

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6431769号
(P6431769)

(45) 発行日 平成30年11月28日(2018.11.28)

(24) 登録日 平成30年11月9日(2018.11.9)

(51) Int. Cl.	F I
C 1 2 Q 1/6827 (2018.01)	C 1 2 Q 1/6827 Z
G 0 6 F 19/20 (2011.01)	G 0 6 F 19/20
C 1 2 N 15/09 (2006.01)	C 1 2 N 15/09 Z

請求項の数 15 (全 173 頁)

(21) 出願番号	特願2014-553484 (P2014-553484)	(73) 特許権者	504159534
(86) (22) 出願日	平成25年1月18日 (2013.1.18)		セクエノム, インコーポレイテッド
(65) 公表番号	特表2015-513392 (P2015-513392A)		アメリカ合衆国 カリフォルニア 921
(43) 公表日	平成27年5月14日 (2015.5.14)		21, サン ディエゴ, ジョン ホブ
(86) 国際出願番号	PCT/US2013/022290		キンズ コート 3595
(87) 国際公開番号	W02013/109981	(74) 代理人	100078282
(87) 国際公開日	平成25年7月25日 (2013.7.25)		弁理士 山本 秀策
審査請求日	平成27年12月17日 (2015.12.17)	(74) 代理人	100113413
(31) 優先権主張番号	61/709,899		弁理士 森下 夏樹
(32) 優先日	平成24年10月4日 (2012.10.4)	(74) 代理人	100181674
(33) 優先権主張国	米国 (US)		弁理士 飯田 貴敏
(31) 優先権主張番号	61/589,202	(74) 代理人	100181641
(32) 優先日	平成24年1月20日 (2012.1.20)		弁理士 石川 大輔
(33) 優先権主張国	米国 (US)	(74) 代理人	230113332
			弁護士 山本 健策

最終頁に続く

(54) 【発明の名称】 実験条件を要因として含める診断プロセス

(57) 【特許請求の範囲】

【請求項1】

胎児の異数性の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、妊婦から得られた循環している無細胞核酸を含む試験試料から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(b) グアニンおよびシトシンの含量に従って(a)における前記カウントを補正し、それによって補正されたカウントを提供するステップであって、前記補正は、

(i) (1) 前記試験試料について前記セクションのそれぞれにマッピングされた前記配列読み取りのカウントと(2) 前記セクションのそれぞれについてのGC含量との間のフィッティングした線形関係に基づいて、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ことであって、前記試験試料についての前記GCの偏り係数が、(i)(1)と(i)(2)との間のフィッティングした線形関係の傾きである、ことと、

(ii) 方程式 :

$$L_i = (m_i - G_i S) / I \quad \text{方程式}$$

(式中、 L_i が前記試験試料についての前記セクションのそれぞれについてのゲノミックセクションの高度であり、 m_i が前記試験試料についての前記セクションのそれぞれにマッピングされた前記配列読み取りのカウントであり、 i が前記試験試料であり、 G_i が(i)で決定された前記試験試料についてのGCの偏り係数であり、 I が、(1)(i)複

数の試料のそれぞれについての前記セクションのそれぞれにマッピングされた配列読み取りのカウントと (i i) 前記セクションのそれぞれについての G C 含量との間のフィッティングした線形関係に基づく、前記複数の試料のそれぞれについての G C の偏り係数であって、前記複数の試料のそれぞれについての前記 G C の偏り係数が、 (i i) (1) (i) と (i i) (1) (i i) との間のフィッティングした線形関係についての傾きである、G C の偏り係数と (2) 前記複数の試料についての前記セクションのそれぞれにマッピングされた前記配列読み取りのカウントとの間の、前記セクションのそれぞれについてのフィッティングした線形関係の切片であり、S が前記セクションのそれぞれについての (i i) (1) と (i i) (2) との間のフィッティングした線形関係の傾きである) に従って前記セクションのそれぞれについてのゲノミックセクションの高度を算出し、それによって、前記補正されたカウントであるゲノミックセクションの高度を提供すること

10

を含むステップと、

(c) 参照ゲノムセクションにマッピングされた、妊婦の群から得られた循環している無細胞核酸を含む試料の群における各試料から得られるヌクレオチド配列読み取りのカウントを得るステップであって、(a) における前記試験試料と (c) における前記試料の群は、共通のフローセルユニット、同じコンテナに由来するフローセル、同じロットに由来するフローセルまたは同じ製造運転で製造されたフローセル、共通の試薬プレートユニット、同じコンテナに由来する試薬プレート、あるいは同じロットに由来する試薬プレートまたは同じ製造運転で製造された試薬プレートから選択される 1 つまたは複数の共通の実験条件に曝露される、ステップであって、(c) において得られる前記カウントが、グアニンおよびシトシンの含量に従って補正され、それによって補正されたカウントが提供され、ここで、(c) において得られる前記カウントが、以下：

20

(i) (1) 前記試料の群における各試料について前記セクションのそれぞれにマッピングされた前記配列読み取りのカウントと (2) 前記セクションのそれぞれについての G C 含量との間のフィッティングした線形関係に基づいて、前記試料の群における各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ことであって、前記試料の群における各試料についての前記 G C の偏り係数が、(i) (1) と (i) (2) との間のフィッティングした線形関係の傾きである、ことと、

30

(i i) 方程式 :

$$L_i = (m_i - G_i S) / I \quad \text{方程式}$$

(式中、 L_i が前記試料の群における各試料についての前記セクションのそれぞれについてのゲノミックセクションの高度であり、 m_i が前記試料の群における各試料についての前記セクションのそれぞれにマッピングされた前記配列読み取りのカウントであり、 i が前記試料の群における各試料であり、 G_i が (i) で決定された前記試料の群における各試料についての G C の偏り係数であり、I が、(1) (i) 前記複数の試料のそれぞれについての前記セクションのそれぞれにマッピングされた配列読み取りのカウントと (i i) 前記セクションのそれぞれについての G C 含量との間のフィッティングした線形関係に基づく、前記複数の試料のそれぞれについての G C の偏り係数であって、前記複数の試料のそれぞれについての前記 G C の偏り係数が、(i i) (1) (i) と (i i) (1) (i i) との間のフィッティングした線形関係についての傾きである、G C の偏り係数と (2) 前記複数の試料についての前記セクションのそれぞれにマッピングされた前記配列読み取りのカウントとの間の、前記セクションのそれぞれについてのフィッティングした線形関係の切片であり、S が前記セクションのそれぞれについての (i i) (1) と (i i) (2) との間のフィッティングした線形関係の傾きである)

40

に従って前記セクションのそれぞれについてのゲノミックセクションの高度を算出し、それによって、前記補正されたカウントであるゲノミックセクションの高度を提供すること

によって補正される、ステップと、

(d) (c) における前記試料の群から得られる前記補正されたカウントに従って得ら

50

れる予測カウントまたは前記予測カウントの誘導値に従って、(b)における前記補正されたカウントを第1のゲノムセクションについて正規化、または(b)における前記補正されたカウントの誘導値を前記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e)前記正規化された試料カウントに基づいて、前記試験試料について胎児の異数性の有無を検出するステップと

を含む、方法。

【請求項2】

前記試料核酸が前記妊婦由来の血漿由来のものである、請求項1に記載の方法。

【請求項3】

前記試料核酸が前記妊婦由来の血清由来のものである、請求項1または2に記載の方法。

【請求項4】

前記胎児の異数性が13トリソミーである、請求項1、2または3に記載の方法。

【請求項5】

前記胎児の異数性が18トリソミーである、請求項1、2または3に記載の方法。

【請求項6】

前記胎児の異数性が21トリソミーである、請求項1、2または3に記載の方法。

【請求項7】

前記予測カウントがカウント中央値である、請求項1から6のいずれか一項に記載の方法。

【請求項8】

前記予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、請求項1から6のいずれか一項に記載の方法。

【請求項9】

前記カウントを正規化する前記ステップが、パーセント表示を決定するステップを含む、請求項1から8のいずれか一項に記載の方法。

【請求項10】

前記正規化されたカウントがzスコアである、請求項1から9のいずれか一項に記載の方法。

【請求項11】

前記正規化されたカウントがロバストなzスコアである、請求項1から10のいずれか一項に記載の方法。

【請求項12】

前記第1のゲノミックセクションについての前記カウントの前記誘導値が前記第1のゲノミックセクションのパーセント表示である、請求項1から11のいずれか一項に記載の方法。

【請求項13】

前記中央値がパーセント表示の中央値である、請求項7に記載の方法。

【請求項14】

前記パーセント表示が染色体表示である、請求項13に記載の方法。

【請求項15】

前記核酸を配列決定し、それによって、核酸配列読み取りを提供するステップ、および前記核酸配列読み取りを、参照ゲノムのゲノミックセクションにマッピングするステップを含む、請求項1から14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】

【0001】

関連特許出願

10

20

30

40

50

この特許出願は、2012年1月20日に出願された表題「DIAGNOSTIC PROCESSES THAT FACTOR EXPERIMENTAL CONDITIONS」、発明者Cosmin Deciuの米国仮特許出願第61/589,202号(代理人管理番号SEQ-6040-PVと指定)の利益を主張し、2012年10月5日に出願された表題「METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS」、発明者Cosmin Deciu、Zeljko Dzakula、Mathias EhrichおよびSung Kimの米国PCT出願第PCT/US2012/059123号(代理人管理番号SEQ-6034-PCと指定)の利益を主張し、2012年10月4日に出願された表題「METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS」、発明者Cosmin Deciu、Zeljko Dzakula、Mathias EhrichおよびSung Kimの米国仮特許出願第61/709,899号(代理人管理番号SEQ-6034-PV3と指定)の利益を主張し、そして2012年6月22日に出願された表題「METHODS AND PROCESSES FOR NON-INVASIVE ASSESSMENT OF GENETIC VARIATIONS」、発明者Zeljko DzakulaおよびMathias Ehrichの米国仮特許出願第61/663,477号(代理人管理番号SEQ-6034-PV2と指定)の利益を主張する。

10

【0002】

20

分野

当該技術は、一部において、遺伝的変異を非侵襲的に評価するための方法、プロセスおよび装置に関する。

【背景技術】

【0003】

生きている生物体(例えば、動物、植物および微生物)の遺伝情報ならびに複製される遺伝情報の他の形態(例えば、ウイルス)は、デオキシリボ核酸(DNA)またはリボ核酸(RNA)にコードされている。遺伝情報とは、化学的または仮説的な核酸の一次構造を示すヌクレオチドまたは修飾されたヌクレオチドの継承である。ヒトでは、完全なゲノムは24の染色体に位置する約30,000遺伝子を含有する(The Human Genome、T. Strachan、BIOS Scientific Publishers、1992年を参照されたい)。各遺伝子は特定のタンパク質をコードし、それが転写および翻訳によって発現されると、生細胞内の特定の生化学的機能が満たされる。

30

【0004】

多くの医学的状態は、1つまたは複数の遺伝的変異によって引き起こされる。ある特定の遺伝的変異は、例えば、血友病、サラセミア、デュシェンヌ型筋ジストロフィー(DMD)、ハンチントン病(HD)、アルツハイマー病および嚢胞性線維症(CF)を含む医学的状態を引き起こす(Human Genome Mutations、D. N. CooperおよびM. Krawczak、BIOS Publishers、1993年)。そのような遺伝病は、特定の遺伝子のDNAにおける単一ヌクレオチドの付加、置換、または欠失によって生じる可能性がある。ある特定の先天性欠損は、異数性とも称される染色体異常によって引き起こされる。例えば、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)など、およびある特定の性染色体異数性、例えば、クラインフェルター症候群(XXY)など。いくつかの遺伝的変異により、例えば、糖尿病、動脈硬化症、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)などのいくつかの疾患のいずれかに対する素因が個体に付される、またはそれらの疾患のいずれかが引き起こされる可能性がある。

40

【先行技術文献】

【非特許文献】

50

【0005】

【非特許文献1】The Human Genome、T. Strachan、BIO S Scientific Publishers、1992年

【非特許文献2】Human Genome Mutations、D. N. CooperおよびM. Krawczak、BIOS Publishers、1993年

【発明の概要】

【発明が解決しようとする課題】

【0006】

1つまたは複数の遺伝的変異または遺伝分散を同定することが、特定の医学的状態の診断、またはそれに対する素因の決定につながる可能性がある。遺伝分散を同定することにより、医学的決定を容易にし、かつ/または有用な医学的手順を使用することができる。

10

【課題を解決するための手段】

【0007】

本明細書では、胎児の異数性の有無を検出するための方法であって、(a)妊婦由来の循環している無細胞(cell-free)核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、(b)ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(c)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(d)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値(derivative)に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(e)正規化された試料カウントから、胎児の異数性の有無を決定するアウトカムをもたらすステップとを含む方法が提供される。いくつかの実施形態では、配列読み取りを参照ゲノムセクションの一部または全部にマッピングする。

20

【0008】

本明細書では、胎児の異数性の有無を検出するための方法であって、(a)妊婦由来の循環している無細胞核酸を含む試料を得るステップと、(b)試料から試料核酸を単離するステップと、(c)試料核酸からヌクレオチド配列読み取りを得るステップと、(d)ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(e)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(f)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(g)正規化された試料カウントから、胎児の異数性の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

30

【0009】

本明細書では、胎児の異数性の有無を検出するための方法であって、(a)妊婦由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(b)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(c)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(d)正規化された試料カウントから、胎児の異数性の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

40

【0010】

本明細書では、遺伝的変異の有無を検出するための方法であって、(a)試験被験体由

50

来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、(b)ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(c)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(d)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(e)正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

10

【0011】

本明細書では、胎児の異数性の有無を検出するための方法であって、(a)試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、(b)試料から試料核酸を単離するステップと、(c)試料核酸からヌクレオチド配列読み取りを得るステップと、(d)ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(e)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(f)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(g)正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

20

【0012】

本明細書では、遺伝的変異の有無を検出するための方法であって、(a)試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(b)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(c)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(d)正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

30

【0013】

本明細書では、遺伝的変異の有無を検出するための方法であって、(a)試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、(b)ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(c)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(d)(c)のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、(e)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(d)の後に残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(f)正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

40

【0014】

50

本明細書では、遺伝的変異の有無を検出するための方法であって、(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、(b) 試料から試料核酸を単離するステップと、(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、(d) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(f) (e) のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、(g) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(f) の後に残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(h) 正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

10

【0015】

本明細書では、遺伝的変異の有無を検出するための方法であって、(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(c) (b) のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(c) の後に残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(e) 正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

20

【0016】

本明細書では、微小欠失の有無を検出するための方法であって、(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、(b) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(d) (c) のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、(e) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(d) の後に残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(f) 正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

30

40

【0017】

本明細書では、微小欠失の有無を検出するための方法であって、(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、(b) 試料から試料核酸を単離するステップと、(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、(d) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(f) (e) のカウントされ、マッピング

50

された配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、(g) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(f)の後に残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(h) 正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法が提供される。

【0018】

本明細書では、微小欠失の有無を検出するための方法であって、(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(c) (b)のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(c)の後に残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(f) 正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

【0019】

いくつかの実施形態では、補正され、カウントされ、マッピングされた配列読み取りを、1つまたは複数の実験条件についてさらに補正した後に、残ったカウントを正規化する。ある特定の実施形態では、遺伝的変異は微小欠失である。いくつかの実施形態では、微小欠失は第22染色体上にある。ある特定の実施形態では、微小欠失は第22染色体の領域22q11.2において起こっている。いくつかの実施形態では、微小欠失は、参照ゲノムhg19による第22染色体のヌクレオチド19,000,000位と22,000,000位の間に起こっている。

【0020】

いくつかの実施形態では、試料核酸は試験被験体由来の血漿由来であり、ある特定の実施形態では、試料核酸は試験被験体由来の血清由来である。いくつかの実施形態では、試験被験体は、ヒト、動物、および植物から選択される。ある特定の実施形態では、ヒト試験被験体は女性、妊婦、男性、胎児、または新生児を含む。

【0021】

いくつかの実施形態では、胎児の異数性は13トリソミーである。ある特定の実施形態では、胎児の異数性は18トリソミーである。いくつかの実施形態では、胎児の異数性は21トリソミーである。

【0022】

ある特定の実施形態では、遺伝的変異は医学的状態に関連付けられる。いくつかの実施形態では、医学的状態はがんである。ある特定の実施形態では、医学的状態は異数性である。

【0023】

いくつかの実施形態では、無細胞試料核酸の配列読み取りはポリヌクレオチド断片の形態である。ある特定の実施形態では、ポリヌクレオチド断片の長さは約20ヌクレオチドから約50ヌクレオチドの間である。いくつかの実施形態では、ポリヌクレオチドの長さは約30ヌクレオチドから約40ヌクレオチドの間である。いくつかの実施形態では、「ポリヌクレオチド断片」という用語は、配列読み取りに関して「配列情報」という用語、または身体DNAのデジタル表示と同義またはそれと互換的であり得、または逆もまた同

10

20

30

40

50

様である。

【 0 0 2 4 】

ある特定の実施形態では、予測カウントは、カウント中央値である。いくつかの実施形態では、予測カウントは、トリムもしくは刈り込み平均 (trimmed or truncated mean)、ウィンザー化平均またはブートストラップ推定値である。ある特定の実施形態では、正規化された試料カウントは、カウントの誘導値を第1のゲノムセクションについて正規化することを含むプロセスによって得られ、誘導値は、第1のゲノムセクションについてのカウントを第1のゲノムセクションを含む複数のゲノムセクションについてのカウントで割ることによって決定される、第1のゲノムセクションカウント表示である。いくつかの実施形態では、第1のゲノムセクションについてのカウントの誘導値を予測カウントの誘導値に従って正規化し、予測カウントの誘導値は、第1のゲノムセクションについての予測カウントを第1のゲノムセクションを含む複数のゲノムセクションについての予測カウントで割ることによって決定される、予測された第1のゲノムセクションカウント表示である。ある特定の実施形態では、第1のゲノムセクションは染色体または染色体の部分であり、複数のゲノムセクションは常染色体を含む。いくつかの実施形態では、染色体は第21染色体、第18染色体または第13染色体である。

10

【 0 0 2 5 】

ある特定の実施形態では、正規化された試料カウントは、第1のゲノムセクションについてのカウントから予測カウントを引き算し、それにより減算値を生成し、減算値をカウントの変動性の推定値で割ることを含むプロセスによって得られる。いくつかの実施形態では、正規化された試料カウントは、第1のゲノムセクションカウント表示から、予測された第1のゲノムセクションカウント表示を引き算し、それにより減算値を生成し、減算値を第1のゲノムセクションカウント表示の変動性の推定値で割ることを含むプロセスによって得られる。ある特定の実施形態では、予測カウントの変動性の推定値は、カウントの中央絶対偏差 (MAD) である。いくつかの実施形態では、カウントの変動性の推定値は、RousseeuwおよびCrouxによって導入されるMADの代替値、またはブートストラップ推定値である。

20

【 0 0 2 6 】

いくつかの実施形態では、1つまたは複数の共通の実験条件はフローセルを含む。ある特定の実施形態では、1つまたは複数の共通の実験条件はフローセル内のチャンネルを含む。いくつかの実施形態では、1つまたは複数の共通の実験条件は試薬プレートを含む。ある特定の実施形態では、試薬プレートを使用して配列決定のために核酸を段階分けする。いくつかの実施形態では、試薬プレートを使用して配列決定のために核酸ライブラリーを調製する。ある特定の実施形態では、1つまたは複数の共通の実験条件は同定タグ指標を含む。

30

【 0 0 2 7 】

ある特定の実施形態では、正規化された試料カウントを、ヌクレオチド配列読み取りまたは試料核酸のグアニンおよびシトシンの含量について補正する。いくつかの実施形態では、本明細書に記載の方法は、カウントまたは正規化された試料カウントを局所重み付け多項式回帰に供するステップを含む。ある特定の実施形態では、局所重み付け多項式回帰は、LOESS回帰またはLOWESS回帰である。いくつかの実施形態では、正規化された試料カウントを、参照ゲノムセクションにおいて反復するヌクレオチド配列について補正する。ある特定の実施形態では、カウントまたは正規化された試料カウントを、参照ゲノムセクションにおいて反復するヌクレオチド配列について補正する。いくつかの実施形態では、方法は、正規化された試料カウントを得る前にカウントをフィルタリングするステップを含む。

40

【 0 0 2 8 】

いくつかの実施形態では、試料核酸は一本鎖核酸を含む。ある特定の実施形態では、試料核酸は二本鎖核酸を含む。いくつかの実施形態では、ヌクレオチド配列読み取りを得るステップは、試料核酸を、配列決定デバイスを使用した配列決定プロセスに供することを

50

含む。ある特定の実施形態では、アウトカムをもたらすステップは、試料核酸中の胎児核酸の分率をファクタリングするステップを含む。いくつかの実施形態では、方法は、試料核酸中の胎児核酸の分率を決定するステップを含む。

【0029】

ある特定の実施形態では、正規化された試料カウントを、ヌクレオチド配列読み取りまたは試料核酸のグアニンおよびシトシンの含量について補正せずに得る。いくつかの実施形態では、正規化された試料カウントを1つの実験条件について得る。ある特定の実施形態では、実験条件はフローセルである。いくつかの実施形態では、正規化された試料カウントを2つの実験条件について得る。ある特定の実施形態では、実験条件はフローセルおよび試薬プレートである。いくつかの実施形態では、実験条件はフローセルおよび同定タグ指標である。いくつかの実施形態では、正規化された試料カウントを3つの実験条件について得る。ある特定の実施形態では、実験条件はフローセル、試薬プレートおよび同定タグ指標である。

10

【0030】

いくつかの実施形態では、正規化された試料カウントを、(i)グアニンおよびシトシンの含量に従って補正し、(i)の後に、(ii)実験条件に従って補正した後に得る。ある特定の実施形態では、正規化された試料カウントを、(i)の前に参照ゲノムセクションにおいて反復するヌクレオチド配列に従って補正した後に得る。いくつかの実施形態では、(ii)は、フローセルに従って補正することからなる。ある特定の実施形態では、(ii)は、同定タグ指標に従って補正し、次いでフローセルに従って補正することからなる。いくつかの実施形態では、(ii)は、試薬プレートに従って補正し、次いでフローセルに従って補正することからなる。ある特定の実施形態では、(ii)は、同定タグ指標および試薬プレートに従って補正し、次いでフローセルに従って補正することからなる。

20

【0031】

ある特定の実施形態では、正規化された試料カウントを、フローセルに従って補正することからなる実験条件に従った補正の後に得る。いくつかの実施形態では、正規化された試料カウントを、同定タグ指標に従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る。ある特定の実施形態では、正規化された試料カウントを、試薬プレートに従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る。いくつかの実施形態では、正規化された試料カウントを、同定タグ指標および試薬プレートに従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る。ある特定の実施形態では、正規化された試料カウントを、参照ゲノムセクションにおいて反復するヌクレオチド配列に従って補正し、その後に実験条件に従って補正した後に得る。

30

【0032】

ある特定の実施形態では、いくつかの方法は、第1のゲノミックセクションについて、試験被験体についてと、他の試料、参照または試料および参照についての正規化された試料カウント、または正規化された試料カウントの誘導値の間の差異の統計的有意性を評価するステップをさらに含む。いくつかの実施形態では、特定の方法は、1つまたは複数のゲノミックセクションについて、試験被験体についてと、他の試料、参照または試料および参照についての正規化された試料カウント、または正規化された試料カウントの誘導値の間の差異の統計的有意性を評価するステップをさらに含む。ある特定の実施形態では、いくつかの方法は、評価に基づいて試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップをさらに含む。いくつかの実施形態では、遺伝的変異は、微小欠失、重複、および異数性から選択される。

40

【0033】

いくつかの実施形態では、コンピュータ可読プログラムコードが組み込まれたコンピュータで使用可能な媒体を含むコンピュータプログラム製品であって、コンピュータ可読プログラムコードが、配列受信モジュール、論理処理モジュール、およびデータディスプレ

50

イ編成モジュールを含む別個のソフトウェアモジュールを含み、かつ、試料核酸における遺伝的変異の有無を同定するための方法であって、(a)配列受信モジュールによって、試料核酸からヌクレオチド配列読み取りを得るステップと、(b)論理処理モジュールによって、ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(c)論理処理モジュールによって、各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(d)論理処理モジュールによって、1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウ
10
ントを得るステップと、(e)論理処理モジュールによって、正規化された試料カウントから、試験被験体における遺伝的変異の有無を決定するアウトカムを生成するステップと、(f)データディスプレイ編成モジュールによって、論理処理モジュールによって決定されるのに応じて、試料核酸における遺伝的変異の有無を示すデータディスプレイを編成するステップとを含む方法の実行が遂行されるように適合されている、コンピュータプログラム製品も提供される。

【0034】

ある特定の実施形態では、本明細書に記載のコンピュータプログラム製品実施形態が記憶されたメモリを含む装置も提供される。いくつかの実施形態では、装置は、本明細書に記載のコンピュータプログラム製品実施形態の1つまたは複数の機能を実行するプロセッ
20
サを含む。ある特定の実施形態では、本明細書において特定されたコンピュータプログラム製品の1つまたは複数の機能をウェブに基づく環境で実行する。

【0035】

ある特定の実施形態では、本明細書において特定されたコンピュータプログラム製品が実行されるウェブに基づくシステムを含む装置も提供される。いくつかの実施形態では、ウェブに基づくシステムは、ウェブに基づく機能性のために十分なコンピュータ、ルーター、および通信機器を含む。ある特定の実施形態では、ウェブに基づくシステムは、ネットワーククラウドコンピューティング、ネットワーククラウドストレージまたはネットワーククラウドコンピューティングおよびネットワーククラウドストレージを含む。

【0036】

いくつかの実施形態では、核酸配列決定装置および処理装置を含むシステムであって、配列決定装置によって試料核酸からヌクレオチド配列読み取りが得られ、処理装置によって配列決定装置からのヌクレオチド配列読み取りが得られ、かつ(a)ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(b)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(c)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウ
30
ントを得るステップと、(d)正規化された試料カウントから、試料核酸における遺伝的
40
変異の有無を決定するアウトカムをもたらすステップとを含む方法が実行されるシステムも提供される。

【0037】

本明細書では、ヒト参照ゲノムhg19による第22染色体のヌクレオチド19,000,000位と22,000,000位の間の22q11.2微小欠失の有無を同定する方法であって、(a)試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、(b)試料から試料核酸を単離するステップと、(c)試料核酸からヌクレオチド配列読み取りを得るステップと、(d)ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、(e)各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、(f
50

) (e) のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、(g) 1 つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(f) の後に残ったカウントを第 1 のゲノムセクションについて正規化、またはカウントの誘導値を第 1 のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、(h) 第 2 2 染色体のヌクレオチド 1 9 , 0 0 0 , 0 0 0 位と 2 2 , 0 0 0 , 0 0 0 位の間に対応する 1 つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体の間の正規化されたカウントまたは正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、(i) (h) における評価から、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法も提供される。

10

【 0 0 3 8 】

ある特定の実施形態が以下の説明、実施例、特許請求の範囲および図においてさらに記載されている。

【 0 0 3 9 】

図は、当該技術の実施形態を例示するものであり、限定するものではない。例示を明白かつ容易にするために、図は定数縮尺で作成されておらず、いくつかの場合には、種々の態様は、特定の実施形態の理解を容易にするために誇張または拡大して示されていることがある。

20

例えば、本発明は、以下の項目を提供する：

(項目 1)

胎児の異数性の有無を検出するための方法であって、

(a) 妊婦由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、

(b) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d) 1 つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第 1 のゲノムセクションについて正規化、または上記カウントの誘導値を上記第 1 のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

30

(e) 上記正規化された試料カウントに基づいて、胎児の異数性の有無を決定するアウトカムをもたらすステップと

を含む、方法。

(項目 2)

胎児の異数性の有無を検出するための方法であって、

(a) 妊婦由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 上記試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

40

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) 1 つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第 1 のゲノムセクションについて正規化、または上記カウントの誘導値を上記第 1 のゲノムセクションについて正規化し、それにより、正規化された試料カウントを

50

得るステップと、

(g) 上記正規化された試料カウントに基づいて、胎児の異数性の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目3)

胎児の異数性の有無を検出するための方法であって、

(a) 妊婦由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 上記正規化された試料カウントに基づいて、胎児の異数性の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目4)

胎児の異数性の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、妊婦由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(c) 上記正規化された試料カウントに基づいて胎児の異数性の有無を検出するステップと

を含む、方法。

(項目5)

上記試料核酸が上記妊婦由来の血漿由来のものである、項目1から4のいずれか一項に記載の方法。

(項目6)

上記試料核酸が上記妊婦由来の血清由来のものである、項目1から4のいずれか一項に記載の方法。

(項目7)

上記胎児の異数性が13トリソミーである、項目1から4のいずれか一項に記載の方法。

(項目8)

上記胎児の異数性が18トリソミーである、項目1から4のいずれか一項に記載の方法。

(項目9)

上記胎児の異数性が21トリソミーである、項目1から3.1のいずれか一項に記載の方法。

(項目10)

上記無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、項目1から4のいずれか一項に記載の方法。

(項目11)

上記ポリヌクレオチド断片の長さが約20ヌクレオチドから約50ヌクレオチドの間で

10

20

30

40

50

ある、項目 10 に記載の方法。

(項目 12)

上記ポリヌクレオチドの長さが約 30 ヌクレオチドから約 40 ヌクレオチドの間である、項目 11 に記載の方法。

(項目 13)

上記予測カウントがカウント中央値である、項目 1 から 12 のいずれか一項に記載の方法。

(項目 14)

上記予測カウントが、トリムもしくは刈り込み平均 (trimmed or truncated mean)、ウィンザー化平均またはブートストラップ推定値である、項目 1 から 12 のいずれか一項に記載の方法。

10

(項目 15)

上記カウントを、GC 含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、項目 1 から 14 のいずれか一項に記載の方法。

(項目 16)

上記カウントを、正規化モジュールによって正規化する、項目 1 から 15 のいずれか一項に記載の方法。

(項目 17)

上記核酸配列読み取りを、配列決定モジュールによって生成する、項目 1 から 16 のいずれか一項に記載の方法。

20

(項目 18)

上記核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム全体にマッピングするステップを含む、項目 1 から 17 のいずれか一項に記載の方法。

(項目 19)

上記核酸配列読み取りをマッピングモジュールによってマッピングする、項目 18 に記載の方法。

(項目 20)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りをカウントモジュールによってカウントする、項目 1 から 19 のいずれか一項に記載の方法。

30

(項目 21)

上記配列読み取りを上記配列決定モジュールから上記マッピングモジュールに移行する、項目 19 または 20 に記載の方法。

(項目 22)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りを上記マッピングモジュールから上記カウントモジュールに移行する、項目 20 または 21 に記載の方法。

(項目 23)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りのカウントを、上記カウントモジュールから上記正規化モジュールに移行する、項目 20 から 22 のいずれか一項に記載の方法。

40

(項目 24)

上記カウントを正規化する上記ステップが、パーセント表示を決定するステップを含む、項目 1 から 23 のいずれか一項に記載の方法。

(項目 25)

上記正規化されたカウントが z スコアである、項目 1 から 24 のいずれか一項に記載の方法。

(項目 26)

上記正規化されたカウントがロバストな z スコアである、項目 1 から 25 のいずれか一項に記載の方法。

50

(項目27)

上記第1のゲノミックセクションについての上記カウントの上記誘導値が上記第1のゲノミックセクションのパーセント表示である、項目1から26のいずれか一項に記載の方法。

(項目28)

上記中央値がパーセント表示の中央値である、項目13から27のいずれか一項に記載の方法。

(項目29)

上記パーセント表示が染色体表示である、項目24から28のいずれか一項に記載の方法。

10

(項目30)

遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、

(b) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

20

(e) 上記正規化された試料カウントに基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目31)

遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 上記試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

30

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

40

(g) 上記正規化された試料カウントに基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目32)

遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参

50

照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 上記正規化された試料カウントに基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目33)

遺伝的変異の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、試験被験体由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 上記正規化された試料カウントに基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目34)

上記試料核酸が上記試験被験体由来の血漿由来のものである、項目30から33のいずれか一項に記載の方法。

(項目35)

上記試料核酸が上記試験被験体由来の血清由来のものである、項目30から33のいずれか一項に記載の方法。

(項目36)

上記遺伝的変異が医学的状態に関連付けられる、項目30から35のいずれか一項に記載の方法。

(項目37)

上記医学的状態ががんである、項目36に記載の方法。

(項目38)

上記医学的状態が異数性である、項目36に記載の方法。

(項目39)

上記試験被験体がヒト、動物、および植物から選択される、項目30から35のいずれか一項に記載の方法。

(項目40)

ヒト試験被験体が、女性、妊婦、男性、胎児、または新生児を含む、項目39に記載の方法。

(項目41)

上記無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、項目30から35のいずれか一項に記載の方法。

(項目42)

上記ポリヌクレオチド断片の長さが約20ヌクレオチドから約50ヌクレオチドの間である、項目41に記載の方法。

(項目43)

上記ポリヌクレオチドの長さが約30ヌクレオチドから約40ヌクレオチドの間である、項目42に記載の方法。

(項目44)

上記予測カウントがカウント中央値である、項目30から43のいずれか一項に記載の

10

20

30

40

50

方法。

(項目 4 5)

上記予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、項目 3 0 から 4 3 のいずれか一項に記載の方法。

(項目 4 6)

上記カウントを、GC 含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、項目 3 0 から 4 5 のいずれか一項に記載の方法。

(項目 4 7)

上記カウントを、正規化モジュールによって正規化する、項目 3 0 から 4 6 のいずれか一項に記載の方法。

10

(項目 4 8)

上記核酸配列読み取りを、配列決定モジュールによって生成する、項目 3 0 から 4 7 のいずれか一項に記載の方法。

(項目 4 9)

上記核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム全体にマッピングするステップを含む、項目 3 0 から 4 8 のいずれか一項に記載の方法。

(項目 5 0)

上記核酸配列読み取りをマッピングモジュールによってマッピングする、項目 4 9 に記載の方法。

20

(項目 5 1)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りをカウントモジュールによってカウントする、項目 3 0 から 5 0 のいずれか一項に記載の方法。

(項目 5 2)

上記配列読み取りを上記配列決定モジュールから上記マッピングモジュールに移行する、項目 5 0 または 5 1 に記載の方法。

(項目 5 3)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りを上記マッピングモジュールから上記カウントモジュールに移行する、項目 5 1 または 5 2 に記載の方法。

30

(項目 5 4)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りのカウントを、上記カウントモジュールから上記正規化モジュールに移行する、項目 5 1 から 5 3 のいずれか一項に記載の方法。

(項目 5 5)

上記カウントを正規化する上記ステップが、パーセント表示を決定するステップを含む、項目 3 0 から 5 4 のいずれか一項に記載の方法。

(項目 5 6)

上記正規化されたカウントが z スコアである、項目 3 0 から 5 5 のいずれか一項に記載の方法。

40

(項目 5 7)

上記正規化されたカウントがロバストな z スコアである、項目 3 0 から 5 6 のいずれか一項に記載の方法。

(項目 5 8)

上記第 1 のゲノミックセクションについての上記カウントの上記誘導値が上記第 1 のゲノミックセクションのパーセント表示である、項目 3 0 から 5 7 のいずれか一項に記載の方法。

(項目 5 9)

上記中央値がパーセント表示の中央値である、項目 4 4 から 5 8 のいずれか一項に記載

50

の方法。

(項目60)

上記パーセント表示が染色体表示である、項目55から59のいずれか一項に記載の方法。

(項目61)

遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、

(b) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d) (c)の上記カウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(e) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(d)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(f) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(g) (f)における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目62)

遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 上記試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) (e)の上記カウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(g) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(f)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(h) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(i) (h)における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目63)

10

20

30

40

50

遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) (b) の上記カウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(c) で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(f) (e) における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目 6 4)

遺伝的変異の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、試験被験体由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(b) (a) の上記カウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(b) で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と上記参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(e) (d) における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目 6 5)

上記補正され、カウントされ、マッピングされた配列読み取りを、1つまたは複数の実験条件についてさらに補正した後に、上記残ったカウントを正規化する、項目 6 1 から 6 4 のいずれか一項に記載の方法。

(項目 6 6)

上記遺伝的変異が微小欠失である、項目 6 1 から 6 5 のいずれか一項に記載の方法。

(項目 6 7)

上記微小欠失が第 2 2 染色体上にある、項目 6 6 に記載の方法。

(項目 6 8)

上記微小欠失が第 2 2 染色体の領域 2 2 q 1 1 . 2 において起こっている、項目 6 7 に記載の方法。

(項目 6 9)

10

20

30

40

50

上記微小欠失が、参照ゲノム hg 19 による第 22 染色体のヌクレオチド 19,000,000 位と 22,000,000 位の間で起こっている、項目 67 に記載の方法。

(項目 70)

上記正規化されたカウントの誘導値が Z スコアである、項目 61 から 69 のいずれか一項に記載の方法。

(項目 71)

上記 Z スコアがロバストな Z スコアである、項目 70 に記載の方法。

(項目 72)

上記試料核酸が上記試験被験体由来の血漿由来のものである、項目 61 から 71 のいずれか一項に記載の方法。

10

(項目 73)

上記試料核酸が上記試験被験体由来の血清由来のものである、項目 61 から 71 のいずれか一項に記載の方法。

(項目 74)

上記遺伝的変異が医学的状態に関連付けられる、項目 61 から 73 のいずれか一項に記載の方法。

(項目 75)

上記医学的状態ががんである、項目 74 に記載の方法。

(項目 76)

上記医学的状態が異数性である、項目 74 に記載の方法。

20

(項目 77)

上記試験被験体がヒト、動物、および植物から選択される、項目 61 から 73 のいずれか一項に記載の方法。

(項目 78)

ヒト試験被験体が、女性、妊婦、男性、胎児、または新生児を含む、項目 77 に記載の方法。

(項目 79)

上記無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、項目 61 から 73 のいずれか一項に記載の方法。

(項目 80)

30

上記ポリヌクレオチド断片の長さが約 20 ヌクレオチドから約 50 ヌクレオチドの間である、項目 79 に記載の方法。

(項目 81)

上記ポリヌクレオチドの長さが約 30 ヌクレオチドから約 40 ヌクレオチドの間である、項目 80 に記載の方法。

(項目 82)

上記予測カウントがカウント中央値である、項目 61 から 81 のいずれか一項に記載の方法。

(項目 83)

上記予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、項目 61 から 81 のいずれか一項に記載の方法。

40

(項目 84)

上記カウントを、GC 含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、項目 61 から 83 のいずれか一項に記載の方法。

(項目 85)

上記カウントを、正規化モジュールによって正規化する、項目 61 から 84 のいずれか一項に記載の方法。

(項目 86)

上記核酸配列読み取りを、配列決定モジュールによって生成する、項目 61 から 85 の

50

いずれか一項に記載の方法。

(項目 87)

上記核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム全体にマッピングするステップを含む、項目 61 から 86 のいずれか一項に記載の方法。

(項目 88)

上記核酸配列読み取りをマッピングモジュールによってマッピングする、項目 87 に記載の方法。

(項目 89)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りをカウントモジュールによってカウントする、項目 61 から 88 のいずれか一項に記載の方法

10

(項目 90)

上記配列読み取りを上記配列決定モジュールから上記マッピングモジュールに移行する、項目 88 または 89 に記載の方法。

(項目 91)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りを上記マッピングモジュールから上記カウントモジュールに移行する、項目 89 または 90 に記載の方法。

(項目 92)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りのカウントを、上記カウントモジュールから上記正規化モジュールに移行する、項目 89 から 91 のいずれか一項に記載の方法。

20

(項目 93)

上記カウントを正規化する上記ステップが、パーセント表示を決定するステップを含む、項目 61 から 92 のいずれか一項に記載の方法。

(項目 94)

上記正規化されたカウントが z スコアである、項目 61 から 93 のいずれか一項に記載の方法。

(項目 95)

上記正規化されたカウントがロバストな z スコアである、項目 61 から 94 のいずれか一項に記載の方法。

30

(項目 96)

上記第 1 のゲノミックセクションについての上記カウントの上記誘導値が上記第 1 のゲノミックセクションのパーセント表示である、項目 61 から 95 のいずれか一項に記載の方法。

(項目 97)

上記中央値がパーセント表示の中央値である、項目 82 から 96 のいずれか一項に記載の方法。

(項目 98)

上記パーセント表示が染色体表示である、項目 93 から 97 のいずれか一項に記載の方法。

40

(項目 99)

微小欠失の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、

(b) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d) (c) の上記カウントされ、マッピングされた配列読み取りを、反復配列および

50

ノまたは過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(e) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(d)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(f) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(g) (f)における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目100)

微小欠失の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 上記試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) (e)の上記カウントされ、マッピングされた配列読み取りを、反復配列およびノまたは過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(g) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(f)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(h) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(i) (h)における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目101)

微小欠失の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) (b)の上記カウントされ、マッピングされた配列読み取りを、反復配列およびノまたは過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(c)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

10

20

30

40

50

(e) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(f) (e)における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目 1 0 2)

微小欠失の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、試験被験体由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(b) (a) の上記カウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(b) で残ったカウントを第 1 のゲノムセクションについて正規化、または上記カウントの誘導値を上記第 1 のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(e) (d)における評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

(項目 1 0 3)

上記補正され、カウントされ、マッピングされた配列読み取りを、1つまたは複数の実験条件についてさらに補正した後に、上記残ったカウントを正規化する、項目 9 9 から 1 0 2 のいずれか一項に記載の方法。

(項目 1 0 4)

上記微小欠失が第 2 2 染色体上にある、項目 1 0 3 に記載の方法。

(項目 1 0 5)

上記微小欠失が第 2 2 染色体の領域 2 2 q 1 1 . 2 において起こっている、項目 1 0 4 に記載の方法。

(項目 1 0 6)

上記微小欠失が、参照ゲノム hg 1 9 による第 2 2 染色体のヌクレオチド 1 9 , 0 0 0 , 0 0 0 位と 2 2 , 0 0 0 , 0 0 0 位の間で起こっている、項目 1 0 4 に記載の方法。

(項目 1 0 7)

上記正規化されたカウントの誘導値が Z スコアである、項目 9 9 から 1 0 6 のいずれか一項に記載の方法。

(項目 1 0 8)

上記 Z スコアがロバストな Z スコアである、項目 1 0 7 に記載の方法。

(項目 1 0 9)

上記試料核酸が上記試験被験体由来の血漿由来のものである、項目 9 9 から 1 0 8 のいずれか一項に記載の方法。

(項目 1 1 0)

上記試料核酸が上記試験被験体由来の血清由来のものである、項目 9 9 から 1 0 8 のいずれか一項に記載の方法。

(項目 1 1 1)

上記遺伝的変異が医学的状態に関連付けられる、項目 9 9 から 1 1 0 のいずれか一項に

10

20

30

40

50

記載の方法。

(項目 1 1 2)

上記医学的状態ががんである、項目 1 1 1 に記載の方法。

(項目 1 1 3)

上記医学的状態が異数性である、項目 1 1 1 に記載の方法。

(項目 1 1 4)

上記試験被験体がヒト、動物、および植物から選択される、項目 9 9 から 1 1 0 のいずれか一項に記載の方法。

(項目 1 1 5)

ヒト試験被験体が、女性、妊婦、男性、胎児、または新生児を含む、項目 1 1 4 に記載の方法。

10

(項目 1 1 6)

上記無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、項目 9 9 から 1 1 0 のいずれか一項に記載の方法。

(項目 1 1 7)

上記ポリヌクレオチド断片の長さが約 2 0 ヌクレオチドから約 5 0 ヌクレオチドの間である、項目 1 1 6 に記載の方法。

(項目 1 1 8)

上記ポリヌクレオチドの長さが約 3 0 ヌクレオチドから約 4 0 ヌクレオチドの間である、項目 1 1 7 に記載の方法。

20

(項目 1 1 9)

上記予測カウントがカウント中央値である、項目 9 9 から 1 1 8 のいずれか一項に記載の方法。

(項目 1 2 0)

上記予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、項目 9 9 から 1 1 8 のいずれか一項に記載の方法。

(項目 1 2 1)

上記カウントを、GC 含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、項目 9 9 から 1 2 0 のいずれか一項に記載の方法。

30

(項目 1 2 2)

上記カウントを、正規化モジュールによって正規化する、項目 9 9 から 1 2 1 のいずれか一項に記載の方法。

(項目 1 2 3)

上記核酸配列読み取りを、配列決定モジュールによって生成する、項目 9 9 から 1 2 2 のいずれか一項に記載の方法。

(項目 1 2 4)

上記核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム全体にマッピングするステップを含む、項目 9 9 から 1 2 3 のいずれか一項に記載の方法。

(項目 1 2 5)

上記核酸配列読み取りをマッピングモジュールによってマッピングする、項目 1 2 4 に記載の方法。

40

(項目 1 2 6)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りをカウントモジュールによってカウントする、項目 9 9 から 1 2 5 のいずれか一項に記載の方法。

(項目 1 2 7)

上記配列読み取りを上記配列決定モジュールから上記マッピングモジュールに移行する、項目 1 2 5 または 1 2 6 に記載の方法。

(項目 1 2 8)

50

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りを上記マッピングモジュールから上記カウントモジュールに移行する、項目 1 2 6 または 1 2 7 に記載の方法。

(項目 1 2 9)

上記参照ゲノムのゲノミックセクションにマッピングされた上記核酸配列読み取りのカウントを、上記カウントモジュールから上記正規化モジュールに移行する、項目 1 2 6 から 1 2 8 のいずれか一項に記載の方法。

(項目 1 3 0)

上記カウントを正規化する上記ステップが、パーセント表示を決定するステップを含む、項目 9 9 から 1 2 9 のいずれか一項に記載の方法。

10

(項目 1 3 1)

上記正規化されたカウントが z スコアである、項目 9 9 から 1 3 0 のいずれか一項に記載の方法。

(項目 1 3 2)

上記正規化されたカウントがロバストな z スコアである、項目 9 9 から 1 3 1 のいずれか一項に記載の方法。

(項目 1 3 3)

上記第 1 のゲノミックセクションについての上記カウントの上記誘導値が上記第 1 のゲノミックセクションのパーセント表示である、項目 9 9 から 1 3 2 のいずれか一項に記載の方法。

20

(項目 1 3 4)

上記中央値がパーセント表示の中央値である、項目 1 1 9 から 1 3 3 のいずれか一項に記載の方法。

(項目 1 3 5)

上記パーセント表示が染色体表示である、項目 1 3 0 から 1 3 4 のいずれか一項に記載の方法。

(項目 1 3 6)

上記正規化された試料カウントが、上記カウントの上記誘導値を上記第 1 のゲノムセクションについて正規化することを含むプロセスによって得られ、上記誘導値が上記第 1 のゲノムセクションについてのカウントを上記第 1 のゲノムセクションを含む複数のゲノムセクションについてのカウントで割ることによって決定される第 1 のゲノムセクションカウント表示である、項目 1 から 1 2 0 のいずれか一項に記載の方法。

30

(項目 1 3 7)

上記第 1 のゲノムセクションについての上記カウントの上記誘導値を上記予測カウントの誘導値に従って正規化し、上記予測カウントの上記誘導値が、上記第 1 のゲノムセクションについての予測カウントを上記第 1 のゲノムセクションを含む複数のゲノムセクションについての予測カウントで割ることによって決定される予測された第 1 のゲノムセクションカウント表示である、項目 1 3 6 に記載の方法。

(項目 1 3 8)

上記第 1 のゲノムセクションが染色体または染色体の部分であり、上記複数のゲノムセクションが常染色体を含む、項目 1 から 1 3 7 のいずれか一項に記載の方法。

40

(項目 1 3 9)

上記染色体が第 2 1 染色体、第 1 8 染色体または第 1 3 染色体である、項目 1 3 8 に記載の方法。

(項目 1 4 0)

上記正規化された試料カウントが、上記第 1 のゲノムセクションについてのカウントから上記予測カウントを引き算し、それにより減算値を生成し、上記減算値を上記カウントの変動性の推定値で割ることを含むプロセスによって得られる、項目 1 から 1 2 0、項目 1 3 8 および項目 1 3 5 のいずれか一項に記載の方法。

(項目 1 4 1)

50

上記予測カウントの変動性の上記推定値が上記カウントの中央絶対偏差 (MAD) である、項目 140 に記載の方法。

(項目 142)

上記カウントの変動性の上記推定値が、Rousseeuw および Croux によって導入される MAD の代替値、またはブートストラップ推定値である、項目 140 に記載の方法。

(項目 143)

上記変動性の上記推定値が、1 つまたは複数の共通の実験条件から生成された試料データについて得られる、項目 140 から 142 のいずれか一項に記載の方法。

(項目 144)

上記変動性の上記推定値が、1 つまたは複数の共通の実験条件から生成されたものではない試料データについて得られる、項目 140 から 142 のいずれか一項に記載の方法。

(項目 145)

上記変動性の上記推定値および上記予測カウントが、1 つまたは複数の共通の実験条件から生成された試料データについて得られる、項目 140 から 144 のいずれか一項に記載の方法。

(項目 146)

上記正規化された試料カウントが、上記第 1 のゲノムセクションカウント表示から上記予測された第 1 のゲノムセクションカウント表示を引き算し、それにより減算値を生成し、上記減算値を上記第 1 のゲノムセクションカウント表示の変動性の推定値で割ることを含むプロセスによって得られる、項目 1 から 139 のいずれか一項に記載の方法。

(項目 147)

上記予測カウント表示の変動性の上記推定値が上記カウント表示の中央絶対偏差 (MAD) である、項目 146 に記載の方法。

(項目 148)

上記カウント表示の変動性の上記推定値が、Rousseeuw および Croux によって導入される MAD の代替値またはブートストラップ推定値である、項目 146 に記載の方法。

(項目 149)

上記予測カウント表示の変動性の上記推定値が、1 つまたは複数の共通の実験条件から生成された試料データについて得られる、項目 146 から 148 のいずれか一項に記載の方法。

(項目 150)

上記予測カウント表示の変動性の上記推定値が、1 つまたは複数の共通の実験条件から生成されたものではない試料データについて得られる、項目 146 から 148 のいずれか一項に記載の方法。

(項目 151)

上記予測カウント表示の変動性の上記推定値および上記予測された第 1 のゲノムセクションカウント表示が、1 つまたは複数の共通の実験条件から生成された試料データについて得られる、項目 146 から 150 のいずれか一項に記載の方法。

(項目 152)

上記 1 つまたは複数の共通の実験条件がフローセルを含む、項目 1 から 151 のいずれか一項に記載の方法。

(項目 153)

上記 1 つまたは複数の共通の実験条件がフローセル内のチャンネルを含む、項目 1 から 151 のいずれか一項に記載の方法。

(項目 154)

上記 1 つまたは複数の共通の実験条件が試薬プレートを含む、項目 1 から 151 のいずれか一項に記載の方法。

(項目 155)

10

20

30

40

50

上記試薬プレートを使用して配列決定のために核酸を段階分けする、項目 1 5 4 に記載の方法。

(項目 1 5 6)

上記試薬プレートを使用して配列決定のために核酸ライブラリーを調製する、項目 1 5 4 に記載の方法。

(項目 1 5 7)

上記 1 つまたは複数の共通の実験条件が同定タグ指標を含む、項目 1 から 1 5 1 のいずれか一項に記載の方法。

(項目 1 5 8)

上記正規化された試料カウントを、上記ヌクレオチド配列読み取りまたは上記試料核酸のグアニンおよびシトシンの含量について補正する、項目 1 から 1 5 7 のいずれか一項に記載の方法。

10

(項目 1 5 9)

上記カウントまたは上記正規化された試料カウントを局所重み付け多項式回帰に供するステップを含む、項目 1 5 8 に記載の方法。

(項目 1 6 0)

上記局所重み付け多項式回帰が L O E S S 回帰である、項目 1 5 9 に記載の方法。

(項目 1 6 1)

上記正規化された試料カウントを、上記参照ゲノムセクションにおいて反復するヌクレオチド配列について補正する、項目 1 から 1 5 9 のいずれか一項に記載の方法。

20

(項目 1 6 2)

上記カウントまたは上記正規化された試料カウントを、上記参照ゲノムセクションにおいて反復するヌクレオチド配列について補正する、項目 1 6 1 に記載の方法。

(項目 1 6 3)

上記正規化された試料カウントを得る前に上記カウントをフィルタリングするステップを含む、項目 1 から 1 6 2 のいずれか一項に記載の方法。

(項目 1 6 4)

上記試料核酸が一本鎖核酸を含む、項目 1 から 1 6 3 のいずれか一項に記載の方法。

(項目 1 6 5)

上記試料核酸が二本鎖核酸を含む、項目 1 から 1 6 3 のいずれか一項に記載の方法。

30

(項目 1 6 6)

上記ヌクレオチド配列読み取りを得るステップが、上記試料核酸を、配列決定デバイスを使用した配列決定プロセスに供することを含む、項目 1 から 1 6 5 のいずれか一項に記載の方法。

(項目 1 6 7)

アウトカムをもたらすステップが、上記試料核酸中の胎児核酸の分率をファクタリングすることを含む、項目 1 から 1 6 6 のいずれか一項に記載の方法。

(項目 1 6 8)

上記試料核酸中の胎児核酸の分率を決定するステップを含む、項目 1 から 1 6 7 のいずれか一項に記載の方法。

40

(項目 1 6 9)

上記正規化された試料カウントを、上記ヌクレオチド配列読み取りまたは上記試料核酸のグアニンおよびシトシンの含量について補正せずに得る、項目 1 から 1 6 8 のいずれか一項に記載の方法。

(項目 1 7 0)

上記正規化された試料カウントを 1 つの実験条件について得る、項目 1 から 1 6 8 のいずれか一項に記載の方法。

(項目 1 7 1)

上記実験条件がフローセルである、項目 1 7 0 に記載の方法。

(項目 1 7 2)

50

上記正規化された試料カウントを2つの実験条件について得る、項目1から168のいずれか一項に記載の方法。

(項目173)

上記実験条件がフローセルおよび試薬プレートである、項目172に記載の方法。

(項目174)

上記実験条件がフローセルおよび同定タグ指標である、項目172に記載の方法。

(項目175)

上記正規化された試料カウントを3つの実験条件について得る、項目1から168のいずれか一項に記載の方法。

(項目176)

上記実験条件がフローセル、試薬プレートおよび同定タグ指標である、項目175に記載の方法。

(項目177)

上記正規化された試料カウントを、(i)グアニンおよびシトシンの含量に従って補正し、(i)の後に、(ii)実験条件に従って補正した後に得る項目1から168のいずれか一項に記載の方法。

(項目178)

上記正規化された試料カウントを、(i)の前に上記参照ゲノムセクションにおいて反復するヌクレオチド配列に従って補正した後に得る、項目177に記載の方法。

(項目179)

(ii)が、フローセルに従って補正することからなる、項目177または175に記載の方法。

(項目180)

(ii)が、同定タグ指標に従って補正し、次いでフローセルに従って補正することからなる、項目177または175に記載の方法。

(項目181)

(ii)が、試薬プレートに従って補正し、次いでフローセルに従って補正することからなる、項目177または175に記載の方法。

(項目182)

(ii)が、同定タグ指標および試薬プレートに従って補正し、次いでフローセルに従って補正することからなる、項目177または175に記載の方法。

(項目183)

上記正規化された試料カウントを、フローセルに従って補正することからなる実験条件に従った補正の後に得る、項目169に記載の方法。

(項目184)

上記正規化された試料カウントを、同定タグ指標に従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る、項目169に記載の方法。

(項目185)

上記正規化された試料カウントを、試薬プレートに従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る、項目169に記載の方法。

(項目186)

上記正規化された試料カウントを、同定タグ指標および試薬プレートに従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る、項目169に記載の方法。

(項目187)

上記正規化された試料カウントを、上記参照ゲノムセクションにおいて反復するヌクレオチド配列に従って補正し、その後上記実験条件に従って補正した後に得る、項目180から186のいずれか一項に記載の方法。

10

20

30

40

50

(項目188)

上記正規化された試料カウントがZスコアである、項目136から186のいずれか一項に記載の方法。

(項目189)

(i)が、

(a)各試料について(i)上記参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウントと(ii)上記部分のそれぞれについてのGC含量との間のフィッティングした関係から、複数の試料について上記参照ゲノムの部分のそれぞれについてのゲアニンおよびシトシン(GC)の偏りを決定することと、

(b)(i)上記GCの偏りと(ii)上記参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウントとの間のフィッティングした関係から上記参照ゲノムの部分のそれぞれについてのゲノミックセクションの高度を算出し、それにより、算出されたゲノミックセクションの高度をもたらし、それにより、上記算出されたゲノミックセクションの高度において、上記参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウントの偏りが減少することと

を含む、項目177から188のいずれか一項に記載の方法。

(項目190)

上記参照ゲノムの部分が染色体内にある、項目189に記載の方法。

(項目191)

上記参照ゲノムの部分が染色体の部分内にある、項目189に記載の方法。

(項目192)

上記染色体が第21染色体である、項目189から191のいずれか一項に記載の方法

(項目193)

上記染色体が第18染色体である、項目189から191のいずれか一項に記載の方法

(項目194)

上記染色体が第13染色体である、項目189から191のいずれか一項に記載の方法

(項目195)

(b)の前に、上記参照ゲノムの部分のいくつかまたは全部にマッピングされた配列読み取りのカウントについて誤差の尺度を算出し、上記参照ゲノムの特定の部分についての配列読み取りのカウントを上記誤差の尺度の閾値に従って除去または重み付けすることを含む、項目189から46のいずれか一項に記載の方法。

(項目196)

上記閾値を、第1のゲノミックセクションの高度と第2のゲノミックセクションの高度との間の標準偏差ギャップ3.5以上に応じて選択する、項目195に記載の方法。

(項目197)

上記誤差の尺度がR因子である、項目195または196に記載の方法。

(項目198)

R因子が約7%~約10%である上記参照ゲノムの部分についての配列読み取りのカウントを(b)の前に除去する、項目197に記載の方法。

(項目199)

(b)の上記フィッティングした関係がフィッティングした線形関係である、項目189から198のいずれか一項に記載の方法。

(項目200)

上記関係の傾きを線形回帰によって決定する、項目199に記載の方法。

(項目201)

各GCの偏りがGCの偏り係数であり、上記GCの偏り係数が(i)上記参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウントと(ii)上記部分のそれぞ

10

20

30

40

50

れについてのGC含量との間の線形関係の傾きである、項目199または200に記載の方法。

(項目202)

(b)の上記フィッティングした関係がフィッティングした非線形関係である、項目189から198のいずれか一項に記載の方法。

(項目203)

各GCの偏りが、GC曲率推定値を含む、項目202に記載の方法。

(項目204)

(c)の上記フィッティングした関係が線形である、項目189から203のいずれか一項に記載の方法。

(項目205)

上記関係の傾きを線形回帰によって決定する、項目204に記載の方法。

(項目206)

(b)の上記フィッティングした関係が線形であり、(c)の上記フィッティングした関係が線形であり、ゲノミックセクションの高度 L_i が、上記参照ゲノムの部分のそれぞれについて、方程式：

$$L_i = (m_i - G_i S) I^{-1} \quad \text{方程式}$$

に従って決定され、式中、 G_i はGCの偏りであり、 I は(c)の上記フィッティングした関係の切片であり、 S は(c)の上記関係の傾きであり、 m_i は測定された、上記参照ゲノムの各部分にマッピングされたカウントであり、そして i は試料である、項目189

(項目207)

上記参照ゲノムの部分の数が約40,000以上の部分である、項目189から206のいずれか一項に記載の方法。

(項目208)

上記参照ゲノムの各部分が所定の長さのヌクレオチド配列を含む、項目189から207のいずれか一項に記載の方法。

(項目209)

上記所定の長さが約50キロベースである、項目208に記載の方法。

(項目210)

(b)の上記GCの偏りをGCの偏りモジュールによって決定する、項目189から209のいずれか一項に記載の方法。

(項目211)

コンピュータ可読プログラムコードが組み込まれたコンピュータで使用可能な媒体を含むコンピュータプログラム製品であって、上記コンピュータ可読プログラムコードが、配列受信モジュール、論理処理モジュール、およびデータディスプレイ編成モジュールを含む別個のソフトウェアモジュールを含み、かつ、試料核酸における遺伝的変異の有無を同定するための方法であって、

(a)上記配列受信モジュールによって、試料核酸からヌクレオチド配列読み取りを得るステップと、

(b)上記論理処理モジュールによって、上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c)上記論理処理モジュールによって、各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d)上記論理処理モジュールによって、1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化し、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e)上記論理処理モジュールによって、上記正規化された試料カウントに基づいて、

10

20

30

40

50

上記試験被験体における遺伝的変異の有無を決定するアウトカムを生成するステップと、
 (f) 上記データディスプレイ編成モジュールによって、上記論理処理モジュールによって決定されるのに応じて上記試料核酸における上記遺伝的変異の有無を示すデータディスプレイを編成するステップと

を含む方法の実行が遂行されるように適合されている、コンピュータプログラム製品。

(項目 2 1 2)

項目 F 1 のコンピュータプログラム製品が記憶されているメモリを含む装置。

(項目 2 1 3)

項目 F 1 に記載のコンピュータプログラム製品の 1 つまたは複数の機能を実行するプロセッサを含む、項目 F 2 に記載の装置。

10

(項目 2 1 4)

核酸配列決定装置および処理装置を含むシステムであって、上記配列決定装置によって試料核酸からヌクレオチド配列読み取りが得られ、上記処理装置によって上記配列決定装置からのヌクレオチド配列読み取りが得られ、かつ

(a) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) 1 つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第 1 のゲノムセクションについて正規化、または上記カウントの誘導値を上記第 1 のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

20

(d) 上記正規化された試料カウントに基づいて上記試料核酸における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む方法が実施される、システム。

(項目 2 1 5)

ヒト参照ゲノム hg 1 9 による第 2 2 染色体のヌクレオチド 1 9 , 0 0 0 , 0 0 0 位と 2 2 , 0 0 0 , 0 0 0 位の間の 2 2 q 1 1 . 2 微小欠失の有無を同定する方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

30

(b) 上記試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) 上記ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) (e) の上記カウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(g) 1 つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って (f) で残ったカウントを第 1 のゲノムセクションについて正規化、または上記カウントの誘導値を上記第 1 のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

40

(h) 第 2 2 染色体のヌクレオチド 1 9 , 0 0 0 , 0 0 0 位と 2 2 , 0 0 0 , 0 0 0 位の間に対応する 1 つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価するステップと、

(i) (h) における上記評価に基づいて、上記試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

50

を含む、方法。

(項目 2 1 6)

上記補正され、カウントされ、マッピングされた配列読み取りを、1つまたは複数の実験条件についてさらに補正した後に、上記残ったカウントを正規化する、項目 F 1 から F 3 のいずれか一項に記載の方法。

(項目 2 1 7)

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、

上記メモリは、上記1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、上記配列読み取りが、試験試料由来の循環している無細胞核酸の読み取りであり、上記1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、そして

(b) 上記正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されている、システム。

(項目 2 1 8)

1つまたは複数のプロセッサおよびメモリを含む装置であって、

上記メモリは、上記1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、上記配列読み取りが、試験試料由来の循環している無細胞核酸の読み取りであり、上記1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、そして

(b) 上記正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されている、装置。

(項目 2 1 9)

コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、

(a) 参照ゲノムのゲノミックセクションにマッピングされた、試験試料由来の循環している無細胞核酸の読み取りである配列読み取りのカウントにアクセスし、

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、そして

(c) 上記正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されている命令を含む、コンピュータプログラム製品。

(項目 2 2 0)

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、

上記メモリは、上記1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、上記配列読み取りが胎児を有する妊婦由来の循環している無細胞核酸の読み取りであり、上記1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参

10

20

30

40

50

照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、そして

(b) 上記正規化された試料カウントに基づいて、上記試験被験体における遺伝的変異の有無を決定する

ように構成されている、システム。

(項目221)

1つまたは複数のプロセッサおよびメモリを含む装置であって、

上記メモリは、上記1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、上記配列読み取りが、胎児を有する妊婦由来の循環している無細胞核酸の読み取りであり、上記1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、そして

(b) 上記正規化された試料カウントに基づいて、上記試験被験体における遺伝的変異の有無を決定する

ように構成されている、装置。

(項目222)

コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、

(a) 参照ゲノムのゲノミックセクションにマッピングされた、胎児を有する妊婦由来の循環している無細胞核酸の読み取りである配列読み取りのカウントにアクセスし、

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、上記カウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、そして

(c) 上記正規化された試料カウントに基づいて、上記試験被験体における遺伝的変異の有無を決定する

ように構成されている命令を含む、コンピュータプログラム製品。

(項目223)

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、

上記メモリは、上記1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、上記配列読み取りが、胎児を有する妊婦由来の循環している無細胞核酸の読み取りであり、上記1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(a)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(c) 1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の、上記正規化されたカウントまたは上記正規化されたカウントの誘導値

10

20

30

40

50

の差異の統計的有意性を評価し、そして

(d)(c)における上記評価に基づいて、上記試験被験体における遺伝的変異の有無を決定する

ように構成されている、システム。

(項目224)

1つまたは複数のプロセッサおよびメモリを含む装置であって、

上記メモリが、上記1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムの部分にマッピングされた、胎児を有する妊婦由来の循環している無細胞核酸の読み取りである配列読み取りのカウントを含み、上記1つまたは複数のプロセッサによって遂行可能な命令が、

10

(a)反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、

(b)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(a)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(c)1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の、上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価し、そして

20

(d)(c)における上記評価に基づいて、上記試験被験体における遺伝的変異の有無を決定する

ように構成されている、装置。

(項目225)

コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、

(a)参照ゲノムの部分にマッピングされた、試験試料由来の循環している無細胞核酸の読み取りである配列読み取りのカウントにアクセスし、

(b)反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、

30

(c)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは上記予測カウントの誘導値に従って、(b)で残ったカウントを第1のゲノムセクションについて正規化、または上記カウントの誘導値を上記第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(d)1つまたは複数の選択されたゲノミックセクションについて、上記試験被験体と参照被験体の間の、上記正規化されたカウントまたは上記正規化されたカウントの誘導値の差異の統計的有意性を評価し、そして

40

(e)(d)における上記評価に基づいて、上記試験被験体における遺伝的変異の有無を決定する

ように構成されている命令を含む、コンピュータプログラム製品。

【図面の簡単な説明】

【0040】

【図1】図1は、妊娠期間に応じてプロットした、選択された試料のそれぞれについての胎児DNA分率を示すグラフである。

【図2】図2は、母体の年齢に応じてプロットした、選択された試料のそれぞれについての胎児DNA分率を示すグラフである。

【図3】図3は、母体の体重に応じてプロットした、選択された試料のそれぞれについて

50

の胎児DNA分率を示すグラフである。

【図4】図4は、フローセルによる第21染色体のマッチした読み取りに応じてプロットした、選択された試料のそれぞれについての第21染色体の百分率を示すグラフである。

【図5】図5は、プレート数による第21染色体のマッチした読み取りに応じてプロットした、選択された試料のそれぞれについての第21染色体の百分率を示すグラフである。

【図6】図6は、配列決定のために使用したIllumina計器に応じてプロットした、選択された試料のそれぞれについての第21染色体の百分率を示すグラフである。

【図7】図7は、妊娠期間に応じてプロットした、選択された試料のそれぞれについての第21染色体のzスコアを示すグラフである。

【図8】図8は、母体の年齢に応じてプロットした、選択された試料のそれぞれについての第21染色体のzスコアを示すグラフである。

10

【図9】図9は、母体の体重に応じてプロットした、選択された試料のそれぞれについての第21染色体のzスコアを示すグラフである。

【図10】図10は、ライブラリー濃度に応じてプロットした、選択された試料のそれぞれについての第21染色体のzスコアを示すグラフである。

【図11】図11は、ライブラリー調製最適化を示すグラフである。図11Aは、半自動化ライブラリー調製方法によって調製した標準化されたライブラリー濃度($n = 287$)と手動ライブラリー調製方法によって調製した標準化されたライブラリー濃度についての比較を示す。図11Bは、93の試料のそれぞれについてのGCRMに基づくzスコアを示す。確認された正倍数性試料($n = 83$)が薄い灰色で示されている。確認された21トリソミー試料($n = 10$)が濃い灰色で示されている。

20

【図12】図12は、zスコアの対応のある比較を示すグラフである。対の試料についてZスコアを算出し、x軸は上記のGCについて正規化し、反復マスキングしたzスコアであり、y軸は12プレックスで配列決定した同じライブラリーからのzスコアである。核型分析によってトリソミーと分類された試料が、図12A(第21染色体)、図12B(第13染色体)、または図12C(第18染色体)で濃い灰色で示されている。各異数性状態に関して影響を受けていない試料が薄い灰色で示されている。各プロット内の横線および縦線は、その染色体についてのそれぞれの分類カットオフを示す(第21染色体については $z = 3$ 、第13染色体および第18染色体については $z = 3.95$)。

【図13】図13は、Zスコア(x軸)対胎児分率(y軸)を示すグラフである。各異数性染色体についての染色体特異的zスコアが胎児DNAの割合(胎児分率)に対してプロットされている。核型分析によってトリソミーと分類された試料が、図13A(第21染色体)、図13B(第13染色体)、または図13C(第18染色体)で濃い灰色で示されている。各異数性状態に関して影響を受けていない試料が薄い灰色で示されている。各プロット内の横線は各染色体についてのそれぞれの分類カットオフを示す(第21染色体については $z = 3$ 、第13染色体および第18染色体については $z = 3.95$)。各パネル内の破線の縦線は影響を受けた試料のロバストな線形フィッティングを示す。各パネル内の破線の横線は影響を受けていない試料全てのロバストな線形フィッティングを示す。

30

【図14】図14は、zスコアの対応のある比較を示す。1269対の試料についてZスコアを算出し、x軸は上記のGCについて正規化し、反復マスキングしたzスコアであり、y軸はハイスループットなアッセイからのzスコアである。核型分析によってトリソミーと分類された試料が図14A(第21染色体)、図14B(第13染色体)、または図14C(第18染色体)で濃い灰色で示されている。各異数性状態に関して影響を受けていない試料が薄い灰色で示されている。各プロット内の横線および縦線は、その染色体についてのそれぞれの分類カットオフを示す(第21染色体については $z = 3$ 、第13染色体および第18染色体については $z = 3.95$)。

40

【発明を実施するための形態】

【0041】

遺伝的変異を同定するために有用な改善された方法、プロセスおよび装置が提供される。1つまたは複数の遺伝的変異または遺伝分散を同定することが、特定の医学的状態の診

50

断、またはそれに対する素因の決定につながる可能性がある。遺伝分散を同定することにより、医学的決定を容易にし、かつ/または有用な医学的手順を使用することができる。

【0042】

遺伝的変異および医学的状態

遺伝分散の有無は、本明細書に記載の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、本明細書に記載の方法および装置によってもたらされるアウトカムに応じて1つまたは複数の遺伝的変異の有無を決定する。遺伝的変異は、一般には、特定の個体に存在する特定の遺伝的表現型であり、多くの場合、遺伝的変異は、個体の統計的に有意な亜集団に存在する。遺伝的変異の非限定的な例としては、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（micro-duplication））、挿入、突然変異、多型（例えば、一塩基多型）、融合、反復（例えば、短いタンデム反復）、別個のメチル化部位、別個のメチル化パターンなど、およびそれらの組合せが挙げられる。挿入、反復、欠失、重複、突然変異または多型は、任意の観察される長さであってよく、いくつかの実施形態では、約1塩基または塩基対（bp）～1,000キロベース（kb）の長さ（例えば、約10bp、50bp、100bp、500bp、1kb、5kb、10kb、50kb、100kb、500kb、または1000kbの長さ）である。いくつかの実施形態では、遺伝的変異は、下でさらに詳細に記載されている染色体異常（例えば異数性）、部分的な染色体異常またはモザイク現象である。

10

【0043】

被験体について有無が同定される遺伝的変異は、ある特定の実施形態では医学的状態に関連付けられる。したがって、本明細書に記載の技術を使用して、医学的状態（medical condition）または医学的容態（medical state）に関連付けられる1つまたは複数の遺伝的変異の有無を同定することができる。医学的状態の非限定的な例としては、知的障害（例えば、ダウン症候群）、異常な細胞増殖（例えば、がん）、微生物核酸の存在（例えば、ウイルス、細菌、真菌、酵母）、および子癩前症に関連するものが挙げられる。

20

【0044】

遺伝的変異、医学的状態および容態（medical conditions and states）の非限定的な例は下に記載されている。

【0045】

胎児の性別

いくつかの実施形態では、胎児の性別の予測を、本明細書に記載の方法または装置によって決定することができる。性別の決定は、一般には、性染色体に基づく。ヒトでは、X染色体およびY染色体の2種の性染色体が存在する。XXの個体は女性であり、XYは男性であり、また、非限定的なバリエーションとしてXO、XYY、XXXおよびXXYが挙げられる。

30

【0046】

染色体異常

いくつかの実施形態では、胎児の染色体異常の有無は、本明細書に記載の方法または装置を使用することによって決定することができる。染色体異常としては、限定することなく、染色体全体または1つまたは複数の遺伝子を含む染色体の領域の増減が挙げられる。染色体異常としては、モノソミー、トリソミー、ポリソミー、異型接合性の欠如、不均衡転座によって引き起こされる欠失および重複を含めた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/または重複が挙げられる。「異数性（aneuploidy）」および「異数性の（aneuploid）」という用語は、本明細書で使用される場合、生物体の細胞内の染色体の数が異常であることを指す。異なる生物体が有する染色体組は広範に変動するので、「異数性」という用語は、染色体の特定の数を指すのではなく、生物体内の所与の1つまたは複数の細胞の染色体含量の状況が異常である状況を指す。

40

【0047】

50

モノソミーとは、一般には、正常な組の一方の染色体が欠如していることである。染色体の部分のみが単一のコピーで存在する部分モノソミーが不均衡転座または欠失で起こり得る。例えば、性染色体(45, X)のモノソミーにより、ターナー症候群が引き起こされる。

【0048】

ダイソミーとは、一般に、染色体の2つのコピーが存在することである。各染色体の2つのコピーを有する(二倍体である、または「正倍数性」である)ヒトなどの生物体については、ダイソミーとは正常な状態である。通常各染色体のコピーを3つ以上有する(三倍体以上である)生物体については、ダイソミーは、異数性染色体の状態である。片親性ダイソミーでは、染色体の両方のコピーが同じ親に由来する(他方の親からの寄与がない)。

10

【0049】

トリソミーとは、一般に、特定の染色体の2つのコピーではなく3つのコピーが存在することである。ヒトダウン症候群に見いだされる、第21染色体が余分に存在することは、「21トリソミー」と称される。18トリソミーおよび13トリソミーは、2種の他のヒト常染色体トリソミーである。性染色体のトリソミーは、女性(例えば、47, XXX)または男性(例えば、クラインフェルター症候群における47, XXY; または47, XYY)において見ることができる。

【0050】

テトラソミーおよびペンタソミーとは、一般に、染色体のコピーがそれぞれ4つまたは5つ存在することである。常染色体ではめったに見られないが、XXXX、XXXXY、XXYY、XYYY、XXXXX、XXXXXY、XXXYY、XXYYYおよびXYYYを含まない性染色体のテトラソミーおよびペンタソミーがヒトにおいて報告されている。

20

【0051】

染色体異常は、種々の機構によって引き起こされ得る。機構としては、これだけに限定されないが、(i)有糸分裂チェックポイントが弱まった結果として起こる不分離、(ii)多数の染色体において不分離を引き起こす不活性化有糸分裂チェックポイント、(iii)1つの動原体が有糸分裂紡錘体極の両方に付着した場合に起こるメロテリック付着、(iv)3つ以上の紡錘体極が形成された場合の多極紡錘の形成、(v)単一の紡錘体極のみが形成された場合の単極紡錘の形成、および(vi)単極紡錘機構の最終結果として生じる四倍体の中間体が挙げられる。

30

【0052】

部分モノソミーまたは部分トリソミーとは、一般には染色体の部分が減少または増加することによって引き起こされる遺伝物質の不均衡である。部分モノソミーまたは部分トリソミーは、個体が2つの異なる染色体の破壊および融合によって形成される誘導染色体を有する不均衡転座によって生じ得る。この状況では、個体は1つの染色体の部分の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部分)ならびに、誘導染色体に關与する他の染色体の部分のただ1つのコピーを有する。

【0053】

モザイク現象とは、一般には、生物体の全ての細胞ではないがいくつかの細胞における異数性である。ある特定の染色体異常がモザイク型染色体異常および非モザイク型染色体異常として存在し得る。例えば、ある特定の21トリソミーの個体はモザイク型ダウン症候群を有し、いくつかは非モザイク型ダウン症候群を有する。異なる機構によりモザイク現象が導かれ得る。例えば、(i)最初の接合体が3つの第21染色体を有する可能性があり、その結果、通常は単純21トリソミーになるが、細胞分裂の過程中、1つまたは複数の細胞株が第21染色体のうちの1つを失う; および(ii)最初の接合体が2つの第21染色体を有する可能性があるが、細胞分裂の過程中、第21染色体のうちの1つが重複する。体細胞モザイク現象は、一般には完全な異数性またはモザイク型異数性を伴う遺伝的症候群に關連付けられるものとは別個の機構によって起こる可能性がある。体細胞モザイク現象は、例えば、ある特定の種類のがんにおいて、およびニューロンにおいて同定

40

50

されている。ある特定の例では、12トリソミーは慢性リンパ球性白血病（CLL）において同定されており、8トリソミーは急性骨髄性白血病（AML）において同定されている。また、個体が染色体の破壊の素因を有する遺伝的症候群（染色体不安定性症候群）には、しばしば、種々の種類のがんのリスクの増加が伴い、したがって、がん発生における体細胞異数性の役割が強調される。本明細書に記載の方法およびプロトコールにより、非モザイク型染色体異常およびモザイク型染色体異常の有無を同定することができる。

【0054】

表1Aおよび表1Bには、本明細書に記載の方法および装置によって潜在的に同定することができる染色体の状態、症候群および/または異常の非限定的な一覧が示されている。表1Bは、2011年10月6日現在のDECIPHERデータベース（例えば、バージョン5.1、GRCh37にマッピングされた位置に基づく；統一資源位置指定子（URL）decipher.sanger.ac.ukにおいて利用可能）からのものである。

【0055】

【表 1 - 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	Xトリソミー症候群
Y	XXXX	XXXX症候群
Y	Xp21欠失	デュシェンヌ/ベッカー症候群、先天性副腎低形成、慢性肉芽腫性疾患(granulomatus disease)
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(X-linked lymphoproliferative disease)
1	1p(体細胞)モノソミー、トリソミー	神経芽細胞腫
2	モノソミー、2qトリソミー	成長遅延、発育および精神遅延、ならびに軽微な身体的異常
3	モノソミー、トリソミー(体細胞)	非ホジキンリンパ腫
4	モノソミー、トリソミー(体細胞)	急性非リンパ球性白血病(ANLL)
5	5p	ネコ鳴き;ルジューヌ症候群
5	5q(体細胞)モノソミー、トリソミー	骨髄異形成症候群
6	モノソミー、トリソミー(体細胞)	明細胞肉腫
7	7q11.23欠失	ウィリアムズ症候群
7	モノソミー、トリソミー	小児期の7モノソミー症候群;体細胞:腎皮質腺腫;骨髄異形成症候群
8	8q24.1欠失	ランガー・ギーディオン症候群(Langer-Giedon syndrome)
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(Warkany syndrome);体細胞:慢性骨髄性白血病
9	9pモノソミー	アルフィ症候群(Alfi's syndrome)
9	9pモノソミー、部分トリソミー	レトレ症候群(Rethore syndrome)

10

20

30

40

【 0 0 5 6 】

【表 1 - 2】

染色体	異常	疾患関連性
9	トリソミー	完全9トリソミー症候群(complete trisomy 9 syndrome);モザイク型9トリソミー症候群(mosaic trisomy 9 syndrome)
10	モノソミー、トリソミー(体細胞)	ALLまたはANLL
11	11p-	無虹彩症;ウィルムス腫瘍
11	11q-	ヤコブセン症候群(Jacobson Syndrome)
11	モノソミー(体細胞)、トリソミー	骨髄系列への影響(ANLL、MDS)
12	モノソミー、トリソミー(体細胞)	CLL、若年性顆粒膜細胞腫(JGCT)
13	13q-	13q症候群;オルベリ症候群(Orbeli syndrome)
13	13q14欠失	網膜芽細胞腫
13	モノソミー、トリソミー	パトー症候群
14	モノソミー、トリソミー(体細胞)	骨髄障害(MDS、ANLL、非定型CML)
15	15q11-q13欠失モノソミー	プラダー・ウィリー、アンジェルマン症候群
15	トリソミー(体細胞)	骨髄系列およびリンパ系系列への影響、例えば、MDS、ANLL、ALL、CLL)
16	16q13.3欠失	ルビンシュタイン・テイビ(Rubenstein-Taybi)
	モノソミー、トリソミー(体細胞)	乳頭状腎細胞癌(悪性)
17	17p-(体細胞)	骨髄性悪性疾患における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)	腎皮質腺腫
17	17p11.2-12 トリソミー	1型シャルコー・マリー・トゥース症候群;HNPP
18	18p-	18p部分モノソミー症候群またはグロウチイ・ラミー・ティーフリイ症候群
18	18q-	グロウチイ・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	

10

20

30

40

50

【 0 0 5 7 】

【 表 1 - 3 】

染色体	異常	疾患関連性
20	20p-	20pトリソミー症候群
20	20p11.2-12欠失	アラジール
20	20q-	体細胞:MDS、ANLL、真性赤血球増加症、慢性好中球性白血病
20	モノソミー、トリソミー(体細胞)	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Opitz G/BBB症候群、ケイラー心臓顔面症候群(Caylor cardiofacial syndrome)
22	モノソミー、トリソミー	完全22トリソミー症候群

10

【 0 0 5 8 】

20

【表 1 - 4】

表1B

症候群	染色体	開始	終了	間隔(Mb)	グレード
12q14微小欠失症候群	12	65,071,919	68,645,525	3.57	
15q13.3微小欠失症候群	15	30,769,995	32,701,482	1.93	
15q24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 (microduplication syndrome)	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 (認知神経科学的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 (認知神経科学的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 (神経発達障害に対する感受性遺伝子座)	1	146,512,930	147,737,500	1.22	3
1q21.1再発性微小重複 (神経発達障害に対する可能性のある感受性遺伝子座)	1	146,512,930	147,737,500	1.22	3
1q21.1血小板減少-橈骨欠損(TAR)症候群に対する感受性遺伝子座	1	145,401,253	145,928,123	0.53	3
22q11欠失症候群(口蓋心臓顔面/ディジョージ症候群)	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末端欠失症候群	22	22,115,848	23,696,229	1.58	
22q13欠失症候群(フェラン・マクダーミド症候群 (Phelan-Mcdermid syndrome))	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	

10

20

30

40

【 0 0 5 9 】

【表 1 - 5】

症候群	染色体	開始	終了	間隔(Mb)	グレード
8p23.1欠失症候群	8	8,119,295	11,765,719	3.65	
9qサブテロメア欠失症候群	9	140,403,363	141,153,431	0.75	1
成人発症型常染色体優性白質ジストロフィー(ADLD)	5	126,063,045	126,204,952	0.14	
アンジェルマン症候群(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(1型)	22	1	16,971,860	16.97	
シャルコー・マリー・トゥース症候群1型A(CMT1A)	17	13,968,607	15,434,038	1.47	1
ネコ鳴き症候群(5p欠失)	5	10,001	11,723,854	11.71	1
脳アミロイド血管症を伴う早期発症型アルツハイマー病		21	27,037,956	27,548,479	0.51
家族性腺腫性ポリポーシス		5	112,101,596	112,221,377	0.12
圧迫性麻痺に対する遺伝的傾向(Hereditary Liability to Pressure Palsies)(HNPP)	17	13,968,607	15,434,038	1.47	1
レリー・ワイル軟骨骨形成異常症(Leri-Weill dyschondroostosis)(LWD)-SHOX欠失	X	751,878	867,875	0.12	
レリー・ワイル軟骨骨形成異常症(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
ペリツェウス・メルツパッヘル病	X	102,642,051	103,131,767	0.49	
ポトツキ・ルプスキ症候群(Potocki-Lupski syndrome)(17p11.2重複症候群)	17	16,706,021	20,482,061	3.78	
ポトツキ・シェーファー症候群(Potocki-Shaffer syndrome)	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

【 0 0 6 0 】

10

20

30

40

【表 1 - 6】

症候群	染色体	開始	終了	間隔(Mb)	グレード
ブラダー・ウィリー症候群(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・ティビ症候群	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/裂足奇形1(SHFM1)	7	95,533,860	96,779,486	1.25	
ステロイドスルファターゼ欠損症(STS)	X	6,441,957	8,167,697	1.73	
WAGR11p13欠失症候群	11	31,803,509	32,510,988	0.71	
ウィリアムズ・ビューレン症候群(WBS)	7	72,332,743	74,616,901	2.28	1
ウォルフ・ヒルシュホーン症候群	4	10,001	2,073,670	2.06	1
Xq28(MECP2)重複	X	152,749,900	153,390,999	0.64	

グレード1の状態では、多くの場合、以下の特性の1つまたは複数を有する；病原となる異常；遺伝学者の間での強力な合意；高度な浸透性；なお変動する表現型を有し得るがいくつかの共通の特徴を有し得る；文献における症例の全てが臨床表現型を有する；異常を有する健康な個体の症例がない；DVGデータベースで報告されていないまたは健康な集団では見いだされない；機能データにより単一遺伝子または多重遺伝子の量効果が確認される；確認されたまたは強力な候補遺伝子；定義済みの臨床管理的意味；サーベイランスの意味がある公知のがんリスク；多数の情報源（OMIM、GeneReviews、Orphanet、Unique、Wikipedia）；および/または診断への使用（生殖カウンセリング）に利用可能である。

【0061】

グレード2の状態では、多くの場合、以下の特性の1つまたは複数を有する；病原となる可能性が高い異常；高度な浸透性；表現型が変動し、DD以外の一貫した特徴がない；文献における症例/報告が少数である；報告された症例の全てが臨床表現型を有する；機能データまたは確認された病原性遺伝子がない；多数の情報源（OMIM、GeneReviews、Orphanet、Unique、Wikipedia）；および/または診断目的および生殖カウンセリングのために使用することができる。

【0062】

グレード3の状態では、多くの場合、以下の特性の1つまたは複数を有する；感受性遺伝子座；健康な個体または記載の発端者の影響を受けていない親；対照母集団において存在する；非浸透性；表現型が軽度であり特異的でない；特徴の一貫性が低いこと；機能データまたは確認された病原性遺伝子がない；データの供給源が限られている；第2の診断の可能性が、大多数から逸脱する症例または新規の臨床的所見が存在する場合の可能性のままである；および/または診断目的で使用する際の注意および生殖カウンセリングのための慎重な助言。

【0063】

子癇前症

いくつかの実施形態では、子癇前症の有無を、本明細書に記載の方法または装置を使用することによって決定する。子癇前症は、妊娠中に高血圧症が生じ（すなわち妊娠誘導性高血圧症）、尿中に著しい量のタンパク質を伴う状態である。いくつかの場合には、子癇前症は、細胞外核酸のレベルの上昇および/またはメチル化パターンの変更にも関連する。例えば、細胞外胎児由来高メチル化RASSF1Aレベルと子癇前症の重症度との間の

10

20

30

40

50

正の相関が観察された。ある特定の例では、子癩前症の胎盤において、正常対照と比較して、H19遺伝子についてDNAのメチル化の増加が観察される。

【0064】

子癩前症は、世界的に、母体および胎児/新生児の死亡率および罹患率の主な原因の一つである。血漿および血清中の循環している無細胞核酸は、出生前診断を含めた種々の医学分野における有望な臨床的適用を伴う新規のバイオマーカーである。切迫子癩前症の指標として母体の血漿中の無細胞胎児(cff)DNAの定量的変化が、種々の試験において、例えば、雄性特異的なSR Y遺伝子座またはDYS14遺伝子座についてのリアルタイム定量的PCRを使用して報告されている。早発性子癩前症の場合、レベルの上昇は第1三半期に見られ得る。症状が発症する前のcffDNAのレベルの上昇は、組織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につながる絨毛間腔内の低酸素/再酸素化に起因する可能性がある。子癩前症では、母体の循環中へのcffDNAの脱落が増加する証拠に加えて、cffDNAの腎クリアランスが低下する証拠も存在する。胎児DNAの量は現在、Y染色体特異的配列を定量することによって決定されるので、全無細胞DNAの測定、または、DNAのメチル化などの性別に依存しない胎児の後成的マーカーの使用などの代替の手法により、代替法がもたらされる。胎盤起源の無細胞RNAは、臨床の実施において子癩前症をスクリーニングし、診断するために使用することができる別の代替のバイオマーカーである。胎児RNAは、それを分解から保護する細胞レベル下の胎盤粒子を伴う。胎児RNAレベルは、時には子癩前症の妊婦において対照と比較して10倍高くなり、したがって、臨床の実施において子癩前症をスクリーニングし、診断するために使用することができる代替のバイオマーカーである。

10

20

【0065】

病原体

いくつかの実施形態では、病原となる状態の有無を本明細書に記載の方法または装置によって決定する。病原となる状態は、宿主が、これだけに限定されないが、細菌、ウイルスまたは真菌を含めた病原体に感染することによって引き起こされる可能性がある。病原体は、一般には、宿主核酸と区別可能であり得る核酸(例えば、ゲノムDNA、ゲノムRNA、mRNA)を保有するので、本明細書において提供される方法および装置を使用して、病原体の有無を決定することができる。多くの場合、病原体は、例えば、後成的な状態および/または1つまたは複数の配列の変動、重複および/または欠失などの特定の病原体に独特の特性を持つ核酸を保有する。したがって、本明細書において提供される方法を使用して、特定の病原体または病原体の変異体(例えば株)を同定することができる。

30

【0066】

がん

いくつかの実施形態では、細胞増殖障害(例えば、がん)の有無を、本明細書に記載の方法または装置を使用することによって決定する。例えば、血清中の無細胞核酸のレベルが、種々の種類のがんの患者において健康な患者と比較して上昇する可能性がある。転移性疾患の患者では、例えば、時には、血清DNAレベルが非転移性患者のおよそ2倍であり得る。転移性疾患の患者は、例えば、がん特異的マーカーおよび/またはある特定の塩基多型または短いタンデム反復によって同定することもできる。循環しているDNAのレベルの上昇と正に相関する可能性があるがんの種類⁴⁰の非限定的な例としては、乳がん、結腸直腸がん、胃腸がん、肝細胞がん、肺がん、黒色腫、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、膵がん、および前立腺がんが挙げられる。種々のがんは、例えば、後成的な状態および/または配列の変動、重複および/または欠失などの、非がん性の健康な細胞由来の核酸と区別可能な特性を持つ核酸を保有する可能性があり、時にはそれらを血流中に放出する可能性がある。そのような特性は、例えば、特定のがんの種類に特異的であり得る。したがって、さらに、本明細書において提供される方法を使用して、特定のがんの種類を同定することができることが意図されている。

40

【0067】

50

他の遺伝的変異

いくつかの実施形態では、遺伝的変異の有無を、本明細書に記載の方法または装置を使用することによって決定することができる。ある特定の実施形態では、遺伝的変異は、コピー数多型(CNV)、微小欠失、重複、または影響を受けていない個体において観察された予測遺伝子の量からの遺伝子の量の変動を引き起こすまたはもたらす任意の状態から選択される1つまたは複数の状態である。いくつかの実施形態では、コピー数多型とは、1つまたは複数のゲノミックセクション、染色体、または染色体の一部の構造的再編成を指し、この再編成は、多くの場合、欠失、重複、逆位、および/または転座によって引き起こされる。CNVは、遺伝性のものまたは新規の突然変異によって引き起こされるものであり得、一般には、1つまたは複数のゲノミックセクションのコピーの数が異常になる(例えば、影響を受けていない試料に対して遺伝子量が異常になる)。いくつかの実施形態では、コピー数多型は1キロベースの小ささから数メガベースまでにわたる領域において起こり得る。CNVは、種々の細胞遺伝学的方法(FISH、CGH、aCGH、核型分析)および/または配列決定方法を使用して検出することができる。

10

【0068】

微小欠失は、一般には、選択されたゲノミックセクションまたはセグメントに位置する遺伝物質(例えば、特定の領域を代表するDNA、遺伝子、核酸)の量が、影響を受けていない領域に対して減少することである。微小欠失、および微小欠失によって引き起こされる症候群は、多くの場合、それが存在しないことにより時には疾患状態が付与される1つまたは複数の遺伝子にわたる1つまたは複数の染色体セグメントの小さな欠失(例えば、一般に5メガベース未満)を特徴とする。微小欠失は、時には、減数分裂の間の染色体の乗換えの誤りによって引き起こされる。多くの場合、微小欠失は、現在利用されている核型分析方法では検出可能ではない。

20

【0069】

染色体重複、または微小重複または重複は、一般に、遺伝物質(例えば、特定の領域を代表するDNA、遺伝子、核酸)の量が、影響を受けていない領域と比較して増加する1つまたは複数の領域である。重複は、しばしば、相同組換えの誤りの結果として起こる、またはレトロトランスポゾン事象に起因する。重複は、いくつかの場合には、小さな領域(何千もの塩基対)から染色体全体までにわたる可能性がある。重複は、ある特定の種類の増殖性疾患に関連付けられている。重複は、ゲノムマイクロアレイまたは比較遺伝学的ハイブリダイゼーション(CGH)を使用して特徴付けることができる。重複は、時には、1回または複数回反復される(例えば、1回、2回、3回、4回、5回、6回、7回、8回、9回または10回反復される)遺伝子領域と特徴付けられる。

30

試料

本明細書に記載の方法および装置において利用する核酸は、多くの場合、被験体から得た試料から単離される。いくつかの実施形態では、被験体は試験被験体と称され、ある特定の実施形態では、被験体は試料被験体または参照被験体と称される。いくつかの実施形態では、試験被験体とは、遺伝的変異の有無について評価される被験体を指す。試料被験体、または参照被験体は、多くの場合、試験被験体と比較するための基礎とし利用される被験体であり、参照被験体は、時には、参照被験体が、試験被験体について評価している遺伝的変異を有さない、またはそれを有することが分かっているという知見に基づいて選択される。被験体は、これだけに限定されないが、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ない生物体または生きていない生物体であってよい。これだけに限定されないが、哺乳動物、爬虫類、鳥類、両生類、魚類、有蹄動物、反芻動物、ウシ亜科の動物(例えば、ウシ)、ウマ科の動物(例えば、ウマ)、ヤギ亜科およびヒツジ属の動物(例えば、ヒツジ、ヤギ)、イノシシ科の動物(例えば、ブタ)、ラクダ科の動物(例えば、ラクダ、ラマ、アルパカ)、サル、類人猿(例えば、ゴリラ、チンパンジー)、クマ科の動物(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た任意のヒトまたは非ヒト動物を選択することができる。被験体は雄であっても雌(例えば女性)であってもよい。

40

50

【0070】

核酸は、任意の種類適切な生物検体または試料から単離することができる。検体の非限定的な例としては、これだけに限定することなく、臍帯血、絨毛膜絨毛、羊水、脳脊髄液 (c e r b r o s p i n a l f l u i d)、脊髄液、洗浄液 (l a v a g e f l u i d) (例えば、気管支肺胞、胃、腹膜、管、耳、関節鏡検査)、生検試料 (例えば、着床前胚由来)、腹腔穿刺試料、胎児有核細胞または胎児細胞残留物、女性生殖器系の洗浄 (w a s h i n g)、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、母乳、乳汁、胚細胞および胎児の細胞 (例えば胎盤細胞) を含めた被験体由来の体液または組織が挙げられる。いくつかの実施形態では、生体試料は、血液であってよく、時には血漿または血清であってよい。本明細書で使用される場合、「血液」とは、一般に、全血、または、例えば、慣習的に定義される血清および血漿などの血液の任意の画分を指す。血漿とは、抗凝固薬を用いて処理した血液を遠心分離することによって生じる全血の画分を指す。血清とは、血液試料が凝固した後に残る体液の水分を多く含む部分を指す。体液または組織試料は、多くの場合、病院または診療所が一般に従う標準のプロトコールに従って採取する。血液については、多くの場合、適量の末梢血 (例えば、3 ~ 40 ミリリットル) を採取し、さらに調製する前に、標準の手順に従って保管することができる。核酸を抽出する体液または組織試料は細胞を含まなくてよい。いくつかの実施形態では、体液または組織試料は、細胞エレメントまたは細胞残留物を含むことがよい。いくつかの実施形態では、胎児の細胞またはがん細胞が試料中に含まれていてよい。

10

20

【0071】

試料は不均一であってよく、これは、2種類以上の核酸種が試料中に存在することを意味する。例えば、不均一な核酸としては、これだけに限定されないが、(i) 胎児に由来する核酸および母体に由来する核酸、(ii) がん核酸および非がん核酸、(iii) 病原体核酸および宿主核酸、より一般的には、(iv) 突然変異した核酸および野生型核酸を挙げることができる。試料は、例えば、胎児の細胞および母体の細胞、がん細胞および非がん細胞、または病原性細胞および宿主細胞など、2つ以上の細胞型が存在することに起因して不均一であってよい。いくつかの実施形態では、少数の核酸種と大多数の核酸種が存在する。

【0072】

本明細書に記載の技術を出生前に適用するために、検査に適した妊娠期間にある女性から、または妊娠の可能性について検査されている女性から体液または組織試料を採取することができる。適切な妊娠期間は、実施される出生前検査に応じて変動し得る。ある特定の実施形態では、妊婦被験体は、時には妊娠第1三半期にあり、時には妊娠第2三半期にあり、時には妊娠第3三半期にある。ある特定の実施形態では、胎児妊娠期間約1週から約45週の間 (例えば、胎児妊娠期間1 ~ 4週、4 ~ 8週、8 ~ 12週、12 ~ 16週、16 ~ 20週、20 ~ 24週、24 ~ 28週、28 ~ 32週、32 ~ 36週、36 ~ 40週または40 ~ 44週) に、および時には胎児妊娠期間約5週から約28週の間 (例えば、胎児妊娠期間6週、7週、8週、9週、10週、11週、12週、13週、14週、15週、16週、17週、18週、19週、20週、21週、22週、23週、24週、25週、26週または27週) に妊婦から体液または組織を採取する。

30

40

【0073】

核酸の単離および処理

核酸は、1つまたは複数の供給源 (例えば、細胞、土壌など) から当技術分野で公知の方法によって得ることができる。細胞溶解手順および試薬は当技術分野で公知であり、一般に、化学的溶解方法、物理的溶解方法、または電解による溶解方法によって実施され得る。例えば、化学的方法では、一般に、溶解剤を使用して細胞を破壊し、細胞から核酸を抽出し、その後、カオトロピック塩で処理する。凍結/解凍し、その後粉砕すること、細胞圧搾の使用などの物理的方法も有用である。高塩濃度溶解手順も一般に使用される。例えば、アルカリ性溶解手順を利用することができる。後者の手順では伝統的にフェノー

50

ル - クロロホルム溶液の使用が組み込まれ、3つの溶液を伴う、フェノール - クロロホルムを含まない代替的な手順を利用することができる。後者の手順では、1つの溶液は15 mMのトリス、pH 8.0; 10 mMのEDTAおよび100 μg/mlのRNA分解酵素Aを含有してよく、第2の溶液は0.2 NのNaOHおよび1%のSDSを含有してよく、第3の溶液は3 MのKOAc、pH 5.5を含有してよい。これらの手順は、その全体が本明細書に組み込まれるCurrent Protocols in Molecular Biology、John Wiley & Sons、N.Y.、6.3.1~6.3.6(1989年)において見ることができる。

【0074】

「核酸」および「核酸分子」という用語は互換的に使用される。これらの用語は、デオキシリボ核酸(DNA、例えば、相補DNA(cDNA)、ゲノムDNA(gDNA)など)、リボ核酸(RNA、例えば、メッセンジャーRNA(mRNA)、低分子阻害性RNA(sRNA)、リボソームRNA(rRNA)、転移RNA(tRNA)、マイクロRNA、胎児または胎盤で高度に発現されるRNAなど)、および/またはDNA類似体またはRNA類似体(例えば、塩基類似体、糖類似体および/または非ネイティブな主鎖などを含有する)、RNA/DNAハイブリッドおよびポリアミド核酸(PNA)などの任意の組成形態の核酸を指し、その全てが一本鎖または二本鎖の形態であってよい。他に限定がない限り、核酸は、天然のヌクレオチドの公知の類似体を含んでよく、そのいくつかは、天然に存在するヌクレオチドと同様に機能し得る。核酸は、本発明のプロセスを行うために有用な任意の形態であってよい(例えば、直鎖状、環状、高次コイル、一本鎖、二本鎖など)。核酸は、ある特定の実施形態では、プラスミド、ファージ、自己複製配列(ARS)、セントロメア、人工染色体、染色体、またはin vitroまたは宿主細胞、細胞、細胞核または細胞質において複製することができるまたは複製される他の核酸であってよい、またはそれ由来であってよい。いくつかの実施形態では、核酸は、単一の染色体由来であってよい(例えば、核酸試料は、二倍体の生物体から得た試料の1つの染色体由来であってよい)。核酸は、一本鎖ポリヌクレオチド(「センス」または「アンチセンス」、「プラス」鎖または「マイナス」鎖、「フォワード」読み枠または「リバーズ」読み枠)および二本鎖ポリヌクレオチドから合成、複製または増幅されたRNAまたはDNAの誘導體、変異体および類似体も含む。デオキシリボヌクレオチドとしては、デオキシアデノシン、デオキシシチジン、デオキシグアノシンおよびデオキシチミジンが挙げられる。RNAについては、塩基シトシンがウラシルに置き換えられており、糖の2'位にヒドロキシル部を含む。核酸は、被験体から得た核酸を鋳型として使用して調製することができる。

【0075】

核酸は、別の核酸と比較して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してよい。核酸は、例えばcDNAライブラリーまたはRNAライブラリーなどの核酸ライブラリー由来であってよい。核酸は、試料由来の核酸分子の核酸精製もしくは単離および/または増幅の結果であってよい。本明細書に記載のプロセスのために提供される核酸は、1つの試料由来の核酸または2つ以上の試料由来(例えば、1以上、2以上、3以上の、4以上、5以上、6以上、7以上、8以上、9以上、10以上、11以上、12以上、13以上、14以上、15以上、16以上、17以上、18以上、19以上、または20以上の試料由来)の核酸を含有してよい。

【0076】

ある特定の実施形態では、核酸は細胞外核酸を含んでよい。細胞外核酸は、多くの場合、実質的に細胞を有さない供給源から単離された核酸である。細胞外核酸は、多くの場合、検出可能な細胞を含まず、細胞エレメントまたは細胞残留物を含有し得る。細胞外核酸の細胞を含まない供給源の非限定的な例は、血漿、血清および尿である。理論によって限定されることなく、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ってよく、多くの場合、大きなスペクトルにわたって一連の長さを有する細胞外核酸の基礎をもたらす(例えば、「ラダー」)。

10

20

30

40

50

【 0 0 7 7 】

細胞外核酸は、異なる核酸種を含んでよく、したがって、本明細書では、ある特定の実施形態では「不均一」と称される。例えば、がんを有する人由来の血清または血漿は、がん細胞由来の核酸および非がん細胞核酸を含んでよい。別の例では、妊婦由来の血清または血漿は、母体核酸および胎児核酸を含んでよい。いくつかの場合には、胎児核酸は、時には核酸全体の約5%～約50%である(例えば、全核酸の約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)。いくつかの実施形態では、核酸中の大多数の胎児核酸が約500塩基対以下の長さである(例えば、胎児核酸の約80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または100%が約500塩基対以下の長さである)。いくつかの実施形態では、核酸中の大多数の胎児核酸が約250塩基対以下の長さである(例えば、胎児核酸の約80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または100%が約250塩基対以下の長さである)。いくつかの実施形態では、核酸中の大多数の胎児核酸が約200塩基対以下の長さである(例えば、胎児核酸の約80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または100%が約200塩基対以下の長さである)。いくつかの実施形態では、核酸中の大多数の胎児核酸が約150塩基対以下の長さである(例えば、胎児核酸の約80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または100%が約150塩基対以下の長さである)。いくつかの実施形態では、核酸中の大多数の胎児核酸が約100塩基対以下の長さである(例えば、胎児核酸の約80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または100%が約100塩基対以下の長さである)。

10

20

【 0 0 7 8 】

ある特定の実施形態では、核酸を含有する試料(複数可)を処理することなく本明細書に記載の方法を行うために核酸を提供することができる。いくつかの実施形態では、核酸を含有する試料(複数可)を処理した後に本明細書に記載の方法を行うために核酸が提供される。例えば、核酸は、試料(複数可)から抽出、単離、精製または増幅することができる。本明細書で使用される場合、「単離された」とは、その元の環境(例えば、天然に存在する場合は天然の環境、または外因的に発現させた場合は宿主細胞)から取り出され、したがって、その元の環境から人為的な介入によって(例えば、「人間の手によって」)変更された核酸を指す。非核酸成分(例えば、タンパク質、脂質)が供給源試料中に存在する成分の量よりも少ない、単離された核酸が提供される。単離された核酸を含む組成物は、非核酸成分を約90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超含まなくてよい。本明細書で使用される場合、「精製された」とは、含有する核酸種が、核酸が由来する試料供給源中の核酸種よりも少ない核酸を指す。核酸を含む組成物は、他の核酸種を約90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または99%超含まなくてよい。増幅された核酸は、多くの場合、試料の核酸を、試料中の核酸またはその一部のヌクレオチド配列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直線的にまたは指数関数的に生成するプロセスに供することによって調製される。

30

40

【 0 0 7 9 】

核酸は、ある特定の実施形態では、本明細書に記載のプロセスのための核酸をもたらす前に、核酸を、核酸断片を生成する方法に供することによって処理することもできる。いくつかの実施形態では、断片化または切断に供された核酸は、公称、アベレージまたは平均の長さが約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塩基対、15塩基対、20塩基対、25塩基対、30塩基対

50

、35塩基対、40塩基対、45塩基対、50塩基対、55塩基対、60塩基対、65塩基対、70塩基対、75塩基対、80塩基対、85塩基対、90塩基対、95塩基対、100塩基対、200塩基対、300塩基対、400塩基対、500塩基対、600塩基対、700塩基対、800塩基対、900塩基対、1000塩基対、2000塩基対、3000塩基対、4000塩基対、5000塩基対、6000塩基対、7000塩基対、8000塩基対または9000塩基対であってよい。断片は、当技術分野で公知の任意の適切な方法によって生成することができ、核酸断片のアベレージ、平均または公称の長さは、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。ある特定の実施形態では、小さな配列の変動を含有する、および/または比較的大きな量の公知のヌクレオチド配列情報を含有する配列を分析するために、長さが比較的短い核酸を利用することができる。いくつかの実施形態では、大きな配列の変動を含有する、および/または比較的少量のヌクレオチド配列情報を含有する配列を分析するために、長さが比較的長い核酸を利用することができる。

10

【0080】

核酸断片は、オーバーラップしているヌクレオチド配列を含有してよく、そのようなオーバーラップ配列により、断片化されていない対応核酸またはその一部のヌクレオチド配列の構築を容易にすることができる。例えば、1つの断片は部分配列(subsequence) x および y を有してよく、別の断片は、部分配列 y および z を有してよく、 x 、 y および z は5ヌクレオチド長以上であってよいヌクレオチド配列である。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、試料由来の核酸における x - y - z ヌクレオチド配列の構築を容易にすることができる。ある特定の実施形態では、核酸を部分的に断片化することもでき(例えば、不完全なまたは終結した特定の切断反応から)、完全に断片化することもできる。

20

【0081】

核酸は、これだけに限定することなく、物理的プロセス、化学的プロセスおよび酵素的プロセスを含む当技術分野で公知のさまざまな方法によって断片化することができる。そのようなプロセスの非限定的な例は、米国特許出願公開第20050112590号に記載されている(2005年5月26日公開、Van Den Boomら、表題「Fragmentation-based methods and systems for sequence variation detection and discovery」)。非特異的に切断された断片または特異的に切断された断片が生成されるようにある特定のプロセスを選択することができる。非特異的に切断された断片核酸を生成することができるプロセスの非限定的な例としては、限定することなく、核酸を、核酸をせん断力に曝露させる装置に接触させること(例えば、核酸を、シリンジ針を通過させること; フレンチプレスの使用); 核酸を照射(例えば、ガンマ、X線、UV照射; 照射の強さによって断片サイズを制御することができる)に曝露させること; 水中で核酸を煮沸すること(例えば、約500塩基対の断片をもたらす)および核酸を酸および塩基加水分解プロセスに曝露させることが挙げられる。

30

【0082】

本明細書で使用される場合、「断片化」または「切断」とは、核酸鋳型遺伝子分子またはその増幅産物などの核酸分子を2つ以上のより小さな核酸分子に切り離す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ってよく、例えば、化学的断片化、酵素的断片化、物理的断片化を含めた種々の方法、試薬または条件のいずれかによって実現することができる。

40

【0083】

本明細書で使用される場合、「断片」、「切断産物」、「切断された産物」またはその文法上の変異形は、核酸鋳型遺伝子分子またはその増幅産物の断片化または切断の結果として生じた核酸分子を指す。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生じた全ての核酸分子を指し得るが、一般には、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または核酸鋳型遺伝子分子の対応するヌクレオチド配列を含有

50

するその増幅産物の一部の断片化または切断の結果として生じた核酸分子のみを指す。例えば、増幅産物は、核酸鋳型配列の増幅されたヌクレオチド領域を超える1つまたは複数のヌクレオチドを含有してよい(例えば、プライマーは、核酸鋳型遺伝子分子と相補的なヌクレオチドに加えて、転写開始配列などの「余分の」ヌクレオチドを含有してよく、その結果、「余分の」ヌクレオチドまたは核酸鋳型遺伝子分子の増幅された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、少なくとも一部において、代表的な核酸鋳型分子に由来するまたはそれに基づくヌクレオチド配列情報を含有する増幅された核酸分子の一部から生じる断片を含んでよい。

【0084】

本明細書で使用される場合、「相補的切断反応」とは、同じ標的核酸もしくはタンパク質または参照核酸もしくはタンパク質の代替の切断パターンが生成されるように、同じ核酸に対して、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更することによって行う切断反応を指す。ある特定の実施形態では、核酸を、1つまたは複数の反応容器において、1種または複数種の特異的な切断剤(例えば、1種、2種、3種、4種、5種、6種、7種、8種、9種、10種、またはそれ以上の特異的な切断剤)を用いて処理することができる(例えば、核酸を別々の容器内で特異的な切断剤のそれぞれを用いて処理する)。

【0085】

核酸を1種または複数種の特異的な切断剤と接触させることによって核酸を特異的に切断することができる。本明細書で使用される場合、「特異的な切断剤」とは、核酸を1つまたは複数の特定の部位で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。特異的な切断剤は、多くの場合、特定の部位における特定のヌクレオチド配列に応じて特異的に切断する。

【0086】

酵素的な特異的切断剤の例としては、これだけに限定することなく、エンドヌクレアーゼ(例えば、DNA分解酵素(例えば、DNA分解酵素I、II);RNA分解酵素(例えば、RNA分解酵素E、F、H、P);Cleavase(商標)酵素;Taq DNAポリメラーゼ;E.coli DNAポリメラーゼIおよび真核生物構造特異的エンドヌクレアーゼ;マウスFEN-1エンドヌクレアーゼ;I型、II型またはIII型制限エンドヌクレアーゼ、例えば、Acc I、Afl III、Alu I、Alw44 I、Apa I、Asn I、Ava I、Ava II、BamH I、Ban II、Bcl I、Bgl I、Bgl II、Bln I、Bsm I、BssH II、BstE II、Cfo I、Cla I、Dde I、Dpn I、Dra I、EclX I、EcoR I、EcoR II、EcoR V、Hae II、Hae III、Hind II、Hind III、Hpa I、Hpa II、Kpn I、Ksp I、Mlu I、MluN I、Msp I、Nci I、Nco I、Nde I、Nde II、Nhe I、Not I、Nru I、Nsi I、Pst I、Pvu I、Pvu II、Rsa I、Sac I、Sal I、Sau3A I、Sca I、ScrF I、Sfi I、Sma I、Spe I、Sph I、Ssp I、Stu I、Sty I、Swa I、Taq I、Xba I、Xho I;グリコシラーゼ(例えば、ウラシル-DNAグリコシラーゼ(glycolysase)(UDG)、3-メチルアデニンDNAグリコシラーゼ、3-メチルアデニンDNAグリコシラーゼII、ピリミジン水和物-DNAグリコシラーゼ、FaPy-DNAグリコシラーゼ、チミンミスマッチ-DNAグリコシラーゼ、ヒポキサンチン-DNAグリコシラーゼ、5-ヒドロキシメチルウラシルDNAグリコシラーゼ(HmUDG)、5-ヒドロキシメチルシトシンDNAグリコシラーゼ、または1,N6-エテノ-アデニンDNAグリコシラーゼ);エキソヌクレアーゼ(例えば、エキソヌクレアーゼIII);リボザイム、およびDNAザイムが挙げられる。核酸は化学作用剤を用いて処理ことができ、修飾された核酸を切断することができる。非限定的な例では、核酸は、(i)アルキル化剤、例えば、アルキルプリンDNA-グリコシラーゼによって認識され、切断され

10

20

30

40

50

る N³-メチルアデニンおよび N³-メチルグアニンを含まないいくつかのアルキル化された塩基を生成するメチルニトロソ尿素など；(i i) DNA 内のシトシン残基の脱アミノ化を引き起こして、ウラシル N-グリコシラーゼによって切断することができるウラシル残基を形成させる亜硫酸水素ナトリウム；および(i i i) グアニンを、ホルムアミドピリミジン DNA N-グリコシラーゼによって切断することができるその酸化形態である 8-ヒドロキシグアニンに変換する化学作用剤を用いて処理することができる。化学的切断プロセスの例としては、これだけに限定することなく、アルキル化(例えば、ホスホチオエートで修飾された核酸のアルキル化)；P 3'-N 5'-ホスホロアミデート(phosphoramidate)を含有する核酸の酸不安定性の切断；および核酸の四酸化オスmiumおよびピペリジン処理が挙げられる。

10

【0087】

いくつかの実施形態では、断片化された核酸をサイズ分画手順に供すことができ、分画されたプールの全部または一部を単離または分析することができる。サイズ分画手順は、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による分離)。

【0088】

本明細書に記載の方法のための核酸をもたらす前に、核酸を核酸内のある特定のヌクレオチドを修飾するプロセスに曝露させることもできる。例えば、核酸をヌクレオチドのメチル化の状態に基づいて選択的に修飾するプロセスを核酸に適用することができる。さらに、高温、紫外線放射、x線などの条件により、核酸分子の配列の変化を誘導することができる。核酸は、本明細書に記載の配列解析または製造プロセスを行うために有用な任意の形態、例えば、固体または液体の形態でもたらすことができる。ある特定の実施形態では、核酸は、任意選択で、これだけに限定することなく、1種または複数種の緩衝液または塩を含めた1つまたは複数の他の成分を含む液体の形態でもたらすことができる。

20

【0089】

配列読み取りを得ること

配列決定、マッピングおよび関連する分析方法は当技術分野で公知である(例えば、参照により組み込まれる米国特許出願公開第2009/0029377号)。そのようなプロセスのある特定の態様は下に記載されている。

【0090】

読み取りは、一般に、本明細書に記載のまたは当技術分野で公知の任意の配列決定プロセスによって作製される短いヌクレオチド配列である。読み取りは、核酸断片の一方の末端から生成することができ(「シングルエンド読み取り」)、時には核酸の両末端から生成される(「ダブルエンド読み取り」)。ある特定の実施形態では、被験体から試料の核酸配列読み取りを「得ること」および/または1つまたは複数の参照人から生物検体の核酸配列読み取りを「得ること」は、核酸について直接配列決定して配列情報を得ることを伴う。いくつかの実施形態では、「得ること」は、別途核酸から直接得た配列情報を受け取ることを伴う。

30

【0091】

いくつかの実施形態では、1個体由来の1つの核酸試料を配列決定する。ある特定の実施形態では、各生体試料が1個体または2個体以上に由来する2つ以上の生体試料由来の核酸試料をプールし、そのプールを配列決定する。後者の実施形態では、各生体試料由来の核酸試料は、多くの場合、1種または複数種の独特の同定タグによって同定される。

40

【0092】

いくつかの実施形態では、ゲノムの画分を配列決定し、それを時には決定されたヌクレオチド配列に包含されるゲノムの量で表す(例えば、1未満の「倍率」カバレッジ)。ゲノムが約1倍カバレッジで配列決定された場合、ゲノムのヌクレオチド配列のおよそ100%が、読み取りによって表される。ゲノムの所与の領域が2つ以上の読み取りまたはオーバーラップしている読み取りによって包含され得る冗長性を用いてゲノムを配列決定することもできる(例えば、1を超える「倍率」カバレッジ)。いくつかの実施形態では、

50

約0.1倍～約100倍カバレッジ、約0.2倍～20倍カバレッジ、または約0.2倍～約1倍カバレッジ(例えば、約0.2倍カバレッジ、0.3倍カバレッジ、0.4倍カバレッジ、0.5倍カバレッジ、0.6倍カバレッジ、0.7倍カバレッジ、0.8倍カバレッジ、0.9倍カバレッジ、1倍カバレッジ、2倍カバレッジ、3倍カバレッジ、4倍カバレッジ、5倍カバレッジ、6倍カバレッジ、7倍カバレッジ、8倍カバレッジ、9倍カバレッジ、10倍カバレッジ、15倍カバレッジ、20倍カバレッジ、30倍カバレッジ、40倍カバレッジ、50倍カバレッジ、60倍カバレッジ、70倍カバレッジ、80倍カバレッジ、90倍カバレッジ)でゲノムを配列決定する。

【0093】

ある特定の実施形態では、1回の実行で配列決定される核酸プールの画分を、さらに副選択した後に配列決定する。ある特定の実施形態では、ハイブリダイゼーションに基づく技法(例えば、オリゴヌクレオチドアレイを使用する)を使用して、まず核酸配列について特定の染色体(例えば、潜在的に異数性の染色体および検査された異数性に關与しない他の染色体(複数可))からの副選択を行うことができる。いくつかの実施形態では、核酸をサイズによって分画することができ(例えば、ゲル電気泳動、サイズ排除クロマトグラフィーによってまたはマイクロフルイディクスに基づく手法によって)、ある特定の例では、分子量が低い核酸(例えば、300塩基対未満、200塩基対未満、150塩基対未満、100塩基対未満)を選択することによって、胎児核酸を濃縮することができる。いくつかの実施形態では、ホルムアルデヒドを添加することによってなど、母体のバックグラウンド核酸を抑制することによって、胎児核酸を濃縮することができる。いくつかの実施形態では、予め選択した核酸のプールの一部またはサブセットについて無作為に配列決定する。いくつかの実施形態では、配列決定の前に核酸を増幅する。いくつかの実施形態では、配列決定の前に核酸の一部またはサブセットを増幅する。

【0094】

本明細書に記載の方法を行うために適した任意の配列決定方法を利用することができる。いくつかの実施形態では、ハイスループットな配列決定方法を使用する。ハイスループットな配列決定方法では、一般に、クローニング的に増幅したDNA鋳型またはフローセルにおいて大規模並列処理様式で配列決定した単一のDNA分子を必要とする(例えば、Metzker M Nature Rev 11巻:31~46頁(2010年); Volkerdingら Clin Chem 55巻:641~658頁(2009年)に記載されている)。そのような配列決定方法では、デジタルの定量情報をもたらすこともでき、各配列読み取りは、個々のクローンDNA鋳型または単一のDNA分子を表す可算「配列タグ」である。ハイスループットな配列決定技術としては、例えば、可逆的ダイターミネーターを用いた1塩基合成反応(sequencing-by-synthesis)、オリゴヌクレオチドプローブライゲーションによる配列決定、パイロシーケンスおよびリアルタイム配列決定が挙げられる。

【0095】

ハイスループットな配列決定方法のために利用するシステムは市販されており、それらとしては、例えば、Roche 454 platform、Applied Biosystems SOLID platform、Helicos True Single Molecule DNA sequencing technology、Affymetrix Inc.のsequencing-by-hybridization platform、Pacific Biosciencesのsingle molecule, real-time (SMRT) technology、454 Life Sciences、Illumina/Solexa and Helicos Biosciencesのsequencing-by-synthesis platform、およびApplied Biosystemsのsequencing-by-ligation platformが挙げられる。Life technologiesのION TORRENT technologyおよびナノポア配列決定も、ハイスループットな配列決定手法において使用することができる。

【0096】

いくつかの実施形態では、第1世代の技術、例えば、自動化サンガー配列決定を含めたサンガー配列決定などを本明細書において提供される方法において使用することができる。発展している核酸イメージング技術（例えば透過型電子顕微鏡（TEM）および原子間力顕微鏡（AFM））の使用を含む追加的な配列決定技術も本明細書において考えられている。種々の配列決定技術の例が下に記載されている。

【0097】

本明細書に記載の方法において使用することができる核酸配列決定技術は、合成による配列決定（sequencing-by-synthesis）および可逆的ターミネーターに基づく配列決定（例えばIlluminaのGenome AnalyzerおよびGenome Analyzer II）である。この技術を用いると、数百万の核酸（例えばDNA）断片について並行して配列決定することができる。この種類の配列決定技術の1つの例では、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）が結合した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する。フローセルは、多くの場合、結合した分析物の上を試薬溶液が順序正しく通過することを保持および/または可能にするように構成することができる固体支持体である。フローセルは、しばしば、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またはミリメートル未満の尺度であり、多くの場合、分析物/試薬の相互作用が起こるチャンネルまたはレーンを有する。

【0098】

合成手順による特定の配列決定では、例えば、鑄型DNA（例えば、循環している無細胞DNA（ccfDNA））を、時には、ライブラリーを生成するための調製中に数百塩基対の長さ10に断片化する。いくつかの実施形態では、鑄型DNA（例えば、ccfDNA）をさらに断片化またはサイズ選択せずにライブラリー調製を実施することができる。ある特定の20実施形態では、ライブラリー生成を、実施例2に記載の通り製造者のプロトコルを改変したものを使用して実施する。ある特定の30実施形態では、自動化された方法および装置を使用して試料の単離およびライブラリー生成を実施する。簡単に述べると、ccfDNAをフィルイン反応（fill-in reaction）、エキソヌクレアーゼ反応またはフィルイン反応とエキソヌクレアーゼ反応の組合せによって末端修復する。生じた平滑末端修復ccfDNAを単一ヌクレオチド分伸長させ、それはアダプタープライマーの3'末端上の一ヌクレオチドオーバーハングと相補的であり、多くの場合、それによりライゲーション効率が上昇する。伸長/オーバーハングヌクレオチドのために任意の相補的なヌクレオチドを使用することができるが（例えば、A/T、C/G）、しばしば、末端修復DNAを伸長させるためにアデニンが使用され、多くの場合、3'末端オーバーハングヌクレオチドとしてチミンが使用される。

【0099】

合成手順によるある特定の配列決定では、例えば、アダプターオリゴヌクレオチドはフローセルアンカーと相補的であり、これを時には、修飾ccfDNA（例えば、末端修復し、単一ヌクレオチド延長させた）を、固体支持体、フローセルの内側表面に結びつけるために利用する。いくつかの実施形態では、アダプタープライマーは、指標ヌクレオチド、または「バーコード」ヌクレオチド（例えば、試料を一義的に同定することを可能にする指標プライマーとして使用可能なヌクレオチドの独特の配列）、1つまた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、シングルエンド配列決定プライマー、ペアエンド配列決定プライマー、多重化配列決定プライマーなどと相補的な配列）、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/指標、アダプター/指標/配列決定）を含む。アダプタープライマーに含有される指標プライマーまたはヌクレオチドは、多くの場合、6ヌクレオチド以上の長さであり、しばしば、プライマー内に、指標ヌクレオチドが配列決定反応の間に配列決定される最初のヌクレオチドになるように位置づけられる。ある特定の40実施形態では、指標ヌクレオチドまたはバーコードヌクレオチドは試料と結びついているが、配列読み取りの50

質が損なわれることを回避するために別々の配列決定反応で配列決定する。その後、バーコードの配列決定および試料の配列決定からの読み取りを一緒に連結させ、読み取りを逆多重化する。連結し逆多重化した後、配列読み取りを本明細書に記載の通りさらに補正または処理することができる。

【0100】

合成手順によるある特定の配列決定では、指標プライマーを利用することにより、フローセルレーンにおける配列反応を多重化することが可能になり、それにより、フローセル1レーン当たり複数の試料を分析することが可能になる。所与のフローセルレーンにおいて分析することができる試料の数は、多くの場合、ライブラリー調製の間に使用する独特の指標プライマーの数に左右される。指標プライマーは、いくつもの商業的な供給源（例えば、Illumina、Life Technologies、NEB）から入手可能である。実施例2に記載の反応を、試験時に利用可能ないくつかの市販のキットの1つを使用して実施し、これは12の独特の指標プライマーを含んだ。市販のマルチプレックス配列決定キットの非限定的な例としては、Illuminaの多重化試料調製オリゴヌクレオチドキットおよび多重化配列決定プライマーおよびPhiX対照キット（例えば、それぞれIlluminaカタログ番号PE-400-1001およびPE-400-1002）が挙げられる。本明細書に記載の方法は、12の指標プライマーに限定されず、任意の数の独特の指標プライマー（例えば、4、8、12、24、48、96、またはそれ以上）を使用して実施することができる。独特の指標プライマーの数が多いほど、単一のフローセルレーンにおいて多重化することができる試料の数が増える。12の指標プライマーを使用した多重化により、96の試料（例えば、96ウェルマイクロウェルプレートのウェルの数と等しい）を8レーンのフローセルで同時に分析することが可能になる。同様に、48の指標プライマーを使用した多重化により、384の試料（例えば、384ウェルマイクロウェルプレートのウェルの数と等しい）を8レーンのフローセルで同時に分析することが可能になる。

【0101】

合成手順によるある特定の配列決定では、アダプター修飾した一本鎖の鋳型DNAをフローセルに加え、限界希釈条件下でアンカーとハイブリダイズさせることによって固定化する。エマルジョンPCRとは対照的に、DNA鋳型をフローセル内で捕捉DNA鎖の「アーチ形成」に依拠する「架橋」増幅によって増幅し、近接するアンカーオリゴヌクレオチドとハイブリダイズさせる。多数の増幅サイクルにより、単一分子DNA鋳型がクローン的に増幅されたアーチ形の「クラスター」に変換され、各クラスターはおよそ1000クローンの分子を含有する。フローセル1個当たりおよそ 50×10^6 個の別々のクラスターを生成することができる。配列決定するために、クラスターを変性させ、その後化学的切断反応および洗浄によりフォワード鎖のみをシングルエンド配列決定のために残す。アダプター配列と相補的なプライマーとハイブリダイズさせることによってフォワード鎖の配列決定を開始し、その後、ポリメラーゼおよび4つの異なるように着色した蛍光可逆的ダイターミネーターの混合物を加えた。ターミネーターはクローンクラスター内の各鎖における配列相補性に応じて組み込まれる。組み込み後、過剰な試薬を洗い流し、クラスターを光学的に調べ、蛍光を記録する。逐次的な化学的ステップを用いて、可逆的ダイターミネーターを遮断解除し、蛍光標識を切断し、洗い流し、次の配列決定サイクルを実施する。この反復的な合成による配列決定プロセスは、時には、36塩基の読み取り長を生成するためにおよそ2.5日を必要とする。フローセル1個当たり 50×10^6 個のクラスターを用いると、全体的な配列アウトプットは、分析の実行1回当たり10億塩基対（Gb）を超え得る。

【0102】

本明細書に記載の方法と一緒に使用することができる別の核酸配列決定技術は454シーケンシング（Roche）である。454シーケンシングでは、実行1回当たり約400~600メガベースのDNAを配列決定することができる大規模並行パイロシーケンシングシステムを使用する。プロセスは、一般には2つのステップを伴う。第1のステッ

10

20

30

40

50

プでは、試料核酸（例えばDNA）を、時には、より小さな断片（300～800塩基対）に分画し、磨く（各末端を平滑にする）。次いで、短いアダプターを断片の末端にライゲーションする。これらのアダプターにより、試料 - ライブラリー断片の増幅および配列決定の両方のためのプライミング配列がもたらされる。1つのアダプター（アダプターB）は、DNAライブラリーをストレプトアビジンでコーティングしたビーズ上に固定化するための5'-ビオチンタグを含有する。ニック修復後、非ビオチン化鎖を放出させ、一本鎖の鋳型DNA（ssDNA）ライブラリーとして使用する。ssDNAライブラリーを、その質について評価し、emPCRのために必要な最適濃度（ビーズ1個当たりのDNAコピー）を滴定によって決定する。ssDNAライブラリーをビーズ上に固定化する。ライブラリー断片を含有するビーズは単一のssDNA分子を有する。ビーズに結合させたライブラリーを、増幅試薬を用いて乳化して油中水混合物にする。各ビーズをPCR増幅が起こるそれ自体のマイクロリアクター内で捕捉する。これにより、ビーズに固定化した、クローン的に増幅されたDNA断片が生じる。

10

【0103】

454シーケンシングの第2のステップでは、一本鎖の鋳型DNAライブラリービーズを、DNAポリメラーゼを含有するインキュベーションミックスに加え、スルフィラーゼおよびルシフェラーゼを含有するビーズと一緒にピコリットルサイズのウェルを含有するデバイス上に積み重ねた。各DNA断片に対してパイロシーケンスを並行して実施する。1つまたは複数のヌクレオチドが付加することにより、光信号が生成し、それがシーケンシング計器内のCCDカメラによって記録される。シグナル強度は、組み込まれたヌクレオチドの数に比例する。パイロシーケンスでは、ヌクレオチドが付加するとピロリン酸（PPi）が放出されることを活用する。PPiを、アデノシン5'ホスホ硫酸の存在下でATPスルフィラーゼによってATPに変換する。ルシフェラーゼはATPを使用してルシフェリンをオキシルシフェリンに変換し、この反応により、識別され、分析される光が生成する（例えば、Margulies, M.ら Nature 437巻: 376～380頁（2005年）を参照されたい）。

20

【0104】

本明細書において提供される方法において使用することができる別の核酸配列決定技術は、Applied BiosystemsのSOLiD（商標）技術である。SOLiD（商標）ライゲーションによる配列決定（sequencing-by-ligation）では、核酸断片のライブラリーを試料から調製し、クローンビーズの集団を調製するために使用する。この方法を用いると、各ビーズ（例えば磁気ビーズ）の表面上に1つの種の核酸断片が存在する。試料核酸（例えばゲノムDNA）をせん断して断片にし、その後、アダプターを断片の5'末端および3'末端に付着させて、断片ライブラリーを生成する。アダプターは、一般には、ユニバーサルアダプター配列であり、したがって、あらゆる断片の出発配列が既知であり、かつ同一である。PCRのために必要な試薬を全て含有するマイクロリアクター内でエマルジョンPCRを行う。次いで、ビーズに付着した、生じたPCR産物をガラススライドに共有結合させる。次いで、プライマーをライブラリー鋳型内のアダプター配列とハイブリダイズさせる。4つの蛍光標識された二塩基（di-base）プローブのセットを配列決定プライマーとのライゲーションについて競合させる。二塩基プローブの特異性は、各ライゲーション反応における最初の塩基および2番目の塩基を調べることによって実現される。ライゲーション、検出および切断の多数のサイクルを実施し、サイクルの数により最終的な読み取り長が決定される。一連のライゲーションサイクルの後、ライゲーションサイクルの第2のラウンドのために伸長産物を取り出し、鋳型をn-1位と相補的なプライマーでリセットする。多くの場合、各配列タグについて5ラウンドのプライマーのリセットを完了する。プライマーのリセットプロセスを通して、各塩基を2つの独立したライゲーション反応において2つの異なるプライマーによって調べる。例えば、読み取り位5の塩基を、ライゲーションサイクル2において2番のプライマーによって、およびライゲーションサイクル1において3番のプライマーによってアッセイする。

30

40

50

【0105】

本明細書に記載の方法において使用することができる別の核酸配列決定技術は、Helicos True Single Molecule Sequencing (tSMS) である。tSMS 技法では、試料由来の各核酸（例えばDNA）鎖の3'末端にポリA配列を付加する。蛍光標識されたアデノシンヌクレオチドを付加することによって各鎖を標識する。次いで、DNA鎖を、フローセル表面に固定化された数百万のオリゴ-T捕捉部位を含有するフローセルとハイブリダイズさせる。鑄型は、 1 cm^2 当たり鑄型約1億個の密度で存在してよい。次いで、フローセルを配列決定装置内にローディングし、フローセルの表面にレーザー照射し、各鑄型の位置を明らかにする。CCDカメラにより、フローセル表面上の鑄型の位置をマッピングすることができる。次いで、鑄型蛍光標識を切断し、洗い流す。DNAポリメラーゼおよび蛍光標識したヌクレオチドを導入することによって配列決定反応を開始する。オリゴ-T核酸がプライマーとしての機能を果たす。ポリメラーゼにより、標識したヌクレオチドが鑄型内のプライマーに指向的に組み込まれる。ポリメラーゼおよび取り込まれなかったヌクレオチドを除去する。蛍光標識したヌクレオチドが指向的に組み込まれた鑄型を、フローセル表面を画像処理することによって検出する。画像処理した後、切断ステップにより蛍光標識を除去し、所望の読み取り長が実現されるまで他の蛍光標識したヌクレオチドを用いてプロセスを繰り返す。各ヌクレオチド付加ステップを用いて配列情報を収集する（例えば、Harris T. D.ら、Science 320巻：106～109頁（2008年）を参照されたい）。

10

【0106】

本明細書において提供される方法において使用することができる別の核酸配列決定技術は、Pacific Biosciencesの単一分子、リアルタイム (single molecular, real-time) (SMRT (商標)) 配列決定技術である。この方法を用いて、4つのDNA塩基のそれぞれを4つの異なる蛍光色素のうちの一つに付着させる。これらの色素をリン酸基によって連結させる (phospholink)。単一のDNAポリメラーゼを、ゼロモード導波管 (ZMW) の底部に鑄型一本鎖DNAの単一分子で固定化する。ZMWは、ZMWの外側に急速に (マイクロ秒で) 拡散する蛍光ヌクレオチドのバックグラウンドに対してDNAポリメラーゼによる一塩基の組み込みを観察することを可能にする閉じ込め構造である。ヌクレオチドが成長している鎖に組み込まれるのには数ミリ秒かかる。この時間中に蛍光標識が励起され、蛍光シグナルが生じ、蛍光タグが切断される。対応する色素の蛍光が検出されることにより、その塩基が組み込まれたことが示される。次いで、このプロセスを繰り返す。

20

30

【0107】

本明細書に記載の方法において使用することができる別の核酸配列決定技術は、半導体技術と単純な配列決定化学を組み合わせ、化学的にコードされる情報 (A、C、G、T) を半導体チップ上のデジタル情報 (0、1) に直接翻訳するION TORRENT (Life Technologies) 単一分子配列決定である。ION TORRENTでは、微小機械処理したウェルの高密度アレイを使用して核酸配列決定を大規模並列処理式で実施する。各ウェルは異なるDNA分子を保持する。ウェルの真下にはイオン感応性層があり、その真下にイオンセンサーがある。一般には、ポリメラーゼによってヌクレオチドがDNAの鎖に組み込まれると、副生成物として水素イオンが放出される。ヌクレオチド、例えばCがDNA鑄型に付加され、次いでDNAの鎖に組み込まれると、水素イオンが放出される。そのイオン由来の電荷により、溶液のpHが変化し、これをイオンセンサーによって検出することができる。シークエンサーにより塩基を呼び出し、直接化学的情報からデジタル情報にすることができる。次いで、シークエンサーにより、チップに次々にヌクレオチドを逐次的に送り込む。チップに送り込まれる次のヌクレオチドがマッチしない場合、電圧の変化は記録されず、塩基は呼び出されない。DNA鎖上に2つの同一の塩基がある場合、電圧は倍増し、チップには呼び出された2つの同一の塩基が記録される。これは直接検出 (すなわち、スキッピング、カメラまたは光を伴わない検出) であるので、各ヌクレオチドの組み込みは数秒の間に記録される。

40

50

【0108】

本明細書に記載の方法において使用することができる別の核酸配列決定技術は、化学感応性電界効果トランジスタ(chemical-sensitive field effect transistor)(CHEMFET)アレイである。この配列決定技法の1つの例では、DNA分子を反応チャンバーに入れ、鑄型分子をポリメラーゼに結合させた配列決定プライマーとハイブリダイズさせることができる。配列決定プライマーの3'末端において1つまたは複数の三リン酸が新しい核酸鎖に組み込まれることは、CHEMFETセンサーにより、電流の変化によって検出することができる。アレイは多数のCHEMFETセンサーを有してよい。別の例では、単一の核酸をビーズに付着させ、その核酸をビーズ上で増幅することができ、個々のビーズをCHEMFETアレイ上の個々の反応チャンバーに写すことができ、各チャンバーはCHEMFETセンサーを有し、核酸について配列決定することができる(例えば、米国特許出願公開第2009/0026082号を参照されたい)。

10

【0109】

本明細書に記載の方法において使用することができる別の核酸配列決定技術は電子顕微鏡である。この配列決定技法の1つの例では、電子顕微鏡を使用して区別可能な金属標識を使用して個々の核酸(例えばDNA)分子を標識する。次いで、これらの分子を平らな表面上に伸ばし、電子顕微鏡を使用して画像処理して、配列を測定する(例えば、Moudrianakis E. N. および Beer M. Proc Natl Acad Sci USA. 1965年3月; 53巻: 564~71頁を参照されたい)。いくつかの場合には、透過型電子顕微鏡(TEM)を使用する(例えばHalcyon MolecularのTEM法)。Individual Molecule Placement Rapid Nano Transfer(IMPRNT)と称されるこの方法は、重原子マーカを用いて選択的に標識した高分子量(例えば約150kb以上)のDNAの単一原子分解透過形電子顕微鏡画像処理を利用すること、およびこれらの分子を超薄膜上に超高密度(鎖間3nm)の並行のアレイに一貫した塩基間の間隔で配置することを含む。重原子マーカ的位置を決定するため、およびDNAから塩基配列情報を抽出するために、電子顕微鏡を使用して膜上の分子を画像処理する(例えば、PCT特許公報WO2009/046445を参照されたい)。

20

【0110】

本明細書の方法を行うために使用することができる他の配列決定方法として、デジタルPCRおよびハイブリダイゼーションによる配列決定が挙げられる。デジタルポリメラーゼ連鎖反応(デジタルPCRまたはdPCR)を使用して、試料中の核酸を直接同定し、定量することができる。いくつかの実施形態では、デジタルPCRをエマルジョン中で実施することができる。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内で分離し、各核酸を個別にPCRによって増幅する。1ウェルあたりに存在する核酸が1つ以下になるように核酸を分離することができる。いくつかの実施形態では、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および母体の対立遺伝子)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。ハイブリダイゼーションによる配列決定では、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ことを伴い、複数のポリヌクレオチドプローブのそれぞれを任意選択で基質につなぎ留めることができる。いくつかの実施形態では、基質は既知のヌクレオチド配列のアレイを伴う平らな表面であってよい。アレイと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。いくつかの実施形態では、各プローブをビーズ、例えば、磁気ビーズなどにつなぎ留める。ビーズとのハイブリダイゼーションを同定することができ、それを使用して試料中の複数のポリヌクレオチド配列を同定することができる。

30

40

【0111】

いくつかの実施形態では、ナノポアシーケンシングを本明細書に記載の方法において使用することができる。ナノポアシーケンシングは、単一分子の配列決定技術であり、

50

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に従い、それが直接配列決定される。ナノポアは、直径1ナノメートル程度の小さな穴またはチャンネルである。ある特定の膜貫通細胞タンパク質がナノポアとしての機能を果たし得る（例えばアルファ溶血素）。いくつかの場合には、ナノポアを合成することができる（例えばシリコンプラットフォームを使用して）。ナノポアを導電性流体に浸漬し、それにわたって電位を印加することにより、ナノポアを通じたイオンの伝導に起因してわずかな電流が生じる。流れる電流の量はナノポアのサイズに対して感応性である。DNA分子がナノポアを通過するに従い、DNA分子上の各ヌクレオチドによってナノポアが異なる程度に閉塞され、電流に特徴的な変化が生じる。したがって、任意の所与の瞬間にナノポアを通過することができる電流の量は、ナノポアがA、C、G、T、またはいくつかの場合にはメチル-Cによって遮断されるかどうかに応じて変動する。DNA分子がナノポアを通過するときにナノポアを通る電流の変化により、DNA配列の直接読み取りが表される。いくつかの場合には、ナノポアを使用して、個々のDNA塩基が正しい順序でナノポアを通過するときにそれらを同定することができる（例えば、Soni GVおよびMeller A. C. Clin. Chem. 53巻：1996～2001頁（2007年）；PCT公開番号WO 2010/004265を参照されたい）。

10

【0112】

ナノポアを使用して核酸分子について配列決定することができるいくつかの方法が存在する。いくつかの実施形態では、デオキシリボヌクレアーゼなどのエキソヌクレアーゼ酵素を使用する。この場合、エキソヌクレアーゼ酵素を使用して、ヌクレオチドを核酸（例えばDNA）分子から逐次的に引き離す。次いで、ナノポアによってヌクレオチドが検出され、それらの放出の順に識別され、したがって、元の鎖の配列が読み取られる。そのような実施形態に関して、DNA分子から放出されるヌクレオチドの一部が、ナノポアのチャンネルに進入し、それと相互作用することができるように、エキソヌクレアーゼ酵素をナノポアに付着させることができる。エキソヌクレアーゼは、ナノポア構造の、チャンネルの開口部を形成するナノポアの一部の極めて近傍にある部位に付着させることができる。いくつかの場合には、エキソヌクレアーゼ酵素は、ナノポア構造に、そのヌクレオチドの出口軌道部位が、開口の一部を形成するナノポアの一部に向かって方向付けられるように付着させることができる。

20

【0113】

いくつかの実施形態では、核酸のナノポアシーケンシングは、核酸（例えばDNA）分子を、ポアを通して押し出すまたは引き寄せる酵素の使用を伴う。この場合、イオン電流はDNA分子内のヌクレオチドがポアを通過するときに変動する。電流の変動により、DNA配列が示される。そのような実施形態については、酵素をナノポア構造に付着させ、したがって、ポアを通るイオン電流の流れに干渉することなく標的核酸をナノポアのチャンネルを通して押し出すまたは引き寄せることができるようにすることができる。酵素は、ナノポア構造の、開口部の一部を形成する構造の一部の極めて近傍にある部位に付着させることができる。酵素は、例えば、その活性部位が開口部の一部を形成する構造の一部に向かって方向付けられるように、サブユニットに付着させることができる。

30

【0114】

いくつかの実施形態では、核酸のナノポアシーケンシングは、ナノポア検出器の極めて近傍においてポリメラーゼ副生成物を検出することを伴う。この場合、ヌクレオシドリン酸（ヌクレオチド）を標識し、したがって、ポリメラーゼがヌクレオチド鎖に付加されるとリン酸標識された種が放出され、リン酸標識された種がポアによって検出される。一般には、リン酸種は、各ヌクレオチドに対して特異的な標識を含有する。核酸鎖にヌクレオチドが逐次的に付加されるに従い、塩基付加の副生成物が検出される。リン酸標識された種が検出される順序を使用して、核酸鎖について配列を決定することができる。

40

【0115】

配列読み取りの長さは、多くの場合、特定の配列決定技術に関連付けられる。ハイスループットな方法では、例えば、サイズが数十塩基対から数百塩基対（bp）まで変動し得

50

る配列読み取りがもたらされる。ナノポアシーケンシングでは、例えば、サイズが数十塩基対から数百塩基対、数千塩基対まで変動し得る配列読み取りをもたらしすることができる。いくつかの実施形態では、配列読み取りは、平均値、中央値またはアベレージの長さ、約15bp~900bp長(例えば約20bp、約25bp、約30bp、約35bp、約40bp、約45bp、約50bp、約55bp、約60bp、約65bp、約70bp、約75bp、約80bp、約85bp、約90bp、約95bp、約100bp、約110bp、約120bp、約130、約140bp、約150bp、約200bp、約250bp、約300bp、約350bp、約400bp、約450bp、または約500bpの読み取りである。いくつかの実施形態では、配列読み取りは平均値、中央値またはアベレージの長さが約1000bp以上の読み取りである。

10

【0116】

いくつかの実施形態では、核酸は、蛍光シグナルまたは配列タグ情報を含んでよい。シグナルまたはタグの定量を、例えば、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、遺伝子チップ解析、マイクロアレイ、質量分析、細胞蛍光分析(cytofluorimetric analysis)、蛍光顕微鏡、共焦点レーザー走査顕微鏡、レーザー走査型サイトメトリー、アフィニティークロマトグラフィー、手動バッチモード分離(manual batch mode separation)、電場サスペンション(electric field suspension)、配列決定、およびその組合せなどのさまざまな技法において使用することができる。

20

【0117】

読み取りをマッピングすること

マッピングヌクレオチド配列読み取り(すなわち、物理的なゲノムの位置が未知の断片からの配列情報)をいくつかの方法で実施することができ、これは多くの場合、得られた配列読み取りを参照ゲノム内のマッチする配列とアラインメントすることを含む(例えば、Liら、「Mapping short DNA sequencing reads and calling variants using mapping quality score」、Genome Res.、2008年8月19日)。そのようなアラインメントでは、一般に、配列読み取りを参照配列に対してアラインメントし、アラインメントした配列読み取りは「マッピングされた」または「配列タグ」と称される。いくつかの場合には、マッピングされた配列読み取りは、「ヒット」と称される。いくつかの実施形態では、マッピングされた配列読み取りを種々のパラメータに応じて一緒に群分けし、下でさらに詳細に考察されている特定のゲノムセクションに割り当てる。

30

【0118】

種々のコンピュータによる方法を使用して、各配列読み取りをゲノムセクションにマッピングすることができる。配列をアラインメントす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は、BLAST、BLITZ、およびFASTA、またはその変形が挙げられる。いくつかの実施形態では、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データベース内の配列を用いて配列読み取りを見だし、かつ/またはアラインメントすることができる。BLASTまたは同様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対して検索することができる。次いで、例えば、検索ヒットを使用して、同定された配列を適切なゲノムセクション(下記)に選別して入れることができる。実施例1、実施例2および実施例3において生成された配列読み取りを、実施例2および実施例3に記載の通り、CASAVAバージョン1.6を使用してUCSChg19ヒト参照ゲノムにマッピングした。いくつかの実施形態では、配列読み取りのマッピングは、反復配列および/またはGC含量について補正する前に実施することができ、ある特定の実施形態では、配列読み取りのマッピングは、反復配列および/またはGC含量について補正した後に実施することができる。

40

【0119】

50

「配列タグ」とは、特定のゲノムセクションおよび/または染色体（すなわち、ヒト被験体については第1染色体～第22染色体、X染色体またはY染色体のうちの1つ）に特異的に割り当てられた核酸（例えばDNA）配列（すなわち読み取り）である。配列タグは、単一の参照ゲノムの部分（例えば、染色体）内で反復性であっても非反復性であってもよい。いくつかの実施形態では、反復配列タグをさらなる分析（例えば定量）から排除する。いくつかの実施形態では、読み取りは、参照ゲノムの部分に一意的または非一意的にマッピングすることができる。読み取りは、参照ゲノム内の単一の配列とアラインメントされている場合、「一意的にマッピングされた」とみなされる。読み取りは、参照ゲノム内の2つ以上の配列とアラインメントされている場合、「非一意的にマッピングされた」とみなされる。いくつかの実施形態では、非一意的にマッピングされた読み取りをさらなる分析（例えば定量）から排除する。ある特定の実施形態では、ある特定の小さな程度 mismatches (0 ~ 1) が参照ゲノムとマッピングされている個々の試料からの読み取りとの間に存在し得る一塩基多型の原因となることが許容され得る。いくつかの実施形態では、参照配列にマッピングされる読み取りにはいかなる程度 mismatches も許容されない。

10

【0120】

参照配列、または参照ゲノムは、多くの場合、一個体または多数の個体から集合させたまたは部分的に集合させたゲノム配列である。試料核酸が妊婦に由来するある特定の実施形態では、参照配列は、時には胎児、胎児の母親または胎児の父親由来ではなく、本明細書では、「外部参照」と称される。いくつかの実施形態では、母体参照を調製し、使用することができる。外部参照に基づいて妊婦由来の参照を調製する場合（「母体参照配列」）、多くの場合、胎児DNAを実質的に含有しない妊婦のDNAからの読み取りを外部参照配列にマッピングし、集合させる。ある特定の実施形態では、外部参照は妊婦と実質的に同じ民族性を有する1つまたは複数の個体のDNA由来である。母体参照配列は、母体のゲノムDNAを完全に包含しなくてよく（例えば、母体のゲノムDNAの約50%、60%、70%、80%、90%またはそれ以上を包含してよく）、母体参照は、母体のゲノムDNA配列と完全にマッチしなくてよい（例えば、母体参照配列は多数 mismatches を含んでよい）。

20

【0121】

ゲノムセクション

30

いくつかの実施形態では、マッピングされた配列読み取り（すなわち配列タグ）を種々のパラメータに応じて一緒に群分けし、特定のゲノムセクションに割り当てる。多くの場合、個々のマッピングされた配列読み取りを使用して、試料中に存在するゲノムセクションの量を同定することができる。いくつかの実施形態では、ゲノムセクションの量は、試料中のより大きな配列（例えば染色体）の量を示す可能性がある。「ゲノムセクション」という用語は、「配列ウィンドウ」、「セクション」、「ピン」、「遺伝子座」、「領域」、「区分」または「セグメント」と互換的に使用することもできる。いくつかの実施形態では、ゲノムセクションは、染色体全体、染色体の部分、多数の染色体の部分、多数の染色体、多数の染色体由来の部分、および/またはそれらの組合せである。いくつかの場合には、ゲノムセクションは、例えば、配列の長さまたは特定の1つまたは複数の特徴を含む1つまたは複数のパラメータに基づいて線引きされる。いくつかの実施形態では、ゲノムセクションは、ゲノム配列の特定の長さに基づく。いくつかの実施形態では、方法は、複数のゲノムセクションにマッピングされた多数の配列読み取りを分析することを含む。ゲノムセクションはほぼ同じ長さであってもよく、ゲノムセクションは異なる長さであってもよい。いくつかの実施形態では、ゲノムセクションは、約10キロベース (kb) ~ 約100 kb、約20 kb ~ 約80 kb、約30 kb ~ 約70 kb、約40 kb ~ 約60 kb、および時には約50 kbである。いくつかの実施形態では、ゲノムセクションは約10 kb ~ 約20 kbである。本明細書で考察されているゲノミックセクションは、連続したひと続きの配列に限定されない。したがって、ゲノムセクションは、連続した配列または連続していない配列で構成されていてよい。本明細書で考察されているゲノミック

40

50

セクションは、単一の染色体に限定されず、いくつかの実施形態では、個々の染色体を超越し得る。いくつかの場合には、ゲノミックセクションは、1つの染色体全体、2つの染色体全体、またはより多くの染色体全体にわたってよい。さらに、ゲノミックセクションは、多数の染色体の接合部または分離部にわたってよい。

【0122】

いくつかの実施形態では、ゲノムセクションは、例えば、遺伝的変異が評価される染色体（例えば第13染色体、第18染色体および/または第21染色体の異数性）などの対象の染色体内の特定の染色体セクションであってよい。ゲノムセクションは、病原体のゲノム（例えば細菌、真菌またはウイルスのゲノム）またはその断片であってよい。ゲノムセクションは、遺伝子、遺伝子断片、調節配列、イントロン、エクソンなどであってよい。

10

【0123】

いくつかの実施形態では、ゲノム（例えばヒトゲノム）を領域の情報量に基づいてゲノムセクションに分割する。生じたゲノム領域は、多数の染色体についての配列を含有してよく、かつ/または、多数の染色体の部分についての配列を含有してよい。いくつかの場合には、分割することにより、ゲノム全体にわたって類似した位置を排除し、独特の領域のみを保持することができる。排除された領域は、単一の染色体内にあってもよく、多数の染色体にわたってもよい。したがって、生じたゲノムは、より速いアラインメントのために切り詰められ、最適化されており、多くの場合、一意的に同定可能な配列に焦点を合わせることが可能になる。いくつかの場合には、分割することにより、類似した領域の重みを減らすことができる。ゲノムセクションの重みを減らすためのプロセスは下でさらに詳細に考察されている。いくつかの実施形態では、ゲノムを、染色体を超越する領域に分割することは、分類の状況において生じた情報取得に基づいてよい。例えば、情報量は、確認された正常な被験体および異常な被験体（例えば正倍数性被験体およびトリソミー被験体）の群間を区別するための特定の遺伝子位置の有意性を測定するp値プロファイルを使用して定量することができる。いくつかの実施形態では、ゲノムを、染色体を超越する領域に分割することは、例えば、タグをアラインメントする間のスピード/利便性、GC含量の高低、GC含量の均一性、反復配列の存在、配列含量の他の尺度（例えば個々のヌクレオチドの分率、ピリミジンまたはプリン分率、天然核酸と非天然核酸の分率、メチル化されたヌクレオチドの分率、およびCpG含量）、メチル化の状態、2重鎖融解温度、配列決定またはPCRに対する従順性、個々のピンに割り当てられる不確実性のレベル、および/または特定の特徴についての標的化検索などの任意の他の基準に基づいてよい。

20

30

【0124】

配列タグ密度

「配列タグ密度」とは、異なる試料を比較するため、およびその後の分析のために配列タグ密度を使用する定義済みのゲノムセクションについての配列タグまたは読み取りの値を指す。いくつかの実施形態では、配列タグの値は、正規化された配列タグの値である。配列タグ密度の値は、時には試料内で正規化し、時には試料の一群（例えば、フローレーションにおいて処理された試料、ライブラリー生成プレートにおいて調製された試料、段階分けプレートにおいて採取された試料など、およびそれらの組合せ）についての中央値に対して正規化する。

40

【0125】

いくつかの実施形態では、正規化は、各ゲノムセクション内に入るタグの数をカウントすること；各染色体について配列タグカウントの総計の中央値、最頻値、アベレージ、または中点値を得ること；常染色体の値の全ての中央値、最頻値、アベレージまたは中点値を得ること；およびこの値を異なる試料について得られた配列タグの総数の差異を説明するための正規化定数として使用することによって実施することができる。ある特定の実施形態では、正規化は、フローセル内の全ての試料について、各ゲノムセクション内に入るタグの数をカウントすること；フローセル内の全ての試料について各染色体についての配

50

列タグカウントの総計の中央値、最頻値、アベレージまたは中点値を得ること、フローセル内の全ての試料について常染色体の値の全ての中央値、最頻値、アベレージまたは中点値を得ること；およびこの値をフローセル内で並行して処理された異なる試料について得られた配列タグの総数の差異を説明するための正規化定数として使用することによって実施することができる。いくつかの実施形態では、正規化は、プレート（例えば、試薬プレート、マイクロウェルプレート）内で調製された全ての試料について、各ゲノムセクション内に入るタグの数をカウントすること；プレート内で調製された全ての試料について各染色体についての配列タグカウントの総計の中央値、最頻値、アベレージまたは中点値を得ること、プレート内で調製された全ての試料について常染色体の値の全ての中央値、最頻値、アベレージまたは中点値を得ること；およびこの値をプレート内で並行して処理された異なる試料について得られた配列タグの総数の差異を説明するための正規化定数として使用することによって実施することができる。

10

【0126】

配列タグ密度は、時には二染色体について約1である。配列タグ密度は、配列決定アーチファクト、とりわけG/Cの偏り、バッチプロセスの影響（例えば、試料の調製）などに応じて変動する可能性があり、外部標準または内部参照（例えば、いくつかの実施形態では、例えば、単一の染色体、全ての常染色体からの算出値、フローセル内で分析された全ての試料（単一の染色体または全ての常染色体）からの算出値、またはプレート内で処理され、1つまたは複数のフローセル内で分析された全ての試料からの算出値であってよい配列タグ（ゲノム配列）の実質的に全てに由来する）を使用することによって補正することができる。したがって、染色体または染色体領域の量不均衡は、検体の他のマッピング可能な配列決定されたタグの間での遺伝子座の百分率表示から推定することができる。したがって、特定の染色体または染色体領域の用量不均衡は、定量的に決定し、正規化することができる。配列タグ密度を正規化および定量するための方法は下でさらに詳細に考察されている。

20

【0127】

いくつかの実施形態では、配列読み取りの全てのうちのある割合は異数性に関与する染色体（例えば、第13染色体、第18染色体、第21染色体）からのものであり、他の配列読み取りは他の染色体からのものである。いくつかの実施形態では、異数性に関与する染色体（例えば、「標的染色体」：第21染色体）の他の染色体と比較した相対的なサイズを考慮に入れることにより、標的染色体に特異的な配列の基準範囲内に入る正規化された頻度を得ることができる。胎児が標的染色体に異数性を有する場合、標的染色体由来の配列の正規化された頻度は、非標的染色体由来の配列の正規化された頻度を統計学的に超え、したがって、異数性を検出することが可能になる。いくつかの実施形態では、正規化された頻度の変化の程度は、分析された試料中の胎児核酸の分画濃度に左右される。

30

【0128】

遺伝的変異の有無のアウトカムおよび決定

いくつかの遺伝的変異は医学的状态に関連付けられる。遺伝的変異は、多くの場合、遺伝的変異がない参照被験体に対して試験被験体のゲノムまたは遺伝情報の検出可能な変化をもたらす遺伝情報（例えば、染色体、染色体の部分、多型領域、転座した領域、変更されたヌクレオチド配列など、または前述のものの組合せ）の増加、減少および/または変更（例えば、重複、欠失、融合、挿入、突然変異、再編成、置換または異常なメチル化）を含む。遺伝的変異の有無は、当技術分野で公知であり、本明細書に記載の通り、ゲノミックセクション（例えば、ゲノムのピン）にマッピングされた配列読み取りを分析し、かつ/または操作することによって決定することができる。いくつかの実施形態では、非限定的な例が表1Aおよび1Bにおいて提供される公知の状態、症候群および/または異常の有無を、本明細書に記載の方法を利用して検出し、かつ/または決定することができる。

40

【0129】

カウントすること

50

いくつかの実施形態では、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分割された配列読み取りを定量して、各ゲノミックセクション（例えば、ピン、区分、ゲノムのセグメントなど）にマッピングされた読み取りの数を決定することができる。ある特定の実施形態では、マッピングされた配列読み取りの総数をマッピングされた配列読み取りの全てをカウントすることによって決定し、いくつかの実施形態では、マッピングされた配列読み取りの総数を、各ピンまたは区分にマッピングされたカウントを合計することによって決定する。いくつかの実施形態では、カウントすることを、読み取りをマッピングするプロセスにおいて実施する。ある特定の実施形態では、マッピングされた配列読み取りのサブセットをマッピングされた配列読み取りの所定のサブセットをカウントすることによって決定し、いくつかの実施形態では、マッピングされた配列読み取りの所定のサブセットを、所定のピンまたは区分のそれぞれにマッピングされたカウントを合計することによって決定する。いくつかの実施形態では、マッピングされた配列読み取りの所定のサブセットは、1～nの配列読み取りを含んでよく、nは、試験被験体試料、1つまたは複数の参照被験体試料、フローセル内で処理された全ての試料、または1つまたは複数のフローセルを使用した分析のためにプレート内で調製された全ての試料から生成された全ての配列読み取りの合計と等しい数を表す。試験被験体試料、1つまたは複数の参照被験体試料、フローセル内で処理された全ての試料、またはプレート内で調製された全ての試料についてマッピングされ、カウントされた配列読み取りは、時には試料カウントと称される。試料カウントは、時には、試料を単離した被験体（例えば、試験被験体試料カウント、参照被験体試料カウントなど）を参照することによってさらに区別される。

10

20

【0130】

いくつかの実施形態では、試験試料を参照試料としても使用する。試験試料を時には参照試料として使用し、遺伝的変異がない（例えば、1つまたは複数の選択されたゲノミックセクションにおいていかなる微小欠失、重複、異数性なども有さない）ことが既知の1つまたは複数の選択されたゲノミックセクション（例えば、第1のゲノミックセクション、第2のゲノミックセクション、第3のゲノミックセクション、5以上のゲノミックセクション、50以上のゲノミックセクション、500以上のゲノミックセクションなど）について、予測カウント中央値および/または予測カウント中央値の誘導値を決定する。遺伝的変異がない1つまたは複数のゲノミックセクションについての予測カウント中央値または予測カウント中央値の誘導値を使用して、試験試料の他の選択されたゲノミックセクション（例えば、参照試料セクションとして利用したものとは異なるゲノミックセクション）から得たカウントの統計的有意性を評価することができる。いくつかの実施形態では、中央絶対偏差も決定し、ある特定の実施形態では、試験試料の他の選択されたゲノミックセクションから得たカウントの統計的有意性を評価するために中央絶対偏差も使用する。

30

【0131】

ある特定の実施形態では、カウントを正規化する正規化プロセスは、予測カウントを使用することを含む。いくつかの実施形態では、試料カウントは、マッピングされた配列読み取りの所定のサブセットから得られる。ある特定の実施形態では、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た配列読み取りの所定のサブセットを選択することができる。いくつかの実施形態では、マッピングされた配列読み取りの所定のセットを比較するための基礎として利用し、これを「予測試料カウント」または「予測カウント」（総称して「予測カウント」）と称することができる。予測カウントは、多くの場合、一部において、1つまたは複数の選択されたゲノミックセクション（例えば、第1のゲノミックセクション、第2のゲノミックセクション、第3のゲノミックセクション、5つ以上のゲノミックセクション、50以上のゲノミックセクション、500以上のゲノミックセクションなど）についてのカウントを合計することによって得られる値である。時には、選択されたゲノミックセクションを、1つまたは複数の変数または特徴の有無に起因して参照、または比較するための基礎として選択する。時には、遺伝的変異（例えば、重複、欠失、挿入、胎児の異数性、トリソミー）がないゲノミックセクション（例えば、1つまたは

40

50

複数のゲノミックセクション、染色体、ゲノムまたはその一部)のカウントから予測カウントを決定する。ある特定の実施形態では、予測カウントは、遺伝的変異(例えば、重複、欠失、挿入、胎児の異数性、トリソミー)を含むゲノミックセクション(例えば、1つまたは複数のゲノミックセクション、染色体、ゲノムまたはその一部)のカウントに由来する。時には、ゲノミックセクションのいくつかは遺伝的変異を含み、ゲノミックセクションのいくつかは遺伝的変異を実質的に含まない1つまたは複数のゲノミックセクションのカウントから予測カウントを決定する。予測カウントは、多くの場合、少なくとも1つの共通の実験条件下で得た、試料の一群からのデータ(例えば、マッピングされた配列読み取りのカウント)を使用して決定する。時には、予測カウントを、カウントに本明細書に記載のまたはそうでなければ当技術分野で公知の1つまたは複数の数学的操作または統計学的操作を適用することによって決定する。そのような数学的操作または統計学的操作によってもたらされる予測カウントまたは予測試料カウント値の非限定的な例としては、中央値、平均値、最頻値、アベレージおよび/または中点、中央絶対偏差、R o u s s e e u wおよびC r o u xによって導入される中央絶対偏差の代替値、ブートストラップ推定値など、およびそれらの組合せが挙げられる。いくつかの実施形態では、予測カウントは、カウント(例えば、ゲノミックセクション、染色体、ゲノムまたはその一部のカウント)の中央値、最頻値、アベレージおよび/または中点である。予測カウントは時には、カウントまたは試料カウントの中央値、最頻値、アベレージおよび/または中点もしくは平均値である。カウントおよび予測カウントの非限定的な例としては、フィルタリングされたカウント、フィルタリングされた予測カウント、正規化されたカウント、正規化された予測カウント、補正されたカウントおよび補正された予測カウントが挙げられる。フィルタリングプロセス、正規化プロセスおよび補正プロセスは、本明細書においてさらに詳細に記載されている。

【0132】

いくつかの実施形態では、予測カウントの誘導値は、正規化および/または操作された(例えば、数学的に操作された)カウントに由来する予測カウントである。正規化および/または操作された(例えば、数学的に操作された)カウントは、時には、カウントの誘導値と称される。カウントの誘導値は時には、第1のゲノミックセクションからのカウントの表示であり、この表示は、多くの場合、第1のゲノミックセクションを含むゲノミックセクションからのカウントと比較した(例えば、それで割った)第1のゲノミックセクションからのカウントである。時には、カウントの誘導値は、パーセント表示または比率表示として表される。時には、表示は、多数のゲノミックセクションに対する1つのゲノミックセクションの表示であり、多数のゲノミックセクションは染色体の全部または一部に由来する。時には、表示は、より多数のゲノミックセクションに対する多数のゲノミックセクションの表示であり、多数のゲノミックセクションは染色体の全部または一部に由来し、より多数のゲノミックセクションは多数の染色体、実質的に全ての常染色体または実質的にゲノム全体に由来する。いくつかの実施形態では、カウントの誘導値を正規化する正規化プロセスは、予測カウントの誘導値を使用することを含む。カウントの誘導値から得た予測カウントは、本明細書では「予測カウントの誘導値」と称される。時には、予測カウントの誘導値は、カウントの表示(例えば、パーセント表示、染色体表示)に由来する予測カウントである。いくつかの実施形態では、予測カウントの誘導値は、カウント表示の中央値、最頻値、アベレージおよび/または中点(例えば、パーセント表示、染色体表示)である。ある特定の実施形態では、中央値は、中央値、平均値、最頻値、中点、アベレージなどである。

【0133】

時には、カウント、予測カウントまたは予測カウントの誘導値について変動性の推定値を決定する。変動性の推定値の非限定的な例としては、カウント、予測カウントまたは予測カウントの誘導値の中央絶対偏差(MAD); R o u s s e e u wおよびC r o u xによって導入されるMADの代替値;ブートストラップ推定値;カウント、予測カウントまたは予測カウントの誘導値の標準偏差など、およびそれらの組合せが挙げられる。変動性

10

20

30

40

50

の推定値は、時には正規化された試料カウントを得るための正規化プロセスにおいて利用される。

【 0 1 3 4 】

ある特定の実施形態では、正規化された試料カウントを得るための正規化プロセスは、第1のゲノムセクションについてのカウントから予測カウントを引き算し、それにより減算値を生成し、減算値をカウントまたは予測カウントの変動性の推定値で割ることを含む。カウントまたは予測カウントの変動性の非限定的な例は、カウントまたは予測カウントの中央絶対偏差 (MAD)、RousseeuwおよびCrouxによって導入されるMADの代替値、またはブートストラップ推定値である。いくつかの実施形態では、正規化された試料カウントを得るための正規化プロセスは、第1のゲノムセクションカウント表示から、予測された第1のゲノムセクションカウント表示を引き算し、それにより減算値を生成し、減算値を第1のゲノムセクションカウント表示または予測された第1のゲノムセクションカウント表示の変動性の推定値で割ることを含む。カウント表示または予測カウント表示の変動性の非限定的な例は、カウント表示または予測カウント表示の中央絶対偏差 (MAD)、RousseeuwおよびCrouxによって導入されるMADの代替値、またはブートストラップ推定値である。いくつかの実施形態では、予測カウントは、第1のゲノムセクションのカウントの中央値、最頻値、アベレージ、平均値および/または中点であり、時には、予測カウント表示は、第1のゲノミックセクションのカウント表示の中央値、平均値、最頻値、アベレージおよび/または中点である。

【 0 1 3 5 】

いくつかの実施形態では、予測カウント、予測カウントの誘導値 (例えば、予測カウント表示)、またはカウント、カウントの誘導値、予測カウントまたは予測カウントの誘導値の変動性の推定値を、1つまたは複数の共通の実験条件下で取得した試料データに従ってそれぞれ独立に決定する。時には、変動性の推定値を、1つまたは複数の共通の実験条件から生成された試料データについて得、時には、変動性の推定値を、1つまたは複数の共通の実験条件から生成されたものではない試料データについて得、時には、予測カウントを、1つまたは複数の共通の実験条件から生成されたものではない試料データについて得、時には、変動性の推定値および予測カウントを、1つまたは複数の共通の実験条件から生成された試料データについて得る。時には、予測カウントの誘導値 (例えば、予測カウント表示) の変動性の推定値を、1つまたは複数の共通の実験条件から生成された試料データについて得、時には、予測カウントの誘導値 (例えば、予測カウント表示) の変動性の推定値を、1つまたは複数の共通の実験条件から生成されたものではない試料データについて得、時には、予測カウントの誘導値 (例えば、予測カウント表示) を、1つまたは複数の共通の実験条件から生成された試料データについて得、時には、予測カウントの誘導値 (例えば、予測カウント表示) の変動性の推定値および予測カウントの誘導値 (例えば、予測カウント表示) を、1つまたは複数の共通の実験条件から生成された試料データについて得る。

【 0 1 3 6 】

いくつかの実施形態では、予測カウントまたは予測カウントの誘導値 (例えば、予測カウント表示) を1つまたは複数の共通の実験条件下で取得した試料データを使用して決定し、カウント、カウントの誘導値、予測カウントまたは予測カウントの誘導値の変動性の推定値を共通の実験条件下で取得したのものではない試料データを使用して決定する。ある特定の実施形態では、カウント、カウントの誘導値、予測カウントまたは予測カウントの誘導値の変動性の推定値を、第1番の試料について取得し、共通の実験条件下で取得したものではない試料データを使用して決定し、予測カウントまたは予測カウントの誘導値 (例えば、予測カウント表示) を、1つまたは複数の共通の実験条件下で取得し、第1番の試料よりも少ない第2番の試料について取得した試料データを使用して決定する。第2番

10

20

30

40

50

の試料は、時には、第1番の試料を取得した時間枠よりも短い時間枠で取得する。

【0137】

1つまたは複数の共通の実験条件下で取得した試料データは、時には、1～約5つの共通の実験条件（例えば、1つ、2つ、3つ、4つまたは5つの共通の実験条件）の下で取得する。共通の実験条件の非限定的な例としては、フローセル内のチャンネル、フローセルユニット、コンテナに共通のフローセル、ロットまたは製造の連続運転に共通のフローセル；試薬プレートユニット、コンテナに共通の試薬プレート、ロットまたは製造の連続運転に共通の試薬プレート；オペレーター；計器（例えば、シーケンシング計器）；湿度、温度；同定タグ指標など、およびそれらの組合せが挙げられる。試薬プレートは、時には、核酸ライブラリー調製および/または核酸の配列決定のために利用される。

10

【0138】

配列読み取りを定量することまたはカウントすることは、これだけに限定されないが、手動のカウント方法および自動化されたカウント方法を含めた任意の適切な様式で実施することができる。いくつかの実施形態では、自動化されたカウント方法は、各染色体および/または1つまたは複数の選択されたゲノミックセクションにマッピングされる配列読み取りまたは配列タグの数を決定またはカウントするソフトウェアにおいて具体化することができる。ソフトウェアは、一般に、コンピュータによって遂行される場合、本明細書に記載の通りコンピュータ操作を実施するコンピュータ可読プログラム命令である。

【0139】

試験被験体および/または参照被験体に由来する試料についての各ピンにマッピングされた配列読み取りの数および配列読み取りの総数をさらに解析および処理して、遺伝的変異の有無を決定するアウトカムをもたらすことができる。カウントされ、マッピングされた配列読み取りは、時には、「データ」または「データセット」と称される。いくつかの実施形態では、データまたはデータセットは、1つまたは複数の特徴または変数（例えば、配列に基づくもの〔例えば、GC含量、特定のヌクレオチド配列など〕、機能に特異的なもの〔例えば、発現された遺伝子、がん遺伝子など〕、位置に基づくもの〔ゲノム特異的、染色体特異的、ゲノミックセクションまたはピン特異的〕、実験条件に基づくもの〔例えば、指標に基づくもの、フローセルに基づくもの、プレートに基づくもの〕など、およびそれらの組合せ）によって特徴付けることができる。ある特定の形態では、データまたはデータセットを、1つまたは複数の特徴または変数（例えば、胎児分率および母体の年齢；胎児分率および地理的位置；パーセント第21染色体表示およびフローセル数；第21染色体のzスコアおよび母体の体重；第21染色体のzスコアおよび妊娠期間など）に基づいて2つ以上の次元を有するマトリックスに組織化および/または層別化することができる。マトリックスへ組織化および/または層別化されたデータは、任意の適切な特徴または変数を使用して組織化および/または層別化されてもよい。マトリックス内のデータの非限定的な例としては、母体の年齢、母体の倍数性、および胎児の寄与によって組織化されたデータが挙げられる。特徴または変数を用いて層別化されるデータの非限定的な例は、図4～45に示されている。ある特定の形態では、1つまたは複数の特徴または変数によって特徴付けられるデータセットを、時には、カウントした後に処理する。

20

30

40

【0140】

高度 (e l e v a t i o n)

いくつかの実施形態では、値は、高度（例えば、数）に帰する。高度は、適切な方法、操作または数学的プロセスによって決定することができる（例えば、処理された高度）。高度は、多くの場合、ゲノミックセクション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、またはそれに由来する。時には、ゲノミックセクションの高度は、ゲノミックセクションにマッピングされたカウントの総数（例えば、正規化されたカウント）と実質的に等しい。多くの場合、高度は、当技術分野で公知の適切な方法、操作または数学的プロセスによって処理、変換または操作されたカウントから決定する。時には、高度は処理されたカウントに由来し、処理されたカウントの非限定的な例としては

50

、重み付けされたカウント、除去されたカウント、フィルタリングされたカウント、正規化されたカウント、補正されたカウント、平均されたカウント、平均として導かれたカウント（例えば、平均高度）、足し算されたカウント、引き算されたカウント、変換されたカウントまたはその組合せが挙げられる。時には、高度は、正規化されたカウント（例えば、ゲノミックセクションの正規化されたカウント）を含む。高度は、適切なプロセスによって正規化されたカウントについてののものであってよく、その非限定的な例としては、ビンに関する（bin-wise）正規化、GC含量による正規化、線形最小二乗回帰および非線形最小二乗回帰、GC LOESS、LOWESS、PERUN、RM、GCRM、cQnなど、および/またはそれらの組合せが挙げられる。高度は、正規化されたカウントまたはカウントの相対量を含んでよい。時には、高度は、平均された2つ以上のゲノミックセクションのカウントまたは正規化されたカウントについてののものであってよく、高度は、アベレージ高度と称される。時には、高度は、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ゲノミックセクションのセットについてのものであり、平均高度と称される。時には、高度は、生のカウントおよび/またはフィルタリングされたカウントを含むゲノミックセクションについて誘導される。いくつかの実施形態では、高度は、生のカウントに基づく。時には、高度は、不確実性値を伴う。ゲノミックセクションについての高度は、時には、「ゲノミックセクションの高度」と称され、これは、本明細書では「ゲノミックセクションレベル」と同義である。

10

【0141】

2つ以上の高度（例えば、プロファイルにおける2つ以上の高度）について正規化されたまたは正規化されていないカウントは、時には、高度に応じて数学的に操作することができる（例えば、足し算すること、掛け算すること、平均すること、正規化することなど、またはその組合せ）。例えば、2つ以上の高度について正規化されたまたは正規化されていないカウントは、プロファイルにおける高度の1つ、いくつか、または全部に応じて正規化することができる。時には、プロファイルにおける全ての高度の正規化されたまたは正規化されていないカウントを、プロファイルにおける1つの高度に応じて正規化する。時には、プロファイルにおける第1の高度の正規化されたまたは正規化されていないカウントをプロファイルにおける第2の高度の正規化されたまたは正規化されていないカウントに応じて正規化する。

20

【0142】

高度（例えば、第1の高度、第2の高度）の非限定的な例は、処理されたカウントを含むゲノミックセクションのセットについての高度、カウントの平均値、中央値、最頻値、中点またはアベレージを含むゲノミックセクションのセットについての高度、正規化されたカウントを含むゲノミックセクションのセットについての高度など、またはそれらの任意の組合せである。いくつかの実施形態では、プロファイルにおける第1の高度および第2の高度は、同じ染色体にマッピングされたゲノミックセクションのカウントに由来する。いくつかの実施形態では、プロファイルにおける第1の高度および第2の高度は、異なる染色体にマッピングされたゲノミックセクションのカウントに由来する。

30

【0143】

いくつかの実施形態では、1つまたは複数のゲノミックセクションにマッピングされた、正規化されたまたは正規化されていないカウントから高度を決定する。いくつかの実施形態では、2つ以上のゲノミックセクションにマッピングされた、正規化されたまたは正規化されていないカウントから高度を決定し、各ゲノミックセクションについて正規化されたカウントは多くの場合ほとんど同じである。高度についてのゲノミックセクションのセットにおけるカウント（例えば、正規化されたカウント）は変動し得る。高度についてのゲノミックセクションのセットでは、そのセットの他のゲノミックセクションにおけるものとは有意に異なるカウントを有する1つまたは複数のゲノミックセクションが存在し得る（例えば、ピークおよび/またはディップ）。任意の適切な数のゲノミックセクションに関連付けられる任意の適切な数の正規化されたまたは正規化されていないカウントにより高度を定義することができる。

40

50

【 0 1 4 4 】

時には、1つまたは複数の高度は、ゲノムのゲノミックセクションの全部またはいくつかの正規化されたまたは正規化されていないカウントから決定することができる。多くの場合、高度は、染色体またはそのセグメントの正規化されたまたは正規化されていないカウントの全部またはいくつかから決定することができる。時には、2つ以上のゲノミックセクション（例えば、ゲノミックセクションのセット）に由来する2つ以上のカウントにより、高度が決定される。時には、2つ以上のカウント（例えば、2つ以上のゲノミックセクションからのカウント）により、高度が決定される。いくつかの実施形態では、2～約100,000のゲノミックセクションからのカウントにより、高度が決定される。いくつかの実施形態では、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のゲノミックセクションからのカウントにより、高度が決定される。いくつかの実施形態では、約10～約50のゲノミックセクションからのカウントにより、高度が決定される。いくつかの実施形態では、約20～約40以上のゲノミックセクションからのカウントにより、高度が決定される。いくつかの実施形態では、高度は、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、またはそれ以上のゲノミックセクションからのカウントを含む。いくつかの実施形態では、高度は、ゲノミックセクションのセット（例えば、参照ゲノムのゲノミックセクションのセット、染色体のゲノミックセクションのセットまたは染色体のセグメントのゲノミックセクションのセット）に対応する。

【 0 1 4 5 】

いくつかの実施形態では、連続したゲノミックセクションの正規化されたまたは正規化されていないカウントについて高度を決定する。時には、連続したゲノミックセクション（例えば、ゲノミックセクションのセット）とは、ゲノムの隣接するセグメントまたは染色体または遺伝子の隣接するセグメントを表す。例えば、2つ以上の連続したゲノミックセクションは、ゲノミックセクションを端から端までマージすることによってアラインメントした場合、各ゲノミックセクションよりも長いDNA配列の配列集合を表し得る。例えば、2つ以上の連続したゲノミックセクションはインタクトなゲノム、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し得る。時には、連続したゲノミックセクションおよび/または連続していないゲノミックセクションの収集物（例えば、セット）から高度を決定する。

【 0 1 4 6 】

有意に異なる高度

いくつかの実施形態では、正規化されたカウントのプロファイルは、プロファイル内に別の高度（例えば、第2の高度）とは有意に異なる高度（例えば、第1の高度）を含む。第1の高度は第2の高度よりも高くても低くてもよい。いくつかの実施形態では、第1の高度は、コピー数多型（例えば、母体のコピー数多型、胎児のコピー数多型、または母体のコピー数多型および胎児のコピー数多型）を含む1つまたは複数の読み取りを含むゲノミックセクションのセットについてのものであり、第2の高度は、コピー数多型を実質的に有さない読み取りを含むゲノミックセクションのセットについてのものである。いくつかの実施形態では、有意に異なるとは、観察可能な差異を指す。時には、有意に異なるとは、統計学的に異なることまたは統計的有意差を指す。統計的有意差は、時には、観察された差異の統計学的評価である。統計的有意差は、当技術分野における適切な方法によって評価することができる。任意の適切な閾値または範囲を使用して、2つの高度が有意に異なることを決定することができる。いくつかの場合には、約0.01パーセント以上異なる（例えば、高度値の一方またはいずれかの0.01パーセント）2つの高度（例えば、平均高度）は有意に異なる。時には、約0.1パーセント以上異なる2つの高度（例え

10

20

30

40

50

ば、平均高度)は有意に異なる。いくつかの場合には、約0.5パーセント以上異なる2つの高度(例えば、平均高度)は有意に異なる。時には、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つの高度(例えば、平均高度)は有意に異なる。時には、2つの高度(例えば、平均高度)は有意に異なり、いずれの高度にもオーバーラップは存在せず、かつ/または一方の高度または両方の高度について算出された不確実性値によって定義される範囲内にオーバーラップは存在しない。いくつかの場合には、不確実性値はシグマとして表される標準偏差である。時には、2つの高度(例えば、平均高度)は有意に異なり、これらは約1倍以上不確実性値が異なる(例えば、1シグマ)。時には、2つの高度(例えば、平均高度)は有意に異なり、これらは約2倍以上(例えば、2シグマ)、約3倍以上の、約4倍以上、約5倍以上、約6倍以上、約7倍以上、約8倍以上、約9倍以上、または約10倍以上不確実性値が異なる。時には、2つの高度(例えば、平均高度)は、約1.1倍、1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2.0倍、2.1倍、2.2倍、2.3倍、2.4倍、2.5倍、2.6倍、2.7倍、2.8倍、2.9倍、3.0倍、3.1倍、3.2倍、3.3倍、3.4倍、3.5倍、3.6倍、3.7倍、3.8倍、3.9倍、または4.0倍以上不確実性値が異なる場合に、有意に異なる。いくつかの実施形態では、2つの高度の間の差異が増加するに従い信頼水準が増加する。いくつかの場合には、2つの高度の間の差異が減少するに従い、かつ/または不確実性値が増加するに従い、信頼水準が減少する。例えば、時には、信頼水準は高度と標準偏差(例えば、MAD)の間の差異の比率と共に増加する。

10

20

【0147】

いくつかの実施形態では、ゲノミックセクションの第1のセットは、多くの場合、ゲノミックセクションの第2のセットとは異なる(例えば、それとオーバーラップしない)ゲノミックセクションを含む。例えば、時には、正規化されたカウントの第1の高度は、プロファイル内の正規化されたカウントの第2の高度とは有意に異なり、第1の高度はゲノミックセクションの第1のセットについてのものであり、第2の高度はゲノミックセクションの第2のセットについてのものであり、ゲノミックセクションの第1のセットおよび第2のセットにおいてオーバーラップしない。いくつかの場合には、ゲノミックセクションの第1のセットは、それぞれ第1の高度および第2の高度が決定されるゲノミックセクションの第2のセットのサブセットではない。時には、ゲノミックセクションの第1のセットは、それぞれ第1の高度および第2の高度が決定されるゲノミックセクションの第2のセットとは異なり、かつ/または別個のものである。

30

【0148】

時には、ゲノミックセクションの第1のセットは、プロファイル内のゲノミックセクションの第2のセットのサブセットである。例えば、時には、プロファイル内のゲノミックセクションの第2のセットについての正規化されたカウントの第2の高度は、プロファイル内の第1の高度についてのゲノミックセクションの第1のセットの正規化されたカウントを含み、ゲノミックセクションの第1のセットは、プロファイル内のゲノミックセクションの第2のセットのサブセットである。時には、アベレージ、平均値、中央値、最頻値または中点の高度は第2の高度に由来し、第2の高度は第1の高度を含む。時には、第2の高度は染色体全体を表すゲノミックセクションの第2のセットを含み、第1の高度はゲノミックセクションの第1のセットを含み、第1のセットは、ゲノミックセクションの第2のセットのサブセットであり、第1の高度により、染色体に存在する母体のコピー数多型、胎児のコピー数多型、または母体のコピー数多型および胎児のコピー数多型が表される。

40

【0149】

いくつかの実施形態では、第2の高度の値は、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、アベレージ最頻値、中点または中央値に第1の高度よりも近い。いくつかの実施形態では、第2の高度は、染色体、染色体の部分またはそのセグ

50

メントの平均高度である。いくつかの実施形態では、第1の高度は、染色体またはそのセグメントを表す優勢な高度（例えば、第2の高度）とは有意に異なる。プロファイルは、第2の高度と有意に異なる多数の第1の高度を含んでよく、各第1の高度はそれぞれ独立に、第2の高度よりも高くても低くてもよい。いくつかの実施形態では、第1の高度および第2の高度は同じ染色体に由来し、第1の高度は第2の高度よりも高いか低いかであり、第2の高度は染色体の優勢な高度である。時には、第1の高度および第2の高度は同じ染色体に由来し、第1の高度により、コピー数多型（例えば、母体および/または胎児のコピー数多型、欠失、挿入、重複）が示され、第2の高度は、染色体またはそのセグメントについてのゲノミックセクションの平均高度または優勢な高度である。

【0150】

いくつかの場合には、第2の高度についてのゲノミックセクションの第2のセット内の読み取りは、遺伝的変異（例えば、コピー数多型、母体および/または胎児のコピー数多型）を実質的に含まない。多くの場合、第2の高度についてのゲノミックセクションの第2のセットは、いくらかの変動性（例えば、高度の変動性、ゲノミックセクションについてのカウントの変動性）を含む。時には、コピー数多型と実質的に関連しない、高度についてのゲノミックセクションのセット内の1つまたは複数のゲノミックセクションは、母体のゲノムおよび/または胎児のゲノムに存在するコピー数多型を有する1つまたは複数の読み取りを含む。例えば、時には、ゲノミックセクションのセットは、小さな染色体のセグメント（例えば、10ゲノミックセクション未満）に存在するコピー数多型を含み、ゲノミックセクションのセットは、コピー数多型と実質的に関連しない高度についてのもの

である。したがって、コピー数多型を実質的に含まないゲノミックセクションのセットは、それでも、ある高度の約10、9、8、7、6、5、4、3、2または1未満のゲノミックセクションに存在するコピー数多型を含んでよい。

【0151】

時には、第1の高度はゲノミックセクションの第1のセットについてのものであり、第2の高度はゲノミックセクションの第2のセットについてのものであり、ゲノミックセクションの第1のセットおよびゲノミックセクションの第2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近接する）。時には、ゲノミックセクションの第1のセットおよびゲノミックセクションの第2のセットは連続していない。

【0152】

胎児核酸と母体核酸の混合物からの比較的短い配列読み取りを利用して、高度および/またはプロファイルに変換することができるカウントをもたらすことができる。カウント、高度およびプロファイルは、電子形態または有形形態で示すことができ、可視化することができる。ゲノミックセクションにマッピングされたカウント（例えば、高度および/またはプロファイルとして示される）により、胎児および/または妊婦に存在する胎児のゲノム、染色体、または染色体の部分もしくはセグメントおよび/もしくは母体のゲノム、染色体、または染色体の部分もしくはセグメントの視覚的な表示をもたらすことができる。

【0153】

データ処理

カウントされた、マッピングされた配列読み取りは、本明細書では、データが操作されていないカウント（例えば、生のカウント）を表すので、生のデータと称される。いくつかの実施形態では、データセット内の配列読み取りデータをさらに補正および/または処理し（例えば、数学的かつ/または統計学的に操作し）、かつ/または提示してアウトカムをもたらすことを容易にすることができる。補正された配列読み取りデータは、多くの場合、配列読み取り、データセット内のデータ、および/または試料核酸の一部または全部を操作することによってもたらされる。任意の適切な操作を使用して配列読み取り、データセット内のデータおよび/または試料核酸の一部または全部を補正することができる。いくつかの実施形態では、配列読み取り、データセット内のデータおよび/または試料核酸に対する補正は、フィルタリング（例えば、選択された特徴または変数に基づいてデ

10

20

30

40

50

ータの一部を除去すること；反復配列を除去すること、情報価値のないピンまたはゼロカウント中央値を有するピンを除去すること）、補正すること（例えば、データの一部または全部を推定量に基づいて再尺度化および／または再重み付けすること；試料カウントをG/C含量に基づいて再重み付けすること、データの一部または全部を胎児分率に基づいて再尺度化および／または再重み付けすること）、1つまたは複数の推定量または統計学的操作を用いて正規化すること（例えば、所与のフローセル内の全てのデータをフローセル内の全てのデータの中央絶対偏差に対して正規化すること）などから選択されるプロセスである。いくつかの実施形態では、推定量はロバストな推定量である。ある特定の実施形態では、配列読み取りデータの一部を補正および／または処理し、いくつかの実施形態では、配列読み取りデータの全てを補正および／または処理する。

10

【0154】

補正または処理された配列読み取り、データセット内のデータおよび／または試料核酸は、時には、誘導値（例えば、カウントの誘導値、データ誘導値、配列読み取りの誘導値など）と称される。カウント、データまたは配列読み取りの誘導値は、多くの場合、カウント、データまたは配列読み取りに対して1つまたは複数の数学的操作および／または統計学的操作を使用することによって生成する。本明細書に記載のまたは当技術分野で公知の任意の適切な数学的操作および／または統計学的操作を使用して、カウント、データ、または配列読み取りの誘導値を生成することができる。カウント、データ、または配列読み取りをフィルタリング、補正、正規化または操作して誘導値を生成するために利用することができる数学的操作および／または統計学的操作の非限定的な例としては、アベレー

20

【0155】

ある特定の実施形態では、より大きなデータセットを含めたデータセットは、さらなる解析を容易にするために前処理することが有効である。データセットの前処理は、時には、冗長であり、かつ／または情報価値のないゲノミックセクションまたはピン（例えば、情報価値のないデータを有するピン、マッピングされた冗長な読み取り、ゼロカウント中央値を有するゲノミックセクションまたはピン、過大表示または過小表示された配列〔例えば、G/C配列〕、反復配列）を除去することを伴う。理論によって限定されることなく、データ処理および／または前処理により、(i)ノイズの多いデータを除去することができ、(ii)情報価値のないデータを除去することができ、(iii)冗長なデータを除去することができ、(iv)より大きなデータセットの複雑さを低下させることができ、(v)実験条件に誘導されるデータの変動性を低下させるまたは排除することができ、(vi)データセット内のデータの一部または全部を再尺度化および／または再重み付けすることができ、かつ／または(vii)ある形態から1つまたは複数の他の形態へのデータの変換を容易にすることができる。「前処理」および「処理」という用語は、データまたはデータセットに関して利用される場合、本明細書では総称して「処理」と称される。処理により、データをさらなる解析により適するものにすることができ、いくつかの実施形態ではアウトカムを生成することができる。

30

40

【0156】

ノイズの多いデータとは、多くの場合、(a)解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有するデータ、(b)有意な標準偏差（例えば、3標準偏差を超える）を有するデータ、(c)有意な平均値の標準誤差を有するデータなど、および前述のものの組合せである。ノイズの多いデータは、時には、出発材料（例えば、核酸試料）の量および／または品質に起因して生じ、時には、配列読み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プロセスの一部として生じる。ある特定の実施形態では、PCRに基づく方法を用いて調製した場合、ノイズは、過大表示されたある特定の配列に由来する。本明細書に記載の方法により、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、したがって、もたらされるアウトカムに対するノイズの多いデータの影響

50

を低下させることができる。

【0157】

情報価値のないデータ、情報価値のないピン、および情報価値のないゲノミックセクションとは、多くの場合、所定のカットオフ閾値と有意に異なるまたは所定の値のカットオフ範囲の範囲外の数値を有するゲノミックセクション、またはそれに由来するデータである。いくつかの実施形態では、カットオフ閾値または値の範囲は、多くの場合、配列読み取りデータ（例えば、参照、被験体、フローセルおよび/またはプレートからのもの）を数学的かつ/または統計学的に操作することによって算出され、ある特定の実施形態では、閾値カットオフ値または値の範囲を生成するために操作された配列読み取りデータは配列読み取りデータ（例えば、参照、被験体、フローセルおよび/またはプレートからのもの）である。いくつかの実施形態では、閾値カットオフ値は、生のまたは正規化されたカウントプロファイルの標準偏差および/または中央絶対偏差（例えば、MADまたはRousseeuwおよびCrouxによって導入されるMADの代替値、またはブートストラップ推定値）を算出し、プロファイルについての標準偏差に、カットオフ閾値として選択された標準偏差の数を表す定数を掛け（例えば、3標準偏差については3を掛ける）、それにより、不確実性についての値を生成することによって得られる。ある特定の実施形態では、算出された不確実性の閾値カットオフ値を超える、または閾値カットオフ値の範囲の外側のゲノミックセクションの一部または全部を、正規化プロセスの一部として、その前に、またはその後に除去する。いくつかの実施形態では、算出された不確実性の閾値カットオフ値を超える、または閾値カットオフ値または生のデータ点の範囲の外側のゲノミックセクションの一部または全部について、正規化または分類プロセスの一部として、またはその前に重み付けする。重み付けの例は本明細書に記載されている。いくつかの実施形態では、冗長なデータ、およびマッピングされた冗長な読み取りとは、すでに遺伝子位置（例えば、塩基位置）に割り当てられ、かつ/またはゲノミックセクションについてカウントされていると同定された試料由来の配列読み取りを指す。

【0158】

実験条件

試料は、時には、共通の実験条件の影響を受ける。実質的に同じ時間にまたは実質的に同じ条件および/または試薬を使用して処理された試料は、時には、異なる時間におよび/または同時に異なる条件および/または試薬を使用して処理された他の試料と比較して、同様の実験条件（例えば、共通の実験条件）に誘導されるデータの変動性を示す。多くの場合、実験手順の間の任意の所与の時間に調製、処理および/または分析することができる試料の数を限定する実施上の考慮すべき問題が存在する。ある特定の実施形態では、アウトカムを生成するために試料を原料から処理するための時間枠は、時には数日、数週間、さらには数ヶ月である。単離と最後の分析の間の時間に起因して、大量の試料を分析するハイスループットな実験により、時には、バッチの影響または実験条件に誘導されるデータの変動性が生じる。実験条件に誘導されるデータの変動性は、多くの場合、試料の単離、保管、調製および/または分析の結果であるあらゆるデータの変動性を含む。実験条件に誘導される変動性の非限定的な例としては、配列の過大表示または過小表示；ノイズの多いデータ；偽データ点または外れ値データ点、試薬の影響、人員の影響、研究所条件の影響などを含む、フローセルに基づく変動性および/またはプレートに基づく変動性が挙げられる。実験条件に誘導される変動性は、時には、データセット内の試料の亜集団に対して生じる（例えば、バッチの影響）。バッチは、多くの場合、実質的に同じ試薬を使用して処理された試料、同じ試料調製プレート（例えば、試料の調製；核酸の単離のために使用するマイクロウェルプレート）において処理された試料、同じ段階分けプレート（例えば、試料をフローセルにローディングする前に組織化するために使用するマイクロウェルプレート）において分析のために段階分けされた試料、実質的に同じ時間に処理された試料、同じ人員によって処理された試料、および/または実質的に同じ実験条件（例えば、温度、CO₂レベル、オゾンレベルなど、またはそれらの組合せ）の下で処理された試料である。実験条件バッチの影響は、時には、同じフローセルで分析され、同じ試薬プ

10

20

30

40

50

レートまたはマイクロウェルプレートにおいて調製され、かつ/または、同じ試薬プレートまたはマイクロウェルプレートにおいて分析のために段階分けされた(例えば、配列決定のために核酸ライブラリーを調製すること)試料に影響を及ぼす。追加的な変動性の原因は、単離された核酸の質、単離された核酸の量、核酸を単離した後保管するまでの時間、保管時間、保管温度など、およびそれらの組合せを含み得る。バッチ(例えば、同時に、かつ/または同じ試薬および/または実験条件を使用して処理されたデータセット内の試料の亜集団)内のデータ点の変動性は、時には、バッチ間で見られるデータ点の変動性を超える。このデータの変動性は、時には、その大きさがデータセット内の他のデータのいくつか、または全ての解釈に影響を及ぼす可能性がある偽データまたは外れ値データを含む。データセットの一部または全部は、本明細書に記載されており、当技術分野で公知のデータ処理ステップを用いて、実験条件について補正することができる;例えば、フローセル内で分析された、またはマイクロウェルプレートにおいて処理された全ての試料について算出された中央絶対偏差に対する正規化。

10

【0159】

実験条件に誘導される変動性は、数週間から数ヶ月または数年(例えば、1週間、1~4週間、1ヶ月、1~3ヶ月、1~6ヶ月)の期間にわたって得られるデータについて観察することができる。時には、1つまたは複数の実験条件が共通の実験条件である多数の実験を数週間から数ヶ月の期間にわたって行う。共通の実験条件の非限定的な例としては、同じ計器、機械またはその一部(例えば、シークエンサー、液体取扱いデバイス、分光光度計、光電池など)の使用、同じデバイス(例えば、フローセル、フローセルチャンネル、プレート、チップなど、またはその一部)の使用、同じプロトコル(操作手順、標準操作手順、レシピ、方法および/または条件(例えば、インキュベーションの時間、温度、圧力、湿度、体積、濃度)の使用、同じオペレーター(例えば、技師、科学者)、ならびに同じ試薬(例えば、ヌクレオチド、オリゴヌクレオチド、配列タグ、同定タグ指標、試料(例えば、c c f DNA試料)、タンパク質(例えば、酵素、緩衝液、塩、水)など)が挙げられる。

20

【0160】

同じデバイス、装置または試薬の使用は、同じ製造者からのデバイス、装置、試薬またはその一部、同じ製造の連続運転、同じロット(例えば、同じプラント、製造者、製造の連続運転または位置に由来する材料、同じ日付が標識された収集物)、同じクリーニングサイクル、同じ調製プロトコル、同じコンテナ(袋、箱、パッケージ、保管ビン、パレット、トレーラー)、同じ輸送(例えば、同じ納期、同じ注文、同じ送り状を有する)、同じ製造プラント、同じ組立てラインなど、またはそれらの組合せを含んでよい。いくつかの実施形態では、同じオペレーターの使用とは、機械、装置またはデバイスの1つまたは複数のオペレーターが同じであることを意味する。

30

【0161】

データセット内のデータを補正することにより、多くの場合、データセットに対する外れ値の影響を低下させるまたは排除すること、データを再尺度化または再重み付けして、アウトカムをもたらすことを容易にすること、および/またはデータセットの複雑さおよび/または次元性を低下させることができる。ある特定の実施形態では、1つまたは複数の共通の実験条件(例えば、使用した試薬、使用したフローセル、使用したプレート、試料を処理した人員、使用した指標配列など、またはそれらの組合せ)に応じてデータを選別(例えば、層別化、組織化)することができる。いくつかの実施形態では、1つまたは複数の共通の実験条件に応じてデータを正規化または補正することができる。

40

【0162】

ロバストな推定量を使用してデータを再尺度化または再重み付けすることができる。ロバストな推定量は、多くの場合、その大きさがアウトカムをもたらすこと(例えば、遺伝的変異の有無を決定すること)に影響を及ぼす可能性がある偽データまたは外れ値データの影響を最小限にするまたは排除する数学的操作または統計学的操作である。任意の適切なロバストな推定量を使用してデータセットを補正することができる。いくつかの実施形

50

態では、ロバストな推定量は、尺度のロバストな推定量であり（例えば、変動性；中央絶対偏差 [MAD] もしくは R o u s s e e u w および C r o u x によって導入される MAD の代替値、またはブートストラップ推定値と同様である、および / またはそれを含む）、ある特定の実施形態では、ロバストな推定量は、位置のロバストな推定量である（例えば、予測値；アベレージまたは中央値と同様である）。尺度および位置のロバストな推定量の非限定的な例は実施例 2 に記載されており、また当技術分野で公知である（例えば、中央値、ANOVA など）。いくつかの実施形態では、データセット内の一部、または全てのデータを、ロバストな推定量を使用して得られる予測カウントまたは予測カウントの誘導値を使用して補正することができる。いくつかの実施形態では、予測カウントは、参照または参照試料（例えば、公知の正倍数性試料）から得られるカウントである。

10

【 0 1 6 3 】

本明細書に記載のデータセットを補正および / または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。データセットを補正するために使用することができる手順の非限定的な例としては、実験条件に基づく補正（例えば、プレートに基づく正規化、フローセルに基づく正規化 [例え、フローセルに基づく中央値の比較]、反復マスキング補正（例えば、反復配列の除去）；G / C 含量の補正；局所重み付け多項式（例えば、L O E S S ）回帰補正、ロバストな推定量（例えば、位置の推定値 [例え、予測値；アベレージと同様である]、尺度の推定値 [例え、変動性] ；および変動性の分析 [例え、ANOVA] ）を使用した正規化が挙げられる。さらに、ある特定の実施形態では、以下のデータ処理方法、フィルタリング、正規化、重み付け、ピーク高さのモニタリング、ピーク面積のモニタリング、ピーク端のモニタリング、面積比の決定、データの数学的処理、データの統計学的処理、統計アルゴリズムの適用、固定された変数を用いた解析、最適化された変数を用いた解析、追加的な処理のためにデータをプロットしてパターンまたは傾向を同定すること、スライディングウィンドウ処理（例えば、スライディングウィンドウ正規化）、スタティックウィンドウ処理（例えば、スタティックウィンドウ正規化）など、および前述のものとの組合せの 1 つまたは複数を利用してデータセットをさらに処理することができる。ある特定の実施形態では、補正ステップの前に、処理方法をデータセットに適用することができる。いくつかの実施形態では、さまざまな特徴（例えば、GC 含量、マッピングされた冗長な読み取り、セントロメア領域、テロメア領域、反復配列など、およびそれらの組合せ）および / または変数（例えば、胎児の性別、母体の年齢、母体の倍数性、胎児核酸のパーセント寄与など、またはそれらの組合せ）に基づいてデータセットを補正および / または処理する。ある特定の実施形態では、本明細書に記載の通りデータセットを処理することにより、大きく、かつ / または複雑なデータセットの複雑さおよび / 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は、年齢および民族的なバックグラウンドが異なる 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した配列読み取りデータが挙げられる。いくつかの実施形態では、データセットは、数千から数百万までの、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の配列読み取りを含んでよい。

20

30

【 0 1 6 4 】

ある特定の実施形態では、データの補正および / または処理は任意の数のステップで実施することができ、2 以上のステップを伴う実施形態では、ステップを任意の順序で実施することができる。例えば、いくつかの実施形態では、単一の補正 / 処理手順のみを使用してデータを補正および / または処理することができ、ある特定の実施形態では、1 以上、5 以上、10 以上または 20 以上の補正 / 処理ステップ（例えば、1 以上の補正 / 処理ステップ、2 つ以上の補正 / 処理ステップ、3 以上の補正 / 処理ステップ、4 以上の補正 / 処理ステップ、5 以上の補正 / 処理ステップ、6 以上の補正 / 処理ステップ、7 以上の補正 / 処理ステップ、8 以上の補正 / 処理ステップ、9 以上の補正 / 処理ステップ、10 以上の補正 / 処理ステップ、11 以上の補正 / 処理ステップ、12 以上の補正 / 処理ステップ、13 以上の補正 / 処理ステップ、14 以上の補正 / 処理ステップ、15 以上の補正 / 処理ステップ、16 以上の補正 / 処理ステップ、17 以上の補正 / 処理ステップ、18

40

50

以上の補正 / 処理ステップ、19 以上の補正 / 処理ステップ、または 20 以上の補正 / 処理ステップ) を使用してデータを補正 / 処理することができる。いくつかの実施形態では、補正 / 処理ステップは、2 回以上繰り返される同じステップであってよく (例えば、2 回以上フィルタリングすること、2 回以上正規化すること)、ある特定の¹⁰実施形態では、補正 / 処理ステップは、2 回以上の異なる補正 / 処理ステップ (例えば、反復マスクング、フローセルに基づく正規化; ピンに関する G / C 含量の補正、フローセルに基づく正規化; 反復マスクング、ピンに関する G / C 含量の補正、プレートに基づく正規化; フィルタリング、正規化; 正規化、ピーク高さおよび端のモニタリング; フィルタリング、正規化、参照に対する正規化、p 値を決定するための統計学的操作など)、を同時にまたは逐次的に行う。いくつかの実施形態では、配列読み取りデータを処理してアウトカムをもたらすことを容易にするために、任意の適切な数および / または組合せの同じまたは異なる補正 / 処理ステップを利用することができる。ある特定の¹⁰実施形態では、本明細書に記載の基準によってデータセットを補正および / または処理することにより、データセットの複雑さおよび / または次元性を低下させることができる。

【0165】

いくつかの実施形態では、1 つまたは複数の補正 / 処理ステップは、本明細書に記載の 1 つまたは複数の実験条件について補正することを含んでよい。時には、データの変動性をもたらす実験条件の非限定的な例としては、配列の過大表示または過小表示 (例えば、偏った増幅に基づく変動性); ノイズの多いデータ; 偽データ点または外れ値データ点; フローセルに基づく変動性 (例えば、あるフローセルで分析された試料では見られるが、²⁰同じバッチ由来の (例えば、同じ試薬プレートまたはマイクロウェルプレートにおいて調製された)) 試料を分析するために使用した他のフローセルでは見られない変動性; および / またはプレートに基づく変動性 (例えば、分析のために使用したフローセルには関係なく、同じ試薬プレートまたはマイクロウェルプレートにおいて調製され、かつ / または同じマイクロウェルプレートにおいて分析のために段階分けされたいくつかのまたは全ての試料において見られる変動性) が挙げられる。

【0166】

いくつかの実施形態では、パーセント表示を、ゲノミックセクション (例えば、ゲノミックセクション、染色体、ゲノムまたはその一部) について算出する。いくつかの実施形態では、パーセント表示を、多数のゲノミックセクションにマッピングされたカウントの³⁰数に対して正規化された (例えば、それで割った)、1 つのゲノミックセクションにマッピングされたカウントの数として決定する。時には、パーセント表示の決定では、性染色体 (例えば、X 染色体および / または Y 染色体) に由来するゲノミックセクションおよび / またはカウントを排除する。時には、パーセント表示の決定には、常染色体に由来するゲノミックセクションおよび / またはカウントのみを含める。時には、パーセント表示の決定には、常染色体および性染色体に由来するゲノミックセクションおよび / またはカウントを含める。例えば、 $perc_i$ は選択されたゲノミックセクション i についてのパーセント表示を示す

【0167】

【化 1】

$$perc_i = \frac{\text{カウント}_i}{\sum_{j=1}^{22} \text{カウント}_j}$$

(式中、 カウント_i は、選択されたゲノミックセクション i にマッピングされた読み取りのカウントであり、 カウント_j は、多数のゲノミックセクション j (例えば、染色体 j 上の多数のゲノミックセクション、全ての常染色体 j 上のゲノミックセクション、ゲノム j のゲノミックセクション) にマッピングされた読み取りのカウントの数である)。例えば、 chr_i は染色体 i についての染色体表示を示す

【0168】

10

20

30

40

50

【化2】

$$chr_i = \frac{\text{カウント}_i}{\sum_{j=1}^{22} \text{カウント}_j}$$

(式中、カウント_jは染色体j上のアラインメントされた読み取りの数である)。いくつかの実施形態では、パーセント表示は、「ゲノムセクションカウント表示」である。時には、パーセント表示は、「ゲノミックセクション表示」または「染色体表示」である。

【0169】

ある特定の実施形態では、1つまたは複数の補正/処理ステップは、実験条件に誘導される変動性について補正することを含んでよい。変動性は、尺度および/または位置のロバストな推定量を使用することによって補正することができる。いくつかの実施形態では、(1)選択されたゲノミックセクションのパーセント表示(例えば、第1のゲノムセクションカウント表示;染色体、例えば第21染色体)、(2)選択されたゲノミックセクションについてのパーセント表示の全ての値の中央値、平均値、最頻値、アベレージおよび/または中点、(3)パーセント表示の全ての値の中央絶対偏差(MAD)を決定し、外れ値の影響を最小限にするまたは排除する、フローセルに基づくロバストな推定量を使用してzスコアを補正することによって、zスコアを実験条件に誘導される変動性について補正することができる。ある特定の実施形態では、標的染色体、標的ゲノム領域または標的ゲノミックセクション(例えば、第21染色体)についての、ロバストなフローセルに基づくzスコアの補正は、以下の式を利用して算出する。

【0170】

【化3】

$$Z_{\text{ロバスト}} = \frac{\text{perc}_i - \text{中央値}(\{\text{perc}_{iec}\})}{\text{MAD}(\{\text{perc}_{iec}\})}$$

書かれている式は、ゲノミックセクションについてのロバストなZスコアを算出するように構成されており、式中、percは選択されたゲノミックセクションi(例えば、任意の適切なゲノミックセクション、染色体、ゲノムまたはその一部)のパーセント表示(例えば、第1のゲノムセクションカウント表示、染色体表示)である。いくつかの実施形態では、実験条件ecについて得られた、選択されたゲノミックセクションiについての1つまたは複数のパーセント表示値から中央値を算出する。実験条件ec'について得られた、選択されたゲノミックセクションiについての1つまたは複数のパーセント表示値からMADを算出する。ある特定の実施形態では、一般式を利用して、選択された標的ゲノミックセクションについての等価な値を代入することによって任意のゲノミックセクションについてのロバストなzスコアを得ることができる。いくつかの実施形態では、選択された試料のセットまたは試料のサブセットについて中央値、平均値、最頻値、アベレージ、中点および/またはMADを算出する。時には、同じ試料のセットについて中央値および/またはMADを算出する。いくつかの実施形態では、異なる試料のセットについて中央値および/またはMADを算出する。いくつかの実施形態では、実験条件ecは同じである。いくつかの実施形態では、実験条件ecは、1つまたは複数の共通の実験条件を含んでよいまたはそれからなつてよい。いくつかの実施形態では、実験条件ecは異なる。いくつかの実施形態では、実験条件ec'は同じである。いくつかの実施形態では、実験条件ec'は、1つまたは複数の共通の実験条件を含んでよいまたはそれからなつてよい。いくつかの実施形態では、実験条件ec'は異なる。時には、実験条件ecおよびec'は異なる。いくつかの実施形態では、実験条件ecおよびec'は、1つまたは複数の共通の実験条件を含んでよい、またはそれからなつてよい。例えば、選択されたゲノミックセクションについてのロバストなZスコアを、(a)選択された試料のセットから収

10

20

30

40

50

集された、1つまたは複数の共通の実験条件下で（例えば、同じフローセルから）得たデータの選択されたセットから得た平均値、および（b）別の選択された試料のセットから収集された、1つまたは複数の共通の実験条件下で（例えば、異なるフローセルおよび選択された試薬の同じロットを使用して）得たデータの別の選択されたセットから得たMADから算出することができる。いくつかの実施形態では、平均値およびMADは、少なくとも1つの共通の実験条件を共有するデータから得られる。時には、平均値およびMADは、共通の実験条件を共有しないデータから得られる。

【0171】

いくつかの実施形態では、正規化された試料カウント（例えば、Zスコア）は、第1のゲノミックセクションのカウント（例えば、カウント、パーセント表示）から予測カウント（例えば、カウントの中央値、パーセント表示の中央値）を引き算し、それにより減算値を生成し、減算値をカウントの変動性の推定値で割ること（例えば、MAD、カウントのMAD、パーセント表示のMAD）を含むプロセスによって得られる。いくつかの実施形態では、予測カウント（例えば、カウントの中央値、パーセント表示の中央値）およびカウントの変動性の推定値（例えば、MAD、カウントのMAD、パーセント表示のMAD）は、少なくとも1つの共通の実験条件を共有するデータから得られる。時には、予測カウント（例えば、カウントの中央値、パーセント表示の中央値）およびカウントの変動性の推定値（例えば、MAD、カウントのMAD、パーセント表示のMAD）は、共通の実験条件を共有しないデータから得られる。いくつかの実施形態では、中央値は、中央値、平均値、最頻値、アベレージおよび/または中点であってよい。

【0172】

ある特定の実施形態では、1つまたは複数の補正/処理ステップは、フローセルに基づく変動性について補正することを含んでよい。フローセルに基づく変動性は、尺度および/または位置のロバストな推定量を使用することによって補正することができる。いくつかの実施形態では、（1）選択された染色体（例えば、第1のゲノムセクションカウント表示；例えば第21染色体）のパーセント表示、（2）フローセルにおいて観察された染色体表示の全ての値の中央値、（3）フローセルにおいて観察された染色体表示の全ての値の中央絶対偏差を決定し、外れ値の影響を最小限にするまたは排除する、フローセルに基づくロバストな推定量を使用してZスコアを補正することによって、フローセルに基づく変動性についてZスコアを補正することができる。ある特定の実施形態では、標的染色体、標的ゲノム領域または標的ゲノミックセクション（例えば、第21染色体）についての、ロバストなフローセルに基づくZスコアの補正は、以下の式を利用して算出する。

【0173】

【化4】

$$\text{ロバスト}_{FC} = \frac{\text{perc.chr21} - \text{中央値}(\{\text{perc.chr21}\})_{FC}}{\text{MAD}(\{\text{perc.chr21}\})_{FC}}$$

書かれている式は第21染色体についてのロバストなZスコアを算出するように構成されており、式中、perc.chr21は、パーセント第21染色体表示（例えば、第1のゲノムセクションカウント表示）であり、MADは中央絶対偏差を表し、FCはフローセルを表す。ある特定の実施形態では、第21染色体参照が指定される（例えば、.chr21）場合、一般式を利用して、選択された標的染色体、標的ゲノム領域または標的ゲノミックセクションについての等価な値を代入することによって任意の染色体についてのロバストなZスコアを得ることができる。

【0174】

いくつかの実施形態では、1つまたは複数の補正/処理ステップは、プレートに基づく変動性について補正することを含んでよい。プレートに基づく変動性は、尺度および/または位置のロバストな推定量を使用することによって補正することができる。ある特定の実施形態では、（1）選択された染色体（例えば、第1のゲノムセクションカウント表示

；例えば第 2 1 染色体)のパーセント表示、(2) 1 つまたは複数のプレートにおいて観察された染色体表示の全ての値の中央値、(3) 1 つまたは複数のプレートにおいて観察された染色体表示の全ての値の中央絶対偏差を決定し、外れ値の影響を最小限にするまたは排除する、プレートに基づくロバストな推定量を使用して z スコアを補正することによって、プレートに基づく変動性について z スコアを補正することができる。ある特定の実施形態では、標的染色体、標的ゲノム領域または標的ゲノミックセクション(例えば、第 2 1 染色体)についてのロバストなプレートに基づく z スコアの補正は、以下の式を利用して算出する。

【 0 1 7 5 】
【化 5】

10

$$\text{ロバストプレート} = \frac{\text{perc.chr21} - \text{中央値}(\{\{\text{perc.chr21}\}\})_{\text{プレート}}}{\text{MAD}(\{\{\text{perc.chr21}\}\})_{\text{プレート}}}$$

書かれている式は第 2 1 染色体についてのロバストな Z スコアを算出するように構成されており、式中、perc.chr21 はパーセント第 2 1 染色体表示(例えば、第 1 のゲノムセクションカウント表示)であり、MAD は中央絶対偏差を表し、プレートは、試料の 1 つまたは複数のプレート(例えば、1 つまたは複数の試薬プレート、1 つまたは複数の試料調製プレート、1 つまたは複数の段階分けプレート)を表す。ある特定の実施形態では、第 2 1 染色体参照が指定される(例えば、.chr21)場合、一般式を利用して、選択された標的染色体、標的ゲノム領域または標的ゲノミックセクションについての等価な値を代入することによって任意の染色体についてのロバストな z スコアを得ることができる。

20

【 0 1 7 6 】

時には、次式を使用して中央絶対偏差(MAD)を算出する：

【 0 1 7 7 】
【化 6】

$$\text{MAD} = 1.4826 * \text{中央値}(\{|X - \text{中央値}(\{X\})|\})$$

30

式中、X は中央絶対偏差が算出される任意のランダムな変数を表し、正規化定数 1 . 4 8 2 6 は 1 / Inv [ファイ] (3 / 4) を表し、ファイは標準のガウス(例えば、正規)分布についての累積分布関数を表し、Inv [ファイ] はその逆関数である(例えば、分位関数に関連する)。Inv [ファイ] は X = 3 / 4 で評価したものであり、1 / 1 . 4 8 2 6 と等しい。「Rコード」では、正規化定数を算出するための方程式は、1 / qnorm (3 / 4) = 1 . 4 8 2 6 である。「Rコード」は、S プログラミング言語に実質的に類似した、種々の統計解析のために使用される非専売のオープンソースプログラミング言語である(例えば、R Development Core Team (2 0 1 0 年) . R : A language and environment for statistical computing . R Foundation for Statistical Computing, Vienna, Austria . ISBN 3 - 9 0 0 0 5 1 - 0 7 - 0、URLワールドワイドウェブ.R-project.org /)。正規化定数 1 . 4 8 2 6 は、正規分布したデータの中央絶対偏差(例えば、MAD)が、大きな試料について、同じデータの標準偏差(例えば、STDEV)と等しくなるように選択され、MAD および STDEV が同じ尺度に有効に置かれる。多くの場合、分位関数を利用して確率分布を規定する。いくつかの実施形態では、確率分布の分位関数はその積分の逆関数であり、多くの場合、所与の確率についてランダムな変数の値またはそれが下回る値を特定する。

40

【 0 1 7 8 】

ある特定の実施形態では、1 つまたは複数の補正 / 処理ステップは、配列の過大表示ま

50

たは過小表示について補正することを含んでよい。本明細書に記載の通り、いくつかの調製および/または配列決定ステップにおいて利用する増幅手順により、時には、配列含量および/または構造に起因して配列の過大表示または過小表示が生じる。配列の過大表示または過小表示により、時には、もたらされるアウトカムの信頼度が低下する。ある特定の実施形態では、ロバストな推定量を使用してデータセットの一部または全部を予測値に関して補正または正規化することにより、過剰配列表示または過小配列表示の影響を最小限にすることまたは排除することができる。いくつかの実施形態では、アベレージ、中央値、アベレージ、中点、最頻値、中央絶対偏差 (MAD)、RousseeuwおよびCrouxによって導入されるMADの代替値、ブートストラップ推定値、標準偏差、zスコア、ロバストなzスコア、ANOVA、LOESS回帰分析(例えば、LOESS平滑化、LOWESS平滑化)などから選択される1つまたは複数の推定量を使用して、染色体の部分、または全てについての予測値を算出する。データセットの一部または全部を補正して配列の過大表示または過小表示の影響を低下させるまたは排除することにより、アウトカムをもたらしやすくすることを容易にし、かつ/またはデータセットの複雑さおよび/または次元性を低下させることができる。

【0179】

いくつかの実施形態では、1つまたは複数の補正/処理ステップは、G/C含量について補正することを含んでよい。本明細書に記載の通り、G/C含量が高い配列は、時には、生のデータセットまたは処理されたデータセットにおいて過大表示または過小表示される。ある特定の実施形態では、ロバストな推定量を使用してデータセットの一部または全部を予測値に関して補正または正規化することにより、データセットの一部または全部(例えば、選択されたピン、染色体の選択された部分、選択された染色体)についてのG/C含量を補正してG/C含量の偏りを最小限にするまたは排除する。いくつかの実施形態では、予測値はヌクレオチド配列読み取りのG/C含量であり、ある特定の実施形態では、予測値は試料核酸のG/C含量である。いくつかの実施形態では、アベレージ、中央値、平均値、最頻値、中点、中央絶対偏差、(MAD)、RousseeuwおよびCrouxによって導入されるMADの代替値、ブートストラップ推定値、標準偏差、zスコア、ロバストなzスコア、ANOVA、LOESS回帰分析(例えば、LOESS平滑化、LOWESS平滑化)などから選択される1つまたは複数の推定量を使用して、染色体の部分、または全てについての予測値を算出する。いくつかの実施形態では、データセットの一部または全部を補正してG/C含量の偏りの影響を低下させるまたは排除することにより、アウトカムをもたらしやすくすることを容易にし、かつ/またはデータセットの複雑さおよび/または次元性を低下させることができる。

【0180】

PERUN

核酸指標に関連する誤差を減少させるために特に有用な正規化方法体系は、本明細書では、パラメータ化された誤差除去および不偏正規化(Parameterized Error Removal and Unbiased Normalization)(PERUN)と称される。PERUN方法体系は、種々の核酸指標(例えば、核酸配列読み取り)に、そ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下させるために適用することができる。

【0181】

例えば、PERUN方法体系は、試料からの核酸配列読み取りに適用し、核酸の高度の決定(例えば、ゲノミックセクションの高度の決定)を損なう可能性がある誤差の影響を低下させることができる。そのような適用は、核酸配列読み取りを使用して、ヌクレオチド配列(例えばゲノミックセクション)の変動する高度として顕在化する、被験体における遺伝的変異の有無を評価するために有用である。ゲノミックセクションにおける変動の非限定的な例は、染色体異数性(例えば、21トリソミー、18トリソミー、13トリソミー)および性染色体(例えば、女性におけるXX対男性におけるXY)の有無である。常染色体(例えば、性染色体以外の染色体)のトリソミーは、影響を受けた常染色体と称

10

20

30

40

50

することができる。ゲノミックセクションの高度における変動の他の非限定的な例としては、微小欠失、微小挿入 (microinsertion)、重複およびモザイク現象が挙げられる。

【0182】

ある特定の適用では、PERUN方法体系により、ピンと称される特定のゲノムの群についての核酸指標を正規化することによって実験上の偏りを減少させることができる。ピンは適切な核酸指標の収集物を含み、その非限定的な例としては、本明細書ではゲノミックセクションまたは参照ゲノムの部分と称される、ある長さの連続したヌクレオチドが挙げられる。ピンは、本明細書に記載の他の核酸指標を含んでよい。そのような適用では、PERUN方法体系により、一般には、3次元においていくつもの試料にわたって特定の

10

【0183】

ある特定の実施形態では、PERUN方法体系は、各ピンについて、(i)配列読み取りをマッピングする参照ゲノムのピンについての実験上の偏りと(ii)ピンにマッピングされた配列読み取りのカウントとの間のフィッティングした関係からゲノミックセクションの高度を算出することを含む。ピンのそれぞれについての実験上の偏りは、複数の試料にわたって、各試料についての(i)ピンのそれぞれにマッピングされた配列読み取りのカウントと(ii)ピンのそれぞれについてのマッピングの特徴との間のフィッティングした関係に応じて決定することができる。この各試料についてのフィッティングした関係を複数の試料について3次元で集合させることができる。ある特定の実施形態では(例えば、図82、実施例4)、この集合体を実験上の偏りに応じて順序づけることができるが、PERUN方法体系は実験上の偏りに応じて集合体を順序づけずに実施することができる。

20

【0184】

関係は当技術分野で公知の方法によって生成することができる。ある特定の実施形態では、各試料について2次元で関係を生成することができ、誤差を証明する、またはおそらく誤差を証明する変数を次元の1つ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、ユーザーによって提供される2つ以上の変数の値を使用してグラフをプロットする当技術分野で公知のグラフ作成ソフトウェアを使用して生成することができる。関係は、

30

【0185】

PERUN方法体系では、フィッティングした関係の1つまたは複数は線形であってよい。実験上の偏りがGCの偏りであり、マッピングの特徴がGC含量である妊婦由来の循環している無細胞核酸を分析するためには、試料についての(i)各ピンにマッピングされた配列読み取りのカウントと(ii)ピンのそれぞれについてのGC含量との間のフィッティングした関係は線形であってよい。後者のフィッティングした関係について、傾きはGCの偏りに関係し、GCの偏り係数は、フィッティングした関係を複数の試料にわたって集合させた際に各ピンについて決定することができる。そのような実施形態では、複数の試料およびピンについての(i)ピンについてのGCの偏り係数と(ii)ピンにマッピングされた配列読み取りのカウントとの間のフィッティングした関係も線形であってよい。切片および傾きは、後者のフィッティングした関係から入手することができる。そのような適用では、傾きはGC含量に基づいて試料に特異的な偏りに対処し、切片は全ての試料に共通するピンに特異的な減弱パターンに対処する。PERUN方法体系では、アウトカム(例えば、遺伝的変異の有無;胎児の性別の決定)をもたらすためにゲノミックセクションの高度を算出する際のそのような試料に特異的な偏りおよびピンに特異的な減

40

50

弱を有意に減少させることができる。

【0186】

したがって、PERUN方法体系を並行して複数の試料にわたって配列読み取りに適用することにより、(i)試料に特異的な実験上の偏り(例えば、GCの偏り)および(ii)試料に共通するピンに特異的な減弱によって引き起こされる誤差を有意に減少させることができる。これらの2つの誤差の原因のそれぞれに別々にまたは逐次的に対処する他の方法では、多くの場合、これらをPERUN方法体系ほど有効に減少させることができない。理論によって限定されることなく、PERUN方法体系では、一部において、その一般に加法的なプロセスでは、他の正規化手法(例えば、GC-LOESS)において利用される一般に乗法的なプロセスほど広がりが大きくなるので、より有効に誤差が減少することが予想される。

10

【0187】

追加的な正規化および統計学的技法をPERUN方法体系と組み合わせて利用することができる。追加的なプロセスは、PERUN方法体系を用いる前、その後、またはその間に適用することができる。PERUN方法体系と組み合わせて使用することができるプロセスの非限定的な例は下に記載されている。

【0188】

いくつかの実施形態では、GC含量についてのゲノミックセクションの高度の二次的な正規化または補正をPERUN方法体系と併せて利用することができる。適切なGC含量の補正または正規化の手順を利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、追加的なGC正規化プロセス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方法体系を適用することにより、各試料についてGCの偏りを決定することができ、GCの偏りがある特定の閾値を上回る試料を追加的なGC正規化プロセスのために選択することができる。そのような実施形態では、所定の閾値の高度を使用して、そのような試料を追加的なGC正規化のために選択することができる。

20

【0189】

ある特定の実施形態では、ピンフィルタリングまたは重み付けプロセスをPERUN方法体系と併せて利用することができる。適切なピンフィルタリングまたは重み付けプロセスを利用することができ、その非限定的な例は本明細書に記載されている。実施例4および実施例5には、ピンフィルタリングのための誤差のR因子(R factor)尺度の利用について記載されている。

30

【0190】

GCの偏りモジュール

GCの偏りの決定(例えば、参照ゲノムの部分(例えば、ゲノミックセクション)のそれぞれについてのGCの偏りの決定)は、GCの偏りモジュールによって(例えば、GCの偏りモジュールを含む装置によって)もたらされる。いくつかの実施形態では、GCの偏りの決定をもたらしためにはGCの偏りモジュールが必要である。時には、GCの偏りモジュールにより、参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウントと各部分のGC含量との間のフィッティングした関係(例えば、フィッティングした線形関係)からGCの偏りの決定がもたらされる。GCの偏り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んでよい。いくつかの実施形態では、GCの偏りの決定(すなわち、GCの偏りのデータ)は、GCの偏りモジュールからの1つまたは複数の命令(例えば、プロセス、ルーチンおよび/またはサブルーチン)を実施および/または実行することができるプロセッサ(例えば、1つまたは複数のプロセッサ)を含む装置によってもたらされる。いくつかの実施形態では、GCの偏りのデータは、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置によってもたらされる。いくつかの実施形態では、GCの偏りモジュールは、1つまたは複数の外部のプロセッサ(例えば、内部または外部のネットワーク、サーバー、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))と一緒に作動する。いくつかの実施形態では、

40

50

GCの偏りのデータは、以下の1つまたは複数を含む装置によってもたらされる：1つまたは複数のフローセル、カメラ、液体取扱い構成部分、プリンター、ディスプレイ（例えば、LED、LCDまたはCRT）など。GCの偏りモジュールにより、適切な装置またはモジュールからデータおよび/または情報を受け取ることができる。時には、GCの偏りモジュールにより、配列決定モジュール、正規化モジュール、重み付けモジュール、マッピングモジュールまたはカウントモジュールからデータおよび/または情報を受け取ることができる。GCの偏りモジュールは、時には、正規化モジュール（例えば、PERUN正規化モジュール）の一部である。いくつかの実施形態では、GCの偏りモジュールにより、配列決定モジュールから配列決定読み取りを受け取ることができ、マッピングモジュールからマッピングされた配列決定読み取りを受け取ることができ、かつ/またはカウントモジュールからカウントを受け取ることができる。多くの場合、GCの偏りモジュールにより、装置または別のモジュール（例えば、カウントモジュール）からデータおよび/または情報が受け取られ、データおよび/または情報が変換され、GCの偏りのデータおよび/または情報（例えば、GCの偏りの決定、線形フィッティングした関係など）がもたらされる。ある特定の实施形態では、GCの偏りのデータおよび/または情報を、GCの偏りモジュールから、予測カウントモジュール、フィルタリングモジュール、比較モジュール、正規化モジュール、重み付けモジュール、範囲設定モジュール、補正モジュール、カテゴリー化モジュール、および/またはアウトカムモジュールに移行することができる。

10

【0191】

20

他のデータ処理

ある特定の实施形態では、1つまたは複数の補正/処理ステップは、反復配列について補正することを含んでよい。本明細書に記載の通り、反復配列は、多くの場合、情報価値のないデータであり、かつ/またはノイズの多いデータに寄与する可能性があり、それにより、時には、もたらされるアウトカムの信頼度が低下する。本明細書に記載のまたは当技術分野で公知の反復配列の影響を低下させるための任意の適切な方法（例えば、反復配列の除去）を使用することができる。反復配列を除去するために利用可能なリソースの非限定的な例は、以下の刊行物に見いだすことができる：URLワールドワイドウェブrepeatmasker.org/papers.htmlおよびワールドワイドウェブbiomedcentral.com/1471-2105/11/80。ある特定の实施形態では、もたらされるアウトカムに対する反復配列の存在の影響を、ロバストな推定量を使用してデータセットの一部または全部を予測値に関して補正または正規化することにより、最小限にすることまたは排除することができる。いくつかの実施形態では、アベレージ、中央値、最頻値、中点、平均値、中央絶対偏差、(MAD)、RousseeuwおよびCrouxによって導入されるMADの代替値、ブートストラップ推定値、標準偏差、zスコア、ロバストなzスコア、ANOVA、LOESS回帰分析（例えば、LOESS平滑化、LOWESS平滑化）などから選択される1つまたは複数の推定量を使用して、染色体の部分、または全てについての予測値を算出する。データセットの一部または全部を補正して反復配列の影響を低下させるまたは排除することにより、アウトカムをもたらしやすくを容易にし、かつ/またはデータセットの複雑さおよび/または次元性を低下

30

40

【0192】

いくつかの実施形態では、1つまたは複数の補正/処理ステップは、指標配列を補正することを含んでよい。本明細書に記載の通り、本明細書に記載の複数の实施形態において利用されるアダプタープライマーは、しばしば指標配列を含む。全ての指標が実質的に同じ性能を有する場合、染色体表示、またはいくつかの他のゲノムに関連する同等の測定基準は、異なる指標によって標識された試料の実質的に全てにわたって同じように分布することになる。しかし実際には、いくつかの指標は、他のものよりも良好に働き、それにより今度は、アルゴリズムによっていくつかの断片が他の断片に対して優先的に分析される（例えば、高く重み付けされる）ということが引き起こされる。さらに、いくつかの指標

50

によってもたらされる検出され、かつ/またはアラインメントされた読み取りの数はより少なく、それにより今度は、他の指標を用いてタグ付けされた試料と比較した場合、それらの指標配列を用いてタグ付けされた試料についての分解がもたらされる。ある特定の実施形態では、1つまたは複数の指標配列に関して、推定量を使用してデータセットの一部または全部を補正または正規化することができ、ある特定の実施形態では、推定量は、アベレージ、中央値、平均値、最頻値、中点、中央絶対偏差(MAD)、RousseeuwおよびCrouxによって導入されるMADの代替値、ブートストラップ推定値、標準偏差、zスコア、ロバストなzスコア、ANOVA、LOESS回帰分析(例えば、LOESS平滑化、LOWESS平滑化)などから選択される。データセットの一部または全部を補正して、1つまたは複数の指標配列に関して低下させることにより、アウトカムを

10

【0193】

データセットの一部または全部を、下記の1つまたは複数の手順を使用してさらに処理することもできる。

【0194】

いくつかの実施形態では、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んでよい。フィルタリングにより、一般に、ゲノミックセクションまたはピンが考察から除去される。これだけに限定されないが、冗長なデータ(例えば、冗長なまたはオーバーラップしているマッピングされた読み取り)、情報価値のないデータ(例えば、ゼロカウント中央値を有するピン)、過大表示または過小表示された配列を有するピン、ノイズの多いデータなど、または前述のものの組合せを含めた任意の適切な基準に基づいて、除去するためにピンを選択することができる。フィルタリングプロセスは、多くの場合、1つまたは複数のピンを考察から除去し、除去に選択された1つまたは複数のピンにおけるカウントを、考察されているピン、1つまたは複数の染色体、またはゲノムについてのカウントまたは合計されたカウントから引き算することを含む。いくつかの実施形態では、ピンを逐次的に(例えば、個々のピンそれぞれの除去の影響を評価することを可能にするために一度に1つ)除去することができる、ある特定の実施形態では、除去に選出された全てのピンを同時に除去することができる。いくつかの実施形態では、分散がある特定のレベルを上回るまたは下回ることを特徴とするゲノミックセクションを除去し、これは、時には、本明細書では、「ノイズの多い」ゲノミックセクションのフィルタリングと称される。ある特定の実施形態では、フィルタリングプロセスは、データセットから、ゲノミックセクション、染色体、または染色体の部分のプロファイルの高度の平均値から、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得ることを含み、ある特定の実施形態では、フィルタリングプロセスは、データセットから、ゲノミックセクション、染色体または染色体の部分のプロファイルの高度の平均値から、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除去することを含む。いくつかの実施形態では、フィルタリングプロセスを利用して、遺伝的変異の有無について分析する候補ゲノミックセクションの数を減少させる。遺伝的変異(例えば、微小欠失、微小重複)の有無について分析される候補ゲノミックセクションの数を減少させることにより、多くの場合、データセットの複雑さおよび/または次元性が低下し、時には、遺伝的変異および/または遺伝子異常(genetic aberration)を検索し、かつ/または同定するスピードが2桁以上増大する。

20

30

40

【0195】

いくつかの実施形態では、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んでよい。正規化は、当技術分野で公知の適切な方法によって実施することができる。時には、正規化は、異なる尺度で測定された値を概念上共通の尺度に対して補正することを含む。時には、正規化は、補正された値の確率分布をアラインメントする複雑な数学的補正を含む。いくつかの場合には、正規化は、分布を正規分布に対してアラインメントすることを含む。時には、正規化は、異なるデータセットについての対応する正

50

正規化された値を、特定の全般的な影響（例えば、誤差および例外）の影響が排除されるように比較することを可能にする数学的補正を含む。時には、正規化は、尺度化を含む。正規化は、時には、1つまたは複数のデータセットを所定の変数または式によって分割することを含む。正規化方法の非限定的な例としては、ビン様式での正規化（bin wise normalization）、GC含量による正規化、線形最小二乗回帰および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS（局所重み付け散布プロット平滑化）、PERUN（以下を参照されたい）、反復マスキング（RM）、GC正規化および反復マスキング（GCRM）、cQnおよび/またはそれらの組合せが挙げられる。いくつかの実施形態では、遺伝的変異の有無（例えば異数性）の決定に、正規化方法（例えば、ビン様式での正規化、GC含量による正規化、線形最小二乗回帰および非線形最小二乗回帰、LOESS、GC LOESS、LOWESS（局所重み付け散布プロット平滑化）、PERUN、反復マスキング（RM）、GC正規化および反復マスキング（GCRM）、cQn、当技術分野で公知の正規化方法および/またはその組合せ）を利用する。

10

【0196】

例えば、LOESSは、k近傍に基づくメタモデルにおいて多数の回帰モデルを組み合わせる当技術分野で公知の回帰モデリング方法である。LOESSは、時には、局所重み付け多項式回帰と称される。いくつかの実施形態では、GC LOESSでは、ゲノミックセクションについての断片カウント（例えば、配列読み取り、カウント）とGC組成との間の関係にLOESSモデルを適用する。LOESSを使用してデータ点のセットを通して滑らかな曲線をプロットすることは、時には、特に平滑化された値のそれぞれが、重み付けされた二次最小二乗回帰によってy軸の散布図基準変数の値の範囲にわたって得られる場合、LOESS曲線と称される。LOESS法では、データセット内の各点について、応答が推定される点の近くの説明的な変数値を用いて低次数の多項式をデータのサブセットにフィッティングする。重み付けされた最小二乗を使用して多項式をフィッティングし、応答が推定される点の近くの点を高く重み付けし、さらに離れた点を低く重み付けする。次いで、ある点についての回帰関数の値を、そのデータ点についての説明的変数値を使用して局所多項式を評価することによって得る。LOESSフィッティングは、時には、データ点のそれぞれについて回帰関数値が計算された後に完了したとみなされる。多項式モデルの次数および重みなどのこの方法の詳細の多くは柔軟である。

20

30

【0197】

ある特定の実施形態では、正規化とは、1つまたは複数のデータセットを所定の変数によって分割することを指す。任意の適切な数の正規化を使用することができる。いくつかの実施形態では、データセットを1回以上、5回以上、10回以上、さらには20回以上正規化することができる。データセットは、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、またはその両方）を表す値（例えば、正規化値）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正規化の種類非限定的な例としては、1つまたは複数の選択された検査ゲノミックセクションまたは参照ゲノミックセクションについての生のカウントデータを、選択された1つまたは複数のゲノミックセクションがマッピングされる染色体またはゲノム全体にマッピングされたカウントの総数に対して正規化すること；1つまたは複数の選択されたゲノムのセグメントについての生のカウントデータを1つまたは複数の選択されたゲノムのセグメントがマッピングされる1つまたは複数のゲノミックセクションまたは染色体についての参照カウント中央値に対して正規化すること；生のカウントデータを予め正規化されたデータまたはその誘導値に対して正規化すること；および予め正規化されたデータを1つまたは複数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。データセットを正規化することには、時には、所定の正規化変数として選択される特徴または性質に応じて統計学的誤差を分離する効果がある。データセットを正規化することにより、時には、データを共通の尺度（例えば、所定の正規化変数）に導くことによって異なる尺度を有するデータのデータ特性を比較することも可能になる。いくつかの実施形態では、統計学的に導かれた値に対する1つまたは複数

40

50

の正規化を利用して、データの差異を最小限にし、特定範囲外のデータの重要性を減らすことができる。正規化値に関してゲノミックセクション、またはピンを正規化することは、時には、「ピン様式での正規化」と称される。

【0198】

ある特定の実施形態では、正規化を含む処理ステップは、スタティックウィンドウに対して正規化することを含み、いくつかの実施形態では、正規化を含む処理ステップは、ムービング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。「ウィンドウ」は、多くの場合、分析のために選択された1つまたは複数のゲノミックセクションであり、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/または他の数学的操作または統計学的操作のために使用される）。スタティックウィンドウに対して正規化することは、多くの場合、正規化プロセスにおいて試験被験体データセットと参照被験体データセットを比較するために選択された1つまたは複数のゲノミックセクションを使用することを伴う。いくつかの実施形態では、選択されたゲノミックセクションを利用してプロファイルを生成する。スタティックウィンドウは、一般に、操作および/または分析の間に変化しない所定のゲノミックセクションのセットを含む。ムービングウィンドウに対する正規化、またはスライディングウィンドウに対する正規化は、多くの場合、選択された検査ゲノミックセクションのゲノム領域に局在しているゲノミックセクション（例えば、遺伝学のごく周囲の近接する1つまたは複数のゲノミックセクションなど）に対して実施される正規化であり、1つまたは複数の選択された検査ゲノミックセクションを、その選択された検査ゲノミックセクションのごく周囲のゲノミックセクションに対して正規化する。ある特定の実施形態では、選択されたゲノミックセクション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ムービングウィンドウ正規化は、多くの場合、近接する検査ゲノミックセクションに繰り返し移動またはスライドさせ、新しく選択された検査ゲノミックセクションを、新しく選択された検査ゲノミックセクションのごく周囲のまたはそれに近接するゲノミックセクションに対して正規化することを含み、近接するウィンドウは、共通の1つまたは複数のゲノミックセクションを有する。ある特定の実施形態では、複数の選択された検査ゲノミックセクションおよび/または染色体をスライディングウィンドウプロセスによって分析することができる。

【0199】

いくつかの実施形態では、スライディングウィンドウまたはムービングウィンドウに対して正規化することにより1つまたは複数の値を生成することができ、各値は、ゲノム（例えば、染色体）の異なる領域から選択される参照ゲノミックセクションの異なるセットに対する正規化を表す。ある特定の実施形態では、生成される1つまたは複数の値は累積和（例えば、選択されたゲノミックセクション、ドメイン（例えば、染色体の部分）、または染色体にわたって正規化されたカウントプロファイルの積分の数値的な推定値）である。スライディングウィンドウプロセスまたはムービングウィンドウプロセスによって生成される値を使用して、プロファイルを生成し、アウトカムに達することを容易にすることができる。いくつかの実施形態では、1つまたは複数のゲノミックセクションの累積和をゲノムの位置に応じて示すことができる。時には、ムービングウィンドウ分析またはスライディングウィンドウ分析を使用して、ゲノムを微小欠失および/または微小挿入の有無について分析する。ある特定の実施形態では、1つまたは複数のゲノミックセクションの累積和を示すことを使用して、遺伝的変異（例えば、微小欠失、微小重複）の領域の有無を同定する。いくつかの実施形態では、ムービングウィンドウ分析またはスライディングウィンドウ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、ある特定の実施形態では、ムービングウィンドウ分析またはスライディングウィンドウ分析を使用して、微小重複を含有するゲノム領域を同定する。

【0200】

いくつかの実施形態では、処理ステップは、重み付けすることを含む。重み付け、または重み関数の実施とは、多くの場合、時には、ある特定のデータセットの特徴または変数の、他のデータセットの特徴または変数に対する影響を変更する（例えば、1つまたは複

10

20

30

40

50

数のゲノミックセクションまたはピンに含有されるデータの有意性および/または寄与を、選択された1つまたは複数のピン内のデータの質または有用性に基づいて増加または減少させる)ために利用されるデータセットの一部または全部の数学的操作である。いくつかの実施形態では、重み関数を使用して、測定分散が比較的小さいデータの影響を増加させ、かつ/または測定分散が比較的大きいデータの影響を減少させることができる。例えば、過小表示されたまたは質が低い配列データを有するピンを「低く重み付け」して、データセットに対する影響を最小限にすることができ、一方で選択されたピンを「高く重み付け」して、データセットに対する影響を増加させることができる。重み関数の非限定的な例は、 $[1 / (\text{標準偏差})^2]$ である。重み付けステップは、時には、正規化するステップと実質的に同様の様式で実施する。いくつかの実施形態では、データセットを所定の変数(例えば、重み付け変数)で割る。所定の変数(例えば、最小化した標的関数、ファイ)は、多くの場合、データセットの異なる一部が違うように重み付けされる(例えば、ある特定のデータ型の影響が増加し、他のデータ型の影響が減少する)ように選択される。

10

【0201】

ある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操作および/または統計学的操作を含んでよい。任意の適切な数学的操作および/または統計学的操作を単独でまたは組み合わせて使用して、本明細書に記載のデータセットを分析し、かつ/または操作することができる。任意の適切な数の数学的かつ/または統計学的操作を使用することができる。いくつかの実施形態では、データセットを、1回以上、5回以上、10回以上または20回以上数学的かつ/または統計学的に操作することができる。使用することができる数学的かつ統計学的な操作の非限定的な例としては、足し算、引き算、掛け算、割り算、代数関数、最小二乗推定量、曲線フィッティング、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、ファイ値、ピークの高度の分析、ピーク端位置の決定、ピーク面積比の算出、染色体の高度の中央値の分析、平均絶対偏差の算出、残差平方和、平均値、標準偏差、標準誤差など、またはそれらの組合せが挙げられる。数学的操作および/または統計学的操作は、配列読み取りデータの全部または一部、またはその処理生成物に対して実施することができる。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は、生のカウント、フィルタリングされたカウント、正規化されたカウント、ピーク高さ、ピーク幅、ピーク面積、ピーク端

20

30

【0202】

いくつかの実施形態では、処理ステップは、1つまたは複数の統計アルゴリズムの使用を含んでよい。任意の適切な統計アルゴリズムを単独でまたは組み合わせて使用して、本明細書に記載のデータセットを分析し、かつ/または操作することができる。任意の適切な数の統計アルゴリズムを使用することができる。いくつかの実施形態では、1種以上、5種以上、10種以上または20種以上の統計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。本明細書に記載の方法と一緒に使用するために適した統計アルゴリズムの非限定的な例としては、決定木、カウンターナル(counter null)、多重比較、オムニバス検定、ベレンス・フィッシャー問題、ブートストラッピング、独立した有意性の検定を組み合わせるためのフィッシャーの方法、帰無仮説、第1種の過誤、第2種の過誤、直接検定、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標本のプールしたt検定、不等分散を有する2標本のプールしていないt検定、1比率z検定(one-proportion z-test)、プールした2比率z検定(two-proportion z-test)、プールしていない2比率z検定、1標本カイ二乗検定、等分散性についての2標本F検定、信頼区間、信用区間、有意性、メタ分析、単純線形回帰、ロバスト線形回帰など、または前述のものの組合せが挙げられる。統計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

40

50

または特徴の非限定的な例としては、生のカウント、フィルタリングされたカウント、正規化されたカウント、ピーク高さ、ピーク幅、ピーク端、側方許容差、P値、中央値の高度、平均高度、ゲノム領域内のカウント分布、核酸種の相対的な表示など、またはそれらの組合せが挙げられる。

【0203】

ある特定の実施形態では、多数（例えば、2つ以上）の統計アルゴリズム（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-最隣接点、ロジスティック回帰および/またはロススムージング（loss smoothing））および/または数学的操作および/または統計学的操作（例えば、本明細書では操作と称される）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。いくつかの実施形態では、多数の操作を使用することにより、アウトカムをもたらすために使用することができるN次元の空間を生成することができる。ある特定の実施形態では、多数の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により、データセットの複雑さおよび/または次元性を低下させることができる。例えば、参照データセットに対して多数の操作を使用することにより、参照試料の遺伝的状態（例えば、選択された遺伝的変異について陽性または陰性）に応じた遺伝的変異の有無を表すために使用することができるN次元の空間（例えば、確率プロット）を生成することができる。実質的に同様の操作のセットを使用した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。試験被験体データセットの複雑さおよび/または次元性を、時には、参照データから生成されたN次元の空間と容易に比較することができる単一の値またはN次元の点まで低下させる。参照被験体データが入るN次元の空間の範囲内に入る試験試料データは、遺伝的状態が参照被験体の遺伝的状態と実質的に同様であることを示す。参照被験体データが入るN次元の空間の範囲外の試験試料データは、遺伝的状態が参照被験体の遺伝的状態と実質的に同様でないことを示す。いくつかの実施形態では、参照は正倍数性であるか、そうでなければ遺伝的変異または医学的状態を有さない。

【0204】

いくつかの実施形態では、補正 / 処理ステップは、任意選択で、データセットまたはその誘導物（例えば、当技術分野で公知であり、かつ/または本明細書に記載の1つまたは複数の数学的かつ/または統計学的なデータ処理ステップの生成物）の種々の態様から1つまたは複数のプロファイル（例えば、プロファイルプロット）を生成するステップを含む。プロファイルの生成は、多くの場合、大量のデータのパターンおよび/または関連の同定を容易にする、データの数学的操作および/または統計学的操作を用いることを伴う。プロファイルは、多くの場合、1つまたは複数の基準に基づく、データまたはデータセットの1つまたは複数の操作によって生じる値である。プロファイルは、多くの場合、多数のデータ点を含む。データセットの本質および/または複雑さに応じて、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、プロファイルは、2個以上のデータ点、3個以上のデータ点、5個以上のデータ点、10個以上のデータ点、24個以上のデータ点、25個以上のデータ点、50個以上のデータ点、100個以上のデータ点、500個以上のデータ点、1000個以上のデータ点、5000個以上のデータ点、10,000個以上のデータ点、または100,000個以上のデータ点を含んでよい。

【0205】

いくつかの実施形態では、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す。プロファイルは、時には、フィルタリングされずにいかなるデータも除去されていないデータを表すデータ点を含む、またはそれから生成され、時には、プロファイルは、フィルタリングされて望ましくないデータが除去されたデータを表すデータ点を含む、またはそれから生成される。いくつかの実施形態では、プロファイルにおけるデータ点は、ゲノミックセクションについてのデータ操作の結果を表す。ある特定の実施形態では、プロファイルにおけるデ

10

20

30

40

50

ータ点は、ゲノミックセクションの群についてのデータ操作の結果を表す。いくつかの実施形態では、ゲノミックセクションの群は互いに近接してよく、ある特定の実施形態では、ゲノミックセクションの群は染色体またはゲノムの異なる一部由来であってよい。

【0206】

データセットから得られるプロファイルにおけるデータ点は、任意の適切なデータの 카테고리化を表し得る。データを群分けしてプロファイルデータ点を生成することができるカテゴリの非限定的な例としては、サイズに基づくゲノミックセクション、配列の特徴（例えば、GC含量、AT含量、染色体上の位置（例えば、短腕、長腕、セントロメア、テロメア）など）に基づくゲノミックセクション、発現のレベル、染色体など、またはそれらの組合せが挙げられる。いくつかの実施形態では、プロファイルを、別のプロファイル（例えば、再正規化されたデータプロファイルを生成するために異なる正規化値に対して再正規化（renormalize）された正規化されたデータプロファイル）から得たデータ点から生成することができる。ある特定の実施形態では、別のプロファイルから得た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さが低下する。データ点の数および/またはデータセットの複雑さが低下することにより、多くの場合、データの解釈が容易になり、かつ/またはアウトカムをもたらすことが容易になる。

【0207】

プロファイルは、頻繁にプロットとして示され、生成することができるプロファイルプロットの非限定的な例としては、生のカウント（例えば、生のカウントプロファイルまたは生のプロファイル）、正規化されたカウント（例えば、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プロファイル）、重み付けされたピン、zスコア、p値、フィッティングした倍数性に対する面積比、フィッティングされた胎児分率と測定された胎児分率との間の比に対する中央値の高度、主成分など、またはそれらの組合せが挙げられる。いくつかの実施形態では、プロファイルプロットにより、操作されたデータを可視化することが可能になる。ある特定の実施形態では、プロファイルプロットを利用して、アウトカム（例えば、フィッティングした倍数性に対する面積比、フィッティングされた胎児分率と測定された胎児分率との間の比に対する中央値の高度、主成分）をもたらすことができる。生のカウントプロファイルプロット、または生のプロファイルプロットは、多くの場合、領域（例えば、ゲノム、染色体、染色体の部分）内の総カウントに対して正規化された、領域内の各ゲノミックセクションにおけるカウントのプロットである。いくつかの実施形態では、スタティックウィンドウプロセスを使用してプロファイルを生成ことができ、ある特定の実施形態では、スライディングウィンドウプロセスを使用してプロファイルを生成することができる。

【0208】

試験被験体について生成されたプロファイルを、時には、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/または統計学的操作の解釈および/またはアウトカムをもたらすことを容易にする。いくつかの実施形態では、1つまたは複数の出発仮定（例えば、核酸の母体寄与（例えば、母体分率）、核酸の胎児の寄与（例えば、胎児分率）、参照試料の倍数性など、またはそれらの組合せ）に基づいてプロファイルを生成する。ある特定の実施形態では、検査プロファイルは、多くの場合、遺伝的変異が存在しないことを表す所定の値に集中し、多くの場合、試験被験体が遺伝的変異を保有する場合には、試験被験体における遺伝的変異が位置する遺伝子位置に対応する領域内の所定の値から逸脱する。遺伝的変異に関連付けられる医学的状態のリスクがある、またはそれに罹患している試験被験体では、選択されたゲノミックセクションについての数値は、影響を受けていない遺伝子位置についての所定の値から有意に変動することが予想される。出発仮定（例えば、固定された倍数性もしくは最適化された倍数性、固定された胎児分率もしくは最適化された胎児分率またはそれらの組合せ）に応じて、遺伝的変異の有無を示す所定の閾値またはカットオフ値または値の範囲は変動し得るが、それでも、遺伝的変異の有無を決定するために有用なアウトカムがもたらされ

10

20

30

40

50

る。いくつかの実施形態では、プロファイルにより、表現型が示され、かつ/または表される。

【0209】

非限定的な例として、生の配列読み取りデータから、(a) 1つまたは複数のフローセルからの全ての試料、または1つまたは複数のプレートからの全ての試料について、全ての染色体、選択された染色体、ゲノミックセクションおよび/またはその一部についての総カウントを得るステップと、(b) (i) 情報価値のないゲノミックセクションおよび/または反復性のゲノミックセクション(例えば、反復マスキング; 実施例2に記載の)(ii) G/C含量の偏り(iii) 過大表示または過小表示された配列、(iv) ノイズの多いデータの1つまたは複数を補正、フィルタリングおよび/または除去するステップと、(c) (b) で残ったデータの一部または全部を、選択された染色体または選択された遺伝子位置について、ロバストな推定量を使用して予測値に対して補正/正規化し、それにより、補正/正規化された値を生成するステップとによって、補正/正規化されたデータセットを生成することができる。ある特定の実施形態では、(c) におけるデータを1つまたは複数の指標配列、1つまたは複数の追加的な推定量、1つまたは複数の追加的な処理ステップなど、またはそれらの組合せに対して任意選択で補正する。いくつかの実施形態では、i) 情報価値のないゲノミックセクションおよび/または反復性のゲノミックセクション(例えば、反復マスキング)(ii) G/C含量の偏り(iii) 過大表示または過小表示された配列、(iv) ノイズの多いデータの1つまたは複数を補正、フィルタリングおよび/または除去するステップは、任意の順序で実施することができる(例えば、(i); (ii); (iii); (iv); (i)、(ii); (ii)、(i); (iii)、(i); (ii)、(iii)、(i); (i)、(iv)、(iii); (ii)、(i)(iii); (i)、(ii)、(iii)、(iv); (ii)、(i)、(iii)、(v); (ii)、(iv)、(iii)、(i)など)。ある特定の実施形態では、残ったデータを本明細書に記載の1つまたは複数の実験条件に基づいて補正することができる。いくつかの実施形態では、1つの方法によって補正された配列は、異なる方法によって実質的に完全に補正された配列の一部に影響を及ぼし得る(例えば、G/C含量の偏りの補正により、時には、反復マスキングによって実質的に完全に除去された配列の最大50%が除去される)。

【0210】

カウントされた、マッピングされた配列読み取りデータの1つまたは複数の操作によって、補正/正規化されたデータセットを生成することができる。配列読み取りをマッピングし、各ゲノムのピンに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)。いくつかの実施形態では、マッピングする前に、データセットについて反復マスキング補正して情報価値のないゲノミックセクションおよび/または反復性のゲノミックセクションを除去し、ある特定の実施形態では、マッピングする前に参照ゲノムについて反復マスキング補正する。いずれのマスキング手順を実施することによっても実質的に同じ結果がもたらされる。ある特定の実施形態では、データセットを、染色体の部分または全てについて予測されたG/C配列表示のロバストな推定量に関してピンに関するG/C正規化することによって、G/C含量の偏りについて補正する。いくつかの実施形態では、データセットについて、反復マスキング補正した後にG/C含量補正し、ある特定の実施形態では、データセットについて、G/C含量補正した後に、反復マスキング補正する。補正後に、残ったカウントを、一般には、合計して、補正されたデータセットを生成する。ある特定の実施形態では、データセットを補正することにより、分類、および/またはアウトカムをもたらすことが容易になる。いくつかの実施形態では、補正されたデータセットから補正されたデータセットプロファイルを生成し、分類、および/またはアウトカムをもたらすことを容易にするために利用する。

【0211】

いくつかの実施形態では、配列読み取りデータをカウントし、反復配列、G/C含量の偏り、または反復配列およびG/C含量の偏りについて補正した後、データセットを1つ

10

20

30

40

50

または複数の指標配列について補正することができる。多数の患者由来の試料を、異なる指標配列で標識し、フローセル上で混合することができる。いくつかの実施形態では、患者および指標の間の配列読み取りのマッピングは同形である（どちらの方向にも独特である）。配列決定測定が完了した後、種々の配列決定された断片を、それが由来する個々の患者に割り当てることができる。種々の配列断片間の分離は、多くの場合、断片配列の指標（バーコード）部分に基づいて実現される。同じ指標（バーコード）を担持する断片の実質的に全てが一緒に群分けされ、その指標に関連付けられる患者に帰する。ある特定の実施形態では、各患者試料について同じ手順を繰り返す。少数の断片は、指標を有さないまたは認識されない指標を有する可能性がある（実験的な誤差に起因して）。認識されない指標が予測された指標のうちの1つと同様だと思われる場合には、任意選択でそれらの断片を同様に許容することができ、それ以外には、指標を有さないまたは認識されない指標を有する断片は割り当てずに残す。所与の患者に割り当てられた断片のみを、参照ゲノムと対照してアラインメントし、その特定の患者の染色体表示にカウントする。補正後に、残ったカウントを、一般には、合計して、補正されたデータセットを生成する。ある特定の実施形態では、データセットを補正することにより、分類、および/またはアウトカムをもたらすことが容易になる。いくつかの実施形態では、補正されたデータセットから補正されたデータセットプロファイルを生成し、分類、および/またはアウトカムをもたらすことを容易にするために利用する。

10

【0212】

配列読み取りデータをカウントし、反復配列、G/C含量の偏り、または反復配列およびG/C含量の偏り、ならびに/または指標配列について補正した後、データセットを補正して、フローセルに基づく実験条件および/またはプレートに基づく実験条件の偏りの影響を最小限にするまたは排除することができる。ある特定の実施形態では、データセットを補正することにより、分類、および/またはアウトカムをもたらすことが容易になる。いくつかの実施形態では、補正されたデータセットから補正されたデータセットプロファイルを生成し、分類、および/またはアウトカムをもたらすことを容易にするために利用する。

20

【0213】

データセットを本明細書に記載の通り補正した後、データセットの一部または全部を、下記の1つまたは複数の手順を用いてさらに処理することもできる。いくつかの実施形態では、データセットの一部または全部の追加的な処理は、本明細書に記載の通り、または当技術分野で公知の通りZスコアを生成することを含む。ある特定の実施形態では、Zスコアを、偽データまたは外れ値データの影響を最小限にするロバストなZスコアとして生成する。

30

【0214】

任意選択で、データセットを正規化して、正規化されたカウント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されたゲノミックセクションを適切な正規化参照値に対して正規化することによってデータセットを正規化することができる。いくつかの実施形態では、正規化参照値は、ゲノミックセクションが選択された1つまたは複数の染色体についての総カウントを表す。ある特定の実施形態では、正規化参照値は、遺伝的変異を保有しないことが分かっている参照被験体のセットから調製した参照データセットからの1つまたは複数の対応するゲノミックセクション、染色体の部分または染色体を表す。いくつかの実施形態では、正規化参照値は、遺伝的変異の有無について分析される試験被験体から調製した試験被験体データセットからの1つまたは複数の対応するゲノミックセクション、染色体の部分または染色体を表す。ある特定の実施形態では、スタティックウィンドウ手法を利用して正規化プロセスを実施し、いくつかの実施形態では、ムービングウィンドウ手法またはスライディングウィンドウ手法を利用して正規化プロセスを実施する。ある特定の実施形態では、正規化されたプロファイルプロットを生成して、分類、および/またはアウトカムをもたらすことを容易にする。正規化されたプロファイル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

50

【0215】

いくつかの実施形態では、データセットを任意選択でフィルタリングし、正規化することができ、処理されたデータセットを、1つまたは複数のフィルタリング手順および/または正規化手順によってさらに操作することができる。ある特定の実施形態では、1つまたは複数のフィルタリング手順および/または正規化手順によってさらに操作されたデータセットを使用して、プロファイルを生成することができる。いくつかの実施形態では、1つまたは複数のフィルタリング手順および/または正規化手順により、時には、データセットの複雑さおよび/または次元性を低下させることができる。複雑さおよび/または次元性が低下した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。

【0216】

いくつかの実施形態では、データセットを、重み付けによってさらに操作することができる。ある特定の実施形態では、1つまたは複数のゲノミックセクションを重み付けのために選択して、選択されたゲノミックセクションに含有されるデータ（例えば、ノイズの多いデータ、情報価値のないデータ）の影響を低下させることができ、いくつかの実施形態では、1つまたは複数のゲノミックセクションを重み付けのために選択して、選択されたゲノムのセグメント内に含有されるデータ（例えば、測定された分散が小さいデータ）の影響を増強または増大させることができる。いくつかの実施形態では、分散が大きいデータの影響を減少させ、分散が小さいデータの影響を増加させる単一の重み関数を利用してデータセットを重み付けする。時には、重み関数を使用して、分散が大きいデータの影響を低下させ、分散が小さいデータの影響を増大させる（例えば、 $[1 / (\text{標準偏差})^2]$ ）。いくつかの実施形態では、重み付けによってさらに操作された、処理されたデータのプロファイルプロットを生成して、分類、および/またはアウトカムをもたらすことを容易にする。重み付けされたデータのプロファイル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。

【0217】

いくつかの実施形態では、データセットを、1つまたは複数の数学的かつ/または統計学的な（例えば、統計関数または統計アルゴリズム）操作によってさらに操作することができる。ある特定の実施形態では、処理されたデータセットを、1つまたは複数の選択されたゲノミックセクション、染色体、または染色体の部分についてZスコアを算出することによってさらに操作することができる。いくつかの実施形態では、処理されたデータセットを、P値を算出することによってさらに操作することができる。ZスコアおよびP値を算出するための式は当技術分野で公知である。ある特定の実施形態では、数学的操作および/または統計学的操作は、倍数性および/または胎児分率に関する1つまたは複数の仮定を含む。いくつかの実施形態では、1つまたは複数の統計学的操作および/または数学的操作によってさらに操作された、処理されたデータのプロファイルプロットを生成して、分類、および/またはアウトカムをもたらすことを容易にする。統計学的かつ/または数学的に操作されたデータのプロファイル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。統計学的かつ/または数学的に操作されたデータのプロファイルプロットに基づいてもたらされたアウトカムは、多くの場合、倍数性および/または胎児分率に関する1つまたは複数の仮定を含む。

【0218】

ある特定の実施形態では、データセットをカウントし、任意選択でフィルタリングし、正規化した後に、処理されたデータセットに対して多数の操作を実施してN次元の空間および/またはN次元の点を生成する。N次元で分析されたデータセットのプロファイル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。

【0219】

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、ピーク高度分析、ピーク幅分析、ピーク端位置分析、ピーク側方許容差など、その派生物、または前述のものの中から選択される1つまたは複数のプロセスを利用することによって、データセットをさらに操作することができる。いくつかの実施形態では、1つまたは複数のピ

10

20

30

40

50

ーク高度分析、ピーク幅分析、ピーク端位置分析、ピーク側方許容差など、その派生物、または前述のものの組合せを利用して処理されたデータのプロファイルプロットを生成して、分類、および/またはアウトカムをもたらすことを容易にする。1つまたは複数のピーク高度分析、ピーク幅分析、ピーク端位置分析、ピーク側方許容差など、その派生物、または前述のものの組合せを利用して処理されたデータのプロファイル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。

【0220】

いくつかの実施形態では、問題の遺伝的変異がない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、参照カウントプロファイル中央値を生成することができ、これにより、遺伝的変異が存在しないことを表す所定の値をもたらすことができ、また、これは多くの場合、試験被験体が遺伝的変異を保有する場合には、試験被験体における遺伝的変異が位置する遺伝子位置に対応する領域内の所定の値から逸脱する。遺伝的変異に関連付けられる医学的状態のリスクがある、またはそれに罹患している試験被験体では、選択された1つまたは複数のゲノミックセクションについての数値は、影響を受けない遺伝子位置についての所定の値から有意に変動することが予想される。ある特定の実施形態では、問題の遺伝的変異を保有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、参照カウントプロファイル中央値を生成することができ、これにより、遺伝的変異が存在することを表す所定の値をもたらすことができ、また、これは多くの場合、遺伝的変異を保有しない試験被験体における遺伝子位置に対応する領域内の所定の値から逸脱する。遺伝的変異に関連付けられる医学的状態のリスクがない、またはそれに罹患していない試験被験体では、選択された1つまたは複数のゲノミックセクションについての数値は、影響を受けた遺伝子位置についての所定の値から有意に変動することが予想される。

【0221】

いくつかの実施形態では、データの解析および処理は、1つまたは複数の仮定を使用することを含んでよい。任意の適切な数または種類の仮定を利用してデータセットを解析または処理することができる。データを処理および/または解析する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は、母体の倍数性、胎児の寄与、参照集団における特定の配列の分布率、民族的なバックグラウンド、選択された医学的状態の血縁の家族構成員における分布率、種々の患者からの生のカウントプロファイル間の相似および/またはGC正規化および反復マスキング(例えば、GCRM)後の実行、PCRアーチファクトを表す同一のマッチ(例えば、同一の塩基位置)、胎児定量器アッセイ(fetal quantifier assay)(例えば、FQA)に固有の仮定、双生児に関する仮定(例えば、双生児2体の1体のみが影響を受けた場合、有効な胎児分率は測定された胎児の総分率の50%のみである(同様に三胎児、四胎児など))、ゲノム全体を均一に網羅する胎児の無細胞DNA(例えば、cfDNA)など、およびそれらの組合せが挙げられる。

【0222】

マッピングされた配列読み取りの質および/または深さにより、所望の信頼水準(例えば、95%以上の信頼水準)で遺伝的変異の有無のアウトカム予測が可能にならない例では、正規化されたカウントプロファイルに基づいて、1つまたは複数の追加的な数学的操作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、データを解析し、かつ/またはアウトカムをもたらすために有用な追加的な数値を生成することができる。正規化されたカウントプロファイルは、多くの場合は、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である。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は本明細書に記載されている。記載の通り、カウントされた、マッピングされた配列読み取りを試験試料カウントまたは参照試料カウントに関して正規化することができる。いくつかの実施形態では、正規化されたカウントプロファイルをプロットとして示すことができる。

【0223】

上記の通り、時には、データをある形態から別の形態に変換する。変換されたデータ、

10

20

30

40

50

または変換は、多くの場合、物理的な出発材料（例えば、試験被験体試料核酸および/または参照被験体試料核酸）からのデータを物理的な出発材料（例えば、配列読み取りデータ）のデジタル表示に変更することであり、いくつかの実施形態では、アウトカムをもたらすために利用することができる1つまたは複数のデジタル表示の数値またはグラフ表示にさらに変換することを含む。ある特定の¹⁰実施形態では、デジタル表示されたデータの数値および/またはグラフ表示の1つまたは複数を利用して、試験被験体の物理的なゲノムの様相を表す（例えば、ゲノムの挿入またはゲノムの欠失の有無を仮想的に表すまたは視覚的に表す；医学的状態に関連付けられる配列の物理量の変動の有無を表す）ことができる。時には、仮想表示を出発材料のデジタル表示の1つまたは複数の数値またはグラフ表示にさらに変換する。これらの手順により、物理的な出発材料を数値もしくはグラフ表示、または試験被験体のゲノムの物理的様相の表示に変換することができる。

【0224】

いくつかの実施形態では、データセットを変換することにより、データの複雑さおよび/またはデータの次元性が低下することによって、アウトカムをもたらすことが容易になる。時には、データセットの複雑さを、物理的な出発材料を出発材料の仮想表示（例えば、物理的な出発材料を表す配列読み取り）に変換するプロセスの間に低下させる。任意の適切な特徴または変数を利用して、データセットの複雑さおよび/または次元性を補正し、かつ/または低下させることができる。データを補正/処理するための標的特徴として使用するために選択することができる特徴の非限定的な例としては、フローセルに基づく実験条件および/またはプレートに基づく実験条件、GC含量、反復配列、指標配列、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、がんの同定、疾患、遺伝性の遺伝子/形質、染色体異常、生物学的カテゴリー、化学的カテゴリー、生化学的カテゴリー、遺伝子またはタンパク質のカテゴリー、遺伝子オントロジー、タンパク質オントロジー、同時調節される遺伝子、細胞シグナル伝達遺伝子、細胞周期遺伝子、前述の遺伝子に関するタンパク質、遺伝子変異体、タンパク質変異体、同時調節される遺伝子、同時調節されるタンパク質、アミノ酸配列、ヌクレオチド配列、タンパク質の構造データなど、および前述のものの組合せが挙げられる。データセットの複雑さおよび/または次元性の低下の非限定的な例としては、複数の配列読み取りをプロファイルプロットに低下させること、複数の配列読み取りを数値（例えば、正規化された値、Zスコア、ロバストなZスコア、p値、中央絶対偏差、または本明細書に記載のMADの代替値）に低下させること；多数の分析方法を確率プロットまたは単一の点に低下させること；誘導量の主成分分析など、またはそれらの組合せが挙げられる。

【0225】

アウトカム

データの分析、補正および処理により、1つまたは複数のアウトカムをもたらすことができる。アウトカムは、多くの場合、被験体が遺伝的変異を有するリスクがあった、またはそのリスクがあるかどうかを決定することを容易にするデータの補正および処理の結果である。アウトカムは、多くの場合、1つまたは複数の確率または推定量の考察に関連して本明細書に記載の補正/処理方法を使用して生成した1つまたは複数の数値を含む。確率の考察としては、これだけに限定されないが、変動性の尺度、信頼水準、感度、特異度、標準偏差、変動係数（CV）および/または信頼水準、Zスコア、ロバストなZスコア、パーセント染色体表示、中央絶対偏差、または中央絶対偏差の代替値、カイ値、ファイ値、倍数性値、胎児分率、フィッティングした胎児分率、面積比、中央値の高度など、またはそれらの組合せが挙げられる。確率の考察により、被験体が遺伝的変異を有するリスクがある、または遺伝的変異を有するかどうかを決定することを容易にすることができ、遺伝的障害の有無を決定するアウトカムは、多くの場合そのような考察を含む。

【0226】

いくつかの実施形態では、アウトカムは、試料核酸中の胎児核酸の分率をファクタリング（factoring）すること（例えば、カウントを補正すること、試料を除去することまたは呼び出しを行わないこと）を含む。胎児分率の決定は、時には、本明細書の実

10

20

30

40

50

施例に記載されており、当技術分野で公知の胎児定量器アッセイ（FQA）を使用して実施する（例えば、その全体が参照により本明細書に組み込まれる「PROCESSES AND COMPOSITIONS FOR METHYLATION-BASED ENRICHMENT OF FETAL NUCLEIC ACIDS」という表題の米国特許出願公開第2010-0105049A1号）。

【0227】

アウトカムは、多くの場合、関連する信頼度の水準を伴う表現型である（例えば、胎児は99%の信頼水準で21トリソミーについて陽性であり、試験被験体は95%の信頼水準で遺伝的変異に関連付けられるがんについて陰性である）。アウトカムの値を生成する異なる方法により、時には、異なる種類の結果が生じ得る。一般に、本明細書に記載の方法を使用して生成したアウトカムの値に基づいて生じさせることができる、可能性のあるスコアまたは呼び出しは、真陽性、偽陽性、真陰性および偽陰性の4種類ある。スコアまたは呼び出しは、多くの場合、被験体/試料において特定の遺伝的変異が存在する確率または存在しない確率を算出することによって生成する。スコアの値を用いて、例えば、遺伝的変異に対応する可能性があるマッピングされた配列読み取りの変動、差異、または比率を決定することができる。例えば、参照ゲノムに対して、データセットから選択された遺伝的変異またはゲノミックセクションについて陽性のスコアを算出することにより、遺伝的変異の有無の同定を導くことができ、遺伝的変異は、時には、医学的状態（例えば、がん、子癇前症、トリソミー、モノソミーなど）に関連付けられる。ある特定の実施形態では、補正されたデータセットからアウトカムを生成する。いくつかの実施形態では、遺伝的変異の有無および/または胎児の異数性を決定する、もたらされるアウトカムは、正規化された試料カウントに基づく。いくつかの実施形態では、アウトカムはプロファイルを含む。アウトカムがプロファイルを含む実施形態では、任意の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することができるプロファイルの非限定的な例としては、zスコアプロファイル、ロバストな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、ファイ値プロファイルなど、およびそれらの組合せが挙げられる。

【0228】

遺伝的変異の有無を決定するために生成されたアウトカムは、時にはヌル結果（例えば、2つのクラスター間のデータ点、遺伝的変異の有無のどちらの値も包含する標準偏差を伴う数値、調査されている遺伝的変異を有する被験体または有さない被験体のプロファイルプロットと同様ではないプロファイルプロットを伴うデータセット）を含む。いくつかの実施形態では、ヌル結果を示すアウトカムは、それでも、決定力のある結果であり、決定は、遺伝的変異の有無を決定するための追加的な情報ならびに/またはデータの生成および/もしくは解析を繰り返す必要性を含む可能性がある。

【0229】

いくつかの実施形態では、本明細書に記載の1つまたは複数の処理ステップを実施した後、アウトカムを生成することができる。ある特定の実施形態では、本明細書に記載の処理ステップのうちの1つの結果としてアウトカムを生成し、いくつかの実施形態では、データセットの統計学的操作および/または数学的操作のそれぞれを実施した後、アウトカムを生成する。遺伝的変異の有無の決定に関わるアウトカムは、任意の適切な形態で表すことができ、その形態は、これだけに限定することなく、被験体または試料について、遺伝的変異の有無に関連付けられる確率（例えば、オッズ比、p値）、尤度、クラスターの範囲内または範囲外の値、閾値を超えるまたは下回る値、分散または信頼度の尺度を伴う値、または危険因子を含む。ある特定の実施形態では、試料間の比較により、試料の同一性を確認することが可能になる（例えば、反復試料および/または混同された（例えば、誤って標識付けられた、混合されたなど）試料を同定することが可能になる）。

【0230】

いくつかの実施形態では、アウトカムは、所定の閾値またはカットオフ値を上回るまたは下回る値（例えば、1超、1未満）、およびその値に関連付けられる不確実性または信

10

20

30

40

50

頼水準を含む。アウトカムにより、データ処理において使用される任意の仮定を説明することもできる。ある特定の実施形態では、アウトカムは、所定の値の範囲内に入るまたはその範囲外の値、および範囲内または範囲外にあるその値に関連する不確実性または信頼水準を含む。いくつかの実施形態では、アウトカムは、所定の値と等しい（例えば、1と等しい、ゼロと等しい）、または所定の値の範囲内の値と等しい値、および範囲と等しいまたは範囲内にあるまたは範囲外にあるその値の関連する不確実性または信頼水準を含む。アウトカムは、時には、プロット（例えば、プロファイルプロット）としてグラフ表示される。

【0231】

上記の通り、アウトカムを、真陽性、真陰性、偽陽性または偽陰性と特徴付けることができる。真陽性とは、遺伝的変異を有すると正確に診断された被験体を指す。偽陽性とは、遺伝的変異を有すると間違っ同定された被験体を指す。真陰性とは、遺伝的変異を有しないと正確に同定された被験体を指す。偽陰性とは、遺伝的変異を有しないと間違っ同定された被験体を指す。任意の所与の方法の性能の2つの尺度を、これらが出現する比率に基づいて算出することができる：(i)一般に、陽性であると正確に同定される予測陽性の分率である感度値；および(ii)一般に、陰性であると正確に同定される予測陰性の分率である特異度値。感度とは、一般に、真陽性の数を、真陽性の数と偽陰性の数を足した数で割った数であり、感度(sens)は0 sens 1の範囲内であり得る。理想的には、偽陰性の数はゼロと等しいまたはほぼゼロであり、したがって、被験体が実際に少なくとも1つの遺伝的変異を有する場合に、少なくとも1つの遺伝的変異を有しないと間違っ同定される被験体はいない。逆に、多くの場合、予測アルゴリズムの、陰性を正確に分類する能力に関して評価を行い、これは感度に対する補完的な測定である。特異度とは、一般に、真陰性の数を、真陰性の数と偽陽性の数を足した数で割った数であり、感度(spec)は0 spec 1の範囲内であり得る。理想的には、偽陽性の数はゼロと等しいまたはほぼゼロであり、したがって、被験体が評価されている遺伝的変異を有さない場合に、少なくとも1つの遺伝的変異を有すると間違っ同定される被験体はいない。

【0232】

ある特定の実施形態では、感度、特異度および/または信頼水準の1つまたは複数 は百分率として表される。いくつかの実施形態では、百分率は、各変数について、それぞれ独立に、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%以上、約99.9%以上、約99.95%以上、約99.99%以上））である。いくつかの実施形態では、変動係数(CV)は百分率として表され、時には、百分率は約10%以下（例えば、約10%、9%、8%、7%、6%、5%、4%、3%、2%または1%、または1%未満（例えば、約0.5%以下、約0.1%以下、約0.05%以下、約0.01%以下））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よらない確率）は、Zスコア、p値、またはt検定の結果で表される。いくつかの実施形態では、本明細書に記載の1つまたは複数のデータ処理操作を用いて、アウトカムについての測定された分散、信頼区間、感度、特異度など（例えば、集合的に信頼度パラメータと称される）を生成することができる。

【0233】

時には、感度および特異度が1と等しい、または100%、またはほぼ1である（例えば、約90%から約99%の間）方法を選択する。いくつかの実施形態では、感度が1と等しい、または100%である方法を選択し、ある特定の実施形態では、感度がほぼ1である方法を選択する（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）。いくつかの実施形態では、特異度が1と等しい、または100%である方法を選択し、ある特定の実施形態では、特異度がほぼ1である方法を選択する（例えば、約90%の特異度、約91%の特異度、約92%の特異度、約9

10

20

30

40

50

3%の特異度、約94%の特異度、約95%の特異度、約96%の特異度、約97%の特異度、約98%の特異度、または約99%の特異度)。

【0234】

いくつかの実施形態では、カウントされた、マッピングされた配列読み取りまたはその誘導物に基づくアウトカムにより、表1Aおよび1Bに列挙されている1つまたは複数の状態、症候群または異常の有無が決定される。ある特定の実施形態では、本明細書に記載の1つまたは複数のデータ処理方法を利用して生成したアウトカムにより、表1Aおよび1Bに列挙されている1つまたは複数の状態、症候群または異常の有無が決定される。いくつかの実施形態では、状態、症候群または異常の有無を決定するアウトカムは、表1Aおよび1Bに列挙されている状態、症候群または異常の検出である、またはそれを含む。

10

【0235】

ある特定の実施形態では、アウトカムは、試験試料と参照試料、試験試料と他の試料、2つ以上の試験試料など、およびそれらの組合せの比較に基づく。いくつかの実施形態では、試料間を比較することにより、アウトカムをもたらすことが容易になる。ある特定の実施形態では、アウトカムは、本明細書に記載の通りまたは当技術分野で公知の通り生成されたZスコアに基づく。いくつかの実施形態では、Zスコアは、正規化された試料カウントを使用して生成する。いくつかの実施形態では、アウトカムをもたらすことを容易にするために生成されたZスコアは、ロバストな推定量を使用して生成したロバストなZスコアである。ある特定の実施形態では、アウトカムは、正規化された試料カウントに基づく。

20

【0236】

1つまたは複数のアウトカムを生成した後、多くの場合、アウトカムを用いて、遺伝的変異の有無および/または関連する医学的状态の決定をもたらす。アウトカムは、一般には、医療専門家(例えば、検査技師または管理者; 医師または補助者)に提供される。いくつかの実施形態では、遺伝的変異の有無を決定するアウトカムは、報告書の形態で医療専門家に提供され、ある特定の実施形態では、報告書は、アウトカムの値および関連する信頼度パラメータの提示を含む。一般に、アウトカムは、遺伝的変異の有無および/または医学的状态の決定を容易にする任意の適切な形態で提示することができる。データセットを報告および/もしくは提示する、またはアウトカムを報告するために使用するために適した形式の非限定的な例としては、デジタルデータ、グラフ、2Dグラフ、3Dグラフ、および4Dグラフ、写真、ピクトグラフ、チャート、棒グラフ、円グラフ、図、フローチャート、散布プロット、マップ、ヒストグラム、密度チャート、関数グラフ、回路図、ブロック図、バブルマップ、信号空間ダイアグラム、コンターダイアグラム、統計地図、スパイダーチャート、ベン図、計算図表など、および前述のものの組合せが挙げられる。アウトカム表示の種々の例が図に示されており、実施例に記載されている。

30

【0237】

アウトカムの使用

遺伝的変異の有無を決定する1つまたは複数のアウトカムを含む報告を受け取る医療専門家、または他の資格のある個体は、報告書において表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関する呼び出しを作成することができる。いくつかの実施形態では、医療専門家は、もたらされたアウトカムに基づいて推奨を行うことができる。いくつかの実施形態では、医療専門家または資格のある個体は、報告書において提供される1つまたは複数のアウトカムの値および関連する信頼度パラメータに基づいて、試験被験体または患者に遺伝的変異の有無に関する呼び出しまたはスコアを提供することができる。ある特定の実施形態では、スコアまたは呼び出しは、医療専門家または資格のある個体によって、提供された報告書を目で観察することを用いて手動で作成される。ある特定の実施形態では、スコアまたは呼び出しは、時にはソフトウェアに埋め込まれた自動化されたルーチンによって作成され、情報を試験被験体または患者に提供する前に、正確を期するために医療専門家または資格のある個体によって精査される。

40

【0238】

50

報告書を受け取ることは、多くの場合、通信手段によって、アウトカムを含むテキスト表示および/またはグラフ表示を入手することを伴い、これにより、医療専門家または他の資格のある個体が試験被験体または患者における遺伝的変異の有無に関する決定を行うことが可能になる。報告書は、コンピュータによってまたは人によるデータ入力によって生成することができ、電子的手段を使用して（例えば、インターネット上で、コンピュータを介して、ファックスを介して、あるネットワークの場所から同じまたは異なる物理的な場所にある別の場所へ）、またはデータを送受信する任意の他の方法によって（例えば、メールサービス、クーリエサービスなど）通信することができる。いくつかの実施形態では、アウトカムを、これだけに限定することなく、口頭、文書、またはファイル形態を含めた適切な媒体で医療専門家に伝達する。ファイルは、例えば、これだけに限定されないが、聴覚的ファイル、コンピュータ可読ファイル、紙ファイル、検査ファイルまたは医療記録ファイルであってよい。アウトカム情報は、検査ファイルから入手することもできる。検査ファイルは、医学的状態の有無を決定するために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行う研究所で生成することができる。研究所は、検査ファイルから医学的状態の有無を同定する人員と同じ場所にあっても異なる場所（例えば、別の国）にあってもよい。例えば、検査ファイルのある場所で生成し、別の場所に伝達し、そこでそのファイル内の情報を妊婦被験体に伝達することができる。ある特定の実施形態では、検査ファイルは、有形であっても電子形態（例えば、コンピュータ可読の形態）であってよい。

10

【0239】

20

医療専門家または資格のある個体は、報告書において提供される1つまたは複数のアウトカムに基づいて、任意の適切な推奨を提供することができる。提供されたアウトカム報告書に基づいて提供することができる推奨の非限定的な例としては、外科手術、放射線療法、化学療法、遺伝相談、出生後治療解決（例えば、生活設計、長期補助介護、医薬品、対症療法）、妊娠中絶、臓器移植、輸血など、または前述のものの組合せが挙げられる。いくつかの実施形態では、推奨は、提供されるアウトカムに基づく分類に左右される（例えば、ダウン症候群、ターナー症候群、T13遺伝的変異に関連付けられる医学的状態、T18遺伝的変異に関連付けられる医学的状態）。

【0240】

これだけに限定されないが、カウントするステップと、データ処理ステップと、アウトカムを生成するステップと、および/または生成されたアウトカムに基づいて1つまたは複数の推奨を提供するステップを含めた本明細書に記載のプロセスの1つまたは複数のステップは、ソフトウェアを使用して実施することができる。

30

【0241】**機械、ソフトウェアおよびインターフェース**

本明細書に記載の方法を行うために、装置、ソフトウェアおよびインターフェースを使用することができる。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、例えば、統計解析アルゴリズム、統計有意性アルゴリズム、統計アルゴリズム、反復ステップと、検証アルゴリズム、およびグラフ表示を実行することを伴ってよい特定の情報、プログラムまたはプロセス（例えば、配列読み取りをマッピングすること、マッピングされたデータを処理すること、および/またはアウトカムをもたらすこと）を使用するためのオプションを入力、要求、問い合わせまたは決定することができる。いくつかの実施形態では、データセットをユーザーがインプット情報として入力することができ、ユーザーは1つまたは複数のデータセットを任意の適切なハードウェア媒体（例えば、フラッシュドライブ）によってダウンロードすることができ、かつ/またはユーザーは、データセットを、その後処理し、かつ/またはアウトカムをもたらすために、あるシステムから別のシステムに送信する（例えば、配列読み取りをマッピングするために、シークエンサーからの配列読み取りデータをコンピュータシステムに送信する；処理し、アウトカムをもたらす、かつ/または報告するために、マッピングされた配列データをコンピュータシステムに送信する）ことができる。

40

50

【0242】

ユーザーは、例えば、ソフトウェアに問い合わせし、次いでそれによりインターネットアクセスを介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により促して、与えられたパラメータに基づいて適切なデータセットを取得することができる。同様に、プログラム可能なプロセッサにより、与えられたパラメータに基づいてプロセッサによって選択された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を促すことができる。プログラム可能なプロセッサにより、インターネット、他の内部または外部の情報などを介して見いだされた情報に基づいてプロセッサによって選択された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を促すことができる。オプションは、1つまたは複数のデータの特徴の選択、1つまたは複数の統計アルゴリズム、1つまたは複数の統計解析アルゴリズム、1つまたは複数の統計有意性アルゴリズム、1つまたは複数のロバストな推定量アルゴリズム、反復ステップと、1つまたは複数の検証アルゴリズム、および方法、装置、またはコンピュータプログラムの1つまたは複数のグラフ表示を選択するように選択することができる。

10

【0243】

本発明で扱われるシステムは、コンピュータシステムの一般的な構成部分、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、手持ち型システム、携帯情報端末、コンピュータキオスクなどを含んでよい。コンピュータシステムは、キーボード、タッチスクリーン、マウス、音声認識またはユーザーがシステムにデータを入力することを可能にする他の手段などの1つまたは複数のインプット手段を含んでよい。システムは、これだけに限定されないが、表示画面（例えば、CRTまたはLCD）、スピーカー、FAX装置、プリンター（例えば、レーザー、インクジェット、インパクト、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/または報告）の視覚的アウトプット、聴覚的アウトプットおよび/またはハードコピーアウトプットをもたらすために有用な他のアウトプットを含めた、1つまたは複数のアウトプットをさらに含んでよい。

20

【0244】

システムにおいて、インプット手段およびアウトプット手段は、他の構成部分の中でも、プログラムの命令を実行するためのマイクロプロセッサおよびプログラムコードおよびデータを記憶するためのメモリを含んでよい中央処理装置に接続されていてよい。いくつかの実施形態では、プロセスは、単一の地理的な場所に位置するシングルユーザーシステムとして実行することができる。ある特定の実施形態では、プロセスは、マルチユーザーシステムとして実行することができる。マルチユーザー実行の場合では、多数の中央処理装置をネットワークによって接続することができる。ネットワークは、建物の部分の単一の部門、建物全体を包含するローカルのもの、多数の建物にわたるもの、地域にわたるもの、国全体にわたるもの、または世界的なものであってよい。ネットワークは、プライベートであっても、プロバイダーによって所有および制御されていても、ユーザーがウェブページにアクセスして情報を入力し、検索するインターネットに基づくサービスとして実行されてもよい。したがって、ある特定の実施形態では、システムは、ユーザーに対してローカルであってもリモートであってもよい1つまたは複数の機械を含む。1つの場所または多数の場所にある2つ以上の機械にユーザーがアクセスすることができ、データを順番に、かつ/または並行してマッピングし、かつ/または処理することができる。したがって、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」計算プラットフォームなどの多数の機械を使用してデータをマッピングし、かつ/または処理するために、任意の適切な設定および制御を利用することができる。

30

40

【0245】

いくつかの実施形態では、装置は、本明細書に記載されているコンピュータプログラム製品が実行される、ウェブに基づくシステムを含んでよい。ウェブに基づくシステムは、時には、ウェブに基づく機能性のために十分なコンピュータ、通信機器（例えば、通信インターフェース、ルーター、ネットワークスイッチ）などを含む。ある特定の実施形態で

50

は、ウェブに基づくシステムは、ネットワーククラウドコンピューティング、ネットワーククラウドストレージまたはネットワーククラウドコンピューティングおよびネットワーククラウドストレージを含む。ネットワーククラウドストレージは、一般に、インターネット上に位置する仮想サーバー上の、ウェブに基づくデータストレージである。ネットワーククラウドコンピューティングは、一般に、リモートネットワーク環境で起こるネットワークに基づくソフトウェアおよび/またはハードウェアの使用である（例えば、リモートサーバー上に位置する少数のために使用するために利用可能なソフトウェア）。いくつかの実施形態では、本明細書に記載されているコンピュータプログラム製品の1つまたは複数の機能は、ウェブに基づく環境で実行される。

【0246】

いくつかの実施形態では、システムは、通信インターフェースを含んでよい。通信インターフェースにより、ソフトウェアおよびデータをコンピュータシステムと1つまたは複数の外部のデバイスの間で移行することが可能になる。通信インターフェースの非限定的な例としては、モデム、ネットワークインターフェース（例えば、イーサネット（登録商標）カードなど）、通信ポート、PCMCIAスロットおよびカードなどが挙げられる。通信インターフェースを介して移行されたソフトウェアおよびデータは、一般に、シグナルの形態であり、電子シグナル、電磁気シグナル、光学シグナルおよび/または通信インターフェースが受け取ることができる他のシグナルであってよい。シグナルは、多くの場合、チャンネルを介して通信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、シグナルを保持し、伝線またはケーブル、光ファイバー、電話線、携帯電話リンク、RFリンクおよび/または他の通信チャンネルによって実行することができる。したがって、ある例では、通信インターフェースを用いて、シグナル検出モジュールによって検出することができるシグナル情報を受け取ることができる。

【0247】

データは、これだけに限定されないが、手動入力デバイスまたは直接データ入力デバイス（DDE）を含めた任意の適切なデバイスおよび/または方法によって入力することができる。手動デバイスの非限定的な例としては、キーボード、コンセプトキーボード、タッチセンシティブスクリーン、ライトペン、マウス、トラックボール、ジョイスティック、グラフィックタブレット、スキャナ、デジタルカメラ、ビデオデジタイザおよび音声認識デバイスが挙げられる。DDEの非限定的な例としては、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。

【0248】

いくつかの実施形態では、配列決定装置からのアウトプットが、入力デバイスによって入力することができるデータとしての機能を果たし得る。ある特定の形態では、マッピングされた配列読み取りが、入力デバイスによって入力することができるデータとしての機能を果たし得る。ある特定の形態では、*in silico* プロセスによってシミュレートされたデータを生成し、シミュレートされたデータが、入力デバイスによって入力することができるデータとしての機能を果たす。本明細書で使用される場合、「*in silico*」とは、コンピュータを使用して実施される試験および実験を指す。*in silico* プロセスは、これだけに限定されないが、本明細書に記載のプロセスに応じて配列読み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読み取りを処理することを含む。

【0249】

システムは、本明細書に記載のプロセスを実施するために有用なソフトウェアを含んでよく、ソフトウェアは、そのようなプロセスを実施するための1つまたは複数のモジュール（例えば、データ取得モジュール、データ処理モジュール、データディスプレイモジュール）を含んでよい。ソフトウェアは、多くの場合、コンピュータによって遂行される場合、コンピュータ操作を実施するコンピュータ可読プログラム命令である。モジュールは、多くの場合、より大きなソフトウェアシステムにおいて使用することができる自蔵式機

10

20

30

40

50

能単位である。例えば、ソフトウェアモジュールは、特定のプロセスまたはタスクを実施するプログラムの一部である。

【0250】

ソフトウェアは、多くの場合、これだけに限定されないが、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めた磁気媒体；ならびにCD-ROMディスク、DVDディスク、光磁気ディスク、フラッシュドライブ、RAM、フロッピー（登録商標）ディスクを含めた光学媒体など、ならびにプログラムの命令を記録することができる他のそのような媒体を含めたコンピュータ可読の媒体上に記録されたプログラムの命令を含有するプログラム製品で提供される。オンライン実行では、組織によって維持されているサーバーおよびウェブサイトを、ソフトウェアダウンロードがリモートユーザーに提供されるように構成することもでき、リモートユーザーが、組織によって維持されているリモートシステムにアクセスして、ソフトウェアにリモートアクセスすることもできる。ソフトウェアは、インプット情報を入手または受け取ることができる。ソフトウェアは、データを特異的に入手または受け取るモジュール（例えば、配列読み取りデータおよび/またはマッピングされた読み取りデータを受け取るデータ受信モジュール）を含んでよく、データを特異的に補正および/または処理するモジュール（例えば、受け取ったデータを補正および/または処理する（例えば、フィルタリングする、正規化する、アウトカムをもたらす、および/または報告する）処理モジュール）を含んでよい。インプット情報を入手および/または受け取ることは、多くの場合、コンピュータ通信手段によってロカールサイトまたはリモートサイトから、人によるデータ入力によって、またはデータを受け取る任意の他の方法によってデータ（例えば、配列読み取り、マッピングされた読み取り）を受け取ることを伴う。インプット情報は、それを受け取る場所と同じ場所で生成することもでき、異なる場所で生成し、受信場所に伝達することもできる。いくつかの実施形態では、インプット情報を処理する前に、それを改変する（例えば、処理に適した形式にする（例えば、表にする））。

【0251】

いくつかの実施形態では、例えば、コンピュータ可読プログラムコードが組み込まれたコンピュータで使用可能な媒体を含むコンピュータプログラム製品であって、コンピュータ可読プログラムコードが、（a）試験被験体由来の試料核酸の配列読み取りを得るステップと、（b）（a）で得られた配列読み取りを、ゲノミックセクションに分けられた既知のゲノムにマッピングするステップと、（c）ゲノミックセクション内にマッピングされた配列読み取りをカウントするステップと、（d）（c）で得られたゲノミックセクションについてのカウントまたはカウントの誘導値を補正することにより、補正されたデータセットを生成するステップと、（e）（d）の補正されたカウントプロファイルから遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む方法の実行が遂行されるように適合されているコンピュータプログラム製品などのコンピュータプログラム製品が提供される。

【0252】

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んでよい。アルゴリズムは、命令の有限列に応じてデータを処理し、かつ/またはアウトカムもしくは報告をもたらすために使用することができる。アルゴリズムは、多くの場合、タスクを完了するための定義済みの命令の一覧である。最初の状態から開始して、命令により、定義済みの一連の連続的な状態を通して進行し、最終的に、最終的な終わりの状態で終結する計算を記述することができる。ある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、いくつかのアルゴリズムにはランダム性が組み込まれている）。例として、これだけに限定することなく、アルゴリズムは、探索アルゴリズム、分類アルゴリズム、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（computational genomic algorithm）、コンビナトリアルアルゴリズム、機械学習アルゴリズム、暗号化アルゴリズム、データ圧縮アルゴリズム、構文解析アルゴリズム

10

20

30

40

50

ムなどであってよい。アルゴリズムは、1つのアルゴリズムまたは組み合わせさせて動作する2つ以上のアルゴリズムを含んでよい。アルゴリズムは、任意の適切な複雑さのクラスおよび/またはパラメータ化された複雑さのものであってよい。アルゴリズムは、算出および/またはデータ処理のために使用することができ、いくつかの実施形態では、確定的手法または確率的/予測的手法において使用することができる。アルゴリズムは、適切なプログラミング言語を使用することによって計算環境において実行することができ、適切なプログラミング言語の非限定的な例は、C、C++、Java（登録商標）、Perl、Python、Fortranなどである。いくつかの実施形態では、アルゴリズムは、誤差限界、統計解析、統計的有意性、および/または他の情報またはデータセットに対する比較（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場合に適用可能）を含むように構成または変更することができる。

10

【0253】

ある特定の実施形態では、いくつかのアルゴリズムは、使用するためにソフトウェアに実装することができる。いくつかの実施形態では、これらのアルゴリズムは、生のデータを用いて訓練することができる。新しい生データ試料のそれぞれについて、訓練されたアルゴリズムにより、代表的な補正および/または処理されたデータセットまたはアウトカムを生じさせることができる。補正または処理されたデータセットは、時には、処理された親データセットと比較して複雑さが低下したものである。いくつかの実施形態では、補正および/または処理されたセットに基づいて、訓練されたアルゴリズムの性能を感度および特異度に基づいて評価することができる。ある特定の実施形態では、感度および/または特異度が最も高いアルゴリズムを同定し、利用することができる。

20

【0254】

ある特定の実施形態では、シミュレートされた（またはシミュレーション）データが、例えば、アルゴリズムを訓練することまたはアルゴリズムを検定することにより、データの補正および/または処理の助けになり得る。いくつかの実施形態では、シミュレートされたデータは、配列読み取りの異なる群分けの仮定上の種々の標本抽出を含む。シミュレートされたデータは、実際の母集団から予測することができるものに基づいてもよく、アルゴリズムを検定し、かつ/または正しい分類を割り当てることに偏っていてもよい。シミュレートされたデータは、本明細書では、「仮想」データとも称される。ある特定の実施形態では、シミュレーションをコンピュータプログラムによって実施することができる。シミュレートされたデータセットの使用における1つの可能性のあるステップは、同定された結果の信頼度、例えば、無作為抽出がどのくらいよく元のデータとマッチするまたはそれを最もよく表すかを評価することである。1つの手法は、選択された標本よりもよいスコアを有する無作為標本の確率を推定する確率の値（P値）を算出することである。いくつかの実施形態では、少なくとも1つの標本が参照標本とマッチする（分解された変動を伴う、または伴わない）ことが想定される経験的なモデルを評価することができる。いくつかの実施形態では、例えばポアソン分布などの別の分布を使用して、確率分布を定義することができる。

30

【0255】

ある特定の実施形態では、システムは1つまたは複数のプロセッサを含んでよい。プロセッサは通信バスに接続されていてよい。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）を含んでよく、補助メモリも含んでよい。補助メモリは、例えば、ハードディスクドライブおよび/または、フロッピー（登録商標）ディスクドライブ、磁気テープドライブ、光ディスクドライブ、メモリカードなどを表す取り外し可能なストレージドライブを含んでよい。取り外し可能なストレージドライブは、多くの場合、取り外し可能な記憶装置からの読み取りおよび/またはそれへの書き込みである。取り外し可能な記憶装置の非限定的な例としては、例えば、取り外し可能なストレージドライブによって読み取ることおよび書き込むことができるフロッピー（登録商標）ディスク、磁気テープ、光ディスクなどが挙げられる。取り外し可能な記憶装置は、コンピュータソフトウェアおよび/またはデータが記憶されたコンピュータ - 使用可能な記

40

50

憶媒体を含んでよい。

【0256】

プロセッサにより、システム内のソフトウェアを実行することができる。いくつかの実施形態では、プロセッサを、ユーザーが実施することができる本明細書に記載のタスクを自動的に実施するようにプログラミングすることができる。したがって、そのようなプロセッサによって行われるプロセッサ、またはアルゴリズムには、ユーザーからの監督またはインプットをほとんど～全く必要なくすることができる（例えば、ソフトウェアを、機能を自動的に実行するようにプログラミングすることができる）。いくつかの実施形態では、プロセスの複雑さは、単一人または人の群がプロセスを遺伝的変異の有無を決定するアウトカムをもたらすために十分に短い時間枠で実施することができないほど大きい。

10

【0257】

いくつかの実施形態では、補助メモリは、コンピュータプログラムまたは他の命令をコンピュータシステムにロードすることを可能にするための他の同様の手段を含んでよい。例えば、システムは、取り外し可能な記憶装置およびインターフェースデバイスを含んでよい。そのようなシステムの非限定的な例としては、プログラムカートリッジおよびカートリッジインターフェース（例えば、テレビゲームデバイスに見いだされるものなど）、取り外し可能なメモリチップ（例えば、EPROM、またはPROMなど）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取り外し可能な記憶装置からコンピュータシステムに移行することを可能にする他の取り外し可能な記憶装置およびインターフェースが挙げられる。

20

【0258】

変換

上記の通り、時には、データのある形態から別の形態に変換する。「変換された（transformed）」、「変換（transformation）」という用語、およびその文法上の派生語または等価な語は、本明細書で使用される場合、物理的な出発材料（例えば、試験被験体試料核酸および/または参照被験体試料核酸）からのデータを物理的な出発材料（例えば、配列読み取りデータ）のデジタル表示に変更することを指し、いくつかの実施形態では、アウトカムをもたらすために利用することができる1つまたは複数のデジタル表示の数値またはグラフ表示にさらに変換することを含む。ある特定の実施形態では、デジタル表示されたデータの数値および/またはグラフ表示の1つまたは複数を利用して、試験被験体の物理的なゲノムの様相を表すことができる（例えば、ゲノムの挿入、重複または欠失の有無を仮想的に表すまたは視覚的に表す；医学的狀態に関連付けられる配列の物理量の変動の有無を表す）。時には、仮想表示を、1つまたは複数の、出発材料のデジタル表示の数値またはグラフ表示にさらに変換する。これらの手順により、物理的な出発材料を数値もしくはグラフ表示、または試験被験体のゲノムの物理的様相の表示に変換することができる。

30

【0259】

いくつかの実施形態では、データセットを変換することにより、データの複雑さおよび/またはデータの次元性が低下することによって、アウトカムをもたらすことが容易になる。時には、データセットの複雑さを、物理的な出発材料を出発材料の仮想表示（例えば、物理的な出発材料を表す配列読み取り）に変換するプロセスの間に低下させる。適切な特徴または変数を利用して、データセットの複雑さおよび/または次元性を低下させることができる。データを処理するための標的特徴として使用するために選択することができる特徴の非限定的な例としては、GC含量、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、がんの同定、疾患、遺伝性の遺伝子/形質、染色体異常、生物学的カテゴリー、化学的カテゴリー、生化学的カテゴリー、遺伝子またはタンパク質のカテゴリー、遺伝子オントロジー、タンパク質オントロジー、同時調節される遺伝子、細胞シグナル伝達遺伝子、細胞周期遺伝子、前述の遺伝子に関するタンパク質、遺伝子変異体、タンパク質変異体、同時調節される遺伝子、同時調節されるタンパク質、アミノ酸配列、ヌクレオチド配列、タンパク質の構造データなど、および前述のもの組合せ

40

50

が挙げられる。データセットの複雑さおよび/または次元性の低下の非限定的な例としては、複数の配列読み取りをプロファイルプロットに低下させること、複数の配列読み取りを数値（例えば、正規化された値、Zスコア、p値）に低下させること；多数の分析方法を確率プロットまたは単一の点に低下させること；誘導量の主成分分析など、またはそれらの組合せが挙げられる。

【0260】

ゲノミックセクション正規化のシステム、装置およびコンピュータプログラム製品

ある特定の態様では、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って遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた試験被験体由来の循環している無細胞試料核酸の配列読み取りのカウントを含み、1つまたは複数のプロセッサによって遂行可能な命令が、(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、(b) 正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されているシステムが提供される。

10

【0261】

ある特定の態様では、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って遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた試験被験体由来の循環している無細胞試料核酸の配列読み取りのカウントを含み、1つまたは複数のプロセッサによって遂行可能な命令が、(a) 反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(a) で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、(c) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価し、(d) (c) における評価に基づいて試験被験体における遺伝的変異の有無を決定するように構成されているシステムが提供される。

20

30

【0262】

ある特定の態様では、1つまたは複数のプロセッサおよびメモリを含む装置であって、メモリが、1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた試験被験体由来の循環している無細胞試料核酸の配列読み取りのカウントを含み、1つまたは複数のプロセッサによって遂行可能な命令が、(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、(b) 正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されている装置も提供される。

40

【0263】

ある特定の態様では、1つまたは複数のプロセッサおよびメモリを含む装置であって、メモリが、1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた試験被験体由来の循環している無細胞試料核酸の配列読み取りのカウントを含み、1つまたは複数のプロセッサによって遂行可能な命令が、(a) 反復配列および/または過大表示もしくは過小表示された配列の影響を最

50

小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(a) で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、(c) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価し、(d) (c) における評価に基づいて試験被験体における遺伝的変異の有無を決定するように構成されている装置も提供される。

10

【0264】

ある特定の態様では、コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、(a) 参照ゲノムのゲノミックセクションにマッピングされた試験被験体由来の循環している無細胞試料核酸の配列読み取りのカウントにアクセスし、(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、(c) 正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されている命令を含むコンピュータプログラム製品も提供される。

20

【0265】

ある特定の態様では、コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、(a) 参照ゲノムの部分にマッピングされた、試験試料由来の循環している無細胞核酸の読み取りである配列読み取りのカウントにアクセスし、(b) 反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(b) で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、(d) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価し、(e) (d) における評価に基づいて試験被験体における遺伝的変異の有無を決定するように構成されている命令を含むコンピュータプログラム製品も提供される。

30

【0266】

ある特定の形態では、システム、装置および/またはコンピュータプログラム製品は、(i) 核酸配列読み取りが得られるように構成された配列決定モジュール；(ii) 核酸配列読み取りが参照ゲノムの部分にマッピングされるように構成されたマッピングモジュール；(iii) ゲノミックセクションが重み付けされるように構成された重み付けモジュール、(iv) ゲノミックセクションまたはゲノミックセクションにマッピングされたカウントがフィルタリングされるように構成されたフィルタリングモジュール、(v) 参照ゲノムの部分にマッピングされた核酸配列読み取りのカウントがもたらされるように構成されたカウントモジュール；(vi) 正規化されたカウントがもたらされるように構成された正規化モジュール；(vii) 予測カウントまたは予測カウントの誘導値がもたらされるように構成された予測カウントモジュール；(viii) 高度および/またはプロファイルがグラフ化され表示されるように構成されたプロットモジュール；(ix) アウトカム（例えば、胎児の異数性の有無を決定するアウトカム）が決定されるように構成されたアウトカムモジュール；(x) セグメント染色体異常または胎児の異数性

40

50

またはその両方の有無が示されるように構成されたデータディスプレイ編成モジュール；
(x i) 配列読み取りのマッピング、マッピングされた配列読み取りのカウント、カウン
トの正規化およびアウトカムの生成の1つまたは複数が実施されるように構成された論理
処理モジュール；または (x i i) 前述のモジュールの2つ以上の組合せを含む。

【 0 2 6 7 】

いくつかの実施形態では、配列決定モジュールおよびマッピングモジュールは、配列読
み取りが配列決定モジュールからマッピングモジュールに移行されるように構成されてい
る。マッピングモジュールおよびカウントモジュールは、時には、マッピングされた配列
読み取りがマッピングモジュールからカウントモジュールに移行されるように構成されて
いる。カウントモジュールおよびフィルタリングモジュールは、時には、カウントがカウ
ントモジュールからフィルタリングモジュールに移行されるように構成されている。カウ
ントモジュールおよび重み付けモジュールは、時には、カウントがカウントモジュールか
ら重み付けモジュールに移行されるように構成されている。マッピングモジュールおよび
フィルタリングモジュールは、時には、マッピングされた配列読み取りがマッピングモジ
ュールからフィルタリングモジュールに移行されるように構成されている。マッピングモ
ジュールおよび重み付けモジュールは、時には、マッピングされた配列読み取りがマッピ
ングモジュールから重み付けモジュールに移行されるように構成されている。時には、重
み付けモジュール、フィルタリングモジュールおよびカウントモジュールは、フィルタリ
ングされたおよび/または重み付けされたゲノミックセクションが重み付けモジュールお
よびフィルタリングモジュールからカウントモジュールに移行されるように構成されてい
る。重み付けモジュールおよび正規化モジュールは、時には、重み付けされたゲノミック
セクションが重み付けモジュールから正規化モジュールに移行されるように構成されてい
る。フィルタリングモジュールおよび正規化モジュールは、時には、フィルタリングされ
たゲノミックセクションがフィルタリングモジュールから正規化モジュールに移行され
るよう構成されている。いくつかの実施形態では、正規化モジュールおよび/または予測
カウントモジュールは、正規化されたカウントがアウトカムモジュールまたはプロティ
ングモジュールに移行されるように構成されている。

【 0 2 6 8 】

モジュール

モジュールは、時には、装置、システムまたはソフトウェアの一部であり、情報および
データの移行および/または処理を容易にすることができる。モジュールの非限定的な例
が下に記載されている。

【 0 2 6 9 】

配列決定モジュール

配列決定モジュールによって、または配列決定モジュールを含む装置によって、配列決
定、配列決定読み取りの入手をもたらすことができる。「配列受信モジュール」は、本明
細書で使用される場合、「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、当技術分野で公知の配列決定技術で核酸の配列を決定する任意の装置であつて
よい。ある特定の実施形態では、配列決定モジュールを含む装置によって、当技術分野で
公知の配列決定反応が実施される。配列決定モジュールにより、一般に、配列決定反応か
らのデータ（例えば、配列決定装置から生成されたシグナル）に応じて核酸配列読み取
りもたらされる。いくつかの実施形態では、配列決定読み取りをもたらすために、配列決
定モジュールまたは配列決定モジュールを含む装置が必要である。いくつかの実施形態で
は、配列決定モジュールにより、別の配列決定モジュール、コンピュータ周辺機器、オペ
レーター、サーバー、ハードドライブ、装置から、または適切な供給源から配列読み取
りを受け取る、入手する、それにアクセスするまたはそれを回収することができる。時に
は、配列決定モジュールにより、配列読み取りを操作することができる。例えば、配列決
定モジュールにより、アラインメント、集合、断片化、相補配列変換 (c o m p l e m e n t)、
逆相補配列変換 (r e v e r s e c o m p l e m e n t)、誤り検査、または配
列読み取りの誤り訂正を行うことができる。配列決定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んでよい。いくつかの実施形態では、配列決定読み取りは、配列決定モジュールからの1つまたは複数の命令（例えば、プロセス、ルーチンおよび/またはサブルーチン）を実施および/または実行することができるプロセッサ（例えば、1つまたは複数のプロセッサ）を含む装置によってもたらされる。いくつかの実施形態では、配列決定読み取りは、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置によってもたらされる。いくつかの実施形態では、配列決定モジュールは、1つまたは複数の外部のプロセッサ（例えば、内部または外部のネットワーク、サーバー、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））と一緒に作動する。時には、配列決定モジュールにより、別のモジュール、装置、周辺機器、構成部分または特殊化された構成部分（例えば、シークエンサー）からデータおよび/または情報を集める、集合させる、および/または受け取る。いくつかの実施形態では、配列決定読み取りは、以下の1つまたは複数を含む装置によってもたらされる：1つまたは複数のフローセル、カメラ、光検出器、光電池、液体取扱い構成部分、プリンター、ディスプレイ（例えば、LED、LCDまたはCRT）など。多くの場合、配列決定モジュールにより、配列読み取りを受け取る、集めるおよび/または集合させる。時には、配列決定モジュールにより、装置のオペレーターからインプットデータおよび/または情報を受け入れ、集める。例えば、時には、装置のオペレーターにより、モジュールに命令、定数、閾値、式または所定の値がもたらされる。時には、配列決定モジュールにより、それが受け取ったデータおよび/または情報を連続した核酸配列に変換することができる。いくつかの実施形態では、配列決定モジュールによってもたらされた核酸配列を印刷または提示する。いくつかの実施形態では、配列読み取りは、配列決定モジュールによってもたらされ、配列決定モジュールから装置、または任意の適切な周辺機器、構成部分または特殊化された構成部分を含む装置に移行される。いくつかの実施形態では、配列決定モジュールから、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置にデータおよび/または情報もたらされる。いくつかの場合には、配列読み取りに関連するデータおよび/または情報を、配列決定モジュールから任意の他の適切なモジュールに移行することができる。いくつかの実施形態では、配列決定モジュールにより、配列読み取りをマッピングモジュールまたはカウントモジュールに移行することができる。

【0270】

マッピングモジュール

マッピングモジュールによって、またはマッピングモジュールを含む装置によって、配列読み取りをマッピングすることができ、マッピングモジュールにより、一般に、読み取りが参照ゲノムまたはそのセグメントにマッピングされる。マッピングモジュールにより、配列決定読み取りを当技術分野で公知の適切な方法によってマッピングすることができる。いくつかの実施形態では、マッピングされた配列読み取りをもたらしするために、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が必要である。マッピング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んでよい。いくつかの実施形態では、マッピングされた配列決定読み取りは、マッピングモジュールからの1つまたは複数の命令（例えば、プロセス、ルーチンおよび/またはサブルーチン）を実施および/または実行することができるプロセッサ（例えば、1つまたは複数のプロセッサ）を含む装置によってもたらされる。いくつかの実施形態では、配列決定読み取りを、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置によってマッピングする。いくつかの実施形態では、マッピングモジュールは、1つまたは複数の外部のプロセッサ（例えば、内部または外部のネットワーク、サーバー、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））と一緒に作動する。装置は、マッピングモジュールおよび配列決定モジュールを含んでよい。いくつかの実施形態では、配列読み取りを、以下の1つまたは複数を含む装置によってマッピングする：1つまたは複数のフローセル、カメラ、液体取扱い構成部分、プリンター、ディスプレイ（例えば、LED、LCDまたはCRT）など。いくつかの実施形態では、マッピングモジュールにより、配列決定モジュールから配列読み取りを受け取ることができる。いくつかの実施形態では、マッ

10

20

30

40

50

ピングされた配列決定読み取りをマッピングモジュールからカウントモジュールまたは正規化モジュールに移行することができる。

【0271】

カウントモジュール

カウントモジュールによって、またはカウントモジュールを含む装置によって、カウントをもたらすことができる。カウントモジュールにより、当技術分野で公知のカウント方法に従ってカウントを決定し、集合させ、かつ/または提示することができる。カウントモジュールにより、一般に、当技術分野で公知のカウント方法体系に従ってカウントを決定するまたは集合させる。いくつかの実施形態では、カウントをもたらすために、カウントモジュールまたはカウントモジュールを含む装置が必要である。カウント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んでよい。いくつかの実施形態では、カウントは、カウントモジュールからの1つまたは複数の命令（例えば、プロセス、ルーチンおよび/またはサブルーチン）を実施および/または実行することができるプロセッサ（例えば、1つまたは複数のプロセッサ）を含む装置によってもたらされる。いくつかの実施形態では、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置によって読み取りをカウントする。いくつかの実施形態では、カウントモジュールは、1つまたは複数の外部のプロセッサ（例えば、内部または外部のネットワーク、サーバー、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））と一緒に作動する。いくつかの実施形態では、以下の1つまたは複数を含む装置によって読み取りをカウントする：配列決定モジュール、マッピングモジュール、1つまたは複数のフローセル、カメラ、液体取扱い構成部分、プリンター、ディスプレイ（例えば、LED、LCDまたはCRT）など。カウントモジュールにより、配列決定モジュールおよび/またはマッピングモジュールからデータおよび/または情報を受け取り、そのデータおよび/または情報を変換し、カウント（例えば、ゲノミックセクションにマッピングされたカウント）をもたらすことができる。カウントモジュールにより、マッピングされた配列読み取りをマッピングモジュールから受け取ることができる。カウントモジュールにより、正規化されたマッピングされた配列読み取りをマッピングモジュールから、または正規化モジュールから受け取ることができる。カウントモジュールにより、カウント（例えば、カウント、集合カウントおよび/またはカウントの表示）に関連するデータおよび/または情報を任意の他の適切な装置、周辺機器、またはモジュールに移行することができる。時には、カウントに関連するデータおよび/または情報を、カウントモジュールから正規化モジュール、プロットングモジュール、カテゴリー化モジュールおよび/またはアウトカムモジュールに移行する。

【0272】

正規化モジュール

正規化モジュールによって（例えば、正規化モジュールを含む装置によって）、正規化されたデータ（例えば、正規化されたカウント）をもたらすことができる。いくつかの実施形態では、配列決定読み取りから得た正規化されたデータ（例えば、正規化されたカウント）をもたらすために、正規化モジュールが必要である。正規化モジュールにより、データ（例えば、カウント、フィルタリングされたカウント、生のカウント）を当技術分野で公知の1つまたは複数の正規化手順によって正規化することができる。正規化モジュールにより、予測カウントの変動性の推定値（例えば、予測カウントのMADおよび/または予測カウント表示のMAD）をもたらすことができる。いくつかの実施形態では、正規化モジュールにより、多数の実験（例えば、時には、異なる実験、時には、1つまたは複数の共通の実験条件に曝露させた実験）から得た予測カウントから多数の中央値を導出し、多数の中央値の絶対誤差（例えば、偏差、変動性、標準偏差、標準誤差）を導出し、算出された絶対誤差の平均値、アベレージ、または中央値を決定することによって予測カウントのMADをもたらすことができる。いくつかの実施形態では、正規化モジュールにより、多数の実験（例えば、時には、異なる実験、時には、1つまたは複数の共通の実験条件に曝露させた実験）から得た予測カウント表示から多数の中央値を導出し、次いで、多

10

20

30

40

50

数の中央値の絶対誤差（例えば、偏差、変動性、標準偏差、標準誤差）を導出することによって予測カウント表示のMADをもたらすことができる。正規化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んでよい。いくつかの実施形態では、正規化されたデータは、正規化モジュールからの1つまたは複数の命令（例えば、プロセス、ルーチンおよび/またはサブルーチン）を実施および/または実行することができるプロセッサ（例えば、1つまたは複数のプロセッサ）を含む装置によってもたらされる。いくつかの実施形態では、正規化されたデータは、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置によってもたらされる。いくつかの実施形態では、正規化モジュールは、1つまたは複数の外部のプロセッサ（例えば、内部または外部のネットワーク、サーバー、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））と一緒に作動する。いくつかの実施形態では、正規化されたデータは、以下の1つまたは複数を含む装置によってもたらされる：1つまたは複数のフローセル、カメラ、液体取扱い構成部分、プリンター、ディスプレイ（例えば、LED、LCDまたはCRT）など。正規化モジュールは、適切な装置またはモジュールからデータおよび/または情報を受け取ることができる。時には、正規化モジュールにより、配列決定モジュール、正規化モジュール、マッピングモジュールまたはカウントモジュールからデータおよび/または情報を受け取ることができる。いくつかの実施形態では、正規化モジュールにより、配列決定モジュールから配列決定読み取りを受け取ることができ、マッピングモジュールからマッピングされた配列決定読み取りを受け取ることができ、かつ/またはカウントモジュールからカウントを受け取ることができる。多くの場合、正規化モジュールにより、別の装置またはモジュールからデータおよび/または情報を受け取り、そのデータおよび/または情報を変換し、正規化されたデータおよび/または情報（例えば、正規化されたカウント、正規化された値、正規化された参照値（NRV）など）をもたらすことができる。ある特定の¹⁰実施形態では、正規化されたデータおよび/または情報を正規化モジュールから比較モジュール、正規化モジュール、範囲設定モジュール、補正モジュール、カテゴリー化モジュール、および/またはアウトカムモジュールに移行することができる。時には、正規化されたカウント（例えば、正規化されたマッピングされたカウント）を正規化モジュールから予測表示モジュールおよび/または実験に基づく表示モジュールに移行する。

【0273】

予測カウントモジュール

予測カウントモジュールによって（例えば、予測カウントモジュールを含む装置によって）、予測カウントまたは予測カウントの誘導値（例えば、パーセント表示）をもたらすことができる。いくつかの実施形態では、配列決定読み取りから得た予測カウントまたは予測カウントの誘導値（例えば、マッピングされた配列読み取りのカウント、マッピングされた配列読み取りの所定のサブセット）をもたらすために、予測カウントモジュールが必要である。予測カウントモジュールにより、1つまたは複数の選択されたゲノミックセクションについてカウントを合計することができる。時には、予測カウントモジュールにより、配列読み取りおよび/またはカウントに1つまたは複数の数学的操作または統計学的操作を適用する。予測カウントモジュールにより、予測カウントの誘導値を、パーセント表示（例えば、カウント表示）を決定することによって決定することができる。予測カウント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んでよい。いくつかの実施形態では、予測カウントまたは予測カウントの誘導値は、予測カウントモジュールからの1つまたは複数の命令（例えば、プロセス、ルーチンおよび/またはサブルーチン）を実施および/または実行することができるプロセッサ（例えば、1つまたは複数のプロセッサ）を含む装置によってもたらされる。いくつかの実施形態では、予測カウントまたは予測カウントの誘導値は、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置によってもたらされる。いくつかの実施形態では、予測カウントモジュールは、1つまたは複数の外部のプロセッサ（例えば、内部または外部のネットワーク、サーバー、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）

）と一緒に作動する。いくつかの実施形態では、予測カウントまたは予測カウントの誘導値は、以下の1つまたは複数を含む装置によってもたらされる：1つまたは複数のフローセル、カメラ、液体取扱い構成部分、プリンター、ディスプレイ（例えば、LED、LCDまたはCRT）など。予測カウントモジュールは、適切な装置またはモジュールからデータおよび/または情報を受け取ることができる。時には、予測カウントモジュールにより、配列決定モジュール、予測カウントモジュール、マッピングモジュール、正規化モジュールまたはカウントモジュールからデータおよび/または情報を受け取ることができる。いくつかの実施形態では、予測カウントモジュールにより、配列決定モジュールから配列決定読み取りを受け取ることができ、マッピングモジュールからマッピングされた配列決定読み取りを受け取ることができ、かつ/またはカウントモジュールからカウントを受け取ることができる。多くの場合、予測カウントモジュールにより、別の装置またはモジュールからデータおよび/または情報を受け取り、そのデータおよび/または情報を変換し、予測カウントまたは予測カウントの誘導値をもたらし。ある特定の実施形態では、予測カウントまたは予測カウントの誘導値を予測カウントモジュールから比較モジュール、予測カウントモジュール、正規化モジュール、範囲設定モジュール、補正モジュール、カテゴリ化モジュール、および/またはアウトカムモジュールに移行することができる。

【0274】

アウトカムモジュール

アウトカムモジュールによって、またはアウトカムモジュールを含む装置によって、遺伝的変異の有無（異数性、胎児の異数性、コピー数多型）を同定することができる。時には、遺伝的変異をアウトカムモジュールによって同定する。多くの場合、異数性の有無の決定をアウトカムモジュールによって同定する。いくつかの実施形態では、遺伝的変異（異数性、コピー数多型）を決定するアウトカムをアウトカムモジュールによって、またはアウトカムモジュールを含む装置によって同定することができる。アウトカムモジュールは、特定の遺伝的変異（例えば、トリソミー、21トリソミー、18トリソミー）を決定するために特殊化することができる。例えば、21トリソミーを同定するアウトカムモジュールは、18トリソミーを同定するアウトカムモジュールとは異なり、かつ/または別個であってよい。いくつかの実施形態では、遺伝的変異または遺伝的変異（例えば異数性、コピー数多型）を決定するアウトカムを同定するために、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が必要である。アウトカム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んでよい。いくつかの実施形態では、遺伝的変異または遺伝的変異を決定するアウトカムは、アウトカムモジュールからの1つまたは複数の命令（例えば、プロセス、ルーチンおよび/またはサブルーチン）を実施および/または実行することができるプロセッサ（例えば、1つまたは複数のプロセッサ）を含む装置によってもたらされる。いくつかの実施形態では、遺伝的変異または遺伝的変異を決定するアウトカムは、協調し、並行して動作するプロセッサなどの多数のプロセッサを含んでよい装置によって同定される。いくつかの実施形態では、アウトカムモジュールは、1つまたは複数の外部のプロセッサ（例えば、内部または外部のネットワーク、サーバー、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））と一緒に作動する。時には、アウトカムモジュールを含む装置によって、別のモジュールまたは装置からデータおよび/または情報を集める、集合させる、および/または受け取る。時には、アウトカムモジュールを含む装置によって、別のモジュールまたは装置にデータおよび/または情報をもたらし、かつ/または移行する。時には、アウトカムモジュールにより、構成部分または周辺機器から、またはそれにデータおよび/または情報を移行する、受け取るまたは集める。多くの場合、アウトカムモジュールにより、カウント、高度、プロファイル、正規化されたデータおよび/または情報、参照高度、予測高度、予測範囲、不確実性値、補正、補正された高度、プロット、カテゴリ化された高度、比較および/または定数を受け取る、集めるおよび/または集合させる。時には、アウトカムモジュールにより、装置のオペレーターからインプットデータおよび/または情報を受け入れ、集める。例えば、時には、装置のオペレーターにより、定数、閾値、式または所定の値がアウトカ

10

20

30

40

50

ムモジュールにもたらされる。いくつかの実施形態では、協調し、並行して動作するプロセッサなどの多数のプロセッサを含む装置によって、データおよび/または情報がもたらされる。いくつかの実施形態では、適切な周辺機器または構成部分を含む装置によって、遺伝的変異または遺伝的変異を決定するアウトカムの同定が提供される。アウトカムモジュールを含む装置によって、正規化モジュール、予測カウントモジュールから正規化されたデータを、範囲設定モジュールから予測高度および/または範囲を、比較モジュールから比較データを、カテゴリー化モジュールからカテゴリー化された高度を、プロットングモジュールからプロットを、および/または補正モジュールから補正データを受け取ることができる。アウトカムモジュールにより、データおよび/または情報を受け取り、そのデータおよび/または情報を変換し、アウトカムをもたらしすることができる。アウトカムモジュールにより、遺伝的変異または遺伝的変異を決定するアウトカムに関連するデータおよび/または情報を適切な装置および/またはモジュールにもたらしまたは移行することができる。本明細書に記載の方法によって同定される遺伝的変異または遺伝的変異を決定するアウトカムは、さらに検査することによって(例えば、母体の核酸および/または胎児核酸の標的化配列決定によって)、それぞれ独立に検証することができる。

10

【0275】

1つまたは複数のアウトカムを生成した後、多くの場合、アウトカムを使用して、遺伝的変異の有無および/または関連する医学的状態の決定をもたらし。アウトカムは、一般には、医療専門家(例えば、検査技師または管理者; 医師または補助者)に提供される。多くの場合、アウトカムはアウトカムモジュールによってもたらされる。時には、アウトカムはプロットングモジュールによってもたらされる。時には、アウトカムは装置の周辺機器または構成部分にもたらされる。例えば、時には、アウトカムはプリンターまたはディスプレイによってもたらされる。いくつかの実施形態では、遺伝的変異の有無を決定するアウトカムは、報告書の形態で医療専門家に提供され、ある特定の実施形態では、報告書は、アウトカムの値および関連する信頼度パラメータの表示を含む。一般に、アウトカムは、遺伝的変異の有無および/または医学的状態の決定を容易にする適切な形式で提示することができる。データセットを報告および/もしくは提示する、またはアウトカムを報告するために使用するために適した形式の非限定的な例としては、デジタルデータ、グラフ、2Dグラフ、3Dグラフ、および4Dグラフ、写真、ピクトグラフ、チャート、棒グラフ、円グラフ、図、フローチャート、散布プロット、マップ、ヒストグラム、密度チャート、関数グラフ、回路図、ブロック図、バブルマップ、信号空間ダイヤグラム、コンターダイヤグラム、統計地図、スパイダーチャート、ベン図、計算図表など、および前述のものの組合せが挙げられる。アウトカム表示の種々の例が図に示されており、実施例に記載されている。

20

30

【0276】

ある特定の実施形態では、アウトカムを生成することは、核酸配列読み取りデータなどを被験体の細胞内核酸の表示に変換することとみなすことができる。例えば、被験体由来の核酸の配列読み取りを分析し、染色体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、比較的小さな配列読み取り断片を比較的大きな染色体構造の表示に変換することとみなすことができる。いくつかの実施形態では、アウトカムは、被験体(例えば、妊婦)からの配列読み取りを、被験体に存在する実在構造(例えば、ゲノム、染色体またはそのセグメント)の表示(例えば、母体の核酸および/または胎児核酸)に変換することによって生じる。いくつかの実施形態では、アウトカムは、第1の被験体(例えば、妊婦)からの配列読み取りの構造(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、および第1の被験体(例えば、妊婦)および/または第2の被験体(例えば、胎児)に存在する構造の表示をもたらし複合表示の第2の変換を含む。

40

【実施例】

【0277】

下記の実施例は、ある特定の実施形態を例示し、当該技術を限定しない。

【0278】

50

(実施例1)

盲検試料を使用した遺伝的変異の有無の決定

ダウン症候群についての有効な出生前スクリーニング検査は、多くの場合、母体の年齢と、第1三半期における頂部浮腫の超音波測定および/または第1三半期および第2三半期に得られるいくつかの母体の血清スクリーニングマーカーの測定からの情報を組み合わせる。これらの出生前スクリーニング検査では、多くの場合、実質的に全ての症例の約90%に至るまでが約2%の偽陽性率で検出される。ダウン症候群の分布率を考慮すると、侵襲的診断検査(例えば、羊水穿刺または絨毛採取)を勧められたスクリーニング陽性の女性16人に1人で妊娠に影響し、15人では影響しない。200例中1例ものそのような侵襲的手順に胎児喪失が伴い、これは出生前診断の著しい有害な結果である。胎児喪失という著しい有害な結果により、時には、偽陽性率が最小限になるように補正されたスクリーニングカットオフが導かれる。実際には、約5%の偽陽性率が一般的である。

10

【0279】

母体の血液中の無細胞DNAの約3~6%が胎児起源のものであるという発見により、ダウン症候群を非侵襲的に検出することができるかどうかを決定するための試験が促された。数百万のDNA断片の最初の36塩基について配列決定して、それらの特異的な染色体起源を決定する技法である大規模並列処理ショットガン配列決定(MPSS)を用いて胎児のダウン症候群を同定した。胎児が第3の第21染色体を有する場合、第21染色体断片の百分率は予測よりもわずかに高くなる。その後の報告により、これらの知見が拡張され、約2%以下の偽陽性率で少なくとも約98%の検出率を実現することができることが示唆されている。これらの試験は有望であるが、以下の因子によって限定された;試験が比較的小さな患者群(13~86のダウン症候群症例および34~410の正倍数性対照試料の範囲)を利用して実施されたこと;DNA配列決定がCLIAによって認定された研究所で実施されなかったこと;ならびに処理量およびターンアラウンドタイムにより臨床上的実施がシミュレートされなかったこと。

20

【0280】

いくつかの実施形態では、本明細書に記載の方法、プロセスおよび装置を利用して、盲検試料を使用し、試験被験体データを正規化するための参照ゲノムデータセットを必要とせず、遺伝的変異(例えば、トリソミー、ダウン症候群)の有無を決定するアウトカムをもたらすことができる。

30

【0281】

材料および方法

全体的な試験デザイン

本明細書で示されている試験(ワールドワイドウェブURL clinicaltrials.gov NCT00877292参照)には、世界中で27か所の出生前診断センター(例えば、以後登録場所と称される)で登録された患者が関与した。母体の年齢、家族歴または同意を得た陽性血清および/もしくは超音波スクリーニング検査、血漿試料、人口統計および妊娠に関連する情報に基づいてダウン症候群のリスクが高い女性。各登録場所で治験審査委員会の認可(または同等のもの)を得た。患者および試料は試験コードによって識別した。侵襲的検査の直前に試料を抜き取り、6時間以内に処理し、-80で保管し、コーディネートセンターにドライアイス輸送した。このコホート内で、ダウン症候群についての盲検DNA検査を用いてコホート内症例対照試験を展開した。妊娠期間(最も近い週;同じ三半期)、登録場所、人種(自己宣言)、および冷凍装置に入っていた時間(1ヶ月以内)に基づいて、各症例に対して7つの正倍数性試料をマッチさせた。偽陰性の結果がないと仮定して、200のダウン症候群妊娠(症例)が80%の検出力を有し、信頼区間(CI)が低いために98%が棄却された。症例は第1三半期および第2三半期の間に同等に分布した。この試験について、ダウン症候群を47、XY、+21または47、XX、+21と定義し、ダウン症候群のモザイクおよび双生児妊娠は排除した。試験のコーディネートおよび試料の保管は独立した大学医療センター(例えば、Women & Infants Hospital)に基づいた。凍結し、コード付けした試料

40

50

(4 mL)を検査するためにSequenom Center for Molecular Medicine (SCMM, San Diego, CA)に送付した。SCMMは、ターンアラウンドタイムを定量することを含めた、核型およびシミュレートされた臨床検査に関する知見を有さなかった。試料のサブセットを、DNA配列決定の経験がある独立した学術的研究所であるロサンゼルスのカリフォルニア大学にあるOrphan Disease Testing Center (UCLA; Los Angeles, CA)において検査するために送付した。どちらの研究もCLIAによって認定されており、どちらもSCMMによって最初に開発された標準化された書面のプロトコルを使用して臨床的解釈をもたらした。

【0282】

試験の完全性

この産業界から資金提供された試験の完全性、信頼度、および独立性を確実にすることを最優先した。3人の監督委員会(承認を参照されたい)を創設し、試験のデザイン、実施、分析、および解釈に対する評価および推奨の提供を委ねた。試験プロトコルは、登録場所の検査、登録場所を試験のスポンサーから切り離すこと、独立した学術的研究所による確認検査、診断検査結果を多数のレベルで盲検化すること、アウトカムデータにリモートコンピュータアクセスしないこと、学術的な検査場所で生のデータ全てにアクセスすること、配列決定および解釈の結果をコーディネートセンターに即時にファイル転送すること、およびファイルチェックサムを使用してその後の変化を同定することを含んだ。SCMMにより、同様の設備、訓練、解釈用ソフトウェア、および標準の操作プロトコルを備えた独立した研究所が提供された。

【0283】

研究所で開発された検査(Laboratory-developed test)

上記の通り、MPPSSを利用して無細胞DNAについて配列決定した。簡単に述べると、循環している無細胞DNA断片を母体の血漿から単離し、胎児の寄与(胎児分率)を決定するアッセイを用いて定量する。残りの単離物を使用して、配列決定ライブラリーを生成し、正規化し、多重化して、4つの試料について単一のフローセルレーン(例えば、フローセル1個当たり8レーン)で実行することを可能にした。マイクロフルイディクスプラットフォーム(Caliper Life Sciences, Hopkinton, MA)を使用してDNAライブラリーを定量し、cBotプラットフォーム(Illumina, Inc, San Diego, CA)を使用してクラスターを生成した。Illumina HiSeq 2000プラットフォームでフローセルについて配列決定し、生じたデータを、Illuminaソフトウェアを使用して解析した。コンピュータ解釈により、中央の推定値(zスコア)を上回るまたは下回る標準偏差(例えば、SD)の口バスタな推定値がもたらされた; zスコア3以上をダウン症候群と一致するとみなした。主要なCLIA研究所(SCMM)の指揮者が結果を精査し、第2の一定分量を検査するための要求を開始し、検査された妊娠の全てについて最終的な「サインアウト」された解釈をもたらした。独立したCLIA研究所(UCLA)の指揮者も同じことをしたが、第2の試料の一定分量を要求することはできなかった。各研究所は自身の結果にのみアクセスした。

【0284】

統計解析

暫定的な解析により、16症例のうち3症例超または112例の対照のうち6例超が誤って分類されたことが示された場合には試験を中断する。試験はマッチさせたが、解析はマッチさせないように計画した。SAS(商標)Analytics Pro(Cary, NC;以前は統計解析システムとして公知であった)およびTrue Epistat(Richardson, TX)を使用して、 X^2 検定、t検定、分散分析(ANOVA)、および線形回帰(適切な変換後)を使用して群および関連性の間の差異を調査した。二項分布を用いて割合の信頼区間(CI)を計算した。P値は両側性であり、有意性は0.05レベルであった。

10

20

30

40

50

【 0 2 8 5 】

結果

試料母集団

2009年4月からおよび2011年2月の間に、登録場所27か所（以下の表1参照）で適格の妊娠中の女性を同定し、インフォームドコンセントを得、試料を採取した。登録者4664人の中で、218人の単胎児ダウン症候群および3930人の単胎児正倍数性妊娠が起こった。図1には、胎児のアウトカム、血漿試料の状態、および279人の女性（6%）が排除された理由に関する詳細が提供される。以前の刊行物または試験に含まれた試料はなかった。合計4385人の女性（94%）が単胎児妊娠、少なくとも2つの適切な血漿試料および診断検査結果を有した。これらのうち、97%が、両端を含めて11週から20週の間妊娠期間であり、34%が第1三半期であった。登録された女性51人を除く全てについて胎児の核型（または同等のもの）が入手可能であった。116人の女性については、血漿試料が検査に適するとみなされなかった（例えば、運搬中に解凍された、凍結されるまでに6時間を超えた、一定分量が1つのみである、および体積が不十分である）。さらに112人の女性を、多胎妊娠または現存する胎児死亡が原因で排除した。4385例の生育可能な単胎児妊娠の中で、34%が第1三半期後期に得られ、66%が第2三半期初期に得られた。合計212症例のダウン症候群を検査のために選択した。それぞれの症例について、マッチする正倍数性妊娠を7例選択した（例えば、1484；正倍数性症例とダウン症候群症例の比率7：1）。237例の他のアウトカムは追加的な常染色体異数性、性染色体異数性、モザイクおよび他の染色体異常であった。1つの対照が後で18トリソミーであることが発見されたが、「正倍数性」対照として含まれていた。

10

20

【 0 2 8 6 】

【 表 1 - 7 】

表1 試験において登録された臨床場所および関連する登録およびアウトカム情報

登録場所	場所	臨床研究者	単胎児妊娠	正常な核型	その他	登録された患者
North York General Hospital	カナダ、トロント	Wendy S. Meschino, MD	41	651	86	778
Istituto G. Gaslini	イタリア、ジェノバ	Pierangela De Blasio, MD	27	492	35	564
Hospital Clinic Barcelona	スペイン、バルセロナ	Antoni Borrell, MD, PhD	24	291	44	369
Centrum Lekarske Genetky	チェコ共和国、チェスケ・ブチヨヴィツェ	David Cufka, MD	14	362	19	395
Hospital Italiano	アルゼンチン、ブエノスアイレス	Lucas Otano, MD, PhD	13	68	14	95
ダルハウジー大学	カナダ、ハリファックス	Michiel Van den Hof, MD	12	115	18	145
Rotunda Hospital	アイルランド、ダブリン	Fergal Malone, MD	12	70	12	94
セルメンヴェイス大学	ハンガリー、ブダペスト	Csaba Papp, MD, PhD	10	64	9	83
IMALAB s.r.o. Medical Laboratories	チェコ共和国、ズリーン	Jaroslav Loucky, RNDr	9	238	8	255
CEMIC	アルゼンチン、ブエノスアイレス	Maria Laura Igarzabal,	8	224	49	281
アイオワ大学	アイオワ州、アイオワシティ	MD Kristi Borowski, MD	8	135	30	173
Women & Infants Hospital	ロードアイランド州、プロビデンス	Barbara O'Brien, MD	6	99	21	126
ペーチ大学	ハンガリー、ペーチ	Bela Veszpremi, MD,	4	172	31	207
アラバマ大学バーミングハム	アラバマ州、バーミングハム	PhD Joseph Biggio, MD	4	169	20	193
Rambam Medical Center	イスラエル、ハイファ	Zeev Weiner, MD	4	133	10	147
Cedars Sinai PDC	カリフォルニア州、ロサンゼルス	John Williams, MD	3	192	28	223
ノースウェスタン大学	イリノイ州、シカゴ	Jeffrey Dungan, MD	3	88	11	102
Henry Ford Hospital	ミシガン州、デトロイト	Jacquelyn Roberson, MD	3	74	14	91
バージニア大学	バージニア州、シャーロッツビル	Devereux N. Saller, Jr,	3	21	8	32
フリティッシュュコロンビア大学	カナダ、バンクーバー	MD Sylvie Langlois, MD	2	67	14	83
Intermountain Healthcare	ユタ州、ソルトレイクシティ	Nancy Rose, MD	2	67	9	78
Brigham and Women's Hospital	マサチューセッツ州、ボストン	Louise Wilkins-Haug, MD	2	21	8	31
ペイラー医科大学	テキサス州、ヒューストン	Anthony Johnson, DO	2	20	0	22
イエール大学	コネティカット州、ニューヘイブン	Maurice J. Mahoney, MD,	1	31	9	41
New Beginnings Perinatal Consultants	ロードアイランド州、プロビデンス	JD	1	7	4	12
カルガリー大学	カナダ、カルガリー	Marshall Carpenter, MD	0	52	5	57
Royal North Shore Hospital	オーストラリア、シドニー	Jo-Ann Johnson, MD	0	7	0	7
総数		Vítomir Tasevski, PhD	218	3,930	516	4,664

以下の表 2 では、人口統計および妊娠に関連する情報が症例と対照の間で比較されている。マッチングは上首尾であった。年齢中央値は両群で約 3 7 歳であり、全員が 1 8 歳以上であった。診断検査の徴候は異なり、症例では超音波異常または多数の徴候を有する可能性がより高かった。試料を採取し、平均して 1 時間以内、全て 6 時間以内に処理し、凍結した。アウトカムは、第 1 三半期の 2 症例（重症の超音波異常を有する生育可能な胎児の中絶後の受胎の産物の、一方は定量的ポリマーゼ連鎖反応、他方は蛍光 *in situ* ハイブリダイゼーション）以外は核型分析に基づいた。

【 0 2 8 7 】

【表 2】

表2 検査した選択されたダウン症候群およびマッチさせた正倍数性試料についての人口統計および妊娠に関連する情報

特性	ダウン症候群	正倍数性	P	
試料の数	212	1,484		
母体の年齢(歳)(アベレージ、SD)	37.0、5.0	36.6、5.1	0.36	
母体の年齢35歳以上(N、%)	160 (75%)	1,036 (70%)	0.12	
妊娠期間(アベレージ、範囲)	15.3 (9.2-21.3)	15.0 (8.1-21.5) 50%/50%	0.21	
第1三半期/第2三半期の妊娠期間(%)	50%/50%	152 (33)	1.0	
ポンド単位の母体の体重(アベレージ、SD)	149 (30)	15%	0.33	
出血(%)	17%		0.44	10
母体の人種(N、%)		1,316 (89%)	1.0	
白人	188 (89%)	35 (2%)		
黒色	5 (2%)	105 (7%)		
アジア人	15 (7%)	28 (2%)		
不明	4 (2%)	303 (20%)		
ラテンアメリカ系白人(N、%)	39 (18%)	42 (3%)	0.92	
アッシュケナジユダヤ人(N、%)	3 (1%)		0.13	
登録の主要な指標(N、%)		327 (22%)	<0.001	
第1三半期検査によるスクリーニング陽性	48 (23%)	118 (8%)		
第2三半期検査によるスクリーニング陽性	11 (5%)	192 (13%)		
統合検査によるスクリーニング陽性	38 (18%)	130 (9%)		20
同定された超音波異常	51 (24%)	543 (37%)		
母体の高齢	24 (12%)	112 (8%)		
2つ以上の徴候	39 (18%)	44 (3%)		
異数性の家族歴	0 (0%)	18 (1%)		
その他または示されていない	1 (<1%)			
診断手順(N、%)		787 (53%)	0.79	
羊水穿刺	114 (54%)	697 (47%)		
絨毛採取	97 (46%)	0 (0%)		
受胎産物の検査	1 (<1%)			
診断検査(N、%)		805 (54%)	<0.001	30
核型単独	95 (46%)	679 (45%)		
核型およびその他	115 (53%)	0 (0%)		
QF-PCRまたはFISH単独	2 (<1%)	45 (3%)		
中程度～全体の溶血(N、%)	8 (4%)	1.2 (0.1-6)	0.60	
時間単位の試料処理(平均値、範囲)	1.1 (0.1-6)		0.63	

循環している遊離のDNAに対する胎児の寄与

M P S Sの前に、抽出されたDNAを検査して、母体の血漿中の胎児起源の遊離のDNAの割合(胎児分率)を決定した。ほぼ全て(1687/1696; 99.5%)が、許容できる限界(4~50%)の範囲内の最終的な胎児分率を有し、幾何平均は13.4%であった。偽陰性の結果を最小限にするために、より低いカットオフを選択した。上のカットオフを、研究所の指揮者にこれが稀な事象を表すことを警告するために選択した。9つが許容されないレベルを有した; 6つが閾値を下回り、3つが閾値を上回った。ダウン症候群を同定することにおけるM P S Sの成功は、胎児分率に高度に左右されるので、16の潜在的な共変量(図4~19、実施例2参照)を探究した(処理時間、溶血、地理的地域、診断検査の徴候、登録場所、妊娠期間、母体の年齢、母体の体重、膈からの出血、母体の人種、白人民族性、胎児の性別、冷凍装置での保管時間、ならびにDNAライブラリー濃度、マッチする配列の数、および胎児のアウトカムに対する胎児分率の影響)。胎児分率と母体の体重の強力な負の関連が症例女性および対照女性において観察され(図11、実施例2参照)、100ポンド、150ポンド、および250ポンドの体重には、それぞれ17.8%、13.2%、および7.3%の予測胎児分率が伴った。妊娠期間、母体の人種、または検査の徴候については関連が見いだされなかった。他の関連は小さく、

大抵有意でなかった。

【0288】

ダウン症候群についての大規模並列処理ショットガン配列決定検査

アッセイプロトコール、および関連する器械使用に関して訓練を受けた30人の科学者、分子技師/技術者により、9週間(2011年の1月から3月まで)にわたって検査が実施された。解釈のために歴史的な基準範囲を使用し、9つが新しいデータのリアルタイムでの精査必要性を伴った。研究所の指揮者による最初のいくつかのフローセルについての精査(サインアウト前)により、参照データに対する補正が必要であったことが明らかになった(実施例2および図20~22参照)。6つのフローセルからのデータを生成した後、結果は監督委員会によって暫定基準に従って評価され、検査を継続することを可能にするために機密決定が行われた。検査の終わりに、しかし非盲検化する前に、SCMMにより、1696の登録者の中で90の検査失敗のうち85について第2の一定分量が要求された(5.3%; 95%CI、4.3~6.5; 実施例2参照)。第2の結果を最終的な解釈のために使用した。

10

【0289】

ダウン症候群の試料では、胎児分率との明白かつ有意な正の関係が示された; 試料の208がカットオフを上回り、4つが下回った。4つのダウン症候群の試料のzスコアはカットオフ3を下回った; 全ての胎児分率が7%未満であった(例えば、7%、7%、5%、および4%)。胎児分率とzスコア間の強力な正の関連が症例については存在したが(対数変換後、傾き=0.676、 $p < 0.001$)、対照については存在しなかった(傾き=0.0022、 $P = 0.50$)。胎児分率が低いダウン症候群の試料のうちの1つの最初のzスコアは5.9であり、1つの境界線上の品質欠陥を伴った; 反復試料のzスコアは2.9であった(最初の陽性の結果と一致する境界線上の値)。反復試料からの情報と最初の試料についての5.9スコア(例えば、境界線上の欠陥)を組み合わせることにより、研究所の指揮者が正しい呼び出しを行うことが可能になる。他の臨床的解釈は全てコンピュータ解釈と合致した。したがって、サインアウトされた結果により、ダウン症候群の胎児212人のうち209人が正確に同定された(検出率98.6%; 95%CI、95.9~99.7)。

20

【0290】

試験に使用した全てのダウン症候群試料および正倍数性試料の臨床的解釈は以下の通りである: 正倍数性妊娠の中では、1471例が陰性であり、3例が陽性であり、13例が第2の一定分量でも失敗した。ダウン症候群妊娠の中では、209例が陽性であり、3例が陰性であった。1471の正倍数性試料の中で、3つのzスコアが胎児分率の範囲にわたって3より大きく、ダウン症候群に不正確に分類され、0.2%の偽陽性率がもたらされた(95%CI、0.1未満~0.6)。女性13人については(13/1696または0.8%; 95%CI、0.4~1.3)、検査結果が利用可能であり、大抵「正常」であったが、最初の試料および反復試料に対する品質管理不足に起因して解釈がもたらされなかった(6人については胎児分率が4%未満であり、1人については胎児分率が50%超であった)(図2B参照)。誤って分類された妊娠についての研究所の結果、試料の取扱い、および妊娠のアウトカムを、潜在的な誤差について広範囲にわたって確認し、同定されなかった(表3、実施例2参照)。zスコアに対する第1の15共変量の解析を実施した(図7~10、実施例2参照)。症例の間で母体の体重について強力な負の関連が存在し、この関連は対照では弱かった。症例において、妊娠期間について小さいが有意な正の関連が存在し(図7、実施例2参照)、11週および19週の妊娠期間でそれぞれ7.2および9.9にzスコアが回帰した。他の関連は小さく大抵有意ではなかった。

30

40

【0291】

独立した研究所による検査性能の確認

独立した大学研究所(例えば、UCLA)において、最初にSCMMによって処理および検査された605の最初の試料一定分量のサブセットについてのクラスター生成、DNA配列決定、および解釈が実施された。このサブセットは、92の患者試料(例えば、プ

50

レート)の完全な群全てからコーディネートセンターによって無作為に選択された。合計578の試料がどちらの場所においても首尾よく検査された(96%)。コンピュータにより解釈されたMPSSの結果は、SCMM値を用いてzスコアとしてとして表されている。合計77例のダウン症候群妊娠および501例の正倍数性妊娠がどちらの場所においても首尾よく検査された。一方の場所または両方の場所において最初の試験で失敗した27の試料は含まれていない。zスコアカットオフ3を使用した。これらの試料の中で、不一致が1つだけ起こった。正倍数性試料がUCLAによっては誤って分類されたが(zスコア=3.46)、SCMMによっては正確に分類された(zスコア=2.02)。どちらの群でも1つのダウン症候群の試料が誤って分類された。77例のダウン症候群妊娠および501例の正倍数性妊娠のどちらの間でも相関は高かった(例えば、それぞれR=0.80および0.83)。この578のサブセットにおいて、SCMMについての検出率、偽陽性率、および最初の失敗率は、それぞれ98.7%、0.0%、および4.4%であった。UCLAについての対応する率は、98.7%、0.2%、および3.9%であった(表3、実施例2参照)。登録者56人の別のサブセットでは、2連の4mLの血漿試料を各研究所で検査した。1つの正倍数性試料が、胎児分率が低いことに起因してどちらの場所においても失敗した。UCLAにおいてさらに2つの正倍数性試料の配列決定が失敗し、そのプロトコールでは再検査が可能でなかった。SCMMおよびUCLAにおける失敗率は、それぞれ1.8%および5.3%であった。残りの53の試料の中で、全ての品質パラメータおよび解釈の結果が2つの場所で合致した(実施例2)。どちらの研究所においても、検出率および偽陽性率はそれぞれ100%および0%であった。

10

20

【0292】

事後解析

大きなサンプルサイズにより、MPSS結果を解釈する代替的方法を調査するための機会がもたらされた。サインアウト後であるが研究所非盲検化の前に、SCMM研究所で、第21染色体のパーセント結果について、MPSSの性能が改善されることが示されているプロセスであるGC含量についての補正を行い、また、RepeatMasker(URLワールドワイドウェブrepeatmasker.org/PreMaskedGenomes.html)に関してフィルタリングし、結果をコーディネートセンターに送って代替の解釈のアルゴリズムをよりよく実施することができるか、よりロバストであるか、またはその両方であるかを決定した。分析により、対照の結果は、フローセルごとまたはプレート(バッチ処理された3つのフローセル)ごとに変動するが(ANOVA、F=13.5、p<0.001)、SDは一定であり(ANOVA、F=1.2、P=0.23)、これにより、GCについて補正された結果をプレート中央値の倍数に変換することが可能になることが示された。ダウン症候群妊娠および正倍数性妊娠におけるプレート中央値の倍数は、1つの持続的な偽陰性の結果以外は完全に分けられた(実施例2参照)。フローセルに特異的なzスコアを補正することによっても、性能が改善され、2つの偽陰性および1つの偽陽性の結果が残った(実施例2参照)。事後解析は、臨床的解釈を行う時点では利用不可能であった。

30

【0293】

臨床的意味

2,116の最初の患者試料(1696が本発明で報告されたものであり、420が他の患者試料である)を、2つのHiSeq2000プラットフォームを使用して、1週間当たり患者235人の処理量で検査した。ターンアラウンドタイム(例えば、試料の解凍からサインアウトまで)が9週間の検査に対して改善され、最終的な20フローセルのうち18について10日の標的に見合った(実施例2参照)。これは、第2の一定分量を必要とした試料の5%を含まないが、失敗は多くの場合、検査プロセスの初期に検出されたので、第2の一定分量を必要とした試料についてのターンアラウンドタイムは倍にはならない。

40

【0294】

有用性を評価するために、単純なモデル(実施例2参照)でダウン症候群についての現

50

行の診断プロトコールを、高リスク妊娠の同定と侵襲的診断の間にM P S Sを挿入する診断プロトコールと比較する。ダウン症候群のリスクが高い女性100,000人を仮定すると、32例の正常な妊娠ごとに1例が妊娠に影響を受け、診断検査費用は患者1人当たり\$1,000であり(実施例2参照)、手順に関連する胎児喪失率は200例に1例である。高リスクの女性による侵襲的検査の完全な取り込みでは、3,000症例が検出され、費用が\$1億、手順に関連する喪失が500例になる。高リスクの女性全てによってM P S S検査が完全に取込まれ、その後M P S S結果が陽性の女性において(検査が失敗した人と一緒に)侵襲的検査を行うことにより、2,958の症例が検出され(42症例が見落とされる)、費用が\$3.9百万、喪失が20例になる。2つのプロトコールの財務費用の差異は、M P S S検査の費用を相殺する助けになり得る。ドル値を480例の潜在的に回避可能な手順に関連する喪失に割り当てることは難しいが、これらは同等に重要な考察である。手順に関連する喪失率は、200人に1人よりも低く、喪失の絶対数は減少するが、減少率は同じままである。

10

【0295】

考察

本明細書において報告されているものを含め、合計350例のダウン症候群および2061例の対照妊娠が報告された。報告されたダウン症候群および対照妊娠の全体で99.0%の感度および特異度が実証され(例えば、95%CI、98.2~99.8%、 $I^2 = 0%$;表5、実施例2参照)、これにより、M P S Sに基づくダウン症候群についての検査の臨床的な有効性の決定的な証拠がもたらされる。陽性の結果では、時には、ダウン症候群のリスクが490倍増加し(例えば、98.6%検出/0.2%偽陽性率)、陰性の結果では、時には、リスクが72分の1低下した(例えば、99.8%/1.4%)。検査は女性1000人に992人で上首尾であった。最初の検査の5.3%で品質確認できなかったが、これらの82%は第2の一定量の検査後に解決された。残りの検査失敗は、多くの場合、胎児分率が少ないことに関連し、これは時には、妊娠の1週間または2週間後に反復試料採取することによって解決することができた。M P S Sの性能は独立した研究所により(例えば、実施例2の表5参照)、元の血漿試料および血漿DNA調製物を使用して確認された。

20

【0296】

本試験では、複数の試料(採取、処理、凍結、および輸送)を27か所の登録場所で取り扱い、予測される臨床上の実施をシミュレートした。我々の発見により、広範な妊娠期間の範囲にわたって、種々の人種/民族群の中で、母体の年齢の全てに対して、および診断検査の徴候の全てに対して、M P S Sの性能が裏付けられる(実施例2参照)。性能は、膣からの出血または試料の溶血の影響を受けず、6時間に至るまでの試料の処理時間に対してロバストである。よく説明されている血液量が増加することの希釈の影響が原因で、15例の検査失敗は体重の重い女性においてより一般的である。解釈における胎児分率の説明は保証することができる。全体的に、スクリーニング結果が偽陽性である大多数の女性は侵襲的検査を回避するが、影響を受けた妊娠のほぼ全ては、従来の侵襲的手段によって確信的に診断される。本試験により、ダウン症候群のリスクが高いと同定された女性に対して、検査の複雑さおよび必要な供給源を考慮に入れて、M P S Sを提案することが支持される。検査は少なくとも週2回起こり、患者の結果の95%についてのターンアラウンドタイムは、羊水細胞および絨毛採取の細胞遺伝学的分析についての現在利用可能なターンアラウンドタイムと同等である。M P S Sの利用可能性により、血清/超音波スクリーニングカットオフの低下も正当化され、それにより、ダウン症候群がより高く検出される。この試験により、初めて、フローセル間の固有の変動性が実証されている。これらの変化を説明することにより、臨床上の実施が改善される。いかに最良にそのような補正を実施するかには、さらなる試験が必要である。

30

40

【0297】

事後解析により、大部分はGC含量についての補正に起因して偽陰性の結果および偽陽性の結果が減少した。これにより、M P S Sの性能が、検査を実務に導入した場合により

50

良いものになることの強力な証拠が構成される。この試験により、M P S Sを研究から臨床的な環境に妥当なターンアラウンドおよび処理量を伴って転換することができることの証拠ももたらされる。ある特定の実行問題は注意に値する。外界温度で無細胞DNAレベルに影響を及ぼすことなく保管および輸送することを可能にする採取管が役立つ。現在、試料は、我々の試験において従ったものと同様のプロトコールで処理し、凍結し、ドライアイス輸送しなければならない。これは観察に基づく試験であったので、臨床的な環境における有効性を示す実証プロジェクトが保証される。患者およびプロバイダーの両方に対する教育材料を開発し、検証して、十分な情報を得た上での決定を確実に行うことに役立てる必要がある。追加的な懸念としては、払い戻しおよび関連性のある専門家ガイドラインの開発が挙げられる。一部では、胎児DNAを検査することにより新しい倫理的問題が生じることが示唆されている。リスクが高い女性のM P S S検査の推奨される設定では、これらの問題の多くは関係ない。

10

【0298】

出生前スクリーニングの分野における主要な目標は、侵襲的手順の必要性を低下させることである。M P S S検査はまだ診断的であるとみなすことはできない。しかし、すでにダウン症候群についてのリスクが高い女性にM P S S検査を提案することにより、高い検出を維持しながら、手順に関連する喪失を96%に至るまで減少させることができる。侵襲的検査による確認はなお必要である。この試験により、以前の報告と一緒に、高い性能が実証されるが、C L I Aによって認定された研究所において、最初の失敗に対して利用可能な第2の一定分量を有し、ターンアラウンドタイムをモニタリングし、オペレーター間および機械間の変動性を評価し、独立した学術臨床研究所における試料結果のサブセットを検証し、遺伝医学者/研究所の指揮者を報告プロセスに統合する検査を実施することによって証拠を拡張した。この報告は双胎妊娠などの他の染色体異常13または事象には対処しない。当該技術が前進するにつれ、そのような改良が利用可能になるであろう。なおいくつかの実行問題に対処する必要があるが、証拠により、この検査を、ダウン症候群のリスクが高い女性に、侵襲的診断検査の前に臨床的に導入することが保証される。

20

【0299】

(実施例2)

盲検試料を使用した遺伝的変異の有無の決定：追加的な材料、方法および結果
試験の完全性

30

試験の独立性および完全性の継続の保証に役立てるために、2009年2月に試験監督委員会を創設した。委員会の構成は、出生前検査および分子遺伝学的方法の臨床的態様および研究的態様の両方の専門知識を有する産科学および遺伝学の学術コミュニティが代表されるように設計した。委員会は、2009年および2010年の間に平均して年に3回、試験Co-Principal Investigators (Co-PI's)と直接または電話によって会合し、2011年2月にその任務を完了し、最後の電話会議を開き、活発な試験登録を終えた。委員会のメンバーは、試験のスポンサー(Sequenom)と秘密保持契約を結ばないことを選択し、したがって、彼らは独占的な方法または結果の知見を有さず、また、試験の過程でSequenomの人員と直接やり取りしなかった。監督委員会のインプットは、1)検査用の試料をコード付けおよび選択する安全な方法、2)試験結果の暫定的確認、および3)試験のスポンサーとコーディネーターの分離、および動員場所の活動を維持するための規則の実行において必須であった。

40

【0300】

試験Co-PIまたはコーディネーターによる各登録場所の検査は、手順に対する順守を精査および評価するための現地訪問、作業空間およびリソースの検査、提出されたデータの検証および試験の目的、方法およびタイムラインに関する質問への回答を伴った。各検査の概要を作成し、特定の試験PIおよび登録場所PIがサインし、患者の識別子またはデータを含有しないコピーを試験のスポンサーに送付した。登録場所は試験のスポンサーとは直接接触せず、また、試料の一部が独立した研究所で検査された。

【0301】

50

同様に、手順を整備して、検出されずに生のデータを変化させることができないこと、および全ての生の結果を独立した研究所で再解析することができることを確実にした。診断検査結果の盲検化を2つのレベルで実現した。コーディネートセンター内では、試料および人口統計情報を Rhode Island に保存し、アウトカムデータは、適切な時間に人口統計データとマージするためにコーディネートセンターの第2の支所に保存した（例えば、Maine）。サーバーはインターネットに接続していなかったため、この情報はいずれも、遠隔地からはアクセス不可能であった。

【0302】

コーディネートセンター

Women & Infants Hospital (WIH) がコーディネートセンターとして機能し、試験の全体的な責任を有した。責任は、試験デザインを実行し、順守すること、登録場所との通信を動員し、確立すること、安全な試験データベースおよびウェブサイトを維持すること、患者データを収集し、検証すること、処理された血漿試料バンクを維持すること、および監督委員会を編成し、利用することを含んだ。センターは2つの場所に位置し、1つはメイン州、スタンディッシュにあり、そこではコンピュータ化されたデータを Co - P I および試験コーディネーターの監督の下で保持し、1つはロードアイランド州、プロビデンスにあり、そこでは、登録場所から試料を入手し、- 80 で保管し、必要に応じて検査研究所に輸送し、また、登録場所に対する行政的支持および供給支持が位置した。試験は WIH により、連邦のガイドラインに従って施行された。WIH と試験のスポンサーの間で守秘義務契約が結ばれ、これにより、Co - P I が試験全体を通して暫定データおよび研究結果にアクセスすることが可能になった。

【0303】

登録場所

多数の患者へのサービス、統合スクリーニング、または第1三半期の診断検査を提供している場所を優先的に探した。27の参加登録場所（表1参照、実施例1）により、第1三半期後期および/または第2三半期初期におけるダウン症候群（または他の常染色体異数性）についての診断検査が提供された。全ての場所が、厳重なプロトコールに従って血漿試料を採取し、処理し、保管し、輸送する能力を有した。場所は治験審査委員会（または同等のもの）の認可を獲得し、試験に登録された各女性のインフォームドコンセントを得た。

【0304】

研究所の場所

Sequenom Center for Molecular Medicine in San Diego (SCMM - SD) は、高複雑度分子遺伝学研究所として CLIA によって認定されている。研究所には2つの Illumina HiSeq 2000 Next Generation Sequencer があり、この試験ではその両方を使用した。カリフォルニア大学 Los Angeles School of Medicine (UCLA) にある Orphan Disease Testing Center も、CLIA によって認定された高複雑度遺伝学研究所であり、この試験の間 Illumina HiSeq 2000 プラットフォームを1つ有した。UCLA は、盲検化された試験試料の大規模並列処理配列決定の実施において SCMM - SD と協力し、SCMM - SD において作成された、Illumina HiSeq 2000 プラットフォームで使用するために更新された標準化された書面のプロトコールに従って臨床的解釈をもたらした。

【0305】

試験母集団

診断検査を予定している妊娠中の女性に関する情報を各登録場所で精査して、試験基準に従って異数性のリスクが高く、胎児が妊娠期間 21 週 6 日以下の女性を同定した。高リスクを、血清検査および/または超音波検査によるダウン症候群または他のトリソミーについてのスクリーニング陽性、分娩時の母体の年齢が 38 歳以上（試験の初期にはこれは

10

20

30

40

50

40歳以上に設定していた)、または異数性の家族歴と定義した。必要条件を満たした女性に、遺伝学カウンセラーまたは医師によって試験に関する情報が与えられ、参加することを選択した場合はサインされたインフォームドコンセントがもたらされた。各女性のサインおよび完全な同意書を現地で保存した。選択された人口統計および妊娠に関連する情報を、標準化された形態で入手し、併せて、診断手順の前に抜き取った静脈血が入った少なくとも2つ(最大5つ)の上部が紫色の10mLチューブを入手した。データ形態上、および処理された血漿チューブ上の検査コードによってのみ参加者を識別した。多胎妊娠および現存する胎児死亡を伴う妊娠は、全ての胎児について診断検査が計画されていたのであれば適格であった。

【0306】

検出力分析

この試験は、現行の実施を変化させるべきかを決定することを意図していた。したがって、検出率(検査陽性のダウン症候群妊娠の割合、または感度)および偽陽性率(検査陽性の影響を受けていない妊娠の割合、または1-特異度)のどちらの推定にも高信頼度の水準が必要とされた。偽陰性がないという仮定の下で、98%よりも有意に高い検出率を見いだすために検出力が少なくとも80%になるように十分な症例を含めるべきである。200症例を分析することにより、90%の検出力がもたらされ、この下限が棄却される。これらの症例のそれぞれに対して7例の正倍数性妊娠(対照)を選択して、偽陽性率の妥当な信頼度を確実にする。

試料/データ採取

羊水穿刺または絨毛採取の前に血漿試料を抜き取り、Ehrlichら、(Am. J. Obstet. Gynecol. (2011年)204巻:205.e1-11頁)のプロトコルに従って処理した。簡単に述べると、10mLの血漿チューブ(EDTAを含有する、上部が紫色)を4、2,500×gで10分遠心分離し、血漿を50mLの遠心管にプールし、4、15,500×gで10分遠心分離した。次いで、血漿を2つ以上の15mLの円錐チューブに、チューブ1個当たり4mLで移し、最後のチューブは残りの体積全てを含有した。これらのチューブを、長期保管するために登録場所で-70以下の冷凍装置に入れた、または、コーディネートセンターへの1~2日の配送のためにドライアイス輸送する前に24時間以下にわたって-20に置いた。-80で保管した場合、試料は、コーディネートセンターへの1~2日の配送のために、通常は月に1回ベースで、バッチでドライアイス輸送した。場所特異的な試験IDを添えた予め印刷したバーコード標識を使用して全ての血漿チューブを識別した。国際輸送のためにQuick International Courier, Inc.を使用して、適切な追跡、パッケージ内のドライアイスの維持、および配送を確実にした。

【0307】

データ収集には標準化された多成分形態を使用し、予め印刷したバーコード試験標識、収集日、妊娠期間、母体の年齢、体重、人種および民族、手順の徴候、胎児の数、胎児の性別、試料抜き取りの日時、抜き取ったチューブの数、研究所で受け取った時間、および冷凍装置内に置かれていた時間を含めた。1つのコピーをその場所で保持し、他を試料と一緒にコーディネートセンターに輸送した。核型情報を得るために、各女性に対して電子申請書を作成し、各申請書は、手順の日付、妊娠期間、手順(例えば、羊水穿刺、CVS)、診断検査(例えば、核型、qfPCR)、解釈された試験結果(ならびに胎児の性別)、および追加的な胎児に関する結果および所見を含めるために十分な空間を含んだ。処理された血漿チューブとデータ形態のどちらについても、検査コードによってのみ参加者を識別した。

【0308】

分析用の試料の選択

選択の判断基準は、完全な4mLの処理された試料へのアクセス、女性の年齢が18歳以上であること、および重要なデータの欠けがないまたは限られていることを含んだ。最後にいくつか登録された第1三半期後期(14週以内の妊娠期間)および第2三半期初期

10

20

30

40

50

(15～22週の妊娠期間)からの症例は、三半期ごとに100症例の標的が妥当なクッションを伴って達せられたので含めなかった。マッチングは、妊娠期間、母体の人種、母体の民族、登録場所、および冷凍装置内にあった時間に基づいた。試料は、研究所で開発された検査(LDT)が最終的な内部検証、刊行物の提出、および監督委員会の同意を通じた後のみに、処理および検査のためにドライアイス輸送した。選択状況(例えば、一定分量の破損、抽出失敗)では、第2の一定分量を要求することができた。第2の一定分量の数および送付の指標を追跡した。

【0309】

研究所での検査

ライブラリー調製

ライブラリーを調製するために、抽出された、循環している無細胞(ccf)DNAを、さらなる断片化またはサイズ選択をせずに使用した。ccf DNAは、一般に、アベレージの長さ約160塩基対に天然に断片化している。DNA溶出液55 μ Lを、抽出後、ライブラリー調製を開始するまで、低結合性エッペンドルフチューブに入れ4で保管した。保管時間は24～72時間にわたった。ライブラリー調製を製造者の仕様書(Illumina)に従い、本明細書に記載のいくつかの変更を伴って行った。酵素および緩衝液はEnzymatics、MA(End Repair Mix - LC; dNTP Mix(各25mM); Exo(-)Klenowポリメラーゼ; 10xBlue Buffer; 100mMのdATP; T4 DNA Ligase; 2xRapid Ligation Buffer)およびNew England Biolabs、MA(Phusion PCR MM)から供給された。アダプターオリゴヌクレオチド、指標オリゴヌクレオチド、およびPCRプライマーはIllumina Inc、CAから入手した。

【0310】

40 μ Lのccf DNAを末端修復のために取得することによってライブラリー調製を開始し、15 μ Lを胎児定量器アッセイ(FQA)品質管理(QC)のために保持した。試料の末端修復を、1x末端修復緩衝液、24.5 μ Mの各dNTP、および末端修復酵素ミックス1 μ Lの最終濃度を用いて実施した。末端修復反応を室温で30分行い、産物をQiagen Qiaquickカラムで浄化し、溶出緩衝液(EB)36 μ L中に溶出させた。末端修復された試料を1xBlue Buffer、192 μ MのdATP、および5UのExo(-)Klenow Polymeraseの最終濃度と混合することによって、末端修復された試料の3'モノ-アデニル化を実施した。反応物を37で30分インキュベートし、Qiagen MinEluteカラムで浄化し、産物を14 μ LのEB中に溶出させた。1xRapid Ligation buffer、48.3nMのIndex PE Adapter Oligos、および600UのT4 DNA Ligaseと一緒に室温で10分インキュベートすることによってアダプターを断片にライゲーションした。ライゲーション反応物をQiaquickカラムで浄化し、試料を23 μ LのEB中に溶出させた。忠実度の高いポリメラーゼを用いて増幅することによって、アダプターで修飾された試料を濃縮した。各試料の溶出液23 μ L全体を、1xPhusion MM、Illumina PE 1.0および2.0プライマー、ならびに12種の指標プライマーのうちの1種と混合して総PCR反応体積50 μ Lにした。本明細書に記載されている方法およびプロセスは、12種の指標プライマーの使用に限定されない。プラットフォームおよび/または製造者の利用可能性に応じて、任意の数の追加的な指標プライマーを本明細書に記載の方法およびプロセスと一緒に使用することができる。指標プライマーの数が多いほど、フローセルレーンで実行することができる試料の数が多くなる。本明細書に記載の方法およびプロセスでは、試験時に市販されていた指標プライマーを利用した。

【0311】

AB GeneAmp PCR System9700サーマルサイクラーを使用し、0.65mLのPCRチューブで試料を増幅した。増幅するために利用したPCR条件は

10

20

30

40

50

、98 で30秒の最初の変性、98 で10秒の変性、65 で30秒のアニーリング、および72 で30秒の伸長を15サイクル含んだ。72 で5分の最終的な伸長の後に4 で保持した。PCR産物をMinEluteカラムで浄化し、ライブラリーを17 μ LのEB中に溶出させた。

【0312】

配列決定ライブラリーの品質管理 (LabChip GX)

マイクロfluidicsプラットフォームにおける電気泳動による分離によってライブラリーを定量した。各ライブラリーを1:100希釈し、Caliper LabChip GX instrumentを用い、HT DNA1K LabChip、v2およびHiSens Reagent kit (Caliper Life Sciences、Hopkinton、MA)を用いて3連で分析した。Caliper LabChip GXソフトウェアv2.2により、200~400bpからのスミア分析を用いて濃度を算出した。

10

【0313】

クラスタリングおよび配列決定

標準のIlluminaプロトコールに従ってクラスタリングおよび配列決定を実施した。個々のライブラリーを、2nMの濃度に対して正規化し、次いで、4プレックス形式で試料1つ当たり1.2pMまたはフローセル1レーン当たり4.8pMの最終的なフローセルローディング濃度にクラスタリングした。cBOT計器およびv4Single-Read cBOT試薬キットを使用した。HiSeq 2000において、v1HiSeq Sequencing Reagentキットおよび補足的なMultiplex Sequencing Primerキットを使用して単一読み取り多重化配列決定を36サイクル実施した。IlluminaのRTA1.7/HCS1.1ソフトウェアを用いて画像解析および塩基呼び出しを実施した。CASAVAバージョン1.6を使用して配列をUCSChg19ヒト参照ゲノム(反復マスキングしていない)に対してアラインメントした。クラスタリングおよび配列決定は、独特の指標プライマーの利用可能性に応じて、8プレックス、12プレックス、16プレックス、24プレックス、48プレックス、96プレックス、またはそれ以上を用いて実施することもできる。

20

【0314】

データ解析

試料を第21染色体トリソミーと二染色体に分類するために、その内容が参照により本明細書に組み込まれるChiuら、(BMJ(2011年)342巻:c7401頁)およびEhrichら、(Am.J.Obstet.Gynecol.(2011年)204巻:205.e1-11頁)に記載されている方法と同様の方法を利用した。これらの試験に使用された方法とは異なり、本明細書で適用する分類は、臨床診察をシミュレートするために、「オンライン」様式で行った。1つのフローセルが処理されたらすぐに試料を呼び出した。この「オンライン」バージョンの分類予測では、ロバストな位置の推定値および染色体表示の尺度を使用することによって、標準化された染色体表示(例えば、フローセルに対してロバストなzスコア、またはFCに対してロバストなzスコア)を確立するために、フローセルに関連する全てのデータを使用した。染色体*i*についての染色体表示を示す chr_i 、

30

40

【0315】

【化7】

$$chr_i = \frac{\text{カウント}_i}{\sum_{j=1}^{22} \text{カウント}_j}$$

(式中、 カウント_j は染色体*j*上のアラインメントされた読み取りの数である)を用いて、染色体*i*を伴う試料*N*についてのFCに対してロバストな染色体のzスコアの方程式は

50

【0316】

【化8】

$$Z_N = \frac{\text{chr}_{iN} - \text{中央値}(\text{chr}_i)}{\text{MAD}(\text{chr}_i)}$$

である。尺度のロバストな推定値について中央絶対偏差 (MAD) の正規化された形態を使用し、

【0317】

【化9】

$$\text{MAD}(X) = \frac{1}{\Phi^{-1}(3/4)} \cdot \text{中央値}(|X - \text{中央値}(X)|),$$

正規分布したランダムな変数の標準偏差に近づくように乗法定数を選択した。試料は、 $Z_N > 3$ の場合は第21染色体に関してトリソミーであると呼び出され、他の場合は二染色体であると呼び出された。

【0318】

反復領域のフィルタリングおよびGC正規化

ヒトゲノムでは、現行の検出方法を用いて推定することができる反復ゲノム配列は、最大でゲノム全体の半分を表す。これらの反復性の領域は、単純な反復、またはタンデムな反復（例えば、大部分は染色体のセントロメアおよびテロメアにおいて見いだされるサテライトDNA、ミニサテライトDNA、マイクロサテライトDNA）、またはセグメント重複および分散反復（例えば、SINES、LINES、DNAトランスポゾン）の形態をとり得る。そのような重複のサイズは、数塩基対 (bp) から、数百bpまで、およびはるか10~300キロベース対までにわたり得る。これらの領域の反復性は、次世代シーケンシング技法のいくつか、例えば大規模並列処理ショットガン配列決定に存在するPCR増幅ステップの変動性の原因であると考えられている。

【0319】

そのような反復性の領域にマッピングされた読み取りの、分類の正確度に対する影響を評価するために、全ての試料を、そのような読み取りを染色体表示の図表化に含めて、または含めずに分析した。試料を、反復ゲノム配列の寄与を除去することの利益を伴って、または伴わずに分析した。効率的なコンピュータ処理のために、短い読み取りをアラインメントするために使用する参照ゲノムは「反復マスキングされた」バージョンではなく、そのような反復性の領域を含むものであった。アラインメント後、基礎を有するフィルタリング手順を Repeat Library 20090604 (URLワールドワイドウェブ repeatmasker.org) に含有される情報に対して利用した。次いで、反復マスキング認識分類のために、反復された領域とオーバーラップしていない読み取りのみを染色体表示の推定について考慮に入れた。

【0320】

GC含量が異なるゲノム配列により、時には、PCRステップの間に異なる増幅効率もたらされ、今度はこれにより、時には、元のゲノムの材料の試料採取が偏ったものになる。この潜在的な増幅の偏りを補償するために、各50Kbのピンについてのカウントを集約し、Alkanら (Nat. Genet. (2009年) 41巻: 1061~1067頁) に記載されているものと同様のLOESS技法を用いることによって、ピンに特異的なGC含量に関してさらに正規化した。次いで、推定のGCの偏りに関して正規化された、フィルタリングされたカウントを、染色体表示を決定するために使用する。

【0321】

本明細書に記載の読み取りのフィルタリングおよびカウントの正規化手順は第21染色

10

20

30

40

50

体の倍数性の「オンライン」分類には使用されなかったが、その後の解析の一部として使用され、全ての試料についてのデータセットが、非盲検化される前にSCMMからコーディネーターに送達された。反復マスキングに関するフィルタリングならびにGC正規化手順の両方を適用した後に算出された染色体表示は、この試験では、「GCについて補正された染色体表示」と称され、そのような染色体表示から算出されたzスコアは「GCについて補正されたzスコア」と称される。

【0322】

SCMM-SD研究所では、1,640の試料の全てについてステップの全てが実施された。UCLA研究所では、これらの試料の約40%についてライブラリー調製物が受け取られ、次いで、検査プロトコルが完了した。7例のダウン症候群症例および対照を含む1つの試料のセット（例えば、1つのプレート；3つのフローセル；約96の試料）について、別々の4mLの処理された血漿試料がSCMM-SD研究所およびUCLA研究所の両方に輸送され、LDT全体が2連で実施された。両方の研究所からの試験結果を有する試料のいずれについても、SCMM-SDからの結果を一次結果とみなした。

【0323】

結果および考察

本明細書の図4～図19に示されている表にしたグラフデータは、212例のダウン症候群妊娠および1,484例の正倍数性妊娠の全てについての胎児分率の共変量分析（胎児に由来する遊離型の循環しているDNAの百分率）を含む。データの可視性を改善するために、カテゴリーデータを標識した目盛の左側および右側に「ディザ処理」した。試験された妊娠の全てについて試料採取時に生育可能であり、また、全てが、診断検査結果（例えば、核型）が入手可能な単胎児妊娠であることが検証された。ディザ処理は、多くの場合、オーバープロットを回避するためにデータ点をランダムにジッターリングするまたはわずかにシフトさせることである。X軸座標をわずかに変動させて、プロットの全体的な見え方は変化させずに、そのカテゴリーについての個々の点を可視化することを可能にした。配列決定の前に胎児分率の検査結果が入手可能であったので、それらを使用して、試料の妥当性を決定した。許容できる胎児分率は両端を含めて4%から50%の間であった（グラフの薄い横破線）。臨床診察では、この範囲の外側の試料を、配列決定するために許容されないとみなすことができる。全体的な胎児分率中央値14.0%（幾何平均13.4%、算術平均15.0%）が図1～図3に薄い横の実線として示されている。胎児分率が4%未満の場合、ダウン症候群由来の循環しているDNAと正倍数性妊娠由来の循環しているDNAの間の小さな差異を分解することが難しい。より高レベルでは、試料の取扱いの潜在的な問題が示される。胎児分率の分布は右側に歪んでいる。このような理由で、表示および分析は対数変換後に行う。回帰分析を用いて探究される共変量については、結果が統計的有意性に達しなかった場合は回帰直線のみが示されている。他の点では、95%予測限界も示されている。

【0324】

胎児分率を、試料の抜き取りと冷凍装置での保管の間の時間に応じて分析した。正倍数性妊娠についての分析の結果を使用すると、冷凍装置に入れるまで1時間、2時間、3時間、4時間および5時間についての予測される胎児分率は、それぞれ13.5%、13.2%、12.8%、12.5%および12.2%になる。

【0325】

試料の溶血の状態を、凍結する前に登録場所で評価した。全くなし、わずか、中程度、および著しいの標準のスキームを使用した。全くなしおよびわずかを、その後「なし」カテゴリーに群分けし、中程度および著しいを「あり」カテゴリーに群分けした。溶血を伴う胎児分率について有意差はなかった（なし、およびありについて、それぞれ、平均=13.2%および13.6%、 $t = -0.46$ 、 $p = 0.64$ ）。ダウン症候群妊娠については、溶血を伴うものについて差異はあったとしてもわずかであった（それぞれ平均=15.4%および15.0%、 $t = 0.14$ 、 $p = 0.89$ ）。

【0326】

地理的地域によって層別化された（左から右に、13.9%、13.1%、12.8% および13.4%の平均胎児分率、ANOVA $F = 1.93$ 、 $p = 0.12$ ）またはダウン症候群妊娠の中で（左から右に17.4%、15.0%、14.5%および15.9%の平均胎児分率、ANOVA $F = 1.45$ 、 $p = 0.23$ ）、パーセント胎児分率（正倍数性妊娠）に有意な関係はなかった。

【0327】

診断検査の徴候によって層別化された（左から右に13.0%、13.2%、13.4%、12.7%、13.1%、14.1%、15.6%、および13.3%の平均胎児分率、ANOVA $F = 0.61$ 、 $p = 0.75$ ）、またはダウン症候群妊娠の中で、（左から右に14.9%、15.0%、15.6%、15.3%、14.8%、NA、13.0%、および15.7%の平均胎児分率、ANOVA $F = 0.11$ 、 $p = 0.99$ ）パーセント胎児分率について有意な関連はなく、再度関連は示されなかった。

10

【0328】

少なくとも50の試料を用いた、登録場所によって層別化されたパーセント胎児分率については有意差があり（10.2%～18.7%にわたる平均胎児分率、ANOVA $F = 5.59$ 、 $p < 0.0001$ ）、ダウン症候群妊娠の中での同じ分析については有意差がない（12.7%～16.9%にわたる平均胎児分率、ANOVA $F = 0.35$ 、 $p = 0.97$ ）。最も高い胎児分率を有する5つの登録場所におけるアベレージ体重は151ポンドであり、それと比較して、より低い胎児分率を有する6つの場所では150ポンドであったので、これは、母体の体重が異なることでは説明されない（図B8参照）。

20

【0329】

図1：x軸は、試料の抜き取り時の妊娠期間を示す。上のパネル（正倍数性妊娠）は、妊娠期間による胎児分率を示す。線形回帰によって有意な関係は見いだされなかった（太い破線、 $p = 0.23$ 、傾き = -0.0024）。ダウン症候群妊娠の分析（下のパネル）では同様の結果が見いだされた、（ $p = 0.10$ 、傾き = 0.0084）。

【0330】

図2：x軸は、推定分娩日における母体の年齢を示す。上のパネル（正倍数性妊娠）は、母体の年齢による胎児分率を示す。線形回帰によって有意な関係は見いだされなかった（太い破線、 $p = 0.23$ 、傾き = -0.0013）。ダウン症候群妊娠の分析（下のパネル）では同様の結果が見いだされた（ $p = 0.26$ 、傾き = -0.0031）。

30

【0331】

図3：x軸は、試料の抜き取り時の母体の体重をポンド単位で示す。上のパネル（正倍数性妊娠）は、正倍数性妊娠からの、母体の体重による胎児分率を示す。線形回帰によって有意な関係が見いだされた（太い破線、薄い破線で示されている95%予測限界を伴う、 $p < 0.0001$ 、傾き = -0.0026）。ダウン症候群妊娠について同様の結果（下のパネル）が見いだされた（ $p = 0.0002$ 、傾き = -0.0017）。例として正倍数性の結果を使用して、体重が100ポンド、150ポンド、200ポンド、250ポンドおよび300ポンドの女性は、それぞれ17.8%、13.2%、9.8%、7.3%および5.4%のアベレージ胎児分率を有することが予測される。

【0332】

膣からの出血が報告された女性（正倍数性妊娠）について、わずかであるが有意な胎児分率の減少があった（なしおよびありについて、それぞれ平均 = 13.3%および12.3%、 $t = 2.04$ 、 $p = 0.04$ ）。ダウン症候群妊娠の中での同じ分析については、出血が報告された女性について有意な増加があった（それぞれ平均 = 14.7%および17.6%、 $t = -2.07$ 、 $p = 0.04$ ）。

40

【0333】

男の正倍数性胎児と女の正倍数性胎児の間で（それぞれ平均13.4%および12.9%、 $t = 1.68$ 、 $p = 0.094$ ）、またはダウン症候群妊娠の中で（平均 = 15.2%および15.3%、それぞれ、 $t = -0.05$ 、 $p = 0.96$ ）、胎児分率に差はなかった。

50

【0334】

ダウン症候群妊娠では正倍数性妊娠よりも胎児分率が高く、これは統計的に有意であった（平均15.2%対13.2%、 $t = -4.11$ 、 $p < 0.0001$ ）。これをダウン症候群についてのスクリーニング検査として用いる場合、5%および10%の偽陽性率で、対応する検出率は、それぞれ9.0%および17.5%になる。これらは、約1.8の累積オッズ比に対応する。

【0335】

胎児分率の共変量分析により、母体の体重が遺伝的変異の決定における有意な因子であることが明らかになった。アベレージ体重100ポンドおよび250ポンドでは、予測される胎児分率は、それぞれ17.8%および7.3%である。母体の体重の影響により、母体の人種および民族性に対して、胎児分率について見いだされた小さいが有意な影響を説明することができる。試料の抜き取りから冷凍装置での保管までの時間も、胎児分率に有意に影響を及ぼし、時間が長いことにより、胎児分率がわずかに低くなる。しかし、試料の抜き取りから冷凍装置での保管までについて見られる影響は、母体の体重についてよりも実質的に小さい。残りの関連は、一般に、小さく、大抵は有意ではない。

【0336】

図4～図6においてグラフで示されているデータには、第21染色体表示（例えば、パーセント第21染色体）とアッセイの変動性との関係が要約されている。患者4人由来の試料は、一般に、単一のフローセルレーンにおいて4ブックスにした（例えば、8レーンは32の患者と等しい）。しかし、大抵は、30の患者試料についてのみ実行し、追加的な位置保持対照を伴った。92の患者について96ウェルプレートで一緒に処理した。各プレートについて3つのフローセルで実行した（例えば、レーンごとに4ブックスおよび4指標プライマーを使用する場合、1つの試料プレートについて3つのフローセルで実行した）。一般に、7つのデータのプレートを一緒に群分けしてバッチを形成した。各バッチは、順不同で配分された試料を含有した。したがって、バッチ内の症例および対照は必ずしも同じ試料プレートまたはフローセルで実行されなかった。症例および対照と一緒に実行することは、時には、マッチさせた分析における総分散の推定の下であり得る。212例のダウン症候群全て、および1,484例のうち13例を除く全ての正倍数性の結果が図4～図6に示されている。試料が、最初は失敗したが、第2の結果は上首尾であった場合は、第2の結果が示されている。反復試料において使用できる結果を生じさせることができなかつた試料は示されていない。試験された妊娠の全てについて試料採取時に生育可能であり、また、全てが、診断検査結果（例えば、核型分析）が入手可能な単胎児妊娠であることが検証された。

【0337】

図4は、フローセルによるC21%の結果を示す。第21染色体のマッチした読み取りの百分率を常染色体の読み取りの総数で割ったものを、正倍数性（小さな丸）およびダウン症候群（より大きな丸）についてフローセル数（x軸）によってプロットした。各フローセルでは32の試料（4ブックスで）検査することができ、それにより、28～30の患者試料が対照試料と一緒にもたらされる（各フローセルにおいて実行した患者試料が全てこの報告に含まれるとは限らない）。一般に、それぞれについて、20～25例の正倍数性妊娠および2～7例のダウン症候群妊娠が示されている。いくつかの場合には（例えば、反復を伴うフローセル）、数ははるかに小さい。全体的に、76のフローセルが、追加的な一定分量の検査を含めた本試験と関連性のあるデータを含有した。フローセルに連続して番号を付し、欠けているフローセルは独立した研究所での検査を含めた他の試験のために使用した。平均レベルでのフローセル間の変化を見ることができる。また、初期フローセルでは正倍数性の平均1.355%を上回る明白な傾向があるが、後期のフローセルでは低くなる傾向がある。フローセルの中で、正倍数性の結果の標準偏差に差異は存在しない。基準線を、正倍数性試料についての全体的なアベレージ胎児分率である1.355%で引いた。平均レベルでのフローセル間の変動性を見ることができるが（ANOVA、 $F = 4.93$ 、 $p < 0.001$ ）、標準偏差は一定である（ $F = 1.1$ 、 $p = 0.3$

1)。

【0338】

図5は、図4と同じデータを含むが、データがフローセルではなくプレートによって層別化されている。処理を96ウェルプレートで実施する。次いで、1つのプレートからの処理された試料について、3つのフローセルで実行する。基準線は1.355%にある。平均レベルでのプレート間の変動性を見ることができるが(ANOVA、 $F = 13.5$ 、 $p < 0.001$)、標準偏差は一定である($F = 1.2$ 、 $p = 0.23$)。この図において、図4で明らかになったものと同じ傾向を見ることができる。全体的な分散の減少は、プレート間の差異を考慮した場合、フローセル間と比較していくらか少ない。しかし、プレートの差異を考慮すると、フローセルの差異について有意な影響はない。図4において見られるように、プレートの中で正倍数性の結果の標準偏差に差異は存在しない。

10

【0339】

図6は、図4および図5と同じデータを含むが、データが配列決定するために使用したIllumina計器に応じて層別化されている。42および34のプレートをそれぞれ2番および3番で処理した。基準線は1.355%にある。正倍数性(それぞれ平均1.355および1.354、 $t = 2.0$ 、 $p = 0.16$)またはダウン症候群妊娠(それぞれ平均1.436および1.438、 $t = 0.32$ 、 $p = 0.57$)において、計器による第21染色体のパーセントに差異は存在しない。2つの機械によりもたらされたC21%に系統的な差異は存在しない。

20

【0340】

212例のダウン症候群の結果全ておよび1,484例のうち13例を除く正倍数性の結果の全てについての15の潜在的な共変量が、臨床的に報告された第21染色体のzスコアに対して要約されている。試験された妊娠の全てについて試料採取時に生育可能であり、また、全てが、診断検査結果(例えば、核型分析)が入手可能な単胎児妊娠であることが検証された。1つのダウン症候群の試料のzスコアが25をわずかに超えたが、24.9にプロットした。正倍数性試料の範囲は-3から+3の間である。症例の中で、カットオフレベル3を使用した。zスコアの分布は、症例では右側に歪んでいるが、対照ではガウス分布である。しかし、それでもデータを線形尺度でプロットした。症例における回帰分析は対数変換後に行った。

30

【0341】

検査のために選択された全ての試料を採取してから6時間以内に処理し、冷凍装置に保管した。試料の抜き取りから冷凍装置での保管までの時間による第21染色体のzスコアについては、正倍数性妊娠またはダウン症候群妊娠のいずれについても、線形回帰によって有意な関係は見いだされなかった(それぞれ $p = 0.90$ 、傾き = -0.0025 ; および $p = 0.50$ 、傾き = -0.20)。

【0342】

溶血の状態を、凍結する前に登録場所で評価した。いずれの群についても溶血の状態によって層別化した後のzスコアに有意差はなかった(正倍数性妊娠およびダウン症候群妊娠について、それぞれ $t = -0.01$ 、 $p = 0.99$ および $t = -0.12$ 、 $p = 0.90$)。

40

【0343】

地理的地域によって層別化されたzスコアについて、正倍数性妊娠について(左から右に、平均zスコア-0.22、-0.14、-0.12および-0.01、ANOVA $F = 1.84$ 、 $p = 0.14$)またはダウン症候群妊娠の中で(左から右に、平均zスコア10.1、9.9、8.9および10.2、ANOVA $F = 1.00$ 、 $p = 0.39$)有意な関係はなかった。

【0344】

診断検査の徴候によって層別化されたzスコアについて、正倍数性妊娠についてはわずかだが有意な影響があったが(左から右に、平均zスコア-0.15、-0.14、-0.24、-0.05、-0.11、0.20、-0.52および-0.20、ANOVA

50

$F = 2.02$ 、 $p = 0.049$) ダウン症候群妊娠については有意な影響はなかった (左から右に、平均 z スコア 8.9 、 9.1 、 9.7 、 9.8 、 10.0 、 n/a 、 10.7 および 9.5 、ANOVA $F = 0.25$ 、 $p = 0.96$)。

【0345】

少なくとも50の試料を用いた、1つまたは複数の登録場所によって層別化された z スコアについては、正倍数性妊娠 ($-0.21 \sim 0.02$ にわたる平均 z スコア、ANOVA $F = 0.57$ 、 $p = 0.84$) またはダウン症候群妊娠 ($6.90 \sim 12.34$ にわたる平均 z スコア、ANOVA $F = 1.45$ 、 $p = 0.16$) について影響はない。

【0346】

図7: x 軸は、試料の抜き取り時の妊娠期間を示す。上のパネル (正倍数性妊娠) は、妊娠期間による z スコアを示す。線形回帰によって有意な関係は見いだされなかった ($p = 0.79$ 、傾き = 0.0023)。ダウン症候群妊娠 (下のパネル参照) では、妊娠期間との有意な正の関連が見いだされた ($p = 0.0023$ 、傾き = 0.017 、 z スコアの対数で)。

10

【0347】

図8: x 軸は、推定分娩日における母体の年齢を示す。上のパネル (正倍数性妊娠) は、母体の年齢による z スコアを示す。線形回帰によって有意な関係は見いだされなかった (太い破線、 $p = 0.62$ 、傾き = -0.0023)。ダウン症候群妊娠の分析 (下のパネル) では同様の結果が見いだされた ($p = 0.14$ 、傾き = -0.0046)。

【0348】

20

図9: x 軸は、試料の抜き取り時の母体の体重をポンド単位で示す。上のパネル (正倍数性妊娠) は、正倍数性妊娠についての試料についての母体の体重による z スコアを示す。線形回帰によって、有意な負の傾きが見いだされた (太い破線、薄い破線によって示されている95%予測限界を伴う、 $p = 0.029$ 、傾き = -0.0016)。ダウン症候群妊娠について、同様であるが、はるかに大きな影響が見られた (下のパネル、 $p = 0.0003$ 、傾き = -0.038)。この後者の影響は、おそらく胎児分率に対する母体の体重の影響に起因する (図11を参照されたい)。

【0349】

正倍数性妊娠については、報告された膈からの出血の状態による z スコアに有意差はなかった (なしおよびありについて、それぞれ平均 = -0.14 および -0.09 、 $t = -0.65$ 、 $p = 0.52$)。ダウン症候群妊娠の中での同じ分析については、出血が報告された女性について有意な増加があった (それぞれ平均 = 9.03 および 11.70 、 $t = -3.14$ 、 $p = 0.0019$)。

30

【0350】

母体の人種によって層別化された z スコアについて、正倍数性妊娠について (左から右に平均 z スコア -0.14 、 -0.15 、 0.28 および -0.21 ; ANOVA $F = 2.44$ 、 $p = 0.063$) またはダウン症候群妊娠について (左から右に平均 z スコア 9.55 、 8.90 、 9.63 および 10.24 、ANOVA $F = 0.12$ 、 $p = 0.95$) 有意な影響はない。

【0351】

40

白人民族性によって層別化された z スコアについて、正倍数性妊娠について (左から右に平均 z スコア -0.16 、 -0.06 および 0.00 、ANOVA $F = 1.70$ 、 $p = 0.18$) またはダウン症候群妊娠について (左から右に平均 z スコア 9.5 、 9.4 および 11.9 、ANOVA $F = 0.38$ 、 $p = 0.68$) 有意な影響はない。

【0352】

胎児の性別によって層別化された z スコアには、正倍数性妊娠について (それぞれ平均 = -0.13 および平均 = -0.13 、 $t = -0.04$ 、 $p = 0.97$) またはダウン症候群妊娠について (それぞれ平均 = 9.25 および平均 = 9.80 、 $t = -0.85$ 、 $p = 0.39$)、男と女との間に差異はない。

【0353】

50

冷凍装置での保管時間による z スコアについて、正倍数性妊娠について（太い破線、 $p = 0.72$ 、傾き = 0.000057 ）またはダウン症候群妊娠について（下のパネル、 $p = 0.25$ 、傾き = -0.0022 ）、線形回帰によって有意な傾きは見いだされなかった。

【0354】

図10：上のパネル（正倍数性妊娠）は、DNAライブラリー濃度に対する z スコアを示す。線形回帰により、統計的に有意な正の傾きが示されている（太い破線、薄い破線によって示されている95% 予測限界を伴う、 $p < 0.0001$ 、傾き = 0.0034 ）。ダウン症候群妊娠については、同様であるが有意ではない影響が見られる（下のパネル、 $p = 0.82$ 、傾き = 0.0024 ）。

10

【0355】

数百万のマッチしたDNA配列による z スコアについての線形回帰では、正倍数性妊娠について（太い破線、 $p = 0.47$ 、傾き = 0.0072 ）およびダウン症候群妊娠について（下のパネル、 $p = 0.94$ 、傾き = 0.0099 ）、有意でない正の傾きが見いだされた。

【0356】

胎児分率の共変量分析について記載の通り、第21染色体の z スコアの共変量分析により、母体の体重も遺伝的変異の決定における有意な因子であることが明らかになったが、見られる影響はダウン症候群妊娠の中でより大きかった。いくつかの場合には、妊娠期間も有意な正の関連を有する。しかし、妊娠期間と共にみられる影響は、母体の体重に関して見られる影響よりも有意に小さい。残りの関連は、一般に、小さく、大抵有意でない。

20

【0357】

以下の表3には、MPS検査によって最初に誤って分類された6つの試料に関する追加的な詳細な情報が提供される。3つの場合では、ダウン症候群と確認された被験体は、最初にダウン症候群を有しないと分類され（試料ID番号162、167および371）、3つの場合では、健康な子であると確認された被験体は、最初にダウン症候群に分類された。

【0358】

【表 3】

表3: MPSS検査による6つの誤分類に関する詳細な情報

	ID = 162	ID = 167	ID = 371	ID = 22	ID = 221	ID = 249
T21 zスコア	+0.83	+1.50	+1.57	+3.82	+4.72	+3.56
MPSSによる解釈	DSでない	DSでない	DSでない	DS	DS	DS
核型	47, XX +21	47, XY +21	47, XY +21	46, XY	46, XX	46, XX
確認	核型により確認された	剖検で確認された	プロバイダーによって確認された	「健康な男児」と確認された	「健康な女児」と確認された	「健康な女児」と確認された
妊娠期間(週)	偽陰性 9.2	偽陰性 14.6	偽陰性 13.0	偽陽性 12.1	偽陽性 10.0	偽陽性 13.6
母体の年齢(歳)	42	43	40	41	33	39
母体の体重(ポンド)	200	165	182	125	174	185
人種/民族性	白人	白人	白人	白人、ラテンアメリカ系	白人	白人
出血	なし	なし	なし	あり	あり	なし
照会理由	母体の年齢およびhX異数性	母体の年齢および統合スクリーニング	第1三半期スクリーニング	母体の年齢が38歳以上	母体の年齢およびhX異数性	第1三半期スクリーニング
処理時間(時間)	1	3	3	1	1	1
試料の体積 (mL)	4.0	4.0	3.8	3.9	4.0	4.0
溶血	わずか	NR	なし	なし	わずか	わずか
胎児分率(%)	4	7	5	19	24	11
注	第1の試料失敗-低胎児DNA			第1の試料失敗-高胎児DNA		

10

20

30

40

大規模並列処理ショットガン配列決定のプロセス全体についてのフローセルによる日数単位の総ターンアラウンドタイム (TAT) を分析した。処理されたフローセルの最初の3分の1について、総ターンアラウンドタイム (TAT) は、我々の刊行物に記載の臨床サインアウトの前にアルゴリズムに行った改変に起因して、コンピュータ解釈時間が優位を占めた。臨床サインアウトのプロセスは経時的に改善された。2つのフローセル (試験の最初から最後まででの約3分の2) は完全に再配列決定する必要があり、この結果、TATが増加した。最後の20フローセルの間、TATは、18について10日標的の範囲内であった (90%)。真の臨床的環境でのTATは、2つの潜在的な改善に基づいていくらかより良い可能性がある: 本試験では、試料を週末にかけて処理せず、また、所与の日

50

に専任の臨床医が常にサインアウトのために対応できたわけではなかった。試料の約 5 % について反復し、それにより、それらの試料については T A T がおよそ倍になった。

【 0 3 5 9 】

正倍数性試料およびダウン症候群の試料の同定についての成功 / 失敗率により、ダウン症候群妊娠由来の 2 1 2 の試料の中で、解釈成功率 (9 2 %) ならびに検査失敗の理由がもたらされた。これらの 1 7 人の女性由来の新しい一定分量の反復検査により、試料の 1 0 0 % が解釈成功を有した。1 , 4 8 4 の検査された正倍数性妊娠について分析を反復した。合計 1 3 の試料が、第 2 の一定分量を試験した後にも検査失敗とみなされた。全体的に、M P S S の実施の成功率は 9 9 . 2 % であり、最初の試料の 5 % で第 2 の一定分量が必要であった。

10

【 0 3 6 0 】

以下に示されている表 4 には、S C M M および U C L A 研究所において試験された 7 9 のダウン症候群試料および 5 2 6 の正倍数性試料についての最終的な M P S S による解釈の比較に関する追加的な詳細な情報が提供される。6 0 5 の試料についての混合ライブラリーが Sequenom Center for Molecular Medicine (S C M M) において調製され、検査され、凍結され、次いで、再検査のために、独立した U C L A 研究所に輸送された。S C M M (それぞれ 9 8 . 7 % および 0 %) における検出率および偽陽性率は、有意ではないがわずかに、U C L A における検出率および偽陽性率 (それぞれ 9 7 . 5 % および 0 . 2 %) よりも良かった。しかし、失敗率は、有意ではないがわずかに、U C L A では S C M M に対して低かった (それぞれ、ダウン症候群において 0 % および 2 . 5 % ; 正倍数性妊娠において 3 . 9 % および 4 . 4 %) 。

20

【 0 3 6 1 】

【表 4】

表4. 2つの研究所において試験された79のダウン症候群の試料および526の正倍数性試料の最終的なMPSSIによる解釈の比較

SCMM

	ダウン症候群			正倍数性			合計
	真陽性	偽陰性	失敗	真陰性	偽陽性	失敗	
UCLA							
ダウン症候群							
真陽性	76	0	1				77
偽陰性	0	1	1				2
検査失敗	0	0	0				0
正倍数性							
真陰性				500	0	4	504
偽陽性				1	0	0	1
検査失敗				2	0	19	21
合計	76	1	2	503	0	23	605

10

20

30

第21染色体のパーセント表示スコアをGC含量およびプレートに基づく実験条件について補正することの影響を分析した。GC補正により、正倍数性妊娠の中で高い（および低い）外れ値の存在が減少し、同時にデータの広がりが減少した。いかなる補正（x軸）も伴わずに、1.38%のカットオフにより、4つの偽陰性の結果および3つの偽陽性の結果がもたらされた。GC補正を用いると、4つの偽陰性の結果のうちの2つ、および3つの偽陽性の結果の全てが、同じ1.38%のカットオフを使用して分解された。しかし、偽陰性の結果のうちの1つおよび新しい偽陽性の結果がカットオフライン上にかかった。残りの第4の偽陰性の解釈は変化しない。MoMを創出するためにプレート補正を加えることによって、3つの偽陽性の全て、および4つの偽陰性のうちの3つが、灰色の帯域の水平な四角形に入る任意のカットオフによって潜在的に分解された。

40

【0362】

1,471の正倍数性の例および212のダウン症候群の症例について、GC含量およびフローセルの変動性について補正した第21染色体のzスコアを使用することにより、

50

2つの偽陰性および3つの元の偽陽性がzスコアカットオフ3(「オンライン」呼び出しアルゴリズムと等しい)を使用して分解された。しかし、新しい偽陽性が1つ生じた。

【0363】

下に示されている表5では、この試験プロトコールおよび結果が、同じくダウン症候群についてスクリーニングするために母体の血漿の大規模並列処理配列決定を用いる以前公開された試験と比較されている。

【0364】

【表5】

特性	本試験	Ehrich 2011	Chiu 2011	Sehnert 2011	
多重化	4プレックス	4プレックス	2プレックス ¹	NR	10
ダウン症候群(N)	212	39	86	13	
正倍数性/非ダウン症候群	1,484	410	146	34	
Illuminaプラットフォーム	HiSeq 2000	GAllx	GAllx	x	
CLIA研究所において実施された	実施された	実施されていない	実施されていない	実施されていない	
シミュレートを実施したか	実施した	実施していない	実施していない	実施していない	
フローセル	76	>15	> 16	NR	20
試験母集団	北米、南米、ヨーロッパ、オーストラリア	米国	香港、オランダ、イギリス	米国	
週単位の妊娠期間(平均値、範囲)	15 (8 - 22)	16 (8 - 36)	13 (NR)	15 (10-28)	
三半期、第1/第2(%)	50/50	NR	88/12	58/42	
失敗(n/N、%)	13/1696 (<1)	18/467 (3.9)	11/764 (1.4)	0/47	
検出率(%)	209/ 212 (98.6)	39/39 (100)	86/86 (100)	13/13 (100)	
偽陽性率(%)	3/1471 (0.2)	1/410 (0.2)	3/146 (2.1)	0/34 (0)	
処理量(試料/週)	250	NR	NR	NR	30
必要体積	>3.5 mL	> 3.5 mL	>2 mL	約4 mL ³	
利用可能な第2の試料	あり	なし	なし	あり	
推定される胎児分率	全て	全て	男のみ	NR	
ターンアラウンドタイム ² (日数)	8.8 ⁴	10 ⁵	NR	NR	
¹ 報告書には8プレックスも含まれるが、2プレックスについての結果のみが示されている ² 処理の開始から配列決定完了まで(アラインメントまたはサインアウトを含まない) ³ 著者らの記述、「単一の[10mL]血液チューブからの血漿が配列決定のために十分であった」 ⁴ 最後の20フローセル[それぞれ32の試料]の平均 ⁵ 著者らの記述、「各バッチ[96の試料]について、DNA抽出から最終的な配列決定の結果までおよそ10日間が必要であった」					

(実施例3)

循環している無細胞DNAを利用した微小欠失の検出

出生前診断の分野は、母体の血漿から単離された循環している無細胞(ccf)胎児DNAの分子キャラクタリゼーションを可能にする技法を実行することを通じて進歩してきた。次世代シーケンシング方法体系を使用して、染色体異常を検出することができることが示されている。21トリソミーの検出は、分析的に、および大規模臨床試験においての両方で検証されている。13トリソミーおよび18トリソミー、性異数性、および他の稀な染色体異常に関する同様の検証がおそらく近い将来後に続くであろう。

【0365】

c c f 胎児DNAを分析物として使用して未だ徹底的には対処されていない遺伝子異常 (g e n e t i c a n o m a l y) の一面は、染色体領域内での (s u b - c h r o m o s o m a l) コピー数多型 (C N V) である。原因不明の発達遅延 / 知的障害 (D D / I D)、自閉症スペクトラム障害 (A S D) または多数の先天性異常 (c o n g e n i t a l a n o m a l y) (M C A) の個体のおよそ12%で、臨床的に関連性のある C N V が診断されている。

【0366】

そのような臨床的に関連性のある状態の1つの例は、ディジョージ症候群、口蓋心臓顔面症候群、および円錐動脈幹異常顔貌症候群を含めた多数の状態で構成される障害である 22q11.2欠失症候群である。これらの状態の正確な顕在化はわずかに変動するが、それぞれが、相同組換えを可能にする反復エレメントが存在することに起因して重複および微小欠失がどちらも高レベルである傾向があることが示されている第22染色体上の約3百万塩基対 (b p) の遺伝子リッチ領域のヘテロ接合性欠失に関連付けられている。染色体22q11.2欠失症候群は、およそ4000件に1件の生児出生に影響を及ぼし、また、頻繁な心臓欠陥、口蓋裂、発達遅延、および学習障害を特徴とする。

【0367】

本明細書には、母体の血漿由来の c c f DNA の配列決定によって染色体領域内での C N V を検出することの技術的な実現性を決定するために実施された調査の結果が記載されている。核型分析によって22q11.2欠失症候群の影響を受けることが確認された、それぞれ胎児を有する2人の女性由来の母体の血漿、および対照として、胎児の異数性のリスクが低い14人の女性由来の母体の血漿を検査した。各試料由来の c c f DNA について、H i s e q 2000計器において2つの個々のレーンを使用して配列決定し、その結果、およそ4xゲノムのカバレッジがもたらされた。2つの検証された症例における既知の影響を受けた領域に対応する第22染色体上の3百万bpの領域の表示の、対照と比較して統計的に有意な減少が検出され、これにより、母体の血漿由来の c c f DNA の配列決定によって染色体領域内での C N V を検出することの技術的な実現性が確認されている。

【0368】

材料および方法

試料の獲得

2つの別々の治験審査委員会 (I n v e s t i g a t i o n a l R e v i e w B o a r d) (I R B) に認可された臨床試験実施計画書 (W e s t e r n I n s t i t u t i o n a l R e v i e w B o a r d I D 20091396およびC o m p a s s I R B 00462) の下で試料を採取した。侵襲的手段の前に、2つの影響を受けた血液試料を採取した。これらの試料に22q11.2微小欠失が存在することを、非経胎盤羊水穿刺によって得られた材料に対する核型分析によって確認した。14の対照試料を、その後の侵襲の手順を伴わずに採取し、したがって、対照試料については核型情報が利用不可能であった。全ての被験体は、E D T A - K 2 噴霧乾燥10mLバキュテナー (B e c t o n D i c k i n s o n , F r a n k l i n L a k e s , N J) 中に全血30~50mLを採取するための静脈穿刺を含めた試験に関連するいずれの手順も書面のインフォームドコンセントをもちた後に受けた。試料を、処理するまで冷蔵した、または湿った氷上で保管した。採血の6時間以内に、母体の全血をE p p e n d o r f 5810Rプラススイングローターを使用して4、2500gで10分遠心分離し、血漿を採取した (例えば、約4mL)。血漿について、E p p e n d o r f 5810Rプラス固定角ローターを使用して、4、15,000gで10分、2回目の遠心分離を行った。2回目の回転後、血漿をチューブの底に形成されたペレットから取り出し、4mLの血漿バーコード一定分量に分配し、すぐに-80で凍結してDNA抽出まで保管した。

【0369】

核酸抽出

QIAamp Circulating Nucleic Acid Kitを製造者のプロトコール(Qiagen)に従って使用して、ccfDNAを母体の血漿から抽出し、Buffer AVE(Qiagen) 55 μ L中に溶出させた。

【0370】

胎児定量器アッセイ

ccfDNAの相対的な品質および量を、当技術分野で公知の方法に従って胎児定量器アッセイ(FQA)によって評価した。FQAでは、母体のccfDNAおよび胎児のccfDNAの間のDNAのメチル化の差異を、定量するための基礎として用いる。16の分析された試料のそれぞれに対して、これによりその全体が参照により組み込まれるEhrlichらおよびPalomakira(Genet Med.(2011年)13巻(11号):913~20頁およびGenetics in Medicine(2012年)14巻:296~305頁)において以前に記載されている通りFQA分析を実施した。

10

【0371】

配列決定ライブラリー調製

TruSeqライブラリー調製(Illumina)用の推奨される製造者のプロトコールの改変バージョンを使用してライブラリーを創出した。抽出されたccfDNA(例えば、約40 μ L)をライブラリー調製用の鋳型として使用した。全てのライブラリーを、液体ハンドリング装置使用(liquid handler instrumentation)(Caliper Zephyr; Caliper Life Sciences)を使用し、末端修復、ライゲーション、およびPCR生化学的プロセスの後に磁気ビーズに基づく(Beckman Coulter)浄化ステップを伴う半自動化プロセスを用いて創出した。ccfDNAは、母体の血漿中に小さな範囲内の断片サイズで存在することがよく特徴付けられているので、抽出されたccfDNAまたは調製されたライブラリーのいずれに対してもサイズ選択は実施しなかった。各ライブラリーのサイズ分布および量を、キャピラリー電気泳動(Caliper LabChip GX; Caliper)を用いて測定し、各ライブラリーを約2 nMの標準濃度に対して正規化した後に、CBot計器(Illumina)を使用してクラスタリングした。各試料を、HiSeq 2000 v3フローセル(Illumina)の2つのレーンを使用した合成による配列決定36サイクルに供した。

20

30

【0372】

データ解析

配列決定データの解析を、これによりその全体が参照により組み込まれるPalomakira(Genet Med.(2011年)13巻(11号):913~20頁およびGenetics in Medicine(2012年)14巻:296~305頁)に記載の通り実施した。簡単に述べると、HiSeq 2000計器からの全てのアウトプットファイル(例えば、.bcl files)をfastq形式に変換し、2009年2月に築かれたヒトゲノム(hg19)に対してCASAVA v1.7(Illumina)を使用してアラインメントした。その後の算出に対する反復配列の影響を最小限にするために、アラインメントした後、Repeat Library 20090604(Universal Resource Locator(URL)ワールドワイドウェブrepeatmasker.org)に含有される情報に基づいてゲノムの反復性の領域とオーバーラップしている読み取りを全て除去した。解析のために、各染色体を別個の50 kbのビンに分け、これらのビンのそれぞれにマッピングされた読み取りの数を合計した。その後の算出に対するG/C含量の偏りの影響を最小限にするために、各ビン内の読み取りを当技術分野で公知のLOESS法を使用してビンに特異的なG/C含量に関して正規化した。次いで、ビンによる反復マスキングされ、G/Cについて正規化された読み取りカウントを、統計的有意性およびカバレッジを算出するために使用する。

40

【0373】

目的の領域にマッピングされたアラインメントされた常染色体の読み取りの総数の分率

50

について、アラインメントされた常染色体の読み取りの総数と比較して z スコアを算出することによって統計的有意性を決定した。ロバストな方法を使用して Z スコアを算出し、それにより、式 $Z_{\text{試料}} = (\text{分率}_{\text{試料}} - \text{分率}_{\text{中央値}_{\text{母集団}}}) / \text{中央絶対偏差}_{\text{母集団}}$ を使用することによって所与の試料についての z スコアを算出した。式、カバレッジ = $L N / G$ (式中、L は読み取り長 (36 bp) であり、N は反復マスキングされ、GC について正規化された読み取りの数である、G は反復マスキングされた一倍体ゲノムのサイズである) によってカバレッジを算出した。

【0374】

結果

妊婦 16 人の血漿から単離された ccf DNA に対して次世代シーケンシングを実施した。16 人のうちの 2 人については、羊水穿刺後の核型分析により、染色体 22q11.2 欠失症候群の影響を受けた胎児を有することが確認された。14 の対照試料の胎児の核型情報は入手不可能であった。影響を受けた試料 2 つから、対照試料と比較して同様の妊娠期間 (19 週および 20 週) に血漿を採取した (中央値 = 20 週間; 下の表 6 参照)。配列決定の前に、総 ccf DNA に対する胎児の寄与を当技術分野で公知の通り測定した。全ての試料が 10% 超の胎児 DNA を含有し、寄与の中央値は 18% であり、胎児の微小欠失を有する 2 つの試料は 17% および 18% の胎児 DNA を含有した (下の表 6 参照)。

【0375】

【表 6】

表 6

試料ID	試料群	血漿の 体積 (mL)	胎児分 率	妊娠期間 (週)	GC正規化 された読み 取りの総数	ゲノムの カバレッ ジ	影響を受 けた領域 におけるG C正規化さ れた分率
12800	低リスク	4	0.42	19	202890379	4.43	0.000755
12801	微小欠失	4	0.17	20	188214827	4.11	0.000732
12802	低リスク	3.9	0.24	24	164976211	3.60	0.000752
12803	低リスク	4	0.13	12	190397481	4.16	0.000753
12804	低リスク	4	0.16	24	175708269	3.84	0.000747
12805	低リスク	4	0.35	17	192035852	4.19	0.000755
12806	低リスク	3.9	0.13	12	189438328	4.14	0.000757
12807	低リスク	3.9	0.18	20	185562643	4.05	0.000755
12808	微小欠失	4	0.18	19	146700048	3.20	0.000726
12809	低リスク	4	0.54	21	154878242	3.38	0.000750
12810	低リスク	4	0.15	16	188121991	4.11	0.000768
12811	低リスク	4	0.16	24	172366695	3.76	0.000757
12812	低リスク	4	0.10	12	180005977	3.93	0.000751
12813	低リスク	4	0.23	25	151510852	3.31	0.000752
12814	低リスク	4	0.20	20	143687629	3.14	0.000752
12815	低リスク	3.9	0.18	12	177482109	3.88	0.000754

各試料について、HiSeq 2000 フローセルの 2 つのレーンを使用して配列決定し、約 3.1x から約 4.4x の間のゲノムのカバレッジがもたらされた (上の表 6 参照)。50 kb のピンサイズを使用して読み取りをピンに入れ、影響を受けた微小欠失試料について第 22 染色体にわたってピンを可視化して、影響を受けた試料について微小欠失の位置を同定した。確認された 22q11.2 微小欠失を保有する試料はどちらも、このゲノムの領域の表示の減少を示した (図 47 を参照されたい)。各試料について、第 22 染色体上の影響を受けた領域について、全ての試料の中央値と比較して Z スコアを算出した

。リスクが低い女性由来の血漿に対応する値は黒色で示されており、既知の22q11.2欠失症候群の症例を示す値は灰色で示されている。-3における破線は、分析された試料全てにわたるこの領域についての中央値表示よりも低い中央絶対偏差の3倍であるzスコアを表し、胎児の異数性の検出において伝統的に使用される分類カットオフである。

【0376】

ゲノムの欠失の正確な位置は、症例ごとにわずかに変動する可能性があるため、Chr 22:19000000から22000000の間に位置する3百万塩基対の領域を検査することを選択した(上の表6参照)。染色体異数性の検出に使用される方法と同様の方法を使用して、標的領域にマッピングされた全ての常染色体の読み取りの分率を算出した。対照試料は、22q11に位置する読み取りの0.075%を含有したが、既知の胎児の微小欠失を伴う影響を受けた試料では、この領域における読み取りの0.073%を示しただけであった。この差異の統計的有意性を検定するために、各試料について、ロバストな方法を使用してzスコアを算出した。影響を受けた試料はどちらも-3未満のzスコアを示したが(例えば、それぞれ-5.4および-7.1)、全ての低リスク対照試料のzスコアが-3よりも高かった(図47を参照されたい)。低リスク試料のうちの1つは+3よりも高いzスコアを示した。22q11のゲノム領域は、以前にゲノムの不安定性に関連付けられており、この結果により、起こることが以前報告された潜在的な重複が示される可能性があるが、低リスク試料については核型情報が入手不可能であったので、観察された結果が胎児のCNVと関連付けられるかどうかは不明のままである。

【0377】

考察

非侵襲的な出生前診断の分野における最近の進歩により、母体の血漿中に存在するccfDNAについて配列決定することによって胎児の異数性を検出することが可能になった。異数性を検出するために使用するものと同様の手法を使用して、本明細書で示されている結果により、発達中の胎児における染色体領域内レベルのCNVを、母体の血漿中の対応するccfDNAについて配列決定することによって非侵襲的に検出することの実現性が確認される。少数の症例ではあるが、本明細書で示されているデータにより、単一の染色体よりも小さな領域、この場合は22q11.2の欠失を母体の血漿から確実に検出することができるが示されている。Petersら(2011年)により、同様の方法体系を使用して検出された第12染色体の4.2Mbの欠失が報告された。Petersらは、妊娠期間の後期(35週)に検出された胎児の微小欠失の単一の症例を検査し、それを、第12染色体および第14染色体について二倍体であることが分かっている7つの試料と比較した。対照的に、上記の試験が公開される前に得られた、より早い妊娠期間(19週および20週)において影響を受けた試料を検査したものである本明細書で示されている結果では、利用した影響を受けた試料および影響を受けていない試料の数が2倍であり、上記よりも28%小さな(3Mb)微小欠失が検出された。さらに、本明細書で示されている結果では、4xゲノムのカバレッジを利用して3Mbの胎児の欠失が首尾よく検出されており、これは、現行の標準の異数性検出に対しておよそ20倍のカバレッジの増加である。潜在的に0.5Mbに至るまでのより小さな欠失、または含有する胎児のccfDNAが少ない試料には、さらに高いカバレッジが必要になり得る。

【0378】

(実施例4)

ライブラリー調製の自動化、多重化レベルの増加およびバイオインフォマティクス

臨床的な正確度を維持しながら、処理量を3倍に増加させ、実践時間を4分の1に低下させるプロセスの変化のセットの実行が以下に提供される。この改変アッセイの3つの主要な変化は、多重化レベルがより高くなったこと(4プレックスから12プレックスへ)、配列決定ライブラリー調製が自動化されたこと、および新規のバイオインフォマティクス方法を実行することを含む。結果により、このプロトコールにより、21トリソミー、18トリソミーおよび13トリソミーを検出するための高い感度および特異度を維持しながら、より多くの処理量に適したより簡易化されたワークフローがもたらされることが確

認される。

【0379】

材料および方法

試料の獲得および血液の処理

3つの別々の治験審査委員会 (Investigational Review Board) (IRB) に認可された臨床試験実施計画書 (BioMed IRB 301-01、Western IRB 20091396、およびCompass IRB 00462) の下で、ハイスループットなアッセイの最初の評価 (ライブラリー調製の開発およびアッセイの検証) のための試料を採取した。全ての被験体は、EDTA-K2 噴霧乾燥 10 mL バキュテナー (EDTA チューブ; Becton Dickinson、Franklin Lakes、NJ) 中に最大 20 mL の全血、および Cell-Free DNA BCT 10 mL バキュテナー (BCT チューブ; Streck、Omaha、NE) 中に 30 mL の全血を採取するための静脈穿刺を含めた試験に関連するいずれの手順も、書面のインフォームドコンセントをもたらされた後に受けた。EDTA チューブ中に採取された試料を、冷蔵、または湿った氷上で保管し、採血の 6 時間以内に血漿を処理した。BCT チューブ中に採取された試料を外界温度で保管し、採血の 72 時間以内に血漿を処理した。EDTA チューブ中の母体の全血を遠心分離し (Eppendorf 5810R プラススイングローター)、2500 g で 10 分冷却し (4 °C)、血漿を採取した。EDTA 血漿について、4 °C、15、500 g で 10 分、2 回目の遠心分離を行った (Eppendorf 5810R プラス固定角ローター)。2 回目の回転後、EDTA 血漿をチューブの底に形成されたペレットから取り出し、4 mL のバーコード付けした血漿一定分量に分配し、すぐに -70 °C 以下で凍結して DNA 抽出まで保管した。BCT チューブ中の母体の全血を遠心分離し (Eppendorf 5810R プラススイングローター)、1600 g で 15 分温め (25 °C)、血漿を採取した。BCT 血漿について、25 °C、2、500 g で 10 分、2 回目の遠心分離を行った (Eppendorf 5810R プラススイングローター)。2 回目の回転後、BCT 血漿をチューブの底に形成されたペレットから取り出し、4 mL のバーコード付けした血漿一定分量に分配し、すぐに凍結して -70 °C 以下で DNA 抽出まで保管した。

【0380】

多重化展開および臨床評価のための試料を以前に記載されている通り採取した (Palomaki GE ら (2012 年) Genet. Med. 14 巻: 296 ~ 305 頁、および Palomaki GE ら (2011 年))。簡単に述べると、登録された患者から、侵襲的手段の前に全血を採取した。全ての試料を、妊娠第 1 三半期または妊娠第 2 三半期において胎児の異数性のリスクが増加している妊婦から、国際共同研究 (ClinicalTrials.gov NCT00877292) の一部として採取した。27 か所の採取場所のそれぞれにおいて、この共同研究について IRB による認可 (または同等のもの) を得た。本発明で 4 プレックス形式で生成し、使用したいいくつかのデータは、本明細書で以前に示されているが、12 プレックス配列決定からの全てのデータは、それぞれ独立に 12 プレックス形式で今回配列決定した同じライブラリーを使用して生成した。さらに、ハイスループットな方法を独立して確認するために、患者 1269 人のそれぞれ由来の血漿一定分量を処理した。これらの患者のそれぞれが、以前に公開された試験に対する別個の血漿一定分量に寄与し、胎児の核型は既知であった。単純な 21 トリソミー、18 トリソミー、および 13 トリソミーであることが確認された単胎児妊娠由来の試料または正倍数性の対照由来の試料のみを使用した。循環している無細胞 DNA を、QIAamp Circulating Nucleic Acid Kit (Qiagen) を本明細書に記載の通り使用して母体の血漿から抽出した。各試料について、胎児定量器アッセイ (FQA) によって ccf DNA の量を評価した。抽出された ccf DNA (40 μL) を全てのライブラリー調製用の鋳型として使用した。最初に増加させた (12 プレックス) マルチプレックス実験についてのライブラリーを、上記の方法を使用して調製した。簡単に述べると、ccf DNA を抽出し、オリゴヌクレオチド (Illum

10

20

30

40

50

ina)、酵素(Enzymatics)、および各酵素反応の間にカラムに基づく方法(Qiagen)を使用した手動の精製プロセスを使用して配列決定ライブラリーを調製した。この試験において使用した、新しく創出されたライブラリーは、TruSeqライブラリー調製(Illumina)についての製造者のプロトコルの改変バージョンならびに液体ハンドリング装置使用(Caliper Zephyr; Caliper Life Sciences)を利用し、末端修復、ライゲーション、およびPCR生化学的プロセス後に磁気ビーズに基づく(AMPure XP; Beckman Coulter)浄化ステップを伴う半自動化プロセスを使用して、96ウェルプレート形式で創出した。ccf DNAは、母体の血漿中に小さな範囲内の断片サイズで存在することがよく特徴付けられているので、抽出されたccf DNAまたは調製されたライブラリーのい
10
ずれにおいてもサイズ選択は実施しなかった。ライブラリーのサイズ分布および定量の評価を、本明細書において以前に記載されている通り実施した。12イソモル(isomolar)の配列決定ライブラリーをプールし、Illumina HiSeq 2000においてIllumina v3つのフローセルの同じレーンと一緒に配列決定した(12プレックス)。合成による配列決定を36サイクル実施し、その後、7サイクルを実施して、各試料指標を読み取った。プールされた、21トリソミーと診断された成人男性ボランティア2人または妊娠していない正倍数性の女性2人の血漿から単離されたccf DNAから配列決定ライブラリーを調製した。ライブラリーを定量し、母体の血漿中のccf 胎児DNAのおおよその寄与に対して2つの濃度(21トリソミー4%および21トリソミー13%)で混合した。これらの対照を臨床評価試験に組み込む前にライブラリー性能を検査した。
20

【0381】

データ解析

HiSeq 2000からのBCL(塩基呼び出し)アウトプットファイルの全てを、FASTQ形式に変換し、2009年2月に築かれたヒトゲノム(hg19)に対してアラインメントした。マルチプレックス展開のためのライブラリーを以前のバージョンの生化学を用いて手動で調製したので、分析方法を以前に記載されている通り適用した(Palomakira、2012年および本明細書)。その後の試験の全てについて、Bowtie 2(Langmead B、Salzberg SL(2012年)Nat. Methods 9巻:357~359頁)を使用して読み取りをhg19に対してアラインメントし、シード配列内の完全な一致のみを可能にした。解析のために、近接する、オーバーラップしていない50kbp長のゲノムのセグメントを含む標準のヒストグラムを使用して各染色体にマッピングされた読み取りを定量した。ビンニング(binning)後、含まれた50kbpのゲノムのセグメントの選択を、以前に記載されている交差検証方法(Brunger AT(1992年)Nature 355巻:472~475頁)を使用して決定した。高い試料間の分散、低いマッピング可能性(Derrien Tら(2012年)PLoS one 7:e30377頁)、または高い反復エレメントの百分率(Repeat Library 20090604; <http://www.repeatmasker.org>)を示すことに基づいて領域をさらなる解析から排除した。最後に、残りの50kbpのゲノムのセグメントに対応するアラインメントされた読み取りを正規化してGCの偏りを考慮に入れ(Alkan Cら(2009年)Nat Genet 41巻:1061~1067頁)、各染色体に由来するアラインメントされた読み取りの分率を算出するために使用した。ロバストなzスコアを、記載の通り、式、Z染色体=(染色体分率試料-染色体分率フローセル中央値)/絶対偏差中央値を使用して算出した。染色体分率中央値は、各フローセルに特異的に算出され、中央絶対偏差(MAD)は静的MADに由来する定数値であった。
30
40

【0382】

結果

非侵襲的に胎児の異数性を検出するためのMPSSを使用したいくつかの臨床試験により、92~100%の範囲の検出率が示され、同時に偽陽性率が1%未満に維持された。
50

我々の目標は、プロトコルを合理化し、試料の処理量を増加させながら、この性能を維持または改善することであった。改善は、3つの態様に焦点を合わせた：I) ロバストな収率が可能になり、処理量が増加するようにライブラリー調製を最適化すること、II) 単一のフローセルレーンと一緒にプールされる個別に分子的に指標された試料の数を増加させること（マルチプレックスレベル）、およびIII) 異数性を分類するための分析的方法を改善すること。

【0383】

従来の配列決定ライブラリー調製は、労働集約的であり、時間がかかり、また、オペレーター間で変動しやすい。これらの問題を緩和するために、96チャンネル液体取扱いプラットフォームを利用する半自動化プロセスを開発した。TruSeqライブラリー調製生化学を、TruSeqライブラリー調製キットに対して推奨されるインプット量である1 μ gの50分の1である、存在量が少ないccf DNA (10~20 ng) を血漿4 mLから回収するために最適化した。さらに、手動の精製手順を、スピード、再現性およびccf DNA回収について最適化された自動化AMPure XPビーズ精製プロセスと交換した。この方法を使用して調製した287ライブラリーのセットと、記載されている（本明細書およびPalomakiら2011年およびPalomakiら2012年）手動の方法を使用して作製したライブラリーを比較することにより、溶出液の体積について標準化した後、ライブラリー濃度中央値が124 nMから225 nMに増加することが明らかになった（図11A）。組み合わせた半自動化プロセスにより、5時間で96ライブラリーが作製され、単一の技師および1.5時間の実践労働時間のみが必要である。これにより、ライブラリー収率または品質を犠牲にすることなく処理量が4倍に増加したと同時に、労力が4分の1に減少した。93のライブラリー（83の確認された正倍数性試料および10の確認された21トリソミー試料；表7）を、この方法を使用して調製し、配列決定し、分析し、この小さなデータセットにおいて正確な分類性能が実証された（図11B；表8）。

【0384】

以前の試験の間に4プレックスで調製し、配列決定したライブラリーについて、12プレックスで配列決定して、多重化の増加の実現性を決定した。Illumina v3つのフローセルおよび配列決定生化学をHCSソフトウェアの改善と組み合わせることにより、1レーン当たりの読み取りカウントの総数が2.23倍増加した（7,200万から1億6,100万）。1629の正倍数性試料、205の21トリソミー試料、54の18トリソミー試料、および12の13トリソミー試料（表7）を含めた1900ライブラリーを12プレックスで配列決定し、分析し、第21染色体、第18染色体、および第13染色体についてのzスコアを4プレックスの結果と比較した（図12）。以前の試験により、zスコアカットオフの上昇を使用したアッセイの性能の増加が示されているので、第18染色体および第13染色体染色体について分類は $z = 3.95$ に基づいた。第21染色体についての分類は、 $z = 3$ のままであった。これらの分類カットオフを使用して、4プレックス配列決定と12プレックス配列決定の間で合計7つの分類結果が一致しなかった。第21染色体について、以前誤って分類された（1つが偽陽性、1つが偽陰性）2つの試料は正確に分類されたが、以前示された真陽性は検出されなかった。第18染色体について4つの試料が偽陽性試料と誤って分類されたが、これらは、以前は正確に分類されていた；これらのライブラリーのそれぞれはGCの偏りが高度であった。13トリソミーの分類については全ての試料が合致した。12プレックスで配列決定した場合、異数性試料の99.3%（204/205の21トリソミー、54/54の18トリソミー、および11/12の13トリソミー）が、21トリソミー、18トリソミー、および13トリソミーについて、それぞれ0%（0/1900）、0.26%（5/1900）、および0.16%（3/1900）の偽陽性率で検出された（表8）。全体的に、これらのデータにより、12プレックス多重化で遂行した場合のアッセイの性能が以前得られた結果と同様であることが示唆される。

【0385】

最適化されたライブラリー調製方法を12プレックス配列決定（ハイスループットなアッセイ設定）と組み合わせ使用して検証試験を実施して、プロセスの完全性を確実にした。合計2856試料からの配列決定の結果のうち、核型が分かっている1269を解析した。これらの1269の臨床的な試料は、1093の正倍数性試料、134の21トリソミー試料、36の18トリソミー試料、および6の13トリソミー試料で構成された（表7）。試料についての胎児DNA分率中央値は0.14（範囲：0.04～0.46）であった。ライブラリーのライブラリー濃度中央値は28.21 nM（範囲：7.53～42.19 nM）であり、これにより、本明細書に記載の他の方法と同様の総収率がもたらされた。最後に、試料1つ当たりのアラインメントされた常染色体の読み取りの数の中央値は16, 291, 390（範囲：8, 825, 886～35, 259, 563）であ

10

【0386】

胎児の核型が分かっている1269の試料から生成されたデータと、以前配列決定された同じ被験体由来の別個の血漿一定分量の最初の比較により、GC含量について正規化し、反復領域とオーバーラップしている読み取りを除去する、以前確立された方法（例えば、GCRM）を用いて分析した場合、差別的な距離（正倍数性試料の95パーセンタイルと21トリソミー試料の5パーセンタイルとの間の差異）が4.9から3.09に減少することが明らかになった。この影響を減らし、同時に全体的な分析時間を減少させるために、ハイスループットなアッセイデータに特異的な新規のバイオインフォマティクスのアルゴリズムを開発した。これらの方法は、個体にわたって安定な表示を有する50 kbpのゲノムのセグメントのみを分類するための算出に基づく。同じハイスループットなデータセットに適用する場合、正倍数性試料と21トリソミー試料との間の差別的な距離は6.49に増加する。全体的に、新規のバイオインフォマティクス手法により、以前に記載された方法と比較して、正倍数性試料と21トリソミー試料との間の差別的な距離が増加する。

20

【0387】

67の対照試料および1269の患者試料についてのハイスループットなアッセイからの結果を、新規の解析方法を使用して解析した。プールされた正倍数性血漿（0% T21ライブラリー）から調製した33のライブラリー、4%の21トリソミーDNAを含有する17の対照ライブラリー、および13%の21トリソミーDNAを含有する17の対照ライブラリーについて配列決定した。全ての場合において、プールされた正倍数性試料のzスコアは3未満であったが、4%および13% 21トリソミー対照試料のzスコアは、3を超えた。次いで、核型情報が分かっている1269の患者試料の分類の正確度を比較した。上記の分類限界に基づいて（第21染色体についてはzスコア=3、第18染色体および第13染色体についてはzスコア=3.95）、確認された胎児の異数性（134の21トリソミー、36の18トリソミー、6の13トリソミー）の全てが、21トリソミー、18トリソミー、および13トリソミーについて、それぞれ0.08%、0%、および0.08%の偽陽性率で検出された（図13；表8）。胎児分率とzスコアの大きさの間に正の相関があったが、正倍数性試料についてはこれらの測定基準に相関はなかった。

30

40

【0388】

1269人のドナーのそれぞれ由来の別個の血漿試料が以前配列決定されており、したがって、性能についての比較として機能する。同等の評価を確実にするために、以前の試験からのzスコアを、中央値を算出するために使用した試料数と等しい96の試料のGCRM値および母集団のサイズを使用して、ハイスループットな分析を使用して算出した（中央値およびMAD算出について）。2つの試験を比較することにより、以前報告された偽陰性の21トリソミー試料および以前報告された偽陽性の21トリソミー試料の正しい分類が明らかになったが、この試験の間に1つの追加的な偽陽性が存在した（図14）。13トリソミー分類および単一の18トリソミー試料についての正しい分類を以前のわずかに3.95を下回るzスコアと比較した場合に、一致しない試料はなかった。対応のあ

50

る z スコアを異数性試料について評価することにより、平均差異が、21トリソミーについては 2.19 であり、18トリソミーについては 1.56 であり、13トリソミーについては 1.64 であることが明らかになり、これには、ハイスループットな方法を使用すると、影響を受けた試料についての z スコアが増加したことが反映されている。ハイスループットなアッセイを使用すると、確認された 21トリソミー試料および 18トリソミー試料について、以前の試験と比較して z スコアが統計的に有意に増加したが（それぞれ $p = 4.24 \times 10^{-12}$ および $p = 0.0002$ ；対応のあるウィルコクソン検定）、確認された 13トリソミー試料については z スコアに有意差はなかった（ $p = 0.31$ ；対応のあるウィルコクソン検定）。非異数性試料について、第 21 染色体、第 18 染色体、または第 13 染色体の z スコアに統計的有意差はなかった（ $p = 0.06$ 、 $p = 0.90$ 、 $p = 0.82$ 、それぞれ；対応のあるウィルコクソン検定）。この正倍数性試料に有意に影響を及ぼすことなく異数性の z スコアが有意に増加することにより、ハイスループットなアッセイ設定および新規のバイオインフォマティクス方法を使用した場合に、第 21 染色体および第 18 染色体について、正倍数性試料と異数性試料との間の分析的な距離が伸びたことがさらに示される。

【0389】

考察

本明細書に示されている開発には研究活動が先行し、その後、CLIA によって認定された研究所において追加的な検証および検証試験が行われた。全体で、新しい研究所検査をもたらす研究から検証までのプロセス全体は、5000 を超える試験試料によって支持される。この試験では、研究、最適化、開発の間に 3400 を超える試料について配列決定した。次いで、1269 の試料を利用して臨床評価試験を実施し、その中で、各トリソミーについて 0.08% 以下の偽陽性率を維持しながら 176 の異数性試料全てを検出した。

【0390】

ライブラリー調製の処理量を 4 倍増加させることを可能にするアッセイを開発し、それを試料多重化による 3 倍の増加と併せて、ハイスループットな ccf DNA 試料処理を可能にした。これらの方法を改善された分析論と組み合わせると同時に、非侵襲的な異数性の検出の感度および特異度を改善し、技師および計器の必要性を減少させた。全体的に、これらのデータにより、開発されたハイスループットなアッセイが技術的に口

【0391】

【表 7】

表7.実施された試験のそれぞれに利用された試料の種類 요약

試験の説明	核型による試料の数				
	未知	正倍数性	21トリソミー	13トリソミー	18トリソミー
ライブラリー最適化	0	83	10	0	0
12プレックス配列決定	0	1629	205	12	54
検証	1587	1093	134	6	36

【0392】

10

20

30

40

【表 8】

表8.実施された試験のそれぞれについての分析結果の要約。Sens=感度;Spec=特異度;NA=該当なし

試験の説明	染色体ごとの分析結果						分析方法
	Spec Chr21	Sens Chr21	Spec Chr13	Sens Chr13	Spec Chr 18	Sens Chr18	
ライブラリー最適化	100	100	NA	NA	NA	NA	GCRM
12プレックス配列決定	100	99.5	99.84	91.7	99.74	100	GCRM
検証	99.92	100	99.92	100	100	100	新規

実施例 5 : 実施形態の例

A 1 . 胎児の異数性の有無を検出するための方法であって、

(a) 妊婦由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、

(b) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと

(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e) 正規化された試料カウントに基づいて、胎児の異数性の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【 0 3 9 3 】

A 2 . 胎児の異数性の有無を検出するための方法であって、

(a) 妊婦由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(g) 正規化された試料カウントに基づいて、胎児の異数性の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【 0 3 9 4 】

A 3 . 胎児の異数性の有無を検出するための方法であって、

(a) 妊婦由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

10

20

30

40

50

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 正規化された試料カウントに基づいて、胎児の異数性の有無を決定するアウトカムをもたらすステップとを含む、方法。

【0395】

A3.1. 胎児の異数性の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、妊婦由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(c) 正規化された試料カウントに基づいて胎児の異数性の有無を検出するステップとを含む、方法。

【0396】

A4. 試料核酸が妊婦由来の血漿由来のものである、実施形態A1からA3.1のいずれか1つに記載の方法。

【0397】

A5. 試料核酸が妊婦由来の血清由来のものである、実施形態A1からA3.1のいずれか1つに記載の方法。

【0398】

A6. 胎児の異数性が13トリソミーである、実施形態A1からA3.1のいずれか1つに記載の方法。

【0399】

A7. 胎児の異数性が18トリソミーである、実施形態A1からA3.1のいずれか1つに記載の方法。

【0400】

A8. 胎児の異数性が21トリソミーである、実施形態A1からA3.1のいずれか1つに記載の方法。

【0401】

A9. 無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、実施形態A1からA3.1のいずれか1つに記載の方法。

【0402】

A10. ポリヌクレオチド断片の長さが約20ヌクレオチドから約50ヌクレオチドの間である、実施形態A9に記載の方法。

【0403】

A11. ポリヌクレオチドの長さが約30ヌクレオチドから約40ヌクレオチドの間である、実施形態A10に記載の方法。

【0404】

A12. 予測カウントがカウント中央値である、実施形態A1からA11のいずれか1つに記載の方法。

【0405】

A13. 予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、実施形態A1からA11のいずれか1つに記載の方法。

【0406】

A14. カウントを、GC含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、実施形態A1からA13のい

10

20

30

40

50

ずれか1つに記載の方法。

【0407】

A15. カウントを、正規化モジュールによって正規化する、実施形態A1からA14のいずれか1つに記載の方法。

【0408】

A16. 核酸配列読み取りを、配列決定モジュールによって生成する、実施形態A1からA15のいずれか1つに記載の方法。

【0409】

A17. 核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム全体にマッピングするステップを含む、実施形態A1からA16のいずれか1つに記載の方法。

10

【0410】

A18. 核酸配列読み取りをマッピングモジュールによってマッピングする、実施形態A17に記載の方法。

【0411】

A19. 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをカウントモジュールによってカウントする、実施形態A1からA18のいずれか1つに記載の方法。

【0412】

A20. 配列読み取りを配列決定モジュールからマッピングモジュールに移行する、実施形態A18またはA19に記載の方法。

20

【0413】

A21. 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをマッピングモジュールからカウントモジュールに移行する、実施形態A19またはA20に記載の方法。

【0414】

A22. 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りのカウントを、カウントモジュールから正規化モジュールに移行する、実施形態A19からA21のいずれか1つに記載の方法。

【0415】

A23. カウントを正規化するステップが、パーセント表示を決定するステップを含む、実施形態A1からA22のいずれか1つに記載の方法。

30

【0416】

A24. 正規化されたカウントがzスコアである、実施形態A1からA23のいずれか1つに記載の方法。

【0417】

A25. 正規化されたカウントがロバストなzスコアである、実施形態A1からA24のいずれか1つに記載の方法。

【0418】

A26. 第1のゲノミックセクションについてのカウントの誘導値が第1のゲノミックセクションのパーセント表示である、実施形態A1からA25のいずれか1つに記載の方法。

40

【0419】

A27. 中央値がパーセント表示の中央値である、実施形態A12からA26のいずれか1つに記載の方法。

【0420】

A28. パーセント表示が染色体表示である、実施形態A23からA27のいずれか1つに記載の方法。

【0421】

B1. 遺伝的変異の有無を検出するための方法であって、

50

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、

(b) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e) 正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

【 0 4 2 2 】

B 2 . 遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(g) 正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

【 0 4 2 3 】

B 3 . 遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

【 0 4 2 4 】

B 3 . 1 . 遺伝的変異の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、試験被験体由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセ

10

20

30

40

50

クションについて正規化し、それにより、正規化された試料カウントを得るステップと、
 (d) 正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を
 決定するアウトカムをもたらすステップと
 を含む、方法。

【0425】

B4. 試料核酸が試験被験体由来の血漿由来のものである、実施形態B1からB3.1のいずれか1つに記載の方法。

【0426】

B5. 試料核酸が試験被験体由来の血清由来のものである、実施形態B1からB3.1のいずれか1つに記載の方法。

10

【0427】

B6. 遺伝的変異が医学的状態に関連付けられる、実施形態B1からB5のいずれか1つに記載の方法。

【0428】

B7. 医学的状態ががんである、実施形態B6に記載の方法。

【0429】

B8. 医学的状態が異数性である、実施形態B6に記載の方法。

【0430】

B9. 試験被験体がヒト、動物、および植物から選択される、実施形態B1からB5のいずれか1つに記載の方法。

20

【0431】

B10. ヒト試験被験体が、女性、妊婦、男性、胎児、または新生児を含む、実施形態B9に記載の方法。

【0432】

B11. 無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、実施形態B1からB5のいずれか1つに記載の方法。

【0433】

B12. ポリヌクレオチド断片の長さが約20ヌクレオチドから約50ヌクレオチドの間である、実施形態B11に記載の方法。

【0434】

B13. ポリヌクレオチドの長さが約30ヌクレオチドから約40ヌクレオチドの間である、実施形態B12に記載の方法。

30

【0435】

B14. 予測カウントがカウント中央値である、実施形態B1からB13のいずれか1つに記載の方法。

【0436】

B15. 予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、実施形態B1からB13のいずれか1つに記載の方法。

【0437】

B14. カウントを、GC含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、実施形態B1からB13のいずれか1つに記載の方法。

40

【0438】

B15. カウントを、正規化モジュールによって正規化する、実施形態B1からB14のいずれか1つに記載の方法。

【0439】

B16. 核酸配列読み取りを、配列決定モジュールによって生成する、実施形態B1からB15のいずれか1つに記載の方法。

【0440】

B17. 核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム

50

全体にマッピングするステップを含む、実施形態 B 1 から B 1 6 のいずれか 1 つに記載の方法。

【 0 4 4 1 】

B 1 8 . 核酸配列読み取りをマッピングモジュールによってマッピングする、実施形態 B 1 7 に記載の方法。

【 0 4 4 2 】

B 1 9 . 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをカウントモジュールによってカウントする、実施形態 B 1 から B 1 8 のいずれか 1 つに記載の方法。

【 0 4 4 3 】

B 2 0 . 配列読み取りを配列決定モジュールからマッピングモジュールに移行する、実施形態 B 1 8 または B 1 9 に記載の方法。

【 0 4 4 4 】

B 2 1 . 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをマッピングモジュールからカウントモジュールに移行する、実施形態 B 1 9 または B 2 0 に記載の方法。

【 0 4 4 5 】

B 2 2 . 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りのカウントを、カウントモジュールから正規化モジュールに移行する、実施形態 B 1 9 から B 2 1 のいずれか 1 つに記載の方法。

【 0 4 4 6 】

B 2 3 . カウントを正規化するステップが、パーセント表示を決定するステップを含む、実施形態 B 1 から B 2 2 のいずれか 1 つに記載の方法。

【 0 4 4 7 】

B 2 4 . 正規化されたカウントが z スコアである、実施形態 B 1 から B 2 3 のいずれか 1 つに記載の方法。

【 0 4 4 8 】

B 2 5 . 正規化されたカウントがロバストな z スコアである、実施形態 B 1 から B 2 4 のいずれか 1 つに記載の方法。

【 0 4 4 9 】

B 2 6 . 第 1 のゲノミックセクションについてのカウントの誘導値が第 1 のゲノミックセクションのパーセント表示である、実施形態 B 1 から B 2 5 のいずれか 1 つに記載の方法。

【 0 4 5 0 】

B 2 7 . 中央値がパーセント表示の中央値である、実施形態 B 1 2 から B 2 6 のいずれか 1 つに記載の方法。

【 0 4 5 1 】

B 2 8 . パーセント表示が染色体表示である、実施形態 B 2 3 から B 2 7 のいずれか 1 つに記載の方法。

【 0 4 5 2 】

C 1 . 遺伝的変異の有無を検出するための方法であって、
 (a) 試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、
 (b) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、
 (c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、
 (d) (c) のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

10

20

30

40

50

(e) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(d)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(f) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(g) (f)における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【0453】

C2. 遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) (e)のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(g) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(f)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(h) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(i) (h)における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【0454】

C3. 遺伝的変異の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) (b)のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(c)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異

10

20

30

40

50

の統計的有意性を評価するステップと、

(f)(e)における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【0455】

C3.1 遺伝的変異の有無を検出するための方法であって、

(a)参照ゲノムセクションにマッピングされた、試験被験体由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(b)(a)のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(c)1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(b)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d)1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(e)(d)における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【0456】

C4. 補正され、カウントされ、マッピングされた配列読み取りを、1つまたは複数の実験条件についてさらに補正した後に、残ったカウントを正規化する、実施形態C1からC3.1のいずれか1つに記載の方法。

【0457】

C5. 遺伝的変異が微小欠失である、実施形態C1からC4のいずれか1つに記載の方法。

【0458】

C6. 微小欠失が第22染色体上にある、実施形態C5に記載の方法。

【0459】

C7. 微小欠失が第22染色体の領域22q11.2において起こっている、実施形態C6に記載の方法。

【0460】

C8. 微小欠失が、参照ゲノムhg19による第22染色体のヌクレオチド19,000,000位と22,000,000位の間で起こっている、実施形態C6に記載の方法。

【0461】

C9. 正規化されたカウントの誘導値がZスコアである、実施形態C1からC8のいずれか1つに記載の方法。

【0462】

C10. ZスコアがロバストなZスコアである、実施形態C9に記載の方法。

【0463】

C11. 試料核酸が試験被験体由来の血漿由来のものである、実施形態C1からC10のいずれか1つに記載の方法。

【0464】

C12. 試料核酸が試験被験体由来の血清由来のものである、実施形態C1からC10のいずれか1つに記載の方法。

10

20

30

40

50

- 【0465】
C13. 遺伝的変異が医学的状態に関連付けられる、実施形態C1からC12のいずれか1つに記載の方法。
- 【0466】
C14. 医学的状態ががんである、実施形態C13に記載の方法。
- 【0467】
C15. 医学的状態が異数性である、実施形態C13に記載の方法。
- 【0468】
C16. 試験被験体がヒト、動物、および植物から選択される、実施形態C1からC12のいずれか1つに記載の方法。 10
- 【0469】
C17. ヒト試験被験体が、女性、妊婦、男性、胎児、または新生児を含む、実施形態C16に記載の方法。
- 【0470】
C18. 無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、実施形態C1からC12のいずれか1つに記載の方法。
- 【0471】
C19. ポリヌクレオチド断片の長さが約20ヌクレオチドから約50ヌクレオチドの間である、実施形態C18に記載の方法。
- 【0472】 20
C20. ポリヌクレオチドの長さが約30ヌクレオチドから約40ヌクレオチドの間である、実施形態C19に記載の方法。
- 【0473】
C21. 予測カウントがカウント中央値である、実施形態C1からC20のいずれか1つに記載の方法。
- 【0474】
C22. 予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、実施形態C1からC20のいずれか1つに記載の方法。
- 【0475】 30
C23. カウントを、GC含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、実施形態C1からC22のいずれか1つに記載の方法。
- 【0476】
C24. カウントを、正規化モジュールによって正規化する、実施形態C1からC23のいずれか1つに記載の方法。
- 【0477】
C25. 核酸配列読み取りを、配列決定モジュールによって生成する、実施形態C1からC24のいずれか1つに記載の方法。
- 【0478】 40
C26. 核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム全体にマッピングするステップを含む、実施形態C1からC25のいずれか1つに記載の方法。
- 【0479】
C27. 核酸配列読み取りをマッピングモジュールによってマッピングする、実施形態C26に記載の方法。
- 【0480】
C28. 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをカウントモジュールによってカウントする、実施形態C1からC27のいずれか1つに記載の方法。
- 【0481】 50

C 2 9 . 配列読み取りを配列決定モジュールからマッピングモジュールに移行する、実施形態 C 2 7 または C 2 8 に記載の方法。

【 0 4 8 2 】

C 3 0 . 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをマッピングモジュールからカウントモジュールに移行する、実施形態 C 2 8 または C 2 9 に記載の方法。

【 0 4 8 3 】

C 3 1 . 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りのカウントを、カウントモジュールから正規化モジュールに移行する、実施形態 C 2 8 から C 3 0 のいずれか 1 つに記載の方法。

【 0 4 8 4 】

C 3 2 . カウントを正規化するステップが、パーセント表示を決定するステップを含む、実施形態 C 1 から C 3 1 のいずれか 1 つに記載の方法。

【 0 4 8 5 】

C 3 3 . 正規化されたカウントが z スコアである、実施形態 C 1 から C 3 2 のいずれか 1 つに記載の方法。

【 0 4 8 6 】

C 3 4 . 正規化されたカウントがロバストな z スコアである、実施形態 C 1 から C 3 3 のいずれか 1 つに記載の方法。

【 0 4 8 7 】

C 3 5 . 第 1 のゲノミックセクションについてのカウントの誘導値が第 1 のゲノミックセクションのパーセント表示である、実施形態 C 1 から C 3 4 のいずれか 1 つに記載の方法。

【 0 4 8 8 】

C 3 6 . 中央値がパーセント表示の中央値である、実施形態 C 2 1 から C 3 5 のいずれか 1 つに記載の方法。

【 0 4 8 9 】

C 3 7 . パーセント表示が染色体表示である、実施形態 C 3 2 から C 3 6 のいずれか 1 つに記載の方法。

【 0 4 9 0 】

D 1 . 微小欠失の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸からヌクレオチド配列読み取りを得るステップと、

(b) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d) (c) のカウントされ、マッピングされた配列読み取りを、反復配列および / または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(e) 1 つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(d) で残ったカウントを第 1 のゲノムセクションについて正規化、またはカウントの誘導値を第 1 のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(f) 1 つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(g) (f) における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

10

20

30

40

50

を含む、方法。

【0491】

D2. 微小欠失の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと

、
(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(f) (e) のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(g) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(f) で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(h) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(i) (h) における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【0492】

D3. 微小欠失の有無を検出するための方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料核酸から得たヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) (b) のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(d) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(c) で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(f) (e) における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む、方法。

【0493】

D3.1. 微小欠失の有無を検出するための方法であって、

(a) 参照ゲノムセクションにマッピングされた、試験被験体由来の循環している無細胞核酸を含む試料核酸から得られるヌクレオチド配列読み取りのカウントを得るステップと、

(b) (a) のカウントされ、マッピングされた配列読み取りを、反復配列および/ま

10

20

30

40

50

たは過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(b)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(e) (d)における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップとを含む、方法。

【0494】

D4. 補正され、カウントされ、マッピングされた配列読み取りを、1つまたは複数の実験条件についてさらに補正した後に、残ったカウントを正規化する、実施形態D1からD3.1のいずれか1つに記載の方法。

【0495】

D5. 微小欠失が第22染色体上にある、実施形態D4に記載の方法。

【0496】

D6. 微小欠失が第22染色体の領域22q11.2において起こっている、実施形態D5に記載の方法。

【0497】

D7. 微小欠失が、参照ゲノムhg19による第22染色体のヌクレオチド19,000,000位と22,000,000位の間で起こっている、実施形態D5に記載の方法。

【0498】

D8. 正規化されたカウントの誘導値がZスコアである、実施形態D1からD8のいずれか1つに記載の方法。

【0499】

D9. ZスコアがロバストなZスコアである、実施形態D8に記載の方法。

【0500】

D10. 試料核酸が試験被験体由来の血漿由来のものである、実施形態D1からD9のいずれか1つに記載の方法。

【0501】

D11. 試料核酸が試験被験体由来の血清由来のものである、実施形態D1からD9のいずれか1つに記載の方法。

【0502】

D12. 遺伝的変異が医学的状態に関連付けられる、実施形態D1からD11のいずれか1つに記載の方法。

【0503】

D13. 医学的状態ががんである、実施形態D12に記載の方法。

【0504】

D14. 医学的状態が異数性である、実施形態D12に記載の方法。

【0505】

D15. 試験被験体がヒト、動物、および植物から選択される、実施形態D1からD11のいずれか1つに記載の方法。

【0506】

D16. ヒト試験被験体が、女性、妊婦、男性、胎児、または新生児を含む、実施形態D15に記載の方法。

10

20

30

40

50

【0507】

D17. 無細胞試料核酸の配列読み取りがポリヌクレオチド断片の形態である、実施形態D1からD11のいずれか1つに記載の方法。

【0508】

D18. ポリヌクレオチド断片の長さが約20ヌクレオチドから約50ヌクレオチドの間である、実施形態D17に記載の方法。

【0509】

D19. ポリヌクレオチドの長さが約30ヌクレオチドから約40ヌクレオチドの間である、実施形態D18に記載の方法。

【0510】

D20. 予測カウントがカウント中央値である、実施形態D1からD19のいずれか1つに記載の方法。

【0511】

D21. 予測カウントが、トリムもしくは刈り込み平均、ウィンザー化平均またはブートストラップ推定値である、実施形態D1からD19のいずれか1つに記載の方法。

【0512】

D22. カウントを、GC含量、ピン様式での正規化、GC LOESS、PERUN、GCRM、またはそれらの組合せによって正規化する、実施形態D1からD21のいずれか1つに記載の方法。

【0513】

D23. カウントを、正規化モジュールによって正規化する、実施形態D1からD22のいずれか1つに記載の方法。

【0514】

D24. 核酸配列読み取りを、配列決定モジュールによって生成する、実施形態D1からD23のいずれか1つに記載の方法。

【0515】

D25. 核酸配列読み取りを、参照ゲノムのゲノミックセクションまたは参照ゲノム全体にマッピングするステップを含む、実施形態D1からD24のいずれか1つに記載の方法。

【0516】

D26. 核酸配列読み取りをマッピングモジュールによってマッピングする、実施形態D25に記載の方法。

【0517】

D27. 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをカウントモジュールによってカウントする、実施形態D1からD26のいずれか1つに記載の方法。

【0518】

D28. 配列読み取りを配列決定モジュールからマッピングモジュールに移行する、実施形態D26またはD27に記載の方法。

【0519】

D29. 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りをマッピングモジュールからカウントモジュールに移行する、実施形態D27またはD28に記載の方法。

【0520】

D30. 参照ゲノムのゲノミックセクションにマッピングされた核酸配列読み取りのカウントを、カウントモジュールから正規化モジュールに移行する、実施形態D27からD29のいずれか1つに記載の方法。

【0521】

D31. カウントを正規化するステップが、パーセント表示を決定するステップを含む、実施形態D1からD30のいずれか1つに記載の方法。

10

20

30

40

50

【0522】

D32. 正規化されたカウントが z スコアである、実施形態 D1 から D31 のいずれか 1 つに記載の方法。

【0523】

D33. 正規化されたカウントがロバストな z スコアである、実施形態 D1 から D32 のいずれか 1 つに記載の方法。

【0524】

D34. 第1のゲノミックセクションについてのカウントの誘導値が第1のゲノミックセクションのパーセント表示である、実施形態 D1 から D33 のいずれか 1 つに記載の方法。

10

【0525】

D35. 中央値がパーセント表示の中央値である、実施形態 D20 から D34 のいずれか 1 つに記載の方法。

【0526】

D36. パーセント表示が染色体表示である、実施形態 D31 から D35 のいずれか 1 つに記載の方法。

【0527】

E1. 正規化された試料カウントが、カウントの誘導値を第1のゲノムセクションについて正規化することを含むプロセスによって得られ、誘導値が第1のゲノムセクションについてのカウントを第1のゲノムセクションを含む複数のゲノムセクションについての

20

カウントで割ることによって決定される第1のゲノムセクションカウント表示である、実施形態 A1 から D21 のいずれか 1 つに記載の方法。

【0528】

E2. 第1のゲノムセクションについてのカウントの誘導値を予測カウントの誘導値に従って正規化し、予測カウントの誘導値が、第1のゲノムセクションについての予測カウントを第1のゲノムセクションを含む複数のゲノムセクションについての予測カウントで割ることによって決定される予測された第1のゲノムセクションカウント表示である、実施形態 E1 に記載の方法。

【0529】

E3. 第1のゲノムセクションが染色体または染色体の部分であり、複数のゲノムセクションが常染色体を含む、実施形態 A1 から E2 のいずれか 1 つに記載の方法。

30

【0530】

E4. 染色体が第21染色体、第18染色体または第13染色体である、実施形態 E3 に記載の方法。

【0531】

E5. 正規化された試料カウントが、第1のゲノムセクションについてのカウントから予測カウントを引き算し、それにより減算値を生成し、減算値をカウントの変動性の推定値で割ることを含むプロセスによって得られる、実施形態 A1 から D21、実施形態 E3 および実施形態 E4 のいずれか 1 つに記載の方法。

【0532】

E5.1. 予測カウントの変動性の推定値がカウントの中央絶対偏差 (MAD) である、実施形態 E5 に記載の方法。

40

【0533】

E5.2. カウントの変動性の推定値が、Rousseeuw および Croux によって導入される MAD の代替値、またはブートストラップ推定値である、実施形態 E5 に記載の方法。

【0534】

E5.3. 変動性の推定値が、1つまたは複数の共通の実験条件から生成された試料データについて得られる、実施形態 E5 から E5.2 のいずれか 1 つに記載の方法。

【0535】

50

E 5 . 4 . 変動性の推定値が、1つまたは複数の共通の実験条件から生成されたものではない試料データについて得られる、実施形態 E 5 から E 5 . 2 のいずれか1つに記載の方法。

【 0 5 3 6 】

E 5 . 5 変動性の推定値および予測カウントが、1つまたは複数の共通の実験条件から生成された試料データについて得られる、実施形態 E 5 から E 5 . 4 のいずれか1つに記載の方法。

【 0 5 3 7 】

E 6 . 正規化された試料カウントが、第1のゲノムセクションカウント表示から予測された第1のゲノムセクションカウント表示を引き算し、それにより減算値を生成し、減算値を第1のゲノムセクションカウント表示の変動性の推定値で割ることを含むプロセスによって得られる、実施形態 A 1 から E 4 のいずれか1つに記載の方法。

10

【 0 5 3 8 】

E 6 . 1 . 予測カウント表示の変動性の推定値がカウント表示の中央絶対偏差 (M A D) である、実施形態 E 6 に記載の方法。

【 0 5 3 9 】

E 6 . 2 . カウント表示の変動性の推定値が、R o u s s e e u w および C r o u s によって導入される M A D の代替値またはブートストラップ推定値である、実施形態 E 6 に記載の方法。

【 0 5 4 0 】

20

E 6 . 3 . 予測カウント表示の変動性の推定値が、1つまたは複数の共通の実験条件から生成された試料データについて得られる、実施形態 E 6 から E 6 . 2 のいずれか1つに記載の方法。

【 0 5 4 1 】

E 6 . 4 . 予測カウント表示の変動性の推定値が、1つまたは複数の共通の実験条件から生成されたものではない試料データについて得られる、実施形態 E 6 から E 6 . 2 のいずれか1つに記載の方法。

【 0 5 4 2 】

E 6 . 5 予測カウント表示の変動性の推定値および予測された第1のゲノムセクションカウント表示が、1つまたは複数の共通の実験条件から生成された試料データについて得られる、実施形態 E 6 から E 6 . 4 のいずれか1つに記載の方法。

30

【 0 5 4 3 】

E 7 . 1つまたは複数の共通の実験条件がフローセルを含む、実施形態 A 1 から E 6 . 6 のいずれか1つに記載の方法。

【 0 5 4 4 】

E 8 . 1つまたは複数の共通の実験条件がフローセルのチャンネルを含む、実施形態 A 1 から E 6 . 6 のいずれか1つに記載の方法。

【 0 5 4 5 】

E 9 . 1つまたは複数の共通の実験条件が試薬プレートを含む、実施形態 A 1 から E 6 . 6 のいずれか1つに記載の方法。

40

【 0 5 4 6 】

E 9 . 1 . 試薬プレートを使用して配列決定のために核酸を段階分けする、実施形態 E 9 に記載の方法。

【 0 5 4 7 】

E 9 . 2 . 試薬プレートを使用して配列決定のために核酸ライブラリーを調製する、実施形態 E 9 に記載の方法。

【 0 5 4 8 】

E 1 0 . 1つまたは複数の共通の実験条件が同定タグ指標を含む、実施形態 A 1 から E 6 . 6 のいずれか1つに記載の方法。

【 0 5 4 9 】

50

E 1 1 . 正規化された試料カウントを、ヌクレオチド配列読み取りまたは試料核酸のグアニンおよびシトシンの含量について補正する、実施形態 A 1 から E 1 0 のいずれか 1 つに記載の方法。

【 0 5 5 0 】

E 1 2 . カウントまたは正規化された試料カウントを局所重み付け多項式回帰に供するステップを含む、実施形態 E 1 1 に記載の方法。

【 0 5 5 1 】

E 1 2 . 1 局所重み付け多項式回帰が L O E S S 回帰である、実施形態 E 1 2 に記載の方法。

【 0 5 5 2 】

E 1 3 . 正規化された試料カウントを、参照ゲノムセクションにおいて反復するヌクレオチド配列について補正する、実施形態 A 1 から E 1 2 のいずれか 1 つに記載の方法。

【 0 5 5 3 】

E 1 4 . カウントまたは正規化された試料カウントを、参照ゲノムセクションにおいて反復するヌクレオチド配列について補正する、実施形態 E 1 3 に記載の方法。

【 0 5 5 4 】

E 1 5 . 正規化された試料カウントを得る前にカウントをフィルタリングするステップを含む、実施形態 A 1 から E 1 4 のいずれか 1 つに記載の方法。

【 0 5 5 5 】

E 1 6 . 試料核酸が一本鎖核酸を含む、実施形態 A 1 から E 1 5 のいずれか 1 つに記載の方法。

【 0 5 5 6 】

E 1 7 . 試料核酸が二本鎖核酸を含む、実施形態 A 1 から E 1 5 のいずれか 1 つに記載の方法。

【 0 5 5 7 】

E 1 8 . ヌクレオチド配列読み取りを得るステップが、試料核酸を、配列決定デバイスを使用した配列決定プロセスに供するステップを含む、実施形態 A 1 から E 1 7 のいずれか 1 つに記載の方法。

【 0 5 5 8 】

E 1 9 . アウトカムをもたらすステップが、試料核酸中の胎児核酸の分率をファクタリングするステップを含む、実施形態 A 1 から E 1 8 のいずれか 1 つに記載の方法。

【 0 5 5 9 】

E 2 0 . 試料核酸中の胎児核酸の分率を決定するステップを含む、実施形態 A 1 から E 1 9 のいずれか 1 つに記載の方法。

【 0 5 6 0 】

E 2 1 . 正規化された試料カウントを、ヌクレオチド配列読み取りまたは試料核酸のグアニンおよびシトシンの含量について補正せずを得る、実施形態 A 1 から E 2 0 のいずれか 1 つに記載の方法。

【 0 5 6 1 】

E 2 2 . 正規化された試料カウントを 1 つの実験条件について得る、実施形態 A 1 から E 2 0 のいずれか 1 つに記載の方法。

【 0 5 6 2 】

E 2 3 . 実験条件がフローセルである、実施形態 E 2 2 に記載の方法。

【 0 5 6 3 】

E 2 4 . 正規化された試料カウントを 2 つの実験条件について得る、実施形態 A 1 から E 2 0 のいずれか 1 つに記載の方法。

【 0 5 6 4 】

E 2 5 . 実験条件がフローセルおよび試薬プレートである、実施形態 E 2 4 に記載の方法。

【 0 5 6 5 】

10

20

30

40

50

E 2 6 . 実験条件がフローセルおよび同定タグ指標である、実施形態 E 2 4 に記載の方法。

【 0 5 6 6 】

E 2 7 . 正規化された試料カウントを 3 つの実験条件について得る、実施形態 A 1 から E 2 0 のいずれか 1 つに記載の方法。

【 0 5 6 7 】

E 2 8 . 実験条件がフローセル、試薬プレートおよび同定タグ指標である、実施形態 E 2 7 に記載の方法。

【 0 5 6 8 】

E 2 9 . 正規化された試料カウントを、(i) グアニンおよびシトシンの含量に従って補正し、(i) の後に、(i i) 実験条件に従って補正した後に得る、実施形態 A 1 から E 2 0 のいずれか 1 つに記載の方法。

10

【 0 5 6 9 】

E 3 0 . 正規化された試料カウントを、(i) の前に参照ゲノムセクションにおいて反復するヌクレオチド配列に従って補正した後に得る、実施形態 E 2 9 に記載の方法。

【 0 5 7 0 】

E 3 1 . (i i) が、フローセルに従って補正することからなる、実施形態 E 2 9 または E 3 0 に記載の方法。

【 0 5 7 1 】

E 3 2 . (i i) が、同定タグ指標に従って補正し、次いでフローセルに従って補正することからなる、実施形態 E 2 9 または E 3 0 に記載の方法。

20

【 0 5 7 2 】

E 3 3 . (i i) が、試薬プレートに従って補正し、次いでフローセルに従って補正することからなる、実施形態 E 2 9 または E 3 0 に記載の方法。

【 0 5 7 3 】

E 3 4 . (i i) が、同定タグ指標および試薬プレートに従って補正し、次いでフローセルに従って補正することからなる、実施形態 E 2 9 または E 3 0 に記載の方法。

【 0 5 7 4 】

E 3 5 . 正規化された試料カウントを、フローセルに従って補正することからなる実験条件に従った補正の後に得る、実施形態 E 2 1 に記載の方法。

30

【 0 5 7 5 】

E 3 6 . 正規化された試料カウントを、同定タグ指標に従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る、実施形態 E 2 1 に記載の方法。

【 0 5 7 6 】

E 3 7 . 正規化された試料カウントを、試薬プレートに従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る、実施形態 E 2 1 に記載の方法。

【 0 5 7 7 】

E 3 8 . 正規化された試料カウントを、同定タグ指標および試薬プレートに従って補正し、次いでフローセルに従って補正することからなる実験条件に従った補正の後に得る、実施形態 E 2 1 に記載の方法。

40

【 0 5 7 8 】

E 3 9 . 正規化された試料カウントを、参照ゲノムセクションにおいて反復するヌクレオチド配列に従って補正し、その後に実験条件に従って補正した後に得る、実施形態 E 3 2 から E 3 8 のいずれか 1 つに記載の方法。

【 0 5 7 9 】

E 4 0 . 正規化された試料カウントが Z スコアである、実施形態 E 1 から E 3 8 のいずれか 1 つに記載の方法。

【 0 5 8 0 】

50

E 4 1 . (i) が、

(a) 各試料について (i) 参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウンと (i i) 部分のそれぞれについての G C 含量との間のフィッティングした関係から、複数の試料について参照ゲノムの部分のそれぞれについてのグアニンおよびシトシン (G C) の偏りを決定することと、

(b) (i) G C の偏りと (i i) 参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウンとの間のフィッティングした関係から参照ゲノムの部分のそれぞれについてのゲノミックセクションの高度を算出し、それにより、算出されたゲノミックセクションの高度をもたらし、それにより、算出されたゲノミックセクションの高度における参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウンの偏りが減少

10

することとを含む、実施形態 E 2 9 から E 4 0 のいずれか 1 つに記載の方法。

【 0 5 8 1 】

E 4 2 . 参照ゲノムの部分が染色体内にある、実施形態 E 4 1 に記載の方法。

【 0 5 8 2 】

E 4 3 . 参照ゲノムの部分が染色体の部分内にある、実施形態 E 4 1 に記載の方法。

【 0 5 8 3 】

E 4 4 . 染色体が第 2 1 染色体である、実施形態 E 4 1 から E 4 3 のいずれか 1 つに記載の方法。

【 0 5 8 4 】

20

E 4 5 . 染色体が第 1 8 染色体である、実施形態 E 4 1 から E 4 3 のいずれか 1 つに記載の方法。

【 0 5 8 5 】

E 4 6 . 染色体が第 1 3 染色体である、実施形態 E 4 1 から E 4 3 のいずれか 1 つに記載の方法。

【 0 5 8 6 】

E 4 7 . (b) の前に、参照ゲノムの部分のいくつかまたは全部にマッピングされた配列読み取りのカウンについて誤差の尺度を算出し、参照ゲノムの特定の部分についての配列読み取りのカウンを誤差の尺度の閾値に従って除去または重み付けするステップを含む、実施形態 E 4 1 から E 4 6 のいずれか 1 つに記載の方法。

30

【 0 5 8 7 】

E 4 8 . 閾値を、第 1 のゲノミックセクションの高度と第 2 のゲノミックセクションの高度との間の標準偏差ギャップ 3 . 5 以上に応じて選択する、実施形態 E 4 7 に記載の方法。

【 0 5 8 8 】

E 4 9 . 誤差の尺度が R 因子である、実施形態 E 4 7 または実施形態 E 4 8 に記載の方法。

【 0 5 8 9 】

E 5 0 . R 因子が約 7 % ~ 約 1 0 % である参照ゲノムの部分についての配列読み取りのカウンを (b) の前に除去する、実施形態 E 4 9 に記載の方法。

40

【 0 5 9 0 】

E 5 1 . (a) のフィッティングした関係がフィッティングした線形関係である、実施形態 E 4 1 から E 5 0 のいずれか 1 つに記載の方法。

【 0 5 9 1 】

E 5 2 . 関係の傾きを線形回帰によって決定する、実施形態 E 5 1 に記載の方法。

【 0 5 9 2 】

E 5 3 . 各 G C の偏りが G C の偏り係数であり、この G C の偏り係数が (i) 参照ゲノムの部分のそれぞれにマッピングされた配列読み取りのカウンと (i i) 部分のそれ

50

それぞれについてのGC含量との間の線形関係の傾きである、実施形態E51またはE52に記載の方法。

【0593】

E54. (b)のフィッティングした関係がフィッティングした非線形関係である、実施形態E41からE50のいずれか1つに記載の方法。

【0594】

E55. 各GCの偏りが、GC曲率推定値を含む、実施形態E54に記載の方法。

【0595】

E56. (b)のフィッティングした関係が線形である、実施形態E41からE55のいずれか1つに記載の方法。

10

【0596】

E57. 関係の傾きを線形回帰によって決定する、実施形態E56に記載の方法。

【0597】

E58. (a)のフィッティングした関係が線形であり、(b)のフィッティングした関係が線形であり、ゲノミックセクションの高度 L_i が、参照ゲノムの部分のそれぞれについて、方程式：

$$L_i = (m_i - G_i S) I^{-1} \quad \text{方程式}$$

(式中、 G_i はGCの偏りであり、 I は(b)のフィッティングした関係の切片であり、 S は(b)の関係の傾きであり、 m_i は測定された、参照ゲノムの各部分にマッピングされたカウントであり、 i は試料である)

20

に従って決定される、実施形態E41からE57のいずれか1つに記載の方法。

【0598】

E59. 参照ゲノムの部分の数が約40,000以上の部分である、実施形態E41からE58のいずれか1つに記載の方法。

【0599】

E60. 参照ゲノムの各部分が所定の長さのヌクレオチド配列を含む、実施形態E41からE59のいずれか1つに記載の方法。

30

【0600】

E61. 所定の長さが約50キロボースである、実施形態E60に記載の方法。

【0601】

E62. (b)のGCの偏りをGCの偏りモジュールによって決定する、実施形態E41からE61のいずれか1つに記載の方法。

【0602】

F1. コンピュータ可読プログラムコードが組み込まれたコンピュータで使用可能な媒体を含むコンピュータプログラム製品であって、コンピュータ可読プログラムコードが、配列受信モジュール、論理処理モジュール、およびデータディスプレイ編成モジュールを含む別個のソフトウェアモジュールを含み、かつ、試料核酸における遺伝的変異の有無を同定するための方法であって、

40

(a)配列受信モジュールによって、試料核酸からヌクレオチド配列読み取りを得るステップと、

(b)論理処理モジュールによって、ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと、

(c)論理処理モジュールによって、各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(d)論理処理モジュールによって、1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウ

50

ントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(e) 論理処理モジュールによって、正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムを生成するステップと、

(f) データディスプレイ編成モジュールによって、論理処理モジュールによって決定されるのに応じて試料核酸における遺伝的変異の有無を示すデータディスプレイを編成するステップと

を含む方法の実行が遂行されるように適合されている、コンピュータプログラム製品。

【0603】

F2. 実施形態F1のコンピュータプログラム製品が記憶されているメモリを含む装置。

10

【0604】

F3. 実施形態F1に記載のコンピュータプログラム製品の1つまたは複数の機能を実行するプロセッサを含む、実施形態F2に記載の装置。

【0605】

F4. 核酸配列決定装置および処理装置を含むシステムであって、配列決定装置によって試料核酸からのヌクレオチド配列読み取りが得られ、処理装置によって配列決定装置からのヌクレオチド配列読み取りが得られ、かつ

(a) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと

20

(b) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(d) 正規化された試料カウントに基づいて試料核酸における遺伝的変異の有無を決定するアウトカムをもたらすステップと

を含む方法が実施されるシステム。

【0606】

30

G1. ヒト参照ゲノムhg19による第22染色体のヌクレオチド19,000,000位と22,000,000位の間の22q11.2微小欠失の有無を同定する方法であって、

(a) 試験被験体由来の循環している無細胞核酸を含む試料を得るステップと、

(b) 試料から試料核酸を単離するステップと、

(c) 試料核酸からヌクレオチド配列読み取りを得るステップと、

(d) ヌクレオチド配列読み取りを参照ゲノムセクションにマッピングするステップと

(e) 各参照ゲノムセクションにマッピングされたヌクレオチド配列読み取りの数をカウントし、それにより、カウントを得るステップと、

40

(f) (e)のカウントされ、マッピングされた配列読み取りを、反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って補正するステップと、

(g) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って(f)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得るステップと、

(h) 第22染色体のヌクレオチド19,000,000位と22,000,000位の間に対応する1つまたは複数の選択されたゲノミックセクションについて、試験被験体

50

と参照被験体についての正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価するステップと、

(i)(h)における評価に基づいて、試験被験体における遺伝的変異の有無を決定するアウトカムをもたらすステップを含む、方法。

【0607】

G2. 補正され、カウントされ、マッピングされた配列読み取りを、1つまたは複数の実験条件についてさらに補正した後に、残ったカウントを正規化する、実施形態F1からF3のいずれか1つに記載の方法。

【0608】

H1. 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って遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、この配列読み取りが、試験試料由来の循環している無細胞核酸の読み取りであり、1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(b) 正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されているシステム。

【0609】

I1. 1つまたは複数のプロセッサおよびメモリを含む装置であって、このメモリが、1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、この配列読み取りが、試験試料由来の循環している無細胞核酸の読み取りであり、1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(b) 正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されている装置。

【0610】

J1. コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、

(a) 参照ゲノムのゲノミックセクションにマッピングされた、試験試料由来の循環している無細胞核酸の読み取りである配列読み取りのカウントにアクセスし、

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(c) 正規化された試料カウントに基づいて胎児の異数性の有無を決定するように構成されている命令を含むコンピュータプログラム製品。

【0611】

K1. 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って遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、この配列読み取りが胎児を有する妊婦由来の循環している無細胞核酸の読み取りであり、1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(b) 正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を決定する

ように構成されているシステム。

【0612】

L1. 1つまたは複数のプロセッサおよびメモリを含む装置であって、このメモリが、1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、この配列読み取りが、胎児を有する妊婦由来の循環している無細胞核酸の読み取りであり、1つまたは複数のプロセッサによって遂行可能な命令が、

10

(a) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(b) 正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を決定する

ように構成されている装置。

20

【0613】

M1. コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、

(a) 参照ゲノムのゲノミックセクションにマッピングされた、胎児を有する妊婦由来の循環している無細胞核酸の読み取りである配列読み取りのカウントにアクセスし、

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、カウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(c) 正規化された試料カウントに基づいて、試験被験体における遺伝的変異の有無を決定する

30

ように構成されている命令を含むコンピュータプログラム製品。

【0614】

N1. 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って遂行可能な命令を含み、かつ、参照ゲノムのゲノミックセクションにマッピングされた配列読み取りのカウントを含み、この配列読み取りが、胎児を有する妊婦由来の循環している無細胞核酸の読み取りであり、1つまたは複数のプロセッサによって遂行可能な命令が、

(a) 反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、

40

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(a)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得、

(c) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体についての、正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価し、

(d) (c)における評価に基づいて、試験被験体における遺伝的変異の有無を決定す

50

る
ように構成されているシステム。

【0615】

O1. 1つまたは複数のプロセッサおよびメモリを含む装置であって、このメモリが、1つまたは複数のプロセッサによって遂行可能な命令を含み、かつ、参照ゲノムの部分にマッピングされた、胎児を有する妊婦由来の循環している無細胞核酸の読み取りである配列読み取りのカウントを含み、1つまたは複数のプロセッサによって遂行可能な命令が

(a) 反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、

10

(b) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(a)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得

(c) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体について、正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価し、

(d) (c)における評価に基づいて、試験被験体における遺伝的変異の有無を決定する
ように構成されている装置。

20

【0616】

P1. コンピュータ可読媒体上に有形的に組み込まれているコンピュータプログラム製品であって、1つまたは複数のプロセッサによって遂行される際に、

(a) 参照ゲノムの部分にマッピングされた、試験試料由来の循環している無細胞核酸の読み取りである配列読み取りのカウントにアクセスし、

(b) 反復配列および/または過大表示もしくは過小表示された配列の影響を最小限にするまたは排除する、選択された変数または特徴に従って、カウントされ、マッピングされた配列読み取りを補正し、

30

(c) 1つまたは複数の共通の実験条件に曝露させた試料、参照、または試料および参照を含む群について得られる予測カウントまたは予測カウントの誘導値に従って、(b)で残ったカウントを第1のゲノムセクションについて正規化、またはカウントの誘導値を第1のゲノムセクションについて正規化し、それにより、正規化された試料カウントを得

(d) 1つまたは複数の選択されたゲノミックセクションについて、試験被験体と参照被験体について、正規化されたカウントまたは正規化されたカウントの誘導値の間の差異の統計的有意性を評価し、

(e) (d)における評価に基づいて、試験被験体における遺伝的変異の有無を決定する
ように構成されている命令を含むコンピュータプログラム製品。

40

【0617】

本明細書において参照されている特許、特許出願、刊行物および文書のそれぞれの全体が、これによって参照により組み込まれる。上記の特許、特許出願、刊行物および文書を引用することは、前述のいずれもが先行技術に係ることを認めるものではなく、これらの刊行物または文書の内容または日付に関してどんな容認も構成するものでもない。

【0618】

当該技術の基本的な態様から逸脱することなく前述のものに改変を行うことができる。当該技術は、1つまたは複数の特定の実施形態を参照してかなり詳細に記載されているが、当業者は、本出願に詳細に開示されている実施形態に変化を生じさせてよく、それでも

50

これらの改変および改良は当該技術の範囲および主旨であることを理解されよう。

【0619】

本明細書において適切に例示的に記載されている技術は、本明細書に詳細に開示されていない要素（複数可）のいずれがなくとも実施することができる。したがって、例えば、本明細書の各例において、「含む（comprising）」、「から本質的になる（consisting essentially of）」および「からなる（consisting of）」という用語はいずれも他の2つの用語と置き換えることができる。使用されている用語および表現は、説明する用語として使用され、限定ではなく、そのような用語および表現を使用することによって、示され、記載されている特徴の等価物またはその一部は排除されず、特許請求された技術の範囲内で種々の改変が可能である。用語「a（1つの）」または「an（1つの）」は、要素のうちの1つ、または要素のうちの2つ以上のいずれかについて記載されていることが文脈上明らかでない限り、それが修飾する要素の1つまたは複数を指す場合がある（例えば、「a（1つの）試薬」という用語は、1つまたは複数の試薬を意味し得る）。「約」という用語は、本明細書で使用される場合、基になるパラメータの10%以内の値（すなわち、プラス10%またはマイナス10%）を指し、一連の値の最初に「約」という用語を使用することにより、値のそれぞれが修飾される（すなわち、「約1、2および3」とは、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グラムから110グラムの間の重量を含んでよい。さらに、値の列挙が本明細書に記載されている場合（例えば、約50%、60%、70%、80%、85%または86%）その列挙は、それらの中間の値および小数値の全てを含む（例えば、54%、85.4%）。したがって、本技術は代表的な実施形態および任意選択の特徴によって詳細に開示されているが、当業者は本明細書に開示されている概念の改変および変形を用いることができ、そのような改変および変形は本技術の範囲内であるとみなされることが理解されるべきである。

【0620】

当該技術のある特定の実施形態は、続く特許請求の範囲に記載されている。

10

20

【図1】

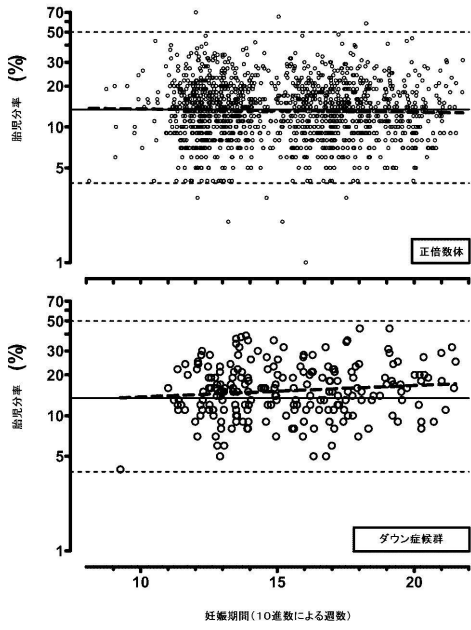


FIG. 1

【図2】

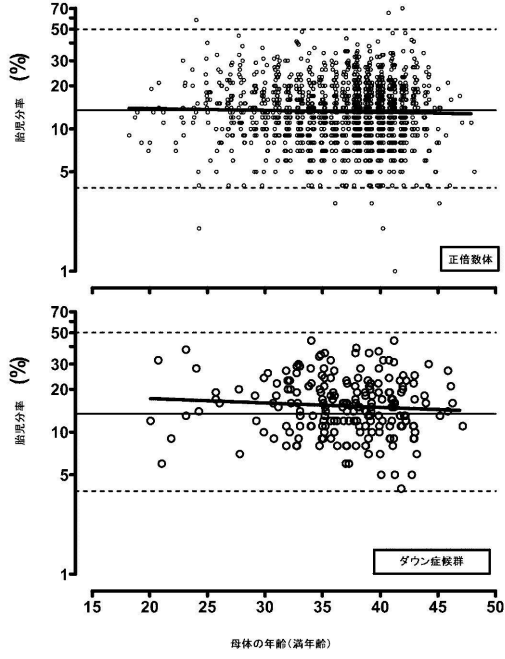


FIG. 2

【図3】

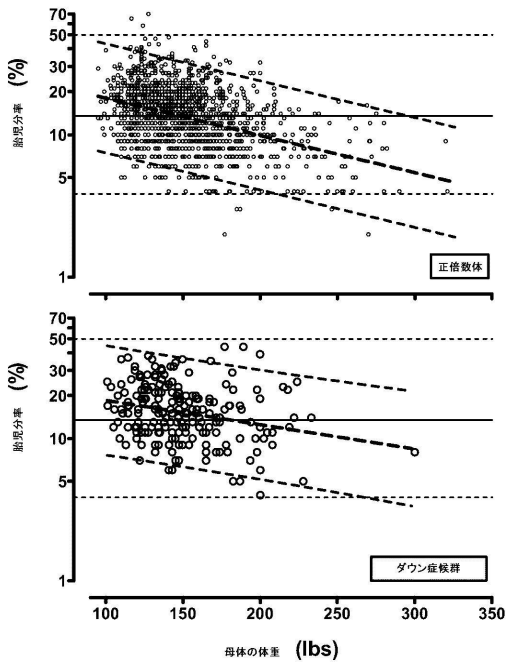


FIG. 3

【図4】

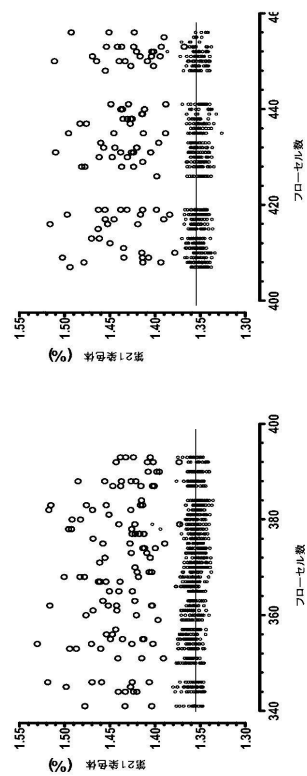


FIG. 4

【 図 5 】

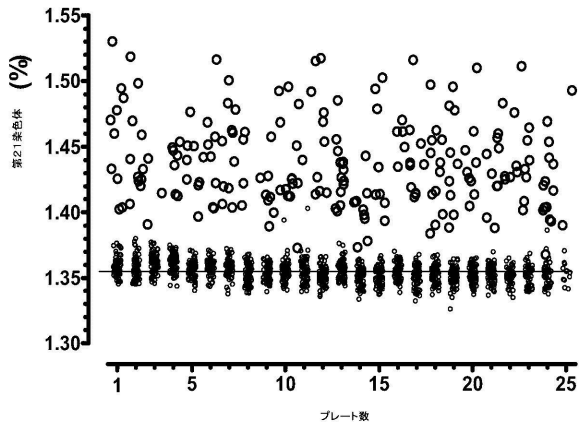
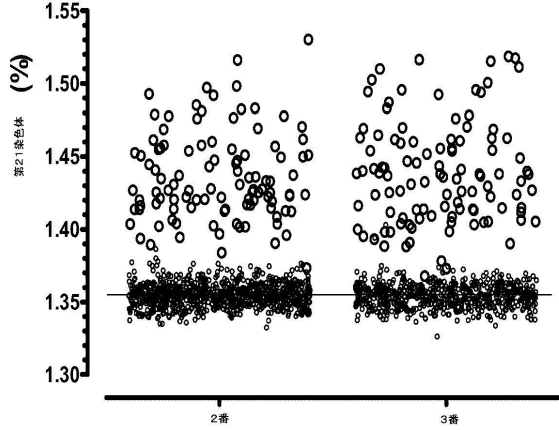


FIG. 5

【 図 6 】



Illumina HiSeq 2000 Instrument

FIG. 6

【 図 7 】

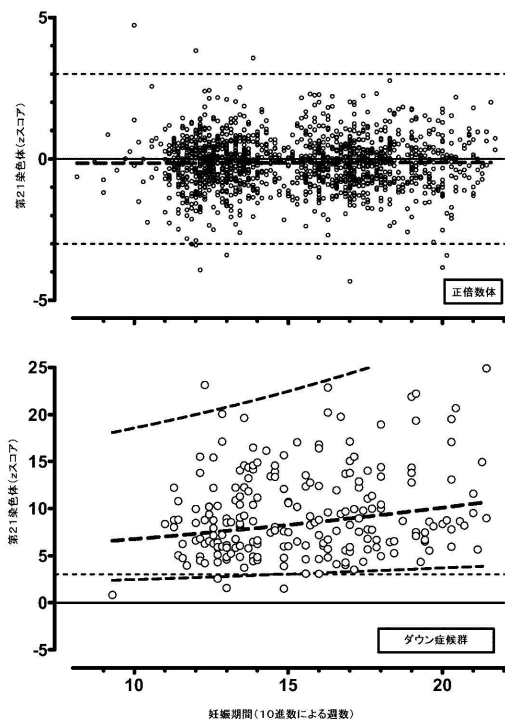


FIG. 7

【 図 8 】

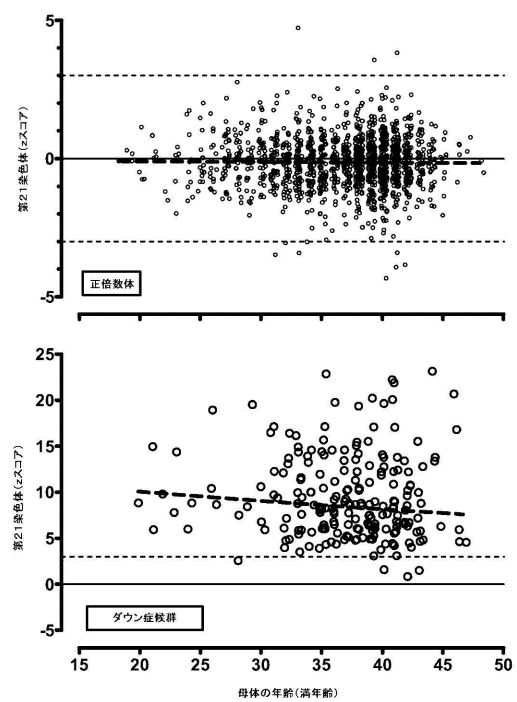


FIG. 8

【 図 9 】

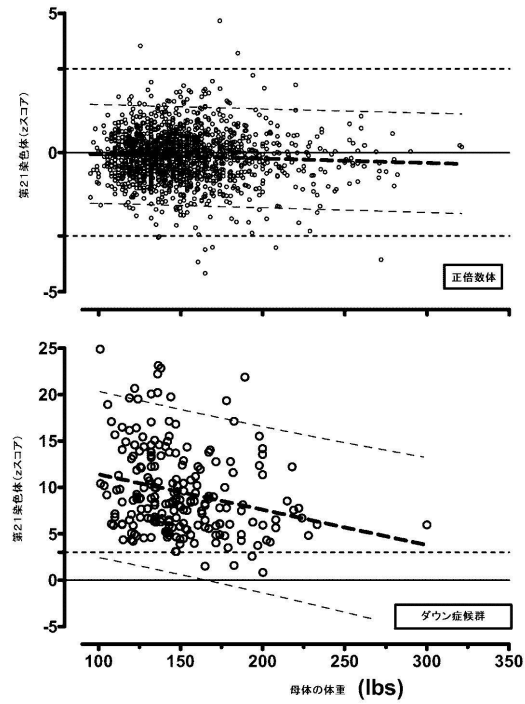


FIG. 9

【 図 10 】

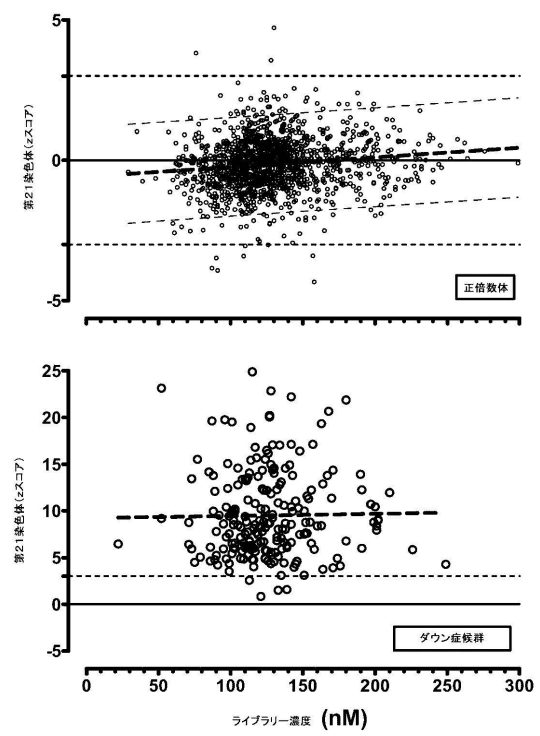


FIG. 10

【 図 11 】

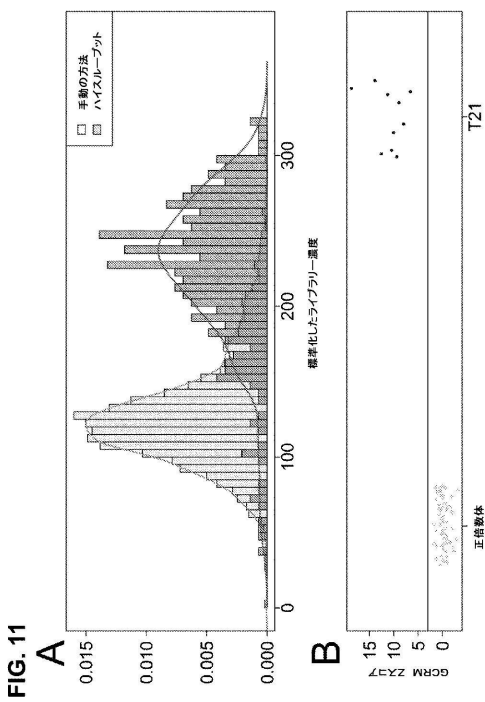


FIG. 11

【 図 12 】

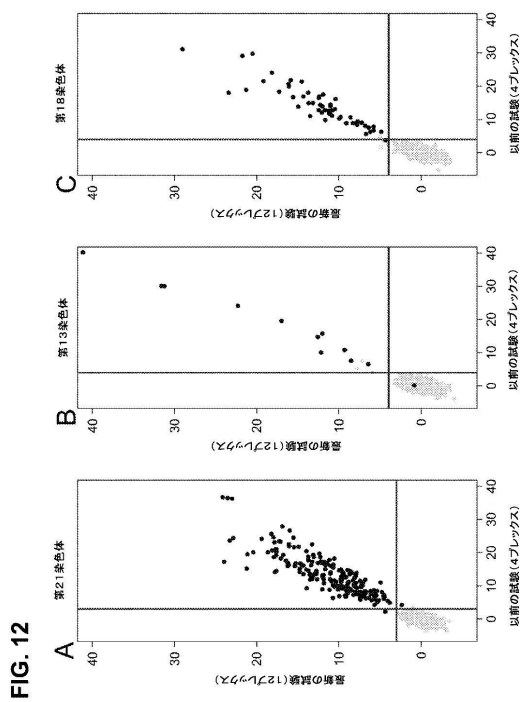
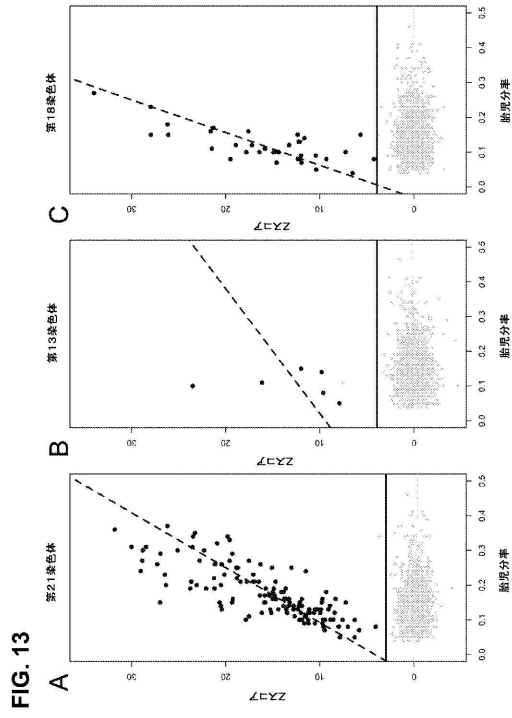
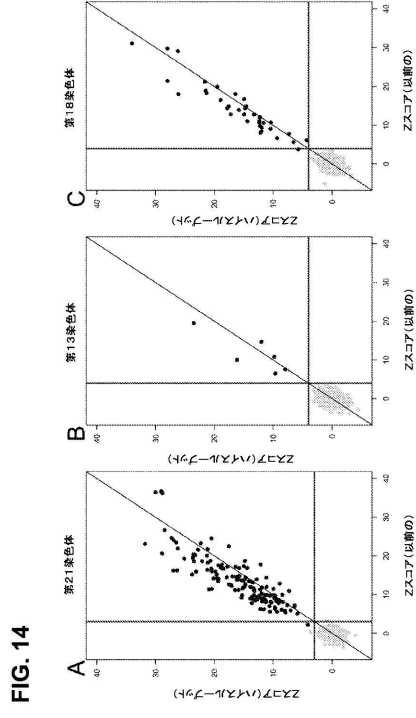


FIG. 12

【 13 】



【 14 】



フロントページの続き

- (31)優先権主張番号 PCT/US2012/059123
 (32)優先日 平成24年10月5日(2012.10.5)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 61/663,477
 (32)優先日 平成24年6月22日(2012.6.22)
 (33)優先権主張国 米国(US)

前置審査

- (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 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
 (72)発明者 ファン デン ボーム, デイルク ヨハネス
 アメリカ合衆国 カリフォルニア 92037, ラ ホヤ, ボネア ウェイ 638, ユニ
 ット エー
 (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

審査官 田ノ上 拓自

- (56)参考文献 米国特許出願公開第2011/0177517(US, A1)
 Genome Research, 2008年, Vol.18, p.839-846
 Clin. Chem., 2009年, Vol.55, No.4, p.641-658
 Genetics in Medicine, 2011年, Vol.13, No.11, p.913-920
 Clin. Chem., 2010年, Vol.56, No.3, p.459-463
 British J. Mathematical and Statistical Psychology, 1999年, Vol.52, p.303-320
 PLOS ONE, 2011年, Vol.6, Issue 7, e21791, p.1-7
 Proc. Natl. Acad. Sci. USA, 2008年, Vol.105, No.42, p.16266-16271
 PLOS ONE, 2010年, Vol.5, Issue 5, e10439, p1-7
 Bioinformatics, 2009年, Vol.25, No.10, p.1244-1250
 American J. Obstetrics and Gynecology, 2011年, 205.e1-e11

(58)調査した分野(Int.Cl., DB名)

C12Q 1/00 - 3/00
 G06F 19/20
 C12N 15/00 - 15/90
 Caplus/MEDLINE/EMBASE/BIOSIS(STN)
 WPIDS/WPIX(STN)
 PubMed