

(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2017-527257

(P2017-527257A)

(43) 公表日 平成29年9月21日 (2017.9.21)

(51) Int.Cl.	F I	テーマコード (参考)
C 1 2 Q 1/68 (2006.01)	C 1 2 Q 1/68 A	2 G 0 4 5
G 0 6 F 19/18 (2011.01)	G 0 6 F 19/18	4 B 0 2 9
G 0 1 N 33/50 (2006.01)	G 0 1 N 33/50 J	4 B 0 6 3
C 1 2 M 1/34 (2006.01)	C 1 2 M 1/34 Z	

審査請求 未請求 予備審査請求 未請求 (全 140 頁)

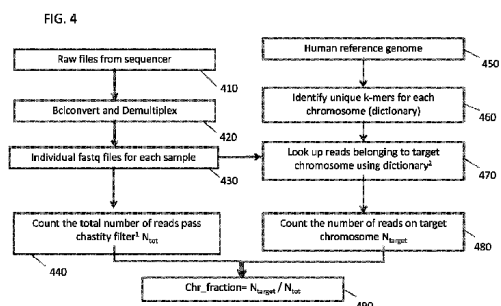
(21) 出願番号	特願2016-569065 (P2016-569065)	(71) 出願人	504159534 セクエノム、 インコーポレイテッド アメリカ合衆国 カリフォルニア 921 21, サン ディエゴ, ジョン ホブ キンズ コート 3595
(86) (22) 出願日	平成27年5月27日 (2015.5.27)	(74) 代理人	100078282 弁理士 山本 秀策
(85) 翻訳文提出日	平成29年1月13日 (2017.1.13)	(74) 代理人	100113413 弁理士 森下 夏樹
(86) 国際出願番号	PCT/US2015/032550	(74) 代理人	100181674 弁理士 飯田 貴敏
(87) 国際公開番号	W02015/183872	(74) 代理人	100181641 弁理士 石川 大輔
(87) 国際公開日	平成27年12月3日 (2015.12.3)	(74) 代理人	230113332 弁護士 山本 健策
(31) 優先権主張番号	62/005,811		
(32) 優先日	平成26年5月30日 (2014.5.30)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 染色体提示の決定

(57) 【要約】

本明細書中に記載される技術は、配列決定プロセスによって生成される配列リードを利用する診断テストに部分的に関する。いくつかの実施形態において、染色体提示を生成するために使用される構成要素は、参照ゲノムにアラインメントされない配列リードのカウンタに基づき得る。ある特定の診断テストは、配列リードの処理を含む。配列リードは、テストサンプル核酸を配列決定プロセスに供することによって生成される比較的短い部分配列（例えば、約20～約40塩基対長）である。



【特許請求の範囲】**【請求項 1】**

診断テストのためにゲノムセグメントの配列リードカウント提示を確定するための方法であって、該方法は、

(a) ゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウント A を提供する工程であって、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、工程；

(b) 該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウント B を提供する工程であって、ここで、該カウント B は、参照ゲノムにアラインメントされない配列リードのカウントである、工程；および

(c) 該カウント A と該カウント B との比率として、該セグメントに対するカウント提示を確定する工程を含む、方法。

【請求項 2】

(b) における前記ゲノムのサブセットが、(a) における前記セグメントより大きい、請求項 1 に記載の方法。

【請求項 3】

前記カウント B が、前記配列リードを参照ゲノムにアラインメントする工程を含まないプロセスによって決定される、請求項 1 または 2 に記載の方法。

【請求項 4】

前記カウント B が、

(i) 前記テストサンプルからの核酸を配列決定するために使用される核酸配列決定プロセスによって生成された全リードのカウント；

(i i) 該テストサンプルからの核酸を配列決定するために使用される核酸配列決定プロセスによって生成された全リードの一部のカウント；

(i i i) 該配列決定プロセスに対する品質管理メトリックに従ってフィルタリングされたリードを除く、(i) の全リードまたは (i i) の全リードの一部のカウント；

(i v) 該配列決定プロセスに対する品質管理メトリックに従って重み付けされた、(i) の全リードまたは (i i) の全リードの一部のカウント；

(v) リードの塩基含有量に従ってフィルタリングされたリードを除く、(i) の全リードまたは (i i) の全リードの一部のカウント；

(v i) リードの塩基含有量に従って重み付けされた、(i) の全リードまたは (i i) の全リードの一部のカウント；または

(v i i) あるリストの中のポリヌクレオチドとマッチするリードのカウントであって、ここで、該リードは、該リストの中のポリヌクレオチドとリードを比較する工程を含むプロセスにおいて、該リストの中のポリヌクレオチドとマッチするかまたはマッチしないと判定され、ここで、該リードは、(i) における全リード、(i i) における全リードの一部、(i i i) の品質管理メトリックに従ってフィルタリングされたリードを除く、(i) の全リードもしくは (i i) の全リードの一部、(i v) の品質管理メトリックに従って重み付けされた、(i) の全リードもしくは (i i) の全リードの一部、(v) のリードの塩基含有量に従ってフィルタリングされたリードを除く、(i) の全リードもしくは (i i) の全リードの一部、または (v i) のリードの塩基含有量に従って重み付けされた、(i) の全リードもしくは (i i) の全リードの一部である、リードのカウントである、請求項 1 ～ 3 のいずれか 1 項に記載の方法。

【請求項 5】

前記一部が、前記全リードからランダムに選択されたリードの一部である、請求項 4 に記載の方法。

【請求項 6】

前記一部が、前記全リードの約 10% ～ 約 90% である、請求項 4 または 5 に記載の方

10

20

30

40

50

法。

【請求項 7】

前記核酸配列決定プロセスが、画像処理を含み、前記品質管理メトリックが、画質に基づく、請求項 4～6 のいずれか 1 項に記載の方法。

【請求項 8】

前記品質管理メトリックが、画像の重なりの評価に基づく、請求項 7 に記載の方法。

【請求項 9】

前記リードの塩基含有量が、グアニンおよびシトシン（G C）含有量である、請求項 4～8 のいずれか 1 項に記載の方法。

【請求項 10】

（v）においてフィルタリングされた前記リードは、第 1 の G C 閾値未満の G C 含有量を有する、請求項 9 に記載の方法。

【請求項 11】

（v）においてフィルタリングされた前記リードが、第 2 の G C 閾値を超える G C 含有量を有する、請求項 9 に記載の方法。

【請求項 12】

（v i i）における前記カウントが、前記リストの中のポリヌクレオチドの配列およびサイズと正確にマッチするリードのカウントである、請求項 4～11 のいずれか 1 項に記載の方法。

【請求項 13】

前記リストの中のポリヌクレオチドが、（a）の前に、参照ゲノムまたは参照ゲノムにおけるサブセットにアラインメントされたものである、請求項 4～12 のいずれか 1 項に記載の方法。

【請求項 14】

前記参照ゲノムにおけるサブセットが、すべての常染色体またはすべての常染色体のサブセットである、請求項 13 に記載の方法。

【請求項 15】

前記比較が、（i）各ポリヌクレオチドがアラインメントする染色体、および／または（i i）各ポリヌクレオチドがアラインメントする染色体の位置番号を追跡する工程を含まない、請求項 13 または 14 に記載の方法。

【請求項 16】

リードを参照ゲノムとアラインメントするアラインメントプロセスに前記リードを供する工程を含み、ここで、前記カウント B は、該リードを該アラインメントプロセスに供する前に決定される、請求項 1～15 のいずれか 1 項に記載の方法。

【請求項 17】

リードを参照ゲノムとアラインメントするアラインメントプロセスに前記リードを供する工程を含み、ここで、前記カウント B は、該アラインメントプロセスによって該参照ゲノムにアラインメントされないリードのカウントである、請求項 1～16 のいずれか 1 項に記載の方法。

【請求項 18】

リードを参照ゲノムとアラインメントするアラインメントプロセスに前記リードを供する工程を含み、ここで、前記カウント A は、該参照ゲノムにおけるセグメントにアラインメントされたリードのカウントである、請求項 1～17 のいずれか 1 項に記載の方法。

【請求項 19】

前記カウント A が、前記配列リードを参照ゲノムにアラインメントする工程を含まないプロセスによって決定される、請求項 1～17 のいずれか 1 項に記載の方法。

【請求項 20】

前記カウント A が、あるリストまたはあるリストのサブセットの中のポリヌクレオチドとマッチするリードのカウントであり、ここで、該リードは、リードを該リストまたは該リストのサブセットの中のポリヌクレオチドと比較する工程を含むプロセスにおいて、該

10

20

30

40

50

リストまたは該リストのサブセットの中のポリヌクレオチドとマッチするかまたはマッチしないと判定される、請求項 19 に記載の方法。

【請求項 21】

前記リストまたは前記リストのサブセットの中のポリヌクレオチドと比較された前記リードが、請求項 4 (i) における全リード；請求項 4 (i i) における全リードの一部；請求項 4 (i i i) の品質管理メトリックに従ってフィルタリングされたリードを除く、請求項 4 (i) の全リードもしくは請求項 4 (i i) の全リードの一部；請求項 4 (i v) の品質管理メトリックに従って重み付けされた、請求項 4 (i) の全リードもしくは請求項 4 (i i) の全リードの一部；請求項 4 (v) のリードの塩基含有量に従ってフィルタリングされたリードを除く、請求項 4 (i) の全リードもしくは請求項 4 (i i) の全リードの一部；または請求項 4 (v i) のリードの塩基含有量に従って重み付けされた、請求項 4 (i) の全リードもしくは請求項 4 (i i) の全リードの一部である、請求項 20 に記載の方法。

10

【請求項 22】

前記カウント A が、前記リストまたは前記リストのサブセットの中のポリヌクレオチドの配列およびサイズと正確にマッチするリードのカウントである、請求項 20 または 21 に記載の方法。

【請求項 23】

前記リストまたは前記リストのサブセットの中のポリヌクレオチドが、(a) の前に、参照ゲノムにおけるセグメントにアラインメントされたものである、請求項 20 ～ 22 のいずれか 1 項に記載の方法。

20

【請求項 24】

前記比較が、(i) 各ポリヌクレオチドがアラインメントする染色体、および／または (i i) 各ポリヌクレオチドがアラインメントする染色体の位置番号の追跡を含まない、請求項 23 に記載の方法。

【請求項 25】

前記配列リードが、(a)、(b) および (c) において前記配列リードを前記参照ゲノムにアラインメントするアラインメントプロセスに供されない、請求項 1 ～ 15 および 19 ～ 24 のいずれか 1 項に記載の方法。

【請求項 26】

前記配列リードが、前記診断テストにおいて該配列リードを前記参照ゲノムにアラインメントするアラインメントプロセスに供されない、請求項 1 ～ 15 および 19 ～ 24 のいずれか 1 項に記載の方法。

30

【請求項 27】

前記セグメントが、染色体である、請求項 1 ～ 26 のいずれか 1 項に記載の方法。

【請求項 28】

前記染色体が、13 番染色体、18 番染色体および 21 番染色体から選択される、請求項 27 に記載の方法。

【請求項 29】

前記セグメントが、染色体のセグメントである、請求項 1 ～ 26 のいずれか 1 項に記載の方法。

40

【請求項 30】

前記セグメントが、微小重複領域または微小欠失領域である、請求項 29 に記載の方法。

【請求項 31】

(c) における前記比率が、前記カウント B で除算された前記カウント A である、請求項 1 ～ 30 のいずれか 1 項に記載の方法。

【請求項 32】

(c) における前記比率が、前記カウント A で除算された前記カウント B である、請求項 1 ～ 30 のいずれか 1 項に記載の方法。

50

【請求項 3 3】

前記核酸が、循環無細胞核酸である、請求項 1 ～ 3 2 のいずれか 1 項に記載の方法。

【請求項 3 4】

前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来である、請求項 1 ～ 3 3 のいずれか 1 項に記載の方法。

【請求項 3 5】

前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてのテストである、請求項 1 ～ 3 3 のいずれか 1 項に記載の方法。

【請求項 3 6】

前記セグメントに対するカウント提示の統計量を決定する工程を含む、請求項 1 ～ 3 5 のいずれか 1 項に記載の方法。

【請求項 3 7】

前記統計量が、 z 得点である、請求項 3 6 に記載の方法。

【請求項 3 8】

前記 z 得点が、(a) (i) 前記テストサンプルに対する前記セグメントに対するカウント提示から (i i) サンプルセットに対する前記セグメントに対するカウント提示の中央値を除いた減算の結果を、(b) 該サンプルセットに対するセグメントに対するカウント提示の M A D で除算した商である、請求項 3 7 に記載の方法。

【請求項 3 9】

前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来であり、前記サンプルセットが、正倍数性胎児の妊娠を有する被験体に対するサンプルのセットである、請求項 1 ～ 3 8 のいずれか 1 項に記載の方法。

【請求項 4 0】

前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来であり、前記サンプルセットが、トリソミー胎児の妊娠を有する被験体に対するサンプルのセットである、請求項 1 ～ 3 8 のいずれか 1 項に記載の方法。

【請求項 4 1】

前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてであり、前記サンプルセットが、該細胞増殖の症状を有する被験体に対するサンプルのセットである、請求項 1 ～ 3 8 のいずれか 1 項に記載の方法。

【請求項 4 2】

前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてであり、前記サンプルセットが、該細胞増殖の症状を有しない被験体に対するサンプルのセットである、請求項 1 ～ 3 8 のいずれか 1 項に記載の方法。

【請求項 4 3】

前記カウント A が、正規化されたカウントのものである、請求項 1 ～ 4 2 のいずれか 1 項に記載の方法。

【請求項 4 4】

前記カウント B が、正規化されたカウントのものである、請求項 1 ～ 4 3 のいずれか 1 項に記載の方法。

【請求項 4 5】

前記正規化されたカウントが、L O E S S 正規化プロセスを含む正規化プロセスによって生成される、請求項 4 3 または 4 4 に記載の方法。

【請求項 4 6】

前記正規化されたカウントが、グアニンおよびシトシン (G C) バイアス正規化を含む正規化プロセスによって生成される、請求項 4 3 ～ 4 5 のいずれか 1 項に記載の方法。

【請求項 4 7】

前記正規化されたカウントが、G C バイアスの L O E S S 正規化 (G C - L O E S S) を含む正規化プロセスによって生成される、請求項 4 3 ～ 4 6 のいずれか 1 項に記載の方法。

10

20

30

40

50

【請求項 48】

前記正規化されたカウントが、主成分正規化を含む正規化プロセスによって生成される、請求項 43～47 のいずれか 1 項に記載の方法。

【請求項 49】

前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来であり、該診断テストは、遺伝的変異の非存在の存在を判定することを含む、請求項 1～48 のいずれか 1 項に記載の方法。

【請求項 50】

前記遺伝的変異が、染色体異数性である、請求項 49 に記載の方法。

【請求項 51】

前記染色体異数性が、染色体全体の 1、3 または 4 コピーである、請求項 50 に記載の方法。

【請求項 52】

前記遺伝的変異が、微小重複または微小欠失である、請求項 49 に記載の方法。

【請求項 53】

前記遺伝的変異が、胎児の遺伝的変異である、請求項 49～52 のいずれか 1 項に記載の方法。

【請求項 54】

前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてであり、該診断テストが、遺伝的変異の非存在の存在を判定することを含む、請求項 1～48 のいずれか 1 項に記載の方法。

【請求項 55】

前記遺伝的変異が、微小重複または微小欠失である、請求項 54 に記載の方法。

【請求項 56】

(a)、(b) および (c) のうちの 1 つもしくはそれを超えるものまたはすべてが、システムにおけるマイクロプロセッサによって行われる、請求項 1～55 のいずれか 1 項に記載の方法。

【請求項 57】

(a)、(b) および (c) のうちの 1 つもしくはそれを超えるものまたはすべてが、システムにおけるメモリーと連動して行われる、請求項 1～56 のいずれか 1 項に記載の方法。

【請求項 58】

(a)、(b) および (c) のうちの 1 つもしくはそれを超えるものまたはすべてが、コンピュータによって行われる、請求項 1～57 のいずれか 1 項に記載の方法。

【請求項 59】

1 つまたはそれを超えるマイクロプロセッサおよびメモリーを備えるシステムであって、該メモリーは、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示を備え、該メモリーは、ヌクレオチド配列リードを備え、該配列リードは、被験体由来のテストサンプルからの核酸のリードであり、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示は、以下：

(a) マイクロプロセッサを用いて、あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウント A を提供することであって、ここで、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、カウント A を提供すること；

(b) マイクロプロセッサを用いて、該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウント B を提供することであって、ここで、該カウント B は、参照ゲノムにアライメントされない配列リードのカウントである、カウント B を提供すること；および

(c) 該カウント A と該カウント B との比率として、該セグメントに対するカウント提示を確定すること

10

20

30

40

50

を行うように構成される、システム。

【請求項 6 0】

1 つまたはそれを超えるマイクロプロセッサおよびメモリーを備える機器であって、該メモリーは、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示を備え、該メモリーは、ヌクレオチド配列リードを備え、該配列リードは、被験体由来のテストサンプルからの核酸のリードであり、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示は、以下：

(a) マイクロプロセッサを用いて、あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウント A を提供することであって、ここで、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、カウント A を提供すること；

(b) マイクロプロセッサを用いて、該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウント B を提供することであって、ここで、該カウント B は、参照ゲノムにアラインメントされない配列リードのカウントである、カウント B を提供すること；および

(c) 該カウント A と該カウント B との比率として、該セグメントに対するカウント提示を確定すること

を行うように構成される、機器。

【請求項 6 1】

実行可能なプログラムが格納されている非一時的なコンピュータ可読記憶媒体であって、該プログラムは、マイクロプロセッサに以下：

(a) 被験体由来のテストサンプルからの核酸のリードであるヌクレオチド配列リードにアクセスすること；

(b) マイクロプロセッサを用いて、あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウント A を提供することであって、ここで、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、カウント A を提供すること；

(c) マイクロプロセッサを用いて、該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウント B を提供することであって、ここで、該カウント B は、参照ゲノムにアラインメントされない配列リードのカウントである、カウント B を提供すること；および

(d) 該カウント A と該カウント B との比率として、該セグメントに対するカウント提示を確定すること

を行うように指示する、非一時的なコンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

関連特許出願

本特許出願は、2014 年 5 月 30 日に出願された、「CHROMOSOME REPRESENTATION DETERMINATIONS」との名称であり、Chen Zhao および Cosmin Deciu を発明者として記載し、そして、代理人管理番号 SEQ-6080-PV によって指定される、米国仮特許出願第 62/005,811 号の利益を主張する。前述の出願の内容全体は、全てのテキスト、表および図面を含めて、本明細書に参考として援用される。

【0 0 0 2】

分野

本明細書中に記載される技術は、配列決定プロセスによって生成された配列リードを利用する診断テストに部分的に関する。いくつかの実施形態において、染色体提示 (chromosome representation) を生成するために使用される構成要素は、参照ゲノムにアラインメントされない配列リードのカウントに基づき得る。

【背景技術】

【0003】

背景

生命体（例えば、動物、植物および微生物）および遺伝情報を複製する他の形態（例えば、ウイルス）の遺伝情報は、デオキシリボ核酸（DNA）またはリボ核酸（RNA）にコードされている。遺伝情報は、化学的核酸または仮説的核酸の1次構造に相当するひと続きのヌクレオチドまたは修飾ヌクレオチドである。ヒトの全ゲノムは、二十四（24）本の染色体上に位置づけられた約30,000種の遺伝子を含んでいる（The Human Genome, T. Strachan, BIOS Scientific Publishers, 1992を参照のこと）。各遺伝子は、特定のタンパク質をコードしており、そのタンパク質は、転写および翻訳を介した発現の後、生細胞内で特定の生化学的機能を果たす。

10

【0004】

多くの医学的症状が、1つまたはそれを超える遺伝的変異によって引き起こされる。ある特定の遺伝的変異は、例えば、血友病、サラセミア、デュシェンヌ型筋ジストロフィー（DMD）、ハンチントン病（HD）、アルツハイマー病および嚢胞性線維症（CF）をはじめとした医学的症状を引き起こす（Human Genome Mutations, D. N. Cooper and M. Krawczak, BIOS Publishers, 1993）。そのような遺伝性疾患は、特定の遺伝子のDNAにおける単一ヌクレオチドの付加、置換または欠失によって生じ得る。ある特定の出生時欠損は、異数性、例えば、トリソミー21（ダウン症候群）、トリソミー13（パター症候群）、トリソミー18（エドワーズ症候群）、モノソミーX（ターナー症候群）、およびある特定の性染色体異数性、例えば、クラインフェルター症候群（XXY）とも称される染色体異常によって引き起こされる。別の遺伝的変異は、性染色体XおよびYに基づいて判定され得ることが多い胎児の性別である。いくつかの遺伝的変異は、個体をいくつかの疾患（例えば、糖尿病、動脈硬化症、肥満症、様々な自己免疫疾患および癌（例えば、直腸結腸癌、乳癌、卵巣癌、肺癌））のいずれかにさせ得るか、またはそれらを引き起こし得る。

20

【0005】

1つまたはそれを超える遺伝的変異（例えば、コピー数変異）または遺伝分散が特定されると、特定の医学的症状を診断することができるか、または特定の医学的症状に対する素因を判定することができる。遺伝分散が特定されると、医学的決断を促すことができ、および／または有益な医学的手技を用いることができる。ある特定の実施形態において、1つまたはそれを超える遺伝的変異または遺伝分散の特定には、無細胞DNAの解析が必要である。無細胞DNA（CF-DNA）は、細胞死に由来し、末梢血の中を循環しているDNAフラグメントから構成されている。高濃度のCF-DNAは、ある特定の臨床症状、例えば、癌、外傷、熱傷、心筋梗塞、脳卒中、敗血症、感染症および他の疾病を示唆することができる。さらに、無細胞胎児DNA（CFF-DNA）は、母体の血流中で検出され得、様々な非侵襲的出生前診断に使用され得る。

30

【先行技術文献】

【非特許文献】

40

【0006】

【非特許文献1】The Human Genome, T. Strachan, BIOS Scientific Publishers, 1992

【非特許文献2】Human Genome Mutations, D. N. Cooper and M. Krawczak, BIOS Publishers, 1993

【発明の概要】

【課題を解決するための手段】

【0007】

要旨

ある特定の態様において、診断テストのためにゲノムセグメントの配列リードカウント

50

提示を確定するための方法が、本明細書中に提供され、その方法は、(a) あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、そのセグメントに対するカウントAを提供する工程であって、その配列リードは、そのゲノムを有する被験体由来のテストサンプルからの核酸のリードである、工程；(b) そのゲノムまたはそのゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、そのゲノムまたはそのゲノムのサブセットに対するカウントBを提供する工程であって、ここで、カウントBは、参照ゲノムにアラインメントされない配列リードのカウントである、工程；および(c) カウントAとカウントBとの比率として、そのセグメントに対するカウント提示を確定する工程を含む。

【0008】

本技術のある特定の態様は、以下の説明、実施例、請求項および図面においてさらに説明される。

【図面の簡単な説明】

【0009】

図面は、本技術の実施形態を例証するものであって、限定するものではない。例証を明確にするためおよび平易にするために、図面は、一定尺度で拡大縮小して作成されておらず、場合によっては、特定の実施形態の理解を促すために、様々な態様が誇張されてまたは拡大されて示されていることがある。

【0010】

【図1】図1は、リードの総数（アラインメント前）とchastityフィルターを通過するリードの総数（アラインメント前）との比較を示している。

【0011】

【図2】図2は、chastityフィルターを通過するリードの総数（アラインメントの前）とすべての常染色体にアラインメントされたリードの総数との比較を示している。

【0012】

【図3】図3A、図3Bおよび図3Cは、21、13および18番染色体に対してSPCA正規化を用いたときの、常染色体を用いて算出された染色体提示およびchastityフィルターを通過するアラインメント前のリードを用いて算出された染色体提示から導かれたz得点の比較を示している。

【0013】

【図4】図4は、特定の標的染色体に対するカウント提示を生成するためにポリヌクレオチドのサブリストを使用する非限定的な例を示している。

【0014】

【図5】図5は、本技術のある特定の実施形態が履行され得るシステムの例証的な実施形態を示している。

【発明を実施するための形態】

【0015】

詳細な説明

ある特定の診断テストは、配列リードの処理を含む。配列リードは、テストサンプル核酸を配列決定プロセスに供することによって生成される比較的短い部分配列（例えば、約20～約40塩基対長）である。いくつかの診断テストが、染色体カウント提示を確定することを必要とし、その染色体カウント提示は、試験染色体に帰されるカウント数の正規化されたバージョンである。染色体カウント提示は、(i) 試験染色体に帰される配列リードの数(Ntest)と、(ii) ゲノム（例えば、ヒト常染色体ならびに性染色体XおよびY）、または染色体（例えば、常染色体）より大きいゲノムのサブセットに対する配列リードの数(NrefまたはNtot)との比率として表現されるときがある。そのNtestおよびNref値は、染色体カウント提示を確定するとき、参照ゲノムに対してアラインメントされたかまたはマッピングされたリードの数をカウントすることによって決定されるときがある。

【0016】

本明細書の以後にさらに詳細に記載されるように、N t e s t および／または N r e f (それぞれカウントAおよびカウントBとも称される)は、配列リードを参照ゲノムとアラインメントせずに決定され得ることが明らかになった。さらに、本明細書中に記載される方法は、ゲノムセグメントに対するカウント提示を生成するために広く使用することができ、ここで、そのセグメントは、標的染色体よりも小さいかもしくは大きい、または標的染色体と同じサイズおよび配列を有する。

【0017】

したがって、ある特定の実施形態において、診断テストのためにゲノムセグメント(すなわち、標的セグメント)の配列リードカウント提示を確定するための方法が提供され、その方法は、(a)ゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、そのセグメントに対するカウントAを提供する工程であって、その配列リードは、そのゲノムを有する被験体由来のテストサンプルからの核酸のリードである、工程；(b)そのゲノムまたはそのゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、そのゲノムまたはそのゲノムのサブセットに対するカウントBを提供する工程であって、ここで、そのカウントAは、参照ゲノムにアラインメントされない配列リードのカウントであり、かつ／またはそのカウントBは、参照ゲノムにアラインメントされない配列リードのカウントである、工程；および(c)カウントAとカウントBとの比率として、そのセグメントに対するカウント提示を確定する工程を含む。

【0018】

任意の好適なサンプルが、本明細書中に記載される方法のために使用され得る。サンプルは、任意の好適な被験体(例えば、ヒト、類人猿、有蹄動物、ウシ、ヒツジ、ウマ、ヤギ、イヌ、ネコ、鳥類、爬虫類、家畜動物など)に由来し得る。サンプルは、任意の妊娠段階(例えば、ヒト被験体の場合、妊娠第1期、第2期または第3期)の、胎児を有する妊婦被験体に由来するときもあるし、出生後の被験体に由来するときもある。サンプルは、すべての染色体が正倍数性の胎児を有する妊娠中の被験体に由来するときもあるし、染色体異数性(例えば、1、3(すなわち、トリソミー(例えば、T21、T18、T13))または4コピーの染色体)または他の遺伝的変異を有する胎児を有する妊娠中の被験体に由来するときもある。サンプルは、細胞増殖の症状(c e l l p r o l i f e r a t i v e c o n d i t i o n)を有する被験体であるときもあるし、細胞増殖の症状を有しない被験体に由来するときもある。細胞増殖の症状の非限定的な例としては、肝臓細胞(例えば、肝細胞)、肺細胞、脾臓細胞、膵臓細胞、結腸細胞、皮膚細胞、膀胱細胞、眼細胞、脳細胞、食道細胞、頭部の細胞、頸部の細胞、卵巣の細胞、精巣の細胞、前立腺細胞、胎盤細胞、上皮細胞、内皮細胞、脂肪細胞、腎臓／腎細胞、心臓細胞、筋細胞、血液細胞(例えば、白血球)、中枢神経系(C N S)細胞などおよび前述の細胞の組み合わせの癌、腫瘍および制御されていない(d i s - r e g u l a t e d)細胞増殖の症状が挙げられる。解析される核酸は、好適なサンプル(例えば、頬側細胞、生検組織または生検細胞、胎児の細胞)から単離された細胞の核酸であるときがある。解析される核酸は、好適なサンプル(例えば、血清、血漿、尿または他の体液)から単離された循環無細胞(c c f)核酸であるときがある。核酸単離プロセスは、当該分野において利用可能であり、公知である。

【0019】

診断テストのために核酸を配列決定するのに適したプロセスは、当該分野で公知であり、大規模並列処理配列決定(M P S)プロセスが用いられるときがある。配列決定プロセスの非限定的な例としては、I l l u m i n a / S o l e x / H i S e q (例えば、I l l u m i n a G e n o m e A n a l y z e r ; G e n o m e A n a l y z e r I I ; H I S E Q 2 0 0 0 ; H I S E Q)、S O L i D、R o c h e / 4 5 4、P A C B I O および／または S M R T、H e l i c o s T r u e S i n g l e M o l e c u l e S e q u e n c i n g、I o n T o r r e n t および I o n 半導体ベースの配列決定、W i l d F i r e、5 5 0 0、5 5 0 0 x 1 W および／または 5 5 0 0 x 1 W G e n e t i c A n a l y z e r ベースの技術；ポロニーシーケンシング、パイロシ

10

20

30

40

50

ーケンシング、大規模並列処理シグネチャ配列決定 (Massively Parallel Signature Sequencing) (MPSS)、RNAポリメラーゼ (RNAP) 配列決定、LaserGenシステムおよび方法、ナノポアベースのプラットフォーム、化学感応性電界効果トランジスタ (CHEM-FET) アレイ、電子顕微鏡法ベースの配列決定 (例えば、ZS Genetics、Halcyon Molecular) ならびにナノボールシーケンシングが挙げられる。ある特定の配列決定プロセスは、1つまたはそれを超える核酸増幅プロセスと組み合わせて履行され、その核酸増幅プロセスの非限定的な例としては、ポリメラーゼ連鎖反応 (PCR; AFLP-PCR、対立遺伝子特異的PCR、Alu-PCR、非対称PCR、コロニーPCR、ホットスタートPCR、インバースPCR (IPCR)、インサイチュPCR (ISH)、配列間特異的PCR (Intersequence-specific PCR) (ISSR-PCR)、ロングPCR、マルチプレックスPCR、ネステッドPCR、定量的PCR、逆転写酵素PCR (RT-PCR)、リアルタイムPCR、単一細胞PCR、固相PCR); ライゲーション増幅 (またはリガーゼ連鎖反応 (LCR)); Q-ベータレプリカーゼまたは鋳型依存性ポリメラーゼの使用に基づく増幅方法; ヘリカーゼ依存性等温増幅; 鎖置換増幅 (SDA); 好熱性SDA核酸配列ベースの増幅 (3SRまたはNASBA); 転写関連増幅 (transcription-associated amplification) (TAA); などおよびそれらの組み合わせが挙げられる。診断テストに対して十分なカバレッジデプス (depth of coverage) を提供する配列決定プロセスが通常使用され、その配列決定プロセスは、サンプルに対して約0.1倍~約60倍のカバレッジ (coverage) (例えば、約0.25倍、0.5倍、0.75倍、1倍、2倍、5倍、10倍、12倍、15倍、20倍、25倍、30倍、35倍、40倍、45倍、50倍、55倍のカバレッジ) を提供するときがある。ある1つのサンプルに対して1回またはそれを超える配列決定ラン (例えば、1、2、3、4または5回のラン) を用いて配列決定プロセスが行われ得る。

【0020】

配列リードは、一般にポリヌクレオチドの提示である。例えば、ポリヌクレオチドにおける配列のATGC描写を含むリードにおいて、「A」は、アデニンヌクレオチドを表し、「T」は、チミンヌクレオチドを表し、「G」は、グアニンヌクレオチドを表し、「C」は、シトシンヌクレオチドを表す。配列リードは、ペアエンド (paired-end) リードであるときもあるし、シングルエンド (single-end) リードであるときもある。シングルエンドリードの名目上の長さ、平均の長さ、長さの平均値、長さの中央値または絶対的な長さは、約15個連続したヌクレオチド~約50個またはそれより多く連続したヌクレオチド、約15個連続したヌクレオチド~約40個連続したヌクレオチドであるときもあるし、約15個連続したヌクレオチド~約36個連続したヌクレオチドであるときもある。シングルエンドリードの名目上の長さ、平均の長さ、長さの平均値、長さの中央値または絶対的な長さは、約20~約30塩基長または約24~約28塩基長であるときもあるし、シングルエンドリードの名目上の長さ、平均の長さ、長さの平均値または絶対的な長さは、約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、21、22、23、24、25、26、27、28もしくは約29塩基長またはそれを超える長さであるときもある。ペアエンドリードの名目上の長さ、平均の長さ、長さの平均値または絶対的な長さは、約10個連続したヌクレオチド~約25個連続したヌクレオチドまたはそれを超えるヌクレオチド (例えば、約10、11、12、13、14、15、16、17、18、19、20、21、22、23、24または25ヌクレオチド長またはそれを超える長さ)、約15個連続したヌクレオチド~約20個連続したヌクレオチドであるときもあるし、約17個連続したヌクレオチドまたは約18個連続したヌクレオチドであるときもある。配列リードに関する情報は、好適な形式を有する1つまたはそれを超えるコンピュータ可読ファイルに含められ得、そのコンピュータ可読ファイルの非限定的な例は、BAM、SAM、SRF、FASTQ、Gzipなどおよびそれらの組み合わせを含む、バイナリー形式および/またはテ

キスト形式である。

【0021】

カウントAは、配列リードを参照ゲノムにアラインメントする工程を含まないプロセスによって決定されるときがあり、カウントBは、配列リードを参照ゲノムにアラインメントする工程を含まないプロセスによって決定されることが多い。診断テストは、カウントBが決定された後に、および／または時折、カウントAが決定された後に、配列リードを参照ゲノムにアラインメントする工程を含み得る。配列リードを参照ゲノムにアラインメントする（例えば、マッピングする）のに適したプロセスは、公知であり、それらのプロセスとしては、BLAST、BLITZ、FASTA、BOWTIE 1、BOWTIE 2、ELAND、MAQ、PROBEMATCH、SOAPまたはSEQMAP、DRAGENなどまたはそれらの変法もしくは組み合わせが挙げられるが、これらに限定されない。参照ゲノムは、当該分野で公知のとおり入手することができ、例えば、GenBank、dbEST、dbSTS、EMBL (European Molecular Biology Laboratory) およびDDBJ (DNA Databank of Japan) データベースにおいて入手することができる。配列リードと参照ゲノムとのアラインメントは、100%の配列マッチであり得る。配列リードのアラインメントは、100%未満の配列マッチ（すなわち、不完全なマッチ、部分的なマッチ、部分的なアラインメント）に対応するときがあり、約99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、89%、88%、87%、86%、85%、84%、83%、82%、81%、80%、79%、78%、77%、76%または75%のマッチであるときがある。したがって、配列リードのアラインメントは、ミスマッチに対応するときがあり、1、2、3、4または5個のミスマッチに対応するときがある。アラインメントプロセスは、配列リードがアラインメントする参照ゲノムの場所（例えば、リードがアラインメントする染色体番号；リードがアラインメントする染色体位置）に関する情報を含むかまたは追跡することが多く、そのような情報は、アラインメントが完了した後に1つまたはそれを超えるコンピュータ可読ファイルに格納され得る。

【0022】

配列リード（例えば、アラインメントされたリードまたはアラインメントされないリード）は、当該分野で公知の任意の好適なカウント法によってカウントされ得る。カウントBは、核酸配列決定プロセスによって生成された全リードであるときがあるか、または核酸配列決定プロセスによって生成された全リードの一部であるときがある。本明細書中で述べられるように、カウントBは、(i) それらのリードの特徴に従ってフィルタリングされたリードを除く、または(ii) それらのリードの特徴に従って重み付けされた、全リードまたは全リードの一部のカウントであるときがある。それらのリードの特徴は、フィルタリングまたは重み付けに対する任意の好適な特徴であり得、その特徴の非限定的な例としては、リードの品質およびリードの塩基含有量が挙げられる。リードの塩基含有量は、リードのヌクレオチド塩基の組成および／またはリードのヌクレオチド塩基の複雑さであるときがある。また、本明細書中で述べられるとき、カウントAおよび／またはカウントBは、ある辞書の中のポリヌクレオチドとマッチするリードのカウントであるときがあり、そのような辞書は、本明細書中でポリヌクレオチドのリストまたはサブリストとも称される。カウントAおよび／またはカウントBは、ある特定の実施形態において、アラインメントされたリードが不釣合に低いカバレッジもしくは不釣合に高いカバレッジを有すると確認された、参照ゲノムの1つもしくはそれを超える領域とアラインメントされたリードを除去するフィルターに従ってフィルタリングされた、全リードまたは全リードの一部のカウントである。

【0023】

いくつかの実施形態において、カウントBは、(i) テストサンプルからの核酸を配列決定するために使用される核酸配列決定プロセスによって生成された全リードのカウント；(ii) テストサンプルからの核酸を配列決定するために使用される核酸配列決定プロセスによって生成された全リードの一部のカウント；(iii) 配列決定プロセスに対す

る品質管理メトリックに従ってフィルタリングされたリードを除く、(i)の全リードまたは(ii)の全リードの一部のカウント；(iv)配列決定プロセスに対する品質管理メトリックに従って重み付けされた、(i)の全リードまたは(ii)の全リードの一部のカウント；(v)リードの塩基含有量に従ってフィルタリングされたリードを除く、(i)の全リードまたは(ii)の全リードの一部のカウント；(vi)リードの塩基含有量に従って重み付けされた、(i)の全リードまたは(ii)の全リードの一部のカウント；(vii)リストの中のポリヌクレオチドとマッチするリードのカウントであって、ここで、それらのリードは、そのリストの中のポリヌクレオチドとリードを比較する工程を含むプロセスにおいて、そのリストの中のポリヌクレオチドとマッチするかまたはマッチしないと判定され、ここで、そのリードは、(i)における全リード、(ii)における全リードの一部、(iii)の品質管理メトリックに従ってフィルタリングされたリードを除く、(i)の全リードもしくは(ii)の全リードの一部、(iv)の品質管理メトリックに従って重み付けされた、(i)の全リードもしくは(ii)の全リードの一部、(v)のリードの塩基含有量に従ってフィルタリングされたリードを除く、(i)の全リードもしくは(ii)の全リードの一部、または(vi)のリードの塩基含有量に従って重み付けされた、(i)の全リードもしくは(ii)の全リードの一部；(viii)など、または(ix)前述のものの組み合わせ（例えば、(i)、(ii)、(iii)、(iv)、(v)、(vi)および(vii)のうちの2つまたはそれを超えるもの）である。

10

【0024】

20

いくつかの実施形態において、カウントAは、あるリストまたはリストのサブセットの中のポリヌクレオチドとマッチするリードのカウントであり、それらのリードは、そのリストまたはそのリストのサブセットの中のポリヌクレオチドとリードを比較する工程を含むプロセスにおいて、そのリストまたはそのリストのサブセットの中のポリヌクレオチドとマッチするかまたはマッチしないと判定される。そのリストまたはそのリストのサブセットの中のポリヌクレオチドとの比較のために使用されるリードは、(i)における全リード、(ii)における全リードの一部、(iii)の品質管理メトリックに従ってフィルタリングされたリードを除く、(i)の全リードもしくは(ii)の全リードの一部、(iv)の品質管理メトリックに従って重み付けされた、(i)の全リードもしくは(ii)の全リードの一部、(v)のリードの塩基含有量に従ってフィルタリングされたリードを除く、(i)の全リードもしくは(ii)の全リードの一部、または(vi)のリードの塩基含有量に従って重み付けされた、(i)の全リードもしくは(ii)の全リードの一部であるリードであるときがあり、ここで、(i)、(ii)、(iii)、(iv)、(v)および(vi)は、前述のパラグラフに記載されている。

30

【0025】

ある特定の実施形態において、カウントAは、参照ゲノムにおける標的セグメントにアラインメントされたリードに従って決定される。その参照ゲノムにおける標的セグメントにアラインメントされたリードの数は、カウントすることができ、得られたそのセグメントに対する総カウントは、カウントAとして使用され得る。全リードのカウントの一部も使用され得、全リードまたは全リードの一部は、カウントAを決定するために本明細書中に記載されるようにフィルタリングされるかまたは重み付けされるときがある。例えば、参照ゲノムの標的セグメントの中の領域にアラインメントされたリードのカバレッジが測定され得、不釣合に少ないまたは不釣合に多数のリードによってカバーされた1つまたはそれを超える領域が、特定され得る。そのような1つまたはそれを超える領域由来のリードは、フィルタリングされ、ある特定の実施形態では、カウントAを決定するために、そのセグメントに対するリードの総カウントから除去される。

40

【0026】

カウントBが配列決定プロセスによって生成された全リードのカウントである実施形態の場合、その全リードは、通常、フィルタリングされない（例えば、それらのリードのいずれもが、1つまたはそれを超える基準に従って除去されない）。そのような実施形態に

50

において、全リードはまた、通常、重み付けされない（例えば、それらのリードのいずれもが、1つまたはそれを超える基準に基づく（base on）重み付け因子によって増大されない）。

【0027】

カウントBが配列決定プロセスによって生成された全リードの一部のカウントである実施形態の場合、その一部は、その全リードからランダムに選択されたリードの一部であることが多い。そのような実施形態において、その一部は、全リードの約10%～約90%（例えば、全リードの約15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%または85%）であるときがある。全リードの約50%～約80%が、カウントBのためにカウントされるときがある。カウントBが配列決定プロセスによって生成された全リードのカウントの一部である実施形態の場合、その全リードの一部は、通常、フィルタリングされず、通常、重み付けされない。

10

【0028】

カウントBが、（i）配列決定プロセスに対する品質管理メトリックに従ってフィルタリングされたリードを除く、または（ii）配列決定プロセスに対する品質管理メトリックに従って重み付けされた、全リードまたは全リードの一部のカウントである、実施形態の場合、それらの配列リードを生成する核酸配列決定プロセスは、画像処理を含むときがあり、その品質管理メトリックは、画質に基づく。画像処理を利用してリードを生成するMPSプロセスの非限定的な例は、Illumina HiSeq/TruSeqプロセスである。簡潔には、固相に捕捉された核酸クラスターの画像が、合成による配列決定プロセスの各合成工程において捕捉される。画質は、必要に応じて、1つのクラスターによって生成された画像が別のクラスターの画像と重なっているかまたは重なっていないかに従う品質管理メトリック（例えば、Chastityフィルターによって使用されるメトリック）によって評価され得る。したがって、いくつかの実施形態において、品質管理メトリックは、画像の重なりの評価に基づくときがある。1つのクラスターが別のクラスターと重なっているかまたは重なっていないかに基づく画像の質は、画像スコアリングモジュールによって割り当てられるスコアを用いて評価され得る。いくつかの実施形態において、不良なスコアを割り当てられたクラスターに帰されるリードを全リードまたは全リードの一部から除外するために、フィルターモジュールが使用される。ある特定の実施形態において、画像スコアリングモジュールによって割り当てられたそれらの関連するスコアによって特定のリードまたは特定のリードのカウントを増大させるために重み付けモジュールが使用され、それにより、リードが重み付けされ、その重み付けされたリードまたは重み付けされたリードカウントは、セグメントカウント提示を生成するために使用され得る。

20

30

【0029】

カウントBが、（i）リードの塩基含有量（例えば、塩基組成）に従ってフィルタリングされたリードを除く、または（ii）リードの塩基含有量に従って重み付けされた、全リードまたは全リードの一部のカウントである、実施形態の場合、任意の好適なタイプのリードの塩基含有量が、使用され得る。DNAにおける4種の塩基（A、T、CまたはG）の各々またはそれらの組み合わせの含有量が、リードの塩基含有量によるフィルタリングまたは重み付けのために使用され得る。フィルタリングまたは重み付けのために使用されるリードの塩基含有量は、グアニンおよびシトシン（GC）含有量であるときがある。塩基含有量（例えば、GC含有量）の量は、塩基含有量モジュールによって各リードに割り当てられ得、その量は、任意の好適な様式（例えば、パーセントGC含有量、GCスコア）で表現され得る。いくつかの実施形態において、塩基含有量は、リードにおける塩基リピートまたはポリヌクレオチドリピートの数（例えば、リードにおける連続した一続きのG塩基；リードにおける3つのGCCGポリヌクレオチドリピート）によって評価され、リピートスコアまたはリピート値（例えば、%反復エレメント）が、リピートスコアリングモジュールによって各リードに割り当てられ得る。塩基含有量モジュールおよびリピ

40

50

ートスコアリングモジュールは、集合的に、塩基含有量モジュールと称される。いくつかの実施形態において、塩基含有量モジュールからの塩基含有量の評価またはスコアに基づいて全リードまたは全リードの一部からリードを除外するために、塩基含有量フィルターモジュールが使用される。いくつかの実施形態において、リードは、それらのリードが、
(i) 第 1 の塩基含有量閾値 (例えば、約 40 % の GC 含有量またはそれ未満という第 1 の閾値 (例えば、約 30 % の GC 含有量という第 1 の閾値)) 未満の塩基含有量 (例えば、GC 含有量) を有するか、および / または (i i) 第 2 の塩基含有量閾値 (例えば、約 60 % の GC 含有量またはそれを超える量という第 2 の閾値 (例えば、約 70 % の GC 含有量という第 2 の閾値)) を超える塩基含有量 (例えば、GC 含有量) を有するかに基づいて、全リードまたは全リードの一部からフィルタリングされて除去される。いくつかの実施形態において、リードは、それらのリードが、リピート含有量閾値 (例えば、約 50 % のリピートという閾値) を超えるリピート含有量 (例えば、塩基リピート含有量) を有するかに基づいて、全リードまたは全リードの一部からフィルタリングされて除去される。ある特定の実施形態において、リピートスコアリングモジュールまたは塩基含有量モジュールによって割り当てられたそれらの関連するスコアまたは値によって、特定のリードまたは特定のリードのカウントを増大させるために重み付けモジュールが使用され、それにより、リードが重み付けされ、その重み付けされたリードまたは重み付けされたリードカウントは、セグメントカウント提示を生成するために使用され得る。

【 0030 】

リードが、あるリストまたはそのリストのサブセット (すなわち、サブリスト) の中のポリヌクレオチドとマッチするかまたはマッチしないか判定される、実施形態の場合、カウント A および / またはカウント B は、そのリストまたはサブリストの中のポリヌクレオチドの配列およびサイズと正確にマッチするリードのカウントであることが多い。サンプル (例えば、テストサンプルではないサンプル) からのリードと参照ゲノムまたは参照ゲノムにおけるサブセットとのアラインメントに基づいて、ポリヌクレオチドをあるリストまたはサブリストのために選択した後、テストサンプルリードをそれらのポリヌクレオチドと比較し、マッチするテストサンプルリードをカウントすることが多い。この事前のアラインメントにおいてアラインメントされたリードは、通常、そのリストまたはサブリストの中のポリヌクレオチドに対応する (例えば、そのリストまたはサブリストの中のポリヌクレオチドと同じである) 。特定のセグメントまたは領域にユニークにアラインメントするリードが、そのリストまたはサブリストの中にポリヌクレオチドとして含めるために選択されることが多い。例えば、参照ゲノムにおける標的セグメント (例えば、標的染色体) にアラインメントし、参照ゲノムにおける他のセグメントにアラインメントしない (例えば、他の染色体にアラインメントしない) リードが、サブリストの中にポリヌクレオチドとして含めるために選択されることが多い。

【 0031 】

カウント B を決定する場合、リストは、事前のアラインメントにおいて参照ゲノムにおけるすべての染色体、すべての常染色体またはすべての常染色体のサブセットにアラインメントしたリードに対応するポリヌクレオチドを含むときがある。カウント A を決定する場合、事前のアラインメントにおいて、参照ゲノムの中の、カウント提示が確定される標的セグメント (例えば、標的セグメントとしての標的染色体) にアラインメントしたリードに対応するポリヌクレオチドを含むサブリストが使用されることが多い。いくつかの実施形態において、あるリストおよびサブリストが使用され、ここで、そのリストは、カウント B を決定するために使用され得るすべての常染色体にマッピングされたポリヌクレオチドを含み、そのサブリストは、カウント A を決定するために使用され得るセグメントにマッピングされたポリヌクレオチドを含む。したがって、ある特定の実施形態において、カウント A およびカウント B は、テストサンプルからのリードを参照ゲノムにアラインメントせずに、標的セグメントに対するカウント提示を生成するために決定され得る。特定の標的染色体に対するカウント提示を生成するためにポリヌクレオチドのサブリストを使用する非限定的な例は、図 4 に例証されており、実施例 2 に記載される。

【0032】

あるリストまたはサブリストの中のポリヌクレオチドとリードを比較するために使用されるプロセス（比較）は、通常、参照ゲノムにリードをアラインメントするために使用されるプロセス（アラインメント）とは異なる。例えば、比較のために使用されるプロセスは、（i）各リードもしくは各ポリヌクレオチドがアラインメントする染色体、および／または（ii）各リードもしくは各ポリヌクレオチドがアラインメントする染色体の位置番号に関する情報を追跡または記録しないことが多い。また、比較のために使用されるプロセスは、バイナリーであることが多く、例えば、リードの配列および長さが、リストおよび／またはサブリストの中のポリヌクレオチドと100%マッチであるかまたはそうではないかを評価し得る。アラインメントプロセスは、複雑度がより高いアルゴリズムを使用することが多いので、バイナリープロセスは、通常、リードを参照ゲノムにアラインメントするためのプロセスよりも複雑でない。

10

【0033】

テストサンプル核酸から生成されたリードは、（i）カウントAおよび／またはカウントBを生成する前に、配列リードを参照ゲノムにアラインメントするアラインメントプロセスに供されないときがあるか；（ii）行われている診断テストにおいて配列リードを参照ゲノムにアラインメントするアラインメントプロセスに供されないときがあるか；または（iii）リードを参照ゲノムとアラインメントするアラインメントプロセスに供されるときがあり、ここで、カウントAおよび／またはカウントBは、リードをアラインメントプロセスに供する前に決定される。いくつかの実施形態において、テストサンプル核酸に対して生成されたリードは、リードを参照ゲノムとアラインメントするアラインメントプロセスに供され、カウントAは、その参照ゲノムにおけるセグメントにアラインメントされたリードのカウントであり、カウントBは、その参照ゲノムにアラインメントされないリードのカウントであるか、または参照ゲノムへのリードのアラインメントの前に決定される。いくつかの実施形態において、カウントAおよび／またはカウントBは、配列リードを参照ゲノムにアラインメントする工程を含まないプロセスによって決定される。

20

【0034】

ある特定の実施形態において、テストサンプルから生成されたリードは、リードを参照ゲノムとアラインメントするアラインメントプロセスに供され、カウントBは、そのアラインメントプロセスによって参照ゲノムにアラインメントされないリードのカウントである。参照ゲノムにアラインメントされ得ないリード（アラインメント不可能なリード）は、反復ポリヌクレオチドを含むおよび／またはセントロメアが起源であるリードであるときがある。

30

【0035】

いくつかの実施形態において、カウント提示が確定される標的セグメントは、染色体であり、その染色体は、13番染色体、18番染色体および21番染色体であるときがある。そのセグメントは、染色体のセグメントであるときがあり、微小重複領域または微小欠失領域であるときがある。

【0036】

ある特定の実施形態において、カウントAは、正規化されたカウントであり、かつ／またはカウントBは、正規化されたカウントである。任意の好適な正規化プロセスまたは好適な正規化プロセスの組み合わせが、正規化されたカウントを生成するために使用され得る。正規化プロセスの非限定的な例としては、部分ごとの（portion-wise）正規化（例えば、ビンごとの（bin-wise）正規化）、GC含有量による正規化、線形および非線形最小二乗回帰、LOESS、GC-LOESS、LOWESS、PERUN、ChAI、RM、GCRM、cQnなど、ならびにそれらの組み合わせが挙げられる。正規化されたカウントは、（i）LOESS正規化プロセスを含む正規化プロセス、（ii）グアニンおよびシトシン（GC）バイアス正規化を含む正規化プロセス、（iii）GCバイアスのLOESS正規化（GC-LOESS）を含む正規化プロセス、（iv）主成分正規化（例えば、ChAI正規化プロセス）を含む正規化プロセスなどおよび

40

50

前述のものの組み合わせによって生成されるときがある。いくつかの実施形態において、正規化プロセスは、GC-L O E S S 正規化に続く主成分正規化を含む。ある特定の正規化プロセス（例えば、C h A I 正規化、主成分正規化、P E R U N 正規化）の具体的な態様は、例えば、2014年5月23日に出願され、W O 2 0 1 4 / 1 9 0 2 8 6 として公開された特許出願番号 P C T / U S 2 0 1 4 / 0 3 9 3 8 9 ; および2014年10月2日に出願され、2015年4月9日にW O 2 0 1 5 / 0 5 1 1 6 3 として公開された特許出願番号 P C T / U S 2 0 1 4 / 0 5 8 8 8 5 に記載されている。

【0037】

いくつかの実施形態において、主成分正規化を含む正規化プロセスは、(a) 複数のサンプルに対して作成されたリード密度分布に従ってフィルタリングすることによって生成され得るリード密度プロファイルを提供する工程、および(b) 主成分分析によって1つまたはそれを超える主成分に従ってテストサンプルに対するリード密度プロファイルを調整し(それらの主成分は、参照サンプルのセットから得られる)、それにより、調整されたリード密度を含むテストサンプルプロファイルを提供する工程を含む。

10

【0038】

いくつかの実施形態において、P E R U N 正規化を含む正規化プロセスは、(1) (i) 各部分にマッピングされた配列リードのカウントと(ii) 各部分に対するGC含有量との間の当てはめ関係(f i t t e d r e l a t i o n)に基づいて、テストサンプルに対するグアニンおよびシトシン(GC) バイアス係数を決定する工程(ここで、そのGC バイアス係数は、線形当てはめ関係の場合、傾きであるか、または非線形当てはめ関係の場合、曲率推定値(c u r v a t u r e e s t i m a t i o n)である); および(2) (a) のカウント、(b) のGC バイアス係数、および(i) 複数のサンプルの各々に対するGC バイアス係数と(ii) それらの複数のサンプルに対する各部分にマッピングされた配列リードのカウントとの間の各部分に対する当てはめ関係に基づいて、各部分に対するゲノム区分レベルを、マイクロプロセッサを用いて算出し、それにより、算出されたゲノム区分レベルを提供する工程を含む。

20

【0039】

いくつかの実施形態において、ある診断方法は、あるセグメントに対するカウント提示の統計量を決定する工程を含み、および/またはあるセグメントに対するカウント提示を用いて統計量を決定する工程を含む。任意の好適な統計量が生成され得、その非限定的な例としては、平均値、中央値、最頻値、平均、p 値、偏差の尺度(例えば、標準偏差(S D)、シグマ、絶対偏差、平均絶対偏差(M A D)、算出された分散など)、好適な誤差の尺度(例えば、標準誤差、平均平方誤差、平方根平均二乗誤差など)、好適な分散の尺度、好適な標準得点(例えば、標準偏差、累積パーセンテージ、パーセンタイル等価物(p e r c e n t i l e e q u i v a l e n t)、Z 得点、T 得点、R 得点、スタンダードナイン(s t a n d a r d n i n e) (スタナイン(s t a n i n e))、スタナインにおけるパーセントなど)またはそれらの組み合わせが挙げられる。任意の好適な統計学的方法は、カウント提示の統計量を生成するためまたはカウント提示を用いて統計量を生成するために使用され得、その非限定的な例としては、正確検定、F 検定、Z 検定、T 検定、不確定性の尺度を算出および/または比較すること、帰無仮説、対立帰無仮説(c o u n t e r n u l l s) など、カイ二乗検定、総括的検定、有意水準(例えば、統計的有意性)を算出および/または比較すること、メタ解析、多変量解析、回帰、線形単回帰、ロバスト線形回帰 最小二乗回帰、主成分分析、線形判別分析、二次判別分析、バギング、ニューラルネットワーク、サポートベクターマシンモデル、ランダムフォレスト、分類木モデル、K 最近隣法、ロジスティック回帰、l o s s 平滑化(l o s s s m o o t h i n g)、ベールンス・フィッシャーアプローチ、ブートストラッピング、独立した有意性検定を組み合わせるためのフィッシャーの方法、ネイマン・ピアソン検定、検証的データ解析、探索的データ解析などまたはそれらの組み合わせが挙げられる。

30

40

【0040】

z 得点が統計量として生成されるときがあり、それは、(a) (i) テストサンプルに

50

対するセグメントに対するカウント提示から (i i) サンプルセットに対するセグメントに対するカウント提示の中央値を除いた減算の結果を、(b) サンプルセットに対するセグメントに対するカウント提示の M A D で除算した商であるときがある。ある特定の実施形態において、診断テストは、出生前遺伝子診断テストであるときがあり、テストサンプルは、胎児を有する妊婦に由来し、サンプルセットは、正倍数性胎児の妊娠を有する被験体に対するサンプルのセットである。いくつかの実施形態において、診断テストは、出生前診断テストであり、テストサンプルは、胎児を有する妊婦に由来し、サンプルセットは、トリソミー胎児の妊娠を有する被験体に対するサンプルのセットである。ある特定の実施形態において、診断テストは、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについての遺伝子テストであり、サンプルセットは、細胞増殖の症状を有する被験体に対するサンプルのセットである。ある特定の実施形態において、診断テストは、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてであり、サンプルセットは、細胞増殖の症状を有しない被験体に対するサンプルのセットである。

10

【0041】

いくつかの実施形態において、診断テストは、出生前遺伝子診断テストであり、テストサンプルは、胎児を有する妊婦に由来し、その診断テストは、遺伝的変異（例えば、胎児の遺伝的変異）の非存在の存在を判定することを含む。遺伝的変異は、染色体異数性であるときがあり、染色体異数性は、染色体全体の 1 コピー（モノソミー）、3 コピー（トリソミー）または 4 コピーであるときがある。ある特定の出生前診断テストの実施形態における遺伝的変異は、微小重複または微小欠失であるときがある。

20

【0042】

ある特定の実施形態において、診断テストは、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについての遺伝子診断テストであり、その診断テストは、遺伝的変異の非存在の存在を判定することを含む。いくつかの癌診断テストの実施形態における遺伝的変異は、微小重複または微小欠失であるときがある。

【0043】

セグメントカウント提示またはそれから導かれる統計量を用いた遺伝的変異の存在もしくは非存在の判定（アウトカム (o u t c o m e) の判定）は、任意の好適な様式で行われ得る。任意の好適な統計量が、アウトカムを判定するために使用され得、その非限定的な例としては、標準偏差、平均絶対偏差、中央絶対偏差、最大絶対偏差、標準得点（例えば、z 値、z 得点、正規得点、標準化変数）などおよびそれらの組み合わせが挙げられる。いくつかの実施形態において、2 つの統計量（例えば、テストサンプルに対する 1 つの統計量（例えば、試験カウント）および参照サンプルに対する別の統計量（例えば、参照カウント））の間の偏差の数が、約 1 より大きい、約 1.5 より大きい、約 2 より大きい、約 2.5 より大きい、約 2.6 より大きい、約 2.7 より大きい、約 2.8 より大きい、約 2.9 より大きい、約 3 より大きい、約 3.1 より大きい、約 3.2 より大きい、約 3.3 より大きい、約 3.4 より大きい、約 3.5 より大きい、約 4 より大きい、約 5 より大きい、または約 6 より大きいとき、アウトカムが判定される。アウトカムの判定は、カウント提示から導かれる統計量（例えば、z 得点）をその統計量に対する所定の閾値（例えば、z 得点閾値；約 3.95 という z 得点閾値）と比較することによって行われるときがある。

30

40

【0044】

アウトカムの判定は、決定分析を用いて行われるときがある。決定分析の非限定的な例は、2014 年 5 月 23 日に出願され、WO 2014/190286 として公開された特許出願番号 P C T / U S 2 0 1 4 / 0 3 9 3 8 9 に記載されている。ある特定の実施形態において、決定分析は、(a) 本明細書中に記載されるようなテストサンプルに対する試験セグメント（例えば、試験染色体）に対するカウント提示を提供する工程；(b) そのテストサンプルに対する胎児分率 (f e t a l f r a c t i o n) を測定する工程；(c) 対数オッズ比 (L O R) を算出する工程（その L O R は、(i) (1) 遺伝的変異を有する条件付き確率と (2) 遺伝的変異を有する事前確率との第 1 の乗算の結果および (

50

i i) (1) その遺伝的変異を有しない条件付き確率と (2) その遺伝的変異を有しない事前確率との第 2 の乗算の結果の商の対数であり、ここで、その遺伝的変異を有する条件付き確率は、(b) の胎児分率および (a) のカウント提示に従って決定される；および (d) L O R およびカウント提示に従ってアウトカム（例えば、遺伝的変異の存在もしくは非存在）を識別する工程を含む。カウント提示は、正規化されたカウント提示であるときがあり、いくつかの実施形態において、遺伝的変異は、染色体異数性、微小重複または微小欠失である。遺伝的変異を有する条件付き確率は、(i) (b) においてテストサンプルに対して測定された胎児分率、(a) におけるテストサンプルに対するカウント提示に対する z 得点、およびそのカウント提示に対する z 得点の胎児分率特異的分布に従って決定される；(i i) 以下の方程式 23 における関係によって決定される：

10

$$Z \sim \text{Normal}\left(\frac{\mu_X f}{\sigma_X^2}, 1\right) \quad (23)$$

【0045】

(式中、f は、胎児分率であり、X は、当該染色体に対する部分の和、 $X \sim f(\mu_X, \sigma_X)$ であり、ここで、 μ_X および σ_X は、それぞれ X の平均値および標準偏差であり、f (・) は、分布関数である)；および／または (i i i) (a) のテストサンプルカウント提示に対する z 得点とそのカウント提示に対する z 得点の胎児分率特異的分布との間の共通部分であるときがある。遺伝的変異を有しない条件付き確率は、(i) (a) のカウント提示および正倍数体に対するカウント提示に従って決定され；および／または (i i) カウント提示の z 得点と、遺伝的変異を有しない被験体におけるカウント提示に対する z 得点の分布との共通部分であるときがある。遺伝的変異を有する事前確率および遺伝的変異を有しない事前確率は、その試験被験体を含まない複数のサンプルから決定されるときがある。決定分析は、(1) L O R がゼロより大きいかわたは小さいかを判定すること；(2) (a) のカウント提示の z 得点の数量化を確定し、それが 3.95 という値より小さいか、大きいかわたは等しいかを判定すること；(3) テストサンプルに対して、(i) そのカウント提示の z 得点の数量化が 3.95 という値より大きいかわたは等しい場合、および (i i) L O R がゼロより大きい場合、遺伝的変異の存在を判定すること；および／または (4) テストサンプルに対して、(i) そのカウント提示の z 得点の定量が 3.95 という値より小さい場合、および／または (i i) L O R がゼロより小さい場合、遺伝的変異の非存在を判定することを含むときがある。

20

30

【0046】

胎児分率は、任意の好適な様式（例えば、テストサンプル中の胎児核酸の量と核酸総量または母体核酸の量との比率）で表現され得、当該分野で公知の任意の好適な方法を用いて決定され得る。ある特定の実施形態において、胎児核酸の量は、男性胎児に特異的なマーカー（例えば、Y 染色体 S T R マーカー（例えば、D Y S 19、D Y S 385、D Y S 392 マーカー）；R h D 陰性女性における R h D マーカー）、多型配列の対立遺伝子の比率に従って、あるいは胎児核酸に特異的であって母体核酸に特異的でない 1 つもしくはそれを超えるマーカー（例えば、母体と胎児との間の差次的なエピジェネティックバイオマーカー（例えば、メチル化）または母体血漿中の胎児 R N A マーカーに従って、測定される。

40

【0047】

いくつかの実施形態において、胎児分率は、フラグメントの長さの情報を組み込んでいる方法（例えば、国際出願公開番号 W O 2013/177086 に記載されているような、フラグメントの長さの比率 (F L R) の解析、胎児の比率の統計量 (F R S) の解析)を用いて測定される。無細胞胎児核酸フラグメントは、通常、母体由来の核酸フラグメン

50

トよりも短く、胎児分率は、いくつかの実施形態において、特定の長さ閾値を下回るフラグメントをカウントし、その数を、例えば、特定の長さ閾値を上回るフラグメントの数および／またはサンプル中の全核酸の量と比較することによって測定され得る。特定の長さの核酸フラグメントをカウントするための方法は、国際出願公開番号WO 2013/177086にさらに詳細に記載されている。

【0048】

ある特定の実施形態において、胎児分率は、メチル化状態に従って胎児核酸を判別するアッセイを用いて測定される（例えば、胎児数量アッセイ（fetal quantifier assay）（FQA）；米国特許出願公開番号2010/0105049を参照のこと）。ある特定のアッセイの実施形態において、母体のテストサンプル中の胎児DNAの濃度は、以下の方法によって測定される：（a）母体のテストサンプル中に存在するDNAの総量を測定し；（b）1つまたはそれを超えるメチル化感受性制限酵素を用いて母体サンプル中の母体DNAを選択的に消化し、それにより、胎児DNAを濃縮し；（c）（b）からの胎児DNAの量を測定し；（d）工程c）からの胎児DNAの量と、（a）からのDNAの総量を比較し、それにより、母体サンプル中の胎児DNAの濃度を決定する。ある特定の実施形態において、母体テストサンプル中の胎児核酸の絶対コピー数は、例えば、質量分析を用いて、および／または絶対コピー数の計測のために競合的PCRアプローチを用いるシステムを用いて、測定され得る。

【0049】

遺伝子テストは、システム内において全体的にまたは部分的に行われるときがある。カウント提示を確定するためのいくつかまたはすべての工程が、（i）システムにおけるマイクロプロセッサによって、（ii）システムにおけるメモリーと連動して、および／または（iii）コンピュータによって、行われるときがある。

サンプル

【0050】

核酸を解析するためのシステム、方法および製品が、本明細書中に提供される。いくつかの実施形態において、核酸フラグメントの混合物中の核酸フラグメントが解析される。核酸の混合物は、異なるヌクレオチド配列、異なるフラグメント長、異なる起源（例えば、ゲノム起源、胎児起源 対 母体起源、細胞または組織の起源、癌 対 非癌の起源、腫瘍 対 非腫瘍起源、サンプル起源、被験体起源など）またはそれらの組み合わせを有する2つまたはそれを超える核酸フラグメント種を含み得る。

【0051】

本明細書中に記載されるシステム、方法および製品において使用される核酸または核酸混合物は、被験体から得られたサンプルから単離されることが多い。被験体は、任意の生命体または非生命体であり得、それらとしては、ヒト、非ヒト動物、植物、細菌、真菌または原生生物が挙げられるが、これらに限定されない。任意のヒトまたは非ヒト動物を選択することができ、それらとしては、哺乳動物、爬虫類、鳥類、両生類、魚類、有蹄動物、反芻動物、牛（例えば、ウシ）、馬（例えば、ウマ）、山羊および羊（例えば、ヒツジ、ヤギ）、豚（例えば、ブタ）、ラクダ科動物（例えば、ラクダ、ラマ、アルパカ）、サル、類人猿（例えば、ゴリラ、チンパンジー）、クマ科動物（例えば、クマ）、家禽、イヌ、ネコ、マウス、ラット、魚類、イルカ、クジラおよびサメが挙げられるが、これらに限定されない。被験体は、雄または雌（例えば、女性、妊婦）であり得る。被験体は、任意の齢であり得る（例えば、胚、胎児、乳児、小児、成体）。

【0052】

核酸は、任意のタイプの好適な生物学的検体またはサンプル（例えば、テストサンプル）から単離され得る。サンプルまたはテストサンプルは、被験体またはその一部（例えば、ヒト被験体、妊婦、胎児）から単離されたまたは得られた任意の検体であり得る。検体の非限定的な例としては、血液または血液製剤（例えば、血清、血漿など）、臍帯血、絨毛膜絨毛、羊水、脳脊髄液、髄液、洗浄液（例えば、気管支肺胞洗浄液、胃洗浄液、腹膜洗浄液、管洗浄液、耳洗浄液、関節鏡下洗浄液）、生検サンプル（例えば、着床前胚；癌

生検材料由来)、腹腔穿刺 (celocentesis) サンプル、細胞 (血液細胞、胎盤細胞、胚、または胎児の細胞、胎児の有核細胞もしくは胎児の細胞レムナント (cellular remnant)) またはそれらの一部分 (例えば、ミトコンドリア、核、抽出物など)、雌性生殖輸管の洗液、尿、便、痰、唾液、鼻粘膜、前立腺液、洗浄液、精液、リンパ液、胆汁、涙、汗、母乳、乳汁などまたはそれらの組み合わせを含むがこれらに限定されない、被験体由来の流体または組織が挙げられる。いくつかの実施形態において、生物学的サンプルは、被験体由来の子宮頸部スワブである。いくつかの実施形態において、生物学的サンプルは、血液であり得、血漿または血清であり得るときがある。用語「血液」は、本明細書中で使用されるとき、妊婦または妊娠の可能性について検査されている女性由来の血液サンプルまたは血液調製物のことを指す。この用語は、従来定義されてきたような、全血、血液製剤、または血液の任意の画分、例えば、血清、血漿、パフイーコートなどを包含する。血液またはその画分は、ヌクレオソーム (例えば、母体および／または胎児のヌクレオソーム) を含むことが多い。ヌクレオソームは、核酸を含み、無細胞であるか、または細胞内のものであるときがある。血液は、パフイーコートも含む。パフイーコートは、フィコール勾配を使用することによって単離されるときがある。パフイーコートは、白血球細胞 (例えば、白血球、T細胞、B細胞、血小板など) を含む得る。ある特定の実施形態において、パフイーコートは、母体および／または胎児の核酸を含む。血漿とは、抗凝固薬で処理された血液の遠心分離から生じる全血の画分のことを指す。血清とは、血液サンプルが凝固した後に残っている流体の水様の部分のことを指す。液体または組織サンプルは、病院またはクリニックが一般に従っている標準的なプロトコルに従って回収されることが多い。血液の場合、適切な量の末梢血 (例えば、3～40ミリリットル) を回収することが多く、それは、調製の前または後に、標準的な手順に従って保管され得る。核酸を抽出する液体または組織サンプルは、細胞を含まない (例えば、無細胞である) 場合がある。いくつかの実施形態において、液体または組織サンプルは、細胞エレメントまたは細胞レムナントを含み得る。いくつかの実施形態において、胎児の細胞または癌細胞が、サンプル中に含まれ得る。

【0053】

サンプルは、液体サンプルであり得る。液体サンプルは、細胞外核酸 (例えば、循環無細胞DNA) を含み得る。液体サンプルの非限定的な例としては、血液または血液製剤 (例えば、血清、血漿など)、臍帯血、羊水、脳脊髄液、髄液、洗浄液 (例えば、気管支肺細胞洗浄液、胃洗浄液、腹膜洗浄液、管洗浄液、耳洗浄液、関節鏡下洗浄液)、生検サンプル (例えば、癌を検出するための液体生検材料)、腹腔穿刺サンプル、雌性生殖輸管の洗液、尿、痰、唾液、鼻粘膜、前立腺液、洗浄液、精液、リンパ液、胆汁、涙、汗、母乳、乳汁などまたはそれらの組み合わせが挙げられる。ある特定の実施形態において、サンプルは、液体生検材料であり、これは、疾患 (例えば、癌) の存在、非存在、進行または緩解についての被験体由来の液体サンプルの評価のことを広く指す。液体生検材料は、固体 (solid) 生検材料 (例えば、腫瘍生検材料) とともに、または固体生検材料の代替物として、使用され得る。ある特定の場合において、細胞外核酸が液体生検材料において解析される。

【0054】

サンプルは、不均一であることが多く、この不均一は、1タイプより多い核酸種がそのサンプル中に存在すること意味する。例えば、不均一な核酸としては、(i) 癌の核酸および癌ではない核酸、(ii) 病原体および宿主の核酸、(iii) 胎児由来および母体由来の核酸、ならびに／またはより一般的には、(iv) 変異型および野生型の核酸が挙げられ得るが、これらに限定されない。1つより多い細胞型、例えば、胎児の細胞および母体の細胞、癌細胞および非癌細胞、または病原体細胞および宿主細胞が存在するので、サンプルは不均一であり得る。いくつかの実施形態において、少数核酸種および多数核酸種が存在する。

【0055】

本明細書中に記載される技術を出生前に適用する場合、液体または組織サンプルが、検

10

20

30

40

50

査に適した在胎期間の女性または妊娠の可能性について検査されている女性から回収され得る。好適な在胎期間は、行われる出生前テストに応じて変動し得る。ある特定の実施形態において、妊婦被験体は、妊娠第1期、妊娠第2期であるときがあるか、または妊娠第3期であるときがある。ある特定の実施形態において、液体または組織は、妊娠約1～約45週（例えば、妊娠1～4、4～8、8～12、12～16、16～20、20～24、24～28、28～32、32～36、36～40または40～44週）の妊婦から回収され、妊娠約5～約28週（例えば、妊娠6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26または27週）の妊婦から回収されるときがある。ある特定の実施形態において、液体または組織サンプルは、出産中または出産直後（例えば、0～72時間後）の妊婦から回収される（例えば、経腔分娩または非経腔分娩（例えば、手術分娩））。

10

血液サンプルの取得およびDNAの抽出

【0056】

いくつかの実施形態において、本明細書中の方法は、被験体のゲノムにおける染色体変化の存在もしくは非存在を検出するおよび／または被験体の健康状態をモニターする非侵襲的手段として、被験体の血液に見られるDNAを分離する工程、濃縮する工程、配列決定する工程および／または解析する工程を含む。

血液サンプルの取得

【0057】

血液サンプルは、本技術の方法を用いて任意の年齢の被験体（例えば、雄または雌の被験体）から得ることができる。血液サンプルは、本技術の方法を用いて検査するのに適した在胎期間の妊婦から得ることができる。好適な在胎期間は、下記で論じられるような検査される障害に応じて変動し得る。被験体（例えば、妊婦）からの血液の回収は、病院またはクリニックが一般に従っている標準的なプロトコルに従って行われることが多い。適切な量の末梢血、例えば、代表的には5～50mlが、回収されることが多く、さらなる調製の前に標準的な手順に従って保管され得る。血液サンプルは、サンプル中に存在する核酸の分解または品質を最小限にする様式で、回収され得るか、保管され得るか、または輸送され得る。

20

血液サンプルの調製

【0058】

被験体の血液に見られるDNAの解析は、例えば、全血、血清または血漿を用いて行われ得る。母体の血液に見られる胎児DNAの解析は、例えば、全血、血清または血漿を用いて行われ得る。患者の血液に見られる腫瘍DNAの解析は、例えば、全血、血清または血漿を用いて行われ得る。被験体（例えば、母体の被験体；癌患者）から得られた血液から血清または血漿を調製するための方法は知られている。例えば、被験体の血液（例えば、妊婦の血液；癌患者の血液）は、血液凝固を防ぐために、EDTAを含むチューブまたはVacutainer SST (Becton Dickinson, Franklin Lakes, N. J.)などの専用の市販品に入れられ得、次いで、遠心分離によって全血から血漿を得ることができる。血清は、遠心分離後の血液凝固ありまたはなしで得られ得る。遠心分離を用いる場合、その遠心分離は、通常、適切な速度、例えば、1,500～3,000×gで行われるが、これに限らない。血漿または血清は、さらなる遠心分離工程に供され、その後、DNA抽出に向けて新しいチューブに移され得る。

30

40

【0059】

全血の細胞を含まない部分に加えて、女性または患者由来の全血サンプルの遠心分離および血漿の除去後に得ることができる Buffy Coat 部分に濃縮された細胞画分からもDNAが回収され得る。

DNAの抽出

【0060】

血液を含む生物学的サンプルからDNAを抽出するための公知の方法が数多くある。DNA調製の一般的な方法（例えば、Sambrook and Russell, Mol

50

ecular Cloning: A Laboratory Manual 3d ed., 2001に記載されているもの)に従うことができ;様々な商業的に入手可能な試薬またはキット、例えば、QiagenのQIAamp Circulating Nucleic Acid Kit、QIAamp DNA Mini KitまたはQIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany)、GenomicPrepTM Blood DNA Isolation Kit (Promega, Madison, Wis.)およびGFXTM Genomic Blood DNA Purification Kit (Amersham, Piscataway, N. J.)も、被験体由来の血液サンプルからDNAを得るために使用してよい。1つより多いこれらの方法の組み合わせを使用してもよい。

10

【0061】

いくつかの実施形態において、被験体から得られたサンプルは、まず、1つまたはそれを超える方法によって、腫瘍核酸について濃縮され得るかまたは相対的に濃縮され得る。例えば、腫瘍DNAと正常な患者DNAとの判別が、本技術の組成物およびプロセスのみを用いて、または他の判別要素と組み合わせて、行われ得る。

【0062】

いくつかの実施形態において、妊婦被験体から得られたサンプルは、まず、1つまたはそれを超える方法によって、胎児核酸について濃縮され得るかまたは相対的に濃縮され得る。例えば、胎児DNAと母体DNAとの判別が、本技術の組成物およびプロセスのみを用いて、または他の判別要素と組み合わせて、行われ得る。これらの要素の例としては、X染色体とY染色体との間の単一ヌクレオチド差異、Y染色体特異的配列、ゲノムのいずれかの箇所に位置する多型、胎児DNAと母体DNAとの間のサイズの差異、および母体組織と胎児組織との間のメチル化パターンの差異が挙げられるが、これらに限定されない。

20

【0063】

サンプルを特定の核酸種について濃縮するための他の方法は、2007年5月30日に出願されたPCT特許出願番号PCT/US07/69991、2007年6月15日に
出願されたPCT特許出願番号PCT/US2007/071232、米国特許仮出願番号60/968,876および60/968,878(本出願人に譲渡された)、(2005年11月28日に
出願されたPCT特許出願番号PCT/EP05/012707)に記載されており、これらのすべてが参照により本明細書に組み込まれる。ある特定の
実施形態において、母体核酸は、サンプルから選択的に除去される(部分的に、実質的に、
ほぼ完全にまたは完全に)。

30

【0064】

用語「核酸」および「核酸分子」は、本開示全体を通じて交換可能に使用され得る。これらの用語は、例えば、DNA(例えば、相補DNA(cDNA)、ゲノムDNA(gDNA)など)、RNA(例えば、メッセージRNA(mRNA)、低分子阻害RNA(sRNA)、リボソームRNA(rRNA)、tRNA、マイクロRNA、胎児または胎盤によって高度に発現されるRNAなど)、ならびに/またはDNAアナログもしくはRNAアナログ(例えば、塩基アナログ、糖アナログおよび/または非天然の骨格などを含むもの)、RNA/DNAハイブリッドおよびポリアミド核酸(PNA)からの任意の組成の核酸のことを指し、これらのすべてが、一本鎖または二本鎖の形態であり得、別段限定されない限り、天然に存在するヌクレオチドと同様の様式で機能し得る、天然のヌクレオチドの公知のアナログを含み得る。核酸は、ある特定の実施形態において、インビトロまたは宿主細胞、細胞、細胞核もしくは細胞の細胞質において複製することができるかまたは複製されることができる、プラスミド、ファージ、ウイルス、自律複製配列(ARS)、セントロメア、人工染色体、染色体または他の核酸であり得るか、またはそれらに由来し得る。いくつかの実施形態において、鋳型核酸は、単一の染色体に由来し得る(例えば、核酸サンプルは、二倍体生物から得られたサンプルの1本の染色体に由来し得る)。具体的に限定されない限り、この用語は、参照核酸と同様の結合特性を有し、および天然

40

50

に存在するヌクレオチドと同様の様式で代謝される、天然のヌクレオチドの公知のアナログを含む核酸を包含する。別段示されない限り、特定の核酸配列は、その保存的に改変されたバリエーション（例えば、縮重コドン置換）、対立遺伝子、オルソログ、単一ヌクレオチド多型（SNP）および相補的配列ならびに明示的に示された配列も暗に包含する。具体的には、縮重コドン置換は、1つまたはそれを超える選択された（またはすべての）コドンの3番目の位置が、混合塩基および／またはデオキシイノシン残基で置換された配列を作製することによって達成され得る。核酸という用語は、遺伝子座、遺伝子、cDNA、および遺伝子によってコードされるmRNAと交換可能に使用される。この用語は、ヌクレオチドアナログから合成されたRNAまたはDNAの等価物、誘導體、バリエーションおよびアナログとして、一本鎖ポリヌクレオチド（「センス」または「アンチセンス」、「プラス」鎖または「マイナス」鎖、「順方向」読み枠または「逆方向」読み枠）および二本鎖ポリヌクレオチドも含み得る。用語「遺伝子」は、ポリペプチド鎖の生成に関わるDNAのセグメントを意味し；この用語は、遺伝子産物の転写／翻訳および転写／翻訳の制御に関わるコード領域の前および後の領域（リーダーおよびトレーラー（trailer））、ならびに個々のコーディングセグメント（エキソン）の間の介在配列（イントロン）を含む。

10

【0065】

デオキシリボヌクレオチドには、デオキシアデノシン、デオキシシチジン、デオキシグアノシンおよびデオキシチミジンが含まれる。RNAの場合、塩基シトシンが、ウラシルで置き換えられる。鋳型核酸は、被験体から得られた核酸を鋳型として用いて調製され得る。

20

核酸の単離および処理

【0066】

核酸は、当該分野で公知の方法によって、1つまたはそれを超える供給源（例えば、細胞、血清、血漿、パフィーコート、リンパ液、皮膚、土壌など）から得られ得る。生物学的サンプル（例えば、血液または血液製剤）からDNAを単離するため、抽出するためおよび／または精製するために、任意の好適な方法を用いることができ、それらの非限定的な例としては、DNA調製の方法（例えば、Sambrook and Russell, Molecular Cloning: A Laboratory Manual 3rd ed., 2001に記載されているもの）、様々な商業的に入手可能な試薬またはキット、例えば、QiagenのQIAamp Circulating Nucleic Acid Kit、QIAamp DNA Mini KitもしくはQIAamp DNA Blood Mini Kit（Qiagen, Hilden, Germany）、GenomicPrepTM Blood DNA Isolation Kit（Promega, Madison, Wis.）およびGFXTM Genomic Blood DNA Purification Kit（Amersham, Piscataway, N. J.）などまたはそれらの組み合わせが挙げられる。

30

【0067】

細胞溶解の手順および試薬は、当該分野で公知であり、一般に、化学的溶解方法（例えば、洗浄剤、低張液、酵素的手順などまたはそれらの組み合わせ）、物理的溶解方法（例えば、フレンチプレス、超音波処理など）または電解による溶解方法によって行われ得る。任意の好適な溶解手順を用いることができる。例えば、化学的方法は、一般に、細胞を破壊するために溶解剤を用い、その細胞から核酸を抽出した後、カオトロピック塩で処理する。凍結／融解の後の粉碎、細胞プレスの使用などのような物理的方法も有用である。高塩溶解手順もよく使用される。例えば、アルカリ溶解手順が使用され得る。後者の手順は、従来よりフェノールクロロホルム溶液の使用を組み込んでおり、3つの溶液を必要とする代替のフェノールクロロホルムフリー手順を用いることもできる。後者の手順では、1つの溶液は、15mM Tris, pH 8.0；10mM EDTAおよび100 μg/ml Rnase Aを含み得；第2の溶液は、0.2N NaOHおよび1% SDSを含み得；第3の溶液は、3M KOAc, pH 5.5を含み得る。これらの手順は

40

50

、全体が本明細書に組み込まれる Current Protocols in Molecular Biology, John Wiley & Sons, N. Y., 6. 3. 1-6. 3. 6 (1989) に見られ得る。

【0068】

核酸は、別の核酸と比べて異なる時点において単離され得、それらのサンプルの各々が、同じまたは異なる供給源由来である。核酸は、例えば、核酸ライブラリー（例えば、cDNAまたはRNAライブラリー）由来であり得る。核酸は、サンプル由来の核酸の精製もしくは単離および／または核酸分子の増幅の結果であり得る。本明細書中に記載されるプロセスのために提供される核酸は、1つのサンプルまたは2つもしくはそれを超えるサンプル（例えば、1つもしくはそれを超える、2つもしくはそれを超える、3つもしくはそれを超える、4つもしくはそれを超える、5つもしくはそれを超える、6つもしくはそれを超える、7つもしくはそれを超える、8つもしくはそれを超える、9つもしくはそれを超える、10個もしくはそれを超える、11個もしくはそれを超える、12個もしくはそれを超える、13個もしくはそれを超える、14個もしくはそれを超える、15個もしくはそれを超える、16個もしくはそれを超える、17個もしくはそれを超える、18個もしくはそれを超える、19個もしくはそれを超えるまたは20個もしくはそれを超えるサンプル）由来の核酸を含み得る。

10

【0069】

ある特定の実施形態において、核酸には、細胞外核酸が含まれ得る。用語「細胞外核酸」は、本明細書中で使用されるとき、実質的に細胞を有しない供給源から単離された核酸のことを指し得、「無細胞」核酸、「循環無細胞核酸」（例えば、CCFフラグメント、ccfDNA）および／または「無細胞循環核酸」とも称される。細胞外核酸は、血液（例えば、ヒトの血液、例えば、妊婦の血液）に存在し得、その血液から得ることができる。細胞外核酸は、検出可能な細胞を含まないことが多く、細胞エレメントまたは細胞レムナントを含むことがある。細胞外核酸に対する細胞を含まない供給源の非限定的な例は、血液、血漿、血清および尿である。本明細書中で使用されるとき、用語「無細胞循環サンプル核酸を得る」には、サンプルを直接得ること（例えば、サンプル、例えば、テストサンプルを回収すること）またはサンプルを回収した別の者からサンプルを得ることが含まれる。理論に限定されるものではないが、細胞外核酸は、ある範囲にわたる一連の長さを有することが多い細胞外核酸（例えば、「ラダー」）の基礎をもたらす、細胞のアポトーシスおよび細胞の破壊の産物であり得る。

20

30

【0070】

細胞外核酸は、種々の核酸種を含み得るので、ある特定の実施形態において、「不均一」と本明細書中で称される。例えば、癌を有する人由来の血清または血漿は、癌細胞（例えば、腫瘍、新形成）由来の核酸および非癌細胞由来の核酸を含み得る。別の例では、妊婦由来の血清または血漿は、母体核酸および胎児核酸を含み得る。場合によっては、癌の核酸または胎児の核酸は、核酸全体の約5%～約50%であるときがある（例えば、全核酸の約4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48または49%が癌の核酸または胎児の核酸である）。いくつかの実施形態において、核酸中の癌の核酸または胎児の核酸の大部分は、約500塩基対またはそれより短い長さである（例えば、癌の核酸または胎児の核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約500塩基対またはそれより短い長さである）。いくつかの実施形態において、核酸中の癌の核酸または胎児の核酸の大部分は、約250塩基対またはそれより短い長さである（例えば、癌の核酸または胎児の核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約250塩基対またはそれより短い長さである）。いくつかの実施形態において、核酸中の癌の核酸または胎児の核酸の大部分は、約200塩基対またはそれより短い長さである（例えば、癌の核酸または胎児の核酸の約80、

40

50

85、90、91、92、93、94、95、96、97、98、99または100%は、約200塩基対またはそれより短い長さである）。いくつかの実施形態において、核酸中の癌の核酸または胎児の核酸の大部分は、約150塩基対またはそれより短い長さである（例えば、癌の核酸または胎児の核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約150塩基対またはそれより短い長さである）。いくつかの実施形態において、核酸中の癌の核酸または胎児の核酸の大部分は、約100塩基対またはそれより短い長さである（例えば、癌の核酸または胎児の核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約100塩基対またはそれより短い長さである）。いくつかの実施形態において、核酸中の癌の核酸または胎児の核酸の大部分は、約50塩基対またはそれより短い長さである（例えば、癌の核酸または胎児の核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約50塩基対またはそれより短い長さである）。いくつかの実施形態において、核酸中の癌の核酸または胎児の核酸の大部分は、約25塩基対またはそれより短い長さである（例えば、癌の核酸または胎児の核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約25塩基対またはそれより短い長さである）。

10

【0071】

核酸は、ある特定の実施形態において、その核酸を含むサンプルの処理なしに、本明細書中に記載される方法を行うために提供され得る。いくつかの実施形態において、核酸は、その核酸を含むサンプルの処理の後に、本明細書中に記載される方法を行うために提供される。例えば、核酸は、サンプルから抽出され得るか、単離され得るか、精製され得るか、部分的に精製され得るか、または増幅され得る。用語「単離された」は、本明細書中で使用されるとき、その元の環境（例えば、それが天然に存在する場合は天然の環境、または外因的に発現される場合は宿主細胞）から取り出された核酸のことを指し、ゆえに、人間の介入によって（例えば、「人間の手によって」）その元の環境から変更されている。用語「単離された核酸」は、本明細書中で使用されるとき、被験体（例えば、ヒト被験体）から取り出された核酸のことを指し得る。単離された核酸は、供給源サンプル中に存在する構成要素の量よりも少ない非核酸構成要素（例えば、タンパク質、脂質）とともに提供され得る。単離された核酸を含む組成物は、非核酸構成要素を約50%から99%超、含まないことがある。単離された核酸を含む組成物は、非核酸構成要素を約90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超、含まないことがある。用語「精製された」は、本明細書中で使用されるとき、核酸を精製手順に供する前に存在する非核酸構成要素の量よりも少ない非核酸構成要素（例えば、タンパク質、脂質、炭水化物）を含む、提供される核酸のことを指し得る。精製された核酸を含む組成物は、他の非核酸構成要素を約80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超、含まないことがある。用語「精製された」は、本明細書中で使用されるとき、核酸が由来するサンプル供給源よりも少ない核酸種を含む、提供される核酸のことを指し得る。精製された核酸を含む組成物は、他の核酸種を約90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超、含まないことがある。例えば、胎児核酸は、母体核酸および胎児核酸を含む混合物から精製され得る。ある特定の例において、胎児核酸の小さいフラグメント（例えば、30～500bpフラグメント）が、胎児核酸フラグメントと母体核酸フラグメントの両方を含む混合物から精製され得るか、または部分的に精製され得る。ある特定の例において、胎児核酸のより小さいフラグメントを含むヌクレオソームが、母体核酸のより大きいフラグメントを含むより大きいヌクレオソーム複合体の混合物から精製され得る。ある特定の例において、癌細胞の核酸は、癌細胞の核酸および非癌細胞の核酸を含む混合物から精製され得る。ある特定の例において、癌細胞の核酸の小さいフラグメントを含むヌクレオソームは、非癌核酸のより大きいフラグメントを含むより大きいヌクレオソーム複合体の混合物から精製され得る。

20

30

40

50

【0072】

いくつかの実施形態において、核酸は、本明細書中に記載される方法の前、方法の最中または方法の後に、剪断されるかまたは切断される。剪断されたまたは切断された核酸は、約5～約10、000塩基対、約100～約1,000塩基対、約100～約500塩基対または約10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000もしくは9000塩基対という名目上の長さ、平均の長さまたは長さの平均値を有し得る。剪断されたまたは切断された核酸は、当該分野で公知の好適な方法によって作製され得、得られる核酸フラグメントの平均の長さ、長さの平均値または名目上の長さは、フラグメントを作製する適切な方法を選択することによって制御され得る。

10

【0073】

いくつかの実施形態において、核酸は、好適な方法によって剪断されるかまたは切断され、その好適な方法の非限定的な例としては、物理的方法（例えば、剪断、例えば、超音波処理、フレンチプレス、加熱、UV照射など）、酵素的プロセス（例えば、酵素的切断剤（例えば、好適なヌクレアーゼ、好適な制限酵素、好適なメチル化感受性制限酵素））、化学的方法（例えば、アルキル化、DMS、ピペリジン、酸加水分解、塩基加水分解、加熱などまたはそれらの組み合わせ）、米国特許出願公開番号20050112590に記載されているプロセスなどまたはそれらの組み合わせが挙げられる。

20

【0074】

本明細書中で使用されるとき、「剪断」または「切断」とは、核酸分子（例えば、核酸鋳型遺伝子分子またはその増幅産物）が2つまたはそれを超えるより小さい核酸分子に切り離され得る手順または条件のことを指す。そのような剪断または切断は、配列特異的、塩基特異的または非特異的であり得、例えば、化学的、酵素的、物理的剪断（例えば、物理的断片化）を含む、種々の方法、試薬または条件のいずれかによって達成され得る。本明細書中で使用されるとき、「切断産物」、「切断された産物」またはその文法上のバリエーションは、核酸の剪断もしくは切断の結果として生じる核酸分子またはその増幅産物のことを指す。

【0075】

用語「増幅された」は、本明細書中で使用されるとき、サンプル中の標的核酸を、その標的核酸と同じもしくは実質的に同じヌクレオチド配列を有するアンプリコン核酸またはそのセグメントを直線的または指数関数的に生成するプロセスに供することを指す。ある特定の実施形態において、用語「増幅された」とは、ポリメラーゼ連鎖反応（PCR）を含む方法のことを指す。例えば、増幅産物は、核酸鋳型配列の増幅されるヌクレオチド領域より多い1つまたはそれを超えるヌクレオチドを含み得る（例えば、プライマーは、核酸鋳型遺伝子分子に相補的なヌクレオチドに加えて、「余分の」ヌクレオチド、例えば、転写開始配列を含み得、その結果、「余分の」ヌクレオチドまたはその核酸鋳型遺伝子分子の増幅されるヌクレオチド領域に対応しないヌクレオチドを含む増幅産物をもたらす）。

30

40

【0076】

本明細書中で使用されるとき、用語「相補的な切断反応」とは、同じ核酸に対して異なる切断試薬を用いて行われるか、あるいは同じ標的核酸もしくは参照核酸または標的タンパク質もしくは参照タンパク質の別の切断パターンが生成されるように、同じ切断試薬の切断特異性を変更することによって、同じ核酸に対して行われる切断反応のことを指す。ある特定の実施形態において、核酸は、1つまたはそれを超える反応容器において、1つまたはそれを超える特異的切断剤（例えば、1、2、3、4、5、6、7、8、9、10個もしくはそれを超える特異的切断剤）で処理され得る（例えば、核酸は、別個の容器において各特異的切断剤で処理される）。用語「特異的切断剤」は、本明細書中で使用されるとき、1つまたはそれを超える特異的部位において核酸を切断できる作用物質のことを

50

指し、1つまたはそれを超える特異的部位において核酸を切断できる化学物質または酵素のことを指すときがある。

【0077】

核酸は、本明細書中に記載される方法のために核酸を提供する前に、その核酸の中のある特定のヌクレオチドを改変するプロセスにも曝露され得る。例えば、核酸の中のヌクレオチドのメチル化状態に基づいてその核酸を選択的に改変するプロセスが、核酸に適用され得る。さらに、高温、紫外線、x線などの条件が、核酸分子の配列の変化を誘導し得る。核酸は、好適な配列解析を行うのに有用な任意の好適な形態で提供され得る。

【0078】

核酸は、一本鎖または二本鎖であり得る。例えば、一本鎖DNAは、例えば、加熱またはアルカリによる処理によって、二本鎖DNAを変性することによって作製され得る。ある特定の実施形態において、核酸は、オリゴヌクレオチドまたはDNA様分子、例えば、ペプチド核酸(PNA)による二重鎖DNA分子のストランド侵入によって形成されるDループ構造である。Dループの形成は、当該分野で公知の方法を用いて、例えば、大腸菌RecAタンパク質の添加および／または塩濃度の変更によって促進され得る。

少数種 対 多数種

【0079】

少なくとも2つの異なる核酸種が、細胞外(例えば、循環無細胞)核酸として、異なる量で存在し得、それらは、少数種および多数種と称されるときがある。ある特定の場
合において、少数種の核酸は、罹患細胞型(例えば、癌細胞、喪失性の(wasting)細胞、免疫系に攻撃された細胞)に由来する。ある特定の実施形態において、染色体変化は、少数種の核酸について判定される。ある特定の実施形態において、染色体変化は、多数種の核酸について判定される。本明細書中で使用されるとき、用語「少数」または「多数」は、任意の点において厳しく定義されると意図されていない。1つの態様において、「少数」と考えられる核酸は、例えば、サンプル中の全核酸の少なくとも約0.1%~サンプル中の全核酸の50%未満の存在量を有し得る。いくつかの実施形態において、少数核酸は、サンプル中の全核酸の少なくとも約1%~サンプル中の全核酸の約40%の存在量を有し得る。いくつかの実施形態において、少数核酸は、サンプル中の全核酸の少なくとも約2%~サンプル中の全核酸の約30%の存在量を有し得る。いくつかの実施形態において、少数核酸は、サンプル中の全核酸の少なくとも約3%~サンプル中の全核酸の約25%の存在量を有し得る。例えば、少数核酸は、サンプル中の全核酸の約1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%または30%の存在量を有し得る。場合によっては、細胞外核酸の少数種は、核酸全体の約1%~約40%であるときがある(例えば、その核酸の約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39または40%が少数種核酸である)。いくつかの実施形態において、少数核酸は、細胞外DNAである。いくつかの実施形態において、少数核酸は、アポトーシス組織由来の細胞外DNAである。いくつかの実施形態において、少数核酸は、細胞増殖性障害によって影響された組織由来の細胞外DNAである。いくつかの実施形態において、少数核酸は、腫瘍細胞由来の細胞外DNAである。いくつかの実施形態において、少数核酸は、細胞外の胎児DNAである。

【0080】

別の態様において、「多数」と考えられる核酸は、例えば、サンプル中の全核酸の50%超~サンプル中の全核酸の約99.9%の存在量を有し得る。いくつかの実施形態において、多数核酸は、サンプル中の全核酸の少なくとも約60%~サンプル中の全核酸の約99%の存在量を有し得る。いくつかの実施形態において、多数核酸は、サンプル中の全核酸の少なくとも約70%~サンプル中の全核酸の約98%の存在量を有し得る。いくつ

かの実施形態において、多数核酸は、サンプル中の全核酸の少なくとも約75%～サンプル中の全核酸の約97%の存在量を有し得る。例えば、多数核酸は、サンプル中の全核酸の少なくとも約70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%または99%の存在量を有し得る。いくつかの実施形態において、多数核酸は、細胞外DNAである。いくつかの実施形態において、多数核酸は、細胞外の母体DNAである。いくつかの実施形態において、多数核酸は、健常組織由来のDNAである。いくつかの実施形態において、多数核酸は、非腫瘍細胞由来のDNAである。

【0081】

いくつかの実施形態において、細胞外核酸の少数種は、約500塩基対またはそれより短い長さである（例えば、少数種核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約500塩基対またはそれより短い長さである）。いくつかの実施形態において、細胞外核酸の少数種は、約300塩基対またはそれより短い長さである（例えば、少数種核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約300塩基対またはそれより短い長さである）。いくつかの実施形態において、細胞外核酸の少数種は、約200塩基対またはそれより短い長さである（例えば、少数種核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約200塩基対またはそれより短い長さである）。いくつかの実施形態において、細胞外核酸の少数種は、約150塩基対またはそれより短い長さである（例えば、少数種核酸の約80、85、90、91、92、93、94、95、96、97、98、99または100%は、約150塩基対またはそれより短い長さである）。

細胞型

【0082】

本明細書中で使用されるとき、「細胞型」とは、別のタイプの細胞と区別され得る細胞のタイプのことを指す。細胞外核酸は、いくつかの異なる細胞型由来の核酸を含み得る。循環無細胞核酸に核酸を与え得る細胞型の非限定的な例としては、肝臓細胞（例えば、肝細胞）、肺細胞、脾臓細胞、膵臓細胞、結腸細胞、皮膚細胞、膀胱細胞、眼細胞、脳細胞、食道細胞、頭部の細胞、頸部の細胞、卵巣の細胞、精巣の細胞、前立腺細胞、胎盤細胞、上皮細胞、内皮細胞、脂肪細胞、腎臓／腎細胞、心臓細胞、筋細胞、血液細胞（例えば、白血球）、中枢神経系（CNS）細胞などおよび前述のものの組み合わせが挙げられる。いくつかの実施形態において、解析される循環無細胞核酸に核酸を与える細胞型には、白血球、内皮細胞および肝細胞肝臓細胞（hepatocyte liver cells）が挙げられる。本明細書中でさらに詳細に記載されるように、マーカーの状態が、医学的症状を有する被験体における細胞型および医学的症状を有しない被験体における細胞型に対して同じまたは実質的に同じである核酸の遺伝子座を同定するおよび選択する一部として、種々の細胞型がスクリーニングされ得る。

【0083】

特定の細胞型は、医学的症状を有する被験体および医学的症状を有しない被験体において、同じまたは実質的に同じままであるときがある。非限定的な例において、特定の細胞型の生細胞または生存細胞の数が、ある細胞変性症状において減少し得、生きている生存細胞が、その医学的症状を有する被験体において、改変されないか、または有意に改変されない。

【0084】

特定の細胞型は、医学的症状の一部として改変されるときがあり、その元の状態とは異なる1つまたはそれを超える特性を有するときがある。非限定的な例において、特定の細胞型は、癌の症状の一部として、正常な速度よりも速い速度で増殖し得、異なる形態を有する細胞に転換し得、1つもしくはそれを超える異なる細胞表面マーカーを発現する細胞に転換し得、および／または腫瘍の一部になり得る。特定の細胞型（すなわち、前駆細胞

）が医学的症状の一部として改変される実施形態において、アッセイされる1つまたはそれを超えるマーカーの各々に対するマーカーの状態は、その医学的症状を有する被験体におけるその特定の細胞型およびその医学的症状を有しない被験体におけるその特定の細胞型に対して同じまたは実質的に同じであることが多い。したがって、用語「細胞型」は、ある医学的症状を有しない被験体における細胞のタイプ、およびその医学的症状を有する被験体におけるその細胞の改変バージョンに関するときがある。いくつかの実施形態において、「細胞型」は、前駆細胞だけであり、前駆細胞から生じる改変バージョンではない。「細胞型」は、前駆細胞、および前駆細胞から生じる改変された細胞に関するときがある。そのような実施形態において、解析されるマーカーに対するマーカーの状態は、ある医学的症状を有する被験体における細胞型およびその医学的症状を有しない被験体における細胞型に対して同じまたは実質的に同じであることが多い。

10

【0085】

ある特定の実施形態において、細胞型は、癌細胞である。ある特定の癌細胞のタイプとしては、例えば、白血病細胞（例えば、急性骨髄性白血病、急性リンパ芽球性白血病、慢性骨髄性白血病、慢性リンパ芽球性白血病）；癌性腎臓／腎細胞（例えば、腎細胞癌（明細胞、1型乳頭状、2型乳頭状、嫌色素、膨大細胞、集合管）、腎腺癌、副腎腫、ウィルムス腫瘍、移行上皮癌腫）；脳腫瘍細胞（例えば、聴神経腫、星状細胞腫（グレードI：毛様細胞性星状細胞腫、グレードII：低悪性度星状細胞腫、グレードIII：未分化星状細胞腫、グレードIV：神経膠芽腫（GBM））、脊索腫、cnsリンパ腫、頭蓋咽頭腫、神経膠腫（脳幹神経膠腫、上衣腫、混合性神経膠腫、視神経膠腫、上衣下腫）、髄芽腫、髄膜腫、転移性脳腫瘍、乏突起膠腫、下垂体腫瘍、原始神経外胚葉性腫瘍（PNET）、神経鞘腫、若年性毛様細胞性星状細胞腫（JPA）、松果体腫瘍、ラブドイド腫瘍）が挙げられる。

20

【0086】

異なる細胞型は、任意の好適な特色によって区別され得、その特色としては、1つまたはそれを超える異なる細胞表面マーカー、1つまたはそれを超える異なる形態学的特徴、1つまたはそれを超える異なる機能、1つまたはそれを超える異なるタンパク質（例えば、ヒストン）修飾および1つまたはそれを超える異なる核酸マーカーが挙げられるが、これらに限定されない。核酸マーカーの非限定的な例としては、単一ヌクレオチド多型（SNP）、核酸遺伝子座のメチル化状態、短タンデム反復、挿入（例えば、微小挿入）、欠失（微小欠失）などおよびそれらの組み合わせが挙げられる。タンパク質（例えば、ヒストン）修飾の非限定的な例としては、アセチル化、メチル化、ユビキチン化、リン酸化、SUMO化などおよびそれらの組み合わせが挙げられる。

30

【0087】

本明細書中で使用されるとき、用語「関連する細胞型」とは、別の細胞型と共通の複数の特色を有する細胞型のことを指す。関連する細胞型において、75%またはそれを超える細胞表面マーカーが、その細胞型と共通しているときがある（例えば、細胞表面マーカーの約80%、85%、90%もしくは95%またはそれを超えるものが、関連する細胞型と共通している）。

核酸の濃縮

40

【0088】

いくつかの実施形態において、核酸（例えば、細胞外核酸）は、核酸の部分集団または種について濃縮されるかまたは相対的に濃縮される。核酸の部分集団としては、例えば、胎児核酸、母体の核酸、癌の核酸、患者の核酸、特定の長さもしくは長さの範囲のフラグメントを含む核酸、または特定のゲノム領域（例えば、単一の染色体、染色体のセットおよび／またはある特定の染色体領域）に由来する核酸が挙げられ得る。そのような濃縮されたサンプルは、本明細書中に提供される方法とともに使用され得る。したがって、ある特定の実施形態において、本技術の方法は、サンプル中の核酸の部分集団、例えば、癌の核酸または胎児の核酸について濃縮するさらなる工程を含む。ある特定の実施形態において、癌または胎児の画分を測定するための方法は、癌の核酸または胎児の核酸を濃縮する

50

ためにも使用され得る。ある特定の実施形態において、母体核酸が、サンプルから選択的に除去される（部分的に、実質的に、ほぼ完全にまたは完全に）。ある特定の実施形態において、特定の低コピー数種の核酸（例えば、癌の核酸または胎児の核酸）について濃縮することにより、定量的感度が改善され得る。サンプルを特定の核酸種について濃縮するための方法は、例えば、米国特許第 6, 927, 028 号、国際特許出願公開番号 WO 2007/140417、国際特許出願公開番号 WO 2007/147063、国際特許出願公開番号 WO 2009/032779、国際特許出願公開番号 WO 2009/032781、国際特許出願公開番号 WO 2010/033639、国際特許出願公開番号 WO 2011/034631、国際特許出願公開番号 WO 2006/056480 および国際特許出願公開番号 WO 2011/143659 に記載されており、本文、表、式および図面のすべてを含むこれらの各々の全内容が、参照により本明細書中に組み込まれる。

10

【0089】

いくつかの実施形態において、核酸は、ある特定の標的フラグメント種および／または参照フラグメント種について濃縮される。ある特定の実施形態において、核酸は、下記に記載される、長さに基づく 1 つまたはそれを超える分離法を用いて、特定の核酸フラグメント長またはフラグメント長の範囲について濃縮される。ある特定の実施形態において、核酸は、本明細書中に記載されるおよび／または当該分野で公知の、配列に基づく 1 つまたはそれを超える分離法を用いて、選択されたゲノム領域（例えば、染色体）由来のフラグメントについて濃縮される。サンプル中の核酸の部分集団（例えば、胎児核酸）について濃縮するためのある特定の方法は、下記に詳細に記載される。

20

【0090】

本明細書中に記載される方法とともに使用され得る核酸の部分集団（例えば、胎児核酸）を濃縮するためのいくつかの方法には、母体核酸と胎児核酸との間のエピジェネティックな差異を利用する方法が含まれる。例えば、胎児核酸は、メチル化の差異に基づいて鑑別され得、母体核酸から分離され得る。メチル化に基づく胎児核酸の濃縮方法は、参照により本明細書中に組み込まれる米国特許出願公開番号 2010/0105049 に記載されている。そのような方法は、サンプル核酸をメチル化特異的結合物質（メチルー C p G 結合タンパク質（MBD）、メチル化特異的抗体など）に結合する工程、および異なるメチル化の状態に基づいて、結合した核酸を未結合の核酸から分離する工程を含むときがある。そのような方法は、母体核酸を選択的かつ完全にまたは実質的に消化する酵素で母体サンプル由来の核酸を選択的に消化することによりそのサンプルを少なくとも 1 つの胎児核酸領域について濃縮することによって母体サンプル中の胎児核酸領域の濃縮を可能にするメチル化感受性制限酵素（上に記載されたような；例えば、HhaI および HpaII）の使用も含み得る。

30

【0091】

本明細書中に記載される方法とともに使用され得る核酸の部分集団（例えば、胎児核酸）を濃縮するための別の方法は、制限エンドヌクレアーゼによって向上される多型配列アプローチ（例えば、参照により本明細書中に組み込まれる米国特許出願公開番号 2009/0317818 に記載されている方法）である。そのような方法は、非標的対立遺伝子を含む核酸を、その非標的対立遺伝子を含む核酸を認識するが標的対立遺伝子を認識しない制限エンドヌクレアーゼで切断すること；および切断されなかった核酸を増幅するが切断されなかった核酸を増幅しないことを含み、ここで、その切断されずに増幅された核酸は、非標的核酸（例えば、母体核酸）に対して濃縮された標的核酸（例えば、胎児核酸）に相当する。ある特定の実施形態において、核酸は、例えば、それが、切断物質による選択的消化に感受性の多型部位を有する対立遺伝子を含むように選択され得る。

40

【0092】

本明細書中に記載される方法とともに使用され得る核酸の部分集団（例えば、胎児核酸）を濃縮するためのいくつかの方法は、選択的酵素分解アプローチを含む。そのような方法は、標的配列をエキソヌクレアーゼ消化から保護する工程を含み、それにより、サンプル中の望まれない配列（例えば、母体 DNA）の排除が促進される。例えば、1 つのアプ

50

ローチでは、サンプル核酸を変性して一本鎖核酸を生成し、一本鎖核酸を少なくとも1つの標的特異的プライマー対と好適なアニーリング条件下で接触させ、アニールしたプライマーをヌクレオチド重合によって伸長して、二本鎖標的配列を生成し、一本鎖（例えば、非標的）核酸を消化するヌクレアーゼを用いて一本鎖核酸を消化する。ある特定の実施形態において、その方法は、さらに少なくとも1サイクル繰り返され得る。ある特定の実施形態では、1回目および2回目のサイクルの各伸長をプライムするために同じ標的特異的プライマー対を使用し、ある特定の実施形態では、1回目および2回目のサイクルに対して異なる標的特異的プライマー対を使用する。

【0093】

本明細書中に記載される方法とともに使用され得る核酸の部分集団（例えば、胎児核酸）を濃縮するためのいくつかの方法は、大規模並列処理シグネチャ配列決定（MPSS）アプローチを含む。MPSSは、通常、アダプター（例えば、タグ）のライゲーションの後、アダプターの解読および核酸配列の読み取りを少しずつ行う固相方法である。各核酸が、ユニークなタグを有するPCR産物を生成するように、タグ化されたPCR産物が通常増幅される。タグは、PCR産物をマイクロビーズに付着させるために用いられることが多い。数回のライゲーショベースの配列決定の後、例えば、配列シグネチャが、各ビーズから特定され得る。MPSSデータセット内の各シグネチャ配列（MPSSタグ）が解析され、他のすべてのシグネチャと比較され、同一のシグネチャのすべてがカウントされる。

【0094】

ある特定の実施形態において、ある特定の濃縮方法（例えば、ある特定のMPSSおよび/またはMPSSに基づく濃縮方法）は、増幅（例えば、PCR）に基づくアプローチを含み得る。ある特定の実施形態において、遺伝子座特異的増幅方法が使用され得る（例えば、遺伝子座特異的増幅プライマーを使用して）。ある特定の実施形態において、マルチプレックスSNP対立遺伝子PCRアプローチが使用され得る。ある特定の実施形態において、マルチプレックスSNP対立遺伝子PCRアプローチが、ユニプレックス配列決定と組み合わせて使用され得る。例えば、そのようなアプローチは、マルチプレックスPCR（例えば、MASSARRAYシステム）を使用すること、および捕捉プローブ配列をアンプリコンに組み込んだ後、例えばIllumina MPSSシステムを用いて配列決定することを含み得る。ある特定の実施形態において、マルチプレックスSNP対立遺伝子PCRアプローチは、3プライマーシステムおよびインデックスシーケンシングと組み合わせて使用され得る。例えば、そのようなアプローチは、ある特定の遺伝子座特異的順方向PCRプライマーに組み込まれた第1の捕捉プローブおよび遺伝子座特異的逆方向PCRプライマーに組み込まれたアダプター配列を有するプライマーを用いるマルチプレックスPCR（例えば、MASSARRAYシステム）を使用することにより、アンプリコンを生成し、その後、例えばIllumina MPSSシステムを用いた、配列決定に向けて逆方向の捕捉配列および分子インデックスバーコードを組み込むための第2のPCRを含み得る。ある特定の実施形態において、マルチプレックスSNP対立遺伝子PCRアプローチは、4プライマーシステムおよびインデックスシーケンシングと組み合わせて使用され得る。例えば、そのようなアプローチは、遺伝子座特異的順方向PCRプライマーと遺伝子座特異的逆方向PCRプライマーの両方に組み込まれたアダプター配列を有するプライマーを用いるマルチプレックスPCR（例えば、MASSARRAYシステム）の後、例えば、Illumina MPSSシステムを用いた、配列決定に向けて順方向捕捉配列と逆方向捕捉配列の両方および分子インデックスバーコードを組み込むための第2のPCRを使用することを含み得る。ある特定の実施形態において、マイクロフルイディクスアプローチが使用され得る。ある特定の実施形態において、アレイベースのマイクロフルイディクスアプローチが使用され得る。例えば、そのようなアプローチは、低プレックス（lowplex）での増幅ならびにインデックスおよび捕捉プローブの組み込みのためにマイクロフルイディクスアレイ（例えば、Fluidigm）を使用した後の配列決定を含み得る。ある特定の実施形態において、エマルジョンマイクロフルイディ

10

20

30

40

50

クスアプローチ、例えば、デジタル液滴PCRが使用され得る。

【0095】

ある特定の実施形態において、ユニバーサル増幅法が使用され得る（例えば、ユニバーサル増幅プライマーまたは非遺伝子座特異的増幅プライマーを用いて）。ある特定の実施形態において、ユニバーサル増幅法は、プルダウンアプローチと組み合わせて使用され得る。ある特定の実施形態において、ある方法は、広く増幅された配列決定ライブラリーからのビオチン化ウルトラマー（ultramer）プルダウン（例えば、AgilentまたはIDT製のビオチン化プルダウンアッセイ）を含み得る。例えば、そのようなアプローチは、標準的なライブラリーの調製、プルダウンアッセイによる選択された領域の濃縮、および第2のユニバーサル増幅工程を含み得る。ある特定の実施形態において、プル
10
ダウンアプローチは、ライゲーションに基づく方法と組み合わせて使用され得る。ある特定の実施形態において、ある方法は、配列特異的アダプターライゲーションを用いるビオチン化ウルトラマープルダウン（例えば、HALOPLEX PCR, Halo Genomics）を含み得る。例えば、そのようなアプローチは、選択プローブを使用して、制限酵素によって消化されたフラグメントを捕捉した後、捕捉された生成物とアダプターとをライゲーションし、ユニバーサル増幅した後、配列決定することを含み得る。ある特定の
20
実施形態において、プルダウンアプローチは、伸長およびライゲーションに基づく方法と組み合わせて使用され得る。ある特定の実施形態において、ある方法は、分子反転プローブ（molecular inversion probe）（MIP）の伸長およびライゲーションを含み得る。例えば、そのようなアプローチは、分子反転プローブを配列アダプターと組み合わせて使用した後、ユニバーサル増幅および配列決定を行うこと
20

【0096】

ある特定の実施形態において、伸長およびライゲーションのアプローチは、プルダウンの構成要素なしに行われ得る。ある特定の実施形態において、ある方法は、遺伝子座特異的順方向プライマーおよび遺伝子座特異的逆方向プライマーのハイブリダイゼーション、伸長およびライゲーションを含み得る。そのような方法は、ユニバーサル増幅、または増幅なしの相補DNA合成の後の配列決定をさらに含み得る。ある特定の実施形態において、そのような方法は、解析中のバックグラウンド配列を減少させ得るか、または排除し得る。
30

【0097】

ある特定の実施形態において、プルダウンアプローチは、随意の増幅構成要素を用いて、または増幅構成要素を用いずに、使用され得る。ある特定の実施形態において、ある方法は、ユニバーサル増幅なしで捕捉プローブの完全な組み込みを伴う、改変されたプルダウンアッセイおよびライゲーションを含み得る。例えば、そのようなアプローチは、改変された選択プローブを使用して、制限酵素によって消化されたフラグメントを捕捉した後の、捕捉された生成物とアダプターとのライゲーション、随意の増幅および配列決定を含み得る。ある特定の実施形態において、ある方法は、環状一本鎖ライゲーションと組み合わせて、アダプター配列の伸長およびライゲーションを用いるビオチン化プルダウンアッセイを含み得る。例えば、そのようなアプローチは、選択プローブを使用して目的の領域（例えば、標的配列）を捕捉すること、そのプローブの伸長、アダプターのライゲーション、一本鎖環状ライゲーション、随意の増幅および配列決定を含み得る。ある特定の
40
実施形態において、配列決定の結果の解析は、標的配列をバックグラウンドから分離し得る。

【0098】

いくつかの実施形態において、核酸は、本明細書中に記載される、配列に基づく1つまたはそれを超える分離方法を用いて、選択されたゲノム領域（例えば、染色体）由来のフラグメントについて濃縮される。配列に基づく分離は、一般に、目的のフラグメント（例えば、標的フラグメントおよび／または参照フラグメント）に存在し、かつサンプルの他のフラグメントには実質的に存在しないかまたはごくわずかな量の他のフラグメント（例
50

えば、5%またはそれ未満)に存在する、ヌクレオチド配列に基づく。いくつかの実施形態において、配列に基づく分離は、分離された標的フラグメントおよび／または分離された参照フラグメントを生成し得る。分離された標的フラグメントおよび／または分離された参照フラグメントは、その核酸サンプル中の残りのフラグメントから単離されることが多い。ある特定の実施形態において、その分離された標的フラグメントおよび分離された参照フラグメントは、互いからも単離される(例えば、別個のアッセイコンパートメントに単離される)。ある特定の実施形態において、その分離された標的フラグメントおよび分離された参照フラグメントは、共に単離される(例えば、同じアッセイコンパートメントに単離される)。いくつかの実施形態において、未結合のフラグメントは、差次的に除去され得るか、または分解され得るか、または消化され得る。

10

【0099】

いくつかの実施形態において、選択的な核酸捕捉プロセスが、標的フラグメントおよび／または参照フラグメントを核酸サンプルから分離するために使用される。商業的に入手可能な核酸捕捉システムとしては、例えば、Nimblegen配列捕捉システム(Roche NimbleGen, Madison, WI); Illumina BEADARRAYプラットフォーム(Illumina, San Diego, CA); Affymetrix GENECHIPプラットフォーム(Affymetrix, Santa Clara, CA); Agilent SureSelect Target Enrichment System(Agilent Technologies, Santa Clara, CA); および関連するプラットフォームが挙げられる。そのような方法は、通常、捕捉オリゴヌクレオチドと、標的フラグメントもしくは参照フラグメントのセグメントまたは標的フラグメントもしくは参照フラグメントのすべてのヌクレオチド配列とのハイブリダイゼーションを含み、固相(例えば、固相アレイ)および／または溶液ベースのプラットフォームの使用を含み得る。捕捉オリゴヌクレオチド(「ベイト」と称されるときがある)は、それらが、選択されたゲノム領域または遺伝子座(例えば、21番、18番、13番、XもしくはY染色体のうちの1つまたは参照染色体)由来の核酸フラグメントに優先的にハイブリダイズするように、選択され得るかまたはデザインされ得る。ある特定の実施形態において、ハイブリダイゼーションに基づく方法(例えば、オリゴヌクレオチドアレイを用いる方法)は、ある特定の染色体(例えば、潜在的に異数性の染色体、参照染色体または他の目的の染色体)由来の核酸配列またはその目的のセグメントを濃縮するために使用され得る。

20

30

【0100】

いくつかの実施形態において、核酸は、1つまたはそれを超える、長さに基づく分離方法を用いて、特定の核酸フラグメントの長さ、長さの範囲、または特定の閾値もしくはカットオフを下回るもしくは上回る長さについて濃縮される。核酸フラグメントの長さは、通常、そのフラグメントにおけるヌクレオチドの数のことを指す。核酸フラグメントの長さは、核酸フラグメントのサイズと称されるときもある。いくつかの実施形態において、長さに基づく分離方法は、個々のフラグメントの長さを計測せずに行われる。いくつかの実施形態において、長さに基づく分離方法は、個々のフラグメントの長さを測定するための方法とともに行われる。いくつかの実施形態において、長さに基づく分離とは、分画されたプールの全部または一部が単離され得る(例えば、保持され得る)および／または解析され得るサイズ分画手順のことを指す。サイズ分画手順は、当該分野で公知である(例えば、アレイ上での分離、モレキュラーシーブによる分離、ゲル電気泳動による分離、カラムクロマトグラフィーによる分離(例えば、サイズ排除カラム)およびマイクロfluidicsスペースのアプローチ)。ある特定の実施形態において、長さに基づく分離アプローチは、例えば、フラグメントの環状化、化学的処理(例えば、ホルムアルデヒド、ポリエチレングリコール(PEG))、質量分析および／またはサイズ特異的核酸増幅を含み得る。

40

【0101】

本明細書中に記載される方法とともに使用され得るある特定の長さに基づく分離方法は

50

、例えば、選択的な配列タギングアプローチを用いる。用語「配列タギング」とは、認識可能な異なる配列を核酸または核酸集団に組み込むことを指す。用語「配列タギング」は、本明細書中で使用されるとき、本明細書中の後に記載される用語「配列タグ」と異なる意味を有する。そのような配列タギング方法において、フラグメントサイズ種（例えば、短いフラグメント）核酸は、長い核酸および短い核酸を含むサンプル中で選択的な配列タギングに供される。そのような方法は、通常、内側プライマーおよび外側プライマーを含むネステッドプライマーのセットを用いて核酸増幅反応を行う工程を含む。ある特定の実施形態において、その内側プライマーの一方または両方が、タグ化されることにより、標的増幅産物にタグが導入され得る。外側プライマーは、一般に、（内側の）標的配列を有する短いフラグメントにアニールしない。内側プライマーは、その短いフラグメントにアニールでき、タグおよび標的配列を有する増幅産物を生成し得る。代表的には、長いフラグメントのタギングは、例えば、外側プライマーの事前のアニーリングおよび伸長による内側プライマーの伸長の阻止を含む、機序の組み合わせによって阻害される。タグ化されたフラグメントの濃縮は、例えば、一本鎖核酸のエキソヌクレアーゼ消化、および少なくとも1つのタグに特異的な増幅プライマーを用いるタグ化されたフラグメントの増幅を含む、種々の方法のいずれかによって達成され得る。

10

【0102】

本明細書中に記載される方法とともに使用され得る別の長さに基づく分離方法は、核酸サンプルをポリエチレングリコール（PEG）沈殿に供する工程を含む。方法の例としては、国際特許出願公開番号WO2007/140417およびWO2010/115016（本文、表、式および図面のすべてを含むこれらの各々の全内容が参照により本明細書中に組み込まれる）に記載されている方法が挙げられる。この方法は、一般に、小さい（例えば、300ヌクレオチド未満の）核酸を実質的に沈殿させずに大きな核酸を実質的に沈殿させるのに十分な条件下において、1つまたはそれを超える一価塩の存在下において、核酸サンプルをPEGと接触させる工程を必要とする。

20

【0103】

本明細書中に記載される方法とともに使用され得る別のサイズに基づく濃縮方法は、ライゲーションによる環状化、例えば、サークリガーゼ（*circ ligase*）を用いる環状化を含む。短い核酸フラグメントは、通常、長いフラグメントよりも高効率で環状化され得る。環状化していない配列は、環状化した配列から分離され得、濃縮された短いフラグメントは、さらなる解析のために使用され得る。

30

核酸ライブラリー

【0104】

いくつかの実施形態において、核酸ライブラリーは、特定のプロセスのために調製され、アセンブルされ、および／または改変された、複数のポリヌクレオチド分子（例えば、核酸のサンプル）であり、そのプロセスの非限定的な例としては、固相（例えば、固体支持体、例えば、フローセル、ビーズ）への固定化、濃縮、増幅、クローニング、検出および／または核酸配列決定が挙げられる。ある特定の実施形態において、核酸ライブラリーは、配列決定プロセスの前または最中に調製される。核酸ライブラリー（例えば、配列決定ライブラリー）は、当該分野で公知であるような好適な方法によって調製され得る。核酸ライブラリーは、標的化されたまたは標的化されない調製プロセスによって調製され得る。

40

【0105】

いくつかの実施形態において、核酸のライブラリーは、核酸を固体支持体に固定化するように構成された化学的部分（例えば、官能基）を含むように改変される。いくつかの実施形態において、核酸のライブラリーは、そのライブラリーを固体支持体に固定化するように構成された、生体分子（例えば、官能基）および／または結合対のメンバーを含むように改変され、それらの非限定的な例としては、チロキシン結合グロブリン、ステロイド結合タンパク質、抗体、抗原、ハプテン、酵素、レクチン、核酸、リプレッサー、プロテインA、プロテインG、アビジン、ストレプトアビジン、ビオチン、補体成分C1q、核

50

酸結合タンパク質、レセプター、炭水化物、オリゴヌクレオチド、ポリヌクレオチド、相補的な核酸配列などおよびそれらの組み合わせが挙げられる。特異的結合対のいくつかの例としては、アビジン部分とビオチン部分；抗原性エピトープと抗体またはその免疫学的に反応性のフラグメント；抗体とハプテン；ジゴキシゲニン（digoxigenin）部分と抗ジゴキシゲニン抗体；フルオレセイン部分と抗フルオレセイン抗体；オペレーターとリプレッサー；ヌクレアーゼとヌクレオチド；レクチンと多糖；ステロイドとステロイド結合タンパク質；活性な化合物と活性な化合物のレセプター；ホルモンとホルモンレセプター；酵素と基質；免疫グロブリンとプロテインA；オリゴヌクレオチドまたはポリヌクレオチドとその対応する相補鎖；などまたはそれらの組み合わせが挙げられるが、これらに限定されない。

10

【0106】

いくつかの実施形態において、核酸のライブラリーは、既知の組成の1つまたはそれを超えるポリヌクレオチドを含むように改変され、その非限定的な例としては、識別子（例えば、タグ、インデックスタグ）、捕捉配列、標識、アダプター、制限酵素部位、プロモーター、エンハンサー、複製起点、ステムループ、相補的な（complementary）配列（例えば、プライマー結合部位、アニーリング部位）、好適なインテグレーション部位（例えば、トランスポゾン、ウイルスインテグレーション部位）、修飾ヌクレオチドなどまたはそれらの組み合わせが挙げられる。既知の配列のポリヌクレオチドは、好適な位置に、例えば、5'末端、3'末端または核酸配列内に付加され得る。既知の配列のポリヌクレオチドは、同じまたは異なる配列であり得る。いくつかの実施形態において、既知の配列のポリヌクレオチドは、表面（例えば、フローセル内の表面）上に固定化された1つまたはそれを超えるオリゴヌクレオチドにハイブリダイズするように構成される。例えば、既知の5'配列を含む核酸分子は、第1の複数のオリゴヌクレオチドにハイブリダイズし得るのに対して、既知の3'配列は、第2の複数のオリゴヌクレオチドにハイブリダイズし得る。いくつかの実施形態において、核酸のライブラリーは、染色体特異的タグ、捕捉配列、標識および／またはアダプターを含み得る。いくつかの実施形態において、核酸のライブラリーは、1つまたはそれを超える検出可能な標識を含む。いくつかの実施形態において、1つまたはそれを超える検出可能な標識は、5'末端、3'末端および／または核酸ライブラリー中の核酸内の任意のヌクレオチド位置において、ライブラリーに組み込まれ得る。いくつかの実施形態において、核酸のライブラリーは、ハイブリダイズされたオリゴヌクレオチドを含む。ある特定の実施形態において、ハイブリダイズされたオリゴヌクレオチドは、標識されたプローブである。いくつかの実施形態において、核酸のライブラリーは、固相上への固定化の前のハイブリダイズされたオリゴヌクレオチドプローブを含む。

20

30

【0107】

いくつかの実施形態において、既知の配列のポリヌクレオチドは、ユニバーサル配列を含む。ユニバーサル配列は、2つもしくはそれを超える核酸分子または核酸分子の2つもしくはそれを超えるサブセットにインテグレートされた特定のヌクレオチド酸配列であり、ここで、そのユニバーサル配列は、それがインテグレートされたすべての分子または分子のサブセットに対して同じである。ユニバーサル配列は、複数の異なる配列にハイブリダイズするように、および／またはユニバーサル配列に相補的な単一のユニバーサルプライマーを用いて複数の異なる配列を増幅するようにデザインされることが多い。いくつかの実施形態において、2つ（例えば、1対）またはそれを超えるユニバーサル配列および／またはユニバーサルプライマーが使用される。ユニバーサルプライマーは、ユニバーサル配列を含むことが多い。いくつかの実施形態において、アダプター（例えば、ユニバーサルアダプター）は、ユニバーサル配列を含む。いくつかの実施形態において、1つまたはそれを超えるユニバーサル配列は、複数の核酸種または核酸サブセットを捕捉するため、同定するため、および／または検出するために使用される。

40

【0108】

核酸ライブラリーを調製するある特定の実施形態において（例えば、合成手順によるあ

50

る特定の配列決定において）、核酸は、サイズ選択されるおよび／または数百塩基対またはそれ未満の長さに断片化される（例えば、ライブラリー作製のための調製において）。いくつかの実施形態において、ライブラリーの調製は、断片化なしに行われる（例えば、c c f DNAを使用するとき）。

【0109】

ある特定の実施形態において、ライゲーションに基づくライブラリー調製方法が使用される（例えば、ILLUMINA TRUSEQ, Illumina, San Diego CA）。ライゲーションに基づくライブラリー調製方法は、最初のライゲーション工程においてインデックス配列を組み込み得るアダプター（例えば、メチル化されたアダプター）デザインを利用することが多く、単一リードの配列決定、ペアエンド配列決定および多重化された配列決定のためにサンプルを調製するために使用され得ることが多い。例えば、核酸（例えば、断片化された核酸またはc c f DNA）は、フィルイン（fill-in）反応、エキソヌクレアーゼ反応またはそれらの組み合わせによって末端が修復されるときがある。いくつかの実施形態において、得られた平滑末端に修復された核酸は、次いで、アダプター／プライマーの3'末端における単一ヌクレオチドのオーバーハングに相補的な単一ヌクレオチドだけ伸長され得る。任意のヌクレオチドが、伸長／オーバーハングヌクレオチドのために使用され得る。いくつかの実施形態において、核酸ライブラリーの調製は、アダプターオリゴヌクレオチドをライゲートすることを含む。アダプターオリゴヌクレオチドは、フローセルアンカーに相補的であることが多く、核酸ライブラリーを固体支持体（例えば、フローセルの内側表面）に固定化するために使用されるときがある。いくつかの実施形態において、アダプターオリゴヌクレオチドは、識別子、1つもしくははそれを超える配列決定プライマーハイブリダイゼーション部位（例えば、ユニバーサル配列決定プライマー、シングルエンド配列決定プライマー、ペアエンド配列決定プライマー、多重化された配列決定プライマーなどに相補的な配列）またはそれらの組み合わせ（例えば、アダプター／配列決定、アダプター／識別子、アダプター／識別子／配列決定）を含む。

【0110】

識別子は、その識別子を含む核酸の検出および／または識別を可能にする核酸（例えば、ポリヌクレオチド）内に組み込まれるかまたはその核酸に付着される好適な検出可能な標識であり得る。いくつかの実施形態において、識別子は、配列決定方法（例えば、ポリメラーゼによるもの）の間に、核酸内に組み込まれるかまたは核酸に付着される。識別子の非限定的な例としては、核酸タグ、核酸インデックスまたはバーコード、放射標識（例えば、同位体）、金属標識、蛍光標識、化学発光標識、リン光性標識、フルオロフォアクエンチャー、色素、タンパク質（例えば、酵素、抗体またはその一部、リンカー、結合対のメンバー）などまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、識別子（例えば、核酸インデックスまたはバーコード）は、ヌクレオチドまたはヌクレオチドアナログのユニークな配列、既知の配列および／または識別可能な配列である。いくつかの実施形態において、識別子は、6つまたはそれを超えて連続したヌクレオチドである。種々の異なる励起スペクトルおよび発光スペクトルを有する多数のフルオロフォアが利用可能である。任意の好適なタイプおよび／または数のフルオロフォアが、識別子として使用され得る。いくつかの実施形態において、1つもしくははそれを超える、2つもしくははそれを超える、3つもしくははそれを超える、4つもしくははそれを超える、5つもしくははそれを超える、6つもしくははそれを超える、7つもしくははそれを超える、8つもしくははそれを超える、9つもしくははそれを超える、10個もしくははそれを超える、20個もしくははそれを超える、30個もしくははそれを超えるまたは50個もしくははそれを超える異なる識別子が、本明細書中に記載される方法（例えば、核酸検出方法および／または配列決定方法）において使用される。いくつかの実施形態において、1つまたは2つのタイプの識別子（例えば、蛍光標識）が、ライブラリー内の各核酸に連結される。識別子の検出および／または定量は、好適な方法、装置または機器によって行われ得、それらの非限定的な例としては、フローサイトメトリー、定量的ポリメラーゼ連鎖反応（qPCR）、ゲル電

10

20

30

40

50

気泳動、ルミノメーター、蛍光光度計、分光光度計、好適な遺伝子チップまたはマイクロアレイ解析、ウエスタンブロット、質量分析、クロマトグラフィー、細胞蛍光分析、蛍光顕微鏡法、好適な蛍光またはデジタルイメージング法、共焦点レーザー走査顕微鏡法、レーザー走査型サイトメトリー、アフィニティークロマトグラフィー、手作業のバッチモード分離、電界懸架、好適な核酸配列決定法および／または核酸配列決定装置などおよびそれらの組み合わせが挙げられる。

【0111】

いくつかの実施形態において、トランスポゾンに基づくライブラリー調製方法が使用される（例えば、E P I C E N T R E N E X T E R A, E p i c e n t r e, M a d i s o n W I）。トランスポゾンに基づく方法は、通常、単一チューブ反応（プラットフォーム特異的タグおよび随意のバーコードの組み込みを可能にすることが多い）においてD N Aを同時に断片化し、タグ化するためにインビトロでの転位を使用し、シークエンサー対応型ライブラリーを調製する。

10

【0112】

いくつかの実施形態において、核酸ライブラリーまたはその一部は、増幅される（例えば、P C Rに基づく方法によって増幅される）。いくつかの実施形態において、配列決定法は、核酸ライブラリーの増幅を含む。核酸ライブラリーは、固体支持体（例えば、フローセル内の固体支持体）上への固定化の前または後に増幅され得る。核酸増幅は、1コピーまたはそれを超えるコピー数の鋳型および／またはその相補鎖を生成することによって、存在する（例えば、核酸ライブラリー中に存在する）核酸鋳型および／またはその相補鎖を増幅するかまたはその数を増加させるプロセスを含む。増幅は、好適な方法によって行われ得る。核酸ライブラリーは、サーモサイクリング法または等温増幅法によって増幅され得る。いくつかの実施形態において、ローリングサークル増幅法が使用される。いくつかの実施形態において、増幅は、核酸ライブラリーまたはその一部が固定化された固体支持体上（例えば、フローセル内）で行われる。ある特定の配列決定法において、核酸ライブラリーが、フローセルに加えられ、好適な条件下におけるアンカーへのハイブリダイゼーションによって固定化される。このタイプの核酸増幅は、固相増幅と称されることが多い。固相増幅のいくつかの実施形態において、増幅産物の全部または一部が、固定化されたプライマーから開始する伸長によって合成される。固相増幅反応は、増幅オリゴヌクレオチド（例えば、プライマー）の少なくとも1つが、固体支持体上に固定化されていることを除いては、標準的な溶相増幅と類似している。

20

30

【0113】

いくつかの実施形態において、固相増幅は、表面に固定化されたただ1種のオリゴヌクレオチドプライマーを含む核酸増幅反応を含む。ある特定の実施形態において、固相増幅は、複数の異なる固定化されたオリゴヌクレオチドプライマー種を含む。いくつかの実施形態において、固相増幅は、固体表面上に固定化された1つの種のオリゴヌクレオチドプライマーおよび溶液中の第2の異なるオリゴヌクレオチドプライマー種を含む核酸増幅反応を含み得る。複数の異なる種の固定化されたプライマーまたは溶液ベースのプライマーが、使用され得る。固相核酸増幅反応の非限定的な例としては、界面増幅、ブリッジ増幅、エマルジョンP C R、W i l d F i r e増幅（例えば、米国特許公開U S 2 0 1 3 0 0 1 2 3 9 9）などまたはそれらの組み合わせが挙げられる。

40

配列決定

【0114】

いくつかの実施形態において、核酸（例えば、核酸フラグメント、サンプル核酸、無細胞核酸）が、配列決定される。ある特定の実施形態において、完全なまたは実質的に完全な配列が得られ、部分的な配列が得られるときがある。

【0115】

いくつかの実施形態において、サンプル中のいくつかまたはすべての核酸が、配列決定の前または配列決定中に、濃縮および／または増幅される（例えば、非特異的に、例えば、P C Rに基づく方法によって）。ある特定の実施形態において、サンプル中の特定の核

50

酸部分またはサブセットが、配列決定の前または配列決定中に、濃縮および／または増幅される。いくつかの実施形態において、予め選択された核酸プールの一部またはサブセットが、ランダムに配列決定される。いくつかの実施形態において、サンプル中の核酸は、配列決定の前または配列決定中に濃縮および／または増幅されない。

【0116】

本明細書中で使用されるとき、「リード (reads)」(例えば、「リード (aread)」)、「配列リード」)は、本明細書中に記載されるまたは当該分野で公知の任意の配列決定プロセスによって生成される短いヌクレオチド配列である。リードは、核酸フラグメントの一方の末端から生成され得(「シングルエンドリード」)、核酸の両方の末端から生成されるときもある(例えば、ペアエンドリード、ダブルエンドリード)。

10

【0117】

配列リードの長さは、特定の配列決定技術に関連することが多い。例えば、ハイスループット法は、サイズが数十から数百塩基対(bp)まで異なり得る配列リードを提供する。例えば、ナノポア配列決定は、サイズが数十、数百から数千の塩基対まで異なり得る配列リードを提供し得る。いくつかの実施形態において、配列リードは、約15bp～約900bp長という長さの平均値、中央値、平均の長さまたは絶対的な長さである。ある特定の実施形態において、配列リードは、約1000bpまたはそれを超える長さの平均値、中央値、平均の長さまたは絶対的な長さである。

【0118】

いくつかの実施形態において、シングルエンドリードの名目上の長さ、平均の長さ、長さの平均値または絶対的な長さは、約15個連続したヌクレオチド～約50個もしくはそれを超えて連続したヌクレオチド、約15個連続したヌクレオチド～約40個もしくはそれを超えて連続したヌクレオチドであるときがあり、約15個連続したヌクレオチドまたは約36個もしくはそれを超えて連続したヌクレオチドであるときがある。ある特定の実施形態において、シングルエンドリードの名目上の長さ、平均の長さ、長さの平均値または絶対的な長さは、約20～約30塩基長または約24～約28塩基長である。ある特定の実施形態において、シングルエンドリードの名目上の長さ、平均の長さ、長さの平均値または絶対的な長さは、約1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、21、22、23、24、25、26、27、28もしくは約29塩基長またはそれを超える長さである。

20

30

【0119】

ある特定の実施形態において、ペアエンドリードの名目上の長さ、平均の長さ、長さの平均値または絶対的な長さは、約10個連続したヌクレオチド～約25個連続したヌクレオチドまたはそれを超えるヌクレオチド(例えば、約10、11、12、13、14、15、16、17、18、19、20、21、22、23、24もしくは25ヌクレオチド長またはそれを超えるヌクレオチド長)、約15個連続したヌクレオチド～約20個連続したヌクレオチドまたはそれを超えるヌクレオチドであるときがあり、約17個連続したヌクレオチドまたは約18個連続したヌクレオチドであるときがある。

【0120】

リードは、一般に、物理的な核酸におけるヌクレオチド配列の提示である。例えば、A T G C描写の配列を含むリードにおいて、物理的な核酸では、「A」は、アデニンヌクレオチドを表し、「T」は、チミンヌクレオチドを表し、「G」は、グアニンヌクレオチドを表し、「C」は、シトシンヌクレオチドを表す。妊婦の血液から得られた配列リードは、胎児核酸と母体核酸との混合物からのリードであり得る。比較的短いリードの混合物は、本明細書中に記載されるプロセスによって、その妊婦および／または胎児に存在するゲノム核酸の提示に変換され得る。比較的短いリードの混合物は、例えば、コピー数変異(例えば、母体および／または胎児のコピー数変異)、遺伝的変異または異数性の提示に変換され得る。母体核酸と胎児核酸との混合物のリードは、母体および胎児の染色体の一方または両方の特徴を含む複合染色体またはそのセグメントの提示に変換され得る。ある特定の実施形態において、被験体からサンプルの核酸配列リードを「得ること」および／ま

40

50

たは1人もしくはそれを超える参照人から生物学的検体の核酸配列リードを「得ること」は、核酸を直接配列決定して配列情報を得ることを含み得る。いくつかの実施形態において、「得ること」は、別のものによって核酸から直接得られた配列情報を受け取ることを含み得る。

【0121】

いくつかの実施形態において、ゲノムの代表的な一部が配列決定され、それは、「カバレッジ」または「倍カバレッジ」と称されるときがある。例えば、1倍カバレッジは、そのゲノムのおおよそ100%のヌクレオチド配列が、リードによって表されたことを示唆する。いくつかの実施形態において、「倍カバレッジ」は、事前の配列決定ランを参照として言及している相対的な用語である。例えば、第2の配列決定ランは、第1の配列決定ランよりも2倍未満のカバレッジを有し得る。いくつかの実施形態において、ゲノムは、重複して配列決定され、ここで、所与のゲノム領域が、2つまたはそれを超えるリードまたはオーバーラップリードによってカバーされ得る（例えば、1より大きい「倍カバレッジ」、例えば、2倍カバレッジ）。

【0122】

いくつかの実施形態において、1つの個体由来の1つの核酸サンプルが、配列決定される。ある特定の実施形態において、2つまたはそれを超えるサンプルの各々からの核酸が、配列決定され、ここで、サンプルは、1つの個体由来であるか、または異なる個体由来である。ある特定の実施形態において、2つまたはそれを超える生物学的サンプル由来の核酸サンプルがプールされ、ここで、各生物学的サンプルは、1つの個体由来であるかまたは2つもしくはそれを超える個体由来であり、そのプールが、配列決定される。後者の実施形態では、各生物学的サンプル由来の核酸サンプルは、1つまたはそれを超えるユニークな識別子によって識別されることが多い。

【0123】

いくつかの実施形態において、配列決定法は、配列決定プロセスにおいて配列決定反応の多重化を可能にする識別子を使用する。ユニークな識別子の数が増えるほど、配列決定プロセスにおいて多重化され得る、例えば、検出のためのサンプルおよび／または染色体の数は増える。配列決定プロセスは、任意の好適な数の（例えば、4、8、12、24、48、96個またはそれを超える）ユニークな識別子を用いて行われ得る。

【0124】

配列決定プロセスは、固相を利用するときがあり、その固相は、ライブラリー由来の核酸が付着され得、試薬が流され得、付着された核酸と接触し得るフローセルを含むときがある。フローセルは、フローセルレーンを備えるときがあり、識別子の使用により、各レーンにおいていくつかのサンプルを解析することが容易になり得る。フローセルは、結合したアナライトの上に試薬溶液を保持するようにおよび／または結合したアナライトの上に試薬溶液を順序正しく通過させることを可能にするように構成され得る、固体支持体であることが多い。フローセルは、しばしば、平面の形状であり、光学的に透明であり、一般に、ミリメートルまたはミリメートル未満のスケールであり、アナライト／試薬相互作用が生じるチャンネルまたはレーンを有することが多い。いくつかの実施形態において、所与のフローセルレーンにおいて解析されるサンプルの数は、ライブラリー調製中および／またはプローブデザイン中に使用されたユニークな識別子の数に依存する。単一フローセルレーン。12個の識別子を使用した多重化は、例えば、8レーンフローセルにおいて、96個のサンプル（例えば、96ウェルのマイクロウェルプレートにおけるウェルの数に等しい）を同時に解析することを可能にする。同様に、48個の識別子を使用した多重化は、例えば、8レーンフローセルにおいて、384個のサンプル（例えば、384ウェルのマイクロウェルプレートにおけるウェルの数に等しい）を同時に解析することを可能にする。商業的に入手可能なマルチプレックス配列決定キットの非限定的な例としては、Illuminaのマルチプレックスサンプル調製オリゴヌクレオチドキットおよびマルチプレックス配列決定プライマーおよびPhiXコントロールキット（例えば、それぞれIlluminaのカatalog番号PE-400-1001およびPE-400-1002）

が挙げられる。

【0125】

核酸を配列決定する任意の好適な方法が使用され得、その非限定的な例としては、Maxim & Gilbert、チェーン・ターミネーション法、合成による配列決定、ライゲーションによる配列決定、質量分析による配列決定、顕微鏡法に基づく手法などまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、第1世代の技術、例えば、マイクロフルイディクスSanger配列決定を含む自動化されたSanger配列決定法を含むSanger配列決定法が、本明細書中に提供される方法において使用され得る。いくつかの実施形態において、核酸イメージング技術（例えば、透過型電子顕微鏡（TEM）および原子間力顕微鏡法（AFM））の使用を含む配列決定技術が使用され得る。いくつかの実施形態において、ハイスループット配列決定法が使用される。ハイスループット配列決定法は、一般に、大規模並列処理形式で、時折、フローセル内において、配列決定される、クローン増幅されたDNA鋳型または単一DNA分子を必要とする。大規模並列処理形式でDNAを配列決定することができる次世代（例えば、第2および第3世代）配列決定法は、本明細書中に記載される方法のために使用され得、本明細書中において集合的に「大規模並列処理配列決定」（MPS）と称される。いくつかの実施形態において、MPS配列決定法は、標的化アプローチを使用し、ここで、特定の染色体、遺伝子または目的の領域が配列決定される。ある特定の実施形態において、サンプル中のほとんどまたはすべての核酸が、ランダムに配列決定される、増幅される、および／または捕捉される、非標的化アプローチが使用される。

10

20

【0126】

いくつかの実施形態において、標的化された濃縮、増幅および／または配列決定アプローチが使用される。標的化アプローチは、配列特異的オリゴヌクレオチドを使用することによってさらに処理するために、サンプル中の核酸のサブセットを単離する、選択する、および／または濃縮することが多い。いくつかの実施形態において、配列特異的オリゴヌクレオチドのライブラリーが、サンプル中の1つまたはそれを超える核酸セットを標的化する（例えば、ハイブリダイズさせる）ために使用される。配列特異的オリゴヌクレオチドおよび／またはプライマーは、1つまたはそれを超える目的の染色体、遺伝子、エキソン、イントロンおよび／または制御領域に存在する特定の配列（例えば、ユニークな核酸配列）に対して選択的であることが多い。任意の好適な方法または方法の組み合わせが、1つまたはそれを超える標的化された核酸のサブセットの濃縮、増幅および／または配列決定のために使用され得る。いくつかの実施形態において、標的化された配列は、1つまたはそれを超える配列特異的アンカーを使用した固相（例えば、フローセル、ビーズ）への捕捉によって、単離および／または濃縮される。いくつかの実施形態において、標的化された配列は、配列特異的プライマーおよび／またはプライマーセットを使用した、ポリメラーゼに基づく方法（例えば、PCRに基づく方法、任意の好適なポリメラーゼに基づく伸長）によって、濃縮および／または増幅される。配列特異的アンカーが、配列特異的プライマーとして使用され得ることが多い。

30

【0127】

MPS配列決定は、合成およびある特定のイメージングプロセスによる配列決定を利用するときがある。本明細書中に記載される方法において使用され得る核酸配列決定技術は、合成による配列決定および可逆的ターミネーターに基づく配列決定（例えば、Illumina's Genome Analyzer; Genome Analyzer II; HISEQ 2000; HISEQ 2500 (Illumina, San Diego CA)）である。この技術を用いれば、数百万の核酸（例えば、DNA）フラグメントを並行して配列決定することができる。このタイプの配列決定技術の1つの例では、オリゴヌクレオチドアンカー（例えば、アダプタープライマー）が結合した表面上に8つの個別のレーンを有する光学的に透明なスライドを備えるフローセルが使用される。フローセルは、結合したアナライトの上に試薬溶液を保持するように、および／または結合したアナライトの上に試薬溶液を順序正しく通過させることを可能にするように、構成され

40

50

得る固体支持体であることが多い。フローセルは、しばしば、平面の形状であり、光学的に透明であり、一般に、ミリメートルまたはミリメートル未満のスケールであり、アナライト／試薬相互作用が生じるチャンネルまたはレーンを有することが多い。

【0128】

合成による配列決定は、いくつかの実施形態において、鋳型特異的な様式で、ヌクレオチドをプライマーまたは既存の核酸鎖に反復して付加すること（例えば、共有結合性の付加によって）を含む。ヌクレオチドの反復付加の各々が検出され、そのプロセスは、核酸鎖の配列が得られるまで複数回繰り返される。得られる配列の長さは、行われる付加工程および検出工程の数に部分的に依存する。合成による配列決定のいくつかの実施形態では、1回のヌクレオチド付加において、同じタイプ（例えば、A、G、CまたはT）の1つ、2つ、3つもしくはそれを超えるヌクレオチドが付加され、検出される。ヌクレオチドは、任意の好適な方法によって（例えば、酵素的または化学的に）付加され得る。例えば、いくつかの実施形態において、ポリメラーゼまたはリガーゼが、鋳型特異的な様式で、プライマーまたは既存の核酸鎖にヌクレオチドを付加する。合成による配列決定のいくつかの実施形態において、異なるタイプのヌクレオチド、ヌクレオチドアナログおよび／または識別子が使用される。いくつかの実施形態において、可逆的ターミネーターおよび／または除去可能な（例えば、切断可能な）識別子が使用される。いくつかの実施形態において、蛍光標識されたヌクレオチドおよび／またはヌクレオチドアナログが使用される。ある特定の実施形態において、合成による配列決定は、切断（例えば、識別子の切断および除去）および／または洗浄工程を含む。いくつかの実施形態において、1つまたはそれを超えるヌクレオチドの付加は、本明細書中に記載されるまたは当該分野で公知の好適な方法によって検出され、その非限定的な例としては、任意の好適なイメージング装置、好適なカメラ、デジタルカメラ、CCD（電荷結合素子）に基づくイメージング装置（例えば、CCDカメラ）、CMOS（相補型金属酸化物半導体（Complementary Metal Oxide Silicon））に基づくイメージング装置（例えば、CMOSカメラ）、フォトダイオード（例えば、光電子増倍管）、電子顕微鏡法、電界効果トランジスタ（例えば、DNA電界効果トランジスタ）、ISFETイオンセンサー（例えば、CHEMFETセンサー）などまたはそれらの組み合わせが挙げられる。本明細書中の方法を行うために使用され得る他の配列決定法としては、デジタルPCRおよびハイブリダイゼーションによる配列決定が挙げられる。

【0129】

本明細書中の方法を行うために使用され得る他の配列決定法としては、デジタルPCRおよびハイブリダイゼーションによる配列決定が挙げられる。デジタルポリメラーゼ連鎖反応（デジタルPCRまたはdPCR）は、サンプル中の核酸を直接識別して定量するために使用され得る。デジタルPCRは、いくつかの実施形態において、エマルジョン内で行われ得る。例えば、個々の核酸が、例えば、マイクロ流体チャンバーデバイスにおいて分離され、各核酸が、PCRによって個々に増幅される。核酸は、1ウェルあたり1つより多い核酸が存在しないように分離され得る。いくつかの実施形態において、異なるプローブが、様々な対立遺伝子（例えば、胎児の対立遺伝子および母体の対立遺伝子）を区別するために使用され得る。対立遺伝子は、コピー数を測定するために数え上げられ得る。

【0130】

ある特定の実施形態において、ハイブリダイゼーションによる配列決定が使用され得る。その方法は、複数のポリヌクレオチド配列を複数のポリヌクレオチドプローブと接触させる工程を含み、ここで、その複数のポリヌクレオチドプローブの各々は、必要に応じて基材に繋ぎ止められ得る。その基材は、いくつかの実施形態において、既知のヌクレオチド配列のアレイを有する平らな表面であり得る。そのアレイへのハイブリダイゼーションのパターンは、そのサンプル中に存在するポリヌクレオチド配列を決定するために使用され得る。いくつかの実施形態において、各プローブは、ビーズ、例えば、磁気ビーズなどに繋ぎ止められる。それらのビーズへのハイブリダイゼーションは識別され得、そのサンプル内の複数のポリヌクレオチド配列を識別するために使用され得る。

【0131】

いくつかの実施形態において、ナノポア配列決定が、本明細書中に記載される方法において使用され得る。ナノポア配列決定は、単一核酸分子（例えば、DNA）が、ナノポアを通過するときに直接配列決定される単一分子配列決定技術である。

【0132】

本明細書中に記載される方法を行うための好適なMPSの方法、システムまたは技術プラットフォームが、核酸配列リードを得るために使用され得る。MPSプラットフォームの非限定的な例としては、Illumina/Solex/HiSeq（例えば、Illumina's Genome Analyzer; Genome Analyzer II; HiSeq 2000; HiSeq）、SOLID、Roche/454、PACBIOおよび／またはSMRT、Helicos True Single Molecule Sequencing、Ion TorrentおよびIon半導体ベースの配列決定（例えば、Life Technologiesが開発したもの）、WildFire、5500、5500xl Wおよび／もしくは5500xl W Genetic

Analyzerに基づく技術（例えば、Life Technologiesが開発し、販売しているもの、米国特許公開番号US20130012399）；ポロニーシーケンシング、パイロシーケンシング、大規模並列処理シグネチャ配列決定（MPSS）、RNAポリメラーゼ（RNAP）配列決定、LaserGenシステムおよび方法、ナノポアベースのプラットフォーム、化学感応性電界効果トランジスタ（CHEMFET）アレイ、電子顕微鏡法ベースの配列決定（例えば、ZS Genetics, Halcyon Molecularが開発したもの）、ナノボールシーケンシングなどまたはそれらの組み合わせが挙げられる。

【0133】

いくつかの実施形態において、染色体特異的配列決定が行われる。いくつかの実施形態において、染色体特異的配列決定は、DANSR（選択された領域のデジタル解析）を用いて行われる。選択された領域のデジタル解析は、PCR鋳型を形成する介在性の「ブリッジ」オリゴヌクレオチドを介した2つの遺伝子座特異的オリゴヌクレオチドのcfDNA依存のカテネーションによる数百の遺伝子座の同時の定量を可能にする。いくつかの実施形態において、染色体特異的配列決定は、染色体特異的配列が濃縮されたライブラリーを作製することによって行われる。いくつかの実施形態において、配列リードは、選択された染色体セットに対してのみ得られる。いくつかの実施形態において、配列リードは、21番、18番および13番染色体に対してのみ得られる。いくつかの実施形態において、配列リードは、参照ゲノム全体もしくはゲノムのセグメントに対して得られるおよび／またはおよび参照ゲノム全体もしくはゲノムのセグメントにマッピングされる。

【0134】

いくつかの実施形態において、配列リードは、配列モジュールによって、作製される、得られる、集められる、アセンブルされる、操作される、変換される、処理される、および／または提供される。配列モジュールを備える機器は、当該分野で公知の配列決定技術を使用して核酸の配列を決定する好適な機器および／または装置であり得る。いくつかの実施形態において、配列モジュールは、アラインメントし得る、アセンブルし得る、断片化し得る、相補鎖生成（complement）し得る、逆相補鎖生成（reverse complement）し得る、および／またはエラーチェックし得る（例えば、配列リードをエラーチェックし得る）。

【0135】

いくつかの実施形態において、サンプルから得られたヌクレオチド配列リードは、部分的なヌクレオチド配列リードである。本明細書中で使用されるとき、「部分的なヌクレオチド配列リード」とは、配列多義性（sequence ambiguity）とも称される不完全な配列情報を有する任意の長さの配列リードのことを指す。部分的なヌクレオチド配列リードは、核酸塩基の同一性および／または核酸塩基の位置もしくは順序に関する情報を欠き得る。部分的なヌクレオチド配列リードは、一般に、単に不完全な配列情報

(またはそれらの塩基のすべてより少ない塩基が配列決定されているかまたは決定されている)が、不注意のまたは故意でない配列決定のエラーに由来する配列リードを含まない。そのような配列決定のエラーは、ある特定の配列決定プロセスに固有であり得、例えば、核酸塩基の同一性に対する不正確なコール、および欠損したまたは余分な核酸塩基を含む。したがって、本明細書中の部分的なヌクレオチド配列リードに対して、その配列に関するある特定の情報は、故意に除外されることが多い。すなわち、すべての核酸塩基より少ない核酸塩基に関する配列情報、または配列決定のエラーとして別途特徴づけられ得るかもしくは配列決定のエラーであり得る配列情報が、故意に得られる。いくつかの実施形態において、部分的なヌクレオチド配列リードは、核酸フラグメントの一部分に及び得る。いくつかの実施形態において、部分的なヌクレオチド配列リードは、核酸フラグメントの長さ全体に及び得る。部分的なヌクレオチド配列リードは、例えば、国際特許出願公開番号WO 2013/052907に記載されており、本文、表、式および図面のすべてを含むこの全内容が、参照により本明細書中に組み込まれる。

10

リードのマッピング

【0136】

配列リードは、マッピングされ得、特定の核酸領域(例えば、染色体、その一部またはセグメント)にマップするリードの数は、カウントと称される。任意の好適なマッピング方法(例えば、プロセス、アルゴリズム、プログラム、ソフトウェア、モジュールなどまたはそれらの組み合わせ)が使用され得る。いくつかの実施形態において、配列リードは、マッピングされない。マッピングプロセスのある特定の態様が、本明細書の以後に記載される。

20

【0137】

ヌクレオチド配列リード(すなわち、物理的なゲノム位置が不明なフラグメントからの配列情報)のマッピングは、いくつかの方法で行われ得、得られた配列リードを参照ゲノム内のマッチする配列とアラインメントすることを含むことが多い。そのようなアラインメントでは、配列リードは、通常、参照配列にアラインメントされ、アラインメントする配列リードは、「マッピングされる」、「マッピングされた配列リード」または「マッピングされたリード」と呼ばれる。ある特定の実施形態において、マッピングされた配列リードは、「ヒット」または「カウント」と称される。いくつかの実施形態において、マッピングされた配列リードは、様々なパラメータに従って共にグループ化され、下記でさらに詳細に論じられる特定の部分に割り当てられる。

30

【0138】

本明細書中で使用されるとき、用語「アラインメントされる」、「アラインメント」または「アラインメントする」とは、マッチ(例えば、100%同一性)または部分的なマッチとして特定され得る2つまたはそれを超える核酸配列のことを指す。アラインメントは、手作業でまたはコンピュータ(例えば、ソフトウェア、プログラム、モジュールまたはアルゴリズム)によって行われ得、その非限定的な例としては、Illumina Genomics Analysisパイプラインの一部として配布されているEfficient Local Alignment of Nucleotide Data (ELAND)コンピュータプログラムが挙げられる。配列リードのアラインメントは、100%配列マッチであり得る。場合によっては、アラインメントは、100%未満の配列マッチである(すなわち、不完全なマッチ、部分的なマッチ、部分的なアラインメント)。いくつかの実施形態において、アラインメントは、約99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、89%、88%、87%、86%、85%、84%、83%、82%、81%、80%、79%、78%、77%、76%または75%マッチである。いくつかの実施形態において、アラインメントは、ミスマッチを含む。いくつかの実施形態において、アラインメントは、1、2、3、4または5つのミスマッチを含む。2つまたはそれを超える配列が、いずれかの鎖を用いてアラインメントされ得る。ある特定の実施形態において、核酸配列は、別の核酸配列の逆相補鎖とアラインメントされる。いくつかの実施形態において、配列リードは、参照配列または参照

40

50

ゲノムにアラインメントされる。いくつかの実施形態において、配列リードは、参照配列または参照ゲノムにアラインメントされない。

【0139】

コンピュータによる様々な方法が、各配列リードをある部分にマッピングするために使用され得る。配列をアラインメントするために使用され得るコンピュータアルゴリズムの非限定的な例としては、BLAST、BLITZ、FASTA、BOWTIE 1、BOWTIE 2、ELAND、MAQ、PROBEMATCH、SOAPもしくはSEQMAPまたはそれらの変法またはそれらの組み合わせが挙げられるが、これらに限定されない。いくつかの実施形態において、配列リードは、参照ゲノム内の配列とアラインメントされ得る。いくつかの実施形態において、配列リードは、当該分野で公知の核酸データベース内に見出され得る配列および／または当該分野で公知の核酸データベース内の配列とアラインメントされ得、その核酸データベースとしては、例えば、GenBank、dbEST、dbSTS、EMBL (European Molecular Biology Laboratory) およびDDBJ (DNA Databank of Japan) が挙げられる。BLASTまたは同様のツールが、特定された配列を配列データベースに対して検索するために使用され得る。次いで、検索のヒットが、例えば、特定された配列を適切な部分 (本明細書の以後に記載される) に選別するために使用され得る。

【0140】

いくつかの実施形態において、マッピングされた配列リードおよび／またはマッピングされた配列リードに関連する情報は、好適なコンピュータ可読形式で、非一時的なコンピュータ可読記憶媒体上に格納されるおよび／または非一時的なコンピュータ可読記憶媒体からアクセスされる。「コンピュータ可読形式」は、本明細書中で広く形式と称されるときがある。いくつかの実施形態において、マッピングされた配列リードは、好適なバイナリー形式、テキスト形式などまたはそれらの組み合わせで格納されるおよび／またはアクセスされる。バイナリー形式は、BAM形式であるときがある。テキスト形式は、配列アラインメント／マップ (SAM) 形式であるときがある。バイナリー形式および／またはテキスト形式の非限定的な例としては、BAM、SAM、SRF、FASTQ、Gzipなどまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、マッピングされた配列リードは、従来の形式 (例えば、SAM形式またはBAM形式) より少ない格納スペース (例えば、より少ないバイト) しか必要としない形式で格納され、および／またはそのような形式に変換される。いくつかの実施形態において、第1の形式のマッピングされた配列リードは、第1の形式よりも小さい格納スペースしか必要としない第2の形式に圧縮される。用語「圧縮される」は、本明細書中で使用されるとき、コンピュータ可読のデータファイルのサイズを小さくする、データ圧縮、情報源符号化および／またはビットレート削減のプロセスのことを指す。いくつかの実施形態において、マッピングされた配列リードは、SAM形式からバイナリー形式に圧縮される。いくつかのデータは、ファイルが圧縮された後に失われるときがある。圧縮プロセスにおいてデータが失われないときがある。いくつかのファイル圧縮の実施形態では、いくつかのデータが、インデックスおよび／または参照で置き換えられて、マッピングされた配列リードに関する情報を含む別のデータファイルにされる。いくつかの実施形態において、マッピングされた配列リードは、リードカウント、染色体識別子 (例えば、リードがマッピングされた染色体を識別する) および染色体位置識別子 (例えば、リードがマッピングされた染色体上の位置を識別する) を含むかまたはそれらからなるバイナリー形式で格納される。いくつかの実施形態において、バイナリー形式は、20バイトアレイ、16バイトアレイ、8バイトアレイ、4バイトアレイまたは2バイトアレイを含む。いくつかの実施形態において、マッピングされたリード情報は、10バイト形式、9バイト形式、8バイト形式、7バイト形式、6バイト形式、5バイト形式、4バイト形式、3バイト形式または2バイト形式でアレイとして格納される。マッピングされたリードデータは、5バイト形式を含む4バイトアレイで格納されるときがある。いくつかの実施形態において、バイナリー形式は、1バイトの染色体順序および4バイトの染色体位置を含む5バイト形式を含む。いくつかの実施

10

20

30

40

50

形態において、マッピングされたリードは、配列アラインメント／マップ（S A M）形式よりも約 100 倍、約 90 倍、約 80 倍、約 70 倍、約 60 倍、約 55 倍、約 50 倍、約 45 倍、約 40 倍または約 30 倍小さい圧縮されたバイナリー形式で格納される。いくつかの実施形態において、マッピングされたリードは、G Z i p 形式よりも約 2 倍～約 50 倍小さい（例えば、約 30、25、20、19、18、17、16、15、14、13、12、11、10、9、8、7、6 または約 5 倍小さい）圧縮された（c o m p r e s s）バイナリー形式で格納される。

【0141】

いくつかの実施形態において、あるシステムは、圧縮モジュールを備える。いくつかの実施形態において、コンピュータ可読形式で非一時的なコンピュータ可読記憶媒体上に格納されたマッピングされた配列リード情報は、圧縮モジュールによって圧縮される。圧縮モジュールは、マッピングされた配列リードを、好適な形式に変換し、および好適な形式から変換するときがある。圧縮モジュールは、いくつかの実施形態において、第 1 の形式のマッピングされた配列リードを受け入れ、それらを圧縮された形式（例えば、バイナリー形式）に変換し、その圧縮されたリードを別のモジュール（例えば、バイアス密度モジュール）に移し得る。圧縮モジュールは、配列リードをバイナリー形式（例えば、B R e a d s 形式）で提供することが多い。圧縮モジュールの非限定的な例としては、G Z I P、B G Z F および B A M などまたはそれらの改変物が挙げられる。

以下は、j a v a を使用して整数を 4 バイトアレイに変換する例を提供している：

【数 2】

```
public static final byte[]
convertToByteArray(int value)
{
    return new byte[] {
        (byte)(value >>> 24),
        (byte)(value >>> 16),
        (byte)(value >>> 8),
        (byte)value;
    }
}
```

【0142】

いくつかの実施形態において、リードは、参照ゲノム内の部分にユニークにまたは非ユニークにマップし得る。あるリードが、参照ゲノム内の単一配列とアラインメントする場合、そのリードは、「ユニークにマッピングされる」と見なされる。あるリードが、参照ゲノム内の 2 つまたはそれを超える配列とアラインメントする場合、そのリードは、「非ユニークにマッピングされる」と見なされる。いくつかの実施形態において、非ユニークにマッピングされたリードは、さらなる解析（例えば、定量）から除外される。ある特定の実施形態では、参照ゲノムと、マッピングされている個々のサンプル由来のリードとの間に存在し得るある特定のわずかなミスマッチ（0～1 個）は、単一ヌクレオチド多型を説明するために許容され得る。いくつかの実施形態において、少しの程度のミスマッチも、参照配列にマッピングされるリードに対して許容されない。

【0143】

本明細書中で使用されるとき、「参照ゲノム」とは、部分的であるかまたは完全であるかを問わず、被験体由来の特定された配列の参照のために使用され得る任意の生物またはウイルスの任意の特定の既知の配列決定されたまたは特徴づけられたゲノムのことを指し得る。例えば、ヒト被験体ならびに他の多くの生物のために使用される参照ゲノムは、W o r l d W i d e W e b U R L n c b i . n l m . n i h . g o v における N a t i o n a l C e n t e r f o r B i o t e c h n o l o g y I n f o r m

ationに見られ得る。「ゲノム」とは、核酸配列として表現される、生物またはウイルスの完全な遺伝情報のことを指す。本明細書中で使用されるとき、参照配列または参照ゲノムは、個々のまたは複数の個体からのアセンブルされたゲノム配列または部分的にアセンブルされたゲノム配列であることが多い。いくつかの実施形態において、参照ゲノムは、1人またはそれを超えるヒト個体からのアセンブルされたまたは部分的にアセンブルされたゲノム配列である。いくつかの実施形態において、参照ゲノムは、染色体に割り当てられた配列を含む。

【0144】

ある特定の実施形態において、サンプル核酸が妊婦由来である場合、参照配列は、胎児、その胎児の母またはその胎児の父由来でないときがあり、それは、本明細書中で「外部参照」と称される。いくつかの実施形態において、母体の参照が、調製され得、使用され得る。外部参照に基づいて、妊婦由来の参照が調製されるとき（「母体参照配列」）、胎児DNAを実質的に含まないその妊婦のDNA由来のリードが、外部参照配列にマッピングされ、アセンブルされることが多い。ある特定の実施形態において、外部参照は、妊婦と実質的に同じ民族性を有する個体のDNA由来である。母体参照配列は、母体のゲノムDNAを完全にカバーしない可能性があり（例えば、母体参照配列は、母体のゲノムDNAの約50%、60%、70%、80%、90%またはそれを超えてカバーし得る）、母体参照は、母体のゲノムDNA配列と完全にマッチしない可能性がある（例えば、母体参照配列は、複数のミスマッチを含み得る）。

【0145】

ある特定の実施形態において、マッピング精度 (m a p p a b i l i t y) は、ゲノム領域（例えば、部分、ゲノム部分、部分）に対して評価される。マッピング精度は、ヌクレオチド配列リードを参照ゲノムの一部と、通常、指定の数のミスマッチ（例えば、0、1つ、2つもしくはそれを超えるミスマッチを含む）まで明確にアラインメントする能力である。所与のゲノム領域の場合、予想されるマッピング精度は、前もってセットされたリード長のスライディングウィンドウ (s l i d d i n g - w i n d o w) アプローチを用い、得られたリードレベルのマッピング精度値を平均して、推定され得る。連続したユニークなヌクレオチド配列を含むゲノム領域は、高いマッピング精度値を有するときがある。

部分

【0146】

いくつかの実施形態において、マッピングされた配列リード（すなわち、配列タグ）は、様々なパラメータに従って共にグループ化され、特定の部分（例えば、参照ゲノムの部分）に割り当てられる。マッピングされた個々の配列リードは、サンプル中に存在する部分（例えば、部分の存在もしくは非存在または量）を特定するために使用され得ることが多い。いくつかの実施形態において、部分の量は、そのサンプル中のより大きい配列（例えば、染色体）の量を指し示している。用語「部分」は、本明細書中で「ゲノム区分 (g e n o m i c s e c t i o n)」、「ビン (b i n)」、「領域」、「区画 (p a r t i t i o n)」、「参照ゲノムの部分」、「染色体の部分」または「ゲノム部分」とも称され得る。いくつかの実施形態において、部分は、染色体全体、染色体のセグメント、参照ゲノムのセグメント、複数の染色体に及ぶセグメント、複数の染色体セグメントおよび/またはそれらの組み合わせである。いくつかの実施形態において、部分は、特定のパラメータに基づいて予め定義される。いくつかの実施形態において、部分は、ゲノムの分割（例えば、サイズ、GC含有量、シーケンシングカバレッジのばらつき、連続した領域、任意に定義されたサイズの連続した領域などによって分割される）に基づいて任意に定義される。

【0147】

いくつかの実施形態において、部分は、例えば、配列の長さまたは特定の特徴を含む1つまたはそれを超えるパラメータに基づいて明示される。部分は、当該分野で公知の (k n o w) または本明細書中に記載される任意の好適な基準を用いて、考慮すべきものから

選択され得、フィルタリングされ得、および／または除去され得る。いくつかの実施形態において、部分は、ゲノム配列の特定の長さに基づく。いくつかの実施形態において、方法は、複数の部分に対する複数のマッピングされた配列リードの解析を含み得る。部分は、ほぼ同じ長さであり得るか、または異なる長さであり得る。いくつかの実施形態において、部分は、ほぼ等しい長さである。いくつかの実施形態において、異なる長さの部分は、調整されるかまたは重み付けされる。いくつかの実施形態において、部分は、約10キロベース(kb)～約20kb、約10kb～約100kb、約20kb～約80kb、約30kb～約70kb、約40kb～約60kbである。いくつかの実施形態において、部分は、約10kb、20kb、30kb、40kb、50kbまたは約60kb長である。部分は、連続した配列の続きに限定されない。したがって、部分は、連続したおよび／または連続していない配列から構成され得る。部分は、単一の染色体に限定されない。いくつかの実施形態において、部分は、1本の染色体の全部もしくは一部または2本またはそれを超える染色体の全部もしくは一部を含む。いくつかの実施形態において、部分は、1本、2本もしくはそれを超える染色体の全体に及び得る。さらに、部分は、複数の染色体の継ぎ合わせたまたはばらばらの領域に及び得る。

10

【0148】

いくつかの実施形態において、部分は、目的の染色体、例えば、コピー数の変動が評価される染色体（例えば、13番、18番および／もしくは21番染色体または性染色体の異数性）における特定の染色体セグメントであり得る。部分は、病原体のゲノム（例えば、細菌、真菌またはウイルスのゲノム）またはそのフラグメントでもあり得る。部分は、

20

【0149】

いくつかの実施形態において、ゲノム（例えば、ヒトゲノム）は、特定の領域の情報量に基づいて部分に分割される。いくつかの実施形態において、ゲノムを分割することにより、そのゲノムにわたる類似の領域（例えば、同一または相同の領域または配列）が排除され得、ユニークな領域だけが維持され得る。分割において除去される領域は、単一の染色体内に存在し得るか、または複数の染色体に及び得る。いくつかの実施形態において、分割されたゲノムは、より速いアラインメントのために切り詰められ、最適化されることにより、ユニークに識別可能な配列に焦点を合わせることが可能になることが多い。

【0150】

いくつかの実施形態において、分割は、類似の領域の重み付けを小さくし得る。部分の重み付けを小さくするためのプロセスは、下記でさらに詳細に論じられる。

30

【0151】

いくつかの実施形態において、染色体を越える領域にゲノムを分割することは、分類の文脈においてもたらされる情報獲得に基づき得る。例えば、情報量は、確認された正常な被験体および異常な被験体（例えば、それぞれ正倍数性の被験体およびトリソミーの被験体）の群間を区別するために特定のゲノム位置の有意性を計測するp値プロファイルを用いて定量化され得る。いくつかの実施形態において、染色体を越える領域にゲノムを分割することは、他の任意の基準、例えば、タグをアラインメントしている間の速度／便利さ、GC含有量（例えば、高いまたは低いGC含有量）、GC含有量の均一性、配列含有量の他の尺度（例えば、個々のヌクレオチドの割合、ピリミジンまたはプリンの割合、天然の核酸と非天然の核酸との割合、メチル化されたヌクレオチドの割合およびCpG含有量）、メチル化の状態、二重鎖の融解温度、配列決定またはPCRに対する適用可能性（amenability）、参照ゲノムの個々の部分に割り当てられる不確定値、および／または特定の特徴に対する標的化された検索に基づき得る。

40

【0152】

染色体の「セグメント」は、通常、染色体の一部であり、代表的には、部分とは異なる染色体の一部である。染色体のセグメントは、部分とは異なる染色体の領域に存在するときにあり、部分とポリヌクレオチドを共有しないときにあり、部分に存在するポリヌクレオチドを含むときがある。染色体のセグメントは、部分より多い数のヌクレオチドを含む

50

ことが多く（例えば、セグメントは、部分を含むときがある）、染色体のセグメントは、部分より少ない数のヌクレオチドを含むときがある（例えば、セグメントは、部分内に存在するときがある）。

部分のフィルタリングおよび／または選択

【0153】

部分は、本明細書中に記載されるまたは当該分野で公知の1つまたはそれを超える特徴、パラメータ、基準および／または方法に従って、処理される（例えば、正規化される、フィルタリングされる、選択されるなどまたはそれらの組み合わせ）ときがある。部分は、任意の好適な方法によって、および任意の好適なパラメータに従って、処理され得る。部分をフィルタリングするためおよび／または選択するために使用され得る特徴および／またはパラメータの非限定的な例としては、カウント、カバレッジ、マッピング精度、ばらつき、不確定性のレベル、グアニン-シトシン（GC）含有量、CCFフラグメントの長さおよび／またはリード長（例えば、フラグメントの長さの比率（FLR）、胎児の比率の統計量（FRS））、DNase I 感度、メチル化の状態、アセチル化、ヒストン分布、クロマチン構造、反復パーセントなどまたはそれらの組み合わせが挙げられる。部分は、本明細書中に列挙されるまたは記載される特徴またはパラメータと相関する任意の好適な特徴またはパラメータに従ってフィルタリングされ得、および／または選択され得る。部分は、部分に特異的な特徴もしくはパラメータ（例えば、複数のサンプルに係る単一の部分に対して測定されるとき）および／またはサンプルに特異的な特徴もしくはパラメータ（例えば、サンプル内の複数の部分に対して測定されるとき）に従ってフィルタリングされ得、および／または選択され得る。いくつかの実施形態において、部分は、比較的低いマッピング精度、比較的大きなばらつき、高レベルの不確定性、比較的長いCCFフラグメント長（例えば、低FRS、低FLR）、比較的高い割合の反復配列、高GC含有量、低GC含有量、低カウント、ゼロカウント、高カウントなどまたはそれらの組み合わせに従ってフィルタリングされ、および／または除去される。いくつかの実施形態において、部分（例えば、部分のサブセット）は、好適なマッピング精度のレベル、ばらつき、不確定性のレベル、反復配列の割合、カウント、GC含有量などまたはそれらの組み合わせに従って選択される。いくつかの実施形態において、部分（例えば、部分のサブセット）は、比較的短いCCFフラグメント長（例えば、高FRS、高FLR）に従って選択される。部分にマッピングされたカウントおよび／またはリードは、部分（例えば、部分のサブセット）をフィルタリングするかまたは選択する前および／または後に処理される（例えば、正規化される）ときがある。いくつかの実施形態において、部分にマッピングされたカウントおよび／またはリードは、部分（例えば、部分のサブセット）をフィルタリングするかまたは選択する前および／または後に処理されない。

【0154】

任意の好適な数のサンプルに由来する配列リードは、本明細書中に記載される1つまたはそれを超える基準、パラメータおよび／または特徴を満たす部分のサブセットを特定するために使用され得る。複数の妊婦由来のサンプル群からの配列リードが、使用されるときがある。複数の妊婦の各々に由来する1つまたはそれを超えるサンプル（例えば、各妊婦由来の1～約20個のサンプル（例えば、約2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18または19個のサンプル））が、対処され得、好適な数の妊婦（例えば、約2～約10,000人の妊婦（例えば、約10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000人の妊婦））が、対処され得る。いくつかの実施形態において、同じ妊婦由来の同じテストサンプルからの配列リードが、参照ゲノムにおける部分にマッピングされ、部分のサブセットを生成するために使用される。

【0155】

妊婦から得られる循環無細胞核酸フラグメント（CCFフラグメント）は、一般に、胎

児の細胞を起源とする核酸フラグメント（すなわち、胎児のフラグメント）および母体の細胞を起源とする核酸フラグメント（すなわち、母体のフラグメント）を含むことが確認されている。胎児を起源とする C C F フラグメントに由来する配列リードは、本明細書中で「胎児のリード」と称される。胎児を有する妊婦（例えば、母）のゲノムを起源とする C C F フラグメントに由来する配列リードは、本明細書中で「母体のリード」と称される。胎児のリードが得られる C C F フラグメントは、本明細書中で胎児の鋳型と称され、母体のリードが得られる C C F フラグメントは、本明細書中で母体の鋳型と称される。

【0156】

C C F フラグメントにおいて、胎児のフラグメントは、一般に比較的短い（例えば、約 200 塩基対長またはそれ未満）こと、および母体のフラグメントは、そのような比較的短いフラグメントおよび比較的より長いフラグメントを含むことも確認されている。比較的短いフラグメントからのかなりの量のリードがマッピングされた部分のサブセットが、選択され得、および／または特定され得る。理論に限定されるものではないが、そのような部分にマッピングされたリードは、胎児のリードについて濃縮され、それにより、胎児の遺伝解析（例えば、胎児のコピー数変異（例えば、胎児の染色体異数性（例えば、T 2 1、T 1 8 および／または T 1 3））の存在もしくは非存在の検出）の精度が改善され得ると予想される。

【0157】

しかしながら、胎児の遺伝解析が、リードのサブセットに基づくとき、かなりの数のリードが考慮されないことが多い。胎児の遺伝解析に向けて、選択された部分のサブセットにマッピングされたリードのサブセットを選択することおよび選択されていない部分におけるリードを除去することにより、例えば、分散が大きくなることに起因して、遺伝解析の精度は低下し得る。いくつかの実施形態において、被験体またはサンプルのマップから得られたシーケンシングリードの約 30 %～約 70 %（例えば、約 35 %、40 %、45 %、50 %、55 %、60 % または 65 %）が、胎児の遺伝解析のために部分のサブセットを選択する際に、考慮すべきものから除去される。ある特定の実施形態において、被験体またはサンプルから得られたシーケンシングリードの約 30 %～約 70 %（例えば、約 35 %、40 %、45 %、50 %、55 %、60 % または 65 %）が、胎児の遺伝解析のために使用される部分のサブセットにマップする。

【0158】

部分は、任意の好適な方法によって選択され得、および／またはフィルタリングされ得る。いくつかの実施形態において、部分は、データ、グラフ、プロットおよび／またはチャートの目視検査に従って選択される。ある特定の実施形態において、部分は、1 つまたはそれを超えるマイクロプロセッサおよびメモリーを備えるシステムまたは機器によって選択され、および／またはフィルタリングされる（例えば、部分的に）。いくつかの実施形態において、部分は、実行可能なプログラムが格納されている非一時的なコンピュータ可読記憶媒体によって選択され、および／またはフィルタリングされ（例えば、部分的に）、ここで、そのプログラムは、マイクロプロセッサにその選択および／またはフィルタリングを行うように指示する。

【0159】

本明細書中に記載される方法によって選択された部分のサブセットは、胎児の遺伝解析のために種々の様式で使用され得る。ある特定の実施形態において、サンプルに由来するリードは、本明細書中に記載される予め選択された部分のサブセットを用いて、および参照ゲノム内のそれらの部分のすべてまたはほとんどを用いずに、マッピングプロセスにおいて使用される。予め選択された部分のサブセットにマップするそれらのリードは、胎児の遺伝解析のさらなる工程において使用されることが多く、予め選択された部分のサブセットにマップしないリードは、胎児の遺伝解析のさらなる工程において使用されないことが多い（例えば、マップしないリードは、除去されるかまたはフィルタリングされる）。

【0160】

いくつかの実施形態において、サンプルに由来する配列リードは、参照ゲノムのすべて

またはほとんどの部分にマッピングされ、その後、本明細書中に記載される予め選択された部分のサブセットが選択される。選択された部分のサブセットからのリードは、胎児の遺伝解析のさらなる工程において使用されることが多い。後者の実施形態において、選択されない部分からのリードは、胎児の遺伝解析のさらなる工程において使用されないことが多い（例えば、選択されない部分におけるリードは、除去されるかまたはフィルタリングされる）。

カウント

【0161】

選択された特徴または変数に基づいてマッピングされたまたは分割された配列リードは、いくつかの実施形態において、1つまたはそれを超える部分（例えば、参照ゲノムの部分）にマッピングされたリードの数を測定するために定量化され得る。ある特定の実施形態において、ある部分にマッピングされた配列リードの量は、カウント（count）（例えば、カウント（acount））と呼ばれる。カウントは、部分に関連することが多い。ある特定の実施形態において、2つまたはそれを超える部分（例えば、部分のセット）に対するカウントは、数学的に操作される（例えば、平均される、加算される、正規化されるなどまたはそれらの組み合わせ）。いくつかの実施形態において、カウントは、部分にマッピングされた（すなわち、部分に関連する）配列リードのいくつかまたはすべてから測定される。ある特定の実施形態において、カウントは、マッピングされた配列リードの予め定義されたサブセットから測定される。マッピングされた配列リードの予め定義されたサブセットは、任意の好適な特徴または変数を用いて定義され得るかまたは選択され得る。いくつかの実施形態において、マッピングされた配列リードの予め定義されたサブセットは、1～n個の配列リードを含み得、ここで、nは、試験被験体サンプルまたは参照被験体サンプルから生成されたすべての配列リードの合計に等しい数である。いくつかの実施形態において、カウントは、部分にマッピングされない配列リードの定量結果である。

【0162】

ある特定の実施形態において、カウントは、当該分野で公知の好適な方法、演算または数学的プロセスによって処理されたまたは操作された配列リードに由来する。カウント（Acount）（例えば、カウント（count））は、好適な方法、演算または数学的プロセスによって測定され得る。ある特定の実施形態において、カウントは、部分に関連する配列リードに由来し、ここで、それらの配列リードのいくつかまたはすべてが、重み付けされるか、除去されるか、フィルタリングされるか、正規化されるか、調整されるか、平均されるか、手段として得られるか、加算されるかもしくは減算されるか、またはそれらの組み合わせによって処理される。いくつかの実施形態において、カウントは、生の配列リードおよびまたはフィルタリングされた配列リードから得られる。ある特定の実施形態において、カウント値は、数学的プロセスによって測定される。ある特定の実施形態において、カウント値は、部分にマッピングされた配列リードの平均、平均値または合計である。カウントは、カウントの平均値の数値であることが多い。いくつかの実施形態において、カウントは、不確定値に関連する。

【0163】

いくつかの実施形態において、カウントは、操作され得るかまたは変換され得る（例えば、正規化される、まとめられる、加算される、フィルタリングされる、選択される、平均される、手段として得られるなどまたはそれらの組み合わせ）。いくつかの実施形態において、カウントは、正規化されたカウントをもたらし変換され得る。カウントは、当該分野で公知のおよび／または本明細書中に記載されるような方法（例えば、部分ごとの正規化、カウントの中央値（median count）（ビンカウントの中央値（median bin count））、部分カウントの中央値（median portion count））の正規化、GC含有量による正規化、線形および非線形最小二乗回帰、LOESS（例えば、GC LOESS）、LOWESS、PERUN、ChAI、主成分正規化、RM、GCRM、cQnおよび／またはそれらの組み合わせ）によって

処理され得る（例えば、正規化され得る）。ある特定の実施形態において、カウントは、L O E S S、カウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化および主成分正規化のうちの1つまたはそれを超えるものによって処理され得る（例えば、正規化され得る）。ある特定の実施形態において、カウントは、L O E S Sに続くカウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化によって処理され得る（例えば、正規化され得る）。ある特定の実施形態において、カウントは、L O E S Sに続くカウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化に続く主成分正規化によって処理され得る（例えば、正規化され得る）。

【0164】

カウント（例えば、生のカウント、フィルタリングされたカウントおよび／または正規化されたカウント）は、処理され得、1つまたはそれを超えるレベルに対して正規化され得る。レベルおよびプロファイルは、本明細書中の以後に、より詳細に記載される。ある特定の実施形態において、カウントは、処理され得、および／または参照レベルに対して正規化され得る。参照レベルは、本明細書中の後で述べられる。あるレベルに従って処理されたカウント（例えば、処理されたカウント）は、不確定値（例えば、算出された分散、誤差、標準偏差、Z得点、p値、平均絶対偏差など）に関連し得る。いくつかの実施形態において、不確定値は、あるレベルより上および下の範囲を定義する。偏差に対する値は、不確定値の代わりに使用され得、偏差の尺度の非限定的な例としては、標準偏差、平均絶対偏差、中央絶対偏差、標準得点（例えば、Z得点、正規得点、標準化変数）などが挙げられる。

【0165】

カウントは、胎児を有する妊婦由来の核酸サンプルから得られることが多い。1つまたはそれを超える部分にマッピングされた核酸配列リードのカウントは、胎児と胎児の母（例えば、妊婦被験体）の両方を代表するカウントであることが多い。ある特定の実施形態において、ある部分にマッピングされたカウントのいくつかは、胎児のゲノムに由来し、同じ部分マッピングされたカウントのいくつかは、母体のゲノムに由来する。

データ処理および正規化

【0166】

マッピングされてカウントされた配列リードおよび／またはマッピングされなかった配列リードは、本明細書中で生データと称される。なぜならそのデータは、操作されていないカウント（例えば、生のカウント）を表しているからである。いくつかの実施形態において、データセット内の配列リードのデータは、アウトカムの提供を容易にするために、さらに処理され得（例えば、数学的および／または統計的に操作され得）、および／または表示され得る。ある特定の実施形態において、より大きいデータセットを含むデータセットは、さらなる解析を容易にするために、前処理から恩恵を受け得る。データセットの前処理は、冗長なおよび／もしくは情報価値のない部分または参照ゲノムの部分（例えば、情報価値のないデータを有する参照ゲノムの部分、冗長なマッピングされたリード、カウントの中央値がゼロである部分、過剰提示または過小提示された配列）の除去を含むときがある。理論に限定されるものではないが、データ処理および／または前処理は、（i）ノイズの多いデータを除去し得、（ii）情報価値のないデータを除去し得、（iii）冗長なデータを除去し得、（iv）より大きいデータセットの複雑さを低下させ得、および／または（v）1つの形態から1つもしくはそれを超える他の形態へのデータの変換を容易にし得る。用語「前処理」および「処理」は、データまたはデータセットに関して使用されるとき、本明細書中で集合的に「処理」と称される。処理は、データをさらなる解析により適用できるようにし得、いくつかの実施形態ではアウトカムを生成し得る。いくつかの実施形態において、1つもしくはそれを超える処理方法またはすべての処理方法（例えば、正規化方法、部分のフィルタリング、マッピング、検証などまたはそれらの組み合わせ）が、メモリーと連動したプロセッサ、マイクロプロセッサ、コンピュータおよび／またはマイクロプロセッサによって制御される装置によって行われる。

【0167】

用語「ノイズの多いデータ」は、本明細書中で使用されるとき、（a）解析されるかまたはプロットされたとき、データポイント間に有意な分散を有するデータ、（b）有意な標準偏差（例えば、3を超える標準偏差）を有するデータ、（c）有意な平均値の標準誤差を有するデータなど、および前述のものの組み合わせのことを指す。ノイズの多いデータは、出発物質（例えば、核酸サンプル）の量および／または質に起因して生じるときがあり、配列リードを生成するために使用されるDNAを調製するためまたは複製するためのプロセスの一部として生じるときがある。ある特定の実施形態において、ノイズは、PCRに基づく方法を用いて調製されたときに過剰提示されるある特定の配列に起因する。本明細書中に記載される方法は、ノイズの多いデータの関与を減少させ得るかまたは排除し得、ゆえに、提供されるアウトカムに対するノイズの多いデータの影響を低減し得る。

10

【0168】

用語「情報価値のないデータ」、「情報価値のない参照ゲノムの部分」および「情報価値のない部分」は、本明細書中で使用されるとき、所定の閾値と有意に異なるかまたは所定の値のカットオフ範囲に入らない数値を有する部分またはそれに由来するデータのことを指す。本明細書中の用語「しきい値」および「閾値」とは、適格なデータセットを用いて算出され、遺伝的変異（例えば、コピー数変異、異数性、微小重複、微小欠失、染色体異常など）の診断の限度として役立つ任意の数字のことを指す。ある特定の実施形態において、しきい値は、本明細書中に記載される方法によって得られる結果によって上回られ、被験体は、コピー数変異（例えば、トリソミー21）と診断される。値の閾値または範囲は、いくつかの実施形態において、配列リードデータ（例えば、参照および／または被験体からの配列リードデータ）を数学的におよび／または統計的に操作することによって算出されることが多く、ある特定の実施形態において、値の閾値または範囲を生成するように操作される配列リードデータは、配列リードデータ（例えば、参照および／または被験体からの配列リードデータ）である。いくつかの実施形態において、不確定値が決定される。不確定値は、一般に、分散または誤差の尺度であり、分散または誤差の任意の好適な尺度であり得る。いくつかの実施形態において、不確定値は、標準偏差、標準誤差、算出された分散、p値または平均絶対偏差（MAD）である。いくつかの実施形態において、不確定値は、本明細書中に記載される式に従って算出され得る。

20

【0169】

任意の好適な手順が、本明細書中に記載されるデータセットを処理するために使用され得る。データセットを処理するために使用するのに適した手順の非限定的な例としては、フィルタリング、正規化、重み付け、ピーク高さのモニタリング、ピーク面積のモニタリング、ピーク端のモニタリング、面積比の測定、データの数学的処理、データの統計的処理、統計的アルゴリズムの適用、固定変数を用いた解析、最適化された変数を用いた解析、さらなる処理のためにパターンまたは傾向を特定するためのデータのプロットなどおよび前述のものの組み合わせが挙げられる。いくつかの実施形態において、データセットは、様々な特徴（例えば、GC含有量、冗長なマッピングされたリード、セントロメア領域、テロメア領域などおよびそれらの組み合わせ）および／または変数（例えば、胎児の性別、母体の年齢、母体の倍数性、胎児核酸の寄与パーセントなどまたはそれらの組み合わせ）に基づいて処理される。ある特定の実施形態において、本明細書中に記載されるようなデータセットの処理は、大きなおよび／または複雑なデータセットの複雑さおよび／または次元を低下させ得る。複雑なデータセットの非限定的な例としては、異なる年齢および民族的バックグラウンドの、1つまたはそれを超える試験被験体および複数の参照被験体から生成された配列リードデータが挙げられる。いくつかの実施形態において、データセットは、各試験被験体および／または各参照被験体に対する数千個から数百万個の配列リードを含み得る。

30

40

【0170】

データ処理は、ある特定の実施形態において、任意の数の工程で行われ得る。例えば、データは、いくつかの実施形態において、ただ1つの処理手順を用いて処理され得、ある特定の実施形態において、データは、1つもしくはそれを超える、5つもしくはそれを超

50

える、10個もしくはそれを超える、または20個もしくはそれを超える処理工程（例えば、1つもしくはそれを超える処理工程、2つもしくはそれを超える処理工程、3つもしくはそれを超える処理工程、4つもしくはそれを超える処理工程、5つもしくはそれを超える処理工程、6つもしくはそれを超える処理工程、7つもしくはそれを超える処理工程、8つもしくはそれを超える処理工程、9つもしくはそれを超える処理工程、10個もしくはそれを超える処理工程、11個もしくはそれを超える処理工程、12個もしくはそれを超える処理工程、13個もしくはそれを超える処理工程、14個もしくはそれを超える処理工程、15個もしくはそれを超える処理工程、16個もしくはそれを超える処理工程、17個もしくはそれを超える処理工程、18個もしくはそれを超える処理工程、19個もしくはそれを超える処理工程または20個もしくはそれを超える処理工程）を用いて処理され得る。いくつかの実施形態において、処理工程は、2回またはそれを超える回数繰り返される同じ工程（例えば、2回またはそれを超える回数のフィルタリング、2回またはそれを超える回数の正規化）であり得、ある特定の実施形態において、処理工程は、同時にまたは連続して行われる2つまたはそれを超える異なる処理工程（例えば、フィルタリング、正規化；正規化、ピーク高さおよびピーク端のモニタリング；フィルタリング、正規化、参照に対する正規化、p値を決定する統計的操作など）であり得る。いくつかの実施形態において、任意の好適な数および／または組み合わせの同じまたは異なる処理工程が、アウトカムの提供を容易にするために配列リードデータを処理するために使用され得る。ある特定の実施形態において、本明細書中に記載される基準によるデータセットの処理は、データセットの複雑さおよび／または次元を低下させ得る。

10

20

【0171】

いくつかの実施形態において、1つまたはそれを超える処理工程は、1つまたはそれを超えるフィルタリング工程を含み得る。用語「フィルタリング」は、本明細書中で使用されるとき、部分または参照ゲノムの部分を考慮すべきものから除去することを指す。参照ゲノムの部分は、任意の好適な基準に基づく除去のために選択され得、その基準としては、冗長なデータ（例えば、冗長なまたはオーバーラップするマッピングされたリード）、情報のないデータ（例えば、カウントの中央値がゼロである参照ゲノムの部分）、過剰提示されるかもしくは過小提示される配列を含む参照ゲノムの部分、ノイズの多いデータなどまたは前述のものの組み合わせが挙げられるが、これらに限定されない。フィルタリングプロセスは、参照ゲノムの1つまたはそれを超える部分を考慮すべきものから除去すること、および参照ゲノム、染色体または検討中のゲノムの部分に対するカウントされたまたは合計されたカウントから、除去のために選択された参照ゲノムの1つまたはそれを超える部分におけるカウントを減算することを含むことが多い。いくつかの実施形態において、参照ゲノムの部分は、連続的に（例えば、各個別の部分の除去の影響の評価を可能にするために1つずつ）除去され得、ある特定の実施形態では、除去のためにマークされた参照ゲノムのすべての部分が、同時に除去され得る。いくつかの実施形態において、ある特定のレベルより上または下の分散を特徴とする参照ゲノムの部分が除去され、それは、本明細書中で、参照ゲノムの「ノイズの多い」部分のフィルタリングと称されるときがある。ある特定の実施形態において、フィルタリングプロセスは、所定の複数のプロファイルの分散毎に、ある部分、染色体、または染色体のセグメントのプロファイルレベルの平均値から逸脱するデータポイントをデータセットから得ることを含み、ある特定の実施形態では、フィルタリングプロセスは、所定の複数のプロファイルの分散毎に、ある部分、染色体または染色体のセグメントのプロファイルレベルの平均値から逸脱しないデータポイントをデータセットから除去することを含む。いくつかの実施形態において、フィルタリングプロセスは、コピー数変異の存在もしくは非存在について解析される参照ゲノムの候補部分の数を減少させるために用いられる。コピー数変異（例えば、微小欠失、微小重複）の存在もしくは非存在について解析される参照ゲノムの候補部分の数の減少は、データセットの複雑さおよび／または次元を低下させることが多く、コピー数変異および／または遺伝的異常を検索するおよび／または識別する速度を2桁またはそれを超える桁数だけ高めるときがある。

30

40

50

【0172】

いくつかの実施形態において、1つまたはそれを超える処理工程は、1つまたはそれを超える正規化工程を含み得る。正規化は、本明細書中に記載されるまたは当該分野で公知の好適な方法によって行われ得る。ある特定の実施形態において、正規化は、異なるスケールで計測された値を概念的に共通のスケールに調整することを含む。ある特定の実施形態において、正規化は、調整された値の確率分布をアラインメントの状態にするための高度な数学的調整を含む。いくつかの実施形態において、正規化は、分布を正規分布に適合させることを含む。ある特定の実施形態において、正規化は、ある特定の全体的な影響（例えば、誤差および例外）の作用を排除するように、異なるデータセットに対する対応する正規化された値の比較を可能にする数学的調整を含む。ある特定の実施形態において、正規化は、スケールリングを含む。正規化は、所定の変数または式による1つまたはそれを超えるデータセットの除算を含むときがある。正規化は、所定の変数または式による1つまたはそれを超えるデータセットの減算を含むときがある。正規化方法の非限定的な例としては、部分ごとの正規化、GC含有量による正規化、カウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化、線形および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS（局所的に重み付けされた散布図平滑化）、PERUN、ChAI、主成分正規化、リピートマスク（RM）、GC-正規化およびリピートマスク（GCRM）、cQnならびに／またはそれらの組み合わせが挙げられる。いくつかの実施形態において、コピー数変異（例えば、異数性、微小重複、微小欠失）の存在もしくは非存在の判定は、正規化方法（例えば、部分ごとの正規化、GC含有量による正規化、カウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化、線形および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS（局所的に重み付けされた散布図平滑化）、PERUN、ChAI、主成分正規化、リピートマスク（RM）、GC-正規化およびリピートマスク（GCRM）、cQn、当該分野で公知の正規化方法、ならびに／またはそれらの組み合わせ）を用いる。いくつかの実施形態において、コピー数変異（例えば、異数性、微小重複、微小欠失）の存在もしくは非存在の判定は、LOESS、カウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化および主成分正規化のうちの1つまたはそれを超えるものを用いる。いくつかの実施形態において、コピー数変異の存在もしくは非存在の判定は、LOESSに続いて、カウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化を用いる。いくつかの実施形態において、コピー数変異の存在もしくは非存在の判定は、LOESSに続いて、カウントの中央値（ビンカウントの中央値、部分カウントの中央値）の正規化に続いて、主成分正規化を用いる。ある特定の正規化プロセス（例えば、ChAI正規化、主成分正規化、PERUN正規化）の態様は、例えば、2014年5月23日に出願され、2014年11月27日にWO2014/190286として公開された特許出願番号PCT/US2014/039389；および2014年10月2日に出願され、2015年4月9日にWO2015/051163として公開された特許出願番号PCT/US2014/058885に記載されている。

【0173】

任意の好適な数の正規化を使用することができる。いくつかの実施形態において、データセットは、1回もしくはそれを超える回数、5回もしくはそれを超える回数、10回もしくはそれを超える回数またはなおも20回もしくはそれを超える回数だけ、正規化され得る。データセットは、任意の好適な特徴または変数（例えば、サンプルデータ、参照データまたはその両方）を代表する値（例えば、正規化値）に対して正規化され得る。使用され得るデータ正規化のタイプの非限定的な例としては、1つまたはそれを超える選択された試験部分または参照部分に対する生のカウントデータを、その選択された部分または区分がマッピングされた染色体またはゲノム全体にマッピングされたカウントの総数に対して正規化すること；1つまたはそれを超える選択された部分に対する生のカウントデータを、1つもしくはそれを超える部分または選択された部分もしくはセグメントがマッピングされた染色体に対する参照カウントの中央値に対して正規化すること；生のカウント

データを、事前に正規化されたデータまたはその微分係数に対して正規化すること；および事前に正規化されたデータを、1つまたはそれを超える他の所定の正規化変数に対して正規化することが挙げられる。データセットの正規化は、所定の正規化変数として選択された特徴または特性に応じて、統計的誤差を切り離す効果を有するときがある。データセットの正規化は、データを通常のスケール（例えば、所定の正規化変数）にすることによって、異なるスケールを有するデータのデータ特性の比較も可能にするときがある。いくつかの実施形態において、統計的に導出された値に対する1つまたはそれを超える正規化は、データの差異を最小にするためおよび範囲外のデータの重要性を低下させるために使用され得る。部分または参照ゲノムの部分を正規化値に対して正規化することは、「部分ごとの正規化」と称されるときがある。

10

【0174】

ある特定の実施形態において、正規化を含む処理工程は、スタティックウィンドウ（static window）に対する正規化を含み、いくつかの実施形態において、正規化を含む処理工程は、ムービング（moving）ウィンドウまたはスライディングウィンドウ（sliding window）に対する正規化を含む。用語「ウィンドウ」は、本明細書中で使用されるとき、解析のために選択され、時折、比較のための参照として使用される（例えば、正規化および／または他の数学的もしくは統計的操作のために使用される）、1つまたはそれを超える部分のことを指す。用語「スタティックウィンドウに対する正規化」は、本明細書中で使用されるとき、試験被験体のデータセットと参照被験体のデータセットとの比較のために選択された1つまたはそれを超える部分を用いる正規化プロセスのことを指す。いくつかの実施形態において、選択された部分は、プロファイルを生成するために使用される。スタティックウィンドウは、一般に、操作中および／または解析中に変化しない所定の部分セットを含む。用語「ムービングウィンドウに対する正規化」および「スライディングウィンドウに対する正規化」は、本明細書中で使用されるとき、選択された試験部分のゲノム領域に局在する部分（例えば、すぐ接して囲んでいる隣接する遺伝的部分または遺伝的区分など）に対して行われる正規化のことを指し、ここで、1つまたはそれを超える選択された試験部分は、その選択された試験部分をすぐ接して囲んでいる部分に対して正規化される。ある特定の実施形態において、これらの選択された部分は、プロファイルを生成するために使用される。スライディングウィンドウ正規化またはムービングウィンドウ正規化は、隣接する試験部分に繰り返してムービングまたはスライディングすること、および新たに選択された試験部分を、その新たに選択された試験部分をすぐ接して囲んでいるかまたはその新たに選択された試験部分に隣接する部分に対して正規化することを含むことが多く、ここで、隣接するウィンドウは、1つまたはそれを超える部分を共通して有する。ある特定の実施形態において、複数の選択された試験部分および／または染色体が、スライディングウィンドウプロセスによって解析され得る。

20

30

【0175】

いくつかの実施形態において、スライディングウィンドウまたはムービングウィンドウに対する正規化は、1つまたはそれを超える値を生成し得、ここで、各値は、異なるゲノム領域（例えば、染色体）から選択される異なる参照部分セットに対する正規化に相当する。ある特定の実施形態において、生成された1つまたはそれを超える値は、累積和（例えば、選択された部分、ドメイン（例えば、染色体の一部）または染色体）に対する正規化されたカウントプロファイルの積分の推定数値）である。スライディングウィンドウまたはムービングウィンドウプロセスによって生成される値は、プロファイルを生成し、アウトカムに到達するのを促進するために使用され得る。いくつかの実施形態において、1つまたはそれを超える部分の累積和は、ゲノム位置の関数として表示され得る。ムービングウィンドウ解析またはスライディングウィンドウ解析は、微小欠失および／または微小挿入の存在もしくは非存在についてゲノムを解析するために使用されるときがある。ある特定の実施形態において、1つまたはそれを超える部分の累積和の表示は、コピー数変異（例えば、微小欠失、微小重複）の領域の存在もしくは非存在を識別するために使用され

40

50

る。いくつかの実施形態において、ムービングウィンドウ解析またはスライディングウィンドウ解析は、微小欠失を含むゲノム領域を識別するために使用され、ある特定の実施形態において、ムービングウィンドウ解析またはスライディングウィンドウ解析は、微小重複を含むゲノム領域を識別するために使用される。

【0176】

使用され得る正規化プロセスのある特定の例、例えば、LOESS、PERUN、ChAIおよび主成分正規化方法が、本明細書の以後に、より詳細に説明される。

【0177】

いくつかの実施形態において、処理工程は、重み付けを含む。用語「重み付けされる」、「重み付けする」もしくは「重み関数」またはそれらの文法上の派生物もしくは等価物は、本明細書中で使用されるとき、他のデータセットの特徴または変数に対してある特定のデータセットの特徴または変数の影響を変化させる（例えば、選択された部分または参照ゲノムの部分におけるデータの質または有用性に基づいて、1つまたはそれを超える部分または参照ゲノムの部分に含まれるデータの有意性および／または寄与を増減させる）ために利用されるときがあるデータセットの一部または全部の数学的操作のことを指す。重み付け関数は、いくつかの実施形態において、測定値の分散が比較的小さいデータの影響を増大させるためおよび／または測定値の分散が比較的大きいデータの影響を減少させるために使用され得る。例えば、過小提示または低品質の配列データを有する参照ゲノムの部分は、データセットに対する影響を最小にするために「重み付けを小さく」され得るのに対して、選択された参照ゲノムの部分は、データセットに対する影響を増大させるために「重み付けを大きく」され得る。重み付け関数の非限定的な例は、 $[1 / (\text{標準偏差})^2]$ である。重み付け工程は、正規化工程と実質的に同様の様式で行われるときがある。いくつかの実施形態において、データセットを所定の変数（例えば、重み付け変数）で除算する。所定の変数（例えば、最小化された目的関数、Phi）は、データセットの異なる一部を異なって重み付けする（例えば、他のデータタイプの影響を減少させつつ、ある特定のデータタイプの影響を増大させる）ために選択されることが多い。

【0178】

ある特定の実施形態において、処理工程は、1つまたはそれを超える数学的操作および／または統計的操作を含み得る。任意の好適な数学的操作および／または統計的操作が、本明細書中に記載されるデータセットを解析するためおよび／または操作するために、単独でまたは組み合わせて使用され得る。任意の好適な数の数学的操作および／または統計的操作を使用することができる。いくつかの実施形態において、データセットは、1回もしくはそれを超える回数、5回もしくはそれを超える回数、10回もしくはそれを超える回数または20回もしくはそれを超える回数だけ、数学的および／または統計的に操作され得る。使用され得る数学的操作および統計的操作の非限定的な例としては、加算、減算、乗算、除算、代数関数、最小二乗推定量、カーブフィッティング、微分方程式、有理多項式、二重多項式（double polynomial）、直交多項式、z得点、p値、カイ値、phi値、ピークレベルの解析、ピーク端位置の決定、ピーク面積比の計算、染色体レベル中央値の解析、平均絶対偏差の算出、平方残差の和、平均値、標準偏差、標準誤差などまたはそれらの組み合わせが挙げられる。数学的操作および／または統計的操作は、配列リードデータの全部もしくは一部またはその処理されたものに対して行われ得る。統計的に操作され得るデータセットの変数または特徴の非限定的な例としては、生のカウント、フィルタリングされたカウント、正規化されたカウント、ピーク高さ、ピーク幅、ピーク面積、ピーク端、側方許容（lateral tolerance）、P値、中央値レベル、平均値レベル、ゲノム領域内のカウントの分布、核酸種の相対的な提示などまたはそれらの組み合わせが挙げられる。

【0179】

いくつかの実施形態において、処理工程は、1つまたはそれを超える統計的アルゴリズムの使用を含み得る。任意の好適な統計的アルゴリズムが、本明細書中に記載されるデータセットを解析するためおよび／または操作するために、単独でまたは組み合わせて使用

され得る。任意の好適な数の統計的アルゴリズムを使用することができる。いくつかの実施形態において、データセットは、1つもしくはそれを超える、5つもしくはそれを超える、10個もしくはそれを超えるまたは20個もしくはそれを超える統計的アルゴリズムを用いて解析され得る。本明細書中に記載される方法とともに使用するのに適した統計的アルゴリズムの非限定的な例としては、決定木、対立帰無仮説、多重比較、総括的検定、ベールン・フィッシャー問題、ブートストラッピング、独立した有意性検定を組み合わせるためのフィッシャーの方法、帰無仮説、第1種の過誤、第2種の過誤、正確検定、1標本Z検定、2標本Z検定、1標本t検定、対応のあるt検定、等しい分散を有するプールされた2標本t検定、不等分散を有するプールされない2標本t検定、1比率z検定、プールされた2比率z検定、プールされない2比率z検定、1標本カイ二乗検定、分散を等しくするための2標本F検定、信頼区間、信用区間、有意性、メタ解析、線形単回帰、ロバスト線形回帰などまたは前述のものの組み合わせが挙げられる。統計的アルゴリズムを用いて解析され得るデータセットの変数または特徴の非限定的な例としては、生のカウント、フィルタリングされたカウント、正規化されたカウント、ピーク高さ、ピーク幅、ピーク端、側方許容、P値、中央値レベル、平均値レベル、ゲノム領域内のカウントの分布、核酸種の相対的な提示などまたはそれらの組み合わせが挙げられる。

10

【0180】

ある特定の実施形態において、データセットは、複数の（例えば、2つもしくはそれを超える）統計的アルゴリズム（例えば、最小二乗回帰、主成分分析、線形判別分析、二次判別分析、バギング、ニューラルネットワーク、サポートベクターマシンモデル、ランダムフォレスト、分類木モデル、K最近隣法、ロジスティック回帰および／またはloss平滑化法）ならびに／または数学的操作および／もしくは統計的操作（例えば、本明細書中で操作と称されるもの）を使用することによって解析され得る。いくつかの実施形態において、複数の操作の使用により、アウトカムを提供するために使用され得るN次元空間が生成され得る。ある特定の実施形態において、複数の操作を用いることによるデータセットの解析は、そのデータセットの複雑さおよび／または次元を低下させ得る。例えば、参照データセットに対して複数の操作を使用することにより、参照サンプルの状態（例えば、選択されたコピー数変異に対して陽性または陰性）に応じて、コピー数変異の存在もしくは非存在を表すために使用され得るN次元空間（例えば、確率プロット）が生成され得る。実質的に類似の操作セットを用いたテストサンプルの解析は、各テストサンプルに対してN次元のポイントを生成するために使用され得る。試験被験体のデータセットの複雑さおよび／または次元は、参照データから生成されたN次元空間と容易に比較され得る単一値またはN次元のポイントにまで低下するときがある。参照被験体のデータによって占有されたN次元空間に入るテストサンプルのデータは、参照被験体の遺伝的状态と実質的に同様の遺伝的状态を示唆する。参照被験体のデータに占有されたN次元空間に入らないテストサンプルのデータは、参照被験体の遺伝的状态と実質的に異なる遺伝的状态を示唆する。いくつかの実施形態において、参照は、正倍数性であるか、またはコピー数変異もしくは医学的症状を別段有しない。

20

30

【0181】

データセットがカウントされ、必要に応じてフィルタリングされ、正規化された後、処理されたデータセットは、いくつかの実施形態において、1つまたはそれを超えるフィルタリング手順および／または正規化手順によってさらに操作され得る。1つまたはそれを超えるフィルタリング手順および／または正規化手順によってさらに操作されたデータセットは、ある特定の実施形態において、プロファイルを生成するために使用され得る。1つまたはそれを超えるフィルタリング手順および／または正規化手順は、いくつかの実施形態において、データセットの複雑さおよび／または次元を低下させ得るときがある。アウトカムは、低下した複雑さおよび／または次元のデータセットに基づいて提供され得る。

40

【0182】

いくつかの実施形態において、部分は、誤差の尺度（例えば、標準偏差、標準誤差、算

50

出された分散、p 値、平均絶対誤差 (MAE)、平均絶対偏差および／または平均絶対偏差 (MAD) に従ってフィルタリングされ得る。ある特定の実施形態において、誤差の尺度とは、カウントのばらつきのことを指す。いくつかの実施形態において、部分は、カウントのばらつきに従ってフィルタリングされる。ある特定の実施形態において、カウントのばらつきは、複数のサンプル (例えば、複数の被験体、例えば、50 もしくはそれを超える、100 もしくはそれを超える、500 もしくはそれを超える、1000 もしくはそれを超える、5000 もしくはそれを超えるまたは 10,000 もしくはそれを超える被験体から得られた複数のサンプル) に対する、参照ゲノムの部分 (すなわち、部分) にマッピングされたカウントに対して決定される誤差の尺度である。いくつかの実施形態において、所定の上部範囲より上のカウントのばらつきを有する部分が、フィルタリングされる (例えば、考慮すべきものから除外される)。いくつかの実施形態において、所定の上部範囲は、約 50、約 52、約 54、約 56、約 58、約 60、約 62、約 64、約 66、約 68、約 70、約 72、約 74 と等しいかもしくはそれを超える MAD 値または約 76 と等しいかもしくはそれを超える MAD 値である。いくつかの実施形態において、所定の下部範囲より下のカウントのばらつきを有する部分が、フィルタリングされる (例えば、考慮すべきものから除外される)。いくつかの実施形態において、所定の下部範囲は、約 40、約 35、約 30、約 25、約 20、約 15、約 10、約 5、約 1 と等しいかもしくはそれ未満の MAD 値または約 0 と等しいかもしくはそれ未満の MAD 値である。いくつかの実施形態において、所定の範囲外のカウントのばらつきを有する部分が、フィルタリングされる (例えば、考慮すべきものから除外される)。いくつかの実施形態において、所定の範囲は、ゼロより大きくかつ約 76 未満、約 74 未満、約 73 未満、約 72 未満、約 71 未満、約 70 未満、約 69 未満、約 68 未満、約 67 未満、約 66 未満、約 65 未満、約 64 未満、約 62 未満、約 60 未満、約 58 未満、約 56 未満、約 54 未満、約 52 未満または約 50 未満の MAD 値である。いくつかの実施形態において、所定の範囲は、ゼロより大きくかつおよび約 67.7 未満の MAD 値である。いくつかの実施形態において、所定の範囲内のカウントのばらつきを有する部分が、選択される (例えば、コピー数変異の存在もしくは非存在を判定するために使用される)。

【0183】

いくつかの実施形態において、部分のカウントのばらつきは、分布 (例えば、正規分布) に相当する。いくつかの実施形態において、その分布の、ある分位点内の部分が、選択される。いくつかの実施形態において、分布に対して約 99.9%、99.8%、99.7%、99.6%、99.5%、99.4%、99.3%、99.2%、99.1%、99.0%、98.9%、98.8%、98.7%、98.6%、98.5%、98.4%、98.3%、98.2%、98.1%、98.0%、97%、96%、95%、94%、93%、92%、91%、90%、85%、80% と等しいかもしくはそれ未満または約 75% の分位点と等しいかもしくはそれ未満の分位点内の部分が、選択される。いくつかの実施形態において、カウントのばらつきの分布の 99% 分位点内の部分が、選択される。いくつかの実施形態において、99% 分位点内の $MAD > 0$ かつ $MAD < 67.725$ を有する部分が選択され、それにより、参照ゲノムの安定した部分のセットが識別される。

【0184】

PERUN に関する部分のフィルタリングの非限定的な例は、例えば、本明細書中および国際特許出願番号 PCT/US12/59123 (WO2013/052913) (本文、表、式および図面のすべてを含むこの全内容が参照により本明細書中に組み込まれる) に提供されている。部分は、誤差の尺度に基づいてまたは誤差の尺度に部分的に基づいて (based on part on) フィルタリングされ得る。R 因子などの偏差の絶対値を含む誤差の尺度は、ある特定の実施形態において、部分の除去または重み付けのために使用され得る。R 因子は、いくつかの実施形態において、実際の測定値から予測されるカウント値で除算された、実際の測定値から予測されるカウント値の絶対偏差の和として定義される (例えば、2012 年 10 月 5 日に出願され、2013 年 4 月 11 日に W

02013/052913として公開された特許出願番号PCT/US2012/059123の228頁の式C)。偏差の絶対値を含む誤差の尺度を使用してもよいが、その代わりに好適な誤差の尺度を使用してもよい。ある特定の実施形態において、偏差の絶対値を含まない誤差の尺度、例えば、平方に基づく散らばりが、使用され得る。いくつかの実施形態において、部分は、マッピング精度の尺度（例えば、マッピング精度スコア）に従ってフィルタリングされるかまたは重み付けされる。ある部分は、その部分にマッピングされた比較的少ない数の配列リード（例えば、その部分にマッピングされた0、1、2、3、4、5個のリード）に従ってフィルタリングされるかまたは重み付けされるときがある。ある部分は、反復配列の割合またはパーセントに従ってフィルタリングされるかまたは重み付けされるときがある。ある特定の実施形態において、部分は、(i) マッピング精度の尺度、(ii) 誤差の尺度（例えば、R因子）および(iii) 反復配列の割合またはパーセントのうちの1つまたはそれを超えるものに従ってフィルタリングされるかまたは重み付けされる。部分は、行われる解析のタイプに従ってフィルタリングされ得るかまたは重み付けされ得る。例えば、13番、18番および/または21番染色体の異数性解析の場合、性染色体が、フィルタリングされ得、常染色体または常染色体のサブセットだけが解析され得る。

10

【0185】

特定の実施形態において、以下のフィルタリングプロセスが、用いられ得る。所与の染色体（例えば、21番染色体）内の同じ部分（例えば、参照ゲノムの部分）のセットが選択され、罹患サンプルおよび非罹患サンプルにおけるリードの数が比較される。そのギャップは、トリソミー21サンプルおよび正倍数性サンプルを関係づけ、21番染色体のほとんどをカバーする部分のセットを含む。その部分のセットは、正倍数性サンプルとT21サンプルとの間で同じである。ある部分が定義され得るとき、部分のセットと単一の区分との間の違いは、重大ではない。異なる患者において、同じゲノム領域が比較される。このプロセスは、トリソミー解析（例えば、T21に加えてまたはT21の代わりにT13またはT18）のために使用され得る。

20

【0186】

データセットがカウントされ、必要に応じてフィルタリングされ、正規化された後、処理されたデータセットは、いくつかの実施形態において、重み付けによって操作され得る。1つまたはそれを超える部分が、ある特定の実施形態において、選択された部分に含まれるデータ（例えば、ノイズの多いデータ、情報価値のないデータ）の影響を減少させるための重み付けに対して選択され得、いくつかの実施形態において、1つまたはそれを超える部分が、選択された部分に含まれるデータ（例えば、分散が小さいと計測されたデータ）の影響を増強するためまたは増大するための重み付けに対して選択され得る。いくつかの実施形態において、データセットは、分散が大きいデータの影響を減少させ、分散が小さいデータの影響を増大する単一の重み付け関数を用いて重み付けされる。重み付け関数は、分散が大きいデータの影響を減少させ、分散が小さいデータの影響を増大するために使用されるときがある（例えば、 $[1 / (\text{標準偏差})^2]$ ）。いくつかの実施形態において、重み付けによってさらに操作された処理されたデータのプロファイルのプロットは、分類および/またはアウトカムの提供を容易にするために生成される。アウトカムは、重み付けされたデータのプロファイルのプロットに基づいて提供され得る。

30

40

【0187】

部分のフィルタリングまたは重み付けは、解析における1つまたはそれを超える好適な時点において行われ得る。例えば、部分は、配列リードが参照ゲノムの部分に対してマッピングされる前またはマッピングされた後に、フィルタリングされ得るかまたは重み付けされ得る。部分は、いくつかの実施形態において、個々のゲノム部分に対する実験上のバイアスが決定される前または決定された後に、フィルタリングされ得るかまたは重み付けされ得る。ある特定の実施形態において、部分は、ゲノム区分レベルが算出される前または算出された後に、フィルタリングされ得るかまたは重み付けされ得る。

【0188】

50

データセットが、カウントされ、必要に応じてフィルタリングされ、正規化され、必要に応じて重み付けされた後、処理されたデータセットは、いくつかの実施形態において、1つまたはそれを超える数学的操作および／または統計的操作（例えば、統計関数または統計的アルゴリズム）操作によって操作され得る。ある特定の実施形態において、処理されたデータセットは、1つまたはそれを超える選択された部分、染色体または染色体の部分に対するZ得点を算出することによってさらに操作され得る。いくつかの実施形態において、処理されたデータセットは、P値を算出することによってさらに操作され得る。ある特定の実施形態において、数学的操作および／または統計的操作は、倍数性および／または胎児分率に関する1つまたはそれを超える仮定を含む。いくつかの実施形態において、1つまたはそれを超える統計的操作および／または数学的操作によってさらに操作された処理されたデータのプロファイルのプロットは、分類および／またはアウトカムの提供を容易にするために生成される。アウトカムは、統計的および／または数学的に操作されたデータのプロファイルのプロットに基づいて提供され得る。統計的および／または数学的に操作されたデータのプロファイルのプロットに基づいて提供されるアウトカムは、倍数性および／または胎児分率に関する1つまたはそれを超える仮定を含むことが多い。

10

【0189】

ある特定の実施形態において、データセットがカウントされ、必要に応じてフィルタリングされ、正規化された後、処理されたデータセットに対して複数の操作が行われることにより、N次元空間および／またはN次元のポイントが生成される。アウトカムは、N次元で解析されたデータセットのプロファイルのプロットに基づいて提供され得る。

20

【0190】

いくつかの実施形態において、データセットは、データセットの一部として、またはデータセットが処理された後および／もしくは操作された後に、1つまたはそれを超えるピークレベル解析、ピーク幅解析、ピーク端位置解析、ピーク側方許容（peak lateral tolerance）など、その微分演算または前述のものの組み合わせを用いて処理される。いくつかの実施形態において、1つまたはそれを超えるピークレベル解析、ピーク幅解析、ピーク端位置解析、ピーク側方許容など、その微分演算または前述のものの組み合わせを用いて処理されたデータのプロファイルのプロットが、分類および／またはアウトカムの提供を容易にするために生成される。アウトカムは、1つまたはそれを超えるピークレベル解析、ピーク幅解析、ピーク端位置解析、ピーク側方許容など、その微分演算または前述のものの組み合わせを用いて処理されたデータのプロファイルのプロットに基づいて提供され得る。

30

【0191】

いくつかの実施形態において、対象のコピー数変異を実質的に含まない1つまたはそれを超える参照サンプルの使用は、参照カウントプロファイルの中央値を生成するために使用され得、その中央値は、コピー数変異が存在しないことを代表する所定の値をもたらし得、試験被験体がコピー数変異を有した場合、コピー数変異がその試験被験体内に位置するゲノム位置に対応する区域における所定の値から逸脱することが多い。コピー数変異に関連する医学的症状のリスクがあるかまたはその医学的症状に罹患している試験被験体では、選択された部分または区分に対する数値は、非罹患のゲノム位置に対する所定の値から有意に変動すると予想される。ある特定の実施形態において、対象のコピー数変異を有すると判明している1つまたはそれを超える参照サンプルの使用は、参照カウントプロファイルの中央値を生成するために使用され得、その中央値は、コピー数変異が存在することを代表する所定の値をもたらし得、試験被験体がコピー数変異を有しないゲノム位置に対応する区域における所定の値から逸脱することが多い。コピー数変異に関連する医学的症状のリスクがないかまたはその医学的症状に罹患していない試験被験体では、選択された部分または区分に対する数値は、罹患ゲノム位置に対する所定の値から有意に変動すると予想される。

40

【0192】

50

いくつかの実施形態において、データの解析および処理は、1つまたはそれを超える仮定の使用を含み得る。好適な数またはタイプの仮定が、データセットを解析するためまたは処理するために使用され得る。データの処理および／または解析のために使用され得る仮定の非限定的な例としては、母体の倍数性、胎児の寄与、参照集団におけるある特定の配列の保有率 (prevalence)、民族的バックグラウンド、関係する家族における選択された医学的症状の有病率、異なる患者由来の生のカウントプロファイル間ならびに／またはGC正規化およびリピートマスク (例えば、GCRM) の後のラン間の類似、完全一致がPCRアーチファクト (例えば、同一の塩基位置) を表すこと、胎児数量アッセイ (例えば、FQA) に固有の仮定、双子に関する仮定 (例えば、双子の2人ともおよび1人だけが罹患している場合、有効な胎児分率は、測定された胎児分率の合計の50% 10 だけである (三つ子、四つ子なども同様に))、胎児の細胞を含まないDNA (例えば、cfDNA) がゲノム全体を均一にカバーすることなどおよびそれらの組み合わせが挙げられる。

【0193】

マッピングされた配列リードの質および／または深度が、正規化されたカウントプロファイルに基づいて所望の信頼水準 (例えば、95%またはそれより高い信頼水準) においてコピー数変異の存在もしくは非存在のアウトカムの予測を可能にしない場合、データ解析および／またはアウトカムの提供にとって有用なさらなる数値を生成するために、1つまたはそれを超えるさらなる数学的操作アルゴリズムおよび／または統計的予測アルゴリズムが使用され得る。用語「正規化されたカウントプロファイル」は、本明細書中で使用 20 されるとき、正規化されたカウントを用いて生成されるプロファイルのことを指す。正規化されたカウントおよび正規化されたカウントプロファイルを生成するために使用され得る方法の例は、本明細書中に記載される。述べるように、マッピングされてカウントされた配列リードは、テストサンプルのカウントまたは参照サンプルのカウントに関して正規化され得る。いくつかの実施形態において、正規化されたカウントプロファイルは、プロットとして示され得る。

LOESS正規化

【0194】

LOESSは、k最近隣法に基づくメタモデルにおいて複数の回帰モデルを組み合わせる当該分野で公知の回帰モデリング法である。LOESSは、局所重み付け多項式回帰と 30 称されるときがある。GC LOESSは、いくつかの実施形態において、LOESSモデルを、参照ゲノムの部分に対するフラグメントカウント (例えば、配列リード、カウント) とGC組成との関係に適用する。LOESSを用いてデータポイントセットを通して滑らかな曲線をプロットすることは、特に、各平滑化値が、y軸の散布図の基準変数の値の範囲にわたって、重み付けされた二次最小二乗回帰によって与えられるとき、LOESS 曲線と呼ばれるときがある。あるデータセットにおける各ポイントに対して、LOESS法は、そのデータのサブセットに低次多項式を当てはめ、説明変数値は、応答を推定しているポイントに近い。その多項式は、重み付き最小二乗を用いて当てはめられ、応答が推定されているポイントに近いポイントにより大きい重みを与えられ、さらに離れているポイントにより小さい重みを与えられる。次いで、あるポイントに対する回帰関数の値は 40 、そのデータポイントに対する説明変数値を用いて局所多項式を評価することによって得られる。LOESSの当てはめは、回帰関数値が各データポイントに対して計算された後、完全であると考えられるときがある。この方法の詳細の多く (例えば、多項式モデルおよび重みの程度) は、フレキシブルである。

PERUN正規化

【0195】

核酸の指標に関連する誤差を減少させるための正規化の方法は、本明細書中および国際特許出願番号PCT/US12/59123 (WO2013/052913) (本文、表、式および図面のすべてを含むこの全内容が、参照により本明細書中に組み込まれる) に記載されているパラメータ化された誤差除去および不偏正規化 (Parameteriz 50

ed Error Removal and Unbiased Normalization) (PERUN)と本明細書中で称される。PERUN法は、そのような指標に基づく予測を混乱させる誤差の影響を減少させる目的で、種々の核酸の指標（例えば、核酸配列リード）に対して適用され得る。

【0196】

例えば、PERUN法は、サンプル由来の核酸配列リードに適用され得、ゲノム区分レベルの判定を損ない得る誤差の影響を減少させ得る。そのような適用は、核酸配列リードを用いて、ヌクレオチド配列の変動するレベル（例えば、部分、ゲノム区分レベル）として示される被験体においてコピー数変異の存在もしくは非存在を判定するために有用である。部分における変異の非限定的な例は、染色体異数性（例えば、トリソミー21、トリソミー18、トリソミー13）および性染色体の存在もしくは非存在（例えば、男性におけるXYに対する女性におけるXX）である。常染色体（例えば、性染色体以外の染色体）のトリソミーは、罹患した常染色体と称され得る。ゲノム区分レベルにおける変異の他の非限定的な例としては、微小欠失、微小挿入、重複およびモザイク現象が挙げられる。

【0197】

ある特定の適用において、PERUN法は、特定の参照ゲノムの部分にマッピングされた核酸リードを正規化することによって実験上のバイアスを減少させ得、その後者は、部分と称され、参照ゲノムの部分と称されるときがある。そのような適用において、PERUN法は、通常、特定の参照ゲノムの部分における核酸リードのカウントをいくつかのサンプルにわたって3次元で正規化する。PERUNおよびその適用の詳細な説明は、国際特許出願番号PCT/US12/59123 (WO2013/052913) および米国特許出願公開番号US20130085681に提供されており、本文、表、式および図面のすべてを含むこれらの全内容は、参照により本明細書中に組み込まれる。

【0198】

ある特定の実施形態において、PERUN法は、(a) テストサンプルに対する参照ゲノムの部分にマッピングされた配列リードカウント、(b) テストサンプルに対する実験上のバイアス（例えば、GCバイアス）、および(c) (i) 配列リードがマッピングされた参照ゲノムの部分に対する実験上のバイアスと(ii) その部分にマッピングされた配列リードのカウントとの当てはめ関係に対する1つまたはそれを超えるフィットパラメータ（例えば、当てはめの推定値）から、参照ゲノムの部分に対するゲノム区分レベルを算出することを含む。参照ゲノムの各部分に対する実験上のバイアスは、(i) 参照ゲノムの各部分にマッピングされた配列リードのカウントと(ii) 参照ゲノムの各部分に対するマッピングの特徴との間の各サンプルに対する当てはめ関係に従って複数のサンプルにわたって決定され得る。各サンプルに対するこの当てはめ関係は、3次元において複数のサンプルに対してアセンブルされ得る。このアセンブリは、ある特定の実施形態において、実験上のバイアスに従って順序づけられ得るが、PERUN法は、実験上のバイアスに従ってアセンブリを順序づけずに行われ得る。各サンプルに対する当てはめ関係および参照ゲノムの各部分に対する当てはめ関係は、独立して、当該分野で公知の好適な当てはめプロセスによって線形関数または非線形関数に当てはめられ得る。

【0199】

いくつかの実施形態において、関係は、幾何的および／またはグラフ的な関係である。いくつかの実施形態において、関係は、数学的関係である。いくつかの実施形態において、関係は、プロットされる。いくつかの実施形態において、関係は、直線関係である。ある特定の実施形態において、関係は、非直線関係である。ある特定の実施形態において、関係は、回帰（例えば、回帰直線）である。回帰は、線形回帰または非線形回帰であり得る。関係は、数学的方程式によって表現され得る。関係は、1つまたはそれを超える定数によって部分的に定義されることが多い。関係は、当該分野で公知の方法によって生成され得る。2次元における関係は、1つまたはそれを超えるサンプルに対して生成され得、ある特定の実施形態では、誤差を立証するかまたは誤差をおそらく立証する変数が、その次元のうちの1つまたはそれを超える次元に対して選択され得る。関係は、例えば、ユー

ザーによって提供される2つまたはそれを超える変数の値を用いてグラフをプロットする当該分野で公知のグラフ化ソフトウェアを用いて生成され得る。関係は、当該分野で公知の方法（例えば、グラフ化ソフトウェア）を用いて当てはめられ得る。ある特定の関係は、線形回帰によって当てはめられ得、その線形回帰は、傾きの値および切片の値をもたらし得る。ある特定の関係は、線形でないときがあり、非線形関数、例えば、放物線関数、双曲線関数または指数関数（例えば、二次関数）によって当てはめられ得る。

【0200】

PERUN法では、1つまたはそれを超える当てはめ関係は、線形であり得る。実験上のバイアスがGCバイアスであり、マッピングの特徴がGC含有量である、妊婦由来の無細胞循環核酸の解析の場合、(i)各部分にマッピングされた配列リードのカウントと(i i)参照ゲノムの各部分に対するGC含有量との間の、サンプルに対する当てはめ関係は、線形であり得る。後者の当てはめ関係の場合、その傾きは、GCバイアスに関し、それらの当てはめ関係が複数のサンプルにわたってアセンブルされるとき、GCバイアス係数は、各サンプルに対して決定され得る。そのような実施形態において、複数のサンプルに対する当てはめ関係、および(i)その部分に対するGCバイアス係数と(i i)部分にマッピングされた配列リードのカウントとの間の部分もまた、線形であり得る。切片および傾きは、後者の当てはめ関係から得ることができる。そのような適用において、その傾きは、GC含有量に基づいてサンプル特異的バイアスを扱い、その切片は、すべてのサンプルに共通の部分特異的減衰パターンを扱う。PERUN法は、あるアウトカム（例えば、コピー数変異の存在もしくは非存在；胎児の性別の判定）を提供するためにゲノム区分レベルを算出するとき、そのようなサンプル特異的バイアスおよび部分特異的減衰を有意に減少させ得る。

【0201】

いくつかの実施形態において、PERUN正規化は、線形関数への当てはめを利用し、方程式I、方程式IIまたはそれらの導関数によって記載される。

方程式I：

$$M = L I + G S \quad (I)$$

方程式II：

$$L = (M - G S) / I \quad (II)$$

【0202】

いくつかの実施形態において、Lは、PERUNによって正規化されたレベルまたはプロファイルである。いくつかの実施形態において、Lは、PERUN正規化手順からの所望のアウトプットである。ある特定の実施形態において、Lは、部分特異的である。いくつかの実施形態において、Lは、複数の参照ゲノムの部分に従って決定され、ゲノム、染色体、部分またはそれらのセグメントの、PERUNによって正規化されたレベルを表す。レベルLは、さらなる解析のため（例えば、Z値、母体の欠失／重複、胎児の微小欠失／微小重複、胎児の性別、性異数性などを判定するため）に使用されることが多い。方程式IIに従った正規化の方法は、パラメータ化された誤差除去および不偏正規化(PERUN)と命名されている。

【0203】

いくつかの実施形態において、Gは、線形モデル、LOESSまたは任意の等価なアプローチを用いて計測されたGCバイアス係数である。いくつかの実施形態において、Gは、傾きである。いくつかの実施形態において、GCバイアス係数Gは、部分iに対するカウントM（例えば、生のカウント）および参照ゲノムから決定される部分iのGC含有量に対する回帰の傾きとして評価される。いくつかの実施形態において、Gは、Mから抽出され、ある関係に従って決定される二次情報を表す。いくつかの実施形態において、Gは、サンプル（例えば、テストサンプル）に対する部分特異的カウントセットおよび部分特異的GC含有量値セットに対する関係を表す。いくつかの実施形態において、部分特異的GC含有量は、参照ゲノムから得られる。いくつかの実施形態において、部分特異的GC含有量は、観察されたまたは計測されたGC含有量から得られる（例えば、サンプルから

計測される)。GC バイアス係数は、サンプル群における各サンプルに対して決定されることが多く、一般に、テストサンプルに対して決定される。GC バイアス係数は、サンプル特異的であることが多い。いくつかの実施形態において、GC バイアス係数は、定数である。ある特定の実施形態において、GC バイアス係数は、いったんサンプルに対して得られると、変化しない。

【0204】

いくつかの実施形態において、I は、切片であり、S は、直線関係から導かれる傾きである。いくつかの実施形態において、I および S が導かれる関係は、G が導かれる関係とは異なる。いくつかの実施形態において、I および S が導かれる関係は、所与の実験上の設定に対して固定される。いくつかの実施形態において、I および S は、カウント（例えば、生のカウント）に従う直線関係および複数のサンプルに従う GC バイアス係数から導かれる。いくつかの実施形態において、I および S は、独立してテストサンプルから導かれる。いくつかの実施形態において、I および S は、複数のサンプルから導かれる。I および S は、部分特異的であることが多い。いくつかの実施形態において、I および S は、正倍数性サンプルにおける参照ゲノムのすべての部分に対して $L = 1$ という仮定を用いて決定される。いくつかの実施形態において、直線関係が、正倍数性サンプルに対して決定され、選択された部分に特異的な I および S 値が決定される ($L = 1$ であると仮定する)。ある特定の実施形態において、同じ手順が、ヒトゲノムにおける参照ゲノムのすべての部分に対して適用され、切片 I および傾き S のセットが、すべての部分に対して決定される。

【0205】

いくつかの実施形態において、交差検証アプローチが適用される。交差検証は、回転推定 (rotation estimation) と称されるときがある。いくつかの実施形態において、交差検証アプローチは、予測モデル（例えば、PERUN）が、テストサンプルを用いて実際にどれくらい正確に行うかを評価するために適用される。いくつかの実施形態において、1 回の交差検証は、データのサンプルを相補的なサブセットに分割すること、1 つのサブセット（例えば、訓練セットと称されるときがある）に対して交差検証解析を行うこと、および別のサブセット（例えば、検証セットまたは試験セットと呼ばれるときがある）を用いてその解析を検証することを含む。ある特定の実施形態において、複数回の交差検証は、異なる分割および／または異なるサブセット）を用いて行われる。交差検証アプローチの非限定的な例としては、一個抜き (leave-one-out)、スライディングエッジ (sliding edges)、K 倍、2 倍、繰り返しランダムサブサンプリング (repeat random sub-sampling) などまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、交差検証は、既知の正倍数性胎児を含むサンプルセットの 90% を含む作業セットをランダムに選択し、そのサブセットを使用してモデルを訓練する。ある特定の実施形態において、ランダムな選択は、100 回繰り返されることにより、すべての部分に対して 100 個の傾きおよび 100 個の切片のセットがもたらされる。

【0206】

いくつかの実施形態において、M の値は、テストサンプルから導かれる計測された値である。いくつかの実施形態において、M は、ある部分に対する計測された生のカウントである。いくつかの実施形態において、値 I および S が、ある部分に対して入手可能である場合、計測値 M は、テストサンプルから決定され、それを用いて、方程式 I I に従って、ゲノム、染色体、セグメントまたはその一部に対する、PERUN によって正規化されたレベル L が決定される。

【0207】

したがって、複数のサンプルにわたって PERUN 法を並行して配列リードに適用することにより、(i) サンプル特異的な実験上のバイアス（例えば、GC バイアス）および (ii) サンプルに共通の部分特異的な減衰によって引き起こされる誤差が有意に減少し得る。これらの 2 つの誤差の供給源の各々が別々にまたは連続的に対処される他の方法は、P

PERUN法ほど効率的にこれらを減少させることができないことが多い。理論に限定されるものではないが、PERUN法の一般的な加法プロセスが、他の正規化アプローチ（例えば、GC-LOESS）において用いられる一般的な乗法プロセスほど散らばりを拡大しないことが1つの理由で、PERUN法は、より効率的に誤差を減少すると予想される。

【0208】

さらなる正規化の手法および統計的手法が、PERUN法と組み合わせて使用され得る。さらなるプロセスは、PERUN法の使用前、使用後および／または使用中に適用され得る。PERUN法と組み合わせて使用され得るプロセスの非限定的な例は、本明細書の以後に記載される。

10

【0209】

いくつかの実施形態において、GC含有量に対するゲノム区分レベルの第2の正規化または調整は、PERUN法とともに使用され得る。好適なGC含有量の調整または正規化の手順が使用され得る（例えば、GC-LOESS、GCRM）。ある特定の実施形態において、さらなるGC正規化プロセスを適用するために、特定のサンプルが識別され得る。例えば、PERUN法の適用により、各サンプルに対するGCバイアスが決定され得、ある特定の閾値より大きいGCバイアスに関連するサンプルが、さらなるGC正規化プロセスに向けて選択され得る。そのような実施形態において、所定のしきい値レベルが、そのようなサンプルをさらなるGC正規化に向けて選択するために使用され得る。

【0210】

ある特定の実施形態において、ある部分をフィルタリングするかまたは重み付けするプロセスが、PERUN法とともに使用され得る。部分をフィルタリングするかまたは重み付けする好適なプロセスが使用され得、非限定的な例は、本明細書中、国際特許出願番号PCT/US12/59123（WO2013/052913）および米国特許出願公開番号US20130085681（これらの本文、表、式および図面のすべてを含む全内容が、参照により本明細書中に組み込まれる）に記載されている。いくつかの実施形態において、母体の挿入、重複および／または欠失（例えば、母体および／または胎児のコピー数変異）に関連する誤差を減少させる正規化の手法が、PERUN法とともに使用され得る。

20

【0211】

PERUN法によって算出されたゲノム区分レベルが、アウトカムを提供するために直接使用され得る。いくつかの実施形態において、ゲノム区分レベルは、胎児分率が約2%～約6%またはそれを超える（例えば、約4%またはそれを超える胎児分率）サンプルに対してアウトカムを提供するために直接使用され得る。PERUN法によって算出されたゲノム区分レベルは、アウトカムを提供するためにさらに処理されるときがある。いくつかの実施形態において、算出されたゲノム区分レベルは、標準化される。ある特定の実施形態において、試験部分（例えば、21番染色体）に対する算出されたゲノム区分レベルの和、平均値または中央値が、試験部分以外の部分（例えば、21番染色体以外の常染色体）に対する算出されたゲノム区分レベルの和、平均値または中央値で除算されることにより、実験上のゲノム区分レベルが生成され得る。実験上のゲノム区分レベルまたは生のゲノム区分レベルが、Z得点の算出などの標準化解析の一部として使用され得る。Z得点は、期待されるゲノム区分レベルを実験上のゲノム区分レベルまたは生のゲノム区分レベルから減算することによってサンプルに対して生成され得、得られた値は、そのサンプルに対する標準偏差で除算され得る。得られたZ得点は、ある特定の実施形態において、種々のサンプルに対して分配され、解析され得るか、または他の変数（例えば、胎児分率など）に関係し得、解析されることにより、アウトカムが提供され得る。

30

40

【0212】

本明細書中で述べられるように、PERUN法は、本質的にGCバイアスおよびGC含有量に従う正規化に限定されず、誤差の他の供給源に関連する誤差を減少させるために使用され得る。非GC含有量バイアスの供給源の非限定的な例は、マッピング精度である。

50

G C バイアスおよび G C 含有量以外の正規化パラメータが扱われるとき、当てはめ関係の 1 つまたはそれを超える関係は、非線形（例えば、双曲的、指数関数的）であり得る。いくつかの実施形態において、実験上のバイアスが、非直線関係から決定される場合、例えば、実験上のバイアスの曲率推定が、解析され得る。

【0213】

P E R U N 法は、種々の核酸の指標に適用され得る。核酸の指標の非限定的な例は、マイクロアレイ上の特定の位置における核酸配列リードおよび核酸レベルである。配列リードの非限定的な例としては、無細胞循環 D N A、無細胞循環 R N A、細胞 D N A および細胞 R N A から得られるものが挙げられる。P E R U N 法は、好適な参照配列（例えば、参照ゲノム D N A、参照細胞 R N A（例えば、トランスクリプトーム）およびそれらの部分（例えば、D N A または R N A トランスクリプトームのゲノム相補鎖の一部、染色体の一部））にマッピングされた配列リードに適用され得る。

10

【0214】

したがって、ある特定の実施形態において、細胞性核酸（例えば、D N A または R N A）が、核酸の指標として役立ち得る。参照ゲノム部分にマッピングされた細胞性核酸リードは、P E R U N 法を用いて正規化され得る。特定のタンパク質に結合した細胞性核酸は、クロマチン免疫沈降（C h I P）プロセスについて言及されるときがある。C h I P によって濃縮された核酸は、細胞性タンパク質と会合した核酸、例えば、D N A または R N A である。C h I P によって濃縮された核酸のリードは、当該分野で公知の技術を用いて得ることができる。C h I P によって濃縮された核酸のリードは、1 つまたはそれを超える参照ゲノムの部分にマッピングされ得、結果は、アウトカムを提供するために P E R U N 法を用いて正規化され得る。

20

【0215】

ある特定の実施形態において、細胞 R N A が、核酸の指標として役立ち得る。細胞 R N A リードは、参照 R N A 部分にマッピングされ得、アウトカムを提供するために P E R U N 法を用いて正規化され得る。トランスクリプトームと称される細胞 R N A に対する既知配列またはそのセグメントは、サンプル由来の R N A リードがマッピングされ得る参照として使用され得る。サンプル R N A のリードは、当該分野で公知の技術を用いて得ることができる。参照にマッピングされた R N A リードの結果は、アウトカムを提供するために P E R U N 法を用いて正規化され得る。

30

【0216】

いくつかの実施形態において、マイクロアレイ核酸レベルが、核酸の指標として役立ち得る。アレイ上の特定のアドレス、すなわちハイブリダイズ核酸に対する、サンプルにわたる核酸レベルが、P E R U N 法を用いて解析され、それによって、マイクロアレイ解析によって提供される核酸の指標が正規化され得る。このように、マイクロアレイ上の特定のアドレスまたはハイブリダイズ核酸は、マッピングされた核酸配列リードに対する部分と類似しており、P E R U N 法は、改善されたアウトカムを提供するためにマイクロアレイデータを正規化するために使用され得る。

C h A I 正規化

【0217】

核酸の指標に関連する誤差を減少させるために使用され得る別の正規化法は、本明細書中で C h A I と称され、主成分分析を利用することが多い。ある特定の実施形態において、主成分分析は、（a）リード密度分布に従って参照ゲノムの部分をフィルタリングし、それによって、フィルタリングされた部分のリード密度を含む、テストサンプルに対するリード密度プロファイルを提供すること（ここで、そのリード密度は、妊婦由来のテストサンプルからの循環無細胞核酸の配列リードを含み、リード密度分布は、複数のサンプルに対する部分のリード密度について測定される）、（b）そのテストサンプルに対するリード密度プロファイルを 1 つまたはそれを超える主成分に従って調整し、それによって、調整されたリード密度を含むテストサンプルプロファイルを提供すること（それらの主成分は、既知の正倍数性サンプルのセットからの主成分分析によって得られる）および（c

40

50

）テストサンプルプロファイルを参照プロファイルと比較することによって、比較結果を提供することを含む。いくつかの実施形態において、主成分分析は、（d）その比較結果に従ってテストサンプルに対するコピー数変異の存在もしくは非存在を判定することを含む。ChAI正規化のある特定の態様は、例えば、2014年10月2日に出願され、2015年4月9日にWO2015/051163として公開された特許出願番号PCT/US2014/058885に記載されている。

部分のフィルタリング

【0218】

ある特定の実施形態において、1つまたはそれを超える部分（例えば、ゲノムの部分）は、フィルタリングプロセスによって、考慮すべきものから除去される。ある特定の実施形態において、1つまたはそれを超える部分は、フィルタリングされる（例えば、フィルタリングプロセスに供される）ことによって、フィルタリングされた部分が提供される。いくつかの実施形態において、フィルタリングプロセスは、ある特定の部分を除去し、部分（例えば、部分のサブセット）を保持する。フィルタリングプロセスの後、保持された部分は、本明細書中で、フィルタリングされた部分と称されることが多い。いくつかの実施形態において、参照ゲノムの部分が、フィルタリングされる。いくつかの実施形態において、フィルタリングプロセスによって除去された参照ゲノムの部分は、コピー数変異の存在もしくは非存在（例えば、染色体異数性、微小重複、微小欠失）の判定に含められない。いくつかの実施形態において、リード密度に関連する部分（例えば、リード密度が、ある部分に対するものである場合）は、フィルタリングプロセスによって除去され、除去された部分に関連するリード密度は、コピー数変異の存在もしくは非存在（例えば、染色体異数性、微小重複、微小欠失）の判定に含められない。いくつかの実施形態において、リード密度プロファイルは、フィルタリングされた部分のリード密度を含み、および／またはフィルタリングされた部分のリード密度からなる。部分は、任意の好適な基準および／または当該分野で公知のもしくは本明細書中に記載される方法を用いて、選択され得、フィルタリングされ得、および／または考慮すべきものから除去され得る。部分をフィルタリングするために使用される基準の非限定的な例としては、冗長なデータ（例えば、冗長なまたはオーバーラップするマッピングされたリード）、情報のないデータ（例えば、マッピングされたカウントがゼロである参照ゲノムの部分）、過剰提示または過小提示された配列を含む参照ゲノムの部分、GC含有量、ノイズの多いデータ、マッピング精度、カウント、カウントのばらつき、リード密度、リード密度のばらつき、不確定性の尺度、反復性の尺度などまたは前述のものの組み合わせが挙げられる。部分は、カウントの分布および／またはリード密度の分布に従ってフィルタリングされるときがある。いくつかの実施形態において、部分は、カウントおよび／またはリード密度の分布に従ってフィルタリングされ、ここで、それらのカウントおよび／またはリード密度は、1つまたはそれを超える参照サンプルから得られる。1つまたはそれを超える参照サンプルは、本明細書中で訓練セットと称されるときがある。いくつかの実施形態において、部分は、カウントおよび／またはリード密度の分布に従ってフィルタリングされ、ここで、それらのカウントおよび／またはリード密度は、1つまたはそれを超えるテストサンプルから得られる。いくつかの実施形態において、部分は、リード密度分布に対する不確定性の尺度に従ってフィルタリングされる。ある特定の実施形態において、リード密度において大きな偏差を示す部分が、フィルタリングプロセスによって除去される。例えば、リード密度の分布（例えば、リード密度の平均の平均値または中央値の分布）が決定され得、ここで、その分布における各リード密度は、同じ部分にマップする。不確定性の尺度（例えば、MAD）は、複数のサンプルに対するリード密度の分布を比較することによって決定され得、ここで、ゲノムの各部分は、不確定性の尺度に関連する。前述の例によると、部分は、各部分に関連する不確定性の尺度（例えば、標準偏差（SD）、MAD）および所定のしきい値に従ってフィルタリングされ得る。所定のしきい値は、ある範囲の許容され得るMAD値を囲む破線の鉛直線によって指摘される。ある特定の場合において、許容され得る範囲内のMAD値を含む部分が保持され、許容され得る範囲外のMAD値を含む部分が、フィルタ

リングプロセスによって考慮すべきものから除去される。いくつかの実施形態において、前述の例によると、所定の不確定性の尺度の外側のリード密度値（例えば、リード密度の中央値、平均または平均値）を含む部分は、フィルタリングプロセスによって考慮すべきものから除去されることが多い。いくつかの実施形態において、ある分布の四分位範囲の外側のリード密度値（例えば、リード密度の中央値、平均または平均値）を含む部分が、フィルタリングプロセスによって考慮すべきものから除去される。いくつかの実施形態において、ある分布の四分位範囲の2倍、3倍、4倍または5倍超外側のリード密度値を含む部分が、フィルタリングプロセスによって考慮すべきものから除去される。いくつかの実施形態において、2シグマ、3シグマ、4シグマ、5シグマ、6シグマ、7シグマまたは8シグマ超外側のリード密度値を含む部分（例えば、シグマは標準偏差によって定義される範囲である）が、フィルタリングプロセスによって考慮すべきものから除去される。

10

【0219】

いくつかの実施形態において、あるシステムは、フィルタリングモジュールを備える。フィルタリングモジュールは、別の好適なモジュールからのものであることが多い、部分（例えば、所定のサイズの部分および／または参照ゲノム内のオーバーラップする部分の位置）および部分に関連するリード密度を受け取る、検索する、および／または格納することが多い。いくつかの実施形態において、選択された部分（例えば、フィルタリングされた部分）は、フィルタリングモジュールによって提供される。いくつかの実施形態において、フィルタリングモジュールは、フィルタリングされた部分を提供するためおよび／または考慮すべきものから部分を除くために、必要とされる。ある特定の実施形態において、フィルタリングモジュールは、考慮すべきものからリード密度を除く、ここで、リード密度は、除去される部分に関連する。フィルタリングモジュールは、選択された部分（例えば、フィルタリングされた部分）を別の好適なモジュールに提供することが多い。

20

バイアス推定値

【0220】

配列決定技術は、複数のバイアス供給源に対して脆弱である。配列決定バイアスは、局所バイアス（例えば、局所ゲノムバイアス）であるときがある。局所バイアスは、配列リードのレベルにおいて明らかにされることが多い。局所ゲノムバイアスは、任意の好適な局所バイアスであり得る。局所バイアスの非限定的な例としては、配列バイアス（例えば、GCバイアス、ATバイアスなど）、DNase Iの感度と相関するバイアス、エントロピー、反復配列バイアス、クロマチン構造バイアス、ポリメラーゼエラー率バイアス、パンドロームバイアス、逆方向反復バイアス、PCR関連バイアスなどまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、局所バイアスの供給源は、明らかにされないか、または既知でない。

30

【0221】

いくつかの実施形態において、局所ゲノムバイアス推定値が決定される。局所ゲノムバイアス推定値は、本明細書中で局所ゲノムバイアス推定と称されるときがある。局所ゲノムバイアス推定値は、参照ゲノム、そのセグメントまたは部分に対して決定され得る。いくつかの実施形態において、局所ゲノムバイアス推定値は、1つまたはそれを超える配列リード（例えば、サンプルのいくつかまたはすべての配列リード）に対して決定される。局所ゲノムバイアス推定値は、参照（例えば、参照ゲノム）の対応する場所および／または位置に対する局所ゲノムバイアス推定に従って配列リードに対して決定されることが多い。いくつかの実施形態において、局所ゲノムバイアス推定値は、配列（例えば、配列リード、参照ゲノムの配列）のバイアスの定量的尺度を含む。局所ゲノムバイアス推定は、好適な方法または数学的プロセスによって決定され得る。いくつかの実施形態において、局所ゲノムバイアス推定値は、好適な分布および／または好適な分布関数（例えば、PDF）によって決定される。いくつかの実施形態において、局所ゲノムバイアス推定値は、PDFの定量的提示を含む。いくつかの実施形態において、局所ゲノムバイアス推定値（例えば、確率密度推定（PDE）、カーネル密度推定）は、局所バイアス含有量の確率密

40

50

度関数（例えば、PDF、例えば、カーネル密度関数）によって決定される。いくつかの実施形態において、密度推定は、カーネル密度推定を含む。局所ゲノムバイアス推定値は、分布の平均、平均値または中央値として表現されるときがある。局所ゲノムバイアス推定値は、好適な分布の和または積分値（例えば、曲線下面積（AUC）として表現されるときがある。

【0222】

PDF（例えば、カーネル密度関数、例えば、Epanechnikovカーネル密度関数）は、バンド幅変数（例えば、バンド幅）を含むことが多い。バンド幅変数は、PDFを使用するとき確率密度推定（PDE）が導かれるウィンドウのサイズおよび／または長さのことを定義することが多い。PDEが導かれるウィンドウは、規定の長さのポリヌクレオチドを含むことが多い。いくつかの実施形態において、PDEが導かれるウィンドウは、部分である。部分（例えば、部分のサイズ、部分の長さ）は、バンド幅変数に従って決定されることが多い。バンド幅変数は、局所ゲノムバイアス推定値；局所ゲノムバイアス推定値が決定されるポリヌクレオチドセグメントの長さ（例えば、ヌクレオチド塩基の連続したセグメント）を決定するために使用されるウィンドウの長さまたはサイズを決定する。PDE（例えば、リード密度、局所ゲノムバイアス推定値（例えば、GC密度））は、任意の好適なバンド幅を用いて決定され得、そのバンド幅の非限定的な例としては、約5塩基～約100,000塩基、約5塩基～約50,000塩基、約5塩基～約25,000塩基、約5塩基～約10,000塩基、約5塩基～約5,000塩基、約5塩基～約2,500塩基、約5塩基～約1,000塩基、約5塩基～約500塩基、約5塩基～約250塩基、約20塩基～約250塩基などのバンド幅が挙げられる。いくつかの実施形態において、局所ゲノムバイアス推定値（例えば、GC密度）は、約400塩基もしくはそれ未満、約350塩基もしくはそれ未満、約300塩基もしくはそれ未満、約250塩基もしくはそれ未満、約225塩基もしくはそれ未満、約200塩基もしくはそれ未満、約175塩基もしくはそれ未満、約150塩基もしくはそれ未満、約125塩基もしくはそれ未満、約100塩基もしくはそれ未満、約75塩基もしくはそれ未満、約50塩基もしくはそれ未満、または約25塩基もしくはそれ未満のバンド幅を用いて決定される。ある特定の実施形態において、局所ゲノムバイアス推定値（例えば、GC密度）は、所与の被験体および／またはサンプルに対して得られる配列リードのリード長の平均、平均値、中央値または最大値に従って決定されたバンド幅を用いて決定される。局所ゲノムバイアス推定値（例えば、GC密度）は、所与の被験体および／またはサンプルに対して得られる配列リードのリード長の平均、平均値、中央値または最大値にほぼ等しいバンド幅を用いて決定されるときがある。いくつかの実施形態において、局所ゲノムバイアス推定値（例えば、GC密度）は、約250、240、230、220、210、200、190、180、160、150、140、130、120、110、100、90、80、70、60、50、40、30、20または約10塩基のバンド幅を用いて決定される。

【0223】

局所ゲノムバイアス推定値は、一塩基の分解能において決定され得るが、局所ゲノムバイアス推定値（例えば、局所GC含有量）は、それより低い分解能において決定され得る。いくつかの実施形態において、局所ゲノムバイアス推定値は、ある局所バイアス含有量に対して決定される。局所ゲノムバイアス推定値（例えば、PDFを用いて決定されるとき）は、あるウィンドウを用いて決定されることが多い。いくつかの実施形態において、局所ゲノムバイアス推定値は、予め選択された数の塩基を含むウィンドウの使用を含む。ウィンドウは、連続した塩基のセグメントを含むときがある。ウィンドウは、連続していない塩基の1つまたはそれを超える部分を含むときがある。ウィンドウは、1つまたはそれを超える部分（例えば、ゲノムの部分）を含むときがある。ウィンドウのサイズまたは長さは、バンド幅によって、かつPDFに従って、決定されることが多い。いくつかの実施形態において、ウィンドウは、バンド幅の長さの約10倍もしくはそれを超えるか、8倍もしくはそれを超えるか、7倍もしくはそれを超えるか、6倍もしくはそれを超えるか、5倍もしくはそれを超えるか、4倍もしくはそれを超えるか、3倍もしくはそれを超え

るか、または約2倍もしくはそれを超える。PDF（例えば、カーネル密度関数）を使用して密度推定値が決定されるとき、ウィンドウは、選択されたバンド幅の長さの2倍であるときがある。ウィンドウは、任意の好適な数の塩基を含み得る。いくつかの実施形態において、ウィンドウは、約5塩基～約100,000塩基、約5塩基～約50,000塩基、約5塩基～約25,000塩基、約5塩基～約10,000塩基、約5塩基～約5,000塩基、約5塩基～約2,500塩基、約5塩基～約1,000塩基、約5塩基～約500塩基、約5塩基～約250塩基または約20塩基～約250塩基を含む。いくつかの実施形態において、ゲノムまたはそのセグメントは、複数のウィンドウに分割される。ゲノムの領域を包含するウィンドウは、オーバーラップしてもよいし、しなくてもよい。いくつかの実施形態において、ウィンドウは、互いから等しい距離に位置される。いくつかの実施形態において、ウィンドウは、互いから異なる距離に位置される。ある特定の実施形態において、ゲノムまたはそのセグメントは、複数のスライディングウィンドウに分割され、ここで、あるウィンドウは、ゲノムまたはそのセグメントにわたって徐々に伸ばしてスライドされ、各刻み幅における各ウィンドウが、局所ゲノムバイアス推定値（例えば、局所GC密度）を含む。ウィンドウは、任意の数値パターンに従って、または任意の非主題的な規定の配列に従って、ゲノムにわたって任意の好適な刻み幅でスライドされ得る。いくつかの実施形態において、局所ゲノムバイアス推定値の決定に向けて、ウィンドウは、ゲノムまたはそのセグメントにわたって、約10,000bpもしくはそれを超えるか、約5,000bpもしくはそれを超えるか、約2,500bpもしくはそれを超えるか、約1,000bpもしくはそれを超えるか、約750bpもしくはそれを超えるか、約500bpもしくはそれを超えるか、約400塩基もしくはそれを超えるか、約250bpもしくはそれを超えるか、約100bpもしくはそれを超えるか、約50bpもしくはそれを超えるか、または約25bpもしくはそれを超える塩基刻みでスライドされる。いくつかの実施形態において、局所ゲノムバイアス推定値の決定に向けて、ウィンドウは、ゲノムまたはそのセグメントにわたって、約25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3、2または約1bp刻みでスライドされる。例えば、局所ゲノムバイアス推定値の決定に向けて、ウィンドウは、約400bp（例えば、200bpのバンド幅）を含み得、ゲノムにわたって1bp刻みでスライドされ得る。いくつかの実施形態において、局所ゲノムバイアス推定値は、カーネル密度関数および約200bpのバンド幅を用いて、ゲノムまたはそのセグメントにおける各塩基に対して決定される。

【0224】

いくつかの実施形態において、局所ゲノムバイアス推定値は、局所GC含有量および／または局所GC含有量の提示である。用語「局所」は、本明細書中で使用されるとき（例えば、局所バイアス、局所バイアス推定値、局所バイアス含有量、局所ゲノムバイアス、局所GC含有量などを記載するために使用されるとき）、10,000bpまたはそれ未満のポリヌクレオチドセグメントのことを指す。いくつかの実施形態において、用語「局所」とは、5000bpもしくはそれ未満、4000bpもしくはそれ未満、3000bpもしくはそれ未満、2000bpもしくはそれ未満、1000bpもしくはそれ未満、500bpもしくはそれ未満、250bpもしくはそれ未満、200bpもしくはそれ未満、175bpもしくはそれ未満、150bpもしくはそれ未満、100bpもしくはそれ未満、75bpもしくはそれ未満、または50bpもしくはそれ未満のポリヌクレオチドセグメントのことを指す。局所GC含有量は、ゲノム、配列リード、配列リードアセンブリ（例えば、コンティグ、プロファイルなど）の局所セグメントに対するGC含有量の提示（例えば、数学的、定量的提示）であることが多い。例えば、局所GC含有量は、局所GCバイアス推定値またはGC密度であり得る。

【0225】

1つまたはそれを超えるGC密度は、参照またはサンプル（例えば、テストサンプル）のポリヌクレオチドに対して決定されることが多い。いくつかの実施形態において、GC密度は、局所GC含有量（例えば、5000bpまたはそれ未満のポリヌクレオチドセグ

10

20

30

40

50

メントに対する)の提示(例えば、数学的、定量的提示)である。いくつかの実施形態において、GC密度は、局所ゲノムバイアス推定値である。GC密度は、本明細書中に記載されるおよび／または当該分野で公知の好適なプロセスを用いて決定され得る。GC密度は、好適なPDF(例えば、カーネル密度関数(例えば、Epanechnikovカーネル密度関数)を用いて決定され得る。いくつかの実施形態において、GC密度は、PDE(例えば、カーネル密度推定)である。ある特定の実施形態において、GC密度は、1つまたはそれを超えるグアニン(G)および／またはシトシン(C)ヌクレオチドの存在もしくは非存在によって定義される。逆に、いくつかの実施形態において、GC密度は、1つまたはそれを超えるアデニン(A)および／またはチミジン(T)ヌクレオチドの存在もしくは非存在によって定義され得る。局所GC含有量に対するGC密度は、いくつかの実施形態において、ゲノム全体またはそのセグメント(例えば、常染色体、染色体のセット、単一の染色体、遺伝子)に対して決定されたGC密度に従って正規化される。1つまたはそれを超えるGC密度は、サンプル(例えば、テストサンプル)または参照サンプルのポリヌクレオチドに対して決定され得る。GC密度は、参照ゲノムに対して決定されることが多い。いくつかの実施形態において、GC密度は、参照ゲノムに従って配列リードに対して決定される。リードのGC密度は、リードがマッピングされる参照ゲノムの対応する場所および／または位置に対して決定されたGC密度に従って決定されることが多い。いくつかの実施形態において、参照ゲノム上の場所に対して決定されたGC密度は、あるリードに対して割り当てられ、および／または提供され、ここで、そのリードまたはそのセグメントは、その参照ゲノム上の同じ場所にマップする。任意の好適な方法を用いることにより、リードに対するGC密度を生成する目的で、マッピングされたリードの参照ゲノム上の場所を決定することができる。いくつかの実施形態において、マッピングされたリードの中央の位置(median position)は、そのリードに対するGC密度が決定された参照ゲノム上の場所を決定する。例えば、あるリードの中央の位置が、12番染色体の、参照ゲノムの塩基番号xにマップする場合、そのリードのGC密度は、12番染色体の、参照ゲノムの塩基番号xまたは塩基番号x付近に位置づけられる位置に対してカーネル密度推定によって決定されたGC密度として提供されることが多い。いくつかの実施形態において、GC密度は、参照ゲノムに従って、リードのいくつかまたはすべての塩基位置に対して決定される。あるリードのGC密度は、参照ゲノム上の複数の塩基位置に対して決定された2つまたはそれを超えるGC密度の平均、和、中央値または積分値を含むときがある。

【0226】

いくつかの実施形態において、局所ゲノムバイアス推定(例えば、GC密度)は、定量され、および／またはある値として提供される。局所ゲノムバイアス推定(例えば、GC密度)は、平均、平均値および／または中央値として表現されるときがある。局所ゲノムバイアス推定(例えば、GC密度)は、PDEの最大ピーク高さとして表現されるときがある。局所ゲノムバイアス推定(例えば、GC密度)は、好適なPDEの和または積分値(例えば、曲線下面積(AUC))として表現されるときがある。いくつかの実施形態において、GC密度は、カーネル重みを含む。ある特定の実施形態において、あるリードに対するGC密度は、カーネル重みの平均、平均値、和、中央値、最大ピーク高さまたは積分値にほぼ等しい値を含む。

バイアス頻度

【0227】

バイアス頻度は、1つまたはそれを超える局所ゲノムバイアス推定値(例えば、GC密度)に従って決定されるときがある。バイアス頻度は、サンプル、参照(例えば、参照ゲノム、参照配列)またはその一部に対する局所ゲノムバイアス推定値の出現数のカウントまたは和であるときがある。バイアス頻度は、サンプル、参照またはその一部に対する局所ゲノムバイアス推定値(例えば、各局所ゲノムバイアス推定値)の出現数のカウントまたは和であるときがある。いくつかの実施形態において、バイアス頻度は、GC密度の頻度である。GC密度の頻度は、1つまたはそれを超えるGC密度に従って決定されること

が多い。例えば、G C 密度の頻度は、値 x の G C 密度がゲノム全体またはそのセグメントにわたって提示される回数に相当し得る。バイアス頻度は、局所ゲノムバイアス推定値の分布であることが多く、ここで、各局所ゲノムバイアス推定値の出現数は、バイアス頻度として表される。バイアス頻度は、数学的に操作されるおよび／または正規化されるときがある。バイアス頻度は、好適な方法によって数学的に操作され得るおよび／または正規化され得る。いくつかの実施形態において、バイアス頻度は、サンプル、参照またはその一部（例えば、常染色体、染色体のサブセット、単一の染色体またはそのリード）に対する各局所ゲノムバイアス推定値の提示（例えば、割合、パーセンテージ）に従って正規化される。バイアス頻度は、サンプルまたは参照のいくつかまたはすべての局所ゲノムバイアス推定値に対して決定され得る。いくつかの実施形態において、バイアス頻度は、テストサンプルのいくつかまたはすべての配列リードに対する局所ゲノムバイアス推定値に対して決定され得る。

10

【0228】

いくつかの実施形態において、あるシステムは、バイアス密度モジュール6を備える。バイアス密度モジュールは、マッピングされた配列リード5および参照配列2を任意の好適な形式で受け取り得、検索し得、および／または格納し得、局所ゲノムバイアス推定値、局所ゲノムバイアス分布、バイアス頻度、G C 密度、G C 密度分布および／またはG C 密度頻度（集合的に、ボックス7によって表される）を生成し得る。いくつかの実施形態において、バイアス密度モジュールは、データおよび／または情報（例えば、7）を別の好適なモジュール（例えば、関係モジュール8）に移す。

20

バイアスの関係

【0229】

いくつかの実施形態において、1つまたはそれを超える関係が、局所ゲノムバイアス推定値とバイアス頻度との間に生成される。用語「関係」は、本明細書中で使用されるとき、2つまたはそれを超える変数または値の間の数学的および／またはグラフ的な関係のことを指す。ある関係は、好適な数学的プロセスおよび／またはグラフ的プロセスによって生成され得る。関係の非限定的な例としては、関数、相関、分布、線形方程式または非線形方程式、線、回帰、適合回帰（*fitted regression*）などまたはそれらの組み合わせの数学的表示および／またはグラフ表示が挙げられる。関係は、当てはめ関係を含むときがある。いくつかの実施形態において、当てはめ関係は、適合回帰を含む。関係は、重み付けされた2つまたはそれを超える変数または値を含むときがある。いくつかの実施形態において、ある関係は、その関係の1つまたはそれを超える変数または値が重み付けされた適合回帰を含む。回帰は、重み付けされる形式で当てはめられるときがある。回帰は、重み付けなしで当てはめられるときがある。ある特定の実施形態において、関係を生成することは、プロットすることまたはグラフで示すことを含む。

30

【0230】

いくつかの実施形態において、好適な関係が、局所ゲノムバイアス推定値とバイアス頻度との間において決定される。いくつかの実施形態において、サンプルに対する（i）局所ゲノムバイアス推定値と（ii）バイアス頻度との間の関係を生成することにより、サンプルバイアス関係が提供される。いくつかの実施形態において、参照に対する（i）局所ゲノムバイアス推定値と（ii）バイアス頻度との間に関係を生成することにより、参照バイアス関係が提供される。ある特定の実施形態において、G C 密度とG C 密度頻度との間に関係が生成される。いくつかの実施形態において、サンプルに対する（i）G C 密度と（ii）G C 密度頻度との間に関係を生成することにより、サンプルG C 密度関係が提供される。いくつかの実施形態において、参照に対する（i）G C 密度と（ii）G C 密度頻度との間の関係を生成することにより、参照G C 密度関係が提供される。いくつかの実施形態において、局所ゲノムバイアス推定値がG C 密度である場合、サンプルバイアス関係は、サンプルG C 密度関係であり、参照バイアス関係は、参照G C 密度関係である。参照G C 密度関係および／またはサンプルG C 密度関係のG C 密度は、局所G C 含有量の提示（例えば、数学的提示または定量的提示）であることが多い。いくつかの実施形態

40

50

において、局所ゲノムバイアス推定値とバイアス頻度との間の関係は、分布を含む。いくつかの実施形態において、局所ゲノムバイアス推定値とバイアス頻度との間の関係は、当てはめ関係（例えば、適合回帰）を含む。いくつかの実施形態において、局所ゲノムバイアス推定値とバイアス頻度との間の関係は、適合された線形または非線形回帰（例えば、多項式回帰）を含む。ある特定の実施形態において、局所ゲノムバイアス推定値および／またはバイアス頻度が好適なプロセスによって重み付けされる場合、局所ゲノムバイアス推定値とバイアス頻度との間の関係は、重み付けされた関係を含む。いくつかの実施形態において、重み付けされた当てはめ関係（例えば、重み付けされた当てはめ）は、内挿を用いた、分位点回帰、パラメータ化された分布または経験的な分布を含むプロセスによって得ることができる。ある特定の実施形態において、局所ゲノムバイアス推定値が重み付けされる場合、テストサンプル、参照またはその一部に対する局所ゲノムバイアス推定値とバイアス頻度との間の関係は、多項式回帰を含む。いくつかの実施形態において、重み付けされた当てはめモデルは、分布の値の重み付けを含む。分布の値は、好適なプロセスによって重み付けされ得る。いくつかの実施形態において、分布の裾の近くに位置する値には、分布の中央値により近い値よりも小さい重みが提供される。例えば、局所ゲノムバイアス推定値（例えば、GC密度）とバイアス頻度（例えば、GC密度頻度）との間の分布の場合、所与の局所ゲノムバイアス推定値に対するバイアス頻度に従って重みが決定され、ここで、分布の平均値により近いバイアス頻度を含む局所ゲノムバイアス推定値には、その平均値からより遠くのバイアス頻度を含む局所ゲノムバイアス推定値よりも大きい重みが提供される。

10

20

【0231】

いくつかの実施形態において、あるシステムは、関係モジュール8を備える。関係モジュールは、関係、ならびに関係を定義する関数、係数、定数および変数を生成し得る。関係モジュールは、データおよび／または情報（例えば、7）を好適なモジュール（例えば、バイアス密度モジュール6）から受け取り得、格納し得、および／または検索し得、関係を生成し得る。関係モジュールは、局所ゲノムバイアス推定値の分布を生成し、比較することが多い。関係モジュールは、データセットを比較することができ、回帰および／または当てはめ関係を生成するときがある。いくつかの実施形態において、関係モジュールは、1つまたはそれを超える分布（例えば、サンプルおよび／または参照の局所ゲノムバイアス推定値の分布）を比較し、配列リードのカウントに対する重み付け因子および／または重み付け割当9を別の好適なモジュール（例えば、バイアス補正モジュール）に提供する。関係モジュールは、配列リードの正規化されたカウントを、それらのカウントが関係および／または比較に従って正規化される分布モジュール21に直接提供するときがある。

30

比較結果の生成およびその使用

【0232】

いくつかの実施形態において、配列リードにおける局所バイアスを減少させるためのプロセスは、配列リードのカウントを正規化することを含む。配列リードのカウントは、テストサンプルと参照との比較結果に従って正規化されることが多い。例えば、配列リードのカウントは、テストサンプルの配列リードの局所ゲノムバイアス推定値を、参照（例えば、参照ゲノムまたはその一部）の局所ゲノムバイアス推定値と比較することによって正規化されるときがある。いくつかの実施形態において、配列リードのカウントは、テストサンプルの局所ゲノムバイアス推定値のバイアス頻度を、参照の局所ゲノムバイアス推定値のバイアス頻度と比較することによって正規化される。いくつかの実施形態において、配列リードのカウントは、サンプルバイアス関係と参照バイアス関係とを比較することによって正規化され、それによって、比較結果が生成される。

40

【0233】

配列リードのカウントは、2つまたはそれを超える関係の比較結果に従って正規化されることが多い。ある特定の実施形態において、2つまたはそれを超える関係が比較され、それによって、配列リードにおける局所バイアスを減少させる（例えば、カウントを正規

50

化する) ために使用される比較結果が提供される。2つまたはそれを超える関係は、好適な方法によって比較され得る。いくつかの実施形態において、比較結果は、第1の関係と第2の関係との加算、減算、乗算および／または除算を含む。ある特定の実施形態において、2つまたはそれを超える関係の比較は、好適な線形回帰および／または非線形回帰の使用を含む。ある特定の実施形態において、2つまたはそれを超える関係の比較は、好適な多項式回帰(例えば、3次多項式回帰)を含む。いくつかの実施形態において、比較結果は、第1の回帰と第2の回帰との加算、減算、乗算および／または除算を含む。いくつかの実施形態において、2つまたはそれを超える関係は、複数の回帰の推論フレームワークを含むプロセスによって比較される。いくつかの実施形態において、2つまたはそれを超える関係は、好適な多変量解析を含むプロセスによって比較される。いくつかの実施形態において、2つまたはそれを超える関係は、基底関数(例えば、ブレンディング関数、例えば、多項式基底、フーリエ基底など)、スプライン、放射基底関数および／またはウェーブレットを含むプロセスによって比較される。

10

【0234】

ある特定の実施形態において、テストサンプルおよび参照に対するバイアス頻度を含む局所ゲノムバイアス推定値の分布は、局所ゲノムバイアス推定値が重み付けされる多項式回帰を含むプロセスによって比較される。いくつかの実施形態において、多項式回帰は、(i) 比率(この比率の各々は、参照の局所ゲノムバイアス推定値のバイアス頻度およびサンプルの局所ゲノムバイアス推定値のバイアス頻度を含む)と(ii) 局所ゲノムバイアス推定値との間において生成される。いくつかの実施形態において、多項式回帰は、(i) 参照の局所ゲノムバイアス推定値のバイアス頻度と、サンプルの局所ゲノムバイアス推定値のバイアス頻度との比率と、(ii) 局所ゲノムバイアス推定値との間において生成される。いくつかの実施形態において、テストサンプルおよび参照のリードに対する局所ゲノムバイアス推定値の分布の比較は、参照およびサンプルに対する局所ゲノムバイアス推定値のバイアス頻度のlog比率(例えば、log2比率)を測定することを含む。いくつかの実施形態において、局所ゲノムバイアス推定値の分布の比較は、参照に対する局所ゲノムバイアス推定値のバイアス頻度のlog比率(例えば、log2比率)を、サンプルに対する局所ゲノムバイアス推定値のバイアス頻度のlog比率(例えば、log2比率)で除算することを含む。

20

【0235】

比較結果に従ってカウントを正規化することは、代表的には、いくつかのカウントを調整し、他のものを調整しない。カウントの正規化は、すべてのカウントを調整するときがあり、配列リードのいかなるカウントも調整しないときがある。配列リードに対するカウントは、重み付け因子を決定することを含むプロセスによって正規化されるときがあり、そのプロセスは、重み付け因子を直接生成し、使用することを含まないときがある。比較結果に従ってカウントを正規化することは、配列リードの各カウントに対して重み付け因子を決定することを含むときがある。重み付け因子は、配列リードに特異的であることが多く、特異的な配列リードのカウントに適用される。重み付け因子は、2つまたはそれを超えるバイアス関係の比較結果(例えば、参照バイアス関係と比較されるサンプルバイアス関係)に従って決定されることが多い。正規化されたカウントは、重み付け因子に従ってカウント値を調整することによって決定されることが多い。重み付け因子に従ってカウントを調整することは、配列リードに対するカウントを加算すること、減算すること、乗算すること、および／または配列リードに対するカウントを重み付け因子で除算することを含むときがある。重み付け因子および／または正規化されたカウントは、回帰(例えば、回帰直線)から決定されるときがある。正規化されたカウントは、参照(例えば、参照ゲノム)の局所ゲノムバイアス推定値のバイアス頻度とテストサンプルの局所ゲノムバイアス推定値のバイアス頻度との比較結果から生じる回帰直線(例えば、当てはめられた回帰直線)から直接得られるときがある。いくつかの実施形態において、サンプルのリードの各カウントには、(ii) 参照の局所ゲノムバイアス推定値のバイアス頻度と比較された、(i) リードの局所ゲノムバイアス推定値のバイアス頻度の比較結果に従って、正規

30

40

50

化されたカウント値が提供される。ある特定の実施形態において、サンプルに対して得られた配列リードのカウントが正規化され、それらの配列リードにおけるバイアスが減少する。

【0236】

あるシステムは、バイアス補正モジュール10を備えるときがある。いくつかの実施形態において、バイアス補正モジュールの機能は、関係モデリングモジュール8によって行われる。バイアス補正モジュールは、マッピングされた配列リードおよび重み付け因子（例えば、9）を、好適なモジュール（例えば、関係モジュール8、圧縮モジュール4）から受け取り得、検索し得、および／または格納し得る。いくつかの実施形態において、バイアス補正モジュールは、マッピングされたリードにカウントを提供する。いくつかの実施形態において、バイアス補正モジュールは、重み付け割当および／またはバイアス補正因子を配列リードのカウントに適用し、それにより、正規化されたおよび／または調整されたカウントが提供される。バイアス補正モジュールは、正規化されたカウントを別の好適なモジュール（例えば、分布モジュール21）に提供することが多い。

【0237】

ある特定の実施形態において、カウントの正規化は、GC密度に加えて1つまたはそれを超える特徴を考慮すること（factoring）、および配列リードのカウントを正規化することを含む。ある特定の実施形態において、カウントの正規化は、1つまたはそれを超える異なる局所ゲノムバイアス推定値を考慮すること、および配列リードのカウントを正規化することを含む。ある特定の実施形態において、配列リードのカウントは、1つまたはそれを超える特徴（例えば、1つまたはそれを超えるバイアス）に従って決定された重み付けに従って重み付けされる。いくつかの実施形態において、カウントは、1つまたはそれを超える組み合された重みに従って正規化される。1つまたはそれを超える組み合された重みに従って、1つまたはそれを超える特徴を考慮することおよび／またはカウントを正規化することは、多変量モデルの使用を含むプロセスによって行われるときがある。任意の好適な多変量モデルを使用することにより、カウントを正規化することができる。多変量モデルの非限定的な例としては、多変量線形回帰、多変量分位点回帰、経験的データの多変量補間、非線形多変量モデルなどまたはそれらの組み合わせが挙げられる。

【0238】

いくつかの実施形態において、あるシステムは、多変量補正モジュール13を備える。多変量補正モジュールは、バイアス密度モジュール6、関係モジュール8および／またはバイアス補正モジュール10の機能を複数回行い、それにより、複数のバイアスに対してカウントが調整され得る。いくつかの実施形態において、多変量補正モジュールは、1つまたはそれを超えるバイアス密度モジュール6、関係モジュール8および／またはバイアス補正モジュール10を含む。多変量補正モジュールは、正規化されたカウント11を別の好適なモジュール（例えば、分布モジュール21）に提供するときがある。

部分の重み付け

【0239】

いくつかの実施形態において、部分は、重み付けされる。いくつかの実施形態において、1つまたはそれを超える部分が重み付けされ、それにより、重み付けされた部分が提供される。部分の重み付けによって、部分依存性が排除されるときがある。部分は、好適なプロセスによって重み付けされ得る。いくつかの実施形態において、1つまたはそれを超える部分は、固有の関数（例えば、固有関数）によって重み付けされる。いくつかの実施形態において、ある固有の関数は、部分を直交固有部分で置き換えることを含む。いくつかの実施形態において、あるシステムは、部分重み付けモジュール42を備える。いくつかの実施形態において、重み付けモジュールは、リード密度、リード密度プロファイルおよび／または調整されたリード密度プロファイルを受け取る、検索する、および／または格納する。いくつかの実施形態において、重み付けされた部分は、部分重み付けモジュールによって提供される。いくつかの実施形態において、重み付けモジュールは、部分を重

みづけするために必要とされる。重み付けモジュールは、当該分野で公知のまたは本明細書中に記載される1つまたはそれを超える重み付け方法によって部分を重み付けし得る。重み付けモジュールは、重み付けされた部分を別の好適なモジュール（例えば、スコアリングモジュール46、PCA統計モジュール33、プロファイル生成モジュール26など）に提供することが多い。

主成分分析

【0240】

いくつかの実施形態において、リード密度プロファイル（例えば、テストサンプルのリード密度プロファイル）は、主成分分析（PCA）に従って調整される。1つまたはそれを超える参照サンプルのリード密度プロファイルおよび／または試験被験体のリード密度プロファイルは、PCAに従って調整され得る。PCA関連プロセスによってリード密度プロファイルからバイアスを除去することは、本明細書中で、プロファイルの調整と称されるときがある。PCAは、好適なPCA方法またはその変法によって行われ得る。PCA方法の非限定的な例としては、正準相関分析（CCA）、Karhunen-Loeve変換（KLT）、Hotelling変換、固有直交分解（POD）、Xの特異値分解（SVD）、XTXの固有値分解（EVD）、因子分析、Eckart-Young定理、Schmidt-Mirsky定理、経験的直交関数（EOF）、経験固有関数分解、経験的成分分析、準調和モード、スペクトル分解、経験的モーダル解析など、それらの変法または組み合わせが挙げられる。PCAは、リード密度プロファイルにおける1つまたはそれを超えるバイアスを特定することが多い。PCAによって特定されたバイアスは、本明細書中で主成分と称されるときがある。いくつかの実施形態において、1つまたはそれを超えるバイアスは、好適な方法を用いて1つまたはそれを超える主成分に従ってリード密度プロファイルを調整することによって除去され得る。リード密度プロファイルは、1つまたはそれを超える主成分を加算すること、1つまたはそれを超える主成分をリード密度プロファイルから減算すること、1つまたはそれを超える主成分を乗算すること、および／または1つまたはそれを超える主成分をリード密度プロファイルで除算することによって、調整され得る。いくつかの実施形態において、1つまたはそれを超えるバイアスは、1つまたはそれを超える主成分をリード密度プロファイルから減算することによって、リード密度プロファイルから除去され得る。リード密度プロファイルにおけるバイアスは、プロファイルのPCAによって特定および／または定量されることが多いが、主成分は、リード密度のレベルにおいてプロファイルから減算されることが多い。PCAは、1つまたはそれを超える主成分を特定することが多い。いくつかの実施形態において、PCAは、第1、第2、第3、第4、第5、第6、第7、第8、第9および第10またはそれを超える主成分を特定する。ある特定の実施形態において、1、2、3、4、5、6、7、8、9、10個もしくはそれを超える主成分が、プロファイルを調整するために使用される。主成分は、PCAにおけるそれらの（there）出現順にプロファイルを調整するために使用されることが多い。例えば、3つの主成分が、リード密度プロファイルから減算される場合、第1、第2および第3主成分が使用される。主成分によって特定されたバイアスは、プロファイルを調整するために使用されないプロファイルの特徴を含むときがある。例えば、PCAは、主成分としてコピー数変異（例えば、異数性、微小重複、微小欠失、欠失、転座、挿入）および／または性差を特定し得る。したがって、いくつかの実施形態において、1つまたはそれを超える主成分は、プロファイルを調整するために使用されない。例えば、第3主成分がプロファイルを調整するために使用されない場合、プロファイルを調整するために、第1、第2および第4主成分が使用されるときがある。主成分は、任意の好適なサンプルまたは参照を用いるPCAから得ることができる。いくつかの実施形態において、主成分は、テストサンプル（例えば、試験被験体）から得られる。いくつかの実施形態において、主成分は、1つまたはそれを超える参照（例えば、参照サンプル、参照配列、参照セット）から得られる。ある特定の場合一つにおいて、PCAは、複数のサンプルを含む訓練セットから得られるリード密度プロファイルの中央値に対して行われ、第1主成分および第2主成分が特定される。いくつかの実施形態において、主成

10

20

30

40

50

分は、対象のコピー数変異を欠くと知られている被験体のセットから得られる。いくつかの実施形態において、主成分は、既知の正倍数体セットから得られる。主成分は、参照（例えば、訓練セット）の1つまたはそれを超えるリード密度プロファイルを用いて行われるPCAに従って特定されることが多い。参照から得られた1つまたはそれを超える主成分は、試験被験体のリード密度プロファイルから減算され、それにより、調整されたプロファイルが提供されることが多い。

【0241】

いくつかの実施形態において、あるシステムは、PCA統計モジュール33を備える。PCA統計モジュールは、別の好適なモジュール（例えば、プロファイル生成モジュール26）からリード密度プロファイルを受け取り得、および／または検索し得る。PCAは、PCA統計モジュールによって行われることが多い。PCA統計モジュールは、参照セット32、訓練セット30および／または1つもしくはそれを超える試験被験体28から、リード密度プロファイルを受け取り、検索し、および／または格納し、リード密度プロファイル进行处理することが多い。PCA統計モジュールは、主成分を生成および／もしくは提供し得、ならびに／または1つもしくはそれを超える主成分に従ってリード密度プロファイルを調整し得る。調整されたリード密度プロファイル（例えば、40、38）は、PCA統計モジュールによって提供されることが多い。PCA統計モジュールは、調整されたリード密度プロファイル（例えば、38、40）を別の好適なモジュール（例えば、部分重み付けモジュール42、スコアリングモジュール46）に提供し得るおよび／または移し得る。いくつかの実施形態において、PCA統計モジュールは、性別コール（gender call）36を提供し得る。性別コールは、PCAおよび／または1つもしくはそれを超える主成分に従って判定される胎児の性別の判定であるときがある。いくつかの実施形態において、PCA統計モジュールは、下記に示されるRコードのいくつか、すべてまたは修正したものを備える。主成分を計算するためのRコードは、一般に、データのクリーニング（例えば、中央値の減算、部分のフィルタリングおよび極値のトリミング）から始まる：

【数3】

```
#Clean the data outliers for PCA
dclean <- (dat - m)[mask,]

for (j in 1:ncol(dclean))
{
  q <- quantile(dclean[,j],c(.25,.75))
  qmin <- q[1] - 4*(q[2]-q[1])
  qmax <- q[2] + 4*(q[2]-q[1])
  dclean[dclean[,j] < qmin,j] <- qmin
  dclean[dclean[,j] > qmax,j] <- qmax
}
```

次いで、主成分が計算される：

【数4】

```
#Compute principal components
pc <- prcomp(dclean)$x
```

最後に、PCAによって調整された各サンプルのプロファイルが、以下を用いて計算され得る：

【数 5】

```
#Compute residuals
mm <- model.matrix(~pc[,1:numpc])
for (j in 1:ncol(dclean))
  dclean[,j] <- dclean[,j] - predict(lm(dclean[,j]~mm))
```

プロファイルの比較

【0242】

いくつかの実施形態において、アウトカムの判定は、比較を含む。ある特定の実施形態において、リード密度プロファイルまたはその一部が、アウトカムを提供するために使用される。いくつかの実施形態において、アウトカムの判定（例えば、コピー数変異の存在もしくは非存在の判定）は、2つまたはそれを超えるリード密度プロファイルの比較を含む。リード密度プロファイルの比較は、ゲノムの選択されたセグメントに対して生成されたリード密度プロファイルの比較を含むことが多い。例えば、試験プロファイルおよび参照プロファイルが、実質的に同じセグメントであるゲノム（例えば、参照ゲノム）のセグメントに対して判定される場合、その試験プロファイルは、参照プロファイルと比較されることが多い。リード密度プロファイルの比較は、リード密度プロファイルの部分の2つまたはそれを超えるサブセットの比較を含むときがある。リード密度プロファイルの部分のサブセットは、ゲノムのセグメント（例えば、染色体またはそのセグメント）に相当し得る。リード密度プロファイルは、任意の量の部分のサブセットを含み得る。リード密度プロファイルは、2つもしくはそれを超えるか、3つもしくはそれを超えるか、4つもしくはそれを超えるかまたは5つもしくはそれを超えるサブセットを含むときがある。ある特定の実施形態において、各部分が、隣接する参照ゲノムのセグメントである場合、リード密度プロファイルは、部分の2つのサブセットを含む。いくつかの実施形態において、試験プロファイルと参照プロファイルの両方が、部分の第1のサブセットおよび部分の第2のサブセットを含み、その第1および第2のサブセットが、ゲノムの異なるセグメントである場合、試験プロファイルは、参照プロファイルと比較され得る。リード密度プロファイルの部分のいくつかのサブセットは、コピー数変異を含むこともあるし、部分の他のサブセットは、コピー数変異を実質的に含まないときもある。プロファイル（例えば、試験プロファイル）の部分のすべてのサブセットが、コピー数変異を実質的に含まないときがある。プロファイル（例えば、試験プロファイル）の部分のすべてのサブセットが、コピー数変異を含むときがある。いくつかの実施形態において、試験プロファイルは、遺伝的変異を含む部分の第1のサブセットおよびコピー数変異を実質的に含まない部分の第2のサブセットを含み得る。

【0243】

いくつかの実施形態において、本明細書中に記載される方法は、比較（例えば、試験プロファイルと参照プロファイルとの比較）を行う工程を含む。2つもしくはそれを超えるデータセット、2つもしくはそれを超える関係および／または2つもしくはそれを超えるプロファイルが、好適な方法によって比較され得る。データセット、関係および／またはプロファイルの比較に適した統計学的方法の非限定的な例としては、Behrens-Fisherアプローチ、ブートストラッピング、独立した有意性検定を組み合わせるためのフィッシャーの方法、Neyman-Pearson検定、検証的データ解析、探索的データ解析、正確検定、F検定、Z検定、T検定、不確定性の尺度の算出および／もしくは比較、帰無仮説、対立帰無仮説など、カイ二乗検定、総括的検定、有意水準（例えば、統計的有意水準）の算出および／もしくは比較、メタ解析、多変量解析、回帰、線形単回帰、ロバスト線形回帰など、または前述のものの組み合わせが挙げられる。ある特定の実施形態において、2つまたはそれを超えるデータセット、関係および／またはプロファイルの比較は、不確定性の尺度の測定および／または比較を含む。「不確定性の尺度」は、本明細書中で使用されるとき、有意性（例えば、統計的有意性）の尺度、誤差の尺度、分散の尺度、信頼の尺度などまたはそれらの組み合わせのことを指す。不確定性の尺度は、

値（例えば、しきい値）またはある範囲の値（例えば、区間、信頼区間、ベイズ信頼区間、しきい値の範囲）であり得る。不確定性の尺度の非限定的な例としては、p 値、好適な偏差の尺度（例えば、標準偏差、シグマ、絶対偏差、平均絶対偏差など）、好適な誤差（例えば、標準誤差、平均平方誤差、平方根平均二乗誤差など）の尺度、好適な分散の尺度、好適な標準得点（例えば、標準偏差、累積パーセンテージ、パーセンタイル等価物、Z 得点、T 得点、R 得点、スタンダードナイン（スタナイン）、スタナインにおけるパーセントなど）などまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、有意水準の決定は、不確定性の尺度（例えば、p 値）の決定を含む。ある特定の実施形態において、2 つまたはそれを超えるデータセット、関係および／またはプロファイルは、複数の（例えば、2 つまたはそれを超える）統計学的方法（例えば、最小二乗回帰、主成分分析、線形判別分析、二次判別分析、バギング、ニューラルネットワーク、サポートベクターマシンモデル、ランダムフォレスト、分類木モデル、K 最近隣法、ロジスティック回帰および／または l o s s 平滑化）ならびに／または任意の好適な数学的操作および／もしくは統計的操作（例えば、本明細書中で操作と称される）を使用することによって解析および／または比較され得る。

10

【0244】

ある特定の実施形態において、2 つまたはそれを超えるリード密度プロファイルの比較は、2 つまたはそれを超えるリード密度プロファイルに対する不確定性の尺度の決定および／または比較を含む。リード密度プロファイルおよび／または関連する不確定性の尺度は、データセットの数学的操作および／もしくは統計的操作の解釈を容易にするため、ならびに／またはアウトカムを提供するために、比較されるときがある。試験被験体に対して生成されたリード密度プロファイルは、1 つまたはそれを超える参照（例えば、参照サンプル、参照被験体など）に対して生成されたリード密度プロファイルと比較されるときがある。いくつかの実施形態において、参照リード密度プロファイルが、コピー数変異を有しないと知られている参照被験体（例えば、参照）のセットから得られる場合、アウトカムは、染色体、その部分またはセグメントについて、試験被験体からのリード密度プロファイルを参照からのリード密度プロファイルと比較することによって提供される。いくつかの実施形態において、参照リード密度プロファイルが、特定のコピー数変異（例えば、染色体異数性、トリソミー、微小重複、微小欠失）を有すると知られている参照被験体のセットから得られる場合、アウトカムは、染色体、その部分またはセグメントについて、試験被験体からのリード密度プロファイルを参照からのリード密度プロファイルと比較することによって提供される。

20

30

【0245】

ある特定の実施形態において、試験被験体のリード密度プロファイルは、コピー数変異が存在しないことを代表する所定の値と比較され、コピー数変異が位置づけられるゲノム位置に対応する 1 つまたはそれを超えるゲノム位置（例えば、部分）において所定の値から逸脱するときがある。例えば、試験被験体（例えば、コピー数変異に関連する医学的症状に対するリスクがあるかまたはその医学的症状に罹患している被験体）において、試験被験体が、対象のコピー数変異を含むとき、リード密度プロファイルは、選択された部分について、参照（例えば、参照配列、参照被験体、参照セット）のリード密度プロファイルと有意に異なると予想される。試験被験体が、対象のコピー数変異を含まないとき、試験被験体のリード密度プロファイルは、選択された部分について、参照（例えば、参照配列、参照被験体、参照セット）のリード密度プロファイルと実質的に同じであることが多い。リード密度プロファイルは、所定のしきい値および／またはしきい値の範囲と比較されることが多い。用語「しきい値」は、本明細書中で使用されるとき、適格なデータセットを用いて算出され、コピー数変異（例えば、コピー数変異、異数性、染色体異常、微小重複、微小欠失など）の診断の限度として役立つ任意の数字のことを指す。ある特定の実施形態において、しきい値は、本明細書中に記載される方法によって得られる結果によって上回られ、被験体は、コピー数変異（例えば、トリソミー）と診断される。いくつかの実施形態において、閾値または値の範囲は、配列リードデータ（例えば、参照および／ま

40

50

たは被験体からの)を数学的におよび／または統計的に操作することによって算出されることが多い。コピー数変異の存在もしくは非存在を示唆する所定の閾値または閾値の範囲は、変動し得るが、コピー数変異の存在もしくは非存在の判定に有用なアウトカムをなおも提供する。ある特定の実施形態において、正規化されたリード密度および／または正規化されたカウントを含むリード密度プロファイルは、分類および／またはアウトカムの提供を容易にするために生成される。アウトカムは、正規化されたカウントを含むリード密度プロファイルのプロットに基づいて(例えば、そのようなリード密度プロファイルのプロットを用いて)提供され得る。

【0246】

いくつかの実施形態において、あるシステムは、スコアリングモジュール46を備える。スコアリングモジュールは、リード密度プロファイル(例えば、調整され、正規化されたリード密度プロファイル)を別の好適なモジュール(例えば、プロファイル生成モジュール26、PCA統計モジュール33、部分重み付けモジュール42など)から受け取り得、検索し得、および／または格納し得る。スコアリングモジュールは、2つまたはそれを超えるリード密度プロファイル(例えば、試験プロファイル、参照プロファイル、訓練セット、試験被験体)を受け取り得、検索し得、格納し得、および／または比較し得る。スコアリングモジュールは、スコア(例えば、プロット、プロファイル統計、比較(例えば、2つまたはそれを超えるプロファイル間の差異)、Z得点、不確定性の尺度、コールゾーン(call zone)、サンプルコール(sample call)50(例えば、コピー数変異の存在もしくは非存在の判定)および／またはアウトカム)を提供し得ることが多い。スコアリングモジュールは、スコアをエンドユーザーおよび／または別の好適なモジュール(例えば、ディスプレイ、プリンターなど)に提供し得る。いくつかの実施形態において、スコアリングモジュールは、特定のテストに対するカイ二乗統計量(例えば、High-chr21カウント)を計算するためのR関数を含む、下記に示されるRコードのいくつか、すべてまたは修正したものを備える。

3つのパラメータは、以下である：

x = サンプルリードデータ (部分×サンプル)

m = 部分に対する中央値

y = テストベクトル (例、chr21に対するTrueを除くすべての部分に対してFalse)

【数6】

```
getChisqP <- function(x,m,y)
{
  ahigh <- apply(x[,y],2,function(x) sum((x>m[,y])))
  alow <- sum(!y)-ahigh
  bhigh <- apply(x[,y],2,function(x) sum((x>m[,y])))
  blow <- sum(y)-bhigh
  p <- sapply(1:length(ahigh), function(i) {
    p <- chisq.test(matrix(c(ahigh[i],alow[i],bhigh[i],blow[i]),2))$p.value/2
    if (ahigh[i]/alow[i] > bhigh[i]/blow[i]) p <- max(p,1-p)
    else p <- min(p,1-p); p})

  return(p)
}
```

ハイブリッド回帰正規化

【0247】

いくつかの実施形態において、ハイブリッド正規化方法が使用される。いくつかの実施形態において、ハイブリッド正規化方法は、バイアス(例えば、GCバイアス)を減少させる。ハイブリッド正規化は、いくつかの実施形態において、(i) 2つの変数(例えば、カウントおよびGC含有量)の関係の解析、ならびに(ii) その解析に従った正規化方法の選択および適用を含む。ハイブリッド正規化は、ある特定の実施形態において、(

i) 回帰 (例えば、回帰分析) ならびに (i i) その回帰に従った正規化方法の選択および適用を含む。いくつかの実施形態において、第 1 のサンプル (例えば、第 1 のサンプルセット) に対して得られたカウントは、別のサンプル (例えば、第 2 のサンプルセット) から得られたカウントとは異なる方法によって正規化される。いくつかの実施形態において、第 1 のサンプル (例えば、第 1 のサンプルセット) に対して得られたカウントは、第 1 の正規化方法によって正規化され、第 2 のサンプル (例えば、第 2 のサンプルセット) から得られたカウントは、第 2 の正規化方法によって正規化される。例えば、ある特定の実施形態において、第 1 の正規化方法は、線形回帰の使用を含み、第 2 の正規化方法は、非線形回帰 (例えば、LOESS、GC-LOESS、LOWESS 回帰、LOESS 平滑化) の使用を含む。

10

【0248】

いくつかの実施形態において、ハイブリッド正規化方法は、ゲノムまたは染色体の部分にマッピングされた配列リード (例えば、カウント、マッピングされたカウント、マッピングされたリード) を正規化するために使用される。ある特定の実施形態において、生のカウントが正規化され、いくつかの実施形態において、調整された、重み付けされた、フィルタリングされたまたは予め正規化されたカウントが、ハイブリッド正規化方法によって正規化される。ある特定の実施形態において、ゲノム区分レベルまたは Z 得点が、正規化される。いくつかの実施形態において、ゲノムまたは染色体の選択された部分にマッピングされたカウントが、ハイブリッド正規化アプローチによって正規化される。カウントは、ゲノムの部分にマッピングされた配列リードの好適な尺度のことを指し得、その非限

20

【0249】

いくつかの実施形態において、正規化方法 (例えば、正規化方法のタイプ) は、回帰 (例えば、回帰分析) および/または相関係数に従って選択される。回帰分析とは、変数 (例えば、カウントおよび GC 含有量) の間の関係を推定するための統計的手法のことを指す。いくつかの実施形態において、回帰は、参照ゲノムの複数の部分の各部分に対するカウントおよび GC 含有量の尺度に従って生成される。好適な GC 含有量の尺度が使用され得、その非限定的な例としては、グアニン、シトシン、アデニン、チミン、プリン (GC) もしくはピリミジン (AT または ATU) 含有量の尺度、融解温度 (T_m) (例えば、変性温度、アニーリング温度、ハイブリダイゼーション温度)、自由エネルギーの尺度などまたはそれらの組み合わせが挙げられる。グアニン (G)、シトシン (C)、アデニン (A)、チミン (T)、プリン (GC) またはピリミジン (AT または ATU) 含有量の尺度は、比率またはパーセンテージとして表現され得る。いくつかの実施形態において、任意の好適な比率またはパーセンテージが使用され、その非限定的な例としては、GC/AT、GC/全ヌクレオチド、GC/A、GC/T、AT/全ヌクレオチド、AT/GC、AT/G、AT/C、G/A、C/A、G/T、G/A、G/AT、C/T などまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、GC 含有量の尺度は、全ヌクレオチド含有量に対する GC の比率またはパーセンテージである。いくつかの実施形態において、GC 含有量の尺度は、参照ゲノムの部分にマッピングされた配列リードに対する全ヌクレオチド含有量に対する GC の比率またはパーセンテージである。ある特定の

30

40

50

は配列リードから決定されない。ある特定の実施形態において、GC含有量の尺度は、1つまたはそれを超える被験体から得られた1つまたはそれを超えるサンプルに対して決定される。

【0250】

いくつかの実施形態において、回帰の生成は、回帰分析または相関分析の生成を含む。好適な回帰が使用され得、その非限定的な例としては、回帰分析（例えば、線形回帰分析）、適合度分析、ピアソン相関分析、順位相関、不明な分散の割合（fraction of variance unexplained）、Nash-Sutcliffeモデル効率解析、回帰モデル検証、損失比例減少（proportional reduction in loss）、根平均二乗偏差などまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、回帰直線が生成される。ある特定の実施形態において、回帰の生成は、線形回帰の生成を含む。ある特定の実施形態において、回帰の生成は、非線形回帰（例えば、LOESS回帰、LOWESS回帰）の生成を含む。

10

【0251】

いくつかの実施形態において、回帰は、例えば、カウントとGC含有量の尺度との間の、相関（例えば、線形相関）の存在もしくは非存在を判定する。いくつかの実施形態において、回帰（例えば、線形回帰）が生成され、相関係数が決定される。いくつかの実施形態において、好適な相関係数が決定され、その非限定的な例としては、決定係数、 R^2 値、ピアソン相関係数などが挙げられる。

【0252】

いくつかの実施形態において、適合度は、回帰（例えば、回帰分析、線形回帰）に対して測定される。適合度は、視覚的解析または数学的解析によって測定されるときがある。評価は、その適合度が非線形回帰に対してより高いかまたは線形回帰に対してより高いかを判定することを含むときがある。いくつかの実施形態において、相関係数は、適合度の尺度である。いくつかの実施形態において、回帰に対する適合度の評価は、相関係数および／または相関係数のカットオフ値に従って明らかにされる。いくつかの実施形態において、適合度の評価は、相関係数と相関係数のカットオフ値とを比較することを含む。いくつかの実施形態において、回帰に対する適合度の評価は、線形回帰を示唆する。例えば、ある特定の実施形態において、適合度は、非線形回帰よりも線形回帰に対してより高く、その適合度の評価は、線形回帰を示唆する。いくつかの実施形態において、評価は、線形回帰を示唆し、カウントを正規化するために線形回帰が使用される。いくつかの実施形態において、回帰に対する適合度の評価は、非線形回帰を示唆する。例えば、ある特定の実施形態において、適合度は、線形回帰よりも非線形回帰に対してより高く、その適合度の評価は、非線形回帰を示唆する。いくつかの実施形態において、評価は、非線形回帰を示唆し、カウントを正規化するために非線形回帰が使用される。

20

30

【0253】

いくつかの実施形態において、相関係数が、相関係数のカットオフと等しいかまたはそれを超えるとき、適合度の評価は、線形回帰を示唆する。いくつかの実施形態において、相関係数が、相関係数のカットオフ未満であるとき、適合度の評価は、非線形回帰を示唆する。いくつかの実施形態において、相関係数のカットオフは、予め決定されている。いくつかの実施形態において、相関係数のカットオフは、約0.5もしくはそれを超えるか、約0.55もしくはそれを超えるか、約0.6もしくはそれを超えるか、約0.65もしくはそれを超えるか、約0.7もしくはそれを超えるか、約0.75もしくはそれを超えるか、約0.8もしくはそれを超えるか、または約0.85もしくはそれを超える。

40

【0254】

例えば、ある特定の実施形態において、相関係数が約0.6と等しいかまたはそれを超えるとき、線形回帰を含む正規化方法が使用される。ある特定の実施形態において、相関係数が、0.6という相関係数のカットオフと等しいかまたはそれを超えるとき、サンプルのカウント（例えば、参照ゲノムの部分ごとのカウント、部分ごとのカウント）は、線形回帰に従って正規化され、そうでなければ、カウントは、非線形回帰に従って正規化さ

50

れる（例えば、その係数が 0.6 という相関係数のカットオフ未満であるとき）。いくつかの実施形態において、正規化プロセスは、参照ゲノムの複数の部分の各部分に対する、
(i) カウントおよび (ii) GC 含有量に対する線形回帰または非線形回帰の生成を含む。ある特定の実施形態において、相関係数が、0.6 という相関係数のカットオフ未満であるとき、非線形回帰を含む正規化方法（例えば、LOWESS、LOESS）が使用される。いくつかの実施形態において、相関係数（例えば、相関係数）が、約 0.7、約 0.65 未満、約 0.6 未満、約 0.55 未満または約 0.5 未満という相関係数のカットオフ未満であるとき、非線形回帰を含む正規化方法（例えば、LOWESS）が使用される。例えば、いくつかの実施形態において、相関係数が、約 0.6 という相関係数のカットオフ未満であるとき、非線形回帰を含む正規化方法（例えば、LOWESS、LOESS）が使用される。

10

【0255】

いくつかの実施形態において、特定のタイプの回帰が選択され（例えば、線形または非線形回帰）、その回帰が生成された後、その回帰をカウントから減算することによって、カウントが正規化される。いくつかの実施形態において、カウントからの回帰の減算は、バイアス（例えば、GC バイアス）が減少した正規化されたカウントを提供する。いくつかの実施形態において、線形回帰が、カウントから減算される。いくつかの実施形態において、非線形回帰（例えば、LOESS、GC-LOESS、LOWESS 回帰）が、カウントから減算される。任意の好適な方法が、カウントから回帰直線を減算するために使用され得る。例えば、カウント x が、0.5 という GC 含有量を含む部分 i （例えば、部分 i ）から得られ、回帰直線が、0.5 という GC 含有量においてカウント y を決定する場合、部分 i に対して、 $x - y =$ 正規化されたカウントである。いくつかの実施形態において、カウントは、回帰の減算の前および／または後に正規化される。いくつかの実施形態において、ハイブリッド正規化アプローチによって正規化されたカウントは、ゲノム区分レベル、Z 得点（Z-scores）、レベルおよび／またはゲノムもしくはそのセグメントのプロファイルを生成するために使用される。ある特定の実施形態において、ハイブリッド正規化アプローチによって正規化されたカウントは、コピー数変異（例えば、胎児における）の存在もしくは非存在を判定するために本明細書中に記載される方法によって解析される。

20

【0256】

いくつかの実施形態において、ハイブリッド正規化方法は、正規化の前または後に、1 つまたはそれを超える部分のフィルタリングまたは重み付けを含む。本明細書中に記載される部分（例えば、参照ゲノムの部分）をフィルタリングする方法を含む、部分をフィルタリングする好適な方法が、使用され得る。いくつかの実施形態において、部分（例えば、参照ゲノムの部分）が、ハイブリッド正規化方法を適用する前にフィルタリングされる。いくつかの実施形態において、選択された部分（例えば、カウントのばらつきに従って選択された部分）にマッピングされたシーケンシングリードのカウントだけが、ハイブリッド正規化によって正規化される。いくつかの実施形態において、フィルタリングされた参照ゲノムの部分（例えば、カウントのばらつきに従ってフィルタリングされた部分）にマッピングされたシーケンシングリードのカウントが、ハイブリッド正規化方法を用いる前に除去される。いくつかの実施形態において、ハイブリッド正規化方法は、好適な方法（例えば、本明細書中に記載される方法）に従って部分（例えば、参照ゲノムの部分）を選択するかまたはフィルタリングすることを含む。いくつかの実施形態において、ハイブリッド正規化方法は、複数のテストサンプルに対する各部分にマッピングされたカウントに対する不確定値に従って部分（例えば、参照ゲノムの部分）を選択することまたはフィルタリングすることを含む。いくつかの実施形態において、ハイブリッド正規化方法は、カウントのばらつきに従って部分（例えば、参照ゲノムの部分）を選択するかまたはフィルタリングすることを含む。いくつかの実施形態において、ハイブリッド正規化方法は、GC 含有量、反復エレメント、反復配列、イントロン、エキソンなどまたはそれらの組み合わせに従って部分（例えば、参照ゲノムの部分）を選択するかまたはフィルタリングす

30

40

50

ることを含む。

【0257】

例えば、いくつかの実施形態において、複数の妊婦被験体由来の複数のサンプルが解析され、部分のサブセット（例えば、参照ゲノムの部分）が、カウントのばらつきに従って選択される。ある特定の実施形態において、妊婦被験体から得られたサンプルに対する選択された各部分に対する（i）カウントおよび（i i）GC含有量に対する相関係数を決定するために、線形回帰が用いられる。いくつかの実施形態において、所定の相関のカットオフ値（例えば、約0.6）より大きい相関係数が決定され、適合度の評価は、線形回帰を示唆し、線形回帰をカウントから減算することによって、カウントが正規化される。ある特定の実施形態において、所定の相関のカットオフ値（例えば、約0.6）未満の相関係数が決定され、適合度の評価は、非線形回帰を示唆し、LOESS回帰が生成され、そのLOESS回帰をカウントから減算することによって、カウントが正規化される。

10

プロファイル

【0258】

いくつかの実施形態において、処理工程は、データセットまたはその微分演算（例えば、当該分野で公知のおよび／または本明細書中に記載される1つまたはそれを超える数学的および／または統計的データ処理工程の結果）の様々な態様から1つまたはそれを超えるプロファイル（例えば、プロファイルプロット）を生成することを含み得る。

【0259】

用語「プロファイル」は、本明細書中で使用されるとき、大量のデータにおけるパターンおよび／または相関の特定を容易にし得る、データの数学的操作および／または統計的操作の結果のことを指す。「プロファイル」は、1つまたはそれを超える基準に基づくデータまたはデータセットの1つまたはそれを超える操作から生じる値を含むことが多い。プロファイルは、複数のデータポイントを含むことが多い。データセットの性質および／または複雑さに応じて、任意の好適な数のデータポイントが、プロファイルに含められ得る。ある特定の実施形態において、プロファイルは、2つもしくはそれを超えるデータポイント、3つもしくはそれを超えるデータポイント、5つもしくはそれを超えるデータポイント、10個もしくはそれを超えるデータポイント、24個もしくはそれを超えるデータポイント、25個もしくはそれを超えるデータポイント、50個もしくはそれを超えるデータポイント、100個もしくはそれを超えるデータポイント、500個もしくはそれを超えるデータポイント、1000個もしくはそれを超えるデータポイント、5000個もしくはそれを超えるデータポイント、10,000個もしくはそれを超えるデータポイントまたは100,000個もしくはそれを超えるデータポイントを含み得る。

20

30

【0260】

いくつかの実施形態において、プロファイルは、データセット全体を代表し、ある特定の実施形態において、プロファイルは、データセットの一部またはサブセットを代表する。つまり、プロファイルは、任意のデータを除去するためにフィルタリングされていないデータを代表するデータポイントを含むかまたはそれらのデータポイントから生成されるときがあり、プロファイルは、望まれないデータを除去するためにフィルタリングされたデータを代表するデータポイントを含むかまたはそれらのデータポイントから生成されるときがある。いくつかの実施形態において、あるプロファイルにおけるデータポイントは、ある部分に対するデータ操作の結果に相当する。ある特定の実施形態において、あるプロファイルにおけるデータポイントは、部分の群に対するデータ操作の結果を含む。いくつかの実施形態において、部分の群は、互いに隣接することもあるし、ある特定の実施形態では、部分の群は、染色体またはゲノムの異なる部分に由来することもある。

40

【0261】

あるデータセットに由来するプロファイルにおけるデータポイントは、任意の好適なデータのカテゴリー化を代表し得る。プロファイルデータポイントを生成するためにデータがグループ化され得るカテゴリーの非限定的な例としては、サイズに基づく部分、配列の特徴（例えば、GC含有量、AT含有量、染色体上の位置（例えば、短腕、長腕、セント

50

ロメア、テロメア)など)に基づく部分、発現レベル、染色体などまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、あるプロファイルは、別のプロファイル(例えば、再正規化されたデータプロファイルを生成するために異なる正規化値に対して再正規化された、正規化されたデータプロファイル)から得られたデータポイントから生成され得る。ある特定の実施形態において、別のプロファイルから得られたデータポイントから生成されたプロファイルは、データポイントの数および/またはデータセットの複雑さを減少させる。データポイントの数および/またはデータセットの複雑さの減少は、データの解釈を容易にするおよび/またはアウトカムの提供を容易にすることが多い。

【0262】

プロファイル(例えば、ゲノムのプロファイル、染色体のプロファイル、染色体のセグメントのプロファイル)は、2つまたはそれを超える部分に対する正規化されたカウントまたは正規化されていないカウントの集合であることが多い。プロファイルは、少なくとも1つのレベル(例えば、ゲノム区分レベル)を含むことが多く、2つまたはそれを超えるレベルを含むことが多い(例えば、あるプロファイルは、複数のレベルを有することが多い)。あるレベルは、一般に、ほぼ同じカウントまたは正規化されたカウントを有する部分のセットに対するものである。レベルは、本明細書中にさらに詳細に記載される。ある特定の実施形態において、プロファイルは、1つまたはそれを超える部分を含み、それらの部分は、重み付けされ得るか、除去され得るか、フィルタリングされ得るか、正規化され得るか、調整され得るか、平均され得るか、平均値として導かれ得るか、加算され得るか、減算され得るか、処理され得るか、またはそれらの任意の組み合わせによって変換され得る。プロファイルは、2つまたはそれを超えるレベルを定義している部分にマッピングされて正規化されたカウントを含むことが多く、ここで、それらのカウントは、好適な方法によって、それらのレベルのうちの1つに従ってさらに正規化される。プロファイル(例えば、プロファイルレベル)のカウントは、不確定値に関連することが多い。

【0263】

1つまたはそれを超えるレベルを含むプロファイルは、パディングされる(例えば、ホールパディング(hole padding))ときがある。パディング(例えば、ホールパディング)とは、母体の微小欠失または母体の重複(例えば、コピー数変異)に起因する、プロファイルにおけるレベルを特定し、調整するプロセスのことを指す。いくつかの実施形態において、胎児の微小重複または胎児の微小欠失に起因するレベルがパディングされる。プロファイルにおける微小重複または微小欠失は、いくつかの実施形態において、染色体異数性(例えば、トリソミー)の偽陽性または偽陰性の判定をもたらすプロファイル(例えば、染色体のプロファイル)のレベル全体を人工的に上昇させるかまたは低下させ得る。いくつかの実施形態において、微小重複および/または欠失に起因するプロファイルにおけるレベルは、パディングまたはホールパディングと称されるときがあるプロセスによって特定され、調整される(例えば、パディングされるおよび/または除去される)。ある特定の実施形態において、あるプロファイルは、そのプロファイル内の第2のレベルと有意に異なる1つまたはそれを超える第1のレベルを含み、その1つまたはそれを超える第1のレベルの各々は、母体のコピー数変異、胎児のコピー数変異、または母体のコピー数変異および胎児のコピー数変異を含み、1つまたはそれを超える第1のレベルが、調整される。

【0264】

1つまたはそれを超えるレベルを含むプロファイルは、第1のレベルおよび第2のレベルを含み得る。いくつかの実施形態において、第1のレベルは、第2のレベルと異なる(例えば、有意に異なる)。いくつかの実施形態において、第1のレベルは、第1の部分セットを含み、第2のレベルは、第2の部分セットを含み、第1の部分セットは、第2の部分セットのサブセットではない。ある特定の実施形態において、第1の部分セットは、第1および第2のレベルが測定される第2の部分セットと異なる。いくつかの実施形態において、あるプロファイルは、そのプロファイル内の第2のレベルと異なる(例えば、有意に異なる、例えば、有意に異なる値を有する)複数の第1のレベルを有し得る。いくつか

の実施形態において、あるプロファイルは、そのプロファイル内の第2のレベルと有意に異なる1つまたはそれを超える第1のレベルを含み、その1つまたはそれを超える第1のレベルが、調整される。いくつかの実施形態において、あるプロファイルは、そのプロファイル内の第2のレベルと有意に異なる1つまたはそれを超える第1のレベルを含み、その1つまたはそれを超える第1のレベルの各々は、母体のコピー数変異、胎児のコピー数変異、または母体のコピー数変異および胎児のコピー数変異を含み、1つまたはそれを超える第1のレベルが、調整される。いくつかの実施形態において、あるプロファイル内の第1のレベルは、そのプロファイルから除去されるか、または調整される（例えば、パディングされる）。あるプロファイルは、1つまたはそれを超える第2のレベルと有意に異なる1つまたはそれを超える第1のレベルを含む複数のレベルを含み得、あるプロファイルにおけるレベルの大部分が、第2のレベルであることが多く、その第2のレベルは、互いにほぼ等しい。いくつかの実施形態において、あるプロファイルにおけるレベルの50%超、60%超、70%超、80%超、90%超または95%超が、第2のレベルである。

10

【0265】

プロファイルは、プロットとして表示されるときがある。例えば、部分のカウント（例えば、正規化されたカウント）を表している1つまたはそれを超えるレベルが、プロットされ得、可視化され得る。生成され得るプロファイルのプロットの非限定的な例としては、生のカウント（例えば、生のカウントプロファイルまたは生のプロファイル）、正規化されたカウント、部分によって重み付けされた、z得点、p値、当てはめられた倍数性に対する面積比、当てはめられた胎児分率と計測された胎児分率との間の比率に対する中央値レベル、主成分などまたはそれらの組み合わせが挙げられる。プロファイルのプロットは、いくつかの実施形態において、操作されたデータの可視化を可能にする。ある特定の実施形態において、プロファイルのプロットは、アウトカム（例えば、当てはめられた倍数性に対する面積比、当てはめられた胎児分率と計測された胎児分率との間の比率に対する中央値レベル、主成分）を提供するために使用され得る。用語「生のカウントプロファイルプロット」または「生のプロファイルプロット」は、本明細書中で使用されるとき、ある領域（例えば、ゲノム、部分、染色体、参照ゲノムの染色体部分または染色体のセグメント）における全カウントに対して正規化された、ある領域における各部分におけるカウントのプロットのことを指す。いくつかの実施形態において、プロファイルは、スタティックウィンドウプロセスを用いて生成され得、ある特定の実施形態において、プロファイルは、スライディングウィンドウプロセスを用いて生成され得る。

20

30

【0266】

試験被験体に対して生成されたプロファイルは、データセットの数学的操作および／もしくは統計的操作の解釈を容易にするため、ならびに／またはアウトカムを提供するために、1つまたはそれを超える参照被験体に対して生成されたプロファイルと比較されるときがある。いくつかの実施形態において、プロファイルは、1つまたはそれを超える開始時の（starting）仮定（例えば、核酸の母体の寄与（例えば、母体分率）、核酸の胎児の寄与（例えば、胎児分率）、参照サンプルの倍数性などまたはそれらの組み合わせ）に基づいて生成される。ある特定の実施形態において、試験プロファイルは、コピー数変異が存在しないことを代表する所定の値の周りに集中することが多く、試験被験体がコピー数変異を有した場合、コピー数変異が試験被験体内に位置するゲノム位置に対応する領域における所定の値から逸脱することが多い。コピー数変異に関連する医学的症状に対するリスクがあるかまたはコピー数変異に関連する医学的症状に罹患している試験被験体では、選択された部分に対する数値は、影響されていないゲノム位置に対する所定の値から有意に変動すると予想される。開始時の仮定（例えば、既定の倍数性もしくは最適化された倍数性、既定の胎児分率もしくは最適化された胎児分率またはそれらの組み合わせ）に応じて、コピー数変異の存在もしくは非存在を示唆する所定の閾値もしくはカットオフ値または閾値範囲は、変動し得るが、コピー数変異の存在もしくは非存在の判定に有用なアウトカムをなおも提供する。いくつかの実施形態において、プロファイルは、表現型

40

50

を示唆し、かつ／または表現型を代表する。

【0267】

非限定的な例として、正規化されたサンプルカウントプロファイルおよび／または正規化された参照カウントプロファイルは、(a) 選択された染色体、その部分またはセグメントに対する参照カウントの中央値を、コピー数変異を有しないと知られている参照のセットから算出し、(b) 参照サンプルの生のカウントから情報価値のない部分を除去し（例えば、フィルタリングし）；(c) 参照ゲノムの残りのすべての部分に対する参照カウントを、参照サンプルの選択された染色体または選択されたゲノム位置に対する残りのカウント総数（例えば、参照ゲノムの情報価値のない部分を除去した後に残ったカウントの合計）に対して正規化し、それにより、正規化された参照被験体プロファイルを生成し；(d) 試験被験体のサンプルから対応する部分を除去し；(e) 1つまたはそれを超える選択されたゲノム位置に対する残りの試験被験体カウントを、選択されたゲノム位置を含む染色体に対する残りの参照カウントの中央値の和に対して正規化し、それにより、正規化された試験被験体プロファイルを生成することによって、生の配列リードデータから得ることができる。ある特定の実施形態において、(b) において、フィルタリングされた部分によって減少される、ゲノム全体に対するさらなる正規化工程は、(c) と (d) との間に含めることができる。

10

【0268】

データセットプロファイルは、マッピングされてカウントされた配列リードデータの1つまたはそれを超える操作によって生成され得る。いくつかの実施形態は、以下を含む。配列リードが、マッピングされ、各ゲノム部分にマップするカウント（すなわち、配列タグ）の数が測定される（例えば、カウントされる）。マッピングされてカウントされた配列リードから、生のカウントプロファイルが生成される。ある特定の実施形態において、コピー数変異を有しないと知られている参照被験体のセットからの染色体、その部分またはセグメントについて、試験被験体からの生のカウントプロファイルを参照カウントプロファイルの中央値とを比較することによって、アウトカムが提供される。

20

【0269】

いくつかの実施形態において、配列リードデータは、必要に応じてフィルタリングされることにより、ノイズの多いデータまたは情報価値のない部分が除去される。フィルタリングの後、残ったカウントは、通常、合計されることにより、フィルタリングされたデータセットが生成される。フィルタリングされたカウントプロファイルは、ある特定の実施形態において、フィルタリングされたデータセットから生成される。

30

【0270】

配列リードデータが、カウントされ、必要に応じてフィルタリングされた後、データセットは、正規化されることにより、レベルまたはプロファイルが生成され得る。データセットは、1つまたはそれを超える選択された部分を好適な正規化参照値に対して正規化することによって正規化され得る。いくつかの実施形態において、正規化参照値は、部分が選択された染色体に対する全カウントを代表する。ある特定の実施形態において、正規化参照値は、コピー数変異を有しないと知られている参照被験体のセットから調製された参照データセットからの、1つまたはそれを超える対応する部分、染色体の部分または染色体を代表する。いくつかの実施形態において、正規化参照値は、コピー数変異の存在もしくは非存在について解析されている試験被験体から調製された試験被験体データセットからの、1つまたはそれを超える対応する部分、染色体の部分または染色体を代表する。ある特定の実施形態において、正規化プロセスは、スタティックウィンドウアプローチを用いて行われ、いくつかの実施形態において、正規化プロセスは、ムービングウィンドウアプローチまたはスライディングウィンドウアプローチを用いて行われる。ある特定の実施形態において、正規化されたカウントを含むプロファイルは、分類および／またはアウトカムの提供を容易にするために生成される。アウトカムは、正規化されたカウントを含むプロファイルのプロットに基づいて（例えば、そのようなプロファイルのプロットを用いて）提供され得る。

40

50

レベル

【0271】

いくつかの実施形態において、値（例えば、数、定量値）は、レベルに帰される。レベルは、好適な方法、演算または数学的プロセスによって決定され得る（例えば、処理されたレベル）。レベルは、部分セットに対するカウント（例えば、正規化されたカウント）であるか、またはそのカウントから導かれることが多い。いくつかの実施形態において、ある部分のレベルは、ある部分にマッピングされたカウント（例えば、カウント、正規化されたカウント）の総数と実質的に等しい。レベルは、当該分野で公知の好適な方法、演算または数学的プロセスによって処理されるか、変換されるか、または操作されたカウントから決定されることが多い。いくつかの実施形態において、あるレベルは、処理されたカウントから導かれ、処理されたカウントの非限定的な例としては、重み付けされたカウント、除去されたカウント、フィルタリングされたカウント、正規化されたカウント、調整されたカウント、平均されたカウント、平均値（例えば、平均値レベル）として導かれたカウント、加算されたカウント、減算されたカウント、変換されたカウントまたはそれらの組み合わせが挙げられる。いくつかの実施形態において、あるレベルは、正規化されたカウント（例えば、部分の正規化されたカウント）を含む。あるレベルは、好適なプロセスによって正規化されたカウントに対するものであり得、そのプロセスの非限定的な例としては、部分ごとの正規化、GC含有量による正規化、カウントの中央値の正規化、線形および非線形最小二乗回帰、LOESS（例えば、GC LOESS）、LOWESS、PERUN、ChAI、主成分正規化、RM、GCRM、cQnなどおよび／またはそれらの組み合わせが挙げられる。あるレベルは、正規化されたカウントまたはカウントの相対量を含み得る。いくつかの実施形態において、あるレベルは、平均された2つまたはそれを超える部分のカウントまたは正規化されたカウントに対するものであり、そのレベルは、平均レベルと称される。いくつかの実施形態において、あるレベルは、平均値レベルと称される、カウントの平均値または正規化されたカウントの平均値を有する部分セットに対するものである。いくつかの実施形態において、あるレベルは、生のカウントおよび／またはフィルタリングされたカウントを含む部分に対して導かれる。いくつかの実施形態において、あるレベルは、生であるカウントに基づく。いくつかの実施形態において、あるレベルは、不確定値（例えば、標準偏差、MAD）に関連する。いくつかの実施形態において、あるレベルは、Z得点またはp値によって表される。

10

20

30

【0272】

1つまたはそれを超える部分に対するレベルは、本明細書中において「ゲノム区分レベル」と同義である。用語「レベル」は、本明細書中で使用されるとき、用語「高さ」と同義であるときがある。ある特定の場において、用語「レベル」は、「配列リードカウント提示」および／または「染色体提示」と同義であり得る。用語「レベル」の意味の判定は、それが使用されている文脈から判定され得る。例えば、用語「レベル」は、ゲノム区分、プロファイル、リードおよび／またはカウントの文脈において使用されているとき、高さを意味することが多い。用語「レベル」は、物質または組成物の文脈において使用されているとき（例えば、RNAのレベル、プレキシングレベル（plexing level）、量のことを指すことが多い。用語「レベル」は、不確定性の文脈において使用されているとき（例えば、誤差のレベル、信頼のレベル、偏差のレベル、不確定性のレベル）、量のことを指すことが多い。

40

【0273】

2つまたはそれを超えるレベル（例えば、あるプロファイルにおける2つまたはそれを超えるレベル）に対する正規化されたまたは正規化されていないカウントは、レベルに従って数学的に操作され得る（例えば、加算され得る、乗算され得る、平均され得る、正規化され得るなどまたはそれらの組み合わせ）ときがある。例えば、2つまたはそれを超えるレベルに対する正規化されたまたは正規化されていないカウントは、あるプロファイルにおけるレベルのうちの1つ、いくつかまたはすべてに従って正規化され得る。いくつかの実施形態において、あるプロファイルにおけるすべてのレベルの正規化されたまたは正

50

規化されていないカウントが、そのプロファイルにおける1つのレベルに従って正規化される。いくつかの実施形態において、あるプロファイルにおける第1の(f i s t)レベルの正規化されたまたは正規化されていないカウントは、そのプロファイルにおける第2のレベルの正規化されたまたは正規化されていないカウントに従って正規化される。

【0274】

レベル(例えば、第1のレベル、第2のレベル)の非限定的な例は、処理されたカウントを含む部分セットに対するレベル、カウントの平均値、中央値または平均を含む部分セットに対するレベル、正規化されたカウントを含む部分セットに対するレベルなどまたはそれらの任意の組み合わせである。いくつかの実施形態において、あるプロファイルにおける第1のレベルおよび第2のレベルは、同じ染色体にマッピングされた部分のカウントから導かれる。いくつかの実施形態において、あるプロファイルにおける第1のレベルおよび第2のレベルは、異なる染色体にマッピングされた部分のカウントから導かれる。

10

【0275】

いくつかの実施形態において、レベルは、1つまたはそれを超える部分にマッピングされた正規化されたまたは正規化されていないカウントから決定される。いくつかの実施形態において、レベルは、2つまたはそれを超える部分にマッピングされた正規化されたまたは正規化されていないカウントから決定され、ここで、各部分に対する正規化されたカウントは、ほぼ同じであることが多い。カウント(例えば、正規化されたカウント)の変動が、あるレベルに対する部分セットに存在し得る。あるレベルに対する部分セットにおいて、そのセットの他の部分(例えば、ピークおよび/またはディップ)と有意に異なるカウントを有する1つまたはそれを超える部分が存在し得る。任意の好適な数の部分に関連する任意の好適な数の正規化されたまたは正規化されていないカウントが、レベルを定義し得る。

20

【0276】

いくつかの実施形態において、あるゲノムの部分のすべてまたはいくつかの正規化されたまたは正規化されていないカウントから、1つまたはそれを超えるレベルが決定され得る。ある染色体またはそのセグメントの正規化されたまたは正規化されていないカウントのすべてまたはいくつかから、レベルが決定され得ることが多い。いくつかの実施形態において、2つまたはそれを超える部分(例えば、部分セット)から導かれた2つまたはそれを超えるカウントが、レベルを決定する。いくつかの実施形態において、2つまたはそれを超えるカウント(例えば、2つまたはそれを超える部分からのカウント)が、レベルを決定する。いくつかの実施形態において、2~約100,000個の部分からのカウントが、レベルを決定する。いくつかの実施形態において、2~約50,000個、2~約40,000個、2~約30,000個、2~約20,000個、2~約10,000個、2~約5000個、2~約2500個、2~約1250個、2~約1000個、2~約500個、2~約250個、2~約100個または2~約60個の部分からのカウントが、レベルを決定する。いくつかの実施形態において、約10~約50個の部分からのカウントが、レベルを決定する。いくつかの実施形態において、約20~約40個もしくはそれを超える部分からのカウントが、レベルを決定する。いくつかの実施形態において、あるレベルは、約2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、45、50、55、60個もしくはそれを超える部分からのカウントを含む。いくつかの実施形態において、あるレベルは、部分セット(例えば、参照ゲノムの部分セット、染色体の部分セットまたは染色体のセグメントの部分セット)に対応する。

30

40

【0277】

いくつかの実施形態において、あるレベルは、連続した部分の正規化されたまたは正規化されていないカウントに対して決定される。いくつかの実施形態において、連続した部分(例えば、部分セット)は、ゲノムの隣接したセグメントまたは染色体もしくは遺伝子の隣接したセグメントに相当する。例えば、2つまたはそれを超えて連続した部分は、そ

50

これらの部分を縦に並べてマージすることによってアラインメントされたとき、各部分よりも長いDNA配列の配列アセンブリであり得る。例えば、2つまたはそれを超えて連続した部分は、インタクトなゲノム、染色体、遺伝子、イントロン、エキソンまたはそれらのセグメントであり得る。いくつかの実施形態において、連続した部分および／または連続していない部分の集合（例えば、セット）からレベルが決定される。

決定分析

【0278】

いくつかの実施形態において、アウトカム（例えば、コールすること）の判定または染色体異数性、微小重複もしくは微小欠失の存在もしくは非存在の判定は、決定分析に従って行われる。ある特定の決定分析の特徴は、国際特許出願公開番号WO 2014/190286（この全体が参照により本明細書中に組み込まれる）に記載されている。例えば、決定分析は、1つまたはそれを超える結果、評価および／またはそれらの決定の起こり得る帰結に基づいて、それらの結果、それらの結果の評価および一連の決定をもたらす1つまたはそれを超える方法を適用すること、ならびに最終的な決定が行われるプロセスのある重大な局面において終了することを含むときがある。いくつかの実施形態において、決定分析は、決定木である。決定分析は、いくつかの実施形態において、1つまたはそれを超えるプロセス（例えば、処理工程、例えば、アルゴリズム）を協調させた使用を含む。決定分析は、人間、システム、装置、ソフトウェア（例えば、モジュール）、コンピュータ、プロセッサ（例えば、マイクロプロセッサ）などまたはそれらの組み合わせによって行われ得る。いくつかの実施形態において、決定分析は、決定分析を使用しない場合（例えば、正規化されたカウントから直接判定が行われる場合）と比較して、偽陰性が少なくかつ偽陽性が少ない判定で、胎児における染色体異数性、微小重複または微小欠失の存在もしくは非存在を判定する方法を含む。いくつかの実施形態において、決定分析は、1つまたはそれを超える微小重複または微小欠失に関連する症状の存在もしくは非存在の判定を含む。例えば、いくつかの実施形態において、決定分析は、被験体由来のテストサンプルに対する、ディジョージ症候群に関連する1つまたはそれを超えるコピー数変異の存在もしくは非存在の判定を含む。いくつかの実施形態において、決定分析は、被験体由来のテストサンプルに対するディジョージ症候群の存在もしくは非存在の判定を含む。

【0279】

いくつかの実施形態において、決定分析は、ゲノムまたはゲノムのセグメント（例えば、染色体またはその一部）に対するプロファイルの生成を含む。プロファイルは、公知のまたは本明細書中に記載される任意の好適な方法によって生成され得、参照ゲノムの部分にマッピングされた配列リードのカウントの入手、カウントの正規化、レベルの正規化、パディングなどまたはそれらの組み合わせを含むことが多い。参照ゲノムにマッピングされた配列リードのカウントの入手は、サンプルの入手（例えば、妊婦被験体から）、サンプル由来の核酸（例えば、循環無細胞核酸）の配列決定、配列リードの入手、参照ゲノムの部分への配列リードのマッピングなどおよびそれらの組み合わせを含み得る。いくつかの実施形態において、プロファイルの生成は、参照ゲノムの部分にマッピングされたカウントの正規化を含み、それにより、算出されたゲノム区分レベルが提供される。

【0280】

いくつかの実施形態において、決定分析は、セグメント化を含む。いくつかの実施形態において、セグメント化は、プロファイルを修正および／または変換し、それにより、プロファイルの1つまたはそれを超える分解レンダリング（*decomposition rendering*）が提供される。セグメント化プロセスに供されたプロファイルは、参照ゲノムまたはその一部（例えば、常染色体および性染色体）における部分（例えば、ピン）にマッピングされた正規化されたカウントのプロファイルであることが多い。本明細書中で述べられるように、それらの部分にマッピングされた生のカウントは、1つまたはそれを超える好適な正規化プロセス（例えば、PERUN、LOESS、GC-LOESS、主成分正規化（ChAI）またはそれらの組み合わせ）によって正規化されることにより、決定分析の一部としてセグメント化されたプロファイルが生成され得る。プロ

ファイルの分解レンダリングは、プロファイルの変換であることが多い。プロファイルの分解レンダリングは、プロファイルから、ゲノム、染色体またはそれらのセグメントの提示への変換であるときがある。

【0281】

ある特定の実施形態において、セグメント化のために用いられるセグメント化プロセスは、プロファイル内の1つまたはそれを超える他のレベルとは異なる（例えば、実質的にまたは有意に異なる）、プロファイル内の1つまたはそれを超えるレベルを位置づけ、特定する。あるプロファイルにおける別のレベルと異なり、そのプロファイルにおける別のレベルと異なるエッジを有する、セグメント化プロセスに従ってそのプロファイルにおいて特定されたレベルは、ウェーブレット、およびより一般的には、不連続のセグメントに対するレベルと本明細書中で称される。セグメント化プロセスは、1つまたはそれを超える不連続のセグメントまたはウェーブレットが特定され得る分解レンダリングを、正規化されたカウントまたはレベルのプロファイルから生成し得る。不連続のセグメントは、通常、セグメント化されるもの（例えば、染色体、複数の染色体、常染色体）より少ない部分（例えば、ピン）をカバーする。

【0282】

いくつかの実施形態において、セグメント化は、あるプロファイル内の不連続のセグメントおよびウェーブレットのエッジを位置づけ、特定する。ある特定の実施形態において、1つまたはそれを超える不連続のセグメントおよびウェーブレットの一方または両方のエッジが、特定される。例えば、セグメント化プロセスは、あるプロファイルにおける不連続のセグメントまたはウェーブレットの右および／または左のエッジの場所（例えば、ゲノム座標、例えば、部分の場所）を特定し得る。不連続のセグメントまたはウェーブレットは、2つのエッジを含むことが多い。例えば、不連続のセグメントまたはウェーブレットは、左のエッジおよび右のエッジを含み得る。いくつかの実施形態において、提示またはビューに応じて、左のエッジは、あるプロファイルにおいて、核酸セグメントの5' - エッジであり得、右のエッジは、3' - エッジであり得る。いくつかの実施形態において、左のエッジは、あるプロファイルにおいて、核酸セグメントの3' - エッジであり得、右のエッジは、5' - エッジであり得る。プロファイルのエッジは、セグメント化の前に既知であることが多く、ゆえに、いくつかの実施形態において、プロファイルのエッジは、あるレベルのどちらのエッジが5' - エッジであり、どちらのエッジが3' - エッジであるかを明らかにする。いくつかの実施形態において、あるプロファイルおよび／または不連続のセグメント（例えば、ウェーブレット）の一方または両方のエッジが、染色体のエッジである。

【0283】

いくつかの実施形態において、不連続のセグメントまたはウェーブレットのエッジは、参照サンプル（例えば、参照プロファイル）に対して生成された分解レンダリングに従って決定される。いくつかの実施形態において、参照プロファイル（例えば、染色体またはそのセグメントのプロファイル）の分解レンダリングに従って、ヌルエッジ高さ分布（null edge height distribution）が明らかにされる。ある特定の実施形態において、あるプロファイルにおける不連続のセグメントまたはウェーブレットのレベルが、ヌルエッジ高さ分布の外側であるとき、その不連続のセグメントまたはウェーブレットのエッジは特定される。いくつかの実施形態において、あるプロファイルにおける不連続のセグメントまたはウェーブレットのエッジは、参照プロファイルに対する分解レンダリングに従って算出されたZ得点に従って特定される。

【0284】

セグメント化は、あるプロファイルにおける2つまたはそれを超える不連続のセグメントまたはウェーブレット（例えば、2つまたはそれを超える断片化されたレベル、2つまたはそれを超える断片化されたセグメント）を生成するときがある。いくつかの実施形態において、セグメント化プロセスから得られる分解レンダリングは、過剰にセグメント化されるかまたは断片化され、複数の不連続のセグメントまたはウェーブレットを含む。セ

グメント化によって生成される不連続のセグメントまたはウェーブレットは、実質的に異なるときがあり、セグメント化によって生成される不連続のセグメントまたはウェーブレットは、実質的に類似であるときがある。実質的に類似の不連続のセグメントまたはウェーブレット（例えば、実質的に類似のレベル）とは、セグメント化されたプロファイルにおける2つまたはそれを超える隣接した不連続のセグメントまたはウェーブレットのことを指すことが多く、その不連続のセグメントまたはウェーブレットの各々は、不確定性の所定のレベル未満だけ異なるゲノム区分レベル（例えば、レベル）を有する。いくつかの実施形態において、実質的に類似の不連続のセグメントまたはウェーブレットは、互いに隣接しており、介在性のセグメントまたはウェーブレットによって分断されていない。いくつかの実施形態において、実質的に類似の不連続のセグメントまたはウェーブレットは、1つまたはそれを超えるより小さいセグメントまたはウェーブレットによって分断されている。いくつかの実施形態において、実質的に類似の不連続のセグメントまたはウェーブレットは、約1～約20個、約1～約15個、約1～約10個または約1～約5個の部分（例えば、ビン）によって分断され、ここで、それらの介在性部分の1つまたはそれを超えるものが、その実質的に類似の不連続のセグメントまたはウェーブレットの各々のレベルと有意に異なるレベルを有する。いくつかの実施形態において、実質的に類似の不連続のセグメントまたはウェーブレットのレベルは、ある不確定性のレベルと約3倍未満、約2倍未満、約1倍未満または約0.5倍未満だけ異なる。実質的に類似の不連続のセグメントまたはウェーブレットは、いくつかの実施形態において、3MAD未満（例えば、3シグマ未満）、2MAD未満、1MAD未満または約0.5MAD未満だけ異なるゲノム区分レベルの中央値を含み、ここで、MADは、そのセグメントまたはウェーブレットの各々のゲノム区分レベルの中央値から算出される。実質的に異なる不連続のセグメントまたはウェーブレットは、いくつかの実施形態において、隣接しないか、あるいは10個もしくはそれを超える、15個もしくはそれを超えるまたは20個もしくはそれを超える部分によって分断されている。実質的に異なる不連続のセグメントまたはウェーブレットは、通常、実質的に異なるレベルを有する。ある特定の実施形態において、実質的に異なる不連続のセグメントまたはウェーブレットは、ある不確定性のレベルの約2.5倍超、約3倍超、約4倍超、約5倍超、約6倍超だけ異なるレベルを含む。実質的に異なる不連続のセグメントまたはウェーブレットは、いくつかの実施形態において、2.5MAD超（例えば、2.5シグマ超）、3MAD超、4MAD超、約5MAD超または約6MAD超だけ異なるゲノム区分レベルの中央値を含み、ここで、MADは、不連続のセグメントまたはウェーブレットの各々のゲノム区分レベルの中央値から算出される。

【0285】

いくつかの実施形態において、セグメント化プロセスは、あるプロファイルまたはそのセグメントにおける1つまたはそれを超える不連続のセグメントまたはウェーブレット（例えば、レベル）に対する、レベル（例えば、定量的値、例えば、平均値または中央値レベル）、不確定性のレベル（例えば、不確定値）、Z得点、Z値、p値などまたはそれらの組み合わせの決定（例えば、算出）を含む。いくつかの実施形態において、レベル（例えば、定量的値、例えば、平均値または中央値レベル）、不確定性のレベル（例えば、不確定値）、Z得点、Z値、p値などまたはそれらの組み合わせは、不連続のセグメントまたはウェーブレットに対して決定される（例えば、算出される）。

【0286】

いくつかの実施形態において、セグメント化は、1つのプロセスまたは複数のサブプロセスを含むプロセスによって達成され、それらの非限定的な例としては、分解生成プロセス（例えば、ウェーブレット分解生成プロセス）、閾値化、平準化、平滑化などまたはそれらの組み合わせが挙げられる。閾値化、平準化、平滑化などは、分解生成プロセスおよび/またはウェーブレット分解レンダリングプロセスとともに行われ得る。

アウトカム

【0287】

本明細書中に記載される方法は、サンプルに対する遺伝的変異（例えば、胎児の異数性

10

20

30

40

50

）の存在もしくは非存在の判定を提供し、それにより、アウトカムが提供され得る（例えば、それにより、遺伝的変異（例えば、胎児の異数性）の存在もしくは非存在を判定するアウトカムが提供され得る）。遺伝的変異は、参照に対して試験被験体のゲノムまたは遺伝情報の検出可能な変化をもたらす、遺伝情報（例えば、染色体、染色体のセグメント、多型領域、転座した領域、変更されたヌクレオチド配列などまたは前述のものの組み合わせ）の獲得、喪失および／または変更（例えば、重複、欠失、融合、挿入、突然変異、再編成、置換または異常なメチル化）を含むことが多い。遺伝的変異の存在もしくは非存在は、部分（例えば、カウント、参照ゲノムのゲノム部分のカウント）にマッピングされた配列リードの変換、解析および／または操作によって判定され得る。アウトカムの判定は、いくつかの実施形態において、妊婦由来の核酸の解析を含む。ある特定の実施形態において、妊婦から得られたカウント（例えば、正規化されたカウント、リード密度、リード密度プロファイル）に従って、アウトカムが判定され、ここで、それらのカウントは、その妊婦から得られた核酸からのカウントである。

10

【0288】

本明細書中に記載される方法は、胎児を有する妊婦由来のテストサンプルに対する、胎児の異数性（例えば、完全な染色体異数性、部分的な染色体異数性または分節性の染色体異常（例えば、モザイク現象、欠失および／または挿入））の存在もしくは非存在を判定するときがある。ある特定の実施形態において、本明細書中に記載される方法は、胎児を有する妊婦由来のサンプルに対して、正倍数性または正倍数性の欠如（非正倍数性）を検出する。本明細書中に記載される方法は、1本またはそれを超える染色体（例えば、13番染色体、18番染色体、21番染色体またはそれらの組み合わせ）またはそれらのセグメントに対するトリソミーを検出するときがある。

20

【0289】

いくつかの実施形態において、遺伝的変異（例えば、胎児の異数性）の存在もしくは非存在は、本明細書中に記載される方法、当該分野で公知の方法またはそれらの組み合わせによって判定される。遺伝的変異の存在もしくは非存在は、通常、参照ゲノムの部分にマッピングされた配列リードのカウントから判定される。

【0290】

参照からのリード密度は、テストサンプルを得た同じ妊婦由来の核酸サンプルに対するものであるときがある。ある特定の実施形態において、参照からのリード密度は、テストサンプルを得た女性とは異なる1人またはそれを超える妊婦由来の核酸サンプルに対するものである。いくつかの実施形態において、試験被験体由来の（form）第1の部分セットからのリード密度および／またはリード密度プロファイルは、第2の部分セットからのリード密度および／またはリード密度プロファイルと比較され、ここで、その第2の部分セットは、第1の部分セットとは異なる。いくつかの実施形態において、試験被験体由来の第1の部分セットからのリード密度および／またはリード密度プロファイルは、第2の部分セットからのリード密度および／またはリード密度プロファイルと比較され、ここで、その第2の部分セットは、その試験被験体由来であるか、またはその試験被験体ではない参照被験体由来である。非限定的な例において、第1の部分セットが、21番染色体またはそのセグメントにおけるものである場合、第2の部分セットは、別の染色体（例えば、1番染色体、13番染色体、14番染色体、18番染色体、19番染色体、それらのセグメントまたは前述のものの組み合わせ）におけるものであることが多い。参照は、通常正倍数性である染色体またはそのセグメントに位置することが多い。例えば、1番染色体および19番染色体の異数性に関連する早期胎児死亡率が高いことから、1番染色体および19番染色体は、胎児において正倍数性であることが多い。試験被験体由来のリード密度および／またはリード密度プロファイルと参照との間の不確定性の尺度が、生成および／または比較され得る。遺伝的変異（例えば、胎児の異数性）の存在もしくは非存在は、試験被験体由来のリード密度および／またはリード密度プロファイルと参照とを比較せずに判定されるときがある。

30

40

【0291】

50

ある特定の実施形態において、参照は、試験被験体と同じ部分セットに対するリード密度および／またはリードプロファイルを含み、ここで、その参照に対するリード密度は、1つまたはそれを超える参照サンプル由来である（例えば、複数の参照被験体由来の複数の参照サンプル由来であることが多い）。参照サンプルは、テストサンプルを得た女性とは異なる1人またはそれを超える妊婦由来であることが多い。

【0292】

試験被験体および／または参照のリード密度および／またはリードプロファイルに対する不確定性の尺度が生成され得る。いくつかの実施形態において、試験被験体のリード密度および／またはリードプロファイルに対する不確定性の尺度が決定される。いくつかの実施形態において、参照被験体のリード密度および／またはリードプロファイルに対する不確定性の尺度が決定される。いくつかの実施形態において、リード密度プロファイル全体またはリード密度プロファイル内の部分のサブセットから不確定性の尺度が決定される。

10

【0293】

いくつかの実施形態において、参照サンプルは、ゲノムの選択されたセグメントに対して正倍数性であり、その選択されたセグメントに対して試験プロファイルと参照プロファイルとの間の不確定性の尺度が評価される。いくつかの実施形態において、遺伝的変異の存在もしくは非存在の判定は、ゲノム（例えば、染色体またはそのセグメント）の選択されたセグメントに対する試験プロファイルと参照プロファイルとの間の偏差の数値（例えば、偏差の尺度、MAD）に従う。いくつかの実施形態において、試験プロファイルと参照プロファイルとの間の偏差の数値が、約1より大きい、約1.5より大きい、約2より大きい、約2.5より大きい、約2.6より大きい、約2.7より大きい、約2.8より大きい、約2.9より大きい、約3より大きい、約3.1より大きい、約3.2より大きい、約3.3より大きい、約3.4より大きい、約3.5より大きい、約4より大きい、約5より大きい、または約6より大きいとき、遺伝的変異が存在すると判定される。例えば、試験プロファイルおよび参照プロファイルは、3という偏差の尺度（例えば、3シグマ、3MAD）を超えて異なるときがあり、遺伝的変異が存在すると判定される。いくつかの実施形態において、妊婦から得られた試験プロファイルは、3より大きい偏差の尺度（例えば、3シグマ、3MAD）だけ参照プロファイルより大きく、胎児の染色体異数性（例えば、胎児のトリソミー）が存在すると判定される。試験プロファイルと参照プロファイルとの間の3より大きい偏差は、ゲノムの選択されたセグメントに対して、試験被験体が非正倍数性であること（例えば、遺伝的変異が存在すること）を示唆することが多い。ゲノムの選択されたセグメントに対する参照プロファイル（この参照は、選択されたセグメントに対して正倍数性である）よりも有意に大きい試験プロファイルは、トリソミーを判定するときがある。いくつかの実施形態において、妊婦から得られたリード密度プロファイルは、3という偏差の尺度（例えば、3シグマ、3MAD）を超えて、選択されたセグメントに対する参照プロファイルより小さく、胎児の染色体異数性（例えば、胎児のモノソミー）が存在すると判定される。参照プロファイルより有意に小さい試験プロファイル（この参照プロファイルは正倍数性を示唆する）は、モノソミーを判定するときがある。

20

30

40

【0294】

いくつかの実施形態において、ゲノムの選択されたセグメントに対する試験プロファイルと参照プロファイルとの間の偏差の数値が、約3.5未満、約3.4未満、約3.3未満、約3.2未満、約3.1未満、約3.0未満、約2.9未満、約2.8未満、約2.7未満、約2.6未満、約2.5未満、約2.0未満、約1.5未満または約1.0未満であるとき、遺伝的変異が存在しないと判定される。例えば、試験プロファイルは、3という偏差の尺度（例えば、3シグマ、3MAD）未満だけ参照プロファイルと異なるときがあり、遺伝的変異が存在しないと判定される。いくつかの実施形態において、妊婦から得られた試験プロファイルは、3という偏差の尺度（例えば、3シグマ、3MAD）未満だけ参照プロファイルと異なり、胎児の染色体異数性が存在しない（例えば、胎児の正倍

50

数性)と判定される。いくつかの実施形態において、(例えば、試験プロファイルと参照プロファイルとの間の3未満の偏差(例えば、標準偏差に対する3シグマ)は、正倍数性である(例えば、遺伝的変異が存在しない)ゲノムのセグメントを示唆することが多い。テストサンプルに対する試験プロファイルと1つまたはそれを超える参照被験体に対する参照プロファイルとの間の偏差の尺度が、プロットされ、可視化され得る(例えば、z得点プロット)。

【0295】

他の任意の好適な参照が、テストサンプルの試験領域(例えば、試験されるゲノムのセグメント)に対する遺伝的変異の存在もしくは非存在の判定(または正倍数性もしくは非正倍数性の判定)のために、試験プロファイルとともに考慮され得る。いくつかの実施形態において、胎児分率の測定は、遺伝的変異の存在もしくは非存在を判定するために、配列リード(例えば、リード密度)のカウントとともに考慮され得る。例えば、リード密度および/またはリード密度プロファイルは、比較および/またはアウトカムの判定の前に、胎児分率に従って正規化され得る。胎児分率を数値化するための好適なプロセスを用いることができ、その非限定的な例としては、質量分析プロセス、配列決定プロセスまたはそれらの組み合わせが挙げられる。

【0296】

いくつかの実施形態において、遺伝的変異(例えば、胎児の異数性)の存在もしくは非存在の判定は、コールゾーン(call zone)に従って判定される。ある特定の実施形態において、ある値(例えば、リード密度プロファイルおよび/または不確定性の尺度)または値の集合が、予め定義された範囲(例えば、ゾーン、コールゾーン)に入るとき、コールが行われる(例えば、遺伝的変異の存在もしくは非存在、例えば、アウトカムを判定するコール)。いくつかの実施形態において、コールゾーンは、同じ患者サンプルから得られた値の集合(例えば、リード密度プロファイルおよび/または不確定性の尺度)に従って定義される。ある特定の実施形態において、コールゾーンは、同じ染色体またはそのセグメントから得られた値の集合に従って定義される。いくつかの実施形態において、遺伝的変異の判定に基づくコールゾーンは、不確定性の尺度(例えば、高い信頼水準、例えば、低い不確定性の尺度)および/または胎児分率に従って定義される。

【0297】

いくつかの実施形態において、コールゾーンは、遺伝的変異の判定、および約2.0%もしくはそれを超えるか、約2.5%もしくはそれを超えるか、約3%もしくはそれを超えるか、約3.25%もしくはそれを超えるか、約3.5%もしくはそれを超えるか、約3.75%もしくはそれを超えるか、または約4.0%もしくはそれを超える胎児分率に従って定義される。例えば、いくつかの実施形態において、試験プロファイルを得たテストサンプルが、胎児を有する妊婦から得られたテストサンプルに対して2%もしくはそれを超えるかまたは4%もしくはそれを超える胎児分率の判定を含む場合、試験プロファイルと参照プロファイルとの比較に基づいて、胎児がトリソミー21を含むというコールが行われる。例えば、いくつかの実施形態において、試験プロファイルを得たテストサンプルが、胎児を有する妊婦から得られたテストサンプルに対して2%もしくはそれを超えるかまたは4%もしくはそれを超える胎児分率の判定を含む場合、試験プロファイルと参照プロファイルとの比較に基づいて、胎児が正倍数性であるというコールが行われる。いくつかの実施形態において、コールゾーンは、約99%もしくはそれを超えるか、約99.1%もしくはそれを超えるか、約99.2%もしくはそれを超えるか、約99.3%もしくはそれを超えるか、約99.4%もしくはそれを超えるか、約99.5%もしくはそれを超えるか、約99.6%もしくはそれを超えるか、約99.7%もしくはそれを超えるか、約99.8%もしくはそれを超えるか、または約99.9%もしくはそれを超える信頼水準によって定義される。いくつかの実施形態において、コールゾーンを用いずに、コールが行われる。いくつかの実施形態において、コールゾーンおよびさらなるデータまたは情報を用いて、コールが行われる。いくつかの実施形態において、コールゾーンを使用しない比較に基づいて、コールが行われる。いくつかの実施形態において、プロファイル

10

20

30

40

50

の目視検査（例えば、リード密度の目視検査）に基づいて、コールが行われる。

【0298】

いくつかの実施形態において、無コールゾーン（no-call zone）は、コールが行われないゾーンである。いくつかの実施形態において、無コールゾーンは、精度が低いこと、リスクが高いこと、誤差が大きいこと、信頼水準が低いこと、不確定性の尺度が大きいことなどまたはそれらの組み合わせを示唆する値または値の集合によって定義される。いくつかの実施形態において、無コールゾーンは、約5%もしくはそれ未満、約4%もしくはそれ未満、約3%もしくはそれ未満、約2.5%もしくはそれ未満、約2.0%もしくはそれ未満、約1.5%もしくはそれ未満または約1.0%もしくはそれ未満の胎児分率によって部分的に定義される。

10

【0299】

遺伝的変異は、医学的症状に関連するときがある。遺伝的変異を判定するアウトカムは、症状（例えば、医学的症状）、疾患、症候群もしくは異常の存在もしくは非存在を判定するアウトカムであるときがあるか、または症状、疾患、症候群もしくは異常（例えば、表1に列挙される非限定的な例）の検出を含む。ある特定の実施形態において、診断は、アウトカムの評価を含む。本明細書中に記載される方法によって症状（例えば、医学的症状）、疾患、症候群または異常の存在もしくは非存在を判定するアウトカムは、独立して、さらなる検査（例えば、核型分析および／または羊水穿刺）によって確認され得るときがある。データの解析および処理は、1つまたはそれを超えるアウトカムを提供し得る。用語「アウトカム」は、本明細書中で使用されるとき、遺伝的変異（例えば、異数性、コピー数変異）の存在もしくは非存在の判定を容易にする、データ処理の結果のことを指し得る。ある特定の実施形態において、用語「アウトカム」は、本明細書中で使用されるとき、遺伝的変異（例えば、異数性、コピー数変異）の存在もしくは非存在を予測するおよび／または判定する結論のことを指す。ある特定の実施形態において、用語「アウトカム」は、本明細書中で使用されるとき、被験体（例えば、胎児）における遺伝的変異（例えば、異数性、コピー数変異）の存在もしくは非存在のリスクまたは確率を予測するおよび／または判定する結論のことを指す。診断は、アウトカムの使用を含むときがある。例えば、健康の専門家（health practitioner）は、アウトカムを解析し得、そのアウトカムに基づいてまたは部分的に基づいて診断を提供し得る。いくつかの実施形態において、症状、症候群または異常（例えば、表1に列挙されるもの）の判定、検出または診断は、遺伝的変異の存在もしくは非存在を判定するアウトカムの使用を含む。いくつかの実施形態において、マッピングされてカウントされた配列リードまたはその変換に基づくアウトカムは、遺伝的変異の存在もしくは非存在を判定する。ある特定の実施形態において、本明細書中に記載される1つまたはそれを超える方法（例えば、データ処理方法）を用いて生成されたアウトカムは、表1に列挙される1つまたはそれを超える症状、症候群または異常の存在もしくは非存在を判定する。ある特定の実施形態において、診断は、症状、症候群または異常の存在もしくは非存在の判定を含む。診断は、症状、症候群または異常の性質および／または原因として、遺伝的変異の判定を含むことが多い。ある特定の実施形態において、アウトカムは、診断ではない。アウトカムは、確率に関する1つまたはそれを超える考慮すべきものの文脈において、本明細書中に記載される処理方法を用いて生成された1つまたはそれを超える数値を含むことが多い。リスクまたは確率に関する考慮すべきものとしては、不確定性の尺度、信頼水準、感度、特異性、標準偏差、変動係数（CV）および／もしくは信頼水準、Z得点、カイ値、ファイ値、倍数性値、当てはめられた胎児分率、面積比、中央値レベルなどまたはそれらの組み合わせが挙げられ得るが、これらに限定されない。確率に関する考慮すべきものは、被験体が、遺伝的変異を有するリスクがあるかまたは遺伝的変異を有するかを判定することを容易にし得、遺伝的障害の存在もしくは非存在を判定するアウトカムは、そのような考慮すべきものを含むことが多い。

20

30

40

【0300】

アウトカムは、表現型であるときがある。アウトカムは、関連する信頼水準を有する表

50

現型であるときがある（例えば、不確定性の尺度、例えば、胎児は、99%の信頼水準でトリソミー21が陽性であり、試験被験体は、95%の信頼水準で遺伝的変異に関連する癌が陰性である）。アウトカムの値を生成する種々の方法が、種々のタイプの結果をもたらし得るときがある。通常、本明細書中に記載される方法を用いて生成されたアウトカムの値に基づいて生成され得る、あり得るスコアまたはコールには4タイプある：真陽性、偽陽性、真陰性および偽陰性。用語「スコア（score）」、「スコア（scores）」、「コール（call）」および「コール（calls）」は、本明細書中で使用されるとき、特定の遺伝的変異が被験体／サンプルに存在するかまたは存在しない確率を算出することを指す。あるスコアの値は、例えば、遺伝的変異に対応し得る、マッピングされた配列リードの変動、差異または比率を決定するために使用され得る。例えば、参照ゲノムに対して、あるデータセットからの選択された遺伝的変異または部分について陽性のスコアを算出することによって、遺伝的変異の存在もしくは非存在が特定され得、その遺伝的変異は、医学的症状（例えば、癌、子癇前症、トリソミー、モノソミーなど）に関連するときがある。いくつかの実施形態において、アウトカムは、リード密度、リード密度プロファイルおよび／またはプロット（例えば、プロファイルプロット）を含む。アウトカムが、あるプロファイルを含むそれらの実施形態において、好適なプロファイルまたはプロファイルの組み合わせが、アウトカムのために使用され得る。アウトカムのために使用され得るプロファイルの非限定的な例としては、z 得点プロファイル、p 値プロファイル、カイ値プロファイル、ファイ値プロファイルなどおよびそれらの組み合わせが挙げられる。

10

20

【0301】

遺伝的変異の存在もしくは非存在を判定するために生成されたアウトカムは、無意味な結果（null result）（例えば、2つのクラスターの間のデータポイント、遺伝的変異の存在と非存在の両方に対する値を包含する標準偏差を伴う数値、調査されている遺伝的変異を有するかまたは有しない被験体に対するプロファイルプロットに似ていないプロファイルプロットを有するデータセット）を含むときがある。いくつかの実施形態において、無意味な結果を示唆するアウトカムは、なおも確定的な結果であり、その判定は、遺伝的変異の存在もしくは非存在を判定するためのさらなる情報および／またはデータ生成の繰り返しおよび／または解析の必要性を含み得る。

【0302】

アウトカムは、いくつかの実施形態において、本明細書中に記載される1つまたはそれを超える処理工程を行った後に生成され得る。ある特定の実施形態において、アウトカムは、本明細書中に記載される処理工程の1つの結果として生成され、いくつかの実施形態において、アウトカムは、データセットの各統計的操作および／または各数学的操作が行われた後に生成され得る。遺伝的変異の存在もしくは非存在の判定に関するアウトカムは、好適な形態で表現され得、その形態としては、被験体またはサンプルに対する遺伝的変異の存在もしくは非存在に関連する、確率（例えば、オッズ比、p 値）、尤度、クラスターの中もしくは外の値、閾値を上回るもしくは下回る値、範囲（例えば、しきい値範囲）内の値、分散もしくは信頼の尺度を伴う値、または危険因子が挙げられるが、これらに限定されない。ある特定の実施形態において、サンプル間の比較は、サンプルの同一性の確認を可能にする（例えば、繰り返されたサンプルおよび／または混同された（例えば、誤って標識された、混合されたなど）サンプルの識別を可能にする）。

30

40

【0303】

いくつかの実施形態において、アウトカムは、所定の閾値もしくはカットオフ値を上回るもしくは下回る値、および／またはその値に関連する不確定性の尺度もしくは信頼水準を含む。ある特定の実施形態において、所定の閾値またはカットオフ値は、予想されるレベルまたは予想されるレベル範囲である。アウトカムは、データ処理において使用される仮定も説明し得る。ある特定の実施形態において、アウトカムは、値の所定の範囲（例えば、しきい値範囲）に入るかまたは入らない値、およびその範囲の内側または外側の値に対する関連する不確定性または信頼水準を含む。いくつかの実施形態において、アウトカ

50

ムは、所定の値に等しい（例えば、1に等しい、ゼロに等しい）かまたは所定の値の範囲内の値に等しい値、およびある範囲に等しいか、範囲内または範囲外である値に対するその関連する不確定性または信頼水準を含む。アウトカムは、プロット（例えば、プロファイルプロット）としてグラフを用いて表されるときがある。

【0304】

上で述べたように、アウトカムは、真陽性、真陰性、偽陽性または偽陰性として特徴づけられ得る。用語「真陽性」は、本明細書中で使用されるとき、遺伝的変異を有すると正しく診断された被験体のことを指す。用語「偽陽性」は、本明細書中で使用されるとき、遺伝的変異を有すると誤って識別された被験体のことを指す。用語「真陰性」は、本明細書中で使用されるとき、遺伝的変異を有しないと正しく識別された被験体のことを指す。用語「偽陰性」は、本明細書中で使用されるとき、遺伝的変異を有しないと誤って識別された被験体のことを指す。任意の所与の方法に対する性能の2つの尺度は、これらの発生の比率に基づいて算出され得る：（i）一般に、陽性であると正しく識別される、予測される陽性の割合である感度値；および（ii）一般に、陰性であると正しく識別される、予測される陰性の割合である特異性値。

【0305】

ある特定の実施形態において、感度、特異性および／または信頼水準の1つまたはそれを超えるものが、パーセンテージとして表現される。いくつかの実施形態において、そのパーセンテージは、各変数に対して独立して、約90%超（例えば、約90、91、92、93、94、95、96、97、98もしくは99%または99%超（例えば、約99.5%またはそれを超える、約99.9%またはそれを超える、約99.95%またはそれを超える、約99.99%またはそれを超える））である。変動係数（CV）は、いくつかの実施形態において、パーセンテージとして表現され、そのパーセンテージは、約10%またはそれ未満（例えば、約10、9、8、7、6、5、4、3、2もしくは1%または1%未満（例えば、約0.5%またはそれ未満、約0.1%またはそれ未満、約0.05%またはそれ未満、約0.01%またはそれ未満））であるときがある。確率（例えば、特定のアウトカムが偶然に起因しない確率）は、ある特定の実施形態において、Z得点、p値、またはt検定の結果として表現される。いくつかの実施形態において、あるアウトカムに対する計測された分散、信頼区間、感度、特異性など（例えば、集散的に信頼パラメータと称される）は、本明細書中に記載される1つまたはそれを超えるデータ処理操作を用いて生成され得る。アウトカムおよび関連する信頼水準を生成する具体例は、実施例の項および国際特許出願番号PCT/US12/59123（WO2013/052913）（本文、表、式および図面のすべてを含むこの全内容が参照により本明細書中に組み込まれる）に記載されている。

【0306】

用語「感度」は、本明細書中で使用されるとき、真陽性の数＋偽陰性の数で除算された真陽性の数のことを指し、ここで、感度（sens）は、 $0 \leq \text{sens} \leq 1$ の範囲内であり得る。用語「特異性」は、本明細書中で使用されるとき、真陰性の数＋偽陽性の数で除算された真陰性の数のことを指し、ここで、感度（spec）は、 $0 \leq \text{spec} \leq 1$ の範囲内であり得る。いくつかの実施形態において、1、すなわち100%に等しいかまたは1に近い（例えば、約90%～約99%の）感度および特異性を有する方法が選択されるときがある。いくつかの実施形態において、1すなわち100%に等しい感度を有する方法が選択され、ある特定の実施形態において、1に近い感度（例えば、約90%の感度、約91%の感度、約92%の感度、約93%の感度、約94%の感度、約95%の感度、約96%の感度、約97%の感度、約98%の感度または約99%の感度）を有する方法が選択される。いくつかの実施形態において、1すなわち100%に等しい特異性を有する方法が選択され、ある特定の実施形態において、1に近い特異性（例えば、約90%の特異性、約91%の特異性、約92%の特異性、約93%の特異性、約94%の特異性、約95%の特異性、約96%の特異性、約97%の特異性、約98%の特異性または約99%の特異性）を有する方法が選択される。

【0307】

いくつかの実施形態において、遺伝的変異の存在もしくは非存在（例えば、染色体異数性）は、胎児に対して判定される。そのような実施形態では、胎児の遺伝的変異（例えば、胎児の染色体異数性）の存在もしくは非存在が判定される。

【0308】

ある特定の実施形態において、遺伝的変異（例えば、染色体異数性）の存在もしくは非存在は、サンプルに対して判定される。そのような実施形態では、サンプル核酸における遺伝的変異（例えば、染色体異数性）の存在もしくは非存在が判定される。いくつかの実施形態において、検出された変動または検出されない変動は、1つの供給源由来のサンプル核酸に存在するが、別の供給源由来のサンプル核酸には存在しない。供給源の非限定的な例としては、胎盤の核酸、胎児の核酸、母体の核酸、癌細胞の核酸、非癌細胞の核酸などおよびそれらの組み合わせが挙げられる。非限定的な例において、検出されるまたは検出されない特定の遺伝的変異は、（i）胎盤の核酸に存在するが、胎児核酸には存在せず、母体核酸にも存在しないか；（i i）胎児核酸に存在するが、母体の核酸には存在しないか；または（i i i）母体の核酸に存在するが、胎児の核酸には存在しない。

【0309】

遺伝的変異および／または関連する医学的症状の存在もしくは非存在（例えば、アウトカム）は、アウトカムモジュールによって提供されることが多い。遺伝的変異（例えば、異数性、胎児の異数性、コピー数変異）の存在もしくは非存在は、いくつかの実施形態において、アウトカムモジュールまたはアウトカムモジュールを備える機器によって識別される。アウトカムモジュールは、特定の遺伝的変異（例えば、トリソミー、トリソミー21、トリソミー18）を判定するために特殊化され得る。例えば、トリソミー21を識別するアウトカムモジュールは、トリソミー18を識別するアウトカムモジュールとは異なり得、かつ／または別個のものであり得る。いくつかの実施形態において、アウトカムモジュールまたはアウトカムモジュールを備える機器は、遺伝的変異または遺伝的変異（例えば、異数性、コピー数変異）を判定するアウトカムを識別するために必要とされる。ある特定の実施形態において、アウトカムは、アウトカムモジュールから表示モジュールに移され、ここで、アウトカムは、表示モジュールによって提供される。

【0310】

本明細書中に記載される方法によって識別される遺伝的変異または遺伝的変異を判定するアウトカムは、独立して、さらなる検査（例えば、母体および／または胎児の核酸の標的化された配列決定）によって確かめられ得る。アウトカムは、通常、ヘルスケアの専門家（例えば、検査技師または管理者；医師または補助者）に提供される。ある特定の実施形態において、アウトカムは、好適な視覚媒体（例えば、機器の周辺装置または構成要素、例えば、プリンターまたはディスプレイ）上に提供される。いくつかの実施形態において、遺伝的変異の存在もしくは非存在を判定するアウトカムは、ヘルスケアの専門家にレポートの形態で提供され、ある特定の実施形態において、そのレポートは、アウトカム値および関連する信頼パラメータの表示を含む。一般に、アウトカムは、遺伝的変異および／または医学的症状の存在もしくは非存在の判定を容易にする好適な形式で表示され得る。データセットを報告するためおよび／もしくは表示するためまたはアウトカムを報告するための使用に適した形式の非限定的な例としては、デジタルデータ、グラフ、2Dグラフ、3Dグラフおよび4Dグラフ、写真（例えば、jpg、bitmap（例えば、bmp）、pdf、tiff、gif、raw、pngなどまたは好適な形式）、ピクトグラフ、チャート、表、棒グラフ、円グラフ、ダイアグラム、フローチャート、散布図、マップ、ヒストグラム、密度チャート、関数グラフ、回路図、ブロック図、バブルマップ、信号空間ダイアグラム、等高線図、統計地図、スパイダーチャート、ベン図、ノモグラムなどおよび前述のものの組み合わせが挙げられる。

【0311】

アウトカムの生成は、ある特定の実施形態において、核酸配列リードデータなどから、被験体の細胞性核酸の提示への変換とみなされ得る。例えば、被験体由来の核酸の配列リ

10

20

30

40

50

ードの解析、ならびに染色体プロファイルおよび／またはアウトカムの生成は、比較的小さい配列リードフラグメントから比較的大きい染色体構造の提示への変換とみなされ得る。いくつかの実施形態において、アウトカムは、被験体（例えば、妊婦）由来の配列リードから、その被験体に存在する既存の構造（例えば、ゲノム、染色体またはそのセグメント）（例えば、母体のおよび／または胎児の核酸）の提示への変換から生じる。いくつかの実施形態において、アウトカムは、第1の被験体（例えば、妊婦）からの配列リードから、構造（例えば、ゲノム、染色体またはそのセグメント）の複合的な提示への変換、ならびに第1の被験体（例えば、妊婦）および／または第2の被験体（例えば、胎児）に存在する構造の提示をもたらすその複合的な提示の第2の変換を含む。

アウトカムの使用

10

【0312】

遺伝的変異の存在もしくは非存在を判定する1つまたはそれを超えるアウトカムを含むレポートを受け取るヘルスケアの専門家または他の資格のある個人は、試験被験体または患者の状態に関するコールを行うために、そのレポートの中に表示されたデータを使用し得る。ヘルスケアの専門家は、いくつかの実施形態において、提供されたアウトカムに基づいて勧告し得る。ヘルスケアの専門家または資格のある個人は、いくつかの実施形態において、レポートに提供されたアウトカム値および関連する信頼パラメータに基づいて、遺伝的変異の存在もしくは非存在に関するコールまたはスコアを試験被験体または患者に提供し得る。ある特定の実施形態において、スコアまたはコールは、提供されたレポートの目視観察によって、ヘルスケアの専門家または資格のある個人によって手作業で生成される。ある特定の実施形態において、スコアまたはコールは、ソフトウェアに組み込まれているときがある自動化されたルーチンによって生成され、試験被験体または患者に情報を提供する前に、ヘルスケアの専門家または資格のある個人によって精度について再検討される。用語「レポートを受け取る」は、本明細書中で使用されるとき、再検討の際にヘルスケアの専門家または他の資格のある個人が試験被験体または患者における遺伝的変異の存在もしくは非存在に関する判定を行うことを可能にする、アウトカムを含む書面および／またはグラフによる提示をコミュニケーション手段によって得ることを指す。そのレポートは、コンピュータ、または人間によるデータ入力によって生成され得、電子的手段を用いて（例えば、インターネットによって、コンピュータを介して、ファックスを介して、同じまたは異なる物理的位置における1つのネットワーク場所から別の場所へ）、またはデータを送受信する他の方法（例えば、メールサービス、クーリエサービスなど）によって、伝達され得る。いくつかの実施形態において、アウトカムは、口頭、文書またはファイル形態を含むがこれらに限定されない好適な媒体でヘルスケアの専門家に伝えられる。そのファイルは、例えば、音声ファイル、コンピュータ可読ファイル、紙のファイル、検査ファイル（laboratory file）または医療記録ファイルであり得るが、これらに限定されない。

20

30

【0313】

用語「アウトカムを提供する」およびその文法上の等価物は、本明細書中で使用されるとき、そのような情報を得るための方法のことも指し得、その方法には、検査室からの情報（例えば、検査ファイル）を得るための方法が含まれるが、これに限定されない。検査ファイルは、医学的症状の存在もしくは非存在を判定するために1つもしくはそれを超えるアッセイまたは1つもしくはそれを超えるデータ処理工程を行った検査室によって生成され得る。その検査室は、検査ファイルから医学的症状の存在もしくは非存在を識別する人員と同じ場所または異なる場所（例えば、別の国）に存在し得る。例えば、検査ファイルが、1つの場所において生成され、別の場所に伝えられ得、その別の場所において、そのファイルの中の情報が妊婦被験体に伝えられる。検査ファイルは、ある特定の実施形態において、有形の形態または電子的形態（例えば、コンピュータ可読形態）であり得る。

40

【0314】

いくつかの実施形態において、アウトカムは、検査室からヘルスケアの専門家、医師または資格のある個人に提供され得、そのヘルスケアの専門家、医師または資格のある個人

50

は、そのアウトカムに基づいて診断を行い得る。いくつかの実施形態において、アウトカムは、検査室からヘルスケアの専門家、医師または資格のある個人に提供され得、そのヘルスケアの専門家、医師または資格のある個人は、さらなるデータおよび／または情報ならびに他のアウトカムとともに、そのアウトカムに部分的に基づいて診断を行い得る。

【0315】

ヘルスケアの専門家または資格のある個人は、そのアウトカムまたはレポートの中に提供されているアウトカムに基づいて、好適な勧告を提供し得る。提供されたアウトカムのレポートに基づいて提供され得る勧告の非限定的な例としては、手術、放射線治療、化学療法、遺伝相談、産後の処置の解決策（例えば、生活設計、長期間にわたる補助介護、医薬、対症療法）、妊娠中絶、臓器移植、輸血などまたは前述のものの組み合わせが挙げられる。いくつかの実施形態において、勧告は、提供されるアウトカムに基づく分類（例えば、ダウン症候群、ターナー症候群、T13における遺伝的変異に関連する医学的症状、T18における遺伝的変異に関連する医学的症状）に依存する。

【0316】

検査室の人員（例えば、検査室の管理者）は、遺伝的変異の存在もしくは非存在の判定（または試験領域に対する正倍数性もしくは非正倍数性の判定）の基礎をなす値（例えば、試験プロファイル、参照プロファイル、偏差のレベル）を解析し得る。近いまたは疑わしい遺伝的変異の存在もしくは非存在に関するコールの場合、検査室の人員は、同じテストを再要求し得、かつ／または試験被験体由来の同じもしくは異なるサンプル核酸を利用した異なるテスト（例えば、胎児の異数性の判定の場合、核型分析および／または羊水穿刺）を要求し得る。

機器、ソフトウェアおよびインターフェース

【0317】

本明細書中に記載されるある特定のプロセスおよび方法（例えば、配列リード、カウント、レベル（例えば、レベル）および／またはプロファイルの定量、マッピング、正規化、範囲設定、調整、カテゴリー化、カウントおよび／または測定）は、コンピュータ、マイクロプロセッサ、ソフトウェア、モジュールまたは他の機器なしでは行うことができないことが多い。本明細書中に記載される方法は、通常、コンピュータによって履行される方法であり、ある方法の1つまたはそれを超える部分が、1つまたはそれを超えるプロセッサ（例えば、マイクロプロセッサ）、コンピュータ、またはマイクロプロセッサによって制御された機器によって行われるときがある。本文書に記載されている方法に関する実施形態は、通常、本明細書中に記載されるシステム、機器およびコンピュータプログラム製品における指示によって履行される同じまたは関連するプロセスに適用可能である。本文書に記載されている方法に関する実施形態は、通常、実行可能なプログラムを格納している非一時的なコンピュータ可読記憶媒体によって履行される、同じまたは関連するプロセスに適用可能であり得、そのプログラムは、マイクロプロセッサに上記方法またはその一部を行うように指示する。いくつかの実施形態において、本明細書中に記載されるプロセスおよび方法（例えば、配列リード、カウント、レベルおよび／またはプロファイルを定量する、カウントする、および／または決定する方法）は、自動化された方法によって行われる。いくつかの実施形態において、1つまたはそれを超える工程および本明細書中に記載される方法は、マイクロプロセッサおよび／もしくはコンピュータによって行われ、かつ／またはメモリーと連動して行われる。いくつかの実施形態において、自動化された方法は、配列リード、カウント、マッピング、マッピングされた配列タグ、レベル、プロファイル、正規化、比較、範囲設定、カテゴリー化、調整、プロット、アウトカム、変換および識別を明らかにする、ソフトウェア、モジュール、マイクロプロセッサ、周辺装置および／または同様のものを備える機器に組み込まれている。本明細書中で使用されるとき、ソフトウェアとは、マイクロプロセッサによって実行されたとき、本明細書中に記載されるようなコンピュータ操作を行うコンピュータ可読プログラムの指示のことを指す。

【0318】

試験被験体（例えば、患者、妊婦）および／または参照被験体から得られた配列リード、カウント、レベルおよびプロファイルは、コピー数変異の存在もしくは非存在を判定するために、さらに解析され、処理され得る。配列リード、カウント、レベルおよび／またはプロファイルは、「データ」または「データセット」と称されるときがある。いくつかの実施形態において、データまたはデータセットは、1つまたはそれを超える特徴または変数（例えば、配列に基づくもの〔例えば、GC含有量、特定のヌクレオチド配列など〕、機能特異的なもの〔例えば、発現された遺伝子、癌遺伝子など〕、場所に基づくもの〔ゲノム特異的、染色体特異的、部分または部分特異的〕などおよびそれらの組み合わせ）によって特徴づけられ得る。ある特定の実施形態において、データまたはデータセットは、1つまたはそれを超える特徴または変数に基づいて、2次元またはそれを超える次元を有する行列に配置され得る。行列に配置されるデータは、任意の好適な特徴または変数を用いて配置され得る。行列の中のデータの非限定的な例としては、母体の年齢、母体の倍数性および胎児の寄与ごとに配置されるデータが挙げられる。ある特定の実施形態において、1つまたはそれを超える特徴または変数によって特徴づけられるデータセットは、カウントした後に処理されるときがある。

10

【0319】

本明細書中に記載される方法を行うために、機器、ソフトウェアおよびインターフェースが使用され得る。機器、ソフトウェアおよびインターフェースを使用するとき、ユーザーは、例えば、統計解析アルゴリズム、統計的有意性アルゴリズム、統計的アルゴリズム、繰り返し工程、検証アルゴリズムおよびグラフ表示の履行を含み得る、特定の情報、プログラムまたはプロセスを使用するためのオプションを入力し得るか、要求し得るか、照会し得るか、または決定し得る（例えば、配列リードのマッピング、マッピングされたデータの処理および／またはアウトカムの提供）。いくつかの実施形態において、データセットは、入力情報としてユーザーによって入力され得、ユーザーは、好適なハードウェア媒体（例えば、フラッシュドライブ）によって1つまたはそれを超えるデータセットをダウンロードし得、および／またはユーザーは、その後の処理および／またはアウトカムの提供のために1つのシステムから別のシステムにデータセットを送信し得る（例えば、配列リードのマッピングのために、シークエンサーからコンピュータシステムに配列リードデータを送信し得る；アウトカムおよび／またはレポートの処理および生成のために、マッピングされた配列データをコンピュータシステムに送信し得る）。

20

30

【0320】

システムは、通常、1つまたはそれを超える機器を備える。各機器は、1つまたはそれを超えるメモリー、1つまたはそれを超えるマイクロプロセッサ、および指示を備える。あるシステムが、2つまたはそれを超える機器を備える場合、それらの機器のいくつかまたはすべてが、同じ場所に存在してもよいし、それらの機器のいくつかまたはすべてが、異なる場所に存在してもよいし、それらの機器のすべてが、1つの場所に存在してもよいし、かつ／またはそれらの機器のすべてが、異なる場所に存在してもよい。あるシステムが、2つまたはそれを超える機器を備える場合、それらの機器のいくつかまたはすべてが、ユーザーと同じ場所に存在してもよいし、それらの機器のいくつかまたはすべてが、ユーザーと異なる場所に存在してもよいし、それらの機器のすべてが、ユーザーと同じ場所に存在してもよいし、かつ／またはそれらの機器のすべてが、ユーザーと異なる1つもしくはそれを超える場所に存在してもよい。

40

【0321】

システムは、計算機および配列決定装置または配列決定機器を備えるときがあり、その配列決定装置または配列決定機器は、物理的な核酸を受け取って配列リードを生成するように構成されており、計算装置は、配列決定装置または配列決定機器からのリードを処理するように構成されている。その計算機は、配列リードから遺伝的変異の存在もしくは非存在（例えば、コピー数変異；胎児の染色体異数性）を判定するように構成されているときがある。

【0322】

50

ユーザーは、例えば、ソフトウェアにクエリーを入れ得、次いで、そのソフトウェアは、インターネットアクセスを介してデータセットを取得し得、ある特定の実施形態では、プログラマブルマイクロプロセッサが、所与のパラメータに基づいて好適なデータセットを取得するように促され得る。プログラマブルマイクロプロセッサはまた、所与のパラメータに基づいてそのマイクロプロセッサによって選択された1つまたはそれを超えるデータセットオプションを選択するようにユーザーに促し得る。プログラマブルマイクロプロセッサは、インターネットを介して見つけ出された情報、他の内部情報または外部情報などに基づいてそのマイクロプロセッサによって選択された1つまたはそれを超えるデータセットオプションを選択するようにユーザーに促し得る。オプションは、方法、機器、装置、コンピュータプログラム、または実行可能なプログラムが格納されている非一時的なコンピュータ可読記憶媒体の、1つまたはそれを超えるデータ特徴の選択、1つまたはそれを超える統計的アルゴリズム、1つまたはそれを超える統計解析アルゴリズム、1つまたはそれを超える統計的有意性アルゴリズム、繰り返し工程、1つまたはそれを超える検証アルゴリズム、および1つまたはそれを超えるグラフ表示を選択するために選択され得る。

10

【0323】

本明細書中で述べられるシステムは、コンピュータシステムの一般的な構成要素（例えば、ネットワークサーバー、ラップトップ型システム、デスクトップ型システム、手持ち式システム、携帯情報端末、コンピューティングキオスクなど）を備え得る。コンピュータシステムは、1つまたはそれを超える入力手段、例えば、キーボード、タッチスクリーン、マウス、音声認識手段、またはユーザーがシステムにデータを入力することを可能にする他の手段を備え得る。システムは、表示画面（例えば、CRTまたはLCD）、スピーカー、FAX機器、プリンター（例えば、レーザー、インクジェット、インパクト、モノクロまたはカラープリンター）、または情報（例えば、アウトカムおよび／またはレポート）の視覚的、聴覚的および／もしくはハードコピー出力を提供するのに有用な他の出力を含むがこれらに限定されない1つまたはそれを超える出力をさらに備え得る。

20

【0324】

あるシステムにおいて、他の構成要素の中でもプログラムの指示を実行するためのマイクロプロセッサならびにプログラムコードおよびデータを格納するためのメモリーを備え得る中央処理装置に入力手段および出力手段が接続されていることがある。いくつかの実施形態において、プロセスは、単一の地理的位置に配置されたシングルユーザーシステムとして履行され得る。ある特定の実施形態において、プロセスは、マルチユーザーシステムとして履行され得る。マルチユーザーによる履行の場合、複数の中央処理装置が、ネットワークによって接続され得る。そのネットワークは、ある建物の一部分における単一の部門、建物全体を包含するローカルであり得るか、複数の建物にまたがり得るか、ある領域にまたがり得るか、一国全体にまたがり得るか、または世界中に及び得る。そのネットワークは、プライベートであり得、プロバイダーに所有され、制御されているか、またはユーザーがウェブページにアクセスして情報を入力して検索する、インターネットに基づくサービスとして履行され得る。したがって、ある特定の実施形態において、システムは、ユーザーに関してローカルまたはリモートであり得る1つまたはそれを超える機器を備える。1つの場所または複数の場所における1つより多い機器は、1人のユーザーによってアクセスされ得、データは、連続しておよび／または並行して、マッピングされ得るおよび／または処理され得る。したがって、好適な配置および制御は、複数の機器（例えば、ローカルネットワーク、リモートネットワークおよび／または「クラウド」コンピューティングプラットフォームにおけるもの）を用いてデータをマッピングするためおよび／または処理するために使用され得る。

30

40

【0325】

いくつかの実施形態において、システムは、通信用インターフェースを備え得る。通信用インターフェースは、コンピュータシステムと1つまたはそれを超える外部デバイスとの間のソフトウェアおよびデータの移動を可能にする。通信用インターフェースの非限定

50

的な例としては、モデム、ネットワークインターフェース（例えば、イーサネット（登録商標）カード）、通信ポート、P C M C I A スロットおよびカードなどが挙げられる。通信用インターフェースを介して移動されるソフトウェアおよびデータは、通常、信号の形態であり、その信号は、通信用インターフェースによって受け取られることが可能な電子信号、電磁信号、光学信号および／または他の信号であり得る。信号は、回線を介して通信用インターフェースに提供されることが多い。回線は、信号を運ぶことが多く、ワイヤもしくはケーブル、光ファイバー、電話回線、携帯電話回線、R F 回線および／または他の通信用回線を用いて構築され得る。したがって、一例において、通信用インターフェースは、信号検出モジュールによって検出され得る信号情報を受け取るために使用され得る。

10

【0326】

データは、好適なデバイスおよび／または方法によって入力され得、それらとしては、手動入力デバイスまたは直接データ入力デバイス（D D E）が挙げられるが、これらに限定されない。手動デバイスの非限定的な例としては、キーボード、コンセプトキーボード、タッチセンシティブスクリーン、ライトペン、マウス、トラックボール、ジョイスティック、グラフィックタブレット、スキャナー、デジタルカメラ、ビデオデジタイザおよび音声認識デバイスが挙げられる。D D E の非限定的な例としては、バーコードリーダー、磁気ストライプコード、スマートカード、磁気インキ文字認識、光学式文字認識、光学的マーク認識およびターンアラウンドドキュメントが挙げられる。

20

【0327】

いくつかの実施形態において、配列決定装置または配列決定機器からの出力は、入力デバイスを介して入力され得るデータとして役立ち得る。ある特定の実施形態において、マッピングされた配列リードが、入力デバイスを介して入力され得るデータとして役立ち得る。ある特定の実施形態において、核酸のフラグメントサイズ（例えば、長さ）が、入力デバイスを介して入力され得るデータとして役立ち得る。ある特定の実施形態において、核酸捕捉プロセスからの出力（例えば、ゲノム領域起源データ）が、入力デバイスを介して入力され得るデータとして役立ち得る。ある特定の実施形態において、核酸のフラグメントサイズ（例えば、長さ）と、核酸捕捉プロセスからの出力（例えば、ゲノム領域起源データ）との組み合わせが、入力デバイスを介して入力され得るデータとして役立ち得る。ある特定の実施形態において、インシリコプロセスによってシミュレーションデータが生成され、そのシミュレーションデータは、入力デバイスを介して入力され得るデータとして役立つ。用語「インシリコ」とは、コンピュータを用いて行われる研究および実験のことを指す。インシリコプロセスとしては、本明細書中に記載されるプロセスに従って、配列リードをマッピングし、マッピングされた配列リードを処理することが挙げられるが、これらに限定されない。

30

【0328】

システムは、本明細書中に記載されるプロセスを行うために有用なソフトウェアを備え得、ソフトウェアは、そのようなプロセスを行うための1つまたはそれを超えるモジュール（例えば、配列決定モジュール、論理処理モジュール、データ表示構築モジュール）を備え得る。用語「ソフトウェア」とは、コンピュータによって実行されたとき、コンピュータ操作を行うコンピュータ可読プログラムの指示のことを指す。1つまたはそれを超えるマイクロプロセッサによって実行可能な指示は、実行されたとき、1つまたはそれを超えるマイクロプロセッサに本明細書中に記載される方法を履行させ得る実行可能なコードとして提供されるときがある。本明細書中に記載されるモジュールは、ソフトウェアとして存在し得、そのソフトウェアに組み入れられた指示（例えば、プロセス、ルーチン、サブルーチン）は、マイクロプロセッサによって履行され得るかまたは行われ得る。例えば、あるモジュール（例えば、ソフトウェアモジュール）は、特定のプロセスまたはタスクを行うプログラムの一部であり得る。用語「モジュール」とは、より大きい機器またはソフトウェアシステムにおいて使用され得る自己完結型の機能単位のことを指す。モジュールは、そのモジュールの機能を行うための指示のセットを備え得る。モジュールは、デー

40

50

タおよび／または情報を変換し得る。データおよび／または情報は、好適な形態で存在し得る。例えば、データおよび／または情報は、デジタルまたはアナログであり得る。ある特定の実施形態において、データおよび／または情報は、パケット、バイト、文字またはビットであり得るときがある。いくつかの実施形態において、データおよび／または情報は、集められた、アセンブルされた、または使用可能な、任意のデータまたは情報であり得る。データおよび／または情報の非限定的な例としては、好適な媒体、写真、ビデオ、音（例えば、周波数、可聴または非可聴）、数字、定数、値、オブジェクト、時間、関数、指示、マップ、参照、配列、リード、マッピングされたリード、レベル、範囲、しきい値、シグナル、表示、提示またはそれらの変換が挙げられる。モジュールは、データおよび／または情報を受け入れ得るかまたは受け取り得、データおよび／または情報を第2の形態に変換し得、その第2の形態を機器、周辺装置、構成要素または別のモジュールに提供し得るかまたは移し得る。モジュールは、以下の非限定的な機能のうちの1つまたはそれを超える機能を行い得る：例えば、配列リードをマッピングすること、カウントを提供すること、部分をアセンブルすること、レベルを提供するかもしくは決定すること、カウントプロファイルを提供すること、正規化すること（例えば、リードを正規化すること、カウントを正規化することなど）、正規化されたカウントプロファイルもしくは正規化されたカウントのレベルを提供すること、2つもしくはそれを超えるレベルを比較すること、不確定値を提供すること、予想されるレベルおよび予想される範囲（例えば、予想されるレベル範囲、しきい値範囲およびしきい値レベル）を提供するかもしくは決定すること、レベルを調整すること（例えば、第1のレベルを調整すること、第2のレベルを調整すること、染色体もしくはそのセグメントのプロファイルを調整すること、および／またはパディングすること）、識別を提供すること（例えば、コピー数変異、遺伝的変異または異数性を識別すること）、カテゴリー化すること、プロットすること、および／またはアウトカムを提供すること。マイクロプロセッサは、ある特定の実施形態において、モジュールにおいて指示を行い得る。いくつかの実施形態において、1つまたはそれを超えるマイクロプロセッサは、モジュールまたはモジュールの群において指示を行うために必要とされる。モジュールは、データおよび／または情報を別のモジュール、機器または供給源に提供し得、データおよび／または情報を別のモジュール、機器または供給源から受け取り得る。

【0329】

コンピュータプログラム製品は、有形のコンピュータ可読媒体に組み入れられているときがあり、非一時的なコンピュータ可読媒体に有形的に組み入れられているときがある。モジュールは、コンピュータ可読媒体（例えば、ディスク、ドライブ）またはメモリー（例えば、ランダムアクセスメモリー）に格納されているときがある。モジュールからの指示を履行することができるモジュールおよびマイクロプロセッサは、1つの機器または異なる機器に存在し得る。モジュールに対する指示を履行することができるモジュールおよび／またはマイクロプロセッサは、ユーザーと同じ場所（例えば、ローカルネットワーク）またはユーザーと異なる場所（例えば、リモートネットワーク、クラウドシステム）に存在し得る。2つまたはそれを超えるモジュールとともに方法が行われる実施形態において、それらのモジュールは、同じ機器に存在し得、1つまたはそれを超えるモジュールは、同じ物理的場所における異なる機器に存在し得、1つまたはそれを超えるモジュールは、異なる物理的場所における異なる機器に存在し得る。

【0330】

機器は、いくつかの実施形態において、モジュールにおいて指示を行うための少なくとも1つのマイクロプロセッサを備える。参照ゲノムの部分にマッピングされた配列リードのカウントは、本明細書中に記載される方法を行うように構成された指示を実行するマイクロプロセッサによってアクセスされるときがある。マイクロプロセッサによってアクセスされるカウントは、システムのメモリー内に存在し得、それらのカウントは、アクセスされ得、それらのカウントを得た後に、そのシステムのメモリーに入れられ得る。いくつかの実施形態において、機器は、マイクロプロセッサ（例えば、1つまたはそれを超える

マイクロプロセッサ)を備え、そのマイクロプロセッサは、モジュールからの1つまたはそれを超える指示(例えば、プロセス、ルーチンおよび/またはサブルーチン)を行うおよび/または履行することができる。いくつかの実施形態において、機器は、複数のマイクロプロセッサ、例えば、協調されていて並行して動くマイクロプロセッサを備える。いくつかの実施形態において、機器は、1つまたはそれを超える外部マイクロプロセッサ(例えば、内部または外部ネットワーク、サーバー、記憶装置および/または記憶ネットワーク(例えば、クラウド))とともに作動する。いくつかの実施形態において、機器は、モジュールを備える。ある特定の実施形態において、機器は、1つまたはそれを超えるモジュールを備える。モジュールを備えている機器は、1つまたはそれを超えるデータおよび/または情報を他のモジュールに移動させ得ることが多く、1つまたはそれを超えるデータおよび/または情報を他のモジュールから受け取り得ることが多い。ある特定の実施形態において、機器は、周辺装置および/または構成要素を備える。ある特定の実施形態において、機器は、データおよび/または情報を他のモジュール、周辺装置および/または構成要素に移動させ得、ならびにデータおよび/または情報を他のモジュール、周辺装置および/または構成要素から移動させ得る、1つまたはそれを超える周辺装置または構成要素を備え得る。ある特定の実施形態において、機器は、データおよび/または情報を提供する周辺装置および/または構成要素とインタラクトする。ある特定の実施形態において、周辺装置および構成要素は、ある機能を行う際に機器を補助するか、またはモジュールと直接インタラクトする。周辺装置および/または構成要素の非限定的な例としては、好適なコンピュータ周辺装置、I/Oまたは記憶方法または記憶装置が挙げられ、それらとしては、スキャナー、プリンター、ディスプレイ(例えば、モニター、LED、LCDまたはCRT)、カメラ、マイクロホン、パッド(例えば、iPad、タブレット)、タッチスクリーン、スマートフォン、携帯電話、USB I/Oデバイス、USB大容量記憶装置、キーボード、コンピュータマウス、デジタルペン、モデム、ハードドライブ、ジャンプドライブ、フラッシュドライブ、マイクロプロセッサ、サーバー、CD、DVD、グラフィックカード、特殊化されたI/Oデバイス(例えば、シークエンサー、フォトセル、光電子増倍管、光学式読取装置、センサーなど)、1つまたはそれを超えるフローセル、流体を取り扱う構成要素、ネットワークインターフェースコントローラ、ROM、RAM、無線伝送方法および無線伝送デバイス(Bluetooth(登録商標)、WiFiなど)、ワールドワイドウェブ(www)、インターネット、コンピュータおよび/または別のモジュールが挙げられるが、これらに限定されない。

【0331】

ソフトウェアは、コンピュータ可読媒体に記録されたプログラム指示を備えるプログラム製品上に提供されていることが多く、そのコンピュータ可読媒体としては、フロッピー(登録商標)ディスク、ハードディスクおよび磁気テープを含む磁気媒体;ならびにCD-ROMディスク、DVDディスク、光磁気ディスク、フラッシュドライブ、RAM、フロッピー(登録商標)ディスクなどを含む光学媒体およびプログラム指示が記録され得る他のそのような媒体が挙げられるが、これらに限定されない。オンラインでの履行において、ある組織が維持しているサーバーおよびウェブサイトは、遠隔ユーザーにソフトウェアのダウンロードを提供するように構成され得るか、または遠隔ユーザーが、組織によって維持されている遠隔システムにアクセスして、ソフトウェアにリモートでアクセスし得る。ソフトウェアは、入力情報を入手し得るかまたは受け取り得る。ソフトウェアは、特にデータを入手するかまたは受け取るモジュール(例えば、配列リードデータおよび/またはマッピングされたリードデータを受け取るデータ受信モジュール)を備え得、そのデータを特に処理するモジュール(例えば、受け取ったデータを処理する(例えば、アウトカムおよび/またはレポートをフィルタリングする、正規化する、提供する)処理モジュール)を備え得る。入力情報を「入手する」および「受け取る」という用語は、ローカルサイトまたはリモートサイトからのコンピュータコミュニケーション手段、人間によるデータ入力またはデータを受信する他の任意の方法によって、データ(例えば、配列リード、マッピングされたリード)を受け取ることを指す。入力情報は、それを受け取った場所と

同じ場所において生成されてもよいし、異なる場所において生成され、受信場所に送信されてもよい。いくつかの実施形態において、入力情報は、処理される（例えば、処理に適した形式にされる（例えば、表にされる））前に修正される。

【0332】

いくつかの実施形態において、コンピュータプログラム製品、例えば、コンピュータ可読プログラムコードが組み入れられている、コンピュータが使用可能な媒体を備えるコンピュータプログラム製品が提供され、そのコンピュータ可読プログラムコードは、ある方法を履行するために実行されるように適合されており、その方法は、（a）ゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、そのセグメントに対するカウントAを提供する工程であって、それらの配列リードは、そのゲノムを有する被験体由来のテストサンプルからの核酸のリードである、工程；（b）そのゲノムまたはそのゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、そのゲノムまたはそのゲノムのサブセットに対するカウントBを提供する工程であって、ここで、そのカウントBは、参照ゲノムにアラインメントされない配列リードのカウントである、工程；および（c）カウントAとカウントBとの比率としての、そのセグメントに対するカウント提示を確定する工程を含む。

【0333】

ソフトウェアは、ある特定の実施形態において、1つまたはそれを超えるアルゴリズムを備え得る。アルゴリズムは、有限列の指示に従ってデータを処理するためおよび／またはアウトカムもしくはレポートを提供するために使用され得る。アルゴリズムは、タスクを完了するための明確な指示のリストであることが多い。初期状態から始まって、それらの指示は、規定の一連の連続状態を介して進み、最終的には最後の終結状態で終了する、計算を記載し得る。1つの状態から次への移行は、必ずしも決定論的でない（例えば、いくつかのアルゴリズムはランダム性を組み込んでいる）。例としてであって限定ではないが、アルゴリズムは、探索アルゴリズム、ソートアルゴリズム、マージアルゴリズム、数値アルゴリズム、グラフアルゴリズム、文字列アルゴリズム、モデリングアルゴリズム、計算幾何（computational genometric）アルゴリズム、組み合わせアルゴリズム、機械学習アルゴリズム、暗号法アルゴリズム、データ圧縮アルゴリズム、構文解析アルゴリズムなどであり得る。アルゴリズムには、1つのアルゴリズムまたは共同して動く2つもしくはそれを超えるアルゴリズムが含まれ得る。アルゴリズムは、任意の好適な複雑さのクラスおよび／またはパラメータ化された複雑さであり得る。アルゴリズムは、算出および／またはデータ処理のために使用され得、いくつかの実施形態では、決定論的アプローチまたは確率的／予測アプローチにおいて使用され得る。アルゴリズムは、コンピュータ環境において、好適なプログラミング言語を使用することによって履行され得、そのプログラミング言語の非限定的な例は、C、C++、Java（登録商標）、Perl、Python、Fortranなどである。いくつかの実施形態において、アルゴリズムは、誤差の範囲、統計解析、統計的有意性および／または他の情報もしくはデータセットとの比較を含むように構成され得るかまたは修正され得る（例えば、ニューラルネットまたはクラスタリングアルゴリズムを用いるときに適用可能である）。

【0334】

ある特定の実施形態において、いくつかのアルゴリズムは、ソフトウェアにおいて使用するために履行され得る。これらのアルゴリズムは、いくつかの実施形態において、生データを用いて訓練され得る。新しい各生データサンプルに対して、訓練されたアルゴリズムが、代表的な処理されたデータセットまたはアウトカムをもたらし得る。処理されたデータセットは、処理された親データセットと比較して複雑さが低下しているときがある。いくつかの実施形態において、処理されたセットに基づいて、訓練されたアルゴリズムの性能は、感度および特異性に基づいて評価され得る。ある特定の実施形態では、最も高い感度および／または特異性を有するアルゴリズムが特定され、使用され得る。

【0335】

ある特定の実施形態において、シミュレートされた（またはシミュレーション）データ

は、例えば、アルゴリズムを訓練するかまたはアルゴリズムをテストすることによって、データ処理を助け得る。いくつかの実施形態において、シミュレーションデータは、配列リードの異なるグルーピングの様々な仮説的サンプリングを含む。シミュレーションデータは、現実の集団から予想され得るものに基づき得るか、またはアルゴリズムをテストするためおよび／もしくは正しい分類に割り当てるために歪められ得る。シミュレーションデータは、本明細書中で「バーチャル」データとも称される。ある特定の実施形態において、シミュレーションは、コンピュータプログラムによって行われ得る。シミュレーションデータセットを使用する際の1つの実行可能な工程は、識別された結果の信頼度、例えば、どれくらい十分にランダムサンプリングがマッチするかまたは最良のものが元のデータに相当するかを評価することである。1つのアプローチは、確率値（p 値）を算出することであり、その確率値は、選択されたサンプルよりも良好なスコアを有するランダムなサンプルの確率を推定する。いくつかの実施形態において、少なくとも1つのサンプルが参照サンプルとマッチすると仮定する経験的モデルが評価され得る（変動の分解ありまたはなしで）。いくつかの実施形態において、例えばポアソン分布などの別の分布が、確率分布を定義するために使用され得る。

10

【0336】

ある特定の実施形態において、システムは、1つまたはそれを超えるマイクロプロセッサを備え得る。マイクロプロセッサは、コミュニケーションバスに接続され得る。コンピュータシステムは、メインメモリーを備え得、ランダムアクセスメモリー（RAM）を備えることが多く、二次メモリーも備え得る。いくつかの実施形態において、メモリーは、非一時的なコンピュータ可読記憶媒体を含む。二次メモリーには、例えば、ハードディスクドライブおよび／または着脱可能な記憶ドライブが含まれ得、それらは、フロッピー（登録商標）ディスクドライブ、磁気テープドライブ、光ディスクドライブ、メモリーカードなどに代表される。着脱可能な記憶ドライブは、着脱可能な記憶装置から読み出し、かつ／または着脱可能な記憶装置に書き込むことが多い。着脱可能な記憶装置の非限定的な例としては、フロッピー（登録商標）ディスク、磁気テープ、光ディスクなどが挙げられ、これらは、例えば、着脱可能な記憶ドライブによって読み出され得、着脱可能な記憶ドライブに書き込まれ得る。着脱可能な記憶装置は、コンピュータソフトウェアおよび／またはデータが格納されている、コンピュータが使用可能な記憶媒体を備え得る。

20

【0337】

マイクロプロセッサは、システムの中にソフトウェアを実装し得る。いくつかの実施形態において、マイクロプロセッサは、ユーザーが行い得る本明細書中に記載されるタスクを自動的に行うようにプログラムされ得る。したがって、マイクロプロセッサ、またはそのようなマイクロプロセッサによって行われるアルゴリズムは、ユーザーの監視または入力を全くかほとんど必要としないことがある（例えば、ソフトウェアは、機能を自動的に履行するようにプログラムされ得る）。いくつかの実施形態において、プロセスの複雑さは、1人の人物または人物の1つの群が、そのプロセスをコピー数変異の存在もしくは非存在の判定にとって十分短い時間枠で行うことができないほど大きい。

30

【0338】

いくつかの実施形態において、二次メモリーは、コンピュータプログラムまたは他の指示がコンピュータシステムに搭載されることを可能にするための他の同様の手段を備え得る。例えば、システムは、着脱可能な記憶装置およびインターフェースデバイスを備え得る。そのようなシステムの非限定的な例としては、プログラムカートリッジおよびカートリッジインターフェース（例えば、ビデオゲームデバイスに見られるもの）、着脱可能なメモリーチップ（例えば、EPROMまたはPROM）および関連ソケット、ならびに他の着脱可能な記憶装置、ならびにソフトウェアおよびデータが、着脱可能な記憶装置からコンピュータシステムに移されることを可能にするインターフェースが挙げられる。

40

【0339】

いくつかの実施形態において、1つの実体が、本明細書中に記載される方法、システム、機器、装置またはコンピュータプログラム製品において、配列リードのカウントを生成

50

し得、それらの配列リードを部分にマッピングし得、それらのマッピングされたリードをカウントし得、マッピングされてカウントされたリードを使用し得る。ある特定の実施形態において、部分にマッピングされた配列リードのカウントは、本明細書中に記載される方法、システム、機器、装置またはコンピュータプログラム製品において、第2の実体による使用のために、1つの実体によって、第2の実体に移されるときがある。

【0340】

いくつかの実施形態において、1つの実体が、配列リードを生成し、いくつかの実施形態において、第2の実体が、それらの配列リードを参照ゲノムにおける部分にマッピングする。第2の実体は、マッピングされたリードをカウントし、本明細書中に記載される方法、システム、機器またはコンピュータプログラム製品において、マッピングされてカウ
10
ントされたリードを使用するときがある。ある特定の実施形態において、第2の実体は、マッピングされたリードを第3の実体に移し、第3の実体は、本明細書中に記載される方法、システム、機器またはコンピュータプログラム製品において、マッピングされたリードをカウントし、マッピングされたリードを使用する。ある特定の実施形態において、第2の実体は、マッピングされたリードをカウントし、マッピングされてカウントされたリ
20
ードを第3の実体に移し、第3の実体は、本明細書中に記載される方法、システム、機器またはコンピュータプログラム製品において、マッピングされてカウントされたリードを使用する。第3の実体が関わる実施形態において、第3の実体は、第1の実体と同じであるときがある。つまり、第1の実体は、配列リードを第2の実体に移すときがあり、その第2の実体は、配列リードを参照ゲノムにおける部分にマッピングし得、および／または
20
マッピングされたリードをカウントし得、第2の実体は、マッピングされたおよび／またはカウントされたリードを第3の実体に移し得る。第3の実体は、本明細書中に記載される方法、システム、機器またはコンピュータプログラム製品において、マッピングされたおよび／またはカウントされたリードを使用し得るときがあり、ここで、その第3の実体は、第1の実体と同じであるときがあり、第3の実体は、第1または第2の実体と異なるときがある。

【0341】

いくつかの実施形態において、1つの実体は、妊婦から血液を入手し、必要に応じてその血液から（例えば、血漿または血清から）核酸を単離し、その血液または核酸を第2の実体に移し、第2の実体は、その核酸から配列リードを生成する。
30

【0342】

図5は、本明細書中に記載される様々なシステム、方法、アルゴリズムおよびデータ構造が履行され得るコンピュータ環境510の非限定的な例を例証している。コンピュータ環境510は、好適なコンピュータ環境の一例にすぎず、本明細書中に記載されるシステム、方法およびデータ構造の用途または機能の範囲に関していかなる限定も提案すると意図されていない。コンピュータ環境510は、コンピュータ環境510に例証されている構成要素のいずれか1つまたは組み合わせに関して、いかなる従属物または必要条件も有しないと解釈されるべきである。ある特定の実施形態において、図5に示されているシステム、方法およびデータ構造のサブセットが使用され得る。本明細書中に記載されるシステム、方法およびデータ構造は、数多くの他の多目的または特殊目的のコンピューティングシステム環境または設定で動作可能である。好適であり得る公知のコンピューティングシステム、環境および／または設定の例としては、パーソナルコンピュータ、サーバーコン
40
ピュータ、シンクライアント、シッククライアント、手持ち式またはラップトップ型デバイス、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能な家庭用電化製品、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、上記システムまたはデバイスのいずれかを備える分散コンピュータ環境などが挙げられるが、これらに限定されない。

【0343】

図5の動作環境510は、処理装置521、システムメモリー522、およびシステムメモリー522を含む様々なシステムの構成要素を処理装置521に動作可能につなぐシ
50

システムバス523を含む、コンピュータ520の形態の多目的のコンピュータデバイスを含む。ただ1つのまたは1つより多い処理装置521が存在し得、コンピュータ520のプロセッサは、単一の中央処理装置（CPU）、または並列処理環境と一般に称される複数の処理装置を備える。コンピュータ520は、従来のコンピュータ、分散コンピュータまたは他の任意のタイプのコンピュータであり得る。

【0344】

システムバス523は、種々のバスアーキテクチャーのいずれかを用いる、メモリーバスまたはメモリーコントローラ、周辺装置用バスおよびローカルバスを含む、いくつかのタイプのバス構造のいずれかであり得る。システムメモリーは、単純にメモリーとも称され得、それには、読出し専用メモリー（ROM）524およびランダムアクセスメモリー（RAM）が含まれる。スタートアップ中などにコンピュータ520内の要素間の情報の移動を助ける基本ルーチンを含む基本入出力システム（BIOS）526は、ROM524に格納されている。コンピュータ520は、ハードディスクから読み出すためおよびハードディスクに書き込むためのハードディスクドライブインターフェース527（図示せず）、着脱可能な磁気ディスク529から読み出すためまたは着脱可能な磁気ディスク529に書き込むための磁気ディスクドライブ528、ならびに着脱可能な光ディスク531（例えば、CD-ROMまたは他の光学媒体）から読み出すためまたは着脱可能な光ディスク531に書き込むための光ディスクドライブ530をさらに備え得る。

【0345】

ハードディスクドライブ527、磁気ディスクドライブ528および光ディスクドライブ530は、それぞれハードディスクドライブインターフェース532、磁気ディスクドライブインターフェース533および光ディスクドライブインターフェース534によって、システムバス523に接続される。上記ドライブおよびそれらの関連するコンピュータ可読媒体は、コンピュータ可読指示、データ構造、プログラムモジュールおよび他のデータの揮発性記憶装置をコンピュータ520に提供する。コンピュータによってアクセス可能なデータを格納し得る任意のタイプのコンピュータ可読媒体（例えば、磁気カセット、フラッシュメモリーカード、デジタルビデオディスク、ベルヌーイカートリッジ、ランダムアクセスメモリー（RAM）、読出し専用メモリー（ROM）など）が、その動作環境において使用され得る。

【0346】

いくつかのプログラムモジュールは、オペレーティングシステム535、1つまたはそれを超えるアプリケーションプログラム536、他のプログラムモジュール537およびプログラムデータ538を備える、ハードディスク、磁気ディスク529、光ディスク531、ROM524またはRAMに格納され得る。ユーザーは、コマンドおよび情報を、キーボード540およびポインティングデバイス542などの入力デバイスを介してパーソナルコンピュータ520に入力し得る。他の入力デバイス（図示せず）としては、マイクロホン、ジョイスティック、ゲームパッド、サテライトディッシュ、スキャナーなどが挙げられ得る。これらのおよび他の入力デバイスは、システムバスにつながったシリアルポートインターフェース546を介して処理装置521に接続されることが多いが、他のインターフェース（例えば、パラレルポート、ゲームポートまたはユニバーサルシリアルバス（USB））によって接続され得る。モニター547または他のタイプの表示デバイスも、ビデオアダプター548などのインターフェースを介してシステムバス523に接続される。コンピュータは、通常、モニターに加えて、スピーカーおよびプリンターなどの他の周辺装置出力デバイス（図示せず）を備える。

【0347】

コンピュータ520は、ネットワーク化された環境において、1つまたはそれを超えるリモートコンピュータ、例えば、リモートコンピュータ549への論理接続を用いて動作し得る。これらの論理接続は、コンピュータ520もしくはその一部につながったコミュニケーションデバイスによってまたは他の様式で達成され得る。リモートコンピュータ549は、別のコンピュータ、サーバー、ルーター、ネットワークPC、クライアント、ピ

10

20

30

40

50

アデバイスまたは他の共通ネットワークノードであり得、代表的には、コンピュータ 5 2 0 に対して上に記載された要素の多くまたはすべてを備えるが、メモリ記憶装置 5 5 0 だけが、図 5 に図示されている。図 5 に描かれている論理接続は、ローカルエリアネットワーク (LAN) 5 5 1 および広域ネットワーク (WAN) 5 5 2 を含む。そのようなネットワーク環境は、オフィスのネットワーク、企業規模のコンピュータネットワーク、イントラネットおよびインターネットにおいてありふれたものであり、これらのすべてが、ネットワークのタイプである。

【0348】

コンピュータ 5 2 0 は、LAN ネットワーク環境において使用されるとき、コミュニケーションデバイスの 1 タイプであるネットワークインターフェースまたはアダプター 5 5 3 を介してローカルネットワーク 5 5 1 に接続される。コンピュータ 5 2 0 は、WAN ネットワーク環境において使用されるとき、モデム 5 5 4、あるタイプのコミュニケーションデバイス、または広域ネットワーク 5 5 2 にわたるコミュニケーションを確立するための他の任意のタイプのコミュニケーションデバイスを備えることが多い。内部または外部のものであり得るモデム 5 5 4 は、シリアルポートインターフェース 5 4 6 を介してシステムバス 5 2 3 に接続される。ネットワーク化された環境において、パーソナルコンピュータ 5 2 0 またはその部分に関して描かれたプログラムモジュールは、リモートメモリ記憶装置に格納され得る。示されているネットワークの接続は非限定的な例であり、コンピュータ間のコミュニケーションリンクを確立するために他のコミュニケーションデバイスを使用してもよいことが認識される。

変換

【0349】

上で述べたように、データは、1 つの形態から別の形態に変換されるときがある。用語「変換される」、「変換」およびそれらの文法上の派生物または等価物は、本明細書中で使用されるとき、物理的な出発物質 (例えば、試験被験体および/または参照被験体のサンプル核酸) からその物理的な出発物質のデジタル提示 (例えば、配列リードデータ) へのデータの変更のことを指し、いくつかの実施形態において、アウトカム (例えば、テストサンプルに対する胎児分率の測定または推定) を提供するために使用され得る、そのデジタル提示の 1 つまたはそれを超える数値またはグラフ表示へのさらなる変換を含む。ある特定の実施形態において、デジタル表示されたデータの 1 つまたはそれを超える数値および/またはグラフ表示は、試験被験体の物理的なゲノムの様子を表すため (例えば、ゲノムの挿入、重複または欠失の存在もしくは非存在を仮想的に表すかまたは視覚的に表すため; 医学的症状に関連する配列の物理量の変動の存在もしくは非存在を表すため) に使用され得る。仮想提示は、出発物質のデジタル提示の 1 つまたはそれを超える数値またはグラフ表示にさらに変換されるときがある。これらの方法は、物理的な出発物質を、数値もしくはグラフ表示または試験被験体のゲノムの物理的な様子の提示に変換し得る。

【0350】

いくつかの実施形態において、データセットの変換は、データの複雑さおよび/またはデータの次元を低下させることによって、アウトカムの提供を容易にする。データセットの複雑さは、物理的な出発物質をその出発物質の仮想提示 (例えば、物理的な出発物質を代表する配列リード) に変換するプロセス中に減少するときがある。好適な特徴または変数が、データセットの複雑さおよび/または次元を低下させるために使用され得る。データ処理に対する標的の特徴として使用するために選択され得る特徴の非限定的な例としては、GC 含有量、胎児の性別予測、フラグメントサイズ (例えば、CCF フラグメント、リードの長さまたはその好適な提示 (例えば、FRS))、フラグメントの配列、染色体異数性の識別、特定の遺伝子またはタンパク質の識別、癌の識別、疾患、遺伝された遺伝子/形質、染色体異常、生物学的カテゴリー、化学的カテゴリー、生化学的カテゴリー、遺伝子またはタンパク質のカテゴリー、遺伝子オントロジー、タンパク質オントロジー、同時制御される遺伝子、細胞シグナル伝達遺伝子、細胞周期遺伝子、前述の遺伝子に関するタンパク質、遺伝子変異体、タンパク質変異体、同時制御される遺伝子、同時制御され

るタンパク質、アミノ酸配列、ヌクレオチド配列、タンパク質構造データなどおよび前述のものの組み合わせが挙げられる。データセットの複雑さおよび／または次元の減少の非限定的な例としては、複数の配列リードからプロファイルプロットへの整理編集（*reduction*）、複数の配列リードから数値（例えば、正規化された値、Z 得点、p 値）への整理編集；複数の解析方法から確率プロットまたは単一ポイントへの整理編集；導出された量の主成分分析などまたはそれらの組み合わせが挙げられる。

遺伝的変異および医学的症状

【0351】

遺伝分散の存在もしくは非存在は、本明細書中に記載される方法、機器または装置を用いて判定され得る。ある特定の実施形態において、1つまたはそれを超える遺伝的変異の存在もしくは非存在は、本明細書中に記載される方法、機器および装置によって提供されるアウトカムに従って判定される。遺伝的変異は、通常、ある特定の個体に存在する特定の遺伝的表現型であり、遺伝的変異は、個体の統計学的に有意な部分母集団に存在することが多い。いくつかの実施形態において、遺伝的変異は、染色体異常（例えば、異数性、1本またはそれを超える染色体の重複、1本またはそれを超える染色体の喪失）、部分的な染色体異常またはモザイク現象（例えば、染色体の1つまたはそれを超えるセグメントの喪失または獲得）、転座、逆位であり、これらの各々が、本明細書中に詳細に記載されている。遺伝的変異の非限定的な例としては、1つまたはそれを超える欠失（例えば、微小欠失）、重複（例えば、微小重複）、挿入、突然変異、多型（例えば、単一ヌクレオチド多型）、融合、反復（例えば、短いタンDEM反復）、異なるメチル化部位、異なるメチル化パターンなどおよびそれらの組み合わせが挙げられる。挿入、反復、欠失、重複、突然変異または多型は、任意の長さであり得、いくつかの実施形態において、約1塩基または塩基対（bp）から約250メガベース（Mb）長である。いくつかの実施形態において、挿入、反復、欠失、重複、突然変異または多型は、約1塩基または塩基対（bp）から約50,000キロベース（kb）長（例えば、約10bp、50bp、100bp、500bp、1kb、5kb、10kb、50kb、100kb、500kb、1000kb、5000kbまたは10,000kb長）である。

【0352】

遺伝的変異は、欠失であるときがある。ある特定の実施形態において、欠失は、染色体またはDNA配列の一部が欠損している突然変異（例えば、遺伝的異常）である。欠失は、遺伝物質の喪失であることが多い。任意の数のヌクレオチドが欠失し得る。欠失は、1本またはそれを超える染色体全体、染色体のセグメント、対立遺伝子、遺伝子、イントロン、エキソン、任意の非コード領域、任意のコード領域、それらのセグメントまたはそれらの組み合わせの欠失を含み得る。欠失は、微小欠失を含み得る。欠失は、単一塩基の欠失を含み得る。

【0353】

遺伝的変異は、遺伝的重複であるときがある。ある特定の実施形態において、重複は、染色体またはDNA配列の一部が複製され、ゲノムの元の場所に挿入された突然変異（例えば、遺伝的異常）である。ある特定の実施形態において、遺伝的重複（例えば、重複）は、DNAの領域の任意の重複である。いくつかの実施形態において、重複は、ゲノム内または染色体内に、しばしばタンDEMで、反復された核酸配列である。いくつかの実施形態において、重複は、1本またはそれを超える染色体全体、染色体のセグメント、対立遺伝子、遺伝子、イントロン、エキソン、任意の非コード領域、任意のコード領域、それらのセグメントまたはそれらの組み合わせの複製を含み得る。重複は、微小重複を含み得る。重複は、重複した核酸の1コピーまたはそれを超えるコピーを含むときがある。重複は、1回またはそれを超える回数反復された（例えば、1、2、3、4、5、6、7、8、9または10回反復された）遺伝的領域として特徴づけられるときがある。場合によっては、重複は、小さい領域（数千の塩基対）から全染色体に及び得る。重複は、相同組換えのエラーの結果として、またはレトロトランスポゾン事象に起因して、頻繁に生じる。重複は、ある特定のタイプの増殖性疾患に関連している。重複は、ゲノムマイクロアレイま

10

20

30

40

50

たは比較ゲノムハイブリダイゼーション (comparative genetic hybridization) (CGH) を用いて特徴づけられ得る。

【0354】

遺伝的変異は、挿入であるときがある。挿入は、核酸配列への1つまたはそれを超えるヌクレオチド塩基対の付加であるときがある。挿入は、微小挿入であるときがある。ある特定の実施形態において、挿入は、ゲノム、染色体またはそれらのセグメントへの染色体のセグメントの付加を含む。ある特定の実施形態において、挿入は、ゲノムまたはそのセグメントへの、対立遺伝子、遺伝子、イントロン、エキソン、任意の非コード領域、任意のコード領域、それらのセグメントまたはそれらの組み合わせの付加を含む。ある特定の実施形態において、挿入は、ゲノム、染色体またはそれらのセグメントへの、起源が不明の核酸の付加（例えば、挿入）を含む。ある特定の実施形態において、挿入は、単一塩基の付加（例えば、挿入）を含む。

10

【0355】

本明細書中で使用されるとき、「コピー数変異」は、通常、遺伝的変異または染色体異常の1つのクラスまたはタイプである。コピー数変異は、欠失（例えば、微小欠失）、重複（例えば、微小重複）または挿入（例えば、微小挿入）であり得る。しばしば、接頭辞「微小」は、本明細書中で使用されるとき、5 Mb 長未満の核酸のセグメントであるときがある。コピー数変異は、染色体のセグメントの1つまたはそれを超える欠失（例えば、微小欠失）、重複および／または挿入（例えば、微小重複、微小挿入）を含み得る。ある特定の実施形態において、重複は、挿入を含む。ある特定の実施形態において、挿入は、重複である。ある特定の実施形態において、挿入は、重複ではない。

20

【0356】

いくつかの実施形態において、コピー数変異は、胎児のコピー数変異である。胎児のコピー数変異は、胎児のゲノムにおけるコピー数変異であることが多い。いくつかの実施形態において、コピー数変異は、母体および／または胎児のコピー数変異である。ある特定の実施形態において、母体および／または胎児のコピー数変異は、妊婦（例えば、胎児を有する女性被験体）、出産した女性被験体または胎児を有することができる女性のゲノム内のコピー数変異である。コピー数変異は、その変異（例えば、重複または欠失）が、ゲノムの一方の対立遺伝子に存在するヘテロ接合性のコピー数変異であり得る。コピー数変異は、その変異がゲノムの両方の対立遺伝子に存在するホモ接合性のコピー数変異であり得る。いくつかの実施形態において、コピー数変異は、ヘテロ接合性またはホモ接合性の胎児のコピー数変異である。いくつかの実施形態において、コピー数変異は、ヘテロ接合性またはホモ接合性の母体および／または胎児のコピー数変異である。コピー数変異は、母体のゲノムおよび胎児のゲノムに存在するときがあるか、母体のゲノムに存在して胎児のゲノムに存在しないときがあるか、または胎児のゲノムに存在して母体のゲノムに存在しないときがある。

30

【0357】

「倍数性」は、胎児または母体に存在する染色体の数に対する言及である。ある特定の実施形態において、「倍数性」は、「染色体倍数性」と同じである。例えば、ヒトにおいて、常染色体は、2本一組で存在することが多い。例えば、遺伝的変異の非存在下では、ほとんどのヒトが、各常染色体（例えば、1～22番染色体）を2本有する。ヒトにおける、2本の常染色体の正常な相補物の存在は、正倍数体または二倍体と称されることが多い。「微小倍数性」は、倍数性と似た意味である。「微小倍数性」は、染色体のセグメントの倍数性のことを指すことが多い。用語「微小倍数性」は、染色体内のコピー数変異（例えば、欠失、重複および／または挿入）の存在もしくは非存在（例えば、ホモ接合性またはヘテロ接合性の欠失、重複もしくは挿入など、またはそれらが存在しないこと）に対する言及であるときがある。

40

【0358】

ある特定の実施形態において、胎児の微小倍数性は、その胎児の母（例えば、妊婦被験体）の微小倍数性と合致する。ある特定の実施形態において、胎児の微小倍数性は、その

50

胎児の母の微小倍数性と合致し、その母と胎児の両方が、同じヘテロ接合性のコピー数変異、ホモ接合性のコピー数変異を有するか、またはその母と胎児の両方が、正倍数性である。ある特定の実施形態において、胎児の微小倍数性は、その胎児の母の微小倍数性と異なる。例えば、胎児の微小倍数性は、コピー数変異についてヘテロ接合性であり、母は、コピー数変異についてホモ接合性であり、胎児の微小倍数性は、特定のコピー数変異について母の微小倍数性と合致しない（例えば、等しくない）ときがある。

【0359】

ある特定の実施形態において、被験体に対して存在もしくは非存在が識別される遺伝的変異は、医学的症状に関連する。したがって、本明細書中に記載される技術は、医学的症状または医学的状态に関連する1つまたはそれを超える遺伝的変異の存在もしくは非存在を識別するために使用され得る。医学的症状の非限定的な例としては、知的障害（例えば、ダウン症候群）、異常な細胞増殖（例えば、癌）、微生物核酸（例えば、ウイルス、細菌、真菌、酵母）の存在、および子癇前症に関連するものが挙げられる。

【0360】

遺伝的変異、医学的症状および医学的状态の非限定的な例は、本明細書の以後に記載される。

胎児の性別

【0361】

いくつかの実施形態において、胎児の性別、または性別に関連する障害（例えば、性染色体異数性）の予測は、本明細書中に記載される方法、機器および／または装置によって判定され得る。性別の判定は、一般に性染色体に基づく。ヒトでは、2本の性染色体、XおよびY染色体が存在する。Y染色体は、遺伝子SR Yを含み、これは、男性としての胚発生を引き起こす。ヒトおよび他の哺乳動物のY染色体は、正常な精子産生に必要な他の遺伝子も含む。XXを有する個体は、女性であり、XYは、男性であり、性染色体異数性と称されることが多い非限定的な変異としては、X0、XYY、XXXおよびXXYが挙げられる。ある特定の実施形態では、男性が、2本のX染色体および1本のY染色体（XXY；クラインフェルター症候群）を有するか、または1本のX染色体および2本のY染色体（XYY症候群；ヤコブ症候群）を有し、一部の女性が、2本の代わりに3本のX染色体（XXX；トリプルX症候群）または1本のX染色体（X0；ターナー症候群）を有する。ある特定の実施形態において、個体における一部の細胞だけが、モザイク現象（例えば、Turnerモザイク現象）と称され得る性染色体異数性の影響を受ける。他の症例としては、SR Yが損傷している（XY女性がもたらされる）症例、またはSR YがXに複製される症例（XX男性がもたらされる）が挙げられる。

【0362】

ある特定の症例では、子宮内において胎児の性別を判定することが有益である場合がある。例えば、1つまたはそれを超える伴性障害の家族歴を有する患者（例えば、妊婦）が、そのような障害を受け継ぐ胎児のリスクの評価を助けるために、宿している胎児の性別の判定を望む場合がある。伴性障害としては、X連鎖障害およびY連鎖障害が挙げられるが、これらに限定されない。X連鎖障害には、X連鎖劣性障害およびX連鎖優性障害が含まれる。X連鎖劣性障害の例としては、免疫障害（例えば、慢性肉芽腫症（CYBB）、ウィスコット・オールドリッチ症候群、X連鎖重症複合免疫不全、X連鎖無ガンマグロブリン血症、高IgM症候群タイプ1、IPEX、X連鎖リンパ増殖性疾患、プロパージン欠損症）、血液障害（例えば、血友病A、血友病B、X連鎖鉄芽球性貧血）、内分泌障害（例えば、アンドロゲン不感性症候群／ケネディ病、KAL1カルマン症候群、X連鎖先天性副腎低形成）、代謝障害（例えば、オルニチントランスカルバミラーゼ欠損症、眼脳腎症候群、副腎脳白質ジストロフィー、グルコース-6-リン酸デヒドロゲナーゼ欠損症、ピルビン酸デヒドロゲナーゼ欠損症、ダノン病／糖原病タイプIIb、ファブリー病、ハンター症候群、レッシュ・ナイハン症候群、メンケス病／後角症候群）、神経系障害（例えば、コフィン・ローリー症候群、MASA症候群、X連鎖アルファサラセミア精神遅滞症候群、シデリウス型X連鎖精神遅滞症候群、色盲、眼白子症、ノリエ病、コロイデレ

10

20

30

40

50

ミア、シャルコー・マリー・トゥース病（CMTX2-3）、ペリツェウス・メルツバッハー病、SMA X2）、皮膚および関連組織の障害（例えば、先天性角化異常症、無汗性外胚葉性形成異常（EDA）、X連鎖魚鱗癬、X連鎖内皮角膜ジストロフィー）、神経筋障害（例えば、ベッカー型筋ジストロフィー／デュシェンヌ、中心核ミオパシー（MTM1）、コンラーディ・ヒューネルマン症候群、エメリ・ドレフュス型筋ジストロフィー1）、泌尿器系障害（例えば、アルポート症候群、デント病、X連鎖腎性尿崩症）、骨／歯の障害（例えば、AMELXエナメル質形成不全症）、および他の障害（例えば、Bartth症候群、McLeod症候群、Smith-Fineman-Myers症候群、Simpson-Golabi-Behmel症候群、Mohr-Tranebjærg症候群、鼻指聴覚（Nasodigitacoustic）症候群）が挙げられるが、これらに限定されない。X連鎖優性障害の例としては、X連鎖性低リン酸血症、巣状皮膚低形成、脆弱X症候群、アイカルディ症候群、色素失調症、レット症候群、チャイルド症候群、Lujan-Fryns症候群および口顔面指症候群1が挙げられるが、これらに限定されない。Y連鎖障害の例としては、男性不妊症、網膜色素変性症および無精子症が挙げられるが、これらに限定されない。

染色体異常

【0363】

いくつかの実施形態において、胎児の染色体異常の存在もしくは非存在は、本明細書中に記載される方法、機器および／または装置を用いることによって判定され得る。染色体異常としては、染色体全体または1つもしくはそれを超える遺伝子を含む染色体の領域の獲得または喪失が挙げられるが、これらに限定されない。染色体異常には、モノソミー、トリソミー、ポリソミー、ヘテロ接合性の消失、不平衡転座によって引き起こされる欠失および重複を含む、1つもしくはそれを超えるヌクレオチド配列（例えば、1つまたはそれを超える遺伝子）の転座、欠失および／または重複が含まれる。用語「染色体異常」または「異数性」は、本明細書中で使用されるとき、主題の染色体の構造と正常な相同染色体の構造との間のずれのことを指す。用語「正常」とは、特定の種の健常個体に見られる優勢である核型または横縞像、例えば、正倍数性のゲノム（例えば、ヒトにおける二倍体、例えば、46, XXまたは46, XY）のことを指す。種々の生物が、広く異なる染色体組を有するので、用語「異数性」とは、特定の数の染色体のことを指さず、生物の所与の細胞内の染色体含有量が異常である状況のことを指す。いくつかの実施形態において、本明細書中の用語「異数性」は、染色体全体または染色体の一部の喪失または獲得によって引き起こされる遺伝物質の不均衡のことを指す。「異数性」とは、染色体のセグメントの1つまたはそれを超える欠失および／または挿入のことを指し得る。いくつかの実施形態において、用語「正倍数性」とは、正常に染色体が全数そろっていることを指す。

【0364】

用語「モノソミー」は、本明細書中で使用されるとき、正常な全数から1本の染色体が不足していることを指す。部分的モノソミーは、染色体のセグメントだけが1コピーで存在する、不平衡転座または欠失において生じ得る。性染色体のモノソミー（45, X）は、例えば、ターナー症候群を引き起こす。用語「ダイソミー」とは、ある染色体が2コピー存在することを指す。各染色体を2コピー有するヒトなどの生物（二倍体または「正倍数体」である生物）の場合、ダイソミーは、正常な状態である。各染色体を正常に3コピーもしくはそれを超えるコピーだけ有する生物（三倍体またはそれを超えるもの）の場合、ダイソミーは、異数体の染色体状態である。片親性ダイソミーでは、ある染色体の両方のコピーが、同じ親に由来する（他方の親からの寄与がない）。

【0365】

用語「トリソミー」は、本明細書中で使用されるとき、特定の染色体が2コピーの代わりに3コピー存在することを指す。ヒトのダウン症候群に見られる、21番染色体が余分に存在することは、「トリソミー21」と称される。トリソミー18およびトリソミー13は、他の2つのヒト常染色体トリソミーである。性染色体のトリソミーは、女性（例えば、トリプルX症候群における47, XXX）または男性（例えば、クラインフェルター

症候群における 47, XX Y ; またはヤコブ症候群における 47, X Y Y) に見られ得る。いくつかの実施形態において、トリソミーは、常染色体の大部分またはすべての重複である。ある特定の実施形態において、トリソミーは、特定のタイプの染色体が 3 本である場合 (例えば、3 コピー) をもたらす (例えば、正倍数性に対する特定のタイプの染色体が 2 本である場合 (例えば、1 対) の代わりに) 染色体全体の異数性である。

【0366】

用語「テトラソミー」および「ペンタソミー」は、本明細書中で使用されるとき、それぞれ、ある染色体が 4 または 5 コピー存在することを指す。常染色体ではめったに見られないが、性染色体テトラソミーおよびペンタソミーは、ヒトにおいて報告されており、それには、XXXX、XXXX Y、XXYY、XYYY、XXXXX、XXXX Y、XXXYY、XXYYY および XYYYY が含まれる。

10

【0367】

染色体異常は、種々の機構によって引き起こされ得る。機構としては、(i) 有糸分裂チェックポイントが弱くなった結果として生じる不分離、(ii) 複数の染色体において不分離を引き起こす不活性な有糸分裂チェックポイント、(iii) 1 つの動原体が両紡錘体極に付着するときに生じるメロテリック結合 (merotelic attachment)、(iv) 2 つより多い紡錘極が形成するときに形成する多極紡錘体、(v) 1 つの紡錘極しか形成されないときに形成する単極紡錘体、および (vi) 単極紡錘体の機構の最終結果として生じる四倍体中間体 (tetraploid intermediate) が挙げられるが、これらに限定されない。

20

【0368】

用語「部分モノソミー」および「部分トリソミー」は、本明細書中で使用されるとき、染色体の一部の喪失または獲得によって引き起こされる遺伝物質の不均衡のことを指す。部分モノソミーまたは部分トリソミーは、不平衡転座に起因し得、ここで、個体は、異なる 2 本の染色体の切断および融合によって形成される誘導染色体を有する。この状況では、その個体は、1 本の染色体の一部を 3 コピー (正常な 2 コピー、および誘導染色体上に存在するセグメント)、および誘導染色体に関わる他方の染色体の一部を 1 コピーだけ有し得る。

【0369】

用語「モザイク現象」は、本明細書中で使用されるとき、生物のいくつかであってすべてではない細胞における異数性のことを指す。ある特定の染色体異常は、モザイク染色体異常および非モザイク染色体異常として存在し得る。例えば、ある特定のトリソミー 21 個体は、モザイクダウン症候群を有し、一部は、非モザイクダウン症候群を有する。種々の機構が、モザイク現象をもたらす得る。例えば、(i) 最初の接合体は、21 番染色体を 3 本有し得、それによって通常、単純なトリソミー 21 がもたらされ得るが、細胞分裂の経過中に、1 つまたはそれを超える細胞系統が、それらの 21 番染色体のうちの 1 本を失った; および (ii) 最初の接合体は、21 番染色体を 2 本有し得るが、細胞分裂の経過中に、それらの 21 番染色体のうちの 1 本が複製された。体細胞モザイク現象は、おそらく、完全な異数性またはモザイク異数性に関わる遺伝的症候群に通常関連する機構と異なる機構によって生じる。体細胞モザイク現象は、例えば、ある特定のタイプの癌およびニューロンにおいて特定されている。ある特定の場において、トリソミー 12 は、慢性リンパ性白血病 (CLL) において特定されており、トリソミー 8 は、急性骨髄性白血病 (AML) において特定されている。また、個体が染色体の切断 (染色体不安定症候群) の素因となる遺伝的症候群は、頻繁に、様々なタイプの癌に対する高リスクと関連し、ゆえに、発癌における体細胞異数性の役割が強調される。本明細書中に記載される方法およびプロトコルは、非モザイク染色体異常およびモザイク染色体異常の存在もしくは非存在を識別し得る。

30

40

【0370】

表 1 A および 1 B は、本明細書中に記載される方法、機器および/または装置によって潜在的に識別され得る、染色体の症状、症候群および/または異常の非限定的なリストを

50

示している。表 1 B は、2011 年 10 月 6 日現在の DECIPHER データベースからのものである（例えば、バージョン 5.1、GRCh37 にマッピングされた位置に基づいて；ユニフォームリソースロケータ（URL）decipher.sanger.ac.ukにおいて入手可能）。

【表 1 A - 1】

表 1A

染色体	異常	疾患関連性
X	XO	ターナー症候群
Y	XXY	クラインフェルター症候群
Y	XYY	二重Y症候群
Y	XXX	トリソミーX症候群
Y	XXXX	XXXX症候群
Y	Xp21欠失	デュシェンヌ型/ベッカー型症候群、先天性副腎低形成、慢性肉芽腫性疾患
Y	Xp22欠失	ステロイドスルファターゼ欠損症
Y	Xq26欠失	X連鎖リンパ球増殖性疾患
1	1p (体細胞) モノソミー トリソミー	神経芽細胞腫
2	モノソミー トリソミー 2q	成長遅滞、発育および精神的遅延、ならびに軽微な身体的異常
3	モノソミー トリソミー (体細胞)	非ホジキンリンパ腫
4	モノソミー トリソミー (体細胞)	急性非リンパ性白血病 (ANLL)
5	5p	猫なきルジューヌ症候群
5	5q (体細胞) モノソミー トリソミー	骨髄異形成症候群
6	モノソミー トリソミー (体細胞)	明細胞肉腫
7	7q11.23 欠失	ウィリアムズ症候群
7	モノソミー トリソミー	小児モノソミー7症候群;体細胞性:腎皮質腺腫;骨髄異形成症候群
8	8q24.1 欠失	Langer-Giedon 症候群
8	モノソミー トリソミー	骨髄異形成症候群; Warkany 症候群; 体細胞性:慢性骨髄性白血病
9	モノソミー 9p	Alfi 症候群

10

20

30

40

【表 1 A - 2】

染色体	異常	疾患関連性
9	モノソミー 9p 部分トリソミー	Rethore 症候群
9	トリソミー	完全トリソミー-9症候群; モザイクトリソミー-9症候群
10	モノソミー トリソミー (体細胞)	ALLまたはANLL
11	11p-	無虹彩症; ウィルムス腫瘍
11	11q-	ヤコブセン症候群
11	モノソミー (体細胞) トリソミー	骨髄細胞系列が影響を受ける(ANLL、MDS)
12	モノソミー トリソミー (体細胞)	CLL、若年性顆粒膜細胞腫 (JGCT)
13	13q-	13q-症候群; Orbeli 症候群
13	13q14 欠失	網膜芽細胞腫
13	モノソミー トリソミー	パトー症候群
14	モノソミー トリソミー (体細胞)	骨髄性障害 (MDS、ANLL、非定型CML)
15	15q11-q13 欠失 モノソミー	ブラダー・ウィリー、アンジェルマン症候群
15	トリソミー (体細胞)	骨髄細胞系列およびリンパ系細胞系列が影響を受ける、例えば、MDS、ANLL、ALL、CLL)
16	トリソミー	完全トリソミー-16 モザイクトリソミー-16
16	16q13.3 欠失	Rubenstein-Taybi
3	モノソミー トリソミー (体細胞)	乳頭状腎細胞癌腫 (悪性)
17	17p-(体細胞)	骨髄性悪性腫瘍における17p症候群
17	17q11.2 欠失	Smith-Magenis
17	17q13.3	Miller-Dieker
17	モノソミー トリソミー (体細胞)	腎皮質腺腫
17	17p11.2-12 トリソミ ー	シャルコー・マリー・トゥース症候群タイプ1; HNPP

10

20

30

40

【表 1 A - 3】

染色体	異常	疾患関連性
18	18p-	18p 部分モノソミー症候群またはGrouchy Lamy Thieffry症候群
18	18q-	Grouchy Lamy Salmon Landry症候群
18	モノソミー トリソミー	エドワーズ症候群
19	モノソミー トリソミー	
20	20p-	トリソミー 20p 症候群
20	20p11.2-12 欠失	アラジール
20	20q-	体細胞性: MDS、ANLL、真性多血症、慢性好中球白血病
20	モノソミー トリソミー (体細胞)	乳頭状腎細胞癌腫 (悪性)
21	モノソミー トリソミー	ダウン症候群
22	22q11.2 欠失	ディジョージ症候群、口蓋心臓顔面症候群、円錐動脈幹奇形顔貌症候群、常染色体優性Opitz G/BBB症候群、Caylor心臓・顔症候群
22	モノソミー トリソミー	完全トリソミー22症候群

【 0 3 7 1 】

10

20

30

【表 1 B - 1】

表1B

症候群	染色体	開始	終了	間隔 (Mb)	グレード
12q14 微小欠失症候群	12	65,071,919	68,645,525	3.57	
15q13.3 微小欠失症候群	15	30,769,995	32,701,482	1.93	
15q24 反復微小欠失症候群	15	74,377,174	76,162,277	1.79	
15q26 過成長症候群	15	99,357,970	102,521,392	3.16	
16p11.2 微小重複症候群	16	29,501,198	30,202,572	0.70	
16p11.2-p12.2 微小欠失症候群	16	21,613,956	29,042,192	7.43	
16p13.11 反復微小欠失 (神経認知障害感受性遺伝子座)	16	15,504,454	16,284,248	0.78	
16p13.11 反復微小重複 (神経認知障害感受性遺伝子座)	16	15,504,454	16,284,248	0.78	
17q21.3 反復微小欠失症候群	17	43,632,466	44,210,205	0.58	1
1p36 微小欠失症候群	1	10,001	5,408,761	5.40	1
1q21.1 反復微小欠失 (神経発達障害に対する感受性遺伝子座)	1	146,512,930	147,737,500	1.22	3
1q21.1 反復微小重複 (神経発達障害に対する可能性がある感受性遺伝子座)	1	146,512,930	147,737,500	1.22	3
1q21.1 血小板減少機骨欠損 (TAR) 症候群に対する感受性遺伝子座	1	145,401,253	145,928,123	0.53	3
22q11 欠失症候群(口蓋心臓顔面/ デイジョージ症候群)	22	18,546,349	22,336,469	3.79	1
22q11 重複症候群	22	18,546,349	22,336,469	3.79	3
22q11.2 末端欠失症候群	22	22,115,848	23,696,229	1.58	
22q13 欠失症候群(フェラン・マクダーミッド症候群)	22	51,045,516	51,187,844	0.14	1
2p15-16.1 微小欠失症候群	2	57,741,796	61,738,334	4.00	
2q33.1 欠失症候群	2	196,925,089	205,206,940	8.28	1
2q37 モソミー	2	239,954,693	243,102,476	3.15	1
3q29 微小欠失症候群	3	195,672,229	197,497,869	1.83	
3q29 微小重複症候群	3	195,672,229	197,497,869	1.83	

10

20

30

40

【表 1 B - 2】

症候群	染色体	開始	終了	間隔 (Mb)	グレード
7q11.23 重複症候群	7	72,332,743	74,616,901	2.28	
8p23.1 欠失症候群	8	8,119,295	11,765,719	3.65	
9q サブテロメア欠失症候群	9	140,403,363	141,153,431	0.75	1
成人発症型常染色体優性白質ジストロフィー (ADLD)	5	126,063,045	126,204,952	0.14	
アンジェルマン症候群 (タイプ1)	15	22,876,632	28,557,186	5.68	1
アンジェルマン症候群 (タイプ2)	15	23,758,390	28,557,186	4.80	1
ATR-16 症候群	16	60,001	834,372	0.77	1
AZFa	Y	14,352,761	15,154,862	0.80	
AZFb	Y	20,118,045	26,065,197	5.95	
AZFb+AZFc	Y	19,964,826	27,793,830	7.83	
AZFc	Y	24,977,425	28,033,929	3.06	
猫の目症候群 (タイプI)	22	1	16,971,860	16.97	
シャルコー・マリー・トゥース症候群タイプ1A (CMT1A)	17	13,968,607	15,434,038	1.47	1
猫なき症候群 (5p 欠失)	5	10,001	11,723,854	11.71	1
脳アミロイドアンギオパチーを伴う早期発症型アルツハイマー病	21	27,037,956	27,548,479	0.51	
家族性腺腫性ポリポーシス	5	112,101,596	112,221,377	0.12	
圧迫性麻痺傾向を伴う遺伝性運動ニューロパシー (Hereditary Liability to Pressure Palsies)(HNPP)	17	13,968,607	15,434,038	1.47	1
Leri-Weill軟骨骨異形成症 (dyschondroostosis)(LWD) - SHOX 欠失	X	751,878	867,875	0.12	
Leri-Weill 軟骨骨異形成症 (LWD) - SHOX 欠失	X	460,558	753,877	0.29	
ミラー・ディーカー症候群 (MDS)	17	1	2,545,429	2.55	1
NF1-微小欠失症候群	17	29,162,822	30,218,667	1.06	1
ペリツェウス・メルツバッハー病	X	102,642,051	103,131,767	0.49	
Potocki-Lupski 症候群 (17p11.2 重複症候群)	17	16,706,021	20,482,061	3.78	
Potocki-Shaffer症候群	11	43,985,277	46,064,560	2.08	1

10

20

30

40

【表 1 B - 3】

症候群	染色体	開始	終了	間隔 (Mb)	グレード
ブラダー・ウィリー症候群 (タイプ1)	15	22,876,632	28,557,186	5.68	1
ブラダー・ウィリー症候群 (タイプ2)	15	23,758,390	28,557,186	4.80	1
RCAD (腎嚢胞・糖尿病)	17	34,907,366	36,076,803	1.17	
ルビンシュタイン・ティビ症候群	16	3,781,464	3,861,246	0.08	1
Smith-Magenis症候群	17	16,706,021	20,482,061	3.78	1
ソトス症候群	5	175,130,402	177,456,545	2.33	1
裂手/足奇形 1 (SHFM1)	7	95,533,860	96,779,486	1.25	
ステロイドスルファターゼ欠損症 (STS)	X	6,441,957	8,167,697	1.73	
WAGR 11p13 欠失症候群	11	31,803,509	32,510,988	0.71	
ウィリアムズ・ビューレン症候群(WBS)	7	72,332,743	74,616,901	2.28	1
ウォルフ・ヒルシュホーン症候群	4	10,001	2,073,670	2.06	1
Xq28 (MECP2) 重複	X	152,749,900	153,390,999	0.64	

【0372】

グレード1の症状は、1つより多い以下の特色を有することが多い；病原性の奇形；遺伝学者の間での強い一致；高い侵入性；不定の表現型であるがいくつかの共通する特徴を有し得ること；文献におけるすべての症例が臨床像を有すること；その異常を有する健常個体の症例がないこと；DVGデータベースに報告されていないかまたは健常集団に見られないこと；単一遺伝子または複数遺伝子の量的効果を確かめる機能的データ；確認されたまたは強い候補遺伝子；定義された臨床管理の意味；サーベイランスに対して意味を有する公知の癌リスク；複数の情報源（OMIM、Gene reviews、Orphanet、Unique、Wikipedia）；および／または診断的使用（生殖カウンセリング）に利用可能であること。

【0373】

グレード2の症状は、1つより多い以下の特色を有することが多い；おそらく病原性の奇形；高い侵入性；DD以外の一貫した特徴を有しない不定の表現型；文献における少数の症例／報告；報告されたすべての症例が臨床像を有すること；機能的データまたは確かめられた病原性遺伝子がないこと；複数の情報源（OMIM、Gene reviews、Orphanet、Unique、Wikipedia）；ならびに／または診断上の目的および生殖カウンセリングのために使用され得ること。

【0374】

グレード3の症状は、1つより多い以下の特色を有することが多い；感受性遺伝子座；発端者の健常個体または無影響の親が記載されること；コントロール集団に存在すること；非侵入性；軽度かつ非特異的な表現型；それほど一貫しない特徴；機能的データまたは確かめられた病原性遺伝子がないこと；より限定的なデータの供給源；大多数から逸脱する症例の場合または新規の臨床所見が存在する場合、第2の診断の可能性が残っていること；ならびに／または診断目的のために使用されるとき警告および生殖カウンセリングのための慎重なアドバイス。

医学的障害および医学的症状

【0375】

10

20

30

40

50

本明細書中に記載される方法は、任意の好適な医学的障害または医学的症状に適用可能であり得る。医学的障害および医学的症状の非限定的な例としては、細胞増殖性障害および症状、消耗性障害および症状、変性障害および症状、自己免疫障害および症状、子癇前症、化学毒性または環境毒性、肝臓の損傷または疾患、腎臓の損傷または疾患、血管性疾患、高血圧ならびに心筋梗塞が挙げられる。

【0376】

いくつかの実施形態において、細胞増殖性障害または症状は、肝臓、肺、脾臓、膵臓、結腸、皮膚、膀胱、眼、脳、食道、頭部、頸部、卵巣、精巣、前立腺などの癌またはそれらの組み合わせである。癌の非限定的な例としては、造血性起源の過形成性／新生物の細胞が関わる疾患である（例えば、骨髄、リンパ系もしくは赤血球の系統またはそれらの前駆細胞から生じる）造血性新生物障害が挙げられ、未分化型急性白血病（例えば、赤芽球性白血病および急性巨核芽球性白血病）から生じ得る。ある特定の骨髄性障害としては、急性前骨髄性白血病（APML）、急性骨髄性白血病（AML）および慢性骨髄性白血病（CML）が挙げられるが、これらに限定されない。ある特定のリンパ系悪性腫瘍としては、急性リンパ芽球性白血病（ALL）（B系統ALLおよびT系統ALLを含む）、慢性リンパ性白血病（CLL）、前リンパ球性白血病（PLL）、ヘアリーセル白血病（HLL）およびワルデンシュトレームマクログロブリン血症（WM）が挙げられるが、これらに限定されない。ある特定の形態の悪性リンパ腫としては、非ホジキンリンパ腫およびその異型、末梢T細胞リンパ腫、成人T細胞白血病／リンパ腫（ATL）、皮膚T細胞性リンパ腫（CTCL）、大顆粒リンパ性白血病（LGF）、ホジキン病およびリード・シュテルンベルク病が挙げられるが、これらに限定されない。細胞増殖性障害は、非内分泌腫瘍または内分泌腫瘍であるときがある。非内分泌腫瘍の例証的な例としては、腺癌、腺房細胞癌腫、腺扁平上皮癌腫、巨細胞腫、管内乳頭粘液性新生物、粘液性嚢胞腺癌、膵芽腫、漿液性嚢胞腺腫、充実性偽乳頭状腫瘍が挙げられるが、これらに限定されない。内分泌腫瘍は、島細胞腫瘍であるときがある。

【0377】

いくつかの実施形態において、消耗性障害もしくは症状または変性障害もしくは症状は、肝硬変、筋萎縮性側索硬化症（ALS）、アルツハイマー病、パーキンソン病、多系統萎縮症、アテローム性動脈硬化症、進行性核上性麻痺、テイ・サックス病、糖尿病、心疾患、円錐角膜、炎症性腸疾患（IBD）、前立腺炎、変形性関節症、骨粗鬆症、関節リウマチ、ハンチントン病、慢性外傷性脳障害、慢性閉塞性肺疾患（COPD）、結核、慢性下痢、後天性免疫不全症候群（AIDS）、上腸間膜動脈症候群などまたはそれらの組み合わせである。

【0378】

いくつかの実施形態において、自己免疫障害または症状は、急性散在性脳脊髄炎（ADEM）、アジソン病、円形脱毛症、強直性脊椎炎、抗リン脂質抗体症候群（APS）、自己免疫性溶血性貧血、自己免疫性肝炎、自己免疫性内耳疾患、水疱性類天疱瘡、セリアック病、シャーガス病、慢性閉塞性肺疾患、クローン病（特発性炎症性腸疾患「IBD」の1タイプ）、皮膚筋炎、真性糖尿病タイプ1、子宮内膜症、グッドパスチャー症候群、グレーヴズ病、ギラン・バレー症候群（GBS）、橋本病、化膿性汗腺炎、特発性血小板減少性紫斑病、間質性膀胱炎、エリテマトーデス、混合結合組織病、モルヘア、多発性硬化症（MS）、重症筋無力症、ナルコレプシー、神経性筋強直症（*euromyotonia*）、尋常性天疱瘡、悪性貧血、多発性筋炎、原発性胆汁性肝硬変、関節リウマチ、統合失調症、強皮症、シェーグレン症候群、側頭動脈炎（「巨細胞性動脈炎」としても知られる）、潰瘍性大腸炎（特発性炎症性腸疾患「IBD」の1タイプ）、脈管炎、白斑、ウェゲナー肉芽腫症などまたはそれらの組み合わせである。

癌

【0379】

いくつかの実施形態において、異常な細胞増殖の症状（例えば、癌、腫瘍、新生物）の存在もしくは非存在は、本明細書中に記載される方法または装置を用いることによって判

10

20

30

40

50

定される。例えば、血清中の無細胞核酸のレベルは、健常患者と比べて、様々なタイプの癌を有する患者において上昇し得る。転移性疾患を有する患者は、例えば、非転移性患者よりもおよそ2倍高い血清DNAレベルを有し得るときがある。転移性疾患を有する患者は、例えば、癌特異的マーカーおよび／またはある特定の単一ヌクレオチド多型もしくは短いタンデム反復によっても識別され得る。高レベルの循環DNAと正に相関し得る癌のタイプの非限定的な例としては、乳癌、直腸結腸癌、消化器癌、肝細胞癌、肺癌、メラノーマ、非ホジキンリンパ腫、白血病、多発性骨髄腫、膀胱癌、ヘパトーマ、子宮頸癌、食道癌、膵癌および前立腺癌が挙げられる。様々な癌が、非癌性の健康な細胞由来の核酸と区別可能な特色（例えば、エピジェネティックな状態ならびに／または配列のバリエーション、重複および／もしくは欠失）を有する核酸を有し得、血流中に放出し得るときがある。そのような特色は、例えば、特定のタイプの癌に特異的であり得る。したがって、本明細書中に提供される方法は、特定のタイプの癌を識別するために使用され得ることがさらに企図される。

10

子癇前症

【0380】

いくつかの実施形態において、子癇前症の存在もしくは非存在は、本明細書中に記載される方法、機器または装置を用いることによって判定される。子癇前症は、妊娠中に高血圧症が生じ（例えば、妊娠誘発性高血圧症）、尿中のかなりの量のタンパク質に関連する症状である。ある特定の実施形態において、子癇前症は、高レベルの細胞外核酸および／またはメチル化パターンの変化にも関連する。例えば、細胞外の胎児由来高メチル化RASFS1Aレベルと子癇前症の重症度との間に正の相関が観察された。ある特定の例において、DNAメチル化の増加が、正常なコントロールと比較して、子癇前症の胎盤においてH19遺伝子に対して観察される。

20

【0381】

子癇前症は、世界中の母体および胎児／新生児の死亡および罹患の主な原因の1つである。血漿中および血清中の循環無細胞核酸は、出生前診断をはじめとした、種々の医学分野において有望な臨床応用を有する新規バイオマーカーである。切迫子癇前症に対する指標としての、母体血漿中の無細胞胎児（cff）DNAの定量的変化は、例えば、男性特異的SRYまたはDYS14遺伝子座に対するリアルタイム定量的PCRを用いた種々の研究において報告されている。早期発症型子癇前症の場合、妊娠第1期に高レベルが見られることがある。症状発生前の高レベルのcffDNAは、絨毛間腔内の低酸素／再酸素化に起因し、組織酸化ストレスならびに胎盤のアポトーシスおよびネクローシスの増加に至ることがある。母体循環へのcffDNAの流出増加に対する証拠に加えて、子癇前症では、cffDNAの腎クリアランスの減少に対する証拠も存在する。胎児DNAの量は、現在、Y染色体特異的配列を定量することによって測定されるので、全無細胞DNAの計測または性別非依存性の胎児エピジェネティックマーカー（例えば、DNAメチル化）の使用などの代替アプローチが、代替法を提供する。胎盤起源の無細胞RNAは、実地臨床において子癇前症のスクリーニングおよび診断に使用され得る別の代替バイオマーカーである。胎児のRNAは、それを分解から保護する細胞内胎盤粒子に関連する。胎児のRNAレベルは、子癇前症を有する妊婦において、コントロールと比較して10倍高いときがあり、ゆえに、実地臨床において子癇前症のスクリーニングおよび診断に使用され得る代替バイオマーカーである。

30

40

病原体

【0382】

いくつかの実施形態において、病原性の症状の存在もしくは非存在は、本明細書中に記載される方法、機器または装置によって判定される。病原性の症状は、細菌、ウイルスまたは真菌を含むがこれらに限定されない病原体による宿主の感染によって引き起こされ得る。病原体は、代表的には、宿主核酸と区別可能であり得る核酸（例えば、ゲノムDNA、ゲノムRNA、mRNA）を有するので、本明細書中に提供される方法、機器および装置を用いることにより、病原体の存在もしくは非存在を判定することができる。病原体は

50

、特定の病原体に特有の特色、例えば、エピジェネティックな状態、ならびに／または1つもしくはそれを超える配列のバリエーション、重複および／もしくは欠失を有する核酸を有することが多い。したがって、本明細書中に提供される方法を用いることにより、特定の病原体または病原体バリエーション（例えば、系統）が特定され得る。

【実施例】

【0383】

下記に示される実施例は、ある特定の実施形態を例証するものであって、本技術を限定するものではない。

実施例1：アラインメントを必要としない染色体カウント正規化の特徴

【0384】

この実施例に記載される方法は、正規化において複数の染色体を使用せずに、全ゲノム配列決定解析に関する染色体提示を算出する代替方法を提供する。様々なタイプの分子診断（例えば、非侵襲性の出生前診断）が、目的のサンプルのゲノム提示の標準化された値を、予め定められたカットオフと比較することに依存している。場合によっては、このゲノム提示は、全ゲノム配列決定実験から得られ、ここで、配列決定されたリードは、まず、参照ゲノムにアラインメントされる。いくつかの配列決定プラットフォームの場合、シーケンシングリードの総数の有意なばらつきは、実験条件自体に応じて存在するが、本質的には固有の生物学的特性に応じては存在しない。このため、そのゲノム提示は、正規化工程を含むことが多く、ある特定の領域にアラインメントされたリードを、他の領域（まさしくその目的の領域も含み得る）にアラインメントされたリードによって除算する。例えば、MaterniT21テスト（Sequenom, Inc., San Diego, California）では、染色体提示は、すべての常染色体上にアラインメントされたリードに対する、目的の染色体上にアラインメントされたリードの比率として算出される。この正規化工程において構築され得る様々なタイプの比率は、これらの比率から導かれる診断の精度全体に対して様々に関連し得る。これまでに、そのような比率は、アラインメントされたリードに基づいて（様々な配列アラインメントツールおよび参照ゲノムを用いて）算出されてきた。

【0385】

一般的な参照ゲノムに対する従来のアラインメント工程の非存在下において染色体提示を推測する方法が、本明細書の以後に記載される。

- a. 染色体提示は、目的の染色体（例えば、chr21）にアラインメントされたリードとシーケンシングリードの数（任意のアラインメントの前）との比率として定義される
- b. 染色体提示は、目的の染色体（例えば、chr21）にアラインメントされたリードと、任意の品質管理メトリックによってフィルタリングされたシーケンシングリードの数（任意のアラインメントの前）（例えば、chastityフィルターを通過するリード）との比率として定義される

【0386】

図1は、最近の研究（LDTv4CE2）において観察された、リードの総数（アラインメント前）とchastityフィルターを通過するリードの総数（アラインメント前）との比較を示している。

【0387】

図2は、最近の研究（LDTv4CE2）において観察された、chastityフィルターを通過するリードの総数（アラインメント前）と、すべての常染色体にアラインメントされたリードとの比較を示している。

【0388】

図3A、図3Bおよび図3Cは、21、13および18番染色体に対してGC-LOESS正規化に続いて主成分正規化を用いたときの、常染色体を用いて算出された染色体提示およびchastityフィルターを通過するアラインメント前のリードを用いて算出された染色体提示から導かれたz得点の比較を示している。

【0389】

フィルターを通過したアラインメント前のリードを用いて算出された染色体提示に基づいて測定される異数性検出の精度は、下記の表 2 ～ 4 に示され、L D T v 4 C E 2 研究の精度と同一であると見出された。

【表 2】

表2

	実際		
LDTv4		T21	正倍数体
	T21	21	0
	正倍数体	0	313
	合計	21	313

10

【表 3】

表3

	実際		
LDTv4		T18	正倍数体
	T18	6	0
	正倍数体	1	328
	合計	7	328

20

【表 4】

表4

	実際		
LDTv4		T13	正倍数体
	T13	7	0
	正倍数体	0	328
	合計	7	328

【0 3 9 0】

実施例 2：アラインメントを必要としないさらなる染色体カウント正規化の特徴

30

【0 3 9 1】

実施例 1 に記載された方法に対する代替法として、一般的な参照ゲノムに対する従来のアラインメント工程の非存在下において染色体提示を推測する方法が本明細書の以後に記載される。これらの方法のいくつかは、アラインメントされたリードが分子と分母の両方に対して使用されることを必要とせずに染色体提示を算出する代替方法を提供する。

a. 染色体提示は、目的の染色体（例えば、chr 21）にアラインメントされたリードのサブセットと、任意の品質管理メトリックによってフィルタリングされるか、またはフィルタリングされない（例えば、chastity フィルターを通過するリード）、所与のサブセットからのシーケンシングリードの数（任意のアラインメントの前）との比率として定義される

40

b. 染色体提示は、目的の染色体（例えば、chr 21）にアラインメントされたリードのサブセットと、ヌクレオチド組成によってフィルタリングされた（例えば、特定の範囲内の GC 含有量を有するリード）、所与のサブセットからのシーケンシングリードの数（任意のアラインメントの前）との比率として定義される

c. 染色体提示は、リードのカスタム辞書にマッチするリードのサブセット（予め配列決定されたサンプルから得られ、目的の染色体に予めアラインメントされたもの）と上記 a ～ d において定義された変数のいずれかとの比率として定義される

d. 染色体提示は、目的の染色体にアラインメントされるかまたはカスタム辞書にマッチするリードと、参照ゲノムのサブセットにアラインメントされないリード（「アラインメント不可能」）との比率として定義される

50

【0392】

図4は、カウントA (N t a r g e t と称される、480)を生成するために、上記(c)および(d)に記載されたカスタム辞書を利用する方法の例を示している。図4に示されているように、分母N t o tに対するリード数は、シークエンサーからリードに対する生ファイル入手することによって生成される(410)。そのプロセスは、N t o t カウントを生成するために、それらのファイルを各テストサンプルに対する個々のF A S T Qファイルに変換すること(430)、およびc h a s t i t y フィルター(画質フィルター、440)に従ってフィルタリングされて除外されたリードを除く、テストサンプルに対するリードの総数をカウントすることを含む。c h a s t i t y フィルターの代わりに、またはc h a s t i t y フィルターに加えて、他のフィルターを使用することができる。例えば、GCパーセンテージ(例えば、30%~60%のGCパーセンテージ)に基づくフィルターが、リードをフィルタリングするために使用され得る(440)。また、複雑さの低いリード(例えば、50%超のリピートを有するリード)を除去するフィルターが、リードをフィルタリングするために使用され得る(440)。

【0393】

図4に示されているように、参照サンプルまたは参照サンプルセットからのリードを、ヒト参照ゲノムにアラインメントし(450)、各染色体に対してリードの辞書(サブリスト)を作成する。それらの各辞書は、その辞書が作成された特定の染色体にユニークにマッピングされたリード(ポリヌクレオチド; k - m e r)を含む(460)。目的の染色体に対する辞書が、標的染色体に対して選択され、テストサンプルからのリード(430)を、その辞書におけるポリヌクレオチドと比較し(470)、その辞書の中のポリヌクレオチドとマッチするリードをカウントする(N t a r g e t 分子、480)。比較(470)は、通常、各リードのマッピングされた位置を返さず、リードが標的染色体に属するかまたは属さないかに関する2値結果を与える。標的染色体に対するカウント提示(染色体割合(c h r o m o s o m e f r a c t i o n)、正規化された染色体カウント)確定のために、N t o t カウントが分母として使用され、N t a r g e t カウントが分子として使用される(490)。

実施例3:ある特定の実施形態の例

【0394】

本技術のある特定の実施形態の非限定的な例が本明細書の以後に列举される。

【0395】

A1. 診断テストのためにゲノムセグメントの配列リードカウント提示を確定するための方法であって、該方法は、

(a) あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウントAを提供する工程であって、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、工程;

(b) 該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウントBを提供する工程であって、ここで、該カウントBは、参照ゲノムにアラインメントされない配列リードのカウントである、工程;および

(c) 該カウントAと該カウントBとの比率として、該セグメントに対するカウント提示を確定する工程を含む、方法。

【0396】

A1. 1. (b)におけるゲノムのサブセットが、(a)におけるセグメントより大きい、実施形態A1に記載の方法。

【0397】

A1. 2. 前記カウントBが、前記配列リードを参照ゲノムにアラインメントする工程を含まないプロセスによって決定される、実施形態A1またはA1. 1に記載の方法。

【0398】

A 2. 前記カウント B が、

(i) 前記テストサンプルからの核酸を配列決定するために使用される核酸配列決定プロセスによって生成された全リードのカウント；

(i i) 該テストサンプルからの核酸を配列決定するために使用される核酸配列決定プロセスによって生成された全リードの一部のカウント；

(i i i) 該配列決定プロセスに対する品質管理メトリックに従ってフィルタリングされたリードを除く、(i) の全リードまたは (i i) の全リードの一部のカウント；

(i v) 該配列決定プロセスに対する品質管理メトリックに従って重み付けされた、(i) の全リードまたは (i i) の全リードの一部のカウント；

(v) リードの塩基含有量に従ってフィルタリングされたリードを除く、(i) の全リードまたは (i i) の全リードの一部のカウント；

(v i) リードの塩基含有量に従って重み付けされた、(i) の全リードまたは (i i) の全リードの一部のカウント；または

(v i i) あるリストの中のポリヌクレオチドとマッチするリードのカウントであって、ここで、該リードは、該リストの中のポリヌクレオチドとリードを比較する工程を含むプロセスにおいて、該リストの中のポリヌクレオチドとマッチするかまたはマッチしないと判定され、ここで、該リードは、(i) における全リード、(i i) における全リードの一部、(i i i) の品質管理メトリックに従ってフィルタリングされたリードを除く、(i) の全リードもしくは (i i) の全リードの一部、(i v) の品質管理メトリックに従って重み付けされた、(i) の全リードもしくは (i i) の全リードの一部、(v) のリードの塩基含有量に従ってフィルタリングされたリードを除く、(i) の全リードもしくは (i i) の全リードの一部、または (v i) のリードの塩基含有量に従って重み付けされた、(i) の全リードもしくは (i i) の全リードの一部である、リードのカウントである、実施形態 A 1 ~ A 1. 2 のいずれか 1 つに記載の方法。

【0399】

A 3. 前記一部が、前記全リードからランダムに選択されたリードの一部である、実施形態 A 2 に記載の方法。

【0400】

A 4. 前記一部が、前記全リードの約 10% ~ 約 90% である、実施形態 A 2 または A 3 に記載の方法。

【0401】

A 5. 前記核酸配列決定プロセスが、画像処理を含み、前記品質管理メトリックが、画質に基づく、実施形態 A 2 ~ A 4 のいずれか 1 つに記載の方法。

【0402】

A 6. 前記品質管理メトリックが、画像の重なりの評価に基づく、実施形態 A 5 に記載の方法。

【0403】

A 7. 前記リードの塩基含有量が、グアニンおよびシトシン (G C) 含有量である、実施形態 A 2 ~ A 6 のいずれか 1 つに記載の方法。

【0404】

A 8. (v) においてフィルタリングされた前記リードが、第 1 の G C 閾値未満の G C 含有量を有する、実施形態 A 7 に記載の方法。

【0405】

A 8. 1. (v) においてフィルタリングされた前記リードが、第 2 の G C 閾値を超える G C 含有量を有する、実施形態 A 7 に記載の方法。

【0406】

A 9. (v i i) における前記カウントが、前記リストの中のポリヌクレオチドの配列およびサイズと正確にマッチするリードのカウントである、実施形態 A 2 ~ A 8. 1 のいずれか 1 つに記載の方法。

【0407】

10

20

30

40

50

A 9. 1. 前記リストの中のポリヌクレオチドが、(a)の前に、参照ゲノムまたは参照ゲノムにおけるサブセットにアラインメントされたものである、実施形態 A 2 ~ A 9 のいずれか 1 つに記載の方法。

【0408】

A 9. 2. 前記参照ゲノムにおけるサブセットが、すべての常染色体またはすべての常染色体のサブセットである、実施形態 A 9. 1 に記載の方法。

【0409】

A 9. 3. 前記比較が、(i) 各ポリヌクレオチドがアラインメントする染色体、および／または (i i) 各ポリヌクレオチドがアラインメントする染色体の位置番号を追跡する工程を含まない、実施形態 A 9. 1 または A 9. 2 に記載の方法。

10

【0410】

A 10. リードを参照ゲノムとアラインメントするアラインメントプロセスに前記リードを供する工程を含み、ここで、前記カウント B は、該リードを該アラインメントプロセスに供する前に決定される、実施形態 A 1 ~ A 9. 3 のいずれか 1 つに記載の方法。

【0411】

A 11. リードを参照ゲノムとアラインメントするアラインメントプロセスに前記リードを供する工程を含み、ここで、前記カウント B は、該アラインメントプロセスによって該参照ゲノムにアラインメントされないリードのカウントである、実施形態 A 1 に記載の方法。

【0412】

A 12. リードを参照ゲノムとアラインメントするアラインメントプロセスに前記リードを供する工程を含み、ここで、前記カウント A は、該参照ゲノムにおけるセグメントにアラインメントされたリードのカウントである、実施形態 A 1 ~ A 11 のいずれか 1 つに記載の方法。

20

【0413】

A 13. 前記カウント A が、前記配列リードを参照ゲノムにアラインメントする工程を含まないプロセスによって決定される、実施形態 A 1 ~ A 11 のいずれか 1 つに記載の方法。

【0414】

A 14. 前記カウント A が、あるリストまたはあるリストのサブセットの中のポリヌクレオチドとマッチするリードのカウントであり、ここで、該リードは、リードを該リストまたは該リストのサブセットの中のポリヌクレオチドと比較する工程を含むプロセスにおいて、該リストまたは該リストのサブセットの中のポリヌクレオチドとマッチするかまたはマッチしないと判定される、実施形態 A 13 に記載の方法。

30

【0415】

A 14. 1. 前記リストまたは前記リストのサブセットの中のポリヌクレオチドと比較された前記リードが、実施形態 A 2 (i) における全リード；実施形態 A 2 (i i) における全リードの一部；実施形態 A 2 (i i i) の品質管理メトリックに従ってフィルタリングされたリードを除く、実施形態 A 2 (i) の全リードもしくは実施形態 A 2 (i i) の全リードの一部；実施形態 A 2 (i v) の品質管理メトリックに従って重み付けされた、実施形態 A 2 (i) の全リードもしくは実施形態 A 2 (i i) の全リードの一部；実施形態 A 2 (v) のリードの塩基含有量に従ってフィルタリングされたリードを除く、実施形態 A 2 (i) の全リードもしくは実施形態 A 2 (i i) の全リードの一部；または実施形態 A 2 (v i) のリードの塩基含有量に従って重み付けされた、実施形態 A 2 (i) の全リードもしくは実施形態 A 2 (i i) の全リードの一部である、実施形態 A 14 に記載の方法。

40

【0416】

A 14. 2. 前記カウント A が、前記リストまたは前記リストのサブセットの中のポリヌクレオチドの配列およびサイズと正確にマッチするリードのカウントである、実施形態 A 14 または A 14. 1 に記載の方法。

50

【0417】

A14. 3. 前記リストまたは前記リストのサブセットの中のポリヌクレオチドが、(a)の前に、参照ゲノムにおけるセグメントにアラインメントされたものである、実施形態A14～A14. 2のいずれか1つに記載の方法。

【0418】

A14. 4. 前記比較が、(i)各ポリヌクレオチドがアラインメントする染色体、および/または(ii)各ポリヌクレオチドがアラインメントする染色体の位置番号の追跡を含まない、実施形態A14. 3に記載の方法。

【0419】

A14. 5. 前記配列リードが、(a)、(b)および(c)において前記配列リードを前記参照ゲノムにアラインメントするアラインメントプロセスに供されない、実施形態A1～A9. 3およびA13～A14. 4のいずれか1つに記載の方法。

10

【0420】

A14. 6. 前記配列リードが、前記診断テストにおいて該配列リードを前記参照ゲノムにアラインメントするアラインメントプロセスに供されない、実施形態A1～A9. 3およびA13～A14. 4のいずれか1つに記載の方法。

【0421】

A15. 前記セグメントが、染色体である、実施形態A1～A14. 6のいずれか1つに記載の方法。

【0422】

A16. 前記染色体が、13番染色体、18番染色体および21番染色体から選択される、実施形態A15に記載の方法。

20

【0423】

A17. 前記セグメントが、染色体のセグメントである、実施形態A1～A14のいずれか1つに記載の方法。

【0424】

A18. 前記セグメントが、微小重複領域または微小欠失領域である、実施形態A17に記載の方法。

【0425】

A19. (c)における前記比率が、前記カウントBで除算された前記カウントAである、実施形態A1～A18のいずれか1つに記載の方法。

30

【0426】

A20. (c)における前記比率が、前記カウントAで除算された前記カウントBである、実施形態A1～A18のいずれか1つに記載の方法。

【0427】

A21. 前記核酸が、循環無細胞核酸である、実施形態A1～A20のいずれか1つに記載の方法。

【0428】

A22. 前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来である、実施形態A1～A21のいずれか1つに記載の方法。

40

【0429】

A23. 前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてのテストである、実施形態A1～A21のいずれか1つに記載の方法。

【0430】

A24. 前記セグメントに対するカウント提示の統計量を決定する工程を含む、実施形態A1～A23のいずれか1つに記載の方法。

【0431】

A25. 前記統計量が、z得点である、実施形態A24に記載の方法。

【0432】

A26. 前記z得点が、(a)(i)前記テストサンプルに対する前記セグメントに対

50

するカウント提示から (i i) サンプルセットに対する該セグメントに対するカウント提示の中央値を除いた減算の結果を、(b) 該サンプルセットに対するセグメントに対するカウント提示の M A D で除算した商である、実施形態 A 2 5 に記載の方法。

【0433】

A 2 7. 前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来であり、前記サンプルセットが、正倍数性胎児の妊娠を有する被験体に対するサンプルのセットである、実施形態 A 2 6 に記載の方法。

【0434】

A 2 8. 前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来であり、前記サンプルセットが、トリソミー胎児の妊娠を有する被験体に対するサンプルのセットである、実施形態 A 2 6 に記載の方法。

10

【0435】

A 2 9. 前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてであり、前記サンプルセットが、該細胞増殖の症状を有する被験体に対するサンプルのセットである、実施形態 A 2 6 に記載の方法。

【0436】

A 3 0. 前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてであり、前記サンプルセットが、該細胞増殖の症状を有しない被験体に対するサンプルのセットである、実施形態 A 2 6 に記載の方法。

【0437】

A 3 1. 前記カウント A が、正規化されたカウントのものである、実施形態 A 1 ~ A 3 0 のいずれか 1 つに記載の方法。

20

【0438】

A 3 2. 前記カウント B が、正規化されたカウントのものである、実施形態 A 1 ~ A 3 1 のいずれか 1 つに記載の方法。

【0439】

A 3 3. 前記正規化されたカウントが、L O E S S 正規化プロセスを含む正規化プロセスによって生成される、実施形態 A 3 1 または A 3 2 に記載の方法。

【0440】

A 3 4. 前記正規化されたカウントが、グアニンおよびシトシン (G C) バイアス正規化を含む正規化プロセスによって生成される、実施形態 A 3 1 ~ A 3 3 のいずれか 1 つに記載の方法。

30

【0441】

A 3 5. 前記正規化されたカウントが、G C バイアスの L O E S S 正規化 (G C - L O E S S) を含む正規化プロセスによって生成される、実施形態 A 3 1 ~ A 3 4 のいずれか 1 つに記載の方法。

【0442】

A 3 6. 前記正規化されたカウントが、主成分正規化を含む正規化プロセスによって生成される、実施形態 A 3 1 ~ A 3 5 のいずれか 1 つに記載の方法。

【0443】

A 3 7. 前記診断テストが、出生前診断テストであり、前記テストサンプルが、胎児を有する妊婦由来であり、該診断テストは、遺伝的変異の非存在の存在を判定することを含む、実施形態 A 1 ~ A 3 6 のいずれか 1 つに記載の方法。

40

【0444】

A 3 8. 前記遺伝的変異が、染色体異数性である、実施形態 A 3 7 に記載の方法。

【0445】

A 3 9. 前記染色体異数性が、染色体全体の 1、3 または 4 コピーである、実施形態 A 3 8 に記載の方法。

【0446】

A 4 0. 前記遺伝的変異が、微小重複または微小欠失である、実施形態 A 3 7 に記載の

50

方法。

【0447】

A 4 1. 前記遺伝的変異が、胎児の遺伝的変異である、実施形態 A 3 7 ~ A 4 0 のいずれか 1 つに記載の方法。

【0448】

A 4 2. 前記診断テストが、細胞増殖の症状の存在、非存在、高リスクまたは低リスクについてであり、該診断テストは、遺伝的変異の非存在の存在を判定することを含む、実施形態 A 1 ~ A 3 6 のいずれか 1 つに記載の方法。

【0449】

A 4 3. 前記遺伝的変異が、微小重複または微小欠失である、実施形態 A 4 2 に記載の方法。

10

【0450】

A 4 4. (a)、(b) および (c) のうちの 1 つもしくはそれを超えるものまたはすべてが、システムにおけるマイクロプロセッサによって行われる、実施形態 A 1 ~ A 4 3 のいずれか 1 つに記載の方法。

【0451】

A 4 5. (a)、(b) および (c) のうちの 1 つもしくはそれを超えるものまたはすべてが、システムにおけるメモリーと連動して行われる、請求項 A 1 ~ A 4 4 のいずれか 1 項に記載の方法。

【0452】

A 4 6. (a)、(b) および (c) のうちの 1 つもしくはそれを超えるものまたはすべてが、コンピュータによって行われる、実施形態 A 1 ~ A 4 5 のいずれか 1 つに記載の方法。

20

【0453】

B 1. 1 つまたはそれを超えるマイクロプロセッサおよびメモリーを備えるシステムであって、該メモリーは、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示を備え、該メモリーは、ヌクレオチド配列リードを備え、該配列リードは、被験体由来のテストサンプルからの核酸のリードであり、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示は、以下：

(a) マイクロプロセッサを用いて、あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウント A を提供することであって、ここで、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、カウント A を提供すること；

30

(b) マイクロプロセッサを用いて、該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウント B を提供することであって、ここで、該カウント B は、参照ゲノムにアラインメントされない配列リードのカウントである、カウント B を提供すること；および

(c) 該カウント A と該カウント B との比率として、該セグメントに対するカウント提示を確定すること

を行うように構成される、システム。

40

【0454】

B 2. 1 つまたはそれを超えるマイクロプロセッサおよびメモリーを備える機器であって、該メモリーは、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示を備え、該メモリーは、ヌクレオチド配列リードを備え、該配列リードは、被験体由来のテストサンプルからの核酸のリードであり、該 1 つまたはそれを超えるマイクロプロセッサによって実行可能な指示は、以下：

(a) マイクロプロセッサを用いて、あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウント A を提供することであって、ここで、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、カウント A を提供すること；

50

(b) マイクロプロセッサを用いて、該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウントBを提供することであって、ここで、該カウントBは、参照ゲノムにアラインメントされない配列リードのカウントである、カウントBを提供すること；および

(c) 該カウントAと該カウントBとの比率として、該セグメントに対するカウント提示を確定すること

を行うように構成される、機器。

【0455】

B3. 実行可能なプログラムが格納されている非一時的なコンピュータ可読記憶媒体であって、該プログラムは、マイクロプロセッサに以下：

(a) 被験体由来のテストサンプルからの核酸のリードであるヌクレオチド配列リードにアクセスすること；

(b) マイクロプロセッサを用いて、あるゲノムセグメントに対する核酸配列リードのカウントを生成し、それにより、該セグメントに対するカウントAを提供することであって、ここで、該配列リードは、該ゲノムを有する被験体由来のテストサンプルからの核酸のリードである、カウントAを提供すること；

(c) マイクロプロセッサを用いて、該ゲノムまたは該ゲノムのサブセットに対する核酸配列リードのカウントを生成し、それにより、該ゲノムまたは該ゲノムのサブセットに対するカウントBを提供することであって、ここで、該カウントBは、参照ゲノムにアラインメントされない配列リードのカウントである、カウントBを提供すること；および

(d) 該カウントAと該カウントBとの比率として、該セグメントに対するカウント提示を確定すること

を行うように指示する、非一時的なコンピュータ可読記憶媒体。

【0456】

図面は、本技術のある特定の実施形態を例証するものであって、限定するものではない。例証を明確にするためおよび平易にするために、図面は、一定尺度で拡大縮小して作成されておらず、場合によっては、特定の実施形態の理解を促すために、様々な態様が誇張されてまたは拡大されて示されていることがある。

【0457】

本明細書中で参照される各特許、特許出願、刊行物および文書の全体が、参照により本明細書に組み込まれる。上記特許、特許出願、刊行物および文書の引用は、前述のいずれもが適切な従来技術であることを自認するものではないし、その引用が、これらの刊行物または文書の内容または日付に関するいずれの自認も構成しない。

【0458】

本技術の基本的な態様から逸脱することなく、前述のものに対して改変が行われ得る。本技術は、1つまたはそれを超える特定の実施形態に照らして実質的に詳細に記載されてきたが、当業者は、本願に具体的に開示された実施形態に対して変更が行われ得るが、これらの改変および改善は、本技術の範囲内および精神の範囲内であることを認識するだろう。

【0459】

本明細書中に例証的に記載された技術は、本明細書中に具体的に開示されていない任意の要素の非存在下において適切に実施され得る。したがって、例えば、本明細書中の各場合において、用語「～を含む」、「～から本質的になる」および「～からなる」のいずれもが、他の2つの用語のいずれかと置き換えられてもよい。用いられてきた用語および表現は、説明の用語であって限定でない用語として使用され、そのような用語および表現の使用は、示されたおよび記載された特徴またはその一部の等価物を排除せず、特許請求される技術の範囲内で様々な改変が可能である。用語「a」または「an」は、それらの要素の1つまたはそれらの要素の1つより多い要素が記載されていることが文脈上明らかでない限り、それが修飾する要素の1つまたは複数のことを指し得る（例えば、「試薬（a reagent）」は、1つまたはそれを超える試薬

10

20

30

40

50

を意味し得る)。用語「約」は、本明細書中で使用されるとき、基礎をなすパラメータの10%以内の値(すなわち、プラスまたはマイナス10%)のことを指し、一連の値の最初における用語「約」の使用は、それらの値の各々を修飾する(すなわち、「約1、2および3」は、約1、約2および約3のことを指す)。例えば、「約100グラム」という重量は、90グラム～110グラムの重量を含み得る。さらに、値のリストが本明細書中に記載されるとき(例えば、約50%、60%、70%、80%、85%または86%)、そのリストは、それらのすべての中間値および小数値(例えば、54%、85.4%)を含む。したがって、本技術は、代表的な実施形態および随意の特徴によって具体的に開示されてきたが、本明細書中に開示される概念の改変およびバリエーションは、当業者によって用いられ得、そのような改変およびバリエーションは、本技術の範囲内であると見なされることが理解されるべきである。

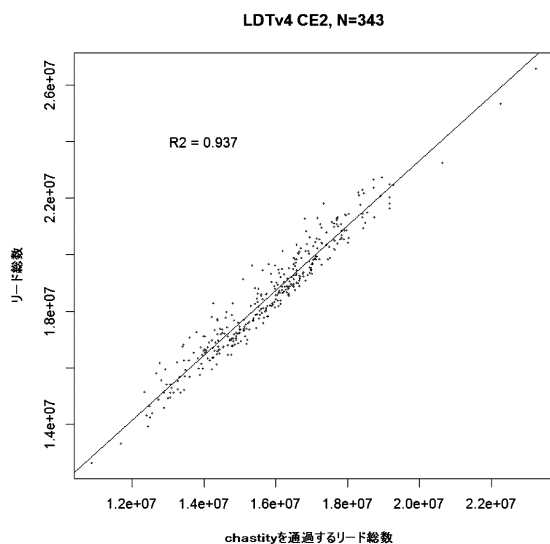
10

【0460】

本技術のある特定の実施形態は、以下の請求項に示される。

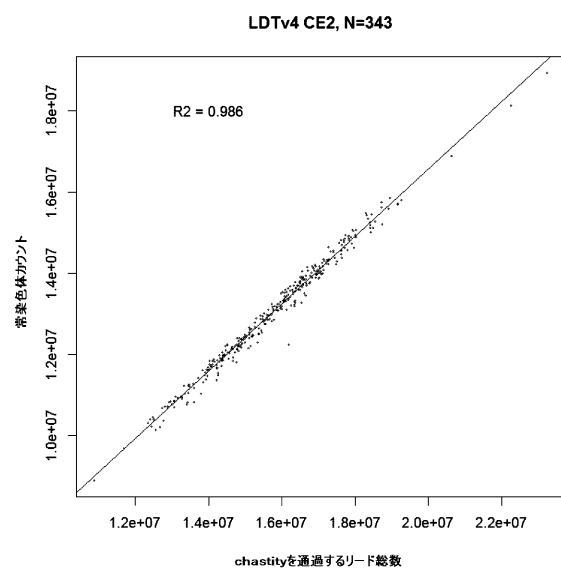
【図1】

FIG. 1

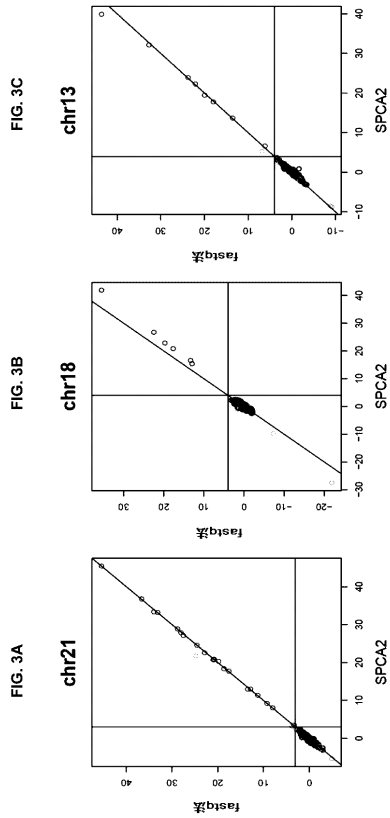


【図2】

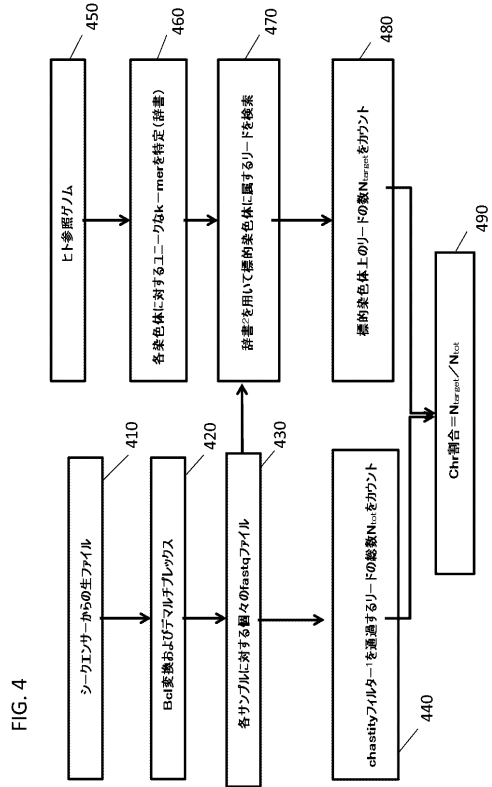
FIG. 2



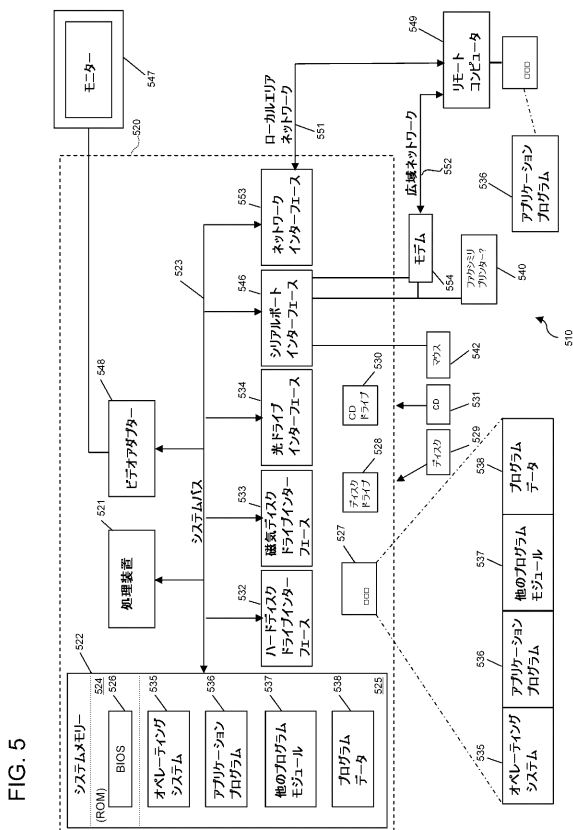
【 図 3 】



【図 4】



【图 5】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/032550

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06F19/18 G06F19/22 C12Q1/68 G06F19/24
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2013/130921 A1 (GAO YANG [CN] ET AL) 23 May 2013 (2013-05-23) paragraphs [0007] - [0008], [0025]; claims 25-30 page 3	1-61
Y	US 2013/150253 A1 (DECIU COSMIN [US] ET AL) 13 June 2013 (2013-06-13) paragraphs [0006], [0007], [0232]	1-61
Y	STEPHANIE C. Y. YU ET AL: "Noninvasive Prenatal Molecular Karyotyping from Maternal Plasma", PLOS ONE, vol. 8, no. 4, 17 April 2013 (2013-04-17), pages 1-8, XP055160833, DOI: 10.1371/journal.pone.0060968 page 3	1-61
----- -/-		

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents :

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

E earlier application or patent but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

Z document member of the same patent family

Date of the actual completion of the international search

24 September 2015

Date of mailing of the international search report

02/10/2015

Name and mailing address of the ISA/

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

Authorized officer

Schmitt, Anja

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/032550

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2014/039556 A1 (GUARDANT HEALTH INC [US]; TALASAZ AMIRALI [US]; ELTOUKHY HELMY [US]) 13 March 2014 (2014-03-13) claims 2,77 -----	1-61
A	Sander Bollen: "BioConductor: Microarray versus Next-Generation Sequencing toolsets", 11 February 2014 (2014-02-11), pages 1-14, XP055215571, Retrieved from the Internet: URL:http://dspace.library.uu.nl/bitstream/handle/1874/290489/Sander_Bollen_writing_assignment.pdf [retrieved on 2015-09-23] the whole document -----	1-61

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2015/032550

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
US 2013130921	A1	23-05-2013	CN	103080336 A		01-05-2013
			EP	2716766 A1		09-04-2014
			JP	2014516541 A		17-07-2014
			US	2013130921 A1		23-05-2013
			WO	2012162884 A1		06-12-2012

US 2013150253	A1	13-06-2013	NONE			

WO 2014039556	A1	13-03-2014	CA	2883901 A1		13-03-2014
			CN	104781421 A		15-07-2015
			EP	2893040 A1		15-07-2015
			HK	1201080 A1		21-08-2015
			KR	20150067161 A		17-06-2015
			WO	2014039556 A1		13-03-2014

フロントページの続き

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

(特許庁注：以下のものは登録商標)

1. iPad

(72)発明者 ザオ, チェン

アメリカ合衆国 カリフォルニア 92122, サン ディエゴ, コスタ ベルデ ブールバード 8950, アpartment 4124

(72)発明者 デチウ, コスミン

アメリカ合衆国 カリフォルニア 92121, サン ディエゴ, シー ミスト ウェイ 10545

Fターム(参考) 2G045 AA24 AA25 CB01

4B029 AA07 BB20 CC01 FA15

4B063 QA19 QQ03 QQ08 QQ43 QQ52 QR08 QR55 QR62 QS25 QS34

QX01