

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5863946号
(P5863946)

(45) 発行日 平成28年2月17日(2016.2.17)

(24) 登録日 平成28年1月8日(2016.1.8)

(51) Int.Cl.	F 1
C 12 Q 1/68 (2006.01)	C 12 Q 1/68 Z N A Z
G 01 N 33/48 (2006.01)	G 01 N 33/48 Z
C 12 N 15/09 (2006.01)	C 12 N 15/00 A

請求項の数 35 (全 67 頁)

(21) 出願番号	特願2014-505313 (P2014-505313)
(86) (22) 出願日	平成24年4月12日 (2012.4.12)
(65) 公表番号	特表2014-512817 (P2014-512817A)
(43) 公表日	平成26年5月29日 (2014.5.29)
(86) 国際出願番号	PCT/US2012/033391
(87) 国際公開番号	W02012/142334
(87) 国際公開日	平成24年10月18日 (2012.10.18)
審査請求日	平成27年3月11日 (2015.3.11)
(31) 優先権主張番号	61/474,362
(32) 優先日	平成23年4月12日 (2011.4.12)
(33) 優先権主張国	米国(US)

早期審査対象出願

(73) 特許権者	513257649 ベリナタ ヘルス インコーポレイテッド アメリカ合衆国 カリフォルニア州 レッドウッドシティ サギノー ドライブ 800
(74) 代理人	100102978 弁理士 清水 初志
(74) 代理人	100102118 弁理士 春名 雅夫
(74) 代理人	100160923 弁理士 山口 裕幸
(74) 代理人	100119507 弁理士 刑部 俊
(74) 代理人	100142929 弁理士 井上 隆一

最終頁に続く

(54) 【発明の名称】多型カウントを用いたゲノム画分の分析

(57) 【特許請求の範囲】

【請求項 1】

妊娠個体の体液から得られたDNA中の胎児DNAの画分を計算する方法であって、以下の段階:

- (a) 前記体液のサンプルを受け取る段階;
- (b) 前記サンプルから、前記体液中に存在する母性ゲノムと胎児ゲノムとの両方のDNAを抽出する条件下でDNAを抽出する段階;
- (c) 前記抽出したDNAを、複数の多型を含むDNA読み取り配列を生成する条件下で、核酸シーケンサーによって配列決定する段階;
- (d) 前記DNA読み取り配列を、前記複数の多型に対応する、参照配列上の複数の多型部位に対してマッピングする段階であって、10
ここで、前記マッピングは、参照配列上の多型部位に対して核酸配列をマッピングするようプログラムされたコンピューター装置を用いて実施される、段階;
- (e) 前記マッピングされたDNA読み取り配列のアレル頻度を、前記複数の多型について決定する段階;
- (f) 以下の接合性の事例:
 - (i) 妊娠個体がホモ接合型であり、かつ胎児がホモ接合型である事例、
 - (ii) 妊娠個体がホモ接合型であり、かつ胎児がヘテロ接合型である事例、
 - (iii) 妊娠個体がヘテロ接合型であり、かつ胎児がホモ接合型である事例、および
 - (iv) 妊娠個体がヘテロ接合型であり、かつ胎児がヘテロ接合型である事例、20

の 2 つ以上の事例における多型について段階(e)において決定された前記アレル頻度を用いて、前記抽出されたDNA中の胎児DNAの画分を計算する段階、
を含み、

ここで、段階(f)は、該妊娠個体におけるヘテロ接合型多型またはホモ接合型多型を同定するための細胞内DNAを用いることなく行われ、

段階(e)～(f)は、前記決定する段階、および計算する段階のためのプログラム命令の下で動く1つまたは複数のプロセッサーにおいて行われる、方法。

【請求項 2】

段階(f)における 2 つ以上の接合性の事例における多型についてのアレル頻度が、事例(ii)における多型についてのアレル頻度を含む、請求項 1 記載の方法。

10

【請求項 3】

段階(f)の計算が、事例(ii)および事例(iii)についてのデータを組み合わせて胎児DNAの画分を計算することを含む、請求項 2 記載の方法。

【請求項 4】

事例(ii)および事例(iii)についてのデータの組み合わせが、事例(iii)のデータを、事例(ii)のデータに変換することを含む、請求項 3 記載の方法。

【請求項 5】

事例(ii)および事例(iii)についてのデータを組み合わせて胎児DNAの画分を計算することが、事例(ii)のデータおよび変換した事例(iii)のデータに回帰法を適用することを含む、請求項 4 記載の方法。

20

【請求項 6】

胎児DNAの画分が、線形回帰モデルの回帰線の傾斜の倍として計算される、請求項 5 記載の方法。

【請求項 7】

事例(iii)のデータを事例(ii)のデータに変換することが、

$$A1=0.5D-A$$

$$D1=D$$

として、データ(D, A)を(D1, A1)に変換することを含む、請求項 4 記載の方法：

ここで、Dは、多型の適用範囲であり、およびAは、多型のマイナーアレルカウントである。

30

【請求項 8】

事例(iii)のデータを事例(ii)のデータに変換することが、三角法変換または回転行列の使用を含む、請求項 4 記載の方法。

【請求項 9】

胎児DNAの画分が、事例(ii)について2A/Dとして、および事例(iii)について1-2A/Dとして計算される、請求項 3 記載の方法：

ここで、Aは、問題の多型についてのマイナーアレル頻度であり、およびDは、問題の多型についての適用範囲である。

【請求項 10】

段階(f)の計算が、該複数の多型をフィルタリングし、該複数の多型のうち、規定の閾値より高いマイナーアレル頻度を有するいかなる多型をも検討から除くことをさらに含む、請求項1記載の方法。

40

【請求項 11】

段階(f)の計算が、該複数の多型をフィルタリングし、該複数の多型のうち、規定の閾値より低いマイナーアレル頻度を有するいかなる多型をも検討から除くことをさらに含む、請求項1記載の方法。

【請求項 12】

段階(f)での計算の前に、段階(e)において決定されたアレル頻度に対し、閾値を適用することをさらに含む、請求項1記載の方法。

【請求項 13】

50

段階(f)の計算が、段階(e)において決定されたアレル頻度を、混合モデルに適用することをさらに含む、請求項1記載の方法。

【請求項14】

混合モデルは、階乗モーメントを使用し、該4つの接合性の事例の2つ以上からのデータを組み合わせて胎児DNAの画分を計算する、請求項13記載の方法。

【請求項15】

階乗モーメントが、

$$F_1 = \frac{1}{n} \sum_{i=1}^n \frac{a_i}{d_i}$$

$$F_2 = \frac{1}{n} \sum_{i=1}^n \frac{a_i(a_i-1)}{d_i(d_i-1)}$$

...

$$F_j = \frac{1}{n} \sum_{i=1}^n \frac{a_i(a_i-1)\cdots(a_i-j+1)}{d_i(d_i-1)(d_i-j+1)}$$

10

として計算される、請求項14記載の方法：

ここで、 F_j は、 j 番目の階乗モーメントであり、 α_i は、 i 番目の多型のマイナーアレル頻度であり、 d_i は、 i 番目の多型の適用範囲であり、および n は、多型の数である。

20

【請求項16】

階乗モーメントと接合性の事例におけるマイナーアレルの可能性との間における以下の関係：

$$F_1 \approx \sum_{i=1}^m \alpha_i p_i^1$$

$$F_2 \approx \sum_{i=1}^m \alpha_i p_i^2$$

...

$$F_j \approx \sum_{i=1}^m \alpha_i p_i^j$$

30

...

$$F_g \approx \sum_{i=1}^m \alpha_i p_i^g$$

に基づいて胎児DNAの画分を計算することをさらに含む、請求項15記載の方法：

ここで、 α_i は、接合性の事例(i)における複数の多型の部分であり、 p_i は、接合性の事例(i)におけるマイナーアレルの二項確率であり、 m は、接合性の事例の数であり、 g は、計算された階乗モーメントの数である。

40

【請求項17】

該複数の多型が、事例(i)および事例(ii)における多型からなり、且つ以下の関係：

$$F_1 \approx \alpha_1 e + (1 - \alpha_1) \left(\frac{f}{2}\right)$$

$$F_2 \approx \alpha_1 e^2 + (1 - \alpha_1) \left(\frac{f}{2}\right)^2$$

$$F_3 \approx \alpha_1 e^3 + (1 - \alpha_1) \left(\frac{f}{2}\right)^3$$

に基づいて胎児DNAの画分を計算することをさらに含む、請求項15記載の方法：

50

ここで、 F_1 、 F_2 および F_3 は、最初の3つの階乗モーメントであり、 $__i$ は、接合性の事例(i)における複数の多型の部分であり、 e は、未知の配列決定のエラー率であり、および f は、胎児DNAの画分である。

【請求項18】

該複数の多型が、事例(i)および事例(ii)における多型からなり、且つ以下の関係：

$$F_1 \approx \alpha_1 e + (1 - \alpha_1) \left(\frac{f}{2}\right)$$

$$F_2 \approx \alpha_1 e^2 + (1 - \alpha_1) \left(\frac{f}{2}\right)^2$$

に基づいて胎児DNAの画分を計算することをさらに含む、請求項15記載の方法：

ここで、 F_1 および F_2 は、最初の2つの階乗モーメントであり、 $__i$ は、接合性の事例(i)における複数の多型の部分であり、 e は、既知の配列決定のエラー率であり、および f は、胎児DNAの画分である。

【請求項19】

該複数の多型が、事例(i)および事例(ii)における多型からなり、且つ以下の関係：

$$F_1 = \alpha_1 \left(\frac{e}{4}\right) + (1 - \alpha_1) \left(e + \frac{f}{2}\right)$$

$$F_2 = \alpha_1 \left(\frac{e}{4}\right)^2 + (1 - \alpha_1) \left(e + \frac{f}{2}\right)^2$$

に基づいて胎児DNAの画分を計算することをさらに含む、請求項15記載の方法：

ここで、 F_1 および F_2 は、最初の2つの階乗モーメントであり、 $__i$ は、接合性の事例(i)における複数の多型の部分であり、 e は、事例(i)および事例(ii)に別個に適用される既知の配列決定のエラー率であり、および f は、胎児DNAの画分である。

【請求項20】

抽出されたDNAが、妊娠個体の血漿から得られた無細胞DNAを含む、請求項1記載の方法。

。

【請求項21】

配列決定する段階が、複数の多型のいずれかを選択的に増幅させることなく実施される、請求項1記載の方法。

【請求項22】

1つまたは複数のプロセッサーにおいてプログラム命令を実行し、段階(f)において計算された胎児DNAの画分をコンピューター読み取り可能媒体に保存された妊娠個体についての患者医療記録に自動的に記録することをさらに含む、請求項1記載の方法。

【請求項23】

前記患者医療記録は、研究室、臨床医のオフィス、病院、健康維持組織、保険会社または個人医療記録ウェブサイトによって保持される、請求項22記載の方法。

【請求項24】

前記方法が、胎児DNAの画分の計算結果の、体液が採取されたヒト対象の処置を処方すること、開始することおよび/または変更することにおける使用のために行われる、請求項1記載の方法。

【請求項25】

前記方法が、胎児DNAの画分の計算結果の、1つまたは複数のさらなる試験を指示することおよび/または行うことにおける使用のために行われる、請求項1記載の方法。

【請求項26】

妊娠個体の体液から得られたDNA中の胎児DNAの画分を計算する方法であって、以下の段階：

(a)前記体液のサンプルを受け取る段階；

(b)前記サンプルから、前記体液中に存在する母性ゲノムと胎児ゲノムとの両方のDNAを抽出する条件下でDNAを抽出する段階；

10

20

30

40

50

(c) 前記抽出したDNAを、DNA読み取り配列を生成する条件下で、核酸シーケンサーによつて配列決定する段階；

(d) 前記DNA読み取り配列について、複数の両アレル性多型を同定する段階であつて、ここで、前記同定は、核酸配列を比較し多型を同定するようプログラムされたコンピューター装置を用いて、前記DNA読み取り配列を参照配列と比較すること、または互いを比較することにより行う、段階；

(e) 前記DNA読み取り配列のアレル頻度を、前記複数の両アレル性多型について決定する段階；

(f) 以下の接合性の事例：

(i) 妊娠個体がホモ接合型であり、かつ胎児がホモ接合型である事例、

10

(ii) 妊娠個体がホモ接合型であり、かつ胎児がヘテロ接合型である事例、

(iii) 妊娠個体がヘテロ接合型であり、かつ胎児がホモ接合型である事例、および

(iv) 妊娠個体がヘテロ接合型であり、かつ胎児がヘテロ接合型である事例、

の2つ以上の事例における多型について段階(e)において決定された前記アレル頻度とを用いて、前記抽出されたDNA中の胎児DNAの画分を計算する段階、を含み、

ここで、段階(f)は、該妊娠個体におけるヘテロ接合型多型またはホモ接合型多型を同定するための細胞内DNAを用いることなく行われ、

段階(e)～(f)は、前記決定する段階、および計算する段階のためのプログラム命令の下で動く1つまたは複数のプロセッサーにおいて行われる、方法。

20

【請求項27】

1または複数の計算が、接合性の事例(ii)における多型について段階(e)で決定されたアレル頻度を用いる、請求項26記載の方法。

【請求項28】

段階(f)の計算が、該複数の両アレル性多型をフィルタリングし、規定の閾値より高いマイナーアレル頻度を有するいかなる多型をも検討から除くことをさらに含む、請求項26記載の方法。

【請求項29】

DNA読み取り配列は、20塩基対と300塩基対との間の長さを有する、請求項26記載の方法。

30

【請求項30】

段階(f)の計算の前に、段階(e)において決定されたアレル頻度に対し、閾値を適用することをさらに含む、請求項26記載の方法。

【請求項31】

段階(f)の計算が、段階(e)で決定されたアレル頻度を、混合モデルに適用することをさらに含む、請求項26記載の方法。

【請求項32】

混合モデルは、階乗モーメントを使用する、請求項31記載の方法。

【請求項33】

抽出されたDNAが、妊娠個体の血漿から得られた無細胞DNAを含む、請求項26記載の方法。

40

【請求項34】

段階(f)の計算が、事例(ii)および事例(iii)についてのデータを組み合わせて胎児DNAの画分を計算することを含む、請求項26記載の方法。

【請求項35】

事例(ii)および事例(iii)についてのデータの組み合わせが、事例(iii)のデータを、事例(ii)のデータに変換することを含む、請求項34記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

50

関連出願の相互参照

本出願は、米国仮特許出願第61/474,362号（2011年4月12日出願）の恩典および優先権を主張する。その内容は、その全体があらゆる目的のために参照により本明細書に組み入れられる。

【背景技術】

【0002】

背景

母性血液中の自由浮遊胎児DNA（時々、「無細胞DNA」または「cfDNA」と呼ばれる）の発見は、染色体異常、異数性および血液サンプルからの異常を検出する可能性を示す。母性血漿における胎児DNAの画分存在量は、一定ではなく、サンプル操作および在胎月齢を含む種々の因子によって変動する。

10

【0003】

DNA配列決定が染色体異常または遺伝的欠陥を同定するために使用される場合、DNAの総数における胎児DNAの相対存在量を知ることが重要である。例えば、胎児画分が分かっている場合、検出力（異常の事例を同定する可能性、すなわち感受性）は、並べ替え方法によって、またはから無限大まで（は、異常なしの帰無仮説の下のスコアの集団の有意性（誤った異常検出の最大尤度）についての臨界点である）の線形結合の積分もしくは非心F分布の畳み込みを介して計算され得る。

【0004】

胎児画分の検出のための既存の方法の弱点は、これらが、性染色体の存在量（雄性体の胚性DNAの相対的存在量を信頼性をもって測定するためにのみ使用され得る）または妊娠組織と胚性組織との間で異なって発現されることで知られる遺伝子のmRNA配列の存在量（これは、在胎月齢または他の因子に起因する発現の変動性を前提とする）の測定値に依存していることである。

20

【0005】

胎児画分の推定は、親の民族差異的集団遺伝学パラメータおよび配列決定の誤差を含むいくつかの邪魔な因子によって困難である場合がある。したがって、これらのおよび他の一般的に存在する混乱因子の存在下で強固である方法を有することが、望ましい。

【発明の概要】

【0006】

30

概要

開示される特定の態様は、母性血液サンプルを配列決定することにより、信頼性をもつて胎児自由浮遊DNAの相対的存在量を測定するコンピューター方法に関する。

【0007】

特定の態様において、本発明は、信頼性をもって、小塩基の変異または挿入欠失などの多型から、胎児画分を推定する方法（これは、親の民族性、胚の性別、在胎月齢および他の環境因子に関して強固である）を提供する。本明細書において開示される多くの例が、適切な多型としてSNPを使用する。本発明は、公知の多型に対して標的化した意図的な事前指定された再配列決定研究の一環として適用され得るか、または母性血漿から作成された重複配列における一致によって見出される変異の邀及的分析（または数人のヒト由来のDNAの混合物が存在する場合、任意の他の設定）において使用され得る。

40

【0008】

この文書は、母性血液サンプル中の胎児DNAの画分存在量の推定のための技術を提示する。開示される特定の技術は、観察されたSNPの対立遺伝子頻度（偶然見出されたか、または胎児画分を推定する目的で指定された既知のSNPのパネルにおいて見出された）を使用する。

【0009】

本開示の多くがサンプル中の胎児核酸の画分の推定に関するが、本発明は、それに限定されない。本明細書において記載される技術および装置は、多くの場合、2つのゲノム（親および子のゲノムに関してても関さなくてもよい）の混合物において1つのゲノム由来の

50

核酸の画分を推定するために使用され得る。

【0010】

本開示の特定の局面は、妊娠個体の体液から得られたDNA中の胎児DNAの画分を測定する方法に関する。このような方法は、以下の操作によって特徴付けられ得る：(a)体液のサンプルを受け取ること；(b)母性ゲノムおよび胎児ゲノムの両方のDNA抽出物が体液中に存在する条件下で、サンプルからDNAを抽出すること；(c)1つまたは複数の多型を含むDNAセグメント配列を生成する条件下で、核酸シーケンサーによって抽出されたDNAを配列決定すること；(d)体液中のDNAの配列決定に由来するDNAセグメント配列を、参照配列における1つまたは複数の指定された多型についてマッピングすること；(e)マッピングされたDNAセグメント配列のアレル頻度を、指定された多型の少なくとも1つについて決定すること；(f)少なくとも1つの指定された多型を、妊娠個体の接合生殖性と胎児の接合生殖性との組み合わせに基づいて分類すること；および(g)妊娠個体から得られたDNAにおける胎児DNAの画分を、(e)において決定されたアレル頻度および(f)からの接合生殖性の組み合わせを用いて推定すること。

【0011】

マッピングは、1つまたは複数の指定された多型に対して核酸配列をマッピングするようにプログラムされたコンピューター装置を用いて行われてもよい。一般に、操作(d)～(g)のいずれかは、プログラム指示の下で動く1つまたは複数のプロセッサーにおいて行われてもよい。

【0012】

特定の態様において、妊娠個体の体液から得られたDNAは、妊娠個体の血漿から得られた無細胞DNAである。典型的には、配列決定は、1つまたは複数の指定された多型のいずれかを選択的に増幅することなく行われる。

20

【0013】

特定の態様において、胎児を孕む個体の血液から得られたDNAセグメントのマッピングは、多型のデータベースに対するセグメントのコンピューター的マッピングを含む。特定の態様において、(f)における分類は、少なくとも1つの指定された多型を、以下の組み合わせのうちの1つに分類する：(i)妊娠個体がホモ接合型であり、胎児がホモ接合型である、(ii)妊娠個体がホモ接合型であり、胎児がヘテロ接合型である、(iii)妊娠個体がヘテロ接合型であり、胎児がホモ接合型である、および(iv)妊娠個体がヘテロ接合型であり、胎児がヘテロ接合型である。

30

【0014】

種々のフィルタリング操作が、使用され得る。これらは、例えば、組み合わせ(i)または組み合わせ(iv)に分類されるいかなる多型も検討しないことを含む。別の例において、この方法は、少なくとも1つの指定された多型をフィルタリングし、規定の閾値より高いマイナーアレル頻度を有する任意の多型を排除することを、さらに含む。なお別の例において、この方法は、少なくとも1つの指定された多型をフィルタリングし、規定の閾値より低いマイナーアレル頻度を有する任意の多型を排除する操作を含む。

【0015】

分類操作は、種々の方法において実施され得る。例えば、(e)において決定されたアレル頻度に対し、閾値を適用することを含んでもよい。別の例において、この分類操作は、複数の多型について得られた(e)からのアレル頻度データを、混合モデルに適用することを含む。1つの実施において、混合モデルは階乗モーメントを使用する。

40

【0016】

本明細書において記載されるように決定された胎児画分は、種々の適用のために使用され得る。いくつかの例において、本明細書において記載される方法は、1つまたは複数のプロセッサーにおいてプログラム指示を実行し、(g)において決定された胎児DNAの画分を、妊娠個体についての患者医療記録に自動的に記録し、コンピューター読み取り可能媒体上に保存する操作を含む。患者医療記録は、研究室、臨床医のオフィス、病院、健康維持組織、保険会社または個体医療記録ウェブサイトによって保持され得る。別の適用におい

50

て、胎児DNAの画分の推定は、母性試験サンプルが採取されたヒト対象の処置を処方するか、開始するか、および/または変更するために使用される。別の適用において、胎児DNAの画分の推定は、1つまたは複数のさらなる試験を指示するか、および/または行うために使用される。

【0017】

本開示の別の局面は、妊娠個体の体液から得られたDNA中の胎児DNA画分を推定するための装置に関する。このような装置は、以下の特徴によって特徴付けられ得る：(a)(i)母性ゲノムおよび胎児ゲノムの両方のDNAを含む体液のサンプルから抽出されたDNAを受け取り、そして(ii)1つまたは複数の指定された多型を含むDNAセグメント配列を生成する条件下で、抽出されたDNAを配列決定するように構成されたシーケンサー；ならびに(b)種々の操作（例えば、本明細書において記載された2つ以上的方法操作と共に記載された操作）を行うよう1つまたは複数のプロセッサーに命令するよう構成された（例えば、プログラムされた）コンピューター装置。いくつかの態様において、このコンピューター装置は、(i)参照配列における1つまたは複数の指定された多型に核酸配列をマッピングし、(ii)少なくとも1つの指定された多型についてマッピングされたDNAセグメント配列のアレル頻度を決定し、(iii)少なくとも1つの指定された多型を、妊娠個体の接合生殖性と胎児の接合生殖性との組み合わせに基づいて分類し、そして(iv)妊娠個体から得られたDNAにおける胎児DNAの画分を、アレル頻度および接合生殖性の組み合わせを用いて推定するように、構成される。

【0018】

特定の態様において、装置はまた、母性ゲノムと胎児ゲノムの両方のDNAを抽出する条件下で、サンプルからDNAを抽出するためのツールをも備える。いくつかの実施において、装置は、シーケンサーにおいて配列決定するために、妊娠個体の血漿から得られた無細胞DNAを抽出するように構成されたモジュールを備える。

【0019】

いくつかの例において、装置は、多型のデータベースを備える。コンピューター装置は、胎児を孕む個体の血液から得られたDNAセグメントを、セグメントを多型のデータベースに対しコンピューター的にマッピングすることによりマッピングするよう、1つまたは複数のプロセッサーに命令するようにさらに構成され得る。データベースにおける配列は、参照配列の一例である。参照配列の他の例は、以下に提示される。

【0020】

特定の態様において、コンピューター装置は、少なくとも1つの指定された多型を、以下の組み合わせのうちの1つに分類するよう、1つまたは複数のプロセッサーに命令するようにさらに構成される：(i)妊娠個体がホモ接合型であり、胎児がホモ接合型である、(ii)妊娠個体がホモ接合型であり、胎児がヘテロ接合型である、(iii)妊娠個体がヘテロ接合型であり、胎児がホモ接合型である、および(iv)妊娠個体がヘテロ接合型であり、胎児がヘテロ接合型である。いくつかの態様において、コンピューター装置は、組み合わせ(i)または組み合わせ(iv)に分類されたいかなる多型をも検討から除くよう、1つまたは複数のプロセッサーに命令するようにさらに構成される。

【0021】

特定の態様において、コンピューター装置は、規定の閾値より高いマイナーアレル頻度を有するいかなる多型をも検討から除くよう、1つまたは複数のプロセッサーに命令するようにさらに構成される。いくつかの態様において、コンピューター装置は、1つまたは複数の指定された多型をフィルタリングし、規定の閾値より低いマイナーアレル頻度を有するいかなる多型をも検討から除くよう1つまたは複数のプロセッサーに命令するように、さらに構成される。特定の態様において、コンピューター装置は、アレル頻度に対し閾値を適用することによって、少なくとも1つの指定された多型を分類するよう、1つまたは複数のプロセッサーに命令するようさらに構成される。

【0022】

特定の態様において、コンピューター装置は、複数の多型について得られたアレル頻度

10

20

30

40

50

データを、混合モデルに対して適用することにより、少なくとも1つの指定された多型を分類するよう、1つまたは複数のプロセッサーに命令するようさらに構成される。混合モデルは、階乗モーメントを使用してもよい。

【0023】

特定の態様において、コンピューター装置は、DNAの胎児の画分を、妊娠個体についての患者医療記録に自動的に記録し、コンピューター読み取り可能媒体に保存するよう、1つまたは複数のプロセッサーに命令するようさらに構成される。患者医療記録は、研究室、臨床医のオフィス、病院、健康維持組織、保険会社または個体医療記録ウェブサイトによって保持され得る。

【0024】

本開示の別の局面は、以下の操作にしたがって、妊娠個体の体液から得られたDNAにおける胎児DNAの画分を推定する方法に関する:(a)複数の多型配列に対し妊娠個体の体液から得られたDNAのセグメント(このDNAは、複数の多型配列を同定する条件下で配列決定されていた)をマッピングし;(b)複数の多型配列のそれぞれについてマッピングした核酸のアレル頻度を決定し;そして(c)アレル頻度を混合モデルに適用し、胎児を孕む個体の血液から得られたDNAにおける胎児DNA画分の推定を得る。操作(a)~(c)の任意の1つまたは複数は、プログラム指示の下で動く1つまたは複数のプロセッサーにおいて行われ得る。特定の態様において、操作(c)は、1つまたは複数のプロセッサーにおいて指示を実行し、複数の多型配列のそれぞれについてのアレル頻度データの階乗モーメントについての一連の等式を解くことを含む。いくつかの態様において、混合モデルは、配列決定のエラーを説明する。

【0025】

特定の態様において、この方法は、胎児および妊娠個体の両方においてヘテロ接合型であると同定された多型についてのアレル頻度のコンピューター的除外を、さらに含む。いくつかの実施において、(c)の前に、この方法は、胎児および妊娠個体の両方においてホモ接合型であると同定された多型についてのアレル頻度のコンピューター的除外の操作を含む。いくつかの実施において、(c)の前に、この方法は、妊娠個体においてヘテロ接合型であると同定された多型についてのアレル頻度のコンピューター的除外の操作を含む。

【0026】

妊娠個体の体液から得られたDNAは、妊娠個体の血漿から得られた無細胞DNAであり得る。体液から得られた核酸のマッピングは、多型のデータベースに対するセグメントのマッピングによって実施され得る。

【0027】

本開示のこの局面の方法は、妊娠個体の体液由来のDNAを、多型配列を含むDNAセグメント配列を生成する条件下で、核酸シーケンサーによって配列決定することをさらに含み得る。

【0028】

いくつかの実施において、(a)におけるマッピングは、複数の両アレル多型配列を同定することを含む。他の態様において、(a)におけるマッピングは、事前に規定された複数の多型配列に対しDNAのセグメントをマッピングすることを含む。

【0029】

いくつかの態様において、この局面の方法は、1つまたは複数のプロセッサーにおいてプログラム指示を実行し、(c)において決定された胎児のDNAの画分を、妊娠個体の患者医療記録に自動的に記録し、コンピューター読み取り可能媒体に保存することをさらに含む。患者医療記録は、研究室、臨床医のオフィス、病院、健康維持組織、保険会社または個体医療記録ウェブサイトによって保持され得る。

【0030】

胎児DNAの画分の推定に基づき、この局面の方法は、母性試験サンプルが採取されたヒト対象の処置を処方すること、開始すること、および/または変更することをさらに含み得る。胎児DNAの画分の推定に基づき、この局面の方法は、1つまたは複数のさらなる試験

10

20

30

40

50

を指示することおよび/または行うことをさらに含み得る。

【0031】

本開示のなお別の局面にしたがい、以下の操作を用いて、妊娠個体の体液から得られたDNAにおける胎児DNAの画分を推定するための方法が、提供される:(a)体液のサンプルを受け取ること;(b)母性ゲノムおよび胎児ゲノムの両方のDNA抽出物が体液中に存在する条件下で、サンプルからDNAを抽出すること;(c)DNAセグメント配列を生成する条件下で、核酸シーケンサーによって抽出されたDNAを配列決定すること;(d)体液に由来するDNAセグメント配列を比較し、この比較から、1つまたは複数の両アレル多型を同定すること;(e)同定された多型の少なくとも1つについてのDNAセグメント配列のアレル頻度を決定すること;(f)少なくとも1つの同定された多型を、妊娠個体の接合生殖性と胎児の接合生殖性との組み合わせに基づいて分類すること;および(g)妊娠個体から得られたDNAにおける胎児DNAの画分を、(e)において決定されたアレル頻度および(f)からの接合生殖性の組み合わせを用いて推定すること。

【0032】

マッピングは、1つまたは複数の指定された多型に対し核酸配列をマッピングするようプログラミングされたコンピューター装置を用いて、行われ得る。一般に、操作(d)~(g)のいずれかが、プログラム指示の下で動く1つまたは複数のプロセッサーにおいて行われ得る。

【0033】

この局面の特定の実施において、DNAセグメント配列は、約20塩基対と約300塩基対との間の長さを有する。

【0034】

この局面の特定の態様において、(f)における分類は、少なくとも1つの同定された多型を、以下の組み合わせの1つに分類する:(i)妊娠個体がホモ接合型であり、胎児がホモ接合型である、(ii)妊娠個体がホモ接合型であり、胎児がヘテロ接合型である、(iii)妊娠個体がヘテロ接合型であり、胎児がホモ接合型である、および(iv)妊娠個体がヘテロ接合型であり、胎児がヘテロ接合型である。この方法はさらに、組み合わせ(i)または組み合わせ(iv)に分類されるいかなる多型をも検討から除くことを含み得る。

【0035】

種々の態様にしたがい、この局面の方法は、本明細書において記載されるフィルタリング操作および/または分類操作を、他の局面に関連して含み得る。例えば、この局面の方法は、1つまたは複数の同定された多型をフィルタリングすることにより、規定の閾値より高いマイナーアレル頻度を有するいかなる多型をも検討から除くことを含み得る。いくつかの場合、少なくとも1つの同定された多型を分類することは、(e)において決定されたアレル頻度に対し閾値を適用することを含む。本明細書において記載される混合モデルの用途は、同定された多型を分類するために使用され得る。

【0036】

本開示の別の局面は、胎児DNAの画分を推定するための、以下の要素を備える装置に関する:(a)(i)母性ゲノムおよび胎児ゲノムの両方のDNAを含む体液のサンプルから抽出されたDNAを受け取り、そして(ii)抽出されたDNAを配列決定し、DNAの配列セグメントを生成するように構成されたシーケンサー;ならびに(b)1つまたは複数のプロセッサーに指示し、(i)妊娠個体の体液から得られたDNAの配列セグメントを複数の多型配列に対してマッピングし、(ii)マッピングされたDNAの配列セグメントからの複数の多型配列のそれぞれについてアレル頻度を決定し、そして(iii)このアレル頻度を、混合モデルに適用し、胎児を孕む個体の血液から得られたDNAにおける胎児DNAの画分の推定を得るよう構成されたコンピューター装置。

【0037】

胎児DNAの画分を推定するためのなお別の装置は、以下の要素を備える:(a)(i)母性ゲノムおよび胎児ゲノムの両方のDNAを含む体液のサンプルから抽出されたDNAを受け取り、そして(ii)DNAセグメント配列を生成する条件下で、抽出されたDNAを配列決定するように構

10

20

30

40

50

成されたシーケンサー；ならびに(b)1つまたは複数のプロセッサーに指示し、(i)体液に由来するDNAセグメント配列を比較し、この比較から、1つまたは複数の両アレル多型を同定し、(ii)同定された多型の少なくとも1つについてDNAセグメント配列のアレル頻度を決定し、(iii)少なくとも1つの同定された多型を、妊娠個体の接合生殖性と胎児の接合生殖性との組み合わせに基づいて分類し、そして(iii)妊娠個体から得られたDNAにおける胎児DNAの画分を、アレル頻度および接合生殖性の組み合わせを用いて推定するよう構成されたコンピューター装置。

【0038】

本明細書において記載される装置の局面において使用される命令および/またはハードウェアは、本明細書において開示される方法の局面の任意の1つまたは複数のコンピューター的 操作またはアルゴリズム的 操作の実行を、このような操作が明白に上で挙げられているか否かに関わらず、提供し得る。

10

【0039】

[本発明1001]

妊娠個体の体液から得られたDNA中の胎児DNAの画分を推定する方法であって、以下の段階：

- (a)前記体液のサンプルを受け取る段階；
- (b)前記サンプルから、前記体液中に存在する母性ゲノムと胎児ゲノムとの両方のDNAを抽出する条件下でDNAを抽出する段階；
- (c)前記抽出したDNAを、1つまたは複数の多型を含むDNAセグメント配列を生成する条件下で、核酸シーケンサーによって配列決定する段階；
- (d)前記体液中のDNAの配列決定に由来する前記DNAセグメント配列を、参照配列における1つまたは複数の指定された多型に対してマッピングする段階であって、ここで、前記マッピングは、1つまたは複数の指定された多型に対して核酸配列をマッピングするようプログラムされたコンピューター装置を用いて実施される、段階；
- (e)前記マッピングされたDNAセグメント配列のアレル頻度を、前記指定された多型の少なくとも1つについて決定する段階；
- (f)前記少なくとも1つの指定された多型を、前記妊娠個体の接合生殖性と前記胎児の接合生殖性との組み合わせに基づいて分類する段階；および
- (g)(e)において決定された前記アレル頻度と(f)からの接合生殖性の組み合わせとを用いて、前記妊娠個体から得られたDNA中の胎児DNAの画分を推定する段階を含み、ここで、(e)～(g)は、前記決定する段階、分類する段階、および推定する段階のためのプログラム命令の下で動く1つまたは複数のプロセッサーにおいて行われる、方法。

20

[本発明1002]

(f)の分類する段階は、少なくとも1つの指定された多型を、以下の組み合わせのうちの1つに分類する、本発明1001の方法：(i)妊娠個体がホモ接合型であり、かつ胎児がホモ接合型である、(ii)妊娠個体がホモ接合型であり、かつ胎児がヘテロ接合型である、(iii)妊娠個体がヘテロ接合型であり、かつ胎児がホモ接合型である、および(iv)妊娠個体がヘテロ接合型であり、かつ胎児がヘテロ接合型である。

40

[本発明1003]

組み合わせ(i)または組み合わせ(iv)に分類されるいかなる多型も、検討から除くことをさらに含む、本発明1002の方法。

[本発明1004]

少なくとも1つの指定された多型をフィルタリングし、規定の閾値より高いマイナーアレル頻度を有するいかなる多型をも検討から除くことをさらに含む、本発明1001の方法。

[本発明1005]

少なくとも1つの指定された多型をフィルタリングし、規定の閾値より低いマイナーアレル頻度を有するいかなる多型をも検討から除くことをさらに含む、本発明1001の方法。

[本発明1006]

50

少なくとも1つの指定された多型を分類する段階は、(e)において決定されたアレル頻度に対し、閾値を適用することを含む、本発明1001の方法。

[本発明1007]

少なくとも1つの指定された多型を分類する段階は、複数の多型について得られた(e)からのアレル頻度データを、混合モデルに適用することを含む、本発明1001の方法。

[本発明1008]

混合モデルは、階乗モーメントを使用する、本発明1007の方法。

[本発明1009]

妊娠個体の体液から得られたDNAは、前記妊娠個体の血漿から得られた無細胞DNAである、本発明1001の方法。

10

[本発明1010]

胎児を孕む個体の血液から得られたDNAセグメントをマッピングする段階は、多型のデータベースに対して前記セグメントをコンピューター的にマッピングすることを含む、本発明1001の方法。

[本発明1011]

配列決定する段階は、1つまたは複数の指定された多型のいずれかを選択的に増幅させることなく実施される、本発明1001の方法。

[本発明1012]

1つまたは複数のプロセッサーにおいてプログラム命令を実行し、(g)において推定された胎児DNAの画分をコンピューター読み取り可能媒体に保存された妊娠個体についての患者医療記録に自動的に記録することをさらに含む、本発明1001の方法。

20

[本発明1013]

前記患者医療記録は、研究室、臨床医のオフィス、病院、健康維持組織、保険会社または個人医療記録ウェブサイトによって保持される、本発明1012の方法。

[本発明1014]

胎児DNAの画分の推定に基づき、母性試験サンプルが採取されたヒト対象の処置を処方すること、開始することおよび/または変更することをさらに含む、本発明1001の方法。

[本発明1015]

胎児DNAの画分の推定に基づき、1つまたは複数のさらなる試験を指示することおよび/または行うことを行ふことをさらに含む、本発明1001の方法。

30

[本発明1016]

(a)(i)母性ゲノムおよび胎児ゲノムの両方のDNAを含む体液のサンプルから抽出されたDNAを受け取り、かつ(ii)1つまたは複数の指定された多型を含むDNAセグメント配列を生成する条件下で、前記抽出されたDNAを配列決定するように構成されたシーケンサーと；

(b)参照配列上の前記1つまたは複数の指定された多型に核酸配列をマッピングし、前記指定された多型の少なくとも1つについてマッピングされたDNAセグメント配列のアレル頻度を決定し、

前記少なくとも1つの指定された多型を妊娠個体の接合生殖性と前記胎児の接合生殖との組み合わせに基づいて分類し、かつ

前記妊娠個体から得られたDNA中の胎児DNAの画分を、前記アレル頻度および前記接合生殖性の組み合わせを用いて推定するよう、1つまたは複数のプロセッサーに命令するよう構成されたコンピューター装置

40

とを備える、妊娠個体の体液から得られたDNA中の胎児DNAの画分を推定するための装置。

[本発明1017]

母性ゲノムと胎児ゲノムとの両方のDNAを抽出する条件下でサンプルからDNAを抽出するためのツールをさらに備える、本発明1016の装置。

[本発明1018]

コンピューター装置は、少なくとも1つの指定された多型を、以下の組み合わせの1つに分類するよう、1つまたは複数のプロセッサーに命令するようさらに構成される、本発明1016の装置：(i)妊娠個体がホモ接合型であり、かつ胎児がホモ接合型である、(ii)妊娠

50

個体がホモ接合型であり、かつ胎児がヘテロ接合型である、(iii)妊娠個体がヘテロ接合型であり、かつ胎児がホモ接合型である、および(iv)妊娠個体がヘテロ接合型であり、かつ胎児がヘテロ接合型である。

[本発明1019]

コンピューター装置は、組み合わせ(i)または組み合わせ(iv)に分類されるいかなる多型も、検討から除くよう1つまたは複数のプロセッサーに命令するようにさらに構成される、本発明1018の装置。

[本発明1020]

コンピューター装置は、規定の閾値より高いマイナーアレル頻度を有するいかなる多型をも検討から除くよう1つまたは複数のプロセッサーに命令するようにさらに構成される、本発明1016の装置。

10

[本発明1021]

コンピューター装置は、1つまたは複数の指定された多型をフィルタリングし、規定の閾値より低いマイナーアレル頻度を有するいかなる多型をも検討から除くよう1つまたは複数のプロセッサーに命令するようにさらに構成される、本発明1016の装置。

[本発明1022]

コンピューター装置は、アレル頻度に対し、閾値を適用することによって少なくとも1つの指定された多型を分類するよう1つまたは複数のプロセッサーに命令するようにさらに構成される、本発明1016の装置。

[本発明1023]

20

コンピューター装置は、複数の多型について得られたアレル頻度データを、混合モデルに適用することによって少なくとも1つの指定された多型を分類するよう1つまたは複数のプロセッサーに命令するようにさらに構成される、本発明1016の装置。

[本発明1024]

混合モデルは、階乗モーメントを使用する、本発明1023の装置。

[本発明1025]

シーケンサーにおける配列決定のために妊娠個体の血漿から得られた無細胞DNAを抽出するための装置をさらに備える、本発明1016の装置。

[本発明1026]

30

コンピューター装置は、多型のデータベースに対してセグメントをコンピューター的にマッピングすることによって胎児を孕む個体の血液から得られたDNAセグメントをマッピングするよう1つまたは複数のプロセッサーに命令するようにさらに構成された、多型のデータベースをさらに備える、本発明1016の装置。

[本発明1027]

コンピューター装置は、DNAの胎児画分を、コンピューター読み取り可能媒体に保存された妊娠個体についての患者医療記録に自動的に記録するよう1つまたは複数のプロセッサーに命令するようにさらに構成される、本発明1016の装置。

[本発明1028]

前記患者医療記録は、研究室、臨床医のオフィス、病院、健康維持組織、保険会社または個人医療記録ウェブサイトによって保持される、本発明1027の装置。

40

[本発明1029]

妊娠個体の体液から得られたDNA中の胎児DNAの画分を推定する方法であって、以下の段階：

(a)前記妊娠個体の体液から得られたDNAのセグメントを、複数の多型配列に対してマッピングする段階であって、ここで、前記複数の多型配列を同定する条件下で、前記DNAを配列決定する、段階；

(b)前記マッピングされた核酸のアレル頻度を、前記複数の多型配列のそれぞれについて決定する段階；および

(c)前記アレル頻度を混合モデルに適用する段階であって、前記胎児を孕む個体の血液から得られたDNA中の前記胎児DNAの画分の推定を得る、段階

50

を含み、ここで、(b)～(c)は、前記決定する段階および適用する段階のためのプログラム命令の下で動く1つまたは複数のプロセッサーにおいて行われる、方法。

[本発明1030]

(c)は、複数の多型配列のそれぞれについてのアレル頻度データの階乗モーメントについての一連の等式を解くために、1つまたは複数のプロセッサーに対する命令を実行することを含む、本発明1029の方法。

[本発明1031]

(c)の前に、胎児および妊娠個体の両方においてヘテロ接合型であると同定されている多型についてのアレル頻度をコンピューター的に除くことをさらに含む、本発明1029の方法。

10

[本発明1032]

(c)の前に、胎児および妊娠個体の両方においてホモ接合型であると同定されている多型についてのアレル頻度をコンピューター的に除くことをさらに含む、本発明1029の方法。

[本発明1033]

(c)の前に、妊娠個体においてヘテロ接合型であると同定されている多型についてのアレル頻度をコンピューター的に除くことをさらに含む、本発明1029の方法。

[本発明1034]

混合モデルは、配列決定エラーを考慮する、本発明1029の方法。

[本発明1035]

妊娠個体の体液から得られるDNAは、前記妊娠個体の血漿から得られる無細胞DNAである、本発明1029の方法。

20

[本発明1036]

妊娠個体の体液から得られる核酸をマッピングする段階は、多型のデータベースに対しセグメントをマッピングすることを含む、本発明1029の方法。

[本発明1037]

妊娠個体の体液のサンプルを得ることをさらに含む、本発明1029の方法。

[本発明1038]

多型配列を含むDNAセグメント配列を生成する条件下で、核酸シーケンサーにより妊娠個体の体液由来のDNAを配列決定することをさらに含む、本発明1029の方法。

30

[本発明1039]

(a)のマッピングする段階は、複数の両アレル性多型配列を同定することを含む、本発明1029の方法。

[本発明1040]

(a)のマッピングする段階は、複数の事前指定された多型配列に対してDNAのセグメントをマッピングすることを含む、本発明1029の方法。

[本発明1041]

コンピューター読み取り可能媒体上に保存された妊娠個体についての患者医療記録において、(c)において推定されたDNAの胎児の画分を自動的に記録するためのプログラム命令を、1つまたは複数のプロセッサーにおいて実行することをさらに含む、本発明1029の方法。

40

[本発明1042]

前記患者医療記録は、研究室、臨床医のオフィス、病院、健康維持組織、保険会社または個人医療記録ウェブサイトによって保持される、本発明1041の方法。

[本発明1043]

胎児DNAの画分の推定に基づき、母性試験サンプルが採取されたヒト対象の処置を処方すること、開始することおよび/または変更することをさらに含む、本発明1029の方法。

[本発明1044]

胎児DNAの画分の推定に基づき、1つまたは複数のさらなる試験を指示することおよび/または行うことを行ふことをさらに含む、本発明1029の方法。

50

[本発明1045]

妊娠個体の体液から得られたDNA中の胎児DNAの画分を推定する方法であって、以下の段階：

- (a)前記体液のサンプルを受け取る段階；
- (b)前記サンプルから、前記体液中に存在する母性ゲノムと胎児ゲノムとの両方のDNAを抽出する条件下でDNAを抽出する段階；
- (c)前記抽出したDNAを、DNAセグメント配列を生成する条件下で、核酸シーケンサーによって配列決定する段階；
- (d)前記体液に由来する前記DNAセグメント配列を比較し、比較から、1つまたは複数の両アレル性多型を同定する段階であって、ここで、前記比較することは、核酸配列を比較し多型を同定するようプログラムされたコンピューター装置を用いて行う、段階；
- (e)前記DNAセグメント配列のアレル頻度を、前記同定された多型の少なくとも1つについて決定する段階；
- (f)前記少なくとも1つの同定された多型を、前記妊娠個体の接合生殖性と前記胎児の接合生殖性との組み合わせに基づいて分類する段階；および
- (g)(e)において決定された前記アレル頻度と(f)からの前記接合生殖性の組み合わせとを用いて、前記妊娠個体から得られたDNA中の胎児DNAの画分を推定する段階を含み、ここで、(e)～(g)は、前記決定する段階、分類する段階、および推定する段階のためのプログラム命令の下で動く1つまたは複数のプロセッサーにおいて行われる、方法。

10

20

[本発明1046]

(f)の分類する段階は、少なくとも1つの同定された多型を、以下の組み合わせのうちの1つに分類する、本発明1045の方法：(i)妊娠個体がホモ接合型であり、かつ胎児がホモ接合型である、(ii)妊娠個体がホモ接合型であり、かつ胎児がヘテロ接合型である、(iii)妊娠個体がヘテロ接合型であり、かつ胎児がホモ接合型である、および(iv)妊娠個体がヘテロ接合型であり、かつ胎児がヘテロ接合型である。

[本発明1047]

組み合わせ(i)または組み合わせ(iv)に分類されるいかなる多型も、検討から除くことをさらに含む、本発明1046の方法。

[本発明1048]

1つまたは複数の同定された多型をフィルタリングし、規定の閾値より高いマイナーアレル頻度を有するいかなる多型をも検討から除くことをさらに含む、本発明1045の方法。

30

[本発明1049]

DNAセグメント配列は、約20塩基対と約300塩基対との間の長さを有する、本発明1045の方法。

[本発明1050]

少なくとも1つの同定された多型を分類する段階は、(e)において決定されたアレル頻度に対し、閾値を適用することを含む、本発明1045の方法。

[本発明1051]

少なくとも1つの同定された多型を分類する段階は、複数の多型について得られた(e)からのアレル頻度データを、混合モデルに適用することを含む、本発明1045の方法。

40

[本発明1052]

混合モデルは、階乗モーメントを使用する、本発明1051の方法。

[本発明1053]

妊娠個体の体液から得られたDNAは、前記妊娠個体の血漿から得られた無細胞DNAである、本発明1045の方法。

[本発明1054]

(a)(i)母性ゲノムおよび胎児ゲノムの両方のDNAを含む体液のサンプルから抽出されたDNAを受け取り、かつ(ii)前記抽出されたDNAを配列決定して、DNAの配列セグメントを生成するように構成されたシーケンサーと；

50

(b) 複数の多型配列に妊娠個体の体液から得られた前記DNAの配列セグメントをマッピングし、

前記複数の多型配列のそれぞれについて、前記DNAのマッピングされた配列セグメントからアレル頻度を決定し、かつ

前記アレル頻度を混合モデルに適用し、前記胎児を孕む個体の血液から得られたDNA中の胎児DNAの画分の推定を得るよう、1つまたは複数のプロセッサーに命令するように構成されたコンピューター装置

とを備える、妊娠個体の体液から得られたDNA中の胎児DNAの画分を推定するための装置。

【本発明1055】

(a)(i) 母性ゲノムおよび胎児ゲノムの両方のDNAを含む体液のサンプルから抽出されたDNAを受け取り、かつ(ii)DNAセグメント配列を生成する条件下で、前記抽出されたDNAを配列決定するように構成されたシーケンサーと；

(b) 前記体液に由来するDNAセグメント配列を比較し、比較から、1つまたは複数の両アレル性多型を同定し；

前記DNAセグメント配列のアレル頻度を、前記同定された多型の少なくとも1つについて決定し；

前記少なくとも1つの同定された多型を、妊娠個体の接合生殖性と前記胎児の接合生殖性との組み合わせに基づいて分類し；かつ

前記アレル頻度と前記接合生殖性の組み合わせとを用いて、前記妊娠個体から得られたDNA中の胎児DNAの画分を推定するよう、1つまたは複数のプロセッサーに命令するように構成されたコンピューター装置

とを備える、妊娠個体の体液から得られたDNA中の胎児DNAの画分を推定するための装置。

本開示の態様のこれらのおよび他の特徴および利点は、添付の図面を参照して、以下でより詳細に説明される。

【図面の簡単な説明】

【0040】

【図1】図1は、所定の遺伝的位置についての胎児および母親の接合生殖性状態の分類を示すブロック図である。

【図2】図2は、本開示の態様のいくつかを実施するための手順フロー例である。

【図3】図3は、ヒトゲノムHG18とアラインメントされた Illumina GA2データの30レーンより多くの配列決定された塩基位置による推定エラーを示す (Elandをデフォルトパラメータを用いて使用する)。

【図4】図4は、ヘテロ接合生殖性の事例1～4についてのマイナーアレルカウントA対適用範囲D (エラーがないと仮定する) のプロットである。

【図5】図5は、事例3データの事例2への変換を示す。

【図6】図6は、回転後のデータを示し、ここで、D1は、事例1および事例2、3は重ならないように選択される。E1は、事例1データの99%信頼区間上限の上限を表す。

【図7】図7は、混合モデルを用いた結果と公知の胎児画分および推定胎児画分との比較を示す。

【図8】図8は、機械エラー率の公知のパラメータとしての使用が、点による上方への偏りを減らすことを示す。

【図9】図9は、機械エラー率を公知のパラメータとして使用したシミュレーションしたデータを示す。事例1および事例2エラーモデルの増大は、0.2未満の胎児画分についての点よりも低く、上方への偏りを大いに減らす。

【図10】図10は、適切に構成され（例えば、プログラムされ）るか、または設計された場合、本開示の態様についての解析装置として寄与し得るコンピューターシステムの概略図である。

【図11A】図11Aは、例において生成される第1番染色体についてマイナーアレル百分率(A/D)における異なる観察の数（頻度）の数のヒストグラムを示す。

【図11B】図11Bは、例において生成される第7番染色体についてマイナーアレル百分率

10

20

30

40

50

(A/D)における異なる観察の数(頻度)の数のヒストグラムを示す。

【図12A】図12Aは、第1番染色体に沿うアレル頻度の分布を示す。

【図12B】図12Bは、第7番染色体に沿うアレル頻度の分布を示す。

【発明を実施するための形態】

【0041】

詳細な説明

序論および概要

本開示の特定の態様は、妊娠雌性体の血液から得られたDNAを解析すること、およびこの分析を、胎児由来のDNAの画分を推定するために用いることに関する。次いで、DNAの胎児画分は、母性血液から得られたDNAの独立した解析に基づいて、胎児の別の測定基準または性質決定にあるレベルの信頼性を与えるために使用され得る。例えば、母性血液から得られた胎児DNAサンプルは、別個に解析されて、この妊娠雌性体が孕む胎児における異常を検出することができる。この別個の解析による異常判定は、母性血液から得られたDNAにおける胎児DNAの画分量に基づく信頼性レベルに統計学的にに基づく。DNA全量において相対的に低い胎児DNAの画分は、胎児DNAに基づく任意の性質決定における低い信頼性を示唆する。

10

【0042】

典型的に、必ずしもではないが、解析された母性血液中のDNAは、無細胞DNAであるが、ある態様において、細胞定着DNAであってもよい。無細胞DNAは、母性血漿から得られる。妊娠雌性体から得られた無細胞DNAカウント中の胎児DNAの量は、胎児の在胎月齢を含む種々の因子に依存して広く変動する。典型的なヒト妊娠女性について、現在、無細胞DNAの約5~20%は、胎児DNAであると信じられている。しかし、胎児画分が非常に低い(例えば、約1%以下である)ことも稀ではない。このような場合、胎児DNAの任意の別個の性質決定が、本質的に疑われ得る。他方、ある研究者らは、母性無細胞DNAサンプルが、胎児DNA画分を40%または50%の高さで有すると報告している。

20

【0043】

本明細書において記載される特定の実施において、母性DNAの胎児画分の決定は、1つまたは複数の多型を有することが公知な配列部位に関する多数のDNA配列に基づく。典型的には、必ずしもではないが、このような多型は、1スクレオチド多型(SNP)である。好適な多型の他の型としては、欠失、STR(短いタンデムリピート)、挿入、インデル(微小インデルを含む)などが挙げられる。さらなる例は、以下に示される。特定の態様において、多型部位は、以下に記載される「参照配列」において見出される。ある態様において、多型部位は、配列タグを互いにおよび/または参照配列とアラインメントした際に発見される。

30

【0044】

本開示の特定の方法は、検討中の多型部位において、胎児のDNA配列は、その母親のDNA配列と一致しない場合があるという事実を使用する。例えば、特定のSNPの部位における母性DNAは、ホモ接合型であってもよく、他方、SNPの胎児版は、ヘテロ接合型であってもよい。したがって、問題のSNPについて得られた配列サンプルの収集は、メジャーアレルを含む配列の大部分と、マイナーアレルを含む残りの画分とによる異種性と考えられる。メジャーアレルおよびマイナーアレルの相対量は、サンプル中の胎児DNAの画分によって決定される。

40

【0045】

ホモ接合型サンプルにおいて、所定のSNPまたは他の多型の両コピーは、同じアレルを含み、他方、ヘテル接合型のSNPまたは他の多型は、1コピーのメジャーアレルと1コピーのマイナーアレルとを含むことが言及される。したがって、ヘテロ接合型の個体からのみ得られたDNAは、50%のメジャーアレルと50%のマイナーアレルとを含むことが分かる。この知見は、下で概説される胎児DNAの画分を説明する際に使用され得る。まとめると、下でより完全に説明されるように、本明細書において開示される種々の方法は、母性DNAおよび胎児DNAにおいて2つのアレルしか存在しない多型のみを検討する。

50

【0046】

ある実施において、母性血液から得られたDNAは、多数回（多型の「適用範囲」と考えられる多型の特定部位に対してマッピングする読み取りの全回数、およびマイナーアレルカウントと考えられる多型についてのマイナーアレルに対してマッピングする読み取りの回数）読み取られる。マイナーアレルカウント対適用範囲の比は、種々の実施において重要である。

【0047】

本明細書において開示される特定の方法は、母性および胎児の両方からのDNAを含むDNAサンプル中の多型の4つの事例を同定し性質決定する。下の図1は、これらの4つの事例を示す。具体的には、比較的興味のない第1の事例において、母性および胎児の両方が、検討される特定の多型においてホモ接合型である。このような事例においては、問題の多型を含むDNAサンプル中の全ての配列が同じアレルを含み、母性および胎児からのDNAの相対量に関して何の情報も得ることができないと考えられる。しかし、この事例は、研究者および技術者が、検討される配列データを作成するために使用されるDNA配列決定装置の相対的エラー率についての何らかの着想を得ることを可能にするという意味で、興味深いともいえることに留意されたい。10

【0048】

解析が遭遇する第2の事例は、妊娠雌性体がホモ接合型で胎児がヘテロ接合型である多型である。この事例において、比較的小さいが、それにも関わらず有為である、検出配列の画分が、マイナーアレルを含むと考えられる。具体的には、この第2の事例において、マイナーアレルの頻度は、名目上、母性の血流における胎児DNAの画分を2で割ることによって得られる。20

【0049】

第3の事例において、検討される多型は、母性DNAにおいてヘテロ接合型であり、胎児DNAにおいてホモ接合型である。この場合、マイナーアレルの頻度は、名目上、0.5からDNAサンプルにおける胎児DNAの画分の2分の1を引くことによって得られる。

【0050】

最後に、第4の事例において、検討される多型は、母性および胎児の両方においてヘテロ接合型である。この事例において、メッジャーアレルおよびマイナーアレルの頻度は、両方とも0.5であると予測される。第1の事例のように、第4の事例は、DNAの胎児画分を決定するためには、比較的興味がない。30

【0051】

研究者、技術者またはソフトウェアが、所定の多型（この多型が4つの事例のどれに属するかが分かっている）についてサンプルにおいて胎児DNAの画分を決定することが課されている場合、検討中の多型が事例2または事例3に入るとすると、胎児DNAの画分は、直接的に推定可能である。しかし、実際には、この知見を実験せずに有することはできない。したがって、コンピューター装置は、本明細書において記載される操作を行う必要がある。

【0052】

本明細書において他所で記載される特定の態様において、1つの多型を4つの事例のうちの1つに分類するために、閾値技術が使用される。多型がこのように分類され、事例2または事例3のどちらかに在ることが見出されると、胎児画分は、推定され得る。他の態様において、この技術は、ゲノムの全てまたは一部を通して分布される多くの多型を検討する。具体的な例において説明されるように、ゲノムに亘る多くの異なるSNPが、この目的のために使用され得る。40

【0053】

特定の態様において、アレル頻度は、母性血液サンプルから得られたDNAサンプルにおいて多くの異なる多型について決定される。この多くの多型について、ある画分は接合生殖性事例1に相当し、別の画分は事例2に相当し、第3の画分は事例3に相当し、最後の画分は事例4に相当すると考えられる。これらの画分は、合計すると1の値になる。混合モデル50

もしくは関連の技術は、これらの4つのカテゴリーのそれぞれにおいて多型の1つまたは複数の統計学的特性を探るために使用され得る。具体的には、混合モデルは、妊娠雌性体の血液から得られるDNAサンプルにおいて遭遇する4つの事例それぞれについての平均および任意で分散を決定するために使用され得る。具体的な態様において、これは、検討中の多型のカウントの総数（適用範囲）に対しマイナーアレルの頻度に関連する平均および分散である。本明細書において他所で詳述されたように、これらの4つのカテゴリー、または少なくとも第2および第3のカテゴリーのそれぞれについての平均値は、母性血液から得られたDNAにおける胎児画分に直接関係する。

【0054】

混合モデルを使用する特定の実施において、多型が考慮される位置それぞれについて、1つまたは複数の階乗モーメントが計算される。例えば、階乗モーメント（または階乗モーメントの集合）は、DNA配列中で考慮される多数のSNP位置を用いて計算される。以下の等式4で示されるように、種々の階乗モーメントのそれぞれは、所定の位置についてのマイナーアレル頻度対適用範囲の比について考慮される種々のSNP位置の全てに亘る合計である。以下の等式5で示されるように、これらの階乗モーメントは、上で記載される4つの接合生殖性の事例のそれぞれに関連したパラメータにも関する。具体的には、これらは、事例それぞれについての可能性、ならびに考慮される多型の集合における4つの事例それぞれについての相対量に関する。説明されるように、この可能性は、母性血液中の無細胞DNAにおける胎児DNA画分に相関する。より完全に以下に説明されるように、充分な数のこれらの階乗モーメント（等式4において示される）を計算することにより、本方法は、全ての未知数を解くために充分な数の式を提供する。この場合の未知数は、考慮される多型の集団における4つの事例それぞれの相対量、ならびに4つの事例それぞれに関連する可能性（したがって、胎児DNA画分）であると考えられる。等式5を参照されたい。類似の結果は、以下の等式7～12において表される他の種類の混合モデルを用いて得られ得る。これらの特定の種類は、事例1および2に当てはまる多型のみを使用し、事例3および4の多型は、閾値技術によってフィルタリングされる。

10

20

【0055】

したがって、階乗モーメントは、接合生殖性の4つの事例の任意の組み合わせの可能性を同定する混合モデルの一部として使用され得る。そして、言及された通り、これらの可能性、または少なくとも第2および第3の事例についての可能性は、母性血液における全無細胞DNAにおける胎児DNAの画分に直接関係する。

30

【0056】

配列決定エラーは、解かれるべき階乗モーメント等式のシステムの複雑性を減らすために使用され得ることもまた言及される。この点において、配列決定エラーは、4つの結果（任意の所定の多型位置における4つの可能性の有る塩基のそれぞれに対応する）のうちの任意の1つを実際に有し得ることが認識されるべきである。

【0057】

特定の態様において、タグが、参照染色体またはゲノムとアラインメントされ、そして両アレル多型が同定される。これらの多型は、事前に規定されているか、さもなければアラインメントの前に同定される。これらは、単に、アラインメントの間に同定され、その後、その本明細書において記載される接合生殖性およびマイナーアレルカウントに基づいて性質決定される。この情報は、本明細書において記載されるように、ゲノム画分を推定するために、使用される。

40

【0058】

本明細書において記載される態様において使用されるタグの長さは、一般に、タグを作るために使用される配列決定方法によって決定される。この方法は、広い範囲のタグの長さに亘って強固である。ある実施において、タグは、約20～300塩基対長（または約30～100塩基対長）の間である。

【0059】

本開示の態様のいくつかを実施するための手順フロー例は、図2に示される。ここで図

50

示されるように、この手順は201にてDNA（無細胞DNAまたは細胞定着DNA）の母性血液または他の体液からの収集によって始まる。このDNAの多数の配列は、参照配列における1つまたは複数の多型に対しマッピングされる。このマッピングは、多型のそれについてのアレル頻度を提供する。ブロック203を参照されたい。

【0060】

より具体的には、ブロック203での手順は、収集したDNAの配列を多数の多型の位置において読み取ることを含み得る。いくつかの場合において、これらは、倍数性決定または胎児DNAについて行われる他の決定についての手順の一部として作成され得る。したがって、ある態様において、別個の配列を作成する必要はない。読み取り配列は、参照配列に対してアラインメントされて、BLASTまたは類似のツールを用いてアラインメントが最大化される。

10

【0061】

参照配列は、多型のデータベースとして提供され得る。ある場合において、これは、全ての多型定義のコンビナトリアルから作成されるアレル探索参照セット（例えば、多型がSNPである場合、全てのSNP配列）である。例えば、添付書類を参照されたい。特定の例において、配列は、約100～150塩基対長である。

【0062】

図2に戻ると、本方法は、ブロック203の操作において考慮される1つまたは複数の多型についての母性/胎児接合生殖性組み合わせを決定する。ブロック205を参照されたい。混合モデルは、特定の態様において、この目的のために使用され得る。言及されたように、組み合わせは、以下の通りである：M&Fホモ接合型、Mホモ接合型およびFヘテロ接合型、Mヘテロ接合型およびFホモ接合型、およびM&Fヘテロ接合型。

20

【0063】

最後に、ブロック207で示されるように、本方法は、1つまたは複数の多型における接合生殖性事例アレル頻度の組み合わせを、母性サンプル由来のDNAにおける胎児成分の画分量を推定するために使用する。

【0064】

定義

以下の議論は、特定の局面および本開示の態様の利益を理解することの助けとして提供される。

30

【0065】

用語「読み取り」は、核酸サンプルの一部からの配列読み取りをいう。典型的には、必ずしもではないが、読み取りは、サンプル中の連続する塩基対の短い配列を表す。読み取りは、サンプル部分の塩基対配列によって（ATCGで）記号的に表され得る。これは、メモリデバイスに保存され、適切に処理されて、参照配列に一致するかまたは他の基準をみたすか否かを決定し得る。読み取りは、配列決定装置から直接的に、またはサンプルに関する保存された配列情報から間接的に、得られ得る。

【0066】

用語「タグ」もまた、核酸サンプルからの短い配列をいう。典型的には、タグは、ゲノム中の配列の位置などの関連する情報を含む。いくつかの目的のために、読み取りおよびタグの用語は、本明細書において相互交換可能である。しかし、典型的には、配列読み取りは、参照配列に対してアラインメントされ、参照ゲノム上の1部位のみにおいてマッピングする読み取りは、タグと呼ばれる。「セグメント配列」は、時々、本明細書において、「タグ」と相互交換可能に使用される。

40

【0067】

しばしば、本明細書において「読み取り」は、36塩基対長（36マー）である核酸の配列として記載される。無論、本開示の態様は、この大きさに限定されない。より小さい読み取りおよびより大きい読み取りが、多くの適用において好適である。ヒトゲノムに対する読み取りをアラインメントする適用について、30塩基対またはそれ以上の大きさの読み取りは、一般に、サンプルを1つの染色体に対してマッピングするために十分であると考え

50

られる。より大きなタグ/読み取りが、いくつかの適用について好適である。全ゲノム配列決定について、1000塩基対またはそれより大きな桁の読み取りが、使用され得る。特定の態様において、読み取りは、約20塩基対と10,000塩基対との間、または約30塩基対と100塩基対との間、または約30塩基対と50塩基対との間の長さを有し得る。

【0068】

「参照配列」は、生物学的分子、しばしば核酸（染色体すなわちゲノム）の配列である。典型的には、多数の読み取りは、所定の参照配列のメンバーである。特定の態様において、読み取りまたはタグは、参照配列と比較されて、この参照配列がこの読み取り配列を含むか否かを決定する。この手順は時々、アラインメントと呼ばれる。

【0069】

種々の態様において、参照配列は、これに対してアラインメントされた読み取りよりも有意に大きい。例えば、これは、少なくとも約100倍大きいか、または少なくとも約1000倍大きいか、または少なくとも約10,000倍大きいか、または少なくとも約 10^5 倍大きいか、または少なくとも約 10^6 倍大きいか、または少なくとも約 10^7 倍大きくてよい。

【0070】

一例において、参照配列は、完全長のヒトゲノムの参照配列である。このような配列は、ゲノム参照配列と呼ばれ得る。別の例において、参照配列は、特定のヒト染色体、例えば第13番染色体に限定される。このような配列は、染色体参照配列と呼ばれ得る。参照配列の他の例としては、他の種のゲノムならびに任意の種の染色体、部分染色体領域（例えば、ストランド）などが挙げられる。

【0071】

種々の態様において、参照配列は、多数の個体に由来する共通配列であるか、または他の組み合わせである。しかし、特定の適用において、参照配列は、特定の個体から得られてもよい。

【0072】

用語「アラインメント」は、読み取りまたはタグを参照配列と比較し、それによって参照配列が読み取り配列を含むか否かを決定する手順をいう。参照配列が読み取りを含む場合、この読み取りは、参照配列に対してマッピングされてもよく、または、特定の態様において、参照配列における特定の位置にマッピングされてもよい。いくつかの場合において、アラインメントは、単に、読み取りが特定の参照配列のメンバーであるか否か（すなわち、この読み取りが参照配列中に存在するか否か）を告げる。例えば、ヒト第13番染色体についての参照配列に対する読み取りのアラインメントは、第13番染色体についての参照配列においてこの読み取りが存在するか否かを告げる。この情報を提供するツールは、セットメンバーシップテスターと呼ばれ得る。いくつかの場合、アラインメントは、さらに、参照配列において読み取りまたはタグがマッピングされる位置を示す。例えば、参照配列が全ヒトゲノム配列である場合、アラインメントは、読み取りが第13番染色体上に存在することを示し得、そしてこの読み取りが、第13番染色体の特定のストランド上にあることをさらに示し得る。

【0073】

「部位」は、読み取りまたはタグに対応する、参照配列における特有の位置である。特定の態様において、これは、染色体のアイデンティティ（例えば、第13番染色体）、染色体のストランド、および染色体における正確な位置を特定する。

【0074】

「多型部位」は、ヌクレオチド配列ダイバージェンスが起こる座である。座は、1塩基対の小ささであってもよい。例示的マーカーは、少なくとも2つのアレルを有し、各出現頻度は、1%より高く、より典型的には、選択された集団の10%または20%よりも高い。多型部位は、1塩基対の小ささであってもよい。用語「多型座」および「多型部位」は、本明細書において、相互交換可能に使用される。

【0075】

本明細書において「多型配列」は、核酸配列、例えば、1つまたは複数の多型部位、例

10

20

30

40

50

えば、1つのSNPまたはタンデムのSNPを含むDNA配列をいう。本技術にしたがう多型配列は、胎児核酸と母性核酸との混合物を含む母性サンプルにおいて、母性アレルと非母性アレルとの間で特異的に区別するために使用され得る。

【 0 0 7 6 】

詳細な態様

典型的には、本明細書において記載されるプロセスは、1つまたは複数の多型に亘り、サンプリングされたDNAに関連する参照配列を使用する。参照配列は、例えば、ヒトのゲノム、染色体または染色体の一領域であってもよい。多型の1つまたは複数は、胎児DNA画分を推定する目的で、指定され得る。胎児画分を決定する際の使用のために指定される多型は、既知である多型である。例えば、既知のSTRにおける参照、事実および配列情報の包括的列挙および関連の集団データは、STRBaseにおいて編纂され、これは、ibm4.carb.net.gov:8800/dna/home.htmにてワールドワイドウェブを介してアクセスされ得る。よく使用されるSTR座についてのGenBank（登録商標）（<http://www2.ncbi.nlm.nih.gov/cgi-bin/genbank>）からの配列情報もまた、STRBaseを通してアクセス可能である。アクセスされ得る既知のSNPの情報は、公にアクセス可能なデータベース（ワールドワイドウェブアドレスwi.mit.eduにおけるヒトSNPデータベース、ワールドワイドウェブアドレスncbi.nlm.nih.govにおけるNCBI dbSNPホームページ、ワールドワイドウェブアドレスlifesciences.perkinelmer.com、ワールドワイドウェブアドレスappliedbiosystems.comにおけるLife Technologies（商標）によるApplied Biosystems（Carlsbad, CA）、ワールドワイドウェブアドレスcelera.comにおけるCelera Human SNPデータベース、ワールドワイドウェブアドレスgan.iarc.frにおけるGenome Analysis Group（GAN）のSNPデータベースを含むが、これらに限定されない）から入手可能である。1つの態様において、胎児画分の決定のために指定されたSNPは、Pakstisら（Pakstisら、Hum Genet 127: 315-324 [2010年]）によって記載される92個体同定SNP（IISNP）（集団の間で頻度において非常に小さな変化（ $F_{st} < 0.06$ ）を有すること、および0.4以上の平均ヘテロ接合生殖性を有する世界において、非常に情報価値のあることが示されている）の群から選択される。本発明の方法に含まれるSNPとしては、連結SNPと非連結SNPとが挙げられる。好適なタンデムSNP配列を指定するために、International HapMap Consortiumデータベースが、探索され得る（The International HapMap Project、Nature 426: 789-796 [2003年]）。このデータベースは、hapmap.orgにおいて、ワールドワイドウェブにおいて利用可能である。

【 0 0 7 7 】

このように使用される多型は、胎児DNA画分を決定するために指定された既知多型のパネルであってもよく、または、他の目的、例えば染色体に対するサンプルDNAタグのマッピングなどのための母性DNAの解析において偶然に見出されていてもよい。

【 0 0 7 8 】

特定の態様において、本方法は、ゲノム、例えば胎児および母性の無細胞DNAを含む母性サンプルの混合物を用いてサンプル中のDNAを配列決定し、参照ゲノム上の既知の多型部位を含む配列に対しマッピングする複数の配列タグを提供し、既知部位においてマッピングされたこのタグを用いて胎児画分を決定することを、含む（下で詳細に記載される）。あるいは、DNAの配列決定後、配列決定技術（例えば、NGS）によって得られた配列タグは、参照ゲノム、例えばhg19に対してマッピングされ、および偶然生じた（すなわち、既知ではない）多型の部位への配列タグマッピングが、胎児画分を決定するために使用される。

【 0 0 7 9 】

既知多型部位に対して配列タグがマッピングされる参照配列は、公開された参照ゲノムであってもよく、または人工的データベースもしくは検討中の多型についての他の予め規定された配列の集合であってもよい。データベース配列のそれぞれは、多型に関連した1つまたは複数のスクレオチドに及ぶ。一例として、下で「添付書類1」において提示される多型配列のリストを参照されたい。

【 0 0 8 0 】

10

20

30

40

50

種々の態様において、胎児DNA画分を推定するために使用される多型の数は、少なくとも2つの多型、より詳細には、それぞれ少なくとも約10の多型、より好ましくは、それぞれ少なくとも約100の多型である。

【0081】

1つの例において、SNP適用範囲およびアレル頻度は、SNP定義のコンビナトリアル展開から構築された参照ゲノムに対して作られた配列をアラインメントすることによって決定される。単位複製配列データベースは、例えば、少なくとも約50塩基の隣接配列によって囲まれた両アレル変動性情報を含む。例えば、変動性情報ストリング「[g/c]」（交替制のアレル「g」と「c」を表す）を有する単位複製配列は、以下のようにあってもよい：

atcg.....accg[g/c]ccgt....

10

【0082】

いくつかの場合において、単位複製配列データベースおよび作られた配列を入力する手順、およびSNP/アレルカウントの出力は、以下の通りである。

1. SNP定義のコンビナトリアル伸長からアレル探索参照セットを作成する。単位複製配列データベースにおける各配列について、変動性情報ストリングにおける各アレルについて、アレルによって置き換わる変動性情報ストリングによってアレル配列を作成する。

a. 例えば、上記の単位複製配列の例を検討すると、2つの配列が作成される：1)atcg....accgGccgt.... および2)atcg.....accgCccgt....

b. 全長アレル探索参照セットの例は、アレル探索データベース配列表において見出され得る。

20

2. アレル探索参照セットに対して配列をマッピングし、この探索セットにおいて1つの配列のみに一致するマッピングを取っておく。

3. アレルカウントを、そのアレル配列に一致する配列の数を計数することにより、決定する。

【0083】

本明細書において開示される方法は、「正常」妊娠、すなわち、母親が1胎児のみを孕む妊娠を想定しており、双子、三つ子などは想定していない。当業者は、非正常妊娠を考慮する改変（特に胎児の数が分かっているもの）を、高く評価する。

【0084】

示される通り、胎児画分を決定する際、この方法は、母性血液からのサンプル中のDNAを配列決定し、検討中の多型の各配列をマッピングする配列タグを数える。各多型について、方法は、マッピングされる読み取りの総数（適用範囲）と各アレルに伴う配列タグの数（アレルカウント）とを符号させる。単純な例において、5の適用範囲を有する多型は、アレルBの3つの読み取りおよびアレルAの2つの読み取りを有し得る。この例において、アレルAは、マイナーアレルとみなされ、アレルBは、メジャーアレルとみなされる。

30

【0085】

いくつかの態様において、この操作は、非常に迅速な配列決定ツール、例えば、大規模パラレルDNA配列決定ツールを使用する。このようなツールの例は、下でより詳細に記載される。いくつかの場合、数千数百万のタグ配列が、1つのサンプルについて読み取られる。好ましくは、配列決定は、配列決定されたDNAの、検討中の多型を有する特定の事前に規定された配列への、素早く且つ直接的な割り当てを可能にする様式で行われる。一般に、30塩基対またはそれより大きな大きさのタグにおいて、この目的のために、十分な情報がある。この大きさのタグは、対象となる配列に対して明確にマッピングされ得る。特定の態様において、プロセスにおいて使用されるタグ配列は、36塩基対長である。

40

【0086】

タグは、参照ゲノムまたはアレル配列データベース中の配列（例えば、以前に言及された添付書類1を参照されたい）に対してマッピングされ、マッピングされたタグの数が、決定される。これは、検討中の各多型についての適用範囲およびマイナーアレルカウントの両方を提供する。いくつかの場合、これは、各タグを23のヒト染色体のうちの1つにマッピングし、同時に染色体あたりのマッピングされたタグの数を決定することによって、

50

なされ得る。

【0087】

言及されたように、適用範囲は、参照配列における所定の多型に対しマッピングされる読み取り配列の総数である。読み取り配列の総数におけるアレルカウントは、1つのアレルを有する多型に対しマッピングされる。全てのアレルカウントの合計は、適用範囲と同じでなければならない。最も高いカウントを有するアレルは、メジャーアレルであり、最も低いカウントを有するアレルは、マイナーアレルである。特定の態様において、胎児DNA画分を推定するために必要とされる唯一の情報は、それぞれ複数の多型についての適用範囲およびマイナーアレルカウントである。いくつかの態様において、DNA配列決定装置の塩基コーリングエラー率もまた、使用される。

10

【0088】

本明細書において開示される特定の方法の数学的または記号的な支持を検討することが、有用である。言及された通り、種々の例において、母性血液から作成された配列は、参照ゲノムまたは他の核酸配列とアラインメントされる（同一の塩基が最大化されるように重ね合わせられる）。ゲノム位置jおよび参照に対してアラインメントされる配列のセットの場合、アラインメントされた配列の中の4つのDNA塩基（「a」、「t」、「g」、および「c」、「アレル」とも呼ばれる）のそれぞれの出現率の数は、それぞれ、 $w(j,1)$ 、 $w(j,2)$ 、 $w(j,3)$ および $w(j,4)$ である。この議論の目的のために、一般性を失わずに、全ての変動性は、両アレル性であるとみなしえる。したがって、以下の表記が使用され得る：

ゲノム位置jにおけるメジャーアレルカウントは、

20

$$B \equiv B_i \equiv \{b_j\} \equiv w_{j,i}^{(1)} = \max_{i \in \{1,2,3,4\}} \{w_{j,i}\}$$

位置jにおけるカウントの一次統計値（メジャーアレルbが、argmaxに相当する。下付き文字は、複数のSNPが検討される場合に使用される）であり、

位置jにおけるマイナーアレルカウントは、

$$A \equiv A_i \equiv \{a_j\} = w_{j,i}^{(2)}$$

位置jにおけるカウントの二次統計値（すなわち、二番目に高いアレルカウント）であり、

位置jにおける適用範囲は、

30

$$D \equiv D_j = \{d_i\} = A_j + B_j$$

であり、そして

配列決定機エラー率は、eと表される。

【0089】

文脈が明らかである場合、簡便にするために、表記は、相互交換可能に使用される；例えば、A、Aiまたは{ai}は、マイナーアレルもしくはマイナーアレルカウントについて相互交換可能に使用され得る。下付き文字は、複数のSNPが検討されるか否かに依存して、使用されてもされなくてもよい。（SNPは、例示の目的でのみ使用される。他の種類の多型が、本明細書において他所で議論されるように使用され得る。）

40

【0090】

図1において、多型接合生殖性の4つの状態についての基礎が、記載される。説明されるように、母親は、所定の多型においてホモ接合型であってもまたはヘテロ接合型であってもよい。同様に、子は、同じ位置においてヘテロ接合型であってもホモ接合型であってもよい。説明されるように、事例1および2は、母親がホモ接合型である多型事例である。子および母親が両方ともホモ接合型である場合、この多型は、事例1の多型である。上で示されるように、この状況は、典型的に、とりたてて興味深くない。母親がホモ接合型であり、子がヘテロ接合型である場合、胎児画分fは、マイナーアレル対適用範囲の比を2倍することによって、名目上得られる。多型事例において、母親がヘテロ接合型であり、子がホモ接合型である場合（図1の事例3）、胎児画分は、名目上、マイナーアレル対適用範囲

50

の比の2倍を1から減算したものである。最後に、母親および胎児の両方がヘテロ接合型である場合の事例において、マイナーアレル画分は、エラーを除いては、常に0.5である。胎児画分は、事例4に当てはまる多型について得られ得ない。

【0091】

4つの事例は、ここで、さらに詳述される。

【0092】

事例1: 母親および子がホモ接合型

この場合、配列決定エラーまたは夾雜物混入を除いては、相違は観察されない。

$$E(\text{最小アレル頻度})=E(A)=0$$

実際に、A～は低npについてポアソン分布によって十分に近似する二項分布（として分布する）。二項もしくはポアソンの分布率パラメータは、配列エラー率eおよび適用範囲Dに関連する。図3は、ヒト参照ゲノムに対してアライメントされ作成された36マー配列の不一致頻度を示す。

この事例は、胎児画分に関して情報を含まない。

【0093】

図3は、ヒトゲノムHG18に対してElandをデフォルトパラメータで用いてアライメントされたIllumina GA2データの30レーンに亘る配列決定された塩基位置によるエラー推定を示す。

【0094】

事例2: 母親がホモ接合型および子がヘテロ接合型

この事例において、小さな胎児画分(f)について、観察されるアレル頻度は、きわめて異なっている。メジャーアレルは、通常、マイナーアレルの数倍多い頻度で出現する。

エラーを除き、1つのSNP位置(D,A), $E(A)=Df/2$ が与えられ、fについての不变推定値は、 $2A/D$ である

エラーを除き、A～二項式($f/2, D$)。平均 $Df/2$ 、分散 $(1-f/2)Df/2$ 。

[$D>15$ である場合、およそ正規分布]

【0095】

事例3: 母親がヘテロ接合型および子がホモ接合型

この事例において、メジャーアレルおよびマイナーアレルについて観察された頻度は、近接しており、 A/D は、0.5のすぐ下である。

エラーを除き、 $E(A)=D(1-f)/2$ 、および $E(1-(2A/D))=f$

エラーを除き、A～二項式($(1-f)/2, D$)。平均 $D((1-f)/2)$ 、分散 $D/4(1-f^2)$ 。

【0096】

事例4: 母親がヘテロ接合型および子がヘテロ接合型

エラーを除き、これには2つの下位事例があることに留意されたい。

事例4.1: 父親由来のアレルが、母親由来のアレルとは異なる

これは、 $E(A)=Df/2$ のマイナーアレルである第3のアレルを導入する。これらの場合は、fについての推定における影響は有さない。なぜなら、配列の単位複製配列に対する割り当てのための手順は、参照SNPが両アレル性である場合に、これらの事例を閉め出すからである。

事例4.2: 父親由来のアレルが、母親のアレルの1つと一致する

この事例において、エラーを除き、2つのアレルは、1:1の比率で現れ、したがって、この事例は、胎児画分推定のために有用ではない。

エラーを除き、 $E(A)=0.5$ 、およびA～二項式($0.5, D$)、0.5で切り捨てる。

【0097】

図4は、ヘテロ接合生殖性事例1から4までについての、マイナーアレルカウントA対適用範囲D（エラーなしと仮定する）のプロットを示す。

【0098】

種々の態様において、本方法は、1つまたは複数のSNP（または他の多型）におけるアレ

10

20

30

40

50

ル頻度の解析を広く考慮し、事例2および/または事例3のいずれかである多型を分類する。このアレル頻度を、この分類と組み合わせて使用し、本方法は、胎児画分を推定できる。

【 0 0 9 9 】

いくつかの場合、個々のSNP位置についての所定のマイナーアレルカウントAおよび適用範囲D、言い換えると一点(D,A)は、一点推定を行う方法を可能にする。例えば、特定の方法は、アレルカウント(D,A)を有するSNPを、1つの事例に分類し、以下のような胎児画分推定を引き出す：

ES1.1 事例決定のための単純閾

個々の位置 (SNP) について、

1. $2A/D < e$ のような決定閾数または二項式(e, D)もしくはポアソン(De)の規定棄却値により、事例1に決定する。また、本発明の範囲内で、代替の分布をも使用し得る。胎児画分(f)は推定されない。

2. $2A/D > (0.5 - e)$ であるか、または二項式の何らかの棄却値($0.5, D$) (または他の好適な近似分布)である場合、事例4に決定する。 f の推定のために、この位置を使用しない。

3. さもなければ、 $2A/D < 0.25$ (または何らかの他の手動設定もしくは自動の推定閾)である場合、事例2に決定する。胎児画分 f は、 $2A/D$ と推定される。

4. さもなければ、事例3である。胎児画分推定 $f = (1 - 2A/D)$ を使用せよ。

【 0 1 0 0 】

正確性は、いくつかのSNPからのアレルカウント情報を組み合わせて胎児画分を推定することによって、得られ得る。

【 0 1 0 1 】

方法EM1: 多数のSNPを、平均化によって組み合わせる。

平均値、中央値、他の中央測定値(例えば、チューキーバイウェイト、M推定量など)を取る。加重平均もまた、使用され得る。重量がいかに規定されるかの例は、下のEM2.4を参照されたい。さらに、ロバスト中央測定値が、使用され得る。

【 0 1 0 2 】

方法EM2変換による事例2および事例3からの同時推定

f がX%よりも低い場合について、事例3の点(D,A)を、事例2の点と同時に起こるように変換し得る。この線から、共通の傾きが原点を通る回帰を介して計算され得る(図5を参照されたい)。

【 0 1 0 3 】

変換に基づく方法の1つの理論上の欠点は、事例2および3の二項分布が、異なる形状を有することである。典型的な胎児画分レベル(<10%)における事例2のデータは、ポアソンに近い右に向かって傾く分布を有し、事例3は、正規に近い分布を有する。

【 0 1 0 4 】

図5は、事例3データの事例2への変換を示す。ここで、単回帰は、両事例から同時に f を推定可能である。

【 0 1 0 5 】

EM2.3を計算するための方法:

工程1: 事例4データを捨てる

各データ点(D,A)について、 $A > (0.5D - T1)$ である場合、さらなる解析から(D,A)を除外する。 $T1(D, A)$ は、実数値関数である。

工程2: 事例3データを変換する

図6を参照されたい。4ではない各データ点(D,A)について、 $A > T2^*D$ である場合、点を新たな座標(D1,A1)に変換する。 $T2(D, A)$ は、実数値関数である。

$$= 2A/D$$

$$A1 = \underline{0.5D - A}$$

$$D1 = D$$

10

20

30

40

50

工程3:事例1データからの汚染を減らすため、閾値DTを確立する。

実数値関数であるT2(D,A)を下回る全てのデータ点は、無視する。

工程4:回帰は、残った変換済み事例2および3のデータについて推定する。

原点を通る回帰を、残った点に適用する。胎児画分推定は、回帰などの傾斜の倍である。

【0106】

事例2および3のデータの同じ一致を達成するために構築可能である変換の多くのクラスが存在することに、留意されたい。例としては、三角法、変換または回転行列の使用が挙げられる。これらの偏差は、本開示の範囲内に含まれることが企図される。さらに、多くのクラスの回帰(L2,L1, ...)または最適化が、使用され得る。最適化アルゴリズムの交換は、些細な変更であり、本開示の範囲内に包含される。

10

【0107】

図6は、回転後のデータを示す。D1の選択により、事例1および事例2ならびに3は、重ならない。E1は、事例1データの99%信頼区間上限の上限を示す。

【0108】

方法EM3加重最小二乗法

EM2.3からの回帰方法は、換算されたデータ点の全てが等しい分散を有すると仮定する。異なるデータソース、およびなお、同じヘテロ接合生殖性パターンからの点の不等分散性を考慮すると、より正確である。

工程1~3は、EM2.3と同じである。

20

工程4:回帰

EM2.3からの回帰において、事例2データからの点は、分散v2(f,D)=[0.5*Df-0.25*Df^2]を有し、事例3データからの点は、分散v3(f,D)=[0.25D(1-f^2)]を有する。EM2.3のように、各点に異なる加重wを与えると仮定し、本発明者らは最小化をもとめる。

$$Q = \sum_{i=1}^n w_i (a_i - s d_i)^2$$

等式1

一次導関数を0に設定し、sを解く：

$$\frac{\partial Q}{\partial s} = \sum_{i=1}^n 2w_i (d_i - s a_i)(-a_i) = 0$$

30

$$\sum_{i=1}^n s a_i^2 - \sum_{i=1}^n 2w_i a_i x_i = 0$$

および

$$s = \frac{\sum_{i=1}^n 2w_i d_i a_i}{\sum_{i=1}^n a_i^2}$$

40

式中、diは、SNPiの適用範囲であり、aiは、SNPiのマイナーアレルカウントである
(事例3については変換されている)。

等式2

この方法は、各点の分散の逆数(適切に、v2(2A/D,D)またはv3(2A/D,D)と推定される)に加重する。胎児画分推定は、 2^*s である。

【0109】

特定の態様において、混合モデルは、多型の集合を2以上の接合生殖性事例に分類し、同時に、これらの事例のそれぞれについての平均アレル頻度から胎児DNA画分を推定するために使用され得る。一般に、混合モデルは、データの特定の集合が、異なる種類のデー

50

タの混合（それぞれそれ自体の予測分布（例えば、正規分布）を有する）に作成されることを仮定する。この手順は、データのそれぞれの種類についての、平均値および他の特性の可能性を見出すことを企図する。本明細書において開示される態様において、検討中の多型についてのマイナーアレル頻度データを作成する、4つまでの異なるデータ種類（接合生殖性事例）が存在する。

【0110】

混合モデルの1つの実施は、以下の節に提示される。この態様において、マイナーアレル頻度Aは、等式3で示される4つの項の合計である。項のそれぞれは、4つの接合生殖性事例のうちの1つと対応する。各項は、多型画分 α とマイナーアレル頻度の二項分布との積である。 α は、4つの事例の各々に当てはまる多型の画分である。それぞれの二項分布は、関連の可能性pおよび適用範囲dを有する。例えば、事例2についてのマイナーアレル可能性は、 $f/2$ で与えられる。

10

【0111】

本開示の態様は、検討中のアレル頻度データについて「階乗モーメント」を使用する。周知であるように、分布の平均は、一次モーメントである。これは、マイナーアレル頻度の予測値である。分散は、二次モーメントである。これは、二乗されたアレル頻度の予測値から計算される。

【0112】

全ての多型に亘るアレル頻度データは、等式4において示されるように、階乗モーメント（一次階乗モーメント、二次階乗モーメントなど）を計算するために使用され得る。これらの等式によって示されるように、階乗モーメントは i （データセット内の個々の多型）に被さる項の和であり、ここで、データセットにおいて、n個のこののような多型が存在する。合計される項は、マイナーアレルカウントの関数、 a_i および適用範囲 d_i である。

20

【0113】

通常、階乗モーメントは、等式5において説明されるような p_i および a_i の値の関係を有する。可能性 p_i からは、胎児画分 f を決定することができる。例えば、 $p_2=f/2$ であり、 $p_3=1-f/2$ である。したがって、信頼性の有る論理が、検討中の複数の多型に亘り、マイナーアレル画分について、階乗モーメント式に対して未知の p_i および a_i に関する等式系を解くことができる。無論、混合モデルを解くために他の技術が、本発明の範囲内で存在する。

30

【0114】

本明細書において開示される混合モデル態様の数学的または記号的な支持をさらに検討することが、有用である。上に記載された4つのヘテロ接合生殖性事例は、点 (a_i, d_i) における a_i の分布についての以下の二項式混合モデルを示唆する：

$$A = \{a_i\} \sim {}_1\text{Bin}(p_1, d_1) + {}_2\text{Bin}(p_2, d_1) + {}_3\text{Bin}(p_3, d_1) + {}_4\text{Bin}(p_4, d_1)$$

式中、

$$1 = {}_1 + {}_2 + {}_3 + {}_4$$

$$m=4$$

等式3

【0115】

p_i 対 胎児画分 および配列決定エラー率に関係する種々のモデルが、下に記載される。パラメータ p_i は、特定の集団パラメータに関係し、これらの値を「フロートする」ようにさせる能力は、これらの方法に、民族性および親の子孫などの因子に関してさらなる強固さを与える。

40

【0116】

種々のヘテロ接合生殖性事例について、上の等式は、胎児画分について解くことができる。恐らく、胎児画分について解くための最も簡単な方法は、混合パラメータが観察されたデータから容易に推定され得るモーメントに関して表され得る階乗モーメントの方法によるものである。

【0117】

50

所定のn個のSNP位置について、階乗モーメントは、以下のように規定される：

$$F_1 = \frac{1}{n} \sum_{i=1}^n \frac{a_i}{d_i}$$

$$F_2 = \frac{1}{n} \sum_{i=1}^n \frac{a_i(a_i-1)}{d_i(d_i-1)}$$

...

$$F_j = \frac{1}{n} \sum_{i=1}^n \frac{a_i(a_i-1)\cdots(a_i-j+1)}{d_i(d_i-1)(d_i-j+1)}$$

10

等式4

階乗モーメントは、以下により、 $\{\alpha_i, p_i\}$ に関連し得る：

$$F_1 \approx \sum_{i=1}^m \alpha_i p_i^1$$

$$F_2 \approx \sum_{i=1}^m \alpha_i p_i^2$$

...

$$F_j \approx \sum_{i=1}^m \alpha_i p_i^j$$

20

等式5

【0118】

解は、 $n > 2^*$ （推定されるパラメータの数）である場合の等式5の関連から引き出される等式系における $\{\alpha_i, p_i\}$ について解くことによって決定され得る。明らかに、この問題は、より多くの $\{\alpha_i, p_i\}$ の推定が必要とされるほどより高くなるgのために、より数学的に難しくなる。

30

【0119】

典型的に、低い胎児画分において、単純閾によって事例1と2との間（または事例3と4との間）のデータを正確に区別することはできない。幸運にも、還元事例モデルの使用について、点 $(2A/D)=T$ の区別により、事例1/2のデータは、事例3/4のデータから容易に分けられる。 $T=0.5$ の使用は、十分に働くことが見出されている。

【0120】

等式4および5を使用する混合モデルの方法は、全ての多型についてのデータを使用するが、配列決定エラーについて別個に考慮しないことに留意されたい。第1および第2の事例についてのデータを第3および第4の事例についてのデータと分ける適切な方法は、配列決定エラーを考慮に入れることができる。

40

【0121】

さらなる例において、混合モデルに対して提供されるデータセットは、事例1および事例2の多型についてのデータのみを含む。これらは、母親がホモ接合型である多型である。閾値技術が、事例3および事例4の多型を除去するために使用され得る。例えば、特定の閾値を超えるマイナーアレル頻度を有する多型が、混合モデルを使用する前に除かれる。適切にフィルタリングされたデータおよび等式7および8に還元された階乗モーメントを用い、等式9において示されるように胎児画分fを計算することができる。等式7は、混合モデルの本実施についての等式3の書き換えであることに留意されたい。また、この特定の例において、機械読み取りに関連する配列決定エラーは、分からぬことにも留意された

50

い。結果として、この等式は、エラーeについて別個に解かなければならない。

【0122】

図7は、この混合モデルおよび既知の胎児画分(x軸)および推定された胎児画分を使用する結果の比較を示す。混合モデルが完全に胎児画分を予測している場合、プロットされた結果は、破線を辿ると考えられる。それでも、推定された画分は、特に、データのほとんどは混合モデルを適用する前に除かれたことを考えると、非常に良好である。

【0123】

さらに詳述すると、いくつかの他の方法が、等式3からのモデルのパラメータ推定のために使用され得る。いくつかの場合において、導関数を二乗統計学の0に設定することによって、扱いやすい解が、見出され得る。直接微分によって容易な解が見出されない場合、二項式PDFのテイラーシリーズ拡大または他の近似多項式が、有効であり得る。最小二乗推定法が、有効であることが周知である。

【0124】

$$\chi^2(\alpha_i, p_i) = \sum_{i=1}^n \frac{\left(P_i - \sum \alpha_i \text{二項式}(p_i, di) \right)^2}{\text{二項式}(n, p)}$$

等式 6

式中、Piは、カウントiの点の数である。Le Camからの代替法[「On the Asymptotic Theory of Estimation and Testing Hypotheses」Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 第1巻、Berkeley CA: University of CA Press, 1956年, pp. 129-156]は、ラルフ・ニュートン反復の公算関数を使用する。等式5からのモーメント解の方法は、反復の開始点として使用され得る。

【0125】

別の適用の下で、近似 分布の混合において操作される期待値最大化方法を含む混合モデルを解く方法が、議論される。

【0126】

モデル事例(1+2)、配列決定エラー未知

ヘテロ接合生殖性1および2のみを説明する還元モデルを考慮されたい。この事例において、混合分布は、以下のように記され得る:

$$A = \{a_i\} \sim \alpha_1 \text{Bin}(e, d_i) + \alpha_2 \text{Bin}(f/2, d_i)$$

式中、

$$\alpha_1 = \alpha_2 = 2$$

$$m = 4$$

等式7

【0127】

および以下の系:

$$F_1 = \alpha_1 e + (1 - \alpha_1) (f/2)$$

$$F_2 = \alpha_1 e^2 + (1 - \alpha_1) (f/2)^2$$

$$F_3 = \alpha_1 e^3 + (1 - \alpha_1) (f/2)^3$$

等式8

は、e(配列決定エラー率)、(事例1点の比)、およびf(胎児画分)について解かれ
る。ここで、Fiは、上の等式4として規定される。胎児画分についての閉じた式の解は、
以下の実数解であるように選択される。

$$F \approx \frac{(F_1 - 1)F_2 \pm \sqrt{F_2} \sqrt{4F_1^3 + F_2 - 3F_1(2 + F_1)F_2 + 4F_2^2}}{2(F_1^2 - F_2)}$$

等式 9

10

20

30

40

50

これは、0と1との間である。

【0128】

エスティメータの性能を測定するために、ハーディ・ワインベルグ平衡点 (ai, di) を、{1%、3%、5%、10%、15%、20%、および25%}に設計した胎児画分および1%で一定の配列決定エラー率によって構築した。1%のエラー率は、本発明者らが使用する配列決定機およびプロトコールについて現在認められている率であり、上の表3で示されたIllumina Genome analyzer IIデータのグラフと一致する。等式9はデータに適用され、上向きの4点バイアスを例外として、「既知の」胎児画分についての一般的な一致が見出される。興味深いことに、配列決定エラー率 e は、1%のすぐ上であると推定される。

【0129】

次の混合モデル例において、閾値または他のフィルタリング技術が、事例3および4に当てはまる多型についてのデータを除くために再び使用される。しかし、この場合において、配列決定エラーは、既知である。これは、胎児DNA画分 f について得られた式を、等式10に示されるように単純化する。図8は、この型の混合モデルが、等式9によって使用されたアプローチと比較して、改善された結果を提供したことを示す。

10

【0130】

類似のアプローチは、等式11および12において示される。このアプローチは、いくつかの配列エラーがマイナーアレルカウントに追加されるのみであることが、認識される。むしろ、4つの配列決定エラー毎に1つのみが、マイナーアレルカウントを増大する。図9は、本技術を使用して、実測の胎児画分と推定胎児画分との間の著しく良好な一致を示す。

20

【0131】

モデル事例(1+2)、配列決定エラー既知

使用される機械の配列決定エラー率は、大部分が既知であり、計算のバイアスおよび複雑性は、解かれるべき変数である e がなくなることによって低下し得る。したがって、本発明者らは、以下の等式系を得る：

$$F_1 = e + (1 - e)(f/2)$$

$$F_2 = e^2 + (1 - e)(f/2)^2$$

等式10

胎児画分 f について解を得るため：

$$F \approx \frac{2(eF_1 - F_2)}{(e - F_1)}$$

30

【0132】

図8は、既知のパラメータとして機械エラー率を用いることにより、一点における上方バイアスが低下することを示す。

【0133】

モデル事例(1+2)、配列決定エラー既知、改善されたエラーモデル

モデルにおけるバイアスを改善するために、本発明者らは、上の等式のエラーモデルを、全ての配列決定エラー事象がマイナーアレルカウントに加えられるわけではない事実を考慮するように、拡張した。ヘテロ接合生殖性事例1において、 $A=ai$ である。さらに、本発明者らは、配列決定エラー事象がヘテロ接合生殖性事例2カウントに貢献し得るという事実を認める。それゆえに、本発明者らは、胎児画分 F を、以下の階乗モーメント関係式の系を解くことにより、決定する。

40

$$F_1 = \alpha_1 e / 4 + (1 - \alpha_1)(e + f/2)$$

$$F_2 = \alpha_1 \left(\frac{e}{4}\right)^2 + (1 - \alpha_1)(e + f/2)^2$$

等式11

これは、以下の解を生じる：

50

$$F \approx \frac{-2(e^2 - 5eF1 + 4F2)}{(e - 4F1)}.$$

等式 12

【 0 1 3 4 】

既知のパラメータとして機械エラー率を使用する模擬データを示す図9において、事例1および2エラー モデルの拡張は、上方バイアスを大いに下げ、胎児画分についての点を0.2未満にする。

【 0 1 3 5 】

実施選択肢

サンプル

本明細書において開示される態様において使用されるサンプルは、細胞性または無細胞性のゲノムDNAを含む。細胞性DNAは、遺伝組成が同じまたは異なる細胞全体から手動でもしくは機械的にゲノムDNAを抽出することにより、細胞全体に由来する。細胞性DNAは、例えば、1対象に由来する同じ遺伝構成の細胞全体に由来してもよく、異なる対象の細胞全体の混合物に由来してもよく、または1対象に由来する遺伝的構成の異なる細胞全体の混合物に由来してもよい。細胞全体からゲノムDNAを抽出する方法は、当技術分野において公知であり、供給源の性質に依存して異なる。

【 0 1 3 6 】

いくつかの場合において、細胞性ゲノムDNAを断片化することが、有益であり得る。断片化は、ランダムであってもよく、または、例えば、制限エンドヌクレアーゼ消化を用いて達成されるように、特異的であってもよい。ランダム断片化のための方法は、当技術分野において周知であり、例えば、限定DNA消化、アルカリ処理および物理的剪断が挙げられる。特定の態様において、サンプル核酸は、約500以上の塩基対の断片への断片化に供され、これが、次世代配列決定（NGS）方法が、容易に適用され得る。1つの態様において、サンプル核酸は、cfDNAとして得られ、これは、断片化に供されない。

【 0 1 3 7 】

無細胞DNAは、対象の生物体液、例えば血液において典型的に見出されるゲノム断片の混合物として、天然に存在するゲノムDNAである。ゲノム混合物は、生物学的プロセス（例えば、アポトーシス）により、天然で破裂してそのゲノム内容物を放出した細胞に由来し得る。cfDNAのサンプルは、同じ種の異なる対象の細胞の混合物に由来するか、異なるゲノム組成の1対象の細胞の混合物に由来するか、または異なる種、例えば対象からの細胞の混合物に由来するcfDNAを含んでもよい。

【 0 1 3 8 】

無細胞核酸（無細胞DNAを含む）は、生物学的サンプル（血漿、血清および尿を含むが、これらに限定されない）から、当技術分野において公知の種々の方法によって得られ得る（Fanら、Proc Natl Acad Sci 105: 16266-16271 [2008年]；Koideら、Prenatal Diagnosis 25: 604-607 [2005年]；Chenら、Nature Med. 2: 1033-1035 [1996年]；Loら、Lancet 350: 485-487 [1997年]；Botezatuら、Clin Chem. 46: 1078-1084, 2000年；およびSuら、J Mol. Diagn. 6: 101-107 [2004年]）。細胞からcfDNAを分離するために、分画法、遠心分離法（例えば、密度勾配遠心分離法）、DNA特異的沈殿法またはハイスループットセルソーティングおよび/もしくは分離方法が、使用され得る。cfDNAの手動または自動の分離のための市販のキットは、入手可能である（Roche Diagnostics, Indianapolis, IN, Qiagen, Valencia, CA, Macherey-Nagel, Duren, DE）。

【 0 1 3 9 】

本明細書において記載される方法が適用される核酸の混合物を含むサンプルは、生物学的サンプル、例えば組織サンプルであっても、生物体液サンプルであっても、または細胞サンプルであってもよい。いくつかの態様において、核酸の混合物は、公知の方法いずれか1つにより、生物学的サンプルから、精製されるか、または単離される。サンプルは、

10

20

30

40

50

精製されたポリヌクレオチドまたは単離されたポリヌクレオチドであり得る。生物体液としては、非限定の例として、血液、血漿、血清、汗、涙、痰、尿、痰、耳液、リンパ液、唾液、脳脊髄液、破碎物、骨髄浮遊液、膿液、大腿骨頸部洗浄液、脳液、腹水、乳汁、呼吸管、腸管および非尿生殖器管の分泌液、羊水および白血球フェレーシス (leukophoresis) のサンプルが挙げられる。いくつかの態様において、サンプルは、非侵襲性手順によって容易に入手可能なサンプル、例えば、血液、血漿、血清、汗、涙、痰、尿、痰、耳液、唾液または糞便である。好ましくは、生物学的サンプルは、末梢血サンプル、または血漿および血清画分である。他の態様において、生物学的サンプルは、スワップもしくはスメア、生検標本または細胞培養物である。別の態様において、サンプルは、2種以上の生物学的サンプルの混合物、例えば、生物体液サンプル、組織サンプルおよび細胞培養物サンプルの2種以上を含み得る生物学的サンプルである。本明細書において使用される用語「血液」、「血漿」および「血清」は、画分またはその処理部分を明確に含む。同様に、サンプルが生検、スワップ、スメアなどから得られる場合、「サンプル」は、生検、スワップ、スメアなどに由来する処理画分または部分を明白に含む。

【0140】

いくつかの態様において、サンプルとしては、異なる個体由来のサンプル、同じまたは異なる個体の異なる発生段階由来のサンプル、異なる疾患の個体（例えば、癌を有する個体または遺伝性障害を有することが疑われる個体）由来のサンプル、正常個体由来のサンプル、一個体における疾患の異なる段階において得られたサンプル、疾患についての異なる処置を受けた個体から得られたサンプル、異なる環境因子を受けた個体由来のサンプル、または病理の素因を有する個体もしくは感染性疾患因子（例えば、HIV）に曝露された個体由来のサンプルが挙げられるが、これらに限定されない。

【0141】

1つの態様において、サンプルは、妊娠雌性体、例えば妊娠女性から得られる母性サンプルである。この場合において、サンプルは、本明細書において記載される方法を用いて分析されて、胎児における染色体異常の可能性の出生前診断を提供し得る。母性サンプルは、組織サンプル、生物体液サンプル、または細胞サンプルであってもよい。生物体液としては、非限定の例として、血液、血漿、血清、汗、涙、痰、尿、痰、耳液、リンパ液、唾液、脳脊髄液、破碎物、骨髄浮遊液、膿液、大腿骨頸部洗浄液、脳液、腹水、乳汁、呼吸管、腸管および非尿生殖器管の分泌液、および白血球フェレーシス (leukophoresis) のサンプルが挙げられる。別の態様において、母性サンプルは、2種以上の生物学的サンプルの混合物、例えば、2種以上の生物体液サンプル、組織サンプルおよび細胞培養物サンプルを含み得る生物学的サンプルである。いくつかの態様において、サンプルは、非侵襲性手順によって容易に入手可能なサンプル（例えば、血液、血漿、血清、汗、涙、痰、尿、痰、耳液、唾液および糞便）である。いくつかの態様において、生物学的サンプルは、末梢血サンプルであるか、または血漿および血清画分である。他の態様において、生物学的サンプルは、スワップもしくはスメア、生検標本、または細胞培養物である。

【0142】

サンプルはまた、インビトロ培養の組織、細胞または他のポリヌクレオチド含有供給源から得られ得る。培養サンプルとしては、異なる培地および条件（例えば、pH、圧力または温度）において維持された培養物（例えば、組織または細胞）、異なる期間の長さに亘って維持された培養物（例えば、組織または細胞）、異なる因子または試薬（例えば、候補薬物またはモジュレーター）によって処理された培養物（例えば、組織または細胞）、または異なる種類の組織もしくは細胞の培養物が挙げられるが、これらに限定されない。生物学的供給源から核酸を単離する方法は、周知であり、上で説明されたような供給源の性質に依存して異なる。

【0143】

ゲノム画分の同定における使用のための多型

説明されるように、多型は、胎児画分を評価するために使用され得る。1つまたは複数の多型のアレル画分および接合生殖性は、この評価において使用される。有用な多型の例

10

20

30

40

50

としては、非限定的に、一ヌクレオチド多型 (SNP)、タンデムSNP、インデルと呼ばれる小規模複数塩基欠失もしくは挿入 (欠失挿入多型すなわちDIPとも呼ばれる)、マルチヌクレオチド多型 (MNP)、短タンデムリピート (STR)、制限断片長多型 (RFLP)、欠失 (微小欠失を含む)、挿入 (微小挿入を含む)、二重重複、反転、転位、増殖、複雑多部位変異、コピー数変異 (CNV) および染色体における配列の任意の他の変化を含む多型が挙げられる。

【 0 1 4 4 】

いくつかの態様において、本開示の方法において使用される多型は、SNPおよび/またはSTRを含む。SNP多型は、単一SNP、タンデムSNPであり得る。単一SNPは、個々のSNPおよびタグSNP (すなわち、ハプロタイプおよび/またはハプロタイプブロックに存在するSNP) を含む。いくつかの態様において、多型の組み合わせが使用される。例えば、コピー数における相違は、1つまたは複数のSNPおよび1つまたは複数のSTRを含む多型配列の組み合わせの比較によって検出され得る。

【 0 1 4 5 】

一般的に、本明細書において記載される配列決定方法により作成された読み取りに含まれ得る任意の多型部位は、異なるゲノムのDNAを含むサンプルにおいてゲノム画分を同定するために使用され得る。本発明の方法を実施するために有用な多型配列は、種々の公にアクセス可能なデータベース (継続的に拡張している) から入手可能である。例えば、有用なデータベースとしては、非限定的に、ワールドワイドウェブアドレス`www.mit.edu`におけるヒトSNPデータベース、ワールドワイドウェブアドレス`ncbi.nlm.nih.gov`におけるNCB¹⁰ I dbSNPホームページ、ワールドワイドウェブアドレス`lifesciences.perkinelmer.com`、ワールドワイドウェブアドレス`celera.com`におけるCeleraヒトSNPデータベース、ワールドワイドウェブアドレス`gan.iarc.fr`におけるGenome Analysis Group (GAN) のSNPデータベース、ワールドワイドウェブアドレス`atcc.org`におけるATCC短タンデムリピート (STR) データベース、およびワールドワイドウェブアドレス`hapmap.org`におけるHapMapデータベースが挙げられる。

【 0 1 4 6 】

胎児画分評価において使用され得る多型の数は、少なくとも、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、200、300、400、500、600、700、800、900、1000以上であり得る。例えば、ヒトゲノムは、少なくとも約1000万のSNPを含むことが推定される。したがって、ヒト対象由来のサンプルにおいてゲノタイピングされ得る利用可能な多型の数は、少なくとも約1000万のSNP、ならびに、任意の1人のヒトゲノム中に存在する多くの他の種の多型であり得る。いくつかの態様において、第1および第2のゲノムのDNA (例えば、cfDNA) の混合物を含むサンプルの第1のゲノムにおける1つまたは複数の多型の同定は、本明細書において記載されるNGS方法を用いる全ゲノム配列決定によって実施される。いくつかの態様において、全ゲノム配列決定方法は、クローン増幅核酸分子の大規模パラレル配列決定または1核酸分子の大規模パラレル配列決定 (すなわち、1分子配列決定) により多型配列を同定するNGS方法である。

【 0 1 4 7 】

適用

サンプル中の2つの異なるゲノム供給源のそれぞれに由来する核酸の画分は、種々の目的のために使用され得る。本明細書において記載される種々の態様において、母性サンプル中の無細胞DNAにおける胎児DNAの画分は、出生前診断を容易にするため、および妊娠の処置を考慮した決定を行うことを助けるために、使用される。他の態様において、検討中のゲノムは、母性のものでも胎児のものでもない。微量ゲノムの存在を決定するためのゲノム供給源の種々の例は、以下に示される。

【 0 1 4 8 】

母性血液中を循環する無細胞胎児DNAおよびRNAは、妊娠管理のためおよび生殖決定の補助のための両方で、数が増大しつつある遺伝的疾患の初期非侵襲性出生前診断(NIPD)のた²⁰ 50

めに使用され得る。少量の循環する胎児DNAが、妊娠期間中、母性血流中に存在する(Loら、Lancet 350: 485-487 [1997年])。死にゆく胎盤細胞に起源すると想定すると、無細胞胎児DNAは、典型的に200bp長よりも短い短断片からなることが示されており(Chanら、ClinChem 50: 88-92 [2004年])、これは妊娠第4週の早さで認められ得(IIlanesら、Early Human Dev 83: 563-566 [2007年])、送達の時間内に母性循環から取り除かれることが知られている(Loら、Am J Hum Genet 64: 218-224 [1999年])。cfDNAに加え、無細胞胎児RNA(cfRNA)の断片もまた、母性血流において認められ得、胎児もしくは胎盤において転写された遺伝子に起源する。母性血液サンプルからのこれらの胎児遺伝因子の抽出およびその後の解析は、NIPDのための新たな機会を提供する。

【0149】

10

説明されるように、本開示の方法は、生物学的サンプル中の第2のゲノムの画分を決定する。この方法は、第1および第2のゲノムのDNA(例えば、cfDNA)の混合物を含む血液サンプルにおいて、多くの障害の存在または非存在を決定してもよい。いくつかの態様において、胎児画分の決定は、(a)cfDNAの混合物の少なくとも一部をゲノム配列決定し、複数の配列タグを得ること;(b)複数の配列タグにおいて、多数の多型の存在または非存在を決定すること、および(c)多数の多型と混合物中の第1および/または第2のゲノムとを関連づけることを、含み得る。好ましい態様において、この混合物は、多数の多型について富化されていない。DNAの混合物における多数の多型の同定は、本明細書において記載されるように、全ゲノム配列決定方法によって得られたマッピングされたタグの配列を、多数の参照多型と比較することによって実施される。

20

【0150】

上で開示された態様において、第1のゲノムは、胎児ゲノムであり、そして第2のゲノムは、母性ゲノムである。別の態様において、第1のゲノムは、非罹患の細胞由来のゲノムであり、第2のゲノムは、罹患した細胞、例えば癌細胞に由来するゲノムである。いくつかの態様において、罹患した細胞および非罹患の細胞は、同じ対象に由来する。例えば、罹患した細胞は、障害によってゲノムが改変されている細胞であり得る。いくつかの態様において、障害は、単一遺伝子性障害である。他の態様において、障害は、多遺伝子性障害である。障害は、1つの多型、すなわちタグSNPによって、またはハプロタイプにおいて存在する多数の多型によって、同定され得る。いくつかの態様において、本方法にしたがって同定される多数の多型は、1つのハプロタイプブロックに存在する。

30

【0151】

本方法の助けによって同定され得る障害は、少なくとも部分的に遺伝子もしくは染色体の異常によって疾病が起きる、遺伝的障害である。サンプル中の胎児画分の知見は、このような障害を出生前の状況において同定することを助け得る。本方法によって同定される障害としては、単一遺伝子性(すなわち1つの遺伝子)の障害および多遺伝子性の(すなわち、複合的)障害が挙げられる。単一遺伝子性障害としては、常染色体優性、常染色体劣性、X染色体優性、X染色体劣性およびY連鎖の障害が挙げられる。

【0152】

40

常染色体優性障害において、障害に罹患したヒトに必須であるのは、突然変異した1コピーの遺伝子のみである。典型的には、罹患した対象は、一方の罹患した親を持ち、子孫が突然変異した遺伝子を受け継ぐ50%の機会が存在する。常染色体優性が時々低下した浸透度を有する状態は、1つの突然変異コピーのみが必須であるが、この突然変異を受け継いだ個体の全てがこの疾患を発症するわけではないことを意味する。本方法によって同定され得る常染色体優性障害の例としては、非限定的に、家族性高コレステロール血症、遺伝性球状赤血球症、マルファン症候群、1型神経線維腫症、遺伝性非ポリポーシス結腸直腸癌および遺伝性多発性外骨腫症、ならびにハンチントン病が挙げられる。

【0153】

本方法を使用して検出される常染色体劣性障害としては、鐮状赤血球貧血、囊胞性纖維症、ティ・サックス病、ティ・サックス病、ムコ多糖症、糖原病およびガラクトース血症が挙げられる。本方法によって検出されるX染色体連鎖障害としては、デュシェンヌ筋ジ

50

ストロフィーおよび血友病が挙げられる。常染色体劣性障害において、常染色体劣性障害に罹患した対象には2コピーの遺伝子が突然変異している必要がある。罹患した対象は、通常、それぞれ1コピーの突然変異遺伝子を保有している非罹患の親（キャリアと呼ばれる）を有する。各々1コピーの突然変異遺伝子を保有する2人の非罹患のヒトは、各妊娠につき、この障害を罹患する子どもを得る25%の機会を有する。本方法によって同定され得るこの種の障害の例としては、囊胞性纖維症、鎌状赤血球貧血、ティ・サックス病、ニーマン・ピック病、脊髄性筋萎縮症およびローベルト症候群が挙げられる。特定の他の表現型（湿性対乾性の耳あかなど）もまた、常染色体劣性様式で決定される。X染色体優性障害は、X染色体における遺伝子の突然変異によって起こる。ほんの少しの障害のみが、この遺伝パターンを有する（主要な例は、X連鎖型低リン血症性くる病である）。男性および女性は両方ともこれらの障害に罹患し、典型的には、雄性体は雌性体よりもより重度に罹患する。いくつかのX優性状態、例えばレット症候群、2型色素失調症およびアイカルディ症候群は、通常、雄性体において致命的であり、したがって、主として雌性体に見られる。クラインフェルター症候群（47,XXY）を有する男の子もまたX優性状態を受け継ぎ、疾患の重篤度の点でこの疾患を有する雌性体と類似した症状を示すという、極めて稀な例を除く。X優性障害を遺伝する機会は、男性と女性との間で異なる。X染色体優性障害を有する男性の息子は、全て非罹患であり（彼らは、父のY染色体を受け継ぐため）、娘は全てこの状態を受け継ぐ。X優性障害を有する女性は、各妊娠について罹患した胎児を有する50%の機会を有するが、色素失調症などの場合においては、雌性体子孫のみが遺伝的に生存可能であることに留意されたい。加えて、これらの状態は受精能力自体を変化させないが、レット症候群またはアルカルディ症候群を有する個体は、滅多に生殖しない。

【0154】

本方法はまた、X連鎖障害に関連する多型の同定を容易にし得る。X連鎖劣性疾患もまた、X染色体上の遺伝子における突然変異によって引き起こされる。雄性体は、雌性体より頻繁に罹患し、この障害を遺伝する機会は、男性と女性との間で異なる。X連鎖劣性障害を有する男性の息子は罹患せず、娘は1コピーの突然変異遺伝子を保有する。X連鎖劣性障害のキャリアである女性（ $X^R X^I$ ）は、罹患する息子を有する50%の機会と、1コピーの突然変異した遺伝子を保有する、したがってキャリアである娘を有する50%の機会を有する。X連鎖劣性状態としては、非限定的に、重症血友病A、デュシェンヌ筋ジストロフィー、およびレッシュ・ナイハン症候群ならびに一般的且つより深刻でない状態、例えば男性型脱毛症および赤緑色盲が挙げられる。X連鎖劣性状態は、時々、雌性体において、偏ったX不活性またはモノソミーX（ターナー症候群）により顕在化し得る。

【0155】

Y連鎖障害は、Y染色体上の突然変異によって引き起こされる。雄性体は、Y染色体をその父から受け継ぐので、罹患した父の全ての息子は、罹患する。雌性体は、その父からX染色体を受け継ぐので、罹患した父の雌性体子孫は、決して罹患しない。Y染色体は比較的小小さく、非常に少ない遺伝子を含むので、比較的少ないY連鎖障害しか存在しない。多くの場合、この症状は、何らかの生殖処置の助けによって回避し得る不妊を含む。例は、雄性不妊および耳介多毛症である。

【0156】

説明されたように、本開示のサンプルにおける遺伝画分を検出するための方法は、母性サンプルからの異数性の検出を容易にするために使用され得る。いくつかの態様において、異数性は、完全染色体トリソミーもしくはモノソミー、または部分的トリソミーもしくはモノソミーである。部分的異数性は、染色体の部分の欠失もしくは獲得によって起こり、不均衡な転移、不均衡な反転、欠失および挿入に起因する染色体不均衡を包含する。飛び抜けて、最も一般的に知られる、生存と両立し得る異数性は、第21番トリソミー、すなわちダウン症候群（DS）であって、これは、第21番染色体の一部分または全ての存在によって引き起こされる。稀に、DSは、遺伝的に受け継いだかまたは散発性の欠損によって起こり得、それにより、第21番染色体の全てもしくは一部分の余剰なコピーが、別の染色体（通常、第14番染色体）に付着し、1つの異常な染色体を形成する。DSは、知能障害、重

10

20

30

40

50

度の学習困難および超過死亡率（長期間の健康問題、例えば心疾患によって起こる）に関係する。公知の臨床的重要性を有する他の異数性としては、エドワード症候群（第18番トリソミー）およびパトー症候群（第13番トリソミー）が挙げられ、これらは、多くの場合、生まれて最初の数ヶ月間で死に至る。性染色体の数に関連する異常もまた公知であり、雌性出産におけるモノソミーX（例えば、ターナー症候群（X0））、三倍性X染色体症候群（XXX）および、雄性出産におけるクラインフェルター症候群（XXY）およびXYY症候群が挙げられ、これらは全て、不妊性および知的技能の低下を含む種々の表現型に関連する。モノソミーX[45,X]は、自然流産の約7%にのぼる妊娠初期死亡の一般的な原因である。45,X（ターナー症候群とも呼ばれる）の1~2/10,000の出生頻度に基づき、45,X受胎産物の1%未満しか、出産まで生き延びられないと推定される。ターナー症候群患者の約30%が、45,X細胞系統と、46,XX細胞系統または再編成されたX染色体を含むもののどちらかとの両方による、モザイクである（HookおよびWarburton 1983年）。生産児における表現型は、高い胚性致死率を考えると、比較的穏やかであり、ターナー症候群を有する恐らく全ての雌性生産児は、2つの性染色体を含む細胞系統を有すると仮定された。モノソミーXは、雌性において45,Xまたは45,X/46XXとして、そして雄性において45,X/46XYとして起こり得る。ヒトにおける常染色体モノソミーは、一般的に、生存適合性でないと示唆される；しかし、生産児における1つの第21番染色体の完全モノソミーを記載する、極めて多くの細胞遺伝学的報告が存在する（Vosranovalら、Molecular Cytogen. 1: 13 [2008年]；Joostenら、Prenatal Diagn. 17: 271-5 [1997年]）。本発明の方法は、これらおよび他の染色体異常を、出生前に診断するために使用され得る。

10

20

【0157】

いくつかの態様にしたがい、胎児画分は、第1~22番、XおよびY染色体のいずれか1つの染色体トリソミーの存在または非存在を決定する際に有用であり得る。本発明にしたがって検出され得る染色体トリソミーの例としては、非限定的に、第21番トリソミー（T21；ダウン症候群）、第18番トリソミー（T18；エドワード症候群）、第16番トリソミー（T16）、第20番トリソミー（T20）、第22番トリソミー（T22；ネコ眼症候群）、第15番トリソミー（T15；プラダー・ウィリー症候群）、第13番トリソミー（T13；パトー症候群）、第8番トリソミー（T8；ウォーカニー症候群）、第9番トリソミーおよびXXY（クラインフェルター症候群）、XYYまたはXXXトリソミーが挙げられる。非モザイク状態で存在する他の常染色体の完全トリソミーは、致死性であるが、モザイク状態で存在する場合は、生存適合性であり得る。モザイク状態で存在するかまたは非モザイク状態で存在するかいずれかである種々の完全トリソミー、および部分的トリソミーは、本発明の技術にしたがって胎児cf DNAにおいて決定され得ることが理解される。

30

【0158】

本方法によって決定可能な部分的トリソミーの非限定の例としては、部分的トリソミー1q32-44、トリソミー9p、トリソミー4モザイク型、トリソミー17p、部分的トリソミー4q26-qter、部分的2pトリソミー、部分的トリソミー1q、および/または部分的トリソミー6p/モノソミー6qが挙げられるが、これらに限定されない。

【0159】

本明細書において開示される方法はまた、染色体モノソミーX、染色体モノソミー-21および部分的モノソミー（モノソミー-13、モノソミー-15、モノソミー-16、モノソミー-21およびモノソミー-22など）（これらは、妊娠流産に関係することが公知である）を決定することを補助するために使用され得る。典型的には完全異数性に含まれる染色体の部分的モノソミーもまた、本発明の方法によって決定され得る。本方法にしたがって決定され得る欠失症候群の非限定の例としては、染色体の部分的欠失によって引き起こされる症候群が、挙げられる。本発明の方法にしたがって決定され得る部分的欠失の例としては、非限定的に、第1、第4、第5、第7、第11、第18、第15、第13、第17、第22および第10番染色体の部分的欠失が挙げられ、これらは、以下に記載される。

40

【0160】

1q21.1欠失症候群または1q21.1（再発性）微小欠失は、稀な第1番染色体の異常である

50

。欠失症候群に次いで、1q21.1二重重複症候群も存在する。特定のスポット上の欠失症候群によるDNAの一部分の欠如が存在する一方、二重重複症候群により、同じスポット上のDNAの類似の部分の2コピーまたは3コピーが存在する。文献は、1q21.1コピー数変動(CNV)として欠失および二重重複の両方を言及する。1q21.1欠失は、TAR(橈骨欠損症を伴う血小板減少)症候群に関連し得る。

【0161】

ウォルフ・ヒルシュホーン症候群(WHS)(OMIM#194190)は、染色体4p16.3のヘミ接合性欠失に関連する連続的遺伝子欠失症候群である。ウォルフ・ヒルシュホーン症候群は、出生前および出生後の成長欠損、種々の程度の発達障害、特徴的な頭蓋顔面容貌(鼻、高い額、張り出した眉間、隔離症、高い三日月眉、突出した眼、内眼角ぜい皮、短い人中、口角をへの字に曲げた独特の口および小顎の「ギリシャ兜」外観)、および発作性疾患によって特徴付けられる先天性奇形症候群である。10

【0162】

第5番染色体の部分的欠失(5p-もしくは5pマイナスとしても公知であり、クリードゥシヤー症候群(OMIM#123450)と呼ばれる)は、第5染色体の短腕(p腕)の欠失(5p15.3-p15.2)によって起こる。この状態を有する胎児は、多くの場合、ネコの鳴き声のような甲高い泣き声を有する。この障害は、幼年時において、知的障害および発達遅延、小さな頭部サイズ(小頭症)、低出生体重および弱い筋緊張(低血圧)により特徴付けられ、明らかな顔の容貌および心疾患の可能性を有する。20

【0163】

染色体7q11.23欠失症候群としても知られるウィリアムズ・ビューレン症候群(OMIM 194050)は、染色体7q11.23上の1.5~1.8Mbのヘミ接合性欠失(約28の遺伝子を有する)によって起こる、多系統障害をもたらす連続的遺伝子欠失症候群である。28

【0164】

11q欠失障害としても知られるヤコブセン症候群は、バンド11q24.1を含む第11番染色体の末端領域における欠失から生じる、稀な先天性障害である。これは、知的障害、明らかな顔の外観および種々の身体的問題(心臓欠陥および出血障害を含む)を引き起こし得る。30

【0165】

第18番染色体の部分的モノソミー(モノソミー-18pとして知られる)は、第18番染色体の短腕(p)の全体または一部分が欠失している(一染色体性である)稀な染色体障害である。この障害は、典型的に、低伸長、種々の程度の精神遅滞、言葉の遅れ、頭蓋および顔(頭蓋顔面)領域の奇形、および/またはさらなる身体的異常によって特徴付けられる。関連の頭蓋顔面欠陥は、場合によって、程度および重症度が大いに変動し得る。38

【0166】

第15番染色体の構造もしくはコピー数の変化によって引き起こされる疾患としては、アンジェルマン症候群およびプラダー・ウィリー症候群が挙げられ、これは、第15番染色体の同じ部位(15p11-q13領域)の遺伝子活性の欠失に関する。いくつかの転位および微小欠失は、キャリア親において無症候性であり得るが、なお、子孫において主要な遺伝的疾患を引き起こし得ることが、理解される。例えば、15q11-q13微小欠失を保有する健康な母は、重症の神経変性障害であるアンジェルマン症候群を有する子どもを出産し得る。したがって、本発明は、胎児におけるこのような部分的欠失および他の欠失を同定するために使用され得る。40

【0167】

部分的モノソミー-13qは、第13番染色体の長腕(q)の一片が失われている(一染色体性である)際に生じる、稀な染色体障害である。部分的モノソミー-13qを有して出生した乳児は、低出生体重、頭部および顔(頭蓋顔面領域)の奇形、骨格異常(特に手および足)、および他の身体的異常を示し得る。精神遅滞は、この状態の特徴である。幼児期の間の死亡率は、この障害を有する出生個体の間で高い。部分的モノソミー-13qのほとんど全ての症例は、明らかな理由なしに、ランダムに生じる(散発性)。50

【0168】

スミス・マゲニス症候群 (SMS、OMIM#182290) は、第17番染色体の1コピー上の遺伝物質の欠失または喪失によって引き起こされる。この周知の症候群は、発達遅延、精神遅滞、先天異常（例えば心臓および腎臓の欠陥）および神経行動学的異常（例えば、重度の睡眠障害および自傷行動）に関連する。スミス・マゲニス症候群 (SMS) は、ほとんどの場合（90%）、染色体17p11.2の3.7Mbの中間部欠失によって引き起こされる。

【0169】

ディジョージ症候群としても知られる22q11.2欠失症候群は、第22番染色体の小片の欠失によって引き起こされる症候群である。欠失 (22q11.2) は、一対の染色体のうちの1つの、長腕上の染色体中央付近で起こる。この症候群の特徴は、同じ家族のメンバーの間であっても広く変動していて、身体の多くの部分に影響を及ぼす。特徴的徵候および症候は、出生時欠損、例えば先天性心疾患、口蓋における欠損（最も一般的には、閉鎖についての神経筋問題に関連する）（口蓋帆咽頭不全）、学習障害、顔の容貌の軽度の相違、および感染の再発を含み得る。染色体領域22q11.2における微小欠失は、統合失調症の危険の20～30倍の増大に関連する。

10

【0170】

第10番染色体の短腕上の欠失は、ディジョージ症候群様の表現型に関連する。染色体10pの部分的モノソミーは、稀であるが、ディジョージ症候群の特徴を示す患者の一部に観察されてきている。

20

【0171】

1つの態様において、本発明の方法は、部分的モノソミー（第1、第4、第5、第7、第11、第18、第15、第13、第17、第22および第10番染色体の部分的モノソミーが挙げられるが、これらに限定されない）を決定するために使用され、例えば、部分的モノソミー1q21.1、部分的モノソミー4p16.3、部分的モノソミー5p15.3-p15.2、部分的モノソミー7q11.23、部分的モノソミー11q24.1、部分的モノソミー18p、第15番染色体の部分的モノソミー（15q11-q13）、部分的モノソミー13q、部分的モノソミー17p11.2、第22番染色体の部分的モノソミー（22q11.2）、および部分的モノソミー10pもまた、本方法を用いて決定され得る。

【0172】

本発明の方法にしたがって決定され得る他の部分的モノソミーとしては、不均衡な転位t(8;11)(p23.2;p15.5);11q23微小欠失;17p11.2欠失;22q13.3欠失;Xp22.3微小欠失;10p14欠失;20p微小欠失、[de1(22)(q11.2q11.23)]、7q11.23および7q36欠失;1p36欠失;2p微小欠失;1型神経線維腫症(17q11.2微小欠失)、Yq欠失;4p16.3微小欠失;1p36.2微小欠失;11q14欠失;19q13.2微小欠失;ルビンシュタイン・ティビ(16p13.3微小欠失);7p21微小欠失;ミラー・ディーカー症候群(17p13.3);および2q37微小欠失が挙げられる。部分的欠失は、染色体の一部分の小さな欠失であってもよく、またはこれらは、1遺伝子の欠失が起こり得る染色体の微小欠失であってもよい。

30

【0173】

染色体腕の一部分の二重重複によって引き起こされるいくつかの二重重複症候群は、同定されている（OMIM [ncbi.nlm.nih.gov/omim]においてオンラインで見られるOnline Mendelian Inheritance in Man]を参照されたい）。1つの態様において、本方法は、第1～22番、XおよびY染色体のいずれか1つのセグメントの二重重複および/もしくは多重重複の存在または非存在を決定するために使用され得る。本方法にしたがって決定され得る二重重複症候群の非限定の例としては、第8、第15、第12および第17番染色体の一部分の二重重複が挙げられ、これは、下で説明される。

40

【0174】

8p23.1二重重複症候群は、ヒト第8番染色体由来の領域の二重重複によって引き起こされる、稀な遺伝障害である。この二重重複症候群は、64,000件の出産に対し1件の有病率と推定されており、8p23.1欠失症候群と相反する。8p23.1二重重複は、言葉の遅れ、発達の遅れ、軽度の異形症、突出した額および三日月眉ならびに先天的心疾患（CHD）の1つま

50

たは複数を含む種々の表現型に関連する。

【0175】

染色体15q二重重複症候群 (Dup15q) は、染色体15q11-13.1の二重重複から生じる、臨床的に同定可能な症候群である。Dup15qを有する子は、通常、筋緊張低下（筋緊張が乏しい）、成長遅滞を有する；彼らは、口唇裂および/もしくは口蓋裂、または心臓、腎臓もしくは他の器官の奇形を持って生まれてくる場合がある；彼らは、ある程度の認識の遅れ/障害（精神遅滞）、言葉および言語の遅れおよび感覚処理障害を示す。

【0176】

パリスター・キリアン症候群は、過剰な第12番染色体物質の結果である。通常、いくつかが過剰な第12番物質を有し、いくつかが正常（過剰な第12番物質を有さない46染色体）である細胞の混合物（モザイク型）が、存在する。この症候群を有する子は、重度の精神遅滞、乏しい筋緊張、「貧相な」顔の容貌および突出した額を含む、多くの問題を有する。彼らは、非常に薄い上唇および厚い下唇、ならびに短い鼻を有する傾向にある。他の健康問題としては、発作、摂食不良、固い関節、成人期の白内障、聴力喪失および心臓欠陥が挙げられる。パリスター・キリアンを有するヒトは、短い寿命を有する。

10

【0177】

dup(17)(p11.2p11.2)もしくはdup17pと呼ばれる遺伝状態を有する個体は、第17番染色体の短腕に過剰な遺伝情報を保有する（二重重複として知られる）。染色体17p11.2の二重重複は、ポトスキ-ルプスキ (Potocki-Lupski) 症候群(PTLS)の根底にあり、この症候群は、医学的文献において数ダースの症例しか報告されていない遺伝状態であると、新たに認識されている。この二重重複を有する患者は、多くの場合、低い筋緊張、乏しい摂食および幼児期の生育不全を有し、そして運動および言語のマイルストーンの発達遅延をも示す。PTLSを有する多くの個体は、発音および言語処理に困難を有する。加えて、患者は、自閉症もしくは自閉症スペクトラム障害を有するヒトにおいて見られるのと同様の行動特性を有し得る。PTLSを有する個体は、心臓欠陥および睡眠時無呼吸を有し得る。染色体17p12における、遺伝子PMP22を含む大きな領域の二重重複は、シャルコー・マリー・トゥース病を引き起こすことが公知である。

20

【0178】

CNVは、死産に関連している。しかし、従来の細胞遺伝学の固有の限界ゆえに、CNVの死産への寄与は、不十分であると考えられる (Harrisら、Prenatal Diagn 31: 932-944 [2011年])。本方法は、部分的異数性（例えば、染色体セグメントの欠失および多重重複）の存在の決定を助ける際に有用であり、同定を補助し死産に関連するCNVの存在または非存在を決定するために使用され得る。

30

【0179】

本方法はまた、複合的、多因子性または多遺伝子性の遺伝障害に関連する（生活様式因子および環境因子と組み合わせた、多数の遺伝子の効果に関連する可能性が高いことを意味する）多型を同定することを補助し得る。多因子性障害としては、例えば、心臓病および糖尿病が挙げられる。複合型障害は多くの場合、家族内で集まるが、これらは、遺伝の明確なパターンを有さない。ある家系において、多遺伝子性疾患は、「家族内で起こる」傾向があるが、遺伝は、メンデルの法則の疾患のように単純ではない。強い環境要素が、多くの複合型障害、例えば血圧と関連する。本方法は、多遺伝子性障害（喘息、多発性硬化症などの自己免疫疾患、癌、纖毛関連疾患、口唇裂、糖尿病、心疾患、高血圧、炎症性腸疾患、精神遅滞、気分障害、肥満、屈折異常および不妊症が挙げられるが、これらに限定されない）に関連する多型を同定するために使用され得る。いくつかの態様において、多型はSNPである。他の態様において、多型はSTRである。なお他の態様において、多型はSNPとSTRとの組み合わせである。

40

【0180】

1つの態様において、障害に関連する多型配列の同定は、cfDNAの混合物における第2のゲノムに相当する細胞性ゲノムの少なくとも1部分の配列決定を含む。第1ゲノムによって与えられた多型配列の同定は、本質的に第2ゲノムのみに由来するDNA分子を含む第1サン

50

プルにおいて、多数の多型部位において配列を決定し、この配列を、第1ゲノムと第2ゲノムとに由来するDNA分子の混合物を含む第2サンプルにおいて、対応する多数の多型部位において決定し、そして両サンプルにおいて決定された多型配列を比較することにより、2つのゲノムの混合物を含むサンプルの第1ゲノムにおいて多数の多型を同定することによって行われる。例えば、胎児ゲノム（すなわち、第1ゲノム）によって与えられた多型配列の同定は、配列を、母性バフィーコートサンプル（すなわち、本質的に第2ゲノムのみに由来するDNA分子を含むサンプル）において、多数の多型部位において決定し、この配列を、精製血漿サンプル（すなわち、胎児ゲノムおよび母性ゲノムに由来するcfDNA分子の混合物を含む第2サンプル）において、対応する多数の多型部位において決定し、そして両サンプルにおいて決定された多型配列を比較することにより、多数の胎児多型を同定することによって行われる。1つの態様において、第1ゲノムは、胎児ゲノムであり、第2ゲノムは、母性ゲノムである。別の態様において、第1ゲノムは、非罹患細胞のゲノムであり、第2ゲノムは、罹患細胞由来のゲノムである。いくつかの態様において、罹患および非罹患の細胞は、同じ対象に由来する。例えば、罹患した細胞は、ゲノムが障害によって変わっている細胞であり得る。

【0181】

1つの態様において、ゲノム画分を推定する本開示の方法は、患者において癌の検出を補助する。種々の例において、癌は、以下を含む方法によって検出される：正常（すなわち、非罹患の）細胞および癌性（すなわち、罹患の）細胞に由来するゲノムの混合物を含む患者由来のサンプルを提供すること；および、癌に関連する多数の多型を同定すること。いくつかの態様において、サンプルは、血液、血漿、血清および尿から選択される。いくつかの態様において、サンプルは、血漿サンプルである。他の態様において、サンプルは、尿サンプルである。

【0182】

1つの態様において、癌に関連する多数の多型の同定は、サンプル中のDNAを、多型標的配列について富化することを含む。他の態様において、サンプルの多型標的配列についての富化は、行われない。いくつかの態様において、癌に関連する多数の多型の同定は、多型配列のコピー数の定量を含む。

【0183】

本発明の方法にしたがって同定され得、および/またはモニタリングされ得る癌としては、 固形腫瘍、ならびに血液腫瘍および/または悪性腫瘍が挙げられる。処置される種々の癌としては、肉腫、癌腫、および腺癌が挙げられ、乳癌、肺癌、結腸直腸癌、膵臓癌、卵巣癌、前立腺癌、腎臓癌種、肝癌、脳腫瘍、黒色腫、多発性骨髄腫、リンパ腫、ホジキンリンパ腫、非ホジキンリンパ腫、小児リンパ腫、ならびにリンパ球起源および皮膚起源のリンパ腫、白血病、小児白血病、有毛細胞白血病、急性リンパ球白血病、急性骨髄細胞白血病、慢性リンパ球白血病、慢性骨髄細胞白血病、慢性骨髄性白血病および肥満細胞白血病、骨髄性新生物、肥満細胞新生物、血液腫瘍およびリンパ腫瘍に限定されず、原発性腫瘍部位から離れた他の組織または器官における転移病変を含む。

【0184】

本発明の方法は、例えば、特定のハプロタイプに関連することが公知である疾患状態における診断または予後の決定において、新規なハプロタイプを決定するために、および薬物に対する応答性に関係するハプロタイプを検出するために、有用である。多数の多型配列と多数の障害との関連は、多数の障害のそれぞれについての1つの多型配列の同一性から決定され得る。あるいは、多数の多型配列と多数の障害との関連は、多数の障害それぞれについての多数の多型配列の同一性から決定され得る。

【0185】

従来のゲノタイピング技術は、多型を数キロ塩基の短いゲノム領域内に同定することに限定されており、ハプロタイプの同定は、コンピューターアルゴリズムを用いた家族データおよび統計学的推定に依存する。全ゲノム配列決定は、多型のゲノム上での直接同定により、ハプロタイプの同定を可能にする。種々の態様にしたがうハプロタイプの同定は、

10

20

30

40

50

多型の間に介在する距離によって限定されない。いくつかの態様において、方法は、母性細胞性DNAを全ゲノム配列決定することを含む。母性細胞性DNAは、胎児ゲノムDNAを欠く生物学的サンプルから得られ得る。例えば、母性DNAは、母性血液のバフィーコート層から得られ得る。全染色体に亘る複数の多型配列を含むハプロタイプが、決定され得る。1つの態様において、胎児ハプロタイプは、公知の障害関連ハプロタイプと比較され、公知の障害関連ハプロタイプのいずれか1つとの胎児ハプロタイプの一致に基づき、胎児が障害を有するか、または胎児が障害に罹患し易いことを示す。胎児ハプロタイプはまた、特定の多型の処置応答性または非応答性と関連するハプロタイプと比較され得る。同定された胎児ハプロタイプと公知のハプロタイプデータベースとの比較は、障害の診断および/または予後診断を可能にする。胎児cfDNAおよび母性cfDNAの混合物を含む任意の生物学的サンプルは、胎児障害の存在または非存在を決定するために使用され得る。好ましくは、生物学的サンプルは、血液またはその画分（血漿を含む）または尿から選択される。1つの態様において、生物学的サンプルは、血液サンプルである。別の態様において、生物学的サンプルは、血漿サンプルである。なお別の態様において、生物学的サンプルは、尿サンプルである。

【0186】

1つの態様において、本発明は、多数の胎児障害の存在または非存在を決定するための方法を提供し、この方法は、(a)胎児DNAと母性DNAとの無細胞混合物を含む母性血液サンプルを得ること、(b)胎児DNAと母性DNAとの無細胞混合物の少なくとも1部分の全ゲノム配列決定をし、それによって、複数の配列タグを得ること；(c)この配列タグにおいて、多数の胎児多型を決定すること、および(d)多数の胎児障害の存在または非存在を決定することを、含む。本方法にしたがって同定され得る多数の胎児障害の例としては、本明細書において記載された単一遺伝子性障害および多遺伝子性障害が挙げられる。

【0187】

1つの態様において、本発明は、多数の胎児障害の存在または非存在を決定するための方法を提供し、この方法は、多数の障害関連ハプロタイプに関連する多数の胎児多型を同定することを含む。いくつかの態様において、ハプロタイプのそれぞれは、少なくとも、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも10または少なくとも15の異なるタグ多型を含む。ハプロタイプにおいて存在するタグ多型は、同種の多型、例えば、全てのタグSNP多型であり得るか、または、多型の組み合わせ、例えば、タグSNPとタグ欠失との組み合わせであり得る。1つの態様において、多型は、タグSNPである。別の態様において、多型は、タグSTRである。なお別の態様において、多型は、タグSNPとタグSTRとの組み合わせである。タグ多型は、ゲノムのコード領域にあってもよく、および/または非コード領域にあってもよい。多型の同定は、本明細書において記載されるNGS技術を用いる全ゲノム配列決定によって行われる。

【0188】

本発明は、2つの異なるゲノム（1つまたは複数の対象となる配列の量において異なることが公知であるかまたは疑われる）に由来する核酸の混合物を含む試験サンプル中の対象となる配列の多型としてコピー数変動（CNV）を同定するための方法を提供する。本発明の方法によって決定されるコピー数変動性は、染色体全体の獲得または喪失、非常に大きな（顕微鏡的に視認可能な）染色体セグメントに関わる変更、およびキロ塩基（kb）からメガ塩基（Mb）までの大きさにおよぶDNAセグメントの超顕微鏡的コピー数変動の非存在が、挙げられる。

【0189】

ヒトゲノムにおけるCNVは、ヒト多様性および疾患の傾向に有意に影響する（Redonら、Nature 23: 444-454 [2006年]、Shaikhら、Genome Res 19: 1682-1690 [2009年]）。CNVは、異なるメカニズムを通して、多くの場合、遺伝子量の不均衡または遺伝子破壊のいずれかによって、遺伝病に寄与することが公知である。遺伝障害とのその直接的相関性に加えて、CNVは、有害であり得る表現型変更を媒介することが公知である。近年、いくつかの研究が、稀なもしくはデノボのCNVの複合障害（自閉症、ADHDおよび統合失調症など）

10

20

30

40

50

において、正常対照と比較した負担の増大を報告しており、稀もしくは独特なCNVの病原性の可能性を強調している (Sebatら、316: 445-449 [2007年]; Walshら、Science 320: 539-543 [2008年])。CNVは、主に欠失、二重重複、挿入および不均衡な転位事象による、ゲノム再編成から生じる。

【0190】

本発明の態様は、2種の異なるゲノム（対象となる1つまたは複数の配列の量において異なることが公知であるかまたは疑われる）に由来する核酸の混合物を含む試験サンプル中の、対象となる配列、例えば臨床的に関連する配列のコピー数変動を評価する方法を提供する。核酸の混合物は、2種以上の細胞に由来する。1つの態様において、核酸の混合物は、医学的状態、例えば癌に罹患している対象に由来する正常細胞および癌性細胞に由来する。

10

【0191】

多くの固形腫瘍、例えば乳癌は、いくつかの遺伝子異常の蓄積を介した転移の開始から進行すると考えられる [Satoら、Cancer Res., 50: 7184-7189 [1990年]; Jongsmaら、J ClinPathol: Mol Path 55: 305-309 [2002年]]。このような蓄積する遺伝子異常は、増殖有利点、遺伝的不安定性および付き添い能力（薬物抵抗性を迅速に進化させる）および血管形成の増大、タンパク質分解および転移を付与し得る。遺伝子異常は、劣性「腫瘍抑制遺伝子」に影響してもよく、または優性的に作用する癌遺伝子に影響してもよい。ヘテロ接合生殖性の喪失 (LOH) をもたらす欠失および組換えは、突然変異した腫瘍抑制アレルを明らかにすることにより、腫瘍進行において、主要な役割を果たすと考えられる。

20

【0192】

cfDNAは、悪性腫瘍（肺癌 (Pathakら、ClinChem 52: 1833-1842 [2006年])、前立腺癌 (Schwartzzenbachら、Clin Cancer Res 15: 1032-8 [2009年]) および乳癌 (Schwartzzenbachら、breast-cancer-research.com/content/11/5/R71においてオンラインで入手可能 [2009年]) が挙げられるがこれらに限定されない）を診断されている患者の循環において見出されている。癌患者における循環するcfDNAにおいて決定され得る癌に関連する遺伝的不安定性の同定は、潜在的診断ツールおよび予後診断ツールである。1つの態様において、本発明の方法は、癌（例えば、癌腫、肉腫、リンパ腫、白血病、生殖細胞腫瘍および芽細胞腫）を有することが疑われるかまたは分かっている対象に由来する核酸の混合物を含むサンプルにおける対象となる配列のCNVを評価する。1つの態様において、サンプルは、末梢血に由来する（処理する）血漿サンプルであり、正常細胞および癌性細胞に由来するcfDNAの混合物を含む。別の態様において、CNVが存在するか否か決定する必要のある生物学的サンプルは、他の生体体液（血清、汗、涙、痰、尿、痰、耳液、リンパ液、唾液、脳脊髄液、破碎物、骨髄浮遊液、涙液、大腿骨頸部洗浄液、脳液、腹水、乳汁、気管、腸管および生殖器管の分泌物および白血球フェレーシス (leukophoresis) サンプルが挙げられるが、これらに限定されない）または組織生検、スワップまたはスメアに由来する癌性細胞および非癌性細胞の混合物に由来する。

30

【0193】

対象となる配列は、癌の発生および/または進行において役割を果たすことが分かっているかまたは疑われている核酸配列である。対象となる配列の例としては、以下で説明されるように、癌性細胞において増幅されるかまたは欠失される核酸配列が挙げられる。

40

【0194】

ヒト固形腫瘍に関連する優性的に作用する遺伝子は、典型的に、過剰発現または改変された発現によってその効果を発揮する。遺伝子増幅は、遺伝子発現の上方制御をもたらす一般的メカニズムである。細胞遺伝学研究の証拠は、ヒト乳癌の50%を超えて生じる有意な増幅を示す。最も留意すべきは、第17番染色体 (17(17q21-q22)) 上に位置する癌原遺伝子ヒト上皮増殖因子レセプター2 (HER2) の増幅は、細胞表面上のHER2レセプターの過剰発現をもたらし、乳癌および他の悪性腫瘍における過剰且つ調節不全のシグナル伝達に導く (Parkら、Clinical Breast Cancer 8: 392-401 [2008年])。種々の癌遺伝子が、他のヒト悪性腫瘍において増幅されることが見出されている。ヒト腫瘍における細胞性癌遺

50

伝子の増幅の例としては、以下の増幅が挙げられる：前骨髄球性白血病細胞株HL60および小細胞肺癌細胞株におけるc-myc、原発性神経芽細胞腫（ステージIIIおよびIV）、神経芽細胞腫細胞株、網膜芽細胞腫細胞株および原発性腫瘍、ならびに小細胞肺癌細胞株および腫瘍におけるN-myc、小細胞肺癌細胞株および腫瘍におけるL-myc、急性骨髄性白血病および結腸癌腫細胞株におけるc-myb、類表皮癌細胞および原発性グリア腫におけるc-erbB、肺、結腸、膀胱および直腸の原発性癌腫におけるc-K-ras-2、乳癌腫細胞株におけるN-ras（Varus H.、*Ann Rev Genetics* 18: 553-612 (1984年)）[Watsonら、*Molecular Biology of the Gene* (第4版、Benjamin/Cummings Publishing Co. 1987年において引用される)】。

【0195】

10

腫瘍抑制遺伝子に関わる染色体欠失は、 固形腫瘍の発生および進行において、重要な役割を果たし得る。網膜芽細胞腫腫瘍抑制遺伝子（Rb-1）（染色体13q14に位置する）は、最も広範囲に亘って特徴付けられた腫瘍抑制遺伝子である。Rb-1遺伝子産物である105 kDaの核リンタンパク質は、細胞周期調節において明らかに重要な役割を果たす（Howeら、*Proc Natl Acad Sci (USA)* 87: 5883-5887 [1990年]）。Rbタンパク質の改変されたかまたは失われた発現は、点変異もしくは染色体欠失のいずれかを介した両遺伝子アレルの不活化によって起こる。Rb-1遺伝子改変は、網膜芽細胞腫においてのみでなく、他の悪性腫瘍（例えば、骨肉腫、小細胞肺癌（Rygaardら、*Cancer Res* 50: 5312-5317 [1990年]）および乳癌においても存在することが見出されている。制限断片長多型（RFLP）研究は、このような腫瘍の種類は、13qにおけるヘテロ接合生殖性を頻繁に失っていることを示しており、このことは、Rb-1遺伝子アレルのうちの1つが、大規模染色体欠失に起因して失われていることを示唆する（Bowcockら、*Am J Hum Genet*, 46: 12 [1990年]）。二重重複、欠失および不均衡な転位（第6番染色体および他のパートナー染色体に関わる）を含む第1番染色体異常は、第1番染色体の領域、特に1q21-1q32および1p11-13は、骨髄増殖性新生物の慢性段階および進行段階の両方に病原的に関連する癌遺伝子もしくは腫瘍抑制遺伝子を、有し得ることを示す（Caramazzaら、*Eur J Hematol* 84: 191-200 [2010年]）。骨髄増殖性新生物もまた、第5番染色体の欠失に関連する。第5番染色体の完全な喪失または中間部欠失は、脊髄形成異常症候群（MDS）の最も一般的な核型異常である。隔離されたdel(5q)/5q-MDS患者は、さらなる核型欠損を有するもの（骨髄増殖性新生物（MPN）および急性骨髄性白血病を発症する傾向がある）よりも好ましい予後診断を有する。不均衡な第5番染色体欠失の頻度は、5qが、1つまたは複数の腫瘍抑制遺伝子（造血幹細胞/前駆細胞（HSC/HPC）における増殖制御において基本的役割を有する）を有するという着想に導いた。一般的欠失領域（CDR）の細胞遺伝学的マッピングは、5q31および5q32上に、候補腫瘍抑制遺伝子（リボソーマルサブユニットRPS14、転写因子Egr1/Krox20および細胞骨格リモデリングタンパク質 - カテニンが挙げられる）を中央に配置する（Eisenmannら、*Oncogene* 28: 3429-3441 [2009年]）。新鮮な腫瘍および腫瘍細胞株の細胞遺伝学および対立形質研究は、染色体3p上のいくつかの異なる領域からのアレル喪失（3p25、3p21-22、3p21.3、3p12-13および3p14を含む）は、広範囲（肺、乳房、腎臓、頭部および頸部、卵巣、子宮頸部、結腸、脾臓、食道、膀胱および他の気管）の主要な上皮癌に関与する、最も早く且つ最も頻繁な遺伝子異常であることを示している。いくつかの腫瘍抑制遺伝子は、染色体3p領域にマッピングされていて、癌腫の発生において、中間部欠失もしくはプロモーター過剰メチル化は、第3番染色体の3pもしくは全体の喪失の前に起こると考えられる（Angeloni D.、*Briefings Functional Genomics* 6: 19-39 [2007年]）。

【0196】

20

30

40

ダウン症（DS）を有する新生児および児童は、多くの場合、先天的一過性白血病を示し、急性骨髄性白血病および急性リンパ球性白血病の増大した危険性を有する。約300の遺伝子を有する第21番染色体は、白血病、リンパ腫および固形腫瘍において、多くの構造的異常、例えば、転位、欠失および増幅に関与し得る。さらに、第21番染色体上に位置する遺伝子は、腫瘍形成において重要な役割を果たすことが同定されている。体細胞性の数的ならびに構造的な第21番染色体異常は、白血病に関連し、特異的遺伝子（21qに位置す

50

るRUNX1、TMPRSS2およびTFFを含む)は、腫瘍形成において役割を果たす (Fonatsch C Gene Chromosomes Cancer 49: 497-508 [2010年])。

【0197】

1つの態様において、本方法は、遺伝子増幅と腫瘍進化の程度との関連を評価する手段を提供する。増幅および/または欠失と癌のステージまたは等級との間の相関性は、予後的に重要であり得る。なぜなら、このような情報は、遺伝ベースの腫瘍等級の定義(最悪の予後を有するより進行した腫瘍を伴う疾患の将来の経過を予測することが好ましい)に寄与し得るからである。加えて、初期増幅および/または欠失事象に関する情報は、これらの事象をその後の疾患進行の予測として関連づける際に有用であり得る。本方法によって同定される遺伝子増幅および欠失は、他の公知のパラメータ、例えば腫瘍等級、組織学、Brd/Urdラベリング指標、ホルモン状態、リンパ節転移、腫瘍サイズ、生存期間および疫学的研究および生物統計学的研究から得られ得る他の腫瘍性質と関連し得る。例えば、本方法によって試験される腫瘍DNAとしては、異形増殖症、腺管癌(インサイチュー)、ステージI~IIIの癌および転移リンパ節が挙げられ、増幅と欠失とステージの間の関連を同定することを可能にする。得られた関連性は、有効な治療的介入を可能にし得る。例えば、連続的に増幅される領域は、過剰発現遺伝子を含み得、その産物は、治療的に攻撃され得る(例えば、成長因子受容体チロシンキナーゼ、p185^{HER2})。

10

【0198】

本方法は、原発性癌からの核酸対他の部位へ転位した細胞の核酸のコピー数変動を決定することにより、薬物抵抗性に関連する増幅および/または欠失事象を同定するために使用され得る。遺伝子増幅および/または欠失が、核型不安定性の顕在化である場合、これは、迅速に薬物抵抗性を発達させ、化学療法抵抗性の患者由来の原発性腫瘍において、化学療法感受性の患者における腫瘍よりも、多くの増幅および/または欠失が予測される。例えば、特定の遺伝子の増幅が、薬物抵抗性の発達に応答性である場合、この遺伝子をとりまく領域は、化学療法抵抗性の患者の肋膜滲出からの腫瘍細胞において継続的に増幅されているが、原発性腫瘍においては増幅されていないことが予測される。遺伝子増幅および/または欠失と薬物抵抗性の発現との間の関連性の発見は、アジュvant治療の利益を得られるかまたは得られない患者の同定を可能にし得る。

20

【0199】

他の態様において、本方法は、トリヌクレオチド反復配列障害(トリヌクレオチド反復配列伸長によって起こる1組の遺伝障害である)に関連する多型を同定するために使用され得る。トリヌクレオチド伸長は、全てのゲノム配列を通して存在する、不安定なミクロサテライト反復配列の部分集団である。健康な遺伝子において、この反復配列が存在する場合、動的突然変異は、反復配列カウントを増大し得、欠損遺伝子を生じ得る。1つの態様において、本方法は、脆弱X症候群に関連するトリヌクレオチド反復配列を同定するために使用され得る。脆弱X症候群に罹患する患者のX染色体の長腕は、230から4000までのC/Gを含み得る(比較すると、キャリアにおいて60~230の反復配列、非罹患の個体において5~54の反復配列である)。このトリヌクレオチド伸長から生じる染色体不安定性は、精神遅滞、明らかな顔の容貌および雄性における巨睾丸症として臨床的に現れる。第2に、関連するDNAトリプレット反復疾患である脆弱X-E症候群もまた、X染色体上に同定されたが、伸長したCCG反復配列の結果であることが見出された。本方法は、他の反復配列伸長障害(カテゴリーI、IIおよびIIIを含む)に関連するトリヌクレオチド反復配列を同定し得る。カテゴリーI障害としては、特定の遺伝子のタンパク質コード部分におけるCAG反復配列伸長によって起こる、ハンチントン病(HD)および脊髄小脳失調症が挙げられる。カテゴリーII伸長は、不均一な伸長により、より表現型が分かれる傾向にあり、一般的に規模は小さいが、遺伝子のエキソンにおいても見られる。カテゴリーIIIとしては、脆弱X症候群、筋強直性ジストロフィー、2種の脊髄小脳失調症、若年性ミオクローヌス癲癇およびフリードリヒ失調症が挙げられる。これらの疾患は、典型的には、最初の2つの群よりも大きな反復配列伸長によって特徴付けられ、この反復配列は、遺伝子のタンパク質コード領域の外側に配置される。

30

40

50

【0200】

他の態様において、本方法は、典型的にはさもなければ関係しないタンパク質のコード領域におけるCAG反復配列の数の増大によって起こることが公知である、少なくとも10の神経障害に関連するCAGトリヌクレオチド反復配列を同定し得る。タンパク質合成の間、伸長されたCAG反復配列は、中断されない連なったグルタミン残基に翻訳され、ポリグルタミン鎖（「polyQ」）として公知のものを形成する。このようなポリグルタミン鎖は、凝集の増大を受け得る。これらの障害は、遺伝の常染色体優性様式（X連鎖遺伝を示す球脊髄性筋萎縮症を除く）、中年期発症、進行性の経過およびCAG反復配列の数と疾患の重症度および発症年齢との相間によって特徴付けられる。病因遺伝子は、公知のポリグルタミン疾患の全てにおいて、広範に発現する。PolyQ疾患の共通の症状は、通常は、人生の10より後半で作用する、進行性の神経細胞の変性によって特徴付けられる。これらの疾患は、同じ反復コドン（CAG）および同じ症状を共有するが、異なるポリグルタミン疾患についての反復配列は、異なる染色体上に存在する。本方法によって同定され得るpolyQ障害の例としては、非限定的に、DRPLA（歯状核赤核淡蒼球ルイ体萎縮症）、HD（ハンチントン病）、SBMA（球脊髄性筋萎縮症またはケネディ病）、SCA1（脊髄小脳失調症1型）、SCA2（脊髄小脳失調症2型）、SCA3（脊髄小脳失調症3型もしくはマシャド・ジョセフ病）、SCA6（脊髄小脳失調症6型）、SCA7（脊髄小脳失調症7型）、SCA17（脊髄小脳失調症17型）が挙げられる。本方法によって同定され得る非polyQ障害の例としては、FRAXA（脆弱X症候群）、FXTAS（脆弱X関連振戦/失調症候群）、FRAXE（脆弱XE精神遅滞）、FRDA（フリードリヒ失調症）、DM（筋強直性ジストロフィー）、SCA8（脊髄小脳失調症8型）、SCA12（脊髄小脳失調症12型）が挙げられる。
20

【0201】

癌におけるCNVの役割に加え、CNVは、通常複合疾患が増えていることに関連しており、これらとしては、ヒト免疫不全ウイルス（HIV）、自己免疫疾患およびある範囲の神経精神病学的障害が挙げられる。

【0202】

現在までに、多くの研究が、炎症に関する遺伝子におけるCNVと免疫応答性ならびにHIV、喘息、クローン病および他の自己免疫障害との間の関連性を報告している（Fanciulliら、Clin Genet 77: 201-213 [2010年]）。例えば、CCL3L1におけるCNVは、HIV/AIDS易罹患性（CCL3L1、17q11.2欠失）、リウマチ性関節炎（CCL3L1、17q11.2欠失）および川崎病（CCL3L1、17q11.2二重重複）において実施されている；HBD-2におけるCNVは、慢性クローン病（HDB-2、8p23.1欠失）および乾癬（HDB-2、8p23.1欠失）にかかり易くすることが報告されている；FCGR3BにおけるCNVは、全身性紅斑性エリトマトーデスにおいて糸球体腎炎にかかり易くすることが示されている（FCGR3B、1q23欠失、1q23二重重複）、抗好中球細胞質抗体（ANCA）関連血管炎（FCGR3B、1q23欠失）およびリウマチ性関節炎発症の危険性の増大。少なくとも2種の、異なる座においてCNVに関連することが示されている炎症性疾患もしくは自己免疫疾患が存在する。例えば、クローン疾患は、HDB-2において低コピー数に関連しているが、IGRM遺伝子の上流の高頻度の欠失多型（p47免疫関連GTPaseファミリーのメンバーをコードする）にも関連している。FCGR3Bコピー数との関連に加え、SEL易罹患性もまた、低コピー数の補体成分C4を有する対象の間で有意に増大することが報告されている。
30
40

【0203】

GSTM1（GSTM1、1q23欠失）およびGSTT1（GSTT1、22q11.2欠失）の座におけるゲノム欠失とアトピー性喘息の危険性の増大との間の関連は、多くの独立した研究において報告されている。いくつかの態様において、本方法は、炎症性疾患および/または自己免疫疾患に関連するCNVの存在もしくは非存在を決定するために使用され得る。例えば、本方法は、HIV、喘息またはクローン疾患を罹患すると疑われる患者におけるCNVの存在を決定するために使用され得る。このような疾患と関連しているCNVの例としては、非限定的に、17q11.2、8p23.1、1q23および22q11.2における欠失および17q11.2および1q23における二重重複が挙げられる。いくつかの態様において、本方法は、CCL3L1、HBD-2、FCGR3B、GSTM、G50

STT1、C4およびIRGMを含むがこれらに限定されない遺伝子においてCNVの存在を決定するために、使用され得る。

【0204】

デノボのおよび遺伝したCNVといくつかの通常の神経疾患および精神疾患との間の関連性は、自閉症、精神分裂症および癲癇、ならびに神経変性疾患のいくつかの症例（例えば、パーキンソン病、筋萎縮性側索硬化症（ALS）および常染色体優性アルツハイマー病）において報告されている（Fanciulliら、Clin Genet 77: 201-213 [2010年]）。細胞遺伝学的異常は、15q11-q13における二重重複を有する自閉症および自閉症スペクトル障害（ASD）を有する患者において観察されている。自閉症ゲノムプロジェクトコンソーシアムにしたがい、いくつかの再発性CNVを含む154のCNVが、染色体15q11-q13上または新たな遺伝子位置（染色体2p16、1q21を含む）および17p12（ASDと重なるスミス・マゲニス症候群に関連する領域）のいずれかにおいて存在する。染色体16p11.2における再発性の微小欠失もしくは微小二重重複は、シナプスの分化を調節しそしてグルタミン作用性神経伝達物質放出を調節することが公知であるSHANK3（22q13.3欠失）、ニューレキシン1（NRXN1、2p16.3欠失）およびニューログリン（neuroglin）（NLGN4、Xp22.33欠失）のような遺伝子についての座において、デノボのCNVが検出される観察に、脚光を当てる。統合失調症はまた、多数のデノボのCNVに関連している。統合失調症に関連している微小欠失および微小二重重複は、神経発生経路およびグルタミン作動性経路に属する遺伝子の過剰出現を含み、このことは、これらの遺伝子に影響を及ぼす多数のCNVが、統合失調症の病理学に直接的に寄与し得ることが示唆される（例えば、ERBB4、2q34欠失、SLC1A3、5p13.3欠失；RAP EGF4、2q31.1欠失；CIT、12.24欠失；およびデノボのCNVを有する多数の遺伝子）。CNVはまた、以下に挙げられる他の神経障害にも関連している：癲癇（CHRNA7、15q13.3欠失）、パーキンソン病（SNCA 4q22二重重複）およびALS（SMN1、5q12.2.-q13.3欠失；およびSMN2欠失）。いくつかの態様において、本方法は、神経系の疾患に関連するCNVの存在または非存在を決定するために使用され得る。例えば、本方法は、自閉症、統合失調症、癲癇、神経変性疾患（例えばパーキンソン病、筋萎縮性側索硬化症（ALS）または常染色体優性のアルツハイマー病）に罹患していることが疑われる患者においてCNVの存在を決定するために使用され得る。本方法は、神経系の疾患（非限定的に、自閉症スペクトル障害（ASD）、統合失調症および癲癇のいずれかを含む）に関連している遺伝子のCNV、ならびに神経変性障害、例えばパーキンソン病に関連している遺伝子のCNVを決定するために使用され得る。このような疾患に関連しているCNVの例としては、非限定的に、15q11-q13、2p16、1q21、17p12、16p11.2および4q22における二重重複、および22q13.3、2p16.3、Xp22.33、2q34、5p13.3、2q31.1、12.24、15q13.3および5q12.2における欠失が、挙げられる。いくつかの態様において、本方法は、SHANK3、NLGN4、NRXN1、ERBB4、SLC1A3、RAPGEF4、CIT、CHRNA7、SNCA、SMN1およびSMN2を含むがこれらに限定されない遺伝子において、CNVの存在を決定するために使用され得る。

【0205】

代謝形質および心臓血管形質（例えば、家族性高コレステロール血症（FH）、アテローム性動脈硬化症および冠状動脈疾患）とCNVとの間の関連性は、多くの研究において報告されている（Fanciulliら、Clin Genet 77: 201-213 [2010年]）。例えば、生殖細胞系再編成、主に欠失は、他のLDLR突然変異を有さない幾人かのFH患者において、LDLR遺伝子（LDLR、19p13.2欠失/二重重複）にて観察されている。別の例は、アポリポタンパク質(a)（apo(a)）をコードするLPA遺伝子であり、その血漿濃度は、冠状動脈疾患、心筋梗塞（MI）および脳卒中の危険性に関連する。リボタンパク質Lp(a)を含むapo(a)の血漿濃度は、個体間で1000倍以上変動し、この変動性の90%が、LPA座において、血漿濃度およびLp(a)イソ型の大きさ（「クリングル4」反復配列の非常に変動する数（5～50の範囲）に対し比例する）により、遺伝的に決定される。これらのデータは、少なくとも2つの遺伝子におけるCNVが、心臓血管についての危険性に関連し得ることを示す。本方法は、CNVの心臓血管障害との関連性について特異的に探索する大型の研究において使用され得る。いくつかの態様において、本方法は、代謝疾患もしくは心臓血管疾患に関連するCNVの存在または

10

20

30

40

50

非存在を決定するために使用され得る。例えば、本方法は、家族性高コレステロール血症を罹患すると疑われる患者においてCNVの存在を決定するために使用され得る。本方法は、代謝疾患または心臓血管疾患（例えば、高コレステロール血症）に関連する遺伝子のCNVを決定するために使用され得る。このような疾患に関連するCNVの例としては、非限定的に、LDLR遺伝子の19p13.2欠失/二重重複およびLPA遺伝子の多重重複が挙げられる。

【0206】

配列決定

種々の態様において、本明細書において記載される方法は、クローン増幅されたDNAテンプレートもしくは1つのDNA分子がフローセル内で大規模パラレル様式により配列決定される、次世代配列決定技術（NGS）を使用する（例えば、Volkerdingら、ClinChem 55: 64 1-658 [2009年]; Metzker M Nature Rev 11: 31-46 [2010年]において記載される）。高処理配列情報に加え、NGSは、デジタル定量可能な情報を提供し、ここでは、各配列読み取りが、個々のクローンのDNAテンプレートまたは1DNA分子を表す、可算「配列タグ」である。NGSの配列決定技術としては、パイロシーケンス、リバーシブル染料ターミネーターを用いた合成による配列決定、オリゴヌクレオチドプローブ連結による配列決定ならびに実時間配列決定が挙げられる。

【0207】

種々の態様において、増幅されていないか、または部分的にしか増幅されていない（標的増幅）サンプルを解析可能である。いくつかの場合、胎児画分を決定する方法は、いかなる種類の標的増幅も必要とせずに達成され得る。

【0208】

配列決定手順の一部分として起こる全ゲノム増幅は、ますます良い有効範囲を提供するために増大した配列決定サイクル数によってまかない得る十分なコピーを提供する。

【0209】

好みしい態様において、2つの異なるゲノムに由来するDNA分子の混合物を含むサンプルは、全ゲノム配列決定の前に、全ゲノム配列について非特異的に富化されている（すなわち、配列決定の前に、全ゲノム増幅が行われる）。

【0210】

サンプルDNAの非特異的富化は、サンプルのゲノムDNA断片の全ゲノム増幅と呼ぶことができ、これは、配列決定によって多型を同定する前にサンプルDNAのレベルを増大するために使用され得る。非特異的富化は、サンプル中に存在する2つのゲノムのうちの1つの選択的富化であってもよい。例えば、非特異的富化は、母性サンプルにおける胎児ゲノムに選択的であってもよく、これは、サンプル中の胎児DNA対母性DNAの相対比を増大させるための公知の方法によって得られ得る。あるいは、非特異的富化は、サンプル中に存在する両ゲノムの非選択的増幅であってもよい。例えば、非特異的増幅は、胎児ゲノムおよび母性ゲノム由来のDNAの混合物を含むサンプル中の、胎児DNAおよび母性DNAのものであってもよい。全ゲノム増幅のための方法は、当技術分野において公知である。縮重オリゴヌクレオチドプライマーによるPCR（DOP）、プライマー伸長PCR技術（PEP）および多置換増幅（MDA）は、全ゲノム増幅方法の例である。いくつかの態様において、異なるゲノム由来のcfDNAの混合物を含むサンプルは、この混合物中に存在するゲノムのcfDNAについて富化されない。他の態様において、異なるゲノム由来のcfDNAの混合物を含むサンプルは、サンプル中に存在するゲノムの任意の1つについて非特異的に富化されている。

【0211】

他の態様において、サンプル中のcfDNAは、特異的に富化される。特異的富化は、DNAサンプルを配列決定する前の増幅のために選択されている特異的配列、例えば多型標的配列についての、ゲノムサンプルの富化をいう。しかし、本開示の態様の利点は、標的増幅が必要ないことである。多型

【0212】

配列決定技術のいくつかは、以下に記載されるように、Affymetrix Inc. (Sunnyvale, CA) 製のハイブリダイゼーションによる配列決定プラットフォーム、ならびに454 Life S

10

20

30

40

50

ciences (Bradford, CT)、Illumina/Solexa (Hayward, CA) およびHelicos Biosciences (Cambridge, MA) 製の合成による配列決定プラットフォーム、ならびにApplied Biosystems (Foster City, CA) 製の連結による配列決定プラットフォームなどのように、市販で入手可能である。Helicos Biosciencesの合成による配列決定を用いて行った1分子配列決定に加え、他の1分子配列決定技術が、本開示の方法によって包含され、これらとしては、Pacific BiosciencesのSMRT (商標) 技術、Ion Torrent (商標) 技術および例えはOxford Nanopore Technologiesによって開発されているナノポア配列決定が挙げられる。

【0213】

自動化サンガー法は、「第1世代」の技術と考えられるが、自動化サンガー配列決定を含むサンガー配列決定もまた、本開示の方法によって使用され得る。核酸画像技術現像 (例えば、原子間力顕微鏡法 (AFM) または透過電子顕微鏡法 (TEM)) の使用を含むさらなる配列決定方法もまた、本開示の方法によって包含される。例示的な配列決定技術は、以下に記載される。

【0214】

1つの態様において、本開示の方法において使用されるDNA配列決定技術は、Helicos True Single Molecule Sequencing (tSMS) (例えば、Harris T.D.ら、Science 320: 106-109 [2008年]において記載される) である。tSMS技術において、DNAサンプルは、約100~200ヌクレオチドの鎖に分裂され、各DNA鎖の3'末端に、ポリA配列が加えられる。各鎖は、蛍光標識アデノシンヌクレオチドの付加によって標識される。次いで、このDNA鎖は、フローセル表面上に固定化されている何百万ものオリゴT捕捉部位を含むフローセルに、ハイブリダイズされる。テンプレートは、約1億テンプレート/cm²の密度であり得る。次いで、このフローセルは、装置 (例えば、Heliscope (商標) シーケンサー) 内にローディングされ、レーザーがこのフローセルの表面を照射し、各テンプレートの位置を明らかにする。CCDカメラが、フローセル表面上にこのテンプレートの位置をマッピングし得る。次いで、テンプレート蛍光標識は、分裂させられて、洗されて除かれる。配列決定反応は、DNAポリメラーゼおよび蛍光標識ヌクレオチドの導入によって始まる。オリゴT核酸は、プライマーとして働く。ポリメラーゼは、標識ヌクレオチドをテンプレート内のプライマーに、直接様式で組み込む。ポリメラーゼおよび組み込まれなかったヌクレオチドは、除かれる。蛍光標識ヌクレオチドの直接組み込みを有するテンプレートは、フローセル表面の画像化によって認められる。画像化後、分裂工程が、蛍光標識を取り除き、この手順は、所望の読み取り長が達成されるまで、他の蛍光標識ヌクレオチドについて繰り返される。配列情報は、各ヌクレオチド付加工程にて収集される。1分子配列決定技術による全ゲノム配列決定は、配列決定ライブラリーの調製におけるPCRベースの増幅を排除し、サンプル調製の直接性は、サンプルのコピーを測定するよりも、直接的なサンプルの測定を可能にする。

【0215】

1つの態様において、本開示の方法において使用されるDNA配列決定技術は、454配列決定 (Roche) (例えば、Margulies, M.ら、Nature 437: 376-380 (2005年)において記載される) である。454配列決定は、2つの工程を含む。第1の工程において、DNAは、約300~800塩基対の断片に剪断され、この断片は、末端平滑化される。次いで、オリゴヌクレオチドアダプターが、断片の末端に連結される。アダプターは、断片の増幅および配列決定のためのプライマーとして働く。断片は、DNA捕捉ビーズ、例えばストレプトアビジン被膜ビーズに、例えば5'ビオチンタグを有するアダプターBを用いて付着し得る。ビーズに付着した断片は、油-水エマルジョンの液滴内でPCR増幅される。結果は、各ビーズ上のクローン増幅されたDNA断片の多数のコピーである。第2工程において、ビーズは、(ピコリットルサイズの) ウェル内に捕捉される。パイロシーケンシングは、各DNA断片で並行して行われる。1つまたは複数のヌクレオチドの付加が、光信号を発生させ、これは、配列決定機器におけるCCDカメラによって記録される。信号強度は、組み込まれたヌクレオチド数に比例する。パイロシーケンシングは、ヌクレオチド付加の際に放出されるピロリン酸 (PPI) を使用する。PPIは、アデノシン5'リン酸の存在下でATPスルフリラーゼによってA

10

20

30

40

50

TPに変換される。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、この反応は、光を発生させ、この光は、認識されて解析される。

【0216】

1つの態様において、本開示の方法において使用されるDNA配列決定技術は、SOLiD（商標）技術（Applied Biosystems）である。SOLiD（商標）連結による配列決定において、ゲノムDNAは、断片に剪断され、この断片の5'末端および3'末端にアダプターが付着して、フラグメントライブラリーを作製する。あるいは、断片の5'末端および3'末端へのアダプターの連結によって断片を環状化させ、この環状化断片を消化して内部アダプターを作製することによって、内部アダプターが導入され得、得られた断片の5'末端および3'末端にアダプターを付着させて、得られた断片によりメイトペアライブラリーを作製し得る。次に、クローン化ビーズ集団が、ビーズ、プライマー、テンプレートおよびPCR成分を含むマイクロリアクターにおいて調製される。PCR後、テンプレートは変性され、ビーズは富化されて、伸ばされたテンプレートを有するビーズが分離される。選択されたビーズ上のテンプレートは、スライドガラスへの結合を可能にする3'修飾を施される。配列は、連続的ハイブリダイゼーションおよび中央決定塩基（または塩基の対）での部分的にランダムなオリゴヌクレオチドの連結によって決定され得、これは、特異的フルオロフォアによって同定される。色が記録された後、連結されたオリゴヌクレオチドは、分裂させられ、除かれて、手順は繰り返される。

【0217】

1つの態様において、本開示の方法において使用されるDNA配列決定技術は、Pacific Biosciencesの1分子実時間（SMRT（商標））配列決定技術である。SMRT配列決定において、染料標識ヌクレオチドの連続的組み込みが、DNA合成の間画像化される。1つのDNAポリメラーゼ分子が、個々の0モード波長アイデンティファイア（ZMWアイデンティファイア）の底表面に付着し、このZMWアイデンティファイアが、リン酸基連結されたヌクレオチドが延長中のプライマー鎖に組み込まれる間に、配列情報を得る。ZMWは、ZMWの外側に迅速に（マイクロ秒で）拡散する蛍光ヌクレオチドの背景に対して、DNAポリメラーゼによる1つのヌクレオチドの組み込みの観察を可能にする閉じ込め構造である。ヌクレオチドを延長する鎖に組み込むために、数ミリ秒かかる。この間に、蛍光標識は励起し、蛍光信号を発して、蛍光タグが分裂する。染料の対応する蛍光の同定は、どの塩基が組み込まれたかを示す。手順は繰り返される。

【0218】

1つの態様において、本開示の方法において使用されるDNA配列決定技術は、ナノポア配列決定（例えばSoni GVおよびMeller A. ClinChem 53: 1996-2001 [2007年]において記載される）である。ナノポア配列決定DNA解析技術は、Oxford Nanopore Technologies（Oxford, United Kingdom）を含む多くの会社によって産業的に開発されている。ナノポア配列決定は、1分子配列決定技術であり、ここで、1分子のDNAが、ナノポアを通過する際に直接的に配列決定される。ナノポアは、直径1ナノメートルの桁の小さな穴である。ナノポアの導伝性流体内への浸漬およびこれを通した電位（電圧）の適用は、ナノポアを通るイオンの伝導に起因する僅かな電流を生じる。流れる電流の量は、ナノポアの大きさおよび形状に感受性である。DNA分子がナノポアを通過する際、DNA分子上の各ヌクレオチドが、ナノポアを異なる程度で塞ぎ、ナノポアを通過する電流の大きさを、異なる程度で変える。したがって、DNA分子がナノポアを通過する際の電流におけるこの変化は、DNA配列の読み取りを表す。

【0219】

1つの態様において、本開示の方法において使用されるDNA配列決定技術は、化学的感受性電界効果トランジスタ（chemFET）アレイである（例えば、2007年12月17日出願の米国特許出願第2009/0026082号に記載される）。この技術の1つの例において、DNA分子は、反応チャンバ内に置かれ得、そしてテンプレート分子は、ポリメラーゼに結合した配列決定プライマーにハイブリダイズし得る。配列決定プライマーの3'末端における新たな核酸鎖への1つまたは複数の三リン酸の組み込みは、chemFETにより、電流における変化によって

10

20

30

40

50

認識され得る。1つのアレイは、多数のchemFETセンサを有し得る。別の例において、1つの核酸は、ビーズに付着し得、そしてこの核酸は、ビーズ上で増幅され得、そして個々のビーズは、chemFETアレイ上の個々の反応チャンバ（それぞれのチャンバがchemFETセンサを有する）に移送され得て、核酸は配列決定され得る。

【0220】

1つの態様において、本開示の方法において使用されるDNA配列決定技術は、透過電子顕微鏡法（TEM）を用いるHalcyon Molecularの方法である。この方法は、個別分子配置迅速ナノ輸送（Individual Molecule Placement Rapid Nano Transfer（IMPRNT））と呼ばれ、重原子マーカーを用いて選択的に標識された高分子量（150kb以上）DNAの1原子解析透過電子顕微鏡画像化を用いることおよびこれらの分子を超薄フィルム上に超高密度（鎖間3nm）平行アレイ内に、一定の塩基間隔で並べることを含む。電子顕微鏡は、フィルム上の分子を画像化し、重原子マーカーの位置を決定して、DNAから塩基配列情報を引き出すために使用される。この方法は、PCT特許公開WO 2009/046445においてさらに記載される。この方法は、10分間未満での完全なヒトゲノムの配列決定を可能にする。

10

【0221】

1つの態様において、DNA配列決定技術は、Ion Torrent1分子配列決定で、これは、半導体技術を単純な配列決定化学と組み合わせて、コードされた情報（A、C、G、T）をデジタル情報（0、1）に半導体チップ上で化学的に直接翻訳する。天然において、ヌクレオチドがポリメラーゼによってDNAの鎖に組み込まれる場合、水素イオンが副産物として放出される。Ion Torrentは、微細加工ウェルの高密度アレイを使用して、大規模平行方法によって、この生化学的手順を行う。各ウェルは、異なるDNA分子を保持する。ウェルの下は、イオン感受性層であり、その下は、イオンセンサである。ヌクレオチド、例えばCが、DNAテンプレートに添加されて、DNAの鎖に組み込まれる場合、水素イオンが放出される。このイオンからの電荷は、溶液のpHを変化させ、これが、Ion Torrentのイオンセンサに同定され得る。シーケンサー（基本的に世界最小のソリッドステートpHメータ）は、塩基を読み上げ、化学的情報からデジタル情報へ直接的に変える。次いで、Ion personal Genome Machine（PGM（商標））シーケンサーは、順次、ヌクレオチドで次々にチップを覆う。チップを覆う次のヌクレオチドが、一致しない場合。電圧変化は記録されず、塩基は読み上げられない。DNA鎖上に2つの同一の塩基が存在する場合、電圧は倍加し、チップは、2つの同一の塩基が呼ばれたことを記録する。直接同定は、数秒でのヌクレオチド組み込みの記録を可能にする。

20

【0222】

いくつかの態様において、本方法は、サンプルヌクレオチド配列を同定するかまたはマッピングする前にこれを増幅するために、PCRまたは関連の技術を使用する。しかし、本明細書において開示されるアルゴリズム技術は、一般的に、ゲノム画分を推定するために増幅を必要とせず、特に使用された多型の標的増幅を必要としない。

30

【0223】

特定の態様は、デジタルPCRおよびハイブリダイゼーションによる配列決定を使用する。デジタルポリメラーゼ連鎖反応（デジタルPCRもしくはdPCR）は、サンプル中の核酸を直接的に同定しそして定量するために使用され得る。デジタルPCRは、エマルジョン内で行われ得る。個々の核酸は、例えば、微少流体チャンバデバイス内に分けられており、各核酸は、PCRによって個々に増幅される。核酸は、平均約0.5個の核酸/ウェルが存在するように分けられてもよく、または1核酸/ウェルを超えないように分けられてもよい。胎児アレルと母性アレルとを識別するために、異なるプローブが、使用され得る。アレルはコピー数を決定するために列挙され得る。ハイブリダイゼーションによる配列決定において、ハイブリダイゼーションは、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含み、ここで、複数のポリヌクレオチドプローブのそれぞれは、基板上に繫留されていてもよい。基板は、公知のヌクレオチド配列のアレイを含む平坦表面であってもよい。アレイに対するハイブリダイゼーションのパターンは、サンプル中に存在するポリヌクレオチド配列を決定するために使用され得る。他の態様において、各

40

50

プローブは、ビーズ、例えば磁気ビーズなどに繫留されている。ビーズへのハイブリダイゼーションは、同定され得、そしてサンプル内の複数のポリヌクレオチド配列を同定するために使用され得る。

【0224】

1つの態様において、本方法は、Illuminaの合成による配列決定およびリバーシブルターミネーターベースの配列決定化学（例えば、Bentleyら、Nature 6: 53-59 [2009年]に記載される）を用いて、何百万ものDNA断片の大規模平行配列決定を使用する。テンプレートDNAは、ゲノムDNA、例えばcfDNAであってもよい。いくつかの態様において、単離された細胞由来のゲノムDNAがテンプレートとして使用され、これは、数百塩基対の長さに断片化される。他の態様において、cfDNAがテンプレートとして使用され、cfDNAは短い断片で存在するため、断片化は必要ではない。例えば、胎児cfDNAは、300bp未満の断片として血流を循環し、母性cfDNAは、約0.5Kbと1Kbとの間の断片として循環すると推測されている（Liら、ClinChem, 50: 1002-1011 (2004年)）。Illuminaの配列決定技術は、断片化ゲノムDNAの、平らな、任意で透明な表面（この上に、オリゴヌクレオチドアンカーが結合している）への付着に依存する。テンプレートDNAは、末端修復されて、5'リン酸化平滑末端を作り、Klenow断片のポリメラーゼ活性が、1つのA塩基を平滑リン酸化DNA断片の3'末端に追加するために使用される。この追加は、連結効率を上げるためにその3'末端における1つのT塩基のぶら下がりを有するオリゴヌクレオチドアダプターへの連結のためのDNA断片を、調製する。アダプターオリゴヌクレオチドは、フローセルアンカーに対して相補的である。限界希釈条件下で、アダプター修飾された1本鎖テンプレートDNAが、フローセルに加えられ、アンカーへのハイブリダイゼーションによって固定化される。付着したDNA断片は、伸長されて架橋増幅され、それぞれ約1,000コピーの同じテンプレートを含む何億ものクラスターを有する超高密度配列決定フローセルを作る。1つの態様において、ランダムに断片化されたゲノムDNA、例えばcfDNAは、クラスター増幅を受ける前にPCRを用いて増幅される。あるいは、増幅していないゲノムライブラリー調製が使用され、ランダムに断片化されたゲノムDNA、例えばcfDNAは、クラスター増幅のみを用いて富化される（Kozarewaら、Nature Methods 6: 291-295 [2009年]）。テンプレートは、除去可能な蛍光色素を有するリバーシブルターミネーターを使用する強固な4色DNA合成による配列決定技術を用いて配列決定される。高感度蛍光同定は、レーザー励起および全内部反射光学を用いて達成される。約20～40bp、例えば36bpの短い配列読み取りは、反復配列をマスクした参照ゲノムに対してアラインメントされ、遺伝的相違は、特に開発したデータ分析パイプラインソフトウェアを用いて呼ばれる。第1の読み取りの完了後、テンプレートは、インサイチューで再編成されて、断片の反対の末端からの第2の読み取りを可能にする。したがって、DNA断片の1端または両端の配列決定が、本方法にしたがって使用される。サンプル中に存在するDNA断片の部分的配列決定が行われ、公知の参照ゲノムに対してマッピングされている予め決定されている長さ、例えば36bpの読み取りを含む配列タグが、カウントされる。

【0225】

配列読み取りの長さは、特定の配列決定技術と関連する。NGS方法は、数十塩基対から数百塩基対までサイズが変動する配列読み取りを提供する。本明細書において記載される方法のいくつかの態様において、配列読み取りは、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である。技術的進歩が、500bpを超える1末端読み取りを可能にし、両端読み取りが行われる場合には約1000bp以上の読み取りを可能にすることが、期待される。1つの態様において、配列読み取りは、36bpである。本開示の方法によって使用され得る他の配列決定方法としては、5000bpを超える核酸分子を配列決定可能である1分子配列決定方法が挙げられる。大容量の配列アウトプットが、シーケンサーからの一次画像アウトプットを塩基の鎖に変換する解析パイプラインによって、移送される。一組の統合化アルゴリズムは、中心的な一次データ変換工程を行

う:画像解析、強度スコアリング、塩基呼び出しおよびアラインメント。

【0226】

マッピング

例えば、特定の遺伝子、染色体、アレルまたは他の構造に対してマッピングされたサンプル中の全ての配列を同定することによって、それぞれ同定された配列を1つの箱にマッピングするために、種々のコンピューター方法が使用され得る。多くのコンピューターアルゴリズムが、配列をアラインメントするために存在し、非限定的に、BLAST (Altschulら、1990年)、BLITZ (MPsrch) (Sturrock&Collins、1993年)、FASTA (Person&Lipman、1988年)、BOWTIE (Langmeadら、Genome Biology 10: R25.1-R25.10 [2009年])、またはELAND (Illumina, Inc.、San Diego, CA, USA) が挙げられる。いくつかの態様において、箱の配列は、当技術分野において公知である核酸データベース（非限定的に、GenBank、dbEST、dbSTS、EMBL（欧州分子生物学研究室）およびDDBJ（日本のDNAデータバンク）が挙げられる）において見出される。BLASTまたは類似のツールが、配列データベースに対して同定された配列を探索するために使用され得、探索ヒットは、同定された配列を適切な箱に選別するために使用され得る。

10

【0227】

装置

配列決定データの解析およびそれに由来する診断は、典型的に、規定のアルゴリズムおよびプログラムにしたがって操作するコンピューターハードウェアを使用して行われる。したがって、特定の態様は、1つまたは複数のコンピューターシステムまたは他の処理システム内に保存されたかまたはそれらを介して移送されたデータを含める手順を使用する。本発明の態様はまた、これらの操作を行うための装置に関連する。この装置は、所望の目的のために特別に構築されていてもよく、または、コンピュータープログラムおよび/またはコンピューター内に保存されたデータ構造によって選択的に作動させたかもしくは再設計された一般目的のコンピューター（またはコンピューター群）であってもよい。いくつかの態様において、プロセッサー群が、いくつかのまたは全ての列挙された解析操作を協同的に（例えば、ネットワークもしくはクラウドコンピューティングを介して）および/または平行して行う。本明細書において記載された方法を行うためのプロセッサーまたはプロセッサー群は、種々の種類のものであってもよく、プログラム可能デバイスのようなマイクロコントローラおよびマイクロプロセッサー（例えば、CPLDおよびFPGA）ならびにゲートアレイASIC、デジタル信号プロセッサー、および/または一般目的のマイクロプロセッサーのような、他のデバイスが挙げられる。

20

【0228】

加えて、特定の態様は、種々のコンピューターによって実施される操作を実行するためのプログラム命令および/またはデータ（データ構造を含む）を含む、実体的および/または非一時的コンピューター読み取り可能媒体もしくはコンピュータープログラム製品に関連する。コンピューター読み取り可能媒体としては、半導体メモリデバイス、ディスクドライブのような磁気媒体、磁気テープ、CDのような光学媒体、光磁気媒体およびプログラム命令を保存し実行するために特別に構築されたハードウェアデバイス（例えば読み取り専用メモリデバイス（ROM）およびランダムアクセスメモリ（RAM））が挙げられるが、これらに限定されない。コンピューター読み取り可能媒体は、エンドユーザーによって直接的に制御されてもよく、または媒体は、エンドユーザーによって間接的に制御されてもよい。直接的に制御された媒体の例としては、ユーザー機能に配置された媒体および/または他のエンティティと共有しない媒体が挙げられる。間接的に制御される媒体の例としては、外部ネットワークを介しておよび/またはサービス提供共有リソース（「クラウド」など）を介してユーザーに間接的にアクセス可能な媒体が挙げられる。プログラム命令の例としては、機械コード（コンパイラによって作成される）およびインターフリタを用いるコンピューターによって実行され得るより高いレベルのコードを含むファイルを含む。

30

【0229】

1つの態様において、コンピュータープログラム製品は、規定のゲノム（例えば、胎児

40

50

のゲノム)に由来する核酸の画分および任意で他の情報(試験サンプル中の胎児の異数性の存在または非存在など)を示すアウトプットを作成するために提供される。コンピューター製品は、特定の生物由来の核酸の画分を決定するための上述の方法の任意の1つまたは複数を行うための命令を含み得る。説明されたように、コンピューター製品は、プロセッサーがゲノム画分を決定することを可能にするためにその上に記録されたコンピューター実行可能な理論もしくはコンピューターコンパイル可能な理論(例えば、命令)、ならびに、いくつかの場合においては、ゲノムにおいて異数性もしくは他の状態が存在するか非存在であるかを有する、非一時的および/または有形コンピューター読み取り可能媒体を含み得る。1つの例において、コンピューター製品は、プロセッサーが胎児画分を決定し胎児異数性を診断することを可能にする、その上に記録されたコンピューター実行可能な理論またはコンピューターコンパイル可能な理論(例えば、命令)を有するコンピューター読み取り可能媒体を含み、この理論は、以下を含む:母性生体サンプル由来の核酸分子の少なくとも一部分からの配列決定データ(この配列決定データは、1つまたは複数の多型の座における配列を含む)を受け取るための受け取り手順;配列を解析し、1つまたは複数の多型についてアレルカウントを決定し、母性生体サンプルにおける核酸の胎児画分を決定するためのコンピューター補助理論;およびサンプル中の核酸の胎児画分を示すアウトプットを作成するためのアウトプット手順。

【0230】

検討中のサンプルからの配列情報は、記載されるように多型参照配列に対してマッピングされ得る。さらに、マッピングされた配列情報は、アレルカウントを作成するためおよび/または多型についての接合生殖性事例を決定するために使用され得る。このような情報は、胎児画分を決定するために使用され得る。種々の態様において、多型参照配列は、データベース(例えば、リレーションナルデータベースまたはオブジェクトデータベースなど)に保存される。ほとんどの場合、助けのないヒトが、これらのコンピューター操作のいずれか1つまたは全てを実施することは実際的でない、または不可能ですらあることが、理解されるべきである。例えば、サンプル由来の1つの30bpの読み取りを多型参照配列のデータベースにマッピングすることは、コンピューター機器の助けなしでは、もしかすると法外に長い期間がかかると考えられる。信頼できる呼び出しが、多くの場合、1つまたは複数の染色体に対し、数千(例えば、少なくとも約10,000)または何百万の読み取りを必要とするので、無論、問題は、いっそうひどくなる。

【0231】

特定の態様において、本開示の方法は、解析した核酸配列を產生した生物についての参照多型に関する保存されたリストまたは他の編成されたデータの収集物を使用する。上で説明されたように、検討下のサンプル由来の配列は、保存された多型に対してアラインメントされ得るか、さもなければマッピングされ得る。個々の多型は、典型的に、核酸サンプルから同定された配列を明確にマッピングするために充分な長さの配列である。典型的に、多型は、各アレルについて1つの群に入る。種々の態様において、参照多型は、多型の配列に加えてその特徴を含むデータベースに保存される。多型に関するこの情報の収集物は、例えば、リレーションナルデータベースまたはオブジェクトデータベースに保存され得る。

【0232】

図10は、適切に構築されたかまたは設計された場合、本発明の解析機器として働くことが可能な、典型的なコンピューターシステムを説明する。コンピューターシステム200は、一次記憶装置206(典型的にランダムアクセスメモリもしくはRAMである)、一次記憶装置204(典型的には読み取り専用メモリもしくはROMである)を含む記憶装置デバイスと組み合わされた、任意の数のプロセッサー202(中央処理ユニットもしくはCPUとも呼ばれる)を備える。CPU 202は、マイクロコントローラーおよびマイクロプロセッサーを含むプログラム可能デバイス(例えば、CPLDおよびFPGA)ならびにプログラム不可能なデバイス(例えば、ゲートアレイASICまたは一般的目的のマイクロプロセッサー)などの種々の種類であり得る。当技術分野において周知であるように、一次記憶装置204は、データおよ

び命令をCPUに移すように働き、一次記憶装置206は、典型的には、双方向性の様式でデータおよび命令を移すために使用される。これらの一次記憶装置デバイスの両方は、上で記載されたような、任意の好適なコンピューター読み取り可能媒体を備え得る。大量記憶装置デバイス208もまた、双方向的に、CPU 202と組み合わされて、さらなるデータ保存容量を提供し、上述のコンピューター読み取り可能媒体のいずれかを備え得る。大量記憶装置208は、プログラム、データなどを保存するために使用され得、典型的に、二次記憶装置媒体、例えばハードディスクである。大量記憶装置208内に保持される情報は、適切な場合、標準的様式で、仮想メモリである一次記憶装置206の一部分として、組み込まれ得ることが理解される。CD-ROM 214のような特定の大量記憶装置デバイスもまた、データを一方的にCPUに渡し得る。

10

【 0 2 3 3 】

CPU 202はまた、1つまたは複数のインプット/アウトプットデバイス（例えば、ビデオモニタ、マウス、キーボード、マイクロホン、接触感知式ディスプレイ、トランステューサーカードリーダー、タブレット、スタイルス、音声または筆記認識機または他の周知のインプットデバイス、例えば、無論、他のコンピューター）に接続するインターフェース210と組み合わされる。最後に、CPU 202は、一般的に212において示される外部接続を用いて、外部デバイス、例えばデータベースまたはコンピューターもしくはテレコミュニケーションネットワークと組み合わせられてもよい。このような接続により、CPUは、本明細書において記載される方法工程を行う過程において、ネットワークから情報を受け取り得るか、またはネットワークに対し情報をアウトプットし得ることが企図される。

20

【 0 2 3 4 】

配列または他のデータが、直接的にまたは間接的に、ユーザーによってコンピューターにインプットされ得る。1つの態様において、コンピューターシステム200は、増幅された核酸の配列を読み取りおよび/または解析する配列決定ツールと直接的に組み合わされる。配列またはこのようなツールからの他の情報は、システム200による解析のために、インターフェース212を介して提供される。あるいは、システム200によって処理される配列は、配列記憶装置供給源、例えばデータベースまたは他のリポジトリから提供される。一旦処理装置200に入ると、一次記憶装置206または大型記憶装置208のようなメモリデバイスは、少なくとも一時的に、核酸の配列をバッファーに入れるかまたは保存する。加えて、メモリデバイスは、種々の染色体または遺伝子についてのタグ数、計算されたコピーカウントなどを保存し得る。メモリはまた、提示された配列またはマッピングされたデータを解析するための種々のルーチンおよび/またはプログラムを保存し得る。このようなプログラム/ルーチンとしては、統計学的解析を行うためのプログラムなどが挙げられ得る。

30

【 0 2 3 5 】

1つの例において、ユーザーは、サンプルを配列決定装置に提供する。データは、コンピューターに接続された配列決定装置によって収集されおよび/または解析される。コンピューター上のソフトウェアは、データ収集および/または解析を可能にする。データは、保存されても、（モニタまたは他の類似のデバイスを介して）表示されても、および/または別の場所に送られてもよい。示されるように、コンピューターは、インターネットに接続され得、インターネットは、データを、遠隔地のユーザー（例えば、臨床医、科学者または解析者）によって使用されるハンドヘルドデバイスに伝達するために使用される。データは、伝達の前に保存されても、および/または解析されてもよいことが、理解される。いくつかの態様において、生データは、収集されて、このデータを解析および/または保存する遠隔地のユーザー（または装置）に送られる。伝達は、インターネットを介して起こり得るが、衛星または他の接続を介しても起こり得る。あるいは、データは、コンピューター読み取り可能媒体（例えば、CDまたは半導体メモリ記憶装置デバイス）において保存され得る、この媒体は、（例えば、メールを介して）エンドユーザーに送られ得る。遠隔地のユーザーは、同じまたは異なる地理学的位置（限定されないが、ビル、町、州、国または大陸が挙げられる）にいてもよい。

40

50

【0236】

いくつかの態様において、本発明の方法は、複数のポリヌクレオチド配列に関するデータを収集し、そしてこのデータをコンピューターに送信することをさらに含む。例えば、コンピューターは、研究室の設備、例えば、サンプル収集装置、ヌクレオチド増幅装置、ヌクレオチド配列決定装置、またはハイブリダイゼーション装置と接続され得る。次いで、コンピューターは、研究室デバイスによって集められた、適用可能なデータを収集してもよい。データは、任意の工程（例えば、実時間で収集されている間、送信の前、送信の間または送信と合わせて、または送信後）において、コンピューター上に保存され得る。データは、コンピューターから引き出され得るコンピューター読み取り可能媒体上に保存され得る。収集されたかまたは保存されたデータは、例えば、ローカルネットワークまたは広範囲ネットワーク（例えば、インターネット）を介して、コンピューターから遠隔地へ伝達され得る。

10

【0237】

1つの局面において、本発明は、少なくとも60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、または少なくとも99%の正確さのヌクレオチド配列決定の定量的解析を実施可能なシステムを、さらに提供する。ヌクレオチド配列決定は、サンガー配列決定、大規模平行配列決定、ハイブリダイゼーションまたは本明細書において記載される他の技術を含み得る。このシステムは、種々の構成要素、例えば、研究室設備およびコンピューターシステムを備え得、本明細書において開示される本発明の方法を実施するように構成され得る。

20

【0238】

いくつかの態様において、装置および/またはプログラミング命令は、本方法に関係する情報（例えば、母性試験サンプルを提供したヒト対象についての患者医療記録における胎児DNA画分および任意で胎児染色体異数性の存在または非存在）を自動的に記録するための命令を、さらに含み得る。患者医療記録は、例えば、研究室、臨床医のオフィス、病院、健康維持組織、保険会社またはウェブサイト上の個人的医療記録によって、維持され得る。さらに、プロセッサー実施解析の結果に基づき、本方法は、母性試験サンプルが採取されたヒト対象の処置を処方すること、開始すること、および/または変えることを、さらに含み得る。これは、対象から採取されたさらなるサンプルにおいて、1つまたは複数のさらなる試験または解析を実施することを含み得る。

30

【実施例】

【0239】

配列決定された変動性から予測された胎児画分:事例2

母性サンプルにおける胎児画分を信頼性をもって推定するために本方法が使用され得ることを実証するために、人工的「母性」サンプルを作成し、塩基変動性を、第1および第7番染色体の全ての座において同定して、マイナー寄与ゲノムの画分を予測した。

【0240】

妊娠雌性体から単離されたcfDNAは、母性cfDNAと胎児cfDNAとの混合物であり、胎児cfDNAのレベルは、全cfDNAの約10%の中央値に相当する（Loら、2010年、「Maternal Plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus」、Prenatal Diagnosis, 2, 1-12）。人工的母性サンプルを作製するため、母親およびその息子から得られたゲノムDNA（gDNA）（母親および息子DNA NA10924およびNA10925; The Coriell Institute for Medical Research, Camden, NJ）を使用し、混合ゲノムのサンプルを作製した。母親のgDNAおよび息子のgDNAのそれぞれの5μgを、約200bpの断片に剪断し、それぞれの濃度を決定した。息子からの10%のDNAおよび母親からの90%のDNAを含む人工的サンプルを作製し、母性血液サンプル（在胎月齢に依存して、典型的には、2～40%の胎児cfDNAを含むと考えられる[Lunら、2008年、「Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma」、Clinical Chemistry, 54, 1664-1672]）を偽装した。配列決定ライブラーを、人工的サンプルのDNAから調製し、IlluminaHiSeq 2000を用いて、4レーンのフローセル上で50配列決定サ

40

50

イクルに供した。およそ8億個の49マー配列読み取りが、作製された。

【0241】

約8億の読み取りを、GSNAPアルゴリズム (<http://research-pub.gene.com/gmap/>) を用いて、反復配列をマスクしたヒト参照配列ゲノム (hg19 build) に対してアラインメントし、1つの不一致および挿入なしおよび欠失なしを得た。ゲノム上の多数の位置に対してマッピングされた配列は、無視した。全ての他のマッピングされた読み取りを、配列タグとして数え、40～100の配列タグがマッピングされた座のみを、さらなる解析について検討した。すなわち、40～100のタグの適用範囲を有する塩基のみを、検討した。

【0242】

各塩基座について、4つの塩基のそれぞれについてマッピングされた多くのタグを、数えた。2つより多くの潜在的塩基を有する座を排除し、1アレルおよび2アレル座に対してマッピングされたタグのみを使用して、人工的胎児画分を予測した。各塩基座においてマッピングされたタグの総数を、その座において、適用範囲(D)で表した。この模擬母性サンプルにおいて、母親のメジャーアレル(B)の寄与は、タグの90%部分を反映し、息子のマイナーアレル(A)の寄与は、タグの10%部分を反映すると予測される。

【0243】

図11AおよびBは、それぞれ第1および第7番染色体についてのマイナーアレル百分率 (A/D) についての、第1および第7番染色体上の改変塩基の数の観察 (頻度) のヒストグラムを示す。マイナーアレル百分率は、所定の座におけるアレルの総数の百分率である。例えば、所定の座 (マイナーアレルAの8回の出現およびマイナーアレルBの56回の出現) について、マイナーアレル百分率は、8%である。このデータは、マイナーアレルについての最も大きな出現数 (頻度) は、マイナーアレルが、5% (胎児画分の半分を表す) で存在する際に観察されたことを示す。したがって、このデータは、10%の胎児画分を含むサンプルを予測し、これは、人工的母性サンプルを作製するために使用されることと一致する。

【0244】

図12AおよびBは、それぞれ第1および第7番染色体に沿ったアレル頻度の分布を示す。両プロットは、5%のマイナーアレル頻度および95%のメジャーアレル頻度で出現する、染色体に沿った改変アレルの最大数を示す。残りのデータ点のいくつかは、母性ゲノムにおいて存在する両アレル性座を示す。一方で、他は、配列決定方法論のノイズを表す。改変アレルが染色体のセントロメアと重なることを示さない各プロットの中央部分は、染色体の反復豊富領域と考えられ、1つより多くの座においてタグがマッピングされ、それゆえ、解析から除かれる。他の領域、例えばセントロメアに隣接する領域およびテロメアに対応する領域において、改変アレルは、過剰提示される。これらの領域の過剰提示は、配列方法論に帰し得、それにより、いくつかの領域を、他より高いレベルにおいて配列決定した。

【0245】

したがって、本方法は、胎児画分を予測するために使用され得る。標的配列 (例えば、SNP) の同定を必要としないので、この方法は特に有用である。任意の染色体の任意の位置における任意の改変体が、胎児画分百分率を予測するために供され得る。

【0246】

他の態様

上は、特定の手順および装置にしたがい、本発明を一般的に説明するが、本発明は、より広範囲な適用範囲を有する。特に、本発明は、妊娠個体から採取されたDNAサンプルにおける胎児DNAの画分を検出する見地から記載されているが、このように限定されず、本明細書において提示された概念および方法はまた、2種以上の異なるゲノムに起源するDNAを有するサンプルにおいてDNA型の相対量が検出される他の文脈においても、適用され得る。無論、当業者は、他の変更、修飾および改変を認識すると考えられる。

【0247】

例えば、本明細書において記載される例および適用のほとんどは、胎児を孕む個体から採取されたDNAサンプルにおけるDNAの胎児画分の推定を考慮するが、本発明はそのように

10

20

30

40

50

限定されない。より一般的には、種々の態様が、試験サンプル中の2つの異なるゲノム（対象となる1つまたは複数の配列の量において、相違することが公知であるかまたは疑われる）からの核酸の混合物を含む試験サンプル中の2つの異なるゲノムからの核酸の相対量を評価することを提供する。核酸の混合物は、2種以上の細胞に由来する。

【0248】

さらに、本明細書において提示される例のほとんどは、ヒト妊娠から採取されたサンプルを考慮するが、本開示はそのように限定されない。例えば、試験されるサンプルを提供する個体は、ポリヌクレオチド配列を含む生物、例えば植物、昆虫例えばハエ、または動物であってもよい。いくつかの態様において、対象は哺乳動物であり、例えば、マウス、ラット、犬、猿またはヒトである。示されるように、対象は、妊娠個体であってもよい。対象は、癌などの疾患を有する個体であってもよく、または、微生物、例えばウイルスのような異物に感染していてもよい。このサンプルは、対象からの体液（例えば、血液、血漿、血清、唾液、尿、排泄物、膿、リンパ液、粘液など）を含み得る。例えば、サンプルは、母性無細胞DNAおよび胎児無細胞DNAの混合物を含む、母性血漿サンプルであってもよい。一般的に、本開示の方法は、サンプルからのDNAの配列決定；配列読み取りの多型に対するマッピング；接合生殖性に基づく多型の分類；およびサンプルの2次供給源由来のDNAの画分を推定することを含み得る。

10

【0249】

添付資料1 アレル探索データベース配列表

>rs560681.1|Chr.1|length=111|allele=A
 CACATGCACA GCCAGCAACC CTGTCAGCAG GAGTTCCAC CAGTTCTTT
 CTGAGAACAT CTGTTCAGGT TTCTCTCCAT CTCTATTTAC TCAGGTCACA
 GGACCTTGGG G

>rs560681.2|Chr.1|length=111|allele=G
 CACATGCACA GCCAGCAACC CTGTCAGCAG GAGTTCCAC CAGTTCTTT
 CTGAGAACAT CTGTTCAGGT TTCTCTCCAT CTCTGTTTAC TCAGGTCACA
 GGACCTTGGG G

10

>rs1109037.1|Chr.2|length=126|allele=A
 TGAGGAAGTG AGGCTCAGAG GGTAAGAAC TTTGTCACAG AGCTGGTGGT
 GAGGGTGGAG ATTTTACACT CCCTGCCTCC CACACCAGTT TCTCCAGAGT
 GGAAAGACTT TCATCTCGCA CTGGCA

>rs1109037.2|Chr.2|length=126|allele=G
 TGAGGAAGTG AGGCTCAGAG GGTAAGAAC TTTGTCACAG AGCTGGTGGT
 GAGGGTGGAG ATTTTACACT CCCTGCCTCC CACACCAGTT TCTCCGGAGT
 GGAAAGACTT TCATCTCGCA CTGGCA

20

>rs9866013.1|Chr.3|length=121|allele=C
 GTGCCTTCAG AACCTTGAG ATCTGATTCT ATTTTAAAG CTTCTTAGAA
 GAGAGATTGC AAAGTGGGTT GTTTCTCTAG CCAGACAGGG CAGGCAAATA
 GGGGTGGCTG GTGGGATGGGA

>rs9866013.2|Chr.3|length=121|allele=T
 GTGCCTTCAG AACCTTGAG ATCTGATTCT ATTTTAAAG CTTCTTAGAA
 GAGAGATTGC AAAGTGGGTT GTTTCTCTAG CCAGACAGGG CAGGTAAATA
 GGGGTGGCTG GTGGGATGGGA

30

>rs13182883.1|Chr.5|length=111|allele=A
 AGGTGTGTCT CTCTTTGTG AGGGGAGGGG TCCCTCTGG CCTAGTAGAG
 GGCCTGGCCT GCAGTGAGCA TTCAAATCCT CAAGGAACAG GGTGGGGAGG
 TGGGACAAAG G

>rs13182883.2|Chr.5|length=111|allele=G
 AGGTGTGTCT CTCTTTGTG AGGGGAGGGG TCCCTCTGG CCTAGTAGAG
 GGCCTGGCCT GCAGTGAGCA TTCAAATCCT CGAGGAACAG GGTGGGGAGG
 TGGGACAAAG G

40

>rs13218440.1|Chr.6|length=139|allele=A
 CCTCGCCTAC TGTGCTGTT CTAACCATCA TGCTTTCCC TGAATCTCTT
 GAGTCTTTT CTGCTGTGGA CTGAAACTTG ATCCTGAGAT TCACCTCTAG
 TCCCTCTGAG CAGCCTCCTG GAATACTCAG CTGGGATGG
 >rs13218440.2|Chr.6|length=139|allele=G
 CCTCGCCTAC TGTGCTGTT CTAACCATCA TGCTTTCCC TGAATCTCTT
 GAGTCTTTT CTGCTGTGGA CTGAAACTTG ATCCTGAGAT TCACCTCTAG
 TCCCTCTGGG CAGCCTCCTG GAATACTCAG CTGGGATGG

10

>rs4606077.1|Chr.8|length=114|allele=C
 GCAACTCCCT CAACTCCAAG GCAGACACCA AAGCCCTCCC TGCCTGTGGC
 TTTGTAGTTC TAGTGTGGGA TCTGACTCCC CACAGCCCAC CCAAAGCCGG
 GGAACACTCCTC ACTG
 >rs4606077.2|Chr.8|length=114|allele=T
 GCAACTCCCT CAACTCCAAG GCAGACACCA AAGCCCTCCC TGCCTGTGGC
 TTTGTAGTTC TAGTGTGGGA TCTGACTCCC CACAGCCTAC CCAAAGCCGG
 GGAACACTCCTC ACTG

20

>rs7041158.1|Chr.9|length=117|allele=C
 AATTGCAATG GTGAGAGGTT GATGGTAAAA TCAAACGGAA CTTGTTATT
 TGTCAATTCTG ATGGACTGGA ACTGAGGATT TTCAATTCC TCTCCAACCC
 AAGACACTTC TCACCTGG
 >rs7041158.2|Chr.9|length=117|allele=T
 AATTGCAATG GTGAGAGGTT GATGGTAAAA TCAAACGGAA CTTGTTATT
 TGTCAATTCTG ATGGACTGGA ACTGAGGATT TTCAATTCC TTTCCAACCC
 AAGACACTTC TCACCTGG

30

>rs740598.1|Chr.10|length=114|allele=A
 GAAATGCCTT CTCAGGTAAT GGAAGGTTAT CCAAATATT TTCGTAAGTA
 TTTCAAATAG CAATGGCTCG TCTATGGTTA GTCTCACAGC CACATTCTCA
 GAACTGCTCA AACCA
 >rs740598.2|Chr.10|length=114|allele=G
 GAAATGCCTT CTCAGGTAAT GGAAGGTTAT CCAAATATT TTCGTAAGTA
 TTTCAAATAG CAATGGCTCG TCTATGGTTA GTCTCGCAGC CACATTCTCA
 GAACTGCTCA AACCA

40

>rs10773760.1|Chr.12|length=128|allele=A

ACCCAAAACA CTGGAGGGGC CTCTTCTCAT TTTCGGTAGA CTGCAAGTGT
 TAGCCGTCGG GACCAGCTTC TGTCTGGAAG TTCGTCAAAT TGCAGTTAAG
 TCCAAGTATG CCACATAGCA GATAAGGG
 >rs10773760.2|Chr.12|length=128|allele=G
 ACCCAAACA CTGGAGGGGC CTCTTCTCAT TTTCGGTAGA CTGCAAGTGT
 TAGCCGTCGG GACCAGCTTC TGTCTGGAAG TTCGTCAAAT TGCAGTTAGG
 TCCAAGTATG CCACATAGCA GATAAGGG

10

>rs4530059.1|Chr.14|length=110|allele=A
 GCACCAGAAT TTAAACAAACG CTGACAATAA ATATGCAGTC GATGATGACT
 TCCCAGAGCT CCAGAAGCAA CTCCAGCACA CAGAGAGGCG CTGATGTGCC
 TGTCAGGTGC
 >rs4530059.2|Chr.14|length=110|allele=G
 GCACCAGAAT TTAAACAAACG CTGACAATAA ATATGCAGTC GATGATGACT
 TCCCAGAGCT CCAGAAGCAA CTCCAGCACA CGGAGAGGCG CTGATGTGCC
 TGTCAGGTGC

20

>rs1821380.1|Chr.15|length=139|allele=C
 GCCCAGATTA GATGGAACCT TTTCCTCTTT TCCAGTGCAA GACAAGCGAT
 TGAAAGAAGT GGATGTGTTA TTGCGGGCAC AATGGAGCCA CTGAACTGCA
 GTGCAAAAT GCAGTAAGGC ATACAGATAG AAGAAGGAG
 >rs1821380.2|Chr.15|length=139|allele=G
 GCCCAGATTA GATGGAACCT TTTCCTCTTT TCCAGTGCAA GACAAGCGAT
 TGAAAGAAGT GGATGTGTTA TTGCGGGCAC AATGGAGCCA CTGAACTGCA
 GTGCAAAAT GCAGTAAGGG ATACAGATAG AAGAAGGAG
 >rs7205345.1|Chr.16|length=116|allele=C

30

TGACTGTATA CCCCAGGTGC ACCCTTGGGT CATCTCTATC ATAGAACTTA
 TCTCACAGAG TATAAGAGCT GATTCTGTG TCTGCCTCTC ACACTAGACT
 TCCACATCCT TAGTGC
 >rs7205345.2|Chr.16|length=116|allele=G
 TGACTGTATA CCCCAGGTGC ACCCTTGGGT CATCTCTATC ATAGAACTTA
 TCTCACAGAG TATAAGAGCT GATTCTGTG TCTGCCTGTC ACACTAGACT
 TCCACATCCT TAGTGC

40

>rs8078417.1|Chr.17|length=110|allele=C
 TGTACGTGGT CACCAGGGGA CGCCTGGCGC TGCGAGGGAG GCCCCGAGCC
 TCGTGCCCCC GTGAAGCTTC AGCTCCCTC CCCGGCTGTC CTTGAGGCTC

TTCTCACACT

>rs8078417.2|Chr.17|length=110|allele=T

TGTACGTGGT CACCAGGGGA CGCCTGGCGC TGCGAGGGAG GCCCCGAGCC
TCGTGCCCCC GTGAAGCTTC AGCTCCCCTC CCTGGCTGTC CTTGAGGCTC
TTCTCACACT

>rs576261.1|Chr.19|length=114|allele=A

CAGTGGACCC TGCTGCACCT TTCCTCCCCT CCCATCAACC TCTTTGTGC
CTCCCCCTCC GTGTACCACC TTCTCTGTCA CCAACCCTGG CCTCACAAC
CTCTCCTTTG CCAC

10

>rs576261.2|Chr.19|length=114|allele=C

CAGTGGACCC TGCTGCACCT TTCCTCCCCT CCCATCAACC TCTTTGTGC
CTCCCCCTCC GTGTACCACC TTCTCTGTCA CCACCCCTGG CCTCACAAC
CTCTCCTTTG CCAC

>rs2567608.1|Chr.20|length=110|allele=A

20

CAGTGGCATA GTAGTCCAGG GGCTCCTCCT CAGCACCTCC AGCACCTTCC
AGGAGGCAGC AGCGCAGGCA GAGAACCCGC TGGAAGAAC TGCGGAAGTT
GTCGGAGAGG

>rs2567608.2|Chr.20|length=110|allele=A

CAGTGGCATA GTAGTCCAGG GGCTCCTCCT CAGCACCTCC AGCACCTTCC
AGGAGGCAGC AGCGCAGGCA GAGAACCCGC TGGAAGGATC GGCGGAAGTT
GTCGGAGAGG

30

>rs2073383.1|Chr.22|length=140|allele=C

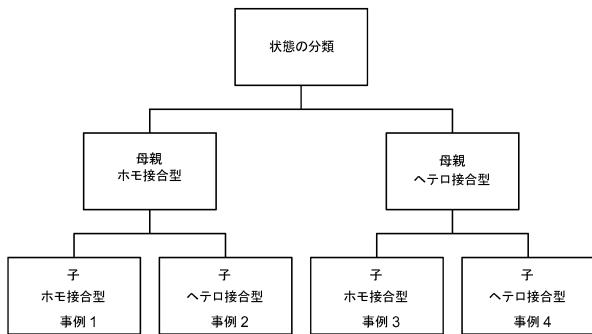
GCTGCAGAAT CCACAGAGCC AGACGCCCGC TGGGCCCGCA GCGCCCCCCT
GCACAAGTGG GGAAACTAGG TCATGGGCC CAGGCAGTGT GGAAGGCAGT
GCAGGAGTTG CCCAGGGCGT GGGGTCTCC AGCCTCAGTG

>rs2073383.2|Chr.22|length=140|allele=T

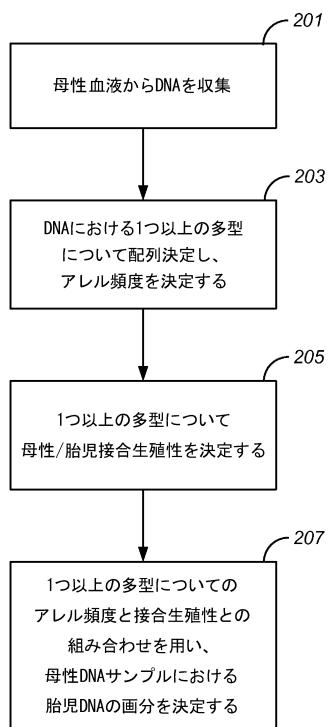
GCTGCAGAAT CCACAGAGCC AGACGCCCGC TGGGCCCGCA GCGCCCCCCT
GCACAAGTGG GGAAACTAGG TCATGGGCC CAGGCAGTGT GGAAGGCAGT
GCAGGAGTTG CCCAGGGGTGT GGGGTCTCC AGCCTCAGTG

40

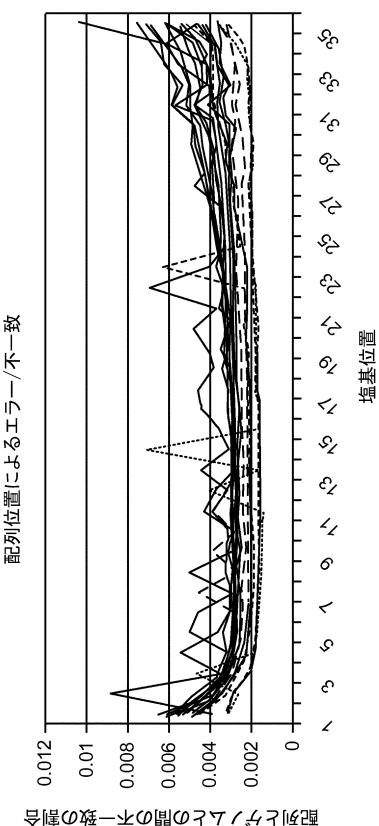
【図1】



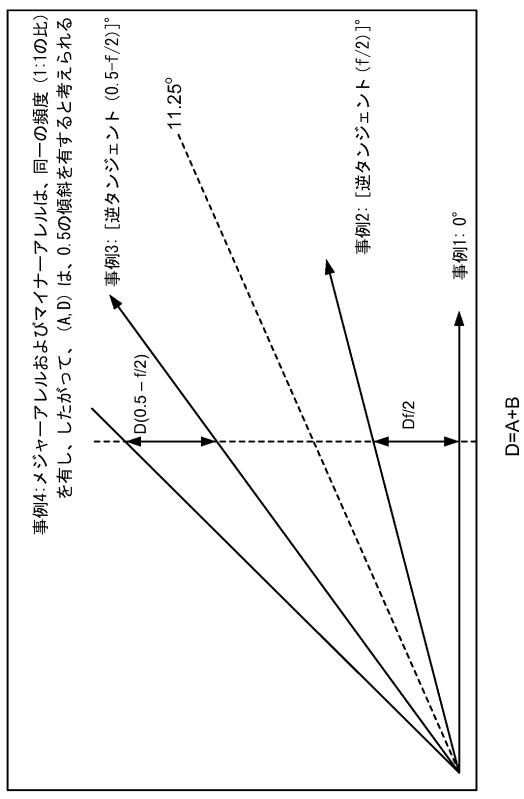
【図2】



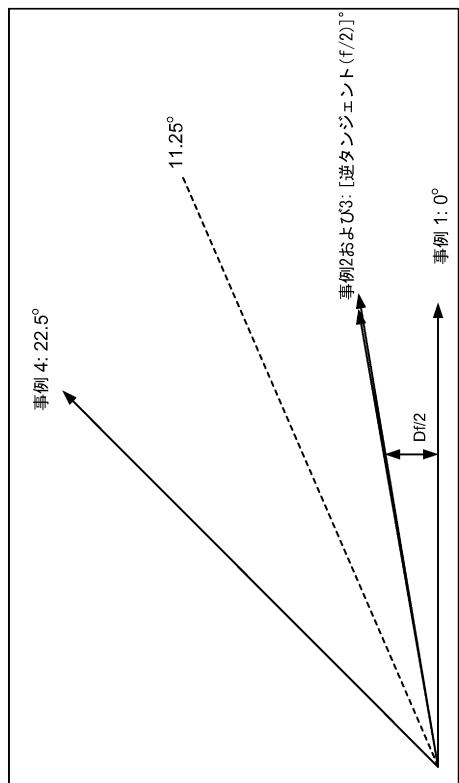
【図3】



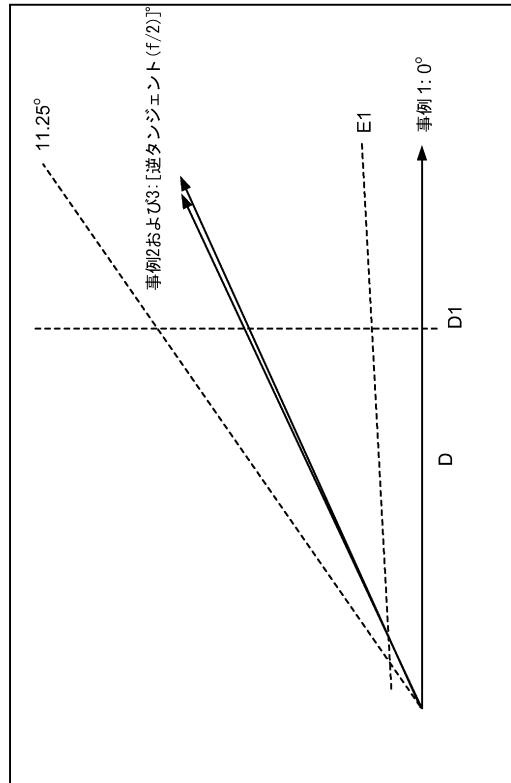
【図4】



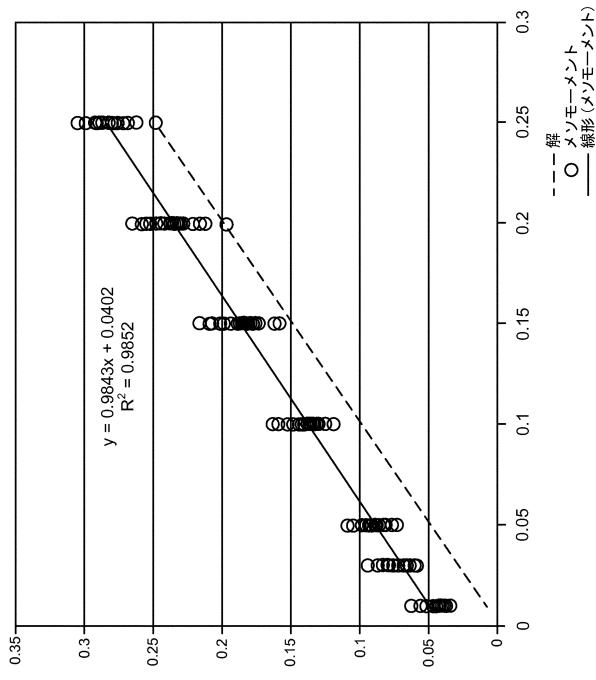
【図5】



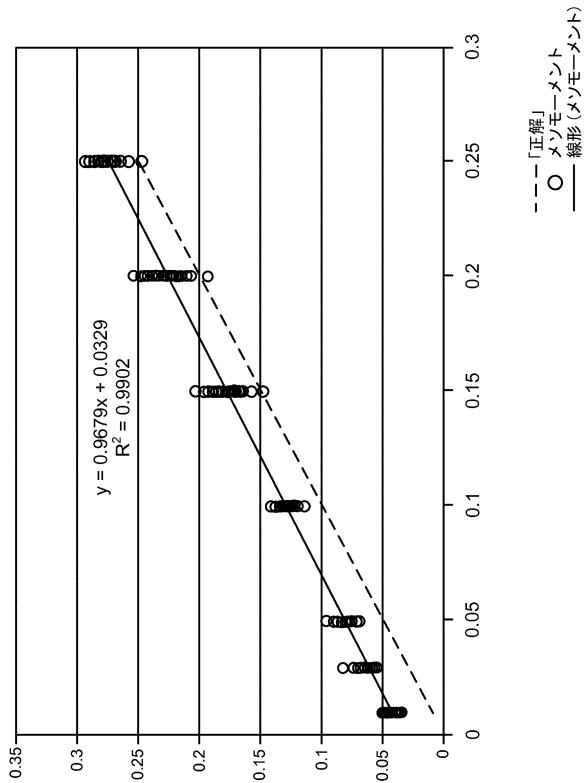
【図6】



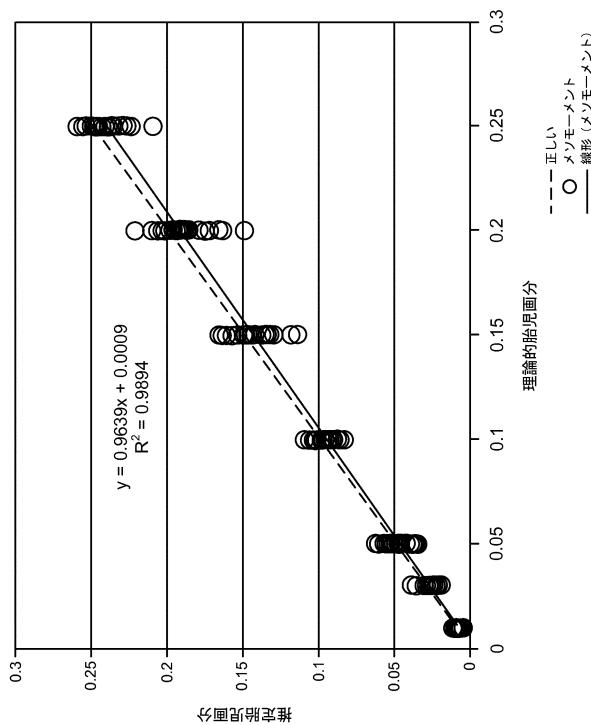
【図7】



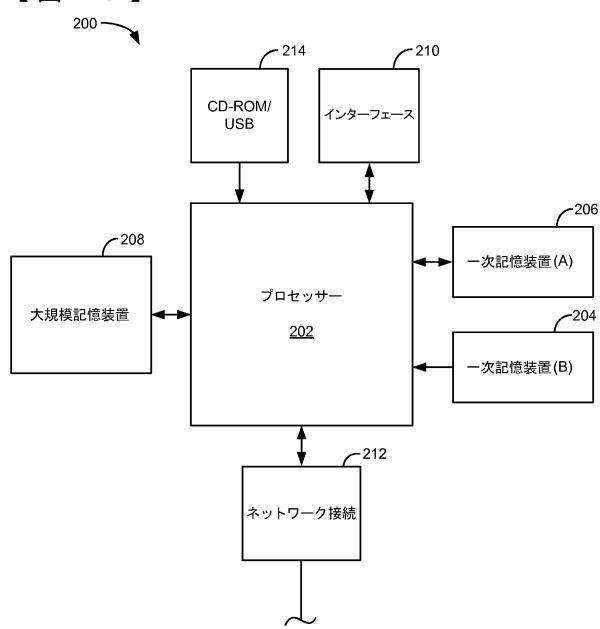
【図8】



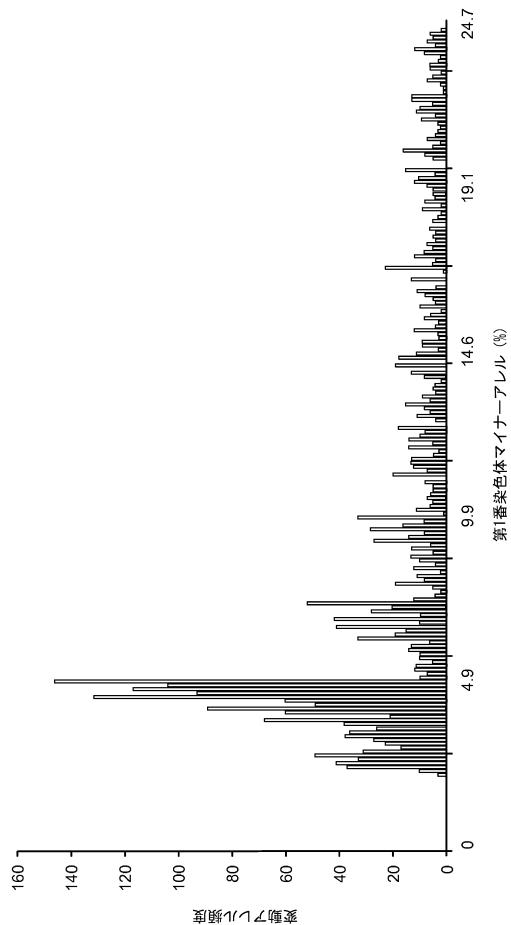
【図9】



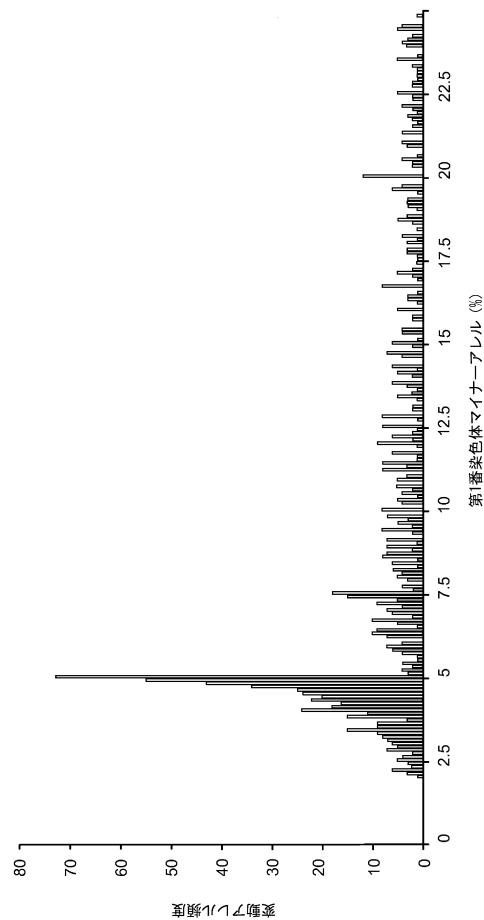
【図10】



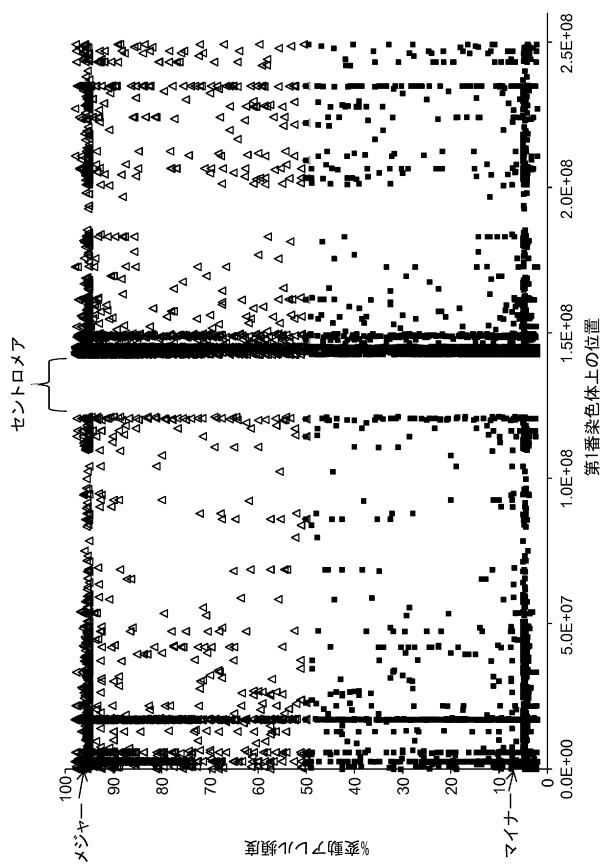
【図11A】



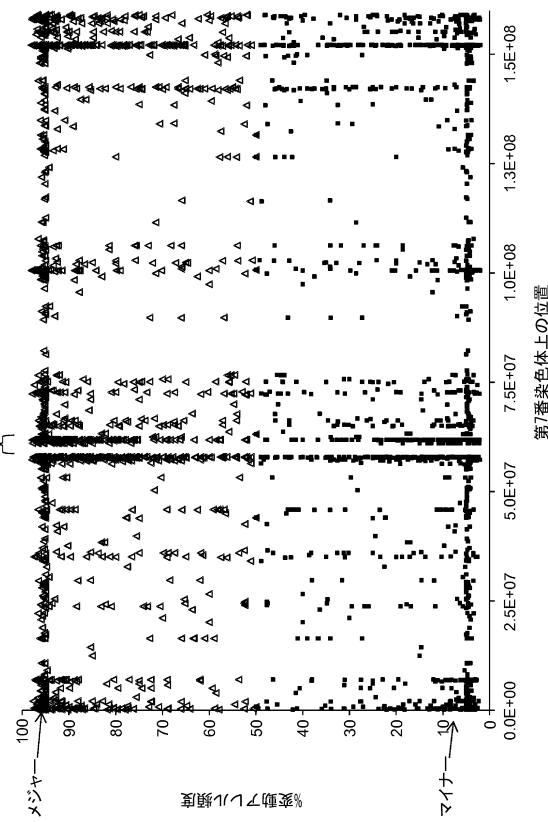
【図11B】



【図 1 2 A】



【図 1 2 B】



【配列表】

0005863946000001.app
0005863946000002.xml

フロントページの続き

(74)代理人 100148699
弁理士 佐藤 利光

(74)代理人 100128048
弁理士 新見 浩一

(74)代理人 100129506
弁理士 小林 智彦

(74)代理人 100114340
弁理士 大関 雅人

(74)代理人 100114889
弁理士 五十嵐 義弘

(74)代理人 100121072
弁理士 川本 和弥

(72)発明者 ラヴァ リチャード ピー。
アメリカ合衆国 カリフォルニア州 レッドウッド シティ エッジウッド ロード 711

(72)発明者 リーズ ブライアン ケイ。
アメリカ合衆国 アリゾナ州 チャンドラー ウエスト シーダー ドライブ 183

(72)発明者 パーク ジョン ピー。
アメリカ合衆国 ネバダ州 リノ リバー フロウ ドライブ 270

審査官 北村 悠美子

(56)参考文献 特表2006-508632 (JP, A)
国際公開第2011/057094 (WO, A1)
The Lancet, 2007年, Vol.369, p.474-481

(58)調査した分野(Int.Cl., DB名)

C 12 Q 1 / 68

C 12 N 15 / 00 - 15 / 90

J S T P l u s / J M E D P l u s / J S T 7 5 8 0 (J D r e a m I I I)

C a p l u s / M E D L I N E / E M B A S E / B I O S I S (S T N)