

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6153874号
(P6153874)

(45) 発行日 平成29年6月28日 (2017. 6. 28)

(24) 登録日 平成29年6月9日 (2017. 6. 9)

(51) Int. Cl.	F I		
C 1 2 Q 1/68 (2006. 01)	C 1 2 Q	1/68	A
C 1 2 M 1/00 (2006. 01)	C 1 2 Q	1/68	Z
G O 6 F 19/18 (2011. 01)	C 1 2 M	1/00	A
C 1 2 N 15/09 (2006. 01)	G O 6 F	19/18	
	C 1 2 N	15/00	A

請求項の数 19 (全 138 頁)

(21) 出願番号	特願2013-553423 (P2013-553423)	(73) 特許権者	513156537
(86) (22) 出願日	平成23年11月18日 (2011. 11. 18)		ナテラ, インコーポレイテッド
(65) 公表番号	特表2014-507141 (P2014-507141A)		アメリカ合衆国 カリフォルニア 940
(43) 公表日	平成26年3月27日 (2014. 3. 27)		70, サン カルロス, インダストリ
(86) 国際出願番号	PCT/US2011/061506		アル ストリート 201, スイート
(87) 国際公開番号	W02012/108920		410
(87) 国際公開日	平成24年8月16日 (2012. 8. 16)	(74) 代理人	100099759
審査請求日	平成26年11月18日 (2014. 11. 18)		弁理士 青木 篤
(31) 優先権主張番号	61/462, 972	(74) 代理人	100077517
(32) 優先日	平成23年2月9日 (2011. 2. 9)		弁理士 石田 敬
(33) 優先権主張国	米国 (US)	(74) 代理人	100087871
(31) 優先権主張番号	61/448, 547		弁理士 福本 積
(32) 優先日	平成23年3月2日 (2011. 3. 2)	(74) 代理人	100087413
(33) 優先権主張国	米国 (US)		弁理士 古賀 哲次

最終頁に続く

(54) 【発明の名称】 非侵襲的出生前倍数性呼び出しのための方法

(57) 【特許請求の範囲】

【請求項 1】

妊娠中の胎児における染色体または染色体セグメントの倍数性状態を決定するための方法であって、当該方法は、

ハイスループット配列決定または遺伝子型決定アレイを用いて、予調製試料中に存在する、該染色体または染色体セグメント上の複数の多型遺伝子座における対立遺伝子を測定し、ここで該予調製試料は、該胎児の母親由来の母系DNAおよび該胎児由来の胎児DNAを含む第1のDNAの試料から該DNAを単離し、該DNAを該複数の多型遺伝子座について優先的に富化することによって調製され、

該予調製試料について得られた測定値から、該複数の多型遺伝子座における対立遺伝子の数をコンピュータで算出し、

複数の倍数性仮説をコンピュータで作製し、ここで該複数の倍数性仮説は各々、該染色体または染色体セグメントの倍数性状態の異なる選択肢に対応し、

各倍数性仮説について、該対立遺伝子がベータ二項分布に従って分布するとの仮定の下、該染色体または染色体セグメント上の該複数の多型遺伝子座における該対立遺伝子の予測数についての同時分布モデルをコンピュータで構築し、

該同時分布モデル、該予調製試料において測定された該対立遺伝子数、および該予調製試料中の胎児DNAの推定される割合を用いて、該倍数性仮説のそれぞれの相対的確率をコンピュータで決定し、

最大の確率を有する仮説に対応する倍数性状態を該胎児の倍数性状態として選択し、こ

10

20

で該倍数性状態について信頼度推定値が算出されることを含む方法。

【請求項 2】

前記第 1 の試料中の前記 DNA が母系の血漿を起源とする、請求項 1 に記載の方法。

【請求項 3】

該複数の多型遺伝子座における DNA を優先的に富化するステップが、
複数のライゲーション媒介性 PCR プロブを取得し、ここで、前記複数の PCR プロブの各々が、該多型遺伝子座のうちの 1 つを標的とし、該 PCR プロブの上流部および下流部が、該遺伝子座の多型部位から少数の塩基で隔てられている DNA の一方の鎖上の DNA の領域とハイブリダイズするように設計されており、該少数が、2～30 塩基対で

10

あり、
該ライゲーション媒介性 PCR プロブを前記第 1 の試料由来の前記 DNA にハイブリダイズさせ、

該ライゲーション媒介性 PCR プロブ末端間のギャップを、DNA ポリメラーゼを用いて充填し、

該ライゲーション媒介性 PCR プロブをライゲーションし、

ライゲーションされた該ライゲーション媒介性 PCR プロブを増幅する

ことを含む、請求項 1 に記載の方法。

【請求項 4】

該複数の多型遺伝子座における DNA を優先的に富化するステップが、
複数の内側フォワードプライマーを取得し、ここで、前記複数のプライマーの各々が、該多型遺伝子座のうちの 1 つを標的とし、該内側フォワードプライマーのそれぞれの 3' 末端が、該遺伝子座の多型部位の上流にあり少数の塩基で該多型部位から隔てられている DNA の領域とハイブリダイズするように設計されており、該少数が、2～30 塩基対であり、

20

必要に応じて、複数の内側リバースプライマーを取得し、ここで、前記複数のプライマーの各々が、該多型遺伝子座のうちの 1 つを標的とし、該内側リバースプライマーのそれぞれの 3' 末端が、該遺伝子座の多型部位の上流にあり少数の塩基で該多型部位から隔てられている DNA の領域とハイブリダイズするように設計されており、該少数が、2～30 塩基対であり、

30

該内側プライマーを該 DNA とハイブリダイズさせ、

ポリメラーゼ連鎖反応を用いて該 DNA を増幅してアンプリコンを形成する

ことを含む、請求項 1 に記載の方法。

【請求項 5】

前記方法が更に、

複数の外側フォワードプライマーを取得し、ここで各プライマーが、前記多型遺伝子座のうちの 1 つを標的とし、該複数の外側フォワードプライマーの各々が、前記内側フォワードプライマーの上流の DNA の領域とハイブリダイズするように設計されており、

必要に応じて、複数の外側リバースプライマーを取得し、ここで各プライマーが、該多型遺伝子座のうちの 1 つを標的とし、該複数の外側 リバースプライマーの各々が、前記内側

40

リバースプライマーのすぐ下流の DNA の領域とハイブリダイズするように設計されてお

り、
前記外側フォワードプライマーを該 DNA とハイブリダイズさせ、

ポリメラーゼ連鎖反応を用いて該 DNA を増幅する

ことをさらに含む、請求項 4 に記載の方法。

【請求項 6】

前記方法が更に、

複数の外側リバースプライマーを取得し、ここで各プライマーが、前記多型遺伝子座のうちの 1 つを標的とし、該複数の外側 リバースプライマーの各々が、対応する前記内側 リバースプライマーのすぐ下流の DNA の領域とハイブリダイズするように設計されており、

50

必要に応じて、複数の外側フォワードプライマーを得るステップであって、それぞれのプライマーが該多型遺伝子座のうちの1つを標的とし、そして該外側フォワードプライマーのそれぞれが、対応する前記内側フォワードプライマーの上流のDNAの領域とハイブリダイズするように設計されており、

前記外側リバースプライマーを該DNAとハイブリダイズさせ、

ポリメラーゼ連鎖反応を用いて該DNAを増幅する

ことをさらに含む、請求項4に記載の方法。

【請求項7】

(a) 前記第1の試料を調製するステップが、該第1の試料中の前記DNAにユニバーサルアダプタを付加し、該第1の試料中の前記DNAを、ポリメラーゼ連鎖反応を用いて増幅することをさらに含む；

10

(b) 前記DNAを増幅するステップが、1つまたは複数の個々の反応容積で行われ、個々の反応容積のそれぞれが、500超の異なるフォワードプライマーとリバースプライマーとの対を含有し；あるいは

(c) 前記内側プライマーが、望ましくないプライマー2重鎖を形成する可能性があるプライマー対を同定し、望ましくないプライマー2重鎖を形成する可能性があると同定されたプライマーの対の少なくとも1つを前記複数のプライマーから除去することによって選択される、請求項4に記載の方法。

【請求項8】

前記方法が、前記胎児の片親または両親から、複数の多型遺伝子座における遺伝子型データを得ることをさらに含む、

20

(a) 前記染色体または染色体セグメント上の前記複数の多型遺伝子座の予測される対立遺伝子数の確率についての同時分布モデルを構築するステップが、前記片親または両親から得られた前記遺伝子データを用いて行われるか；あるいは

(b) 前記第1の試料が母系の血漿から単離されたものであり、前記母親から遺伝子型データを得るステップが、前記予調製試料に対して得られた前記測定値から該母系の遺伝子型データを推定することによって行われる、請求項1に記載の方法。

【請求項9】

前記優先的な富化により、前記予調製試料と前記第1の試料の間に、2倍以下の平均の対立遺伝子の偏りがもたらされる、請求項4に記載の方法。

30

【請求項10】

前記複数の多型遺伝子座が一塩基多型である、請求項1に記載の方法。

【請求項11】

前記胎児の前記倍数性状態を呼び出すステップが、前記同時分布モデルおよび前記対立遺伝子数の確率を用いて決定される前記倍数性仮説のそれぞれの前記相対的確率と、読み取り数解析、ヘテロ接合率の比較、親の遺伝子情報を使用する場合にのみ利用可能な統計量、特定の親の状況に対して正規化された遺伝子型シグナルの確率、前記第1の試料または前記予調製試料の推定される胎児の割合を用いて算出される統計量、およびそれらの組合せからなる群から選択される一つ以上の統計学的技法を用いて算出される該倍数性仮説のそれぞれの相対的確率とを組み合わせることをさらに含む、請求項1に記載の方法。

40

【請求項12】

前記方法が、倍数性状態を決定するために、標的化配列決定およびベイズ最尤推定(Maximum Likelihood Estimation: MLE)を用いることを含む、請求項1に記載の方法。

【請求項13】

請求項4に記載の方法で使用するために設計された、妊娠中の胎児における標的染色体または染色体セグメントの倍数性状態を決定するためのキットであって、前記複数の内側フォワードプライマーと、必要に応じて前記複数の内側リバースプライマーとを含み、該プライマーのそれぞれが前記標的染色体または染色体セグメント上の前記多型部位のうちの1つのすぐ上流および／または下流のDNAの領域とハイブリダイズす

50

るように設計されており、該ハイブリダイズする領域が、少数の塩基によって該多型部位から隔てられており、該少数が、2～30塩基対である、プライマーを含むキット。

【請求項14】

妊娠中の胎児における染色体の倍数性状態の決定に役立つ診断ボックスであって、請求項1に記載の方法における調製するステップおよび測定するステップを実行することができる診断ボックス。

【請求項15】

前記胎児の前記呼び出された倍数性状態に基づいた臨床的措置がとられるかどうかの指標となり、該臨床的措置は、妊娠中絶すること、または妊娠を維持することの一方から選択される、請求項1に記載の方法。

10

【請求項16】

母系遺伝子データが、天然では排他的またはほぼ排他的に母系である遺伝物質の測定によって決定されるものではない、請求項8に記載の方法。

【請求項17】

母系遺伝子データが、母系DNAと胎児DNAの混合物を含む母系血漿について実施される遺伝子測定から推定される、請求項8に記載の方法。

【請求項18】

前記方法が更に、胎児DNAの推測を用いて、倍数性の呼び出しと、倍数性の呼び出しの信頼度推定値を計算することを含む、請求項1に記載の方法。

【請求項19】

20

前記同時分布モデルの構築が、染色体内の異なる場所における染色体乗換えの確率に関するデータを用いて、前記染色体上の多型対立遺伝子間の依存性をモデリングすることにより行われる、請求項1に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本願は、2011年2月9日に出願された米国仮特許出願第61/462,972号；2011年3月2日に出願された米国仮特許出願第61/448,547号；2011年4月12日に出願された米国仮特許出願第61/516,996号；2011年5月18日に出願された米国特許出願第13/110,685号；および2011年6月23日に出願された米国仮特許出願第61/571,248号の利益を主張し、これらの特許出願のすべては、それらのすべての教示のために本明細書中に参考として援用される。

30

【0002】

分野

本開示は、一般に、非侵襲的出生前倍数性呼び出しのための方法に関する。

【背景技術】

【0003】

背景

出生前診断の現行の方法では、医師および親に、成長している胎児の異常を警告することができる。出生前診断をしないと、50人に1人の乳児が重大な身体的または精神的ハンディキャップを持って生まれ、30人に1人もが先天性奇形のいくつかの形態を有する。残念ながら、標準の方法は、正確度が乏しいか、または流産のリスクを有する侵襲的な手順を伴う。母系の血中ホルモンレベルまたは超音波測定に基づく方法は非侵襲的であるが、同様に正確度が低い。羊水穿刺、絨毛膜絨毛生検および胎児の血液試料採取などの方法は正確度が高いが、侵襲的であり、著しいリスクを有する。米国では全妊娠のおよそ3%に対して羊水穿刺が実施されたが、その使用頻度は過去15年にわたって減少している。

40

【0004】

最近、無細胞胎児DNAおよびインタクトな胎児細胞が母系の血液循環に進入し得るこ

50

とが発見された。したがって、この遺伝材料を分析することにより、早期の非侵襲的出生前遺伝子診断（NPD）が可能になり得る。

【0005】

正常なヒトは、健康な二倍体細胞の全てに23種の染色体を2組有し、1つのコピーが各親に由来する。多すぎる染色体および／または少なすぎる染色体を有する核細胞における状態である異数性が、着床の失敗、流産、および遺伝病の大部分に関与すると考えられる。染色体異常を検出することにより、とりわけ、上首尾の妊娠の機会が増すことに加えて、ダウン症候群、クラインフェルター症候群、およびターナー症候群などの状態を有する個体または胚を同定することができる。染色体異常を検査することは、母親の年齢が35歳から40歳の間では胚の少なくとも40%が異常であり、40歳を超えると、胚の半分超が異常であるということが推定されるので、特に重要である。

10

【0006】

出生前スクリーニングのために用いられるいくつかの検査

妊娠初期に母系の血清において測定される妊娠関連血漿タンパク質A（PAPP-A）のレベルが低いことは、13トリソミー、18トリソミー、および21トリソミーを含めた胎児の染色体異常に関連し得る。さらに、妊娠初期のPAPP-Aレベルが低いことにより、胎内発育遅延（SGA）の乳児または死産を含めた有害な妊娠転帰を予測することができる。妊娠中の女性は、多くの場合、妊娠初期での血清スクリーニングを受け、これは、一般に、女性を、ホルモンであるPAPP-Aおよびベータヒト絨毛性ゴナドトロピン（ベータ-hCG）の血中レベルについて検査することを伴う。いくつかの場合には、女性は、可能性のある生理的欠陥を探すための超音波も受ける。特に、項部浮腫（nuchal translucency）（NT）測定により、胎児における異数性のリスクを示すことができる。多くの地域では、出生前スクリーニングのための標準の処置は、妊娠初期での血清スクリーニングとNT検査の組合せを包含する。

20

【0007】

トリプルテストは、トリプルスクリーニング、Kettinger検査またはBart検査とも称され、患者を、染色体異常（および神経管欠損）について高リスクまたは低リスクのいずれかに分類するために、妊娠中、妊娠中期に実施される調査である。「多マーカースクリーニング検査」という用語が代わりに使用される場合がある。「トリプルテスト」という用語は、「ダブルテスト」、「クアドループルテスト」、「クアッドテスト」および「ペンタテスト」という用語を包含し得る。

30

【0008】

トリプルテストでは、アルファフェトプロテイン（AFP）、コンジュゲートしていないエストリオール（UE₃）、ベータヒト絨毛性ゴナドトロピン（ベータ-hCG）、浸潤性トロホプラスト抗原（ITA）および／またはインヒビンの血清レベルを測定する。陽性検査とは、染色体異常（および神経管欠損）の危険性が高いことを意味し、そのような患者は、決定的な診断を受けるために、より感度が高く特異度が高い手順、主に羊水穿刺のような侵襲的手順が照会される。トリプルテストを用いて、21トリソミー（ダウン症候群）を含めたいくつかの状態をスクリーニングすることができる。ダウン症候群に加えて、トリプルテストおよびクアドループルテストでは、エドワーズ症候群、開放性神経管欠損としても公知である胎児の18トリソミーがスクリーニングされ、ターナー症候群、三倍体性、16トリソミーモザイク現象、胎児の死亡、スミスレムリオピッツ症候群、およびステロイドスルファターゼ欠損の危険性の増大を検出することもできる。

40

【発明の概要】

【課題を解決するための手段】

【0009】

要旨

本明細書には、妊娠中の胎児における染色体の倍数性状態を決定するための方法が開示されている。本明細書に例示されている態様によると、ある実施形態では、妊娠中の胎児における染色体の倍数性状態を決定するための方法は、胎児の母親由来の母系DNAおよ

50

び胎児由来の胎児DNAを含む第1のDNAの試料を得るステップと、調製された試料が得られるようにDNAを単離することによって第1の試料を調製するステップと、染色体上の複数の多型遺伝子座における、調製された試料中のDNAを測定するステップと、調製された試料に対して行ったDNA測定から、複数の多型遺伝子座における対立遺伝子数をコンピュータで算出するステップと、それぞれが、染色体における可能性のある異なる倍数性状態に関する、複数の倍数性仮説をコンピュータで作製するステップと、各倍数性仮説について、染色体上の複数の多型遺伝子座における予測される対立遺伝子数についての同時分布モデルをコンピュータで構築するステップと、同時分布モデルおよび調製された試料において測定された対立遺伝子数を用いて、倍数性仮説のそれぞれの相対的確率 (conditional probability) をコンピュータで決定するステップと、最大の確率を有する仮説に対応する倍数性状態を選択することによって胎児の倍数性状態を呼び出すステップとを含む。

10

【0010】

いくつかの実施形態では、第1の試料中のDNAは母系の血漿を起源とする。いくつかの実施形態では、第1の試料を調製するステップは、DNAを増幅するステップをさらに含む。いくつかの実施形態では、第1の試料を調製するステップは、複数の多型遺伝子座における第1の試料中のDNAを優先的に富化するステップをさらに含む。

【0011】

いくつかの実施形態では、複数の多型遺伝子座における第1の試料中のDNAを優先的に富化するステップは、複数の環状化前プローブであって、それぞれのプローブが多型遺伝子座のうちの1つを標的とし、該プローブの3'末端および5'末端が遺伝子座の多型部位から少数の塩基で隔てられているDNAの領域とハイブリダイズするように設計されており、該少数が、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21~25、26~30、31~60、またはそれらの組合せであるプローブを得るステップと、環状化前プローブと第1の試料由来のDNAをハイブリダイズさせるステップと、ハイブリダイズしたプローブ末端間のギャップを、DNAポリメラーゼを用いて埋めるステップと、環状化前プローブを環状化するステップと、環状化されたプローブを増幅するステップとを含む。

20

【0012】

いくつかの実施形態では、複数の多型遺伝子座におけるDNAを優先的に富化するステップは、複数のライゲーション媒介性PCRプローブであって、それぞれのPCRプローブが多型遺伝子座のうちの1つを標的とし、その上流および下流PCRプローブが、遺伝子座の多型部位から少数の塩基で隔てられているDNAの一方の鎖上のDNAの領域とハイブリダイズするように設計されており、該少数が、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21~25、26~30、31~60、またはそれらの組合せであるPCRプローブを得るステップと、ライゲーション媒介性PCRプローブと第1の試料由来のDNAをハイブリダイズさせるステップと、ライゲーション媒介性PCRプローブ末端間のギャップを、DNAポリメラーゼを用いて埋めるステップと、ライゲーション媒介性PCRプローブをライゲーションするステップと、ライゲーションされたライゲーション媒介性PCRプローブを増幅するステップとを含む。

30

40

【0013】

いくつかの実施形態では、複数の多型遺伝子座におけるDNAを優先的に富化するステップは、多型遺伝子座を標的とする複数のハイブリッド捕捉プローブを得るステップと、ハイブリッド捕捉プローブを、第1の試料中のDNAとハイブリダイズさせるステップと、DNAに関する第1の試料からハイブリダイズしていないDNAの一部または全部を物理的に除去するステップとを含む。

【0014】

いくつかの実施形態では、ハイブリッド捕捉プローブは、多型部位と隣接しているがオーバーラップはしていない領域とハイブリダイズするように設計されている。いくつかの

50

実施形態では、ハイブリッド捕捉プローブは、多型部位と隣接しているがオーバーラップはしていない領域とハイブリダイズするように設計されており、隣接捕捉プローブの長さは、約120塩基未満、約110塩基未満、約100塩基未満、約90塩基未満、約80塩基未満、約70塩基未満、約60塩基未満、約50塩基未満、約40塩基未満、約30塩基未満、および約25塩基未満からなる群から選択することができる。いくつかの実施形態では、ハイブリッド捕捉プローブは、多型部位とオーバーラップする領域とハイブリダイズするように設計されており、複数のハイブリッド捕捉プローブは、各多型遺伝子座に対する少なくとも2つのハイブリッド捕捉プローブを含み、各ハイブリッド捕捉プローブが、一方の多型遺伝子座において別の対立遺伝子と相補的であるように設計されている。

10

【0015】

いくつかの実施形態では、複数の多型遺伝子座のDNAを優先的に富化するステップは、複数の内側のフォワードプライマーであって、それぞれのプライマーが多型遺伝子座のうちの1つを標的とし、該内側のフォワードプライマーの3'末端が、多型部位の上流にあり、少数の塩基で多型部位から隔てられているDNAの領域とハイブリダイズするように設計されており、少数が、1塩基対、2塩基対、3塩基対、4塩基対、5塩基対、6～10塩基対、11～15塩基対、16～20塩基対、21～25塩基対、26～30塩基対または31～60塩基対からなる群から選択されるプライマーを得るステップと、必要に応じて、複数の内側のリバースプライマーであって、それぞれのプライマーが多型遺伝子座のうちの1つを標的とし、内側のリバースプライマーの3'末端が、多型部位の上流にあり、少数の塩基で多型部位から隔てられているDNAの領域とハイブリダイズするように設計されており、少数が、1塩基対、2塩基対、3塩基対、4塩基対、5塩基対、6～10塩基対、11～15塩基対、16～20塩基対、21～25塩基対、26～30塩基対または31～60塩基対からなる群から選択されるプライマーを得るステップと、内側のプライマーをDNAとハイブリダイズさせるステップと、ポリメラーゼ連鎖反応を用いてDNAを増幅してアンプリコンを形成するステップとを含む。

20

【0016】

いくつかの実施形態では、該方法は、複数の外側のフォワードプライマーであって、それぞれのプライマーが多型遺伝子座のうちの1つを標的とし、内側のフォワードプライマーの上流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、必要に応じて、複数の外側のリバースプライマーであって、それぞれのプライマーが該多型遺伝子座のうちの1つを標的とし、内側のリバースプライマーのすぐ下流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、第1のプライマーをDNAとハイブリダイズさせるステップと、ポリメラーゼ連鎖反応を用いてDNAを増幅するステップとをさらに含む。

30

【0017】

いくつかの実施形態では、該方法は、複数の外側のリバースプライマーであって、それぞれのプライマーが多型遺伝子座のうちの1つを標的とし、内側のリバースプライマーのすぐ下流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、必要に応じて、複数の外側のフォワードプライマーであって、それぞれのプライマーが多型遺伝子座のうちの1つを標的とし、内側のフォワードプライマーの上流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、第1のプライマーをDNAとハイブリダイズさせるステップと、ポリメラーゼ連鎖反応を用いてDNAを増幅するステップとをさらに含む。

40

【0018】

いくつかの実施形態では、第1の試料を調製するステップは、第1の試料中のDNAにユニバーサルアダプタを付加するステップと、第1の試料中のDNAを、ポリメラーゼ連鎖反応を用いて増幅するステップとをさらに含む。いくつかの実施形態では、増幅されたアンプリコンの少なくとも小部分が100bp未満、90bp未満、80bp未満、70bp未満、65bp未満、60bp未満、55bp未満、50bp未満、または45bp

50

未満であり、小部分とは10%、20%、30%、40%、50%、60%、70%、80%、90%または99%である。

【0019】

いくつかの実施形態では、DNAを増幅するステップは、1つまたは複数の個々の反応容積で行われ、個々の反応容積のそれぞれは、100超の異なるフォワードプライマーとリバースプライマーの対、200超の異なるフォワードプライマーとリバースプライマーの対、500超の異なるフォワードプライマーとリバースプライマーの対、1,000超の異なるフォワードプライマーとリバースプライマーの対、2,000超の異なるフォワードプライマーとリバースプライマーの対、5,000超の異なるフォワードプライマーとリバースプライマーの対、10,000超の異なるフォワードプライマーとリバースプライマーの対、20,000超の異なるフォワードプライマーとリバースプライマーの対、50,000超の異なるフォワードプライマーとリバースプライマーの対、または、100,000超の異なるフォワードプライマーとリバースプライマーの対を含有する。

10

【0020】

いくつかの実施形態では、第1の試料を調製するステップは、第1の試料を複数の部分に分割するステップであって、複数の多型遺伝子座のサブセットにおいて各部分内のDNAが優先的に富化されるステップをさらに含む。いくつかの実施形態では、望ましくないプライマー2重鎖を形成する可能性があるプライマー対を同定するステップ、および望ましくないプライマー2重鎖を形成する可能性があると同定されたプライマーの対の少なくとも1つを複数のプライマーから除去するステップによって内側のプライマーを選択する。いくつかの実施形態では、内側のプライマーは、標的の多型遺伝子座の上流または下流のいずれかとハイブリダイズするように設計された領域を含有し、必要に応じて、PCR増幅が可能になるように設計されたユニバーサルプライミング配列 (priming sequence) を含有する。いくつかの実施形態では、プライマーの少なくとも一部は、個々のプライマー分子各々について異なるランダムな領域をさらに含有する。いくつかの実施形態では、プライマーの少なくとも一部は分子バーコードをさらに含有する。

20

【0021】

いくつかの実施形態では、該方法は、胎児の一方の親または両親から遺伝子型データを得るステップも包含する。いくつかの実施形態では、胎児の一方の親または両親から遺伝子型データを得るステップは、親由来のDNAを調製するステップであって、複数の多型遺伝子座におけるDNAを優先的に富化し、調製された親のDNAを得ることを含むステップと、必要に応じて、調製された親のDNAを増幅するステップと、複数の多型遺伝子座における、調製された試料中の親のDNAを測定するステップとを含む。

30

【0022】

いくつかの実施形態では、染色体上の複数の多型遺伝子座における予測される対立遺伝子数の確率についての同時分布モデルを構築するステップを、一方の親または両親から得られた遺伝子データを用いて行う。いくつかの実施形態では、第1の試料を母系の血漿から単離し、母親から遺伝子型データを得るステップを、調製された試料に対して行ったDNA測定から母系の遺伝子型データを推定することによって行う。

40

【0023】

いくつかの実施形態では、優先的な富化により、調製された試料と第1の試料の間に、2倍以下、1.5倍以下、1.2倍以下、1.1倍以下、1.05倍以下、1.02倍以下、1.01倍以下、1.005倍以下、1.002倍以下、1.001倍以下および1.0001倍以下からなる群から選択される係数の程度の、平均の対立遺伝子の偏りがもたらされる。いくつかの実施形態では、複数の多型遺伝子座はSNPである。いくつかの実施形態では、調製された試料中のDNAを測定するステップを配列決定によって行う。

【0024】

いくつかの実施形態では、妊娠中の胎児における染色体の倍数性状態の決定に役立つ診断ボックス (diagnostic box) であって、請求項1に記載の方法における調製するステップおよび測定するステップを実行することができる診断ボックスが開示さ

50

れている。

【0025】

いくつかの実施形態では、対立遺伝子数は、バイナリーではなく確率的なものである。いくつかの実施形態では、複数の多型遺伝子座における、調製された試料中のDNAの測定値を、胎児がハプロタイプに関連づけられる遺伝性の1つまたは複数の疾患を有するか否かを決定するためにも用いる。

【0026】

いくつかの実施形態では、対立遺伝子数の確率についての同時分布モデルを構築するステップを、染色体内の異なる場所における染色体乗換えの確率に関するデータを使用して、染色体上の多型対立遺伝子間の依存性をモデリングすることによって行う。いくつかの実施形態では、対立遺伝子数についての同時分布モデルを構築するステップおよび各仮説の相対的確率を決定するステップを、参照染色体を使用することを必要としない方法を用いて行う。

【0027】

いくつかの実施形態では、各仮説の相対的確率を決定するステップに、調製された試料中の胎児DNAの推定される割合 (estimated fraction) を使用する。いくつかの実施形態では、対立遺伝子数の確率を算出するステップおよび各仮説の相対的確率を決定するステップにおいて使用する、調製された試料からのDNA測定値は、一次遺伝子データを含む。いくつかの実施形態では、最大の確率を有する仮説に対応する倍数性状態を選択するステップを、最尤推定または最大事後推定を用いて行う。

【0028】

いくつかの実施形態では、胎児の倍数性状態を呼び出すステップは、同時分布モデルおよび対立遺伝子数の確率を用いて決定される倍数性仮説のそれぞれの相対的確率と、読み取り数解析 (read count analysis)、ヘテロ接合率の比較、親の遺伝子情報を使用する場合にのみ利用可能な統計量、特定の親の状況に対して正規化された遺伝子型シグナルの確率、第1の試料または調製された試料の推定される胎児の割合 (fetal fraction) を用いて算出される統計量、およびそれらの組合せからなる群から選択される統計学的技法を用いて算出される倍数性仮説のそれぞれの相対的確率とを組み合わせるステップも包含する。

【0029】

いくつかの実施形態では、呼び出された倍数性状態について信頼度推定値を算出する。いくつかの実施形態では、該方法は、呼び出された胎児の倍数性状態に基づいて、妊娠中絶すること、または妊娠を維持することの一方から選択される臨床的措置をとるステップも包含する。

【0030】

いくつかの実施形態では、該方法は、妊娠4週から5週の間；妊娠5週から6週の間；妊娠6週から7週の間；妊娠7週から8週の間；妊娠8週から9週の間；妊娠9週から10週の間；妊娠10週から12週の間；妊娠12週から14週の間；妊娠14週から20週の間；妊娠20週から40週の間；妊娠初期；妊娠中期；妊娠後期；またはそれらの組合せにおいて、胎児に対して実施することができる。

【0031】

いくつかの実施形態では、該方法を用いて、妊娠中の胎児における決定された染色体の倍数性状態を示す報告を作製する。いくつかの実施形態では、請求項9に記載の方法で使用するために設計された、妊娠中の胎児における標的染色体の倍数性状態を決定するためのキットであって、複数の内側のフォワードプライマーおよび、必要に応じて複数の内側のリバースプライマーであって、該プライマーのそれぞれが標的染色体上の多型部位のうちの1つのすぐ上流および／または下流のDNAの領域とハイブリダイズするように設計されているプライマーと、必要に応じてさらに別の染色体であって、ハイブリダイズする領域が、少数の塩基によって該多型部位から隔てられており、該少数が、1、2、3、4、5、6～10、11～15、16～20、21～25、26～30、31～60、およ

10

20

30

40

50

びそれらの組合せからなる群から選択される染色体とを含むキットが開示されている。

【0032】

いくつかの実施形態では、胎児のゲノムDNAおよび母系のゲノムDNAを含む母系の組織試料において胎児の異数性の存在または不在を決定するための方法であって、(a) 前記母系の組織試料から、胎児のゲノムDNAと母系のゲノムDNAの混合物を得るステップと、(b) ステップa) の胎児のゲノムDNAと母系のゲノムDNAの混合物から無作為に選択されたDNA断片の大規模並行DNA配列決定を行って、前記DNA断片の配列を決定するステップと、(c) ステップb) で得られた配列が属する染色体を同定するステップと、(d) ステップc) のデータを用いて、前記母系のゲノムDNAと胎児のゲノムDNAの混合物中の少なくとも1つの第1の染色体の量を決定するステップであって、前記少なくとも1つの第1の染色体が、胎児において正倍数性であると推定されるステップと、(e) ステップc) のデータを用いて、前記母系のゲノムDNAと胎児のゲノムDNAの混合物中の第2の染色体の量を決定するステップであって、前記第2の染色体が、胎児において異数体であることが疑われるステップと、(f) 胎児DNAと母系DNAの混合物中の胎児DNAの割合を算出するステップと、(g) 第2の標的染色体が正倍数性である場合、ステップd) の数を用いて第2の標的染色体の量の予測される分布を算出するステップと、(h) 第2の標的染色体が異数性である場合、ステップd) の第1の数およびステップf) で算出された、胎児DNAと母系DNAの混合物中の胎児DNAの割合を用いて第2の標的染色体の量の予測される分布を算出するステップと、(i) 最尤法または最大事後法を用いて、ステップe) で決定された第2の染色体の量がステップg) で算出された分布またはステップh) で算出された分布の一部である可能性がより高いかを決定し、それにより、胎児の異数性の存在または不在を示すステップとを含む方法が開示されている。

10

20

本発明の好ましい実施形態において、例えば以下の項目が提供される。

(項目1)

妊娠中の胎児における染色体の倍数性状態を決定するための方法であって、
該胎児の母親由来の母系DNAおよび該胎児由来の胎児DNAを含む第1のDNAの試料を得るステップと、
調製された試料が得られるように該DNAを単離することによって該第1の試料を調製するステップと、
該染色体上の複数の多型遺伝子座における該調製された試料中の該DNAを測定するステップと、
該調製された試料に対して行った該DNA測定から、該複数の多型遺伝子座における対立遺伝子数をコンピュータで算出するステップと、
それぞれが、該染色体における可能性のある異なる倍数性状態に関する、複数の倍数性仮説をコンピュータで作製するステップと、
各倍数性仮説について、該染色体上の該複数の多型遺伝子座における予測される該対立遺伝子数についての同時分布モデルをコンピュータで構築するステップと、
該同時分布モデルおよび該調製された試料において測定された該対立遺伝子数を用いて、該倍数性仮説のそれぞれの相対的確率をコンピュータで決定するステップと、
最大の確率を有する該仮説に対応する該倍数性状態を選択することによって該胎児の該倍数性状態を呼び出すステップと
を含む方法。

30

40

(項目2)

前記第1の試料中の前記DNAが母系の血漿を起源とする、項目1に記載の方法。

(項目3)

前記第1の試料を調製する前記ステップが、前記DNAを増幅するステップをさらに含む、項目1に記載の方法。

(項目4)

前記第1の試料を調製する前記ステップが、複数の多型遺伝子座における前記第1の試

50

料中の前記DNAを優先的に富化するステップをさらに含む、項目1に記載の方法。

(項目5)

前記複数の多型遺伝子座における前記第1の試料中の前記DNAを前記優先的に富化するステップが、

複数の環状化前プローブであって、それぞれのプローブが該多型遺伝子座のうちの1つを標的とし、該プローブの3'末端および5'末端が該遺伝子座の多型部位から少数の塩基で隔てられているDNAの領域とハイブリダイズするように設計されており、該少数が、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21～25、26～30、31～60、またはそれらの組合せであるプローブを得るステップと、

10

該環状化前プローブと該第1の試料由来のDNAをハイブリダイズさせるステップと、
該ハイブリダイズしたプローブ末端間のギャップを、DNAポリメラーゼを用いて埋めるステップと、

該環状化前プローブを環状化するステップと、
該環状化されたプローブを増幅するステップと
を含む、項目4に記載の方法。

(項目6)

前記複数の多型遺伝子座における前記DNAを前記優先的に富化するステップが、
複数のライゲーション媒介性PCRプローブであって、それぞれのPCRプローブが該多型遺伝子座のうちの1つを標的とし、該PCRプローブの上流部および下流部が、該遺伝子座の多型部位から少数の塩基で隔てられているDNAの一方の鎖上のDNAの領域とハイブリダイズするように設計されており、該少数が、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21～25、26～30、31～60、またはそれらの組合せであるPCRプローブを得るステップと、

20

該ライゲーション媒介性PCRプローブと前記第1の試料由来の前記DNAをハイブリダイズさせるステップと、

該ライゲーション媒介性PCRプローブ末端間のギャップを、DNAポリメラーゼを用いて埋めるステップと、

該ライゲーション媒介性PCRプローブをライゲーションするステップと、
ライゲーションされた該ライゲーション媒介性PCRプローブを増幅するステップと
を含む、項目4に記載の方法。

30

(項目7)

前記複数の多型遺伝子座における前記DNAを前記優先的に富化するステップが、
該多型遺伝子座を標的とする複数のハイブリッド捕捉プローブを得るステップと、
該ハイブリッド捕捉プローブを、前記第1の試料中の前記DNAとハイブリダイズさせるステップと、

DNAに関する該第1の試料からハイブリダイズしていないDNAの一部または全部を物理的に除去するステップと

を含む、項目4に記載の方法。

40

(項目8)

前記複数のハイブリッド捕捉プローブが、前記多型部位と隣接しているがオーバーラップはしていない領域とハイブリダイズするように設計されている、項目7に記載の方法。

(項目9)

前記複数のハイブリッド捕捉プローブが、前記多型部位と隣接しているがオーバーラップはしていない領域とハイブリダイズするように設計されており、該隣接捕捉プローブの長さが、約120塩基未満、約110塩基未満、約100塩基未満、約90塩基未満、約80塩基未満、約70塩基未満、約60塩基未満、約50塩基未満、約40塩基未満、約30塩基未満、および約25塩基未満からなる群から選択することができる、項目7に記載の方法。

50

(項目 1 0)

前記複数のハイブリッド捕捉プローブが、前記多型部位とオーバーラップする領域とハイブリダイズするように設計されており、該複数のハイブリッド捕捉プローブが、各多型遺伝子座に対する少なくとも2つのハイブリッド捕捉プローブを含み、各ハイブリッド捕捉プローブが、一方の多型遺伝子座において別の対立遺伝子と相補的であるように設計されている、項目7に記載の方法。

(項目 1 1)

複数の多型遺伝子座における前記DNAを前記優先的に富化するステップが、
複数の内側のフォワードプライマーであって、それぞれのプライマーが該多型遺伝子座のうちの1つを標的とし、該内側のフォワードプライマーの3'末端が、前記多型部位の上流にあり少数の塩基で該多型部位から隔てられているDNAの領域とハイブリダイズするように設計されており、該少数が、1塩基対、2塩基対、3塩基対、4塩基対、5塩基対、6～10塩基対、11～15塩基対、16～20塩基対、21～25塩基対、26～30塩基対または31～60塩基対からなる群から選択されるプライマーを得るステップと、

10

必要に応じて、複数の内側のリバースプライマーであって、それぞれのプライマーが該多型遺伝子座のうちの1つを標的とし、該内側のリバースプライマーの3'末端が、該多型部位の上流にあり少数の塩基で該多型部位から隔てられているDNAの領域とハイブリダイズするように設計されており、該少数が、1塩基対、2塩基対、3塩基対、4塩基対、5塩基対、6～10塩基対、11～15塩基対、16～20塩基対、21～25塩基対、26～30塩基対または31～60塩基対からなる群から選択されるプライマーを得るステップと、

20

該内側のプライマーを該DNAとハイブリダイズさせるステップと、
ポリメラーゼ連鎖反応を用いて該DNAを増幅してアンプリコンを形成するステップとを含む、項目4に記載の方法。

(項目 1 2)

複数の外側のフォワードプライマーであって、それぞれのプライマーが前記多型遺伝子座のうちの1つを標的とし、前記内側のフォワードプライマーの上流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、

必要に応じて、複数の外側のリバースプライマーであって、それぞれのプライマーが該多型遺伝子座のうちの1つを標的とし、前記内側のリバースプライマーのすぐ下流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、

30

第1のプライマーを該DNAとハイブリダイズさせるステップと、
前記ポリメラーゼ連鎖反応を用いて該DNAを増幅するステップとをさらに含む、項目11に記載の方法。

(項目 1 3)

複数の外側のリバースプライマーであって、それぞれのプライマーが前記多型遺伝子座のうちの1つを標的とし、前記内側のリバースプライマーのすぐ下流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、

必要に応じて、複数の外側のフォワードプライマーであって、それぞれのプライマーが該多型遺伝子座のうちの1つを標的とし、前記内側のフォワードプライマーの上流のDNAの領域とハイブリダイズするように設計されているプライマーを得るステップと、

40

前記第1のプライマーを該DNAとハイブリダイズさせるステップと、
前記ポリメラーゼ連鎖反応を用いて該DNAを増幅するステップとをさらに含む、項目11に記載の方法。

(項目 1 4)

前記第1の試料を調製する前記ステップが、
該第1の試料中の前記DNAにユニバーサルアダプタを付加するステップと、
該第1の試料中の前記DNAを、前記ポリメラーゼ連鎖反応を用いて増幅するステップと

50

をさらに含む、項目 1 1 に記載の方法。

(項目 1 5)

増幅された前記アンプリコンの少なくとも小部分が 1 0 0 b p 未満、9 0 b p 未満、8 0 b p 未満、7 0 b p 未満、6 5 b p 未満、6 0 b p 未満、5 5 b p 未満、5 0 b p 未満または 4 5 b p 未満であり、前記小部分が 1 0 %、2 0 %、3 0 %、4 0 %、5 0 %、6 0 %、7 0 %、8 0 %、9 0 % または 9 9 % である、項目 1 1 に記載の方法。

(項目 1 6)

前記 D N A を増幅する前記ステップが、1 つまたは複数の個々の反応容積で行われ、個々の反応容積のそれぞれが、1 0 0 超の異なるフォワードプライマーとリバースプライマーの対、2 0 0 超の異なるフォワードプライマーとリバースプライマーの対、5 0 0 超の異なるフォワードプライマーとリバースプライマーの対、1, 0 0 0 超の異なるフォワードプライマーとリバースプライマーの対、2, 0 0 0 超の異なるフォワードプライマーとリバースプライマーの対、5, 0 0 0 超の異なるフォワードプライマーとリバースプライマーの対、1 0, 0 0 0 超の異なるフォワードプライマーとリバースプライマーの対、2 0, 0 0 0 超の異なるフォワードプライマーとリバースプライマーの対、5 0, 0 0 0 超の異なるフォワードプライマーとリバースプライマーの対、または、1 0 0, 0 0 0 超の異なるフォワードプライマーとリバースプライマーの対を含有する、項目 1 1 に記載の方法。

10

(項目 1 7)

前記第 1 の試料を調製する前記ステップが、該第 1 の試料を複数の部分に分割するステップであって、前記複数の多型遺伝子座のサブセットにおける各部分内の前記 D N A が優先的に富化されるステップをさらに含む、項目 1 1 に記載の方法。

20

(項目 1 8)

前記内側のプライマーを、望ましくないプライマー 2 重鎖を形成する可能性があるプライマー対を同定するステップ、および望ましくないプライマー 2 重鎖を形成する可能性があると同定されたプライマーの対の少なくとも 1 つを前記複数のプライマーから除去するステップによって選択する、項目 1 1 に記載の方法。

(項目 1 9)

前記内側のプライマーが、前記標的の多型遺伝子座の上流または下流のいずれかとハイブリダイズするように設計された領域を含有し、必要に応じて、P C R 増幅が可能になるように設計されたユニバーサルプライミング配列を含有する、項目 1 1 に記載の方法。

30

(項目 2 0)

前記プライマーの少なくとも一部が、個々のプライマー分子各々について異なるランダムな領域をさらに含有する、項目 1 1 に記載の方法。

(項目 2 1)

前記プライマーの少なくとも一部が、分子バーコードをさらに含有する、項目 1 1 に記載の方法。

(項目 2 2)

前記胎児の一方の親または両親から遺伝子型データを得るステップをさらに含む、項目 1 に記載の方法。

40

(項目 2 3)

前記胎児の一方の親または両親から遺伝子型データを得る前記ステップが、該両親由来の前記 D N A を調製するステップであって、前記複数の多型遺伝子座における前記 D N A を優先的に富化し、調製された親の D N A を得るステップを含むステップと

、必要に応じて、該調製された親の D N A を増幅するステップと、

該複数の多型遺伝子座における該調製された試料中の該親の D N A を測定するステップとを含む、項目 2 2 に記載の方法。

(項目 2 4)

50

前記染色体上の前記複数の多型遺伝子座の前期予測される対立遺伝子数の確率についての同時分布モデルを構築する前記ステップを、前記一方の親または両親から得られた前記遺伝子データを用いて行う、項目 2 2 に記載の方法。

(項目 2 5)

前記第 1 の試料を母系の血漿から単離し、前記母親から遺伝子型データを得る前記ステップを、前記調製された試料に対して行った前記 DNA 測定から該母系の遺伝子型データを推定するステップによって行う、項目 2 2 に記載の方法。

(項目 2 6)

前記優先的な富化により、前記調製された試料と前記第 1 の試料の間に、2 倍以下、1.5 倍以下、1.2 倍以下、1.1 倍以下、1.05 倍以下、1.02 倍以下、1.01 倍以下、1.005 倍以下、1.002 倍以下、1.001 倍以下および 1.0001 倍以下からなる群から選択される係数の程度の、平均の対立遺伝子の偏りがもたらされる、項目 4 に記載の方法。

10

(項目 2 7)

前記複数の多型遺伝子座が SNP である、項目 1 に記載の方法。

(項目 2 8)

前記調製された試料中の前記 DNA を測定する前記ステップを配列決定するステップによって行う、項目 1 に記載の方法。

(項目 2 9)

妊娠中の胎児における染色体の倍数性状態の決定に役立つ診断ボックスであって、項目 1 に記載の方法における調製するステップおよび測定するステップを実行することができる診断ボックス。

20

(項目 3 0)

前記対立遺伝子数がバイナリーではなく確率的なものである、項目 1 に記載の方法。

(項目 3 1)

前記複数の多型遺伝子座における前記調製された試料中の前記 DNA の測定値を、前記胎児がハプロタイプに関連づけられる遺伝性の 1 つまたは複数の疾患を有するか否かを決定するためにも用いる、項目 1 に記載の方法。

(項目 3 2)

対立遺伝子数の確率についての同時分布モデルを構築する前記ステップを、染色体内の異なる場所における染色体乗換えの確率に関するデータを使用して、前記染色体上の多型対立遺伝子間の依存性をモデリングすることによって行う、項目 1 に記載の方法。

30

(項目 3 3)

対立遺伝子数についての同時分布モデルを構築する前記ステップと各仮説の前記相対的確率を決定する前記ステップをどちらも、参照染色体を使用することを必要としない方法を用いて行う、項目 1 に記載の方法。

(項目 3 4)

各仮説の前記相対的確率を決定する前記ステップが、前記調製された試料中の胎児 DNA の推定される小部分を使用する、項目 1 に記載の方法。

(項目 3 5)

対立遺伝子数の確率を算出するステップおよび各仮説の前記相対的確率を決定するステップにおいて使用する前記調製された試料からの前記 DNA 測定値が、一次遺伝子データを含む、項目 1 に記載の方法。

40

(項目 3 6)

最大の確率を有する前記仮説に対応する前記倍数性状態を選択するステップを、最尤推定または最大事後推定を使用して行う、項目 1 に記載の方法。

(項目 3 7)

前記胎児の前記倍数性状態を呼び出す前記ステップが、

前記同時分布モデルおよび前記対立遺伝子数の確率を用いて決定される前記倍数性仮説のそれぞれの前記相対的確率と、読み取り数解析、ヘテロ接合率の比較、親の遺伝子情報

50

を使用する場合にのみ利用可能な統計量、特定の親の状況に対して正規化された遺伝子型シグナルの確率、前記第1の試料または前記調製された試料の推定される胎児の割合を用いて算出される統計量、およびそれらの組合せからなる群から選択される統計学的技法を用いて算出される該倍数性仮説のそれぞれの相対的確率とを組み合わせるステップをさらに含む、項目1に記載の方法。

(項目38)

前記呼び出された倍数性状態について信頼度推定値を算出する、項目1に記載の方法。

(項目39)

前記胎児の前記呼び出された倍数性状態に基づいて、妊娠中絶すること、または妊娠を維持することの一方から選択される臨床的措置をとるステップをさらに含む、項目1に記載の方法。

(項目40)

妊娠4週から5週の間；妊娠5週から6週の間；妊娠6週から7週の間；妊娠7週から8週の間；妊娠8週から9週の間；妊娠9週から10週の間；妊娠10週から12週の間；妊娠12週から14週の間；妊娠14週から20週の間；妊娠20週から40週の間；妊娠初期；妊娠中期；妊娠後期；またはそれらの組合せにおいて実施することができる、項目1に記載の方法。

(項目41)

項目1に記載の方法を用いて生成した、決定された妊娠中の胎児における染色体の倍数性状態を示す報告。

(項目42)

項目11に記載の方法で使用するために設計された、妊娠中の胎児における標的染色体の倍数性状態を決定するためのキットであって、

前記複数の内側のフォワードプライマーおよび、必要に応じて前記複数の内側のリバースプライマーであって、該プライマーのそれぞれが前記標的染色体上の前記多型部位のうちの1つのすぐ上流および／または下流のDNAの領域とハイブリダイズするように設計されているプライマーと、必要に応じてさらに別の染色体であって、該ハイブリダイズする領域が、少数の塩基によって該多型部位から隔てられており、該少数が、1、2、3、4、5、6～10、11～15、16～20、21～25、26～30、31～60、およびそれらの組合せからなる群から選択される染色体

とを含むキット。

(項目43)

胎児のゲノムDNAおよび母系のゲノムDNAを含む母系の組織試料において胎児の異数性の存在または不在を決定するための方法であって、

a) 該母系の組織試料から、胎児のゲノムDNAと母系のゲノムDNAの混合物を得るステップと、

b) ステップa)の胎児のゲノムDNAと母系のゲノムDNAの混合物から無作為に選択されたDNA断片の大規模並行DNA配列決定を行って、該DNA断片の配列を決定するステップと、

c) ステップb)で得られた配列が属する染色体を同定するステップと、

d) ステップc)のデータを用いて、母系のゲノムDNAと胎児のゲノムDNAの該混合物中の少なくとも1つの第1の染色体の量を決定するステップであって、該少なくとも1つの第1の染色体が、該胎児において正倍数性であると推定されるステップと、

e) ステップc)のデータを用いて、母系のゲノムDNAと胎児のゲノムDNAの該混合物中の第2の染色体の量を決定するステップであって、該第2の染色体が、該胎児において異数性であることが疑われるステップと、

f) 胎児DNAと母系DNAの該混合物中の胎児DNAの割合を算出するステップと、

g) 該第2の標的染色体が正倍数性である場合、ステップd)の数を用いて該第2の標的染色体の量の予測される分布を算出するステップと；

h) 該第2の標的染色体が異数性である場合、ステップd)の第1の数およびステップ

10

20

30

40

50

f) で算出された、胎児DNAと母系DNAの該混合物中の胎児DNAの前記割合を用いて該第2の標的染色体の量の予測される分布を算出するステップと、

i) 最尤法または最大事後法を用いて、ステップe) で決定された該第2の染色体の量がステップg) で算出された該分布またはステップh) で算出された該分布のどちらの一部である可能性がより高いかを決定し、それにより、胎児の異数性の存在または不在を示すステップと

を含む方法。

【0033】

ここで開示されている実施形態は、添付図を参照してさらに説明され、同様の構造はいくつかの概観を通じて同様の数字で参照される。示されている図は必ずしも一定の縮尺ではなく、概して、ここで開示されている実施形態の原理の例示が強調されている。

【図面の簡単な説明】

【0034】

【図1】図1は、直接多重mini-PCR法の図表示である。

【図2】図2は、セミネステッドmini-PCR法の図表示である。

【図3】図3は、完全ネステッドmini-PCR法の図表示である。

【図4】図4は、ヘミネステッドmini-PCR法の図表示である。

【図5】図5は、3重ヘミネステッドmini-PCR法の図表示である。

【図6】図6は、片側ネステッドmini-PCR法の図表示である。

【図7】図7は、片側mini-PCR法の図表示である。

【図8】図8は、逆セミネステッドmini-PCR法の図表示である。

【図9】図9は、セミネステッド法のいくつかの可能性のあるワークフローである。

【図10】図10は、ループライゲーションアダプタの図表示である。

【図11】図11は、内部にタグを付けたプライマーの図表示である。

【図12】図12は、内部のタグを有するいくつかのプライマーの例である。

【図13】図13は、ライゲーションアダプタ結合領域を有するプライマーを使用する方法の図表示である。

【図14】図14は、2つの異なる分析技法を用いた計数方法についてのシミュレートされた倍数性呼び出しの正確度を示す。

【図15】図15は、実験4の細胞系における複数のSNPについての2つの対立遺伝子の比を示す。

【図16】図16は、染色体により分けた、実験4の細胞系における複数のSNPについての2つの対立遺伝子の比を示す。

【図17】図17は、染色体により分けた、4人の妊娠中の女性の血漿試料における複数のSNPについての2つの対立遺伝子の比を示す。

【図18】図18は、データ補正の前後の二項分散によって説明することができるデータの割合を示す。

【図19】図19は、短いライブラリー調製プロトコル後の試料中の胎児DNAの相対的な富化を示すグラフである。

【図20】図20は、直接PCRとセミネステッド法を比較した読み取りグラフの深さを示す。

【図21】図21は、3つのゲノム試料の直接PCRについての読み取りの深さの比較を示す。

【図22】図22は、3つの試料のセミネステッドmini-PCRについての読み取りの深さの比較を示す。

【図23】図23は、1, 200プレックス反応および9, 600プレックス反応についての読み取りの深さの比較を示す。

【図24】図24は、6つの細胞について3つの染色体における読み取り計数比を示す。

【図25】図25は、3つの染色体における、2つの3細胞反応についての、および1 ngのゲノムDNAに対して行った第3の反応についての対立遺伝子の比を示す。

10

20

30

40

50

【図 2 6】図 2 6 は、3 つの染色体における、2 つの単一細胞反応についての対立遺伝子の比を示す。

【発明を実施するための形態】

【0035】

上記の図には、ここで開示されている実施形態が記載されているが、考察において言及されている通り、他の実施形態も意図されている。本開示は、例示により実施形態を代表として示しており、限定として示しているのではない。当業者は、ここで開示されている実施形態の原理の範囲および趣旨の範囲内に入る多数の他の改変および実施形態を考案することができる。

【0036】

詳細な説明

ある実施形態では、本開示は、DNA の混合試料（すなわち、胎児の母親由来の DNA、および胎児由来の DNA）から測定された遺伝子型データから、および必要に応じて、母親由来の遺伝物質（genetic material）および場合によっては同様に父親由来の遺伝物質の試料から測定された遺伝子型データから、妊娠中の胎児における染色体の倍数性状態を決定するための *ex vivo* 方法であって、該決定を、同時分布モデルを用い、親の遺伝子型データを考慮して、胎児における可能性のある異なる倍数性状態についての予測される対立遺伝子分布の集合を作製し、予測される対立遺伝子分布と、混合試料において測定された実際の対立遺伝子分布とを比較し、予測される対立遺伝子分布パターンが観察された対立遺伝子分布パターンと最も厳密に一致する倍数性状態を選択することによって行う方法を提供する。ある実施形態では、混合試料は、母系の血液または母系の血清もしくは血漿に由来する。ある実施形態では、DNA の混合試料を、複数の多型遺伝子座で優先的に富化することができる。ある実施形態では、優先的な富化は、対立遺伝子の偏りが最小限になるように行う。ある実施形態では、本開示は、複数の遺伝子座において対立遺伝子の偏りが少なくなるように優先的に富化された DNA の組成に関する。ある実施形態では、対立遺伝子分布（複数可）を、混合試料由来の DNA について配列決定することによって測定する。ある実施形態では、同時分布モデルにより、対立遺伝子が二項様式で分布することが仮定される。ある実施形態では、種々の供給源からの現存の組換え頻度を考慮して、例えば、International HapMap Consortium からのデータを使用して、遺伝的に連鎖している遺伝子座について予測同時対立遺伝子分布の集合を作製する。

【0037】

ある実施形態では、本開示は、非侵襲的な出生前診断（NPD）の方法、詳細には、DNA 混合物について測定された遺伝子型データにおいて複数の多型遺伝子座における対立遺伝子測定値を観察することによって胎児の異数性状態を決定するための方法であって、ある特定の対立遺伝子測定値により異数体の胎児が示され、一方、他の対立遺伝子測定値により正倍数性の胎児が示される示す方法を提供する。ある実施形態では、遺伝子型データを、母系の血漿に由来する DNA 混合物について配列決定することによって測定する。ある実施形態では、DNA 試料を、対立遺伝子分布を算出する複数の遺伝子座に対応する DNA 分子について優先的に富化することができる。ある実施形態では、母親由来の遺伝物質のみを含む、または、ほぼ母親由来の遺伝物質のみを含む DNA の試料を測定し、場合によっては、父親由来の遺伝物質のみを含む、または、ほぼ父親由来の遺伝物質のみを含む DNA の試料も測定する。ある実施形態では、一方の親または両親の遺伝子測定値を推定される胎児の割合と一緒に使用して、胎児における可能性のある異なる基礎をなす遺伝子の状態に対応する複数の予測される対立遺伝子分布を作製し、該予測される対立遺伝子分布は、仮説と称することができる。ある実施形態では、母系の遺伝子データは、天然で排他的またはほぼ排他的に母系のものである遺伝物質を測定することによって決定するのではなく、母系 DNA と胎児 DNA の混合物を含む母系の血漿に対して行われる遺伝子測定から推定する。いくつかの実施形態では、仮説は、1 つまたは複数の染色体における胎児の倍数性、胎児のどの染色体のどのセグメントがどちらの親から遺伝したか、および

10

20

30

40

50

それらの組合せを含んでよい。いくつかの実施形態では、胎児の倍数性状態は、観察された対立遺伝子測定値と、異なる仮説であって、該仮説の少なくとも一部が、異なる倍数性状態に対応する仮説を比較し、観察された対立遺伝子測定値を考慮して、真である可能性が最も高い仮説に対応する倍数性状態を選択することによって決定する。ある実施形態では、この方法は、遺伝子座がホモ接合性であるかヘテロ接合性であるかにかかわらず、測定されたSNPの一部または全部からの対立遺伝子測定データの使用を伴い、したがって、ヘテロ接合性のみである遺伝子座の対立遺伝子の使用は伴わない。この方法は、遺伝子データがただ1つの多型遺伝子座に関係する状況には適さない場合がある。この方法は、遺伝子データが、標的染色体に対して10超の多型遺伝子座、または20超の多型遺伝子座についてのデータを含む場合に特に有利である。この方法は、遺伝子データが、標的染色体に対して50超の多型遺伝子座、100超の多型遺伝子座、または標的染色体に対して200超の多型遺伝子座についてのデータを含む場合に特に有利である。いくつかの実施形態では、遺伝子データは、標的染色体に対して500超の多型遺伝子座、1,000超の多型遺伝子座、2,000超の多型遺伝子座、または、標的染色体に対して5,000超の多型遺伝子座についてのデータを含んでよい。

【0038】

ある実施形態では、本明細書に開示されている方法は、元のDNAの試料中、多型遺伝子座の集合からの各多型遺伝子座に存在する相対的な対立遺伝子頻度を保存する選択的富化技法を用いる。いくつかの実施形態では、増幅および／または選択的富化技法は、ライゲーション媒介性PCRなどのPCR、ハイブリダイゼーションによる断片の捕捉、分子反転プローブまたは他の環状化プローブを伴い得る。いくつかの実施形態では、増幅または選択的な富化のための方法は、標的配列と正確にハイブリダイズした際に、ヌクレオチドプローブの3'末端または5'末端が、少数のヌクレオチドで対立遺伝子の多型部位から隔てられるようなプローブの使用を伴ってよい。この隔たりにより、対立遺伝子の偏りと称される、一方の対立遺伝子の優先的な増幅が減少する。これは、正確にハイブリダイズしたプローブの3'末端または5'末端が対立遺伝子の多型部位と直接隣接する、またはそれと非常に近くなるようなプローブの使用を伴う方法よりも改善されている。ある実施形態では、ハイブリダイズ領域が、多型部位を含有する可能性がある、またはそれを確実に含有するプローブは排除される。ハイブリダイゼーションの部位に多型部位があることにより、一部の対立遺伝子において不均等なハイブリダイゼーションが引き起こされ得る、または、ハイブリダイゼーションが全体で阻害され得、その結果、特定の対立遺伝子が優先的に増幅される。これらの実施形態は、試料が単一の個体由来の純粋なゲノム試料であろうが個体の混合物であろうが、試料の各多型遺伝子座における元の対立遺伝子頻度をより良好に保存するという点で、標的化増幅および／または選択的な富化を伴う他の方法よりも改善されている。

【0039】

ある実施形態では、本明細書に開示されている方法は、非常に効率的な高度多重標的化PCRを使用して、DNAを増幅し、その後ハイスループット配列決定を行って各標的遺伝子座における対立遺伝子頻度を決定する。約50または100を超えるPCRプライマーを1回の反応で、生じた配列読み取りの大部分が、標的の遺伝子座にマッピングされるように多重化できることは、新規かつ非自明である。非常に効率的な様式で実施するための高度多重標的化PCRを可能にする1つの技法は、互いとハイブリダイズする可能性が低いプライマーの設計を伴う。一般にはプライマーと称されるPCRプローブは、少なくとも500、少なくとも1,000、少なくとも5,000、少なくとも10,000、少なくとも20,000、少なくとも50,000または少なくとも100,000の潜在的なプライマー対間の潜在的に有害な相互作用、または、プライマーと試料DNAの間の意図されたものではない相互作用の熱力学的モデルを作製し、次いで、このモデルを使用して、プール内の他の設計物と適合しない設計物を排除することによって選択する。非常に効率的な様式で実施するための高度多重標的化PCRを可能にする別の技法は、部分的な、または完全なネスティング手法を用いて標的化PCRを行うことである。これらの

手法の1つまたはその組合せを用いることにより、単一のプール内の少なくとも300個、少なくとも800個、少なくとも1,200個、少なくとも4,000個または少なくとも10,000個のプライマーを多重化することが可能になり、生じた増幅されたDNAは、配列決定すると、標的の遺伝子座にマッピングされる大多数のDNA分子を含む。これらの手法の1つまたはその組合せを用いることにより、単一のプール内の多数のプライマーを多重化することが可能になり、生じた増幅されたDNAは、50%超、80%超、90%超、95%超、98%超または99%超の、標的の遺伝子座にマッピングされるDNA分子を含む。

【0040】

ある実施形態では、本明細書に開示されている方法は、多型遺伝子座の各対立遺伝子の独立した観察の数の定量的尺度を提供する。これは、マイクロアレイまたは定性的PCRなどの、2つの対立遺伝子の比に関する情報をもたらすが、いずれかの対立遺伝子の独立した観察の数を定量化しない大多数の方法とは異なる。独立した観察の数に関する定量的情報をもたらす方法では、倍数性の算出には比のみを利用し、一方、定量的情報はそれ自体では有用ではない。独立した観察の数に関する情報を保持することの重要性を例示するために、2つの対立遺伝子、AおよびBを有する試料の遺伝子座について考察する。第1の実験では20の対立遺伝子Aおよび20の対立遺伝子Bを観察し、第2の実験では200の対立遺伝子Aおよび200の対立遺伝子Bを観察する。どちらの実験でも、比(A/(A+B))は0.5と等しいが、第2の実験は、第1の実験よりも対立遺伝子AまたはBの頻度の確実性に関する多くの情報を伝える。先行技術で公知のいくつかの方法は、個々の対立遺伝子からの対立遺伝子の比(チャンネル比)(すなわち x_i/y_i)を平均または合計し、この比を、参照染色体と比較するか、またはこの比が特定の状況でどのように挙動すると予想されるかに関する規則を用いるかのいずれかで解析することを伴う。当技術分野で公知のそのような方法では、対立遺伝子の重み付けを伴わず、各対立遺伝子についてほぼ同じ量のPCR産物を確実にすることができ、および全ての対立遺伝子が同じように挙動するはずであることが想定される。そのような方法にはいくつかの不都合があり、より重要なことに、本開示の他の箇所に記載されているいくつかの改善を用いることが妨げられる。

【0041】

ある実施形態では、本明細書に開示されている方法は、ダイソミーにおいて予測される対立遺伝子頻度分布ならびに減数分裂Iの間の染色体不分離、減数分裂IIの間の染色体不分離、および/または胎児発生の初期の有糸分裂の間の染色体不分離によって生じるトリソミーの場合に予測され得る複数の対立遺伝子頻度分布を明確にモデリングする。なぜこれが重要であるかを例示するために、乗換えがない場合を考える：減数分裂Iの間の染色体不分離により、2つの異なる相同体が一方の親から遺伝によって受け継がれたトリソミーがもたらされ、対照的に、減数分裂IIの間、または胎児発生の初期の有糸分裂の間の染色体不分離により、一方の親由来の同じ相同体の2つのコピーがもたらされることになる。各筋書きにより、各多型遺伝子座において、また遺伝連鎖に起因して、共同して考えられる全ての遺伝子座において、予測される対立遺伝子の異なる頻度がもたらされることになる。相同体間での遺伝物質の交換をもたらす乗換えにより、遺伝様式がより複雑になり、ある実施形態では、当該方法は、遺伝子座間の物理的な距離に加えて、組換え率の情報を使用することによってこれに適応する。ある実施形態では、減数分裂I時の染色体不分離と減数分裂IIまたは有糸分裂時の染色体不分離との間の区別の改善を可能にするために、当該方法では、モデルに、セントロメアからの距離が増加するにつれて上昇する乗換えの確率を組み入れる。減数分裂IIおよび有糸分裂時の染色体不分離は、有糸分裂時の染色体不分離により、一般には、1つの相同体の同一またはほぼ同一のコピーがもたらされるが、一方、減数分裂II時の染色体不分離事象の後に存在する2つの相同体は、多くの場合、配偶子形成の間の1つまたは複数の乗換えに起因して異なるという事実によって区別することができる。

【0042】

いくつかの実施形態では、本明細書に開示されている方法は、観察された対立遺伝子測定値を、可能性のある胎児の遺伝子異数性に対応する理論的仮説と比較するステップを包含し、ヘテロ接合性遺伝子座における対立遺伝子の比を定量するステップは包含しない。遺伝子座の数が約20未満の場合、ヘテロ接合性遺伝子座における対立遺伝子の比を定量するステップを含む方法を用いて行った倍数性の決定と、観察された対立遺伝子測定値を、可能性のある胎児の遺伝子の状態に対応する理論的な対立遺伝子分布の仮説と比較することを含む方法を用いて行った倍数性の決定は、同様の結果をもたらす得る。しかし、遺伝子座の数が50超である場合、これらの2つの方法は、有意に異なる結果をもたらす可能性があり、遺伝子座の数が400超、1,000超または2,000超である場合、これらの2つの方法は、ますます有意に異なる結果をもたらす可能性が高い。これらの差は、各対立遺伝子の大きさを独立に測定すること、および比を総計または平均することを伴わずにヘテロ接合性遺伝子座における対立遺伝子の比を定量するステップを含む方法が、同時分布モデルを用いること、連鎖解析を実施すること、二項分布モデルを用いること、および／または他の高度な統計学的技法を含めた技法を用いることを妨げるが、観察された対立遺伝子測定値を、可能性のある胎児の遺伝子の状態に対応する理論的な対立遺伝子分布の仮説と比較するステップを含む方法を用いると、決定の正確度を実質的に上昇させることができるこれらの技法を用いることができるという事実に起因する。

10

【0043】

ある実施形態では、本明細書に開示されている方法は、観察された対立遺伝子測定値の分布により、同時分布モデルを用いて正倍数性または異数体の胎児が示されるかどうかを決定するステップを包含する。同時分布モデルの使用は、方法であって、多型遺伝子座を独立に処理することによってヘテロ接合率を決定する方法とは、得られた決定の正確度が有意に高いという点で異なり、それよりも有意に改善されている。いかなる特定の理論にも縛られることなく、それらの正確度が高い1つの理由は、同時分布モデルでは、SNP間の連鎖、および成長して胎児になる胚を形成する配偶子を生じる減数分裂の間に起こった乗換えの尤度を考慮に入れることであると考えられる。1つまたは複数の仮説について対立遺伝子測定値の予測される分布を作製する際に連鎖の概念を用いる目的は、それにより、連鎖を用いない場合よりも相当よい現実に対応する予測される対立遺伝子測定値分布の作製を可能にすることである。例えば、2つのSNPが存在し、1および2は、互いに近くに位置し、母親は、一方の相同体上のSNP1がAであり、SNP2がAであり、相同体2上のSNP1がBであり、SNP2がBであると考ええる。父親が、両方の相同体上の両方のSNPについてAであり、胎児のSNP1についてBが測定された場合、これは、相同体2を胎児が遺伝によって受け継いだこと、したがって、胎児のSNP2にBが存在する尤度があるかに高いことを示す。連鎖を考慮に入れたモデルではこれが予測されるが、連鎖を考慮に入れないモデルでは予測されない。あるいは、母親のSNP1がABであり、近くのSNP2がABである場合、その場所における母系トリソミーに対応する2つの仮説—一致コピーエラー（減数分裂IIまたは胎児発生初期の有糸分裂における染色体不分離）を伴うもの、および不一致コピーエラー（減数分裂Iにおける染色体不分離）を伴うものを用いることができる。一致コピーエラートリソミーの場合には、胎児が、SNP1において母親からAAを遺伝によって受け継いだ場合、胎児は、SNP2において母親から、ABではなく、AAまたはBBのいずれかを遺伝によって受け継ぐ可能性はあるかに高い。不一致コピーエラーの場合には、胎児は、両方のSNPにおいて母親からABを遺伝によって受け継ぐことになる。連鎖を考慮に入れた、倍数性呼び出し方法によって立てられた対立遺伝子分布の仮説により、これらの予測がなされ、したがって、連鎖を考慮に入れなかった倍数性呼び出し方法よりも相当に大きな程度で、実際の対立遺伝子測定値に対応する。連鎖手法は、対立遺伝子の比を算出することおよびそれらの対立遺伝子の比を総計することに依拠する方法を用いる場合には不可能であることに留意されたい。

20

30

40

【0044】

観察された対立遺伝子測定値を、可能性のある胎児の遺伝子の状態に対応する理論的仮説と比較するステップを含む方法を用いる倍数性の決定の正確度がより高いと考えられる

50

1つの理由は、配列決定を使用して対立遺伝子を測定する場合、この方法では、読み取りの総数が他の方法よりも少ない場合に、対立遺伝子からのデータから、より多くの情報を収集することができることであり、例えば、対立遺伝子の比を算出することおよび総計することに依拠する方法では、不釣り合いに重み付けられた確率論的ノイズが生じる。例えば、配列決定を用いて対立遺伝子を測定することを伴う場合であって、各遺伝子座について配列読み取りが5つのみ検出された遺伝子座の集合が存在する場合を考える。ある実施形態では、対立遺伝子のそれぞれについて、データを、仮定された対立遺伝子分布と比較し、配列読み取りの数に従って重み付けることができ、したがって、これらの測定からのデータは、適切に重み付けられ、全体的な決定に組み入れられる。これは、ヘテロ接合性遺伝子座における対立遺伝子の比を定量することを伴う方法が、可能性のある対立遺伝子の比として0%、20%、40%、60%、80%または100%の比しか算出することができず、これらはいずれも予測される対立遺伝子の比には近づくことができないので、上記方法とは対照的である。この後者の場合、算出された対立遺伝子の比は、読み取りが不十分なので棄却しなければならないか、あるいは、不相応に重み付けされ、確率論的ノイズが決定に導入され、それにより、決定の正確度が低下する。ある実施形態では、個々の対立遺伝子測定を、独立した測定として処理することができ、この場合、同じ遺伝子座の対立遺伝子に対して行った測定間の関係が、異なる遺伝子座の対立遺伝子に対して行った測定間の関係と異なる。

10

【0045】

ある実施形態では、本明細書に開示されている方法は、任意のメトリックを、ダイソミーであることが予想される参照染色体において観察された対立遺伝子測定値と比較するステップ（RC法と称される）を包含せずに、観察された対立遺伝子測定値の分布により、正倍数性または異数体の胎児が示されるかどうかを決定するステップを包含する。これは、疑わしい染色体から無作為に配列決定された断片の割合を、1つまたは複数の推測ダイソミー参照染色体と比較して評価することによって異数性を検出する、ショットガン配列決定を用いる方法などの方法に対する有意な改善である。このRC法では、推測ダイソミー参照染色体が実際にはダイソミーではない場合、不正確な結果がもたらされる。これは、異数性が、単一染色体のトリソミーより実質的である場合、または胎児が三倍体であり、全ての常染色体がトリソミーである場合に起こり得る。雌性三倍体（69、XXX）胎児の場合には、実際は、ダイソミー染色体は全く存在しない。本明細書に記載の方法は、参照染色体を必要とせず、雌性三倍体胎児におけるトリソミー染色体を正確に同定することができる。染色体、仮説、子の割合（child fraction）およびノイズレベルのそれぞれについて、同時分布モデルを、参照染色体のデータ、全体的な子の割合の見積もりまたは固定された参照仮説のいずれも伴わずに適合させることができる。

20

30

【0046】

ある実施形態では、本明細書に開示されている方法は、観察されている多型遺伝子座における対立遺伝子分布をどのように使用して、先行技術の方法よりも高い正確度で胎児の倍数性状態を決定することができるかを実証している。ある実施形態では、該方法は、標的化配列決定を用いて、複数のSNPにおける混合母体-胎児遺伝子型、および必要に応じて、母親の遺伝子型および／または父親の遺伝子型を得て、まず異なる仮説の下での種々の予測される対立遺伝子頻度分布を確立すること、次いで、母体-胎児混合物において得られる定量的な対立遺伝子の情報を観察すること、および、どの仮説がデータに最もよく適合するかを評価することを用い、データに最もよく適合する仮説に対応する遺伝子の状態を正確な遺伝子の状態として呼び出す。ある実施形態では、本明細書に開示されている方法は、呼び出された遺伝子の状態が正確な遺伝子の状態であることの信頼度を生成するために、適合の程度も用いる。ある実施形態では、本明細書に開示されている方法は、親の状況が異なる遺伝子座に関して見いだされる対立遺伝子の分布を解析するアルゴリズムを使用するステップ、および異なる親の状況（異なる親の遺伝子型のパターン）についての、異なる倍数性状態について、観察された対立遺伝子分布を、予測される対立遺伝子分布と比較するステップを包含する。これは、母体-胎児混合試料中の各遺伝子座にお

40

50

る各対立遺伝子の独立した事例の数を推定することができる方法を用いない方法とは異なり、それよりも改善されている。ある実施形態では、本明細書に開示されている方法は、観察された対立遺伝子測定値の分布により、母親がヘテロ接合性である遺伝子座において測定された、観察された対立遺伝子分布を用いて、正倍数性または異数体の胎児が示されるかどうかを決定するステップを包含する。これは、その特定の標的個体に対して情報価値が高いことが知られていない遺伝子座についてDNAが優先的に富化されていない場合、または優先的に富化されている場合、倍数性の決定において、配列データの集合から約2倍の遺伝子測定データを用いることが可能になり、それにより、より正確な決定がもたらされるので、母親がヘテロ接合性である遺伝子座における観察された対立遺伝子分布を用いない方法とは異なり、それよりも改善されている。

10

【0047】

ある実施形態では、本明細書に開示されている方法は、天然では、各遺伝子座における対立遺伝子頻度が多項式（したがって、SNPが二対立遺伝子である場合は二項式）であると仮定する同時分布モデルを用いる。いくつかの実施形態では、同時分布モデルは、ベータ二項分布を使用する。各遺伝子座に存在する各対立遺伝子についての定量的尺度を提供する配列決定などの測定技法を用いる場合、二項モデルを、各遺伝子座に適用することができ、対立遺伝子頻度の基礎をなす程度およびその頻度の信頼度を確かめることができる。対立遺伝子の比から倍数性呼び出しを生成する当技術分野で公知の方法または定量的な対立遺伝子情報が棄却される方法を用いて、観察された比の確実性を確かめることができない。当該方法は、特定の遺伝子座における対立遺伝子の比を算出し、次いでそれらの比を総計することを伴う任意の方法では、任意の所与の対立遺伝子または遺伝子座からのDNAの量を示す測定された強度または計数がガウス様式で分布することを必ず仮定するので、対立遺伝子の比を算出し、それらの比を総計して倍数性呼び出しを行う方法とは異なり、それよりも改善されている。本明細書に開示されている方法は、対立遺伝子の比を算出することを伴わない。いくつかの実施形態では、本明細書に開示されている方法は、複数の遺伝子座の各対立遺伝子の観察結果の数をモデルに組み入れるステップを包含し得る。いくつかの実施形態では、本明細書に開示されている方法は、予測される分布自体を算出するステップであって、それにより、対立遺伝子測定値のガウス分布を仮定するモデルのいずれよりも正確であり得る同時二項分布（joint binomial distribution）モデルを用いることが可能になるステップを包含し得る。二項分布モデルがガウス分布よりも有意に正確である尤度は、遺伝子座の数が増加するにつれて増大する。例えば、20未満の遺伝子座を調べる場合、二項分布モデルが有意にすぐれている尤度は低い。しかし、100超、または特に400超、または特に1,000超、または特に2,000超の遺伝子座を使用すると、二項分布モデルの、ガウス分布モデルよりも有意に正確である尤度は非常に高くなり、それにより、より正確な倍数性の決定がもたらされる。二項分布モデルがガウス分布よりも有意に正確である尤度は、同様に、各遺伝子座における観察結果の数が増加するにつれて増大する。例えば、各遺伝子座において10未満の別個の配列を観察する場合、二項分布モデルが有意にすぐれている尤度は低い。しかし、各遺伝子座について50超の配列読み取り、または特に100超の配列読み取り、または特に200超の配列読み取り、または特に300超の配列読み取りを使用すると、二項分布モデルの、ガウス分布モデルよりも有意に正確である尤度は非常に高くなり、それにより、より正確な倍数性の決定がもたらされる。

20

30

40

【0048】

ある実施形態では、本明細書に開示されている方法では、配列決定を用いて、DNA試料中の各遺伝子座における各対立遺伝子の事例の数を測定する。配列決定の読み取りのそれぞれは、特定の遺伝子座にマッピングし、バイナリーの配列読み取りとして処理することができる、あるいは、読み取りおよび／またはマッピングの同一性の確率を、配列読み取りの一部として組み入れることができ、その結果、確率的な配列読み取り、すなわち所与の遺伝子座にマッピングされる配列読み取りの推定の整数または分数がもたらされる。バイナリーの計数または計数の確率を使用すると、測定値の各集合について二項分布を用

50

いることが可能であり、これにより、計数の数の範囲の (around the number of counts) 信頼区間を算出することが可能になる。二項分布を用いることができることにより、より正確な倍数性の推定およびより精度の高い信頼区間を算出することが可能になる。これは、存在する対立遺伝子の量を測定するために強度を用いる方法、例えば、マイクロアレイを用いる方法、または電気泳動のバンドにおいて蛍光性タグを付けたDNAの強度を測定するために蛍光リーダーを用いて測定を行う方法とは異なり、それよりも改善されている。

【0049】

ある実施形態では、本明細書に開示されている方法では、本データの集合の態様を用いて、そのデータの集合についての推定される対立遺伝子頻度分布のパラメータを決定する。これは、本予測される対立遺伝子頻度分布または場合によっては予測される対立遺伝子の比のパラメータを設定するために、トレーニングデータ集合または事前データ集合を利用する方法よりも改善されている。これは、あらゆる遺伝子試料の収集および測定に關する異なる状態の集合が存在することが原因であり、したがって、当該データの集合からのデータを使用して、その試料についての倍数性の決定に使用するためのものである同時分布モデルのパラメータを決定する方法がより正確になりやすい。

【0050】

ある実施形態では、本明細書に開示されている方法は、観察された対立遺伝子測定値の分布により、最尤法を用いて、正倍数性または異数体の胎児が示されるかどうかを決定するステップを包含する。最尤法を用いることは、得られる決定の正確度が有意に高いという点で、単一仮説棄却法を用いる方法とは異なり、それよりも有意に改善されている。1つの理由は、単一仮説棄却法では、2つの測定値分布ではなく、ただ1つの測定値分布に基づいてカットオフ閾値が設定される、つまり、閾値が通常は最適ではないことである。別の理由は、最尤法では、個々の試料の各々の特定の特性にかかわらず全ての試料に対して使用されるカットオフ閾値を決定するのではなく、個々の試料の各々についてカットオフ閾値を最適化することが可能になることである。別の理由は、最尤法を用いることにより、各倍数性呼び出しについて信頼度を算出することが可能になることである。各呼び出しに対して信頼度の算出を行うことができることにより、実践者が、どの呼び出しが正確であるか、およびどれが誤りである可能性がより高いかを知ることが可能になる。いくつかの実施形態では、多種多様な方法を最尤推定法と組み合わせて、倍数性呼び出しの正確度を増強することができる。ある実施形態では、最尤法を、米国特許第7, 888, 017号に記載の方法と組み合わせて用いることができる。ある実施形態では、最尤法を、標的化PCR増幅を用いて混合試料中のDNAを増幅し、その後、読み取り計数方法、例えば、2011年10月のMontrealでのInternational Congress of Human Genetics 2011年において発表されたTANDEM DIAGNOSTICSを用いて配列決定し分析する方法と組み合わせて用いることができる。ある実施形態では、本明細書に開示されている方法は、混合試料中のDNAの胎児の割合を推定するステップ、およびその推定値を用いて倍数性呼び出しと倍数性呼び出しの信頼度の両方を算出するステップを包含する。これは、推定される胎児の割合を十分な胎児の割合のスクリーニングとして用い、その後、胎児の割合を考慮に入れず、呼び出しについての信頼度の算出も生じない単一仮説棄却法を用いて倍数性呼び出しを行う方法とは異なり、かつその方法とは区別されることに留意されたい。

【0051】

ある実施形態では、本明細書に開示されている方法は、各測定値に確率を付与することによってデータがノイズを伴い、エラーを含有する傾向を考慮に入れる。付与された確率的な推定値を伴う測定データを使用して立てられた仮説の集合から正確な仮説を選択するために最尤法を用いることにより、不正確な測定値が考慮に入れられない可能性が高くなり、倍数性呼び出しを導く算出において正確な測定値が用いられる。より精度が高くなるために、この方法では、倍数性の決定において不正確に測定されたデータの影響を系統的に低下させる。これは、全てのデータが同等に正確であると仮定される方法または範囲外

10

20

30

40

50

のデータが倍数性呼び出しを導く算出から任意に排除される方法よりも改善されている。チャンネル比測定値を用いる現行の方法は、個々のSNPチャンネル比を平均することによってこの方法を多数のSNPに拡張することを主張する。SNPの質および観察された読み取りの深さに基づいて予測される測定値の分散によって個々のSNPに重み付けをしないことにより、生じた統計量の正確度が低下し、その結果、倍数性呼び出しの正確度が、特に境界の場合に有意に低下する。

【0052】

ある実施形態では、本明細書に開示されている方法は、胎児においてどのSNPまたは他の多型遺伝子座がヘテロ接合性であるかの知見を前提としない。この方法により、父系の遺伝子型の情報が入手不可能である場合に倍数性呼び出しを行うことが可能になる。これは、標的とする遺伝子座を適切に選択するため、または混合胎児DNA／母系DNA試料に対して得た遺伝子測定値を解釈するために、どのSNPがヘテロ接合性であるかの知見が前もって知られていなければならない方法よりも改善されている。

【0053】

本明細書に記載の方法は、利用可能なDNAが少量である試料、または胎児DNAのパーセントが低い試料に対して用いる場合に特に有利である。これは、少量のDNAしか利用可能でない場合に生じる、対立遺伝子ドロップアウト率が相応して高いこと、および／または胎児DNAと母系DNAの混合試料中の胎児DNAのパーセントが低い場合に胎児の対立遺伝子ドロップアウト率が相応して高いことに起因する。対立遺伝子ドロップアウト率が高いこと、つまり、標的個体について、対立遺伝子の大部分が測定されなかったことにより、不十分に正確な胎児の割合の算出、および不十分に正確な倍数性の決定がもたらされる。本明細書に開示されている方法は、SNP間の遺伝様式における連鎖を考慮に入れた同時分布モデルを用いることができるので、有意により正確な倍数性の決定を行うことができる。本明細書に記載の方法により、混合物中の胎児性のDNA分子のパーセントが40%未満、30%未満、20%未満、10%未満、8%未満、さらには6%未満である場合に、正確な倍数性の決定を行うことが可能になる。

【0054】

ある実施形態では、個体のDNAが関連する個体のDNAと混在している場合の測定値に基づいて個体の倍数性状態を決定することが可能である。ある実施形態では、DNAの混合物は、母系の血漿中に見いだされる浮動性DNAであり、これは、既知の核型および既知の遺伝子型を有する母親由来のDNAを含み得、また、未知の核型および未知の遺伝子型を有する胎児DNAと混在し得る。一方の親または両親からの既知の遺伝子型の情報を用いて、混合試料中のDNAの複数の潜在的な遺伝子の状態を、異なる倍数性状態、各親から胎児への異なる染色体の寄与、および必要に応じて、混合物中の異なる胎児DNAの割合について予測することが可能である。潜在的な組成のそれぞれは、仮説と称することができる。次いで、胎児の倍数性状態を、実際の測定値について調べ、観察されたデータを考慮してどの潜在的な組成が最も可能性が高いかを決定することによって決定することができる。

【0055】

いくつかの実施形態では、本明細書に開示されている方法は、存在するDNAが非常に少量である状況において、例えば、*in vitro*での受精または、1つまたは少数の細胞（一般には、細胞10個未満、細胞20個未満または細胞40個未満）が利用可能である法医学的な状況において用いることができる。これらの実施形態では、本明細書に開示されている方法は、他のDNAが混入していないが、DNAが少量であるので倍数性呼び出しが非常に難しい場合に少量のDNAから倍数性呼び出しを行うために役立つ。いくつかの実施形態では、本明細書に開示されている方法は、標的DNAに別の個体のDNAが混入している状況において、例えば、出生前診断、父子試験との関連における母系の血液または受胎検査の産物において用いることができる。これらの方法が特に有利になるいくつかの他の状況は、より大量の正常な細胞の中でただ1つまたは少数の細胞が存在するがん検査の場合である。これらの方法の一部として用いる遺伝子測定は、DNAまたはR

NAを含む任意の試料、例えば、これらに限定されないが血液、血漿、体液、尿、毛髪、涙、唾液、組織、皮膚、指の爪、割球、胚、羊水、絨毛膜絨毛試料、糞便、胆汁、リンパ液、頸管粘液、精液または核酸を含む他の細胞または材料に対して行うことができる。ある実施形態では、本明細書に開示されている方法は、核酸検出方法、例えば、配列決定、マイクロアレイ、qPCR、デジタルPCRまたは核酸を測定するために用いられる他の方法と一緒に実行することができる。何らかの理由で望ましいことが見いだされた場合、遺伝子座における対立遺伝子数の確率の比を算出することができ、対立遺伝子の比を、本明細書に記載の方法のいくつかと、それらの方法に適合性がある限りにおいて組み合わせ用いて倍数性状態を決定することができる。いくつかの実施形態では、本明細書に開示されている方法は、加工された試料に対して行ったDNA測定から、複数の多型遺伝子座における対立遺伝子の比をコンピュータで算出するステップを包含する。いくつかの実施形態では、本明細書に開示されている方法は、本開示に記載の他の改善の任意の組合せと一緒に、加工された試料に対して行ったDNA測定から、複数の多型遺伝子座における対立遺伝子の比をコンピュータで算出するステップを包含する。

10

【0056】

上記の点のさらなる考察は、本文書の他の箇所に見いだすことができる。

【0057】

非侵襲的な出生前診断（NPD）

非侵襲的な出生前診断のプロセスは、いくつかのステップを伴う。ステップのいくつかとしては、（１）胎児から遺伝物質を得るステップと、（２）混合試料中に存在する可能性のある胎児の遺伝物質をex vivoで富化するステップと、（３）遺伝物質をex vivoで増幅するステップと、（４）遺伝物質の特定の遺伝子座をex vivoで優先的に富化するステップと、（５）遺伝物質をex vivoで測定するステップと、（６）遺伝子型データを、ex vivoで、コンピュータで分析するステップとを挙げることができる。これらの６つおよび他の関連性のあるステップの実施を減少させるための方法が本明細書に記載されている。該方法のステップの少なくとも一部は、直接体には適用されない。ある実施形態では、本開示は、体から単離され、分離された組織および他の生物材料に適用される処置および診断の方法に関する。該方法のステップの少なくとも一部は、コンピュータで実行される。

20

【0058】

本開示のいくつかの実施形態により、臨床医は母親が妊娠中の胎児の遺伝子の状態を非侵襲的に決定することが可能になり、それにより、胎児の遺伝物質を採取することによって乳児の健康が危険にさらされることがなく、また、母親が侵襲的手順を受ける必要がない。さらに、ある特定の態様では、本開示により、胎児の遺伝子の状態を、高い正確度、例えば、出生前ケアに広く用いられているトリプルテストの、非侵襲的な母系の血清分析物に基づくスクリーニングよりも有意に高い正確度で決定することが可能になる。

30

【0059】

本明細書に開示されている方法の正確度が高いことは、本明細書に記載の、遺伝子型データを分析するためのインフォマティクス手法の結果である。現代の技術的な進歩により、ハイスループット配列決定および遺伝子型決定アレイなどの方法を用いて遺伝子試料から大量の遺伝子情報を測定することができるようになった。本明細書に開示されている方法により、臨床医は利用可能な大量のデータをより大きく活用すること、および胎児の遺伝子の状態のより正確な診断を行うことが可能になる。いくつかの実施形態の詳細が下に示されている。異なる実施形態は、上述のステップの異なる組合せを包含し得る。異なるステップの異なる実施形態の種々の組合せを互換的に用いることができる。

40

【0060】

ある実施形態では、妊娠中の母親から血液試料を取得し、母体起源のDNA、および胎児起源のDNAの両方の混合物を含有する母親の血液の血漿中の浮動性DNAを単離し、胎児の倍数性状態を決定するために使用する。ある実施形態では、本明細書に開示されている方法は、多型対立遺伝子に対応するDNAの混合物中のDNA配列を、対立遺伝子の

50

比および／または対立遺伝子分布が、富化に際してほとんど変わらないままであるように、優先的に富化するステップを包含する。ある実施形態では、本明細書に開示されている方法は、生じた分子の非常に高い百分率が、標的の遺伝子座に対応するように、非常に効率的な標的化PCRに基づく増幅を伴う。ある実施形態では、本明細書に開示されている方法は、母体起源のDNA、および胎児起源のDNAの両方を含有するDNAの混合物について配列決定するステップを包含する。ある実施形態では、本明細書に開示されている方法は、測定された対立遺伝子分布を用いて、母親が妊娠中の胎児の倍数性状態を決定するステップを包含する。ある実施形態では、本明細書に開示されている方法は、決定された倍数性状態を臨床医に報告するステップを包含する。ある実施形態では、本明細書に開示されている方法は、臨床的措置をとるステップ、例えば、絨毛膜絨毛採取または羊水穿刺の侵襲的検査の経過観察を実施するステップ、トリソミーの個体の誕生の準備をするステップ、またはトリソミーの胎児の選択的中絶を包含する。

10

【0061】

本出願は、2006年11月28日出願の米国実用新案出願第11/603,406号（米国特許出願公開第：20070184467）；2008年3月17日出願の米国実用新案出願第12/076,348号（米国特許出願公開第：20080243398）；2009年8月4日出願のPCT実用新案出願第PCT/US09/52730号（PCT公開第WO/2010/017214号）；2010年9月30日出願のPCT実用新案出願第PCT/US10/050824号（PCT公開第WO/2011/041485号）、および2011年5月18日出願の米国実用新案出願第13/110,685号を参照する。本出願において使用される語彙のいくつかは、これらの参考文献にその前例を有し得る。本明細書に記載の概念のいくつかは、これらの参考文献に見いだされる概念に照らして、よりよく理解することができる。

20

【0062】

浮動性胎児DNAを含む母系の血液のスクリーニング

本明細書に記載の方法を用いて、標的の遺伝物質が、ある量の他の遺伝物質の存在下で見いだされる、子、胎児または他の標的個体の遺伝子型の決定を補助することができる。いくつかの実施形態では、遺伝子型とは、1つまたは複数の染色体の倍数性状態を指し得、1つまたは複数の疾患に関連づけられる対立遺伝子またはそのいくつかの組合せを指し得る。本開示では、考察は、胎児DNAが母系の血液に見いだされる場合に胎児の遺伝子の状態を決定することに焦点が当てられるが、この例は、この方法を適用することができる可能性のある状況に限定することを示していない。さらに、該方法は、標的DNAの量が非標的DNAに対していかなる割合で存在する場合にも適用可能であり、例えば、標的DNAは、存在するDNAの0.000001%から99.999999%の間のいずれを構成してもよい。さらに、非標的DNAは、関連性のある非標的個体（複数可）の一部または全部からの遺伝子データが既知である限りは、必ずしも1つの個体由来である必要はなく、さらには関連する個体由来である必要はない。ある実施形態では、本明細書に開示されている方法を用いて、胎児DNAを含有する母系の血液から胎児の遺伝子型データを決定することができる。該方法は、妊娠中の女性の子宮内に複数の胎児がいる場合、または他の混入DNA、例えば、他の既に生まれている同胞由来のDNAが試料に存在する可能性がある場合にも用いることができる。

30

40

【0063】

この技法は、胎児の血液細胞が胎盤絨毛を通じて母系の循環に進入する現象を用いることができる。普通は、胎児の細胞の非常に少数のみが、このように母系の循環に入る（胎児母体間出血についてのKleihauer-Betke検査で陽性になるには不十分である）。胎児の細胞を選別し、さまざまな技法によって解析して特定のDNA配列を探すことができるが、侵襲的手順が本質的に有するリスクは伴わない。この技法は、問題の胎盤組織が胎児と同じ遺伝子型のDNAを含有する場合、胎盤組織のアポトーシス後のDNA放出によって浮動性胎児DNAが母系の循環に進入する現象も用いることができる。母系の血漿中に見いだされる浮動性DNAは、30～40%の胎児DNAと同程度の割合で

50

胎児DNAを含有することが示されている。

【0064】

ある実施形態では、血液を妊娠中の女性から抜き取ることができる。研究により、母系の血液は、母体起源の浮動性DNAに加えて、胎児由来の少量の浮動性DNAを含有し得ることが示された。さらに、母体起源の多くの血液細胞に加えて、胎児起源のDNAを含む脱核胎児血液細胞も存在し得、これは、一般には、核DNAを含有しない。胎児DNAを単離するまたは胎児DNAが富化された画分を作製するための当技術分野で公知の多くの方法が存在する。例えば、クロマトグラフィーにより、胎児DNAが富化された特定の画分が作製されることが示されている。

【0065】

比較的非侵襲的に抜き取られ、ある量の胎児DNAを、細胞性または浮動性のいずれかで、その母系DNAに対する割合に富化されて、またはその元の比率のいずれかで含有する母系の血液、血漿または他の体液の試料を手にしたら、前記試料中に見いだされるDNAの遺伝子型を決定することができる。いくつかの実施形態では、血液は、血液を静脈、例えば、尺側皮静脈から回収するための針を使用して抜き取ることができる。本明細書に記載の方法を用いて、胎児の遺伝子型データを決定することができる。例えば、該方法を用いて、1つまたは複数の染色体における倍数性状態を決定することができ、該方法を用いて、挿入、欠失、および転座を含め、1つのSNPまたはSNPの集合の同一性を決定することができる。該方法を用いて、1つまたは複数の遺伝子型の形体の起源である親を含めた1つまたは複数のハプロタイプを決定することができる。

【0066】

この方法は、任意の遺伝子型決定および／または配列決定方法、例えば、ILLUMINA INFINIUM ARRAYプラットフォーム、AFFYMETRIX GENCHIP、ILLUMINA GENOME ANALYZERまたはLIFE TECHNOLOGIES' SOLID SYSTEMに使用することができる任意の核酸を用いて機能する点に留意されたい。これは、血漿から抽出された浮動性DNAまたはその増幅物（例えば、全ゲノム増幅、PCR）；他の細胞型（例えば、全血由来のヒトリンパ球）由来のゲノムDNAまたはその増幅物を含む。DNAを調製するために、これらのプラットフォームのうちの1つに適したゲノムDNAを生成する任意の抽出または精製方法も同様に機能する。この方法は、RNAの試料を用いて同等に良好に機能し得る。ある実施形態では、試料の保管は、分解が最小限になるように行ない得る（例えば、約-20℃またはそれよりも低い温度で凍結下）。

【0067】

親支援

いくつかの実施形態は、PARENTAL SUPPORT（商標）（PS）法と組み合わせて用いることができ、PARENTAL SUPPORT（商標）（PS）法の複数の実施形態は、その全体が参照により本明細書に組み込まれる、米国特許出願第11/603,406号（米国特許出願公開第20070184467号）、米国特許出願第12/076,348号（米国特許出願公開第20080243398号）、米国特許出願第13/110,685号、PCT出願第PCT/US09/52730号（PCT公開第WO/2010/017214号）、およびPCT出願第PCT/US10/050824号（PCT公開第WO/2011/041485号）に記載されている。PARENTAL SUPPORT（商標）は、遺伝子データを解析するために使用することができる、インフォマティクスに基づく手法である。いくつかの実施形態では、本明細書に開示されている方法は、PARENTAL SUPPORT（商標）法の一部とみなすことができる。いくつかの実施形態では、PARENTAL SUPPORT（商標）法は、標的個体の遺伝子データを高い正確度で、その個体由来の1つまたは少数の細胞の遺伝子データ、または標的個体由来のDNAおよび1つまたは複数の他の個体由来のDNAからなるDNAの混合物の遺伝子データを決定するため、詳細には、標的個体における疾患関連対立遺伝子、他の対象の対立遺伝子、および／または1つまたは複数の染色体の倍数性状

10

20

30

40

50

態を決定するために使用することができる方法の集合である。PARENTAL SUPPORT（商標）とは、これらの方法のいずれも指し得る。PARENTAL SUPPORT（商標）は、インフォマティクスに基づく方法の例である。

【0068】

PARENTAL SUPPORT（商標）法では、既知の親の遺伝子データ、すなわち母親および／または父親のハプロタイプおよび／または二倍体の遺伝子データを、減数分裂の機構および標的DNAの不完全な測定、および場合によっては1つまたは複数の関連する個体の知見と共に、集団に基づく乗換え頻度と一緒に、in silicoで、複数の対立遺伝子における遺伝子型、および／または胚または任意の標的細胞（複数可）の倍数性状態、および重要な遺伝子座に対して位置を決める標的DNAを高い程度の信頼度で再構築するために使用する。PARENTAL SUPPORT（商標）法により、不十分に測定された一塩基多型（SNP）だけでなく、挿入および欠失、ならびに全く測定されなかったSNPまたはDNAの全領域も再構築することができる。さらに、PARENTAL SUPPORT（商標）法により、単一細胞から、複数の疾患に関連づけられる遺伝子座の測定ならびに異数性についてのスクリーニングの両方を行うことができる。いくつかの実施形態では、PARENTAL SUPPORT（商標）法を用いて、1つまたは複数の細胞の遺伝子の状態を決定するために、IVFサイクルの間に生検された胚由来の1つまたは複数の細胞を特徴付けることができる。

10

【0069】

PARENTAL SUPPORT（商標）法により、ノイズを伴う遺伝子データをクリーニングすることが可能になる。これは、関連する個体（親）の遺伝子型を参照として用いて、標的ゲノム（胚）における正確な遺伝子の対立遺伝子を推定することによって行ない得る。PARENTAL SUPPORT（商標）は、少量の遺伝物質しか利用可能でない場合（例えば、PGD）、および遺伝物質の量が限られていることに起因して遺伝子型の直接的な測定が本質的にノイズを伴う場合に特に適し得る。PARENTAL SUPPORT（商標）は、利用可能な遺伝物質のごく一部のみが標的個体由来である場合（例えば、NPD）、および別の個体由来の混入DNAシグナルに起因して遺伝子型の直接的な測定が本質的にノイズを伴う場合に特に適し得る。PARENTAL SUPPORT（商標）法により、従来の順序づけられていない二倍体測定値は対立遺伝子のドロップアウト、ドロップイン、可変性の増幅の偏りおよび他のエラーの率が高いことによって特徴付けることができるが、胚上に非常に正確な規則正しい二倍体の対立遺伝子配列を、染色体セグメントのコピー数と共に再構築することができる。該方法では、基礎をなす遺伝子モデルおよび基礎をなす測定エラーのモデルの両方を使用することができる。遺伝子モデルにより、各SNPにおける対立遺伝子の確率およびSNP間の乗換え確率の両方を決定することができる。対立遺伝子の確率は、各SNPにおいて親から得られたデータに基づいてモデリングすることができ、International HapMap Projectにより開発されたHapMapデータベースから得られたデータに基づいてSNP間の乗換え確率をモデリングすることができる。適切な基礎をなす遺伝子モデルおよび測定エラーモデルを考慮すると、最大事後（MAP）推定を、計算的に効率的にするための改変を伴って用いて、正確な、胚の各SNPにおける規則正しい対立遺伝子値を推定することができる。

20

30

40

【0070】

上で概説した技法により、いくつかの場合には、個体の遺伝子型を、その個体に由来する非常に少量のDNAを考慮して決定することができる。これは、1つまたは少数の細胞由来のDNAであってよい、または、母系の血液中に見いだされる少量の胎児DNA由来であってよい。

【0071】

定義

一塩基多型（SNP）とは、同じ種の2つのメンバーのゲノム間で異なる可能性がある一塩基を指す。この用語の使用は、各変異体が存在する頻度に対するいかなる限定も意味

50

するべきではない。

【0072】

配列とは、DNA配列または遺伝子配列を指す。配列とは、個体のDNA分子または鎖の一次の物理的構造を指し得る。配列とは、DNA分子またはDNA分子の相補鎖に見いだされるヌクレオチドの配列を指し得る。配列とは、*in silico*で表示される、DNA分子に含有される情報を指し得る。

【0073】

遺伝子座とは、個体のDNA上の対象の特定の領域を指し、可能性のある挿入もしくは欠失の部位またはいくつかの他の関連性のある遺伝的変異の部位である、SNPを指し得る。疾患に関連づけられるSNPとは、疾患に関連づけられる遺伝子座を指す場合もある。

10

【0074】

多型対立遺伝子、同様に「多型遺伝子座」とは、所与の種内の個体間で遺伝子型が変動する対立遺伝子または遺伝子座を指す。多型対立遺伝子のいくつかの例としては、一塩基多型、短いタンデム反復、欠失、重複、および逆位が挙げられる。

【0075】

多型部位とは、個体間で変動する多型領域に見いだされる特異的なヌクレオチドを指す。

【0076】

対立遺伝子とは、特定の遺伝子座を占有する遺伝子を指す。

20

【0077】

遺伝子データ、同様に「遺伝子型データ」とは、1つまたは複数の個体のゲノムの態様を記載するデータを指す。これは、1つの遺伝子座または遺伝子座の集合、部分配列または全配列、染色体の部分もしくは染色体の全体、またはゲノム全体を指し得る。これは、1つまたは複数のヌクレオチドの同一性を指し得、これは、逐次的なヌクレオチドの集合またはゲノム内の異なる場所由来のヌクレオチド、またはそれらの組合せを指し得る。遺伝子型データは、一般には*in silico*であるが、化学的にコードされる遺伝子データとして配列内に物理的なヌクレオチドを考えることも可能である。遺伝子型データは、個体（複数可）「on（に関する）」、「of（の）」、「at（における）」、「from（からの）」または「on（に関する）」とすることができる。遺伝子型データとは、これらの測定を遺伝物質に対して行う場合、遺伝子型決定プラットフォームからの出力測定値を指し得る。

30

【0078】

遺伝物質、同様に「遺伝子試料」とは、DNAまたはRNAを含む1つまたは複数の個体由来の組織または血液などの物理的物質を指す。

【0079】

ノイズを伴う遺伝子データとは、以下のいずれかを伴う遺伝子データを指す：対立遺伝子ドロップアウト、不確実な塩基対測定値、不正確な塩基対測定値、塩基対測定値の欠落、挿入または欠失の不確実な測定値、染色体セグメントコピー数の不確実な測定値、偽のシグナル、測定値の欠落、他のエラー、またはそれらの組合せ。

40

【0080】

信頼度とは、呼び出されたSNP、対立遺伝子、対立遺伝子の集合、倍数性呼び出しまたは決定された染色体セグメントコピーの数が個体の実際の遺伝子の状態を正確に示す統計学的尤度を指す。

【0081】

倍数性呼び出し、同様に「染色体コピー数呼び出し」または「コピー数呼び出し」（CNC）、とは、細胞内に存在する1つまたは複数の染色体の量および／または染色体の同一性を決定する行為を指し得る。

【0082】

異数性とは、誤った数の染色体が細胞に存在する状態を指す。ヒト体細胞の場合には、

50

異数性とは、細胞が、22対の常染色体および1対の性染色体を含有しない場合を指し得る。ヒト配偶子の場合には、異数性とは、細胞が、23種の染色体のそれぞれのうちの1つを含有しない場合を指し得る。単一染色体型の場合には、異数性とは、大体2つの相同であるが同一ではない染色体コピーが存在する場合、または同じ親を起源とする2つの染色体コピーが存在する場合を指し得る。

【0083】

倍数性状態とは、細胞における1つまたは複数の染色体型の量および／または染色体の同一性を指す。

【0084】

染色体とは、単一染色体コピーを指し得、これは正常な体細胞に46個存在するDNAの単一分子を意味し、その例は「母体由来の第18染色体」である。染色体とは、正常なヒト体細胞に23個存在する染色体型を指す場合もあり、例は「第18染色体」である。

【0085】

染色体の同一性とは、指示対象の染色体数、すなわち染色体型を指し得る。正常なヒトは、22種類の番号が付された常染色体型、および2種類の性染色体を有する。染色体の同一性とは、親起源の染色体を指す場合もある。染色体の同一性とは、親から遺伝によって受け継がれる特定の染色体を指す場合もある。染色体の同一性とは、染色体の他の同定される形体 (feature) を指す場合もある。

【0086】

遺伝物質の状態または単に「遺伝子の状態」とは、DNA上のSNPの集合の同一性、相が特定された (phased) 遺伝物質のハプロタイプ、および挿入、欠失、反復および変異を含めたDNAの配列を指し得る。これは、1つまたは複数の染色体の倍数性状態、染色体セグメントまたは染色体セグメントの集合を指す場合もある。

【0087】

対立遺伝子データとは、1つまたは複数の対立遺伝子の集合に関する遺伝子型データの集合を指す。対立遺伝子データとは、相が特定されたハプロタイプデータを指し得る。対立遺伝子データとは、SNPの同一性を指し得、対立遺伝子データとは、挿入、欠失、反復および変異を含めたDNAの配列データを指し得る。対立遺伝子データとは、親起源の各対立遺伝子を包含し得る。

【0088】

対立遺伝子の状態とは、1つまたは複数の対立遺伝子の集合内の遺伝子の実際の状態を指す。対立遺伝子の状態とは、対立遺伝子データに記載されている遺伝子の実際の状態を指し得る。

【0089】

対立遺伝子の比 (allellic ratio) または対立遺伝子の比 (allelle ratio) とは、試料または個体に存在する遺伝子座における各対立遺伝子の量の間の比を指す。試料を配列決定によって測定した場合、対立遺伝子の比とは、遺伝子座における各対立遺伝子にマッピングされる配列読み取りの比を指し得る。試料を強度に基づく測定方法によって測定した場合、対立遺伝子の比とは、測定方法によって推定される遺伝子座に存在する各対立遺伝子の量の比を指し得る。

【0090】

対立遺伝子数とは、特定の遺伝子座にマッピングされる配列の数を指し、その遺伝子座が多型である場合、対立遺伝子数とは、対立遺伝子のそれぞれにマッピングされる配列の数を指す。各対立遺伝子がバイナリー様式でカウントされる場合は、対立遺伝子数は整数になる。対立遺伝子が確率的にカウントされる場合は、対立遺伝子数は分数であり得る。

【0091】

対立遺伝子数確率とは、マッピングの確率と組み合わせた、特定の遺伝子座または多型遺伝子座における対立遺伝子の集合にマッピングされる可能性がある配列の数を指す。カウントされた配列のそれぞれについてのマッピングの確率がバイナリーである (0または1) 対立遺伝子数は、対立遺伝子数の確率と等しいことに留意されたい。いくつかの実施

10

20

30

40

50

形態では、対立遺伝子数の確率はバイナリーであってよい。いくつかの実施形態では、対立遺伝子数の確率は、DNA測定値と等しくなるように設定することができる。

【0092】

対立遺伝子分布または「対立遺伝子数分布」とは、遺伝子座の集合内の各遺伝子座に存在する各対立遺伝子の相対量を指す。対立遺伝子分布は、個体、試料または試料に対して得た測定値の集合を指す場合がある。配列決定との関連において、対立遺伝子分布とは、多型遺伝子座の集合内の各対立遺伝子についての、特定の対立遺伝子にマッピングされる読み取りの数または見込み数を指す。対立遺伝子測定値は、確率的に処理することができる、すなわち所与の対立遺伝子が所与の配列読み取りを示す尤度は、0から1の間の分数である、または対立遺伝子測定値は、バイナリー様式で処理することができる、すなわち任意の所与の読み取りは、特定の対立遺伝子のちょうど0コピーまたは1コピーであると考えられる。

10

【0093】

対立遺伝子分布パターンとは、異なる親の状況についての異なる対立遺伝子分布の集合を指す。特定の対立遺伝子分布パターンにより、特定の倍数性状態が示され得る。

【0094】

対立遺伝子の偏りとは、ヘテロ接合性遺伝子座における測定された対立遺伝子の比が、元のDNAの試料に存在していた比と異なる程度を指す。特定の遺伝子座における対立遺伝子の偏りの程度は、その遺伝子座において観察された対立遺伝子比を測定し、その遺伝子座における元のDNA試料中の対立遺伝子の比で割ったものと等しい。対立遺伝子の偏りは1より大きいと定義することができ、したがって、対立遺伝子の偏りの程度の算出によって1未満の値xが生じる場合、対立遺伝子の偏りの程度は $1/x$ と言い換えることができる。対立遺伝子の偏りは、増幅の偏り、精製の偏りまたは異なる対立遺伝子に違うように影響を及ぼすいくつかの他の現象に起因し得る。

20

【0095】

プライマー、同様に「PCRプローブ」とは、単一のDNA分子（DNAオリゴマー）またはDNA分子（DNAオリゴマー）の集団を指し、DNA分子は同一またはほぼ同一であり、プライマーは、標的の多型遺伝子座とハイブリダイズするように設計された領域を含有しており、PCR増幅が可能になるように設計されたプライミング配列を含有してよい。プライマーは、分子バーコードも含有してよい。プライマーは、個々の分子それぞれについて異なるランダムな領域を含有してよい。

30

【0096】

ハイブリッド捕捉プローブとは、PCRまたは直接的な合成などの種々の方法によって生成され、試料中の特異的な標的DNA配列の一方の鎖と相補的であることが意図された、場合によっては改変された任意の核酸配列を指す。外因性のハイブリッド捕捉プローブを調製された試料に加え、変性-再アニーリングプロセスを通じてハイブリダイズさせて、外因性断片-内因性断片の2重鎖を形成することができる。次いで、これらの2重鎖を、種々の手段によって試料から物理的に分離することができる。

【0097】

配列読み取りとは、クローン配列決定法を用いて測定したヌクレオチド塩基の配列を示すデータを指す。クローン配列決定により、単一の1つの元のDNA分子または1つの元のDNA分子のクローンまたは1つの元のDNA分子のクラスターを示す配列データを生じることができる。配列読み取りは、配列の各塩基の位置において、ヌクレオチドが正確に呼び出されている確率を示す関連する品質スコアも有し得る。

40

【0098】

配列読み取りのマッピングとは、特定の生物体のゲノム内配列における配列読み取りの開始場所を決定するプロセスである。配列読み取りの開始場所は、読み取りとゲノム配列のヌクレオチド配列の類似性に基づく。

【0099】

一致コピーエラー、同様に「一致染色体異数性」（MCA）とは、1つの細胞が、2つ

50

の同一またはほぼ同一の染色体を含有する異数性の状態を指す。この種類の異数性は、減数分裂における配偶子形成の間に生じる可能性があり、減数分裂不分離エラーと称することができる。この種類のエラーは、有糸分裂において生じる可能性がある。一致トリソミーとは、所与の染色体の3つのコピーが個体に存在し、コピーのうちの2つが同一である場合を指し得る。

【0100】

不一致コピーエラー、同様に「独自の染色体異数性 (unique chromosome aneuploidy)」(UCA)とは、1つの細胞が同じ親由来であり、かつ相同であるが同一ではない可能性がある2つの染色体を含有する異数性の状態を指す。この種類の異数性は、減数分裂の間に生じる可能性があり、減数分裂エラーと称することができる。不一致トリソミーとは、所与の染色体の3つのコピーが個体に存在し、コピーのうちの2つが同じ親由来であり、かつ相同であるが同一ではない場合を指し得る。不一致トリソミーとは、一方の親由来の2つの相同染色体が存在し、染色体の一部のセグメントは同一であるが他のセグメントはただ単に相同なだけである場合を指し得ることに留意されたい。

10

【0101】

相同染色体とは、通常は減数分裂の間に対合する同じ遺伝子の集合を含有する染色体コピーを指す。

【0102】

同一染色体とは、同じ遺伝子の集合を含有する染色体コピーを指し、各遺伝子について、同一染色体は同一またはほぼ同一である同じ対立遺伝子の集合を有する。

20

【0103】

対立遺伝子ドロップアウト (ADO)とは、所与の対立遺伝子における相同染色体由来の塩基対の集合内の塩基対の少なくとも一方が検出されない状況を指す。

【0104】

遺伝子座ドロップアウト (LDO)とは、所与の対立遺伝子における相同染色体由来の塩基対の集合内の塩基対の両方が検出されない状況を指す。

【0105】

ホモ接合性とは、対応する染色体の遺伝子座と同様の対立遺伝子を有することを指す。

【0106】

ヘテロ接合性とは、対応する染色体の遺伝子座と同様でない対立遺伝子を有することを指す。

30

【0107】

ヘテロ接合性率とは、集団内の個体が所与の遺伝子座においてヘテロ接合性対立遺伝子を有する率を指す。ヘテロ接合性率は、個体またはDNAの試料中の所与の遺伝子座における予測された対立遺伝子の比または測定された対立遺伝子の比を指す場合もある。

【0108】

情報価値が高い一塩基多型 (HISNP)とは、胎児が母親の遺伝子型には存在しない対立遺伝子を有するSNPを指す。

【0109】

染色体領域とは、染色体のセグメントまたは完全な染色体を指す。

40

【0110】

染色体のセグメントとは、1つの塩基対から染色体全体までサイズに幅があり得る染色体のセクションを指す。

【0111】

染色体とは、完全な染色体または染色体のセグメントもしくはセクションのいずれかを指す。

【0112】

コピーとは、染色体セグメントのコピーの数を指す。それは、染色体セグメントの同一のコピー、または同一ではない相同なコピーを指す場合があり、染色体セグメントの異な

50

るコピーは実質的に類似した遺伝子座の集合を含有し、対立遺伝子のうちの1つまたは複数異なる。M2コピーエラーなどの異数性のいくつかの場合には、所与の染色体セグメントの同一であるいくつかのコピーならびに同じ染色体セグメントの同一ではないいくつかのコピーを有する可能性があることに留意されたい。

【0113】

ハプロタイプとは、一般には、同じ染色体上で一緒に遺伝によって受け継がれる、複数の遺伝子座における対立遺伝子の組合せを指す。ハプロタイプとは、所与の遺伝子座の集合の間で起こった組換え事象の数に応じて、わずか2つの遺伝子座または染色体全体を指し得る。ハプロタイプとは、統計学的に関連する単一の染色分体上の一塩基多型（SNP）の集合も指す場合がある。

10

【0114】

ハプロタイプデータ、同様に「相が特定されたデータ」または「順序づけられた遺伝子データ」とは、二倍体ゲノムまたは倍数体ゲノムの単一染色体、すなわち、二倍体ゲノムの染色体の分離された母系のコピーまたは父系のコピーのいずれかからのデータを指す。

【0115】

相の特定（Phasing）とは、順序づけられていない、二倍体（または倍数性）遺伝子データを考慮して個体のハプロタイプ遺伝子データを決定する行為を指す。相の特定とは、1つの染色体上に見いだされる対立遺伝子の集合について、対立遺伝子の2つの遺伝子のどちらが、個体の2つの相同染色体のそれぞれと関連するかを決定する行為を指し得る。

20

【0116】

相が特定されたデータとは、1つまたは複数のハプロタイプが決定された遺伝子データを指す。

【0117】

仮説とは、所与の染色体の集合における可能性のある倍数性状態または所与の遺伝子座の集合における可能性のある対立遺伝子の状態の集合を指す。可能性の集合は、1つまたは複数のエレメントを含んでよい。

【0118】

コピー数仮説、同様に「倍数性状態仮説」とは、個体の染色体のコピーの数に関する仮説を指す。これは、各染色体の起源となる親、および親の2つの染色体のどちらが個体に存在するかを含めた、染色体のそれぞれの同一性に関する仮説を指す場合もある。これは、もしあれば、関連する個体由来の染色体または染色体セグメントのいずれが、個体由来の所与の染色体に遺伝的に対応するかに関する仮説を指す場合もある。

30

【0119】

標的個体とは、遺伝子の状態が決定される個体を指す。いくつかの実施形態では、限られた量のDNAのみが標的個体から入手可能である。いくつかの実施形態では、標的個体は胎児である。いくつかの実施形態では、2体以上の標的個体が存在し得る。いくつかの実施形態では、一对の親から生まれた胎児をそれぞれ標的個体とみなすことができる。いくつかの実施形態では、決定される遺伝子データは、1つの対立遺伝子の呼び出しまたは対立遺伝子の呼び出しの集合である。いくつかの実施形態では、決定される遺伝子データは、倍数性呼び出しである。

40

【0120】

関連する個体とは、標的個体と遺伝的に関連する、したがって、標的個体とハプロタイプブロックを共有する任意の個体を指す。ある状況では、関連する個体は、標的個体の遺伝学的な親、または親に由来する任意の遺伝物質、例えば、精子、極体、胚、胎児または子であってよい。関連する個体とは、同胞、親または祖父母を指す場合もある。

【0121】

同胞とは、遺伝学的な親が問題の個体と同じである任意の個体を指す。いくつかの実施形態では、同胞とは、生まれた子、胚もしくは胎児、または生まれた子、胚もしくは胎児に由来する1つまたは複数の細胞を指し得る。同胞とは、親の一方を起源とする一倍体個

50

体、例えば、精子、極体または任意の他のハプロタイプの遺伝物質の集合を指す場合もある。個体は、それ自体を同胞とみなすことができる。

【0122】

胎児の (fetal) とは、「胎児の (of the fetus)」、または「胎児と遺伝的に同様である胎盤の領域の (of the region of the placenta that is genetically similar to the fetus)」を指す。妊娠中の女性では、胎盤のいくらかの部分は胎児と遺伝的に同様であり、母系の血液に見いだされる浮動性胎児DNAは、遺伝子型が胎児と一致する胎盤の部分を起源とし得る。胎児の染色体の半分の遺伝子情報は、胎児の母親から遺伝によって受け継がれることに留意されたい。いくつかの実施形態では、胎児の細胞に由来するこれらの母系的に遺伝によって受け継がれた染色体からのDNAは、「母体起源の (of maternal origin)」ものではなく「胎児起源の (of fetal origin)」ものと考えられる。

10

【0123】

胎児起源のDNAとは、元々は遺伝子型が基本的に胎児の遺伝子型と等しい細胞の一部であったDNAを指す。

【0124】

母体起源のDNAとは、元々は遺伝子型が基本的に母親の遺伝子型と等しい細胞の一部であったDNAを指す。

【0125】

20

子とは、胚、割球または胎児を指し得る。ここで開示されている実施形態では、記載されている概念は、生まれた子、胎児、胚またはそれら由来の細胞の集合である個体に同等に良好に当てはまることに留意されたい。子という用語の使用は、単に子と称される個体が親の遺伝学的子孫であることを内包することを意味する。

【0126】

親とは、個体の遺伝学的母親または父親を指す。個体は、一般には、2体の親、母親および父親を有するが、これは、例えば、遺伝子または染色体のキメラ現象において、必ずしもそうではない。親は個体とみなすことができる。

【0127】

親の状況とは、標的の2体の親の一方または両方について、2つの関連性のある染色体のそれぞれにおける所与のSNPの遺伝子の状態を指す。

30

【0128】

所望の通り発生させる、同様に「正常に発生させる」とは、成長可能な胚を子宮内に着床させ、妊娠をもたらすこと、および／または、妊娠を継続し、出生をもたらすこと、および／または、生まれた子に染色体異常がないこと、および／または、生まれた子に他の望ましくない遺伝子の状態、例えば、疾患に関連づけられる遺伝子がないことを指す。「所望の通り発生させる」という用語は、親または健康管理の補助者により所望され得るいかなるものも包含することを意味する。いくつかの場合には、「所望の通り発生させる」とは、医学的な研究または他の目的に有用である成長できない胚または成長可能な胚を指し得る。

40

【0129】

子宮への挿入とは、in vitroでの受精との関連において胚を子宮腔に移入するプロセスを指す。

【0130】

母系の血漿とは、妊娠中の女性由来の血液の血漿部分を指す。

【0131】

臨床的決定とは、個体の健康または生存に影響を及ぼす転帰を有する措置を取るか取らないかの任意の決定を指す。出生前診断との関連において、臨床的決定とは、胎児を流産するか流産しないかの決定を指し得る。臨床的決定とは、さらなる検査を行うこと、望ましくない表現型を減ずるための措置を取ることを、または異常を持つ子の誕生の準備をする

50

ための措置を取ることに決定を指す場合もある。

【0132】

診断ボックスとは、本明細書に開示されている方法の1つまたは複数の態様を実施するために設計された1つの機械またはその機械の組合せを指す。ある実施形態では、診断ボックスは、患者をケアする所に置くことができる。ある実施形態では、診断ボックスにより、標的化増幅、その後の配列決定を実施することができる。ある実施形態では、診断ボックスは、単独で、または技師の補助で機能し得る。

【0133】

インフォマティクスに基づく方法とは、大量のデータを解明するために、統計量に大きく依拠する方法を指す。出生前診断との関連において、インフォマティクスに基づく方法とは、1つまたは複数の染色体における倍数性状態または1つまたは複数の対立遺伝子における対立遺伝子の状態を、状態を直接物理的に測定することによってではなく、例えば、分子アレイまたは配列決定からの大量の遺伝子データを考慮して、最も可能性が高い状態を統計学的に推論することによって決定するために設計された方法を指す。本開示のある実施形態では、インフォマティクスに基づく技法は、本特許に開示されているものであってよい。本開示のある実施形態では、インフォマティクスに基づく技法はPARENTAL SUPPORT（商標）であってよい。

【0134】

一次遺伝子データとは、遺伝子型決定プラットフォームから出力されるアナログの強度シグナルを指す。SNPアレイとの関連において、一次遺伝子データとは、いかなる遺伝子型呼び出しも行われる前の強度シグナルを指す。配列決定との関連において、一次遺伝子データとは、いかなる塩基対の同一性も決定される前、および配列がゲノムにマッピングされる前にシーケンサーから生じる、クロマトグラムと類似しているアナログ測定値を指す。

【0135】

二次遺伝子データとは、遺伝子型決定プラットフォームから出力される加工された遺伝子データを指す。SNPアレイとの関連において、二次遺伝子データとは、SNPアレイリーダーに付随するソフトウェアによって行われる対立遺伝子呼び出しを指し、該ソフトウェアにより、試料中に所与の対立遺伝子が存在するか存在しないかの呼び出しが行われる。配列決定との関連において、二次遺伝子データとは、配列の塩基対の同一性が決定されることを指し、場合によっては、同様に、配列がゲノムにマッピングされたことを指す。

【0136】

非侵襲的な出生前診断（NPД）または、同様に「非侵襲的な出生前スクリーニング」（NP S）とは、母親の血液に見いだされる遺伝物質を用いて、母親が妊娠中の胎児の遺伝子の状態を決定する方法を指し、遺伝物質は、母親の静脈内血液を抜き取ることによって得る。

【0137】

遺伝子座に対応するDNAを優先的に富化すること、または遺伝子座におけるDNAを優先的に富化することは、その遺伝子座に対応する富化後のDNA混合物中のDNA分子の百分率を、その遺伝子座に対応する富化前のDNA混合物中のDNA分子の百分率よりも高くする任意の方法を指す。該方法は、遺伝子座に対応するDNA分子の選択的増幅を包含し得る。該方法は、遺伝子座に対応しないDNA分子を除去するステップを包含し得る。該方法は、方法の組合せを包含し得る。富化の程度は、その遺伝子座に対応する富化後の混合物におけるDNA分子の百分率を、その遺伝子座に対応する富化前の混合物におけるDNA分子の百分率で割ったものと定義される。優先的な富化は、複数の遺伝子座において行うことができる。本開示のいくつかの実施形態では、富化の程度は20を超える。本開示のいくつかの実施形態では、富化の程度は200を超える。本開示のいくつかの実施形態では、富化の程度は2,000を超える。優先的な富化を複数の遺伝子座において行う場合、富化の程度とは、遺伝子座の集合内の全ての遺伝子座の平均の富化の程度を

10

20

30

40

50

指し得る。

【0138】

増幅とは、DNA分子のコピーの数を増加させる方法を指す。

【0139】

選択的な増幅とは、特定のDNA分子またはDNAの特定の領域に対応するDNA分子のコピーの数を増加させる方法を指し得る。選択的な増幅とは、特定の標的のDNA分子または標的のDNAの領域のコピーの数を、DNAの標識していない分子または領域を増大させるよりも増加させる方法を指す場合もある。選択的な増幅は、優先的に富化する方法であってよい。

【0140】

ユニバーサルプライミング配列とは、標的DNA分子の集団に、例えば、ライゲーション、PCRまたはライゲーション媒介性PCRによって付加することができるDNA配列を指す。標的分子の集団に付加したら、ユニバーサルプライミング配列に特異的なプライマーを用いて、増幅プライマーの単一の対を使用して標的集団を増幅することができる。ユニバーサルプライミング配列は、一般には、標的配列に関連しない。

【0141】

ユニバーサルアダプタまたは「ライゲーションアダプタ」または「ライブラリータグ」は、標的二本鎖DNA分子の集団の5'末端および3'末端に共有結合的に連結することができるユニバーサルプライミング配列を含有するDNA分子である。アダプタを付加することにより、そこからPCR増幅を行うことができる標的集団の5'末端および3'末端にユニバーサルプライミング配列がもたらされ、標的集団由来の全ての分子を、増幅プライマーの単一の対を使用して増幅する。

【0142】

標的化とは、DNAの混合物中の遺伝子座の集合に対応するDNA分子を選択的に増幅する、または別の方法で優先的に富化するために使用される方法を指す。

【0143】

同時分布モデルとは、複数のランダムな変数に関して定義済みの事象の確率を、変数の確率が関連づけられている、同じ確率空間に対して定義済みの複数のランダムな変数を考慮して、定義するモデルを指す。いくつかの実施形態では、変数の確率が関連づけられていない退化事例を用いることができる。

【0144】

仮説

本開示との関連において、仮説とは、可能性のある遺伝子の状態を指す。仮説とは、可能性のある倍数性状態を指し得る。仮説とは、可能性のある対立遺伝子の状態を指し得る。仮説の集合とは、可能性のある遺伝子の状態の集合、可能性のある対立遺伝子の状態の集合、可能性のある倍数性状態の集合、またはそれらの組合せを指し得る。いくつかの実施形態では、仮説の集合は、集合からの1つの仮説が、任意の所与の個体の実際の遺伝子の状態に対応するように設計することができる。いくつかの実施形態では、仮説の集合は、あらゆる可能性のある遺伝子の状態が、集合からの少なくとも1つの仮説によって記載することができるように設計することができる。本開示のいくつかの実施形態では、方法の一態様は、どの仮説が問題の個体の実際の遺伝子の状態に対応するかを決定することである。

【0145】

本開示の別の実施形態では、1つのステップは仮説を作製するステップを包含する。いくつかの実施形態では、仮説は、コピー数仮説であってよい。いくつかの実施形態では、仮説は、関連する個体のそれぞれ由来のどの染色体のセグメントが、もしあれば、他の関連する個体のどのセグメントに遺伝的に対応するかに関する仮説を包含する。仮説を作製することとは、変数の限度を、考慮中の可能性のある遺伝子の状態の全集合がそれらの変数に包含されるように設定する行為を指し得る。

【0146】

「コピー数仮説」は、「倍数性仮説」または「倍数性状態仮説」とも称され、標的個体における所与の染色体コピー、染色体型または染色体のセクションについての可能性のある倍数性状態に関する仮説を指し得る。これは、個体の2種以上の染色体型における倍数性状態を指す場合もある。コピー数仮説の集合とは、仮説の集合を指し得、各仮説は、個体における可能性のある異なる倍数性状態に対応する。仮説の集合は、可能性のある倍数性状態の集合、可能性のある親のハプロタイプの寄与の集合、混合試料中の可能性のある胎児DNAの百分率の集合、またはそれらの組合せに関し得る。

【0147】

正常な個体は、各親由来の各染色体型のうちの1つを含有する。しかし、減数分裂および有糸分裂におけるエラーに起因して、個体が、各親由来の所与の染色体型を0個、1個、2個、またはそれより多くを有する可能性がある。実際には、親由来の所与の染色体が3つ以上認められることはまれである。本開示では、いくつかの実施形態では、所与の染色体の0コピー、1コピーまたは2コピーが親に由来する、可能性のある仮説のみを考慮し、親を起源とするいくつかの可能性のあるコピーを考慮することは自明の拡張である。いくつかの実施形態では、所与の染色体に対して、可能性のある仮説が9つある：母体起源の0個の染色体、1個の染色体または2個の染色体に関する3つの可能性のある仮説に父系起源の0個の染色体、1個の染色体または2個の染色体に関する3つの可能性のある仮説を掛け合わせたもの。(m, f)を、mが母親から遺伝によって受け継がれた所与の染色体の数であり、およびfが父親から遺伝によって受け継がれた所与の染色体の数である仮説を指すものとする。したがって、9つの仮説は、(0, 0)、(0, 1)、(0, 2)、(1, 0)、(1, 1)、(1, 2)、(2, 0)、(2, 1)、および(2, 2)である。これらは、 H_{00} 、 H_{01} 、 H_{02} 、 H_{10} 、 H_{12} 、 H_{20} 、 H_{21} 、および H_{22} と記載することもできる。異なる仮説は、異なる倍数性状態に対応する。例えば、(1, 1)とは、通常のダイソミー染色体を指し、(2, 1)とは、母系トリソミーを指し、(0, 1)とは、父系モノソミーを指す。いくつかの実施形態では、2つの染色体が一方の親から遺伝によって受け継がれ、1つの染色体が他方の親から遺伝によって受け継がれる場合は、2つの場合にさらに分けられ得る：2つの染色体が同一である場合（一致コピーエラー）と、2つの染色体が相同であるが同一ではない場合（不一致コピーエラー）。これらの実施形態では、可能性のある仮説が16ある。他の仮説の集合、および異なる数の仮説を使用することが可能であることが理解されるべきである。

【0148】

本開示のいくつかの実施形態では、倍数性仮説とは、他の関連する個体由来の染色体のいずれが、標的個体のゲノムに見いだされる染色体に対応するかに関する仮説を指す。いくつかの実施形態では、方法の鍵となるのは、関連する個体がハプロタイプブロックを共有することが予測され得るという事実であり、関連する個体から測定された遺伝子データを、どのハプロタイプブロックが標的個体と関連する個体との間で一致するかの知見と一緒に用いると、標的個体の遺伝子測定値を単独で用いるよりも高い信頼度で標的個体についての正確な遺伝子データを推論することが可能である。従って、いくつかの実施形態では、倍数性仮説は、染色体の数だけでなく、関連する個体のどの染色体が、標的個体の1つまたは複数の染色体と同一またはほぼ同一であるかに関し得る。

【0149】

仮説の集合が定義されたら、アルゴリズムが入力遺伝子データに対して作動すると、考慮中の仮説のそれぞれについて、決定された統計学的な確率が出力され得る。種々の仮説の確率は、種々の仮説のそれぞれについて、専門技法、アルゴリズム、および／または本開示の他の箇所に記載されている方法のうちの1つまたは複数により示された確率が等しい値を、関連性のある遺伝子データを入力として用いて数学的に算出することによって決定することができる。

【0150】

複数の技法によって決定された通り、異なる仮説の確率が推定されたら、それらを組み合わせることができる。これは、各仮説について、各技法によって決定された確率を掛け

10

20

30

40

50

算することを必要とし得る。仮説の確率の積を正規化することができる。1つの倍数性仮説は、染色体についての1つの可能性のある倍数性状態を指す。

【0151】

「確率の組合せ」プロセスは、「仮説の組合せ」または専門技法の結果の組合せとも称され、線形代数の当業者によく知られているはずの概念である。1つの可能性のある確率の組み合わせ方は以下の通りである：専門技法を用いて、遺伝子データの集合を考慮して仮説の集合を評価する場合、方法の出力は、仮説の集合内の各仮説と1対1に関連する確率の集合である。そのそれぞれが集合内の仮説のうちの1つと関連づけられる、第1の専門技法によって決定された確率の集合を、そのそれぞれが同じ仮説の集合と関連づけられる、第2の専門技法によって決定された確率の集合と組み合わせる場合、確率の2つの集合を掛け算する。これは、集合内の各仮説について、2つの専門方法によって決定された、その仮説と関連づけられる2つの確率を掛け合わせ、対応する積が出力確率であることを意味する。このプロセスは、任意の数の専門技法に拡大することができる。ただ1つの専門技法を用いる場合、出力確率は入力確率と同じである。3つ以上の専門技法を用いる場合、関連性のある確率を同時に掛け算することができる。積は、仮説の集合内の仮説の確率が合計で100%になるように正規化することができる。

10

【0152】

いくつかの実施形態では、所与の仮説についての複合確率が他の仮説のいずれかについての複合確率を超える場合、その仮説が、最も可能性が高いと決定されるとみなすことができる。いくつかの実施形態では、正規化された確率が閾値を超えた場合、仮説を、最も可能性が高いと決定することができ、倍数性状態または他の遺伝子の状態を呼び出すことができる。ある実施形態では、これは、その仮説に関連づけられる染色体の数および同一性を、倍数性状態として呼び出すことができることを意味し得る。ある実施形態では、これは、その仮説に関連づけられる対立遺伝子の同一性を、対立遺伝子の状態として呼び出すことができることを意味し得る。いくつかの実施形態では、閾値は、約50%から約80%の間であり得る。いくつかの実施形態では、閾値は、約80%から約90%の間であり得る。いくつかの実施形態では、閾値は、約90%から約95%の間であり得る。いくつかの実施形態では、閾値は、約95%から約99%の間であり得る。いくつかの実施形態では、閾値は、約99%から約99.9%の間であり得る。いくつかの実施形態では、閾値は、約99.9%超であり得る。

20

30

親の状況

親の状況とは、標的の2体の親の一方または両方についての、2つの関連性のある染色体のそれぞれの所与の対立遺伝子の遺伝子の状態を指す。ある実施形態では、親の状況とは、標的の対立遺伝子の状態を指すのではなく、親の対立遺伝子の状態を指すことに留意されたい。所与のSNPについての親の状況は、父系の2つと母系の2つの、4塩基対からなっており、これらは互いに同じであってよい、または異なってよい。「 $m_1 m_2 \mid f_1 f_2$ 」と書くことが一般的であり、ここで m_1 および m_2 は、2つの母系染色体上の所与のSNPの遺伝子の状態であり、 f_1 および f_2 は2つの父系染色体上の所与のSNPの遺伝子の状態である。いくつかの実施形態では、親の状況は、「 $f_1 f_2 \mid m_1 m_2$ 」と書くことができる。下付き文字の「1」および「2」は、第1の染色体および第2の染色体の所与の対立遺伝子における遺伝子型を示すことに留意されたい。どの染色体を「1」とし、どの染色体を「2」とするかの選択は任意であることにも留意されたい。

40

【0153】

本開示では、塩基対の同一性を一般的に示すために、多くの場合、AおよびBを使用することに留意されたい；AまたはBは、C（シトシン）、G（グアニン）、A（アデニン）またはT（チミン）を同等に上手く示すことができる。例えば、所与のSNPに基づく対立遺伝子において、母親の遺伝子型が1つの染色体上のそのSNPにおいてTであり、相同染色体上のそのSNPにおいてGであり、その対立遺伝子における父親の遺伝子型が、相同染色体の両方のそのSNPにおいてGであった場合、標的個体の対立遺伝子が親の状況 $AB \mid BB$ を有するということができ、対立遺伝子が親の状況 $AB \mid AA$ を有すると

50

いうこともできる。理論上、4種の可能性のあるヌクレオチドはいずれも所与の対立遺伝子に存在し得、したがって、例えば、所与の対立遺伝子において母親が遺伝子型ATを有し、父親が遺伝子型GCを有する可能性があることに留意されたい。しかし、経験的なデータにより、ほとんどの場合、所与の対立遺伝子において4種の可能性のある塩基対のうち2種のみが観察されることが示されている。例えば、単一のタンDEM反復を用いた場合、2超、4超、さらには10超の親の状況を有する可能性がある。本開示の考察では、所与の対立遺伝子において2種の可能性のある塩基対のみが観察されると仮定するが、本明細書に開示されている実施形態は、この仮定が当てはまらない場合を考慮に入れるように変更することができる。

【0154】

「親の状況」とは、同じ親の状況を有する標的SNPの集合またはサブセットを指し得る。例えば、標的個体の所与の染色体上の1000個の対立遺伝子を測定する場合、状況AA|BBとは、標的の母親の遺伝子型がホモ接合性であり、標的の父親の遺伝子型がホモ接合性であるが、その遺伝子座における母系の遺伝子型と父系の遺伝子型が同様でない、1,000個の対立遺伝子群内の全ての対立遺伝子の集合を示し得る。親のデータについて相が特定されない、したがって、AB=BAである場合は、可能性のある親の状況は9ある：AA|AA、AA|AB、AA|BB、AB|AA、AB|AB、AB|BB、BB|AA、BB|AB、およびBB|BB。親のデータについて相が特定される、したがって、AB≠BAである場合は、可能性のある異なる親の状況が16ある：AA|AA、AA|AB、AA|BA、AA|BB、AB|AA、AB|AB、AB|BA、AB|BB、BA|AA、BA|AB、BA|BA、BA|BB、BB|AA、BB|AB、BB|BA、およびBB|BB。性染色体上の一部のSNPを除いて、染色体上のあらゆるSNP対立遺伝子が、これらの親の状況のうちの1つを有する。一方の親についての親の状況がヘテロ接合性であるSNPの集合は、ヘテロ接合性の状況と称することができる。

【0155】

NPDにおける親の状況の使用

非侵襲的な出生前診断は、非侵襲的に、例えば、妊娠中の母親に対する採血によって得られる遺伝物質から胎児の遺伝子の状態を決定するために用いることができる重要な技法である。血液を分離し、血漿単離し、その後血漿DNAを単離することができる。サイズ選択を用いて、適切な長さのDNAを単離することができる。DNAを遺伝子座の集合において優先的に富化することができる。次いで、このDNAを、遺伝子型決定アレイにハイブリダイズさせ、蛍光を測定することによって、またはハイスループットシーケンサーで配列決定することによる、いくつかの手段によって測定することができる。

【0156】

非侵襲的な出生前診断との関連において胎児の倍数性呼び出しのために配列決定を使用する場合、配列データを使用するいくつかのやり方がある。配列データを使用することができる最も一般的なやり方は、単に所与の染色体にマッピングされる読み取りの数をカウントすることである。例えば、胎児の第21染色体の倍数性状態を決定しようとするときを考える。さらに、試料中のDNAの10%が胎児起源のDNAで構成され、90%が母体起源のDNAで構成されるときを考える。この場合、ダイソミーであることが予測され得る染色体、例えば、第3染色体の読み取りの平均の数を調べ、それを、読み取りを独特の配列の一部である染色体上の塩基対の数について調整した第21染色体上の読み取りの数と比較する。胎児が正倍数性であった場合、ゲノムの単位当たりのDNAの量は全ての場所においてほぼ同等であることが予想される（確率的変動を受けやすい）。他方では、胎児が第21染色体においてトリソミーであった場合、第21染色体由来の遺伝単位当たりのDNAがゲノムの他の場所よりもわずかに多いことが予想される。詳細には、混合物中の第21染色体由来のDNAが約5%多いことが予想される。配列決定を使用してDNAを測定する場合、独特のセグメント当たりの第21染色体由来の独自にマッピング可能な読み取りが他の染色体由来のものよりも約5%多いことが予想される。ある特定の閾値よりも多い量の特定の染色体由来のDNAの観察を、その染色体に独自にマッピング可能な配列の

数について調整した場合に、異数性を診断するための基礎として使用することができる。異数性を検出するために使用することができる別の方法は、親の状況を考慮に入れることができること以外は上記のものと同様である。

【0157】

どの対立遺伝子を標的とするかを考える際、一部の親の状況が、他よりも情報価値がある可能性がある尤度を考慮に入れることができる。例えば、 $AA \mid BB$ および対称の状況 $BB \mid AA$ では、胎児が母親とは異なる対立遺伝子を保有することが既知であるので、最も情報価値のある状況である。対称性の理由で、 $AA \mid BB$ 状況と $BB \mid AA$ 状況はどちらも $AA \mid BB$ と称することができる。情報価値のある親の状況の別の集合は $AA \mid AB$ および $BB \mid AB$ であり、これは、これらの場合、胎児が、母親が有さない対立遺伝子を保有する見込みが50%であるからである。対称性の理由で、 $AA \mid AB$ 状況と $BB \mid AB$ 状況はどちらも、 $AA \mid AB$ と称することができる。情報価値のある親の状況の第3の集合は $AB \mid AA$ および $AB \mid BB$ であり、これは、これらの場合、胎児が既知の父系対立遺伝子を保有し、その対立遺伝子が母系ゲノムにも存在するからである。対称性の理由で、 $AB \mid AA$ 状況と $AB \mid BB$ 状況は、 $AB \mid AA$ と称することができる。第4の親の状況は $AB \mid AB$ であり、ここでは胎児は未知の対立遺伝子の状態を有し、対立遺伝子の状態がいかなるものでも、それは、母親が同じ対立遺伝子を有するものである。第5の親の状況は $AA \mid AA$ であり、ここでは母親および父親がヘテロ接合性である。

【0158】

ここで開示されている実施形態の異なる実行

標的個体の倍数性状態を決定するための方法が本明細書に開示されている。標的個体は、割球、胚または胎児であってよい。本開示のいくつかの実施形態では、標的個体における1つまたは複数の染色体の倍数性状態を決定するための方法は、本文書に記載のステップのいずれか、およびそれらの組合せを包含し得る：

いくつかの実施形態では、胎児の遺伝子の状態を決定することにおいて使用する遺伝物質の供給源は、母系の血液から単離された胎児有核赤血球などの胎児の細胞であってよい。該方法は、妊娠中の母親由来の血液試料を得るステップを包含し得る。該方法は、視覚的な技法を用いて、色の特定の組合せは有核赤血球と独自に関連づけられ、色の同様の組合せは母系の血液中に存在する任意の他の細胞には関連づけられないというアイデアに基づいて胎児の赤血球を単離するステップを包含し得る。有核赤血球に関連づけられる色の組合せは、染色することによってより区別可能にすることができる核の周りのヘモグロビンの赤色、および、例えば青色に染色することができる核材料の色を含んでよい。母系の血液から細胞を単離し、それをスライドに広げ、次いで、赤色（ヘモグロビン由来）と青色（核材料由来）の両方が認められる点を同定することにより、有核赤血球の場所を同定することが可能となり得る。次いで、これらの有核赤血球を、マイクロマニピュレーターを使用して抽出し、遺伝子型決定および／または配列決定技法を用いて、これらの細胞の遺伝物質の遺伝子型の態様を測定することができる。

【0159】

ある実施形態では、胎児のヘモグロビンの存在下でのみ蛍光を発し、母系のヘモグロビンの存在下では蛍光を発しない色素を用いて有核赤血球を染色し、したがって、有核赤血球が、母親に由来するかまたは胎児に由来するかの多義性を除くことができる。本開示のいくつかの実施形態は、染色または他のやり方で核材料に印をつけることを伴ってよい。本開示のいくつかの実施形態は、胎児の細胞に特異的な抗体を使用して胎児核材料に特異的に印をつけることを伴ってよい。

【0160】

胎児の細胞を母系の血液から単離するため、または胎児DNAを母系の血液から単離するため、または母系遺伝物質の存在下で胎児の遺伝物質の試料を富化するための多くのやり方がある。これらの方法のいくつかがここに列挙されているが、これは網羅的な列挙を意図したものではない。便宜上、一部の適切な技法がここに列挙されている：蛍光で、または別のやり方でタグを付けた抗体、サイズ排除クロマトグラフィー、磁気で、または他

のやり方で標識したアフィニティータグ、後成的な差異、例えば、特定の対立遺伝子における母系の細胞と胎児の細胞の間の示差的なメチル化、密度勾配遠心分離に続くC D 4 5 / 1 4 枯渇およびC D 4 5 / 1 4 陰性細胞からのC D 7 1 陽性選択、異なる重量オスモル濃度を用いた一重または二重のP e r c o l l 勾配またはガラクトース特異的レクチン法。

【0161】

本開示のある実施形態では、標的個体は胎児であり、胎児由来の複数のDNA試料に対して異なる遺伝子型測定を行う。本開示のいくつかの実施形態では、胎児DNA試料は単離された胎児の細胞由来であり、その胎児の細胞は、母系の細胞と混在している可能性がある。本開示のいくつかの実施形態では、胎児DNA試料は浮動性胎児DNA由来であり、その胎児DNAは、浮動性母系DNAと混在している可能性がある。いくつかの実施形態では、胎児DNA試料は、母系DNAと胎児DNAの混合物を含有する母系の血漿または母系の血液から得ることができる。いくつかの実施形態では、胎児DNAは、母系DNAと、99.9 : 0.1 % ~ 99 : 1 % ; 99 : 1 % ~ 90 : 10 % ; 90 : 10 % ~ 80 : 20 % ; 80 : 20 % ~ 70 : 30 % ; 70 : 30 % ~ 50 : 50 % ; 50 : 50 % ~ 10 : 90 % ; または10 : 90 % ~ 1 : 99 % ; 1 : 99 % ~ 0.1 : 99.9 % の範囲の母体：胎児比で混在している可能性がある。

【0162】

いくつかの実施形態では、遺伝子試料を調製し、かつ／または精製することができる。そのような目的を実現するための当技術分野で公知のいくつかの標準の手順がある。いくつかの実施形態では、試料を遠心分離して、種々の層に分離することができる。いくつかの実施形態では、濾過を用いてDNAを単離することができる。いくつかの実施形態では、DNAの調製は、増幅、分離、クロマトグラフィーによる精製、液液分離、単離、優先的な富化、優先的な増幅、標的化増幅または当技術分野で公知であるか、または本明細書に記載されているいくつかの他の技法のいずれかを伴ってよい。

【0163】

いくつかの実施形態では、本開示の方法は、DNAを増幅するステップを包含し得る。DNAの増幅は、少量の遺伝物質を、同様の遺伝子データの集合を含む、より大量の遺伝物質に変換するプロセスであり、これに限定されないが、ポリメラーゼ連鎖反応（PCR）を含めた多種多様な方法によって行うことができる。DNAを増幅する1つの方法は、全ゲノム増幅（WGA）である。WGAに利用可能ないくつかの方法がある：ライゲーション媒介性PCR（LM-PCR）、縮重オリゴヌクレオチドプライマーPCR（DOP-PCR）、および多置換増幅（MDA）。LM-PCRでは、アダプタと称される短いDNA配列をDNAの平滑末端にライゲーションする。これらのアダプタはユニバーサル増幅配列を含有し、これを使用して、PCRによってDNAを増幅する。DOP-PCRでは、同様にユニバーサル増幅配列を含有するランダムプライマーが第1ラウンドのアニーリングおよびPCRにおいて使用されている。次いで、第2ラウンドのPCRを使用して、さらにユニバーサルプライマー配列を用いて配列を増幅する。MDAでは、DNAを複製する高度にプロセッシブかつ非特異的な酵素であり、単一細胞分析のために使用されているphi-29ポリメラーゼを用いる。単一細胞由来の材料の増幅に対する主要な限定は、（1）極度に希釈したDNA濃度または非常に小さな体積の反応混合物を使用する必要性、および（2）全ゲノムにわたってDNAをタンパク質から確実に解離することの難しさである。それにもかかわらず、単一細胞全ゲノム増幅は、何年にもわたる種々の適用のために首尾よく用いられてきた。DNAの試料からDNAを増幅する他の方法がある。DNA増幅では、最初のDNAの試料を、配列の集合が同様であるが、はるかに量が多いDNAの試料に変換する。いくつかの場合には、増幅は必要ない可能性がある。

【0164】

いくつかの実施形態では、DNAを、ユニバーサル増幅、例えば、WGAまたはMDAを用いて増幅することができる。いくつかの実施形態では、DNAを、標的化増幅、例えば、標的化PCRまたは環状化プローブを用いることによって増幅することができる。い

くつかの実施形態では、DNAを、標的化増幅方法または所望のDNAと望ましくないDNAの完全なまたは部分的な分離をもたらす方法、例えば、ハイブリダイゼーション手法による捕捉を用いて、優先的に富化することができる。いくつかの実施形態では、DNAを、ユニバーサル増幅方法と優先的な富化方法の組合せを用いることによって増幅することができる。これらの方法のいくつかについてのより充実した記載は本文書の他の箇所に見いだすことができる。

【0165】

標的個体および／または関連する個体の遺伝子データは、これらに限定されないが、遺伝子型決定マイクロアレイ、およびハイスループット配列決定を含めた群から選択されるツールおよび、または技法を用いて適切な遺伝物質を測定することによって、分子の状態から電子の状態に変換することができる。いくつかのハイスループット配列決定方法としては、サンガーDNA配列決定、パイロシークエンシング、ILLUMINA SOLEXAプラットフォーム、ILLUMINAのGENOME ANALYZERまたはAPPLIED BIOSYSTEMの454配列決定プラットフォーム、HELICOSのTRUE SINGLE MOLECULE SEQUENCINGプラットフォーム、HALCYON MOLECULARの電子顕微鏡配列決定法または任意の他の配列決定法が挙げられる。これらの方法は全て、DNAの試料に保存されている遺伝子データを、一般には、途中でメモリデバイスに保存されて加工される遺伝子データの集合に物理的に変換する。

【0166】

関連性のある個体の遺伝子データは、これらに限定されないが、個体のバルク二倍体組織、個体由来の1つまたは複数の二倍体細胞、個体由来の1つまたは複数の一倍体細胞、標的個体由来の1つまたは複数の割球、個体において見いだされる細胞外遺伝物質、母系の血液中に見いだされる個体由来の細胞外遺伝物質、母系の血液中に見いだされる個体由来の細胞、関連する個体由来の配偶子（複数可）から作製される1つまたは複数の胚、そのような胚から取得した1つまたは複数の割球、関連する個体において見いだされる細胞外遺伝物質、関連する個体を起源とすることが既知である遺伝物質、およびそれらの組合せを含めた群から選択される物質を分析することによって測定することができる。

【0167】

いくつかの実施形態では、標的個体の対象の染色体型のそれぞれについて、少なくとも1つの倍数性状態仮説の集合を作製することができる。倍数性状態仮説はそれぞれ、標的個体の染色体または染色体セグメントの1つの可能性のある倍数性状態を指し得る。仮説の集合は、標的個体の染色体が有すると予測することができる、可能性のある倍数性状態の一部または全部を含んでよい。可能性のある倍数性状態のいくつかは、零染色体性、モノソミー、ダイソミー、片親性ダイソミー、正倍数性、トリソミー、一致トリソミー、不一致トリソミー、母系トリソミー、父系トリソミー、テトラソミー、平衡（2：2）テトラソミー、不平衡（3：1）テトラソミー、ペンタソミー、ヘキサソミー、他の異数性、およびそれらの組合せを含んでよい。これらの異数性状態はいずれも、混在してよい、または、部分的な異数性、例えば、不平衡転座、平衡転座、ロバートソン転座、組換え、欠失、挿入、乗換え、およびそれらの組合せであってよい。

【0168】

いくつかの実施形態では、決定された倍数性状態の知見を使用して、臨床的決定を行うことができる。この知見は、一般には、事項の物理的配列としてメモリデバイスに保存され、次いで、報告に変換することができる。次いで、報告は実行され得る。例えば、臨床的決定は、妊娠中絶することであってよい、あるいは、臨床的決定は、妊娠を継続することであってよい。いくつかの実施形態では、臨床的決定は、遺伝的障害の表現型の発現の重症度を低下させるために設計された介入、または特別支援児（special needs child）に対する準備をするための関連性のあるステップを取る決定を伴ってよい。

【0169】

本開示のある実施形態では、本明細書に記載の任意の方法は、複数の標的が、同じ標的個体、例えば、同じ妊娠中の母親からの複数の採血に由来することを可能にするために改変することができる。これにより、複数の遺伝子測定によって標的遺伝子型を決定することができるより多くのデータがもたらされ得るので、モデルの正確度を改善することができる。ある実施形態では、1つの標的遺伝子データの集合は、報告された一次データとしての機能を果たし、他の標的遺伝子データの集合は、一次標的遺伝子データを再確認するためのデータとしての機能を果たす。ある実施形態では、標的個体から取得した遺伝物質からそれぞれ測定された複数の遺伝子データの集合を並行して考慮し、したがって、両方の標的遺伝子データの集合は、高い正確度で測定された親の遺伝子データのどのセクションが胎児のゲノムを構成するかを決定するための助けとして機能する。

10

【0170】

ある実施形態では、該方法を、父子試験のために使用することができる。例えば、母親から、および遺伝学的父親である、またはそうでない可能性がある男性からのSNPに基づく遺伝子型の情報ならびに混合試料から測定された遺伝子型の情報を考慮すると、その男性の遺伝子型の情報が実際に妊娠中の胎児の実際の遺伝学的父親を表しているかどうかを決定することが可能である。これを行うための単純なやり方は、単に、母親がAAであり、可能性のある父親がABまたはBBである状況について検査することである。これらの場合、それぞれ、父親が2分の1回(AA | AB)または常に(AA | BB)寄与することを予想することができる。予測ADOを考慮に入れると、観察される胎児のSNPが、可能性のある父親のSNPと相関するかどうかを決定することは簡単である。

20

【0171】

本開示の一実施形態は以下の通りであってよい：妊娠中の女性が、自身の胎児がダウン症候群を患っているかどうか、および／または嚢胞性線維症を患っているかどうか知ることを望んでおり、その女性はこれらの状態のいずれかを患っている子を産むことを望んでいない。医師はその女性の血液を採り、ヘモグロビンを1つのマーカーではっきり赤色であらわれるように染色し、核材料を別のマーカーではっきり青色であらわれるように染色する。母系の赤血球は、一般には無核であるが、高い割合の胎児の細胞が核を含有することが公知であるので、医師は、赤色および青色の両方を示す細胞を同定することにより、いくつかの有核赤血球を視覚的に単離することができる。医師は、これらの細胞を、マイクロマニピュレーターでスライドから取り出し、検査室に送り、そこで10個の個々の細胞を増幅し、遺伝子型決定する。遺伝子測定を使用することによって、PARENTAL

30

SUPPORT (商標) 法で、細胞10個のうち6個が母系の血液細胞であり、細胞10個のうち4個が胎児の細胞であることを決定することができる。妊娠中の母親に既に子が生まれている場合、PARENTAL SUPPORT (商標) は、胎児の細胞に対して信頼できる対立遺伝子呼び出しを行い、それらが生まれた子の対立遺伝子と同様でないことを示すことによって、胎児の細胞が生まれた子の細胞と別個のものであることを決定するためにも使用することができる。この方法は、本開示の父系検査実施形態と同様の概念であることに留意されたい。胎児の細胞から測定された遺伝子データは質が非常に悪い可能性があり、単一細胞の遺伝子型決定の難しさに起因して、多くの対立遺伝子ドロップアウトを含む。臨床医は、測定された胎児DNAを親の信頼できるDNA測定値と一緒に用い、PARENTAL SUPPORT (商標) を使用して胎児のゲノムの態様を高い正確度で推定し、それにより、胎児由来の遺伝物質に含有される遺伝子データを、コンピュータ上に保存される、予測される胎児の遺伝子の状態に変換することができる。臨床医は、胎児の倍数性状態と、複数の疾患に関連づけられる対象の遺伝子の存在または不在の両方とを決定することができる。胎児は正倍数性であり、嚢胞性線維症の保有者ではないことが分かり、母親は妊娠を継続することを決定する。

40

【0172】

本開示のある実施形態では、妊娠中の母親は、自身の胎児がいずれかの全染色体異常を患っているかどうかを決定することを望んでいる。その女性は担当医師の所に行き、自身の血液の試料を提供し、また、その女性とその女性の夫は、頬スワブにより自身のDNA

50

の試料を提供する。検査室の研究者は、親のDNAを増幅するためのMDAプロトコールを使用し、多数のSNPにおける親の遺伝子データを測定するためのILLUMINA INFINIUMアレイを使用して、親のDNAの遺伝子型決定を行う。次いで、研究者は血液を遠心沈澱し、血漿を採り、サイズ排除クロマトグラフィーを使用して浮動性DNAの試料を単離する。あるいは、研究者は、1つまたは複数の蛍光抗体、例えば、胎児のヘモグロビンに特異的な抗体を使用して、胎児有核赤血球を単離する。次いで、研究者は、単離または富化された胎児の遺伝物質を取得し、それを、各オリゴヌクレオチドの2つの末端が標的対立遺伝子のいずれかの側の隣接配列に対応するように適切に設計された70-マー（mer）のオリゴヌクレオチドのライブラリーを使用して増幅する。ポリメラーゼ、リガーゼ、および適切な試薬を添加すると、オリゴヌクレオチドはギャップ充填環状化、所望の対立遺伝子の捕捉を受けた。エキソヌクレアーゼを加え、熱失活させ、産物を直接PCR増幅の鋳型として使用した。PCR産物について、ILLUMINA GENOME ANALYZERで配列決定した。配列読み取りをPARENTAL SUPPORT（商標）法のための入力として使用し、次いで、それにより胎児の倍数性状態を予測した。

【0173】

別の実施形態では、母親が妊娠中であり、高齢出産である夫婦が、妊娠中の胎児がダウン症候群、ターナー症候群、プラダーウィリー症候群またはいくつかの他の全染色体異常を有するかどうかを知りたいことを望んでいる。産科医は、母親および父親から採血を行う。血液を検査室に送り、そこで、技師が母体試料を遠心分離して血漿および Buffy Coat を単離する。Buffy Coat 内のDNAおよび父系の血液試料を増幅によって変換し、増幅された遺伝物質にコードされる遺伝子データを、遺伝物質をハイスループットシーケンサーにかけることによって、分子的に保存された遺伝子データから電子的に保存された遺伝子データにさらに変換して、親の遺伝子型を測定する。血漿試料を、5,000プレックスヘミネステッド標的化PCR法を用いて、遺伝子座の集合において優先的に富化する。DNA断片の混合物を、配列決定に適したDNAライブラリーに調製する。次いで、DNAを、ハイスループット配列決定方法、例えば、ILLUMINA GAIIx GENOME ANALYZERを用いて配列決定する。配列決定により、DNA内に分子的にコードされている情報をコンピュータハードウェアに電子的にコードされる情報に変換する。ここで開示されている実施形態を含むインフォマティクスに基づく技法、例えば、PARENTAL SUPPORT（商標）を使用して、胎児の倍数性状態を決定することができる。これは、調製された試料に対して行ったDNA測定から、複数の多型遺伝子座における対立遺伝子数の確率をコンピュータで算出するステップと、それぞれが、染色体における可能性のある異なる倍数性状態に関連する、複数の倍数性仮説をコンピュータで作製するステップと、各倍数性仮説について、染色体上の複数の多型遺伝子座における予測される対立遺伝子数についての同時分布モデルをコンピュータで構築するステップと、同時分布モデルおよび調製された試料において測定された対立遺伝子数を用いて、倍数性仮説のそれぞれの相対的確率をコンピュータで決定するステップと、最大的確率を有する仮説に対応する倍数性状態を選択することによって胎児の倍数性状態を呼び出すステップとを包含し得る。胎児がダウン症候群を有することが決定される。報告を印刷する、または妊娠中の女性の担当産科医に電子的に送信し、その産科医が診断をその女性に伝達する。その女性、その女性の夫、および医師は腰を据えて選択肢を議論する。夫婦は、胎児がトリソミーの状態を患っているという知見に基づいて、妊娠中絶することを決定する。

【0174】

ある実施形態では、企業が、母系の採血から妊娠中の胎児における異数性を検出するために設計された診断技術を提供することを決定し得る。その産物により、母親が、該母親の血液を採取することができる担当産科医に来診することを必要とし得る。産科医は、胎児の父親からも遺伝子試料を収集することができる。臨床医は、母親の血液から血漿を単離し、血漿からDNAを精製することができる。臨床医は、母親の血液から Buffy Coat 層を単離し、Buffy Coat からDNAを調製することもできる。臨床医は、父親の遺

伝子試料からDNAを調製することもできる。臨床医は、本開示に記載の分子生物学技法を用いて、血漿試料に由来するDNAにおいてDNAにユニバーサル増幅タグを付加することができる。臨床医は、ユニバーサルタグが付けられたDNAを増幅することができる。臨床医は、ハイブリダイゼーションによる捕捉および標的化PCRを含めたいくつもの技法によってDNAを優先的に富化することができる。標的化PCRは、ネスティング、ヘミネスティングまたはセミネスティングまたは血漿由来DNAの効率的な富化をもたらす任意の他の手法を伴ってよい。標的化PCRにより、例えば、1回の反応で10,000個のプライマーを用いて大規模に多重化することができ、ここで、プライマーは、第13染色体、第18染色体、第21染色体、X染色体上のSNPおよびXおよびYの両方に共通し、必要に応じて、他の染色体にも共通する遺伝子座を標的とする。選択的な富化および／または増幅は、個々の分子それぞれに、異なるタグ、分子バーコード、増幅用タグ、および／または配列決定用タグを用いてタグ付けすることを伴ってよい。次いで、臨床医は血漿試料について配列決定し、また、場合によっては、調製された母系DNAおよび／または父系DNAを配列決定することができる。分子生物学的ステップを、診断ボックスによって完全にまたは部分的に実行することができる。配列データを、単一のコンピュータに、または別の種類の計算プラットフォーム、例えば、「クラウド」において見出すことができるものに供給することができる。計算プラットフォームにより、シーケンサーによって行われた測定から標的の多型遺伝子座における対立遺伝子数を算出することができる。計算プラットフォームにより、第13染色体、第18染色体、第21染色体、X染色体およびY染色体のそれぞれについての零染色体性、モノソミー、ダイソミー、一致トリソミー、および不一致トリソミーに関係する、複数の倍数性仮説を作製することができる。計算プラットフォームにより、調べられている5つの染色体のそれぞれに対して、各倍数性仮説に対して、染色体上の標的の遺伝子座における予測される対立遺伝子数についての同時分布モデルを構築することができる。計算プラットフォームにより、同時分布モデルおよび血漿試料に由来する優先的に富化されたDNAに対して測定された対立遺伝子数を用いて、倍数性仮説のそれぞれが真である確率を決定することができる。計算プラットフォームにより、第13染色体、第18染色体、第21染色体、X染色体およびY染色体のそれぞれについて、最大の確率を有する適切な仮説に対応する倍数性状態を選択することによって胎児の倍数性状態を呼び出すことができる。呼び出された倍数性状態を含む報告を作製することができ、それを産科医に電子的に送ること、出力デバイスに表示すること、または印刷した報告のハードコピーを産科医に送達することができる。産科医は、患者、および必要に応じて胎児の父親に知らせることができ、彼らは、どの臨床的な選択肢を受け入れられるか、およびどれが最も望ましいかを決定することができる。

【0175】

別の実施形態では、以後「母親」と称される妊娠中の女性は、自身の胎児（複数可）がいずれかの遺伝子異常または他の状態を保有するか否か知りたいと決めることができる。母親は、いかなる全体的異常もないことを確実にしてから妊娠の継続を確信することを希望することができる。母親は、担当産科医のもとに行くことができ、担当産科医は、母親の血液の試料も採ることができる。担当産科医は、母親の頬からの頬スワブなどの遺伝子試料を採ることもできる。担当産科医は、胎児の父親からも遺伝子試料、例えば、頬スワブ、精子試料または血液試料を採ることができる。担当産科医は、試料を臨床医に送ることができる。臨床医は、母系の血液試料中の浮動性胎児DNAの画分を富化することができる。臨床医は、母系の血液試料中の脱核胎児血液細胞の画分を富化することができる。臨床医は、本明細書に記載の方法の種々の態様を用いて、胎児の遺伝子データを決定することができる。その遺伝子データは、胎児の倍数性状態、および／または胎児における1つまたはいくつもの疾患に関連づけられる対立遺伝子の同一性を含み得る。出生前診断の結果が要約されている報告を作製することができる。報告は、医師に送達または郵送することができ、医師は、母親に胎児の遺伝子の状態を告げることができる。母親は、胎児が1つまたは複数の染色体もしくは遺伝子の異常または望ましくない状態を有するという事実に基づいて、妊娠を中止することを決定することができる。母親は、同様に、胎児が、

10

20

30

40

50

いかなる染色体全体もしくは遺伝子の異常またはいかなる対象の遺伝子の状態も有さないという事実に基づいて、妊娠を継続することを決定することができる。

【0176】

別の例は、精子ドナーにより人工受精し、妊娠中である妊娠中の女性に関し得る。その女性は、自身が保有している胎児が遺伝病を有するリスクを最小限にすることを希望している。その女性は、静脈滲血土による採血を受け、本開示に記載の技法を用いて、3つの胎児有核赤血球を単離し、組織試料も、母親および遺伝学的父親から採取する。胎児由来の遺伝物質ならびに母親および父親由来の遺伝物質を、必要に応じて増幅し、ILLUMINA INFINIUM BEADARRAYを用いて遺伝子型決定し、本明細書に記載の方法により親の遺伝子型および胎児の遺伝子型をきれいにし、高い正確度で相を特定し、ならびに、胎児についての倍数性呼び出しを行う。胎児が正倍数性であることが見いだされ、再構築された胎児の遺伝子型から表現型による罹病性を予測し、報告を作製し、母親の担当医師に送り、したがって、彼らはどんな臨床的決定が最良であり得るかを決定することができる。

【0177】

ある実施形態では、母親および父親の未処理の遺伝物質を、増幅によって、ある量の、配列は同様であるが量がより多いDNAに変換する。次いで、遺伝子型決定方法により、核酸によりコードされる遺伝子型データを、上記のものなどのメモリデバイスに物理的かつ／または電子的に保存することができる遺伝子測定値に変換する。PARENTAL SUPPORT (商標) アルゴリズムを構成する関連性のあるアルゴリズムを、プログラミング言語を用いてコンピュータプログラムに翻訳するが、そのアルゴリズムの関連性のある部分は本明細書において詳細に考察されている。次いで、物理的にコードされるビットおよびバイトではなく、生の測定データを示すパターンに整理されているコンピュータハードウェアでコンピュータプログラムを実行することにより、胎児の倍数性状態の高い信頼度の決定を示すパターンに変換される。この変換の詳細は、本明細書に記載の方法を実行するために使用するデータ自体およびコンピュータ言語およびハードウェアシステムに依拠する。次いで、高い質の胎児の倍数性の決定を示すように物理的に構成されたデータを、健康管理実践者に送ることができる報告に変換する。この変換は、プリンタまたはコンピュータディスプレイを使用して行うことができる。報告は、紙または他の適切な媒体に印刷されたコピーであってよい、あるいは、報告は、電子的なものであってよい。電子報告の場合は、伝達することができ、健康管理実践者が利用できるコンピュータに位置するメモリデバイス上に物理的に保存することができる。電子報告は、読み取ることができるようにスクリーン上に表示することもできる。スクリーン表示の場合には、データは、表示デバイス上でピクセルの物理的変換を引き起こすことによって可読の形式に変換することができる。変換は、リン光性スクリーンに電子を物理的に発射することによって、光子を放出または吸収する基材の前に置くことができるスクリーン上のピクセルの特定の集合の透明度を物理的に変化させる電気的な電荷を変更することによって実現することができる。この変換は、ピクセルの特定の集合において、液晶中のナノスケールの分子の配向を、例えば、ネマティック相からコレステリック相またはスメクチックな相に変化させることによって実現することができる。この変換は、意味のあるパターンに配置された複数の発光ダイオードで構成されたピクセルの特定の集合からの光子の放出を引き起こす電流によって実現することができる。この変換は、情報を表示するために使用される任意の他の方法、例えば、コンピュータスクリーンまたはいくつかの他の出力デバイスまたは情報伝達法によって実現することができる。次いで、健康管理実践者は、報告にあるデータを措置に変換するように、報告に基づいて行動することができる。措置は、妊娠を継続または中止することであってよく、その場合、遺伝子の異常を有する妊娠中の胎児は、非生存胎児に変換される。本明細書において列挙されている変換は、例えば、妊娠中の母親および父親の遺伝物質を、本開示において概説されているいくつかのステップを通じて、遺伝子の異常を有する胎児を流産することからなる、または妊娠を継続することからなる医学的な決定に変換することができるように総計することができる。あるいは、遺伝子型の

10

20

30

40

50

測定値の集合を、医師が妊娠中の患者を処置することに役立つ報告に変換することができる。

【0178】

本開示のある実施形態では、本明細書に記載の方法を用いて、宿主母親、すなわち妊娠中の女性が、自身が保有している胎児の生物学的母親ではない場合にでも、胎児の倍数性状態を決定することができる。本開示のある実施形態では、本明細書に記載の方法を用いて、母系の血液試料のみを使用し、父系の遺伝子試料を必要とせずに胎児の倍数性状態を決定することができる。

【0179】

ここで開示されている実施形態における数学のいくつかにより、限られた数の異数性の状態に関する仮説が立てられる。いくつかの場合には、例えば、0、1つまたは2つの染色体のみが、各親を起源とすることが予測される。本開示のいくつかの実施形態では、本開示の基本的な概念を変化させることなく、数学的な導出を拡大して異数性の他の形態、例えば、3つの染色体が一方の親を起源とするクアドロソミー (quadrosomy)、ペンタソミー、ヘキサソミーを考慮に入れることができる。同時に、より小さな数の倍数性状態、例えば、トリソミーおよびダイソミーのみに焦点を当てることが可能である。整数でない染色体を示す倍数性の決定は、遺伝物質の試料中のモザイク現象を示し得ることに留意されたい。

【0180】

いくつかの実施形態では、遺伝子の異常は、ダウン症候群 (または21トリソミー)、エドワーズ症候群 (18トリソミー)、パトー症候群 (13トリソミー)、ターナー症候群 (45X)、クラインフェルター症候群 (2つのX染色体を持つ男性)、プラダーウィリー症候群、およびディジョージ症候群 (UPD15) などの異数性の一種である。前文に列挙されているものなどの先天性障害は、一般に望ましくなく、胎児が1つまたは複数の表現型の異常を患っているという知見により、妊娠中絶すること、特別支援児の誕生のための準備をするために必要な対策をとること、または染色体異常の重症度を減らすことを意図したいくつかの治療的手法をとることを決定するための基礎を提供することができる。

【0181】

いくつかの実施形態では、本明細書に記載の方法を、非常に初期の妊娠期間、例えば、早ければ4週、早ければ5週、早ければ6週、早ければ7週、早ければ8週、早ければ9週、早ければ10週、早ければ11週、および早ければ12週において用いることができる。

【0182】

宿主において生存しているがんを起源とするDNAを、宿主の血液中找到することができることが実証されていることに留意されたい。母系の血液中找到される混合DNAを測定することによって遺伝子診断を行うことができるのと同様に、宿主血液中找到される混合DNAを測定することによって同等に良好に遺伝子診断を行うことができる。遺伝子診断は、異数性状態または遺伝子変異を含み得る。母系の血液に対して行った測定からの胎児の倍数性状態または遺伝子の状態を決定することにおいて読み取る当該開示における任意の主張は、宿主血液に対する測定からがんの倍数性状態または遺伝子の状態を決定することにおいて、同等に良好に読み取ることができる。

【0183】

いくつかの実施形態では、本開示の方法により、がんの倍数性状態を決定することが可能になり、該方法は、宿主由来の遺伝物質を含有する混合試料、およびがん由来の遺伝物質を得るステップと、混合試料中のDNAを測定するステップと、混合試料中のがん起源のDNAの割合を算出するステップと、混合試料に対して得た測定値および算出された割合を用いてがんの倍数性状態を決定するステップとを含む。いくつかの実施形態では、該方法は、がんの倍数性状態の決定に基づいてがん治療を施すステップをさらに含んでよい。いくつかの実施形態では、該方法は、がんの倍数性状態の決定に基づいてがん治療を施

すステップをさらに含んでよく、がん治療は、医薬品、生物学的治療薬、および抗体に基づく治療およびそれらの組合せを含む群から選択される。

【0184】

いくつかの実施形態では、本明細書に開示されている方法を、*in vitro*での受精の間に胚を選択するための着床前遺伝子診断（PGD）との関連において使用し、ここで、標的個体は胚であり、3日目の胚由来の単一細胞または2つの細胞の生検または5日目または6日目の胚の栄養外胚葉生検材料からの配列決定データから、胚に関する倍数性の決定を行うために、親の遺伝子型データを使用することができる。PGDの環境では、子のDNAのみを測定し、ほんの少数の細胞、一般に、1～5個であるが、多くて10個、20個または50個を検査する。次いで、対立遺伝子AおよびB（SNPにおける）の開始時のコピーの総数が子の遺伝子型および細胞の数によって自明に決定される。NPDでは、開始時のコピーの数は非常に多く、したがって、PCRの後の対立遺伝子の比が開始時の比を正確に反映することが予想される。しかし、PGDにおける開始時のコピーが少数であることは、コンタミネーションおよび不完全なPCR効率が、PCR後の対立遺伝子の比に対する非自明の効果を有することを意味する。この効果は、配列決定後に測定された対立遺伝子の比における分散を予測することにおける読み取りの深さよりも重要であり得る。既知の子の遺伝子型を考慮して測定された対立遺伝子の比の分布は、PCRプローブの効率およびコンタミネーションの確率に基づいたPCRプロセスのMonte Carloシミュレーションによって作製することができる。可能性のある子の遺伝子型のそれぞれについて対立遺伝子の比の分布を考慮して、種々の仮説の尤度を、NIPDについて記載されているのと同様に算出することができる。

【0185】

本明細書に開示されている実施形態はいずれも、デジタル電子回路、集積回路、特別に設計されたASIC（特定用途向け集積回路）、コンピュータハードウェア、ファームウェア、ソフトウェアにおいて、またはそれらの組合せにおいて実行することができる。ここで開示されている実施形態の装置は、プログラム可能なプロセッサによって実行するための機械可読記憶デバイスに実体的に具体化されたコンピュータプログラム産物において実行することができ、ここで開示されている実施形態の方法のステップは、入力データを操作し、出力を生成することによってここで開示されている実施形態の機能を実施するための命令のプログラムを実行するプログラム可能なプロセッサによって実施することができる。ここで開示されている実施形態は、少なくとも1つのプログラム可能なプロセッサを含むプログラム可能なシステムにおいて実行可能かつ／または解釈可能な1つまたは複数のコンピュータプログラムで有利に実行することができ、該少なくとも1つのプログラム可能なプロセッサは、特別または汎用であり得る、データおよび命令を受け、データおよび命令を伝達するための記憶システム、少なくとも1つの入力デバイス、および少なくとも1つの出力デバイスと連結している。各コンピュータプログラムは、所望であれば、高レベルの手続き型のまたはオブジェクト指向のプログラミング言語で、またはアセンブリ言語または機械語で実装することができ、どんな場合でも、言語はコンパイラ型言語またはインタープリタ型言語であってよい。コンピュータプログラムは、独立型プログラムとして、またはモジュール、コンポーネント、サブルーチンまたはコンピュータ環境において使用するために適した他のユニットとしてのものを含めた、任意の形態で展開することができる。コンピュータプログラムは、1か所の、または複数か所にわたって分布した、および通信網により相互接続された1台のコンピュータまたは複数のコンピュータで実行または解釈されるように展開することができる。

【0186】

コンピュータ可読の記憶媒体とは、本明細書で使用される場合、物理的なまたは有形の記憶装置（シグナルとは対照的に）を指し、これらに限定することなく、情報の有形の記憶装置、例えば、コンピュータ可読の命令、データ構造、プログラムモジュールまたは他のデータの任意の方法または技術において実装される揮発性のおよび不揮発性の、取り外し可能なおよび取り外し不可能な媒体を指す。コンピュータ可読の記憶媒体としては、こ

れらに限定されないが、所望の情報またはデータまたは命令を実体的に保存するために使用することができ、コンピュータまたはプロセッサがアクセスすることができるRAM、ROM、EPROM、EEPROM、フラッシュメモリまたは他のソリッドステート記憶技術、CD-ROM、DVDまたは他の光学記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置または他の磁気記憶デバイスまたは任意の他の物理的なまたは有形の媒体が挙げられる。

【0187】

本明細書に記載の方法のいずれかは、コンピュータスクリーンまたは紙への印刷などの物理的形式でデータを出力するステップを含んでよい。本文書の他の箇所の任意の実施形態の記載では、記載されている方法を、医師がそれに基づいて行動することができる形式のすぐに使用可能なデータを出力するステップと組み合わせることができることが理解されるべきである。さらに、記載されている方法を、臨床処置、または措置を行わないという臨床的決定の実行をもたらす臨床的決定の実際の実行と組み合わせることができる。標的個体に関係する遺伝子データを決定するための本文書に記載の実施形態のいくつかを、IVFとの関連において、1つまたは複数の胚を移入するために選択する決定と組み合わせ、必要に応じて、胚を将来の母親の子宮に移入するプロセスと組み合わせることができる。標的個体に関係する遺伝子データを決定するための本文書に記載の実施形態のいくつかを、潜在的な染色体異常またはそれが無いことの医療専門家による通知と組み合わせ、必要に応じて、出生前診断との関連において、胎児を流産するか流産しないかの決定と組み合わせることができる。本明細書に記載の実施形態のいくつかを、すぐに使用可能なデータを出力すること、および臨床処置、または措置を行わないという臨床的決定の実行をもたらす臨床的決定を実行することと組み合わせることができる。

【0188】

標的化富化および配列決定

非侵襲的な出生前対立遺伝子呼び出しまたは倍数性呼び出しのための方法の一部としてDNAの試料を標的遺伝子座の集合において富化し、その後、配列決定する技法を使用することにより、いくつかの予想外の利点が付与され得る。本開示のいくつかの実施形態では、該方法は、インフォマティクスに基づく方法、例えば、PARENTAL SUPPORT（商標）（PS）で使用するための遺伝子データを測定するステップを包含する。実施形態のいくつかの最終的転帰は、胚または胎児のすぐに使用可能な遺伝子データである。具体化された方法の一部として、個体および／または関連する個体の遺伝子データを測定するために用いることができる多くの方法が存在する。ある実施形態では、標的の対立遺伝子の集合の濃度を富化するための方法が本明細書に開示されており、該方法は、以下のステップの1つまたは複数を含む：遺伝物質を標的化増幅するステップ、遺伝子座に特異的なオリゴヌクレオチドプローブを添加するステップ、特定のDNA鎖をライゲーションするステップ、所望のDNAの集合を単離するステップ、反応の望ましくない構成成分を除去するステップ、ハイブリダイゼーションによって特定のDNAの配列を検出するステップ、およびDNAの配列決定方法によって1つまたは複数のDNA鎖の配列を検出するステップ。いくつかの場合には、DNA鎖とは標的遺伝物質を指し得、いくつかの場合には、DNA鎖とはプライマーを指し得、いくつかの場合には、DNA鎖とは合成された配列、またはそれらの組み合わせを指し得る。これらのステップは、いくつかの異なる順序で行うことができる。分子生物学の高度な可変性を考慮すると、どの方法、およびどのステップの組み合わせが、種々の状況において上手く実施されないか、上手く実施されるか、または最も上手く実施されるかは通常明白ではない。

【0189】

例えば、標的化増幅の前のDNAのユニバーサル増幅ステップにより、いくつかの有利な点、例えば、ボトルネックのリスクの除去および対立遺伝子の偏りの低減が付与され得る。DNAを、標的配列の両側の2つの隣接する領域とハイブリダイズすることができるオリゴヌクレオチドプローブと混合することができる。ハイブリダイゼーション後、プローブの末端を、ライゲーションの手段であるポリメラーゼ、およびプローブの環状化を可

10

いくつかの実施形態では、標的遺伝物質の検出は多重様式で行うことができる。並行して実行することができる遺伝子の標的配列の数は、1~10、10~100、100~1,000、1,000~10,000、10,000~100,000、100,000~1,000,000、1,000,000~10,000,000または10,000,000~100,000,000にわたり得る。先行技術は、最大で約50または100個以下のプライマーのプールを伴う上首尾の多重PCR反応の開示を含むことに留意されたい。プール当たり100個超のプライマーを多重化するための先の試みにより、プライマーの二量体形成などの望ましくない副反応を伴う重大な問題が生じた。

20

【 0 1 9 2 】

30

40

ローブを添加するステップ、特定のDNA鎖をライゲーションするステップ、所望のDNAの集合を単離するステップ、反応の望ましくない構成成分を除去するステップ、ハイブリダイゼーションによって特定のDNAの配列を検出するステップ、DNAの配列決定方法によって1つまたは複数のDNA鎖の配列を検出するステップを伴う。いくつかの場合には、DNA鎖とは、標的遺伝物質を指し得、いくつかの場合には、DNA鎖とは、プライマーを指し得、いくつかの場合には、DNA鎖とは、合成された配列、またはそれらの組み合わせを指し得る。これらのステップは、いくつかの異なる順序で行うことができる。分子生物学の高度な可変性を考慮すると、どの方法、およびどのステップの組み合わせが、種々の状況において上手く実施されないか、上手く実施されるか、または最も上手く実施されるかは通常明白ではない。

10

【0194】

理論上はゲノム内の任意の数の遺伝子座、1個の遺伝子座から100万超までのいずれかの遺伝子座を標的とすることが可能であることに留意されたい。DNAの試料を標的化、次いで配列決定に供する場合、シーケンサーによって読み取られる対立遺伝子の割合は、それらが試料中に天然に存在する量に対して富化される。富化の程度は、1パーセント（またはさらに低い）から10倍、100倍、1,000倍またはさらに多く100万倍までのいずれであってもよい。ヒトゲノムには、およそ7,500万の多型遺伝子座を含む、およそ30億の塩基対、およびヌクレオチドが存在する。標的とされる遺伝子座が多いほど、より少ない富化の程度が可能である。標的とされる遺伝子座の数が少ないほど、より大きな富化の程度が可能であり、それらの遺伝子座において、所与の数の配列読み取りに対してより大きな読み取りの深さを実現することができる。

20

【0195】

本開示のある実施形態では、標的化または優先（preferential）は、完全にSNPに焦点を当てることができる。ある実施形態では、標的化または優先は、任意の多型部位に焦点を当てることができる。エクソンを富化するためのいくつかの商業的な標的化産物が利用可能である。驚いたことに、排他的にSNPを、または排他的に多型遺伝子座を標的化することは、対立遺伝子分布に依拠するNPDのための方法を用いる場合に特に有利である。配列決定を用いるNPDのための公開された方法も存在し、例えば、米国特許第7,888,017号は、読み取りの計数が、所与の染色体にマッピングされる読み取りの数をカウントすることに焦点が当てられる読み取り数解析を伴い、分析された配列読み取りは、多型のゲノムの領域には焦点を当てていない。多型対立遺伝子に焦点を当てないこれらの種類の方法体系は、対立遺伝子の集合を標的化または優先的に富化することほど役立たない。

30

【0196】

本開示のある実施形態では、遺伝子試料をゲノムの多型領域において富化するためにSNPに焦点を当てると標的化方法を用いることが可能である。ある実施形態では、少数のSNP、例えば、1から100の間のSNP、またはそれよりも多数、例えば、100から1,000の間、1,000から10,000の間、10,000から100,000の間、または100,000超のSNPに焦点を当てることが可能である。ある実施形態では、生存するトリソミーでの出生と相関する1つまたは少数の染色体、例えば、第13染色体、第18染色体、第21染色体、X染色体およびY染色体またはそのいくつかの組み合わせに焦点を当てることが可能である。ある実施形態では、標的のSNPを小さな係数、例えば、1.01倍から100倍の間、またはそれよりも大きな係数、例えば、100倍から1,000,000倍の間、または1,000,000超倍まで富化することが可能である。本開示のある実施形態では、標的化方法を用いて、ゲノムの多型領域において優先的に富化されたDNAの試料を作製することが可能である。ある実施形態では、この方法を用いて、これらの特性のいずれかを有するDNAの混合物を作製することが可能であり、ここで、DNAの混合物は、母系DNAと、浮動性胎児DNAも含有する。ある実施形態では、この方法を用いて、これらの係数の任意の組み合わせを有するDNAの混合物を作製することが可能である。例えば、本明細書に記載の方法を用いて、母系DNAお

40

50

よび胎児DNAを含み、200 SNPであって、全てが第18染色体または第21染色体のいずれかに位置し、平均で1,000倍に富化される200 SNPに対応するDNAにおいて優先的に富化されたDNAの混合物を生成することができる。別の例では、該方法を用いて、10,000 SNPであって、全てまたはほとんどが第13染色体、第18染色体、第21染色体、X染色体およびY染色体に位置し、遺伝子座当たりの平均の富化が500倍を超える10,000 SNPにおいて優先的に富化されたDNAの混合物を作製することが可能である。本明細書に記載の標的化方法のいずれかを用いて、特定の遺伝子座において優先的に富化されたDNAの混合物を作製することができる。

【0197】

いくつかの実施形態では、本開示の方法は、ハイスループットDNAシーケンサーを使用して混合画分中のDNAを測定するステップであって、混合画分中のDNAが、不相応な数の1つまたは複数の染色体由来の配列を含有し、1つまたは複数の染色体が第13染色体、第18染色体、第21染色体、X染色体、Y染色体およびそれらの組み合わせを含む群から選択されるステップをさらに含む。

【0198】

本明細書には3つの方法、多重PCR、ハイブリダイゼーションによる標的化捕捉、および連結逆方向プローブ(LIP)が記載されており、それを用いて、胎児の異数性を検出するために、母系の血漿試料由来の十分な数の多型遺伝子座から測定値を得、解析することができる。これは、標的の遺伝子座を選択的に富化する他の方法を排除するものではない。該方法の核心を変化させることなく他の方法を同等に良好に用いることができる。それぞれの場合において、アッセイされる多型は、一塩基多型(SNP)、小さな挿入欠失またはSTRを含んでよい。好ましい方法は、SNPの使用を伴う。各手法により、対立遺伝子頻度データが生じ、各標的の遺伝子座についての対立遺伝子頻度データおよび／またはこれらの遺伝子座からの同時対立遺伝子頻度分布を解析して、胎児の倍数性を決定することができる。各手法は、供給源材料が限られていること、および母系の血漿が母系DNAと胎児DNAの混合物からなるという事実起因して、それ自体の考慮すべき事柄を有する。この方法は、より正確な決定をもたらすための他の手法と組み合わせることができる。ある実施形態では、この方法を、米国特許第7,888,017号に記載のものなどの配列計数手法と組み合わせることができる。記載されている手法は、胎児の父系性を非侵襲的に、母系の血漿試料から検出するために用いることもできる。さらに、各手法は、異数性染色体の存在または不在を検出するため、分解されたDNA試料由来の多数のSNPについて遺伝子型決定するため、セグメントに分かれたコピー数の変動(CNV)を検出するため、他の対象の遺伝子型の状態またはそのいくつかの組み合わせを検出するために、他のDNAの混合物または純粋なDNA試料に適用することができる。

【0199】

試料中の対立遺伝子分布の正確な測定

現行の配列決定手法を用いて、試料中の対立遺伝子の分布を推定することができる。そのような方法の1つは、ショットガン配列決定と称される、プールDNAから配列を無作為にサンプリングするステップを包含する。配列決定データにおける特定の対立遺伝子の割合は、一般には、非常に低く、単純統計量によって決定することができる。ヒトゲノムは、およそ30億の塩基対を含有する。したがって、使用した配列決定方法により100bpの読み取りが生じた場合、特定の対立遺伝子は、およそ3,000万回の配列読み取りごとに1回測定される。

【0200】

ある実施形態では、本開示の方法を用いて、DNAの試料中の同じ遺伝子座の集合を含有する2種以上の異なるハプロタイプの存在または不在を、その染色体由来の遺伝子座の測定された対立遺伝子分布から決定する。異なるハプロタイプは、1つの個体由来の2つの異なる相同染色体、トリソミーの個体由来の3つの異なる相同染色体、母親および胎児由来の3つの異なる相同なハプロタイプであって、該ハプロタイプのうちの1つが母親と胎児の間で共有されるハプロタイプ、母親および胎児由来の3つまたは4つのハプロタイ

プであって、該ハプロタイプの1つまたは2つが母親と胎児の間で共有されるハプロタイプ、または他の組み合わせを示し得る。ハプロタイプ間で多型である対立遺伝子はより情報価値がある傾向があるが、母親および父親がどちらも同じ対立遺伝子についてホモ接合性ではない任意の対立遺伝子により、測定された対立遺伝子分布を通じて、単純読み取り数解析から入手可能である情報を越えた有用な情報がもたらされる。

【0201】

しかし、そのような試料のショットガン配列決定は、それにより、試料中の異なるハプロタイプ間で多型ではない領域、または対象ではない染色体についての多くの配列がもたらされ、したがって、標的ハプロタイプの割合に関する情報を示さないの、非常に非効率的である。本明細書には、ゲノム内で多型である可能性がより高い、試料中のDNAのセグメントを特異的に標的とし、かつ／または優先的に富化して、配列決定によって得られる対立遺伝子の情報の収量を上昇させる方法が記載されている。標的個体に存在する実際の量を真に表すことになる富化された試料において測定された対立遺伝子分布について、標的のセグメント内の所与の遺伝子座における他の対立遺伝子と比較して1つの対立遺伝子の優先的な富化がわずかである、または存在しないことが重大であることに留意されたい。多型対立遺伝子を標的とするための現行の当技術分野で公知の方法は、存在する任意の対立遺伝子の少なくとも一部が検出されることが確実になるように設計されている。しかし、これらの方法は、元の混合物に存在する多型対立遺伝子の不偏の対立遺伝子分布を測定する目的では設計されていなかった。標的富化の任意の特定の方法により、富化された試料を生成することができ、測定された対立遺伝子分布が元の増幅されていない試料に存在する対立遺伝子分布を、任意の他の方法よりも良好に正確に示すことは自明ではない。理論上は、多くの富化方法がそのような目的を実現することが予測され得るが、当業者は、現行の増幅、標的化および他の優先的な富化方法には相当量の確率論的または決定論的な偏りがあることをよく理解している。本明細書に記載の方法の一実施形態により、ゲノム内の所与の遺伝子座に対応するDNAの混合物に見いだされる複数の対立遺伝子を、対立遺伝子のそれぞれの富化の程度がほぼ同じになるように増幅または優先的に富化することが可能になる。別の言い方では、該方法により、各遺伝子座に対応する対立遺伝子間の比は元のDNAの混合物における比と基本的に同じままで、混合物に存在する対立遺伝子の相対的な量を全体として増大させることが可能になる。遺伝子座を優先的に富化する先行技術の方法により、1%超、2%超、5%超、さらには10%超の対立遺伝子の偏りがもたらされ得る。この優先的な富化は、ハイブリダイゼーション手法による捕捉を用いた場合の捕捉の偏り、または各サイクルに関しては小さい可能性があるが、20サイクル、30サイクルまたは40サイクルにわたって組み立てる(compounded)と大きくなり得る増幅の偏りに起因し得る。本開示の目的で、比が基本的に同じままであるとは、元の混合物における対立遺伝子の比を、生じた混合物における対立遺伝子の比で割ったものが、0.95から1.05の間、0.98から1.02の間、0.99から1.01の間、0.995から1.005の間、0.998から1.002の間、0.999から1.001の間、または0.9999から1.0001の間であることを意味する。本明細書で提示された対立遺伝子の比の算出は、標的個体の倍数性状態の決定には使用することができず、単に対立遺伝子の偏りを測定するために使用されるメトリックであり得ることに留意されたい。

【0202】

ある実施形態では、混合物が標的遺伝子座の集合において優先的に富化されたら、クローン試料(単一分子から生成される試料;例としては、ILLUMINA GAIIx、ILLUMINA HISEQ、LIFE TECHNOLOGIES SOLID、5500XLが挙げられる)について配列決定する、以前の、現行のまたは次世代の配列決定計器のうちの任意の1つを用いて配列決定することができる。比は、標的の領域内の特定の対立遺伝子を通して配列決定することによって評価することができる。これらの配列決定の読み取りを、対立遺伝子の型、および、したがって決定される異なる対立遺伝子の割り当てに応じて分析し、カウントすることができる。1から数塩基の長さである変動に

10

20

30

40

50

ついて、対立遺伝子の検出は配列決定によって実施し、捕捉された分子の対立遺伝子の組成を評価するために、配列決定の読み取りが問題の対立遺伝子にわたることが必須である。遺伝子型についてアッセイする捕捉された分子の総数は、配列決定の読み取りの長さが増加することによって増加することができる。全ての分子の完全な配列決定により、富化されたプールにおいて利用可能な最大量のデータの収集が保証される。しかし、配列決定は、現在は費用がかかり、少数の配列読み取りを用いて対立遺伝子分布を測定することができる方法は非常に価値がある。さらに、読み取りの長さが増加すると、可能性のある読み取りの最大長に対する技術的な限界ならびに正確度の限界がでてくる。有用性が最大である対立遺伝子は、1～数塩基の長さのものであるが、理論的には、配列決定の読み取りの長さよりも短い任意の対立遺伝子を使用することができる。対立遺伝子の変動は全ての型で生じるが、本明細書において提供される実施例は、ほんの数個隣接する塩基対に含有されるSNPまたは変異体に焦点を当てる。より大きな変異体、例えば、セグメントに分かれたコピー数の変異体は、多くの場合、セグメントの内部のSNPの全体的な集団が重複しているので、これらのより小さな変動を総計することによって検出することができる。数塩基よりも大きな変異体、例えば、STRは、特別な考慮およびいくつかの標的化手法研究を必要とするが、他のものは必要としない。

10

【0203】

ゲノム内の1つまたは複数の変異体の位置を特異的に単離し、富化するために使用することができる複数の標的化手法が存在する。一般には、これらは、変異体配列に隣接している変異していない配列を利用することに依拠する。基材が母系の血漿である、配列決定との関連において標的化に関連する先行技術が存在する（例えば、Liaoら、Clin. Chem. 2011年；57巻（1号）：92～101頁を参照されたい）。しかし、先行技術における手法は全て、エクソンを標的とする標的化プローブを使用し、ゲノムの多型領域を標的とすることには焦点を当てていない。ある実施形態では、本開示の方法は、多型領域に排他的またはほぼ排他的に焦点を当てた標的化プローブを使用するステップを包含する。ある実施形態では、本開示の方法は、SNPに排他的またはほぼ排他的に焦点を当てる標的化プローブを使用するステップを包含する。本開示のいくつかの実施形態では、標的的多型部位は、少なくとも10%のSNP、少なくとも20%のSNP、少なくとも30%のSNP、少なくとも40%のSNP、少なくとも50%のSNP、少なくとも60%のSNP、少なくとも70%のSNP、少なくとも80%のSNP、少なくとも90%のSNP、少なくとも95%のSNP、少なくとも98%のSNP、少なくとも99%のSNP、少なくとも99.9%のSNPまたは排他的にSNPからなる。

20

30

【0204】

ある実施形態では、本開示の方法を用いて、遺伝子型（特定の遺伝子座におけるDNAの塩基組成）およびDNA分子の混合物由来のこれらの遺伝子型の相対的な割合を決定することができる。これらのDNA分子は、1つまたはいくつもの遺伝的に別個の個体を起源とし得る。ある実施形態では、本開示の方法を用いて、多型遺伝子座の集合における遺伝子型、およびこれらの遺伝子座に存在する異なる対立遺伝子の量の相対的な比を決定することができる。ある実施形態では、多型遺伝子座は、完全にSNPからなっておりよい。ある実施形態では、多型遺伝子座は、SNP、単一のタンDEM反復、および他の多型を含んでよい。ある実施形態では、本開示の方法を用いて、DNAの混合物における多型遺伝子座の集合における対立遺伝子の相対的な分布を決定することができ、DNAの混合物は、母親を起源とするDNA、および胎児を起源とするDNAを含む。ある実施形態では、妊娠中の女性由来の血液から単離されたDNAの混合物について同時対立遺伝子分布を決定することができる。ある実施形態では、遺伝子座の集合における対立遺伝子分布を使用して、妊娠中の胎児について1つまたは複数の染色体の倍数性状態を決定することができる。

40

【0205】

ある実施形態では、DNA分子の混合物は、1つの個体の複数の細胞から抽出したDNAに由来しておりよい。ある実施形態では、個体がモザイク（生殖系列または体細胞）である場合、DNAが由来する元の細胞の集団は、同じ遺伝子型または異なる遺伝子型の二倍体

50

細胞または一倍体細胞の混合物を含み得る。ある実施形態では、DNA分子の混合物は、単一細胞から抽出したDNAに由来してもよい。ある実施形態では、DNA分子の混合物は、同じ個体の2つ以上の細胞または異なる個体の2つ以上の細胞の混合物から抽出したDNAに由来してもよい。ある実施形態では、DNA分子の混合物は、無細胞DNAを含有することが公知である血漿などの、既に細胞から遊離した生物材料から単離されたDNAに由来してもよい。ある実施形態では、この生物材料は、胎児DNAが混合物中に存在することが示されている妊娠中の場合と同様に、1つまたは複数の個体由来のDNAの混合物であってよい。ある実施形態では、生物材料は、母系の血液に見いだされた細胞の混合物由来であってよく、細胞のいくつかは胎児を起源とする。ある実施形態では、生物材料は、胎児の細胞において富化された妊娠中の血液由来の細胞であってよい。

10

【0206】

環状化プローブ

本開示のいくつかの実施形態は、文献において以前に記載されている「連結逆方向プローブ」(LIP)を使用することを伴う。LIPとは、環状DNA分子を作製することを伴う技術を包含することを意味する総称であり、プローブは、標的の対立遺伝子の両側の標的のDNAの領域とハイブリダイズするように設計されており、したがって、適切なポリメラーゼおよび／もしくはリガーゼ、および適切な条件、緩衝液および他の試薬の添加により、標的の対立遺伝子をわたるDNAの相補的な逆方向領域が完成し標的の対立遺伝子に見いだされる情報を捕捉するDNAの環状ループを作製される。LIPは、環状化前プローブ(pre-circularized probe)、環状化前プローブ(pre-circularizing probe)または環状化プローブとも称される。LIPプローブは、長さが50ヌクレオチドから500ヌクレオチドの間の直鎖DNA分子であってよく、ある実施形態では、長さが70ヌクレオチドから100ヌクレオチドの間であってよく、いくつかの実施形態では、本明細書に記載されているよりも長くてよい、または短くてよい。本開示の他の複数の実施形態は、LIP技術の異なる具体化、例えば、Padlockプローブおよび分子逆方向プローブ(MIP)を伴う。

20

【0207】

配列決定するために特定の場所を標的とする1つの方法は、プローブの3'末端および5'末端が標的DNAと、標的の領域に近接し、その両側の場所で、逆方向様式でアニーリングし、したがって、DNAポリメラーゼおよびDNAリガーゼを添加することにより、3'末端からの伸長がもたらされ、標的分子と相補的な一本鎖プローブに塩基が付加され(ギャップ充填)、その後、新しい3'末端が元のプローブの5'末端とライゲーションし、その結果、後でバックグラウンドDNAから単離することができる環状DNA分子がもたらされるようなプローブを合成することである。プローブ末端は、対象の標的の領域に隣接するように設計されている。この手法の一態様は、一般に、MIPSと称され、充填される配列の性質を決定するために、アレイ技術と併せて用いられている。対立遺伝子の比を測定する状況においてMIPを用いることの1つの欠点は、ハイブリダイゼーションステップ、環状化ステップおよび増幅ステップが、同じ遺伝子座における異なる対立遺伝子について同等の率で起こらないことである。その結果、元の混合物に存在する実際の対立遺伝子の比を表さない対立遺伝子の比が測定される。

30

40

【0208】

ある実施形態では、環状化プローブは、標的の多型遺伝子座の上流とハイブリダイズするように設計されているプローブの領域および標的の多型遺伝子座の下流とハイブリダイズするように設計されているプローブの領域が、非核酸骨格を通じ共有結合的に接続するように構築される。この骨格は、任意の生体適合性分子または生体適合性分子の組み合わせであってよい。可能性のある生体適合性分子のいくつかの例は、ポリ(エチレングリコール)、ポリカーボネート、ポリウレタン、ポリエチレン、ポリプロピレン、スルホンポリマー、シリコン、セルロース、フルオロポリマー、アクリル化合物、スチレンブロック共重合体、および他のブロック共重合体である。

【0209】

50

本開示のある実施形態では、この手法は、配列内の充填を調べる手段として配列決定を容易に受けられるように改変されている。元の試料の元の対立遺伝子の割合を保持するために、少なくとも1つの重要な考慮すべき事柄を考慮に入れなければならない。ギャップ充填領域内の異なる対立遺伝子の間の可変性の位置は、変異体の鑑別をもたらすDNAポリメラーゼによる開始の偏りがあり得るので、プローブ結合部位に近すぎないようにすべきである。別の考慮すべき事柄は、異なる対立遺伝子からの不均等な増幅をもたらす得るギャップ充填領域内の変異体と相関があるプローブ結合部位にさらなる変動が存在する可能性があることである。本開示のある実施形態では、環状化前プローブの3'末端および5'末端を、標的の対立遺伝子の変異の位置（多型部位）と1つまたは少数の位置だけ離れている塩基とハイブリダイズするように設計する。多型部位（SNPまたは他の種類のもの）と、環状化前プローブの3'末端および／または5'末端がハイブリダイズするように設計されている塩基との間の塩基の数は、1塩基であってよく、2塩基であってよく、3塩基であってよく、4塩基であってよく、5塩基であってよく、6塩基であってよく、7～10塩基であってよく、11～15塩基であってよく、または、16～20塩基、20～30塩基または30～60塩基であってよい。フォワードプライマーおよびリバースプライマーは、多型部位から離れた異なる数の塩基とハイブリダイズするように設計することができる。現行のDNA合成技術を用いて環状化プローブを多数生成することができ、これにより、非常に多数のプローブを生成し、潜在的にプールすることが可能になり、多くの遺伝子座を同時に調べることができる。300, 000超のプローブで作業することが報告されている。標的個体のゲノムのデータを測定するために使用することができる環状化プローブを伴う方法を考察している2つの論文としては、Porrecaら、Nature Methods、2007年、4巻（11号）、931～936頁；および同様にTurnerら、Nature Methods、2009年、6巻（5号）、315～316頁が挙げられる。これらの論文に記載されている方法は、本明細書に記載の他の方法と組み合わせて用いることができる。これらの2つの論文からの方法の特定のステップは、本明細書に記載の他の方法からの他のステップと組み合わせて用いることができる。

【0210】

本明細書に開示されている方法のいくつかの実施形態では、標的個体の遺伝物質を、必要に応じて増幅し、その後、環状化前プローブとハイブリダイズさせ、ギャップ充填を実施してハイブリダイズしたプローブの2つの末端間の塩基を充填し、2つの末端をライゲーションして環状化されたプローブを形成し、環状化されたプローブを、例えば、ローリングサークル増幅を用いて増幅する。所望の標的対立遺伝子の遺伝子情報が適切に設計された環状化オリゴヌクレオチド性プローブ、例えば、LIP系において捕捉されたら、環状化されたプローブの遺伝子配列を測定して、所望の配列データをもたらすことができる。ある実施形態では、適切に設計されたオリゴヌクレオチドプローブを、増幅されなかった標的個体の遺伝物質において直接環状化し、その後増幅することができる。ローリングサークル増幅、MDAまたは他の増幅プロトコルを含めたいくつもの増幅手順を使用して、元の遺伝物質を増幅することまたはLIPを環状化することができることに留意されたい。異なる方法を用いて、例えば、ハイスループット配列決定、サンガー配列決定、他の配列決定方法、ハイブリダイゼーションによる捕捉、環状化による捕捉、多重PCR、他のハイブリダイゼーション方法、およびそれらの組み合わせを用いて、標的ゲノム上の遺伝子情報を測定することができる。

【0211】

上記の方法の1つまたは組み合わせ、インフォマティクスに基づく方法、例えば、PARALLEL SUPPORT（商標）法を、適切な遺伝子測定と一緒に用いて個体の遺伝物質を測定したら、次いで、それを用いて、個体における1つまたは複数の染色体の倍数性状態、および／または対立遺伝子の1つもしくは対立遺伝子の集合（詳細には、対象の疾患または遺伝子の状態と相関する対立遺伝子）の遺伝子の状態を決定することができる。遺伝子配列を多重化捕捉し、その後、配列決定を用いて遺伝子型決定するためのLI

Pの使用が報告されていることに留意されたい。しかし、LIPに基づく戦略によって生じる配列決定データを、単一細胞、少数の細胞または細胞外DNAにおいて見いだされる遺伝物質を増幅するために使用することは、標的個体の倍数性状態を決定するためには用いられていない。

【0212】

ハイブリダイゼーションアレイ、例えば、ILLUMINA INFINIUMアレイまたはAFFYMETRIX遺伝子チップによって測定された遺伝子データから個体の倍数性状態を決定するためのインフォマティクスに基づく方法の適用は、本文書の他の箇所の参考文献に記載されている。しかし、本明細書に記載の方法は、以前に文献に記載された方法に対する改善を示す。例えば、LIPに基づく手法、その後のハイスループット配列決定により、この手法は、多重化についての能力がより優れ、捕捉特異性がより優れ、均一性がより優れ、対立遺伝子の偏りが少ないので、予想外に、より良好な遺伝子型データがもたらされる。多重化がより大きいことにより、より多くの対立遺伝子を標的とすることが可能になり、より正確な結果がもたらされる。均一性がより優れていることにより、より多くの標的の対立遺伝子が測定され、より正確な結果がもたらされる。対立遺伝子の偏りの率がより低いことにより、誤った呼び出しの率が低下し、より正確な結果がもたらされる。より正確な結果により、臨床転帰が改善され、より良い医療がもたらされる。

10

【0213】

LIPを、配列決定以外の方法によって遺伝子型決定するために、DNAの試料における特定の遺伝子座を標的とするための方法として用いることができることに留意することが重要である。例えば、SNPアレイまたは他のDNAもしくはRNAに基づくマイクロアレイを用いて遺伝子型決定するために、LIPを用いてDNAを標的とすることができる。

20

【0214】

ライゲーション媒介性PCR

ライゲーション媒介性PCRは、DNAの混合物における1つまたは複数の遺伝子座を増幅することによってDNAの試料を優先的に富化するために用いるPCRの方法であり、該方法は、プライマー対の集合を得るステップであって、対の各プライマーが標的特異的配列および非標的配列を含有し、標的特異的配列が、標的領域であって、1つが多型部位の上流、および1つが多型部位の下流である標的領域とアニーリングするように設計されており、該標的特異的配列が、多型部位から0、1、2、3、4、5、6、7、8、9、10、11～20、21～30、31～40、41～50、51～100、または、100超隔てられていてよいステップと、上流のプライマーの3'末端からDNAを重合させて、それと、標的分子と相補的なヌクレオチドを有する下流のプライマーの5'末端との間の一本鎖領域を充填するステップと、上流のプライマーの最後の重合した塩基を、近接する下流のプライマーの5'塩基とライゲーションさせるステップと、重合し、ライゲーションした分子のみを、上流のプライマーの5'末端および下流のプライマーの3'末端を含有する非標的配列を使用して増幅するステップとを含む。別個の標的に対するプライマー対を同じ反応において混合することができる。非標的配列は、ユニバーサル配列としての機能を果たし、したがって、首尾よく重合し、ライゲーションした全てのプライマー対を、増幅プライマーの単一の対を用いて増幅することができる。

30

40

【0215】

ハイブリダイゼーションによる捕捉

標的ゲノムにおける特異的な配列の集合を優先的に富化することは、いくつかのやり方で実現することができる。本文書の他の箇所に、特異的な配列の集合を標的とするためにLIPをどのように用いることができるかについての記載があるが、これらの適用の全てにおいて、他の標的化および／または優先的な富化方法を、同じ目的のために同等に良好に用いることができる。別の標的化方法の1つの例はハイブリダイゼーション手法による捕捉である。商業的なハイブリダイゼーション技術による捕捉のいくつかの例としては、AGILENTのSURE SELECT、およびILLUMINAのTRUSEQが挙

50

げられる。ハイブリダイゼーションによる捕捉では、所望の標的の配列と相補的またはほぼ相補的なオリゴヌクレオチドの集合をDNAの混合物とハイブリダイズさせ、次いで混合物から物理的に分離することが可能になる。所望の配列が標的化オリゴヌクレオチドとハイブリダイズしたら、標的化オリゴヌクレオチドを物理的に取り出す作用により、標的の配列も取り出されることになる。ハイブリダイズしたオリゴを取り出したら、それらを、それらの融解温度を上回るまで加熱し、増幅することができる。標的化オリゴヌクレオチドを物理的に取り出すためのいくつかのやり方は、標的化オリゴを固体支持体、例えば磁気ビーズまたはチップと共有結合させることによる。標的化オリゴヌクレオチドを物理的に取り出すための別のやり方は、標的化オリゴヌクレオチドを、別の分子部分に対する強力な親和性を有する分子部分と共有結合させることによる。そのような分子対の例は、例えばSURF SELECTIONにおいて使用されるビオチンおよびストレプトアビジンである。したがって、その標的の配列をビオチン分子に共有結合的に付着させ、ハイブリダイゼーション後に、ストレプトアビジンを付加した固体支持体を使用して、標的の配列がハイブリダイズしたビオチン化オリゴヌクレオチドをプルダウンすることができる。

10

【0216】

ハイブリッド捕捉は、対象の標的と相補的なプローブを標的分子とハイブリダイズさせることを伴う。ハイブリッド捕捉プローブはもともと、標的間に相対的な均一性を有するゲノムの大部分を標的とし、富化するために開発された。その適用では、増幅される全ての標的が、全ての領域を配列決定によって検出することができる十分な均一性を有することが重要であったが、元の試料における対立遺伝子の割合を保持することには注意が払われなかった。捕捉した後、試料中に存在する対立遺伝子を、捕捉された分子の直接配列決定によって決定することができる。これらの配列決定の読み取りを、対立遺伝子の型に応じて分析し、カウントすることができる。しかし、現行の技術を用いると、測定された捕捉された配列の対立遺伝子分布は、一般には、元の対立遺伝子分布を表さない。

20

【0217】

ある実施形態では、配列決定によって対立遺伝子の検出を実施する。多型部位における対立遺伝子の同一性を捕捉するために、捕捉された分子の対立遺伝子の組成を評価するために、配列決定の読み取りが問題の対立遺伝子にわたることが必須である。捕捉分子は多くの場合、長さが変動するので、配列決定の際に、分子全体が配列決定されなければ、変異の位置がオーバーラップすることを保証することができない。しかし、最大の可能性のある長さおよび配列決定の読み取りの正確度に関する費用検討ならびに技術的な限界により、分子全体の配列決定は実行できない。ある実施形態では、約30塩基から約50塩基または約70塩基まで増加させることができる読み取りの長さにより、標的の配列内の変異の位置とオーバーラップする読み取りの数を著しく増加させることができる。

30

【0218】

対象の位置を調べる読み取りの数を増加させるための別のやり方は、基礎をなす富化された対立遺伝子の偏りをもたらしなす限りはプローブの長さを減少させることである。合成されたプローブの長さは、1つの遺伝子座において見いだされた2つの異なる対立遺伝子とハイブリダイズするように設計された2種のプローブが元の試料中の種々の対立遺伝子とほぼ同等の親和性でハイブリダイズするために十分な長さであるべきである。現在、当該技術分野で公知の方法には、一般には120塩基より長いプローブが記載されている。現行の実施形態において、対立遺伝子が1つまたは少数の塩基である場合、捕捉プローブは、約110塩基未満、約100塩基未満、約90塩基未満、約80塩基未満、約70塩基未満、約60塩基未満、約50塩基未満、約40塩基未満、約30塩基未満、および約25塩基未満であり得、これは、全ての対立遺伝子からの同等の富化を確実にするために十分である。ハイブリッド捕捉技術を用いて富化するDNAの混合物が、血液、例えば母系の血液から単離された浮動性DNAを含む混合物である場合、DNAの平均長はかなり短く、一般には、200塩基未満である。より短いプローブを使用することにより、ハイブリッド捕捉プローブが所望のDNA断片を捕捉する見込みが大きくなる。より大きな変動は、より長いプローブを必要とする場合がある。ある実施形態では、対象の変動は、1

40

50

(S N P) ～数塩基の長さである。ある実施形態では、ゲノム内の標的の領域を、ハイブリッド捕捉プローブを使用して優先的に富化することができ、ここで、ハイブリッド捕捉プローブの長さは90塩基未満であり、80塩基未満、70塩基未満、60塩基未満、50塩基未満、40塩基未満、30塩基未満または25塩基未満であってよい。ある実施形態では、所望の対立遺伝子が配列決定される見込みを増大させるために、多型の対立遺伝子の場所に隣接している領域とハイブリダイズするように設計されているプローブの長さを、90塩基超から、約80塩基まで、または約70塩基まで、または約60塩基まで、または約50塩基まで、または約40塩基まで、または約30塩基まで、または約25塩基まで減少させることができる。

【0219】

捕捉を可能にするために、合成されたプローブと標的分子の間に最小のオーバーラップが存在する。この合成されたプローブは、できるだけ短くすることができるが、それでもこの最小の必要なオーバーラップよりも大きい。多型領域を標的とするためにより短いプローブ長を用いることの効果は、より多くの分子が標的対立遺伝子領域とオーバーラップすることである。元のDNA分子の断片化の状態も、標的の対立遺伝子とオーバーラップする読み取りの数に影響を及ぼす。血漿試料などの一部のDNA試料は、*in vivo*で起こる生物学的プロセスに起因して既に断片化されている。しかし、より長い断片を有する試料には、配列決定ライブラリーの調製および富化の前の断片化が有益である。プローブと断片の両方が短い(約60～80bp)場合、最大の特異性は、対象の重要な領域とオーバーラップできない比較的少ない配列読み取りで実現することができる。

【0220】

ある実施形態では、ハイブリダイゼーション条件を調整して、元の試料中に存在する異なる対立遺伝子の捕捉における均一性を最大にすることができる。ある実施形態では、対立遺伝子間のハイブリダイゼーションの偏りの差異を最小限にするためにハイブリダイゼーション温度を低下させる。当技術分野で公知の方法では、温度を低下させることには、プローブと、意図されたものではない標的とのハイブリダイゼーションを増大させる効果があるので、ハイブリダイゼーションのためにより低い温度を用いることを回避する。しかし、目的が、最大の忠実度で対立遺伝子の比を保存することである場合、現行技術の教示がこの手法を避けているという事実にもかかわらずより低いハイブリダイゼーション温度を用いる手法により、最適に正確な対立遺伝子の比がもたらされる。ハイブリダイゼーション温度は、標的の領域の実質的なオーバーラップを有する標的のみが捕捉されるように、標的と合成されたプローブとの間のより大きなオーバーラップを必要とするために上昇させることもできる。本開示のいくつかの実施形態では、ハイブリダイゼーション温度を、通常のハイブリダイゼーション温度から、約40℃まで、約45℃まで、約50℃まで、約55℃まで、約60℃まで、約65℃までまたは約70℃まで低下させる。

【0221】

ある実施形態では、ハイブリッド捕捉プローブは、多型の対立遺伝子に隣接している領域に見いだされるDNAと相補的なDNAを有する捕捉プローブの領域が多型部位のすぐ隣ではないように設計することができる。その代わりに、捕捉プローブは、標的の多型部位に隣接しているDNAとハイブリダイズするように設計されている捕捉プローブの領域が、多型部位とファンデルワールスにより接触する捕捉プローブの部分と、1つまたは少数の塩基と等しい長さの小さな距離で隔てられているように設計することができる。ある実施形態では、ハイブリッド捕捉プローブは、多型の対立遺伝子と隣接しているが、それとは交差していない領域とハイブリダイズするように設計されており、これは、隣接捕捉プローブと称される。隣接捕捉プローブの長さは、約120塩基未満、約110塩基未満、約100塩基未満、約90塩基未満であってよく、約80塩基未満、約70塩基未満、約60塩基未満、約50塩基未満、約40塩基未満、約30塩基未満、または約25塩基未満であってよい。隣接捕捉プローブの標的となるゲノムの領域は、多型遺伝子座と1塩基対、2塩基対、3塩基対、4塩基対、5塩基対、6塩基対、7塩基対、8塩基対、9塩基対、10塩基対、11～20塩基対、または、20超塩基対で隔てられていてよい。

【0222】

標的化配列捕捉を用いる、標的化捕捉に基づく疾患スクリーニング検査についての記載。現在AGILENT (SURE SELECT)、ROCHE-NIMBLEGENまたはILLUMINAから提供されているものなどの特別注文の標的化配列捕捉。捕捉プローブは、種々の種類の変異の捕捉を確実にするために特別注文で設計することができる。点変異については、点変異とオーバーラップする1つまたは複数のプローブが、変異を捕捉し、配列決定するために十分であるはずである。

【0223】

小さな挿入または欠失については、変異とオーバーラップする1つまたは複数のプローブが、変異を含む断片を捕捉し、配列決定するために十分であり得る。ハイブリダイゼーションは、一般には、ゲノム配列を参照するために設計されるプローブ限定捕捉効率の間で効率が低い可能性がある。変異を含む断片の捕捉を確実にするために、正常な対立遺伝子と一致するものと、変異対立遺伝子と一致するものの2種のプローブを設計することができる。より長いプローブにより、ハイブリダイゼーションを増強することができる。複数のオーバーラッププローブにより捕捉を増強することができる。最後に、プローブをすぐ隣であるがオーバーラップはしていないところに置くと、変異は、正常な対立遺伝子と変異対立遺伝子の比較的同様の捕捉効率が可能となり得る。

10

【0224】

単純タンデム反復 (STR) については、これらの高度に可変性の部位とオーバーラップしているプローブは、断片を上手く捕捉する可能性が低い。捕捉を増強するために、プローブを、可変性部位と近接しているがオーバーラップはしていないところに置くことができる。次いで、断片について通常通り配列決定して、STRの長さおよび組成を示すことができる。

20

【0225】

大規模な欠失については、現在エクソーム捕捉系において用いられている一般的な手法である一連のオーバーラッププローブが機能し得る。しかし、この手法を用いると、個体がヘテロ接合性であるか否かを決定することが難しい場合がある。捕捉された領域内のSNPを標的とし、評価することにより、個体が保有者であることを示す、その領域にわたるヘテロ接合性の損失を潜在的に示すことができる。ある実施形態では、非オーバーラッププローブまたはシングルトンプローブを、潜在的に欠失した領域にわたって置き、捕捉された断片の数をヘテロ接合性の尺度として使用することが可能である。個体が大規模な欠失を有する場合には、非欠失 (二倍体) 参照遺伝子座と比較して、断片の数の2分の1を捕捉のために利用可能であることが予測される。したがって、欠失した領域から得られた読み取りの数は、正常な二倍体の遺伝子座から得られた読み取りの数のおよそ半分であるはずである。潜在的に欠失した領域にわたる複数のシングルトンプローブからの配列決定の読み取りの深さを総計し、平均することにより、シグナルを増強し、診断の信頼度を改善することができる。2つの手法、SNPを標的として、ヘテロ接合性の損失を同定すること、および複数のシングルトンプローブを使用して、その遺伝子座から基礎をなす断片の量の定量的尺度を得ることを組み合わせることもできる。これらの戦略のいずれか、または両方を、他の戦略と組み合わせて、同じ結果をよりよく得ることができる。

30

40

【0226】

試験の間に、同じ試験において捕捉され、配列決定されるY染色体断片の存在によって示される、男の胎児のcfDNAの検出、ならびに母親および父親が影響されないX連鎖優性変異または母親が影響されない優性変異のいずれかにより、胎児に対するリスクが高まることが示される。影響のない母親における同じ遺伝子内に2つの変異劣性対立遺伝子が検出されることは、胎児が、変異対立遺伝子を父親から、および潜在的に第2の変異対立遺伝子を母親から遺伝によって受け継いだことを意味する。全ての場合において、羊水穿刺または絨毛膜絨毛採取による追跡検査も示され得る。

【0227】

標的化捕捉に基づく疾患スクリーニング検査は、異数性についての標的化捕捉に基づく

50

非侵襲的な出生前診断検査と組み合わせることができる。

【0228】

読み取りの深さ（DOR）の変動性を減少させるためのいくつかのやり方が存在する：例えば、プライマー濃度を上昇させることができる、より長い標的化増幅プローブを使用することができる、または、S T A サイクルをより多く実行することができる（例えば、25超、30超、35超、または、さらには40超）。

【0229】

標的化PCR

いくつかの実施形態では、PCRを用いて、ゲノムの特定の場所を標的とすることができる。血漿試料において、元のDNAは高度に断片化されている（一般には、500bp未満、平均長200bp未満）。PCRでは、増幅を可能にするために、フォワードプライマーとリバースプライマーの両方が同じ断片とアニーリングしなければならない。したがって、断片が短い場合、PCRアッセイでは、同様に比較的短い領域を増幅しなければならない。MIP Sのように、多型の位置がポリメラーゼ結合部位と近すぎると、異なる対立遺伝子からの増幅に偏りが生じる。現在、SNPを含有するものなどの多型領域を標的とするPCRプライマーは、一般には、プライマーの3'末端が1つまたは複数の多型の塩基のすぐ隣の塩基とハイブリダイズするように設計される。本開示のある実施形態では、フォワードPCRプライマーおよびリバースPCRプライマーの両方の3'末端が、標的の対立遺伝子の変異の位置（多型部位）と1つまたは少数の位置だけ離れている塩基とハイブリダイズするように設計する。多型部位（SNPまたは他の種類のもの）と、プライマーの3'末端がハイブリダイズするように設計された塩基との間の塩基の数は、1塩基であってよい、2塩基であってよい、3塩基であってよい、4塩基であってよい、5塩基であってよい、6塩基であってよい、7～10塩基であってよい、11～15塩基であってよい、または、16～20塩基であってよい。フォワードプライマーおよびリバースプライマーは、多型部位から離れた異なる数の塩基とハイブリダイズするように設計することができる。

【0230】

PCRアッセイを多数生成することができるが、異なるPCRアッセイ間の相互作用により、約100アッセイを越えてそれらを多重化することが難しくなる。種々の複雑な分子的手法を用いて、多重化のレベルを上昇させることができるが、それでも反応当たり100未満、おそらく200、またはことによると500アッセイに限られ得る。多量のDNAを有する試料は、複数の副次反応に分割し、次いで配列決定の前に組み換えることができる。DNA分子の全体的な試料または一部の垂集団のいずれかが限られている試料については、試料を分割することにより統計学的なノイズが導入されることになる。ある実施形態では、少ないまたは限られた量のDNAとは、10pg未満、10pgから100pgの間、100pgから1ngの間、1ngから10ngの間または10ngから100ngの間の量を指し得る。この方法は、複数のプールに分割するステップを包含する他の方法では確率的なノイズの導入に関連する重大な問題が引き起こされ得る少量のDNAに対して特に有用であるが、この方法は、いかなる量のDNAの試料に対して実行された場合でも、偏りを最小化する利益をもたらすことに留意されたい。これらの状況では、ユニバーサル前増幅ステップを使用して、全体的な試料の量を増大させることができる。理想的には、この前増幅ステップでは、対立遺伝子分布を感知できるほどには変えるべきでない。

【0231】

ある実施形態では、本開示の方法により、体液由来の単一細胞またはDNAなどの限られた試料から、配列決定またはいくつかの他の遺伝子型決定方法によって遺伝子型決定するために、多数の標的の遺伝子座、詳細には1,000～5,000の遺伝子座、5,000～10,000の遺伝子座、または10,000超の遺伝子座に特異的なPCR産物を生成することができる。現在、5超～10個の標的の多重PCR反応を実施することにより、大きな課題が示され、多くの場合、プライマー副産物、例えば、プライマー二量体

10

20

30

40

50

、および他のアーチファクトが妨害となる。ハイブリダイゼーションプローブを用いたマイクロアレイを使用して標的配列を検出する場合、プライマー二量体および他のアーチファクトは検出されないので、これらは無視することができる。しかし、検出の方法として配列決定を用いる場合、配列決定の読み取りの大部分は、そのようなアーチファクトを配列決定し、試料中の所望の標的配列は配列決定しない。1回の反応において50超または100の反応を多重化し、その後に配列決定するために用いられる先行技術に記載の方法により、一般には20%超、および多くの場合50%超、多くの場合80%超およびいくつかの場合には90%超のオフターゲットの配列読み取りがもたらされる。

【0232】

一般に、試料における多数の(n)標的(50超、100超、500超または1,000超)に対し標的化配列決定を実施するために、試料をいくつもの並行した反応物へと分割し、1つの個体標的を増幅することができる。これは、PCR多ウェルプレートにおいて実施されている、または商業的なプラットフォーム、例えば、FLUIDIGM ACCESS ARRAY (マイクロ流体チップにおいて試料当たり48の反応)またはRAIN DANCE TECHNOLOGYからのDROPLET PCR (100~数千もの標的)において実行することができる。残念ながら、これらのスプリットアンドプール(split-and-pool)方法は、多くの場合、存在するゲノムのコピーが、各ウェル中にゲノムの各領域の1つのコピーが存在することを確実にするためには不十分であるので、DNAの量が限られている試料に対しては問題がある。これは、多型遺伝子座を標的とする場合に特に重大な問題であり、分割およびプールすることによって導入される確率論的ノイズにより、元のDNAの試料に存在していた対立遺伝子の割合の非常に不十分に正確な測定が引き起こされるので、多型遺伝子座における対立遺伝子の相対的な割合が必要である。限られた量のDNAのみが利用可能である場合に適用可能な、多くのPCR反応物を有効かつ効率的に増幅するための方法が本明細書に記載されている。ある実施形態では、単一細胞、体液、DNAの混合物、例えば、母系の血漿中に見いだされる浮動性DNA、生検材料、環境試料および/または法医学試料を分析するために該方法を適用することができる。

【0233】

ある実施形態では、標的化配列決定は、以下のステップの1つ、複数または全てを伴い得る。a) DNA断片の両末端にアダプタ配列を有するライブラリーを生成し、増幅するステップ。b) ライブラリー増幅後に複数の反応物に分割するステップ。c) DNA断片の両末端にアダプタ配列を有するライブラリーを生成し、必要に応じて増幅するステップ。d) 標的当たり1つの標的特異的「フォワード」プライマーおよび1つのタグ特異的プライマーを使用して選択された標的の1000~10,000プレックス増幅を実施するステップ。e) この産物から、「リバーサ」標的特異的プライマーおよび1つ(または複数)の、第1ラウンドにおいて標的特異的フォワードプライマーの一部として導入されたユニバーサルタグに特異的なプライマーを使用して第2の増幅を実施するステップ。f) 選択された標的の1000プレックス前増幅を、限られたサイクル数で実施するステップ。g) 産物を複数の一定分量に分け、個々の反応における標的のサブプールを増幅するステップ(例えば、50~500プレックス、しかし、これは単一プレックスに至るまで使用することができる。h) 並行サブプール反応の産物をプールするステップ。i) これらの増幅の間、プライマーは、産物を配列決定できるように配列決定適合性タグ(部分長、または完全長)を担持してよい。

【0234】

高度多重PCR

本明細書には、血漿から得られたゲノムDNA由来の100~数万をも超える標的配列(例えば、SNP遺伝子座)の標的化増幅を可能にする方法が開示されている。増幅された試料は、プライマー二量体産物を比較的含まず、標的遺伝子座における対立遺伝子の偏りが少ない。増幅の間または増幅後に、産物に配列決定適合性アダプタを付加する場合、これらの産物の分析を配列決定によって実施することができる。

【0235】

当技術分野で公知の方法を用いて高度多重PCR増幅を実施することにより、所望の増幅産物が過剰であり、配列決定に適さないプライマー二量体産物が生成する。これらは、経験的に、これらの産物を形成するプライマーを排除することによって、またはプライマーの *in silico* 選択を実施することによって減少させることができる。しかし、アッセイの数が多くなるほど、この問題はより困難になる。

【0236】

1つの解法は、5,000プレックス反応をいくつかの低プレックス増幅、例えば、100回の50プレックス反応または50回の100プレックス反応に分割すること、またはマイクロフレイディクスを使用すること、または、さらには、試料を個々のPCR反応に分割することである。しかし、妊娠血漿からの非侵襲的な出生前診断においてなど試料DNAが限られている場合は、多数の反応間に試料を分割することは、これによりボトルネッキングが生じるので、回避するべきである。

【0237】

本明細書には、まず試料の血漿DNAを全体的に増幅し、次いで試料を、反応当たり、より中程度の数の標的配列を伴う複数の多重化標的富化反応に分割するための方法が記載されている。ある実施形態では、本開示の方法は、DNA混合物を複数の遺伝子座で優先的に富化するために用いることができ、該方法は以下のステップの1つまたは複数を含む：DNAの混合物からライブラリーを生成し、増幅するステップであって、ライブラリー内の分子がDNA断片の両末端にライゲーションされたアダプタ配列を有するステップ、増幅されたライブラリーを複数の反応に分割するステップ、選択された標的の多重増幅の第1ラウンドを、標的当たり1つの標的特異的「フォワード」プライマーおよび1つまたは複数のアダプタ特異的なユニバーサル「リバーサ」プライマーを使用して実施するステップ。ある実施形態では、本開示の方法は、「リバーサ」標的特異的プライマー、および1つまたは複数の、第1ラウンドにおいて標的特異的フォワードプライマーの一部として導入されたユニバーサルタグに特異的なプライマーを使用して第2の増幅を実施するステップをさらに含む。ある実施形態では、該方法は、完全ネステッドPCR手法、ヘミネステッドPCR手法、セミネステッドPCR手法、片側完全ネステッドPCR手法、片側ヘミネステッドPCR手法または片側セミネステッドPCR手法を伴ってよい。ある実施形態では、DNA混合物を複数の遺伝子座において優先的に富化するために本開示の方法を用い、該方法は、選択された標的の多重化前増幅を、限られたサイクル数で実施するステップと、産物を複数の一定分量に分けるステップと、個々の反応における標的のサブプールを増幅するステップと、並行サブプール反応の産物をプールするステップとを含む。この手法は、50～500遺伝子座について、500～5,000遺伝子座について、5,000～50,000遺伝子座について、または、さらには50,000～500,000遺伝子座について、対立遺伝子の偏りが低レベルになるように標的化増幅を実施するために用いることができることに留意されたい。ある実施形態では、プライマーは、部分長、または完全長の配列決定適合性タグを担持する。

【0238】

ワークフローは、(1)血漿DNAを抽出するステップと、(2)断片の両末端にユニバーサルアダプタを有する断片ライブラリーを調製するステップと、(3)ライブラリーを、アダプタに特異的なユニバーサルプライマーを使用して増幅するステップと、(4)増幅された試料「ライブラリー」を複数の一定分量に分けるステップと、(5)一定分量に対して多重化(例えば、標的当たり1つの標的特異的プライマーおよびタグ特異的プライマーを用いた約100プレックス、1,000または10,000プレックス)増幅を実施するステップと、(6)1つの試料の一定分量をプールするステップと、(7)試料についてバーコーディングを行うステップと、(8)試料を混合し、濃度を調整するステップと、(9)試料について配列決定するステップとを伴ってよい。ワークフローは、列挙されているステップのうちの1つを含有する複数のサブステップを含んでよい(例えば、ステップ(2)のライブラリーを調製するステップは、3つの酵素的ステップ(平滑末

端化、dAテリングおよびアダプタライゲーション) および3つの精製ステップを伴ってよい)。ワークフローのステップは、組み合わせることができる、分けることができる、または異なる順序で実施することができる(例えば、試料のバーコーディングおよびプール)。

【0239】

ライブラリーの増幅は、短い断片をより効率的に増幅することに偏りがあるように実施することができることに留意することが重要である。このように、より短い配列、例えば、モノヌクレオソームのDNA断片を妊娠中の女性の循環中に見いだされる無細胞の胎児DNA(胎盤起源の)として優先的に増幅することが可能である。PCRアッセイは、タグ、例えば配列決定タグ(通常15~25塩基の切断形態)を有してよいことに留意されたい。多重化した後、試料のPCR多重化産物をプールし、次いで、タグ特異的PCRによって(ライゲーションによって行うこともできる)タグ付けを完了する(バーコーディングを含む)。また、多重化として完全な配列決定タグを同じ反応に加えることができる。第1のサイクルでは、標的を、標的特異的プライマーを用いて増幅することができ、その後、タグ特異的プライマーが優勢になって完全なSQアダプタ配列を完成させる。PCRプライマーはタグを担持しなくてよい。配列決定タグはライゲーションによって増幅産物に付加することができる。

10

【0240】

ある実施形態では、高度多重PCR、その後クローン配列決定を用いて増幅された材料を評価することによって、胎児の異数性を検出することができる。従来の多重PCRでは最大で50遺伝子座を同時に評価するが、本明細書に記載の手法を使用して、50超の遺伝子座を同時に、100超遺伝子座を同時に、500超遺伝子座を同時に、1,000超遺伝子座を同時に、5,000超遺伝子座を同時に、10,000超遺伝子座を同時に、50,000超遺伝子座を同時に、および100,000超遺伝子座を同時に、同時評価することを可能にし得る。実験により、10,000まで、10,000を含めて、および10,000超の別個の遺伝子座を、単一反応において、十分に優良な効率および特異性で同時に評価して、非侵襲的な出生前異数性診断および/またはコピー数の呼び出しを高い正確度で行うことができることが示された。アッセイは、単一反応において、母系の血漿から単離されたcfDNA試料全体、その画分またはcfDNA試料のさらに加工した誘導体と組み合わせることができる。cfDNAまたは誘導体は、複数の並行の多重反応に分割することもできる。最適な試料の分割および多重化を、種々の性能仕様のトレードオフによって決定する。材料の量が限られているので、試料を複数の画分に分割することにより、サンプリングノイズ、取扱い時間、およびエラーの可能性の増大がもたらされる可能性がある。逆に、高多重化の結果、偽の増幅の量が増え、増幅の不等性が増す可能性があり、どちらによっても検査性能が低下する。

20

30

【0241】

本明細書に記載の方法の適用における2つの極めて重要な関連する考慮すべき事柄は、限られた量の元の血漿および対立遺伝子頻度または他の測定値を得る材料内の元の分子の数である。元の分子の数が特定のレベルを下回る場合、ランダムサンプリングノイズが著しくなり、検査の正確度に影響を及ぼす可能性がある。一般には、標的遺伝子座当たり500~1000個の元の分子相当を含む試料に対して測定を行う場合、非侵襲的な出生前異数性診断を行うために十分な品質のデータを得ることができる。別個の測定値の数を増加させる、例えば、試料の体積を増加させるいくつかのやり方が存在する。試料に適用される各操作によっても、潜在的に材料が損失する。検査の性能を低下させ得る損失を回避するために、種々の操作によって受けた損失を特徴付けること、および特定の操作を回避するまたは必要に応じてその収量を改善することが必須である。

40

【0242】

ある実施形態では、元のcfDNA試料の全てまたはある割合を増幅することによって、その後のステップにおける潜在的な損失を減ずることが可能である。試料中の遺伝物質の全てを増幅し、それにより下流の手順のために利用可能な量を増大させるために、種々

50

の方法が利用可能である。ある実施形態では、ライゲーション媒介性PCR (LM-PCR) DNA断片を、1つの別個のアダプタ、2つの別個のアダプタまたは多くの別個のアダプタのいずれかをライゲーションした後に、PCRによって増幅する。ある実施形態では、多置換増幅(MDA) phi-29ポリメラーゼを使用して、全てのDNAを等温的に増幅する。DOP-PCRおよびその変形では、ランダムプライミングを使用して元の材料DNAを増幅する。各方法は、特定の特性、例えば、代表的なゲノムの領域全てにわたる増幅の均一性、元のDNAの捕捉および増幅の効率、および断片の長さに応じた増幅性能を有する。

【0243】

ある実施形態では、LM-PCRを、3'チロシンを有する単一のヘテロ二本鎖アダプタと一緒に使用することができる。ヘテロ二本鎖アダプタにより、PCRの第1ラウンドの間に元のDNA断片の5'末端および3'末端上の2つの別個の配列に変換することができる単一のアダプタ分子を使用することが可能になる。ある実施形態では、サイズ分離によって増幅されたライブラリー、またはAMPURE、TASS、もしくは他の同様の方法などの産物を分画することが可能である。ライゲーションの前に、試料DNAを平滑末端化し、次いで、単一のアデノシン塩基を3'末端に付加する。ライゲーションの前に、DNAを、制限酵素またはいくつかの他の切断方法を用いて切断することができる。ライゲーションの間、試料断片の3'アデノシンおよびアダプタの相補的な3'チロシンオーバーハングにより、ライゲーション効率が増強され得る。PCR増幅の伸長ステップは、約200bp、約300bp、約400bp、約500bpまたは約1,000bpより長い断片からの増幅を低下させるための時間の観点から、限られ得る。母系の血漿中に見いだされるより長いDNAはほぼ排他的に母系であるので、これにより、胎児DNAの10~50%の富化および検査性能の改善がもたらされ得る。市販のキットによって規定されている条件を用いていくつもの反応を実行し、試料DNA分子の10%未満の上首尾のライゲーションがもたらされた。これに対する反応条件の一連の最適化により、ライゲーションがおおよそ70%に改善された。

【0244】

Mini-PCR

従来のPCRアッセイ設計により、明確な胎児の分子が著しく損失するが、この損失は、mini-PCRアッセイと称される非常に短いPCRアッセイを設計することによって著しく低下させることができる。母系の血清中の胎児のcfDNAは高度に断片化されており、断片サイズはほぼガウス様式で分布しており、平均が160bpであり、標準偏差が15bpであり、最小サイズが約100bpであり、最大サイズが約220bpである。標的の多型に関する断片の開始位置および終了位置の分布は、必ずしもランダムではないが、個々の標的の間で、および集合的に全ての標的の間で広範に変動し、1つの特定の標的遺伝子座の多型部位は、その遺伝子座を起源とする種々の断片の中で開始から終了までの任意の位置を占有し得る。mini-PCRという用語は、さらなる制限または限定なく、通常のPCRを等しく良好に指し得ることに留意されたい。

【0245】

PCRの間、増幅はフォワードプライマー部位とリバースプライマー部位の両方を含む鋳型DNA断片のみから起こる。胎児のcfDNA断片は短いので、両方のプライマー部位が存在する尤度であって、フォワードプライマー部位とリバースプライマー部位の両方を含む長さLの胎児の断片の尤度は、アンプリコンの長さsと断片の長さLの比である。理想的な条件下では、アンプリコンが45bp、50bp、55bp、60bp、65bpまたは70bpであるアッセイにより、それぞれ、利用可能な鋳型断片分子の72%、69%、66%、63%、59%または56%から首尾よく増幅される。アンプリコンの長さは、フォワードプライミング部位およびリバースプライミング部位の5'末端の間の距離である。当業者により一般に使用されるものよりも短い長さのアンプリコンにより、必要な短い配列読み取りのみによる所望の多型遺伝子座のより効率的な測定がもたらされ得る。ある実施形態では、アンプリコンの実質的な画分は100bp未満、90bp未満、8

0 b p 未満、7 0 b p 未満、6 5 b p 未満、6 0 b p 未満、5 5 b p 未満、5 0 b p 未満、または4 5 b p 未満であるべきである。

【0 2 4 6】

先行技術で公知の方法において、本明細書に記載のものなどの短いアッセイは、必要ではなく、また、プライマーの設計に対して、プライマーの長さの限定、アニーリング特性、およびフォワードプライマーとリバースプライマーの間の距離によってかなりの制約を課すので、通常は回避されることに留意されたい。

【0 2 4 7】

いずれかのプライマーの3'末端が多型部位のおよそ1~6塩基の範囲内である場合、増幅の偏りが潜在的に存在することにも留意されたい。最初にポリメラーゼが結合する部位におけるこの一塩基の差異により、一方の対立遺伝子の優先的な増幅がもたらされる可能性があり、これにより、観察される対立遺伝子頻度に変更され、性能が低下する可能性がある。これらの制約の全てにより、特定の遺伝子座を首尾よく増幅するプライマーを同定すること、およびそれに加えて、同じ多重化反応に適合するプライマーの大きな集合を設計することが非常に困難になる。ある実施形態では、内側のフォワードプライマーおよびリバースプライマーの3'末端を、多型部位の上流にあり、少数の塩基で多型部位から隔てられているDNAの領域とハイブリダイズするように設計する。理想的には、塩基の数は6塩基から10塩基の間であってよいが、同等に良好に、4塩基から15塩基の間、3塩基から20塩基の間、2塩基から30塩基の間または1塩基から60塩基の間であってよく、実質的に同じ結果が実現され得る。

【0 2 4 8】

多重PCRは、全ての標的が増幅される単回ラウンドのPCRを伴ってよい、または、1ラウンドのPCR、その後の1または複数のラウンドのネステッドPCRまたはネステッドPCRのいくつかの変形を伴ってよい。ネステッドPCRは、前のラウンドで使用されたプライマーよりも少なくとも1つの塩基対だけ内部に結合する1つまたは複数の新しいプライマーを使用した、次の1または複数のラウンドのPCR増幅からなる。ネステッドPCRにより、正確な内部の配列を有する前の反応からの増幅産物のみを、その後の反応において、増幅することによって偽の増幅標的の数が減少する。偽の増幅標的が減少することにより、特に配列決定において得ることができる有用な測定値の数が改善される。ネステッドPCRは、一般には、前のプライマー結合部位よりも完全に内部のプライマーを設計することを伴い、必然的に、増幅のために必要な最小のDNAセグメントサイズが増大する。DNAが高度に断片化されている母系の血漿cfDNAなどの試料については、より大きなアッセイサイズにより、測定値を得ることができる別個のcfDNA分子の数が減少する。ある実施形態では、この影響を相殺するために、第2ラウンドのプライマーの一方または両方が、いくつかの数の塩基を内部に伸長させている第1の結合部位とオーバーラップしている部分的なネスティング手法を用いて、全アッセイサイズの拡大を最小限にしながらさらに別の特異性を実現することができる。

【0 2 4 9】

ある実施形態では、PCRアッセイの多重プールを、潜在的にヘテロ接合性である1つまたは複数の染色体上のSNPまたは他の多型遺伝子座または非多型遺伝子座を増幅するように設計し、これらのアッセイを単一反応において用いてDNAを増幅する。PCRアッセイの数は、50回から200回の間のPCRアッセイ、200回から1,000回の間のPCRアッセイ、1,000回から5,000回の間のPCRアッセイまたは5,000回から20,000回の間のPCRアッセイ(50~200プレックス、200~1,000プレックス、1,000~5,000プレックス、5,000~20,000プレックス、20,000超プレックスそれぞれ)であってよい。ある実施形態では、約10,000PCRアッセイ(10,000プレックス)の多重プールを、X染色体、Y染色体、第13染色体、第18染色体、および第21染色体および第1染色体または第2染色体上の潜在的にヘテロ接合性であるSNP遺伝子座を増幅するように設計し、これらのアッセイを単一反応において用いて、材料血漿試料、絨毛膜絨毛試料、羊水穿刺試料、単

一または少数の細胞、他の体液または組織、がんまたは他の遺伝物質から得られた c f D N A を増幅する。各遺伝子座の S N P 頻度は、アンプリコンの配列決定のクローンによる方法またはいくつかの他の方法によって決定することができる。対立遺伝子頻度分布の統計分析または全てのアッセイの比を使用して、試料が、検査に含まれる染色体のうちの 1 つまたは複数のトリソミーを含有するかどうかを決定することができる。別の実施形態では、元の c f D N A 試料を 2 つの試料に分割し、並行した 5, 0 0 0 プレックスアッセイを実施する。別の実施形態では、元の c f D N A 試料を n 個の試料に分割し、並行した (約 1 0, 0 0 0 / n) プレックスアッセイを実施し、ここで n は、2 から 1 2 の間または 1 2 から 2 4 の間または 2 4 から 4 8 の間または 4 8 から 9 6 の間である。データを収集し、既に記載されているものと同様に分析する。この方法は、転座、欠失、重複、および他の染色体異常を検出することに同等に良好に適用可能であることに留意されたい。

10

【0250】

ある実施形態では、標的ゲノムに対する相同性を有さない尾部をプライマーのいずれかの 3' 末端または 5' 末端に付加することもできる。これらの尾部により、その後の操作、手順または測定が容易になる。ある実施形態では、尾部配列は、標的特異的フォワードプライマーと標的特異的リバースプライマーに対して同じであってよい。ある実施形態では、標的特異的フォワードプライマーと標的特異的リバースプライマーのために異なる尾部を用いることができる。ある実施形態では、異なる遺伝子座または遺伝子座の集合に対して複数の異なる尾部を用いることができる。全ての遺伝子座の間で、または遺伝子座のサブセットの間で特定の尾部が共有されてよい。例えば、現行の配列決定プラットフォームのいずれかに必要なフォワード配列およびリバース配列に対応するフォワード尾部およびリバース尾部を用いて、増幅後の直接配列決定が可能になる。ある実施形態では、尾部を、全ての増幅された標的の間で、他の有用な配列を付加するために使用することができる一般的なプライミング部位として使用することができる。いくつかの実施形態では、内側のプライマーは、標的の多型遺伝子座の上流または下流のいずれかとハイブリダイズするように設計された領域を含有してよい。いくつかの実施形態では、プライマーは、分子バーコードを含有してよい。いくつかの実施形態では、プライマーは、P C R 増幅が可能になるように設計されたユニバーサルプライミング配列を含有してよい。

20

【0251】

ある実施形態では、1 0, 0 0 0 プレックス P C R アッセイプールを、フォワードプライマーおよびリバースプライマーが、ハイスループット配列決定計器、例えば、I L L U M I N A から入手可能な H I S E Q、G A I I X または M Y S E Q に必要な所要のフォワード配列およびリバース配列に対応する尾部を有するように作製する。さらに、配列決定尾部に対して 5' 側には、その後の P C R においてアンプリコンにヌクレオチドバーコード配列を付加するためのプライミング部位として用いることができるさらに別の配列が含まれ、それにより、複数の試料をハイスループット配列決定計器の単一のレーンで多重化配列決定することが可能になる。

30

【0252】

ある実施形態では、1 0, 0 0 0 プレックス P C R アッセイプールを、リバースプライマーが、ハイスループット配列決定計器に必要な所要のリバース配列に対応する尾部を有するように作製する。第 1 の 1 0, 0 0 0 プレックスアッセイで増幅した後、その後の P C R 増幅を、全ての標的に対して部分ネステッドフォワードプライマー (例えば、6 塩基ネステッド)、および第 1 ラウンドに含まれたリバース配列決定尾部に対応するリバースプライマーを有する別の 1 0, 0 0 0 プレックスプールを使用して実施することができる。この、ただ 1 つの標的特異的プライマーおよびユニバーサルプライマーを用いた部分ネステッド増幅の次のラウンドでは、アッセイに必要なサイズが限られ、サンプリングノイズが低下するが、偽のアンプリコンの数が著しく減少する。配列決定タグを、付加したライゲーションアダプタに、および/または P C R プローブの一部として付加することができ、したがって、タグは最後のアンプリコンの一部になる。

40

【0253】

50

胎児の割合は検査の性能に影響を及ぼす。母系の血漿中に見いだされるDNAの胎児の割合を富化するためのいくつかのやり方が存在する。胎児の割合は、既に考察されている上記のLM-PCR法によって、ならびに長い母系断片の標的化除去によって増大させることができる。ある実施形態では、標的遺伝子座の多重PCR増幅の前に、追加の多重PCR反応を行って、その後の多重PCRにおいて標的とされる遺伝子座に対応する長くて大きい母系断片を選択的に除去することができる。追加のプライマーを、無細胞の胎児DNA断片の間に存在することが予想されるよりも多型からの距離が長い部位とアニーリングするように設計する。これらのプライマーは、標的多型遺伝子座の多重PCRの前の1サイクル多重PCR反応において用いることができる。これらの遠位プライマーには、タグを付けたDNAの小片の選択的な認識を可能にし得る分子または部分でタグ付けする。ある実施形態では、これらのDNAの分子は、1サイクルのPCR後に、これらのプライマーを含む新しく形成された二本鎖DNAを除去することを可能にするビオチン分子を用いて共有結合的に修飾することができる。その第1ラウンドの間に形成された二本鎖DNAは、母体起源であるようである。ハイブリッド材料の除去は、磁性ストレプトアビジンビーズを使用することにより実現することができる。同等に良好に機能し得るタグ付けの他の方法が存在する。ある実施形態では、サイズ選択方法を用いて、試料を、DNAのより短い鎖；例えば、約800bp未満、約500bp未満、または約300bp未満の鎖について富化することができる。次いで、短い断片の増幅を通常通り進める。

【0254】

本開示に記載のmini-PCR法により、単一反応において、単一の試料から数百～数千、またはさらに数百万もの遺伝子座の高度に多重化された増幅および分析が可能になる。同時に、増幅されたDNAの検出を多重化することができ、バーコーディングPCRを用いることによって、数十～数百の試料を1つの配列決定レーンにおいて多重化することができる。この多重化検出は、最大49プレックスまで首尾よく試験されており、はるかに高い程度の多重化が可能である。事実上、これにより、単回の配列決定の実行で、数百の試料について数千ものSNPにおける遺伝子型決定することが可能になる。これらの試料について、該方法により、遺伝子型およびヘテロ接合性率を決定すること、また、同時に、コピー数を決定することが可能になり、これらはどちらも異数性を検出するために用いることができる。この方法は、母系の血漿中に見いだされる浮動性DNAから妊娠中の胎児の異数性を検出することにおいて特に有用である。この方法は、胎児の性判別をし、かつ／または胎児の父系性を予測するための方法の一部として用いることができる。この方法は、変異量決定ための方法の一部として用いることができる。この方法は、任意の量のDNAまたはRNAに対して用いることができ、標的の領域はSNP、他の多型領域、非多型領域、およびそれらの組み合わせであってよい。

【0255】

いくつかの実施形態では、断片化されたDNAのライゲーション媒介性ユニバーサルPCR増幅を用いることができる。ライゲーション媒介性ユニバーサルPCR増幅を用いて、血漿DNAを増幅することができ、次いで、それを複数の並行した反応へと分割することができる。ライゲーション媒介性ユニバーサルPCR増幅は、短い断片を優先的に増幅し、それにより、胎児の割合を富化するためにも用いることができる。いくつかの実施形態では、ライゲーションによって断片にタグを付加することにより、より短い断片を検出すること、プライマーのより短い標的配列特異的部分を使用すること、および／または非特異的な反応を減少させるより高温でアニーリングすることを可能にし得る。

【0256】

本明細書に記載の方法は、ある量の混入DNAと混在している標的DNAの集合が存在する場合の、いくつかの目的のために用いることができる。いくつかの実施形態では、標的DNAおよび混入DNAは、遺伝的に関連する個体由来であってよい。例えば、胎児（標的）における遺伝子の異常は、胎児（標的）のDNAおよび同様に母系の（混入）DNAを含有する母系の血漿から検出することができ、異常としては、全染色体異常（例えば、異数性）、部分的な染色体異常（例えば、欠失、重複、逆位、転座）、ポリヌクレオチ

ド多型（例えば、STR）、一塩基多型、および／または他の遺伝子の異常または差異が挙げられる。いくつかの実施形態では、標的DNAおよび混入DNAは、同じ個体由来であってよいが、例えば、がんの場合には、標的DNAと混入DNAが1つまたは複数の変異によって異なる（例えば、H. Mamónら *Preferential Amplification of Apoptotic DNA from Plasma: Potential for Enhancing Detection of Minor DNA Alterations in Circulating DNA*, *Clinical Chemistry* 54巻：9号（2008年）を参照されたい）。いくつかの実施形態では、DNAは、細胞培養物（アポトーシス性）上清に見いだすことができる。いくつかの実施形態では、その後のライブラリーの調製、増幅および／または配列決定のために、生体試料（例えば、血液）においてアポトーシスを誘導することが可能である。この目的を実現するためのいくつかの可能となるワークフローおよびプロトコールが本開示の他の箇所を示されている。

10

【0257】

いくつかの実施形態では、標的DNAは、単一細胞、標的ゲノムの1つのコピー未満からなるDNAの試料、低量のDNA、混合起源のDNA（例えば、妊娠血漿：胎盤性DNAおよび母系DNA；がん患者の血漿および腫瘍：健康なDNAとがんDNAの混合物、移植片など）、他の体液、細胞培養物、培養物の上清、DNAの法医学的試料、DNAの古代試料（例えば、琥珀中に閉じ込められた昆虫）、他のDNAの試料、およびそれらの組み合わせを起源としてよい。

20

【0258】

いくつかの実施形態では、短いアンプリコンサイズを用いることができる。短いアンプリコンサイズは、断片化されたDNAに特に適している（例えば、A. Sikoraら *Detection of increased amounts of cell-free fetal DNA with short PCR amplicons*, *Clin Chem*, 2010年1月；56巻（1号）：136～8頁を参照されたい）。

【0259】

短いアンプリコンサイズを用いることにより、いくつかの重要な利益がもたらされ得る。短いアンプリコンサイズにより、最適化された増幅効率もたらされ得る。短いアンプリコンサイズにより、一般には、より短い産物が生じ、したがって、非特異的なプライミングの見込みは少ない。より短い産物は、クラスターがより小さくなるほど、配列決定フローセル上により高密度にクラスター化することができる。本明細書に記載の方法は、より長いPCRアンプリコンに対して同等に良好に機能し得ることに留意されたい。アンプリコンの長さは、必要であれば、例えば、より大きな配列の範囲について配列決定する場合には増大させることができる。ネステッドPCRプロトコールの第1のステップとして100bp～200bp長のアッセイを伴う146プレックス標的化増幅を用いた実験を、単一細胞およびゲノムDNAに対して行い、陽性の結果を得た。

30

【0260】

いくつかの実施形態では、本明細書に記載の方法を用いて、SNP、コピー数、ヌクレオチドのメチル化、mRNAのレベル、他の種類のRNAの発現レベル、他の遺伝子の形体および／または後成的な形体を増幅し、かつ／または検出することができる。本明細書に記載のmini-PCR法は、次世代配列決定と一緒に用いることができ、該方法は、他の下流の方法、例えば、マイクロアレイ、デジタルPCRによる計数、リアルタイムPCR、質量分析などと一緒に用いることができる。

40

【0261】

いくつかの実施形態では、本明細書に記載のmini-PCR増幅方法は、少数集団を正確に定量化するための方法の一部として用いることができる。該方法は、スパイク較正物質を使用した絶対的定量化のために用いることができる。該方法は、超ディープシーケンシングによる、変異／微量な対立遺伝子の定量化のために用いることができ、高度に多

50

重化された様式で実行することができる。該方法は、ヒト、動物、植物または他の生き物における近親者または祖先の標準の父系性および同一性検査のために用いることができる。該方法は、法医学的試験のために用いることができる。該方法は、任意の種類の材料、例えば、羊水およびCVS、精子、受胎産物（POC）に対する迅速な遺伝子型決定およびコピー数解析（CN）のために用いることができる。該方法は、胚からの生検試料に対する遺伝子型決定などの単一細胞分析のために用いることができる。該方法は、min-PCRを用いた標的化配列決定による迅速な胚分析（生検から1日未満、1日、または2日以内）のために用いることができる。

【0262】

いくつかの実施形態では、該方法を、腫瘍分析のために用いることができる：腫瘍生検材料は、多くの場合、健康細胞と腫瘍細胞の混合物である。標的化PCRにより、近くにバックグラウンド配列がないSNPおよび遺伝子座のディープシーケンシングが可能になる。該方法は、腫瘍DNAに対するコピー数およびヘテロ接合性の損失の分析のために用いることができる。前記腫瘍DNAは、腫瘍患者の多くの異なる体液または組織に存在し得る。該方法は、腫瘍の再発の検出および／または腫瘍スクリーニングのために用いることができる。該方法は、種子の品質管理試験のために用いることができる。該方法は、育種または漁業のために用いることができる。これらの方法はいずれも、倍数性呼び出しのための非多型の遺伝子座の標的化に同等に良好に用いることができることに留意されたい。

【0263】

本明細書に開示されている方法の基礎をなす基本的な方法のいくつかが記載されているいくつかの文献としては以下が挙げられる（1）Wang HY、Luo M、Tereshchenko IV、Frikker DM、Cui X、Li JY、Hu G、Chu Y、Azaro MA、Lin Y、Shen L、Yang Q、Kambooris ME、Gao R、Shih W、Li H. Genome Res. 2005年2月；15巻（2号）：276～83頁。Department of Molecular Genetics、Microbiology and Immunology/The Cancer Institute of New Jersey、Robert Wood Johnson Medical School、New Brunswick、New Jersey 08903、USA。（2）High-throughput genotyping of single nucleotide polymorphisms with high sensitivity. Li H、Wang HY、Cui X、Luo M、Hu G、Greenawalt DM、Tereshchenko IV、Li JY、Chu Y、Gao R. Methods Mol Biol. 2007年；396頁— PubMed PMID:18025699。（3）配列決定のために平均9つのアッセイの多重化を含む方法は、Nested Patch PCR enables highly multiplexed mutation discovery in candidate genes. Varley KE、Mitra RD. Genome Res. 2008年11月；18巻（11号）：1844～50頁、Epub 2008年10月10日に記載されている。本明細書に開示されている方法により、上記の参考文献におけるものよりも桁の大きい多重化が可能になることに留意されたい。

【0264】

プライマーの設計

高度多重PCRにより、多くの場合、プライマー二量体形成などの非生産的な副反応がもたらす産物DNAが非常に高い割合で産生され得る。ある実施形態では、非生産的な副反応を引き起こす可能性が最も高い特定のプライマーを、プライマーライブラリーから除去して、ゲノムにマッピングされる増幅されたDNAを高い割合でもたらすプライマーライブラリーを得ることができる。問題のあるプライマー、すなわち、特に二量体を安定させる可能性があるプライマーを除去するステップにより、予想外に、その後の配列決定による分析のための非常に高いPCR多重化レベルが可能になった。プライマー二量体およ

び／または他の悪影響を及ぼす産物によって性能が著しく低下する配列決定などの系では、他に記載されている多重化よりも10倍超、50倍超、および100倍超高度な多重化が実現された。これは、過剰なプライマー二量体が感知できるほど結果に影響を及ぼさないプローブに基づく検出方法、例えば、マイクロアレイ、TaqMan、PCRとは対照的であることに留意されたい。当技術分野における一般的な考えでは、配列決定するための多重化PCRは、同じウェルでは約100アッセイに限られることにも留意されたい。例えば、FluidigmおよびRainDanceは、1つの試料について並行した反応で48または1000のPCRアッセイを実施するためのプラットフォームを提供する。

【0265】

非マッピングプライマー二量体または他の悪影響を及ぼすプライマー産物の量を最小限にしたライブラリーのためのプライマーを選択するためのいくつかのやり方が存在する。経験的なデータにより、少数の「悪い」プライマーは大量の非マッピングプライマー二量体副反応に関与することが示されている。これらの「悪い」プライマーを除去することにより、標的の遺伝子座に対して位置を決める配列読み取りのパーセントを上昇させることができる。「悪い」プライマーを同定するための1つのやり方は、標的化増幅によって増幅されたDNAの配列決定データを調べることであり、最大の頻度で認められるプライマー二量体を除去して、ゲノムにマッピングされない副産物DNAをもたらす可能性が有意に低いプライマーライブラリーを生じることができる。種々のプライマーの組み合わせの結合エネルギーを算出することができる公的に入手可能なプログラムも存在し、結合エネルギーが最も高いプライマーの組み合わせを除去することにより、同様に、ゲノムにマッピングされない副産物DNAをもたらす可能性が有意に低いプライマーライブラリーが生じる。

【0266】

多数のプライマーを多重化することにより、含めることができるアッセイにかなりの制約が課される。意図せずに相互作用するアッセイにより、偽の増幅産物がもたらされる。miniPCRのサイズの制約により、さらなる制約がもたらされ得る。ある実施形態では、非常に多数の潜在的なSNP標的（約500から100万超の間）で開始し、各SNPを増幅するためのプライマーを設計することを試みるのが可能である。プライマーを設計することができる場合、可能性のあるプライマーの対の全ての間で偽のプライマー2重鎖が形成される尤度を、DNA2重鎖形成についての公開された熱力学的なパラメータを使用して評価することによって、偽の産物を形成する可能性があるプライマー対を同定することを試みるのが可能である。プライマー相互作用は、相互作用に関連するスコア関数によって順位づけ、所望のプライマーの数に見合うまで相互作用スコアが最も悪いプライマーを排除することができる。ヘテロ接合性の可能性があるSNPが最も有用である場合には、同様にアッセイの一覧を順位付け、最もヘテロ接合性に適合するアッセイを選択することが可能である。実験により、相互作用スコアが高いプライマーが、プライマー二量体を形成する可能性が最も高いことが検証された。高度な多重化においては、全ての偽の相互作用を排除することは不可能であるが、in silicoで相互作用スコアが最も高いプライマーまたはプライマー対は、全体の反応の優位を占め、意図された標的からの増幅を著しく限定するので、これらを除去することが必須である。この手順を実施して、10,000プライマーに至る多重化プライマー集合を作製した。この手順による改善は、実質的なものであり、全てのPCR産物の配列決定によって決定された通り、最も悪いプライマーを除去しなかった反応からの10%と比較して、標的産物の80%超、90%超、95%超、98%超、およびさらには99%超の増幅を可能にする。以前に記載されている部分的なセミネステッド手法と組み合わせると、90%超、および95%超までのアンプリコンを標的の配列にマッピングすることができる。

【0267】

どのPCRプローブが二量体を形成する可能性があるかを決定するための他の方法が存在することに留意されたい。ある実施形態では、最適化されていないプライマーの集合を

10

20

30

40

50

使用して増幅したDNAのプールの分析が、問題のあるプライマーを決定するために十分であり得る。例えば、分析は配列決定を用いて行うことができ、最大の数で存在する二量体を、二量体を形成する可能性が最も高いものであると決定し、除去することができる。

【0268】

この方法には、いくつかの潜在的な適用、例えば、SNP遺伝子型決定、ヘテロ接合性率決定、コピー数測定、および他の標的化配列決定への適用がある。ある実施形態では、プライマーを設計する方法を、本文書の他の箇所に記載されているmini-PCR法と組み合わせて用いることができる。いくつかの実施形態では、プライマーの設計方法を、大規模な多重PCR法の一部として用いることができる。

【0269】

プライマーにタグを使用することにより、プライマー二量体産物の増幅および配列決定を減少させることができる。タグプライマーを用いて、必要な標的特異的配列を20塩基対未満、15塩基対未満、12塩基対未満、さらには10塩基対未満までに短縮することができる。これは、プライマー結合部位内で標的配列が断片化される場合に標準のプライマーの設計に伴って偶然発見され得る、または、または、プライマー設計へと企画することができる。この方法の利点としては、特定の最大のアンプリコンの長さに対して設計することができるアッセイの数が増加すること、および「情報価値のない」プライマー配列の配列決定が短縮されることが挙げられる。該方法は、内部のタグ付けと組み合わせて用いることもできる（本文書の他の箇所を参照されたい）。

【0270】

ある実施形態では、多重標的化PCR増幅における非生産的な産物の相対量を、アニーリング温度を上昇させることによって減少させることができる。標的特異的プライマーと同じタグを有するライブラリーを増幅する場合には、タグがプライマー結合に寄与するので、アニーリング温度をゲノムDNAと比較して増大させることができる。いくつかの実施形態では、以前報告されたものよりも相当低いプライマー濃度を用い、それと一緒に、他の箇所で報告されているものよりも長いアニーリング時間を用いる。いくつかの実施形態では、アニーリング時間は、10分超、20分超、30分超、60分超、120分超、240分超、480分超、およびさらには960分超であってよい。ある実施形態では、以前の報告よりも長いアニーリング時間を用い、これにより、より低いプライマー濃度が可能になる。いくつかの実施形態では、プライマー濃度は、50 nM、20 nM、10 nM、5 nM、1 nM、および1 μM未満までの低さである。驚いたことに、これにより、高度に多重化された反応、例えば、1,000プレックス反応、2,000プレックス反応、5,000プレックス反応、10,000プレックス反応、20,000プレックス反応、50,000プレックス反応、およびさらには100,000プレックス反応に対して頑強な性能がもたらされる。ある実施形態では、増幅には、長いアニーリング時間で実行する1サイクル、2サイクル、3サイクル、4サイクルまたは5サイクルを用い、その後、タグを付けたプライマーを用いて通常のアニーリング回数より多いPCRサイクルを行う。

【0271】

標的場所を選択するために、候補となるプライマー対設計物のプールを用いて着手し、プライマー対間の潜在的に有害な相互作用の熱力学的モデルを作製し、次いで、このモデルを用いてプール内の他の設計物と適合しない設計物を排除することができる。

【0272】

標的化PCRの変形物—ネスティング

PCRを行う場合に可能である多くのワークフローが存在し、本明細書に開示されている方法に典型的ないくつかのワークフローが記載されている。本明細書において概説されているステップは、他の可能性のあるステップを排除することを意図しておらず、かつ、方法が適正に機能するために本明細書に記載のステップいずれかが必要であることも意味しない。多数のパラメータの変形または他の改変が文献において公知であり、本発明の核心に影響を及ぼすことなく行うことができる。1つの特定の一般的なワークフローが下に

示され、その後にくつもの可能性のある変形物 (variant) が続く。変形物とは、一般には、可能性のある二次PCR反応、例えば、行うことができる異なる種類のネスティング (ステップ3) を指す。変形物は、本明細書に明確に記載されているものと違う時間において、または異なる順序で行うことができることに留意することが重要である。

【0273】

1. 試料中のDNAには、多くの場合ライブラリータグまたはライゲーションアダプタタグ (LT) と称されるライゲーションアダプタを付加することができ、ライゲーションアダプタはユニバーサルプライミング配列を含有し、その後ユニバーサル増幅が続く。ある実施形態では、これは、断片化後に、配列決定ライブラリーを作製するために設計された標準のプロトコールを使用して行うことができる。ある実施形態では、DNA試料を平滑末端化し、次いで、Aを3'末端に付加することができる。T-オーバーハングを有するY-アダプタを付加し、ライゲーションすることができる。いくつかの実施形態では、AまたはTオーバーハング以外の他の粘着末端を使用することができる。いくつかの実施形態では、他のアダプタ、例えば、ループライゲーションアダプタを付加することができる。いくつかの実施形態では、アダプタは、PCR増幅のために設計されたタグを有してよい。

【0274】

2. 特異的標的増幅 (STA) : 数百、数千、数万、さらには数十万もの標的を1回の反応において前増幅で多重化することができる。STAは、一般には、10~30サイクル実行されるが、5~40サイクル、2~50サイクル、およびさらには1~100サイクル実行することができる。例えば、より単純なワークフローのため、または大部分の二量体の配列決定を回避するために、プライマーに尾部を付けることができる。一般には、同じタグを保有する両方のプライマーの二量体は効率的に増幅または配列決定されないことに留意されたい。いくつかの実施形態では、1サイクルから10サイクルの間のPCRを行うことができ、いくつかの実施形態では、10サイクルから20サイクルの間のPCRを行うことができ、いくつかの実施形態では、20サイクルから30サイクルの間のPCRを行うことができ、いくつかの実施形態では、30サイクルから40サイクルの間のPCRを行うことができ、いくつかの実施形態では、40サイクル超のPCRを行うことができる。増幅は、線形増幅であってよい。PCRサイクルの数を最適化して、最適な読み取りの深さ (DOR) プロファイルをもたらすことができる。異なるDORプロファイルは異なる目的のために望ましい場合がある。いくつかの実施形態では、全てのアッセイ間の読み取りのより均一な分布が望ましい；いくつかのアッセイについてDORが非常に小さい場合、データが非常に有用であるためには確率論的ノイズが高すぎる可能性があるが、読み取りの深さが非常に高い場合、各追加の読み取りの限界有用性は比較的小さい。

【0275】

プライマー尾部により、普遍的にタグを付けたライブラリーからの断片化されたDNAの検出を改善することができる。ライブラリータグおよびプライマー尾部が相同な配列を含有する場合、ハイブリダイゼーションを改善することができ (例えば、融解温度 (T_M) を下げる)、プライマー標的配列の一部が試料のDNA断片内にある場合にのみ、プライマーを伸長することができる。いくつかの実施形態では、13以上の標的的特異的塩基対を用いることができる。いくつかの実施形態では、10~12の標的的特異的塩基対を用いることができる。いくつかの実施形態では、8~9の標的的特異的塩基対を用いることができる。いくつかの実施形態では、6~7の標的的特異的塩基対を用いることができる。いくつかの実施形態では、STAは、前増幅されたDNA、例えば、MDA、RCA、他の全ゲノム増幅またはアダプター媒介性ユニバーサルPCRに対して実施することができる。いくつかの実施形態では、STAは、例えば、サイズ選択、標的捕捉、指向性分解によって特定の配列および集団が富化された、または枯渇した試料に対して実施することができる。

【0276】

3. いくつかの実施形態では、二次的な多重PCRまたはプライマー伸長反応を実施し

て、特異性を増大させ、望ましくない産物を減少させることが可能である、例えば、完全なネスティング、セミネスティング、ヘミネスティング、および／またはより小さなアッセイプールの並行した反応への細分化は、全て、特異性を増大させるために用いることができる技法である。実験により、試料を3回の400プレックス反応に分割することにより、正確に同じプライマーを用いた1回の1,200プレックス反応よりも高い特異性で産物DNAがもたらされることが示された。同様に、実験により、試料を4回の2,400プレックス反応に分割することにより、正確に同じプライマーを用いた1回の9,600プレックス反応よりも高い特異性で産物DNAがもたらされることが示された。ある実施形態では、同じ方向性および反対の方向性の標的特異的プライマーおよびタグ特異的プライマーを用いることが可能である。

10

【0277】

4. いくつかの実施形態では、STA反応によって産生されるDNA試料（希釈、精製またはその他）をタグ特異的プライマーおよび「ユニバーサル増幅」を用いて増幅すること、すなわち、前増幅し、タグを付けた標的の多くまたは全てを増幅することが可能である。プライマーは、ハイスループット配列決定プラットフォームにおける配列決定に必要な追加の機能的な配列、例えば、バーコードまたは完全なアダプタ配列を含有してよい。

【0278】

これらの方法は、任意のDNAの試料を分析するために用いることができ、DNAの試料が特に少ない場合、または、それが、DNAが2つ以上の個体を起源とするDNAの試料である場合、例えば、母系の血漿の場合に特に有用である。これらの方法は、単一または少数の細胞、ゲノムDNA、血漿DNA、増幅された血漿ライブラリー、増幅されたアポトシス性の上清ライブラリーまたは他の混合DNAの試料などのDNA試料に対して用いることができる。ある実施形態では、これらの方法は、遺伝子の構成が異なる細胞が、単一の個体に存在する可能性がある場合、例えば、がんまたは移植片に用いることができる。

20

【0279】

プロトコールの変形物（上記のワークフローに対する変形物および／または追加物）

直接多重mini-PCR：タグを付けたプライマーを用いた複数の標的配列の特異的標的増幅（STA）が図1に示されている。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを示す。103は、PCRプライマーがハイブリダイズした、ユニバーサル増幅された一本鎖DNAを示す。104は、最終のPCR産物を示す。いくつかの実施形態では、STAは、100超、200超、500超、1,000超、2,000超、5,000超、10,000超、20,000超、50,000超、100,000超、または200,000超の標的に対して行うことができる。その後の反応において、タグ特異的プライマーにより全ての標的配列を増幅し、サンプリングインデックスを含めた、配列決定するために必要な全ての配列を含むタグを伸長する。ある実施形態では、プライマーにタグ付けしなくてよい、または特定のプライマーのみにタグを付けてよい。配列決定アダプタは、従来のアダプタライゲーションによって付加することができる。ある実施形態では、最初のプライマーはタグを担持してよい。

30

40

【0280】

ある実施形態では、プライマーを、増幅されるDNAの長さが予想外に短くなるように設計する。先行技術により、当業者が一般には、100+bpのアンプリコンを設計することが実証されている。ある実施形態では、アンプリコンを、80bp未満になるように設計することができる。ある実施形態では、アンプリコンを、70bp未満になるように設計することができる。ある実施形態では、アンプリコンを、60bp未満になるように設計することができる。ある実施形態では、アンプリコンを、50bp未満になるように設計することができる。ある実施形態では、アンプリコンを、45bp未満になるように設計することができる。ある実施形態では、アンプリコンを、40bp未満になるように設計することができる。ある実施形態では、アンプリコンを、35bp未満になるように

50

設計することができる。ある実施形態では、アンプリコンを、40bpから65bpの間になるように設計することができる。

【0281】

実験を、このプロトコルを使用して、1200プレックス増幅を用いて実施した。ゲノムDNAと妊娠血漿の両方を使用した；配列読み取りの約70%が標的の配列にマッピングされた。詳細は本文書の他の箇所に示されている。アッセイの設計および選択を伴わない1042プレックスの配列決定により、配列の>99%がプライマー二量体産物となった。

【0282】

逐次的なPCR：STA1の後、産物の複数の一定分量を、同じプライマーを有する複雑さが低下したプールを用いて並行して増幅することができる。第1の増幅により、分割するために十分な材料が生じ得る。この方法は、少ない試料、例えば、約6~100pg、約100pg~1ng、約1ng~10ngまたは約10ng~100ngの試料に対して特に優良である。1200プレックスを3回の400プレックスにしたプロトコルを実施した。配列決定の読み取りのマッピングは、1200プレックス単独における約60~70%から95%超まで増大した。

【0283】

セミネステッドmini-PCR：(図2参照) STA1の後、内側のネステッドフォワードプライマー(103B、105b)の多重のセットおよび1つ(または少数)のタグ特異的リバースプライマー(103A)で構成される第2のSTAを実施する。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを示す。103は、フォワードプライマーBおよびリバースプライマーAがハイブリダイズした、ユニバーサル増幅された一本鎖DNAを示す。104は、103からのPCR産物を示す。105は、ハイブリダイズしたネステッドフォワードプライマーb、および既に103と104の間に生じたPCRからの分子の一部であるリバースタグAを有する104からの産物を示す。106は、最終のPCR産物を示す。このワークフローを用いると、通常、配列の95%超が意図された標的にマッピングされる。ネステッドプライマーは外側のフォワードプライマー配列とオーバーラップしてよいが、追加の3'末端塩基を導入する。いくつかの実施形態では、1から20個の間の余分の3'塩基を用いることが可能である。実験により、1200プレックス設計物において9個以上の余分の3'塩基を用いると良好に機能することが示された。

【0284】

完全ネステッドmini-PCR：(図3参照) STAステップ1の後、第2の多重PCR(または複雑さが低下した並行のm.p. PCR)を、タグ(A、a、B、b)を保有する2つのネステッドプライマーを用いて実施することが可能である。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを示す。103は、フォワードプライマーBおよびリバースプライマーAがハイブリダイズした、ユニバーサル増幅された一本鎖DNAを示す。104は、103からのPCR産物を示す。105は、ネステッドフォワードプライマーbおよびネステッドリバースプライマーaがハイブリダイズした、104からの産物を示す。106は、最終のPCR産物を示す。いくつかの実施形態では、2つのプライマーの完全なセットを用いることが可能である。完全ネステッドmini-PCRプロトコルを使用した実験を用いて、単一細胞および3つの細胞に対して、ユニバーサルライゲーションアダプタを付加し、増幅するステップ102を伴わずに146プレックス増幅を実施した。

【0285】

ヘミネステッドmini-PCR：(図4参照) 断片の末端にアダプタを有する標的DNAを用いることが可能である。フォワードプライマー(B)の多重セットおよび1つ(または少数)のタグ特異的リバースプライマー(A)で構成されるSTAを実施する。第

2のSTAを、ユニバーサルタグ特異的フォワードプライマーおよび標的特異的リバープライマーを使用して実施することができる。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを示す。103は、リバープライマーAがハイブリダイズした、ユニバーサル増幅された一本鎖DNAを示す。104は、リバープライマーAおよびライゲーションアダプタタグプライマーLTを使用して増幅した、103からのPCR産物を示す。105は、フォワードプライマーBがハイブリダイズした、104からの産物を示す。106は、最終のPCR産物を示す。このワークフローでは、標的特異的フォワードプライマーおよびリバープライマーを別々の反応において使用し、それにより、反応の複雑さが減少し、フォワードプライマーとリバープライマーの二量体形成が防がれる。この例では、プライマーAおよびBを、第1のプライマーとみなすことができ、プライマー「a」および「b」を、内側のプライマーとみなすことができることに留意されたい。この方法は、直接PCRと同等に優良であるが、プライマー二量体を回避するので、直接PCRに対する大きな改善である。第1ラウンドのヘミネステッドプロトコルの後、一般には、約99%の非標的DNAが認められるが、第2ラウンドの後には一般には、大きく改善される。

10

【0286】

三重ヘミネステッドmini-PCR：(図5参照)断片の末端にアダプタを有する標的DNAを用いることが可能である。フォワードプライマー(B)の多重セットおよび1つ(または少数)のタグ特異的なリバープライマー(A)および(a)で構成されるSTAを実施する。第2のSTAを、ユニバーサルタグ特異的フォワードプライマーおよび標的特異的リバープライマーを使用して実施することができる。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを示す。103は、リバープライマーAがハイブリダイズした、ユニバーサル増幅された一本鎖DNAを示す。104は、リバープライマーAおよびライゲーションアダプタタグプライマーLTを使用して増幅した、103からのPCR産物を示す。105は、フォワードプライマーBがハイブリダイズした、104からの産物を示す。106は、リバープライマーAおよびフォワードプライマーBを使用して増幅した、105からのPCR産物を示す。107は、リバープライマー「a」がハイブリダイズした、106からの産物を示す。108は、最終のPCR産物を示す。この例では、プライマー「a」およびBを、内側のプライマーとみなすことができ、Aを、第1のプライマーとみなすことができることに留意されたい。必要に応じて、AとBの両方を第1のプライマーとみなすことができ、「a」を、内側のプライマーとみなすことができる。リバープライマーおよびフォワードプライマーの名称は切り換えることができる。このワークフローでは、標的特異的フォワードプライマーおよびリバープライマーを別々の反応において使用し、それにより、反応の複雑さが減少し、フォワードプライマーとリバープライマーの二量体形成が防がれる。この方法は、直接PCRと同等に優良であるが、プライマー二量体を回避するので、直接PCRに対する大きな改善である。第1ラウンドのヘミネステッドプロトコルの後、一般には、約99%の非標的DNAが認められるが、第2ラウンドの後には一般には、大きく改善される。

20

30

40

【0287】

片側ネステッドmini-PCR：(図6参照)断片の末端にアダプタを有する標的DNAを用いることが可能である。STAを、ネステッドフォワードプライマーの多重セットを用い、リバープライマーとしてライゲーションアダプタタグを使用して実施することもできる。次いで、ネステッドフォワードプライマーおよびユニバーサルリバープライマーのセットを使用して第2のSTAを実施することができる。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを示す。103は、フォワードプライマーAがハイブリダイズした、ユニバーサル増幅された一本鎖DNAを示す。104は、フォワードプライマーAおよびライゲーションアダプタタグリバープライマーLTを使用して増

50

幅した、103からのPCR産物を示す。105は、ネステッドフォワードプライマーがハイブリダイズした、104からの産物を示す。106は、最終のPCR産物を指示す。この方法では、第1のSTAおよび第2のSTAにおいてオーバーラップしているプライマーを使用することにより、標準のPCRによるよりも短い標的配列を検出することができる。該方法は、一般には、既に上記のSTAステップ1ーユニバーサルタグの付加および増幅を受けたDNAの試料を差し引いて実施し、2つのネステッドプライマーは一方の側にのみあり、他方の側にはライブラリータグを使用する。該方法を、アポトーシス性の上清および妊娠血漿のライブラリーに対して実施した。このワークフローを用いると、配列の約60%が意図された標的にマッピングされた。リバースアダプタ配列を含有した読み取りはマッピングしておらず、したがって、リバースアダプタ配列を含有する読み取りをマッピングした場合にはこの数字は大きくなることが予想されることに留意されたい。

10

【0288】

片側のみのmini-PCR：断片の末端にアダプタを有する標的DNAを用いることが可能である（図7参照）。STAを、フォワードプライマーの多重セットおよび1つ（または少数）のタグ特異的なリバースプライマーを用いて実施することができる。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを指示す。103は、フォワードプライマーAがハイブリダイズした一本鎖DNAを示す。104は、フォワードプライマーAおよびライゲーションアダプタタグリバースプライマーLTを使用して増幅した、103からのPCR産物を示し、これは最終のPCR産物である。この方法により、標準のPCRによるよりも短い標的配列を検出することができる。しかし、ただ1つの標的特定のプライマーを使用するので、比較的非特異的であり得る。このプロトコルの有効性は片側ネステッドmini-PCRの半分である。

20

【0289】

リバースセミネステッドmini-PCR：断片の末端にアダプタを有する標的DNAを用いることが可能である（図8参照）。STAを、フォワードプライマーの多重セットおよび1つ（または少数）のタグ特異的なリバースプライマーを用いて実施することができる。101は、Xに対象の多型遺伝子座を有する二本鎖DNAを示す。102は、ユニバーサル増幅のためにライゲーションアダプタを付加した二本鎖DNAを示す。103は、リバースプライマーBがハイブリダイズした一本鎖DNAを示す。104は、リバースプライマーBおよびライゲーションアダプタタグフォワードプライマーLTを使用して増幅した、103からのPCR産物を示す。105は、フォワードプライマーA、および内側のリバースプライマー「b」がハイブリダイズした、PCR産物104を示す。106は、フォワードプライマーAおよびリバースプライマー「b」を使用して105から増幅されたPCR産物を示し、これは最終のPCR産物である。この方法により、標準のPCRによるよりも短い標的配列を検出することができる。

30

【0290】

上記の方法の単に反復または組み合わせであるさらなる変形物、例えば、プライマーの3つのセットを使用する二重ネステッドPCRも存在し得る。別の変形物は片側半ネステッドmini-PCRであり、STAをネステッドフォワードプライマーの多重セットおよび1つ（または少数）のタグ特異的なリバースプライマーを用いて実施することもできる。

40

【0291】

これらの変形物の全てにおいて、フォワードプライマーおよびリバースプライマーの同一性は交換することができることに留意されたい。いくつかの実施形態では、ネステッド変形物は、アダプタタグを付加すること、およびユニバーサル増幅ステップを含む最初のライブラリーの調製を伴わずに同等に良好に実行することができることに留意されたい。いくつかの実施形態では、追加のフォワードプライマーおよび／またはリバースプライマーおよび増幅ステップを伴ってPCRの追加のラウンドを含めることができ、これらの追加のステップは、標的の遺伝子座に対応するDNA分子のパーセントをさらに増大させる

50

ことが望ましい場合に特に有用であり得ることに留意されたい。

【0292】

ネスティングワークフロー

異なる程度のネスティング、および異なる程度の多重化を伴って増幅を実施するための多くのやり方が存在する。図9では、フローチャートが、可能性のあるワークフローのいくつかと共に示されている。10, 000プレックスPCRの使用は単なる例であり、これらのフローチャートは他の多重化の程度に対しても同等に良好に機能することに留意されたい。

【0293】

ループライゲーションアダプタ

例えば、配列決定するためのライブラリーを作出するためにユニバーサルタグを付けたアダプタを付加する場合、アダプタをライゲーションするためのいくつかのやり方が存在する。1つのやり方は、試料DNAを平滑末端化し、A-テーリングを実施し、T-オーバーハングを有するアダプタとライゲーションすることである。アダプタをライゲーションするための、いくつかの他のやり方が存在する。ライゲーションすることができるアダプタもいくつか存在する。例えば、DNAの2つの鎖からなり、一方の鎖が二本鎖領域、およびフォワードプライマー領域によって指定される領域を有し、他方の鎖が第1の鎖上の二本鎖領域と相補的な二本鎖領域、およびリバースプライマーを伴う領域によって指定されるY-アダプタを使用することができる。アニーリングする場合、二本鎖領域は、A-オーバーハングを有する二本鎖DNAとライゲーションするために、T-オーバーハングを含有してよい。

【0294】

ある実施形態では、アダプタは、末端領域が相補的であって、フォワードプライマーでタグを付けた領域(LFT)、リバースプライマーでタグを付けた領域(LRT)、およびその2つの間の切断部位を含有する、DNAのループであってよい(図10参照)。101は、二本鎖の平滑末端の標的DNAを指す。102は、A尾部をもつ標的DNAを指す。103は、T-オーバーハング「T」および切断部位「Z」を有するループライゲーションアダプタを指す。104は、ループライゲーションアダプタが付加された標的DNAを指す。105は、切断部位において切断された、ライゲーションアダプタが付加された標的DNAを指す。LFTはライゲーションアダプタフォワードタグを指し、LRTはライゲーションアダプタリバースタグを指す。相補的な領域はT-オーバーハング、または標的DNAとライゲーションするために使用することができる他の形体で終わってよい。切断部位は、UNGに沿った切断のための一連のウラシルであり得るか、あるいは制限酵素もしくは他の切断方法または単に基本的な増幅によって認識され切断され得る配列であり得る。これらのアダプタは、例えば、配列決定するための任意のライブラリーを調製するために使用することができる。これらのアダプタは、本明細書に記載の他の方法のいずれか、例えば、mini-PCR増幅方法と組み合わせて用いることができる。

【0295】

内部にタグを付けたプライマー

所与の多型遺伝子座に存在する対立遺伝子を決定するために配列決定を用いる場合、配列読み取りは、一般には、プライマー結合部位(a)の上流で開始され、次いで、多型部位(X)が読まれる。タグは一般には、図11の左側に示されている通り配置される。101は、対象の多型遺伝子座「X」およびタグ「b」が付加されたプライマー「a」を有する一本鎖標的DNAを指す。非特異的なハイブリダイゼーションを回避するために、プライマー結合部位(「a」と相補的な標的DNAの領域)は、一般には、18~30bpの長さである。配列タグ「b」は、一般には約20bpであり、理論上は、これらは約15bpより長い任意の長さであってよいが、多くの人々は配列決定プラットフォームの企業から販売されているプライマー配列を使用する。「a」と「X」の間の距離「d」は、対立遺伝子の偏りを回避するために少なくとも2bpであってよい。多重PCR増幅を、過剰なプライマー間相互作用を回避するために慎重なプライマーの設計が必要である、本

10

20

30

40

50

明細書に開示されている方法または他の方法を用いて実施する場合、許容できる「a」と「X」の間の距離「d」のウィンドウは、相当に変動し得る：2 b p ~ 1 0 b p、2 b p ~ 2 0 b p、2 b p ~ 3 0 b pまたは、さらには2 b p ~ 3 0 b p超。したがって、図11の左側に示されているプライマーの配置を用いる場合、配列読み取りは、多型遺伝子座を測定するために十分に長い読み取りを得るために、最小の4 0 b pでなければならず、また、「a」および「d」の長さに応じて配列読み取りは6 0 b pまたは7 5 b pまでが必要になる場合がある。通常、配列読み取りが長いほど、所与の数の読み取りについて配列決定するための費用および時間が増し、したがって、必要な読み取りの長さを最小化することにより、時間と金の両方を節約することができる。さらに、平均で、読み取りの初期の塩基の読み取りは、読み取り後期の読み取りよりも正確な読み取りであるので、必要な配列読み取りの長さを減らすことにより、多型領域の測定の正確度を上げることもできる。

10

【0296】

ある実施形態では、図11の103に示されている通り、内部にタグを付けたプライマーと称されるプライマー結合部位(a)を複数のセグメント(a'、a''、a'''...)に分割し、配列タグ(b)を、2つのプライマー結合部位の中央のDNAのセグメント上に置く。この配置により、シーケンサーがより短い配列読み取りを行うことが可能になる。ある実施形態では、a' + a''は少なくとも約18 b pであるべきであり、30 b p、40 b p、50 b p、60 b p、80 b p、100 b p、または100 b p超の長さであってよい。ある実施形態では、a''は少なくとも約6 b pであるべきであり、ある実施形態では、約8 b pから16 b pの間である。全ての他の因子も同等であり、内部にタグを付けたプライマーを使用することにより、必要な配列読み取りの長さを、少なくとも6 b p、8 b p、10 b p、12 b p、15 b pと同程度、さらには20 b pまたは30 b pと同程度に切り詰めることができる。この結果、かなりの金、時間および正確度の利点をもたらされ得る。内部にタグを付けたプライマーの例は図12に示されている。

20

【0297】

ライゲーションアダプタ結合領域を有するプライマー

断片化されたDNAに伴う1つの問題は、その長さが短いので、多型がDNA鎖の末端の近くにある見込みが長い鎖よりも高いことである(例えば、101、図10)。多型をPCRによって捕捉するためには、多型の両側に適切な長さのプライマー結合部位が必要であるので、プライマーと標的の結合部位の間のオーバーラップが不十分であることに起因して、標的の多型を有するかなりの数のDNAの鎖が捕捉し損なわれる。ある実施形態では、標的DNA101にはライゲーションアダプタ102を付加することができ、標的プライマー103は、設計された結合領域(a)の上流に付加したライゲーションアダプタタグ(lt)と相補的な領域(cr)を有し得る(図13参照)；したがって、結合領域(aと相補的な101の領域)が一般にハイブリダイゼーションのために必要な18 b pよりも短い場合には、ライブラリタグと相補的なプライマーの領域(cr)により、PCRが進行することができるまで結合エネルギーを増大させることができる。より短い結合領域に起因して失われる任意の特異性は、適切に長い標的結合領域を有する他のPCRプライマーによって補うことができることに留意されたい。この実施形態は、直接PCRまたは本明細書に記載の他の方法のいずれか、例えば、ネステッドPCR、セミネステッドPCR、ヘミネステッドPCR、片側ネステッドまたはセミネステッドまたはヘミネステッドPCRまたは他のPCRプロトコールと組み合わせて用いることができることに留意されたい。

30

40

【0298】

配列決定データを用い、種々の仮説について、観察された対立遺伝子データと予測される対立遺伝子分布を比較することを伴う分析的な方法と組み合わせて倍数性を決定する場合、読み取りの深さが低い対立遺伝子からの追加の読み取りのそれぞれにより、読み取りの深さが高い対立遺伝子からの読み取りよりも多くの情報をもたらされる。したがって、理想的には、各遺伝子座が同様の数の代表的な配列読み取りを有する均一な読み取りの深

50

さ（DOR）が認められることが望まれる。したがって、DORの分散を最小限にすることが望ましい。ある実施形態では、アニーリング時間を増加させることによってDORの変動係数（これは、DORの標準偏差／平均DORと定義することができる）を減少させることが可能である。いくつかの実施形態では、アニーリング温度は、2分超、4分超、10分超、30分超、および1時間超またはさらに長くてよい。アニーリングは平衡プロセスであるので、アニーリング時間の増加に伴うDORの分散の改善に限界はない。ある実施形態では、プライマー濃度を増加させることにより、DORの分散が減少する。

【0299】

診断ボックス

ある実施形態では、本開示は、本開示に記載の方法のいずれかを部分的にまたは完全に実行することができる診断ボックスを含む。ある実施形態では、診断ボックスは、診察室、病院の検査室または患者をケアする場所に合理的に近い任意の適切な場所に置かれてよい。ボックスは、方法全体を完全に自動化された様式で実行することを可能にし得る、またはボックスは、技師が手動で完了するための、1つまたはいくつものステップを必要とする場合がある。ある実施形態では、ボックスは、少なくとも母系の血漿について測定された遺伝子型データを解析することを可能にし得る。ある実施形態では、ボックスは、診断ボックスで測定された遺伝子型データを、次いで遺伝子型データを解析し、場合によっては報告の作製も行う外部の計算設備に伝達する手段と連結することができる。診断ボックスは、水性試料または液体試料を1つの容器から別の容器に移すことができるロボットユニットを含んでよい。診断ボックスは、固体と液体の両方のいくつかの試薬を含んでよい。診断ボックスは、ハイスループットシーケンサーを含んでよい。診断ボックスは、コンピュータを含んでよい。

【0300】

プライマーキット

いくつかの実施形態では、本開示に記載の方法を実現するために設計された複数のプライマーを含むキットを処方することができる。プライマーは、本明細書に開示されている外側のフォワードプライマーおよびリバースプライマー、内側のフォワードプライマーおよびリバースプライマーであってよく、プライマーの設計のセクションに開示されている通り、キット内の他のプライマーに対する結合親和性が低いように設計されたプライマーであってよく、関連するセクションに記載のとおりハイブリッド捕捉プローブまたは環状化前プローブであるかまたはそのいくつかの組合せであってよい。ある実施形態では、本明細書に開示されている方法で使用するために設計された、妊娠中の胎児における標的染色体の倍数性状態を決定するためのキットであって、複数の内側のフォワードプライマー、および必要に応じて複数の内側のリバースプライマー、および必要に応じて、外側のフォワードプライマーおよび外側のリバースプライマーであって、該プライマーのそれぞれが、標的染色体および必要に応じて、さらに別の染色体上の多型部位のうちの1つのすぐ上流および／または下流のDNAの領域とハイブリダイズするように設計されているプライマー、を含むキットを処方することができる。ある実施形態では、プライマーキットは、本文書の他の箇所に記載されている診断ボックスと組み合わせて用いることができる。

【0301】

DNAの組成物

胎児に関するゲノムの情報、例えば、胎児の倍数性状態を決定するために、胎児の血液と母系の血液の混合物について測定された配列決定データに対してインフォマティクス分析を実施する場合、対立遺伝子の集合における対立遺伝子分布を測定することが有利であり得る。残念ながら母系の血液試料の血漿において見いだされるDNA混合物から胎児の倍数性状態を決定することを試みる場合などの多くの場合、利用可能なDNAの量は、混合物において優良な忠実度で対立遺伝子分布を直接測定するためには十分でない。これらの場合には、DNA混合物を増幅することにより、所望の対立遺伝子分布を優良な忠実度で測定することができる十分な数のDNA分子がもたらされる。しかし、配列決定するためのDNAの増幅に一般に用いられる増幅の現行の方法は、多くの場合、非常に偏りがあ

る、つまり、多型遺伝子座の両方の対立遺伝子が同じ量で増幅されない。偏りのある増幅の結果、元の混合物における対立遺伝子分布とかなり異なる対立遺伝子分布がもたらされ得る。ほとんどの目的のためには、多型遺伝子座に存在する対立遺伝子の相対的な量を非常に正確に測定することは必要とされない。対照的に、本開示のある実施形態では、多型対立遺伝子の特異的に富化し、対立遺伝子の比を保存する増幅または富化方法は有利である。

【0302】

対立遺伝子の偏りが最小限になるようにDNAの試料を複数の遺伝子座で優先的に富化するために用いることができるいくつかの方法が本明細書に記載されている。いくつかの例では、複数の遺伝子座を標的とするために環状化プローブを使用し、環状化前プローブの3'末端および5'末端が、標的の対立遺伝子の多型部位から1つまたは少数の位置離れた塩基とハイブリダイズするように設計されている。別の例は、3'末端PCRプローブが標的の対立遺伝子の多型部位から1つまたは少数の位置離れた塩基とハイブリダイズするように設計されているPCRプローブを使用するというものである。別の例は、スプリットアンドプール手法を用いて、優先的に富化された遺伝子座が、対立遺伝子の偏りが少なく、直接的な多重化の欠点を伴わずに富化されているDNAの混合物を作製するというものである。別の例は、標的の多型部位に隣接しているDNAとハイブリダイズするように設計されている捕捉プローブの領域が、多型部位と1つまたは少数の塩基で隔てられるように捕捉プローブが設計されているハイブリッド捕捉手法を用いるというものである。

【0303】

多型遺伝子座の集合における測定された対立遺伝子分布を使用して個体の倍数性状態を決定する場合には、遺伝子測定のためにDNAの試料が調製されるとき該DNA試料における対立遺伝子の相対的な量を保存することが望ましい。この調製は、WGA増幅、標的化増幅、選択的富化技法、ハイブリッド捕捉法、環状化プローブまたはDNAの量を増幅し、かつ／または特定の対立遺伝子に対応するDNA分子の存在を選択的に増強することを意図した他の方法を包含し得る。

【0304】

本開示のいくつかの実施形態では、マイナー対立遺伝子頻度が最大である遺伝子座を標的とするように設計されたDNAプローブの集合が存在する。本開示のいくつかの実施形態では、胎児がそれらの遺伝子座において情報価値が高いSNPを有する尤度が最大である遺伝子座を標的とするように設計されたプローブの集合が存在する。本開示のいくつかの実施形態では、プローブが所与の母集団サブグループに対して最適化された遺伝子座を標的とするように設計されたプローブの集合が存在する。本開示のいくつかの実施形態では、プローブが母集団サブグループの所与の混合物に対して最適化された遺伝子座を標的とするように設計されたプローブの集合が存在する。本開示のいくつかの実施形態では、プローブが、異なるマイナー対立遺伝子頻度プロファイルを有する異なる母集団サブグループに由来する所与の親の対に対して最適化された遺伝子座を標的とするように設計されたプローブの集合が存在する。本開示のいくつかの実施形態では、胎児起源のDNAの小部分とアニーリングした少なくとも1つの塩基対を含む環状化DNA鎖が存在する。本開示のいくつかの実施形態では、胎盤起源のDNAの小部分とアニーリングした少なくとも1つの塩基対を含む環状化DNA鎖が存在する。本開示のいくつかの実施形態では、環状化し、その一方でヌクレオチドの少なくとも一部が胎児起源のDNAとアニーリングした環状化DNA鎖が存在する。本開示のいくつかの実施形態では、環状化し、その一方でヌクレオチドの少なくとも一部が胎盤起源のDNAとアニーリングした環状化DNA鎖が存在する。本開示のいくつかの実施形態では、いくつかが単一のタンデム反復を標的とし、いくつかが一塩基多型を標的とするプローブの集合が存在する。いくつかの実施形態では、非侵襲的な出生前診断のために遺伝子座を選択する。いくつかの実施形態では、非侵襲的な出生前診断のためにプローブを使用する。いくつかの実施形態では、環状化プローブ、MIP、ハイブリダイゼーションプローブによる捕捉、SNPアレイ上のプローブ、ま

たはそれらの組合せを含んでよい方法を用いて遺伝子座を標的とする。いくつかの実施形態では、プローブを環状化プローブ、MIP、ハイブリダイゼーションプローブによる捕捉、SNPアレイ上のプローブ、またはそれらの組合せとして使用する。いくつかの実施形態では、非侵襲的な出生前診断のために遺伝子座について配列決定する。

【0305】

配列の相対的 (relative) 情報価値が、関連する親の状況と組み合わせるとより大きくなる場合には、親の状況が既知であるSNPを含有する配列読み取りの数を最大にすることにより、混合試料についての配列決定の読み取りの集合の情報価値が最大になり得る。ある実施形態では、親の状況が既知であるSNPを含有する配列読み取りの数は、qPCRを用いて特定の配列を優先的に増幅することによって増強することができる。ある実施形態では、親の状況が既知であるSNPを含有する配列読み取りの数を、環状化プローブ（例えば、MIP）を用いて特定の配列を優先的に増幅することによって増強することができる。ある実施形態では、親の状況が既知であるSNPを含有する配列読み取りの数を、ハイブリダイゼーション法による捕捉（例えば、SURESELECT）を用いて特定の配列を優先的に増幅することによって増強することができる。異なる方法を用いて、親の状況が既知であるSNPを含有する配列読み取りの数を増強することができる。ある実施形態では、標的化は、伸長ライゲーション、伸長を伴わないライゲーション、ハイブリダイゼーションによる捕捉またはPCRによって実現することができる。

10

【0306】

断片化されたゲノムDNAの試料において、DNA配列のある小部分 (fraction) が個々の染色体に独自にマッピングされ、他のDNA配列は異なる染色体上に見いだされる。血漿中に見いだされるDNAは、母体起源であろうと胎児起源であろうと、一般には、多くの場合、500bpを下回る長さに断片化されていることに留意されたい。典型的なゲノム試料では、マッピング可能な配列のおよそ3.3%が第13染色体にマッピングされ、マッピング可能な配列の2.2%が第18染色体にマッピングされ、マッピング可能な配列の1.35%が第21染色体にマッピングされ、女性ではマッピング可能な配列の4.5%がX染色体にマッピングされ、マッピング可能な配列の2.25%がX染色体にマッピングされ（男性では）、マッピング可能な配列の0.73%がY染色体にマッピングされる（男性では）。これらは胎児において異数性である可能性が最も高い染色体である。また、dbSNPに含まれるSNPを使用すると、短い配列の中では20配列のうちおよそ1つがSNPを含有する。この割合は、発見されていない多くのSNPが存在し得るとすれば、より高くなり得る。

20

30

【0307】

本開示のある実施形態では、標的化方法を用いて、所与の染色体にマッピングされるDNAの試料中のDNAの小部分を、その小部分が、上に列挙されているゲノム試料に典型的な百分率を有意に超えるように増強することができる。本開示のある実施形態では、標的化方法を用いて、DNAの試料中のDNAの小部分を、SNPを含有する配列の百分率が、ゲノム試料に典型的に見出され得る百分率を有意に超えるように増強することができる。本開示のある実施形態では、出生前診断のために、標的化方法を用いて、母系DNAと胎児DNAの混合物中の染色体由来のDNAまたはSNPの集合由来のDNAを標的と

40

【0308】

疑わしい染色体にマッピングされる読み取りの数を計数し、それを、参照染色体にマッピングされる読み取りの数と比較し、疑わしい染色体上の読み取りの存在量が過剰であることは、その染色体における胎児の三倍体性に対応するという仮定を用いることによって胎児の異数性を決定するための方法が報告されていることに留意されたい（米国特許第7,888,017号）。これらの出生前診断のための方法では、いかなる種類の標的化も使用されず、また、出生前診断のための標的化の使用については記載されていない。

【0309】

混合試料の配列決定において標的化手法を用いることにより、少ない配列読み取りを用

50

いて特定のレベルの正確度を実現することが可能であり得る。正確度とは感度を指す場合があり、正確度とは特異度を指す場合があり、または正確度はそのいくつかの組合せを指す場合がある。正確度の所望のレベルは、90%から95%の間であり得、正確度の所望のレベルは、95%から98%の間であり得、正確度の所望のレベルは、98%から99%の間であり得、正確度の所望のレベルは、99%から99.5%の間であり得、正確度の所望のレベルは、99.5%から99.9%の間であり得、正確度の所望のレベルは、99.9%から99.99%の間であり得、正確度の所望のレベルは、99.99%から99.999%の間であり得、正確度の所望のレベルは、99.999%から100%の間であり得る。95%を上回る正確度のレベルを、高い正確度と称することができる。

【0310】

母系DNAと胎児DNAの混合試料からどのように胎児の倍数性状態を決定することができるかについて実証している、いくつかの公開された先行技術の方法、例えば：G. J. W. LiaoらClinical Chemistry 2011年；57巻（1号）92～101頁が存在する。これらの方法は、各染色体に沿った数千もの場所に焦点を当てる。標的とすることができ、一方ではNAの混合試料から、所与の数の配列読み取りについて、胎児における倍数性の決定を高い正確度でもたらず染色体に沿った場所の数は予想外に少ない。本開示のある実施形態では、任意の標的化の方法、例えば、qPCR、リガンド媒介性PCR、他のPCR法、ハイブリダイゼーションによる捕捉または環状化プローブを用いた標的化配列決定を用いることによって正確な倍数性の決定を行うことができ、ここで、標的とする必要がある染色体に沿った遺伝子座の数は、5,000個から2,000個の間の遺伝子座であり得、2,000個から1,000個の間の遺伝子座であり得、1,000個から500個の間の遺伝子座であり得、500個から300個の間の遺伝子座であり得、300個から200個の間の遺伝子座であり得、200個から150個の間の遺伝子座であり得、150個から100個の間の遺伝子座であり得、100個から50個の間の遺伝子座であり得、50個から20個の間の遺伝子座であり得、20個から10個の間の遺伝子座であり得る。最適には、標的とする必要がある染色体に沿った遺伝子座の数は、100個から500個の間の遺伝子座であり得る。高レベルの正確度は、少数の遺伝子座を標的とし予想外に少数の配列読み取りを実行することによって実現することができる。読み取りの数は、1億個から5000万個の間の読み取りであり得、読み取りの数は、5000万個から2000万個の間の読み取りであり得、読み取りの数は、2000万個から1000万個の間の読み取りであり得、読み取りの数は、1000万個から500万個の間の読み取りであり得、読み取りの数は、500万個から200万個の間の読み取りであり得、読み取りの数は、200万個から100万個の間であり得；読み取りの数は、100万個から500,000個の間であり得；読み取りの数は、500,000個から200,000個の間であり得、読み取りの数は、200,000個から100,000個の間であり得、読み取りの数は、100,000個から50,000個の間であり得、読み取りの数は、50,000個から20,000個の間であり得、読み取りの数は、20,000個から10,000個の間であり得、読み取りの数は、10,000個未満であり得る。より大量の入力DNAに対してはより少数の読み取りが必要である。

【0311】

いくつかの実施形態では、胎児起源のDNAと母体起源のDNAの混合物を含む組成物であって、第13染色体に独自にマッピングされる配列のパーセントが4%超、5%超、6%超、7%超、8%超、9%超、10%超、12%超、15%超、20%超、25%超または30%超である組成物が存在する。本開示のいくつかの実施形態では、胎児起源のDNAと母体起源のDNAの混合物を含む組成物であって、第18染色体に独自にマッピングされる配列のパーセントが3%超、4%超、5%超、6%超、7%超、8%超、9%超、10%超、12%超、15%超、20%超、25%超または30%超である組成物が存在する。本開示のいくつかの実施形態では、胎児起源のDNAと母体起源のDNAの混合物を含む組成物であって、第21染色体に独自にマッピングされる配列のパーセントが

2 % 超、3 % 超、4 % 超、5 % 超、6 % 超、7 % 超、8 % 超、9 % 超、10 % 超、12 % 超、15 % 超、20 % 超、25 % 超または30 % 超である組成物が存在する。本開示のいくつかの実施形態では、胎児起源のDNAと母体起源のDNAの混合物を含む組成物であって、独自にX染色体にマッピングされる配列のパーセントが6 % 超、7 % 超、8 % 超、9 % 超、10 % 超、12 % 超、15 % 超、20 % 超、25 % 超または30 % 超である組成物が存在する。本開示のいくつかの実施形態では、胎児起源のDNAと母体起源のDNAの混合物を含む組成物であって、独自にY染色体にマッピングされる配列のパーセントが1 % 超、2 % 超、3 % 超、4 % 超、5 % 超、6 % 超、7 % 超、8 % 超、9 % 超、10 % 超、12 % 超、15 % 超、20 % 超、25 % 超または30 % 超である組成物が存在する。

【0312】

いくつかの実施形態では、組成物は、胎児起源のDNAと母体起源のDNAの混合物を含むと記載され、ある染色体に独自にマッピングされ少なくとも1つの一塩基多型を含有する配列のパーセントは、0.2 % 超、0.3 % 超、0.4 % 超、0.5 % 超、0.6 % 超、0.7 % 超、0.8 % 超、0.9 % 超、1 % 超、1.2 % 超、1.4 % 超、1.6 % 超、1.8 % 超、2 % 超、2.5 % 超、3 % 超、4 % 超、5 % 超、6 % 超、7 % 超、8 % 超、9 % 超、10 % 超、12 % 超、15 % 超または20 % 超であり、染色体は13、18、21、XまたはYの群から選択される。本開示のいくつかの実施形態では、胎児起源のDNAと母体起源のDNAの混合物を含む組成物であって、ある染色体に独自にマッピングされ一塩基多型の集合からの少なくとも1つの一塩基多型を含有する配列のパーセントは0.15 % 超、0.2 % 超、0.3 % 超、0.4 % 超、0.5 % 超、0.6 % 超、0.7 % 超、0.8 % 超、0.9 % 超、1 % 超、1.2 % 超、1.4 % 超、1.6 % 超、1.8 % 超、2 % 超、2.5 % 超、3 % 超、4 % 超、5 % 超、6 % 超、7 % 超、8 % 超、9 % 超、10 % 超、12 % 超、15 % 超または20 % 超であり、染色体は第13染色体、第18染色体、第21染色体、X染色体およびY染色体の集合から選択され、一塩基多型の集合内の一塩基多型の数は、1個から10個の間、10個から20個の間、20個から50個の間、50個から100個の間、100個から200個の間、200個から500個の間、500個から1,000個の間、1,000個から2,000個の間、2,000個から5,000個の間、5,000個から10,000個の間、10,000個から20,000個の間、20,000個から50,000個の間、および50,000個から100,000個の間である組成物が存在する。

【0313】

理論上は、増幅の各サイクルにより、存在するDNAの量が倍增するが、実際には、増幅の程度は2倍よりわずかに低い。理論上は、標的化増幅を含めた増幅により、偏りのないDNA混合物の増幅がもたらされるが、実際には、異なる対立遺伝子は他の対立遺伝子と異なる程度で増幅される傾向がある。DNAを増幅する場合、対立遺伝子の偏りの程度は、一般には増幅ステップの数に伴って上昇する。いくつかの実施形態では、本明細書に記載の方法は、低レベルの対立遺伝子の偏りでDNAを増幅するステップを包含する。対立遺伝子の偏りはさらに別のサイクルのそれぞれで複合されるので、全体的な偏りのn乗根を算出することによってサイクル当たりの対立遺伝子の偏りを決定することができ、ここで、nは富化の程度の、底を2とする対数である。いくつかの実施形態では、第2のDNAの混合物を含む組成物が存在し、該第2のDNAの混合物は、第1のDNAの混合物からの複数の多型遺伝子座に優先的に富化されており、富化の程度は、少なくとも10、少なくとも100、少なくとも1,000、少なくとも10,000、少なくとも100,000または少なくとも1,000,000であり、第2のDNAの混合物での各遺伝子座における対立遺伝子の比は、第1のDNAの混合物でのその遺伝子座における対立遺伝子の比とは、平均で、1,000 % 未満、500 %、200 %、100 %、50 %、20 %、10 %、5 %、2 %、1 %、0.5 %、0.2 %、0.1 %、0.05 %、0.02 % または0.01 % の係数だけ異なる。いくつかの実施形態では、第2のDNAの混合物を含む組成物が存在し、第2のDNAの混合物は、第1のDNAの混合物からの複数の多型遺伝子座に優先的に富化されており、ここで、サイクル当たりの複数の多型遺伝子座

についての対立遺伝子の偏りは、平均で、10%未満、5%、2%、1%、0.5%、0.2%、0.1%、0.05%または0.02%である。いくつかの実施形態では、複数の多型遺伝子座は、少なくとも10個の遺伝子座、少なくとも20個の遺伝子座、少なくとも50個の遺伝子座、少なくとも100個の遺伝子座、少なくとも200個の遺伝子座、少なくとも500個の遺伝子座、少なくとも1,000個の遺伝子座、少なくとも2,000個の遺伝子座、少なくとも5,000個の遺伝子座、少なくとも10,000個の遺伝子座、少なくとも20,000個の遺伝子座、または少なくとも50,000個の遺伝子座を含む。

【0314】

最尤推定

10

生物学的な現象または医学的状态の存在または不在を検出するための当技術分野で公知の大多数の方法は、状態と相関するメトリックを測定し、メトリックが所与の閾値の一方の側にあれば、その状態が存在し、メトリックが閾値の他方の側にあれば、その状態は存在しないという単一仮説棄却検定を用いることを包含する。単一仮説棄却検定では、帰無仮説と対立仮説の間の決定を行う際に帰無分布を調べるだけである。対立分布を考慮に入れなければ、観察されたデータを考慮して各仮説の尤度を推定することはできず、したがって、呼び出しに対する信頼度を算出することができない。したがって、単一仮説棄却検定を用いて、特定の場合と関連する信頼度についての感受性を伴わずにyesまたはnoの答えを得る。

【0315】

20

いくつかの実施形態では、本明細書に開示されている方法では、生物学的な現象または医学的状态の存在または不在を、最尤法を用いて検出することができる。最尤法は、状態の不在または存在を呼び出すための閾値をそれぞれの場合について適切に調整することができるので、単一仮説棄却法を用いる方法に対する実質的な改善である。これは、母系の血漿中に見いだされる浮動性DNAに存在する胎児DNAと母系DNAの混合物から入手可能な遺伝子データから、妊娠中の胎児における異数性の存在または不在を決定することを目的とする診断技法に特に関連性がある。これは、血漿中の胎児DNAの小部分により割合の変化(fraction change)が導かれると、異数性対正倍数性を呼び出すための最適な閾値が変化することに起因する。胎児の割合が降下すると、異数性に関連づけられるデータの分布がますます正倍数性に関連づけられるデータの分布と同様になる。

30

【0316】

最尤推定法では、各仮説に関連づけられる分布を使用して、各仮説に対して条件づけたデータの尤度を推定する。次いで、これらの条件的確率を、仮説呼び出しおよび信頼度に変換することができる。同様に、最大事後推定法では、最尤推定と同じ条件的確率を使用するが、最良の仮説を選択し、信頼度を決定する際に前の母集団も組み入れる。

【0317】

したがって、最尤推定(MLE)技法または密接に関連する最大事後(MAP)技法を用いることにより、2つの利点が生じ、まず、正確な呼び出しの見込みが増大し、また、各呼び出しに対して信頼度を算出することが可能になる。ある実施形態では、最大の確率を有する仮説に対応する倍数性状態を選択するステップを、最尤推定または最大事後推定を使用して行う。ある実施形態では、妊娠中の胎児の倍数性状態を決定するための方法であって、単一仮説棄却法を用いる現在当技術分野で公知の任意の方法をとり、それをMLE技法またはMAP技法を用いるように再公式化することを含む方法が開示されている。これらの技法を適用することによって有意に改善することができる方法のいくつかの例は、米国特許第8,008,018号、米国特許第7,888,017号または米国特許第7,332,277号に見いだすことができる。

40

【0318】

ある実施形態では、胎児のゲノムDNAおよび母系のゲノムDNAを含む母系の血漿試料における胎児の異数性の存在または不在を決定するための方法であって、母系の血漿試

50

料を得るステップと、血漿試料中に見いだされるDNA断片を、ハイスループットシーケンサーを用いて測定するステップと、配列を染色体にマッピングし、各染色体にマッピングされる配列読み取りの数を決定するステップと、血漿試料中の胎児DNAの割合を算出するステップと、第2の標的染色体が正倍数性である場合に存在すると予測される標的染色体の量の予測される分布、およびその染色体が異数性である場合に予測される1つまたは複数の予測される分布を、胎児の割合および正倍数性であることが予測される1つまたは複数の参照染色体にマッピングされる配列読み取りの数を使用して算出するステップと、MLEまたはMAPを用いて、どの分布が、正確である可能性が最も高いかを決定し、それにより、胎児の異数性の存在または不在を示すステップとを含む方法が記載されている。ある実施形態では、血漿由来のDNAを測定するステップは、大規模な並行のショットガン配列決定を行うことを包含し得る。ある実施形態では、血漿試料由来のDNAを測定するステップは、例えば、標的化増幅によって、複数の多型遺伝子座または非多型遺伝子座において優先的に富化されたDNAを配列決定することを包含し得る。複数の遺伝子座を、1つまたは少数の異数性が疑わしい染色体および1つまたは少数の参照染色体を標的とするように設計することができる。優先的に富化することの目的は、倍数性を決定するために情報価値のある配列読み取りの数を増加させることである。

10

【0319】

倍数性呼び出しのインフォマティクスによる方法

本明細書には、配列データを考慮して胎児の倍数性状態を決定するための方法が記載されている。いくつかの実施形態では、この配列データは、ハイスループットシーケンサーで測定することができる。いくつかの実施形態では、配列データは、母系の血液から単離された浮動性DNAを起源とするDNAについて測定することができ、ここで、浮動性DNAは、いくらかの母体起源のDNA、およびいくらかの胎児／胎盤起源のDNAを含む。このセクションでは、分析された混合物中の胎児DNAの割合は未知であり、データから推定されると仮定して胎児の倍数性状態を決定する本開示の一実施形態が記載される。混合物中の胎児DNAの割合（「胎児の割合」）または胎児DNAの百分率を別の方法によって測定することができ、また、それが胎児の倍数性状態の決定において既知であると仮定される実施形態も記載される。いくつかの実施形態では、胎児DNAと母系DNAの混合物である母系の血液試料自体に対して行った遺伝子型決定測定値のみを使用して胎児の割合を算出することができる。いくつかの実施形態では、測定されたか、または別の方法で既知である母親の遺伝子型および／または測定されたか、または別の方法で既知である父親の遺伝子型を用いてその割合を算出することもできる。別の実施形態では、胎児の倍数性状態は、単に、問題の染色体について算出された胎児DNAの割合に基づいて、ダイソミーであると仮定される参照染色体について算出された胎児DNAの割合と比較して決定することができる。

20

30

【0320】

好ましい実施形態では、特定の染色体について、以下が得られるN個のSNPを観察し、解析するとする：

- ・NR個の浮動性DNA配列測定値の集合 $S = (s_1, \dots, s_{NR})$ 。この方法では、SNP測定値を利用するので、非多型の遺伝子座に対応する配列データは全て無視することができる。簡易型では、各SNPに対して(A, B)計数を得、AおよびBが、所与の遺伝子座に存在する2つの対立遺伝子に対応する場合、Sは $S = ((a_1, b_1), \dots, (a_N, b_N))$ と書くことができ、式中、 a_i はSNP i 上のA計数であり、 b_i はSNP i 上のB計数であり、 $\sum_{i=1:N} (a_i + b_i) = NR$ である、および

40

- ・以下からなる親のデータ

- SNPマイクロアレイまたは他の強度に基づく遺伝子型決定プラットフォームからの遺伝子型：母親 $M = (m_1, \dots, m_N)$ 、父親 $F = (f_1, \dots, f_N)$ 、 $m_i, f_i \in (AA, AB, BB)$ および／または

- 配列データ測定値：NRM個の母親の測定値 $SM = (sm_{1r}, \dots, sm_{nr_m})$ 、NRF個の父親測定値 $SF = (sf_{1r}, \dots, sf_{nr_f})$ 。上記の単純化と同様、各SNP

50

について (A, B) 計数が得られる場合、 $SM = ((am_1, bm_1), \dots, (am_N, bm_N))$ 、 $SF = ((af_1, bf_1), \dots, (af_N, bf_N))$ 。

【0321】

集合的に、母親、父親、子のデータは $D = (M, F, SM, SF, S)$ で示される。親のデータが望ましく、それによりアルゴリズムの正確度が上昇するが、特に父親のデータは必須ではないことに留意されたい。これは、母親および／または父親のデータの不在下でさえ、非常に正確なコピー数の結果を生じることが可能であることを意味する。

【0322】

データの対数尤度 $LIK(D|H)$ を、考えられる仮説 (H) 全てにわたって最大にすることによって最良のコピー数の推定値 (H^*) を導くことが可能である。特に、倍数性仮説のそれぞれの相対的確率は、同時分布モデルおよび調製された試料において測定された対立遺伝子数を用いて、および以下のとおり正確である可能性が最も高い仮説を決定するためのこれらの相対的確率を用いて決定することが可能である：

【0323】

【化1】

$$H^* = \underset{H}{\operatorname{argmax}} LIK(D|H)$$

同様に事後仮説尤度は、データを考慮すると：

【0324】

【化2】

$$H^* = \underset{H}{\operatorname{argmax}} LIK(D|H) * \operatorname{priorprob}(H)$$

と書くことができ、式中、 $\operatorname{priorprob}(H)$ はモデル設計および以前の知見に基づいて各仮説 H に割り当てられた事前確率である。

【0325】

事前確率を用いて最大事後推定値を見いだすことも可能である：

【0326】

【化3】

$$H_{MA} = \underset{H}{\operatorname{argmax}} LIK(D|H)$$

ある実施形態では、考慮に入れることができるコピー数仮説は、以下である：

・モノソミー：

○母系 $H10$ (母親由来の1つのコピー)

○父系 $H01$ (父親由来の1つのコピー)

・ダイソミー： $H11$ (母親および父親それぞれにつき1つのコピー)

・単純なトリソミー、乗換えは考慮しない：

○母系： $H21$ __一致 (母親由来の2つの同一のコピー、父親由来の1つのコピー)、
 $H21$ __不一致 (母親由来の両方のコピー、父親由来の1つのコピー)

○父系： $H12$ __一致 (母親由来の1つのコピー、父親由来の2つの同一のコピー)、
 $H12$ __不一致 (母親由来の1つのコピー、父親由来の両方のコピー)

・複合トリソミー、乗換えを考慮する (同時分布モデルを用いる)：

○母系 $H21$ (母親由来の2つのコピー、父親由来の1つのコピー)、

○父系 $H12$ (母親由来の1つのコピー、父親由来の2つのコピー)

他の実施形態では、他の倍数性状態、例えば、零染色体性 ($H00$)、片親性ダイソミー ($H20$ および $H02$)、およびテトラソミー ($H04$ 、 $H13$ 、 $H22$ 、 $H31$ および $H40$) を考慮することができる。

【0327】

乗換えがない場合、起源が有糸分裂、減数分裂 I または減数分裂 II のいずれにしろ、

10

20

30

40

50

各トリソミーは、一致トリソミーまたは不一致トリソミーのうち的一方になる。乗換えに起因して、真のトリソミーは通常、2つの組合せである。まず、単純仮説について仮説の尤度を導く方法が記載されている。次に、個々のSNP尤度を乗換えと組み合わせた複合仮説について仮説の尤度を導く方法が記載されている。

【0328】

単純仮説についてのLIK(D|H)

ある実施形態では、単純仮説について、以下の通りLIK(D|H)を決定することができる。単純仮説Hについて、染色体全体についての仮説Hの対数尤度LIK(H)を、既知のまたは導かれた子の割合cfを仮定して、個々のSNPの対数尤度の合計として算出することができる。ある実施形態では、データからcfを導くことが可能である。

10

【0329】

【化4】

$$LIK(D|H) = \sum_i LIK(D|H, cf, i)$$

この仮説では、いかなるSNP間の連鎖も仮定せず、したがって、同時分布モデルを利用しない。

【0330】

いくつかの実施形態では、SNPごとに対数尤度を決定することができる。特定のSNP_iについて、胎児の倍数性仮説Hおよびパーセント胎児DNA_{cf}を仮定すると、観察されたデータDの対数尤度は、

20

【0331】

【化5】

$$LIK(D|H, i) = \log P(D|H, cf, i) = \log \left(\sum_{m, f, c} P(D|m, f, c, H, cf, i) P(c|m, f, H) P(m|i) P(f|i) \right)$$

と定義され、式中、mは可能性のある真の母親の遺伝子型であり、fは可能性のある真の父親の遺伝子型であり、m, f ∈ {AA, AB, BB}であり、cは、仮説Hを考慮した、可能性のある子の遺伝子型である。特に、モノソミーについてはc ∈ {A, B}であり、ダイソミーについてはc ∈ {AA, AB, BB}であり、トリソミーについてはc ∈ {AAA, AAB, ABB, BBB}である。

30

【0332】

遺伝子型事前頻度：p(m|i)は、SNP_iにおける既知の母集団頻度に基づくSNP_iにおける母親の遺伝子型mの一般的な事前確率であり、pA_iで示される。具体的には、

$$p(AA | pA_i) = (pA_i)^2, \quad p(AB | pA_i) = 2(pA_i) * (1 - pA_i), \quad p(BB | pA_i) = (1 - pA_i)^2$$

父親の遺伝子型の確率、p(f|i)は同様に決定することができる。

【0333】

40

真の子の確率：p(c|m, f, H)は、親m、f、および仮定している仮説Hを考慮して真の子の遺伝子型=cが得られる確率であり、これは容易に算出することができる。例えば、H11、H21一致およびH21不一致についてのp(c|m, f, H)が以下に示されている。

【0334】

【表 1】

p(c m,f,H)													
		H11			H21一致				H21不一致				
m	f	AA	AB	BB	AAA	AAB	ABB	BBB	AAA	AAB	ABB	BBB	
AA	AA	1	0	0	1	0	0	0	1	0	0	0	
AB	AA	0.5	0.5	0	0.5	0	0.5	0	0	1	0	0	
BB	AA	0	1	0	0	0	1	0	0	0	1	0	
AA	AB	0.5	0.5	0	0.5	0.5	0	0	0.5	0.5	0	0	
AB	AB	0.25	0.5	0.25	0.25	0.25	0.25	0.25	0	0.5	0.5	0	
BB	AB	0	0.5	0.5	0	0	0.5	0.5	0	0	0.5	0.5	
AA	BB	0	1	0	0	1	0	0	0	1	0	0	
AB	BB	0	0.5	0.5	0	0.5	0	0.5	0	0	1	0	
BB	BB	0	0	1	0	0	0	1	0	0	0	1	

データの尤度： $P(D|m, f, c, H, i, cf)$ は、真の母親の遺伝子型 m 、真の父親の遺伝子型 f 、真の子の遺伝子型 c 、仮説 H および子の割合 cf を考慮したSNP i における所与のデータ D の確率である。これは、以下の通り母親、父親および子データの確率に分解することができる：

$$P(D|m, f, c, H, cf, i) = P(SM|m, i) P(M|m, i) P(SF|f, i) P(F|f, i) P(S|m, c, H, cf, i) \quad 20$$

母親のSNPアレイデータの尤度：SNPアレイ遺伝子型が正確であると仮定して、真の遺伝子型 m と比較した、SNP i における母親のSNPアレイ遺伝子型データの確率 m_i は、単に、

【0335】

【化6】

$$P(M|m, i) = \begin{cases} 1 & m_i = m \\ 0 & m_i \neq m \end{cases}$$

である。

【0336】

母親の配列データの尤度：SNP i における母親の配列データの確率は、計数 $S_i = (am_i, bm_i)$ の場合には、余分のノイズまたは偏りを伴わず、 $P(SM|m, i) = P_{X|m}(am_i)$ と定義される二項確率であり、 $X|m \sim \text{Binom}(p_m(A), am_i + bm_i)$ 、 $p_m(A)$ は

【0337】

【表2】

m	AA	AB	BB	A	B	呼び出しなし
p(A)	1	0.5	0	1	0	0.5

と定義される。

【0338】

父親のデータの尤度：同様の方程式が父親のデータの尤度に当てはまる。親のデータ、特に父親のデータを用いずに子の遺伝子型を決定することが可能であることに留意されたい。例えば、父親の遺伝子型データ F が利用可能でない場合、単に $P(F|f, i) = 1$ を使用することができる。父親の配列データ SF が利用可能でない場合、単に $P(SF|f, i) = 1$ を使用することができる。

【0339】

いくつかの実施形態では、該方法は、各倍数性仮説について、染色体上の複数の多型遺

10

20

30

40

50

伝子座において、予測される対立遺伝子数についての同時分布モデルを構築することを伴い、そのような目的を実現するための1つの方法がここに記載されている。遊離の胎児DNAデータの尤度： $P(S|m, c, H, cf, i)$ は、SNP i における遊離の胎児DNA配列データの確率であり、真の母親の遺伝子型 m 、真の子の遺伝子型 c 、子のコピー数仮説 H を考慮し、子の割合 cf を仮定する。これは、実際、SNP i におけるA含量の真の確率 $\mu(m, c, cf, H)$ を考慮したSNP i における配列データ S の確率 $P(S|m, c, H, cf, i) = P(S|\mu(m, c, cf, H), i)$ である。

【0340】

計数について、 $S_i = (a_i, b_i)$ であり、データに余分のノイズまたは偏りを伴わない場合、

$$P(S|\mu(m, c, cf, H), i) = P_x(a_i)$$

式中、 $X \sim \text{Binom}(p(A), a_i + b_i)$ 、 $p(A) = \mu(m, c, cf, H)$ 。正確なアラインメントおよびSNP当たりの (A, B) 計数が未知である、より複雑な場合には、 $P(S|\mu(m, c, cf, H), i)$ は積分した二項式の組合せである。

【0341】

真のA含量の確率： $\mu(m, c, cf, H)$ 、この母親／子混合物におけるSNP i におけるA含量の真の確率は、真の母親の遺伝子型 $=m$ 、真の子の遺伝子型 $=c$ 、および全体的な子の割合 $=cf$ と仮定して、

【0342】

【化7】

$$\mu(m, c, cf, H) = \frac{\#A(m) * (1 - cf) + \#A(c) * cf}{n_m * (1 - cf) + n_c * cf}$$

と定義され、式中、 $\#A(g) =$ 遺伝子型 g におけるAの数であり、 $n_m = 2$ は、母親のソミーであり、 n_c は仮説 H の下での子の倍数性である（1はモノソミーについてであり、2はダイソミーについてであり、3はトリソミーについてである）。

【0343】

同時分布モデルの使用：複合仮説についてのLIK(D|H)

いくつかの実施形態では、方法は、各倍数性仮説について、染色体上の複数の多型遺伝子座における予測される対立遺伝子数についての同時分布モデルを構築することを包含し、そのような目的を実現するための1つの方法がここに記載されている。多くの場合、トリソミーは、通常、乗換えに起因して、純粋に一致または不一致ではなく、したがって、このセクションでは、可能性のある乗換えを考慮に入れて、一致トリソミーと不一致トリソミーが組み合わされた複合仮説 H_{21} （母系トリソミー）および H_{12} （父系トリソミー）についての結果が導かれる。

【0344】

トリソミーの場合には、乗換えがなければ、トリソミーは単に一致トリソミーまたは不一致トリソミーになる。一致トリソミーとは、子が一方の親由来の同一染色体セグメントの2つのコピーを遺伝によって受け継ぐ場合である。不一致トリソミーとは、子が親由来の各相同染色体セグメントの1つのコピーを遺伝によって受け継ぐ場合である。乗換えにより、染色体の一部のセグメントが一致トリソミーを有し得、他の部分が、不一致トリソミーを有し得る。このセクションには、対立遺伝子の集合について、ヘテロ接合率について、すなわち、1つまたは複数の仮説について、いくつかの遺伝子座における予測される対立遺伝子数の同時分布モデルをどのように構築するかが記載されている。

【0345】

SNP i に対し、LIK(D|H_m, i)は、一致仮説 H_m に対する適合であり、LIK(D|H_u, i)は、不一致仮説 H_u に対する適合であり、 $p_c(i) = \text{SNP } i - 1$ とSNP i の間の乗換えの確率であると仮定する。このとき、完全な尤度を以下の通り算出することができる：

【0346】

【化8】

$$\text{LIK}(D|H) = \sum_E \text{LIK}(D|E, 1:N)$$

式中、 $\text{LIK}(D|E, 1:N)$ は、SNP 1:Nについての仮説Eの結末の尤度である。E=最後のSNPの仮説であり、 $E \in (H_m, H_u)$ である。再帰的に、以下を算出することができる：

【0347】

【化9】

$$\begin{aligned} \text{LIK}(D|E, 1:i) &= \text{LIK}(D|E, i) + \log(\exp(\text{LIK}(D|E, 1:i-1)) * (1 - \text{pc}(i)) \\ &\quad + \exp(\text{LIK}(D|\sim E, 1:i-1)) * \text{pc}(i)) \end{aligned}$$

10

式中、 $\sim E$ はE以外の仮説（非E）であり、考慮される仮説は H_m および H_u である。詳細には、1:i SNPの尤度を、1~(i-1) SNPの尤度に基づいて、同じ仮説で乗換えなしに、または逆の仮説で乗換えありのいずれかを用い、SNP iの尤度を掛けて算出することができる

SNP 1について、 $i = 1$ 、 $\text{LIK}(D|E, 1:1) = \text{LIK}(D|E, 1)$ 。

【0348】

SNP 2について、 $i = 2$ 、 $\text{LIK}(D|E, 1:2) = \text{LIK}(D|E, 2) + \log(\exp(\text{LIK}(D|E, 1)) * (1 - \text{pc}(2)) + \exp(\text{LIK}(D|\sim E, 1)) * \text{pc}(2))$

20

および $i = 3:N$ について同様である。

【0349】

いくつかの実施形態では、子の割合(fraction)を決定することができる。子の割合とは、DNAの混合物における子を起源とする配列の割合(proportion)を指し得る。非侵襲的な出生前診断との関連において、子の割合とは、母系の血漿における、胎児または胎児の遺伝子型を有する胎盤の部分の起源とする配列の割合を指し得る。子の割合とは、母系の血漿から調製したDNAの試料中の子の割合を指し得、胎児DNAに関して富化されていてよい。DNAの試料中の子の割合を決定する1つの目的は、胎児に関して倍数性呼び出しを行うことができるアルゴリズムにおいて使用するためであり、したがって、子の割合とは、非侵襲的な出生前診断のために配列決定によって分析した

30

いかなるDNAの試料をも指し得る。

【0350】

非侵襲的な出生前異数性診断の方法の一部である本開示において示されているアルゴリズムのいくつかは、既知の子の割合を仮定し、これはいつでも当てはまるわけではない。ある実施形態では、親のデータの存在を伴って、または伴わずに、選択された染色体上のダイソミーの尤度を最大にすることによって、最も可能性が高い子の割合を見いだすことが可能である。

【0351】

40

詳細には、ダイソミー仮説について、および染色体chr上の子の割合cfについて、上記の通り $\text{LIK}(D|H11, cf, chr) = \text{対数尤度}$ と仮定する。正倍数性であると仮定した、Csetにおいて選択された染色体（通常、1:16）についての完全な尤度は：

【0352】

【化10】

$$\text{LIK}(cf) = \sum_{chr \in Cset} \text{Lik}(D|H11, cf, chr)$$

である。

50

最も可能性が高い子の割合 ($c f^*$) は、 $c f^* = \arg \max_{c f} L I K (c f)$ として導かれる。

【0353】

染色体の任意の集合を使用することが可能である。参照染色体における正倍数性を仮定せずに子の割合を導くことも可能である。この方法を用いて、以下の状況のいずれかについて子の割合を決定することが可能である：（１）親に関するアレイデータおよび母系の血漿に関するショットガン配列決定データを有する；（２）親に関するアレイデータおよび母系の血漿に関する標的化配列決定データを有する；（３）親と母系の血漿の両方に関する標的化配列決定データを有する；（４）母親と母系の血漿画分の両方に関する標的化配列決定データを有する；（５）母系の血漿画分に関する標的化配列決定データを有する；（６）親の画分および子の画分の測定値の他の組合せ。

10

【0354】

いくつかの実施形態では、インフォマティクスによる方法により、データドロップアウトを組み入れることができる；これにより、より正確度が高い倍数性の決定をもたらすことができる。本開示の他の箇所において、Aが生じる確率は、真の母親の遺伝子型、真の子の遺伝子型、混合物中の子の割合、および子のコピー数の一次関数であると仮定されている。例えば、混合物中の真の子ABを測定する代わりに、母親または子の対立遺伝子がドロップアウトし得る可能性もあり、これは対立遺伝子Aにマッピングされる配列のみを測定する場合であり得る。ゲノムのイルミナデータについての親のドロップアウト率を $d_{p g}$ 、配列データについての親のドロップアウト率を $d_{p s}$ および配列データについての子のドロップアウト率を $d_{c s}$ で示すことができる。いくつかの実施形態では、母親のドロップアウト率を0、子のドロップアウト率を比較的低いと仮定することができ、この場合、結果はドロップアウトの影響を著しくは受けない。いくつかの実施形態では、対立遺伝子ドロップアウトの可能性は、それにより予測される倍数性呼び出しに有意な影響がもたらされるのに十分に大きい場合がある。そのような場合、対立遺伝子ドロップアウトをここでアルゴリズムに組み入れる：

20

親のSNPアレイデータドロップアウト：母親のゲノムのデータMについて、ドロップアウト後の遺伝子型を m_d とすると、

【0355】

【化11】

30

$$P(M|m, i) = \sum_{m_d} P(M|m_d, i) P(m_d|m)$$

式中、上記と同様に

【0356】

【化12】

$$P(M|m_d, i) = \begin{cases} 1 & m_i = m_d \\ 0 & m_i \neq m_d \end{cases}$$

40

であり、 $P(m_d | m)$ は、ドロップアウト率dについて以下の通り定義される真の遺伝子型mを考慮した可能性のあるドロップアウト後の遺伝子型 m_d の尤度である

【0357】

【表 3】

m	md					
	AA	AB	BB	A	B	呼び出しなし
AA	(1-d) ²	0	0	2d(1-d)	0	d ²
AB	0	(1-d) ²	0	d(1-d)	d(1-d)	d ²
BB	0	0	(1-d) ²	0	2d(1-d)	d ²

同様の方程式が父親の S N P アレイデータに当てはまる。

【0 3 5 8】

親の配列データドロップアウト：母親の配列データ S M について

【0 3 5 9】

【化 1 3】

$$P(SM|m,i) = \sum_{m_d} P_{X|m_d}(am_i)P(m_d|m)$$

式中、 $P(m_d|m)$ は前のセクションで定義された通りであり、二項分布からの $P_{x|m_d}(am_i)$ 確率は、親のデータの尤度セクションにおいて上記の通り定義される。同様の方程式が父系の配列データに当てはまる。

【0 3 6 0】

浮動性 D N A 配列データドロップアウト：

【0 3 6 1】

【化 1 4】

$$P(S|m,c,H,cf,i) = \sum_{m_d,c_d} P(S|\mu(m_d,c_d,cf,H),i)P(m_d|m)P(c_d|c)$$

式中、 $P(S|\mu(m_d,c_d,cf,H),i)$ は浮動性のデータの尤度に関するセクションにおいて定義されている通りである。

【0 3 6 2】

ある実施形態では、 $p(m_d|m)$ は、真の母親の遺伝子型 m を考慮し、ドロップアウト率 d_{ps} を仮定した、観察された母親の遺伝子型 m_d の確率であり、 $p(c_d|c)$ は、真の子の遺伝子型 c を考慮し、ドロップアウト率 d_{cs} を仮定した、観察された子の遺伝子型 c_d の確率である。 n_{AT} = 真の遺伝子型 c における対立遺伝子 A の数であり、 n_{AD} = 観察された遺伝子型 c_d における対立遺伝子 A の数であり、 $n_{AT} \geq n_{AD}$ であり、同様に、 n_{BT} = 真の遺伝子型 c における対立遺伝子 B の数であり、 n_{BD} = 観察された遺伝子型 c_d における対立遺伝子 B の数であり、 $n_{BT} \geq n_{BD}$ であり、 d = ドロップアウト率であるとする、

【0 3 6 3】

【化 1 5】

$$p(c_d|c) = \binom{n_{AT}}{n_{AD}} * d^{n_{AT}-n_{AD}} * (1-d)^{n_{AD}} * \binom{n_{BT}}{n_{BD}} * d^{n_{BT}-n_{BD}} * (1-d)^{n_{BD}}$$

である。

【0 3 6 4】

ある実施形態では、インフォマティクスによる方法により、ランダムな偏りおよび一貫した偏りが組み入れられる可能性がある。理想的に言えば、配列計数の数に S N P 当たりの一貫したサンプリングの偏りまたはランダムなノイズ（二項分布の変動に加えて）は存在しない。詳細には、S N P i において、母親の遺伝子型 m 、真の子の遺伝子型 c および子の割合 cf 、および $X = S N P \quad i$ における $(A+B)$ 読み取りの集合内の A の数に

対して、 X は $X \sim \text{Binomial}(p, A+B)$ として作用し、 $p = \mu(m, c, cf, H) = A$ 含量の真の確率である。

【0365】

ある実施形態では、インフォマティクスによる方法により、ランダムな偏りが組み入れられる可能性がある。大抵の場合、測定値に偏りがあると仮定し、したがってこのSNPにおいてAが生じる確率は、上で定義された p とは少し異なる q に等しい。 p と q がどのくらい異なるかは、測定プロセスの正確度および他の因子の数に左右され、 p から離れた q の標準偏差によって定量化することができる。ある実施形態では、 q を、ベータ分布を有するとして p に集中したその分布の平均に応じたパラメータ α 、 β 、および一部の特定の標準偏差 s を用いてモデリングすることが可能である。詳細には、これにより $X|q \sim \text{Bin}(q, D_i)$ がもたらされ、 $q \sim \text{Beta}(\alpha, \beta)$ である。 $E(q) = p$ 、 $V(q) = s^2$ とすれば、パラメータ α 、 β は $\alpha = pN$ 、 $\beta = (1-p)N$ として導くことができ、

【0366】

【化16】

$$N = \frac{p(1-p)}{s^2} - 1$$

である。

【0367】

これはベータ二項分布の定義であり、可変性のパラメータ q を伴う二項分布からサンプリングし、 q は平均の p を有するベータ分布に従う。したがって、偏りが無いセットアップでは、SNP i において、真の母親の遺伝子型(m)を仮定し、SNP i 上の母親の配列A計数(a_{m_i})およびSNP i 上の母親の配列B計数(b_{m_i})を考慮した親の配列データ(SM)確率を以下の通り算出することができる：

$P(SM|m, i) = P_{X|m}(a_{m_i})$ 、ここで、 $X|m \sim \text{Binom}(p_m(A), a_{m_i} + b_{m_i})$ 。

【0368】

ここで、標準偏差 s を有するランダムな偏りを含めると、以下になる：

$X|m \sim \text{BetaBinom}(p_m(A), a_{m_i} + b_{m_i}, s)$ 。

【0369】

偏りが無い場合には、真の母親の遺伝子型(m)、真の子の遺伝子型(c)、子の割合(cf)を仮定し、子仮説 H を仮定し、SNP i 上の浮動性DNA配列A計数(a_i)およびSNP i 上の浮動性配列B計数(b_i)を考慮した母系の血漿DNA配列データ(S)確率を以下の通り算出することができる

$P(S|m, c, cf, H, i) = P_X(a_i)$

式中、 $X \sim \text{Binom}(p(A), a_i + b_i)$ 、 $p(A) = \mu(m, c, cf, H)$ 。

【0370】

ある実施形態では、標準偏差 s を有するランダムな偏りを含めると、これは $X \sim \text{BetaBinom}(p(A), a_i + b_i, s)$ になり、余分の変動の量は偏差パラメータ s または同等に N によって指定される。 s の値が小さいほど(または N の値が大きいほど)この分布は標準的な二項分布に近づく。偏りの量を推定すること、すなわち上記の N を、明白な状況 $AA|AA$ 、 $BB|BB$ 、 $AA|BB$ 、 $BB|AA$ から、上記の確率における推定

【0371】

【化17】

$\hat{N}$

を用いて推定することが可能である。データの挙動に応じて、 N を、読み取りの深さ $a_i + b_i$ または $a_i + b_i$ の関数に関係なく一定になるようにし、これにより、より大きな

読み取りの深さに対して偏りをより小さくすることができる。

【0372】

ある実施形態では、インフォマティクスによる方法により、一貫したSNPごとの偏りが組み入れられる可能性がある。配列決定プロセスのアーチファクトに起因して、一部のSNPは、真のA含量に関係なく、一貫して低いまたは高い計数を有し得る。SNP i が一貫してA計数の数に対して w_i パーセントの偏りを加えると仮定する。いくつかの実施形態では、この偏りを、同じ条件で導かれたトレーニングデータの集合から推定し、親の配列データ推定値に戻すことができる：

$$P(SM|m, i) = P_{X|m}(am_i), X|m \sim \text{BetaBinom}(p_m(A) + w_i, am_i + bm_i, s),$$

浮動性DNA配列データ確率推定値：

$$P(S|m, c, cf, H, i) = P_X(a_i), X \sim \text{BetaBinom}(p(A) + w_i, a_i + b_i, s).$$

【0373】

いくつかの実施形態では、該方法は、特にさらに別のノイズ、示差的な試料の質、示差的なSNPの質、およびランダムサンプリングの偏りを考慮に入れて書くことができる。この例はここに示されている。この方法は、データを大規模に多重化されたmini-PCRプロトコールを使用して生成した状況において特に有用であることが示されており、これを実験7〜13において用いた。該方法は、それぞれが最終のモデルに異なる種類のノイズおよび／または偏りを導入するいくつかのステップを伴う：

(1) 母系DNAと胎児DNAの混合物を含む第1の試料は、サイズ = N_0 分子、通常、1,000〜40,000の範囲、 p = 真の%refの元のDNAの量を含むと仮定する。

【0374】

(2) ユニバーサルライゲーションアダプタを使用した増幅において、 N_1 分子がサンプリングされると仮定する；通常、サンプリングに起因して $N_1 \sim N_0 / 2$ 分子およびランダムサンプリングの偏りが導入される。増幅された試料は、いくつかの分子 N_2 を含有してよく、 $N_2 \gg N_1$ である。 X_1 は N_1 サンプリングされた分子のうちの参照遺伝子座の量 (SNPごと) を示し、プロトコールの残り全体を通してランダムサンプリングの偏りを導入する $p_1 = X_1 / N_1$ の変動を伴うとする。このサンプリングの偏りを、単純な二項分布モデルを使用する代わりにベータ二項 (BB) 分布を使用することによってモデルに含める。ベータ二項分布のパラメータ N を後で、試料ごとに、 $0 < p < 1$ のSNPについて漏れおよび増幅の偏りを調整した後に、トレーニングデータから推定することができる。漏れは、SNPが不正確に読み取られる傾向である。

【0375】

(3) 増幅ステップにより、対立遺伝子のあらゆる偏りが増幅され、したがって、可能性のある一様でない増幅によって増幅の偏りが導入される。遺伝子座における一方の対立遺伝子が f 倍に増幅され、その遺伝子座における他方の対立遺伝子が g 倍に増幅されると仮定すると、 $f = g e^b$ であり、 $b = 0$ は偏りがないことを示す。偏りパラメータ b は0に集中し、特定のSNPにおいて対立遺伝子Aが対立遺伝子Bと対照的にどのくらい多くまたは少なく増幅されたかを示す。パラメータ b は、SNPによって異なってよい。偏りパラメータ b は、SNPごとに、例えば、トレーニングデータから推定することができる。

【0376】

(4) 配列決定ステップは、増幅された分子の試料について配列決定するステップを包含する。このステップでは、漏れが存在する可能性があり、漏れとは、SNPが不正確に読み取られる状況である。漏れは、任意の数の問題に起因し得、漏れの結果、SNPは、正確な対立遺伝子Aとして読み取られないが、その遺伝子座において見いだされる別の対立遺伝子Bとして、または一般にはその遺伝子座において見いだされない対立遺伝子CまたはDとして読み取られる。配列決定により、サイズ N_3 、 $N_3 < N_2$ の増幅された試料

からいくつかのDNA分子の配列データが測定されると仮定する。いくつかの実施形態では、 N_3 は、20, 000~100, 000; 100, 000~500, 000; 500, 000~4, 000, 000; 4, 000, 000~20, 000, 000; または20, 000, 000~100, 000, 000の範囲内であってよい。サンプリングされた各分子は正確に読み取られた確率 p_g を有し、その場合、正確に対立遺伝子Aとして示される。試料は、確率 $1 - p_g$ で元の分子と無関係の対立遺伝子として不正確に読み取られ、確率 p_r で対立遺伝子Aのようであり、確率 p_m で対立遺伝子Bのようであり、または確率 p_o で対立遺伝子Cまたは対立遺伝子Dのようであり、 $p_r + p_m + p_o = 1$ である。パラメータ p_g 、 p_r 、 p_m 、 p_o は、SNPごとに、トレーニングデータから推定する。

10

【0377】

異なるプロトコルは、同様のステップを包含し、分子生物学的ステップに変動を伴ってよく、その結果、異なる量のランダムサンプリング、異なるレベルの増幅および異なる漏れによる偏りがもたらされる。以下のモデルを、これらの場合のそれぞれに同等に良好に適用することができる。サンプリングされたDNAの量のモデルは、SNPごとに、以下によって示される：

$$X_3 \sim \text{Beta Binomial} (L(F(p, b), p_r, p_g), N * H(p, b))$$

式中、 p = 参照DNAの真の量であり、 b = SNP当たりの偏りであり、上記のように、 p_g は正確な読み取りの確率であり、 p_r は、上記の通り、悪い読み取りの場合、不正確に読み取られたが、思いがけなく正確な対立遺伝子に見える読み取りの確率であり：

20

$$F(p, b) = p e^b / (p e^b + (1 - p)), H(p, b) = (e^b p + (1 - p))^2 / e^b, L(p, p_r, p_g) = p * p_g + p_r * (1 - p_g) \text{ である。}$$

【0378】

いくつかの実施形態では、該方法では単純な二項分布の代わりにベータ二項分布を使用する；これは、ランダムサンプリングの偏りに対処する。ベータ二項分布のパラメータ N は、試料ごとに、必要に応じて推定する。単に p の代わりに偏り補正 $F(p, b)$ 、 $H(p, b)$ を用いて、増幅の偏りに対処する。偏りのパラメータ b は、前もって、SNPごとに、トレーニングデータから推定する。

【0379】

30

いくつかの実施形態では、該方法では、単に p の代わりに漏れ補正 $L(p, p_r, p_g)$ を用い；これは漏れによる偏り、すなわちSNPおよび試料の質の変動に対処する。いくつかの実施形態では、パラメータ p_g 、 p_r 、 p_o は、前もって、SNPごとに、トレーニングデータから推定する。いくつかの実施形態では、パラメータ p_g 、 p_r 、 p_o は、実行されている現行の試料を用いて更新して、試料の質の変動を明らかにすることができる。

【0380】

本明細書に記載のモデルは、かなり一般的であり、示差的な試料の質と示差的なSNPの質の両方を明らかにすることができる。異なる試料およびSNPは、いくつかの実施形態では平均および分散が元のDNAの量、ならびに試料およびSNPの質の関数であるベータ二項分布を用いるという事実によって例証されるように、異なって処理される。

40

【0381】

プラットフォームのモデリング

血漿中に存在する予測される対立遺伝子の比が r である（母系の遺伝子型および胎児の遺伝子型に基づいて）単一のSNPを考慮に入れる。予測される対立遺伝子の比は、母系DNAと胎児DNAの組合せにおいて、予測される対立遺伝子Aの割合と定義される。母系の遺伝子型 g_m および子の遺伝子型 g_c について、予測される対立遺伝子の比は、遺伝子型が同様に対立遺伝子の比で示されると仮定して、方程式1によって示される。

$$r = f g_c + (1 - f) g_m \quad (1)$$

SNPにおける観察は、存在する各対立遺伝子でマッピングされた読み取りの数、 n_a

50

および n_b からなり、合計して読み取りの深さ d になる。閾値が既にマッピング確率および $phred$ スコアに適用されており、したがって、マッピングおよび対立遺伝子の観察を正確であるとみなすことができると仮定する。 $phred$ スコアとは、特定の塩基における特定の測定値が誤りである確率に関する数値尺度である。ある実施形態では、塩基を配列決定によって測定した場合、 $phred$ スコアは、呼び出された塩基に対応する色素の強度と他の塩基の色素の強度の比から算出することができる。尤度を観察するための最も単純なモデルは、 d 読み取りのそれぞれが、対立遺伝子の比 r を有する大規模なプールからそれぞれ独立に抜き取られたと仮定する二項分布である。方程式 2 によりこのモデルが説明される。

【0382】

【化18】

$$P(n_a, n_b | r) = \text{pbinom}(n_a; n_a + n_b, r) = \binom{n_a + n_b}{n_a} r^{n_a} (1 - r)^{n_b} \quad (2)$$

二項式モデルは、いくつかのやり方で拡張することができる。母系の遺伝子型および胎児の遺伝子型が全て A であるかまたは全て B であるかのいずれかの場合、血漿における予測される対立遺伝子の比は 0 または 1 になり、二項確率は明確に定義されない。実際には、時には実施において予想外の対立遺伝子が観察される。ある実施形態では、補正した対立遺伝子の比

【0383】

【化19】

$$\hat{r} = 1 / (n_a + n_b)$$

を用いて、予想外の対立遺伝子を少数にすることが可能である。ある実施形態では、トレーニングデータを用いて、各 SNP 上に現れる予想外の対立遺伝子の率をモデリングすること、およびこのモデルを使用して予測される対立遺伝子の比を補正することが可能である。予測される対立遺伝子の比が 0 または 1 ではない場合、観察された対立遺伝子の比は、増幅の偏りまたは他の現象に起因して、予測される対立遺伝子の比に十分に高い読み取りの深さに収束し得ない。次いで、対立遺伝子の比を、予測される対立遺伝子の比に集中したベータ分布としてモデリングし、二項分布よりも分散が大きい $P(n_a, n_b | r)$ についてのベータ二項分布をもたらすことができる。

【0384】

単一の SNP における応答についてのプラットフォームモデルは、 $F(a, b, g_c, g_m, f)$ と定義される (3)、または観察されている $n_a = a$ および $n_b = b$ の確率は、母系の遺伝子型および胎児の遺伝子型を考慮すると、同様に方程式 1 による胎児の割合に左右される。F の関数形式は、二項分布、ベータ二項分布または上記と同様の関数であってよい。

$$F(a, b, g_c, g_m, f) = P(n_a = a, n_b = b | g_c, g_m, f) = P(n_a = a, n_b = b | r(g_c, g_m, f)) \quad (3)$$

ある実施形態では、子の割合を以下の通り決定することができる。出生前検査のための胎児の割合 f の最尤推定値は、父系の情報を使用することなく導くことができる。これは、父系の遺伝子型データが入手不可能である場合、例えば、記録の父親が実際には胎児の遺伝学的父親ではない場合に関連性があり得る。胎児の割合は、母系の遺伝子型が 0 または 1 である場合に SNP の集合から推定し、その結果、可能性のある胎児の遺伝子型 2 つのみの集合がもたらされる。 S_0 を、母系の遺伝子型が 0 である SNP の集合と定義し、 S_1 を、母系の遺伝子型が 1 である SNP の集合と定義する。 S_0 における可能性のある胎児の遺伝子型は 0 および 0.5 であり、可能性のある対立遺伝子の比の集合 $R_0(f) = \{0, f/2\}$ がもたらされる。同様に、 $R_1(f) = \{1 - f/2, 1\}$ である。この方法は、母系の遺伝子型が 0.5 である SNP を含むように自明に拡張され得るが、これらの SNP は、可能性のある対立遺伝子の比のより大きな集合に起因して情報価値が低い

10

20

30

40

50

。

【0385】

N_{a0} および N_{b0} を、 S_0 における SNP について n_{as} および n_{bs} によって形成されるベクトルと定義し、 N_{a1} および N_{b1} を同様に S_1 について定義する。 f の最尤推定値

【0386】

【化20】

 $\hat{f}$

は方程式 4 によって定義される。

10

【0387】

【化21】

$$\hat{f} = \arg \max_f P(N_{a0}, N_{b0}|f) P(N_{a1}, N_{b1}|f) \quad (4)$$

各 SNP における対立遺伝子数を SNP の血漿対立遺伝子の比に対して独立して条件づけたと仮定して、確率は、各集合内の SNP に関する積として表すことができる (5)。

【0388】

【化22】

$$P(N_{a0}, N_{b0}|f) = \prod_{s \in S_0} P(n_{as}, n_{bs}|f) \quad (5)$$

20

$$P(N_{a1}, N_{b1}|f) = \prod_{s \in S_1} P(n_{as}, n_{bs}|f)$$

f への依存は、可能性のある対立遺伝子の比の集合 $R_0(f)$ および $R_1(f)$ による。 SNP 確率 $P(n_{as}, n_{bs} | f)$ は、 f に対して条件づけた最尤遺伝子型を仮定することによって概算することができる。合理的に高い胎児の割合および読み取りの深さにおける最尤遺伝子型の選択は信頼度が高くなる。例えば、胎児の割合 10 パーセントおよび読み取りの深さ 1000 において、母親が遺伝子型 0 を有する SNP を考慮する。予測される対立遺伝子の比は 0 パーセントおよび 5 パーセントであり、これは十分に高い読み取りの深さにおいて容易に区別可能である。推定される子の遺伝子型を方程式 5 に代入することにより、胎児の割合を推定するための完全な方程式 (6) がもたらされる。

30

【0389】

【化23】

$$\hat{f} = \arg \max_f \left[\prod_{s \in S_0} \left(\max_{r_s \in R_0(f)} P(n_{as}, n_{bs}|r_s) \right) \prod_{s \in S_1} \left(\max_{r_s \in R_1(f)} P(n_{as}, n_{bs}|r_s) \right) \right] \quad (6)$$

胎児の割合は、範囲 $[0, 1]$ でなければならず、したがって、条件付き次元検索によって最適化を容易に実行することができる。

【0390】

低い読み取りの深さまたは高いノイズレベルの存在下では、不自然に高い信頼度をもたらし得る最尤遺伝子型を仮定しないことが好ましい場合がある。別の方法では、各 SNP における可能性のある遺伝子型にわたって合計し、その結果、 S_0 における SNP についての $P(n_a, n_b | f)$ について以下の表現 (7) がもたらされる。事前確率 $P(r)$ は、 $R_0(f)$ にわたって一様であると仮定することができる、または母集団頻度に基づいてよい。群 S_1 への拡大は自明である。

40

【0391】

【化24】

$$P(n_a, n_b|f) = \sum_{r \in R_0(f)} P(n_a, n_b|r) P(r) \quad (7)$$

50

いくつかの実施形態では、確率を以下の通り導くことができる。2つの仮説 H_t および H_f のデータ尤度から信頼度を算出することができる。各仮説の尤度を、応答モデル、推定される胎児の割合、母親の遺伝子型、対立遺伝子の母集団頻度、および血漿対立遺伝子数に基づいて導く。

【0392】

以下の表記を定義する：

G_m 、 G_c 真の母系の遺伝子型および子の遺伝子型

G_{af} 、 G_{tf} 父親とされる人の真の遺伝子型および真の父親の真の遺伝子型

$G(g_c, g_m, g_{tf}) = P(G_c = g_c | G_m = g_m, G_{tf} = g_{tf})$ 遺伝形質確率

$P(g) = P(G_{tf} = g)$ 特定のSNPにおける遺伝子型 g の母集団頻度

各SNPにおける観察は血漿対立遺伝子の比に対して独立して条件づけられると仮定して、父系性仮説の尤度はSNPにおける尤度の積である。以下の方程式により、単一のSNPについての尤度が導かれる。方程式8は、任意の仮説 h の尤度についての一般的な表現であり、次いで、 H_t および H_f の特定の場合に分解される。

【0393】

【化25】

$$\begin{aligned} P(n_a, n_b | h, G_m, G_{tf}, f) &= \sum_{g_c \in (0,0.5,1)} P(n_a, n_b | G_c = g_c, G_m, G_{tf}, h, f) P(G_c = g_c, G_m, G_{tf}, h, f) \\ &= \sum_{g_c \in (0,0.5,1)} P(n_a, n_b | G_c = g_c, G_m, f) P(G_c = g_c | G_m, G_{tf}, h) \\ &= \sum_{g_c \in (0,0.5,1)} F(n_a, n_b, g_c, g_m, f) P(G_c = g_c | G_m, G_{tf}, h) \quad (8) \end{aligned}$$

H_t の場合には、父親とされる人は真の父親であり、胎児の遺伝子型は、方程式9に従って母系の遺伝子型および父親とされる人の遺伝子型から遺伝によって受け継がれる。

【0394】

【化26】

$$\begin{aligned} P(n_a, n_b | H_t, G_m, G_{tf}, f) &= \sum_{g_c \in (0,0.5,1)} F(n_a, n_b, g_c, g_m, f) P(G_c = g_c | G_m, G_{tf}, H_t) \quad (9) \\ &= \sum_{g_c \in (0,0.5,1)} F(n_a, n_b, g_c, g_m, f) G(g_c, G_m, G_{tf}) \end{aligned}$$

H_f の場合には、父親とされる人は真の父親ではない。真の父親の遺伝子型の最良の推定値は、各SNPにおける母集団頻度によって生じる。したがって、子の遺伝子型の確率を、方程式10の場合と同様に既知の母親の遺伝子型および母集団頻度によって決定する。

【0395】

【化27】

$$\begin{aligned} P(n_a, n_b | H_f, G_m, G_{tf}, f) &= \sum_{g_c \in (0,0.5,1)} F(n_a, n_b, g_c, g_m, f) P(G_c = g_c | G_m, G_{tf}, H_f) \\ &= \sum_{g_c \in (0,0.5,1)} F(n_a, n_b, g_c, g_m, f) P(G_c = g_c | G_m) \\ &= \sum_{g_c \in (0,0.5,1)} \sum_{g_{tf} \in (0,0.5,1)} F(n_a, n_b, g_c, g_m, f) P(G_c = g_c | G_m, G_{tf} = g_{tf}) P(G_{tf} = g_{tf}) \\ &= \sum_{g_c \in (0,0.5,1)} \sum_{g_{tf} \in (0,0.5,1)} F(n_a, n_b, g_c, g_m, f) G(g_c, G_m, g_{tf}) P(g_{tf}) \end{aligned}$$

正確な父系性に対する信頼度 C_p を、ベイズの法則(11)を用いて、2つの尤度のSNPに関する積から算出する。

【0396】

【化 28】

$$C_p = \frac{\prod_s P(n_{as}, n_{bs} | H_t, G_{ms}, G_{tf}, f)}{\prod_s P(n_{as}, n_{bs} | H_t, G_{ms}, G_{tf}, f) + \prod_s P(n_{as}, n_{bs} | H_f, G_{ms}, G_{tf}, f)} \quad (11)$$

パーセントによる胎児の割合を用いた最尤モデル

母系の血清中に含有される浮動性DNAを測定することによって、または任意の混合試料中の遺伝子型の材料を測定することによって胎児の倍数性状態を決定することは、非自明の作業である。いくつかの方法であって、例えば、推測が、胎児が特定の染色体においてトリソミーである場合は、母系の血液に見いだされるその染色体由来の全体的なDNAの量が参照染色体に対して上昇するというものである読み取り数解析を実施する方法が存在する。そのような胎児においてトリソミーを検出するための1つのやり方は、各染色体について予測されるDNAの量を、例えば、所与の染色体に対応する分析集合内のSNPの数に従って、または染色体の独自にマッピング可能な部分の数に従って正規化することである。測定値が正規化されたら、特定の閾値を超えるDNAの量が測定された任意の染色体をトリソミーであると決定する。この手法は、Fanら PNAS、2008年；105巻（42号）；16266～16271頁に記載されており、Chiuら BMJ 2011年；342巻：c7401頁にも記載されている。Chiuらの論文では、正規化は、以下の通りZスコアを算出することによって実現された：

検査例における第21染色体の百分率についてのZスコア＝（（検査例における第21染色体の百分率）－（参照対照における第21染色体の百分率の平均））／（参照対照における第21染色体の百分率の標準偏差）。

【0397】

これらの方法では、単一仮説棄却法を用いて胎児の倍数性状態を決定する。しかし、これらは、いくつかの著しい欠点を被る。胎児における倍数性を決定するためのこれらの方法は試料中の胎児DNAの百分率に従って不変であるので、1つのカットオフ値を使用し、その結果、決定の正確度は最適ではなく、混合物中の胎児DNAの百分率が比較的低い場合は、正確度が最も悪くなる。

【0398】

ある実施形態では、胎児の倍数性状態を決定するために用いる本開示の方法は、試料中の胎児DNAの割合を考慮に入れるステップを包含する。本開示の別の実施形態では、該方法は、最尤推定の使用を包含する。ある実施形態では、本開示の方法は、試料中の胎児起源または胎盤起源のDNAのパーセントを算出するステップを包含する。ある実施形態では、異数性を呼び出すための閾値は、算出されたパーセント胎児DNAに基づいて適応調整する。いくつかの実施形態では、DNAの混合物中の胎児起源のものであるDNAの百分率を推定するための方法は、母親由来の遺伝物質、および胎児由来の遺伝物質を含む混合試料を得るステップと、胎児の父親由来の遺伝子試料を得るステップと、混合試料中のDNAを測定するステップと、父親の試料中のDNAを測定するステップと、混合試料のDNA測定値、および父親の試料のDNA測定値を使用して、混合試料中の胎児起源のものであるDNAの百分率を算出するステップとを含む。

【0399】

本開示のある実施形態では、混合物中の胎児DNAの割合または胎児DNAの百分率を測定することができる。いくつかの実施形態では、胎児DNAと母系DNAの混合物である母系の血漿試料自体に対して行った遺伝子型決定の測定値のみを使用して割合を算出することができる。いくつかの実施形態では、測定されたか、または別の方法で既知である母親の遺伝子型および／または測定されたか、または別の方法で既知である父親の遺伝子型を使用して割合を算出することもできる。いくつかの実施形態では、母系DNAと胎児DNAの混合物に対して得た測定値を親の状況の知見と一緒に使用して、パーセント胎児DNAを算出することができる。ある実施形態では、特定の対立遺伝子測定値についての確

率についてのモデルを調整するために母集団頻度を使用して胎児DNAの割合を算出することができる。

【0400】

本開示のある実施形態では、胎児の倍数性状態の決定の正確度について信頼度を算出することができる。ある実施形態では、最大の尤度 (H_{major}) の仮説の信頼度を $(1 - H_{major}) / \Sigma$ (全てのH) として算出することができる。仮説の全ての分布が既知である場合、仮説の信頼度を決定することが可能である。親の遺伝子型情報が既知である場合、仮説の全ての分布を決定することが可能である。正倍数性の胎児についての予測されるデータの分布および異数性の胎児についての予測されるデータの分布の知見が既知である場合、倍数性の決定の信頼度を算出することが可能である。親の遺伝子型データが既知である場合、これらの予測される分布を算出することが可能である。ある実施形態では、正常な仮説の周りの検定統計量の分布および異常な仮説の周りの検定統計量の分布の知見を用いて、呼び出しの信頼性を決定すること、ならびに閾値を改良してより信頼できる呼び出しを行うことができる。これは、混合物中の胎児DNAの量および／またはパーセントが低い場合に特に有用である。これは、Z統計量などの検定統計量が、胎児DNAが高いパーセントで存在する場合に対して最適化された閾値に基づいて設けた閾値を超えないことが原因で、実際には異数性である胎児が正倍数性であると見いだされる状況を回避するために役立つ。

【0401】

ある実施形態では、本明細書に開示されている方法を用いて、母系の遺伝物質と胎児の遺伝物質の混合物中の母系の標的染色体および胎児の標的染色体のコピーの数を決定することによって胎児の異数性を決定することができる。この方法は、母系の遺伝物質と胎児の遺伝物質の両方を含む母系の組織を得るステップを伴ってよく、いくつかの実施形態では、この母系の組織は、母系の血液から単離された母系の血漿または組織であってよい。この方法は、上述の母系の組織を加工することによって、前記母系の組織から母系の遺伝物質と胎児の遺伝物質の混合物を得るステップも伴ってよい。この方法は、標的染色体由来の標的配列を含む個々の反応試料および標的染色体由来の標的配列を含まない個々の反応試料を無作為にもたらすために、得られた遺伝物質を複数の反応試料に分配するステップ、例えば、試料に対してハイスループット配列決定を実施するステップも伴ってよい。この方法は、前記個々の反応試料中に存在するまたは存在しない遺伝物質の標的配列を分析して、反応試料中の正倍数性であると推測される胎児の染色体の存在または不在を示すバイナリーの結果の第1の数、および反応試料中の異数性である可能性がある胎児の染色体の存在または不在を示すバイナリーの結果の第2の数をもたらすステップを伴ってよい。例えば、特定の染色体、染色体の特定の領域、特定の遺伝子座または遺伝子座の集合にマッピングされる配列読み取りをカウントするインフォマティクス技法によってバイナリーの結果の数のいずれかを算出することができる。この方法は、集団内の染色体の長さ、染色体の領域の長さまたは遺伝子座の数に基づいてバイナリーの事象の数を規格化するステップを包含し得る。この方法は、反応試料中の正倍数性であると推測される胎児の染色体について、第1の数を用いてバイナリーの結果の数の予測される分布を算出するステップを伴ってよい。この方法は、反応試料中の異数性であることが推測される胎児の染色体についてのバイナリーの結果の数の予測される分布を算出するステップであって、第1の数、および混合物において見いだされる胎児DNAの推定される割合を、例えば、正倍数性であると推測される胎児の染色体についてのバイナリーの結果の数の予測読み取り計数分布に $(1 + n / 2)$ (n は推定される胎児の割合である) を掛けることによって用いて算出するステップを伴ってよい。いくつかの実施形態では、配列読み取りを、バイナリーの結果ではなく、確率的なマッピングで処理することができ、この方法では、より高い正確度がもたらされるが、さらなる計算能力が必要である。胎児の割合は、複数の方法によって推定することができ、そのいくつかは、本開示の他の箇所に記載されている。この方法は、最尤手法を用いて、第2の数が、正倍数性であるまたは異数性である、異数性である可能性がある胎児の染色体に対応するかどうかを決定するステップを伴ってよい。この

10

20

30

40

50

方法は、測定されたデータを考慮して、正確である尤度が最大である仮説に対応する倍数性状態である胎児の倍数性状態を呼び出すステップを包含し得る。

【0402】

最尤モデルを用いて、胎児の倍数性状態を決定する任意の方法の正確度を上昇させることができることに留意されたい。同様に、胎児の倍数性状態を決定する任意の方法について信頼度を算出することができる。最尤モデルを用いることにより、単一仮説棄却法を用いて倍数性の決定を行う任意の方法の正確度が改善される。最尤モデルは、正常な場合と異常な場合の両方について尤度分布を算出することができる任意の方法に用いることができる。最尤モデルを用いることは、倍数性呼び出しについての信頼度を算出する能力を意味する。

10

【0403】

方法のさらなる考察

ある実施形態では、本明細書に開示されている方法では、多型遺伝子座の各対立遺伝子の独立した観察の数の定量的尺度を利用し、ここで、これには対立遺伝子の比を算出するステップは包含されない。これは、遺伝子座の2つの対立遺伝子の比に関する情報をもたらすが、いずれかの対立遺伝子の独立した観察の数を定量化しない、一部のマイクロアレイに基づく方法などの方法とは異なる。当技術分野で公知のいくつかの方法では、独立した観察の数に関する定量的情報をもたらされ得るが、倍数性の決定をもたらす算出には対立遺伝子の比のみを利用し、定量的情報は利用しない。独立した観察の数に関する情報を保持することの重要性を例示するために、2つの対立遺伝子、AおよびBを有する試料の遺伝子座を考慮に入れる。第1の実験では20の対立遺伝子Aおよび20の対立遺伝子Bを観察し、第2の実験では200の対立遺伝子Aおよび200の対立遺伝子Bを観察する。どちらの実験でも比 $(A / (A + B))$ は0.5と等しいが、第2の実験は、第1の実験よりも対立遺伝子AまたはBの頻度の確実性に関するより多くの情報を伝える。当該方法では、対立遺伝子の比を利用するのではなく、定量的データを使用して、各多型遺伝子座における最も可能性が高い対立遺伝子頻度をより正確にモデリングする。

20

【0404】

ある実施形態では、当該方法では、複数の多型遺伝子座からの測定値を総計するための遺伝子モデルを構築して、トリソミーとダイソミーをよりよく区別し、トリソミーの種類も決定する。さらに、当該方法では、遺伝連鎖情報を組み入れて、方法の正確度を増強する。これは、対立遺伝子の比を染色体上の多型遺伝子座の全てにわたって平均する、当技術分野で公知のいくつかの方法とは対照的である。本明細書に開示されている方法により、ダイソミーにおいて予測される対立遺伝子頻度分布、ならびに、減数分裂Iの間の染色体不分離、減数分裂IIの間の染色体不分離、および胎児発生の初期の有糸分裂の間の染色体不分離によって生じるトリソミーが明確にモデリングされる。なぜこれが重要であるかを例示すると、乗換えがなければ、減数分裂Iの間の染色体不分離により、2つの異なる相同体が一方の親から遺伝によって受け継がれたトリソミーがもたらされ、減数分裂IIの間、または胎児発生の初期の有糸分裂の間の染色体不分離により、一方の親由来の同じ相同体の2つのコピーがもたらされることになる。各筋書きにより、各多型遺伝子座において、および、一緒に考慮に入れたすべての物理的に連鎖した遺伝子座（すなわち同じ染色体上の遺伝子座）においても予測される対立遺伝子の異なる頻度がもたらされる。相同体間での遺伝物質の交換をもたらす乗換えにより、遺伝様式がより複雑になるが、当該方法は、遺伝連鎖情報、すなわち組換え率の情報および遺伝子座間の物理的な距離を使用することによってこれに対して適応する。減数分裂I時の染色体不分離と減数分裂IIまたは有糸分裂時の染色体不分離をよりよく区別するために、当該方法では、乗換えの確率の上昇をセントロメアからの距離の増加としてモデルに組み入れる。減数分裂IIおよび有糸分裂時の染色体不分離は、有糸分裂時の染色体不分離により、一般には、1つの相同体の同一またはほぼ同一のコピーがもたらされるが、減数分裂II時の染色体不分離事象の後に存在する2つの相同体は、多くの場合、配偶子形成の間の1つまたは複数の乗換えに起因して異なるという事実によって区別することができる。

30

40

50

【0405】

ある実施形態では、本開示の方法では、ダイソミーを仮定する場合、親のハプロタイプを決定することができない。ある実施形態では、トリソミーの場合、当該方法により、血漿が一方の親由来の2つのコピーを取り、親の相情報は、2つのコピーが問題の親から遺伝によって受け継がれたいづれによっても決定することができないという事実を用いることによって、一方の親または両親のハプロタイプに関する決定を行うことができる。詳細には、子は、親の2つの同じコピー（一致トリソミー）または親の両方のコピー（不一致トリソミー）のいずれかを遺伝によって受け継ぐことができる。各SNPにおいて、一致トリソミーの尤度および不一致トリソミーの尤度を算出することができる。乗換えを考慮している連鎖モデルを使用しない倍数性呼び出し方法では、トリソミーの全体的な尤度を、染色体全体にわたって一致トリソミーおよび不一致トリソミーの単純な重み付けられた平均として算出する。しかし、乗換えが存在する場合にのみ、分離エラーおよび乗換えをもたらす生物学的な機構に起因して、トリソミーは、染色体上で一致から不一致に変化し得る（および逆もまた同じ）。当該方法は、確率的に、乗換えの尤度を考慮に入れ、その結果、乗換えの尤度を考慮に入れない方法よりも正確度が高い倍数性呼び出しがもたらされる。

10

【0406】

ある実施形態では、参照染色体を使用して子の割合およびノイズレベルの量または確率分布を決定する。ある実施形態では、子の割合、ノイズレベル、および／または確率分布を、倍数性の状態が決定される染色体から入手可能な遺伝子情報のみを使用して決定する。当該方法は、参照染色体を伴わずに、ならびに特定の子の割合またはノイズレベルの固定を伴わずに機能する。これは、子の割合および染色体の挙動を較正するために参照染色体由来の遺伝子データが必要な、当技術分野で公知の方法の有意な改善であり、また、それと異なる点である。

20

【0407】

胎児の割合を決定するために参照染色体を必要としないある実施形態では、仮説の決定を以下の通り行う：

$$H^* = \arg \max_H \text{LIK}(D | H) * \text{prior prob}(H)。$$

【0408】

参照染色体を用いるアルゴリズムでは、一般には、参照染色体はダイソミーであると仮定し、次いで、（a）最も可能性が高い子の割合およびランダムなノイズレベルNを、この仮定および参照染色体のデータに基づいて固定し：

30

【0409】

【化29】

$$[cfr^*, N^*] = \underset{cfr, N}{\operatorname{argmax}} \text{LIK}(D(\text{ref. chrom}) | H11, cfr, N)$$

次いで、換算する

$$\text{LIK}(D | H) = \text{LIK}(D | H, cfr^*, N^*)$$

か、または、

40

（b）この仮定および参照染色体のデータに基づいて、子の割合およびノイズレベルの分布を推定する。詳細には、cfrおよびNについてただ1つの値に固定しないが、確率p（cfr, N）を可能性のあるcfr、N値の広い範囲に割り当てる：

$$p(cfr, N) \sim \text{LIK}(D(\text{ref. chrom}) | H11, cfr, N) * \text{prior prob}(cfr, N)$$

式中、prior prob（cfr, N）が特定の子の割合およびノイズレベルの事前確率であり、以前の知見および実験によって決定される。所望であれば、単にcfr, Nの範囲にわたって一様にする。次いで、

【0410】

【化 3 0】

$$\text{LIK}(D|H) = \sum_{\text{cfr}, N} \text{LIK}(D|H, \text{cfr}, N) * p(\text{cfr}, N)$$

と書くことができる。

上記のどちらの方法も優良な結果をもたらす。

【0 4 1 1】

いくつかの場合には、参照染色体を使用することは望ましくない、可能である、または実行可能であることに留意されたい。そのような場合には、各染色体について最良の倍数性呼び出しを別々に導くことが可能である。詳細には：

【0 4 1 2】

【化 3 1】

$$\text{LIK}(D|H) = \sum_{\text{cfr}, N} \text{LIK}(D|H, \text{cfr}, N) * p(\text{cfr}, N|H)$$

$p(\text{cfr}, N|H)$ は、各染色体について別々に、単に参照染色体についてダイソミーであると仮定するのではなく、仮説Hを仮定して、上記の通り決定することができる。この方法を用いて、固定されたノイズのパラメータと子の画分のパラメータ両方を保持すること、いずれかのパラメータを固定すること、または両方のパラメータを各染色体および各仮説について確率的な形態で保持することが可能である。

【0 4 1 3】

DNAの測定、特にDNAの量が少ない場合、またはDNAが混入DNAと混在している場合の測定は、ノイズが入りかつ／または誤りがちである。このノイズにより、正確度が低い遺伝子型データ、および正確度が低い倍数性呼び出しがもたらされる。いくつかの実施形態では、プラットフォームのモデリングまたはいくつかの他のノイズモデリングの方法を用いて、倍数性の決定に対するノイズの有害作用をカウントすることができる。当該方法では、両チャンネルの同時モデル (joint model) を用い、入力DNAの量、DNAの質、および／またはプロトコルの質に起因するランダムなノイズを考慮する。

【0 4 1 4】

これは、遺伝子座における対立遺伝子の強度の比を用いて倍数性の決定を行う、当技術分野で公知のいくつかの方法とは対照的である。この方法は、正確なSNPノイズモデリングを妨げる。詳細には、測定におけるエラーは、一般には、測定されたチャンネル強度比に特異的に依存せず、モデルを、一次元の情報を使用するように縮小する。ノイズ、チャンネルの質およびチャンネルの相互作用の正確なモデリングには、対立遺伝子の比を用いてモデリングすることができない2次元の同時モデルが必要である。

【0 4 1 5】

詳細には、2つのチャンネル情報を、 $f(x, y)$ が $r = x/y$ である比 r に投影すること自体は、正確なチャンネルノイズおよび偏りのモデリングに役立たない。特定のSNPにおけるノイズは、比率の関数ではない、すなわちノイズ $(x, y) \neq f(x, y)$ であるが、実際には、両方のチャンネルの同時関数 (joint function) である。例えば、二項式モデルでは、測定された比率のノイズは $r(1-r)/(x+y)$ の分散を有し、これは、純粋に r の関数ではない。任意のチャンネルの偏りまたはノイズが包含されるモデルでは、SNP i において、観察されたチャンネルX値は $x = a_i X + b_i$ であると仮定し、ここで、 X は真のチャンネル値であり、 b_i は余分のチャンネルの偏りおよびランダムなノイズである。同様に、 $y = c_i Y + d_i$ と仮定する。 $(a_i X + b_i) / (c_i Y + d_i)$ は X/Y の関数ではないので、観察された比 $r = x/y$ では、真の比率 X/Y を正確に予測することまたは残りのノイズをモデリングすることができない。

【0 4 1 6】

本明細書に開示されている方法には、個々の測定チャンネルの全ての同時二項分布を使用したノイズおよび偏りの有効なモデリングのやり方が記載されている。関連性のある方程式は、文書の他の箇所、SNPの挙動を有効に調整するSNP当たりの一貫した偏り、P(good)およびP(ref|bad)、P(mut|bad)について記載されているセクションに見いだすことができる。ある実施形態では、本開示の方法では、対立遺伝子の比のみに依拠する、実施による限定を回避し、その代わりに、挙動を両方のチャンネル計数に基づいてモデリングするベータ二項分布を使用する。

【0417】

ある実施形態では、本明細書に開示されている方法により、全ての利用可能な測定値を使用することによって、母系の血漿中に見いだされる遺伝子データから妊娠中の胎児の倍数性を呼び出すことができる。ある実施形態では、本明細書に開示されている方法により、親の状況のサブセットのみからの測定値を使用することによって、母系の血漿中に見いだされる遺伝子データから妊娠中の胎児の倍数性を呼び出すことができる。当技術分野で公知のいくつかの方法では、親の状況がAA|BB状況からのものである場合、すなわち所与の遺伝子座において親がどちらもホモ接合性であるが、対立遺伝子が異なる場合に測定された遺伝子データのみを使用する。この方法に伴う1つの問題は、AA|BB状況からの多型遺伝子座の割合が小さく、一般には、10%未満であることである。本明細書に開示されている方法のある実施形態では、方法は、親の状況がAA|BBである遺伝子座において行われた母系の血漿の遺伝子測定値を使用しない。ある実施形態では、当該方法では、親の状況がAA|AB、AB|AA、およびAB|ABである多型遺伝子座についてのみ血漿測定値を使用する。

【0418】

当技術分野で公知のいくつかの方法は、両親の遺伝子型が存在するAA|BB状況におけるSNPからの対立遺伝子の比を平均し、これらのSNPにおける平均の対立遺伝子の比から倍数性呼び出しの決定を主張するステップを包含する。この方法は、示差的なSNPの挙動に起因して著しく不正確である。この方法では、両親の遺伝子型が既知であることを仮定することに留意されたい。対照的に、いくつかの実施形態では、当該方法では、親のいずれかの存在を仮定せず、均一なSNPの挙動を仮定しない、同時チャンネル分布モデルを使用する。いくつかの実施形態では、当該方法では、異なるSNPの挙動／重み付けを考慮に入れる。いくつかの実施形態では、当該方法は、一方の親または両親の遺伝子型の知見を必要としない。当該方法がどのようにこれを実現するかは以下の通りである：

いくつかの実施形態では、仮説の対数尤度をSNPごとに決定することができる。特定のSNP i について、胎児の倍数性についての仮説Hおよびパーセント胎児DNA cf を仮定すると、観察されたデータDの対数尤度は、：

【0419】

【化32】

$$LIK(D|H, i) = \log P(D|H, cf, i) = \log \left(\sum_{m, f, c} P(D|m, f, c, H, cf, i) P(c|m, f, H) P(m|i) P(f|i) \right)$$

と定義され、式中、mは可能性のある真の母親の遺伝子型であり、fは可能性のある真の父親の遺伝子型であり、 $m, f \in \{AA, AB, BB\}$ であり、cは、仮説Hを考慮した、可能性のある子の遺伝子型である。詳細には、モノソミーについては $c \in \{A, B\}$ であり、ダイソミーについては $c \in \{AA, AB, BB\}$ であり、トリソミーについては $c \in \{AAA, AAB, ABB, BBB\}$ である。親の遺伝子型データを含めることにより、一般には、より正確な倍数性の決定がもたらされるが、当該方法が良好に機能するために親の遺伝子型データは必須ではないことに留意されたい。

【0420】

当技術分野で公知のいくつかの方法は、母親がホモ接合性であるが、血漿において異な

る対立遺伝子が測定される（ $AA \mid AB$ または $AA \mid BB$ の状況）SNPからの対立遺伝子の比を平均し、これらのSNPにおける平均の対立遺伝子の比から倍数性呼び出しの決定を主張するステップを包含する。この方法は、父系の遺伝子型が入手不可能である場合を意図している。ホモ接合性で反対の父親BBの存在を伴わずに、血漿が特定のSNPにおいてヘテロ接合性であることをどのくらい正確に主張することができるかどうかは疑問であることに留意されたい：子の割合が少ない場合、対立遺伝子Bが存在するように見えるのは、単にノイズの存在であり得、さらに、Bが存在しないように見えるのは、胎児の測定値のうちの単純対立遺伝子のドロップアウトであり得る。さらには、血漿のヘテロ接合性を実際に決定することができる場合には、この方法では、父系トリソミーを区別することができない。詳細には、母親がAAであるSNP、および血漿においていくらかのBが測定されるSNPについて、父親がGGである場合、生じた子の遺伝子型はAGGであり、平均の比は33%のAになる（子の割合＝100%）。しかし、父親がAGである場合には、生じた子の遺伝子型は、一致トリソミーについてはAGGであり得、33%のAの比を与える、または不一致トリソミーについてはAAGであり得、平均の比についてさらに66%のA近くが得られる。多くのトリソミーが乗換えを伴って染色体上にあるとすれば、全体的な染色体は、不一致トリソミーが全くないところと全てが不一致トリソミーであるところとの間のいずれをも有し得、この比率は、33～66%の間のいずれにも変動し得る。通常のダイソミーについては、比率は約50%であるはずである。連鎖モデルまたは平均の正確なエラーモデルを使用しないと、この方法では多くの場合、父系トリソミーが見落とされる。対照的に、本明細書に開示されている方法では、利用可能な遺伝子型の情報および母集団頻度に基づいて、各親の遺伝子型の候補に対して親の遺伝子型の確率を割り当て、親の遺伝子型を明確に必要としない。さらに、本明細書に開示されている方法により、親の遺伝子型データの不在下または存在下でさえもトリソミーを検出することができ、また、連鎖モデルを使用して可能性のある一致トリソミーから不一致トリソミーへの乗換えの点を同定することによって補償することができる。

【0421】

当技術分野で公知のいくつかの方法は、母系の遺伝子型も父系の遺伝子型も未知であるSNPからの対立遺伝子の比を平均するため、および、これらのSNPにおける平均の比から倍数性呼び出しを決定するための方法を主張する。しかし、これらの目的を実現する方法は開示されていない。本明細書に開示されている方法により、そのような状況において正確な倍数性呼び出しを行うことができ、同時確率最尤法を用い、必要に応じて、SNPノイズおよび偏りのモデル、ならびに連鎖モデルを利用する実施化が本文書の他の箇所に開示されている。

【0422】

当技術分野で公知のいくつかの方法は、対立遺伝子の比を平均し、1つまたは少数のSNPにおける平均の対立遺伝子の比からの倍数性呼び出しの決定を主張するステップを包含する。しかし、そのような方法では、連鎖の概念を利用しない。本明細書に開示されている方法にはこれらの欠点がない。

【0423】

DNAの起源を決定するための事前として配列の長さを使用すること

配列の長さの分布は母系DNAと胎児DNAで異なり、胎児の方が一般に短いことが報告されている。本開示のある実施形態では、以前の知見を経験的なデータの形態で用い、母親のDNA（ $P(X \mid \text{母系})$ ）と胎児DNA（ $P(X \mid \text{胎児})$ ）の両方の予測される長さの事前分布を構築することが可能である。長さ x の新しい未確認のDNA配列を考慮すると、母系または胎児のいずれかを考慮した x の事前尤度に基づいて、DNAの所与の配列が母系DNAまたは胎児DNAのいずれかである確率を定めることが可能である。詳細には、 $P(x \mid \text{母系}) > P(x \mid \text{胎児})$ である場合は、DNA配列を母系に分類することができ、 $P(x \mid \text{母系}) = P(x \mid \text{母系}) / [(P(x \mid \text{母系}) + P(x \mid \text{胎児}))]$ であり、 $p(x \mid \text{母系}) < p(x \mid \text{胎児})$ である場合は、DNA配列を胎児に分類することができ、 $P(x \mid \text{胎児}) = P(x \mid \text{胎児}) / [(P(x \mid \text{母系}) + P(x \mid \text{胎児}))]$ である

。本開示のある実施形態では、その試料に対して特異的である母系の配列の長さおよび胎児の配列の長さの分布を、高い確率で母系または胎児に割り当てることができる配列を考慮に入れることによって決定することができ、次いで、その試料に特異的な分布を、その試料についての予測されるサイズ分布として用いることができる。

【0424】

配列決定の費用を最小限にするための可変性の読み取りの深さ

診断薬に関する多くの臨床試験、例えば、Chiuら BMJ 2011年：342巻：c7401頁では、いくつかのパラメータを用いるプロトコルを設定し、次いで、試験における患者のそれぞれに対して同じパラメータを用いて同じプロトコルを実行する。遺伝物質を測定するための方法として配列決定を用いて母親が妊娠中の胎児の倍数性の状態を決定する場合には、1つの関係するパラメータは読み取りの数である。読み取りの数とは、実際の読み取りの数、意図された読み取りの数、シーケンサーの分割レーン、完全なレーンまたは完全なフローセルを指し得る。これらの試験では、読み取りの数は、一般には、全てまたはほぼ全ての試料が正確度の所望のレベルを実現することを確実にするレベルで設定する。配列決定は、現在のところ費用のかかる技術であり、マッピング可能な100万の読み取り5回当たりおよそ\$200の費用がかかり、一方価格が下がると、同様のレベルの正確度で作動するが読み取りが少ない配列決定に基づく診断を可能にする任意の方法により、かなりの量の金が必要節約される。

10

【0425】

倍数性の決定の正確度は、一般には、読み取りの数および混合物中の胎児DNAの割合を含めたいくつもの因子に左右される。正確度は、一般には、混合物中の胎児DNAの割合がより多いほどより高い。同時に、正確度は、一般には、読み取りの数がより多いほどより高い。匹敵する正確度で倍数性の状態を決定する2つの場合を伴う状況を有することが可能であり、第1の場合には第2の場合よりも混合物中の胎児DNAの割合がより少なく、第1の場合には第2の場合よりも多くの読み取りが配列決定される。混合物中の胎児DNAの推定される割合を、所与のレベルの正確度を実現するために必要な読み取りの数を決定することにおけるガイドとして使用することが可能である。

20

【0426】

本開示のある実施形態では、試料の集合を、集合内の異なる試料が異なる読み取りの深さに配列決定される場合に実行することができ、試料のそれぞれに対して実行される読み取りの数は、各混合物において算出された胎児DNAの割合を考慮して、所与のレベルの正確度が実現されるように選択する。本開示のある実施形態では、これは、混合物中の胎児DNAの割合を決定するために混合試料の測定を行うことを伴ってよく、この胎児の割合の推定は、配列決定を用いて行うことができ、TaqManを用いて行うことができ、qPCRを用いて行うことができ、SNPアレイを用いて行うことができ、所与の遺伝子座における異なる対立遺伝子を区別することができる任意の方法を用いて行うことができる。胎児の割合を推定することの必要性は、実際の測定データと比較する際に考慮される仮説の集合内の全てのまたは選択された胎児の割合の集合を包含する仮説を含めることによって排除することができる。混合物中の胎児DNAの割合を決定した後、各試料について読み取られる配列の数を決定することができる。

30

40

【0427】

本開示のある実施形態では、妊娠中の女性100人が各人のOBに来診し、抗lysan t (anti-lysan t) および／またはDNAアーゼを不活化するものが入った血液チューブ中に各人の血液を採取する。該女性はそれぞれ、自身が妊娠中の胎児の父親が唾液試料を提供するためのキットを家に持ち帰る。100組の夫婦全てについての両者の遺伝物質の集合を検査室に送り返し、そこで母親の血液を遠心沈澱させ、血漿だけでなく Buffyコートも単離する。血漿は、母系DNAならびに胎盤に由来するDNAの混合物を含む。母系の Buffyコートおよび父系の血液についてSNPアレイを使用して遺伝子型決定し、母系の血漿試料中のDNAを、SURESELECTハイブリダイゼーションプロンプを用いて標的とする。プロンプを用いてプルダウンされたDNAを使用して、

50

母体試料のそれぞれに対するものであり、各試料に異なるタグでタグ付けした、タグ付けしたライブラリーを100個生成する。各ライブラリーからの小部分を取り出し、それらの小部分のそれぞれを一緒に混合し、ILLUMINA HISEQ DNAシーケンサーの2つのレーンに多重様式で加え、各レーンからおおよそ5,000万のマッピング可能な読み取りがもたらされ、100の多重化混合物においておおよそ1億のマッピング可能な読み取り、または試料当たりおおよそ100万の読み取りがもたらされた。配列読み取りを使用して、各混合物中の胎児DNAの割合を決定した。50の試料が、混合物中15%超の胎児DNAを有し、100万の読み取りが、99.9%の信頼度で胎児の倍数性の状態を決定するために十分であった。

【0428】

10

残りの混合物のうち、25個が10%から15%の間の胎児DNAを有し、これらの混合物から調製された関連性のあるライブラリーのそれぞれの小部分を多重化し、HISEQの1つのレーンに流し、各試料についてさらなる200万の読み取りを生成した。10%から15%の間の胎児DNAを有する混合物のそれぞれについての配列データの2つの集合を一緒に加え、生じた試料当たり300万の読み取りは、それらの胎児の倍数性の状態を99.9%の信頼度で決定するために十分であった。

【0429】

残りの混合物のうち、13個が6%から10%の間の胎児DNAを有し、これらの混合物から調製された関連性のあるライブラリーのそれぞれの画分を多重化し、HISEQの1つのレーンに流し、各試料についてさらなる400万の読み取りを生成した。6%から10%の間の胎児DNAを有する混合物のそれぞれについての配列データの2つの集合を一緒に加え、生じた混合物当たり500万の総読み取りは、それらの胎児の倍数性の状態を99.9%の信頼度で決定するために十分であった。

20

【0430】

残りの混合物のうち、8つが4%から6%の間の胎児DNAを有し、これらの混合物から調製された関連性のあるライブラリーのそれぞれの小部分を多重化し、HISEQの1つのレーンに流し、各試料についてさらなる600万の読み取りを生成した。4%から6%の間の胎児DNAを有する混合物のそれぞれについての配列データの2つの集合を一緒に加え、生じた混合物当たり700万の総読み取りは、それらの胎児の倍数性の状態を99.9%の信頼度で決定するために十分であった。

30

【0431】

残りの4つの混合物のうちの全てが2%から4%の間の胎児DNAを有し、これらの混合物から調製された関連性のあるライブラリーのそれぞれの小部分を多重化し、HISEQの1つのレーンに流し、各試料についてさらなる1200万の読み取りを生成した。2%から4%の間の胎児DNAを有する混合物のそれぞれについての配列データの2つの集合を一緒に加え、生じた混合物当たり1,300万の総読み取りは、それらの胎児の倍数性の状態を99.9%の信頼度で決定するために十分であった。

【0432】

この方法では、試料100個にわたって99.9%の正確度を実現するために、HISEQ機械で配列決定するための6レーンが必要であった。あらゆる試料に対して同じ数の実行が必要である場合、あらゆる倍数性の決定が99.9%の正確度で行われることを確実にするためには、配列決定に25レーンが取られ、4%の呼び出しなしの比または誤差率を許容する場合、14レーンの配列決定で実現することができた。

40

【0433】

未加工の遺伝子型決定データの使用

母系の血液に見いだされる胎児DNAにおいて測定された胎児の遺伝子情報を使用してNPDを実現できるいくつかの方法が存在する。これらの方法のいくつかは、SNPアレイを使用して胎児DNAの測定を行うことを包含し、いくつかの方法は非標的化配列決定を包含し、いくつかの方法は標的化配列決定を包含する。標的化配列決定ではSNPを標的とすることができ、STRを標的とすることができ、他の多型遺伝子座を標的とする

50

ことができ、非多型の遺伝子座またはそのいくつかの組み合わせを標的とすることができ
る。これらの方法のいくつかは、測定を行う機械のセンサーによってもたらされる強度デ
ータから対立遺伝子の同一性を呼び出す、商業的なまたは専有の対立遺伝子呼び出し元を
使用することを含めてよい。例えば、ILLUMINA INFINIUMシステムまたはAFFYMETRIX
GENECHIPマイクロアレイシステムは、DNAの相補的なセグメントとハイブリダイズする
ことができるDNA配列を付着させたビーズまたはマイクロチップを含み、ハイブリダイ
ゼーションすると、センサー分子の蛍光性が変化し、それを検出することができる。配
列決定方法、例えば、ILLUMINA SOLEXA GENOME SEQUENCERまたはABI SOLID
GENOME SEQUENCERもあり、これは、DNAの断片の遺伝子配列について配列決定し、配
列決定される鎖と相補的なDNAの鎖が伸長すると、伸長したヌクレオチドの同一性が、一
般には、相補的なヌクレオチドに付加した蛍光性タグまたは放射性タグを介して検出され
る。これらの方法の全てにおいて、遺伝子型データまたは配列決定データは、一般には、
蛍光もしくは他のシグナルまたはそれが無いことに基づいて決定される。これらのシステ
ムは、一般には、蛍光または他の検出デバイスのアナログ出力（一次遺伝子データ）から
特定の対立遺伝子の呼び出し（二次遺伝子データ）を行う低レベルのソフトウェアパッケ
ージと組み合わせる。例えば、SNPアレイ上の所与の対立遺伝子の場合には、該ソフト
ウェアにより、蛍光強度がある特定の閾値を上回る、または下回る量である場合に、特
定のSNPが存在するまたは存在しないという呼び出しを行う。同様に、シーケンサーの
出力は、色素のそれぞれについて検出された蛍光のレベルを示すクロマトグラムであり、
該ソフトウェアにより、特定の塩基対が、AもしくはTまたはCもしくはGであるという
呼び出しを行う。ハイスループットシーケンサーにより、一般には、読み取りと称され
る一連のそのような測定が行われ、配列決定された、最も可能性が高いDNA配列の構造
が示される。クロマトグラムの直接的なアナログ出力は、本明細書では一次遺伝子デ
ータであると定義され、該ソフトウェアによって行われる塩基対／SNPの呼び出しは、
本明細書では二次遺伝子データとみなされる。ある実施形態では、一次データとは、
遺伝子型決定プラットフォームの加工されていない出力である生の強度データを指し、
遺伝子型決定プラットフォームとは、SNPアレイまたは配列決定プラットフォーム
を指し得る。二次遺伝子データとは、加工された遺伝子データであって、対立遺伝子
の呼び出しが行われている、または配列データが塩基対に割り当てられている、かつ
／または配列読み取りがゲノムにマッピングされているデータを指す。

【0434】

多くのより高レベルの適用では、これらの対立遺伝子の呼び出し、SNPの呼び出しお
よび配列読み取り、すなわち遺伝子型決定ソフトウェアにより生じる二次遺伝子デ
ータを活用する。例えば、DNA NEXUS、ELANDまたはMAQで配列決定の読み取り
を取得し、それらをゲノムにマッピングする。例えば、非侵襲的な出生前診断との
関連において、複雑なインフォマティクス、例えば、PARENTAL SUPPORT（商標）
は、個体の遺伝子型を決定するための多数のSNPの呼び出しに影響を及ぼし得る。ま
た、着床前遺伝子診断との関連において、ゲノムにマッピングされる配列読み取り
の集合を取得することが可能であり、各染色体または染色体のセクションにマッピ
ングされる正規化された読み取りの計数を取得することにより、個体の倍数性の状
態を決定することが可能であり得る。非侵襲的な出生前診断との関連において、母
系の血漿中に存在するDNAにおいて測定された配列読み取りの集合を取得し、それ
らをゲノムにマッピングすることが可能であり得る。次いで、各染色体または染色
体のセクションにマッピングされる正規化された読み取りの計数を取得し、そのデ
ータを使用して、個体の倍数性の状態を決定することができる。例えば、不釣り
合いに多数の読み取りを有する染色体は、血液を抜き取った母親が妊娠中の胎児
においてトリソミーであると結論づけることが可能であり得る。

【0435】

しかし、実際には、測定計器からの最初の出力は、アナログシグナルである。特定の塩

基対を、配列決定ソフトウェアに関連するソフトウェア、例えば、塩基対Tを呼び出すことができるソフトウェアによって呼び出す場合、実際には、その呼び出しは、該ソフトウェアにより可能性が最も高いと考えられる呼び出しである。しかし、いくつかの場合には、呼び出しは低信頼度であり得、例えば、アナログシグナルにより、特定の塩基対が、Tである可能性が90%だけであり、Aである可能性が10%であることが示され得る。別の例では、SNPアレイリーダーに付随する遺伝子型呼び出しソフトウェアにより、特定の対立遺伝子がGであることが呼び出され得る。しかし、実際には、基礎をなすアナログシグナルにより、対立遺伝子がGである可能性が70%だけであり、対立遺伝子がTである可能性が30%であることが示され得る。これらの場合には、より高レベルの適用においてより低レベルのソフトウェアによって行った遺伝子型の呼び出しおよび配列の呼び出しを用いる場合、一部の情報が失われる。すなわち、遺伝子型決定プラットフォームによって直接測定される一次遺伝子データは、添付のソフトウェアパッケージによって決定される二次遺伝子データよりも厄介であり得るが、それはより多くの情報を含む。二次遺伝子データ配列のゲノムへのマッピングにおいて、一部の塩基が十分に明瞭に読み取られていないので、または、マッピングが明瞭ではないので、多くの読み取りが捨てられる。一次遺伝子データ配列読み取りを使用する場合、二次遺伝子データ配列読み取りに最初に変換された際に捨てられた可能性がある読み取りの全てまたはその多くを、読み取りを確率的に処理することによって使用することができる。

10

【0436】

本開示のある実施形態では、より高レベルのソフトウェアは、より低いレベルのソフトウェアによって決定される対立遺伝子の呼び出し、SNPの呼び出しまたは配列読み取りに依拠しない。その代わりに、より高レベルのソフトウェアは、遺伝子型決定プラットフォームから直接測定されたアナログシグナルにその算出の基礎を置く。本開示のある実施形態では、インフォマティクスに基づく方法、例えば、PARENTAL SUPPORT（商標）を、胚／胎児／子の遺伝子データを再構築するその能力が、遺伝子型決定プラットフォームによって測定された一次遺伝子データを直接使用するように工学的に操作されるように改変する。本開示のある実施形態では、インフォマティクスに基づく方法、例えば、PARENTAL SUPPORT（商標）では、一次遺伝子データを使用し、二次遺伝子データを使用せずに、対立遺伝子の呼び出し、および／または染色体コピー数の呼び出しを行うことができる。本開示のある実施形態では、遺伝子の呼び出し、SNPの呼び出し、すべての配列読み取り、配列マッピングを、一次遺伝子データを二次的な遺伝子の呼び出しに変換するのではなく、遺伝子型決定プラットフォームによって直接測定された生の強度データを使用することによって確率的に処理する。ある実施形態では、対立遺伝子数の確率を算出するステップおよび各仮説の相対的確率を決定するステップにおいて使用する調製された試料からのDNA測定値は、一次遺伝子データを含む。

20

30

【0437】

いくつかの実施形態では、該方法により、少なくとも1つの関連する個体の遺伝子データを組み入れる、標的個体の遺伝子データの正確度を上昇させることができ、該方法は、標的個体のゲノムに特異的な一次遺伝子データおよび関連する個体（複数可）のゲノム（複数可）に特異的な遺伝子データを取得するステップと、関連する個体（複数可）由来のどの染色体のセグメントが、標的個体のゲノム内のそれらのセグメントに対応する可能性があるかに関する1つまたは複数の仮説の集合を作製するステップと、標的個体の一次遺伝子データおよび関連する個体（複数可）の遺伝子データを考慮して仮説のそれぞれの確率を決定するステップと、各仮説に関連する確率を用いて、実際の標的個体の遺伝物質の最も可能性が高い状態を決定するステップとを含む。いくつかの実施形態では、該方法により、標的個体のゲノム内の染色体のセグメントのコピーの数を決定することができ、該方法は、どのくらいの染色体セグメントのコピーが標的個体のゲノム内に存在するかに関するコピー数についての仮説の集合を作製するステップと、標的個体からの一次遺伝子データおよび1つまたは複数の関連する個体からの遺伝子情報をデータ集合に組み入れるステップと、データ集合に関連するプラットフォームの応答の特性を推定するステップであって

40

50

、プラットフォームの応答が、ある実験と別の実験で変動し得るステップと、データ集合およびプラットフォームの応答特性を考慮して、コピー数についての仮説のそれぞれの相対的確率を計算するステップと、最も可能性の高いコピー数についての仮説に基づいて染色体セグメントのコピー数を決定するステップとを含む。ある実施形態では、本開示の方法は、標的個体の少なくとも1つの染色体の倍数性の状態を決定することができ、該方法は、標的個体から、および1つまたは複数の関連する個体から、一次遺伝子データを得るステップと、標的個体の染色体のそれぞれについて、少なくとも1つの倍数性の状態についての仮説の集合を作製するステップと、1つまたは複数の専門技法を用いて、集合内の倍数性の状態についての仮説のそれぞれの統計的確率を決定するステップと、使用した専門技法のそれぞれについて、得られた遺伝子データを考慮して、倍数性の状態についての仮説のそれぞれについて、1つまたは複数の専門技法によって決定された統計的確率を組み合わせるステップと、標的個体の染色体のそれぞれについて、倍数性の状態についての仮説のそれぞれについての複合統計確率に基づいて倍数性の状態を決定するステップとを含む。ある実施形態では、本開示の方法は、対立遺伝子の集合において、標的個体において、および標的個体の一方の親または両親から、および必要に応じて、1つまたは複数の関連する個体から、対立遺伝子の状態を決定することができ、該方法は、標的個体から、および一方の親または両親から、および任意の関連する個体から一次遺伝子データを得るステップと、標的個体について、および一方の親または両親について、および必要に応じて、1つまたは複数の関連する個体について、少なくとも1つの対立遺伝子についての仮説の集合を作製するステップであって、仮説が対立遺伝子の集合における可能性のある対立遺伝子の状態を記載するステップと、仮説の集合内の各対立遺伝子についての仮説について統計的確率を、得られた遺伝子データを考慮して決定するステップと、対立遺伝子の集合内の対立遺伝子のそれぞれについて、標的個体について、および一方の親または両親について、および必要に応じて、1つまたは複数の関連する個体について、対立遺伝子についての仮説のそれぞれの統計的確率に基づいて対立遺伝子の状態を決定するステップとを含む。

【0438】

いくつかの実施形態では、混合試料の遺伝子データは、配列データを含んでよく、ここで、配列データは、ヒトゲノムに独自にマッピングされない場合がある。いくつかの実施形態では、混合試料の遺伝子データは、配列データを含んでよく、ここで、配列データは、ゲノム内の複数の場所にマッピングされ、ここで、可能性のあるマッピングのそれぞれは、所与のマッピングが正確である確率を伴う。いくつかの実施形態では、配列読み取りは、ゲノム内の特定の位置に関連づけられると仮定されない。いくつかの実施形態では、配列読み取りは、ゲノム内の複数の位置に関連づけられ、付随する確率はその位置に属する。

【0439】

出生前診断の組み合わせ方法

異数性または他の遺伝的欠陥を出生前診断または出生前スクリーニングするために用いることができる多くの方法が存在する。本文書の他の箇所、ならびに、2006年11月28日に出願された米国実用新案出願第11/603,406号；2008年3月17日に
出願された米国実用新案出願第12/076,348号、およびPCT実用新案出願第PCT/S09/52730号に、関連する個体の遺伝子データを使用して、胎児などの標的個体の遺伝子データが公知であるまたは推定される正確度を上昇させる方法の1つが記載されている。出生前診断のために用いる他の方法は、母系の血液中の、種々の遺伝子の異常と相関する特定のホルモンのレベルを測定するステップを包含する。この例はトリプルテストと称される、母系の血液中のいくつか（一般に、2つ、3つ、4つまたは5つ）の異なるホルモンのレベルを測定する検査である。複数の方法を用いて所与の転帰の尤度を決定し、どの方法もそれ自体が決定的でない場合には、これらの方法によって生じる情報を組み合わせて、個々の方法のいずれよりも正確な予測を行うことが可能である。トリプルテストでは、3つの異なるホルモンから生じる情報を組み合わせることにより、個

々のホルモンレベルにより予測され得るよりも正確な遺伝子の異常の予測がもたらされ得る。

【0440】

本明細書には、胎児の遺伝子の状態、詳細には、胎児における遺伝子の異常の可能性に関してより正確な予測を行うための方法であって、種々の方法を用いて行った胎児における遺伝子の異常の予測を組み合わせること含む方法が開示されている。「より正確な」方法とは、所与の偽陽性率において、偽陰性率がより低い、異常を診断するための方法を指し得る。好ましい本開示の実施形態では、予測のうちの1つまたは複数、胎児に関する公知の遺伝子データに基づいて行い、遺伝子の知見はPARENTAL SUPPORT (商標)法を使用して決定した、すなわち、胎児に関連する個体の遺伝子データを使用して、胎児の遺伝子データをより高い正確度で決定した。いくつかの実施形態では、遺伝子データは、胎児の倍数性の状態を含んでよい。いくつかの実施形態では、遺伝子データとは、胎児のゲノムにおける対立遺伝子の呼び出しの集合を指し得る。いくつかの実施形態では、予測のいくつかは、トリプルテストを用いて行われた。いくつかの実施形態では、予測のいくつかを、母系の血液中の他のホルモンレベルの測定値を用いて行った。いくつかの実施形態では、診断を考慮する方法によって行われる予測を、スクリーニングを考慮する方法によって行われる予測と組み合わせることができる。いくつかの実施形態では、該方法は、アルファフェトプロテイン (AFP) の母系の血中レベルを測定するステップを包含する。いくつかの実施形態では、該方法は、コンジュゲートしていないエストリオール (UE_3) の母系の血中レベルを測定するステップを包含する。いくつかの実施形態では、該方法は、ベータヒト絨毛性ゴナドトロピン (ベータ-hCG) の母系の血中レベルを測定するステップを包含する。いくつかの実施形態では、該方法は、浸潤性トロホブラスト抗原 (IT A) の母系の血中レベルを測定するステップを包含する。いくつかの実施形態では、該方法は、インヒビンの母系の血中レベルを測定するステップを包含する。いくつかの実施形態では、該方法は、妊娠関連血漿タンパク質 A (PAPP-A) の母系の血中レベルを測定するステップを包含する。いくつかの実施形態では、該方法は、他のホルモンまたは母系の血清マーカーの母系の血中レベルを測定するステップを包含する。いくつかの実施形態では、予測のいくつかは、他の方法を用いて行なわれてもよい。いくつかの実施形態では、予測のいくつかは、完全に組み込まれた検査、例えば、妊娠の約12週における超音波検査および血液検査ならびに約16週における第2の血液検査を組み合わせた検査を用いて行なわれてもよい。いくつかの実施形態では、該方法は、胎児の項部浮腫 (NT) を測定するステップを包含する。いくつかの実施形態では、該方法は、予測を行うために、測定された上述のホルモンのレベルを使用するステップを包含する。いくつかの実施形態では、該方法は、上述の方法の組み合わせを包含する。

【0441】

予測を組み合わせるための多くのやり方が存在し、例えば、ホルモンの測定値を、中央値の倍数 (multiple of the median) (MoM) に変換し、次いで、尤度比 (LR) に変換することができる。同様に、他の測定値を、NT分布の混合モデルを使用してLRに変換することができる。NTおよび生化学的マーカーについてのLRに、年齢および妊娠に関連するリスクを掛けて、21トリソミーなどの種々の状態に対するリスクを導くことができる。検出率 (DR) および偽陽性率 (FPR) を、所与のリスク閾値を上回るリスクを有する割合を取ることによって算出することができる。

【0442】

ある実施形態では、倍数性の状態を呼び出すための方法は、同時分布モデルおよび対立遺伝子数の確率を用いて決定される倍数性についての仮説のそれぞれの相対的確率と、これらに限定されないが、読み取り数解析、ヘテロ接合率の比較、親の遺伝子情報を使用する場合にのみ利用可能な統計量、特定の親の状況に対して正規化された遺伝子型シグナルの確率、第1の試料または調製された試料における推定される胎児の割合を用いて算出される統計量、およびそれらの組み合わせを含めた、胎児がトリソミーであるリスクスコアを決定する他の方法から選択される統計学的技法を用いて算出された倍数性についての仮

説のそれぞれの相対的確率とを組み合わせるステップを包含する。

【0443】

別の方法は、4つの測定されたホルモンレベルを伴う状況であって、これらのホルモンのまわりの確率分布が既知である状況を包含し得：正倍数性の場合には $p(x_1, x_2, x_3, x_4 | e)$ 、異数性の場合には $p(x_1, x_2, x_3, x_4 | a)$ である。次いで、DNA測定値についての確率分布を測定することができ、正倍数性の場合および異数性の場合、それぞれ $g(y | e)$ および $g(y | a)$ である。これらは、正倍数性／異数性の仮定を考慮して、独立していると仮定すると、 $p(x_1, x_2, x_3, x_4 | a) g(y | a)$ および $p(x_1, x_2, x_3, x_4 | e) g(y | e)$ として組み合わせ、次いで、母系の年齢を考慮して、それぞれに事前 $p(a)$ および $p(e)$ を掛けることができる。次いで、最も高いものを選択することができる。

10

【0444】

ある実施形態では、中心極限定理を惹起して、 $g(y | a)$ または e の分布がガウス分布であると仮定し、多数の試料について調べることによって平均値および標準偏差を測定することが可能である。別の実施形態では、転帰を考慮して、これらが独立していないと仮定し、同時分布 $p(x_1, x_2, x_3, x_4 | a)$ または e を推定するために十分な試料を収集することができる。

【0445】

ある実施形態では、標的個体が倍数性の状態であると決定されるための倍数性の状態は、最大の確率を有する仮説に関連づけられる。いくつかの場合には、1つの仮説は、90%超の正規化された複合確率を有する。各仮説は1つの倍数性の状態または倍数性の状態の集合に関連づけられ、その正規化された複合確率が90%超であるかまたはいくつかの他の閾値、例えば、50%、80%、95%、98%、99%または99.9%を超えている仮説に関連づけられる倍数性の状態を、決定された倍数性の状態として仮説が呼び出されるのに必要な閾値として選択することができる。

20

【0446】

母系の血液中の、以前の妊娠由来の子由来のDNA

非侵襲的な出生前診断の1つの難しさは、現行の妊娠由来の胎児の細胞と以前の妊娠由来の胎児の細胞を鑑別することである。一部では、前の妊娠由来の遺伝物質 (genetic matter) はいくらかの時間の後に消えると考えられているが、決定的な証拠は示されていない。本開示のある実施形態では、PARENTAL SUPPORT (商標) (PS) 法、および父系のゲノムの知見を使用して、母系の血液中に存在する父系起源の胎児DNA (すなわち胎児が父親から遺伝によって受け継いだDNA) を決定することが可能である。この方法では、相が特定された親の遺伝子情報を利用することができる。相が特定されていない遺伝子型の情報から、祖父母の遺伝子データ (例えば、祖父の精子から測定された遺伝子データ) を使用して、または、他の生まれた子からの遺伝子データ、または流産の試料から親の遺伝子型の相を特定することが可能である。父系の細胞のHi-C Mapに基づく相の特定またはハプロタイピングによって、相が特定されていない遺伝子情報の相を特定することもできる。上首尾のハプロタイピングは、染色体が緊密な束である有糸分裂の相にある細胞を静止させ、マイクロfluidicsを使用して別々の染色体を別々のウェルに入れることによって実証されている。別の実施形態では、相が特定された親のハプロタイプデータを使用して、父親由来の2種以上の相同体の存在を検出することが可能であり、これは、2人以上の子由来の遺伝物質が血液中に存在することを意味する。胎児において正倍数性であることが予測される染色体に焦点を当てることにより、胎児がトリソミーを患っている可能性を除外することができる。また、胎児DNAが現在の父親由来でないかどうかを決定することが可能であり、その場合、他の方法、例えば、トリプルテストを用いて遺伝子の異常を予測することができる。

30

40

【0447】

採血以外の方法によって入手可能な、胎児の遺伝物質の他の供給源があり得る。母系の血液において入手可能な胎児の遺伝物質の場合には、2つの主要なカテゴリー：(1) 胎

50

児の細胞全体、例えば、胎児有核赤血球または赤芽球、および（２）浮動性胎児DNAがある。胎児の細胞全体の場合には、胎児の細胞が母系の血液中で長期間存続することができ、したがって、妊娠中の女性から、前の妊娠由来の子または胎児由来のDNAを含有する細胞を単離することが可能であるといういくつかの証拠が存在する。浮動性胎児DNAは、数週間のうちに系から取り除かれるという証拠も存在する。１つの難題は、その遺伝物質が細胞に含有される個体の同一性をどのように決定するか、すなわち、測定された遺伝物質が前の妊娠由来の胎児由来でないことをどのように確実にするかである。本開示のある実施形態では、母系遺伝物質の知見を用いて、問題の遺伝物質が母系遺伝物質ではないことを確実にすることができる。本文書または本文書において参照されているいずれかの特許に記載の通り、インフォマティクスに基づく方法、例えば、PARENTAL SUPPORT（商標）を含めた、この目的を実現するためのいくつかの方法が存在する。

10

【0448】

本開示のある実施形態では、妊娠中の母親から抜き取った血液を、浮動性胎児DNAを含む画分、および有核赤血球を含む画分に分離することができる。浮動性DNAは必要に応じて富化することができ、DNAの遺伝子型の情報を測定することができる。浮動性DNAから測定された遺伝子型の情報から、母系の遺伝子型の知見を使用して、胎児の遺伝子型の態様を決定することができる。これらの態様は、倍数性の状態、および／または対立遺伝子の集合の同一性を指し得る。次いで、個々の有核赤血球について、本文書の他の箇所および他の参考特許に記載されている方法、特に本文書の最初のセクションに記載の方法を用いて遺伝子型決定することができる。母系ゲノムの知見により、任意の所与の単一の血液細胞が遺伝的に母系かどうかを決定することが可能になる。また、上記の通り決定された胎児の遺伝子型の態様により、単一の血液細胞が、現在妊娠中の胎児に遺伝的に由来するかどうかを決定することが可能になる。本質的に、本開示のこの態様により、母親の遺伝子の知見、および場合によっては他の関連する個体、例えば、父親からの遺伝子情報を、母系の血液に見いだされる浮動性DNAから測定された遺伝子情報と一緒に使用して、母系の血液に見いだされる単離された有核細胞が、（a）遺伝的に母系であるか、（b）遺伝的に現在妊娠中の胎児由来であるか、または（c）遺伝的に前の妊娠由来の胎児由来であるかのいずれかを決定することが可能になる。

20

【0449】

出生前の性染色体異数性の決定

30

当技術分野で公知の方法では、妊娠中の胎児の性別を、母親の血液から決定することを試みる人は、胎児の浮動性DNA（fffDNA）が母親の血漿中に存在するという事実を用いている。母系の血漿中のY特異的遺伝子座を検出することができれば、これは、妊娠中の胎児が男であることを意味する。しかし、先行技術で公知の方法を用いる場合、いくつかの場合には、fffDNAの量が、男の胎児の場合にY特異的遺伝子座が検出されることを確実にするには低すぎるので、血漿中のY特異的遺伝子座が検出されないことでは、妊娠中の胎児が女であることは必ずしも保証されない。

【0450】

本明細書では、Y特異的核酸、すなわち排他的に父系的に由来する遺伝子座由来であるDNAを測定することを必要としない新規の方法が提示される。以前に開示されたParental Support方法では、乗換え頻度データ、親の遺伝子型データ、および妊娠中の胎児の倍数性の状態を決定するためのインフォマティクス技法を用いる。胎児の性別は、単に性染色体における胎児の倍数性の状態である。XXである子は女であり、およびXYは男である。本明細書に記載の方法により、胎児の倍数性の状態を決定することもできる。性判別は性染色体の倍数性の決定と有効に同義であることに留意されたい；性判別の場合には、仮定は、多くの場合、子が正倍数性であるとして立てられ、したがって、可能性のある仮説が少ない。

40

【0451】

本明細書に開示されている方法は、X染色体とY染色体の両方に共通する遺伝子座について調べて、胎児について予測される存在する胎児DNAの量に関するベースラインを作

50

製するステップを包含する。次いで、X染色体のみに特異的な領域を調べて、胎児が女であるか男であるかを決定することができる。男の場合には、X染色体に特異的な遺伝子座由来の胎児DNAが、XとYの両方に特異的な遺伝子座由来の胎児DNAよりも少ないと認められることが予想される。対照的に、女の胎児では、これらの群のそれぞれのDNAの量が同じであることが予想される。問題のDNAは、試料に存在するDNAの量を定量することができる任意の技法、例えば、qPCR、SNPアレイ、遺伝子型決定アレイまたは配列決定によって測定することができる。排他的に1個体に由来するDNAについては、以下が認められることが予想される：

【0452】

【表4】

	Xに特異的なDNA	XおよびYに特異的なDNA	Yに特異的なDNA
男 (XY)	A	2A	A
女 (XX)	2A	2A	0

10

胎児由来のDNAが母親由来のDNAと混在しており、混合物中の胎児DNAの割合がFであり、混合物中の母系DNAの割合がMであり、したがって $F + M = 100\%$ である場合、以下が認められることが予想される：

【0453】

【表5】

	Xに特異的なDNA	XおよびYに特異的なDNA	Yに特異的なDNA
男の胎児 (XY)	$M + \frac{1}{2}F$	$M + F$	$\frac{1}{2}F$
女の胎児 (XX)	$M + F$	$M + F$	0

20

FおよびMが既知である場合には、予測比を計算することができ、観察されたデータを、予測データと比較することができる。MおよびFが未知である場合には、閾値を過去のデータに基づいて選択することができる。どちらの場合でも、XとYの両方に特異的な遺伝子座において測定されたDNAの量をベースラインとして用いることができ、胎児の性別の検査は、X染色体のみに特異的な遺伝子座において観察されたDNAの量に基づいてよい。その量がベースラインよりも、およそ $1/2F$ と等しい量だけ低い、またはそれが予め定義された閾値未満になる量だけ低ければ、胎児は男であることが決定され、その量がベースラインとほとんど等しい、またはそれが予め定義された閾値未満になる量だけ低くなければ、胎児は女であることが決定される。

30

【0454】

別の実施形態では、多くの場合Z染色体と称される、X染色体とY染色体に共通である遺伝子座のみを調べることができる。Z染色体上の遺伝子座のサブセットは、一般には、常にX染色体上のA、およびY染色体上のBである。Z染色体由来のSNPがB遺伝子型を有することが見いだされた場合は、胎児は男であると呼び出され、Z染色体由来のSNPがA遺伝子型のみを有することが見いだされた場合には、胎児は女であると呼び出される。別の実施形態では、X染色体においてのみ見いだされる遺伝子座を調べることができる。AA|Bなどの状況は、Bが存在することにより、胎児が父親由来のX染色体を有することが示されるので、特に情報価値がある。AB|Bなどの状況も、女の胎児の場合には、男の胎児と比較して、多くの場合、Bが半分しか存在しないことが認められると予想されるので、情報価値がある。別の実施形態では、対立遺伝子AとBの両方がX染色体とY染色体の両方に存在し、どのSNPが父系のY染色体由来であるか、およびどれが父系

40

50

のX染色体由来であるかが既知であるZ染色体上のSNPを調べることができる。

【0455】

ある実施形態では、Y染色体とX染色体によって共有される相同な非組換え（HNR）領域間で変動することが公知の一塩基位置を増幅することが可能である。このHNR領域内の配列は、X染色体とY染色体の間でほとんど同一である。この同一の領域内に、母集団内のX染色体の間およびY染色体の間では不変であるが、X染色体とY染色体の間では異なる一塩基位置がある。各PCRアッセイによりX染色体とY染色体の両方に存在する遺伝子座由来の配列を増幅することができる。増幅された配列のそれぞれの内部に、配列決定またはいくつかの他の方法を用いて検出することができる単一の塩基がある。

【0456】

ある実施形態では、胎児の性別を、母系の血漿中に見いだされる胎児の浮動性DNAから決定することができ、方法は以下のステップの一部または全部を含む：1）HNR領域内のX/Y変異体一塩基位置を増幅するPCR（通常のPCRまたはmini-PCRのいずれか、所望であればそれに加えて多重化）プライマーを設計するステップ、2）母系の血漿を得るステップ、3）母系の血漿由来の標的を、HNR X/Y PCRアッセイを用いてPCR増幅するステップ、4）アンプリコンについて配列決定するステップ、5）配列データを、増幅された配列のうちの1つまたは複数の内部のY対立遺伝子の存在について検査するステップ。1つまたは複数の存在により、男の胎児が示される。全てのアンプリコン由来の全てのY対立遺伝子が存在しないことにより、女の胎児が示される。

【0457】

ある実施形態では、標的化配列決定を用いて、母系の血漿中のDNAおよび／または親の遺伝子型を測定することができる。ある実施形態では、父系的に供給されたDNAを起源とすることが明白な配列を全て無視することができる。例えば、状況AA | ABでは、A配列の数を計数し、B配列の全てを無視することができる。上記のアルゴリズムについてヘテロ接合性率を決定するために、所与のプロープについて、観察されたA配列の数と総配列の予測数を比較することができる。試料ごとに各プロープについて配列の予測数を算出することができる多くのやり方がある。ある実施形態では、過去のデータを使用して、全ての配列読み取りのどの画分が特異的なプロープのそれぞれに属するかを決定し、次いで、この経験的な画分を配列読み取りの総数と組み合わせて使用して、各プロープにおける配列の数を推定することが可能である。別の手法では、一部の公知のホモ接合性の対立遺伝子を標的とし、次いで、過去のデータを使用して、各プロープにおける読み取りの数と既知のホモ接合性の対立遺伝子における読み取りの数を関連づけることができる。次いで、各試料について、ホモ接合性の対立遺伝子における読み取りの数を測定し、次いで、この測定値を経験的に導かれた関連性と一緒を使用して、各プロープにおける配列読み取りの数を推定することができる。

【0458】

いくつかの実施形態では、複数の方法によって行われた予測を組み合わせることによって胎児の性別を決定することが可能である。いくつかの実施形態では、複数の方法は、本開示に記載の方法から選択される。いくつかの実施形態では、複数の方法の少なくとも1つは本開示に記載の方法から選択される。

【0459】

いくつかの実施形態では、本明細書に記載の方法を用いて、妊娠中の胎児の倍数性の状態を決定することができる。ある実施形態では、倍数性呼び出し方法では、X染色体に特異的な遺伝子座またはX染色体とY染色体の両方に共通する遺伝子座を使用するが、いかなるY特異的な遺伝子座も使用しない。ある実施形態では、倍数性呼び出し方法では、以下の1つまたは複数を使用する：X染色体に特異的な遺伝子座、X染色体とY染色体の両方に共通する遺伝子座、およびY染色体に特異的な遺伝子座。ある実施形態では、性染色体の比が同様である場合、例えば、45, X（ターナー症候群）、46, XX（正常な女性）および47, XXX（Xトリソミー）、鑑別は、種々の仮説に従って対立遺伝子分布と予測される対立遺伝子分布とを比較することによって実現することができる。別の実施形

10

20

30

40

50

態では、これは、性染色体についての配列読み取りの相対的な数と、正倍数性であることが仮定される1つまたは複数の参照染色体とを比較することによって実現することができる。これらの方法は、異数性の場合を含むように拡大することができることにも留意されたい。

【0460】

単一遺伝子疾患スクリーニング

ある実施形態では、胎児の倍数性の状態を決定するための方法は、単一遺伝子障害についての同時検査が可能になるように拡張することができる。単一遺伝子疾患の診断は、異数性試験のために用いる同じ標的化手法に影響を及ぼし、さらなる特異的な標的を必要とする。ある実施形態では、単一遺伝子NP D診断は連鎖解析による。多くの場合、c f D N A試料の直接的な試験は、母系DNAが存在することにより、胎児が母親の変異を遺伝によって受け継いだかどうかを決定することが実質的に不可能になるので、信頼できない。独自の父系的に由来する対立遺伝子を検出することは困難が少ないが、疾患が優性であり、父親が保有する場合にのみ完全に情報価値があり、それによりこの手法の有用性が限定される。ある実施形態では、該方法は、PCRまたは関連する増幅手法を包含する。

【0461】

いくつかの実施形態では、該方法は、親において、周囲に非常にしっかりと連鎖したSNPがある異常な対立遺伝子について、第一度近親者からの情報を用いて相を特定するステップを包含する。次いで、これらのSNPから得られた標的化配列決定データに対してParental Supportを実行して、正常な相同体または異常な相同体のいずれを、両親から胎児が遺伝によって受け継いだかを決定することができる。SNPが十分に連鎖している限りは、胎児の遺伝子型の遺伝を非常に確実に決定することができる。いくつかの実施形態では、方法は、(a) 日常病の特定の集合を密に隣接させるためのSNP遺伝子座の集合を、異数性を試験するための本発明者らの多重プールに付加するステップと；(b) 正常な対立遺伝子および異常な対立遺伝子を有する、付加したこれらのSNPから、種々の近親者からの遺伝子データに基づいて、対立遺伝子について確実に相の特定をするステップと、(c) 疾患遺伝子座の周囲の領域内の遺伝によって受け継がれた母系の相同体および父系の相同体における胎児のディプロタイプまたは相が特定されたSNP対立遺伝子の集合を再構築して、胎児の遺伝子型を決定するステップとを含む。いくつかの実施形態では、疾患に関連づけられる遺伝子座に密接に関連するさらなるプローブを、異数性試験のために用いる多型遺伝子座の集合に加える。

【0462】

試料は母系DNAと胎児DNAの混合物であるので、胎児のディプロタイプを再構築することは困難である。いくつかの実施形態では、該方法では、近親者の情報を組み入れて、SNPおよび疾患対立遺伝子の相を特定し、次いで、場所特異的な組換え尤度からのSNPおよび組換えデータと母系の血漿の遺伝子測定値から観察されたデータの物理的な距離を考慮に入れて、最も可能性が高い胎児の遺伝子型を得る。

【0463】

ある実施形態では、疾患に関連づけられる遺伝子座あたりいくつかのさらなるプローブを標的の多型遺伝子座の集合に含める；疾患に関連づけられる遺伝子座あたりのさらなるプローブの数は、4個から10個の間、11個から20個の間、21個から40個の間、41個から60個の間、61個から80個の間、またはそれらの組み合わせであってよい。

【0464】

試料中のDNA分子の数の決定

第1ラウンドのDNA増幅の間に試料中の元のDNA分子のそれぞれについて独自に同定された分子を生成することによって、試料中のDNA分子の数を決定するための方法が本明細書に記載されている。上記の目的を実現し、その後単一分子配列決定法またはクローン配列決定法が続く手順が本明細書に記載されている。

【0465】

該手法は、1つまたは複数の特定の遺伝子座を標的とするステップ、および、各標的の遺伝子座が独特のタグを有し、このバーコードをクローン配列決定または単一分子配列決定を用いて配列決定した際に互いに区別することができるように元の分子のタグを付けたコピーを生成するステップを包含する。独特の配列決定されたバーコードのそれぞれは元の試料における独特の分子を示す。同時に、配列決定データを使用して、分子が由来する遺伝子座を確認する。この情報を使用して、各遺伝子座について、元の試料中の独特の分子の数を決定することができる。

【0466】

この方法は、元の試料中の分子の数の定量的評価が必要な任意の適用のために用いることができる。さらに、1つまたは複数の標的の独特の分子の数を、1つまたは複数の他の標的に対する独特の分子の数に関連づけて、相対的なコピー数、対立遺伝子分布または対立遺伝子の比を決定することができる。あるいは、元の標的のコピーの最も可能性の高い数を同定するために、種々の標的から検出されたコピーの数を分布によってモデリングすることができる。適用としては、これらに限らないが、挿入および欠失、例えば、デュシェンヌ型筋ジストロフィーの保有者に見いだされるものの検出；染色体のセグメントの欠失または重複、例えば、コピー数変異体において観察されたものの定量；生まれた個体由来の試料の染色体コピー数；生まれていない個体、例えば、胚または胎児由来の試料の染色体コピー数が挙げられる。

【0467】

該方法は、配列による標的化に含有される同時に起こる変動の評価と組み合わせることができる。これを用いて、元の試料における各対立遺伝子を示す分子の数を決定することができる。このコピー数法は、SNPまたは他の配列の変動の評価と組み合わせて、生まれた個体および生まれていない個体の染色体コピー数を決定することができる；短い配列の変動を有するが、その中でPCRにより複数の標的領域から増幅することができる遺伝子座由来のコピーの識別および定量化、例えば、棘筋萎縮の保有者検出の目的で；異なる個体の混合物からなる試料由来の分子の種々の供給源のコピー数の決定、例えば、母系の血漿から得られた浮動性DNAからの胎児の異数性の検出の目的で。

【0468】

ある実施形態では、単一の標的遺伝子座に関係する方法は、以下のステップの1つまたは複数を含んでよい：（1）特定の遺伝子座をPCR増幅するための、オリゴマーの標準の対を設計するステップ。（2）合成の間に、標的特異的オリゴマーのうちの1つの5'末端側に、標的遺伝子座またはゲノムに対する相補性を有さない、または最小の相補性を有する特定の塩基の配列を付加するステップ。尾部と称されるこの配列は既知の配列であって、その後の増幅のために用いられるものであり、ランダムなヌクレオチドの配列が後に続く。これらのランダムなヌクレオチドはランダムな領域を含む。ランダムな領域は、各プローブ分子間で確率的に異なる、ランダムに生成した核酸の配列を含む。したがって、合成した後、尾部を付けたオリゴマープールは、既知の配列から始まり、その後に分子間で異なる未知の配列が続き、その後に標的特異的配列が続くオリゴマーの集団からなる。（3）尾部を付けたオリゴマーのみを使用して1ラウンドの増幅（変性、アニーリング、伸長）を実施するステップ。（4）エキソヌクレアーゼを反応物に加え、PCR反応を有効に停止させ、反応物を適切な温度でインキュベートして、鋳型とアニーリングしなかったフォワード一本鎖オリゴを除去し、伸長させて二本鎖産物を形成するステップ。（5）反応物を高い温度でインキュベートして、エキソヌクレアーゼを変性させ、その活性を排除するステップ（6）第1の反応において使用したオリゴマーの尾部と相補的な新しいオリゴヌクレオチドを他の標的特異的オリゴマーと一緒に反応物に加えて、PCRの第1ラウンドで生成した産物のPCR増幅を可能にするステップ。（7）増幅を継続して下流のクローン配列決定のために十分な産物を生成させるステップ。（8）十分な数の塩基を配列に結んだ、増幅されたPCR産物を、多数の方法、例えば、クローン配列決定によって測定するステップ。

【0469】

ある実施形態では、本開示の方法は、多数の遺伝子座を並行してまたは別の方法で標的とするステップを包含する。異なる標的遺伝子座に対するプライマーを独立に生成し、混合して多重PCRプールを作製することができる。ある実施形態では、元の試料を、サブプールに分けることができ、各サブプールにおいて異なる遺伝子座を標的とした後に組み換え、配列決定することができる。ある実施形態では、プールを細分する前にタグを付けるステップおよびいくつかの増幅サイクルを実施して全ての標的の効率的な標的化を確実にした後に、分割し、改善し、その後、細分されたプールにおけるより小さなプライマーの集合を使用して増幅を継続することによって増幅することができる。

【0470】

この技術が特に有用になる適用の1つの例は、非侵襲的な出生前異数性診断であり、所与の遺伝子座における対立遺伝子の比またはいくつかの遺伝子座における対立遺伝子の分布を用いて、胎児に存在する染色体のコピーの数の決定に役立てることができる。この状況では、最初の試料中に存在するDNAを、種々の対立遺伝子の相対的な量を維持しながら増幅することが望ましい。一部の場合、特に、存在するDNAが非常に少量である、例えば、5,000未満のゲノムのコピー、1,000未満のゲノムのコピー、500未満のゲノムのコピー、および100未満のゲノムのコピーである場合には、ボトルネックと称される現象が起こり得る。これは、最初の試料中に任意の所与の対立遺伝子の少数のコピーが存在し、増幅の偏りの結果、増幅されたDNAのプールの有するそれらの対立遺伝子の比が最初のDNAの混合物におけるそれらの対立遺伝子の比とは有意に異なるということである。標準のPCR増幅の前に、各DNAの鎖にバーコードの独特のまたはほぼ独特の集合を適用することにより、同じ元の分子を起源とする配列決定されたDNAのn個の同一の分子の集合からDNAのn-1コピーを排除することが可能である。

【0471】

例えば、個体のゲノム内のヘテロ接合性であるSNP、および元のDNAの試料中に各対立遺伝子が10分子存在する個体由来のDNAの混合物を考える。増幅した後、その遺伝子座に対応する100,000分子のDNAが存在し得る。確率論的なプロセスに起因して、DNAの比は1:2~2:1のいずれであってもよいが、元の分子のそれぞれに独特のタグでタグを付けたので、増幅されたプール内のDNAが正確に各対立遺伝子由来の10分子のDNAを起源とすることを決定することが可能である。したがって、この方法により、この手法を用いない方法よりも正確な測度の相対的な各対立遺伝子の量がもたらされる。対立遺伝子の偏りの相対量を最小限にすることが望ましい方法に対して、この方法により正確なデータがもたらされる。

【0472】

配列決定された断片と標的遺伝子座の関連づけは、いくつかのやり方で実現することができる。ある実施形態では、標的配列に対応する分子バーコード、同様に、十分な数の独特の塩基を標的断片に結んで十分な長さの配列を得て、標的遺伝子座を明白に同定することを可能にする。別の実施形態では、ランダムに生成した分子バーコードを含有する分子バーコーディングプライマーは、それが関連づけられる標的を同定する遺伝子座に特異的なバーコード（遺伝子座バーコード）も含有することができる。この遺伝子座バーコードは、個々の標的の各々に対する全ての分子バーコーディングプライマー間で、したがって、生じたアンプリコンの全ての間で同一であるが、他の全ての標的とは異なる。ある実施形態では、本明細書に記載のタグ付け方法を、片側のネスティングプロトコールと組み合わせることができる。

【0473】

ある実施形態では、分子バーコーディングプライマーの設計および生成は、以下の通り実施化することができる：分子バーコーディングプライマーは、標的配列と相補的でない配列、それに続くランダムな分子バーコード領域、それに続く標的特異的配列からなっており、分子バーコードの5'の配列は部分配列PCR増幅のために用いることができ、配列決定するためにアンプリコンをライブラリーに変換することにおいて有用な配列を含み得る。ランダムな分子バーコード配列は多数のやり方で生成することができる。好ましい方

法では、分子タグを付けたプライマーを、バーコード領域を合成する間の反応のために4種の塩基全てを含むように合成する。塩基の全てまたは塩基の種々の組み合わせは、IUPAC DNA多義コードを使用して明記することができる。このように、合成された分子の集団は、分子バーコード領域内の配列のランダムな混合物を含有する。バーコード領域の長さにより、どのくらい多くのプライマーが独特のバーコードを含有するかが決定される。独特の配列の数は、 N^L としてバーコード領域の長さに関連づけられ、ここで、 N は塩基の数であり、一般には4であり、 L はバーコードの長さである。5塩基のバーコードにより、最大1024個の独特の配列をもたらしことができ、8塩基のバーコードにより、65536個の独特のバーコードをもたらしことができる。ある実施形態では、DNAを、配列決定方法によって測定することができ、配列データは単一分子の配列を示す。これは、単一分子について直接配列決定する方法、または単一分子を増幅して、配列決定計器によって検出可能なクローンを形成するが、なお単一分子を示す、本明細書ではクローン配列決定と称される方法を含むことができる。

10

【0474】

いくつかの実施形態

いくつかの実施形態では、妊娠中の胎児における染色体についての決定された倍数性の状態が開示されている報告を作製するための方法が本明細書に開示されており、該方法は、胎児の母親由来のDNAおよび胎児由来のDNAを含有する第1の試料を得るステップと、胎児の一方の親または両親から遺伝子型データを得るステップと、調製された試料が得られるようにDNAを単離することによって第1の試料を調製するステップと、複数の多型遺伝子座において調製された試料中のDNAを測定するステップと、調製された試料に対して得たDNA測定値から、対立遺伝子数または複数の多型遺伝子座における対立遺伝子数の確率をコンピュータで算出するステップと、染色体における可能性のある異なる倍数性の状態について、染色体上の複数の多型遺伝子座における予測される対立遺伝子数の確率に関する、倍数性についての複数の仮説をコンピュータで作製するステップと、倍数性についての仮説のそれぞれについて、胎児の一方の親または両親からの遺伝子型データを使用して、染色体上の各多型遺伝子座の対立遺伝子数確率についての同時分布モデルをコンピュータで構築するステップと、調製された試料についての同時分布モデルおよび算出された対立遺伝子数の確率を用いて、倍数性についての仮説のそれぞれの相対的確率をコンピュータで決定するステップと、最大的確率を有する仮説に対応する倍数性の状態を選択することによって胎児の倍数性の状態を呼び出すステップと、決定された倍数性の状態が開示されている報告を作製するステップとを含む。

20

30

【0475】

いくつかの実施形態では、該方法を用いて、複数のそれぞれの母親における複数の妊娠中の胎児の倍数性の状態を決定し、該方法は、調製された試料のそれぞれにおける胎児起源のDNAのパーセントを決定するステップをさらに含み、ここで、調製された試料中のDNAを測定するステップは、各調製された試料中のいくつものDNA分子について配列決定することによって行い、より大きな胎児DNAの画分を有する調製された試料よりも、より小さな胎児DNAの画分を有する調製された試料由来のDNA分子について多く配列決定する。

40

【0476】

いくつかの実施形態では、該方法を用いて、複数のそれぞれの母親における複数の妊娠中の胎児の倍数性の状態を決定し、ここで、調製された試料中のDNAを測定するステップは、各胎児に対して、DNAの調製された試料の第1の画分について配列決定して第1の測定値の集合を得ることによって行い、該方法は、第1のDNA測定値の集合を考慮して、各胎児の倍数性についての仮説のそれぞれに対して第1の相対的確率の決定を行うステップと、倍数性についての仮説のそれぞれに対する第1の相対的確率の決定が、異数体の胎児に対応する倍数性についての仮説が有意であるが決定的ではない確率を有することを示す、その胎児からの調製された試料の第2の画分について再び配列決定して、第2の測定値の集合を得るステップと、第2の測定値の集合および必要に応じて、第1の測定値

50

の集合も使用して、胎児の倍数性についての仮説に対して第2の相対的確率の決定を行うステップと、第2の相対的確率の決定によって決定された通り最大の確率を有する仮説に対応する倍数性の状態を選択することによって第2の試料を再び配列決定した、その胎児の倍数性の状態を呼び出すステップとをさらに含む。

【0477】

いくつかの実施形態では、優先的に富化されたDNAの試料を含む組成物であって、優先的に富化されたDNAの試料が、第1のDNAの試料からの複数の多型遺伝子座において優先的に富化されており、第1のDNAの試料が母系の血漿に由来する母系DNAと胎児DNAの混合物からなり、富化の程度が少なくとも2倍であり、第1の試料と優先的に富化された試料の間の対立遺伝子の偏りが、平均で、2%未満、1%未満、0.5%未満、0.2%未満、0.1%未満、0.05%未満、0.02%未満、および0.01%未満からなる群から選択される組成物が開示されている。いくつかの実施形態では、そのような優先的に富化されたDNAの試料を作製するための方法が開示されている。

【0478】

いくつかの実施形態では、胎児のゲノムDNAおよび母系のゲノムDNAを含む母系の組織試料において胎児の異数性の存在または不在を決定するための方法であって、(a) 前記母系の組織試料から、胎児のゲノムDNAと母系のゲノムDNAの混合物を得るステップと、(b) 胎児DNAと母系DNAの混合物を複数の多型対立遺伝子において選択的に富化するステップと、(c) ステップaにおける胎児のゲノムDNAと母系のゲノムDNAの混合物から選択的に富化された断片を分布させて、単一のゲノムDNA分子または単一のゲノムDNA分子の増幅産物を含む反応試料をもたらすステップと、(d) ステップc)における反応試料中の選択的に富化されたゲノムDNAの断片についての大規模並行DNA配列決定を行って、前記選択的に富化された断片の配列を決定するステップと、(e) ステップd)において得られた配列が属する染色体を同定するステップと、(f) ステップd)からのデータを分析して、i) 母親および胎児の両方において二倍体であると推測される、少なくとも1つの最初の標的の染色体に属する、ステップd)からのゲノムDNAの断片の数、およびii) 胎児において異数体であることが疑われる第2の標的染色体に属する、ステップd)からのゲノムDNAの断片の数を決定するステップと、(g) 第2の標的染色体が正倍数性である場合、第2の標的染色体について、ステップf)パートi)において決定された数を使用してステップd)からのゲノムDNAの断片の数の予測される分布を算出するステップと、(h) 第2の標的染色体が異数体である場合、第2の標的染色体について、ステップf)パートii)である第1の数およびステップb)の混合物において見いだされる胎児DNAの推定される割合を用いてステップd)からのゲノムDNAの断片の数の予測される分布を算出するステップと、(i) 最尤法または最大事後法を用いて、ステップf)パートii)において決定されたゲノムDNAの断片の数が、ステップg)で算出された分布またはステップh)で算出された分布のどちらの一部である可能性がより高いかを決定し、それにより、胎児の異数性の存在または不在を示すステップとを含む方法が開示されている。

【0479】

実験セクション

ここで開示されている実施形態は、以下の実施例に記載されており、これらは本開示の理解を補助するために記載され、その後続く特許請求の範囲において定義されている本開示の範囲をいかなる形でも限定するものと解釈されるべきではない。以下の実施例は、当業者に、本記載した実施形態をどのように用いるかについての完全な開示および記載を提供するために提示されており、本開示の範囲を限定するものではなく、以下の実験が、実施した全ての実験または唯一の実験であることを示すものでもない。使用される数字(例えば、量、温度など)に関して正確さを確実にするための試みが行われているが、いくつかの実験的な誤差および偏差が考慮されるべきである。別段の指定のない限り、部分は体積による部分であり、温度は摂氏度である。記載されている方法の変形は、実験が例示することを意味する基本的な態様を変化させることなく行うことができることが理解され

るべきである。

【0480】

実験1

目的は、親の遺伝子型を使用して胎児の割合を算出するベイズ最尤推定（MLE）アルゴリズムにより、公開された方法と比較して非侵襲的な出生前トリソミー診断の正確度が改善されることを示すことであった。

【0481】

母系のcfDNAについてシミュレートされた配列決定データを、21トリソミーおよびそれぞれの母親の細胞系において得られた読み取りをサンプリングすることによって作製した。正確なダイソミーおよびトリソミーの呼び出しの率を、公開された方法（Chiuら BMJ 2011年：342巻：c7401頁）および本発明者らのMLEに基づくアルゴリズムについての種々の胎児の割合における500のシミュレーションから決定した。IRBに承認されたプロトコールの下で収集した、4人の妊娠中の母親およびそれぞれの父親由来の500万のショットガン読み取りを得ることによってシミュレーションを検証した。親の遺伝子型を290K SNPアレイで得た（図14参照）。

【0482】

シミュレーションでは、MLEに基づく手法により、9%の低さの胎児の割合に対して99.0%の正確度が実現され、全体的な正確度によく対応した信頼度が報告された。本発明者らは、これらの結果を、4つの実際の試料を用いて検証し、全て正確な呼び出しが得られ、計算された信頼度は99%を超えた。対照的に、Chiuらの公開されたアルゴリズムを実行したところ、99.0%の正確度を実現するためには18%の胎児の割合が必要であり、9%の胎児DNAでは87.8%の正確度しか実現されなかった。

【0483】

MLEに基づく手法と併せて、親の遺伝子型から胎児の割合を決定することにより、妊娠初期および妊娠中期の早期に予測される胎児の割合において、公開されているアルゴリズムよりも高い正確度が実現される。さらに、本明細書に開示されている方法は、結果の信頼性の決定において、特に倍数性の検出がより難しい低胎児の割合において、極めて重要な信頼度メトリックを生じる。公開された方法では、偽陽性率を予め定義する手法である、ダイソミートレーニングデータの大きな集合に基づく、倍数性を呼び出すための正確度が低い閾値方法を用いる。さらに、信頼度メトリックなしでは、公開された方法で呼び出しを行うには胎児のcfDNAが不十分である場合に、偽陰性の結果を報告する危険性がある。いくつかの実施形態では、呼び出された倍数性の状態について信頼度推定値を算出する。

【0484】

実験2

目的は、ベイズ最尤推定（MLE）アルゴリズムにおいて、標的化配列決定手法を親の遺伝子型およびHapmapデータと組み合わせて用いることによって、特に低胎児の割合からなる試料における、胎児の18トリソミー、21トリソミー、およびXトリソミーの非侵襲的な検出を改善することであった。

【0485】

4つの正倍数性妊娠および2つのトリソミー陽性妊娠由来の母体試料およびそれぞれの父系の試料を、IRBに承認されたプロトコールの下で、胎児の核型が既知である患者から得た。母系のcfDNAを血漿から抽出し、標的の特定のSNPを優先的に富化した後、およそ1,000万の配列読み取りを得た。親試料について同様に配列決定して、遺伝子型を得た。

【0486】

記載されているアルゴリズムにより、正倍数性の試料および異数性の試料の正常な染色体の全てについて第18染色体ダイソミーおよび第21染色体ダイソミーが正確に呼び出された。18トリソミーおよび21トリソミーの呼び出しは正確であり、男の胎児および女の胎児におけるX染色体コピー数も正確であった。アルゴリズムによって生じる信頼度

は、全ての場合において98%を超えた。

【0487】

記載されている方法により、妊娠初期の試料および妊娠中期の早期の試料のおよそ30%を占める、12%未満の胎児DNAで構成される試料を含めた、6つの試料由来の試験した染色体の全てについて倍数性が正確に報告された。当該MLEアルゴリズムと公開された方法の間の極めて重要な差異は、MLEアルゴリズムでは親の遺伝子型およびHapmapデータを活用して、正確度を改善し、信頼度メトリックを生成することである。低胎児の割合では、全ての方法の正確度が低くなる；十分な胎児のcfDNAがない試料を正確に同定して、信頼できる呼び出しを行うことが重要である。他のものは男の胎児の胎児の割合を推定するためにY染色体特異的プローブを使用した。同時に親の遺伝子型決定を行うことにより、両方の性について胎児の割合を推定することができる。非標的化ショットガン配列決定を用いた公開された方法の別の固有の限界は、GCリッチなどの因子が異なることにより、倍数性呼び出しの正確度が染色体間で変動することである。当該標的化配列決定手法は、そのような染色体規模の変動とはほとんど無関係であり、染色体間でより一貫した性能をもたらす。

10

【0488】

実験3

目的は、母系の血漿中の浮動性胎児DNAのSNP遺伝子座を解析するための新規のインフォマティクスを用いて、三倍体の胎児においてトリソミーが高い信頼度で検出可能であるかどうかを決定することであった。

20

【0489】

超音波異常の後に、血液20mLを妊娠中の患者から抜き取った。遠心分離した後、母系DNAをパフィーコートから抽出し(DNEASY、QIAGEN)、無細胞DNAを血漿から抽出した(QIAAMP QIAGEN)。両方のDNA試料において、第2染色体、第21染色体、およびX染色体上のSNP遺伝子座に標的化配列決定を適用した。最尤ベイズ推定により、全ての可能性のある倍数性の状態の集合から、最も可能性が高い仮説を選択した。該方法により、胎児DNA割合、倍数性の状態および倍数性の決定における明確な信頼度を決定する。参照染色体の倍数性に関する仮定は行わない。診断では、現在の技術水準である、配列読み取りの計数と無関係の検定統計量を使用する。

30

【0490】

当該方法により、第2染色体および第21染色体のトリソミーが正確に診断された。子の割合は11.9%[CI 11.7~12.1]と推定された。胎児は、第2染色体および第21染色体の1つの母系のコピーおよび2つの父系のコピーを有することが見いだされ、信頼度は有効に1(エラー確率 $<10^{-30}$)であった。これは、第2染色体および第21染色体のそれぞれ92,600読み取りおよび258,100読み取りを用いて実現された。

【0491】

これは、中期の核型によって確認されるように、母系の血液に由来するトリソミー染色体であって、その胎児が三倍体であるという非侵襲的な出生前診断の最初の実証である。非侵襲的な診断の現存の方法では、この試料において異数性は検出されない。現行の方法は、ダイソミー参照染色体と比較した、トリソミー染色体における余剰な配列読み取りに依拠するが、三倍体の胎児はダイソミー参照を有さない。さらに、現存の方法では、この胎児DNAの割合および配列読み取りの数を用いては同様に信頼度が高い倍数性の決定は実現されない。該手法を24の染色体全てに拡張することは簡単である。

40

【0492】

実験4

正倍数性の妊娠由来の母系の血漿から単離されたDNA、および同様に21三倍体性細胞系由来のゲノムDNAの、標準のPCR(ネスティングを使用しなかったことを意味する)を使用した800プレックス増幅のために以下のプロトコルを使用した。ライブラリーの調製および増幅は、単一チューブ平滑末端化、その後のA-テーリングを伴った。

50

AGILENT SURESELECTキットに見いだされるライゲーションキットを使用してアダプタライゲーションを実行し、PCRを7サイクル実行した。次いで、第2染色体、第21染色体およびX染色体上のSNPを標的とする800の異なるプライマー対を使用して、STAを15サイクル行った（95℃で30秒間；72℃で1分間；60℃で4分間；65℃で1分間；72℃で30秒間）。12.5nMのプライマー濃度で反応を実行した。次いで、ILLUMINA IIGAXシーケンサーを用いてDNAについて配列決定した。シーケンサーにより190万の読み取りが出力され、その92%がゲノムにマッピングされ、ゲノムにマッピングされた読み取りのうち99%超が、標的のプライマーにより標的とされた領域のうちの1つにマッピングされた。数は血漿DNAとゲノムDNAの両方で基本的に同じであった。図15は、第21染色体において既知のトリソミーを有する細胞系から取得したゲノムDNAにおける、シーケンサーによって検出された約780SNPについての2つの対立遺伝子の比を示す。対立遺伝子分布は視覚的に読み取ることが簡単ではないので、ここでは可視化を容易にするために対立遺伝子の比がプロットされていることに留意されたい。丸印はダイソミー染色体上のSNPを示し、星印はトリソミー染色体上のSNPを示す。図16は、図Xの場合と同様に、同じデータの別の表示であり、Y軸は各SNPについて測定されたAとBの相対的な数であり、X軸は染色体によってSNPを分けたSNP数である。図16では、SNP1～312は、第2染色体上に見いだされ、SNP313～605は、トリソミーである第21染色体上に見いだされ、SNP606～800はX染色体上に見出される。第2染色体およびX染色体からのデータは、相対的な配列計数が3つのクラスター内にあるとおり、ダイソミー染色体を示す：グラフの一番上がAAであり、グラフの一番下がBBであり、グラフの中央がABである。トリソミーである第21染色体からのデータは4つのクラスターを示す：グラフの一番上がAAAであり、0.65の線（2/3）の周辺がAABであり、.35の線（1/3）の周辺がABBであり、グラフの一番下がBBBである。

【0493】

図17は、同じ800プレックスプロトコールについてのデータであって、妊娠中の女性由来の4つの血漿試料から増幅したDNAに対して測定されたデータを示す。これらの4つの試料について、点について7つのクラスターが認められることが予想される：（1）グラフの一番上に沿っているのは、母親および胎児がどちらもAAである遺伝子座であり、（2）グラフの一番上のわずかに下は、母親がAAであり、胎児がABである遺伝子座であり、（3）0.5の線のわずかに上は、母親がABであり、胎児がAAである遺伝子座であり、（4）0.5の線に沿っているのは、母親および胎児がどちらもABである遺伝子座であり、（5）0.5の線のわずかに下は、母親がABであり、胎児がBBである遺伝子座であり、（6）グラフの一番下のわずかに上は、母親がBBであり、胎児がABである遺伝子座であり、（7）グラフの一番下に沿っているのは、母親および胎児がどちらもBBである遺伝子座である。胎児の割合が小さいほど、クラスター（1）と（2）の間、クラスター（3）、（4）および（5）の間、ならびにクラスター（6）と（7）の間の分離が小さくなる。分離は、胎児起源のDNAの画分の半分であることが予想される。例えば、DNAの20%が胎児性であり、80%が母系である場合、（1）～（7）は、それぞれ1.0、0.9、0.6、0.5、0.4、0.1および0.0に集中することが予想される；例えば、図17、POOL1__BC5__ref__rateを参照されたい。その代わりに、DNAの8%が胎児性であり、92%が母系である場合、（1）～（7）は、それぞれ1.00、0.96、0.54、0.50、0.46、0.04および0.00に集中することが予想される；例えば、図17、POOL1__BC2__ref__rateを参照されたい。胎児DNAが検出されない場合は、（2）、（3）、（5）または（6）が認められることは予想されない；あるいは、分離は0であると言える、したがって（1）および（2）は互いに一番上にあり、（3）、（4）および（5）、ならびに、同様に（6）および（7）も同様である；例えば、図17、POOL1__BC7__ref__rateを参照されたい。図17、POOL1__BC1__ref__rateについて胎児の割合は約25%であることに留意されたい。

10

20

30

40

50

【0494】

実験5

DNAの増幅および測定の方法の大多数により、一般に遺伝子座において見いだされる2つの対立遺伝子が、DNAの試料中の対立遺伝子の実際の量を表さない強度または計数で検出される、いくつかの対立遺伝子の偏りが生じる。例えば、単一の個体について、ヘテロ接合性遺伝子座において、ヘテロ接合性遺伝子座について予測される理論的な比である、2つの対立遺伝子の1:1の比が認められることが予想されるが、対立遺伝子の偏りに起因して、55:45または、さらには60:40が認められ得る。配列決定との関連において、読み取りの深さが低い場合には、単純な確率論的ノイズにより、有意な対立遺伝子の偏りがもたらされる可能性があることにも留意されたい。ある実施形態では、各SNPの挙動をモデリングすることが可能であり、したがって、特定の対立遺伝子について一貫した偏りが観察される場合、この偏りを補正することができる。図18は、偏り補正の前後の、二項分散によって説明することができるデータの割合を示す。図18では、星印は、800プレックス実験について、生の配列データにおいて観察された対立遺伝子の偏りを示し、丸印は、補正後の対立遺伝子の偏りを示す。対立遺伝子の偏りが全くない場合には、データが $x=y$ の線に沿うことが予想されることに留意されたい。150プレックス標的化増幅を用いてDNAを増幅することによって生じる同様のデータの集合により、偏り補正後に1:1の線のごく近傍に包含されるデータが生じた。

10

【0495】

実験6

プライマーのアニーリングおよび伸長の時間が数分に限られている、アダプタタグに特異的なプライマーとライゲーションしたアダプタを使用したDNAのユニバーサル増幅は、より短いDNA鎖の割合を富化する効果を有する。配列決定に適したDNAライブラリーを作製するために設計された大多数のライブラリープロトコールはそのようなステップを含み、プロトコールの例は公開されており、当業者に周知である。本発明のいくつかの実施形態では、ユニバーサルタグを有するアダプタを血漿DNAにライゲーションし、アダプタタグに特異的なプライマーを使用して増幅する。いくつかの実施形態では、ユニバーサルタグは、配列決定のために用いるものと同じタグであってよく、それはPCR増幅のためだけのユニバーサルタグであってよい、またはそれはタグの集合であってよい。胎児DNAは一般には、天然では短く、一方、母系DNAは天然では短いものと長いものの両方であり得るので、この方法は、混合物中の胎児DNAの割合を富化する効果を有する。アポトーシス性の細胞由来のDNAであると考えられ、胎児DNAと母系DNAの両方を含有する浮動性DNAは短く、大部分は200bp未満である。静脈切開後の一般的な現象である細胞溶解によって放出される細胞性DNAは、一般には、ほぼ排他的に母系であり、同様にかかなり長く、大部分が500bpを超える。したがって、数分超放置した血液試料は、短い（胎児性+母系）DNAおよびより長い（母系）DNAの混合物を含有する。母系の血漿に対してユニバーサル増幅を比較的短い伸長時間で実施し、その後、標的化増幅することにより、胎児DNAの相対的な割合が、標的化増幅を単独で用いて増幅した血漿と比較して増大する傾向がある。これは、入力が血漿DNA（垂直方向の軸）である場合に測定された胎児のパーセント対入力DNAがILLUMINA GAIIxライブラリー調製プロトコールを用いて調製したライブラリーを有する血漿DNAである場合に測定された胎児のパーセントを示す図19において認めることができる。線の下に入る点は全て、ライブラリーの調製ステップにより胎児起源のDNAの割合（fraction）が富化されることを示す。赤色であった2つの血漿の試料は溶血を示し、したがって、細胞溶解によって存在する長い母系DNAの量が増大したことを示し、これは、標的化増幅の前にライブラリーの調製を実施した場合に、胎児の割合（fetal fraction）が特に有意に富化されることを示す。本明細書に開示されている方法は、溶血があるまたは比較的長い鎖の混入DNAを含む細胞が溶解し、短いDNAと長いDNAが混合した試料に混入するいくつかの他の状況が生じている場合に特に有用である。一般には、比較的短いアニーリング時間および伸長時間は30秒間から2分間の間であるが、5秒

20

30

40

50

または10秒以下の長さであってよく、または5分間または10分間の長さであってよい。

【0496】

実験7

正倍数性の妊娠由来の母系の血漿から単離されたDNA、および同様に21三倍体性細胞系由来のゲノムDNAの、直接PCRプロトコール、および同様にセミネステッド手法を用いた1,200プレックス増幅のために以下のプロトコールを使用した。ライブラリーの調製および増幅は、単一チューブ平滑末端化、その後のA-ターミングを伴った。AGILENT SURESELECTキットに見いだされるライゲーションキットの改変を用いてアダプタライゲーションを実行し、PCRを7サイクル実行した。標的のプライマープールでは、第21染色体由来のSNPについての550のアッセイ、ならびに第1染色体およびX染色体のそれぞれ由来のSNPについての325のアッセイを行った。どちらのプロトコールも、16nMのプライマー濃度を用いてSTAの15サイクルを伴った(95℃で30秒間;72℃で1分間;60℃で4分間;65℃で30秒間;72℃で30秒間)。セミネステッドPCRプロトコールは、29nMの内側のフォワードタグ濃度、および1μMまたは0.1μMのリバースタグ濃度を用いたSTA15サイクルの第2の増幅を伴った(95℃で30秒間;72℃で1分間;60℃で4分間;65℃で30秒間;72℃で30秒間)。次いで、ILLUMINA IIGAXシーケンサーを用いてDNAについて配列決定した。直接PCRプロトコールについては、読み取りの73%がゲノムにマッピングされ、セミネステッドプロトコールについては、配列読み取りの97.2%がゲノムにマッピングされる。したがって、セミネステッドプロトコールにより、およそ30%多くの情報がもたらされ、これは、主に、プライマー二量体を引き起こす可能性が最も高いプライマーが排除されたことに起因すると推測される。

【0497】

読み取りの深さの変動性は、セミネステッドプロトコールを使用した場合、直接PCRプロトコールを使用した場合よりも高い傾向があり(図20参照)、ひし形はセミネステッドプロトコールを用いて実行した遺伝子座についての読み取りの深さを指し、四角はネスティングなしで実行した遺伝子座についての読み取りの深さを指す。SNPは、ひし形について読み取りの深さによって配置されており、したがって、ひし形は全て曲線上に置かれ、一方四角は、ゆるく相関するようである;SNPの配置は任意であり、読み取りの深さを指すのは、ドットの左から右への場所ではなく、ドットの高さである。

【0498】

いくつかの実施形態では、本明細書に記載の方法により、優れた読み取りの深さ(DOR)の分散を実現することができる。例えば、1,200アッセイの、ゲノムDNAの1,200プレックス直接PCR増幅を用いたこの実験の1つのバージョン(図21)において、1186アッセイでは、DORが10超であり、平均の読み取りの深さが400であり、1063アッセイ(88.6%)では、読み取りの深さが200から800の間であり、各対立遺伝子についての読み取りの数は、意味のあるデータを得るために十分に高いが、各対立遺伝子についての読み取りの数は、これらの読み取りの限界使用が特に小さい場合、それほど高くない、理想的なウィンドウを有した。12の対立遺伝子のみが、高い読み取りの深さを有し、1035の読み取りにおいて一番高かった。DORの標準偏差は290であり、平均DORは453であり、DORの変動係数は64%であり、950,000の総読み取りが存在し、およそ読み取りの63.1%がゲノムにマッピングされた。1,200プレックスセミネステッドプロトコールを使用した別の実験(図22)では、DORはより高かった。DORの標準偏差は583であり、平均DORは630であり、DORの変動係数は93%であり、870,000の総読み取りが存在し、読み取りの96.3%がゲノムにマッピングされた。これらの場合のどちらにおいても、SNPは母親についての読み取りの深さによって配置され、したがって、曲線は母系の読み取りの深さを示すことに留意されたい。子と父親の間の鑑別は重要でなく、それは単にこの説明のために重要であるトレンドである。

【0499】

実験8

実験において、セミネステッド1, 200プレックスPCRプロトコルを用いて、1つの細胞由来のDNAおよび3つの細胞由来のDNAを増幅した。この実験は、母系の血液から単離された胎児の細胞を使用した出生前異数性試験と関連する、または生検割球または栄養外胚葉試料を使用した着床前遺伝子診断のためのものである。条件ごとに2個体(46XYおよび47XX+21)由来の1つの細胞および3つの細胞の3つの複製物が存在した。アッセイは、第1染色体、第21染色体およびX染色体を標的とした。3つの異なる溶解方法を使用した: ARCTURUS、MPER v2およびアルカリ溶解。1つの配列決定レーンにおいて多重化48試料に配列決定を実行した。アルゴリズムにより、3つの染色体のそれぞれについて、および複製物のそれぞれについての正確な倍数性呼び出しが生じた。

10

【0500】

実験9

1つの実験では、4つの母系の血漿試料を調製し、ヘミネステッド9, 600プレックスプロトコルを使用して増幅した。試料を以下のように調製した: 母系の血液最大40mLを遠心分離して、buffyコートおよび血漿を単離した。母系のゲノムDNAをbuffyコートから調製し、父系のDNAを血液試料または唾液試料から調製した。母系の血漿中の無細胞DNAを、QIAGEN CIRCULATING NUCLEIC ACIDキットを使用して単離し、TE緩衝液45μLで製造者の指示に従って溶出した。ユニバーサルライゲーションアダプタを、精製された血漿DNA35μLの各分子の末端に付加し、ライブラリーを、アダプタ特異的プライマーを使用して7サイクルにわたって増幅した。ライブラリーを、AGENCOURT AMPUREビーズを使用して精製し、水50μLで溶出した。

20

【0501】

9, 600個の標的特異的タグを付けたリバースプライマーのプライマー濃度14.5nMおよび1つのライブラリーアダプタ特異的フォワードプライマーの500nMを使用し、15サイクルのSTAを用いて、3μLのDNAを増幅した(最初のポリメラーゼ活性化のために95℃で10分間、次いで、95℃で30秒間; 72℃で10秒間; 65℃で1分間; 60℃で8分間; 65℃で3分間および72℃で30秒間; を15サイクル、および72℃で2分間の最終の伸長)。

30

【0502】

ヘミネステッドPCRプロトコルは、第1のSTA産物の希釈物の、1000nMのリバースタグ濃度、および9, 600個の標的特異的フォワードプライマーのそれぞれについて16.6μnMの濃度を用いたSTAを15サイクルにわたる第2の増幅を伴った(最初のポリメラーゼ活性化のために95℃で10分間、次いで95℃で30秒間; 65℃で1分間; 60℃で5分間; 65℃で5分間および72℃で30秒間; の15サイクル、および72℃で2分間の最終の伸長)。

【0503】

次いで、STA産物の一定分量を、標準のPCRによって、1μMのタグ特異的なフォワードプライマーおよびバーコードを付けたリバースプライマーを用いて10サイクルにわたって増幅し、バーコードを付けた配列決定ライブラリーを生成した。各ライブラリーの一定分量を異なるバーコードのライブラリーと混合し、スピンカラムを使用して精製した。

40

【0504】

このように、単一ウェル反応において9, 600個のプライマーを使用した; プライマーは、第1染色体、第2染色体、第13染色体、第18染色体、第21染色体、X染色体およびY染色体上に見いだされるSNPを標的とするように設計した。次いで、アンプリコンについて、ILLUMINA GAII Xシーケンサーを用いて配列決定した。試料当たり、およそ390万の読み取りがシーケンサーによって生成され、370万の読み取

50

りがゲノムにマッピングされ（94％）、それらのうち、290万の読み取り（74％）が標的のSNPにマッピングされ、平均の読み取りの深さは344であり、読み取りの深さの中央値は255であった。4つの試料についての胎児の割合は、9.9％、18.9％、16.3％、および21.2％であることが見いだされた。

【0505】

関連性のある母系のゲノムDNA試料および父系のゲノムDNA試料を、セミネステッド9600プレックスプロトコルを使用して増幅し、配列決定した。セミネステッドプロトコルは、第1のSTAにおいて9,600個の外側のフォワードプライマーおよびタグを付けたリバースプライマーを7.3nMで適用するという点で異なる。サーモサイクリング条件および第2のSTAの組成、およびバーコーディングPCRはヘミネステッドプロトコルについてのもと同じであった。

10

【0506】

配列決定データを、本明細書に開示されているインフォマティクス方法を用いて解析し、DNAが4つの母系の血漿試料中に存在した胎児について、6つの染色体において倍数性の状態を呼び出した。集団内の28個の染色体全てについての倍数性呼び出しは正確に呼び出され、信頼度は、正確に呼び出されたが、信頼度が83％であった1つの染色体以外は99.2％を超えた。

【0507】

図23は、9,600プレックスヘミネスティング手法の読み取りの深さを、実験7に記載の1,200プレックスセミネステッド手法の読み取りの深さと一緒に示すが、読み取りの深さが100超、200超および400超であるSNPの数は1,200プレックスプロトコルにおけるそれよりも有意に多かった。第90パーセンタイルにおける読み取りの数を第10パーセンタイルにおける読み取りの数で割って、無次元のメトリックを得ることができ、それにより、読み取りの深さの均一性が示され；その数が小さいほど、読み取りの深さがより均一である（狭い）。平均の第90パーセンタイル／第10パーセンタイル比は、実験9において行った方法では11.5であるが、実験7において行った方法では5.6である。特定の読み取りの百分率が読み取り数の閾値を超えることを確実にするために、より少ない配列読み取りが必要であるので、配列決定効率のためには、所与のプロトコルプレキシティ（plexity）に対してより狭い読み取りの深さがより良い。

20

30

【0508】

実験10

1つの実験では、4つの母系の血漿試料を調製し、セミネステッド9,600プレックスプロトコルを使用して増幅した。実験10の詳細は実験9と非常に類似しており、例外はネスティングプロトコルであること、および4つの試料の同一性を含めたことであった。集団内の28個の染色体全てについての倍数性呼び出しは正確に呼び出され、信頼度は99.7％を超えた。760万（97％）の読み取りがゲノムにマッピングされ、読み取りの630万（80％）が標的のSNPにマッピングされた。平均の読み取りの深さは751であり、読み取りの深さの中央値は396であった。

【0509】

実験11

1つの実験では、3つの母系の血漿試料を5つの均等な部分に分割し、各部分を、2,400個の多重化プライマー（4つの部分）または1,200個の多重化プライマー（1つの部分）のいずれかを使用して増幅し、合計10,800個のプライマーについて、セミネステッドプロトコルを使用して増幅した。増幅した後、配列決定するために該部分を一緒にプールした。実験11の詳細は実験9と非常に類似しており、例外は、ネスティングプロトコル、およびスプリットアンドプール手法であった。集団内の21個の染色体の全てについて、倍数性呼び出しは正確に呼び出され、信頼度は、信頼度83％で呼び出しが上手く行かなかった1つ以外は99.7％を超えた。340万の読み取りが標的のSNPにマッピングされ、平均の読み取りの深さは404であり、読み取りの深さの中央

40

50

値は258であった。

【0510】

実験12

1つの実験では、4つの母系の血漿試料を、4つの均等な部分に分割し、各部分を、2, 400個の多重化プライマーを使用して増幅し、合計9, 600個のプライマーについて、セミネステッドプロトコールを使用して増幅した。増幅した後、配列決定するために該部分を一緒にプールした。実験12の詳細は実験9と非常に類似しており、例外は、ネスティングプロトコール、およびスプリットアンドプール手法であった。集団内の28個の染色体全てについての倍数性呼び出しは正確に呼び出され、信頼度は、信頼度78%で呼び出しが上手く行かなかった1つ以外は97%を超えた。450万の読み取りが標的のSNPにマッピングされ、平均の読み取りの深さは535であり、読み取りの深さの中央値は412であった。

10

【0511】

実験13

1つの実験では、4つの母系の血漿試料を調製し、合計9, 600個のプライマーについて、9, 600プレックス3重ヘミネステッドプロトコールを使用して増幅した。実験12の詳細は実験9と非常に類似しており、例外は、増幅の3つのラウンドを伴うネスティングプロトコールであり；該3つのラウンドは、それぞれ15サイクルのSTA、10サイクルのSTAおよび15サイクルのSTAを伴った。集団内の28個の染色体のうち27個についての倍数性呼び出しは正確に呼び出され、信頼度は、94.6%で正確に呼び出された1つ、および信頼度80.8%で呼び出しが上手く行かなかった1つ以外は99.9%を超えた。350万の読み取りが標的のSNPにマッピングされ、平均の読み取りの深さは414であり、読み取りの深さの中央値は249であった。

20

【0512】

実験14

1つの実験では、細胞の集合45個を、1, 200プレックスセミネステッドプロトコールを使用して増幅し、配列決定し、倍数性の決定を3つの染色体において行った。この実験は、3日目の胚由来の単一細胞生検材料または5日目の胚由来の栄養外胚葉生検材料において着床前遺伝子診断を実施する条件をシミュレートすることを意図していることに留意されたい。個々の単一細胞15個および3つの細胞の集合30個を、合計45の反応のために、45個の個々の反応チューブに入れ、各反応は、ただ1つの細胞系由来の細胞を含有したが、異なる反応は異なる細胞系由来の細胞を含有した。細胞を洗浄バッファー5μl中に調製し、ARCTURUS PICOPURE溶解緩衝液(APPLIED BIOSYSTEMS)5μlを加えることによって溶解させ、56℃で20分間、95℃で10分間インキュベートした。

30

【0513】

単一の細胞／3つの細胞のDNAを、1200個の標的特異的フォワードプライマーおよびタグを付けたリバースプライマーを50nMのプライマー濃度を使用して、25サイクルのSTAを用いて増幅した（最初のポリメラーゼ活性化のために95℃で10分間、次いで95℃で30秒間；72℃で10秒間；65℃で1分間；60℃で8分間；65℃で3分間および72℃で30秒間；を25サイクル、および72℃で2分間の最終の伸長）。

40

【0514】

セミネステッドPCRプロトコールは、1000nMの濃度のリバースタグ特異的プライマー、および、それぞれ60nMの濃度の400個の標的特異的ネステッドフォワードプライマーを使用したSTAの20サイクル（最初のポリメラーゼ活性化のために95℃で10分間、次いで95℃で30秒間；65℃で1分間；60℃で5分間；65℃で5分間および72℃で30秒間；を15サイクル、および72℃で2分間の最終の伸長）にわたる、第1のSTA産物の希釈物の3つの並行した第2の増幅を伴った。したがって、3つの並行400プレックス反応では、第1のSTAにおいて増幅された合計1200個の

50

標的が増幅された。

【0515】

次いで、S T A産物の一定分量を、標準のP C Rによって、1 μ Mのタグ特異的なフォワードプライマーおよびバーコードを付けたリバースプライマーを用いて15サイクルにわたって増幅し、バーコードを付けた配列決定ライブラリーを生成した。各ライブラリーの一定分量を異なるバーコードのライブラリーと混合し、スピンカラムを使用して精製した。

【0516】

このように、単一細胞反応において1, 200個のプライマーを使用した；プライマーは、第1染色体、第21染色体およびX染色体上に見いだされるS N Pを標的とするように設計した。次いで、アンプリコンについて、I L L U M I N A G A I I Xシーケンサーを使用して配列決定した。試料当たりおよそ390万の読み取りがシーケンサーによって生成され、5000億~8000億の読み取りがゲノムにマッピングされた（試料当たりの全ての読み取りの74%~94%）。

【0517】

細胞系由来の関連性のある母系のゲノムD N A試料および父系のゲノムD N A試料を、同じセミネステッド1200プレックスアッセイプールを使用して、同様のプロトコルを用い、より少ないサイクルおよび1200プレックスの第2のS T Aを用いて解析し、配列決定した。

【0518】

配列決定データを、本明細書に開示されているインフォマティクス方法を用いて解析し、試料について3つの染色体において倍数性の状態を呼び出した。

【0519】

図24は、6つの試料について、3つの染色体（1=第1染色体；2=第21染色体；3=X染色体）における正規化された読み取りの深さの比（垂直方向の軸）を示す。比は、その染色体にマッピングされる読み取りの数と等しくなるように設定し、正規化し、それぞれが3つの46XY細胞を含む3つのウェルにわたって平均した、その染色体にマッピングされる読み取りの数で割った。46XY反応に対応する3つのデータ点の集合は、1:1の比を有することが予測される。47XX+21細胞に対応する3つのデータ点の集合は、第1染色体については1:1、第21染色体については1.5:1、およびX染色体については2:1の比を有することが予測される。

【0520】

図25は、3つの反応に関して3つの染色体（1、21、X）についてプロットした対立遺伝子の比を示す。左下の反応は、3つの46XY細胞における反応を示す。左側の領域は第1染色体についての対立遺伝子の比であり、中央の領域は第21染色体についての対立遺伝子の比であり、右側の領域はX染色体についての対立遺伝子の比である。46XY細胞に関して、第1染色体については、S N P遺伝子型A A、A BおよびB Bに対応する1、0.5および0の比が認められることが予想される。46XY細胞に関して、第21染色体については、S N P遺伝子型A A、A BおよびB Bに対応する1、0.5および0の比が認められることが予想される。46XY細胞に関して、X染色体については、S N P遺伝子型A、およびBに対応する1および0の比が認められることが予想される。右下の反応は、3つの47XX+21細胞における反応を示す。対立遺伝子の比は、左下のグラフの場合と同様に染色体によって分離される。47XX+21細胞に関して、第1染色体については、S N P遺伝子型A A、A BおよびB Bに対応する1、0.5および0の比が認められることが予想される。47XX+21細胞に関して、第21染色体については、S N P遺伝子型A A A、A A B、A B BおよびB B Bに対応する1、0.67、0.33および0の比が認められることが予想される。47XX+21細胞に関して、X染色体については、S N P遺伝子型A A、A B、およびB Bに対応する1、0.5および0の比が認められることが予想される。右上のプロットは、47XX+21細胞系由来のゲノムD N Aを1 n g含む反応に対して行われた。図26は、図25と同じグラフを示すが、

ただ1つの細胞に対して実施された反応についてのものである。左側のグラフは、4 7 X X + 2 1細胞を含有する反応についてのグラフであり、右側のグラフは4 6 X X細胞を含有する反応についてのグラフであった。

【0521】

図25および図26に示されているグラフから、1および0の比が認められることが予想される染色体については点のクラスターが2つあること、1、0.5、および0の比が認められることが予想される染色体については点のクラスターが3つあること、および1、0.67、0.33および0の比が認められることが予想される染色体については点のクラスターが4つあることが視覚的に明白である。parental support アルゴリズムにより、45反応全ての3つの染色体の全てについて正確な呼び出しを行うことができた。

10

【0522】

本明細書において引用されている全ての特許、特許出願および刊行参考文献は、その全体が参照により本明細書に組み込まれる。本開示の方法はその特定の実施形態とともに記載されているが、さらに改変することができることが理解されよう。さらに、本出願は、本開示の方法が関する当技術分野における公知または通例の実施の範囲内に入る、および添付の特許請求の範囲の範囲内に入る本開示からの逸脱を含めた、本開示の方法の任意の変動、使用または適応を包含するものとする。

【図1】

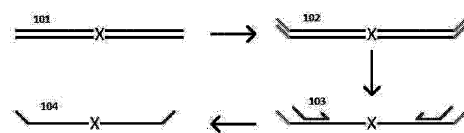


Figure 1

【図2】

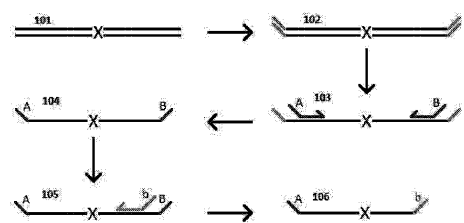


Figure 2

【図3】

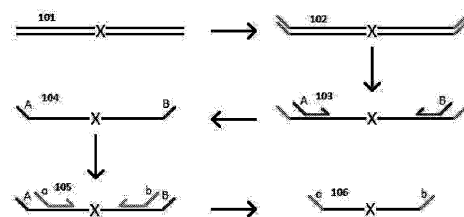


Figure 3

【図4】

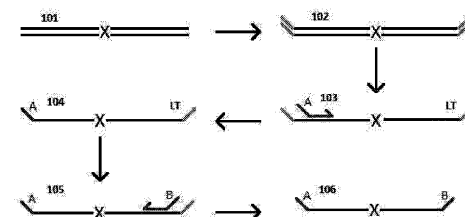


Figure 4

【図5】

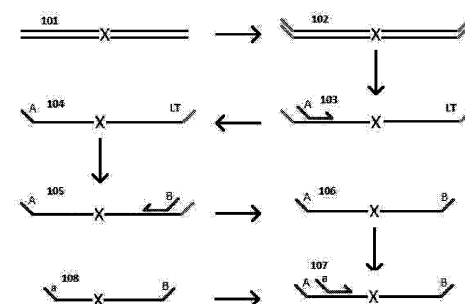


Figure 5

【图 6】

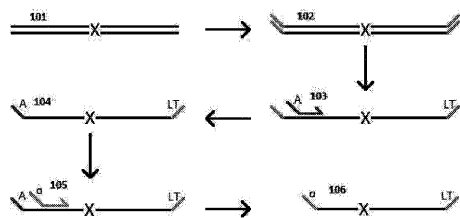


Figure 6

【図 7】

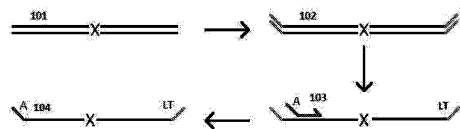


Figure 7

【图 8】

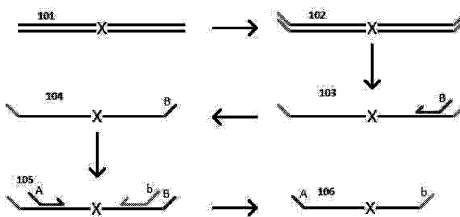


Figure 8

【図 1 1】

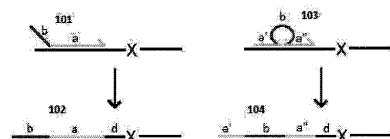


Figure 11

【图 1 2】

配列決定アダプタ配列をプライマー配列内に配置し、両側を標的の特定の配列に隣接させる。各プライマーの3' 末端の10塩基の相補的特徴である。プライマーはリアルタイムPCRで上流に試験された。これについて配列決定するために、配列決定する必要があるプライマー塩基の数減らす。

プライマーID	プライマー配列	ターゲット配列
n8130564	int-tag 1.10	AACCTCATACATGACGCTGACGCTCTTCGATCTTCGATGACGACATCCTCTCTGTGAGACGAGGCTCTTCGATCTCCGCTCTG
n2812909	int-tag 2.10	TCTCTCTGTGAGACGAGGCTCTTCGATCTCCGCTCTG
n12011281	int-tag 3.10	TCTCTCTGTGAGACGAGCTCTTCGATCTCCGCTCTG
n6191561	int-tag 4.10	TACATCTCTTGAGACGAGCTCTTCGATCTCCGATCTGTCAGT
n1061070	int-tag 5.10	TTGAGTGGAGTGGTGGACGAGGCTCTTCGATCTCCGATCTGAGAGTTG
n10604081	int-tag 6.10	GCTCTGATGTCGACGAGCTCTTCGATCTCCGATCTGAGAGCTG

配列決定アダプタ配列をプライマー配列内に配置し、両側を標的の特定の配列に隣接させる。内部のタグを、両末端に基づいて10塩基の相補的特徴としてペプチド構造内に形成する。これにより、プライマーの相補的特徴が非常に近い。内部のタグの非相補的特徴が知られる。各プライマーの3' 末端の10塩基の相補的特徴である。プライマーはリアルタイムPCRで上流に試験された。

プライマーID	プライマー配列	ターゲット配列
n8130564	loop-int-tag 1.10	AACCTCATACATGACGCTGACGCTCTTCGATCTTCGATGACGACATCCTCTCTGTGAGACGAGGCTCTTCGATCTCCGCTCTG
n2812909	loop-int-tag 2.10	TCTCTCTGTGAGACGAGGCTCTTCGATCTCCGCTCTG
n12011281	loop-int-tag 3.10	TCTCTCTGTGAGACGAGCTCTTCGATCTCCGCTCTG
n6191561	loop-int-tag 4.10	TACATCTCTTGAGACGAGCTCTTCGATCTCCGATCTGTCAGT
n1061070	loop-int-tag 5.10	TTGAGTGGAGTGGTGGACGAGGCTCTTCGATCTCCGATCTGAGAGTTG
n10604081	loop-int-tag 6.10	GCTCTGATGTCGACGAGCTCTTCGATCTCCGATCTGAGAGCTG

Figure 12

【图 13】

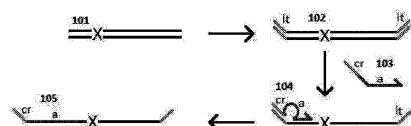


Figure 13

【图 9】

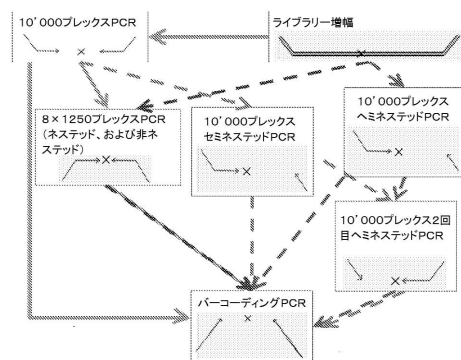


Figure 9

【図 10】

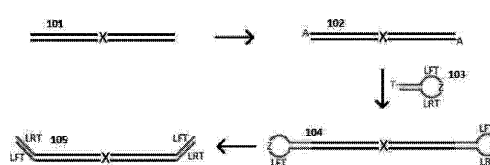


Figure 10

【図 14】

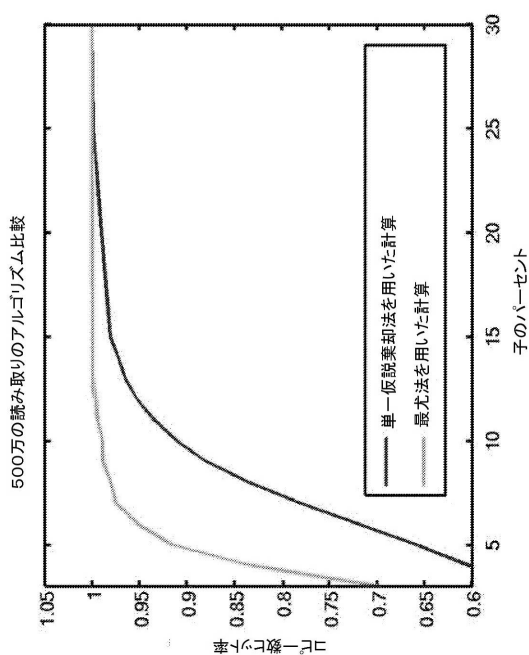


Figure 14

【図 15】

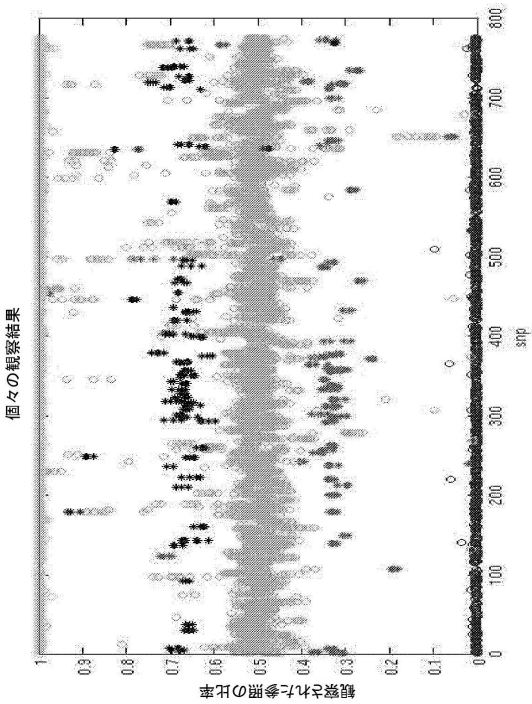


Figure 15

【図 16】

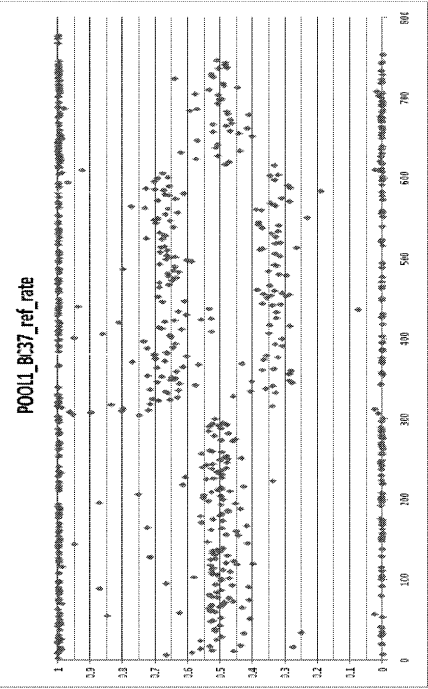


Figure 16

【図 17】

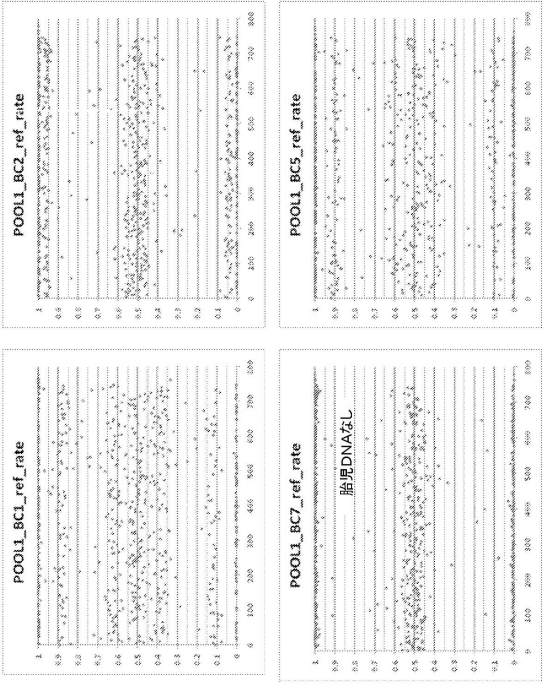


Figure 17

【図 18】

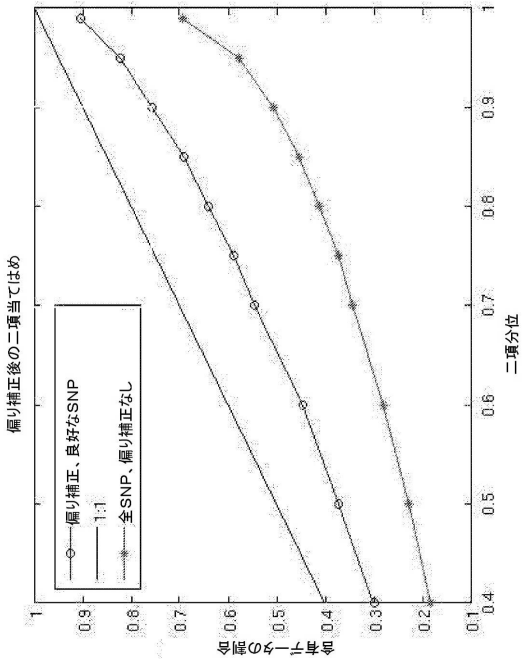


Figure 18

【図 19】

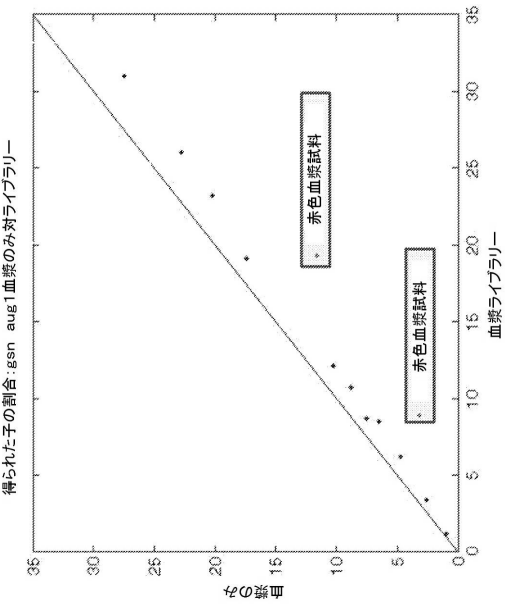


Figure 19

【図 20】

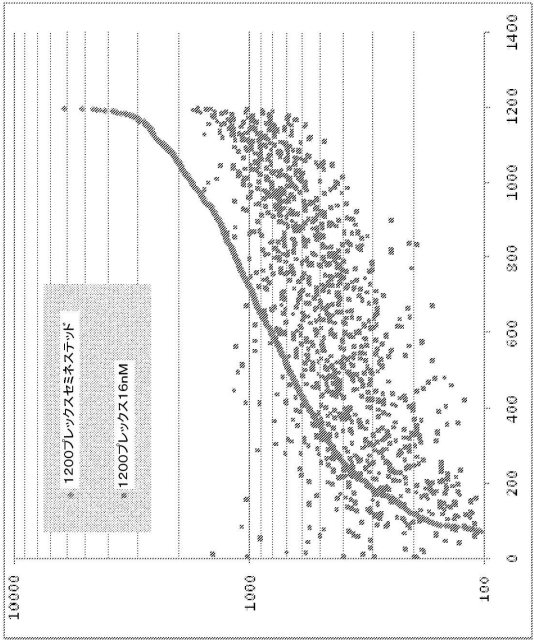


Figure 20

【図 21】

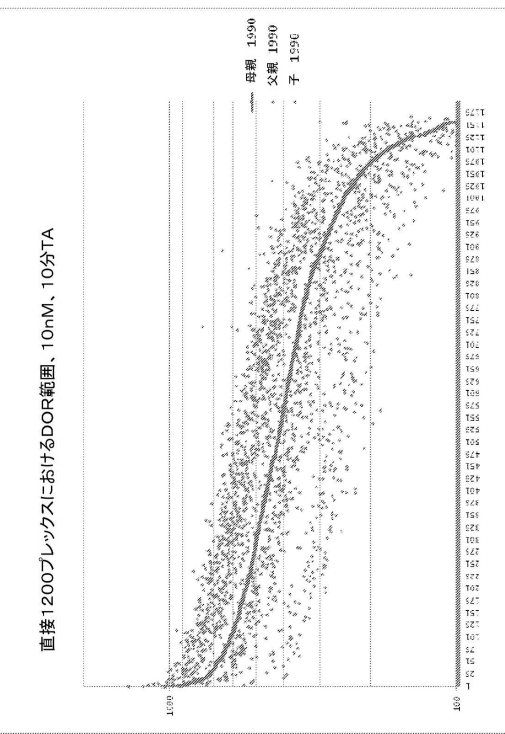


Figure 21

【図 22】

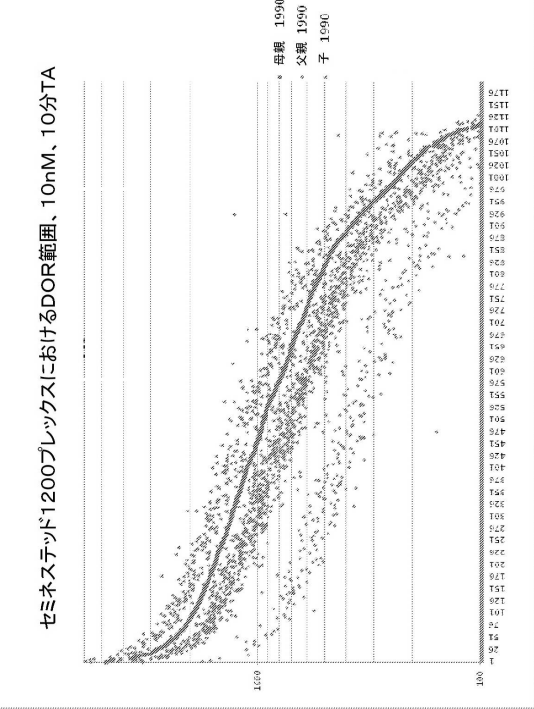


Figure 22

【図 23】

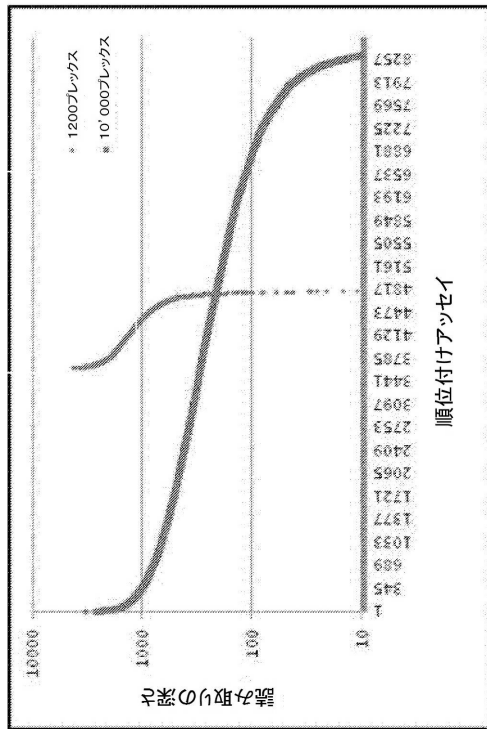


Figure 23

【図 24】

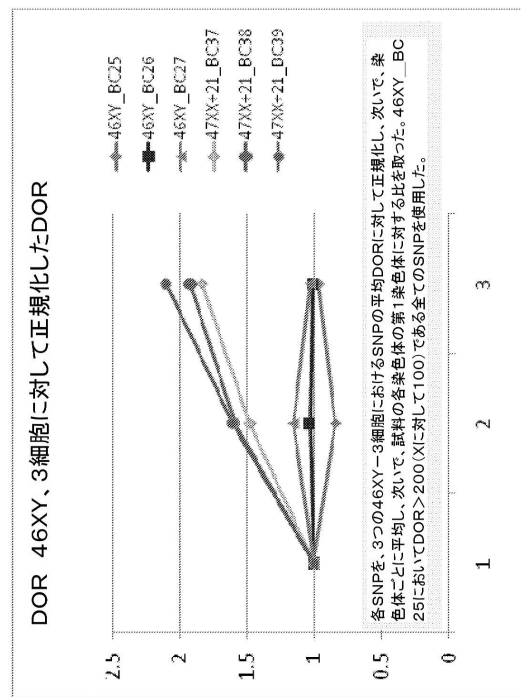


Figure 24

【図 25】

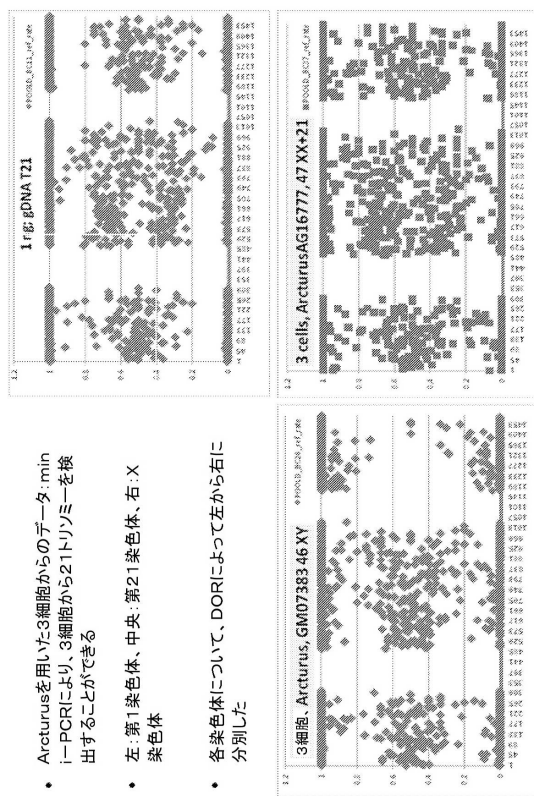


Figure 25

- Arcturusを用いた3細胞からのデータ: min PCRにより、3細胞から21トリソミーを検出することができる
- 左: 第1染色体、中央: 第21染色体、右: X染色体
- 各染色体について、DORによって左から右に分別した

【図 26】

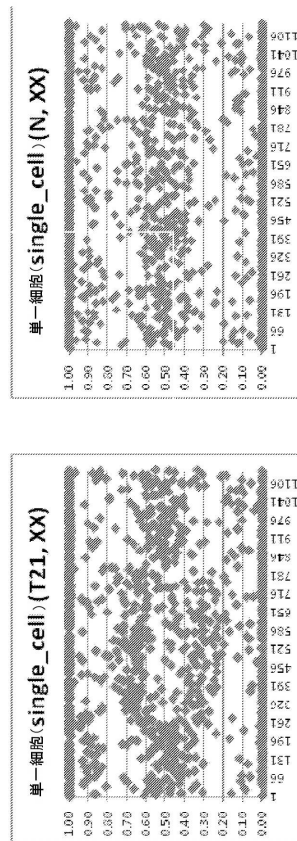


Figure 26

【配列表】

0006153874000001.app

フロントページの続き

- (31)優先権主張番号 61/516,996
(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)
- (31)優先権主張番号 13/110,685
(32)優先日 平成23年5月18日(2011.5.18)
(33)優先権主張国 米国(US)
- (31)優先権主張番号 61/571,248
(32)優先日 平成23年6月23日(2011.6.23)
(33)優先権主張国 米国(US)
- (31)優先権主張番号 13/300,235
(32)優先日 平成23年11月18日(2011.11.18)
(33)優先権主張国 米国(US)
- (31)優先権主張番号 61/542,508
(32)優先日 平成23年10月3日(2011.10.3)
(33)優先権主張国 米国(US)
- (74)代理人 100117019
弁理士 渡辺 陽一
- (74)代理人 100150810
弁理士 武居 良太郎
- (74)代理人 100141977
弁理士 中島 勝
- (72)発明者 ラビノビッツ, マシュー
アメリカ合衆国 カリフォルニア 94110, サンフランシスコ, ドロレス ストリート
1133
- (72)発明者 ゲメロス, ジョージ
アメリカ合衆国 カリフォルニア 95129, サン ホセ, カメオ ドライブ 1546
- (72)発明者 バンジェビック, ミレナ
アメリカ合衆国 カリフォルニア 94022, ロス アルトス ヒルズ, エレナ ロード
26799
- (72)発明者 ライアン, アリソン
アメリカ合衆国 カリフォルニア 94063, レッドウッド シティ, コロンビア アベ
ニュー 77
- (72)発明者 デムコ, ザカリー
アメリカ合衆国 カリフォルニア 94022, ロス アルトス ヒルズ, エレナ ドライブ
26799
- (72)発明者 ヒル, マシュー
アメリカ合衆国 カリフォルニア 94025, メンロー パーク, ロヨラ アベニュー 5
6
- (72)発明者 ツィーマーマン, ベルンハルト
アメリカ合衆国 カリフォルニア 94401, サン マテオ, ロディ アベニュー 166
2
- (72)発明者 パネル, ヨハン
アメリカ合衆国 カリフォルニア 94115, サンフランシスコ, ブロードウェイ ストリ
ート 2000, アpartment 309

- (56)参考文献 米国特許出願公開第2007/0184467 (US, A1)
PLoS One, 2010, Vol. 5, No. 10, e13184, doi: 10.1371/journal.pone.0013184

- (58)調査した分野(Int.Cl., DB名)

C12Q 1/68

C12N 15/00-15/90

JSTPlus/JMEDPlus/JST7580 (JDreamIII)

CAPLUS/MEDLINE/BIOSIS/WPIDS (STN)