

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5770737号
(P5770737)

(45) 発行日 平成27年8月26日(2015.8.26)

(24) 登録日 平成27年7月3日(2015.7.3)

(51) Int.Cl.	F 1
C 12 Q 1/68	(2006.01)
C 12 N 15/09	(2006.01)
G 06 F 19/18	(2011.01)
C 12 Q	1/68
C 12 N	15/00
G 06 F	19/18

請求項の数 47 (全 49 頁)

(21) 出願番号	特願2012-537410 (P2012-537410)
(86) (22) 出願日	平成22年11月5日 (2010.11.5)
(65) 公表番号	特表2013-509870 (P2013-509870A)
(43) 公表日	平成25年3月21日 (2013.3.21)
(86) 国際出願番号	PCT/EP2010/066935
(87) 国際公開番号	W02011/054936
(87) 国際公開日	平成23年5月12日 (2011.5.12)
審査請求日	平成25年1月10日 (2013.1.10)
(31) 優先権主張番号	61/259,076
(32) 優先日	平成21年11月6日 (2009.11.6)
(33) 優先権主張国	米国(US)
(31) 優先権主張番号	61/360,399
(32) 優先日	平成22年6月30日 (2010.6.30)
(33) 優先権主張国	米国(US)

早期審査対象出願

最終頁に続く

(54) 【発明の名称】 サイズに基づくゲノム分析

(57) 【特許請求の範囲】

【請求項 1】

胎児を身ごもっている女性対象から得られた生物試料中に配列不均衡が存在するかどうかを決定するための方法であって、ここで、生物試料は、ゲノムの核酸配列の部分である核酸分子を含み、該生物試料は胎児及び女性対象由来の核酸分子を含み、該方法は、

生物試料中の複数の核酸分子の各々について、

該核酸分子のサイズを測定し；

該核酸がゲノムのどの核酸配列に由来するかを同定し；

該ゲノムの核酸配列の第1の配列に由来する核酸分子のサイズ分布の値をコンピュータシステムにより決定し；及び

該サイズ分布の値とカットオフ値を比較し、それによって、配列不均衡が生物試料中の該第1の配列に存在するかどうかの分類を決定することを含む方法。

【請求項 2】

遺伝子座で妊娠女性対象の母親遺伝子型を決定し、ここで、第1の配列が該遺伝子座を含み；及び

該遺伝子座で妊娠女性対象の母親遺伝子型と決定されたサイズ分布に基づいて、該遺伝子座での母親遺伝子型と比較して、配列不均衡が生物試料中の第1の配列に存在するかどうかの分類を決定することをさらに含む、請求項1に記載の方法。

【請求項 3】

前記分類を用いて、該遺伝子座で胎児遺伝子型を決定することをさらに含む、請求項 2 に記載の方法。

【請求項 4】

妊娠女性対象の母親ハプロタイプを決定し、ここで、第 1 の配列は母親ハプロタイプの少なくとも一部を含み；及び

妊娠女性対象の母親ハプロタイプと決定されたサイズ分布に基づいて、母親ハプロタイプと比較して、配列不均衡が生物試料中の第 1 の配列に存在するかどうかの分類を決定する

ことをさらに含む、請求項 1 に記載の方法。

10

【請求項 5】

第 1 の配列における遺伝子座で胎児の父親の父親遺伝子型を決定し；及び

該遺伝子座での父親遺伝子型を用いて、母親ハプロタイプと比較して、試験される配列不均衡が過剰発明又は発現不足であるかどうかを決定する

ことをさらに含む、請求項 4 に記載の方法。

【請求項 6】

前記分類を用いて、該第 1 の配列に対応する胎児のハプロタイプを決定することをさらに含む、請求項 4 に記載の方法。

【請求項 7】

第 1 の配列が染色体であり、配列不均衡が胎児染色体異数性である、請求項 1 に記載の方法。

20

【請求項 8】

複数の核酸分子がサイズに基づいて予め選択されている、請求項 7 に記載の方法。

【請求項 9】

複数の核酸分子が 200 bp 未満となるように予め選択されるか、又は 150 bp 未満となるように予め選択される、請求項 8 に記載の方法。

【請求項 10】

複数の核酸分子が少なくとも 2 つのグループに予め選択され、各々は異なる範囲のサイズを有し、該方法は、

第 1 の染色体に由来する各グループの核酸分子のサイズ分布を決定し；

30

サイズ分布を比較し；及び

該比較に基づいて、胎児染色体異数性が第 1 の染色体に存在するかどうかを決定する

ことをさらに含む、請求項 8 に記載の方法。

【請求項 11】

第 1 の染色体に由来する核酸分子のサイズ分布の値を決定することが、

複数の染色体の各々について、染色体に由来する核酸分子のサイズから統計値を計算することを含み；及び

ここで、胎児染色体異数性が第 1 の染色体に存在するかどうかの分類を比較及び決定することが、

該統計値に基づいて染色体のランキングを決定し；

40

第 1 の染色体の決定されたランキングと参照生物試料から得られた第 1 の染色体の別のランキングとを比較し；及び

該比較に基づいて、胎児染色体異数性が第 1 の染色体に存在するかどうかの分類を決定する

ことを含む、請求項 7 に記載の方法。

【請求項 12】

第 1 の染色体に由来する核酸分子のサイズ分布の値を決定することが、

第 1 の染色体に由来する核酸分子のサイズから第 1 の統計値を計算し；

1 以上の第 2 の染色体に由来する核酸分子のサイズから第 2 の統計値を計算することを含み、

50

ここで、胎児染色体異数性が第1の染色体に存在するかどうかの分類を比較及び決定する

第1の統計値と第2の統計値との間の分離値を決定し、該分離値は、該第1の統計値と第2の統計値の間の差異又は第1の統計値と第2の統計値の比率を含み；

該分離値と1以上のカットオフ値を比較し；及び

該比較に基づいて、胎児染色体異数性が第1の染色体に存在するかどうかの分類を決定する

ことを含む、請求項7に記載の方法。

【請求項13】

分離値が、第1の統計値と第2の統計値との間の差異、又は該差異に由来する結果である、請求項12に記載の方法。 10

【請求項14】

第1の統計値と第2の統計値が、各染色体に由来する核酸分子について測定されたサイズの平均サイズを含む、請求項12に記載の方法。

【請求項15】

第1の統計値と第2の統計値が、各々、第1のサイズカットオフを下回るサイズを有する対応する核酸分子のサイズの第1の合計を含む、請求項12に記載の方法。

【請求項16】

第1の統計値が第2の合計で割った第1の合計を含み、第2の合計が、対応する核酸分子の全てのサイズについてのものであるか、又は第2のサイズカットオフを下回るサイズを有する対応する核酸分子のサイズについてのものであり、第2のサイズカットオフは第1のサイズカットオフよりも大きい、請求項15に記載の方法。 20

【請求項17】

第1の統計値が第2の合計で割った第1の合計を含み、第2の合計が、第2のサイズカットオフよりも大きい、対応する核酸分子のサイズについてのものである、請求項15に記載の方法。

【請求項18】

第2のサイズカットオフが第1のサイズカットオフであり、第2の合計が第3のサイズカットオフを下回るサイズのものであり、第3のサイズカットオフが第2のサイズカットオフよりも大きい、請求項17に記載の方法。 30

【請求項19】

複数の第1のサイズカットオフに対する第1の統計値と第2の統計値を計算し、及び第1のサイズカットオフの各々に対する第1と第2の統計値との間の分離値を決定することを繰り返し；並びに

該分離値の極値を同定する

ことをさらに含む、請求項15に記載の方法。

【請求項20】

前記極値と1以上のカットオフ値を比較することをさらに含む、請求項19に記載の方法。

【請求項21】

前記極値に対応する第1のサイズカットオフを同定し；及び

該極値に対応する第1のサイズカットオフと1以上のカットオフ値を比較することをさらに含む、請求項19に記載の方法。 40

【請求項22】

生物試料は、血液、血漿、血清、胎児細胞を含む母親血液、母親血液から得られた胎児細胞、尿、唾液、及び子宮内洗浄液を含む、請求項7に記載の方法。

【請求項23】

妊娠女性対象から得られた生物試料中に配列不均衡が存在するかどうかを決定するための操作を実行するためのプロセッサを制御するための複数の指示を記憶するコンピュータ読み取り可能な媒体であって、該生物試料は、核酸配列の一部である核酸分子を含み、該 50

生物試料が胎児及び女性対象由来の核酸分子を含み、該指示が請求項 1 ~ 2 2 のいずれか 1 項に記載のステップのうち 1 又は複数を含むコンピュータ読み取り可能な媒体。

【請求項 2 4】

胎児を身ごもっている女性対象から得られた生物試料中に染色体異数性が存在するかどうかを決定するための方法であって、該生物試料が胎児及び女性対象由来の核酸分子を含み、該方法は、

生物試料中の複数の核酸分子の各々について、

核酸分子のサイズを測定し；

核酸分子がどの染色体に由来するのかを同定し；

複数の染色体の各々について、

染色体に由来する核酸分子のサイズから統計値をコンピュータシステムにより計算し；
統計値に基づいて染色体のランクイングを決定し；

第 1 の染色体の決定されたランクイングと参照生物試料から得られた第 1 の染色体の別のランクイングとを比較し、それによって、胎児染色体異数性が第 1 の染色体に存在するかどうかの分類を決定する

ことを含む方法。

【請求項 2 5】

第 1 の染色体の決定されたランクイングと参照生物試料から得られた第 1 の染色体の別のランクイングとを比較することが、

該染色体の決定されたランクイングと生物試料についての参照染色体のランクイングとの間の第 1 の差異を決定し；

染色体の別のランクイングと参照生物試料の参照染色体のランクイングとの間の第 2 の差異を決定し；及び

該差異を比較する

ことを含む、請求項 2 4 に記載の方法。

【請求項 2 6】

複数の核酸分子の各々のサイズを測定することが、

生物試料を受入れ；及び

生物試料に含まれる複数の核酸分子の少なくとも一部を配列決定する
ことを含む、請求項 2 4 に記載の方法。

【請求項 2 7】

各核酸分子の配列決定された部分が各核酸分子の両末端を含む、請求項 2 6 に記載の方法。

【請求項 2 8】

統計値が各染色体に由来する核酸分子について測定されたサイズの中央値サイズを含む
、請求項 2 4 に記載の方法。

【請求項 2 9】

統計値が各染色体に由来する核酸分子について測定されたサイズの平均サイズを含む、
請求項 2 4 に記載の方法。

【請求項 3 0】

測定されたサイズが、長さ若しくは分子量、又はその長さと関連した測定されたパラメータである、請求項 2 4 に記載の方法。

【請求項 3 1】

測定されたサイズが該長さに対応する蛍光強度である、請求項 3 0 に記載の方法。

【請求項 3 2】

ランクイングが統計値の最低から最大である、請求項 2 4 に記載の方法。

【請求項 3 3】

ランクイングが統計値の最大から最低である、請求項 2 4 に記載の方法。

【請求項 3 4】

胎児を身ごもっている女性対象から得られた生物試料中に染色体異数性が存在するかど

10

20

30

40

50

うかを決定するための操作を実行するためのプロセッサを制御するための複数の指示を記憶するコンピュータ読み取り可能な媒体であって、該生物試料が胎児及び女性対象由来の核酸分子を含み、該指示が、請求項24～33のいずれか1項に記載のステップのうち1又は複数を含むコンピュータ読み取り可能な媒体。

【請求項35】

胎児を身ごもっている女性対象から得られた生物試料中に染色体異数性が存在するかどうかを決定するための方法であって、ここで、生物試料は胎児及び女性対象由来の核酸分子を含み、該方法は、

生物試料における複数の核酸分子の各々について、

核酸分子のサイズを測定し；

核酸分子がどの染色体に由来するかを同定し；

第1の染色体に由来する核酸分子のサイズから第1の統計値をコンピュータシステムにより計算し；

1以上の第2の染色体に由来する核酸分子のサイズから第2の統計値をコンピュータシステムにより計算し；

第1の統計値と第2の統計値との間の分離値を決定し；

分離値と1以上のカットオフ値とを比較し、それによって、胎児染色体異数性が第1の染色体について存在するかどうかの分類を決定する

ことを含む方法。

【請求項36】

分離値が第1の統計値と第2の統計値との間との差異であるか、又は該差異に由来する結果である、請求項35に記載の方法。

【請求項37】

該分離値が、第1の統計値と第2の統計値の比であるか、又は該比由来の結果である、請求項35に記載の方法。

【請求項38】

第1の統計値と第2の統計値が、各染色体に由来する核酸分子について測定されたサイズの平均サイズを含む、請求項35に記載の方法。

【請求項39】

第1の統計値と第2の統計値が、各々、第1のサイズカットオフを下回るサイズを有する、対応する核酸分子のサイズの第1の合計を含む、請求項35に記載の方法。

【請求項40】

第1の統計値が第2の合計で割った第1の合計を含み、第2の合計が、対応する核酸分子の全てのサイズについてのものであるか、又は第2のサイズカットオフを下回るサイズを有する対応する核酸分子のサイズについてのものであり、第2のサイズカットオフが第1サイズカットオフよりも大きい、請求項39に記載の方法。

【請求項41】

第1の統計値が第2の合計で割った第1の合計を含み、第2の合計が、第2のサイズカットオフよりも大きい、対応する核酸分子のサイズについてのものである、請求項39に記載の方法。

【請求項42】

複数の第1のサイズカットオフに対する第1の統計値と第2の統計値を計算し、及び第1のサイズカットオフの各々に対する第1と第2の統計値との間の分離値を決定することを繰り返し；並びに

該分離値の極値を同定する

ことをさらに含む、請求項39に記載の方法。

【請求項43】

前記極値と1以上のカットオフ値を比較することをさらに含む、請求項42に記載の方法。

【請求項44】

10

20

30

40

50

前記極値に対応する第1のサイズカットオフを同定し；及び
該極値に対応する第1のサイズカットオフと1以上のカットオフ値を比較する
ことをさらに含む、請求項42に記載の方法。

【請求項45】

1以上のカットオフ値が、生物試料における胎児DNAの割合の測定に基づいている、
請求項35に記載の方法。

【請求項46】

胎児を身ごもっている女性対象から得られた生物試料中に染色体異数性が存在するかどうかを決定するための操作を実行するためのプロセッサを制御するための複数の指示を記憶するコンピュータ読み取り可能な媒体であって、該生物試料が胎児及び女性対象由來の核酸分子を含み、該指示が請求項35～45のいずれか1項に記載のステップのうち1又は複数

10

を含むコンピュータ読み取り可能な媒体。

【請求項47】

請求項1～22、24～33、及び35～45のいずれか1項に定義される工程のうち1又は複数を実行するための手段を含み、さらに、少なくともプロセッサ及びメモリを含むコンピュータシステム。

【発明の詳細な説明】

【技術分野】

【0001】

20

関連出願への相互参照

本出願は、2009年11月6日に出願された「Detection of Chromosomal Imbalance」と題する米国仮特許出願第61/259076号、及び2010年6月30日に出願された「Size-Based Genomic Analysis」と題する米国仮特許出願第61/360399号からの優先権を主張する非仮特許出願であり、それらの全内容は、全ての目的で参照により本明細書中に援用される。

【0002】

また、本出願は、2008年7月23日に出願された、「Diagnosing Fetal Chromosomal Aneuploidy Using Massive Parallel Genomic Sequencing」と題する米国特許出願第12/178,181号（代理人受任番号第016285-005220US）；「Diagnosing Fetal Chromosomal Aneuploidy Using Genomic Sequencing With Enrichment」と題する米国特許出願第12/614350号（代理人受任番号第016285-005221US）；及び「Size-Based Genomic Analysis」と題する米国特許継続出願（代理人受任番号第016285-0066710US）に関連し、それらの全内容は、全ての目的で参照により本明細書中に援用される。

30

【背景技術】

【0003】

40

1997年において、母親血漿中に細胞不含胎児核酸の発見が、非侵襲性の出生前診断の新たな可能性を切り開いた（Lo YMら、Lancet 1997；350：485-487）。この技術は、臨床応用に急速に変えられ、これは、胎児由來の父方遺伝の遺伝子又は配列の決定、例えば、胎児性別決定、胎児RhD状態の決定、及び胎児が父方遺伝の突然変異を遺伝したかどうかの決定を行う。しかしながら、母親ゲノム及び父親ゲノムの両方に存在するゲノム標的、例えば、トリソミー21に関わる出生前診断は非常に多くのチャレンジとなっている。

【0004】

最近、優れた定量的正確さを有する単一分子カウント技術がこの課題に対する期待できる解決法となり得ることが示されている（Lo et al Proc Natl Ac

50

ad Sci USA 2007; 104: 13116 - 13121; Fan et al Anal Chem 2007; 79: 7576 - 7579; 米国特許出願第111/701,686号; Chiu et al Trends Genet 2009; 25: 324 - 331; Chiu et al Proc Natl Acad Sci USA 2008; 105: 20458 - 20463; Fan et al Proc Natl Acad Sci USA 2008; 105: 16266 - 16271)。このような方法は、死と正との間の選択されたゲノム位置からの分子数における定量的差異の観察を通じて診断目的を達成する。例えば、胎児ダウン症の診断について、第21染色体由来の分子数は、トリソミー21を胎児が持っている場合に増加する (Chiu et al Proc Natl Acad Sci USA 2008; 105: 20458 - 20463; Fan et al Proc Natl Acad Sci USA 2008; 105: 16266 - 16271)。
10

【0005】

しかしながら、このようなカウント技術は、限定された数のデータ点又は他の欠点を被る場合がある。したがって、既存技術と比較したある種の利点を有する出生前診断を行うための新規な方法、システム、及び装置を提供することが望まれる。

【発明の概要】

【課題を解決するための手段】

【0006】

本発明のある種の態様は、妊娠女性対象から得られた生物試料における配列不均衡（例えば、胎児染色体異数性）の出生前診断を行うためのサイズに基づく分析を用いることができるシステム、方法、及び装置を提供することができる。例えば、危険性のある染色体についての核酸分子の断片のサイズ分布を用いて、胎児染色体異数性を決定することができる。また、いくつかの態様では、生物試料（母親と胎児のDNAを含む）における配列不均衡などの他の配列不均衡を検出することができ、この場合、不均衡は母親遺伝子型、突然変異状態、又はハプロタイプに相関する。このような不均衡は、試料が胎児と母親からではなく、純粋に母親からであった場合に期待されるはずのサイズ分布と比較して、特定配列に対応する断片（核酸分子）のサイズ分布を介して決定され得る。（小さいサイズ分布への）シフトはある種の環境における不均衡を表すことができる。
20

【0007】

一態様では、互いに比較した各染色体からの断片のサイズ分布のランキング（例えば、サイズ分布を表す統計値）を用いて不均衡を決定する。例えば、試験試料における危険性のある染色体の断片のサイズのランキングは、参照生物試料から得られた危険性のある染色体についてのランキングと対比することができる。診断は、対比に基づいて実施可能である。例として、特定量によってランキングが変化する場合（例えば、核酸断片のサイズの減少を示す）、胎児染色体異数性が危険性のある染色体に存在するという診断がなされ得る。このようなランキング分析を用いた種々の態様では、全ての22個の常染色体と性染色体が用いられるか、又は一部の染色体が用いられる。
30

【0008】

別の態様では、危険性のある染色体の断片のサイズ（例えば、サイズ分布を表す統計値）と参照染色体の断片のサイズとの間の相違が用いられる。例えば、このサイズの相違がカットオフ（又は閾値と呼ばれる）よりも大きいか又は小さい場合、胎児染色体異数性が危険性のある染色体に存在するという診断がなされ得る。
40

【0009】

一例の態様によれば、妊娠女性対象から得られた生物試料における配列不均衡の出生前診断を行うための方法が提供される。生物試料は、核酸配列の一部である核酸分子を含む。生物試料中の複数の核酸分子の各々について、核酸分子のサイズが測定され、どの核酸配列に由来するかが同定される。コンピュータシステムは、第1の配列に対応する核酸分子のサイズ分布を決定する。決定されたサイズ分布に基づいて、配列不均衡が第1の配列に存在するかどうかの分類が決定される。
50

【0010】

別の例の態様によれば、妊娠女性対象から得られた生物試料中の胎児染色体異数性の出生前診断を行うための方法が提供される。生物試料中の複数の核酸分子の各々について、核酸分子のサイズが測定され、核酸分子がどの染色体に由来するかが同定される。コンピュータシステムは、染色体に対応する核酸分子のサイズから統計値を計算する。統計値は複数の染色体の各々について計算される。染色体のランキングは統計値に基づいて決定される。第1の染色体の決定されたランキングは、参照生物試料から得られた第1の染色体の別のランキングと対比される。対比に基づいて、胎児染色体異数性が第1の染色体に存在するかどうかの分類が決定される。

【0011】

10

別の例の態様によれば、妊娠女性対象から得られた生物試料中の胎児染色体異数性の出生前診断を行うための方法が提供される。生物試料中の複数の核酸分子の各々について、核酸分子のサイズが測定され、核酸分子がどの染色体に由来するかが同定される。コンピュータシステムは、第1の染色体に対応する核酸分子のサイズから第1統計値を計算する。コンピュータシステムは、1以上の第2染色体に対応する核酸分子のサイズから第2統計値を計算する。第1統計値と第2統計値との間の分離が決定される。分離値は1以上のカットオフ値と対比される。この対比に基づいて、胎児染色体異数性が第1の染色体に存在するかどうかの分類が決定される。

【0012】

20

本発明の他の態様は、本明細書に記載されている方法と関連したシステム及びコンピュータ読み取り可能な媒体に関する。一態様では、コンピュータ読み取り可能な媒体は、データ受入れ、データを分析するための指示を含むが、データを作製する（例えば、核酸分子を配列決定する）ように機械を指向するための指示を含まない。別の態様では、コンピュータ読み取り可能な媒体は、データを作製するように機械を指向するための指示を含む。一態様では、コンピュータプログラム製品は、本明細書に記載されている方法のための操作を行うためのプロセッサを制御するための複数の指示を記憶するコンピュータ読み取り可能な媒体を含む。また、態様は、潜在的には、各工程又は工程の各グループを実行するための異なるコンポーネントとともに、本明細書に記載されている方法のいずれかの工程を実施するように構成されたコンピュータシステムに関する。

【0013】

30

図面及び特許請求の範囲を含む本明細書の残り部分への参照は、本発明の態様の他の特徴及び利点を実現するであろう。さらなる特徴及び利点、並びに本発明の種々の態様の構造及び操作は、添付の図面を参照して以下に詳細に記載される。図面では、参照番号は同一又は機能的に類似したエレメントを指示し得る。

【図面の簡単な説明】

【0014】

【図1】本発明の態様による妊娠女性対象から得られた生物試料中の配列不均衡の出生前診断を行うための方法100を例証するフローチャートである。

【図2】本発明の態様による（Illumina Cluster Generation Reagent Kitのバージョン1を用いた場合）整列された配列のサイズの寒天で染色体の中央値ランキングを示すプロットである。

40

【図3】本発明の態様による母親血漿における異なる染色体に由来する配列のサイズ分析は、胎児の第21染色体異数性の非侵襲性出生前検出のために用いることができる。

【図4】本発明の態様によるサイズの統計値のランキングを用いて、妊娠女性対象から得られた生物試料中の胎児染色体異数性の出生前診断を行うための方法を例証するフローチャートである。

【図5】本発明の態様に従って整列されたサイズについて、第7染色体と第14染色体の比較を例証する表である。

【0015】

【図6】本発明の態様に従ってゲノム位置に関する断片のサイズの統計値の比較を用いて

50

、妊娠女性対象から得られた生物資料における配列不均衡の出生前診断を行うための方法を例証するフローチャートである。

【図7】本発明の態様に従ってカットオフサイズ(w)に対する短い断片(F)によって寄与された全長の画分のプロット700を示す。

【図8】本発明の態様に従って正倍数体に関するカットオフサイズに対する第21染色体(実線)と参照染色体(第13、18及び21を除く全ての常染色体)(破線)によって寄与された全長の画分のプロット800を示す。

【図9】本発明の態様に従って、トリソミー21妊娠に関するカットオフサイズに対する第21染色体(実線)と参照染色体(第13、18及び21を除く全ての常染色体)(破線)に関する F 値のプロット900を示す。

【図10】本発明の態様に従って、正倍数体とトリソミー21妊娠に関するカットオフサイズに対する第21染色体と参照染色体(第13、18及び21を除く全ての常染色体)との間の $F_{(21-Ref)}$ のプロット1000を示す。

【0016】

【図11】本発明に態様に従って120人の妊娠の性(gender)と分類のための表1100を示す。

【図12】本発明の態様に従って異なる疾患又は非疾患状態の4つの資料についての異なる染色体のランキングを例証する。

【図13】本発明の態様に従って、常染色体、トリソミー13、トリソミー18、及びトリソミー21の120症例についての第13染色体のランキングを示す。

【図14】本発明の態様に従って、常染色体、トリソミー13、トリソミー18、及びトリソミー21の120症例についての第18染色体のランキングを示す。

【図15】本発明の態様に従って、常染色体、トリソミー13、トリソミー18、及びトリソミー21の120症例についての第21染色体のランキングを示す。

【0017】

【図16】本発明の態様に従って整列された配列のサイズに関して第13染色体と第5及び第6染色体との比較を例証する表である。第5及び第6染色体と第13染色体との比較は、トリソミー13の妊娠における第21染色体配列のサイズの変化の検出に使用される。常染色体及びトリソミー18妊娠の結果は比較について含まれる。

【図17】本発明の態様に従って整列された配列のサイズに関して第18染色体と第12及び第14染色体との比較を例証する表である。常染色体及びトリソミー18妊娠の結果は本発明の態様による比較について含まれる。

【図18】本発明の態様に従って、150bpの第18染色体と参照染色体($F_{(18-Ref)}$)との間の短い断片によって寄与された全長の画分における相違を示す。

【図19】本発明の態様に従って、150bpの第21染色体と参照染色体($F_{(21-Ref)}$)との間の短い断片によって寄与された全長の画分における相違を示す。

【図20】本発明の態様に従って、異なる染色体(BCB1build 36、バージョン48)のGC含量のリストを示す。

【0018】

【図21】本発明の態様に従って、150bpの第13染色体と参照染色体($F_{(13-Ref)}$)との間の短い断片によって寄与された全長の画分における相違を示す。

【図22】本発明の態様に従って、第21染色体に対して整列された配列の中央値サイズとY染色体に対して整列された配列の割合との相関を示すプロットである。

【図23A】本発明の態様に従って、第18、第13及び第21染色体に対して各々整列された配列の中央値サイズとY染色体に対して整列された配列の割合との相関を示すプロットである。

【図23B】本発明の態様に従って、第18、第13及び第21染色体に対して各々整列された配列の中央値サイズとY染色体に対して整列された配列の割合との相関を示すプロットである。

【図23C】本発明の態様に従って、第18、第13及び第21染色体に対して各々整列

10

20

30

40

50

された配列の中央値サイズとY染色体に対して整列された配列の割合との相関を示すプロットである。

【図24】母親血漿DNA分析を用いた胎児異数性(トリソミー13及びトリソミー18)の非侵襲性検出のための本発明の態様及び別の方法の精度の比較を示す。

【図25A】本発明に従った妊娠女性と胎児遺伝子型の異なるシナリオについての略図を示す。

【図25B】本発明に従った妊娠女性と胎児遺伝子型の異なるシナリオについての略図を示す。

【図25C】本発明に従った妊娠女性と胎児遺伝子型の異なるシナリオについての略図を示す。 10

【0019】

【図26】本発明の態様に従って、母親がヘテロ接合えあり、父親がホモ接合である例を示す。

【図27】本発明の態様に従って、親のハプロタイプが図26に示されるものである場合に、胎児が母親からHaplotype Iを遺伝する例を示す。

【図28】本発明の態様に従って、第22染色体上の单一ヌクレオチド多型(SNP)タイプについてのサイズ分析を例証する表を示す。

【図29】本発明の態様に従って、第22染色体上のタイプSNPについてのサイズ分析を例証する表を示す。

【図30】本発明の態様に従って、第22染色体上のタイプとタイプSNPについての $F_{(Haplotype\ I - Haplotype\ II)}$ のプロットを示す。 20

【0020】

【図31A】本発明の態様に従って、標的富化なしに血漿DNAのサイズ分析を与える表である。

【図31B】本発明の態様に従って、標的富化なしに血漿DNAのサイズ分析を与える表である。

【図32】本発明の態様に従って、標的富化なしにT21及び正倍数体試料についてのFのプロットである。

【図33】本発明の態様に従うシステムと方法を用いた使用に適した例示的なコンピュータ装置のブロック図を示す。 30

【発明を実施するための形態】

【0021】

定義

用語「生物試料」とは、本明細書で使用するとき、対象(例えば、妊娠女性などのヒト)から採取され、対象とする1以上の核酸分子を含むいずれかの試料を意味する。

【0022】

用語「核酸」又は「ポリヌクレオチド」は、デオキシリボ核酸(DNA)又はリボ核酸(RNA)、及び一本又は二本鎖のいずれかの形態であるそのポリマーのことである。具体的に限定されない限り、当該用語は、基準核酸と同程度の結合特性を有する、既知の天然ヌクレオチドアナログを含む核酸を含み、そして天然に存在するヌクレオチドに同様の方法で代謝される。他に示されない限り、特定の核酸配列はまた、その保存的に改変された改変体(例えば、縮重コドン置換)、対立遺伝子、オルソログ、SNPs、及び相補的配列、ならびに明確に示された配列を暗黙的に含む。具体的には、縮重コドン置換は、一つ以上選択された(又は全ての)コドンの第三の位置が混合塩基及び/又はデオキシノシン残基で置換されている配列を生じさせることによって達成され得る(Batzler et al., Nucleic Acid Res. 19: 5081 (1991); Ohstsuka et al., J. Biol. Chem. 260: 2605-2608 (1985); 及びRossolini et al., Mol. Cell. Probes 8: 91-98 (1994))。用語、核酸は、遺伝子、cDNA、mRNA、低分子ノンコーディングRNA、マイクロRNA(miRNA)、Piwi相互作用RNA、及び遺伝子又は 40

遺伝子座によってコード化されたショートヘアピンRNA(s h RNA)と交換可能に使用される。

【0023】

用語「遺伝子」は、ポリペプチド鎖を産生することに關与するDNAのセグメント又は転写されたRNA産物のセグメントを意味する。それは、コード領域に先行する、及びコード領域に続く領域(リーダー及びトレイラー)、ならびに個々のコード断片(エキソン)間の介在配列を含み得る。

【0024】

本明細書において使用される用語「臨床的に重要な核酸配列」は、その潜在的不均衡が試験される、より大きな遺伝子配列断片に相当するポリヌクレオチド配列のことか、又はそのより大きな遺伝子配列自体に相当するポリヌクレオチド配列のことであり得る。一つの実施例は、第21番染色体の配列である。他の例は、第18番、第13番、X及びY染色体を含む。さらに他の例は、胎児がその親の片方又は両方から受け継いでいるかもしれない変異遺伝子配列、又は遺伝子多型、又はコピー数変異を含む。さらに他の実施例は、悪性腫瘍中において変異した、消失した、又は増幅された配列、例えばヘテロ接合性の欠損又は遺伝子重複が起こっている配列を含む。幾つかの実施形態において、複数の臨床的に重要な核酸配列、又は臨床的に重要な核酸配列の同等の複数のメーカーが、当該不均衡を検出するためのデータを提供するために使用され得る。例えば、第21番染色体上の5つの不連続配列からのデータは、可能性のある第21番染色体不均衡の決定のための付加的な方法において使用され得、効果的に試料の容量を1/5へと減少させる。

10

【0025】

用語「参照核酸配列」とは、本明細書において使用するとき、サイズ分布が標的配列に対して比較するために使用される核酸配列を指す。参照核酸配列の例には、染色体、染色体の一部、特定の遺伝子座(例えば、母親のもの)、特定のハプロタイプ、ゲノム、又は人工的に合成された核酸配列が含まれる。このような参照核酸配列は、試料中に内因的に存在してもよく、又は試料加工若しくは分析中に外因的に添加されてもよい。

いくつかの態様では、参照核酸配列は、疾患のない健常な状態を代表するサイズプロファイルを示す。

【0026】

本明細書において使用される用語「に基づく」は、「少なくとも一部に基づく」ことを意味し、そして、一つの方法の入力とその方法の出力との関係において生じる、一つの値の決定において使用される別の値(又は結果)のことをいう。本明細書において使用される用語「導く」はまた、一つの方法の入力とその方法の出力との関係のことであり、例えば誘導が公式計算であるときに起こる。

30

【0027】

本明細書において使用される用語「パラメータ」は、定量的データのセットを特徴づける数値、及び/又は定量的データセット間における数的関係を意味する。例えば、第一の核酸配列の第一の量と第二の核酸配列の第二の量との間における比率(又は比率の関数)はパラメータである。

【0028】

本明細書において使用するとき、用語「遺伝子座(locus)」又はその複数形の「遺伝子座(loci)」とは、ゲノム全体での改変を有する任意の長さのヌクレオチド(又は塩基対)の局在又は所在を意味する。

40

【0029】

用語「配列不均衡」とは、本明細書で使用するとき、参照量からの臨床的に関連した核酸配列における少なくとも1つのカットオフ値によって定義される任意の有意な逸脱を意味する。配列不均衡は、染色体用量不均衡、対立遺伝子不均衡、突然変異用量不均衡、ハプロタイプ用量不均衡、及び他の類似した不均衡を含むことができる。一例として、対立遺伝子不均衡又は突然変異用量不均衡は、胎児が母親由来の異なる遺伝子型を有する場合に生じ、それにより試料中の特定の遺伝子座で不均衡を創出することができる。

50

【0030】

本明細書において使用される用語「染色体異数性」は、2倍体ゲノムのもの由来の染色体量における変異を意味する。当該変異は、増加又は欠失であり得る。それは、一つの染色体の全体、又は一つの染色体の領域に関連し得る。

【0031】

用語「ハプロタイプ」とは、本明細書において使用するとき、同じ染色体又は染色体領域上の一緒に伝えられる、複数の遺伝子座での対立遺伝子の組み合わせを意味する。ハプロタイプは、僅かに一対の遺伝子座、染色体領域、又は全染色体を意味する場合がある。用語「対立遺伝子」とは、異なる表現型形質に帰着するか又は帰着しない可能性がある同じ物理的ゲノム遺伝子座での代替のDNA配列を意味する。任意の特定の2倍体生物では、各染色体（男性対象における性染色体を除く）の2つのコピーを有し、各遺伝子についての遺伝子型は、ホモ接合体においては同じであり、ヘテロ接合体においては異なる、その遺伝子座に存在する対立遺伝子の対を含む。生物の集団又は種は、典型的には、種々の個体間で各遺伝子座に複数の対立遺伝子を含む。1を超える対立遺伝子がこの集団に見られるゲノム遺伝子座は多形性部位と呼ばれる。遺伝子座での対立遺伝子改变体は、存在する対立遺伝子の数（すなわち、多型の程度）、又は集団におけるヘテロ接合体の比率（すなわち、ヘテロ接合性率）として測定可能である。本明細書において使用するとき、用語「多型」とは、ヒトゲノムにおける任意の個体間改变を意味し、その頻度とは関係しない。このような改变体の例には、限定されないが、単一のヌクレオチド多型、単一のタンデムリピート多型、挿入・欠失多型、突然変異（疾患原因となり得る）、及びコピー数の改变体が含まれる。

10

【0032】

妊娠女性の血漿中に存在する胎児DNA分子は、一般的には、母親由来の分子よりも短いことが見出されている（Chan et al Clin Chem 2004；50：88-92；Li et al Clin Chem 2004；50：1002-1011；米国特許出願第20050164241号）。本発明のある種の態様は、ゲノムの一部に由来するDNA分子のサイズ変化によって、ゲノムの特定の一部の過剰発現又は過少発現を胎児が有するかどうかを母親血漿DNAから決定することができる。胎児DNAは母親血漿中のDNAの少数画分を示すので、母親血漿中のサイズ変化の程度全体は小さい可能性があり、従って、検出が困難となり得る。いくつかの態様では、多数の分子のサイズは、疾患と健康との間の統計的に有意な差に達することが測定されている。

20

【0033】

I. 断片のサイズ決定

多数のDNA分子のサイズが測定可能な1つの方法は、大規模並列ゲノム配列決定による。これは、例えば、Illumina Genome Analyzerプラットフォーム（合成による配列決定を用いる）（Bentley DR et al Nature 2008；456：53-59）、ABI SOLiD（連結による配列決定を用いる）（McKernan et al Genome Res 2009；19：1527-1541）、Roche 454プラットフォーム（Marguelis et al Nature 2005；437：376-380）、及びHelicos單分子配列決定プラットフォーム（Harris et al Science 2008；320：106-109）によって実施され得る。また、他の大規模並列配列決定もまた用いることができる事が期待され、例えば、Pacific Biosciences（單分子、リアルタイム（SMRT（商標））技術（Eid et al Science 2009；323：133-138）、ナノポア配列決定（Clarke J et al. Nat Nanotechnol 2009；4：465-470）、半導体配列決定（例えば、Ion Torrent（www.iontorrent.com）による）が挙げられる。

30

【0034】

このような配列決定からDNAサイズ情報を得るための1つの方法は、対末端（PE）

40

50

配列決定を行うことであり、この場合、DNA分子の両末端が配列される。次に、分子の両末端に対応する配列は、参照ゲノム（例えば、参照ヒトゲノム又は参照ウマゲノム、あるいは対象とする任意の動物のゲノム）に戻ってマッピングすることができる。一態様では、両末端は、各々、各末端について個別に、参照ヒトゲノム（例えば、約10～24塩基又は25～36塩基）に戻ってマッピングするには十分に長い長さで配列決定される。別の態様では、配列の比率だけが、ヒトゲノムの非繰り返し領域に対してミスマッチを含まないで戻ってマッピングされ得る。一局面では、このマッピングは、両配列がともにマッピングに使用される場合、明白であってもよい。このシナリオでは、末端の各々が非常に短いため、信頼をもってマッピングされないとしても、両配列を用いることは明白なマッピングを提供することができる。分子のサイズは、2つの配列の末端のゲノム座標の差し引きによって実施され得る。

【0035】

別の態様では、分子のサイズは、ちょうど2つの末端の代りに、全DNA分子の完全な又は完全に近い配列決定によって得ることができる。これは、相対的に長い読み取り長さを用いた配列決定プラットフォーム、例えば、Roche 454 プラットフォーム、Pacific Biosciences 単一分子、リアルタイプ（SMRT（商標））技術、及びIon Torrent 技術（www.iontorrent.com）によって効率的に行うことができる。

【0036】

上記の配列決定に基づく方法のスループットは、インデックス又はバーコード化の使用を用いて増加させることができる（Cronn et al. Nucleic Acids Res 2008; 36: e122）。したがって、試料又は患者特異的なインデックス若しくはバーコードは、特定の核酸配列決定ライブラリーにおいて核酸断片に添加し得る。次に、多数のこのようなライブラリーは、各々が試料又は患者特異的なインデックス若しくはバーコードを有するものであり、ともに混ぜられ、ともに配列決定される。配列決定反応後、配列決定データは、各試料又はバーコード若しくはインデックスに基づく患者から回収することができる。この戦略は、スループットを増加させ、したがって、現発明のコストパフォーマンスを増大させ得る。

【0037】

別の態様では、生物試料中の核酸分子は、サイズ分析前に選択されるか又は分画され得る。1つの変形では、核酸分子は、ゲノム（例えば、第21、第18、第13、又はX染色体の1つ）における選択された遺伝子座からの核酸分子を選択的に結合するデバイス（例えば、プローブを含むマイクロアレイ又は溶液）で処理され、その後、サイズ分析は核酸分子の結合サブセット上で実施され得る。このような態様では、Nimblegen 配列捕捉システム（www.nimblegen.com/products/seqcap/index.html）又はAgilent SureSelect Target Enrichment System（www.opengenomics.com/SureSelect_Target_Enrichment_System）あるいは類似のプラットフォームを使用することができる。別の態様では、未結合核酸分子サブセットは、別々に除去されるか又は分解されるか又は消化され得る。

【0038】

少なくともいくつかの態様は、任意の単一分子分析プラットフォームを用いて実施することができ、この場合、分子の染色体起源及び長さは、例えば、電気泳動、光学的方法（例えば、光学的マッピングとその変形、en.wikipedia.org/wiki/Optical_mapping#cite_note-Nanocoding-3，及びJou et al. Proc Natl Acad Sci USA 2007; 104: 2673-2678）、蛍光に基づく方法、プローブに基づく方法、デジタルPCR（マイクロ流体に基づく、又はエマルションに基づく、例えば、BEAMing（Dressman et al. Proc Natl Acad Sci USA 2003; 100: 8817-8822）、RainDance（www.raindancete

10

20

30

40

50

ch.com/technology/pcr-genomics-research.a.s.p)、ローリングサイクル増幅、質量分析、融解分析(又は融解曲線分析)、分子ふるいなどにより分析することができる。質量分析の例として、長い分子が大きな分子量(サイズ値の例)を有するであろう。

【0039】

一例では、血漿DNA分子は、対末端の配列決定プロトコールを用いた Illumina Genome Analyzerシステムによって無作為に配列決定された。この実験では、Illumina対末端(PE) Cluster Generation試薬キットのバージョン1を用いた。各々の末端36bpについて配列決定された。各配列の2つの末端は、Illuminaによって提供されるGAPipeline-1.0ソフトウェアパッケージを用いて、繰り返しマスクされたヒトゲノム(NCBI Build 36,バージョン48)に対して整列された。各末端からの36bpのうちほんの32bpだけがアライメント目的で使用された。

【0040】

いくつかの態様では、PEは、以下の基準が一連の分析について使用され得ることを満たすように読む:(1)各々が示唆された対の個別のメンバーとともに、配列決定フローセル上で同じ染色体で配列決定され、ヒト参照ゲノムについて期待されるのと同じである、正しい配向を有する染色体に対して整列され得た;(2)対の両メンバーの配列決定された読み取りは、任意のヌクレオチドミスマッチないしに繰り返しマスクされたヒト参照ゲノムに対して整列され得た;(3)対の各メンバーの配列決定された読み取りは一意性スコア>4であった;及び対は600bp未満の挿入サイズを示す。次に、各々の整列された配列のサイズは、2つの末端の各々の位置に従って計算された。

【0041】

I I . 異数性状態を決定するためのサイズ分布の使用

図1は、本発明の態様による、妊娠女性対象から得られた生物試料における配列不均衡の出生前診断を行うための方法100を例証するフローチャートである。方法100は、胎児染色体異数性を分析することに関して主に記載される一方で、方法100の他の態様、及び本明細書の他の方法は、他の配列不均衡(例えば、遺伝子型又は突然変異の同定)に適用され得る。方法100及び他の方法は、1以上のプロセッサを含むコンピュータシステムによって、全体的又は部分的に実施されてもよい。

【0042】

方法100及び本明細書に記載されている方法のいずれかは、ステップを実行するように構成され得る、プロセッサを含むコンピュータシステムを用いて全体的に又は部分的に実施されてもよい。したがって、態様は、各ステップ又はステップの各グループを実行する異なるコンポーネントを潜在的に用いて、本明細書に記載されている方法のいずれかのステップを実行するように構成されたコンピュータシステムに指向される。番号付けされたステップとして示されているが、本明細書中の方法のステップは、同時に又は異なる順番で実行されてもよい。さらに、これらのステップの一部は、他の方法からの他のステップの一部を用いて使用されてもよい。また、全ての又は一部のステップは任意であってよい。さらに、方法のいずれかのステップのいずれかは、これらのステップを実行するためのモジュール、回路、又は他の手段を用いて実行され得る。

【0043】

ステップ110では生物試料中の核酸分子(例えば、DNA又はRNA)の少なくともいくつかのサイズが測定される。また、核酸分子は、全ゲノムの断片であるため、断片とも呼ばれる。このサイズは、任意の適切な方法、例えば、上記の方法を介して測定されてもよい。

【0044】

ステップ120は、核酸分子の各々が由来するゲノムにおける位置を同定する。この位置は、ゲノムの任意の部分であってもよく、提供される例についてヒトであるが、他のゲノムであってもよい。例えば、位置は、断片がどの染色体番号に由来し、ゲノム座標(

10

20

30

40

50

例えば、特定の座標又は座標の範囲)によって画定され得るように染色体のどの部分であり、さらに、染色体の2つ(倍数体を想定する)のうちのどの1つが断片に由来(配向)するかであってもよい。

【0045】

一態様では、同定は、配列情報を配列決定し、それと参照ヒトゲノム配列とを比較することによって実行されてもよい。別の態様では、この同定は、既知の染色体起源を用いて一連のプローブにハイブリダイゼーションさせることによって実行され得る。プローブは、マイクロアレイフォーマット又は溶液中で、1以上の蛍光標識を用いて標識され得る。さらに別の態様では、核酸分子は、溶液中又は固体表面上で一連のプローブによって捕捉されてもよく、次に、捕捉された(又は残りの捕捉されていない)核酸分子が配列決定される。いくつかの態様では、染色体異数性以外の配列不均衡が同定されるべきであるが、断片が起源とする染色体を同定するステップは任意であってもよい。

【0046】

ステップ130では、第1位置(例えば、第1の染色体)に対応する核酸分子のサイズ分布が決定される。種々の態様は、様々なサイズ分布を用いることができる。いくつかの態様では、サイズ分布は、他の染色体の断片と比較して、1つの染色体の断片のサイズのランキング(例えば、平均(average)、中央値、又は平均(mean))に関する。他の態様では、サイズ分布は、染色体の断片の実際のサイズの統計値に関する。一実施では、統計値は、染色体の断片の任意の平均(average)、平均(mean)、又は中央値のサイズを含むことができる。別の実施では、統計値は、全ての断片の全長によって割ってもよい、カットオフ値を下回る断片の全長を含むことができ、又は少なくとも断片は大きなカットオフ値を下回ってもよい。

【0047】

配列不均衡(例えば、胎児染色体異数性)は第1位置に存在するかどうかの分類は、決定されたサイズ分布に基づいて決定される。一態様では、染色体のランキングは、参照ランキング(例えば、正倍数体試料の1つ)と比較される。変化が有意である(例えば、閾値を超える)場合、試料は異数性として分類することができる。別の態様では、実際のサイズの統計値は、2つの染色体間、又は染色体グループ間で比較される。例えば、各統計値間の相違を採用してもよく、この相違はカットオフと比較される。

【0048】

I I . サイズ分布(ランク)

態様は、染色体不均衡が存在するかどうかを決定するために、試料の核酸断片のサイズの値を用いることができる。また、断片は、配列決定が行われた後に配列と呼ばれる。一態様では、複数の染色体についての断片のサイズの分布が決定され、その染色体が分布の統計値(例えば、平均(average)、平均(mean)、又は中央値)に基づいてランキングされる。便宜上、用語「サイズ」は、本明細書においては、同意語としてサイズの統計値とともに使用され得る。用語「サイズ」は、特定の断片のサイズ、及び一連の断片のサイズの統計的測定を指すときは明確であるべきである。

【0049】

A . ランク

図2は、母親血漿中の異なる染色体に由来するDNA断片の測定されたサイズ分布を例証するプロット200である。測定サイズは、インビオでのDNA断片サイズを反映し得るだけでなく、分析ステップの寄与も反映し得るので、サイズ分布は、(例えば、Illumina Genome Analyzer及びABI SOLiDプラットフォームについて)プラットフォームごとに変化し得た。しかしながら、参照試料及び試験試料は同じプラットフォーム又は試薬タイプを用いて分析される限り、態様はプラットフォーム/試薬に独立して用いられ得る。また、いくつかの態様は、例えば、任意のエラーが決定及び修正され得る場合、又はプラットフォーム及び/又は試薬タイプが密接に合致した分析性能をもつことを示し得る場合、異なるプラットフォーム及び/又は試薬タイプを用いて用いられてもよい。

10

20

30

40

50

【0050】

図2では、第2常染色体及びX染色体は、それらに整列された断片配列のサイズについて比較された。Y軸上で、ランクは、配列のサイズの降順であり、即ち、最長配列を有する染色体はランク1位となり、最短配列を有する染色体はランク23位となる。一態様では、この比較（ランクの分散の Kruskal-Wallis One Way Analysis、続く Bonferroni 修正対比較による）は、SigmaStat (SPSS) ソフトウェアを用いて非パラメトリック的になされた。この比較は、サイズを表す任意の統計値のものであり得て、個々のサイズの各々のランクイング、及び各染色体の配列の個別のランクイングの統計分析が含まれる。一実施では、ランク番号付けにおける同点及びスキップが許される。

10

【0051】

いくつかの態様では、配列の各々は、特定の染色体に対してマッピングされる。次に、各染色体について、1以上の統計値は、その染色体に対してマッピングされた配列について決定される。即ち、異なるタイプの統計値（例えば、平均及び中央値）は、各染色体のサイズについて計算し得る。次に、対応する統計値をランクイングすることができる。例えば、各染色体についての平均サイズを互いに比較することができる。1を超える統計値が用いられる場合、（例えば、いくつかの式に従って、例えば、重量平均）統計値を組み合わせることができ、この組み合わせた統計値をランクイングすることができる。一態様では、特定の染色体についての統計値のランクイングは、（統計値について上記されるように）組み合わせることができ、組み合わせたランクイングは互いに比較され得る。

20

【0052】

他の態様では、配列の全ては、サイズに従って個別にランクイングされる。即ち、1,000,000個の配列が存在する場合、ランクイングは1～100万あり、ランクイング番号付けにおいて同点及びスキップがあり得る。次に、同じ染色体に対してマッピングされた全ての配列についてのランクイングを一緒に加えることができる。ランクイングの合計は、染色体についての配列ランクイングに到達する特定の染色体に対して整列された配列の数によって割ることができる。最大平均配列ランクイングである染色体は、最長（ランクイング1位、Y軸上で最高）として記され、最低平均配列ランクイングである染色体は最短（ランクイング23、Y軸上で最低）として記され得る。別の態様では、中央値ランクを決定することができる。正倍数体の雄性胎児、正倍数体の雌性胎児、及びトリソミー21雄性胎児であるケースの中央値ランクを図2に示す。

30

【0053】

X軸上で、染色体は、各染色体（X染色体を除く（後記参照））由来のDNA断片のサイズ分布の降順で整列される。一態様では、このようなランクイングについて、正倍数体のケースだけを用いた。最長の測定サイズ（例えば、長さ）を有する染色体が左側に整列される。X染色体は、胎児の性によってそのランクイングが支配されるため、略図の右に置かれている。

【0054】

上記の通り、測定サイズは、プラットフォームごとに変化してもよい（例えば、Illuminaシステムから別のシステムに変更する場合）。このようにして、一局面では、サイズは、実際のサイズとは対照的に、「測定」サイズと称することができる。さらに、サイズは、Illuminaキットの1バージョンから別のものにスイッチする場合、例えば、バージョン1の対末端クラス生成キットからバージョン2に変更する場合、変更してもよい。一態様では、ユーザーは、彼らの特定のシステムについてランクイングを行うことができる。

40

【0055】

X染色体の平均ランクは、雄性胎児を持つ妊娠女性と比較した場合、雌性胎児を持つ妊娠女性の血漿中では低い（即ち、短くなっている）ことが図2に示され得る。この観察の説明は、胎児によって放出されるDNA断片が母親からのDNA断片よりも短いことである。したがって、X染色体の二重用量を放出することによって、雌性胎児は、母親血漿中

50

の X 染色体由来の全体の測定サイズ断片を減少させるであろう。反対に、雄性胎児は、X 染色体の単回用量を放出することができるだけであろう。

【 0 0 5 6 】

また、第 21 染色体のランクは、正倍数体胎児を有する妊娠女性と比較した場合、トリソミー 21 胎児を有する妊娠女性の血漿中で減少（即ち、短くなる）ことが図 2 に示され得る。この観察の説明は、再度、母親血漿中の第 21 染色体に由来する測定サイズ断片に由来し得る。反対に、正倍数体胎児は、胎児細胞あたり第 21 染色体の 2 用量を放出することができるだけであろう。

【 0 0 5 7 】

図 3 は、本発明の態様による母親血漿における異なる染色体由来の配列のサイズ分析が 10 、胎児の第 21 染色体の異数性の非侵襲性出生前検出について用いることができることを示すグラフ 300 である。この例では、染色体サイズは、他の染色体と比較した場合、サイズランクによって表される。このようにして、より大きなサイズのランク番号は、染色体が母親血漿においてより短い DNA 断片サイズを有することを示す。

【 0 0 5 8 】

グラフ 300 は、胎児が正倍数体（ランク 9 ~ 18 のランキング）である場合と比較して、トリソミー 21 胎児（ランク 18 ~ 21 のランキング）を伴う妊娠について、母親血漿中の第 21 染色体のサイズランク数が大きい（即ち、より短い DNA 断片を指示する）ことを示す。この観察の説明は、胎児 DNA が母親 DNA よりも短く、トリソミー 21 胎児由来の第 21 染色体の追加用量が母親血漿中の第 21 染色体配列の統計値の全短縮をもたらすことである。 20

【 0 0 5 9 】

B . ランクを用いた方法

図 4 は、本発明の態様による、統計値のランキングを用いた妊娠女性対象から得られた生物試料における胎児染色体異数性の出生前診断を行うための方法を例証するためのフローチャートである。

【 0 0 6 0 】

ステップ 410 では、生物試料から得られた複数の核酸分子（断片）の各々のサイズを測定する。この複数の核酸分子は、得られた核酸分子の全ての小集団となり得ることに気付かれない。複数の核酸分子は、さらに、配列決定がサイズ測定の一部として行われる場合、配列決定される全ての核酸分子の小集団となり得る。 30

【 0 0 6 1 】

ステップ 420 では、複数の核酸分子の各々がどの染色体に由来するかが同定される。種々の態様では、ステップ 410 と 420 の順番を反対にしてもよく、又は同時に生じさせてもよい。例えば、対末端の配列決定状況では、配列のゲノムアラインメントは、その染色体位置、及びその長さ（開始と終了のゲノム座標の差し引きによる）を提供することができる。一態様では、染色体はステップ 120 として同定され得る。

【 0 0 6 2 】

ステップ 430 では、複数の染色体の各々について、統計値は、染色体に対応する核酸分子のサイズから計算される。統計値は、本明細書に記載されている方法のいずれかにおいて計算され得る。例えば、統計値は、上述されるように、初期のランキング段階の結果を含むことができた。一態様では、任意の特定の染色体に対応する核酸分子のほんの一部を用いてもよい。 40

【 0 0 6 3 】

ステップ 440 では、染色体のランキングは、統計値に基づいて決定される。一態様では、ランキングは、基本のソーティングアルゴリズムを用いて決定されてもよい。別の態様では、より複雑な比較を行うことができ、例えば、ランクの相違の Kruskal - Wallis 一方向分析、続く Bonferroni 修正対末端比較、又は他の適した方法が挙げられる。種々の態様では、ランキングは、整数、画分、実数（例えば、ランクにおいて）、又は朱書きに基づく英数字ランキング（例えば、A - X ）であってもよい。 50

【0064】

ステップ450では、第1の染色体の決定されたランキングは、参照生物試料から得られた第1の染色体の別のランキングと比較される。一態様では、この比較は、1以上の参照生物試料のランキングから決定される、カットオフ閾値（例えば、単一の値又は範囲）に対する決定されたランキングの比較である。このようなカットオフ値は、図3から決定され得るように、ランクが18（若しくは19）又はそれより高い場合であってもよい。別の態様では、2つの試料間の第1の染色体のランキングにおける相違を決定することができ、この相違はカットオフと比較され得る。一実施では、参照生物試料は、対象とする疾患を含まないことが分析により決定され、さらに、試料が、ランキングにより問題を引き起こす可能性がある任意の関連疾患を含まないことが決定されてもよい。

10

【0065】

ステップ460では、比較に基づいて、胎児染色体異数性が第1の染色体に存在するかどうかの分類が決定される。一態様では、この分類は疾患の又は疾患でないという二元分類であってもよい。別の態様では、分類は、予測できない分類が用いられ得るという点で三元的であってもよい。さらに別の態様では、分類は、特定の分類の可能性を含むことができ、したがって、ちょうど3つの分類を超える分類を効果的に含むことができる。

【0066】

I I I . サイズ分布（サイズの統計値）

他の態様では、サイズ分布は、ランキングとは対照的に、断片サイズの統計値（例えば、特定のゲノム位置に対する実際値又は絶対値の統計値）を含むことができる。一態様では、第1の染色体の実際のサイズは、同じ試験試料についての1以上の参照染色体の実際のサイズと比較し得る。例えば、第1の染色体と1以上の参照染色体のこれらの実際のサイズ間の分離値（例えば、相違又は比率）をカットオフと比較することができる。一態様では、カットオフは参照試料から決定し得る。別の態様では、試験試料と参照生物試料との間の染色体の断片の実際のサイズ間の分離値は、カットオフとともに用いられてもよい。さらに別の態様では、染色体の断片の実際のサイズは、信頼できる分類を得るためのカットオフに対して比較されてもよい。

20

【0067】

A . 絶対サイズ

いくつかの例は、第21染色体由来の断片の絶対サイズと、1以上の参照染色体由来の断片の絶対サイズとを比較することによって、トリソミー21の非浸襲性出生前検出を達成することができることを示す。一態様では、第7と第14染色体は、第21染色体に相対的に近い母親血漿におけるサイズ値（例えば、絶対サイズ又はサイズランキング）を有することができるので、参照染色体として選択され得る。実際には、参照染色体は、正倍数体試料についての第21染色体（又は他の対象染色体）に関して、例えば、特定の分析プラットフォーム及び/又は試薬タイプ上で、一致するする断片サイズを有する任意の染色体であってもよい。

30

【0068】

図5は、本発明の態様による、各染色体に対して整列された配列のサイズに関して、第21染色体と第染色体及び第14染色体との比較を例証する表500である。表500のデータは、16個の試験試料から得られる。各試料について、第7、第21、及び第21の各々の断片について平均サイズが示される。また、平均値間の相違が提供される。p値は、各相違が健常試料において生じるであろう可能性を示す。

40

【0069】

図5の表500から分かるように、第21染色体に対して整列された配列は、全てのトリソミー21の妊娠についての第7染色体と第14染色体に対して整列された配列よりも（例えば、平均サイズで）有意に短かった（Mann - Whitneyランク - sum検定、p値 < 0.001）。この程度の統計的有意性での短縮は、正倍数体妊娠において観察されなかった。したがって、表500は、第21と第7染色体間の平均断片サイズの相違が1 bpよりも大きかった一方で、正倍数体症例のいずれも1 bpよりも大きい相違を

50

示した。したがって、1 bp は、分類を決定するための正確なカットオフを提供し得る。同様に、トリソミー 21 妊娠の全てについて、第 14 染色体からの平均断片サイズが、第 21 染色体からの平均断片サイズよりも一貫して大きい。確かに、0.5 bp のカットオフが、第 21 染色体と比較した場合に第 14 染色体断片の観察された「伸長」について用いられる場合、全てのトリソミー 21 症例は、非トリソミー 21 症例と区別することができる。このようにして、一態様では、カットオフ値は、1 以上の参照試料から決定することができる。

【0070】

B. 絶対サイズを用いた方法

図 6 は、本発明の態様による、ゲノム位置についての断片のサイズの統計値の比較を用いた、妊娠女性対象から得られた生物試料における配列不均衡の出生前診断を行うための方法を例証するフローチャートである。一局面では、方法 600 は、第 1 の染色体の断片のサイズと 1 以上の参照染色体の断片のサイズについての分離値（例えば、相違又は比率）に基づく配列不均衡の分類の決定に指向され得る。

10

【0071】

ステップ 610 では、生物試料から得られた複数の核酸分子のサイズが測定される。複数の核酸分子が得られ、ステップ 410 に記されたのと同じ断片を含んでもよいことに気付くべきである。

【0072】

ステップ 620 では、核酸の各々が由来するゲノム位置を同定する。この位置は、ステップ 120 及び他のところで記載されているように、ゲノムの任意の部分であってもよい。例えば、複数の核酸分子の各々がどの染色体に由来するかが同定される。この決定は、本明細書で記載されるように、参照ゲノムに対するマッピングによって行うことができる。

20

【0073】

ステップ 639 では、第 1 統計値は、第 1 ゲノム位置（例えば、第 1 の染色体）に対応する核酸分子のサイズから計算される。一態様では、第 1 統計値は、第 1 の染色体に対応する断片の平均（average）、平均（mean）、又は中央値サイズであってもよい。別の態様では、第 1 統計値は、カットオフのタイプであってもよい、第 1 サイズを下回る断片の長さの合計を含むことができる。例えば、200 bp よりも小さい断片の各々は、合計されたそれらの長さを有することができる。この合計は、（第 1 サイズと同じであってもよい）第 1 の染色体に対応する全ての断片の長さの合計、又は第 2 サイズカットオフよりも大きな断片の長さの合計などの別の数によって割られてもよい。例えば、第 1 統計値は、断片の全長と比較した第 1 サイズカットオフを下回る断片の全長の比率、又は大きな断片の全長と比較した小断片の全長の比率であってもよい。

30

【0074】

ステップ 640 では、第 2 統計値は、第 2 ゲノム位置（例えば、第 2 染色体）に対応する核酸分子のサイズから計算される。第 2 染色体は参照染色体と考えることができる。一態様では、複数の参照染色体についての統計値を計算することができる。一実施では、統計値は、該統計値が 1 以上の第 2 染色体のものとなり得るように組み合わせることができる。別の態様では、複数の参照染色体についての統計値は、以下に示すように、個別に比較されてもよい。

40

【0075】

ステップ 650 では、第 1 統計値と第 2 統計値は分離値を得るために比較される。一態様では、分離値は、第 1 統計値と第 2 統計値との間の相違であり得て、決定される。別の態様では、分離値は、2 つの統計値の比率であってもよい。なお別の態様では、複数の分離値は決定され得て、例えば、各参照染色体について計算され得る、各第 2 統計値についてのものであってもよい。

【0076】

ステップ 660 では、分離値は、1 以上のカットオフ値と比較される。一態様では、こ

50

の比較は、複数の分離値の各々について行うことができる。例えば、上記されるように、異なる分離値は、第1統計値と各第2統計値との間で決定することができる。種々の実施では、各分離値は、同一又は異なったカットオフ値と比較されてもよい。別の態様では、分離値は、該分離値が特定の範囲内にあるかどうかを決定するために2つのカットオフ値と比較される。この範囲は、非正常なデータ点（異数性）が生じるかどうかを決定するために1つのカットオフを含むことができ、第2カットオフは、データ点が、測定又は分析におけるエラーによって引き起こされる可能性があるかどうか（例えば、分離値が、疾患試料についてでさえ、なおも期待されたであろうものよりも大きいかどうか）を決定するために用いることができた。

【0077】

10

ステップ670では、配列不均衡が（例えば、胎児染色体異数性）が第1ゲノム位置に存在するかどうかの分類が、比較に基づいて決定される。一態様では、複数のカットオフ（例えば、N個のカットオフ）が単一の分離値に用いることができる。このような態様では、N+1個の分類を決定することができる。例えば、2つのカットオフを用いて、正倍数体（正常又は健常）、中間、及び正倍数体（疾患又は不健康）の分類を決定してもよい。複数の比較が行われる（例えば、各分離値についてのもの）別の態様では、分類は比較の各々に基づくことができる。例えば、ルールに基づく方法は、該比較の各々から得られる分類を調べることができる。一実施では、最終的な分類は、全ての分類が一致している場合に与えられるだけである。別の実施では、大多数の分類が用いられる。なお別の実施では、より複雑な式は、分離値の各々が各カットオフ値にどれくらい近いかに基づいて用いることができ、これらの近さを分析して、分類を決定することができる。例えば、この近さを（他の因子、例えば、標準化とともに）まとえることができ、結果を別のカットオフ値と比較することができた。

【0078】

20

他の態様では、方法600の変形は、参照試料に由来し得る、第1の染色体についての統計値とカットオフ値との直接的比較に適用され得る。いくつかの態様はまた、非妊娠個体からの生物試料を分析するために用いることができる。このような分析は、ちょうど、試料の断片の全てのサイズについての統計値を見ることができ、配列不均衡が存在し得るかどうかを決定するために、統計値と分離値とを比較することができる。不均衡が存在するものとして分類される場合、不均衡の位置のさらなる分析は、例えば、特定のゲノム位置（例えば、染色体）についての統計値及び/又は分離値を分析することによって行うことができる。

【0079】

30

C. 短い断片の全長の使用

上述したように、いくつかの態様では、血漿DNAのサイズ分布はまた、短いDNA断片によって寄与される全DNA長の画分によって反映され得る。例えば、サイズ分布は、カットオフ値を下回る断片の全長を含むことができ、それは、全ての断片の全長、又はより大きなカットオフ値を下回る少なくとも断片で割ってもよい。反対に、血漿DNAのサイズ分布はまた、長いDNA断片によって寄与される全DNA長の画分によって反映され得る。例えば、サイズ分布は、カットオフ値を上回る断片の全長を含むことができ、それは、全ての断片の全長、又は類似のカットオフ値を下回る少なくとも断片で割ってもよい。別の例として、小比率v/s大比率はまた用いることができる。1つの態様は、短い血漿DNA断片を定義するためにカットオフとして150bpを用いる。しかしながら、任意のカットオフ、例えば、130bp、140bp、160bp及び170bpはまた、短いDNA断片を定義するためのカットオフとして用いることができる。本明細書で使用するとき、塩基対はまた、一本鎖断片の長さを参照して、多数のヌクレオチド(nt)と同義語として、言及されてもよい。

【0080】

40

一態様では、短いDNA断片によって寄与されるDNA長の断片の計算は以下の通りであり得る：F = 短いDNA断片によって寄与されるDNA長の画分；S = 全ての短いDN

50

A断片の長さ（カットオフに等しいか又は下回る長さ）の合計；及びそれらの長さに関わらずに試料中の全てのDNA断片の全長。したがって、画分は、サイズの統計値の一例である、 $F = S / T$ によって与えることができる。Fの計算は、試料由来の全ての断片、又はゲノムの特定の位置、例えば、特定の染色体についてであってもよい。

【0081】

一実施では、試料中の全てのDNA断片の全長が決定され得る。次に、DNA断片が「短い断片」として定義される下回るカットオフサイズ(w)が選択され得る。カットオフサイズは、異なる診断目的に併せるために変更及び選択され得る。短いDNA断片の全長は、カットオフサイズに等しいか又は短い全てのDNA断片の長さを合計することによって計算することができる。短いDNA断片によって寄与される全長の画分、以下のように計算することができる：

$$F = \frac{w \text{ 長さ}}{600 \text{ 長さ}}$$

ここで、 w は、カットオフ(w)($b p$)に等しいか又は小さい長さを有するDNA断片の長さの合計を表し；及び

600 は、 $600 b p$ に等しいか又はそれより小さいDNA断片の長さの合計を表す。また、統計値は、ランキングを用いた態様において使用され得る。例えば、Fは、一連の各ゲノム位置（例えば、染色体）からの断片について計算することができる。

【0082】

例証の目的で、全長は、以下の実施例における $600 b p$ よりも短い断片の合計によって計算された。しかしながら、他のサイズ限定、例えば、 $400 b p$ 、 $500 b p$ 及び $700 b p$ は、「全長」を計算するために用いることができる。この例では、全長は、 $600 b p$ 以下のDNA断片に基づいて計算された。これは、*Illumina Genome Analyzer (Solexa)*システムが、 $600 b p$ よりも長いDNA断片の增幅及び配列決定に有効でないためである。さらに、 $600 b p$ よりも短いDNA断片に分析を限定することはまた、ゲノムの構造的变化から生じるバイアスを避けることができる。構造的变化の存在、例えば、再配列(Kidd JM et al, *Nature* 2008; 453: 56-64)では、DNA断片のサイズは、サイズが参照ゲノムに対してDNA断片の末端をマッピングすることによって生物情報学的に推定される場合に過大評価され得る。さらに、首尾よく配列決定され、参照ゲノムに対してマッピングされた99.9%を超える全てのDNA断片は $600 b p$ 未満であり、したがって、 $600 b p$ に等しいか又はそれより短い全ての断片は、試料におけるDNA断片のサイズ分布の典型的な見積りを与えるであろう。

【0083】

上記されるように、DNA断片について、より短いサイズ分布へのシフトは、胎児が副本を有する染色体について観察され得る。一局面では、異数性の危険性のある染色体（標的染色体）と異数性の危険性がない染色体（参照染色体）との間の短いDNA断片によって寄与された全長の画分の差異の測定は定量的測定であり、これらの染色体に由来するDNA断片についてのサイズ分布が異なるかどうかを決定することができる。

【0084】

一態様では、本発明者らは、それぞれ、異数性の危険性がある染色体と参照染色体（単数又は複数）についての短いDNA断片によって寄与される全長の画分として $F(T_{ar})$ と $F(R_{ef})$ を定義する。標的染色体と参照染色体（単数又は複数）との間の短いDNA断片によって寄与される全長の断片の差異（ $F(T_{ar} - R_{ef})$ ）は、 $F(T_{ar} - R_{ef}) = F(T_{ar}) - F(R_{ef})$ として計算することができる。例えば、 $F(21-1)$ は、第21染色体と第1の染色体についての短いDNA断片によって寄与された全長の画分間の差異である。胎児染色体異数性の出生前診断における $F(T_{ar} - R_{ef})$ の適用は、以下の節において検討される。別の態様では、 $F(T_{ar}) / F(R_{ef})$ の比率はまた、 $F(T_{ar} - R_{ef})$ が用いられるのと同じようにして、分離値として使用され得る。

【0085】

10

20

30

30

40

50

標的染色体と参照染色体の F 値間の差異は、胎児が標的染色体についてトリソミックであるか否かを決定するための統計値として用いることができる。胎児が標的染色体についてトリソミックである場合、胎児のトリソミー染色体は、母親血漿への短い胎児DNAの追加用量に寄与するであろう。よって、標染色体配列のサイズ分布の見掛けの短縮をもたらす。標的染色体配列のサイズ分布のこの短縮は、標的配列の短いDNA断片 (F_{target}) によって寄与された配列長の画分において増加へと導くであろう。結果として、 F_{target} 及び F_{ref} 間の差異 F において増加となるであろう。

【0086】

図7は、雄性胎児による母親血漿試料のためのカットオフサイズ (w) に対する短い断片 (F) によって寄与された全長の画分のプロット700を示す。常染色体とY染色体に対して整列されたDNA断片についての F 値は、「短いDNA断片」を定義するために使用されるカットオフサイズに対してY軸上にプロットされる。雄性妊娠では、Y染色体に整列されたDNA分子は、雄性胎児から放出されたDNAを表す。母親血漿における大部分の循環DNA分子は母親に由来するため、常染色体に対して整列されたDNA断片は選択的に母親DNA断片を表す。 F 値は、試料における全てのDNA断片がカットオフサイズより短いか又は等しい場合、カットオフサイズとともに増加し、1.0の値に到達する。DNA分子の2つの種間のサイズ分布における差異は、それらの F 値における差異によって反映され得る。より高い F 値は、全長のより高い画分が短い画分によって寄与されることを指示し、したがって、DNA断片のより短いサイズ分布を指示する。

【0087】

プロット700に示されるように、Y染色体からのDNA分子についてのサイズ分布は、常染色体からのDNAのサイズ分布よりも短い。具体的には、Y染色体についての F 値は、常染色体よりも先に生じ、80bp ~ 350bp間で $F_{常染色体}$ よりも高い F_Y をもたらす。Y染色体と常染色体との間の F 値の差異 ($F_{(Y-常染色体)}$) は、カットオフサイズに対してさらにプロットされ、破線によって表され、80bp ~ 350bp間でポジティブである。 $F_{(Y-常染色体)}$ の最大値は0.23であり、150bp周辺で生じる。下記の実施例に例証されるように、危険性のある染色体と参照染色体(単数又は複数)との間の短い断片 (F) によって寄与された全断片長の画分における差異は、それらのサイズ分布における差異を定量するための有用な分離値である。さらに例として、任意のサイズカットオフを用いて F を決定することができ、例えば、130bp ~ 170bpの間である。

【0088】

図8は、正倍数体妊娠についてのカットオフサイズに対して、第21染色体(実線)及び参照染色体(第13、第18及び第21染色体を除く全ての常染色体)(破線)に関する短い断片 (F) によって寄与された全長の断片のプロット800を示す。2つの F 値 ($F_{(21-Ref)} = F_{(chr\ 21)} - F_{(Ref)}$) における差異は、破線によって示されている。第21染色体と参照染色体からのDNA断片のサイズ分布が正倍数体妊娠において同じであるため、 $F_{(21-Ref)}$ はいずれのカットオフサイズについてもゼロに近い。

【0089】

図9は、トリソミー21妊娠についてのカットオフサイズに対する第21染色体(実線)と参照染色体(第13、第18及び第21染色体を除く全ての常染色体)(破線)に関する F 値のプロット900を示す。 $F_{(21-Ref)}$ は破線によって示されている。胎児からの第21染色体の追加用量によって、第21染色体についての母親血漿におけるDNA断片のサイズ分布は、参照染色体よりも短い。DNA断片のサイズ分布のこの差異は、約150bpで最大である0.016に到達する、 $F_{(21-Ref)}$ のポジティブ値によって反映される。

【0090】

図10は、正倍数体とトリソミー21妊娠についてのサイズカットオフに対する第21染色体と参照染色体(第13、第18及び第21染色体を除く全ての常染色体)との間の

10

20

30

40

50

$F_{(21-Ref)}$ のプロット 1000 を示す。 $F_{(21-Ref)}$ の増加はトリソミー 21 において観察されるが、 $F_{(21-Ref)}$ は正倍数体ケースについての各サイズカットオフで約ゼロである。最大 $F_{(21-Ref)}$ は 150 bp 周辺で観察されるので、150 bp での差異は、第 21 染色体配列のサイズ分布の任意の有意な短縮があるかどうかを決定するために分離値として用いることができる。しかしながら、正倍数体とトリソミーケース間の有意差が生じる任意のサイズを用いることができ、例えば、限定されないが、140 bp、145 bp、155 bp 及び 160 bp である。この例では、150 bp で観察された第 21 染色体と参照染色体の全長の画分における差異は、それぞれトリソミー 21 と正倍数体妊娠について 0.016 と -0.002 である。

【0091】

10

短い DNA 断片によって寄与された全 DNA 長の画分の形状におけるこの差異は、正倍数体と異数性胎児間を区別するために使用することができる。種々の方法で差異を試験することができる。一態様では、特定のサイズカットオフで F の値（例えば、分離値の例）は、試料のカテゴリー化（分類）を決定するためにカットオフ値と比較可能である。別の態様では、 F のピーク値を見出すことができ、その値は 1 以上のカテゴリー化カットオフ値と比較され得る。種々の態様では、ピーク値は、最大若しくは最小値、最大 / 最小値近くの平均値、又は最大 / 最小値と関連しているか若しくはそれに由来する他の値であつてもよい。分離値（例えば、 F ）の他の統計値も用いることができ、例えば、ピーク幅、又はピークに対応する特定のサイズカットオフの位置である。

【0092】

20

一態様では、特定のゲノム位置又は全ゲノムについての複数の断片から得られた F 値（又は本明細書に記載されている他の統計値、例えば、大きな断片の長さで割った小さな断片の長さ）は、病因が存在するかどうかを決定するために用いることができる。例えば、統計値がカットオフ値を超える場合、病因は、小さな断片量が正常範囲外であるため、存在するものとして同定され得る。これは、妊娠男子以外の患者について行われ、胎児以外の疾患を同定することができた。

【0093】

30

いくつかの態様では、物理的サイズの分画が、断片のサイズ分析前に行うことができる。一態様では、核酸分子は、2 つのサイズ画分（例えば、200 bp より大きいもの、及び 200 bp より小さいか又はそれに等しいもの）に分けることができ、これらのサイズ画分の各々に対する選択された染色体（例えば、第 21 染色体）のサイズ分布を比較することができる。胎児トリソミー（例えば、トリソミー 21）の存在下で、より小さい分子サイズのサイズ画分は、より大きな分子サイズのサイズ画分と比較して、相対的に多量に増加するであろう。

【0094】

40

他の態様では、長さカットオフ値を下回る多数の画分はサイズ分布の代りに用いることができる。例えば、長さカットオフ値の数は、1 以上の参照染色体と比較して、標的染色体（例えば、第 21 染色体）について（例えば、差異又は比率）比較することができる。一態様では、長さカットオフ値を下回る断片の数は、割合を得るために、断片の総数によって割られ、この割合は、パラメータを与えるために、染色体と 1 以上の参照染色体との間で比較され得る。得られたパラメータ（例えば、差異又は比率）は、カットオフ値（例えば、1 %）と比較することができる。一局面では、長さカットオフは長さで選択されて もよく、この場合、上記割合は最大である。

【0095】

V. ランクを用いた例

トリソミー 21 に加えて、母親血漿における断片分析は、他の胎児染色体異数性、例えば、トリソミー 13、トリソミー 21、トリソミー 18 及び性染色体の異数性（例えば、ターナー症候群、クラインフェルター症候群及び XYY など）の非侵襲性出生前検出に用いることができる。また、態様は、染色体異常が特定の染色体（例えば、染色体転座によって引き起こされるトリソミー 21）の部分だけを伴う場合に用いることができる。この

50

ようなシナリオでは、断片サイズ以上は、影響を受けた染色体領域からのDNA断片について観察されるであろう。

【0096】

図11は、製造業者の使用説明書に従って、多重試料調製キット(Illumina)を用いて構築された母親血漿DNAのライブラリーについての表1100を示す。区別可能なバーコードを付した全ての2つの試料は、 Illumina Genome Analyzer II 上で1つのレーンに導入され、その後、標準の多重化された対末端配列決定に供された。試料はバーコードに基づいて区別され得た。120人の妊娠女性の血漿試料を分析した。胎児の性別及び染色体異数性状態を表1100に示す。

【0097】

図12は、本発明に態様に従って、異なる疾患の4つの試料又は非疾患状態について、異なる染色体のランキングを示す。第22染色体及びX染色体は、先に記載されたようにそれらの断片のサイズに従ってランキングされた。見ることができるように、 Illumina Cluster Generation Reagent Kit のバージョン2に基づく染色体の相対的ランキングは、このキットのバージョン1を用いた場合と異なる(図2)。以下の示すように、態様(例えば、方法400)は、異数性を正倍数体症例とは区別され得る。

【0098】

A. トリソミー13

この例では、本発明者らは、トリソミー13の出生前診断についての態様の使用を示す。図13は、例えば、方法400に起因する場合のように、120人の妊娠について第13染色体に関する第13染色体のランクのグラフを示す。このグラフでは、T13、T18、T21及びEuは、それぞれ、トリソミー13、トリソミー18、トリソミー21及び正倍数体妊娠を指す。23人のトリソミー13妊娠のうちの18人(78.3%)において、第13染色体は、22又はそれを下回ってランキングされる一方で、97人の非トリソミー13妊娠のほんの2人(2.1%)において、第13染色体は22又はそれを下回ってランキングされた。したがって、22のカットオフランクを用いると、胎児のトリソミー13の出生前診断について染色体断片のランク分析の感受性及び特異性は、それぞれ78.3% > 及び97.9%であった。

【0099】

図13では、第13染色体のランクは正倍数体について高く(即ち、小さい数ランクを示す)、トリソミー18と21妊娠は、トリソミー13妊娠よりも高い。換言すると、他の染色体上の配列と比較すると、第13染色体配列は、非トリソミー13妊娠よりもトリソミー13妊娠について短いようである。トリソミー13妊娠における第13染色体配列の見掛けの短縮は、胎児の過剰な第13染色体により、第13染色体に対して胎児DNAの寄与の増加が原因している。

【0100】

B. トリソミー18

図14は、120症例について染色体18のランキングを示す。30人のトリソミー18のうち26人(86.7%)では、第18染色体のランクは13よりも低く(即ち、大きな数がランクを示す)、一方、90人の非トリソミー18症例のいずれも13よりも低いランクはなかった。したがって、カットオフとして13のランクを用いると、胎児トリソミー18の出生前診断についての染色体断片のランク分析の感受性及び特異性は、それぞれ86.7%及び100%である。

【0101】

この分析では、本発明者らは、トリソミー18、トリソミー13、トリソミー21、及び正倍数体妊娠における第18染色体のサイズランクを比較している。第18染色体に照らして、後者の3つのグループは、第18染色体に関わる用量異常性がないため、「正常な」対照としてみなすことができる。図14から分かるように、正倍数体とトリソミー13についての第18染色体のランクは約1~3に集中している。一方、トリソミー1

10

20

30

40

50

8症例についての第18染色体のランクは、13～22であり、第18染色体の断片のサイズが正倍数体、トリソミー21、及びトリソミー13症例よりも短いことを示す。再度、過剰な第18染色体は、これらの観察を説明する。

【0102】

C. トリソミー21

図15は、120症例についての第21染色体のランクを示す。9人のトリソミー21症例のうち8人(88.9%)では、第21染色体は22又はそれを下回ってランクされ、一方、111人の非トリソミー21症例のうちいずれも、第21染色体は22又はそれを下回るランクを有した。したがって、カットオフとして22のランクを用いると、胎児トリソミー21の出生前診断についての染色体断片サイズのランク分析の感受性及び特異性は、それぞれ88.9% > 及び100%であったであろう。

10

【0103】

V. サイズの差異を用いた例

A. トリソミー13

これらの次の例では、本発明者らは、例えば、方法600について記載されるように、第13染色体由来の断片の全体サイズと、1以上の参照染色体由来の断片のものと比較することによって、トリソミー13の非浸襲性出生前検出を達成できることを示す。この例は、先の例におけるものと同じデータセットを用いる。例証するように、トリソミー13検出に関しては、参照染色体として第5染色体と第6染色体を選択している。

20

【0104】

図16の表1600から分かるように、同じ試料内で、第13染色体に対して整列された配列は、全てのトリソミー13妊娠について、第5染色体と第6染色体に対して整列された配列よりも有意に短かった(Mann-Whitneyランク-合計検定、p値<0.001)。対照として、正倍数体とトリソミー18妊娠を含む。正倍数体とトリソミー18妊娠の両方において、第13染色体の用量は正常であった。表1600から分かるように、第13染色体由来の配列についてこのような統計的有意のサイズ異常性は、正倍数体とトリソミー18妊娠において観察されなかった。

【0105】

さらに、全てのトリソミー13妊娠について、同じ試料内で、第13染色体と第5染色体との間の平均断片サイズの差異は0.4bpより大きく、一方、非トリソミー13のいずれも0.4bpを超える差異を示さなかった。同様に、全てのトリソミー13妊娠について、第13染色体と第6染色体との間の平均断片サイズの差異は、0.5bpより多く、一方、非トリソミー13症例のいずれも0.5bpを超える差異を示さなかった。

30

【0106】

B. トリソミー18

例証として、トリソミー18の検出について、第14染色体を参照染色体として選択した。非トリソミー18妊娠について、図17の表1700において、第18染色体由来の配列は、第14染色体由来の配列よりも統計的有意に長いことが分かる(Mann-Whitneyランク-合計検定、p値<0.005)。しかしながら、トリソミー18症例に関して、第18染色体配列は、第14染色体に整列された配列よりも有意に長くなかった。正倍数体とトリソミー18症例は、0bpのカットオフ値を用いて、第18と第14染色体からのDNA断片ntの平均サイズ間の差異に基づいて決定することができる。これらの観察は、母親よりも短い胎児由来の第18染色体の追加用量は、胎児がトリソミー18を有する場合、このような配列の全サイズを減少させたであろうという事実によって説明され得る。次に、これは、第14染色体の分布に近い第18染色体の全サイズ分布をもたらす。

40

【0107】

V. 全序寄与を用いた例

以下の例において、参照染色体は、第13、第18、及び第21染色体を除く全ての常染色体からなっていた。

50

【0108】

図18は、150bpで第18染色体と参照染色体との短い断片によって寄与された全長の画分の差異($F_{(18-Ref)}$)を示す。 $F_{(18-Ref)}$ について0.0003の診断カットオフ値を用いて、トリソミー18妊娠は93.3%感受性と100%特異性で検出可能である。

【0109】

図19は、150bpで第21染色体と参照染色体との間の短い断片によって寄与された全長の画分における差異($F_{(21-Ref)}$)を示す。 $F_{(21-Ref)}$ について0.007の診断カットオフ値を用いて、トリソミー21妊娠は100%感受性と100%特異性で検出可能である。

10

【0110】

VII. 参照染色体(単数又は複数)の選択

1以上の参照染色体は、母親血漿DNAのサイズ分析によって、例えば、方法600を用いて、胎児染色体異数性の非侵襲性出生前検出について様々な方法において選択され得る。様々な態様では、異なる参照染色体を選択することができる。

【0111】

第1タイプの参照染色体は、母親血漿に潜在的に関与する染色体(例えば、第21、第18、又は第13染色体)由来のものとして、特異的分析プラットフォーム(又は密接に合致した分析実行を用いる分析プラットフォーム)上での類似したサイズ分布を、参照染色体由来のDNA断片が提示するものである。一態様では、このタイプの分析では、胎児染色体異数性は、危険性のある染色体のサイズランキング又は絶対サイズが、参照染色体(単数又は複数)からの統計的有意な減少を示す場合に検出される。

20

他の態様では、危険性のある染色体由来の断片と参照染色体(単数又は複数)由来の断片との間の平均又は中央値サイズ差異を測定することができる。

【0112】

第2タイプの参照染色体は、胎児が正倍数体である場合、母親血漿中の異数性に潜在的に関与する染色体(例えば、第21、第18、又は第13)由来の断片よりも、参照染色体由来のDNA断片が統計的に短いものである。このタイプのシナリオは、特定のプラットフォームを用いて測定する場合、危険性の染色体が最も長いものの1つであるときに直面することがある。例えば、図12において、第18染色体由来のDNA断片は、母親血漿において測定される場合、常染色体のうち最も長い。したがって、第18染色体よりも断片が統計的有意に短い参照染色体を選択することができる。一態様では、この体報の分析では、胎児染色体異数性は、危険性のある染色体のサイズランキング又は絶対サイズは、参照染色体(単数又は複数)とは統計的有意に異なることが分からぬ場合に検出される。例えば、この戦略は、上記されるように、図17についても分析に用いられている。

30

【0113】

第3のタイプの参照染色体は、胎児が正倍数体である場合、母親血漿中の異数性に潜在的に関与する染色体(例えば、第21、第18、又は第13)由来の断片よりも、参照染色体由来のDNA断片が統計的に長いものである。このタイプのシナリオは、特定のプラットフォームを用いて測定する場合、危険性の染色体が最も短いものの1つであるときに直面することがある。したがって、断片が危険性のある染色体よりも統計的有意に長い参照染色体を選択することができる。一態様では、このタイプの分析では、胎児染色体異数性は、サイズランキング、ランキングにおける差異、又は参照染色体と危険性のある染色体との間の絶対サイズが増加する場合に検出される。

40

【0114】

第4タイプの参照染色体は類似のGC含量を有するものである。染色体のGC含量は、配列決定反応において定量的読み出しに影響を及ぼす可能性がある。染色体間のGC含量における差異から得られる潜在的なバイアスを最小にする1つの方法は、類似のGC含量を有する適切な参照染色体を選択することである。図20は、異なる染色体のGC含量のリストを示す(NCBI build 36, バージョン48)。染色体はGC含量の昇

50

順で列挙される。ここで、類似のG C含量を有する参照染色体を用いた例を示す。

【0115】

図21は、150bpで第13染色体と参照染色体との間の短い断片によって寄与された全長の画分における差異($F_{(13-Ref)}$)を示す。ここで、例証として、本発明者らは、 $F_{(21-Ref)}$ 分析について参照染色体として第3、第4、第5及び第6染色体を用いた。図20に示されるように、第3～6染色体はG C含量が36.53%～38.79%であり、第13染色体の39.07%のG C含量に類似している。0.0038の診断カットオフを用いると、 F 分析は、トリソミー13症例を検出し、感受性95.7%、特異性99.0%であった。

【0116】

GCバイアス結果は、異なる程度に異なる配列決定プラットフォームに影響を及ぼす可能性があると期待される。例えば、事前增幅を要求しないプラットフォーム、*Helicos* プラットフォーム(Harris TD, et al. Single-molecule DNA sequencing of a viral genome. *Science* 2008; 320: 106-9)、ナノ細孔(Lund J, Parviz BA. Scanning probe and nanopore DNA sequencing: core techniques and possibilities. *Methods Mol Biol* 2009; 578: 113-22)、又はPacific Biosciencesの単一分子リアルタイプシステム(Eid J, et al. Real-time DNA sequencing from single polymerase molecules. *Science* 2009; 323: 133-8)の使用は、参照群の染色体のより広い選択を可能にし得た。

10

【0117】

IV. 胎児DNA濃度の使用

トリソミー21(又は他のトリソミー)妊娠では、胎児は、母親細胞由来のものよりは短いであろう第21染色体断片の過剰用量を母親血漿に放出するであろう。これらの短い断片の濃度は、母親血漿中の胎児DNAの濃度と相関するであろうと期待される。換言すると、Y染色体由来の配列の画分濃度が高くなるにつれて、トリソミー21妊娠における第21染色体由来の配列の測定サイズは短くなるであろう。

【0118】

20

図22は、男子トリソミー21胎児を伴う妊娠に油浦することを示す。図22に示されるように、第21染色体配列の中央値サイズと、Y染色体に整列された配列の割合との間の負の相関が確かにある($r = -0.942$ 、ピアソン相関)。また、類似の傾向は、第21染色体のサイズランクを用いる場合に期待され、即ち、第21染色体のランク数が増加し、画分胎児DNA濃度が増加する場合により短い断片を指示する。相関があるため、態様は、本明細書に記載されている方法のいずれかにおいてパラメータとして胎児DNA濃度の測定を用いることができる。

【0119】

このタイプの分析では、母親血漿中の胎児DNAの画分及び絶対濃度は、当業者に知られている任意の方法によって測定可能である。胎児が雄性である場合、胎児DNA濃度は、母親血漿中のY染色体由来の配列の画分濃度によって測定され得る。別の例は、單一ヌクレオチド多型又は单一タンデムリピート多型又は挿入-欠失多型などの父親遺伝の遺伝子マーカーの使用である。別の例は、胎児と母親のDNA間で別々にメチル化される領域などのエピジェネティックマーカーの使用である(Poon et al. Clin Chem 2002; 48: 35-41; Chiu et al. Am J Pathol 2007; 170: 941-950; Chan et al. Clin Chem 2006; 52: 2211-2218; US米国特許第6,927,028号)。上記マーカーは、当業者に知られている方法を用いて分析することができ、例えば、ポリメラーゼ連鎖反応(PCR)、デジタルPCR、配列決定、大規模な並列配列決定及び標的化された大規模な並列配列決定が挙げられる。

30

40

50

【0120】

一態様では、母親血漿中の測定された胎児DNA濃度に関連して、染色体異数性を検出するための診断敷地（例えば、本明細書に記載されている任意の方法についてのカットオフ値）を変更することができた。したがって、相対的に高い胎児DNAを含む母親血漿試料について、異数性を潜在的に伴う染色体由来の血漿DNA分子の短縮の程度は、相対的に低い胎児DNA濃度を含む母親血漿試料よりも劇的であることが期待されたであろう。

【0121】

Fはサイズ分布と関連するため、Fもまた胎児DNA濃度と相関を示す。図23Aは、 $F_{(18-Ref)}$ （第18染色体vs参考配列）と胎児DNA濃度との間の相関を示す。13人のT18症例のうち10は雄性胎児を有し、胎児DNAの画分濃度は、これらの試料における染色体Y配列の画分濃度によって評価され得る。 $F_{(18-Ref)}$ とY染色体配列の画分濃度との間に有意な相関があった（ $r = 0.879$ 、スピアマン相関）。これらの結果は、母親血漿中の第18染色体配列のサイズ分布における短縮の程度が、トリソミー18妊娠における母親血漿中の胎児DNAの画分濃度と相関することを示唆する。

【0122】

診断カットオフを下回る症例を白丸で表し、一方、カットオフより大きな差異を有する症例を黒丸で表す。診断カットオフ（0.0003）より小さいF値を有する2つの症例（白丸で表される）は、0.0003より大きなF値を有する症例（黒丸で表される）と比較した場合、胎児DNAの相対的に低い画分濃度を有した。低い画分胎児DNA濃度は、これらの2つの症例が図18において実行された分析において失われていた理由であり得た。したがって、一態様では、胎児濃度は試料において低い場合、分類は無視されるか又はやり直しされ得る。

【0123】

図23Bは、 $F_{(21-Ref)}$ （第21染色体vs参考染色体）と胎児DNA濃度との間の相関を示す。9個のT21症例のうち5つは雄性胎児を有していた。 $F_{(21-Ref)}$ の値とY染色体配列の画分濃度との間に有意な相関があった（ $r = 0.9$ 、スピアマン相関）。

【0124】

図23Cは、 $F_{(13-Ref)}$ （第13染色体vs第3、第4、第5及び第6染色体）と胎児DNA濃度との間の相関を示す。23個のトリソミー13症例のうちの14は雄性胎児を有していた。胎児DNAの画分濃度は、試料におけるY染色体配列の画分濃度によって推定され得る。 $F_{(13-Ref)}$ とY染色体配列の画分濃度との間のポジティブな相関があった（ $r = 0.644$ 、スピアマン相関）。診断カットオフ（0.0038）を下回る $F_{(13-Ref)}$ を有する症例を白丸で表す。低い画分胎児DNA濃度は、この症例が、図21において実行された分析において見当たらなかった理由であり得た。

【0125】

IX. サイズ分析と分子カウンティング法の比較

図24は、母親血漿DNA分析を用いて、胎児異数性（トリソミー13とトリソミー18）の非浸襲性検出についての本発明の態様の精度と別の方法の精度との比較を示す。この例は、サイズを用いた態様vs分子カウンティングに基づく方法の比較を例証する（米国特許出願第11/701,686号；Chiu et al Trends Genet 2009；25:324-331；Chiu et al Proc Natl Acad Sci USA 2008；105:20458-20463；Fan et al Proc Natl Acad Sci USA 2008；105:16266-16271；米国特許出願公開第2009/0029377号）。8つの母親血漿試料（2つの正倍数体、2つのトリソミー18、及び4つのトリソミー13）は、Chiuら（Proc Natl Acad Sci USA 2008；105:20458-20463）によって報告されている方法を用いて分析された。各々の症例について、分

10

20

30

40

50

子カウンティングを用いた Chiu による z スコアを用いた従前に報告されている分子カウンティング法は、サイズ分析に基づく態様の結果と比較された。

【 0 1 2 6 】

z スコアの計算について、対象とする染色体の割合表現は、最初に各症例について計算された。次に、染色体表現の平均及び標準偏差を参照症例について計算した。このデータセットは、症例 1、2、5、6、8 及び 8 が、第 18 染色体表現の平均及び標準偏差の計算について参照群として用いられた。症例 1、2、3 及び 4 は、第 13 染色体表現の平均及び標準偏差の計算についての参照群として用いられた。z スコアは、参照群の平均からの標準偏差の数として定義される。染色体の有意な過剰発現を $z \text{ スコア} > 3$ として定義する。症例 3 及び 4 はトリソミー 18 を担持し、第 18 染色体断片はそれらの血漿中で過剰発現された。症例 5、6、7 及び 8 はトリソミー 13 症例を担持したが、症例 5 及び 7 だけがそれらの血漿中で第 13 染色体の過剰発現を示した。症例 6 及び 8 は、トリソミー 13 胎児を持つにも関わらず、母親血漿中に第 13 染色体の有意な過剰発現を示さなかった。

【 0 1 2 7 】

染色体の DNA 断片サイズの有意な短縮は、特定の染色体に対して整列された全ての断片のサイズと、参照染色体に対して整列されたサイズとを比較することによって検出された。Mann-Whitney 試験はこの比較に用いられ、P 値 < 0.0001 は、有意差の存在として定義された。第 13 染色体断片のサイズの分析について、参照染色体は第 5 染色体であった。正倍数体又はトリソミー 18 胎児を有する全ての症例について、第 13 染色体の断片サイズは、第 5 染色体とは有意に異なっていなかった。トリソミー 13 の胎児を有する 4 つの症例について、第 13 染色体談判は、第 5 染色体断片よりも有意に短く、これは、第 13 染色体の断片サイズが非トリソミー 13 症例と比較して短くなかったことを意味した。したがって、すべての 4 つのトリソミー 13 症例は、z スコア法によって、4 つのうちの 2 つと比較して、本発明によって正確に同定された。

【 0 1 2 8 】

第 18 染色体断片のサイズの分析について、参照染色体は第 12 染色体であった。正倍数体又はトリソミー 13 胎児を持つ全ての症例について、第 18 染色体の断片は第 12 染色体胎児よりも有意に長かった。トリソミー 18 を有する 2 つの症例について、第 18 染色体断片のサイズは、第 12 染色体断片と有意に相違しなかった。これは、第 18 染色体の断片サイズは、非トリソミー 18 胎児を有する症例と比較した場合に短かった。換言すると、両者のトリソミー 18 症例は、正確に分類された。

【 0 1 2 9 】

X . 遺伝子疾患の多型の検出及び診断

また、母親血漿 DNA のサイズ分析は胎児遺伝子型の非侵襲性検出に使用可能である。胎児遺伝子型は、胎児が突然変異遺伝子を遺伝したかどうか、特定の対立遺伝子の不均衡を有するかどうか、又は他の配列不均衡若しくは目的を有するかどうかを決定するためには用いることができる。このような態様では、1 つの対立遺伝子が参照ゲノム位置（配列）であり得て、異なる対立遺伝子は試験中のゲノム位置であり得る。このようにして、参照配列を用いる方法のいずれかはまた、遺伝子型と他の配列不均衡の決定に適用され得る。

【 0 1 3 0 】

一態様では、配列不均衡（したがって、可能には遺伝子型）は、サイズ差異（不均衡）が母親試料（例えば、母親が対立遺伝子でヘテロ接合であるとき）中の対立遺伝子間に存在するか又は存在しないことによって決定され得る。例えば、試料中の対立遺伝子間にサイズプロフィールの差異がない場合、胎児は、母親と同じ遺伝子型を有することが決定される場合もある。別の例として、試料中の対立遺伝子間のサイズプロフィールに差異がある場合、胎児は、母親とは異なる遺伝子型を有することが決定され得る。

【 0 1 3 1 】

以下の例えは、母親は、特定の遺伝子座についてヘテロ接合（即ち、N M によって記される N 対立遺伝子の 1 つのコピーと M 対立遺伝子の 1 つのコピーを有する）である。文字

10

20

30

40

50

NとMは、通常、野生型（正常についてのN）と突然変異（突然変異についてのM）対立遺伝子をそれぞれ表す。しかしながら、N及びM対立遺伝子は、任意の2つの異なる対立遺伝子に対応することができるが、必ずしも野生型及び/又は突然変異である必要はない。一態様では、Mは危険性のあるゲノム配列であり、Nは参照配列であると考えることができる。これに関連して、遺伝子型の決定に参照配列を用いる上記方法の応用を理解することができる。

【0132】

非妊娠女性では、2つの対立遺伝子を担持する分子の平均サイズは同じであったであろう。しかしながら、妊娠女性の血漿には、母親と胎児のDNA分子の混合物が存在する。母親由来のDNA分子は、胎児由来のDNAよりも長い。母親と胎児の両方が2つの対立遺伝子（即ち、NとMの対立遺伝子）を有する場合、これらの対立遺伝子はともに、長い及び短いDNA分子の等しい寄与を有する。よって、得られたNとMの対立遺伝子のサイズ分布は同じであろう。反対に、母親と胎児遺伝子型が異なる場合、例えば、母親がNMであり、胎児がMMである場合、NとMの対立遺伝子のサイズ分布は異なるであろう。換言すると、血漿中の2つの対立遺伝子のサイズ分布は、胎児遺伝子型によって影響されたであろう。図25A～25Cは、本発明の態様に従って、妊娠女性と胎児のいでの異なるシナリオについての略図を示す。

【0133】

図25Aでは、胎児はNの遺伝子型を有し、母親遺伝子型はNMである。バー2510の長さは、それぞれ母親と胎児の2つの対立遺伝子の1つに由来する断片の平均（average）（平均（mean））サイズの指標である。上記されるように、胎児は、母親よりも小さな平均サイズを有する。したがって、長いバーは母親DNAを表し、短いバーは胎児DNAを表す。

【0134】

母親と胎児は対立遺伝子Nに寄与し、母親だけが対立遺伝子Mに寄与するので、対立遺伝子Nを持つ分子のサイズ分布は対立遺伝子Mを持つ分子よりも短かったであろう。換言すると、NMの遺伝子型を有する妊娠女性では、対立遺伝子Mと比較して対立遺伝子Nに対して短いサイズ分布は、胎児遺伝子型がNNであることを意味するであろう。よって、Nのサイズ分布（例えば、平均サイズ）がある種のカットオフ（例えば、割合又は絶対値）によってMよりも小さい場合、2つの野生型（N）を有することが同定し得る。

【0135】

図25Bでは、胎児はNMの遺伝子型である。母親と胎児はともに対立遺伝子NとMに寄与する。結果として、対立遺伝子MとNを有する分子についてのサイズ分布は同じである。NMの遺伝子型を有する妊娠女性では、対立遺伝子MとNについて同じサイズ分布がNMの胎児遺伝子型を指示したであろう。よって、Nのサイズ（例えば、平均サイズ）がある種のカットオフ（例えば、割合又は値）内でMにほぼ等しい場合、胎児は1つの野生型（N）対立遺伝子Mと1つの突然変異（M）を有することが同定し得る。

【0136】

図25Cでは、胎児は、MMの遺伝子型を有する。母親と胎児はともに対立遺伝子Mに寄与し、一方、母親だけが対立遺伝子Nに寄与するため、対立遺伝子Mを有する分子についてのサイズ分布は対立遺伝子Nを有する分子よりも短かったであろう。NMの遺伝子型を有する妊娠女性では、対立遺伝子Nよりも対立遺伝子Mについて短いサイズ分布は、胎児遺伝子型がMMであることを指示したであろう。よって、Mのサイズ分布（例えば、平均サイズ）はある種のカットオフ（例えば、割合又は値）により小さい場合、胎児は2つの突然変異（M）対立遺伝子を有することが同定し得る。

【0137】

また、本方法は、母親がホモ接合、例えば、NN又はMMである状況を分析するために用いることができる。胎児が異なる遺伝子型を有する場合、母親試料からのサイズ分布はまた変化し、よって、胎児遺伝子型を決定することができる。さらに、サイズ分布が変化しない場合、遺伝子型は、母親がヘテロ接合ケースである場合について上記した通り、母

10

20

30

40

50

親と同じであることが決定され得る。

【0138】

いくつかの態様では、不均衡などが存在するかどうかの決定は、母親遺伝子型と比較して（例えば、1つのMについて1つのNが不均衡でないことを意味する）、カットオフ値で行うことができる。例えば、母親遺伝子型からの十分に大きな逸脱（例えば、割合で）が存在する場合、胎児は、より小さなサイズ分布を有する対立遺伝子を遺伝したことが決定され得る。一態様では、カットオフ値は、母親試料中の胎児核酸の割合に依存し得る。胎児核酸のより高い割合が存在する場合、大きな逸脱が期待され、よって、より大きなカットオフを（例えば、別の対立遺伝子と比較して1つの対立遺伝子に関するサイズ分布の差異について）用いることができる。胎児核酸のより低い割合が存在する場合、より小さな逸脱が期待され、よって、より小さなカットオフを用いることができる。

10

【0139】

一態様では、父親の遺伝子型は、母親のどの対立遺伝子が胎児核酸により変化されるそのサイズ分布を有するかを決定するために用いられてもよい。胎児が遺伝子型についてホモ接合である場合において、これは、対立遺伝子が母親から遺伝されるちょうどそのものに、可能な胎児遺伝子型を狭くするのを可能にし得る。それは、父親からどれが来るかが知られているためである。このように狭めた後、胎児genomeの決定は、ほんの2つの可能性を試験するのに必要であるため、より正確になる場合がある。一実施において、同じカットオフ値は、父親からどの遺伝子型が遺伝されるかに関わらずに用いられる。別の実施では、異なるカットオフ値を用いることができる。

20

【0140】

種々の態様では、本明細書に言及されているサイズ分布のいずれかがこのような配列不均衡決定において用いられてもよい。いくつかの態様では、精度レベルがまた提供されてもよい。例えば、「未決定」の分類はまた、不均衡及び均衡分類の他に用いられてもよい。この方法では、いくつかの決定がより信頼度をもって決定され得て、一方、中間領域の値はさらにデータ点を要求してもよい。

【0141】

X I . 胎児ハプロタイプ遺伝のサイズ分析

サイズ分析の応用は、母親のどのハプロタイプが胎児に移動するかを決定するために、さらに拡張され得る。ハプロタイプは複数の遺伝子座での対立遺伝子を意味することができる。用語「ハプロタイプ」の定義は、本出願の定義節において見出すことができる。胎児ハプロタイプは、胎児が突然変異した遺伝子を遺伝したかどうか、特定の対立遺伝子の不均衡を有するかどうか、又は他の目的であるかどうかを決定するために用いることができる。よって、ハプロタイプは遺伝子型と同じように用いられてもよいが、より多くの遺伝子座が存在するため、より少量の血液試料が、胎児ハプロタイプの決定の同じか、又はさらには良好な統計的信頼度を達成するために用いられてもよい。一局面では、配列不均衡は、母親ハプロタイプと比較して決定され得る。本明細書では、ハプロタイプは、一連の多型、例えば、各々がゲノムにおける特定位置にあるSNPとして表すことができ、そこでは配列バリエーションが知られている。

30

【0142】

一態様では、母親血漿中の配列不均衡について分析することによって胎児のハプロタイプを決定する方法が提供される。一局面では、ハプロタイプ間のサイズプロフィールにおける分離値（例えば、差異）を用いて配列不均衡を決定する。一態様では、母親ハプロタイプ（例えば、Hap I 及び Hap II）は、分析（例えば、母親試料（例えば、胎児核酸を含まない試料）の配列決定（He D et al. Bioinformatics 2010; 26: i183 - i190）又は単一分子ハプロタイプング（Ding C et al. Proc Natl Acad Sci USA 2003; 100: 7449 - 7453、及び Xiao M et al. Nat Methods 2009; 6: 199 - 201））によって決定される。別の態様では、母親ハプロタイプ（単数又は複数）は、母親の親、兄弟姉妹、以前の子供、又は他の親戚の分析を用いて決定され

40

50

得る。なお別の態様では、強力な系統不均衡での多型について、1つの遺伝子座での母親遺伝子型を知ることは、例えば、対立遺伝子が同じ配列で、即ち、ハプロタイプで通常見られる場合、別の遺伝子座で遺伝子型を含むことができる。このようにして、母親ハプロタイプは、ゲノムの1つの測定から暗に決定され得る。1を超える遺伝子座での遺伝子型が決定され得て、この場合、各々の決定された遺伝子型は、他の遺伝子座で遺伝子型を暗示し得て、即ち、ハプロタイプが推定される。

【0143】

一態様では、父親の遺伝子型も決定される。この情報は、父親が特定のS N Pでホモ接合であるか又はヘテロ接合であるかを決定するために用いることができる。各遺伝子座での父親の対立遺伝子の直接的な決定がなされ得る。したがって、各父親の対立遺伝子がタイプ 又はタイプ のS N Pと称される母親のH a p I又はH a p I I上の対立遺伝子と同じであるかどうかを決定することができる。

10

【0144】

別の態様では、父親のハプロタイプも決定される。父親のハプロタイプは、親の試料の分析(例えば、配列決定(He D et al. Bioinformatics 2010; 26: i183 - i190)又は單一分子ハプロタイプング(Ding C et al. Proc Natl Acad Sci USA 2003; 100: 7449 - 7453、及びXiao M et al. Nat Methods 2009; 6: 199 - 201))によって、あるいは父親の親、兄弟姉妹、以前の子供、又は他の親戚の分析を通じて、決定され得る。あるいは、強力な系統不均衡な多型について、1つの遺伝子座での父親の遺伝子型を知ることは、例えば、対立遺伝子が同じ配列、即ち、ハプロタイプで通常見られる場合、別の遺伝子座で遺伝子型を含むことができる。したがって、父親のハプロタイプは、遺伝子型の1測定から暗に決定され得る。また、1を超える遺伝子座での遺伝子型が決定され得て、この場合、各々の決定された遺伝子型は他の遺伝子座での遺伝子座を暗に意味することができ、よって、ハプロタイプが推定される。

20

【0145】

このような態様の1つの応用は、配列不均衡(例えば、サイズプロフィールの差異を検出することによる)が、父親と母親がともにヘテロ接合である場合のS N Pを分析することによって、胎児核酸を含む母親試料に存在するか又は存在しないかを決定するためであり得る。それによって、胎児遺伝子型又はハプロタイプは、配列不均衡が存在するか又は存在しないかにより推定される。

30

【0146】

図26は、分析されるべきS N P遺伝子座で母親がヘテロ接合であり、父親がホモ接合である例を示す。サイズ分析を実施するために、母親がヘテロ接合であり、父親がホモ接合であるS N Pの小集団に焦点を当てることができる。母親の2つの相同的な染色体はそれぞれH a p IとH a p I Iと呼ばれる。これらのS N Pの各々について、本発明者らは、2つの母親対立遺伝子のうちのいずれの1つがH a p I上に位置するか、及びH a p I I上ではどうかを決定することができる。S N Pは、父親対立遺伝子がH a p I I上の母親対立遺伝子と同一である場合にタイプ 、父親対立遺伝子がH a p I I上の母親対立遺伝子に同一である場合にタイプ として定義される。胎児ゲノムの決定の更なる記載は、上述した出願「Fetal Genomic Analysis From A Maternal Biological Sample」において見出すことができる。

40

【0147】

母親ハプロタイプと父親遺伝子型又はハプロタイプが分かると、各S N Pに関連した断片のサイズ分布は、S N P小集団に配列不均衡が存在するかどうかを同定することによって父親ハプロタイプを決定するために用いることができる。代替の態様では、父親の遺伝子型又はハプロタイプが確実に分からぬ場合、それにも関わらずに、例えば、試験された集団の既知の遺伝子型又はハプロタイプの頻度に基づいて、可能な父親の遺伝子型又はハプロタイプを、例えば、統計的手法を用いて推定することができた。次に、配列不均衡(サイズ不均衡を介して決定される)は、例えば、X節で検討したように、各S N Pにつ

50

いて決定される。

【0148】

例えば、父親がHap I上の対立遺伝子（タイプ₁）についてホモ接合である場合、サイズ不均衡が存在しないか（胎児は母親からHap I Iを遺伝し、したがって、まさに母親のようにHap IとHap I Iについてヘテロ接合である）、又はHap Iが類似にサイズ分布を有するサイズ不均衡であろう（胎児が母親からHap Iを遺伝子、したがって、Hap Iについてホモ接合である）のいずれかである。父親がHap I I（後述するようにタイプ₂）についてホモ接合である場合、不均衡が存在しないか（胎児は母親からHap Iを遺伝し、したがって、まさに母親のようにHap IとHap I Iについてヘテロ接合である）、又はHap I Iが類似にサイズ分布を有する不均衡であろう（胎児が母親からHap I Iを遺伝子、したがって、Hap Iについてホモ接合である）のいずれかである。また、中間分類を用いてもよく、あるいは、均衡と不均衡との間の確実性のレベルが変更し得る。典型的には、類似のカットオフはいずれかの対応について用いられるであろう。種々の態様では、本明細書に言及されたサイズ分布のいずれかを用いてもよい。

10

【0149】

一態様では、複数のSNP局在の各々についての断片に関するサイズ分布は、胎児の2つのハプロタイプを決定するために分析され得る。例えば、1つのハプロタイプ上のSNP対立遺伝子について断片のサイズ分布間の相違は、他のハプロタイプ上のSNP対立遺伝子のサイズ分布と比較可能である。統計は、種々の方法、例えば、各SNPについて決定し、ハプロタイプで過半数（均衡、不均衡、及び中間を含んでもよい）を採用することによって分析され得る。別の例として、サイズは、SNP全体で（例えば、カットオフ値と比較される平均又は中央値のサイズ分布を得るために）集められ得る。あるいは、この2つの組み合わせを用いることができる。別の実施は、データ点の極値、例えば、特定のSNPについての最小差を用いることであり得た。

20

【0150】

図27は、父親ハプロタイプが図26に示されるものである場合、母親からのHap Iを胎児が遺伝する例を示す。タイプ₁のSNP（斜線なしのボックスによって囲まれている）について、胎児は、父親から母親のHap I上に位置されたものと同一である対立遺伝子を遺伝したであろう。よって、胎児はHap I腕の対立遺伝子についてホモ接合であったろう。結果として、Hap I上の対立遺伝子のサイズ分布は、母親試料中のHap I I上の対立遺伝子よりも短かったであろう。タイプ₂のSNP（斜線ボックスによって囲まれている）について、胎児は父親から母親Hap I I上のものと同一である対立遺伝子を遺伝したであろう。よって、胎児はヘテロ接合であったろう。結果として、Hap IとHap I I上の対立遺伝子のサイズ分布は母親試料において同じサイズ分布を有したであろう。

30

【0151】

一態様では、同じタイプ（タイプ₁又はタイプ₂）のSNPをともに分析することができる。タイプ₁のSNPについて、Hap I上の対立遺伝子を担持する分子のサイズ分布は、Hap I I上の対立遺伝子を担持する分子のサイズ分布よりも短いであろう。タイプ₂のSNPについて、Hap I及びHap I I上の対立遺伝子を担持する分子のサイズ分布は同じであろう。換言すると、Hap Iを担持する分子のサイズ分布はHap I Iを担持する分子よりも小さい場合、胎児はHap Iについてホモ接合である。Hap IとHap I Iを担持する分子のサイズ分布が同じである場合、胎児はヘテロ接合である。

40

【実施例】

【0152】

下記の実験は、胎児ハプロタイプ遺伝分析の精度を試験するために使用された。サラセミアの出生前診断について産科クリニックに来ているカップルを採用した。血液試料を父親と母親から採取した。母親については、血液試料は、妊娠12週で絨毛膜標本サンプリング（CVS）前に採取された。CVS後、一部を実験のために保存した。

【0153】

50

D N A は、父親と母親の軟膜、C V S 試料を抽出した。これらのD N A 試料は、父親、母親及び胎児遺伝子型を決定するために、Affymetrix Genome-Wide Human SNP Array 6.0 システムによって分析に供された。この実験では、C V S データは、母親ハプロタイプを推定するために用いられた。しかしながら、試験の臨床的実施において、母親ハプロタイプは、上記されるように他の手段によって推定され得た。また、C V S データは、本明細書中の方法を用いた決定の精度を確認するために用いられ、これはC B S を必要としない。

【0154】

この例証では、本発明者らは、母親がヘテロ接合であり、父親がホモ接合であるS N P の情報を与える小集団に焦点を合わせる。この小集団のS N P では、カップルと胎児遺伝子型は、母親ハプロタイプを構築するために用いられる。本発明者らは、母親が胎児に移動した一連の対立遺伝子をハプロタイプI (H a p I) と定義し、一方、胎児が母親から受領しなかった一連の対立遺伝子をハプロタイプI I (H a p I I) と定義する。

【0155】

次に、本発明者らは、情報を与えるS N P を2つの小集団、即ち、タイプ I とタイプ II に分けた。タイプ I のS N P について、父親の対立遺伝子はH a p I 上の母親の対立遺伝子と同一であった。これらのS N P について、胎児は、親から同じ対立遺伝子 (H a p I 上の対立遺伝子) を遺伝したであろうし、即ち、胎児は、H a p I 上にS N P についてホモ接合であったろう。タイプ II のS N P について、父親の対立遺伝子はH a p I I 上の母親の対立遺伝子と同一であった。これらのS N P について、胎児は、母親からH a p I 上の対立遺伝子と、父親から異なる対立遺伝子 (H a p I I 上のものと同一) とを遺伝したであろうし、胎児はヘテロ接合であったろう。

【0156】

各配列の2つの末端は、ShenzhenのBeijing Genomics Institute からのSOAP2 プログラムを用いて非繰り返しマスクされたヒトゲノム (UCSC genome . ucsc . edu からダウンロードされるHg18 NCB I . 26 ダウンロードされる) に対して整列された (soap . genomics . org . cn /) (Liら . Bioinformatics 2009, 25 (15) : 1966 - 7)。

【0157】

一態様では、ハプロタイプについての短い断片によって寄与された全長の画分の統計値は、どの母親ハプロタイプが胎児に移動したかを決定するために使用された。例として、第22染色体は、どの母親ハプロタイプが胎児に移動するかを推定するために、サイズ分析をどのように使い得るかを例証するために用いられる。第1に、本発明者らは、第22染色体をいくつかのセグメントに分け、各々のセグメントは、50の情報を与えるS N P からなり、そこでは、母親はヘテロ接合であり、父親はホモ接合であった (第22染色体とセグメントはともに配列の例である)。各セグメントについて、これらの情報を与えるS N P をカバーするD N A 断片 (配列の部分である分子の例) は、2つの群、即ち、H a p I とH a p I I に分けられ、これらの断片は2つの母親ハプロタイプのうちのいずれかに寄与した。各セグメントについて、母親H a p I と母親H a p I I に対してマッピングしている全ての断片の全長の合計が決定された。次に、同時に、各セグメントにおける母親H a p I とH a p I I に対してマッピングしている短い断片の全長の合計が決定された。例証の目的で、150bp又はそれより短い断片は、短い断片の全長の計算について、この例において短いものとして定義される。これらの長さから、短い断片によって寄与された全長の断片は、各セグメント内でH a p I とH a p I I に対してマッピングしているD N A 断片について計算され得る。

【0158】

図28は、本発明の態様による第22染色体上のタイプ I のS N P についてのサイズ分析を例証する表を示す。F (H a p I - H a p I I) は、H a p I とH a p I I との間の短い断片によって寄与された全長の画分の差を表す。ほんの28のS N P からなる最後

10

20

30

40

50

のセグメントを除いた場合、 $F_{(HapI - HapII)}$ は 0.0288 ~ 0.0701 の範囲であった。 $F_{(HapI - HapII)}$ のポジティブ値は、HapI の F 値が HapII よりも通常大きいことを表す。F 値が短い断片によって寄与された全長の画分として定義されるので、より高い F 値は、全長のより高い画分が短い断片によって寄与されることを指示する。換言すると、これらの結果は、HapI 上に対立遺伝子を担持する DNA 断片が、分析される 50 の SNP からなる領域の各々について HapII 上の対立遺伝子を担持する DNA 断片よりも短いことを指示する。これは、HapI のサイズ分析が HapII よりも短いことを指示する。したがって、本発明者らは、胎児が HapI 上で対立遺伝子についてホモ接合であると推定できる。上記の通り、一態様では、分類は、HapI と HapII における 2 つのゲノム位置間の分離値（例えば、F）に基づいて、各セグメントについて行うことができ、次に、全分類は各セグメント分類に基づいて行うことができる。別の態様では、全分離値（例えば、平均分離値）は、各セグメントについての分離値から決定することができ、この全統計値は分類を決定するために用いることができる。
10

【0159】

一態様では、約 0.02 のカットオフは、不均衡が存在するかどうかを決定するために用いてもよい。F の中央値又は平均を用いた場合、より大きなカットオフを使用することができ、それに正確であった。また、カットオフは、中間結果に用いてもよく、例えば、0.015 と 0.025 との間の領域が中間であってもよく、更なる分析を必要としてもよい。
20

【0160】

図 29 は、本発明の態様による第 22 染色体上のタイプ の SNP についてのサイズ分析を例証する表を示す。 $F_{(HapI - HapII)}$ は、HapI と HapII についての短い断片によって寄与された全長の画分における差を表す。 $F_{(HapI - HapII)}$ 値は、-0.0203 ~ 0.0207 の範囲であり、中央値は 0.003 であった。小さい $F_{(HapI - HapII)}$ 値は、HapI と HapII に対してマッピングしている断片に対して同じサイズ分布と匹敵する。よって、本発明者らは、HapI と HapII について胎児がヘテロ接合であると推定できる。タイプ の SNP は、父親の対立遺伝子が HapII 上の母親対立遺伝子と同一である SNP として定義されるので、この結果は、胎児が母親から HapI を遺伝したことを暗に意味する。
30

【0161】

図 30 は、本発明の態様による第 22 染色体上のタイプ とタイプ の SNP についての $F_{(HapI - HapII)}$ のプロットを示す。タイプ の SNP について、HapI 断片のサイズ分布は HapII 断片よりも短く、それは、ゼロより大きい $F_{(HapI - HapII)}$ の値をもたらす。タイプ の SNP について、HapI と HapII 断片のサイズ分布間の差がない；したがって、 $F_{(HapI - HapII)}$ 値は 0 周辺に集まる。カットオフ 0.025 を用いて、 $F_{(HapI - HapII)}$ 分析は、13 のタイプ のセグメントと 21 のタイプ のセグメントの全てについて、胎児による HapI の遺伝を正確に推定することができる。
40

【0162】

XII. 標的化配列の使用の例

以下の例は、本発明の態様のサイズに基づく診断アプローチが標的化された配列決定フォーマットに有用であり得ることを示す。このようなフォーマットは、診断対象のゲノム領域が配列決定について特異的に標的化される。このフォーマットは、ランダム配列決定に関わる状況とは対照的に、対象とする領域に焦点を併せるという利点を有し、この場合、配列決定パワーのいくつかは、診断応用と即座に関連しない領域について用いられる。よって、標的化された配列決定フォーマットは、スループットを増加させ、システムコストを削減することが期待され得る。標的化された配列決定は、当業者に知られている任意のフォーマットを用いて行うことができ、例えば、液相捕捉システム（例えば、Agilent SureSelect システム）、固相捕捉システム（例えば、Roche N 50

imble Genシステム)又は標的特異的増幅(例えば、RainDanceシステム)が挙げられる。

【0163】

血液試料は妊娠第1期中の8人の妊娠女性から回収した。血漿3.2mLからDNAをDSP DNA Blood Mini Kit (Qiagen)によって各ケースについて抽出された。絨毛膜試料(母親試料が採取された後に回収される)について行った染色体分析は、4人の胎児がT21 (UK229、UK510、UK807、PW421)を有し、他の4人が正倍数体男子(PW226、PW263、PW316、PW370)であることを示した。

【0164】

各ケースについて5~30ngの血漿DNAは、Chromatin Immunoprecipitation Sequencing試料調製の製造業者の使用説明書に従って、対末端試料調製キット(Illumina)によってDNAライブラリー構築のために用いられた。アダプター-リガンド化されたDNAは、さらにサイズ選択することなしに、QIAquick PCR精製キット(Qiagen)において与えられるスピンカラムを用いて直接精製された。次に、アダプター-リガンド化されたDNAは、標準プライマーを用いて15サイクルのPCRにより増幅された。プライマーは、IlluminaのPCR Primer PE 1.0と2.0であった。DNAライブラリーは、NanoDrop ND-1000分光光度計(NanoDrop Technologies)を用いて定量され、DNA 1000キット(Agilent)を用いて2100 Bioanalyzer上で稼働され、サイズ分布を調べた。0.6~1μgの増幅血漿DNAライブラリーは、約290bpの平均サイズで各試料について生成された。

10

【0165】

SureSelect Human ALL Exon捕捉ライブラリーは、Agilentから得られ、ヒトエクソンの37.8Mbをカバーした(カタログ番号:5190-2310)。本研究の8つ全てのケースについて、各ケースの500ngの増幅血漿DNAライブラリーを捕捉プローブとともに24時間65にて、製造業者の使用説明書に従ってインキュベートした。ハイブリダイゼーション後、捕捉された標的は、ストレプトアビジン被覆された磁気ビーズ(Dynal DynaMag-2 Invitrogen)を用いることによって、ビオチン化されたプローブ/標的ハイブリダイズを引くことによって選択され、MinElute PCR精製キット(Qiagen)を用いて精製された。最後に、標的化されたDNAライブラリーは、AgilentからのSureSelect GA PEプライマーを用いた12サイクルのPCR増幅によって濃縮された。PCR産物は、QIAquick PCR精製キット(Qiagen)によって精製された。

20

【0166】

標的富化の有無による8対のライブラリーは、2つのフローセルの16レーンに装填され、次に、36bp×2の対末端フォーマットを用いたGenome Analyzer IIx (Illumina)によって配列決定された。全36bpの配列決定された読み出しが、Short Oligonucleotide Alignment Program 2 (http://genome.ucsc.edu)を用いて、マスクされていないヒト参照ゲノム(Hg18)に対して整列された。対末端読み出しの断片サイズは、40bp~600bpの範囲として画定された。各配列決定されたDNA断片のサイズは、両末端で最外部のヌクレオチドの座標から推定された。

30

【0167】

胎児トリソミー21の親検出のためのサイズ分析

この実施例では、短い断片に寄与された全長の画分は、第21染色体と参照染色体について計算され、それぞれ F_{21} と F_{ref} で表される。参照染色体は、第13、第18及び第21染色体を除く全ての常染色体からなっていた。全長は、600bp以下のDNA断片の全てについて長さの合計によって計算された。第21染色体と参照染色体間の未時

40

50

間断片によって寄与された長さの画分の差 (F) は $F_{21} - F_{ref}$ として計算された。

【 0 1 6 8 】

図 3 1 A は、本発明の態様による、標的富化なしでの血漿 D N A のサイズ分析を与える表である。種々のカラムは、第 21 染色体について 150 b p 未満の断片についての全長、第 21 染色体について 600 b p 未満の断片についての全長、及び F_{21} として 2 つの比を与える。他のカラムは、参照染色体についての 150 b p 未満の断片についての全長、参照染色体についての 600 b p 未満の断片についての全長、及び F_{ref} として 2 つの比を与える。最後のカラムは、2 つの画分間の差 F である。

【 0 1 6 9 】

図 3 1 B は、本発明の態様による、標的富化を有する血漿 D N A のサイズ分析を与える表である。図 3 1 B のカラムは、図 3 1 B の表と同じデータフォーマットを有する。

【 0 1 7 0 】

図 3 2 は、標的富化の有無による T 21 と正倍数体についての F のプロットである。標的富化なしの試料について、 F についてカットオフ値 0.005 を用いて、T 21 と正倍数体妊娠からの血漿試料は、100 % 精度で分別され得る。標的富化を有する試料について、 F についてカットオフ値を用いて、T 21 と正倍数体妊娠からの血漿試料は 100 % 精度で分別され得る。この実施例は、サイズに基づく分析が標的化された配列決定を用いて行うことができる事を示す。T 21 の検出について、第 21 染色体と参照染色体についての標的化された配列決定を用いることは有利であり得て、その結果、50 % の配列決定は前者に指向され、残りは後者に指向される。このような設計は、T 21 検出に即座に関連しない領域について配列決定パワーの浪費を削減し得る。このような設計は、多重配列決定を用いて（例えば、インデックス又はバーコードによる配列決定を用いることによって）、多数の患者からの試料を配列決定することができる。

【 0 1 7 1 】

本明細書において記載されている任意のソフトウェア・コンポーネント又はソフトウェア関数は、任意の適切なコンピュータ言語、例えば Java、C++ 又は Perl を使用して、例えば従来の又はオブジェクト指向の技術を使用するプロセッサによって実行されるソフトウェアコードとして実行され得る。ソフトウェアコードは、記憶及び / 又は送信のためのコンピュータ読み取り可能媒体、ランダムアクセスメモリー (R A M) 、読み取り専用メモリー (R O M) 、磁気媒体、例えばハードドライブ若しくはフロッピーディスク、又は光学媒体、例えばコンパクトディスク (C D) 若しくは D V D (デジタル多用途ディスク) 、フラッシュメモリーなどを含む好ましい媒体における、一連の指令又はコマンドとして蓄積され得る。コンピュータ読み取り可能媒体は、かかる記憶又は送信装置の任意の組み合わせであり得る。

【 0 1 7 2 】

また、このようなプログラムは、コード化され、インターネットを含む種々のプロトコールに適合する有線、光、及び / 又は無線ネットワークを経由した送信のために適用されるキャリアー信号を使用して送信され得る。したがって、本発明の態様のコンピュータ読み取り可能媒体は、このようなプログラムでコード化されたデータ信号を使用して作られ得る。プログラムコードでコード化されたコンピュータ読み取り可能媒体は、互換性のあるデバイスとともにパッケージされるか、又は他のデバイスからは分離して（例えばインターネットダウンロードを通じて）提供され得る。任意のこのようなコンピュータ読み取り可能媒体は、単一のコンピュータプログラム製品（例えばハードドライブ若しくは全コンピュータシステム）上に又はその中に装備されてもよく、システム若しくはネットワーク内における異なるコンピュータプログラム製品上若しくはその中に存在してもよい。コンピュータシステムは、ユーザーに本明細書において言及された任意の結果を提供するためのモニター、プリンター、又は他の好ましいディスプレイを含んでもよい。

【 0 1 7 3 】

コンピュータシステムの例を図 3 3 に示す。図 3 3 に示されたサブシステムは、シス

10

20

30

40

50

ムバス 3375 を通じて相互接続される。更なるサブシステム、例えばプリンター 3374、キーボード 3378、固定ディスク 3379、ディスプレイヤダプター 3382 と連結されたモニター 3376、及びその他のものが示される。I/O コントローラー 3371 と連結する、周辺機器及び入力 / 出力 (I/O) 装置は、当技術分野において既知である任意の数の手段、例えばシリアルポート 3377 によってコンピュータシステムと接続され得る。例えば、シリアルポート 3377 又は外部インターフェース 3381 は、コンピュータ装置をインターネットのような広域ネットワーク、マウス入力装置、又はスキャナーと接続するために使用され得る。システムバス経由の相互接続により、セントラルブロセッサ 3373 を個々のサブシステムと通信させ、システムメモリー 4372 又は固定ディスク 3379 からの指示の実行、及びサブシステム間の情報の交換をさせることができる。システムメモリー 3372 及び / 又は固定ディスク 3379 は、コンピュータ読み取り可能媒体を具現化し得る。本明細書に言及されている値のいずれかは、1 つのコンポーネントから別のコンポーネントに出力することができ、ユーザーに出力することができる。10

【0174】

コンピュータシステムは、例えば、外部インターフェース 3381 又は内部インターフェースによってともに連結される、複数の同じコンポーネント又はサブシステムを含むことができる。いくつかの態様では、コンピュータシステム、サブシステム、又は装置はネットワークを介して通信され得る。このような例において、1 つのコンピュータがクライアントであり、別のコンピュータがサーバーであると考えることができ、その場合、各々は同じコンピュータシステムの一部であり得る。クライアント及びサーバーは、各々、多数のシステム、サブシステム、又はコンポーネントを含むことができる。20

【0175】

特定の態様の特定の詳細は、本発明の精神及び範囲から逸脱することなし、任意の適切な方法で組み合わせてもよい。しかしながら、本発明の他の態様は、各個々の側面、又はこれらの個々の側面の特定の組み合わせに関連する態様に指向されてもよい。

【0176】

本発明の例示的態様の上記記載は、例証及び説明の目的のために存在する。記載された明確な形態に本発明を網羅する、又は限定することは意図されておらず、そして多くの変更及び変形は、上記教示の観点において可能である。該態様は、本発明の原理、並びにそれにより他の当業者が本発明を種々の実施形態で本発明を最もよく利用することができるよう、意図された特定の使用に好適である種々の変形を有する、その実用的な利用を最もよく説明するために選択かつ記載された。30

【0177】

本明細書において引用された全ての出願、特許及び特許出願は、全ての目的のために、それら全体で参照により本明細書に援用される。

【図1】

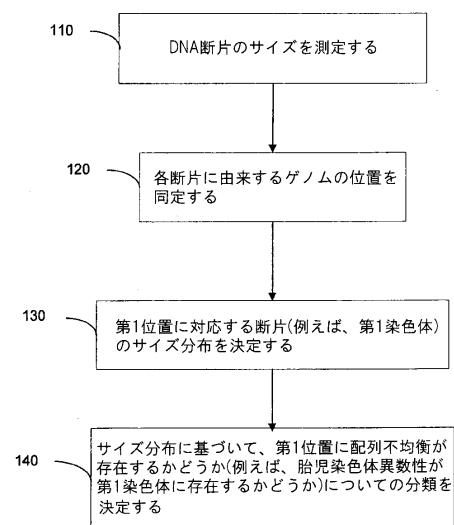


FIG. 1

【図2】

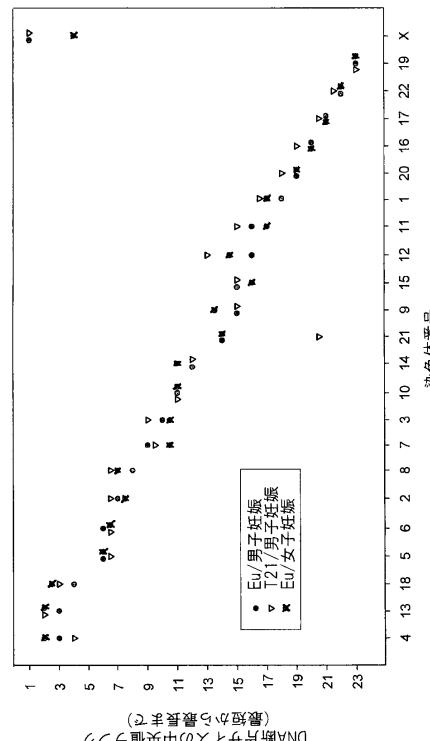


FIG. 2

【図3】

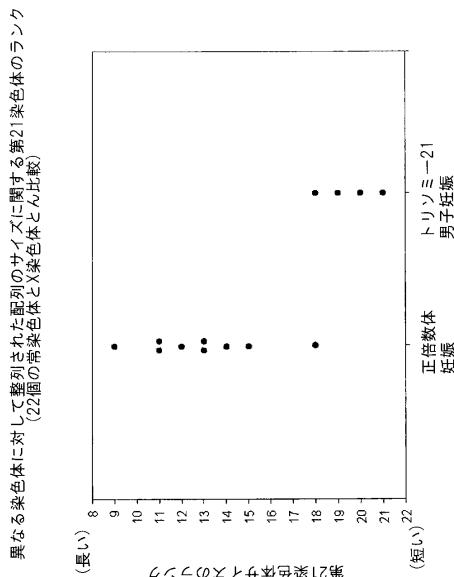


FIG. 3

【図4】

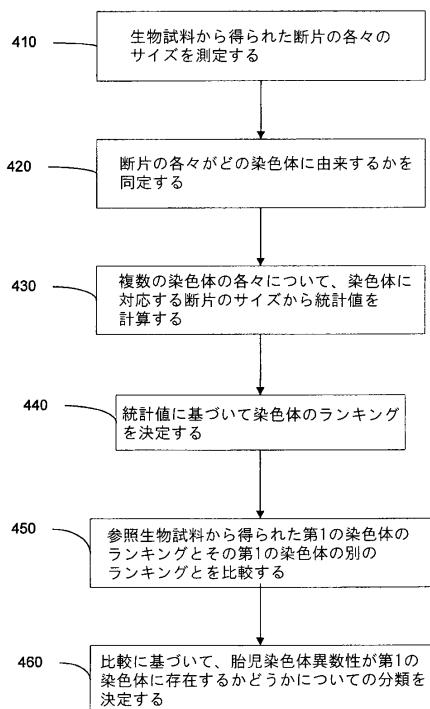


FIG. 4

【図5】

試料名	妊娠期	胎児 核型	第7染色体 断片の平均サイズ		第21染色体 断片の平均サイズ		第7と第21 染色体の差 平均間の差		P値* (第7 vs 第2染色体)		P値* (第14 vs 第21染色体)
			第7染色体 断片の平均サイズ	第21染色体 断片の平均サイズ	第7染色体 断片の平均サイズ	第21染色体 断片の平均サイズ	第7と第21 染色体の差 平均間の差	第7と第21 染色体の差 平均間の差	第7染色体 断片の平均サイズ	第21染色体 断片の平均サイズ	
MA800-男子	3rd	正倍数体	156.2	151.0	0.8	0.08	156.1	-0.1	0.822	0.822	
MA801-男子	3rd	正倍数体	151.2	151.7	0.5	0.085	151.5	0.3	0.564	0.564	
MA814-女子	3rd	正倍数体	157.8	157.3	-0.5	0.106	156.8	-0.9	0.022	0.022	
PW005-Eu-男子	1st	正倍数体	145.0	146.1	0.6	0.527	145.9	-0.1	0.988	0.988	
PW007-Eu-男子	1st	正倍数体	152.1	152.7	0.6	0.033	152.3	0.2	0.679	0.679	
PW008-Eu-男子	1st	正倍数体	147.2	147.3	0.1	0.538	147.0	-0.2	0.688	0.688	
PW012-Eu-男子	1st	正倍数体	146.7	146.8	0.1	0.333	146.7	0.1	0.68	0.68	
PW020-Eu-男子	1st	正倍数体	139.7	140.0	0.3	0.174	139.7	0.0	0.741	0.741	
PW009-Eu-女子	1st	正倍数体	151.9	152.2	0.3	0.068	152.1	0.2	0.405	0.405	
PW010-Eu-女子	1st	正倍数体	148.3	148.5	0.2	0.101	148.4	0.1	0.285	0.285	
PW016-Eu-女子	1st	正倍数体	141.5	141.1	-0.4	0.782	141.2	-0.3	0.842	0.842	
PW022-Eu-女子	1st	正倍数体	140.8	140.7	-0.1	0.362	140.5	-0.4	0.84	0.84	
M2B49-T21-男子	1st	T21	145.9	147.0	1.1	<0.001	146.8	0.8	<0.001	<0.001	
MA396-T21-男子	1st	T21	144.3	146.3	2.0	<0.001	145.9	1.6	<0.001	<0.001	
MA446-T21-男子	1st	T21	141.0	142.7	1.7	<0.001	142.3	1.3	<0.001	<0.001	
MA620-T21-男子	1st	T21	148.7	151.3	1.6	<0.001	150.9	1.2	<0.001	<0.001	

FIG. 5

【図7】

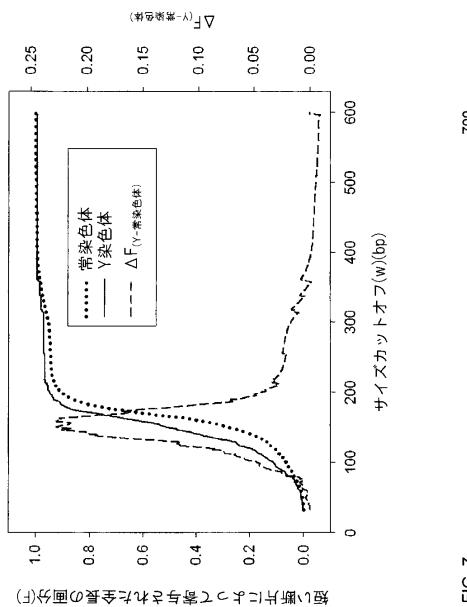


FIG. 7

【図6】

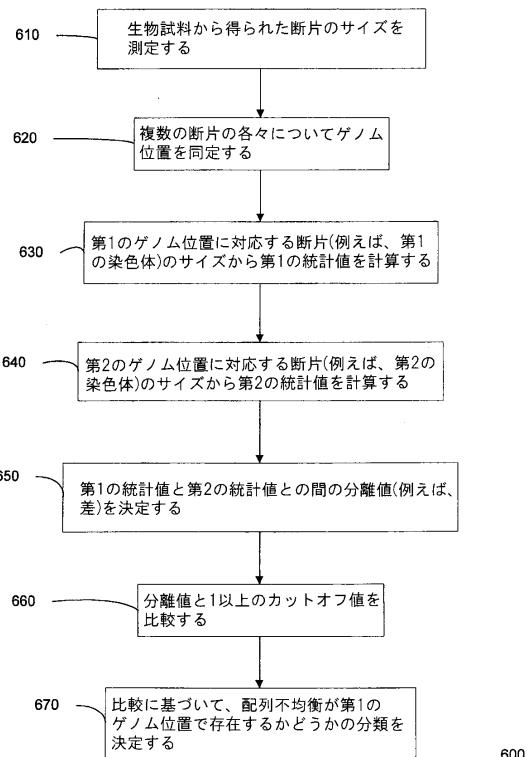


FIG. 6

【図8】

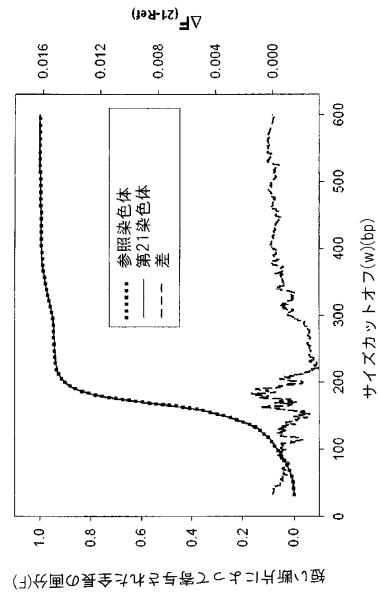


FIG. 8

【図9】

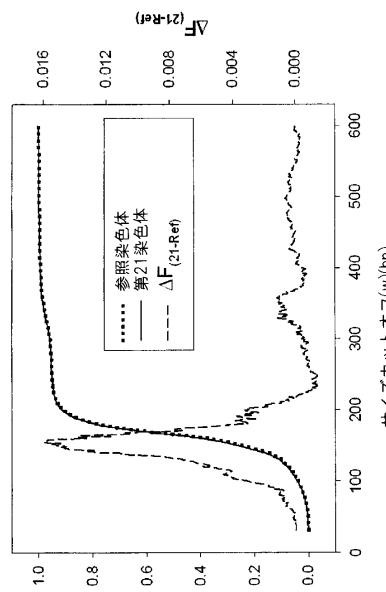


FIG. 9

【図10】

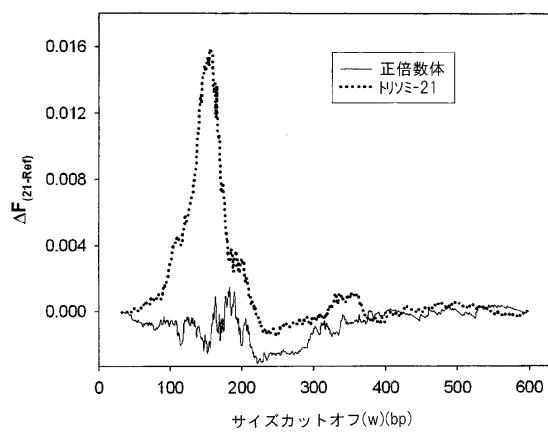


FIG. 10

1000

【図11】

	症例番号		
	女子	男子	合計
トリソミー13(T13)	9	14	23
トリソミー18(T18)	20	10	30
トリソミー21(T21)	4	5	9
正倍数体	23	35	58
合計	56	64	120

FIG. 11

1100

【図12】

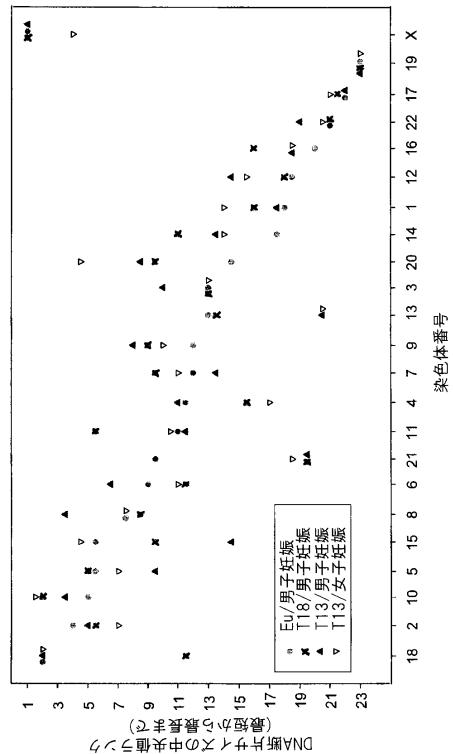


FIG. 12

【図 1 3】

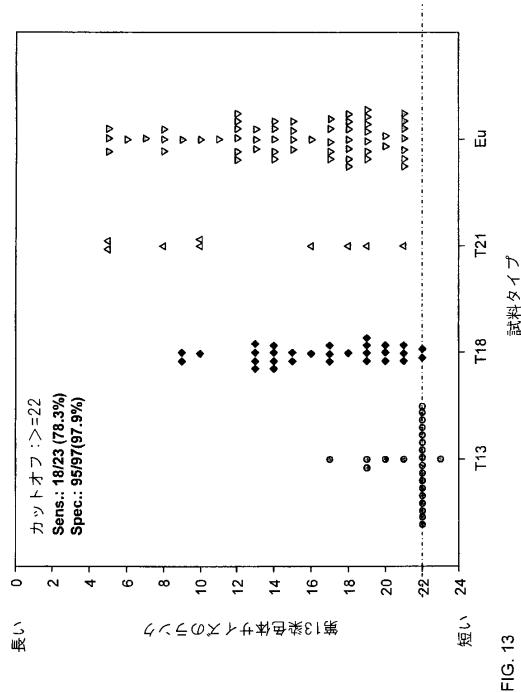


FIG. 13

【図 1 4】

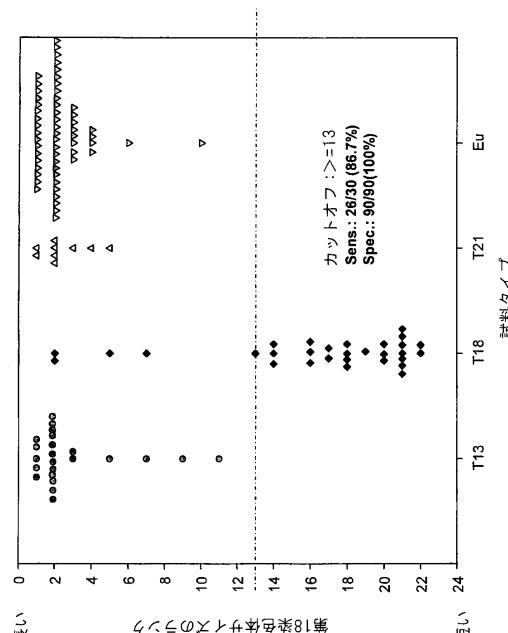


FIG. 14

【図 1 5】

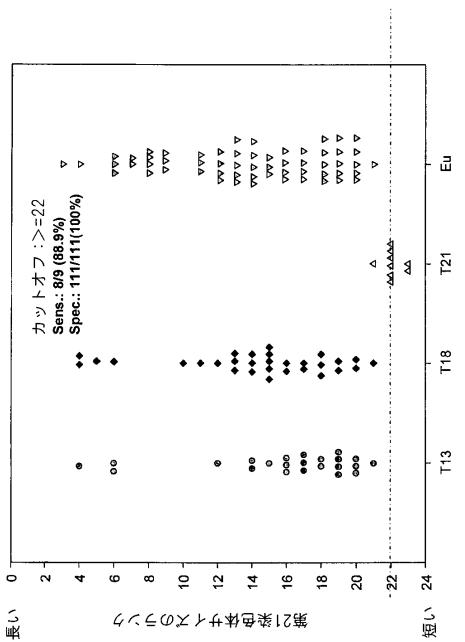


FIG. 15

【図 1 6】

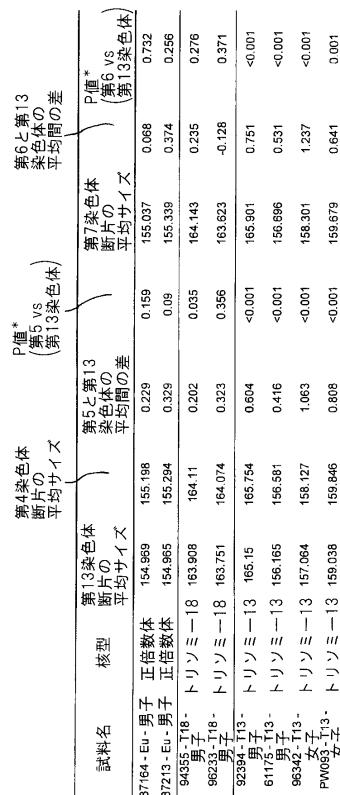


FIG. 16

【図17】

試料名	核型	第18染色体 断片の平均サイズ	第14染色体 断片の平均サイズ	P値 (第18 vs 第14染色体)
87184- Eu- 男子	正倍数体	155.595	154.614	<0.001
87213- Eu- 男子	正倍数体	155.977	155.142	<0.001
94355- T18- 男子	トリソミー-18	163.637	163.914	0.835
96233- T18- 男子	トリソミー-18	163.897	164.378	0.461
92394- T13- 男子	トリソミー-13	166.163	165.715	0.448
61175- T13- 男子	トリソミー-13	157.186	156.606	<0.001
96342- T13- 女子	トリソミー-13	158.326	158.013	0.313
99093- T13- 女子	トリソミー-13	160.469	159.503	0.966

FIG. 17

【図19】

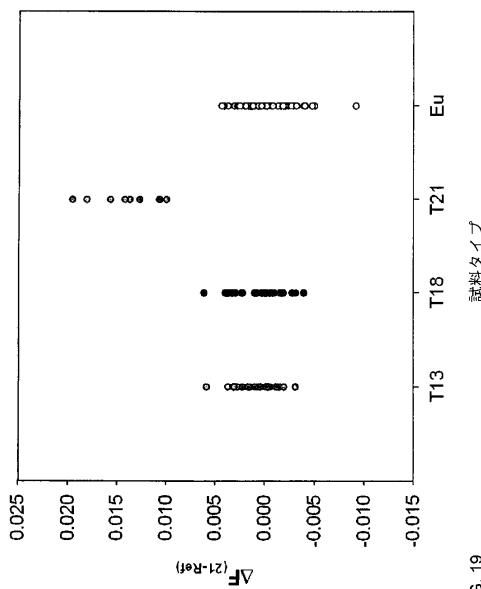


FIG. 19

【図18】

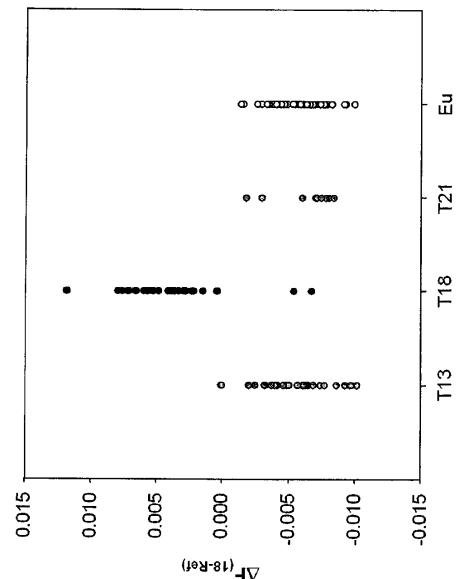


FIG. 18

【図20】

Chr.	GC含量(%)
4	36.53
13	37.10
6	38.60
5	38.61
3	38.79
X	38.96
18	39.07
8	39.45
2	39.67
12	39.80
7	39.94
21	40.00
14	40.33
9	41.03
10	41.41
1	41.52
11	41.89
15	42.35
20	44.65
16	45.22
17	46.05
22	49.89
19	50.65

FIG. 20

【図 2 1】

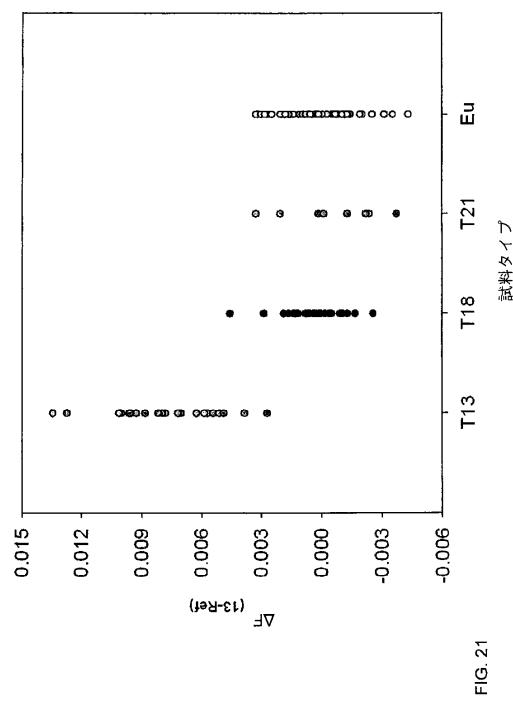


FIG. 21

【図 2 2】

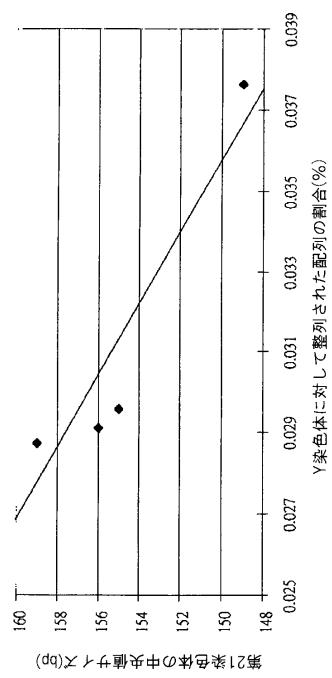


FIG. 22

【図 2 3 A】

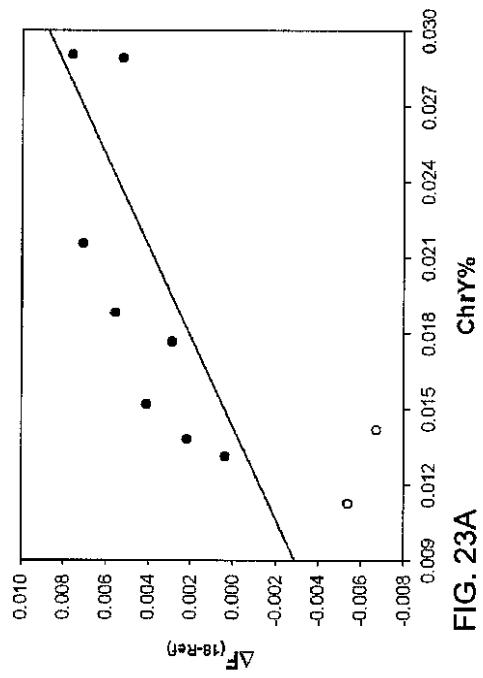


FIG. 23A

【図 2 3 B】

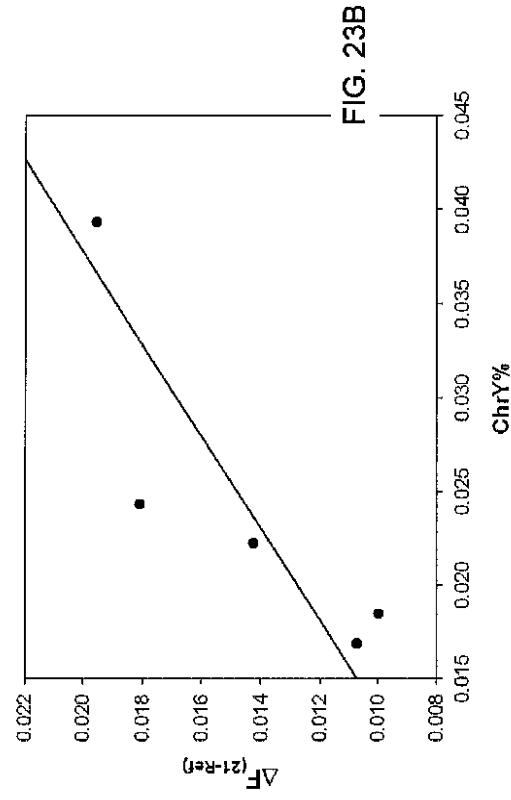
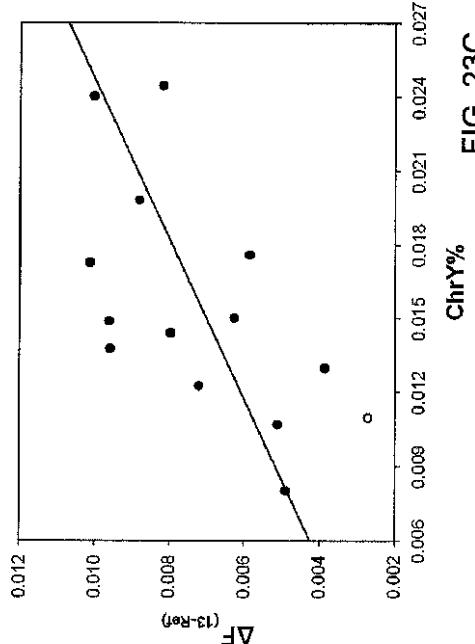


FIG. 23B

【図 23C】

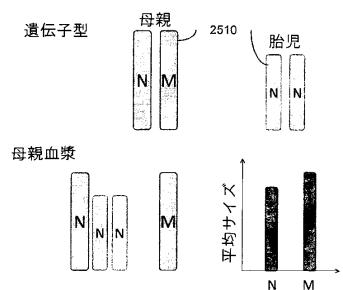


【図 24】

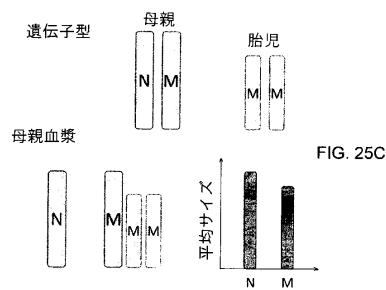
試料ID	核型	検出された第18染色体のDNA断片の有意な短縮		検出された第13染色体のDNA断片の有意な短縮	
		第18染色体についてのZスコア	第13染色体についてのZスコア		
1	正倍数体	1.07	0.84	no	no
2	正倍数体	-0.64	0.08	no	no
3	トリソミー-18	3.62	-1.43	yes	no
4	トリソミー-18	6.07	0.51	yes	no
5	トリソミー-13	0.61	3.55	no	yes
6	トリソミー-13	-1.04	2.06	no	yes
7	トリソミー-13	2.67	5.72	no	yes
8	トリソミー-13	-1.39	1.55	no	yes

FIG. 24

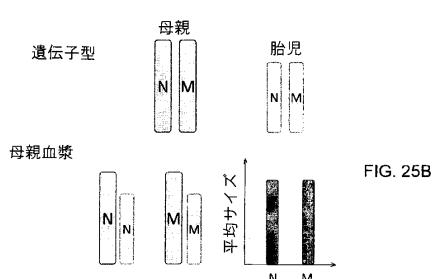
【図 25A】



【図 25C】



【図 25B】



【図26】

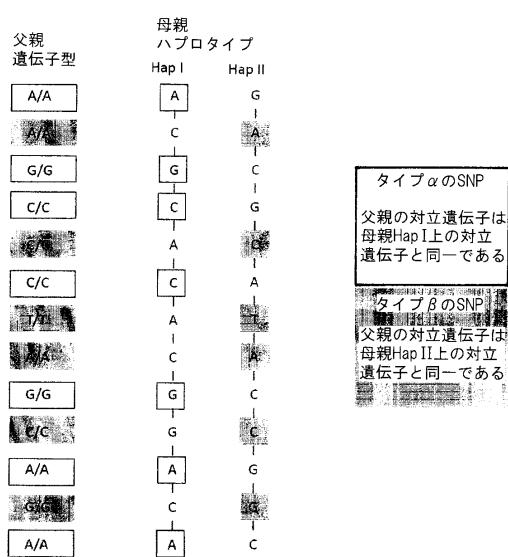


FIG. 26

【図27】

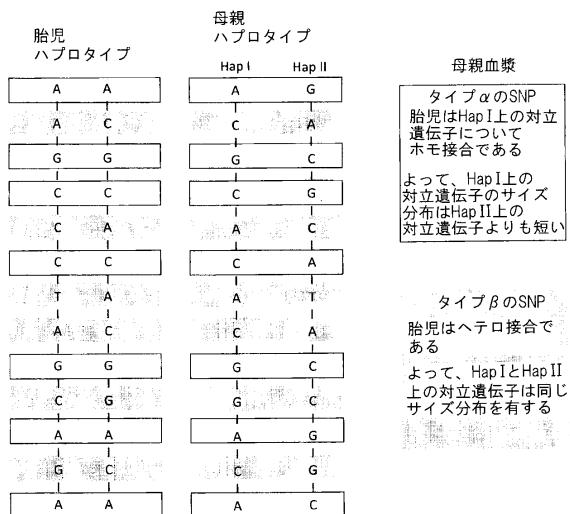


FIG. 27

【図28】

セグメント内 のSNP数	開始SNP	終了SNP	タイプαのSNP		$\Delta F_{(Hap\ I - Hap\ II)}$
			Hap I	Hap II	
50	rs2027649	rs5993883	0.2595	0.2096	0.0499
50	rs2239395	rs1002286	0.2648	0.2219	0.0429
50	rs4822458	rs3747134	0.2905	0.2204	0.0701
50	rs5761557	rs4410	0.2659	0.2037	0.0623
50	rs134784	rs4822998	0.2569	0.2221	0.0348
50	rs5762936	rs5998473	0.2882	0.2302	0.0580
50	rs5734086	rs8140669	0.2599	0.2311	0.0288
50	rs5999854	rs8756540	0.2659	0.2249	0.0411
50	rs229535	rs2413637	0.2953	0.2425	0.0528
50	rs4820431	rs11090087	0.2793	0.2199	0.0594
50	rs1023469	rs1972489	0.2884	0.2438	0.0445
50	rs5764858	rs5769218	0.2644	0.2001	0.0643
50	rs6009043	rs5769452	0.2481	0.2118	0.0363
28	rs17825762	rs131815	0.2648	0.2407	0.0241

FIG. 28

【図29】

セグメント内 のSNP数	開始SNP	終了SNP	タイプβのSNP		$\Delta F_{(Hap\ I - Hap\ II)}$
			Hap I	Hap II	
50	rs2159071	rs873387	0.2565	0.2414	0.0151
50	rs11917	rs78908	0.2548	0.2577	-0.0029
50	rs589089	rs7288450	0.2495	0.2589	-0.0093
50	rs5759868	rs2301492	0.2431	0.2477	-0.0046
50	rs2301497	rs713911	0.2411	0.2204	0.0207
50	rs4820740	rs5752850	0.2181	0.2195	-0.0014
50	rs5752851	rs5752964	0.2344	0.2372	-0.0028
50	rs1894473	rs135472	0.2603	0.2510	0.0093
50	rs135475	rs5754558	0.2324	0.2457	-0.0133
50	rs4821148	rs11705488	0.2381	0.2380	0.0001
50	rs490647	rs362246	0.2270	0.2224	0.0046
50	rs362214	rs6000130	0.2404	0.2401	0.0003
50	rs739206	rs12628179	0.2341	0.2193	0.0148
50	rs17298479	rs2143921	0.2597	0.2643	-0.0047
50	rs5758913	rs926542	0.2655	0.2582	0.0073
50	rs7291629	rs133761	0.2156	0.2360	-0.0203
50	rs133755	rs6007919	0.2359	0.2332	0.0027
50	rs135570	rs136646	0.2496	0.2466	0.0030
50	rs80454	rs8136986	0.2354	0.2372	-0.0018
50	rs5768117	rs133662	0.2339	0.2330	0.0009
49	rs6007851	rs739365	0.2376	0.2372	0.0004

FIG. 29

【図30】

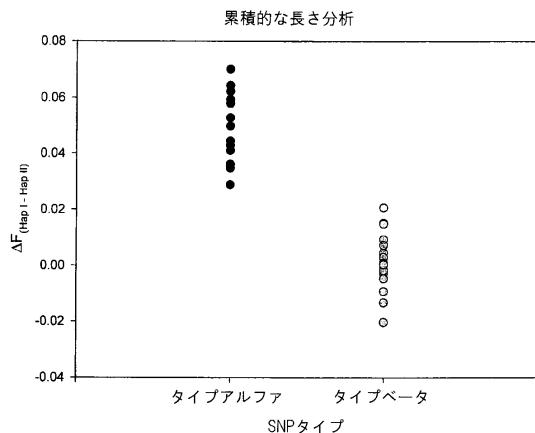


FIG. 30

【図31A】

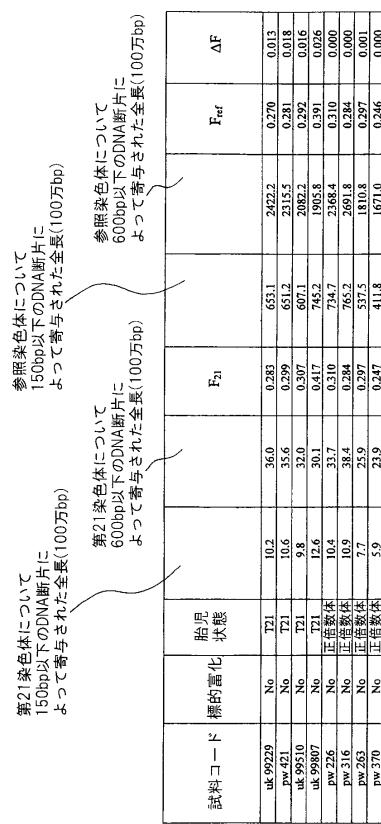


FIG. 31A

【図31B】

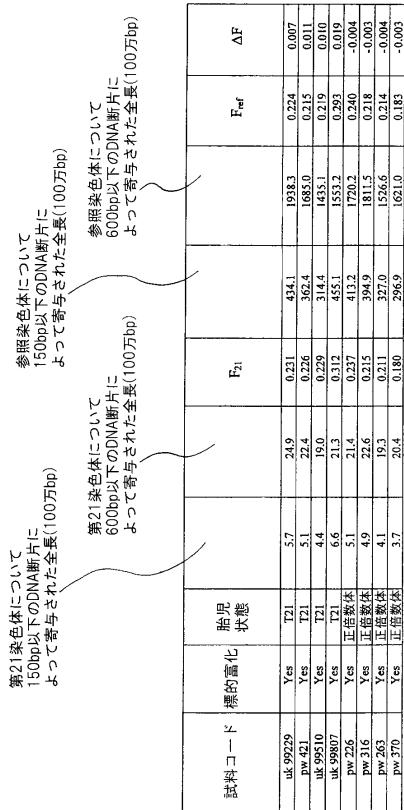


FIG. 31B

【図32】

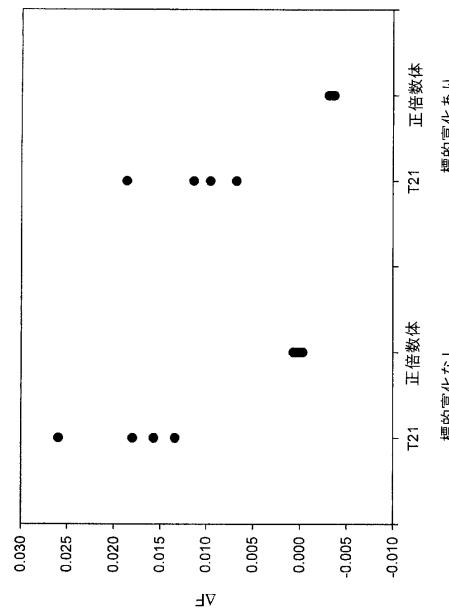


FIG. 32

【図33】
標的富化なし
正倍数体

【図 3 3】

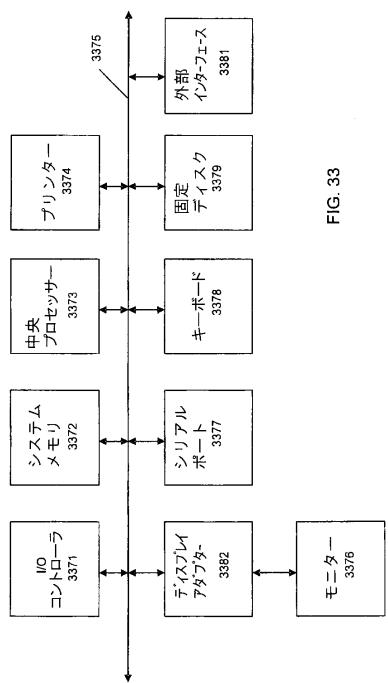


FIG. 33

フロントページの続き

(74)代理人 100087413
弁理士 古賀 哲次

(74)代理人 100117019
弁理士 渡辺 陽一

(74)代理人 100141977
弁理士 中島 勝

(74)代理人 100150810
弁理士 武居 良太郎

(72)発明者 ユク ミン デニス 口
中華人民共和国,香港特別行政区,クーロン,ホマンティン,キン タク ストリート 7,フォース フロア

(72)発明者 チャン クワン チー
中華人民共和国,香港特別行政区,クーロン,メイ フー シン チエン,ブロードウェイ ストリート,ブロック 34,フラット エー,13 フロア

(72)発明者 チエン ウエンリ
中華人民共和国,香港特別行政区,ニューテリトリーズ,シャーティン,ザ チャイニーズ ユニバーシティ オブ ホンコン ピージーエイチ2,アール2601

(72)発明者 ワイ クウン ロッサ チウ
中華人民共和国,香港特別行政区,マ ロック パス 52,ダブル ヘブン,ハウス 31

審査官 柴原 直司

(56)参考文献 国際公開第2007/028155 (WO, A1)

(58)調査した分野(Int.Cl., DB名)

C 12 N 15 / 00 - 15 / 90
C 12 Q 1 / 68
G 06 F 19 / 18
J S T P l u s / J M E D P l u s / J S T 7 5 8 0 (J D r e a m I I I)
P u b M e d