

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6623400号
(P6623400)

(45) 発行日 令和1年12月25日 (2019. 12. 25)

(24) 登録日 令和1年12月6日 (2019. 12. 6)

(51) Int. Cl.	F I
C 1 2 Q 1/6827 (2018. 01)	C 1 2 Q 1/6827 Z
C 1 2 M 1/00 (2006. 01)	C 1 2 M 1/00 A
C 1 2 M 1/34 (2006. 01)	C 1 2 M 1/34 Z

請求項の数 15 (全 32 頁)

(21) 出願番号	特願2018-509952 (P2018-509952)	(73) 特許権者	519181180
(86) (22) 出願日	平成27年5月6日 (2015. 5. 6)		チョージャン アンノロード バイオーテ
(65) 公表番号	特表2018-514234 (P2018-514234A)		クノロジー カンパニー リミテッド
(43) 公表日	平成30年6月7日 (2018. 6. 7)		Zhejiang Annoroad B
(86) 国際出願番号	PCT/CN2015/078422		io-Technology Co.,
(87) 国際公開番号	W02016/176847		Ltd.
(87) 国際公開日	平成28年11月10日 (2016. 11. 10)		中華人民共和国 チョージャン プロヴィ
審査請求日	平成30年4月4日 (2018. 4. 4)		ンス イウ シティ チョウジアン サブ
			ディストリクト ガオシン ロード ナン
			バー10 スタンダード ワークショップ
			ビルディング 2 1/EF
			1/F Standard Worksh
			op Building 2, No 1
			O Gaoxin Rd, Chouji
			ang Subdistrict, Yiw
			最終頁に続く

(54) 【発明の名称】 染色体異数性を測定するためのキット、装置及び方法

(57) 【特許請求の範囲】

【請求項 1】

胎児の染色体異数性を検出する方法であって、検出対象妊婦由来の末梢血遊離DNAに対してハイスループットシーケンシングを行うことで、すべての染色体を含むシークエンスデータを得ること；前記シークエンスデータ中のすべての染色体をウィンドウに分割することにより、該染色体のカバレッジを計算し、前記染色体の各々の補正前カバレッジを得ること；前記検出対象妊婦の各ウィンドウ中に含まれる固有の配列の数量に対してZ検定を行ってZcnv値を得、さらに得られたZcnv値の大きさに基づいて前記検出対象妊婦のコピー数異常染色体フラグメントを得ること、ここで、前記検出対象妊婦のコピー数異常染色体フラグメントは、前記シークエンスデータ中の300Kb以上の長さを有するフラグメントであって、当該フラグメントに含まれるすべてのウィンドウのうち80%以上のウィンドウのZcnv値が4以上であるか、または-4以下であるフラグメントであり；前記検出対象妊婦のコピー数異常染色体フラグメントによる前記染色体の補正前カバレッジに対する影響を利用することにより、各染色体の補正前カバレッジに対して補正を行い、これにより各染色体の補正後カバレッジを得ること；及び各染色体の補正後カバレッジを用いて各染色体のZ検定を行うことによりZaneu値を得、さらにZaneu値の絶対値が3以上であるか否かによって前記染色体が異数性を有するか否かを判定することを含むことを特徴とし、

10

20

ここで、前記 Z a n e u 値の絶対値が 3 以上である場合、前記染色体は異数性を有すると判定され、

ここで、前記検出対象妊婦のコピー数異常染色体フラグメントによる前記染色体の補正前カバレッジに対する影響は以下のパラメータ によって示され、

【数 1】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \quad \cdot \cdot \cdot (2)$$

前記式 (2) において、m は前記コピー数異常染色体フラグメントが存在する染色体の有効配列長であって、単位は M b で表され； n は前記検出対象妊婦の前記コピー数異常染色体フラグメントの長さであって、単位は M b で表され； c n は前記妊婦の前記コピー数異常染色体フラグメントの出現する回数であり、

前記式 (2) において、 f は前記検出対象妊婦由来の末梢血遊離 D N A に含まれる胎児遊離 D N A の濃度であり、且つ前記胎児遊離 D N A の濃度 f は 5 0 % 未満であり、

ここで、前記染色体の補正前カバレッジに対する補正は、

【数 2】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて行われ、

式中、

【数 3】

$$\hat{x}$$

は各前記染色体の補正前カバレッジを示し、 x ' は各前記染色体の補正後染色体カバレッジを示すものである、方法。

【請求項 2】

前記シークエンスデータ中のすべての染色体を同じサイズのウィンドウに分割する形でカバレッジを計算することで、前記染色体の補正前カバレッジを得ることを特徴とする、請求項 1 に記載の方法。

【請求項 3】

それぞれの前記ウィンドウのサイズは 1 0 0 K b であり、且つ隣り合う二つの前記ウィンドウ間の重複度は 5 0 % であることを特徴とする、請求項 2 に記載の方法。

【請求項 4】

前記検出対象妊婦の各ウィンドウに含まれる固有の配列の数量に対して Z 検定を行って Z c n v 値を得、さらに得られた Z c n v 値の大きさに基づいて前記検出対象妊婦のコピー数異常染色体フラグメントを得るステップにおいて、

各前記染色体の G C 含有量とマッピング率に基づいて前記ウィンドウの前記固有の配列の数量について計算を行い、各前記ウィンドウの前記固有の配列の数量の補正前カバレッジを得ること、及び

前記ウィンドウの前記固有の配列の数量の補正前カバレッジに対して標準化処理を行い、前記ウィンドウの前記固有の配列の数量の Z c n v 値を得て、さらに前記 Z c n v 値の大小により前記検出対象妊婦が前記コピー数異常染色体フラグメントを有するか否かを判定すること、

前記シークエンスデータ中に 3 0 0 K b 以上のフラグメントが存在し、且つ前記 3 0 0 K b 以上のフラグメント中のすべてのウィンドウのうち 8 0 % 以上のウィンドウにおける

前記固有の配列の数量の Z_{cnv} 値が 4 と等しいかそれより大きく、または - 4 と等しいかそれより小さい場合、300 K b 以上のフラグメントが前記検出対象妊婦のコピー数異常染色体フラグメントであると判定することを含むことを特徴とする、請求項 1 に記載の方法。

【請求項 5】

前記染色体の補正後カバレッジを用いて各前記染色体に対して Z_{aneu} 検定を行い、 Z_{aneu} 値を得るステップにおいて、前記 Z_{aneu} 値は

【数 4】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

10

に従って計算され、そのうち、

【数 5】

$$\bar{x}$$

は LOESS アルゴリズムによるものであり、既知の陰性サンプル群によって得られる補正前カバレッジであり、 s は陰性サンプル群における

【数 6】

$$(x' - \bar{x})$$

20

の標準偏差であることを特徴とする、請求項 1 に記載の方法。

【請求項 6】

胎児の染色体異数性を検出する装置であって、シークエンスデータ取得モジュール、第一カバレッジ計算モジュール、 Z_{cnv} 値計算モジュール、コピー数異常染色体フラグメント検索モジュール、コピー数異常染色体フラグメント認定モジュール、計算モジュール、補正モジュール、第二カバレッジ計算モジュール、 Z_{aneu} 値判定モジュール、染色体異数性認定モジュールを備え、前記シークエンスデータ取得モジュールは、妊婦由来の末梢血遊離 DNA に対してハイスループットシーケンシングを行うことで、すべての染色体を含むシークエンスデータを得るためのものであり、前記第一カバレッジ計算モジュールは、前記シークエンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、各前記染色体の補正前カバレッジを得るためのものであり、

30

前記 Z_{cnv} 値計算モジュールは、前記検出対象妊婦の前記ウィンドウに含まれる固有の配列の数量の Z_{cnv} 値を計算するためのものであり、

前記コピー数異常染色体フラグメント検索モジュールは、前記シークエンスデータ中において 300 K b 以上のフラグメントであって、且つすべてのウィンドウのうち 80 % 以上のウィンドウの染色体フラグメントの Z_{cnv} 値は 4 以上であるか、または - 4 以下であるフラグメントを検索するものであり、

40

前記コピー数異常染色体フラグメント認定モジュールは、前記シークエンスデータにおける検索により得られる、前記 300 K b 以上のフラグメントで且つすべてのウィンドウのうち 80 % 以上のウィンドウの染色体フラグメントの Z_{cnv} 値が 4 以上であるか、または - 4 以下であるフラグメントを検出対象妊婦のコピー数異常染色体フラグメントとして認定するものであり、

計算モジュールは、式 (2) に示される計算式に従ってパラメータ を計算するものであり、前記パラメータ は妊婦のコピー数異常染色体フラグメントの各前記染色体の補正前カバレッジに対する影響であり、

【数 7】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \cdot \cdot \cdot (2)$$

前記式(2)において、 m は前記コピー数異常染色体フラグメントが存在する染色体の有効配列長であり、単位はMbである； n は前記検出対象妊婦の前記コピー数異常染色体フラグメントの長さであり、単位はMbである； cn は前記妊婦の前記コピー数異常染色体フラグメントの出現する回数である、 f は前記検出対象妊婦由来の末梢血遊離DNAに含まれる胎児遊離DNAの濃度であり、且つ前記胎児遊離DNAの濃度 f は50%未満であり、前記補正モジュールは、

10

【数 8】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて、各前記染色体の補正前カバレッジに対して補正を行い、各前記染色体の補正後カバレッジを得るものであり、そのうち、

【数 9】

20

$$\hat{x}$$

は各染色体の補正前カバレッジを示し、 x' は各前記染色体の補正後染色体カバレッジを示している、前記第二カバレッジ計算モジュールは、各前記染色体の補正後カバレッジを用いて各染色体のZaneu値を計算するものであり、前記Zaneu値判定モジュールは、前記Zaneu値の絶対値が3と等しくもしくはそれより大きいかを判定するものであり、前記染色体異数性認定モジュールは、前記Zaneu値の絶対値が3と等しくもしくはそれより大きい場合、前記染色体が異数性を有すると認定するものであることを特徴とする、前記装置。

30

【請求項 7】

前記第一カバレッジ計算モジュールは、前記シーケンスデータ中のすべての染色体を同じサイズのウィンドウに分割する染色体ウィンドウ分割サブモジュール、

前記同じサイズのウィンドウの形でカバレッジを計算することで、各前記染色体の補正前カバレッジを得る第一カバレッジ計算サブモジュールを含むことを特徴とする、請求項6に記載の装置。

【請求項 8】

前記染色体ウィンドウ分割サブモジュールにおいて、それぞれの前記ウィンドウのサイズは100Kbであり、且つ隣り合う二つの前記ウィンドウの間の重複度は50%である、請求項7に記載の装置。

40

【請求項 9】

Zcnv値計算モジュールは、固有の配列カバレッジ計算ユニット、固有の配列Zcnv値計算ユニットを備え、前記固有の配列カバレッジ計算ユニットは、前記各染色体のGC含有量及びマッピング率に基づいて各前記ウィンドウの前記固有の配列の数量について計算を行い、前記ウィンドウの前記固有の配列の数量の補正前カバレッジを得るためのものであり、前記固有の配列Zcnv値計算ユニットは、各前記ウィンドウの前記固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各前記ウィンドウの前記固有の配列の数量のZcnv値を得るためのものであることを特徴とする、請求項6に記載の装置。

【請求項 10】

50

前記第二カバレッジ計算モジュールにおいて、
前記 Z a n e u は
【数 1 0】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

に従って計算され、そのうち、
【数 1 1】

$$\bar{x}$$

は L O E S S アルゴリズムにより、既知の陰性サンプル群から得られるカバレッジの値であり、s は陰性サンプル群における

【数 1 2】

$$(x' - \bar{x})$$

の標準偏差であることを特徴とする、請求項 6 に記載の装置。

【請求項 1 1】

胎児の染色体異数性を検出するためのキットであり、検出試薬及び検出機器、第一カバレッジ計算機器、Z c n v 値計算機器、コピー数異常染色体フラグメント検索機器、コピー数異常染色体フラグメント認定機器、計算機器、補正機器、第二カバレッジ計算機器、Z a n e u 値判定機器、染色体異数性認定機器を備え、前記検出試薬及び検出機器は、検出対象妊婦由来の末梢血遊離 DNA に対してハイスループットシーケンシングを行うことで、すべての染色体を含むシークエンスデータを得るためのものであり、前記第一カバレッジ計算機器は、前記シークエンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、各前記染色体の補正前カバレッジを得るためのものであり、

前記 Z c n v 値計算機器は、前記検出対象妊婦の前記ウィンドウに含まれる固有の配列の数量を用いて Z 検定を行い、Z c n v 値を得るためのものであり、

前記コピー数異常染色体フラグメント検索機器は、前記シークエンスデータ中において 3 0 0 K b 以上のフラグメントであって、且つすべてのウィンドウのうち 8 0 % 以上のウィンドウ中の染色体フラグメントの Z c n v 値は 4 以上であるか、または - 4 以下であるフラグメントを検索するものであり、

前記コピー数異常染色体フラグメント認定機器は、前記 Z c n v 値の大小により前記検出対象妊婦のコピー数異常染色体フラグメントを得るためのものであり、

前記 計算機器は、式 (2) に示される計算式に従ってパラメータ を計算するものであり、そのうち、前記パラメータ は妊婦のコピー数異常染色体フラグメントの各前記染色体の補正前カバレッジに対する影響である、

【数 1 3】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \cdot \cdot \cdot (2)$$

m は前記コピー数異常染色体フラグメントが所在する染色体の有効配列長であり、単位は M b である； n は前記検出対象妊婦の前記コピー数異常染色体フラグメントの長さであり、単位は M b である； c n は前記妊婦の前記コピー数異常染色体フラグメントの出現する回数である、 f は前記検出対象妊婦由来の末梢血遊離 DNA に含まれる胎児遊離 DNA の濃度であり且つ前記胎児遊離 DNA の濃度 f は 5 0 % 未満であり、前記補正機器は、

【数 1 4】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて、各前記染色体の補正前カバレッジに対して補正を行い、各前記染色体の補正後カバレッジを得るものであり、そのうち、

【数 1 5】

$$\hat{x}$$

10

は各前記染色体の補正前カバレッジを示し、 x' は各前記染色体の補正後染色体カバレッジを示している、前記第二カバレッジ計算機器は、各前記染色体の補正後カバレッジを用いて各染色体のZaneu値を計算するものであり、前記Zaneu値判定機器は、前記Zaneu値の絶対値が3と等しくもしくはそれより大きいかなかを判定するものであり、前記染色体異数性認定機器は、前記Zaneu値の絶対値が3と等しくもしくはそれより大きい場合、前記染色体が異数性を有すると認定するものであることを特徴とする、前記キット。

【請求項 1 2】

20

前記第一カバレッジ計算機器は、前記シークエンズデータ中のすべての染色体を同じサイズのウィンドウに分割するための染色体ウィンドウ分割部材、前記同じサイズのウィンドウの形でカバレッジを計算し、各前記染色体の補正前カバレッジを得る第一カバレッジ計算部材を含むことを特徴とする、請求項 1 1に記載のキット。

【請求項 1 3】

前記染色体ウィンドウ分割部材において、前記ウィンドウのサイズは100Kbであり、且つ隣り合う二つの前記ウィンドウの間の重複度は50%であることを特徴とする、請求項 1 2に記載のキット。

【請求項 1 4】

前記固有の配列Zcnv値計算機器は、前記染色体のGC含有量とマッピング率を用いて各前記固有の配列の数量について計算を行い、各前記ウィンドウの前記固有の配列の数量の補正前カバレッジを得る固有の配列カバレッジ計算部材、及び、各前記ウィンドウの前記固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各ウィンドウの前記固有の配列の数量のZcnv値を得る固有の配列Zcnv値計算部材を含むことを特徴とする、請求項 1 1に記載のキット。

30

【請求項 1 5】

前記第二カバレッジ計算機器において、Zaneuは

【数 1 6】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

40

に従って計算され、そのうち、

【数 1 7】

$$\bar{x}$$

はLOESSアルゴリズムにより、既知の陰性サンプル群により得られる補正前カバレッジであり、 s は陰性サンプル群における

【数 18】

$$(x' - \bar{x})$$

の標準偏差であることを特徴とする、前記請求項 11 に記載のキット。

【発明の詳細な説明】

【技術分野】

【0001】

本発明はバイオ医薬分野に属し、具体的に言えば、染色体異数性を測定するためのキット、装置及び方法に係る。

10

【背景技術】

【0002】

胎児遊離DNA(cff-DNA)は1997年にLoによって発見されてから二十数年になるが、この発見は非侵襲性の出産前診断方法(NIPT)に可能性を提供している。非侵襲性の出産前診断の方法は主に以下の二つの優位性がある：第一に、NIPTは流産のリスクを負うことがなく、臨床上において羊水検査及び臍帯穿刺(Cordocentesis)などの侵襲的方式によって行われる染色体核型分析は1/2000の流産リスクがあり、研究により早すぎる臍帯穿刺は胎児の位置の傾斜をもたらす可能性がある。もう一方で、NIPTは最も早い場合は妊娠8週目で測定を行うことができ、より早期にリスクを判定することができ、中絶が妊婦にもたらすリスクを低減させることができる。

20

【0003】

これらのメリットにより、非侵襲の出産前関連の研究法は日進月歩であり、応用の範囲はますます広くなり、現在の方法例えばNIPT胎児染色体異数性検出、NIPT胎児単一遺伝子障害検出、NIPT胎児(Copy Number Variation, CNV)検出、NIPT胎児全ゲノム測定、NIPT胎児親子鑑定などがある。

【0004】

現在、すべてのNIPT応用において、応用がもっとも広く比較的最も成熟しているのは胎児染色体異数性検出である。胎児染色体に対する異数性検出の多くのアルゴリズムにおいて、Chuiが2008年に発明したハイスループットシーケンシング(MPS)法は臨床的に適用されるものであり、頑健性を示している。ダウン症候群について、偽陽性率(FPR)は0.443%に、偽陰性率(FNR)は0.004%にまで低下する。エドワーズ症候群について、FPRは0.22%、FNRは0.025%である。

30

【0005】

前記方法は非常に低いエラー率に達しているものの、判定ミスリスクは依然存在する。そのため、現在の方法に対して改善を行い、測定のエラー率をできるだけ低下させる必要がある。

【発明の概要】

【発明が解決しようとする課題】

【0006】

本出願の主な目的は染色体異数性を検出するキット、装置及び方法を提供することで、検出の偽陽性率を低下させることである。

40

【課題を解決するための手段】

【0007】

前記目的を実現するために、本出願は一方では、染色体異数性を検出する方法を提供するものであり、該方法は以下のステップを含む：妊婦由来の末梢血遊離DNAに対してハイスループットシーケンシングを行うことで、すべての染色体を含むシーケンスデータを得る、前記シーケンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、各染色体の補正前カバレッジを得る、前記検出対象妊婦の各ウィンドウ中の固有の配列の数量に対してZ検定を行い、 Z_{cnv} 値を得て、さらに前記 Z_{cnv} 値の大小に基づいて検出対象妊婦のコピー数異常染色体フラグメントを得る、前記検出対象妊

50

婦のコピー数異常染色体フラグメントはシークエンスデータにおける300Kb以上のフラグメントであり、且つ300Kb以上のフラグメントにおいて、すべてのウィンドウのうち80%以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値は4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントであり、検出対象妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響を利用することにより、各染色体の補正前カバレッジに対して補正を行い、これにより各染色体の補正後カバレッジを得る、及び各染色体の補正後カバレッジを用いて各染色体に対して Z 検定を行い、 Z_{aneu} 値を得て、さらに Z_{aneu} 値の絶対値が3と等しくもしくはそれより大きいかなにかによって前記染色体が異数性を有するかを判定する、前記 Z_{aneu} 値の絶対値が3と等しくもしくはそれより大きい場合、前記染色体は異数性を有する、そのうち、前記検出対象妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響はパラメータによって示され、胎児が母体のコピー数異常染色体フラグメントを遺伝している場合、パラメータの計算式は式(1)に示す通りである：

【0008】

【数1】

$$\alpha = \frac{(m-n) \cdot 2 + n \cdot cn}{m \cdot 2} \dots (1)$$

【0009】

胎児が母体のコピー数異常染色体フラグメントを遺伝していない場合、パラメータの計算式は式(2)に示す通りである：

【0010】

【数2】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \dots (2)$$

【0011】

式(1)及び(2)において、 m はコピー数異常染色体フラグメントが存在する染色体の有効長さであり、単位はMbである； n は検出対象妊婦の前記コピー数異常染色体フラグメントの長さであり、単位はMbである； cn は妊婦のコピー数異常染色体フラグメントの出現する回数である；式(2)において、 f は検出対象妊婦由来の末梢血遊離DNAに含まれる胎児遊離DNAの濃度であり且つ仮に胎児遊離DNAの濃度 f が50%未満であり、さらに

【0012】

【数3】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて、各染色体の補正前カバレッジに対して補正を行う、そのうち

【0013】

【数4】

$$\hat{x}$$

10

20

30

40

50

は各記染色体の補正前カバレッジを示し、 x' は各染色体の補正後染色体カバレッジを示すものである。

【0014】

さらに、シークエンスデータ中のすべての染色体を同じサイズのウインドウに分割する形でカバレッジを計算することで、各染色体の補正前カバレッジを得る。

【0015】

さらに、それぞれのウインドウのサイズは100Kbであり、且つ隣り合う二つのウインドウ間の重複度は50%である。

【0016】

さらに、検出対象妊婦の各ウインドウ中の固有の配列の数量に対してZ検定を行い、 Z_{cnv} 値を得て、さらに Z_{cnv} 値に基づいて検出対象妊婦のコピー数異常染色体フラグメントを得るステップにおいて、前記シークエンスデータ中の各配列のシークエンス深度に基づいて、各ウインドウの固有の配列の数量を統計する、各染色体のGC含有量とマッピング率に基づいて各ウインドウの固有の配列の数量について計算を行い、各ウインドウの単一配列の数量の補正前カバレッジを得る、及び各ウインドウの固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各ウインドウの固有の配列の数量の Z_{cnv} 値を得て、さらに Z_{cnv} 値の大小により検出対象妊婦がコピー数異常染色体フラグメントを有するか否かを判定する、シークエンスデータ中に300Kb以上のフラグメントが存在し、且つ前記300Kb以上のフラグメント中のすべてのウインドウのうち80%以上のウインドウにおける固有の配列の数量の Z_{cnv} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さい場合、300Kb以上のフラグメントが検出対象妊婦のコピー数異常染色体フラグメントであると判定する。

【0017】

さらに、各染色体の補正後カバレッジを用いて各染色体に対してZ検定を行い、 Z_{aneu} 値を得るステップにおいて、 Z_{aneu} 値は

【0018】

【数5】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

に従って計算され、そのうち、

【0019】

【数6】

$$\bar{x}$$

はLOESSアルゴリズムによるものであり、既知の陰性サンプル群によって得られる補正前カバレッジであり、 s は陰性サンプル群における

【0020】

【数7】

$$(x' - \bar{x})$$

の標準偏差である。

【0021】

前記目的を実現するために、本出願は一方において、染色体異数性を検出する装置を提供し、該装置は以下のモジュールを含む：シークエンスデータ取得モジュール、第一カバレッジ計算モジュール、 Z_{cnv} 値計算モジュール、コピー数異常染色体フラグメント検索モジュール、コピー数異常染色体フラグメント認定モジュール、第一計算モジュール

、第二計算モジュール、補正モジュール、第二カバレッジ計算モジュール、 Z_{aneu} 値判定モジュール、染色体異数性認定モジュール。前記シーケンスデータ取得モジュールは、妊婦由来の末梢血遊離DNAに対してハイスループットシーケンシングを行うことで、すべての染色体を含むシーケンスデータを得るためのものであり、前記第一カバレッジ計算モジュールは、シーケンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、各染色体の補正前カバレッジを得るためのものであり、前記 Z_{cnv} 値計算モジュールは、検出対象妊婦のウィンドウ中の固有の配列の数量の Z_{cnv} 値を計算するためのものであり、前記コピー数異常染色体フラグメント検索モジュールは、前記シーケンスデータ中において300Kb以上のフラグメントを検索するものであり、且つ前記300Kb以上のフラグメントにおいて、すべてのウィンドウのうち80%以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値は4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントであり、コピー数異常染色体フラグメント認定モジュールは、シーケンスデータにおける検索により得られる300Kb以上のフラグメントで且つすべてのウィンドウのうち80%以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントを検出対象妊婦のコピー数異常染色体フラグメントとして認定するものであり、第一計算モジュールは、胎児が母体のコピー異常フラグメントを遺伝している場合、式(1)に示される計算式に従ってパラメータを計算するものであり、そのうち、前記パラメータは妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響である。

【0022】

【数8】

$$\alpha = \frac{(m-n) \cdot 2 + n \cdot cn}{m \cdot 2} \dots (1)$$

m はコピー数異常染色体フラグメントが存在する染色体の有効長さであり、単位はMbである； n は妊婦のコピー数異常染色体フラグメントの長さであり、単位はMbである； cn は妊婦のコピー数異常染色体フラグメントの出現する回数である、第二計算モジュールは、胎児が母体のコピー数異常の染色体を遺伝していない場合において、式(2)に示す計算式に従って前記パラメータを計算するものであり、

【0023】

【数9】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \dots (2)$$

m はコピー数異常染色体フラグメントが存在する染色体の有効長さであり、単位はMbである； n は検出対象妊婦のコピー数異常染色体フラグメントの長さであり、単位はMbである； cn は妊婦の前記コピー数異常染色体フラグメントの出現する回数である、 f は検出対象妊婦由来の末梢血遊離DNAに含まれる胎児遊離DNAの濃度であり且つ仮に胎児遊離DNAの濃度 f が50%未満であり、補正モジュールは、

【0024】

【数 1 0】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて、各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得るものであり、そのうち、

【0 0 2 5】

【数 1 1】

10

$$\hat{x}$$

は各染色体の補正前カバレッジを示し、 x' は各染色体の補正後染色体カバレッジを示している、第二カバレッジ計算モジュールは、各染色体の補正後カバレッジを用いて各染色体の Z_{aneu} 値を計算するものであり、 Z_{aneu} 値判定モジュールは、 Z_{aneu} 値の絶対値が 3 と等しくもしくはそれより大きいかを判定するものであり、染色体異数性認定モジュールは、 Z_{aneu} 値の絶対値が 3 と等しくもしくはそれより大きい場合、染色体が異数性を有すると認定するものである。

【0 0 2 6】

20

さらに、第一カバレッジ計算モジュールは、シーケンスデータ中のすべての染色体を同じサイズのウィンドウに分割する染色体ウィンドウ分割サブモジュール、同じサイズのウィンドウの形でカバレッジを計算することで、各染色体の補正前カバレッジを得る第一カバレッジ計算サブモジュールを含む。

【0 0 2 7】

染色体ウィンドウ分割サブモジュールにおいて、それぞれのウィンドウのサイズは 1 0 0 K b であり、且つ隣り合う二つのウィンドウの間の重複度は 5 0 % である。

【0 0 2 8】

さらに、固有の配列計算モジュールは、固有の配列統計ユニット、固有の配列カバレッジ計算ユニット、固有の配列 Z_{cnv} 値計算ユニットを備え、固有の配列統計ユニットは、シーケンスデータ中の各配列のシーケンス深度に基づいて、各ウィンドウの固有の配列の数量を統計するためのものであり、固有の配列カバレッジ計算ユニットは、各染色体の GC 含有量及びマッピング率に基づいて各ウィンドウの固有の配列の数量について計算を行い、ウィンドウの固有の配列の数量の補正前カバレッジを得るためのものであり、固有の配列 Z_{cnv} 値計算ユニットは、各ウィンドウの固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各ウィンドウの固有の配列の数量の Z_{cnv} 値を得るためのものである。

30

【0 0 2 9】

さらに、第二カバレッジ計算モジュールにおいて、 Z_{aneu} は

【0 0 3 0】

40

【数 1 2】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

に従って計算され、そのうち、

【0 0 3 1】

【数 1 3】

$$\bar{x}$$

は L O E S S アルゴリズムにより、既知の陰性サンプル群から得られるカバレッジ値であり、 s は陰性サンプル群における

【0 0 3 2】

【数 1 4】

$$(x' - \bar{x})$$

10

の標準偏差である。

【0 0 3 3】

本発明はもう一方において、染色体異数性を検出するためのキットを提供し、該キットは検出試薬及び検出機器、第一カバレッジ計算機器、 Z_{cnv} 値計算機器、コピー数異常染色体フラグメント検索機器、コピー数異常染色体フラグメント認定機器、第一計算機器、第二計算機器、補正機器、第二カバレッジ計算機器、 Z_{aneu} 値判定機器、染色体異数性認定機器を備え、検出試薬及び検出機器は、検出対象妊婦由来の末梢血遊離 DNA に対してハイスループットシーケンシングを行うことで、すべての染色体を含むシークエンスデータを得るためのものであり、第一カバレッジ計算機器は、シークエンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、各染色体の補正前カバレッジを得るためのものであり、 Z_{cnv} 値計算機器は、検出対象妊婦の前記ウィンドウ中の固有の配列の数量を用いて Z 検定を行い、 Z_{cnv} 値を得るためのものであり、コピー数異常染色体フラグメント検索機器は、シークエンスデータ中において 3 0 0 K b 以上のフラグメントを検索するものであり、且つ 3 0 0 K b 以上のフラグメントにおいて、すべてのウィンドウのうち 8 0 % 以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値は 4 と等しいかそれより大きく、または - 4 と等しいかそれより小さいフラグメントであり、コピー数異常染色体フラグメント認定機器は、 Z_{cnv} 値の大小により検出対象妊婦のコピー数異常染色体フラグメントを得るためのものであり、第一計算機器は、胎児が母体の前記コピー異常フラグメントを遺伝している場合、式 (1) に示される計算式に従ってパラメータ を計算するものであり、そのうち、パラメータ は妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響である、

20

30

【0 0 3 4】

【数 1 5】

$$\alpha = \frac{(m - n) \cdot 2 + n \cdot cn}{m \cdot 2} \cdot \cdot \cdot (1)$$

40

m はコピー数異常染色体フラグメントが存在する染色体の有効長さであり、単位は M b である； n は検出対象妊婦の前記コピー数異常染色体フラグメントの長さであり、単位は M b である； cn は妊婦のコピー数異常染色体フラグメントの出現する回数である、第二計算機器は、胎児が母体のコピー数異常の染色体を遺伝していない場合において、式 (2) に示す計算式に従ってパラメータ を計算するものであり、

【0 0 3 5】

【数 1 6】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \cdot \cdot \cdot (2)$$

mはコピー数異常染色体フラグメントが存在する染色体の有効長さであり、単位はMbである；nは検出対象妊婦のコピー数異常染色体フラグメントの長さであり、単位はMbである；cnは妊婦のコピー数異常染色体フラグメントの出現する回数である、fは検出対象妊婦由来の末梢血遊離DNAに含まれる胎児遊離DNAの濃度であり且つ仮に胎児遊離DNAの濃度fが50%未満であり、補正機器は、

10

【0036】

【数 1 7】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて、各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得るものであり、そのうち、

【0037】

20

【数 1 8】

$$\hat{x}$$

は各染色体の補正前カバレッジを示し、x'は各染色体の補正後染色体カバレッジを示している、

第二カバレッジ計算機器は、各染色体の補正後カバレッジを用いて各染色体のZ_{aneu}値を計算するものであり、

Z_{aneu}値判定機器は、Z_{aneu}値の絶対値が3と等しくもしくはそれより大きいかが否かを判定するものであり、

30

染色体異数性認定機器は、Z_{aneu}値の絶対値が3と等しくもしくはそれより大きい場合、染色体が異数性を有すると認定するものである。

【0038】

さらに、第一カバレッジ計算機器は、シーケンスデータ中のすべての染色体を同じサイズのウィンドウに分割するための染色体ウィンドウ分割部材、同じサイズのウィンドウの形でカバレッジを計算し、各染色体の補正前カバレッジを得る第一カバレッジ計算部材を含む。

【0039】

さらに、染色体ウィンドウ分割部材において、ウィンドウのサイズは100Kbであり、且つ隣り合う二つのウィンドウの間の重複度は50%である。

40

【0040】

さらに、固有の配列Z_{cnv}値計算機器は、シーケンスデータ中の各配列のシーケンス深度により、各ウィンドウの固有の配列の数量を統計する固有の配列統計部材、染色体のGC含有量とマッピング率を用いて各前記固有の配列の数量について計算を行い、各ウィンドウの前記固有の配列の数量の補正前カバレッジを得る固有の配列カバレッジ計算部材、

各ウィンドウの前記固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各ウィンドウの固有の配列の数量のZ_{cnv}値を得る固有の配列Z_{cnv}値計算部材を含む。さらに、第二カバレッジ計算機器において、Z_{aneu}は

【0041】

50

【数 1 9】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

に従って計算され、そのうち、

【 0 0 4 2】

【数 2 0】

$$\bar{x}$$

10

は L O E S S アルゴリズムにより、既知の陰性サンプル群により得られる補正前カバレッジであり、s は陰性サンプル群における

【 0 0 4 3】

【数 2 1】

$$(x' - \bar{x})$$

の標準偏差である。

【発明の効果】

20

【 0 0 4 4】

本出願の技術構成を用いて、母体染色体上に存在する特定サイズのコピー数異常染色体フラグメントをスクリーニングし、さらに染色体に異数性が存在するかを判定し、該母体のコピー数異常染色体フラグメントによる各染色体のカバレッジに対する影響を除去することにより、各染色体の補正後のカバレッジを得て、本出願の補正後のカバレッジの計算及び判定により得られた染色体異数性の結果はより正確である。

【 0 0 4 5】

本出願の一部となる図面を用いて本出願に対してさらに理解を深め、本出願の例示的実施例及びその説明を本出願の解釈に用いるが、これらは本出願を不当に限定するものではない。図面は以下の通りである。

30

【図面の簡単な説明】

【 0 0 4 6】

【図 1】本出願の典型的な実施形態における染色体異数性を測定する方法のフロー模式図である。

【図 2】本出願の典型的な実施形態における染色体異数性の検出装置の構造模式図である。

【図 3】図 3 A、図 3 B 及び図 3 C はそれぞれ本出願の実施例 1 の 1 3 号染色体、1 8 号染色体及び 2 1 号染色体異数性検出の補正結果の模式図である。

【図 4】本出願の実施例 2 のサンプル E K 0 1 8 7 5 及び B D 0 1 4 6 2 の 2 1 号染色体上における異数性の補正結果の模式図である。

40

【図 5】本出願の実施例 3 のサンプル E K 0 1 8 7 5 の 2 1 号染色体の異数性検出の補正結果の模式図である。

【図 6】本出願の実施例 4 のサンプル B D 0 1 4 6 2 の 2 1 号染色体異数性検出の補正結果である。

【発明を実施するための形態】

【 0 0 4 7】

説明すべきことは、矛盾しない前提下において、本出願の実施例及び実施例中の特徴は互いに組み合わせることができる。以下において実施例を参照しながら本出願を説明する。

【 0 0 4 8】

50

本出願中において、 Z_{cnv} または Z_{aneu} は統計学中の Z 検定 の計算値であり、大量サンプル（即ちサンプル容量が 30 より多い）の平均値差異性検出に用いられる方法である。それは標準正規分布の理論により差異発生の確率を推測することで、二つの平均数の差異が顕著かを比較するものである。

【0049】

マッピング率とはウインドウ内のシーケンス配列をゲノム参照配列上にマッピングした場合の比率である。シーケンス配列は同時にゲノム参照配列上の複数の位置にマッピングでき、唯一の配列ではない可能性があり、そのためウインドウ内のマッピング率は単一配列のマッピング率より大きいものである。

【0050】

説明しておきたいことは、本出願の出願人は従来の方法を用いて大量に分析を行ったところ、少なくとも以下三種類の N I P T 判定ミスに繋がる可能性が存在することを発見した。

【0051】

まず、Lo は 1998 年に c f f - D N A が胎盤に由来することを発見しているが、これは胎盤モザイク (c o n f i n e d p l a c e n t a l m o s a i c i s m : C P M)) が出現した際、我々は N I P T の結果により胎児の状況を正確に予測することができず、結果が正確性を失いやすいことを意味している。次に、妊婦自身に一定の C N V が存在する場合、M P S に基づいてカバレッジを統計することにより、Z 値に変換する方法は正確性に欠ける。妊婦に重複フラグメントが存在する場合、染色体上にマッピングした場合対応する 固有の配列 数が増え、カバレッジの向上により Z 値が大きくなり、偽陽性のリスクが大きくなる。逆に、妊婦にフラグメントの欠失が存在する場合、Z 値が低下し、偽陰性のリスクが大きくなる。以前の研究でもわかるように、胎盤にモザイク (C P M) が出現し及び妊婦コピー数フラグメント異常 (C N V) が出現することは、偽陽性判定の重要な原因となる。最後に、染色体カバレッジの計算または G C 含有量を利用したカバレッジ補正プロセスにはデータ変動が生じ、これにより誤差が生じる場合がある。

【0052】

そのため、前記判定ミスの原因に対して総合的に分析した上で、本出願は染色体異数性を測定する方法を提供し、図 1 に示すように、該方法は以下のステップを含む：検出対象妊婦由来の末梢血遊離 D N A に対してハイスループットシーケンシングを行い、すべての染色体を含むシーケンスデータを得る。シーケンスデータ中のすべての染色体をウインドウに分割する形でカバレッジを計算し、各染色体の補正前カバレッジを得る。検出対象妊婦の各ウインドウにおける 固有の配列 の数量の Z_{cnv} 値について計算を行い、さらに Z_{cnv} データの大小により検出対象妊婦のコピー数異常染色体フラグメントを得る。検出対象妊婦のコピー数異常染色体フラグメントとは、シーケンスデータ中の 300 K b 以上のフラグメントである。さらに 300 K b 以上のフラグメントにおいて、すべてのウインドウのうち 80 % 以上のウインドウ中の染色体フラグメントの Z_{cnv} 値が 4 と等しいかそれより大きく、または - 4 と等しいかそれより小さいのである。検出対象妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響を利用して、各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得る。及び各染色体の補正後カバレッジを用いて、各染色体の Z_{aneu} 値を計算し、さらに Z_{aneu} 値の絶対値が 3 と等しいかそれより大きいかによって染色体が異数性を有するかを判定する。 Z_{aneu} 値の絶対値が 3 と等しいかそれより大きい場合、染色体は異数性である。そのうち、検出対象妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響をパラメータ で表示し、胎児が母体のコピー数異常染色体フラグメントを遺伝した場合、パラメータ の計算式は式 (1) に示す通りである：

【0053】

10

20

30

40

【数 2 2】

$$\alpha = \frac{(m-n) \cdot 2 + n \cdot cn}{m \cdot 2} \dots (1)$$

胎児が母体のコピー数異常染色体フラグメントを遺伝していない場合、パラメータ の計算式は式 (2) に示す通りである：

【 0 0 5 4】

【数 2 3】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \dots (2)$$

式 (1) 及び式 (2) において、mはコピー数異常染色体フラグメントが所在する染色体の有効長さを示すものであり、単位はMbである。nは検出対象妊婦のコピー数異常染色体フラグメントの長さを示し、単位はMbである。cnは妊婦のコピー数異常染色体フラグメントの出現する回数である。式 (2) において、fは検出対象妊婦由来の末梢血遊離DNAに含まれる胎児遊離DNAの濃度であり、且つ仮に胎児遊離DNAの濃度fが50%未満の場合、

【 0 0 5 5】

【数 2 4】

$$x' = \frac{\hat{x}}{\alpha}$$

を利用して、各染色体のカバレッジに対して補正を行い、そのうち、

【 0 0 5 6】

【数 2 5】

$$\hat{x}$$

は各染色体の補正前カバレッジであり、x'は各染色体の補正後の染色体カバレッジである。

【 0 0 5 7】

本出願の前記方法は、従来技術と同じように、シーケンスデータ中の母体にコピー数異常染色体フラグメントを直接除去して考慮しないではなく、母体染色体上に存在する特定サイズのコピー数異常染色体フラグメントをスクリーニングし、さらに染色体に異数性が存在するか否かを判定する際において、コピー数異常染色体フラグメントの各染色体のカバレッジに対する影響を除き、これにより染色体補正後のカバレッジを得て、これにより本出願の方法により測定される染色体異数性の結果はより正確である。

【 0 0 5 8】

本出願の前記方法において、前記検出対象妊婦由来の末梢血の遊離DNAに含まれる胎児遊離DNAの濃度fの計算方法は本分野の通常の計算方法である。例えば、胎児が男性で、且つコピー数異常染色体フラグメントがX染色体にある場合、胎児遊離DNAの濃度は

【 0 0 5 9】

10

20

30

40

50

【数 2 6】

$$f = 2(1 - \frac{\bar{N}_{23}}{\bar{N}})$$

によって計算され、そのうち、

【0 0 6 0】

【数 2 7】

$$\frac{\bar{N}_{23}}{\bar{N}}$$

10

はX染色体上のウインドウ平均固有の配列及びすべてのウインドウ平均固有の配列の比であり、コピー数異常染色体フラグメントが2 1、1 8または1 3号染色体にある場合、胎児遊離DNAの濃度は

【0 0 6 1】

【数 2 8】

$$f = 2(\frac{\bar{N}_i}{\bar{N}} - 1)$$

20

によって計算され、そのうち、

【0 0 6 2】

【数 2 9】

$$\frac{\bar{N}_i}{\bar{N}}$$

30

は2 1、1 8または1 3号染色体のウインドウの平均固有の配列数及びすべてのウインドウの平均固有の配列数の比である。胎児が女性である場合、妊婦由来の末梢血の遊離DNAに対して特定遺伝子のメチル化測定を行う。その原理は以下である：特定の遺伝子の妊婦DNA中及び胎児DNA中におけるメチル化の形は異なるものである。例えば、胎児と胎盤由来のRASSF1A（3号染色体上）遺伝子は高度にメチル化されたものであり、母体自身由来のRASSF1A遺伝子是非メチル化されたものであり、メチル化感受性酵素、例えばHhaI、BstUI（30U）及びHpaIIを用いてcffDNAを処理することで、非メチル化の遺伝子は消化されるが、メチル化された遺伝子は消化されなく、これによりQ-PCRによって胎児cffDNAの含有量の測定を行うことができる。具体的なステップは以下の文献を参照できる：PLOS ONE 9：71-7（2014）、Quantification of Cell-Free DNA in Normal and Complicated Pregnancies：Overcoming Biological and Technical Issues。

40

【0 0 6 3】

本出願の前記方法において、各染色体の補正前カバレッジに対して計算を行う場合、染色体をウインドウに分割する形で計算を行うことで、相対的に安定的な染色体カバレッジを得ることができる。そのため、本出願の好ましい実施例において、シークエンスデータ中のすべての染色体を同じサイズのウインドウに分割する形でカバレッジを計算し、これにより染色体の補正前カバレッジを得る。

50

【 0 0 6 4 】

本出願のさらに好しい実施例において、ウィンドウに分割した形でカバレッジを計算し、それぞれのウィンドウのサイズは100Kbであり、且つ隣り合う二つのウィンドウ間の重複度は50%である。それぞれのウィンドウのサイズを100Kbに制御してさらに隣り合う二つのウィンドウ間の重複度を50%に制御することで、相対的に安定的な染色体カバレッジが得られるだけでなく、さらにウィンドウ間の重複度を向上させることでコピー数異常染色体フラグメントの正確度を向上させることができ、これにより妊婦コピー数異常染色体フラグメントの検出効率を向上させることができる。

【 0 0 6 5 】

本出願の前記方法において、通常のコピー数異常染色体フラグメントを計算する方法とステップに基づいて、シークエンスデータ品質または検出精度の違いにより、適宜コピー数異常染色体フラグメントが満たす条件を調整することで得られるようにする。本出願の好ましい実施例において、検出対象妊婦の各ウィンドウ中の固有の配列数の Z_{cnv} 値に対して計算を行い、さらに Z_{cnv} 値により妊婦のコピー数異常染色体フラグメントを得るステップは以下を含む：シークエンスデータ中の各配列のシークエンス深度に基づいて、各ウィンドウの固有の配列の数量を統計する。各染色体のGC含有量及びマッピング率に基づいて各ウィンドウの固有の配列の数量について計算を行い、各ウィンドウの固有の配列の数量の補正前カバレッジを得る。及び各ウィンドウの固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各ウィンドウの固有の配列の数量の Z_{cnv} 値を得て、さらに Z_{cnv} 値の大小のフラグメントにより検出対象妊婦がコピー数異常染色体フラグメントを有するか判定する。シークエンスデータ中に300Kb以上のフラグメントが存在し、且つ300Kb以上のフラグメント中のすべてのウィンドウのうち80%以上のウィンドウの固有の配列の数量の Z_{cnv} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さい場合、300Kb以上のフラグメントは検出対象妊婦のコピー数異常染色体フラグメントとする。

【 0 0 6 6 】

本出願の前記各ウィンドウの固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各ウィンドウの固有の配列の数量の Z_{cnv} 値を得るステップにおいて、標準化処理は各ウィンドウの固有の配列の数量の補正值に対して、 $(x - u) / sd(x - u)$ を行い、そのうち x が補正後の値で、 u は x の平均値であり、 sd は標準偏差である。前記検出対象妊婦のコピー数異常染色体フラグメントを測定するステップにおいて、「少なくとも300Kbの領域及び該領域中の80%のウィンドウの Z_{cnv} 値は4と等しいかそれより大きく、または-4と等しいかそれより小さい」という条件に設定することで、本出願の前記検出ステップは信頼度の高い妊婦コピー数異常染色体フラグメントを検出することができ、さらにこれらのコピー数異常染色体フラグメントを利用してその所在する染色体の Z_{cnv} 値に対して補正を行うことで、妊婦のコピー数異常染色体フラグメントの測定結果のエラーにより偽陰性の判定となることを回避できる。

【 0 0 6 7 】

本出願の前記方法において、各染色体の補正後カバレッジを用いて各染色体に対して Z 検出を行い、 Z_{aneu} 値を得るステップにおいて、 Z_{aneu} 値は

【 0 0 6 8 】

【数30】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

によって計算され、そのうち、

【 0 0 6 9 】

【数 3 1】

$$\bar{x}$$

は L O E S S アルゴリズムにより、既知の陰性サンプル群により得られるカバレッジであり、s は陰性サンプル群における

【0 0 7 0】

【数 3 2】

$$(x' - \bar{x})$$

10

の標準偏差である。前記式の計算により得られる補正後の Z_{aneu} 値は染色体異数性をより正確に反応することができ、これにより測定結果はより正確となる。

【0 0 7 1】

本出願のもう一つの典型的な実施形態において、染色体異数性を検出する装置を提供し、図 2 に示すように、該装置は以下のモジュールを含む：シークエンスデータ測定モジュール：検出対象妊婦由来の末梢血遊離 DNA に対してハイスループットシーケンシングを行い、これにより染色体のシークエンスデータが得られる。第一カバレッジ計算モジュール：シークエンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、これにより各染色体の補正前カバレッジが得られる。 Z_{cnv} 値計算モジュール：検出対象妊婦の各ウィンドウ中の固有の配列の数量の Z_{cnv} 値に対して計算を行う。コピー数異常染色体フラグメント検索モジュール：シークエンスデータ中において 300 K b 以上のフラグメントを検索するものであり、且つ 300 K b 以上のフラグメントにおいて、すべてのウィンドウのうち 80 % 以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値は 4 と等しいかそれより大きく、または - 4 と等しいかそれより小さいものである。コピー数異常染色体フラグメント決定モジュール：シークエンスデータから検索により得られる 300 K b 以上のフラグメントですべてのウィンドウのうち 80 % 以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値が 4 と等しいかそれより大きく、または - 4 と等しいかそれより小さいフラグメントを検出対象妊婦のコピー数異常染色体フラグメントとする。第一計算モジュール：胎児が母体のコピー数異常染色体フラグメントを遺伝している状況において、式 (1) に示す計算式によりパラメータ α を計算し、そのうち、パラメータ α は妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響である。m はコピー数異常染色体フラグメントが存在する染色体の有効長さであり、単位は Mb である。n は妊婦のコピー数異常染色体フラグメントの長さであり、単位は Mb である。cn は妊婦のコピー数異常染色体フラグメントが出現する回数である。

20

30

【0 0 7 2】

【数 3 3】

$$\alpha = \frac{(m - n) \cdot 2 + n \cdot cn}{m \cdot 2} \cdot \cdot \cdot (1)$$

40

【0 0 7 3】

第二計算モジュール：胎児が母体のコピー数異常染色体を遺伝していない場合、式 (2) に示す計算式でパラメータ α を計算し、m はコピー数異常染色体フラグメントの染色体における有効長さであり、単位は Mb である。n は妊婦のコピー数異常染色体フラグメントの長さであり、単位は Mb である。cn は妊婦のコピー数異常染色体フラグメントの出現回数である。f は検出対象妊婦由来の末梢血遊離 DNA に含まれる胎児遊離 DNA の濃度であり、且つ胎児遊離 DNA の濃度 f は 50 % 未満である。

50

【 0 0 7 4 】

【 数 3 4 】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \cdot \cdot \cdot (2)$$

【 0 0 7 5 】

補正モジュール：

【 0 0 7 6 】

【 数 3 5 】

10

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて、各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得る、そのうち、

【 0 0 7 7 】

【 数 3 6 】

20

$$\hat{x}$$

は染色体の補正前カバレッジを示し、 x' は各染色体の補正後の染色体カバレッジを示している。第二カバレッジ計算モジュール：各染色体の補正後カバレッジを用いて各染色体の Z_{aneu} 値を得る。 Z_{aneu} 値判定モジュール： Z_{aneu} 値の絶対値が3と等しいかそれより大きいかを判定する。染色体異数性認定モジュール： Z_{aneu} 値の絶対値が3と等しいかそれより大きい場合において、染色体が異数性を有すると認定する。

【 0 0 7 8 】

本出願の前記装置において、コピー数異常染色体フラグメント検索モジュール、コピー数異常染色体フラグメント認定モジュール及び補正モジュールを追加し、本出願は母体の染色体上に存在する少なくとも300Kbの領域で該領域中の80%のウィンドウの Z_{cnv} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントをスクリーニングすることにより本出願の前記装置は信頼できる妊婦コピー数異常染色体フラグメントを検出し、さらにこれらのコピー数異常染色体フラグメントをその所在する染色体の Z 検定値に対して補正を行い、さらに妊婦のコピー数異常染色体フラグメントの検出結果の間違いによって偽陰性の判定となることを回避できる。補正モジュールにより、該コピー数異常染色体フラグメントの各染色体のカバレッジに対する影響について補正を行い、これにより本出願の染色体異数性認定モジュールにより染色体異数性の認定結果をより正確にする。本出願の前記装置の補正モジュールにおいて、パラメータの計算式中の胎児濃度は本分野の通常の計算方法であり、具体的には前記の通りであり、ここでは詳しく説明しない。

30

40

【 0 0 7 9 】

説明しなければならないことは、本出願の前記モジュールは装置の一部として一つの計算端末で稼働することができ、該計算機端末が提供するプロセッサにより前記シークエンスデータ検出モジュール、第一カバレッジ計算モジュール、固有の配列計算モジュール、コピー数異常染色体フラグメント検索モジュール、コピー数異常染色体フラグメント認定モジュール、第一計算モジュール、第二計算モジュール、補正モジュール、第二カバレッジ計算モジュール及び染色体異数性認定モジュールにより実現する技術構成であり、明らかにわかるように該計算機端末はハードウェアが実現する設備であり、プロセッサはプログラムを実行するためのハードウェア装置である。且つ本出願が提供する前記各機能

50

性サブモジュールは移動端末、計算機端末または類似の計算装置中で稼働でき、記録媒体の一部として記憶することもできる。

【 0 0 8 0 】

本出願の前記装置において、前記第一カバレッジ計算モジュールは本分野の通常の計算モジュールに基づいて、シークエンスデータの違いにより適宜調整することで得られるものである。本出願の好ましい実施例において、前記第一カバレッジ計算モジュールは以下を含む：染色体ウインドウ分割サブモジュール：シークエンスデータ中のすべての染色体を同じサイズのウインドウに分割する；第一カバレッジ計算サブモジュール：同じサイズのウインドウの形でカバレッジを計算し、これにより各染色体の補正前カバレッジを得る。本出願の染色体ウインドウ分割サブモジュール及び第一カバレッジ計算サブモジュールの第一カバレッジ計算モジュールを用いて同じサイズのウインドウに分割する形で計算を行い、これにより比較的安定的なカバレッジを得ることができる。

10

【 0 0 8 1 】

本出願のより好ましい実施例において、前記染色体ウインドウ分割サブモジュールにおいて、それぞれのウインドウの大きさは 1 0 0 K b であり、且つ隣り合う二つのウインドウの間の重複度は 5 0 % である。それぞれのウインドウを 1 0 0 K b の大きさに分割する形で計算する計算モジュールは比較的安定的なカバレッジを得ることができ、もう一方ではウインドウ間のカバレッジを向上することでコピー数異常染色体フラグメントの精密度を向上させ、さらには妊婦のコピー数異常染色体フラグメントの検出効率を向上させることができる。

20

【 0 0 8 2 】

本出願の前記装置において、固有の配列計算モジュールは通常の計算モジュールで得ることができる。本出願の好ましい実施例において、前記固有の配列計算モジュールはさらに以下を含む：固有の配列統計ユニット：シークエンスデータ中の各配列のシークエンス深度に基づいて、各ウインドウの固有の配列の数量を統計する；固有の配列のカバレッジ計算ユニット：各染色体の G C 含有量及びマッピング率に基づいて各ウインドウの固有の配列の数量について補正を行い、各ウインドウの固有の配列の数量の補正值を得る；固有の配列 Z_{cnv} 値計算ユニット：各ウインドウの固有の配列の数量の補正值に対して標準化処理を行い、各ウインドウの固有の配列の数量の Z_{cnv} 値を得る。

30

【 0 0 8 3 】

本出願の前記固有の配列の計算モジュールは、まず固有の配列統計ユニットにより、シークエンスデータ中の各配列のシークエンス深度に基づいて、各ウインドウの固有の配列の数量を統計し、それから固有の配列カバレッジ計算ユニットを実行し、各染色体の G C 含有量とマッピング率に基づいて各ウインドウの固有の配列の数量について計算を行い、各ウインドウの固有の配列の数量の補正前カバレッジを得て、それから固有の配列 Z_{cnv} 値計算サブユニットを実行し、各ウインドウの固有の配列の数量の補正前カバレッジについて標準化処理を行い、各ウインドウの固有の配列の数量の Z_{cnv} 値を得る。前記ユニットは本分野の通常の計算ユニットに基づいて適宜調整したものであり、これはコピー数異常染色体フラグメント検索モジュールで検索及びコピー数異常染色体フラグメント認定モジュールで認定を行う根拠及び前提であり、これは検出対象サンプル中の母体 DNA コピー数異常染色体フラグメントの存在の正確な認定に根拠を提供している。

40

【 0 0 8 4 】

本出願の前記装置において、第二カバレッジ計算モジュールにおいて、 Z_{aneu} は

【 0 0 8 5 】

【 数 3 7 】

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

50

によって計算され、そのうち、

【 0 0 8 6 】

【 数 3 8 】

$$\bar{x}$$

は L O E S S アルゴリズムによるもので、既知の陰性サンプル群により得られる補正前のカバレッジであり、s は陰性サンプル群における

【 0 0 8 7 】

【 数 3 9 】

$$(x' - \bar{x})$$

10

の標準偏差である。前記式の計算により得られる補正後の Z_{aneu} 値はより正確に染色体異数性を反映することができ、これにより検出結果はより正確である。

【 0 0 8 8 】

本出願のもう一つの典型的な実施形態において、さらに染色体異数性を検出するキットを提供し、キットは以下を含む：シークエンスデータ取得機器：検出対象妊婦由来の末梢血遊離 DNA に対してハイスループットシークエンシングを行い、これによりすべての染色体を含むシークエンスデータを得る；第一カバレッジ計算機器：シークエンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、これにより各染色体の補正前カバレッジを得る； Z_{cnv} 値計算機器：検出対象妊婦の各ウィンドウ中の単一配列の数量の Z_{cnv} 値に対して計算を行う；コピー数異常染色体フラグメント検索機器：シークエンスデータ中において 3 0 0 K b 以上のフラグメントを検索し、且つ 3 0 0 K b 以上のフラグメントにおいて、すべてのウィンドウのうち 8 0 % 以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値は 4 と等しいかそれより大きく、または - 4 と等しいかそれより小さいフラグメントであり、コピー数異常染色体フラグメント認定機器：シークエンスデータ中から検索される 3 0 0 K b 以上のフラグメントで且つすべてのウィンドウのうち 8 0 % 以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値は 4 と等しいかそれより大きく、または - 4 と等しいかそれより小さいフラグメントは検出対象妊婦のコピー数異常染色体フラグメントである；第一計算機器：胎児が母体のコピー数異常染色体フラグメントを遺伝している場合において、式 (1) に示される計算式でパラメータ α を計算し、そのうち、パラメータ α は妊婦のコピー数異常染色体フラグメントの各染色体の補正前カバレッジに対する影響である；m はコピー数異常染色体フラグメントが所在する染色体の有効長さであり、単位は M b である；n は妊婦のコピー数異常染色体フラグメントの長さであり、単位は M b である；cn は妊婦のコピー数異常染色体フラグメントが出現する回数である；

【 0 0 8 9 】

【 数 4 0 】

$$\alpha = \frac{(m - n) \cdot 2 + n \cdot cn}{m \cdot 2} \cdot \cdot \cdot (1)$$

40

第二計算機器：胎児が母体のコピー数異常染色体を遺伝していない場合、式 (2) に示されるようにパラメータ β を計算し、m はコピー数異常染色体フラグメントが所在する染色体の有効長さであり、単位は M b である；n は妊婦のコピー数異常染色体フラグメントの長さであり、単位は M b である；cn は妊婦のコピー数異常染色体フラグメントの出現する回数である；f は検出対象妊婦由来の末梢血遊離 DNA に含まれる胎児遊離 DNA の

50

濃度を示し且つ仮に胎児遊離DNAの濃度 f は 50%未満である；

【0090】

【数41】

$$\alpha = \frac{(m-n) \cdot 2 + f \cdot n \cdot 2 + (1-f) \cdot n \cdot cn}{m \cdot 2} \dots (2)$$

補正機器：

【0091】

【数42】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得る、そのうち、

【0092】

【数43】

$$\hat{x}$$

は各染色体の補正前カバレッジを示し、 x' は各染色体の補正後の染色体カバレッジを示す；第二カバレッジ計算機器：各染色体の補正後カバレッジを用いて各染色体の Z_{ane} 値を得る； Z_{ane} 値判定機器： Z_{ane} 値の絶対値が3と等しいかそれより大きいかを判定する；染色体異数性第一認定機器： Z_{ane} 値の絶対値が3と等しいかそれより大きい場合、染色体が異数性を有すると認定する。

【0093】

本出願の前記キットは、コピー数異常染色体フラグメント検索機器、コピー数異常染色体フラグメント認定機器及び補正機器を追加し、本出願は母体染色体上に存在する少なくとも300Kbの領域で且つ該領域中の80%のウィンドウの Z_{cnv} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントであり、これにより本出願の前記キットは信頼できる妊婦コピー数異常染色体フラグメントを検出することができ、さらにこれらのコピー数異常染色体フラグメントを用いてその所在する染色体の Z_{cnv} 値に対して補正を行うことで、妊婦のコピー数異常染色体フラグメントの測定結果の間違いにより偽陰性となる判定を回避できる。補正機器により該コピー数異常染色体フラグメントの各染色体のカバレッジに対する影響に対して補正を行い、これにより本出願の染色体異数性認定機器により染色体異数性の認定結果をより正確にする。本出願の前記キットの補正機器において、パラメータの計算式中の胎児濃度は本分野の通常の計算方法であり、具体的には前記の通りであり、ここでは再度説明しない。

【0094】

説明したいことは、本出願の前記部材、部品は機器の一部として一つの計算端末で稼働することができ、該計算機端末が提供するプロセッサを用いて前記シークエンスデータ検出機器、第一カバレッジ計算機器、固有の配列計算機器、コピー数異常染色体フラグメント検索機器、コピー数異常染色体フラグメント認定機器、第一計算機器、第二計算機器、補正機器、第二カバレッジ計算機器及び染色体異数性認定機器により実現する技術構成であり、明らかなように、該計算機端末はハードウェアで実現する設備であり、プロセッサはプログラムの実行に用いるハードウェアキットである。さらに、本出願が提供する前記各機能サブ機器は移動端末、計算機端末または類似の演算キット中で実行でき、記憶媒体の一部としても記憶することもできる。

【 0 0 9 5 】

本出願の前記キットにおいて、前記第一カバレッジ計算機器は、本分野の通常の計算機器に、シークエンスデータの違いにより適切に調整することで得られるものである。本出願の好ましい実施例において、前記第一カバレッジ計算機器は以下を含む：染色体ウインドウ分割サブ機器：シークエンスデータ中のすべての染色体を同じサイズのウインドウに分割する；第一カバレッジ計算サブ機器：同じサイズのウインドウの形でカバレッジを計算することで、各染色体の補正前カバレッジを得る。本出願の染色体ウインドウ分割サブ機器及び第一カバレッジ計算サブモジュールの第一カバレッジ計算モジュールを同じサイズのウインドウに分割する形で計算を行うことで、比較的安定的なカバレッジを得ることに有利である。

10

【 0 0 9 6 】

本出願のさらなる実施例において、前記染色体ウインドウ分割サブモジュールにおいて、それぞれのウインドウの大きさは100Kbであり、且つ隣り合う二つのウインドウの間の重複度は50%である。それぞれのウインドウを100Kbの大きさに分割する形で計算を行う計算機器は相対的に安定的なカバレッジを得ることができ、もう一方ではウインドウ間のカバレッジを増加させることでコピー数異常染色体フラグメントの精密度を向上させることができ、これにより妊婦コピー数異常染色体フラグメントの検出効率を向上させることができる。

【 0 0 9 7 】

本出願の前記キットにおいて、固有の配列計算機器は通常の計算機器を用いることで得られる。本出願の好ましい実施例において、前記固有の配列計算機器は以下を含む：固有の配列統計ユニット：シークエンスデータ中の各配列のシークエンス深度に基づいて、各ウインドウの固有の配列の数量を統計する；固有の配列のカバレッジ計算ユニット：各染色体のGC含有量及びマッピング率に基づいて各ウインドウの固有の配列の数量について計算を行い、各ウインドウの固有の配列の数量の補正前カバレッジを得る；固有の配列 Z_{cnv} 値計算モジュール：各ウインドウの固有の配列の数量の補正前カバレッジに対して標準化処理を行い、各ウインドウの固有の配列の Z_{cnv} 値を得る。

20

【 0 0 9 8 】

本出願の前記固有の配列計算機器は、まず固有の配列統計ユニットを稼働させることで、シークエンスデータ中の各配列のシークエンス深度に基づいて、各ウインドウの固有の配列の数量を統計でき、それから固有の配列カバレッジ計算モジュールを実行し、各染色体のGC含有量とマッピング率に基づいて各ウインドウの固有の配列の数量に対して補正を行い、各ウインドウの固有の配列の数量の補正值を得て、それから固有の配列 Z_{cnv} 値計算サブユニットを実行し、各ウインドウの固有の配列の数量の補正值に対して標準化処理を行い、各ウインドウの固有の配列の数量の Z_{cnv} 値を得る。前記ユニットは本分野の通常の計算及び補正ユニットに基づいて適切に調整を行うものであり、コピー数異常染色体フラグメント検索装置に対して検索及びコピー数異常染色体フラグメント認定機器について認定する根拠と前提であり、正確に検出対象サンプル中の母体DNAコピー数異常染色体フラグメントの存在に根拠を提供する。

30

【 0 0 9 9 】

本出願の前記キットにおいて、第二カバレッジ計算機器において、 Z_{aneu} 値は

40

【 0 1 0 0 】

【 数 4 4 】

$$Z_{aneu} = \frac{x' - \bar{x}}{S}$$

によって計算され、そのうち、

【 0 1 0 1 】

50

【数 4 5】

$$\bar{x}$$

は LOESS アルゴリズムであり、陰性サンプル群により得られる補正前カバレッジであり、 s は陰性サンプル群における

【0 1 0 2】

【数 4 6】

$$(x' - \bar{x})$$

10

の標準差である。前記式の計算により得られる補正後の Z_{aneu} 値は、より正確に染色体異数性を反映することができ、検出結果はより正確である。

【0 1 0 3】

以下に具体的な実施例を参照しながら本出願の有益な効果について説明する。

【実施例 1】

【0 1 0 4】

本出願の妊婦コピー数異常染色体フラグメント補正における染色体異数性検出に対する補正効果を測定するために、本実施例はポアソン分布に基づいて妊婦のシミュレーションデータを生成し、該シミュレーションデータにおいて、それぞれ単独に 13、18 及び 21 号染色体に定量のコピー数異常染色体フラグメントを添加し、コピー数異常染色体フラグメントのサイズは 0.5 Mb ~ 5 Mb、歩長は 0.25 Mb である。それから、コピー数異常染色体フラグメントを含むシミュレーションデータ中に 3 種類の異なる濃度の健常人 DNA (5%、10%、15%) をそれぞれ混入する。全体のプロセスは異なる胎児濃度下において、異なるコピー数異常染色体フラグメントサイズの 13、18 及び 21 号染色体カバレッジに対する影響をシミュレーションするものであり、さらにこれにより妊婦のコピー数異常染色体フラグメントの染色体異数性検出の補正効果を検出する。すべての計算は、胎児が妊婦コピー数異常染色体フラグメントを遺伝していない仮定下で行われたものである。

20

【0 1 0 5】

測定結果は図 3 A、3 B 及び 3 C を参照する。前記 3 つの図において、横の座標はサンプルが所在する妊婦コピー数異常染色体フラグメントのサイズであり、縦座標はこのサンプルの染色体 Z 値である。図中の実線が示すのは補正前の染色体 Z 値であり、点線は妊婦コピー数異常染色体フラグメントの補正後の染色体カバレッジの計算による Z 値であり、即ち Z_{aneu} 値である。四角形、円形及び三角形はそれぞれ該サンプルの胎児濃度が 5%、10% 及び 15% であることを示している。

30

【0 1 0 6】

図 3 A、3 B 及び 3 C から明らかに分かるように、直接染色体カバレッジで Z 値を計算する場合、妊婦のコピー数異常染色体フラグメントの向上により、サンプルの Z 値も大きくなる。21 号染色体を例として、10% 胎児濃度下において、妊婦に 3 Mb の 21 号染色体上の重複が存在するであれば、仮に胎児が 21 トリソミー症候群患児でなくとも、以前のカバレッジで計算された Z 値は 3 より大きく、陽性と判定されてしまう。図中の点線は、即ち本出願の方法で補正した後のカバレッジで計算された染色体の Z 値であり、即ち Z_{aneu} 値であり、いずれも 0 のベースライン附近に安定している。これは、各種状況下において、本出願の妊婦コピー数異常染色体フラグメントを用いたものにより補正した染色体異数倍の検出方法は極めて有効である。

40

【0 1 0 7】

さらに本出願の思想により提供した検出方法、検出装置及びキットは実際の患者例のサンプル中において染色体異数倍を測定する効果であり、さらに本出願の検出方法、検出装置及びキットを用いてそれぞれ以下の患者例サンプルを検出することができ、具体的には実施例 2 及び実施例 3 である。

50

【実施例 2】

【0108】

6615 件の検出対象妊婦由来の末梢血遊離 DNA サンプルに対してハイスループットシーケンシングを行い、各サンプルにすべての染色体が含まれているシーケンスデータを得た。

【0109】

各サンプルのシーケンスデータ中の各配列のシーケンス深度に基づいて、各ウインドウの固有の配列の数量を統計する。各染色体の GC 含有量及びマッピング率に基づいて各ウインドウの固有の配列の数量に対して補正を行い、各ウインドウの固有の配列の数量の補正值を得る。各ウインドウの固有の配列の数量の補正值に対して標準化処理を行い、各ウインドウの固有の配列の数量の Z_{cnv} 値を得て、さらに Z_{cnv} 値の大小に基づいて検出対象妊婦がコピー数異常染色体フラグメントを有するかについて判定する。シーケンスデータ中には 300 Kb 以上のフラグメントが存在し、且つ 300 Kb 以上のフラグメント中のすべてのウインドウのうち 80 % 以上のウインドウの固有の配列の数量の Z_{cnv} 値は 4 と等しいかそれより大きく、または -4 と等しいかそれより小さい場合、300 Kb 以上のフラグメントは検出対象妊婦のコピー数異常染色体フラグメントであると考えられる。

10

【0110】

検出対象妊婦のコピー数異常染色体フラグメントを用いて各染色体の補正前カバレッジの影響を利用して、さらに

20

【0111】

【数 47】

$$x' = \frac{\hat{x}}{\alpha}$$

を利用して、各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得る；そのうち、

【0112】

30

【数 48】

$$\hat{x}$$

は各染色体の補正前カバレッジであり、 x' は各染色体の補正後染色体カバレッジである；そのうち、検出対象妊婦のコピー数異常染色体フラグメントの各染色体に対する補正前カバレッジの影響は、パラメータ α を用いて式 (1) または (2) によって計算される；

各染色体の補正後カバレッジを用いて各染色体の Z_{aneu} 値を計算し、具体的には式

【0113】

【数 49】

40

$$Z_{aneu} = \frac{x' - \bar{x}}{s}$$

の計算により Z_{aneu} 値を得て、さらに Z_{aneu} 値の絶対値が 3 と等しいかそれより大きいことで染色体が異数性を有するかを判定する； Z_{aneu} 値の絶対値が 3 と等しいかそれより大きい場合、染色体は異数性を有する； Z_{aneu} 値の絶対値が 3 未満の場合、染色体は異数性を有しない；

本出願の前記検出方法によりサンプル EK01875 及び BD01462 の 21 号染色

50

体上に存在する妊婦コピー数異常染色体フラグメントを発見し、さらにこの二つのサンプルを以前の陽性の結果から陰性の結果に補正し、具体的な結果は図4を参照すること。

【0114】

図4の左図（カラー図面参照）は現在の検出方法によって検出されるすべてのサンプルの21号染色体のZ値統計図であり、これから分かるように陰性サンプルのZ値はほとんど3未満であり、正規分布に近い。図中の円形はサンプルEK01875であり、Z値は4.66である。三角形はサンプルBD01462であり、Z値は3.87である。

図4における右図は本出願の検出方法によって検出することで得られる21号染色体のZ値統計図であり、この場合のサンプルEK01875は $Z_{aneu} = 2.36$ で、サンプルBD01462は $Z_{aneu} = 1.83$ である。

10

【実施例3】

【0115】

本出願の染色体異数性の検出装置を用いて前記サンプルの添加を行い（サンプルEK01875、妊婦年齢29歳、妊娠約18週）、該装置は以下を含む：

シークエンスデータ検出モジュール：検出対象妊婦由来の末梢血遊離DNAに対してハイスループットシーケンシングを行い、これによりすべての染色体のシークエンスデータを得る；第一カバレッジ計算モジュール：シークエンスデータ中のすべての染色体のウィンドウに分割する形でカバレッジを計算することで、各染色体の補正前カバレッジを得る； Z_{cnv} 値計算モジュール：検出対象妊婦の各ウィンドウ中の固有の配列の数量の Z_{cnv} 値に対して計算を行う；コピー数異常染色体フラグメント検索モジュール：シークエンスデータ中において300Kb以上のフラグメントを検索し、且つ300Kb以上のフラグメントにおいて、すべてのウィンドウのうち80%以上のウィンドウ中の染色体フラグメントが、 Z_{cnv} が4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントである。

20

【0116】

コピー数異常染色体フラグメント認定モジュール：シークエンスデータ中から検索することで得られる300Kb以上のフラグメントで、且つすべてのウィンドウのうち80%以上のウィンドウ中の染色体フラグメントの Z_{cnv} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントを、検出対象妊婦のコピー数異常染色体フラグメントと認定する。

30

第一計算モジュール：胎児が母体のコピー数異常染色体フラグメントを遺伝している場合、式(1)に示す計算式でパラメータを計算する；

第二計算モジュール：胎児が母体のコピー数異常の染色体を遺伝していない場合、式(2)に示す計算式でパラメータを計算する；

補正モジュール：

【0117】

【数50】

$$x' = \frac{\hat{x}}{\alpha}$$

40

を用いて、各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得る；

第二カバレッジ計算モジュール：各染色体の補正後カバレッジで各染色体の Z_{aneu} 値を計算する；

Z_{aneu} 値判定モジュール： Z_{aneu} 値の絶対値が3と等しいかそれより大きいかの判定に用いる；

染色体異数性認定モジュール： Z_{aneu} 値の絶対値が3と等しいかそれより大きい場合、染色体が異数性を有すると認定する。

50

【 0 1 1 8 】

本出願の前記染色体異数性を検出する装置を用いて分析した後、妊婦は21号染色体上に850 Kbの重複が存在することを検出した。図5に示すように、コピー数に重複が存在する領域はそれぞれ500 Kbの21q22.11(32361194bp~32861193bp)及び350 Kbの21q22.12(37261194bp~37611193bp)であり、コピー数はいずれも3である。

【 0 1 1 9 】

それから、従来技術中のAffymetrix CytoScan 750k SNPチップを用いて該妊婦の染色体コピー異常フラグメントの結果に対して検証を行い、同じように、21q22.11(32399114bp~32811202bp)領域と21q22.12(37292432bp~37602701bp)において、コピー数は3であると検出した。

10

【 0 1 2 0 】

これから分かるように、チップ検出結果の位置と本出願の装置により検出される位置はほとんど百パーセントマッチする。本出願の装置において、妊婦コピー数異常染色体フラグメントの該染色体カバレッジ計算に対する影響のパラメータ値は1.012であり、染色体が異数性を有するかを示すZ値は元の4.66から2.36に補正され、これにより判定結果は陰性となる。

【 実施例 4 】

【 0 1 2 1 】

20

本出願の染色体異数性を検出するキットを用いて前記サンプル(サンプルBD01462、妊婦年齢24歳、妊娠約24週)を添加して検出を行い、該キットは以下を含む：

シーケンスデータ検出試薬及び機器：検出対象妊婦由来の末梢血遊離DNAに対してハイスループットシーケンシングを行い、これによりすべての染色体を含むシーケンスデータを得る；

第一カバレッジ計算機器：シーケンスデータ中のすべての染色体をウィンドウに分割する形でカバレッジを計算し、これにより各染色体の補正前カバレッジを得る；

固有の配列計算機器：検出対象妊婦の各ウィンドウにおける固有の配列の数量の Z_{cn} について計算を行う；

コピー数異常染色体フラグメント検索機器：シーケンスデータ中で300 Kb以上のフラグメントを検索し、且つ300 Kb以上のフラグメントにおいて、すべてのウィンドウのうち80%以上のウィンドウ中の染色体フラグメントの Z_{cn} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントである。

30

【 0 1 2 2 】

コピー数異常染色体フラグメント認定機器：シーケンスデータ中から検索により得られる300 Kb以上のフラグメントで且つすべてのウィンドウのうち80%以上のウィンドウ中の染色体フラグメントの Z_{cn} 値が4と等しいかそれより大きく、または-4と等しいかそれより小さいフラグメントを検出対象妊婦のコピー数異常染色体フラグメントと認定する。

【 0 1 2 3 】

40

第一計算機器：胎児が母体のコピー数異常染色体フラグメント遺伝している場合、式(1)に示す計算式でパラメータを計算する。

第二計算機器：胎児が母体のコピー数異常染色体フラグメントの染色体を遺伝していない場合、式(2)に示す計算式に従ってパラメータを計算する。

【 0 1 2 4 】

補正機器：

【 0 1 2 5 】

【数 5 1】

$$x' = \frac{\hat{x}}{\alpha}$$

を用いて、各染色体の補正前カバレッジに対して補正を行い、各染色体の補正後カバレッジを得る；

第二カバレッジ計算機器：各染色体の補正後カバレッジを用いて各染色体の Z_{aneu} 値を計算する；

Z_{aneu} 値判定機器： Z_{aneu} 値の絶対値が 3 と等しいかそれより大きいかを判定する；

染色体異数性認定機器： Z_{aneu} 値の絶対値が 3 と等しいかそれより大きい場合、染色体が異数性を有すると判定する。

【0 1 2 6】

本出願の前記検出キットを用いて検出分析した後、図 6 に示すように、妊婦が 2 1 号染色体上に全部で 7 0 0 K b の重複が存在し、領域は 2 1 q 2 3 . 1 (2 8 9 1 1 1 9 4 b p ~ 2 9 6 1 1 9 3 0) であり、コピー数は 3 であると検出した。

【0 1 2 7】

同じように、A f f y m e t r i x C y t o S c a n 7 5 0 k S N P チップ検証結果が示すように、2 1 q 2 1 . 3 (2 8 9 7 3 7 9 2 b p ~ 2 9 5 4 2 4 0 0) に重複が存在する。

【0 1 2 8】

検出されるコピー数は 4 であるが、本出願の検出結果と異なって、該結果の位置と本出願のキットによって検出される位置はほとんど百パーセントマッチするものであり、同じように本出願の検出方法の正確性を示している。本出願のキット中の妊婦コピー数異常染色体フラグメントの該染色体のカバレッジ計算の影響パラメータ の値は 1 . 0 0 9 であり、染色体が異数性を有するかを示す Z 値は元の 3 . 8 7 から 1 . 8 3 に補正され、このため判定結果は陰性である。

【0 1 2 9】

以上の記載から分かるように、本出願の前記実施例は以下の技術効果を実現している：本出願は妊婦自身のコピー数異常染色体フラグメントの染色体異数性を計算する際の影響を考慮して、従来技術におけるシーケンスタデータ中の母体中のコピー数異常染色体フラグメントを直接除去して考慮しないという考え方に基づいて、創造的に母体染色体上に存在する特定の大きさのコピー数異常染色体フラグメントの染色体異数性に対する影響をパラメータ で示し、さらにパラメータ により各染色体のカバレッジに対して補正を行い、これによりコピー数異常染色体フラグメントの異数性に対する判定の影響を低減できるのであって、該コピー数異常染色体フラグメントの存在を無視することではない。本出願の方法によって検出することで得られる染色体異数性の結果はより正確である。

【0 1 3 0】

本出願の方法、装置またはキットは、ほとんど妊婦のコピー数異常染色体フラグメントの影響を受けない N I P T 胎児染色体異数性に影響されない検出方法を提供し、これにより検出の制度を向上させ、大規模使用に適している。

【0 1 3 1】

明らかなように、当業者は以下のように理解すべきである。前記の本出願のモジュール、部品またステップは通用の計算装置によって実現され、これらは単一の計算装置に集中することもできるが、複数の計算装置からなるネットワークに分布することもでき、また、これらは計算装置の実行できるプログラムコードによっても実現することもでき、このため、これらを記憶装置中に記憶して計算装置によって実行でき、またはこれらをそれぞれの集積回路モジュールにして、またはこれら中の複数のモジュールまたはステップを単

10

20

30

40

50

一の集積回路モジュールにすることで実現することもできる。このようにすることで、本出願はいかなる特定のハードウェア及びソフトウェアの組み合わせに制限されるものではない。

【 0 1 3 2 】

以上は本出願の好ましい実施例に過ぎず、本出願を制限するものではない。当業者にとって、本出願は各種変更及び変化を有することができる。本出願の趣旨の原則に背かないであれば、いかなる修正、均等的な置換、改善も、いずれも本出願の保護範囲に含まれるものである。

【 図 1 】

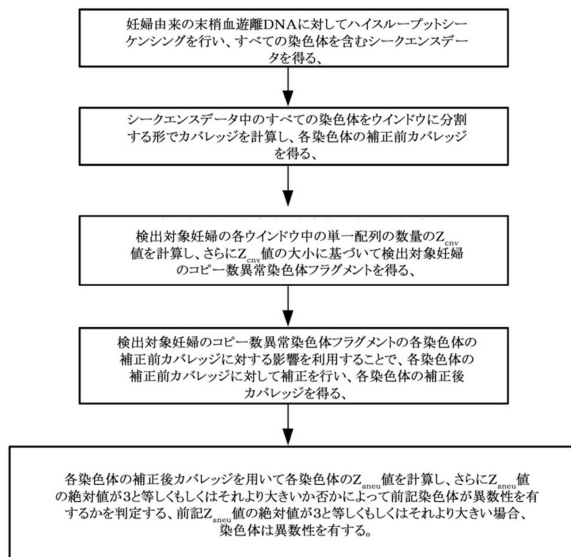


図1

【 図 2 】

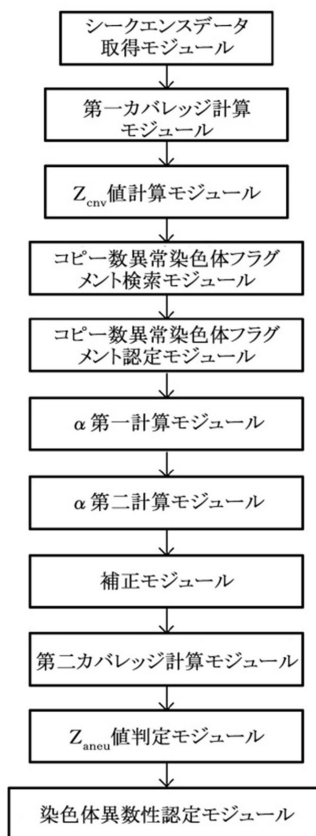


図2

【図3】

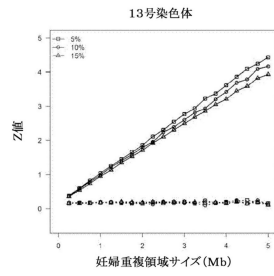


図3A

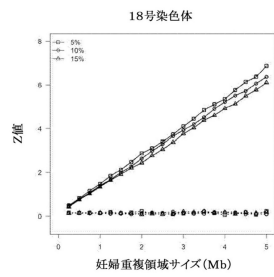


図3B

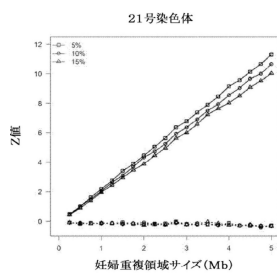


図3C

【図6】

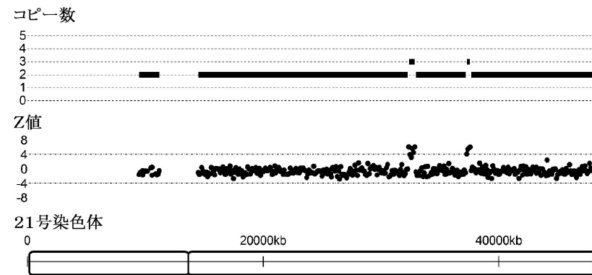


図6

【図4】

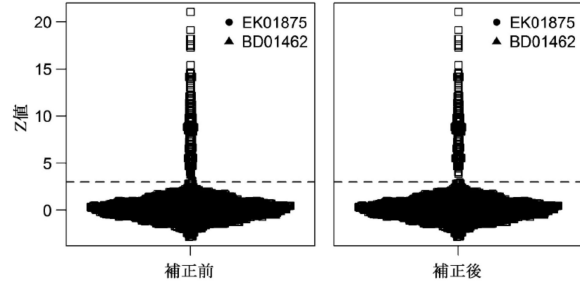


図4

【図5】

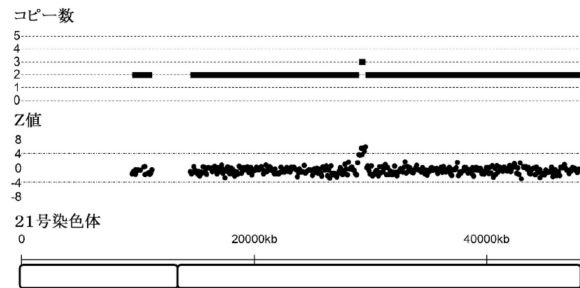


図5

フロントページの続き

(73)特許権者 519181180

チョージャン アンノロード バイオ-テクノロジー カンパニー リミテッド
 Zhejiang Annoroad Bio-Technology Co., Ltd.
 中華人民共和国 チョージャン プロヴィンス イウ シティ チョウジアン サブディストリクト
 ガオシン ロード ナンバー10 スタンダード ワークショップ ビルディング 2 1/
 エフ
 1/F Standard Workshop Building 2, No 10 Gaoxin Rd, Choujiang Subdistrict, Yiwu City, Zhejiang Province CHINA

(73)特許権者 517387498

アンノロード ジーン テクノロジー (ペキン) カンパニー リミテッド
 Annoroad Gene Technology (Beijing) Co., Ltd
 中華人民共和国 100176 ペキン、ダーシーン ディストリクト、エコノミック-テクノロ
 ジカル ディベロップメント エリア、ケチュアン 6 ロード、ヤード 88、ビルディング
 B2

(74)代理人 110000578

名古屋国際特許業務法人

(74)代理人 100120293

弁理士 中谷 智子

(72)発明者 チェン チョンジアン

中華人民共和国 100176、ペキン エコノミック-テクノロジカル ディベロップメント
 エリア、ケチュアン 6 ロード、ヤード 88、ビルディング B2

(72)発明者 リアン ジュンピン

中華人民共和国 100176、ペキン エコノミック-テクノロジカル ディベロップメント
 エリア、ケチュアン 6 ロード、ヤード 88、ビルディング B2

(72)発明者 シュアン ザオリン

中華人民共和国 100176、ペキン エコノミック-テクノロジカル ディベロップメント
 エリア、ケチュアン 6 ロード、ヤード 88、ビルディング B2

(72)発明者 リ ダウェイ

中華人民共和国 100176、ペキン エコノミック-テクノロジカル ディベロップメント
 エリア、ケチュアン 6 ロード、ヤード 88、ビルディング B2

審査官 北村 悠美子

(56)参考文献 Proc. Natl. Acad. Sci. USA, 2008年, Vol.105, No.51, p.20458-20463

Prenatal Diagnosis, 2013年, Vol.33, p.602-608

Nucleic Acids Research, 2014年, Vol.42, No.5, e31

European Journal of Human Genetics, 2015年 1月, Vol.23, p.1286-1293

Plos One, 2010年, Vol.5, No.5, e10439

(58)調査した分野(Int.Cl., DB名)

C12Q 1/00 - 3/00

C12N 15/00 - 15/90

CAplus/MEDLINE/EMBASE/BIOSIS/WPIDS(STN)

PubMed