

(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2017-522866

(P2017-522866A)

(43) 公表日 平成29年8月17日 (2017.8.17)

(51) Int.Cl. F 1 テーマコード (参考)
C 1 2 Q 1/68 (2006.01) C 1 2 Q 1/68 Z 4 B 0 6 3

審査請求 未請求 予備審査請求 未請求 (全 73 頁)

(21) 出願番号	特願2016-575069 (P2016-575069)	(71) 出願人	516300726
(86) (22) 出願日	平成27年6月26日 (2015.6.26)		1 O エックス ジェノミクス, インコーポレイテッド
(85) 翻訳文提出日	平成28年12月22日 (2016.12.22)		アメリカ合衆国 カリフォルニア 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
(86) 国際出願番号	PCT/US2015/038141	(74) 代理人	100078282
(87) 国際公開番号	W02015/200869		弁理士 山本 秀策
(87) 国際公開日	平成27年12月30日 (2015.12.30)	(74) 代理人	100113413
(31) 優先権主張番号	62/017,808		弁理士 森下 夏樹
(32) 優先日	平成26年6月26日 (2014.6.26)	(74) 代理人	100181674
(33) 優先権主張国	米国 (US)		弁理士 飯田 貴敏
(31) 優先権主張番号	62/072,214	(74) 代理人	100181641
(32) 優先日	平成26年10月29日 (2014.10.29)		弁理士 石川 大輔
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 核酸配列の分析

(57) 【要約】

本開示は、ハプロタイプフェージング及びコピー数変異アッセイの方法、組成物、及び系に関する。複数の独立したパーティションにおいてビーズを含むバーコードを試料と組み合わせるための方法及び系、ならびにバーコード化された試料を処理、配列決定、及び分析する方法が本開示内に含まれる。本開示は、遺伝物質の特徴付けに大きな進歩をもたらすのに有用であり得る方法及び系を提供する。これらの方法及び系は、例えば、ハプロタイプフェージング、構造的変異、例えば、欠失、重複、コピー数変異体、挿入、反転、転座、長縦列型反復 (LTR)、短縦列型反復 (STR)、及び様々な他の有用な特徴付けの特定を含む、一般的に利用可能な技術の使用が実質的に困難であり得る遺伝的特徴付けを提供するのに有用であり得る。

【選択図】 図 1 3 B

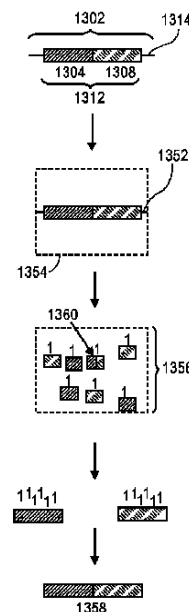


Figure 13B

【特許請求の範囲】**【請求項 1】**

核酸における 1 つ以上の変異の特定方法であって、

(a) 前記核酸の第 1 の断片であって、10 キロベース (k b) を超える長さを有する、前記第 1 の断片を提供することと、

(b) 前記第 1 の断片の複数の第 2 の断片を配列決定して、共通のバーコード配列を共有する、複数の断片配列を提供することと、

(c) 前記共通のバーコード配列の存在により、前記複数の断片配列を前記第 1 の断片に帰属させることと、

(d) 前記複数の断片配列を用いて前記第 1 の断片の核酸配列を決定することであって、前記核酸配列が 1 % 未満の誤差率で決定される、前記決定することと、 10

(e) (d) で決定された前記第 1 の断片の前記核酸配列における前記 1 つ以上の変異を特定し、それにより前記核酸内の前記 1 つ以上の変異を特定することと、を含む、前記方法。

【請求項 2】

前記第 1 の断片が複数の別個のパーティションの中の 1 つの別個のパーティション内にある、請求項 1 に記載の方法。

【請求項 3】

前記別個のパーティションがエマルジョン中の液滴である、請求項 2 に記載の方法。

【請求項 4】

前記特定が、前記第 1 の断片の前記核酸配列から前記核酸におけるフェーズ化変異体を特定することを含む、請求項 1 に記載の方法。 20

【請求項 5】

前記特定が、前記第 1 の断片の前記核酸配列から前記核酸における 1 つ以上の構造的変異を特定することを含む、請求項 1 に記載の方法。

【請求項 6】

前記第 1 の断片が 15 k b を超える長さを有する、請求項 1 に記載の方法。

【請求項 7】

前記第 1 の断片が 20 k b を超える長さを有する、請求項 1 に記載の方法。

【請求項 8】

前記決定が、前記複数の断片配列を参照にマッピングすることを含む、請求項 1 に記載の方法。 30

【請求項 9】

前記決定が、前記共通のバーコード配列を有する前記複数の断片配列をアセンブリングすることを含む、請求項 1 に記載の方法。

【請求項 10】

長さが少なくとも 10 k b である前記核酸の複数の第 1 の断片を提供することを更に含み、前記特定が、前記複数の第 1 の断片の各々から核酸配列を決定し、前記複数の第 1 の断片の各々からの前記核酸配列から前記核酸における前記 1 つ以上の変異を特定することを含む、請求項 1 に記載の方法。 40

【請求項 11】

前記複数の第 1 の断片の 2 つ以上の核酸配列を、前記 2 つ以上の核酸配列の重複する核酸配列に基づいて、推測コンティグに連結することを更に含み、最大推測コンティグ長が少なくとも 10 k b である、請求項 10 に記載の方法。

【請求項 12】

前記最大推測コンティグ長が少なくとも 20 k b である、請求項 11 に記載の方法。

【請求項 13】

前記最大推測コンティグ長が少なくとも 40 k b である、請求項 12 に記載の方法。

【請求項 14】

前記最大推測コンティグ長が少なくとも 50 k b である、請求項 13 に記載の方法。 50

【請求項 15】

前記コンティグ長が少なくとも100kbである、請求項14に記載の方法。

【請求項 16】

前記最大推測コンティグ長が少なくとも200kbである、請求項15に記載の方法。

【請求項 17】

前記最大推測コンティグ長が少なくとも500kbである、請求項16に記載の方法。

【請求項 18】

前記最大推測コンティグ長が少なくとも750kbである、請求項17に記載の方法。

【請求項 19】

前記最大推測コンティグ長が少なくとも1メガベース (Mb) である、請求項18に記載の方法。 10

【請求項 20】

前記最大推測コンティグ長が少なくとも1.75Mbである、請求項19に記載の方法。

【請求項 21】

前記最大推測コンティグ長が少なくとも2.5Mbである、請求項20に記載の方法。

【請求項 22】

前記複数の第1の断片の2つ以上の核酸配列を、前記複数の第1の断片の前記2つ以上の核酸配列内の重複するフェーズ化変異体に基づいて、フェーズブロックに連結することを更に含み、最大フェーズブロック長が少なくとも10kbである、請求項10に記載の方法。 20

【請求項 23】

前記最大フェーズブロック長が少なくとも20kbである、請求項22に記載の方法。

【請求項 24】

前記最大フェーズブロック長が少なくとも40kbである、請求項23に記載の方法。

【請求項 25】

前記最大フェーズブロック長が少なくとも50kbである、請求項24に記載の方法。

【請求項 26】

前記最大フェーズブロック長が少なくとも100kbである、請求項25に記載の方法。 30

【請求項 27】

前記最大フェーズブロック長が少なくとも200kbである、請求項26に記載の方法。

【請求項 28】

前記最大フェーズブロック長が少なくとも500kbである、請求項27に記載の方法。

【請求項 29】

前記最大フェーズブロック長が少なくとも750kbである、請求項28に記載の方法。

【請求項 30】

前記最大フェーズブロック長が少なくとも1Mbである、請求項29に記載の方法。 40

【請求項 31】

前記最大フェーズブロック長が少なくとも1.75Mbである、請求項30に記載の方法。

【請求項 32】

前記最大フェーズブロック長が少なくとも2.5Mbである、請求項31に記載の方法。

【請求項 33】

前記複数の第1の断片の2つ以上の核酸配列を、前記2つ以上の核酸配列の重複する核酸配列に基づいて、推測コンティグに連結し、それにより推測コンティグの集団を作るこ 50

とを更に含み、前記推測コンティグの集団のN50が少なくとも10kbである、請求項10に記載の方法。

【請求項34】

前記推測コンティグの集団のN50が少なくとも20kbである、請求項33に記載の方法。

【請求項35】

前記推測コンティグの集団のN50が少なくとも40kbである、請求項34に記載の方法。

【請求項36】

前記推測コンティグの集団のN50が少なくとも50kbである、請求項35に記載の方法。

10

【請求項37】

前記推測コンティグの集団のN50が少なくとも100kbである、請求項36に記載の方法。

【請求項38】

前記推測コンティグの集団のN50が少なくとも200kbである、請求項37に記載の方法。

【請求項39】

前記推測コンティグの集団のN50が少なくとも500kbである、請求項38に記載の方法。

20

【請求項40】

前記推測コンティグの集団のN50が少なくとも750kbである、請求項39に記載の方法。

【請求項41】

前記推測コンティグの集団のN50が少なくとも1Mbである、請求項40に記載の方法。

【請求項42】

前記推測コンティグの集団のN50が少なくとも1.75Mbである、請求項41に記載の方法。

【請求項43】

前記推測コンティグの集団のN50が少なくとも2.5Mbである、請求項42に記載の方法。

30

【請求項44】

前記複数の第1の断片の2つ以上の核酸配列を、前記複数の第1の断片の前記2つ以上の核酸配列内の重複するフェーズ化変異体に基づいて、フェーズブロックに連結し、それによりフェーズブロックの集団を作ることを含み、前記フェーズブロックの集団のN50が少なくとも10kbである、請求項10に記載の方法。

【請求項45】

前記フェーズブロックの集団のN50が少なくとも20kbである、請求項44に記載の方法。

40

【請求項46】

前記フェーズブロックの集団のN50が少なくとも40kbである、請求項45に記載の方法。

【請求項47】

前記フェーズブロックの集団のN50が少なくとも50kbである、請求項46に記載の方法。

【請求項48】

前記フェーズブロックの集団のN50が少なくとも100kbである、請求項47に記載の方法。

【請求項49】

50

前記フェーズブロックの集団のN50が少なくとも200kbである、請求項48に記載の方法。

【請求項50】

前記フェーズブロックの集団のN50が少なくとも500kbである、請求項49に記載の方法。

【請求項51】

前記フェーズブロックの集団のN50が少なくとも750kbである、請求項50に記載の方法。

【請求項52】

前記フェーズブロックの集団のN50が少なくとも1Mbである、請求項51に記載の方法。

【請求項53】

前記フェーズブロックの集団のN50が少なくとも1.75Mbである、請求項52に記載の方法。

【請求項54】

前記フェーズブロックの集団のN50が少なくとも2.5Mbである、請求項53に記載の方法。

【請求項55】

核酸の構造的変異の存在の決定方法であって、

(a) 前記核酸の複数の第1の断片分子を提供することであって、前記複数の第1の断片分子の所与の第1の断片分子が前記構造的変異を含む、前記提供することと、

(b) 前記複数の第1の断片分子の各々の複数の第2の断片分子を配列決定して複数の断片配列を提供することであって、所与の第1の断片分子に対応する前記複数の断片配列の各々が共通のバーコード配列を共有する、前記提供することと、

(c) (i) 前記複数の断片配列を参照配列にマッピングすること、(ii) 前記共通のバーコード配列を共有する前記複数の断片配列を特定すること、及び(iii) 前記所与の第1の断片分子の長さよりも更に離れている位置の前記参照配列にマッピングする、前記共通のバーコード配列を共有する上昇した量の前記複数の断片配列の存在に基づき前記構造的変異を特定することであって、前記上昇した量が前記構造的変異を欠如する配列に対してである、前記特定することにより、前記構造的変異の存在を決定することと、を含む、前記方法。

【請求項56】

前記上昇した量が、前記構造的変異を有する前記核酸の領域に由来する前記第1の断片分子の総数に対して1%以上である、請求項55に記載の方法。

【請求項57】

前記上昇した量が、前記構造的変異を有する前記核酸の領域に由来する前記第1の断片分子の総数に対して2%以上である、請求項56に記載の方法。

【請求項58】

前記位置が少なくとも約100塩基離れている、請求項55に記載の方法。

【請求項59】

前記位置が少なくとも約500塩基離れている、請求項58に記載の方法。

【請求項60】

前記位置が少なくとも約1キロベース(kb)離れている、請求項59に記載の方法。

【請求項61】

前記位置が少なくとも約10kb離れている、請求項60に記載の方法。

【請求項62】

前記複数の断片配列から前記所与の第1の断片分子のアセンブリを作ることにより前記構造的変異を特定することを更に含み、前記複数の断片配列が前記共通のバーコード配列の存在に基づいて前記アセンブリの入力として選択される、請求項55に記載の方法。

【請求項63】

10

20

30

40

50

前記アセンブリが、前記複数の断片配列からコンセンサス配列を生成することにより作られる、請求項 6 2 に記載の方法。

【請求項 6 4】

前記構造的変異が転座を含む、請求項 5 5 に記載の方法。

【請求項 6 5】

変異体核酸配列の特徴付け方法であって、

(a) 変異体核酸を断片化して 1 0 キロベース (k b) を超える長さを有する複数の第 1 の断片を提供することと、

(b) 前記複数の第 1 の断片を別個のパーティションに分離することと、

(c) 各第 1 の断片からそのそれぞれのパーティション内に複数の第 2 の断片を作ることであって、前記複数の第 2 の断片がそれに付加されるバーコード配列を有し、所与のパーティション内の前記バーコード配列が共通のバーコード配列である、前記作ることと、

(d) 前記複数の第 2 の断片及びそれに付加される前記バーコード配列を配列決定して複数の第 2 の断片の配列を得ることと、

(e) 少なくとも部分的に前記共通のバーコード配列の存在に基づき、前記第 2 の断片の配列を元の第 1 の断片に帰属させて、前記第 2 の断片の配列のための第 1 の断片の配列環境 (sequence context) を提供することと、

(f) 前記第 1 の断片の配列環境から前記変異体核酸の変異体部分を特定し、それにより前記変異体核酸配列を特徴付けることと、を含む、前記方法。

【請求項 6 6】

前記帰属が、少なくとも部分的に前記共通のバーコード配列の存在に基づき、前記複数の第 2 の断片の配列から前記複数の第 1 の断片からの個々の断片の配列の少なくとも一部をアセンブリングすることを含む、請求項 6 5 に記載の方法。

【請求項 6 7】

前記帰属が、少なくとも部分的に前記共通のバーコード配列に基づき、前記複数の第 2 の断片の配列を前記複数の第 1 の断片からの個々の第 1 の断片にマッピングすることを含む、請求項 6 5 に記載の方法。

【請求項 6 8】

前記複数の第 1 の断片のうちの 2 つ以上を、前記複数の第 1 の断片のうちの前記 2 つ以上の間の重複する配列に基づき、推測コンティグに連結することを更に含む、請求項 6 5 に記載の方法。

【請求項 6 9】

前記特定が、前記第 1 の断片の配列環境から 1 つ以上のフェーズ化変異体を特定することを含む、請求項 6 5 に記載の方法。

【請求項 7 0】

前記複数の第 1 の断片のうちの 2 つ以上を、前記複数の第 1 の断片のうちの前記 2 つ以上の間の重複するフェーズ化変異体に基づき、フェーズブロックに連結することを更に含む、請求項 6 5 に記載の方法。

【請求項 7 1】

前記特定が、前記第 1 の断片の配列環境から 1 つ以上の構造的変異を特定することを含む、請求項 7 0 に記載の方法。

【請求項 7 2】

前記 1 つ以上の構造的変異が、挿入、欠失、転座、レトロトランスポゾン、反転、及び重複から独立して選択される、請求項 7 1 に記載の方法。

【請求項 7 3】

前記構造的変異が挿入または転座を含み、前記第 1 の断片配列環境が前記挿入または転座の存在を示す、請求項 7 2 に記載の方法。

【請求項 7 4】

核酸の配列における変異体の特定方法であって、

前記核酸の複数の個々の断片分子の核酸配列であって、各々少なくとも 1 キロベース (

10

20

30

40

50

k b) の長さを有する、前記複数の個々の断片分子の核酸配列を得ることと、

前記複数の個々の断片分子のうちの1つ以上の配列を、1つ以上の推測コンティグに連結することと、

前記1つ以上の推測コンティグから1つ以上の変異体を特定することと、を含む、前記方法。

【請求項75】

前記得ることが、長さが10 kbを超える複数の断片分子の前記核酸配列を得ることを含む、請求項74に記載の方法。

【請求項76】

前記得ることが、

前記複数の個々の断片分子の各個々の断片分子の複数のバーコード化された断片を提供することであって、所与の個々の断片分子の前記バーコード化された断片が共通のバーコードを有する、前記提供することと、

前記複数の個々の断片分子の前記複数のバーコード化された断片を配列決定することであって、前記配列決定が1%未満の配列決定誤差率を提供する、前記決定することと、

前記複数のバーコード化された断片の配列及びそれらの関連バーコードから前記複数の個々の断片分子の配列を決定することと、を含む、請求項75に記載の方法。

【請求項77】

前記連結が、2つ以上の個々の断片分子間における1つ以上の重複配列を特定して、前記2つ以上の個々の断片分子を前記1つ以上の推測コンティグに連結することを含む、請求項76に記載の方法。

【請求項78】

前記連結が、2つ以上の個々の断片分子間における1つ以上の共通の変異体を特定して、前記2つ以上の個々の断片分子を前記1つ以上の推測コンティグに連結することを含む、請求項77に記載の方法。

【請求項79】

前記1つ以上の共通の変異体がフェーズ化変異体であり、前記1つ以上の推測コンティグが少なくとも100 kbの最大フェーズブロック長を含む、請求項78に記載の方法。

【請求項80】

前記特定において特定された前記1つ以上の変異体が構造的変異を含む、請求項74に記載の方法。

【請求項81】

前記構造的変異が、挿入、欠失、転座、レトロトランスポゾン、反転、及び重複から選択される、請求項80に記載の方法。

【請求項82】

核酸の特徴付け方法であって、

少なくとも10キロベース(kb)の長さを有する複数の断片分子の核酸配列を得ることと、

前記複数の断片分子の前記核酸配列における1つ以上のフェーズ化変異体の位置を特定することと、

少なくとも第1の断片分子の前記核酸配列を少なくとも第2の断片分子に、前記第1及び第2の断片分子内の1つ以上の共通のフェーズ化変異体位置の存在に基づいて連結して、少なくとも10 kbの最大フェーズブロック長のフェーズブロックを提供することと、

前記少なくとも10 kbの最大フェーズブロック長の前記フェーズブロックから1つ以上のフェーズ化変異体を特定することと、を含む、前記方法。

【請求項83】

前記フェーズブロックから1つ以上の追加のフェーズ化変異体を特定することを更に含む、請求項82に記載の方法。

【請求項84】

前記複数の断片分子が別個のパーティション内にある、請求項82に記載の方法。

10

20

30

40

50

【請求項 85】

前記別個のパーティションがエマルジョン中の液滴である、請求項 84 に記載の方法。

【請求項 86】

前記複数の断片分子の前記長さが少なくとも 50 kb である、請求項 82 に記載の方法。

【請求項 87】

前記複数の断片分子の前記長さが少なくとも 100 kb である、請求項 86 に記載の方法。

【請求項 88】

前記最大フェーズブロック長が少なくとも 50 kb である、請求項 82 に記載の方法。

10

【請求項 89】

前記最大フェーズブロック長が少なくとも 100 kb である、請求項 88 に記載の方法。

【請求項 90】

前記最大フェーズブロック長が少なくとも 1 Mb である、請求項 89 に記載の方法。

【請求項 91】

前記最大フェーズブロック長が少なくとも 2 Mb である、請求項 90 に記載の方法。

【請求項 92】

前記最大フェーズブロック長が少なくとも 2.5 Mb である、請求項 91 に記載の方法。

20

【請求項 93】

方法であって、

a) 第 1 の核酸であって、生物の第 1 の染色体に由来する標的配列を含む、前記第 1 の核酸を第 1 のパーティションに分配することと、

b) 第 2 の核酸であって、前記生物の第 2 の染色体に由来する標的配列を含む、前記第 2 の核酸を第 2 のパーティションに分配することと、

c) 前記第 1 のパーティションにおいて、第 1 のバーコード配列を前記第 1 の核酸の断片または前記第 1 の核酸の一部のコピーに付加して、第 1 のバーコード化された断片を提供することと、

d) 前記第 2 のパーティションにおいて、第 2 のバーコード配列であって、前記第 1 のバーコード配列とは異なる、前記第 2 のバーコード配列を前記第 2 の核酸の断片または前記第 2 の核酸の一部のコピーに付加して、第 2 のバーコード化された断片を提供することと、

30

e) 前記第 1 及び第 2 のバーコード化された断片の前記核酸配列を決定し、前記第 1 及び第 2 の核酸の核酸配列をアセンブリングすることと、

f) 前記第 1 及び第 2 の核酸の前記核酸配列を比較して、それぞれ、第 1 及び第 2 の染色体に由来するとして前記第 1 及び第 2 の核酸を特徴付けることと、を含む前記方法。

【請求項 94】

前記第 1 のバーコード配列を含むオリゴヌクレオチドが前記第 1 の核酸と共分配され (copartitioned)、前記第 2 のバーコード配列を含むオリゴヌクレオチドが前記第 2 の核酸と共分配される、請求項 93 に記載の方法。

40

【請求項 95】

前記第 1 のバーコード配列を含む前記オリゴヌクレオチドが第 1 のビーズに放出可能に付加され、前記第 2 のバーコード配列を含む前記オリゴヌクレオチドが第 2 のビーズに放出可能に付加され、前記共分配が、それぞれ、前記第 1 及び第 2 のビーズを前記第 1 及び第 2 のパーティションに共分配することを含む、請求項 94 に記載の方法。

【請求項 96】

前記第 1 及び第 2 のパーティションがエマルジョン中の液滴を含む、請求項 93 に記載の方法。

【請求項 97】

50

前記第 1 の染色体が父方染色体であり、前記第 2 の染色体が母方染色体である、請求項 9 3 に記載の方法。

【請求項 9 8】

前記第 1 の染色体及び前記第 2 の染色体が相同染色体である、請求項 9 3 に記載の方法。

【請求項 9 9】

前記第 1 の核酸及び前記第 2 の核酸が 1 つ以上の変異を含む、請求項 9 3 に記載の方法。

【請求項 1 0 0】

方法であって、

10

a) 第 1 の核酸であって、生物の第 1 の染色体に由来する標的配列を含む、前記第 1 の核酸を第 1 のパーティションに分配することと、

b) 第 2 の核酸であって、前記生物の第 2 の染色体に由来する標的配列を含む、前記第 2 の核酸を第 2 のパーティションに分配することと、

c) 前記第 1 のパーティションにおいて、第 1 のバーコード配列を前記第 1 の核酸の断片または前記第 1 の核酸の一部のコピーに付加して、第 1 のバーコード化された断片を提供することと、

d) 前記第 2 のパーティションにおいて、第 2 のバーコード配列であって、前記第 1 のバーコード配列とは異なる、前記第 2 のバーコード配列を前記第 2 の核酸の断片または前記第 2 の核酸の一部のコピーに付加して、第 2 のバーコード化された断片を提供することと、

20

e) 前記第 1 及び第 2 のバーコード化された断片の前記核酸配列を決定し、前記第 1 及び第 2 の核酸の核酸配列をアセンブリングすることと、

f) 前記第 1 及び第 2 の核酸の前記核酸配列を比較して、前記第 1 及び第 2 の核酸の前記核酸配列間の任意の変異を特定することと、を含む、前記方法。

【請求項 1 0 1】

前記第 1 のバーコード配列を含むオリゴヌクレオチドが前記第 1 の核酸と共分配され、前記第 2 のバーコード配列を含むオリゴヌクレオチドが前記第 2 の核酸と共分配される、請求項 1 0 0 に記載の方法。

【請求項 1 0 2】

30

前記第 1 のバーコード配列を含む前記オリゴヌクレオチドが第 1 のビーズに放出可能に付加され、前記第 2 のバーコード配列を含む前記オリゴヌクレオチドが第 2 のビーズに放出可能に付加され、前記共分配が、それぞれ、前記第 1 及び第 2 のビーズを前記第 1 及び第 2 のパーティションに共分配することを含む、請求項 1 0 1 に記載の方法。

【請求項 1 0 3】

前記第 1 及び第 2 のパーティションがエマルジョン中の液滴を含む、請求項 1 0 0 に記載の方法。

【請求項 1 0 4】

前記第 1 の染色体が父方染色体であり、前記第 2 の染色体が母方染色体である、請求項 1 0 0 に記載の方法。

40

【請求項 1 0 5】

前記第 1 の染色体及び前記第 2 の染色体が相同染色体である、請求項 1 0 0 に記載の方法。

【請求項 1 0 6】

前記第 1 の核酸及び前記第 2 の核酸が 1 つ以上の変異を含む、請求項 1 0 0 に記載の方法。

【請求項 1 0 7】

前記第 1 及び第 2 の染色体が胎児に由来する、請求項 9 3 または請求項 1 0 0 に記載の方法。

【請求項 1 0 8】

50

前記第 1 及び第 2 の核酸が妊婦から採取した試料から得られる、請求項 9 3 または請求項 1 0 0 に記載の方法。

【請求項 1 0 9】

前記第 1 の染色体が染色体 2 1、1 8、または 1 3 である、請求項 9 3 または請求項 1 0 0 に記載の方法。

【請求項 1 1 0】

前記第 2 の染色体が染色体 2 1、1 8、または 1 3 である、請求項 9 3 または請求項 1 0 0 に記載の方法。

【請求項 1 1 1】

前記第 1 または第 2 の染色体の相対量を決定することを更に含む、請求項 9 3 または請求項 1 0 0 に記載の方法。

10

【請求項 1 1 2】

参照染色体に対する前記第 1 または第 2 の染色体の量を決定することを更に含む、請求項 9 3 または請求項 1 0 0 に記載の方法。

【請求項 1 1 3】

前記第 1 の染色体もしくは第 2 の染色体またはその両方のコピー数が増加する、請求項 9 3 または請求項 1 0 0 に記載の方法。

【請求項 1 1 4】

前記コピー数の増加が癌または異数体の結果である、請求項 1 1 3 に記載の方法。

【請求項 1 1 5】

前記第 1 の染色体もしくは第 2 の染色体またはその両方のコピー数が減少する、請求項 9 3 または請求項 1 0 0 に記載の方法。

20

【請求項 1 1 6】

前記コピー数の減少が癌または異数体の結果である、請求項 1 1 5 に記載の方法。

【請求項 1 1 7】

胎児核酸配列の特徴付け方法であって、

(a) 母方核酸配列を決定することであって、前記母方核酸が、胎児の妊娠中の母親に由来し、

(i) 母方核酸を断片化して複数の第 1 の母方断片を提供すること、

(i i) 前記複数の第 1 の母方断片を母方パーティション内に分離すること、

30

(i i i) 前記第 1 の母方断片の各々からそれらのそれぞれの母方パーティション内に複数の第 2 の母方断片を作ることであって、前記複数の第 2 の母方断片がそれに付加される第 1 のバーコード配列を有し、前記母方パーティションの所与の母方パーティション内で、前記第 2 の母方断片がそれに付加される第 1 の共通のバーコード配列を含む、前記作ること、

(i v) 前記複数の第 2 の母方断片を配列決定して複数の母方断片配列を提供すること、

(v) 少なくとも部分的に前記第 1 の共通のバーコード配列の存在に基づき、前記母方断片配列を元の第 1 の母方断片に帰属させて前記母方核酸配列を決定することにより、前記決定することと、

(b) 父方核酸配列を決定することであって、前記父方核酸が、前記胎児の父親に由来し、

40

(i) 父方核酸を断片化して複数の第 1 の父方断片を提供すること、

(i i) 前記複数の第 1 の父方断片を父方の別個のパーティション内に分離すること、

(i i i) 各第 1 の父方断片からそのそれぞれのパーティション内に複数の第 2 の父方断片を作ることであって、前記複数の第 2 の父方断片がそれに付加される第 2 のバーコード配列を有し、所与の父方パーティション内で、前記第 2 の父方断片がそれに付加される第 2 の共通のバーコード配列を含む、作ること、

(i v) 前記複数の第 2 の父方断片及びそれに付加される前記第 2 のバーコード配列を配列決定して複数の父方断片配列を提供すること、

(v) 少なくとも部分的に前記第 2 の共通のバーコード配列の存在に基づき、前記父方

50

断片配列を元の第1の父方断片に帰属させて前記父方核酸配列を決定することにより、前記決定することと、

(c) 前記妊娠中の母親から胎児核酸を得、前記母方核酸配列及び前記父方核酸配列を使用して、前記胎児核酸の配列及び／または前記胎児核酸の前記配列の1つ以上の遺伝的変異を決定することと、を含む、前記方法。

【請求項118】

前記父方断片配列及び前記母方断片配列が各々、配列を1つ以上の推測コンティグに連結するために使用される、請求項117に記載の方法。

【請求項119】

前記推測コンティグが、母方及び父方のフェーズブロックを構築するために使用される、請求項118に記載の方法。

【請求項120】

前記胎児核酸の前記配列が、胎児フェーズブロックを構築するために前記母方及び父方のフェーズブロックと比較される、請求項119に記載の方法。

【請求項121】

前記父方断片配列が前記複数の第1の父方断片の配列の少なくとも一部を生成するためにアセンブリングされ、それにより前記父方核酸配列を決定し、前記母方断片配列が前記複数の第1の母方断片の配列の少なくとも一部を生成するためにアセンブリングされ、それにより前記母方核酸配列を決定する、請求項120に記載の方法。

【請求項122】

前記父方核酸配列の前記決定が前記父方断片配列を父方参照にマッピングすることを含み、前記母方核酸配列の前記決定が前記母方断片配列を母方参照にマッピングすることを含む、請求項117に記載の方法。

【請求項123】

前記胎児核酸の前記配列が少なくとも99%の精度で決定される、請求項117に記載の方法。

【請求項124】

前記胎児核酸の前記配列の前記1つ以上の遺伝的変異が少なくとも99%の精度で決定される、請求項117に記載の方法。

【請求項125】

前記1つ以上の遺伝的変異が、構造的変異及び一塩基多型(SNP)からなる群から選択される、請求項117に記載の方法。

【請求項126】

前記1つ以上の遺伝的変異が、コピー数変異、挿入、欠失、転座、レトロトランスポゾン、反転、再編成、反復配列伸長、及び重複からなる群から選択される構造的変異である、請求項117に記載の方法。

【請求項127】

(c)において、前記母方核酸配列及び前記父方核酸配列について決定された1つ以上の遺伝的変異を使用して、前記胎児核酸の前記配列の前記1つ以上の遺伝的変異を決定することを更に含む、請求項117に記載の方法。

【請求項128】

(c)において、前記胎児核酸の1つ以上の新規突然変異を決定することを更に含む、請求項117に記載の方法。

【請求項129】

(c)の間または後に、前記胎児核酸に関連する異数体を決定することを更に含む、請求項117に記載の方法。

【請求項130】

(a)の(v)の間または後に、前記母方核酸配列をハプロタイプ分析(haplotyping)してハプロタイプが解明された母方核酸配列を提供することと、(b)の(v)の間または後に、前記父方核酸配列をハプロタイプ分析してハプロタイプが解明され

10

20

30

40

50

た父方核酸配列を提供することと、を更に含む、請求項 1 1 7 に記載の方法。

【請求項 1 3 1】

(c) において、前記ハプロタイプが解明された母方核酸配列及び前記ハプロタイプが解明された父方核酸配列を使用して、前記胎児核酸の前記配列及び／または前記 1 つ以上の遺伝的変異を決定することを更に含む、請求項 1 3 0 に記載の方法。

【請求項 1 3 2】

前記母方核酸及び前記父方核酸のうちの 1 つ以上がゲノムデオキシリボ核酸 (DNA) である、請求項 1 1 7 に記載の方法。

【請求項 1 3 3】

(c) において、前記胎児核酸が無細胞核酸を含む、請求項 1 1 7 に記載の方法。

10

【請求項 1 3 4】

(a) において、少なくとも 99% の精度で前記母方核酸配列を決定することを更に含む、請求項 1 1 7 に記載の方法。

【請求項 1 3 5】

(b) において、少なくとも 99% の精度で前記父方核酸配列を決定することを更に含む、請求項 1 1 7 に記載の方法。

【請求項 1 3 6】

前記母方核酸配列及び／または前記父方核酸配列が 10 キロベース (kb) を超える長さを有する、請求項 1 1 7 に記載の方法。

【請求項 1 3 7】

前記母方及び父方のパーティションがエマルジョン中の液滴を含む、請求項 1 1 7 に記載の方法。

20

【請求項 1 3 8】

(a) において、前記第 1 のバーコード配列が、第 1 の粒子に放出可能に付加された前記所与の母方パーティションにおいて提供される、請求項 1 1 7 に記載の方法。

【請求項 1 3 9】

(b) において、前記第 2 のバーコード配列が、第 2 の粒子に放出可能に付加された前記所与の父方パーティションにおいて提供される、請求項 1 3 8 に記載の方法。

【請求項 1 4 0】

試料核酸の特徴付け方法であって、

30

(a) 生物学的試料であって、無細胞試料核酸を含む、前記生物学的試料を対象から得ることと、

(b) 液滴において、バーコード配列を前記無細胞試料核酸の断片または前記試料核酸の一部のコピーに付加してバーコード化された試料断片を得ることと、

(c) 前記バーコード化された試料断片の核酸配列を決定し、前記バーコード化された試料断片の前記核酸配列に基づき試料核酸配列を提供することと、

(d) プログラム化されたコンピュータプロセッサを使用して前記試料核酸配列の参照核酸配列との比較を生成することであって、前記参照核酸配列が 10 キロベース (kb) を超える長さ及び少なくとも 99% の精度を有する、前記生成することと、

(e) 前記比較を使用して前記試料核酸配列における 1 つ以上の遺伝的変異を特定し、それにより前記試料核酸を疾患と関連付けることと、を含む、前記方法。

40

【請求項 1 4 1】

前記試料核酸配列における前記 1 つ以上の遺伝的変異が、構造的変異及び一塩基多型 (SNP) からなる群から選択される、請求項 1 4 0 に記載の方法。

【請求項 1 4 2】

前記試料核酸配列の前記 1 つ以上の遺伝的変異が、コピー数変異、挿入、欠失、レトロトランスポゾン、転座、反転、再編成、反復配列伸長、及び重複からなる群から選択される構造的変異である、請求項 1 4 0 に記載の方法。

【請求項 1 4 3】

(c) において、前記試料核酸配列が少なくとも 99% の精度で提供される、請求項 1

50

40に記載の方法。

【請求項144】

(b)において、前記バーコード配列が粒子に放出可能に付加された前記液滴に提供され、(b)が、前記バーコード配列を前記付加する前に、前記粒子から前記液滴中に前記バーコード配列を放出することを更に含む、請求項140に記載の方法。

【請求項145】

(b)において、前記バーコード配列が前記粒子に放出可能に付加されたプライマー配列の一部として提供され、前記プライマー配列がランダムN-mer配列も含み、(b)が、前記バーコード配列を前記付加する前に、前記粒子から前記液滴中に前記プライマー配列を放出することを更に含む、請求項144に記載の方法。

10

【請求項146】

(b)において、前記プライマーを使用して、増幅反応において、前記バーコード配列を前記無細胞試料核酸の前記断片または前記無細胞試料核酸の一部の前記コピーに付加することを更に含む、請求項145に記載の方法。

【請求項147】

(i)追加の液滴において、追加のバーコード配列を参照核酸の断片または前記参照核酸の一部のコピーに付加してバーコード化された参照断片を提供すること、及び(ii)前記バーコード化された参照断片の核酸配列を決定し、前記バーコード化された参照断片の前記核酸配列に基づき前記参照核酸配列を決定することを更に含む、請求項140に記載の方法。

20

【請求項148】

前記参照核酸配列の前記決定が、前記バーコード化された参照断片の前記核酸配列をアセンブリングすることを含む、請求項147に記載の方法。

【請求項149】

粒子に放出可能に付加された前記追加の液滴に前記追加のバーコード配列を提供することと、前記追加のバーコード配列を前記付加する前に、前記粒子から前記追加のパーティション中に前記追加のバーコード配列を放出することと、を更に含む、請求項147に記載の方法。

【請求項150】

前記粒子に放出可能に付加されたプライマー配列の一部として前記追加のバーコード配列を提供することであって、前記プライマー配列がランダムN-mer配列も含む、前記提供することと、前記追加のバーコード配列を前記付加する前に、前記粒子から前記追加の液滴中に前記プライマーを放出することと、を更に含む請求項149に記載の方法。

30

【請求項151】

前記プライマーを使用して、増幅反応において、前記追加のバーコード配列を前記参照核酸の前記断片または前記参照核酸の一部の前記コピーに付加することを更に含む、請求項150に記載の方法。

【請求項152】

前記参照核酸配列において1つ以上の遺伝的変異を決定することを更に含む、請求項147に記載の方法。

40

【請求項153】

前記参照核酸配列における前記1つ以上の遺伝的変異が、構造的変異及び一塩基多型(SNP)からなる群から選択される、請求項152に記載の方法。

【請求項154】

前記参照核酸配列における前記1つ以上の遺伝的変異が、コピー数変異、挿入、欠失、レトロトランスポゾン、転座、反転、再編成、反復配列伸長、及び重複からなる群から選択される構造的変異である、請求項152に記載の方法。

【請求項155】

前記参照核酸が生殖系列核酸配列を含む、請求項147に記載の方法。

【請求項156】

50

前記参照核酸が癌核酸配列を含む、請求項 1 4 7 に記載の方法。

【請求項 1 5 7】

前記試料核酸配列が 1 0 k b を超える長さを有する、請求項 1 4 0 に記載の方法。

【請求項 1 5 8】

前記参照核酸が疾患状態の不在を示すゲノムに由来する、請求項 1 4 7 に記載の方法。

【請求項 1 5 9】

前記参照核酸が疾患状態を示すゲノムに由来する、請求項 1 4 7 に記載の方法。

【請求項 1 6 0】

前記疾患状態が癌を含む、請求項 1 5 8 または 1 5 9 のいずれか一項に記載の方法。

【請求項 1 6 1】

前記疾患状態が異数体を含む、請求項 1 5 8 または 1 5 9 のいずれか一項に記載の方法

10

【請求項 1 6 2】

前記無細胞試料核酸が腫瘍核酸を含む、請求項 1 4 0 に記載の方法。

【請求項 1 6 3】

前記腫瘍核酸が循環腫瘍核酸を含む、請求項 1 6 2 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

関連出願

20

本出願は、2 0 1 4 年 6 月 2 6 日に出願された米国仮特許出願第 6 2 / 0 1 7 , 8 0 8 号、及び 2 0 1 4 年 1 0 月 2 9 日に出願された米国仮特許出願第 6 2 / 0 7 2 , 2 1 4 号に対して優先権を主張し、それらの各々は、すべての目的において参照によりその全体が本明細書に組み込まれる。

【背景技術】

【0 0 0 2】

特定のヒトゲノムの基本認識は、突然変異などのある特定の遺伝的変異の存在または不在を単に特定する以上のことを要する場合がある。ある特定の遺伝的変異が同じまたは異なる染色体上に現れるか（フェージングとしても知られる）を決定することも重要である。遺伝子のコピー数についての情報と同様、ハプロタイプなどの遺伝的変異のパターンについての情報も重要である。

30

【0 0 0 3】

「ハプロタイプ」という用語は、連続ブロックと一緒に受け継がれる D N A 配列変異体（対立遺伝子）の組を指す。一般に、ヒトゲノムは、母方コピーと父方コピーの各遺伝子の 2 つのコピーを含有する。各々が 2 つの可能な対立遺伝子、例えば遺伝子の対立遺伝子「A」と「a」及び遺伝子の対立遺伝子「B」と「b」を有する一对の遺伝子において、所与の個体のゲノムは、A 及び B の対立遺伝子が同じ染色体上に存在する「A B / a b」（「シス」構成）、または A 及び B の対立遺伝子が異なる染色体上に存在する「A b / a B」（「トランス」構成）の 2 つのハプロタイプのうちの 1 つを含む。フェージング方法またはアッセイは、指定の対立遺伝子の組が同じまたは異なる染色体上に存在するかを決定するために使用され得る。いくつかの場合では、ハプロタイプを定義するいくつかの連結対立遺伝子は特定の疾患の表現型と相関または関連する可能性があり、このような場合において、いずれか 1 つの特定の遺伝的変異体よりもむしろハプロタイプが、患者が疾患を呈するかに関して最も決定的な要因であり得る。

40

【0 0 0 4】

遺伝子コピー数もいくつかの疾患の表現型において役割を担う。大半の遺伝子は通常 2 つのコピーで存在するが、増幅遺伝子は 2 つを超える機能性コピーで存在する遺伝子である。いくつかの場合では、遺伝子は機能性コピーの減少を受ける場合もある。遺伝子コピー数の減少または増加は異常レベルの m R N A 及びタンパク質発現の生成につながる可能性があり、癌性状態または他の障害につながる可能性がある。癌及び他の遺伝的障害は、

50

異常な（増加または減少した）染色体数（「異数体」）と相関することが多い。蛍光インサイツハイブリダイゼーションまたは比較ゲノムハイブリダイゼーションなどの細胞遺伝学的手法が異常遺伝子数または異常染色体コピー数の存在を検出するために使用され得る。遺伝的フェージングの情報、ハプロタイプ、またはコピー数変異を検出する改善された方法が当該技術分野において必要である。

【発明の概要】

【課題を解決するための手段】

【0005】

本開示は、遺伝物質の特徴付けに大きな進歩をもたらすのに有用であり得る方法及び系を提供する。これらの方法及び系は、例えば、ハプロタイプフェージング、構造的変異、例えば、欠失、重複、コピー数変異体、挿入、反転、転座、長縦列型反復（LTR）、短縦列型反復（STR）、及び様々な他の有用な特徴付けの特定を含む、一般的に利用可能な技術の使用が実質的に困難であり得る遺伝的特徴付けを提供するのに有用であり得る。

【0006】

本開示の態様は、核酸における1つ以上の変異を特定するための方法を提供し、a) 核酸の第1の断片であって、10キロベース（kb）を超える長さを有する、第1の断片を提供すること、（b）第1の断片の複数の第2の断片を配列決定して、共通のバーコード配列を共有する、複数の断片配列を提供すること、（c）共通のバーコード配列の存在により、複数の断片配列を第1の断片に帰属させること、（d）複数の断片配列を使用して第1の断片の核酸配列を決定することであって、該核酸配列が1%未満の誤差率で決定される、決定すること、及び（e）（d）で決定された第1の断片の核酸配列における1つ以上の変異を特定し、それにより核酸内の1つ以上の変異を特定することを含む。

【0007】

いくつかの場合では、第1の断片は複数の別個のパーティションの中の1つの別個のパーティション内にある。いくつかの場合では、別個のパーティションはエマルジョン中の液滴である。いくつかの場合では、特定は、第1の断片の核酸配列におけるフェーズ化変異体を特定することを含む。いくつかの場合では、特定は、第1の断片の核酸配列から核酸における1つ以上の構造的変異を特定することを含む。いくつかの場合では、第1の断片は15 kbを超える長さを有する。いくつかの場合では、第1の断片は20 kbを超える長さを有する。いくつかの場合では、決定は、複数の断片配列を参照にマッピングすることを含む。いくつかの場合では、決定は、共通のバーコード配列を有する複数の断片配列をアセンブリングすることを含む。

【0008】

いくつかの場合では、1つ以上の変異を特定するための方法は、長さが少なくとも10 kbである核酸の複数の第1の断片を提供することを更に含み、該特定は、複数の第1の断片の各々から核酸配列を決定し、複数の第1の断片の各々からの核酸配列から核酸における1つ以上の変異を特定することを含む。

【0009】

いくつかの場合では、1つ以上の変異を特定するための方法は、2つ以上の核酸配列の重複する核酸配列に基づいて、複数の第1の断片の2つ以上の核酸配列を推測コンティグに連結することを更に含み、最大推測コンティグ長は少なくとも10 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも20 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも40 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも50 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも100 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも200 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも500 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも750 kbである。いくつかの場合では、最大推測コンティグ長は少なくとも1メガベース（Mb）である。いくつかの場合では、最大推測コンティグ長は少なくとも1.75 Mbである。いくつかの場合では、最大推測コンティグ長は少なくとも2.5 Mbである。

【0010】

いくつかの場合では、1つ以上の変異を特定するための方法は、複数の第1の断片の2つ以上の核酸配列内の重複するフェーズ化変異体に基づいて、複数の第1の断片の2つ以上の核酸配列をフェーズブロックに連結することを更に含み、最大フェーズブロック長は少なくとも10 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも20 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも40 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも50 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも100 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも200 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも500 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも750 kbである。いくつかの場合では、最大フェーズブロック長は少なくとも1 Mbである。いくつかの場合では、最大フェーズブロック長は少なくとも1.75 Mbである。いくつかの場合では、最大フェーズブロック長は少なくとも2.5 Mbである。

【0011】

いくつかの場合では、1つ以上の変異を特定するための方法は、2つ以上の核酸配列の重複する核酸配列に基づいて、複数の第1の断片の2つ以上の核酸配列を推測コンティグに連結し、それにより推測コンティグの集団を作ること更に含み、推測コンティグの集団のN50は少なくとも10 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも20 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも40 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも50 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも100 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも200 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも500 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも750 kbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも1 Mbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも1.75 Mbである。いくつかの場合では、推測コンティグの集団のN50は少なくとも2.5 Mbである。

【0012】

いくつかの場合では、1つ以上の変異を特定するための方法は、複数の第1の断片の2つ以上の核酸配列内の重複するフェーズ化変異体に基づいて、複数の第1の断片の2つ以上の核酸配列をフェーズブロックに連結し、それによりフェーズブロックの集団を作ること更に含み、フェーズブロックの集団のN50は少なくとも10 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも20 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも40 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも50 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも100 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも200 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも500 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも750 kbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも1 Mbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも1.75 Mbである。いくつかの場合では、フェーズブロックの集団のN50は少なくとも2.5 Mbである。

【0013】

本開示の更なる態様は、核酸の構造的変異の存在を決定する方法を提供する。本方法は、(a)核酸の複数の第1の断片分子を提供することであって、複数の第1の断片分子の所与の第1の断片分子が構造的変異を含む、提供すること、(b)複数の第1の断片分子の各々の複数の第2の断片分子を配列決定して複数の断片配列を得ることであって、所与の第1の断片分子に対応する複数の断片配列の各々が共通のバーコード配列を共有する、

得ること、ならびに (c) (i) 複数の断片配列を参照配列にマッピングすること、(i i) 共通のバーコード配列を共有する複数の断片配列を特定すること、及び (i i i) 所与の第 1 の断片分子の長さよりも更に離れている位置で参照配列にマッピングする、共通のバーコード配列を共有する上昇した量の複数の断片配列の存在に基づき構造的変異を特定することであって、該上昇した量が構造的変異を欠如する配列に対してである、特定することにより、構造的変異の存在を決定することを含み得る。

【0014】

いくつかの場合では、上昇した量は、構造的変異を有する核酸の領域に由来する第 1 の断片分子の総数に対して 1 % 以上である。いくつかの場合では、上昇した量は、構造的変異を有する核酸の領域に由来する第 1 の断片分子の総数に対して 2 % 以上である。いくつかの場合では、位置は少なくとも約 100 塩基離れている。いくつかの場合では、位置は少なくとも約 500 塩基離れている。いくつかの場合では、位置は少なくとも約 1 キロベース (k b) 離れている。いくつかの場合では、位置は少なくとも約 10 k b 離れている。

10

【0015】

いくつかの場合では、核酸の構造的変異の存在を決定する方法は、複数の断片配列から所与の第 1 の断片分子のアセンブリを作ることにより構造的変異を特定することを更に含み、複数の断片配列は、共通のバーコード配列の存在に基づいてアセンブリの入力として選択される。いくつかの場合では、アセンブリは複数の断片配列からコンセンサス配列を生成することにより作られる。いくつかの場合では、構造的変異は転座を含む。

20

【0016】

本開示の更なる態様は、変異体核酸配列を特徴付ける方法を提供する。いくつかの場合では、本方法は、(a) 変異体核酸を断片化して 10 キロベース (k b) を超える長さを有する複数の第 1 の断片を得ること、(b) 複数の第 1 の断片を別個のパーティションに分離すること、(c) 各第 1 の断片からそのそれぞれのパーティション内に複数の第 2 の断片を作ることであって、複数の第 2 の断片がそれに付加されるバーコード配列を有し、所与のパーティション内のバーコード配列が共通のバーコード配列である、作ること、(d) 複数の第 2 の断片及びそれに付加されるバーコード配列を配列決定して複数の第 2 の断片の配列を得ること、(e) 少なくとも一部共通のバーコード配列の存在に基づき、第 2 の断片の配列を元の第 1 の断片に帰属させて、第 2 の断片の配列に関する第 1 の断片の配列環境 (sequence context) を得ること、及び (f) 第 1 の断片の配列環境から変異体核酸の変異体部分を特定し、それにより変異体核酸配列を特徴付けることを含み得る。いくつかの場合では、帰属は、少なくとも一部共通のバーコード配列の存在に基づき、複数の第 2 の断片の配列から複数の第 1 の断片からの個々の断片の配列の少なくとも一部をアセンブリングすることを含む。いくつかの場合では、帰属は、少なくとも部分的に共通のバーコード配列に基づき、複数の第 2 の断片の配列を複数の第 1 の断片からの個々の第 1 の断片にマッピングすることを含む。

30

【0017】

いくつかの場合では、変異体核酸配列を特徴付ける方法は、複数の第 1 の断片のうちの 2 つ以上の間の重複する配列に基づき、複数の第 1 の断片のうちの 2 つ以上を推測コンテ
ィグに連結することを更に含む。いくつかの場合では、特定は、第 1 の断片の配列環境から 1 つ以上のフェーズ化変異体を特定することを含む。いくつかの場合では、変異体核酸配列を特徴付けする方法は、複数の第 1 の断片のうちの 2 つ以上の間の重複するフェーズ化変異体に基づき、複数の第 1 の断片のうちの 2 つ以上をフェーズブロックに連結することを更に含む。いくつかの場合では、特定は、第 1 の断片の配列環境から 1 つ以上の構造的変異を特定することを含む。いくつかの場合では、1 つ以上の構造的変異は、挿入、欠失、転座、レトロトランスポゾン、反転、及び重複から独立して選択される。いくつかの場合では、構造的変異は挿入または転座を含み、第 1 の断片の配列環境は挿入または転座の存在を示す。

40

【0018】

50

本開示の更なる態様は、核酸の配列において変異体を特定する方法を提供する。いくつかの場合では、本方法は、核酸の複数の個々の断片分子の核酸配列を得ることであって、複数の個々の断片分子の核酸配列が各々、少なくとも1キロベース(kb)の長さを有する、得ることと、複数の個々の断片分子のうちの1つ以上の配列を1つ以上の推測コンティグに連結することと、1つ以上の推測コンティグから1つ以上の変異体を特定することとを含む。いくつかの場合では、得ることは、長さが10kbを超える複数の断片分子の核酸配列を得ることを含む。いくつかの場合では、得ることは、複数の個々の断片分子の各個々の断片分子の複数のバーコード化された断片を提供することであって、所与の個々の断片分子のバーコード化された断片が共通のバーコードを有する、提供することと、複数の個々の断片分子の複数のバーコード化された断片を配列決定することであって、該配列決定が1%未満の配列決定誤差率を提供する、配列決定することと、複数のバーコード化された断片の配列及びそれらの関連するバーコードから複数の個々の断片分子の配列を決定することと、を含む。

10

【0019】

いくつかの場合では、連結は、2つ以上の個々の断片分子間の1つ以上の重複する配列を特定して、2つ以上の個々の断片分子を1つ以上の推測コンティグに連結することを含む。いくつかの場合では、連結は、2つ以上の個々の断片分子間の1つ以上の共通の変異体を特定して、2つ以上の個々の断片分子を1つ以上の推測コンティグに連結することを含む。いくつかの場合では、1つ以上の共通の変異体はフェーズ化変異体であり、1つ以上の推測コンティグは少なくとも100kbの最大フェーズブロック長を含む。いくつかの場合では、特定において特定された1つ以上の変異体は構造的変異を含む。いくつかの場合では、構造的変異は、挿入、欠失、転座、レトロトランスポゾン、反転、及び重複から選択される。

20

【0020】

本開示の更なる態様は、核酸を特徴付ける方法を提供する。いくつかの場合では、本方法は、少なくとも10キロベース(kb)の長さを有する複数の断片分子の核酸配列を得ることと、複数の断片分子の核酸配列における1つ以上のフェーズ化変異体の位置を特定することと、少なくとも第1の断片分子の核酸配列を少なくとも第2の断片分子に、第1及び第2の断片分子内の1つ以上の共通するフェーズ化変異体の位置の存在に基づいて連結して、少なくとも10kbの最大フェーズブロック長を有するフェーズブロックを提供することと、少なくとも10kbの最大フェーズブロック長を有するフェーズブロックから1つ以上のフェーズ化変異体を特定することと、を含む。いくつかの場合では、核酸を特徴付ける方法は、フェーズブロックから1つ以上の追加のフェーズ化変異体を特定することを更に含む。いくつかの場合では、複数の断片分子は別個のパーティションにある。いくつかの場合では、別個のパーティションはエマルジョン中の液滴である。いくつかの場合では、複数の断片分子の長さは少なくとも50kbである。いくつかの場合では、複数の断片分子の長さは少なくとも100kbである。いくつかの場合では、最大フェーズブロック長は少なくとも50kbである。いくつかの場合では、最大フェーズブロック長は少なくとも100kbである。いくつかの場合では、最大フェーズブロック長は少なくとも1Mbである。いくつかの場合では、最大フェーズブロック長は少なくとも2Mbである。いくつかの場合では、最大フェーズブロック長は少なくとも2.5Mbである。

30

40

【0021】

本開示の更なる態様は、(a)生物の第1の染色体に由来する標的配列を含む、第1の核酸を第1のパーティションに分配すること、(b)生物の第2の染色体に由来する標的配列を含む、第2の核酸を第2のパーティションに分配すること、(c)第1のパーティションにおいて、第1のバーコード配列を第1の核酸の断片または第1の核酸の一部のコピーに付加して第1のバーコード化された断片を提供すること、(d)第2のパーティションにおいて、第1のバーコード配列とは異なる第2のバーコード配列を第2の核酸の断片または第2の核酸の一部のコピーに付加して、第2のバーコード化された断片を提供すること、(e)第1及び第2のバーコード化された断片の核酸配列を決定し、第1及び第

50

2の核酸の核酸配列をアセンブリングすること、及び(f)第1及び第2の核酸の核酸配列を比較して、それぞれ、第1及び第2の核酸を第1及び第2の染色体に由来すると特徴付けることを含む方法を提供する。いくつかの場合では、第1のバーコード配列を含むオリゴヌクレオチドは第1の核酸と共分配され、第2のバーコード配列を含むオリゴヌクレオチドは第2の核酸と共分配される。いくつかの場合では、第1のバーコード配列を含むオリゴヌクレオチドは第1のビーズに放出可能に付加され、第2のバーコード配列を含むオリゴヌクレオチドは第2のビーズに放出可能に付加され、共分配は、それぞれ、第1及び第2のビーズを第1及び第2のパーティションに共分配することを含む。いくつかの場合では、第1及び第2のパーティションはエマルジョン中の液滴を含む。いくつかの場合では、第1の染色体は父方染色体であり、第2の染色体は母方染色体である。いくつかの場合では、第1の染色体及び第2の染色体は相同染色体である。いくつかの場合では、第1の核酸及び第2の核酸は1つ以上の変異を含む。

10

【0022】

いくつかの場合では、第1及び第2の染色体は胎児に由来する。いくつかの場合では、第1及び第2の核酸は、妊婦から採取した試料から得られる。いくつかの場合では、第1の染色体は、染色体21、18、または13である。いくつかの場合では、第2の染色体は、染色体21、18、または13である。いくつかの場合では、本方法は、第1または第2の染色体の相対量を決定することを更に含む。いくつかの場合では、本方法は、参照染色体に対する第1または第2の染色体の量を決定することを更に含む。いくつかの場合では、第1の染色体もしくは第2の染色体またはその両方のコピー数は増加する。いくつかの場合では、コピー数の増加は癌または異数体の結果である。いくつかの場合では、第1の染色体もしくは第2の染色体またはその両方のコピー数は減少する。いくつかの場合では、コピー数の減少は癌または異数体の結果である。

20

【0023】

本開示の更なる態様は、(a)生物の第1の染色体に由来する標的配列を含む、第1の核酸を第1のパーティションに分配すること、(b)生物の第2の染色体に由来する標的配列を含む、第2の核酸を第2のパーティションに分配すること、(c)第1のパーティションにおいて、第1のバーコード配列を第1の核酸の断片または第1の核酸の一部のコピーに付加して第1のバーコード化された断片を提供すること、(d)第2のパーティションにおいて、第1のバーコード配列とは異なる第2のバーコード配列を第2の核酸の断片または第2の核酸の一部のコピーに付加して、第2のバーコード化された断片を提供すること、(e)第1及び第2のバーコード化された断片の核酸配列を決定し、第1及び第2の核酸の核酸配列をアセンブリングすること、及び(f)第1及び第2の核酸の核酸配列を比較して、第1及び第2の核酸の核酸配列間の任意の変異を特定することを含む方法を提供する。いくつかの場合では、第1のバーコード配列を含むオリゴヌクレオチドは第1の核酸と共分配され、第2のバーコード配列を含むオリゴヌクレオチドは第2の核酸と共分配される。いくつかの場合では、第1のバーコード配列を含むオリゴヌクレオチドは第1のビーズに放出可能に付加され、第2のバーコード配列を含むオリゴヌクレオチドは第2のビーズに放出可能に付加され、共分配は、それぞれ、第1及び第2のビーズを第1及び第2のパーティションに共分配することを含む。いくつかの場合では、第1及び第2のパーティションはエマルジョン中の液滴を含む。いくつかの場合では、第1の染色体は父方染色体であり、第2の染色体は母方染色体である。いくつかの場合では、第1の染色体及び第2の染色体は相同染色体である。いくつかの場合では、第1の核酸及び第2の核酸は1つ以上の変異を含む。いくつかの場合では、第1及び第2の染色体は胎児に由来する。いくつかの場合では、第1及び第2の核酸は、妊婦から採取した試料から得られる。いくつかの場合では、第1の染色体は、染色体21、18、または13である。いくつかの場合では、第2の染色体は、染色体21、18、または13である。いくつかの場合では、本方法は、第1または第2の染色体の相対量を決定することを更に含む。いくつかの場合では、本方法は、参照染色体に対する第1または第2の染色体の量を決定することを更に含む。いくつかの場合では、第1の染色体もしくは第2の染色体またはその両方のコ

30

40

50

ピー数は増加する。いくつかの場合では、コピー数の増加は癌または異数体の結果である。いくつかの場合では、第1の染色体もしくは第2の染色体またはその両方のコピー数は減少する。いくつかの場合では、コピー数の減少は癌または異数体の結果である。

【0024】

本開示の更なる態様は、胎児核酸配列を特徴付けるための方法を提供する。いくつかの場合では、本方法は、(a) 胎児の妊娠中の母親から得られる、母方核酸の配列を、(i) 母方核酸を断片化して複数の第1の母方断片を提供すること、(ii) 複数の第1の母方断片を母方パーティションに分離すること、(iii) 第1の母方断片の各々からそれぞれのそれぞれの母方パーティション内に複数の第2の母方断片を作ることであって、該複数の第2の母方断片がそれに付加される第1のバーコード配列を有し、母方パーティションの所与の母方パーティション内において、第2の母方断片がそれに付加される第1の共通のバーコード配列を含む、作ること、(iv) 複数の第2の母方断片を配列決定して複数の母方断片配列を提供すること、(v) 少なくとも部分的に第1の共通のバーコード配列の存在に基づき母方断片配列を元の第1の母方断片に帰属させて母方核酸配列を決定することにより、決定すること、(b) 胎児の父親から得られる、父方核酸の配列を、(i) 父方核酸を断片化して複数の第1の父方断片を提供すること、(ii) 複数の第1の父方断片を父方の別個のパーティションに分離すること、(iii) 各第1の父方断片からそのそれぞれのパーティション内に複数の第2の父方断片を作ることであって、該複数の第2の父方断片がそれに付加される第2のバーコード配列を有し、所与の父方パーティション内において、第2の父方断片がそれに付加される第2の共通のバーコード配列を含む、作ること、(iv) 複数の第2の父方断片及びそれに付加される第2のバーコード配列を配列決定して、複数の父方断片配列を提供すること、(v) 少なくとも部分的に第2の共通のバーコード配列の存在に基づき、父方断片配列を元の第1の父方断片に帰属させて父方核酸配列を決定することにより、決定すること、(c) 妊娠中の母親から胎児核酸を得、母方核酸配列及び父方核酸配列を使用して、胎児核酸の配列及び／または胎児核酸の配列の1つ以上の遺伝的変異を決定することを含む。

【0025】

いくつかの場合では、父方断片配列及び母方断片配列は各々、配列を1つ以上の推測コンティグに連結するために使用される。いくつかの場合では、推測コンティグは母方及び父方のフェーズブロックを構築するために使用される。いくつかの場合では、胎児核酸の配列は、胎児フェーズブロックを構築するために母方及び父方のフェーズブロックと比較される。いくつかの場合では、父方断片配列は複数の第1の父方断片の配列の少なくとも一部を生成するためにアセンブリングされ、それにより父方核酸配列を決定し、母方断片配列は複数の第1の母方断片の配列の少なくとも一部を生成するためにアセンブリングされ、それにより母方核酸配列を決定する。いくつかの場合では、父方核酸配列の決定は父方断片配列を父方参照にマッピングすることを含み、母方核酸配列の決定は母方断片配列を母方参照にマッピングすることを含む。

【0026】

いくつかの場合では、胎児核酸の配列は少なくとも99%の精度で決定される。いくつかの場合では、胎児核酸の配列の1つ以上の遺伝的変異は少なくとも99%の精度で決定される。いくつかの場合では、1つ以上の遺伝的変異は、構造的変異及び一塩基多型(SNP)からなる群から選択される。いくつかの場合では、1つ以上の遺伝的変異は、コピー数変異、挿入、欠失、転座、レトロトランスポゾン、反転、再編成、反復配列伸長、及び重複からなる群から選択される構造的変異である。

【0027】

いくつかの場合では、胎児核酸配列を特徴付けるための方法は、(c)において、母方核酸配列及び父方核酸配列について決定された1つ以上の遺伝的変異を使用して、胎児核酸の配列の1つ以上の遺伝的変異を決定することを更に含む。いくつかの場合では、胎児核酸配列を特徴付けるための方法は、(c)において、胎児核酸の1つ以上の新規突然変異を決定することを更に含む。いくつかの場合では、胎児核酸配列を特徴付けるための方

法は、(c)の間または後に、胎児核酸に関連する異数体を決定することを更に含む。

【0028】

いくつかの場合では、胎児核酸配列を特徴付けるための方法は、(a)の(v)の間または後に、母方核酸配列をハプロタイプ分析(haplotyping)してハプロタイプが解明された母方核酸配列を提供することと、(b)の(v)の間または後に、父方核酸配列をハプロタイプ分析してハプロタイプが解明された父方核酸配列を提供することと、を更に含む。いくつかの場合では、胎児核酸配列を特徴付けるための方法は、(c)において、ハプロタイプが解明された母方核酸配列及びハプロタイプが解明された父方核酸配列を使用して、胎児核酸の配列及び／または1つ以上の遺伝的変異を決定することを更に含む。いくつかの場合では、母方核酸及び父方核酸のうちの1つ以上はゲノムデオキシリボ核酸(DNA)である。いくつかの場合では、(c)において、胎児核酸は無細胞核酸を含む。いくつかの場合では、胎児核酸配列を特徴付けるための方法は、(a)において、少なくとも99%の精度で母方核酸配列を決定することを更に含む。いくつかの場合では、胎児核酸配列を特徴付けるための方法は、(b)において、少なくとも99%の精度で父方核酸配列を決定することを更に含む。

10

【0029】

いくつかの場合では、母方核酸配列及び／または父方核酸配列は10キロベース(kb)を超える長さを有する。いくつかの場合では、母方及び父方のパーティションはエマルジョン中の液滴を含む。いくつかの場合では、(a)において、第1のバーコード配列は、第1の粒子に放出可能に付加された所与の母方パーティションにおいて提供される。いくつかの場合では、(b)において、第2のバーコード配列は、第2の粒子に放出可能に付加された所与の父方パーティションにおいて提供される。

20

【0030】

本開示の更なる態様は、試料核酸を特徴付けるための方法を提供する。いくつかの場合では、本方法は、(a)無細胞試料核酸を含む、生物学的試料を対象から得ること、(b)液滴において、バーコード配列を無細胞試料核酸の断片または試料核酸の一部のコピーに付加してバーコード化された試料断片を提供する、(c)バーコード化された試料断片の核酸配列を決定し、バーコード化された試料断片の核酸配列に基づき、試料核酸配列を提供すること、(d)プログラムされたコンピュータプロセッサを使用して試料核酸配列の参照核酸配列への比較を生成することであって、該参照核酸配列が10キロベース(kb)を超える長さ及び少なくとも99%の精度を有する、生成すること、及び(e)該比較を使用して試料核酸配列における1つ以上の遺伝的変異を特定し、それにより試料核酸を疾患と関連付けることを含む。いくつかの場合では、試料核酸配列における1つ以上の遺伝的変異は、構造的変異及び一塩基多型(SNP)からなる群から選択される。いくつかの場合では、試料核酸配列の1つ以上の遺伝的変異は、コピー数変異、挿入、欠失、レトロトランスポゾン、転座、反転、再編成、反復配列伸長、及び重複からなる群から選択される構造的変異である。いくつかの場合では、(c)において、試料核酸配列は少なくとも99%の精度で提供される。いくつかの場合では、(b)において、バーコード配列は粒子に放出可能に付加された液滴に提供され、(b)は、バーコード配列を付加する前に、粒子から液滴中にバーコード配列を放出することを更に含む。いくつかの場合では、(b)において、バーコード配列は、粒子に放出可能に付加されたプライマー配列の一部として提供され、該プライマー配列はランダムN-merも含み、(b)は、バーコード配列を付加する前に、粒子から液滴中にプライマー配列を放出することを更に含む。いくつかの場合では、(b)において、プライマーを使用して、増幅反応において、バーコード配列を無細胞試料核酸の断片または無細胞試料核酸の一部のコピーに付加する。

30

40

【0031】

いくつかの場合では、試料核酸を特徴付けるための方法は、(i)追加の液滴において、追加のバーコード配列を参照核酸の断片または参照核酸の一部のコピーに付加してバーコード化された参照断片を提供すること、及び(ii)バーコード化された参照断片の核酸配列を決定し、バーコード化された参照断片の核酸配列に基づき参照核酸配列を決定す

50

ることを更に含む。いくつかの場合では、参照核酸配列の決定は、バーコード化された参照断片の核酸配列をアセンブリングすることを含む。いくつかの場合では、試料核酸を特徴付けるための方法は、粒子に放出可能に付加された追加の液滴に追加のバーコード配列を提供することと、追加のバーコード配列を付加する前に、粒子から追加のパーティション中に追加のバーコード配列を放出することを更に含む。いくつかの場合では、試料核酸を特徴付けるための方法は、粒子に放出可能に付加されたプライマー配列の一部として追加のバーコード配列を提供することであって、該プライマー配列がランダム N-mer 配列も含む、提供することと、追加のバーコード配列を付加する前に、粒子から追加の液滴中にプライマーを放出することとを更に含む。いくつかの場合では、試料核酸を特徴付けるための方法は、プライマーを使用して、増幅反応において、追加のバーコード配列を参照核酸の断片または参照核酸の一部のコピーに付加することを更に含む。いくつかの場合では、試料核酸を特徴付けるための方法は、参照核酸配列において 1 つ以上の遺伝的変異を決定することを更に含む。

【0032】

いくつかの場合では、参照核酸配列における 1 つ以上の遺伝的変異は、構造的変異及び一塩基多型 (SNP) からなる群から選択される。いくつかの場合では、参照核酸配列における 1 つ以上の遺伝的変異は、コピー数変異、挿入、欠失、レトロトランスポゾン、転座、反転、再編成、反復配列伸長、及び重複からなる群から選択される構造的変異である。いくつかの場合では、参照核酸は生殖系列核酸配列を含む。いくつかの場合では、参照核酸は癌核酸配列を含む。いくつかの場合では、試料核酸配列は 10 kb を超える長さを有する。いくつかの場合では、参照核酸は疾患状態の不在を示すゲノムに由来する。いくつかの場合では、参照核酸は疾患状態を示すゲノムに由来する。いくつかの場合では、疾患状態は癌を含む。いくつかの場合では、疾患状態は異数体を含む。いくつかの場合では、無細胞試料核酸は腫瘍核酸を含む。いくつかの場合では、腫瘍核酸は循環腫瘍核酸を含む。

【0033】

本開示の更なる態様及び利点は、本開示の例示的な実施形態のみが示され、説明される以下の発明を実施するための形態から当業者には容易に明らかとなろう。理解されるように、本開示は他の及び異なる実施形態が可能であり、そのいくつかの詳細は、すべて本開示から逸脱することなく、様々な明らかな観点において修正が可能である。したがって、図面及び説明は性質上図示的であり、制限するものとみなされないものとする。

【0034】

参照による組み込み

本明細書に言及されるすべての刊行物、特許、及び特許出願は、各個々の刊行物、特許、または特許出願が参照により組み込まれるように具体的かつ個々に示されるのと同程度まで、参照によりその全体が本明細書に組み込まれる。

【0035】

本発明の新規特徴は、付属の特許請求の範囲に詳細に記述される。本発明の特徴及び利点のより良い理解は、本発明の原理が利用される図示的な実施形態を記述する以下の発明を実施するための形態及び添付の図面を参照することにより得られるだろう。

【図面の簡単な説明】

【0036】

【図 1】従来のプロセス对本明細書に記載される例示的プロセス及び系を用いたフェーズ化変異体の特定及び分析の模式図を提供する。

【図 2】従来のプロセス对本明細書に記載される例示的プロセス及び系を用いた構造的変異の特定及び分析の模式図を提供する。

【図 3】本明細書に開示される方法及び組成物を用いたコピー数またはハプロタイプを検出するためのアッセイを実施するための例示的なワークフローを図示する。

【図 4】核酸試料をビーズと組み合わせ、核酸とビーズを別個の液滴に分配するための例示的なプロセスの模式図を提供する。

【図 5】染色体核酸断片をバーコード化及び増幅するための例示的なプロセスの模式図を提供する。

【図 6】配列データを個々の染色体に帰属させる際の染色体核酸断片のバーコード化の例示的な使用の模式図を提供する。

【図 7】フェーズ化配列決定プロセスの例の模式図を提供する。

【図 8】健康な患者（上のパネル）及びハプロタイプコピー数が増加した（中央のパネル）またはハプロタイプコピー数の減少（下のパネル）を有する癌患者のゲノムの例示的なサブセットの模式図を提供する。

【図 9 A】図 9 A 及び B は、（a）腫瘍 DNA の相対的寄与を示す模式図、及び（b）通常の配列決定法によるこのようなコピーの増加及び減少を検出する表示を提供する。

【図 9 B】図 9 A 及び B は、（a）腫瘍 DNA の相対的寄与を示す模式図、及び（b）通常の配列決定法によるこのようなコピーの増加及び減少を検出する表示を提供する。

【図 10】単一変異体の位置（左のパネル）及び組み合わせた変異体の位置（右のパネル）を用いたコピーの増加及び減少を検出する例の模式図を提供する。

【図 11】コピー数の増加及び減少を特定するための記載される方法及び系の可能性の模式図を提供する。

【図 12】本明細書に記載される方法及び組成物を用いた染色体数及びコピー数変異の決定に基づく異数体試験を実施するための例示的なワークフローを図示する。

【図 13 A】図 13 A 及び B は、遺伝子試料における転座及び遺伝子融合などの構造的変異を特定するためのプロセスの例示的な概要を図示する。

【図 13 B】図 13 A 及び B は、遺伝子試料における転座及び遺伝子融合などの構造的変異を特定するためのプロセスの例示的な概要を図示する。

【図 14】本明細書に記載される方法及び組成物を用いたコピー数変異の決定に基づく癌診断試験を実施するための例示的なワークフローを図示する。

【図 15】NCI-H2228 癌細胞系からの EML-4-ALK 構造的変異の模式図を提供する。

【図 16 A】図 16 A 及び 16 B は、陰性対照細胞系（図 16 B）と比較したとき、癌細胞系（図 16 A）において、図 15 に示される EML-4-ALK 変異体構造の存在を特定するための本明細書に記載される系を用いたバーコードマッピングデータを提供する。

【図 16 B】図 16 A 及び 16 B は、陰性対照細胞系（図 16 B）と比較したとき、癌細胞系（図 16 A）において、図 15 に示される EML-4-ALK 変異体構造の存在を特定するための本明細書に記載される系を用いたバーコードマッピングデータを提供する。

【図 17】本明細書に記載されるように父方核酸配列を分析する例示的なワークフローを模式的に示す。

【図 18】本明細書に記載されるように母方核酸配列を分析する例示的なワークフローを模式的に示す。

【図 19】本明細書に記載されるように胎児核酸配列を分析する例示的なワークフローを模式的に示す。

【図 20】本明細書に記載されるように参照核酸配列を分析する例示的なワークフローを模式的に示す。

【図 21】本明細書に記載されるように試料核酸配列を分析する例示的なワークフローを模式的に示す。

【図 22】例示的なコンピュータ制御システムを模式的に示す。

【発明を実施するための形態】

【0037】

本発明の様々な実施形態が本明細書において示され、説明されるが、このような実施形態が単に例として提供されることは当業者には明らかであろう。当業者は、多くの変形、変更、及び置換を、本発明から逸脱することなく想到し得る。本明細書に記載される本発明の実施形態に対する様々な代替が利用され得ることを理解するべきである。

【0038】

本明細書で使用される、「生物」という用語は、概して、継続する生体系を指す。生物の非限定的な例としては、動物（例えば、ヒト、他の種類の哺乳類、鳥、爬虫類、昆虫、本明細書の他の箇所に記載される他の例示的な種類の動物）、植物、菌類、及び細菌が挙げられる。

【0039】

本明細書で使用される、「コンティグ」という用語は、概して、所与の長さの連続する核酸配列を指す。連続する配列は、短いもしくは長い配列リードのいずれかの配列リードを含む個々の配列リード、またはリード内の重複する配列に基づき整合及びアセンブリングされる、もしくは他の既知の連鎖データ、例えば本明細書の他の箇所に記載される共通のバーコードによるタグ付けに基づき断片内で連結されると定義される配列リードのアセンブリから得ることができる。これらの重複する配列リードは、同様に、短いリード、例えば500塩基未満、例えば、いくつかの場合では約100～500塩基、及びいくつかの場合では100～250塩基を含むか、またはより長い配列リード、例えば500塩基超、1000塩基超、または更には10,000塩基超に基づく。

【0040】

I. 概要

本開示は、遺伝物質の特徴付けに大きな進歩をもたらすのに有用な方法及び系を提供する。いくつかの場合では、本方法及び系は、例えば、ハプロタイプフェージング、構造的変異、例えば、欠失、重複、コピー数変異体、挿入、反転、レトロトランスポゾン、転座、LTR、STR、及び様々な他の有用な特徴付けの特定を含む、一般的に利用可能な技術の使用が非常に困難であるか、または更には不可能である遺伝的特徴付けを提供するのに有用であり得る。

【0041】

一般に、本明細書に記載される方法及び系は、長い範囲の変異体情報の特定及び使用（例えば、他の変異を含有する配列セグメントを含む、異なる配列セグメントであって、発生元の配列においてかなりの距離、例えば短いリード配列決定技術によって提供されるよりも長い距離によって分離される、配列セグメントに変異を関連付ける）を可能にする長い個々の核酸分子の配列決定を提供することにより上記の目標を達成する。しかしながら、これらの方法及び系は、短いリード配列決定技術の非常に低い配列決定誤差率の利点を伴って、及び報告された長いリード長配列決定技術、例えばSMRT配列決定及びナノ細孔配列決定技術などの単一分子配列決定のものよりもはるかに下回ってこれらの目的を達成する。

【0042】

一般に、本明細書に記載される方法及び系は、高処理高精度の短いリード配列決定技術を使用して配列可能（sequenceable）であるより小さい断片に長い核酸分子を分割するが、より小さい断片から得られる配列情報を発生元のより長い個々の核酸分子に帰属させることを可能にする方法でこのような分割を行う。配列リードを発生元のより長い核酸分子に帰属させることにより、一般的に短い配列リードだけでは得ることができない、そのより長い核酸配列の有意な特徴付け情報を得ることができる。述べたように、このような特徴付け情報は、ハプロタイプフェージング、構造的変異の特定、及びコピー数変異の特定を含み得る。

【0043】

本明細書に記載される方法及び系の利点は、いくつかの一般的な例に対して記載される。第1の例では、フェーズ化配列変異体は、本明細書に記載される方法及び系を使用して特定及び特徴付けされる。図1は、本明細書に記載される方法により提示されるフェーズ化変異体呼び出しの課題及び解決策を模式的に図示する。示されるように、パネルIの核酸102及び104は、異なる染色体、例えば母方及び父方由来の染色体の同じ領域の2つの一倍体配列を表す。各配列は、各一倍体配列を特徴付ける異なる対立遺伝子で、一連の変異体、例えば核酸102上に変異体106～114及び核酸104上に変異体116～122を含む。それらの非常に短い配列リードのため、大半の配列決定技術は、同じ一

倍体配列上の他の変異体に対して個々の変異体の環境を提供することができない。加えて、それらは個々の分子成分、例えば各一倍体配列を分離しない試料調製技術に依存するため、様々な変異体、例えば変異体が由来する一倍体配列のフェージングを特定することができない。結果として、これらの短いリード技術は、これらの変異体をそれらの発生元の分子に解明することができない。このアプローチによる問題は、パネル I I a 及び I I I a に模式的に図示される。簡潔に、パネル I I a に示される両方の一倍体配列からのプールされた断片を配列決定して多数の短い配列リード 1 2 4 を得、得られた配列 1 2 6 がアセンブリされる（パネル I I I a に示される）。示されるように、パネル I I a においてより短い配列リードのいずれの相対的なフェージング環境も有しないため、アセンブリプロセスにおいて 2 つの異なる一倍体配列との関係では変異体を解明することができないだろう。したがって、パネル I I I a に示される得られたアセンブリは、変異体 1 0 6 ~ 1 2 2 のすべてを含む単一のコンセンサス配列アセンブリ 1 2 6 をもたらす。

【0044】

対照的に、及び図 1 のパネル I I b に示されるように、本明細書に記載される方法及び系は、上述のアプローチのように、より長い核酸 1 0 2 及び 1 0 4 をより短い配列可能な断片に分解または分割するが、これらの断片を用いて、それらをそれらの発生元の分子環境に帰属させる能力を維持する。これは、異なる断片がそれらの発生元の分子環境により分類または「区画化」されるパネル I I b に模式的に図示される。本開示の文脈において、この分類は、分子環境を維持する群に断片を物理的に分配すること、ならびにその環境を後に解明することを可能にするためにこれらの断片をタグ付けすることのうちの 1 つまたはその両方を通して達成され得る。

【0045】

この分類は、それぞれ、核酸 1 0 2 及び 1 0 4 からの短い配列リードを表す群 1 2 8 及び 1 3 0 との関係ではより短い配列リードの割り付けとして模式的に図示される。発生元の配列環境は配列決定プロセスを通して維持されるため、それぞれ、配列 1 0 2 及び 1 0 4 との関係では様々な変異体 1 0 6 ~ 1 1 4 及び 1 1 6 ~ 1 2 2 の発生元の分子環境の解明、例えばフェージングの際に、その環境を利用することができる。

【0046】

別の例示的な利点のある用途では、本方法及び系は、短いリード配列技術を使用して一般的に識別不可能である、または識別が少なくとも困難である構造変異体の特徴付けに有用である。

【0047】

これは、図 2 において単純な転座事象に関して模式的に図示される。示されるように、ゲノム試料は、転座事象、例えば、配列 2 0 2 から配列 2 0 4 への遺伝要素 2 0 6 の転座を含む核酸を含む。このような転座は、例えば同じまたは異なる領域への異なる染色体間、同じ染色体の異なる領域間での転座を含む、様々な異なる転座の種類のいずれかである。

【0048】

再び、上記の図 1 に図示される例のように、従来の配列決定は、パネル I の配列 2 0 2 及び 2 0 4 を小さい断片に分解し、パネル I I a に示されるように、これらの断片から短い配列リード 2 0 8 を生成することにより開始する。これらの配列断片 2 0 8 は比較的に短いため、転座した配列 2 0 6 の環境は、すなわち、同じまたは異なる配列上の変異体の位置に由来するため、アセンブリプロセス中に簡単に失われる。更に、それらの短いリード長のため、配列アセンブリは、その定義からして、構造的変異を反映しないであろう参照配列の使用を前提とすることが多い。そのため、短い配列リード 2 0 8 は、必然的に、転座した配列 2 0 6 の適切な位置を無視してアセンブリされ、代わりに、パネル I I I a に示されるように、非変異体配列 2 1 0 及び 2 1 2 がアセンブリされるだろう。

【0049】

対照的に、本明細書に記載される方法及び系を使用することにより、配列 2 0 2 及び 2

04から得られる短い配列リードは、より小さい配列断片の元の分子分類を維持する群214及び216としてパネルI I bに示されるように区画化により提供され、パネルI I bに示される配列218及び220としてのそれらのアセンブリを可能にし、帰属を発生元の配列202及び204に戻し、パネルI I I bに図示されるように、正しい配列アセンブリ218及び220において、転座変異、例えば転座した配列断片206aを特定することを可能にする。

【0050】

上述のように、本明細書に記載される方法及び系は、より長い核酸の短い配列リードに対する個々の分子環境を提供する。本明細書で使用されるとき、個々の分子環境は、配列リード自体内に含まれず、そのため、一般的に、それらが短い配列リード、例えば約150塩基、または対のリードの約300塩基のリード全体またはその一部に含まれないような特定の配列リードを超えた配列環境、例えば、隣接または近接する配列に対する関係を指す。いくつかの態様では、本方法及び系は、短い配列リードのために長い範囲の配列環境を提供する。そのような長い範囲の環境は、互いの距離が1キロベース(kb)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、もしくは更には100kb以上長い範囲内である配列リードに対する所与の配列リードの関係または連結を含む。より長い範囲の個々の分子環境を提供することにより、本明細書に記載される方法及び系は、非常に長い推測分子環境も提供する。本明細書に記載されるように、配列環境は、例えば、短い配列リードを個々のより長い分子または連結した分子のコンティグにマッピングすることからの低分析能の環境、ならびに例えば、個々の分子の連続する決定された配列を有する(このような決定された配列は、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、または更には100kbより長い)、例えば、より長い個々の分子の大部分の長い範囲の配列決定からの高分析能の配列環境を含み得る。配列環境のように、短い配列の、より長い核酸、例えば、個々の長い核酸分子または連結された核酸分子もしくはコンティグの集合体の両方への帰属は、高レベルの配列環境を提供するためにより長い核酸伸長に対して短い配列をマッピングすること、ならびにこれらのより長い核酸を通して短い配列からアセンブリされた配列を提供することの両方を含み得る。

【0051】

更に、長い個々の分子に関連する長い範囲の配列環境を利用することができるが、このような長い範囲の配列環境を有することにより、更に長い範囲の配列環境を推測することも可能にする。一例として、上述の長い範囲の分子環境を提供することにより、異なる発生元の分子からの長い配列の中で重複する変異体部分、例えば、フェーズ化変異体、転座された配列などを特定することができ、これらの分子間の推測される連結を可能にする。このような推測される連結または分子環境は、本明細書において「推測コンティグ」と称される。いくつかの場合では、フェーズ化配列の文脈において論じられる場合、推測コンティグは、一般にフェーズ化配列を表す場合があり、例えば重複するフェーズ化変異体に基づく場合、個々の発生元の分子よりも実質的に長い長さのフェーズ化コンティグを推測することができる。これらのフェーズ化コンティグは、本明細書において「フェーズブロック」と称される。

【0052】

より長い単一分子リードを用いて開始することにより、短いリード配列決定技術またはフェーズ化配列決定に対する他のアプローチを使用してさもなければ達成可能であろうよりも長い推測コンティグまたはフェーズブロックを導くことができる。例えば、公表された米国特許公開第2013/0157870号(その全開示は参照によりその全体が本明細書に組み込まれる)を参照。特に、本明細書に記載される方法及び系を使用することにより、少なくとも約10キロベース(kb)、少なくとも約20kb、少なくとも約50

k b の N 5 0 (その長さ以上のすべてのフェーズブロックまたはコンティグの集合体がすべてのコンティグまたはフェーズブロックの長さの合計の少なくとも半分を含有し、その長さ以下のすべてのコンティグまたはフェーズブロックの集合体がすべてのコンティグまたはフェーズブロックの長さの合計の少なくとも半分も含有するコンティグまたはフェーズブロック長)、モード、平均、または中央値を有する推測コンティグまたはフェーズブロック長を得ることができる。いくつかの態様では、推測コンティグまたはフェーズブロック長は、少なくとも約 1 0 0 k b、少なくとも約 1 5 0 k b、少なくとも約 2 0 0 k b の N 5 0、モード、平均、または中央値を有し、いくつかの場合では、少なくとも約 2 5 0 k b、少なくとも約 3 0 0 k b、少なくとも約 3 5 0 k b、少なくとも約 4 0 0 k b、いくつかの場合では、少なくとも約 5 0 0 k b、少なくとも約 7 5 0 k b、少なくとも約 1 M b、少なくとも約 1. 7 5 M b、少なくとも約 2. 5 M b 以上が達成される。また更に他の場合では、少なくとも約 2 0 k b、4 0 k b、5 0 k b、1 0 0 k b、2 0 0 k b、3 0 0 k b、4 0 0 k b、5 0 0 k b、7 5 0 k b、1 メガベース (M b)、1. 7 5 M b、2 M b、もしくは 2. 5 M b、またはそれらを超えた最大推測コンティグまたはフェーズブロック長を得ることができる。また更に他の場合では、推測コンティグまたはフェーズブロック長は、少なくとも約 2 0 k b、少なくとも約 4 0 k b、少なくとも約 5 0 k b、少なくとも約 1 0 0 k b、少なくとも約 2 0 0 k b、いくつかの場合では、少なくとも約 5 0 0 k b、少なくとも約 7 5 0 k b、少なくとも約 1 M b、いくつかの場合では、少なくとも約 1. 7 5 M b、少なくとも約 2. 5 M b 以上であってよい。

【0053】

一態様では、本明細書に記載される方法及び系は、試料核酸またはその断片の、別個のコンパートメントまたはパーティション (本明細書において互換的にパーティションと称される) への区画化、堆積、または分配を提供し、各パーティションはその自身の内容物を他のパーティションの内容物から分離した状態を維持する。固有の識別子、例えばバーコードは、特徴、例えば核酸配列情報を、特定のコンパートメント内に含まれる試料核酸、具体的にはパーティション内に最初に堆積され得る連続する試料核酸の比較的長い伸長に後に帰属させることを可能にするために、区画化または分配した試料核酸を保持するパーティションに事前に、後に、または同時に送達され得る。

【0054】

試料核酸は、連続する核酸分子の比較的長い断片または伸長で核酸がパーティションに存在するように分配され得る。これらの断片は、分析される全体的な試料核酸、例えば染色体全体、エクソーム、または他の大きなゲノム断片のいくつかの重複する断片を表し得る。これらの試料核酸は、全ゲノム、個々の染色体、エクソーム、アンブリコン、または様々な異なる関心の核酸のいずれかを含み得る。いくつかの場合では、試料核酸のこれらの断片は、1 0 0 塩基よりも長い、5 0 0 塩基より長い、1 k b より長い、5 k b より長い、1 0 k b より長い、1 5 k b より長い、2 0 k b より長い、3 0 k b より長い、4 0 k b より長い、5 0 k b より長い、6 0 k b より長い、7 0 k b より長い、8 0 k b より長い、9 0 k b より長い、または更には 1 0 0 k b より長くてよく、これは上述のより長い範囲の分子環境を可能にする。

【0055】

試料核酸はまた、所与のパーティションが開始試料核酸の 2 つの重複する断片を含む確率が非常に低いレベルで分配され得る。これは、分配プロセス中、低入力量及び／または濃度で試料核酸を提供することにより達成され得る。結果として、いくつかの場合では、所与のパーティションは、いくつかの長い重複しない開始試料核酸の断片を含み得る。次に、異なるパーティション中の試料核酸は、固有の識別子と関連し、任意の所与のパーティションに関して、その中に含有される核酸は同じ固有の識別子を有するが、異なるパーティションは異なる固有の識別子を含み得る。更に、分配は、試料成分を非常に少容積のパーティションまたは液滴に割り付けるため、上述のように割り付けを達成するために、高容積プロセス、例えばチューブまたは多重ウェルプレートのウェルに必要とされ得るように、試料の実質的な希釈を行う必要がないことを理解する。更に、本明細書に記載さ

れる系は、このような高レベルのバーコード多様性を利用するため、上記に提供されるように、多数のゲノム等価物の中で多様なバーコードを割り付けることができる。特に、前述のように、多重ウェルプレートアプローチ（例えば、米国特許公開第2013/0079231号及び第2013/0157870号を参照、それらの全開示は参照によりそれらの全体が本明細書に組み込まれる）は、百～数百の異なるバーコード配列を用いて行うのみである可能性があり、バーコードを異なる細胞／核酸に帰属させることを可能にするために、それらの試料の限界希釈プロセスを利用する。そのため、それらは一般的に、約1:10、及び確実に1:100をゆうに上回るゲノム：（バーコードの種類）の比率を提供し得る100細胞よりも大幅に少ない細胞を用いて行う。一方、本明細書に記載される系は、高レベルのバーコード多様性、例えば10,000、100,000、500,000などを超える多様なバーコードの種類のため、約1:50以下、1:100以下、1:1000以下、または更には小さい比率のゲノム：（バーコードの種類）で行うことができ、同時にゲノム当たり大幅に改善されたバーコードの多様性を尚も提供しながら多数のゲノム（例えば、アッセイ当たり約100ゲノムを超える、アッセイ当たり500ゲノムを超える、アッセイ当たり1000ゲノムを超える、または更にはそれ以上）の充填を可能にする。

10

【0056】

多くの場合、試料は、分配の前にビーズに放出可能に付加される一組のオリゴヌクレオチドタグと組み合わせられる。オリゴヌクレオチドは少なくとも第1及び第2の領域を含み得る。第1の領域は、所与のパーティション内のオリゴヌクレオチドとの関係では、実質的に同じバーコード配列であり得るが、異なるパーティションとの関係では、異なるバーコード配列であり得、大半の場合、そうである。第2の領域は、パーティション内の試料内の核酸をプライムするために使用され得るN-mer（例えば、ランダムN-merまたは特定の配列を標的とするように設計されたN-merのいずれか）であり得る。いくつかの場合では、N-merが特定の配列を標的とするように設計される場合、特定の染色体（例えば、染色体1、13、18、または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領域を標的とするように設計され得る。いくつかの場合では、N-merは、疾患もしくは障害（例えば、癌）に関連する遺伝子または領域など、特定の遺伝子または遺伝領域を標的とするように設計され得る。パーティション内で、増幅反応は、核酸の長さに沿った異なる場所で核酸試料をプライムするために第2のN-merを用いて行われ得る。増幅の結果として、各パーティションは、同一またはほぼ同一のバーコードに付加され、各パーティションにおける核酸の重複するより小さい断片を表し得る核酸の増幅された生成物を含有し得る。バーコードは、一組の核酸が同じパーティションに由来し、よって、同じ核酸の鎖に由来する可能性もあることを表すマーカーとして機能し得る。増幅後、核酸は、プールされ、配列決定され、配列決定アルゴリズムを使用して整合される。より短い配列リードは、それらの関連するバーコード配列により、試料核酸の単一の長い断片と整合され、それに帰属され得るため、その配列上の特定された変異体のすべては、単一の発生源の断片及び単一の発生源の染色体に帰属され得る。更に、複数の長い断片にわたって複数の共存変異体を整合することにより、その染色体寄与を更に特徴付けることができる。したがって、特定の遺伝子変異体のフェージングに関する結論がその後導かれ得る。このような情報は、一般的に、同じ核酸鎖または異なる核酸鎖上に存在する特定の組の遺伝子変異体であるハプロタイプの特定に有用であり得る。コピー数変異もこの方法で特定され得る。

20

30

40

【0057】

記載される方法及び系は、現在の核酸配列決定技術及びそれらの関連する試料調製方法に対して大きな利点を提供する。ハプロタイプフェージング及びコピー数変異データは一般的に、細胞の集団から遺伝物質を抽出するために、生物学的試料（例えば、血液、細胞、または組織試料）がまとめて処理され、それを所与の配列決定技術用に特別に構成される配列決定ライブラリに変換するため、ゲノムDNAを配列決定することにより入手できない。この集団試料処理アプローチの結果として、配列決定データは一般的に、遺伝的情

50

報が同じまたは異なる染色体上に存在するかを決定することが不可能であるフェーズ化されない遺伝子型をもたらす。

【0058】

遺伝的特徴を特定の染色体に帰属させることが不可能なことに加え、このような集団試料調製及び配列決定方法は、試料中の大部分の構成成分を主に特定及び特徴付けする傾向にもあり、少数の構成成分、例えば、抽出した試料中の全DNAの数パーセントを構成する1つの染色体によって、または1つもしくは2～3つの細胞によって、または血流中に循環する断片化された腫瘍細胞DNA分子によって寄与される遺伝物質を特定及び特徴付けするように設計されていない。記載される方法及び系は、より大きな試料中に存在する小集団を検出するための大きな利点も提供する。そのため、多くの場合、臨床試料のごく一部しかコピー数変異を有する組織を含有していないため、それらは試料中のコピー数変異を評価するのに有用であり得る。例えば、試料が妊婦からの血液試料である場合、試料のごく一部しか循環無細胞胎児DNAを含有していないだろう。

【0059】

本明細書に開示されるバーコード化技法の使用は、所与の組の遺伝マーカーに個々の分子環境を提供する固有の能力を付与する、すなわち、所与の組の遺伝マーカー（単一マーカーとは対照的に）を個々の試料核酸分子に帰属させ、変異体協調アセンブリングを通して、複数の試料核酸分子の中で及び／または特定の染色体に対して広いまたは更には長い範囲の推測される個々の分子環境を提供する。これらの遺伝マーカーは、特定の遺伝子座、例えばSNPなどの変異体を含み得るか、またはそれらは短い配列を含み得る。更に、バーコード化の使用は、試料から抽出した全核酸集団の少数構成成分と多数構成成分とを区別する能力、例えば、血流中の循環腫瘍DNAの検出及び特徴付けを容易にする更なる利点を付与し、また任意の増幅中の増幅の偏りも減少させるまたは排除する。加えて、マイクロ流体フォーマットでの実施は、非常に小さい試料体積及び低入力量のDNAで作業する能力、ならびにゲノム全体のタグ付けを容易にするために多数の試料パーティション（例えば液滴）を迅速に処理する能力を付与する。

【0060】

前述のように、本明細書に記載される方法及び系の利点は、普遍的に利用可能な短いリード配列決定技術の使用を通して結果を得ることができることである。このような技術は、十分に特徴付けされ、高度に有効であるプロトコル及び試薬系を用いて容易に利用可能であり、研究団体内に広く普及している利点を有する。これらの短いリード配列決定技術は、例えば、Illumina, Inc.（例えば、GXII、NextSeq、MiSeq、HiSeq、X10）、Thermo-FisherのIon Torrent事業部（例えば、Ion Proton及びIon PGM）、ピロシーケンシング（pyrosequencing）法、ならびにその他から利用可能なものを含む。

【0061】

得に利点なのは、本明細書に記載の方法及び系がこれらの短いリード配列決定技術を利用し、それらの関連する低誤差率で行うことである。特に、本明細書に記載される方法及び系は、上述のように個々の分子リード長または環境を得るが、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、または更に短い個々の配列決定リード（交配对伸長を除く）を用いて、及びこのような個々の分子リード長において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、または更には0.001%未満の配列決定誤差率でそれらを得る。

【0062】

II. ワークフローの概要

1つの例示的な態様では、本開示に記載される方法及び系は、個々の試料（例えば核酸）を別個のパーティションに堆積または分配することを提供し、各パーティションはその自身の内容物を他のパーティション中の内容物から分離した状態を維持する。本明細書で使用されるとき、パーティションは、様々な異なる形態、例えばウェル、チューブ、マク

ロもしくはナノウェル、貫通孔などを含み得る容器または管を指す。しかしながら、いくつかの態様では、パーティションは流体の流れの中で流動性である。これらの管は、例えば内部の流体中心またはコアを包囲する外側バリアを有するマイクロカプセルまたは微小胞から構成され得るか、またはそれらはそのマトリックス内の物質を連行及び／または維持することが可能である多孔質マトリックスであり得る。しかしながら、いくつかの態様では、これらのパーティションは非水性連続相、例えば油相内の水性流体の液滴を含み得る。様々な異なる管は、例えば、2013年8月13日に出願された米国特許出願第13/966,150号に説明されている。同様に、非水性または油性連続相中に安定した液滴を作るためのエマルジョン系は、例えば、米国特許公開第2010/0105112号に詳細に説明されており、その全開示は参照によりその全体が本明細書に組み込まれる。ある特定の場合では、マイクロ流体チャネルネットワークが本明細書に記載されるパーティションを生成するのに適している場合がある。このようなマイクロ流体デバイスの例としては、2014年4月10日に出願された米国仮特許出願第61/977,804号に詳細に説明されているものが挙げられ、その全開示はすべての目的において参照によりその全体が本明細書に組み込まれる。細胞の水性混合物が非水性流体中に押し出される多孔膜を含む、代替の機構も個々の細胞の分配に利用され得る。このような系は、例えば Nanomi, Inc. から一般的に入手可能である。

【0063】

エマルジョン中の液滴の場合では、試料物質、例えば核酸の、別個のパーティション中への分配は、一般的に、水性の試料含有流を、水性液滴が流動流分配流体内に作られるように、非水性流の分配流体、例えばフッ素化油も流れる連結部に流すことにより達成され得、このような液滴は試料物質を含む。以下に説明されるように、パーティション、例えば液滴は、共分配されたバーコードオリゴヌクレオチドも含み得る。任意の特定のパーティション内の試料物質の相対量は、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量などを含む、系の様々な異なるパラメータを制御することにより調整され得る。本明細書に記載されるパーティションは、多くの場合、非常に小体積を有することを特徴とする。例えば、液滴系パーティションの場合では、液滴は、1000ピコリットル (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、または更には1 pL 未満の総体積を有し得る。ビーズと共分配される場合、パーティション内の試料流体体積は、上述の体積の90%未満、80%未満、70%未満、60%未満、50%未満、40%未満、30%未満、20%未満、または更には上述の体積の10%未満であり得ることを理解する。いくつかの場合では、低反応体積パーティションの使用は、非常に少量の開始試薬、例えば入力核酸を用いた反応を行う際に有利であり得る。低入力核酸の試料を分析するための方法及び系は、2014年6月26日に提出された米国仮特許出願第62/017,580号に提示されており、その全開示は参照によりその全体が本明細書に組み込まれる。

【0064】

試料がそれらのそれぞれのパーティションの中に導入されたら、本明細書に記載される方法及び系により、パーティション内の試料核酸は、一般的に、これらの核酸を特徴付けるときに、それらがそれらのそれぞれの起源に由来するとして帰属され得るように、固有の識別子と共に提供される。したがって、試料核酸は、固有の識別子（例えばバーコード配列）と共分配され得る。いくつかの態様では、固有の識別子は、これらの試料に付加され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、所与のパーティションにおけるオリゴヌクレオチドとの関係では、その中に含まれる核酸バーコード配列は同じであるが、異なるパーティションとの関係では、オリゴヌクレオチドは異なるバーコード配列を有し得るよう分配される。いくつかの態様では、1つの核酸バーコード配列のみが所与のパーティションと関連し得るが、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得る。

【0065】

核酸バーコード配列は、オリゴヌクレオチドの配列内に6～約20以上のヌクレオチドを含み得る。これらのヌクレオチドは、完全に連続している、すなわち、隣接したヌクレオチドの単一列であり得るか、またはそれらは1つ以上のヌクレオチドによって分離される2つ以上の別個のサブ配列に分離され得る。いくつかの場合では、分離されたサブ配列は、長さが約4～約16のヌクレオチドであり得る。

【0066】

共分配されたオリゴヌクレオチドは、分配した核酸の処理に有用な他の機能性配列も含み得る。これらの配列は、例えば、パーティション内の個々の核酸からゲノムDNAを増幅するための標的またはランダム／汎用増幅プライマー配列を含む一方で、例えば、配列の存在を特定するため、またはバーコード化された核酸もしくはいくつかの他の潜在的な機能性配列のいずれかをプルダウンするために、関連バーコード配列、配列決定プライマー、ハイブリダイゼーションまたはプローブ配列が付加される。同様に、試料物質と共にオリゴヌクレオチド及び関連バーコードならびに他の機能性配列の共分配は、例えば、2014年2月7日出願された米国仮特許出願第61/940,318号、2014年5月9日出願された第61/991,018号、及び2014年6月26日出願された米国特許出願第14/316,383号、ならびに2014年2月7日出願された米国特許出願第14/175,935号に記載されており、それらの全開示は参照によりそれらの全体が本明細書に組み込まれる。

【0067】

簡潔に、例示的なプロセスの1つでは、各々がビーズに放出可能に付加される多数の上述のオリゴヌクレオチドを含み得るビーズが提供され、特定のビーズに付加されたオリゴヌクレオチドのすべてが同じ核酸バーコード配列を含み得るが、多数の多様なバーコード配列が使用されるビーズの集団にわたって示され得る。いくつかの場合では、ビーズの集団は、少なくとも1000の異なるバーコード配列、少なくとも10,000の異なるバーコード配列、少なくとも100,000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を提供し得る。加えて、各ビーズは多数のオリゴヌクレオチド分子が付加されて提供され得る。特に、個々のビーズ上にバーコード配列を含むオリゴヌクレオチドの分子の数は、少なくとも約10,000オリゴヌクレオチド、少なくとも100,000オリゴヌクレオチド分子、少なくとも1,000,000オリゴヌクレオチド分子、少なくとも10億のオリゴヌクレオチド分子であり得る。

【0068】

オリゴヌクレオチドは、ビーズへの特定の刺激を適用するときに、ビーズから放出可能であり得る。いくつかの場合では、刺激は、例えばオリゴヌクレオチドを放出し得る感光性連結の切断を通じた光刺激であり得る。いくつかの場合では、熱刺激が使用される場合があり、ビーズ環境の温度の上昇がオリゴヌクレオチドの連結の切断またはビーズからの他の放出をもたらし得る。いくつかの場合では、オリゴヌクレオチドのビーズへの連結を切断するか、ないしは別の方法でビーズからオリゴヌクレオチドの放出をもたらし得る化学刺激が使用され得る。

【0069】

本明細書に記載される方法及び系によると、付加されたオリゴヌクレオチドを含むビーズは、単一のビーズ及び単一の試料が個々のパーティション内に含有されるように、個々の試料と共分配され得る。いくつかの場合では、単一のビーズパーティションが望ましい場合、流体の相対流量は、占有されるこれらのパーティションが主に単独で占有されることを確実にするために、平均して、パーティションが、パーティション当たり1つのビーズ未満を含有するように制御され得る。同様に、より高いパーセンテージのパーティションが占有されること、例えば占有されていないパーティションが数パーセントのみであることを可能にすることを提供するように流量を制御することを希望してもよい。いくつかの態様では、所望の数の単独占有パーティション、ある特定のレベル未満の占有されない

パーティション、及びある特定のレベル未満の複数の占有されたパーティションを確実にするために、流れ及びチャネル構造が制御される。

【0070】

図3は、コピー数変異またはハプロタイプアッセイに使用するためなど、試料核酸をバーコード化し、続いて配列決定するための例示的な方法を図示する。第1に、核酸を含む試料は供給源から得ることができ300、一組のバーコード化されたビーズも得ることができる310。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出するために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例において、核酸を含む低量の試料305、バーコード化されたビーズ315、及びいくつかの場合では、他の試薬、例えば還元剤320が組み合わせられ、分配を受ける。例として、このような分配は、マイクロ流体デバイス325などの液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス325の補助により、油中水型エマルジョン330が形成され得、エマルジョンは、試料核酸305、還元剤320、及びバーコード化されたビーズ315を含有する水性液滴を含有している。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる335。ランダムN-merは次に、試料核酸の異なる領域をプライムし、増幅後に試料の増幅されたコピーをもたらすことができ、各コピーはバーコード配列でタグ付けされている340。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダムN-mer配列を含有する一組のオリゴヌクレオチドを含有する。続いて、エマルジョンを破壊し345、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法350（例えばPCR）を介して付加され得る。次に、配列決定が実施され得355、配列決定データを解釈するためにアルゴリズムが適用される360。配列決定アルゴリズムは一般的に、例えば、配列決定リードを整合し、かつ／または特定の配列リードが属する試料を特定するために、バーコードの分析を実施することができる。

【0071】

上述のように、単一ビーズの占有が所望され得るが、複数の占有されたパーティション、または占有されていないパーティションが多くの場合存在し得ることを理解する。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共分配するためのマイクロ流体チャネル構造の例が図4に模式的に図示される。示されるように、チャネルのセグメント402、404、406、408、及び410がチャネル連結部412で流体連通で提供される。個々の試料414を含む水性流は、チャネル連結部412に向かってチャネルセグメント402を流れる。本明細書の他の箇所に記載されるように、これらの試料は、分配プロセスの前に水性流体内に懸濁され得る。

【0072】

同時に、バーコードを保有するビーズ416を含む水性流は、チャネル連結部412に向かってチャネルセグメント404を流れる。非水性分配流体は側方チャネル406及び408の各々からチャネル連結部412内に導入され、組み合わせられた流れは、出口チャネル410に流れる。チャネル連結部412内で、チャネルセグメント402及び404からの2つの組み合わせられた水性流が組み合わせられ、共分配された試料414及びビーズ416を含む液滴418内に分配される。前述のように、チャネル連結部412で組み合わせられる流体の各々の流れ特徴を制御すること、ならびにチャネル連結部の形状を制御することにより、生成されるパーティション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得るために、組み合わせ及び分配を最適化することができる。

【0073】

理解されるように、例えば、化学刺激、核酸伸長、転写、及び／または増幅試薬（ポリメラーゼ、逆転写酵素、ヌクレオシド三リン酸、またはNTP類似体など）、プライマー配列、ならびに追加の補助因子（このような反応に使用される二価の金属イオンなど）、連結反応試薬（リガーゼ酵素及び連結配列など）、染料、標識、または他の標識化試薬を含む、いくつかの他の試薬が試料及びビーズと共に共分配され得る。

【0074】

共分配されたら、ビーズに配備されたオリゴヌクレオチドは、分配された試料をバーコード化し、増幅するために使用され得る。試料の増幅及びバーコード化においてこれらのバーコードオリゴヌクレオチドを使用するための例示的なプロセスは、2014年2月7日に出願された米国特許出願第61/940,318号、2014年5月9日に
10
出願された第61/991,018号、及び2014年6月26日に
出願された米国特許出願第14/316,383号に詳細に記載されており、それらの全開示は参照によりそれらの全体が本明細書に組み込まれる。簡潔に、一態様では、オリゴヌクレオチドは、試料と共に分配され、それらのビーズから試料を有するパーティションの中に放出されるビーズ上に存在する。オリゴヌクレオチドは、バーコード配列と共に、その5'端でプライマー配列を含み得る。このプライマー配列は、試料の多くの異なる領域をランダムにプライムするように意図されたランダムオリゴヌクレオチド配列であるか、または試料の特定の標的領域の上流をプライムすることを標的とする特異的プライマー配列であり得る。

【0075】

放出されたら、オリゴヌクレオチドのプライマー部分は試料の相補領域にアニールすることができる。試料及びビーズとも共分配される伸長反応試薬、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）は、次に、テンプレートとして試料を使用してプライマー配列を伸長して、プライマーがアニールされたテンプレートの鎖に対する相補的断片を生成し、相補的断片はオリゴヌクレオチド及びその関連するバーコード配列を含む。複数のプライマーのアニリング及び試料の異なる部分への伸長は、試料の重複する相補的断片の大きなプールをもたらす得、各々、それが作られたパーティションを示すその自身のバーコード配列を保有する。いくつかの場合では、これらの相補的断片はそれら自体、パーティションに存在するオリゴヌクレオチドによりプライムされたテンプレートとして使用されて、再びバーコード配列を含む相補体の相補体を生成する。いくつかの場合では、この複製プロセスは、最初の相補体が複写
20
30
されるとき、分子が更なる反復コピーを生成するための基盤となる能力を減少させるヘアピン構造または部分的なヘアピン構造の形成を可能にするために、その末端でまたはその付近で2つの相補的配列を生成するように構成される。この一例の模式図は図5に示される。

【0076】

図が示すように、バーコード配列を含むオリゴヌクレオチドは、試料核酸504と共に、例えば、エマルジョン中の液滴502に共分配される。本明細書の他の箇所に記述されるように、オリゴヌクレオチド508は、パネルAに示されるように、オリゴヌクレオチドがビーズ506から放出可能であり得る、試料核酸504と共に分配されるビーズ506上に提供され得る。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、ならびに所与の配列決定系の付加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ける結合に使用されるP5配列を含むように示される。示されるように、オリゴヌクレオチドは、試料核酸504の一部の複製をプライムするためのランダムまたは標的N-merを含み得るプライマー配列516も含む。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をプライムするために使用される、「リード1」またはR1プライミング領域などの配列決定プライミング領域を提供し得る配列514もオリゴヌクレオチド508内に含まれる。いくつかの場合では、バーコード配列512、固定化配列510、及びR1配列514は、所与のビーズに付加さ
40
50

れるオリゴヌクレオチドのすべてに共通してよい。プライマー配列 5 1 6 はランダム N－mer プライマーにより変動し得るか、またはある特定の標的用途により所与のビーズ上のオリゴヌクレオチドに共通し得る。

【0077】

プライマー配列 5 1 6 の存在に基づき、オリゴヌクレオチドは、パネル B に示されるように試料核酸をプライムすることができ、これは、ビーズ 5 0 6 及び試料核酸 5 0 4 とともに共分割されるポリメラーゼ酵素及び他の伸長試薬を用いて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 に示されるように、ランダム N－mer プライマーに関して、試料核酸 5 0 4 の複数の異なる領域にアニールするだろうオリゴヌクレオチドの伸長後、複数の重複する相補体または核酸の断片、例えば断片 5 1 8 及び 5 2 0 が作られる。試料核酸の部分に相補的である配列部分、例えば配列 5 2 2 及び 5 2 4 を含むが、これらの構築物は一般的に、本明細書において、付加されたバーコード配列を有する試料核酸 5 0 4 の断片を含むと称される。理解されるように、上述のようにテンプレート配列の複製部分は、多くの場合、本明細書において、そのテンプレート配列の「断片」と称される。しかしながら、前述にも関わらず、「断片」という用語は、例えば酵素的、化学的、または機械的断片化による配列の所与の分子の実際の断片化など、テンプレート配列の部分を提供する他の機構により作られたものを含む、発生源の核酸配列の部分のいずれの表現、例えばテンプレートまたは試料核酸を包含する。しかしながら、いくつかの態様では、テンプレートまたは試料核酸配列の断片は、根底にある配列またはその相補体の複製された部分を示す場合がある。

【0078】

バーコード化された核酸断片は、次に、例えば配列分析により特徴付けされるか、またはそれらは、パネル D に示されるように、プロセスにおいて更に増幅されてよい。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 3 0 6 から放出され、断片 5 1 8 及び 5 2 0 をプライムし得る。特に、同様に、オリゴヌクレオチド 5 0 8 b におけるランダム N－mer プライマー 5 1 6 b の存在に基づき（いくつかの場合では、所与のパーティションにおいて他のランダム N－mer とは異なる場合がある、例えばプライマー配列 5 1 6）、オリゴヌクレオチドは、断片 5 1 8 とアニールし、試料核酸配列の一部の重複を含む、配列 5 2 8 を含む断片 5 1 8 の少なくとも一部に相補体 5 2 6 を作るように伸長される。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 を通して複製されるまで続く。本明細書の他の箇所に記述されるように、及びパネル D に図示されるように、オリゴヌクレオチドは、例えば断片 5 1 8 内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 を通して複製された後、所望の地点でポリメラーゼによる複製の停止を促すように構成され得る。本明細書に記載されるように、これは、例えば使用されるポリメラーゼ酵素によって処理することが不可能な異なるヌクレオチド及び／またはヌクレオチド類似体の組み込みを含む、異なる方法によって達成され得る。例えば、これは、非ウラシル耐性ポリメラーゼがその領域の複製を中止するのを妨げるために配列領域 5 1 2 内のウラシル含有ヌクレオチドの包含を含み得る。結果として、バーコード配列 5 1 2、付加配列 5 1 0、R 1 プライマー領域 5 1 4、及びランダム N－mer 配列 5 1 6 b を含む、完全長オリゴヌクレオチド 5 0 8 b を一端に含む断片 5 2 6 が作られる。配列のもう一方の端に、第 1 のオリゴヌクレオチド 5 0 8 のランダム N－mer に対する相補体 5 1 6 ' ならびに配列 5 1 4 ' として示される R 1 配列のすべてまたは一部に対する相補体が含まれ得る。R 1 配列 5 1 4 及びその相補体 5 1 4 ' は、次に、共にハイブリダイズして、部分的ヘアピン構造 5 2 8 を形成する。理解されるように、ランダム N－mer は異なるオリゴヌクレオチドの間で異なるため、これらの配列及びそれらの相補体はヘアピン形成に関与するとは考えられず、例えば、ランダム N－mer 5 1 6 に対する相補体である配列 5 1 6 ' はランダム N－mer 配列 5 1 6 b に相補的であるとは考えられないだろう。これは、他の用途、例えば N－mer が所与のパーティション内のオリゴヌクレオチドの中で一般的であろう標的プライマーに関して、そうではないだろう。

10

20

30

40

50

【0079】

これらの部分的なヘアピン構造を形成することにより、試料配列の第1レベルの重複を更なる複製から除去する、例えばコピーの反復コピーを妨げることを可能にする。部分的なヘアピン構造は、作られた断片、例えば断片526の後続処理に有用な構造も提供する。

【0080】

次に、複数の異なるパーティションからの断片のすべては、本明細書に記載される高処理シーケンサーで配列決定するためにプールされ得る。各断片はその起源のパーティションについてコード化されるため、その断片の配列はバーコードの存在に基づきその起源に帰属し戻ることができる。これは図6に模式的に図示される。一例に示されるように、第1の供給源600（例えば、個々の染色体、核酸の鎖など）に由来する核酸604及び異なる染色体602または核酸の鎖に由来する核酸606は各々、上述のようにそれら自体のバーコードオリゴヌクレオチドの組と共に分配される。

【0081】

次に、各パーティション内で、各核酸604及び606は、第1の断片（複数可）の重複する第2の断片の組、例えば第2の断片の組608及び610を別個に提供するように処理される。この処理は、特定の第1の断片に由来する第2の断片の各々についても同じであるバーコード配列を有する第2の断片も提供する。示されるように、第2の断片組608のバーコード配列は「1」と示される一方で、断片組610のバーコード配列は「2」と示される。多数の異なる断片組を示差的にバーコード化するために、バーコードの多様なライブラリが使用され得る。しかしながら、異なる第1の断片からの第2の断片組のすべてが異なるバーコード配列でバーコード化される必要はない。いくつかの場合では、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得る。本明細書の他の箇所で、多様なバーコードライブラリが詳細に記載される。

【0082】

次に、例えば断片組608及び610からのバーコード化された断片は、例えばIlluminaまたはThermo Fisher, Inc.のIon Torrent事業部から入手可能な合成技術による配列を使用して、配列決定のためにプールされ得る。配列決定されたら、配列リード612は、少なくとも部分的に含まれるバーコードに基づき、いくつかの場合では、部分的に断片自体の配列に基づき、例えば集められたリード614及び616に示されるように、それらのそれぞれの断片に帰属され得る。次に、各断片組の帰属された配列リードがアセンブリングされて、各試料断片のアセンブリングされた配列、例えば配列618及び620を提供し、これは続いてそれらのそれぞれの元の染色体（600及び602）に更に帰属し戻ることができる。ゲノム配列をアセンブリングするための方法及び系は、2014年6月26日に出願された米国仮特許出願第62/017,589号に記載されており、その全開示は参照によりその全体が本明細書に組み込まれる。いくつかの例では、ゲノム配列は、新規アセンブリング及び／または参照に基づくアセンブリングによりアセンブリングされる（例えば参照にマッピングする）。

【0083】

III. フェージング及びコピー数アッセイへの方法及び系の適用

本明細書に記載される系及び方法の一態様では、配列についてのフェーズ情報を決定する際に、配列リードをより長い発生元の分子に帰属させる能力が使用される。一例では、2つ以上の特定の遺伝子変異体配列（例えば、対立遺伝子、遺伝マーカー）を明らかにする配列と関連するバーコードが比較されて、その遺伝マーカーの組が試料において同じ染色体または異なる染色体上に存在するか否かが決定される。このようなフェージング情報は、試料においてある特定の標的染色体または遺伝子の相対的コピー数を決定するために使用され得る。記載される方法及び症状の利点は、フェージング及びコピー数情報を決定するために、複数の位置、遺伝子座、変異体などが、それらが由来する個々の染色体または核酸鎖を特定するために使用され得ることである。多くの場合、明細書に記載されるフェージング、ハプロタイプ、及びコピー数変異情報を決定するために、染色体に沿って複

数の位置（例えば、2、3、4、5、6、7、8、9、10、20、30、40、50、100、500、1000、5000、10000、50000、100000、または500000を超える）が使用される。

【0084】

例として、上述されるように、本明細書に記載される方法及び系は、上述の分配及び帰属態様により、比較的短い配列リードを提供し得る配列決定技術を利用するにも関わらず、個々の核酸断片、例えば個々の核酸分子から有効な長い配列リードを提供するときに有用であり得る。これらの長い配列リードは単一の開始断片または分子に帰属され得るため、配列における変異体の位置は、同様に、単一分子及び外挿により単一染色体に帰属させることができる。加えて、同じ染色体に由来すると推測され得る整合した配列を提供するために、隣接した断片の整合特徴として任意の所与の断片上の複数の位置を利用することができる。例として、第1の断片が配列決定され得、上述の帰属方法及び系により、その配列上に存在する変異体はすべて単一染色体に帰属され得る。次に、1つの染色体にのみ存在すると決定される複数のこれらの変異体を共有する第2の断片は、同じ染色体に由来すると仮定され、よって、2つの断片のフェーズ化整合を作るために、第1の断片と整合され得る。これを繰り返すことにより、長い範囲のフェーズ情報の特定を可能にする。単一染色体上の変異体の特定は、既知の参照、例えばHapMap、または例えばさもないければ同一の配列伸長上で異なる変異体を示す、配列決定データの集計のいずれかから得ることができる。

【0085】

図7はフェーズ化配列決定プロセスの例の模式図を提供する。示されるように、例えば染色体、染色体断片、エクソーム、または他の大きい単一核酸分子などの発生源の核酸702は、複数の大きい断片704、706、708に断片化され得る。発生源の核酸702は、特定の核酸分子、例えば染色体に特異的であるいくつかの配列変異体（A、B、C、D、E、F、及びG）を含み得る。本明細書に記載のプロセスによると、発生源の核酸は、関連する配列変異体のサブセットを含む複数の大きい重複する断片704、706、及び708に断片化され得る。次に、本明細書に記載されるように、各断片は分配され、サブ断片に更に断片化され、バーコード化されて、より大きい断片の複数の重複するバーコード化されたサブ断片を提供し、所与のより大きい断片のサブ断片は同じバーコード配列を有する。例えば、バーコード配列「1」及びバーコード配列「2」と関連するサブ断片は、それぞれ、パーティション710及び712に示される。次に、バーコード化されたサブ断片はプールされ、配列決定され、そして配列決定されたサブ断片がアセンブリングされて、長い断片配列714、716、及び717を提供することができる。長い断片配列714、716、及び717のうちの1つ以上は複数の変異体を含み得る。次に、長い断片配列は、配列714、716、及び717からの重複するフェーズ化変異体情報に基づき更にアセンブリングされて、フェーズ化位置が決定され得るフェーズ化配列718を提供する。

【0086】

フェーズ化位置が決定されたら、様々な方法でその情報を更に利用することができる。例えば、ある特定の障害の遺伝的リスクを評価する、父方対母方の特徴を特定する、異数体を特定する、またはハプロタイプ分析情報を特定する際にフェーズ化変異体の知識を利用することができる。

【0087】

本明細書に開示される系及び方法のいくつかの態様では、コピー数変異アッセイは、コピー数計数の精度を改善するために、2つ以上のフェーズ化遺伝マーカーの同時検出を使用して実施される。フェージングの情報を利用することにより、複数の遺伝子座に対する、及びハプロタイプにわたるリードの計数のみに基づく素朴な方法による変化と比較してシグナルの相対強度を増加させることができる。加えて、フェージングの情報を利用することにより、位置特異的偏りの正規化、シグナルを実質的に更に高めることを可能にする。コピー数変異（CNV）の精度は、配列決定深度、CNVの長さ、コピーの数などを

含む無数の要因に依存し得る。本明細書に提供される方法及び系は、少なくとも70%、80%、85%、90%、91%、92%、93%、94%、95%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%、99.95%、99.99%、99.995%、または99.999%の精度でCNVを決定し得る。いくつかの場合では、本明細書に提供される方法及び系は、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%、0.0001%、または0.00005%未満の誤差率でCNVを決定する。同様に、本明細書に提供される方法及び系は、少なくとも70%、80%、85%、90%、91%、92%、93%、94%、95%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%、99.95%、99.99%、99.995%、または99.999%の精度で2つ以上の遺伝子変異体のフェージング／ハプロタイプ情報を検出し得る。いくつかの場合では、本明細書に提供される方法及び系は、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%、0.0001%、または0.00005%未満の誤差率でフェージングまたはハプロタイプ情報を決定する。本開示は、遺伝子座特異的偏りを取り除く方法も提供し、遺伝子座特異的变化は、少なくとも2倍、3倍、4倍、5倍、10倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍、100倍、200倍、500倍、1000倍、5000倍、または10000倍減少される。本明細書に提供される方法及び系は、コピー数の変化が染色体の数または染色体の一部の変化を反映するなど、コピー数の変動を検出するために使用され得る。いくつかの場合では、本明細書に提供される方法及び系は、同じ染色体上に存在する遺伝子のコピー数の変動を検出するために使用され得る。

【0088】

図8（上のパネル）は、健康な患者のゲノムのサブセットの模式図である。この患者は、示される遺伝子座にヘテロ接合の遺伝子型、及び別個の染色体上に位置する2つの別個のハプロタイプ（1及び2）805、810を有する。患者の天然に生じる変異（SNPまたはインデルなど）は丸として示される。図8は癌815を有する患者のゲノムも示す。ある特定の癌はハプロタイプのコピー数の増加と関連する。中央のパネルはハプロタイプ2、810の増加を示す。癌は、ハプロタイプ2、820の減少を示す図8の下のパネルに示されるように、ハプロタイプの数の減少と関連し得る。通常の配列決定技法は、ハプロタイプコピーのこの減少または増加を正確に決定することができない。図9aに示されるように、これは一部、患者の血液中の腫瘍寄与DNA910が全DNAのごく一部であり、その大部分が正常な組織905によって寄与されたDNAであるという事実による。この低濃度の腫瘍DNAは、通常の配列決定技法を用いたコピー数の不正確な検出をもたらす。図9bを参照。コピー変動がない平均深度D 935での予想される計数のピーク920とコピー減少925（940）及びコピー増加930（945）のピークにおける差は検出が困難である。任意の所与の個々のマーカーに関して、複製試験におけるコピー数アッセイの結果の分布は、ポアソン分布を近似する方法で正しい答えの周りに分布され得、分布の幅はアッセイにおける様々な偶然誤差源に依存する。所与の試料に関して、コピー数の変化は試料の比較的小さい部分であり得るため、単一遺伝マーカーを監視するための幅広い確率分布は正確な結果を隠す可能性がある。この問題は、図10に示されるように（左側のパネル）、通常の配列決定技法が一度にハプロタイプの1つの変異体位置のみを検討するという事実による。このような技法を使用することにより、コピー減少1025、正常なコピー1020、及びコピー増加1030を表すピーク間で大幅に重複し得る。本明細書に開示される技法は、ハプロタイプ全体（または一部）の検出を可能にし、分析能を高め、ならびにコピー増加及び減少の検出を改善する。図10（右側のパネル）。この改善は図11に模式的に図示され、正常な検出1100は広がった重複するピークをもたらすが、本明細書の技法1110は、より鋭いピーク及び改善されたコピー増加

または減少の分析能を高める。2つ以上のフェーズ化遺伝マーカー、特に単一染色体上に共存することが知られ、したがって同期化されたランダムではない様式で、多いまたは少ない数で大概常に現れる可能性があるマーカーの同時監視を使用することにより、予想された結果分布の幅を狭小し、同時に計数の精度を改善する効果を有する。

【0089】

癌を検出し、診断する際の利点に加えて、本明細書に提供される方法及び系は、胎児異数体を検出するためのより正確で高感度のプロセスも提供する。

【0090】

胎児異数体は胎児染色体数における異常である。異数体は、一般的に、顕著な身体的及び神経性障害をもたらす。例えば、X染色体の数の減少はターナー症候群に關与する。染色体番号21のコピー数の増加はダウン症候群を引き起こす。羊水穿刺または絨毛検査（CVS）などの侵襲性試験は流産のリスクにつながる場合があり、母体血液を検査するあまり侵襲性ではない方法がここでは使用される。

【0091】

本明細書に記載される方法は、胎児異数体を非侵襲的に検出する際に有用であり得る。例示的なプロセスは図12に示される。異数体ゲノムを有する胎児を妊娠しているリスクにある妊婦が検査される1200。胎児遺伝物質を含有する母体の血液試料を回収する1205。次に、遺伝物質（例えば無細胞核酸）を血液試料から抽出する1210。一組のバーコード化されたビーズも得ることができる1215。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出するために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例では、試料1210、バーコード化されたビーズ1220、及びいくつかの場合では、他の試薬、例えば還元剤が組み合わせられ、分配が行われる。例として、このような分配は、マイクロ流体デバイス1225などの液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス1225の補助により、油中水型エマルジョン1230が形成され得、エマルジョンは、試料核酸1210、バーコード化されたビーズ1215、及びいくつかの場合では、還元剤を含有する水性液滴を含有する。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる1235。ランダムN-merは次に、試料核酸の異なる領域をプライムし、増幅後に試料の増幅されたコピーをもたらすことができ、各コピーはバーコード配列でタグ付けされている1240。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダムN-mer配列を含有する一組のオリゴヌクレオチドを含有する。他の実施形態では、個々の液滴は、固有のバーコード配列を含むか、またはいくつかの場合では、液滴の全集団のある特定の割合は固有の配列を有する。続いて、エマルジョンを破壊し1245、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法（例えばPCR）を介して付加され得る。次に、配列決定は、任意の好適な種類の配列決定プラットフォーム（例えば、Illumina、Ion Torrent、Pacific Biosciences SMRT、Roche 454配列決定 SOLiD配列決定など）1250、及び配列決定データを解釈するために適用されるアルゴリズム1255を介して実施される。配列決定アルゴリズムは一般的に、例えば、配列決定リードを整合し、かつ／または特定の配列リードが属する試料を特定するために、バーコードの分析を実施することができる。整合された配列は、付加された固有のバーコードに基づき、それらのそれぞれの遺伝子起源（例えば染色体）に更に帰属させることができる。次に、染色体コピーの数は正常な二倍体染色体のものと比較される1260。患者は、異なる染色体の任意のコピー数異常及び関連リスク／疾患を通知される1265。

10

20

30

40

50

【0092】

フェージング、例えば、遺伝子変異体に関連するか、または異なる染色体上に存在するかの決定は、様々な用途に有用な情報を提供し得る。例として、フェージングは、疾患に関連するゲノムのある特定の転座が存在するかを決定するのに有用である。このような転座の検出は鑑別診断及び処置の修正を可能にする。ゲノムのどの対立遺伝子に関連するかの決定は、遺伝子がどのように受け継がれたかを考慮するのに有用である。

【0093】

多くの場合、染色体対の各個々の染色体の対立遺伝子のパターンであるハプロタイプを知るために有用であり得る。例えば、1つの染色体上に存在する不活性化突然変異の2つのコピーは、限定的な結果のものであり得るが、2つの染色体間で分散された場合、例えばいずれの染色体も活性遺伝子産物を供給しない場合、大きな作用を有し得る。これらの作用は、例えば疾患のリスクの増加またはある特定の薬剤への応答の欠如で表される場合がある。

【0094】

I V. 構造的変異の特定／特徴付けへの方法及び系の適用

他の用途では、本明細書に記載される方法及び系は、広範囲の異なる遺伝子構造的変異の特定及び特徴付けに関する長い範囲の分子配列情報を得る際に高度に有用である。上述のように、これらの変異は、挿入、欠失、重複、レトロトランスポゾン、転座、反転、短縦列型反復及び長縦列型反復などを含む、多種多様な異なる変異体事象を含む。これらの構造的変異は、様々な多様な遺伝子疾患と関連すると考えられているため、大きな学術的関心のものである。

【0095】

これらの変異における関心にも関わらず、これらの構造的変異を特定及び特徴付けする効果的かつ効率的な方法はあまりない。一部、これは、これらの変異が異常な配列セグメントの存在を特徴としないが、代わりに、正常な配列セグメントと考えられるであろう異常配列環境を伴うか、または単純に配列情報が欠落しているためである。それらの比較的短いリード長のため、大半の配列決定技術は、有意な環境、そして特にそれらが生成する配列リードの、例えばそれらのリード長を超えた長い範囲の配列環境を提供できず、よって、アセンブリングプロセスにおいてこれらの変異の特定を失う。これらの変異を特定する際の問題は、多くの分子、例えば複数の染色体が組み合わされて、変異を含む及び変異を含まない両方のゲノム物質を含み得るコンセンサス配列をもたらすこれらの技術の集合アプローチにより更に複雑化される。

【0096】

しかしながら、ここに記載される方法及び系の文脈において、個々の発生元の核酸分子に帰属される長い範囲の配列情報を導くために短いリード配列決定技術を利用し、よって、これらの個々の分子全体または一部に含有される長い範囲の変異体領域の配列環境を維持することができる。

【0097】

上述のように本明細書に記載される方法及び系は、個々の発生元の核酸分子に帰属される長い範囲の配列情報を提供し、更に、この長い範囲の配列情報を有する際に、これらのより長い配列情報の比較及び重複を通して更に長い範囲の配列環境を推測することができる。そのような長い範囲の配列情報及び／または推測される配列環境は、利用可能な技法を使用して容易に特定されない多数の構造的変異の特定及び特徴付けを可能にする。

【0098】

図2に簡潔した様式で図示されるが、図13A及び13Bは、本明細書に記載される方法及び系を使用して、ある特定の種類の構造的変異を特定するためのより詳細な例示的プロセスを提供する。示されるように、生物のゲノムまたは生物からの組織は、通常、図13Aに図示される第1の遺伝子型を含み得、第1の遺伝子1304を含む第1の遺伝子領域1302は、第2の遺伝子1208を含む第2の遺伝子領域1306から分離されている。この分離は、例えば同じエクソンにおける異なる領域、同じ染色体上の異なるエクソ

ン、異なる染色体などを含む、遺伝子間の距離の範囲を反映し得る。しかしながら、図 1 3 B に示されるように、変異体配列 1 3 1 4 において遺伝子融合 1 3 1 2 として遺伝子 1 3 0 4 と 1 3 0 8 との間で遺伝子融合体を作るように遺伝子 1 3 0 8 が遺伝子領域 1 3 0 4 に挿入されることが生じた転座事象を反映する遺伝子型が示される。

【0099】

大きなゲノム構造変異体（大きな反転または転座など）を検出するための現在の方法は、変異体の切断点（例えば、転座した部分が一緒に融合するゲノム遺伝子座）に及ぶリード対に依存する。そのようなリード対が配列決定実験中に観察されることを確実にするために、非常に詳細な配列決定が要求され得る。標的配列決定（エクソーム配列決定など）において、現在の配列決定技術を用いた構造変異体の検出は、切断点が標的指向された領域（例えばエクソン）内でない限り（この可能性はかなり低い）、ほぼ不可能である。

10

【0100】

しかしながら、本明細書に記載されるバーコード方法及び系により提供される情報は、構造変異体を検出する能力を大幅に改善することができる。直観的に、切断点の左方及び右方への遺伝子座は、ゲノム DNA の共通の断片上にある傾向にあり、したがって単一のパーティション内に維持され、よって共通のまたは共有のバーコード配列でバーコード化され得る。剪断の確率的性質により、バーコードのこの共有は、配列が切断点から遠くなるにつれて減少する。統計手法を使用して、2 つのゲノム遺伝子座間でのバーコードの重複が偶然に予測されるであろうものよりも大幅に大きいかを決定することができる。そのような重複は切断点の存在を示唆する。重要なことは、バーコード情報は、このような情報 20 が利用可能である場合、従来の配列決定（切断点に及ぶリードからの情報など）により提供される情報を補足する。

【0101】

本明細書に記載される方法の文脈において、適切な遺伝子領域を含む、生物からのゲノム物質は、上述のように比較的長い断片を含むように断片化される。これは、図 1 3 A の非転座遺伝子型に関して図示される。示されるように、それぞれ、遺伝子領域 1 3 0 2 及び 1 3 0 6 を含む 2 つの長い個々の第 1 の分子断片 1 3 1 6 及び 1 3 1 8 が作られる。これらの断片は、それぞれ、パーティション 1 3 2 0 及び 1 3 2 2 に別個に分配され、第 1 の断片の各々は、パーティション内でそれぞれいくつかの第 2 の断片 1 3 2 4 及び 1 3 2 6 に断片化され、この断片化プロセスは、所与のパーティション内の第 2 の断片のすべて 30 に共通である固有の識別子またはバーコード配列を第 2 の断片に付加する。タグまたはバーコードは、それぞれ、パーティション 1 3 2 0 及び 1 3 2 2 の各々に関して「1」または「2」により示される。結果として、完全に分離した遺伝子 1 3 0 4 及び 1 3 0 8 は、異なって分配され、異なってバーコード化される第 2 の断片群をもたらし得る。

30

【0102】

バーコード化されたら、次に、第 2 の断片はプールされ、第 2 の断片の配列ならびにその断片のバーコード配列の両方を提供することができる核酸配列決定プロセスを受けることができる。特定のバーコード、例えば 1 または 2 の存在に基づき、次に、第 2 の断片の配列は、バーコードを各配列に帰属させることにより示されるように、ある特定の発生元の配列、例えば、遺伝子 1 3 0 4 または 1 3 0 8 に帰属させることができる。いくつかの 40 場合では、発生元の第 1 の断片配列を分離するようにバーコード化された第 2 の断片の配列のマッピングは、転座が生じなかったことを決定するために十分に決定的であり得る。しかしながら、いくつかの場合では、例えば、アセンブリされた配列 1 3 3 0 及び 1 3 3 2 により示されるように、発生元の第 1 の断片の配列のすべてまたは一部についてアセンブリされた配列を提供するために、第 2 の断片の配列をアセンブリすることができる。

40

【0103】

図 1 3 A に示される転座されなかった遺伝子型の例とは対照的に、図 1 3 B は転座含有遺伝子型に適用された同じプロセスの模式図を示す。示されるように、第 1 の長い核酸断片 1 3 5 2 は変異体配列 1 3 1 4 から生成され、転座変異体の少なくとも一部、例えば 50

50

遺伝子融合体 1 3 1 2 を含む。次に、第 1 の断片 1 3 5 2 は別個のパーティション 1 3 5 4 に分配される。パーティション 1 3 5 4 内で、第 1 の断片 1 3 5 2 は、同様にパーティション 1 3 5 4 内のすべての第 2 の断片 1 3 5 6 に関して同じである固有のバーコード（バーコード「1」として示される）を含む第 2 の断片 1 3 5 6 に更に断片化される。上記のように、第 2 の断片をプールし、配列決定することにより、第 2 の断片の根底にある配列ならびにそれらの関連するバーコードを提供する。次に、これらのバーコード化された配列は、それらのそれぞれの遺伝子配列に帰属させることができる。しかしながら、示されるように、両遺伝子は同じバーコード配列を含む帰属された第 2 の断片の配列を反映し得、それらが同じパーティションに由来することを示し、同じ発生源の分子の可能性があり、遺伝子融合体であることを示す。これは、遺伝子融合体の少なくとも一部も含むが、異なるバーコードで異なるパーティションに処理されるいくつかの重複する第 1 の断片を提供することにより更に検証され得る。

10

【0104】

いくつかの場合では、もともと分離された遺伝子の各々に帰属する複数の異なるバーコード配列（及びそれらの根底にある断片配列）の存在は、遺伝子融合体または他の転座事象の存在を示し得る。いくつかの場合では、少なくとも 2 つのバーコード、少なくとも 3 つの異なるバーコード、少なくとも 4 つの異なるバーコード、少なくとも 5 つの異なるバーコード、少なくとも 10 の異なるバーコード、少なくとも 20 以上の異なるバーコードを、参照配列に基づき分離されたと予想されたとであろう 2 つの遺伝子領域に帰属させることにより、これらの領域を互いに近位に、隣接して配置した、ないしは別の方法で一体化した転座事象の指標を提供し得る。いくつかの場合では、分配される断片のサイズは、変異体の連結を特定することができる感度を示し得る。特に、所与の液滴中の断片の長さが 10 kb である場合、その 10 kb サイズ範囲内である連結が検出可能であろうことが予想されるだろう。

20

【0105】

同様に、変異体及び野生型構造の両方が同じ 10 kb 断片内に入る場合、両方ともおそらく共通のまたは共有のバーコードを通して連結を示すため、その変異体の特定はより困難であろうことが予想されるだろう。そのため、断片サイズの選択は、野生型または変異体としてに関わらず、検出された連結された配列の相対的近接を調整するために使用され得る。しかしながら、一般に、100 塩基超、500 塩基超、1 kb 超、10 kb、20 kb 超、30 kb 超、40 kb 超、50 kb 超、60 kb 超、70 kb 超、80 kb 超、90 kb 超、100 kb 超、200 kb 超、または更にはそれ以上で通常分離される近位配列をもたらず構造変異体は、本明細書において、連結が共有のまたは共通のバーコードにより、及び／または記述されるように、切断点に及ぶ配列データにより示される変異体ゲノムにおけるこれらの連結されない配列セグメント間の連結を特定することにより容易に特定され得る。このような連結は一般的に、これらの連結された配列が 50 kb 未満、40 kb 未満、30 kb 未満、20 kb 未満、10 kb 未満、5 kb 未満、4 kb 未満、3 kb 未満、2 kb 未満、1 kb 未満、500 塩基未満、200 塩基未満、または更にはそれ未満でゲノム配列内で分離されるとき、特定可能である。

30

【0106】

いくつかの場合では、通常、例えば、10 kb 超、20 kb 超、30 kb 超、40 kb 超、または 50 kb 以上で分離されるであろう、互いに近位に位置付けられた、または連結された 2 つの配列をもたらず構造的変異は、2 つの配列部分に共通のバーコードを含むマッピング可能なバーコード化された配列の総数のパーセンテージにより特定され得る。

40

【0107】

理解されるように、いくつかの場合では、本明細書に記載されるプロセスは、ある特定の配列距離内の配列が、野生型または変異体配列としてに関わらず、単一パーティション内、例えば単一核酸断片として含まれることを確実にすることができる。例えば、共通のまたは重複するバーコード配列が 2 つの配列にマッピングされたバーコードの総数の 1 % を超える場合、2 つの配列セグメントとの関係では、そして特に通常連結されないであ

50

う2つの配列セグメント、例えば構造的変異との関係では、連結を特定するために使用され得る。いくつかの場合では、共有のまたは共通のバーコードは、ゲノム内の構造的変異を構成する構造連結を特定するために、2つの通常分離された配列に対して合計マッピング可能バーコードの2%超、3%超、4%超、5%超、6%超、7%超、8%超、及びいくつかの場合では、9%超、または更には10%超であり得る。いくつかの場合では、共有のまたは共通のバーコードは、構造的変異を有しないと知られる対照ゲノムより統計的に有意に大きな割合または数で検出され得る。加えて、第2の配列の断片が、例えば第2の断片1358にあるような、変異体配列が「正常な」配列または「切断点」に接触する地点に及ぶ場合、この情報を遺伝子融合体の更なる証拠として使用することができる。

【0108】

再び、上記のように、第2の断片の配列をアセンブリングして、アセンブリングされた配列1360として示される遺伝子融合体1312のアセンブリングされた配列を得ることにより、遺伝子融合体1312の構造を更に解明することができる。

【0109】

更に、バーコード配列の存在は短い配列をより長い発生元の断片の配列にアセンブリングすることを可能にするが、これらのより長い断片は、異なる重複する発生元の長い断片からアセンブリングされた重複する長い断片から、より長い範囲の配列情報を推測することも可能にする。この得られたアセンブリは、より長い範囲の配列レベルの特定及び遺伝子融合体1312の特徴付けを可能にする。

【0110】

いくつかの場合では、上述の方法はレトロトランスポゾンの存在の特定に有用である。レトロトランスポゾンは、転写、続いてスプライスしたメッセンジャーRNA (mRNA) の逆転写、及びゲノムの新しい位置への挿入により作ることができる。よって、これらの構造変異体はイントロンを欠如し、多くの場合、染色体間であるが、そうでなければ多様な特徴を有する。レトロトランスポゾンが遺伝子の機能性コピーを導入する場合、それらは、ヒト及びショウジョウバエのゲノムに報告されているレトロ遺伝子と称される。他の場合では、レトロコピー (retrocopies) は、全転写産物、特定の転写産物イソ型、または不完全な転写産物を含有し得る。加えて、代替の転写開始部位及びプロモーター配列は時折、転写産物内に存在し、そのためレトロトランスポゾンは時折、下流配列の発現を導く可能性があるゲノムの再挿入領域内にプロモーター配列を導入する。

【0111】

縦列重複と異なり、レトロトランスポゾンはエクソンまたはイントロン内の親の遺伝子から遠く離れて挿入される。遺伝子の近くに挿入される場合、レトロトランスポゾンは発現のために局所制御配列を利用することができる。遺伝子の近くに挿入することにより、受入れ遺伝子を不活性化するか、または新しいキメラ転写産物を作ることにもできる。レトロトランスポゾン媒介キメラ遺伝子転写産物は、ヒト試料からのRNA-Seqデータにおいて報告されている。

【0112】

レトロトランスポゾンの重要性にも関わらず、それらの検出は、交配対ライブラリからの対合リード補助、全ゲノム配列決定 (WGS) におけるエクソン-エクソン接合部の発見、またはレトロトランスポゾンキメラのRNA-Seq認識に依存する直接的なアプローチに限定され得る。これらの方法のすべては、分析を複雑にする偽陽性を有する可能性がある。

【0113】

レトロトランスポゾンは、本明細書に記載される系及び方法を使用して全ゲノムライブラリから特定することができ、それらの挿入部位は上記に論じられるバーコードマッピングを使用してマッピングすることができる。例えば、Ceph NA12878ゲノムはSKA3-DDX10キメラレトロトランスポゾンを有する。SKA3イントロン無しの転写産物はDDX10のエクソン10と11との間に挿入される。更に、CBX3-C15ORF17レトロトランスポゾンは、本明細書に記載される方法を使用して、NA12

10

20

30

40

50

878においても検出することができる。CBX3のイソ型2はC15ORF17のエクソン2と3との間に挿入される。このキメラ転写産物は、HapMapプロジェクト(D. R. Schrider et al. PLoS Genetics 2013)からのヨーロッパ人RNA-Seq試料の20%において観察されている。

【0114】

レトロトランスポゾン、本明細書に記載される方法及び系を使用して調製された全エクソームライブラリにおいても検出することができる。レトロトランスポゾンはエクソーム標的により容易に富化されるが、イントロンは捕捉中に除去されるため、転座事象とレトロトランスポゾンを区別することは困難であるか、または不可能である。しかしながら、本明細書に記載の系及び方法を使用して、疑われるレトロトランスポゾンのイントロンベイトを導入することにより、全エクソーム配列決定(WES)ライブラリにおいてレトロトランスポゾンを特定することができる(例えば、2014年10月29日に出願された米国仮特許出願第62/072,164号も参照、すべての目的において参照によりその全体が本明細書に組み込まれる)。イントロンシグナルの欠如はレトロトランスポゾン構造変異体を示す一方で、イントロンシグナルは転座を示し得る。

【0115】

理解されるように、上述の変異を特定及び特徴付ける際により長い範囲の配列環境を使用する能力は、バーコードを変異内の領域にマッピングする、及び／または変異に及ぶことにより、挿入、欠失、レトロトランスポゾン、反転などを含む他の構造的変異の範囲を特定するのに同じように適用可能である。

【0116】

V. コピー数変異から生じる疾患及び障害

本方法及び系は、広範囲の疾患及び障害の診断及び／または検出に対して高度に正確かつ感度が高いアプローチを提供する。コピー数変異と関連する疾患は、例えば、ディジョージ/口蓋心臓顔面症候群(22q11.2欠失)、プラダー・ウィリー症候群(15q11-q13欠失)、ウィリアムス・ボイレン症候群(7q11.23欠失)、ミラー・ディーカー症候群(MDLS)(17p13.3微小欠失)、スミス・マゲニス症候群(SMS)(17p11.2微小欠失)、神経線維腫症1型(NF1)(17q11.2微小欠失)、フェラン・マクアーミッド症候群(22q13欠失)、レット症候群(染色体Xq28上のMECP2における機能損失突然変異)、メルツバッハー病(PLP1のCNV)、脊髄性筋萎縮症(SMA)(染色体5q13上のテロメアSMN1のホモ接合体不在)、ポトツキ・ラップスキー症候群(PTLS、染色体17p.11.2の重複)を含み得る。PMP22遺伝子の更なるコピーは、シャルコー・マリー・トゥース病IA型(CMT1A)及び圧迫麻痺性遺伝性ニューロパチー(HNPP)に関連し得る。疾患は、Lupski J. (2007) Nature Genetics 39:S43-S47に記載される疾患であり得る。

【0117】

本明細書に提供される方法及び系は、広範囲の胎児異数体を正確に検出または診断することもできる。多くの場合、本明細書に提供される方法は、試料内の胎児核酸を評価するために、妊婦から採取した試料(例えば血液試料)を分析することを含む。胎児異数体は、例えば、13番染色体トリソミー(パター症候群)、18番染色体トリソミー(エドワーズ症候群)、21番染色体トリソミー(ダウン症候群)、クラインフェルター症候群(XXY)、1つ以上の染色体のモノソミー(X染色体モノソミー、ターナー症候群)、X染色体トリソミー、1つ以上の染色体のトリソミー、1つ以上の染色体のテトラソミーまたはペンタソミー(例えば、XXXX、XXYY、XXXY、XYYY、XXXXX、XXXXY、XXXYY、XYYYY、及びXXYYYY)、三倍体(各染色体が3本、例えばヒトにおいて69染色体)、四倍体(各染色体が4本、例えばヒトにおいて92染色体)、及び高次倍数体(multiploidy)を含み得る。いくつかの実施形態では、異数体は部分異数体であり得る。部分異数体は、例えば、1p36重複、dup(17)(p11.2p11.2)症候群、ダウン症候群、ペリツェウス・メルツバッヘル病、d

up (22) (q11.2q11.2) 症候群、及びネコ眼症候群を含み得る。いくつかの場合では、異常な遺伝子型、例えば胎児遺伝子型は、ネコ鳴き症候群、ウォルフ・ヒルショルン症候群、ウィリアムス・ボイレン症候群、シャルコー・マリー・トゥース病、圧迫麻痺性遺伝性ニューロパチー、スミス・マゲニス症候群、神経線維腫症、アラジール症候群、口蓋心臓顔面症候群、ディジョージ症候群、ステロイドスルファターゼ欠損症、カルマン症候群、線状皮膚欠損を伴う小眼球症、副腎低形成症、グリセロールキナーゼ欠損症、ペリツェウス・メルツバッヘル病、Y染色体短腕上の精巣決定因子、無精子症（a因子）、無精子症（b因子）、無精子症（c因子）、または1p36欠失などの状態をもたらし得る性または常染色体の1つ以上の欠失による。いくつかの実施形態では、染色体数の減少はXO症候群をもたらす。

10

【0118】

過剰なゲノムDNAコピー数変異は、リ・フラウメニ癌素因症候群とも関連する（Shlien et al. (2008) PNAS 105:11264-9）。CNVは、CHARGE（欠損、心奇形、後鼻孔閉鎖、遅滞、性器、及び耳の異常）を含む奇形症候群、ピータース・プラス症候群（Peters-Plus）、ピット・ホプキンス症候群、及び血小板減少-橈骨欠損症候群と関連する（例えば、Ropers HH (2007) Am J of Hum Genetics 81:199-207を参照）。コピー数変異と癌との関係は、例えば、Shlien A. and Malkin D. (2009) Genome Med. 1(6):62に記載されている。コピー数変異は、例えば、自閉症、統合失調症、及び特発性学習障害と関連する。例えば、Sebat J., et al. (2007) Science 316:445-9; Pinto J. et al. を参照。

20

【0119】

本明細書に記載されるように、本明細書に提供される方法及び系は、異なる種類の癌と関連するCNVを検出するのに有用である。例えば、本方法及び系は、非小細胞肺癌において増加し得るEGFRコピー数を検出するために使用され得る。

【0120】

本明細書に提供される方法及び系は、病原体（例えば、ウイルス、細菌、微生物、真菌など）からの感染に対する感受性を含む、特定の疾患または障害に対する対象の感受性のレベルを決定するためにも使用され得る。例えば、比較的高レベルのCCL3L1がHIV感染に対する低感受性と関連することを考えれば、本方法は、CCL3L1のコピー数を分析することにより、対象のHIV感染の感受性を決定するために使用され得る（Gonzalez E. et al. (2005) Science 307:1434-1440）。別の例では、本方法は、全身性エリテマトーデスに対する患者の感受性を決定するために使用され得る。このような場合、例えば、本方法は、FCGR3B（CD16細胞表面免疫グロブリン受容体）のコピー数を決定するために使用され得、これはこの分子の低コピー数が全身性エリテマトーデスに対する増加した感受性と関連するためである（Aitman T. J. et al. (2006) Nature 439:851-855）。本明細書に提供される方法及び系は、自閉症、統合失調症、特発性学習障害と関連するCNVなど、他の疾患または障害と関連するCNVを検出するためにも使用され得る（Kinght et al., (1999) The Lancet 354(9191):1676-81.）。同様に、本方法及び系は、染色体4p16のコピー数可変領域の5つの縦列コピーに関連する常染色体優性小耳症を検出するために使用され得る（Balikova I. (2008) Am J. Hum Genet. 82:181-187）。

30

40

【0121】

VI. 疾患及び障害の検出、診断、及び治療

本明細書に提供される方法及び系は、疾患または障害の検出、診断、及び治療を補助することもできる。いくつかの場合では、方法は、本明細書に記載される系または方法を使用して疾患または障害を検出し、疾患の検出に基づき対象に治療を更に提供することを含

50

む。例えば、癌が検出された場合、対象は、外科的介入により、このような癌を治療するように設計された薬物を投与することにより、ホルモン療法を提供することにより、及び／または放射線もしくはより全身的な化学療法を施すことにより治療され得る。

【0122】

多くの場合、本方法及び系は鑑別診断も可能にし、標的療法で患者を治療することを更に含み得る。一般に、疾患または障害（またはその不在）の鑑別診断は、疾患または障害を有する疑いがある対象から得た試料核酸の配列を決定及び特徴付けし、それを、障害または疾患状態の存在（または不在）を示す参照核酸の配列及び／または配列の特徴付けと比較することにより、障害または疾患状態（またはその不在）の指標として試料核酸を更に特徴付けることにより達成され得る。

【0123】

参照核酸配列は、疾患または障害の不在を示すゲノム（例えば生殖系列核酸）から得ることができるか、または疾患または障害状態を示すゲノム（例えば、癌核酸、異数体を示す核酸など）から得ることができる。更に、参照核酸配列（例えば、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、または更には100 kb より長い長さを有する）は、非限定的な例として、特定の配列の存在（または不在）の決定、特定のハプロタイプの存在（または不在）の決定、1つ以上の遺伝的変異（例えば、構造的変異（例えばコピー数変異、挿入、欠失、転座、反転、レトロトランスポゾン、再編成、反復配列伸長、重複など）、一塩基多型（SNP）など）の存在（または不在）の決定、及びこれらの組み合わせを含む、1つ以上の点で特徴付けされ得る。更に、参照配列の任意の好適な種類及び数の配列特徴が、試料核酸の配列を特徴付けるために使用され得る。例えば、参照核酸配列の1つ以上の遺伝的変異（もしくはその欠如）または構造的変異（もしくはその欠如）は、障害または疾患状態の存在（または不在）の指標として参照核酸を特定するための配列記号として使用され得る。利用される参照核酸配列の特徴付けに基づき、試料核酸配列は類似する方法で特徴付けされ、参照核酸配列に類似する形質を表すか否かに基づき、障害または疾患を示す核酸に由来する（または由来しない）として更に特徴付け／特定され得る。いくつかの場合では、試料核酸配列及び／または参照核酸配列の特徴付けならびにそれらの比較は、プログラムされたコンピュータプロセッサの補助により完了し得る。いくつかの場合では、このようなプログラムされたコンピュータプロセッサは、本明細書の他の箇所に記載される例示的なコンピュータ制御システムなどのコンピュータ制御システムに含まれてよい。

【0124】

試料核酸は、本明細書の他の箇所に記載される試料供給源及び生物学的試料供給源を含む、任意の好適な供給源から得ることができる。いくつかの場合では、試料核酸は、無細胞核酸を含み得る。いくつかの場合では、試料核酸は、腫瘍核酸（例えば腫瘍DNA）を含み得る。いくつかの場合では、試料核酸は、循環腫瘍核酸（例えば循環腫瘍DNA（ctDNA））を含み得る。循環腫瘍核酸は、循環腫瘍細胞（CTC）に由来してもよく、かつ／または例えば対象の血液、血漿、他の体液、または組織から得てもよい。

【0125】

図20～21は、疾患の検出及び診断の文脈における試料核酸を特徴付けるための例示的な方法を図示する。図20は、長い範囲の配列環境が、例えば図6に示されるものに類似する方法などで、より短いバーコード化された断片から参照核酸（例えば生殖系列核酸（例えば生殖系列ゲノムDNA）、特定の障害または疾患状態と関連する核酸）に関して決定され得る例示的な方法を示す。図20に関して、参照核酸を得ることができ2000、一組のバーコード化されたビーズも得ることができる2010。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出す

るために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例では、参照核酸2005、バーコード化されたビーズ2015、及びいくつかの場合では、他の試薬、例えば還元剤2020が組み合わせられ、分配が行われる。いくつかの場合では、参照核酸2000は、分配前に断片化されてよく、得られた断片の少なくとも一部はバーコード化のために2005として分配される。例として、このような分配は、マイクロ流体デバイス2025などの液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス2025の補助により、油中水型エマルジョン2030が形成され得、エマルジョンは、参照核酸2005、還元剤2020、及びバーコード化されたビーズ2015を含有する水性液滴を含有している。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる2035。ランダムN-merは次に、参照核酸の異なる領域をプライムし、増幅後に参照核酸の増幅されたコピーをもたらすことができ、各コピーはバーコード配列でタグ付けされている2040。いくつかの場合では、増幅2040は、本明細書の他の箇所に記載されるものと類似する方法によって達成され得、図5に模式的に示される。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダムN-mer配列を含有する一組のオリゴヌクレオチドを含有する。続いて、エマルジョンを破壊し2045、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法2050（例えばPCR）を介して付加され得る。次に、配列決定が実施され得2055、配列決定データを解釈するためにアルゴリズムが適用される2060。いくつかの場合では、配列決定データの解釈2060は、参照核酸の少なくとも一部の配列を提供することを含み得る。いくつかの場合では、参照核酸の長い範囲の配列環境は、例えば、参照核酸が疾患状態（例えば、本明細書の他の箇所に記載されるように1つ以上のハプロタイプの決定、1つ以上の構造的変異（例えばコピー数変異、挿入、欠失、転座、反転、再編成、反復配列伸長、重複、レトロトランスポゾン、遺伝子融合体など）の決定、1つ以上のSNPの呼び出しなど）に由来する場合などにおいて得られ、特徴付けされる。いくつかの場合では、変異体は、図7に関して本明細書の他の箇所に記載されるなど、より長い範囲の配列環境を提供するために生成された供給源及び推測コンティグから得た様々な参照核酸について呼び出されてよい。

【0126】

図21は、図20に示されるように得た参照の2060特徴付けから試料核酸配列を特徴付ける例を示す。長い範囲の配列環境は、例えば図6に模式的に示される方法などを介して、本明細書の他の箇所に記載されるように、より短いバーコード化された断片の配列決定から試料核酸に関して得ることができる。図21に示されるように、核酸試料（例えば循環腫瘍核酸を含む試料）は、障害または疾患（例えば癌）を有する疑いがある対象から得ることができ2100、一組のバーコード化されたビーズも得ることができる2110。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出するために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例では、試料核酸2105、バーコード化されたビーズ2115、及びいくつかの場合では、他の試薬、例えば還元剤2120が組み合わせられ、分配が行われる。いくつかの場合では、胎児試料2100が分配前に断片化され、得られた断片の少なくとも一部はバーコード化のために2105として分配される。例として、このような分配は、マイクロ流体デバイス2125などの液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス2125の補助により、油中水型エマルジ

ョン 2 1 3 0 が形成され得、エマルジョンは、試料核酸 2 1 0 5、還元剤 2 1 2 0、及びバーコード化されたビーズ 2 1 1 5 を含有する水性液滴を含有している。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる 2 1 3 5。ランダム N - m e r は次に、試料核酸の異なる領域をプライムし、増幅後に試料核酸の増幅されたコピーをもたらしことができ、各コピーはバーコード配列でタグ付けされている 2 1 4 0。いくつかの場合では、増幅 2 1 4 0 は、本明細書の他の箇所に記載されるものと類似する方法によって達成され得、図 5 に模式的に示される。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダム N - m e r 配列を含有する一組のオリゴヌクレオチドを含有する。続いて、エマルジョンを破壊し 2 1 4 5、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法 2 1 5 0（例えば P C R）を介して付加され得る。次に、配列決定が実施され得 2 1 5 5、配列決定データを解釈するためにアルゴリズムが適用される 2 1 6 0。いくつかの場合では、配列決定データの解釈 2 1 6 0 は、試料核酸の配列を提供することを含み得る。いくつかの場合では、核酸試料の長い範囲の配列環境が得られる。試料核酸配列は、参照核酸配列の特徴付け 2 0 6 0 を使用して、特徴付けされ得る 2 1 6 0（例えば、本明細書の他の箇所に記載されるように 1 つ以上のハプロタイプの決定、1 つ以上の構造的変異（例えば、コピー数変異、挿入、欠失、転座、反転、再編成、反復配列伸長、重複、レトロトランスポゾン、遺伝子融合体など）の決定。試料核酸配列及びその特徴付けの、参照核酸の配列及び特徴付けとの比較に基づき、障害または疾患状態の存在（または不在）に関する鑑別診断 2 1 7 0 を行うことができる。

【0 1 2 7】

理解されるように、参照核酸及び試料核酸の分析は別個の分配分析として完了することができるか、または単一の分配分析の一部として完了することができる。例えば、図 2 0 及び 2 1 に従って、試料及び参照核酸は、同じデバイスに添加され、バーコード化された試料及び参照断片は液滴において生成され、エマルジョンは両種類の核酸の液滴を含んでいる。次に、本明細書の他の箇所に記載されるように、エマルジョンは破壊され、液滴の内容物がプールされ、更に処理され（例えば P C R を介した追加の配列のバルク添加）、配列決定され得る。バーコード化された断片からの個々の配列決定リードは、バーコード配列を介してそれらのそれぞれの試料配列に帰属され得る。試料核酸から得た配列は、参照核酸配列の特徴付けに基づき特徴付けされ得る。

【0 1 2 8】

本明細書の方法及び系を利用することにより、本明細書に記載されるように、参照及び試料核酸配列の長い範囲の配列環境を含む、核酸の長い範囲の配列環境を決定する際の精度を改善することができる。本明細書に提供される方法及び系は、少なくとも 7 0 %、8 0 %、8 5 %、9 0 %、9 1 %、9 2 %、9 3 %、9 4 %、9 5 %、9 9 %、9 9 . 1 %、9 9 . 2 %、9 9 . 3 %、9 9 . 4 %、9 9 . 5 %、9 9 . 6 %、9 9 . 7 %、9 9 . 8 %、9 9 . 9 %、9 9 . 9 5 %、9 9 . 9 9 %、9 9 . 9 9 5 %、または 9 9 . 9 9 9 % の精度で参照及び／または試料核酸の長い範囲の配列環境を決定することができる。いくつかの場合では、本明細書に提供される方法及び系は、1 0 %、9 %、8 %、7 %、6 %、5 %、4 %、3 %、2 %、1 %、0 . 1 %、0 . 0 5 %、0 . 0 1 %、0 . 0 0 5 %、0 . 0 0 1 %、0 . 0 0 0 5 %、0 . 0 0 0 1 %、0 . 0 0 0 0 5 %、0 . 0 0 0 0 1 %、または 0 . 0 0 0 0 0 5 % 未満の誤差率で参照及び／または試料核酸の長い範囲の配列環境を決定することができる。

【0 1 2 9】

更に、本明細書の方法及び系は、1 つ以上の態様において、参照核酸配列及び／または試料核酸配列を特徴付ける（例えば、配列の決定、1 つ以上の遺伝的変異の決定、ハプロタイプの決定など）際の精度も改善することができる。したがって、本明細書に提供される方法及び系は、少なくとも 7 0 %、8 0 %、8 5 %、9 0 %、9 1 %、9 2 %、9 3 %、9 4 %、9 5 %、9 9 %、9 9 . 1 %、9 9 . 2 %、9 9 . 3 %、9 9 . 4 %、9 9 .

5 %、99.6 %、99.7 %、99.8 %、99.9 %、99.95 %、99.99 %、99.995 %、または99.999 %の精度で、1つ以上の態様において、参照核酸配列及び／または試料核酸配列を特徴付けることができる。いくつかの場合では、本明細書に提供される方法及び系は、10 %、9 %、8 %、7 %、6 %、5 %、4 %、3 %、2 %、1 %、0.1 %、0.05 %、0.01 %、0.005 %、0.001 %、0.0005 %、0.0001 %、0.00005 %、0.00001 %、または0.000005 %未満の誤差率で、1つ以上の態様において、参照核酸配列及び／または試料核酸配列を特徴付けることができる。

【0130】

更に、上記に論じられるように、参照核酸の長い範囲の配列環境及びその特徴付けを決定する際の改善された精度は、試料核酸の配列決定及び特徴付け、ならびに後続の障害または疾患の鑑別診断における使用における改善された精度をもたらし得る。したがって、試料核酸配列（長い範囲の配列環境を含む）は、10 %、9 %、8 %、7 %、6 %、5 %、4 %、3 %、2 %、1 %、0.1 %、0.05 %、0.01 %、0.005 %、0.001 %、0.0005 %、0.0001 %、0.00005 %、0.00001 %、または0.000005 %未満の誤差率で、参照核酸配列の分析から提供され得る。いくつかの場合では、試料核酸配列は、少なくとも70 %、80 %、85 %、90 %、91 %、92 %、93 %、94 %、95 %、99 %、99.1 %、99.2 %、99.3 %、99.4 %、99.5 %、99.6 %、99.7 %、99.8 %、99.9 %、99.95 %、99.99 %、99.995 %、または99.999 %の精度で、参照核酸の配列と比較し、かつ／またはその配列を特徴付けることにより、障害または疾患（またはその不在）の鑑別診断に使用され得る。いくつかの場合では、試料核酸配列は、10 %、9 %、8 %、7 %、6 %、5 %、4 %、3 %、2 %、1 %、0.1 %、0.05 %、0.01 %、0.005 %、0.001 %、0.0005 %、0.0001 %、0.00005 %、0.00001 %、または0.000005 %未満の誤差率で、参照核酸の配列と比較し、かつ／またはその配列を特徴付けることにより、障害または疾患（またはその不在）の鑑別診断に使用され得る。

【0131】

例において、本方法及び系は、肺癌がEGFR遺伝子の変異と関連する非小細胞肺癌であるかを決定するために、肺癌の患者におけるコピー数変異を検出するために使用され得る。このような診断後、患者の治療レジメンは、鑑別診断と相関するように洗練され得る。標的療法または分子による標的療法は、癌の主な医療行為の様式（薬物療法）のうちの1つであり、他はホルモン療法及び細胞毒性化学療法である。標的療法は、すべての迅速に分裂する細胞に単純に干渉するよりも（例えば従来の化学療法）、むしろ発癌及び腫瘍成長に必要とされる特定の標的分子に干渉することにより癌細胞の成長を遮断する。

【0132】

図14は、非小細胞肺癌を鑑別診断するための例示的なプロセスを示す。慢性咳、体重減少、及び息切れを有する患者が肺癌について検査される1400。血液を患者から得1405、試料（例えば、循環腫瘍細胞、無細胞DNA、循環核酸（例えば、循環腫瘍核酸）など）は血液に由来する1410。一組のバーコード化されたビーズも得ることができる1415。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出するために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例では、試料1410、バーコード化されたビーズ1420、及びいくつかの場合では、他の試薬、例えば還元剤が組み合わせられ、分配が行われる。例として、このような分配は、マイクロ流体デバイス1425などの液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス1425の補助に

より、油中水型エマルジョン 1430 が形成され得、エマルジョンは、試料核酸 1410、バーコード化されたビーズ 1415、及びいくつかの場合では、還元剤を含有する水性液滴を含有している。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる 1435。ランダム N-mer は次に、試料核酸の異なる領域をプライムし、増幅後に試料の増幅されたコピーをもたらすことができ、各コピーはバーコード配列でタグ付けされている 1440。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダム N-mer 配列を含有する一組のオリゴヌクレオチドを含有する。続いて、エマルジョンを破壊し 1445、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法（例えば PCR）を介して付加され得る。次に、配列決定が実施され得 1450、配列決定データを解釈するためにアルゴリズムが適用される 1455。配列決定アルゴリズムは一般的に、例えば、配列決定リードを整合し、かつ／または特定の配列リードが属する試料を特定するために、バーコードの分析を実施することができる。

【0133】

次に、異なる遺伝子の CNV を決定するために、分析した配列を既知のゲノム参照配列と比較する 1460。DNA における EGFR コピー数が通常より高い場合、患者は、小細胞肺癌の代わりに非小細胞肺癌（NSCLC）と鑑別診断され得る 1465。非小細胞肺癌の CTC は小細胞肺癌と更に区別され得る他のコピー数変異も有する。癌の病期により、外科手術、化学療法、または放射線療法が処方される 1470。いくつかの場合では、NSCLC と診断された患者は、ALK 阻害剤（例えばクリゾチニブ）などの、このような癌を標的とする薬物を投与される。EGFR における変異のいくつかの場合では、患者は、セツキシマブ、パニツムマブ、ラパチニブ、及び／またはカペシタビンを投与される。異なる状況では、標的は ERBB2 などの異なる遺伝子であってよく、療法はトラスツズマブ（ハーセプチン）を含む。（2010）Nature 466:368-72; Cook E. H. and Scherer S. W. (2008) Nature 455:919-923。

【0134】

標的療法の主なカテゴリーは、小分子、小分子薬物複合体、及びモノクローナル抗体である。小分子は、メシル酸イマチニブ（グリベック、STI-571 としても知られる）（慢性骨髄性白血病、消化管間質腫瘍、及び一部の他の種類の癌に認可されている）、ゲフィチニブ（イレッサ、ZD1839 としても知られる）（上皮成長因子受容体（EGFR）チロシンキナーゼを標的とし、小細胞肺癌に対して米国で認可されている）、エルロチニブ（タルセバとして販売されている）、ボルテゾミブ（ベルケイド）（タンパク質に干渉することにより癌細胞に細胞死を起こさせるアポトーシス誘導プロテアソーム阻害剤薬物である）、タモキシフェン、JAK 阻害剤（例えばトファシチニブ）、ALK 阻害剤（例えばクリゾチニブ）、Bcl-2 阻害剤（例えば臨床試験中のオパトクラックス、ABT-263、及びゴシポール）、PARP 阻害剤（例えばイニパリブ、臨床試験中のオラパリブ）、PI3K 阻害剤（例えば第 III 相試験中のペリホシン）、アパチニブ（選択的 VEGF 受容体 2 阻害剤である）、[D-Lys(6)]-LHRH に連結された AN-152（AEZS-108）ドキシソルビシン、Braf 阻害剤（ベムラフェニブ、ダブラフェニブ、LGX818）（BRAF V600E 突然変異を保有する転移性黒色腫を治療するために使用される）、MEK 阻害剤（トラメチニブ、MEK162）、CDK 阻害剤、例えば D-0332991、臨床試験中の LEE011、Hsp90 阻害剤、及びサリノマイシンなどのチロシンキナーゼ阻害剤を含み得る。

【0135】

他の療法は、葉酸受容体を標的とする小分子の小分子薬物複合体であるビンタフォリドなどの小分子薬物複合体を含む。モノクローナル抗体は本明細書に提供される方法の一部として投与され得る別の種類の療法である。モノクローナル薬物複合体も投与され得る。例示的なモノクローナル抗体は、

10

20

30

40

50

リツキシマブ（M a b T h e r aまたはリツキサンとして販売されている）（B細胞上に見られるC D 2 0を標的とし、非ホジキンリンパ腫を標的とする）、トラスツズマブ（ハーセプチン）（一部の種類の乳癌に発現されるH e r 2 / n e u（E r b B 2としても知られる）受容体を標的とする）、セツキシマブ（アービタックスとして販売されている）、及びパニツムマブ、ベバシズマブ（アバスチンとして販売されている）（V E G Fリガンドを標的とする）を含む。

【0136】

V I I . 親の核酸からの胎児核酸の特徴付け

本明細書の他の箇所に記述されるように、本明細書に記載される方法及び系は、対象の血液または血漿内の循環核酸を特徴付けするためにも使用され得る。このような分析は、患者における潜在的な疾患状態の特定に使用するための循環腫瘍D N A、または非侵襲的な方法で、例えば羊水穿刺または他の侵襲性手順を通した直接試料採取に頼ることなく胎児D N Aを特徴付けるために、妊娠している女性の血液または血漿内の循環胎児D N Aの分析を含む。

【0137】

いくつかの場合では、本方法は、少なくとも部分的に親の核酸配列の分析に基づき、胎児核酸配列、例えば、循環胎児D N Aを特徴付けるために使用され得る。例えば、長い範囲の配列環境は、本明細書に記載される方法及び系を使用して、より短いバーコード化された断片から父方及び母方の両方の核酸（例えば、1 k bより長い、5 k bより長い、1 0 k bより長い、1 5 k bより長い、2 0 k bより長い、3 0 k bより長い、4 0 k bより長い、5 0 k bより長い、6 0 k bより長い、7 0 k bより長い、8 0 k bより長い、9 0 k bより長い、または更には1 0 0 k bより長い長さを有する）に関して決定され得る。長い範囲の配列環境は、父方及び母方核酸配列の両方における一塩基多型（S N P）、構造的変異（例えば、コピー数変異、挿入、欠失、転座、反転、再編成、反復配列伸長、レトロトランスポゾン、重複、遺伝子融合など）を含む、1つ以上のハプロタイプ及び1つ以上の遺伝的変異を決定するために使用され得る。更に、父方及び母方核酸の長い範囲の配列環境ならびに任意の決定されたS N P、ハプロタイプ、及び／または構造的変異情報は、妊娠中の母親から得た胎児核酸（例えば、例えば無細胞胎児核酸などの循環胎児核酸）の配列を特徴付けるために使用され得る。いくつかの場合では、母方及び父方の配列との比較及び特徴付けによる胎児核酸の特徴付けは、プログラムされたコンピュータプロセッサの補助により完了することができる。いくつかの場合では、このようなプログラムされたコンピュータプロセッサは、本明細書の他の箇所に記載される例示的なコンピュータ制御システムなどのコンピュータ制御システムに含まれてよい。

【0138】

例えば、親及び／もしくは母方核酸の配列ならびに／または長い範囲の配列環境は、胎児核酸配列を含む胎児核酸を特徴付けることにより参照として使用され得る。実際、本明細書に記載される方法及び系により得られた長い範囲の配列環境は、胎児核酸配列が特徴付けされ得る父方及び母方核酸に関する改善された長い範囲の配列環境情報を提供し得る。いくつかの場合では、参照としての親の核酸からの胎児核酸配列の特徴付けは、胎児核酸の配列において、胎児核酸の少なくとも一部の配列を決定する、かつ／または胎児核酸配列の1つ以上のS N Pを呼び出す、胎児核酸配列の1つ以上の新しい突然変異を決定する、胎児核酸配列の1つ以上のハプロタイプを決定する、ならびに／または1つ以上の構造的変異を決定及び特徴付けるなどを含み得る。

【0139】

図17～19は、より短いバーコード化された断片の配列決定を介して、父方及び母方核酸に関して得られたより長い範囲の配列環境から胎児核酸を特徴付けるための例示的な方法を図示する。図17は、例えば、図6に示されるものに類似する方法などで、より短いバーコード化された断片からより長い範囲の配列環境が父方核酸試料（例えば、父方のゲノムD N A）に関して決定され得る例示的な方法を示す。図17に関して、父方核酸を含む試料は胎児の父親から得ることができ1700、一組のバーコード化されたビーズも

得ることができる1710。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出するために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例では、核酸を含む父方の試料1705、バーコード化されたビーズ1715、及びいくつかの場合では、他の試薬、例えば還元剤1720が組み合わせられ、分配が行われる。いくつかの場合では、父方の試料1700は分配前に断片化され、得られた断片の少なくとも一部はバーコード化のために1705として分配される。例として、このような分配は、マイクロ流体デバイス1725などの液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス1725の補助により、油中水型エマルジョン1730が形成され得、エマルジョンは、父方の試料核酸1705、還元剤1720、及びバーコード化されたビーズ1715を含有する水性液滴を含有している。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる1735。ランダムN-merは次に、父方の試料核酸の異なる領域をプライムし、増幅後に父方試料の増幅されたコピーをもたらすことができ、各コピーはバーコード配列でタグ付けされている1740。いくつかの場合では、増幅1740は、本明細書の他の箇所に記載されるものと類似する方法によって達成され得、図5に模式的に示される。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダムN-mer配列を含有する一組のオリゴヌクレオチドを含有する。続いて、エマルジョンを破壊し1745、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法1750（例えばPCR）を介して付加され得る。次に、配列決定が実施され得1755、配列決定データを解釈するためにアルゴリズムが適用される1760。いくつかの場合では、例えば、配列決定データの解釈1760は、父方核酸の少なくとも一部の配列を提供することを含み得る。いくつかの場合では、父方核酸試料の長い範囲の配列環境が得られ、特徴付けされ得る（例えば、本明細書の他の箇所に記載される1つ以上のハプロタイプの決定、1つ以上の構造的変異（例えば、コピー数変異、挿入、欠失、転座、反転、再編成、反復配列伸長、重複、レトロトランスポゾン、遺伝子融合など）の決定、1つ以上のSNPの呼び出し、1つ以上の他の遺伝的変異の決定など）。いくつかの場合では、変異体は、図7に関して本明細書の他の箇所に記載されるなど、より長い範囲の配列環境を提供するために生成された様々な父方核酸及び推測コンティグにいついて呼び出されてよい。

【0140】

図18は、例えば、図6に示されるものに類似する方法などで、より短いバーコード化された断片から長い範囲の配列環境が母方核酸試料（例えば、母方のゲノムDNA）に関して決定され得る例示的な方法を示す。図18に関して、母方核酸を含む試料は胎児の妊娠中の母親から得ることができる1800、一組のバーコード化されたビーズも得ることができる1810。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出するために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例では、核酸を含む母方の試料1805、バーコード化されたビーズ1815、及びいくつかの場合では、他の試薬、例えば還元剤1820が組み合わせられ、分配が行われる。いくつかの場合では、母方の試料1800は分配前に断片化され、得られた断片の少なくとも一部はバーコード化のために1805として分配される。例として、このような分配は、マイクロ流体デバイス1825などの

液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス1825の補助により、油中水型エマルジョン1830が形成され得、エマルジョンは、母方の試料核酸1805、還元剤1820、及びバーコード化されたビーズ1815を含有する水性液滴を含有している。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる1835。ランダムN-merは次に、母方の試料核酸の異なる領域をプライムし、増幅後に母方試料の増幅されたコピーをもたらすことができ、各コピーはバーコード配列でタグ付けされている1840。いくつかの場合では、増幅1840は、本明細書の他の箇所に記載されるものと類似する方法によって達成され得、図5に模式的に示される。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダムN-mer配列を含有する一組のオリゴヌクレオチドを含有する。続いて、エマルジョンを破壊し1845、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法1850（例えばPCR）を介して付加され得る。次に、配列決定が実施され得1855、配列決定データを解釈するためにアルゴリズムが適用される1860。いくつかの場合では、例えば、配列決定データの解釈1860は、母方核酸の少なくとも一部の配列を提供することを含み得る。いくつかの場合では、母方核酸試料の長い範囲の配列環境が得られ、特徴付けされ得る（例えば、本明細書の他の箇所に記載される1つ以上のハプロタイプの決定、1つ以上の構造的変異（例えば、コピー数変異、挿入、欠失、転座、反転、再編成、反復配列伸長、重複、レトロトランスポゾン、遺伝子融合など）の決定、1つ以上のSNPの呼び出し、1つ以上の他の遺伝的変異の決定など。いくつかの場合では、変異体は、図7に関して本明細書の他の箇所に記載されるなど、より長い範囲の配列環境を提供するために生成された試料及び推測コンティグから得た様々な母方核酸について呼び出されてよい。

【0141】

図19は、それぞれ、図17及び図18に示されるように得た父方1760及び母方1860の特徴付けから胎児試料配列を特徴付ける例を示す。図19に示されるように、胎児核酸試料は、妊娠中の母親から得ることができる1900。長い範囲の配列環境は、例えば図6に模式的に示される方法などを介して、本明細書の他の箇所に記載されるように、より短いバーコード化された断片の配列決定から胎児核酸について得ることができる。いくつかの場合では、胎児核酸試料は、例えば妊娠中の母親の血液、血漿、他の体液、または組織から得ることができる循環胎児DNA及び／または無細胞DNAであってよい。一組のバーコード化されたビーズも得ることができる1910。ビーズは、1つ以上のバーコード配列、ならびにランダムN-merまたは他のプライマーなどのプライマーを含有するオリゴヌクレオチドに連結され得る。いくつかの場合では、バーコード配列は、例えば、バーコードとビーズとの間の連結の切断を通して、もしくはバーコードを放出するために下部のビーズの分解を通して、またはこの2つの組み合わせで、バーコード化されたビーズから放出可能である。例えば、いくつかの態様では、バーコード化されたビーズは、バーコード配列を放出するために、還元剤などの薬剤により分解または溶解され得る。この例では、核酸を含む胎児試料1905、バーコード化されたビーズ1915、及びいくつかの場合では、他の試薬、例えば還元剤1920が組み合わせられ、1905として分配が行われる。いくつかの場合では、胎児試料1900は分配前に断片化され、得られた断片の少なくとも一部はバーコード化のために1905として分配される。例として、このような分配は、マイクロ流体デバイス1925などの液滴生成系に成分を導入することを伴う場合がある。マイクロ流体デバイス1925の補助により、油中水型エマルジョン1930が形成され得、エマルジョンは、母方の試料核酸1905、還元剤1920、及びバーコード化されたビーズ1915を含有する水性液滴を含有している。還元剤は、バーコード化されたビーズを溶解または分解し、それにより液滴内のビーズから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放出することができる1935。ランダムN-merは次に、胎児試料核酸の異なる領域をプライムし、増幅後に胎児試料の増幅されたコピーをもたらすことができ、各コピーはバーコード配列でタグ付けされ

10

20

30

40

50

ている1940。いくつかの場合では、増幅1940は、本明細書の他の箇所に記載されるものと類似する方法によって達成され得、図5に模式的に示される。いくつかの場合では、各液滴は、同一のバーコード配列及び異なるランダムN-mer配列を含有する一組のオリゴヌクレオチドを含有する。続いて、エマルジョンを破壊し1945、更なる配列（例えば、特定の配列決定法に役立つ配列、追加のバーコードなど）が、例えば増幅法1950（例えばPCR）を介して付加され得る。次に、配列決定が実施され得1955、配列決定データを解釈するためにアルゴリズムが適用される1960。一般に、胎児核酸試料のより長い範囲の配列環境は、配列決定されるより短いバーコード化された断片から得ることができる。いくつかの場合では、例えば、配列決定データの解釈1960は、胎児核酸の少なくとも一部の配列を提供することを含み得る。胎児核酸配列は、長い範囲の配列環境ならびに／または父方1760及び母方1860試料の特徴付けを使用して、特徴付けされ得る1960（例えば、本明細書の他の箇所に記載されるように1つ以上のハプロタイプの決定、1つ以上の構造的変異（例えば、コピー数変異、挿入、欠失、転座、反転、再編成、反復配列伸長、重複、レトロトランスポゾン、遺伝子融合体など）の決定、1つ以上の新しい突然変異の決定、1つ以上のSNPの呼び出しなど）。いくつかの場合では、胎児核酸のフェーズブロックは、胎児核酸配列を母方及び父方のフェーズブロックと比較することにより決定され得る。

【0142】

理解されるように、父方核酸、母方核酸、及び／または胎児核酸の分析は、別個の分配分析の一部として完了することができるか、または1つ以上の組み合わせられた分配分析の一部として完了することができる。例えば、図17～19に従って、父方、母方、及び胎児の核酸は、同じデバイスに添加され、バーコード化された父方、母方、及び胎児の断片は液滴において生成され、エマルジョンは3つの種類の核酸の液滴を含んでいる。次に、本明細書の他の箇所に記載されるように、エマルジョンは破壊され、液滴の内容物がプールされ、更に処理され（例えばPCRを介した追加の配列のバルク添加）、配列決定され得る。バーコード化された断片からの個々の配列決定リードは、バーコード配列を介してそれらのそれぞれの試料配列に帰属され得る。

【0143】

いくつかの場合では、胎児ゲノムの配列を含む胎児核酸の配列、及び／または胎児核酸配列における遺伝的変異は、本明細書に記載される方法及び系を使用して得た長い範囲の父方及び母方の配列環境及び特徴付けから決定され得る。例えば、循環胎児核酸の配列決定と共に、父方及び母方ゲノムのゲノム配列決定は、対応する胎児ゲノム配列を決定するために使用され得る。親のゲノム及び無細胞胎児核酸の配列分析からゲノム胎児核酸の配列を決定する例は、Kitzman et al. (2012 June 6) Sci Transl. Med. 4 (137): 137ra76に見出すことができ、参照により全体が本明細書に組み込まれる。胎児ゲノムの決定は、例えば胎児異数体を含む、胎児のゲノム障害の出生前決定及び診断に有用である。本明細書の他の箇所に論じられるように、本明細書に提供される方法及び系は、核酸配列におけるハプロタイプを解明する際に有用であり得る。ハプロタイプが解明された父方及び母方の配列は、父方及び母方のサンプル試料核酸配列に対して決定され得、そのそれぞれは、胎児ゲノムの配列の決定及び／またはその特徴付けをより正確に行うことに役立ち得る。

【0144】

本明細書の方法及び系を利用することにより、親の核酸配列（例えば、母方核酸配列、父方核酸配列）の長い範囲の配列環境を含む、核酸の長い範囲の配列環境を決定する際の精度を改善することができる。本明細書に提供される方法及び系は、少なくとも70%、80%、85%、90%、91%、92%、93%、94%、95%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%、99.95%、99.99%、99.995%、または99.999%の精度で、親の核酸の長い範囲の配列環境を決定することができる。いくつかの場合では、本明細書に提供される方法及び系は、10%、9%、8%、7%、6%、5%、4

%、3%、2%、1%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%、0.0001%、0.00005%、0.00001%、または0.000005%未満の誤差率で、親の核酸の長い範囲の配列環境を決定することができる。更に、本明細書の方法及び系は、1つ以上の態様において、父方核酸配列を特徴付ける（例えば、配列の決定、1つ以上の遺伝的変異の決定、1つ以上の構造変異体の決定、ハプロタイプの決定など）際の精度も改善することができる。したがって、本明細書に提供される方法及び系は、少なくとも70%、80%、85%、90%、91%、92%、93%、94%、95%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%、99.95%、99.99%、99.995%、または99.999%の精度で、1つ以上の態様において、父方核酸配列を特徴付けることができる。いくつかの場合では、本明細書に提供される方法及び系は、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%、0.0001%、0.00005%、0.00001%、または0.000005%未満の誤差率で、1つ以上の態様において、親の核酸配列を特徴付けることができる。

10

【0145】

更に、上記に論じられるように、親の核酸の長い範囲の配列環境及びその特徴付けを決定する際の改善された精度は、胎児核酸の配列決定及び特徴付けにおいて改善された精度をもたらし得る。したがって、いくつかの場合では、胎児核酸配列（長い範囲の配列環境を含む）は、少なくとも70%、80%、85%、90%、91%、92%、93%、94%、95%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%、99.95%、99.99%、99.995%、または99.999%の精度で、親の核酸配列の分析から提供され得る。いくつかの場合では、胎児核酸配列（長い範囲の配列環境を含む）は、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%、0.0001%、0.00005%、0.00001%、または0.000005%未満の誤差率で、親の核酸配列の分析から提供され得る。いくつかの場合では、胎児核酸配列は、1つ以上の態様において、少なくとも70%、80%、85%、90%、91%、92%、93%、94%、95%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%、99.95%、99.99%、99.995%、または99.999%の精度で、本明細書に記載される親の核酸配列の分析（例えば、配列の決定、1つ以上の遺伝的変異の決定、1つ以上の構造的変異の決定、ハプロタイプの決定など）を介して特徴付けされ得る。いくつかの場合では、胎児核酸配列は、1つ以上の態様において、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%、0.0001%、0.00005%、0.00001%、または0.000005%未満の誤差率で、本明細書に記載される親の核酸配列の分析（例えば、配列の決定、1つ以上の遺伝的変異の決定、ハプロタイプの決定、1つ以上の構造的変異の決定など）を介して特徴付けされ得る。

20

30

40

【0146】

V I I I . 試料

疾患または障害の検出は、患者から試料を得ることによって開始し得る。本明細書で使用されるとき、「試料」という用語は、一般的に、生物学的試料を指す。生物学的試料の例としては、核酸分子、アミノ酸、ポリペプチド、タンパク質、炭水化物、脂肪、またはウイルスを含む。例において、生物学的試料は、1つ以上の核酸分子を含む核酸試料である。例示的な試料としては、ポリヌクレオチド、核酸、オリゴヌクレオチド、無細胞核酸（例えば、無細胞DNA（cfDNA））、循環無細胞核酸、循環腫瘍核酸（例えば、循環腫瘍DNA（ctDNA））、循環腫瘍細胞（CTC）核酸、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）、2本鎖DNA（d

50

sDNA)、1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などを挙げることができる。要約すると、使用される試料は、特定の処理ニーズにより変動し得る。

【0147】

核酸を含むいずれ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。流体物質は、血液、臍帯血、唾液、尿、汗、血清、精液、腔液、胃液及び消化液、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を含み得るが、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、正常な健康な組織、疾患組織、または健康な組織と疾患組織の混合を含み得る。いくつかの場合では、物質は腫瘍を含み得る。腫瘍は、良性(非癌性)または悪性(癌)であり得る。腫瘍の非限定的な例としては、線維肉腫、粘液肉腫、脂肪肉腫、軟骨肉腫、骨肉腫、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、ユーイング肉腫、平滑筋肉腫、横紋筋肉腫、胃腸管系細胞癌、結腸細胞癌、膀胱癌、乳癌、泌尿生殖器系細胞癌、卵巣癌、前立腺癌、扁平上皮癌、基底細胞癌、腺癌、汗腺癌、脂腺癌、乳頭癌、乳頭状腺癌、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、セミノーマ、胚性癌腫、ウィルムス腫瘍、子宮頸癌、内分泌系細胞癌、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、髄芽腫、頭蓋咽頭腫、上衣腫、松果体腫、血管芽腫、聴神経腫瘍、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を挙げることができる。物質は様々な種類の臓器と関連し得る。臓器の非限定的な例としては、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、喉頭、食道、胃、またはこれらの組み合わせを挙げることができる。いくつかの場合では、物質は、真核細胞、原核細胞、菌細胞、心臓細胞、肺細胞、腎細胞、肝細胞、膵細胞、生殖細胞、幹細胞、人工多能性幹細胞、胃腸細胞、血液細胞、癌細胞、菌体、ヒトのマイクロバイオーーム試料から単離した菌体などを含むが、これらに限定されない様々な細胞を含み得る。いくつかの場合では、物質は、例えば、単一細胞の内容物または複数細胞の内容物などの細胞の内容物を含み得る。個々の細胞を分析するための方法及び系は、2014年6月26日に出願された米国仮特許出願第62/017,558号に記載されており、その全開示は参照によりその全体が本明細書に組み込まれる。

【0148】

試料は様々な対象から得ることができる。対象は生存対象または死亡対象であってよい。対象の例としては、ヒト、哺乳類、非ヒト哺乳類、ゲッ歯類、両生類、爬虫類、イヌ、ネコ、ウシ、ウマ、ヤギ、ヒツジ、メンドリ、鳥(avines)、マウス、ウサギ、昆虫、ナメクジ、微生物、細菌、寄生虫、または魚を含み得るが、これらに限定されない。いくつかの場合では、対象は、疾患または障害を有する、有する疑いがある、または発症するリスクにある患者であり得る。いくつかの場合では、対象は妊婦であり得る。いくつかの場合では、対象は正常な健康な妊婦であり得る。いくつかの場合では、対象は、ある特定の先天性欠損を有する乳児を有するリスクにある妊婦であり得る。

【0149】

試料は様々なアプローチにより対象から得ることができる。例えば、試料は、循環系へのアクセス(例えば、注射器または他の装置を介して静脈内または動脈内)、分泌された生物学的試料を回収する(例えば、唾液、痰尿、糞便など)、外科的に(例えば、生検)生物学的試料を取得する(例えば、術中試料、術後試料など)、拭き取り(例えば、口腔内拭き取り、口腔咽頭拭き取り)、またはピペット操作により対象から得ることができる。

10

20

30

40

50

【0150】

CNVは療法の有効性と関連し得る。例えば、HER2遺伝子コピー数の増加は、進行した非小細胞肺癌におけるゲフィチニブ療法に対する応答を増強することができる。Cappuzzo F. et al. (2005) J. Clin. Oncol. 23:5007-5018を参照。高EGFR遺伝子コピー数は、ラパチニブ及びカペシタビンに対する感受性の増加について予測することができる。Fabi et al. (2010) J. Clin. Oncol. 28:15s (2010 ASCO Annual Meeting)を参照。高EGFR遺伝子コピー数は、セツキシマブ及びパニツムマブに対する感受性の増加と関連する。

【0151】

コピー数変異は、癌患者のある特定の治療に対する耐性と関連し得る。例えば、チミジル酸シンターゼの増幅は、転移性結腸直腸癌患者において5-フルオロウラシル治療に対する耐性をもたらし得る。Wang et al. (2002) PNAS USA vol. 99, pp. 16156-61を参照。

【0152】

IX. コンピュータ制御システム

本開示は、本明細書に提供される方法、例えば、本明細書に記載される、遺伝的変異の核配列決定及び決定、参照核酸配列の保存、配列分析の実施、ならびに／または試料と参照核酸配列の比較のための方法を実施するようにプログラムされるか、ないしは別の方法でそのように構成されるコンピュータシステムを提供する。このようなコンピュータシステムの例は図22に示される。図22に示されるように、コンピュータシステム2201は、シングルコアもしくはマルチコアプロセッサ、または並列処理用の複数のプロセッサであり得る中央処理装置(CPU、また本明細書において「プロセッサ」及び「コンピュータプロセッサ」)2205を含む。コンピュータシステム2201は、メモリまたはメモリ場所2210(例えば、ランダムアクセスメモリ、読み取り専用メモリ、フラッシュメモリ)、電子保存ユニット2215(例えばハードディスク)、1つ以上の他のシステムと通信するための通信インターフェース2220(例えばネットワークアダプタ)、ならびにキャッシュメモリ、他のメモリ、データ保存、及び／または電子表示アダプタなどの周辺機器2225も含む。メモリ2210、保存ユニット2215、インターフェース2220、及び周辺機器2225は、マザーボードなどの通信バス(実線)を通してCPU2205と通信する。保存ユニット2215は、データを保存するためのデータ保存ユニット(またはデータ保管庫)であり得る。コンピュータシステム2201は、通信インターフェース2220の補助により、コンピュータネットワーク(「ネットワーク」)2230に動作可能に接続され得る。ネットワーク2230は、インターネット、インターネット、及び／もしくはエクストラネット、インターネットと通信するイントラネット及び／もしくはエクストラネットであり得る。いくつかの場合では、ネットワーク2230は、電気通信及び／またはデータネットワークである。ネットワーク2230は、クラウドコンピューティングなどの分散コンピューティングを可能にし得る1つ以上のコンピュータサーバーを含み得る。ネットワーク2230は、いくつかの場合ではコンピュータシステム2201の補助により、コンピュータシステム2201に接続されたデバイスがクライアントまたはサーバーとして機能することを可能にし得るピア・ツー・ピアネットワークを実施することができる。

【0153】

CPU2205は、プログラムまたはソフトウェアにおいて具現化され得る機械読み取り可能命令のシーケンスを実行することができる。命令はメモリ2210などのメモリ場所に保存され得る。CPU2205により実施される動作の例としては、取り出し、解読、実行、及び書き戻しを含み得る。

【0154】

保存ユニット2215は、ドライバ、ライブラリ、及び保存済みプログラムなどのファイルを保存することができる。保存ユニット2215は、ユーザデータ、例えばユーザプ

10

20

30

40

50

リファレンス及びユーザプログラムを保存することができる。コンピュータシステム 2201 は、いくつかの場合では、イントラネットまたはインターネットを通してコンピュータシステム 2201 と通信するリモートサーバー上に位置するなど、コンピュータシステム 2201 の外部の 1 つ以上の追加のデータ保存ユニットを含み得る。

【0155】

コンピュータシステム 2201 は、ネットワーク 2230 を通して 1 つ以上のリモートコンピュータシステムと通信することができる。例えば、コンピュータシステム 2201 は、ユーザ（例えばオペレータ）のリモートコンピュータシステムと通信することができる。リモートコンピュータシステムの例としては、パーソナルコンピュータ（例えば、ポータブル PC）、スレートまたはタブレット PC（例えば、Apple（登録商標）iPad（登録商標）、Samsung（登録商標）Galaxy Tab）、電話、スマートフォン（例えば、Apple（登録商標）iPhone（登録商標）、アンドロイド使用可能デバイス、Blackberry（登録商標））、または携帯情報端末が挙げられる。ユーザは、ネットワーク 2230 を介してコンピュータシステム 2201 にアクセスすることができる。

【0156】

本明細書に記載される方法は、コンピュータシステム 2201 の電子保存場所、例えばメモリ 2210 または電子保存ユニット 2215 上に保存された機械（例えばコンピュータプロセッサ）実行可能コードにより実施され得る。機械実行可能または機械読み取り可能コードはソフトウェアの形態で提供され得る。使用中、コードはプロセッサ 2205 によって実行され得る。いくつかの場合では、コードは、保存ユニット 2215 から検索され、プロセッサ 2205 によって容易にアクセスするためにメモリ 2210 上に保存され得る。いくつかの状況では、電子保存ユニット 2215 は除外されてよく、機械実行可能命令がメモリ 2210 上に保存される。

【0157】

コードは予めコンパイルされ、コードを実行するように適応されたプロセッサを有する機械で使用するために構成され得るか、または実行中にコンパイルされ得る。コードは、コードが予めコンパイルされたまたはコンパイルされてすぐの様式で実行することを可能にするように選択され得るプログラム言語で供給され得る。

【0158】

コンピュータシステム 2201 などの本明細書に提供されるシステム及び方法の態様はプログラミングにおいて具現化され得る。技術の様々な態様は、典型的には、一種の機械読み取り可能媒体において保有または具現化される機械（またはプロセッサ）実行可能コード及び／または関連データの形態で、「製品」または「製造物品」として考えられ得る。機械実行可能コードは、電子保存ユニット、このようなメモリ（例えば、読み取り専用メモリ、ランダムアクセスメモリ、フラッシュメモリ）、またはハードディスク上に保存され得る。「保存」式媒体は、コンピュータ、プロセッサなどの有形メモリのいずれかもしくははすべて、またはそれらの関連モジュール（様々な半導体メモリ、テープドライブ、ディスクドライブなど）を含み得、これは常にソフトウェアプログラミングの非一時的な保存を提供することができる。ソフトウェアのすべてまたは一部は、時折、インターネットまたは様々な他の電気通信ネットワークを通して通信され得る。例えば、このような通信は、1 つのコンピュータまたはプロセッサから別のコンピュータまたはプロセッサに、例えば管理サーバーまたはホストコンピュータからアプリケーションサーバーのコンピュータプラットフォームへの読み込みを可能にし得る。よって、ソフトウェア要素を有し得る別の種類の媒体は、有線及び光学地上通信ネットワークを通して、ならびに様々な無線リンク上で、ローカルデバイス間の物理インターフェースなどにわたって使用される光学波、電波、及び電磁波を含む。有線またはワイヤレスリンク、光学リンクなどのこのような波を運ぶ物理要素もソフトウェアを保有する媒体と考えられ得る。本明細書で使用时、非一時的な有形「保存」媒体に限定されない限り、コンピュータまたは機械「読み取り可能媒体」などの用語は、実行のための命令をプロセッサに提供する際に関与する

任意の媒体を指す。

【0159】

よって、コンピュータ実行可能コードなどの機械読み取り可能媒体は、有形保存媒体、搬送波媒体、または物理伝送媒体を含むが、これらに限定されない多くの形態を取ることができる。不揮発性媒体は、例えば、図面に示されるデータベースなどを実施するためなどに使用され得る任意のコンピュータ（複数可）などの保存デバイスのいずれかなどの光学または磁気ディスクを含む。揮発性保存媒体は、このようなコンピュータプラットフォームのメインメモリなどの動的メモリを含む。有形伝送媒体は、同軸ケーブル；コンピュータシステム内にバスを含むワイヤを含む、銅線及び光学ファイバを含む。搬送波伝送媒体は、電気もしくは電磁シグナル、または高周波（RF）赤外（IR）データ通信中に生成されるものなどの音もしくは光波の形態を取ることができる。したがって、コンピュータ読み取り可能媒体の一般的な形態は、例えば、フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、CD-ROM、DVD、もしくはDVD-ROM、任意の他の光学媒体、パンチカード、紙テープ、穴のパターンを有する任意の他の物理保存媒体、RAM、ROM、PROM、及びEPROM、FLASH-EPROM、任意の他のメモリチップもしくはカートリッジ、搬送波輸送データもしくは命令、このような搬送波などを輸送するケーブルもしくはリンク、またはコンピュータがプログラミングコード及び／もしくはデータを読むことができる任意の他の媒体を含む。コンピュータ読み取り可能媒体のこれらの形態の多くは、実行のためにプロセッサに対する1つ以上の命令の1つ以上のシーケンスの保有に参与し得る。

【0160】

コンピュータシステム2201は、例えば、コンピュータシステム2201に接続された核酸配列決定計器の出力または読み出しを提供するためのユーザインターフェース（UI）を含む電子表示2235を含むか、またはそれと通信することができる。このような読み出しは、所与の核酸試料を含む核酸塩基の配列など、核酸配列決定読み出しを含み得る。UIは、このような読み出しを利用する分析の結果を表示するためにも使用され得る。UIの例としては、グラフィカルユーザインターフェース（GUI）及びウェブベースユーザインターフェースが挙げられるが、これらに限定されない。電子表示2235は、コンピュータのモニタ、または容量もしくは抵抗タッチスクリーンであり得る。

【実施例】

【0161】

実施例1：フェーズ化変異体の特定

NA12878ヒト細胞系からのゲノムDNAは、長さが約10kbの断片を回収するために、Blue Pippin DNAサイジング法を用いて断片のサイズ基準分離を受けた。次に、マイクロ流体分配系を使用して（例えば、2014年4月10日に出願された米国仮特許出願第61/977,804号を参照、すべての目的においてその全体が参照により本明細書に組み込まれる）、フッ素化油連続相内の水性液滴中にサイズ選択された試料核酸をバーコードビーズと共分配し、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行うための他の試薬、ならびにビーズからバーコードオリゴヌクレオチドを放出するための化学活性剤も含む。これは、1ngの総DNA投入量及び2ng総DNA投入量の両方に対して繰り返された。700,000を超える異なるバーコード配列のバーコード多様性を表すストックライブラリのサブセットとしてバーコードビーズを得た。バーコード含有オリゴヌクレオチドは追加の配列成分を含み、一般構造式：

ビーズ-P5-BC-R1-Nmer

を有し、式中、P5及びR1は、それぞれ、Illumina付加及びリード1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、テンプレート核酸をプライムするために使用されたランダム10塩基Nmerプライミング配列を示す。例えば、2014年6月26日に出願された米国特許出願第14/316,383号を参照（その全開示はすべての目的においてその全体が参照により本明細書に

組み込まれる)。

【0162】

ビーズの溶解後、各液滴内の試料核酸のテンプレートに対してバーコードオリゴのプライマー伸長を可能にするために、液滴を熱サイクル (thermocycled) した。これは、上述の他の含まれた配列に加えて、発生元のパーティションを表すバーコード配列を含む試料核酸のコピー断片をもたらした。

【0163】

コピー断片のバーコード標識後、増幅されたコピー断片を含む液滴のエマルジョンが破壊され、追加のシーケンサー必要成分、例えばリード2プライマー配列及び Illumina シーケンサー用の P7 付加配列が、コピー断片のもう一端にこれらの配列を付加した更なる増幅を通して、コピー断片に添加された。

【0164】

次に、配列決定ライブラリが 10X 読み深度、20X 読み深度、及び 30X 読み深度で、Illumina HiSeq システムで配列決定され、次に得られた配列リード及びそれらの関連バーコード配列が分析された。共通のバーコードを共有した近接マッピング配列はより大きいコンティグにアセンブリされ、フェーズ化 SNP を特定するために、それらの関連バーコード及び配列マッピングに基づき、一塩基多型が特定され、個々の開始分子と関連付けられた。次に、重複するフェーズ化 SNP を含んだ配列は、重複するフェーズ化 SNP に基づいてフェーズ化配列データのフェーズブロックまたは推測コンティグにアセンブリされた。得られたデータは、比較のための細胞系の既知のハプロタイプマップと比較された。

【0165】

少なくとも 1 つのアプローチでは、一連のヘテロ接合変異体の各対立遺伝子は、2 つのハプロタイプに対して 2 つのうちの 1 つに割り当てられる。一組の変異体と仮定すると、観察されたリード及びバーコードデータの対数尤度、ならびにヘテロ接合変異体のフェージング割り当てを返す対数尤度関数対数 P (バーコード化されたリード | フェージング割り当て、変異体) が定義される。対数尤度関数の形態は、バーコード化された配列リードデータについての 2 つの主な観測から導かれる：(1) 1 つのバーコードからのリードが一倍体ゲノムのごく一部を網羅するため、ゲノムの所与の領域における両ハプロタイプに関するリードを含有する 1 つのバーコードの確率は小さい。逆に、ゲノムの局所領域における 1 つのバーコードに関するリードは単一のハプロタイプに由来する可能性が高い；(2) 観察された塩基が、それが由来したハプロタイプにおける真の塩基と異なる確率は、シーケンサーにより割り当てられた観察された塩基の Phred QV により示される。

【0166】

次に、所与の一組のバーコード化されたリード及び変異体の対数尤度関数を最大限にするフェージング構成が報告される。次に、ハプロタイプ構成を点数化する最大尤度は、構造化探索方法により見出される。最初に、隣接変異体の小ブロック (例えば約 50 の変異体) の最適なフェージング構成を見つけるために、ビーム探索が使用される。第 2 に、ブロックの相対的フェージングがブロック連結部にわたる掃引において決定される。この時点で、全体的な近最適フェージング構成が見出され、更なる最適化の開始点として使用される。次に、個々の変異体のハプロタイプ割り当てはフェージングに対する局所改善を見出すために反転され、交換した構成間の対数尤度における差はそのフェージング割り当ての信頼度の推定を提供する。最後に、フェージング構成は、内部的に正しい確率が高いフェーズブロックに分割される。次に、最適構成の対数尤度を、現在の SNP の右のすべての SNP がそれらのハプロタイプの割り当てを反転させる構成と比較することにより、各 SNP でフェーズブロックを破壊するかが試験される。

【0167】

下の表は、NA 12878 ゲノムに関して得られたフェージング計量を提供する。明らかに、非常に長いフェーズブロックは、短いリード配列データから得られ、非常に低い短いまたは長いスイッチ誤差で、顕著なパーセンテージのフェーズ化 SNP を正確

に特定する。

【表 1】

	1 0 X 読み 深度	2 0 X 読み 深度	3 0 X 読み 深度	3 0 X ビー ム探索
N 5 0 フェ ーズブロッ ク	1 9 3 k b	3 8 5 k b	4 2 8 k b	4 8 9 k b
最長フェー ズブロック	2 1 2 1 k b	2 5 1 4 k b	2 5 1 4 k b	3 0 2 7 k b
長いスイッ チ誤差	0 . 0 0 5 3	0 . 0 0 2 1	0 . 0 0 1 8	0 . 0 0 1 5
短いスイッ チ誤差	0 . 0 0 4	0 . 0 0 1 7	0 . 0 0 1 4	0 . 0 0 1 2
S N P フェ ーズ化	8 3 %	9 4 %	9 5 %	9 5 . 2 %

10

【0 1 6 8】

NA 1 2 8 7 8 三つ組 (NA 1 2 8 7 8、NA 1 2 8 8 2、及び NA 1 2 8 7 7)、グ
ジャラート人 (NA 2 0 8 4 7)、メキシコ人 (NA 1 9 6 6 2)、アフリカ人 (NA 1
9 7 0 1) 細胞系試料を含むいくつかの追加の試料からフェーズ化 SNP を更に実験する
。約 1 MB の N 5 0 フェーズブロック長は、0 . 3 % 未満のスイッチ誤差で、9 5 % を超
えるフェーズ化された SNP で達成された。同じ試料の全エクソーム配列決定、例えば、
バーコード化に従った標的プルダウンは、同様に 0 . 3 % 未満のスイッチ誤差で、約 9 0
% の遺伝子 SNP フェージングを示した。

20

【0 1 6 9】

実施例 2 : EML-4 / ALK 遺伝子反転 / 転座の特定

本明細書に記載される方法及びプロセスは、特徴付けされた癌細胞系から構造的変異を
検出するために使用された。特に、NCI-H 2 2 2 8 肺癌細胞系は、そのゲノム内に E
ML 4 - ALK 融合転座を有することで知られている。野生型と比較した変異の構造は図
1 5 に図示される。上のパネルに示されるように、変異体構造において、EML-4 遺伝
子は、同じ染色体上にあるが、ALK 遺伝子から比較的離れているか、または距離があり
、代わりに転座され、ALK 遺伝子に融合される (例えば、Choi, et al., I
dentification of Novel Isoforms of the E
ML 4 - LK Transforming Gene in Non-Small Ce
ll Lung Cancer, J. Cancer Res., 68 : 4971 (Jul
y 2008) を参照)。転座と併せて、EML 4 遺伝子も反転される。転座は、野生型
構造と比較したとき、パネル I I に更に図示され、転座は、EML-4 のエクソン 1 ~ 6
(黒のボックスとして示される) の、ALK のエクソン 20 ~ 29 (白のボックスとして
示される) への融合、ならびに EML-4 のエクソン 1 ~ 19 に融合した ALK のエクソ
ン 7 ~ 23 の融合をもたらす。

30

40

【0 1 7 0】

この変異を特定するために、NCI-H 2 2 2 8 細胞系からのゲノム DNA は、長さが
約 10 kb の断片を選択するために Blue Pippin (登録商標) システム (Sa
ge Sciences, Inc.) を使用してサイズ分離を受けた。

【0 1 7 1】

次に、サイズ選択された試料核酸は、DNA が、バーコード化後及び配列決定前に、A
gilent SureSelect Exome 捕捉キットを使用してハイブリッド捕
捉を受けたことを除き、実施例 1 に関して上述されるように、バーコードビーズで共分配
され、増幅され、配列決定ライブラリに加工された。次に、配列決定ライブラリは、I I

50

lumina HiSeq システムで、約 80X 読み深度まで配列決定され、次に得られた配列リード及びそれらの関連バーコード配列が分析された。野生型と比較したとき、転座事象に及ぶゲノムの部分の間でのより高い数の共有バーコードが明らかに明白であり、野生型に存在しない融合成分間の構造的近接を示す。特に、及び図 16A に示されるように、融合構造は、EML-4 エクソン 1~6 と ALK エクソン 20~29 との間で 12 バーコード、そして EML-4 エクソン 7~23 と ALK エクソン 1~19 との間で 20 バーコードのバーコード重複を示し、これはヘテロ接合細胞系の野生型構築物の重複するバーコードに匹敵するものであった。

【0172】

対照的に、非変異体細胞系 (NA12878) を用いた陰性対照実行は、図 16B に示されるように、実質的に、野生型対変異体構築物のバーコード重複のみを示し、配列読み深度は約 140X であり、3ng の開始 DNA を使用した。

【0173】

特に、様々な配列セグメントに対して多数の合計マッピングバーコードを示すが、非常に高い数の共通のまたは重複するバーコードを示した野生型構造と比較して、非常に小さいパーセンテージの重複するバーコードのみ、例えば合計マッピングバーコード配列の 0.5% 未満が融合構造に対して見られた。結果として、融合または転座切断点にわたって一般的なマッピングバーコードはこれらの転座事象を特定するための強力な基盤を提供する。

【0174】

最初に顕著なバーコード交差／重複を伴うゲノム遺伝子座のすべての対を検索し、この検索を効率的な疎行列乗算としてコードする SV 検出のためのアルゴリズムも利用された。次に、この第 1 段階からの候補は、リード対、分割リード、及びバーコードデータを組み込む確率モデルを利用してフィルタ処理される。NA12878 及び NA20847 に対する SV 呼び出しは、複数の大規模な欠失及び反転の呼び出し、及び隣接するフェーズブロックに対するそれらのフェージングをもたらし、上述の核三つ組における遺伝パターンを有するフェージングの一貫性を示す。

【0175】

実施例 3：CNV スクリーニングを介した狼瘡に対する増加した感受性の検出

患者は狼瘡に対する感受性について検査される。患者から血液を得る。無細胞 DNA 試料は本明細書に列挙される技法を用いて配列決定される。次に、異なる遺伝子の CNV を決定するために、この配列を既知のゲノム参照配列と比較する。FCGR3B (CD16 細胞表面免疫グロブリン受容体) の低コピー数は、全身性エリテマトーデスに対する増加した感受性を示す。患者は、任意のコピー数異常及び関連リスク／疾患を通知される。

【0176】

実施例 4：CNV スクリーニングを介した増加した神経芽細胞腫の素因の検出

患者は神経芽細胞腫の素因について検査される。患者から血液を得る。無細胞 DNA 試料は本明細書に列挙される技法を用いて配列決定される。次に、異なる遺伝子の CNV を決定するために、この配列を既知のゲノム参照配列と比較する。1q21.1 での CNV は増加した神経芽細胞腫の素因を示す。患者は、任意のコピー数異常及び関連リスク／疾患を通知される。

【0177】

実施例 5：CNV スクリーニングを介した肺癌の鑑別診断

慢性咳、体重減少、及び息切れを有する患者が肺癌について検査される。患者から血液を得る。循環腫瘍細胞 (CTC) または無細胞 DNA 試料は本明細書に列挙される技法を用いて配列決定される。次に、異なる遺伝子の CNV を決定するために、CTC 配列を既知のゲノム参照配列と比較する。DNA における EGFR コピー数が通常より高い場合、患者は、小細胞肺癌の代わりに非小細胞肺癌 (NSCLC) と鑑別診断され得る。非小細胞肺癌の CTC は小細胞肺癌と更に区別され得る他のコピー数変異も有する。癌の病期により、外科手術、化学療法、または放射線療法が処方される。

10

20

30

40

50

【0178】

小細胞肺癌は、大半の場合、非小細胞肺癌よりも迅速かつ広範囲に転移性である（よって、異なっており病期分類される）。NSCLCは、通常、化学療法及び／または放射線にあまり感受性ではないため、早期で診断された場合、外科手術が選択治療であり、多くの場合、補助（付随的）化学療法はシスプラチンを伴う。非小細胞肺癌（NSCLC）を有する患者に対しては、標的療法、例えばクリゾチニブなどのALK阻害剤も利用可能である。標的療法は、すべての迅速に分裂する細胞に単純に干渉するよりも（例えば従来の化学療法）、むしろ発癌及び腫瘍成長に必要とされる特定の標的分子に干渉することにより癌細胞の成長を遮断する。

【0179】

実施例6：フェージングを介した胎児異数体の鑑別診断

胎児異数体は染色体数における異常である。異数体は、一般的に、顕著な身体的及び神経性障害をもたらす。X染色体の数の減少はターナー症候群に關与する。染色体番号21のコピー数の増加はダウン症候群を引き起こす。羊水穿刺または絨毛検査（CVS）などの侵襲性試験は流産のリスクにつながる場合があり、母体血液を検査するあまり侵襲性ではない方法がここでは使用される。

【0180】

ダウン症候群またはターナー症候群の家族歴を有する妊娠中の患者が検査される。胎児遺伝物質を含有する母体の血液試料を回収する。次に、異なる染色体からの核酸は、本明細書に記載されるように、バーコード化されたタグ分子と共に異なるパーティションに分離される。次に、試料を配列決定し、各染色体コピーの数を正常な二倍体染色体の配列と比較する。患者は、異なる染色体の任意のコピー数異常及び関連リスク／疾患を通知される。

【0181】

実施例7：パーキットリンパ腫の鑑別診断のためのフェージングを介した染色体転座の検出

パーキットリンパ腫は染色体のt(8;14)転座を特徴とする。一般的にリンパ腫と診断された患者はパーキットリンパ腫について検査される。腫瘍生検組織はリンパ節から回収される。異なる染色体からの核酸は、本明細書に記載されるように、バーコード化されたタグ分子と共に異なるパーティションに分離される。次に、試料は、染色体転座を検出するために、配列決定され、対照DNA試料と比較される。患者がパーキットリンパ腫を有すると診断される場合、CHOPまたはR-CHOPレジメンを含む、他の種類のリンパ腫よりも集中的な化学療法レジメンが必要とされる可能性がある。CHOPは、DNAに結合し、架橋結合の形成をもたらすことによりDNAを損傷するアルキル化剤であるシクロホスファミド、DNA塩基間に自身を挿入することによりDNAを損傷する挿入剤であるヒドロキシダウノルビシン（ダウノルビシンまたはアドリアマイシンとも呼ばれる）、タンパク質チューブリンに結合することにより細胞の重複を阻止するオンコピン（ビンクリスチン）、コルチコステロイドであるプレドニゾンまたはプレドニゾンからなる。このレジメンは、パーキットリンパ腫がB細胞起源のものであるため、モノクローナル抗体リツキシマブとも併用され得、この併用はR-CHOPと呼ばれる。

【0182】

実施例8：親のゲノムとの比較による無細胞DNAに由来する胎児ゲノム配列のフェージング

妊娠中の患者からの母方DNAを含む試料、及び胎児の父親からの父方DNAを含む試料が回収される。各試料からの核酸は、本明細書に記載されるように、分子バーコード化されたタグと共に異なるパーティションに分離される。次に、試料を配列決定し、分配された母方及び父方の断片の各々に関して推測コンティグを生成するために配列を使用する。推測コンティグは、母方及び父方の染色体の各々の部分についてのハプロタイプブロックを構築するために使用される。

【0183】

胎児遺伝物質を含有する母体の血液試料を回収する。無細胞DNAを配列決定して、母方循環DNA及び胎児循環DNAの両方の配列を生成する。リードを上記で生成した父方及び母方のフェーズブロックと比較する。一部のフェーズブロックは減数分裂中に組換えを受けた。父方フェーズブロックと一致し、母方フェーズブロックと一致しない胎児物質が特定される。いくつかの場合では、胎児物質は父方フェーズブロックの全体と一致し、父方由来の染色体において、胎児がその父方フェーズブロックを有すると決定される。他の場合では、胎児物質はフェーズブロックの一部と一致し、次に第2のフェーズブロックと一致し、2つのフェーズブロックは父方ゲノムにおいて相同染色体領域上にある。減数分裂期組換え事象がこの領域で生じたことが決定され、組換えの可能性が高い地点が決定され、2つの父方フェーズブロックの組み合わせである新しい胎児フェーズブロックが生成される。

10

【0184】

循環DNAの配列を母方フェーズブロックと比較する。母方フェーズブロックにおけるヘテロ接合性部位は、母方由来の胎児染色体の可能性が高いフェーズを決定するために使用される。循環DNA配列は、母方ゲノムのヘテロ接合部位でのコピー数を決定するために使用される。特定の母方フェーズブロックの上昇したコピー数は、母方由来の胎児の染色体が上昇したフェーズブロックの配列を含有することを示す。いくつかの場合では、父方の場合において記載されるものと類似して、最初は、相同領域の1つのフェーズブロックが上昇したように見え、その後、同じ領域の別のフェーズブロックの一部が上昇したように見え、減数分裂期組換えが生じたことを示す。これらの場合では、組換えの可能性が高い領域が決定され、新しい胎児フェーズブロックが2つの母方フェーズブロックから構築される。

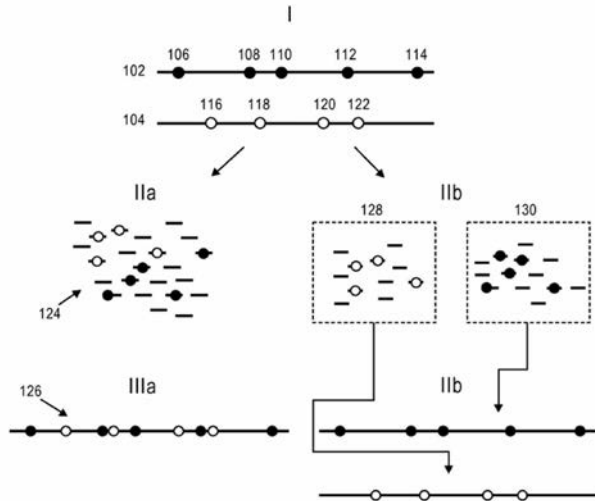
20

【0185】

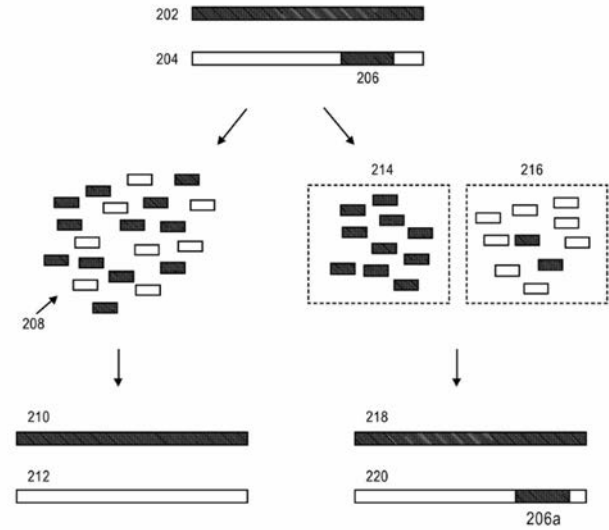
本発明の好ましい実施形態が本明細書において示され、説明されるが、このような実施形態が単に例として提供されることは当業者には明らかであろう。本発明は本明細書内に提供される特定の実施例によって制限されることが意図されない。本発明は前述の明細書に関して説明されるが、本明細書の実施形態の説明及び図は、限定的な意味で解釈されるように意図されていない。ここで、当業者は、多くの変形、変更、及び置換を、本発明から逸脱することなく想到するだろう。更に、本発明のすべての態様は、様々な条件及び変数に依存する、本明細書に記述される特定の描写、構成、または相対的割合に限定されないことを理解するべきである。本明細書に記載される本発明の実施形態に対する様々な代替が本発明を実践する際に利用され得ることを理解するべきである。したがって、本発明は任意のそのような代替、修正、変形、または等価物を網羅することも企図される。以下の特許請求の範囲は本発明の範囲を定義し、これらの特許請求の範囲の範囲内の方法及び構造ならびにそれらの等価物はそれにより網羅されることが意図される。

30

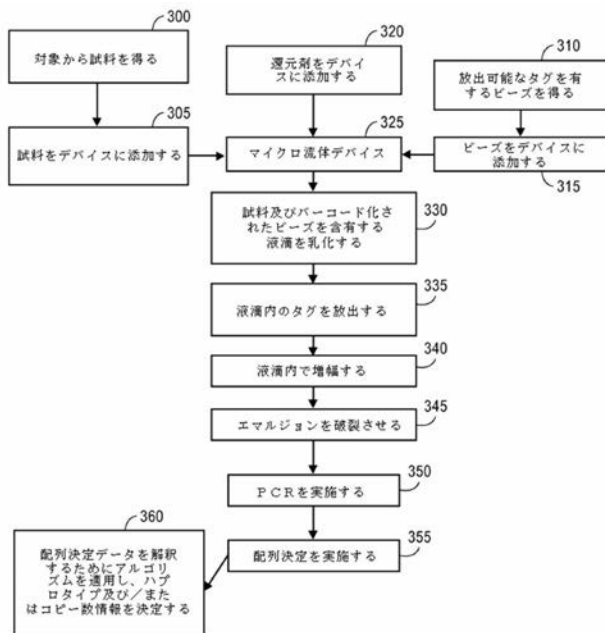
【図 1】



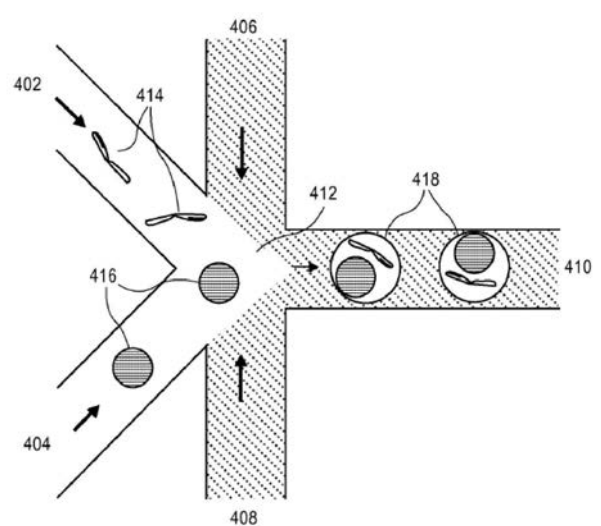
【図 2】



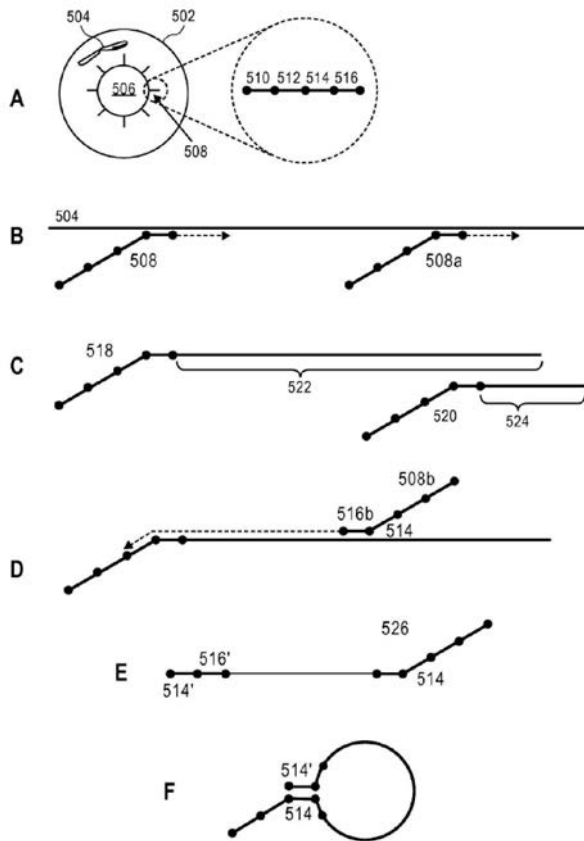
【図 3】



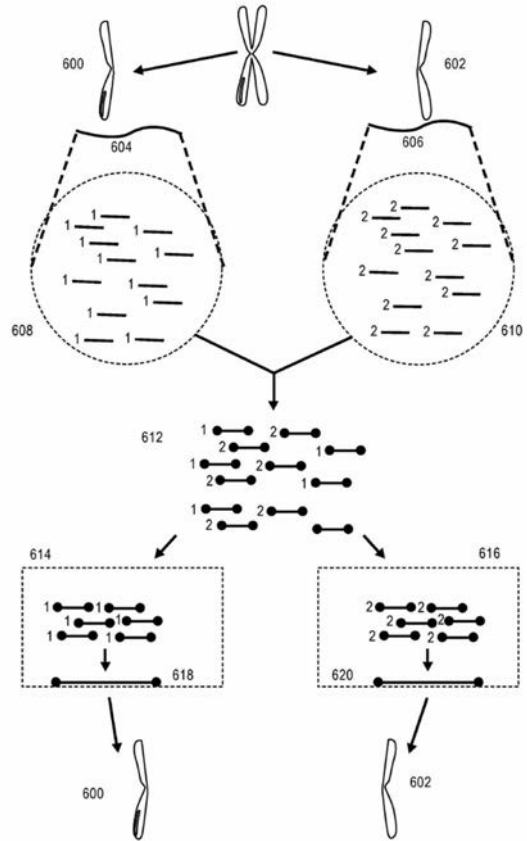
【図 4】



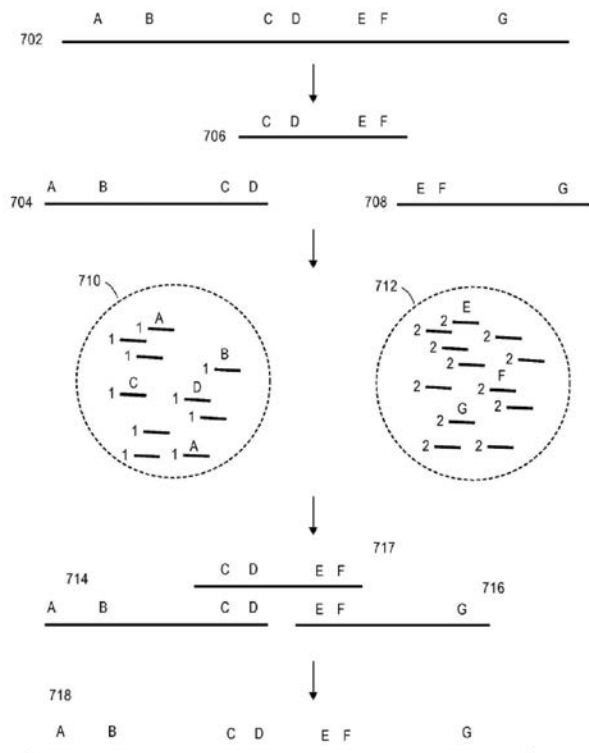
【図 5】



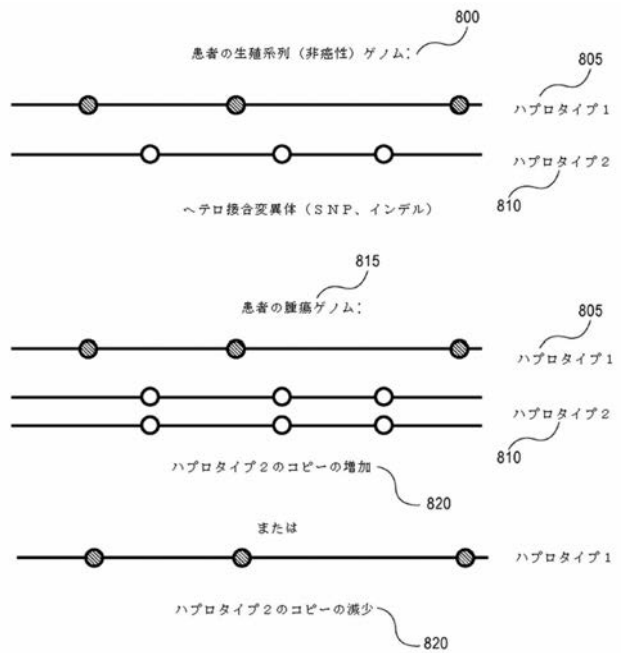
【図 6】



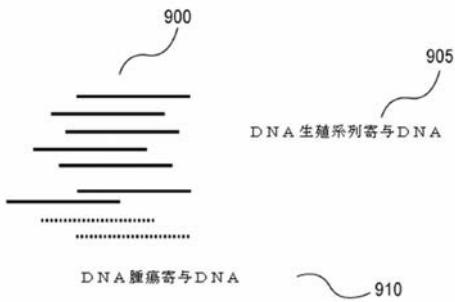
【図 7】



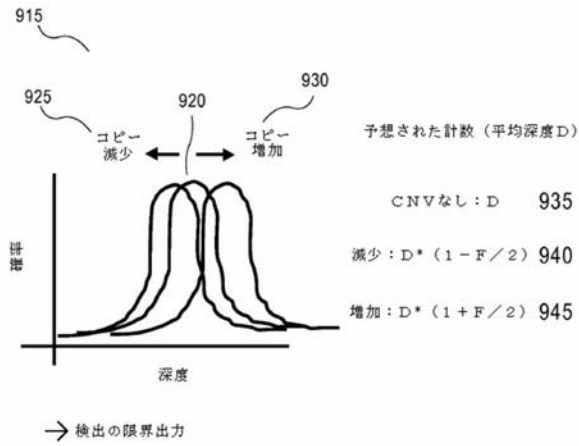
【図 8】



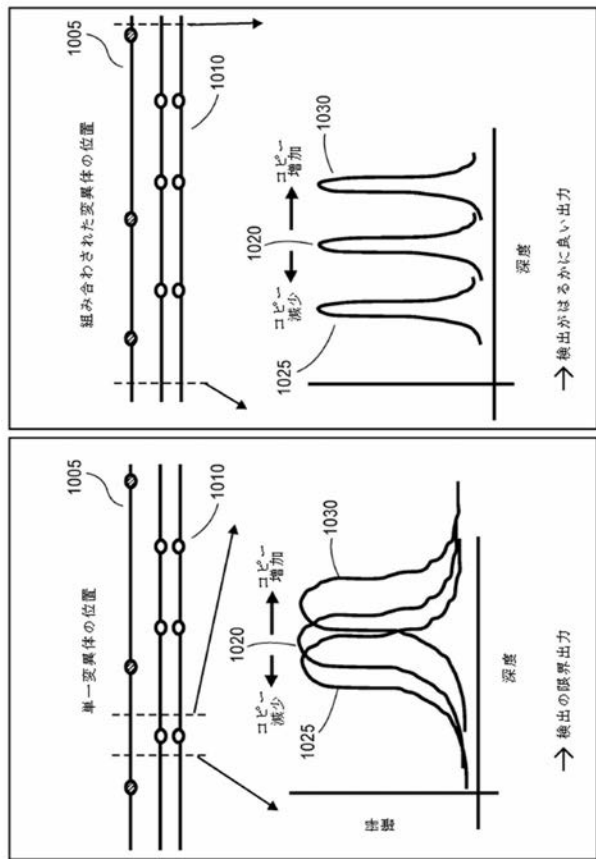
【図 9 A】



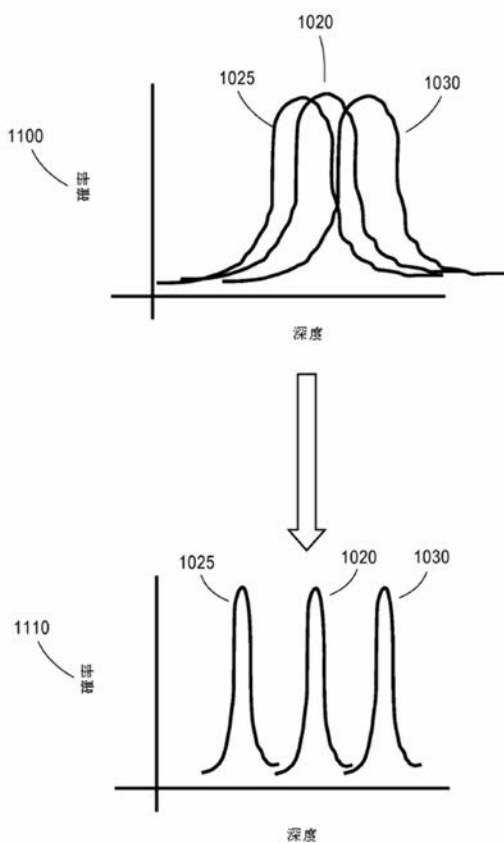
【図 9 B】



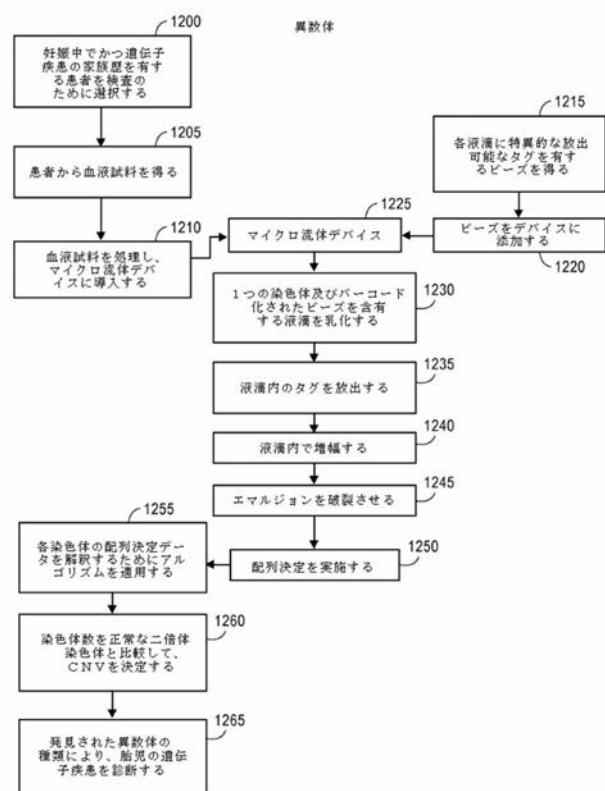
【図 10】



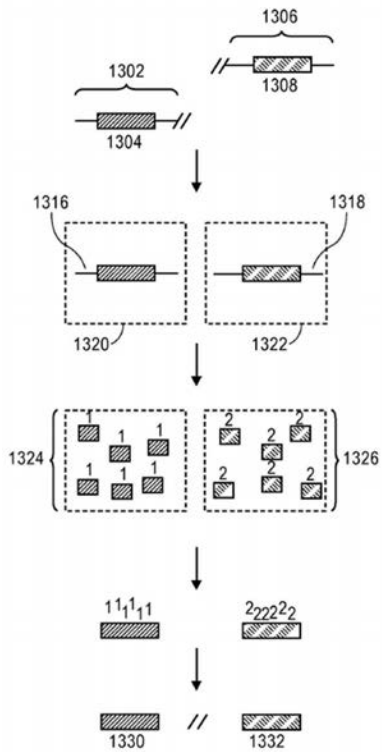
【図 11】



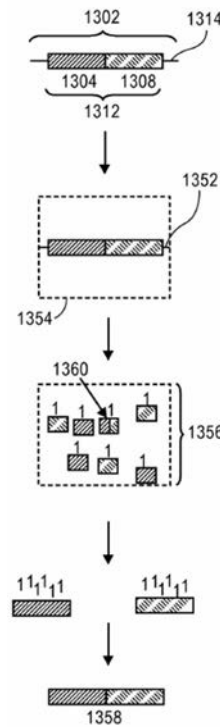
【図 12】



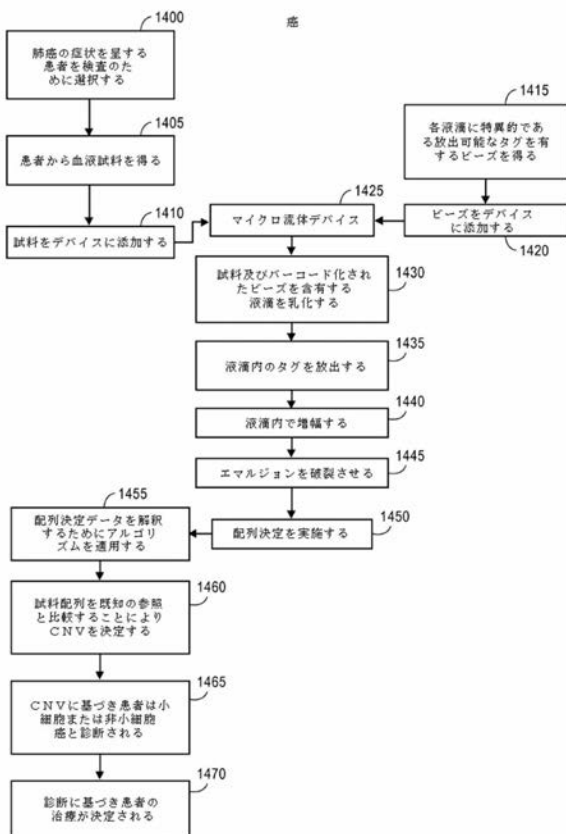
【図 13 A】



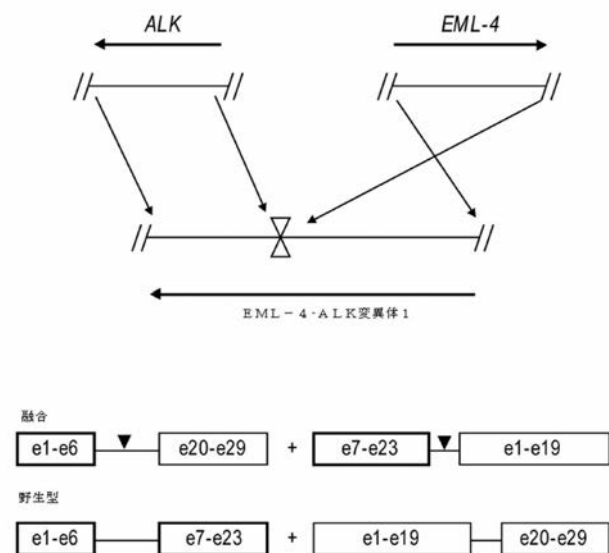
【図 13 B】



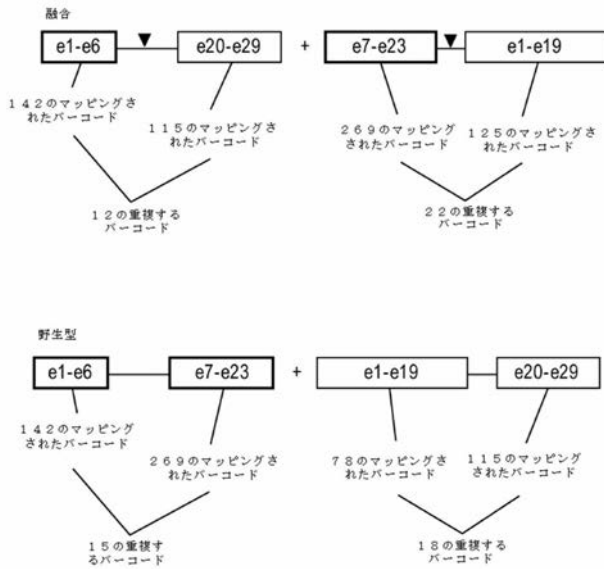
【図 14】



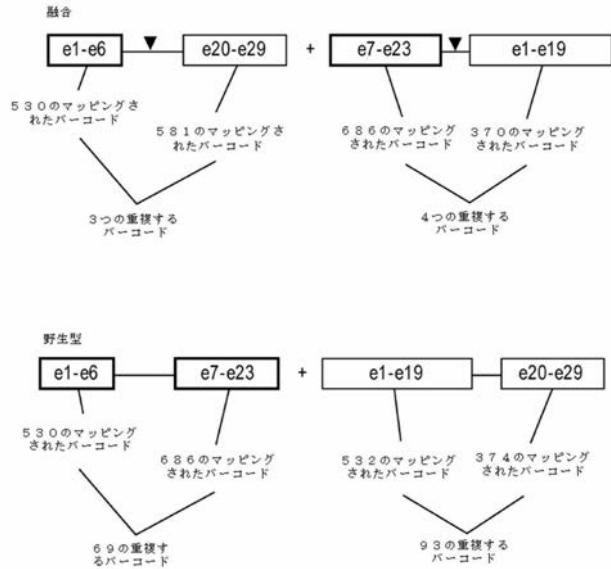
【図 15】



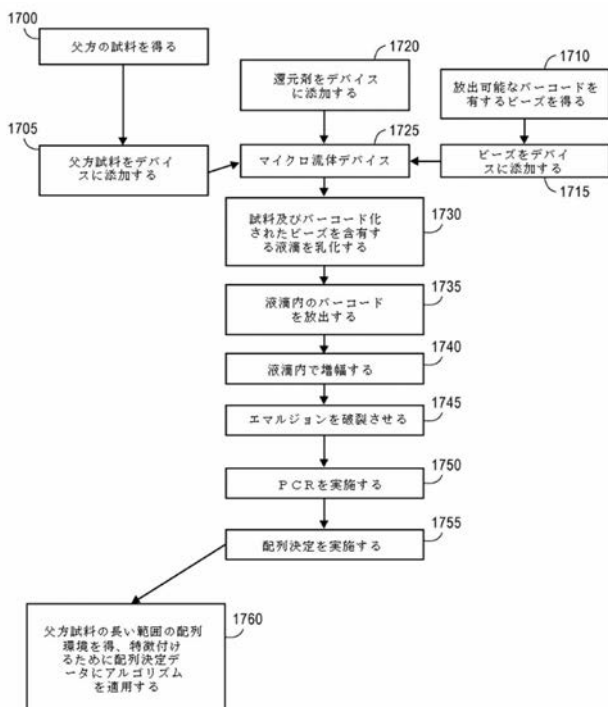
【図 16 A】



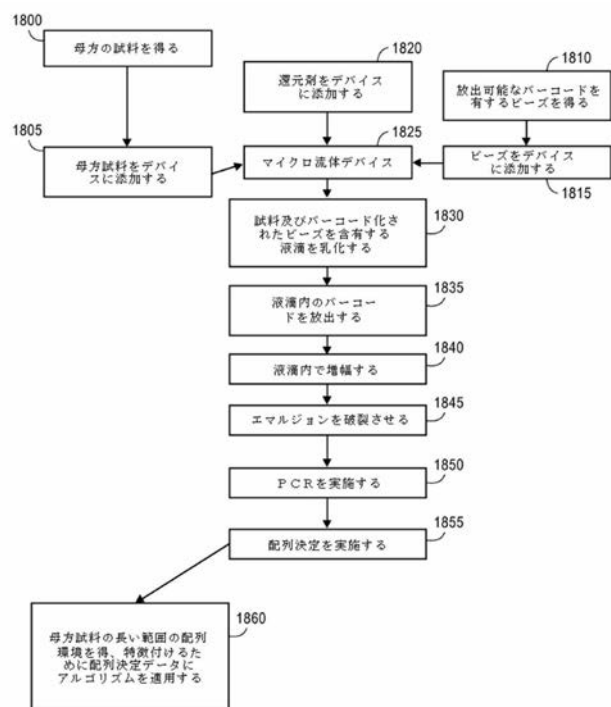
【図 16 B】



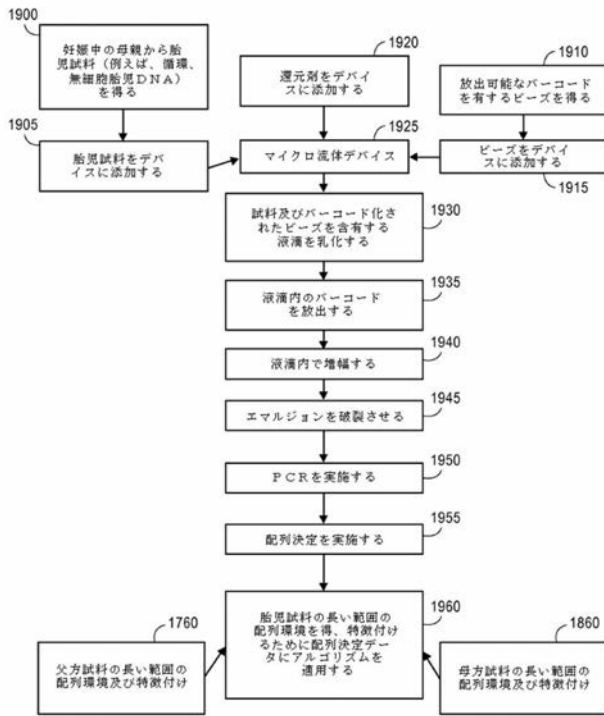
【図 17】



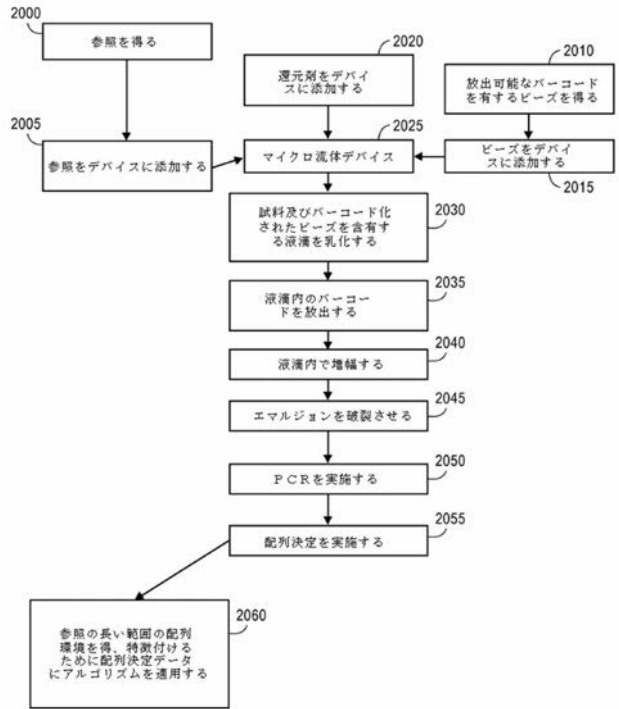
【図 18】



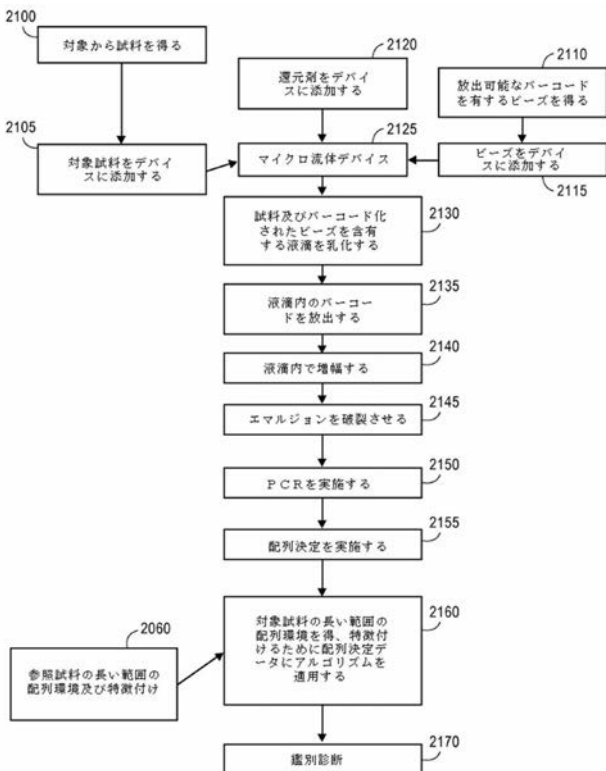
【図 19】



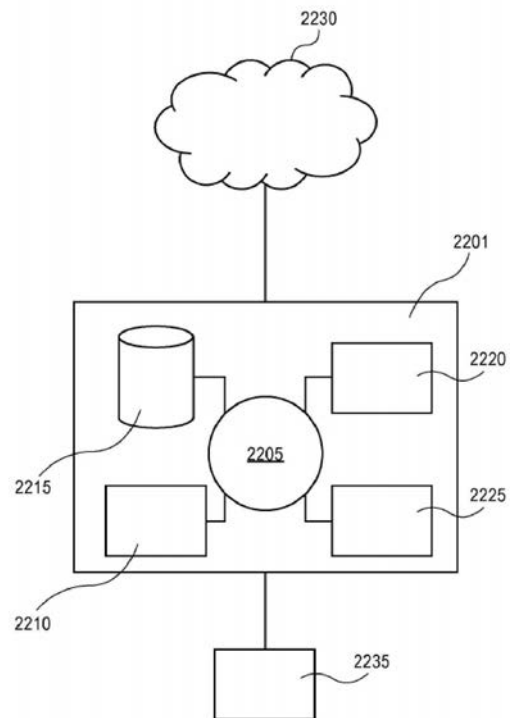
【図 20】



【図 21】



【図 22】



【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US15/38141												
A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - C12Q1/68, C12N15/10, C40B40/06 (2015.01) CPC - C12Q1/6869, G06F19/22, C12Q1/6806 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) IPC(8) Classification(s): C12Q1/68, C12N15/10, C40B40/06 (2015.01) CPC Classification(s): C12Q1/6869, G06F19/22, C12Q1/6806 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) PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); EBSCO; PubMed; Google/Google Scholar; 'haplotype phasing', 'copy number variation', barcode, 'multiple partition, N50, 'phase block', sequencing														
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X — Y</td> <td>US 2013/0157870 A1 (PUSHKAREV, D et al.) June 20, 2013; abstract; figures 1, 4, 7, 8 paragraphs [0015], [0026], [0027], [0129], [0137], [0143], [0157]; [0158]-[0161], Claim 24</td> <td>1-41, 44-51, 65-81 42, 43, 52-54</td> </tr> <tr> <td>Y</td> <td>WO 2012/106546 A2 (UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION) August 09, 2012; abstract; paragraphs [0081], [00233]</td> <td>42, 43, 52-54</td> </tr> <tr> <td>A</td> <td>BROWNING SR et al. Haplotype Phasing: Existing Methods And New Developments. NatRev Genet. 2011 Sept 16; Vol. 12, No. 10, pp 703-714; DOI: 10.1038/nrg3054. entire document</td> <td>1-54, 65-81</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X — Y	US 2013/0157870 A1 (PUSHKAREV, D et al.) June 20, 2013; abstract; figures 1, 4, 7, 8 paragraphs [0015], [0026], [0027], [0129], [0137], [0143], [0157]; [0158]-[0161], Claim 24	1-41, 44-51, 65-81 42, 43, 52-54	Y	WO 2012/106546 A2 (UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION) August 09, 2012; abstract; paragraphs [0081], [00233]	42, 43, 52-54	A	BROWNING SR et al. Haplotype Phasing: Existing Methods And New Developments. NatRev Genet. 2011 Sept 16; Vol. 12, No. 10, pp 703-714; DOI: 10.1038/nrg3054. entire document	1-54, 65-81
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.												
X — Y	US 2013/0157870 A1 (PUSHKAREV, D et al.) June 20, 2013; abstract; figures 1, 4, 7, 8 paragraphs [0015], [0026], [0027], [0129], [0137], [0143], [0157]; [0158]-[0161], Claim 24	1-41, 44-51, 65-81 42, 43, 52-54												
Y	WO 2012/106546 A2 (UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION) August 09, 2012; abstract; paragraphs [0081], [00233]	42, 43, 52-54												
A	BROWNING SR et al. Haplotype Phasing: Existing Methods And New Developments. NatRev Genet. 2011 Sept 16; Vol. 12, No. 10, pp 703-714; DOI: 10.1038/nrg3054. entire document	1-54, 65-81												
☐ 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 "&" document member of the same patent family														
Date of the actual completion of the international search 26 October 2015 (26.10.2015)		Date of mailing of the international search report 25 NOV 2015												
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300		Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774												

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US15/38141

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

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

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

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

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

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

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

-***Please See Supplemental Page-***

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Group I: Claims 1-54, 65-81

Remark on Protest

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

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/US15/38141

Continued From Box III: Lack of unity of invention.

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

Group I: Claims 1-54 and 65-81 are directed toward a method for identifying one or more variations in a nucleic acid, comprising: (a) providing a first fragment of the nucleic acid, wherein the first fragment has a length greater than 10 kilobases (kb); (b) sequencing a plurality of second fragments of the first fragment to provide a plurality of fragment sequences, which plurality of fragment sequences share a common barcode sequence; and determining a nucleic acid sequence of the first fragment using the plurality of fragment sequences, wherein the nucleic acid sequence is determined at an error rate of less than 1%.

Group II: Claims 55-64 are directed toward a method of determining a presence of a structural variation of a nucleic acid, comprising: identifying the structural variation based on a presence of an elevated amount of the plurality of fragment sequences sharing the common barcode sequence that map to the reference sequence at locations that are further apart than a length of the given first fragment molecule, which elevated amount is relative to a sequence lacking the structural variation.

Group III: Claims 82-92 are directed toward a method of characterizing nucleic acids, comprising: linking the nucleic acid sequences of at least a first fragment molecule to at least a second fragment molecule based upon a presence of one or more common phased variant positions within the first and second fragment molecules, to provide a phase block with a maximum phase block length of at least 10 kb; and identifying one or more phased variants from the phase block with the maximum phase block length of at least 10 kb.

Group IV: Claims 93-116 are directed toward a method comprising: a) partitioning a first nucleic acid into a first partition, where the first nucleic acid comprises the target sequence derived from a first chromosome of an organism; b) partitioning a second nucleic acid into a second partition, where the second nucleic acid comprises the target sequence derived from a second chromosome of the organism, and comparing the nucleic acid sequence of the first and second nucleic acids to identify any variation between the nucleic acid sequence of the first and second nucleic acids.

Group V: Claims 117-139 are directed toward a method for characterizing a fetal nucleic acid sequence, comprising: obtaining a fetal nucleic acid from the pregnant mother and determining a sequence of the fetal nucleic acid and/or one or more genetic variations of the sequence of the fetal nucleic acid using the maternal nucleic acid sequence and the paternal nucleic acid sequence.

Group VI: Claims 140-163 are directed toward a method for characterizing a sample nucleic acid, comprising: (d) using a programmed computer processor to generate a comparison of the sample nucleic acid sequence to a reference nucleic acid sequence, which reference nucleic acid sequence has a length greater 10 kilobases (kb) and an accuracy of at least 99%; and (e) using the comparison to identify one or more genetic variations in the sample nucleic acid sequence, thereby associating the sample nucleic acid with a disease.

The inventions listed as Groups I-VI do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include determining a nucleic acid sequence of a first fragment at an error rate of less than 1%, which is not present in any other Group, the special technical features of Group II including determining locations that are further apart than a length of the given first fragment molecule, which is not present in any other Group, the special technical features of Group III including phased variant positions, which is not present in any other Group, the special technical features of Group IV including first and second chromosomes of an organism, which are not present in any other Group, the special technical features of Group V including a fetal nucleic acid sequence, which is not present in any other Group, the special technical features of Group VI including associating a sample nucleic acid with a disease.

Groups I-VI share the technical features including characterizing nucleic acids, including variants or variations in nucleic acids; nucleic acid fragments; and determining nucleic acid sequences of the fragments. Groups I, III, and VI share the technical features including a length greater than 10 kilobases. Groups I, II and V share the technical features including sequencing a plurality of second fragments of a first fragment; and wherein a plurality of fragments or fragment sequences share a common barcode sequence. Groups I, II, and IV-VI share the technical features including a barcode sequence. Groups I and III-VI share the technical features including attributing the plurality of fragment sequences to the first fragment, or assembling the plurality of fragment sequences. Groups I and IV-VI share the technical features including discrete partitions. Groups I and V share the technical features including separating a plurality of fragments into discrete partitions. Groups II-IV and VI share the technical features including mapping or comparing the sequences to a first sequence or reference sequence.

However, these shared technical features are previously disclosed by US 2013/0157870 A1 to Pushkarev, et al. (hereinafter "Pushkarev").

Pushkarev discloses characterizing nucleic acids (obtaining a nucleic acid sequence (characterizing nucleic acids); abstract), including variants or variations in nucleic acids (including haplotype sequences (including variants or variations in nucleic acids); abstract); nucleic acid fragments (figure 1; paragraph [0007]); and determining nucleic acid sequences of the fragments (paragraph [0007]); a length greater than 10 kilobases (long continuous fragments and intermediate sized fragments of about 10,000 bp (a length greater than 10 kilobases), paragraph [0007]; wherein a long continuous fragment is of a size greater than the intermediate length of 10 kb fragments); sequencing a plurality of second fragments of a first fragment (sequencing a plurality of fragments generated from an amplified fragment (sequencing a plurality of second fragments of a first fragment); paragraphs [0010], [0011]); a barcode sequence (paragraph [0008]); wherein a plurality of fragments or fragment sequences share a common barcode sequence (wherein the fragments in a partition are labeled with partition-specific barcodes (paragraph [0013]); attributing the plurality of fragment sequences to the first fragment or assembling the plurality of fragment sequences (assembling a consensus sequence from the sequence reads; paragraph [0007]); discrete partitions (paragraph [0007]); separating a plurality of fragments into discrete partitions (paragraph [0007]); and mapping or comparing the sequences to a first sequence or reference sequence (comparing assembled long fragments with reference sequences; paragraph [0033]).

Since none of the special technical features of the Groups I-VI inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Pushkarev reference, unity of invention is lacking.

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

(特許庁注：以下のものは登録商標)

1. アンドロイド

(74)代理人 230113332

弁護士 山本 健策

(72)発明者 シュナールーレビン, マイケル

アメリカ合衆国 カリフォルニア 94306, パロ アルト, リッキーズ ウェイ 4238 ユニット ティー

(72)発明者 ヤロシュ, ミルナ

アメリカ合衆国 カリフォルニア 94041, マウンテン ビュー, パロ アルト アベニュー 272

(72)発明者 ハインドソン, クリストファー

アメリカ合衆国 カリフォルニア 94566, プレザントン, バーナル アベニュー 4868, アpartment ビー

(72)発明者 ネス, ケビン

アメリカ合衆国 カリフォルニア 94566, プレザントン, キャンピニア プレイス 4306

(72)発明者 サクソノフ, サージ

アメリカ合衆国 カリフォルニア 94619, オークランド, バルモラル ドライブ 5799

(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 94566, プレザントン, テラ コート 1789

(72)発明者 ゼン, グレース エックス. ワイ.

アメリカ合衆国 カリフォルニア 94041, マウンテン ビュー, ブッシュ ストリート 245, アpartment 18

(72)発明者 マークス, パトリック

アメリカ合衆国 カリフォルニア 94110, サンフランシスコ, 21エスティー 3025 アpartment 5

(72)発明者 シュテュルプナーゲル, ジョン

アメリカ合衆国 カリフォルニア 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

Fターム(参考) 4B063 QA01 QA05 QA08 QA13 QA17 QA18 QA19 QA20 QQ02 QQ03

QQ08 QQ42 QQ52 QR32 QR35