

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6561046号
(P6561046)

(45) 発行日 令和1年8月14日 (2019.8.14)

(24) 登録日 令和1年7月26日 (2019.7.26)

(51) Int. Cl.	F I
C 1 2 Q 1/6869 (2018.01)	C 1 2 Q 1/6869 Z
G 1 6 B 30/00 (2019.01)	G 1 6 B 30/00
C 1 2 N 15/12 (2006.01)	C 1 2 N 15/12

請求項の数 14 (全 233 頁)

(21) 出願番号	特願2016-515122 (P2016-515122)	(73) 特許権者	504159534
(86) (22) 出願日	平成26年5月23日 (2014.5.23)		セクエノム, インコーポレイテッド
(65) 公表番号	特表2016-526879 (P2016-526879A)		アメリカ合衆国 カリフォルニア 921
(43) 公表日	平成28年9月8日 (2016.9.8)		21, サン ディエゴ, ジョン ホブ
(86) 国際出願番号	PCT/US2014/039389		キンズ コート 3595
(87) 国際公開番号	W02014/190286	(74) 代理人	100078282
(87) 国際公開日	平成26年11月27日 (2014.11.27)		弁理士 山本 秀策
審査請求日	平成29年5月2日 (2017.5.2)	(74) 代理人	100113413
(31) 優先権主張番号	61/827, 385		弁理士 森下 夏樹
(32) 優先日	平成25年5月24日 (2013.5.24)	(74) 代理人	100181674
(33) 優先権主張国・地域又は機関	米国 (US)		弁理士 飯田 貴敏
		(74) 代理人	100181641
			弁理士 石川 大輔
		(74) 代理人	230113332
			弁護士 山本 健策

最終頁に続く

(54) 【発明の名称】 遺伝子の変動の非侵襲性評価のための方法および処理

(57) 【特許請求の範囲】

【請求項 1】

胎児中の微小重複または微小欠失の存在または非存在を検出する方法であって、

(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得るステップであって、これらの配列の読取りが、胎児を出産する妊娠中の雌に由来する試験試料から得られた循環無細胞核酸の読取りであるステップと、

(b) 各部分に対してマッピングした該カウントを正規化し、それにより、計算されたゲノム区分レベルを提供するステップと、

(c) 該計算されたゲノム区分レベルに従って、ゲノムのセグメントについてのプロファイルを生成するステップと、

(d) セグメント化処理によって該プロファイルをセグメント化し、該セグメント化処理から該プロファイル中の1つまたはそれより多い他のレベルとは異なり、かつ、該プロファイル中の他のレベルよりも大きい曲線下面積 (AUC) を有する該プロファイル中のレベルを同定するステップと、

(e) (d) において同定された異なるレベルに従って、微小重複または微小欠失の存在または非存在を検出するステップと

を含み、該方法が、コンピュータが実施する方法である方法。

【請求項 2】

(b) における前記正規化が、グアニンおよびシトシン (GC) の偏りの LOESS 正規化 (GC-LOESS 正規化) を含むか、主成分正規化を含むか、またはグアニンおよ

びシトシン（GC）の偏りのLOESS正規化（GC-LOESS正規化）、およびそれに続く、主成分正規化を含む、請求項1に記載の方法。

【請求項3】

前記参照ゲノムの1つまたは複数の部分を検討事項からフィルタリングし、該1つまたは複数の部分におけるカウントを減算することを含む、請求項1または2に記載の方法。

【請求項4】

（d）における前記セグメント化処理が、サーキュラーバイナリセグメンテーションを含む、請求項1から3のいずれか一項に記載の方法。

【請求項5】

前記配列の読取りが、1倍のカバレッジまたはそのフラクションにより行われる配列決定プロセスによって生成された、請求項1から4のいずれか一項に記載の方法。

10

【請求項6】

対数オッズ比（LOR）を決定することを含み、LORが、（i）（1）遺伝子の変動を有する条件付き確率と（2）該遺伝子の変動を有する事前確率との第1の乗算の積と、（ii）（1）該遺伝子の変動を有さない条件付き確率と（2）該遺伝子の変動を有さない事前確率との第2の乗算の積との商の対数である、請求項1から5のいずれか一項に記載の方法。

【請求項7】

前記異なるレベルの定量を行うことを含む、請求項6に記載の方法。

【請求項8】

20

前記定量が、前記異なるレベルについてのカウント表示、前記異なるレベルが位置する染色体の染色体表示、または前記異なるレベルについてのカウント表示および前記異なるレベルが位置する染色体の染色体表示を含む、請求項7に記載の方法。

【請求項9】

前記定量が、前記カウント表示、前記染色体表示、または前記カウント表示および前記染色体表示のzスコアによる定量であり、

該zスコアは、任意選択で、前記異なるレベルについて、（i）試験試料カウント表示から、（ii）正倍数体カウント表示の中央値を減算して得た差を、（iii）該正倍数体カウント表示のMADで除算して得た減算の結果であり、ここで、該（i）試験試料カウント表示が、該試験試料について、全カウントを、全常染色体カウントで除算して得た比であり、該（ii）正倍数体カウント表示の中央値が、正倍数体試料について、全カウントを、全常染色体カウントで除算して得た比の中央値であり、

30

該zスコアは、任意選択で、前記染色体について、（i）試験試料カウント表示から、（ii）正倍数体カウント表示の中央値を減算して得た差を、（iii）該正倍数体カウント表示のMADで除算して得た減算の結果であり、ここで、該（i）試験試料カウント表示が、該試験試料について、前記異なるレベルが位置する該染色体中の全カウントを、全常染色体カウントで除算して得た比であり、該（ii）正倍数体カウント表示の該中央値が、正倍数体試料について、該異なるレベルが位置する該染色体中の全カウントを、全常染色体カウントで除算して得た比の中央値である

請求項8に記載の方法。

40

【請求項10】

前記遺伝子の変動を有する前記条件付き確率が、前記試験試料について決定された胎児フラクション、該試験試料について決定された前記異なるレベルについての前記カウント表示のzスコア、および該異なるレベルについての該カウント表示についてのzスコアの該胎児フラクションに関する分布に従って決定され、

前記遺伝子の変動を有する前記条件付き確率が、任意選択で、等式23：

【化38】

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

50

[式中、 f は胎児フラクションであり、 X は、該遺伝子の変動をカバーする前記異なるレベルについての部分のカウントの合計 $X \sim f(\mu X, \sigma X)$ であり、ここで、 μX および σX はそれぞれ、 X の平均値および標準偏差であり、 $f(\cdot)$ は分布関数である。]に示す関係により決定され、

前記遺伝子の変動を有する前記条件付き確率が、任意選択で、前記異なるレベルについての前記カウント表示の、前記試験試料に関する前記 z スコアと、該異なるレベルについての該カウント表示の z スコアの、前記胎児フラクションに関する分布との交差部分である

請求項9に記載の方法。

【請求項11】

10

前記遺伝子の変動を有さない前記条件付き確率が、前記試験試料について決定された前記異なるレベルについてのカウント表示の z スコアと、正倍数体中の該異なるレベルについての該カウント表示の z スコアの分布との交差部分である、請求項6に記載の方法。

【請求項12】

前記遺伝子の変動を有する前記事前確率および該遺伝子の変動を有さない該事前確率が、前記試験被験体を含まない複数の試料から決定される、請求項6、10または11のいずれか一項に記載の方法。

【請求項13】

前記試験試料について、(i)前記異なるレベルについての前記カウント表示の前記 z スコアによる定量結果が、値3.95超またはそれに等しく、かつ(ii)前記LORが、

20

ゼロ超である場合に、微小欠失または微小重複が存在することを検出すること、および前記試験試料について、(i)前記異なるレベルについての前記カウント表示の前記 z スコアによる定量結果が、値3.95未満であり、かつ/または(ii)前記LORが、ゼロ未満である場合に、微小欠失または微小重複が存在しないことを検出することを含む、請求項9から12のいずれか一項に記載の方法。

【請求項14】

前記核酸配列の読取りがマッピングされた前記参照ゲノムの部分が、ほぼ同じ長さのゲノム配列である、請求項1から13のいずれか一項に記載の方法。

【発明の詳細な説明】

【技術分野】

30

【0001】

本出願は、2013年5月24日に出願され、遺伝子の変動の非侵襲性評価のための方法および処理というタイトルであり、発明者としてZeljko Dzakulaらが挙げられ、代理人整理番号SEQ-6068-PVが指定されている、米国仮特許出願第61/827,385号の利益を請求する。本出願は、2012年11月5日に出願され、遺伝子の変動の非侵襲性評価のための方法および処理というタイトルであり、発明者としてCosmin Deciu、Zeljko Dzakula、Mathias EhrichおよびSung Kimが挙げられ、代理人整理番号SEQ-6034-CTtが指定されている、米国特許出願第13/669,136号に関連し、この出願は、2012年10月5日に出願され、遺伝子の変動の非侵襲性評価のための方法および処理というタイトルであり、発明者としてCosmin Deciu、Zeljko Dzakula、Mathias EhrichおよびSung Kimが挙げられ、代理人整理番号SEQ-6034-PCが指定されている、国際PCT出願PCT/US2012/059123の継続であり、この出願は、(i)2012年10月4日に出願され、遺伝子の変動の非侵襲性評価のための方法および処理というタイトルであり、発明者としてCosmin Deciu、Zeljko Dzakula、Mathias EhrichおよびSung Kimが挙げられ、代理人整理番号SEQ-6034-PV3が指定されている、米国仮特許出願第61/709,899号の利益を請求し、(ii)2012年6月22日に出願され、遺伝子の変動の非侵襲性評価のための方法および処理というタイトルであり、発明者としてZeljko DzakulaおよびMathias Ehrichが挙げられ、代理人整理番号SEQ-6034-PV2が指定されている、米国仮特許出願第61/663,477号の利益を請求し、(iii)2011年10月6日に出願され、遺伝子の変動の非侵襲性評価のための方法および処理というタイトルであり、

40

50

発明者としてZeljko DzakulaおよびMathias Ehrichが挙げられ、代理人整理番号SEQ-60 34-PVが指定されている、米国仮特許出願第61/544, 251号の利益を請求する。前述の出願、本文、表および図面を含むすべての内容は、参照によって本明細書に組み込まれる。

【0002】

本明細書において提供する技術は一部分、遺伝子の変動を非侵襲性に評価するための方法、処理（process）および装置に関する。

【背景技術】

【0003】

生きている生物（例えば、動物、植物および微生物）ならびに遺伝情報を複製するその他の形態（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）中にコードされる。遺伝情報は連続的なヌクレオチドまたは改変ヌクレオチドであり、これらは化学的なまたは仮定上の核酸の一次構造を示す。ヒトの場合、完全なゲノムは、24本の染色体上に位置する約30, 000個の遺伝子を含む（The Human Genome, T. Strachan, BIOS Scientific Publishers, 1992年を参照されたい）。各遺伝子が特定のタンパク質をコードし、該タンパク質は、生きている細胞内で転写および翻訳を経て発現した後、特定の生化学的機能を果たす。

【0004】

多くの医学的状態が、1つまたは複数の遺伝子の変動により引き起こされる。特定の遺伝子の変動が医学的状態を引き起こし、これらとして、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）が挙げられる（Human Genome Mutations, D. N. CooperおよびM. Krawczak, BIOS Publishers, 1993年）。そのような遺伝性疾患は、特定の遺伝子のDNA中の単一ヌクレオチドの付加、置換または欠失の結果生じ得る。例えば、特定の出生時欠損が、異数性とも呼ばれる染色体異常、例として、21トリソミー（ダウン症候群）、13トリソミー（パトー症候群）、18トリソミー（エドワーズ症候群）、Xモノソミー（ターナー症候群）、および特定の性染色体異数性、例として、クラインフェルター症候群（XXY）により引き起こされる。別の遺伝子の変動が胎児の性別であり、これはしばしば、性染色体のXおよびYに基づいて決定され得る。いくつかの遺伝子の変動により、例えば、糖尿病、動脈硬化、肥満、種々の自己免疫疾患およびがん（例えば、結腸直腸がん、乳がん、卵巣がん、肺がん）等のいくつかの疾患のうちのいずれかに、個体が、罹患しやすくなる恐れ、またはそうした疾患を発症する恐れがある。

【発明の概要】

【発明が解決しようとする課題】

【0005】

1つまたは複数の遺伝子の変動または分散の同定が、特定の医学的状態の診断またはそうした状態に対する素因の決定につながり得る。遺伝子の分散の同定は、医学的決定の促進および／または有用な医学的手順の利用をもたらすことができる。特定の実施形態では、1つまたは複数の遺伝子の変動または分散の同定が、無細胞DNAの分析を含む。無細胞DNA（CF-DNA）は、細胞死から生じ、末梢血中を循環するDNA断片から構成される。高い濃度のCF-DNAは、特定の臨床状態、例として、がん、外傷、熱傷、心筋梗塞、脳卒中、敗血症、感染およびその他の疾病の指標となり得る。さらに、無細胞胎児DNA（CFF-DNA）を、母体の血流中で検出し、種々の非侵襲性の出生前診断法のために使用することもできる。

【課題を解決するための手段】

【0006】

本明細書において、特定の態様では、胎児中の染色体の異数性、微小重複または微小欠失の存在または非存在を決定して、より少ない擬陰性およびより少ない擬陽性の確定（d

10

20

30

40

50

termination) をもたらす方法であって、(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得るステップであって、これらの配列の読取りが、妊娠中の雌から得られた循環無細胞核酸の読取りであるステップと、(b) それぞれの部分に対してマッピングしたカウントを正規化し、それにより、計算されたゲノム区分のレベルを得るステップと、(c) 計算されたゲノム区分のレベルに従って、ゲノムのセグメントについてのプロファイルを生成するステップと、(d) このプロファイルをセグメント化し、それにより、2つまたはそれ超の分解レンダリングを得るステップと、(e) これらの2つまたはそれ超の分解レンダリングに従って、胎児中の染色体の異数性、微小重複または微小欠失の存在または非存在を決定して、より少ない擬陰性およびより少ない擬陽性の確定をもたらすステップとを含む、方法を提供する。

10

【0007】

また、本明細書において、特定の態様では、ウェーブレット事象の存在または非存在を決定して、より少ない擬陰性およびより少ない擬陽性の確定をもたらすための方法であって、(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得るステップであって、これらの配列の読取りが、妊娠中の雌から得られた循環無細胞核酸の読取りであるステップと、(b) これらの部分のそれぞれに対してマッピングしたカウントを正規化し、それにより、計算されたゲノム区分のレベルを得るステップと、(c) 一連の部分を複数のサブセットの部分にセグメント化するステップと、(d) 計算したゲノム区分のレベルに従って、これらのサブセットのそれぞれについてのレベルを決定するステップと、(e) これらのレベルのそれぞれについて有意性のレベルを決定するステップと、(f) これらのレベルのそれぞれについて決定した有意性のレベルに従って、ウェーブレット事象の存在または非存在を決定して、より少ない擬陰性およびより少ない擬陽性の確定をもたらすステップとを含む、方法も提供する。

20

【0008】

また、本明細書において、特定の態様では、胎児中の染色体の異数性、微小重複または微小欠失の存在または非存在を決定して、より少ない擬陰性およびより少ない擬陽性の確定をもたらす方法であって、(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得るステップであって、これらの配列の読取りが、妊娠中の雌から得られた循環無細胞核酸の読取りであるステップと、(b) それぞれの部分に対してマッピングしたカウントを正規化し、それにより、計算されたゲノム区分のレベルを得るステップと、(c) ゲノムのセグメントを選択し、それにより、一連の部分を得るステップと、(d) 一連の部分を再帰的に区分化し、それにより、2つまたはそれ超のサブセットの部分を得るステップと、(e) それらの2つまたはそれ超のサブセットの部分のそれぞれについてのレベルを決定するステップと、(f) (e) において決定したレベルに従って、試料について、胎児中の染色体の異数性、微小重複または微小欠失の存在または非存在を決定して、より少ない擬陰性およびより少ない擬陽性の確定をもたらすステップとを含む、方法も提供する。

30

【0009】

また、本明細書では、1つまたは複数のプロセッサおよびメモリを含むシステムであって、このメモリが、1つまたは複数のプロセッサによる実行可能命令を含み、参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを含み、これらの配列の読取りが、妊娠中の雌から得られた循環無細胞核酸の読取りであり、これらの1つまたは複数のプロセッサによる実行可能命令が、(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得(これらの配列の読取りは、妊娠中の雌から得られた循環無細胞核酸の読取りである)、(b) それぞれの部分に対してマッピングしたカウントを正規化し、それにより、ゲノム区分のレベルの計算を行い、(c) 計算したゲノム区分のレベルに従って、ゲノムのセグメントについてのプロファイルを生成し、(d) このプロファイルをセグメント化し、それにより、2つまたはそれ超の分解レンダリングを得、(e) これらの2つまたはそれ超の分解レンダリングに従って、胎児中の染色体の異数性、微小重複または微小欠失の存在または非存在を決定して、より少ない擬陰性およびより少

40

50

ない擬陽性の確定をもたらすように構成される、システムを提供する。

本発明は、例えば、以下の項目を提供する。

(項目 1)

胎児中の染色体の異数性、微小重複または微小欠失の存在または非存在を決定するための方法であって、

(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを正規化し、それにより、正規化されたカウントを提供するステップであって、これらの配列の読取りが、胎児を出産する妊娠中の雌に由来する試験試料から得られた循環無細胞核酸の読取りであるステップと、

(b) 該部分の該正規化されたカウントまたは該部分のサブセット中の該正規化されたカウントをセグメント化し、それにより、1つまたは複数の個別のセグメントを提供するステップと、

(c) 該1つまたは複数の個別のセグメントから候補セグメントを同定するステップと、

(d) 該候補セグメントにより、染色体の異数性、微小重複または微小欠失の存在または非存在を決定するステップと

を含む方法。

(項目 2)

前記セグメント化が閾化を含む、項目 1 に記載の方法。

(項目 3)

前記セグメント化がレベル化を含む、項目 1 または 2 に記載の方法。

(項目 4)

前記レベル化が、胎児フラクション、カバレッジ、最小セグメント長さ、またはそれらの組合せに従って行われる、項目 3 に記載の方法。

(項目 5)

閾化およびレベル化が行われ、該閾化が、該レベル化の前に行われる、項目 1 から 4 のいずれか一項に記載の方法。

(項目 6)

(b) における前記セグメント化が、前記部分の前記正規化されたカウントにおいて行われる、項目 1 から 5 のいずれか一項に記載の方法。

(項目 7)

(b) における前記セグメント化が、前記部分のサブセット中の前記正規化されたカウントにおいて行われる、項目 1 から 5 のいずれか一項に記載の方法。

(項目 8)

前記部分の前記サブセットが、染色体の全部分または染色体の全部分のサブセットである、項目 7 に記載の方法。

(項目 9)

前記正規化されたカウントが、レベルを有するプロファイル中にあり、該プロファイルが、(b) においてセグメント化される、項目 1 から 8 のいずれか一項に記載の方法。

(項目 10)

前記セグメント化が、前記個別のセグメントを含む分解レンダリングを生成する、請求項 1 から 9 のいずれか一項に記載の方法。

(項目 11)

(a) における前記正規化が、グアニンおよびシトシン (GC) の偏りの LOESS 正規化 (GC-LOESS 正規化) を含む、項目 1 から 10 のいずれか一項に記載の方法。

(項目 12)

(a) における前記正規化が、主成分正規化を含む、項目 1 から 11 のいずれか一項に記載の方法。

(項目 13)

(a) における前記正規化が、GC-LOESS 正規化、およびそれに続く、主成分正

10

20

30

40

50

規化を含む、項目 1 から 12 のいずれか一項に記載の方法。

(項目 14)

(a) における前記正規化が、

(1) (i) 前記部分のそれぞれに対してマッピングした前記配列の読取りの前記カウントと、(i i) 該部分のそれぞれについての GC 含有量との間で適合させた関係式に基づいて、前記試験試料についてのグアニンおよびシトシン (GC) の偏り係数を決定するステップであって、該 GC の偏り係数が、線形適合関係式の場合の勾配、または非線形適合関係式の場合の曲率の推定値であるステップと、

(2) マイクロプロセッサを使用して、(a) の該カウント、(b) の該 GC の偏り係数、および該部分のそれぞれについて、(i) 複数の試料のそれぞれについての該 GC の偏り係数と、(i i) 該複数の試料についての、該部分のそれぞれに対してマッピングした該配列の読取りの該カウントとの間で適合させた関係式に基づいて、該部分のそれぞれについてのゲノム区分のレベルを計算し、それにより、計算されたゲノム区分のレベルを提供するステップと

を含む、項目 1 から 13 のいずれか一項に記載の方法。

(項目 15)

(b) における前記セグメント化が、2 つまたはそれ超の異なるセグメント化処理の適用を含む、項目 1 から 14 のいずれか一項に記載の方法。

(項目 16)

前記 2 つまたはそれ超の異なるセグメント化処理のそれぞれが、ハールウェーブレットセグメンテーション、サーキュラーバイナリセグメンテーション、最大エントロピーセグメンテーション、エッジ検出カーネルを用いるコンボリューション、ジェンセンシャノンダイバージェンス、バイナリ再帰的セグメンテーション、およびフーリエ変換から独立に選択される、項目 15 に記載の方法。

(項目 17)

前記 2 つまたはそれ超の異なるセグメント化処理のうちの 1 つが、サーキュラーバイナリセグメンテーションである、項目 15 または 16 に記載の方法。

(項目 18)

前記 2 つまたはそれ超の異なるセグメント化処理のうちの 1 つが、ハールウェーブレットセグメンテーションである、項目 15 から 17 のいずれか一項に記載の方法。

(項目 19)

(b) における前記セグメント化が、ハールウェーブレットセグメント化処理およびサーキュラーバイナリセグメント化処理を含む、項目 15 から 18 のいずれか一項に記載の方法。

(項目 20)

前記 2 つまたはそれ超のセグメント化処理が、平行して行われる、項目 15 から 19 のいずれか一項に記載の方法。

(項目 21)

前記セグメント化が、分解レンダリング中の隣接する断片化されたレベルを統合することを含む仕上げ処理を含む、項目 1 から 20 のいずれか一項に記載の方法。

(項目 22)

前記候補セグメントの 1 つまたは複数のエッジを決定することを含む、項目 1 から 21 のいずれか一項に記載の方法。

(項目 23)

前記候補セグメントがカバーする部分の数を決定することを含む、項目 1 から 22 のいずれか一項に記載の方法。

(項目 24)

前記候補セグメントのレベルを決定することを含む、項目 1 から 23 のいずれか一項に記載の方法。

(項目 25)

10

20

30

40

50

前記候補セグメントが、曲線下面積（AUC）の分析により同定される、項目1から24のいずれか一項に記載の方法。

（項目26）

前記AUC分析が、前記候補セグメントがカバーする部分の数、および／または該候補セグメントについての前記レベルに関する、項目25に記載の方法。

（項目27）

前記候補セグメントを検証し、それにより、検証された候補セグメントを提供することを含む、項目1から26のいずれか一項に記載の方法。

（項目28）

前記検証することが、スライディングエッジ処理を行うことを含む、項目27に記載の方法。

（項目29）

前記検証することが、リーブワンアウト処理を行うことを含む、項目27または28に記載の方法。

（項目30）

前記検証することが、前記スライディングエッジ処理および前記リーブワンアウト処理を行うことを含む、項目29に記載の方法。

（項目31）

前記検証することが、前記候補セグメントについての有意性のレベルを生成することを含む、項目27から30のいずれか一項に記載の方法。

（項目32）

前記検証することが、複合候補セグメントについての有意性のレベルを生成することを含む、項目27から31のいずれか一項に記載の方法。

（項目33）

第1のセグメンテーションから第1の候補セグメントを同定し、該第1のセグメンテーションとは異なる第2のセグメンテーションから第2の候補セグメントを同定することを含む、項目27から32のいずれか一項に記載の方法。

（項目34）

前記第1の候補セグメントと前記第2の候補セグメントとが実質的に同じであるまたは実質的に異なるかどうかを決定することを含む、項目33に記載の方法。

（項目35）

前記第1の候補セグメントと前記第2の候補セグメントとが実質的に異なる場合に、微小欠失または微小重複が存在しないと決定することを含む、項目33または34に記載の方法。

（項目36）

前記候補セグメントまたは前記検証された候補セグメントの定量を行うことを含む、請求項1から35のいずれか一項に記載の方法。

（項目37）

前記定量が、前記候補セグメントまたは前記検証された候補セグメントについてのカウント表示である、項目36に記載の方法。

（項目38）

前記定量が、前記候補セグメントまたは前記検証された候補セグメントについての前記カウント表示のzスコアによる定量である、項目37に記載の方法。

（項目39）

前記zスコアが、前記候補セグメントまたは前記検証された候補セグメントについて、（i）試験試料カウント表示から、（ii）正倍数体カウント表示の中央値を減算して得た差を、（iii）該正倍数体カウント表示のMADで除算して得た減算の結果であり、ここで、該（i）試験試料カウント表示が、該試験試料について、全カウントを、全常染色体カウントで除算して得た比であり、該（ii）正倍数体カウント表示の中央値が、正倍数体試料について、全カウントを、全常染色体カウントで除算して得た比の中央値であ

10

20

30

40

50

る、項目 38 に記載の方法。

(項目 40)

前記候補セグメントまたは前記検証された候補セグメントが位置する染色体の該染色体表示の定量を行うことを含む、項目 36 から 39 のいずれか一項に記載の方法。

(項目 41)

前記染色体表示の前記定量が、 z スコアによる定量である、項目 40 に記載の方法。

(項目 42)

前記 z スコアが、前記染色体について、(i) 試験試料カウント表示から、(ii) 正倍数体カウント表示の中央値を減算して得た差を、(iii) 該正倍数体カウント表示の MAD で除算して得た減算の結果であり、ここで、該 (i) 試験試料カウント表示が、該試験試料について、前記候補セグメントが位置する該染色体中の全カウントを、全常染色体カウントで除算して得た比であり、該正倍数体カウント表示の該 (ii) 中央値が、正倍数体試料について、該候補セグメントが位置する該染色体中の全カウントを、全常染色体カウントで除算して得た比の中央値である、項目 41 に記載の方法。

(項目 43)

前記候補セグメントまたは前記検証された候補セグメントの前記定量が、前記染色体表示の前記定量と比較される、項目 36 から 42 のいずれか一項に記載の方法。

(項目 44)

第 1 の候補セグメントまたは第 1 の前記検証された候補セグメントの z スコアによる定量が行われ、第 2 の候補セグメントまたは第 2 の前記検証された候補セグメントの z スコアによる定量が行われ、該第 1 の候補セグメントおよび該第 2 の候補セグメントが、2 つの異なるタイプのセグメンテーションから同定される、項目 43 に記載の方法。

(項目 45)

(i) 1 未満の係数で乗算した、前記第 1 の候補セグメントまたは前記検証された第 1 の候補セグメントの前記 z スコアによる定量、および (ii) 該係数で乗算した、前記第 2 の候補セグメントまたは前記検証された第 2 の候補セグメントの該 z スコアによる定量の最小値を決定することを含む、項目 44 に記載の方法。

(項目 46)

前記染色体表示の前記 z スコアによる定量結果が、前記最小値未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、項目 45 に記載の方法。

(項目 47)

前記染色体表示の前記 z スコアによる定量結果が、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、項目 44 に記載の方法。

(項目 48)

前記試験試料について、(i) 前記染色体表示の前記 z スコアによる定量結果が、値 3.95 超またはそれに等しく、(ii) 該染色体表示の該 z スコアによる定量結果が、前記最小値超またはそれに等しい場合に、染色体異数性が存在すると決定することを含む、項目 47 に記載の方法。

(項目 49)

前記試験試料について、(i) 前記染色体表示の前記 z スコアによる定量結果が、値 3.95 未満であり、かつ／または (ii) 該染色体表示の該 z スコアによる定量結果が、前記最小値未満である場合に、染色体異数性が存在しないと決定することを含む、項目 47 に記載の方法。

(項目 50)

前記染色体異数性が、トリソミーまたはモノソミーである、項目 48 または 49 に記載の方法。

(項目 51)

前記第 1 の候補セグメントまたは前記検証された第 1 の候補セグメントの前記 z スコアによる定量結果が、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定し、前記第 2 の候補セグメントまたは前記検証された第 2 の候補セグメントの前記 z ス

10

20

30

40

50

コアによる定量結果が、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定することを含み、項目 49 に記載の方法。

(項目 52)

前記第 1 の候補セグメントおよび前記第 2 の候補セグメント、または検証されたそれらのセグメントが、実質的に同じであるかどうかを決定することを含み、項目 51 に記載の方法。

(項目 53)

前記試験試料について、(i) 前記第 1 の候補セグメントまたは前記検証された第 1 の候補セグメントの前記 z スコアによる定量結果が、値 3.95 超またはそれに等しく、前記第 2 の候補セグメントまたは前記検証された第 2 の候補セグメントの前記 z スコアによる定量結果が、値 3.95 超またはそれに等しく、(ii) 該第 1 の候補セグメントおよび該第 2 の候補セグメント、または検証されたそれらのセグメントが、実質的に同じである場合に、微小欠失または微小挿入が存在すると決定することを含み、項目 52 に記載の方法。

(項目 54)

前記試験試料について、(i) 前記第 1 の候補セグメントもしくは前記検証された第 1 の候補セグメントの前記 z スコアによる定量結果が、値 3.95 未満であり、かつ／または前記第 2 の候補セグメントもしくは前記検証された第 2 の候補セグメントの前記 z スコアによる定量結果が、値 3.95 未満であり、かつ／または(ii) 該第 1 の候補セグメントおよび該第 2 の候補セグメント、もしくは検証されたそれらのセグメントが、実質的に同じではない場合に、微小欠失または微小挿入が存在しないと決定することを含み、請求項 52 に記載の方法。

(項目 55)

前記候補セグメントまたは検証された候補セグメントについての前記カウント表示の z スコアによる定量結果を決定し、それが、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定することを含み、項目 36 から 42 のいずれか一項に記載の方法。

(項目 56)

前記染色体表示の z スコアによる定量結果を決定し、それが、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定することを含み、項目 36 から 42 のいずれか一項に記載の方法。

(項目 57)

対数オッズ比 (LOR) を計算することを含み、LOR が、(i) (1) 遺伝子の変動を有する条件付き確率と (2) 該遺伝子の変動を有する事前確率との第 1 の乗算の積と、(ii) (1) 該遺伝子の変動を有さない条件付き確率と (2) 該遺伝子の変動を有さない事前確率との第 2 の乗算の積との商の対数である、項目 55 および／または 56 に記載の方法。

(項目 58)

前記遺伝子の変動を有する前記条件付き確率が、前記試験試料について決定された胎児フラクション、該試験試料について決定された前記セグメントについての前記カウント表示の z スコア、および該セグメントについての該カウント表示についての z スコアの、該胎児フラクションに関する分布に従って決定される、項目 57 に記載の方法。

(項目 59)

前記遺伝子の変動を有する前記条件付き確率が、等式 23 :

【化 38】

$$Z \sim \text{正規}\left(\frac{\mu_x f}{\sigma_x^2}, 1\right) \quad (23)$$

10

20

30

40

[式中、 f は胎児フラクションであり、 X は、該遺伝子の変動をカバーする前記セグメントについての部分のカウントの合計 $X \sim f(\mu X, \sigma X)$ であり、ここで、 μX および σX はそれぞれ、 X の平均値および標準偏差であり、 $f(\cdot)$ は分布関数である。]

に示す関係により決定される、項目58に記載の方法。

(項目60)

前記遺伝子の変動を有する前記条件付き確率が、前記セグメントについての前記カウント表示の、前記試験試料に関する前記 z スコアと、該セグメントについての該カウント表示の z スコアの、前記胎児フラクションに関する分布との交差部分である、項目58または59に記載の方法。

(項目61)

前記遺伝子の変動を有さない前記条件付き確率が、前記試験試料について決定された前記セグメントについての前記カウント表示の前記 z スコアと、正倍数体中の該セグメントについての該カウント表示の z スコアの分布との交差部分である、項目57に記載の方法。

(項目62)

前記遺伝子の変動を有する前記事前確率および該遺伝子の変動を有さない該事前確率が、前記試験被験体を含まない複数の試料から決定される、項目57から61のいずれか一項に記載の方法。

(項目63)

前記LORが、ゼロ超またはゼロ未満のいずれであるかを決定することを含む、項目57から62のいずれか一項に記載の方法。

(項目64)

前記試験試料について、(i)前記染色体表示の前記 z スコアによる定量結果が、値3.95超またはそれに等しく、(ii)前記LORが、ゼロ超である場合に、染色体異数性が存在すると決定することを含む、項目55から63のいずれか一項に記載の方法。

(項目65)

前記試験試料について、(i)前記染色体表示の前記 z スコアによる定量結果が、値3.95未満であり、かつ／または(ii)前記LORが、ゼロ未満である場合に、染色体異数性が存在しないと決定することを含む、項目55から63のいずれか一項に記載の方法。

(項目66)

前記染色体異数性が、トリソミーまたはモノソミーである、項目64または65に記載の方法。

(項目67)

前記試験試料について、(i)前記候補セグメントまたは検証された候補セグメントについての前記カウント表示の前記 z スコアによる定量結果が、値3.95超またはそれに等しく、(ii)前記LORが、ゼロ超である場合に、微小欠失または微小重複が存在すると決定することを含む、項目55から63のいずれか一項に記載の方法。

(項目68)

前記試験試料について、(i)前記候補セグメントまたは検証された候補セグメントについての前記カウント表示の前記 z スコアによる定量結果が、値3.95未満であり、かつ／または(ii)前記LORが、ゼロ未満である場合に、微小欠失または微小重複が存在しないと決定することを含む、項目55から63のいずれか一項に記載の方法。

(項目69)

前記微小欠失が、ディジョージ症候群と関連する、項目67または68に記載の方法。

(項目70)

前記カウント表示が、正規化されたカウント表示である、項目1から69のいずれか一項に記載の方法。

(項目71)

10

20

30

40

50

(a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、システム中のマイクロプロセッサにより行われる、項目1から70のいずれか一項に記載の方法。

(項目72)

(a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、コンピュータにより行われる、項目1から71のいずれか一項に記載の方法。

(項目73)

(a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、メモリと併せて行われる、項目1から72のいずれか一項に記載の方法。

(項目74)

(a)に先立って、前記妊娠中の雌から得られた試料中の核酸の配列決定を行い、それにより、該核酸配列の読取りを提供することを含む、項目1から73のいずれか一項に記載の方法。

(項目75)

(a)に先立って、前記核酸配列の読取りを、前記参照ゲノムの前記部分に対してマッピングすることを含む、項目1から74のいずれか一項に記載の方法。

(項目76)

胎児中の染色体異数性の存在または非存在を決定するための方法であって、

(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントにより、染色体カウント表示を決定するステップであって、これらの配列の読取りが、胎児を出産する妊娠中の雌に由来する試験試料についての循環無細胞核酸の読取りであるステップと、

(b) 該試験試料について、胎児フラクションを決定するステップと、

(c) 対数オッズ比(LOR)を計算するステップであって、LORが、(i)(1)染色体異数性を有する条件付き確率と(2)該染色体異数性を有する事前確率との第1の乗算の積と、(ii)(1)該染色体異数性を有さない条件付き確率と(2)該染色体異数性を有さない事前確率との第2の乗算の積との商の対数であり、ここで、該染色体異数性を有する該条件付き確率が、(b)の該胎児フラクションおよび(a)の該カウント表示に従って決定されるステップと、

(d) 該LORおよび該染色体カウント表示により、染色体異数性の存在または非存在を同定するステップとを含む方法。

(項目77)

前記染色体カウント表示が、前記染色体中の全部分についての前記カウントを、常染色体中の全部分についての該カウントで除算した結果である、項目76に記載の方法。

(項目78)

前記染色体カウント表示のzスコアによる定量を行うことを含む、項目76または77に記載の方法。

(項目79)

前記zスコアが、(i)試験試料染色体カウント表示から、(ii)正倍数体カウント表示の中央値を減算して得た差を、(iii)該正倍数体カウント表示のMADで除算して得た減算の結果であり、ここで、該(i)試験試料染色体カウント表示が、該染色体中の部分中のカウントを、前記常染色体中の部分中のカウントで除算して得た比であり、該(ii)該正倍数体カウント表示の中央値が、正倍数体について、該染色体中の部分中のカウントを、常染色体中の部分中のカウントで除算して得た比の中央値である、項目78に記載の方法。

(項目80)

前記遺伝子の変動を有する前記条件付き確率が、(b)の前記試験試料について決定された胎児フラクション、(a)の該試験試料についての前記染色体カウント表示のzスコア、および該染色体カウント表示のzスコアの胎児フラクションに特異的な分布に従って決定される、項目76から79のいずれか一項に記載の方法。

10

20

30

40

50

(項目 8 1)

前記遺伝子の変動を有する前記条件付き確率が、等式 2 3：

【化 3 9】

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

[式中、f は胎児フラクションであり、X は、前記染色体についての部分の合計 $X \sim f(\mu_X, \sigma_X)$ 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、 $f(\cdot)$ は分布関数である。]

に示す関係により決定される、項目 8 0 に記載の方法。

(項目 8 2)

前記遺伝子の変動を有する前記条件付き確率が、(a) の前記試験試料染色体カウント表示についての前記 z スコアと、該染色体カウント表示の z スコアの胎児フラクションに特異的な分布との交差部分である、項目 8 0 または 8 1 に記載の方法。

(項目 8 3)

前記染色体異数性を有さない前記条件付き確率が、(a) の前記染色体カウント表示、および正倍数体についてのカウント表示に従って決定される、項目 7 6 から 8 2 のいずれか一項に記載の方法。

(項目 8 4)

前記遺伝子の変動を有さない前記条件付き確率が、前記染色体カウント表示の前記 z スコアと、正倍数体中の前記染色体カウント表示の z スコアの分布との交差部分である、請求項 8 3 に記載の方法。

(項目 8 5)

前記遺伝子の変動を有する前記事前確率および該遺伝子の変動を有さない該事前確率が、前記試験被験体を含まない複数の試料から決定される、項目 7 6 から 8 4 のいずれか一項に記載の方法。

(項目 8 6)

前記 L O R が、ゼロ超またはゼロ未満のいずれであるかを決定することを含む、項目 7 6 から 8 5 のいずれか一項に記載の方法。

(項目 8 7)

参照ゲノムの部分に対してマッピングした核酸配列の読取りの前記カウントが、正規化されたカウントである、項目 7 6 から 8 6 のいずれか一項に記載の方法。

(項目 8 8)

前記カウントが、G C - L O E S S 正規化を含む正規化により正規化される、項目 8 7 に記載の方法。

(項目 8 9)

前記カウントが、主成分正規化を含む正規化により正規化される、項目 8 7 または 8 8 に記載の方法。

(項目 9 0)

カウントが、G C - L O E S S 正規化を含む正規化、およびそれに続く、主成分正規化により正規化される、項目 8 7 から 8 9 のいずれか一項に記載の方法。

(項目 9 1)

前記カウントが、

(1) (i) 前記部分のそれぞれに対してマッピングした前記配列の読取りの前記カウントと、(i i) 該部分のそれぞれについての G C 含有量との間で適合させた関係式に基づいて、前記試験試料についてのグアニンおよびシトシン (G C) の偏り係数を決定するステップであって、該 G C の偏り係数が、線形適合関係式の場合の勾配、または非線形適合関係式の場合の曲率の推定値であるステップと、

10

20

30

40

50

(2) マイクロプロセッサを使用して、(a)の該カウント、(b)の該GCの偏り係数、および該部分のそれぞれについて、(i)複数の試料のそれぞれについての該GCの偏り係数と、(ii)該複数の試料についての、該部分のそれぞれに対してマッピングした該配列の読取りの該カウントとの間で適合させた関係式に基づいて、該部分のそれぞれについてのゲノム区分のレベルを計算し、それにより、計算されたゲノム区分のレベルを提供するステップと

を含む正規化により正規化される、項目87から90のいずれか一項に記載の方法。

(項目92)

前記染色体カウント表示のzスコアによる定量結果を決定し、それが、値3.95未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、項目76から91のいずれか一項に記載の方法。

10

(項目93)

前記試験試料について、(i)前記染色体カウント表示の前記zスコアによる定量結果が、値3.95超またはそれに等しく、(ii)前記LORが、ゼロ超である場合に、染色体異数性が存在すると決定することを含む、項目92に記載の方法。

(項目94)

前記試験試料について、(i)前記染色体表示の前記zスコアによる定量結果が、値3.95未満であり、かつ/または(ii)前記LORが、ゼロ未満である場合に、染色体異数性が存在しないと決定することを含む、項目92に記載の方法。

20

(項目95)

前記染色体異数性が、トリソミーまたはモノソミーである、項目93または94に記載の方法。

(項目96)

前記カウント表示が、正規化されたカウント表示である、項目76から95のいずれか一項に記載の方法。

(項目97)

(a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、システム中のマイクロプロセッサにより行われる、項目76から96のいずれか一項に記載の方法。

(項目98)

30

(a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、コンピュータにより行われる、項目76から97のいずれか一項に記載の方法。

(項目99)

(a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、メモリと併せて行われる、項目76から98のいずれか一項に記載の方法。

(項目100)

(a)に先立って、前記妊娠中の雌から得られた試料中の核酸の配列決定を行い、それにより、該核酸配列の読取りを提供することを含む、項目76から99のいずれか一項に記載の方法。

(項目101)

40

(a)に先立って、前記核酸配列の読取りを、前記参照ゲノムの前記部分に対してマッピングすることを含む、項目76から100のいずれか一項に記載の方法。

【0010】

この技術の特定の態様を、以下の説明、実施例、特許請求の範囲および図面においてさらに記載する。

【0011】

図面は、本技術の実施形態を説明し、制限するものではない。説明を明確にし、また分かりやすくするために、図面は一定の縮尺で作成されておらず、一部の事例では、様々な態様が、特定の実施形態を理解しやすくするために、誇張または拡大して示される場合もある。

50

【図面の簡単な説明】

【0012】

【図1】図1は、ウェーブレット法の概略図を示す。正規化された部分のカウントデータ（右上パネル）がウェーブレット変換されて、ウェーブレット平滑化プロファイル（右下）を生成する。不均一なイベントが、ウェーブレットノイズ除去後に明確に視認される。

【0013】

【図2】図2は、閾化せずにレベル化したときの効果を示す。所望のイベントサイズ毎に最適なレベルを決めることができる。

【0014】

【図3】図3は、第13染色体に関する試料について、不均一なプロファイル（上段）を含むプロファイルおよびウェーブレット変換されたプロファイル（中段）を示す。下段のパネルは、第13染色体に関する複数の正倍数体参照試料から得られたヌルエッジ高さ分布を示す。中段のパネルでは、2つの大きな差異（円内）は、不均一なイベントの境界に対応する。

10

【0015】

【図4】図4は、ウェーブレットまたはCBS後のセグメントの統合に関する例を示す。当初区切られた3つのセグメント（左側パネル、染色体の右半分）は、単一の長いストレッチ（右側パネル、染色体の右半分）に統合され、微小重複を明確に視認できるようにした。

【0016】

20

【図5】図5A～5Eは、ウェーブレット平滑化後（図5B）、CBS平滑化後（図5C）、およびセグメント統合後（図5Dおよび図5E）の染色体プロファイルを示す。2つの方法に由来する2つの最良セグメントを比較し、そして相互を「交差確認」する。

【0017】

【図6A】図6Aおよび図6Bは、決定分析の非限定的な例を示す。フロー図に示すいくつかの要素（例えば、ボックス）は、任意選択である。一部の実施形態では、追加の要素が付加される（例えば、検証）。

【図6B】図6Aおよび図6Bは、決定分析の非限定的な例を示す。フロー図に示すいくつかの要素（例えば、ボックス）は、任意選択である。一部の実施形態では、追加の要素が付加される（例えば、検証）。

30

【0018】

【図7】図7は、650から拡張された比較の非限定的な例を示す。

【0019】

【図8】図8は、631および632により表された2つのウェーブレット事象の比較の非限定的な例を示す。

【0020】

【図9】図9は、ウェーブレット平滑化後（A）、および統合後（B）、およびCBS平滑化および統合後（C）の染色体プロファイルを示す。比較後、2つの方法に由来する2つの最良セグメントは、相互を「交差棄却」する。

【0021】

40

【図10-1】図10は、ディジョージ症候群と関連する遺伝子の変動と関係のある第22染色体のセグメントのプロファイルを示す。ディジョージ症候群と関連した遺伝子の微小欠失および微小重複を、この領域に対してマッピングした。左側（パネルA～G）のプロファイルを、ハールウェーブレットおよびCBSによりセグメント化し、平滑化、統合化、および比較した。複合プロファイルを右側パネル（A'～G'）に示す。フローセル毎の試料負荷量の差異を、A-A'、0.5-plex；B-B'、1-plex；C-C'、2-plex；D-D'、3-plex；E-E'、4-plex；F-F'、5-plex、およびG-G'、6-plexの各パネルに示す。試料読取りカバレッジが10分の1に減少しても、ディジョージ微小欠失が検出された（例えば、パネルF'を参照）。

50

【図10-2】図10は、ディジョージ症候群と関連する遺伝子の変動と関係のある第22染色体のセグメントのプロファイルを示す。ディジョージ症候群と関連した遺伝子の微小欠失および微小重複を、この領域に対してマッピングした。左側（パネルA～G）のプロファイルを、ハールウェーブレットおよびCBSによりセグメント化し、平滑化、統合化、および比較した。複合プロファイルを右側パネル（A'～G'）に示す。フローセル毎の試料負荷量の差異を、A-A'、0.5-plex；B-B'、1-plex；C-C'、2-plex；D-D'、3-plex；E-E'、4-plex；F-F'、5-plex、およびG-G'、6-plexの各パネルに示す。試料読取りカバレッジが10分の1に減少しても、ディジョージ微小欠失が検出された（例えば、パネルF'を参照）。

10

【0022】

【図11】図11Aは、第1染色体のプロファイル中における微小欠失の検出を表す、複合ウェーブレット事象を示す。図11Bは、第2染色体のプロファイル中における微小重複の検出を表す、複合ウェーブレット事象を示す。

【0023】

【図12】図12は、最大エントロピー法を利用した、第12染色体中における微小重複場所の検出を表す代表例を示す。

【0024】

【図13-1】図13は、プレート上の試料の場所を示す1対の番号で標識された16の試料について、ディジョージ領域の拡大図を示す。試料対3__4（下から2番目）および9__10（下から5番目）はディジョージ妊娠胎児に属する。全ての他の試料は、正倍数体として核型分類された。強調されたボックス（グレー領域）は、ディジョージ領域とPERUN部分選択の間の重複部について、概要を示す（参照ゲノムchr22__368～chr22__451の部分）。

20

【図13-2】図13は、プレート上の試料の場所を示す1対の番号で標識された16の試料について、ディジョージ領域の拡大図を示す。試料対3__4（下から2番目）および9__10（下から5番目）はディジョージ妊娠胎児に属する。全ての他の試料は、正倍数体として核型分類された。強調されたボックス（グレー領域）は、ディジョージ領域とPERUN部分選択の間の重複部について、概要を示す（参照ゲノムchr22__368～chr22__451の部分）。

30

【0025】

【図14】図14は、ディジョージ領域に関するZスコアを示す。各データポイントは、各患者由来の2つの別々の血漿アリコートから得られた2つのプロファイルの合計に由来する。2つの罹患した症例を含む全16患者に基づき、Z-標準化を実施した。

【0026】

【図15】図15～16は、試料3__4（ディジョージ）および1__2（正倍数体）それぞれに関する代表的なヒストグラムを示す。各ヒストグラムは、ディジョージ領域内に含有される15×15グリッドの領域上で得られたZスコアの分布を示す。外側のエッジから開始し、内側に向かって移動しながら、1つの部分ずつ、ディジョージ領域の左側および右側のエッジの両方をスライドすることにより、領域を選択した。試料3__4および9__10のヒストグラム（図示せず）からは、一貫して欠損が認められ、Z=-3を上回る3__4に関するZスコアは、ほんのわずかであった。試料13__14（図示せず）に関するヒストグラムからは、一貫して過剰表示が示唆され、Zスコアが3未満となる領域はほんのわずかであった。他の全試料（例えば、1__2）は、Zスコアの[-3、3]セグメント内に限定されたままであった。

40

【0027】

【図16】図15～16は、試料3__4（ディジョージ）および1__2（正倍数体）それぞれに関する代表的なヒストグラムを示す。各ヒストグラムは、ディジョージ領域内に含有される15×15グリッドの領域上で得られたZスコアの分布を示す。外側のエッジから開始し、内側に向かって移動しながら、1つの部分ずつ、ディジョージ領域の左側およ

50

び右側のエッジの両方をスライドすることにより、領域を選択した。試料 3 __ 4 および 9 __ 1 0 のヒストグラム（図示せず）からは、一貫して欠損が認められ、 $Z = -3$ を上回る 3 __ 4 に関する Z スコアは、ほんのわずかであった。試料 1 3 __ 1 4（図示せず）に関するヒストグラムからは、一貫して過剰表示が示唆され、 Z スコアが 3 未満となる領域はほんのわずかであった。他の全試料（例えば、1 __ 2）は、 Z スコアの $[-3, 3]$ セグメント内に限定されたままであった。

【0028】

【図17】図17は、16個の試料それぞれについて、 Z スコア中央値およびその ± 3 MAD 信頼区間を示す。各 Z スコア中央値を、スライディングエッジから得た 15×15 グリッドの領域（225領域）から決定した。公知のディジョージ試料（3 __ 4 および 9 __ 1 0）に関する Z スコアは、圧倒的多数のディジョージサブ領域について、 -3 未満のままであった。試料 1 3 __ 1 4 では重複が明らかなことが、その Z スコアが大部分について 3 を上回るという事実により確認された。他の全試料の Z スコアは、 $[-3, 3]$ セグメント内に常に限定された。

10

【0029】

【図18】図18～図19は、試料 3 __ 4（ディジョージ）および 1 __ 2（正倍数体）のそれぞれについて、代表的なヒストグラムを示す。各ヒストグラムは、ディジョージ領域について得た Z スコアの分布を示す。「リーブワンアウト」法（“leave one out” method）を使用して、各 Z スコアを、異なる 16 セットの参照試料を使用して計算した。試料 9 __ 1 0（図示せず）に関するヒストグラムより、欠損が確認された。参照セットに依存して、試料 3 __ 4 は明確に欠損していた、または境界線上に Z スコアを有した。試料 1 3 __ 1 4（図示せず）に関するヒストグラムから、過剰表示が示唆され、境界線上のスコアはわずかであった。1 __ 2 を含め他の全試料は、 Z スコアの $[-3, 3]$ セグメント内に限定されたままであった。

20

【0030】

【図19】図18～図19は、試料 3 __ 4（ディジョージ）および 1 __ 2（正倍数体）のそれぞれについて、代表的なヒストグラムを示す。各ヒストグラムは、ディジョージ領域について得た Z スコアの分布を示す。「リーブワンアウト」法（“leave one out” method）を使用して、各 Z スコアを、異なる 16 セットの参照試料を使用して計算した。試料 9 __ 1 0（図示せず）に関するヒストグラムより、欠損が確認された。参照セットに依存して、試料 3 __ 4 は明確に欠損していた、または境界線上に Z スコアを有した。試料 1 3 __ 1 4（図示せず）に関するヒストグラムから、過剰表示が示唆され、境界線上のスコアはわずかであった。1 __ 2 を含め他の全試料は、 Z スコアの $[-3, 3]$ セグメント内に限定されたままであった。

30

【0031】

【図20】図20は、各試料の Z スコア中央値を示し、その代表的なものを図18～19に示す。 Z スコア中央値およびその ± 3 MAD 信頼区間を、「リーブワンアウト」法を使用して決定した異なる 16 セットの参照試料から計算した。公知のディジョージ試料（3 __ 4 および 9 __ 1 0）に関する Z スコアは、参照試料の圧倒的多数のサブセットについて -3 未満のままであった。試料 1 3 __ 1 4 では重複が明らかなことが、その Z スコアが大部分について 3 を上回るという事実により確認された。他の全試料の Z スコアは、 $[-3, 3]$ セグメント内に常に限定された。

40

【0032】

【図21】図21は、16個の試料それぞれについて、 15×15 グリッドのディジョージサブ領域を使用して得た Z スコア中央値（ x 軸）と、「リーブワンアウト」技法により生成した Z スコア中央値（ y 軸）との比較を示す。対角線は理想的な一致を表す（勾配 = 1、切片 = 0）。

【0033】

【図22】図22～23は、試料 3 __ 4（ディジョージ）および 1 __ 2（正倍数体）のそれぞれに関する代表的なヒストグラムを示す。各ヒストグラムは、異なる 16 セットの参

50

照試料を使用して、ディジョージ領域のサブ領域について得たZスコアの分布を示す。サブ領域は、 15×15 グリッドの225個のサブ領域からランダムに選択した。「リーブワンアウト」分析より、試料3__4（図37）および9__10（図示せず）について欠損を確認した。試料13__14に関するヒストグラムより過剰表示を確認した（図示せず）。1__2を含め他の全試料は、Zスコアの $[-3, 3]$ セグメント内に限定されたままであった。

【0034】

【図23】図22～23は、試料3__4（ディジョージ）および1__2（正倍数体）のそれぞれに関する代表的なヒストグラムを示す。各ヒストグラムは、異なる16セットの参照試料を使用して、ディジョージ領域のサブ領域について得たZスコアの分布を示す。サブ領域は、 15×15 グリッドの225個のサブ領域からランダムに選択した。「リーブワンアウト」分析より、試料3__4（図37）および9__10（図示せず）について欠損を確認した。試料13__14に関するヒストグラムより過剰表示を確認した（図示せず）。1__2を含め他の全試料は、Zスコアの $[-3, 3]$ セグメント内に限定されたままであった。

10

【0035】

【図24】図24は、「リーブワンアウト」法を使用した16個の試料それぞれについて、ランダムに選択されたディジョージ領域のサブ領域に関するZスコア中央値およびその ± 3 MAD信頼区間を示す。公知のディジョージ試料（3__4および9__10）に関するZスコアは、ほとんどの参照試料について-3未満のままであった。試料13__14では重複が明らかなことが、そのZスコアが大部分について3を上回るという事実により示された。他の全試料のZスコアは、 $[-3, 3]$ セグメント内に常に限定されたが、ただし試料17__18を除く。

20

【0036】

【図25】図25～26は、試料3__4（ディジョージ）および試料1__2（正倍数体）に関する代表的なヒストグラムをそれぞれ示し、異なる16セットの参照試料を使用して、ディジョージ領域の225個のサブ領域全てについて得たZスコアの分布を表す。各試料について、 15×15 グリッド上でスライディングエッジ法を使用して、225個のサブ領域を生成した。スライディングエッジを「リーブワンアウト」分析と組み合わせた。結果により、罹患試料3__4および9__10（図示せず）の両方について欠損が確認される。試料13__14に関するヒストグラムにより、過剰表示が確認された（図示せず）。1__2を含め他の全試料は、Zスコアの $[-3, 3]$ セグメント内に限定されたままであったが、17__18に散発的な例外が認められた（図示せず）。

30

【0037】

【図26】図25～26は、試料3__4（ディジョージ）および試料1__2（正倍数体）に関する代表的なヒストグラムをそれぞれ示し、異なる16セットの参照試料を使用して、ディジョージ領域の225個のサブ領域全てについて得たZスコアの分布を表す。各試料について、 15×15 グリッド上でスライディングエッジ法を使用して、225個のサブ領域を生成した。スライディングエッジを「リーブワンアウト」分析と組み合わせた。結果により、罹患試料3__4および9__10（図示せず）の両方について欠損が確認される。試料13__14に関するヒストグラムにより、過剰表示が確認された（図示せず）。1__2を含め他の全試料は、Zスコアの $[-3, 3]$ セグメント内に限定されたままであったが、17__18に散発的な例外が認められた（図示せず）。

40

【0038】

【図27】図27は、「リーブワンアウト」技法と組み合わせて、 15×15 グリッドのディジョージサブ領域を使用して得たZスコア中央値と、 15×15 グリッド単独に由来するZスコア中央値との比較を示す。対角線は理想的な一致を表す（勾配=1、切片=0）。

【0039】

【図28】図28は、「リーブワンアウト」技法と組み合わせて、 15×15 グリッドの

50

ディジョージサブ領域を使用して得たZスコアのMADを、 15×15 グリッド単独に由来するZスコアのMADと比較して示す。対角線は理想的な一致を表す（勾配 = 1、切片 = 0）。

【0040】

【図29】図29は、「リーブワンアウト」法と組み合わせて、カノニックディジョージ領域について、その 15×15 グリッドの完全なサブ領域を評価したときのZスコア中央値およびその ± 3 MAD信頼区間を示す。公知のディジョージ試料（3__4および9__10）に関するZスコアは、ほとんどの参照試料について-3未満のままであった。試料13__14では重複が明らかだが、そのZスコアが大部分について3を上回るという事実により示された。他の全試料のZスコアは、 $[-3, 3]$ セグメント内に常に限定されたが、ただし試料17__18を除く。

10

【0041】

【図30】図30は、本技術のある特定の実施形態が実施可能であるシステムの例示的な実施形態を示す。

【0042】

【図31】図31は、LDTV2雄試料に関する、対数オッズ比（LOR）法を使用した分類結果を示す。

【0043】

【図32】図32は、実施例6に記載する等式23のある特定の側面を図解的に表す。

【0044】

20

【図33】図33は、Epanechnikovカーネルにより得たGC密度の実施形態を示す（バンド幅 = 200bp）。

【0045】

【図34】図34は、HTRA1遺伝子に関するGC密度（y軸）のプロットを示し、この場合、GC密度がゲノム全体にわたり正規化されている。ゲノム位置をx軸に示す。

【0046】

【図35】図35は、参照ゲノム（実線）、および試料から得た配列の読取り（破線）について、ローカルなゲノムの偏り（例えば、GC密度、x軸）の推定値の分布を示す。偏り頻度（例えば、密度頻度）をy軸に示す。GC密度の推定値は、ゲノム全体にわたり正規化されている。この例では、試料は、参照から予想されるよりも高いGC含有量の読取りをより多く有する。

30

【0047】

【図36】図36は、重み付けした3次多項式で適合した関係を使用して、参照ゲノムに関するGC密度の推定値、および試料についての配列の読取りのGC密度の推定値の分布の比較を示す。GC密度の推定値（x軸）を、ゲノム全体にわたり正規化した。GC密度頻度を、参照の密度頻度を試料の密度頻度で割った比の対数（ $\log_2$ ）として、y軸に表す。

【0048】

【図37】図37Aは、ゲノムの全部分に関するGC密度中央値（x軸）の分布を示す。図37Bは、複数の試料に関するGC密度分布により決定した絶対偏差中央値（MAD）（x軸）を示す。GC密度頻度をy軸に示す。各部分は、複数の参照試料（例えば、トレーニングセット）に関するGC密度中央値の分布によりフィルタリングし、またMAD値は、複数の試料のGC密度分布により決定した。確定された閾（例えば、MADの四分位数間範囲の4倍）から外れたGC密度を含む部分は、フィルタリング処理により、検討事項から除去した。

40

【0049】

【図38】図38Aは、読取り密度中央値を含むゲノムに関する試料の読取り密度プロファイル（y軸、例えば読取り密度／部分）、およびゲノム内の各ゲノム部分の相対的な位置（x軸、部分インデックス）を示す。図38Bは、第1の主成分（PC1）を示し、図38Cは、500個の正倍数体のトレーニングセットから得た読取り密度プロファイルの

50

主成分分析から得られた第2の主成分（PC2）を示す。

【0050】

【図39】図39A～Cは、第21染色体のトリソミーを含むゲノムに関する試料の読取り密度プロファイルの例を示す（例えば、2本の縦線で挟まれた部分）。各ゲノム部分の相対的な位置をx軸に示す。読取り密度をy軸に提示する。図39Aは、未処理の（例えば、未調整の）読取り密度プロファイルを示す。図39Bは、プロファイル中央値の減算を含む第1の調整を含む39Aのプロファイルを示す。図39Cは、第2の調整を含む39Bのプロファイルを示す。第2の調整は、主成分プロファイルを8倍してこれ減算を含み、この試料中に見出される表示に基づき重み付けされている（例えば、モデルが構築される）。例えば、 $\text{Sample Profile} = A * PC1 + B * PC2 + C * PC3 \dots$ および、例えば39Cに示すような補正後のプロファイル $= \text{Sample Profile} - A * PC1 + B * PC2 + C * PC3 \dots$

10

【0051】

【図40】図40は、T21試験に関するブートストラップされたトレーニング試料に由来する検定p値のQQプロットを示す。QQプロットは、一般的に2つの分布を比較する。図40は、試験試料に由来するChAIスコア（y軸）と均一分布（すなわち、予想されるp値の分布、x軸）との比較を示す。各ポイントは、単一試験試料のlog-pP値スコアを表す。試料を振り分け、そして均一分布に基づき「予想される」値（x軸）を割り当てる。下側の破線は、対角線を表し、また上側の線は、ボンフェローニ閾を表す。均一分布に従う試料は、下側の対角線（下側破線）上に乗るものと予想される。部分内の相関に起因してデータ値が対角線からかなり離れており（例えば、偏り）、予想よりも高スコアリング（低p値）の試料が示される。本明細書に記載する方法（例えばChAI、例えば実施例7を参照）は、この観察された偏りを補正することができる。

20

【0052】

【図41】図41Aはトレーニングセット内の男性および女性に関するPC2係数の差異を表す読取り密度プロットを示す。図41Bは、PC2係数を用いた性別判定に関する受信者動作特性（ROC）プロットを示す。配列決定により行われた性別判定を、真否参照用として使用した。

【0053】

【図42A】図42A～42Bは、システムの実施形態を示す。

30

【図42B】図42A～42Bは、システムの実施形態を示す。

【発明を実施するための形態】

【0054】

本明細書では、胎児中の胎児期遺伝子の変動（例えば、染色体の異数性、微小重複または微小欠失）の存在または非存在を決定するための方法であって、ここでは、一部および／または全部において、核酸配列に従って決定が下される、方法を提供する。一部の実施形態では、核酸配列を、妊娠中の雌から得られた試料（例えば、妊娠中の雌の血液）から得る。また、本明細書では、改善されたデータ操作法、ならびにシステム、装置およびモジュールも提供し、これらは一部の実施形態では、本明細書に記載する方法を実施する。一部の実施形態では、本明細書に記載する方法により遺伝子の変動を同定することによって、特定の医学的状態の診断をもたらすこと、または特定の医学的状態の素因を決定することができる。遺伝子の分散を同定することによって、医学的決定の促進および／または有用な医学的手順の利用をもたらすことができる。

40

試料

【0055】

本明細書では、核酸を分析するための方法および組成を提供する。一部の実施形態では、核酸断片の混合物中の核酸断片を分析する。核酸の混合物は、異なるヌクレオチド配列、異なる断片長、異なる起源（例えば、ゲノム起源、胎児起源対母体起源、細胞起源もしくは組織起源、試料起源、被験体起源等）、またはそれらの組合せを有する2つまたはそれ超の核酸断片種を含むことができる。

50

【0056】

しばしば、本明細書に記載する方法および装置において利用する核酸または核酸混合物を、被験体から得られた試料から単離する。被験体は、これらに限定されないが、ヒト、非ヒト動物、植物、細菌、真菌または原生生物を含めた、任意の生きているまたは生きていない生物であり得る。これらに限定されないが、哺乳動物、爬虫類、トリ、両生類、魚、有蹄動物、反芻動物、ウシ科（例えば、ウシ）、ウマ科（例えば、ウマ）、ヤギ（c a p r i n e）およびヒツジ（o v i n e）（例えば、ヒツジ、ヤギ）、ブタ（s w i n e）（例えば、ブタ）、ラクダ科（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（例えば、クマ）、家禽、イヌ、ネコ、マウス、ラット、魚、イルカ、クジラおよびサメを含めて、任意のヒトまたは非ヒト動物を選択することができる。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、任意の年齢（例えば、胚、胎児、乳児、小児、成人）であり得る。

10

【0057】

核酸を、任意のタイプの適切な生物学的検体または試料（例えば、試験試料）から単離することができる。試料または試験試料は、被験体またはその一部分（p a r t）（例えば、ヒト被験体、妊娠中の雌、胎児）から単離されるまたは得られる任意の検体であり得る。検体の非限定的な例として、被験体から得られた体液または組織が挙げられ、これらには、非限定的に、血液または血液生成物（例えば、血清、血漿等）、臍帯血、絨毛膜絨毛、羊水、脳脊髄液、脊髄液、洗浄した液（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、関節鏡検査洗浄液）、生検試料（例えば、着床前胚生検試料から得られた試料）、腹腔穿刺試料（c e l o c e n t e s i s s a m p l e）、細胞（血液細胞、胎盤細胞、胚もしくは胎児細胞、胎児有核細胞もしくは胎児細胞残余物）またはそれらの一部分（例えば、ミトコンドリア、核、抽出物等）、雌の生殖器系の洗浄物、尿、糞便、痰、唾液、鼻粘液、前立腺液、洗浄液（l a v a g e）、精液、リンパ液、胆汁、涙、汗、乳汁、乳房液等、あるいはそれらの組合せが含まれる。一部の実施形態では、生物学的試料は、被験体から得られた子宮頸部スワブである。一部の実施形態では、生物学的試料は、血液であり得、時には、血漿または血清であり得る。用語「血液」は、本明細書で使用する場合、妊婦または妊娠の可能性について試験されている女性から得られた血液の試料または調製物を指す。この用語は、全血、血液生成物または血液の任意の画分、例として、従来の定義どおりの血清、血漿、パフィーコート等を包含する。血液またはその画分はしばしば、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。ヌクレオソームは、核酸を含み、時には、無細胞または細胞内ヌクレオソームである。血液はまた、パフィーコートも含む。パフィーコートを時には、フィコール勾配を利用することによって単離する。パフィーコートは、白血球細胞（例えば、白血球、T細胞、B細胞、血小板等）を含むことができる。特定の実施形態では、パフィーコートは、母体核酸および／または胎児核酸を含む。血漿は、抗凝固剤で処理した血液の遠心分離の結果得られた、全血の画分を指す。血清は、血液試料が凝固した後に残存する水性の液体部分を指す。体液試料または組織試料をしばしば、病院またはクリニックが一般に従う標準的なプロトコールに従って収集する。血液の場合、末梢血の適切な量（例えば、3～40ミリリットル）をしばしば収集し、調製する前または調製した後に標準的な手順に従って保存することができる。核酸を抽出する体液試料または組織試料は、無細胞の場合がある（例えば、無細胞）。一部の実施形態では、体液試料または組織試料は、細胞要素または細胞残余物を含む場合がある。一部の実施形態では、胎児細胞またはがん細胞を、試料中に含む場合がある。

20

30

40

【0058】

しばしば、試料は不均一であり、これは、1つ超のタイプの核酸種が試料中に存在することを意味する。例えば、不均一核酸として、これらに限定されないが、（i）胎児由来の核酸および母体由来の核酸、（i i）がんの核酸および非がんの核酸、（i i i）病原体の核酸および宿主の核酸、より一般的には、（i v）変異した核酸および野生型の核酸を挙げることができる。試料は、不均一であり得、これは、1つ超の細胞型、例として、

50

胎児細胞および母体細胞、がん細胞および非がん細胞、または病原体細胞および宿主細胞が存在するからである。一部の実施形態では、少量の核酸種および多量の核酸種が存在する。

【0059】

本明細書に記載する技術を出生前に適用する場合、体液試料または組織試料を、試験するのに適切な在胎齢において雌から、または妊娠の可能性について試験されている雌から収集することができる。適切な在胎齢は、実施されている出生前試験に応じて様々であり得る。特定の実施形態では、妊娠中の雌の被験体は、時には妊娠第一期にあり、時には妊娠第二期にあり、または時には妊娠第三期にある。特定の実施形態では、体液または組織を、妊娠中の雌から、在胎約1～約45週（例えば、在胎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）において、時には、在胎約5～約28週（例えば、在胎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）において収集する。特定の実施形態では、体液試料または組織試料を、妊娠中の雌から、出産（例えば、経膈分娩または非経膈分娩（例えば、外科的分娩））の間または直後（例えば、0～72時間後）に収集する。

血液試料の入手およびDNAの抽出

【0060】

本明細書の方法はしばしば、妊娠中および時には妊娠後に、母体および／もしくは胎児の遺伝子の変動の存在または非存在を検出するため、ならびに／または胎児および／もしくは妊娠中の雌の健康状態をモニターするための非侵襲性手段として、母体の血液に見出される胎児のDNAを分離すること、富化することおよび分析することを含む。したがって、本明細書の特定の方法を実行する最初のステップはしばしば、妊婦から血液試料を得ること、および試料からDNAを抽出することを含む。

血液試料の入手

【0061】

血液試料を、本技術による方法を使用して、妊婦から試験するのに適切な在胎齢において得ることができる。適切な在胎齢は、下記に論じるように、試験する障害に応じて変化させることができる。女性からの血液の収集はしばしば、病院またはクリニックが一般に従う標準的なプロトコルに従って実施される。末梢血の適切な量、例えば、典型的には5～50mlをしばしば収集し、さらに調製する前に、標準的な手順に従って保存することができる。血液試料は、試料中に存在する核酸の品質の劣化を最小限に留める様式で、収集し、保存し、または輸送することができる。

血液試料の調製

【0062】

母体の血液に見出される胎児のDNAの分析を、例えば、全血、血清または血漿を使用して行うことができる。母体の血液から血清または血漿を調製する方法が公知である。例えば、妊婦の血液を、EDTAまたはVacutainer SST (Becton Dickinson, Franklin Lakes, N. J.)等の特殊な市販製品を含有するチューブ中に入れて、血液凝固を阻止することができ、次いで、血漿を、全血から遠心分離により得ることができる。血清は、血液凝固後の遠心分離有りまたは無しで得ることができる。遠心分離を使用する場合には、典型的には、適切なスピード、例えば、1,500～3,000回転gで実施するが、必ずしもそうではない。血漿または血清を、DNA抽出のための新しいチューブに移す前に、追加の遠心分離のステップに付してもよい。

【0063】

全血の、無細胞の部分に加えて、また、DNAも、細胞画分から回収し、パフィーコート部分中で富化することができ、このパフィーコート部分は、女性から得られた全血試料を遠心分離し、血漿を除去して得ることができる。

DNAの抽出

【0064】

血液を含めた、生物学的試料からDNAを抽出するための多数の公知の方法がある。DNAの調製の一般的な方法（例えば、SambrookおよびRussell、Molecular Cloning: A Laboratory Manual 3d ed. 2001年による記載）に従うことができ、また、種々の市販されている試薬またはキット、例として、QiagenのQIAamp Circulating Nucleic Acid Kit、QIAamp DNA Mini Kit、またはQIAamp DNA Blood Mini Kit（Qiagen、Hilden、ドイツ）、Genomic Prep（商標）Blood DNA Isolation Kit（Promega、Madison、Wis.）、およびGFX（商標）Genomic Blood DNA Purification Kit（Amersham、Piscataway、N. J.）を使用して、妊婦から得られた血液試料からDNAを得ることもできる。また、これらの方法のうちの1つ超の組合せを使用することもできる。

10

【0065】

一部の実施形態では、最初に、1つまたは複数の方法により、試料を、胎児核酸について富化またはある程度まで富化することもできる。例えば、本技術の組成および処理を、単独で、またはその他の識別因子と組み合わせて使用して、胎児のDNAと母体のDNAとの識別を行うことができる。これらの因子の例として、X染色体とY染色体との間の単一ヌクレオチドの差、Y染色体に特異的な配列、ゲノム中の他の箇所に位置する多型、胎児のDNAと母体のDNAとの間のサイズの差、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が挙げられるが、これらに限定されない。

20

【0066】

試料を核酸の特定の種について富化するためのその他の方法が、2007年5月30日出願のPCT特許出願第PCT/US07/69991号、2007年6月15日出願のPCT特許出願第PCT/US2007/071232号、米国仮出願第60/968,876号および同第60/968,878号（本出願人に譲渡）（2005年11月28日出願のPCT特許出願第PCT/EP05/012707号）に記載されており、これらは全て、参照により本明細書に援用されている。特定の実施形態では、母体核酸を、試料から、選択的に（部分的に、実質的に、ほとんど完全に、または完全に）除去する。

【0067】

用語「核酸」および「核酸分子」を、本開示全体を通して交換可能に使用することができる。これらの用語は、DNA（例えば、相補的DNA（cDNA）、ゲノムDNA（gDNA）等）、RNA（例えば、メッセンジャー（message）RNA（mRNA）、低分子阻害RNA（siRNA）、リボゾームRNA（rRNA）、tRNA、マイクロRNA、胎児または胎盤が高度に発現するRNA等）、ならびに／またはDNAもしくはRNAのアナログ（例えば、塩基のアナログ、糖のアナログおよび／もしくは外来の骨格等を含有するもの）、RNA/DNAのハイブリッドおよびポリアミド核酸（PNA）等に由来する任意の組成の核酸を指し、これらは全て、一本鎖または二本鎖の形態であり得、別段の限定がない限り、天然に存在するヌクレオチドに類似する様式で機能することができる天然ヌクレオチドの公知のアナログを包含することができる。特定の実施形態では、核酸は、プラスミド、ファージ、自律複製性配列（ARS）、セントロメア、人工染色体、染色体、あるいはin vitroで、または宿主細胞、細胞、細胞核もしくは細胞の細胞質中で、複製し得るまたは複製され得るその他の核酸であってもよく、あるいはそれらに由来してもよい。鋳型核酸は、一部の実施形態では、単一の染色体に由来し得る（例えば、核酸試料は、二倍体生物から得られた試料の1つの染色体に由来し得る）。特段の限定がない限り、この用語は、参照核酸に類似する結合特性を有し、天然に存在するヌクレオチドに類似する様式で代謝される天然ヌクレオチドの公知のアナログを含有する核酸を包含する。別段の記載がない限り、特定の核酸配列は、明確に示す配列のみならず、また、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、一塩基多型（SNP）および相補配列も暗に包含する。具体的には、1つま

30

40

50

たは複数の選択された（または全ての）コドンの第3の位置が、混合性塩基の残基および／またはデオキシイノシン残基で置換されている配列を生成することによって、縮重コドン置換を得ることができる。核酸という用語は、座位、遺伝子、cDNA、および遺伝子がコードするmRNAと交換可能に使用する。この用語はまた、均等物として、ヌクレオチドのアナログから合成されたRNAまたはDNAの誘導体、バリエーションおよびアナログ、一本鎖（「センス」鎖または「アンチセンス」鎖、「プラス」鎖または「マイナス」鎖、「フォワード」リーディングフレームまたは「リバース」リーディングフレーム）、および二本鎖ポリヌクレオチドも含むことができる。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、これは、遺伝子産物の転写／翻訳および転写／翻訳の調節に関わる、コード領域に先行する領域およびコード領域に続く領域（リーダーおよびトレーラー）、ならびに個々のコードセグメント（エクソン）間の介在配列（イントロン）を含む。

10

【0068】

デオキシリボヌクレオチドは、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を含む。RNAの場合、塩基シトシンが、ウラシルで置き換えられる。被験体から得られた核酸を鋳型として使用して、鋳型核酸を調製することができる。

核酸の単離および処理

【0069】

核酸を、1つまたは複数の供給源（例えば、細胞、血清、血漿、パフィーコート、リンパ液、皮膚、土壌等）から、当技術分野で公知の方法により得ることができる。任意の適切な方法を使用して、生物学的試料（例えば、血液または血液生成物から）からのDNAを単離する、抽出するおよび／または精製することができ、それらの非限定的な例として、DNAの調製の方法（例えば、SambrookおよびRussell、Molecular Cloning: A Laboratory Manual 3d ed. 2001年による記載）、種々の市販されている試薬またはキット、例として、QiagenのQIAamp Circulating Nucleic Acid Kit、QIAamp DNA Mini Kit、またはQIAamp DNA Blood Mini Kit（Qiagen、Hilden、ドイツ）、Genomic Prep（商標）Blood DNA Isolation Kit（Promega、Madison、Wis.）、およびGFX（商標）Genomic Blood DNA Purification Kit（Amersham、Piscataway、N. J.）等、またはそれらの組合せが挙げられる。

20

30

【0070】

細胞溶解の手順および試薬は、当技術分野で公知であり、一般に、化学的方法（例えば、洗剤、低張溶液、酵素による手順等、もしくはそれらの組合せ）、物理的方法（例えば、フレンチプレス、超音波処理等）、または電解質による溶解方法により行うことができる。任意の適切な溶解手順を利用することができる。例えば、化学的方法は一般に、溶解剤を利用して、細胞を破壊し、細胞から核酸を抽出し、続いて、カオトロピック塩を用いて処理する。物理的方法、例として、凍結／解凍、それに続く、粉碎；細胞プレスの使用等もまた有用である。高い塩濃度による溶解の手順もまた、一般に使用される。例えば、アルカリによる溶解の手順を利用することができる。後者の手順には従来、フェノールクロロホルム溶液の使用を組み入れており、3つの溶液が関与する、代替のフェノールクロロホルムを用いない手順も利用することができる。後者の手順の場合、1つの溶液が、15mMトリス、pH 8.0；10mM EDTA、および100μg/mlリボヌクレアーゼAを含有することができ；第2の溶液が、0.2N NaOHおよび1% SDSを含有することができ；第3の溶液が、3M KOAc、pH 5.5を含有することができる。これらの手順は、Current Protocols in Molecular Biology、John Wiley & Sons、N.Y.、6.3.1～6.3.6（1989年）に見出すことができ、その全体が本明細書に援用されている。

40

50

【0071】

核酸を、別の核酸と比較する場合、異なる時点で単離することができ、試料のそれぞれが、同じ供給源または異なる供給源に由来する。例えば、核酸は、核酸ライブラリー、例として、cDNAライブラリーまたはRNAライブラリーに由来し得る。核酸は、核酸の精製もしくは単離、および／または試料から得られた核酸分子の増幅の結果であり得る。本明細書に記載する処理に提供される核酸は、1つの試料に由来する核酸、あるいは2つまたはそれ超の試料（例えば、1つもしくは複数、2つもしくはそれ超、3つもしくはそれ超、4つもしくはそれ超、5つもしくはそれ超、6つもしくはそれ超、7つもしくはそれ超、8つもしくはそれ超、9つもしくはそれ超、10個もしくはそれ超、11個もしくはそれ超、12個もしくはそれ超、13個もしくはそれ超、14個もしくはそれ超、15個もしくはそれ超、16個もしくはそれ超、17個もしくはそれ超、18個もしくはそれ超、19個もしくはそれ超、または20個もしくはそれ超の試料）に由来する核酸を含有することができる。

10

【0072】

特定の実施形態では、核酸は、細胞外核酸を含むことができる。用語「細胞外核酸」は、本明細書で使用する場合、実質的に細胞を有さない供給源から単離された核酸を指すことができ、また、「無細胞」核酸および／または「無細胞循環」核酸とも呼ぶ。細胞外核酸は、血液（例えば、妊娠中の雌の血液）中に存在し、そこから得ることができる。細胞外核酸はしばしば、検出可能な細胞を含まず、細胞要素または細胞残余物を含有する場合がある。細胞外核酸のための、無細胞の供給源の非限定的な例が、血液、血漿、血清および尿である。本明細書で使用する場合、用語「無細胞循環試料核酸を得る」は、試料を直接得ること（例えば、試料、例えば、試験試料を収集すること）、または試料を収集した他者から試料を得ることを含む。理論により制限されることなく、細胞外核酸は、細胞アポトーシスおよび細胞分解の産物であり得、これらは、スペクトル（例えば、「ラダー」）にわたる一連の長さをしばしば有する細胞外核酸の基を提供する。

20

【0073】

特定の実施形態では、細胞外核酸は、異なる核酸種を含むことができ、したがって、本明細書では、「不均一である」と呼ばれる。例えば、がんを有する人から得られた血清または血漿は、がん細胞に由来する核酸および非がん細胞に由来する核酸を含む場合がある。別の例では、妊娠中の雌から得られた血清または血漿は、母体核酸および胎児核酸を含む場合がある。一部の事例では、胎児核酸は時には、核酸全体の約5%～約50%である（例えば、全ての核酸の約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、胎児核酸である）。一部の実施形態では、核酸中の胎児核酸の大半の長さが、約500塩基対もしくはそれ未満、約250塩基対もしくはそれ未満、約200塩基対もしくはそれ未満、約150塩基対もしくはそれ未満、約100塩基対もしくはそれ未満、約50塩基対もしくはそれ未満、または約25塩基対もしくはそれ未満である。

30

【0074】

特定の実施形態では、核酸を含有する試料（複数可）を処理せずに、核酸を提供して、本明細書に記載する方法を実施することができる。一部の実施形態では、核酸を含有する試料（複数可）を処理してから、核酸を提供して、本明細書に記載する方法を実施する。例えば、核酸を、試料（複数可）から、抽出し、単離し、精製し、部分的に精製し、または増幅することができる。用語「単離（isolated）」は、本明細書で使用する場合、核酸をその元々の環境（例えば、核酸が天然に存在する場合の天然の環境、または外因性に発現させる場合の宿主細胞）から取り出すことを指し、したがって、ヒトの介入により（例えば、「人の手により」）、核酸は、その元々の環境から変化している。用語「単離核酸」は、本明細書で使用する場合、被験体（例えば、ヒト被験体）から取り出された核酸を指すことができる。単離核酸は、供給源の試料中に存在する成分の量よりも少な

40

50

い非核酸成分（例えば、タンパク質、脂質）を伴って提供され得る。単離核酸を含む組成は、その約50%～99%超が非核酸成分を含有しない場合がある。単離核酸を含む組成は、その約90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超が非核酸成分を含有しない場合がある。用語「精製（purified）」は、本明細書で使用する場合、核酸を精製手順に付す前に存在する非核酸成分（例えば、タンパク質、脂質、炭水化物）の量よりも少ない非核酸成分を含有する核酸を提供することを指すことができる。精製核酸を含む組成は、その約80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超がその他の非核酸成分を含有しない場合がある。用語「精製」は、本明細書で使用する場合、核酸が由来する試料供給源中よりも少ない核酸種を含有する核酸を提供することを指すことができる。精製核酸を含む組成は、その約90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超がその他の核酸種を含有しない場合がある。例えば、胎児核酸を、母体核酸および胎児核酸を含む混合物から精製することができる。特定の例では、胎児核酸の小さな断片を含むヌクレオソームを、母体核酸のより大きな断片を含むより大きなヌクレオソーム複合体の混合物から精製することができる。

10

【0075】

一部の実施形態では、本明細書に記載する方法の前、間または後に、核酸を断片化または切断する。断片化または切断した核酸は、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塩基対の基準の（nominal）、平均（average）または平均値（mean）の長さを有することができる。断片を、当技術分野で公知の適切な方法により生成することができ、核酸断片の平均、平均値または基準の長さを、適切な断片生成手順を選択することによって制御することができる。

20

【0076】

核酸断片は、オーバーラップするヌクレオチド配列を含有することができ、そのようなオーバーラップする配列は、断片化されていない、対応する核酸のヌクレオチド配列、またはそのセグメントの構築を促進することができる。例えば、1つの断片が、サブ配列xおよびyを有する場合があります、別の断片が、サブ配列yおよびzを有する場合があります、x、yおよびzは、5ヌクレオチド長またはそれ超であり得るヌクレオチド配列である。特定の実施形態では、オーバーラップ配列yを利用して、試料に由来する核酸中のx-y-zのヌクレオチド配列の構築を促進することができる。特定の実施形態では、核酸は、部分的に（例えば、不完全なもしくは終結した特定の切断反応から）断片化させてもよく、または完全に断片化させてもよい。

30

【0077】

一部の実施形態では、核酸を、適切な方法により断片化または切断し、それらの非限定的な例として、物理的方法（例えば、せん断、例えば、超音波処理、フレンチプレス、加熱、UV照射等）、酵素処理（例えば、酵素切断剤（例えば、適切なヌクレアーゼ、適切な制限酵素、適切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、ペペリジン、酸加水分解、塩基加水分解、加熱等、もしくはそれらの組合せ）、米国特許出願公開第20050112590号に記載されている処理等、またはそれらの組合せが挙げられる。

40

【0078】

本明細書で使用する場合、「断片化」または「切断」は、核酸分子、例として、核酸鋳型遺伝子分子またはその増幅産物を、2つまたはそれ超のより小さな核酸分子に断片化することができる手順または条件を指す。そのような断片化または切断は、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、物理的断片化を含めた、多

50

様な方法、試薬または条件のうちのいずれかにより達成することができる。

【0079】

本明細書で使用する場合、「断片」、「切断産物」、「切断された産物」、またはそれらの文法上の変型は、核酸鋳型遺伝子分子の断片化もしくは切断の結果として得られた核酸分子、またはそれらの増幅産物を指す。そのような断片または切断された産物は、切断反応の結果として得られた全ての核酸分子を指す場合があるが、典型的には、そのような断片または切断された産物は、核酸鋳型遺伝子分子のうちの対応するヌクレオチド配列を含有する、核酸鋳型遺伝子分子の断片化もしくは切断の結果として得られた核酸分子またはそれらの増幅産物セグメントのみを指す。用語「増幅 (amplified)」は、本明細書で使用する場合、試料中の標的核酸を、標的核酸またはそのセグメントと同じまた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を線形にまたは指数関数的に生成する処理に付すことを指す。特定の実施形態では、用語「増幅」は、ポリメラーゼ連鎖反応 (PCR) を含む方法を指す。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もヌクレオチドを1つまたは複数多く含有することができる (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、「余分な」ヌクレオチド、例として、転写開始配列を含有することができ、その結果、「余分な」ヌクレオチド、または核酸鋳型遺伝子分子のうち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有する増幅産物が生じる)。したがって、断片は、表示される核酸鋳型分子から得られたまたはそれに基づくヌクレオチド配列情報を、少なくとも一部において含有する、増幅された核酸分子のセグメントまたは一部分から生じる断片を含むことができる。

10

20

【0080】

本明細書で使用する場合、用語「補完的切断反応」は、異なる切断試薬を使用して、または同じ切断試薬の切断特異性を変化させることによって、同じ核酸に対して行われる切断反応を指し、したがって、同じ標的または参照の核酸またはタンパク質の代替の切断パターンを生成させる。特定の実施形態では、核酸を、1つまたは複数の反応槽中で、1つまたは複数の特異的切断剤 (例えば、1、2、3、4、5、6、7、8、9、10個またはそれ超の特異的切断剤) を用いて処理することができる (例えば、核酸を、別個の槽中でそれぞれの特異的切断剤を用いて処理する)。用語「特異的切断剤」は、本明細書で使用する場合、核酸を1つまたは複数の特異的な部位において切断することができる作用剤、時には、化学物質または酵素を指す。

30

【0081】

また、本明細書に記載する方法に核酸を提供する前に、核酸中の特定のヌクレオチドを改変する処理に、核酸を曝露させることができる。例えば、核酸を、その中のヌクレオチドのメチル化状況に基づいて選択的に改変する処理を、核酸に適用することができる。加えて、高温、紫外線照射、X線照射等の条件が、核酸分子の配列中に変化を引き起こすことができる。核酸を、適切な配列分析を行うのに有用な任意の適切な形態で提供することができる。

【0082】

核酸は、一本鎖であっても、または二本鎖であってもよい。例えば、二本鎖DNAを、例えば、加熱またはアルカリを用いる処理により変性させることによって、一本鎖DNAを生成することができる。特定の実施形態では、核酸は、二重鎖DNA分子の鎖へオリゴヌクレオチドを侵入させることによって形成されるDループ構造で存在するか、またはDNA様分子、例として、ペプチド核酸 (PNA) で存在する。Dループの形成は、E. Coli RecAタンパク質を添加すること、および/または塩濃度を、例えば、当技術分野で公知の方法を使用して変化させることによって促進することができる。

40

胎児核酸の含有量の決定

【0083】

一部の実施形態では、核酸中の胎児核酸の量 (例えば、濃度、相対量、絶対量、コピー数等) を決定する。特定の実施形態では、試料中の胎児核酸の量を、「胎児フラクション

50

」と呼ぶ。一部の実施形態では、「胎児フラクション」は、妊娠中の雌から得られた試料（例えば、血液試料、血清試料、血漿試料）中の循環無細胞核酸中の胎児核酸のフラクションを指す。特定の実施形態では、雄の胎児に特異的なマーカー（例えば、Y染色体STRマーカー（例えば、DYS19、DYS385、DYS392マーカー）；RhD陰性の雌中のRhDマーカー）、多型配列の対立遺伝子の比に従って、または胎児核酸に特異的であり、母体核酸にはそうでない1つもしくは複数のマーカー（例えば、母親と胎児との間のエピジェネティックなバイオマーカーの差（例えば、メチル化；下記にさらに詳細に記載する）、もしくは母体の血漿中の胎児のRNAマーカー（例えば、Lo、2005年、Journal of Histochemistry and Cytochemistry、53巻（3号）：293～296頁を参照されたい））に従って、胎児核酸の量を決定する。

10

【0084】

胎児核酸の含有量（例えば、胎児フラクション）の決定は時には、例えば、参照により本明細書に援用されている米国特許出願公開第2010/0105049号の記載に従って、胎児数量アッセイ（fetal quantifier assay）（FQA）を使用して行う。このタイプのアッセイにより、母体試料中の胎児核酸を、試料中の核酸のメチル化状況に基づいて検出および定量することが可能になる。特定の実施形態では、母体試料に由来する胎児核酸の量を、存在する核酸の総量に比して決定することができ、それにより、試料中の胎児核酸のパーセントが得られる。特定の実施形態では、母体試料中の胎児核酸のコピー数を決定することができる。特定の実施形態では、配列に特異的（または部分に特異的）な様式で、時には、正確な染色体量分析を可能にする（例えば、胎児の異数性、微小重複または微小欠失の存在または非存在を検出する）のに十分な感度を伴って、胎児核酸の量を決定することができる。

20

【0085】

胎児数量アッセイ（FQA）を、本明細書に記載する方法のうちのいずれかと併せて行うことができる。任意の当技術分野で公知の方法、および／または米国特許出願公開第2010/0105049号の記載により、例えば、差次的メチル化状況に基づいて母体のDNAと胎児のDNAとを区別し、胎児のDNAを定量する（すなわち、その量を決定する）ことができる方法等により、そのようなアッセイを行うことができる。メチル化状況に基づいて核酸を差別化するための方法として、これらに限定されないが、メチル化感受性による、例えば、MBD2-Fc断片（MBD2のメチル結合ドメインが、抗体のFc断片に融合している（MBD-Fc））を使用する捕捉（Gebhardら（2006年）Cancer Res. 66巻（12号）：6118～28頁）；メチル化特異的抗体；亜硫酸水素塩により変換する方法、例えば、MSP（メチル化感受性PCR）、COBRA、メチル化感受性単一ヌクレオチドによるプライマーの伸長（Ms-SNuPE）、またはSequenom MassCLEAVE（商標）技術；およびメチル化感受性制限酵素の使用（例えば、母体試料中の母体のDNAを、1つまたは複数のメチル化感受性制限酵素を使用して消化し、それにより、胎児のDNAを富化する）が挙げられる。また、メチル感受性酵素を使用して、メチル化状況に基づいて核酸を差別化することもでき、これらの酵素は、例えば、後者がメチル化されていない場合には、それらのDNA認識配列において優先的または実質的に切断または消化を行うことができる。したがって、非メチル化DNA試料は、メチル化DNA試料よりも小さな断片に切られ、高度メチル化DNA試料は切断されない。明確な記述がない場合には、メチル化状況に基づいて核酸を差別化するための任意の方法を、本明細書の技術の組成および方法と共に使用することができる。胎児のDNAの量を、増幅反応の間に、例えば、1つまたは複数の競合物質を既知の濃度で導入することによって決定することができる。胎児のDNAの量の決定はまた、例えば、RT-PCR、プライマーの伸長、配列決定および／または計数により行うこともできる。特定の事例では、核酸の量は、米国特許出願公開第2007/0065823号の記載に従ってBEAming技術を使用して決定することができる。特定の実施形態では、制限効率を決定することができ、効率の比率を使用して、胎児のDNAの量をさらに

30

40

50

決定する。

【0086】

特定の実施形態では、胎児数量アッセイ（FQA）を使用して、母体試料中の胎児のDNAの濃度を、例えば、以下の方法により決定することができる：a）母体試料中に存在するDNAの総量を決定し；b）母体試料中の母体のDNAを、1つまたは複数のメチル化感受性制限酵素を使用して選択的に消化し、それにより、胎児のDNAを富化し；c）ステップb）から得られた胎児のDNAの量を決定し；d）ステップc）から得られた胎児のDNAの量を、ステップa）から得られたDNAの総量と比較し、それにより、母体試料中の胎児のDNAの濃度を決定する。特定の実施形態では、母体試料中の胎児核酸の絶対コピー数を、例えば、質量分析および／または絶対コピー数を測定するために競合PCRのアプローチを使用するシステムを使用して決定することができる。例えば、いずれも参照により本明細書に援用されているDingおよびCantor（2003年）Proc. Natl. Acad. Sci. USA, 100巻：3059～3064頁、ならびに米国特許出願公開第2004/0081993号を参照されたい。

10

【0087】

特定の実施形態では、多型配列（例えば、一塩基多型（SNP））の対立遺伝子の比に基づいて、例えば、参照により本明細書に援用されている米国特許出願公開第2011/0224087号に記載の方法等を使用して、胎児フラクションを決定することができる。そのような方法では、ヌクレオチド配列の読取りを、母体試料について得、参照ゲノム中の参考にする多型の部位（例えば、SNP）において、第1の対立遺伝子に対してマッピングされるヌクレオチド配列の読取りの総数と、第2の対立遺伝子に対してマッピングされるヌクレオチド配列の読取りの総数とを比較することによって、胎児フラクションを決定する。特定の実施形態では、例えば、試料中の胎児核酸と母体核酸との混合物に対して、母体核酸はそうした混合物に大きく寄与し、これと比較して、胎児の対立遺伝子の寄与は相対的に小さいことにより、胎児の対立遺伝子を同定する。したがって、母体試料中の胎児核酸の相対的な存在量を、多型の部位のそれら2つの対立遺伝子のそれぞれについての参照ゲノム上の標的核酸配列に対してマッピングしたユニークな配列の読取りの総数のパラメータとして決定することができる。

20

【0088】

一部の実施形態では、断片長情報（例えば、参照により本明細書に援用されている国際出願公開第WO2013/177086号の記載に従う断片長比（fragment length ratio）（FLR）の分析、胎児比統計値（fetal ratio statistic）（FRS）の分析）を組み入れる方法を使用して、胎児フラクションを決定することができる。無細胞胎児核酸の断片は一般に、母体に由来する核酸の断片よりも短い（例えば、Chanら、（2004年）Clin. Chem. 50巻：88～92頁；Loら（2010年）Sci. Transl. Med. 2巻：61ra91を参照されたい）。したがって、一部の実施形態では、特定の長さの閾を下回る断片を計数し、それらのカウントを、例えば、特定の長さの閾を上回る断片から得られたカウント、および／または試料中の全ての核酸の量と比較することによって、胎児フラクションを決定することができる。特定の長さの核酸断片を計数するための方法が、国際出願公開第WO2013/177086号にさらに詳細に記載されている。

30

40

【0089】

一部の実施形態では、部分特異的胎児フラクションの推定値に従って、胎児フラクションを決定することができる。理論に制限されることなく、胎児のCCF断片（例えば、特定の長さまたは範囲の長さの断片）から得られる読取りの量はしばしば、部分に対する距離度（例えば、同じ試料内、例えば、同じ配列決定のラン内）を用いてマッピングされる。また、理論に制限されることなく、特定の部分は、複数の試料間で比較する場合、胎児のCCF断片（例えば、特定の長さまたは範囲の長さの断片）から得られる、読取りの類似の表示を示し、その表示は、部分特異的胎児フラクション（例えば、胎児を起源とするCCF断片の相対量、パーセントまたは比）と相関する傾向を示す。

50

【0090】

一部の実施形態では、部分特異的胎児フラクションの推定値を、一部分、部分特異的パラメータ、および胎児フラクションとのそれらの関係式に基づいて決定する。部分特異的パラメータは、部分中の特定のサイズ（例えば、サイズ範囲）のCCF断片長から得られた読取りの量または比率を反映する（例えば、それと相関する）任意の適切なパラメータであり得る。部分特異的パラメータは、複数の試料について決定された部分特異的パラメータの平均、平均値または中央値であり得る。任意の適切な部分特異的パラメータを使用することができる。部分特異的パラメータの非限定的な例として、FLR（例えば、FRS）、選択された断片長未満の長さを有する読取りの量、ゲノムのカバレッジ（すなわち、カバレッジ）、マッピング可能性、カウント（例えば、部分に対してマッピングされた配列の読取りのカウント、例えば、正規化されたカウント、PERUNにより正規化されたカウント、ChAIにより正規化されたカウント）、デオキシリボヌクレアーゼI感受性、メチル化状況、アセチル化、ヒストンの分布、グアニン-シトシン（GC）含有量、クロマチン構造等、またはそれらの組合せが挙げられる。部分特異的パラメータは、FLRおよび／またはFRSと、部分に特異的な様式で相関する任意の適切なパラメータであり得る。一部の実施形態では、一部または全部の部分特異的パラメータが、部分についての、FLRの直接的または間接的な表示である。一部の実施形態では、部分特異的パラメータは、グアニン-シトシン（GC）含有量ではない。

10

【0091】

一部の実施形態では、部分特異的パラメータは、CCF断片から得られた読取りの量を表示する、それと相関する、またはそれに比例する任意の適切な値であり、この場合、部分に対してマッピングされる読取りは、選択された断片長未満の長さを有する。特定の実施形態では、部分特異的パラメータは、部分に対してマッピングされる比較的短いCCF断片（例えば、約200塩基対もしくはそれ未満）から得られた読取りの量の表示である。選択された断片長未満の長さを有するCCF断片はしばしば、比較的短いCCF断片であり、時には、選択された断片長は、約200塩基対またはそれ未満（例えば、約190、180、170、160、150、140、130、120、110、100、90または80塩基長であるCCF断片）である。任意の適切な方法（例えば、配列決定法、ハイブリダイゼーションのアプローチ）により、CCF断片の長さ、またはCCF断片から得られる読取りを決定（例えば、推定または推測）することができる。一部の実施形態では、CCF断片の長さを、ペアエンドシーケンシング法から得られた読取りにより決定（例えば、推定または推測）する。特定の実施形態では、CCF断片の鋳型の長さを、CCF断片から得られた読取り（例えば、シングルエンドリード）の長さから直接決定する。

20

30

【0092】

1つまたは複数の重み付け係数により、部分特異的パラメータに重み付けする、または部分特異的パラメータを調整することができる。一部の実施形態では、重み付けしたまたは調整した部分特異的パラメータは、試料（例えば、試験試料）についての、部分特異的胎児フラクションの推定値を提供することができる。一部の実施形態では、重み付けまたは調整は一般に、部分のカウント（例えば、部分に対してマッピングされた読取り）、または別の部分特異的パラメータを、部分特異的胎児フラクションの推定値へと変換し、そのような変換は時には、転換とみなされる。

40

【0093】

一部の実施形態では、重み付け係数は一部分、胎児フラクション（例えば、複数の試料から決定した胎児フラクション）と、複数の試料（例えば、トレーニングセット）についての部分特異的パラメータとの間の関係式を記載および／または定義する係数または定数である。一部の実施形態では、重み付け係数を、複数の、胎児フラクションの確定と、複数の部分特異的パラメータとについての関係式に従って決定する。1つの関係式を、1つまたは複数の重み付け係数により定義することができ、1つまたは複数の重み付け係数を、1つの関係式から決定することができる。一部の実施形態では、重み付け係数（例えば、1つまたは複数の重み付け係数）を、（i）複数の試料のそれぞれについて決定した胎

50

児核酸のフラクションと (i i) 複数の試料についての部分特異的パラメータとに従って適合させた、部分についての関係式から決定する。

【0094】

重み付け係数は、適切な関係式（例えば、適切な数学的关系式、代数関係式、適合させた関係式 (fitted relation)、回帰、回帰分析、回帰モデル）から得られる、任意の適切な係数、推定係数または定数であり得る。適切な関係式に従って、そこから誘導して、またはそれから推定して、重み付け係数を決定することができる。一部の実施形態では、重み付け係数は、適合させた関係式から推定された係数である。複数の試料について、関係式を適合させることを時には、モデルをトレーニングすると呼ぶ。関係 (relationship) を適合させる（例えば、モデルをトレーニングして、トレーニングセットを得る）任意の適切なモデルおよび／または方法を使用することができる。使用することができる適切なモデルの非限定的な例として、回帰モデル、線形回帰モデル、単純回帰モデル、通常の最小二乗回帰モデル、重回帰モデル、一般的な重回帰モデル、多項式回帰モデル、一般線形モデル、一般化線形モデル、離散選択回帰モデル、ロジスティック回帰モデル、多項ロジットモデル、混合ロジットモデル、プロビットモデル、多項プロビットモデル、順序ロジットモデル、順序プロビットモデル、ポアソンモデル、多変量応答回帰モデル、マルチレベルモデル、固定効果モデル、ランダム効果モデル、混合モデル、非線形回帰モデル、ノンパラメトリックモデル、セミパラメトリックモデル、ロバストモデル、クォンタイルモデル、アイソトニックモデル、主成分モデル、最小角モデル、ローカルモデル、セグメント化モデル、および変数誤差モデルが挙げられる。一部の実施形態では、適合させた関係式は、回帰モデルではない。一部の実施形態では、適合させた関係式は、決定木モデル、サポートベクターマシンモデル、およびニューラルネットワークモデルから選択される。モデルをトレーニングした結果（例えば、回帰モデル、関係式）はしばしば、数学的に記載することができる関係式となり、この関係式は、1つまたは複数の係数（例えば、重み付け係数）を含む。より複雑な多変量モデルは、1、2、3つまたはそれ超の重み付け係数を決定することができる。一部の実施形態では、複数の試料から得られた胎児フラクションおよび2つまたはそれ超の部分特異的パラメータ（例えば、係数）に従って、モデルをトレーニングする（例えば、複数の試料に、例えば、行列により適合させた関係）。

【0095】

重み付け係数は、適切な方法により、適切な関係式（例えば、適切な数学的关系式、代数関係式、適合させた関係式、回帰、回帰分析、回帰モデル）から得ることができる。一部の実施形態では、適合させた関係式に、推定により適合させ、この非限定的な例として、最小二乗法、通常の最小二乗法、線形回帰、部分回帰、全回帰、一般化回帰、加重回帰、非線形回帰、反復加重回帰、リッジ回帰、最小絶対偏差、ベイズ、ベイズ多変量、縮小ランク、LASSO、Weighted Rank Selection Criteria (WRSC)、Rank Selection Criteria (RSC)、エラスティックネット推定法（例えば、エラスティックネット回帰）、およびそれらの組合せが挙げられる。

【0096】

重み付け係数を、ゲノムの任意の適切な部分について決定する、またはそれと関連付けることができる。重み付け係数を、任意の適切な染色体の任意の適切な部分について決定する、またはそれと関連付けることができる。一部の実施形態では、重み付け係数を、ゲノム中の一部または全部の部分について決定する、またはそれらと関連付ける。一部の実施形態では、重み付け係数を、ゲノム中の一部または全部の染色体の部分について決定する、またはそれらと関連付ける。重み付け係数を時には、選択された染色体の部分について決定する、またはそれらと関連付ける。重み付け係数を、1つまたは複数の常染色体の部分について決定する、またはそれらと関連付けることができる。重み付け係数を、常染色体またはそれらのサブセットの中の部分を含む複数の部分中の部分について決定する、またはそれらと関連付けることができる。一部の実施形態では、重み付け係数を、性染色

体（例えば、Chr Xおよび／またはChr Y）の部分について決定する、またはそれらと関連付ける。重み付け係数を、1つまたは複数の常染色体および1つまたは複数の性染色体の部分について決定する、またはそれらと関連付けることができる。特定の実施形態では、重み付け係数を、全ての常染色体ならびにX染色体およびY染色体中の複数の部分中の部分について決定する、またはそれらと関連付ける。重み付け係数を、X染色体および／またはY染色体中の部分を含まない複数の部分中の部分について決定する、またはそれらと関連付けることができる。特定の実施形態では、重み付け係数を、ある染色体の部分について決定する、またはそれらと関連付け、この染色体は、異数性（例えば、全染色体異数性）を含む。特定の実施形態では、重み付け係数を、ある染色体の部分について決定する、またはそれらのみと関連付け、この染色体は、異数体ではない（例えば、正倍数体染色体である）。重み付け係数を、第13、18および／または21染色体中の部分を含まない複数の部分中の部分について決定する、またはそれらと関連付けることができる。

10

【0097】

一部の実施形態では、重み付け係数を、1つまたは複数の試料（例えば、トレーニングセットの試料）に従って、部分について決定する。重み付け係数はしばしば、部分に特異的である。一部の実施形態では、1つまたは複数の重み付け係数を、部分に独立に割り当てる。一部の実施形態では、重み付け係数を、複数の試料についての胎児フラクションの確定（例えば、試料に特異的な胎児フラクションの確定）のための関係式および複数の試料に従って決定した部分特異的パラメータに従って決定する。重み付け係数はしばしば、複数の試料、例えば、約20個～約100、000個もしくはそれ超、約100個～約100、000個もしくはそれ超、約500個～約100、000個もしくはそれ超、約1000個～約100、000個もしくはそれ超、または約10、000個～約100、000個もしくはそれ超の試料から決定する。重み付け係数を、正倍数体である試料（例えば、正倍数体の胎児を含む被験体から得られた試料、例えば、異数体染色体が存在しない試料）から決定することができる。一部の実施形態では、重み付け係数を、異数体染色体を含む試料（例えば、正倍数体の胎児を含む被験体から得られた試料）から得る。一部の実施形態では、重み付け係数を、正倍数体の胎児を有する被験体およびトリソミーの胎児を有する被験体から得られた複数の試料から決定する。重み付け係数を、複数の試料から得ることができ、これらの試料は、雄の胎児および／または雌の胎児を有する被験体から得られる。

20

30

【0098】

胎児フラクションをしばしば、トレーニングセットの1つまたは複数の試料について決定し、そこから、重み付け係数を誘導する。重み付け係数を決定する胎児フラクションは時には、試料に特異的な胎児フラクションの確定である。重み付け係数を決定する胎児フラクションは、本明細書に記載するまたは当技術分野で公知である任意の適切な方法により決定することができる。一部の実施形態では、胎児核酸の含有量（例えば、胎児フラクション）の決定を、本明細書に記載するまたは当技術分野で公知である適切な胎児数量アッセイ（FQA）を使用して行い、それらの胎児フラクションの確定の非限定的な例として、雄の胎児に特異的なマーカーに従う確定、多型配列の対立遺伝子の比に基づく確定、胎児核酸に特異的であり、母体核酸にはそうでない1つもしくは複数のマーカーに従う確定、メチル化に基づくDNAの識別の使用による確定（例えば、A. Nygrenら（2010年）Clinical Chemistry、56巻（10号）：1627～1635頁）、競合PCRのアプローチを使用する質量分析の方法および／もしくはシステムによる確定、参照により本明細書に援用されている米国特許出願公開第2010/0105049号に記載の方法による確定等、またはそれらの組合せが挙げられる。しばしば胎児フラクションを、一部分、Y染色体のレベル（例えば、1つまたは複数のゲノム区分のレベル；プロファイルのレベル）に従って決定する。一部の実施形態では、Y染色体の適切なアッセイに従って、胎児フラクションを決定する（例えば、定量的リアルタイムPCRを使用することによって、胎児特異的座位（例として、雄を妊娠している場合のY染色体上のSRY座位）の量

40

50

を、母親および胎児の両方に共通する任意の常染色体上の座位の量と比較する（例えば、Lo YMら（1998年）Am J Hum Genet、62巻：768～775頁））。

【0099】

（例えば、試験試料についての）部分特異的パラメータに、1つまたは複数の重み付け係数（例えば、トレーニングセットから誘導した重み付け係数）により重み付けまたは調整を行うことができる。例えば、重み付け係数を、部分について、トレーニングセットの複数の試料についての、部分特異的パラメータと胎児フラクションの確定との関係式に従って誘導することができる。次いで、試験試料の部分特異的パラメータの調整および／または重み付けを、トレーニングセットから誘導した重み付け係数に従って行うことができる。一部の実施形態では、重み付け係数を誘導する部分特異的パラメータが、調整または重み付けを行う（例えば、試験試料の）部分特異的パラメータと同じである（例えば、両方のパラメータがFLRである）。特定の実施形態では、重み付け係数を誘導する部分特異的パラメータが、調整または重み付けを行う（例えば、試験試料の）部分特異的パラメータと異なる。例えば、重み付け係数を、トレーニングセットの試料についての、カバレッジ（すなわち、部分特異的パラメータ）と胎児フラクションとの間の関係式から決定することができる。試験試料の部分についてのFLR（すなわち、別の部分特異的パラメータ）を、カバレッジから誘導した重み付け係数に従って調整することができる。理論により制限されることなく、（例えば、試験試料についての）部分特異的パラメータに時には、それぞれの部分特異的パラメータと共通の部分特異的FLRとの間の関係および／または相関関係に起因して、（例えば、トレーニングセットの）異なる部分特異的パラメータから誘導された重み付け係数により調整および／または重み付けを行うことができる。

【0100】

部分特異的胎児フラクションの推定値を、試料（例えば、試験試料）について、部分特異的パラメータに対して、その部分について決定した重み付け係数により重み付けすることによって決定することができる。重み付けは、任意の適切な数学的操作を適用することによって、部分特異的パラメータを、重み付け係数により調整、変換および／または転換することを含むことができ、それらの非限定的な例として、乗算、除算、加算、減算、積分、記号計算、代数的計算、アルゴリズム、三角関数もしくは幾何関数、転換（例えば、フーリエ変換）等、またはそれらの組合せが挙げられる。重み付けは、適切な数学的モデルによって、部分特異的パラメータを、重み付け係数により調整、変換および／または転換することを含むことができる。

【0101】

一部の実施形態では、胎児フラクションを、試料について、1つまたは複数の部分特異的胎児フラクションの推定値に従って決定する。一部の実施形態では、胎児フラクションを、試料（例えば、試験試料）について、1つまたは複数の部分についての部分特異的パラメータの重み付けまたは調整に従って決定（例えば、推定）する。特定の実施形態では、試験試料についての胎児核酸のフラクションを、調整したカウントまたは調整したサブセットのカウントに基づいて推定する。特定の実施形態では、試験試料についての胎児核酸のフラクションを、部分についての、調整したFLR、調整したFRS、調整したカバレッジおよび／または調整したマッピング可能性に基づいて推定する。一部の実施形態では、約1～約500、000個、約100～約300、000個、約500～約200、000個、約1000～約200、000個、約1500～約200、000個、または約1500～約50、000個の部分特異的パラメータの重み付けまたは調整を行う。

【0102】

（例えば、試験試料についての）胎児フラクションを、任意の適切な方法により、（例えば、同じ試験試料についての）複数の部分特異的胎児フラクションの推定値に従って決定することができる。一部の実施形態では、妊娠中の雌から得られたある試験試料中の胎児核酸のフラクションの推定の精度を向上させるための方法は、1つまたは複数の部分特異的胎児フラクションの推定値を決定するステップを含み、この試料についての胎児フラクションの推定値は、これら1つまたは複数の部分特異的胎児フラクションの推定値に従

って決定される。一部の実施形態では、胎児核酸のフラクションを、試料（例えば、試験試料）について推定または決定するステップは、1つまたは複数の部分特異的胎児フラクションの推定値を合計するステップを含む。合計するステップは、複数の部分特異的胎児フラクションの推定値に従って、平均、平均値、中央値、AUCまたは積分値を決定することを含むことができる。

【0103】

一部の実施形態では、妊娠中の雌から得られた試験試料中の胎児核酸のフラクションの推定の精度を向上させるための方法は、参照ゲノムの部分に対してマッピングした配列の読取りのカウントを得るステップを含み、これらの配列の読取りは、妊娠中の雌に由来する試験試料から得られた循環無細胞核酸の読取りであり、得られたカウントの少なくとも1つのサブセットは、ゲノムのある領域から得られ、この領域が提供する、この領域に由来する全カウントと比べた胎児核酸から得られたカウントは、ゲノムの別の領域の全カウントと比べた胎児核酸のカウントよりも多い。一部の実施形態では、胎児核酸のフラクションの推定値を、部分のあるサブセットに従って決定し、部分のこのサブセットは、別の部分の胎児核酸のカウントよりも多い数の、胎児核酸から得られたカウントがマッピングされる部分に従って選択される。一部の実施形態では、部分のこのサブセットは、別の部分の非胎児核酸と比べた胎児核酸のカウントよりも多い数の、非胎児核酸と比べた胎児核酸から得られたカウントがマッピングされる部分に従って選択される。部分の全てまたはサブセットに対してマッピングされたカウントに重み付けすることができ、それにより、重み付けしたカウントを提供する。重み付けしたカウントを利用して、胎児核酸のフラクションを推定することができ、別の部分の胎児核酸のカウントよりも多い数の、胎児核酸から得られたカウントがマッピングされる部分に従って、カウントに重み付けすることができる。一部の実施形態では、別の部分の非胎児核酸と比べた胎児核酸のカウントよりも多い数の、非胎児核酸と比べた胎児核酸から得られたカウントがマッピングされる部分に従って、カウントに重み付けする。

【0104】

胎児フラクションを、試料（例えば、試験試料）について、該試料についての複数の部分特異的胎児フラクションの推定値に従って決定することができ、部分に特異的な推定値は、ゲノムの任意の適切な領域またはセグメントの部分から得られる。部分特異的胎児フラクションの推定値を、適切な染色体（例えば、1つもしくは複数の選択された染色体、1つもしくは複数の常染色体、性染色体（例えば、Chr Xおよび／もしくはChr Y）、異数体染色体、正倍数体染色体等、またはそれらの組合せ）の1つまたは複数の部分について決定することができる。

【0105】

一部の実施形態では、胎児フラクションを決定するステップは、（a）参照ゲノムの部分に対してマッピングした配列の読取りのカウントを得るサブステップ（これらの配列の読取りは、妊娠中の雌に由来する試験試料から得られた循環無細胞核酸の読取りである）と、（b）マイクロプロセッサを使用して、胎児核酸の部分特異的フラクションに関して、（i）それぞれの部分に対してマッピングした配列の読取りのカウントに対してかまたは（i i）その他の部分特異的パラメータに対して、それぞれの部分と独立に関連付けた重み付け係数に従って重み付けし、それにより、重み付け係数に従う、部分特異的胎児フラクションの推定値を得るサブステップ（複数の試料について、重み付け係数のそれぞれは、（i）複数の試料のそれぞれについての胎児核酸のフラクションと、（i i）それぞれの部分に対してマッピングした配列の読取りのカウントまたはその他の部分特異的パラメータとの間でそれぞれの部分について適合させた関係式から決定されている）と、（c）試験試料についての胎児核酸のフラクションを、部分特異的胎児フラクションの推定値に基づいて推定するサブステップとを含む。

【0106】

本明細書に提供する方法と併せて、細胞外核酸中の胎児核酸の量を、定量し、使用することができる。したがって、特定の実施形態では、本明細書に記載する技術の方法は、胎

10

20

30

40

50

児核酸の量を決定する追加のステップを含む。被験体から得られた核酸試料中の胎児核酸の量を、試料核酸を調製するための処理の前または後で決定することができる。特定の実施形態では、試料核酸を処理し、調製した後で、試料中の胎児核酸の量を決定し、この量を利用して、さらなる評価を行う。一部の実施形態では、アウトカムは、試料核酸中の胎児核酸のフラクションの寄与の程度を加減する（例えば、カウントを調整する、試料を除去する、判定を行う、または判定を行わない）ことを含む。

【0107】

決定のステップを、本明細書に記載する方法の前、その間、その中の任意の一点、または本明細書に記載する特定（例えば、異数性の検出、微小重複もしくは微小欠失の検出、胎児の性別の決定）の方法の後に行うことができる。例えば、胎児の性別または異数性、微小重複もしくは微小欠失の決定方法を所与の感度または特異性で行うために、胎児核酸を定量する方法を、胎児の性別または異数性、微小重複もしくは微小欠失の決定の前、間または後に実行して、約2%超、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%またはそれ超の胎児核酸を有する試料を同定することができる。一部の実施形態では、例えば、特定の閾量の胎児核酸（例えば、約15%またはそれ超の胎児核酸；約4%またはそれ超の胎児核酸）を有すると決定された試料を、胎児の性別または異数性、微小重複もしくは微小欠失の決定のために、あるいは異数性または遺伝子の変動の存在または非存在について、さらに分析する。特定の実施形態では、試料が、特定の閾量の胎児核酸（例えば、約15%またはそれ超の胎児核酸；約4%またはそれ超の胎児核酸）を有する場合のみに、例えば、胎児の性別または異数性、微小重複もしくは微小欠失の存在または非存在の確定を選択する（例えば、選択し、患者に伝える）。

【0108】

一部の実施形態では、染色体の異数性、微小重複または微小欠失の存在または非存在を確認するために、胎児フラクションの決定または胎児核酸の量の決定が、要求されることも、必要になることもない。一部の実施形態では、染色体の異数性、微小重複または微小欠失の存在または非存在の確認が、胎児のDNAと母体のDNAとの配列の差別化を必要としない。特定の実施形態では、この理由は、特定の染色体、染色体部分またはそのセグメントにおける母体配列および胎児配列の両方の合計された寄与を分析するからである。一部の実施形態では、染色体の異数性、微小重複または微小欠失の存在または非存在の確認は、胎児のDNAと母体のDNAとを区別するであろう先験的な配列情報に依存しない。

核酸の富化

【0109】

一部の実施形態では、核酸（例えば、細胞外核酸）を、富化し、または相対的に富化して、核酸の亜集団または種を得る。核酸の亜集団は、例えば、胎児核酸、母体核酸、特定の長さもしくは特定の範囲の長さの断片を含む核酸、または特定のゲノム領域（例えば、単一の染色体、一連の染色体および／もしくは特定の染色体領域）に由来する核酸を含むことができる。そのような富化試料は、本明細書に提供する方法と併せて使用することができる。したがって、特定の実施形態では、本技術の方法は、試料中の核酸の亜集団、例えば、胎児核酸等について富化する追加のステップを含む。特定の実施形態では、富化して、胎児核酸を得るために、上記に記載した、胎児フラクションを決定するための方法もまた使用することができる。特定の実施形態では、母体核酸を、試料から、選択的に（部分的に、実質的に、ほとんど完全に、または完全に）除去する。特定の実施形態では、富化して、特定の低いコピー数の種の核酸（例えば、胎児核酸）を得ることによって、定量的感度を改善することができる。試料を核酸の特定の種について富化するための方法が、例えば、米国特許第6,927,028号、国際特許出願公開第WO2007/140417号、国際特許出願公開第WO2007/147063号、国際特許出願公開第WO2009/032779号、国際特許出願公開第WO2009/032781号、国際特許

出願公開第W O 2 0 1 0 / 0 3 3 6 3 9 号、国際特許出願公開第W O 2 0 1 1 / 0 3 4 6 3 1 号、国際特許出願公開第W O 2 0 0 6 / 0 5 6 4 8 0 号および国際特許出願公開第W O 2 0 1 1 / 1 4 3 6 5 9 号に記載されており、これらは全て、参照により本明細書に援用されている。

【0110】

一部の実施形態では、核酸を富化して、特定の標的断片種および／または参照断片種を得る。特定の実施形態では、下記に記載する1つまたは複数の、長さに基づく分離の方法を使用して、核酸を富化して、特定の核酸の断片長または特定の範囲の断片長を得る。特定の実施形態では、本明細書に記載するおよび／または当技術分野で公知である1つまたは複数の、配列に基づく分離方法を使用して、核酸を富化して、選択ゲノム領域（例えば、染色体）に由来する断片を得る。下記に、試料中の核酸の亜集団（例えば、胎児核酸）について富化するための特定の方法を詳細に記載する。

10

【0111】

本明細書に記載する方法と共に使用することができる、核酸の亜集団（例えば、胎児核酸）について富化するためのいくつかの方法は、母体核酸と胎児核酸との間のエピジェネティックな差を活用する方法を含む。例えば、メチル化の差に基づいて、胎児核酸を、母体核酸と差別化し、それから分離することができる。メチル化に基づく胎児核酸の富化方法が、参照により本明細書に援用されている米国特許出願公開第2010/0105049号に記載されている。そのような方法は時には、試料核酸を、メチル化特異的結合剤（メチル-CpG結合タンパク質（MBD）、メチル化特異的抗体等）に結合させるステップと、差次的メチル化状況に基づいて、未結合の核酸から、結合した核酸を分離するステップとを含む。そのような方法はまた、メチル化感受性制限酵素（上記に記載のとおり；例えば、HhaIおよびHpaII）の使用を含むこともでき、この方法により、母体核酸を選択的かつ完全または実質的に消化して、試料を少なくとも1つの胎児核酸の領域について富化する酵素を用いて、母体試料に由来する核酸を選択的に消化することによって、母体試料中の胎児核酸の領域の富化が可能になる。

20

【0112】

本明細書に記載する方法と共に使用することができる、核酸の亜集団（例えば、胎児核酸）について富化するための別の方法が、参照により本明細書に援用されている米国特許出願公開第2009/0317818号に記載の方法等の制限エンドヌクレアーゼにより多型配列を増強するアプローチである。そのような方法は、非標的対立遺伝子を含む核酸を、非標的対立遺伝子を含むが、標的対立遺伝子は含まない核酸を認識する制限エンドヌクレアーゼを用いて切断するステップと、切断された核酸は増幅せずに、未切断の核酸を増幅するステップとを含み、ここで、未切断の、増幅された核酸は、非標的核酸（例えば、母体核酸）と比べて富化された標的核酸（例えば、胎児核酸）を表す。特定の実施形態では、例えば、切断剤による選択的消化を受けやすい多型の部位を有する対立遺伝子を含むように、核酸を選択することができる。

30

【0113】

本明細書に記載する方法と共に使用することができる、核酸の亜集団（例えば、胎児核酸）について富化するためのいくつかの方法は、選択的酵素分解のアプローチを含む。そのような方法は、エキソヌクレアーゼ消化から標的配列を保護し、それにより、試料中の望まれない配列（例えば、母体のDNA）の排除を促進するステップを含む。例えば、1つのアプローチでは、試料核酸を変性させて、一本鎖核酸を生成し、一本鎖核酸を、適切なアニリング条件下で、少なくとも1つの、標的特異的プライマーの対と接触させ、アニールさせたプライマーを、ヌクレオチドの重合により伸長して、二本鎖標的配列を生成し、一本鎖（すなわち、非標的）の核酸を消化するヌクレアーゼを使用して、一本鎖核酸を消化する。特定の実施形態では、少なくとも1回の追加のサイクルにおいて、この方法を繰り返すことができる。特定の実施形態では、同じ、標的特異的プライマーの対を使用して、第1サイクルおよび第2サイクルのそれぞれにおいてプライマーの伸長を行い、特定の実施形態では、第1サイクルおよび第2サイクルのために、異なる、標的特異的プラ

40

50

イマーの対を使用する。

【0114】

本明細書に記載する方法と共に使用することができる、核酸の垂集団（例えば、胎児核酸）について富化するためのいくつかの方法は、大規模並行シグネチャー配列決定（MPSS）のアプローチを含む。MPSSは典型的には、アダプター（すなわち、タグ）のライゲーションを使用し、続いて、アダプターのデコーディングを行い、核酸配列をこざみに読み取る固相法である。典型的には、タグを付けたPCR産物が増幅され、結果として、それぞれの核酸から、ユニークなタグを有するPCR産物が生成する。しばしば、PCR産物をマイクロビーズにつなぐために、タグを使用する。ライゲーションに基づく配列決定を数ラウンド行った後に、例えば、配列のシグネチャーを、それぞれのビーズから同定することができる。MPSSデータセット中のそれぞれのシグネチャー配列（MPSSタグ）を、分析し、全てのその他のシグネチャーと比較し、全ての同一のシグネチャーを計数する。

10

【0115】

特定の実施形態では、特定の富化方法（例えば、特定の、MPSSおよび／またはMPSSに基づく富化方法）は、増幅（例えば、PCR）に基づくアプローチを含むことができる。特定の実施形態では、座位に特異的な増幅方法を使用することができる（例えば、座位に特異的な増幅プライマーを使用する）。特定の実施形態では、マルチプレックスSNP対立遺伝子PCRのアプローチを使用することができる。特定の実施形態では、マルチプレックスSNP対立遺伝子PCRのアプローチを、ユニプレックス配列決定と組み合わせて使用することができる。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）の使用、および捕捉プローブ配列のアンプリコン中への組み入れ、続いて、例えば、Illumina MPSSシステムを使用する配列決定を含むことができる。特定の実施形態では、マルチプレックスSNP対立遺伝子PCRのアプローチを、3つのプライマーからなるシステムおよびインデックス化配列決定と組み合わせて使用することができる。例えば、そのようなアプローチは、例えば、Illumina MPSSシステムを使用する配列決定のために、特定の座位に特異的なフォワードPCRプライマー中に組み入れた第1の捕捉プローブ、および座位に特異的なリバーズPCRプライマー中に組み入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）を使用し、それにより、アンプリコンを生成し、続いて、リバーズ捕捉配列および分子インデックスバーコードを組み入れるための第2のPCRを行うことを含むことができる。特定の実施形態では、マルチプレックスSNP対立遺伝子PCRのアプローチを、4つのプライマーからなるシステムおよびインデックス化配列決定と組み合わせて使用することができる。例えば、そのようなアプローチは、例えば、Illumina MPSSシステムを使用する配列決定のために、座位に特異的なフォワードPCRプライマーおよび座位に特異的なリバーズPCRプライマーの両方中に組み入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）を使用し、続いて、フォワード捕捉配列およびリバーズ捕捉配列の両方ならびに分子インデックスバーコードを組み入れるための第2のPCRを行うことを含むことができる。特定の実施形態では、マイクロ流体技術のアプローチを使用することができる。特定の実施形態では、アレイに基づくマイクロ流体技術のアプローチを使用することができる。例えば、そのようなアプローチは、マイクロ流体技術によるアレイ（例えば、Fluidigm）を使用して、低いプレックス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入れを行い、続いて、配列決定を行うことを含むことができる。特定の実施形態では、例えば、デジタル小滴PCR等のエマルジョンマイクロ流体技術のアプローチを使用することができる。

20

30

40

【0116】

特定の実施形態では、（例えば、ユニバーサルプライマーまたは座位に特異的でない増幅プライマーを使用して）ユニバーサル増幅法を使用することができる。特定の実施形態では、ユニバーサル増幅法を、プルダウンのアプローチと組み合わせて使用することがで

50

きる。特定の実施形態では、方法は、ユニバーサルに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によるプルダウン（例えば、AgilentまたはIDT製のビオチン化プルダウンアッセイ）を含むことができる。例えば、そのようなアプローチは、標準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についての富化、および第2のユニバーサル増幅のステップを含むことができる。特定の実施形態では、プルダウンのアプローチは、ライゲーションに基づく方法と組み合わせて使用することができる。特定の実施形態では、方法は、配列特異的アダプターのライゲーションを用いるビオチン化ウルトラマーによるプルダウン（例えば、HALOPLEX PCR、Halo Genomics）を含むことができる。例えば、そのようなアプローチは、制限酵素消化断片を捕捉するためのセクタープローブの使用、続いて、捕捉された産物のアダプターへのライゲーション、およびユニバーサル増幅、続いて、配列決定を含むことができる。特定の実施形態では、プルダウンのアプローチを、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することができる。特定の実施形態では、方法は、分子反転プローブ（MIP）による伸長およびライゲーションを含むことができる。例えば、そのようなアプローチは、配列アダプターと組み合わせた分子反転プローブの使用、続いて、ユニバーサル増幅および配列決定を含むことができる。特定の実施形態では、相補的DNAを、合成し、増幅せずに配列決定することができる。

10

【0117】

特定の実施形態では、伸長およびライゲーションのアプローチを、プルダウンのコンポーネントなしで行うことができる。特定の実施形態では、方法は、座位に特異的なフォワードプライマーおよびリバースプライマーによるハイブリダイゼーション、伸長、ならびにライゲーションを含むことができる。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的DNA合成、続いて、配列決定をさらに含むことができる。特定の実施形態では、そのような方法は、分析の間のバックグラウンドの配列を低下させるまたは排除することができる。

20

【0118】

特定の実施形態では、プルダウンのアプローチを、任意選択の増幅コンポーネントを伴わせて、または増幅コンポーネントなしで使用することができる。特定の実施形態では、方法は、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むことができ、捕捉プローブを十分に組み入れ、ユニバーサル増幅は行わない。例えば、そのようなアプローチは、制限酵素消化断片を捕捉するための、改変されたセクタープローブの使用、続いて、捕捉された産物のアダプターへのライゲーション、任意選択の増幅、および配列決定を含むことができる。特定の実施形態では、方法は、環状一本鎖ライゲーションと組み合わせた、アダプター配列の伸長およびライゲーションを伴う、ビオチン化プルダウンアッセイを含むことができる。例えば、そのようなアプローチは、目的の捕捉領域（すなわち、標的配列）に対するセクタープローブの使用、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、任意選択の増幅、および配列決定を含むことができる。特定の実施形態では、配列決定の結果の分析により、バックグラウンドから標的配列を分離することができる。

30

【0119】

一部の実施形態では、本明細書に記載する1つまたは複数の、配列に基づく分離方法を使用して、核酸を富化して、選択ゲノム領域（例えば、染色体）に由来する断片を得る。配列に基づく分離は一般に、ヌクレオチド配列が、目的の断片（例えば、標的断片および／または参照断片）中には存在し、試料のその他の断片中に実質的に存在しない、またはその他の断片はごくわずかな量でしか存在しない（例えば、5%もしくはそれ未満）ことに基づく。一部の実施形態では、配列に基づく分離は、標的断片の分離および／または参照断片の分離をもたらすことができる。分離された標的断片および／または分離された参照断片をしばしば、核酸試料中の残存する断片から単離し、取り出す。特定の実施形態では、また、分離された標的断片と分離された参照断片とを、相互に単離し、取り出す（例えば、分離アッセイのコンパートメントとして単離する）。特定の実施形態では、分離さ

40

50

れた標的断片と分離された参照断片とを、一緒に単離する（例えば、同じアッセイコンパートメントとして単離する）。一部の実施形態では、未結合断片を、分別的に除去または分解または消化することができる。

【0120】

一部の実施形態では、選択的に核酸を捕捉する処理を使用して、核酸試料から、標的断片および／または参照断片を分離し、取り出す。市販されている、核酸を捕捉するシステムとして、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）；および関連のプラットフォームが挙げられる。そのような方法は典型的には、標的断片または参照断片のヌクレオチド配列のセグメントまたは全てに対する捕捉オリゴヌクレオチド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／または溶液に基づくプラットフォームの使用を含むことができる。選択されたゲノム領域または座位（例えば、第21、18、13、XもしくはY染色体のうちの1つ、または参照の染色体）に由来する核酸断片に優先的にハイブリダイズするように、捕捉オリゴヌクレオチド（時には、「おとり」と呼ぶ）を、選択するまたは設計することができる。特定の実施形態では、（例えば、オリゴヌクレオチドアレイを使用する）ハイブリダイゼーションに基づく方法を使用し、富化して、特定の染色体（例えば、潜在的に異数体の染色体、参照の染色体、もしくは目的のその他の染色体）、またはそれらの目的のセグメントに由来する核酸配列を得ることができる。

【0121】

一部の実施形態では、1つまたは複数の、長さに基づく分離の方法を使用して、核酸を、特定の核酸断片の長さ、特定の範囲の長さ、または特定の閾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富化する。核酸断片の長さは典型的には、断片中のヌクレオチドの数を指す。また、核酸断片の長さは時には、核酸断片のサイズとも呼ぶ。一部の実施形態では、長さに基づく分離の方法を、個々の断片の長さを測定することなく実施する。一部の実施形態では、長さに基づく分離の方法を、個々の断片の長さを決定するための方法と併せて実施する。一部の実施形態では、長さに基づく分離は、サイズ分画の手順を指し、分画されたプールの全部または一部を、単離（例えば、保持）および／または分析することができる。サイズ分画の手順は、当技術分野で公知である（例えば、アレイ上での分離、分子ふるいによる分離、ゲル電気泳動による分離、カラムクロマトグラフィー（例えば、分子ふるいカラム）による分離、およびマイクロ流体技術に基づくアプローチ）。特定の実施形態では、長さに基づく分離のアプローチとして、例えば、断片の環状化、化学物質による処理（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／またはサイズに特異的な核酸増幅を挙げることができる。

【0122】

本明細書に記載する方法と共に使用することができる、特定の長さに基づく分離の方法は、例えば、選択的な配列によるタグ付けのアプローチを利用する。用語「配列によるタグ付け」は、認識可能であり、かつ明確に異なる配列を、核酸または核酸の集団中に組み入れることを指す。用語「配列によるタグ付け」は、本明細書で使用する場合、本明細書で後に記載する用語「配列タグ」とは異なる意味を有する。そのような配列によるタグ付けの方法では、ある断片サイズの種（例えば、短い断片）の核酸を、長い核酸および短い核酸を含む試料中で、選択的な配列によるタグ付けに付す。そのような方法は典型的には、核酸増幅反応を、内側プライマーおよび外側プライマーを含むセットのネステッドプライマーを使用して実施するステップを含む。特定の実施形態では、内側プライマーの一方または両方にタグを付け、それにより、タグを標的の増幅産物上に導入することができる。外側プライマーは一般に、（内側の）標的配列を担持する短い断片にはアニールしない

。内側プライマーは、短い断片にアニールし、タグおよび標的配列を担持する増幅産物を生成することができる。典型的には、長い断片のタグ付けは、例えば、外側プライマーの以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遮断を含む、機構の組合せを通して阻害される。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使用する、タグを付けた断片の増幅を含めた、多様な方法のうちのいずれかにより、タグを付けた断片についての富化を行うことができる。

【0123】

本明細書に記載する方法と共に使用することができる、別の、長さに基づく分離の方法は、核酸試料を、ポリエチレングリコール（PEG）沈殿に付すステップを含む。方法の例として、国際特許出願公開第WO2007/140417号および同第WO2010/115016号に記載されているものが挙げられる。この方法は一般に、小さな（例えば、300ヌクレオチド未満の）核酸を実質的に沈殿させることなく、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つまたは複数の一価の塩の存在下で、核酸試料をPEGと接触させることを必要とする。

10

【0124】

本明細書に記載する方法と共に使用することができる、別の、サイズに基づく富化方法は、ライゲーション、例えば、*circ ligase*を使用するライゲーションによる環状化を含む。短い核酸断片は典型的には、長い断片よりも高い効率で環状化させることができる。環状化しなかった配列を、環状化した配列から分離することができ、富化した短い断片を使用して、さらなる分析を行うことができる。

20

核酸ライブラリー

【0125】

一部の実施形態では、核酸ライブラリーは、特定の処理（それらの非限定的な例として、固相（例えば、固体の支持体、例えば、フローセル、ビーズ）上への固定化、富化、増幅、クローニング、検出が挙げられる）のために、および／または核酸の配列決定のために、調製され、アセンブルされ、かつ／または改変される複数のポリヌクレオチド分子（例えば、核酸の試料）である。特定の実施形態では、核酸ライブラリーを、配列決定の処理の前または間に調製する。核酸ライブラリー（例えば、配列決定ライブラリー）を、当技術分野で公知の適切な方法により調製することができる。核酸ライブラリーを、標的化する調製処理または標的化しない調製処理により調製することができる。

30

【0126】

一部の実施形態では、核酸のライブラリーを改変して、固体の支持体への核酸の固定化のために構成される化学的部分（例えば、官能基）を含める。一部の実施形態では、核酸のライブラリーを改変して、固体の支持体へのライブラリーの固定化のために構成される、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め、それらの非限定的な例として、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、核酸結合タンパク質、受容体、炭水化物、オリゴヌクレオチド、ポリヌクレオチド、相補的核酸配列等、およびそれらの組合せが挙げられる。特異的な結合対のいくつかの例として、非限定的に、アビジン部分とビオチン部分；抗原性エピトープと、抗体もしくはその免疫学的反応性断片；抗体とハプテン；ジゴキシゲニン（*digoxigen*）部分と抗ジゴキシゲニン（*anti-digoxigen*）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；活性化化合物と活性化化合物の受容体；ホルモンとホルモン受容体；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もしくはポリヌクレオチドと、それに対応する相補体等、またはそれらの組合せが挙げられる。

40

【0127】

一部の実施形態では、核酸のライブラリーを改変して、既知の組成の1つまたは複数の

50

ポリヌクレオチドを含め、それらの非限定的な例として、識別子（例えば、タグ、インデックス化タグ）、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、複製開始点、ステムループ、相補配列（例えば、プライマー結合部位、アニーリング部位）、適切な組込み部位（例えば、トランスポゾン、ウイルス組込み部位）、改変ヌクレオチド等、またはそれらの組合せが挙げられる。既知の配列のポリヌクレオチドを、適切な位置、例えば、核酸配列の5'末端、3'末端または内部に付加することができる。既知の配列のポリヌクレオチドは、同じ配列であっても、または異なる配列であってもよい。一部の実施形態では、既知の配列のポリヌクレオチドを、表面（例えば、フローセル中の表面）上に固定化された1つまたは複数のオリゴヌクレオチドにハイブリダイズするように構成する。例えば、5'既知配列を含む核酸分子を、第1の、複数のオリゴヌクレオチドにハイブリダイズさせることができ、一方、その分子の3'既知配列を、第2の、複数のオリゴヌクレオチドにハイブリダイズさせることができる。一部の実施形態では、核酸のライブラリーは、染色体に特異的なタグ、捕捉配列、標識および／またはアダプターを含むことができる。一部の実施形態では、核酸のライブラリーは、1つまたは複数の検出可能な標識を含む。一部の実施形態では、1つまたは複数の検出可能な標識を、核酸ライブラリー中に、5'末端において、3'末端において、かつ／または該ライブラリー中の核酸の内部の任意のヌクレオチドの位置において組み入れることができる。一部の実施形態では、核酸のライブラリーは、ハイブリダイズさせたオリゴヌクレオチドを含む。特定の実施形態では、ハイブリダイズさせたオリゴヌクレオチドは、標識されたプローブである。一部の実施形態では、核酸のライブラリーは、固相上への固定化する前のハイブリダイズさせたオリゴヌクレオチドプローブを含む。

【0128】

一部の実施形態では、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つもしくはそれ超の核酸分子、または核酸分子の2つもしくはそれ超のサブセット中に組み込む特異的なヌクレオチド配列であり、ユニバーサル配列は、それが組み込まれている分子またはサブセットの分子全てについて同じである。ユニバーサル配列はしばしば、ユニバーサル配列に対して相補性である単一のユニバーサルプライマーを使用して、複数の異なる配列にハイブリダイズし、かつ／またはそれらを増幅するように設計される。一部の実施形態では、2つ（例えば、対）またはそれ超の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を使用する。ユニバーサルプライマーはしばしば、ユニバーサル配列を含む。一部の実施形態では、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。一部の実施形態では、1つまたは複数のユニバーサル配列を使用して、核酸の複数の種またはサブセットを、捕捉、同定および／または検出する。

【0129】

核酸ライブラリーの調製の特定の実施形態では（例えば、合成の手順による特定の配列決定の場合には）、核酸を、サイズにより選択および／または断片化して、数百塩基対またはそれ未満の長さにする（例えば、ライブラリーの生成のための調製の場合）。一部の実施形態では、ライブラリーの調製を、断片化せずに行う（例えば、ccfDNAを使用する場合）。

【0130】

特定の実施形態では、ライゲーションに基づくライブラリーの調製方法を使用する（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の調製方法はしばしば、アダプター（例えば、メチル化アダプター）の設計を活用し、この設計は、最初のライゲーションのステップにおいて、インデックス配列を組み入れることができ、しばしば、シングルリードシーケンシング、ペアエンドシーケンシング、およびマルチプレックスシーケンシングのための試料を調製するために使用することができる。例えば、fill-in反応、エキソヌクレアーゼ反応、またはそれらの組合せにより、時には、核酸（例えば、断片化核酸またはccfDNA）の末端の修復をもたらす。一部の実施形態では、次いで、得られた平滑末端

修復核酸を、アダプター／プライマーの3'末端上の単一ヌクレオチドのオーバーハングに対して相補性である単一ヌクレオチドにより伸長することができる。任意のヌクレオチドを、伸長／オーバーハングヌクレオチドのために使用することができる。一部の実施形態では、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のライゲーションを含む。アダプターオリゴヌクレオチドはしばしば、フローセルアンカーに対して相補性であり、時には、例えば、核酸ライブラリーを、固体の支持体、例として、フローセルの内側表面に固定化するために利用される。一部の実施形態では、アダプターオリゴヌクレオチドは、識別子、1つもしくは複数の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に対して相補性である配列、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、マルチプレックス配列決定プライマー等）、またはそれらの組合せ（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。

10

【0131】

識別子は、核酸（例えば、ポリヌクレオチド）中に組み入れるまたはそれにつなぐ、適切な検出可能な標識であり、識別子により、識別子を含む核酸の検出および／または同定が可能になる。一部の実施形態では、識別子を、配列決定法の間に、（例えば、ポリメラーゼにより）核酸中に組み入れるまたはそれにつなぐ。識別子の非限定的な例として、核酸タグ、核酸のインデックスもしくはバーコード、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、抗体もしくはその一部分、リンカー、結合対のメンバー）等、またはそれらの組合せが挙げられる。一部の実施形態では、識別子（例えば、核酸のインデックスまたはバーコード）は、ユニークな、既知のおよび／または同定可能な配列のヌクレオチドまたはヌクレオチド類似体である。一部の実施形態では、識別子は、6つまたはそれ超の連続ヌクレオチドである。多様な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入手可能である。任意の適切なタイプおよび／または数のフルオロフォアを、識別子として使用することができる。一部の実施形態では、1つもしくは複数、2つもしくはそれ超、3つもしくはそれ超、4つもしくはそれ超、5つもしくはそれ超、6つもしくはそれ超、7つもしくはそれ超、8つもしくはそれ超、9つもしくはそれ超、10個もしくはそれ超、20個もしくはそれ超、30個もしくはそれ超、または50個もしくはそれ超の異なる識別子が、本明細書に記載する方法（例えば、核酸の検出および／または配列決定法）において利用される。一部の実施形態では、1つまたは2つのタイプの識別子（例えば、蛍光標識）を、ライブラリー中のそれぞれの核酸に連結する。識別子の検出および／または定量を、適切な方法または装置により行うことができ、それらの非限定的な例として、フローサイトメトリー、定量的ポリメラーゼ連鎖反応（qPCR）、ゲル電気泳動、ルミノメーター、蛍光光度計、分光光度計、適切な遺伝子チップもしくはマイクロアレイによる分析、ウェスタンブロット、質量分析、クロマトグラフィー、細胞蛍光測定法による分析、蛍光顕微鏡法、適切な蛍光法もしくはデジタル撮像法、共焦点レーザー走査顕微鏡法、レーザー走査細胞数測定、親和性クロマトグラフィー、手作業バッチモードによる分離、電場懸濁、適切な核酸配列決定法および／または核酸シーケンサー等、ならびにそれらの組合せが挙げられる。

20

30

40

【0132】

一部の実施形態では、トランスポゾンに基づくライブラリーの調製方法を使用する（例えば、EPICENTRE NEXTERA、Epicentre、Madison WI）。トランスポゾンに基づく方法は典型的には、*in vitro*での転位を使用して、単一チューブ中での反応においてDNAの断片化およびタグ付けを同時に行い（しばしば、プラットフォームに特異的なタグおよび任意選択のバーコードの組み入れが可能である）、シーケンサーで使用するライブラリーを調製する。

【0133】

一部の実施形態では、核酸ライブラリーまたはその一部分を増幅する（例えば、PCRに基づく方法により増幅する）。一部の実施形態では、配列決定法は、核酸ライブラリー

50

の増幅を含む。核酸ライブラリーを、固体の支持体（例えば、フローセル中の固体の支持体）上への固定化の前または後に増幅することができる。核酸増幅は、（例えば、核酸ライブラリー中に）存在する核酸鋳型および／またはその相補体の数を、鋳型および／またはその相補体の1つまたは複数のコピーを生成することによって増幅するまたは増加させる処理を含む。増幅は、適切な方法により行うことができる。核酸ライブラリーを、サーモサイクリング法または等温増幅法により増幅することができる。一部の実施形態では、ローリングサークル増幅法を使用する。一部の実施形態では、増幅は、核酸ライブラリーまたはその部分が固定化されている、固体の支持体（例えば、フローセルの内部）上で起きる。特定の配列決定法では、核酸ライブラリーを、フローセルに添加し、適切な条件下でのハイブリダイゼーションによりアンカーに固定化する。このタイプの核酸増幅をしばしば、固相増幅と呼ぶ。固相増幅の一部の実施形態では、全部または一部の増幅産物を、固定化されたプライマーから開始する伸長により合成する。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）のうちの少なくとも1つを固体の支持体上に固定化する点を除き、標準的な溶液相の増幅に類似する。

10

【0134】

一部の実施形態では、固相増幅は、表面に固定化された、1つの種のオリゴヌクレオチドプライマーのみを含む核酸増幅反応を含む。特定の実施形態では、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。一部の実施形態では、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むことができる。固定化されたプライマーまたは溶液に基づくプライマーの複数の異なる種を使用することができる。固相核酸増幅反応の非限定的な例として、界面増幅、ブリッジ増幅、エマルジョンPCR、WildFire増幅（例えば、米国特許公開第US20130012399号）等、またはそれらの組合せが挙げられる。

20

配列決定

【0135】

一部の実施形態では、核酸（例えば、核酸断片、試料核酸、無細胞核酸）の配列決定を行う。特定の実施形態では、完全または実質的に完全な配列を得、時には、部分的な配列を得る。

【0136】

一部の実施形態では、試料中の一部または全部の核酸を、配列決定の前または間に（例えば、非特異的に、例えば、PCRに基づく方法により）富化および／または増幅する。特定の実施形態では、試料中の特異的な、核酸の部分またはサブセットを、配列決定の前または間に富化および／または増幅する。一部の実施形態では、核酸のあらかじめ選択されたプールの部分またはサブセットの配列決定をランダムに行う。一部の実施形態では、配列決定の前または間に、試料中の核酸の富化および／または増幅を行わない。

30

【0137】

本明細書で使用する場合、「読取り」(reads) (すなわち、「読取り」(aread)、「配列の読取り」(asequenceread))は、本明細書に記載するまたは当技術分野で公知である、任意の配列決定の処理により生成された短いヌクレオチド配列である。読取りは、核酸断片の一方の末端から生成させることができ（「シングルエンドリード」）、時には、核酸の両方の末端から生成させる（例えば、ペアードエンドリード、ダブルエンドリード(double-endread)）。

40

【0138】

配列の読取りの長さはしばしば、特定の配列決定の技術と関連する。例えば、高スループット法は、塩基対(bp)のサイズが数十から数百まで様々であり得る配列の読取りを提供する。例えば、ナノポア配列決定は、塩基対のサイズが数十から数百または数千まで様々であり得る配列の読取りを提供することができる。一部の実施形態では、配列の読取りの平均値、中央値、平均の長さまたは絶対長が、約15bp～約900bp長である。特定の実施形態では、配列の読取りの平均値、中央値、平均の長さまたは絶対長が、約1

50

000bpまたはそれ超である。

【0139】

一部の実施形態では、シングルエンドリードの基準の、平均、平均値の長さまたは絶対長が、時には、約15個の連続ヌクレオチド～約50個もしくはそれ超の連続ヌクレオチド、約15個の連続ヌクレオチド～約40個もしくはそれ超の連続ヌクレオチドであり、時には、約15個の連続ヌクレオチド、または約36個もしくはそれ超の連続ヌクレオチドである。特定の実施形態では、シングルエンドリードの基準の、平均、平均値の長さまたは絶対長が、約20～約30塩基長、または約24～約28塩基長である。特定の実施形態では、シングルエンドリードの基準の、平均、平均値の長さまたは絶対長が、約1、約2、約3、約4、約5、約6、約7、約8、約9、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約21、約22、約23、約24、約25、約26、約27、約28、もしくは約29塩基長またはそれ超である。

10

【0140】

特定の実施形態では、ペアードエンドリードの基準の、平均、平均値の長さまたは絶対長が、時には、約10個の連続ヌクレオチド～約25個の連続ヌクレオチドもしくはそれ超（例えば、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約20、約21、約22、約23、約24もしくは約25ヌクレオチド長もしくはそれ超）、約15個の連続ヌクレオチド～約20個の連続ヌクレオチドもしくはそれ超であり、時には、約17個の連続ヌクレオチド、または約18個の連続ヌクレオチドである。

20

【0141】

読取りは一般に、ヌクレオチド配列の、物理的な核酸で示す表示である。例えば、ATGCと描写される配列を含有する読取りでは、物理的な核酸として、「A」はアデニンヌクレオチドを表示し、「T」はチミンヌクレオチドを表示し、「G」はグアニンヌクレオチドを表示し、「C」はシトシンヌクレオチドを表示する。妊娠中の雌の血液から得られた配列の読取りは、胎児核酸と母体核酸との混合物に由来する読取りであり得る。比較的短い読取りの混合物を、本明細書に記載する処理により、妊娠中の雌および／または胎児中に存在するゲノム核酸の表示に転換することができる。比較的短い読取りの混合物を、例えば、コピー数の変動（例えば、母体および／もしくは胎児のコピー数の変動）、遺伝子の変動、または異数性、微小重複もしくは微小欠失の表示に転換することができる。母体核酸と胎児核酸との混合物の読取りを、母体の染色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表示に転換することができる。特定の実施形態では、被験体から得られた試料の核酸配列の読取りを「得」、かつ／または1人もしくは複数の参照の人から得られた生物学的検体の核酸配列の読取りを「得る」には、核酸の配列決定を直接行って、配列情報を得ることを含むことができる。一部の実施形態では、「得る」は、他者が核酸から直接得た配列情報を受け取ることを含むことができる。

30

【0142】

一部の実施形態では、ゲノムの表示される割合が、配列決定され、時には、「カバレッジ」または「カバレッジ倍率」と呼ばれる。例えば、1倍のカバレッジは、ゲノムのヌクレオチド配列のおおよそ100%が、読取りにより表示されることを示す。一部の実施形態では、「カバレッジ倍率」は、参照としての以前の配列決定のランを参照して比較する用語である。例えば、第2の配列決定のランが、第1の配列決定のランのカバレッジの1/2である場合がある。一部の実施形態では、冗長性をもたせて、ゲノムの配列決定を行い、この場合、ゲノムの所与の領域を、2つもしくはそれ超の読取り、またはオーバーラップする読取りがカバーすることができる（例えば、1超の「カバレッジ倍率」、例えば、2倍のカバレッジ）。

40

【0143】

一部の実施形態では、1つの個体から得られた1つの核酸試料の配列決定を行う。特定の実施形態では、2つまたはそれ超の試料のそれぞれから得られた核酸の配列決定を行い

50

、この場合、試料は、1つの個体から得られるか、または異なる個体から得られる。特定の実施形態では、2つまたはそれ超の生物学的試料から得られた核酸試料をプールし、この場合、それぞれの生物学的試料が、1つの個体、または2つもしくはそれ超の個体から得られ、プールした試料の配列決定を行う。後者の実施形態では、それぞれの生物学的試料から得られた核酸試料をしばしば、1つまたは複数のユニークな識別子により同定する。

【0144】

一部の実施形態では、配列決定法は、配列決定の処理における配列決定反応 (sequence reaction) のマルチプレックス化を可能にする識別子を利用する。ユニークな識別子の数が多くなるほど、例えば、配列決定の処理においてマルチプレックス化することができる、検出される試料および／または染色体の数が増える。任意の適切な数 (例えば、4、8、12、24、48、96個またはそれ超) のユニークな識別子を使用して、配列決定の処理を行うことができる。

【0145】

配列決定の処理は、時には固相を使用し、固相は、時にはフローセルを含み、フローセルの上に、ライブラリーに由来する核酸をつなぐことができ、試薬を、流し、つなげた核酸と接触させることができる。フローセルは時には、フローセルのレーンを含み、識別子の使用により、それぞれのレーン中のいくつかの試料の分析を促進することができる。フローセルはしばしば、結合させた被検体を保持し、かつ／または結合させた被検体上を試薬溶液が順序正しく通過するのを可能にするように構成することができる固体の支持体である。フローセルは、多くの場合、平面形状であり、光学的に透明であり、一般に、ミリメートルのまたはミリメートルを下回るスケールであり、しばしば、チャンネルまたはレーンを有し、それらの中で、被検体と試薬との相互作用が発生する。一部の実施形態では、フローセルの所与のレーン中の分析される試料の数は、ライブラリーの調製および／またはプローブの設計の間に利用されるユニークな識別子の数に依存する。単一のフローセルのレーン。例えば、12個の識別子を使用するマルチプレックス化により、8レーンのフローセル中の (例えば、96ウェルのマイクロウェルプレート中のウェルの数に等しい) 96個の試料を同時に分析するのが可能になる。同様に、例えば、48個の識別子を使用するマルチプレックス化により、8レーンのフローセル中の (例えば、384ウェルのマイクロウェルプレート中のウェルの数に等しい) 384個の試料を同時に分析するのも可能になる。市販されている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、Illuminaのマルチプレックス化試料調製オリゴヌクレオチドキット、ならびにマルチプレックス化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。

【0146】

核酸の配列決定を行う任意の適切な方法を使用することができ、それらの非限定的な例として、Maxim & Gilbert、鎖停止法、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、顕微鏡法に基づく技法等、またはそれらの組合せが挙げられる。一部の実施形態では、本明細書に提供する方法では、第一世代の技術、例えば、サンガー配列決定法等 (これらとして、マイクロ流体サンガー配列決定を含めた、自動化サンガー配列決定法が挙げられる) を使用することができる。一部の実施形態では、核酸の撮像技術 (例えば、透過型電子顕微鏡法 (TEM) および原子間力顕微鏡法 (AFM)) の使用を含む配列決定の技術を使用することができる。一部の実施形態では、高スループット配列決定法を使用する。高スループット配列決定法は一般に、DNA鋳型または単一のDNA分子をクローン的に増幅させることを含み、これらのクローン増幅させた鋳型または分子の配列決定を、大規模に並行して、時にはフローセルの内部で行う。大規模に並行してDNAの配列決定を行うことが可能な次世代 (例えば、第2世代および第3世代) の配列決定の技法を、本明細書に記載する方法のために使用することができ、本明細書では、これらをまとめて「大規模並行配列決定」 (MPS) と呼ぶ。一部

の実施形態では、M P S 配列決定法は、標的化のアプローチを利用し、この場合、特定の染色体、遺伝子、または目的の領域の配列決定を行う。特定の実施形態では、標的化しないアプローチを使用し、この場合、ランダムに、試料中のほとんどまたは全ての核酸の配列決定を行い、それらを増幅し、かつ／または捕捉する。

【0147】

一部の実施形態では、富化、増幅および／または配列決定の標的化アプローチを使用する。標的化のアプローチはしばしば、試料中の核酸のサブセットを単離、選択および／または富化して、配列に特異的なオリゴヌクレオチドの使用によりさらなる処理を行う。一部の実施形態では、配列に特異的なオリゴヌクレオチドのライブラリーを利用して、試料中の核酸の1つまたは複数のセットを標的にする（例えば、それらにハイブリダイズさせる）。しばしば、配列に特異的なオリゴヌクレオチドおよび／またはプライマーは、目的の染色体、遺伝子、エクソン、イントロンおよび／または調節領域の1つまたは複数中に存在する特定の配列（例えば、ユニークな核酸配列）に選択的である。任意の適切な方法または方法の組合せを使用して、標的とされる核酸の1つまたは複数のサブセットの富化、増幅および／または配列決定を行うことができる。一部の実施形態では、標的とされる配列を、1つまたは複数の配列特異的なアンカーを使用して固相（例えば、フローセル、ビーズ）に捕捉することにより単離および／または富化する。一部の実施形態では、配列に特異的なプライマーおよび／またはプライマーセットを使用する、ポリメラーゼに基づく方法（例えば、任意の適切なポリメラーゼに基づく伸長による、P C Rに基づく方法）により、標的とされる配列を富化および／または増幅する。配列特異的なアンカーはしばしば、配列特異的なプライマーとして使用することができる。

【0148】

M P S 配列決定時には、合成による配列決定および特定の可視化処理を使用する。本明細書に記載する方法において使用することができる核酸の配列決定の技術は、合成による配列決定および可逆的なターミネーターに基づく配列決定（例えば、I l l u m i n a の 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 2 5 0 0 (I l l u m i n a , S a n D i e g o C A)) である。この技術を用いれば、数百万個の核酸（例えば、DNA）断片に対して、並行して配列決定を行うことができる。このタイプの配列決定の技術の1つの例では、8つの個々のレーンを有する光学的に透明なスライドを含有するフローセルを使用し、それらの表面上に、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）が結合している。フローセルはしばしば、結合させた被検体を保持し、かつ／または結合させた被検体上を試薬溶液が順序正しく通過するのを可能にするように構成することができる固体の支持体である。フローセルは、多くの場合、平面形状であり、光学的に透明であり、一般に、ミリメートルのまたはミリメートルを下回るスケールであり、しばしば、チャンネルまたはレーンを有し、それらの中で、被検体と試薬との相互作用が発生する。

【0149】

一部の実施形態では、合成による配列決定は、鋳型指向性の様式で、プライマーまたは既存の核酸鎖に、ヌクレオチドを反復して付加すること（例えば、共有結合による付加により）を含む。ヌクレオチドが反復付加される度に、検出を行い、核酸鎖の配列が得られるまで、この処理を複数回繰り返す。得られる配列の長さは一部分、実施される付加および検出のステップの数に依存する。合成による配列決定の一部の実施形態では、1ラウンドのヌクレオチド付加で、同じタイプ（例えば、A、G、CまたはT）の1、2、3つまたはそれ超のヌクレオチドを、付加し、検出する。ヌクレオチドは、任意の適切な方法により（例えば、酵素にまたは化学的に）付加することができる。例えば、一部の実施形態では、ポリメラーゼまたはリガーゼが、鋳型指向性の様式で、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一部の実施形態では、異なるタイプのヌクレオチド、ヌクレオチド類似体および／または識別子を使用する。一部の実施形態では、可逆的なターミネーターおよび／または除去可能（例えば、切断可能）な識別子を使用する。一部の実施形態では、蛍光標識されたヌクレオチドおよび／またはヌクレオチ

ド類似体を使用する。特定の実施形態では、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）ならびに／または洗浄ステップを含む。一部の実施形態では、1つまたは複数のヌクレオチドの付加を、本明細書に記載するまたは当技術分野で公知である適切な方法により検出し、それらの非限定的な例として、任意の適切な撮像装置、適切なカメラ、デジタルカメラ、CCD（チャージカップルデバイス）に基づく撮像装置（例えば、CCDカメラ）、CMOS（相補型金属酸化シリコン（Complementary Metal Oxide Silicon））に基づく撮像装置（例えば、CMOSカメラ）、光ダイオード（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）等、またはそれらの組合せが挙げられる。本明細書の方法を実施するために使用することができるその他の配列決定法には、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含まれる。

10

【0150】

本明細書の方法を実施するために使用することができるその他の配列決定法には、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含まれる。デジタルポリメラーゼ連鎖反応（デジタルPCRまたはdPCR）を使用して、試料中の核酸の同定および定量を直接行うことができる。一部の実施形態では、デジタルPCRを、エマルジョン中で行うことができる。例えば、個々の核酸を、例えば、マイクロ流体チャンバーデバイス中で分離し、それぞれの核酸を、PCRにより個々に増幅する。1個のウェル当たり1つ以下の核酸が存在するように核酸を分離することができる。一部の実施形態では、異なるプローブを使用して、種々の対立遺伝子（例えば、胎児の対立遺伝子と母体の対立遺伝子と）を区別することができる。対立遺伝子を数え上げて、コピー数を決定することができる。

20

【0151】

特定の実施形態では、ハイブリダイゼーションによる配列決定を使用することができる。この方法は、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ステップを含み、複数のポリヌクレオチドプローブのそれぞれを、基材に任意選択でつなぎ止めることができる。一部の実施形態では、基材は、一群の既知のヌクレオチド配列を有する平らな表面であり得る。アレイへのハイブリダイゼーションのパターンを使用して、試料中に存在するポリヌクレオチド配列を決定することができる。一部の実施形態では、それぞれのプローブを、ビーズ、例えば、磁性ビーズ等につなぎ止める。ビーズへのハイブリダイゼーションを同定し、試料内の複数のポリヌクレオチド配列を同定するために使用することができる。

30

【0152】

一部の実施形態では、本明細書に記載する方法において、ナノポア配列決定を使用することができる。ナノポア配列決定は、単一分子の配列決定の技術であり、それにより、単一の核酸分子（例えば、DNA）がナノポアを通過する度に、その配列を直接決定する。

【0153】

本明細書に記載する実施方法に適切なMPSの方法、システムまたは技術プラットフォームを使用して、核酸を配列決定した読取りを得ることができる。MPSプラットフォームの非限定的な例として、Illumina/Solex/HiSeq（例えば、IlluminaのGenome Analyzer；Genome Analyzer II；HISEQ2000；HISEQ）、SOLID、Roche/454、PACBIOおよび／またはSMRT、Helicos True Single Molecule Sequencing、Ion Torrentおよびイオン半導体に基づく配列決定（例えば、Life Technologiesが開発したもの）、WildFire、5500、5500xl Wおよび／または5500xl W Genetic Analyzerに基づく技術（例えば、Life Technologiesが開発し、販売するもの、米国特許公開第US20130012399号）；ポロニー配列決定、パイロシーケンシング、大規模並行シグネチャー配列決定（MPSS）、RNAポリメラーゼ（R

40

50

N A P) 配列決定、L a s e r G e n のシステムおよび方法、ナノポアに基づくプラットフォーム、化学感応性電界効果トランジスタ (C H E M F E T) アレイ、電子顕微鏡法に基づく配列決定 (例えば、Z S G e n e t i c s、H a l c y o n M o l e c u l a r が開発したもの)、ナノボール配列決定が挙げられる。

【0154】

一部の実施形態では、染色体に特異的な配列決定を行う。一部の実施形態では、D A N S R (選択された領域のデジタル分析) を利用して、染色体に特異的な配列決定を行う。選択された領域のデジタル分析を行うことによって、P C R 鋳型を形成するための、介在「ブリッジ」オリゴヌクレオチドを介する、2つの座位特異的オリゴヌクレオチドの c f D N A に依存するカテナーションにより、数百個の座位を同時に定量することが可能になる。一部の実施形態では、染色体に特異的な配列が富化されたライブラリーを生成することによって、染色体に特異的な配列決定を行う。一部の実施形態では、配列の読取りを、選択された一連の染色体についてのみ得る。一部の実施形態では、配列の読取りを、第 2 1、1 8 および 1 3 染色体についてのみ得る。

マッピングの読取り

【0155】

配列の読取りをマッピングすることができ、特定の核酸領域 (例えば、染色体、その部分またはセグメント) に対してマッピングする読取りの数を、カウントと呼ぶ。任意の適切なマッピングの方法 (例えば、処理、アルゴリズム、プログラム、ソフトウェア、モジュール等、またはそれらの組合せ) を使用することができる。下記に、マッピング処理の特定の態様を記載する。

【0156】

ヌクレオチド配列の読取り (すなわち、ゲノムの物理的な位置が不明である断片から得られた配列情報) のマッピングを、いくつかの方法で実施することができ、これはしばしば、得られた配列の読取りの、参照ゲノム中の一致する配列とのアラインメントを含む。そのようなアラインメントでは、配列の読取りを一般に、参照配列に対して整列させ、整列させた読取りを、「マッピング」されている、「マッピングされた配列の読取り」または「マッピングされた読取り」と呼ぶ。特定の実施形態では、マッピングされた配列の読取りを、「ヒット」または「カウント」と呼ぶ。一部の実施形態では、マッピングされた配列の読取りを、種々のパラメータに従って、一緒にしてグループ化し、特定の部分に割り当てが、これに関しては、下記にさらに詳細に論じる。

【0157】

本明細書で使用する場合、用語「整列させた (a l i g n e d)」、「アラインメント (a l i g n m e n t) 」または「整列する (a l i g n i n g) 」により、一致 (例えば、1 0 0 % 同一) または部分一致と同定され得る 2 つまたはそれ超の核酸配列について言及する。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、モジュールもしくはアルゴリズム) により行うことができ、それらの非限定的な例として、I l l u m i n a G e n o m i c s A n a l y s i s パイプラインの一部として流通されている E f f i c i e n t L o c a l A l i g n m e n t o f N u c l e o t i d e D a t a (E L A N D) コンピュータプログラムが挙げられる。配列の読取りのアラインメントは、1 0 0 % 配列一致であり得る。場合によっては、アラインメントは、1 0 0 % 配列一致よりも低い (すなわち、不完全一致、部分一致、部分アラインメント)。一部の実施形態では、アラインメントは、約 9 9 %、9 8 %、9 7 %、9 6 %、9 5 %、9 4 %、9 3 %、9 2 %、9 1 %、9 0 %、8 9 %、8 8 %、8 7 %、8 6 %、8 5 %、8 4 %、8 3 %、8 2 %、8 1 %、8 0 %、7 9 %、7 8 %、7 7 %、7 6 % または 7 5 % 一致である。一部の実施形態では、アラインメントは、不一致を含む。一部の実施形態では、アラインメントは、1、2、3、4 または 5 つの不一致を含む。2 つまたはそれ超の配列は、いずれかの鎖を使用して整列させることができる。特定の実施形態では、核酸配列を、別の核酸配列の逆相補体と整列させる。

【0158】

種々の計算方法を使用して、それぞれの配列の読取りをある部分に対してマッピングすることができる。配列を整列させるために使用することができるコンピュータアルゴリズムの非限定的な例として、BLAST、BLITZ、FASTA、BOWTIE1、BOWTIE2、ELAND、MAQ、PROBEMATCH、SOAPもしくはSEQMAP、またはそれらの変更形態もしくはそれらの組合せが挙げられるが、これらに限定されない。一部の実施形態では、配列の読取りを、参照ゲノム中の配列と整列させることができる。一部の実施形態では、配列の読取りを、例えば、GenBank、dbEST、dbSTS、EMBL (European Molecular Biology Laboratory) およびDDBJ (DNA Databank of Japan) を含めた、当技術分野で公知の核酸のデータベース中に見出し、かつ／またはそれらの中の配列と整列させることができる。BLASTまたは類似のツールを使用して、同定された配列を配列データベースに照らして検索することができる。次いで、例えば、(下記に記載するように) 検索ヒットを使用して、同定された配列を適切な部分へと選別することができる。

10

【0159】

一部の実施形態では、マッピングされた配列の読取りおよび／またはマッピングされた配列の読取りと関連する情報は、適切なコンピュータ可読フォーマットの非一時的なコンピュータ可読記憶媒体上に記憶させ、かつ／またはそこからアクセスされる。本明細書では、「コンピュータ可読フォーマット」は時には、大まかにフォーマットと呼ぶ。一部の実施形態では、マッピングされた配列の読取りは、適切なバイナリフォーマット、テキストフォーマット等またはそれらの組合せで記憶させ、かつ／またはアクセスされる。バイナリフォーマットは時には、BAMフォーマットである。テキストフォーマットは時には、配列アラインメント／マップ (SAM) フォーマットである。バイナリフォーマットおよび／またはテキストフォーマットの非限定的な例として、BAM、SAM、SRF、FASTQ、Gzip等、またはそれらの組合せが挙げられる。一部の実施形態では、マッピングされた配列の読取りは、従来のフォーマット (例えば、SAMフォーマットまたはBAMフォーマット) よりも少ない記憶空間 (例えば、より少ないバイト) を必要とするフォーマットで記憶させ、かつ／またはそれに変換される。一部の実施形態では、第1のフォーマットのマッピングされた配列の読取りは、第1のフォーマットよりも少ない記憶空間を必要とする第2のフォーマットに圧縮される。用語「圧縮される」は、本明細書で使用する場合、コンピュータ可読データファイルのサイズを低下させる、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処理を指す。一部の実施形態では、マッピングされた配列の読取りは、バイナリフォーマットのSAMフォーマットから圧縮される。ファイルを圧縮すると、いくつかのデータが時には失われる。時には、圧縮処理でデータは失われない。ファイル圧縮の一部の実施形態では、いくつかのデータは、マッピングされた配列の読取りに関する情報を含む別のデータファイルへのインデックスおよび／またはリファレンスで置き換えられる。一部の実施形態では、マッピングされた配列の読取りを、読取りのカウント、(例えば、読取りがマッピングされる染色体を識別する) 染色体の識別子、および(例えば、読取りがマッピングされる染色体上の位置を識別する) 染色体位置の識別子を含むまたはそれらからなるバイナリフォーマットで記憶させる。一部の実施形態では、バイナリフォーマットは、20バイト配列、16バイト配列、8バイト配列、4バイト配列または2バイト配列を含む。一部の実施形態では、マッピングされた読取り情報を、10バイトフォーマット、9バイトフォーマット、8バイトフォーマット、7バイトフォーマット、6バイトフォーマット、5バイトフォーマット、4バイトフォーマット、3バイトフォーマットまたは2バイトフォーマットの配列で記憶させる。時には、マッピングされた読取りデータを、5バイトフォーマットを含む4バイト配列で記憶させる。一部の実施形態では、バイナリフォーマットは、1バイトの染色体の順序数および4バイトの染色体の位置を含む5バイトのフォーマットを含む。一部の実施形態では、マッピングされた読取りを、配列アラインメント／マップ (SAM) フォーマットの約1／100、約1／90、約1／80、約1／70、約1／60、約1／55、約1／50

20

30

40

50

、約 1 / 4 5、約 1 / 4 0 または約 1 / 3 0 である圧縮されたバイナリフォーマットで記憶させる。一部の実施形態では、マッピングされた読取りを、G Z i p フォーマットの約 1 / 2 ~ 約 1 / 5 0 (例えば、約 1 / 3 0、1 / 2 5、1 / 2 0、1 / 1 9、1 / 1 8、1 / 1 7、1 / 1 6、1 / 1 5、1 / 1 4、1 / 1 3、1 / 1 2、1 / 1 1、1 / 1 0、1 / 9、1 / 8、1 / 7、1 / 6 または約 1 / 5) である圧縮バイナリフォーマットで記憶させる。

【0160】

一部の実施形態では、システムは、圧縮モジュールを含む(例えば、図 4 2 A の 4)。一部の実施形態では、コンピュータ可読フォーマットの非一時的なコンピュータ可読記憶媒体上に記憶させたマッピングされた配列の読取り情報を、圧縮モジュールにより圧縮する。圧縮モジュールは時には、マッピングされた配列の読取りを、適切なフォーマットに変換したり、適切なフォーマットから変換したりする。一部の実施形態では、圧縮モジュールは、第 1 のフォーマットのマッピングされた配列の読取りを受け取り(例えば、図 4 2 A の 1)、これらを圧縮されたフォーマット(例えば、バイナリフォーマット、5)に変換し、圧縮された読取りを別のモジュール(例えば、偏り密度モジュール、6)に移すことができる。圧縮モジュールはしばしば、配列の読取りをバイナリフォーマット、5 (例えば、B R e a d s フォーマット)で提供する。圧縮モジュールの非限定的な例として、G Z I P、B G Z F および B A M 等、またはそれらの改変形態が挙げられる。

10

【0161】

以下に、j a v a (登録商標)を使用する、整数の 4 バイト配列への変換の例を示す。

20

【化 1】

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

30

【0162】

一部の実施形態では、読取りを、参照ゲノム中の部分に対してユニークまたは非ユニークにマッピングすることができる。参照ゲノム中の単一配列との整列の場合であれば、読取りは、「ユニークにマッピングされる」とみなされる。参照ゲノム中の 2 つまたはそれ超の配列との整列の場合であれば、読取りは、「非ユニークにマッピングされる」とみなされる。一部の実施形態では、非ユニークにマッピングされた読取りは、さらなる分析(例えば、定量)から排除される。特定の実施形態では、特定の、低い程度の不一致(0 ~ 1 つ)は、参照ゲノムと、マッピングされている、個々の試料から得られた読取りとの間に存在し得る一塩基多型であると説明することができる場合がある。一部の実施形態では、参照配列に対してマッピングされる読取りには、いかなる程度の不一致も許されない。

40

【0163】

本明細書で使用する場合、用語「参照ゲノム」は、部分であれ、完全であれ、任意の生物またはウイルスの任意の特定の公知の配列決定されたまたは特徴付けられたゲノムであって、被験体由来の同定された配列を照会するために使用することができるゲノムを指すことができる。例えば、ヒト被験体および多くのその他の生物のために使用する参照ゲノムを、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

50

けるNational Center for Biotechnology Informationにおいて見出すことができる。「ゲノム」は、核酸配列として表される、生物またはウイルスの完全な遺伝情報を指す。本明細書で使用する場合、参照配列または参照ゲノムはしばしば、1つの個体または複数の個体から得られた、アセンブルしたまたは部分的にアセンブルしたゲノム配列である。一部の実施形態では、参照ゲノムは、1つまたは複数のヒト個体から得られた、アセンブルしたまたは部分的にアセンブルしたゲノム配列である。一部の実施形態では、参照ゲノムは、染色体に割り当てられた配列を含む。

【0164】

特定の実施形態では、試料核酸が妊娠中の雌に由来する場合、参照配列が時には、胎児にも、胎児の母親にも、胎児の父親にも由来せず、これを本明細書では「外部参照」と呼ぶ。一部の実施形態では、母体の参照を調製し、使用することができる。外部参照に基づいて、妊娠中の雌からの参照（「母体の参照配列」）を調製する場合、胎児のDNAを実質的に含有しない、妊娠中の雌のDNAから得られた読取りをしばしば、外部参照配列に対してマッピングし、アセンブルする。特定の実施形態では、外部参照は、妊娠中の雌と実質的に同じ民族性を有する個体のDNAに由来する。母体の参照配列は、母体のゲノムDNAを完全にはカバーしない場合があり（例えば、母体のゲノムDNAの約50%、60%、70%、80%、90%またはそれ超をカバーする場合がある）、母体の参照は、母体のゲノムDNA配列と完全には一致しない場合がある（例えば、母体の参照配列は、複数の不一致を含む場合がある）。

【0165】

特定の実施形態では、マッピング可能性を、ゲノム領域（例えば、部分、ゲノム部分、部分）について評価する。マッピング可能性は、ヌクレオチド配列の読取りを、参照ゲノムのある部分に対して、典型的には、例えば、0、1、2つまたはそれ超の不一致を含めた、特定の数の不一致が存在するだけで、明確に整列させることができることである。所与のゲノム領域について、事前にセットされた、読取りの長さのスライディングウィンドウのアプローチを使用し、得られた、読取りレベルのマッピング可能性の値を平均化して、予想されるマッピング可能性を推定することができる。ユニークなヌクレオチド配列のストレッチを含むゲノム領域が時には、高いマッピング可能性の値を有する。

部分

【0166】

一部の実施形態では、マッピングされる配列の読取り（すなわち、配列タグ）を、種々のパラメータに従って、一緒にしてグループ化し、特定の部分（例えば、参照ゲノムの部分）に割り当てる。しばしば、個々のマッピングされる配列の読取りを使用して、試料中に存在する、ある部分（例えば、ある部分の存在、不在または量）を同定することができる。一部の実施形態では、部分の量は、試料中のより大きな配列（例えば、染色体）の量を示す。用語「部分」はまた、本明細書では、「ゲノム区分」、「ピン」、「領域」、「区画」、「参照ゲノムの部分」、「染色体の部分」または「ゲノム部分」と呼ぶこともできる。一部の実施形態では、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、複数の染色体に広がるセグメント、複数の、染色体のセグメント、および／またはそれらの組合せである。一部の実施形態では、部分は、特定のパラメータに基づいてあらかじめ定義される。一部の実施形態では、部分は、ゲノムの区分化（例えば、サイズ、GC含有量、連続した領域、恣意的に定義されたサイズの連続した領域等による区分化）に基づいて恣意的に定義される。

【0167】

一部の実施形態では、部分は、例えば、配列の長さまたは1つもしくは複数の特定の特徴を含む、1つまたは複数のパラメータに基づいて描写される。当技術分野で公知であるまたは本明細書に記載する任意の適切な判定基準を使用して、部分は、選択し、フィルタリングし、かつ／または検討事項から除去することができる。一部の実施形態では、部分は、ゲノム配列の特定の長さに基づく。一部の実施形態では、方法は、複数の部分に対し

てマッピングされた、複数の配列の読取りの分析を含むことができる。部分はおよそ同じ長さであってもよく、または部分は異なる長さであってもよい。一部の実施形態では、部分は、ほぼ等しい長さのものである。一部の実施形態では、異なる長さの部分を調整する、またはそれらに重み付けする。一部の実施形態では、部分は、約10キロベース(kb)～約100kb、約20kb～約80kb、約30kb～約70kb、約40kb～約60kb、時には、約50kbである。一部の実施形態では、部分は、約10kb～約20kbである。部分は、配列の連続するランに限定されない。したがって、部分は、連続するおよび／または連続し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。一部の実施形態では、部分は、1つの染色体の全部もしくは一部、または2つもしくはそれ超の染色体の全部もしくは一部を含む。一部の実施形態では、部分は、1、2つまたはそれ超の染色体全体に広がり得る。さらに、部分は、複数の染色体のつながっているまたは離れた領域にも広がり得る。

10

【0168】

一部の実施形態では、部分は、目的の染色体、例えば、遺伝子の変動（例えば、第13、18および／もしくは21染色体、または性染色体の異数性）を評価する染色体等における特定の染色体のセグメントであり得る。部分はまた、病原体のゲノム（例えば、細菌の、真菌の、もしくはウイルスの）、またはその断片であり得る。部分は、遺伝子、遺伝子の断片、調節配列、イントロン、エクソン等であり得る。

【0169】

一部の実施形態では、ゲノム（例えば、ヒトゲノム）を、特定の領域の情報内容に基づいて、部分に区分化する。一部の実施形態では、ゲノムの区分化は、ゲノムにわたって類似の領域（例えば、同一または相同な領域または配列）を排除し、ユニークな領域のみを保持することができる。区分化する間に除去される領域は、単一の染色体内にある場合または複数の染色体に広がる場合がある。一部の実施形態では、区分化されたゲノムを、より迅速なアラインメントのために、切り詰め、最適化して、しばしば、ユニークに同定することが可能な配列に焦点を当てるのを可能にする。

20

【0170】

一部の実施形態では、区分化して、類似の領域の重み付けを減らすことができる。下記に、部分の重み付けを減らすための処理について、さらに詳細に論じる。

【0171】

30

一部の実施形態では、染色体の範囲を超える領域へのゲノムの区分化は、分類の状況で生成した情報のゲインに基づいて行うことができる。例えば、正常と確認された被験体群と異常と確認された被験体群と（例えば、それぞれ、正倍数体の被験体とトリソミーの被験体と）を区別するための特定のゲノムの場所の有意性を測定するp値プロファイルを使用して、情報内容を定量することができる。一部の実施形態では、例えば、タグを整列させる間のスピード／利便性、GC含有量（例えば、高いもしくは低いGC含有量）、GC含有量の様性、配列の含有量のその他の尺度（例えば、個々のヌクレオチドの割合、ピリミジンもしくはプリン割合、天然核酸対非天然核酸の割合、メチル化ヌクレオチドの割合、およびCpG含有量）、メチル化状況、二重鎖の融解温度、配列決定もしくはPCRへの適用可能性、参照ゲノムの個々の部分に割り当てられた不確実性の値、ならびに／または特定の特徴を標的とする検索等の任意のその他の判定基準に基づいて、染色体の範囲を超える領域へのゲノムの区分化を行うことができる。

40

【0172】

染色体の「セグメント」は、一般に染色体の一部分であり、典型的には部分とは異なる染色体の一部分である。染色体のセグメントは、時には部分とは異なる染色体の領域中にあり、時には部分とはポリヌクレオチドを共有せず、時には部分中にあるポリヌクレオチドを含む。染色体のセグメントは、しばしば部分よりも大きな数のヌクレオチドを含有し（例えば、セグメントは、時には部分を含む）、染色体のセグメントは、時には部分よりも小さな数のヌクレオチドを含有する（例えば、セグメントは、時には部分内にある）。
カウント

50

【0173】

一部の実施形態では、選択された特徴または変数に基づいてマッピングまたは区分化される配列の読取りを定量して、1つまたは複数の部分（例えば、参照ゲノムの部分）に対してマッピングされる読取りの数を決定することができる。特定の実施形態では、部分に対してマッピングされる配列の読取りの分量をカウントと呼ぶ（例えば、1カウント）。しばしば、カウントを、部分と関連付ける。特定の実施形態では、2つまたはそれ超の部分（例えば、一連の部分）についてのカウントは、数学的に操作される（例えば、平均化、加算、正規化等、またはそれらの組合せ）。一部の実施形態では、カウントは、部分に対してマッピングされる（すなわち、部分と関連付けられる）配列の読取りの一部または全部から決定される。特定の実施形態では、カウントは、マッピングされた配列の読取りのあらかじめ定義されたサブセットから決定される。任意の適切な特徴または変数を利用して、マッピングされる配列の読取りのあらかじめ定義されるサブセットを定義または選択することができる。一部の実施形態では、マッピングされる配列の読取りのあらかじめ定義されたサブセットは、1～n個の配列の読取りを含むことができ、ここで、nは、試験被験体または参照被験体の試料から生成された全ての配列の読取りの合計に等しい数を表わす。

10

【0174】

特定の実施形態では、カウントは、当技術分野で公知の適切な方法、演算または数学的处理により処理または操作される配列の読取りから誘導される。カウント（a count）（例えば、カウント（counts））は、適切な方法、演算または数学的处理により決定することができる。特定の実施形態では、カウントを、ある部分と関連付けた配列の読取りから誘導し、この場合、配列の読取りの一部または全部に対して、重み付け、除去、フィルタリングすること、正規化、調整、平均化、平均値として誘導すること、加算もしくは減算、またはそれらの組合せによる処理が行われる。一部の実施形態では、カウントを、未加工の配列の読取りおよび／またはフィルタリングした配列の読取りから誘導する。特定の実施形態では、カウントの値を、数学的处理により決定する。特定の実施形態では、カウントの値は、ある部分に対してマッピングされた配列の読取りの平均、平均値または合計である。しばしば、カウントは、カウントの平均の数である。一部の実施形態では、カウントは、不確実性の値と関連付けられる。

20

【0175】

一部の実施形態では、カウントを操作または転換することができる（例えば、正規化する、組み合わせる、加算する、フィルタリングする、選択する、平均化する、平均値として誘導する等、またはそれらの組合せ）。一部の実施形態では、カウントを転換して、正規化したカウントを生成することができる。当技術分野で公知の方法により、かつ／または本明細書の記載のとおり（例えば、部分に関する（portion-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、GC LOESS、LOWESS、PERUN、ChAI、RM、GCRM、cQn、および／またはそれらの組合せにより）、カウントを処理する（例えば、正規化する）ことができる。

30

【0176】

カウント（例えば、未加工の、フィルタリングした、および／または正規化したカウント）を、1つまたは複数のレベルに対して処理し、正規化することができる。下記に、レベルおよびプロファイルについてより詳細に記載する。特定の実施形態では、カウントを、参照レベルに対して処理し、かつ／または正規化することができる。本明細書では後に、参照レベルについて扱う。レベルに従って処理したカウント（例えば、処理したカウント）を、不確実性の値（例えば、計算した分散、誤差、標準偏差、Zスコア、p値、平均絶対偏差（mean absolute deviation）等）と関連付けることができる。一部の実施形態では、不確実性の値が、あるレベルを上回る範囲および下回る範囲を定義する。偏差についての値を、不確実性の値の代わりに使用することができ、偏差の尺度の非限定的な例として、標準偏差、平均絶対偏差（average absolute deviation）、絶対偏差の中央値、標準スコア（例えば、Zスコア、Zス

40

50

コア、正常スコア、標準化した変数）等が挙げられる。

【0177】

カウントはしばしば、胎児を出産する妊娠中の雌に由来する核酸試料から得られる。1つまたは複数の部分に対してマッピングされた核酸配列の読取りのカウントはしばしば、胎児および胎児の母親（例えば、妊娠中の雌の被験体）の両方を表示するカウントである。特定の実施形態では、ある部分に対してマッピングされたカウントの一部は、胎児のゲノムに由来し、同じ部分に対してマッピングされたカウントの一部は、母体のゲノムに由来する。

データの処理および正規化

【0178】

本明細書では、計数されるに至った、マッピングされた配列の読取りを、未加工データと呼び、その理由は、これらのデータが、操作されていないカウント（例えば、未加工カウント）を表示するからである。一部の実施形態では、データセット中の配列の読取りのデータを、さらに処理し（例えば、数学的および／もしくは統計学的に操作し）、かつ／または示して、アウトカムの提供を促進することができる。特定の実施形態では、より大きなデータセットを含めて、データセットは、さらなる分析を促進するために、前処理が役立つ場合がある。データセットの前処理は時には、余分の、かつ／または有益でない部分または参照ゲノムの部分（例えば、有益でないデータを有する参照ゲノムの部分、余分の、マッピングされた読取り、カウントの中央値がゼロである部分、大きな比率を占めるまたは少ない比率を占める配列）の除去を含む。理論により制限されることなく、データの処理および／または前処理は、（i）ノイズの多いデータ（noisy data）を除去し、（ii）有益でないデータを除去し、（iii）余分のデータを除去し、（iv）より大きなデータセットの複雑性を低下させ、かつ／または（v）1つの形態から1つもしくは複数のその他の形態へのデータの転換を促進することができる。本明細書では、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用いる場合には、まとめて「処理」と呼ぶ。処理は、データをさらなる分析に、より適用可能にすることができ、一部の実施形態では、アウトカムをもたらすことができる。一部の実施形態では、1つまたは複数または全ての処理方法（例えば、正規化の方法、部分をフィルタリングすること、マッピング、検証等、またはそれらの組合せ）が、メモリと併せたプロセッサ、マイクロプロセッサ、コンピュータにより、かつ／またはマイクロプロセッサが制御する装置により行われる。

【0179】

用語「ノイズの多いデータ」は、本明細書で使用する場合、（a）分析またはプロットした場合にデータ点間に有意な分散を示すデータ、（b）有意な標準偏差を有する（例えば、3標準偏差よりも大きい）データ、（c）平均値の有意な標準誤差を有するデータ等、および上記の組合せを指す。ノイズの多いデータは、時には出発材料（例えば、核酸試料）の分量および／または品質に起因して発生し、時には配列の読取りを生成するために使用するDNAを調製または複製するための処理の一部から発生する。特定の実施形態では、ノイズは、PCRに基づく方法を使用して調製する場合の、大きな比率を占める特定の配列から生じる。本明細書に記載する方法は、ノイズの多いデータの寄与を低下させるまたは排除することができ、したがって、ノイズの多いデータの、提供されるアウトカムに対する作用を低下させる。

【0180】

用語「有益でないデータ」、「有益でない、参照ゲノムの部分」、および「有益でない部分」は、本明細書で使用する場合、所定の閾値とは有意に異なる数値、または値のあらかじめ定義された値の限界範囲の外側に存在する数値を有する部分、またはそこから誘導されたデータを指す。用語「閾」および「閾値」は、本明細書では、適格なデータセットを使用して計算される任意の数を指し、遺伝子の変動（例えば、コピー数の変動、異数性、微小重複（microduplication）、微小欠失、染色体異常等）の診断の限界として役立つ。特定の実施形態では、本明細書に記載する方法により得られた結果が

閾を上回り、被験体が、遺伝子の変動（例えば、21トリソミー）を有すると診断される。一部の実施形態では、閾値または値の範囲はしばしば、（例えば、参照および／または被験体から得られた）配列の読取りのデータを数学的および／または統計学的に操作することによって計算され、特定の実施形態では、閾値または値の範囲を生成するために操作される配列の読取りのデータは、（例えば、参照および／または被験体から得られた）配列の読取りのデータである。一部の実施形態では、不確実性の値を決定する。不確実性の値は、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適切な尺度であり得る。一部の実施形態では、不確実性の値は、標準偏差、標準誤差、計算した分散、p値または平均絶対偏差（MAD）である。一部の実施形態では、不確実性の値を、実施例4の方式に従って計算することができる。

10

【0181】

本明細書に記載するデータセットを処理するために、任意の適切な手順を利用することができる。データセットを処理するために使用するのに適切な手順の非限定的な例として、フィルタリングすること、正規化すること、重み付けすること、ピークの高さをモニタリングすること、ピークの面積をモニタリングすること、ピークのエッジをモニタリングすること、面積比を決定すること、データを数学的に処理すること、データを統計学的に処理すること、統計学的アルゴリズムを適用すること、固定変数を用いて分析すること、最適化された変数を用いて分析すること、データをプロットし、パターンまたは傾向を確認して、さらなる処理を行うこと等、および上記の組合せが挙げられる。一部の実施形態では、種々の特徴（例えば、GC含有量、余分の、マッピングされた読取り、セントロメア領域、テロメア領域等、およびそれらの組合せ）、ならびに／または変数（例えば、胎児の性別、母体の年齢、母体の倍数性、胎児核酸のパーセント寄与等、またはそれらの組合せ）に基づいて、データセットは処理される。特定の実施形態では、本明細書の記載のとおりデータセットを処理することによって、大きいおよび／または複雑なデータセットの複雑性および／または次元性を低下させることができる。複雑なデータセットの非限定的な例として、年齢および民族性の背景が異なる1つまたは複数の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の読取りのデータが挙げられる。一部の実施形態では、データセットは、それぞれの試験被験体および／または参照被験体について、数千～数百万個の配列の読取りを含むことができる。

20

【0182】

特定の実施形態では、データ処理を、任意の数のステップで行うことができる。例えば、一部の実施形態では、単一の処理手順のみを使用して、データを処理することができる。特定の実施形態では、1つもしくは複数、5つもしくはそれ超、10個もしくはそれ超、または20個もしくはそれ超の処理ステップ（例えば、1つもしくは複数の処理ステップ、2つもしくはそれ超の処理ステップ、3つもしくはそれ超の処理ステップ、4つもしくはそれ超の処理ステップ、5つもしくはそれ超の処理ステップ、6つもしくはそれ超の処理ステップ、7つもしくはそれ超の処理ステップ、8つもしくはそれ超の処理ステップ、9つもしくはそれ超の処理ステップ、10個もしくはそれ超の処理ステップ、11個もしくはそれ超の処理ステップ、12個もしくはそれ超の処理ステップ、13個もしくはそれ超の処理ステップ、14個もしくはそれ超の処理ステップ、15個もしくはそれ超の処理ステップ、16個もしくはそれ超の処理ステップ、17個もしくはそれ超の処理ステップ、18個もしくはそれ超の処理ステップ、19個もしくはそれ超の処理ステップ、または20個もしくはそれ超の処理ステップ）を使用して、データを処理することができる。一部の実施形態では、処理ステップは、2回またはそれ超の回数繰り返される同じステップであり得（例えば、2回またはそれ超の回数フィルタリングする、2回またはそれ超の回数正規化する）、特定の実施形態では、処理ステップは、同時または順次に行われる2つまたはそれ超の異なる処理ステップであり得る（例えば、フィルタリングし、正規化する；正規化し、ピークの高さおよびエッジをモニタリングする；フィルタリングし、正規化し、参照に対して正規化し、統計学的に操作して、p値を決定する等）。一部の実施形態では、同じまたは異なる処理ステップの任意の適切な数および／または組合せを利用し、

30

40

50

配列の読取りのデータを処理して、アウトカムの提供を促進することができる。特定の実施形態では、本明細書に記載する判定基準によりデータセットを処理することによって、データセットの複雑性および／または次元性を低下させることができる。

【0183】

一部の実施形態では、1つまたは複数の処理ステップは、1つまたは複数のフィルタリングステップを含むことができる。用語「フィルタリングする」は、本明細書で使用する場合、部分または参照ゲノムの部分を検討事項から除去することを指す。これらに限定されないが、余分のデータ（例えば、余分なまたはオーバーラップする、マッピングされた読取り）、有益でないデータ（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）、大きな比率を占めるもしくは少ない比率を占める配列を有する参照ゲノムの部分、ノイズの多いデータ等、または上記の組合せを含めた、任意の適切な判定基準に基づいて、参照ゲノムの部分を選択して、除去することができる。フィルタリング処理はしばしば、参照ゲノムの1つまたは複数の部分を検討事項から除去すること、および、除去するために選択された参照ゲノムの1つまたは複数の部分におけるカウントを、参照ゲノム、1つもしくは複数の染色体、または検討下のゲノムの部分について計数したかまたは合計したカウントから減算することを含む。一部の実施形態では、参照ゲノムの部分を、逐次的に除去する（例えば、1つずつ除去して、それぞれの個々の部分の除去の作用の評価を可能にする）ことができ、特定の実施形態では、除去するためにマークされた、参照ゲノムの部分全てを、同時に除去することができる。一部の実施形態では、特定のレベルを上回るまたは下回る分散により特徴付けられた参照ゲノムの部分を除去し、本明細書では、これを時には、参照ゲノムの「ノイズの多い」部分をフィルタリングすると呼ぶ。特定の実施形態では、フィルタリング処理は、部分、染色体または染色体のセグメントの平均プロファイルレベルから、プロファイルの分散の所定の倍数だけ逸脱するデータ点を、データセットから得ることを含み、特定の実施形態では、フィルタリング処理は、部分、染色体または染色体のセグメントの平均プロファイルレベルから、プロファイルの分散の所定の倍数だけ逸脱しないデータ点を、データセットから除去することを含む。一部の実施形態では、フィルタリング処理を利用して、遺伝子の変動の存在または非存在について分析する、参照ゲノムの候補となる部分の数を低下させる。遺伝子の変動（例えば、微小欠失、微小重複）の存在または非存在について分析する、参照ゲノムの候補となる部分の数を低下させることによって、しばしばデータセットの複雑性および／または次元性を低下させ、時には遺伝子変動および／または遺伝子異常の検索および／または同定のスピードを2桁またはそれ超だけ増加させる。

【0184】

一部の実施形態では、1つまたは複数の処理ステップは、1つまたは複数の正規化ステップを含むことができる。正規化は、本明細書に記載するまたは当技術分野で公知である適切な方法により行うことができる。特定の実施形態では、正規化は、異なるスケールで測定された値を、概念的に（*notionally*）共通のスケールに調整することを含む。特定の実施形態では、正規化は、調整された値の確率分布をアラインメントに至らせるための高度な数学的調整を含む。一部の実施形態では、正規化は、分布を正規分布にそろえることを含む。特定の実施形態では、正規化は、特定の全体的な影響（例えば、誤差および異常）の作用を排除する方法で、異なるデータセットについて正規化した対応する値を比較するのを可能にする数学的調整を含む。特定の実施形態では、正規化は、スケールリングを含む。正規化は時には、所定の変数または式による1つまたは複数のデータセットの除算を含む。正規化の方法の非限定的な例として、部分に関する正規化、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（局所重み付け散布図平坦化）、PERUN、ChAI、リポートマスクキング（RM）、GC正規化およびリポートマスクキング（GCRM）、cQn、ならびに／またはそれらの組合せが挙げられる。一部の実施形態では、遺伝子の変動の存在または非存在（例えば、異数性、微小重複、微小欠失）の決定は、正規化の方法（例えば、部分に関する正規化、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

10

20

30

40

50

LOESS、LOWESS（局所重み付け散布図平坦化）、PERUN、CHAI、リピートマスクキング（RM）、GC正規化およびリピートマスクキング（GCRM）、cQn、当技術分野で公知の正規化の方法、ならびに／またはそれらの組合せ）を利用する。

【0185】

任意の適切な数の正規化を使用することができる。一部の実施形態では、データセットを、1回もしくは複数回、5回もしくはそれ超の回数、10回もしくはそれ超の回数、または20回またはそれ超の回数さえ正規化することができる。データセットを、任意の適切な特徴または変数（例えば、試料データ、参照データ、または両方）を表示する値（例えば、正規化する値）に対して正規化することができる。使用することができるデータの正規化のタイプの非限定的な例として、1つまたは複数の選択された試験部分または参照部分についての未加工カウントデータを、その上で、選択された部分または区分がマッピングされる染色体またはゲノム全体に対してマッピングされるカウントの総数に対して正規化すること；1つまたは複数の選択された部分についての未加工カウントデータを、その上で、選択された部分またはセグメントがマッピングされる1つもしくは複数の部分または染色体についての参照のカウントの中央値に対して正規化すること；未加工カウントデータを、あらかじめ正規化されたデータまたはそれらの誘導値に対して正規化すること；および先に正規化されたデータを、1つまたは複数のその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時には、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、統計学的誤差を単離する作用を有する。また、データセットの正規化は時には、異なるスケールを有するデータのデータとしての特徴の比較を、データに共通のスケール（例えば、所定の正規化変数）を与えることによって可能にする。一部の実施形態では、統計学的に誘導された値に対する1回または複数回の正規化を利用して、データの差を最小化し、外れたデータの重要性を減少させることができる。部分または参照ゲノムの部分を正規化する値に関して正規化することを時には、「部分に関する正規化」と呼ぶ。

【0186】

特定の実施形態では、正規化を含む処理ステップは、静止したウィンドウに対して正規化することを含み、一部の実施形態では、正規化を含む処理ステップは、移動するウィンドウまたはスライディングウィンドウに対して正規化することを含む。用語「ウィンドウ」は、本明細書で使用する場合、分析のために選ばれた1つまたは複数の部分を指し、時には、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／またはその他の数学的もしくは統計学的な操作のために使用される）。用語「静止したウィンドウに対して正規化する」は、本明細書で使用する場合、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を比較するために選択された1つまたは複数の部分を使用する正規化の処理を指す。一部の実施形態では、選択された部分を利用して、プロファイルを生成する。静止したウィンドウは一般に、操作および／または分析の間に変化しない所定の一連の部分を含む。用語「移動するウィンドウに対して正規化する」および「スライディングウィンドウに対して正規化する」は、本明細書で使用する場合、選択された試験部分のゲノム領域に限局される部分（例えば、遺伝子の直近の周囲の、隣接する部分または区分等）に対して行われる正規化を指し、この場合、1つまたは複数の選択された試験部分は、選択された試験部分の直近の周囲の部分に対して正規化される。特定の実施形態では、選択された部分を利用して、プロファイルを生成する。スライディングウィンドウまたは移動するウィンドウの正規化はしばしば、隣接する試験部分に向けて繰り返し移動またはスライディングさせ、新たに選択された試験部分を、新たに選択された試験部分の直近の周囲の部分または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み、この場合、隣接するウィンドウは、共通する1つまたは複数の部分を有する。特定の実施形態では、複数の選択された試験部分および／または染色体を、スライディングウィンドウ処理により分析することができる。

【0187】

一部の実施形態では、スライディングウィンドウまたは移動するウィンドウに対して正規化することによって、1つまたは複数の値を生成することができ、この場合、それぞれ値は、ゲノムの異なる領域（例えば、染色体）から選択された異なる一連の参照部分に対する正規化を表示する。特定の実施形態では、生成された1つまたは複数の値は、累積合計（例えば、選択された部分、ドメイン（例えば、染色体の一部）または染色体にわたり正規化されたカウントプロファイルの積分の数的な推定値）である。スライディングウィンドウまたは移動するウィンドウの処理により生成された値を使用して、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ことができる。一部の実施形態では、1つまたは複数の部分の累積合計を、ゲノムの位置の関数として示すことができる。時には、移動するウィンドウまたはスライディングウィンドウの分析を使用して、ゲノムを微小欠失および／または微小挿入の存在または非存在について分析する。特定の実施形態では、1つまたは複数の部分の累積合計を示すことを使用して、遺伝子の変動（例えば、微小欠失、微小重複）の領域の存在または非存在を同定する。一部の実施形態では、移動するウィンドウまたはスライディングウィンドウの分析を使用して、微小欠失を含有するゲノム領域を同定し、特定の実施形態では、移動するウィンドウまたはスライディングウィンドウの分析を使用して、微小重複を含有するゲノム領域を同定する。

10

【0188】

下記に、利用することができる正規化の処理の特定の例、例えば、LOESS、PER UN、ChAI および主成分正規化の方法等をより詳細に記載する。

【0189】

20

一部の実施形態では、処理ステップは、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「重み付け関数」、またはそれらの文法上の派生語もしくは相当語句は、本明細書で使用する場合、特定のデータセットの特徴または変数の影響を、その他のデータセットの特徴または変数に関して変化させる（例えば、1つもしくは複数の部分または参照ゲノムの部分中に含有されるデータの有意性および／または寄与を、参照ゲノムの選択された1つまたは複数の部分中のデータの品質または有用性に基づいて増加または減少させる）ために利用するデータセットの一部または全部の数学的操作を指す。一部の実施形態では、重み付け関数を使用して、比較的小さな測定値の分散を有するデータの影響を増加させること、および／または比較的大きな測定値の分散を有するデータの影響を減少させることができる。例えば、少ない比率を占めるまたは低い品質の配列データを有する参照ゲノムの部分の「重み付けを減らし」て、データセットに対する影響を最小化することができ、一方、参照ゲノムの選択された部分の「重み付けを増やし」て、データセットに対する影響を増加させることもできる。重み付け関数の非限定的な例が、 $[1 / (\text{標準偏差})^2]$ である。重み付けステップは時には、正規化ステップに実質的に類似する様式で行われる。一部の実施形態では、データセットは、所定の変数（例えば、重み付け変数）により除算される。しばしば、所定の変数（例えば、最小化標的関数、Phi）を選択して、データセットの異なる一部分に異なる重み付けを加える（例えば、特定のデータのタイプの影響を増加させ、一方、その他のデータのタイプの影響を減少させる）。

30

【0190】

40

特定の実施形態では、処理ステップは、1つまたは複数の数学的および／または統計学的な操作を含むことができる。任意の適切な数学的および／または統計学的な操作を、単独でまたは組み合わせを使用して、本明細書に記載するデータセットを分析および／操作することができる。任意の適切な数の数学的および／または統計学的な操作を使用することができる。一部の実施形態では、データセットを、数学的および／または統計学的に、1回もしくは複数回、5回もしくはそれ超の回数、10回もしくはそれ超の回数、または20回もしくはそれ超の回数操作することができる。使用することができる数学的および統計学的な操作の非限定的な例として、加算、減算、乗算、除算、代数関数、最小二乗推定量、曲線近似、微分方程式、有理多項式、二重多項式、直交多項式、zスコア、p値、カイ値、phi値、ピークレベルの分析、ピークのエッジの場所の決定、ピーク面積比の

50

計算、染色体レベルの中央値の分析、平均絶対偏差の計算、残余の二乗の合計、平均値、標準偏差、標準誤差等、またはそれらの組合せが挙げられる。数学的および／または統計学的な操作を、配列の読取りのデータまたはそれらの処理された生成物の全部または一部に対して行うことができる。統計学的に操作することができ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、未加工カウント、フィルタリングしたカウント、正規化したカウント、ピークの高さ、ピークの幅、ピークの面積、ピークのエッジ、ラテラルトレランス (lateral tolerance)、P値、レベルの中央値、平均レベル、ゲノム領域内のカウントの分布、核酸種の相対的な表示等、またはそれらの組合せが挙げられる。

【0191】

一部の実施形態では、処理ステップは、1つまたは複数の統計学的アルゴリズムの使用を含むことができる。任意の適切な統計学的アルゴリズムを、単独でまたは組み合わせて使用して、本明細書に記載するデータセットを分析および／操作することができる。任意の適切な数の統計学的アルゴリズムを使用することができる。一部の実施形態では、1つもしくは複数、5つもしくはそれ超、10個もしくはそれ超、または20個もしくはそれ超の統計学的アルゴリズムを使用して、データセットを分析することができる。本明細書に記載する方法と共に使用するのに適切な統計学的アルゴリズムの非限定的な例として、決定木、対立帰無、多重比較、オムニバス検定、ペーレンスーフィッシャー問題、ブートストラッピング、有意性の独立性検定を組み合わせるためのフィッシャー法、帰無仮説、第一種の過誤、第二種の過誤、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、等分散を有する2標本プールt検定、不等分散を有する2標本非プールt検定、1比率z検定、2比率z検定プール、2比率z検定非プール、1標本カイ二乗検定、分散の一樣性についての2標本F検定、信頼区間、信頼区間 (credible interval)、有意性、メタ分析、単一線形回帰、ロバスト線形回帰等、または上記のものの組合せが挙げられる。統計学的アルゴリズムを使用して分析することができ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、未加工カウント、フィルタリングしたカウント、正規化したカウント、ピークの高さ、ピークの幅、ピークのエッジ、ラテラルトレランス、P値、レベルの中央値、平均レベル、ゲノム領域内のカウントの分布、核酸種の相対的な表示等、またはそれらの組合せが挙げられる。

【0192】

特定の実施形態では、複数 (例えば、2つもしくはそれ超) の統計学的アルゴリズム (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最近傍、ロジスティック回帰および／もしくは損失平滑化)、ならびに／または (例えば、本明細書では操作と呼ぶ) 数学的および／もしくは統計学的な操作を利用することによって、データセットを分析することができる。一部の実施形態では、複数の操作の使用により、アウトカムをもたらすために使用することができるN次元空間を生成することができる。特定の実施形態では、複数の操作を利用することによりデータセットを分析することによって、データセットの複雑性および／または次元性を低下させることができる。例えば、複数の操作を参照データセットに対して使用することによって、参照試料の遺伝子の状況 (例えば、選択された遺伝子の変動について陽性または陰性) に応じて、遺伝子の変動の存在または非存在を表示するために使用することができるN次元空間 (例えば、確率プロット) を生成することができる。実質的に類似する一連の操作を使用する試験試料の分析を使用して、試験試料のそれぞれについてN次元の点を生成することができる。試験被験体のデータセットの複雑性および／または次元性は時には、参照データから生成されたN次元空間と容易に比較することができる単一の値またはN次元の点に低減される。参照被験体のデータが存在するN次元空間に存在する試験試料データは、参照被験体の遺伝子の状況に実質的に類似する遺伝子の状況を示す。参照被験体のデータが存在するN次元空間の外側に存在する試験試料データは、参照被験体の遺伝子の状況に実質的に類似しない遺伝子の状況を示す。一部の実施形態では、参照は、正倍数体であるかまたは、別

10

20

30

40

50

段に、遺伝子の変動も医学的状态も有しない。

【0193】

一部の実施形態では、データセットを、計数し、任意選択でフィルタリングして正規化した後で、フィルタリングし、かつ／または正規化する1つまたは複数の手順により、これらの処理されたデータセットをさらに操作することができる。特定の実施形態では、フィルタリングし、かつ／または正規化する1つまたは複数の手順によりさらに操作されているデータセットを使用して、プロファイルを生成することができる。一部の実施形態では、時には、フィルタリングし、かつ／または正規化する1つまたは複数の手順により、データセットの複雑性および／または次元性を低下させることができる。低下させた複雑性および／または次元性のデータセットに基づいて、アウトカムをもたらすことができる。

10

【0194】

一部の実施形態では、誤差の尺度（例えば、標準偏差、標準誤差、計算した分散、p値、平均絶対誤差（mean absolute error）（MAE）、平均絶対偏差および／または平均絶対偏差（MAD））に従って、部分をフィルタリングすることができる。特定の実施形態では、誤差の尺度は、カウントの可変性を指す。一部の実施形態では、カウントの可変性に従って、部分をフィルタリングする。特定の実施形態では、カウントの可変性は、複数の試料（例えば、複数の被験体、例えば、50人／匹もしくはそれ超、100人／匹もしくはそれ超、500人／匹もしくはそれ超、1000人／匹もしくはそれ超、5000人／匹もしくはそれ超、または10,000人／匹もしくはそれ超の被験体から得られた複数の試料）について、参照ゲノムのある部分（すなわち、部分）に対してマッピングされたカウントについて決定した誤差の尺度である。一部の実施形態では、所定の上方範囲を上回るカウントの可変性を有する部分をフィルタリングする（例えば、検討事項から排除する）。一部の実施形態では、所定の上方範囲は、約50に等しいもしくはそれ超、約52に等しいもしくはそれ超、約54に等しいもしくはそれ超、約56に等しいもしくはそれ超、約58に等しいもしくはそれ超、約60に等しいもしくはそれ超、約62に等しいもしくはそれ超、約64に等しいもしくはそれ超、約66に等しいもしくはそれ超、約68に等しいもしくはそれ超、約70に等しいもしくはそれ超、約72に等しいもしくはそれ超、約74に等しいもしくはそれ超、または約76に等しいもしくはそれ超のMAD値である。一部の実施形態では、所定の下方範囲を下回るカウントの可変性を有する部分をフィルタリングする（例えば、検討事項から排除する）。一部の実施形態では、所定の下方範囲は、約40に等しいもしくはそれ未満、約35に等しいもしくはそれ未満、約30に等しいもしくはそれ未満、約25に等しいもしくはそれ未満、約20に等しいもしくはそれ未満、約15に等しいもしくはそれ未満、約10に等しいもしくはそれ未満、約5に等しいもしくはそれ未満、約1に等しいもしくはそれ未満、または約0に等しいもしくはそれ未満のMAD値である。一部の実施形態では、所定の範囲の外側にあるカウントの可変性を有する部分をフィルタリングする（例えば、検討事項から排除する）。一部の実施形態では、所定の範囲は、ゼロ超から、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未満までのMAD値である。一部の実施形態では、所定の範囲は、ゼロ超から約67.7未満までのMAD値である。一部の実施形態では、所定の範囲内のカウントの可変性を有する部分を選択する（例えば、遺伝子の変動の存在または非存在を決定するために使用する）。

20

30

40

【0195】

一部の実施形態では、部分のカウントの可変性が、分布（例えば、正規分布）を示す。一部の実施形態では、部分は、分布のクォンタイル内で選択される。一部の実施形態では、分布の約99.9%に等しいもしくはそれ未満、約99.8%に等しいもしくはそれ未満、約99.7%に等しいもしくはそれ未満、約99.6%に等しいもしくはそれ未満、約99.5%に等しいもしくはそれ未満、約99.4%に等しいもしくはそれ未満、約9

50

9. 3%に等しいもしくはそれ未満、約99.2%に等しいもしくはそれ未満、約99.1%に等しいもしくはそれ未満、約99.0%に等しいもしくはそれ未満、約98.9%に等しいもしくはそれ未満、約98.8%に等しいもしくはそれ未満、約98.7%に等しいもしくはそれ未満、約98.6%に等しいもしくはそれ未満、約98.5%に等しいもしくはそれ未満、約98.4%に等しいもしくはそれ未満、約98.3%に等しいもしくはそれ未満、約98.2%に等しいもしくはそれ未満、約98.1%に等しいもしくはそれ未満、約98.0%に等しいもしくはそれ未満、約97%に等しいもしくはそれ未満、約96%に等しいもしくはそれ未満、約95%に等しいもしくはそれ未満、約94%に等しいもしくはそれ未満、約93%に等しいもしくはそれ未満、約92%に等しいもしくはそれ未満、約91%に等しいもしくはそれ未満、約90%に等しいもしくはそれ未満、約85%に等しいもしくはそれ未満、約80%に等しいもしくはそれ未満、または約75%に等しいもしくはそれ未満のクォンタイル内の部分が選択される。一部の実施形態では、カウントの可変性の分布の99%クォンタイル内の部分が選択される。一部の実施形態では、99%クォンタイル内で、 $MAD > 0$ および $MAD < 67.725$ を有する部分が選択され、その結果、参照ゲノムの一連の安定な部分が同定される。

【0196】

PERUNに関する、部分をフィルタリングすることの非限定的な例が、例えば、本明細書および国際特許出願第PCT/US12/59123 (WO2013/052913)号に示されており、後者は、全ての文書、表、等式および図面を含めた、その内容全体が、参照により本明細書に援用されている。誤差の尺度に基づいて、または誤差の尺度の一部に基づいて、部分をフィルタリングすることができる。特定の実施形態では、R因子等の偏差の絶対値を含む誤差の尺度を使用して、部分の除去または部分への重み付けを行うことができる。R因子は、一部の実施形態では、実際の測定値から予測されるカウントの値の絶対偏差の合計を、実際の測定値から予測されるカウントの値で除算した結果と定義する（例えば、本明細書の等式B）。偏差の絶対値を含む誤差の尺度を使用することができるが、誤差の適切な尺度もそれに代わって利用することができる。特定の実施形態では、偏差の絶対値を含まない誤差の尺度、例として、二乗に基づくばらつきを利用することができる。一部の実施形態では、マッピング可能性の尺度（例えば、マッピング可能性スコア）に従って、部分をフィルタリングするまたは重み付けする。時には、部分に対してマッピングされた、比較的低い数の配列の読取り（例えば、部分に対してマッピングされた、0、1、2、3、4、5つの読取り）に従って、その部分をフィルタリングするまたは重み付けする。実施している分析のタイプに従って、部分をフィルタリングするまたは重み付けすることができる。例えば、第13、18および/または21染色体の異数性の分析の場合、性染色体をフィルタリングすることができ、常染色体のみまたは常染色体のサブセットを分析することができる。

【0197】

特定の実施形態では、以下のフィルタリングする処理を利用することができる。所与の染色体（例えば、第21染色体）内の同じ一連の部分（例えば、参照ゲノムの部分）を選択し、読取りの数を、罹患試料と非罹患試料とで比較する。ギャップにより、21トリソミー試料と正倍数体試料とを関係付け、これには、ほとんどの第21染色体をカバーする一連の部分を含める。これらの一連の部分は、正倍数体試料とT21試料との間で同じである。部分を定義することができるので、一連の部分と単一区分との区別はあまり重要でない。同じゲノム領域を、異なる患者において比較する。この処理を、トリソミーの分析、例として、T21に加えてまたはその代わりに、T13またはT18について利用することができる。

【0198】

一部の実施形態では、データセットを、計数し、任意選択でフィルタリングし正規化した後で、重み付けすることによって、これらの処理されたデータセットを操作することができる。特定の実施形態では、1つまたは複数の部分を選択し、それらに重み付けして、選択された部分中に含有されるデータ（例えば、ノイズの多いデータ、有益でないデータ

）の影響を低下させることができ、一部の実施形態では、1つまたは複数の部分を選択し、それらに重み付けして、選択された部分中に含有されるデータ（例えば、小さな分散が測定されたデータ）の影響を増強または増大させることができる。一部の実施形態では、大きな分散を有するデータの影響を減少させ、小さな分散を有するデータの影響を増加させる単一の重み付け関数を利用して、データセットに重み付けする。時には、重み付け関数を使用して、大きな分散を有するデータの影響を低下させ、小さな分散を有するデータの影響を増大させる（例えば、 $[1 / (\text{標準偏差})^2]$ ）。一部の実施形態では、重み付けによりさらに操作した処理済データのプロファイルのプロットを生成して、分類、および／またはアウトカムの提供を促進する。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、アウトカムをもたらすことができる。

10

【0199】

部分をフィルタリングすることまたは重み付けすることは、分析における1つまたは複数の適切な点で行うことができる。例えば、配列の読取りを、参照ゲノムの部分に対してマッピングする前または後に、部分をフィルタリングするまたは重み付けすることができる。一部の実施形態では、個々のゲノム部分についての実験の偏りを決定する前または後に、部分をフィルタリングするまたは重み付けすることができる。特定の実施形態では、ゲノム区分のレベルを計算する前または後に、部分をフィルタリングするまたは重み付けすることができる。

【0200】

一部の実施形態では、データセットを、計数し、任意選択でフィルタリングし、正規化し、任意選択で重み付けした後に、これらの処理されたデータセットを、1つまたは複数の数学的および／または統計学的な（例えば、統計学的関数または統計学的アルゴリズムによる）操作により操作することができる。特定の実施形態では、1つまたは複数の選択された部分、染色体、または染色体の部分についてZスコアを計算することによって、処理されたデータセットをさらに操作することができる。一部の実施形態では、P値を計算することによって、処理されたデータセットをさらに操作することができる。Zスコアおよびp値を計算するための等式の一実施形態を、等式1（実施例2）に示す。特定の実施形態では、数学的および／または統計学的な操作は、倍数性および／または胎児フラクションに関する1つまたは複数の仮定を含む。一部の実施形態では、1つまたは複数の統計学的小さい／または数学的な操作によりさらに操作した処理済データのプロファイルのプロットを生成して、分類、および／またはアウトカムの提供を促進する。統計学的小さい／または数学的に操作したデータのプロファイルのプロットに基づいて、アウトカムをもたらすことができる。統計学的小さい／または数学的に操作したデータのプロファイルのプロットに基づいてもたらされたアウトカムはしばしば、倍数性および／または胎児フラクションに関する1つまたは複数の仮定を含む。

20

30

【0201】

特定の実施形態では、データセットを、計数し、任意選択でフィルタリングし正規化した後で、複数の操作を、処理されたデータセットに対して行って、N次元空間および／またはN次元の点を生成する。N次元で分析したデータセットのプロファイルのプロットに基づいて、アウトカムをもたらすことができる。

40

【0202】

一部の実施形態では、データセットの処理および／または操作の一部としてまたはその後、1つまたは複数のピークレベルの分析、ピーク幅の分析、ピークのエッジの場所の分析、ピークのラテラルトレランス等、それらの誘導物、または上記の組合せを利用して、データセットを処理する。一部の実施形態では、1つまたは複数のピークレベルの分析、ピーク幅の分析、ピークのエッジの場所の分析、ピークのラテラルトレランス等、それらの誘導物、または上記の組合せを利用して処理したデータのプロファイルのプロットを生成して、分類、および／またはアウトカムの提供を促進する。1つまたは複数のピークレベルの分析、ピーク幅の分析、ピークのエッジの場所の分析、ピークのラテラルトレランス等、それらの誘導物、または上記の組合せを利用して処理してあるデータのプロファ

50

イルのプロットに基づいて、アウトカムをもたらすことができる。

【0203】

一部の実施形態では、問題の遺伝子変動を実質的に含有しない1つまたは複数の参照試料を使用して、参照のカウントプロファイルの中央値を得ることができ、この中央値は、遺伝子の変動の不在を表示する所定の値をもたらし得、しばしば、もし試験被験体が遺伝子の変動を保有するならば、その遺伝子の変動が試験被験体において位置するゲノムの場所に対応する領域における所定の値から、当該中央値は逸脱する。遺伝子の変動と関連する医学的状態のリスクがある、またはそうした医学的状態に罹患している試験被験体において、選択された部分または区分についての数値は、罹患していない場合のゲノムの場所についての所定の値とは有意に異なるものになることが予想される。特定の実施形態では、問題の遺伝子変動を担持することが分かっている1つまたは複数の参照試料を使用して、参照のカウントプロファイルの中央値を生成することができ、この中央値は、遺伝子の変動の存在を表示する所定の値をもたらし得、しばしば、試験被験体とその遺伝子の変動を担持しないゲノムの場所に対応する領域における所定の値から、当該中央値は逸脱する。遺伝子の変動と関連する医学的状態のリスクがない、またはそうした医学的状態に罹患していない試験被験体においては、選択された部分または区分についての数値は、罹患している場合のゲノムの場所についての所定の値とは有意に異なることが予想される。

10

【0204】

一部の実施形態では、データの分析および処理は、1つまたは複数の仮定の使用を含むことができる。適切な数またはタイプの仮定を利用して、データセットを分析または処理することができる。データの処理および／または分析のために使用することができる仮定の非限定的な例として、母体の倍数性、胎児の寄与、参照集団中の特定の配列の存在率、民族性背景、血縁の家族における選択された医学的状態の存在率、異なる患者から得られた未加工カウントのプロファイル間の平行度および／またはGC正規化およびリピートマスキング（GC-normalization and repeat masking）（例えば、GCRM）後のラン、PCRの人工産物を表わす同一の一致（例えば、同一塩基の位置）、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、双子の両方のうち、一方のみが罹患している場合、有効な胎児フラクションは、測定された全胎児フラクションの50%のみである（三つ子、四つ子等についても同様））、ゲノム全体を一樣にカバーする胎児の無細胞DNA（例えば、cfDNA）等、ならびにそれらの組合せが挙げられる。

20

30

【0205】

正規化されたカウントプロファイルに基づいて、遺伝子の変動の存在または非存在のアウトカムを所望の信頼性のレベル（例えば、95%またはそれ超の信頼性のレベル）で予測することが、マッピングされた配列の読取りの品質および／または深さでは可能でない事例では、1つまたは複数の追加の数学的操作のアルゴリズムおよび／または統計学的予測アルゴリズムを利用して、データ分析および／またはアウトカムの提供に有用な追加の数値を生成することができる。用語「正規化されたカウントプロファイル」は、本明細書で使用する場合、正規化されたカウントを使用して生成されたプロファイルを指す。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することができる方法の例を、本明細書に記載する。上記で述べたように、計数されるに至った、マッピングされた配列の読取りを、試験試料のカウントまたは参照試料のカウントに関して正規化することができる。一部の実施形態では、正規化されたカウントプロファイルは、プロットして示すことができる。

40

【0206】

LOESS正規化

LOESSとは、当技術分野で公知の回帰モデル化法であって、多重回帰モデルを、k最近傍法ベースのメタモデル内で組み合わせる回帰モデル化法である。LOESSは、場合によって、局所重み付け多項式回帰と称する。一部の実施形態では、GC LOESSでは、LOESSモデルを、断片のカウント（例えば、配列の読取り、カウント）と、参

50

照ゲノムの部分についてのGC組成との間の関係へと適用する。データ点のセットを通る滑らかな曲線のプロットングであって、LOESSを使用するプロットングは、場合によって、LOESS曲線と呼ばれ、特に、各平滑値が、y軸の散布図基準変数の値の区間にわたる、重み付き二次最小二乗回帰により与えられる場合、そう呼ばれる。データセット中の各点について、LOESS法は、低次多項式を、説明変数値がその応答が推定される点の近傍にあるデータのサブセットへと適合させる。多項式は、その応答が推定される点の近傍の点には大きな重みを与え、遠く離れた点には小さな重みを与える、重み付き最小二乗法を使用して適合させる。次いで、点についての回帰関数値を、そのデータ点についての説明変数値を使用して、局所多項式の値を評価することにより得る。LOESS適合は、場合によって、回帰関数値を、データ点の各々について計算した後において、完全であると考えられる。多項式モデルの次数および重みなど、この方法の詳細の多くは、適応性がある。

10

【0207】

PERUN正規化

本明細書では、核酸指標と関連する誤差を低減するための正規化法を、パラメータ化誤差除去および不偏正規化 (PERUN: parameterized error removal and unbiased normalization) と称するが、これは、本明細書ならびに本文、表、等式、および図面の全てを含むその全内容が参照により本明細書に援用される、国際特許出願第PCT/US12/59123号 (WO2013/052913) において記載されている。PERUN法は、このような指標に基づく予測を交絡させる誤差の影響を低減する目的で、様々な核酸指標 (例えば、核酸配列の読取り) へと適用することができる。

20

【0208】

例えば、PERUN法を、試料に由来する核酸配列の読取りへと適用し、ゲノム区分のレベルの決定を損ないうる誤差の影響を低減することができる。このような適用は、核酸配列の読取りを使用して、被験体における遺伝子の変動の存在または非存在であって、ヌクレオチド配列の様々なレベル (例えば、部分レベル、ゲノム区分のレベル) として顕在化される存在または非存在を決定するのに有用である。部分中の変動の非限定的な例は、染色体の異数性 (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または非存在 (例えば、女性におけるXX対男性におけるXY) 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することができる。ゲノム区分のレベルにおける変動の他の非限定的な例は、微小欠失、微小挿入、重複、およびモザイク現象を含む。

30

【0209】

ある特定の適用では、PERUN法は、それを読取り部分と称し、場合によって、参照ゲノムの部分とも称する、参照ゲノムの特定の部分へとマッピングした核酸の読取りを正規化することにより、実験上の偏りを低減することができる。このような適用では、PERUN法は一般に、多数の試料にわたり、参照ゲノムの特定の部分における核酸の読取りのカウントを、三次元で正規化する。PERUNについての詳細な記載およびその適用については、本明細書の実施例の節、本文、表、等式、および図面の全てを含むその全内容が参照により本明細書に援用される、国際特許出願第PCT/US12/59123号 (WO2013/052913) および米国特許出願公開US20130085681において提示されている。

40

【0210】

ある特定の実施形態では、PERUN法は、参照ゲノムの部分についてのゲノム区分のレベルを、(a) 試験試料についての、参照ゲノムの部分へとマッピングした配列の読取りのカウント、(b) 試験試料についての、実験上の偏り (例えば、GCの偏り)、および(c) (i) 配列の読取りがマッピングされる参照ゲノムの部分についての実験上の偏りと、(ii) 部分へとマッピングした配列の読取りのカウントとの適合させた関係についての、1つまたは複数の適合させたパラメータ (例えば、適合の推定値) から計算する

50

ステップを含む。参照ゲノムの部分の各々についての実験上の偏りは、複数の試料にわたり、各試料についての適合させた関係であって、(i) 参照ゲノムの部分の各々へとマッピングした配列の読取りのカウントと、(ii) 参照ゲノムの部分の各々についてのマッピング特徴との関係に従って決定することができる。この各試料についての適合させた関係は、複数の試料について、三次元でアセンブルすることができる。ある特定の実施形態では、アセンブリーを、実験上の偏りに従って整序することもできるが、PERUN法は、実験上の偏りに従ってアセンブリーを整序することなく実施することもできる。各試料についての適合させた関係と、参照ゲノムの各部分についての適合させた関係とは、当技術分野で公知の適切な適合させた処理により、線形関数または非線形関数へと独立に適合させることができる。

10

【0211】

一部の実施形態では、関係は、幾何学的関係および／またはグラフ的关系である。一部の実施形態では、関係は、数学的关系である。一部の実施形態では、関係は、プロットされる。一部の実施形態では、関係は、線形関係である。ある特定の実施形態では、関係は、非線形関係である。ある特定の実施形態では、関係は、回帰（例えば、回帰直線）である。回帰は、線形回帰の場合もあり、非線形回帰の場合もある。関係は、数式により表すことができる。関係は一部分、1つまたは複数の定数により規定されることが多い。関係は、当技術分野で公知の方法により生成することができる。ある特定の実施形態では、1つまたは複数の試料について、二次元の関係を生成することができ、誤差の証明となる変数、または誤差の証明となる可能性が高い変数を、次元のうちの1または複数について選択することができる。関係は、例えば、当技術分野で公知のグラフ作成ソフトウェアであって、使用者により用意される2つまたはそれ超の変数の値を使用してグラフをプロットするソフトウェアを使用して、生成することができる。関係は、当技術分野で公知の方法（例えば、グラフ作成ソフトウェア）を使用して適合させることができる。ある特定の関係には、線形回帰を適合させることができ、線形回帰により、傾き値および切片値を生成することができる。ある特定の関係は、場合によって、線形ではなく、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）などの非線形関数を適合させることができる。

20

【0212】

PERUN法では、適合させた関係のうちの1または複数のは、線形でありうる。妊娠中の雌に由来する無細胞循環核酸についての分析であって、実験上の偏りをGCの偏りとし、マッピング特徴をGC含有量とする分析では、試料についての適合させた関係であって、(i) 各部分へとマッピングした配列の読取りのカウントと、(ii) 参照ゲノムの部分の各々についてのGC含有量との間の関係は、線形でありうる。この適合させた関係では、傾きは、GCの偏りに関連し、適合させた関係を、複数の試料にわたりアセンブルする場合、GCの偏り係数は、各試料について決定することができる。このような実施形態では、複数の試料および部分についての適合させた関係であって、(i) 部分についてのGCの偏り係数と、(ii) 部分へとマッピングした配列の読取りのカウントとの間の関係もまた、線形でありうる。切片および傾きは、この適合させた関係から得ることができる。このような適用では、傾きは、GC含有量に基づく試料特異的な偏りに対処し、切片は、全ての試料に共通する、部分特異的な減衰パターンに対処する。PERUN法により、このような試料特異的な偏りおよび部分特異的な減衰であって、アウトカム（例えば、遺伝子の変動の存在または非存在；胎児の性別の決定）をもたらすためにゲノム区分のレベルを計算する場合の偏りおよび減衰を有意に低減することができる。

30

40

【0213】

一部の実施形態では、PERUN正規化に、線形関数への適合を使用し、これを、等式A、等式B、またはその派生形により記載する。

等式A：

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

等式B：

50

$$L = (M - GS) / I \quad (B)$$

【0214】

一部の実施形態では、Lは、PERUNにより正規化されたレベルまたはPERUNにより正規化されたプロファイルである。一部の実施形態では、Lは、PERUN正規化手順からの所望の出力である。ある特定の実施形態では、Lは、部分特異的である。一部の実施形態では、Lは、参照ゲノムの複数の部分に従って決定され、ゲノム、染色体、その部分またはセグメントの、PERUNにより正規化されたレベルを表示する。レベルLは、さらなる分析（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、性異数性などを決定する分析）に使用されることが多い。等式Bに従う正規化法を、PERUN (parameterized error removal and unbiased normalization) と名付ける。

10

【0215】

一部の実施形態では、Gは、線形モデル、LOESS、または任意の同等の手法を使用して測定されたGCの偏り係数である。一部の実施形態では、Gは、傾きである。一部の実施形態では、GCの偏り係数であるGは、部分iについてのカウントM（例えば、未加工のカウント）および参照ゲノムから決定された部分iのGC含有量についての回帰の傾きとして評価する。一部の実施形態では、Gは、Mから抽出された副次情報を表し、関係に従って決定される。一部の実施形態では、Gは、試料（例えば、試験試料）についての、部分特異的なカウントのセットと、部分特異的なGC含有量値のセットとの関係を表わす。一部の実施形態では、部分特異的なGC含有量は、参照ゲノムから導出される。一部の実施形態では、部分特異的なGC含有量は、観察または測定されたGC含有量（例えば、試料から測定されたGC含有量）から導出される。GCの偏り係数は、試料群中の各試料について決定されることが多く、一般に、試験試料について決定される。GCの偏り係数は、試料特異的であることが多い。一部の実施形態では、GCの偏り係数は、定数である。ある特定の実施形態では、GCの偏り係数は、試料について導出されたら変化しない。

20

【0216】

一部の実施形態では、Iは、線形関係から導出される切片であり、Sは、線形関係から導出される傾きである。一部の実施形態では、IおよびSが導出される関係は、Gが導出される関係と異なる。一部の実施形態では、IおよびSが導出される関係は、所与の実験設定について一定である。一部の実施形態では、IおよびSは、複数の試料に従って、カウント（例えば、未加工のカウント）と、GCの偏り係数とに従う線形関係から導出される。一部の実施形態では、IおよびSは、試験試料とは独立に導出される。一部の実施形態では、IおよびSは、複数の試料から導出される。IおよびSは、部分特異的であることが多い。一部の実施形態では、IおよびSを、正倍数体試料中の参照ゲノムの全ての部分についてL=1であるという仮定により決定する。一部の実施形態では、線形関係を、正倍数体試料について決定し、選択部分に特異的なI値およびS値を決定する（L=1と仮定する）。ある特定の実施形態では、同じ手順を、ヒトゲノム中の参照ゲノムの全ての部分へと適用し、切片Iおよび傾きSのセットを、あらゆる部分について決定する。

30

【0217】

一部の実施形態では、交差検証法を適用する。交差検証は、場合によって、回転推定とも称する。一部の実施形態では、交差検証法を適用して、試験試料を使用する実践において、予測モデル（例えば、PERUNなど）が、どのくらい正確に機能するのかについて評価する。一部の実施形態では、1ラウンドの交差検証は、データの試料を、相補サブセットへとパーティショニングするステップと、1つのサブセット（例えば、場合によって、訓練セットと称する）に対して交差検証分析を実施するステップと、別のサブセット（例えば、場合によって、検証セットまたは試験セットと呼ばれる）を使用して、分析の検証するステップとを含む。ある特定の実施形態では、異なるパーティションおよび／または異なるサブセットを使用して、複数ラウンドの交差検証を実施する。交差検証法の非限定的な例は、リーブワンアウト、スライディングエッジ、K分割、二分割、反復ランダム

40

50

サブサンプリングなど、またはこれらの組合せを含む。一部の実施形態では、交差検証では、既知の正倍数性胎児を含む試料のセットのうちの90%を含有する作業セットをランダムに選択し、こ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では、ランダムな選択を100回にわたり反復し、あらゆる部分について、100の傾きおよび100の切片のセットを得る。

【0218】

一部の実施形態では、Mの値は、試験試料に由来する実測値である。一部の実施形態では、Mは、部分についての測定された未加工のカウントである。値IおよびSが部分について利用可能である一部の実施形態では、測定値Mを試験試料から決定し、それを使用して、等式Bに従って、ゲノム、染色体、そのセグメントまたは部分についてのPERUNにより正規化されたレベルであるLを決定する。

10

【0219】

したがって、PERUN法の、複数の試料にわたり平行した、配列の読取りへの適用により、(i) 試料特異的な実験上の偏り（例えば、GCの偏り）および(ii) 試料に共通する部分特異的な減衰により引き起こされる誤差を有意に低減することができる。これらの2つの誤差の発生源の各々に、個別にまたは逐次的に対処する他の方法は、PERUN法ほど有効にはこれらを低減することが可能でないことが多い。理論に制約されずに述べると、一部分、その一般に加法的な処理が、他の正規化法（例えば、GC-LOESS）で活用される、一般に乗算的な処理ほど広がり拡大しないため、PERUN法は、誤差をより有効に低減することが期待される。

20

【0220】

さらなる正規化技法および統計学的技法も、PERUN法と組み合わせて活用することができる。さらなる処理は、PERUN法の使用の前、使用の後、および／または使用の間に適用することができる。PERUN法と組み合わせて使用されうる処理の非限定的な例については、本明細書の下記で記載される。

【0221】

一部の実施形態では、ゲノム区分のレベルの、GC含有量についての二次的正規化または調整は、PERUN法と共に活用することができる。適切なGC含有量の調整手順またはGC含有量の正規化手順を活用することができる（例えば、GC-LOESS、GCRM）。ある特定の実施形態では、さらなるGCの正規化処理を適用するための特定の試料を同定することができる。例えば、PERUN法を適用することにより、各試料についてのGCの偏りを決定することができ、ある特定の閾を上回るGCの偏りと関連する試料を、さらなるGCの正規化処理のために選択することができる。このような実施形態では、所定の閾レベルを使用して、このような試料をさらなるGCの正規化のために選択することができる。

30

【0222】

ある特定の実施形態では、部分のフィルタリング処理または重み付き処理を、PERUN法と共に活用することができる。適する部分のフィルタリング処理または重み付き処理を活用することができ、非限定的な例は、本明細書、本文、表、等式、および図面の全てを含むその全内容が参照により本明細書に援用される、国際特許出願第PCT/US12/59123号(WO2013/052913)および米国特許出願公開第US20130085681号において記載されている。一部の実施形態では、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の変動）と関連する誤差を低減する正規化技法を、PERUN法と共に活用する。

40

【0223】

PERUN法により計算されたゲノム区分のレベルは、アウトカムを提示するために直接活用することができる。一部の実施形態では、ゲノム区分のレベルは、胎児フラクションが約2%～約6%またはそれ超の（例えば、胎児フラクションが約4%またはそれ超の）試料についてのアウトカムを提示するのに直接活用することができる。PERUN法により計算されたゲノム区分のレベルは、場合によって、アウトカムを提示するためにさら

50

に処理される。一部の実施形態では、計算されたゲノム区分のレベルを標準化する。ある特定の実施形態では、試験部分（例えば、第21染色体）について計算されたゲノム区分のレベルの合計、平均値、または中央値を、試験部分以外の部分（例えば、第21染色体以外の常染色体）について計算されたゲノム区分のレベルの合計、平均値、または中央値で除して、試験ゲノム区分のレベルを生成することができる。試験ゲノム区分のレベルまたは未加工のゲノム区分のレベルは、ZスコアまたはZスコアの計算などの標準化分析の一部として使用することができる。Zスコアは、期待されたゲノム区分のレベルを、試験ゲノム区分のレベルまたは未加工のゲノム区分のレベルから減算することにより、試料について生成することができ、結果として得られる値を、試料についての標準偏差で除算することができる。ある特定の実施形態では、結果として得られるZスコアを、異なる試料について分布させ、分析することもできるか、または胎児フラクションおよび他の変数など、他の変数と関係づけ、分析して、アウトカムを提示することもできる。

10

【0224】

本明細書で注目される通り、PERUN法は、GCの偏りおよびGC含有量に従う正規化自体に限定されず、他の誤差の発生源と関連する誤差を低減するのにも使用することができる。GC含有量以外の偏りの発生源の非限定的な例は、マッピング可能性である。GCの偏りおよびGC含有量以外の正規化されたパラメータに対処する場合、適合させた関係のうちの1または複数は、非線形（例えば、双曲線、指数）でありうる。一部の実施形態では、実験上の偏りを、非線形関係から決定する場合、例えば、実験上の偏りの曲率の推定について分析することができる。

20

【0225】

PERUN法は、様々な核酸指標へと適用することができる。核酸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の読取りおよび核酸レベルである。配列の読取りの非限定的な例は、無細胞循環DNA、無細胞循環RNA、細胞DNAおよび細胞RNAから得られる読取りを含む。PERUN法は、参照ゲノムDNA、参照細胞RNA（例えば、トランスクリプトーム）、およびこれらの部分（例えば、ゲノムDNAの相補体またはRNAトランスクリプトームの一部分（複数可）、染色体の一部分（複数可））など、適切な参照配列へとマッピングした配列の読取りへと適用することができる。

【0226】

したがって、ある特定の実施形態では、細胞核酸（例えば、DNAまたはRNA）は、核酸指標として役立ちうる。参照ゲノムの部分へとマッピングした細胞核酸の読取りは、PERUN法を使用して正規化することができる。細胞核酸の、特定のタンパク質への結合は、場合によって、クロマチン免疫沈降（ChIP: chromatin immunoprecipitation）過程と称する。ChIPに富む核酸は、例えば、DNAまたはRNAなど、細胞タンパク質と会合する核酸である。ChIPに富む核酸の読取りは、当技術分野で公知の技術を使用して得ることができる。ChIPに富む核酸の読取りは、1つまたは複数の参照ゲノムの部分へとマッピングすることができ、結果は、アウトカムを提示するためのPERUN法を使用して正規化することができる。

30

【0227】

ある特定の実施形態では、細胞RNAは、核酸指標として用いられうる。細胞RNA読取りは、参照RNA部分へとマッピングし、アウトカムを提示するためのPERUN法を使用して正規化することができる。トランスクリプトームと称する細胞RNAまたはそのセグメントについての公知の配列を、試料に由来するRNA読取りをマッピングしうる参照として使用することができる。試料RNAの読取りは、当技術分野で公知の技術を使用して得ることができる。参照へとマッピングしたRNA読取りの結果は、アウトカムを提示するためのPERUN法を使用して正規化することができる。

40

【0228】

一部の実施形態では、マイクロアレイによる核酸レベルは、核酸指標として役立ちうる。試料にわたり、アレイ上の特定のアドレスの核酸レベルまたはアレイ上でハイブリダイ

50

ズしている核酸を、P E R U N法を使用して分析し、これにより、マイクロアレイ分析によりもたらされる核酸指標を正規化することができる。このようにして、マイクロアレイ上の特定のアドレスまたはマイクロアレイ上でハイブリダイズしている核酸は、マッピングした核酸配列の読取りの部分と類義であり、P E R U N法を使用して、マイクロアレイデータを正規化して、アウトカムの改善をもたらすことができる。

【0229】

C h A Iによる正規化

本明細書では、核酸指標と関連する誤差を低減するのに使用されうる別の正規化法を、C h A Iと称し、これには主成分分析を使用することが多い。ある特定の実施形態では、主成分分析は、(a) 読取り密度分布に従って、参照ゲノムの部分をフィルタリングし、これにより、試験試料についての読取り密度プロファイルであって、フィルタリングされた部分の読取り密度を含み、読取り密度が、妊娠中の雌による試験試料に由来する循環無細胞核酸の配列の読取りを含み、読取り密度分布が、複数の試料についての部分の読取り密度について決定されるプロファイルを提示するステップと、(b) 1つまたは複数の主成分であって、主成分分析により公知の正倍数体試料のセットから得られる主成分に従って、試験試料についての読取り密度プロファイルを調整し、これにより、調整された読取り密度を含む試験試料プロファイルを提示するステップと、(c) 試験試料プロファイルを、参照プロファイルと比較し、これにより、比較をもたらすステップとを含む。一部の実施形態では、主成分分析は、(d) 比較に従って、遺伝子の変動の存在または非存在を、試験試料について決定するステップを含む。

【0230】

部分のフィルタリング

ある特定の実施形態では、1つまたは複数の部分(例えば、ゲノム部分)を、フィルタリング処理により検討事項から除外する。ある特定の実施形態では、1つまたは複数の部分をフィルタリングし(例えば、フィルタリング処理にかけ)、これにより、フィルタリングされた部分を提示する。一部の実施形態では、フィルタリング処理により、ある特定の部分を除外し、部分(例えば、部分のサブセット)を保持する。本明細書では、フィルタリング処理の後で保持された部分を、フィルタリングされた部分と称することが多い。一部の実施形態では、参照ゲノムの部分をフィルタリングする。一部の実施形態では、フィルタリング処理により除外された参照ゲノムの部分は、遺伝子の変動(例えば、染色体の異数性、微小重複、微小欠失)の存在または非存在の決定に含まれない。一部の実施形態では、読取り密度と関連する部分(例えば、読取り密度が部分についての読取り密度である場合)は、フィルタリング処理により除外され、除外された部分と関連する読取り密度は、遺伝子の変動(例えば、染色体の異数性、微小重複、微小欠失)の存在または非存在の決定に含まれない。一部の実施形態では、読取り密度プロファイルは、フィルタリングされた部分の読取り密度を含み、かつ／またはこれからなる。部分は、任意の適切な判定基準および／または当技術分野で公知の方法もしくは本明細書で記載される方法を使用して、選択し、フィルタリングし、かつ／または検討事項から除外することができる。部分のフィルタリングに使用される判定基準の非限定的な例は、冗長データ(例えば、マッピングした読取りの冗長または重複)、有益でないデータ(例えば、マッピングしたカウントがゼロである参照ゲノムの部分)、大きな比率を占める配列または少ない比率を占める配列を有する参照ゲノムの部分、G C含有量、ノイズの多いデータ、マッピング可能性、カウント、カウントの可変性、読取り密度、読取り密度の可変性、不確定性の尺度、再現性の尺度など、または前出の組合せを含む。部分は、場合によって、カウントの分布および／または読取り密度の分布に従ってフィルタリングされる。一部の実施形態では、部分を、カウントおよび／または読取り密度が、1つまたは複数の参照試料から得られる場合の、カウントの分布および／または読取り密度に従ってフィルタリングする。本明細書では、場合によって、1つまたは複数の参照試料を、訓練セットと称する。一部の実施形態では、部分を、カウントおよび／または読取り密度が、1つまたは複数の試験試料から得られる場合の、カウントの分布および／または読取り密度に従ってフィルタリングする

。一部の実施形態では、部分を、読取り密度分布についての不確定性の尺度に従ってフィルタリングする。ある特定の実施形態では、読取り密度の大きな偏差を裏付ける部分を、フィルタリング処理により除外する。例えば、分布中の各読取り密度が、同じ部分へとマッピングされる場合は、読取り密度の分布（例えば、読取り密度の平均、読取り密度の平均値、または読取り密度の中央値の分布；例えば、図37Aの分布）を決定することができる。ゲノムの各部分が、不確定性の尺度と関連する場合は、読取り密度の分布を複数の試料について比較することにより、不確定性の尺度（例えば、MAD）を決定することができる。前出の例によれば、部分は、各部分と関連する不確定性の尺度（例えば、標準偏差（SD）、MAD）および所定の閾に従ってフィルタリングすることができる。図37Bは、部分についてのMAD値の分布であって、複数の試料についての読取り密度分布に従って決定される分布を示す。所定の閾を、許容可能なMAD値の範囲を取り囲む垂直方向の破線で指し示す。図37Bの例では、許容可能な範囲中のMAD値を含む部分を保持し、許容可能な範囲外のMAD値を含む部分を、フィルタリング処理により検討事項から除外する。一部の実施形態では、前出の例に従って、所定の不確定性の尺度外の読取り密度値（例えば、読取り密度の中央値、平均、または平均値）を含む部分を、フィルタリング処理により検討事項から除外することが多い。一部の実施形態では、分布の四分位範囲外の読取り密度値（例えば、読取り密度の中央値、平均、または平均値）を含む部分を、フィルタリング処理により検討事項から除外する。一部の実施形態では、分布の四分位範囲を2倍、3倍、4倍、または5倍を超えて外れる読取り密度値を含む部分を、フィルタリング処理により検討事項から除外する。一部の実施形態では、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（例えば、シグマが、標準偏差により規定される範囲である場合）を超えて外れる読取り密度値を含む部分を、フィルタリング処理により検討事項から除外する。

【0231】

一部の実施形態では、システムは、フィルタリングモジュールを含む（18、図42A）。フィルタリングモジュールは、部分（例えば、所定のサイズおよび／または重複の部分、参照ゲノム中の部分の位置）および部分と関連する読取り密度であって、別の適切なモジュール（例えば、分布モジュール12、図42A）に由来することが多い読取り密度を、受容、回収、および／または保存することが多い。一部の実施形態では、選択部分（例えば、20（図42A）、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り提示される。一部の実施形態では、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示し、かつ／または部分を検討事項から除外するように必要とされる。ある特定の実施形態では、読取り密度が除外された部分と関連する場合は、フィルタリングモジュールにより、読取り密度を検討事項から除外する。フィルタリングモジュールは、選択部分（例えば、フィルタリングされた部分）を、別の適切なモジュール（例えば、分布モジュール12、図42A）へと提示することが多い。フィルタリングモジュールの非限定的な例を、実施例7に提示する。

【0232】

偏りの推定値

配列決定技術は、複数の偏り発生源に対して脆弱でありうる。場合によって、配列決定の偏りは、局所的な偏り（例えば、局所的なゲノムの偏り）である。局所的な偏りは、配列の読取りのレベルで顕在化することが多い。局所的なゲノムの偏りは、任意の適切な局所的な偏りでありうる。局所的な偏りの非限定的な例は、配列の偏り（例えば、GCの偏り、ATの偏りなど）、DNアーゼI感度、エントロピー、反復配列の偏り、クロマチン構造の偏り、ポリメラーゼエラー率の偏り、回分配列の偏り、逆位リピートの偏り、PCR関連の偏りなど、またはこれらの組合せと相関する偏りを含む。一部の実施形態では、局所的な偏りの発生源は、決定されていないか、または公知ではない。

【0233】

一部の実施形態では、局所的なゲノムの偏りの推定値を決定する。本明細書では、場合によって、局所的なゲノムの偏りの推定値を、局所的なゲノムの偏りの推定と称する。局

所的なゲノムの偏りの推定値は、参照ゲノム、そのセグメントまたは部分について決定することができる。一部の実施形態では、局所的なゲノムの偏りの推定値を、1つまたは複数の配列の読取り（例えば、試料の一部または全部の配列の読取り）について決定する。局所的なゲノムの偏りの推定値は、参照（例えば、参照ゲノム）の対応する位置（location）および／または場所（position）についての局所的なゲノムの偏りの推定に従って、配列の読取りについて決定することが多い。一部の実施形態では、局所的なゲノムの偏りの推定値は、配列（例えば、参照ゲノムの配列の読取り、配列）の偏りの定量的尺度を含む。局所的なゲノムの偏りの推定は、適切な方法または数学的处理により決定することができる。一部の実施形態では、局所的なゲノムの偏りの推定値を、適切な分布および／または適切な分布関数（例えば、PDF）により決定する。一部の実施形態では、局所的なゲノムの偏りの推定値は、PDFの定量的表示を含む。一部の実施形態では、局所的なゲノムの偏りの推定値（例えば、確率密度推定（PDE：probability density estimation）、カーネル密度推定）を、局所的な偏りの含有量の確率密度関数（例えば、PDF：probability density function、例えば、カーネル密度関数）により決定する。一部の実施形態では、密度推定は、カーネル密度推定を含む。局所的なゲノムの偏りの推定値は、場合によって、分布の平均、平均値、または中央値として表す。場合によって、局所的なゲノムの偏りの推定値は、適切な分布の合計または積分（例えば、曲線下面積（AUC：area under a curve））として表す。

【0234】

PDF（例えば、カーネル密度関数、例えば、エパネクニコフカーネル密度関数）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を使用する場合の確率密度推定値（PDE）を導出するウィンドウのサイズおよび／または長さを規定することが多い。PDEを導出するウィンドウは、規定された長さのポリヌクレオチドを含むことが多い。一部の実施形態では、PDEを導出するウィンドウは、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することが多い。バンド幅変数により、局所的なゲノムの偏りの推定値を決定するのに使用されるウィンドウの長さまたはサイズであって、そこから局所的なゲノムの偏りの推定値を決定する、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的なセグメント）である、ウィンドウの長さまたはサイズを決定する。その非限定的な例が、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1000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を含む、任意の適切なバンド幅を使用して、PDE（例えば、読取り密度、局所的なゲノムの偏りの推定値（例えば、GC密度））を決定することができる。一部の実施形態では、局所的なゲノムの偏りの推定値（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使用して決定する。ある特定の実施形態では、局所的なゲノムの偏りの推定値（例えば、GC密度）は、所与の被験体および／または試料について得られた配列の読取りの平均読取り長さ、平均値読取り長さ、中央値読取り長さ、または最大読取り長さに従って決定されたバンド幅を使用して決定する。場合によって、局所的なゲノムの偏りの推定値（例えば、GC密度）は、所与の被験体および／または試料について得られた配列の読取りの平均読取り長さ、平均値読取り長さ、中央値読取り長さ、または最大読取り長さとはほぼ等しいバンド幅を使用して決定する。一部の実施形態では、局所的なゲノムの偏りの推定値（例えば、GC密度）は、約250、240、230、220、210、200、190、180、160、150、140、130、120、110、100、

10

20

30

40

50

90、80、70、60、50、40、30、20、または約10塩基のバンド幅を使用して決定する。

【0235】

局所的なゲノムの偏りの推定値は、単一塩基分解で決定しうるが、局所的なゲノムの偏りの推定値（例えば、局所的なGC含有量）は、低分解度でも決定することができる。一部の実施形態では、局所的なゲノムの偏りの推定値を、局所的な偏りの含有量について決定する。局所的なゲノムの偏りの推定値（例えば、PDFを使用して決定される）は、ウィンドウを使用して決定することが多い。一部の実施形態では、局所的なゲノムの偏りの推定値は、あらかじめ選択された数の塩基を含むウィンドウの使用を含む。場合によって、ウィンドウは、連続的な塩基のセグメントを含む。場合によって、ウィンドウは、1つまたは複数の非連続的な塩基の部分を含む。場合によって、ウィンドウは、1つまたは複数の部分（例えば、ゲノム部分）を含む。ウィンドウのサイズまたは長さは、バンド幅により、かつ、PDFに従って決定することが多い。一部の実施形態では、ウィンドウは、バンド幅の長さの約10倍もしくはそれ超、8倍もしくはそれ超、7倍もしくはそれ超、6倍もしくはそれ超、5倍もしくはそれ超、4倍もしくはそれ超、3倍もしくはそれ超、または約2倍もしくはそれ超である。PDF（例えば、カーネル密度関数）を使用して、密度推定値を決定する場合、ウィンドウは、場合によって、選択されたバンド幅の長さの2倍である。ウィンドウは、任意の適切な数の塩基を含みうる。一部の実施形態では、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1000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。一部の実施形態では、ゲノムまたはそのセグメントを、複数のウィンドウへとパーティショニングする。ゲノムの領域を包含するウィンドウは、重複する場合もあり、重複しない場合もある。一部の実施形態では、互いから等距離にウィンドウを配置する。一部の実施形態では、互いから異なる距離にウィンドウを配置する。ある特定の実施形態では、ゲノムまたはそのセグメントを、ウィンドウを、ゲノムまたはそのセグメントにわたり徐々にスライドさせる、複数のスライディングウィンドウへとパーティショニングする。各インクリメントの各ウィンドウは、局所的なゲノムの偏りの推定値（例えば、局所GC密度）を含む。ウィンドウは、ゲノムにわたり、任意の適切なインクリメントでスライドさせることもでき、任意の数値パターンに従ってスライドさせることもでき、任意の無主題の規定配列に従ってスライドさせることもできる。一部の実施形態では、局所的なゲノムの偏りの推定値を決定するために、ゲノムまたはそのセグメントにわたり、約10、000bpまたはそれ超、約5、000bpまたはそれ超、約2、500bpまたはそれ超、約1、000bpまたはそれ超、約750bpまたはそれ超、約500bpまたはそれ超、約400塩基またはそれ超、約250bpまたはそれ超、約100bpまたはそれ超、約50bpまたはそれ超、または約25bpまたはそれ超のインクリメントでウィンドウをスライドさせる。一部の実施形態では、局所的なゲノムの偏りの推定値を決定するために、ゲノムまたはそのセグメントにわたり、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のインクリメントでウィンドウをスライドさせる。例えば、局所的なゲノムの偏りの推定値を決定するために、ウィンドウは、約400bp（例えば、200bpのバンド幅）を含むことが可能であり、1bpのインクリメントでゲノムにわたりスライドさせることができる。一部の実施形態では、カーネル密度関数および約200bpのバンド幅を使用して、局所的なゲノムの偏りの推定値を、ゲノム内またはそのセグメント中の各塩基について決定する。

【0236】

一部の実施形態では、局所的なゲノムの偏りの推定値は、局所的なGC含有量および／または局所的なGC含有量の表示である。本明細書で使用される「局所」という用語（例えば、局所的な偏り、局所的な偏りの推定値、局所的な偏りの含有量、局所的なゲノムの

10

20

30

40

50

偏り、局所的なGC含有量などについて記載するのに使用される)は、10,000bpまたはそれ未満のポリヌクレオチドセグメントを指す。一部の実施形態では、「局所」という用語は、5000bpまたはそれ未満、4000bpまたはそれ未満、3000bpまたはそれ未満、2000bpまたはそれ未満、1000bpまたはそれ未満、500bpまたはそれ未満、250bpまたはそれ未満、200bpまたはそれ未満、175bpまたはそれ未満、150bpまたはそれ未満、100bpまたはそれ未満、75bpまたはそれ未満、または50bpまたはそれ未満のポリヌクレオチドセグメントを指す。局所的なGC含有量は、ゲノム、配列の読取り、配列の読取りアセンブリー(例えば、コンテイング、プロファイルなど)の局所セグメントについてのGC含有量の表示(例えば、数学的表示、定量的表示)であることが多い。例えば、局所的なGC含有量は、局所的なGCの偏りの推定値の場合もあり、局所的なGC密度の場合もある。

10

【0237】

1つまたは複数のGC密度は、参照または試料(例えば、試験試料)のポリヌクレオチドについて決定することが多い。一部の実施形態では、GC密度は、局所的な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メントについての)の表示(例えば、数学的表示、定量的表示)である。一部の実施形態では、GC密度は、局所的なゲノムの偏りの推定値である。GC密度は、本明細書で記載される適切な処理および/または当技術分野で公知の適切な処理を使用して決定することができる。GC密度は、適切なPDF(例えば、カーネル密度関数(例えば、エパネクニコフカーネル密度関数、例えば、図33を参照されたい))を使用して決定することができる。一部の実施形態では、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では、GC密度は、1つまたは複数のグアニン(G)ヌクレオチドおよび/またはシトシン(C)ヌクレオチドの存在または非存在により規定する。逆に、一部の実施形態では、GC密度は、1つまたは複数のアデニン(A)ヌクレオチドおよび/またはチミジン(T)ヌクレオチドの存在または非存在により規定することもできる。一部の実施形態では、局所的なGC含有量についてのGC密度を、全ゲノム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;例えば、図34を参照されたい)について決定されたGC密度に対して正規化する。1つまたは複数のGC密度は、試料(例えば、試験試料)または参照試料のポリヌクレオチドについて決定することができる。GC密度は、参照ゲノムについて決定することが多い。一部の実施形態では、GC密度を、参照ゲノムに従って、配列の読取りについて決定する。読取りのGC密度は、読取りがマッピングされる参照ゲノムの対応する位置および/または場所について決定されたGC密度に従って決定することが多い。一部の実施形態では、参照ゲノム上の位置について決定されたGC密度を、読取りについて割り当て、かつ/または提示するが、ここで、読取りまたはそのセグメントは、同じ参照ゲノム上の位置へとマッピングされる。任意の適切な方法を使用して、読取りについてのGC密度を生成する目的で、参照ゲノム上にマッピングした読取りの位置を決定することができる。一部の実施形態では、マッピングした読取りの場所の中央値により、参照ゲノム上の位置であって、それに由来する読取りについてのGC密度を決定する位置が決定される。例えば、読取りの場所の中央値が、参照ゲノムの塩基番号xにおける第12染色体へとマッピングされる場合、読取りのGC密度は、参照ゲノムの塩基番号xまたはその近傍における第12染色体上に位置する場所についてのカーネル密度推定により決定されるGC密度として提示されることが多い。一部の実施形態では、GC密度を、参照ゲノムに従った、読取りの一部または全部の塩基の場所について決定する。場合によって、読取りのGC密度は、参照ゲノム上の複数の塩基の場所について決定された、2つまたはそれ超のGC密度の平均、合計、中央値、または積分を含む。

20

30

40

【0238】

一部の実施形態では、局所的なゲノムの偏りの推定(例えば、GC密度)は、値として定量および/または提示される。局所的なゲノムの偏りの推定(例えば、GC密度)は、場合によって、平均、平均値、および/または中央値として表される。局所的なゲノムの偏りの推定(例えば、GC密度)は、場合によって、PDEの最大ピーク高さとして表さ

50

れる。場合によって、局所的なゲノムの偏りの推定（例えば、G C 密度）は、適切な P D E の合計または積分（例えば、曲線下面積（A U C））として表される。一部の実施形態では、G C 密度は、カーネル重みを含む。ある特定の実施形態では、読取りの G C 密度は、カーネル重みの平均、平均値、合計、中央値、最大ピーク高さ、または積分とほぼ等しい値を含む。

【0239】

偏り頻度

偏り頻度は、場合によって、1つまたは複数の局所的なゲノムの偏りの推定値（例えば、G C 密度）に従って決定される。偏り頻度は、場合によって、試料、参照（例えば、参照ゲノム、参照配列）、またはこれらの一部についての局所的なゲノムの偏りの推定値の発生数のカウントまたは合計である。偏り頻度は、場合によって、試料、参照、またはこれらの一部についての、局所的なゲノムの偏りの推定値（例えば、各局所的なゲノムの偏りの推定値）の発生数のカウントまたは合計である。一部の実施形態では、偏り頻度は、G C 密度頻度である。G C 密度頻度は、1つまたは複数の G C 密度に従って決定することが多い。例えば、G C 密度頻度は、値 x の G C 密度が全ゲノムまたはそのセグメントにわたり表示される回数を表示しうる。偏り頻度は、局所的なゲノムの偏りの推定値の分布であることが多く、ここで、各局所的なゲノムの偏りの推定値の発生数は、偏り頻度として表示される（例えば、図 35 を参照されたい）。偏り頻度は、場合によって、数学的に操作および／または正規化される。偏り頻度は、適切な方法により数学的に操作および／または正規化することができる。一部の実施形態では、偏り頻度を、試料、参照、またはこれらの一部についての、局所的なゲノムの偏りの推定値（例えば、常染色体、染色体のサブセット、単一の染色体、またはこれらの読取り）の表示（例えば、フラクション、百分率）に対して正規化する。偏り頻度は、試料または参照の、一部または全部の局所的なゲノムの偏りの推定値について決定することができる。一部の実施形態では、偏り頻度は、試験試料の、一部または全部の配列の読取りについての、局所的なゲノムの偏りの推定値について決定することができる。

【0240】

一部の実施形態では、システムは、偏り密度モジュール 6 を含む。偏り密度モジュールは、マッピングした配列の読取り 5 および参照配列 2 を、任意の適切なフォーマットで受容、回収、および／または保存し、局所的なゲノムの偏りの推定値、局所的なゲノムの偏り分布、偏り頻度、G C 密度、G C 密度分布、および／または G C 密度頻度（併せて、ボックス 7 により表示された）を生成することが可能である。一部の実施形態では、偏り密度モジュールにより、データおよび／または情報（例えば、7）を、別の適切なモジュール（例えば、関係モジュール 8）へと転送する。

【0241】

関係

一部の実施形態では、1つまたは複数の関係を、局所的なゲノムの偏りの推定値と、偏り頻度との間で生成する。本明細書で使用される「関係」という用語は、2つまたはそれ超の変数または値の間の数学的関係および／またはグラフ的关系を指す。関係は、適切な数学的処理および／またはグラフ的処理により生成することができる。関係の非限定的な例は、関数、相関、分布、線形等式または非線形等式、直線、回帰、適合させた回帰など、またはこれらの組合せの数学的表示および／またはグラフ表示を含む。場合によって、関係は、適合させた関係を含む。一部の実施形態では、適合させた関係は、適合させた回帰を含む。場合によって、関係は、2つまたはそれ超の変数または値であって、重み付き変数または重み付き値を含む。一部の実施形態では、関係は、適合させた回帰を含み、ここで、関係の 1つまたは複数の変数または値が重み付けされている。場合によって、回帰は、重み付き様式で適合させる。場合によって、回帰は、重み付けされずに適合させる。ある特定の実施形態では、関係の生成は、プロットングまたはグラフ作成を含む。

【0242】

一部の実施形態では、適切な関係を、局所的なゲノムの偏りの推定値と、偏り頻度との

間で決定する。一部の実施形態では、試料についての (i) 局所的なゲノムの偏りの推定値と、(i i) 偏り頻度との関係を生成することにより、試料偏り関係を提示する。一部の実施形態では、参照についての (i) 局所的なゲノムの偏りの推定値と、(i i) 偏り頻度との関係を生成することにより、参照偏り関係を提示する。ある特定の実施形態では、関係を、GC密度とGC密度頻度との間で生成する。一部の実施形態では、試料についての (i) GC密度と、(i i) GC密度頻度との関係を生成することにより、試料GC密度関係を提示する。一部の実施形態では、参照についての (i) GC密度と、(i i) GC密度頻度との関係を生成することにより、参照GC密度関係を提示する。一部の実施形態では、局所的なゲノムの偏りの推定値がGC密度である場合、試料偏り関係は、試料GC密度関係であり、参照偏り関係は、参照GC密度関係である。参照GC密度関係および／または試料GC密度関係のGC密度は、局所的なGC含有量についての表示（例えば、数学的表示または定量的表示）であることが多い。一部の実施形態では、局所的なゲノムの偏りの推定値と偏り頻度との関係は、分布を含む。一部の実施形態では、局所的なゲノムの偏りの推定値と偏り頻度との関係は、適合させた関係（例えば、適合させた回帰）を含む。一部の実施形態では、局所的なゲノムの偏りの推定値と偏り頻度との関係は、線形適合回帰または非線形適合回帰（例えば、多項式回帰）を含む。ある特定の実施形態では、局所的なゲノムの偏りの推定値と偏り頻度との関係は、重み付き関係を含み、ここで、局所的なゲノムの偏りの推定値および／または偏り頻度は、適切な処理により重み付けされる。一部の実施形態では、重み付き適合させた関係（例えば、重み付き適合）は、四分位回帰、パラメータ付きの確率分布、または補間を有する経験的分布を含む処理により得ることができる。ある特定の実施形態では、試験試料、参照、またはこれらの一部についての、局所的なゲノムの偏りの推定値と偏り頻度との関係は、多項式回帰を含み、局所的なゲノムの偏りの推定値は、重み付けされている。一部の実施形態では、重み付き適合モデルは、分布値を重み付けすることを含む。分布値は、適切な処理により重み付けすることができる。一部の実施形態では、分布のテールの近傍に位置する値には、分布中央値に近い値より小さな重みを施す。例えば、局所的なゲノムの偏りの推定値（例えば、GC密度）と、偏り頻度（例えば、GC密度頻度）との分布については、重みを、所与の局所的なゲノムの偏りの推定値についての偏り頻度に従って決定し、ここで、分布の平均値に近接した偏り頻度を含む局所的なゲノムの偏りの推定値には、平均値から遠い偏り頻度を含む局所的なゲノムの偏りの推定値より大きな重みを施す。

【0243】

一部の実施形態では、システムは、関係モジュール8を含む。関係モジュールにより、関係のほか、関係を規定する関数、係数、定数、および変数を生成することができる。関係モジュールにより、データおよび／または情報（例えば、7）を、適切なモジュール（例えば、偏り密度モジュール6）から受容、保存、および／または回収し、関係を生成することができる。関係モジュールにより、局所的なゲノムの偏りの推定値の分布を生成および比較することが多い。関係モジュールにより、データセットを比較し、場合によって、回帰および／または適合させた関係を生成することができる。一部の実施形態では、関係モジュールにより、1つまたは複数の分布（例えば、試料および／または参照の局所的なゲノムの偏りの推定値の分布）を比較し、配列の読取りのカウントについての重み付け係数および／または重み割当て9を、別の適切なモジュール（例えば、偏り補正モジュール）へと提示する。場合によって、関係モジュールにより、正規化された配列の読取りのカウントを、分布モジュール21へと直接提示し、ここで、カウントを、関係および／または比較に対して正規化する。

【0244】

比較の生成およびその使用

一部の実施形態では、配列の読取り中の局所的な偏りを低減するための処理は、配列の読取りのカウントを正規化することを含む。配列の読取りのカウントは、試験試料の参照との比較に対して正規化されることが多い。例えば、場合によって、配列の読取りのカウントは、試験試料の配列の読取りの局所的なゲノムの偏りの推定値を、参照（例えば、参

照ゲノムまたはその一部)の局所的なゲノムの偏りの推定値と比較することにより正規化する。一部の実施形態では、配列の読取りのカウントは、試験試料の局所的なゲノムの偏りの推定値の偏り頻度を、参照の局所的なゲノムの偏りの推定値の偏り頻度と比較することにより正規化する。一部の実施形態では、配列の読取りのカウントは、試料偏り関係と参照偏り関係とを比較することにより正規化し、これにより、比較を生成する。

【0245】

配列の読取りのカウントは、2つまたはそれ超の関係の比較に対して正規化されることが多い。ある特定の実施形態では、2つまたはそれ超の関係について比較し、これにより、配列の読取り中の局所的な偏りを低減する(例えば、カウントを正規化する)ために使用される比較を提示する。適切な方法により、2つまたはそれ超の関係について比較することができる。一部の実施形態では、比較は、第1の関係に第2の関係を加算すること、第1の関係から第2の関係を減算すること、第1の关系到第2の関係を乗算すること、および/または第1の関係を第2の関係で除算することを含む。ある特定の実施形態では、2つまたはそれ超の関係の比較は、適切な線形回帰および/または非線形回帰の使用を含む。ある特定の実施形態では、2つまたはそれ超の関係の比較は、適切な多項式回帰(例えば、三次多項式回帰)を含む。一部の実施形態では、比較は、第1の回帰に第2の回帰を加算すること、第1の回帰から第2の回帰を減算すること、第1の回帰に第2の回帰を乗算すること、および/または第1の回帰を第2の回帰で除算することを含む。一部の実施形態では、2つまたはそれ超の関係について、多重回帰の推論フレームワークを含む処理により比較する。一部の実施形態では、2つまたはそれ超の関係について、適切な多変量分析を含む処理により比較する。一部の実施形態では、2つまたはそれ超の関係について、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処理により比較する。

【0246】

ある特定の実施形態では、試験試料および参照についての偏り頻度を含む、局所的なゲノムの偏りの推定値の分布を、多項式回帰を含む処理により比較するが、ここで、局所的なゲノムの偏りの推定値は、重み付けされている。一部の実施形態では、多項式回帰を、(i)比の各々が、参照の局所的なゲノムの偏りの推定値の偏り頻度および試料の局所的なゲノムの偏りの推定値の偏り頻度を含む比と、(i i)局所的なゲノムの偏りの推定値との間で生成する。一部の実施形態では、多項式回帰を、(i)参照の局所的なゲノムの偏りの推定値の偏り頻度の、試料の局所的なゲノムの偏りの推定値の偏り頻度に対する比と、(i i)局所的なゲノムの偏りの推定値との間で生成する。一部の実施形態では、試験試料および参照の読取りについての局所的なゲノムの偏りの推定値の分布の比較は、参照および試料についての、局所的なゲノムの偏りの推定値の偏り頻度の対数比(例えば、 $\log 2$ 比)を決定することを含む。一部の実施形態では、局所的なゲノムの偏りの推定値の分布の比較は、参照についての、局所的なゲノムの偏りの推定値の偏り頻度対数比(例えば、 $\log 2$ 比)を、試料についての局所的なゲノムの偏りの推定値の偏り頻度の対数比(例えば、 $\log 2$ 比)で除算することを含む(例えば、実施例7および図36を参照されたい)。

【0247】

比較に従ったカウントを正規化することでは、あるカウントは調整されるが、他のカウントは調整されないことが典型的である。カウントを正規化することでは、ある場合には、全カウントが調整され、ある場合には、いかなる配列の読取りのカウントも調整されない。配列の読取りについてのカウントは、ある場合には、重み付け係数を決定することを含む処理により正規化し、ある場合には、処理は、重み付け係数の直接的な生成および活用を含まない。比較に従ったカウントを正規化することは、場合によって、各配列の読取りのカウントについての重み付け係数を決定することを含む。重み付け係数は、配列の読取りに特異的であり、特異的配列の読取りのカウントへと適用されることが多い。重み付け係数は、2つまたはそれ超の偏り関係の比較(例えば、参照偏り関係と比較した試料偏

10

20

30

40

50

り関係)に従って決定することが多い。正規化されたカウントは、カウント値を、重み付け係数に従って調整することにより決定することが多い。重み付け係数に従ったカウントの調整は、場合によって、配列の読取りについてのカウントに重み付け係数を加算すること、配列の読取りについてのカウントから重み付け係数を減算すること、配列の読取りについてのカウントに重み付け係数を乗算すること、および／または配列の読取りについてのカウントを重み付け係数で除算することを含む。重み付け係数および／または正規化されたカウントは、場合によって、回帰(例えば、回帰直線)から決定する。正規化されたカウントは、場合によって、参照の局所的なゲノムの偏りの推定値の偏り頻度(例えば、参照ゲノム)と、試験試料の局所的なゲノムの偏りの推定値の偏り頻度との比較の結果として得られる、回帰直線(例えば、適合させた回帰直線)から直接得る。一部の実施形態では、試料の読取りの各カウントを、(i)読取りの局所的なゲノムの偏りの推定値の偏り頻度の、(ii)参照の局所的なゲノムの偏りの推定値の偏り頻度と比較した比較に対して、正規化されたカウント値として提示する。ある特定の実施形態では、試料について得られる配列の読取りのカウントを正規化し、配列の読取り中の偏りを低減する。

10

【0248】

場合によって、システムは、偏り補正モジュール10を含む。一部の実施形態では、偏り補正モジュールの機能は、関係モデル化モジュール8により果たされる。偏り補正モジュールにより、マッピングした配列の読取りおよび重み付け係数(例えば、9)を、適切なモジュール(例えば、関係モジュール8、圧縮モジュール4)から受容、回収、および／または保存することができる。一部の実施形態では、偏り補正モジュールにより、マッピングした読取りへとカウントを提示する。一部の実施形態では、偏り補正モジュールにより、重み割当ておよび／または偏り補正因子を、配列の読取りのカウントへと適用し、これにより、正規化および／または調整されたカウントを提示する。偏り補正モジュールにより、正規化されたカウントを、別の適切なモジュール(例えば、分布モジュール21)へと提示することが多い。

20

【0249】

ある特定の実施形態では、カウントを正規化することは、GC密度に加えた、1つまたは複数の特徴を因子分解することと、配列の読取り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では、カウントを正規化することは、1つまたは複数の異なる局所的なゲノムの偏りの推定値を因子分解することと、配列の読取り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では、配列の読取りのカウントを、1つまたは複数の特徴(例えば、1つまたは複数の偏り)に従って決定された重み付けに従って重み付けする。一部の実施形態では、カウントを、1つまたは複数の組み合わせられた重みに対して正規化する。場合によって、1つまたは複数の組み合わせられた重みに従って、1つまたは複数の特徴を因子分解することおよび／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処理を介する。任意の適切な多変量モデルを使用して、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は、多変量線形回帰、多変量四分位回帰、経験データの多変量補間、非線形多変量モデルなど、またはこれらの組合せを含む。

30

【0250】

一部の実施形態では、システムは、多変量補正モジュール13を含む。多変量補正モジュールは、偏り密度モジュール6、関係モジュール8、および／または偏り補正モジュール10の機能を、複数回にわたり果たし、これにより、複数の偏りについてのカウントを調整することができる。一部の実施形態では、多変量補正モジュールは、1つまたは複数の偏り密度モジュール6、関係モジュール8、および／または偏り補正モジュール10を含む。場合によって、多変量補正モジュールにより、正規化されたカウント11を、別の適切なモジュールへと提示する(例えば、分布モジュール21)。

40

【0251】

重み付き部分

一部の実施形態では、部分を、重み付けする。一部の実施形態では、1つまたは複数の部分を、重み付けし、これにより、重み付き部分を提示する。重み付き部分は、場合によ

50

って、部分依存性を除去する。部分は、適切な処理により重み付けすることができる。一部の実施形態では、1つまたは複数の部分を、固有関数 (eigen function (またはeigenfunction)) により重み付けする。一部の実施形態では、固有関数は、部分を直交固有部分により置きかえることを含む。一部の実施形態では、システムは、部分重み付けモジュール42を含む。一部の実施形態では、重みモジュールにより、読取り密度、読取り密度プロファイル、および／または調整された読取り密度プロファイルを受容、回収、および／または保存する。一部の実施形態では、重み付け部分を、部分重み付けモジュールにより提示する。一部の実施形態では、重みモジュールは、部分を重み付けするように必要とされる。重みモジュールでは、当技術分野で公知であるかまたは本明細書で記載される1つまたは複数の重み付け法により、部分を重み付けすることができる。重みモジュールにより、重み付け部分を、別の適切なモジュール (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) へと提示することが多い。

【0252】

主成分分析

一部の実施形態では、読取り密度プロファイル (例えば、試験試料 (例えば、図39A) の読取り密度プロファイル) を、主成分分析 (PCA: principal component analysis) に従って調整する。1もしくは複数の参照試料の読取り密度プロファイルおよび／または試験被験体の読取り密度プロファイルは、PCAに従って調整することができる。本明細書では、場合によって、PCA関連処理を介する、読取り密度プロファイルからの偏りの除去を、プロファイルの調整と称する。PCAは、適切なPCA法またはその変化形により実施することができる。PCA法の非限定的な例は、カノニカル相関分析 (CCA)、KL (Karhunen-Loeve) 変換 (KLT)、ホテリング変換、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、エッカートヤングの定理、シュミットミルスキーの定理、経験的直交関数 (EOF)、経験的固有関数分解、経験的成分分析、準調和モード、スペクトル分解、経験的モード分析など、これらの変化形または組合せを含む。PCAにより、読取り密度プロファイル中の1つまたは複数の偏りを同定することが多い。本明細書では、場合によって、PCAにより同定された偏りを、主成分と称する。一部の実施形態では、適切な方法を使用して、1つまたは複数の主成分に従って読取り密度プロファイルを調整することにより、1つまたは複数の偏りを除外することができる。読取り密度プロファイルは、読取り密度プロファイルに1つまたは複数の主成分を加算すること、読取り密度プロファイルから1つまたは複数の主成分を減算すること、読取り密度プロファイルに1つまたは複数の主成分を乗算すること、および／または読取り密度プロファイルを1つまたは複数の主成分で除算することにより調整することができる。一部の実施形態では、1つまたは複数の主成分を、読取り密度プロファイルから減算することにより、1つまたは複数の偏りを、読取り密度プロファイルから除外することができる。読取り密度プロファイル中の偏りは、プロファイルのPCAにより同定および／または定量されることが多いが、主成分は、読取り密度のレベルでプロファイルから減算されることが多い。PCAにより、1つまたは複数の主成分を同定することが多い。一部の実施形態では、PCAにより、第1、第2、第3、第4、第5、第6、第7、第8、第9、および第10、またはそれ超の順位の主成分を同定する。ある特定の実施形態では、1、2、3、4、5、6、7、8、9、10またはそれ超の主成分を使用して、プロファイルを調整する。主成分は、PCA中のそれらの出現の順序でプロファイルを調整するのに使用することが多い。例えば、3つの主成分を、読取り密度プロファイルから減算する場合、第1、第2、および第3の主成分を使用する。場合によって、主成分により同定される偏りは、プロファイルの特徴であって、プロファイルを調整するのに使用されない特徴を含む。例えば、PCAにより、主成分としての遺伝子の変動 (例えば、異数性、微小重複、微小欠失、欠失、転位、挿入) および／または性差 (例えば、図38Cで見られる) を同定する。したがって、一部の実施形態では、1つまたは複数の主成分は、プロファイルを調整するのに使用さ

10

20

30

40

50

れない。例えば、場合によって、第1、第2、および第4の主成分を使用して、プロファイル进行调整するが、ここで、第3の主成分は、プロファイル进行调整するのに使用されない。主成分は、任意の適切な試料または参照を使用して、PCAから得ることができる。一部の実施形態では、主成分を、試験試料（例えば、試験被験体）から得る。一部の実施形態では、主成分を、1つまたは複数の参照（例えば、参照試料、参照配列、参照セット）から得る。例えば、図38A～Cに示される通り、PCAは、第1の主成分（図38B）および第2の主成分（図38C）の同定を結果としてもたらす複数の試料を含む訓練セット（図38A）から得られる中央値読取り密度プロファイルに対して実施される。一部の実施形態では、主成分を、問題の遺伝子の変動を欠くことが既知である被験体のセットから得る。一部の実施形態では、主成分を、公知の正倍数体のセットから得る。主成分は、参照の1つまたは複数の読取り密度プロファイル（例えば、訓練セット）を使用して実施されるPCAに従って同定することが多い。参照から得られる1つまたは複数の主成分を、試験被験体の読取り密度プロファイル（例えば、図39B）から減じ、これにより、調整プロファイル（例えば、図39C）を提示することが多い。

10

【0253】

一部の実施形態では、システムは、PCA統計モジュール33を含む。PCA統計モジュールにより、読取り密度プロファイルを、別の適切なモジュール（例えば、プロファイル生成モジュール26）から受容するおよび／または回収することができる。PCAは、PCA統計モジュールにより実施することが多い。PCA統計モジュールにより、読取り密度プロファイルを受容、回収、および／または保存し、読取り密度プロファイルを、参照セット32、訓練セット30、および／または1もしくは複数の試験被験体28から処理することが多い。PCA統計モジュールにより、主成分を生成および／もしくは提示し、かつ／または1つまたは複数の主成分に従って、読取り密度プロファイルを調整することができる。調整された読取り密度プロファイル（例えば、40、38）は、PCA統計モジュールによりもたらされることが多い。PCA統計モジュールにより、調整された読取り密度プロファイル（例えば、38、40）を、別の適切なモジュール（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）へと提示および／または転送することができる。一部の実施形態では、PCA統計モジュールにより、性別判定36を提示することができる。性別決定は、場合によって、PCAに従って、かつ／または1もしくは複数の主成分に従って決定された、胎児の性別の決定である。一部の実施形態では、PCA統計モジュールは、下記に示されるRコードの一部、全部、または1つの修飾を含む。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、中央値を減算すること、部分をフィルタリングすること、および極値をトリミングすること）で始まる。

20

30

【化 2】

#PCAのために、データの外れ値をクリーニングする
 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
}
```

10

次いで、主成分を計算する。

```
#主成分を計算する
pc <- prcomp(dclean)$x
```

最後に、各試料のPCA調整プロファイルは、

```
#残差を計算する
mm <- model.matrix(~pc[,1:numpc])
for (j in 1:ncol(dclean))
  dclean[,j] <- dclean[,j] - predict(lm(dclean[,j]~mm))
```

20

で計算することができる。

【0 2 5 4】

プロファイルの比較

一部の実施形態では、アウトカムの決定は、比較を含む。ある特定の実施形態では、読取り密度プロファイルまたはその部分を活用して、アウトカムを提示する。一部の実施形態では、アウトカムの決定（例えば、遺伝子の変動の存在または非存在の決定）は、2つまたはそれ超の読取り密度プロファイルの比較を含む。読取り密度プロファイルの比較は、選択されたゲノムのセグメントについてなされた読取り密度プロファイルの比較を含むことが多い。例えば、試験プロファイルは、参照プロファイルと比較することが多く、試験プロファイルおよび参照プロファイルを、実質的に同じセグメントであるゲノムのセグメント（例えば、参照ゲノム）について決定した。読取り密度プロファイルの比較は、場合によって、読取り密度プロファイルの部分の2つまたはそれ超のサブセットの比較を含む。読取り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）を表しうる。読取り密度プロファイルは、部分の任意の量のサブセットを含みうる。場合によって、読取り密度プロファイルは、2つもしくはそれ超、3つもしくはそれ超、4つもしくはそれ超、または5つもしくはそれ超のサブセットを含む。ある特定の実施形態では、読取り密度プロファイルは、部分の2つのサブセットを含み、ここで、各部分は、隣接する参照ゲノムのセグメントを表示する。一部の実施形態では、試験プロファイルを、参照プロファイルと比較することができ、ここで、試験プロファイルおよび参照プロファイルはいずれも、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、ここで、第1のサブセットおよび第2のサブセットは、ゲノムの異なるセグメントを表示する。読取り密度プロファイルの部分のあるサブセットは、遺伝子の変動を含むことが可能であり、他の部分のサブセットは、場合によって、遺伝子の変動を実質的に含まない。場合によって、プロファイル（例えば、試験プロファイル）の部分の全てのサブセットは、遺伝子の変動を実質的に含まない。場合によって、プロファイル（例えば、試験プロファイル）の部分の全てのサブセットは、遺伝子の変動を含む。一部の

30

40

50

実施形態では、試験プロファイルは、遺伝子の変動を含む部分の第1のサブセット、および遺伝子の変動を実質的に含まない部分の第2のサブセットを含みうる。

【0255】

一部の実施形態では、本明細書で記載される方法は、比較（例えば、試験プロファイルを参照プロファイルと比較すること）をあらかじめ形成することを含む。適切な方法により、2つもしくはそれ超のデータセット、2つもしくはそれ超の関係、および／または2つもしくはそれ超のプロファイルについて比較することができる。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切な統計学的方法の非限定的な例は、ベールンフィッシャー法、ブートストラップ法、独立の有意性検定を組み合わせるためのフィッシャー法、ネイマンピアソン検定、確認的データ分析、調査的データ分析、正確検定、F検定、Z検定、T検定、不確定性の尺度、帰無仮説、対立帰無（*counternull*）などの計算および／もしくは比較、カイ二乗検定、オムニバス検定、有意性（例えば、統計学的有意性）のレベルの計算および／もしくは比較、メタ分析、多変量分析、回帰、単一線形回帰、頑健な線形回帰など、または前出の組合せを含む。ある特定の実施形態では、2つまたはそれ超のデータセット、関係、および／またはプロファイルの比較は、不確定性の尺度の決定および／または比較を含む。本明細書で使用される「不確定性の尺度」とは、有意性（例えば、統計学的有意性）の尺度、誤差の尺度、分散の尺度、信頼性の尺度など、またはこれらの組合せを指す。不確定性の尺度は、値（例えば、閾）の場合もあり、値の範囲（例えば、区間、信頼区間、ベイズ信頼区間、閾範囲）の場合もある。不確定性の尺度の非限定的な例は、p値、偏差の適切な尺度（例えば、標準偏差、シグマ、絶対偏差、平均絶対偏差など）、適切な誤差の尺度（例えば、標準誤差、二乗平均誤差、二乗平均平方根誤差など）、分散の適切な尺度、適切な標準スコア（例えば、標準偏差、累積百分率、百分位数同等物、Zスコア、Tスコア、Rスコア、標準的9段階法（スタナイン）、スタナインパーセントなど）など、またはこれらの組合せを含む。一部の実施形態では、有意性のレベルの決定は、不確定性の尺度（例えば、p値）を決定することを含む。ある特定の実施形態では、2つまたはそれ超のデータセット、関係、および／またはプロファイルは、複数の（例えば、2つまたはそれ超の）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、バグギング、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近傍法、ロジスティック回帰および／またはLOESSスムージング）、ならびに／または任意の適切な数学的操作および／もしくは統計学的操作（例えば、本明細書では操作と称する）を活用することにより分析および／または比較することができる。

【0256】

ある特定の実施形態では、2つまたはそれ超の読取り密度プロファイルの比較は、2つまたはそれ超の読取り密度プロファイルについての、不確定性の尺度の決定および／または比較を含む。場合によって、読取り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を比較して、データセットの数学的操作および／もしくは統計学的操作の解釈を容易とし、かつ／またはアウトカムを提示する。場合によって、試験被験体について生成された読取り密度プロファイルは、1つまたは複数の参照（例えば、参照試料、参照被験体など）について生成された読取り密度プロファイルと比較する。一部の実施形態では、アウトカムを、試験被験体に由来する読取り密度プロファイルの、染色体、部分、またはこれらのセグメントについての参照に由来する読取り密度プロファイルとの比較により提示し、ここで、参照の読取り密度プロファイルは、遺伝子の変動を保有しないことが既知である、参照被験体のセット（例えば、参照）から得る。一部の実施形態では、アウトカムを、試験被験体に由来する読取り密度プロファイルの、染色体、部分、またはこれらのセグメントについての参照に由来する読取り密度プロファイルとの比較により提示し、ここで、参照の読取り密度プロファイルは、特異的遺伝子の変動（例えば、染色体の異数性、トリソミー、微小重複、微小欠失）を保有することが既知である、参照被験体のセットから得られる。

【0257】

ある特定の実施形態では、試験被験体の読取り密度プロファイルは、遺伝子の変動の非存在を表示する所定の値と比較され、場合によって、遺伝子の変動が位置するゲノム位置に対応する1つまたは複数のゲノム位置（例えば、部分）において、所定の値から逸脱する。例えば、試験被験体（例えば、遺伝子の変動と関連する医学的状態の危険性があるか、またはこれを患っている被験体）では、読取り密度プロファイルは、試験被験体が、問題の遺伝子の変動を含む場合の選択部分について、参照の読取り密度プロファイル（例えば、参照配列、参照被験体、参照セット）から有意に異なることが期待される。試験被験体の読取り密度プロファイルは、試験被験体が、問題の遺伝子の変動を含まない場合の選択部分について、参照の読取り密度プロファイル（例えば、参照配列、参照被験体、参照セット）と実質的に同じであることが多い。読取り密度プロファイルは、所定の閾および／または閾範囲と比較されることが多い（例えば、図40を参照されたい）。本明細書で使用される「閾」という用語は、定性的データセットを使用して計算され、遺伝子の変動（例えば、コピー数の変動、異数性、染色体の異常、微小重複、微小欠失など）についての診断の限界として用いられる、任意の数を指す。ある特定の実施形態では、閾は、本明細書で記載される方法により得られる結果により超えられ、被験体は、遺伝子の変動（例えば、トリソミー）を有すると診断される。一部の実施形態では、閾値または閾値の範囲は、配列の読取りデータ（例えば、参照および／または被験体に由来する）を、数学的および／または統計学的に操作することを介して計算されることが多い。遺伝子の変動の存在または非存在を指し示す所定の閾または閾の範囲は、遺伝子の変動の存在または非存在を決定するために有用なアウトカムをやはり提示しながらも、変化しうる。ある特定の実施形態では、正規化された読取り密度および／または正規化されたカウントを含む読取り密度プロファイルを生成して、アウトカムの分類および／または提示を容易とする。アウトカムは、正規化されたカウントを含む読取り密度プロファイルのプロットに基づき（例えば、このような読取り密度プロファイルのプロットを使用して）提示することができる。

【0258】

一部の実施形態では、システムは、スコアリングモジュール46を含む。スコアリングモジュールは、読取り密度プロファイル（例えば、調整された、正規化された読取り密度プロファイル）を、別の適切なモジュール（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）から受容、回収、および／または保存しうる。スコアリングモジュールは、2つまたはそれ超の読取り密度プロファイル（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）を受容、回収、保存、および／または比較しうる。スコアリングモジュールにより、スコア（例えば、プロット、プロファイル統計、比較（例えば、2つまたはそれ超のプロファイルの間の差違）、Zスコア、不確定性の尺度、判定域、試料判定50（例えば、遺伝子の変動の存在または非存在の決定）、および／またはアウトカム）を提示しうることが多い。スコアリングモジュールにより、スコアを、末端使用者および／または別の適切なモジュール（例えば、ディスプレイ、プリンターなど）へと提示することができる。一部の実施形態では、スコアリングモジュールは、下記に示されるRコードであって、具体的な検定（例えば、第21染色体カウントが大きいこと）のためのカイ二乗統計を計算するためのR関数を含むRコードの一部、全部、または1つの修飾を含む。

3つのパラメータは、

x = 試料の読取りデータ（部分xの試料）

m = 部分についての中央値

y = 検定ベクター（例えば、第21染色体について真であることを除き、全ての部分について偽）

である。

【化 3】

```

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)
}

```

10

【0 2 5 9】

回帰のハイブリッド正規化

一部の実施形態では、ハイブリッド正規化法を使用する。一部の実施形態では、ハイブリッド正規化法により、偏り（例えば、GCの偏り）を低減する。一部の実施形態では、ハイブリッド正規化は、（i）2つの変数（例えば、カウントおよびGC含有量）の関

係についての分析と、（ii）分析に従った正規化法の選択および適用とを含む。ある特定の

の実施形態では、ハイブリッド正規化は、（i）回帰（例えば、回帰分析）と、（ii）

回帰に従った正規化法の選択および適用とを含む。一部の実施形態では、第1の試料につ

いて得られたカウント（例えば、第1の試料セット）を、別の試料（例えば、第2の試料

セット）から得られるカウントとは異なる方法により正規化する。一部の実施形態では、

第1の試料について得られたカウント（例えば、第1の試料セット）を、第1の正規化法

により正規化し、第2の試料（例えば、第2の試料セット）から得られるカウントを、第

2の正規化法により正規化する。例えば、ある特定の実施形態では、第1の正規化法は、

線形回帰の使用を含み、第2の正規化法は、非線形回帰（例えば、LOESS、GC-LOESS、LOWESS回帰、LOESSスムージング）の使用を含む。

20

【0 2 6 0】

一部の実施形態では、ハイブリッド正規化法を使用して、ゲノムまたは染色体の部分へとマッピングした配列の読取り（例えば、カウント、マッピングしたカウント、マッピングした読取り）を正規化する。ある特定の実施形態では、未加工のカウントを正規化し、一部の実施形態では、調整されるか、重み付けされるか、フィルタリングされるか、または既に正規化されたカウントを、ハイブリッド正規化法により正規化する。ある特定の

の実施形態では、ゲノム区分のレベルまたはZスコアを、正規化する。一部の実施形態では、

選択されたゲノム部分または染色体へとマッピングしたカウントを、ハイブリッド正規化

法により正規化する。カウントは、ゲノムの部分へとマッピングした配列の読取りの適切

な尺度であって、その非限定的な例が、未加工のカウント（例えば、処理されていないカ

ウント）、正規化されたカウント（例えば、PERUN、ChAI、または適切な方法に

より正規化された）、部分レベル（例えば、平均レベル、平均値レベル、中央値レベルな

ど）、Zスコアなど、またはこれらの組合せを含む尺度を指す場合がある。カウントは、

1つまたは複数の試料（例えば、試験試料、妊娠中の雌による試料）に由来する未加工の

カウントの場合もあり、処理されたカウントの場合もある。一部の実施形態では、カウ

ントを、1つまたは複数の被験体から得られる1つまたは複数の試料から得る。

30

40

【0 2 6 1】

一部の実施形態では、正規化法（例えば、正規化法の種類）を、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する。回帰分析とは、変数（例えば、カウントおよびGC含有量）間の関係を推定するための統計学的技法を指す。一部の実施形態では、回帰を、参照ゲノムの複数の部分のうちの各部分についてのGC含有量のカウントおよ

50

び尺度に従って生成する。G C含有量の適切な尺度であって、その非限定的な例が、グアニン含有量、シトシン含有量、アデニン含有量、チミン含有量、プリン（G C）含有量、またはピリミジン（A TまたはA T U）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、またはこれらの組合せを含む尺度を使用することができる。グアニン（G）含有量、シトシン（C）含有量、アデニン（A）含有量、チミン（T）含有量、プリン（G C）含有量、またはピリミジン（A TまたはA T U）含有量の尺度は、比または百分率として表すことができる。一部の実施形態では、任意の適する比または百分率であって、その非限定的な例が、G C / A T、G C / 全ヌクレオチド、G C / A、G C / T、A T / 全ヌクレオチド、A T / G C、A T / G、A T / C、G / A、C / A、G / T、G / A、G / A T、C / Tなど、またはこれらの組合せを含む比または百分率を使用する。一部の実施形態では、G C含有量の尺度は、G C含有量の、全ヌクレオチド含有量に対する比または百分率である。一部の実施形態では、G C含有量の尺度は、参照ゲノムの部分へとマッピングした配列の読取りについての、G C含有量の、全ヌクレオチド含有量に対する比または百分率である。ある特定の実施形態では、G C含有量は、各参照ゲノムの部分へとマッピングした配列の読取りに従って、かつ／または各参照ゲノムの部分へとマッピングした配列の読取りから決定し、配列の読取りは、試料（例えば、妊娠中の雌から得られた試料）から得る。一部の実施形態では、G C含有量の尺度は、配列の読取りに従って、かつ／または配列の読取りから決定されない。ある特定の実施形態では、G C含有量の尺度を、1つまたは複数の被験体から得られる1つまたは複数の試料について決定する。

【0262】

一部の実施形態では、回帰を生成することは、回帰分析または相関分析を生成することを含む。その非限定的な例が、回帰分析、（例えば、線形回帰分析）、適合の良さについての分析、ピアソン相関分析、ランク相関、説明されていない分散の割合、N S（N a s h－S u t c l i f f e）モデルによる効率解析、回帰モデルの検証、P R L（p r o p o r t i o n a l r e d u c t i o n i n l o s s）、二乗平均平方根偏差など、またはこれらの組合せを含む、適切な回帰を使用することができる。一部の実施形態では、回帰直線を生成する。ある特定の実施形態では、回帰を生成することは、線形回帰を生成することを含む。ある特定の実施形態では、回帰を生成することは、非線形回帰（例えば、L O E S S回帰、L O W E S S回帰）を生成することを含む。

【0263】

一部の実施形態では、回帰により、例えば、G C含有量のカウントおよび尺度の間の相関（例えば、線形相関）の存在または非存在を決定する。一部の実施形態では、回帰（例えば、線形回帰）を生成し、相関係数を決定する。一部の実施形態では、その非限定的な例が、決定係数、 R^2 値、ピアソン相関係数などを含む、適切な相関係数を決定する。

【0264】

一部の実施形態では、適合の良さを、回帰（例えば、回帰分析、線形回帰）について決定する。適合の良さは、場合によって、目視分析または数学的分析により決定する。評価は、場合によって、適合の良さが、非線形回帰で大きいのか、線形回帰で大きいのかについて決定することを含む。一部の実施形態では、相関係数は、適合の良さの尺度である。一部の実施形態では、回帰についての適合の良さの評価を、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決定する。一部の実施形態では、適合の良さの評価は、相関係数と相関係数のカットオフ値との比較を含む。一部の実施形態では、回帰についての適合の良さの評価は、線形回帰を指し示す。例えば、ある特定の実施形態では、適合の良さは、非線形回帰についてより、線形回帰について大きく、適合の良さの評価は、線形回帰を指し示す。一部の実施形態では、評価は、線形回帰を指し示し、線形回帰を使用して、カウントを正規化する。一部の実施形態では、回帰についての適合の良さの評価は、非線形回帰を指し示す。例えば、ある特定の実施形態では、適合の良さは、線形回帰についてより、非線形回帰について大きく、適合の良さの評価は、非線形回帰を指し示す。一部の実施形態では、評価は、非線形回帰を指し示し、非線形回帰を使用して、カウントを正規

化する。

【0265】

一部の実施形態では、適合の良さの評価は、相関係数が、相関係数カットオフに等しいかまたはそれ超の場合に線形回帰を指し示す。一部の実施形態では、適合の良さの評価は、相関係数が相関係数カットオフ未満である場合に非線形回帰を指し示す。一部の実施形態では、相関係数カットオフは、所定のカットオフである。一部の実施形態では、相関係数カットオフは、約0.5もしくはそれ超、約0.55もしくはそれ超、約0.6もしくはそれ超、約0.65もしくはそれ超、約0.7もしくはそれ超、約0.75もしくはそれ超、約0.8もしくはそれ超、または約0.85もしくはそれ超である。

【0266】

例えば、ある特定の実施形態では、相関係数が、約0.6に等しいかまたはそれ超の場合に、線形回帰を含む正規化法を使用する。ある特定の実施形態では、相関係数が、0.6の相関係数カットオフに等しいかまたはそれ超の場合は、試料（例えば、参照ゲノムの部分1つ当たりのカウント、部分1つ当たりのカウント）のカウントを、線形回帰に従って正規化し、そうでない場合は、カウントを、非線形回帰に従って正規化する（例えば、係数が、0.6の相関係数カットオフ未満である場合）。一部の実施形態では、正規化処理は、(i) カウントおよび(ii) GC含有量、参照ゲノムの複数の部分のうちの各部分についての、線形回帰または非線形回帰を生成することを含む。ある特定の実施形態では、相関係数が、0.6の相関係数カットオフ未満である場合に、非線形回帰（例えば、LOWESS、LOESS）を含む正規化法を使用する。一部の実施形態では、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未満の相関係数カットオフ未満である場合に、非線形回帰（例えば、LOWESS）を含む正規化法を使用する。例えば、一部の実施形態では、相関係数が約0.6の相関係数カットオフ未満である場合に、非線形回帰（例えば、LOWESS、LOESS）を含む正規化法を使用する。

【0267】

一部の実施形態では、回帰の具体的な種類（例えば、線形または非線形回帰）を選択し、回帰を生成した後で、回帰をカウントから減算することにより、カウントを正規化する。一部の実施形態では、回帰をカウントから減算することにより、偏り（例えば、GCの偏り）の低減された、正規化されたカウントを提示する。一部の実施形態では、線形回帰をカウントから減算する。一部の実施形態では、非線形回帰（例えば、LOESS、GC-LOESS、LOWESS回帰）をカウントから減算する。任意の適切な方法を使用して、回帰直線をカウントから減算することができる。例えば、カウント x を、0.5のGC含有量を含む部分 i から導出し、回帰直線により、GC含有量を0.5とするときのカウント y を決定し、よって、 $x - y$ = 部分 i についての正規化されたカウントである。一部の実施形態では、回帰を減算する前に、かつ／または回帰を減算した後で、カウントを正規化する。一部の実施形態では、ハイブリッド正規化法により正規化されたカウントを使用して、ゲノム区分のレベル、Zコア、ゲノムまたはそのセグメントのレベルおよび／またはプロファイルを生成する。ある特定の実施形態では、ハイブリッド正規化法により正規化されたカウントを、本明細書で記載される方法により分析して、遺伝子の変動（例えば、胎児における）の存在または非存在を決定する。

【0268】

一部の実施形態では、ハイブリッド正規化法は、正規化の前または後における、1つまたは複数の部分をフィルタリングすることまたは重み付けすることを含む。本明細書で記載される部分（例えば、参照ゲノムの部分）のフィルタリング法を含む、適切な部分のフィルタリング法を使用することができる。一部の実施形態では、部分（例えば、参照ゲノムの部分）は、ハイブリッド正規化法を適用する前にフィルタリングする。一部の実施形態では、選択部分（例えば、カウントの可変性に従って選択された部分）へとマッピングした配列決定読取りのカウントだけを、ハイブリッド正規化により正規化する。一部の実施形態では、ハイブリッド正規化法を活用する前に、フィルタリングされた参照ゲノムの

10

20

30

40

50

部分（例えば、カウントの可変性に従ってフィルタリングされた部分）へとマッピングした配列決定読取りのカウントを除外する。一部の実施形態では、ハイブリッド正規化法は、適切な方法（例えば、本明細書で記載される方法）に従った、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。一部の実施形態では、ハイブリッド正規化法は、複数の試験試料について部分の各々へとマッピングしたカウントについての不確定値に従った、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。一部の実施形態では、ハイブリッド正規化法は、カウントの可変性に従った、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。一部の実施形態では、ハイブリッド正規化法は、GC含有量、反復エレメント、反復配列、イントロン、エクソンなど、またはこれらの組合せに従った、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。

10

【0269】

例えば、一部の実施形態では、複数の妊娠中の雌被験体から由来する複数の試料を分析し、部分（例えば、参照ゲノムの部分）のサブセットを、カウントの可変性に従って選択する。ある特定の実施形態では、線形回帰を使用して、(i) カウントおよび (ii) GC含有量についての相関係数を、妊娠中の雌被験体から得られた試料についての選択部分の各々について決定する。一部の実施形態では、所定の相関カットオフ値（例えば、約0.6の相関カットオフ値）を超える相関係数を決定し、適合の良さの評価により、線形回帰を指し示し、線形回帰をカウントから減算することによりカウントを正規化する。ある特定の実施形態では、所定の相関カットオフ値（例えば、約0.6の相関カットオフ値）未満の相関係数を決定し、適合の良さの評価により、非線形回帰を指し示し、LOESS回帰を生成し、LOESS回帰をカウントから減算することによりカウントを正規化する。

20

【0270】

プロファイル

一部の実施形態では、処理するステップは、データセットまたはその派生形の多様な側面（例えば、当技術分野で公知であり、かつ／または本明細書で記載される、1つまたは複数の数学的データ処理ステップおよび／または統計学的データ処理ステップの産物）からの、1つまたは複数のプロファイルの生成（例えば、プロファイルのプロット）を含む。

30

【0271】

本明細書で使用される「プロファイル」という用語は、大量のデータ中のパターンおよび／または相関の同定を容易としうるデータに対する数学的操作および／または統計学的操作の産物を指す。「プロファイル」は、データまたはデータセットに対する、1つまたは複数の判定基準に基づく、1つまたは複数の操作から結果として得られる値を含むことが多い。プロファイルは、複数のデータ点を含むことが多い。データセットの性格および／または複雑性に依拠して、任意の適切な数のデータ点を、プロファイルに含めることができる。ある特定の実施形態では、プロファイルには、2つまたはそれ超のデータ点、3つもしくはそれ超のデータ点、5つもしくはそれ超のデータ点、10もしくはそれ超のデータ点、24もしくはそれ超のデータ点、25もしくはそれ超のデータ点、50もしくはそれ超のデータ点、100もしくはそれ超のデータ点、500もしくはそれ超のデータ点、1000もしくはそれ超のデータ点、5000もしくはそれ超のデータ点、10,000もしくはそれ超のデータ点、または100,000もしくはそれ超のデータ点を含むことができる。

40

【0272】

一部の実施形態では、プロファイルは、データセットの全体を表し、ある特定の実施形態では、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表わす。すなわち、プロファイルは、ある場合には、いかなるデータも除外するようにフィルタリングされていないデータを表示するデータ点を含むかまたはこれらから生成されており、プロファイルは、ある場合には、望ましくないデータを除外するようにフィルタリングされたデータ

50

を表示するデータ点を含むかまたはこれらから生成されている。一部の実施形態では、プロファイル中のデータ点は、部分についてのデータ操作の結果を表示する。ある特定の実施形態では、プロファイル中のデータ点は、部分の群についてのデータ操作の結果を含む。一部の実施形態では、部分の群は、互いと隣接することが可能であり、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一部分に由来しうる。

【0273】

データセットから導出されたプロファイル中のデータ点は、任意の適切なデータの類別を表示しうる。プロファイルデータ点を生成するようにデータを群分けしうるカテゴリーの非限定的な例は、サイズに基づく部分、配列特徴（例えば、GC含有量、AT含有量、染色体上の場所（例えば、短腕部、長腕部、セントロメア、テロメア）など）に基づく部分、発現のレベル、染色体など、またはこれらの組合せを含む。一部の実施形態では、プロファイルは、別のプロファイルから得られるデータ点から生成することができる（例えば、再正規化データプロファイルを生成するように、異なる正規化する値に対して再正規化された正規化データプロファイル）。ある特定の実施形態では、別のプロファイルから得られるデータ点から生成されたプロファイルにより、データ点の数および／またはデータセットの複雑性を低減する。データ点の数および／またはデータセットの複雑性の低減により、データの解釈が容易となり、かつ／またはアウトカムの提示が容易となることが多い。

10

【0274】

プロファイル（例えば、ゲノムプロファイル、染色体プロファイル、染色体のセグメントのプロファイル）は、2つまたはそれ超の部分のための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コレクションであることが多い。プロファイルは、少なくとも1つのレベル（例えば、ゲノム区分のレベル）を含むことが多く、2つまたはそれ超のレベルを含むことが多い（例えば、プロファイルは、複数のレベルを有することが多い）。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ついてのレベルである。レベルについては、本明細書でより詳細に記載される。ある特定の実施形態では、プロファイルは、1つまたは複数の部分であって、重み付けするか、除外するか、フィルタリングするか、正規化するか、調整するか、平均するか、平均値として導出するか、加算するか、減算するか、処理するか、またはこれらの任意の組合せにより変換しうる部分を含む。プロファイルは、2つまたはそれ超のレベルを規定する部分へとマッピングした正規化されたカウントを含むことが多く、ここで、カウントは、適切な方法により、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、不確定値と関連することが多い。

20

30

【0275】

1つまたは複数のレベルを含むプロファイルは、場合によって、穴埋め（例えば、ホールの穴埋め）される。穴埋め（padding）（例えば、ホールの穴埋め）とは、母体の微小欠失または母体の重複（例えば、コピー数の変動）に起因するプロファイル中のレベルを同定および調整する処理を指す。一部の実施形態では、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を穴埋めする。一部の実施形態では、プロファイル中の微小重複または微小欠失により、プロファイル（例えば、染色体プロファイル）の全体的なレベルを人工的に上昇または低下させ、染色体の異数性（例えば、トリソミー）についての、偽陽性または偽陰性の決定をもたらすことができる。一部の実施形態では、微小重複および／または欠失に起因するプロファイル中のレベルを同定し、場合によって、穴埋めまたはホールの穴埋めと称する処理により調整する（例えば、穴埋めおよび／または除外する）。ある特定の実施形態では、プロファイルは、プロファイル中の第2のレベルと有意に異なる、1つまたは複数の第1のレベルを含み、1つまたは複数の第1のレベルの各々は、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動を含み、第1のレベルのうちの1または複数調整する。

40

【0276】

1つまたは複数の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

50

みうる。一部の実施形態では、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）。一部の実施形態では、第1のレベルは、第1の部分のセットを含み、第2のレベルは、第2の部分のセットを含み、第1の部分のセットは、第2の部分のセットのサブセットではない。ある特定の実施形態では、第1の部分のセットは、第2の部分のセットと異なり、これらから第1のレベルおよび第2のレベルが決定される。一部の実施形態では、プロファイルは、プロファイル中の第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のレベルを有しうる。一部の実施形態では、プロファイルは、プロファイル中の第2のレベルと有意に異なる、1つまたは複数の第1のレベルを含み、第1のレベルのうちの1または複数を調整する。一部の実施形態では、プロファイルは、プロファイル中の第2のレベルと有意に異なる、1つまたは複数の第1のレベルを含み、1つまたは複数の第1のレベルの各々は、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動を含み、第1のレベルのうちの1または複数を調整する。一部の実施形態では、プロファイル中の第1のレベルを、プロファイルから除外するかまたは調整する（例えば、穴埋めする）。プロファイルは、1つまたは複数の第2のレベルと有意に異なる、1つまたは複数の第1のレベルを含む複数のレベルを含むことが可能であり、プロファイル中のレベルの大半は、互いとほぼ等しい第2のレベルであることが多い。一部の実施形態では、プロファイル中のレベルのうちの50%超、60%超、70%超、80%超、90%超または95%超は、第2のレベルである。

【0277】

プロファイルは、場合によって、プロットとして示される。例えば、部分のカウント（例えば、正規化されたカウント）を表示する1つまたは複数のレベルは、プロットし、視覚化することができる。生成されうるプロファイルのプロットの非限定的な例は、未加工のカウント（例えば、未加工のカウントプロファイルまたは未加工のプロファイル）、正規化されたカウント、部分重み付け、zスコア、p値、適合させた倍数性と対比した面積比、適合させた胎児フラクションと測定した胎児フラクションとの間の比と対比した中央値レベル、主成分など、またはこれらの組合せを含む。一部の実施形態では、プロファイルのプロットにより、操作データの視覚化が可能となる。ある特定の実施形態では、プロファイルのプロットを活用して、アウトカム（例えば、適合させた倍数性と対比した面積比、適合させた胎児フラクションと測定した胎児フラクションとの間の比と対比した中央値レベル、主成分）を提示することができる。本明細書で使用される「未加工のカウントプロファイルのプロット」または「未加工のプロファイルのプロット」という用語は、領域中の全カウントに対して正規化された、領域中の各部分（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）中のカウントのプロットを指す。一部の実施形態では、プロファイルは、スタティックウィンドウ処理を使用して生成することができ、ある特定の実施形態では、プロファイルは、スライディングウィンドウ処理を使用して生成することができる。

【0278】

試験被験体について生成されたプロファイルは、場合によって、1つまたは複数の参照被験体について生成されたプロファイルと比較して、データセットの数学的操作および／もしくは統計学的操作の解釈を容易とし、かつ／またはアウトカムを提示する。一部の実施形態では、プロファイルは、1つまたは複数の出発仮定（例えば、母体の核酸寄与（例えば、母体のフラクション）、胎児の核酸寄与（例えば、胎児フラクション）、参照試料の倍数性など、またはこれらの組合せ）に基づき生成する。ある特定の実施形態では、試験プロファイルは、遺伝子の変動の非存在を表示する所定の値を中心とすることが多く、試験被験体が遺伝子の変動を保有したとする場合に、試験被験体において遺伝子の変動が位置するゲノム位置に対応するエリア中の所定の値からは逸脱することが多い。遺伝子の変動と関連する医学的状態の危険性があるか、またはこれを患っている試験被験体では、選択部分についての数値が、罹患していないゲノム位置についての所定の値から有意に変化することが期待される。出発仮定（例えば、一定の倍数性もしくは最適化された倍数性

、一定の胎児フラクションもしくは最適化された胎児フラクション、またはこれらの組合せ）に応じて、遺伝子の変動の存在または非存在を指し示す所定の閾もしくはカットオフ値またはの閾範囲は、遺伝子の変動の存在または非存在を決定するために有用なアウトカムをやはり提示しながらも、変化しうる。一部の実施形態では、プロファイルは、表現型を指し示し、かつ／またはこれを表示する。

【0279】

非限定的な例として述べると、正規化された試料および／または参照カウントプロファイルは、（a）遺伝子の変動を保有しないことが既知である参照基準のセットから選択された、染色体、部分、またはこれらのセグメントについての、参照中央値カウントを計算すること、（b）有益でない部分の、参照試料の未加工のカウントからの除外（例えば、フィルタリング）、（c）残りの参照ゲノムの全ての部分についての参照カウントを、参照試料、選択された染色体、または選択されたゲノム位置についての、残りカウントの総数（例えば、有益でない参照ゲノムの部分を除外した後の残りのカウントの合計）に対して正規化し、これにより、正規化された参照被験体プロファイルを生成すること、（d）対応する部分を試験被験体試料から除外すること、および（e）1つまたは複数の選択されたゲノム位置についての、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有する1つまたは複数の染色体についての、残りの参照中央値カウントの合計に対して正規化し、これにより、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り、未加工の配列の読取りデータから得ることができる。ある特定の実施形態では、（b）における、フィルタリングした部分により低減された全ゲノムに関する、さらなる正規化ステップを、（c）と（d）との間に含めることができる。

【0280】

データセットプロファイルは、カウントされたマッピングした配列の読取りデータに対する1つまたは複数の操作により生成することができる。一部の実施形態は、以下を含む：配列の読取りをマッピングし、各ゲノム部分へとマッピングされる配列タグの数を決定する（例えば、カウントする）。未加工のカウントプロファイルを、カウントされたマッピングした配列の読取りから生成する。ある特定の実施形態では、試験被験体に由来する未加工のカウントプロファイルを、遺伝子の変動を保有しないことが既知である、参照被験体のセットに由来する、染色体、部分、またはこれらのセグメントについての、参照中央値カウントプロファイルと比較することにより、アウトカムを提示する。

【0281】

一部の実施形態では、配列の読取りデータは、ノイズの多いデータまたは有益でない部分を除外するように、任意選択でフィルタリングする。フィルタリングの後、残りのカウントを足し合わせて、フィルタリングされたデータセットを生成することが典型的である。ある特定の実施形態では、フィルタリングされたカウントプロファイルを、フィルタリングされたデータセットから生成する。

【0282】

配列の読取りデータをカウントし、任意選択でフィルタリングした後で、データセットを正規化して、レベルまたはプロファイルを生成することができる。1つまたは複数の選択部分を、適切な正規化された参照値に対して正規化することにより、データセットを正規化することができる。一部の実施形態では、正規化された参照値は、部分が選択される1つまたは複数の染色体についての全カウントを表示する。ある特定の実施形態では、正規化された参照値は、遺伝子の変動を保有しないことが既知である、参照被験体のセットから調製された、参照データセットに由来する染色体の部分または染色体である、1つまたは複数の対応する部分を表示する。一部の実施形態では、正規化された参照値は、遺伝子の変動の存在または非存在について分析される試験被験体から調製された、試験被験体データセットに由来する、染色体の部分または染色体である、1つまたは複数の対応する部分を表示する。ある特定の実施形態では、正規化処理は、スタティックウィンドウ法を活用して実施し、一部の実施形態では、正規化処理は、ムービングウィンドウ法またはスライディングウィンドウ法を活用して実施する。ある特定の実施形態では、正規化された

カウントを含むプロファイルを生成して、アウトカムの分類および／または提示を容易とする。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き（例えば、このようなプロファイルのプロットを使用して）提示することができる。

【0283】

レベル

一部の実施形態では、値（例えば、数、定量的値）を、レベルに帰する。レベルは、適切な方法、演算、または数学的処理（例えば、処理されたレベル）により決定することができる。レベルは、部分のセットについてのカウント（例えば、正規化されたカウント）であるか、またはこれから導出されることが多い。一部の実施形態では、部分のレベルは、部分へとマッピングしたカウント（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、当技術分野で公知の適切な方法、演算、または数学的処理により処理、変換、または操作されたカウントから決定することが多い。一部の実施形態では、レベルは、処理されたカウントから導出し、処理されたカウントの非限定的な例は、重み付けされるか、除外されるか、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、平均値として導出される（例えば、平均レベル）か、加算されるか、減算されるか、変換されたカウント、またはこれらの組合せを含む。一部の実施形態では、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。レベルは、その非限定的な例が、部分に関する正規化、GC含有量による正規化、線形最小二乗回帰および非線形最小二乗回帰、GC LOESS、LOWESS、PERUN、ChAI、RM、GCRM、cQnなど、および／またはこれらの組合せを含む、適切な処理により正規化されたカウントについてのレベルでありうる。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うる。一部の実施形態では、レベルは、平均された、2つもしくはそれ超の部分のカウントまたは正規化されたカウントについてのレベルであり、レベルを、平均レベルと称する。一部の実施形態では、レベルは、平均カウントまたは正規化されたカウントの平均値を有する部分のセットについてのレベルであり、これを、平均レベルと称する。一部の実施形態では、レベルを、未加工のカウントおよび／またはフィルタリングされたカウントを含む部分について導出する。一部の実施形態では、レベルは、未加工のカウントであるカウントに基づく。一部の実施形態では、レベルは、不確定値（例えば、標準偏差、MAD）と関連する。一部の実施形態では、レベルを、Zスコアまたはp値により表示する。本明細書では、1つまたは複数の部分についてのレベルは、「ゲノム区分のレベル」と同義である。

【0284】

2つまたはそれ超のレベル（例えば、2つまたはそれ超のプロファイル中のレベル）についての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は、場合によって、レベルに従って、数学的に操作する（例えば、これに加算する、これに乗算する、これを平均する、これを正規化するなど、またはこれらの組合せ）ことができる。例えば、2つまたはそれ超のレベルについての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は、プロファイル中のレベルの1つ、一部、または全部に従って正規化することができる。一部の実施形態では、プロファイル中の全てのレベルについての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を、プロファイル中の1つのレベルに従って正規化する。一部の実施形態では、プロファイル中の第1のレベルについての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を、プロファイル中の第2のレベルについての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に従って正規化する。

【0285】

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のセットについてのレベル、カウントの平均値、中央値、もしくは平均を含む部分のセットについてのレベル、正規化されたカウントを含む部分のセットについてのレベルなど、またはこれらの任意の組合せである。一部の実施形態では、プロファイル中の第1のレベルおよび第2のレベルは、同じ染色体へとマッピングした部分のカウントから導出する。一部の実施形態では、プロファイル中の第1のレベルおよび第2のレベルは

、異なる染色体へとマッピングした部分のカウントから導出する。

【0286】

一部の実施形態では、レベルを、1つまたは複数の部分へとマッピングした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から決定する。一部の実施形態では、レベルを、2つまたはそれ超の部分へとマッピングした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から決定するが、ここで、各部分の正規化されたカウントは、ほぼ同じであることが多い。レベルについての部分のセット中のカウント（例えば、正規化されたカウント）には、変動があり得る。レベルについての部分のセット内には、セットの他の部分（例えば、ピークおよび／またはディップ）内とは、カウントが有意に異なる1つまたは複数の部分が存在し得る。任意の適切な数の部分と関連する、任意の適切な数の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は、レベルを規定しうる。

10

【0287】

一部の実施形態では、1つまたは複数のレベルは、ゲノムの部分の全部または一部の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から決定することができる。レベルは、染色体またはそのセグメントの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全部または一部から決定しうる人が多い。一部の実施形態では、2つまたはそれ超の部分（例えば、部分のセット）から導出された、2つまたはそれ超のカウントにより、レベルを決定する。一部の実施形態では、2つまたはそれ超のカウント（例えば、2つまたはそれ超の部分に由来するカウント）により、レベルを決定する。一部の実施形態では、2～約100、000の部分に由来するカウントにより、レベルを決定する。一部の実施形態では、2～約50、000、2～約40、000、2～約30、000、2～約20、000、2～約10、000、2～約5000、2～約2500、2～約1250、2～約1000、2～約500、2～約250、2～約100、または2～約60の部分に由来するカウントにより、レベルを決定する。一部の実施形態では、約10～約50の部分に由来するカウントにより、レベルを決定する。一部の実施形態では、約20～約40またはそれ超の部分に由来するカウントにより、レベルを決定する。一部の実施形態では、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またはそれ超の部分に由来するカウントを含む。一部の実施形態では、レベルは、部分のセット（例えば、参照ゲノムの部分のセット、染色体の部分のセット、または染色体のセグメントの部分のセット）に対応する。

20

30

【0288】

一部の実施形態では、レベルを、連続的な部分の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について決定する。一部の実施形態では、連続的な部分（例えば、部分のセット）は、ゲノムの隣接セグメントまたは染色体もしくは遺伝子の隣接セグメントを表示する。例えば、2つまたはそれ超の連続的な部分は、部分を末端から末端へとマージすることにより整列させる場合、各部分より長いDNA配列の配列アセンブリーを表示し得る。例えば、2つまたはそれ超の連続的な部分は、無傷ゲノム、染色体、遺伝子、イントロン、エクソン、またはそのセグメントを表示しうる。一部の実施形態では、レベルを、連続的な部分および／または非連続的な部分のコレクション（例えば、セット）から決定する。

40

【0289】

異なるレベル

一部の実施形態では、正規化されたカウントのプロファイルは、プロファイル中の別のレベル（例えば、第2のレベル）と有意に異なるレベル（例えば、第1のレベル）を含む。第1のレベルは、第2のレベルより高レベルの場合もあり、低レベルの場合もある。一部の実施形態では、第1のレベルは、コピー数の変動（例えば、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動）を含む1つまたは複数の読取りを含む部分のセットについてのレベルであり、第2のレベルは

50

、コピー数の変動を実質的に有さない読取りを含む部分のセットについてのレベルである。一部の実施形態では、「有意に異なる」とは、観察可能な差違を指す。一部の実施形態では、「有意に異なる」とは、「統計学的に異なる」または「統計学的な有意差」を指す。統計学的な有意差は、場合によって、観察された差違についての統計学的評価である。統計学的な有意差は、当技術分野で適切な方法により評価することができる。任意の適切な閾または範囲を使用して、2つのレベルが有意に異なることを決定することができる。ある特定の実施形態では、約0.01パーセント（例えば、レベル値のうちの1つまたは一方の0.01パーセント）またはそれ超異なる2つのレベル（例えば、平均レベル）は、有意に異なる。一部の実施形態では、約0.1パーセントまたはそれ超異なる2つのレベル（例えば、平均レベル）は、有意に異なる。ある特定の実施形態では、約0.5パーセントまたはそれ超異なる2つのレベル（例えば、平均レベル）は、有意に異なる。一部の実施形態では、約0.5、0.75、1、1.5、2、2.5、3、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5%、または約10%超異なる2つのレベル（例えば、平均レベル）は、有意に異なる。一部の実施形態では、2つのレベル（例えば、平均レベル）は、有意に異なり、いずれのレベルにも重複はなく、かつ／または一方もしくは両方のレベルについて計算された不確定値により規定される範囲に重複はない。ある特定の実施形態では、不確定値は、シグマとして表される標準偏差である。一部の実施形態では、2つのレベル（例えば、平均レベル）は、有意に異なり、不確定値の約1倍（例えば、1シグマ）またはそれ超異なる。一部の実施形態では、2つのレベル（例えば、平均レベル）は、有意に異なり、不確定値の約2倍（例えば、2シグマ）もしくはそれ超、不確定値の約3倍もしくはそれ超、約4倍もしくはそれ超、約5倍もしくはそれ超、約6倍もしくはそれ超、約7倍もしくはそれ超、約8倍もしくはそれ超、約9倍もしくはそれ超、または約10倍もしくはそれ超異なる。一部の実施形態では、2つのレベル（例えば、平均レベル）は、不確定値の約1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2.0、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、もしくは4.0倍、またはそれ超異なる場合に有意に異なる。一部の実施形態では、信頼性レベルは、2つのレベルの間の差違が増加するとともに増加する。ある特定の実施形態では、信頼性レベルは、2つのレベルの間の差違が減少するとともに、かつ／または不確定値が増加するとともに減少する。例えば、場合によって、信頼性レベルは、レベル間の差違と標準偏差（例えば、MAD）との比に応じて増加する。

【0290】

1つまたは複数の予測アルゴリズムを使用して、互いに対して非依存的に重み付けされる場合もあり、依存的に重み付けされる場合もある可変条件下で収集された検出データの有意性を決定するか、またはこれに意味を与えることができる。本明細書で使用される「変数」という用語は、値または値のセットを有するアルゴリズムの因子、量、または関数を指す。

【0291】

一部の実施形態では、第1の部分のセットは、第2の部分のセットと異なる（例えば、第2の部分のセットと重複しない）部分を含むことが多い。例えば、場合によって、正規化されたカウントの第1のレベルは、プロファイル中の正規化されたカウントの第2のレベルと有意に異なり、第1のレベルは、第1の部分のセットについてのレベルであり、第2のレベルは、第2の部分のセットについてのレベルであり、部分は、第1の部分のセットおよび、第2の部分のセットにおいて重複しない。ある特定の実施形態では、第1の部分のセットは、第2の部分のセットのサブセットではなく、これらから、それぞれ、第1のレベルおよび第2のレベルが決定される。一部の実施形態では、第1の部分のセットは、第2の部分のセットと異なり、かつ／または別個であり、これらから、それぞれ、第1のレベルおよび第2のレベルが決定される。

【0292】

一部の実施形態では、第1の部分のセットは、プロファイル中の第2の部分のセットのサブセットである。例えば、場合によって、プロファイル中の第2の部分のセットについての正規化されたカウントの第2のレベルは、プロファイル中の第1のレベルについての、第1の部分のセットの正規化されたカウントを含み、第1の部分のセットは、プロファイル中の第2の部分のセットのサブセットである。一部の実施形態では、平均レベル、平均値レベル、または中央値レベルは、第2のレベルから導出され、ここで、第2のレベルは、第1のレベルを含む。一部の実施形態では、第2のレベルは、全染色体を表示する第2の部分のセットを含み、第1のレベルは、第1の部分のセットを含み、ここで、第1のセットは、第2の部分のセットのサブセットであり、第1のレベルは、染色体内に存在する、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動を表示する。

10

【0293】

一部の実施形態では、第2のレベルの値は、第1のレベルより、染色体またはそのセグメントについてのカウントプロファイルの、平均値、平均、または中央値の値に近い。一部の実施形態では、第2のレベルは、染色体、染色体の部分またはそのセグメントの平均レベルである。一部の実施形態では、第1のレベルは、染色体またはそのセグメントを表示する主要レベル（例えば、第2のレベル）と有意に異なる。プロファイルには、第2のレベルと有意に異なる、複数の第1のレベルを含むことができ、各第1のレベルは独立に、第2のレベルより高レベルの場合もあり、低レベルの場合もある。一部の実施形態では、第1のレベルおよび第2のレベルは、同じ染色体から導出し、第1のレベルは、第2のレベルより高レベルであるか、低レベルであり、第2のレベルは、染色体の主要レベルである。一部の実施形態では、第1のレベルおよび第2のレベルは、同じ染色体から導出し、第1のレベルは、コピー数の変動（例えば、母体および／または胎児のコピー数の変動、欠失、挿入、重複）を指し示し、第2のレベルは、染色体またはそのセグメントについての部分の平均レベルまたは主要レベルである。

20

【0294】

ある特定の実施形態では、第2のレベルについての第2の部分のセット中の読取りは、遺伝子の変動（例えば、コピー数の変動、母体および／または胎児のコピー数の変動）を実質的に含まない。第2のレベルについての第2の部分のセットは、何らかの可変性（例えば、部分についてのレベルの可変性、カウントの可変性）を含むことが多い。一部の実施形態では、実質的にコピー数の変動がないことと関連するレベルについての部分のセット中の1つまたは複数の部分は、母体および／または胎児のゲノム内に存在するコピー数の変動を有する1つまたは複数の読取りを含む。例えば、場合によって、部分のセットは、小さな染色体のセグメント（例えば、10未満の部分）内に存在するコピー数の変動を含み、部分のセットは、実質的にコピー数の変動がないことと関連するレベルについての部分のセットである。したがって、実質的にコピー数の変動を含まない部分のセットはやはり、レベルの約10、9つ、8つ、7つ、6つ、5つ、4つ、3つ、2つ、または1つ未満の部分に存在するコピー数の変動を含みうる。

30

【0295】

一部の実施形態では、第1のレベルは、第1の部分のセットについてのレベルであり、第2のレベルは、第2の部分のセットについてのレベルであり、第1の部分のセットおよび第2の部分のセットは、連続している（例えば、染色体またはそのセグメントの核酸配列に関して隣接する）。一部の実施形態では、第1の部分のセットおよび第2の部分のセットは、連続していない。

40

【0296】

胎児核酸と母体核酸との混合物に由来する比較的短い配列の読取りを活用して、レベルおよび／またはプロファイルへと変換されうるカウントを提示することができる。カウント、レベル、およびプロファイルは、電子的形態で描示することもでき、実体的形態で描示することもでき、視覚化することができる。部分へとマッピングした（例えば、レベルおよび／またはプロファイルとして表された）カウントは、胎児および／または妊娠中の

50

雌において存在する胎児および／または母体のゲノム、染色体、または染色体の部分もしくはセグメントについての視覚的表示をもたらさう。

【0297】

参照レベルおよび正規化された参照値

一部の実施形態では、プロファイルは、参照レベル（例えば、参照として使用されるレベル）を含む。正規化されたカウントのプロファイルにより、期待レベルおよび期待範囲（期待レベルおよび期待範囲についての下記の議論を参照されたい）が決定される参照レベルを提示することが多い。参照レベルは、母親および胎児の両方に由来するマッピングした読取りを含む部分の正規化されたカウントについての参照レベルであることが多い。参照レベルは、胎児および母親（例えば、妊娠中の雌）に由来するマッピングした読取りの正規化されたカウントの合計であることが多い。一部の実施形態では、参照レベルは、正倍数性の母親および／または正倍数性の胎児に由来するマッピングした読取りを含む部分についての参照レベルである。一部の実施形態では、参照レベルは、胎児および／または母体の遺伝子の変動（例えば、異数性（例えば、トリソミー）、コピー数の変動、微小重複、微小欠失、挿入）を有するマッピングした読取りを含む部分についての参照レベルである。一部の実施形態では、参照レベルは、母体および／または胎児の遺伝子の変動（例えば、異数性（例えば、トリソミー）、コピー数の変動、微小重複、微小欠失、挿入）を実質的に含まない部分についての参照レベルである。一部の実施形態では、第2のレベルは、参照レベルとして使用されるレベルである。ある特定の実施形態では、プロファイルは、正規化されたカウントの第1のレベルおよび正規化されたカウントの第2のレベルを含み、第1のレベルは、第2のレベルと有意に異なり、第2のレベルは、参照レベルである。ある特定の実施形態では、プロファイルは、第1の部分のセットについての正規化されたカウントの第1のレベル、第2の部分のセットについての正規化されたカウントの第2のレベルを含み、第1の部分のセットは、母体および／または胎児のコピー数の変動を有するマッピングした読取りを含み、第2の部分のセットは、母体および／または胎児のコピー数の変動を実質的に有さないマッピングした読取りを含み、第2のレベルは、参照レベルである。

【0298】

一部の実施形態では、プロファイルについての1つまたは複数のレベルについての部分へとマッピングしたカウントを、参照レベルのカウントに従って正規化する。一部の実施形態では、参照レベルのカウントに従った、レベルのカウントを正規化することは、レベルのカウントを、参照レベルのカウントまたはその倍数もしくは分数で除算することを含む。参照レベルのカウントに従って正規化されたカウントは、別の処理（例えば、PERUN、ChAI）に従って正規化されていることが多く、参照レベルのカウントもまた正規化されている（例えば、PERUN、ChAIにより）ことが多い。一部の実施形態では、レベルのカウントを、参照レベルのカウントに従って正規化し、参照レベルのカウントは、正規化する前に、または正規化した後で、適切な値へとスケール可能である。参照レベルのカウントのスケール処理は、任意の適切な定数（すなわち、数）を含むことが可能であり、任意の適切な数学的操作を、参照レベルのカウントへと適用することができる。

【0299】

正規化された参照値（NRV: normalized reference value）は、参照レベルの正規化されたカウントに従って決定することが多い。NRVの決定は、参照レベルのカウントへと適用された任意の適切な正規化処理（例えば、数学的操作）を含むことが可能であり、ここでは、同じプロファイル内の他のレベルのカウントを正規化するのに、同じ正規化処理を使用する。NRVの決定は、参照レベルを、参照レベル自体で除算することを含むことが多い。NRVの決定は、参照レベルを、参照レベル自体の倍数で除算することを含むことが多い。NRVの決定は、参照レベルを、参照レベルと定数（例えば、任意の数）との和または差で除算することを含むことが多い。

【0300】

NRVは、場合によって、ヌル値と称する。NRVは、任意の適切な値でありうる。一部の実施形態では、NRVは、ゼロ以外の任意の値である。一部の実施形態では、NRVは、整数 (whole number) である。一部の実施形態では、NRVは、正の整数 (integer) である。一部の実施形態では、NRVは、1、10、100、または1000である。NRVは、1に等しいことが多い。一部の実施形態では、NRVは、ゼロに等しい。参照レベルのカウントを、任意の適切なNRVに対して正規化することができる。一部の実施形態では、参照レベルのカウントを、ゼロであるNRVに対して正規化する。参照レベルのカウントは、1であるNRVに対して正規化することが多い。

【0301】

期待レベル

10

期待レベルは、場合によって、あらかじめ規定されたレベル（例えば、理論レベル、予測レベル）である。本明細書では、場合によって、「期待レベル」を、「所定のレベル値」と称する。一部の実施形態では、期待レベルは、コピー数の変動を含む部分のセットについての正規化されたカウントのレベルについての予測値である。ある特定の実施形態では、期待レベルを、実質的にコピー数の変動を含まない部分のセットについて決定する。期待レベルは、染色体の倍数性（例えば、0、1つ、2つ（すなわち、二倍体）、3つ、または4つの染色体）または微小倍数性（ホモ接合性またはヘテロ接合性の欠失、重複、挿入、またはこれらの非存在）について決定することができる。期待レベルは、母体の微小倍数性（例えば、母体および／または胎児のコピー数の変動）について決定することが多い。

20

【0302】

遺伝子の変動またはコピー数の変動についての期待レベルは、任意の適切な様式で決定することができる。期待レベルは、レベルの適切な数学的操作（例えば、レベルについての部分のセットへとマッピングしたカウント）により決定することが多い。一部の実施形態では、期待レベルを、場合によって、期待レベル定数と称する定数を活用することにより決定する。コピー数の変動についての期待レベルは、場合によって、参照レベル、参照レベルの正規化されたカウント、もしくはNRVに、期待レベル定数を乗算すること、参照レベル、参照レベルの正規化されたカウント、もしくはNRVに期待レベル定数を加算すること、参照レベル、参照レベルの正規化されたカウント、もしくはNRVから期待レベル定数を減算すること、参照レベル、参照レベルの正規化されたカウント、もしくはNRVを期待レベル定数で除算すること、またはこれらの組合せにより計算する。同じ被験体、試料、または試験群について決定された期待レベル（例えば、母体および／または胎児のコピー数の変動の期待レベル）は、同じ参照レベルまたはNRVに従って決定することが多い。

30

【0303】

期待レベルは、参照レベル、参照レベルの正規化されたカウント、またはNRVに、期待レベル定数を乗算することにより決定することが多く、ここで、参照レベル、参照レベルの正規化されたカウント、またはNRVは、ゼロに等しくない。一部の実施形態では、期待レベル定数を、参照レベル、参照レベルの正規化されたカウント、またはゼロに等しいNRVへと加算することにより、期待レベルを決定する。一部の実施形態では、期待レベル、参照レベルの正規化されたカウント、NRVおよび期待レベル定数は、スケール可能である。スケール処理は、任意の適切な定数（すなわち、数）および任意の適切な数学的操作を含むことが可能であり、同じスケール処理を、検討される全ての値へと適用する。

40

【0304】

期待レベル定数

期待レベル定数は、適切な方法により決定することができる。一部の実施形態では、期待レベル定数を任意に決定する。期待レベル定数は、経験的に決定することが多い。一部の実施形態では、期待レベル定数を、数学的操作に従って決定する。一部の実施形態では、期待レベル定数を、参照（例えば、参照ゲノム、参照試料、参照試験データ）に従って

50

決定する。一部の実施形態では、期待レベル定数は、遺伝子の変動またはコピー数の変動（例えば、重複、挿入、または欠失）の存在または非存在を表示するレベルについての、所定の期待レベル定数である。一部の実施形態では、期待レベル定数は、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動の存在または非存在を表示するレベルについての、所定の期待レベル定数である。コピー数の変動についての期待レベル定数は、任意の適切な定数または定数のセットでありうる。

【0305】

一部の実施形態では、ホモ接合性の重複（例えば、ホモ接合性の重複）についての期待レベル定数は、約1.6～約2.4、約1.7～約2.3、約1.8～約2.2、または約1.9～約2.1でありうる。一部の実施形態では、ホモ接合性の重複についての期待レベル定数は、約1.6、1.7、1.8、1.9、2.0、2.1、2.2、2.3、または約2.4である。ホモ接合性の重複についての期待レベル定数は、約1.90、1.92、1.94、1.96、1.98、2.0、2.02、2.04、2.06、2.08、または約2.10であることが多い。ホモ接合性の重複についての期待レベル定数は、約2であることが多い。

10

【0306】

一部の実施形態では、ヘテロ接合性の重複（例えば、ホモ接合性の重複）についての期待レベル定数は、約1.2～約1.8、約1.3～約1.7、または約1.4～約1.6である。一部の実施形態では、ヘテロ接合性の重複についての期待レベル定数は、約1.2、1.3、1.4、1.5、1.6、1.7、または約1.8である。ヘテロ接合性の重複についての期待レベル定数は、約1.40、1.42、1.44、1.46、1.48、1.5、1.52、1.54、1.56、1.58、または約1.60であることが多い。一部の実施形態では、ヘテロ接合性の重複についての期待レベル定数は、約1.5である。

20

【0307】

一部の実施形態では、コピー数の変動の非存在（例えば、母体および／または胎児のコピー数の変動の非存在）についての期待レベル定数は、約1.3～約0.7、約1.2～約0.8、または約1.1～約0.9である。一部の実施形態では、コピー数の変動の非存在についての期待レベル定数は、約1.3、1.2、1.1、1.0、0.9、0.8、または約0.7である。コピー数の変動の非存在についての期待レベル定数は、約1.09、1.08、1.06、1.04、1.02、1.0、0.98、0.96、0.94、または約0.92であることが多い。一部の実施形態では、コピー数の変動の非存在についての期待レベル定数は、約1である。

30

【0308】

一部の実施形態では、ヘテロ接合性の欠失（例えば、母体の、胎児の、または母体および胎児のヘテロ接合性の欠失）についての期待レベル定数は、約0.2～約0.8、約0.3～約0.7、または約0.4～約0.6である。一部の実施形態では、ヘテロ接合性の欠失についての期待レベル定数は、約0.2、0.3、0.4、0.5、0.6、0.7、または約0.8である。ヘテロ接合性の欠失についての期待レベル定数は、約0.40、0.42、0.44、0.46、0.48、0.5、0.52、0.54、0.56、0.58、または約0.60であることが多い。一部の実施形態では、ヘテロ接合性の欠失についての期待レベル定数は、約0.5である。

40

【0309】

一部の実施形態では、ホモ接合性の欠失（例えば、ホモ接合性の欠失）についての期待レベル定数は、約-0.4～約0.4、約-0.3～約0.3、約-0.2～約0.2、または約-0.1～約0.1でありうる。一部の実施形態では、ホモ接合性の欠失についての期待レベル定数は、約-0.4、-0.3、-0.2、-0.1、0.0、0.1、0.2、0.3、または約0.4である。ホモ接合性の欠失についての期待レベル定数は、約-0.1、-0.08、-0.06、-0.04、-0.02、0.0、0.02、

50

0.04、0.06、0.08、または約0.10であることが多い。ホモ接合性の欠失についての期待レベル定数は、約0であることが多い。

【0310】

期待レベルの範囲

一部の実施形態では、遺伝子の変動またはコピー数の変動の存在または非存在（例えば、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動）を、期待レベルの範囲内または範囲外にあるレベルにより決定する。期待レベルの範囲は、期待レベルに従って決定することが多い。一部の実施形態では、期待レベルの範囲を、遺伝子の変動を実質的に含まないかまたはコピー数の変動を実質的に含まないレベルについて決定する。適切な方法を使用して、期待レベルの範囲を決定することができる。

10

【0311】

一部の実施形態では、期待レベルの範囲を、レベルについて計算された適切な不確定値に従って規定する。不確定値の非限定的な例は、標準偏差、標準誤差、計算された分散、p値、および平均絶対偏差（MAD: mean absolute deviation）である。一部の実施形態では、遺伝子の変動またはコピー数の変動についての期待レベルの範囲は、一部分、レベル（例えば、第1のレベル、第2のレベル、第1のレベルおよび第2のレベル）についての不確定値を計算することにより決定する。一部の実施形態では、期待レベルの範囲を、プロファイル（例えば、染色体またはそのセグメントについての正規化されたカウントのプロファイル）について計算された不確定値に従って規定する。一部の実施形態では、不確定値を、遺伝子の変動を実質的に含まないかまたはコピー数の変動を実質的に含まないレベルについて計算する。一部の実施形態では、不確定値を、第1のレベル、第2のレベルまたは第1のレベルおよび第2のレベルについて計算する。一部の実施形態では、不確定値を、第1のレベル、第2のレベル、または第1のレベルを含む第2のレベルについて決定する。

20

【0312】

期待レベルの範囲は、場合によって、一部分、不確定値に定数（例えば、所定の定数） n を乗算すること、不確定値に定数（例えば、所定の定数） n を加算すること、不確定値から定数（例えば、所定の定数） n を減算すること、または不確定値を定数（例えば、所定の定数） n で除算することにより計算する。適切な数学的手順または手順の組合せを使用することができる。定数 n （例えば、所定の定数 n ）は、場合によって、信頼区間と称する。選択された信頼区間は、選択された定数 n に従って決定する。定数 n （例えば、所定の定数 n 、信頼区間）は、適切な様式で決定することができる。定数 n は、数またはゼロ超の数の分数でありうる。定数 n は、整数でありうる。定数 n は、10未満の数であることが多い。一部の実施形態では、定数 n は、約10未満、約9未満、約8未満、約7未満、約6未満、約5未満、約4未満、約3未満、または約2未満の数である。一部の実施形態では、定数 n は、約10、9.5、9、8.5、8、7.5、7、6.5、6、5.5、5、4.5、4、3.5、3、2.5、2、または1である。定数 n は、遺伝的素質が既知である被験体（妊娠中の雌および／または胎児）に由来するデータから経験的に決定することができる。

30

40

【0313】

不確定値および定数 n により、範囲（例えば、不確定カットオフ）を規定することが多い。例えば、場合によって、不確定値は、標準偏差（例えば、 ± 5 ）であり、これに、定数 n （例えば、信頼区間）を乗じ、これにより、範囲または不確定カットオフ（uncertainty cutoff）（例えば、 $5n \sim -5n$ ）を規定する。

【0314】

一部の実施形態では、遺伝子の変動（例えば、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動）についての期待レベルの範囲は、期待レベルに、不確定値を乗算した定数 n （例えば、 $n \times \text{シグマ}$ （例えば、6シグマ））を加えた和である。一部の実施形態では、 k と命名される遺伝子の変動また

50

はコピー数の変動についての期待レベルの範囲は、式：

式 R：（期待レベルの範囲） $_k$ = （期待レベル） $_k$ + $n \sigma$

【0315】

〔式中、 σ は、不確定値であり、 n は、定数（例えば、所定の定数）であり、期待レベルの範囲および期待レベルは、遺伝子の変動 k （例えば、 k ＝ヘテロ接合性の欠失、例えば、 k ＝遺伝子の変動の非存在）についての期待レベルの範囲および期待レベルである。例えば、1に等しい期待レベル（例えば、コピー数の変動の非存在）、 ± 0.05 に等しい不確定値（すなわち、 σ ）、および $n=3$ について、期待レベルの範囲を、 $1.15 \sim 0.85$ と規定する〕により規定することができる。一部の実施形態では、ヘテロ接合性の重複についての期待レベルを 1.5 とし、 $n=3$ とし、不確定値 σ を ± 0.05 とする場合、ヘテロ接合性の重複についての期待レベルの範囲を、 $1.65 \sim 1.35$ と決定する。一部の実施形態では、ヘテロ接合性の重複についての期待レベルを 0.5 とし、 $n=3$ とし、不確定値 σ を ± 0.05 とする場合、ヘテロ接合性の欠失についての期待レベルの範囲を、 $0.65 \sim 0.35$ と決定する。一部の実施形態では、ヘテロ接合性の重複についての期待レベルを 2.0 とし、 $n=3$ とし、不確定値 σ を ± 0.05 とする場合、ホモ接合性の重複についての期待レベルの範囲を、 $2.15 \sim 1.85$ と決定する。一部の実施形態では、ヘテロ接合性の重複についての期待レベルを 0.0 とし、 $n=3$ とし、不確定値 σ を ± 0.05 とする場合、ホモ接合性の欠失についての期待レベルの範囲を、 $0.15 \sim -0.15$ と決定する。

【0316】

一部の実施形態では、ホモ接合性のコピー数の変動についての期待レベルの範囲（例えば、母体の、胎児のまたは母体および胎児のホモ接合性のコピー数の変動）は一部分、対応するヘテロ接合性のコピー数の変動についての期待レベルの範囲に従って決定する。例えば、場合によって、ホモ接合性の重複についての期待レベルの範囲は、ヘテロ接合性の重複についての期待レベルの範囲の上限を超える全ての値を含む。一部の実施形態では、ホモ接合性の重複についての期待レベルの範囲は、ヘテロ接合性の重複についての期待レベルの範囲の上限を超えるかまたはこれに等しい全ての値を含む。一部の実施形態では、ホモ接合性の重複についての期待レベルの範囲は、ヘテロ接合性の重複についての期待レベルの範囲の上限を超え、かつ、式 R〔式中、 σ は、不確定値であり、正の値であり、 n は、定数であり、 k は、ホモ接合性の重複である〕により規定される上限未満である全ての値を含む。一部の実施形態では、ホモ接合性の重複についての期待レベルの範囲は、ヘテロ接合性の重複についての期待レベルの範囲の上限を超えるかまたはこれに等しく、かつ、式 R〔式中、 σ は、不確定値であり、 σ は、正の値であり、 n は、定数であり、 k は、ホモ接合性の重複である〕により規定される上限未満であるかまたはこれに等しい全ての値を含む。

【0317】

一部の実施形態では、ホモ接合性の欠失についての期待レベルの範囲は、ヘテロ接合性の欠失についての期待レベルの範囲の下限未満の全ての値を含む。一部の実施形態では、ホモ接合性の欠失についての期待レベルの範囲は、ヘテロ接合性の欠失についての期待レベルの範囲の下限未満であるかまたはこれに等しい全ての値を含む。一部の実施形態では、ホモ接合性の欠失についての期待レベルの範囲は、ヘテロ接合性の欠失についての期待レベルの範囲の下限未満であり、かつ、式 R〔式中、 σ は、不確定値であり、 σ は、負の値であり、 n は、定数であり、 k は、ホモ接合性の欠失である〕により規定される下限を超える全ての値を含む。一部の実施形態では、ホモ接合性の欠失についての期待レベルの範囲は、ヘテロ接合性の欠失についての期待レベルの範囲の下限未満であるかまたはこれに等しく、かつ、式 R〔式中、 σ は、不確定値であり、 σ は、負の値であり、 n は、定数であり、 k は、ホモ接合性の欠失である〕により規定される下限を超えるかまたはこれに等しい全ての値を含む。

【0318】

不確定値を活用して、閾値を決定することができる。一部の実施形態では、範囲（例え

ば、閾範囲)は、未加工のカウント、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から決定された不確定値を計算することにより得られる。一部の実施形態では、範囲は、範囲を生成するレベルについての不確定値(例えば、レベルの正規化されたカウント)に、カットオフ閾(例えば、3標準偏差では、3を乗算する)として選択された、不確定値の倍数(例えば、ある数だけの標準偏差)を表わす所定の定数(例えば、1、2、3、4、5、6など)を乗算することにより決定することができる。一部の実施形態では、範囲は、値(例えば、所定の値、不確定値、所定の定数を乗じられた不確定値)を、範囲を生成するレベルに加算することおよび／または範囲を生成するレベルから減算することにより決定することができる。例えば、レベルを1に等しいとし、標準偏差を±0.2とし、ここで、所定の定数を3とすると、範囲は、(1+3(0.2))～(1+3(−0.2))、または1.6～0.4と計算することができる。範囲は、場合によって、コピー数の変動についての期待範囲または期待レベルの範囲を規定しうる。ある特定の実施形態では、値の範囲外であれ、値の範囲内であれ、閾値を超える部分の一部または全部を、正規化処理の一部として、正規化処理の前に、または正規化処理の後で除外する。一部の実施形態では、範囲外であれ、範囲内であれ、計算された閾値を超える部分の一部または全部を、正規化処理もしくは分類処理の一部として、正規化処理もしくは分類処理の前に、重み付けするかまたは調整する。重み付けの例については、本明細書で記載される。本明細書で使用される「冗長データ」および「冗長なマッピングした読取り」という用語は、試料に由来する配列の読取りであって、既にゲノム位置(例えば、塩基の場所)へと割り当てられ、かつ／または部分についてカウントされたものとして同定される配列の読取りを指す。

【0319】

一部の実施形態では、不確定値を、下記の式：

【化4】

$$Z = \frac{L_A - L_O}{\sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_O^2}{N_O}}}$$

【0320】

[式中、Zは、2つのレベルの間の標準化した偏差を表示し、Lは、平均値(または中央値)レベルであり、シグマは、標準偏差(またはMAD)である。添え字Oは、プロファイルのセグメント(例えば、第2のレベル、染色体、NRV、「正倍数性レベル」、コピー数の変動が存在しないレベル)について描示し、Aは、別のプロファイルのセグメント(例えば、第1のレベル、コピー数の変動を表示するレベル、異数性(例えば、トリソミー)を表示するレベル)について描示する。変数N_Oは、添え字Oにより描示されるプロファイルのセグメント中の部分の総数を表示する。N_Aは、添え字Aにより描示されるプロファイルのセグメント中の部分の総数を表示する]に従って決定する。

【0321】

コピー数の変動の類別

別のレベル(例えば、第2のレベル)と有意に異なるレベル(例えば、第1のレベル)は、期待レベルの範囲に従って、コピー数の変動(例えば、母体および／または胎児のコピー数の変動、胎児のコピー数の変動、欠失、重複、挿入)として類別しうることが多い。一部の実施形態では、第1のレベルが、第2のレベルと有意に異なり、第1のレベルが、コピー数の変動についての期待レベルの範囲内にある場合に、コピー数の変動の存在を類別する。例えば、コピー数の変動(例えば、母体および／または胎児のコピー数の変動、胎児のコピー数の変動)は、第1のレベルが、第2のレベルと有意に異なり、第1のレベルが、コピー数の変動についての期待レベルの範囲内にある場合に類別することができる。一部の実施形態では、ヘテロ接合性の重複(例えば、母体もしくは胎児の、または母

体および胎児の、ヘテロ接合性の重複)またはヘテロ接合性の欠失(例えば、母体または胎児の、または母体および胎児の、ヘテロ接合性の欠失)は、第1のレベルが、第2のレベルと有意に異なり、第1のレベルが、ヘテロ接合性の重複またはヘテロ接合性の欠失のそれぞれについての期待レベルの範囲内にある場合に類別される。一部の実施形態では、ホモ接合性の重複またはホモ接合性の欠失は、第1のレベルが、第2のレベルと有意に異なり、第1のレベルが、ホモ接合性の重複またはホモ接合性の欠失のそれぞれについての期待レベルの範囲内にある場合に類別される。

【0322】

レベルの調整

一部の実施形態では、1つまたは複数のレベルを調整する。レベルを調整するための処理は、穴埋めと称することが多い。一部の実施形態では、プロファイル中の複数のレベル(例えば、ゲノムのプロファイル、染色体のプロファイル、染色体の部分またはセグメントのプロファイル)を調整する。一部の実施形態では、プロファイル中の約1つ~約10,000またはそれ超のレベルを調整する。一部の実施形態では、プロファイル中の約1つ~約1000、1つ~約900、1つ~約800、1つ~約700、1つ~約600、1つ~約500、1つ~約400、1つ~約300、1つ~約200、1つ~約100、1つ~約50、1つ~約25、1つ~約20、1つ~約15、1つ~約10、または1つ~約5つのレベルを調整する。一部の実施形態では、1つのレベルを調整する。一部の実施形態では、第2のレベルと有意に異なるレベル(例えば、正規化されたカウントプロファイルの第1のレベル)を調整する。一部の実施形態では、コピー数の変動として類別されたレベルを調整する。一部の実施形態では、第2のレベルと有意に異なるレベル(例えば、正規化されたカウントプロファイルの第1のレベル)を、コピー数の変動(例えば、コピー数の変動、例えば、母体のコピー数の変動)として類別し、調整する。一部の実施形態では、レベル(例えば、第1のレベル)は、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動についての期待レベルの範囲内にあり、そのレベルを調整する。一部の実施形態では、1つまたは複数のレベル(例えば、プロファイル中のレベル)を調整しない。一部の実施形態では、レベル(例えば、第1のレベル)は、コピー数の変動についての期待レベルの範囲外にあり、そのレベルを調整しない。コピー数の変動の非存在についての期待レベルの範囲中のレベルは、調整しないことが多い。任意の適切な数の調整を、プロファイル中の1つまたは複数のレベルに対して施すことができる。一部の実施形態では、1つまたは複数のレベルを調整する。一部の実施形態では、2またはそれ超、3またはそれ超、5またはそれ超、6またはそれ超、7またはそれ超、8またはそれ超、9またはそれ超、場合によって、10またはそれ超のレベルを調整する。

【0323】

一部の実施形態では、第1のレベルの値を、第2のレベルの値に従って調整する。一部の実施形態では、コピー数の変動を表示するものとして同定される第1のレベルを、第2のレベルの値に対して調整し、ここで、第2のレベルは、コピー数の変動が存在しないことと関連することが多い。ある特定の実施形態では、コピー数の変動を表示するものとして同定される第1のレベルの値を、第1のレベルの値が、第2のレベルの値とほぼ等しくなるように調整する。

【0324】

調整は、適切な数学的演算を含みうる。一部の実施形態では、調整は、1つまたは複数の数学的演算を含む。一部の実施形態では、レベルは、それを正規化すること、それをフィルタリングすること、それを平均すること、それに乗算すること、それを除算すること、それに加算すること、もしくはそれから減算すること、またはこれらの組合せにより調整される。一部の実施形態では、所定の値または定数によりレベルを調整する。一部の実施形態では、レベルの値を、別のレベルの値へと改変することによりレベルを調整する。例えば、第1のレベルは、その値を第2のレベルの値へと改変することにより調整することができる。このような場合の値は、処理値(例えば、平均値、正規化した値など)であ

りうる。

【0325】

一部の実施形態では、レベルを、コピー数の変動（例えば、母体のコピー数の変動）として類別し、本明細書では所定の調整値（PAV: predeterminded adjustment value）と称する、所定の値に従って調整する。PAVは、特異的コピー数の変動について決定することが多い。特異的コピー数の変動（例えば、ホモ接合性の重複、ホモ接合性の欠失、ヘテロ接合性の重複、ヘテロ接合性の欠失）について決定されたPAVは、特異的コピー数の変動（例えば、ホモ接合性の重複、ホモ接合性の欠失、ヘテロ接合性の重複、ヘテロ接合性の欠失）として類別されたレベルを調整するのに使用することが多い。ある特定の実施形態では、レベルを、コピー数の変動として類別し、次いで、類別されたコピー数の変動の種類に特異的なPAVに従って調整する。一部の実施形態では、レベル（例えば、第1のレベル）を、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動として類別し、PAVを、レベルへと加算すること、またはレベルから減算することにより調整する。レベル（例えば、第1のレベル）を、母体のコピー数の変動として類別し、PAVを、レベルへと加算することにより調整することが多い。例えば、重複（例えば、母体の、胎児の、または母体および胎児のホモ接合性の重複）として類別されたレベルは、特異的重複（例えば、ホモ接合性の重複）について決定されたPAVを加算することにより調整し、これにより、調整されたレベルを提示することができる。コピー数の重複について決定されるPAVは、負の値であることが多い。一部の実施形態では、重複について決定されたPAVを活用することにより、重複を表示するレベルに調整を施す結果として、レベルの値が低減される。一部の実施形態では、第2のレベルと有意に異なるレベル（例えば、第1のレベル）を、コピー数の欠失（例えば、ホモ接合性の欠失、ヘテロ接合性の欠失、ホモ接合性の重複、ホモ接合性の重複）として類別し、コピー数の欠失について決定されたPAVを加算することにより、第1のレベルを調整する。コピー数の欠失について決定されたPAVは、正の値であることが多い。一部の実施形態では、欠失について決定されたPAVを活用することにより、欠失を表示するレベルに調整を施す結果として、レベルの値が増加する。

【0326】

PAVは、任意の適切な値でありうる。PAVは、コピー数の変動（例えば、類別されたコピー数の変動）に従って決定され、これに特異的であることが多い。ある特定の実施形態では、PAVを、コピー数の変動（例えば、類別されたコピー数の変動）についての期待レベルおよび／またはPAV係数に従って決定する。PAVは、場合によって、期待レベルにPAV係数を乗算することにより決定する。例えば、コピー数の変動についてのPAVは、コピー数の変動（例えば、ヘテロ接合性の欠失）について決定された期待レベルに、同じコピー数の変動（例えば、ヘテロ接合性の欠失）について決定されたPAV係数を乗算することにより決定することができる。例えば、PAVは、コピー数の変動 k （例えば、 k ＝ヘテロ接合性の欠失）についての下記の式：

$$PAV_k = (\text{期待レベル})_k \times (PAV\text{係数})_k$$

により決定することができる。

【0327】

PAV係数は、任意の適切な値でありうる。一部の実施形態では、ホモ接合性の重複についてのPAV係数は、約 -0.6 ～約 -0.4 の間である。一部の実施形態では、ホモ接合性の重複についてのPAV係数は、約 -0.60 、 -0.59 、 -0.58 、 -0.57 、 -0.56 、 -0.55 、 -0.54 、 -0.53 、 -0.52 、 -0.51 、 -0.50 、 -0.49 、 -0.48 、 -0.47 、 -0.46 、 -0.45 、 -0.44 、 -0.43 、 -0.42 、 -0.41 、および -0.40 である。ホモ接合性の重複についてのPAV係数は、約 -0.5 であることが多い。

【0328】

例えば、NRVを約1とし、ホモ接合性の重複の期待レベルを約2に等しいとすると、

ホモ接合性の重複についてのPAVは、上記の式に従って、約-1と決定される。この場合、例えば、約-1を、第1のレベルの値へと加算することにより、ホモ接合性の重複として類別された第1のレベルを調整する。

【0329】

一部の実施形態では、ヘテロ接合性の重複についてのPAV係数は、約-0.4～約-0.2の間である。一部の実施形態では、ヘテロ接合性の重複についてのPAV係数は、約-0.40、-0.39、-0.38、-0.37、-0.36、-0.35、-0.34、-0.33、-0.32、-0.31、-0.30、-0.29、-0.28、-0.27、-0.26、-0.25、-0.24、-0.23、-0.22、-0.21、および-0.20である。ヘテロ接合性の重複についてのPAV係数は、約-0.33

10

【0330】

例えば、NRVを約1とし、ヘテロ接合性の重複の期待レベルを約1.5に等しいとすると、ホモ接合性の重複についてのPAVは、上記の式に従って、約-0.495と決定される。この場合、例えば、約-0.495を、第1のレベルの値へと加算することにより、ヘテロ接合性の重複として類別された第1のレベルを調整する。

【0331】

一部の実施形態では、ヘテロ接合性の欠失についてのPAV係数は、約0.4～約0.2の間である。一部の実施形態では、ヘテロ接合性の欠失についてのPAV係数は、約0.40、0.39、0.38、0.37、0.36、0.35、0.34、0.33、0.32、0.31、0.30、0.29、0.28、0.27、0.26、0.25、0.24、0.23、0.22、0.21、および0.20である。ヘテロ接合性の欠失についてのPAV係数は、約0.33であることが多い。

20

【0332】

例えば、NRVを約1とし、ヘテロ接合性の欠失の期待レベルを約0.5に等しいとすると、ヘテロ接合性の欠失についてのPAVは、上記の式に従って、約0.495と決定される。この場合、例えば、約0.495を、第1のレベルの値へと加算することにより、ヘテロ接合性の欠失として類別された第1のレベルを調整する。

【0333】

一部の実施形態では、ホモ接合性の欠失についてのPAV係数は、約0.6～約0.4の間である。一部の実施形態では、ホモ接合性の欠失についてのPAV係数は、約0.60、0.59、0.58、0.57、0.56、0.55、0.54、0.53、0.52、0.51、0.50、0.49、0.48、0.47、0.46、0.45、0.44、0.43、0.42、0.41、および0.40である。ホモ接合性の欠失についてのPAV係数は、約0.5であることが多い。

30

【0334】

例えば、NRVを約1とし、ホモ接合性の欠失の期待レベルを約0に等しいとすると、ホモ接合性の欠失についてのPAVは、上記の式に従って、約1と決定される。この場合、例えば、約1を、第1のレベルの値へと加算することにより、ホモ接合性の欠失として類別された第1のレベルを調整する。

40

【0335】

ある特定の実施形態では、PAVは、コピー数の変動についての期待レベル（例えば、コピー数の変動の期待レベル）にほぼ等しいかまたは等しい。

【0336】

一部の実施形態では、調整を施す前にレベルのカウントを正規化する。ある特定の実施形態では、調整を施す前に、プロファイル中の一部または全部のレベルのカウントを正規化する。例えば、参照レベルのカウントまたはNRVに従って、レベルのカウントを正規化することができる。ある特定の実施形態では、調整を施す前に、レベルのカウント（例えば、第2のレベル）を、参照レベルのカウントまたはNRVに従って正規化し、プロファイル中の他の全てのレベル（例えば、第1のレベル）のカウントを、同じ参照レベルの

50

カウントまたはNRVと比べて正規化する。

【0337】

一部の実施形態では、プロファイルのレベルは、1つまたは複数の調整の結果として得られる。ある特定の実施形態では、プロファイルのレベルを、プロファイル中の1つまたは複数のレベルを調整した後で決定する。一部の実施形態では、1つまたは複数の調整を施した後で、プロファイルのレベルを再計算する。

【0338】

一部の実施形態では、コピー数の変動（例えば、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動）を、調整により決定する（例えば、直接的または間接的に決定する）。例えば、プロファイル中の調整されたレベル（例えば、調整された第1のレベル）は、母体のコピー数の変動として同定することができる。一部の実施形態では、調整の大きさは、コピー数の変動の種類（例えば、ヘテロ接合性の欠失、ホモ接合性の重複など）を指し示す。ある特定の実施形態では、プロファイル中の調整されたレベルを、コピー数の変動を表示するものとして、コピー数の変動についてのPAVの値に従って同定することができる。例えば、所与のプロファイルについて、PAVは、ホモ接合性の重複では約-1であり、ヘテロ接合性の重複では約-0.5であり、ヘテロ接合性の欠失では約0.5であり、ホモ接合性の欠失では約1である。前出の例では、約-1で調整されるレベルは、例えば、ホモ接合性の重複として同定することができる。一部の実施形態では、1つまたは複数のコピー数の変動は、1つまたは複数の調整を含むプロファイルまたはレベルから決定することができる。

【0339】

ある特定の実施形態では、プロファイル内の調整されたレベルを比較する。一部の実施形態では、異常および誤差は、調整されたレベルを比較することにより確認する。例えば、プロファイル中の1つまたは複数の調整されたレベルについて比較し、特定のレベルを、異常または誤差として確認することができることが多い。一部の実施形態では、異常または誤差を、レベルを構成する1つまたは複数の部分内で確認する。異常または誤差は、同じレベル内で（例えば、プロファイル内で）確認することもでき、隣接する（adjacent、contiguous、adjoining、またはabutting）部分を表示する1つまたは複数のレベルにおいて確認することもできる。一部の実施形態では、1つまたは複数の調整されたレベルは、隣接する（adjacent、contiguous、adjoining、またはabutting）部分のレベルであり、ここで、1つまたは複数の調整されたレベルについて比較し、異常または誤差を確認する。異常または誤差は、プロファイル内またはレベル中のピークまたはディップである可能性があり、ここで、ピークまたはディップの原因は、既知または未知である。ある特定の実施形態では、調整されたレベルについて比較し、異常または誤差を確認し、ここで、異常または誤差は、確率誤差、系統誤差、偶然誤差、または使用者誤差に起因する。一部の実施形態では、調整されたレベルについて比較し、異常または誤差を、プロファイルから除外する。ある特定の実施形態では、調整されたレベルについて比較し、異常または誤差を調整する。

【0340】

レベルに基づく胎児フラクションの決定

一部の実施形態では、胎児フラクションを、母体および／または胎児のコピー数の変動を表示するものとして類別されたレベルに従って決定する。例えば、胎児フラクションの決定は、胎児フラクションを決定するために活用される、母体および／または胎児のコピー数の変動についての期待レベルの評価を含むことが多い。一部の実施形態では、胎児フラクションを、コピー数の変動を表示するものとして類別されたレベル（例えば、第1のレベル）について、同じ種類のコピー数の変動について決定された期待レベルの範囲に従って決定する。胎児フラクションは、期待レベルの範囲内にある観察レベルに従って決定し、これにより、母体および／または胎児のコピー数の変動として類別することが多い。一部の実施形態では、胎児フラクションを、母体および／または胎児のコピー数の変動と

して類別された観察レベル（例えば、第1のレベル）が、同じ母体および／または胎児のコピー数の変動について決定された期待レベルと異なる場合に決定する。

【0341】

一部の実施形態では、レベル（例えば、第1のレベル、観察レベル）は、第2のレベルと有意に異なり、第1のレベルを、母体および／または胎児のコピー数の変動として類別し、胎児フラクションを、第1のレベルに従って決定する。一部の実施形態では、第1のレベルは、プロファイル中の第2のレベルと有意に異なる観察レベルおよび／または実験的に得られたレベルであり、胎児フラクションを、第1のレベルに従って決定する。一部の実施形態では、第1のレベルは、平均レベル、平均値レベル、または合計レベルであり、胎児フラクションを、第1のレベルに従って決定する。ある特定の実施形態では、第1のレベルおよび第2のレベルは、観察レベルおよび／または実験的に得られたレベルであり、胎児フラクションを、第1のレベルに従って決定する。場合によって、第1のレベルは、第1の部分のセットについての正規化されたカウントを含み、第2のレベルは、第2の部分のセットについての正規化されたカウントを含み、胎児フラクションを、第1のレベルに従って決定する。一部の実施形態では、第1のレベルの第1の部分のセットは、コピー数の変動（例えば、第1のレベルは、コピー数の変動を表示する）を含み、胎児フラクションを、第1のレベルに従って決定する。一部の実施形態では、第1のレベルの第1の部分のセットは、ホモ接合性またはヘテロ接合性の母体のコピー数の変動を含み、胎児フラクションを、第1のレベルに従って決定する。一部の実施形態では、プロファイルは、第1の部分のセットについての第1のレベルおよび第2の部分のセットについての第2のレベルを含み、第2の部分のセットは、実質的にコピー数の変動（例えば、母体のコピー数の変動、胎児のコピー数の変動、または母体のコピー数の変動および胎児のコピー数の変動）を含まず、胎児フラクションを、第1のレベルに従って決定する。

【0342】

一部の実施形態では、レベル（例えば、第1のレベル、観察レベル）は、第2のレベルと有意に異なり、第1のレベルを、母体および／または胎児のコピー数の変動に関して類別し、胎児フラクションを、第1のレベルおよび／またはコピー数の変動の期待レベルに従って決定する。一部の実施形態では、第1のレベルを、コピー数の変動についての期待レベルに従って、コピー数の変動に関して類別し、胎児フラクションを、第1のレベルと期待レベルとの差に従って決定する。ある特定の実施形態では、レベル（例えば、第1のレベル、観察レベル）を、母体および／または胎児のコピー数の変動として類別し、胎児フラクションを、第1のレベルとコピー数の変動の期待レベルとの差の2倍として決定する。一部の実施形態では、レベル（例えば、第1のレベル、観察レベル）を、母体および／または胎児のコピー数の変動として類別し、第1のレベルを、期待レベルから減じ、これにより、差を提供し、胎児フラクションを、差の2倍として決定する。一部の実施形態では、レベル（例えば、第1のレベル、観察レベル）を、母体および／または胎児のコピー数の変動として類別し、期待レベルを、第1のレベルから減じ、これにより、差を提供し、胎児フラクションを、差の2倍として決定する。

【0343】

胎児フラクションは、パーセントとして提示することが多い。例えば、胎児フラクションを、100で除算することができ、これにより、パーセント値を提供する。例えば、母体のホモ接合性の重複を表示し、155のレベルである第1のレベルと、母体のホモ接合性の重複についての期待レベルであって、150のレベルである期待レベルとでは、胎児フラクションは、10%（例えば、 $(\text{胎児フラクション} = 2 \times (155 - 150))$ ）として決定することができる。

【0344】

一部の実施形態では、胎児フラクションを、プロファイル中の2つまたはそれ超のレベルであって、コピー数の変動として類別されたレベルから決定する。例えば、場合によって、プロファイル中の2つまたはそれ超のレベル（例えば、2つまたはそれ超の第1のレベル）を、参照レベル（例えば、実質的にコピー数の変動を含まないレベルである、第2

のレベル)と有意に異なるものとして同定し、2つまたはそれ超のレベルを、母体および／または胎児のコピー数の変動を表示するものとして類別し、胎児フラクションを、2つまたはそれ超のレベルの各々から決定する。一部の実施形態では、胎児フラクションを、プロファイル中の約3つもしくはそれ超、約4つもしくはそれ超、約5つもしくはそれ超、約6つもしくはそれ超、約7つもしくはそれ超、約8つもしくはそれ超、または約9つもしくはそれ超の胎児フラクションの決定から決定する。一部の実施形態では、胎児フラクションを、プロファイル中の約10もしくはそれ超、約20もしくはそれ超、約30もしくはそれ超、約40もしくはそれ超、約50もしくはそれ超、約60もしくはそれ超、約70もしくはそれ超、約80もしくはそれ超、または約90もしくはそれ超の胎児フラクションの決定から決定する。一部の実施形態では、胎児フラクションを、プロファイル中の約100もしくはそれ超、約200もしくはそれ超、約300もしくはそれ超、約400もしくはそれ超、約500もしくはそれ超、約600もしくはそれ超、約700もしくはそれ超、約800もしくはそれ超、約900もしくはそれ超、または約1000もしくはそれ超の胎児フラクションの決定から決定する。一部の実施形態では、胎児フラクションを、プロファイル中の約10～約1000、約20～約900、約30～約700、約40～約600、約50～約500、約50～約400、約50～約300、約50～約200、または約50～約100の胎児フラクションの決定から決定する。

10

【0345】

一部の実施形態では、胎児フラクションを、プロファイル中の複数の胎児フラクションの決定の平均または平均値として決定する。ある特定の実施形態では、複数の胎児フラクションの決定から決定された胎児フラクションは、複数の胎児フラクションの決定の平均(例えば、平均、平均値、標準平均、中央値など)である。複数の胎児フラクションの決定から決定された胎児フラクションは、当技術分野で公知であるか、または本明細書で記載される適切な方法により決定される平均値であることが多い。一部の実施形態では、胎児フラクションの決定の平均値は、重み付き平均値である。一部の実施形態では、胎児フラクションの決定の平均値は、非重み付き平均値である。複数の胎児フラクションの決定から生成された平均値、中央値、または平均胎児フラクションの決定(すなわち、平均値、中央値、または平均胎児フラクションの決定値)は、場合によって、不確定値(例えば、分散、標準偏差、MADなど)と関連する。一部の実施形態では、複数の決定に由来する、平均値、中央値、または平均胎児フラクション値を決定する前に、1つまたは複数の逸脱した決定を除外する(本明細書でより詳細に記載される)。

20

30

【0346】

プロファイル中の一部の胎児フラクションの決定は、場合によって、全体的な胎児フラクションの決定(例えば、平均値または平均胎児フラクションの決定)に含まれない。一部の実施形態では、胎児フラクションの決定を、プロファイル中の第1のレベル(例えば、第2のレベルと有意に異なる第1のレベル)から導出し、第1のレベルは、遺伝子の変動を指し示さない。例えば、プロファイル中のいくつかの第1のレベル(例えば、スパイクまたはディップ)は、異常または未知の原因から生成される。このような値からは、真のコピー数の変動から得られる他の胎児フラクションの決定から有意に異なる胎児フラクションの決定を生じることが多い。一部の実施形態では、プロファイル中の他の胎児フラクションの決定から有意に異なる胎児フラクションの決定を確認し、胎児フラクションの決定から除外する。例えば、異常なスパイクおよびディップから得られる一部の胎児フラクションの決定は、それらをプロファイル中の他の胎児フラクションの決定と比較することにより確認し、全体的な胎児フラクションの決定から排除する。

40

【0347】

一部の実施形態では、平均値、中央値、または平均胎児フラクションの決定と有意に異なる、独立の胎児フラクションの決定は、確認され、認識され、かつ／または観察可能な差違である。ある特定の実施形態では、「有意に異なる」という用語は、「統計学的に異なる」および／または「統計学的な有意差」を意味し得る。「独立の」胎児フラクションの決定は、コピー数の変動として類別された特異的レベルから決定(例えば、一部の実施

50

形態では、単一の決定)された胎児フラクションでありうる。任意の適切な閾または範囲を使用して、胎児フラクションの決定が、平均値、中央値、または平均胎児フラクションの決定と有意に異なることを決定することができる。ある特定の実施形態では、胎児フラクションの決定は、平均値、中央値、または平均胎児フラクションの決定と有意に異なり、決定は平均または平均値からの逸脱パーセントとして表すことができる。ある特定の実施形態では、胎児フラクションの決定は、平均値、中央値、または平均胎児フラクションの決定と有意に異なり、約10パーセントまたはそれ超異なる。一部の実施形態では、胎児フラクションの決定は、平均値、中央値、または平均胎児フラクションの決定と有意に異なり、約15パーセントまたはそれ超異なる。一部の実施形態では、胎児フラクションの決定は、平均値、中央値、または平均胎児フラクションの決定と有意に異なり、約15%~約100%またはそれ超異なる。

10

【0348】

ある特定の実施形態では、胎児フラクションの決定は、平均値または平均胎児フラクションの決定と関連する複数の不確定値に従った、平均値、中央値、または平均胎児フラクションの決定と有意に異なる。不確定値および定数 n (例えば、信頼区間)により、範囲(例えば、不確定カットオフ)を規定することが多い。例えば、場合によって、不確定値は、胎児フラクションの決定についての標準偏差(例えば、 ± 5)であり、これに、定数 n (例えば、信頼区間)を乗じ、これにより、範囲または不確定カットオフ(例えば、 $5n \sim -5n$ 、場合によって、 5σ と称する)を規定する。一部の実施形態では、独立の胎児フラクションの決定は、不確定カットオフにより規定される範囲外にあり、平均値、中央値、または平均胎児フラクションの決定と有意に異なると考えられる。例えば、平均を10とし、不確定カットオフを3とすると、13超かまたは7未満である独立の胎児フラクションは、有意に異なる。一部の実施形態では、胎児フラクションの決定は、平均値、中央値、または平均胎児フラクションの決定と有意に異なり、不確定値の n 倍(例えば、 $n \times \sigma$)を超えて異なり、ここで、 n は、1、2、3、4、5、6、7、8、9、または10にほぼ等しいか、またはそれ超である。一部の実施形態では、胎児フラクションの決定は、平均値、中央値、または平均胎児フラクションの決定と有意に異なり、不確定値の n 倍(例えば、 $n \times \sigma$)を超えて異なり、ここで、 n は、1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2.0、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、または4.0にほぼ等しいか、またはそれ超である。

20

30

【0349】

一部の実施形態では、レベルは、胎児および/または母体の微小倍数性を表示する。一部の実施形態では、レベル(例えば、第1のレベル、観察レベル)は、第2のレベルと有意に異なり、第1のレベルは、母体および/または胎児のコピー数の変動として類別され、第1のレベルおよび/または第2のレベルは、胎児の微小倍数性および/または母体の微小倍数性を表示する。ある特定の実施形態では、第1のレベルは、胎児の微小倍数性を表示し、一部の実施形態では、第1のレベルは、母体の微小倍数性を表示する。第1のレベルは、胎児の微小倍数性および母体の微小倍数性を表示することが多い。一部の実施形態では、レベル(例えば、第1のレベル、観察レベル)は、第2のレベルと有意に異なり、第1のレベルは、母体および/または胎児のコピー数の変動として類別され、第1のレベルは、胎児および/または母体の微小倍数性を表示し、胎児フラクションは、胎児および/または母体の微小倍数性に従って決定される。場合によって、第1のレベルは、母体および/または胎児のコピー数の変動として類別され、第1のレベルは、胎児の微小倍数性を表示し、胎児フラクションは、胎児の微小倍数性に従って決定される。一部の実施形態では、第1のレベルは、母体および/または胎児のコピー数の変動として類別され、第1のレベルは、母体の微小倍数性を表示し、胎児フラクションは、母体の微小倍数性に従って決定される。一部の実施形態では、第1のレベルは、母体および/または胎児のコピー数の変動として類別され、第1のレベルは、母体および胎児の微小倍数性を表示し、胎

40

50

児フラクシオンは、母体および胎児の微小倍数性に従って決定される。

【0350】

一部の実施形態では、胎児フラクシオンの決定は、胎児および／または母体の微小倍数性を決定することを含む。一部の実施形態では、レベル（例えば、第1のレベル、観察レベル）は、第2のレベルと有意に異なり、第1のレベルを、母体および／または胎児のコピー数の変動として類別し、胎児および／または母体の微小倍数性を、第1のレベルおよび／または第2のレベルに従って決定し、胎児フラクシオンを決定する。一部の実施形態では、第1のレベルを、母体および／または胎児のコピー数の変動として類別し、胎児の微小倍数性を、第1のレベルおよび／または第2のレベルに従って決定し、胎児フラクシオンを、胎児の微小倍数性に従って決定する。ある特定の実施形態では、第1のレベルを、母体および／または胎児のコピー数の変動として類別し、母体の微小倍数性を、第1のレベルおよび／または第2のレベルに従って決定し、胎児フラクシオンを、母体の微小倍数性に従って決定する。一部の実施形態では、第1のレベルを、母体および／または胎児のコピー数の変動として類別し、母体および胎児の微小倍数性を、第1のレベルおよび／または第2のレベルに従って決定し、胎児フラクシオンを、母体および胎児の微小倍数性に従って決定する。

10

【0351】

胎児フラクシオンは、所与のレベルについて、またはコピー数の変動として類別されたレベルについて、母親の微小倍数性が、胎児の微小倍数性と異なる（例えば、胎児の微小倍数性と同じではない）場合に決定することが多い。一部の実施形態では、胎児フラクシオンは、母親が、重複についてホモ接合性（例えば、2の微小倍数性）であり、胎児が、同じ重複についてヘテロ接合性（例えば、1.5の微小倍数性）である場合に決定する。一部の実施形態では、胎児フラクシオンは、母親が、重複についてヘテロ接合性（例えば、1.5の微小倍数性）であり、胎児が、同じ重複についてホモ接合性（例えば、2の微小倍数性）であるか、または胎児において重複が存在しない（例えば、1の微小倍数性）場合に決定する。一部の実施形態では、胎児フラクシオンは、母親が、欠失についてホモ接合性（例えば、0の微小倍数性）であり、胎児が、同じ欠失についてヘテロ接合性（例えば、0.5の微小倍数性）である場合に決定する。一部の実施形態では、胎児フラクシオンは、母親が、欠失についてヘテロ接合性（例えば、0.5の微小倍数性）であり、胎児が、同じ欠失についてホモ接合性（例えば、0の微小倍数性）であるか、または胎児において欠失が存在しない（例えば、1の微小倍数性）場合に決定する。

20

30

【0352】

ある特定の実施形態では、母親の微小倍数性が、コピー数の変動として確認された所与のレベルについて、胎児の微小倍数性と同じである（例えば、同じとして確認された）場合は、胎児フラクシオンを決定することができない。一部の実施形態では、例えば、母親および胎児の両方が、同じコピー数のコピー数の変動を保有する場合の所与のレベルについて、胎児フラクシオンは決定されない。例えば、母親および胎児の両方が、同じ欠失についてホモ接合性であるか、または同じ重複についてホモ接合性である場合は、胎児フラクシオンを、コピー数の変動として類別されたレベルについて決定することができない。ある特定の実施形態では、母親および胎児の両方が、同じ欠失についてヘテロ接合性であるか、または同じ重複についてヘテロ接合性である場合は、胎児フラクシオンを、コピー数の変動として類別されたレベルについて決定することができない。複数の胎児フラクシオンの決定を試料について行う実施形態では、平均値、中央値、または平均の値から有意に逸脱する決定は、母体の倍数性が、胎児の倍数性に等しいコピー数の変動の結果として得られる場合があり、このような決定は、検討事項から除外することができる。

40

【0353】

一部の実施形態では、母体のコピー数の変動および胎児のコピー数の変動の微小倍数性は未知である。一部の実施形態では、コピー数の変動についての、胎児の微小倍数性および／または母体の微小倍数性の決定がなされない場合は、胎児フラクシオンを生成し、平均値、中央値、または平均胎児フラクシオンの決定と比較する。コピー数の変動について

50

の胎児フラクシヨンの決定であって、平均値、中央値、または平均胎児フラクシヨンの決定と有意に異なる決定は、場合によって、母親および胎児の微小倍数性が、コピー数の変動について同じであるためである。胎児フラクシヨンの決定であって、平均値、中央値、または平均胎児フラクシヨンの決定と有意に異なる決定は、差違の発生源または原因に関わらず、全体的な胎児フラクシヨンの決定から排除することが多い。一部の実施形態では、母親および／または胎児の微小倍数性を、当技術分野で公知の方法（例えば、標的化配列決定法）により決定および／または確かめる。

【0354】

胎児の倍数性

一部の実施形態では、胎児の倍数性の決定を使用して、一部分、遺伝子の変動（例えば、染色体異数性、トリソミー）の存在または非存在の決定を行う。胎児の倍数性は、一部分、本明細書で記載される方法を含む、胎児フラクシヨンの決定の適切な方法により決定された胎児フラクシヨンの尺度から決定することができる。一部の実施形態では、胎児の倍数性を、胎児フラクシヨンの決定および等式（8）、（20）、（21）、またはこれらの変化形もしくは派生形に従って決定する（実施例2）。一部の実施形態では、胎児の倍数性を、下記に記載される方法により決定する。一部の実施形態では、下記に記載される各方法は、複数の試料について、ゲノムの部分（すなわち、部分 i ）について決定された計算参照カウント F_i （場合によって、 f_i としても表示される）を必要とし、ここで、ゲノムの部分 i についての胎児の倍数性は、正倍数性である。一部の実施形態では、不確定値（例えば、標準偏差、 σ ）を、参照カウント f_i について決定する。一部の実施形態では、参照カウント f_i 、不確定値、試験試料カウントおよび／または測定された胎児フラクシヨン（ F ）を、下記に記載される方法に従って、胎児の倍数性を決定するのに使用する。一部の実施形態では、参照カウント（例えば、平均、平均値、または中央値による参照カウント）を、本明細書で記載される方法（例えば、部分に関する正規化、GC含有量による正規化、線形最小二乗回帰および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS、PERUN、ChAI、RM、GCRMおよび／またはこれらの組合せ）により正規化する。一部の実施形態では、参照カウントを、PERUNにより正規化する場合、正倍数体であるゲノムのセグメントの参照カウントは、1に等しい。一部の実施形態では、ゲノムの部分またはセグメントについての参照カウント（例えば、正倍数体であることが既知の胎児についての）および試験試料のカウントの両方を、PERUNにより正規化し、参照カウントは、1に等しい。同様に、一部の実施形態では、カウントを、参照カウントの中央値により正規化する（すなわち、参照カウントの中央値で除算する）場合も、正倍数体であるゲノムの部分またはセグメントの参照カウントは、1に等しい。例えば、一部の実施形態では、ゲノムの部分またはセグメントについての、参照カウント（例えば、正倍数体である胎児についての）および試験試料のカウントの両方を、中央値参照カウントにより正規化し、正規化された参照カウントは、1に等しく、試験試料カウントは、中央値参照カウントにより正規化する（例えば、中央値参照カウントで除算する）。一部の実施形態では、ゲノムの部分またはセグメントについての、参照カウント（例えば、正倍数体である胎児についての）および試験試料のカウントの両方を、GCRM、GC、RM、または適切な方法により正規化する。一部の実施形態では、参照カウントは、平均、平均値、または中央値による参照カウントである。参照カウントは、部分についての正規化されたカウント（例えば、正規化されたゲノム区分のレベル）であることが多い。一部の実施形態では、参照カウントおよび試験試料についてのカウントは、未加工のカウントである。一部の実施形態では、参照カウントを、平均、平均値、または中央値によるカウントプロファイルから決定する。一部の実施形態では、参照カウントは、計算されたゲノム区分のレベルである。一部の実施形態では、参照試料の参照カウントおよび試験試料のカウント（例えば、患者試料、例えば、 y_i ）を、同じ方法または処理により正規化する。

【0355】

一部の実施形態では、胎児フラクシヨン（ F ）の測定値を決定する。次いで、この胎児

10

20

30

40

50

フラクション値を、等式（８）、これらの派生形、または変化形に従って胎児の倍数性を決定するのに使用する。一部の実施形態では、胎児が正倍数体であれば負の値となり、胎児が正倍数体でなければ正の値となる。一部の実施形態では、負の値は、胎児が、検討されるゲノムのセグメントについて正倍数体であることを指し示す。ある特定の実施形態では、負ではない値は、胎児が、異数性（例えば、重複）を含むことを指し示す。ある特定の実施形態では、負ではない値は、胎児が、トリソミーを含むことを指し示す。ある特定の実施形態では、任意の正の値は、胎児が、異数性（例えば、トリソミー、重複）を含むことを指し示す。

【０３５６】

一部の実施形態では、残差平方和を決定する。例えば、残差平方和を表示する等式であって、等式（８）から導出された等式を、等式（１３）に例示する。一部の実施形態では、残差平方和を、等式（８）から、１の値へと設定した倍数性値 X （等式（９）を参照されたい）および $3/2$ の値へと設定した倍数性値（等式（１３）を参照されたい）について決定する。一部の実施形態では、残差平方和（等式（９）および（１３））を、ゲノムまたは染色体のセグメントについて（例えば、ゲノムのセグメント中の参照ゲノムの部分 i 全てについて）決定する。例えば、残差平方和（例えば、等式（９）および（１３））を、第２１染色体、第１３染色体、第１８染色体、またはこれらの部分について決定することができる。一部の実施形態では、胎児の倍数性状態を決定するために、等式（１３）の結果を、等式（９）から減じて、値 F_i （例えば、等式（１４）を参照されたい）に到達する。ある特定の実施形態では、値 F_i の符号（すなわち、正または負）により、胎児の異数性の存在または非存在を決定する。ある特定の実施形態では、負である F_i の値（例えば、等式（１４）に由来する）は、異数性の非存在を指し示し（例えば、胎児は、参照ゲノムの部分 i について正倍数体であり）、負ではない F_i の値は、異数性の存在（例えば、トリソミー）を指し示す。

【０３５７】

一部の実施形態では、参照カウント f_i 、参照カウントについての不確定値 σ 、および／または測定された胎児フラクション（ F ）を、等式（９）および（１３）において使用して、参照ゲノムの部分 i の全ての合計についての残差平方和を決定する。一部の実施形態では、参照カウント f_i 、参照カウントについての不確定値 σ 、および／または測定された胎児フラクション（ F ）を、等式（９）および（１３）において使用して、胎児の倍数性を決定する。一部の実施形態では、試験試料についての、部分 i についての y_i により表示されるカウント（例えば、正規化されたカウント、例えば、計算されたゲノム区分のレベル）を使用して、部分 i についての胎児の倍数性状態を決定する。例えば、ある特定の実施形態では、ゲノムのセグメントについての倍数性状態を、試験試料について決定された参照カウント f_i 、不確定値（例えば、参照カウントに由来する）、胎児フラクション（ F ）、および試験試料について決定されたカウント y_i に従って決定し、ここで、倍数性状態は、等式（１４）、またはこれらの派生形もしくは変化形に従って決定する。一部の実施形態では、カウント y_i および／または参照カウントを、本明細書で記載される方法（例えば、部分に関する正規化、GC含有量による正規化、線形最小二乗回帰および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS、PERUN、ChAI、RM、GCRM、およびこれらの組合せ）により正規化する。一部の実施形態では、ゲノムまたは染色体の部分またはセグメントについての胎児の倍数性状態（例えば、正倍数体、異数体、トリソミー）を、上記および実施例の節で記載される非限定的な例により決定する。

【０３５８】

一部の実施形態では、胎児フラクションを、試験試料から決定し、カウント y を、試験試料について決定し、これらの両方を使用して、胎児についての倍数性を、試験試料から決定する。本明細書で記載される方法についてのある特定の実施形態では、 X により表示された胎児の倍数性値は、一定値または仮定値ではない。本明細書で記載される方法についてのある特定の実施形態では、胎児フラクション F は、一定である。一部の実施形態で

10

20

30

40

50

は、倍数性（例えば、倍数性値）を、ゲノムの部分またはセグメントについて、等式（20）または（21）に従って決定する（実施例2）。この方法についての一部の実施形態では、倍数性値を決定し、ここで、値は、1、 $3/2$ 、または $5/4$ に近い。一部の実施形態では、約1の倍数性値は、正倍数体の胎児を指し示し、約 $3/2$ の値は、胎児のトリソミーを指し示し、双子の場合は、約 $5/4$ の値は、検討されるゲノムの部分またはセグメントについて、一方の胎児が、トリソミーを含み、他方の胎児が正倍数体であることを指し示す。胎児の倍数性の決定から、胎児の異数性の存在または非存在を決定することに関するさらなる情報については、下記の別の節で論じる。

【0359】

一部の実施形態では、その決定値で一定の胎児フラクションを決定し、胎児の倍数性を、回帰から決定する。任意の適切な回帰であって、その非限定的な例が、線形回帰、非線形回帰（例えば、多項式回帰）などを含む、任意の適切な回帰を活用することができる。一部の実施形態では、線形回帰を、等式（8）、（20）、（21）、および／またはこれらの派生形もしくは変化形に従って使用する。一部の実施形態では、線形回帰は、等式（8）、（20）、（21）、および／またはこれらの派生形もしくは変化形から導出される残差平方和に従って使用する。一部の実施形態では、胎児の倍数性を、等式（8）、（20）、（21）、および／またはこれらの派生形もしくは変化形に従って決定し、回帰は使用しない。一部の実施形態では、胎児の倍数性を、等式（8）、（20）、（21）、および／またはこれらの派生形もしくは変化形から導出される残差平方和に従って、参照ゲノムの複数の部分 i について決定し、回帰は使用しない。等式の派生形とは、等式の数学的証明から得られる、等式の任意の変化形である。

【0360】

一部の実施形態では、参照カウント f_i （本明細書で既に記載した）、不確定値 σ 、および／または測定された胎児フラクション（ F ）を、等式（20）および（21）で使用して、胎児の倍数性を決定する。一部の実施形態では、参照カウント f_i 、不確定値 σ 、および／または測定された胎児フラクション（ F ）を、等式（20）または（21）で使用して、胎児の倍数性 X を、部分 i について、または参照ゲノムの複数の部分 i の合計について（例えば、染色体またはそのセグメントについての、参照ゲノムの部分 i の全ての合計について）決定する。一部の実施形態では、試験試料の部分 i について、 y_i により表示されるカウント（例えば、正規化されたカウント、計算されたゲノム区分のレベル）を、参照ゲノムの複数の部分 i により表示されるゲノムのセグメントについての、胎児の倍数性を決定するのに使用する。例えば、ある特定の実施形態では、ゲノムのセグメントについての倍数性 X を、試験試料について決定された参照カウント f_i 、不確定値、胎児フラクション（ F ）、および試験試料について決定されたカウント y_i に従って決定し、ここで、倍数性を、等式（20）、（21）、またはこれらの派生形もしくは変化形に従って決定する。一部の実施形態では、カウント y_i および／または参照カウントを、本明細書で記載される方法（例えば、部分に関する正規化、GC含有量による正規化、線形最小二乗回帰および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS、PERUN、ChAI、RM、GCRM、およびこれらの組合せ）により正規化する。一部の実施形態では、カウント y_i および／または参照カウントを、同じ方法（例えば、部分に関する正規化、GC含有量による正規化、線形最小二乗回帰および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS、PERUN、ChAI、RM、GCRM、本明細書で記載される方法またはこれらの組合せ）により正規化および／または処理する。一部の実施形態では、カウント y_i および f_i は、ゲノムまたは染色体の同じ部分またはセグメントへとマッピングしたカウントである。

【0361】

不確定値 σ は、その非限定的な例が、標準偏差、標準誤差、計算された分散、 p 値、および／または平均絶対偏差（MAD）を含む、適切な誤差の尺度でありうる。不確定値 σ は、その非限定的な例が、Zスコア、Z値、 t 値、 p 値、交差検証誤差、ゲノム区分のレベル、計算されたゲノム区分のレベル、レベル、カウントなど、またはこれらの組合せを

含む、任意の適切な測定値について決定することができる。一部の実施形態では、 σ を、1の値に設定する。一部の実施形態では、 σ を、1の値に設定しない。一部の実施形態では、 σ の値を推定し、場合によって、測定および／または計算する。

【0362】

一部の実施形態では、 M_i とは、ゲノムの部分*i*についての母親の倍数性（すなわち、母体の倍数性）である。一部の実施形態では、 M_i を、 y_i を決定する同じ患者（例えば、同じ試験試料）について決定する。一部の実施形態では、母体の倍数性 M_i は、既知であるか、または本明細書で記載される方法に従って決定する。一部の実施形態では、穴埋めの前に、または穴埋めの後で（例えば、レベルの調整を施した後で）、母体の倍数性を決定する。ある特定の実施形態では、 M_i を、プロファイルの視覚化から推定または決定する。一部の実施形態では、母体の倍数性 M_i は、既知ではない。一部の実施形態では、母体の倍数性 M_i を、仮定する。例えば、一部の実施形態では、母親が、評価されるゲノムのセグメントに欠失および／または重複を有さないことが仮定されるかまたは既知である。一部の実施形態では、母体の倍数性が1であることが仮定されるかまたは既知である。一部の実施形態では、穴埋めの後で（例えば、レベルの調整を施した後で）、母体の倍数性を1の値に設定する。一部の実施形態では、母体の倍数性を無視し、1の値に設定する。一部の実施形態では、母親が、評価されるゲノムのセグメントに欠失および／または重複を有しないと仮定して、等式（21）を、等式（20）から導出する。

10

【0363】

一部の実施形態では、胎児の倍数性を決定するための方法は、妊娠中の雌から得られた試験試料についての核酸配列の読取りに従った方法である。一部の実施形態では、配列の読取りは、試料（例えば、試験試料）に由来する循環無細胞核酸についての読取りである。一部の実施形態では、胎児の倍数性を決定するための方法は、参照ゲノムの部分へとマッピングした配列の読取りのカウントを得るステップを含む。一部の実施形態では、配列の読取りを、参照ゲノムの部分のサブセットへとマッピングする。一部の実施形態では、胎児の倍数性の決定は、胎児フラクションを決定することを含む。一部の実施形態では、胎児の倍数性の決定は、ゲノム区分のレベルの計算または決定を含む。ある特定の実施形態では、胎児の倍数性の決定は、胎児フラクションの決定およびゲノム区分のレベルの計算または決定を含む。一部の実施形態では、胎児フラクションおよび計算されたゲノム区分のレベルは、同じ試験試料（例えば、試験試料の同じ部分）から決定する。一部の実施形態では、胎児フラクションおよび計算されたゲノム区分のレベルは、同じ試験試料（例えば、試験試料の同じ部分）から得られる同じ読取りから決定する。一部の実施形態では、胎児フラクションおよび計算されたゲノム区分のレベルは、同じ配列決定の実行および／または同じフローセルから得られる同じ読取りから決定する。一部の実施形態では、胎児フラクションおよび計算されたゲノム区分のレベルは、同じ器具および／または機器（例えば、配列決定装置、フローセルなど）により決定する。

20

30

【0364】

一部の実施形態では、胎児の倍数性を決定するための方法を、胎児フラクションの決定および正規化されたカウント（例えば、計算されたゲノム区分のレベル）に従って決定し、ここで、胎児フラクションの決定および正規化されたカウント（例えば、計算されたゲノム区分のレベル）は、試験試料の異なる一部分（例えば、異なるアリコート、または、例えば、同じ被験体もしくは患者からほぼ同時に採取された異なる試験試料）から決定する。例えば、場合によって、胎児フラクションを、試験試料の第1の一部分から決定し、正規化されたカウントおよび／またはゲノム区分のレベルは、試験試料の第2の部分から決定する。一部の実施形態では、胎児フラクションおよび計算されたゲノム区分のレベルは、同じ被験体（例えば、患者）から採取される、異なる試験試料（例えば、試験試料の異なる一部分）から決定する。一部の実施形態では、胎児フラクションおよび計算されたゲノム区分のレベルは、異なる時点において得られた読取りから決定する。一部の実施形態では、胎児フラクションの決定および正規化されたカウント（例えば、計算されたゲノム区分のレベル）は、異なる器具および／または異なる機器（例えば、配列決定装置、フ

40

50

ローセルなど)により決定する。

【0365】

決定 (d e c i s i o n) 分析の特徴

一部の実施形態では、アウトカム決定 (d e t e r m i n a t i o n) (例えば、判定を下すこと) 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決定 (d e t e r m i n a t i o n) は、決定 (d e c i s i o n) 分析に従って下す。例えば、決定 (d e c i s i o n) 分析は、場合によって、1つまたは複数の結果、結果の評価、ならびに決定 (d e c i s i o n) の結果、評価、および／または可能な帰結に基づく一連の決定 (d e c i s i o n) をもたらし、処理のいくつかの岐路であって、最終決定 (d e c i s i o n) が下される岐路で終結する、1つまたは複数の方法の適用を含む。一部の実施形態では、決定分析は、決定木である。一部の実施形態では、決定分析は、1つまたは複数の処理 (例えば、処理ステップ、例えば、アルゴリズム) の協調的な使用を含む。決定分析は、関係者、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、またはこれらの組合せにより実施することができる。一部の実施形態では、決定 (d e c i s i o n) 分析は、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または非存在を決定する (d e t e r m i n i n g) 方法であって、決定分析を活用しない (例えば、決定 (d e t e r m i n a t i o n) を正規化されたカウントから直接下す) 場合と比較して、偽陰性の決定 (d e t e r m i n a t i o n) が低減され、偽陽性の決定 (d e t e r m i n a t i o n) が低減された方法を含む。一部の実施形態では、決定 (d e c i s i o n) 分析は、1つまたは複数の微小重複または微小欠失と関連する状態の存在または非存在を決定すること (d e t e m i n i n g) を含む。例えば、一部の実施形態では、決定分析は、被験体由来する試験試料についての、ディジョージ症候群と関連する1つまたは複数の遺伝子の変動の存在または非存在を決定することを含む。一部の実施形態では、決定分析は、被験体由来する試験試料についての、ディジョージ症候群の存在または非存在を決定することを含む。

【0366】

一部の実施形態では、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) についてのプロファイル生成することを含む。プロファイルは、公知であるかまたは本明細書に記載され、参照ゲノムの部分へとマッピングした配列の読取りのカウントを得るステップ、カウントを正規化するステップ、レベルを正規化するステップ、穴埋めするステップなど、またはこれらの組合せを含むことが多い、任意の適切な方法により生成することができる。参照ゲノムへとマッピングした配列の読取りのカウントを得るステップは、試料を得る (例えば、妊娠中の雌被験体から) こと、試料由来する核酸 (例えば、循環無細胞核酸) の配列決定、配列の読取りを得ること、配列の読取りを参照ゲノムの部分へとマッピングすることなど、およびこれらの組合せを含みうる。一部の実施形態では、プロファイルの生成は、参照ゲノムの部分へとマッピングしたカウントを正規化し、これにより、計算されたゲノム区分のレベルを提示することを含む。

【0367】

一部の実施形態では、決定分析は、セグメント化することを含む。一部の実施形態では、セグメント化することにより、プロファイルを改変および／または変換し、これにより、プロファイルの1つまたは複数の分解レンダリングを提示する。セグメント化処理にかけられたプロファイルは、参照ゲノム内またはこれらの部分 (例えば、常染色体および性染色体) 中の部分 (例えば、ビン) へとマッピングし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。本明細書で対処される通り、1つまたは複数の適切な正規化処理 (例えば、P E R U N、L O E S S、G C - L O E S S、主成分正規化 (C h A I)、またはこれらの組合せ) により、部分へとマッピングした未加工のカウントを正規化して、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生成することができる。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、場合によって、プロファイルの、ゲノム、染色体またはそのセグメン

トの表示への変換である。

【0368】

ある特定の実施形態では、セグメント化のために活用されるセグメント化処理により、プロファイル中の1つまたは複数のレベルであって、プロファイル中の1つまたは複数の他のレベルと異なる（例えば、実質的または有意に異なる）レベルを位置特定および同定する。本明細書では、プロファイル内でセグメント化処理に従って同定されるレベルであって、プロファイル中の別のレベルと異なり、プロファイル中の別のレベルと異なるエッジを有するレベルを、ウェーブレットと称し、より一般に、個別セグメントについてのレベルと称する。セグメント化処理により、正規化されたカウントまたは正規化されたレベルのプロファイルから、1つまたは複数の個別セグメントまたはウェーブレットを同定し、分解レンダリングを生成することができる。個別セグメントは一般に、セグメント化されるもの（例えば、染色体、染色体（複数）、常染色体）より少数の部分（例えば、

10

【0369】

一部の実施形態では、セグメント化することにより、プロファイル中の個別セグメントおよびウェーブレットのエッジを位置特定および同定する。ある特定の実施形態では、1つまたは複数の個別セグメントのエッジおよび1つまたは複数のウェーブレットのエッジの一方または両方を同定する。例えば、セグメント化処理により、プロファイル中の個別セグメントまたはウェーブレットの右エッジおよび／または左エッジの位置（例えば、ゲノム座標、例えば、部分の位置）を同定することができる。個別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、個別セグメントまたはウェーブレットは、左エッジおよび右エッジを含みうる。一部の実施形態では、表示または図示に応じて、左エッジは、5' - エッジであることが可能であり、右エッジは、プロファイル中の核酸セグメントの3' - エッジでありうる。一部の実施形態では、左エッジは、3' - エッジであることが可能であり、右エッジは、プロファイル中の核酸セグメントの5' - エッジでありうる。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、したがって、一部の実施形態では、プロファイルのエッジにより、レベルのどのエッジが、5' - エッジであり、どのエッジが3' - エッジであるのかを決定する。一部の実施形態では、プロファイルのエッジおよび／または個別セグメント（例えば、ウェーブレット）の一方または両方は、染色体のエッジである。

20

30

【0370】

一部の実施形態では、個別セグメントまたはウェーブレットのエッジを、参照試料（例えば、参照プロファイル）について生成された分解レンダリングに従って決定する。一部の実施形態では、ヌルエッジの高さの分布を、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決定する（例えば、図3を参照されたい）。ある特定の実施形態では、プロファイル中の個別セグメントまたはウェーブレットのエッジを、個別セグメントまたはウェーブレットのレベルが、ヌルエッジの高さの分布の外側にある場合に同定する。一部の実施形態では、プロファイル中の個別セグメントまたはウェーブレットのエッジを、参照プロファイルについての分解レンダリングに従って計算されたZスコアに従って同定する。

40

【0371】

場合によって、セグメント化することにより、プロファイル中の、2つまたはそれ超の個別セグメントまたはウェーブレット（例えば、2つまたはそれ超の断片化レベル、2つまたはそれ超の断片化セグメント）を生成する。一部の実施形態では、セグメント化処理から導出された分解レンダリングは、過剰セグメント化または断片化されており、複数の個別セグメントまたはウェーブレットを含む。場合によって、セグメント化することにより生成される個別セグメントまたはウェーブレットは、実質的に異なり、場合によって、セグメント化することにより生成される個別セグメントまたはウェーブレットは、実質的に同様である。実質的に同様な個別セグメントまたはウェーブレット（例えば、実質的に同様なレベル）とは、各々が、所定の不確定性のレベル未満異なるゲノム区分のレベル（

50

例えば、レベル)を有する、セグメント化されたプロファイル中の、2つまたはそれ超の隣接する個別セグメントまたはウェーブレットを指すことが多い。一部の実施形態では、実質的に同様な個別セグメントまたはウェーブレットとは、互いと隣接し、介在セグメントまたは介在ウェーブレットで隔てられていない。一部の実施形態では、実質的に同様な個別セグメントまたはウェーブレットは、1つまたは複数の小型のセグメントまたはウェーブレットで隔てられている。一部の実施形態では、実質的に同様な個別セグメントまたはウェーブレットは、約1つ〜約20、約1つ〜約15、約1つ〜約10、または約1つ〜約5つの部分(例えば、ビン)で隔てられており、ここで、介在部分のうちの1つまたは複数のレベルは、実質的に同様な個別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。一部の実施形態では、実質的に同様な個別セグメントまたはウェーブレットのレベルは、不確定性のレベルの約3倍未満、約2倍未満、約1倍未満、または約0.5倍未満異なる。一部の実施形態では、実質的に同様な個別セグメントまたはウェーブレットは、3 MAD未満(例えば、3シグマ未満)、2 MAD未満、1 MAD未満、または約0.5 MAD未満異なる中央値ゲノム区分のレベルを含み、ここで、MADは、セグメントまたはウェーブレットの各々の中央値ゲノム区分のレベルから計算する。一部の実施形態では、実質的に異なる個別セグメントまたはウェーブレットは、隣接しないか、または10またはそれ超、15またはそれ超、もしくは20またはそれ超の部分で隔てられている。実質的に異なる個別セグメントまたはウェーブレットのレベルは一般に、実質的に異なる。ある特定の実施形態では、実質的に異なる個別セグメントまたはウェーブレットは、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倍超異なるレベルを含む。一部の実施形態では、実質的に異なる個別セグメントまたはウェーブレットは、2.5 MAD超(例えば、2.5シグマ超)、3 MAD超、4 MAD超、約5 MAD超、または約6 MAD超異なる中央値ゲノム区分のレベルを含み、ここで、MADは、個別セグメントまたはウェーブレットの各々の中央値ゲノム区分のレベルから計算する。

10

20

【0372】

一部の実施形態では、セグメント化処理は、プロファイル内またはそのセグメント中の1つまたは複数の個別セグメントまたはウェーブレット(例えば、レベル)についての、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スコア、Z値、p値など、またはこれらの組合せの決定(例えば、計算)を含む。一部の実施形態では、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スコア、Z値、p値など、またはこれらの組合せを、個別セグメントまたはウェーブレットについて決定する(例えば、計算する)。

30

【0373】

一部の実施形態では、セグメント化を、1つの処理または複数の下位処理であって、その非限定的な例が、分解生成処理(例えば、ウェーブレット分解生成処理)、閾化、レベル化、スムージングなど、またはこれらの組合せを含む下位処理を含む処理により達成する。閾化、レベル化、スムージングなどは、分解生成処理と共に実施することができ、ウェーブレット分解レンダリング処理に言及する、本明細書の下記で記載される。

40

【0374】

ウェーブレットセグメンテーション処理

一部の実施形態では、セグメント化を、ウェーブレット分解生成処理に従って実施する。一部の実施形態では、セグメント化を、2つまたはそれ超のウェーブレット分解生成処理に従って実施する。一部の実施形態では、ウェーブレット分解生成処理により、プロファイル中の1つまたは複数のウェーブレットを同定し、プロファイルの分解レンダリングを提示する。

【0375】

セグメント化は、本明細書で記載されるかまたは当技術分野で公知である、任意の適切なウェーブレット分解生成処理により、完全にまたは一部分実施することができる。ウェ

50

ーブレット分解生成処理の非限定的な例は、ハールウェーブレットセグメンテーション (Haar, Alfred (1910年)、「Zur Theorie der orthogonalen Funktionensysteme」、Mathematische Annalen、69巻(3号):331~371頁;Nason, G.P. (2008年)、「Wavelet methods in Statistics」、R. Springer、New York.) (例えば、WaveThresh) であるWaveThresh、適切なバイナリ再帰的セグメンテーション処理であるサーキュラーバイナリセグメンテーション(CBS) (Olshen, AB、Venkatraman, ES、Lucito, R、Wigler, M (2004年)、「Circular binary segmentation for the analysis of array-based DNA copy number data」、Biostatistics、5巻、4号:557~72頁;Venkatraman, ES、Olshen, AB (2007年)、「A faster circular binary segmentation algorithm for the analysis of array CGH data」、Bioinformatics、23巻、6号:657~63頁)、MODWT (Maximal Overlap Discrete Wavelet Transform) (L. Hsu, S. Self、D. Grove、T. Randolph、K. Wang、J. Delrow、L. Loo、およびP. Porter、「Denoising array-based comparative genomic hybridization data using wavelets」、Biostatistics (Oxford、England)、6巻、2号、211~226頁、2005年)、定常ウェーブレット(SWT) (Y. WangおよびS. Wang、「A novel stationary wavelet denoising algorithm for array-based DNA copy number data」、International Journal of Bioinformatics Research and Applications、3巻、2号、206~222頁、2007年)、双対木複素ウェーブレット変換(DTCWT) (Nha, N.、H. Heng、S. Orintara、およびW. Yuhang (2007年)、「Denoising of Array-Based DNA Copy Number Data Using The Dual-tree Complex Wavelet Transform」、137~144頁)、最大エントロピーセグメンテーション(maximum entropy segmentation)、エッジ検出カーネルによるコンボリューション(convolution with edge detection kernel)、ジェンセンシャノンダイバージェンス(Jensen Shannon Divergence)、カルバックライブラーダイバージェンス(Kullback-Leibler divergence)、バイナリ再帰的セグメンテーション(Binary Recursive Segmentation)、フーリエ変換など、またはこれらの組合せを含む。

【0376】

ウェーブレット分解生成処理は、その非限定的な例が、UNIX (登録商標)、Linux (登録商標)、Oracle、Windows (登録商標)、Ubuntu、ActionScript、C、C++、C#、Haskell、Java (登録商標)、JavaScript (登録商標)、Objective-C、Perl、Python、Ruby、Smalltalk、SQL、Visual Basic、COBOL、Fortran、UML、HTML (例えば、PHPによる)、PGP、G、R、Sなど、またはこれらの組合せを含む、適切な言語 (例えば、当技術分野で公知のコンピュータプログラミング言語) および/またはオペレーティングシステムで書かれた適切なソフトウェア、モジュール、および/またはコードにより表示または実施することができる。一部の実施形態では、適するウェーブレット分解生成処理を、SコードもしくはRコードまたはパッケージ (例えば、Rパッケージ) で表示する。CRANまたはCRANミラーサイト (例えば、Comprehensive R Archive Network (CRAN); インターネットURL: cran.us.r-project.org) からのダウンロードのためには、ウェーブレット分解生成処理のためのR、Rソースコード、Rプログラム、Rパッケージ、およびRドキュメンテーションが利用可能である。CRANは、世界中のRのためのコードおよびドキュメンテーションの、同一の最新バージョンを保存する、ftpサーバーおよびウェブサーバーのネットワークである。例えば、WaveThresh (WaveThresh: Wavelets statistics and transforms; インターネットURL: cran.r-project.org/web/packages/wavethresh/index.html) およびWaveThreshについての詳細な記載 (「WaveThresh」パッケージ; インターネットURL: cran.r-project.org/web/package/wavethresh/wavethresh.pdf) が、ダウンロードのために利用可能で

ありうる。一部の実施形態では、ウェーブレット分解生成処理のためのRコード（例えば、最大エントロピーセグメンテーション）については、実施例4で記載する。CBS法のためのRコードの例は、ダウンロードすることができる（例えば、DNAcopy；インターネットURL：bioconductor.org/packages/2.12/bioc/html/DNAcopy.htmlまたは「DNAcopy」パッケージ；インターネットURL：bioconductor.org/packages/release/bioc/manuals/DNAcopy/man/DNAcopy.pdf）。

【0377】

一部の実施形態では、ウェーブレット分解生成処理（例えば、ハールウェーブレットセグメンテーション、例えば、WaveThresh）は、閾化を含む。一部の実施形態では、閾化により、シグナルをノイズから識別する。ある特定の実施形態では、閾化により、どのウェーブレット係数（例えば、ノード）が、シグナルを指し示し、保持すべきであり、どのウェーブレット係数が、ノイズの反映を指し示し、除外すべきであるのかを決定する。一部の実施形態では、閾化は、1つまたは複数の変数パラメータを含み、ここで、使用者は、パラメータの値を定める。一部の実施形態では、閾化パラメータ（例えば、閾化パラメータ、ポリシーパラメータ）により、ウェーブレット分解生成処理で活用されるセグメント化の量について記載または規定することができる。任意の適切なパラメータ値を使用することができる。一部の実施形態では、閾化パラメータを使用する。一部の実施形態では、閾化パラメータ値は、ソフトな閾化である。ある特定の実施形態では、ソフトな閾化を活用して、小さな係数および有意でない係数を除外する。ある特定の実施形態では、ハードな閾化を活用する。ある特定の実施形態では、閾化は、ポリシーパラメータを含む。任意の適切なポリシー値を使用することができる。一部の実施形態では、使用されるポリシーは、「ユニバーサル」ポリシーであり、一部の実施形態では、使用されるポリシーは、「シュア」ポリシーである。

【0378】

一部の実施形態では、ウェーブレット分解生成処理（例えば、ハールウェーブレットセグメンテーション、例えば、WaveThresh）は、レベル化を含む。一部の実施形態では、閾化の後で、いくつかの高レベルの係数が残る。これらの係数は、元のシグナル中の急勾配の変化または大きなスパイクを表示し、ある特定の実施形態では、レベル化により除外される。一部の実施形態では、レベル化は、値の、分解レベル c として公知のパラメータへの割当てを含む。ある特定の実施形態では、最適の分解レベルを、染色体の長さ（例えば、プロファイルの長さ）、所望のウェーブレット長さなど、1つまたは複数の決定値に従って決定して、胎児フラクション、配列カバレッジ（例えば、プレックスレベル）、および正規化されたプロファイルのノイズレベルを検出する。ゲノム、染色体、またはプロファイルのセグメントの所与の長さ（ N_{chr} ）について、ウェーブレット分解レベル c は、場合によって、等式 $N_{micro} = N_{chr} / 2^{c+1}$ に従って、最小ウェーブレット長さ N_{micro} と関係づけられる。一部の実施形態では、サイズ N_{micro} 。またはそれ超の微小欠失を検出するために、所望の分解レベル c を、以下の等式： $c = \log_2 (N_{chr} / N_{micro}) - 1$ に従って決定する。例えば、 $N_{chr} = 4096$ 参照ゲノムの部分であり、 $N_{micro} = 128$ 参照ゲノムの部分であれば、分解レベル c は4であり、ある特定の場合には、 $c \pm 1$ レベル（すなわち、約3～約5）を使用することができる。一部の実施形態では、分解レベル c は、約1、2、3、4、5、6、7、8、9、または10である。一部の実施形態では、 N_{micro} を検出するのに所望される最小ウェーブレット長さは、約1Mb、2Mb、3Mb、4Mb、5Mb、6Mb、7Mb、8Mb、10Mb、15Mbであるか、または約20Mbを超える。一部の実施形態では、 N_{micro} は、所定の N_{micro} である。一部の実施形態では、配列カバレッジの量（例えば、プレックスレベル）および胎児フラクションは、 N_{micro} に反比例する。例えば、試料中の胎児フラクションの量が増加すると、検出に所望される最小ウェーブレット長さは減少する（すなわち、分解能が増加する）。一部の実施形態では、カバレッジが増加する（例えば、プレックスレベルが低下する）と、検出に所望される最小ウェーブレット長さは減少する（すなわち、分解能が増加する）。例えば、約10%の

胎児フラクションを含む試料では、4プレックスで、約1Mbまたはそれ超の N_{micr} 。が得られ、12プレックスで、約3Mbまたはそれ超の N_{micr} 。が得られる。一部の実施形態では、閾化は、レベル化の前に実施し、場合によって、閾化は、レベル化の後で実施する。

【0379】

最大エントロピーセグメンテーション処理

一部の実施形態では、適切な分解生成処理は、最大エントロピーセグメンテーション処理を含む。一部の実施形態では、最大エントロピーセグメンテーションは、分解レンダリングを決定することを含む。一部の実施形態では、最大エントロピーセグメンテーションは、染色体内異常（例えば、微小重複、微小欠失）の存在または非存在を決定することを含む。

10

【0380】

ある特定の実施形態では、最大エントロピーセグメンテーションは、ゲノムのセグメント（例えば、部分のセット、プロファイル）を再帰的にパーティショニングすることを含む。ある特定の実施形態では、最大エントロピーセグメンテーション処理は、ゲノムのセグメントを、レベル（例えば、ゲノム区分のレベル）に従ってパーティショニングする。ある特定の実施形態では、最大エントロピーセグメンテーションは、プロファイルのセグメント化部分についてのレベルを決定することを含む。一部の実施形態では、最大エントロピーセグメンテーションは、ゲノムのセグメントを、2つのセグメント（例えば、2つの部分のセット）へと分割し、2つのセグメントについてのレベルを計算する。一部の実施形態では、2つのセグメントについてのレベルを、分割（例えば、セグメント化）を施す前に、またはこれを施した後で計算する。一部の実施形態では、パーティショニング部位（例えば、セグメント化の位置、分割の位置）を、結果として得られる2つのセグメントのレベルの間の差を最大化するように選択する。一部の実施形態では、最大エントロピーセグメンテーションは、プロファイル中のあらゆる可能なパーティショニング部位（例えば、セグメント）について、仮説的セグメント化イベントの結果として得られる、2つの仮説的セグメント間のレベル差を決定し、最大のレベル差が予測される部位を選択し、次いで、プロファイルを、2つのセグメントへと分割する（例えば、パーティショニングする）。一部の実施形態では、新たに分割された2つの隣接セグメントを、その非限定的な例が、t検定、tベースの判定基準などを含む、適切な統計学的方法により、有意に異なるか、または有意に異ならないと決定する。一部の実施形態では、部分の第1のサブセットのレベルが、部分の第2のサブセットのレベルと有意に異なる場合、最大エントロピーセグメンテーションは、部分の第1のサブセットと第2のサブセットとをパーティショニングすることを含む。一部の実施形態では、部分の第1のサブセットと第2のサブセットとは、互いに隣接する。

20

30

【0381】

一部の実施形態では、新たに分割された2つの隣接セグメントを、有意に異なる決定し、セグメントの各々を、最大エントロピーセグメンテーションに従って（例えば、最大のレベル差を結果としてもたらすパーティショニング部位に従って）再度パーティショニングする。一部の実施形態では、最大エントロピーセグメンテーションは、部分のセット（例えば、プロファイル）を再帰的にパーティショニングし、これにより、2つまたはそれ超の部分のサブセットをもたらすことを含み、ここで、結果として得られるサブセットの各々は、隣接する部分のサブセットのレベルと有意に異なるレベルを含む。

40

【0382】

一部の実施形態では、最大エントロピーセグメンテーションは、1つまたは複数の個別セグメントを同定することを含む。一部の実施形態では、最大エントロピーセグメンテーションは、第2のレベルと有意に異なる第1のレベルを同定することを含む。個別セグメントは、プロファイル中のセグメントの第2のレベル（例えば、参照レベル）と有意に異なる第1のレベルを有することが多い。ある特定の実施形態では、個別セグメントを、参照レベル（例えば、ヌルレベル、ヌルプロファイル）に従って決定する。一部の実施形態

50

では、参照レベルは、全プロファイルまたはその一部のレベルである。一部の実施形態では、参照レベルは、正倍数性であるものとして既知であるか、またはコピー数の変動（例えば、微小重複または微小欠失）を欠くものとして既知である、参照プロファイルまたは参照プロファイルの部分（例えば、またはセグメント）のレベルである。一部の実施形態では、個別セグメントは、第2のレベル（例えば、参照レベル）と有意に異なる第1のレベル（例えば、ウェーブレット）を有し、第2のレベルは、参照レベルである。一部の実施形態では、最大エントロピーセグメンテーションは、同定された個別セグメントに従って、かつ／または第2のレベルと有意に異なる第1のレベルに従って、偽陰性の決定が低減され、偽陽性の決定が低減された試料について、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または非存在を決定することを含む。

10

【0383】

一部の実施形態では、最大エントロピーセグメンテーションは、セグメント化された（例えば、分割された）部分の2つのサブセットを再接合することを含む。一部の実施形態では、分割された2つのセグメントは、有意に異ならず、2つのセグメントは、再接合される。一部の実施形態では、セグメント化された部分の2つのサブセットの各々のレベルは、有意に異ならず（例えば、あらかじめ規定された閾、例えば、Zスコアおよび／または不確定性のレベル、例えば、MADに従って）、サブセットは、再接合される。一部の実施形態では、再接合されたセグメントは、再度パーティショニングしない。

【0384】

一部の実施形態では、決定分析は、2つまたはそれ超の分解レンダリングを結果としてもたらす、2つまたはそれ超のセグメント化処理を含む。ある特定の実施形態では、決定分析は、分解レンダリングを独立に生成する、2つまたはそれ超の異なるセグメント化処理（例えば、分解生成処理）を援用することを含む。一部の実施形態では、決定分析は、第1のセグメント化処理および第2のセグメント化処理を含み、第1のおよび第2のセグメント化処理は、平行して実施する。ある特定の実施形態では、第1のおよび第2のセグメント化処理は、逐次的に実施する。ある特定の実施形態では、決定分析は、分析される試料および援用されるセグメント化処理の種類に応じて実質的に同じであるかまたは実質的に異なる、分解レンダリングを独立に生成する、2つまたはそれ超の異なるセグメント化処理を含む。一部の実施形態では、第1のセグメント化処理は、ウェーブレットセグメンテーション処理（例えば、ハールウェーブレット処理）を含み、第2のセグメント化処理は、サーキュラーバイナリセグメンテーション処理を含む。

20

30

【0385】

仕上げ

一部の実施形態では、分解レンダリングを仕上げし、これにより、仕上げされた分解レンダリングをもたらす。一部の実施形態では、分解レンダリングを、2回またはそれ超にわたり仕上げする。一部の実施形態では、セグメント化処理の1つまたは複数のステップの前に、かつ／またはこれらの後で、分解レンダリングを仕上げする。一部の実施形態では、決定分析は、2つまたはそれ超のセグメント化処理を含み、各セグメント化処理は、1つまたは複数の仕上げ処理を含む。分解レンダリングは、仕上げされた分解レンダリングを指す場合もあり、仕上げされない分解レンダリングを指す場合もある。

40

【0386】

したがって、一部の実施形態では、セグメント化処理は、仕上げを含む。一部の実施形態では、仕上げ処理により、2つまたはそれ超の実質的に同様な個別セグメントまたはウェーブレット（例えば、分解レンダリング中の）を同定し、それらを、単一の個別セグメントまたはウェーブレットへと統合する（merge）（例えば、図4）。一部の実施形態では、仕上げ処理により、実質的に同様な、2つまたはそれ超の隣接セグメントまたはウェーブレットを同定し、それらを、単一のレベル、セグメント、またはウェーブレットへと統合する。したがって、一部の実施形態では、仕上げ処理は、統合処理を含む。ある特定の実施形態では、隣接する断片化された個別セグメントまたはウェーブレットを、それらのゲノム区分のレベルに従って統合する。一部の実施形態では、2つまたはそれ超の

50

隣接する個別セグメントまたはウェーブレットの統合は、最終的に統合する、2つまたはそれ超の隣接する個別セグメントまたはウェーブレットについての中央値レベルの計算を含む。一部の実施形態では、実質的に同様な、2つまたはそれ超の隣接する個別セグメントまたはウェーブレットを統合し、これにより、仕上げする結果として、単一のセグメント、ウェーブレット、またはレベルをもたらす。ある特定の実施形態では、2つまたはそれ超の隣接する個別セグメントまたはウェーブレットを、WillenbrockおよびFridly (Willenbrock H、Fridlyand J、A comparison study: applying segmentation to array CGH data for downstream analyses、Bioinformatics (2005年11月15日)、21巻(22号):4084~91頁)により記載されている処理により統合する。一部の実施形態では、2つまたはそれ超の隣接する個別セグメントまたはウェーブレットを、GLADとして公知であり、Hupe、P.ら(2004年)、「Analysis of array CGH data:from signal ratio to gain and loss of DNA regions」、Bioinformatics、20巻、3413~3422頁において記載されている処理により統合する。

10

【0387】

候補セグメントまたはウェーブレット事象の同定

一部の実施形態では、決定分析は、分解レンダリング中の候補セグメントまたはウェーブレット事象を同定することを含む。候補セグメントを、分解レンダリング内で最も有意な個別セグメントであると決定し、ウェーブレット事象を、ウェーブレット分解レンダリング内で同定された、最も有意なウェーブレットであると決定する。「候補セグメント」はまた、任意の種類のセグメント化処理および分解レンダリングを使用して、セグメント化から結果として得られる、分解レンダリング内で最も有意な個別セグメントでもある。ウェーブレットセグメンテーション処理を活用する場合、候補セグメントは、「ウェーブレット事象」と同義である。候補セグメントは一般に、分解レンダリング中の最も有意な個別セグメントであり、場合によって、セグメントによりカバーされる部分(例えば、ビン)の数の点で、かつ/またはセグメントについての正規化されたカウントのレベルの絶対値の点で最も有意である。候補セグメントは、場合によって、大きく、場合によって、分解レンダリング中の他の個別セグメントより実質的に大きい。一部の実施形態では、分解レンダリング中の1つだけの候補セグメントを同定する。一部の実施形態では、分解レンダリング中の1つまたは複数の個別セグメントを同定し、1つまたは複数の個別セグメントのうちの1つを、候補セグメントとして同定する。一部の実施形態では、候補セグメントは、第2の個別セグメントのレベルより実質的に大きなレベルを有する第1の個別セグメントであり、ここで、第1の個別レベルは、分解レンダリング内で最大のレベルである。適切な方法により候補セグメントを同定することができる。一部の実施形態では、候補セグメントを、曲線下面積(AUC: area under the curve)分析により同定する。一部の実施形態では、決定分析は、AUC分析を含む。第1の個別セグメントが、分解レンダリング中の別の個別セグメントについての場合より、実質的に大きなレベルを有し、かつ/または実質的に大きな多数の部分のカバーする、ある特定の実施形態では、第1のセグメントは、大きなAUCを含む。レベルをAUCについて分析する場合、レベルの絶対値(例えば、正規化されたカウントに対応するレベルは、欠失では負の値をとり、重複では正の値をとりうる)を活用することが多い。ある特定の実施形態では、AUCを、計算されたAUCの絶対値(例えば、結果として得られる正の値)として決定する。ある特定の実施形態では、候補セグメントは、同定されたら(例えば、AUC分析により、または適切な方法により)、かつ、任意選択で、検証された後で、zスコアの計算などについて選択して、候補セグメントが、遺伝子の変動(例えば、異数性、微小欠失、または微小重複)を表示するのかどうかについて決定する。

20

30

40

【0388】

対数オッズ比分析

場合によって、試料についての比較における使用のために、かつ/または決定(例えば、遺伝子の変動の存在または非存在の決定)における使用のために、オッズ比または対数オッズ比(LOR: log odds ratio)を計算する。LORは、場合によ

50

て、(A)と(B)との商の対数として計算し、ここで、(A)は、(1)遺伝子の変動を有する条件付き確率と、(2)遺伝子の変動を有する事前確率との第1の乗算の積であり、(B)は、(1)遺伝子の変動を有さない条件付き確率と、(2)遺伝子の変動を有さない事前確率との第2の乗算の積である。遺伝子の変動は、場合によって、染色体異数性(例えば、全染色体の1つ、3つ、4つのコピー)、微小欠失、または微小挿入である。

【0389】

LORの計算は、場合によって、試験試料について決定された胎児フラクションの適用を含み、場合によって、試験試料について同定された染色体または候補セグメントについてのカウント表示の適用を含む。一部の実施形態では、染色体異数性を有する条件付き確率を、胎児フラクションおよびカウント表示に従って決定する。したがって、一部の実施形態では、ある特定の方法は、参照ゲノムの部分へとマッピングした核酸配列の読取りのカウントに従って、染色体のカウント表示および/または候補セグメントのカウント表示を決定するステップを含み、配列の読取りは、胎児を出産する妊娠中の雌による試験試料についての、循環無細胞核酸についての読取りであることが多い。候補セグメントは、場合によって、検証された候補セグメント(本明細書で記載される)である。

【0390】

染色体のカウント表示は、場合によって、染色体中の部分(例えば、ビン)へとマッピングしたカウントであって、ゲノムまたはそのサブセットの部分であり、染色体(例えば、全ての常染色体)より大きい部分中のカウントで除算して得たカウントである。染色体のカウント表示は、場合によって、定量し、任意の適切な定量(例えば、zスコア)を活用することができる。zスコアにより染色体のカウント表示を定量する実施形態では、zスコアは、場合によって、差(A)を値(B)で除算して得た商である。差(A)は、場合によって、(i)試験試料の染色体のカウント表示から、(ii)正倍数性染色体のカウント表示の中央値を減算して得た差である。値(B)は、場合によって、正倍数性染色体のカウント表示のMADである。試験試料の染色体のカウント表示は、場合によって、試験試料についての、(a)染色体中の部分中のカウントの、(b)常染色体中の部分中のカウントに対する比である。正倍数性染色体のカウント表示の中央値は、場合によって、正倍数性についての、(a)染色体中の部分中のカウントの、(b)常染色体中の部分中のカウントに対する比の中央値である。カウントは、場合によって、正規化されたカウントであり、これにより、ゲノム部分へとマッピングしたカウントを、1つまたは複数の適切な正規化処理により正規化することができる。活用される正規化処理の非限定的な例は、当技術分野で公知であり、本明細書でも記載されている(例えば、LOESS、GC-LOESS、PERUN、ChAI、主成分正規化処理)。

【0391】

候補セグメントのカウント表示は、場合によって、候補セグメント中の部分(例えば、ビン)または候補セグメントによりカバーされた部分(例えば、ビン)へとマッピングしたカウントであって、ゲノムまたはそのサブセットの部分であり、候補セグメントより大きい部分(例えば、全ての常染色体)中のカウントで除算して得たカウントである。候補セグメントのカウント表示は、場合によって、定量し、任意の適切な定量(例えば、zスコア)を活用することができる。zスコアにより候補セグメントのカウント表示を定量する実施形態では、zスコアは、場合によって、差(A)を値(B)で除算して得た商である。差(A)は、場合によって、(i)試験試料の候補セグメントのカウント表示から、(ii)正倍数性の候補セグメントのカウント表示の中央値を減算して得た差である。値(B)は、場合によって、正倍数性の候補セグメントのカウント表示のMADである。試験試料の候補セグメントのカウント表示は、場合によって、試験試料についての、(a)候補セグメント中の部分中のカウントの、(b)常染色体中の部分中のカウントに対する比である。正倍数性の候補セグメントのカウント表示の中央値は、場合によって、正倍数性についての、(a)候補セグメント中の部分中のカウントの、(b)常染色体中の部分中のカウントに対する比の中央値である。カウントは、場合によって、正規化されたカウ

ントであり、これにより、ゲノム部分へとマッピングしたカウントを、1つまたは複数の適切な正規化処理により正規化することができる。活用される正規化処理の非限定的な例は、当技術分野で公知であり、本明細書でも記載されている（例えば、LOESS、GC-LOESS、PERUN、ChAI、主成分正規化処理）。

【0392】

LORの計算を有する方法は、場合によって、試験試料についての胎児フラクションを決定することを含む。胎児フラクションは、その非限定的な例が本明細書で記載される（例えば、Y染色体遺伝子座（例えば、SRY遺伝子座）の定量、FRSの定量）、当技術分野で公知の任意の適切な方法を使用して決定することができる。

【0393】

LORの計算についてのある特定の実施形態では、遺伝子の変動を有する条件付き確率は、試験試料について決定された胎児フラクション、試験試料についての、染色体のカウント表示、または候補セグメントのカウント表示についてのzスコア、および染色体のカウント表示、または候補セグメントのカウント表示についての、zスコアの胎児フラクションに特異的な分布に従って評価する。一部の実施形態では、遺伝子の変動を有する条件付き確率を、本明細書の下記の実施例6で示される等式23における関係により決定し、ここで、fは、胎児フラクションであり、Xは、染色体または候補セグメントについての部分の合計であり、 $X \sim f(\mu X, \sigma X)$ [式中、 μX および σX は、それぞれ、Xの平均値および標準偏差であり、 $f(\mu X, \sigma X)$ は、分布関数である]である。遺伝子の変動を有する条件付き確率は、場合によって、試験試料についての、染色体のカウント表示、または候補セグメントのカウント表示についてのzスコアと、染色体のカウント表示、または候補セグメントのカウント表示についての、zスコアの胎児フラクションに特異的な分布との交差部分である（例えば、T21についての例についての図32を参照されたい）。実施例6は、図32に言及しながら、微小重複イベントまたは微小欠失イベントの存在または非存在を決定する状況における正倍数体分布に言及する分布シフトについて記載する。

【0394】

染色体異数性を有さない条件付き確率は、場合によって、染色体のカウント表示、または候補セグメントのカウント表示、および正倍数性についてのカウント表示に従って決定する。遺伝子の変動を有さない条件付き確率は、場合によって、染色体のカウント表示のzスコアと、正倍数体中の染色体のカウント表示についてのzスコアの分布との交差部分である（例えば、T21についての例についての図32を参照されたい）。

【0395】

遺伝子の変動を有する事前確率および遺伝子の変動を有さない事前確率は、例えば、1つまたは複数の患者集団について、当技術分野で公知の統計学的データを使用して決定することが多い。T21発生の確率およびT21が発生しない確率は、例えば、特定の地域中の集団についてたやすく決定することができる。事前確率は、試験被験体を含まない複数の試料から決定することが多い。

【0396】

比較および決定分析

一部の実施形態では、決定分析は、比較を含む。一部の実施形態では、比較は、少なくとも2つの分解レンダリングの比較を含む。一部の実施形態では、比較は、少なくとも2つの候補セグメントの比較を含む。ある特定の実施形態では、少なくとも2つの候補セグメントの各々は、異なる分解レンダリングに由来する。例えば、第1の候補セグメントは、第1の分解レンダリングに由来することが可能であり、第2の候補セグメントは、第2の分解レンダリングに由来しうる。一部の実施形態では、比較は、2つの分解レンダリングが実質的に同じであるのか、異なるのかを決定することを含む。一部の実施形態では、比較は、2つの候補セグメントが実質的に同じであるのか、異なるのかを決定することを含む。

【0397】

一部の実施形態では、各レンダリングが、候補セグメントを含み、各分解レンダリングに由来する候補セグメントが、実質的に同じであると決定される場合、2つの分解レンダリングは、実質的に同じである。2つの候補セグメントは、その非限定的な例が、目視、2つの候補セグメントのレベルもしくはZスコアの比較、2つの候補セグメントのエッジの比較、2つの候補セグメントもしくはそれらの対応する分解レンダリングの重合せなど、またはこれらの組合せを含む、適切な比較方法により、実質的に同じであるかまたは異なると決定することができる。一部の実施形態では、2つの候補セグメントのエッジは、実質的に同じであり、2つの候補セグメントは、実質的に同じである。ある特定の実施形態では、候補セグメントのエッジは、別の候補セグメントのエッジと実質的に同じであり、2つのエッジは、10未満、9未満、8未満、7未満、6未満、5未満、4未満、3未満、2未満、または1未満の部分（例えば、ピン）で隔てられている。一部の実施形態では、2つのエッジは、実質的に同じであり、同じ位置（例えば、同じ部分）にある。一部の実施形態では、実質的に同じである2つの候補セグメントは、実質的に同じ（例えば、不確定性のレベル中の、例えば、不確定性のレベルの約3、2、1倍、またはそれ未満の）レベル、Zスコアなどを含む。一部の実施形態では、2つの候補セグメントは、実質的に異なるエッジおよび／または実質的に異なるレベルを含み、比較に従って、実質的に同じでない（例えば、異なる）と決定される。

10

【0398】

ある特定の実施形態では、比較は、1つまたは複数の複合候補セグメントを生成し、または1つもしくは複数の複合候補セグメントの比較を含む比較に基づき（例えば、これに一部分または単に基づき）、異数性、微小欠失、または微小重複の存在または非存在の決定を下すことを含む。複合候補セグメントは、任意の適切な方法により生成することができる。一部の実施形態では、複合候補セグメントは、2つまたはそれ超の候補セグメント（例えば、レベル、AUC、および／またはエッジ）を平均することにより生成する。一部の実施形態では、複合候補セグメントは、2つまたはそれ超の候補セグメントの重合せにより生成する。一部の実施形態では、2つまたはそれ超の候補セグメントは、実質的に同じであり、複合候補セグメントを生成する（例えば、図11）。

20

【0399】

比較は、場合によって、本明細書の下記で記載される通り、2つまたはそれ超の分解レンダリングから導出された候補セグメント（例えば、ウェーブレット事象）を定量し、比較を活用して、試料中の遺伝子の変動（例えば、染色体異数性、微小重複、または微小欠失）の存在または非存在を決定することを含む。

30

【0400】

ある特定の実施形態では、比較は、2つまたはそれ超の分解レンダリング内で同定される候補セグメント（例えば、ウェーブレット事象）に由来する、複合候補セグメント（例えば、複合ウェーブレット事象）の存在または非存在を決定することを含む。一部の実施形態では、2つまたはそれ超の候補セグメント（例えば、2つまたはそれ超の分解レンダリングから導出される、例えば、ウェーブレット事象）は、重複するか、または実質的に同じであり、複合候補セグメント（例えば、複合ウェーブレット事象）の存在が決定される（図11）。複合ウェーブレット事象の存在または非存在は、任意の適切な方法により決定することができる。一部の実施形態では、複合候補セグメント（例えば、複合ウェーブレット事象）の存在または非存在を、2つまたはそれ超の候補セグメント（例えば、複合ウェーブレット事象、例えば、レベル、AUC、および／またはエッジ）を平均することにより決定する。一部の実施形態では、複合候補セグメント（例えば、複合ウェーブレット事象）の存在または非存在を、2つまたはそれ超の候補セグメント（例えば、ウェーブレット事象）の重合せにより決定する。ある特定の実施形態では、2つまたはそれ超の候補セグメント（例えば、ウェーブレット事象）が、重複するか、または実質的に同じである場合に、複合候補セグメント（例えば、複合ウェーブレット事象）の存在を決定する。

40

【0401】

50

一部の実施形態では、2つまたはそれ超の候補セグメント（例えば、複合ウェーブレット事象、例えば、2つまたはそれ超の分解レンダリングから導出された）は、重複しないか、または異なり（例えば、実質的に異なり）、複合候補セグメントの非存在（例えば、複合ウェーブレット事象の非存在）が決定される。一部の実施形態では、複合候補セグメント（例えば、複合ウェーブレット事象）の非存在は、染色体異数性、微小重複、または微小欠失の非存在を指し示す。

【0402】

一部の実施形態では、決定分析は、アウトカムを決定すること（*determining*）（例えば、胎児における、例えば、遺伝子の変動の存在または非存在を決定すること（*determining*））を含む。一部の実施形態では、決定分析は、染色体異数性、微小重複、または微小欠失の存在または非存在を決定する方法を含む。一部の実施形態では、決定分析は、遺伝子の変動の存在または非存在（例えば、胎児における）を決定する方法であって、本明細書で記載される決定分析を使用しない（例えば、1もしくは複数の候補セグメントの存在もしくは非存在をセグメント化せず、同定せずに、かつ／または1もしくは複数の候補セグメントを定量しない）遺伝子の変動の存在または非存在の決定と比較して、偽陰性の決定が低減され、偽陽性の決定が低減された方法を含む。一部の実施形態では、決定分析は、一連の方法または方法ステップを含む。決定分析の非限定的な例は、図6～8に示され、本明細書で記載される。ある特定の実施形態では、決定分析は、カウントを得ることと、プロファイルを生成し、かつ／またはプロファイルを得ることを含む。一部の実施形態では、決定分析は、プロファイルのセグメント化および分解レンダリングを生成することを含む。一部の実施形態では、分解レンダリングまたはそのセグメント（例えば、染色体、レベル、個別セグメントまたはウェーブレット、候補セグメントまたはウェーブレット事象、複合セグメントまたは複合ウェーブレットを表示するセグメント）を、適切な方法により定量する。適切な定量法の非限定的な例は、当技術分野で公知であり、本明細書でも一部分記載され、例えば、Zスコア、p値、t値、レベル、AUC、倍数性、不確定性のレベルなど、またはこれらの組合せを決定する方法を含む。

【0403】

一部の実施形態では、決定分析は、2つまたはそれ超のセグメント化法によりプロファイルをセグメント化することを含む。一部の実施形態では、決定分析は、50またはそれ超のセグメント化法を含む。ある特定の実施形態では、決定分析は、50もしくはそれ未満、40もしくはそれ未満、30もしくはそれ未満、20もしくはそれ未満、10もしくはそれ未満、または約5もしくはそれ未満のセグメント化法を含む。ある特定の実施形態では、決定分析は、約10、9つ、8つ、7つ、6つ、5つ、4つ、3つ、または2つのセグメント化法を含む。一部の実施形態では、セグメント化の各方法（例えば、2つの方法を活用する場合である、例えば、図6A、611および612）により、プロファイルの分解レンダリングを提示する。一部の実施形態では、セグメント化の2つまたはそれ超の方法によりもたらされる分解レンダリングは、同じであるか、実質的に同じであるか、または異なる。

【0404】

一部の実施形態では、セグメント化の後で、仕上げ（例えば、図6A、621および622；図6B、623）を施す。一部の実施形態では、1つまたは複数のセグメント化処理の適用から導出される1つまたは複数の分解レンダリングは、場合によって、同じ仕上げ法により仕上げする。一部の実施形態では、1つまたは複数のセグメント化ステップから導出される1つまたは複数の分解レンダリングを、異なる仕上げ法により仕上げする。一部の実施形態では、分解レンダリングを、1つ、2つ、3つ、またはそれ超の仕上げ法により仕上げする。一部の実施形態では、各分解レンダリングを、1つの方法により仕上げし、方法は、各分解レンダリングについて同じである。

【0405】

一部の実施形態では、候補セグメント（例えば、ウェーブレット事象）の存在または非存在を、セグメント化の後で、かつ、任意選択で、仕上げの後で同定する（例えば、図6

10

20

30

40

50

A、631および632；図6B、623）。一部の実施形態では、仕上げ処理を省き、候補セグメント（例えば、ウェーブレット事象）を、セグメント化から導出された分解レンダリングから直接同定する。一部の実施形態では、候補セグメント（例えば、ウェーブレット事象）を、仕上げされた分解レンダリング内で、かつ／または仕上げされた分解レンダリングから同定する。一部の実施形態では、候補セグメント（例えば、ウェーブレット事象）は、1つまたは複数の分解レンダリング内で同定されず、遺伝子の変動の非存在が決定される。候補セグメント（例えば、ウェーブレット事象）が、1つまたは複数の分解レンダリング（例えば、仕上げされた分解レンダリング）のうちの1つにおいて同定されない、一部の実施形態では、決定分析を終結させる。

【0406】

一部の実施形態では、候補セグメント（例えば、ウェーブレット事象）を同定したら、定量する（例えば、図6A、641および642；図6B、644（例えば、ZスコアまたはLORによる定量））。候補セグメント（例えば、ウェーブレット事象）は、その非限定的な例が、Zスコアの計算、p値の計算、t値の決定、レベルの決定、倍数性の決定、計算された不確定性のレベルなど、またはこれらの組合せを含む、適切な方法により定量することができる。

【0407】

一部の実施形態では、決定分析は、比較を含む（例えば、図6A、6B、および8における650、651、810）。一部の実施形態では、比較は、定量の後で行う（例えば、図6A、641、642および643；図6B、651）。一部の実施形態では、比較は、ウェーブレットまたは候補セグメントの同定の後で行う（例えば、図6A、631および632；図6B、633）。場合によって、比較は、染色体の定量の後で行う（例えば、図6A、643；図6B、645（例えば、ZスコアまたはLORによる定量））。一部の実施形態では、決定を下すことは、比較の後で行う（例えば、図6A、660；図6B、661）。

【0408】

ある特定の実施形態では、検証された候補セグメントを含む候補セグメント（併せて、「候補セグメント」と称する）を定量する。本明細書で記載される通り、候補セグメントは、場合によって、候補セグメントのカウント表示として定量し、候補セグメントのカウント表示は、場合によって、Zスコアにより定量する。場合によって、候補セグメントが位置する染色体について、染色体のカウント表示を生成し、定量する。染色体のカウント表示については、本明細書で記載されており、Zスコアにより定量することができ、これについてもまた、本明細書で記載されている。本明細書で記載される通り、候補セグメントのカウント表示および／または染色体のカウント表示のためのカウントは、場合によって、正規化されたカウントである。

【0409】

ある特定の実施形態では、第1の候補セグメントのカウント表示のZスコアによる定量を生成し、第2の候補セグメントのカウント表示のZスコアによる定量を生成し、ここで、第1の候補セグメントと第2の候補セグメントとは、2つの異なる種類のセグメント化から同定する。一部の実施形態は、(i) 1未満（例えば、約0.6～約0.8）の係数を乗算した、第1の候補セグメントのカウント表示のZスコアによる定量、および(ii) 係数を乗算した、第2の候補セグメントのカウント表示のZスコアによる定量のうちの最小値を決定することを含む。

【0410】

一部の実施形態では、候補セグメントが位置する染色体について、候補セグメントのカウント表示の定量を、染色体のカウント表示の定量と比較する。ある特定の実施形態は、染色体表示のZスコアによる定量が、前出の段落で言及された最小値未満、それ超、またはそれに等しいのいずれであるかを決定することを含む。一部の実施形態は、染色体のカウント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））未満、それ超、またはそれに等しいのいずれであるかを決定す

10

20

30

40

50

ることを含む。

【0411】

ある特定の実施形態は、試験試料について、(i) 染色体のカウント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））超またはそれに等しく、(ii) 染色体のカウント表示のZスコアによる定量が、前出の段落で言及された最小値超またはそれに等しい場合に、染色体異数性の存在を決定することを含む。一部の実施形態は、試験試料について、(i) 染色体のカウント表示のZスコアによる定量は、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））未満であり、かつ／または(ii) 染色体のカウント表示のZスコアによる定量は、最小値未満である場合に、染色体異数性の非存在を決定することを含む。染色体異数性は、場合によって、トリソミーまたはモノソミーであり、場合によって、1つ、3つ、または4つの染色体の発生である。

10

【0412】

一部の実施形態は、第1の候補セグメントのカウント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））未満、それ超、またはそれに等しいのいずれであるかを決定し、第2の候補セグメントのカウント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））未満、それ超、またはそれに等しいのいずれであるかを決定することを含む。ある特定の実施形態は、第1の候補セグメントおよび第2の候補セグメントが、実質的に同じであるかまたは重複するかどうかを決定することを含む。

20

【0413】

一部の実施形態は、試験試料について、(i) 第1の候補セグメントのカウント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））超またはそれに等しく、第2の候補セグメントのカウント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））超またはそれに等しく、(ii) 第1の候補セグメントおよび第2の候補セグメントが、実質的に同じであるかまたは重複する場合に、微小欠失または微小挿入の存在を決定することを含む。ある特定の実施形態は、試験試料について、(i) 第1の候補セグメントのカウント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））未満であり、かつ／もしくは第2の候補セグメントのカウ
ント表示のZスコアによる定量が、閾のZスコア値（例えば、約3.95の値（例えば、約3.5～約4.5））未満であり、かつ／または(ii) 第1の候補セグメントおよび第2の候補セグメントが、実質的に同じでないかもしくは重複しない場合に、微小欠失または微小挿入の非存在を決定することを含む。

30

【0414】

一部の実施形態では、比較では、2つまたはそれ超の値（例えば、定量、例えば、プロファイルの定量および／または候補セグメント（例えば、ウェーブレット事象）の定量から導出された値）を比較する。一部の実施形態では、比較では、候補セグメント（例えば、ウェーブレット事象）またはプロファイルの定量を、所定の値または閾と比較する。比較の非限定的な例を、図7に示す。一部の実施形態では、比較は、Zスコアの比較を含む。ある特定の実施形態では、比較は、全染色体（染色体プロファイル）についての全染色体表示についてのZスコアの絶対値（すなわち、 $|Z_{chr}|$ ）の比較を含む。値 $|Z_{chr}|$ は、場合によって、所定の値、閾、または比較特徴（例えば、図7、710の閾3.95）と比較される。一部の実施形態では、Zスコアを比較するために使用される閾、所定の値、または比較特徴は、約2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.75、3.8、3.85、3.9、3.95、4.0、4.05、4.1、4.15、4.2、4.3、4.4、または約4.5である。値 $|Z_{chr}|$ は、場合によって、分解レンダリングによる候補セグメントおよびその部分のカウント表示についてのZスコアの絶対値（例えば、実施例3の $|Z_{wave}|$ および $|Z_{cb}|$ ならびに図7の $|Z_{A4}|$ および $|Z_{B4}|$ ）と比較される。

40

50

【0415】

一部の実施形態では、比較の結果は、別の比較についての決定 (decision) またはアウトカムの決定 (decision) である。一部の実施形態では、第1の比較の結果 (例えば、図7、710) は、一連の比較における次の比較を決定する (determine) 決定 (decision) である。例えば、第1の比較 (例えば、図7、710) により、 $|Z_{chr}|$ が、所定の値超またはそれに等しいことを決定する (determine) ことができ、第2の比較 (例えば、図7、721) により、 $|Z_{chr}|$ を、 $|Z_{A4}|$ および/または $|Z_{B4}|$ と比較する。代替的に、第1の比較 (例えば、図7、710) により、 $|Z_{chr}|$ が、所定の値未満であることを決定することができ、第2の比較 (例えば、図7、722) により、決定 (decision) 分析 (例えば、図6A、631および632) で既に同定された候補セグメント (例えば、ウェーブレット事象) が、実質的に同じであるのか、異なるのかを決定する。

10

【0416】

一部の実施形態では、第1の比較の結果 (例えば、図7、710) は、第2の一連の比較を決定する (determine) 決定 (decision) であり、第2の比較から導出された決定 (decision) により、第3の比較などを決定する (determine)。一部の実施形態では、第1の比較により、 $|Z_{chr}|$ が、所定の値超またはそれに等しいことを決定することができ、第2の比較 (例えば、図7、721) により、 $|Z_{chr}|$ が、 $|Z_{A4}|$ および/もしくは $|Z_{B4}|$ またはそれらの分数 (例えば、所定の値 α を乗算された $|Z_{A4}|$ および/または $|Z_{B4}|$) 超であることを決定することができ、全染色体異数性の存在が決定される。トリソミーおよびモノソミーは、適切な方法で識別することができる。

20

【0417】

一部の実施形態では、第1の比較により、 $|Z_{chr}|$ が、所定の値超またはそれに等しいことを決定することができ、第2の比較 (例えば、図7、721) により、 $|Z_{chr}|$ が、 $|Z_{A4}|$ および/もしくは $|Z_{B4}|$ またはそれらの分数 (例えば、所定の値 α を乗算された $|Z_{A4}|$ および/または $|Z_{B4}|$) 未満であることを決定することができ、第3の比較が実施される。ある特定の実施形態では、第1の比較により、 $|Z_{chr}|$ が、所定の値未満であることを決定することができ、第2の比較により、同定された候補セグメント (例えば、ウェーブレット事象) が、重複するかまたは実質的に同じ (複合候補セグメント) であることを決定し、第3の比較により、 $|Z_{A4}|$ および $|Z_{B4}|$ が、所定の値 (例えば、3.95) 超またはそれに等しいことを決定し、微小重複および/または微小欠失の存在を決定する。微小重複および微小欠失は、適切な方法で識別することができる。例えば、微小重複は、正のZスコアを有する可能性があり、微小欠失は、負のZスコアを有する可能性がある。

30

【0418】

一部の実施形態では、比較により、2つまたはそれ超の候補セグメント (例えば、ウェーブレット事象) が、重複するかまたは実質的に同じではなく (例えば、それらは、実質的に異なる; 例えば、図8、822)、遺伝子の変動がプロファイル中に存在しないことを決定することができる。一部の実施形態では、比較により、2つまたはそれ超の候補セグメント (例えば、ウェーブレット事象、例えば、1つまたは複数の分解レンダリング内で同定された全候補セグメント (例えば、ウェーブレット事象)) が、重複または実質的に同じである (例えば、図8、821) ことを決定することができ、微小重複または微小欠失の存在または非存在が決定される。一部の実施形態では、微小重複または微小欠失の存在または非存在を、複合候補セグメント (例えば、複合ウェーブレット事象) の定量に従って決定する。

40

【0419】

一部の実施形態では、決定分析は、候補セグメント (例えば、ウェーブレット事象) のセグメント化、仕上げ、および同定のうちの2つまたはそれ超を含む。一部の実施形態では、決定分析は、2つまたはそれ超の候補セグメント (例えば、ウェーブレット事象) の

50

定量を含みうる。一部の実施形態では、決定分析は、染色体プロファイルの定量を含みうる。一部の実施形態では、決定分析は、1つまたは複数の比較を含む。一部の実施形態では、決定 (decision) 分析は、遺伝子の変動の存在または非存在の決定 (determination) を含む。

【0420】

一部の実施形態では、決定 (decision) 分析は、候補セグメント (例えば、ウェーブレット事象) のセグメント化、仕上げ、同定、1つまたは複数の比較、および遺伝子の変動の存在または非存在の決定 (determination) を含み、かつ／またはこれからなる。一部の実施形態では、決定 (decision) 分析は、候補セグメント (例えば、ウェーブレット事象) のセグメント化、仕上げ、同定、定量、1つまたは複数の比較、および遺伝子の変動の存在または非存在の決定 (determination) を含み、かつ／またはこれからなる。一部の実施形態では、決定 (decision) 分析は、候補セグメント (例えば、ウェーブレット事象) のセグメント化、仕上げ、同定、複合候補セグメント (例えば、複合ウェーブレット事象) の存在または非存在の決定 (determination)、複合候補セグメント (例えば、複合ウェーブレット事象) の定量、1つまたは複数の比較、および遺伝子の変動の存在または非存在の決定 (determination) を含み、かつ／またはこれからなる。一部の実施形態では、決定 (decision) 分析は、候補セグメント (例えば、ウェーブレット事象) のセグメント化、仕上げ、同定、候補セグメント (例えば、ウェーブレット事象) の定量、染色体プロファイルの定量、比較および遺伝子の変動の存在または非存在の決定 (determination) を含み、かつ／またはこれからなる。一部の実施形態では、決定分析は、検証を含む。

【0421】

一部の実施形態では、比較または決定分析は、オッズ比または対数オッズ比 (LOR) との比較を含む。ある特定の実施形態では、比較または決定は、計算された LOR が、ゼロ超またはゼロ未満のいずれであることを決定することを含む。

【0422】

一部の実施形態では、比較または決定 (decision) は、染色体のカウント表示の z スコアによる定量を生成することと、染色体のカウント表示が、値 (例えば、約 3.95 の z スコア値 (例えば、約 3.5 ~ 約 4.5)) 未満、それ超、またはそれに等しいのいずれであることを決定すること (determining) とを含む。ある特定の実施形態では、決定 (decision) は、試験試料について、(i) 染色体のカウント表示の z スコアによる定量が、値 (例えば、約 3.95 の) 超またはそれに等しく、(ii) LOR が、ゼロ超である場合に、染色体異数性の存在を決定すること (deciding (determining)) を含む。一部の実施形態では、決定 (decision) は、試験試料について、(i) 染色体のカウント表示の z スコアによる定量が、値 (例えば、約 3.95 の) 未満であり、かつ／または (ii) LOR が、ゼロ未満である場合に、染色体異数性の非存在を決定すること (deciding (determining)) を含む。染色体異数性は、場合によって、トリソミーまたはモノソミーであり、場合によって、1つ、3つ、または4つの染色体のコピーである。

【0423】

一部の実施形態では、比較または決定 (decision) は、候補セグメントのカウント表示の z スコアによる定量を生成することと、候補セグメントのカウント表示が、値 (例えば、約 3.95 の z スコア値 (例えば、約 3.5 ~ 約 4.5)) 未満、それ超、またはそれに等しいのいずれであることを決定すること (determining) とを含む。ある特定の実施形態では、決定 (decision) は、試験試料について、(i) 候補セグメントのカウント表示の z スコアによる定量が、値 (例えば、約 3.95 の) 超またはそれに等しく、(ii) LOR が、ゼロ超である場合に、微小欠失または微小挿入の存在イベントを決定すること (deciding (determining)) を含む。一部の実施形態では、決定 (decision) は、試験試料について、(i) 候補セグ

メントのカウント表示の z スコアによる定量が、値（例えば、約 3.95 の）未満であり、かつ／または (i i) L O R が、ゼロ未満である場合に、微小欠失または微小挿入の非存在イベントを決定すること (d e c i d i n g (d e t e r m i n i n g)) を含む。微小欠失イベントは、場合によって、ディジョージ症候群と関連するイベントである。

【0424】

アウトカム

本明細書で記載される方法により、試料についての、遺伝子の変動の存在または非存在の決定（例えば、胎児の異数性）をもたらすことができ、これにより、アウトカムを提示する（例えば、これにより、遺伝子の変動（例えば、胎児の異数性）の存在または非存在の決定因であるアウトカムを提示する）ことができる。遺伝子の変動は、遺伝子情報（例えば、染色体、染色体のセグメント、多型領域、転座領域、ヌクレオチド配列の変化など、または前出の組合せ）の獲得、喪失、および／または変化（例えば、重複、欠失、融合、挿入、変異、再構成、置換、または異常なメチル化）であって、参照に対する、試験被験体のゲノム情報または遺伝子情報の検出可能な変化を結果としてもたらす、遺伝子情報の獲得、喪失、および／または変化を含むことが多い。遺伝子の変動の存在または非存在は、部分へとマッピングした配列の読取り（例えば、カウント、参照ゲノムの、ゲノムの部分のカウント）を変換、分析、および／または操作することにより決定することができる。一部の実施形態では、アウトカムを決定することは、妊娠中の雌に由来する核酸を分析することを含む。ある特定の実施形態では、アウトカムを、妊娠中の雌から得られたカウント（例えば、正規化されたカウント）であって、妊娠中の雌から得られた核酸からのカウントに従って決定する。

【0425】

本明細書で記載される方法は、場合によって、胎児を出産する妊娠中の雌からの試験試料について、胎児の異数性の存在または非存在（例えば、完全な染色体異数性、部分的な染色体異数性、または部分的染色体異常（例えば、モザイク現象、欠失、および／または挿入））を決定する。ある特定の実施形態では、本明細書で記載される方法により、胎児を出産する妊娠中の雌からの試料について、正倍数性または正倍数性の欠如（非正倍数性）を検出する。本明細書で記載される方法では、場合によって、1つもしくは複数の染色体（例えば、第13染色体、第18染色体、第21染色体またはこれらの組合せ）またはそのセグメントについて、トリソミーを検出する。

【0426】

一部の実施形態では、遺伝子の変動（例えば、胎児の異数性）の存在または非存在を、本明細書で記載される方法、当技術分野で公知の方法、またはこれらの組合せにより決定する。遺伝子の変動の存在または非存在は一般に、参照ゲノムの部分へとマッピングした配列の読取りのカウントから決定する。遺伝子の変動の存在または非存在を決定するのに活用される配列の読取りのカウントは、場合によって、未加工のカウントおよび／またはフィルタリングされたカウントであり、正規化されたカウントであることが多い。1つまたは複数の適切な正規化処理を使用して、その非限定的な例が、部分に関する正規化、GC含有量による正規化、線形最小二乗回帰および非線形最小二乗回帰、L O E S S、G C L O E S S、L O W E S S、P E R U N、C h A I、R M、G C R M、およびこれらの組合せを含む、正規化されたカウントを生成することができる。正規化されたカウントは、場合によって、特定のセットまたは部分のセットについての1つまたは複数のレベルまたはプロファイル中のレベルとして表される。正規化されたカウントは、場合によって、遺伝子の変動の存在または非存在を決定する前に、調整または穴埋めされる。

【0427】

一部の実施形態では、アウトカムを、1つまたは複数のレベルに従って決定する。一部の実施形態では、遺伝子の変動（例えば、染色体異数性）の存在または非存在の決定を、1つまたは複数の調整されたレベルに従って決定する。一部の実施形態では、遺伝子の変動（例えば、染色体異数性）の存在または非存在の決定を、1つ〜約10,000の調整されたレベルを含むプロファイルに従って決定する。遺伝子の変動（例えば、染色体異数

性)の存在または非存在の決定は、約1つ～約1000、1つ～約900、1つ～約800、1つ～約700、1つ～約600、1つ～約500、1つ～約400、1つ～約300、1つ～約200、1つ～約100、1つ～約50、1つ～約25、1つ～約20、1つ～約15、1つ～約10、または1つ～約5つの調整を含むプロファイルに従って決定することが多い。一部の実施形態では、遺伝子の変動(例えば、染色体異数性)の存在または非存在の決定を、約1つの調整(例えば、1つの調整されたレベル)を含むプロファイルに従って決定する。一部の実施形態では、アウトカムを、1つもしくは複数、2つもしくはそれ超、3つもしくはそれ超、5つもしくはそれ超、6つもしくはそれ超、7つもしくはそれ超、8つもしくはそれ超、9つもしくはそれ超、または、場合によって、10もしくはそれ超の調整を含む、1つまたは複数のプロファイル(例えば、染色体またはそのセグメントのプロファイル)に従って決定する。一部の実施形態では、遺伝子の変動(例えば、染色体異数性)の存在または非存在の決定を、プロファイルに従って決定するが、ここで、プロファイル中の一部のレベルは調整しない。一部の実施形態では、遺伝子の変動(例えば、染色体異数性)の存在または非存在の決定を、プロファイルに従って決定するが、ここで、調整は施さない。

10

【0428】

一部の実施形態では、プロファイル中のレベル(例えば、第1のレベル)の調整により、偽決定または偽アウトカムを低減する。一部の実施形態では、プロファイル中のレベル(例えば、第1のレベル)の調整により、偽決定または偽アウトカムの頻度および/または確率(例えば、統計学的確率、尤度)を低減する。偽決定または偽アウトカムは、正確ではない決定またはアウトカムでありうる。偽決定または偽アウトカムは、被験体(例えば、妊娠中の雌、胎児、および/またはこれらの組合せ)の、実際の遺伝子構成もしくは真の遺伝子構成または実際の遺伝的素質もしくは真の遺伝的素質(例えば、遺伝子の変動の存在または非存在)を反映しない決定またはアウトカムでありうる。一部の実施形態では、偽決定または偽アウトカムは、偽陰性決定である。一部の実施形態では、陰性決定または陰性アウトカムとは、遺伝子の変動(例えば、異数性、コピー数の変動)の非存在である。一部の実施形態では、偽決定または偽アウトカムは、偽陽性決定または偽陽性アウトカムである。一部の実施形態では、陽性決定または陽性アウトカムとは、遺伝子の変動(例えば、異数性、コピー数の変動)の存在である。一部の実施形態では、決定またはアウトカムを診断で活用する。一部の実施形態では、決定またはアウトカムは、胎児についての決定またはアウトカムである。

20

30

【0429】

遺伝子の変動(例えば、胎児の異数性)の存在または非存在は、場合によって、部分のセットについてのカウントを参照と比較せずに決定する。本明細書では、試験試料について測定されたカウントであり、試験領域(例えば、目的の部分のセット)中のカウントを、「試験カウント」と称する。試験カウントは、場合によって、本明細書で記載される、処理されたカウント、平均されたカウントもしくは合計されたカウント、表示、正規化されたカウント、あるいは1つもしくは複数のレベルまたは複数のレベルである。ある特定の実施形態では、部分のセットについて、試験カウントを平均または合計し(例えば、平均、平均値、中央値、モード、または合計を計算し)、平均されたカウントまたは合計されたカウントを、閾または範囲と比較する。試験カウントは、場合によって、第1の部分のセットについてのカウントの、第2の部分のセットについてのカウントに対する比または百分率として表されうる、表示として表される。ある特定の実施形態では、第1の部分のセットは、1つまたは複数の試験染色体(例えば、第13染色体、第18染色体、第21染色体、またはこれらの組合せ)についてのセットであり、場合によって、第2の部分のセットは、ゲノムまたはゲノムの部分(例えば、常染色体または常染色体および性染色体)についてのセットである。ある特定の実施形態では、表示を、閾または範囲と比較する。ある特定の実施形態では、試験カウントを、部分のセットにわたり正規化されたカウントについての1つもしくは複数のレベルまたは複数のレベルとして表し、1つもしくは複数のレベルまたは複数のレベルを、閾または範囲と比較する。特定の閾を上回るかまた

40

50

は下回り、場合によって、特定の範囲内または特定の範囲外にある、試験カウント（例えば、平均されたカウントまたは合計されたカウント、表示、正規化されたカウント、1つもしくは複数のレベルまたは複数のレベル）は、遺伝子の変動の存在または正倍数性の欠如（例えば、非正倍数性）の決定因である。特定の閾を上回るかまたは下回り、場合によって、特定の範囲内または特定の範囲外にある、試験カウント（例えば、平均されたカウントまたは合計されたカウント、表示、正規化されたカウント、1つもしくは複数のレベルまたは複数のレベル）は、遺伝子の変動または正倍数性の非存在の決定因である。

【0430】

遺伝子の変動（例えば、胎児の異数性）の存在または非存在は、場合によって、その非限定的な例が、試験カウント、参照カウント、未加工のカウント、フィルタリングされたカウント、平均されたカウントまたは合計されたカウント、表示（例えば、染色体表示）、正規化されたカウント、1つもしくは複数のレベルまたは複数のレベル（例えば、部分のセットについて、例えば、ゲノム区分のレベル、プロファイル）、Zスコアなど、またはこれらの組合せを含む、カウントを比較することにより決定する。一部の実施形態では、試験カウントを、参照（例えば、参照カウント）と比較する。参照（例えば、参照カウント）は、その非限定的な例が、未加工のカウント、フィルタリングされたカウント、平均されたカウントまたは合計されたカウント、表示（例えば、染色体表示）、正規化されたカウント、1つもしくは複数のレベルまたは複数のレベル（例えば、部分のセットについて、例えば、ゲノム区分のレベル、プロファイル）、Zスコアなど、またはこれらの組合せを含む、カウントの適切な決定でありうる。参照カウントは、正倍数体の試験領域についてのカウントまたは正倍数性であるゲノムもしくは染色体のセグメントからのカウントであることが多い。一部の実施形態では、参照カウントおよび試験カウントを、同じ試料および／または同じ被験体から得る。一部の実施形態では、参照カウントは、異なる試料および／または異なる被験体からのものである。一部の実施形態では、参照カウントは、試験カウントを導出および／または決定する、対応するゲノムのセグメントから決定し、かつ／またはそれと比較する。対応するセグメントとは、参照ゲノムの同じ位置へとマッピングされる、セグメント、部分、または部分のセットを指す。一部の実施形態では、参照カウントは、試験カウントを導出および／または決定する、異なるゲノムのセグメントから決定し、かつ／またはそれと比較する。

【0431】

ある特定の実施形態では、試験カウントは、場合によって、第1の部分のセットについてのカウントであり、参照は、第1の部分のセットと異なる、第2の部分のセットについてのカウントを含む。参照カウントは、場合によって、試験試料を得る同じ妊娠中の雌に由来する核酸試料についてのカウントである。ある特定の実施形態では、参照カウントは、試験試料を得た雌と異なる、1例または複数例の妊娠中の雌に由来する核酸試料についてのカウントである。一部の実施形態では、第1の部分のセットは、第13染色体中、第18染色体中、第21染色体中、これらのセグメント中、または前出の組合せ中にあり、第2の部分のセットは、別の1つまたは複数の染色体中またはそのセグメント中にある。第1の部分のセットが、第21染色体中またはそのセグメント中にある、非限定的な例では、第2の部分のセットは、別の染色体（例えば、第1染色体、第13染色体、第14染色体、第18染色体、第19染色体、そのセグメント、または前出の組合せ）中にあることが多い。参照は、正倍数体であることが典型的な染色体中またはそのセグメント中に位置することが多い。例えば、第1染色体および第19染色体は、胎児では、第1染色体異数性および第19染色体異数性と関連する、早期の胎児の死亡率が高率であることに起因して、正倍数体であることが多い。試験カウントと参照カウントとの偏差の尺度を、生成することができる。

【0432】

ある特定の実施形態では、参照は、試験カウントの場合と同じ部分のセットについてのカウントを含み、参照についてのカウントは、1つまたは複数の参照試料（例えば、複数の参照被験体に由来する複数の参照試料であることが多い）からのカウントである。参照

10

20

30

40

50

試料は、試験試料を得る雌と異なる、1例または複数例の妊娠中の雌に由来することが多い。試験カウントと参照カウントとの偏差の尺度（例えば、不確定性の尺度、不確定値）を、生成することができる。一部の実施形態では、偏差の尺度を、試験カウントから決定する。一部の実施形態では、偏差の尺度を、参照カウントから決定する。一部の実施形態では、偏差の尺度を、全プロファイルまたはプロファイル中の部分のサブセットから決定する。

【0433】

偏差の適切な尺度であって、その非限定的な例が、標準偏差、平均絶対偏差、中央値絶対偏差、最大絶対偏差、標準スコア（例えば、 z 値、 z スコア、正規スコア、標準化された変数）などを含む尺度を選択することができる。一部の実施形態では、参照試料は、試験領域について正倍数体であり、試験カウントと参照カウントとの偏差を評価する。一部の実施形態では、遺伝子の変動の存在または非存在の決定は、ゲノムまたは染色体のセグメントまたは部分についての、試験カウントと参照カウントとの偏差（例えば、偏差の尺度、 MAD ）の数に従う。一部の実施形態では、試験カウントと参照カウントとの偏差の数が、約1超、約1.5超、約2超、約2.5超、約2.6超、約2.7超、約2.8超、約2.9超、約3超、約3.1超、約3.2超、約3.3超、約3.4超、約3.5超、約4超、約5超、または約6超である場合に、遺伝子の変動の存在を決定する。例えば、場合によって、試験カウントが、参照カウントと、偏差の尺度（例えば、3シグマ、3 MAD ）で3超異なれば、遺伝子の変動の存在を決定する。一部の実施形態では、妊娠中の雌から得られる試験カウントが、参照カウントより、偏差の尺度（例えば、3シグマ、3 MAD ）で3超大きければ、胎児の染色体異数性（例えば、胎児のトリソミー）の存在が決定される。試験カウントと参照カウントとの3超の偏差は、非正倍数体の試験領域（例えば、遺伝子の変動の存在）を指し示すことが多い。場合によって、正倍数性を指し示す参照カウントを有意に上回る試験カウントは、トリソミーの決定因である。一部の実施形態では、妊娠中の雌から得られる試験カウントが、参照カウントより偏差の尺度（例えば、3シグマ、3 MAD ）で3超小さければ、胎児の染色体異数性（例えば、胎児のモノソミー）の存在が決定される。場合によって、正倍数性を指し示す参照カウントを有意に下回る試験カウントは、モノソミーの決定因である。

【0434】

一部の実施形態では、試験カウントと参照カウントとの偏差の数が、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未満である場合に、遺伝子の変動の非存在を決定する。例えば、場合によって、試験カウントが、参照カウントと、偏差の尺度（例えば、3シグマ、3 MAD ）で3未満異なれば、遺伝子の変動の非存在が決定される。一部の実施形態では、妊娠中の雌から得られる試験カウントが、参照カウントと、偏差の尺度（例えば、3シグマ、3 MAD ）で3未満異なれば、胎児の染色体異数性の非存在（例えば、胎児の正倍数体）が決定される。一部の実施形態では、（例えば、試験カウントと参照カウントとの3未満の偏差（例えば、標準偏差では、3シグマ）は、正倍数体の試験領域（例えば、遺伝子の変動の非存在）を指し示すことが多い。試験試料についての試験カウントと、1つまたは複数の参照被験体についての参照カウントとの偏差の尺度は、プロットし、視覚化する（例えば、 z スコアプロット）ことができる。

【0435】

他の任意の適切な参照は、試験試料の試験領域について、遺伝子の変動の存在または非存在を決定する（または正倍数体もしくは非正倍数体の決定の）ための試験カウントで因子分解することができる。例えば、胎児フラクションの決定は、試験カウントで因子分解して、遺伝子の変動の存在または非存在を決定することができる。胎児フラクションを定量するための適切な処理であって、その非限定的な例が、質量分析処理、配列決定処理、またはこれらの組合せを含む処理を活用することができる。

【0436】

一部の実施形態では、胎児の染色体異数性（例えば、トリソミー）の存在または非存在は、一部分、胎児の倍数性の決定から決定される。一部の実施形態では、胎児の倍数性を、本明細書で記載される適切な方法により決定する。一部のある特定の実施形態では、約 1.20 もしくはそれ超、1.25 もしくはそれ超、1.30 もしくはそれ超、約 1.35 もしくはそれ超、約 1.4 もしくはそれ超、または約 1.45 もしくはそれ超の胎児の倍数性の決定は、胎児の染色体異数性の存在（例えば、胎児のトリソミーの存在）を指し示す。一部の実施形態では、約 1.20～約 2.0、約 1.20～約 1.9、約 1.20～約 1.85、約 1.20～約 1.8、約 1.25～約 2.0、約 1.25～約 1.9、約 1.25～約 1.85、約 1.25～約 1.8、約 1.3～約 2.0、約 1.3～約 1.9、約 1.3～約 1.85、約 1.3～約 1.8、約 1.35～約 2.0、約 1.35～約 1.9、約 1.35～約 1.8、約 1.4～約 2.0、約 1.4～約 1.85、または約 1.4～約 1.8 の胎児の倍数性の決定は、胎児の染色体異数性の存在（例えば、胎児のトリソミーの存在）を指し示す。一部の実施形態では、胎児の異数性は、トリソミーである。一部の実施形態では、胎児の異数性は、第 13 染色体、第 18 染色体、および／または第 21 染色体のトリソミーである。

【0437】

一部の実施形態では、約 1.35 未満、約 1.30 未満、約 1.25 未満、約 1.20 未満、または約 1.15 未満の胎児の倍数性は、胎児の異数性の非存在（例えば、胎児のトリソミーの非存在、例えば、正倍数体）を指し示す。一部の実施形態では、約 0.7～約 1.35、約 0.7～約 1.30、約 0.7～約 1.25、約 0.7～約 1.20、約 0.7～約 1.15、約 0.75～約 1.35、約 0.75～約 1.30、約 0.75～約 1.25、約 0.75～約 1.20、約 0.75～約 1.15、約 0.8～約 1.35、約 0.8～約 1.30、約 0.8～約 1.25、約 0.8～約 1.20、または約 0.8～約 1.15 の胎児の倍数性の決定は、胎児の染色体異数性の非存在（例えば、胎児のトリソミーの非存在、例えば、正倍数体）を指し示す。

【0438】

一部の実施形態では、約 0.8 未満、約 0.75 未満、約 0.70 未満、または約 0.6 未満の胎児の倍数性は、胎児の異数性の存在（例えば、染色体欠失の存在）を指し示す。一部の実施形態では、約 0～約 0.8、約 0～約 0.75、約 0～約 0.70、約 0～約 0.65、約 0～約 0.60、約 0.1～約 0.8、約 0.1～約 0.75、約 0.1～約 0.70、約 0.1～約 0.65、約 0.1～約 0.60、約 0.2～約 0.8、約 0.2～約 0.75、約 0.2～約 0.70、約 0.2～約 0.65、約 0.2～約 0.60、約 0.25～約 0.8、約 0.25～約 0.75、約 0.25～約 0.70、約 0.25～約 0.65、約 0.25～約 0.60、約 0.3～約 0.8、約 0.3～約 0.75、約 0.3～約 0.70、約 0.3～約 0.65、約 0.3～約 0.60 の胎児の倍数性の決定は、胎児の染色体異数性の存在（例えば、染色体欠失の存在）を指し示す。一部の実施形態では、決定される胎児の異数性は、全染色体欠失である。

【0439】

一部の実施形態では、胎児の異数性の存在または非存在の決定（例えば、上記の倍数性の決定の範囲のうちの 1 または複数に従う）を、判定域（call zone）に従って決定する。ある特定の実施形態では、値（例えば、倍数性値、胎児フラクション値、不確定性のレベル）または値のコレクションが、あらかじめ規定された範囲（例えば、帯域、判定域）内にある場合に、判定（例えば、遺伝子の変動の存在または非存在を決定する判定、例えば、アウトカム）を下す。一部の実施形態では、判定域を、同じ患者試料から得られる値のコレクションに従って規定する。ある特定の実施形態では、判定域を、同じ染色体またはそのセグメントから導出される値のコレクションに従って規定する。一部の実施形態では、倍数性の決定に基づく判定域を、信頼性レベル（例えば、高い信頼性レベル、例えば、低い不確定性のレベル）および／または胎児フラクションに従って規定する。一部の実施形態では、判定域を、倍数性の決定および約 2.0% もしくはそれ超、約 2.5% もしくはそれ超、約 3% もしくはそれ超、約 3.25% もしくはそれ超、約 3.5%

もしくはそれ超、約 3.75%もしくはそれ超、または約 4.0%もしくはそれ超の胎児フラクションに従って規定する。例えば、一部の実施形態では、胎児を出産する妊娠中の雌から得られた試料についての、2%もしくはそれ超または4%もしくはそれ超の胎児フラクションの決定を伴う、1.25超の倍数性の決定に基づき、胎児は、トリソミー 21を含むという判定を下す。ある特定の実施形態では、例えば、胎児を出産する妊娠中の雌から得られた試料についての、2%もしくはそれ超または4%もしくはそれ超の胎児フラクションの決定を伴う、1.25未満の倍数性の決定に基づき、胎児は、正倍数体であるという判定を下す。一部の実施形態では、判定域は、約 99%もしくはそれ超、約 99.1%もしくはそれ超、約 99.2%もしくはそれ超、約 99.3%もしくはそれ超、約 99.4%もしくはそれ超、約 99.5%もしくはそれ超、約 99.6%もしくはそれ超、約 99.7%もしくはそれ超、約 99.8%もしくはそれ超、または約 99.9%もしくはそれ超の信頼性レベルにより規定する。一部の実施形態では、判定域を使用せずに判定を下す。一部の実施形態では、判定域およびさらなるデータまたは情報を使用して判定を下す。一部の実施形態では、判定域の使用を伴わずに、倍数性値に基づき判定を下す。一部の実施形態では、倍数性値を計算せずに判定を下す。一部の実施形態では、プロファイルの目視（例えば、ゲノム区分のレベルの目視）に基づき判定を下す。判定は、その非限定的な例が、胎児の倍数性の決定、胎児フラクションの決定、母体の倍数性、不確定性および／または信頼性決定、部分レベル、レベル、プロファイル、Zスコア、期待された染色体表示、測定された染色体表示、カウント（例えば、正規化されたカウント、未加工のカウント）、胎児のまたは母体のコピー数の変動（例えば、類別されたコピー数の変動）、有意に異なるレベル、調整されたレベル（例えば、穴埋め）など、またはこれらの組合せを含む、本明細書で記載される方法により得られた決定、値、および／またはデータに完全に、または一部分基づく任意の適切な方法により下すことができる。

【0440】

一部の実施形態では、判定を下さない場合、判定域は存在しない。一部の実施形態では、判定域が存在しないことは、低い精度、高い危険性、大きな誤差、低い信頼性レベル、高い不確定性のレベルなど、またはこれらの組合せを指し示す値または値のコレクションにより規定される。一部の実施形態では、判定域が存在しないことは、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フラクションにより一部分規定される。

【0441】

遺伝子の変動は、場合によって、医学的状态と関連する。遺伝子の変動の決定因であるアウトカムは、場合によって、状態（例えば、医学的状态）、疾患、症候群、もしくは異常の存在または非存在の決定因であるアウトカムであるか、または状態、疾患、症候群、もしくは異常（例えば、表 1 に列挙された非限定的な例）の検出を含む。ある特定の実施形態では、診断は、アウトカムについての評価を含む。本明細書で記載される方法による状態（例えば、医学的状态）、疾患、症候群、または異常の存在または非存在の決定因であるアウトカムは、場合によって、さらに調べることにより（例えば、核型分析および／または羊水穿刺により）、独立に確かめることができる。データの分析および処理は、1 つまたは複数のアウトカムを提示しうる。本明細書で使用される「アウトカム」という用語は、遺伝子の変動（例えば、異数性、コピー数の変動）の存在または非存在を決定することを容易とする、データ処理の結果を指すことができる。ある特定の実施形態では、本明細書で使用される「アウトカム」という用語は、遺伝子の変動（例えば、異数性、コピー数の変動）の存在または非存在を予測および／または決定する結論を指す。ある特定の実施形態では、本明細書で使用される「アウトカム」という用語は、被験体（例えば、胎児）における遺伝子の変動の存在または非存在（例えば、異数性、コピー数の変動）の危険性または確率を予測および／または決定する結論を指す。診断は、場合によって、アウトカムの使用を含む。例えば、医療従事者は、アウトカムを分析し、アウトカムに基づくか、またはアウトカムに一部分基づき、診断を提示することができる。一部の実施形態で

は、状態、症候群、または異常（例えば、表1に列举された）についての決定、検出、または診断は、遺伝子の変動の存在または非存在の決定因であるアウトカムの使用を含む。一部の実施形態では、カウントされた、マッピングした配列の読取りまたはその変換に基づくアウトカムは、遺伝子の変動の存在または非存在の決定因である。ある特定の実施形態では、本明細書で記載される1つまたは複数の方法（例えば、データ処理法）を活用して生成されたアウトカムは、表1に列举された1つまたは複数の状態、症候群、または異常の存在または非存在の決定因である。ある特定の実施形態では、診断は、状態、症候群、または異常の存在または非存在の決定を含む。診断は、状態、症候群、または異常の性質および／または原因としての遺伝子の変動の決定を含むことが多い。ある特定の実施形態では、アウトカムは、診断ではない。1つまたは複数の確率の検討事項の文脈では、アウトカムは、本明細書で記載される処理法を使用して生成される1つまたは複数の数値を含むことが多い。危険性または確率の検討事項は、不確定値、可変性の尺度、信頼性レベル、感度、特異性、標準偏差、変動係数（CV）および／または信頼性レベル、Zスコア、カイ値、phi値、倍数性値、適合させた胎児フラクション、面積比、中央値レベルなど、またはこれらの組合せを含みうるがこれらに限定されない。確率の検討事項により、被験体に遺伝子の変動を有する危険性があるかまたは被験体が遺伝子の変動を有するのかどうかを決定することを容易とすることが可能になり、遺伝子障害の存在または非存在の決定因であるアウトカムは、このような検討事項を含むことが多い。

10

【0442】

アウトカムは、場合によって、表現型である。アウトカムは、場合によって、関連する信頼性レベル（例えば、不確定値、例えば、胎児は、99%の信頼性レベルでトリソミー21について陽性であり、試験被験体は、95%の信頼性レベルで、遺伝子の変動と関連するがんについて陰性である）を有する表現型である。アウトカム値を生成する異なる方法は、場合によって、異なる種類の結果をもたらさう。一般に、本明細書で記載される方法を使用して生成されるアウトカム値に基づき下されうる4種類の可能なスコアまたは判定：真陽性、偽陽性、真陰性、および偽陰性が存在する。本明細書で使用される「スコア（score）」、「スコア（scores）」、「判定（call）」、および「判定（calls）」という用語は、特定の遺伝子の変動が、被験体／試料に存在するかまたは非存在である確率を計算することを指す。スコアの値を使用して、例えば、遺伝子の変動に対応しうる、マッピングした配列の読取りの変動、差違、または比を決定することができる。例えば、データセットに由来する、選択された遺伝子の変動または部分について、参照ゲノムに対して正のスコアを計算することにより、場合によって、医学的状态（例えば、がん、子癇前症、トリソミー、モノソミーなど）と関連する、遺伝子の変動の存在または非存在の同定をもたらすことができる。一部の実施形態では、アウトカムは、レベル、プロファイル、および／またはプロット（例えば、プロファイルのプロット）を含む。アウトカムが、プロファイルを含む実施形態では、適切なプロファイルまたはプロファイルの組合せを、アウトカムのために使用することができる。アウトカムのために使用されうる、プロファイルの非限定的な例は、Zスコアプロファイル、p値プロファイル、カイ値プロファイル、phi値プロファイルなど、およびこれらの組合せを含む。

20

30

【0443】

遺伝子の変動の存在または非存在を決定するために生成されたアウトカムは、場合によって、ヌルの結果（例えば、2つのクラスター間のデータ点、遺伝子の変動の存在および非存在の両方についての値を包含する標準偏差を有する数値、調査される遺伝子の変動を有するかまたは含まない被験体についてのプロファイルのプロットと同様ではないプロファイルのプロットを有するデータセット）を含む。一部の実施形態では、ヌルの結果を指し示すアウトカムもやはり決定因の結果であり、決定は、遺伝子の変動の存在または非存在を決定するためのさらなる情報および／またはデータ生成の反復および／または分析に対する必要を含みうる。

40

【0444】

一部の実施形態では、アウトカムは、本明細書で記載される、1つまたは複数の処理ス

50

テップを実施した後で生成することができる。ある特定の実施形態では、アウトカムは、本明細書で記載される処理ステップのうちの1つの結果として生成し、一部の実施形態では、アウトカムは、データセットの各統計学的操作および／または各数学的操作を実施した後で生成することができる。遺伝子の変動の存在または非存在の決定に関するアウトカムは、限定せずに述べると、確率（例えば、オッズ比、p 値）、尤度、クラスター中またはクラスター外の値、閾を上回る値または閾を下回る値、範囲（例えば、閾範囲）内の値、分散または信頼性の尺度を有する値、または被験体もしくは試料についての遺伝子の変動の存在もしくは非存在と関連する危険性因子を含む、適切な形態で表すことができる。ある特定の実施形態では、試料間の比較は、試料の識別の確認を可能とする（例えば、反復された試料および／または混合された試料（例えば、誤表示された試料、組み合わせられた試料など）の同定を可能とする）。

10

【0445】

一部の実施形態では、アウトカムは、所定の閾またはカットオフ値を上回るかまたは下回る値（例えば、1 超の値、1 未満の値）、およびその値と関連する不確定性のレベルまたは信頼性レベルを含む。ある特定の実施形態では、所定の閾値またはカットオフ値は、期待レベルまたは期待レベルの範囲である。アウトカムはまた、データ処理において使用される仮定についても記載しうる。ある特定の実施形態では、アウトカムは、所定の値の範囲（例えば、閾範囲）内または範囲外にある値、および範囲内または範囲外にあるその値についての、関連する不確定性のレベルまたは信頼性レベルを含む。一部の実施形態では、アウトカムは、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、および等しいかまたは範囲内にあるかもしくは範囲外にあるその値についての、その関連する不確定性のレベルまたは信頼性レベルを含む。アウトカムは、場合によって、プロット（例えば、プロファイルのプロット）としてグラフィ的に表される。

20

【0446】

上記で注目した通り、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ることができる。本明細書で使用される「真陽性」という用語は、遺伝子の変動を有するとして被験体が正しく診断されたことを指す。本明細書で使用される「偽陽性」という用語は、遺伝子の変動を有するとして被験体が誤って同定されたことを指す。本明細書で使用される「真陰性」という用語は、遺伝子の変動を有さないとして被験体が正しく同定されたことを指す。本明細書で使用される「偽陰性」という用語は、遺伝子の変動を有さないとして被験体が誤って同定されたことを指す。任意の所与の方法についての性能の2つの尺度は、(i) 一般に、予測された陽性の割合であって、陽性として正しく同定された割合である感度値；および (ii) 一般に、予測された陰性の割合であって、陰性として正しく同定された割合である特異性値の発生比に基づき計算することができる。

30

【0447】

ある特定の実施形態では、感度、特異性、および／または信頼性レベルのうちの1または複数は、百分率として表される。一部の実施形態では、百分率は、各変数について独立に、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%またはそれ超、約99.9%またはそれ超、約99.95%またはそれ超、約99.99%またはそれ超））である。一部の実施形態では、変動係数(CV)は、百分率として表され、場合によって、百分率は、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。ある特定の実施形態では、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、Zスコア、p 値、またはt 検定の結果として表される。一部の実施形態では、アウトカムについての、測定された分散、信頼区間、感度、特異性など（例えば、併せて、信頼性パラメータと称する）は、本明細書で記載される、1つまたは複数のデータ処理操作を使用して生成することができる。アウトカムおよび関連する信頼性レベルを生成することの具体例は

40

50

、実施例の節ならびに本文、表、等式、および図面の全てを含むその全内容が参照により本明細書に援用される、国際特許出願第 P C T / U S 1 2 / 5 9 1 2 3 号 (W O 2 0 1 3 / 0 5 2 9 1 3) において記載されている。

【0448】

本明細書で使用される「感度」という用語は、真陽性の数を、真陽性の数に偽陰性の数を加算して得た数で除算して得たものを指し、ここで感度 (s e n s) は、 $0 \leq s e n s \leq 1$ の範囲内でありうる。本明細書で使用される「特異性」という用語は、真陰性の数を、真陰性の数に偽陽性の数を加算して得た数で除算して得たものを指し、ここで感度 (s p e c) は、 $0 \leq s p e c \leq 1$ の範囲内でありうる。一部の実施形態では、場合によって、感度および特異性が1もしくは100%に等しいか、または1の近傍にある（例えば、約90%～約99%間にある）方法を選択する。一部の実施形態では、感度が1または100%に等しい方法を選択し、ある特定の実施形態では、感度が1の近傍にある（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%の感度である）方法を選択する。一部の実施形態では、特異性が1または100%に等しい方法を選択し、ある特定の実施形態では、特異性が1の近傍にある（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%の特異性である）方法を選択する。

10

【0449】

一部の実施形態では、遺伝子の変動の存在または非存在（例えば、染色体異数性）を、胎児について決定する。このような実施形態では、胎児の遺伝子の変動（例えば、胎児の染色体異数性）の存在または非存在を決定する。

20

【0450】

ある特定の実施形態では、試料についての、遺伝子の変動（例えば、染色体異数性）の存在または非存在を決定する。このような実施形態では、試料核酸中の、遺伝子の変動（例えば、染色体異数性）の存在または非存在を決定する。一部の実施形態では、検出される変動または検出されない変動は、1つの供給源に由来する試料核酸中には存在するが、別の供給源に由来する試料核酸中には存在しない。供給源の非限定的な例は、胎盤の核酸、胎児核酸、母体核酸、がん細胞の核酸、非がん細胞の核酸など、およびこれらの組合せを含む。非限定的な例では、検出されるまたは検出されない、特定の遺伝子の変動は、（i）胎盤の核酸中には存在するが、胎児核酸中には存在せず、母体核酸中にも存在しないか、（i i）胎児核酸中には存在するが、母体核酸中には存在しないか、または（i i i）母体核酸中には存在するが、胎児核酸中には存在しない。

30

【0451】

1つまたは複数のアウトカムを生成した後で、アウトカムを使用して、遺伝子の変動の存在もしくは非存在および／または関連する医学的状态の決定をもたらすことが多い。アウトカムは、医療従事者（例えば、検査室技師または管理者；医師または助手）へと提示することが典型的である。アウトカムは、アウトカムモジュールにより提示することが多い。ある特定の実施形態では、アウトカムを、プロットングモジュールにより提示する。ある特定の実施形態では、アウトカムは、装置の周辺機器上またはコンポーネント上に提示される。例えば、場合によって、アウトカムを、プリンターまたはディスプレイにより提示する。一部の実施形態では、遺伝子の変動の存在または非存在の決定因であるアウトカムは、医療従事者へと、レポートの形態で提示され、ある特定の実施形態では、レポートは、アウトカム値および関連する信頼性パラメータの提示を含む。一般に、アウトカムは、遺伝子の変動の存在もしくは非存在および／または医学的状态の決定を容易とする、適切なフォーマットで示すことができる。データセットを報告および／もしくは提示するか、またはアウトカムを報告するための使用に適するフォーマットの非限定的な例は、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、ピクトグラフ、チャート、棒グラフ、円グラフ、概略図、フローチャート、散布図、マップ、ヒ

40

50

ストグラム、密度図、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、コンターダイアグラム、カルトグラム、レーダーチャート、ベン図、ノモグラムなど、および前出の組合せを含む。アウトカム表示の多様な例については、図面で示し、実施例で記載する。

【0452】

ある特定の実施形態では、アウトカムの生成は、核酸配列の読取りデータなどの、被験体の細胞核酸の表示への変換と考えることができる。例えば、被験体に由来する核酸の配列の読取りを分析し、染色体のプロファイルおよび／またはアウトカムを生成することは、比較的小さな配列の読取り断片の、比較的大きな染色体構造の表示への変換と考えることができる。一部の実施形態では、アウトカムは、被験体（例えば、妊娠中の雌）に由来する配列の読取りの、被験体（例えば、母体核酸および／または胎児核酸）内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）の表示への変換の結果として得られる。一部の実施形態では、アウトカムは、第1の被験体（例えば、妊娠中の雌）に由来する配列の読取りの、構造（例えば、ゲノム、染色体またはそのセグメント）の複合表示への変換、ならびに第1の被験体（例えば、妊娠中の雌）内および／または第2の被験体（例えば、胎児）内に存在する構造の表示をもたらす複合表示の第2の変換を含む。

10

【0453】

ある特定の実施形態では、アウトカムは、1つまたは複数の候補セグメントの分析に従って生成することができる。一部の実施形態では、遺伝子の変動の存在または非存在を、個別セグメント、候補セグメント、または複合候補セグメント（例えば、個別セグメント、候補セグメント、または複合候補セグメントの存在または非存在）に従って決定する。一部の実施形態では、同じプロファイルの2つの分解レンダリングから導出された2つの候補セグメントは、実質的に同じであり（例えば、比較に従う）、染色体異数性、微小重複、または微小欠失の存在が決定される。一部の実施形態では、複合候補セグメントの存在により、染色体異数性、微小重複、または微小欠失の存在を指し示す。一部の実施形態では、全染色体異数性の存在は、プロファイル中の個別セグメント、候補セグメント、または複合候補セグメントの存在に従って決定され、プロファイルは、ゲノムのセグメント（例えば、染色体より大きいセグメント、例えば、2つまたはそれ超の染色体を表示するセグメント、全ゲノムを表示するセグメント）である。一部の実施形態では、全染色体異数性の存在は、プロファイル中の個別セグメント、候補セグメント、または複合候補セグメントの存在に従って決定され、個別セグメントのエッジは、染色体のエッジと実質的に同じである。ある特定の実施形態では、プロファイル中の個別セグメント、候補セグメント、または複合候補セグメントの少なくとも1つのエッジが、染色体および／または染色体中の個別セグメントのエッジと異なる場合に、微小重複または微小欠失の存在を決定する。一部の実施形態では、微小重複の存在が決定され、個別セグメント、候補セグメント、または複合候補セグメントについてのレベルまたはAUCは、参照レベル（例えば、正倍数体の領域）より実質的に大きい。一部の実施形態では、微小欠失の存在が決定され、個別セグメント、候補セグメント、または複合候補セグメントについてのレベルまたはAUCは、実質的に参照レベル（例えば、正倍数体の領域）未満である。一部の実施形態では、2つまたはそれ超の異なる分解レンダリング中で同定される候補セグメントは、実質的に同じでなく（例えば、異なり）、染色体異数性、微小重複、および／または微小欠失の非存在が決定される。一部の実施形態では、プロファイル中の個別セグメント、候補セグメント、もしくは複合候補セグメント、またはプロファイルの分解レンダリングの非存在により、染色体異数性、微小重複、または微小欠失の非存在を指し示す。

20

30

40

【0454】

検証

一部の実施形態では、本明細書で記載される方法は、検証を含む。一部の実施形態では、決定（decision）分析（例えば、決定（decision）木）、遺伝子の変動（例えば、コピー数の変動、微小重複、微小欠失、異数性）の存在または非存在の決定

50

(d e t e r m i n a t i o n)、判定を下すこと、および／またはアウトカムの決定 (d e t e r m i n a t i o n) は、検証を含む。任意の適切な検証処理を活用して、本明細書で記載される方法、判定、またはアウトカムを検証することができる。

【0455】

一部の実施形態では、検証は、分解レンダリング中で同定される候補セグメントを検証することまたは無効にする (i n v a l i d a t i n g) ことを含む。検証された候補セグメントにより、候補セグメントの存在を確認する。無効にされた候補セグメントにより、候補セグメントの存在を指し示す判定を、候補セグメントの非存在を指し示す判定へと変化させる。例えば、一部の実施形態では、セグメント化処理による候補セグメントの同定に続き、検証を実施することができ、ここで、候補セグメントを検証するか、または無効にする。無効にされた候補セグメントは、プロファイル中の染色体異数性、微小重複、または微小欠失の非存在を指し示す。一部の実施形態では、検証は、偽陰性の決定が低減され、かつ／または偽陽性の決定が低減された候補セグメントの存在または非存在の決定を含む。候補セグメントは、その非限定的な例が、「スライディングエッジ」処理、「リープワンアウト」処理など、またはこれらの組合せを含む、適切な方法により検証することができる。

10

【0456】

一部の実施形態では、検証は、候補セグメントまたは複合候補セグメントについての有意性のレベルを生成することを含む。一部の実施形態では、有意性のレベルは、Zスコア、z 値、p 値などである。一部の実施形態では、検証は、不確定性のレベルを生成することを含む。一部の実施形態では、不確定性のレベルは、有意性のレベルに関連する。例えば、場合によって、平均、平均値、または中央値による有意性のレベルを決定し、不確定性のレベルを、平均、平均値、または中央値による有意性のレベルについて決定する。

20

【0457】

一部の実施形態では、有意性のレベルおよび／または不確定値に従って、候補セグメントを検証するか、または無効にする。検証されるかまたは無効にされた個別セグメントは、検証されるかまたは無効にされた複合候補セグメントでありうる。一部の実施形態では、検証された候補セグメントの存在または非存在を、候補セグメントについての有意性のレベルおよび／または不確定性のレベルに従って決定する。一部の実施形態では、検証された候補セグメントの非存在により、染色体異数性、微小重複、または微小欠失の非存在を指し示す。一部の実施形態では、検証された候補セグメントの存在により、候補セグメントの存在を確認する。一部の実施形態では、2つまたはそれ超の検証された候補セグメントの存在により、複合候補セグメントの決定または生成がもたらされる。一部の実施形態では、1つまたは複数の検証された候補セグメントの存在により、一部分、信頼性レベルを増加させて、染色体異数性、微小重複、または微小欠失の存在を決定する。一部の実施形態では、候補セグメントの存在により、一部分、ディジョージ症候群の存在を指し示す。一部の実施形態では、検証された候補セグメントの非存在により、染色体異数性、微小重複、または微小欠失の非存在を指し示す。

30

【0458】

スライディングエッジによる検証

40

一部の実施形態では、検証は、「スライディングエッジ」処理を含む。適切な「スライディングエッジ」処理は、分解レンダリング中のセグメントを検証するために直接使用することもでき、そのために適応させることもできる。一部の実施形態では、「スライディングエッジ」処理は、候補セグメント（例えば、部分のセットにより表示される候補セグメント）、または候補セグメントを含むかもしくは候補セグメントであることが疑われるセグメントを、複数の部分のサブセットへとセグメント化することを含む。一部の実施形態では、候補セグメントは、全染色体または染色体のセグメントについての部分のセットである。一部の実施形態では、候補セグメントは、公知の遺伝子の変動または公知の遺伝子障害と関連する領域を含む部分のセットである。一部の実施形態では、候補セグメントは、ディジョージ領域を含む。

50

【0459】

ある特定の実施形態では、「スライディングエッジ」処理は、同定された候補セグメント（部分のセット）を、複数の部分のサブセットへとセグメント化することを含み、ここで、部分のサブセットの各々は、同様であるが異なるエッジを有する候補セグメントを表示する。一部の実施形態では、元の同定された候補セグメントを、分析に組み込む。例えば、元の同定された候補セグメントを、複数の部分のサブセットのうちの1つとして組み込む。部分のサブセットは、元の同定された個別セグメントの一方または両方のエッジを、任意の適切な方法により変化させることにより決定することができる。一部の実施形態では、左エッジを変化させ、これにより、異なる左エッジを有する個別セグメントを生成することができる。一部の実施形態では、右エッジを変化させ、これにより、異なる右エッジを有する個別セグメントを生成することができる。一部の実施形態では、右エッジおよび左エッジのいずれも変化させることができる。一部の実施形態では、エッジを、1つまたは複数の隣接する参照ゲノムの部分だけ、元のエッジの左または右へと移動させることにより、エッジを変化させる。

10

【0460】

実施例5で記載されるスライディングエッジ法の実施形態では、両方のエッジを移動させることにより、元の個別セグメントを、15参照ゲノムの部分だけ変化させ、これにより、個別セグメントによる15×15グリッド（例えば、225の異なる部分のサブセット）を創出する。例えば、右エッジを安定に保ちながら、左エッジを、右に7参照ゲノムの部分だけ移動させ、次いで、左に7参照ゲノムの部分だけ移動させ、これにより、15の可能な左エッジを生成することができる。15の左エッジの各々を安定に保ちながら、右エッジを、右へと7参照ゲノムの部分だけ移動させ、左へと7参照ゲノムの部分だけ移動させ、これにより、15の可能な右エッジを生成することができる、結果として得られるサブセットは、225の異なる個別セグメント（例えば、参照ゲノムの部分のサブセット）を含む。

20

【0461】

一部の実施形態では、一方または両方のエッジを、5～30参照ゲノムの部分だけ変化させる。一部の実施形態では、エッジを、いずれかの方向に、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の参照ゲノムの部分だけ移動させる。一部の実施形態では、部分のサイズに関わらず、一方または両方のエッジについて、エッジを変化させて、約100,000～約2,000,000塩基対、250,000～約1,500,000塩基対、または約500,000～約1,000,000塩基対の範囲のエッジを生成する。一部の実施形態では、部分のサイズに関わらず、一方または両方のエッジについて、エッジを変化させて、約500,000、600,000、700,000、750,000、800,000、900,000、または約1,000,000塩基対の範囲のエッジを生成する。

30

【0462】

一部の実施形態では、同定された個別セグメントは、第1の末端および第2の末端を含み、セグメント化は、(i)再帰的除外により、1つまたは複数の部分を、部分のセットの第1の末端から除外し、これにより、部分のサブセットに各々の再帰的除外を施すことと、(ii)n回にわたる反復の後で、(i)の再帰的除外を終結させ、これにより、n+1の部分のサブセットをもたらすことと、ここで、部分のセットは、サブセットであり、各サブセットは、異なる数の部分、第1のサブセットの末端、および第2のサブセットの末端を含むことと、(iii)1つまたは複数の部分を、(ii)で再帰的除外によりもたらされた、n+1の部分のサブセットの各々のうちの、第2のサブセットの末端から除外することと、(iv)n回にわたる反復の後で、(iii)の再帰的除外を終結させ、これにより、複数の部分のサブセットをもたらすこととを含む。一部の実施形態では、複数のサブセットは、(n+1)²サブセットに等しい。一部の実施形態では、nは、5～30の間の整数に等しい。一部の実施形態では、nは、1、2、3、4、5、6、7、8、9

40

50

、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、または30に等しい。

【0463】

スライディングエッジ法のある特定の実施形態では、有意性のレベル（例えば、Zスコア、p値）を、参照ゲノムの部分のサブセットの各々について決定し、平均、平均値、または中央値による有意性のレベルを、サブセットの全てについて決定された有意性のレベルに従って決定する。

一部の実施形態では、有意性のレベルは、Zスコアまたはp値である。一部の実施形態では、Zスコアを、以下の式：

$$Z_i = (E_i - \text{Med. } E_{(n)}) / \text{MAD}$$

10

【0464】

〔式中、 E_i は、個別セグメントiのレベルの定量的決定であり、 $\text{Med. } E_{(n)}$ は、スライディングエッジ処理により生成された全ての個別セグメントについての中央値レベルであり、MADは、 $\text{Med. } E_{(n)}$ についての中央値絶対偏差であり、 Z_i は、個別セグメントiについて結果として得られるZスコアである〕に従って計算する。一部の実施形態では、MADは、任意の適切な不確定性の尺度で置きかえることができる。一部の実施形態では、 E_i は、その非限定的な例が、部分についてのカウントの、中央値レベル、平均レベル、平均値レベル、合計などを含む、レベルの任意の適切な尺度である。

【0465】

一部の実施形態では、中央値、平均値、または平均によるZスコアを、スライディングエッジ処理により生成された全ての個別セグメントについて決定し、不確定性のレベル（例えば、MAD）を、中央値、平均値、または平均によるZスコアから生成する。一部の実施形態では、スライディングエッジ処理により生成された全ての個別セグメントについて決定された、中央値、平均値、または平均によるZスコア、および中央値、平均値、または平均によるZスコアについての不確定性のレベルに従って、個別セグメント（例えば、同定された元の個別セグメント）を検証するかまたは無効にする。一部の実施形態では、有意性のレベル（例えば、Zスコア）についての所定の範囲（例えば、閾範囲）を、あらかじめ決定する。一部の実施形態では、候補セグメントの非存在についてのZスコアについての所定の範囲は、約3.5～約-3.5、約3.25～約-3.25、約3.0～約-3.0、約2.75～約-2.75、または約2.5～約-2.5である。一部の実施形態では、中央値、平均値、または平均によるZスコアであって、所定の範囲外の値を有するZスコアにより、「スライディングエッジ」法に従って、検証された個別セグメントの存在を確認する。一部の実施形態では、中央値、平均値、または平均によるZスコアであって、所定の範囲内の値を有するZスコアにより、「スライディングエッジ」法に従って、候補セグメントを無効にし、かつ／または候補セグメントの非存在（例えば、検証された候補セグメントの非存在）を決定する。一部の実施形態では、中央値、平均値、または平均によるZスコアであって、約2.25、2.5、2.75、3.0、3.25、または3.5超の絶対値を有するZスコアにより、「スライディングエッジ」法に従って、個別セグメントの存在を確認し、かつ／または検証する。一部の実施形態では、中央値、平均値、または平均によるZスコアであって、約2.25、2.5、2.75、3.0、3.25、または3.5未満の絶対値を有するZスコアにより、「スライディングエッジ」法に従って、候補セグメントの非存在を決定し、かつ／または無効にする。一部の実施形態では、中央値Zスコアと関連する不確定値により、個別セグメントが検証されるのか、無効にされるのかを一部分決定する。一部の実施形態では、中央値、平均値、または平均によるZスコアが、閾範囲外にあり、不確定値（例えば、MAD）が、閾範囲と、不確定値の0%（例えば、重複しない）、5%、10%、20%、25%、30%、35%、または40%未満重複する場合に、候補セグメントを検証する。一部の実施形態では、中央値、平均値、または平均によるZスコアが、閾範囲外にあり、不確定値（例えば、MAD）が、閾範囲と、不確定値の約25%、30%、40%、50%、60%、または約70%超重複する場合に、候補セグメントを検証する。

20

30

40

50

【0466】

一部の実施形態では、分布を、スライディングエッジ処理により生成された全ての個別セグメントについて決定された有意性のレベル（例えば、Zスコア）について生成する（例えば、図13～14を参照されたい）。ある特定の実施形態では、中央値、平均値、または平均による有意性のレベルおよび／または有意性のレベルの分布に従って、個別セグメントを検証するか、または無効にする。一部の実施形態では、約50%、60%、70%、75%、80%、85%、90%、または約95%またはそれ超の分布が、有意性のレベルについて所定の範囲外にあれば、個別セグメントを検証する。例えば、Zスコアの3.0～-3.0の所定の範囲について、検証された候補セグメントは中央値のZスコアを有する可能性があり、Zスコアの分布のうちの70%またはそれ超は、絶対値が3.0超でありうる。

10

【0467】

リーブワンアウトによる検証

一部の実施形態では、検証は、「リーブワンアウト」処理を含む。適切な「リーブワンアウト」処理を使用することができる。一部の実施形態では、「リーブワンアウト」処理により、参照試料のえり抜きのセットと関連する信頼性レベルを提示する。一部の実施形態では、「リーブワンアウト」処理により、参照試料のえり抜きのセットと関連する不確定性のレベルを提示する。一部の実施形態では、「リーブワンアウト」処理により、参照試料のえり抜きのセットに従って決定された、信頼性レベルおよび／または不確定性のレベルに従って、候補セグメントを検証するかまたは無効にする。

20

【0468】

一部の実施形態では、「リーブワンアウト」処理を、試験試料および2つまたはそれ超の参照試料（例えば、本明細書では、場合によって、元のセットと称する、参照試料のセット）について実施する。一部の実施形態では、試験試料を、2つまたはそれ超の参照試料のうちの1つとして組み込む。一部の実施形態では、試験試料を、2つまたはそれ超の参照試料のうちの1つとして組み込まない。一部の実施形態では、「リーブワンアウト」処理は、2つまたはそれ超の参照試料のうちの1つを、元の試料セットから除外し、これにより、参照試料のサブセットをもたらすことを含む。ある特定の実施形態では、参照試料を元のセットから除外する処理を、セット中の各参照試料について反復する。参照試料を元のセットから除外する場合、既に除外された参照試料があれば、これを元のセットへと戻すことが多い。一部の実施形態では、1つだけの参照試料を、任意の1つのサブセットから除外する。結果は、複数の参照試料のサブセット（本明細書では、場合によって、試料の複数のサブセットと称する）であることが多く、ここで、各サブセットは、参照試料のうちの1つを元のセットから逸失している。

30

【0469】

ある特定の実施形態では、「リーブワンアウト」処理は、参照試料のサブセットの各サブセットに従って、有意性のレベルを決定することを含む。ある特定の実施形態では、次いで、平均値、平均、または中央値による有意性のレベルを、サブセットの全てについて決定された有意性のレベル値から計算する。一部の実施形態では、不確定性のレベル（例えば、MAD）を、平均値、平均、または中央値による有意性のレベルに従って計算する。一部の実施形態では、「リーブワンアウト」処理に従って生成された、中央値、平均値、または平均による有意性のレベルおよび／または不確定性のレベルに従って、個別セグメントを検証するか、または無効にする。

40

【0470】

「リーブワンアウト」処理のある特定の実施形態では、有意性のレベルは、Zスコアまたはp値である。一部の実施形態では、「リーブワンアウト」処理についてのZスコアを、以下の式：

$$Z_i = (E_i - \text{Med. } E_{(n)}) / \text{MAD}$$

〔式中、 E_i は、セグメントiのレベルの定量的決定であり、 $\text{Med. } E_{(n)}$ は、参照試料のサブセットのセグメントiについての中央値レベルであり、MADは、 $\text{Med. } E$

50

(n) についての中央値絶対偏差であり、 Z_i は、セグメント i について結果として得られる Z スコアである] に従って計算する。一部の実施形態では、 MAD は、任意の適切な不確定性の尺度で置きかえることができる。一部の実施形態では、 E_i は、その非限定的な例が、部分についてのカウントの、中央値レベル、平均レベル、平均値レベル、合計などを含む、レベルの任意の適切な尺度である。

【0471】

一部の実施形態では、検証は、「スライディングエッジ」処理および「リーブワンアウト」処理を含む。例えば、一部の実施形態では、参照試料のサブセット（例えば、「リーブワンアウト」処理から生成された）は、「スライディングエッジ処理」により生成された参照試料のセットから生成される。例えば、所与の試験試料について、「スライディングエッジ」処理により、セグメント化処理から同定された個別セグメントについて、225のセグメントをもたらすことができ、次いで、10の参照試料のセットを使用して、「リーブワンアウト」処理を実施する。上記の例では、複合中央値、複合平均値、または複合平均による有意性のレベル（例えば、複合中央値による Z スコア）および複合不確定性レベル（例えば、複合 MAD ）を、結果として得られる2250の Z スコアから計算する。一部の実施形態では、複合中央値による有意性のレベル（例えば、複合中央値による Z スコア）および／または複合不確定性レベル（例えば、複合 MAD ）に従って、セグメント化処理により同定された個別セグメントを検証するか、または無効にする。

【0472】

一部の実施形態では、決定分析は、候補セグメント（例えば、複合候補セグメント）についての Z スコアまたは複合 Z スコアに従って、染色体異数性、微小重複、または微小欠失の存在または非存在を決定することを含む。一部の実施形態では、候補セグメントは、トリソミーを指し示し、候補セグメントは、全染色体を表示する部分のセットについての候補セグメントである。ある特定の実施形態では、候補セグメントは、全染色体を表示する部分のセットについての絶対 Z スコアが、所定の値超もしくはそれに等しいか、または所定の閾超もしくはそれに等しい場合に、全染色体異数性を指し示す（例えば、図7を参照されたい）。ある特定の実施形態では、候補セグメントは、全染色体を表示する部分のセットについての絶対 Z スコアが、約2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.75、3.8、3.85、3.9、3.95、4.0、4.05、4.1、4.15、4.2、4.3、4.4、または約4.5の所定の値超またはそれに等しい場合に、全染色体異数性を指し示す。ある特定の実施形態では、候補セグメントは、全染色体を表示する部分のセットについての絶対 Z スコアが、3.95超またはそれに等しい場合に、トリソミーを指し示す。ある特定の実施形態では、候補セグメントは、全染色体を表示する部分のセットについての絶対 Z スコアが、(i) ハールウェーブレット分解処理に従って同定された個別セグメント、または(ii) CBS処理に従って同定された個別セグメントについて決定された Z スコアの絶対値超またはそれに等しい場合に、トリソミーを指し示す。ある特定の実施形態では、候補セグメントは、全染色体を表示する部分のセットについての絶対 Z スコアが、(i) ハールウェーブレット分解処理に従って同定された個別セグメント、または(ii) CBS処理に従って同定された個別セグメントについて決定された複数の Z スコアの絶対値超またはそれに等しい場合に、トリソミーを指し示す。一部の実施形態では、複数の Z スコアの絶対値は、約0.4、0.5、0.6、0.7、0.8、または約0.9を乗算された Z スコアの絶対値である。

【0473】

ある特定の実施形態では、候補セグメント（例えば、有意な候補セグメント）は、全染色体を表示する部分のセットについての絶対 Z スコアが、3.95超またはそれに等しく、(i) ハールウェーブレット分解処理に従って同定された個別セグメント、または(ii) CBS処理に従って同定された個別セグメントについて決定された Z スコアの絶対値超またはそれに等しい場合に、トリソミーを指し示す。ある特定の実施形態では、候補セグメントは、全染色体を表示する部分のセットについての絶対 Z スコアが、3.95超ま

たはそれに等しく、(i) ハールウェーブレット分解処理に従って同定された個別セグメント、または(ii) CBS 処理に従って同定された個別セグメントについて決定された複数の Z スコアの絶対値超またはそれに等しい場合に、トリソミーを指し示す。一部の実施形態では、複数の Z スコアの絶対値は、約 0.4、0.5、0.6、0.7、0.8、または約 0.9 を乗算された Z スコアの絶対値である。

【0474】

一部の実施形態では、候補セグメントは、(i) ハールウェーブレット分解処理に従って同定された個別セグメント、および(ii) CBS 処理に従って同定された個別セグメントについて決定された Z スコアの絶対値が、約 2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.75、3.8、3.85、3.9、3.95、4.0、4.05、4.1、4.15、4.2、4.3、4.4、もしくは約 4.5 超、またはそれに等しい場合に、トリソミーを指し示さず、微小欠失または微小重複の存在を決定する。一部の実施形態では、候補セグメントは、トリソミーを指し示さず、微小欠失または微小重複の存在を決定する。一部の実施形態では、候補セグメントは、(i) ハールウェーブレット分解処理に従って同定された個別セグメント、および(ii) CBS 処理に従って同定された個別セグメントについて決定された Z スコアの絶対値が、3.95 超またはそれに等しい場合に、トリソミーを指し示さず、微小欠失または微小重複の存在を決定する。一部の実施形態では、候補セグメントは、トリソミーを指し示さず、微小欠失または微小重複の存在を決定し、ハールウェーブレット分解処理に従って同定された個別セグメントは、CBS 処理に従って同定された個別セグメントと実質的に同じである。

【0475】

一部の実施形態では、アウトカムを決定すること(determining)(例えば、胎児における、例えば、遺伝子の変動の存在または非存在を決定すること(determining))は、決定(decision)分析を含む。一部の実施形態では、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または非存在を決定する(determining)方法であって、偽陰性の決定が低減され、偽陽性の決定が低減された方法は、決定(decision)分析を含む。一部の実施形態では、決定分析は、一連の方法または方法ステップを含む。決定分析の非限定的な例は、図 6～8 に示され、本明細書で記載される。

【0476】

アウトカムの使用

遺伝子の変動の存在または非存在の決定因の 1 つまたは複数のアウトカムを含むレポートを受け取る医療従事者または他の有資格者は、レポート内に示されたデータを使用して、試験被験体または患者の状態についての判定を下すことができる。一部の実施形態では、医療従事者は、提示されたアウトカムに基づき、推奨を行うことができる。一部の実施形態では、医療従事者または有資格者は、レポートで提示された、1 つまたは複数のアウトカム値および関連する信頼性パラメータに基づき、試験被験体または患者に、遺伝子の変動の存在または非存在に関する判定またはスコアを提示することができる。ある特定の実施形態では、提示されたレポートの目視観察を使用して、医療従事者または有資格者が、手作業でスコアを作成するかまたは判定を下す。ある特定の実施形態では、場合によって、ソフトウェア内に埋め込まれた自動式のルーチンにより、スコアを作成するかまたは判定を下し、試験被験体または患者へと情報を提供する前に、医療従事者または有資格者が、精度について精査する。本明細書で使用される「レポートを受け取ること」という用語は、精査されると、医療従事者または他の有資格者が、試験被験体または患者における遺伝子の変動の存在または非存在について決定することを可能とする、アウトカムを含む通信手段、書面表示、および／またはグラフ表示により得ることを指す。レポートは、コンピュータにより作成することもでき、手作業によるデータ入力により作成することもでき、電子的手段(例えば、インターネットを介する、コンピュータを介する、ファックスを介する、同じ物理的施設または異なる物理的施設における 1 つのネットワーク拠点から

別の拠点への)を使用して通信することもでき、データを送付または受領する別の方法(例えば、郵便、宅急便(登録商標)など)により通信することもできる。一部の実施形態では、アウトカムは、限定せずに述べると、言語形態、文書形態、またはファイル形態を含む適切な媒体により、医療従事者へと伝送する。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、書類ファイル、検査室ファイル、または医療記録ファイルでありうるがこれらに限定されない。

【0477】

本明細書で使用される、「アウトカムを提示すること」という用語およびその文法的な同等物はまた、このような情報を得るための方法であって、限定せずに述べると、情報を検査室から得る(例えば、検査室ファイル)ステップを含む方法も指す場合がある。検査室ファイルは、医学的状态の存在または非存在を決定するための、1つまたは複数のアッセイまたは1つまたは複数のデータ処理ステップを実行した検査室により作成することができる。検査室は、医学的状态の存在または非存在を検査室ファイルから確認する職員と同じ場所にある場合もあり、異なる場所(例えば、別の国)にある場合もある。例えば、検査室ファイルは、1つの場所で作成し、その中の情報が妊娠中の雌被験体へと伝送される別の場所へと伝送することができる。ある特定の実施形態では、検査室ファイルは、実体的形態(tangible form)の場合もあり、電子的形態(例えば、コンピュータ可読形態)の場合もある。

【0478】

一部の実施形態では、アウトカムは、検査室から、医療従事者、医師、または有資格者へと提示することができ、医療従事者、医師、または有資格者は、アウトカムに基づき、診断を下すことができる。一部の実施形態では、アウトカムは、検査室から、医療従事者、医師、または有資格者へと提示することができ、医療従事者、医師、または有資格者は、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共に、アウトカムに一部分基づき、診断を下すことができる。

【0479】

医療従事者または有資格者は、レポートで提示された1つまたは複数のアウトカムに基づき、適切な推奨を提示することができる。提示されたアウトカムレポートに基づき提示されうる、推奨の非限定的な例は、手術、放射線療法、化学療法、遺伝子カウンセリング、生後処置解決手段(after birth treatment solutions)(例えば、人生設計、長期にわたる介護ケア、医薬、対症的処置)、妊娠中絶、臓器移植、輸血など、または前出の組合せを含む。一部の実施形態では、推奨は、提示されたアウトカムベースの分類(例えば、ダウン症候群、ターナー症候群、T13における遺伝子の変動と関連する医学的状态、T18における遺伝子の変動と関連する医学的状态)に依存する。

【0480】

検査室関係者(例えば、検査室管理者)は、遺伝子の変動の存在または非存在の決定(または試験領域についての正倍数体もしくは非正倍数体の決定)の根底をなす値(例えば、試験カウント、参照カウント、偏差のレベル)を分析することができる。遺伝子の変動の存在または非存在に関する判定であって、微妙であるかまたは疑わしい判定について、検査室関係者は、同じ試験を再発注することもでき、かつ/または試験被験体に由来する同じ試料核酸または異なる試料核酸を使用する、異なる試験(例えば、胎児の異数性の決定の場合における核型分析および/または羊水穿刺)を発注することもできる。

遺伝子の変動および医学的状态

【0481】

遺伝子の変動(genetic variance)の存在または非存在は、本明細書に記載する方法または装置を使用して決定することができる。ある特定の実施形態では、1つまたは複数の遺伝子の変動(genetic variation)の存在または非存在は、本明細書に記載する方法および装置により提供されるアウトカムにより決定される。遺伝子の変動は、一般的に、ある特定の個体中に存在する特定の遺伝的表現型であり

、多くの場合、遺伝子の変動は、個体の統計的に有意な部分母集団の中に存在する。一部の実施形態では、遺伝子の変動は、染色体異常（例えば、異数性）、部分的染色体異常、またはモザイク現象であり、そのそれぞれについて、本明細書でより詳細に記載する。遺伝子の変動の非限定的な例として、1つまたは複数の欠失（例えば、微小欠失）、重複（例えば、微小重複）、挿入、変異、多型（例えば、一塩基多型）、融合、リピート（例えば、短いタンデムリピート）、異なるメチル化部位、異なるメチル化パターン等、およびその組合せが挙げられる。挿入、リピート、欠失、重複、変異、または多型は、任意の長さのものであり得、一部の実施形態では、長さ約1塩基または塩基対（bp）～約250メガ塩基（Mb）である。一部の実施形態では、挿入、リピート、欠失、重複、変異、または多型は、長さ約1塩基または塩基対（bp）～約1,000キロ塩基（kb）である（例えば、長さ約10bp、50bp、100bp、500bp、1kb、5kb、10kb、50kb、100kb、500kb、または1000kb）。

10

【0482】

遺伝子の変動は、欠失の場合もある。ある特定の実施形態では、欠失は染色体またはDNA配列の一部分が欠損している変異である（例えば、遺伝子異常）。欠失は、多くの場合、遺伝物質の喪失である。任意の数のヌクレオチドが欠失し得る。欠失は、1つもしくは複数の染色体全体、染色体のセグメント、対立遺伝子、遺伝子、イントロン、エクソン、任意の非コード領域、任意のコード領域、そのセグメント、またはその組合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。

【0483】

20

遺伝子の変動は、遺伝子の重複の場合もある。ある特定の実施形態では、重複は染色体またはDNA配列の一部分がコピーされ、ゲノムへと挿入される変異（例えば、遺伝子異常）である。ある特定の実施形態では、遺伝子の重複（すなわち、重複）は、DNA領域の任意の重複である。一部の実施形態では、重複は、ゲノムまたは染色体内の、多くの場合タンデムに反復した核酸配列である。一部の実施形態では、重複は、1つもしくは複数の染色体全体、染色体のセグメント、対立遺伝子、遺伝子、イントロン、エクソン、任意の非コード領域、任意のコード領域、そのセグメント、またはその組み合わせのコピーを含み得る。重複は、微小重複を含み得る。重複は、1つまたは複数の重複した核酸のコピーを含む場合もある。重複は、1回または複数回反復した（例えば、1、2、3、4、5、6、7、8、9、または10回反復した）遺伝子領域として特徴付けられる場合もある。重複は、小領域（数千塩基対）から一部の事例では染色体全体の範囲であり得る。重複は、相同組換えにおける誤差の結果として、またはレトロトランスポゾンイベントに起因して高頻度で生ずる。重複は、ある特定の種類の増殖性疾患と関連していた。重複は、ゲノムマイクロアレイまたは比較遺伝子交雑法（CGH）を使用して特徴付けできる。

30

【0484】

遺伝子の変動は、挿入の場合もある。挿入は、1つまたは複数のヌクレオチド塩基対の核酸配列への付加の場合もある。挿入は、微小挿入の場合もある。ある特定の実施形態では、挿入は、染色体のセグメントのゲノム、染色体、またはそのセグメントへの付加を含む。ある特定の実施形態では、挿入は、対立遺伝子、遺伝子、イントロン、エクソン、任意の非コード領域、任意のコード領域、そのセグメントまたはその組合せの、ゲノムまたはそのセグメントへの付加を含む。ある特定の実施形態では、挿入は、起源が不明の核酸の、ゲノム、染色体、またはそのセグメントへの付加（すなわち、挿入）を含む。ある特定の実施形態では、挿入は、単一塩基の付加（すなわち、挿入）を含む。

40

【0485】

本明細書で使用する場合、「コピー数の変動」は、一般的に遺伝子の変動または染色体異常のクラスまたは種類である。コピー数の変動は、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。多くの場合、本明細書で時に使用される接頭辞「微小」は、長さ5Mb未満の核酸のセグメントである。コピー数の変動は、染色体のセグメントの1つまたは複数の欠失（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

50

では、重複は挿入を含む。ある特定の実施形態では、挿入は重複である。ある特定の実施形態では、挿入は重複ではない。例えば、多くの場合、ある部分で配列が重複すると、重複が見出される部分に関するカウントが増加する。多くの場合、ある部分で配列が重複するとレベルが高まる。特定の実施形態では、第1のレベルを構成する部分に重複が存在すると、重複が存在しない第2のレベルと比較してレベルが高まる。ある特定の実施形態では、挿入は、部分のカウントを増加させ、挿入を表す配列が同一部分内の別の位置に存在する（すなわち、重複される）。ある特定の実施形態では、挿入は、部分のカウント、またはレベルを有意に増加させず、挿入された配列は、同一部分内の配列の重複ではない。ある特定の実施形態では、挿入は重複として検出または表示されず、挿入を表す重複配列は、同一部分に存在しない。

10

【0486】

一部の実施形態では、コピー数の変動は、胎児のコピー数の変動である。多くの場合、胎児のコピー数の変動は、胎児のゲノム内のコピー数の変動である。一部の実施形態では、コピー数の変動は、母体および／または胎児のコピー数の変動である。ある特定の実施形態では、母体および／または胎児のコピー数の変動は、妊娠中の雌（例えば、胎児を出産する雌の被験体）、分娩経験のある雌の被験体、または胎児を出産する能力を有する雌のゲノム内のコピー数の変動である。コピー数の変動は、ヘテロ接合性のコピー数の変動であり得、この場合、変動（例えば、重複または欠失）は、ゲノムの1方の対立遺伝子上に存在する。コピー数の変動は、ホモ接合性のコピー数の変動であり得、この場合、変動は、ゲノムの両方の対立遺伝子に存在する。一部の実施形態では、コピー数の変動はヘテロ接合性またはホモ接合性の胎児のコピー数の変動である。一部の実施形態では、コピー数の変動は、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の変動である。コピー数の変動は、母体ゲノムおよび胎児ゲノムに存在する、母体ゲノムに存在するが胎児ゲノムに存在しない、または胎児ゲノムに存在するが母体ゲノムに存在しない場合がある。

20

【0487】

「倍数性」とは、胎児または母親中に存在する染色体の数への言及である。ある特定の実施形態では、「倍数性」は、「染色体倍数性」と同じである。ヒトでは、例えば常染色体は、多くの場合、対で存在する。例えば、遺伝子の変動が存在しない場合、ほとんどのヒトは各常染色体（例えば、第1～22染色体）を2つ有する。ヒトにおける2つの常染色体の正常な相補の存在は、これは多くの場合、正倍数体と呼ばれる。「微小倍数性」は、意味上では、倍数性に類似する。「微小倍数性」は、多くの場合、染色体のセグメントの倍数性を指す。用語「微小倍数性」とは、染色体内のコピー数の変動（例えば、欠失、重複、および／または挿入）の存在または非存在（例えば、ホモ接合性またはヘテロ接合性の欠失、重複、または挿入等またはその非存在）への言及の場合もある。「倍数性」および「微小倍数性」は、プロファイル内のレベルのカウントを正規化した後に決定される場合もある。したがって、常染染色体の対を表すレベル（例えば、正倍数体）は、多くの場合、倍数性1に正規化される。同様に、重複、欠失、または挿入が存在しないことを表す染色体のセグメント内のレベルは、多くの場合、微小倍数性1に正規化される。倍数性および微小倍数性は、多くの場合、部分—特異的（例えば、部分特異的）および試料—特異的である。倍数性は、多くの場合、1/2の整数倍として規定され、正倍数体（例えば、2つの染色体）、染色体1つ存在（例えば、染色体欠失）、染色体非存在、染色体3つ（例えば、トリソミー）、および染色体4つをそれぞれ表す、1、1/2、0、3/2、および2の値を有する。同様に、微小倍数性は、多くの場合、1/2の整数倍として規定され、正倍数体（例えば、コピー数の変動無し）、ヘテロ接合性の欠失、ホモ接合性の欠失、ヘテロ接合性の重複、およびホモ接合性の重複をそれぞれ表す、1、1/2、0、3/2、および2の値を有する。胎児に関する倍数性値についての一部の例を表2に提示する。

30

40

【0488】

ある特定の実施形態では、胎児の微小倍数性は、胎児の母親（すなわち、妊娠中の雌の

50

被験体)の微小倍数性と一致する。ある特定の実施形態では、胎児の微小倍数性は、胎児の母親の微小倍数性と一致し、母親および胎児いずれも、同一のヘテロ接合性のコピー数の変動、ホモ接合性のコピー数の変動を担持する、または両方とも正倍数体である。ある特定の実施形態では、胎児の微小倍数性は、胎児の母親の微小倍数性と異なる。例えば、胎児の微小倍数性は、コピー数の変動についてヘテロ接合性であり、母親は、コピー数の変動についてホモ接合性であり、胎児の微小倍数性は、特定のコピー数の変動に関して母親の微小倍数性と一致しない(例えば、等しくない)場合もある。

【0489】

微小倍数性は、多くの場合、期待されるレベルと関連する。例えば、レベル(例えば、プロファイル内のレベル、時にコピー数の変動を実質的に含まないレベル)は、値1に正

10

【0490】

被験体について存在または非存在が同定された遺伝子の変動は、ある特定の実施形態では医学的状态と関連する。したがって、本明細書に記載する技術は、医学的状态または病状と関連する1つまたは複数の遺伝子の変動の存在または非存在を同定するのに使用することができる。医学的状态の非限定的な例として、知的障害(例えば、ダウン症候群)、異常な細胞増殖(例えば、がん)、微生物核酸(例えば、ウイルス、細菌、真菌、酵母)の存在、および子癰前症と関連した状態が挙げられる。

20

【0491】

遺伝子の変動、医学的状态および病状の非限定的な例は、以下に記載されている。

胎児の性別

【0492】

一部の実施形態では、胎児の性別または性別関連の障害(例えば、性染色体異数性)の予測は、本明細書に記載する方法または装置により決定することができる。性別の決定は、性染色体に一般的に基づく。ヒトでは、2つの性染色体、X染色体およびY染色体が存在する。Y染色体は、雄としての胚発生を引き起こす遺伝子、SR Yを含有する。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とされる他の遺伝子も含有する。XXを有する個体は雌であり、XYは雄であり、多くの場合、性染色体異数性と呼ばれる非限定的な変動として、X0、XYY、XXX、およびXXYが挙げられる。ある特定の実施形態では、雄は、2つのX染色体および1つのY染色体(XXY; クラインフェルター症候群)、または1つのX染色体および2つのY染色体(XYY症候群; ジェイコブス症候群)を有し、ならびに一部の雌は、3つのX染色体(XXX; トリプルX症候群)または2つではなく単一の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では、個体内の一部の細胞のみが、性染色体異数性により影響を受け、モザイク現象(例えば、ターナーモザイク現象)と呼ばれる場合もある。他の症例として、SR Yが損傷を受けている症例(XYの雌となる)、またはXにコピーされた症例(XXの雄となる)が挙げられる。

30

【0493】

ある特定の症例では、子宮内の胎児の性別を決定することが有益な場合もある。例えば、1つまたは複数の伴性障害の家族歴を有する患者(例えば、妊娠中の雌)は、かかる障害を受け継ぐ胎児のリスクを評価するのに役立つように、身ごもっている胎児の性別を決定したいと欲する場合がある。伴性障害として、非限定的に、X連鎖およびY連鎖障害が挙げられる。X連鎖障害として、X連鎖劣性障害およびX連鎖優性障害が挙げられる。X関連劣性障害の例として、非限定的に、免疫障害(例えば、慢性肉芽腫性疾患(CYBB)、ヴィスコット・アルドリッチ症候群、X連鎖重症複合型免疫不全症、X連鎖無ガンマグロブリン血症、1型高IgM症候群、IPEX、X連鎖リンパ増殖性疾患、プロパーン欠損症)、血液学的障害(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

40

50

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/I I b型グリコーゲン蓄積症、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/オクシピタル・ホーン症候群)、神経系障害(例えば、コフィンローリー症候群、MA S A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X連鎖精神遅滞症候群、色盲、眼球白皮症、ノリエ病、コロイデレミア、シャルコー-マリー-トゥース病(C M T X 2-3)、ペリツェウス-メルツバッハー病、S M A X 2)、皮膚および関連組織の障害(例えば、先天性角化不全症、低汗性外胚葉形成不全(E D A)、X連鎖魚鱗癬、X連鎖角膜内皮ジストロフィー)、神経筋障害(例えば、ベッカー型筋ジストロフィー/デュシェンヌ型筋ジストロフィー、中心核ミオパシー(M T M 1)、コンラディー-ヒューネルマン症候群、エメリー-ドレフ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連鎖腎原性尿崩症)、骨/歯の障害(例えば、A M E L Xエナメル質形成不全症)、および他の障害(例えば、バース症候群、マクロード症候群、スミス-ファインマン-マイヤーズ症候群、シンプソン-グラビー-ベーム症候群、M o h r -T r a n e b j a e r g症候群、鼻指聴覚症候群)。X連鎖優性障害の例として、非限定的に、X連鎖低リン酸血症、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、R e t t症候群、C H I L D症候群、L u j a n -F r y n s症候群、および口腔・顔面・指趾症候群1が挙げられる。Y連鎖障害の例として、非限定的に、雄不妊症、網膜色素変性、および無精子症が挙げられる。

10

20

染色体異常

【0494】

一部の実施形態では、胎児染色体異常の存在または非存在は、本明細書に記載する方法または装置を使用して決定することができる。染色体異常として、非限定的に、染色体全体または1つもしくは複数の遺伝子を含む染色体の領域の取得または喪失が挙げられる。染色体異常には、モノソミー、トリソミー、ポリソミー、ヘテロ接合性の喪失、転座、不均衡な転座により引き起こされた欠失および重複を含む、1つまたは複数のヌクレオチド配列(例えば、1つまたは複数の遺伝子)の欠失および/または重複が含まれる。用語「染色体異常」または「異数性」は、本明細書で使用する場合、被験体の染色体構造と正常な相同染色体構造の間の乖離を指す。用語「正常」とは、特定の種の健康な個体に見出される優勢な核型またはバンディングパターン、例えば正倍数体ゲノム(ヒトでは、46、X Xまたは46、X Y)を指す。生物が異なれば染色体の相補性も幅広く異なるので、用語「異数性」は特定の染色体の数を指すものではなく、生物の所与の細胞の1つまたは複数内の染色体含有量が異常である状況を指す。一部の実施形態では、用語「異数性」は、本明細書では、染色体の全部または染色体の一部の喪失または取得により引き起こされた遺伝物質の不均衡を指す。「異数性」は、染色体のセグメントの1つまたは複数の欠失および/または挿入を指し得る。用語「正倍数体」は、一部の実施形態では、染色体の正常な相補を指す。

30

【0495】

用語「モノソミー」は、本明細書で使用する場合、正常な相補の1つの染色体が欠如していることを指す。単一のコピー内に染色体のセグメントのみが存在する、不均衡な転座または欠失においては、部分的モノソミーが生じ得る。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」は、染色体のコピーが2つ存在することを指す。各染色体の2つのコピーを有するヒト等の生物(二倍体または「正倍数体」の生物)の場合、ダイソミーは正常な状態である。各染色体の3つまたはそれ超のコピーを通常有する生物(三倍体またはそれ超の生物)の場合、ダイソミーは異数体の染色体の状態である。片親性のダイソミーでは、染色体の両方のコピーは同一の親に由来する(他方の親の寄与はない)。

40

【0496】

用語「トリソミー」は、本明細書で使用する場合、特定の染色体の2つのコピーではな

50

く3つのコピーが存在することを指す。ヒトのダウン症候群に見出される余分な第21染色体の存在は、「トリソミー21」と呼ばれる。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、雌（例えば、トリプルX症候群の47、XXX）または雄（例えば、クラインフェルター症候群の47、XXY；またはジェイコブス症候群の47、XYY）に認められる場合がある。一部の実施形態では、トリソミーは、ほとんどまたは全ての常染色体の重複である。ある特定の実施形態では、トリソミーは全染色体異数性であり、特定の種類の染色体について3つのインスタンス（例えば、3つのコピー）をもたらす（例えば、正倍数体についての特定の種類の染色体の2つのインスタンス（すなわち対）ではなく）。

【0497】

用語「テトラソミー」および「ペンタソミー」は、本明細書で使用する場合、4つまたは5つの染色体のコピーがそれぞれ存在することを指す。常染色体ではほとんど認められないが、性染色体のテトラソミーおよびペンタソミーが、XXXX、XXX Y、XXYY、XYYY、XXXXX、XXXXY、XXXY Y、XXYYY、およびXYYYYを含め、ヒトで報告されている。

【0498】

染色体異常は、様々な機構により引き起こされ得る。機構には、(i) 有糸分裂チェックポイントが脆弱化した結果として生ずる不分離、(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つの動原体が両方の有糸分裂紡錘体極に結合したときに生ずるメロテリック結合、(iv) 2つ超の紡錘体極が形成されたときの多極紡錘体形成、(v) 単一の紡錘体極しか形成されなかったときの単極紡錘体形成、および(vi) 単極紡錘体機構の最終結果として四倍体中間体が生ずることが含まれるが、これらに限定されない。

【0499】

用語「部分的モノソミー」および「部分的トリソミー」は、本明細書で使用する場合、染色体の一部分の喪失または取得により引き起こされた遺伝物質の不均衡を指す。部分的モノソミーまたは部分的トリソミーは、不均衡な転座に起因し得るが、この場合、個体は2つの異なる染色体の破断および融合により形成された誘導染色体を担持する。この状況では、個体は1つの染色体の一部分の3つのコピー（2つの正常なコピー、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与する他の染色体の一部分の1つのコピーのみを有する。

【0500】

用語「モザイク現象」は、本明細書で使用する場合、生物の全ての細胞ではなく、一部の細胞内の染色体異数性を指す。ある特定の染色体異常は、モザイク性および非モザイク性の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。異なる機構が、モザイク現象をもたらし得る。例えば、(i) 最初の接合体は、3つの第21染色体を有すると考えられ、これは単純なトリソミー21を通常もたすが、細胞分裂の過程で、1つまたは複数の細胞系が、第21染色体の1つを喪失する；および(ii) 最初の接合体は、2つの第21染色体を有すると考えられるが、細胞分裂の過程で、第21染色体の1つが重複した。体細胞モザイク現象は、完全なまたはモザイク性の異数性を伴う遺伝的症候群と一般的に関連する機構とは異なる機構を通じて生ずる可能性がある。体細胞モザイク現象は、例えばある特定の種類のがんやニューロンにおいて同定された。ある特定の事例では、トリソミー12は、慢性リンパ球性白血病（CLL）において同定され、トリソミー8は、急性骨髄性白血病（AML）において同定された。また、個体が染色体の破断しやすい傾向を有するような遺伝的症候群（染色体不安定症候群）では、様々な種類のがんに対するリスクの増大と高頻度に関連し、したがって発癌性における体細胞異数性の役割が注目される。本明細書に記載する方法およびプロトコールは、非モザイク性およびモザイク性の染色体異常の存在または非存在を同定することができる。

【0501】

表 1 A および 1 B は、本明細書に記載する方法および装置により同定される可能性があり得る染色体の状態、症候群、および／または異常の非限定的なリストを提示する。表 1 B は、2011 年 10 月 6 日時点の DECIPHER データベースに由来する（例えば、バージョン 5.1、GRCh37 に対してマッピングされた場所に基づく；ユニフォームリソースロケーター（URL）`dechipher.sanger.ac.uk`にて入手可能）。

【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	YY 症候群
Y	XXX	トリソミーX 症候群
Y	XXXX	4X 症候群
Y	Xp21 欠失	デュシェンヌ型筋ジストロフィー/ベッカー症候群、先天性副腎低形成、慢性肉芽腫性 (granulomatous disease) 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 連鎖リンパ増殖性疾患
1	1p(体細胞性) 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、発達・精神遅滞、および軽微な身体的異常
3	モノソミー トリソミー(体細胞性)	非ホジキンリンパ腫
4	モノソミー トリソミー(体細胞性)	急性非リンパ球性白血病(ANLL)
5	5p	猫鳴き病;ルジューヌ症候群
5	5q (体細胞性) 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(体細胞性)	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	小児期のモノソミー7 症候群;体細胞性:腎皮質腺腫; 骨髄異形成症候群

10

20

30

40

【表 1 A - 2】

8	8q24.1 欠失	ランガー-ギデオン症候群
8	モノソミー トリソミー	骨髄異形成症候群;ワーカニー症候群;慢性骨髄性白血病
9	モノソミー-9p	アルフィー症候群
9	モノソミー-9p 部分的トリソミー	レトレ症候群
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 性)	ALL または ANLL
11	11p-	無虹彩;ウィルムス腫瘍
11	11q-	ヤコブセン症候群
11	モノソミー (体細胞性)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 性)	CLL、若年性顆粒膜細胞腫瘍(JGCT)
13	13q-	13q-症候群;オルペリ症候群
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 性)	骨髄系障害 (MDS、ANLL、異型 CML)
15	15q11-q13 欠失 モノソミー	プラダー-ウィリー症候群、アンジェルマン症候群
15	トリソミー(体細胞 性)	骨髄細胞系列およびリンパ球系列が影響を受ける(例え ば、MDS、ANLL、ALL、CLL)
16	16q13.3 欠失	ルーベンシュタイン-タイビ

10

20

30

40

【表 1 A - 3】

3	モノソミー トリソミー(体細胞性)	乳頭状腎細胞癌(悪性)
17	17p-(体細胞性)	骨髄系悪性腫瘍における 17p 症候群
17	17q11.2 欠失	スミス-マゲニス
17	17q13.3	ミラー-ディッカー
17	モノソミー トリソミー(体細胞性)	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 1 型;HNPP
18	18p-	18p 部分的モノソミー症候群またはグロウチー・ラミー・ティーフリイ症候群
18	18q-	グロウチー・ラミー・サルモン・ランドリー症候群
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、真性赤血球増加症、慢性好中性白血病
20	モノソミー トリソミー(体細胞性)	乳頭状腎細胞癌(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、円錐動脈幹異常顔貌症候群、常染色体優性オピッツ G/BBB 症候群、ケイラー心顔症候群
22	モノソミー トリソミー	完全トリソミー22 症候群

【 0 5 0 2 】

10

20

30

40

【表 1 B - 1】

表1B

症候群	染色体	スタート	エンド	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 再発性微小欠失症候群	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 再発性微小欠失(神経認知障害感受性遺伝子座)	16	15,504,454	16,284,248	0.78	
16p13.11 再発性微小重複(神経認知障害感受性遺伝子座)	16	15,504,454	16,284,248	0.78	
17q21.3 再発性微小欠失症候群	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 再発性微小欠失(神経発達障害に関する感受性遺伝子座)	1	146,512,930	147,737,500	1.22	3

10

20

30

40

【表 1 B - 2】

症候群	染色体	スタート	エンド	間隔(Mb)	グレード
1q21.1 再発性微小重複(神経発達障害に関する感受性遺伝子座の可能性)	1	146,512,930	147,737,500	1.22	3
1q21.1 血小板減少-橈骨欠損(TAR)症候群に関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/ディジョージ症候群)	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 遠位欠失症候群	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 モノソミー	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	
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ロイコジストロフィー(ADLD)	5	126,063,045	126,204,952	0.14	

10

20

30

40

【表 1 B - 3】

症候群	染色体	スタート	エンド	間隔(Mb)	グレード
アンジェルマン症候群(1型)	15	22,876,632	28,557,186	5.68	1
アンジェルマン症候群(2型)	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	
ネコ眼症候群(I型)	22	1	16,971,860	16.97	
シャルコー-マリー-トウース症候群 1A型(CMT1A)	17	13,968,607	15,434,038	1.47	1
猫鳴き症候群(5p 欠失)	5	10,001	11,723,854	11.71	1
脳アミロイド血管障害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ジス	5	112,101,596	112,221,377	0.12	
遺伝性圧脆弱性ニューロパチー(HNPP)	17	13,968,607	15,434,038	1.47	1
レリー・ワイル軟骨骨異形成症 (Leri-Weill dyschondroostosis) (LWD)-SHOX 欠失	X	751,878	867,875	0.12	
レリー・ワイル軟骨骨異形成症 (LWD)-SHOX 欠失	X	460,558	753,877	0.29	
ミラー-ディッカー症候群(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

10

20

30

40

【表 1 B - 4】

症候群	染色体	スタート	エンド	間隔(Mb)	グレード
ペリツェウス-メル ツバッハー病	X	102,642,051	103,131,767	0.49	
ポトキ-ルプスキ症 候群(17p11.2 重複 症候群)	17	16,706,021	20,482,061	3.78	
ポトキ-シェーファ ー症候群	11	43,985,277	46,064,560	2.08	1
プラダー-ウィリー 症候群(1 型)	15	22,876,632	28,557,186	5.68	1
プラダー-ウィリー 症候群(2 型)	15	23,758,390	28,557,186	4.80	1
RCAD(腎嚢胞およ び糖尿病)	17	34,907,366	36,076,803	1.17	
ルビンシュタイン- テイビ症候群	16	3,781,464	3,861,246	0.08	1
スミス-マゲニス症 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 / 裂足症 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(STS)	X	6,441,957	8,167,697	1.73	
WAGR11p13 欠失 症候群	11	31,803,509	32,510,988	0.71	
ウィリアムス-ボイ 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	

【 0 5 0 3 】

グレード 1 の状態は、多くの場合、1 つまたは複数の以下の特徴を有する；病源的異常；遺伝学者の間で強く合意されている；高い浸透性；なおも多様な表示型を有し得るが、いくつかの一般的な特性も有する；文献中の全ての症例は臨床表示型を有する；異常を有する健康な個体の症例を認めない；D V G データベースに報告されていない、または健常母集団では見出されない；単一遺伝子または多重遺伝子の量的効果を確認する機能的データ；確認済みまたは強固な候補遺伝子；臨床マネジメント案が規定済み；がんのリスクが

10

20

30

40

50

公知でサーベイの案を有する；複数の情報源（OMIM、GeneReviews、Orphanet、Unique、Wikipedia）；および／または診断用途で利用可能（妊娠カウンセリング）。

【0504】

グレード2の状態は、多くの場合、1つまたは複数の下記の特徴を有する；病源的異常の可能性；高い浸透性；DDを除き一貫した特性を有さない多様な表示型；文献では症例／報告の数が少ない；報告された全ての症例は臨床表示型を有する；機能的データまたは確認済みの病原性遺伝子を認めない；複数の情報源（OMIM、GeneReviews、Orphanet、Unique、Wikipedia）；および／または診断目的および妊娠カウンセリングのために使用できる。

10

【0505】

グレード3の状態は、多くの場合、1つまたは複数の下記の特徴を有する；感受性遺伝子座；健常な個体または発端者の未罹患の両親が記載されている；対照母集団中に存在する；非浸透性；表示型が軽度で特異的ではない；特性はあまり一貫していない；機能的データまたは確認済みの病原性遺伝子を認めない；データの供給源がより限定的；大部分から乖離している症例に関して、または新規臨床所見が存在する場合、第2の診断の可能性は、可能性の状態のままである；および／または診断目的で使用する際には要注意、および妊娠カウンセリングの場合、助言には慎重を期す。

子癇前症

【0506】

20

一部の実施形態では、子癇前症の存在または非存在は、本明細書に記載する方法または装置を使用して決定される。子癇前症は、妊娠中に高血圧症が発生する状態（すなわち、妊娠誘発性高血圧症）であり、尿中の相当量のタンパク質と関連する。ある特定の実施形態では、子癇前症は、細胞外核酸のレベル上昇および／またはメチル化パターン変化とも関連する。例えば、細胞外の胎児由来過剰メチル化RASSF1Aレベルと子癇前症の重症度の間に正の相関が認められた。ある特定の例では、子癇前症の胎盤内のH19遺伝子について、正常な対照と比較してDNAのメチル化の増加が認められる。

【0507】

子癇前症は、世界的に、母体および胎児／新生児の死亡率および疾病率の主因の1つである。血漿および血清中の循環無細胞核酸は新規バイオマーカーであり、出生前診断を含む異なる医学分野における臨床用途として有望である。母体血漿中の無細胞胎児（cff）DNAの定量的変化は、例えば雄特異的SRYまたはDYS14遺伝子座に関するリアルタイム定量的PCRを使用して、その変化が切迫した子癇前症に関する指標となること、異なる試験で報告されている。早期発症型の子癇前症の症例では、最初の三半期にレベルの上昇が認められる場合がある。症状発現前のcffDNAのレベルの上昇は、組織の酸化ストレスおよび胎盤のアポトーシスおよび壊死の増加をもたらす絨毛間腔内の低酸素／再酸素化に起因する場合もある。cffDNAの母体循環への排出増加に関する証拠に加えて、子癇前症では、cffDNAの腎臓クリアランスの低下に関する証拠も存在する。胎児DNAの量は、現在のところ、Y-染色体特異的配列の定量により決定されるので、代替的アプローチ、例えば無細胞総DNAの測定または性別に依存しない胎児エピジェネティックマーカー、例えばDNAメチル化の使用により、代替法が提供される。胎盤起源の無細胞RNAは、臨床診療において子癇前症をスクリーニングおよび診断するのに使用できる別の代替的バイオマーカーである。胎児RNAは、これを分解から保護する細胞内胎盤粒子と関連する。胎児のRNAレベルは、対照と比較して子癇前症の妊娠中の雌では10倍高い場合があり、したがって、臨床診療において子癇前症をスクリーニングおよび診断するのに使用できる代替的バイオマーカーである。

30

40

病原体

【0508】

一部の実施形態では、病態の存在または非存在は、本明細書に記載する方法または装置により決定される。病態は、細菌、ウイルス、または真菌を含むが、これらに限定されな

50

い病原体に宿主が感染することにより引き起こされ得る。病原体は宿主の核酸と区別可能な核酸（例えば、ゲノムDNA、ゲノムRNA、mRNA）を一般的に有するので、本明細書において提供される方法および装置が、病原体の存在または非存在を決定するのに使用できる。多くの場合、病原体は、例えばエピジェネティックな状態および／または1つもしくは複数の配列の変動、重複、および／または欠失等の、特定の病原体に固有の特徴を持つ核酸を有する。したがって、本明細書において提供される方法は、特定の病原体または病原体の変異体（例えば、株）を同定するのに使用できる。

がん

【0509】

一部の実施形態では、細胞増殖障害（例えば、がん）の存在または非存在が、本明細書に記載する方法または装置を使用して決定される。例えば、血清中の無細胞核酸のレベルは、健康な患者と比較して様々な種類のがんを有する患者で上昇し得る。例えば、転移性の疾患を有する患者は、非転移性の患者の約2倍高い血清DNAレベルを有する場合があります。転移性の疾患を有する患者は、がん特異的マーカー、および／または、例えばある特定の塩基多型または短いタンデムリピートによっても同定され得る。循環DNAのレベル上昇と正に相関し得るがんの種類の特異的な例として、乳がん、結腸直腸がん、消化器がん、肝細胞がん、肺がん、メラノーマ、非ホジキンリンパ腫、白血病、多発性骨髄腫、膀胱がん、ヘパトーマ、子宮頸がん、食道がん、膵臓がん、および前立腺がんが挙げられる。様々ながんは、非がん性の健康な細胞に由来する核酸から区別可能な特徴、例えばエピジェネティックな状態、ならびに／または配列の変動、重複、および／もしくは欠失等を伴う核酸を有し得る、および、時には、これを血流中に放出し得る。かかる特徴は、例えば特定の種類のがんに特異的であり得る。したがって、本明細書において提供される方法は、特定の種類のがんを同定するのに使用できることがさらに想定される。

【0510】

本明細書において以後より詳細に記載するように、ソフトウェアが、本明細書に記載する処理において、下記を含むが、これに限定されない1つまたは複数のステップを行うために使用できる；カウント計測、データ処理、アウトカムの生成、および／または生成されたアウトカムに基づく1つもしくは複数の推奨の提供。

機械、ソフトウェア、およびインターフェース

【0511】

本明細書に記載するある特定の処理および方法（例えば、定量、マッピング、正規化、範囲の設定、調整、分類、カウント計測、ならびに／または配列の読取り、カウント、レベル（例えば、レベル）、および／もしくはプロファイルの決定）は、多くの場合、コンピュータ、プロセッサ、ソフトウェア、モジュールまたは他の装置なしで行うことができない。本明細書に記載する方法は、一般的にコンピュータが実施する方法であり、方法の1つまたは複数の部分が、1つまたは複数のプロセッサ（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制御式装置により行われる場合がある。本文書に記載されている方法に関連する実施形態は、一般的に、本明細書に記載するシステム、装置、およびコンピュータプログラム産物におけるインストラクションにより実施される同一のまたは関連する処理に適用可能である。一部の実施形態では、本明細書に記載する処理および方法（例えば、定量、カウント計測、ならびに／または配列の読取り、カウント、レベル、および／もしくはプロファイルの決定）は、自動化された方法により行われる。一部の実施形態では、本明細書に記載する1つまたは複数のステップおよび方法は、プロセッサおよび／もしくはコンピュータにより行われる、および／またはメモリと併せて行われる。一部の実施形態では、自動化された方法は、配列の読取り、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、範囲の設定、分類、調整、プロット、アウトカム、変換、および同定を決定するソフトウェア、モジュール、プロセッサ、周辺機器、および／またはそのようなものを含む装置に組み入れる。本明細書で使用する場合、ソフトウェアとは、本明細書に記載するように、プロセッサにより実行されたときにコンピュータの操作を行う、コンピュータ可読なプログ

ラムインストラクションを指す。

【0512】

試験被験体（例えば、患者、妊娠中の雌）に由来する、および／または参照被験体に由来する配列の読取り、カウント、レベル、およびプロファイルは、遺伝子の変動の存在または非存在を決定するためにさらに分析および処理することができる。配列の読取り、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」と呼ばれる場合もある。一部の実施形態では、データまたはデータセットは、1つまたは複数の特性または変数（例えば、配列に基づく〔例えば、GC含有量、特異的ヌクレオチド配列等〕、機能特異的〔例えば、発現した遺伝子、がん遺伝子等〕、位置に基づく〔ゲノム特異的、染色体特異的、部分または部分特異的〕特性または変数等およびその組合せ）により特徴付けることができる。ある特定の実施形態では、データまたはデータセットは、1つまたは複数の特性または変数に基づく2次元またはそれ超の次元を有するマトリックスに組織化され得る。マトリックスに組織化されたデータは、任意の適する特性または変数を使用して組織化され得る。マトリックス中のデータの非限定的な例として、母体の年齢、母体の倍数性、および胎児の寄与により組織化されるデータが挙げられる。ある特定の実施形態では、1つまたは複数の特性または変数により特徴付けられるデータセットは、カウント計測後に処理される場合もある。

10

【0513】

装置、ソフトウェア、およびインターフェースが、本明細書に記載する方法を実施するのに使用できる。装置、ソフトウェア、およびインターフェースを使用して、ユーザーは、特定の情報、プログラム、または処理（例えば、配列の読取りのマッピング、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）を使用するためのオプションを入力、要求、照会、または決定することができ、例えば統計分析アルゴリズム、統計的有意性アルゴリズム、統計的アルゴリズム、反復ステップ、検証アルゴリズム、および図形表示の実施が含まれ得る。一部の実施形態では、データセットは、インプット情報としてユーザーが入力可能であり、ユーザーは、適するハードウェア媒体（例えば、フラッシュドライブ）により1つもしくは複数のデータセットをダウンロードすることができ、ならびに／またはユーザーは、後続する処理のために、および／もしくはアウトカムを提供するために、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することができる（例えば、シーケンサーからコンピュータシステムに、配列の読取りのマッピング用として配列の読取りデータを送信する；マッピングされた配列データを、処理して、ならびにアウトカムおよび／またはレポートの取得用としてコンピュータシステムに送信する）。

20

30

【0514】

システムは、1つまたは複数の装置を一般的に含む。各装置は、1つまたは複数のメモリ、1つまたは複数のプロセッサ、およびインストラクションを含む。システムが2つまたはそれ超の装置を含む場合、装置の一部または全部は同一の場所に位置し得る、装置の一部または全部は異なる場所に位置し得る、全ての装置は1つの場所に位置し得る、および／または全ての装置は異なる場所に位置し得る。システムが2つまたはそれ超の装置を含む場合、装置の一部もしくは全部はユーザーと同じ場所に位置し得る、装置の一部もしくは全部はユーザーと異なる場所に位置し得る、全ての装置はユーザーと同じ場所に位置し得る、および／または全ての装置はユーザーとは異なる1つもしくは複数の場所に位置し得る。

40

【0515】

システムは、演算装置およびシーケンサーを含む場合があり、この場合、シーケンサーは、身体由来の核酸を入手し、配列の読取りを生成するように構成され、演算装置は、シーケンサーから得られた読取りを処理するように構成される。演算装置は、配列の読取りから遺伝子の変動（例えば、コピー数の変動；胎児染色体異数性）の存在または非存在を決定するように構成される場合がある。

【0516】

ユーザーは、例えばソフトウェアに照会を行うことができ、ソフトウェアは、次にイン

50

ターネットにアクセスしてデータセットを取得することができ、ある特定の実施形態では、プログラム可能なプロセッサは、与えられたパラメータに基づいて、適するデータセットを取得するように催促され得る。また、プログラム可能なプロセッサは、与えられたパラメータに基づいてプロセッサにより選択された1つまたは複数の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す場合もある。プログラム可能なプロセッサは、インターネット、他の内部または外部の情報等を経由して見出される情報に基づき、プロセッサにより選択された1つまたは複数の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、1つまたは複数のデータ特性セレクション、1つまたは複数の統計的アルゴリズム、1つまたは複数の統計分析アルゴリズム、1つまたは複数の統計的有意性アルゴリズム、反復ステップ、1つまたは複数の検証アルゴリズム、ならびに方法、装置、またはコンピュータプログラムの1つまたは複数の図形表示を選択するために選ばれ得る。

10

【0517】

本明細書が取り上げるシステムは、コンピュータシステムの一般的なコンポーネント、例えばネットワークサーバー、ラップトップシステム、デスクトップシステム、ハンドヘルドシステム、パーソナルデジタルアシスタント、コンピュータキオスク (computing kiosk) 等を含み得る。コンピュータシステムは、ユーザーがデータをシステムに入力できるようにする1つまたは複数のインプット手段、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、または他の手段を含み得る。システムは、ディスプレイスクリーン (例えば、CRTまたはLCD)、スピーカー、ファックス機、プリンター (例えば、レーザー式、インクジェット式、インパクト式、白黒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アウトプットを提供するのに有用な他のアウトプットを含むが、これらに限定されない、1つまたは複数のアウトプットをさらに含み得る。

20

【0518】

システムでは、インプットおよびアウトプット手段は、コンポーネントの中でもとりわけ、プログラムインストールを実行する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保管するメモリを含み得る中央処理装置と接続され得る。一部の実施形態では、処理は、単一の地理的箇所に所在する単一のユーザーシステムとして実施され得る。ある特定の実施形態では、処理は、マルチユーザーシステムとして実施され得る。マルチユーザーで実施される場合、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。ネットワークは、建物の一部内の一部門、建物全体を範囲に含むようにローカルであり得、複数の建物にまたがり得、1つの領域にまたがり得、国全体にまたがり得、または世界規模であり得る。ネットワークは個人的であり得、プロバイダーにより所有、および管理され得る、またはユーザーが情報を入力および引き出すためにウェブページにアクセスするような、インターネットに基づくサービスとして実施され得る。したがって、ある特定の実施形態では、システムは、ユーザーにとってローカルまたはリモートであり得る1つまたは複数の機械を含む。1つの場所または複数の場所にある1つ超の機械に、ユーザーはアクセスでき、データは、連続しておよび／または並行してマッピングおよび／または処理され得る。したがって、適するコンフィグレーションおよび制御を利用して、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等において、複数の機械を使用してデータをマッピングおよび／または処理することができる。

30

40

【0519】

システムは、一部の実施形態では、コミュニケーションインターフェースを含み得る。コミュニケーションインターフェースは、コンピュータシステムと1つまたは複数の外部デバイス間で、ソフトウェアおよびデータを伝送できるようにする。コミュニケーションインターフェースの非限定的な例として、モデム、ネットワークインターフェース (イーサネットカード等)、コミュニケーションポート、PCMCIAスロットおよびカード等が挙げられる。コミュニケーションインターフェース経由で伝送したソフトウェアお

50

よびデータは、一般的にシグナルの形態を取り、これは、電子シグナル、電磁気シグナル、光学シグナル、および／またはコミュニケーションインターフェースにより受信される他のシグナルであり得る。シグナルは、多くの場合、チャンネルを介してコミュニケーションインターフェースに提供される。チャンネルは、多くの場合、シグナルを担持し、ワイヤーまたはケーブル、光ファイバー、電話線、携帯電話リンク、RFリンク、および／または他のコミュニケーションチャンネルを使用して実施され得る。したがって、1つの例では、コミュニケーションインターフェースは、シグナル検出モジュールにより検出できるシグナル情報を受信するのに使用できる。

【0520】

データは、マニュアルインプットデバイスまたはダイレクトデータ入力デバイス（DDE）を含むが、これらに限定されない、適するデバイスおよび／または方法によりインプットできる。マニュアルデバイスの非限定的な例として、キーボード、コンセプトキーボード、タッチ感応式スクリーン、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、ビデオデジタイザー、および音声認識デバイスが挙げられる。DDEの非限定的な例として、バーコードリーダー、磁気ストリップコード、スマートカード、磁気インク文字認識、光学式文字認識、光学式マーク認識、およびターンアラウンドドキュメントが挙げられる。

【0521】

一部の実施形態では、シーケンサーからのアウトプットは、インプットデバイス経由のインプットとなり得るデータとしての役割を果たすことができる。ある特定の実施形態では、マッピングされた配列の読取りは、インプットデバイス経由のインプットとなり得るデータとしての役割を果たすことができる。ある特定の実施形態では、シミュレートしたデータは、インシリコ処理により生成され、またシミュレートしたデータは、インプットデバイス経由のインプットとなり得るデータとしての役割を果たす。用語「インシリコ」とは、コンピュータを使用して行う研究および実験を指す。インシリコ処理は、本明細書に記載する処理により、配列の読取りをマッピングすること、およびマッピングされた配列の読取りを処理することを含むが、これらに限定されない。

【0522】

システムには、本明細書に記載する処理を行うために有用なソフトウェアを含むことができ、ソフトウェアは、かかる処理を行う1つまたは複数のモジュールを含み得る（例えば、配列決定モジュール、論理処理モジュール、データディスプレイ組織化モジュール）。用語「ソフトウェア」は、コンピュータにより実行されると、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インストラクションを指す。1つまたは複数のプロセッサにより実行可能なインストラクションは、実行されると、1つまたは複数のプロセッサに本明細書に記載する方法を実施させることができる実行可能なコードとして提供される場合もある。本明細書に記載するモジュールは、ソフトウェアとして存在し得、ソフトウェアに組み入れたインストラクション（例えば、プロセス、ルーチン、サブルーチン）が、プロセッサにより実施または行われ得る。例えば、モジュール（例えば、ソフトウェアモジュール）は、特定の処理またはタスクを行うプログラムの一部分であり得る。用語「モジュール」は、より大型の装置またはソフトウェアシステムで使用できる自己内蔵機能ユニットを指す。モジュールは、モジュールの機能を実施する一連のインストラクションを含み得る。モジュールは、データおよび／または情報を変換することができる。データおよび／または情報は、適する形態であり得る。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では、データおよび／または情報は、パケット、バイト、符号、またはビットであり得る。一部の実施形態では、データおよび／または情報は、任意の収集された、集積された、または使用可能な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、適する媒体、画像、ビデオ、音声（例えば、周波数、可聴または非可聴）、番号、定数、値、物体、時間、機能、インストラクション、マップ、参照、配列、読取り、マッピングされた読取り、レベル、範囲、閾、シグナル、ディスプレイ、表示、またはそれらの変換物が挙げられる。モ

10

20

30

40

50

ジュールは、データおよび／または情報を受け入れまたは受信し、データおよび／または情報を第2の形態に変換し、第2の形態を装置、周辺機器、コンポーネント、または別のモジュールに提供または伝送することができる。モジュールは、1つまたは複数の下記の非限定的な機能を行うことができる：例えば、配列の読取りをマッピングする、カウントを提供する、部分を集積する、レベルを提供するまたは決定する、カウントプロファイルを提供する、正規化する（例えば、読取りを正規化する、カウントを正規化する等）、正規化されたカウントプロファイルまたは正規化されたカウントのレベルを提供する、2つまたはそれ超のレベルを比較する、不確実性値を得る、期待されるレベルおよび期待される範囲（例えば、期待されるレベル範囲、閾範囲、および閾レベル）を提供するまたは決定する、レベルに調整を施す（例えば、第1のレベルの調整、第2のレベルの調整、染色体もしくはそのセグメントのプロファイルの調整、および／またはパディング）、識別情報を提供する（例えば、コピー数の変動、遺伝子の変動、または異数性を同定する）、分類する、プロットする、および／またはアウトカムを決定する。プロセッサは、ある特定の実施形態では、モジュール内でインストラクションを実施することができる。一部の実施形態では、1つまたは複数のプロセッサは、モジュールまたはモジュール群内でインストラクションを実施するように要求される。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、装置、またはソースに提供することができ、ならびにデータおよび／または情報を別のモジュール、装置、またはソースから受信することができる。

10

【0523】

コンピュータプログラム産物は、実体的なコンピュータ可読媒体に組み入れる場合もある。また非一時的コンピュータ可読媒体に実体的に組み入れる場合もある。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）上またはメモリ（例えば、ランダムアクセスメモリ）内に保管される場合もある。モジュールからのインストラクションを実施することができるモジュールおよびプロセッサは、ある装置内または異なる装置内に所在し得る。モジュールに関するインストラクションを実施することができるモジュールおよび／またはプロセッサは、ユーザーと同じ場所（例えば、ローカルネットワーク）、またはユーザーとは異なる場所（例えば、リモートネットワーク、クラウドシステム）に所在し得る。方法が、2つまたはそれ超のモジュールと併せて実施される複数の実施形態では、モジュールは、同一装置内に所在してもよく、1つまたは複数のモジュールは、物理的な場所が同一である異なる装置内に所在してもよく、1つまたは複数のモジュールは、物理的な場所が異なる、異なる装置内に所在してもよい。

20

30

【0524】

装置は、一部の実施形態では、モジュール内のインストラクションを実施する少なくとも1つのプロセッサを含む。参照ゲノムの部分に対してマッピングされた配列の読取りのカウントには、本明細書に記載する方法を実施するように構成されたインストラクションを実行するプロセッサからアクセスする場合がある。プロセッサがアクセスするカウントは、システムのメモリ内にあってもよく、カウントは、その取得後にアクセス可能およびシステムのメモリ内に配置可能である。一部の実施形態では、装置はプロセッサ（例えば、1つまたは複数のプロセッサ）を含み、プロセッサは、モジュールからの1つまたは複数のインストラクション（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／または実施することができる。一部の実施形態では、装置は、並行同調作業型のプロセッサ（`processors coordinated and working in parallel`）等の複数のプロセッサを含む。一部の実施形態では、装置は、1つまたは複数の外部プロセッサ（例えば、内部または外部のネットワーク、サーバー、保管デバイス、および／または保管ネットワーク（例えば、クラウド））と共に稼働する。一部の実施形態では、装置はモジュールを含む。ある特定の実施形態では、装置は、1つまたは複数のモジュールを含む。モジュールを含む装置は、多くの場合、1つまたは複数のデータおよび／または情報を、他のモジュールから受信し、またそれに対して伝送することができる。ある特定の実施形態では、装置は周辺機器および／またはコンポーネントを含む。ある特定の実施形態では、装置は、データおよび／または情報を、他の

40

50

モジュール、周辺機器、および／またはコンポーネントに対して、およびこれらから伝送することができる1つまたは複数の周辺機器またはコンポーネントを含み得る。ある特定の実施形態では、装置は、データおよび／または情報を提供する周辺機器および／またはコンポーネントと相互作用する。ある特定の実施形態では、周辺機器およびコンポーネントは、装置がある機能を実施するのを支援する、またはモジュールと直接相互作用する。周辺機器および／またはコンポーネントの非限定的な例として、適したコンピュータ周辺機器、I/Oもしくは保管方法、またはデバイス挙げられ、これにはスキャナー、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、マイクロフォン、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大容量記憶デバイス、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、プロセッサ、サーバー、CD、DVD、グラフィックカード、特殊I/Oデバイス（例えば、シーケンサー、フォトセル、光電子増倍管、光学読取り装置、センサー等）、1つまたは複数のフローセル、流体ハンドリングコンポーネント、ネットワークインターフェースコントローラー、ROM、RAM、無線転送方法およびデバイス（ブルートゥース、Wi-Fi等）、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含まれるが、これらに限定されない。

10

【0525】

ソフトウェアは、多くの場合、コンピュータ可読媒体に記録されているプログラムインストールを含むプログラム産物上に提供され、そのような媒体として、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、光磁気ディスクを含む光学式媒体、フラッシュドライブ、RAM、フロッピー（登録商標）ディスク等、およびプログラムインストールが記録可能であ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。オンラインで実施する際には、組織により維持されるサーバーおよびウェブサイトは、ソフトウェアダウンロードをリモートユーザーに提供するように構成され得る、またはリモートユーザーは、組織により維持されるリモートシステムにアクセスして、遠隔的にソフトウェアにアクセスすることができる。ソフトウェアはインプット情報を取得または受信することができる。ソフトウェアは、データを具体的に取得または受信するモジュール（例えば、配列の読取りデータおよび／またはマッピングされた読取りデータを受信するデータ受信モジュール）を含み得、データを具体的に処理するモジュール（例えば、受信したデータを処理する処理モジュール（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））を含み得る。用語、インプット情報を「取得する」および「受信する」とは、ローカルもしくはリモートサイトからコンピュータコミュニケーション手段により、ヒトがデータ入力することにより、または任意の他のデータ受信方法により、データ（例えば、配列の読取り、マッピングされた読取り）を受信することを指す。インプット情報は、受信した場所と同一の場所で生成される場合もあれば、異なる場所で生成され、受信場所に送られる場合もある。一部の実施形態では、インプット情報は、処理される前に修正される（例えば、処理しやすいフォーマット（例えば、表形式）に配置される）。

20

30

40

【0526】

一部の実施形態では、コンピュータプログラム産物、例えばコンピュータ可読プログラムコードを組み入れたコンピュータ使用可能媒体を含むコンピュータプログラム産物等が提供され、コンピュータ可読プログラムコードは、実行されたときに、下記ステップを含む方法を実施するように適合されている：（a）試験被験体から得た試料核酸の配列の読取りを取得するステップ；（b）（a）で得られた配列の読取りを公知のゲノムに対してマッピングするステップであって、公知のゲノムが部分に分割されているステップ；（c）部分内のマッピングされた配列の読取りをカウント計測するステップ；（d）（c）で得られた部分についてのカウントを正規化することにより、試料正規化カウントプロファイルを生成するステップ；および（e）（d）の試料正規化カウントプロファイルから

50

遺伝子の変動の存在または非存在を決定するステップ。

【0527】

ある特定の実施形態では、ソフトウェアは1つまたは複数のアルゴリズムを含み得る。アルゴリズムは、データを処理するのに、および／または有限列のインストラクションにより、アウトカムまたはレポートを提供するのに使用できる。アルゴリズムは、多くの場合、タスクを完了するための規定されたインストラクションのリストである。初期状態から開始し、インストラクションは、規定された一連の連続した状態を経由して進行し、最終的に最終エンディング状態で終了する演算について記載し得る。1つの状態から次の状態への移行は必ずしも確定的ではない（例えば、一部のアルゴリズムには、偶然性を取り入れている）。例として、アルゴリズムは、非限定的にサーチアルゴリズム、ソーティングアルゴリズム、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、ストリングアルゴリズム、モデリングアルゴリズム、計算幾何学アルゴリズム、コンビナトリアルアルゴリズム、機械学習アルゴリズム、クリプトグラフィアルゴリズム、データ圧縮アルゴリズム、パーキングアルゴリズム等であり得る。アルゴリズムは、1つのアルゴリズムまたは組み合わせて作業する2つもしくはそれ超のアルゴリズムを含み得る。アルゴリズムは、任意の適する複雑性クラス、および／またはパラメータ化された複雑性のものであってもよい。アルゴリズムは計算および／またはデータ処理するのに使用することができる。一部の実施形態では、確定的または確率的／予測的なアプローチで使用する事ができる。アルゴリズムは、適するプログラミング言語を使用することにより、演算環境内で実施可能であり、そのような言語の非限定的な例として、C、C++、Java（登録商標）、Perl、Python、Fortran等がある。一部の実施形態では、アルゴリズムは、許容誤差、統計分析、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（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使用する際に適用可能）を含むように構成または修正され得る。

【0528】

ある特定の実施形態では、いくつかのアルゴリズムが、ソフトウェア内で使用するために実施され得る。これらのアルゴリズムは、一部の実施形態では、未加工データを用いてトレーニング可能である。新しい未加工データ試料毎に、トレーニングされたアルゴリズムは、代表的な処理済みデータセットまたはアウトカムを生成し得る。処理済みのデータセットは、処理された親データセットと比較して複雑性が低減されたものの場合もある。処理済みのセットに基づき、一部の実施形態では、感度および特異性に基づきトレーニングされたアルゴリズムの性能を評価することができる。最高の感度および／または特異性を有するアルゴリズムが、ある特定の実施形態では、同定および利用され得る。

【0529】

ある特定の実施形態では、シミュレートした（またはシミュレーション）データが、例えばアルゴリズムをトレーニングするまたはアルゴリズムを試験することによりデータ処理を補助することができる。一部の実施形態では、シミュレートしたデータには、配列の読取りの異なるグルーピングの、仮想的な様々なサンプリングが含まれる。シミュレートしたデータでは、何が真の母集団から期待されか、またはアルゴリズムを試験する、および／または正しい分類を割り当てる際に何に歪みが生じ得るか、が基準となり得る。また、シミュレートしたデータは、本明細書では、「仮想」データとも呼ばれる。シミュレーションは、ある特定の実施形態では、コンピュータプログラムにより行われ得る。シミュレートしたデータセットを使用する際の1つの可能なステップは、確認された結果の信頼度を評価すること、例えばランダムサンプリングが、どのくらい良好にオリジナルデータと一致するか、またはオリジナルデータを最好に代表するか、評価することである。1つのアプローチは、確率値（p値）を計算することであり、この値は、ランダム試料が選択された試料より良好なスコアを有する確率を推定する。一部の実施形態では、経験的モデルが評価される場合があり、この場合、少なくとも1つの試料が参照試料と一致することを前提とする（分解変動（resolved variation）の有りまたは無しを問わない）。一部の実施形態では、例えばポアソン分布等の別の分布が、確率分布を規定

するのに使用することができる。

【0530】

システムは、ある特定の実施形態では、1つまたは複数のプロセッサを含み得る。プロセッサは、コミュニケーションバスと接続され得る。コンピュータシステムは、メインメモリ、多くの場合ランダムアクセスメモリ（RAM）を含み得、二次メモリも含むことができる。一部の実施形態では、メモリは、非一時的コンピュータ可読記憶媒体を含む。二次メモリは、例えばハードディスクドライブおよび／またはリムーバブル記憶ドライブを含み得、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、光学式ディスクドライブ、メモリカード等がこれに該当し得る。リムーバブル記憶ドライブは、多くの場合、リムーバブルストレージユニットから読み取る、および／またはこれに書き込む。リムーバブルストレージユニットの非限定的な例として、フロッピー（登録商標）ディスク、磁気テープ、光学式ディスク等が挙げられ、例えばリムーバブル記憶ドライブにより、読取りおよび書き込み可能である。リムーバブルストレージユニットは、コンピュータソフトウェアおよび／またはデータを内蔵するコンピュータ使用可能記憶媒体を含み得る。

10

【0531】

プロセッサは、システム内でソフトウェアを実施可能である。一部の実施形態では、プロセッサは、ユーザーが行うことができる、本明細書に記載す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、プロセッサまたはかかるプロセッサにより実施されるアルゴリズムは、ユーザーによる監視またはインプットを、ほとんどまたはまったく必要としないと考えられる（例えば、ソフトウェアは、機能を自動的に実施するようにプログラムされ得る）。一部の実施形態では、処理はあまりにも複雑であり、一人の個人であっても、また個人の群であっても、遺伝子の変動の存在または非存在を決定するのに十分短いタイムフレーム内で処理を行うことは不可能である。

20

【0532】

一部の実施形態では、二次メモリは、コンピュータプログラムまたは他のインストラクションをコンピュータシステムにロードできるようにするために、他の類似した手段を含み得る。例えば、システムは、リムーバブルストレージユニットおよびインターフェースデバイスを含み得る。かかるシステムの非限定的な例として、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（ビデオゲームデバイスに見出されるもの等）、リムーバブルメモリチップ（EPROMまたはPROM等）、および関連するソケット、ならびにソフトウェアおよびデータをリムーバブルストレージユニットからコンピュータシステムに伝送できるようにする、他のリムーバブルストレージユニットおよびインターフェースが挙げられる。

30

【0533】

一部の実施形態では、1つの実体は、配列の読取りのカウントを生成すること、配列の読取りを部分に対してマッピングすること、マッピングされた読取りをカウント計測すること、およびカウントが計測されマッピングされた読取りを、本明細書に記載する方法、システム、装置、またはコンピュータプログラム産物において利用することができる。ある特定の実施形態では、部分に対してマッピングされた配列の読取りのカウントは、本明細書に記載する方法、システム、装置、またはコンピュータプログラム産物において、第2の実体を使用するために、1つの実体により、第2の実体に伝送される場合もある。

40

【0534】

一部の実施形態では、1つの実体は配列の読取りを生成し、一部の実施形態では、第2の実体はその配列の読取りを参照ゲノム内の部分に対してマッピングする。第2の実体は、マッピングされた読取りをカウント計測し、カウントが計測されマッピングされた読取りを、本明細書に記載する方法、システム、装置、またはコンピュータプログラム産物において利用する場合がある。ある特定の実施形態では、第2の実体は、マッピングされた読取りを第3の実体に伝送し、第3の実体は、マッピングされた読取りをカウント計測し、マッピングされた読取りを、本明細書に記載する方法、システム、装置、またはコンピ

50

ュータプログラム産物において利用する。ある特定の実施形態では、第2の実体は、マッピングされた読取りをカウント計測し、カウントが計測されマッピングされた読取りを第3の実体に伝送し、第3の実体は、カウントが計測されマッピングされた読取りを、本明細書に記載する方法、システム、装置、またはコンピュータプログラム産物において利用する。第3の実体が関与する実施形態では、第3の実体は、第1の実体と同一である場合もある。すなわち、第1の実体は、配列の読取りを第2の実体に伝送する場合があります、この第2の実体は、参照ゲノム内の部分に対して配列の読取りをマッピングする、および／またはマッピングされた読取りをカウント計測することができ、第2の実体は、マッピングされおよび／またはカウントが計測された読取りを第3の実体に伝送することができる。第3の実体は、マッピングされおよび／またはカウントが計測された読取りを本明細書に記載する方法、システム、装置、またはコンピュータプログラム産物において利用することができる場合もあり、この場合、第3の実体は第1の実体と同一である場合もあれば、第3の実体は第1または第2の実体とは異なる場合もある。

10

【0535】

一部の実施形態では、1つの実体は、妊娠中の雌から血液を取得し、任意選択で血液から（例えば、血漿または血清から）核酸を単離し、核酸から配列の読取りを生成する第2の実体に血液または核酸を移送する。

【0536】

図30は、本明細書に記載する様々なシステム、方法、アルゴリズム、およびデータ構造の実施が可能である演算環境510の非限定的な例を示す。演算環境510は、適する演算環境の1つの例に過ぎず、本明細書に記載するシステム、方法、およびデータ構造の使用の範囲または機能性について何らかの制限を示唆するようには意図されない。また、演算環境510は、演算環境510に示すコンポーネントの任意の1つまたはその組合せと関連する何らかの依存性または要件を有するものと解釈してはならない。図30に示すシステム、方法、およびデータ構造のサブセットは、ある特定の実施形態で利用可能である。本明細書に記載するシステム、方法、およびデータ構造は、非常に多くの他の汎用または専用の演算システム環境またはコンフィギュレーションと共に運用可能である。適すると考えられる公知の演算システム、環境、および／またはコンフィギュレーションの例として、パーソナルコンピュータ、サーバーコンピュータ、シンクライアント、シッククライアント、携帯式またはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサに基づくシステム、セットトップボックス、プログラム可能な民生用電子機器、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含む分散型演算環境等が挙げられるが、これらに限定されない。

20

30

【0537】

図30のオペレーティング環境510はコンピュータ520の形態の汎用演算デバイスを含み、これには、処理ユニット521、システムメモリ522、およびシステムメモリ522を含む様々なシステムコンポーネントを処理ユニット521に作動可能に連結させるシステムバス523が含まれる。コンピュータ520のプロセッサが、単一の中央処理装置（CPU）または並列処理環境と一般的に呼ばれる複数の処理ユニットを含むように、処理ユニット521は1つのみ存在し得る、または1つ超存在し得る。コンピュータ520は、従来型コンピュータ、分散型コンピュータ、またはあらゆる他の種類のコンピュータであり得る。

40

【0538】

システムバス523は、メモリバスまたはメモリコントローラー、周辺バス、および様々なバスアーキテクチャのいずれかを使用するローカルバスを含む、任意の数種類のバス構造であり得る。また、システムメモリは、単にメモリと呼ばれる場合もあり、リードオンリメモリ（ROM）524およびランダムアクセスメモリ（RAM）を含む。立ち上げの間等に、コンピュータ520内のエレメント間の情報伝送に役立つ基本ルーチンを含む基本入出力システム（BIOS）526は、ROM524に保管される。コンピュータ

50

520は、図示しないがハードディスクから読み出し、これに書き込むハードディスクドライバインターフェース527、リムーバブル磁気ディスク529から読み出し、これに書き込む磁気ディスクドライバ528、およびリムーバブル光学式ディスク531、例えばCD-ROMまたは他の光学式媒体から読み出し、これに書き込む光学式ディスクドライバ530をさらに含み得る。

【0539】

ハードディスクドライバ527、磁気ディスクドライバ528、および光学式ディスクドライバ530は、ハードディスクドライバインターフェース532、磁気ディスクドライバインターフェース533、および光学式ディスクドライバインターフェース534により、システムバス523とそれぞれ接続される。ドライバおよびその関連するコンピュータ可読媒体は、コンピュータ可読インストラクション、データ構造、プログラムモジュール、およびコンピュータ520用の他のデータの揮発性の保管を提供する。コンピュータがアクセス可能なデータを保管することができる、あらゆる種類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리카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(RAM)、リードオンリーメモリ(ROM)等が、オペレーティング環境内で使用することができる。

【0540】

いくつかのプログラムモジュールが、オペレーティングシステム535、1つまたは複数の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を含む、ハードディスク、磁気ディスク529、光学式ディスク531、ROM524、またはRAM上に保管され得る。ユーザーは、コマンドおよび情報を、インプットデバイス、例えばキーボード540およびポインティングデバイス542を通じてパーソナルコンピュータ520に入力することができる。他のインプットデバイス(図示せず)として、マイクロフォン、ジョイスティック、ゲームパッド、サテライトディッシュ、スキャナー等を挙げることができる。これらおよび他のインプットデバイスが、多くの場合、システムバスに連結したシリアルポートインターフェース546を経由して処理ユニット521と接続されるが、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)により接続される場合もある。モニター547または他の種類のディスプレイデバイスも、インターフェース、例えばビデオアダプター548を介してシステムバス523と接続される。モニターに加えて、コンピュータは、他の周辺アウトプットデバイス(図示せず)、例えばスピーカーおよびプリンターを一般的に含む。

【0541】

コンピュータ520は、1つまたは複数のリモートコンピュータ、例えばリモートコンピュータ549との論理接続を使用して、ネットワーク化した環境内で作動可能である。これらの論理接続は、コンピュータ520もしくはその一部と連結しているコミュニケーションデバイスにより、または他の方式で達成され得る。図30ではメモリストレージデバイス550しか示さなかったが、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、ピアデバイス、もしくは他の一般的なネットワークノードであり得、コンピュータ520と関連して上記エレメントの多くまたは全てを一般的に含む。図30に示す論理接続として、ローカルエリアネットワーク(LAN)551およびワイドエリアネットワーク(WAN)552が挙げられる。かかるネットワーク環境は、オフィスネットワーク、企業全体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では普通であり、そのいずれも典型的なネットワークである。

【0542】

LAN-ネットワーク環境で使用する場合、コンピュータ520は、コミュニケーションデバイスの一種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553を介してローカルネットワーク551と接続される。WAN-ネットワーク環境で使用する場合、コンピュータ520は、多くの場合、コミュニケーションデバイスの一種であるモデム5

10

20

30

40

50

5 4、またはワイドエリアネットワーク 5 5 2 全体にわたりコミュニケーションを確立するために他の任意の種類のコミュニケーションデバイスを含む。モデム 5 5 4 は、内部または外部であってもよいが、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 と接続される。ネットワーク化された環境では、パーソナルコンピュータ 5 2 0 またはその一部と関連して示されるプログラムモジュールは、リモートメモリストレージデバイス内に保管され得る。示すようなネットワーク接続は非限定的な例であり、また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の他のコミュニケーションデバイスも使用することができると認識される。

モジュール

【0 5 4 3】

1 つまたは複数のモジュールが本明細書に記載する方法で利用可能であり、その非限定的な例として、論理処理モジュール、配列決定モジュール、マッピングモジュール、カウント計測モジュール、フィルタリングモジュール、重み付けモジュール、正規化モジュール、G C 偏りモジュール、レベルモジュール、比較モジュール、範囲設定モジュール、分類モジュール、プロットングモジュール、表示モジュール、関係モジュール、アウトカムモジュール、および／またはデータディスプレイ組織化モジュール等、またはその組み合わせが挙げられる。モジュールは、マイクロプロセッサにより管理される場合もある。ある特定の実施形態では、モジュールまたは 1 つもしくは複数のモジュールを含む装置は、別のモジュール、装置、コンポーネント、周辺機器、または装置のオペレーターに、またはそれらから、データおよび／または情報を収集、集積、受信、取得、アクセス、回収、提供、および／または伝送する。一部の実施形態では、データおよび／または情報（例えば、配列決定の読取り）は、下記の 1 つまたは複数の含む装置によりモジュールに提供される：1 つまたは複数のフローセル、カメラ、検出器（例えば、光検出器、フォトセル、電気的検出器（例えば、振幅変調検出器、周波数および位相変調検出器、位相ロックループ検出器）、カウンタ、センサー（例えば、圧力、温度、容積、フロー、重量のセンサー）、流体ハンドリングデバイス、プリンター、ディスプレイ（例えば、LED、LCD、またはCRT）等またはその組合せ。例えば、装置のオペレーターは、定数、閾値、式、または所定の値をモジュールに提供する場合もある。モジュールは、多くの場合、データおよび／または情報を、別のモジュールもしくは装置に、またはそれから伝送するように構成される。モジュールは、別のモジュールからデータおよび／または情報を受信することができ、その非限定的な例として、論理処理モジュール、配列決定モジュール、マッピングモジュール、カウント計測モジュール、フィルタリングモジュール、重み付けモジュール、正規化モジュール、G C 偏りモジュール、レベルモジュール、比較モジュール、範囲設定モジュール、分類モジュール、プロットングモジュール、表示モジュール、関係モジュール、アウトカムモジュール、および／またはデータディスプレイ組織化モジュール等またはその組合せが挙げられる。モジュールは、データおよび／または情報を操作および／または変換することができる。モジュールに由来する、またはモジュールにより変換されたデータおよび／または情報は、別の適する装置および／またはモジュールに伝送することができ、その非限定的な例として、論理処理モジュール、配列決定モジュール、マッピングモジュール、カウント計測モジュール、フィルタリングモジュール、重み付けモジュール、正規化モジュール、G C 偏りモジュール、レベルモジュール、比較モジュール、範囲設定モジュール、分類モジュール、プロットングモジュール、表示モジュール、関係モジュール、アウトカムモジュール、および／またはデータディスプレイ組織化モジュール等またはその組合せが挙げられる。モジュールを含む装置は、少なくとも 1 つのプロセッサを含み得る。一部の実施形態では、データおよび／または情報は、モジュールを含む装置により受信および／または提供される。モジュールを含む装置は、プロセッサを含むことができ（例えば、1 つまたは複数のプロセッサ）、そのようなプロセッサは、モジュールの 1 つまたは複数のインストラクション（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／または実施することができる。一部の実施形態では、モジュールは、1 つまたは複数の外部プロセッサ（例えば、内部または外部の

10

20

30

40

50

ネットワーク、サーバー、ストレージデバイス、および／またはストレージネットワーク（例えば、クラウド））と共に作動する。

論理処理モジュール

【0544】

ある特定の実施形態では、論理処理モジュールは、データおよび／もしくは情報を、またはデータおよび／もしくは情報の、1つもしくは複数のその他のモジュール、周辺機器、もしくはデバイスへの、およびそれからの伝送を、統合、管理、制限、組織化、命令、分配、分割、変換、および／もしくは制御する。

データディスプレイ組織化モジュール

【0545】

ある特定の実施形態では、データディスプレイ組織化モジュールは、データおよび／または情報を適する可視的媒体へと処理および／または変換するが、その媒体の非限定的な例として、画像、ビデオおよび／またはテキスト（例えば、数字、文字、およびシンボル）が挙げられる。一部の実施形態では、データディスプレイ組織化モジュールは、適するディスプレイ（例えば、モニター、LED、LCD、CRT等、またはその組合せ）、プリンター、適する周辺機器、またはデバイス上に表示するために、データおよび／または情報を、処理、変換、および／または伝送する。一部の実施形態では、データディスプレイ組織化モジュールは、胎児または母体のゲノム、染色体、またはその一部分のデータおよび／または情報を可視的表示に処理、変換する。

配列決定モジュール

【0546】

一部の実施形態では、配列モジュールは、配列の読取りを取得、生成、収集、集積、操作、変換、処理、変換、および／または伝送する。「配列受信モジュール」は、本明細書で使用する場合、「配列決定モジュール」と同じである。配列決定モジュールを含む装置は、当技術分野において公知の配列決定技術を利用して核酸の配列を決定するあらゆる装置であり得る。一部の実施形態では、配列決定モジュールは、配列の読取りを整列、集積、断片化、相補、逆相補、エラーチェック、またはエラー修正することができる。

マッピングモジュール

【0547】

配列の読取りは、マッピングモジュールにより、またはマッピングモジュールを含む装置によりマッピング可能であり、このマッピングモジュールは、一般的に、参照ゲノムまたはそのセグメントに対して読取りをマッピングする。マッピングモジュールは、配列決定の読取りを、当技術分野において公知の適する方法によりマッピング可能である。一部の実施形態では、マッピングモジュールまたはマッピングモジュールを含む装置は、マッピングされた配列の読取りを提供するように要求される。

カウント計測モジュール

【0548】

カウントは、カウント計測モジュールまたはカウント計測モジュールを含む装置により提供され得る。一部の実施形態では、カウント計測モジュールは、参照ゲノムに対してマッピングされた配列の読取りをカウント計測する。一部の実施形態では、カウント計測モジュールは、当技術分野において公知のカウント計測法により、カウントを生成、集積、および／または提供する。一部の実施形態では、カウント計測モジュールまたはカウント計測モジュールを含む装置は、カウントを提供するように要求される。

フィルタリングモジュール

【0549】

フィルタリング部分（例えば、参照ゲノムの部分）は、フィルタリングモジュールにより（例えば、フィルタリングモジュールを含む装置により）提供され得る。一部の実施形態では、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のデータ（例えば、フィルタリングされた部分）を提供する、および／または検討事項から部分を除去するように要求される。ある特定の実施形態では、フィルタリングモジュールは、部分に対してマ

10

20

30

40

50

ッピングされたカウントを検討事項から除去する。ある特定の実施形態では、フィルタリングモジュールは、部分に対してマッピングされたカウントを、レベルまたはプロファイルの決定から除去する。フィルタリングモジュールは、当技術分野において公知の、または本明細書に記載する1つまたは複数のフィルタリング法により、データ（例えば、カウント、部分に対してマッピングされたカウント、部分、部分のレベル、正規化されたカウント、未加工のカウント等）をフィルタリングすることができる。

重み付けモジュール

【0550】

重み付け部分（例えば、参照ゲノムの部分）は、重み付けモジュールにより（例えば、重み付けモジュールを含む装置により）提供され得る。一部の実施形態では、重み付けモジュールは、ゲノム区分を重み付けする、および／または重み付けされた部分の値を提供するように要求される。重み付けモジュールは、当技術分野において公知の、または本明細書に記載する1つまたは複数の重み付け法により、部分を重み付けすることができる。

10

正規化モジュール

【0551】

正規化されたデータ（例えば、正規化されたカウント）は、正規化モジュールにより（例えば、正規化モジュールを含む装置により）提供され得る。一部の実施形態では、正規化モジュールは、配列決定の読取りから得られた正規化されたデータ（例えば、正規化されたカウント）を提供するように要求される。正規化モジュールは、本明細書に記載する、または当技術分野において公知の1つまたは複数の正規化法（例えば、PERUN、ChAI、ハイブリッド式の正規化等またはその組合せ）により、データ（例えば、カウント、フィルタリングされたカウント、未加工のカウント）を正規化することができる。

20

GC 偏りモジュール

【0552】

GCの偏りを決定すること（例えば、参照ゲノムの部分（例えば、部分、参照ゲノムの部分）のそれぞれについてGCの偏りを決定すること）は、GC偏りモジュールにより（例えば、GC偏りモジュールを含む装置により）提供され得る。一部の実施形態では、GC偏りモジュールは、GCの偏りの決定を提供するように要求される。一部の実施形態では、GC偏りモジュールは、参照ゲノムの部分のそれぞれに対してマッピングした配列の読取りのカウントと各部分のGC含有量との間で適合させた関係（例えば、線形適合関係）からGCの偏りの決定を提供する。GC偏りモジュールは、正規化モジュール（例えば、PERUN、ChAI正規化モジュール）の一部分である場合もある。

30

レベルモジュール

【0553】

参照ゲノムの部分についてレベル（例えば、レベル）を決定すること、および／またはゲノム区分のレベルを計算することでは、レベルモジュールにより（例えば、レベルモジュールを含む装置により）提供され得る。一部の実施形態では、レベルモジュールは、レベルまたは計算されたゲノム区分のレベル（例えば、等式A、B、L、M、N、O、および／またはQによる）を提供するように要求される。一部の実施形態では、レベルモジュールは、GCの偏りと参照ゲノムの部分のそれぞれに対してマッピングした配列の読取りのカウントとの間で適合させた関係（例えば、線形適合関係）からレベルを提供する。一部の実施形態では、レベルモジュールは、PERUNの一部分としてゲノム区分のレベルを計算する。一部の実施形態では、レベルモジュールは、等式 $L_i = (m_i - G_i S) I^{-1}$ により、ゲノム区分のレベル（すなわち、 L_i ）を提供し、式中 G_i はGCの偏り、 m_i は参照ゲノムの各部分に対してマッピングした測定カウントであり、 i は試料であり、 I は、GCの偏りと参照ゲノムの部分のそれぞれに対してマッピングした配列の読取りのカウントとの間で適合させた関係（例えば、線形適合関係）の切片、 S は、その勾配である。

40

比較モジュール

【0554】

50

第1のレベルは、比較モジュールまたは比較モジュールを含む装置により、第2のレベルとは有意に異なるものとして同定され得る。一部の実施形態では、比較モジュールまたは比較モジュールを含む装置は、2つレベル間の比較を提供するように要求される。

範囲設定モジュール

【0555】

様々なコピー数の変動（例えば、重複、挿入、および／または欠失）に関する期待される範囲（例えば、期待されるレベル範囲）、またはコピー数の変動が存在しない範囲は、範囲設定モジュールまたは範囲設定モジュールを含む装置により提供され得る。ある特定の実施形態では、期待されるレベルは、範囲設定モジュールまたは範囲設定モジュールを含む装置により提供される。一部の実施形態では、範囲設定モジュールまたは範囲設定モジュールを含む装置は、期待されるレベルおよび／または範囲を提供するように要求される。

10

分類モジュール

【0556】

コピー数の変動（例えば、母体および／または胎児のコピー数の変動、胎児のコピー数の変動、重複、挿入、欠失）は、分類モジュールまたは分類モジュールを含む装置により分類され得る。ある特定の実施形態では、コピー数の変動（例えば、母体および／または胎児のコピー数の変動）は、分類モジュールにより分類される。ある特定の実施形態では、別のレベル（例えば、第2のレベル）とは有意に異なると決定されたレベル（例えば、第1のレベル）は、分類モジュールによりコピー数の変動を表わすものとして同定される。ある特定の実施形態では、コピー数の変動の非存在が分類モジュールにより決定される。一部の実施形態では、コピー数の変動の決定は、分類モジュールを含む装置により決定され得る。分類モジュールは、母体および／または胎児のコピー数の変動、胎児のコピー数の変動、重複、欠失、または挿入もしくはその欠如、または上記のものの組合せを分類することに特化し得る。例えば、母体の欠失を同定する分類モジュールは、胎児の重複を同定する分類モジュールとは異なるおよび／または相違し得る。一部の実施形態では、分類モジュールまたは分類モジュールを含む装置は、コピー数の変動を同定すること、またはコピー数の変動を決定するアウトカムが要求される。

20

プロットिंगモジュール

【0557】

一部の実施形態では、プロットिंगモジュールは、データおよび／または情報を適する可視的媒体へと処理および／または変換するが、その非限定的な例として、チャート、プロット、グラフ等またはその組合せが挙げられる。一部の実施形態では、プロットिंगモジュールは、適するディスプレイ（例えば、モニター、LED、LCD、CRT等またはその組合せ）、プリンター、適する周辺機器、またはデバイス上に表示するために、データおよび／または情報を処理、変換、および／または伝送する。ある特定の実施形態では、プロットिंगモジュールは、カウント、レベル、および／またはプロファイルのビジュアルディスプレイを提供する。一部の実施形態では、データディスプレイ組織化モジュールが、データおよび／または情報を、胎児または母体のゲノム、染色体、またはその一部分について、可視的表示へと処理、変換する。

30

40

【0558】

一部の実施形態では、プロットングモジュールまたはプロットングモジュールを含む装置は、カウント、レベル、またはプロファイルのプロットするように要求される。

関係モジュール

【0559】

ある特定の実施形態では、関係モジュールが、データおよび／または情報を、関係へと処理および／または変換する。ある特定の実施形態では、関係は、関係モジュールにより生成されるおよび／またはこれから伝送される。

アウトカムモジュール

【0560】

50

遺伝子の変動の存在または非存在（異数性、胎児の異数性、コピー数の変動）は、一部の実施形態では、アウトカムモジュールまたはアウトカムモジュールを含む装置により同定される。ある特定の実施形態では、遺伝子の変動は、アウトカムモジュールにより同定される。多くの場合、異数性の存在または非存在の決定は、アウトカムモジュールにより同定される。一部の実施形態では、遺伝子の変動（異数性、コピー数の変動）の決定因のアウトカムは、アウトカムモジュールまたはアウトカムモジュールを含む装置により同定され得る。アウトカムモジュールは、特異的な遺伝子の変動（例えば、トリソミー、トリソミー 2 1、トリソミー 1 8）の決定に特化し得る。例えば、トリソミー 2 1 を同定するアウトカムモジュールは、トリソミー 1 8 を同定するアウトカムモジュールとは異なるおおよび／または相違し得る。一部の実施形態では、アウトカムモジュールまたはアウトカムモジュールを含む装置は、遺伝子の変動または遺伝子の変動の決定因のアウトカム（例えば、異数性、コピー数の変動）を同定するように要求される。本明細書に記載する方法により同定される遺伝子の変動または遺伝子の変動の決定因のアウトカムは、さらなる試験により（例えば、母体おおよび／または胎児の核酸の標的化配列決定法により）独立して確かめ得る。

10

変換

【0561】

上記のように、データは 1 つの形態から別の形態に変換される場合もある。用語「変換された」、「変換」、およびその文法的な派生物または同等物は、本明細書で使用する場合、身体の出発材料（例えば、試験被験体おおよび／または参照被験体試料の核酸）から身体の出発材料のデジタル表示（例えば、配列の読取りデータ）へのデータの変更を指し、一部の実施形態では、アウトカムを提供するのに利用できる 1 つもしくは複数の数値への、またはデジタル表示の図形表示へのさらなる変換を含む。ある特定の実施形態では、1 つまたは複数の数値おおよび／またはデジタル的に表示されたデータの図形表示は、試験被験体の身体のゲノムの状況を表すのに利用できる（例えば、ゲノムの挿入、重複、または欠失の存在または非存在を仮想的に表すまたは可視的に表す；医学的状态と関連した配列の物理量の変動の存在または非存在を表す）。仮想表示は、1 つもしくは複数の数値、または出発材料のデジタル表示の図形表示にさらに変換される場合もある。これらの方法は、身体の出発材料を、数値もしくは図形表示に、または試験被験体ゲノムの物理的状況表示に変換することができる。

20

30

【0562】

一部の実施形態では、データセットを変換すると、データの複雑性おおよび／またはデータの次元数が低減し、これによりアウトカムの提供が容易になる。データセットの複雑性は、身体の出発材料を出発材料の仮想表示に変換する処理の間に低減する場合もある（例えば、身体の出発材料を表わす配列の読取り）。適する特性または変数が、データセットの複雑性おおよび／または次元数を低減するのに利用できる。データ処理するための標的特性として使用するのに選択できる特性の非限定的な例として、GC 含有量、胎児の性別予測、染色体異数性の同定、特定の遺伝子またはタンパク質の同定、がん、疾患、遺伝性の遺伝子／特性、染色体異常の同定、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、共制御された遺伝子、細胞シグナル伝達遺伝子、細胞周期遺伝子、上記遺伝子に関連するタンパク質、遺伝子変異体、タンパク質変異体、共制御された遺伝子、共制御されたタンパク質、アミノ酸配列、ヌクレオチド配列、タンパク質構造データ等、おおよび上記のものの組合せが挙げられる。データセットの複雑性おおよび／または次元数の低減に関する非限定的な例として；複数の配列読取りをプロファイルプロットに低減化すること、複数の配列読取りを数値に低減化すること（例えば、値、Z スコア、p 値の正規化）；複数の分析方法を確率プロットまたは単一ポイントに低減化すること；導き出された数量の主成分分析等、またはその組合せが挙げられる。

40

【実施例】

【0563】

50

下記の実施例は、もっぱら例示として提示されていて、制限するものではない。したがって、下記の実施例はある特定の実施形態について説明し、本技術に制限を設けるものではない。当業者は、本質的に同一または類似のアウトカムを得るために変更または修正可能な様々な非クリティカルパラメータを容易に認識する。

【0564】

(実施例1)

PERUNおよび遺伝子の変動と関連した状態を検出するための一般的方法。

【0565】

本明細書に記載する方法および基礎理論は、遺伝子の変動と関連した様々な状態を検出するのに利用可能であり、遺伝子の変動の存在または非存在の決定的なアウトカムを提供する、またはその存在または非存在を決定する。

10

参照ゲノムの有益でない部分の除去

【0566】

参照ゲノムの有益でない部分を除去するための複数の試みにより、部分を選択することが分類を改善し得ると判明した。

等式A：

$$M = L I + G S$$

(A)

等式Aの各項は下記の意味を有する：

M：測定されたカウント、不必要な変動により影響を受けた一次情報を表す。

20

L：染色体レベルこれは、データ処理手順に由来する所望のアウトプットである。Lは、胎児および／または母体の正倍数体からの逸脱を示す。これは、確率誤差および系統的偏りの両方によりマスクされる数量である。染色体レベルLは、試料特異的かつ部分特異的である。

G：線形モデルのLOESS、または任意の同等のアプローチを使用して測定されたGCの偏り係数。Gは、Mおよび一連の部分特異的GC含有量の値から抽出される二次情報を表し、通常参照ゲノムに由来する（ただし、実際に観察されたGC含有量に由来する場合もある）。Gは、試料特異的であり、またゲノム位置によらず不変である。Gは不必要な変動のある部分を包含する。

I：線形モデルの切片。このモデルパラメータは所与の実験設定について一定であり、試料から独立しており、部分特異的である。

30

S：線形モデルの勾配。このモデルパラメータは所与の実験設定について一定であり、試料から独立しており、部分特異的である。

【0567】

MおよびGの数量を測定する。最初に、部分特異的値IおよびSが未知である。未知のIおよびSを評価するために、正倍数体試料中の参照ゲノムの全ての部分についてL=1と仮定しなければならない。仮定は必ずしも真ではないが、欠失／重複を有する試料は、いずれも正常な染色体レベルを有する試料に劣後するものと、合理的に予想することができる。正倍数体試料に適用される線形モデルにより、選択された部分に対して特異的なIおよびSパラメータ値が抽出される（L=1と仮定）。同一の手順が、ヒトゲノム中の参照ゲノムの全ての部分に適用されて、ゲノムの位置毎に切片Iおよび勾配Sがセットで得られる。交差検証では、全LDTv2CE正倍数体の90%を含有するワークセットがランダムに選択され、モデルをトレーニングするのにそのサブセットを使用する。ランダム選択は100回繰り返され、部分毎に100個の勾配および100個の切片がセットで得られる。

40

測定されたカウントからの染色体レベルの抽出

【0568】

部分毎にモデルパラメータ値IおよびSが入手可能であると仮定し、新しい試験試料について収集された測定値Mが、下記の等式Bに従って、染色体レベルを評価するのに使用される：

50

$$L = (M - GS) / I \quad (B)$$

【0569】

等式Aと同様に、GCの偏り係数Gを、部分毎に測定された未加工のカウントMと参照ゲノムのGC含有量との間の回帰式の勾配として評価する。次に、染色体レベルLをさらなる分析で使用する（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、性染色体異数性等）。等式Bに包含される手順は、PERUN（parameterized error removal and unbiased normalization）と命名されている。

（実施例2）

式の例

10

【0570】

以下の記載は、本明細書に記載する方法で 사용할 ことができる数学的および／または統計的な式の非限定的な例である。

【0571】

Zスコア、およびZスコアから計算され、予想レベルの1からの偏差と関連するp値は、次に、平均レベルにおける不確実性に関する推定値に照らし評価可能である。p値はt分布に基づき、t分布の次数はピークにおける参照ゲノムの部分の数によって決定する。所望の信頼度レベルに応じて、カットオフはノイズを抑えることができ、実際のシグナルを確実に検出できるようにする。

等式1：

20

【化5】

$$Z = \frac{\Delta_1 - \Delta_2}{\sqrt{\sigma_1^2 \left(\frac{1}{N_1} + \frac{1}{n_1} \right) + \sigma_2^2 \left(\frac{1}{N_2} + \frac{1}{n_2} \right)}} \quad (1)$$

【0572】

等式1は、2つの異なる試料に由来するピークレベルを直接比較するのに使用することができ、この場合、Nおよびnは、染色体全体中および異常部分内の参照ゲノムの部分の数をそれぞれ指す。2つの試料間の類似性を測定するp値をもたらしt-検定の次数は、2つの異常なストレッチのうち短い方における参照ゲノムの部分の数によって決定する。

30

【0573】

等式8は、胎児の異数性について遺伝子の変動の存在または非存在を決定するために、胎児フラクション（fetal fraction）、母体の倍数性、および参照カウント中央値を、分類スキームに組み入れるのに利用可能である。

等式8：

$$y_i = (1 - F) M_i f_i + F X f_i \quad (8)$$

式中、 Y_i は試験試料内の部分に関する測定されたカウントを表し、カウントプロファイル中央値内の部分に対応し、Fは胎児フラクションを表し、Xは胎児の倍数性を表し、 M_i は各部分に割り当てられた母体の倍数性を表す。等式（8）のXに使用される可能な値は：胎児が正倍数体の場合1；胎児が三倍体の場合3／2；および双胎胎児が存在し、一方が罹患し、他方はそうではない場合5／4である。双胎の症例では、一方の胎児が罹患しており、他方はそうではない場合、5／4が使用されるが、その理由は、等式（8）の項Fは総胎児DNAを表し、したがって全ての胎児DNAが考慮されなければならないためである。一部の実施形態では、母体ゲノムにおける大規模な欠失および／または重複は、各部分または部分に母体の倍数性、 M_i を割り当てることにより説明され得る。母体の倍数性は、多くの場合、1／2の倍数として割りあてられ、一部の実施形態では、部分に関する正規化を使用して推定可能である。母体の倍数性は、多くの場合、1／2の倍数なので、母体の倍数性は容易に説明をつけることがつき、したがって導関数を単純化するためのさらなる等式に含まれることはない。

40

【0574】

50

X = 1で等式(8)を評価すると、(例えば、正倍数体を仮定)、胎児フラクシオンは相殺され、下記の等式は残差平方和をもたらす。

等式9：

【化6】

$$\phi_z = \sum_{i=1}^N \frac{1}{q_i} (y_i - f_i)^2 = \sum_{i=1}^N \frac{y_i^2}{q_i} - 2 \sum_{i=1}^N \frac{y_i f_i}{q_i} + \sum_{i=1}^N \frac{f_i^2}{q_i} = \Xi_{yy} - 2\Xi_{fy} + \Xi_{ff} \quad (9)$$

【0575】

等式(9)および後続する計算を単純化するために、下記の等式が利用される。

等式10：

【化7】

$$\Xi_{yy} = \sum_{i=1}^N \frac{y_i^2}{q_i} \quad (10)$$

等式11：

【化8】

$$\Xi_{ff} = \sum_{i=1}^N \frac{f_i^2}{q_i} \quad (11)$$

等式12：

【化9】

$$\Xi_{fy} = \sum_{i=1}^N \frac{y_i f_i}{q_i} \quad (12)$$

【0576】

X = 3/2で等式(8)を評価する場合(例えば、三倍体を仮定)、下記の等式は残差平方和をもたらす。

等式13：

【化10】

$$\phi_r = \sum_{i=1}^N \frac{1}{q_i} \left(y_i - f_i - \frac{1}{2} F f_i \right)^2 = \Xi_{yy} - 2\Xi_{fy} + \Xi_{ff} + F(\Xi_{ff} - \Xi_{fy}) + \frac{1}{4} F^2 \Xi_{ff} \quad (13)$$

【0577】

等式(9)と(13)の間の差異は、機能的アウトカム(例えば、phi)をなすが、この機能的アウトカムは、代替仮説(例えば、トリソミー単生児、X = 3/2)に対して帰無仮説(例えば、正倍数体、X = 1)を検定するのに使用することができる：

等式14：

【化11】

$$\phi = \phi_z - \phi_r = F(\Xi_{fy} - \Xi_{ff}) - \frac{1}{4} F^2 \Xi_{ff} \quad (14)$$

等式18：

10

20

30

40

【化 1 2】

$$\begin{aligned}\varphi &= \sum_{i=1}^N \frac{1}{\sigma_i^2} [y_i - (1-F)M_i f_i - F X f_i]^2 \\ &= \sum_{i=1}^N \frac{1}{\sigma_i^2} [y_i^2 - 2(1-F)M_i f_i y_i - 2FX f_i y_i + (1-F)^2 M_i^2 f_i^2 + 2F(1-F)X M_i f_i^2 + F^2 X^2 f_i^2]\end{aligned}$$

(18)

【0 5 7 8】

最適な倍数性値は、等式 2 0 から得られる場合もある：

【化 1 3】

10

$$X = \frac{\sum_{i=1}^N \frac{f_i y_i}{\sigma_i^2} - (1-F) \sum_{i=1}^N \frac{M_i f_i^2}{\sigma_i^2}}{F \sum_{i=1}^N \frac{f_i^2}{\sigma_i^2}} \quad (20)$$

【0 5 7 9】

母体の倍数性に関する項 M_i は、一部の数学的導関数から省略され得る。Xに関する得られた表示は、母親が評価の目的となる染色体または複数の染色体に欠失または重複を有さないような、比較的単純で、そして多くの場合最も頻繁に生ずる特殊なケースに対応する。

20

等式 2 1：

【化 1 4】

$$X = \frac{E_{fx} - (1-F)E_{ff}}{F E_{ff}} = \frac{E_{fx}}{F E_{ff}} - \frac{1-F}{F} = 1 + \frac{1}{F} \left(\frac{E_{fx}}{E_{ff}} - 1 \right) \quad (21)$$

【0 5 8 0】

X_{iff} および X_{ify} は、等式 (1 1) および (1 2) より、それぞれ与えられる。全ての実験誤差が無視し得る実施形態では、等式 (2 1) を解くことにより、 $X_{iff} = X_{ify}$ の場合、正倍数体について1の値が得られる。全ての実験誤差が無視し得るある特定の実施形態では、等式 (2 1) を解くことにより、三倍体について $3/2$ の値が得られる (X_{iff} と X_{ify} の間の三倍体の関係については等式 (1 5) を参照)。

30

【表 2】

表 2

妊娠状態	胎児 Chr21	胎児 Chr18	胎児 Chr13	胎児 ChrX	胎児 ChrY
雌 T21	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
雌 T18	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
雌 T13	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 0$
雄 T21	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
雄 T18	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
雄 T13	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
雄正倍数体	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1/2$
ターナー	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 0$
ジェイコブス	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$	$P_{ij}^F = 1$
クラインフェルター	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1/2$
トリプルX	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 1$	$P_{ij}^F = 3/2$	$P_{ij}^F = 0$

【0581】

(実施例3)

デシジョンツリー分析

【0582】

未報告またはこれまで未知の染色体異数性を含め、あらゆる染色体について胎児の染色体異数性を検出することができるデシジョンツリー法を開発した。さらに、ゲノム内の不均一なカバレッジイベントが、正規化手順（例えば、PERUN）適用後に検出可能である。不均一なイベントは、図1に示す通り（左上のパネル）、微小欠失／重複を示し得る。これらのイベントを独立に検出することができる。

方法

【0583】

事前に定義されたゲノム座標を有するT21、T18、およびT13を検出する場合とは異なり、不均一なカバレッジイベントは、ゲノムのどこでも生じ得る。事前に定義された位置で遺伝子の変動を検出するには、示された位置において有意レベルを決定することが必要となるだけである。この実施例に記載されているアルゴリズムは、部分カウント／カバレッジが一貫して上昇または低下している領域について調査し、かかるイベントの境界を正確に決定する。この実施例に記載されるものとして、2つの直交法の検出力を利用する方法がある。

ウェーブレット分解

【0584】

第1のアルゴリズムは、ウェーブレット変換を活用する。ウェーブレット変換は、シグナル処理の目的に特に有用な数学的なツールである。この改良されたアプリケーションでは、全ゲノム配列決定データを最初に整列、区分し、GCの偏りを除去するために正規化した。GCの偏りを低減するのに、PERUN正規化（本明細書に記載する）を利用したが、他のGCの偏り低減処理も利用できる（例えば、本明細書に記載するCHAI処理）

。その後、データ内のノイズを低減するために、ウェーブレット平滑化法を、正規化したプロファイルに適用し、それにより、微小欠失／微小重複イベントが明確に見えるようにした。ウェーブレット法に関する箱髭図を図1に示す。

【0585】

標準的なハールウェーブレットをウェーブレット分解に使用した。原則として、より複雑なウェーブレット変換も、平滑化後、イベントの場所を同定するのに使用することができる。ノイズからシグナルを区別するためには、どのウェーブレット係数がシグナルを示し、保持されるべきか、またどれがノイズの反映であり得、除去されるべきか決定する必要がある。このステップは、閾化と呼ばれる。公知のように、マグニチュードが大きくレベルが低い係数は、シグナルの傾向を保持し、一方マグニチュードが小さくレベルが高い係数は、シグナルの詳細を保持する。小さく有意でない係数を除去するのに、「ソフトな」閾化法を使用した [DonohoおよびJohnstone、(1995年) WaveLab and Reproducible Research]。閾化後、高いレベル係数が一部残留する可能性がある。これらの係数は、オリジナルのシグナル中の急激な変化または大きなスパイクを表し、除去される。このステップは「レベル化」と呼ばれ、図2は閾化を行わないレベル化の効果を示す(例えば、より多くの詳細が保持され、レベルが高まる)。レベル化を最適に選択するには、多くの要因、例えば染色体の長さ、検出するのに所望のイベント長、および正規化されたプロファイルのノイズレベルに依存し得る。染色体の長さを N_{chr} (2の累乗に最も近い数まで延長される) およびウェーブレット分解レベルを c とすれば、ウェーブレットプロファイルの最低セグメント長は、 $L = N_{chr} / 2^{c+1}$ である。したがって、サイズ N_{micro} の微小欠失を検出するには、所望の分解レベルは、 $c = \log_2 (N_{chr} / N_{micro}) - 1$ である。例えば、参照ゲノムの部分 $N_{chr} = 4096$ 個、および微小欠失が有するサイズとして参照ゲノムの部分 $N_{micro} = 128$ 個である場合、分解レベルは、 $c = 4$ であるはずである。 $c \pm 1$ の分解レベルも利用できる。

サーキュラーバイナリセグメンテーション (CBS) 法

【0586】

ウェーブレット法は、微小欠失／微小重複の可能性のある場所を同定することができる。しかし、それ単独では、真のイベントの存在は保証されない。プロファイルが十分に正規化されない場合、ウェーブレットアルゴリズムは、GC残差により生じるローカルな変動により陥穽に陥る可能性がある。さらに、検出されたエッジの精度には、ウェーブレット係数のトランケーション次数により限界がある。偽陽性を低減するために、独立した方法を利用してウェーブレットの所見を検証した。サーキュラーバイナリセグメンテーション (CBS) は、アレイCGHデータを使用したコピー数の変動 (CNV) 検出のために当初提案された方法である。CBSは、変化のポイントを正確に指摘することができる。CBSは、尤度比統計量を使用して、染色体を等しいコピー数領域に繰り返し区切ることにより機能する [Olshen AB, Venkatraman ES, Lucito R, Wigler M., Biostatistics (2004年) 10月; 5巻(4号): 557~72頁]。CBSは一般的に良好に機能するが、シグナル内のノイズが高い場合、ゲノムを過剰に区切る傾向を有する [Lai, WR, Johnson, MD, Kucherlapati, R, Park, PJ, Bioinformatics, (2005年) 21巻, 19号: 3763~70頁]。この実施例では、PERUN—正規化部分カウントデータと共に機能するように構成されており、ウェーブレット所見を検証するための独立した方法として使用される。図5は、CBSアルゴリズムについて例示する。

ウェーブレットまたはCBS平滑化プロファイルに関するセグメントの統合

【0587】

ウェーブレットまたはCBSは、目的の染色体を等しいコピー数の領域／セグメントに区切る。各セグメントは、CNVの潜在的候補を代表する。これまでのセクションで議論したように、CBSは染色体を過剰に区切る傾向を有し、したがって広いスパンのCNV領域が、いくつかのより小さなピースに分離され得る。類似した状況は、ウェーブレット法についても生じ得る。図4はこれを説明し、この場合、大きな重複は、CBS法により3つのピースに当初区切られ、CNV幅は過小評価されることとなった(図4、左下のパ

ネル)。この欠点を克服するために、WillenbrockおよびFridlyand (Willenbrock H, Fridlyand J., Bioinformatics (2005年) 11月15日; 21巻(22号): 4084~91頁)により提案されたアルゴリズムを、ウェーブレットまたはCBS平滑化プロファイルのコピー数領域が等しいより長いストレッチにさらに統合するために適用した。2つのセグメントに対してマッピングされた部分カウントが、有意に異なる場合、または予測されるセグメント値が動的に決定された閾より接近している場合、これら2つのセグメントを統合する [WillenbrockおよびFridlyand、2005年]。図4の右下のパネルは、セグメントの統合の効果を説明し、セグメントの統合後に、明瞭な微小重複が認められるようになった。

ウェーブレットおよびCBSアルゴリズムに由来する統計量

10

【0588】

各染色体に、3つの主要なZスコア統計量が、ウェーブレット/CBS平滑化、セグメント統合化プロファイルから推測することができる。

(1) ウェーブレット平滑化プロファイルおよびその部分カウント表示に由来する最良のセグメント(候補セグメント)であるZスコア(Z_{wave})。候補セグメントの試料カウント表示は、セグメント内の正規化された総カウントを、試験試料に関する正規化された総常染色体カウントで除算したものである。候補セグメントに関するカウント表示の中央値は、正倍数体試料セットについて生成され、MADは候補セグメントに関する正倍数体カウント表示について決定される。セグメントに関する Z_{wave} 統計量は、試験試料のカウント表示ー正倍数体カウント表示の中央値であり、また減算した結果はMADで除算される。

20

(2) CBS平滑化プロファイルおよびその部分カウント表示に由来する最良のセグメントであるZスコア(Z_{cbs})。候補セグメントの試料カウント表示は、セグメント内の正規化された総カウントを、試験試料に関する正規化された総常染色体カウントで除算したものである。候補セグメントに関するカウント表示の中央値は、正倍数体試料セットについて生成され、MADは、候補セグメントに関する正倍数体カウント表示について決定される。セグメントに関する Z_{cbs} 統計量は、試験試料のカウント表示ー正倍数体カウント表示の中央値であり、また減算した結果はMADで除算される。

(3) 染色体全体(Z_{chr})に関する全体的な染色体の表示。試料カウント表示は、候補セグメントが内在する染色体内の正規化された総カウントを、試験試料に関する正規化された総常染色体カウントで除算したものである。染色体に関するカウント表示の中央値は、正倍数体試料セットについて生成され、MADは、正倍数体カウント表示について決定される。染色体に関する Z_{chr} 統計量は、試験試料のカウント表示ー正倍数体カウント表示の中央値であり、減算した結果は、MADで除算される。

30

【0589】

最良のセグメントは、染色体上の全セグメントの中でも曲線下面積(AUC)が最も大きなセグメントである場合もある。かかるセグメントは、目的の染色体上で最も重要な所見を代表する。例えば、図4の右下のパネルは2つのセグメントを有するが、両者のうち第2のセグメントは最も大きなAUCを有する。図5は、各染色体についてのウェーブレットおよびCBS平滑化、ならびに導出した統計量について要約する。

40

CNV検出に関するデシジョンツリー

【0590】

染色体のそれぞれについて3つの主要な統計量を計算したら、統計量は、所与の試料中のトリソミー、微小欠失、または微小重複の存在を決定するのに使用することができる。デシジョンツリーを以下に示す:

1. 下記の場合、染色体はトリソミーまたはモノソミーとして分類される:

a. $|Z_{chr}| \geq 3.95$ 、および

b. $|Z_{chr}| \geq \min(|Z_{wave}|, |Z_{cbs}|)$

2. 下記の場合、染色体は微小欠失/微小重複を有するものと分類される:

a. トリソミーまたはモノソミーに該当しない

50

- b. $|Z_{wave}| \geq 3.95$ 、および $|Z_{CBS}| \geq 3.95$
 c. ウェーブレットおよびCBSの最良セグメントが重複する

【0591】

条件1では、染色体全体のZスコアが有意であること、およびそのマグニチュードが、ウェーブレットまたはCBSいずれかの最良セグメントに匹敵すべきことが、本質的に必要であることに留意すること。条件2では、ウェーブレットおよびCBSの最良セグメントの両方が有意であることが必要であり、両セグメントが相互に重複しており、所見が相互確認される必要がある（図5）。場合によっては、CBSおよびウェーブレット法により同意亭されたウェーブレット事象は重複せず（図9）、微小重複または微小欠失が存在しないことが示される。ほとんどの適用例では、Zスコアカットオフ（例えば、所定の閾

10

結果

【0592】

本検出方法を、第22染色体上に3MB未満で延在する微小欠失22q11の2つの症例に適用した（図10）。3MBの分解能を実現するために、試料を0.5-plexで最初に配列決定した。カバレッジを1/10に下げても、22q11微小欠失が検出された（図10、F'）。1つの試料の検出結果を図10に示すが、強調された領域は、微小欠失イベント（約2.5MB）を示す。

20

【0593】

アルゴリズムを、異なる試験に由来する試料にも適用し、微小欠失／微小重複に関する19例の推定症例が検出された。検出された症例のうち2例を図11Aおよび11Bに示す。

（実施例4）

最大エントロピー

【0594】

最大エントロピーは、本明細書に記載する自動化されたアルゴリズムであり、ゲノムを均一なレベルのセグメントに区切る。このアルゴリズムは、ヒトゲノムのGC含有量分布を特徴付けるために、Cohenら（Cohen N, Dagan T, Stone L, Graur D., (2005 年) Mol. Biol. Evol., 5月; 22巻(5号): 1260~72頁）が使用した手順に基づく。Cohenの方法は、微小欠失／重複の場所を検出する目的にふさわしい（例えば、微小欠失／微小重複のエッジ）。修正事項には、セグメンテーションを最低の長さに制限する能力、およびセグメンテーションを終了するのに使用されるt-値に基づく判定基準が含まれた。t-値は、染色体（例えば、セグメント、区切り、または領域）内で同定された全てのセグメントが均質であるか決定するのに使用される。新しく同定されたセグメントが、t-値に基づく有意検定で不合格となった場合、セグメントは再度統合されて、セグメンテーションは停止する。これらの操作を行ったコードは、下記に含まれる：

30

【化 1 5】

```
# PATH: C:\SEQUENOM_PROJECTS\2013\2013_01_10_PARTITIONING_MAX_ENTROPY\  
# FILE: PARTITIONING_MAX_ENTROPY.r  
# AUTHOR: Zeljko Jovan Dzakula  
# See: Cohen et al. (2005): GC Composition of the Human Genome: In Search  
of Isochores. Mole Biol. Evol. 22(5):1260-1272.  
# DATE: 1/10/2013
```

```
rm( list=ls() );
```

```
shannonEntropy <- function( values ) {  
  meanValue <- mean( values, na.rm=TRUE );  
  entropy <- -meanValue * log2( meanValue ) -  
    ( 1 - meanValue ) * log2( 1 - meanValue );  
  return( entropy );  
} # shannonEntropy
```

10

```
partitionShannonEntropy <- function( values, edge ) {  
  nValues <- length( values );  
  if ( edge < 1 | edge >= nValues ) {  
    return( NA );  
  } # if edge  
  overallEntropy <- shannonEntropy( values );  
  leftEntropy <- shannonEntropy( values[ 1:edge ] );  
  rightEntropy <- shannonEntropy( values[ ( edge + 1 ):nValues ] );
```

20

【化 16】

```

    return( overallEntropy - ( leftEntropy * edge + rightEntropy * (
nValues - edge ) ) / nValues );
} # partitionShannonEntropy

findBestEdge <- function( values ) {
  epsilon <- 1e-6;
  bestEdge <- 1;
  bestEntropy <- -1e7;
  values <- values[ !is.na( values ) ];
  minValue <- min( values, na.rm=TRUE ) - epsilon;
  maxValue <- max( values, na.rm=TRUE ) + epsilon;
  values <- ( values - minValue ) / ( maxValue - minValue );
  nValues <- length( values );
  for ( candidateEdge in 1:( nValues - 1 ) ) {
    candidateEntropy <- partitionShannonEntropy( values,
candidateEdge );
    #print( paste( "", candidateEdge, candidateEntropy ) );
    if ( candidateEntropy > bestEntropy ) {
      bestEdge <- candidateEdge;
      bestEntropy <- candidateEntropy;
    } # if candidateEntropy
  } # for candidateEdge
  return( bestEdge );
} # findBestEdge

isPartitionHomogeneous <- function( values, edge, cutoff=10 ) {
  nValues <- length( values );
  if ( edge < 1 | edge >= nValues ) {
    print( paste( "Illegal argument: edge=", edge, sep="" ) );
    return( TRUE );
  } # if edge
  leftMean <- median( values[ 1:edge ], na.rm=TRUE );
  leftMAD <- mad( values[ 1:edge ], na.rm=TRUE );
  rightMean <- median( values[ ( edge + 1 ):nValues ], na.rm=TRUE );
  rightMAD <- mad( values[ ( edge + 1 ):nValues ], na.rm=TRUE );
  tValue <- abs( leftMean - rightMean ) / sqrt(
    leftMAD * leftMAD / edge + rightMAD * rightMAD / ( nValues - edge
) );
  print( tValue );
  return( tValue < cutoff );
} # isPartitionHomogeneous

partitionRegions <- function( values, regions, minLength, cutoff=10 ) {
  nValues <- length( values );
  rownames( regions ) <- c( "Start", "End", "Homogeneous",
"MedianLevel", "MAD" );
  finished <- FALSE;
  while ( !finished ) {
    nRegions <- ncol( regions );
    for ( region in 1:nRegions ) {
      if ( regions[ "Homogeneous", region ] == FALSE ) {
        startPoint <- regions[ "Start", region ];

```

【化 17】

```

        endPoint <- regions[ "End", region ];
        relativeEdge <- findBestEdge( values[ startPoint:endPoint
] );

        absoluteEdge <- startPoint + relativeEdge - 1;
        homogeneous <- isPartitionHomogeneous(
            values[ startPoint:endPoint ], relativeEdge,
cutoff=cutoff );
        if ( ( homogeneous == TRUE ) |
            ( edge == endPoint ) |
            ( relativeEdge < minLength ) |
            ( endPoint - absoluteEdge ) < minLength ) {
            print( paste( "Homogeneous: ", startPoint, "-",
absoluteEdge, "-", endPoint, sep="" ) );
            regions[ "Homogeneous", region ] <- TRUE;
        } else {
            print( paste( "Inhomogeneous: ", startPoint, "-",
absoluteEdge, "-", endPoint, sep="" ) );
            regions[ , region ] <- c( startPoint, absoluteEdge,
FALSE, NA, NA );
            regions <- cbind( regions,
                c( absoluteEdge + 1, endPoint, FALSE, NA, NA ) );
        } # if homogeneous else
    } # if !homogeneous
} # for region
finished <- TRUE;
nRegions <- ncol( regions );
for ( region in 1:nRegions ) {
    finished <- finished & regions[ "Homogeneous", region ];
} # for region
if ( finished ) {
    regions <- t( regions );
    nRegions <- nrow( regions );
    rownames( regions ) <- 1:nRegions;
    for ( region in 1:nRegions ) {
        startPoint <- regions[ region, "Start" ];
        endPoint <- regions[ region, "End" ];
        regions[ region, "MedianLevel" ] <-
            median( values[ startPoint:endPoint ], na.rm=TRUE );
        regions[ region, "MAD" ] <-
            mad( values[ startPoint:endPoint ], na.rm=TRUE );
    } # for region
    return( regions );
} # if done
} # while TRUE
} # partition

partition <- function( values, minLength, cutoff=10 ) {
    nValues <- length( values );
    regions <- data.frame( "InitialEdges"=c( 1, nValues, FALSE, NA, NA )
);
    #regions <- data.frame( "FirstHalf"=c( 1, floor( nValues / 2 ),
FALSE, NA, NA ) );

```


【表 3】

表 3. 最大エントロピー法により検出された微小重複および微小欠失イベント

検出されたイベント	染色体
微小欠失	5
微小重複	22
微小欠失	22
微小重複	10
微小重複	21
微小重複	12
微小欠失	15
微小欠失	1
微小重複	19
微小重複	2
微小欠失	7
微小欠失	6
微小欠失	4
微小重複	6
微小重複	4

10

20

30

【0597】

(実施例5)

検証法

【0598】

「リーブワンアウト」および「スライディングエッジ」と呼ばれる2つの検証法を、P E R U Nと連結させて、口蓋心臓顔面症候群（ディジョージ症候群、22q11）として公知の、第22染色体上のサブ染色体異常を同定し、22q11領域で観察された正倍数体（すなわち「正常」）染色体レベルからの偏差の統計的有意性を評価した。この方法論を、欠失／重複について標的化した検出および標的化しない検出の両方に適用した。

40

【0599】

16個の妊娠中の雌の血漿試料を、これまでの記載（Jensen TJ、Dzakula Z、Deciu C、van den Boom D、Ehrich M.（2012年）：Detection of microdeletion 22q11.2 in a fetus by next-generation sequencing of maternal plasma.、Clin.Chem.、58巻（7号）：1148～51頁）に従い収集、処理、および配列決定した。核型分類により確認されたように、2つの試料が、ディジョージ胎児を身ごもった母親に属する。残りの14個の試料は正倍数体妊娠（すなわち、「正常な胎児」）に対応し

50

、22q11の検出および特徴付けのための参照として用いる。

【0600】

16個の試料全てに関するオリジナルの未処理のカウント（1試料当たり2データセット）を、ELANDアラインメントに最適化したPERUN部分パラメータを使用して再処理した。パラメータをLDT2CEデータ上でトレーニングし、部分の選択は、これまでに記載したように、相互の検証に基づいた。マッピング可能性に基づくフィルタリングは適用しなかった。二次LOESS正規化により、PERUN正規化後に残留するあらゆるGCの偏りが除去された。各試料について2回測定が行われたので、全部で32個のプロファイルが得られたが、対をなすプロファイルを単一のPERUNプロファイルに併合し、16個のプロファイル（1試料1個当たりプロファイル）を得た。プロファイルを併合する前は、その標準偏差は0.020～0.030の範囲であった。一致したプロファイルを加えることにより、可変性は約1/1.2（約1/1.14～1/1.27）に低下したが、期待した改善の1/1.414（2の平方根）ほどではなかった。正規化されたカウントプロファイルは、未処理およびGCRMプロファイル、ならびに14個の正倍数体試料に基づく参照プロファイル中央値について行った正規化の結果と比較して有意に改善した。プロファイルの標準偏差が低下し、またゲノム全体を通じて均一度が向上したことから、改善は明白である。一方、16個の未処理のプロファイルに関する標準偏差（総カウントに応じてスケール調整し、参照ゲノムの部分の数を乗算した）は、0.55～0.64の範囲であるが、対応する16個のPERUNプロファイルの標準偏差は、0.016～0.026の範囲である。図13は、全16個のプロファイルを示し、ディジョージ領域（chr22_368～chr22_451の範囲）は、背景をグレーリボン付きにして表す。交差検証に基づく部分のフィルタリングにより、参照ゲノムの多くの部分が22q11微小欠失から除去されており、参照ゲノムの部分の下記のセットのみが残されていることに留意すること：chr22_371、chr22_372、chr22_380、chr22_381、chr22_382、chr22_383、chr22_384、chr22_385、chr22_386、chr22_387、chr22_388、chr22_389、chr22_390、chr22_391、chr22_392、chr22_393、chr22_394、chr22_395、chr22_396、chr22_397、chr22_398、chr22_399、chr22_400、chr22_401、chr22_402、chr22_403、chr22_404、chr22_415、chr22_416、chr22_417、chr22_418、chr22_419、chr22_422、chr22_423、chr22_424、chr22_426、chr22_427、chr22_428、chr22_439、chr22_440、chr22_441、chr22_442、chr22_443、chr22_444、chr22_445、およびchr22_446。

【0601】

図13は、詳細な目視検査が可能となるように、ディジョージ領域内のPERUNプロファイルを拡大して示す。22q11欠失が、罹患症例（3_4および9_10）において明白である。

【0602】

22q11欠失の存在または非存在に関する判定の信頼度を定量するために、Zスコアを、第22染色体の位置18,546,349～22,336,469をカバーするカニックディジョージ領域について評価した。この領域内の参照ゲノムの部分に関するPERUNレベルを、試料毎に個別に合算した。全ての試料について、技術的複製物を測定したので、ディジョージ領域内の染色体物質の表示として2つの合計の平均を使用した。全ての表示（正倍数体および罹患試料の両方を含む）の中央値を個々の表示から差し引き、差異を全ての表示のMADで除算してZスコアを得た。結果を図14に示す。罹患症例3_4および9_10の2例は-3未満のZスコアを有し、これら2試料中に欠失が存在することが確認された。さらに、試料（13_14）の1例では、正の高いZスコアが認められ、目的領域における重複の可能性が示された。図13の目視検査により、13_1

4 P E R U N プロファイルの中央部分が、ディジョージ領域内の過剰表示区分を含有することが確認された。

【0603】

図13に示すプロファイルから、罹患症例(3__4)の1例に認められた欠失は、カノニックディジョージ欠失とは異なることが示唆された。3__4プロファイルは、一部分欠損しているにすぎず、欠失の右側エッジは、異常部分の予想されたエッジの左側に、1 M b p (20個の参照ゲノムの部分)を超えて、部分c h r 2 2 __4 2 6に接近して位置した。また、3__4内の欠失の左側エッジは、2 2 q 1 1の予想される左側エッジよりも左側にシフトしているようでもあった。欠失の真の範囲は、明らかにZ値に影響を及ぼした。本試験の第1の目標は、欠失/重複の開始エッジと最終エッジの場所が既知として、欠失/重複の判定の信頼度を評価することであった。

10

【0604】

欠失のエッジの場所の影響を評価し、信頼区間をZスコアと関連付けるために、カノニックディジョージ領域と一部分重複する(またはこれに含有される)225個の異なる領域の表示に基づき、Zスコアの評価を繰り返した。これらの領域のうち最大の領域は、c h r 2 2 __3 7 1より開始し、c h r 2 2 __4 4 6で終了した。欠失の最後の部分のc h r 2 __4 4 7は、P E R U N交差検証の確認によりフィルター除去されているので、計算に含まれなかった。残りの領域は、カノニックディジョージ領域の左側エッジから第1番目と第15番目の部分の間の任意の箇所で開始した。さらに、領域は、カノニックディジョージ領域の右側エッジとそれに先行する第15番目の部分の間の任意の箇所で終了した。これは、15×15グリッドの領域開始/終了ポイントを生み出した。全試料の表示をこのグリッド上で評価し、グリッドポイント毎にZ値を取得した。全ての試料が、Z-標準化で使った中央値およびMAD表示に寄与した。225個の考え得る領域(試料3__4および1__2に関する)全てについて得られたZ値の代表的なヒストグラムを図15~16に示す。

20

【0605】

図17の散布図は、全試料に関する225個の考え得る領域全てについて、得られたZ値のヒストグラムを要約する。図17は、試料当たりのZ値中央値およびそのZスコア中央値周辺の3つのMAD信頼区間のみを示す。欠失エッジの選択に依存して、試料13__14は、概ね、過剰表示しているように見えた(15×15グリッド内のほとんどの領域について、Z>3)。2つの罹患試料(3__4および9__10)に関する3つのMAD信頼区間は、「正常」領域(-3~3)と一部分重複するが、両試料は概ね-3未満のままであった。

30

【0606】

図20では、微小欠失エッジの選択に由来する可変性について調べた。別の考え得る可変性の原因として、Zスコア標準化のための参照試料の選択があり得る。Zスコアにおける可変性に対する参照試料の選択の寄与、およびこれに起因する欠失/重複判定の信頼度(または信頼度の欠如)を評価するために、単一の領域、ならびに15×15グリッドの領域について「リーブワンアウト」分析を行った。カノニックディジョージ領域のみに適用したとき、「リーブワンアウト」分析では、参照の選択がZスコアの可変性に対して有意に寄与することを示すことができなかった(図18~20)。しかし、より徹底した「リーブワンアウト」分析を、15×15グリッドの領域について行い、参照試料の選択は、Zスコアの可変性に対して有意に影響を及ぼすことを確認した(図21~23)。

40

【0607】

図21では、15×15グリッドのディジョージサブ領域を使用して得られたZ値中央値と、「リーブワンアウト」技法により生成されたZ値中央値との間の一致性が実証された。

【0608】

図22~24では、15×15グリッドのサブ領域からランダムに取り出されたディジョージサブ領域への「リーブワンアウト」技法の使用について提示した。カノニックディ

50

ジョージ領域の場合と同様の結論が導かれ得る。

【0609】

図25～29は、「リーブワンアウト」技法と領域エッジのスライディングの組合せの結果を示す。15×15サブ領域のそれぞれに、16個の参照セットは、1試料当たり3,600セットのZスコアに併合されたZスコアを生成した。代表的な図25～26に示す分布は、追加の詳細を提供する一方、図14～24から導かれた結論と一般的に一致した。一部のヒストグラムに認められる鋭いピークを説明し得る事実として、試料のサブセットは、Zスコアの標準化に関する中央値に高頻度で寄与するということが挙げられる。

【0610】

「スライディングエッジ」および「リーブワンアウト」の方法は、判定を検証するのに、同時にまたは独立して使用することができる。また2つの異なる手順は、導出されたZスコアの不確実性についてより徹底した実像を得るのにも併用される。2つの技法は、本明細書では、標的化した欠失／重複の検出において適用されるが、ゲノム全体を通じて、これまでに未知の欠失／重複を目的とした、標的化されない「調査」まで原則として拡張可能である。ウェーブレット法、最大エントロピー法、サーキュラーバイナリセグメンテーション法、エッジ検出カーネルによるコンボリューション、または他のいくつかの適する方法を使用して罹患エリアを概略把握したら、スライディングエッジ技法およびリーブワンアウト技法が、新たに検出された欠失／重複の範囲および信頼性を確認するのに適用可能である。

【0611】

スライディングエッジ分析を行おうとする動機は、患者3__4に認められた22q11欠失の範囲とカノニックディジョージ欠失の範囲との間で認められた不一致に起因する。結局のところ、3__4試料に認められた2.5Mbの欠失は、この計算が行われた時に既に知られていた欠失よりも代表的であった。典型的な欠失は、下記の2つの出典によるカノニック8Mbよりはむしろ約3Mbである：

(C.Carlsonら、(1997年) The American Journal of Human Genetics、61巻、3号、620～629頁)

(Schwinger E、Devriendt K、Rauch A、Philip N.、Clinical utility gene card for: DiGeorge syndrome、velocardiofacial syndrome、Shprintzen syndrome、chromosome 22q11.2 deletion syndrome、(22q11.2、TBX1). Eur J Hum Genet.、2010年9月；18巻(9号)。doi: 10.1038/ejhg.2010年5月Epub 2010年2月3日.PubMed PMID: 20125192; PubMed Central PMCID: PMC 2987430.)

【0612】

ディジョージ欠失は、若干7～8%の症例では、1.5Mbほどの短さで報告されている。染色体異常の予想されるサイズと、(高い)臨床的意義を有する実際のサイズとの間のかかる解離は、現在のデータが示唆する頻度よりも高頻度であり得る。そのような理由から、特定の異常に特別に対応した標的化した方法、および最初に異常を発見し、次に臨床的アノテーションのデータベースに照会する、より一般的な標的化しないアプローチの両方が、スライディングエッジおよびリーブワンアウト分析から利益を得る。

(実施例6)

遺伝子の変動の対数オッズ比検出

【0613】

胎児フラクションは、遺伝子の変動に関する非侵襲的出生前試験(NIPT)においてその役割を演ずる。胎児フラクションが高い(例えば、24%)が、Zスコアがわずかに高め(例えば、 $z=3.2$)の試料は、偽陽性分類となり得ることが認められた。かかる試料の場合、例えばそれが真にトリソミーである場合、Zスコアは3よりかなり大きいはずである。この問題に対処し、偽陽性の判定が生ずる可能性を抑えるために、対数オッズ比(LOR)処理が開発された。

【0614】

LORは、等式22に従って、試験試料について、その観察されたZスコアおよび胎児

10

20

30

40

50

フラクシオンを既知として、トリソミー 21 (T21) と非 T21 の確率により計算可能であり、：

【化 19】

$$\log \frac{P(T21|Z, \hat{f})}{P(\overline{T21}|Z, \hat{f})} = \log \frac{P(Z|T21, \hat{f})P(T21, \hat{f})}{P(Z|\hat{f})P(\overline{T21}, \hat{f})} = \log \frac{P(Z|T21, \hat{f})P(T21)}{P(Z|\overline{T21}, \hat{f})P(\overline{T21})} \quad (22)$$

式中、

【化 20】

$\hat{f}$

10

は、本明細書では $\hat{f}$ とも表わされ、測定された胎児フラクシオンであり（例えば、雄試料の場合 chr Y、または公知の他の胎児フラクシオン測定技法による）、

【化 21】

$$P(T21|Z, \hat{f}) \text{ および } P(\overline{T21}|Z, \hat{f})$$

は、それぞれ

【化 22】

$$T21 \text{ (T21を有する) および } \overline{T21} \text{ (非T21)}$$

20

の事後確率であり、以下に記載するように Z および $\hat{f}$ が既知として、

式中、

【化 23】

$$P(T21) \text{ および } P(\overline{T21})$$

は、それぞれ

T21 および 非 T21 に関する事前確率であり、

30

【化 24】

$$P(Z|T21, \hat{f}) \text{ および } P(Z|\overline{T21}, \hat{f})$$

は、それぞれ T21 および 非 T21 に関する条件付き確率であり、本明細書において導出される。

【0615】

条件付き確率は、試験試料について、z スコア (Z) および計算して得た胎児フラクシオン ($\hat{f}$) により決定することができる (等式 22 の右側の部分を参照)。罹患していない正倍数体試料の場合、X はイベント領域に関するビンカウントの合計を表すものとする。配列決定には本質的にランダム性が内在するので、X は $X \sim f(\mu_X, \sigma_X)$ のランダム変数であり、この場合、 μ_X および σ_X はそれぞれ平均値および標準偏差であり、 $f(\cdot)$ は分布関数である。同様に、罹患したトリソミー試料の場合、罹患領域に関するビンカウントは、 $Y \sim f(\mu_Y, \sigma_Y)$ であり、この場合、 $\mu_Y = \mu_X (1 + f/2)$ であり、 f は胎児フラクシオンである。 $\sigma_Y = \sigma_X$ と仮定すれば、z スコア分布は、等式 23 のように記載することができる：

40

【化 2 5】

$$Z \sim \text{正規} \left(\frac{\mu_x f}{\sigma_x^2}, 1 \right) \quad (23)$$

式中、 μ_x および σ_x は、大規模なプールの正倍数体試料から経験的に評価され得る。

【0616】

正倍数体試料の場合、その z スコアは胎児フラクションとは独立であり、標準正規分布に従う。

【0617】

等式 23 を図 32 に図式的に示す。正倍数体の被験体に関する z スコア分布は、ゼロを中心とするが、分布の中心は、胎児フラクションに対して感度を有さない。等式 23 を図式的に表すと、T21 に関する個々の z スコア分布それぞれの中心は異なる z スコア上に位置するが、この場合、個々の分布それぞれは異なる胎児フラクションに対応する。試験試料について決定した胎児フラクションを適用すれば、T21 についてどの z スコア分布を評価するかが決定される。試験試料について決定した z スコアを適用すれば、T21 について条件付き確率が特定される。T21 に関する条件付き確率は、試験試料について決定した z スコアと、当該試験試料について決定した胎児フラクションにより選択された T21 に関する z スコア分布との交点である。試験試料について決定した z スコアを適用すれば、非 T21 に関する条件付き確率も特定される。非 T21 に関する条件付き確率は、試験試料について決定した z スコアと非 T21 に関する z スコア分布との交点である。

【0618】

T21 および非 T21 に関する条件付き確率を等式 22 に代入すると、試験試料に関する LOR 計算結果が得られる。3.95 を上回る z スコアおよびゼロを上回る LOR を有する試験試料は、遺伝子の変動が存在する（例えば、T21 が存在する）ものとして分類される。3.95 未満の z スコアおよび／またはゼロ未満の LOR を有する試験試料は、遺伝子の変動が存在しないものとして分類される。

【0619】

図 31 は、LOR 法を使用した、LDT v2 雄試料に関する分類結果を示す。このように、胎児フラクションが増加するに従い、LOR は急速に発散する。したがって、胎児フラクションが大きい境界線上の試料でも確実に分類可能であり、またそうであった。特に、ゼロを上回る LOR を有する白丸で表わされた試料は、正確に T21 として分類され、ゼロ未満の LOR を有する白丸で表わされた試料は、正確に非 T21 として分類された。

【0620】

等式 22 は、多くの場合該当しないが、測定された胎児フラクションは真の胎児フラクションと同一であることを仮定する。測定の不確実性を補償するために、等式 24 に従って、発展型の LOR 法を開発した：

【化 2 6】

$$\begin{aligned} \frac{P(T21|Z, \hat{f})}{P(\bar{T}21|Z, \hat{f})} &= \frac{P(T21, Z, \hat{f})}{P(\bar{T}21, Z, \hat{f})} = \frac{\sum_f P(T21, Z, f, \hat{f})}{\sum_f P(\bar{T}21, Z, f, \hat{f})} \\ &= \frac{\sum_f P(Z|T21, f, \hat{f}) P(T21|f, \hat{f}) P(f|\hat{f}) P(\hat{f})}{\sum_f P(Z|\bar{T}21, f, \hat{f}) P(\bar{T}21|f, \hat{f}) P(f|\hat{f}) P(\hat{f})} \\ &= \frac{\sum_f P(Z|T21, f, \hat{f}) P(f|\hat{f})}{\sum_f P(Z|\bar{T}21, f, \hat{f}) P(f|\hat{f})} \frac{P(T21)}{P(\bar{T}21)} = \frac{\sum_f P(Z|T21, f) P(f|\hat{f})}{\sum_f P(Z|\bar{T}21, f) P(f|\hat{f})} \frac{P(T21)}{P(\bar{T}21)} \quad (24) \end{aligned}$$

式中、 f は真の胎児フラクションであり、 $\hat{f}$ は測定された胎児フラクションである。等

式(22)と比較して、等式(24)は、 f^h が既知として f の条件付き確率について平均化する。発展型のLORは、こうして条件付き確率の重みつき平均の使用を可能にし、測定された胎児フラクションの値により近い考え得る胎児フラクションの値に対してより大きな重みを付ける。したがって、等式24に基づき、発展型のLORは真の陽性試料を回収した。したがって、発展型のLORは、試験試料中の遺伝子の変動の存在または非存在を決定するのに使用することができる。

【0621】

LOR法は、複数の種類の染色体異数性(例えば、T21以外)の存在または非存在を決定するのに適用可能であり、また複数の種類のその他の遺伝子の変動(例えば、第21染色体以外の染色体の染色体異数性、微小重複、微小欠失)の存在または非存在を決定するのに適用され得る。Zスコアが3.95に等しいまたはそれ超、およびLORがゼロを上回るとき、陽性イベント(例えば、染色体異数性(例えば、モノソミー、トリソミー)の存在;微小重複の存在、微小欠失の存在)と決定される。微小重複または微小欠失の存在または非存在を決定する場合、例えばウェーブレット平滑化分解レンダリングまたはCBS平滑化分解レンダリングにそれぞれ起因する Z_{wave} または Z_{CBS} のZスコア値が利用可能である(例えば、実施例3を参照)。図32に示す関係は、微小欠失または微小重複について類似するが異なる。微小重複イベントの場合、重複セグメントに関する μ_x および σ_x は、染色体トリソミーの場合より値は小さく、図32では、微小重複のZスコア分布は、正倍数体Zスコア分布に近づくように左側にシフトする。図32では、微小重複イベントに関するZスコア分布は、正倍数体Zスコア分布の右側に留まる。しかし、微小欠失イベントに関するZスコア分布は、図32では、正倍数体Zスコア分布の左側にシフトする。染色体異数性の存在または非存在を決定する場合(例えば、第21染色体以外の染色体について)、例えばZスコア値、 Z_{chr} が利用可能である(例えば、実施例3を参照)。

(実施例7)

ChAI正規化処理

【0622】

ChAIは、試験被験体(例えば、妊娠中の雌)から得られた配列の読取りから、胎児内の遺伝子の変動(例えば、染色体異数性、微小重複、微小欠失)の存在または非存在を決定するのに使用することができるシステムである。ChAIに関するシステムのフロー図の例を、図42Aおよび42Bに示す。妊娠中の雌の試験被験体、および本明細書でトレーニングセットと呼ぶときもある、1例または複数例の参照被験体から、配列の読取りを取得した。トレーニングセットの妊娠中の雌被験体は、他の試験法により正倍数体であることが確認された胎児を有した。

【0623】

配列の読取りを、ChAIが一層迅速に稼働できるように、まずSAMまたはBAMフォーマットからバイナリ読取りフォーマット(BReadフォーマット)に圧縮した。BReadフォーマットは、染色体および参照ゲノムにより決定された塩基対の位置を含め、読取り毎にゲノムの場所を保管し、他の情報を廃棄する。BReadファイルは、含有される読取りのカウントから開始する。これにより、メモリのリロケーションが不要となり、ローディング時間が改善する。値をディスク上に4バイトアレイとして保管した。読取りを、次に5バイトフォーマットを使用して保管し、1つは染色体序数(1~22、X、Y、Mのゼロインデックス)、4つは染色体位置に関する。最初の4バイトから配列の読取りカウントを最初に読み取って、BReadファイルをロードした。各配列の読取りを、次に一度に5バイトずつロードし、最初のバイトは染色体序数を示し、次の4バイトは整数位置に変換する。読取りのランダムサンプリングは、特定の読取りインデックスに対するディスクスキップコマンドを使用することにより、迅速に行うことができる。

【0624】

例として、マッピングされた読取り、17,673,732について、異なるフォーマットのディスク使用状況を、BReadフォーマットのディスク使用状況と表Iで比較す

る。

【0625】

【表4】

表 4: 17,673,732 の読取りを有する試料に基づき、異なるフォーマットのディスク使用状況。

フォーマット	使用スペース
SAM	4.0 GB
マッピングされた読取り位置	247 MB
GZip 読取り位置	97 MB
BReads	85 MB

10

【0626】

BReadフォーマットは、オリジナルのSAMファイルよりも約1/50ほど小さく、その使用スペースは、GZipフォーマットよりも約12%小さい。BReadは、ワнтаイムメモリアロケーションの冒頭部で、読取り数を保管するという長所も有し、また読取りは順番に読み取られる必要はないので、迅速にサンプリング可能である。これらの特性は他のフォーマットでは不可能であった。

20

GCの偏りのモデリング

【0627】

各試料についてGCの偏りモデルを次に習得した。部分フィルターを作成するために、およびGCの偏り単独では十分に説明されない他のゲノムの偏りを習得するために、トレーニング用に指定された試料を、一部分使用した。最終的に、トレーニング統計量を、試験試料をフィルタリングおよびスコア化するのに使用した。

【0628】

ChAIにより、ローカルなGC含有量の密度推定値を使用して、GCの偏りをモデル化した。Epanechnikovカーネル等のカーネル機能を使用して、参照ゲノムからGC密度を見積もった(図33)。ガウシアンまたはトリウェイトカーネル(triweight kernel)を含む、他のカーネルも適する。バンド幅を200bpとして選択したが、バンド幅パラメータはフレキシブルである。

30

【0629】

カーネルを使用して、参照ゲノムについてGC密度を塩基対分解能で見積もった(例えば、図34に示す通り)。参照のGC密度推定値を使用して、試料に由来する各読取りのローカルなGC含有量を決定した。試料に関するGC密度推定値の分布を、次に参照ゲノム全体にわたる分布と比較して、GCの偏りを決定した(図35)。Aターリッチ領域(GC密度=0)に対してマッピングする読取りおよび参照値を廃棄した。

【0630】

40

試料のGC密度分布と参照の同分布との間の差異を、試料の分布密度で除算した参照の分布密度の対数比に適合した多項式を使用してモデル化した(図36)。重み付けの方式でモデルを適合させ、各重みには、所与のGC密度値に対する試料の分布密度値が採用された。こうすることにより、分布のテールによって、適合が過剰になりすぎないことを保証した。他の適合モデル、例えば分位点回帰分析モデルまたはパラメータ付きの確率分布が、偏りの分布に適するのと同様に使用することができる。

【0631】

参照と比較して試料が過剰にまたは過少に表示される場合、これを調整するために、GC適合モデルを使用して、試料に関する配列の読取りの各カウントを重み付けした。これらの重みを読取り密度の推定値に組み入れることにより、ChAIアルゴリズムは、GC

50

の偏りを補正することができた。

多次元偏り補正

【0632】

GCの偏りは、ゲノム内の読取りパターンに影響を及ぼすいくつかの偏りの1つに過ぎなかった。一般化された多変量モデルを使用して読取りの重みを見積もるために、さらなる偏りをモデル化および補正する場合もあった。この補正は以下に行った：

1. ゲノム位置のサブセットのそれぞれについて、試験試料および参照ゲノムについて、N偏り値を見積もった。
2. N次元平滑化カーネルまたは適するパラメトリック機能を使用して、偏り値の密度をモデル化した。
3. 参照および試験密度から得た一連の密度値について、対数比を計算した。
4. 選択したポイントを使用して、多変量モデルにより、密度の対数比をモデル化した（例えば、各次元について重み付きの3次多項式）。
5. 参照と比較して所与の読取りに関する頻度の比を見積もるのに当該モデルを使用し、そしてしかるべき重みを割り当てた。

部分のフィルタリング

【0633】

ゲノム上の配列の読取り（例えば、カウント）の表示に基づき、試料を染色体異常についてスコア化した。この表示は、ローカルなGCの推定で使用したものと類似した密度関数を使用して決定した。読取り密度カーネルは、一般的にはるかに大きなバンド幅を有し、デフォルトは50,000bpである。読取りの各カウントは、GCの偏りモデルに由来するその重みに等しい値である密度に寄与する。読取り密度は、任意または全部の塩基対において評価され得るが、演算性能上の理由から、ある特定の場所のみを使用した。この位置を「部分」と呼んだ。部分は、読取り密度を見積もるのに最も重要であればどこにも位置し得る。染色体異数性を分類する場合、部分には、最初に（例えば、フィルタリング前に）、ゲノム全体にわたり均等な間隔が設けられる。各部分は、50,000bpウィンドウから構成され、これを、フィルタリング前に、隣接する次の部分と25,000bp重複させた。

【0634】

一部の部分は、十分にマッピングされないゲノム領域を含み、試料から試料へと読取り密度に極度の乱れを引き起こした。ChAIは、トレーニングセットを使用したフィルタリング処理により、この部分を同定および除去した。中央値（例えば、図37A）および/またはMAD値（例えば、図37B）において大きな偏差を示した部分は、検討事項から除去した。この偏差の閾は、トレーニング母集団の四分位値の外側に、四分位数範囲の4倍を超えて存在するあらゆる値として設定した（図37A、37B）。この閾は、ChAIパラメータの特定のセットについて試験成績が最大化するように微調整可能である。トレーニングおよびスコアリング

【0635】

フィルタリングされた部分に対してマッピングする読取りのみを使用して、各試料のゲノム読取り密度プロファイルを計算した。トレーニングセットの一部であった試料を、次にトレーニング統計量を見積もるのに使用したが、同統計量は試験セットをスコア化するのに使用された。この統計量は、部分中央値、主成分、およびスコアリング検定統計量に関するヌル分布からなった。部分中央値および主成分を、任意の数の生物学的および技術的アーチファクトに由来して存在し得る、ゲノム全体に及ぶ読取りの偏りをモデリングするのに使用した（図38A～C）。極端な部分値が他の試料に与える影響を最低限に抑えるために、試料中の他の部分にまたがり、4×IQRの外部にある各値は、4×IQRまで切り捨てた。

【0636】

最初にトレーニングされた中央値を試験部分の値から減じることにより、試験試料を隠れた偏りについて補正した。トップトレーニングされた主成分と相関性を有する試料値の

10

20

30

40

50

コンポーネントも除去した。これは、主成分の項に基づき、多変量直線回帰を使用して、部分の値をモデリングすることにより実施した（図39A～C）。モデルにより予測された値を試料の値から差し引き、偏りのない残差のみを残した。使用される主成分の数は任意選択であり、デフォルトは8である。

【0637】

補正後、試料を、フィッシャー正確検定法を使用してスコア化した。この検定では、目的の染色体領域内のトレーニングされた中央値よりも大きいまたは小さい値を有する部分の数を比較した。このカウントを、ゲノム内の残りの部分に対して評価した。スコアリング統計量を、 $-\log_{10}$ (p 値) として設定した。他のスコアリング統計量、例えばウィルコクソン符号順位検定またはF検定も、このステップで使用することができる。

10

【0638】

部分間の残差相関に起因して、検定統計量が、トレーニングおよび試験試料の両方において増加した。この増加を、トレーニングセットのブートストラップから見積もった（図40）。

【0639】

試験試料に関するスコアを、実験的バックグラウンドとしてこのヌル分布を使用して補正した。実験的分布内のスコアよりはるかに大きなスコアを、ヌル分布のテール部についてパレート外挿を使用して補正した。

性別の判定

【0640】

20

性別を、試料の主成分プロファイルから決定した。トレーニングデータセットでは、第2の主成分（例えば、PC2）が性別と高度に相関した。このコンポーネントの回帰係数を検定統計量として使用すると、それは非常に正確な性別検定となった（図41A、41B）。

部分依存性の除去

【0641】

本アプローチの予知力を向上させるために、ChAI作動期間中にさらなるステップを実施した。これは、部分一試料マトリックス内の相関構造の量を低減することを含み、可変独立性の検定仮説をより適切に裏付け、ヌル順列内の有意スコア頻度を抑制した。本アプローチは、部分を、ほぼ全ての同一情報を含むが相関構造を有さない直交した固有部分と置き換えることを含んだ。

30

【0642】

第1のステップは、一連のトレーニング部分Mについて、変換マトリックスMeigを習得することであった：

1. SVD分解： $M = U * D * V^T$
2. 独立した固有部分Nの数を選択する：（例えば、DのN対角エレメントの積算分画が95%を上回るように）
3. 一般逆行列を計算する： $Meig = pinv(U[\dots, 1:N] * D[1:N, 1:N])$

【0643】

40

部分マトリックスMの任意のサブセットについて、その対応するMeigにより左乗法を行うと、その結果、当該サブセットにつき、次元が低下した相関を有さない表示が得られた。このように、Meigは、トレーニングデータセットに基づき導出され、さらに修正を加えずに試験試料に適用した。

【0644】

Meigは、試験変数を変換する際にも使用した。試験変数を全てのゼロからなるベクトルとして表し、ゼロを予想される偏差の場所に配置した（例えば、Chr21部分）。変換された部分データが適正に一致するように、このベクトルを、左乗法によるMeigを用いて変換した。

【0645】

50

このアプローチが構築し得る独立した固有部分の多さは、せいぜいトレーニングセット内に存在する試料と同じほどに過ぎない。例えば、50,000個の部分および1,000個の試料からなるトレーニングセットでは、変換されたデータは、最大でも1,000個の部分しか含有しない。これは過剰補正の可能性があり、部分の数が大幅に低下する。本アプローチは、部分データのより小さなサブセットについて、個別のMeig変換を計算し、これを個別に適用することにより、より緩やかに行うことができる。これは、隣接した部分からローカルな相関構造を除去するのに特に有用であった。

【0646】

他のアプローチも、部分の相関構造を低減するのに使用することができる。例えば、多くのクラスタリング法が、部分をグループ化し、そしてこれを集合した部分のより小さなセットに置き換えるのに使用することができる（例えば、群の平均またはセントロイドに基づき）。

10

分布／プロファイル生成モジュール

【0647】

配列の読取りデータ（例えば、BRead）から読取り密度プロファイルを生成するために、スクリプトをjava（登録商標）形式で書いた。下記のコードは、各配列の読取りについて読取りデータを収集し、また適する読取り密度ウィンドウ（例えば、部分に関する個々の読取り密度）で密度プロファイルを更新するように設計されており、部分中央値または中間点からの読取りの距離によって重み付けがなされ、試料のGCの偏り補正に基づいた。下記のスクリプトは、関係モジュールまたは偏り補正モジュールから生成された重み付きのカウントおよび／または正規化されたカウントを判定または利用することができる。一部の実施形態では、分布モジュールは、以下に示すjava（登録商標）スクリプトの一部もしくは全部、またその変形形態を含み得る。一部の実施形態では、プロファイル生成モジュールは、以下に示すjava（登録商標）スクリプトの一部もしくは全部、またその変形形態を含み得る：

20

【化 2 7】

```
package utilities.genome;
import java.util.Iterator;

import utilities.data.VectorUtil;
import utilities.text.DataFormatter;

public class ChromDensScaleRunnable implements Runnable{

    private GenomeScaleBoolean mask;
    private GenomeScaleFloat density;
    private final String modelPath;
    private final String brPath;
    private final int bandwidth;
    private final GenomeFloat gcdens;
    private final int stepSize;
    private final int sampleSize;
    private final int shift;
    private final String report;

    public ChromDensScaleRunnable(String modelPath, String brPath, int bandwidth,
    GenomeFloat gcdens, int stepSize, GenomeScaleBoolean mask, String report, int sampleSize, int
    shift)
    {
        this.modelPath = modelPath;
        this.brPath = brPath;
        this.bandwidth = bandwidth;
        this.gcdens = gcdens;
        this.stepSize = stepSize;
        this.mask = mask;
        this.report = report;
        this.sampleSize = sampleSize;
        this.shift = shift;
    }
}
```

【化 2 8】

```

    }

    public void run()
    {
        double[] mdat = (gcdens==null) ? null : VectorUtil.loadDoubleFromFile(modelPath,
6);

        //Build density
        density = new GenomeScaleFloat(stepSize);
10

        double correction = 0;

        try
        {
            Iterator<GenomicPosition> readIterator = (sampleSize==1) ?
GenomeIO.scanBReads(brPath) : new BReadsSampler(brPath, sampleSize,true);

            while (readIterator.hasNext())
            {
                GenomicPosition gp = readIterator.next().shift(shift);
20
                int pos = gp.pos;
                int start = Math.max(0, pos-bandwidth);
                int end = Math.min(pos+bandwidth, GenomeUtil.chromosomeSize(gp.chr)-
1);

                int cindex = gp.chr.ordinal();

                double weight;
                if (gcdens!=null)
                {
                    float gc = gcdens.values[cindex][pos-1];
                    if (gc==0) continue;
30

                    weight = modelWeight(mdat, gc);
                }else weight = 1;

                int[][] gpoints = density.getScalePoints(cindex, start, end, mask);

                if (gpoints[0].length==0 || Double.isNaN(weight)) continue;
                if (weight>2) weight = 2;
                if (weight<.5) weight = .5;
40

                correction += weight;

                for (int i=0;i<gpoints[0].length;i++)
                    density.values[cindex][gpoints[0][i]] += kernel((gpoints[1][i]-
pos)/(double)bandwidth) * weight;
            }

```

【化 2 9】

```

    }catch (Exception e)
    {
        System.out.println("THROW!");
        e.printStackTrace();
        System.exit(0);
    }

    //System.out.println(GenomeIO.countReadsFromBReads(brPath));
    //System.out.println(correction);

    //Normalize intensity
    for (int i=0;i<density.values.length;i++)
        for (int j=0;j<density.values[i].length;j++)
        {
            float blah = density.values[i][j];

            density.values[i][j] /= correction;

            if (Double.isNaN(density.values[i][j]) ||
Double.isInfinite(density.values[i][j]))
            {
                System.out.println("NA val2: "+modelPath+",
"+density.values[i][j]+", "+blah+", "+correction);
                System.exit(0);
            }
        }

        if (report!=null) System.out.println(report);

    }

    public GenomeScaleFloat density()
    {
        return density;
    }

    private static double kernel(double x)
    {
        return .75 * (1.0 - x*x);
    }

    public static double modelWeight(double[] mdat, double gcdens)
    {
        if (mdat[5]==1) gcdens = Math.log(gcdens);

        double x2 = gcdens * gcdens;

```

【化 3 0】

```

        return Math.pow(2, mdat[0] + mdat[1] * gcdens + mdat[2] * x2 + mdat[3] * x2 *
gcdens);
    }

}

```

フィルタリングモジュール

【0 6 4 8】

読取り密度プロファイルの部分をフィルタリングするために、スクリプトをR形式で書いた。このコードは、試料全体にわたり読取り密度プロファイルを検査し、そして保持される部分および／または廃棄される（例えば、分析から除去される）部分を四分位数間範囲に基づき同定する。一部の実施形態では、フィルタリングモジュールは、以下に示すRスクリプトの一部もしくは全部、またその変形形態を含む：

【化 3 1】

```

rcodepath <- "I:/ghannum/Projects/Binless/RCode"
mdistpath <-
"I:/ghannum/Projects/Binless/Reference/MarkerDistribution_LDTv2_200_50000_50000.txt"
outpath <-
"I:/ghannum/Projects/Binless/Reference/LDTv2_200_50000_50000_MarkerMask.txt"

args <- commandArgs(trailingOnly = TRUE)
rcodepath <- args[1]
mdistpath <- args[2]
outpath <- args[3]

source(paste(rcodepath,"/src/utilities/scanmatrix.R",sep=""))

dat <- scanMatrix(mdistpath,rownames=FALSE,colnames=TRUE)

m <- apply(dat,1,median)
v <- apply(dat,1,mad)

qm <- quantile(m,c(.25,.75))
qv <- quantile(v,c(.25,.75))

scalem <- qm[2]-qm[1]
scalev <- qv[2]-qv[1]

ok <- m > qm[1]-4*scalem & m < qm[2]+4*scalem & v > qv[1]-4*scalev & v < qv[2]+4*scalev

write.table(matrix(as.integer(ok),1),row.names=F,col.names=F,quote=F,file=outpath,sep="")

```

偏り密度モジュール、関係モジュール、偏り補正モジュールおよびプロットングモジュール

【0 6 4 9】

偏り密度を生成し、関係を生成および比較し、配列の読取り内の偏りを補正するために、スクリプトをR形式で書いた。このコードは、各試料および参照について、ローカルなゲノムの偏りの推定値（例えば、GC密度）に基づき、1つまたは複数の試料を分析するように、ならびに偏りモデル（例えば、関係および／または関係の比較）を構築するように、一般的にマイクロプロセッサに指示する。下記のスクリプトは、全てではないが、下記事項を目的として、1つまたは複数のプロセッサに指示する：試験試料の配列の読取り

に関して、(i) グアニンおよびシトシン (G C) 密度と (i i) G C 密度頻度との間の関係を生成し、それにより、試料の G C 密度関係を生成する、(b) 試料の G C 密度関係と参照の G C 密度関係を比較し、それにより、比較を生成するが、この場合、参照の G C 密度関係は、参照に関する (i) G C 密度と (i i) G C 密度頻度との間の関係であり、スクリプトのしかるべき修正を含む、(c) (b) で決定した比較により、試料に関する配列の読取りのカウンタを正規化するが、この場合、試料に関する配列の読取りの偏りは低下している。一部の実施形態では、偏り密度モジュール、関係モジュール、偏り補正モジュール、および／またはプロットモジュールは、以下に示す一部または全部のスクリプトについて、その一部もしくは全部、またその変形形態を含む：

【化 3 2】

10

```
gcpath <- "I:/ghannum/Projects/Binless/Reference/BiasMaps/DnaseDensity_200_dist.txt"
inpath <- "I:/ghannum/Projects/Binless/Models/LDTv2_DNase-200"
outpath <- "I:/ghannum/Projects/Binless/Models/LDTv2_DNase-200"
makePlots <- TRUE
logTransform <- TRUE
```

```
args <- commandArgs(trailingOnly = TRUE)
gcpath <- args[1]
inpath <- args[2]
outpath <- args[3]
makePlots <- args[4]
logTransform <- as.logical(args[5])
paths <- dir(inpath)
paths <- paths[grep("_BiasDistr.txt$",paths)]
```

20

```
gcref <- scan(gcpath,0)
gcref <- gcref[gcref!=0]
```

30

```
if (logTransform) gcref <- log(gcref)
```

【化 3 3】

```

from <- quantile(gcref,.005)
to <- quantile(gcref,.995)

x <- seq(from,to,length.out=100);

d1y <- predict(smooth.spline(density(gcref,from=from,to=to)),x)$y
if (!logTransform) d1y <- sapply(d1y,function(x){max(x,0)})
10

print(paste("Processing",length(paths),"models."))

for (f in paths)
{
  distr <- scan(paste(inpath,"/",f,sep=""),0)
  distr <- distr[distr!=0]
  if (logTransform) distr <- log(distr)
  d2y <- predict(smooth.spline(density(distr,from=from,to=to)),x)$y
  if (!logTransform) d2y <- sapply(d2y,function(x){max(x,0)})
  pp <- log2(d1y / d2y)
  pp[pp > 2] <- 2; pp[pp < -2] <- -2
  ok <- !is.na(pp)
  mod <- lm(pp[ok]~x+l(x^2)+l(x^3), data=list(x=x[ok]), w=d2y[ok])
  w <- 2^predict(mod,list(x=distr))
  fname <- substr(f,1,nchar(f)-14)
  out <- c(mod$coefficients,mean(w))
  out[out==Inf] <- "Infinity"
  out[out==-Inf] <- "-Infinity"
  write.table(matrix(c(out,as.integer(logTransform)),ncol=1),file=paste(outpath,"/",fname,
e,"_BiasMod.txt",sep=""),row.names=F,col.names=F,quote=F)

  if (makePlots)
  {
    png(units="in",height=4,width=4,res=300,file=paste(outpath,"/",fname,"_BiasMod.png",sep=""))
    if (logTransform)
    {

```

10

20

30

40

【化 3 4】

```

        plot(x[ok],pp[ok],ylim=c(-4,4),xlab="Bias Density",ylab="Log2 Ratio
(Reference / Sample)")
    }else plot(x[ok],pp[ok],ylim=c(-4,4),xlab="Log-Bias Density",ylab="Log2
Ratio (Reference / Sample)")
        abline(h=0,lty=2)
        lines(x[ok],predict(mod),col=3)
        dev.off()
    }
}

```

10

```

##Demo transformation
#load("I:/ghannum/Projects/Binless/2012_11_13_cewi_PERUN_19FCs_AltGCbias_chrFracti
ons.RData")
#
#d <- dir("I:/ghannum/Projects/Binless/GCDistribution/LDTv2/")
#d <- d[grep("_GCDistr.txt",d)]
#v <- as.numeric(df.cewi.GCbiasTable[, "gcBiasRobust"])[1:length(d)]
#
#a <-
scan(paste("I:/ghannum/Projects/Binless/GCDistribution/LDTv2/",d[which.min(v)],sep=""),0);
a <- sort(a)
#b <-
scan(paste("I:/ghannum/Projects/Binless/GCDistribution/LDTv2/",d[which.max(v)],sep=""),0);
d <- sort(d)
#
#r <- scan("I:/ghannum/Projects/Binless/Reference/GCDensity_200_density.txt",0)
#
#plot(density(r),ylim=c(0,1e10),xlab="GC Density"); lines(density(a),col=3);
lines(density(b),col=2)
#
#a <- a[a!=0]
#b <- b[b!=0]
#r <- r[r!=0]
#
#plot(density(r),ylim=c(0,1e10),xlab="GC Density"); lines(density(a),col=3);
lines(density(b),col=2)
#

```

20

30

40

【化 3 5】

```

#modA <-
as.numeric(scan(paste("I:/ghannum/Projects/Binless/GCDistribution/LDTv2/",substr(d[which.
min(v)],1,nchar(d[which.min(v)])-12],"_GCMod.txt",sep=""),""))
#modB <-
as.numeric(scan(paste("I:/ghannum/Projects/Binless/GCDistribution/LDTv2/",substr(d[which.
max(v)],1,nchar(d[which.max(v)])-12],"_GCMod.txt",sep=""),""))
#
#wa <- sapply(a,function(x){2^sum(c(1,x,x^2,x^3)*modA[1:4])})
#wb <- sapply(b,function(x){2^sum(c(1,x,x^2,x^3)*modB[1:4])})
#
#wa <- wa/(length(wa)*modA[5])
#wb <- wb/(length(wb)*modB[5])
#
#plot(density(r),ylim=c(0,1e10),xlab="GC Density"); lines(density(a,weights=wa),col=3);
lines(density(b,weights=wb),col=2)

```

10

20

(実施例 8) : 実施形態の例

【0 6 5 0】

下記の例は、ある特定の実施形態について説明し、本技術に制限を設けるものではない。

A 1. 偽陰性および偽陽性の決定が低減した、胎児中の染色体異数性、微小重複または微小欠失の存在または非存在を決定する方法であって、

(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得るステップであり、配列の読取りが、胎児を出産する妊娠中の雌に由来する循環無細胞核酸の読取りであるステップと、

(b) 各部分に対してマッピングしたカウントを正規化し、それにより、計算されたゲノム区分のレベルを得るステップと、

(c) 計算されたゲノム区分のレベルにより、ゲノムのセグメントについてプロファイルを生成するステップと、

(d) プロファイルをセグメント化し、それにより、2つまたはそれ超の分解レンダリングを提供するステップと

(e) 2つまたはそれ超の分解レンダリングにより、偽陰性および偽陽性の決定が低下した、染色体異数性、微小重複または微小欠失の存在または非存在を決定するステップを含む方法。

30

【0 6 5 1】

A 2. セグメント化が、閾化を含む、実施形態 A 1 に記載の方法。

40

【0 6 5 2】

A 3. 閾化が、ソフトな閾化を含む、実施形態 A 2 に記載の方法。

【0 6 5 3】

A 4. 閾化が、ポリシーを含む、実施形態 A 2 または A 3 に記載の方法。

【0 6 5 4】

A 5. ポリシーが、一般的である、実施形態 A 4 に記載の方法。

【0 6 5 5】

A 6. ポリシーが、明確である、実施形態 A 4 に記載の方法。

【0 6 5 6】

A 7. 閾化が、Wave Thresholdにより行われる、実施形態 A 2 から A 6 のいずれ

50

か1つに記載の方法。

【0657】

A8．セグメント化が、レベル化を含む、実施形態A1からA7のいずれか1つに記載の方法。

【0658】

A9．レベル化が、胎児フラクションにより行われる、実施形態A8に記載の方法。

【0659】

A10．レベル化が、カバレッジにより行われる、実施形態A8またはA9に記載の方法。

【0660】

A11．レベル化が、最小セグメント長さの検出により行われる、実施形態A8からA10のいずれか1つに記載の方法。

【0661】

A12．閾化およびレベル化が行われ、閾化が、レベル化の前に行われる、実施形態A8からA11のいずれか1つに記載の方法。

【0662】

A13．(d)におけるセグメント化が、2つまたはそれ超の異なる分解生成処理により行われる、実施形態A1からA12のいずれか1つに記載の方法。

【0663】

A14．2つまたはそれ超の異なる分解生成処理のそれぞれが、ハールウェーブレットセグメンテーション、サーキュラーバイナリセグメンテーション、最大エントロピーセグメンテーション、エッジ検出カーネルによるコンボリューション、ジェンセンシャノンダイバージェンス、バイナリ再帰的セグメンテーション、およびフーリエ変換から独立に選択される、実施形態A13に記載の方法。

【0664】

A15．2つまたはそれ超の異なる分解生成処理のうちの1つが、サーキュラーバイナリセグメンテーションである、実施形態A13またはA14に記載の方法。

【0665】

A16．2つまたはそれ超の異なる分解生成処理のうちの1つが、ハールウェーブレットである、実施形態A13からA15のいずれか1つに記載の方法。

【0666】

A17．(d)におけるセグメント化が、ハールウェーブレットおよびサーキュラーバイナリセグメンテーションを含む、実施形態A13からA16のいずれか1つに記載の方法。

【0667】

A18．2つまたはそれ超の分解生成処理が、並行して適用される、実施形態A13からA17のいずれか1つに記載の方法。

【0668】

A19．2つまたはそれ超の分解生成処理が、連続して適用される、実施形態A13からA17のいずれか1つに記載の方法。

【0669】

A20．2つまたはそれ超の分解レンダリングのうちの1つまたは複数を仕上げ処理し、それにより、1つまたは複数の仕上げ処理された分解レンダリングを得ることを含む、実施形態A1からA19のいずれか1つに記載の方法。

【0670】

A21．仕上げ処理することが、分解レンダリング中の隣接する断片化されたレベルを統合することを含む、実施形態A20に記載の方法。

【0671】

A22．隣接する断片化されたレベルが、そのゲノム区分のレベルにより統合される、実施形態A20またはA21に記載の方法。

10

20

30

40

50

【0672】

A23. 2つまたはそれ超の分解レンダリングのうちの1つまたは複数の中で候補セグメントを同定することを含む、実施形態A1からA22のいずれか1つに記載の方法。

【0673】

A23. 1. 前記候補セグメントが、前記1つまたは複数の仕上げ処理された分解レンダリング中で同定される、実施形態A23に記載の方法。

【0674】

A24. 候補セグメントのエッジを決定することを含む、実施形態A23またはA23.1に記載の方法。

【0675】

A25. 候補セグメントのレベルを決定することを含む、実施形態A23からA24のいずれか1つに記載の方法。

【0676】

A26. 前記候補セグメントが、マルチプロファイルにより同定される、実施形態A23またはA25のいずれか1つに記載の方法。

【0677】

A27. (a)におけるカウントが、妊娠中の雌から得られた試料から得られる、実施形態A1からA26のいずれか1つに記載の方法。

【0678】

A28. マルチプロファイルが、試料から生成される、実施形態A27に記載の方法。

【0679】

A29. マルチプロファイルが、参照試料から生成される、実施形態A26またはA27に記載の方法。

【0680】

A30. 候補セグメントが、曲線下面積(AUC)の分析により同定される、実施形態A23からA29のいずれか1つに記載の方法。

【0681】

A31. 少なくとも2つの候補セグメントを比較し、それにより、比較を得ることを含む、実施形態A23からA30のいずれか1つに記載の方法。

【0682】

A32. 第1の候補セグメントが、第1の分解レンダリングに由来し、第2の候補セグメントが、第2の分解レンダリングに由来する、実施形態A31に記載の方法。

【0683】

A33. 少なくとも2つの候補セグメントが、比較により実質的に同じであると決定される、実施形態A31またはA32に記載の方法。

【0684】

A33. 1. 少なくとも2つの候補セグメントが、比較により異なると決定される、実施形態A31またはA32に記載の方法。

【0685】

A33. 2. 染色体異数性の存在または非存在が、比較により決定される、実施形態A31からA33.1のいずれか1つに記載の方法。

【0686】

A34. 比較が、少なくとも2つの候補セグメントを重ね合わせることを含む、実施形態A31からA33.2のいずれか1つに記載の方法。

【0687】

A34. 1. 比較により、複合候補セグメントの存在または非存在を決定することを含む、実施形態A31またはA34のいずれか1つに記載の方法。

【0688】

A35. 第1の候補セグメントが、第2の候補セグメントと実質的に重複し、複合候補セグメントが存在すると決定される、実施形態A34またはA34.1に記載の方法。

10

20

30

40

50

【0689】

A35. 1. 第1の候補セグメントが、第2の候補セグメントとは実質的に重複せず、複合候補セグメントが存在しない決定される、実施形態34または34. 1に記載の方法。

【0690】

A36. 染色体異数性の存在または非存在が、複合候補セグメントの存在または非存在により(e)で決定される、実施形態A34. 1からA35. 1のいずれか1つに記載の方法。

【0691】

A37. 分解レンダリングにおいて同定された候補セグメントを検証し、それにより、検証された候補セグメントを得ることを含む、実施形態A23からA36のいずれか1つに記載の方法。

10

【0692】

A38. 検証することが、スライディングエッジ処理を行うことを含む、実施形態A37に記載の方法。

【0693】

A39. 検証することが、リーブワンアウト処理(leave one out process)を行うことを含む、実施形態A37またはA38に記載の方法。

【0694】

A40. 検証することが、スライディングエッジ処理およびリーブワンアウト処理を行うことを含む、実施形態A39に記載の方法。

20

【0695】

A41. 検証することが、候補セグメントについて有意性のレベルを生成することを含む、実施形態A37からA40のいずれか1つに記載の方法。

【0696】

A42. 検証することが、複合候補セグメントについて有意性のレベルを生成することを含む、実施形態A37からA41のいずれか1つに記載の方法。

【0697】

A43. 有意性のレベルが、Zスコアである、実施形態A41またはA42に記載の方法。

30

【0698】

A44. 不確実性のレベルが、有意性のレベルと関連する、実施形態A41からA43のいずれか1つに記載の方法。

【0699】

A45. 検証された候補セグメントの存在または非存在が、候補セグメントについての有意性のレベルおよび不確実性のレベルにより決定される、実施形態A44に記載の方法。

【0700】

A46. 染色体異数性、微小重複または微小欠失の存在が、有意性のレベルおよび不確実性のレベルにより決定され、有意性のレベルおよび不確実性のレベルの両方が、複合候補セグメントについて生成される、実施形態A44またはA45に記載の方法。

40

【0701】

A47. 染色体異数性、微小重複または微小欠失の存在が、ZスコアおよびZスコアと関連した不確実性のレベルにより決定され、Zスコアおよび不確実性のレベルの両方が、複合候補セグメントについて生成される、実施形態A46に記載の方法。

【0702】

A47. 1. Zスコアが、約3.95より大きいまたはこれに等しい絶対値を有する、実施形態A47に記載の方法。

【0703】

A48. 染色体異数性の存在または非存在が決定される、実施形態A1からA47. 1

50

のいずれか 1 つに記載の方法。

【0704】

A 48. 1. 染色体異数性が、トリソミーである、実施形態 A 1 から A 48 のいずれか 1 つに記載の方法。

【0705】

A 48. 2. 染色体異数性が、モノソミーである、実施形態 A 1 から A 48. 1 のいずれか 1 つに記載の方法。

【0706】

A 49. 微小重複の存在または非存在が決定される、実施形態 A 1 から A 48. 3 のいずれか 1 つに記載の方法。

【0707】

A 50. 微小欠失の存在または非存在が決定される、実施形態 A 1 から A 48. 3 のいずれか 1 つに記載の方法。

【0708】

A 51. ディジョージ症候群を示唆する微小欠失の存在または非存在が決定される、実施形態 A 1 から A 50 のいずれか 1 つに記載の方法。

【0709】

A 52. (a)、(b)、(c)、(d)、および (e) のうちの 1 つまたは複数または全てが、プロセッサにより行われる、実施形態 A 1 から A 51 のいずれか 1 つに記載の方法。

【0710】

A 53. プロセッサが、マイクロプロセッサである、実施形態 A 52 に記載の方法。

【0711】

A 54. (a)、(b)、(c)、(d)、および (e) のうちの 1 つまたは複数または全てが、コンピュータにより行われる、実施形態 A 1 から A 53 のいずれか 1 つに記載の方法。

【0712】

A 55. (a)、(b)、(c)、(d)、および (e) のうちの 1 つまたは複数または全てが、メモリと併せて行われる、実施形態 A 1 から A 54 のいずれか 1 つに記載の方法。

【0713】

A 56. (a)、(b)、(c)、(d)、および (e) のうちの 1 つまたは複数または全てが、マイクロプロセッサ制御式装置により行われる、実施形態 A 1 から A 55 のいずれか 1 つに記載の方法。

【0714】

A 57. (a) に先立って、妊娠中の雌から得られた試料中の核酸の配列決定を行い、それにより、核酸配列決定の読取りを得ることを含む、実施形態 A 1 から A 56 のいずれか 1 つに記載の方法。

【0715】

A 58. (a) に先立って、参照ゲノムの部分または参照ゲノムの全部に対して、核酸配列の読取りをマッピングすることを含む、実施形態 A 1 から A 57 のいずれか 1 つに記載の方法。

【0716】

B 1. 偽陰性および偽陽性の決定が低減した候補セグメントの存在または非存在を決定する方法であって、

(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得るステップであり、配列の読取りが、妊娠中の雌に由来する循環無細胞核酸の読取りであるステップと、

(b) 部分のそれぞれに対してマッピングしたカウントを正規化し、それにより、計算されたゲノム区分のレベルを得るステップと、

10

20

30

40

50

(c) 一連の部分を部分の複数のサブセットにセグメント化するステップと、
(d) 計算されたゲノム区分のレベルにより、サブセットのそれぞれにレベルを決定するステップと、
(e) レベルのそれぞれに有意性のレベルを決定するステップと、
(f) 各レベルに決定された有意性のレベルにより、偽陰性および偽陽性の決定が低減した候補セグメントの存在または非存在を決定するステップとを含む、方法。

【0717】

B1. 一連の部分が、候補セグメントを含むと疑われる、実施形態 B1 に記載の方法。

10

【0718】

B2. 部分のサブセット全てについて、レベルのそれぞれに決定された有意性のレベルにより、有意性のレベル中央値を生成することを含む、実施形態 B1 または B1. 1 に記載の方法。

【0719】

B3. 部分のサブセット全てについて、レベルのそれぞれに決定された有意性のレベルの分布を生成することを含む、実施形態 B1 から B2 のいずれか 1 つに記載の方法。

【0720】

B4. 部分のサブセット全てに関するレベル全てについて決定された有意性のレベルにより、不確実性の値を生成することを含む、実施形態 B1 から B3 のいずれか 1 つに記載の方法。

20

【0721】

B5. (f) における決定することが、有意性のレベル中央値および不確実性の値による、実施形態 B4 に記載の方法。

【0722】

B6. (f) における決定することが、有意性のレベル中央値および有意性のレベルの分布による、実施形態 B3 または B4 に記載の方法。

【0723】

B6. 1. (f) における決定することが、有意性のレベルについて決定された所定の範囲による、実施形態 B1 から B6 のいずれか 1 つに記載の方法。

30

【0724】

B6. 2. B3 で生成された有意性のレベルの分布の 75% またはそれ超が、有意性のレベルについて所定の範囲の外側にあるとき、候補セグメントの存在が決定される、実施形態 B6. 1 に記載の方法。

【0725】

B6. 3. 実施形態 B4 で生成された不確実性の値の 75% またはそれ超が、有意性のレベルについて所定の範囲の外側にあるとき、候補セグメントの存在が決定される、実施形態 B6. 1 または B6. 2 に記載の方法。

【0726】

B7. 有意性のレベルが、Z スコアである、実施形態 B1 から B6. 3 のいずれか 1 つに記載の方法。

40

【0727】

B7. 1. 所定の範囲が、約 3 ～ 約 -3 の間の Z スコアである、実施形態 B7 に記載の方法。

【0728】

B8. 不確実性の値が、絶対偏差中央値である、実施形態 B4 から B7. 1 のいずれか 1 つに記載の方法。

【0729】

B9. 一連の部分が、第 1 の末端および第 2 の末端を含み、(c) のセグメントが (i) 1 つまたは複数の部分を、再帰的除去により、一連の部分の第 1 の末端から除去し

50

、それにより、各再帰的除去によって得られた部分のサブセットを得るステップと、
 (i i) n 回繰り返した後に (i) の再帰的除去を終了し、それにより、部分のサブセットを $n+1$ 個得るステップであって、一連の部分がサブセットであり、各サブセットが、異なる数の部分、第 1 のサブセット末端、および第 2 のサブセット末端を含むステップと、

(i i i) (i i) で得た $n+1$ 個の部分のサブセットそれぞれの第 2 のサブセット末端から、再帰的除去により 1 つまたは複数の部分を除去するステップと、

(i v) n 回繰り返した後に (i i i) の再帰的除去を終了し、それにより、部分のサブセットを複数得るステップと

を含む、実施形態 B 1 から B 8 のいずれか 1 つに記載の方法。

10

【0730】

B 10. 複数のサブセットが、 $(n+1)$ 2 個のサブセットに等しい、実施形態 B 9 に記載の方法。

【0731】

B 11. n が、5 ~ 30 の間の整数に等しい、実施形態 B 9 または B 10 に記載の方法。

【0732】

B 12. n が、15 に等しい、実施形態 B 9 から B 11 のいずれか 1 つに記載の方法。

【0733】

B 13. 一連の部分が、染色体内にある、実施形態 B 1 から B 12 のいずれか 1 つに記載の方法。

20

【0734】

B 14. 一連の部分が、公知の遺伝子の変動または公知の遺伝子の障害と関連した領域を含む、実施形態 B 13 に記載の方法。

【0735】

B 14. 一連の部分が、ディジョージ領域を含む、実施形態 B 13 または B 14 に記載の方法。

【0736】

B 15. (a) ~ (e) が、試験試料および 2 つまたはそれ超の参照試料について行われる、実施形態 B 1 から B 14 のいずれか 1 つに記載の方法。

30

B 16. (i) (a) に先立って、2 つまたはそれ超の参照試料のうちの 1 つを除去し、それにより、参照試料のサブセットを得るステップと、

(i i) 参照試料のサブセットのそれぞれについて (a) ~ (e) を行うステップと、

(i i i) 参照試料のサブセットのそれぞれについて、実施形態 B 2 により、有意性のレベル中央値を生成するステップと、

(i v) (i i i) で生成された中央値により、有意性の複合レベル中央値を生成するステップと、

(v) (i v) の有意性の複合レベル中央値について、不確実性の複合レベルを生成するステップであって、(f) における決定するステップが、有意性の複合レベル中央値および不確実性の複合レベルによるステップと

40

を含む、実施形態 B 15 に記載の方法。

【0737】

B 17. 参照試料のサブセットのそれぞれが、異なる一連の参照試料を含む、実施形態 B 16 に記載の方法。

【0738】

B 18. 除去される 2 つまたはそれ超の参照試料の 1 つが、それぞれサブセットの 1 つのみから除去される、実施形態 B 16 または B 17 に記載の方法。

【0739】

B 19. (a)、(b)、(c)、(d)、(e)、および (f) のうちの 1 つまたは複数または全てが、プロセッサにより行われる、実施形態 B 1 から B 18 のいずれか 1 つ

50

に記載の方法。

【0740】

B20. プロセッサが、マイクロプロセッサである、実施形態B19に記載の方法。

【0741】

B21. (a)、(b)、(c)、(d)、(e)、および(f)のうちの1つまたは複数または全てが、コンピュータにより行われる、実施形態B1からB20のいずれか1つに記載の方法。

【0742】

B22. (a)、(b)、(c)、(d)、(e)、および(f)のうちの1つまたは複数または全てが、メモリと併せて行われる、実施形態B1からB21のいずれか1つに記載の方法。

10

【0743】

B23. (a)、(b)、(c)、(d)、(e)、および(f)のうちの1つまたは複数または全てが、マイクロプロセッサ制御式装置により行われる、実施形態B1からB22のいずれか1つに記載の方法。

【0744】

B24. (a)に先立って、妊娠中の雌から得られた試料中の核酸の配列決定を行い、それにより、核酸配列決定の読取りを得ることを含む、実施形態B1からB23のいずれか1つに記載の方法。

【0745】

20

B25. (a)に先立って、参照ゲノムの部分または参照ゲノムの全部に対して核酸配列の読取りをマッピングすることを含む、実施形態B1からB24のいずれか1つに記載の方法。

【0746】

C1. 偽陰性および偽陽性の決定が低減した、胎児中の染色体異数性、微小重複または微小欠失の存在または非存在を決定する方法であって、

(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを得るステップであり、配列の読取りが妊娠中の雌に由来する循環無細胞核酸の読取りであるステップと、

(b) 各部分に対してマッピングしたカウントを正規化し、それにより、計算されたゲノム区分のレベルを得るステップと、

30

(c) ゲノムのセグメントを選択し、それにより、一連の部分を得るステップと、

(d) 一連部分を再帰的に区切り、それにより、部分の2つまたはそれ超のサブセットを提供するステップと、

(e) 部分の2つまたはそれ超のサブセットのそれぞれについてレベルを決定するステップと、

(f) (e)で決定されたレベルにより、偽陰性および偽陽性の決定が低減した試料について、胎児中の染色体異数性、微小重、または微小欠失の存在または非存在を決定するステップと

を含む、方法。

40

【0747】

C2. (e)で決定された部分の2つまたはそれ超のサブセットのそれぞれについてレベルが有意に異なるか決定することを含む、実施形態C1に記載の方法。

【0748】

C3. (d)における区切ることが、再帰的に区切ることを含む、実施形態C1またはC2に記載の方法。

【0749】

C4. 再帰的に区切ることが、バイナリ再帰的に区切ることを含む、実施形態C3に記載の方法。

【0750】

50

C 5. 再帰的に区切ることが、最大エントロピーに基づき区切ることを含む、実施形態 C 3 に記載の方法。

【0751】

C 6. 部分の第 1 のサブセットのレベルが部分の第 2 のサブセットのレベルと有意に異なるとき、部分の第 1 および第 2 のサブセットを区切ることを含む、部分の第 1 および第 2 のサブセットが相互に隣接している、実施形態 C 2 から C 5 のいずれか 1 つに記載の方法。

【0752】

C 7. 部分の第 3 のサブセットおよび部分の第 4 のサブセットのレベルが有意に異ならないとき、部分の第 3 のサブセットと部分の第 4 のサブセットとを再結合させ、それにより、再結合した部分のサブセットを得ることを含む、部分の第 3 のサブセットおよび部分の第 4 番のサブセットが相互に隣接し、および再結合した部分が、再び区切られない、実施形態 C 2 から C 6 のいずれか 1 つに記載の方法。

【0753】

C 8. (a)、(b)、(c)、(d)、(e)、および(f)のうちの1つまたは複数または全てが、プロセッサにより行われる、実施形態 C 1 から C 7 のいずれか 1 つに記載の方法。

【0754】

C 9. プロセッサが、マイクロプロセッサである、実施形態 C 8 に記載の方法。

【0755】

C 10. (a)、(b)、(c)、(d)、(e)、および(f)のうちの1つまたは複数または全てが、コンピュータにより行われる、実施形態 C 1 から C 9 のいずれか 1 つに記載の方法。

【0756】

C 11. (a)、(b)、(c)、(d)、(e)、および(f)のうちの1つまたは複数または全てが、メモリと併せて行われる、実施形態 C 1 から C 10 のいずれか 1 つに記載の方法。

【0757】

C 12. (a)、(b)、(c)、(d)、(e)、および(f)のうちの1つまたは複数または全てが、マイクロプロセッサ制御式装置により行われる、実施形態 C 1 から C 11 のいずれか 1 つに記載の方法。

【0758】

C 13. (a) に先立って、妊娠中の雌から得られた試料中の核酸の配列決定を行い、それにより、核酸配列決定の読取りを得ることを含む、実施形態 C 1 から C 12 のいずれか 1 つに記載の方法。

【0759】

C 14. (a) に先立って、核酸配列の読取りを参照ゲノムの部分または参照ゲノムの全部に対してマッピングすることを含む、実施形態 C 1 から C 13 のいずれか 1 つに記載の方法。

【0760】

D 1. 胎児中の染色体異数性、微小重複または微小欠失の存在または非存在を決定するための方法であって、

(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントを正規化し、それにより、正規化されたカウントを得るステップであり、これらの配列の読取りが、胎児を出産する妊娠中の雌に由来する試験試料から得られた循環無細胞核酸の読取りであるステップと、

(b) 部分の正規化されたカウントまたは部分のサブセット中の正規化されたカウントをセグメント化し、それにより、1つまたは複数の個別のセグメントを得るステップと、

(c) 1つまたは複数の個別のセグメントから候補セグメントを同定するステップと、

10

20

30

40

50

(d) 候補セグメントにより、染色体異数性、微小重複または微小欠失の存在または非存在を決定するステップとを含む方法。

【0761】

D2. セグメント化が閾化を含む、実施形態D1に記載の方法。

【0762】

D3. セグメント化がレベル化を含む、実施形態D1またはD2のいずれか1つに記載の方法。

【0763】

D4. レベル化が、胎児フラクション、カバレッジ、最小セグメント長さ、またはそれらの組合せに従って行われる、実施形態D3に記載の方法。 10

【0764】

D5. 閾化およびレベル化が行われ、閾化が、レベル化の前に行われる、実施形態D1からD4のいずれか1つに記載の方法。

【0765】

D5. 1. (b)におけるセグメント化が、部分の正規化されたカウントに対して行われる、実施形態D1からD5のいずれか1つに記載の方法。

【0766】

D5. 2. (b)におけるセグメント化が、部分のサブセット中の正規化されたカウントに対して行われる、実施形態D1からD5のいずれか1つに記載の方法。 20

【0767】

D5. 3. 部分のサブセットが、染色体の全部分または染色体の全部分のサブセットである、実施形態D5. 2に記載の方法。

【0768】

D5. 4. 正規化されたカウントが、レベルを有するプロファイル中にあり、プロファイルが、(b)においてセグメント化される、実施形態D1からD5. 3のいずれか1つに記載の方法。

【0769】

D5. 5. セグメント化が、個別のセグメントを含む分解レンダリングを生成する、実施形態D1からD5. 4のいずれか1つに記載の方法。 30

【0770】

D5. 6. (a)における正規化が、グアニンおよびシトシン(GC)の偏りのLOESS正規化(GC-LOESS正規化)を含む、実施形態D1からD5. 5のいずれか1つに記載の方法。

【0771】

D5. 7. (a)における正規化が、主成分正規化を含む、実施形態D1からD5. 6のいずれか1つに記載の方法。

【0772】

D5. 8. (a)における正規化が、GC-LOESS正規化、およびそれに続く、主成分正規化を含む、実施形態D1からD5. 7のいずれか1つに記載の方法。 40

【0773】

D5. 9. (a)における正規化が、
(1) (i) 部分のそれぞれに対してマッピングした配列の読取りのカウントと、(ii) 部分のそれぞれについてのGC含有量との間で適合させた関係式に基づいて、試験試料についてのグアニンおよびシトシン(GC)の偏り係数を決定するサブステップであって、GCの偏り係数が、線形適合関係式の場合の勾配、または非線形適合関係式の場合の曲率の推定値であるサブステップと、

(2) マイクロプロセッサを使用して、(a)のカウント、(b)のGCの偏り係数、および部分のそれぞれについて、(i) 複数の試料のそれぞれについてのGCの偏り係数と、(ii) 複数の試料についての、部分のそれぞれに対してマッピングした配列の読取り 50

のカウンタとの間で適合させた関係式に基づいて、部分のそれぞれについてのゲノム区分のレベルを計算し、それにより、計算されたゲノム区分のレベルを得るサブステップとを含む、実施形態 D 1 から D 5. 8 のいずれか 1 つに記載の方法。

【0774】

D 6. (b) におけるセグメント化が、2 つまたはそれ超の異なるセグメント化処理の適用を含む、実施形態 D 1 から D 5. 9 のいずれか 1 つに記載の方法。

【0775】

D 7. 2 つまたはそれ超の異なるセグメント化処理のそれぞれが、ハールウェーブレットセグメンテーション、サーキュラーバイナリセグメンテーション、最大エントロピーセグメンテーション、エッジ検出カーネルによるコンボリューション、ジェンセンシャノンダイバージェンス、バイナリ再帰的セグメンテーション、およびフーリエ変換から独立に選択される、実施形態 D 6 に記載の方法。

10

【0776】

D 8. 2 つまたはそれ超の異なるセグメント化処理のうちの 1 つが、サーキュラーバイナリセグメンテーションである、実施形態 D 6 または D 7 に記載の方法。

【0777】

D 9. 2 つまたはそれ超の異なるセグメント化処理のうちの 1 つが、ハールウェーブレットセグメンテーションである、実施形態 D 6 から D 8 のいずれか 1 つに記載の方法。

【0778】

D 10. (b) におけるセグメント化が、ハールウェーブレットセグメント化処理およびサーキュラーバイナリセグメント化処理を含む、実施形態 D 6 から D 9 のいずれか 1 つに記載の方法。

20

【0779】

D 11. 2 つまたはそれ超のセグメント化処理が、平行して行われる、実施形態 D 6 から D 10 のいずれか 1 つに記載の方法。

【0780】

D 12. セグメント化が、分解レンダリング中の隣接する断片化されたレベルを統合することを含む仕上げ処理を含む、実施形態 D 1 から D 11 のいずれか 1 つに記載の方法。

【0781】

D 13. 候補セグメントの 1 つまたは複数のエッジを決定することを含む、実施形態 D 1 から D 12 のいずれか 1 つに記載の方法。

30

【0782】

D 14. 候補セグメントがカバーする部分の数を決定することを含む、実施形態 D 1 から D 13 のいずれか 1 つに記載の方法。

【0783】

D 15. 候補セグメントのレベルを決定することを含む、実施形態 D 1 から D 14 のいずれか 1 つに記載の方法。

【0784】

D 15. 1. 候補セグメントが、曲線下面積 (AUC) の分析により同定される、実施形態 D 1 から D 15 のいずれか 1 つに記載の方法。

40

【0785】

D 16. AUC 分析が、候補セグメントがカバーする部分の数、および／または候補セグメントについてのレベルに関する、実施形態 D 15. 1 に記載の方法。

【0786】

D 16. 1. 候補セグメントを検証し、それにより、検証された候補セグメントを得ることを含む、実施形態 D 1 から D 16 のいずれか 1 つに記載の方法。

【0787】

D 16. 2. 検証することが、スライディングエッジ処理を行うことを含む、実施形態 D 16. 1 に記載の方法。

【0788】

50

D 1 6 . 3 . 検証することが、リーブワンアウト処理を行うことを含む、実施形態 D 1 6 . 1 または D 1 6 . 2 に記載の方法。

【0 7 8 9】

D 1 6 . 4 . 検証することが、スライディングエッジ処理およびリーブワンアウト処理を行うことを含む、実施形態 D 1 6 . 3 に記載の方法。

【0 7 9 0】

D 1 6 . 5 . 検証することが、候補セグメントについての有意性のレベルを生成することを含む、実施形態 D 1 6 . 1 から D 1 6 . 4 のいずれか 1 つに記載の方法。

【0 7 9 1】

D 1 6 . 6 . 検証することが、複合候補セグメントについての有意性のレベルを生成することを含む、実施形態 D 1 6 . 1 から D 1 6 . 5 のいずれか 1 つに記載の方法。

10

【0 7 9 2】

D 1 6 . 7 . 第 1 のセグメンテーションから第 1 の候補セグメントを同定し、第 1 のセグメンテーションとは異なる第 2 のセグメンテーションから第 2 の候補セグメントを同定することを含む、実施形態 D 1 から D 1 6 . 6 のいずれか 1 つに記載の方法。

【0 7 9 3】

D 1 6 . 8 . 第 1 の候補セグメントと第 2 の候補セグメントとが実質的に同じであるまたは実質的に異なるかどうかを決定することを含む、実施形態 D 1 6 . 7 に記載の方法。

【0 7 9 4】

D 1 6 . 9 . 第 1 の候補セグメントと第 2 の候補セグメントとが実質的に異なる場合に、微小欠失または微小重複が存在しないと決定することを含む、実施形態 D 1 6 . 7 または D 1 6 . 8 に記載の方法。

20

【0 7 9 5】

D 1 7 . 候補セグメントまたは検証された候補セグメントの定量を行うことを含む、実施形態 D 1 から D 1 6 . 9 のいずれか 1 つに記載の方法。

【0 7 9 6】

D 1 8 . 定量が、候補セグメントまたは検証された候補セグメントについてのカウント表示である、実施形態 D 1 7 に記載の方法。

【0 7 9 7】

D 1 9 . 定量が、候補セグメントまたは検証された候補セグメントについてのカウント表示の z スコアによる定量である、実施形態 D 1 8 に記載の方法。

30

【0 7 9 8】

D 2 0 . z スコアが、候補セグメントまたは検証された候補セグメントについて、(i) 試験試料カウント表示から、(i i) 正倍数体カウント表示の中央値を減算して得た差を、(i i i) 正倍数体カウント表示の M A D で除算して得た商であり、ここで、(i) 試験試料カウント表示が、試験試料について、全カウントを、全常染色体カウントで除算して得た比であり、(i i) 正倍数体カウント表示の中央値が、正倍数体試料について、全カウントを、全常染色体カウントで除算して得た比の中央値である、実施形態 D 1 9 に記載の方法。

【0 7 9 9】

40

D 2 1 . 候補セグメントまたは検証された候補セグメントが位置する染色体の染色体表示の定量を行うことを含む、実施形態 D 1 7 から D 2 0 のいずれか 1 つに記載の方法。

【0 8 0 0】

D 2 2 . 染色体表示の定量が、z スコアによる定量である、実施形態 D 2 1 に記載の方法。

【0 8 0 1】

D 2 3 . z スコアが、染色体について、(i) 試験試料カウント表示から、(i i) 正倍数体カウント表示の中央値を減算して得た差を、(i i i) 正倍数体カウント表示の M A D で除算して得た商であり、ここで、(i) 試験試料カウント表示が、試験試料について、候補セグメントが位置する染色体中の全カウントを、全常染色体カウントで除算して

50

得た比であり、正倍数体カウント表示の (i i) 中央値が、正倍数体試料について、候補セグメントが位置する染色体中の全カウントを、全常染色体カウントで除算して得た比の中央値である、実施形態 D 2 2 に記載の方法。

【0802】

D 2 4. 候補セグメントまたは検証された候補セグメントの定量が、染色体表示の定量と比較される、実施形態 D 1 7 から D 2 3 のいずれか 1 つに記載の方法。

【0803】

D 2 5. 第 1 の候補セグメントまたは第 1 の検証された候補セグメントの z スコアによる定量が行われ、第 2 の候補セグメントまたは第 2 の検証された候補セグメントの z スコアによる定量が行われ、第 1 の候補セグメントおよび第 2 の候補セグメントが、2 つの異なるタイプのセグメンテーションから同定される、実施形態 D 2 4 に記載の方法。

10

【0804】

D 2 6. (i) 1 未満の係数で乗算した、第 1 の候補セグメントまたは検証された第 1 の候補セグメントの z スコアによる定量、および (i i) 係数で乗算した、第 2 の候補セグメントまたは検証された第 2 の候補セグメントの z スコアによる定量の最小値を決定することを含む、実施形態 D 2 5 に記載の方法。

【0805】

D 2 7. 染色体表示の z スコアによる定量結果が、最小値未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、実施形態 D 2 6 に記載の方法。

【0806】

20

D 2 8. 染色体表示の z スコアによる定量結果が、値 3. 9 5 未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、実施形態 D 2 5 に記載の方法。

【0807】

D 2 9. 試験試料について、(i) 染色体表示の z スコアによる定量結果が、値 3. 9 5 超またはそれに等しく、(i i) 染色体表示の z スコアによる定量結果が、最小値超またはそれに等しい場合に、染色体異数性が存在すると決定することを含む、実施形態 D 2 8 に記載の方法。

【0808】

D 3 0. 試験試料について、(i) 染色体表示の z スコアによる定量結果が、値 3. 9 5 未満であり、かつ/または (i i) 染色体表示の z スコアによる定量結果が、最小値未満である場合に、染色体異数性が存在しないと決定することを含む、実施形態 D 2 8 に記載の方法。

30

【0809】

D 3 1. 染色体異数性が、トリソミーまたはモノソミーである、実施形態 D 2 9 または D 3 0 に記載の方法。

【0810】

D 3 2. 第 1 の候補セグメントまたは検証された第 1 の候補セグメントの z スコアによる定量結果が、値 3. 9 5 未満、それ超、またはそれに等しいのいずれであるかを決定し、第 2 の候補セグメントまたは検証された第 2 の候補セグメントの z スコアによる定量結果が、値 3. 9 5 未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、実施形態 D 3 0 に記載の方法。

40

【0811】

D 3 4. 第 1 の候補セグメントおよび第 2 の候補セグメント、または検証されたそれらのセグメントが、実質的に同じであるかどうかを決定することを含む、実施形態 D 3 2 に記載の方法。

【0812】

D 3 5. 試験試料について、(i) 第 1 の候補セグメントまたは検証された第 1 の候補セグメントの z スコアによる定量結果が、値 3. 9 5 超またはそれに等しく、第 2 の候補セグメントまたは検証された第 2 の候補セグメントの z スコアによる定量結果が、値 3. 9 5 超またはそれに等しく、(i i) 第 1 の候補セグメントおよび第 2 の候補セグメント

50

、または検証されたそれらのセグメントが、実質的に同じである場合に、微小欠失または微小挿入が存在すると決定することを含む、実施形態 D 3 4 に記載の方法。

【0813】

D 3 6. 試験試料について、(i) 第 1 の候補セグメントもしくは検証された第 1 の候補セグメントの z スコアによる定量結果が、値 3.95 未満であり、かつ／または第 2 の候補セグメントもしくは検証された第 2 の候補セグメントの z スコアによる定量結果が、値 3.95 未満であり、かつ／または (ii) 第 1 の候補セグメントおよび第 2 の候補セグメント、もしくは検証されたそれらのセグメントが、実質的に同じではない場合に、微小欠失または微小挿入が存在しないと決定することを含む、実施形態 D 3 4 に記載の方法。

10

【0814】

D 3 7. 候補セグメントまたは検証された候補セグメントについてのカウント表示の z スコアによる定量結果を決定し、それが、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、実施形態 D 1 7 から D 2 3 のいずれか 1 つに記載の方法。

【0815】

D 3 7. 1. 染色体表示の z スコアによる定量結果を決定し、それが、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、実施形態 D 1 7 から D 2 3 のいずれか 1 つに記載の方法。

【0816】

20

D 3 8. 対数オッズ比 (LOR) を計算することを含み、LOR が、(i) (1) 遺伝子の変動を有する条件付き確率と (2) 遺伝子の変動を有する事前確率との第 1 の乗算の積と、(ii) (1) 遺伝子の変動を有さない条件付き確率と (2) 遺伝子の変動を有さない事前確率との第 2 の乗算の積との商の対数である、実施形態 D 3 7 および／または D 3 7. 1 に記載の方法。

【0817】

D 3 9. 遺伝子の変動を有する条件付き確率が、試験試料について決定された胎児フラクション、試験試料について決定されたセグメントについてのカウント表示の z スコア、およびセグメントについてのカウント表示についての z スコアの、胎児フラクションに関する分布に従って決定される、実施形態 D 3 8 に記載の方法。

30

【0818】

D 3 9. 1. 遺伝子の変動を有する条件付き確率が、等式 2 3 :

【化 3 6】

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

[式中、f は胎児フラクションであり、X は、遺伝子の変動をカバーするセグメントについての部分のカウントの合計 $X \sim f(\mu_X, \sigma_X)$ 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、 $f(\cdot)$ は分布関数である。]

40

に示す関係により決定される、実施形態 D 3 9 に記載の方法。

【0819】

D 4 0. 遺伝子の変動を有する条件付き確率が、セグメントについてのカウント表示の、試験試料に関する z スコアと、セグメントについてのカウント表示についての z スコアの、胎児フラクションに関する分布との交差部分である、実施形態 D 3 9 または D 3 9. 1 に記載の方法。

【0820】

D 4 1. 遺伝子の変動を有さない条件付き確率が、試験試料について決定されたセグメントについてのカウント表示の z スコアと、正倍数体中のセグメントについてのカウント表示についての z スコアの分布との交差部分である、実施形態 D 3 8 に記載の方法。

50

【0821】

D42. 遺伝子の変動を有する事前確率および遺伝子の変動を有さない事前確率が、試験被験体を含まない複数の試料から決定される、実施形態D38からD41のいずれか1つに記載の方法。

【0822】

D43. LORが、ゼロ超またはゼロ未満のいずれであるかを決定することを含む、実施形態D38からD42のいずれか1つに記載の方法。

【0823】

D44. 試験試料について、(i) 染色体表示のzスコアによる定量結果が、値3.95超またはそれに等しく、(ii) LORが、ゼロ超である場合に、染色体異数性が存在すると決定することを含む、実施形態D37からD43のいずれか1つに記載の方法。

10

【0824】

D45. 試験試料について、(i) 染色体表示のzスコアによる定量結果が、値3.95未満であり、かつ/または(ii) LORが、ゼロ未満である場合に、染色体異数性が存在しないと決定することを含む、実施形態D37からD43のいずれか1つに記載の方法。

【0825】

D46. 染色体異数性が、トリソミーまたはモノソミーである、実施形態D44またはD45に記載の方法。

【0826】

20

D47. 試験試料について、(i) 候補セグメントまたは検証された候補セグメントについてのカウント表示のzスコアによる定量結果が、値3.95超またはそれに等しく、(ii) LORが、ゼロ超である場合に、微小欠失または微小重複が存在すると決定することを含む、実施形態D37からD43のいずれか1つに記載の方法。

【0827】

D48. 試験試料について、(i) 候補セグメントまたは検証された候補セグメントについてのカウント表示のzスコアによる定量結果が、値3.95未満であり、かつ/または(ii) LORが、ゼロ未満である場合に、微小欠失または微小重複が存在しないと決定することを含む、実施形態D37からD43のいずれか1つに記載の方法。

【0828】

30

D49. 微小欠失が、ディジョージ症候群と関連する、実施形態D47またはD48に記載の方法。

【0829】

D49. 1. カウント表示が、正規化されたカウント表示である、実施形態D1からD49のいずれか1つに記載の方法。

【0830】

D50. (a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、システム中のマイクロプロセッサにより行われる、実施形態D1からD49. 1のいずれか1つに記載の方法。

【0831】

40

D51. (a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、コンピュータにより行われる、実施形態D1からD50のいずれか1つに記載の方法。

【0832】

D52. (a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、メモリと併せて行われる、実施形態D1からD51のいずれか1つに記載の方法。

【0833】

D53. (a)に先立って、妊娠中の雌から得られた試料中の核酸の配列決定を行い、それにより、核酸配列の読取りを得ることを含む、実施形態D1からD52のいずれか1つに記載の方法。

【0834】

50

D 5 4. (a) に先立って、核酸配列の読取りを、参照ゲノムの部分に対してマッピングすることを含む、実施形態 D 1 から D 5 3 のいずれか 1 つに記載の方法。

【0 8 3 5】

E 1. 胎児中の染色体異数性の存在または非存在を決定するための方法であって、
(a) 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントにより、染色体カウント表示を決定するステップであり、これらの配列の読取りが、胎児を出産する妊娠中の雌に由来する試験試料についての循環無細胞核酸の読取りであるステップと、

(b) 試験試料について、胎児フラクションを決定するステップと、

(c) 対数オッズ比 (L O R) を計算するステップであり、L O R が、(i) (1) 染色体異数性を有する条件付き確率と (2) 染色体異数性を有した事前確率との第 1 の乗算の積と、(i i) (1) 染色体異数性を有さない条件付き確率と (2) 染色体異数性を有さなかった事前確率との第 2 の乗算の積との商の対数であり、ここで、染色体異数性を有する条件付き確率が、(b) の胎児フラクションおよび (a) のカウント表示に従って決定されるステップと、

(d) L O R および染色体カウント表示により、染色体異数性の存在または非存在を同定するステップと

を含む方法。

【0 8 3 6】

E 1. 1. 染色体カウント表示が、染色体中の全部分についてのカウントを、常染色体中の全部分についてのカウントで除算した結果である、実施形態 E 1 に記載の方法。

【0 8 3 7】

E 2. 染色体カウント表示の z スコアによる定量を行うことを含む、実施形態 E 1 または E 1. 1 に記載の方法。

【0 8 3 8】

E 3. z スコアが、(i) 試験試料染色体カウント表示から、(i i) 正倍数体カウント表示の中央値を減算して得た差を、(i i i) 正倍数体カウント表示の M A D で除算して得た商であり、ここで、(i) 試験試料染色体カウント表示が、染色体中の部分中のカウントを、常染色体中の部分中のカウントで除算して得た比であり、正倍数体カウント表示の (i i) 中央値が、正倍数体について、染色体中の部分中のカウントを、常染色体中の部分中のカウントで除算して得た比の中央値である、実施形態 E 2 に記載の方法。

【0 8 3 9】

E 4. 遺伝子の変動を有する条件付き確率が、(b) の試験試料について決定された胎児フラクション、(a) の試験試料についての染色体カウント表示についての z スコア、および染色体カウント表示についての z スコアの胎児フラクションに特異的な分布に従って決定される、実施形態 E 1 から E 3 のいずれか 1 つに記載の方法。

【0 8 4 0】

E 5. 遺伝子の変動を有する条件付き確率が、等式 2 3 :

【化 3 7】

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

[式中、f は胎児フラクションであり、X は、染色体についての部分の合計 $X \sim f (\mu_X, \sigma_X)$ 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、 $f(\cdot)$ は分布関数である。]

に示す関係により決定される、実施形態 E 4 に記載の方法。

【0 8 4 1】

E 6. 遺伝子の変動を有する条件付き確率が、(a) の試験試料染色体カウント表示についての z スコアと、染色体カウント表示についての z スコアの胎児フラクションに特異

10

20

30

40

50

的な分布との交差部分である、実施形態 E 4 または E 5 に記載の方法。

【0842】

E 7. 染色体異数性を有さない条件付き確率が、(a) の染色体カウント表示、および正倍数体についてのカウント表示に従って決定される、実施形態 E 1 から E 6 のいずれか 1 つに記載の方法。

【0843】

E 8. 遺伝子の変動を有さない条件付き確率が、染色体カウント表示の z スコアと、正倍数体中の染色体カウント表示の z スコアの分布との交差部分である、実施形態 E 7 に記載の方法。

【0844】

E 9. 遺伝子の変動を有する事前確率および遺伝子の変動を有さない事前確率が、試験被験体を含まない複数の試料から決定される、実施形態 E 1 から E 8 のいずれか 1 つに記載の方法。

【0845】

E 10. LOR が、ゼロ超またはゼロ未満のいずれであるかを決定することを含む、実施形態 E 1 から E 9 のいずれか 1 つに記載の方法。

【0846】

E 11. 参照ゲノムの部分に対してマッピングした核酸配列の読取りのカウントが、正規化されたカウントである、実施形態 E 1 から E 10 のいずれか 1 つに記載の方法。

【0847】

E 12. カウントが、GC-LOESS 正規化を含む正規化により正規化される、実施形態 E 11 に記載の方法。

【0848】

E 13. カウントが、主成分正規化を含む正規化により正規化される、実施形態 E 11 または E 12 に記載の方法。

【0849】

E 14. カウントが、GC-LOESS 正規化を含む正規化、およびそれに続く、主成分正規化により正規化される、実施形態 E 11 から E 13 のいずれか 1 つに記載の方法。

【0850】

E 14. 1. カウントが、
(1) (i) 部分のそれぞれに対してマッピングした配列の読取りのカウントと、(ii) 部分のそれぞれについての GC 含有量との間で適合させた関係式に基づいて、試験試料についてのグアニンおよびシトシン (GC) の偏り係数を決定するサブステップであって、GC の偏り係数が、線形適合関係式の場合の勾配、または非線形適合関係式の場合の曲率の推定値であるサブステップと、
(2) マイクロプロセッサを使用して、(a) のカウント、(b) の GC の偏り係数、および部分のそれぞれについて、(i) 複数の試料のそれぞれについての GC の偏り係数と、(ii) 複数の試料についての、部分のそれぞれに対してマッピングした配列の読取りのカウントとの間で適合させた関係式に基づいて、部分のそれぞれについてのゲノム区分のレベルを計算し、それにより、ゲノム区分のレベルの計算を行うサブステップと
を含む正規化により正規化される、実施形態 E 11 から E 14 のいずれか 1 つに記載の方法。

【0851】

E 15. 染色体カウント表示の z スコアによる定量結果を決定し、それが、値 3.95 未満、それ超、またはそれに等しいのいずれであるかを決定することを含む、実施形態 E 1 から E 14. 1 のいずれか 1 つに記載の方法。

【0852】

E 16. 試験試料について、(i) 染色体カウント表示の z スコアによる定量結果が、値 3.95 超またはそれに等しく、(ii) LOR が、ゼロ超である場合に、染色体異数性が存在すると決定することを含む、E 15 に記載の方法。

【0853】

E17. 試験試料について、(i) 染色体表示のzスコアによる定量結果が、値3.95未満であり、かつ／または(ii) LORが、ゼロ未満である場合に、染色体異数性が存在しないと決定することを含む、実施形態E15に記載の方法。

【0854】

E18. 染色体異数性が、トリソミーまたはモノソミーである、実施形態E16またはE17に記載の方法。

【0855】

E18. 1. カウント表示が、正規化されたカウント表示である、実施形態E1からE18. 1のいずれか1つに記載の方法。

10

【0856】

E19. (a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、システム中のマイクロプロセッサにより行われる、実施形態E1からE18. 1のいずれか1つに記載の方法。

【0857】

E20. (a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、コンピュータにより行われる、実施形態E1からE19のいずれか1つに記載の方法。

【0858】

E21. (a)、(b)、(c)および(d)のうちの1つまたは複数または全てが、メモリと併せて行われる、実施形態E1からE20のいずれか1つに記載の方法。

20

【0859】

E22. (a)に先立って、妊娠中の雌から得られた試料中の核酸の配列決定を行い、それにより、核酸配列の読取りを得ることを含む、実施形態E1からE21のいずれか1つに記載の方法。

【0860】

E23. (a)に先立って、核酸配列の読取りを、参照ゲノムの部分に対してマッピングすることを含む、実施形態E1からE22のいずれか1つに記載の方法。

【0861】

本明細書において参照される特許、特許出願、出版物、および文書それぞれについて、その全体を、本明細書により参照によって援用する。上記特許、特許出願、出版物、および文書の引用は、上記資料のいずれかが、関連する先行技術であることを承認するものではなく、またこれらの出版物または文書の内容または日付に関していかなる承認となるものでもない。

30

【0862】

本技術の基本的な態様から逸脱せずに、上記について修正を行うことができる。本技術は、1つまたは複数の特定の実施形態を参照しながら、かなり詳細に記載されており、当業者は、本出願で具体的に開示されている実施形態に変更を行うことが可能であることを認識するであろうが、これらの修正および改良は、依然として本技術の範囲および精神内である。

【0863】

本明細書に例示として記載する本技術は、本明細書に特に開示されないエレメント（複数可）のいずれかが存在しなくても好適に実践可能である。したがって、例えば、本明細書の各事例において、用語「を含む (comprising)」、「本質的に～からなる (consisting essentially of)」、および「からなる (consisting of)」のいずれも、他方の2つの用語と置き換え可能である。採用された用語および語句は、制限ではなく説明の用語として使用され、またかかる用語および語句の使用が、示され記載された特性、またはそのセグメントと等価なものをいずれも除外するものではなく、様々な修正が、特許請求された技術の範囲内で可能である。用語「1つの(a)」または「1つの(an)」は、エレメントのうちの1つ、またはエレメントのうちの1つ超が記載されていることが文脈上明白でない限り、それが修飾する1つ

40

50

または複数のエレメントを指し得る（例えば、「試薬（a reagent）」は、1つまたは複数の試薬を意味し得る）。用語「約（about）」は、本明細書で使用する場合、基礎となるパラメータの10%以内の値を指す（すなわち、プラスまたはマイナス10%）、および連なった値の最初で用語「約」を使用する場合、その用語は値のそれぞれを修飾する（すなわち、「約1、2、および3」は、約1、約2、および約3を指す）。例えば、「約100グラム」の重量は、90グラム～110グラムの間の重量を含み得る。さらに、値の列挙が本明細書に記載される場合（例えば、約50%、60%、70%、80%、85%、または86%）、列挙には、全ての中間の値およびその分数の値（例えば、54%、85.4%）が含まれる。したがって、本技術は、代表的な実施形態および任意選択的な特性により具体的に開示されているものの、本明細書で開示する概念の修正および変更は当業者により実施可能であると理解すべきであり、かかる修正および変更は本技術の範囲内とみなされる。

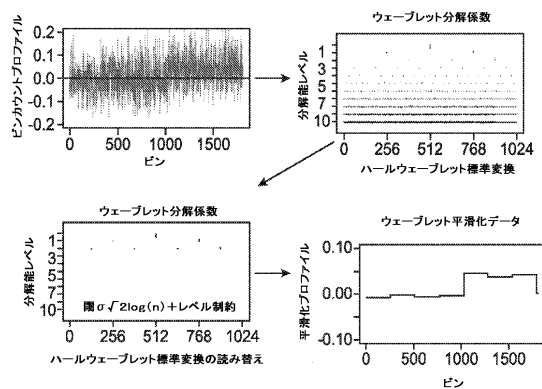
10

【0864】

本技術のある特定の実施形態を、後続する特許請求の範囲に記載する。

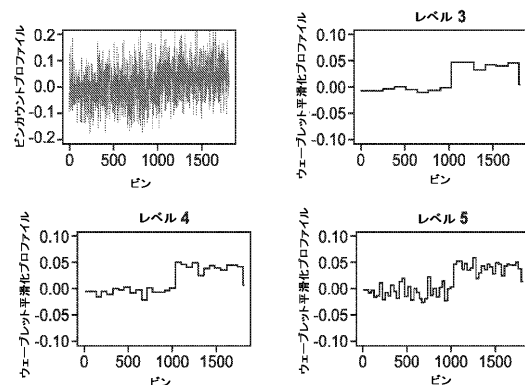
【図1】

FIG. 1



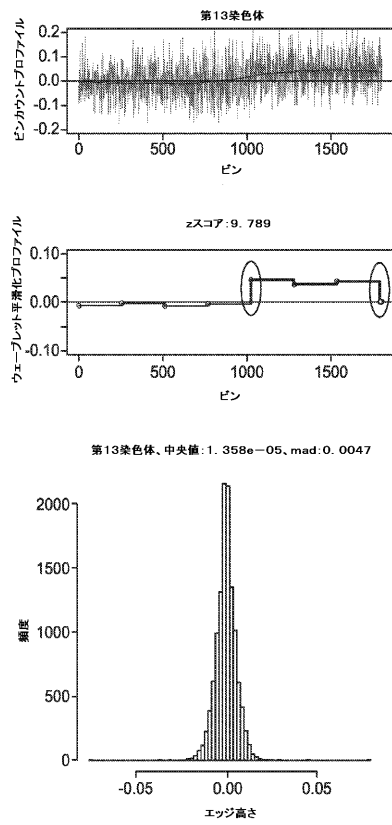
【図2】

FIG. 2



【図 3】

FIG. 3



【図 4】

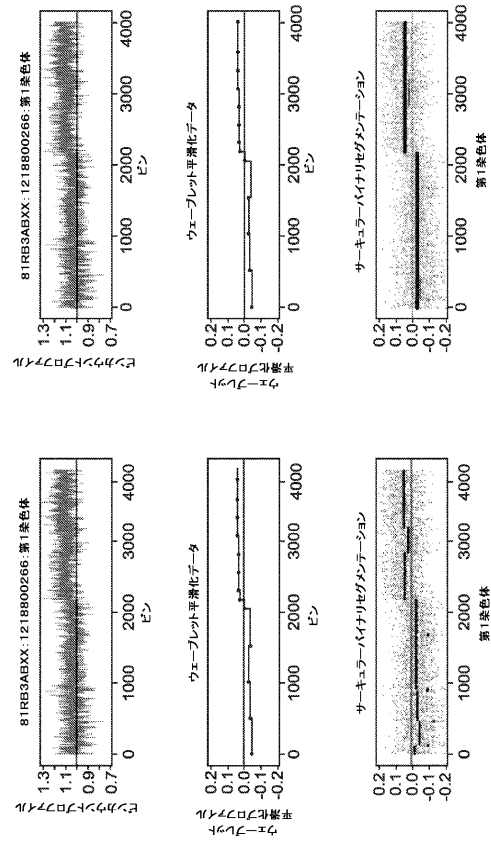
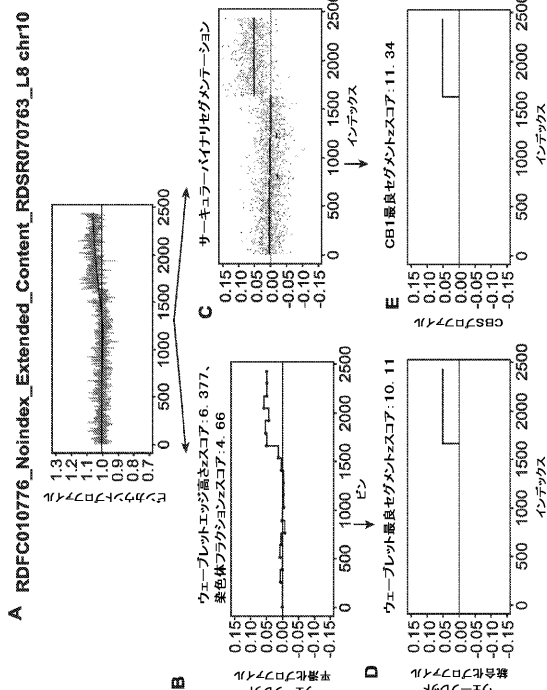


FIG. 4

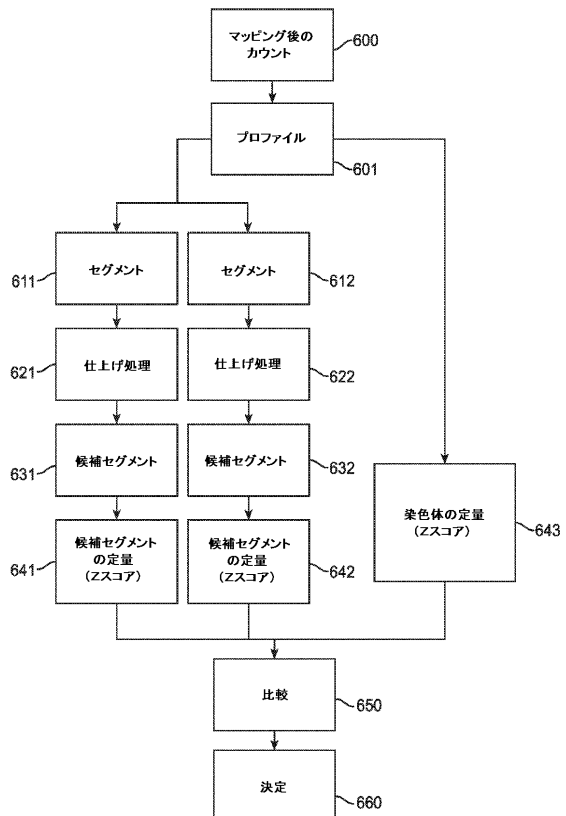
【図 5】

FIG. 5



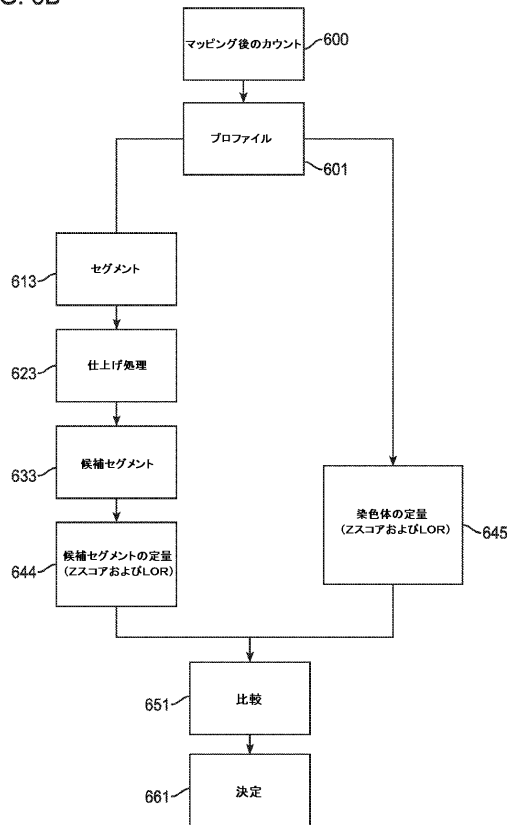
【图 6 A】

FIG. 6A



【図 6 B】

FIG. 6B



【図 7】

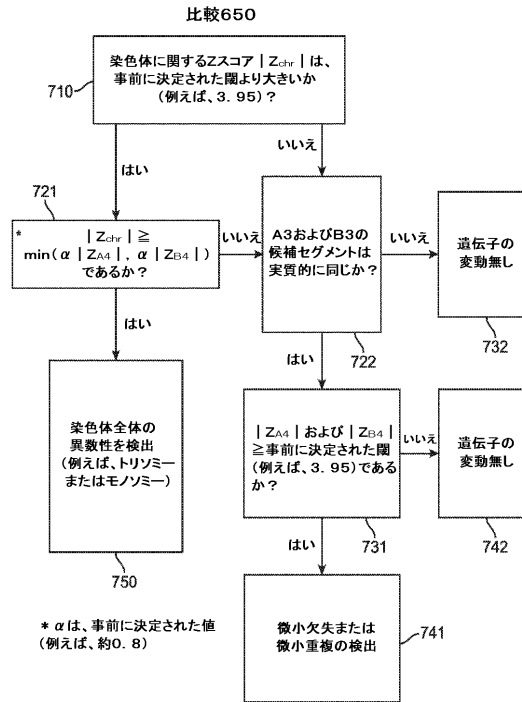


FIG. 7

【図 8】

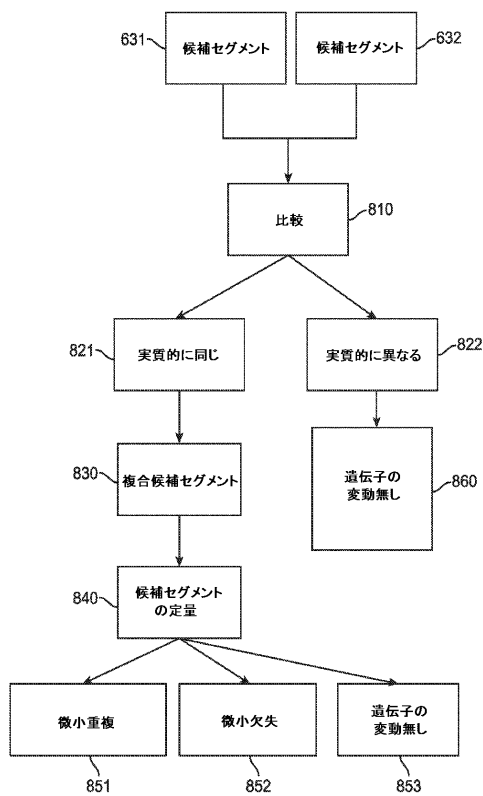


FIG. 8

【図 9】

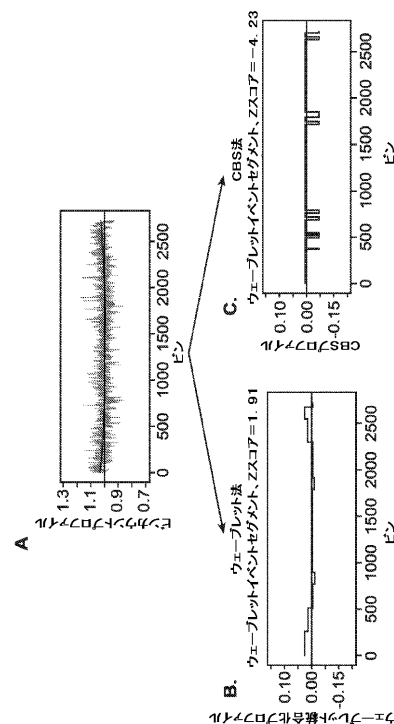


FIG. 9

【図 10-1】

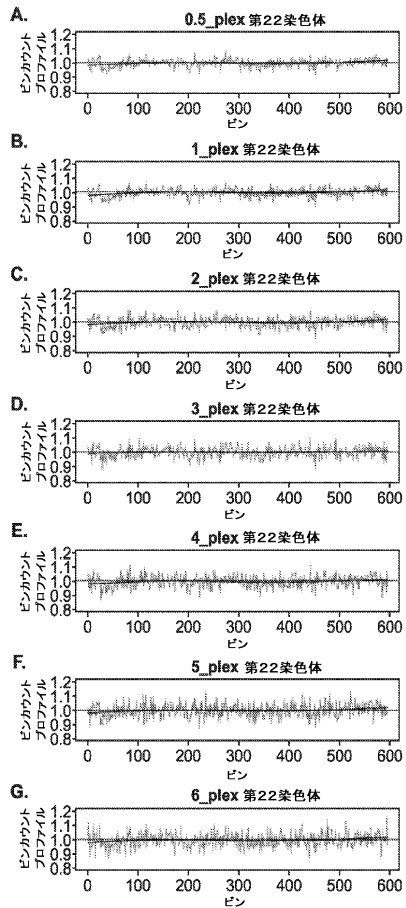


FIG. 10

【図 10-2】

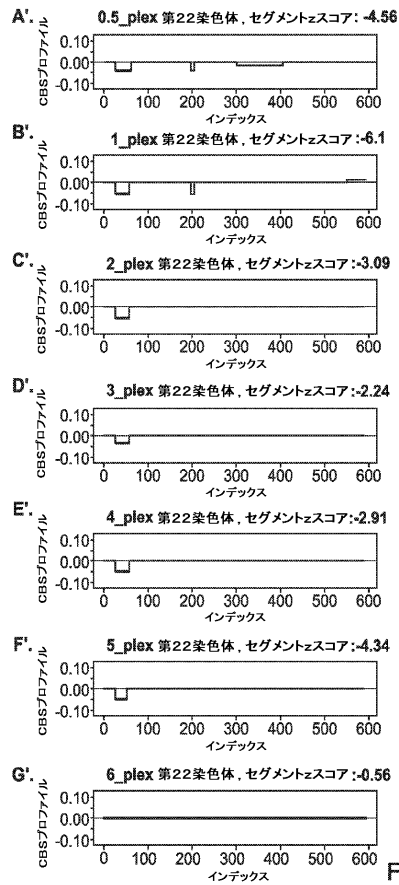
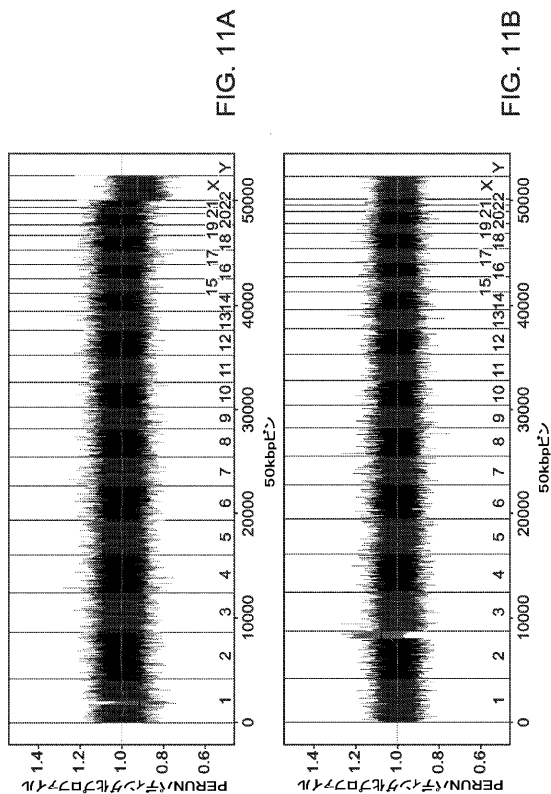


FIG. 10 (続き)

【図 11】



【図 12】

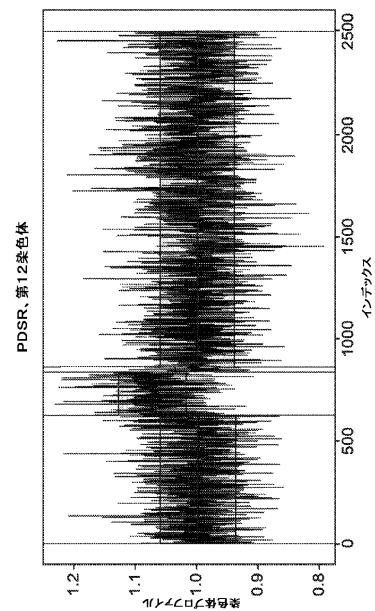


FIG. 12

【図 13-1】

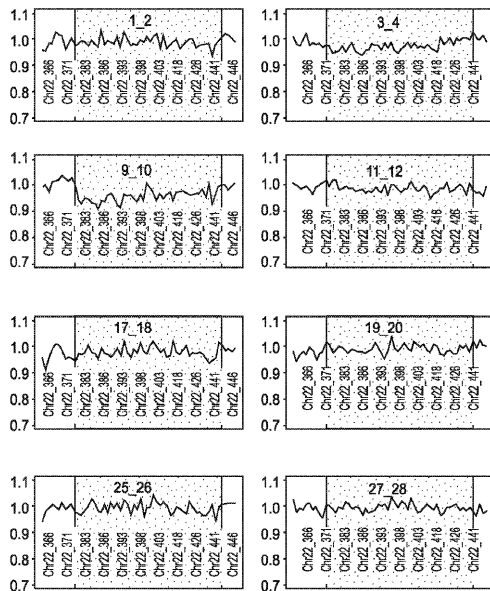


FIG. 13

【図 13-2】

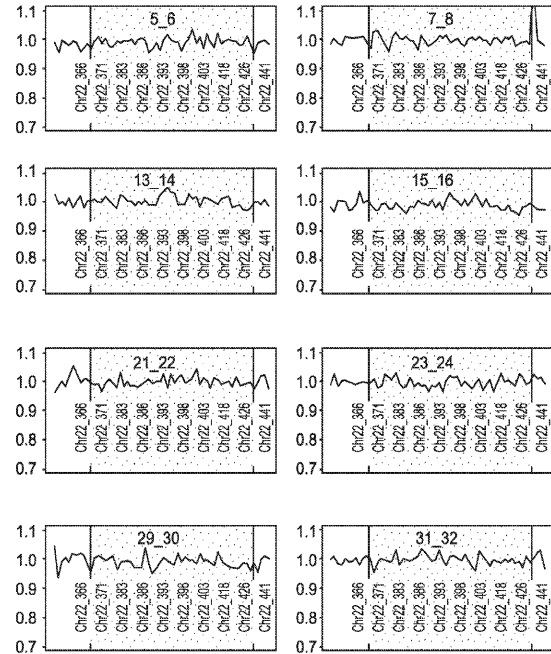


FIG. 13 (続き)

【図 14】

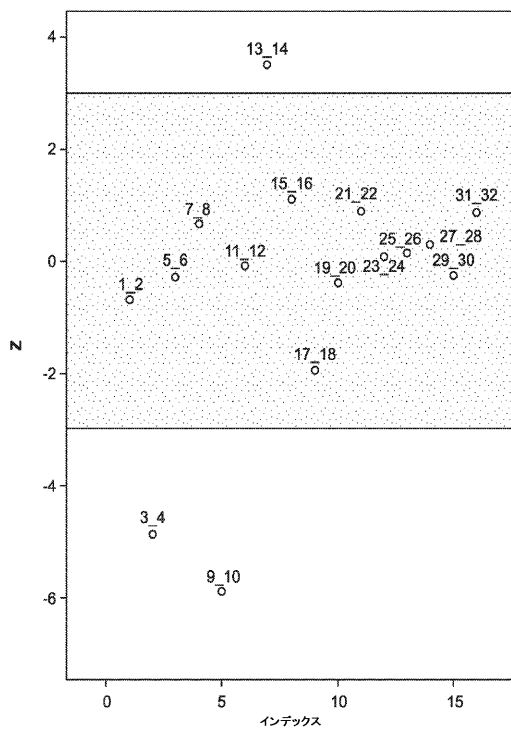


FIG. 14

【図 15】

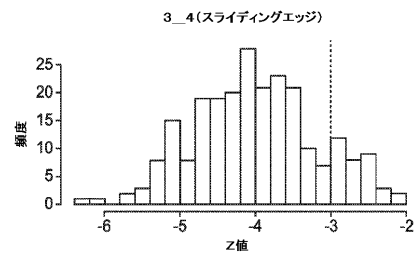


FIG. 15

【図 16】

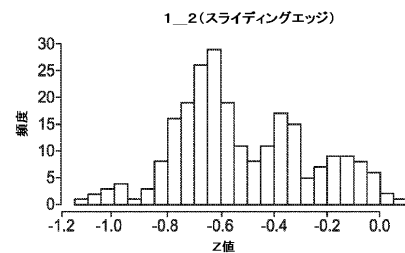


FIG. 16

【図 17】

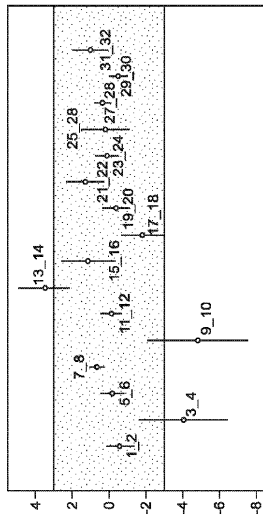
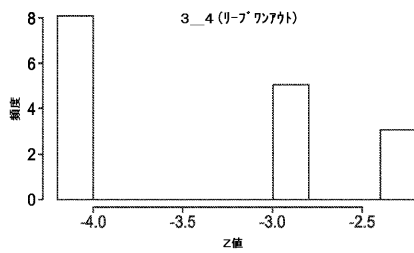


FIG. 17

【図 18】

FIG. 18



【図 21】

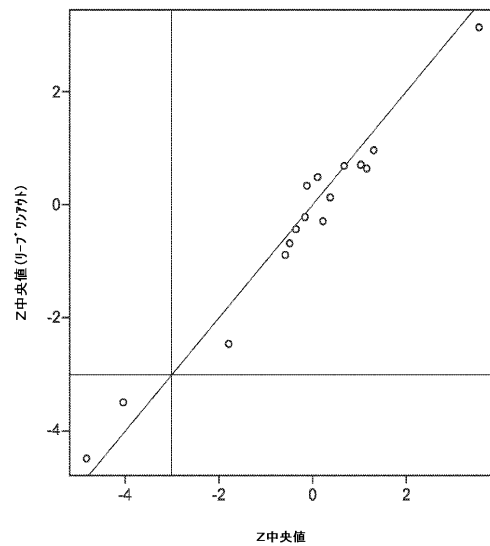
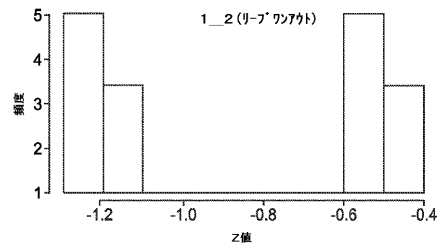


FIG. 21

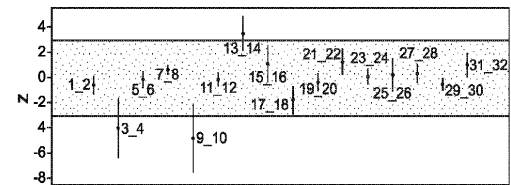
【図 19】

FIG. 19



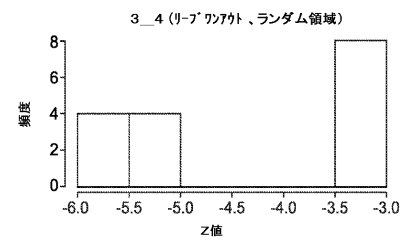
【図 20】

FIG. 20



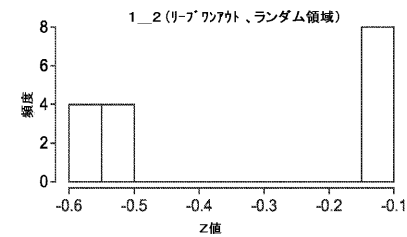
【図 22】

FIG. 22



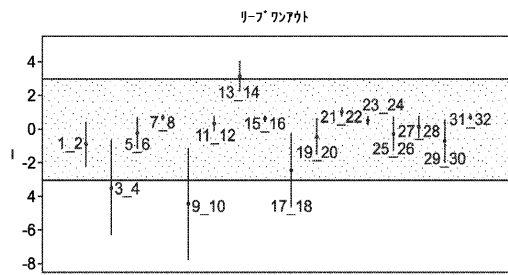
【図 23】

FIG. 23



【図 24】

FIG. 24



【図 25】

3_4 (リーフ' ワンアクト + スライディングエッジ)

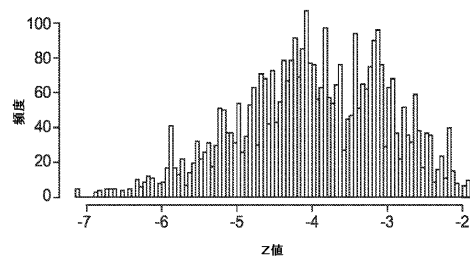


FIG. 25

【図 26】

1_2 (リーフ' ワンアクト + スライディングエッジ)

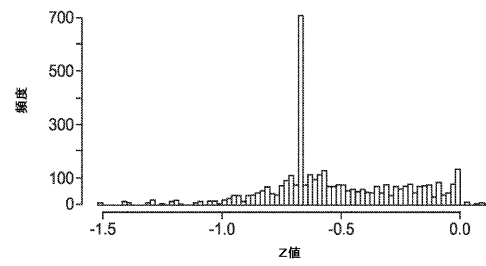


FIG. 26

【図 27】

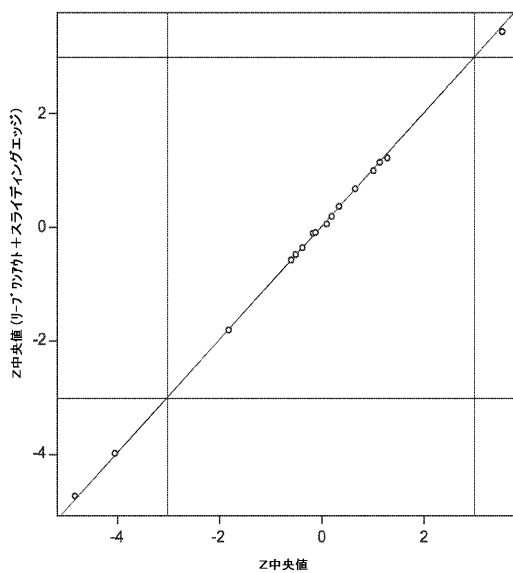


FIG. 27

【図 28】

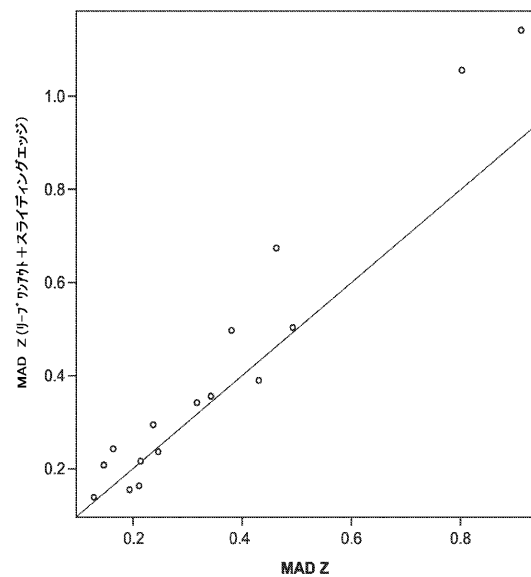


FIG. 28

【図 29】

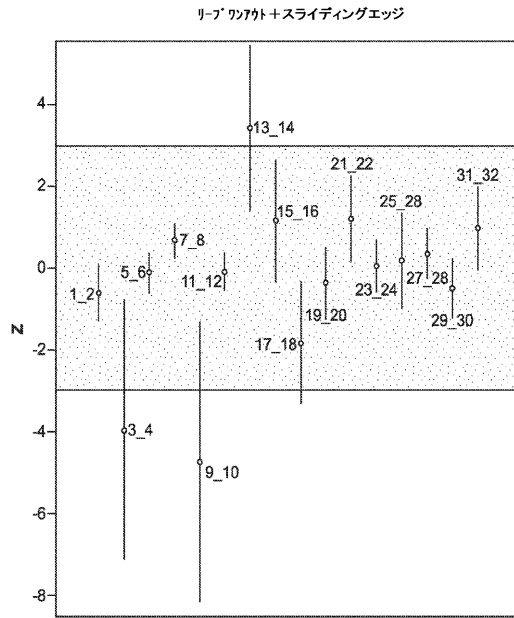


FIG. 29

【図 30】

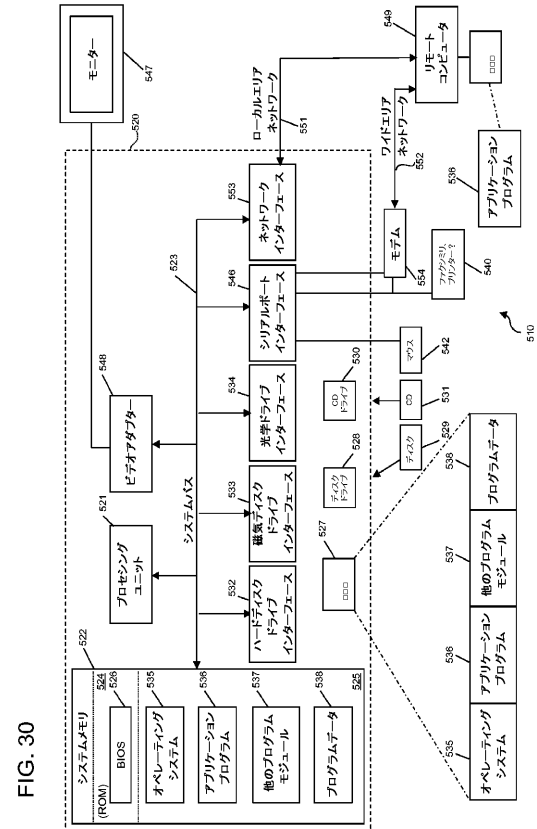
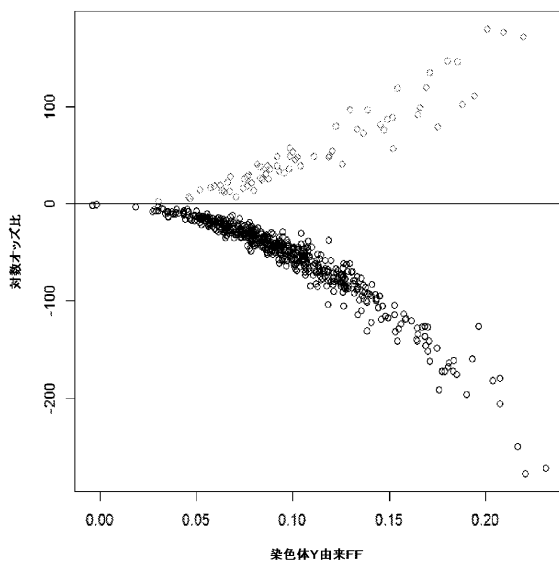


FIG. 30

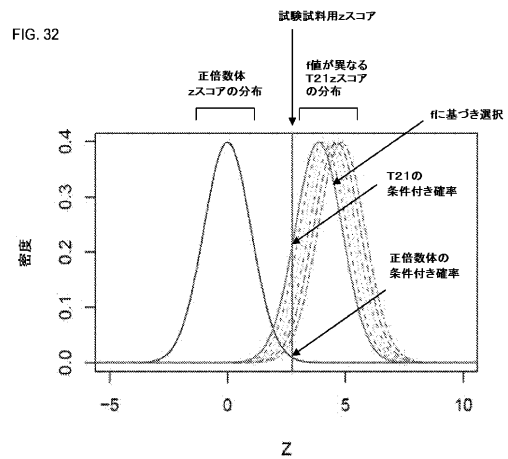
【図 31】

FIG. 31



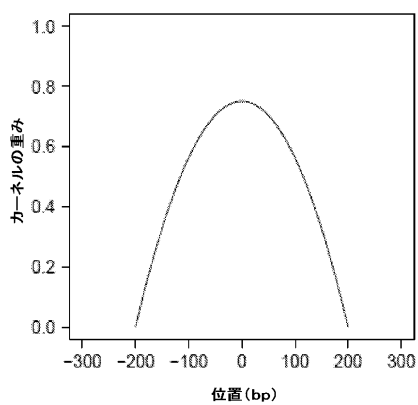
【図 32】

FIG. 32



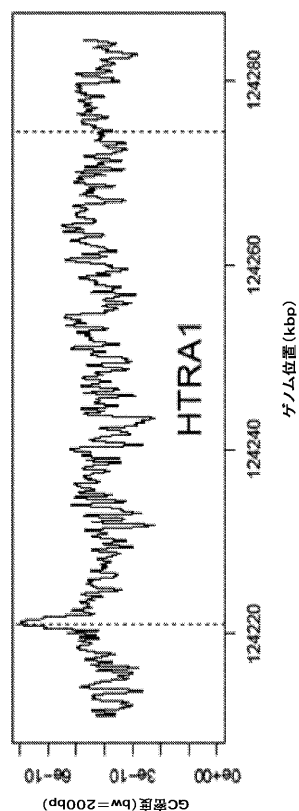
【図 3 3】

FIG. 33



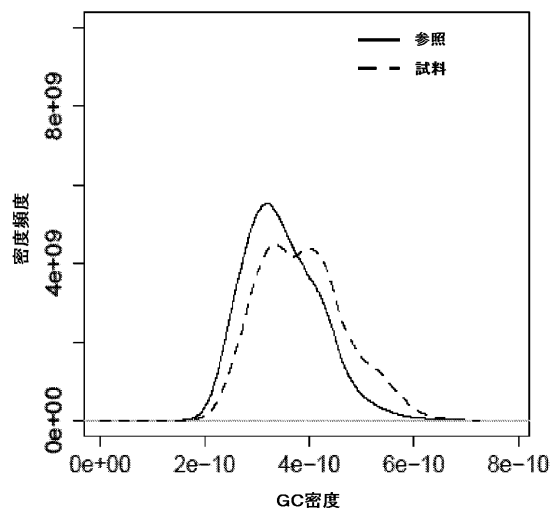
【図 3 4】

FIG. 34



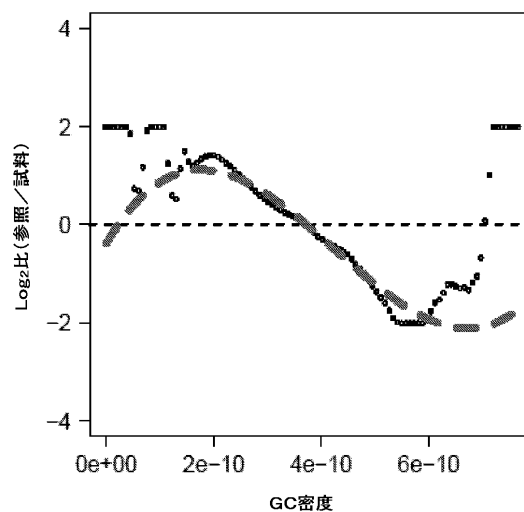
【図 3 5】

FIG. 35

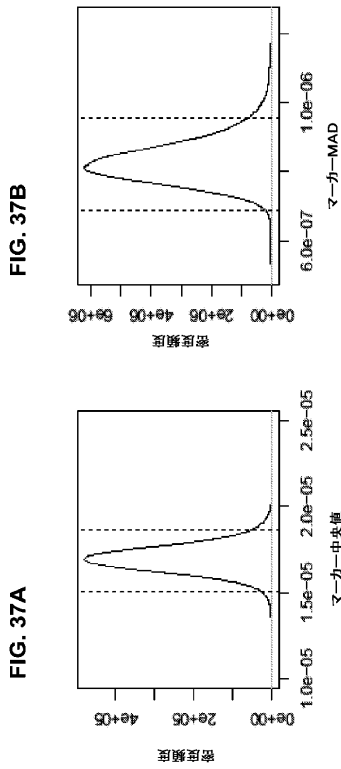


【図 3 6】

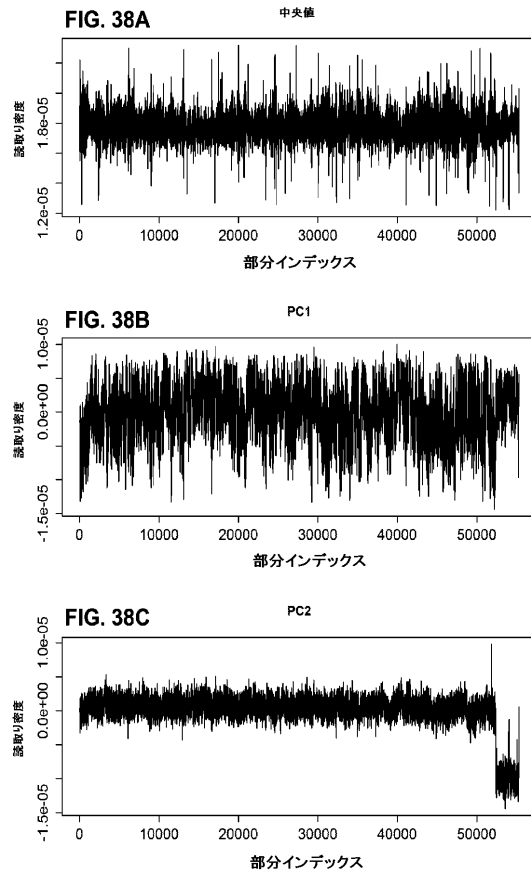
FIG. 36



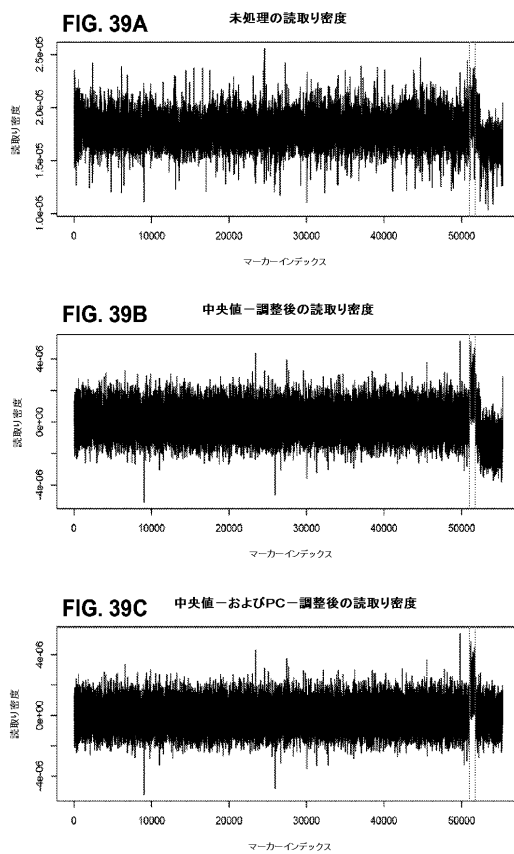
【図 37】



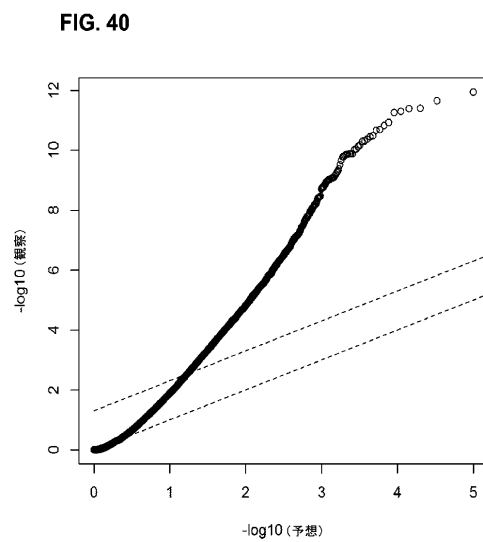
【図 38】



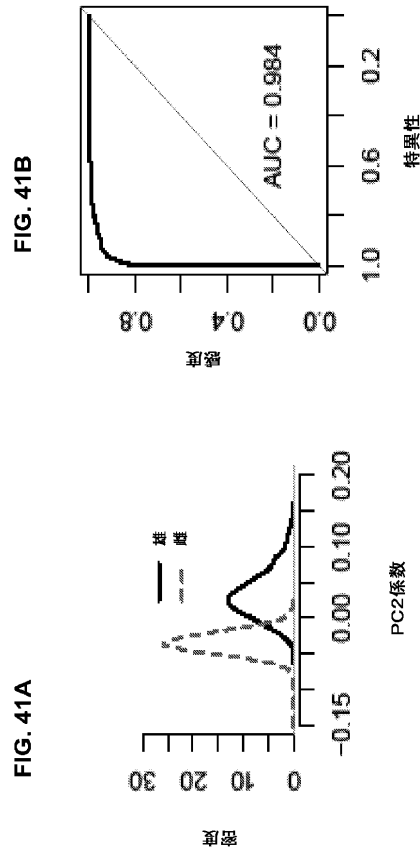
【図 39】



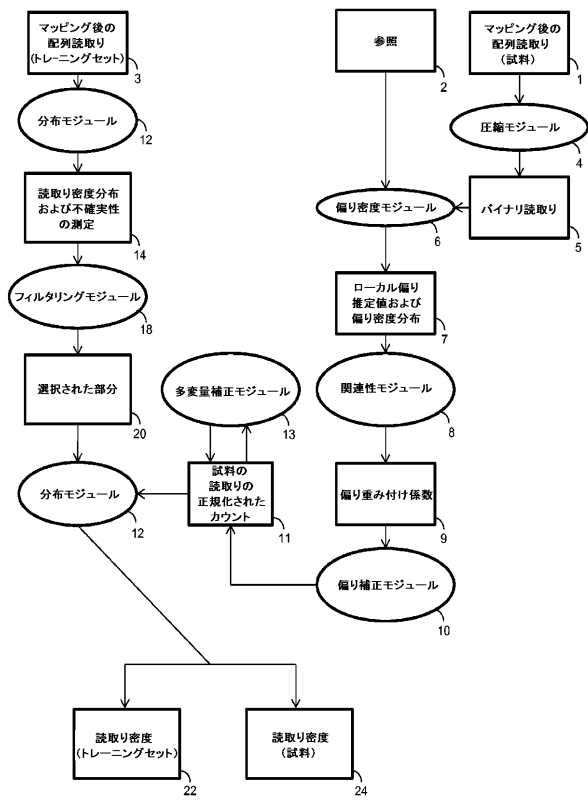
【図 40】



【図 4 1】

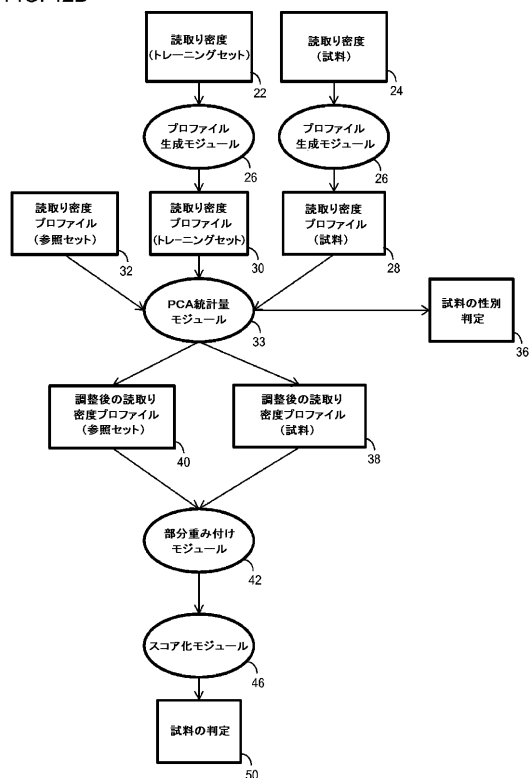


【図 4 2 A】



【図 4 2 B】

FIG. 42B



フロントページの続き

- (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 アパートメント 4124
- (72)発明者 ザクラ, ジェリコ
アメリカ合衆国 カリフォルニア 92129, サン ディエゴ, サンダンス アベニュー
12830
- (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 1
0545
- (72)発明者 キム, サン キュン
アメリカ合衆国 カリフォルニア 91206, グレンデール, グレンモア ブールバード
662
- (72)発明者 マズルーム, アミン
アメリカ合衆国 カリフォルニア 92128, サン ディエゴ, ケープウッド レーン 1
4013
- (72)発明者 ハンナム, グレゴリー
アメリカ合衆国 カリフォルニア 92129, サン ディエゴ, ババリアン ドライブ 1
3125
- (72)発明者 エーリック, マシアス
アメリカ合衆国 カリフォルニア 92109, サン ディエゴ, ダイヤモンド ストリート
1947

審査官 戸来 幸男

- (56)参考文献 国際公開第2013/052913 (WO, A2)
BMJ, 2011年, vol.342, pp.C7401(1-9)

- (58)調査した分野(Int.Cl., DB名)
C12Q 1/68-1/6897
G16B 30/00-30/20
C12N 15/00-15/90
CAPLUS/MEDLINE/EMBASE/BIOSIS/
WPIDS (STN)
JSTPLUS/JMEDPLUS/JST7580 (JDreamIII)
PubMed