包括度を提供する。

10

【0034】

増幅は、一般に、ゲノムの選択領域の中の、または近くの配列に相補的なプライマーの伸長にわたって実施される。いくつかの場合において、目的の領域にわたってタイル状になるように設計されているプライマーライブラリーが使用され、換言すると、プライマーライブラリーは、ゲノムの選択領域に沿って特定の距離で領域を増幅するように設計される。いくつかの場合において、選択的増幅は、ゲノムの選択領域に沿って、10、15、20、25、50、100、200、250、500、750、1000、または10000塩基毎に相補性があるプライマーを利用する。なお更なる例において、タイル状のプライマー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得る、または選択領域の特定の部分もしくは割合が異なるプライマー対により増幅されるように、知的に設計され得る。

20

【0035】

一般に、本明細書に記載されている方法及び系は、ゲノムの選択領域の配列の決定を提供することによって標的化ゲノム配列決定を達成し、この配列決定情報は、極めて低い配列決定誤差率及び高い処理量の短い読み込み配列決定技術という利点を有する方法を使用して得られる。

【0036】

核酸の配列決定は、配列読み込みの分子構成または配列読み込みの一部を保存する方法で典型的に実施される。このことは、多重配列読み込み、または配列読み込みの複数の部分が、核酸の単一由来分子に属し得ることを意味する。「属する」とは、配列読み込みが核酸の特定の由来分子の塩基の直線状配列内にあると特定され得ることを意味し、換言すると、断片1及び2が由来核酸分子Aから生成される場合、配列決定は、断片1、2、3及び4の配列読み込みが分子構成を保持するように実施され、断片1及び2は、由来分子Aから誘導されることが容易に確認されることを意味する。

30

【0037】

この核酸の単一分子は、様々な長さの任意のものであり得るが、好ましい態様において、長い範囲の分子構成の保存を可能にするように、比較的長い分子である。特に、単一由来分子は、好ましくは、典型的な短い読み込み配列長さより実質的に長いもの、例えば、200塩基を超えるものであり、多くの場合、少なくとも1000塩基以上、5000塩基以上、10,000塩基以上、20,000塩基以上、30,000塩基以上、40,000塩基以上、50,000塩基以上、60,000塩基以上、70,000塩基以上、80,000塩基以上、90,000塩基以上、または100,000塩基以上、いくつかの場合では、最長1メガベース以上である。

40

【0038】

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、分子構成保持しながら、核酸、特にゲノムの選択領域の核酸を特徴決定することができる。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

50

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（例えば、同じ核酸分子に由来している）が、いくつかの場合では、これらの領域は、互いに相対的に分別されていることがあり、例えば、1 kbを超える間隔、5 kbを超える間隔、10 kbを超える間隔、20 kbを超える間隔、30 kbを超える間隔、40 kbを超える間隔、50 kbを超える間隔、いくつかの場合では100 kb、またはそれ以上の間隔であり得る。目的の領域は、一般にゲノムの分離及び分別部分であり、いくつかの場合では、そのような領域は不十分に特徴決定された領域である。目的の領域は、また、個別の遺伝子、遺伝子群、エクソンを示すこともある。示されているように、それぞれの核酸102及び104は分別されている。図1に例示されているように、それぞれの核酸は、それ自体の区分114及び116にそれぞれ分別されているが、理解されるように、本明細書に記載されている方法は、そのような区分における使用に限定されず、核酸分子を分別する任意の方法を使用することができ、これらの分別された核酸分子を、本明細書に開示されている任意の方法に従ってさらに処理することができる。本明細書の他の箇所に示されているように、図1の114及び116などの区分は、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。そのような分子構成は、断片を、それが誘導された元の核酸分子に属させるのを可能にする任意の方法を使用して、保存することができる。図1に示されているように、これが達成される1つの方法は、由来の断片が区分された液滴を表しているバーコード配列、例えば例示されているバーコード配列「1」または「2」を、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を分析して、ゲノムの科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。分子構成の保存を助ける他の配列決定方法には、Pacific Biosciencesから入手可能なSMRT配列決定、例えばOxford Nanoporeにより記載されているナノポア配列決定及びIllumina, Inc.から入手可能なTruseq SLRプロセスなどの単一分子配列決定プロセスが含まれる。

#### 【0039】

上記によると、バーコード付けステップに加えて、核酸102または104が目的の選択ゲノム領域を含有する場合、これらの領域のアンプリコンが区分114及び116それぞれにおける断片の大きな割合を形成するように、1つ以上の選択的増幅のステップがあってもよい。この増幅ステップは、一般に、本明細書に記載されている方法によるバーコードの結合の前、または結合と同時に実施されるが、いくつかの実施形態において、増幅ステップは、バーコードの結合に続いて生じることもある。

#### 【0040】

ライブラリー118内のプールされた断片が、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（時には、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により、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る。そ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ができる。選択的増幅が、ゲノムの選択領域の部分を含む核酸断片の量を増加するために利用される状況では、分子構成を特定する能力は、ゲノムの選択領域を、多くの場合にこれらの領域の新規会合を可能にする深さの包括度で配列決定する方法も提供する。

#### 【0041】

特定の状況において、本明細書に記載されている配列決定方法は、選択領域の深い包括

10

20

30

40

50



度と、ゲノムの長い範囲にわたる低レベルの連結読み込みとの組み合わせを含む。理解されるように、この新規と再配列決定との組み合わせは、ゲノム全体及び／またはゲノム大きな部分を配列決定する効率的な方法を提供する。本明細書に記載されている選択的増幅方法による、不十分に特徴決定されている及び／または高度に多形性の領域の標的化包括度は、これらの領域の新規配列会合に必要な包括度レベルで核酸材料の量を提供し、一方、ゲノムの他の領域にわたる連結ゲノム配列決定は、分子構成の保存を介して一緒に連結している分離領域について配列情報を提供することによって、残りのゲノムの高い処理量での分析を可能にする。本明細書に記載されている方法及び組成物は、同じ配列決定プラットフォーム及び配列決定ライブラリーを両方の種類の包括度に使用できるので、新規と連結読み込み配列決定との組み合わせを可能にすることに独自に適応可能である。本明細書に記載されている方法に従って配列決定される核酸及び／または核酸断片の個体群は、新規配列決定用のゲノム領域及び再配列決定用のゲノム領域の両方からの配列を含有し、新規配列決定用の目的の領域を網羅する核酸の割合は、本明細書においてさらに詳細に記載される標的化増幅方法に起因して、ゲノムの他の領域を網羅する核酸よりも多い。そのような方法は、本明細書に記載されている方法が会合の際に位相情報の保持を可能にするので、ハプロタイプの新規会合にさらに適応可能である。

10

20

30

40

50

#### 【0042】

ゲノムの選択領域から配列情報を取得する能力を提供することに加えて、本明細書に記載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、米国特許出願第14/752,589号及び同第14/752,602号に記載されており、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。

#### 【0043】

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、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号及び同第14/316,463号においてもさら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、ならびにゲノム材料の配列決定及び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。

#### 【0044】

一般に、本発明の方法は、本明細書においてさら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2に例示されているステップを含む。理解されるように、図2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、変更及び修正され得る例示的な実施形態である。

#### 【0045】

図2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、目的の標的化領域を含有する試料核酸が、例えば区分に分別されるステップ(201)を含む。一般に、目的のゲノム領域の核酸を含有する区分は、それぞれ標的化豊富化を受けて、大きな割合が選択ゲノム領域の配列を含有する断片の個体群を産生する(202)。次にこれらの断片を、通常は、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、断片の元の分子構成を保存するようにさらに断片化またはコピーする(203)が、断片の元の分子構成に属させる任意の他の方法を使用することができる。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203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る、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。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る。試料の増幅及び試料の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、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号及び同第14/316,463号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、ならびにゲノム材料の配列決定及び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。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、試料の重複相補的な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、これらの相補的な断片は、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つの相補的な配列を末端または末端の近くに産生して、分子がさら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。

#### 【0046】

図2に例示されている方法に戻ると、区分特異的なバーコードをコピー断片に結合すると、次にバーコード付き断片をプールする(204)。次にプールされた断片を配列決定し(205)、目的の標的化領域が由来分子構成により特定され、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属させる(206)。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なバーコードを、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分子に属させることを可能にすることである。

#### 【0047】

上記のワークフローに加えて、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さらに豊富化、単離または分別される、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な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、などの処理をさらに受けてもよい。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さらに増幅してもよい。そのような標的化ブルダウン豊富化方法の例示的な方法は、2015年10月29日出願の米国特許出願第14/

927, 297号に記載されており、これは、全ての目的において、特に、記載、図面及び実施例を含む、標的化プルダウン豊富化方法及び配列決定方法に関する全ての教示について、その全体が参照として本明細書に組み込まれる。

#### 【0048】

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みのために、個別の分子構成を提供する。そのような個別の分子構成は、短い配列読み込みを、由来する個別の核酸に属させることを可能にする任意の方法または組成物によって、提供することができる。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。

10

20

#### 【0049】

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、個別の分子の近接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、もしくはコンティグに属させることは、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の両方を含んでもよい。

30

#### 【0050】

さらに、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長い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。

40

#### 【0051】

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

50

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10kb、少なくとも約20kb、少なくとも約50kb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100kb、少なくとも約150kb、少なくとも約200kb、多くの場合において、少なくとも約250kb、少なくとも約300kb、少なくとも約350kb、少なくとも約400kb、いくつかの場合において、少なくとも約500kb、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、200kbを超える、300kbを超える、400kbを超える、500kbを超える、1Mbを超える、またはさらには2Mb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

30

40

50

#### 【0052】

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、本明細書に記載されている方法のいずれかに従う更なる処理によって、試料核酸の分別を提供する。そのような分別は、核酸が、分別される他の核酸から相対的に単離されるように、更なる処理及び反応を受けることを受けることを可能にする任意の形態であり得る。分別は、他の全ての核酸からそれぞれ分別される単一核酸に関するものであり得る、または2つ以上の核酸の群にして、次に核酸の他の群から分別することであり得る。いくつかの例示的な実施形態において、そのような分別には、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)への試料核酸またはその断片の区画化、沈着化もしくは区分化が含まれ、それぞれの区分はその含有物を、他の区分の含有物から分別したまま維持する。独自の識別子または他の属させる手段(いくつかの例では、バーコード)を、分別された核酸に前もって、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、情報が誘導される試料核酸に、後に属させることを可能にすることができる。核酸が区画または区分に分別される特定の例示的な実施形態において、識別子は、特定の区画、特に、区分に元々沈着されていてよい近接試料核酸の比較的長い伸展の中に含まれ得る、または伸展に導入され得る。

#### 【0053】

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、または他の大きなゲノム部分を表す。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近接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、1kbより長く、5kbより長く、10kbより長く、15kbより長く、20kbより長く、30kbより長く、40kbより長く、50kbより長く、60kbより長く、70kbより長く、80kbより長く、90kbより長く、またはさらに100kb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。

#### 【0054】

試料核酸も、所定の区分が非常に低い確率でゲノム座の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、異なる独自の識別子を含むことができる。さらに、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。さらに、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、試料の制限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/核酸に属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）の比で操作され得るが、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、またはさらにそれ以上のオーダー）を装填することも可能である。

#### 【0055】

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。核酸にバーコードを付ける方法は、当該技術において知られており、本明細書に記載されている。いくつかの例において、方法は、Amini et al, 2014, Nature Genetics（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、オリゴヌクレオチドは、少なくとも第1及び第2の領域を含むことができる。第1の領域は、所定の区分内のオリゴヌクレオチドの間にあり、実質的に同じバーコード配列であってもよいバーコード領域であるが、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配列である。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）であり得る。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得る。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得る。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むことができ、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属し得るので、配列において特定された変異体を、全て、単一の由来断片及び単一の由来染色体に属させることができる。さらに、複数の共同配置変異体を複数の長い断片にわたって整列させることによって、染色体の寄与についてさらに特徴決定することができる。したがって、長い範囲のゲノム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プロタイプの特定にも、有用であり得る。コピー数変動も、この方法によって特定することがで

10

20

30

40

50

きる。

【0056】

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、ゲノムの不十分に特徴決定された、もしくは高度に多形性の領域からの1つの染色体、または1個もしくは数個の細胞からの材料、または血流中に循環する断片化腫瘍細胞DNA分子が寄与する遺伝子材料を特定及び特徴決定するように設計されていない。本明細書に記載されている方法は、少数の構成要素からの遺伝子材料を増加する選択的増幅方法を含み、この遺伝子材料の分子構成を保持する能力は、これらの構成要素の遺伝子特徴決定をさらに提供する。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、ハプ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。

10

【0057】

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のセットを（単一のマーカーとは対照的に）個別の試料核酸分子に属させる、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。さらに、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少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。

20

【0058】

前記に記載されたように、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. (GAIIx、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。

30

【0059】

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、関連する低い誤差率及び高い処理量を伴って行われることが、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、メイトペア (mate pair) 伸長を除いて、1000 bpより短い、500 bpより短い、300 bpより短い、200 bpより短い、150 bp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。

40

【0060】

50

## II. ワークフロー概観

本開示に記載されている方法及び系は、分別された核酸が1つ以上の他の核酸から相対的に単離されて更なる処理及び/または反応を受けることができるように、異なる群または異なる領域への核酸の分別を提供する。そのような分別は、特定の例示的な場合において、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含むことができ、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分別し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、中空などが含まれ得る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得る。様々な異なる器は、例えば、2013年8月13日出願の米国特許出願第13/966,150号に記載されている。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ション系は、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の例には、2015年4月9日出願の米国特許出願第14/682,952号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。

10

20

30

40

50

### 【0061】

エマルションの液滴を利用する方法において、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、1000 pL未満、900 pL未満、800 pL未満、700 pL未満、600 pL未満、500 pL未満、400 pL未満、300 pL未満、200 pL未満、100 pL未満、50 pL未満、20 pL未満、10 pL未満、またはさらに1 pL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%未満であり得る。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14/752,589号及び同第14/752,602号に表されており、これら全開示は、その全体が参照として本明細書に組み込まれる。

### 【0062】

試料が、本明細書に記載されている方法及び系によって対応する区分に導入されると、区分内の試料核酸は、一般に、標的化包括度が新規配列決定を可能にするための目的のゲノム領域がゲノムの他の領域と比較して高い割合で存在するように、選択的増幅を受ける（しかし、理解されるように、これらの他のゲノム領域も、新規包括度に対する目的のものではないので少ない程度であるが、増幅されてもよい）。特定の実施形態において、目

的のゲノム領域は、ゲノムの選択領域の少なくとも1x、2x、5x、10x、20x、30x、40x、または50xの包括度を提供するために増幅される。更なる実施形態において、区画内の全ての核酸が増幅されるが、選択ゲノム領域は、少なくとも1~5、2~10、3~15、4~20、5~25、6~30、7~35、8~40、9~45、または10~50倍多いアンプリコンが、ゲノムの他の部分と比べてこれらの選択ゲノム領域から産生されるように標的化されて、増幅される。

#### 【0063】

ゲノムの選択領域の選択的増幅と同時に、またはその後、区分内の核酸（または、その断片）には、これらの核酸が特徴決定されたときに、これらに対応する起源から誘導されたように属させ得るように、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と共に同時区分化される。いくつかの例示的な実施形態において、そのような独自の識別子はバーコード配列である。明確さのため、本明細書における考察の多くは、バーコード配列を含む識別子を対象としているが、理解されるように、配列読み取りのために分子構成を保持するのに使用できる任意の独自の識別子を、本明細書に記載されている方法に従って使用することができる。いくつかの好ましい態様において、独自の識別子は、核酸試料に結合し得る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得るように、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に関連するが、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。

#### 【0064】

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、完全に近接的、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分別されている、2つ以上の分別部分配列に分別されていてもよい。典型的には、分別部分配列は、典型的には約4~約16個のヌクレオチドの長さであり得る。

#### 【0065】

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、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号及び同第14/316,463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。

#### 【0066】

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た上記に記載されたオリゴヌクレオチドを含み得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得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分子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 100,000,000 個のオリゴヌクレオチド分子、いくつかの場合では少なくとも 10 億個のオリゴヌクレオチド分子であり得る。

#### 【0067】

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得る。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得る光解離性連結の切断を介した、光刺激であり得る。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らすことができる、化学刺激を使用してもよい。

10

#### 【0068】

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得る。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。

20

#### 【0069】

図 3 は、特にコピー数変動またはハプ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し、続いて配列決定する 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から得ることができ (300)、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 - m e r または他のプライマーなど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、還元剤など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 - m e r を有するオリゴヌクレオチドを、液滴内でビーズから遊離させることができる (335)。次にランダム N - m e r は、試料核酸の異なる領域を刺激して、増幅後に試料の増幅コピーをもたらす、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 - m e r 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理すること、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

30

40

50

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さらに使用して、コピーの配列を由来分子構成に属させることもできる。

#### 【0070】

理解されるように、バーコード配列をタグ付けする前に、またはタグ付けすると同時に、(340)試料を本明細書に記載されているいずれかの方法によって増幅して、ゲノムの選択領域の標的化包括度を提供することができる。この標的化包括度は、一般に、ゲノムの他の領域のアンプリコンと比較して、ゲノムの選択領域を含有する区分中の核酸(または、その部分)の配列を表すアンプリコンの大きな個体群をもたらす。その結果、ゲノムの選択領域の区分内の、バーコード配列を含有する増幅コピーの数が、ゲノムの他の領域よりも大きくなる(340)。

10

#### 【0071】

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、異なる増幅プロトコルを使用して、断片へのバーコード配列の結合に使用されるプロトコルよりも、ゲノムの選択領域の部分を含む断片の増幅を優先させる。1つの非限定例において、標的化PCRプライマーを使用する選択的増幅は、標準的なPCR増幅熱サイクル条件下で実施され、一方、バーコード結合のための増幅は、ランダムN-merの刺激及び伸長を可能にするため、温度の急激な低下、続いてゆっくりとした増加温度の配向を伴って実施される。

#### 【0072】

上記に示されたように、単一占有が最も望ましい状態であり得るが、複数占有区分または非占有区分が多くの場合に存在し得ることが理解される。試料及びバーコードオリゴヌクレオチドを含むビ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。本明細書の他の箇所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、水性流体内に浮遊されていてもよい。

20

#### 【0073】

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。非水性区分化流体が、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。

30

#### 【0074】

理解されるように、例えば、化学刺激、核酸伸長、転写、及び/またはポリメラーゼ、逆転写酵素、ヌクレオシド三リン酸もしくは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。プライマー配列には、ゲノムの選択領域またはその組み合わせの増幅に向けられた、ランダムプライマー配列または標的化PCRプライマーが含まれ得る。

40

#### 【0075】

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。これらのバーコードオリゴヌクレオチドを試料の増幅及びバーコード付けに使用する特に洗練されたプロセスは、米国特許

50

出願第 14 / 3 1 6 , 3 8 3 号、同第 14 / 3 1 6 , 3 9 8 号、同第 14 / 3 1 6 , 4 1 6 号、同第 14 / 3 1 6 , 4 3 1 号、同第 14 / 3 1 6 , 4 4 7 号及び同第 14 / 3 1 6 , 4 6 3 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。プライマー配列は、ランダムであり得る、または構造化され得る。ランダムプライマー配列は、一般に、試料の多数の異なる領域を無作為に刺激することが意図される。構造化プライマー配列には、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 G C N - m e r ）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、が含まれ得る。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のいずれかを、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。

#### 【0076】

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、ヌクレオシド三リン酸、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、次に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生して、分子がさら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。この1つの例の概略図が、図5に示されている。

#### 【0077】

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、エマルションの液滴502の中に、試料核酸504と共に同時区分化されている。本明細書の他の箇所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、1つ以上の機能性配列、例えば、配列510、514及び516に加えて、バーコード配列512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得る配列510、例えば、Illumina HiSeqまたはMiSeq系のフローセルにおいて結合のために使用される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 - m e r 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性(directed)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得る。プライマー配列516は、ランダ

10

20

30

40

50

ム N - m e r プライマーでは変わり得るか、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に共通のものであり得る。

【 0 0 7 8 】

プライマー配列 5 1 6 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 5 0 6 及び試料核酸 5 0 4 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 5 0 8 及び 5 0 8 a 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 - m e r プライマーでは、試料核酸 5 0 4 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 5 1 8 及び 5 2 0 である。試料核酸の部分に相補的な配列部分、例えば配列 5 2 2 及び 5 2 4 を含んでいるが、これらの構造体は、本明細書において、結合バーコード配列を有する試料核酸 5 0 4 の断片を含むものと一般に呼ばれる。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、本明細書において、そのテンプレート配列の「断片」と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、試料核酸配列の断片またはテンプレートは、基礎配列またはその補体の複製部分を示す。

10

【 0 0 7 9 】

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、プロセスによってさら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 5 0 8 b も、ビーズ 5 0 6 から遊離されて、断片 5 1 8 及び 5 2 0 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 5 0 8 b におけるランダム N - m e r プライマー 5 1 6 b (多くの場合では、所定の区分内の他のランダム N - m e r、例えばプライマー配列 5 1 6 と異なる)の存在に基づいて、オリゴヌクレオチドは断片 5 1 8 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 5 2 8 を含む断片 5 1 8 の少なくとも一部まで、補体 5 2 6 作り出す。オリゴヌクレオチド 5 0 8 b の伸長は、断片 5 1 8 のオリゴヌクレオチド部分 5 0 8 を介して複製されるまで続く。本明細書の他の箇所に示されており、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 5 1 8 に含まれるオリゴヌクレオチド 5 0 8 の配列 5 1 6 及び 5 1 4 を介する複製の後、所望の地点でポリメラーゼによる複製に終止を促すように形成され得る。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することができ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、ウラシル含有ヌクレオチドを配列領域 5 1 2 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得る。その結果、一方の末端に完全長オリゴヌクレオチド 5 0 8 b を含み、バーコード配列 5 1 2、結合配列 5 1 0、R 1 プライマー領域 5 1 4 及びランダム N - m e r 配列 5 1 6 b を含む、断片 5 2 6 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 5 0 8 のランダム N - m e r の補体 5 1 6 '、ならびに R 1 配列の全てまたは一部の補体が配列 5 1 4 ' と示されて含まれる。次に R 1 配列 5 1 4 及びその補体 5 1 4 ' を一緒にハイブリッド形成させて、部分ヘアピン構造 5 2 8 を形成することができる。理解されるように、ランダム N - m e r 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することが予測されず、例えば、ランダム N - m e r 5 1 6 の補体である配列 5 1 6 ' は、ランダム N - m e r 配列 5 1 6 b の補体であることが予測されない。このことは、他の用途、例えば、N - m e r が所定の区分内のオリゴヌクレオチドに共通している標的化プライマーには当てはまらない。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 5 2 6 の後の処理にとって有用な構造も提供する

20

30

40

50

。

## 【 0 0 8 0 】

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属させることができる。これは、図 6 A 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 6 0 0 (例えば、個別の染色体、核酸の鎖など)に由来する核酸 6 0 4 及び異なる染色体 6 0 2 または核酸の鎖から誘導される核酸 6 0 6 は、それぞれ、上記に記載されたように、それ自体の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。

10

## 【 0 0 8 1 】

それぞれの区分内において、それぞれの核酸 6 0 4 及び 6 0 6 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 6 0 8 及び 6 1 0 を別個にもたらす。またこの処理は、特定の第 1 の断片から誘導される第 2 の断片のそれぞれにおいて同じ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 6 0 8 のバーコード配列は、「 1 」と示されており、断片セット 6 1 0 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、本明細書の他の箇所に詳細に記載されている。

20

## 【 0 0 8 2 】

次に、例えば、断片セット 6 0 8 及び 6 1 0 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、プールした断片 6 1 2 の配列読み込みを、集合読み込み 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属させることができる。次にそれぞれの断片セットの属する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、これらに対応する元の染色体または供給源核酸分子 (6 0 0 及び 6 0 2 ) にさらに属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、2015 年 6 月 26 日出願の米国特許出願第 14 / 752, 773 号に記載されており、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。

30

## 【 0 0 8 3 】

いくつかの実施形態において、図 6 B に例示されているように、プライマーセット 6 1 3 が、断片セット 6 0 8 または 6 1 0 を含有する区分に含まれる。プライマーセット 6 1 3 は、更なる実施形態において、バーコード配列 (6 0 8 にはバーコード「 1」、6 1 0 にはバーコード「 2」) を提供する前に、提供するのと同時に、または提供した後、断片セット 6 0 8 及び 6 1 0 が増幅されて、ゲノムの選択領域がゲノムの他の領域にわたる追加の伸長を網羅するように、ゲノムの選択領域に向けられる。図 6 B に描写されている例示的な実施形態において、断片セット 6 0 8 は、プライマーセット 6 1 3 が向けられるゲノムの選択領域の配列を含有するが、断片セット 6 1 0 は、ゲノムの選択領域の配列を含有しない。このように、セット 6 1 0 よりも、セット 6 0 8 の断片の包括度が増加している (例えば、コピー数が多い)。したがって、プールされた断片 6 1 2 は、標的化様式で増幅された断片を含有するバーコード付き断片を含有し、断片セット 6 1 0 (「 2 」バーコード付き断片) よりも、断片セット 6 0 8 (「 1 」バーコード付き断片) からの大きな割合の配列読み込みを可能にする。加えて、バーコードに起因して、セット 6 0 8 からの大きな割合の配列読み込みは、プールされたセット 6 1 2 の断片を、プールされたセット

40

50

612における残りの断片のように、対応する元の供給源核酸分子600及び602に属させることができる(図6Aに示されている)。

#### 【0084】

#### III. 核酸配列決定への方法及び系の適用

本明細書に記載されている方法、組成物及び系は、核酸配列決定技術における使用に特に適応可能である。そのような配列決定技術には、短い読み込み及び長い読み込み配列決定技術を含む、当該技術に既知の任意の技術が含まれ得る。特定の態様において、本明細書に記載されている方法、組成物及び系は、短い読み込みで高い精度の配列決定技術に使用される。

#### 【0085】

本明細書に記載されている方法、組成物及び系は、不十分に特徴決定されている、高度に多形性がある、及び/または基準配列から逸脱している、ゲノム領域の遺伝子特徴決定を可能にする。特に、本明細書に記載されている方法、組成物及び系は、追加の多重配列情報がゲノムのこれらの選択部分から取得され得るように、ゲノムの選択部分の増加した多重包括度を提供する。特定の場合において、この追加の配列情報(例えば、ゲノムの標的化領域の増加した包括度)は、ゲノムのこれらの選択部分の新規配列決定を可能にするのに十分な情報を提供する。この新規配列決定は、不十分に特徴決定されている、高度に多形性がある、及び/または基準配列から逸脱している、ゲノム領域において特定の使用される。理解されるように、有意な率(例えば、Altmeose 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、マルチメガベース異質染色性間隙と注釈する。この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、構造的な特徴が含まれる。正確な特徴決定に抵抗性がある追加の例示的な領域には、限定されることなく、密接な相同性がある偽遺伝子(例えば、SMN1/2 Cyp2d6)を有する区域、トランスポゾン(SINE、LINEなど)が限定されることなく含まれる、ゲノム全体にわたって実質的に反復している配列を有する区域、同様に、基準配列が不十分な先導としてしか機能しない、途方もない変動を有する区域(ヒト白血球抗原(HLA)複合体の遺伝子をコードする領域など)が含まれる。本明細書に記載されている方法、組成物及び系は、目的の領域の選的増幅を、分子構成を維持する能力と組み合わせ、それによって、一般的に不十分に特徴決定されているゲノム領域の新規配列決定を可能にする。

#### 【0086】

特定の場合において、本明細書に記載されている方法は、ゲノムの選択領域が配列決定される前に選択的に増幅されるステップを含む。この増幅は、一般に当該技術に既知の方法(PCR増幅が限定されることなく含まれる)を使用して実施され、ゲノムの選択領域の少なくとも1x、2x、3x、4x、5x、6x、7x、8x、9x、10x、11x、12x、13x、14x、15x、16x、17x、18x、19x、または20xの包括度を提供し、それによって、これらの選択領域の新規配列決定を可能にする量の核酸を提供する。更なる実施形態において、増幅は、ゲノムの選択領域の少なくとも1x~30x、2x~25x、3x~20x、4x~15x、または5x~10xの包括度を提供する。

#### 【0087】

増幅は、一般に、ゲノムの選択領域の中の、または近くの配列に相補的なプライマーの伸長にわたって実施される。いくつかの場合において、目的の領域にわたってタイル状になるように設計されているプライマーライブラリーが使用され、換言すると、プライマーライブラリーは、ゲノムの選択領域に沿って特定の距離で領域を増幅するように設計される。いくつかの場合において、選択的増幅は、ゲノムの選択領域に沿って、10、15、

20、25、50、100、200、250、500、750、1000、または10000塩基毎に相補性があるプライマーを利用する。なお更なる例において、タイル状のプライマー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得る、または選択領域の特定の部分もしくは割合が異なるプライマー対により増幅されるように、知的に設計され得る。更なる実施形態において、プライマー対は、それぞれの対がゲノムの選択部分における任意の近接領域の約1～5%、2～10%、3～15%、4～20%、5～25%、6～30%、7～35%、8～40%、9～45%、または10～50%を増幅するように設計される。

#### 【0088】

特定の実施形態において、上記に記載されたうちのいずれかによると、増幅は、少なくとも3メガベース対長さ(Mb)のゲノム領域にわたって生じる。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って選択的に増幅されるゲノムの選択領域は、少なくとも3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5、または10Mbである。なお更なる実施形態において、ゲノムの選択領域は、約2～20、3～18、4～16、5～14、6～12、または7～10Mbの長さである。上記において考察されたように、増幅は、これらの領域の末端における、または末端に近い配列に相補的な単一プライマー対を使用して、これらの領域にわたって生じ得る。他の実施形態において、増幅は、規則的なセグメント、ランダムセグメントまたは増幅される領域に沿った異なるセグメント距離のいくつかの組み合わせが、上記に記載された包括度の程度で増幅されるように、領域の長さにわたってタイル状であるプライマー対ライブラリーによって実施される。

#### 【0089】

いくつかの実施形態において、ゲノムの選択領域の選択的増幅に使用されるプライマーは、プライマー自体が増幅されないようにウラシルを含有する。

#### 【0090】

一般に、本明細書に記載されている方法及び系は、ゲノムの選択領域の配列の決定を提供することによって標的化ゲノム配列決定を達成し、この配列決定情報は、一般に、極めて低い配列決定誤差率及び高い処理量の短い読み込み配列決定技術という利点を有する方法を使用して得られる。前記に記載されたように、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. (GAIIx、NextSeq、MiSeq、HiSeq、X10)、Ion Torrent division of Thermo-Fisher (Ion PGM)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。

#### 【0091】

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すること及び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。

#### 【0092】

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 N14/316,383、14/316,398、14/316,416、14/316,431、14/316,447及び14/316,463においてもさらに詳細に記載

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、ならびにゲノム材料の配列決定及び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。

#### 【0093】

使用される配列決定プラットフォームにかかわらず、一般に、本明細書に記載されている方法のいずれかによると、核酸の配列決定は、配列読み込みの分子構成または配列読み込みの部分の保存する方法で典型的に実施される。このことは、多重配列読み込み、または配列読み込みの複数の部分が、核酸の単一由来分子に属し得ることを意味する。「属する」とは、配列読み込みが核酸の特定の由来分子の塩基の直線状配列内にあると特定され得ることを意味し、換言すると、図7を参照して、断片703、704、705及び706が由来核酸分子701及び702から生成される場合、全ての断片が配列決定反応のために一緒にプールされている場合であっても、配列決定は、断片703、704、705及び706の配列読み込みが分子構成を保持するように実施され、断片703及び704は由来分子701から誘導され、一方、705及び706は由来分子702から誘導されることが容易に確認されることを意味する。加えて、配列決定は、一般に、由来分のみならず、直線状分子に沿ったそれぞれの断片の相対位置も確認されるように実施され、例えば、断片703が由来核酸701の直線状配列に沿って断片704の「上流」にあると決定することができる。一般に、分子構成は、1つ以上の断片を他の断片と区別する任意の識別子または任意の他の方法の使用によって保持される。一般に、そのような識別子は、群または個別の実体に分別されている断片に使用される。いくつかの例において、そのような分別は分離区分への分別であるが、理解されるように、分子を分別する任意の他の方法を使用することができる。なお更なる例において、使用される識別子はバーコードであり、直線位置は、バーコード付けと、重複断片の配列読み込みのアルゴリズムアセンブリ(algorithmic assembly)の両方によって決定される。明確さのために、本明細書における多くの考察は、区分への分別及び/またはバーコード付けに関するものであるが、由来核酸分子を分別する任意の方法及び断片を特定する、そうでなければ属させる任意の方法は、本明細書に記載されている方法及び系に使用されることが理解される。

#### 【0094】

理解されるように、核酸の単一由来分子は、様々な長さの任意のものであり得るが、好ましい態様において比較的長い分子であり、長い範囲の分子構成の保存を可能にする。特に、単一由来分子は、好ましくは、典型的な短い読み込み配列長さより実質的に長いもの、例えば、200塩基を超えるものであり、多くの場合、少なくとも1000塩基以上、5000塩基以上、10,000塩基以上、20,000塩基以上、30,000塩基以上、40,000塩基以上、50,000塩基以上、60,000塩基以上、70,000塩基以上、80,000塩基以上、90,000塩基以上、または100,000塩基以上、いくつかの場合では、1メガベース以上である。

#### 【0095】

特定の状況において、本明細書に記載されている配列決定方法は、選択領域の深い包括度と、ゲノムの長い範囲にわたる低レベルの連結読み込みとの組み合わせを含む。理解されるように、この新規と再配列決定との組み合わせは、ゲノム全体及び/またはゲノム大きな部分を配列決定する効率的な方法を提供する。本明細書に記載されている選択的増幅方法による、不十分に特徴決定されている及び/または高度に多形性のある領域の標的化包括度は、新規配列会合に必要な核酸材料の量を提供し、一方、ゲノムの他の領域にわたる連結ゲノム配列決定は、残りのゲノムの高い処理量での配列決定を維持する。本明細書に記載されている方法及び組成物は、同じ配列決定プラットフォームを両方の種類の包括度に使用できるので、この新規と連結読み込み配列決定との組み合わせを可能にすることに独自に適応可能である。本明細書に記載されている方法に従って配列決定される核酸及び/または核酸断片の個体群は、新規配列決定用のゲノム領域及び再配列決定用のゲノム領域の両方からの配列を含有し、新規配列決定用の目的の領域を網羅する核酸の割合は、



本明細書においてさらに詳細に記載される標的化増幅方法に起因して、ゲノムの他の領域を網羅する核酸よりも多い。

【0096】

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、分子構成保持しながら、核酸、特にゲノムの選択領域の核酸を特徴決定することができる。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（例えば、同じ核酸分子に由来している）が、いくつかの場合では、これらの領域は、互いに相対的に分別されていることがあり、例えば、1 kbを超える間隔、5 kbを超える間隔、10 kbを超える間隔、20 kbを超える間隔、30 kbを超える間隔、40 kbを超える間隔、50 kbを超える間隔、いくつかの場合では100 kbまでの間隔であり得る。目的の領域は、一般にゲノムの分離及び分別部分であり、いくつかの場合では、そのような領域は不十分に特徴決定された領域である。目的の領域は、また、個別の遺伝子、遺伝子群、エクソンを示すこともある。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にそれぞれ分別されている。本明細書の他の箇所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しているバーコード配列、例えば例示されているバーコード配列「1」または「2」を、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を分析して、ゲノムの科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。

【0097】

上記によると、バーコード付けステップに加えて、核酸102または104が目的の選択ゲノム領域を含有する場合、これらの領域のアンプリコンが区分114及び116それぞれにおける断片の大きな割合を形成するように、1つ以上の選択的増幅のステップがあってもよい。この増幅ステップは、一般に、本明細書に記載されている方法によるバーコードの結合の前、または結合と同時に実施されるが、いくつかの実施形態において、増幅ステップは、バーコードの結合に続いて生じることもある。

【0098】

ライブラリー118内のプールされた断片が、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（時には、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により、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る。選択的増幅が、ゲノムの選択領域の部分を含む核酸断片の量を増加するために利用される状況では、分子構成を特定する能力は、ゲノムの選択領域を、多くの場合にこれらの領域の新規会合を可能にする深さで配列決定する方法も提供する。

【0099】

一般に、本発明の方法は、本明細書においてさら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2に例示されているステップを含む。理解されるように、図2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、変更及び修正され得る例示的な実施形態である。

【0100】

10

20

30

40

50

図 2 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 (201) を含む。一般に、目的のゲノム領域の核酸を含有する区分は、それぞれ標的化豊富化を受けて、大きな割合が選択ゲノム領域の配列を含有する断片の個体群を産生する (202)。次にこれらの断片を、通常は、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、断片の元の分子構成を保存するようにさらに断片化またはコピーする (203)。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 203 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。試料の増幅及び試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、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 号及び同第 14 / 316, 463 号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、ならびにゲノム材料の配列決定及び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。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。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、分子がさら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。

#### 【0101】

図 2 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、次にバーコード付き断片をプールする (204)。次にプールされた断片を配列決定し (205)、目的の標的化領域が由来分子構成により特定され、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属させる (206)。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属させることができることである。

#### 【0102】

上記のワークフローに加えて、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含む方法を使用して、更なる分析、特に配列決定のためにさらに豊富化、単離または分別される、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、などの処理をさらに受けてもよい。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さらに増幅してもよい。そのような標的化プルダウン豊富化方法の例示的な方法は、2015年10月29日出願のUS 2015/0192777, 2015年10月29日出願のUS 2015/0192778に記載されており、これは、全ての目的において、特に、記載、図面及び実施例を含む、標的化プルダウン豊富化方法及び配列決定方法に関する全ての教示について、その全体が参照として本明細書に組み込まれる。

10

#### 【0103】

いくつかの態様において、ゲノムの選択領域の包括度についての方法は、これらの選択領域の核酸分子及び／またはその断片を含有する分離区分が更なる処理のためにそれ自体選別される方法を含む。理解されるように、この分離区分の選別は、本明細書に記載されている目的のゲノム領域の選択的増幅及び／または標的化プルダウンの他の方法との任意の組み合わせ、特に、上記に記載されたワークフローステップの任意の組み合わせによって実施され得る。

#### 【0104】

一般に、そのような分離区分の選別方法は、ゲノムの1つ以上の選択部分の少なくとも一部を含有する区分が、ゲノムのこれらの部分の任意の配列を含有しない区分から分別されるステップを含む。これらの方法は、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の配列が豊富化された個体群を、ゲノムのこれらの部分の配列を含有する分離区分内に提供するステップを含む。そのような豊富化は、ゲノムの1つ以上の選択部分の少なくとも一部を含む分離区分内の断片に定方向PCR増幅を使用して、個体群を産生することによって、一般に達成される。したがって、この定方向PCR増幅は、ゲノムの1つ以上の選択部分の少なくとも一部を含むアンプリコンを産生する。特定の実施形態において、これらのアンプリコンは、検出可能標識に結合され、いくつかの非限定実施形態では、蛍光分子が含まれ得る。一般に、そのような結合は、ゲノムの1つ以上の選択部分を含有する断片から生成されたアンプリコンのみが、検出可能標識に結合されるように生じる。いくつかの実施形態において、検出可能標識の結合は、ゲノムの1つ以上の選択部分の選択的増幅の際に生じる。そのような検出可能標識には、更なる実施形態において、蛍光標識、電気化学標識、磁気ビーズ及びナノ粒子が限定されることなく含まれ得る。この検出可能標識の結合は、当該技術に既知の方法を使用して達成することができる。なお更なる実施形態において、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含有する分離区分は、これらの区分内のアンプリコンに結合した検出可能標識から放出されるシグナルに基づいて選別される。

20

30

#### 【0105】

更なる実施形態において、ゲノムの選択部分を含有する分離区分を、そのような配列を含有しないものから選別するステップは、(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、それぞれの分離区分が第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)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の配列が豊富化されている少なくともいくつかの分離区分内に、個体群を提供するステップ、(d)共通バーコード配列を、それぞれの断片が、それが含有されていた分離区分に属するように、それぞれの分離区分内の断片に結合するステップ、(e)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含有する分離区分を、ゲノムの1つ以上の選択部分を含む断片を含有しない分離区分から分別するステップ、(f)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るステップを含む。

40

50

## 【0106】

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、断片から配列情報を取得する前に、分離区分は合わされ、断片は一緒にプールされる。更なる実施形態において、断片から配列情報を取得するステップは、確定することが同じ第1の個別の核酸分子から誘導される断片を確定することをさらに含むように、断片の配列の分子構成を維持するように実施される。なお更なる実施形態において、この配列情報の取得は、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配列決定反応を含む。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。

## 【0107】

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、分離区分は、エマルションの液滴を含む。更なる実施形態において、分離区分内のバーコード付き断片は、ゲノムの1つ以上の選択部分の約 $1 \times \sim 10 \times$ の包括度を表す。なお更なる実施形態において、分離区分内のバーコード付き断片は、ゲノムの1つ以上の選択部分の約 $2 \times \sim 5 \times$ の包括度を表す。なお更なる実施形態において、分離区分内のアンブリコンのバーコード付き断片は、ゲノムの1つ以上の選択部分の少なくとも $1 \times$ の包括度を表す。なお更なる実施形態において、分離区分内のバーコード付き断片は、ゲノムの1つ以上の選択部分の少なくとも $2 \times$ または $5 \times$ の包括度を表す。

## 【0108】

ゲノムの選択領域から配列情報を取得する能力を提供することに加えて、本明細書に記載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、米国特許出願第14/752,589号及び同第14/752,602号に記載されており、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。

## 【0109】

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みのために、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1 kbより長い、5 kbより長い、10 kbより長い、15 kbより長い、20 kbより長い、30 kbより長い、40 kbより長い、50 kbより長い、60 kbより長い、70 kbより長い、80 kbより長い、90 kbより長い、もしくはさらに100 kb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。

## 【0110】

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、個別の分子の近接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、1 kbより長い、5 kbより長い、10 kbより長い、15 kbより長い、20 kbより長い、30 kbより長い、40 kbより長い、50 kbより長い

、60 kbより長い、70 kbより長い、80 kbより長い、90 kbより長い、またはさらには100 kbより長い。配列構成と同様に、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、もしくはコンティグに属させることは、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の両方を含んでもよい。

#### 【0111】

さらに、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長い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。

10

#### 【0112】

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数値より長い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、またはさらには2 Mbを超える最大位相ブロック長さを得ることもできる。

20

30

#### 【0113】

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、それぞれの区分は、それ自体の内容物を他の区分の内容物から分別し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、区分に元々沈着されていてもよい近接試料核酸の比較的長い伸展に、後に属させることを可能にすることができる。

40

#### 【0114】

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、または他の大きなゲノム部分を表す。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近接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

50

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。

#### 【0115】

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、異なる独自の識別子を含むことができる。さらに、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。さらに、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/核酸に属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、600,000、700,000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）の比で操作され得るが、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、またはさらにそれ以上のオーダー）を装填することも可能である。

#### 【0116】

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。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。

#### 【0117】

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の間にあり、実質的に同じバーコード配列であってもよいバーコード領域であるが、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 - m e r ( ランダム N - m e r または特定の配列を標的にするように設計された N - m e r ) であり得る。N - m e r 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得る。本明細書において考察されているように、N - m e r を、不十分に特徴決定される傾向がある、または高度の多形成がある、もしくは基準配列から逸脱している、ゲノムの選択領域のために設計することもできる。いくつかの場合において、N - m e r 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得る。区分内では、増幅反応は、第 2 の N - m e r 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、より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属し得るので、配列において特定された変異体を、全て、単一の由来断片及び単一の由来染色体に属させることができる。さらに、複数の共配置変異体を複数の長い断片にわたって整列させることによって、染色体の寄与についてさらに特徴決定することができる。したがって、長い範囲のゲノム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プロタイプの特定にも、有用であり得る。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

20

30

40

50

#### 【 0 1 1 8 】

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、ゲノムの不十分に特徴決定された、もしくは高度に多形性の領域から、または 1 個もしくは数個の細胞の材料から、または血流中に循環する断片化腫瘍細胞 DNA 分子からの 1 つの染色体が寄与する遺伝子材料を特定及び特徴決定するように設計されていない。本明細書に記載されている方法は、少数の構成要素からの遺伝子材料を増加する選択的増幅を含み、この遺伝子材料の分子構成を保持する能力は、これらの構成要素の遺伝子特徴決定をさらに提供する。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、ハプ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。

#### 【 0 1 1 9 】

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のセットを ( 単一のマーカーとは対照的に ) 個別の試料核酸分子に属させる、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。さらに、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少数の構成成分と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、マイクロ流体力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。

#### 【0120】

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みのために、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によって、低い解像度構成、ならびに、例えば個別の分子の近接決定配列を有する長い個別の分子の大きな部分を、例えば長い範囲で配列決定することによって、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、もしくはコンティグに属させることは、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の両方を含んでもよい。

#### 【0121】

##### 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、scRNA、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、特定の処理の必要性に応じて変わり得る。

#### 【0122】

核酸を含む任意の物質が、試料の供給源であり得る。物質は、流体、例えば、生物学的流体であり得る。流体物質には、血液、臍帯血、唾液、尿、汗、血清、精液、膿液、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得るが、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得る。物質は、正常な健康組織、疾患組織、または健康及び疾患組織の混合物から構成され得る。いくつかの場合において、物質は腫瘍から構成され得る。腫瘍は、良性（非癌）または悪性（癌）であり得る。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、膵癌、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、肝細胞腫、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得る。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得る。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、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得る。いくつかの場合において、物質は、例えば単一細胞の内容物または複数の細胞の内容物などの細胞の内容物を含んでもよい。個別の細胞を分析する方法及び系は、例えば、2015年6月26日出願の米国特許出願第14/752,641号に提示されており、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。

10

20

#### 【0123】

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対象であり得る。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、または魚が含まれ得るが、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得る。いくつかの場合において、対象は妊婦であり得る。いくつかの場合において、対象は正常で健康な妊婦であり得る。いくつかの場合において、対象は、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、妊婦であり得る。

30

#### 【0124】

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、循環系から (例えば、シリンジもしくは他の器具により静脈内もしくは動脈内から) 入手することによって、分泌された生物学的試料 (例えば、唾液、痰、尿、糞便など) を収集することによって、生物学的試料 (例えば、術中試料、術後試料など) の外科的な (例えば、生検による) 取得によって、スワブ拭き取り (swabbing) (例えば、頬側スワブ、口腔咽頭スワブ) によって、またはピペット操作を介して、対象から得てもよい。

#### 【0125】

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これらの等価物が網羅されることが、意図される。

40

#### 【実施例】

#### 【0126】

実施例：TP53遺伝子の標的化包括度

TP53遺伝子を標的にする増幅反応を実施した。腫瘍タンパク質p53は、p53、

50

細胞腫瘍抗原 p 5 3 ( U n i P r o t 名称 )、リンタンパク質 p 5 3、腫瘍抑制因子 p 5 3、抗原 N Y - C O - 1 3、または形質転換関連タンパク質 5 3 ( T R P 5 3 ) としても知られており、ヒトにおいて T P 5 3 遺伝子によりコードされるタンパク質である。p 5 3 タンパク質は、多細胞生物体において重要であり、細胞周期を調節し、したがって腫瘍抑制因子として機能して、癌を予防する。このように、p 5 3 は、ゲノム突然変異を防止することによって安定性を保存する役割のため、「ゲノムの守護者 ( t h e g u a r d i a n o f t h e g e n o m e ) 」と記載されてきた。したがって、T P 5 3 は、腫瘍抑制遺伝子と分類される。

#### 【 0 1 2 7 】

T P 5 3 遺伝子を含むゲノムの領域 ( 長さが約 1 9 1 4 9 b p ) の標的化増幅を、遺伝子全体にまたがるように合計 9 6 プライマーを使用して、多重反応によって実施した。プライマーは、約 4 0 0 b p の間隔で、このゲノム領域にわたってタイル状になるように設計した。増幅反応は、1 4 サイクルのアニーリングステップに温度勾配及び約 3 n g の D N A の投入量を用いて実施した。この例に使用した熱サイクルプロトコールは、以下であった。

【表 1】

初期変性	98°C	30 秒
18 サイクル	98°C	10 秒
	30-55°C	15 秒
	72°C	15 秒
	72°C	2 分
最終伸長	72°C	2 分
保持	4C	

#### 【 0 1 2 8 】

この種類の反応の例示的なワークフローを図 8 に描写する。理解されるように、これは本明細書に記載されている本発明による方法の例示的な実施形態であり、既知の方法を使用して変更または拡張することができる。図 8 に示されているように、ゲノムの選択領域 ( この場合では、T P 5 3 遺伝子 ) は、8 0 2 及び 8 0 3 と描写されているように、標的的特異的プライマーを使用して増幅される。加えて、バーコード 8 0 1 を有するプライマーもアンプリコンに組み込まれ、このことは、本明細書に記載されているように、特定の実施形態では、後に続く配列読み込み ( 8 0 8 ) に分子構成を提供することができる。

#### 【 0 1 2 9 】

プライマー 8 0 2 及び 8 0 3 は、この実験では、「尾部 ( t a i l s ) 」R 1 及び R 2 を有し、このことは、得られるアンプリコンを、I l l u m i n a プラットフォームなどの特定のプラットフォームによる配列決定に適応可能にした。S I プライマー ( 8 0 6 ) を用いた増幅は、I l l u m i n a プラットフォームにも使用される試料指標をさらに提供した。理解されるように、他の配列決定プラットフォームに有用な配列を、R 1 及び R 2、ならびに S 1 プライマーの代わりに使用することができる。

#### 【 0 1 3 0 】

図 9 は、増幅反応が、テンプレートのない対照 ( N T C ) が生成物を示さなかったもので、特異的であることを示している。図 1 0 は、上記記載のプロトコールの結果として見られる折り畳みの豊富化を、一連の温度にわたって提示している。

#### 【 0 1 3 1 】

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。さらに、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

10

20

30

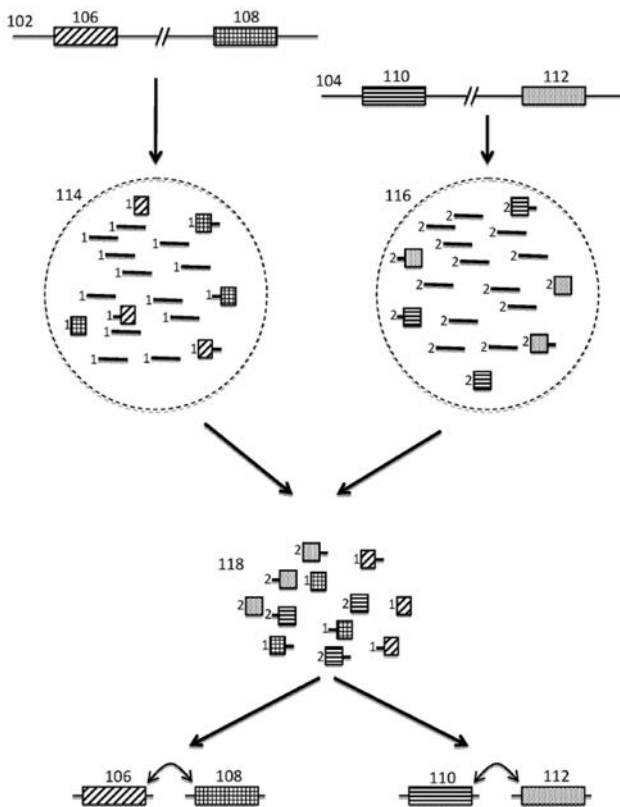
40

50

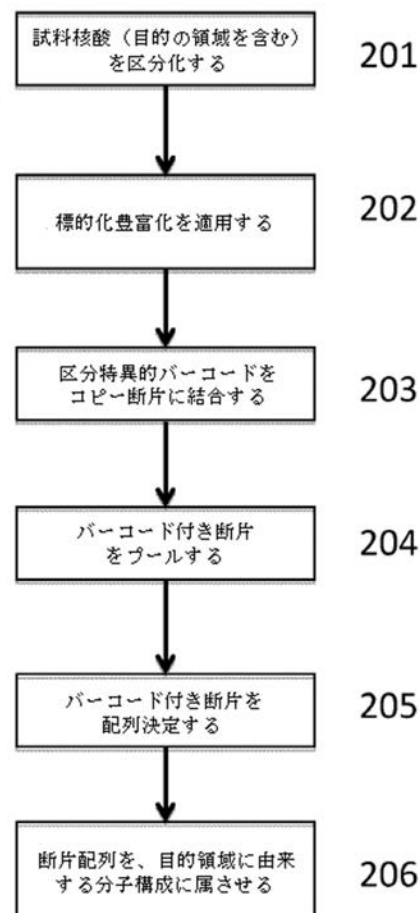
り、任意の順番で実施できることが理解されるべきである。上記の記載に含有され、添付の図面に示されている全ての事項は、特定の態様の例示のためのみであり、示されている実施形態を制限するものではないと解釈されることが意図される。文脈から明白ではない限り、または明確に記述されていない限り、本明細書に提供されている任意の濃度値は、一般に、混合物の特定の成分の添加の際、または添加後に生じる任意の変換を無視している混合物の値または百分率として提示されている。既に明白に本明細書に組み込まれていない程度に、本開示において参照されている全ての公開参考文献及び特許文献は、全ての目的において、それらの全体が参照として本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。

10

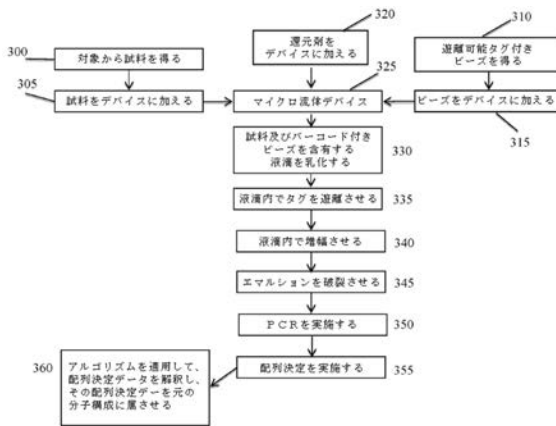
【図 1】



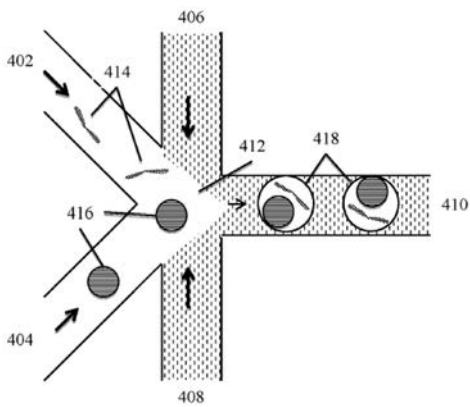
【図 2】



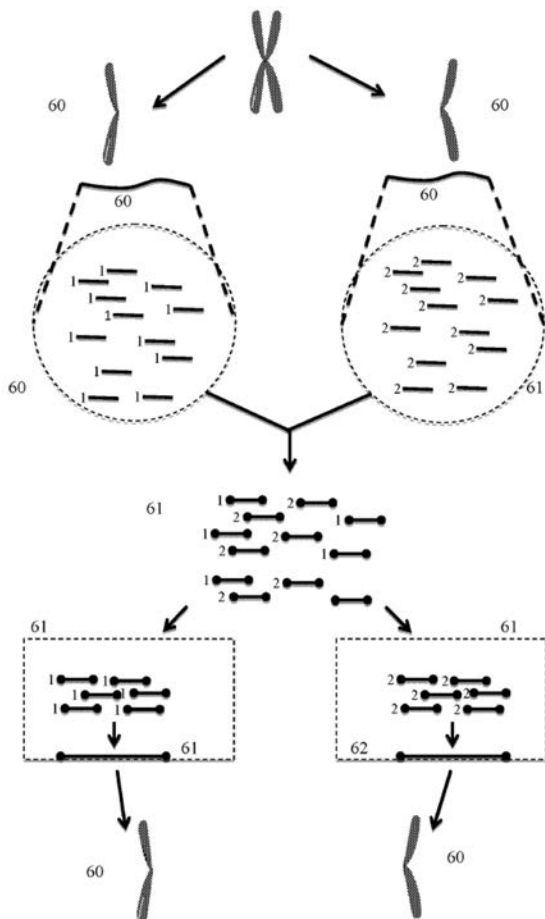
【図 3】



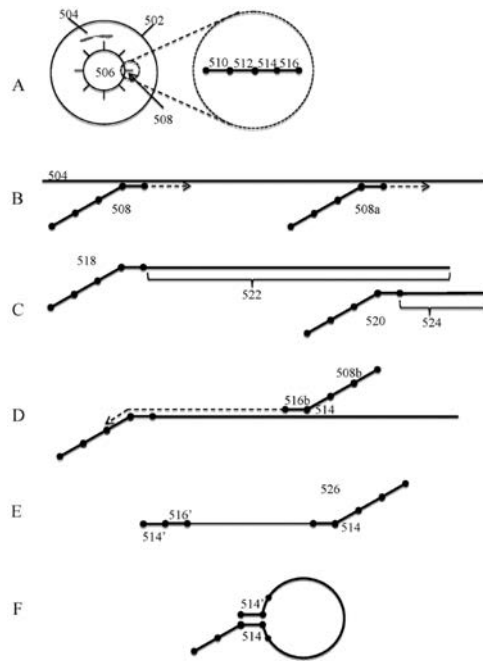
【図 4】



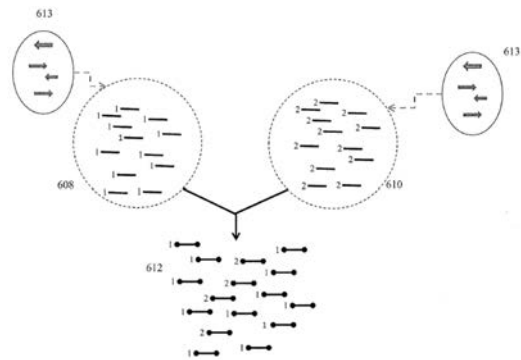
【図 6 A】



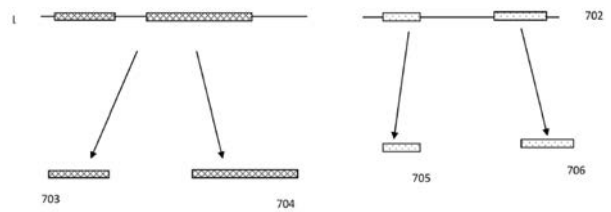
【図 5】



【図 6 B】



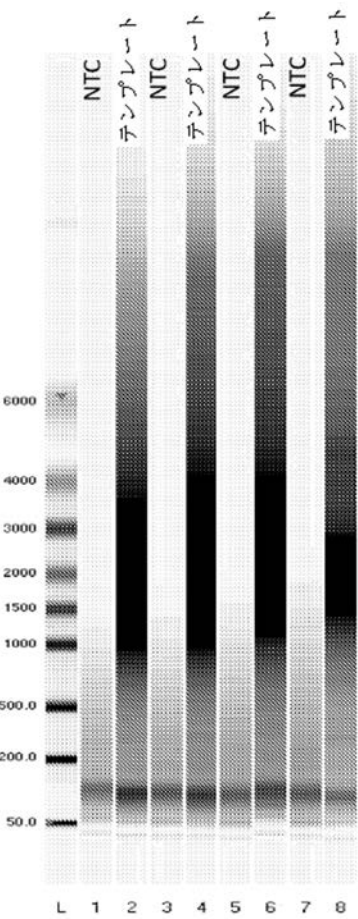
【図 7】



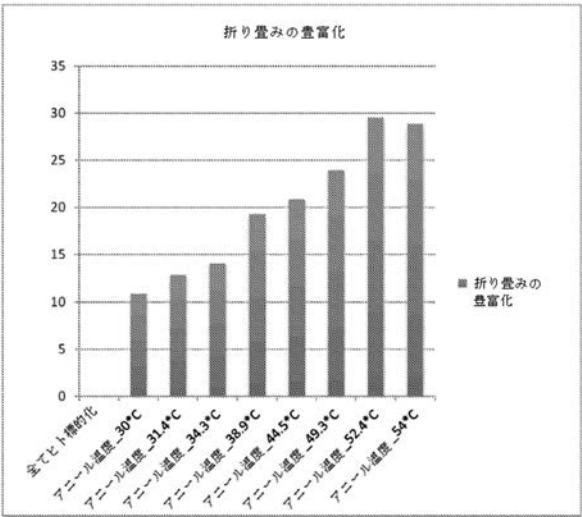
【 図 8 】



【 図 9 】



【 図 10 】



## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2016/019382

## A. CLASSIFICATION OF SUBJECT MATTER

INV. C12Q1/68  
ADD.

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)  
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, BIOSIS, EMBASE, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014/378345 A1 (HINDSON BENJAMIN [US] ET AL) 25 December 2014 (2014-12-25) abstract; claims 64-88; figures 1,39 paragraph [0007] - paragraph [0008] paragraph [0039] - paragraph [0048] paragraph [0095] - paragraph [0097] paragraph [0190] paragraph [0286] - paragraph [0291] paragraph [0296] - paragraph [0297] paragraph [0401] - paragraph [0426] paragraph [0440]	1-65, 67-93
A	WO 2013/177220 A1 (SCRIPPS RESEARCH INST [US]; HEAD STEVEN ROBERT [US]; ORDOUKHANIAN PHIL) 28 November 2013 (2013-11-28) the whole document ----- -/-	1-93

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

## \* Special categories of cited documents :

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

"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

"Z" document member of the same patent family

Date of the actual completion of the international search

25 May 2016

Date of mailing of the international search report

03/06/2016

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel: (+31-70) 340-2040,  
Fax: (+31-70) 340-3016

Authorized officer

Tilkorn, A

## INTERNATIONAL SEARCH REPORT

International application No

PCT/US2016/019382

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>JE HYUK LEE ET AL: "Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues",  NATURE PROTOCOLS,  vol. 10, no. 3,  12 February 2015 (2015-02-12), pages  442-458, XP55272042,  GB  ISSN: 1754-2189, DOI:  10.1038/nprot.2014.191  the whole document</p> <p>-----</p>	1-93
X,P	<p>GRACE X Y ZHENG ET AL: "Haplotyping germline and cancer genomes with high-throughput linked-read sequencing",  NATURE BIOTECHNOLOGY,  vol. 34, no. 3,  1 February 2016 (2016-02-01), pages  303-311, XP055272039,  US  ISSN: 1087-0156, DOI: 10.1038/nbt.3432  the whole document</p> <p>-----</p>	1,26,51, 76,91

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No

PCT/US2016/019382

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014378345	A1	25-12-2014	NONE
-----	-----	-----	-----
WO 2013177220	A1	28-11-2013	AU 2013266394 A1 26-02-2015
			CA 2874413 A1 28-11-2013
			CN 104736722 A 24-06-2015
			EP 2852687 A1 01-04-2015
			US 2015265995 A1 24-09-2015
			WO 2013177220 A1 28-11-2013
-----	-----	-----	-----



## 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 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 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

Fターム(参考) 4B063 QA08 QQ42 QQ52 QR56 QR62 QS25 QS34